

Increased Efficiency in the Initial Stages of Drug Discovery

Contents

Background

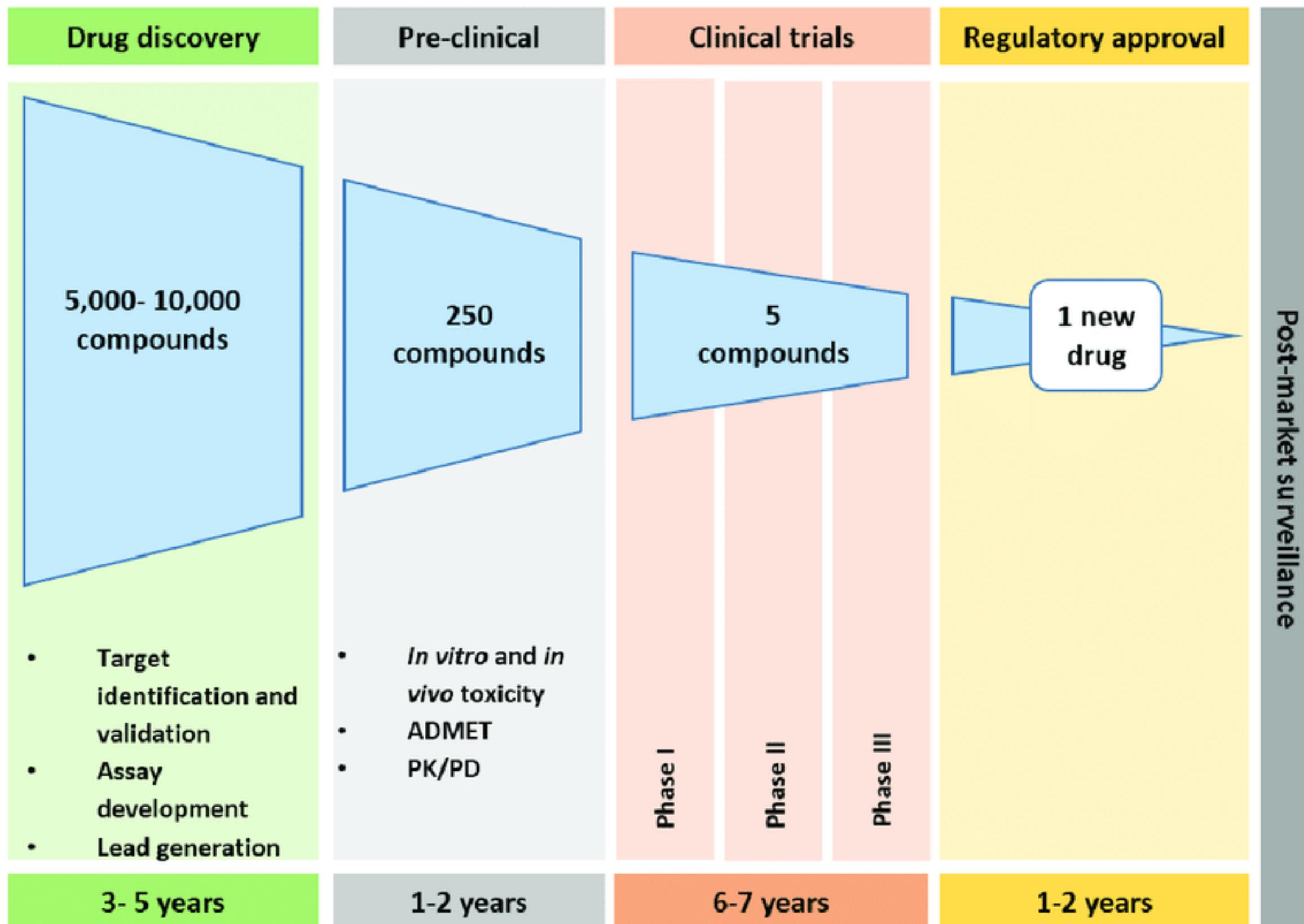
Business Problem

Data & Methods

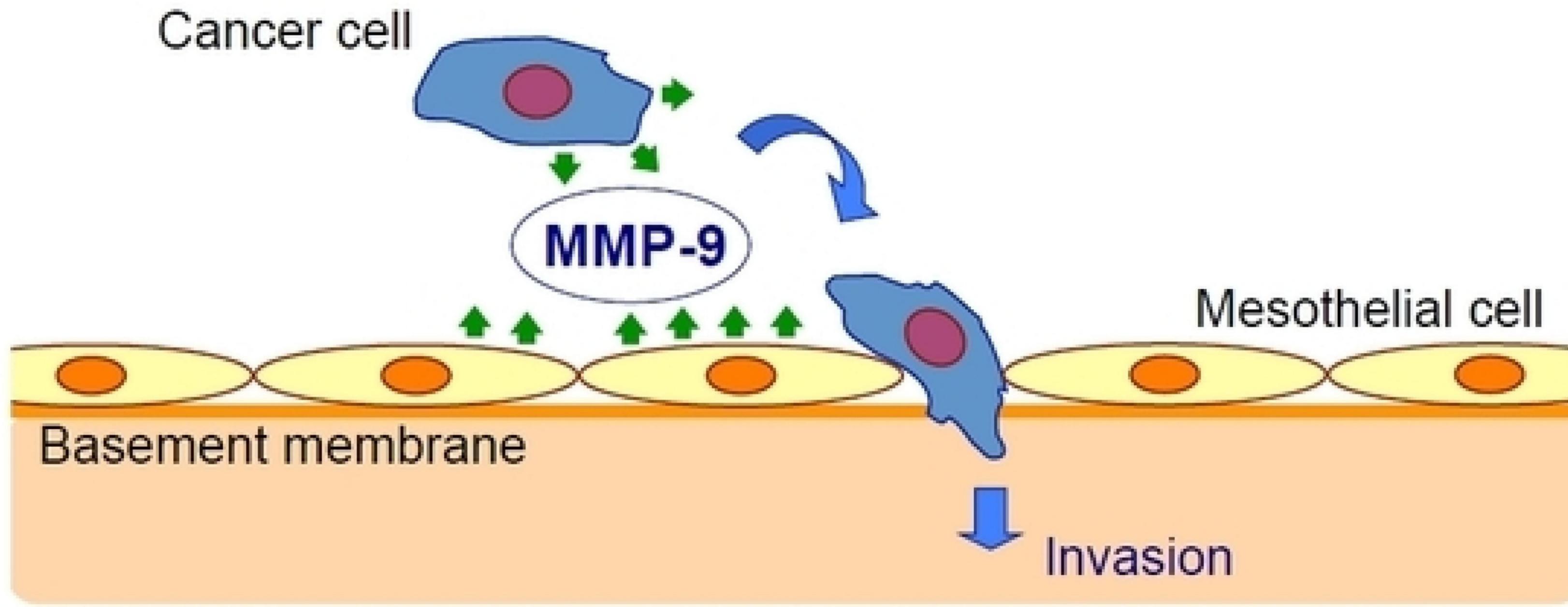
Model

Conclusions & Future Directions

Drug Discovery Process



Target: MMP9



Business Problem

Screening small molecule libraries is expensive and time consuming.

Matrix Metalloproteinase 9
- promotes cancer metastasis

Can machine learning models be taught to distinguish between active and inactive compounds?

Data & Methods

ChEMBL Database

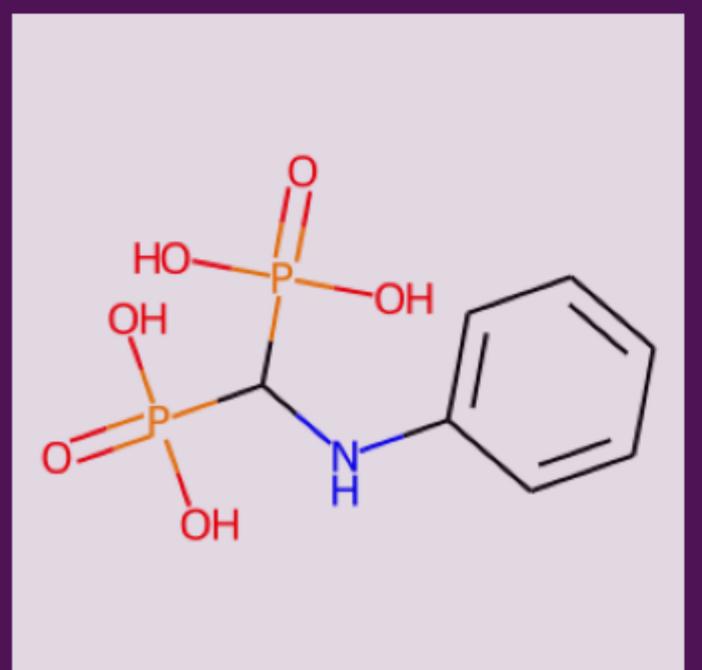
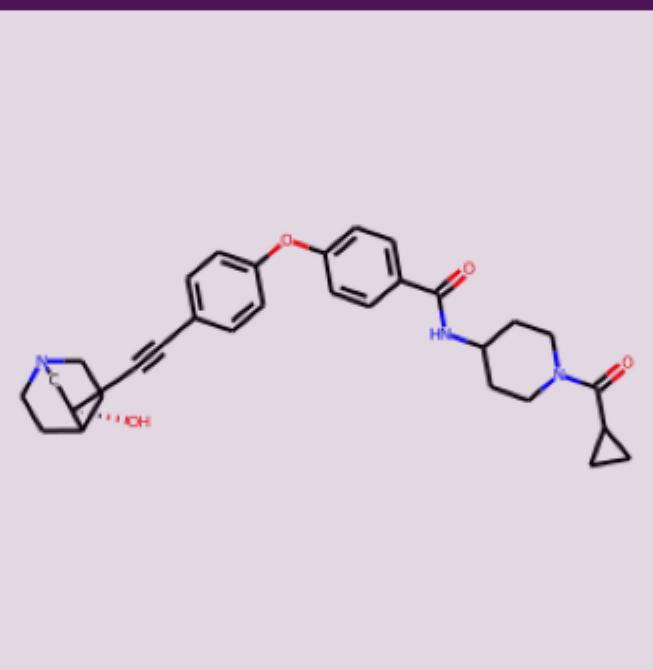
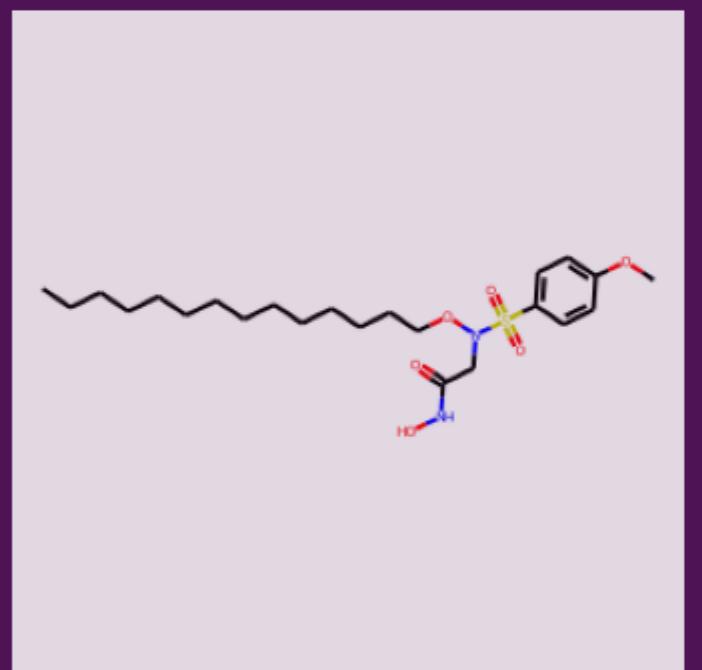
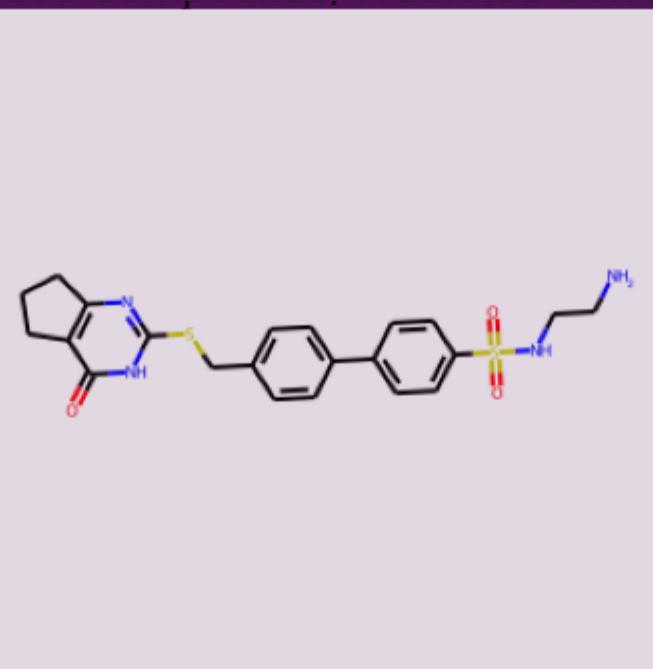
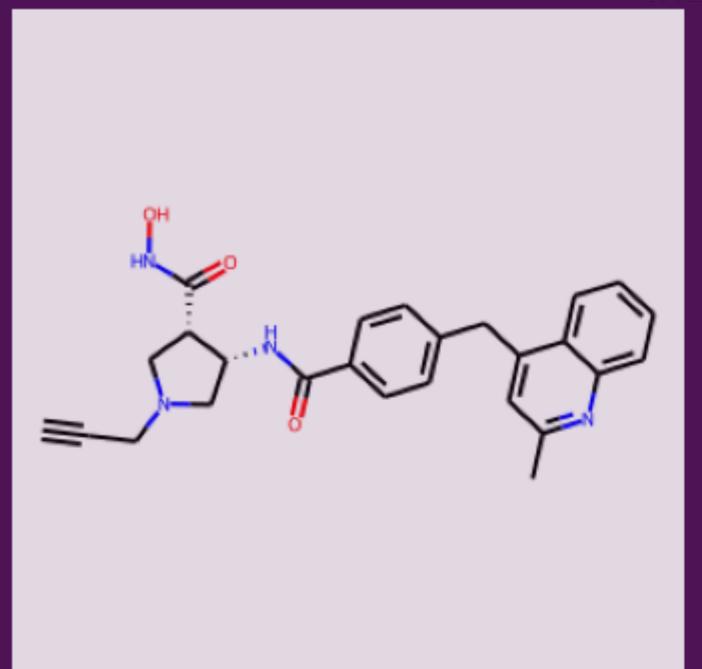
- 2263 compounds

PubChem Fingerprints generated by PaDEL-Descriptor

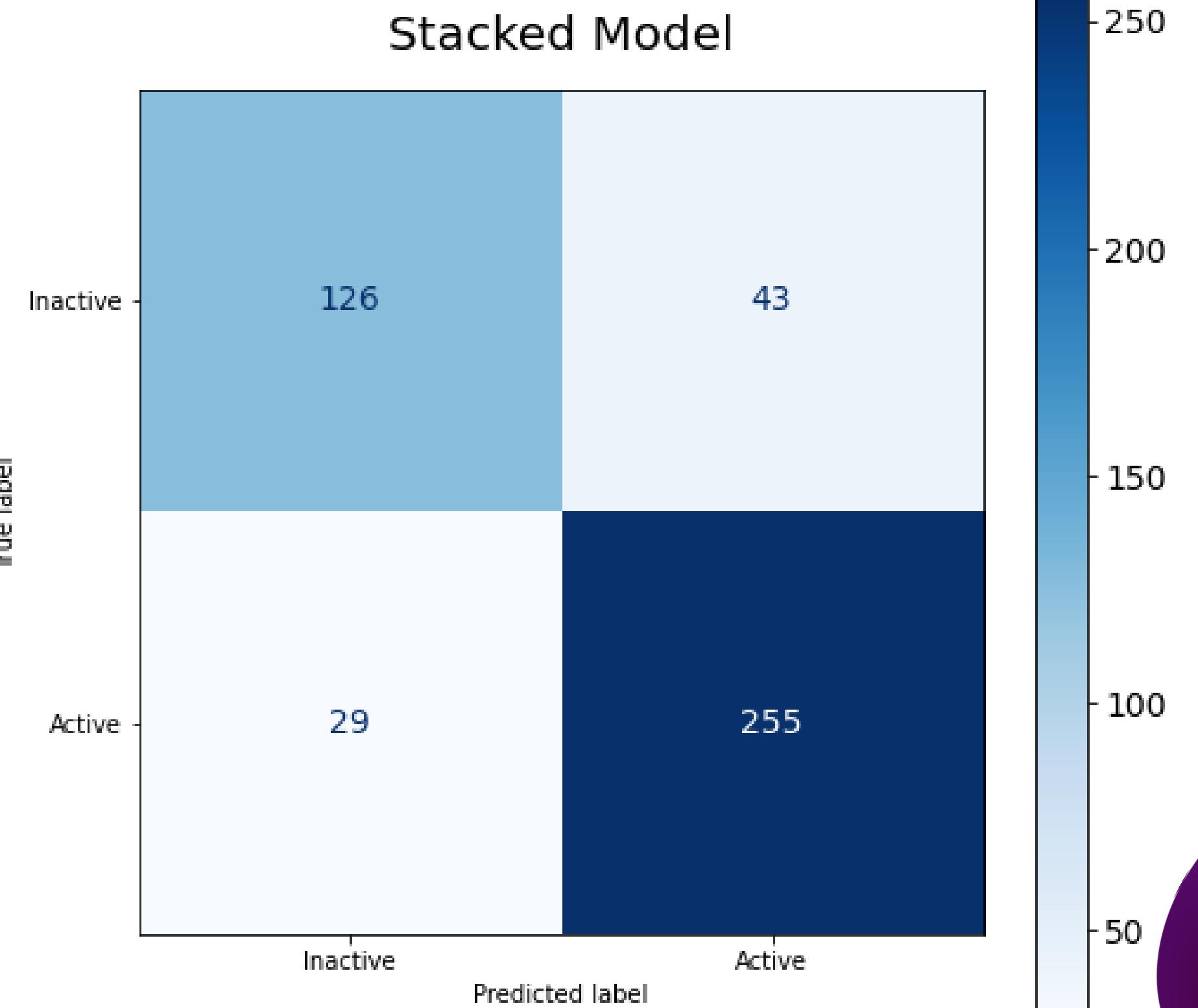
- 147 Fingerprints used in analysis

Classes: Active & Inactive

Model: Stack of 3 classification models



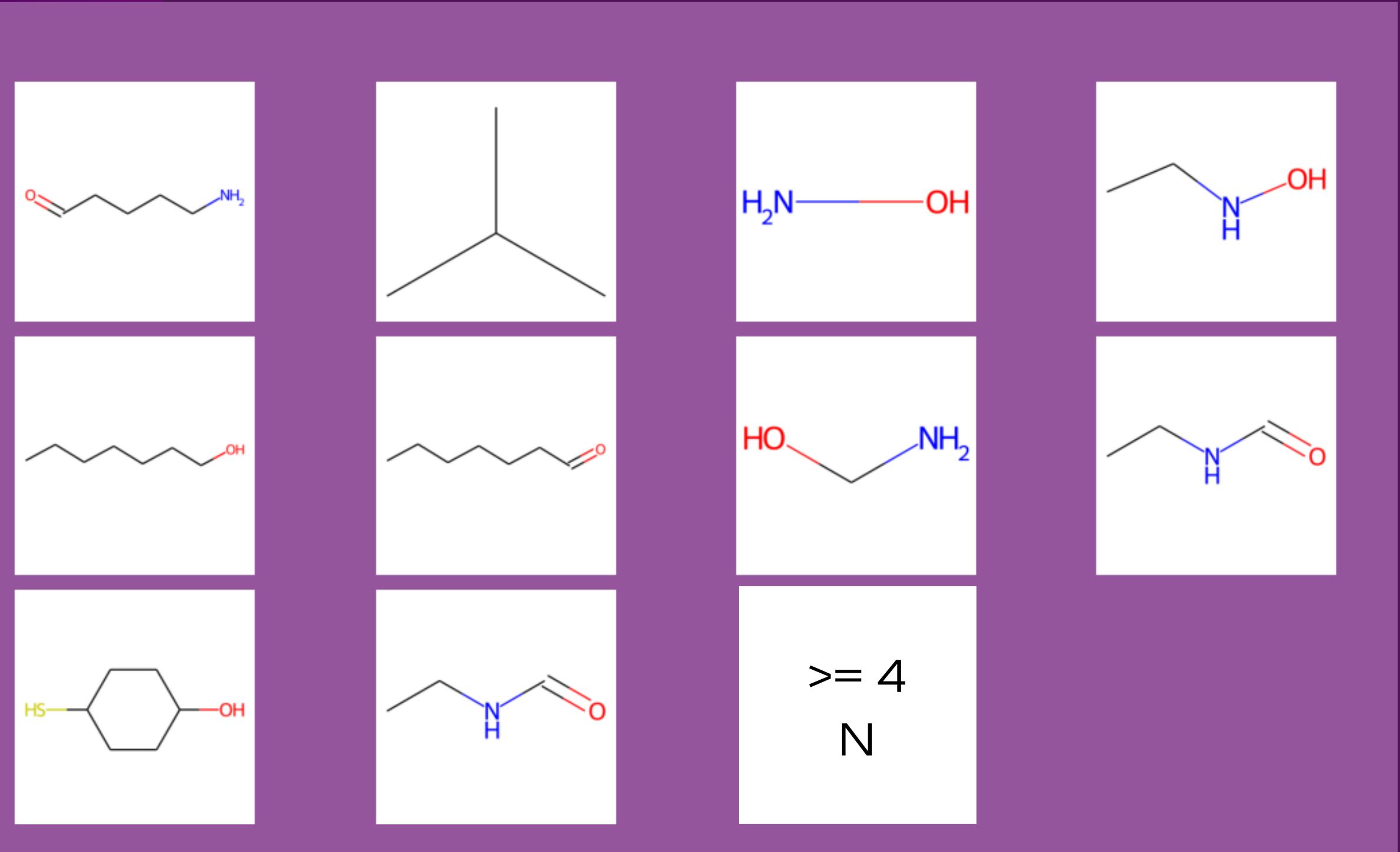
Model Performance



Important Components of Active Compounds

Important in
at least two models

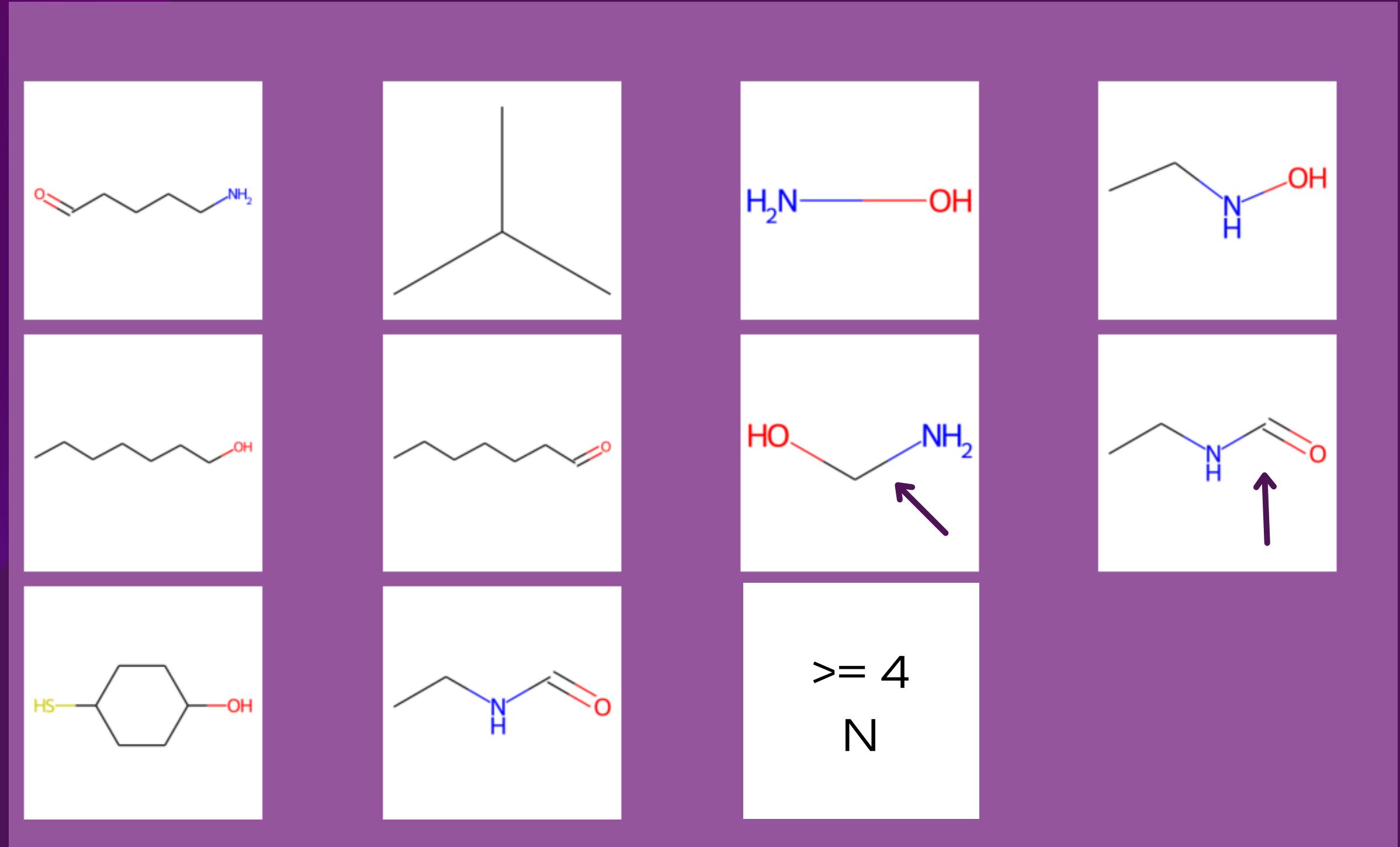
Most important
for each model



Important Components of Active Compounds

Important in
at least two models

Most important
for each model



Conclusions & Future Directions

Current Model

Works well enough to improve the efficiency of screening small molecule libraries

Improved Model

Make this a multi-label classification task to find drugs specific to MMP9 over the other MMPs



Thank you for listening!

Questions?

Contact me at cindylund@gmail.com