DEEP LEARNING FOR DRUG DISCOVERY

DANIIL POLYKOVSKIY

daniil@insilico.com

INSILICO MEDICINE

GOALS

1. What is drug discovery and where is machine learning there?

GOALS

1. What is drug discovery and where is machine learning there?

2. What are representation specific-methods (how to generate a graph?)

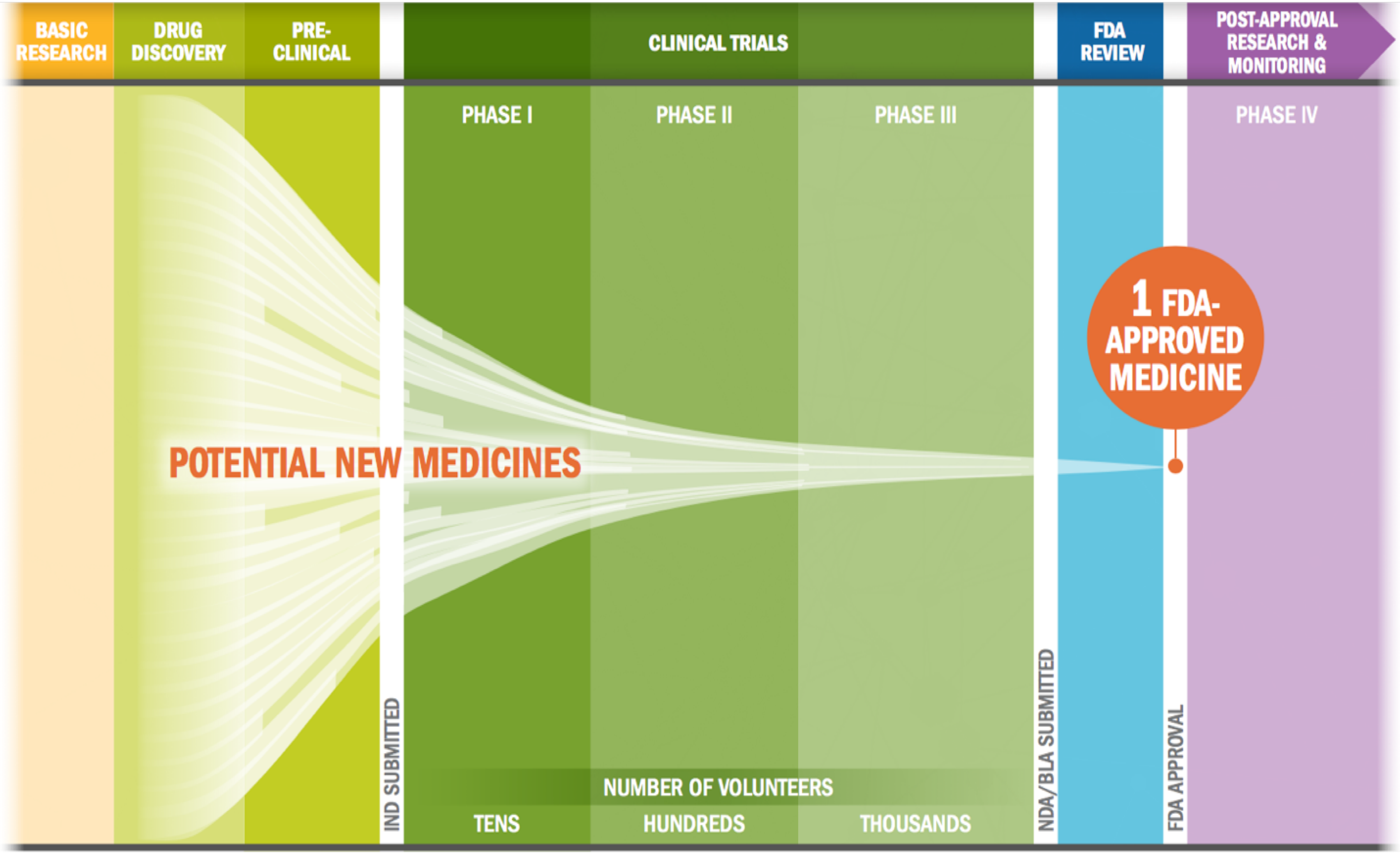
GOALS

1. What is drug discovery and where is machine learning there?

2. What are representation specific-methods (how to generate a graph?)

3. What are approaches for generating useful molecules?

BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT

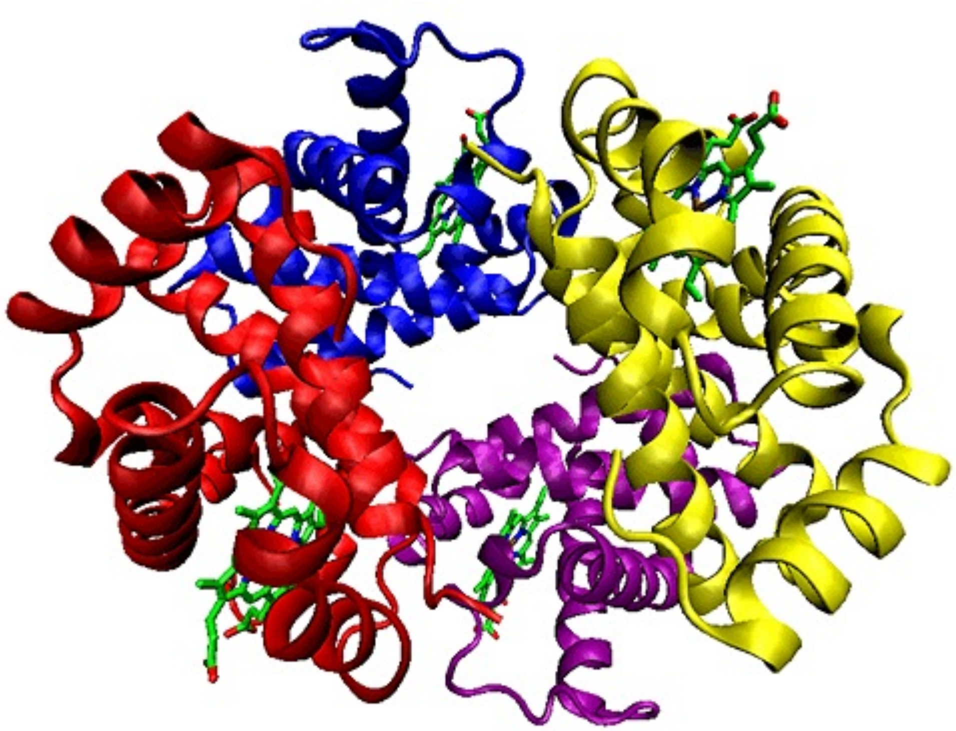


Key: IND: Investigational New Drug Application, NDA: New Drug Application, BLA: Biologics License Application

Source: Biopharmaceutical Research & Development, PRMA http://phrma-docs.phrma.org/sites/default/files/pdf/rd_brochure_022307.pdf

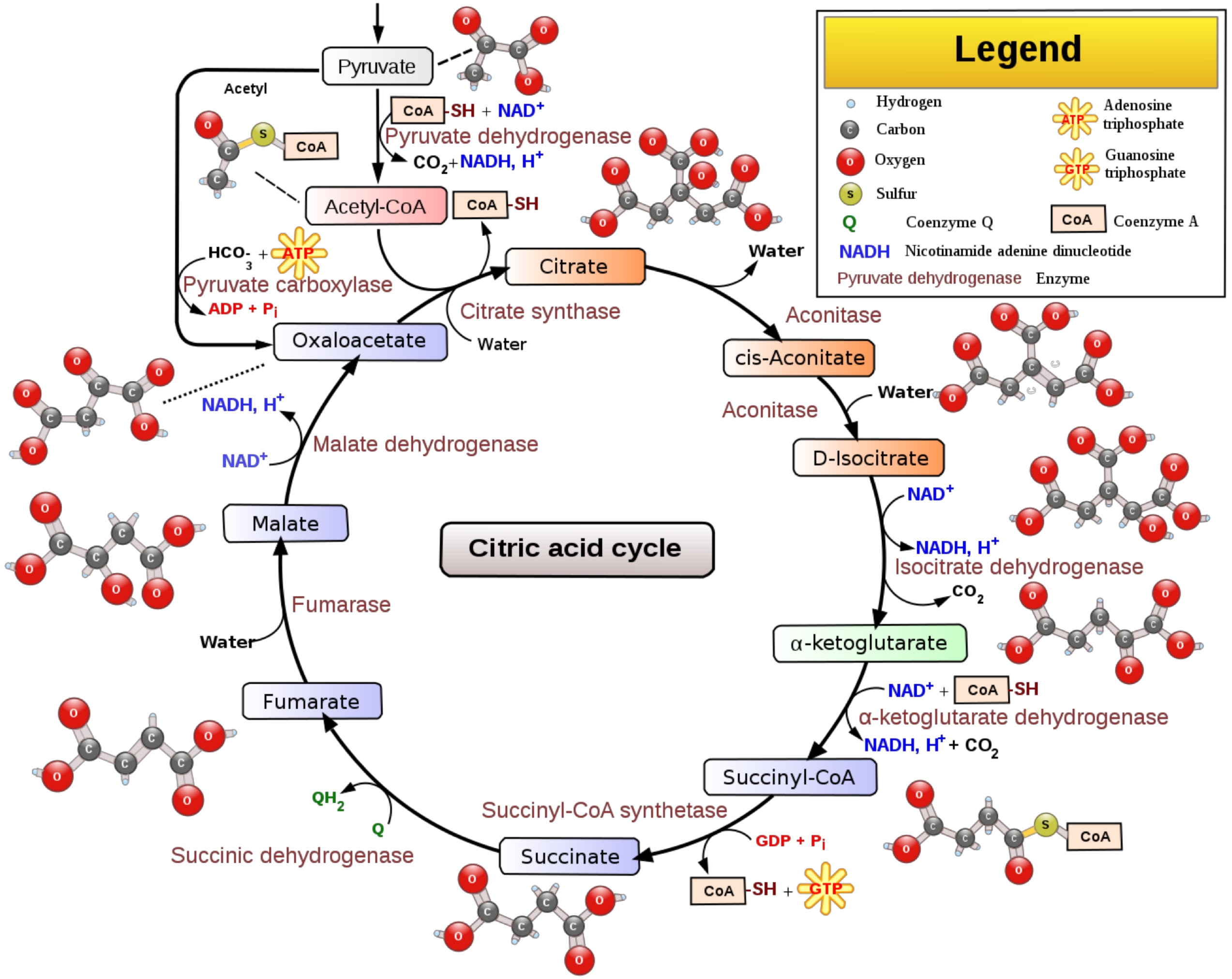
PROTEINS

- Large molecules (chains of amino acids)
- E.g. accelerate specific reactions



PROTEINS

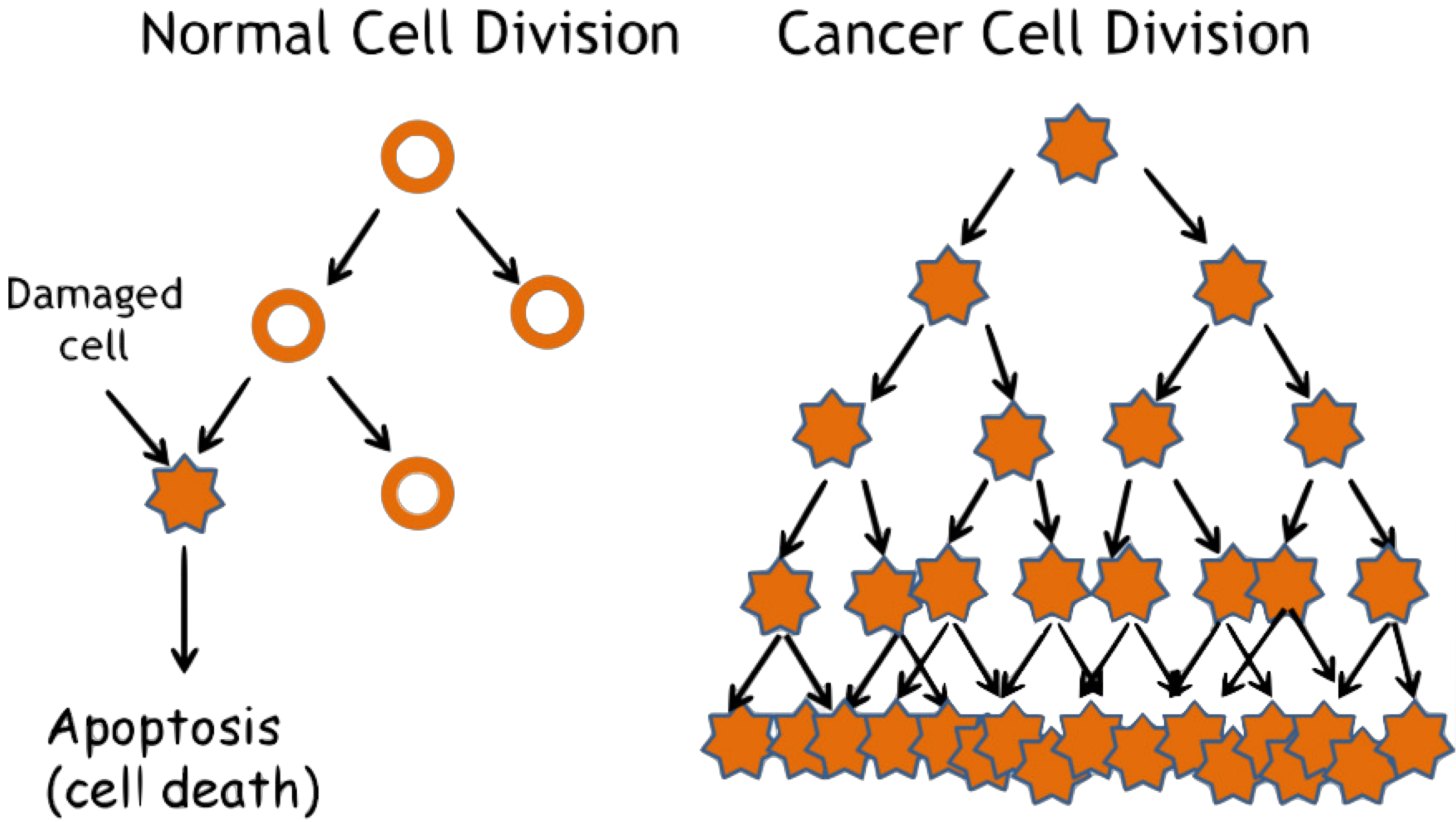
- Healthy proteins work well together



PROTEINS

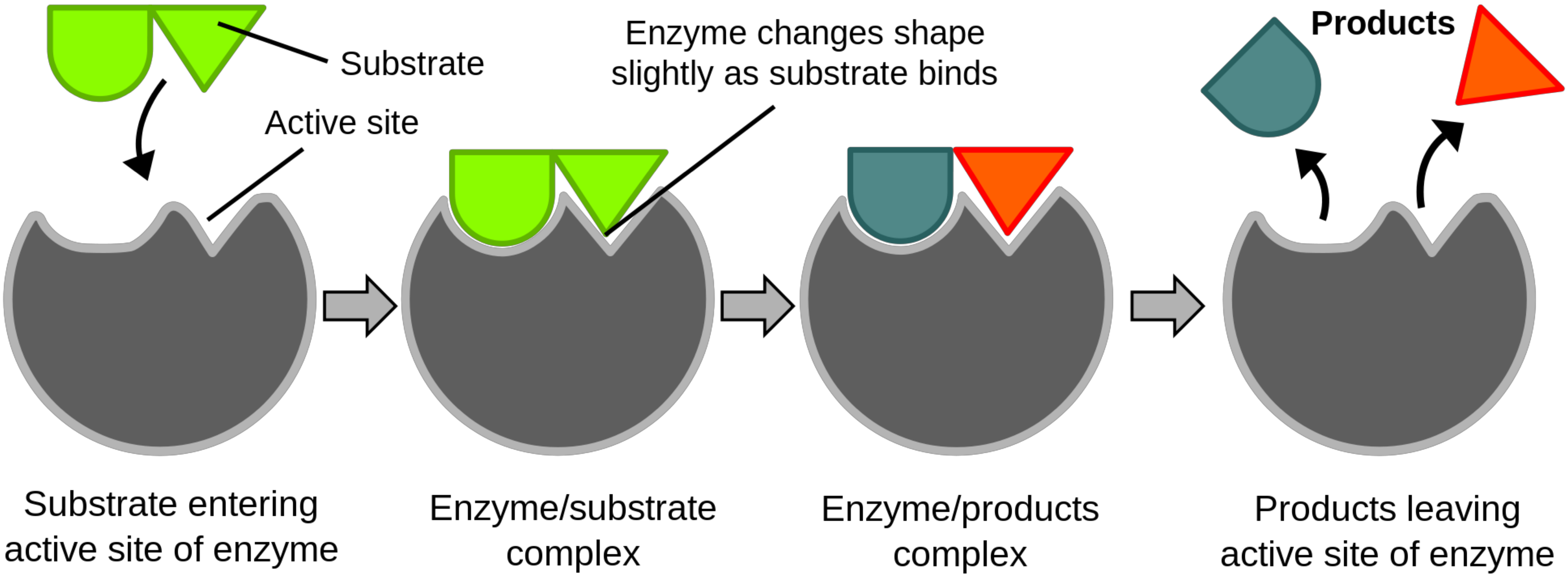
- Malicious proteins may be constantly active
- What can go wrong:
  - Constantly activated cell division
  - Resistance to apoptosis

  = Cancer

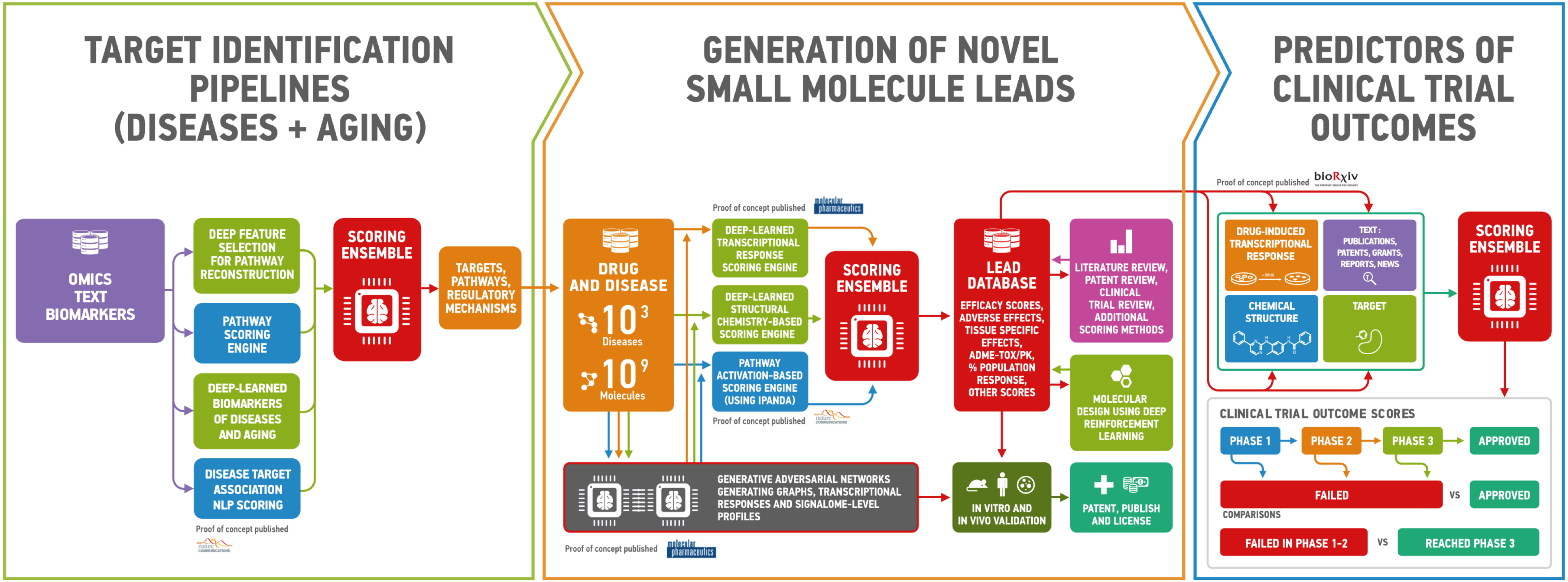


ACTION

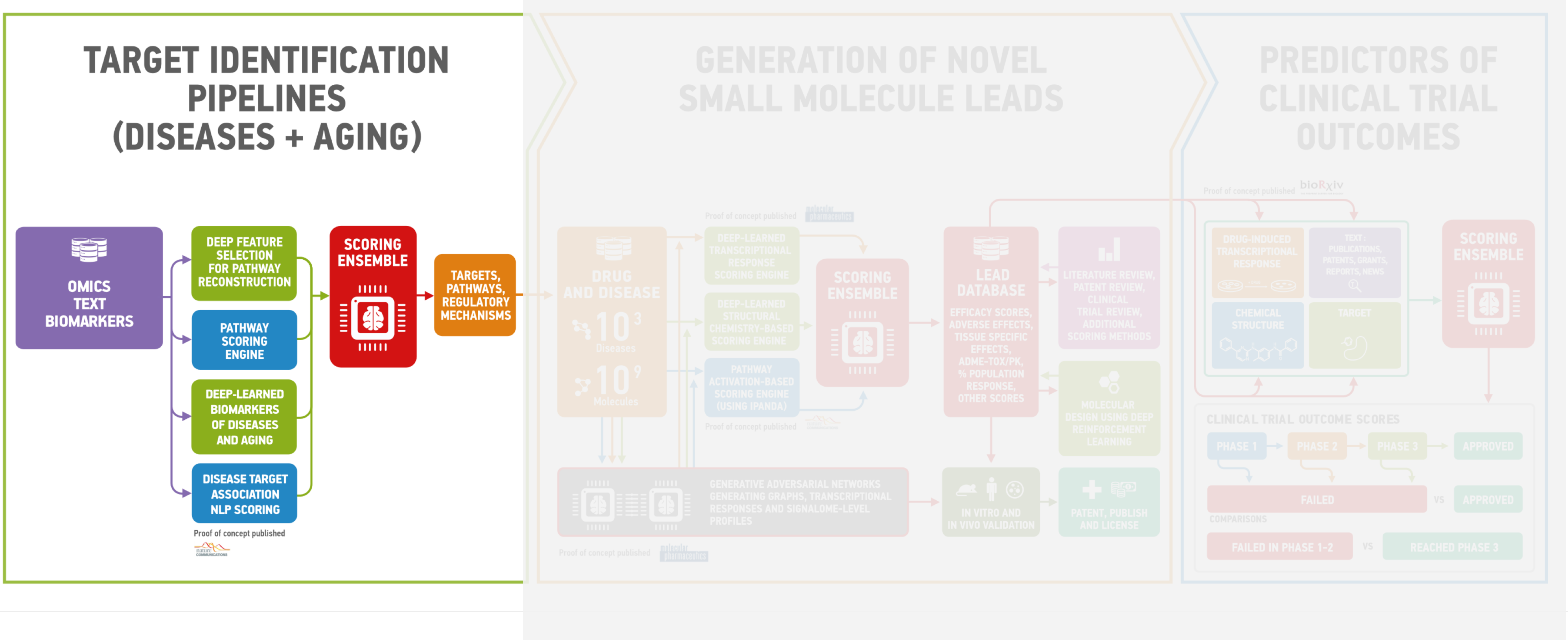
- Protein transforms a substrate into products
- Should have some regulation mechanism…



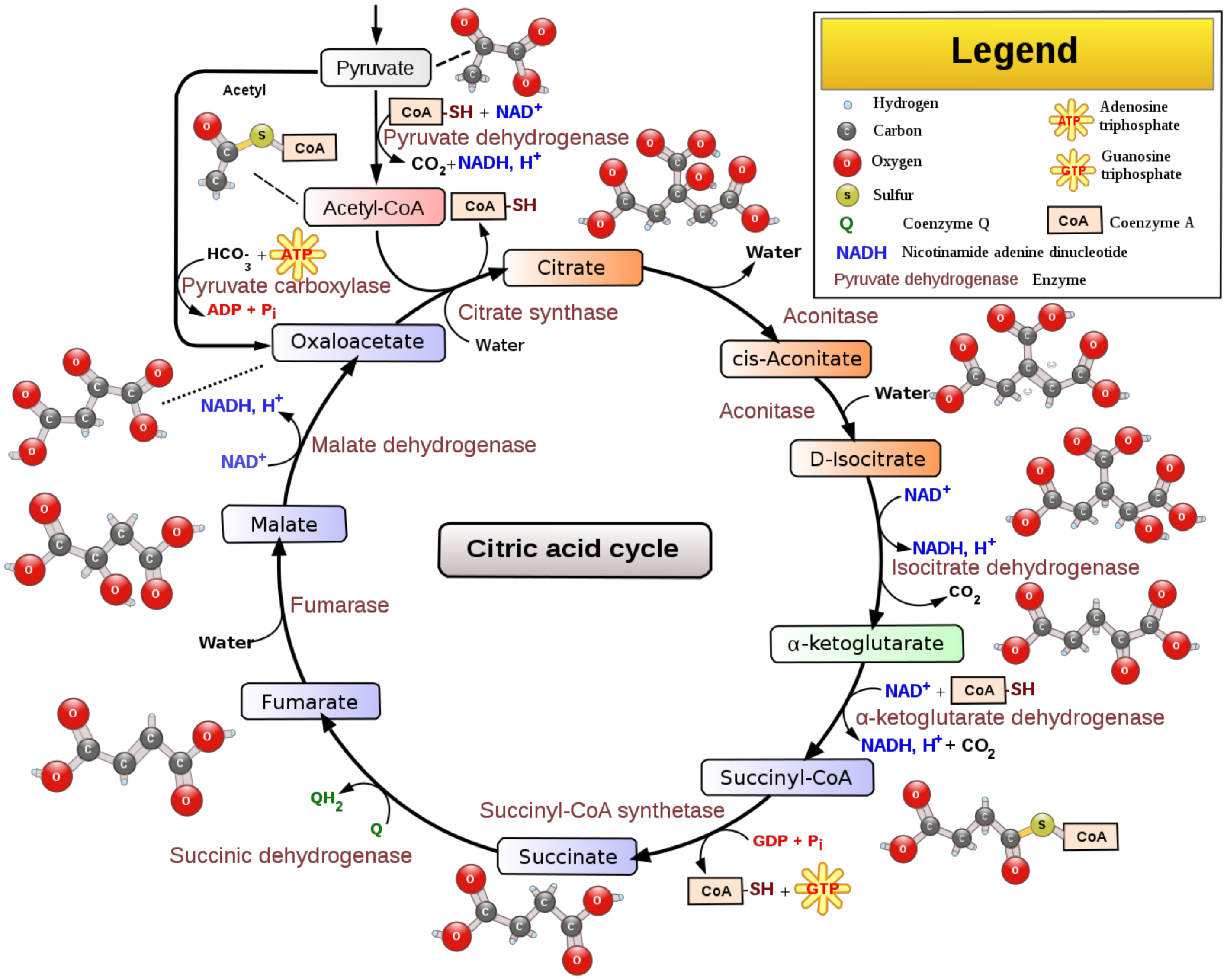
FULL DRUG DISCOVERY PIPELINE



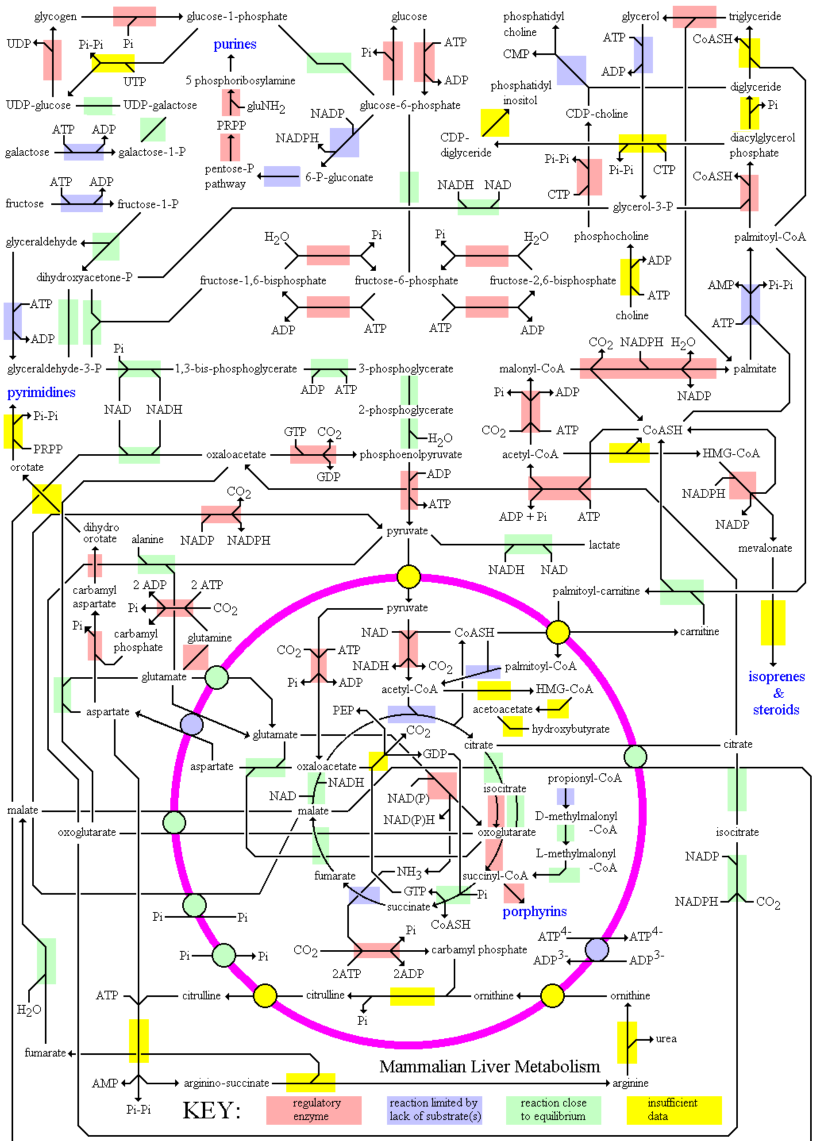
FULL DRUG DISCOVERY PIPELINE



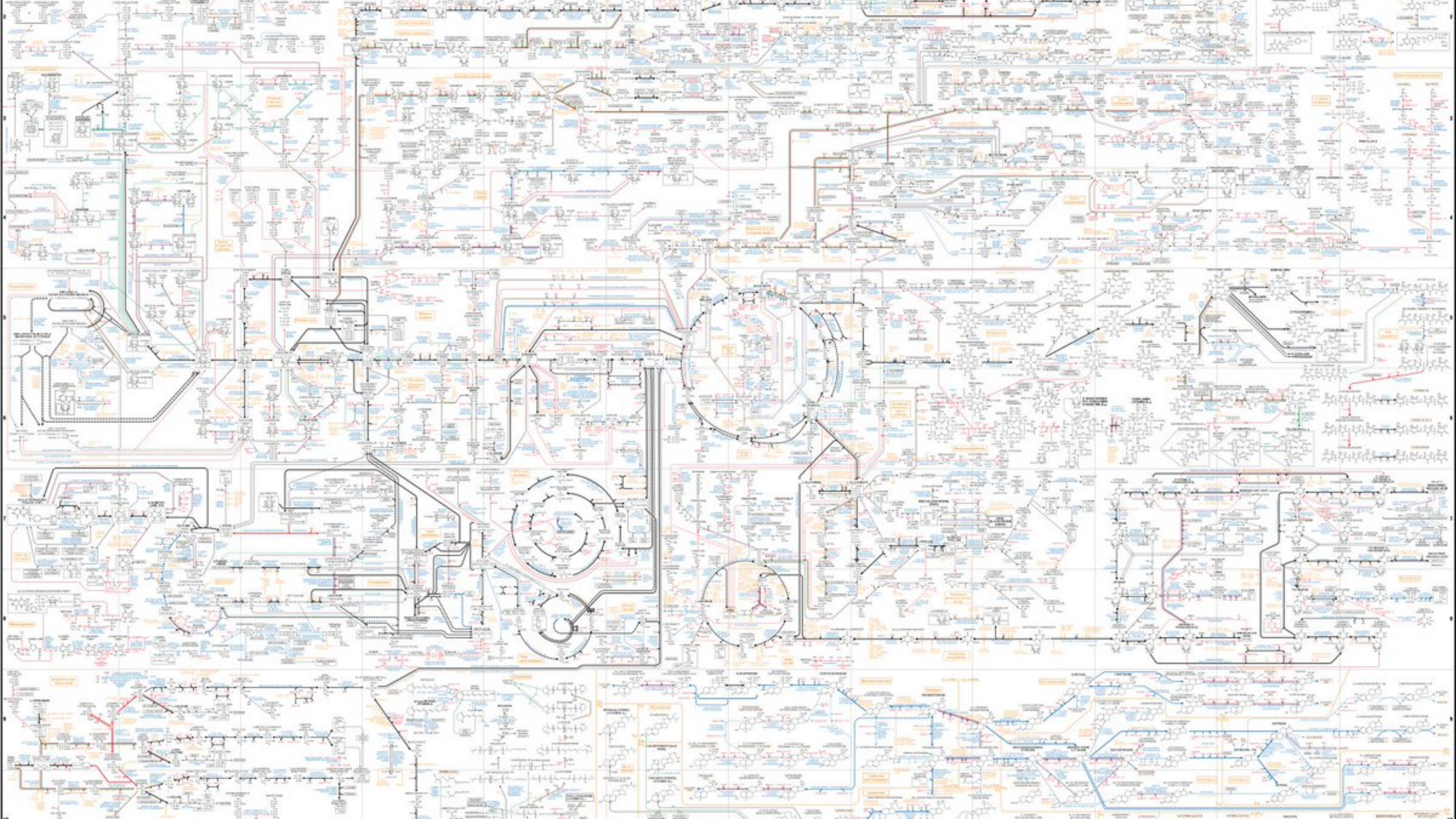
TARGET ID


Legend
Hydrogen
Carbon
Oxygen
Sulfur
Coenzyme Q
Nicotinamide adenine dinucleotide
Pyruvate dehydrogenase Enzyme
Adenosine triphosphate
Guanosine triphosphate
Coenzyme A
Citric acid cycle
Pyruvate
Acetyl
CoA-SH + NAD+
Pyruvate dehydrogenase
CO2+NADH, H+
Acetyl-CoA
CoA-SH
HCO3- + ATP
Pyruvate carboxylase
ADP + Pi
Citrate
Water
Citrate synthase
Water
Oxaloacetate
Aconitase
cis-Aconitate
Water
Aconitase
D-Isocitrate
NAD+
NADH, H+
Isocitrate dehydrogenase
CO2
α-ketoglutarate
NAD+ + CoA-SH
α-ketoglutarate dehydrogenase
NADH, H+ + CO2
Succinyl-CoA
Succinyl-CoA synthetase
GDP + Pi
CoA-SH + GTP
Succinate
QH2
Q
Succinic dehydrogenase
Fumarate
Water
Fumarase
Malate
NADH, H+
NAD+
Malate dehydrogenase

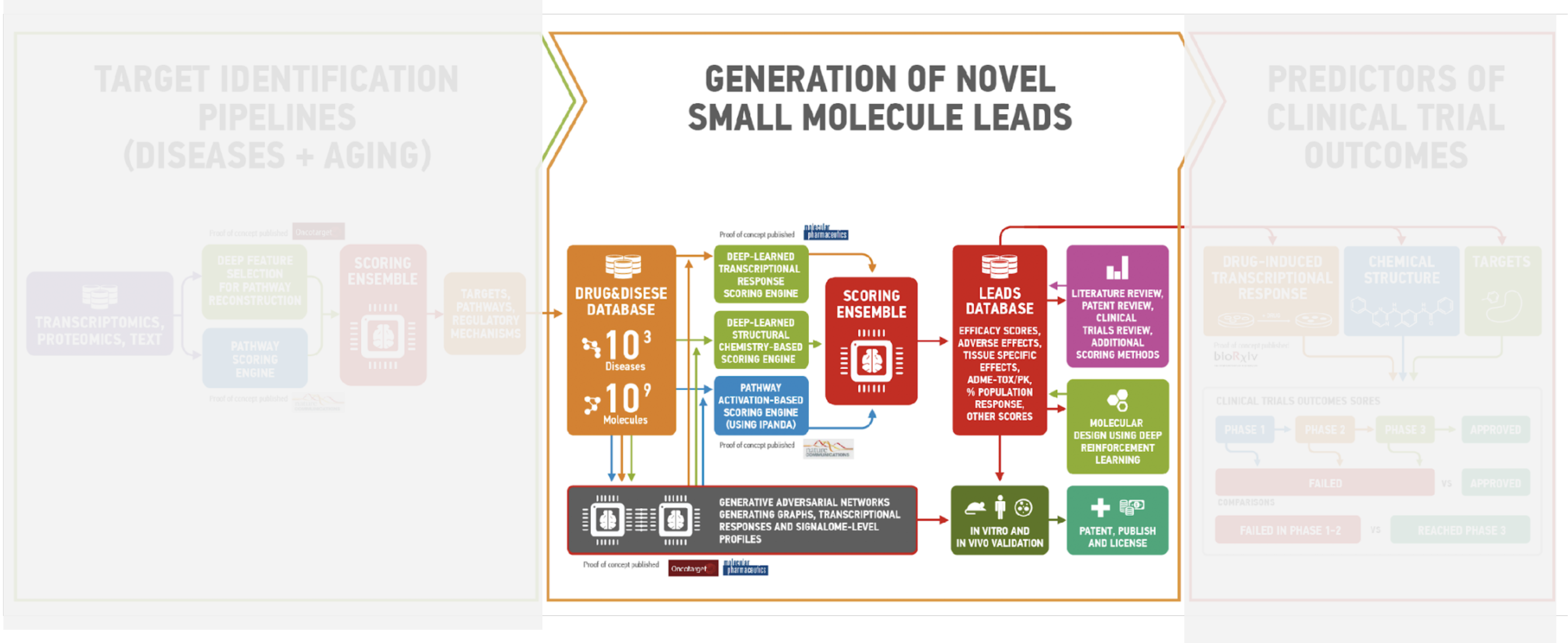
TARGET ID



Mammalian Liver Metabolism
KEY:
regulatory enzyme
reaction limited by lack of substrate(s)
reaction close to equilibrium
insufficient data



FULL DRUG DISCOVERY PIPELINE



GENERATIVE MODELS

MACHINE LEARNING MODELS

Discriminative: Who painted a picture?

Generative: Paint a new picture

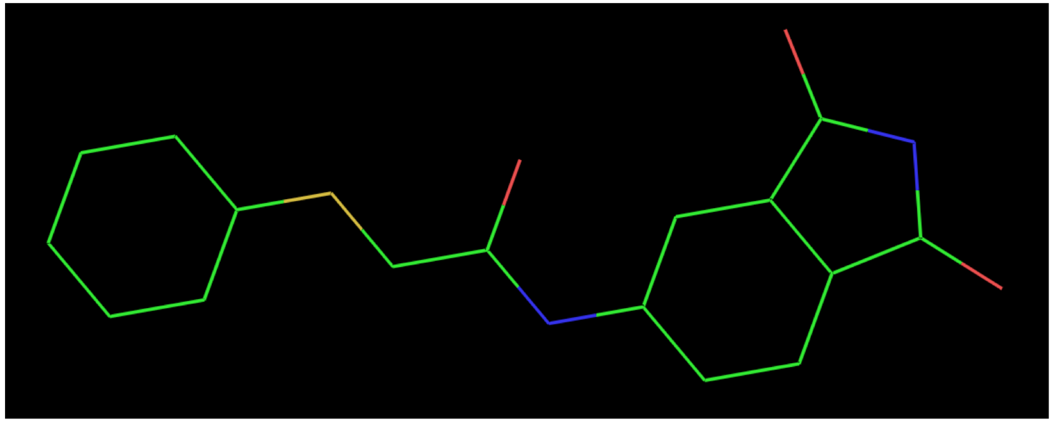
Van Gogh, Starry night

Daniil Polykovskiy, Late at night

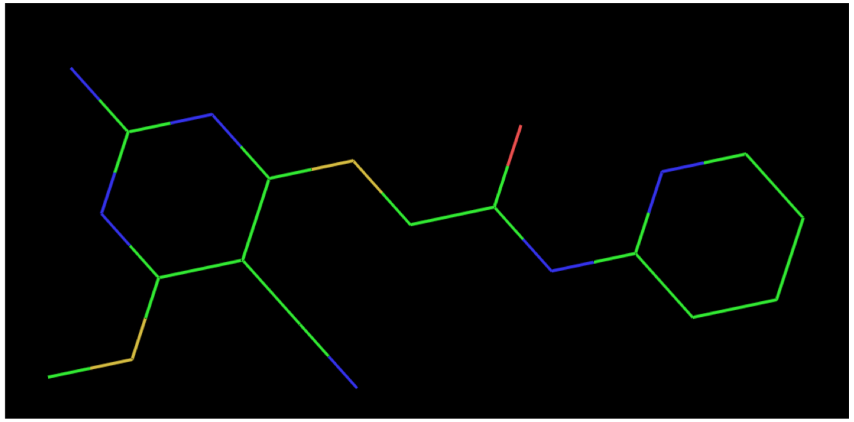
MACHINE LEARNING MODELS

Discriminative: Is a molecule active?

Generative: Create a new structure



Known active molecules



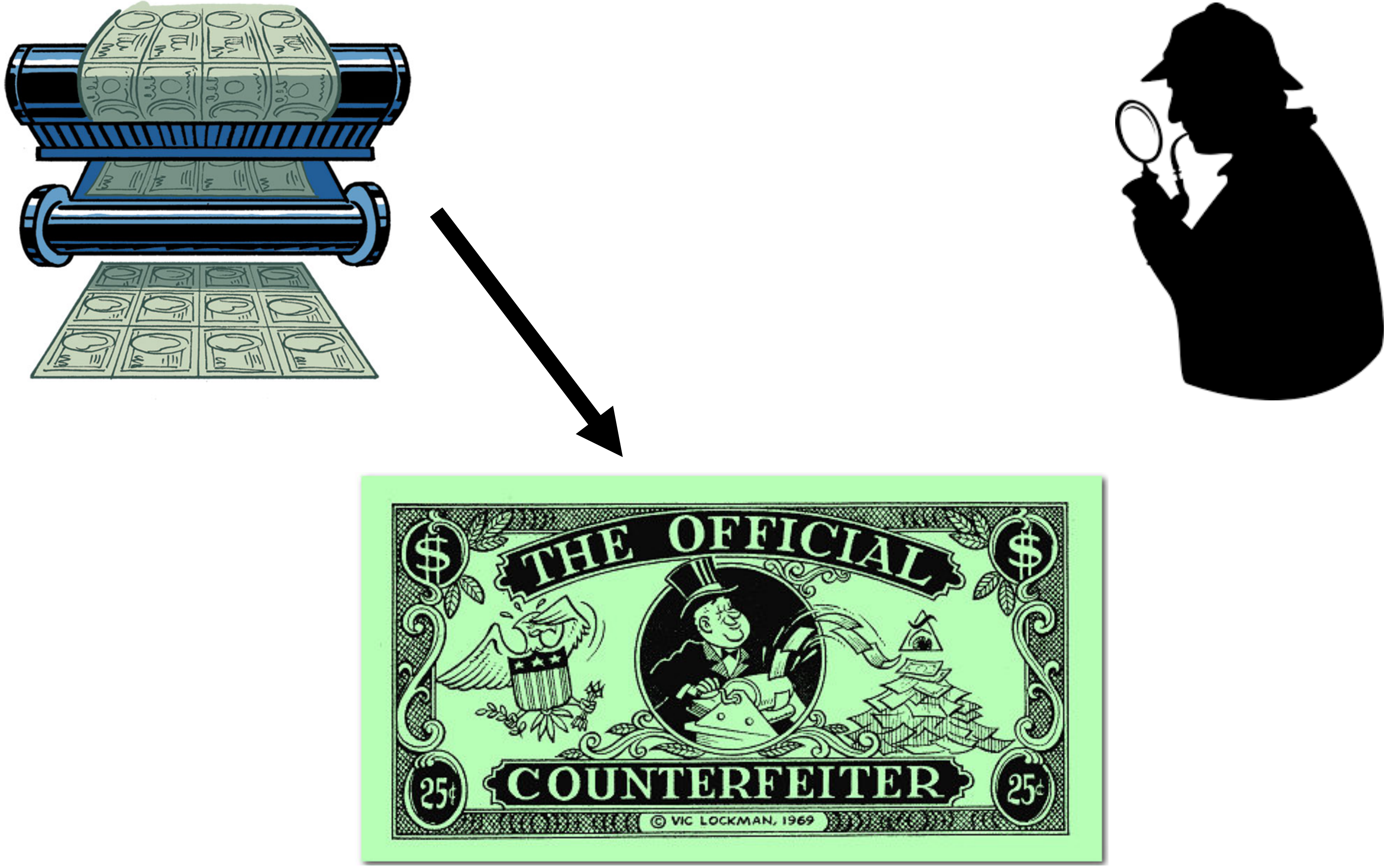
New active molecule

IMAGE GENERATION SOTA

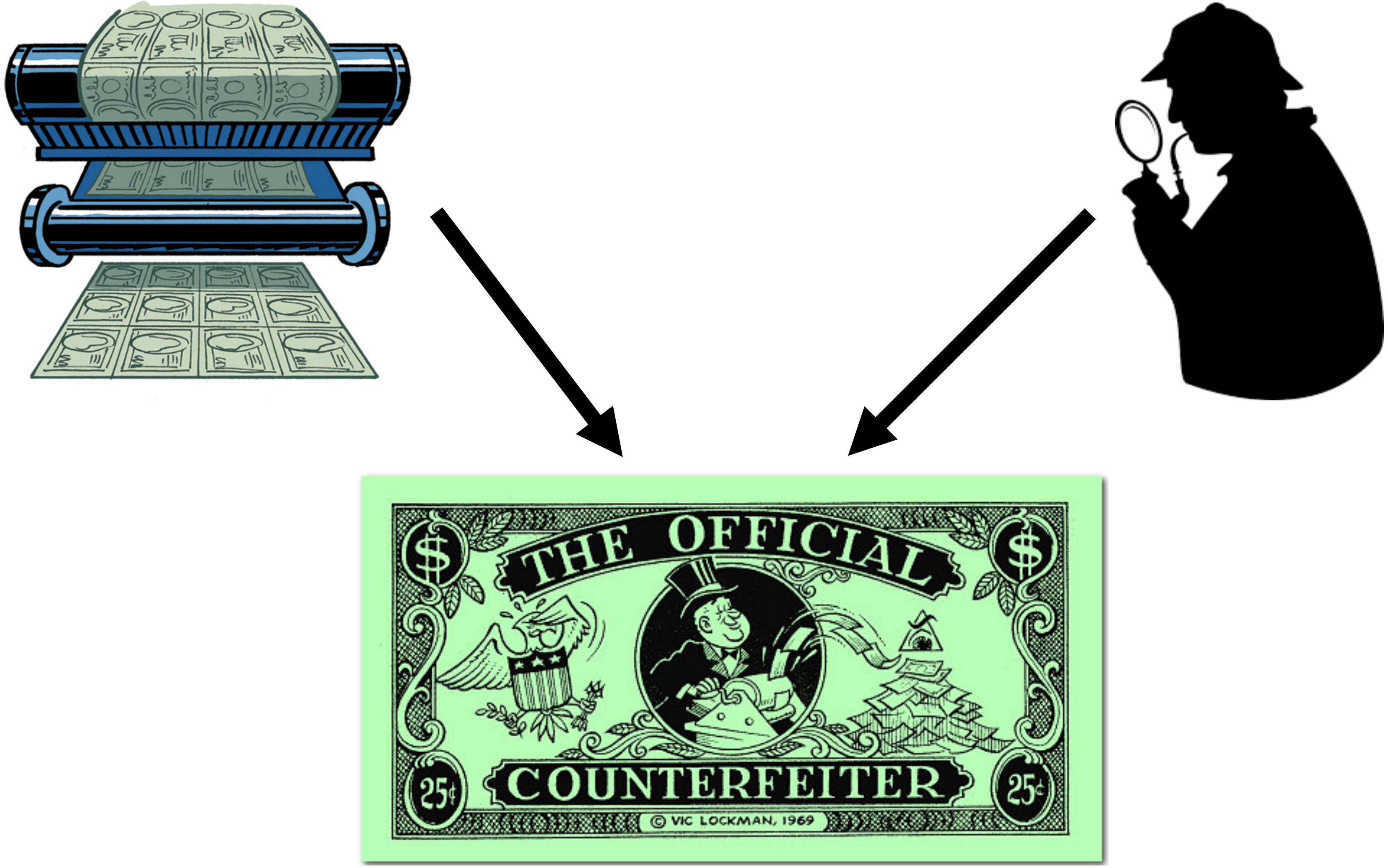
Karras et al., Progressive Growing of GANs for Improved Quality, Stability, and Variation

GENERATIVE ADVERSARIAL NETWORKS

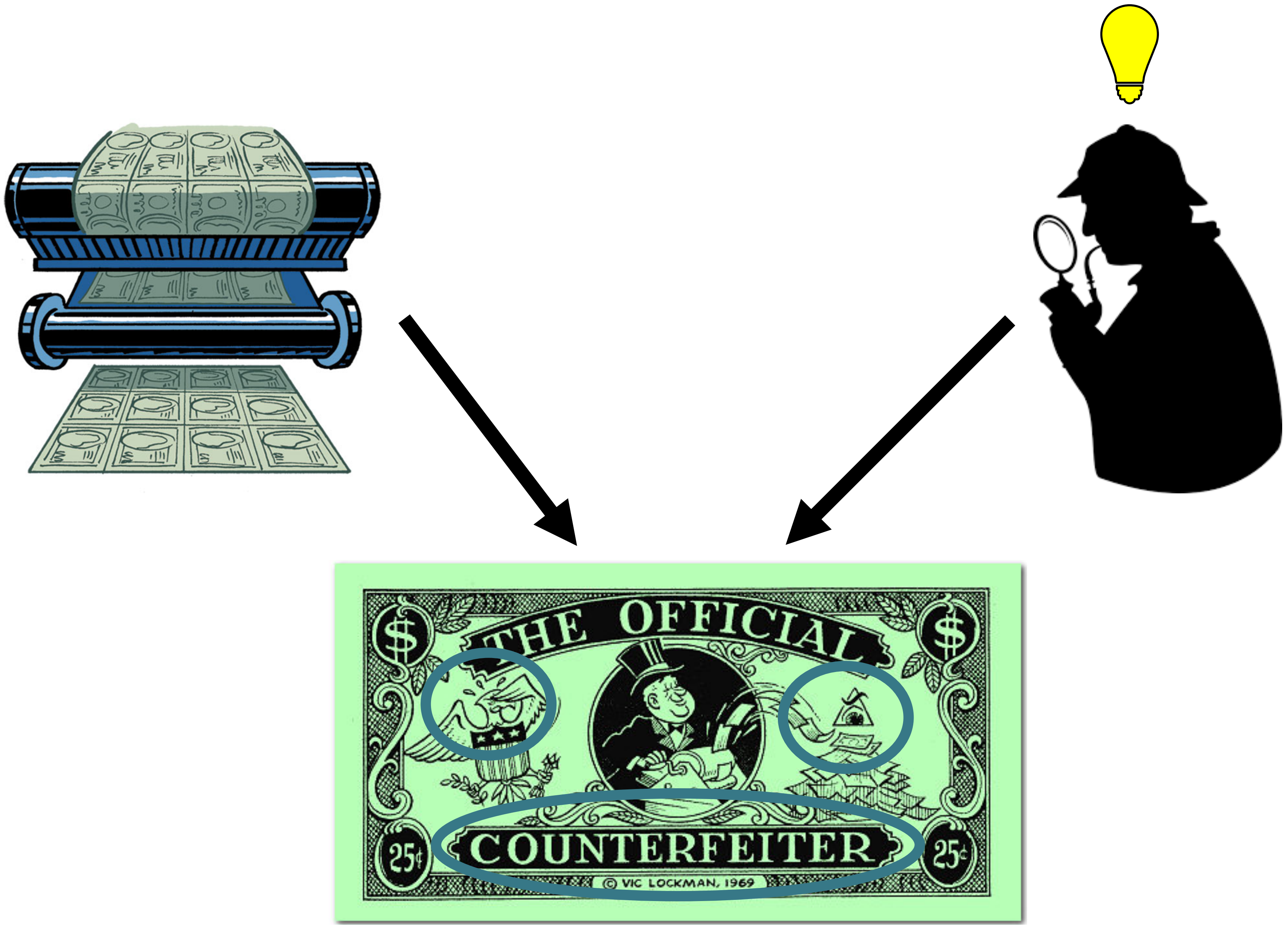
GENERATIVE ADVERSARIAL NETWORKS



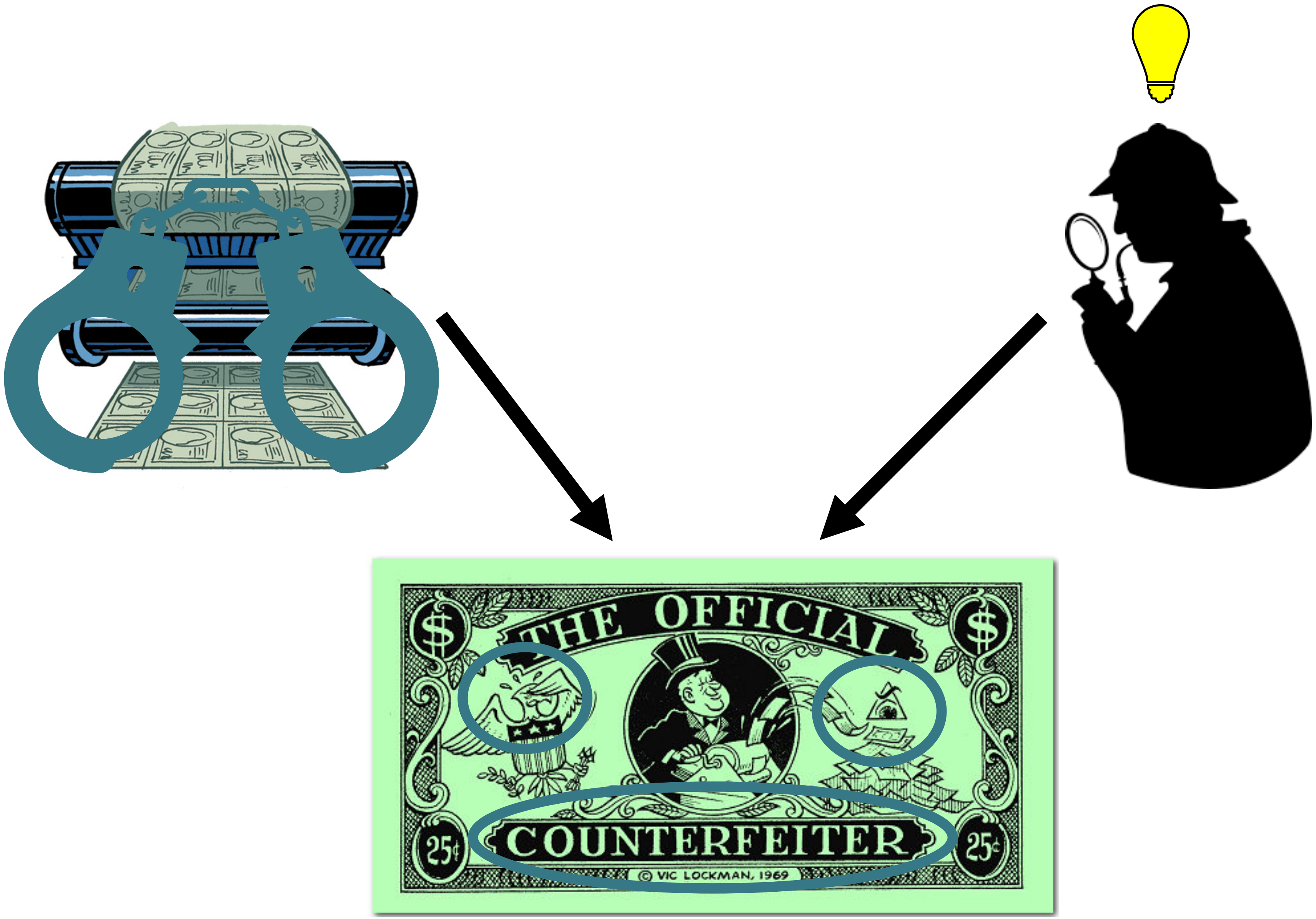
GENERATIVE ADVERSARIAL NETWORKS



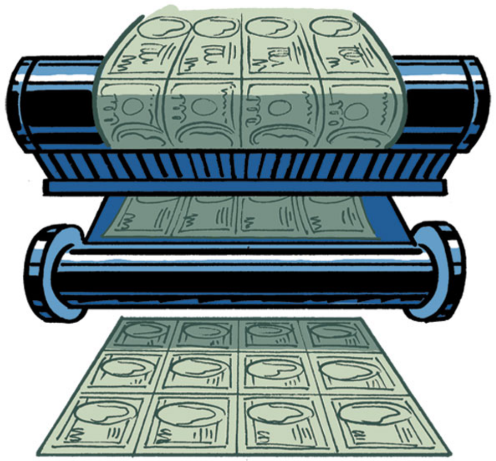
GENERATIVE ADVERSARIAL NETWORKS



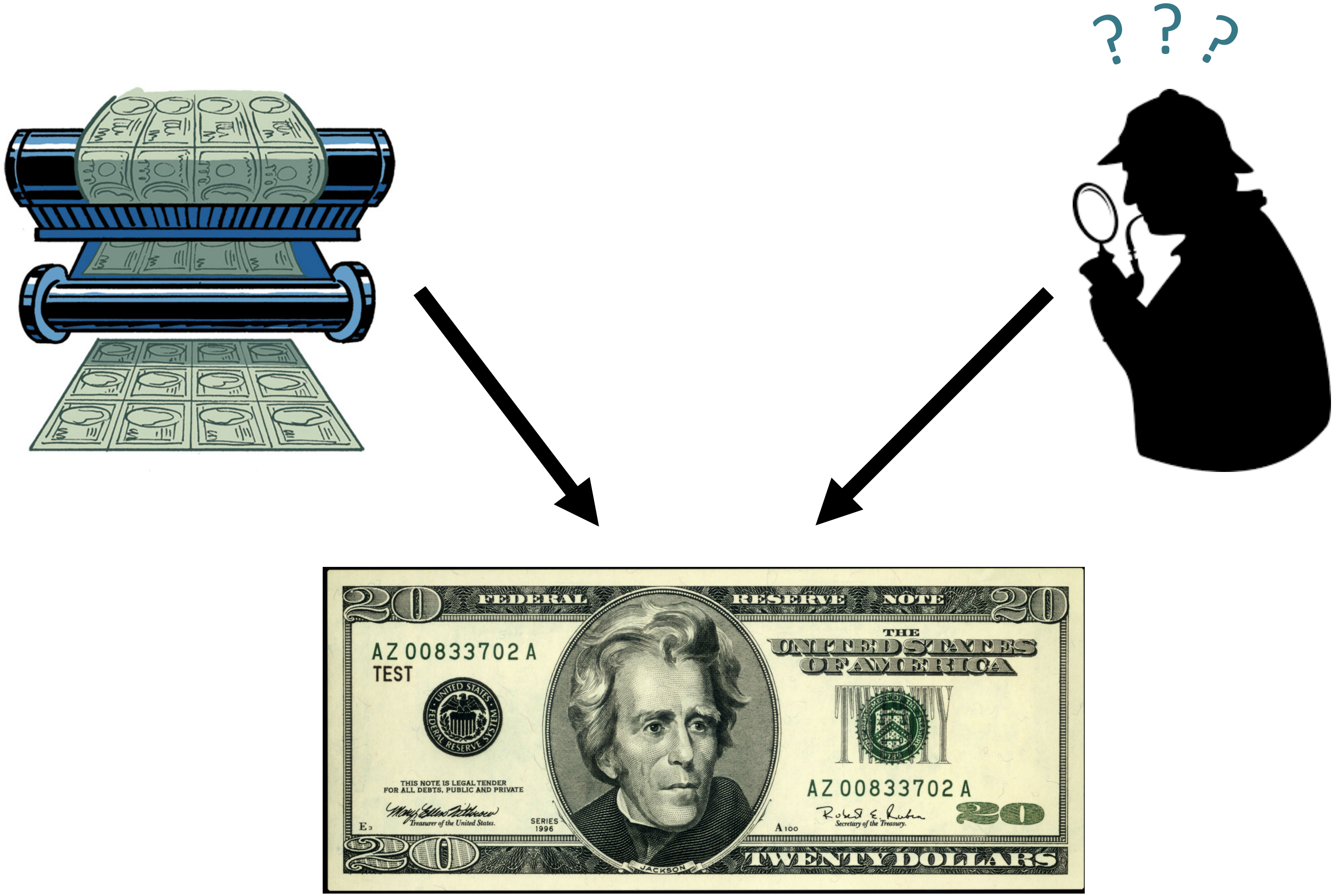
GENERATIVE ADVERSARIAL NETWORKS



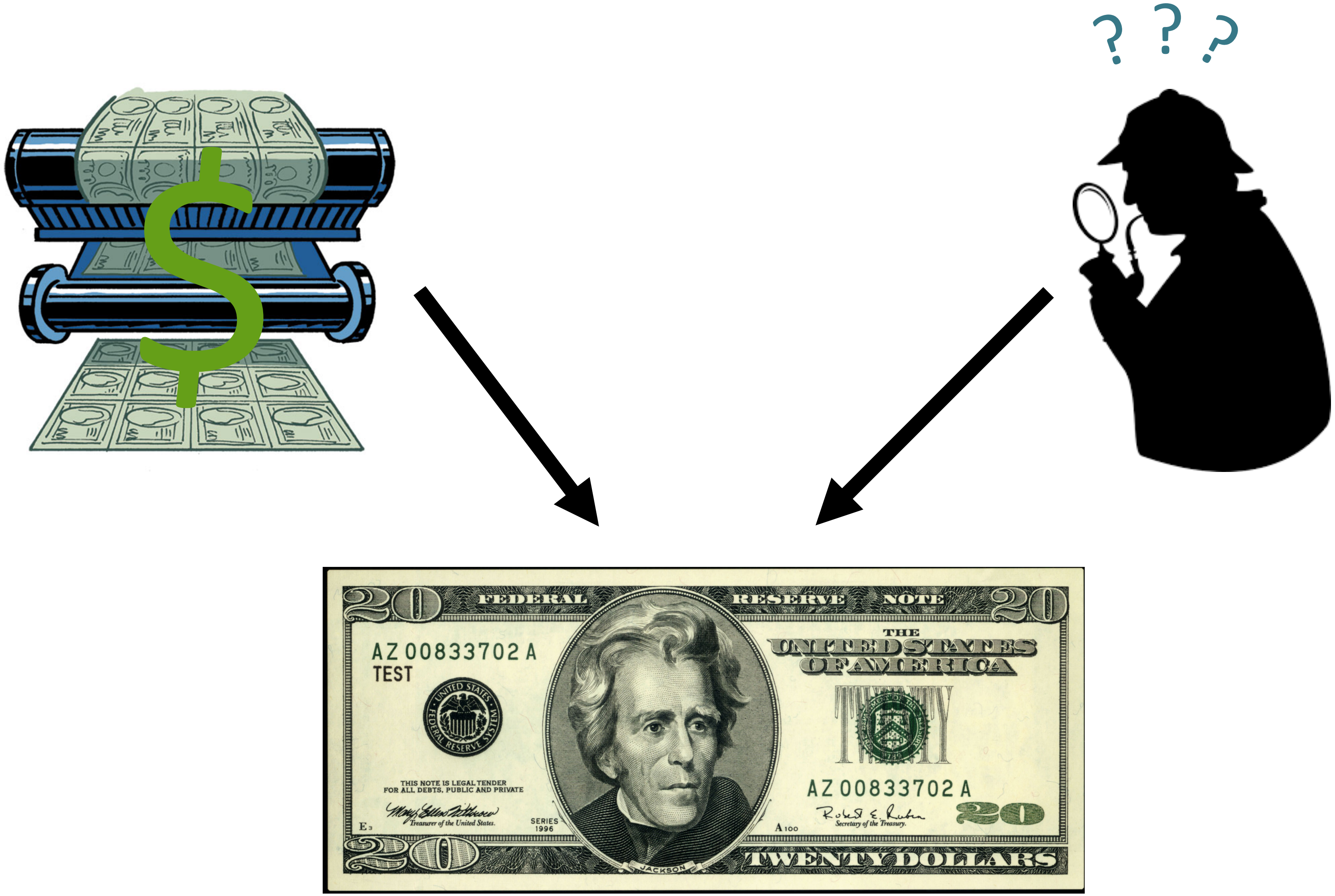
GENERATIVE ADVERSARIAL NETWORKS



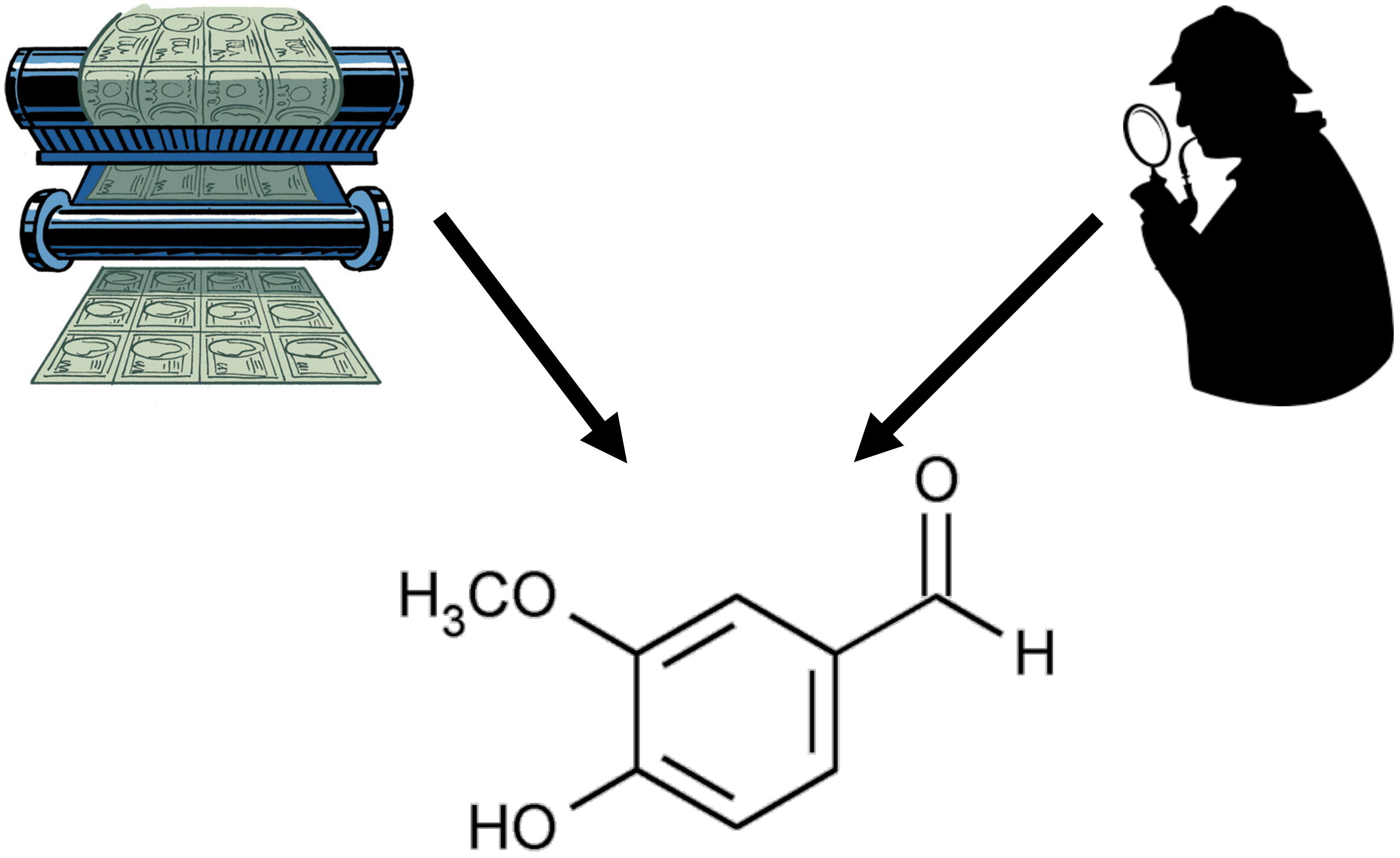
GENERATIVE ADVERSARIAL NETWORKS



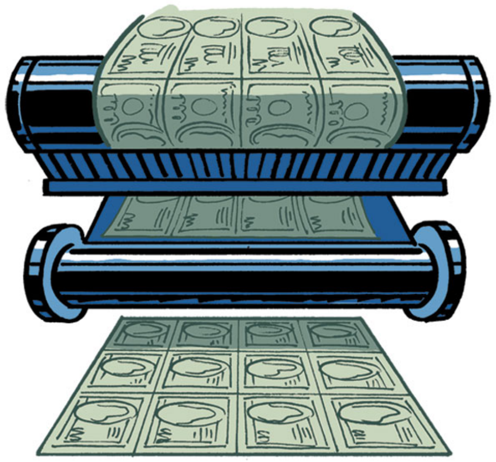
GENERATIVE ADVERSARIAL NETWORKS



GENERATIVE ADVERSARIAL NETWORKS



GENERATIVE ADVERSARIAL NETWORKS



Generator

Discriminator

- The discriminator learns to distinguish generated objects from real ones
- The generator tries to fool the discriminator into believing that generated objects are real

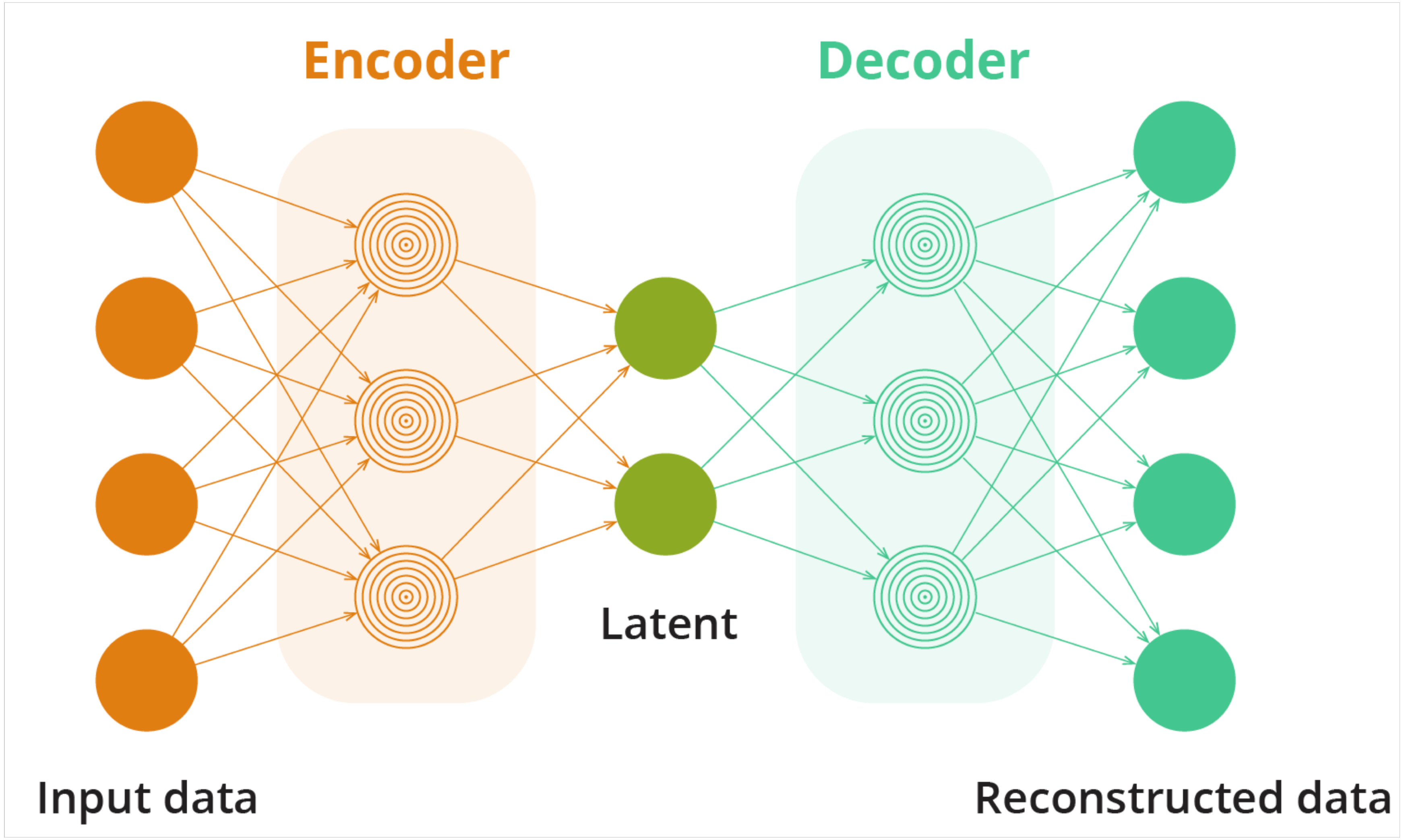
GENERATIVE ADVERSARIAL NETWORKS

- **Mode collapse:** some classes of data are not generated
- Can’t work with discrete data out of the box

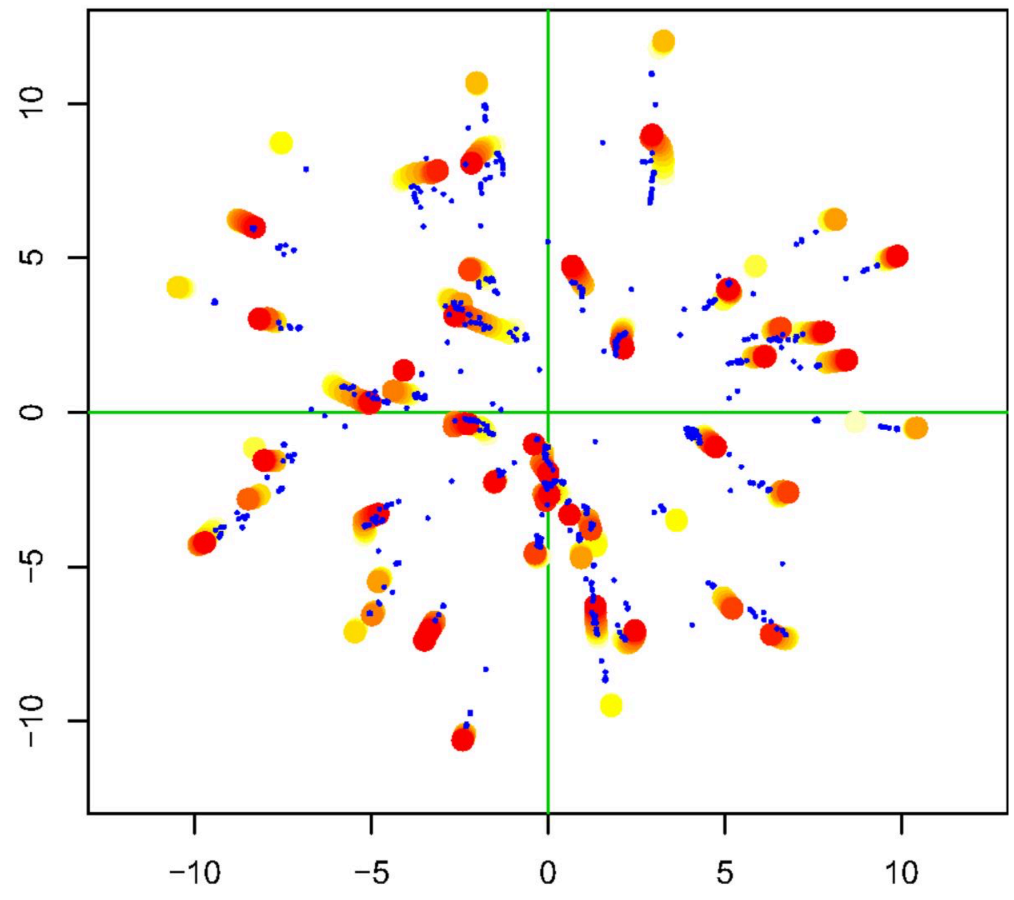
GENERATIVE ADVERSARIAL NETWORKS

- **Mode collapse:** some classes of data are not generated
- ~~Can’t work with discrete data out of the box~~

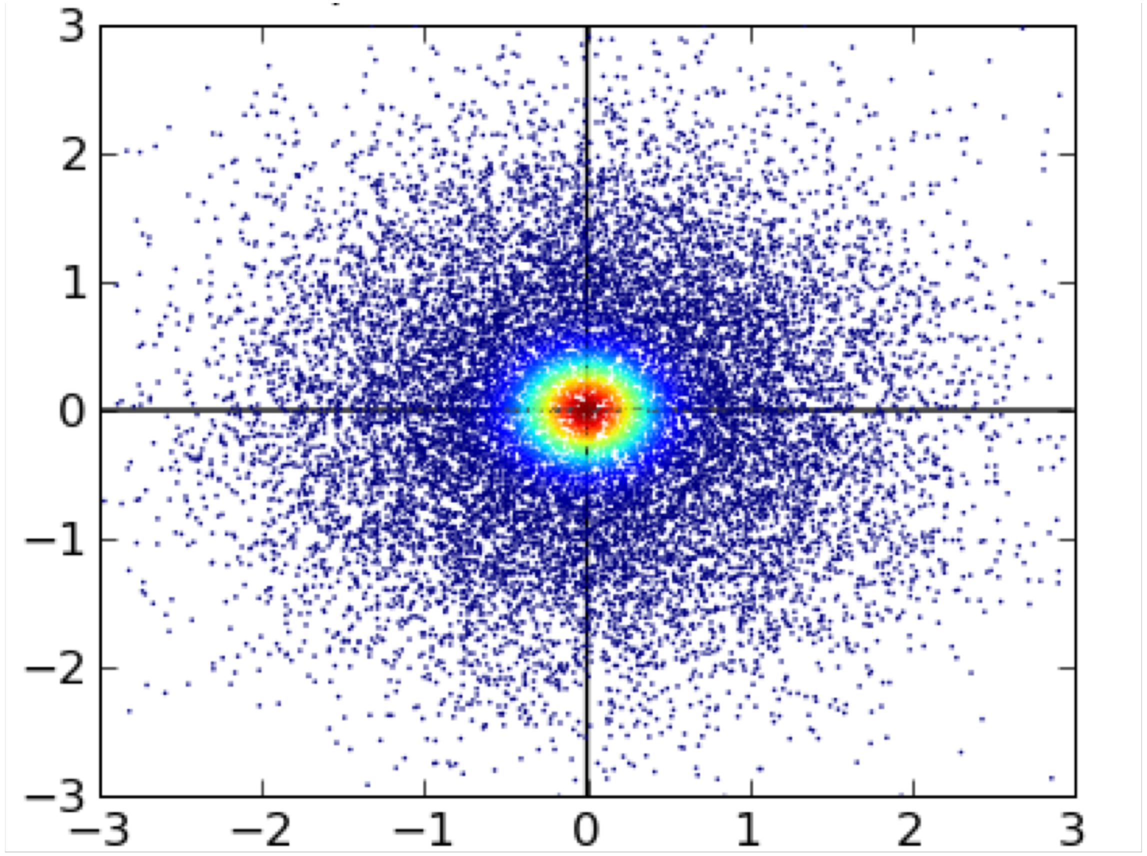
AUTOENCODER



LATENT CODES

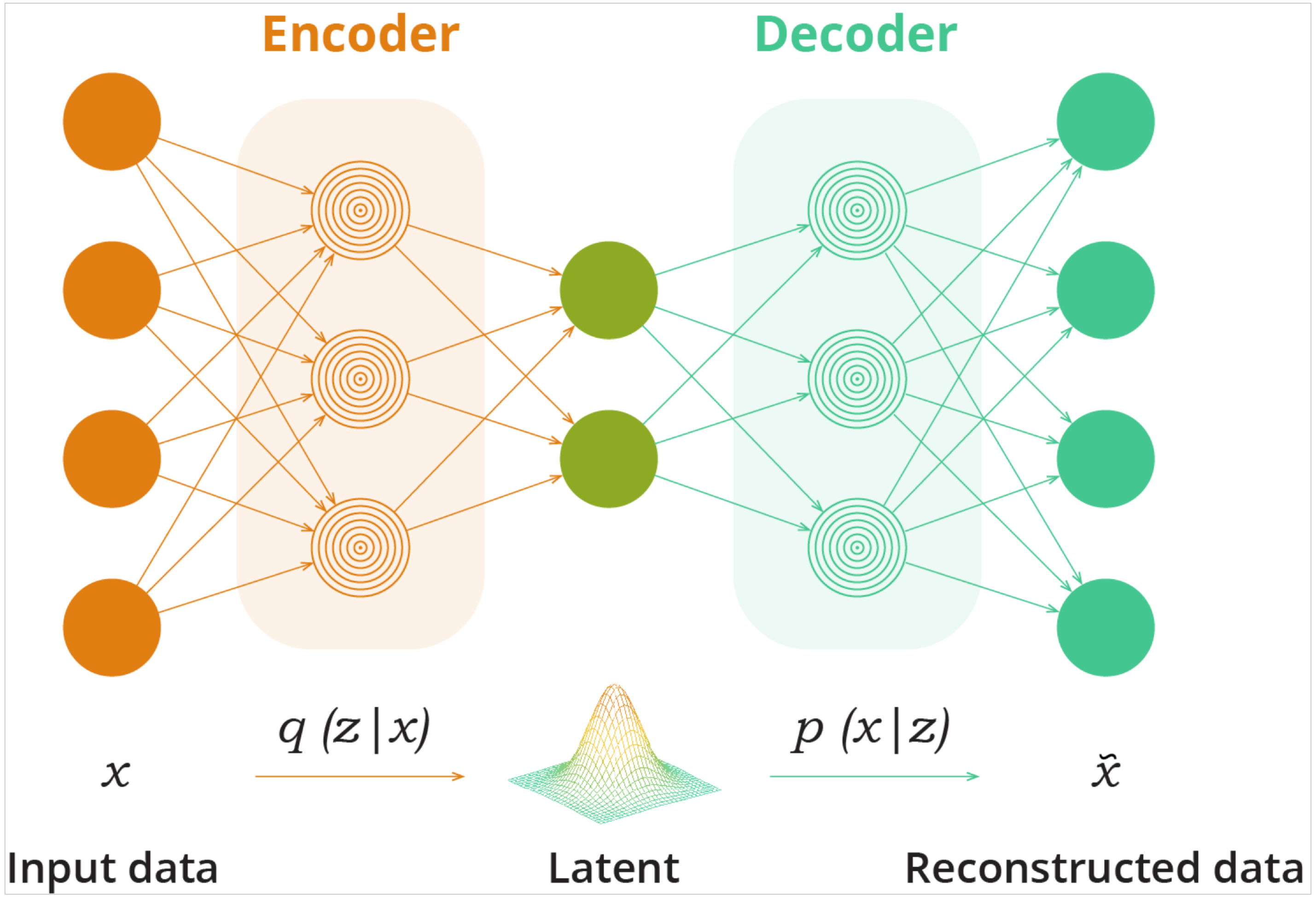


Autoencoder



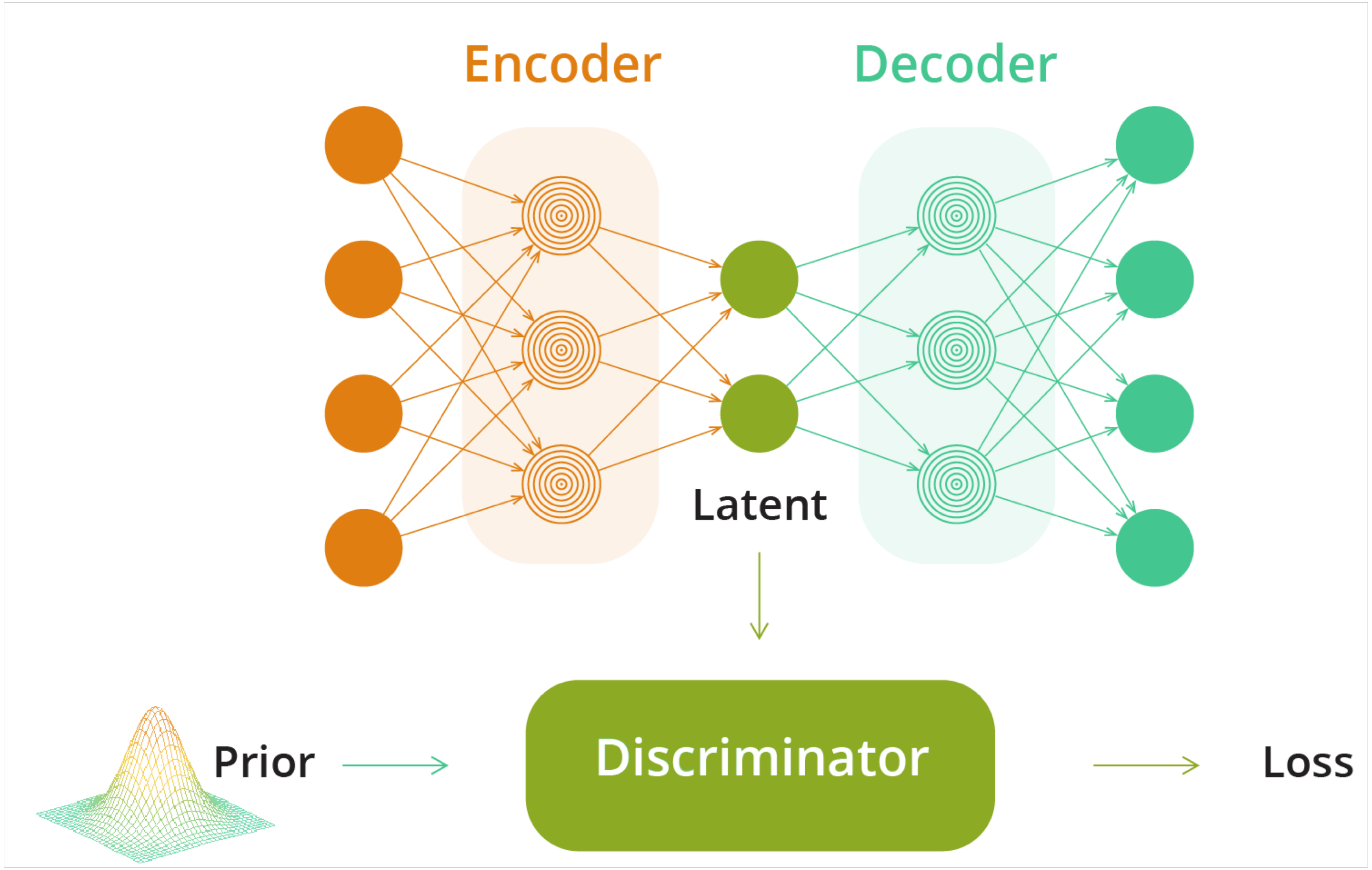
Variational Autoencoder

VARIATIONAL AUTOENCODER



Kingma & Welling 2013, Auto-Encoding Variational Bayes, arxiv.org/abs/1312.6114

ADVERSARIAL AUTOENCODER



Makhzani et al. 2015, Adversarial Autoencoders, arxiv.org/abs/1511.05644

AUTOENCODERS

- No mode collapse (otherwise decoder won’t work)
- Can work with discrete data out of the box

DRUG DISCOVERY

REPRESENTATIONS

REPRESENTATIONS

Fingerprints

00010001000101000…

Kadurin et al, The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology

Can be used to search known molecules by Tanimoto similarity:

$$T(x,y) = \frac{|\mathrm{fp}(x) \& \mathrm{fp}(y)|}{|\mathrm{fp}(x) \vee \mathrm{fp}(y)|}$$

Good when there is a lack of data

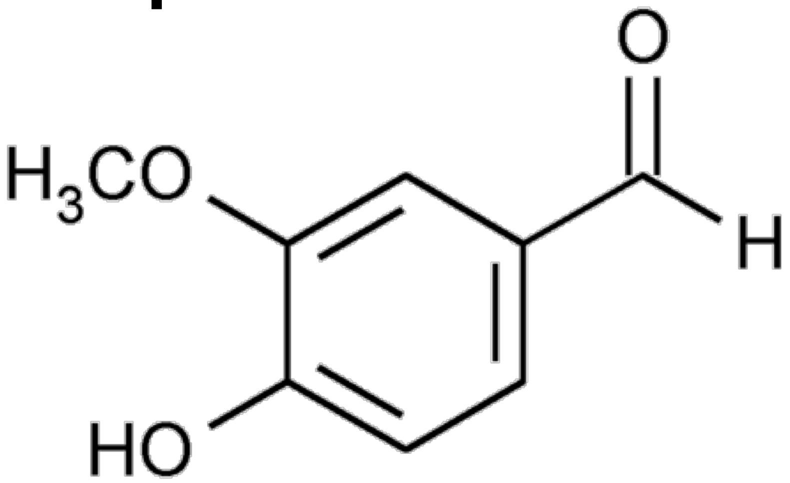
REPRESENTATIONS

- Train a message passing neural network

OR

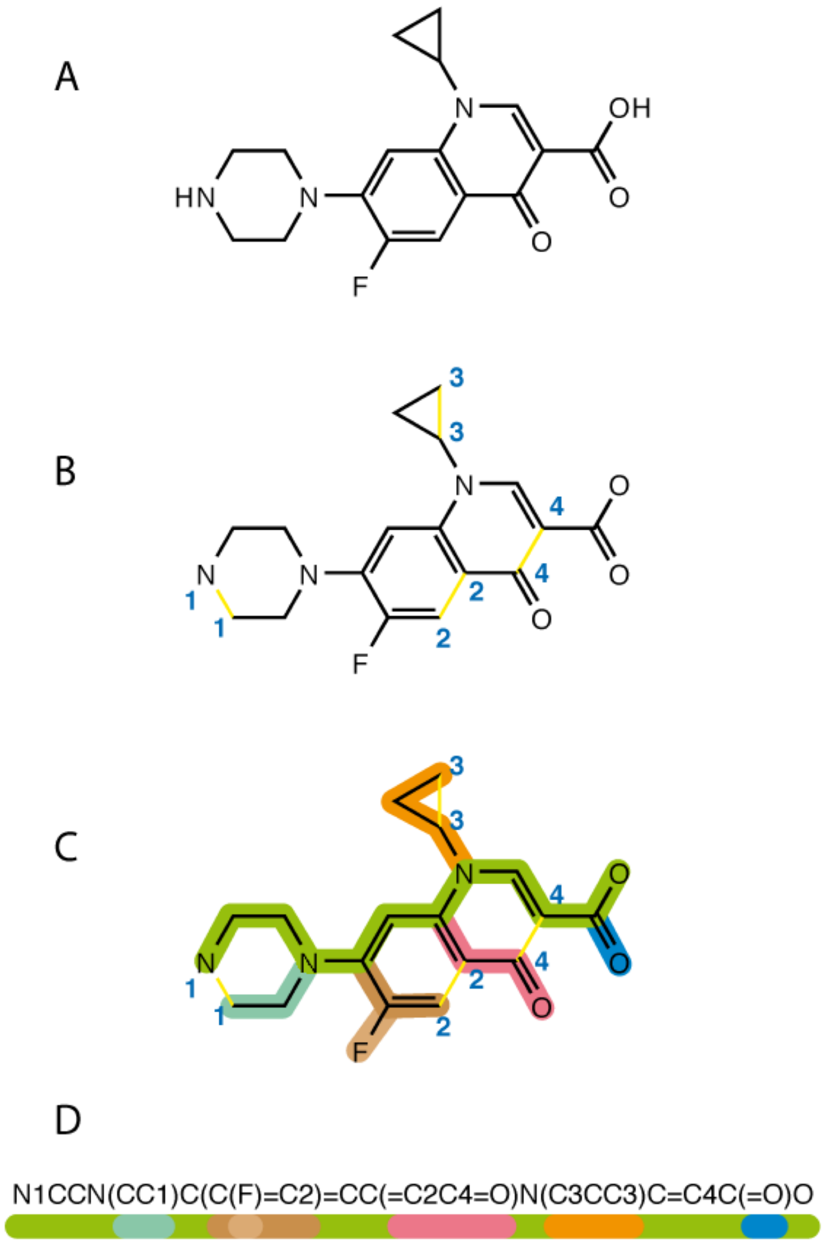
- Train a CNN on the adjacency matrix

Graphs



Jin et al, Junction Tree Variational Autoencoder for Molecular Graph Generation

REPRESENTATIONS



SMILES Strings

c1(C=O)cc(OC)c(O)cc1

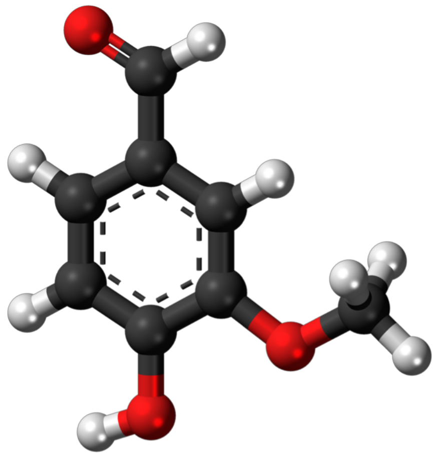
Dai et al, Syntax-Directed Variational Autoencoder for Structured Data

- Build a spanning tree
- Write atoms in the depth-first search order

REPRESENTATIONS

- \+ Can be used together with an active site render
- – Voxels are very sparse—can’t even train an CNN-autoencoder on them

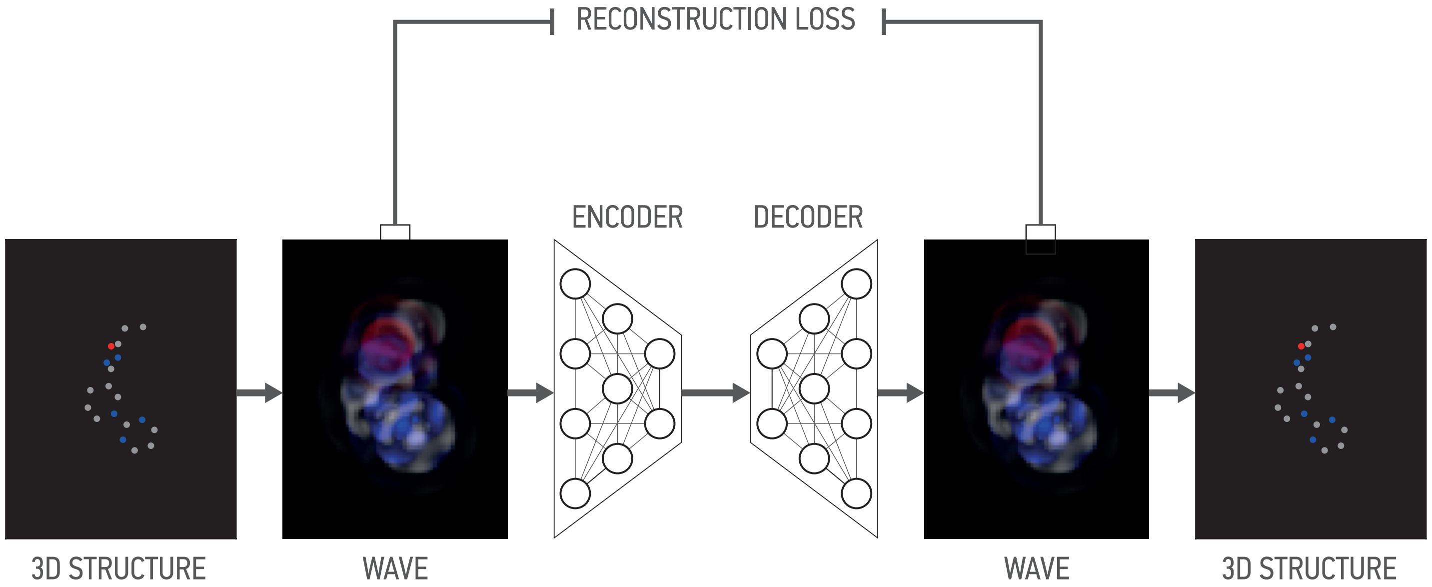
3D



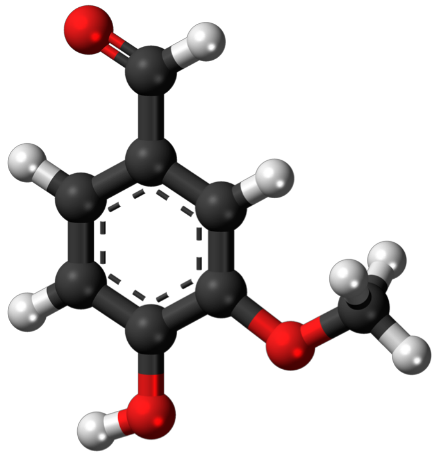
Kuzminykh et al, 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks

REPRESENTATIONS

+ Can be used together with an active site render
– Voxels are very sparse—can't even train an CNN-autoencoder on them



3D



Kuzminykh et al, 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks

REPRESENTATIONS

Fingerprints

00010001000101000...

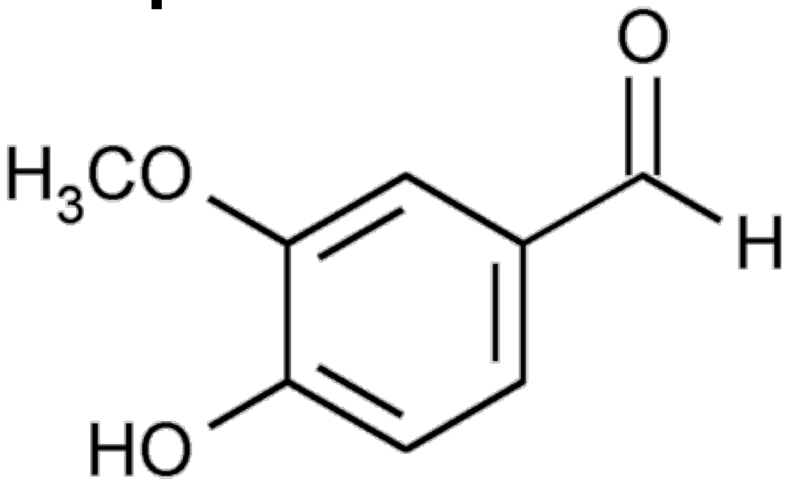
Kadurin et al, The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology

SMILES Strings

c1(C=O)cc(OC)c(O)cc1

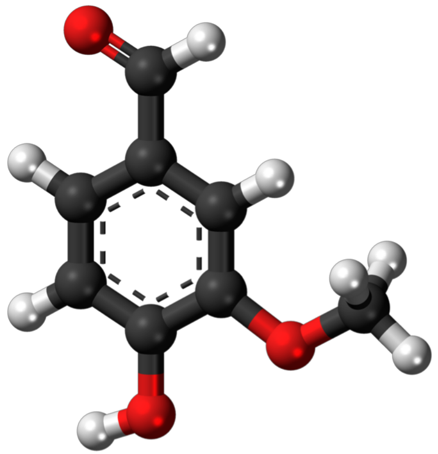
Dai et al, Syntax-Directed Variational Autoencoder for Structured Data

Graphs



Jin et al, Junction Tree Variational Autoencoder for Molecular Graph Generation

3D



Kuzminykh et al, 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks

Grammar and Graphs

GRAMMAR VAE

- Write a formal grammar that can represent any SMILES
- Train an autoencoder

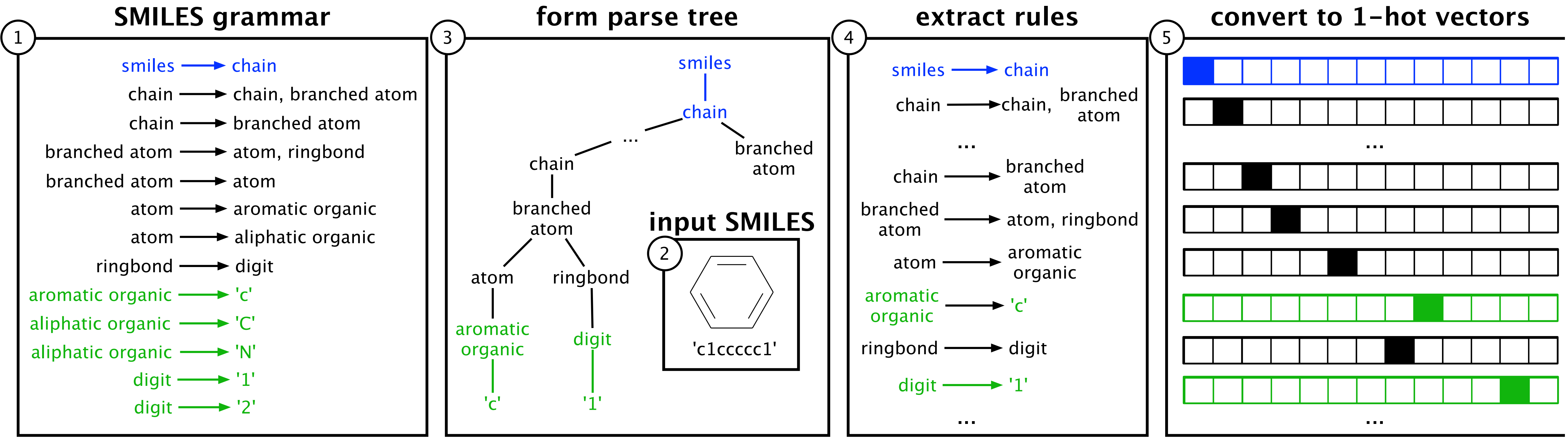
Context-free grammar (V, Σ, R, S)

- V — non-terminal symbols
- Σ — terminal symbols
- R — production rules $V \rightarrow (V \cup \Sigma)^*$
- S — non-terminal start symbol

Kusner et al., Grammar Variational Autoencoder

GRAMMAR VAE

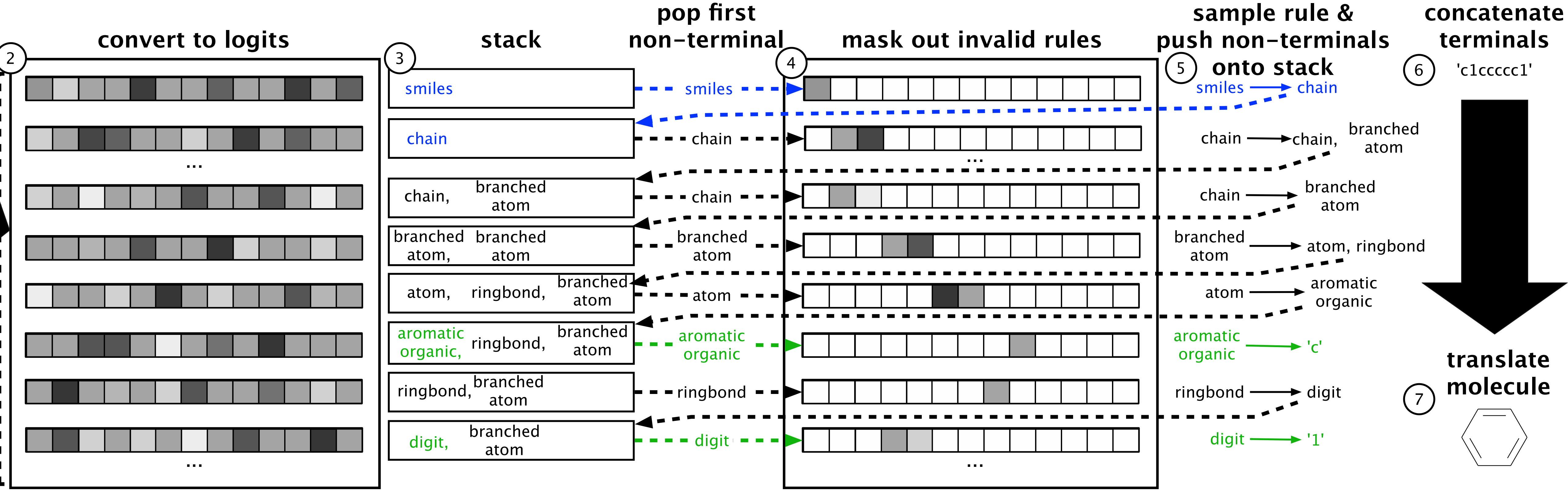
- Write a formal grammar that can represent any SMILES
- Train an autoencoder



Kusner et al., Grammar Variational Autoencoder

GRAMMAR VAE

- Write a formal grammar that can represent any SMILES
- Train an autoencoder

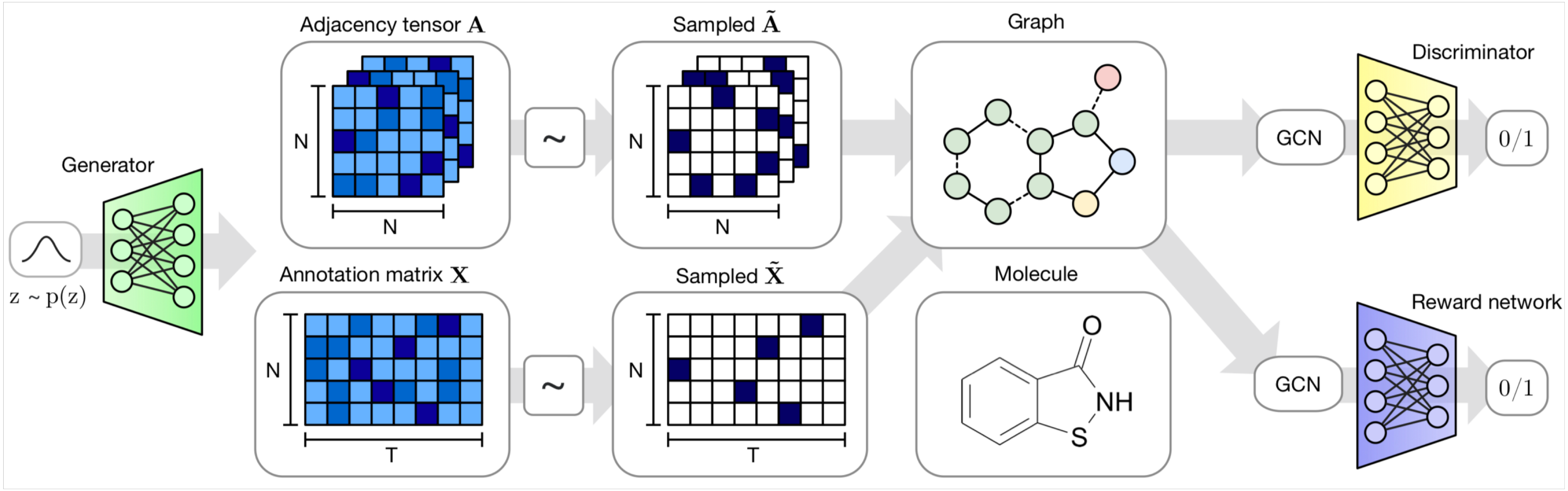


Kusner et al., Grammar Variational Autoencoder

SMILES VS GRAPHS

- 1D (SMILES) cycles are spread across the whole string
- 2D (Graphs) decisions are made more locally

GRAPHS: MOLGAN

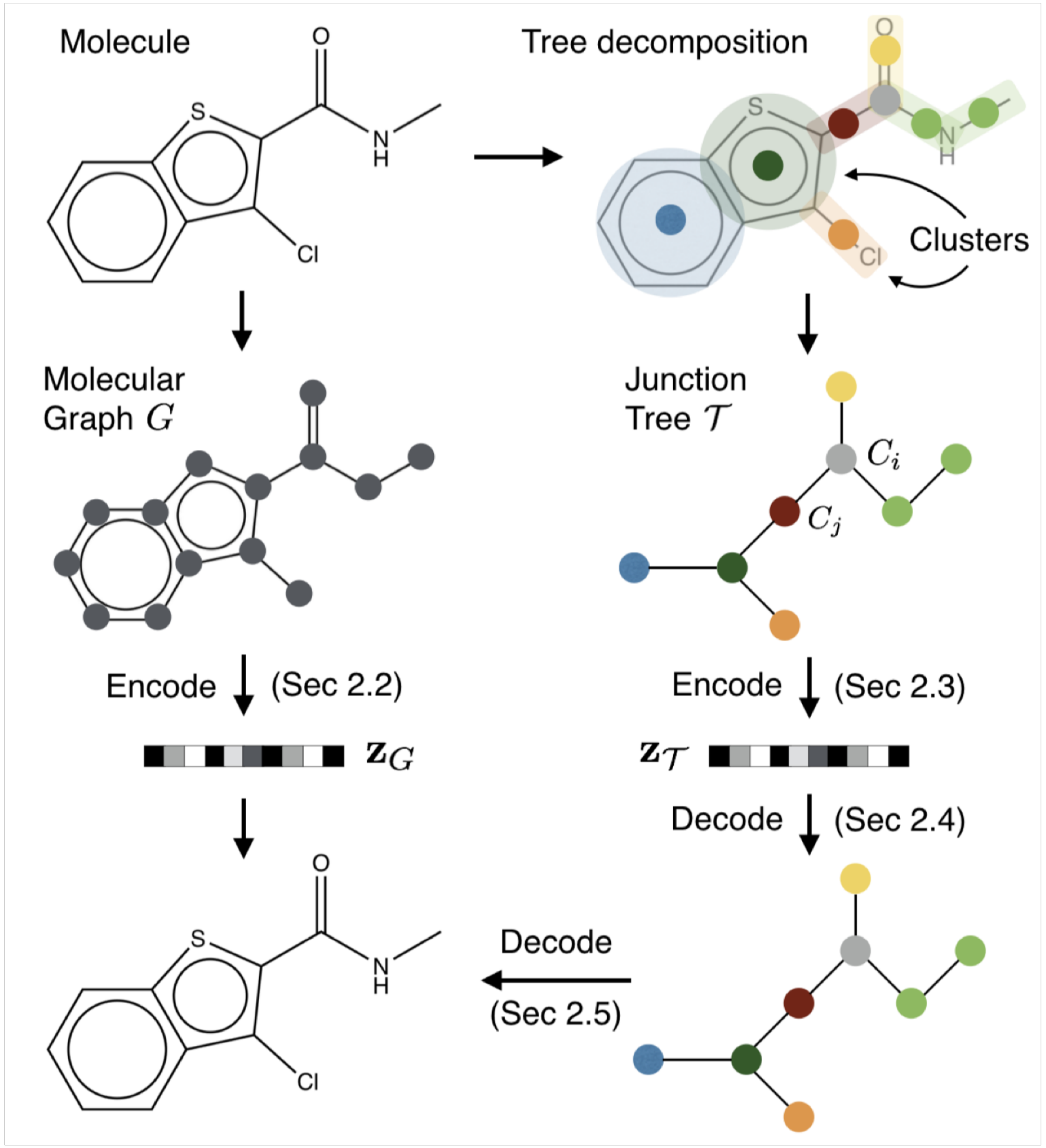


Graph Convolutional Networks (GCN)

- $h^{l+1} = \tanh\left[f_s^l(h_i^l, x_i) + \sum_{j=1}^{N} \sum_{y=1}^{Y} \frac{A_{i,j,y}}{|N_i|} f_y^l(h_j^l, x_i)\right]$

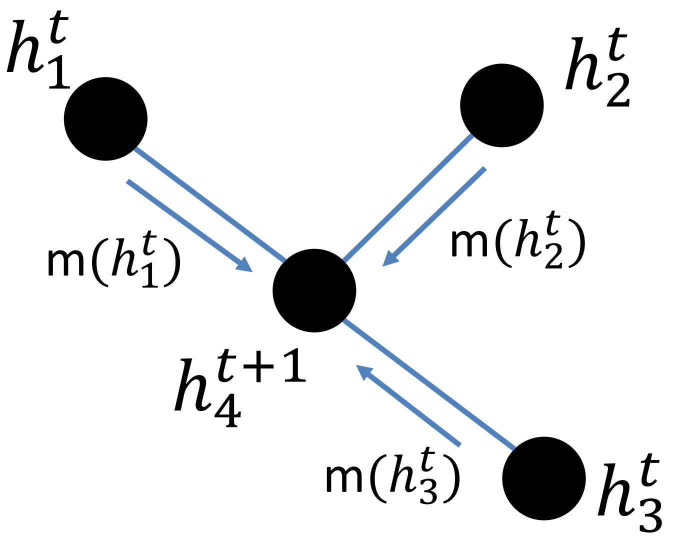
De Caom & Kipf, MolGAN: An implicit generative model for small molecular graphs

GRAPHS: JUNCTION TREE VAE



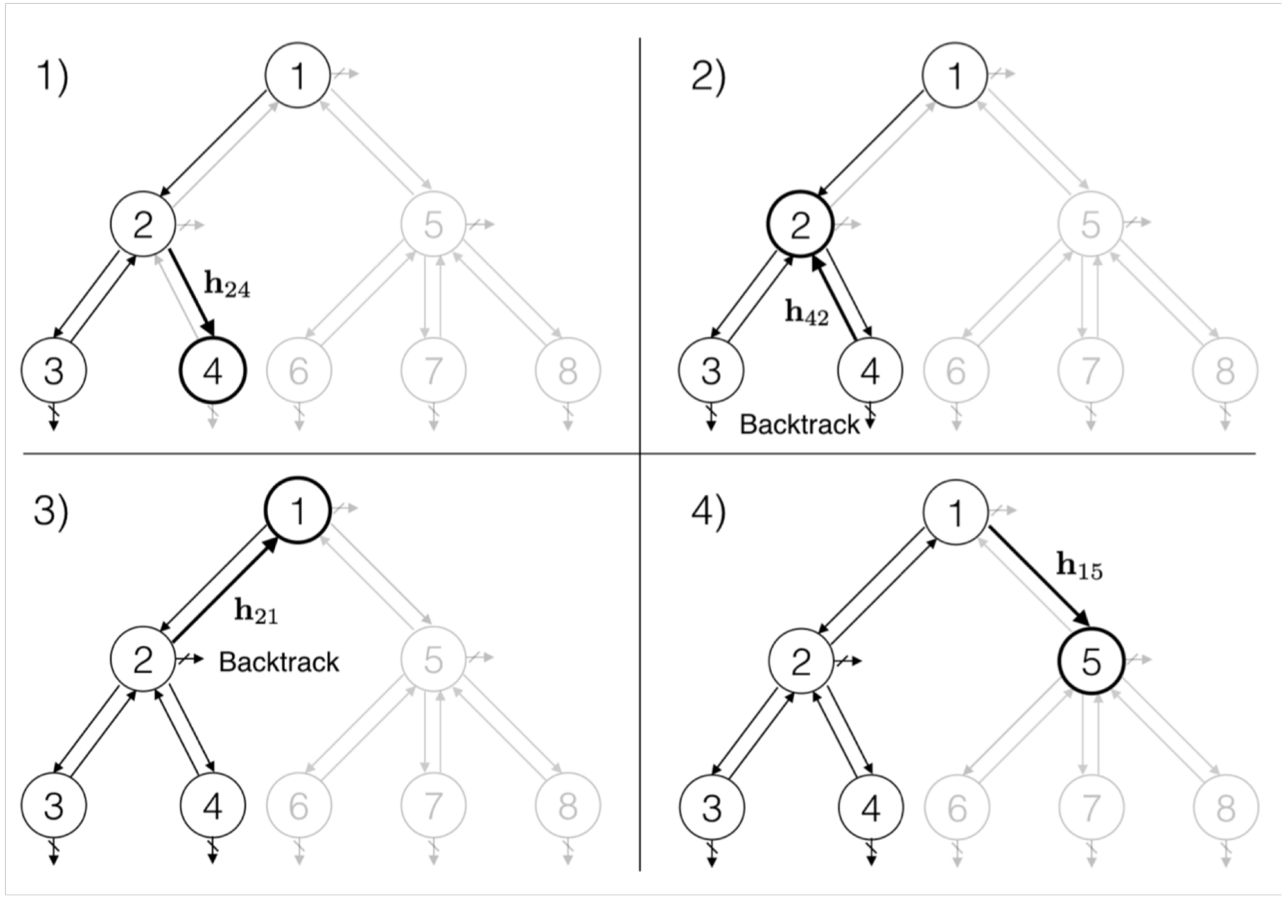
Jin et al, Junction Tree Variational Autoencoder for Molecular Graph Generation

GRAPHS: JUNCTION TREE VAE



$$z_4^{t+1} = \tanh(W^t \sum_{i \in ch(4)} m(z_i^t))$$

Message-Passing Neural Networks



Decoding

Jin et al, Junction Tree Variational Autoencoder for Molecular Graph Generation

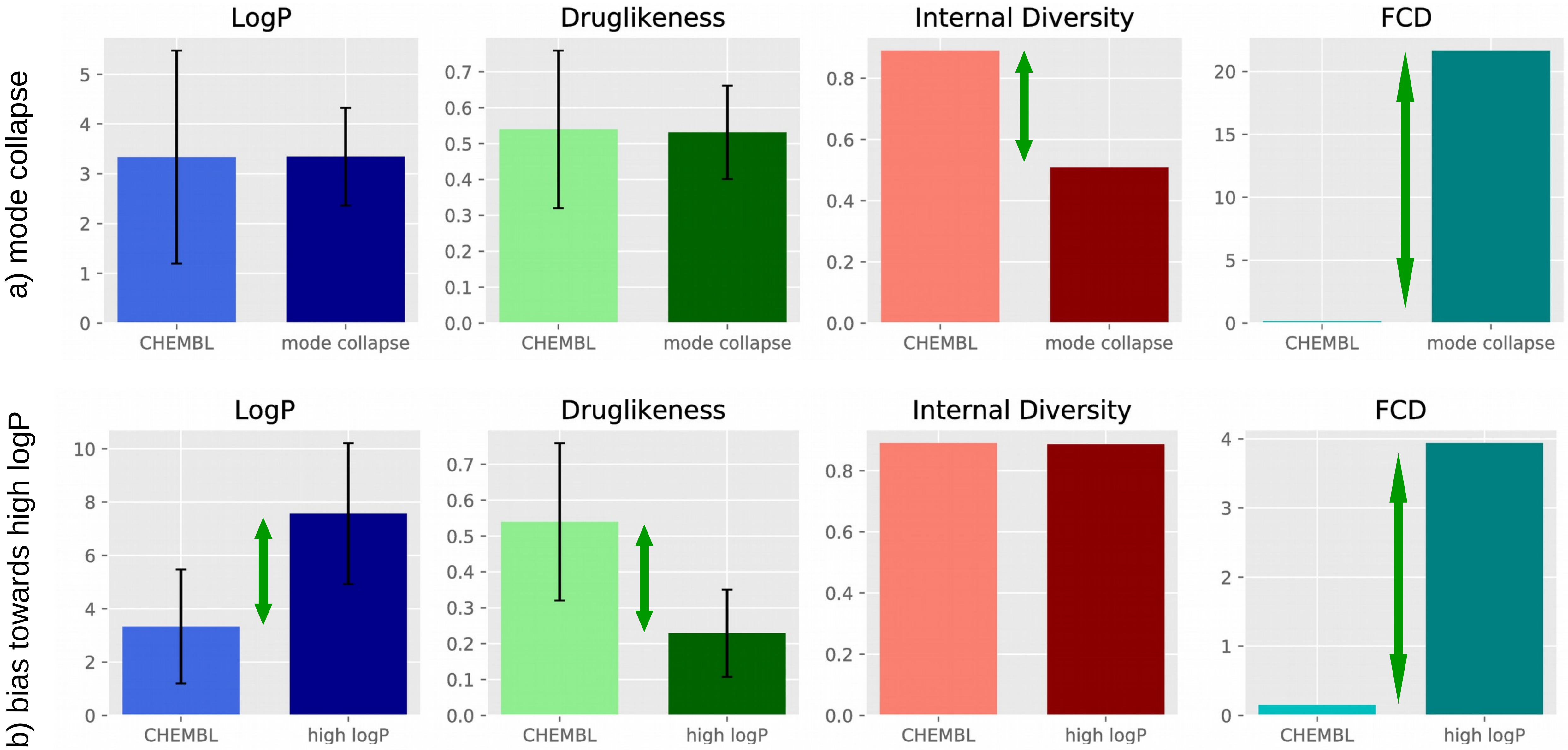
HOW TO EVALUATE A MODEL

Metric	Description
%Unique@k	Measures diversity in terms of uniqueness
%Valid	Measures model quality in terms of validity (e.g. correct valences)
Internal Diversity	Average pairwise similarity between generated molecules
External Diversity	Average similarity to the nearest known molecule
Fréchet ChemNet Distance	Similarity of distributions at the penultimate layer of ChemNet

Similarity between molecules:

$$T(x, y) = \frac{|\mathrm{fp}(x) \& \mathrm{fp}(y)|}{|\mathrm{fp}(x) \vee \mathrm{fp}(y)|}$$

FRÉCHET CHEMNET DISTANCE



Preuer et al., Fréchet ChemNet Distance: A Metric for Generative Models for Molecules in Drug Discovery

DRUG DISCOVERY

- Can generate novel molecular structures GANs, VAEs or AAEs
- Would this useful?

DRUG DISCOVERY

- Can generate novel molecular structures GANs, VAEs or AAEs
- Would this useful?
- We need to generate molecular structures with specific properties

DRUG DISCOVERY

- Can generate novel molecular structures GANs, VAEs or AAEs
- Would this useful?
- We need to generate molecular structures with specific properties

Looking for a needle in a haystack

Generate perfect needles

HOW TO DISCOVER A MOLECULE?

- How to generate molecules with given properties (e.g. given activity or synthetic accessibility)

CONDITIONAL GENERATION

$$x \sim p(x \mid \text{properties})$$

HOW TO DISCOVER A MOLECULE?

- How to generate molecules with given properties (e.g. given activity or synthetic accessibility)

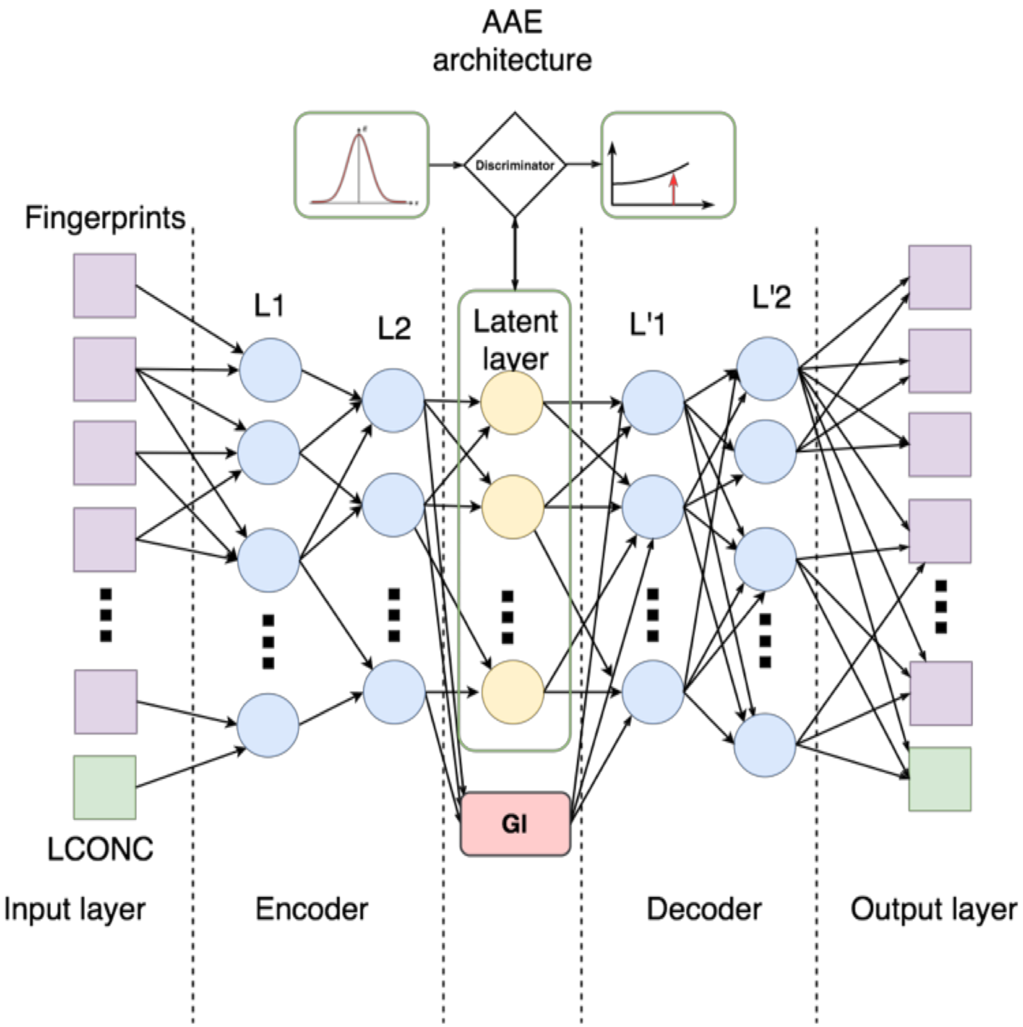
CONDITIONAL GENERATION

$$x \sim p(x \mid \text{properties})$$

OPTIMIZATION

$$\text{quality}(x) \to \max_{x}$$

CONDITIONAL GENERATION



- Specified GI $\sim N(5,1)$
- Reconstructed molecules and selected those with LCONC < -5
- Top 10 similar compounds for each of 32 molecules were selected from PubChem
- Many from resulting 69 compounds were studied as anticancer compounds

Kadurin, et al, The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology.

CAN NEURAL NETWORKS EXTRAPOLATE?

- If the largest value of activity in the training dataset is X, can we generate objects with a higher activity than X?

CAN NEURAL NETWORKS EXTRAPOLATE?

- If the largest value of activity in the training dataset is X, can we generate objects with a higher activity than X?

- NO: Why would a model be able to infer outside the values it was trained on?

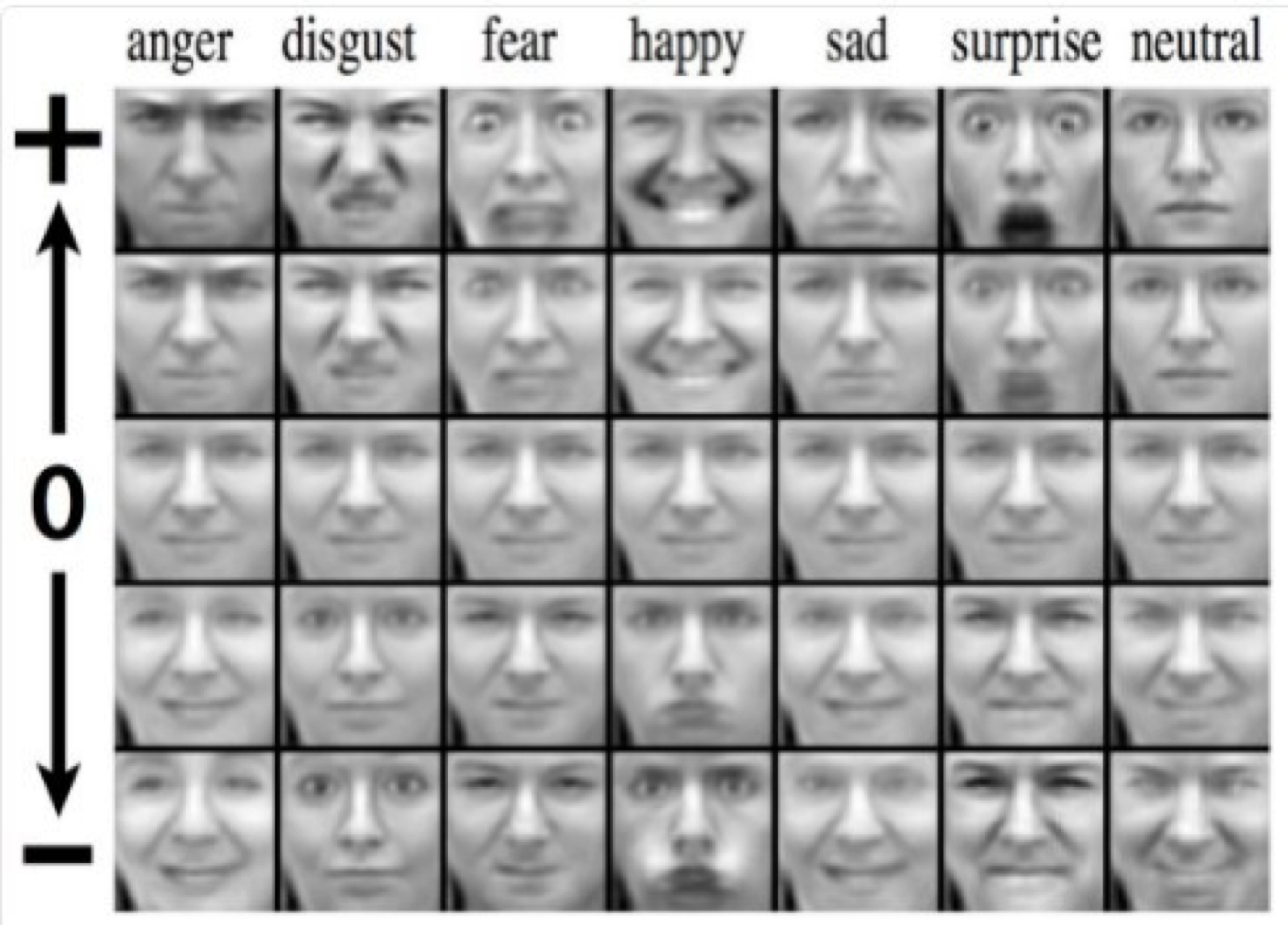
CAN NEURAL NETWORKS EXTRAPOLATE?

- If the largest value of activity in the training dataset is X, can we generate objects with a higher activity than X?

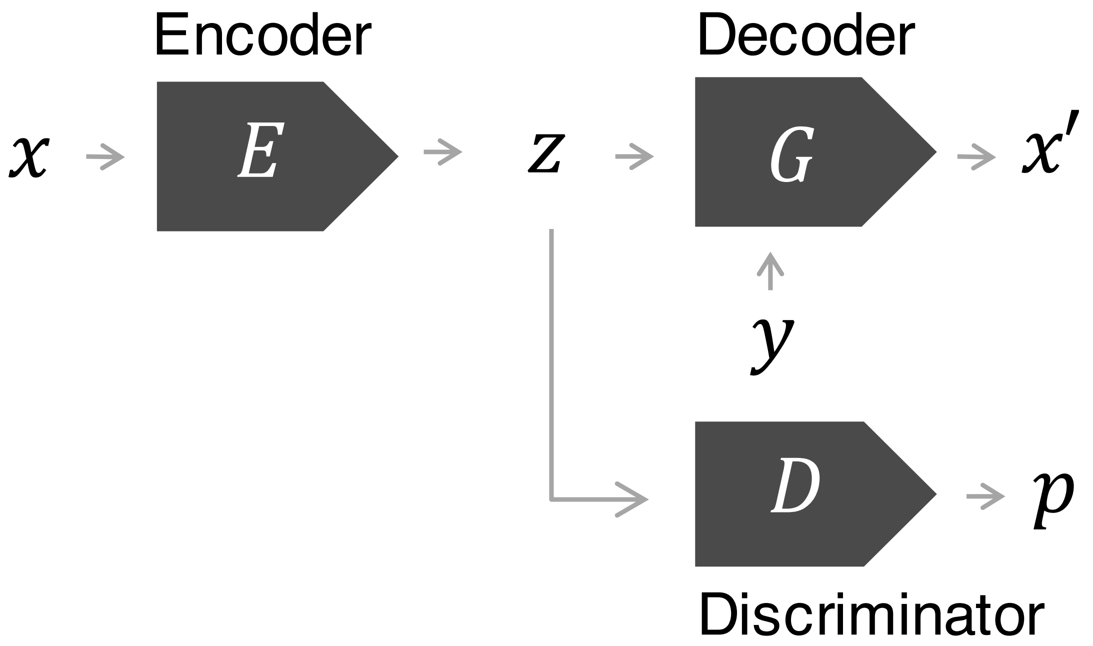
- NO: Why would a model be able to infer outside the values it was trained on?
- YES: There are some patterns that can allow the model to extrapolate

CAN NEURAL NETWORKS EXTRAPOLATE?

- Train a conditional generative model on faces labeled with emotions
- 0 — absence, 1 — present
- Generate molecules with conditions in [-5, 5] range



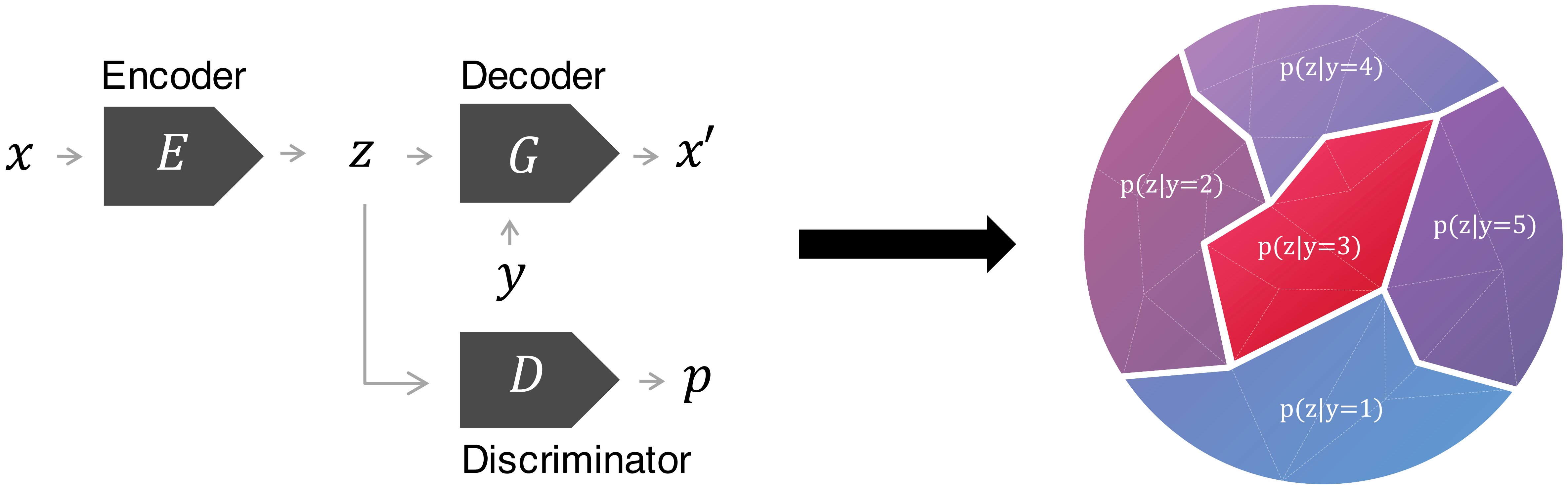
SUPERVISED (CONDITIONAL) AAE



Supervised (Conditional) Adversarial Autoencoder

Makhzani et al. 2015, Adversarial Autoencoders, arxiv.org/abs/1511.05644

SUPERVISED (CONDITIONAL) AAE



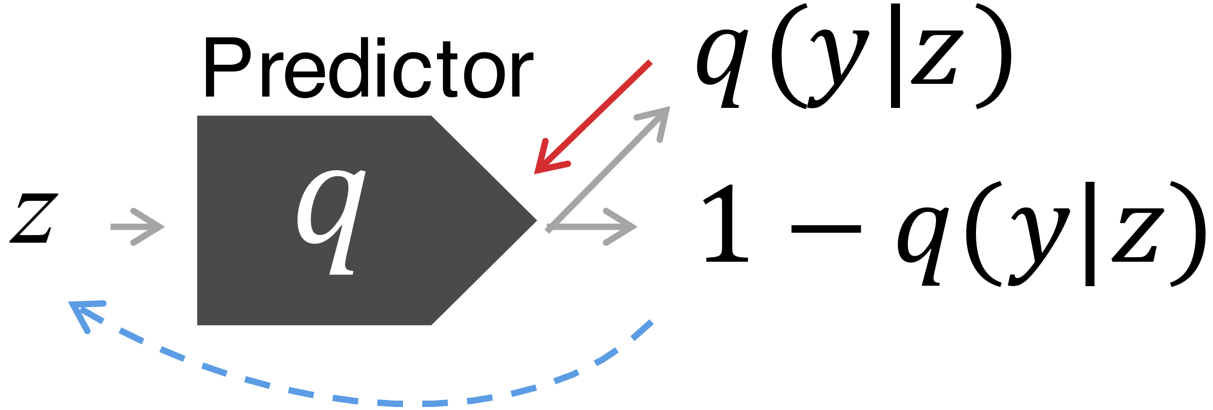
Supervised (Conditional) Adversarial Autoencoder

PREDICTIVE DISENTANGLEMENT

- Independence: minimize mutual information

$$\mathcal{I}(z,y) = \mathcal{H}(y) + \underbrace{\mathbb{E}_{p(y,z)} \log p(y|z)}_{=-\mathcal{H}(y|z)} + \underbrace{\max_q -\mathbb{E}_{p(z)} \mathcal{KL}(p(y|z) \| q(y|z))}_{=0,\ \text{achieved at } q(y|z)=p(y|z)}$$

$$= \mathcal{H}(y) + \max_q \mathbb{E}_{p(y,z)} \log q(y|z)$$



Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

PREDICTIVE DISENTANGLEMENT

- Problem: sometimes we can't estimate $q(y|z)$

$$Q = \left\{ q(y \mid z) \,\middle|\, q(y \mid z) = \prod_{i=1}^{d} q(y_i \mid z) \right\}$$

$$I(z, y) = \mathcal{H}(y) + \max_{q} \mathbb{E}_{p(y,z)} \log q(y \mid z) \geq \mathcal{H}(y) + \max_{q \in Q} \mathbb{E}_{p(y,z)} \log q(y \mid z)$$

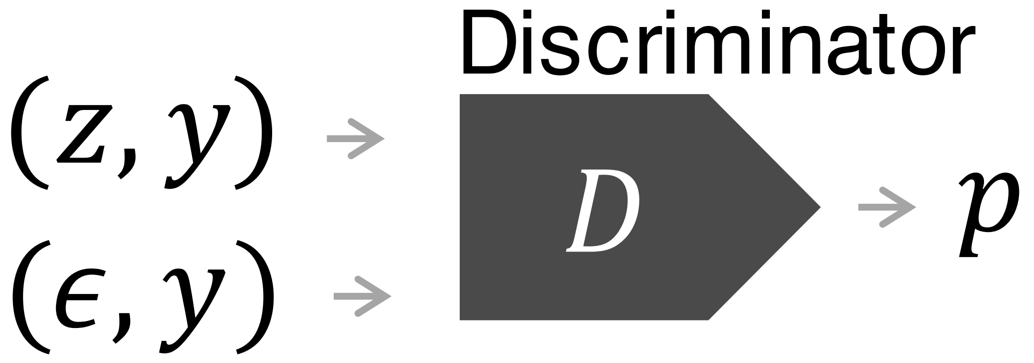
- Mutual information is underestimated
- Model will only match priors: $q(y_i|z) = p(y_i)$

Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

JOINT DISENTANGLEMENT

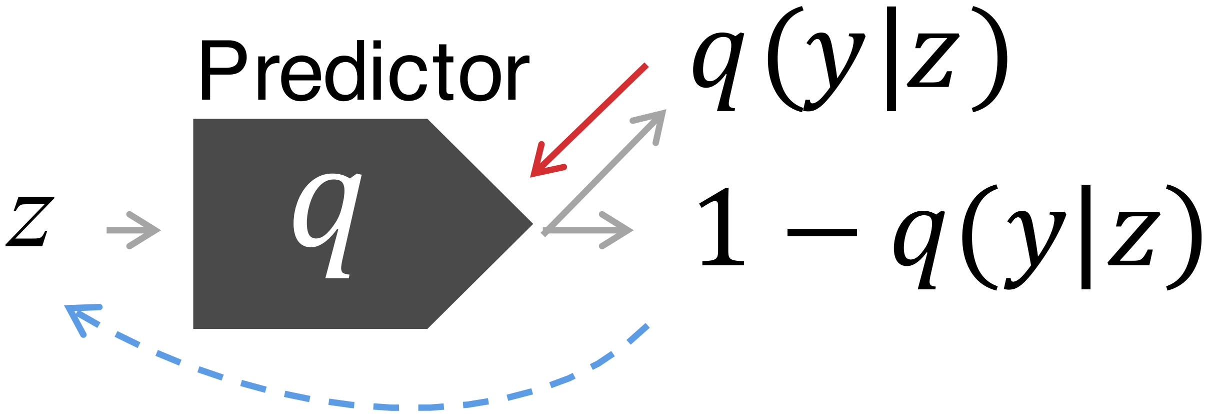
- "Real examples": samples from prior $p(z)$ and y
- "Fake examples": tuples of z and y

- Target distribution is $p(z)p(y)$

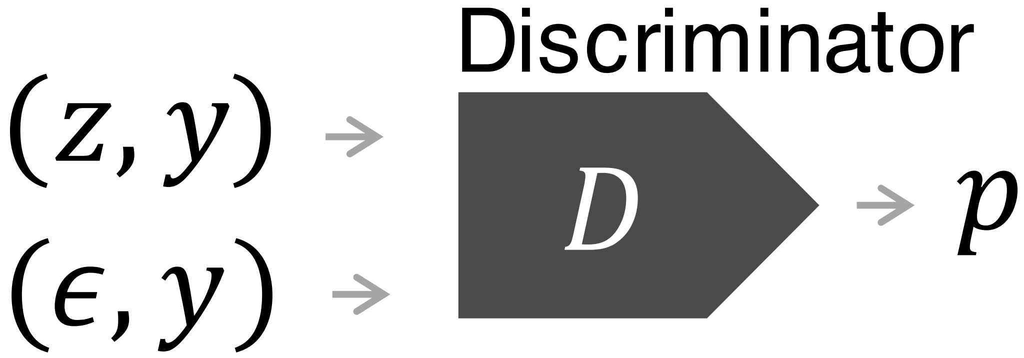


Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

DISENTANGLEMENTS



Predictive disentanglement



Joint disentanglement

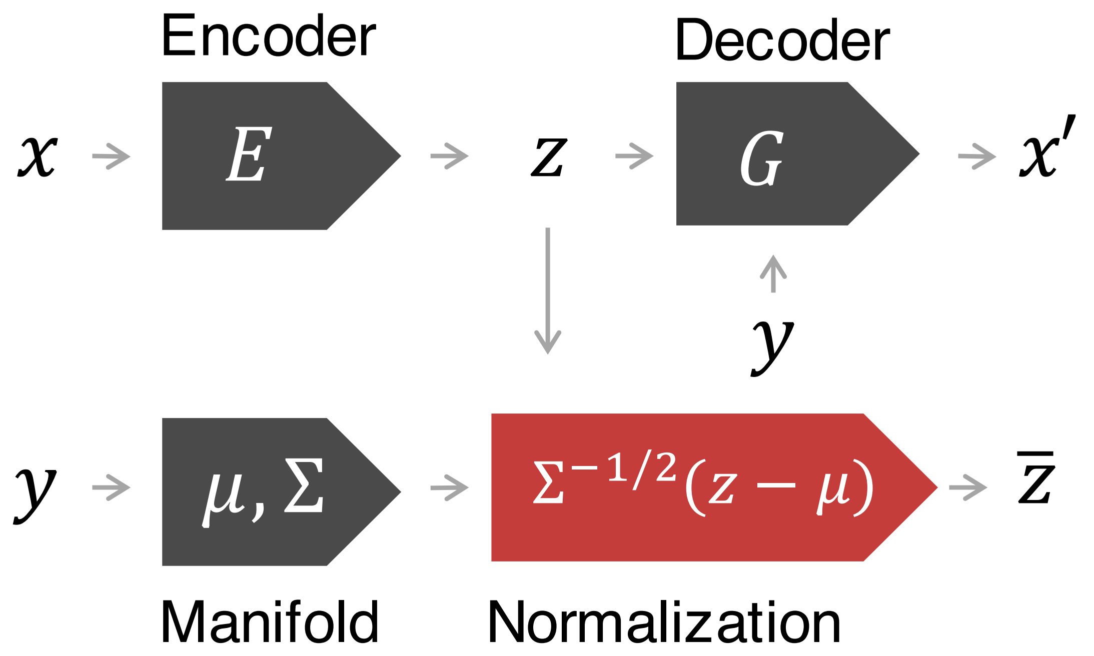
- Predictive: more stable
- Joint: can eliminate more mutual information

- Combined: use both techniques

Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

ENTANGLED

- Instead of disentanglement, learn $p(z|y) = \mathcal{N}(\mu(y), \Sigma(y))$
- Discriminator: $N\mathrm{F}\mu\mathrm{F}y\mathrm{G}, \Sigma\mathrm{F}y\mathrm{GG}$ vs real latent codes + y
- Problem: both distributions are learnable
- Idea: reparameterization! $N(0, 1)$ vs reparameterized codes + y



Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

EXPERIMENTS

- Condition on MACCS fingerprints (166 bits)

Disentanglement	Tanimoto, %	Hamming	Exact, %	Remaining MI
No	80.0	10.49	4.4	2.75
Predictive	86.2	7.13	11.4	0.64
Joint	88.7	5.78	17.4	1.56
Combined	91.8	4.18	27.8	0.32
Entangled, no Predictive	93.5	3.31	40.9	2.51
Entangled	**93.6**	**3.28**	**41.3**	1.30

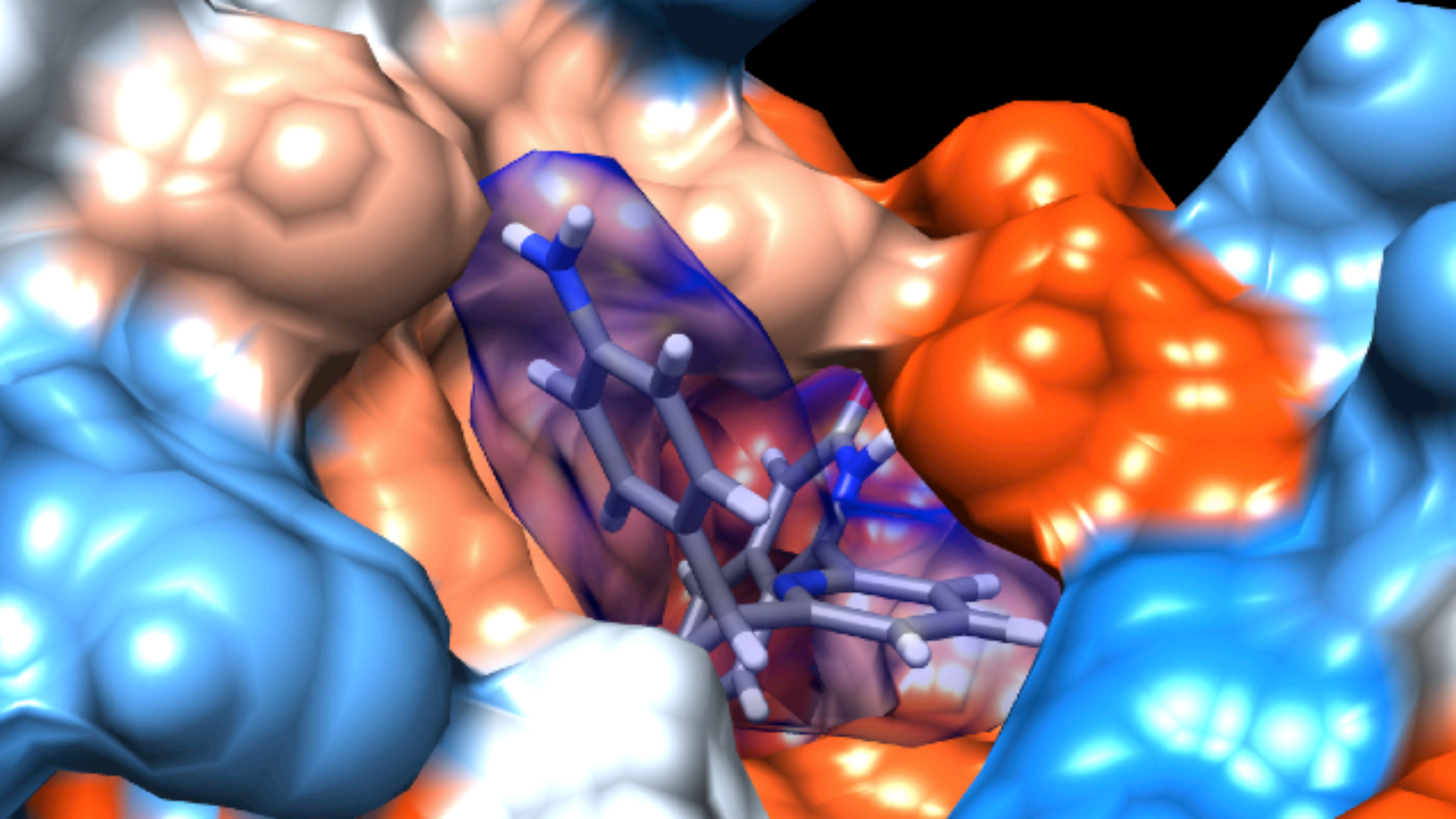
Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

EXPERIMENTS

- Condition on logP, SA, Docking

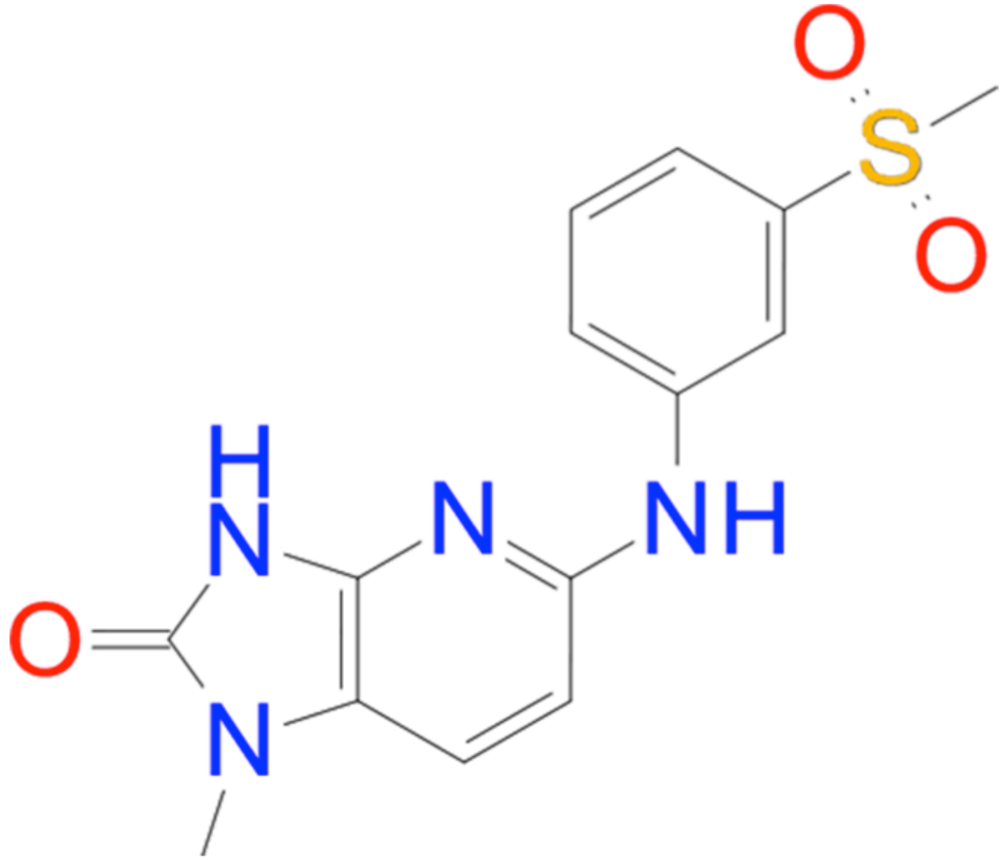
Disentanglement	logP, r	SA, r	E, r
No	0.311 ± 0.01	0.0522 ± 0.009	0.02 ± 0.04
Predictive	0.687 ± 0.006	0.0893 ± 0.008	0.063 ± 0.05
Joint	0.595 ± 0.007	0.0838 ± 0.008	0.109 ± 0.04
Combined	0.677 ± 0.007	0.0896 ± 0.007	0.116 ± 0.04
Entangled	$\mathbf{0.804 \pm 0.005}$	$\mathbf{0.593 \pm 0.007}$	$\mathbf{0.406 \pm 0.04}$

Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018



EXPERIMENTS

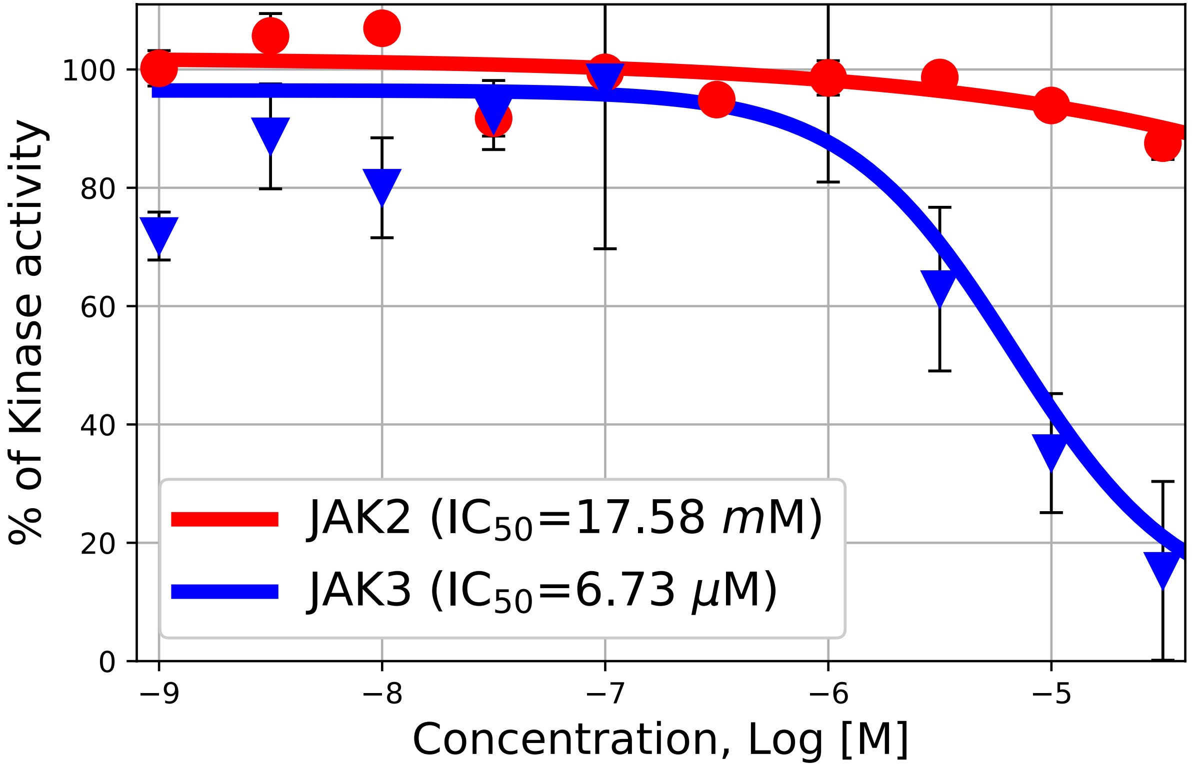
- JAK3 inhibitor (rheumatoid arthritis, psoriasis, and vitiligo)
- Specified high activity against JAK3 and low against JAK2

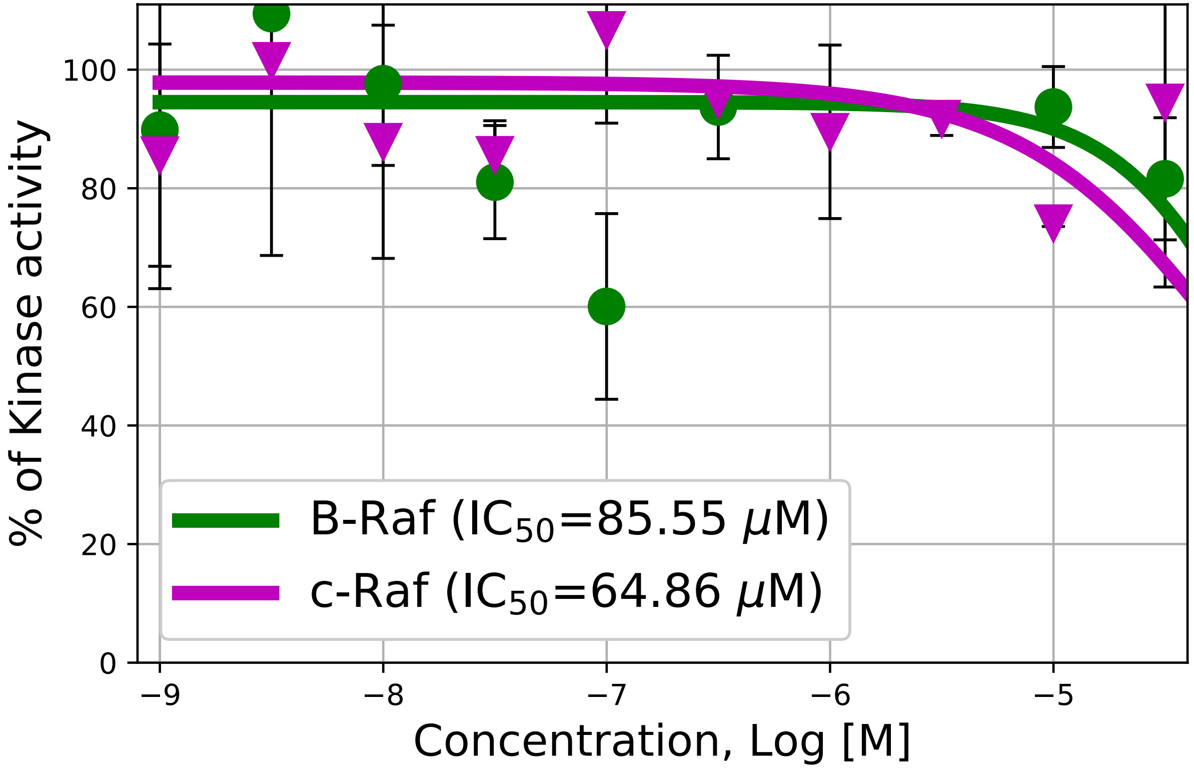


Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

EXPERIMENTS

- JAK3 inhibitor (rheumatoid arthritis, psoriasis, and vitiligo)





Polykovskiy et al, Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery, 2018

HOW TO DISCOVER A MOLECULE?

- How to generate molecules with given properties (e.g. given activity or synthetic accessibility)

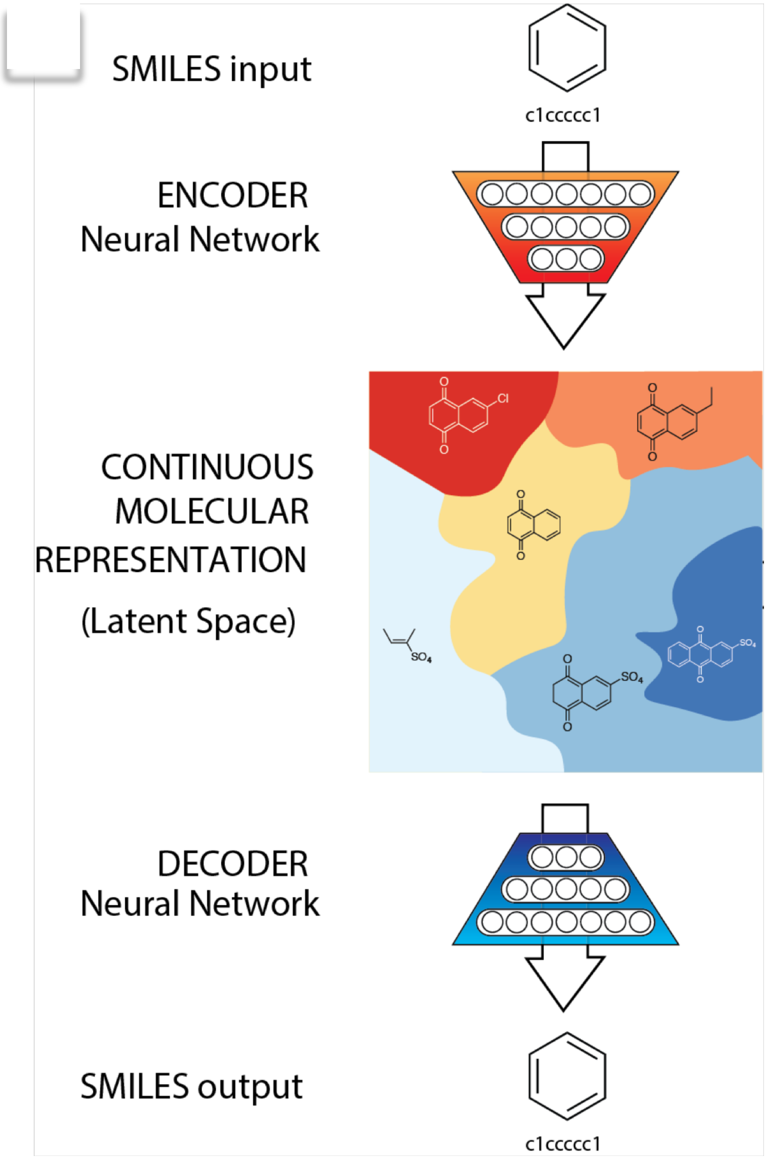
CONDITIONAL GENERATION

$$x \sim p(x \mid \text{properties})$$

OPTIMIZATION

$$\text{quality}(x) \to \max_x$$

OPTIMIZATION PERSPECTIVE



- Train a Variational Autoencoder (VAE)
- Latent space becomes “dense”
- We can travel around the latent space and reconstruct valid molecules from all points

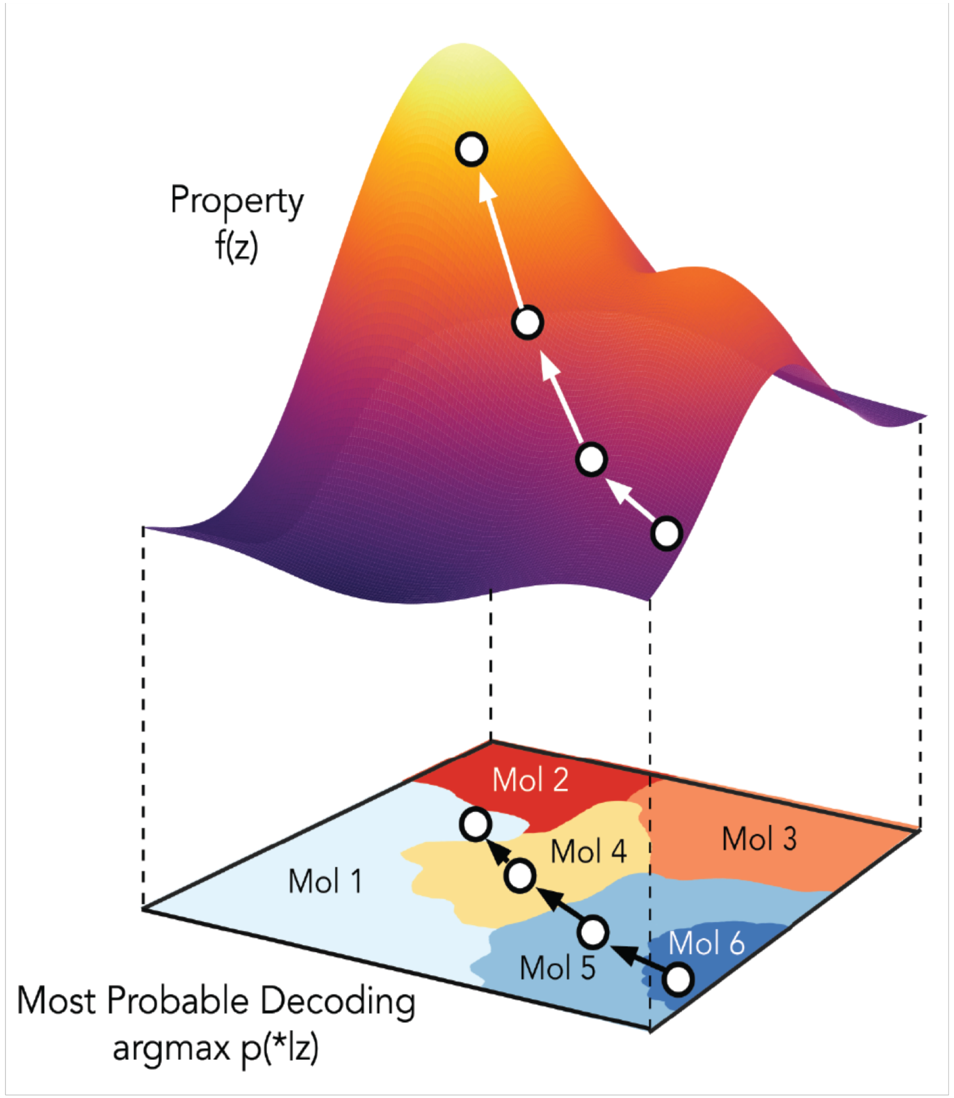
Gómez-Bombarelli, et al., Automatic chemical design using a data-driven continuous representation of molecules

OPTIMIZATION PERSPECTIVE

- Perform Bayesian Optimization (BO) on the latent space

- Synthesize sampled molecules and provide experimental results back to the model

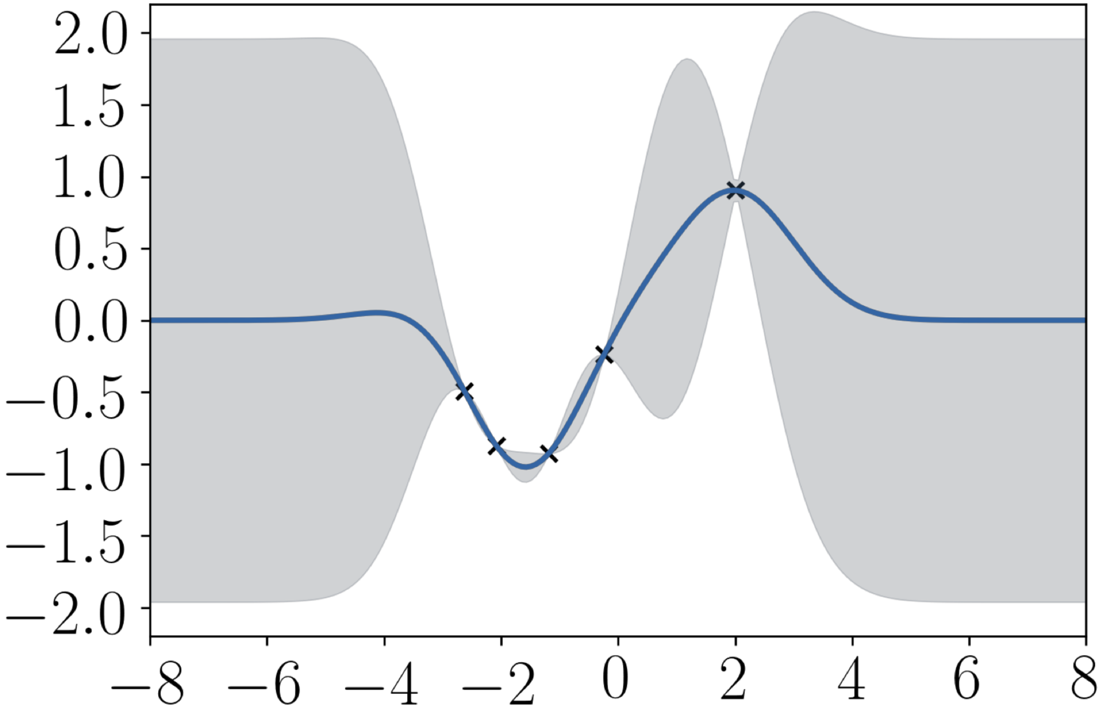
Objective: 5 × QED – SAS



Gómez-Bombarelli, et al., Automatic chemical design using a data-driven continuous representation of molecules

BAYESIAN OPTIMIZATION

- Used when evaluating a function is very expensive: e.g. hyperparameter search
- Train a predictive model based on known experimental values (usually a Gaussian Process)
- Estimate expected value in new points along with the confidence
- Test the point with the highest expected profit



REINFORCEMENT LEARNING

Agent

REINFORCEMENT LEARNING

Agent

Environment

REINFORCEMENT LEARNING

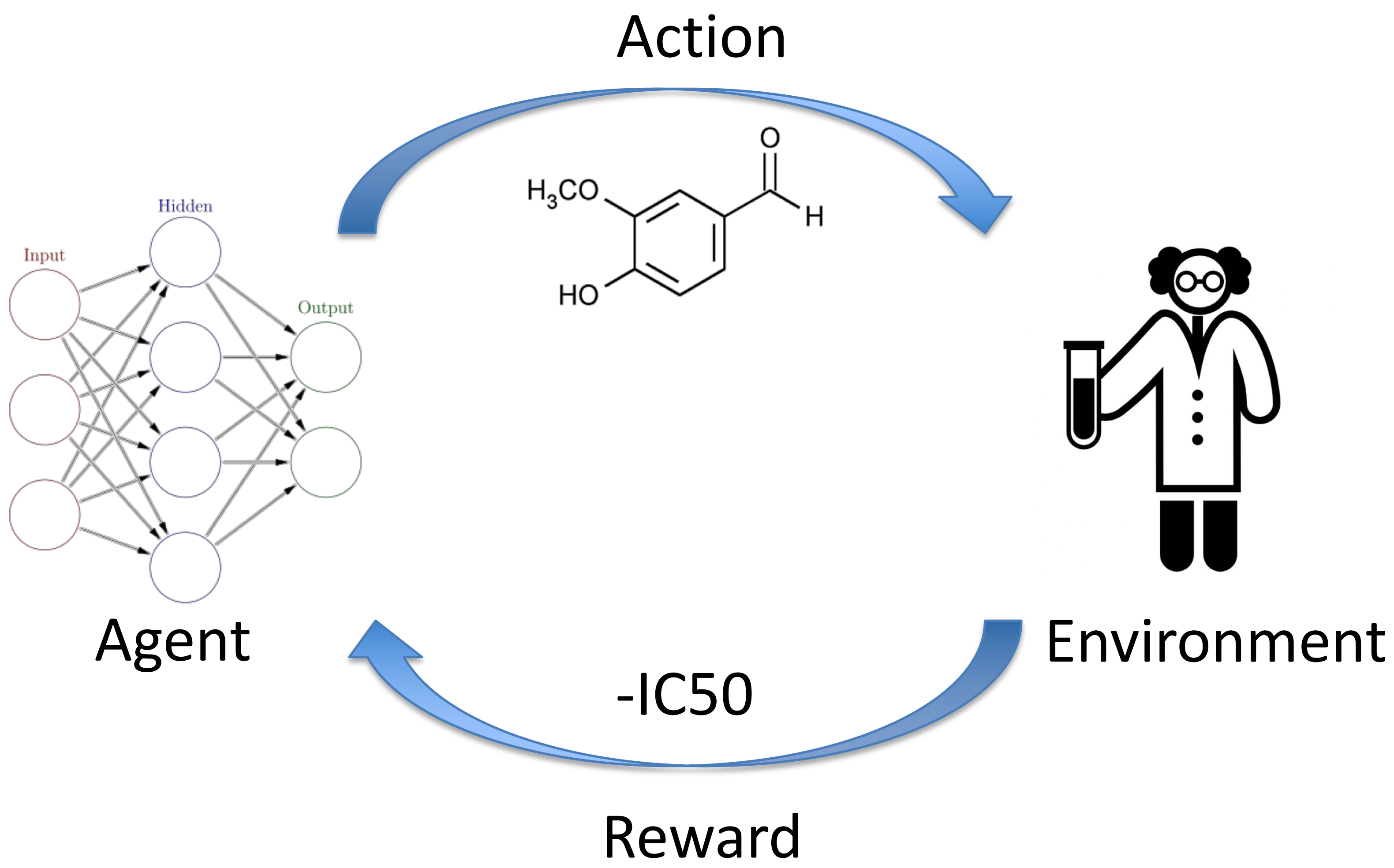
Action

Agent

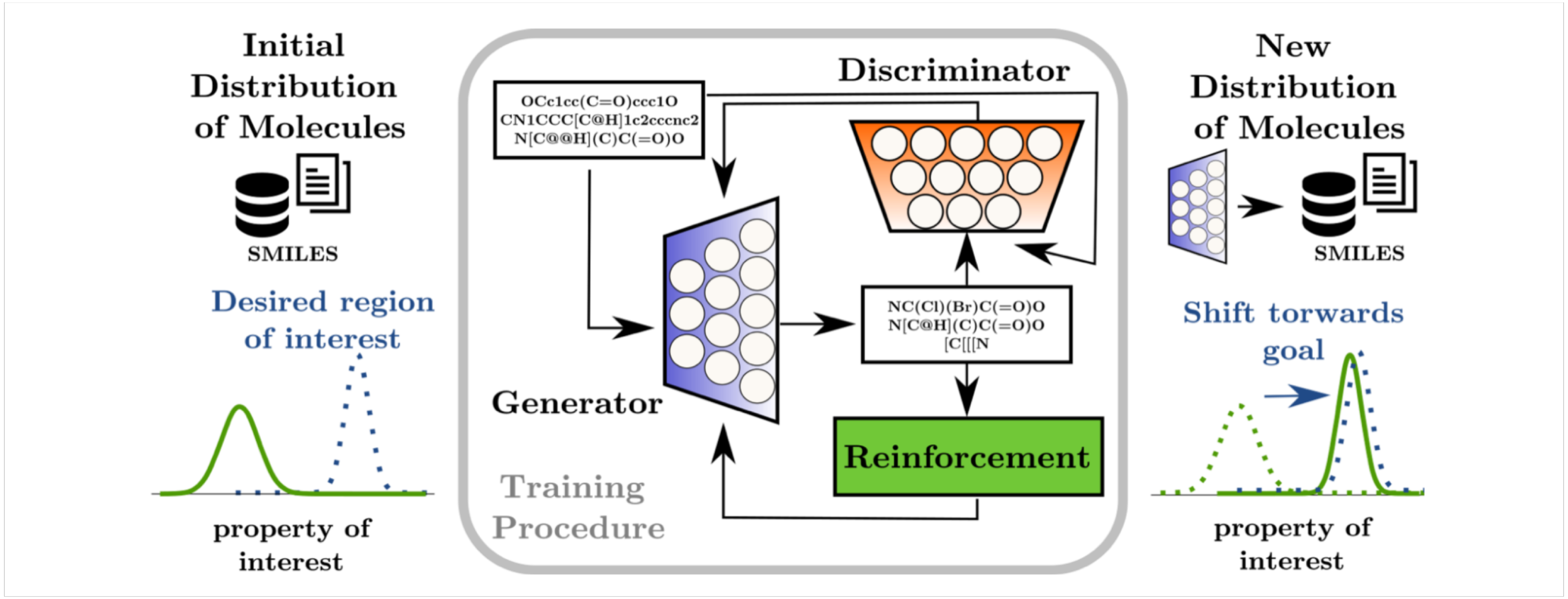
Environment

REINFORCEMENT LEARNING

REINFORCEMENT LEARNING



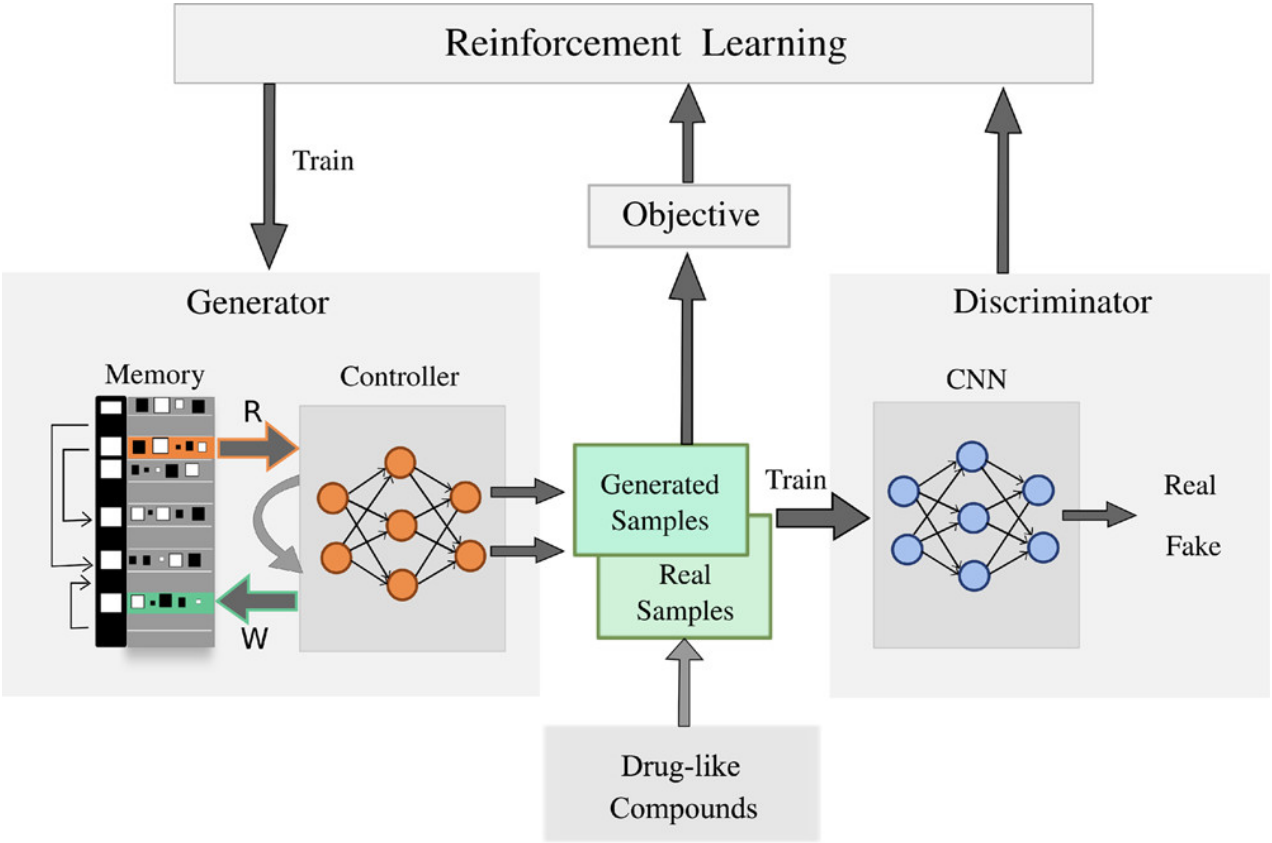
REINFORCEMENT LEARNING



- Guide a distribution of molecules towards more optimal structures using RL
- Reward: Synthetic Accessibility (SA), Rule of Five (RO5)

Sanchez-Lengeling, el al., Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC)

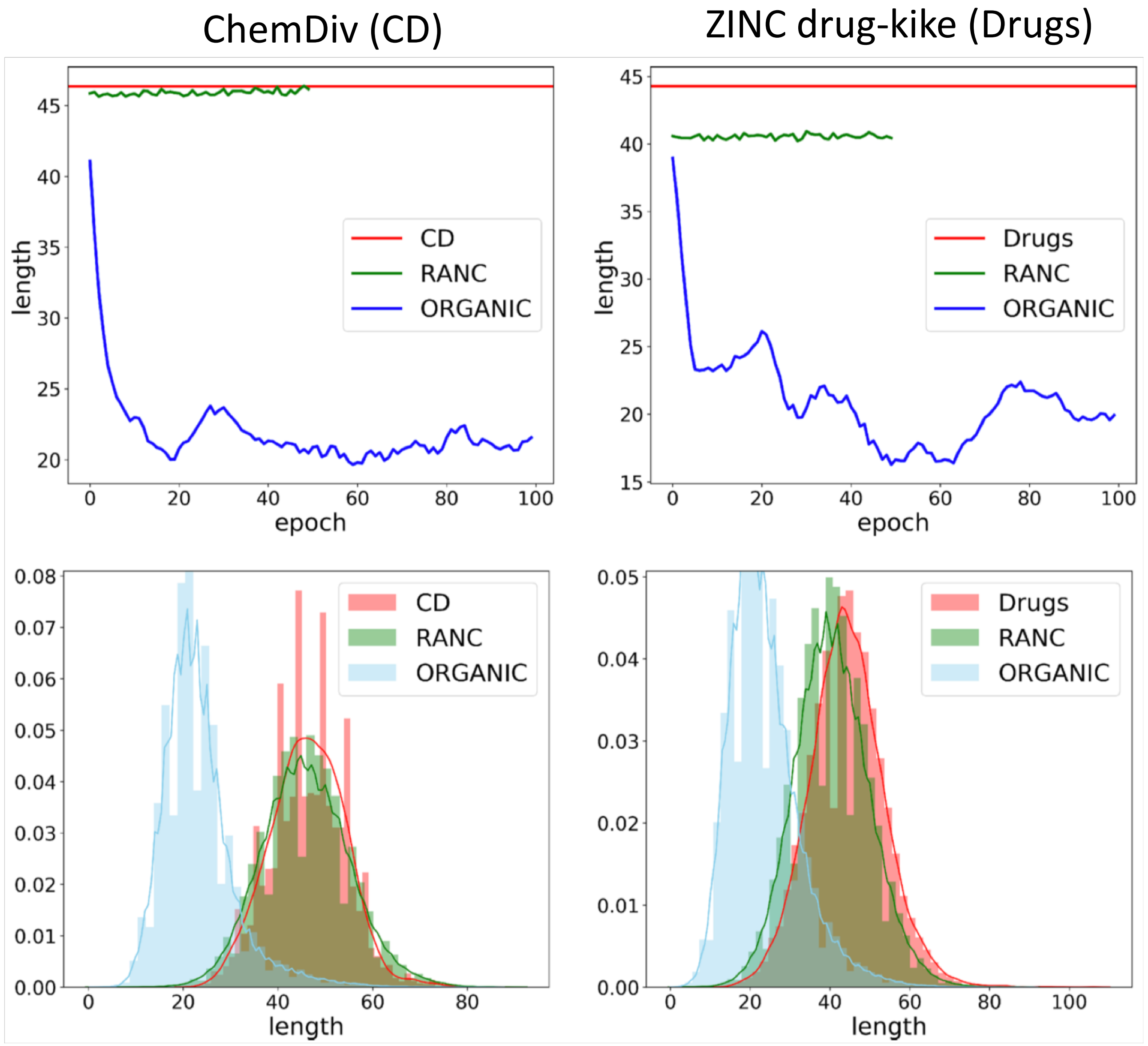
REINFORCED ADVERSARIAL NEURAL COMPUTER



- Differentiable Neural Computer is used as a Generator
- Molecules are optimized using Reinforcement Learning, optimizing predefined reward

Putin et al, Reinforced Adversarial Neural Computer for de Novo Molecular Design

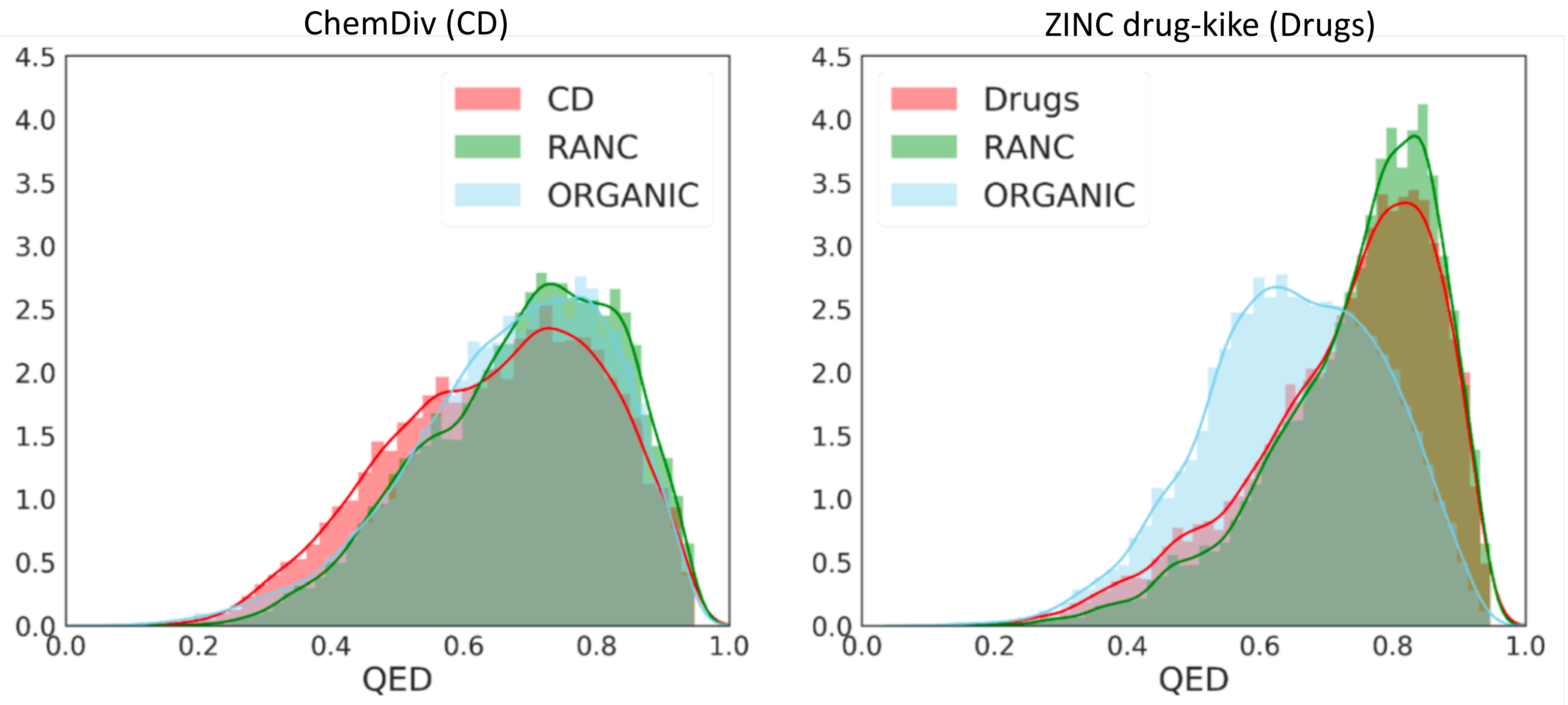
REINFORCED ADVERSARIAL NEURAL COMPUTER



Putin et al, Reinforced Adversarial Neural Computer for de Novo Molecular Design

REINFORCED ADVERSARIAL NEURAL COMPUTER

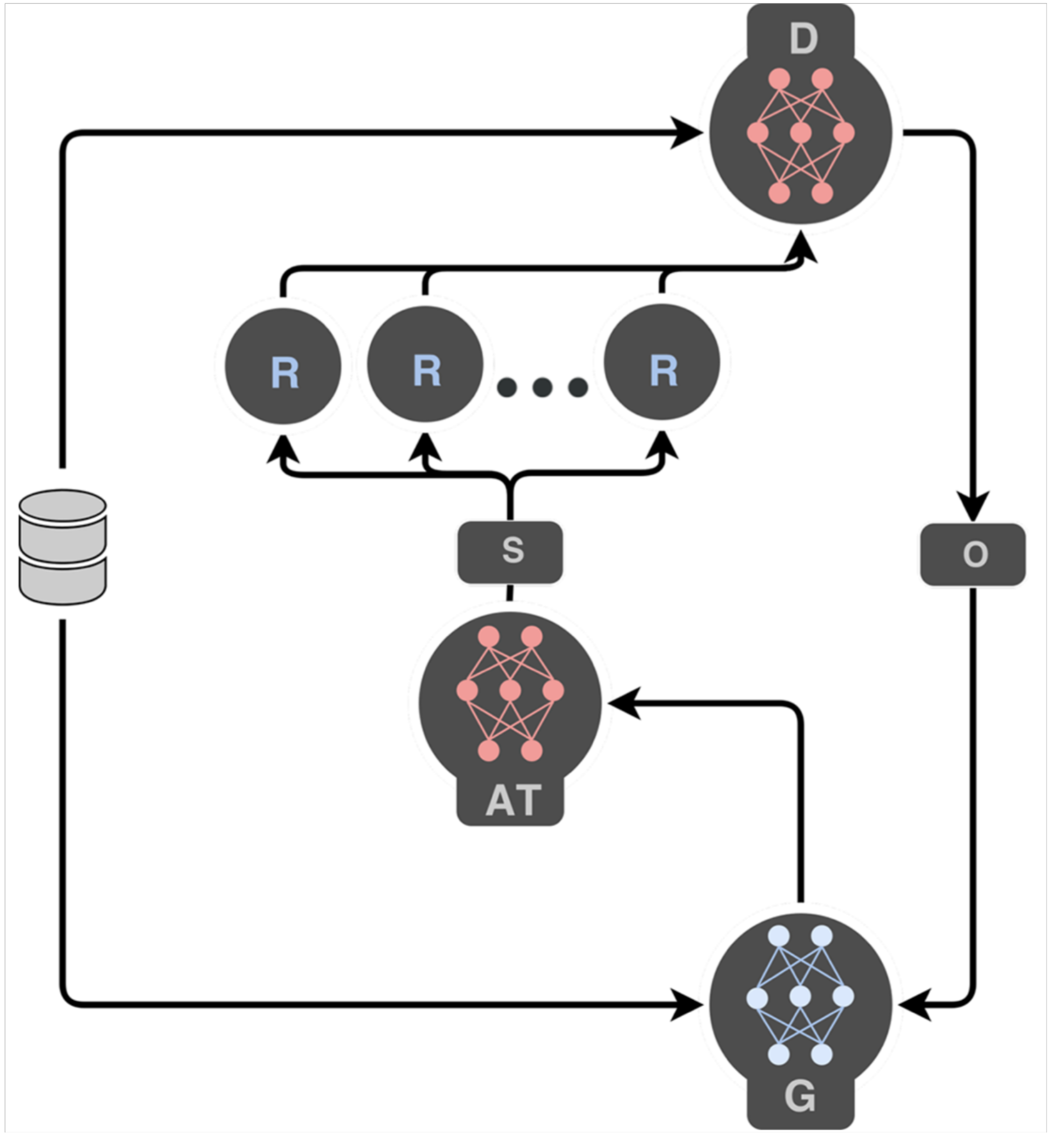
Quantitative Estimate of Drug-likeness



Putin et al, Reinforced Adversarial Neural Computer for de Novo Molecular Design

ADVERSARIAL THRESHOLD NEURAL COMPUTER

- Adversarial threshold is a Discriminator with a fixed lag
- Adversarial threshold selects molecules with highest predicted reward
- Only selected molecules are evaluated to provide gradient



Putin et al, Adversarial Threshold Neural Computer for Molecular de Novo Design

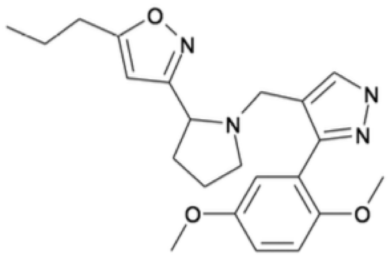
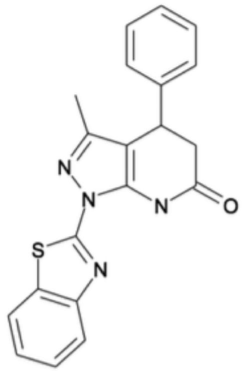
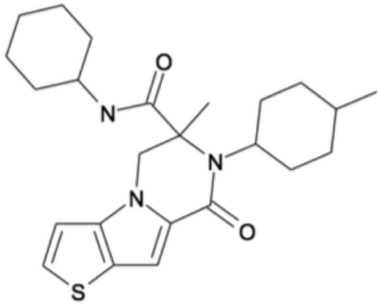
ADVERSARIAL THRESHOLD NEURAL COMPUTER

Setup:

- Train on known kinase inhibitors
- Generate many compounds using ATNC
- Retain **50** most promising molecules
- Intersect with available in-house molecules

- Test selected compounds *in vitro* against a set multiple kinases
- Measure Inhibition % at 10μM

ADVERSARIAL THRESHOLD NEURAL COMPUTER

Compound	Kinase (example)	Inhibition (%)	Supplier
	SGK1	101	AKos Consulting & Solutions AKOS005014423
	Aurora A/B	95	AKos Consulting & Solutions AKOS022051824
	SGK1	99	AKos Consulting & Solutions AKOS004939857

Putin et al, Reinforced Adversarial Neural Computer for de Novo Molecular Design

CHALLENGING TARGETS

CHALLENGING TARGETS

- Filter molecules using deterministic rules like RO5, MCF
- Eliminate molecules with bad predicted activity and other key indicators

RESULT

CHALLANGES

- Novelty
  - generate something previously unseen

CHALLANGES

- Novelty
  - generate something previously unseen
- Diversity
  - generate different molecules

CHALLANGES

- Novelty
  - generate something previously unseen
- Diversity
  - generate different molecules
- Activity
  - should produce good molecules, not good metrics

DEEP LEARNING FOR DRUG DISCOVERY

Insilico Medicine, Inc
Emerging Technology Centers
Johns Hopkins University
9601 Medical Center Dr, Suite 127
Rockville, MD 20850

- **Drug Discovery**
- **Drug Repurposing**
- **Biomarker Development**
- **Clin. Trials Predictors**
- **Aging Research**

INSILICO MEDICINE

Daniil Polykovskiy
daniil@insilico.com

www.insilico.com