



Intellegens

Imputation of protein activity data using deep learning

Gareth Conduit and Tom Whitehead

Neural network algorithm to



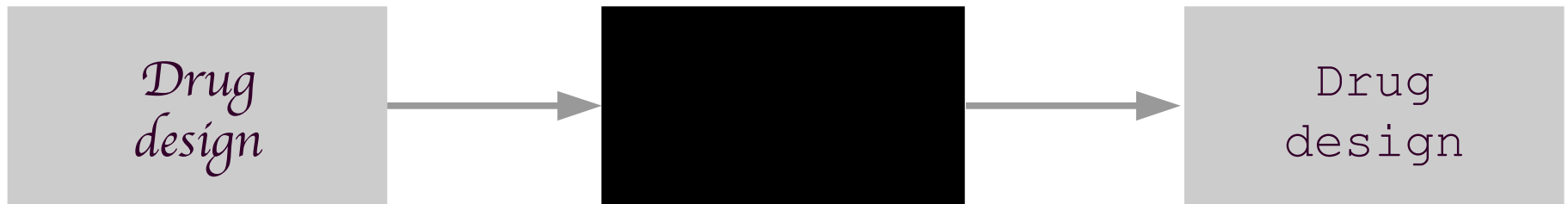
Merge simulations, physical laws, and experimental data

Reduce the need for expensive experimental development

Accelerate drugs and materials discovery

Generic with experimentally **proven** applications in materials discovery

A black box



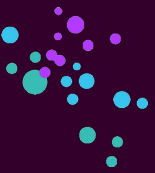
Train on complete data



Digitize handwriting



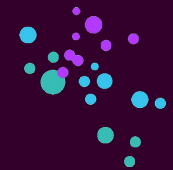
Train on fragmented data



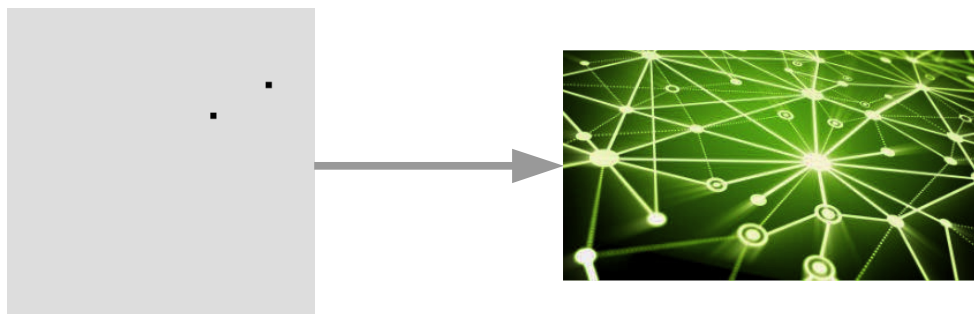
Predict on fragmented data



Enhancing drug discovery



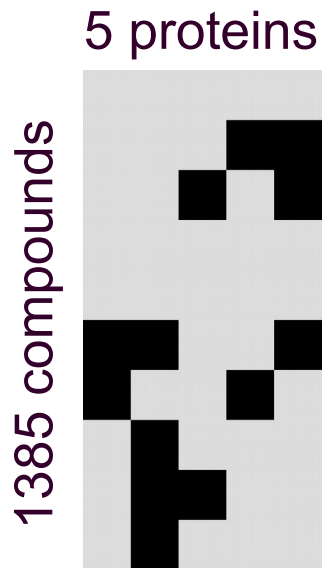
ChEMBL dataset just 0.1% complete



Data available for Adrenergic receptors



5 proteins with 1731 compounds, dataset 37% complete



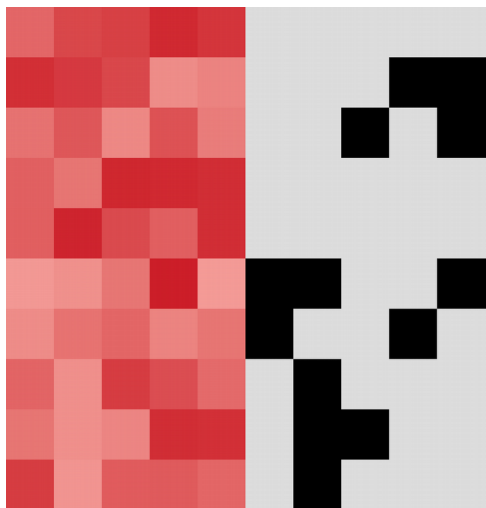
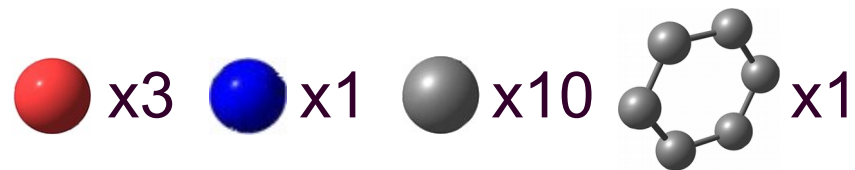
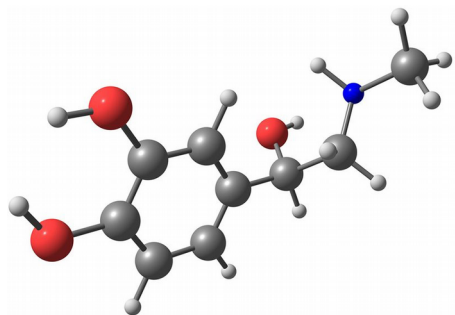
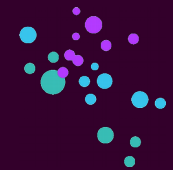
Impute the Adrenergic receptors dataset



Original dataset 37% complete, filled 65% of entries



Additional descriptors for Adrenergic receptors



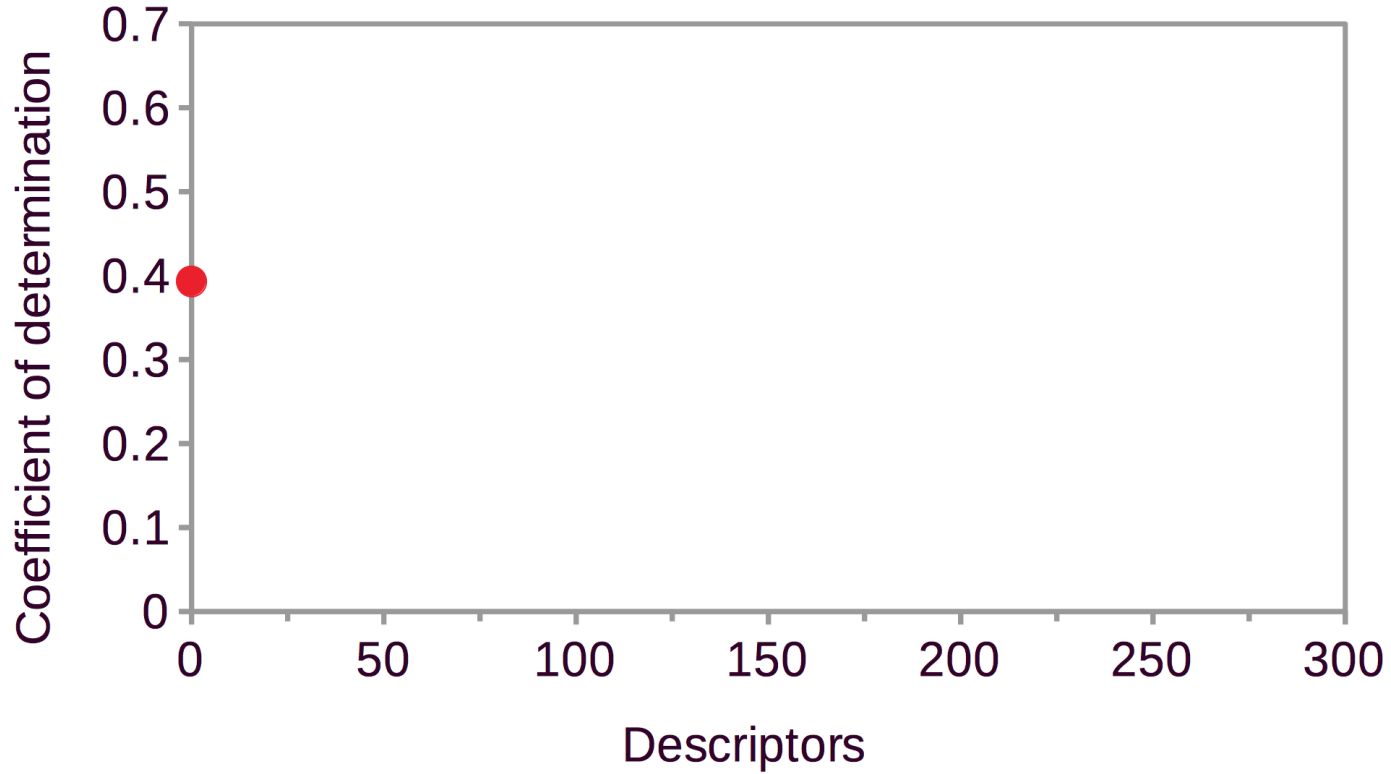
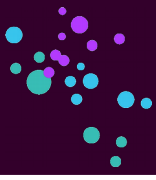
Improved predictions



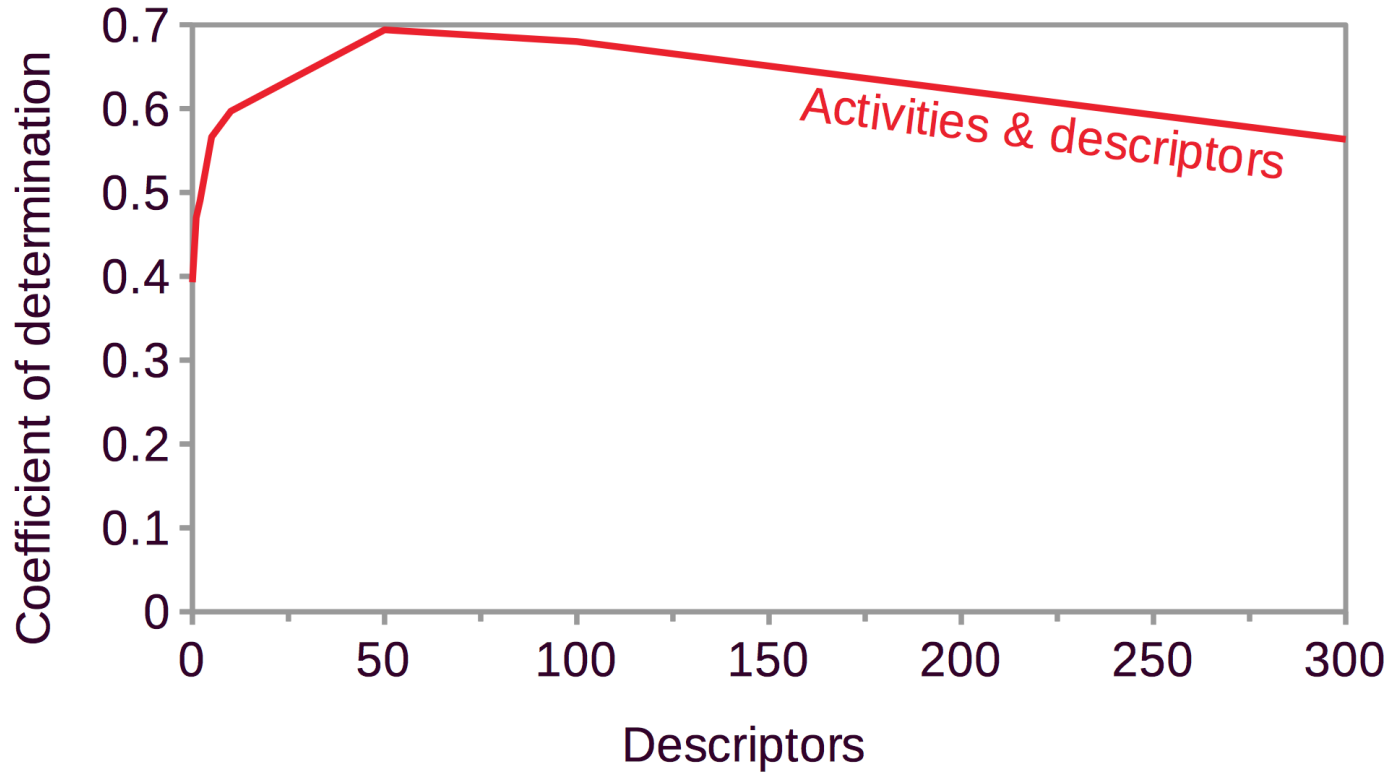
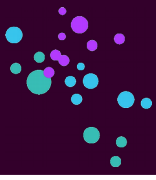
Include structural information to fill to 82%



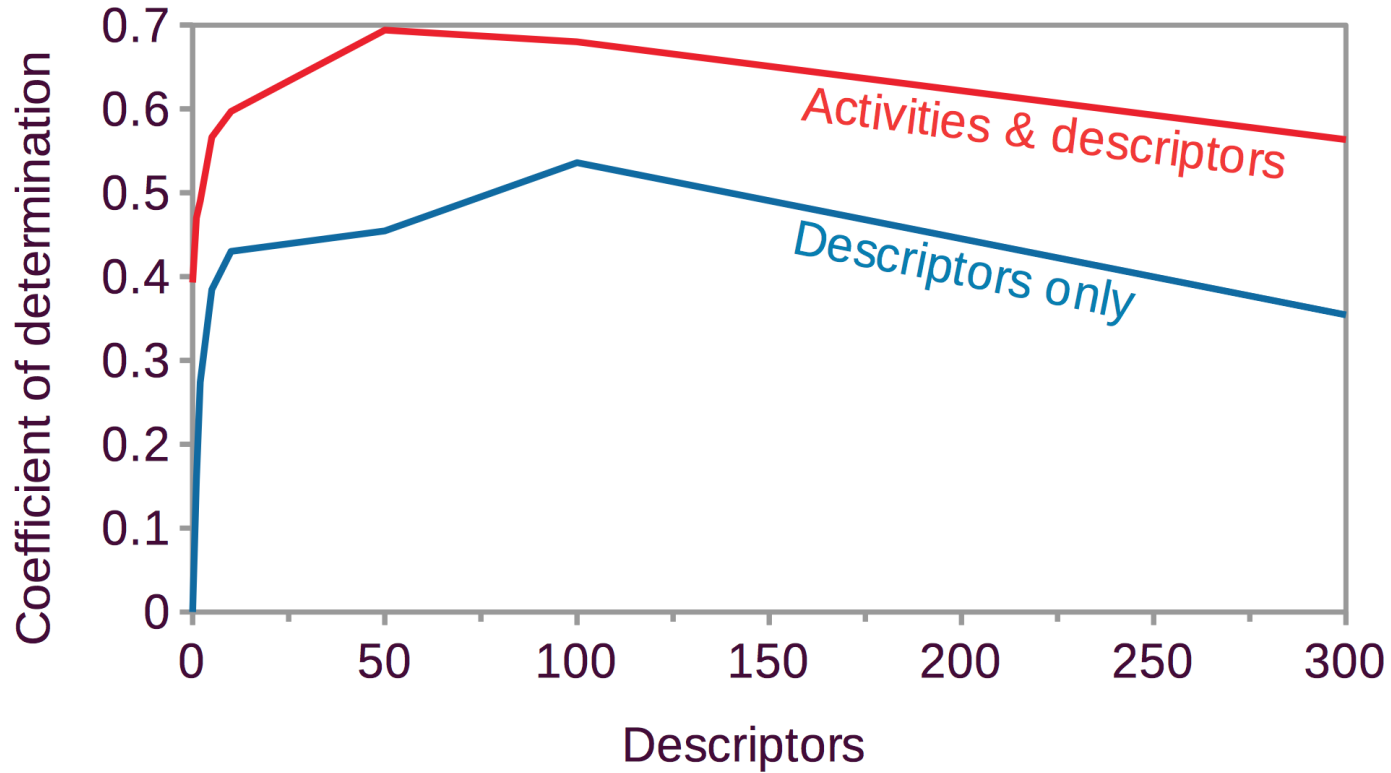
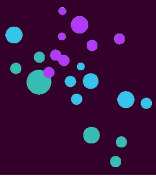
Predictions from just activities



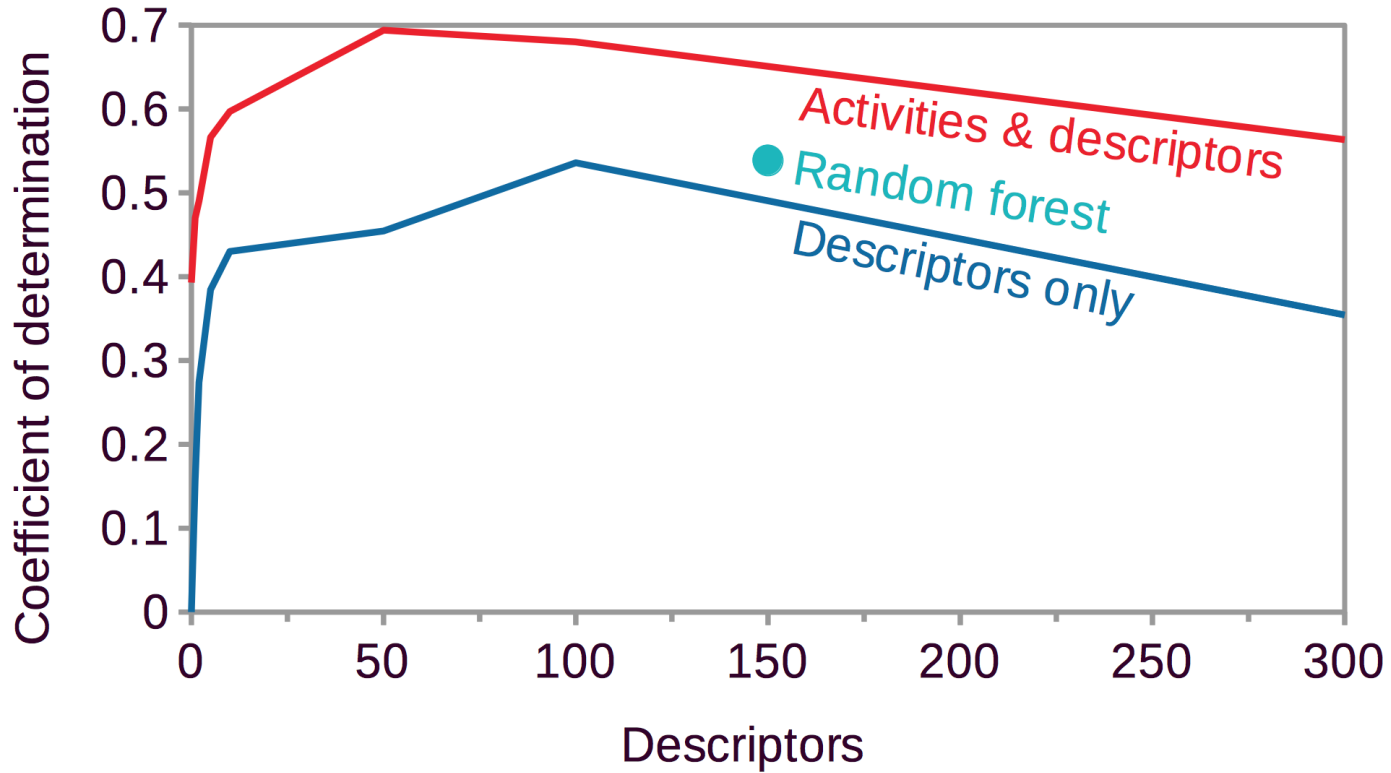
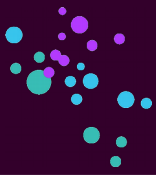
Predictions from activities and descriptors



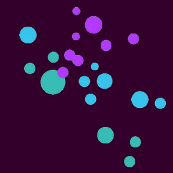
Predictions with just descriptors



Comparison to random forest



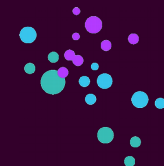
Future prospects



Enhance ChEMBL dataset from 0.1% to 20% complete



Even more materials designed



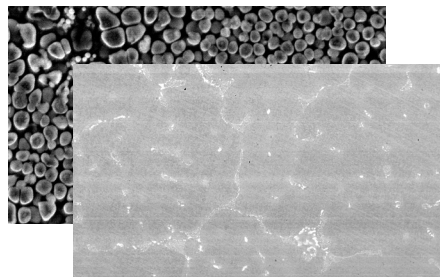
Battery design
with DFT and
experimental data



Designing lubricants
with DFT and
experimental data



Nickel and
molybdenum
alloys





Apply deep learning to high-value **fragmented** data

Merge experiments and simulations into **holistic** design tool

Experimentally **proven** applications in materials design,
founded start-up **intellegens**

Scientists establish all possible **SOURCES** of information