Intellegens

Imputation of protein activity data using deep learning

Gareth Conduit

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 **proven** applications in drug design and materials discovery

Neural network: train on complete data

Proposed neural network: train on fragmented data

Protein activity

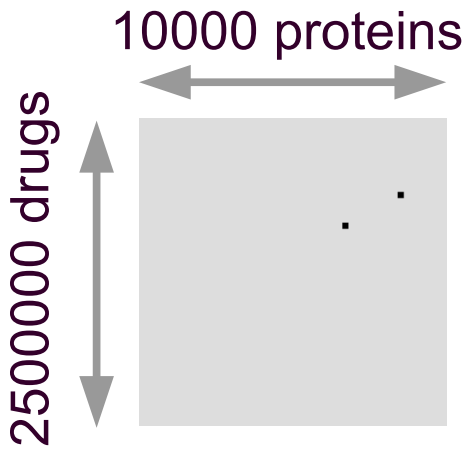
Protein activity

Proposed neural network: predict on fragmented data

Data available for drug discovery

10,000 proteins with 2,500,000 compounds

Original dataset 0.05% complete



e-therapeutics

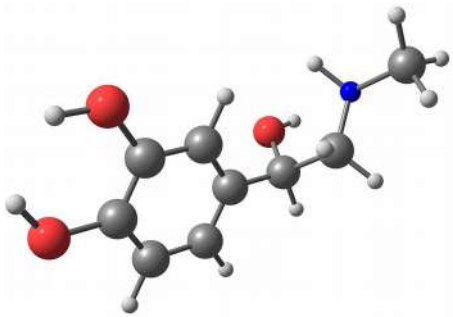
Impute the database used for drug discovery

10,000 proteins with 2,500,000 compounds

Original dataset 0.05% complete, filled 32% of entries

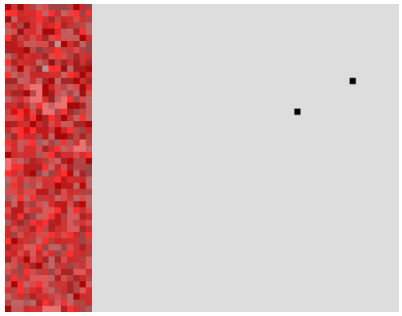
e-therapeutics

Drug discovery with additional descriptors



1101 Morse descriptor

200



e-therapeutics

Improved drug discovery

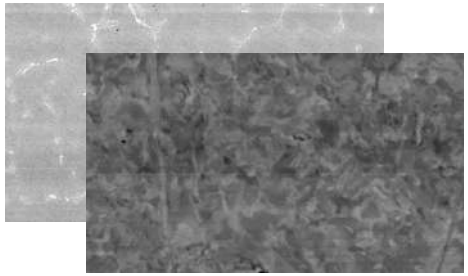
Include drug structural information to fill to 46%

Saved >$1billion in experimental costs

e-therapeutics

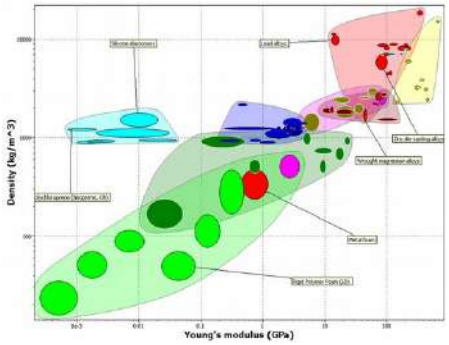
Materials designed

Molybdenum forging alloys



3D printed alloy designed from 10 data entries

Found errors in materials databases



ROLLS ROYCE

ROLLS ROYCE Materials Solutions

GRANTA MATERIAL INSPIRATION

Even more materials designed

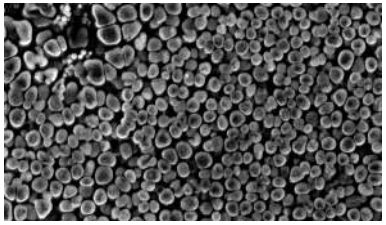
Battery design with DFT and experimental data

SAMSUNG

Designing lubricants with DFT and experimental data

bp

Nickel disc alloys



ROLLS ROYCE

Summary

Apply deep learning to high-value fragmented data

Merge experiments and simulations into holistic design tool

Experimentally proven drugs and materials design with 7 companies, founded startup intellegens

Scientists establish all possible sources of information