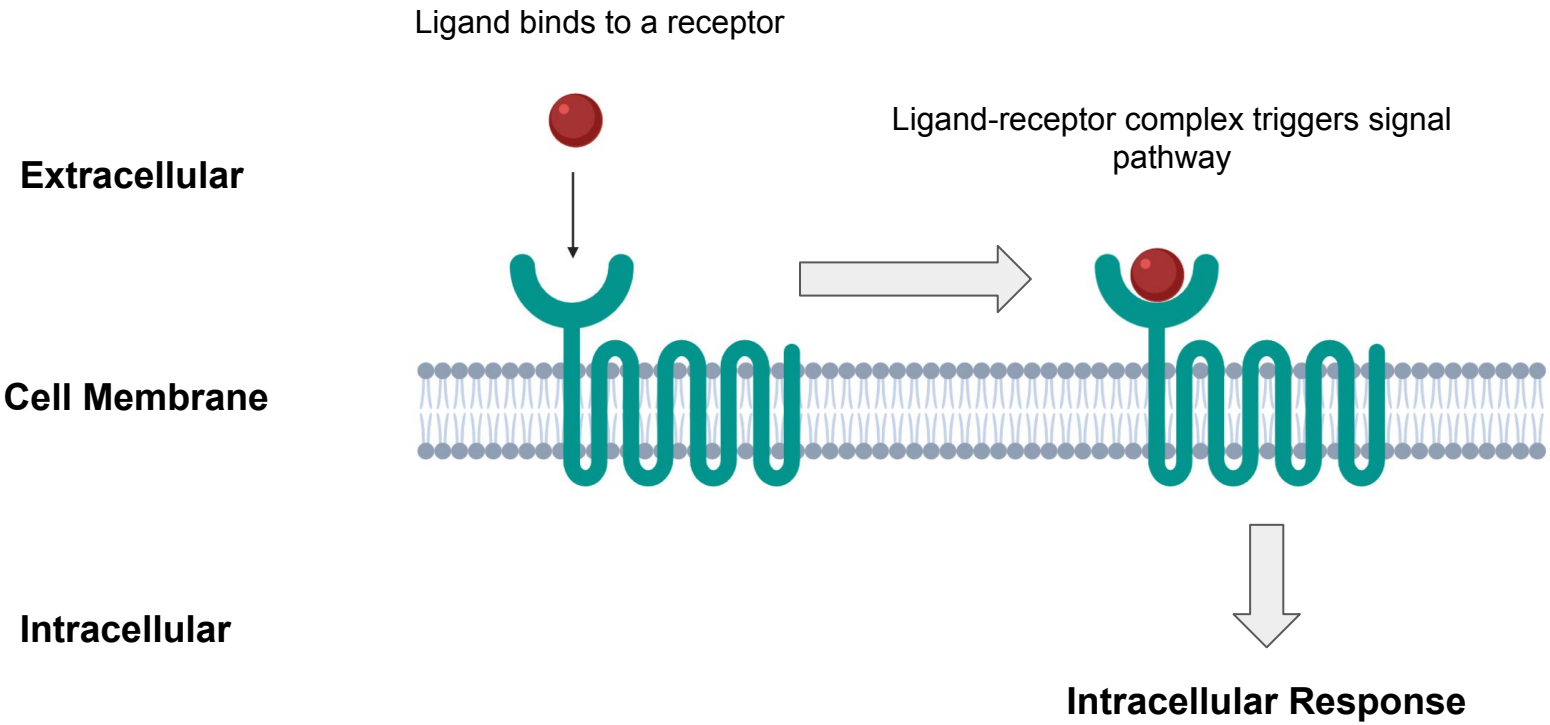
Latent representations of chemical ligands to predict combinatorial receptor-ligand interactions

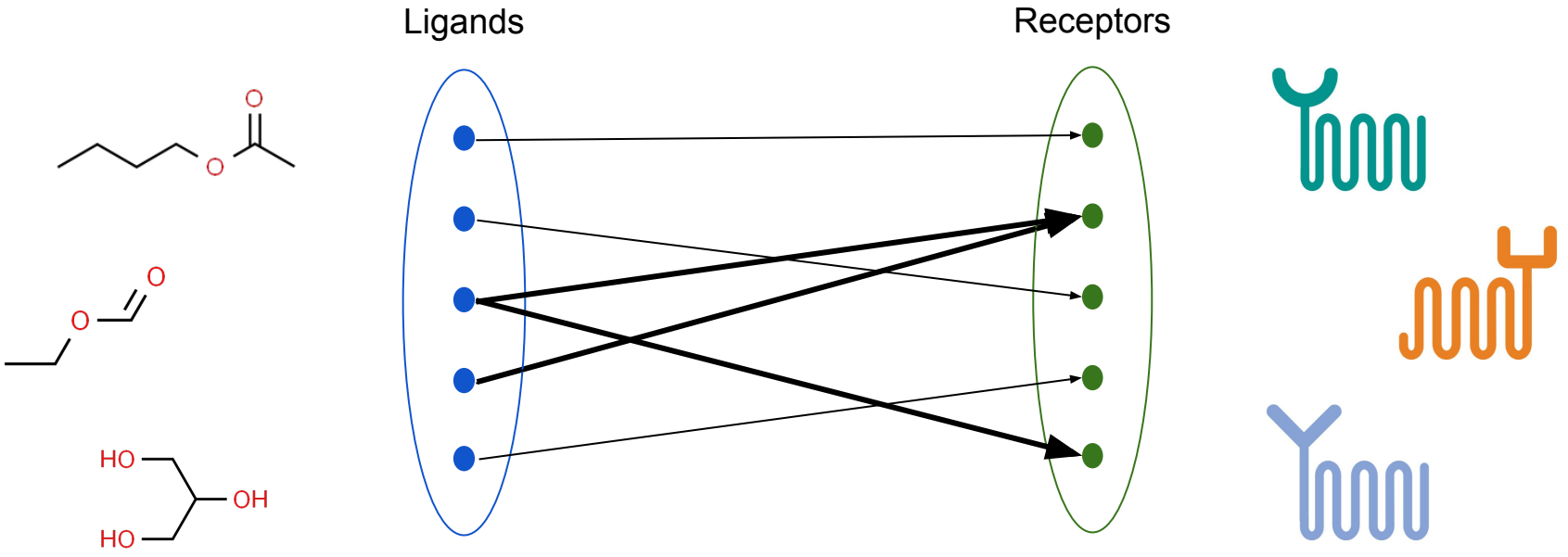
Jonathan Yin
Mentors: Dr. Hattie Chung, Michael Truell, Regev Group, Broad Institute

MIT PRIMES Conference, October 20, 2019

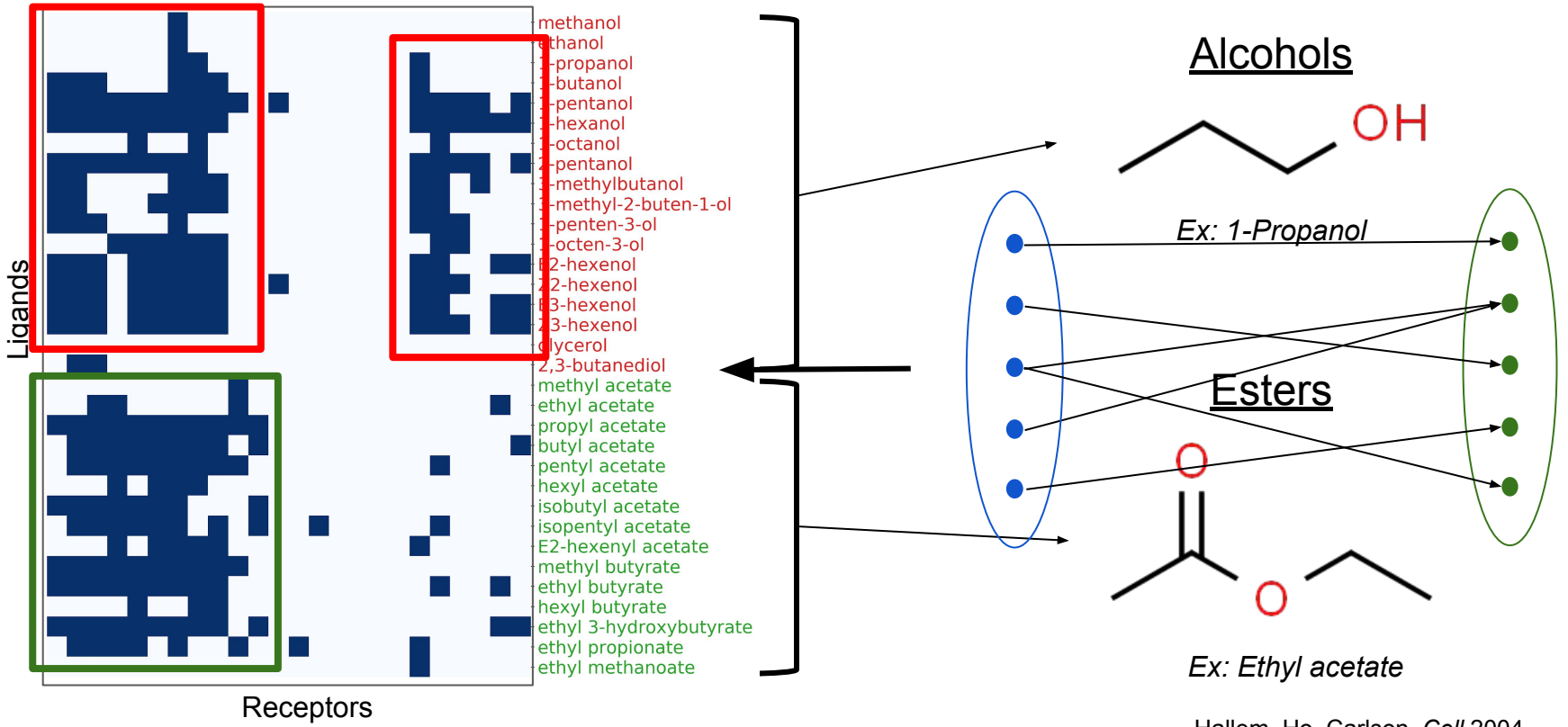
Receptors process signals from the environment



Understanding combinatorial receptor-ligand interactions

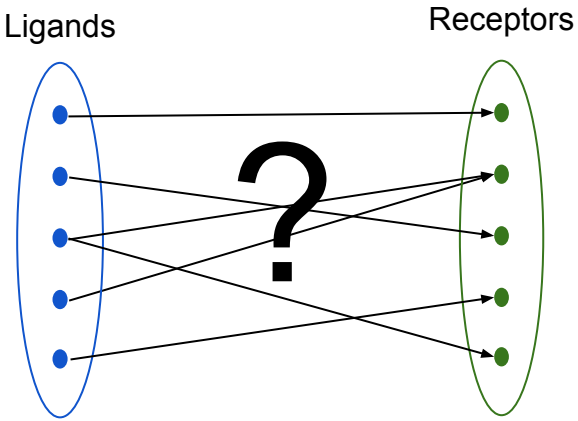


Similar chemicals activate similar sets of receptors



Hallem, Ho, Carlson. *Cell* 2004.

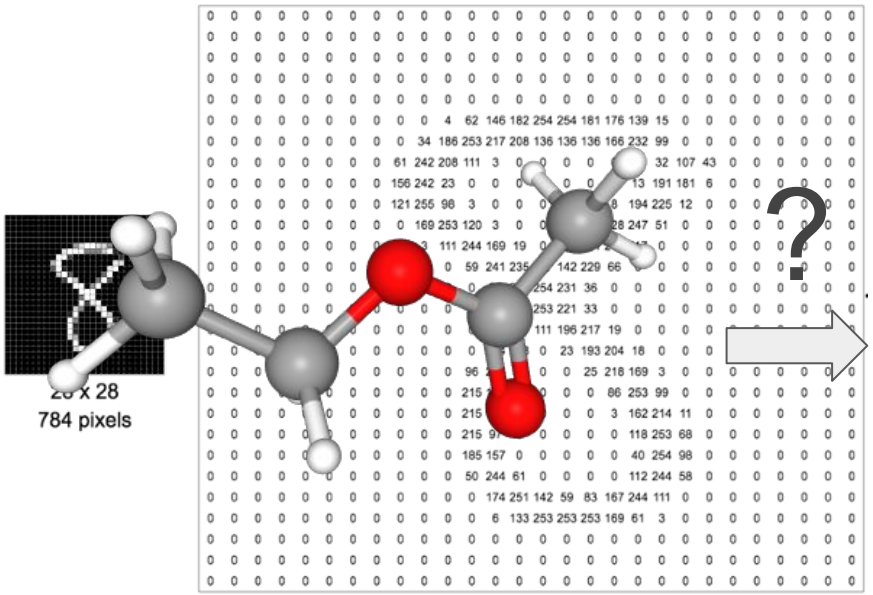
Motivation



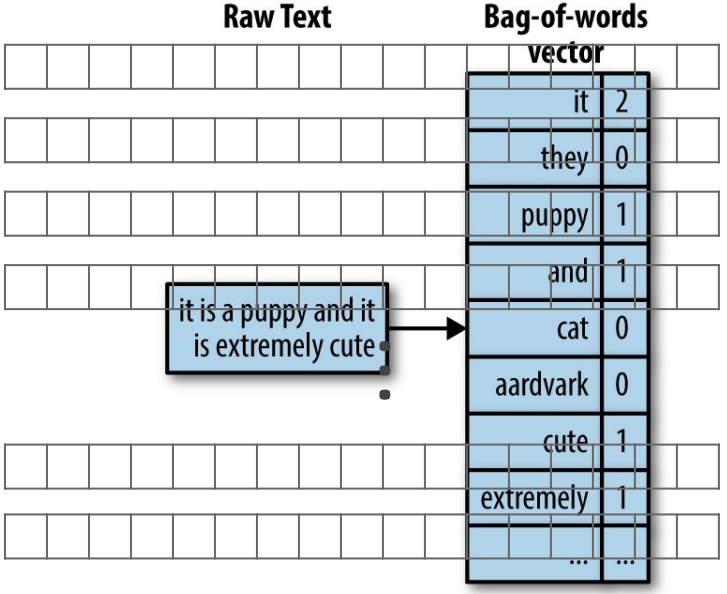
Predicting ligand-receptor interactions

Ability to control intracellular behaviors

Challenge: representing inputs for predictive algorithms

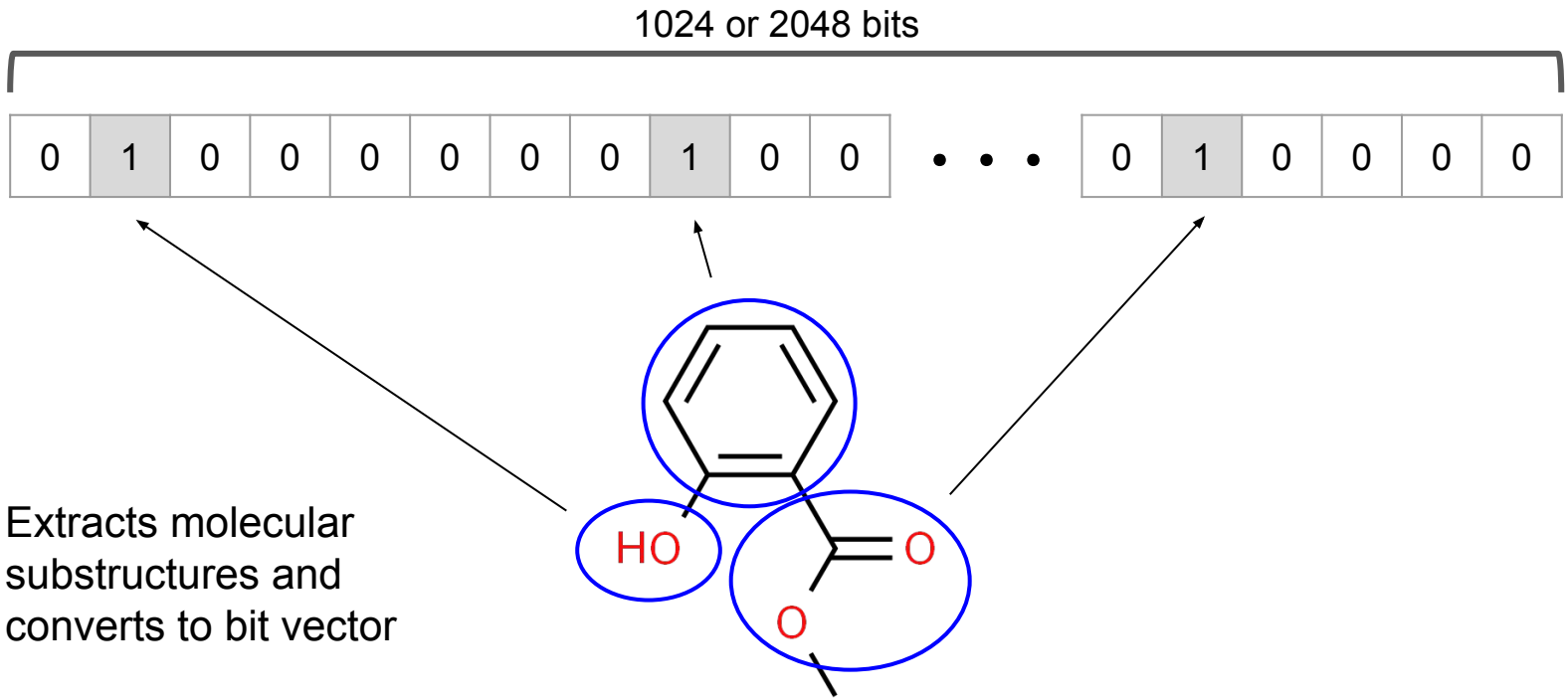


Images



Text

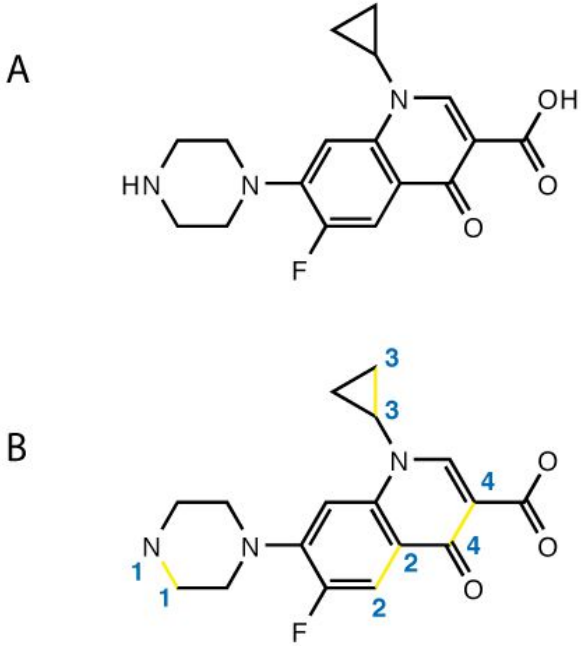
Low-level representation schemes: molecular fingerprints



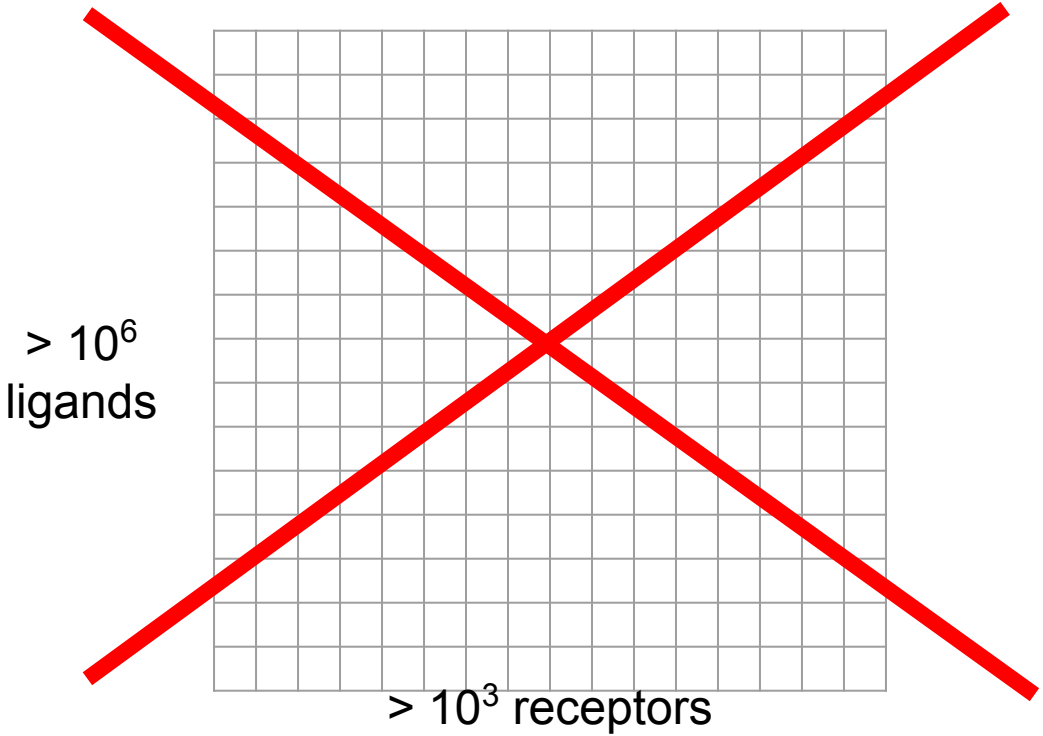
Extracts molecular substructures and converts to bit vector

Low-level representation schemes: SMILES

Simplified Molecular-Input Line-Entry System

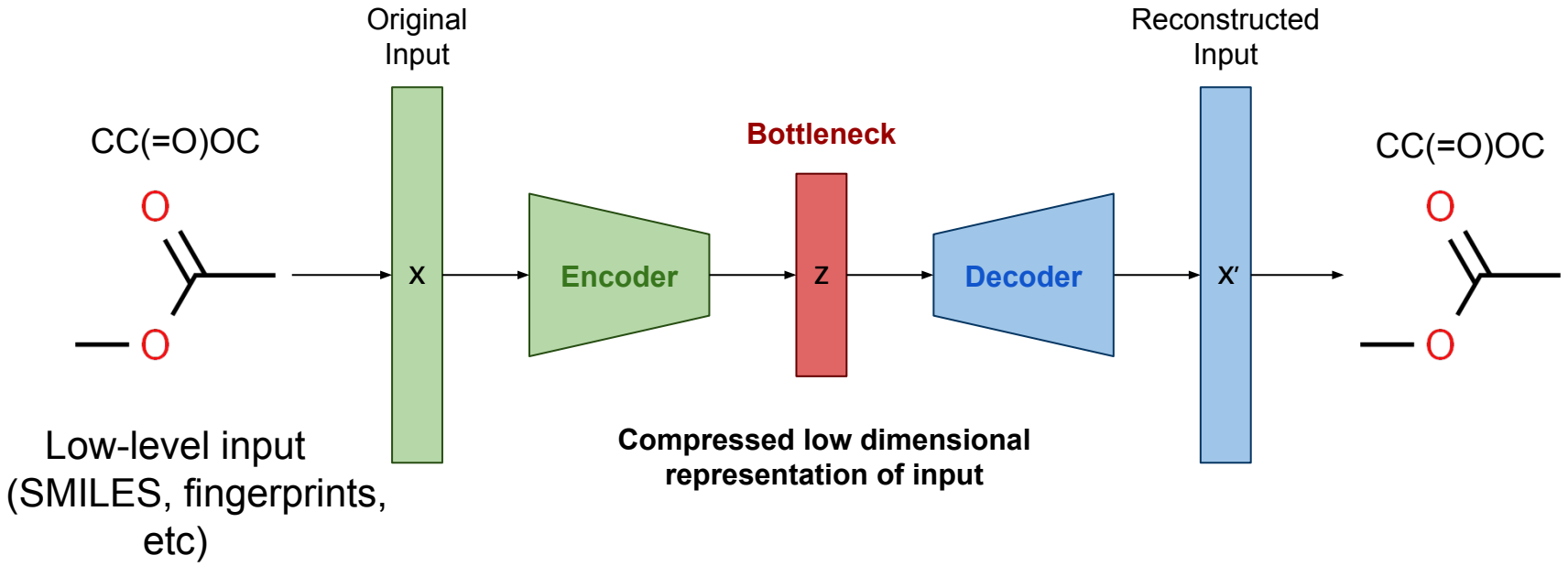


Machine learning models typically require high quantities of data



In the absence of a large dataset, feature abstraction is necessary

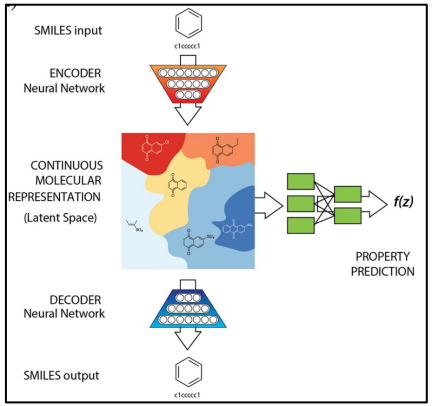
Recent trend: feature abstraction with variational autoencoders (deep neural network)



Current models

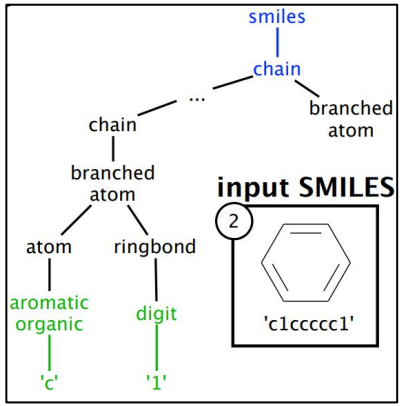
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Gómez-Bombarelli et al. 2016



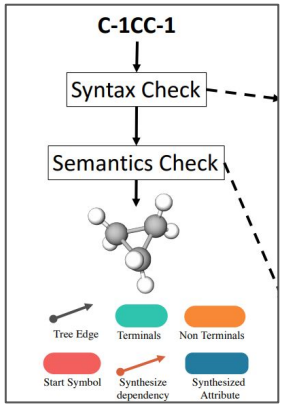
Grammar Variational Autoencoder

Kushner et al. 2017



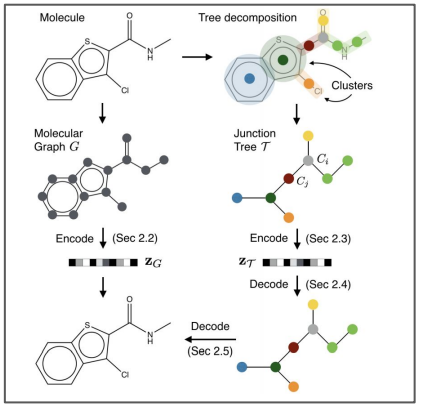
Syntax-Directed Variational Autoencoder for Structured Data

Dai et al. 2018

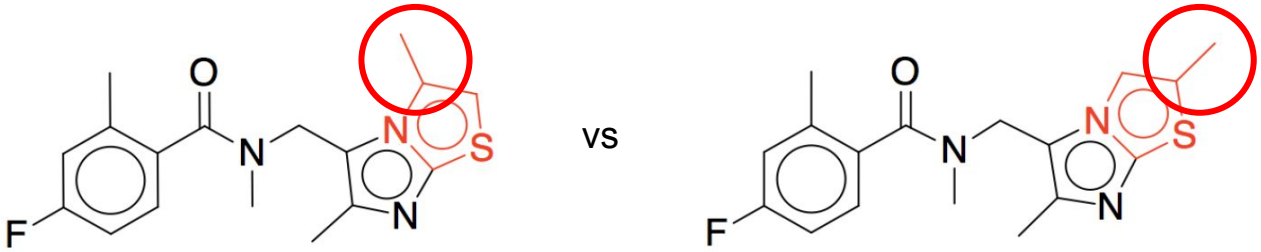


Junction Tree Variational Autoencoder for Molecular Graph Generation

Jin et al. 2018



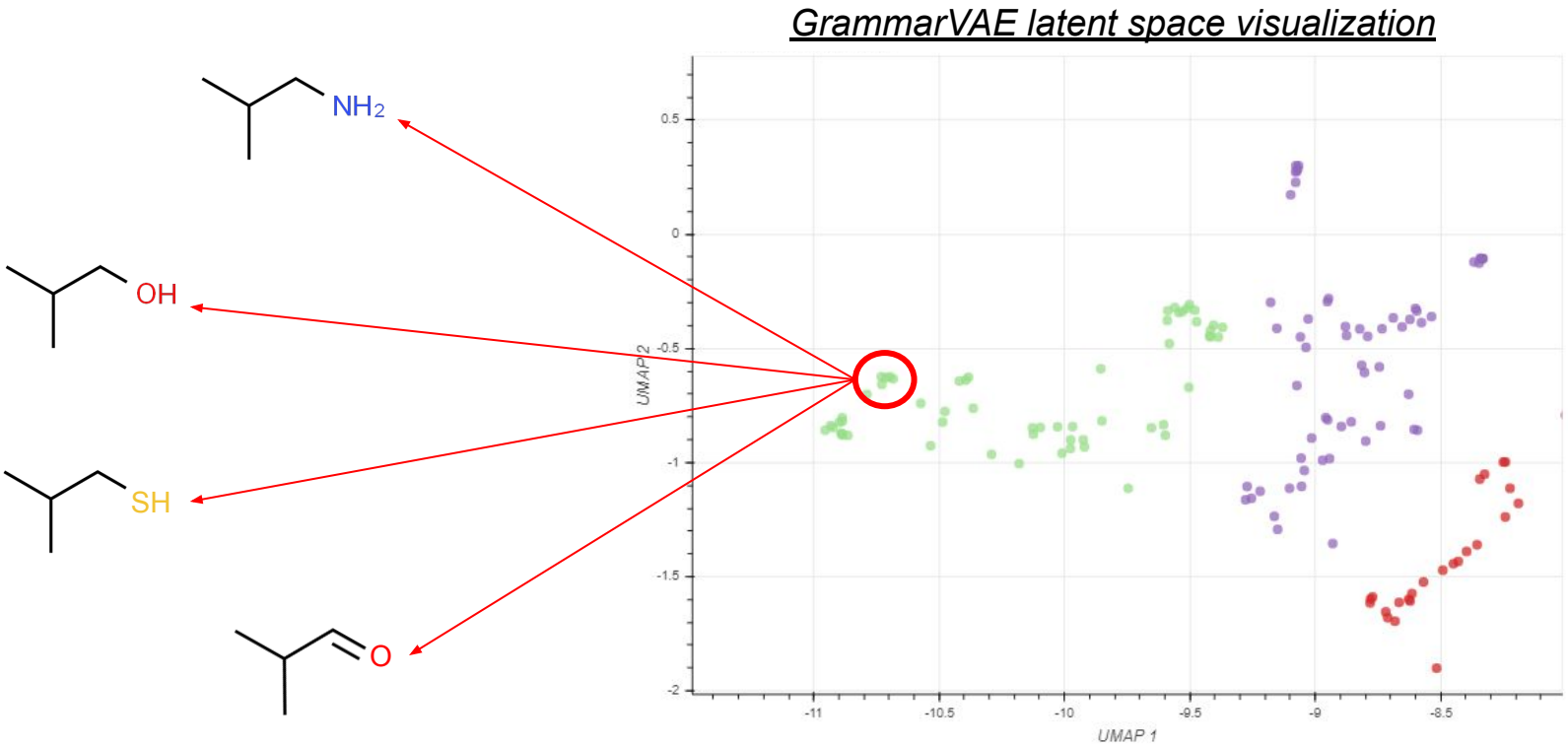
Issues with existing SMILES-based models



Cc1cn2c(CN(C)C(=O)c3ccc(F)cc3C)c(C)nc2s1

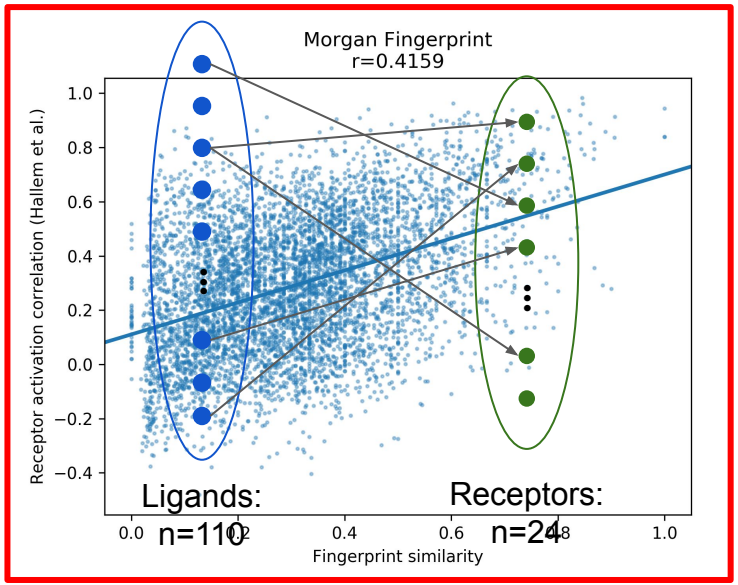
Cc1cc(F)ccc1C(=O)N(C)Cc1c(C)nc2scc(C)n12

Current models overemphasize molecular geometry

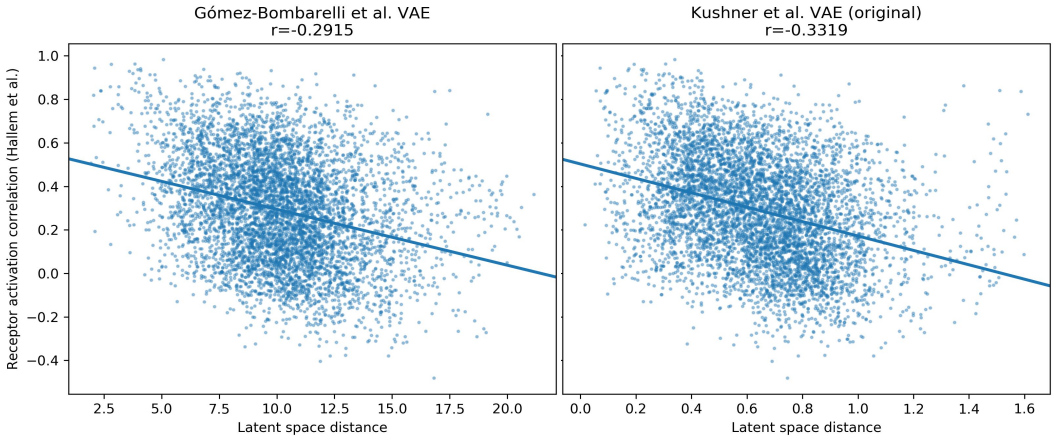


Kushner et al. 2017

Evaluating the latent space of current models



Molecular fingerprint still performs the best

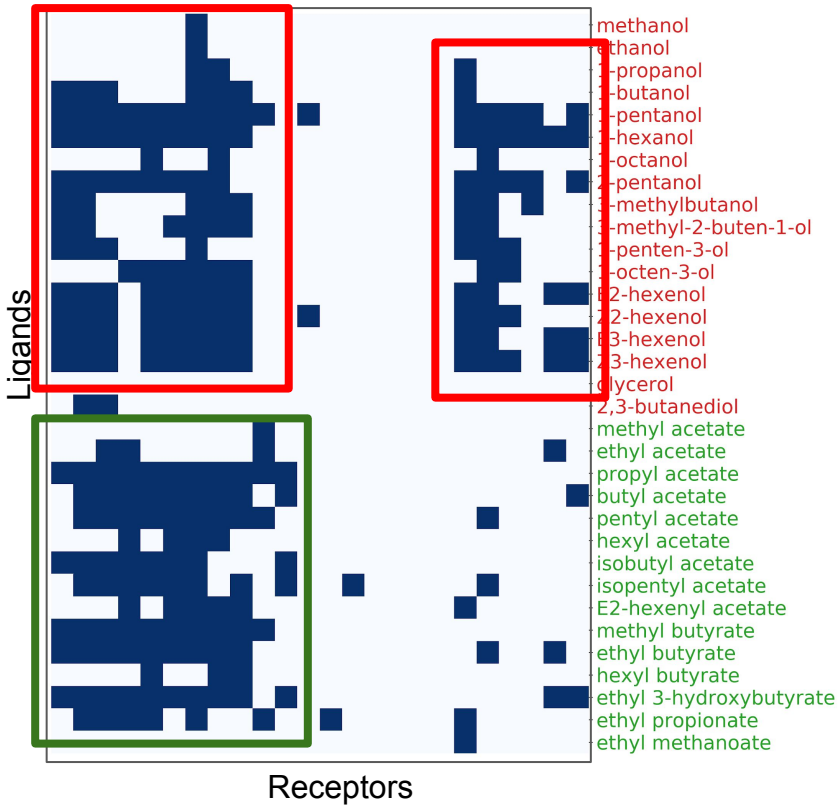


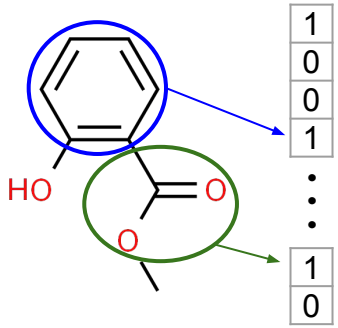
Correlation between distance in latent space and receptor activity

Data from Hallem, Ho, Carlson. *Cell* 2004.

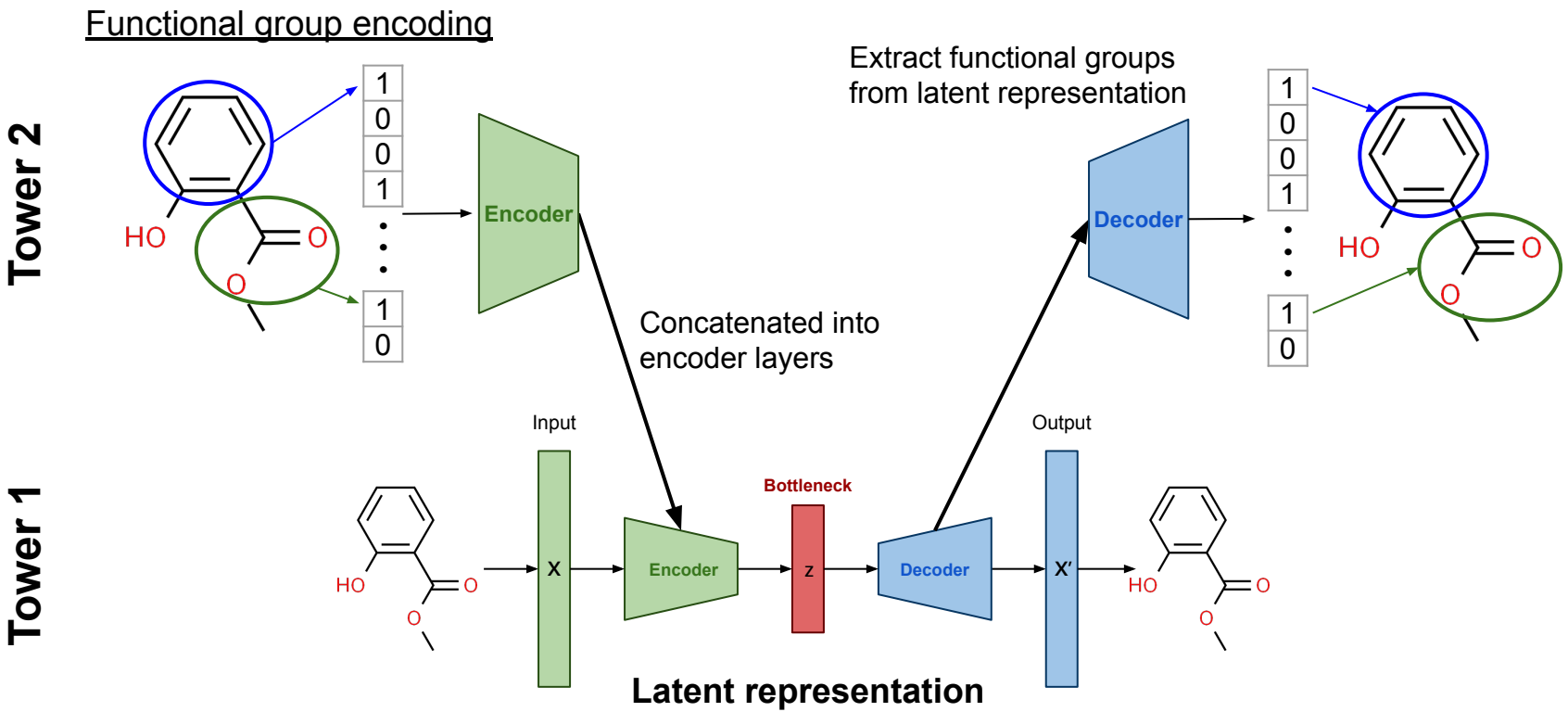
Our Approach

Incorporating prior knowledge helps the model

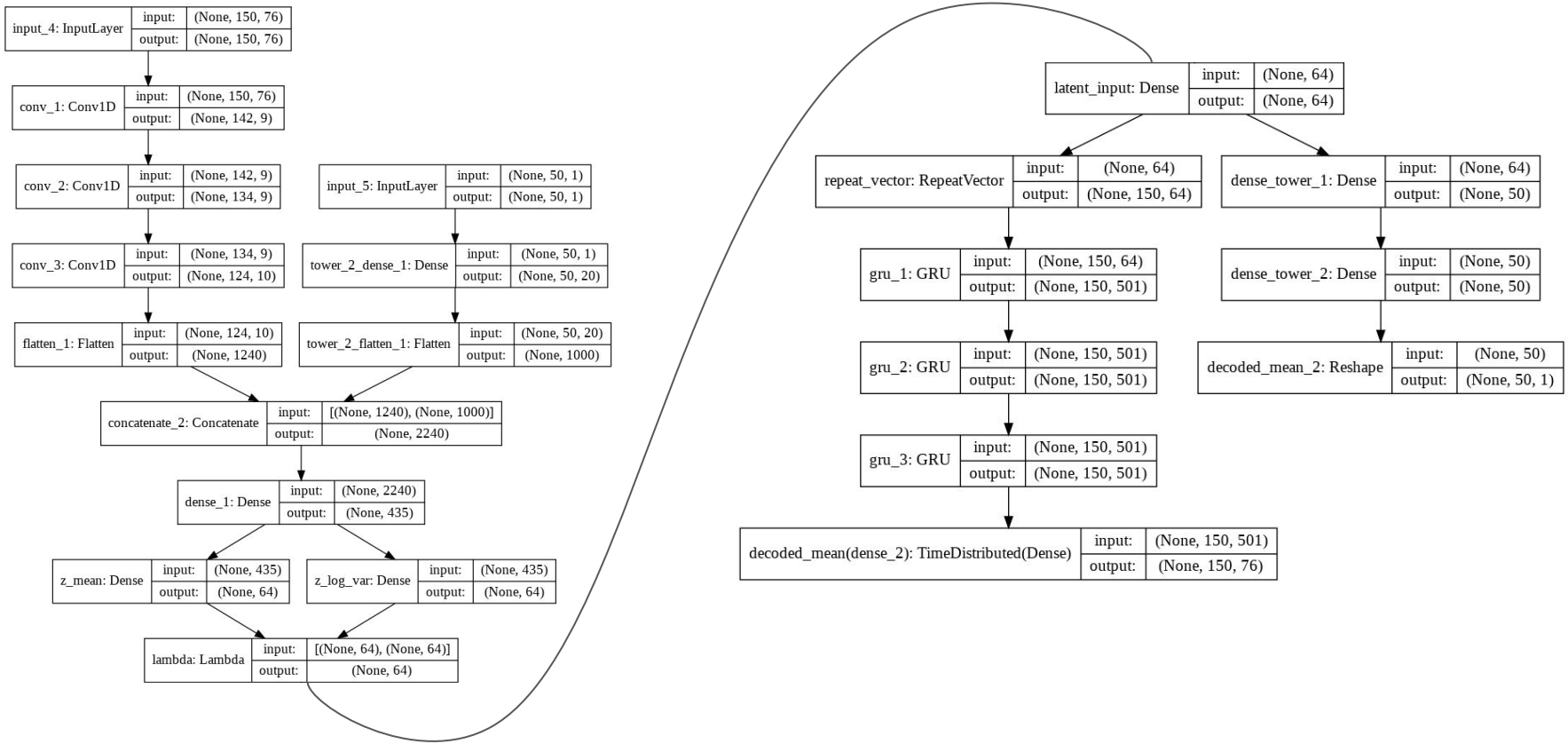




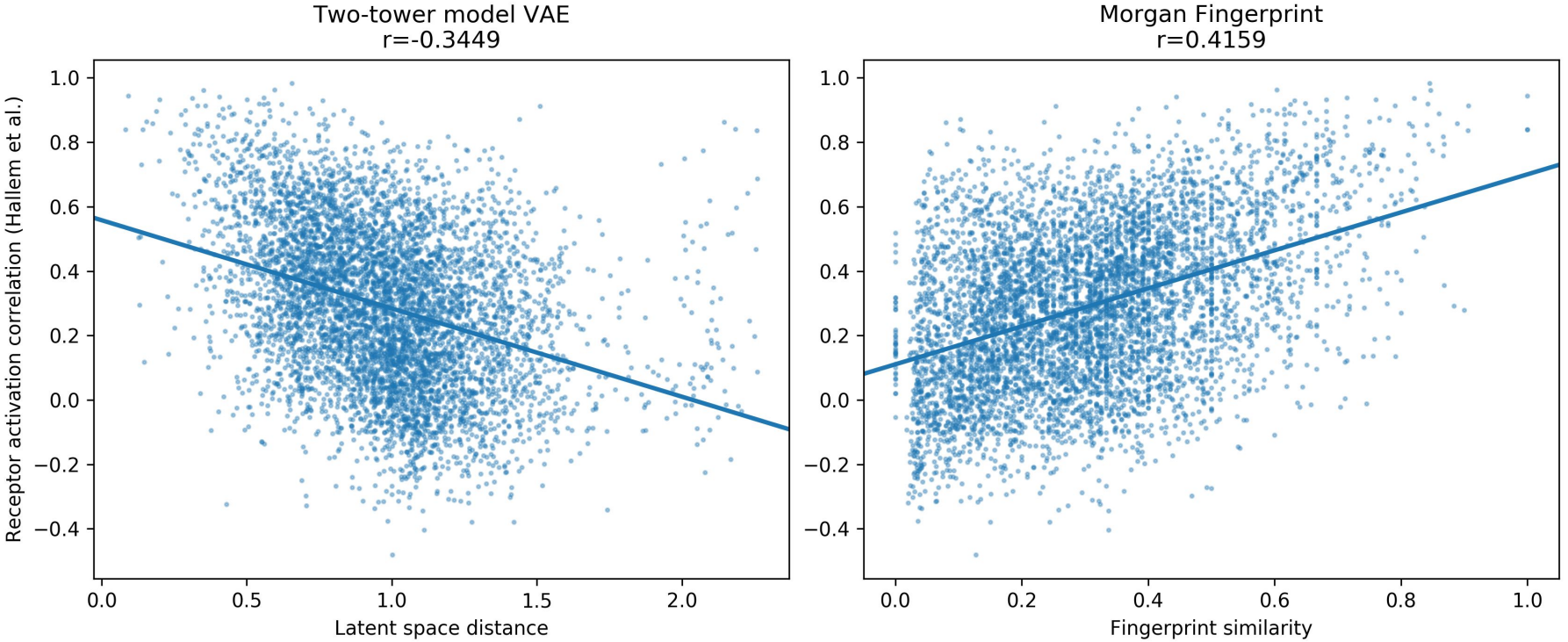
Our model: two-tower approach



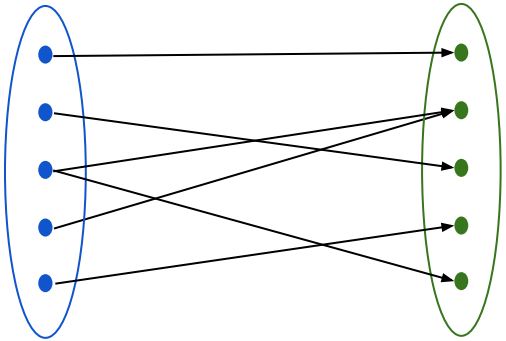
Two-tower architecture

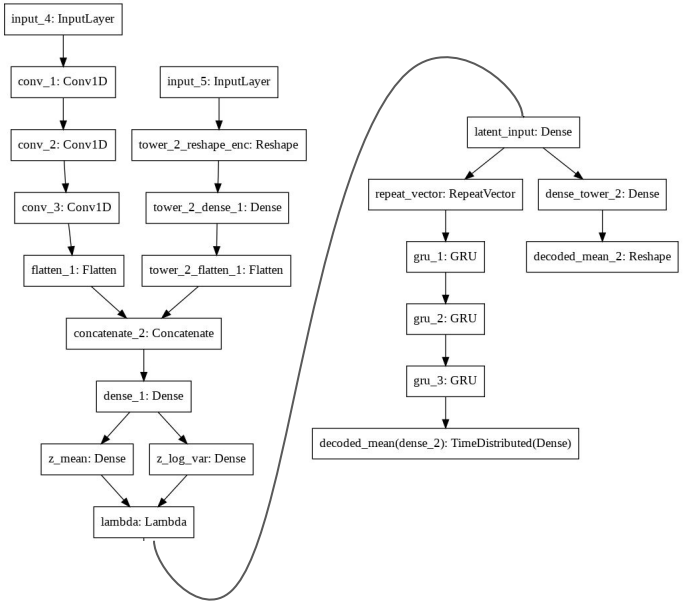


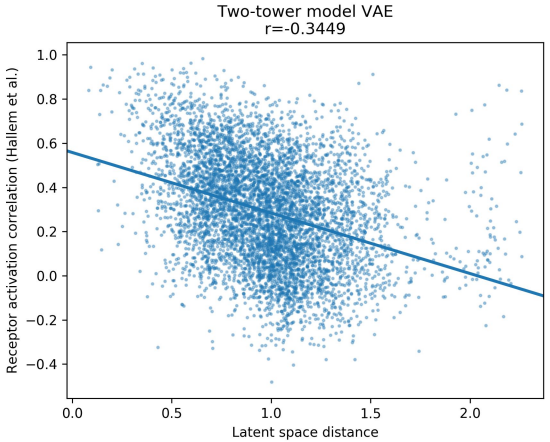
Our results


Two-tower model VAE
r=-0.3449
Morgan Fingerprint
r=0.4159
Receptor activation correlation (Hallem et al.)
Latent space distance
Fingerprint similarity
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.0
0.5
1.0
1.5
2.0
0.2
0.4
0.6
0.8

Recap







Combinatorial Ligand-receptor binding

Feature abstraction with VAEs: Two-tower approach

Predicting receptor activities