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

Protein activity

Protein activity

Digitize handwriting

Drug design

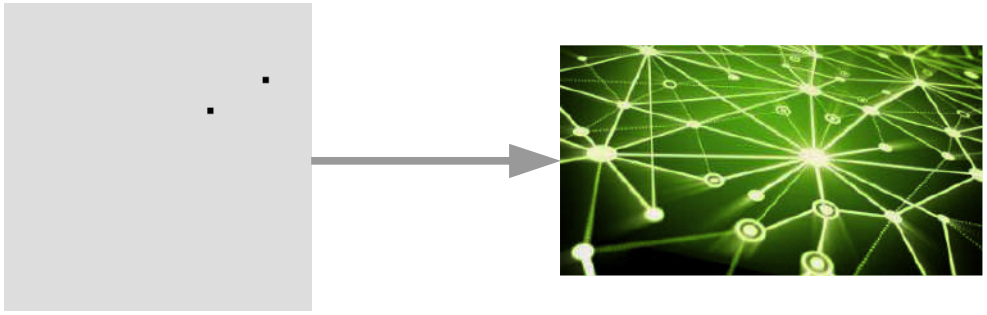
Drug design

Train on fragmented data

Predict on fragmented data

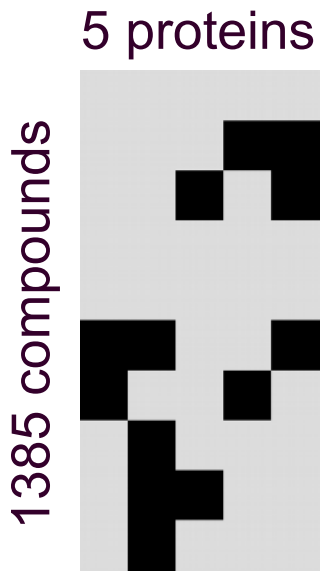
Enhancing drug discovery

ChEMBL dataset just 0.1% complete



Data available for Adrenergic receptors

5 proteins with 1731 compounds, dataset 37% complete

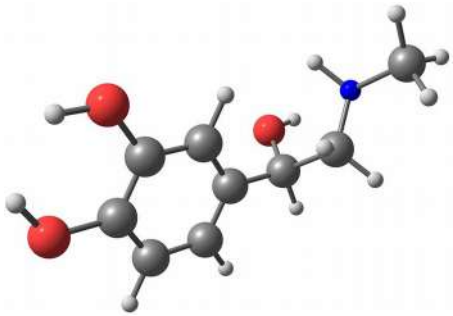


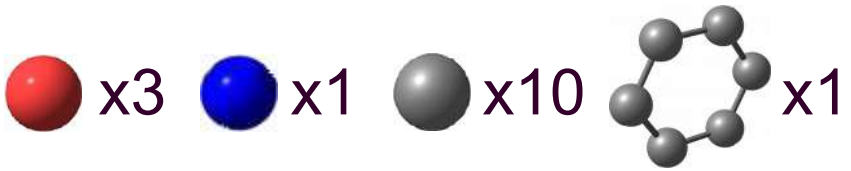
Impute the Adrenergic receptors dataset

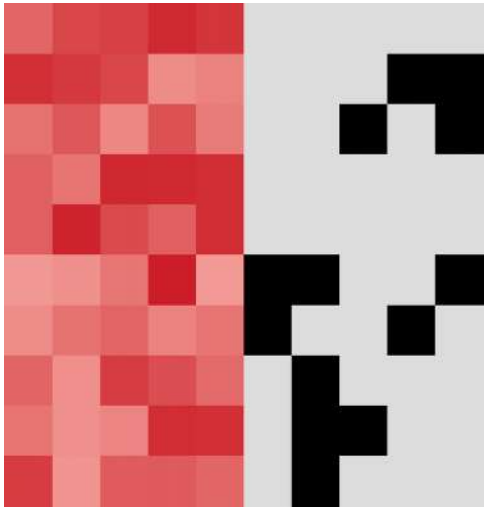
Original dataset 37% complete, filled 65% of entries

optibrium

Additional descriptors for Adrenergic receptors







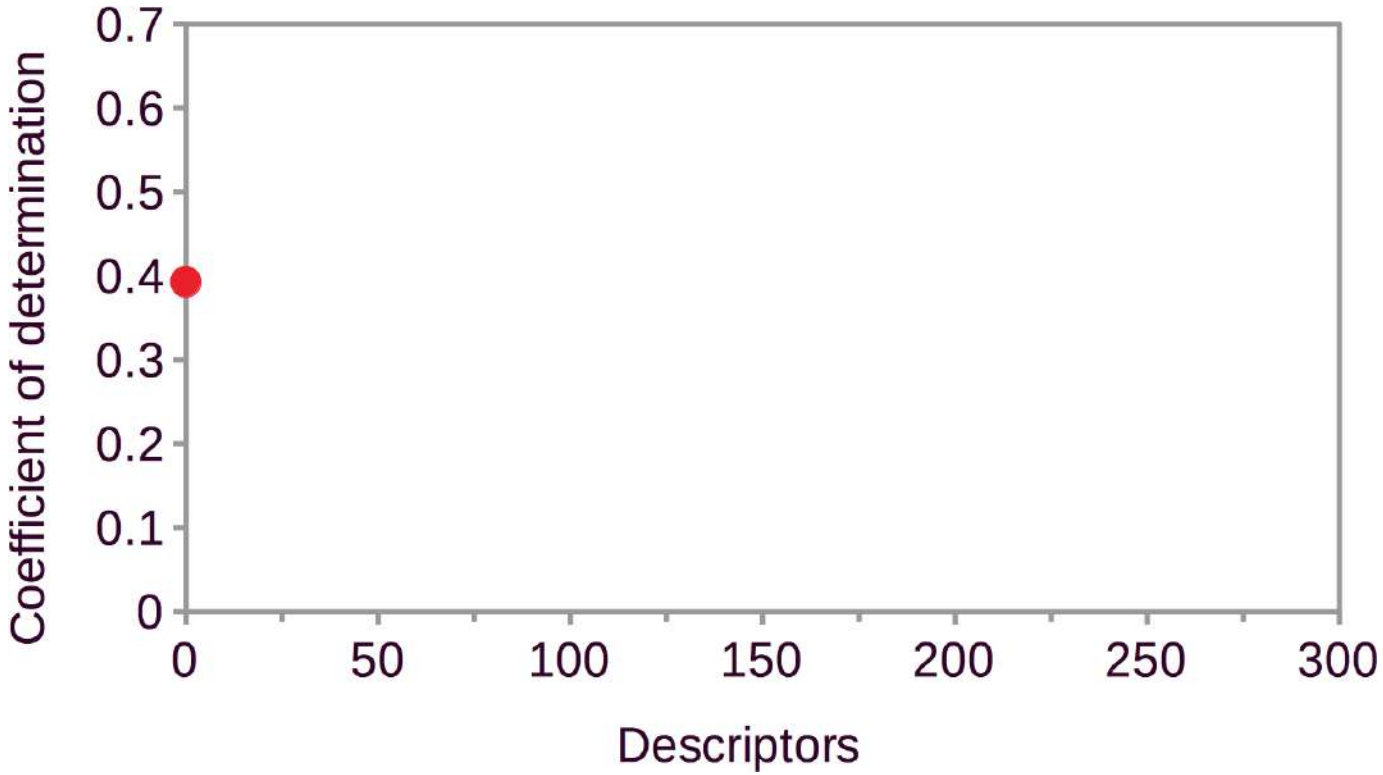
optibrium

Improved predictions

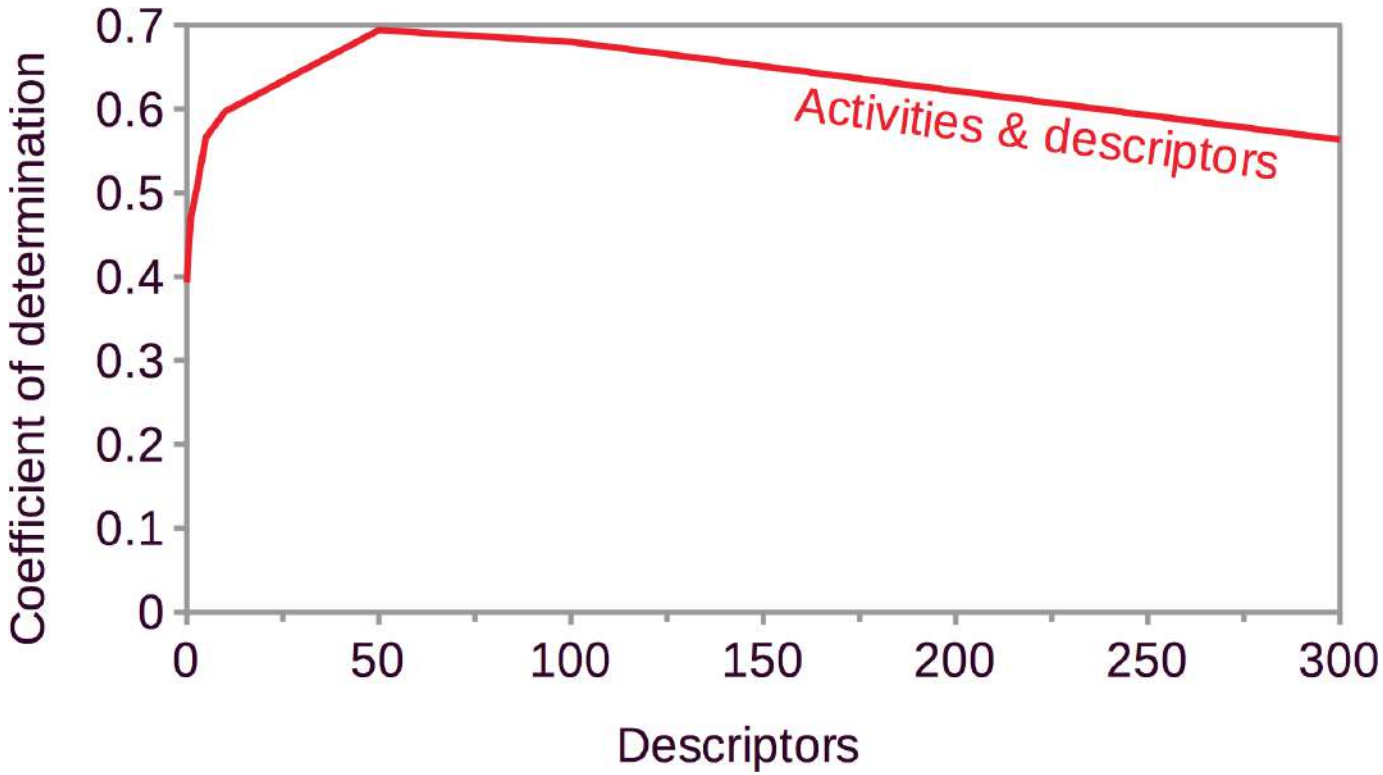
Include structural information to fill to 82%

optibrium

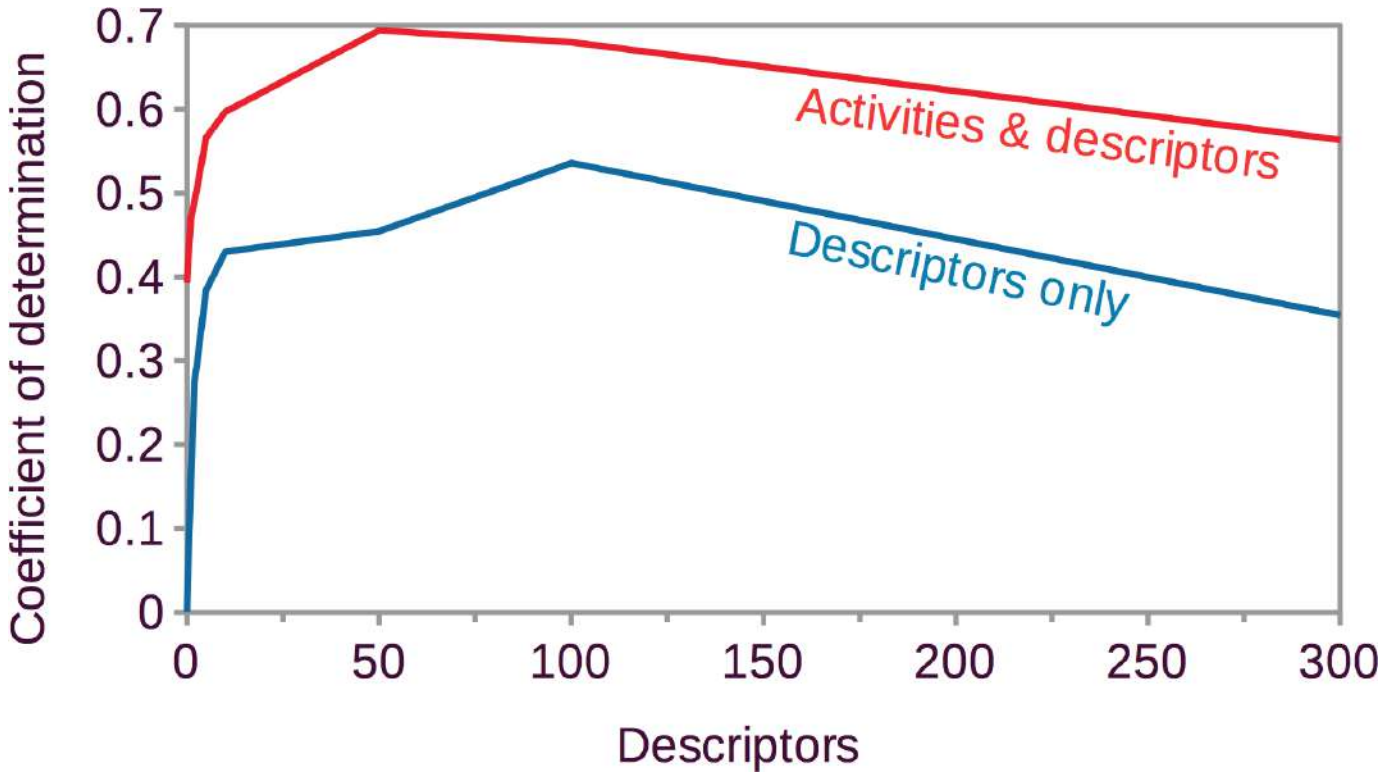
Predictions from just activities



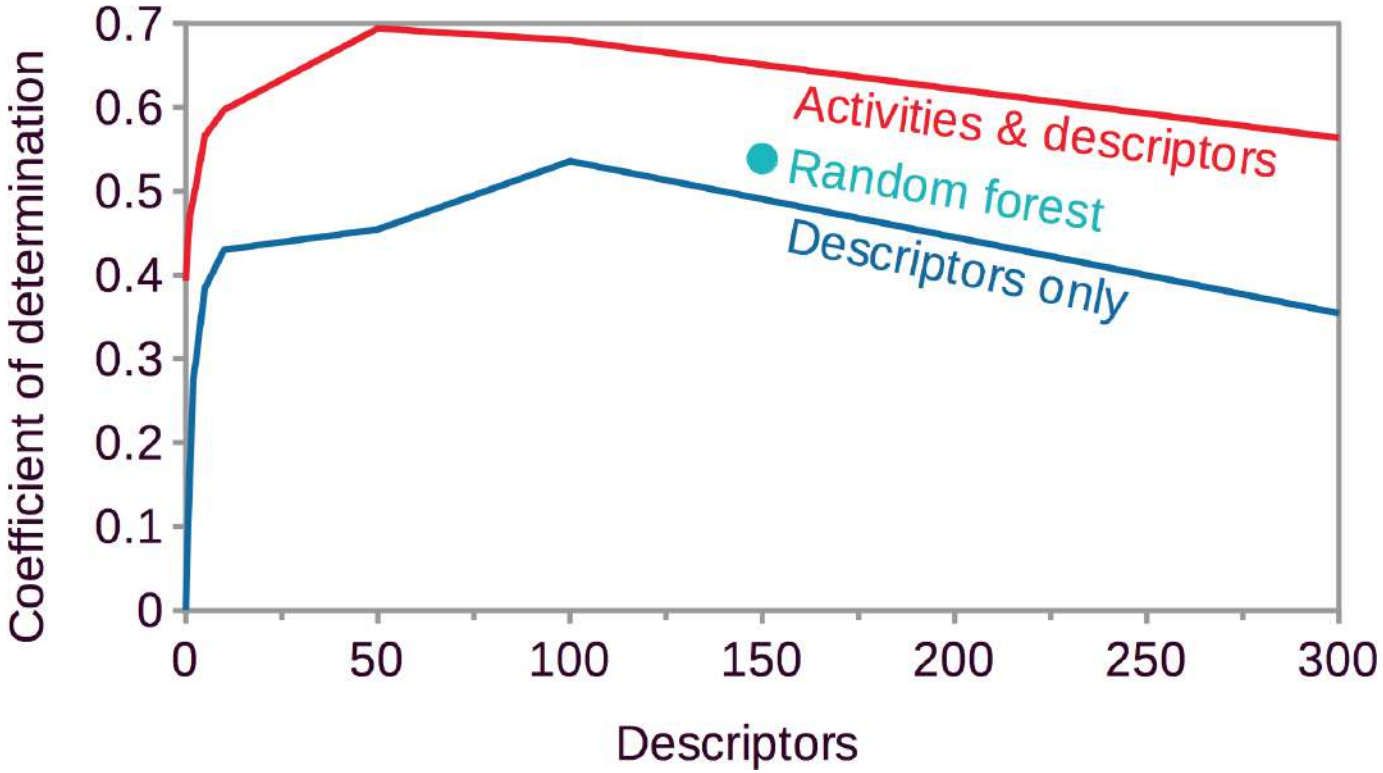
Predictions from activities and descriptors



Predictions with just descriptors



Comparison to random forest

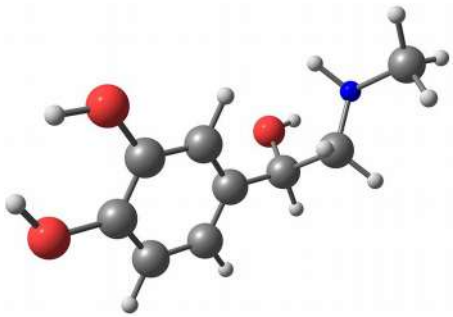


Future prospects

Enhance ChEMBL dataset from 0.1% to 20% complete

Materials designed

Drug discovery



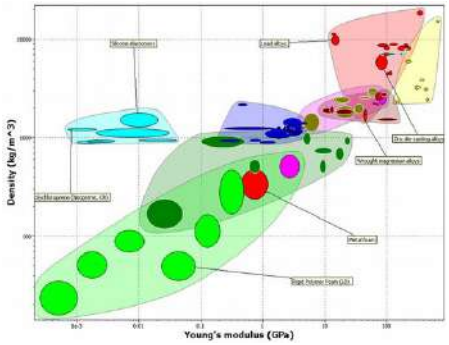
e-therapeutics

3D printed alloy designed from 10 data entries

ROLLS ROYCE

Materials Solutions

Found errors in materials databases



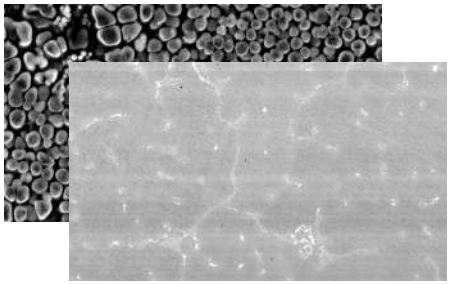
GRANTA
MATERIAL INSPIRATION

Even more materials designed

Battery design with DFT and experimental data

Designing lubricants with DFT and experimental data

Nickel and molybdenum alloys



Summary

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