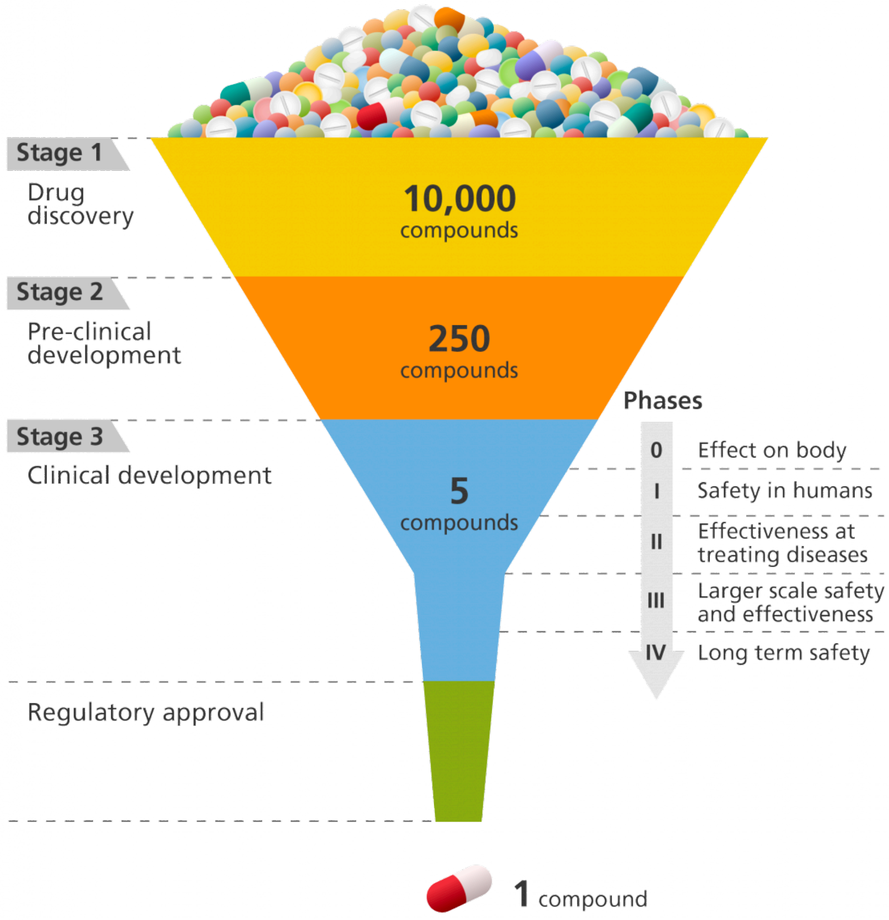
LIMO: Latent Inceptionism for Targeted Molecule Generation

Peter Eckmann, Kunyang Sun, Bo Zhao, Mudong Feng, Michael K. Gilson, Rose Yu
UC San Diego

ICML 2022
ID: 16741
Session 5 Track 3
Poster: 5:45pm PDT

Introduction

- Central goal of drug discovery is finding drug-like molecules with high binding affinity to a target protein
- Currently, identifying new molecules is a resource-intensive process, involving large and inefficient experimental compound screens
- Generative models suggest new compounds to synthesize and test
- We present a novel approach to targeted molecule generation called Latent Inceptionism on Molecules (LIMO), a fast generative modeling framework for *de novo* design



Related work

Reinforcement learning (RL)

- Directly optimizes molecular properties by systematically altering or constructing a molecular graph (e.g. GCPN[1], GraphDF[2])
- Effective at generating molecules with desired properties
- Prohibitively slow when optimizing for expensive-to-compute properties, e.g. binding affinity

Generative modeling

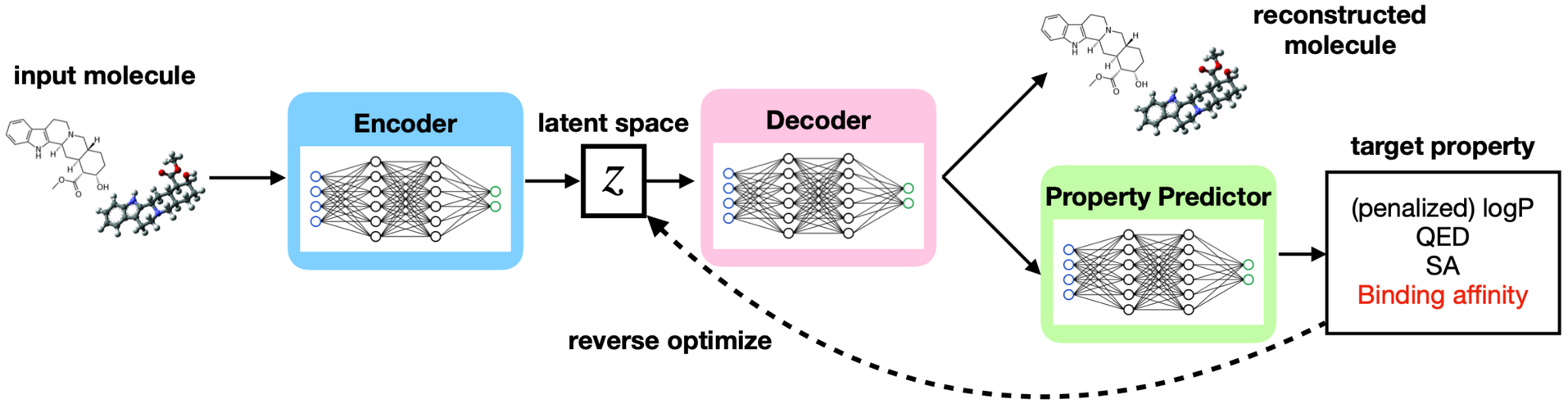
- Uses a surrogate model of molecular properties on top of a learned latent space mapping to discrete molecular structures (JT-VAE[3])
- Very fast at generating molecules
- Technique has so far not reached the property scores of RL-generated molecules

[1]You, J., Liu, B., Ying, R., Pande, V., and Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *32nd Conference on Neural Information Processing Systems*, 2018.

[2]Luo, Y., Yan, K., and Ji, S. GraphDF: A discrete flow model for molecular graph generation. In *International Conference on Machine Learning*. PMLR 139, 2021b.

[3]Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In *Proceedings of the 35th International Conference on Ma- chine Learning*, 2018.

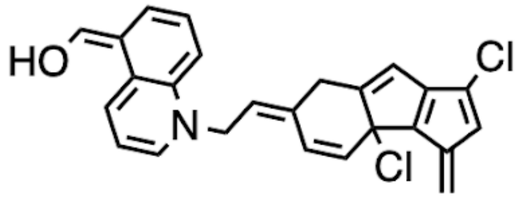
Methodology



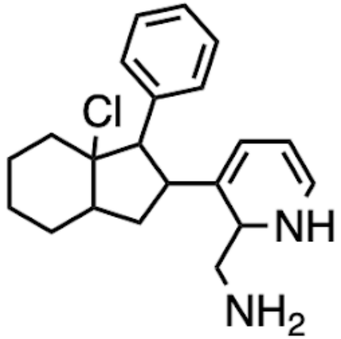
- Directly optimize in the **latent space** to generate new molecules that have a variety of target properties
- Instead of predicting properties from the latent space, LIMO uses a **property predictor** stacked *on top of* the VAE decoder
- Fully differentiable, LIMO allows for more targeted molecule generation and accurate property prediction

Target Protein

ESR1

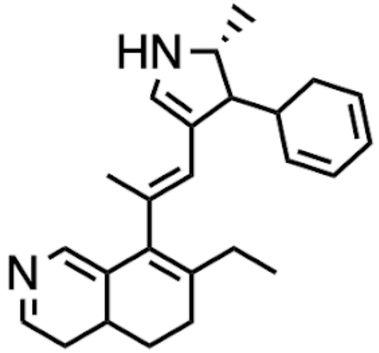


$K_D = 4.6 \cdot 10^{-9}$ M

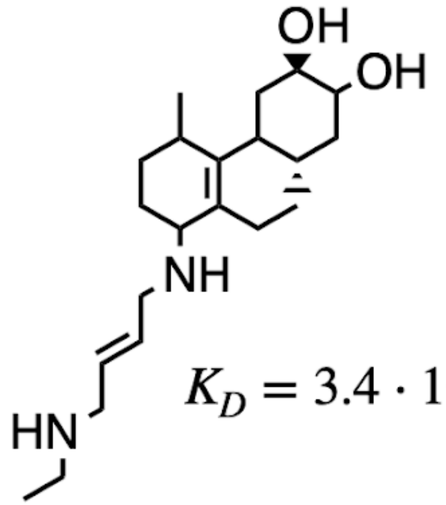


$K_D = 2.8 \cdot 10^{-9}$ M

ACAA1



$K_D = 5.0 \cdot 10^{-8}$ M



$K_D = 3.4 \cdot 10^{-8}$ M

- **Human estrogen receptor (ESR1)**: a target of drugs used to treat breast cancer, has many known binders. Used a crystal structure of the protein (PDB 1ERR) for docking calculations and the location of the binding site.
- **Human peroxisomal acetyl-CoA acyl transferase 1 (ACAA1)**: challenge from Structural Genomics Consortium, no known binders but does have a crystal structure (PDB 2IIK) with a potential drug-binding pocket.

Experiments

Single-objective binding affinity optimization

Method	ESR1			ACAA1			Time
	1st	2nd	3rd	1st	2nd	3rd	(hrs)
GCPN	6.4	6.6	8.5	75	83	84	6
MolDQN	373	588	1062	240	337	608	6
GraphDF	25	47	51	370	520	590	12
MARS	17	64	69	163	203	236	6
LIMO	**0.72**	**0.89**	**1.4**	**37**	**37**	**41**	**1**

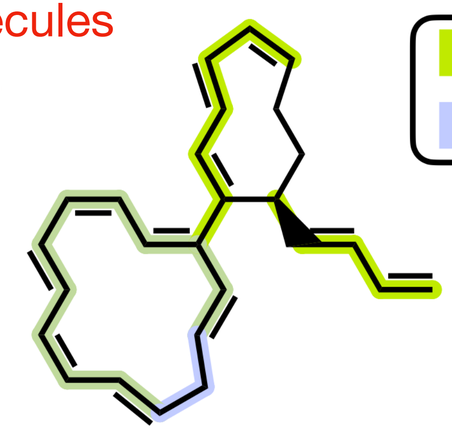
- **Task:** generate molecules with high binding affinity to a target protein, computed by a physics-based affinity estimator (AutoDock-GPU)
  - Metric: K_D, a measure of binding affinity, lower is better
  - Human estrogen receptor (ESR1) and peroxisomal acetyl-CoA acyl transferase 1 (ACAA1) are disease-relevant proteins

Experiments

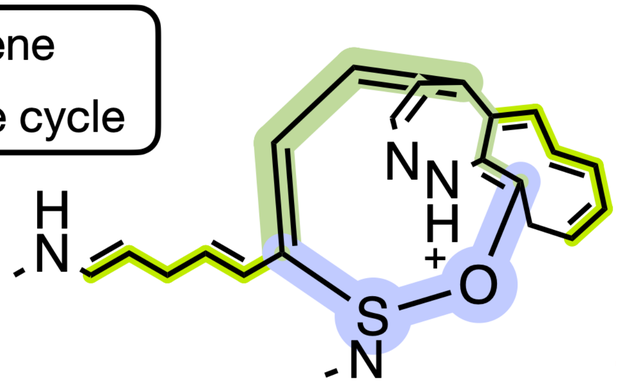
Single-objective binding affinity optimization

- Examples of generated molecules:

Generated molecules are toxic…



$K_D = 3.7 \cdot 10^{-8}$ M



$K_D = 4.1 \cdot 10^{-8}$ M

…and hard to synthesize!

Experiments

Multi-objective binding affinity optimization

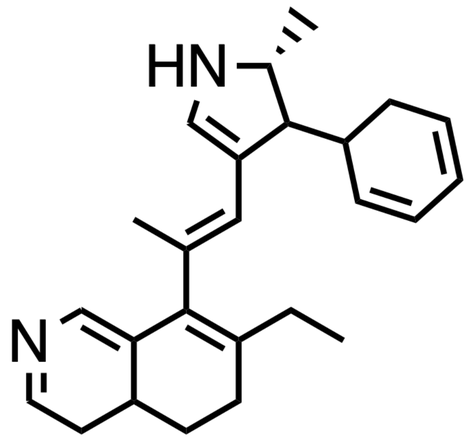
Ligand	Optimized prop.			Non-optimized prop.				
	K_D (AD) (↓)	QED (↑)	SA (↓)	K_D (ABFE) (↓)	Lipinski	PAINS (↓)	Fsp^3 (↑)	MCE-18 (↑)
ESR1								
LIMO mol. #1	4.6	0.43	4.8	$\mathbf{6 \cdot 10^{-5}}$	✓	0	0.16	90
LIMO mol. #2	**2.8**	0.64	4.9	1000	✓	0	0.52	76
GCPN mol. #1	810	0.43	4.2	-	✓	0	0.29	22
GCPN mol. #2	$2.7 \cdot 10^4$	0.80	3.7	-	✓	0	0.56	47
Tamoxifen	87	0.45	2.0	1.5*	✓	0	0.23	16
Raloxifene	$7.9 \cdot 10^6$	0.32	2.4	0.030*	✓	0	0.25	59
ACAA1								
LIMO mol. #1	**28**	0.57	5.5	$\mathbf{4 \cdot 10^4}$	✓	0	0.52	52
LIMO mol. #2	31	0.44	4.9	No binding	✓	0	0.81	45
GCPN mol. #1	8500	0.69	4.2	-	✓	0	0.52	61
GCPN mol. #2	8500	0.54	4.3	-	✓	0	0.52	30

- **Solution:** incorporate measures of molecule quality (synthesizability and drug-likeness) into optimization process

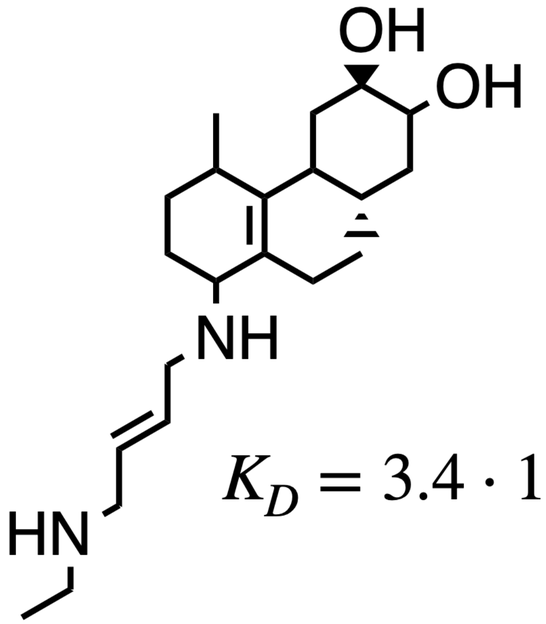
Experiments

Multi-objective binding affinity optimization

New molecules are more
synthesizable and drug-like!



$K_D = 5.0 \cdot 10^{-8}$ M

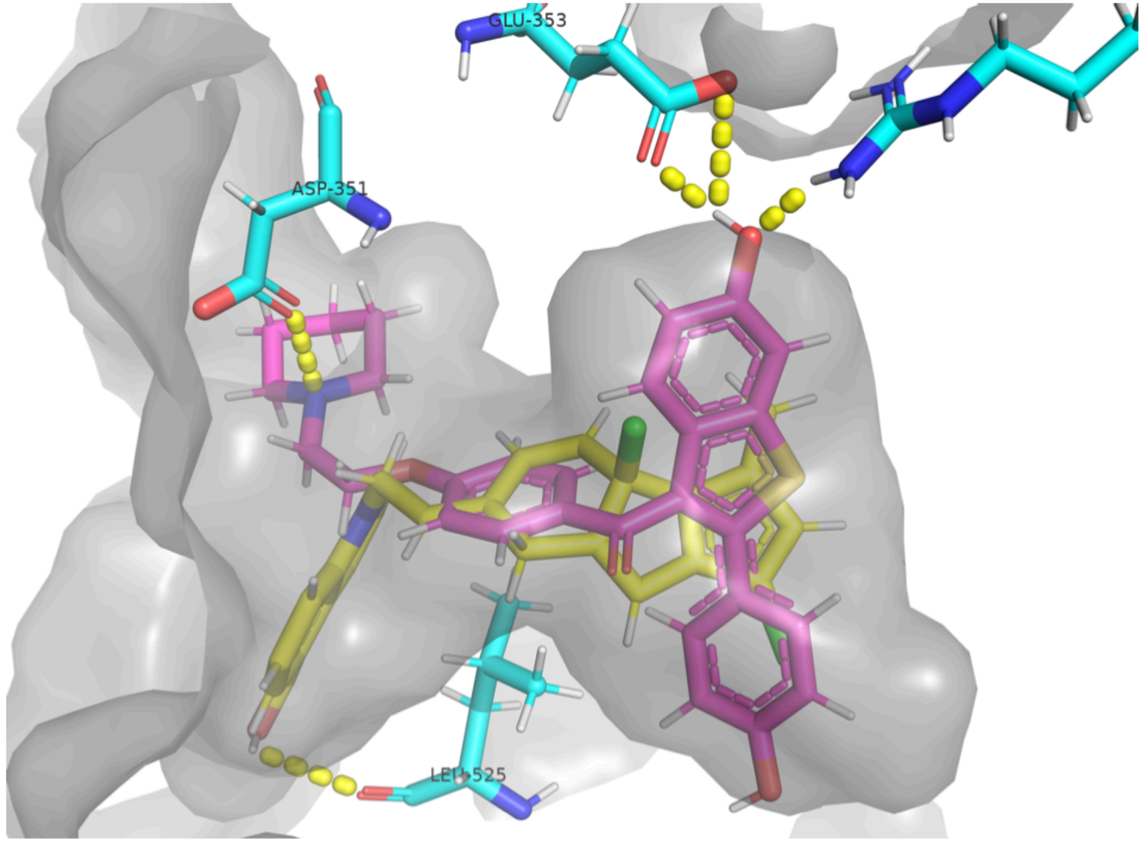


$K_D = 3.4 \cdot 10^{-8}$ M

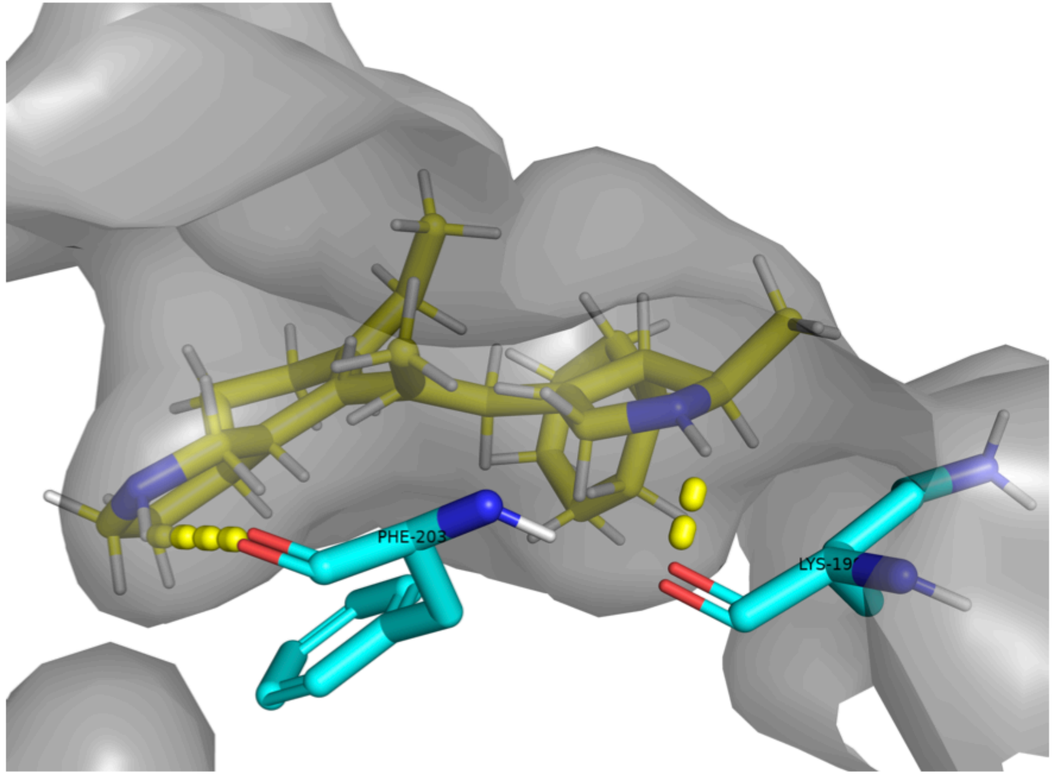
Experiments

Multi-objective binding affinity optimization

Human estrogen receptor



Human peroxisomal acetyl-CoA acyl transferase 1



Takeaways

- Two neural networks (decoder and property predictor) in sequence enables faster gradient-based reverse-optimization of molecular properties.
- Multi-objective optimization including drug-likeness and synthesizability is important for molecule generation with high binding affinity

Code available at https://github.com/Rose-STL-Lab/LIMO

Thank you for your attention!

LIMO: Latent Inceptionism for Targeted Molecule Generation
ICML 2022