UNIVERSITY OF CAMBRIDGE

Deep learning for materials design and drug discovery

Gareth Conduit

Theory of Condensed Matter group

Neural network algorithm to

Merge simulations, physical laws, and experimental data

Reduce the need for expensive experimental development

Accelerate materials and drugs discovery

Generic with proven applications in materials discovery and drug design

A black box

Train with complete data

Materials design

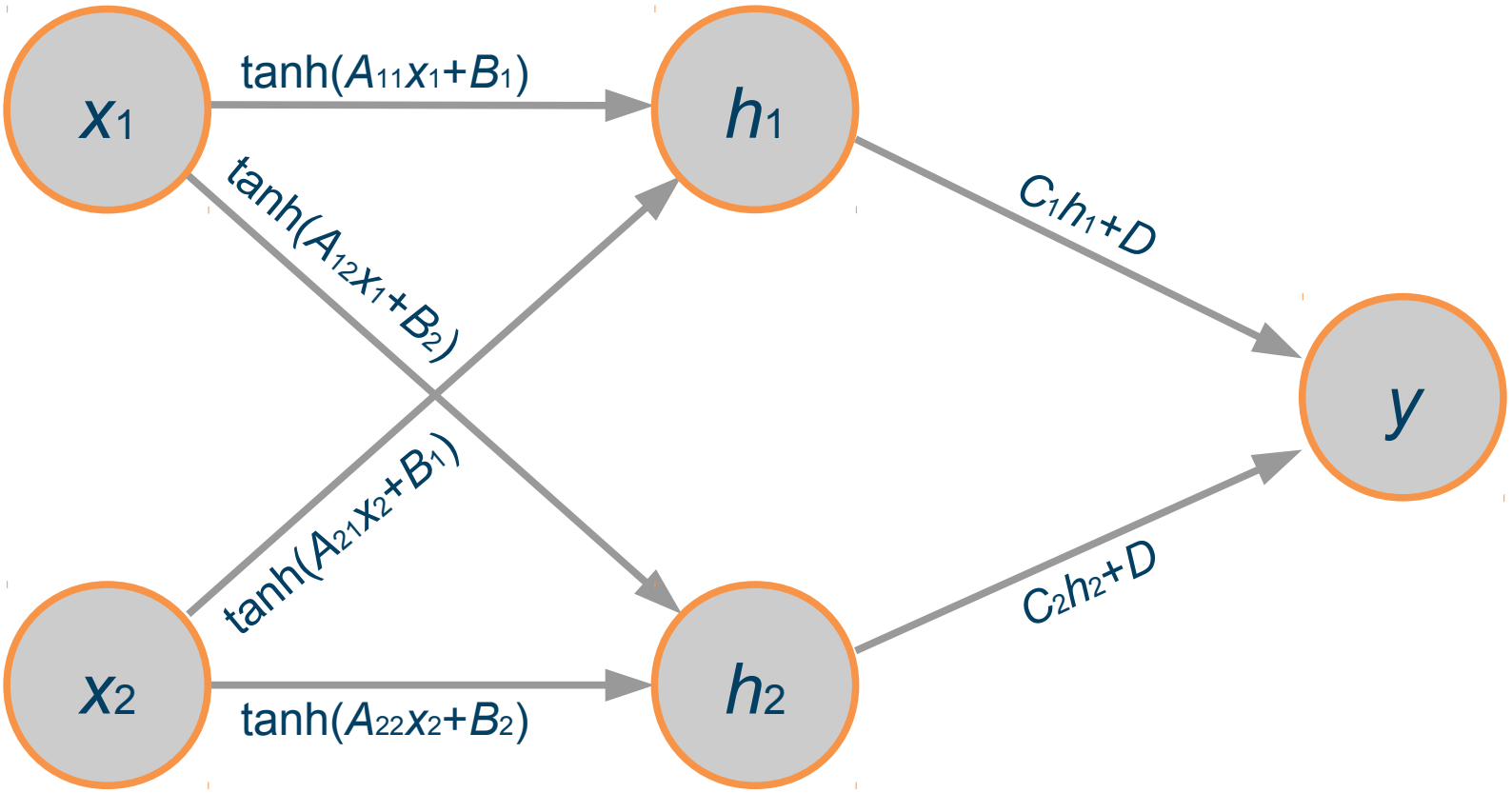
Materials design

Predict with complete data

Crystal structure

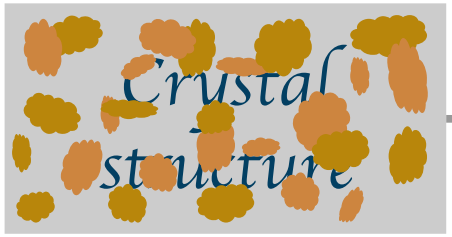
Crystal structure

Architecture



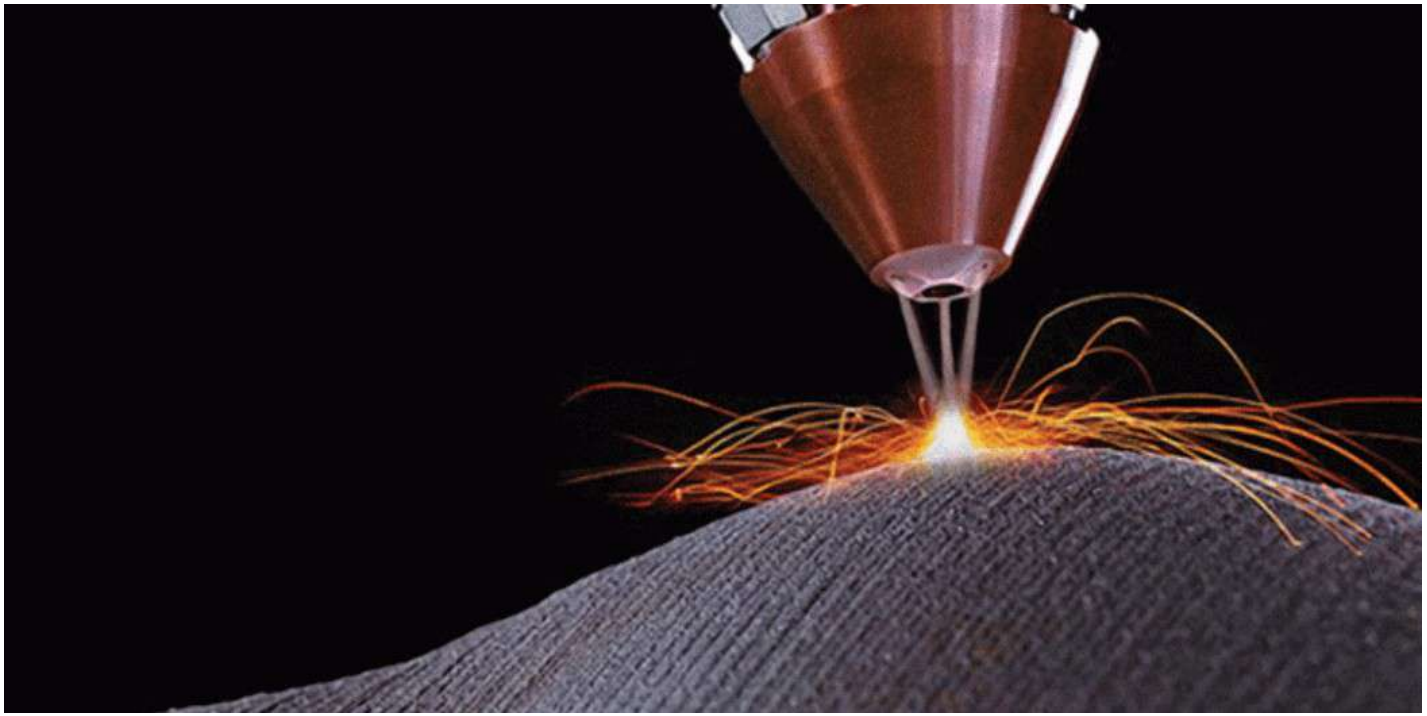
Train with fragmented data

Predict with fragmented data



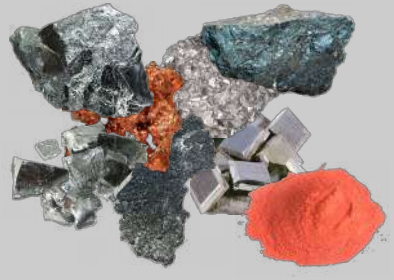
Crystal
structure

Direct laser deposition requires new alloys



Neural networks for materials design

Composition



Properties

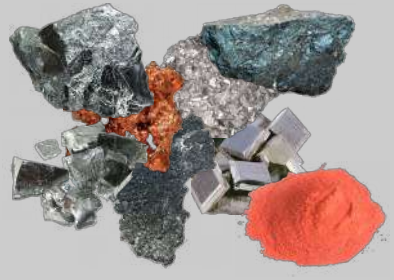
Process

Fatigue

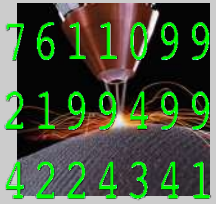
Welding

Neural networks for materials design

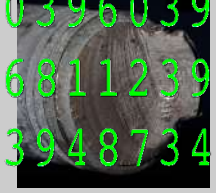
Composition



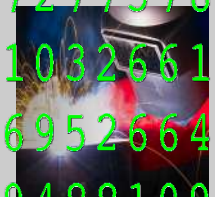
Properties



Process



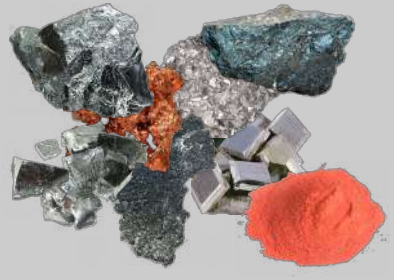
Fatigue



Welding

Neural networks for materials design

Composition



Properties

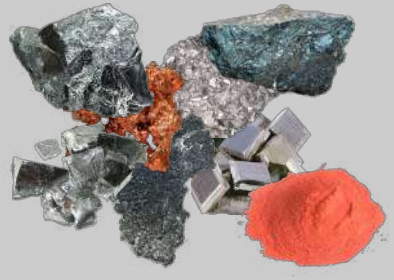
Process

Fatigue

Welding

Neural networks for materials design

Composition



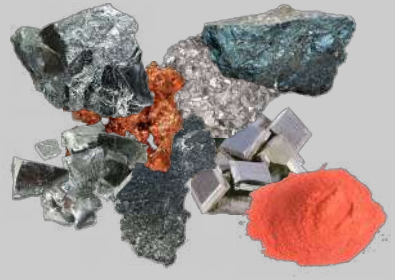
Properties

Process

Fatigue

Welding

Composition



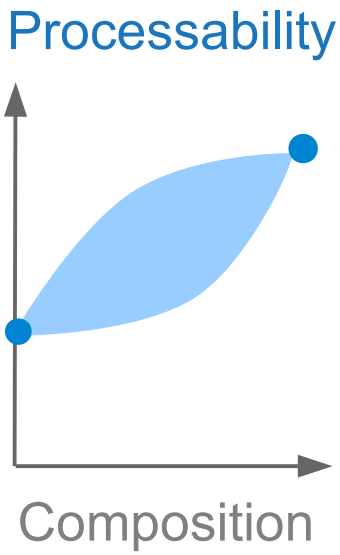
Properties

Process

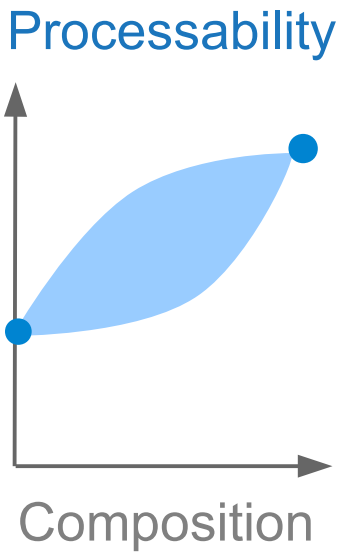
Fatigue

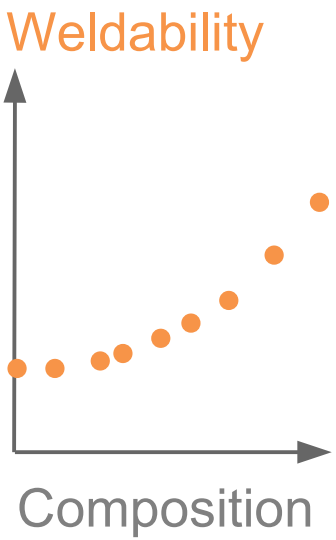
Welding

Insufficient processability results

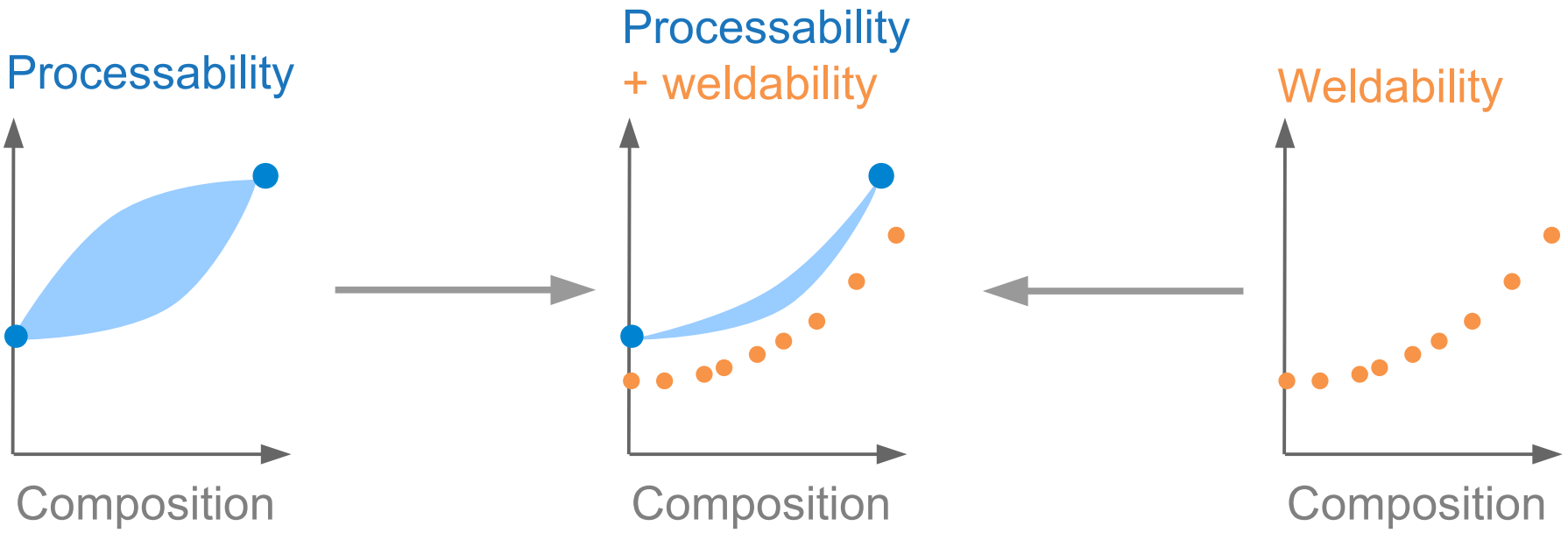


Welding is analogous to direct laser deposition

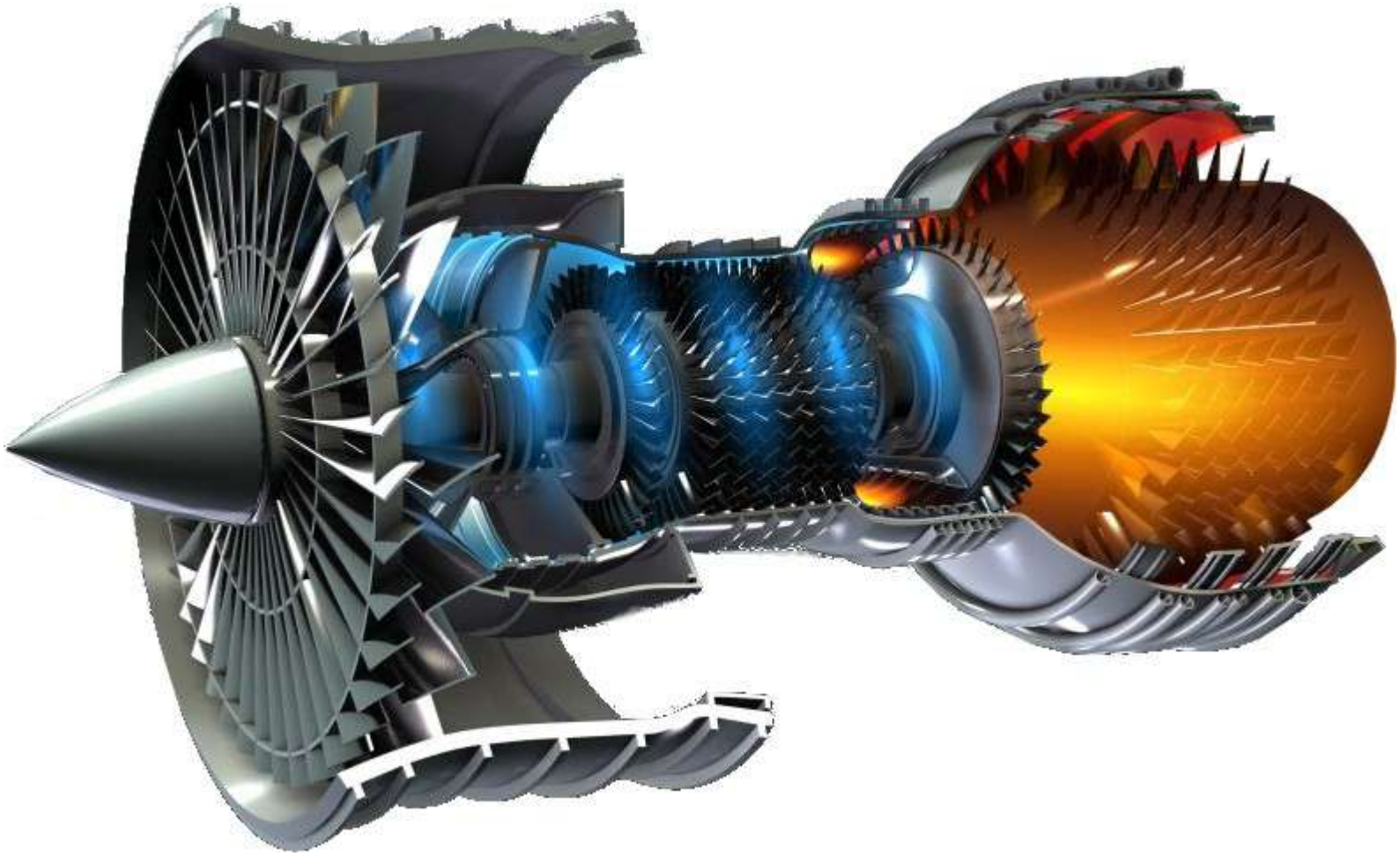




Merging properties with the neural network



Schematic of a jet engine



Target properties

Elemental cost	< 25 $\$kg^{-1}$
Density	< 8500 kgm^{-3}
γ' content	< 25 wt%
Oxidation resistance	< 0.3 $mgcm^{-2}$
Processability	< 0.15% defects
Phase stability	> 99.0 wt%
γ' solvus	> 1000˚C
Thermal resistance	> 0.04 $K\Omega^{-1}m^{-3}$
Yield stress at 900˚C	> 200 MPa
Tensile strength at 900˚C	> 300 MPa
Tensile elongation at 700˚C	> 8%
1000hr stress rupture at 800˚C	> 100 MPa
Fatigue life at 500 MPa, 700˚C	> 10^5 cycles

Composition

Cr: 19%

Co: 4%

Mo: 4.9%

W: 1.2%

Zr: 0.05%

Nb: 3%

Al: 2.9%

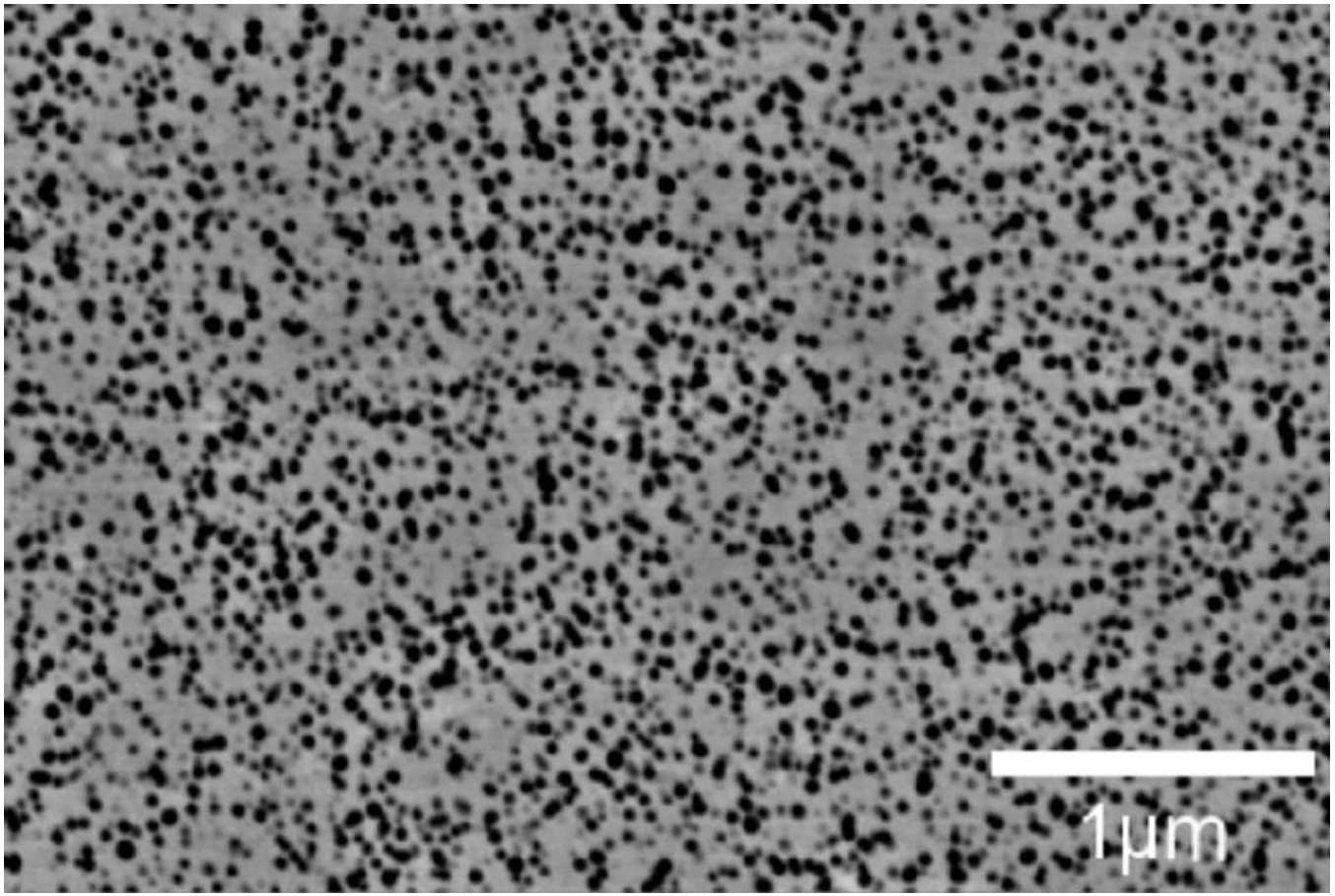
C: 0.04%

B: 0.01%

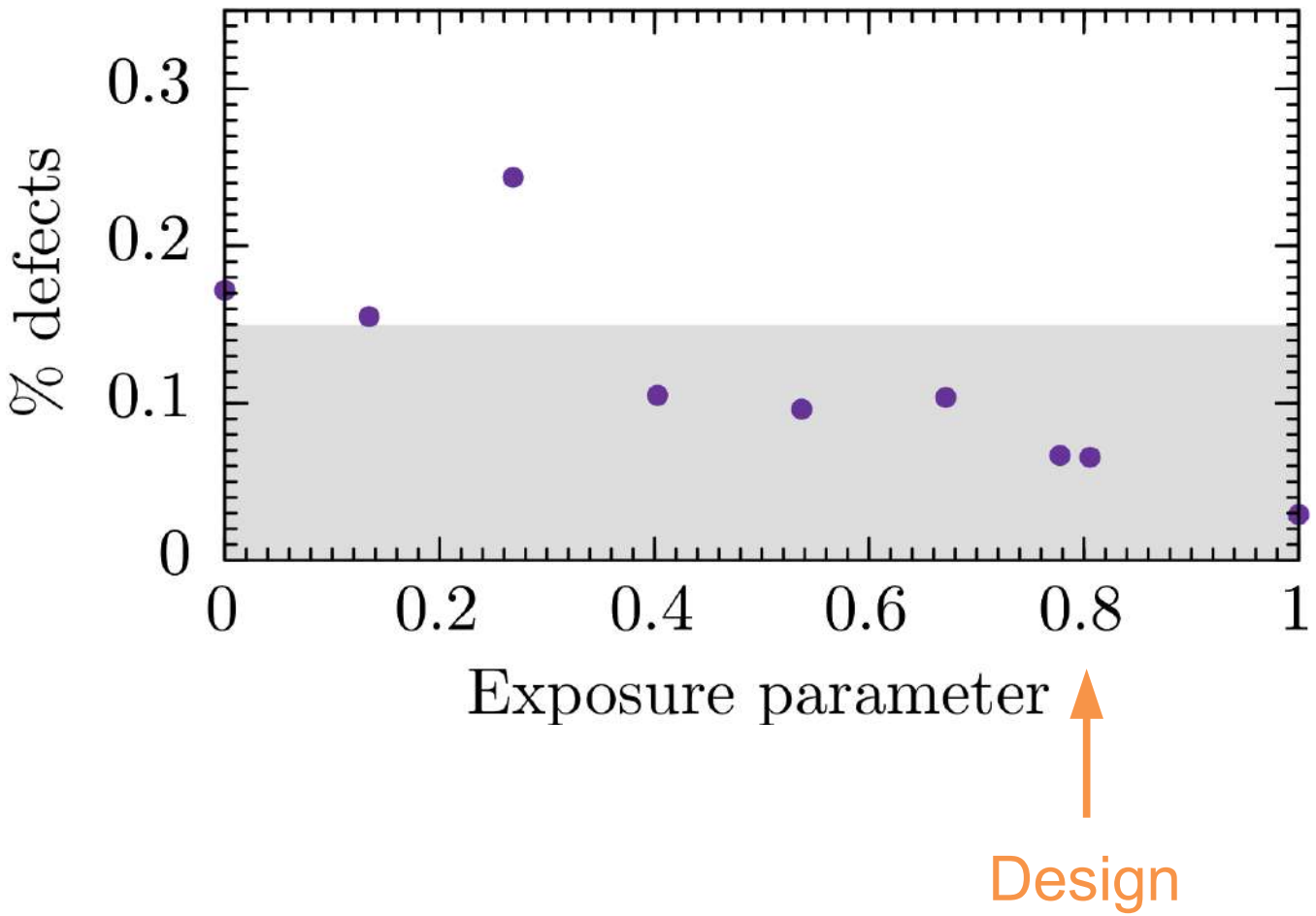
Ni

Expose 0.8

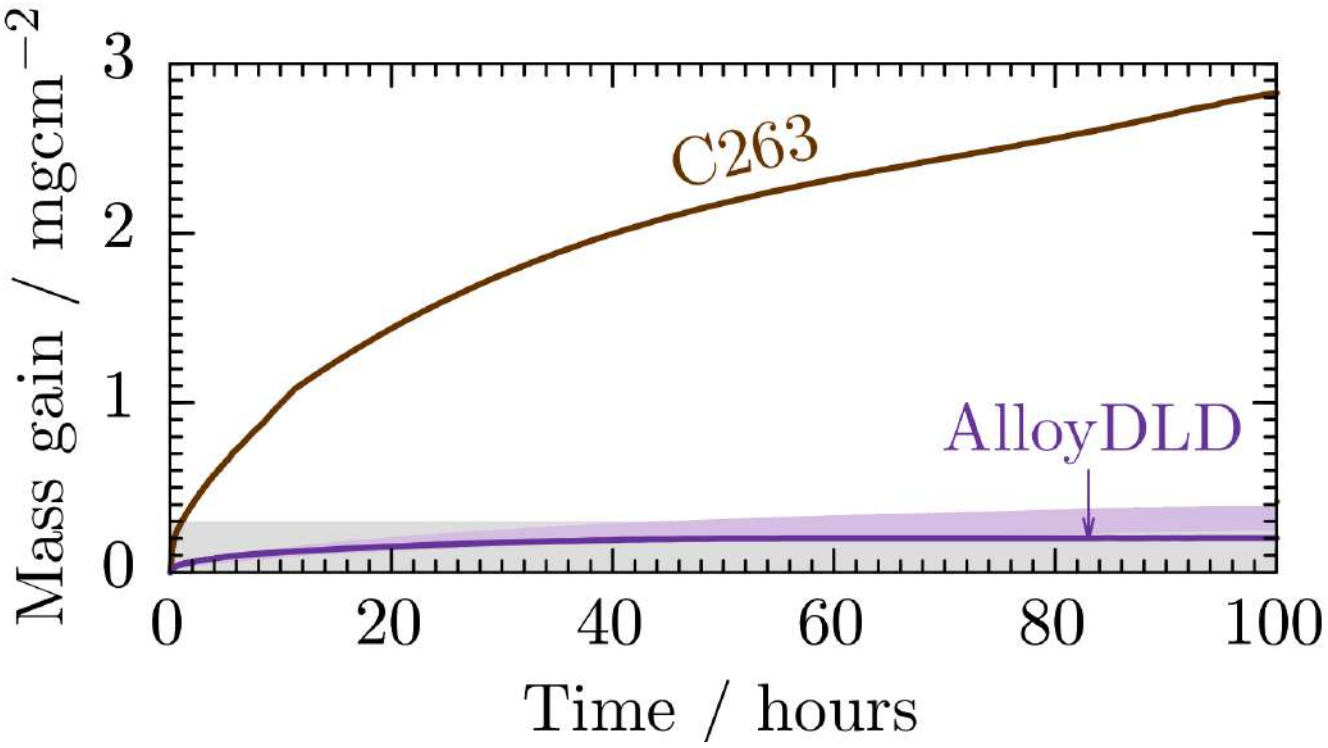
Microstructure



Testing the processability: horizontal printing



Testing the oxidation resistance

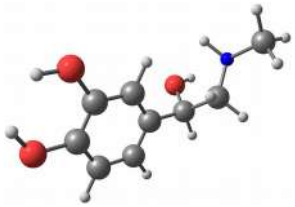


Printing a component for an engine

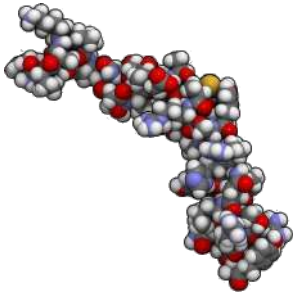
Materials Solutions

ROLLS ROYCE

Action of a drug



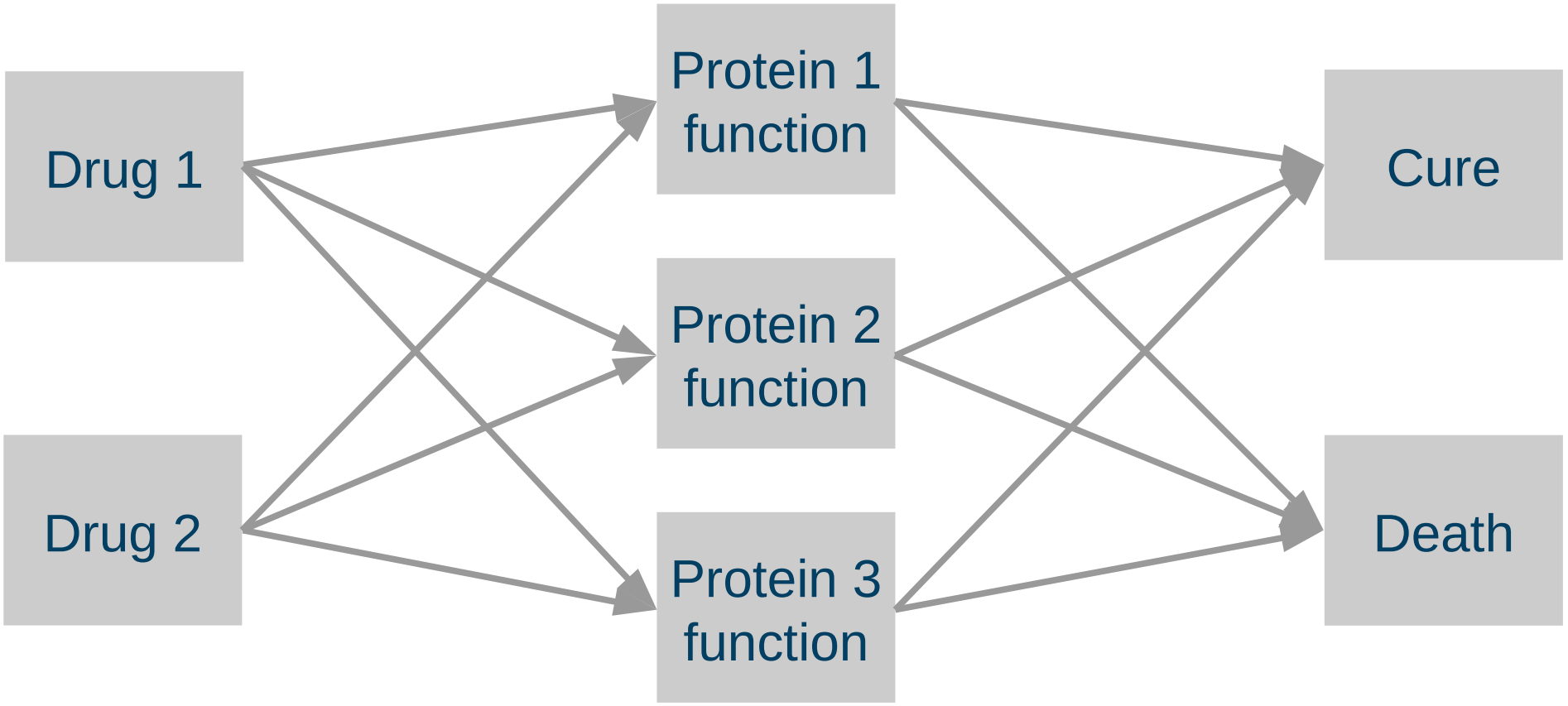
Drug



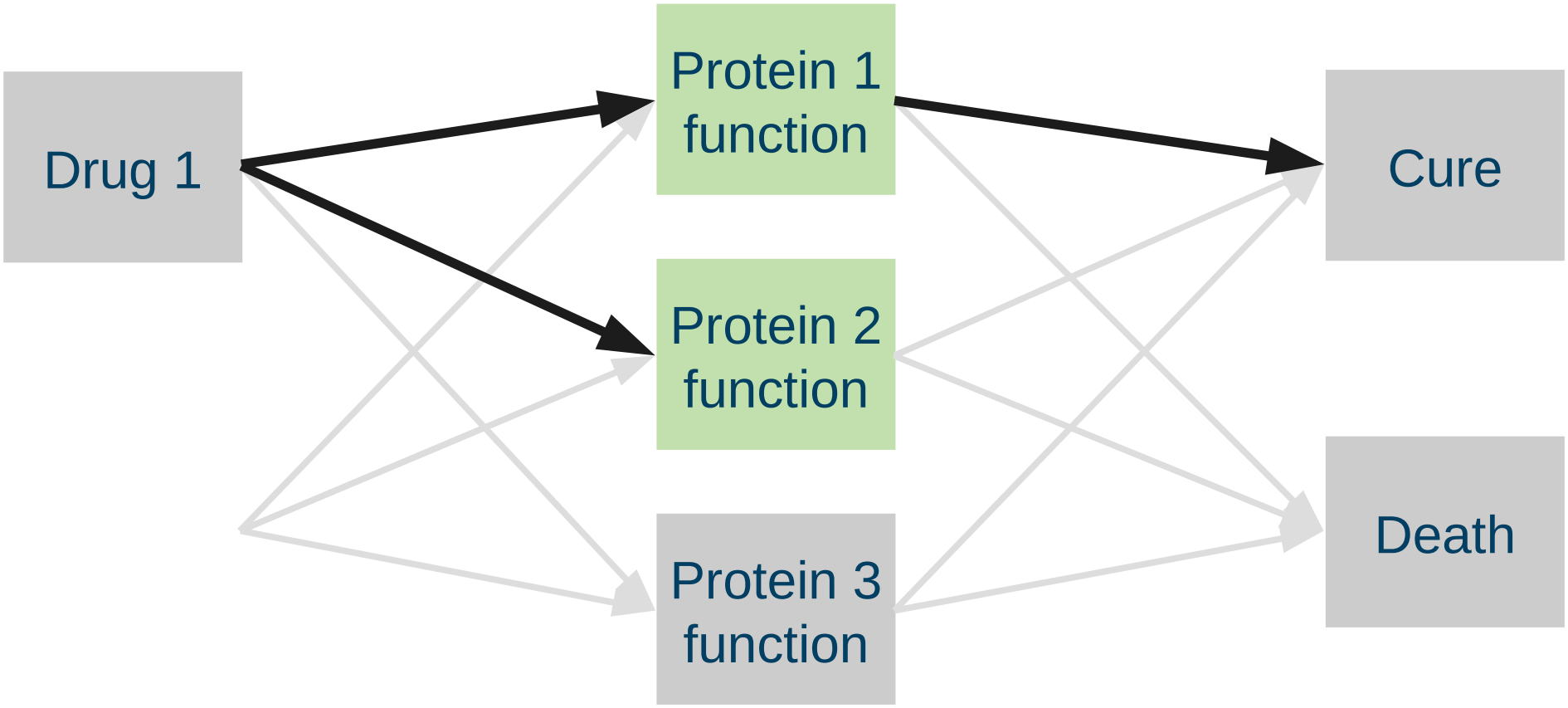
Protein

Effect

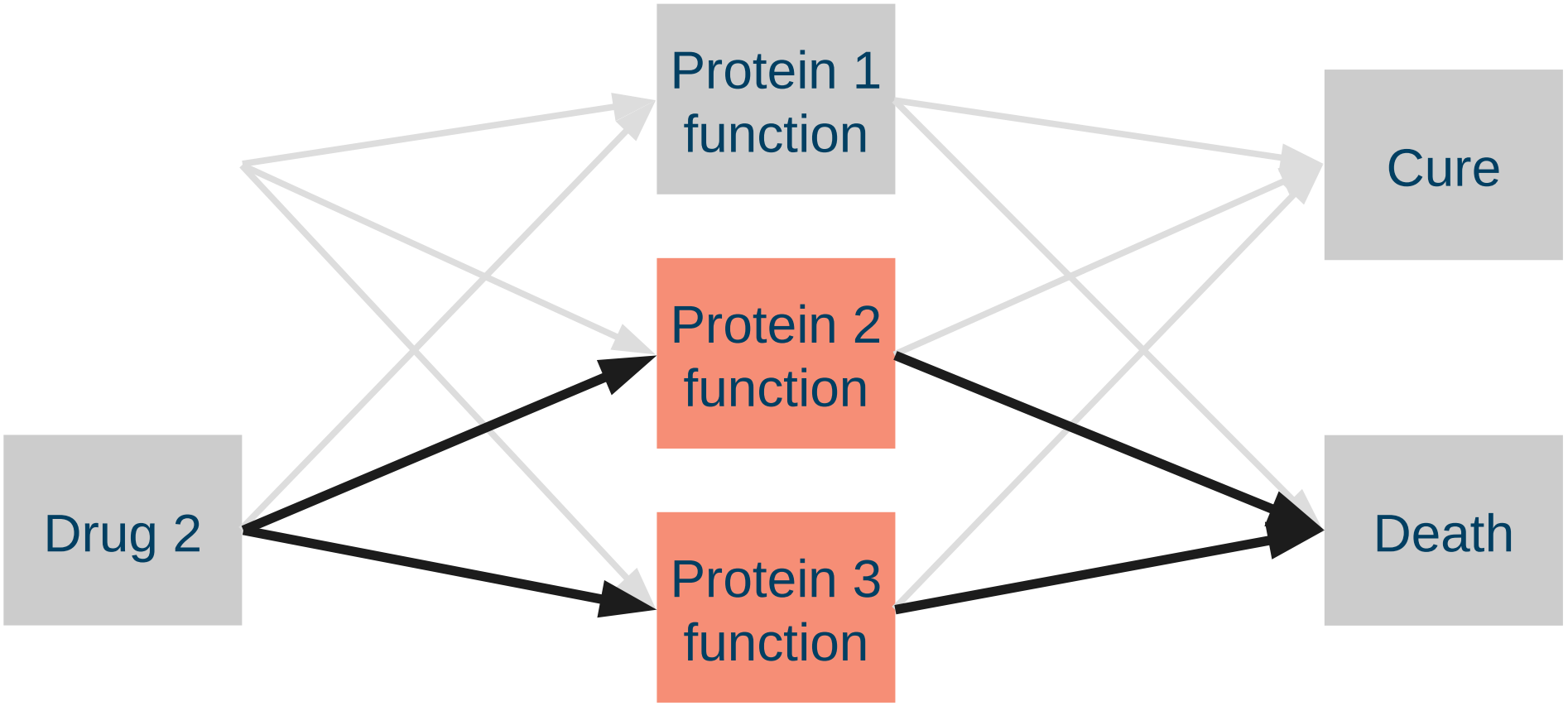
Drug discovery



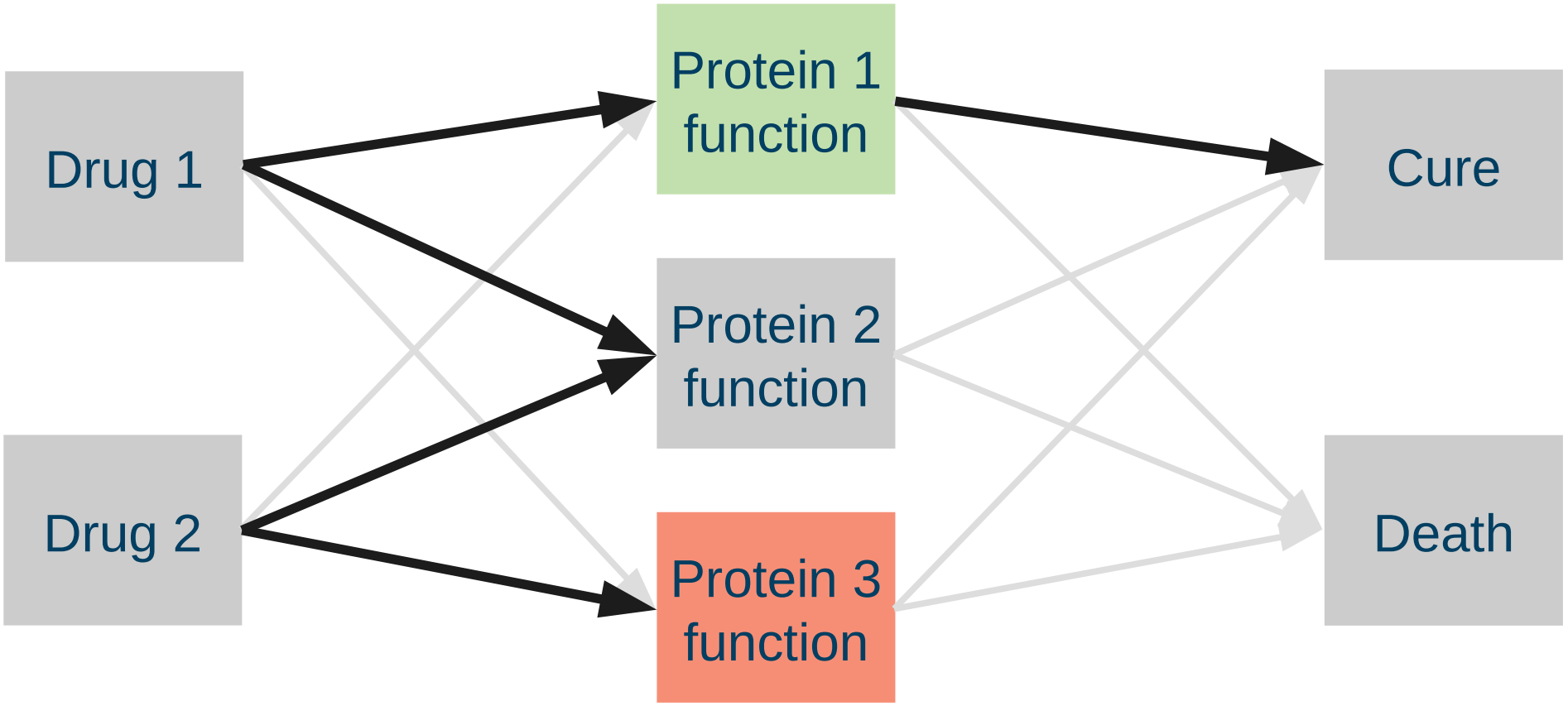
Drug discovery



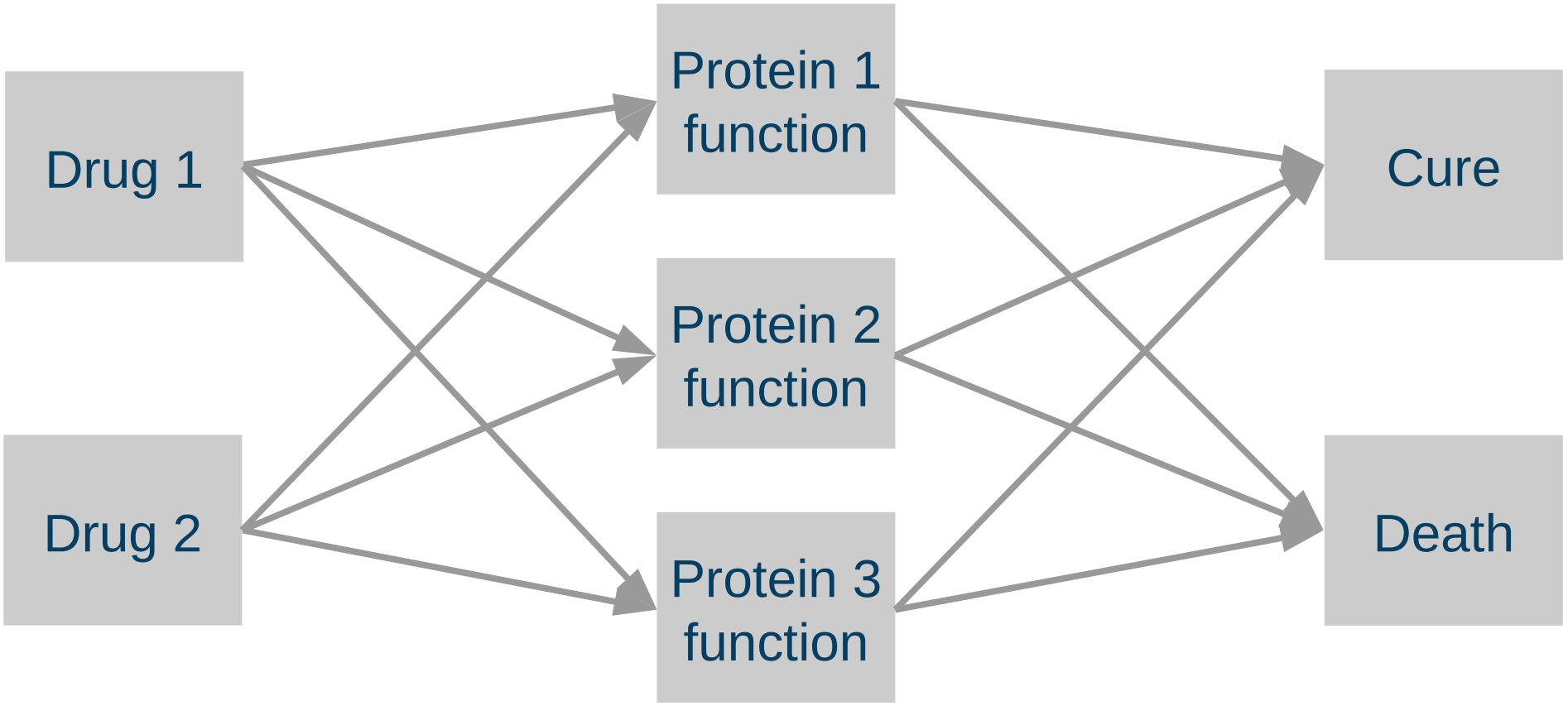
Drug discovery



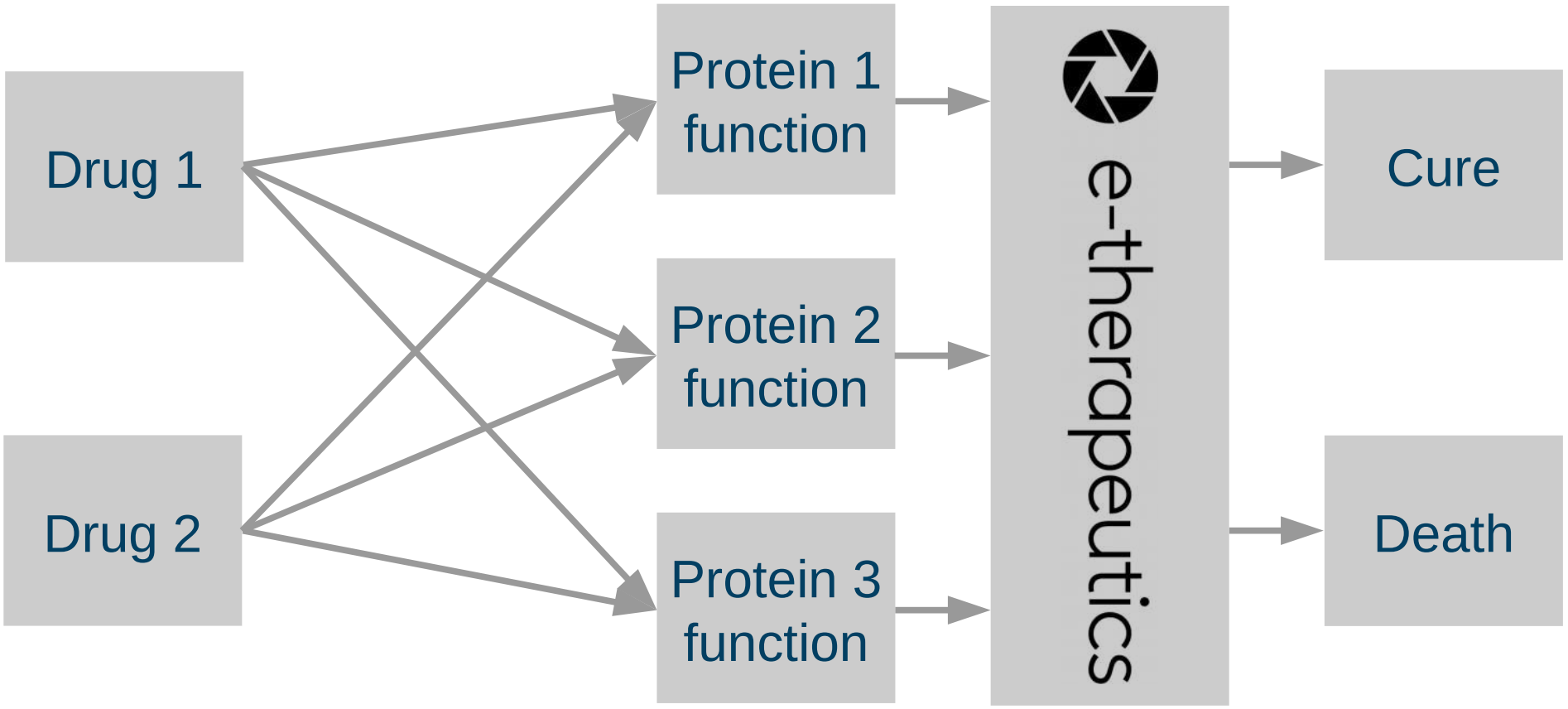
Drug discovery



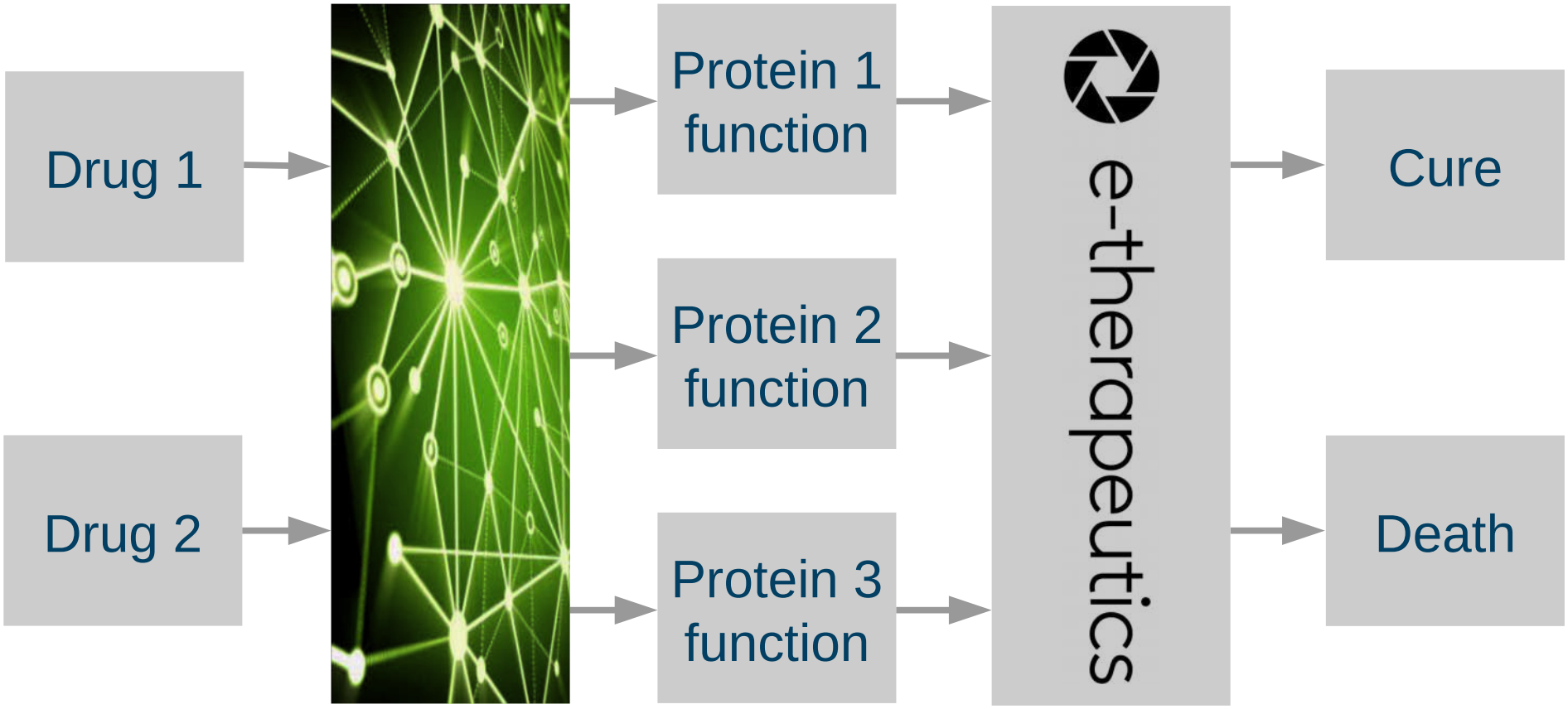
Drug discovery



Drug discovery

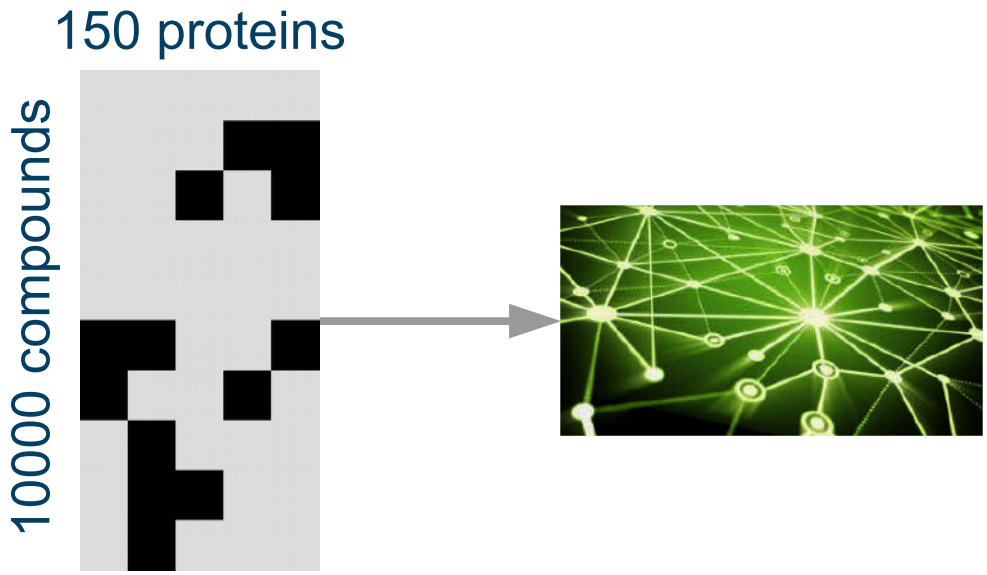


Drug discovery



Novartis dataset for benchmarking machine learning

150 proteins with 10000 compounds, data set 5% complete



Martin, Valery, Polyakov, Tian, and Perez
J. Chem. Inf. Model. **57**, 2077 (2017)

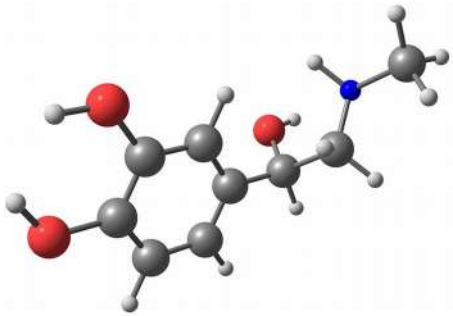
optibrium

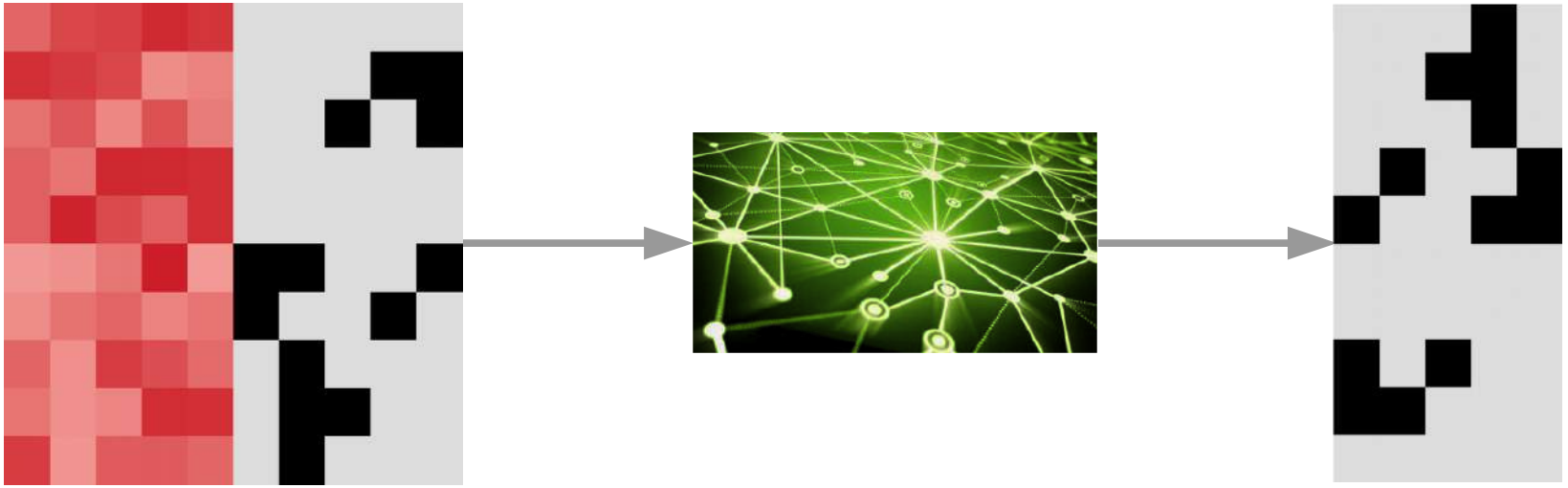
Want to impute missing entries

Validate using a holdout data set

optibrium

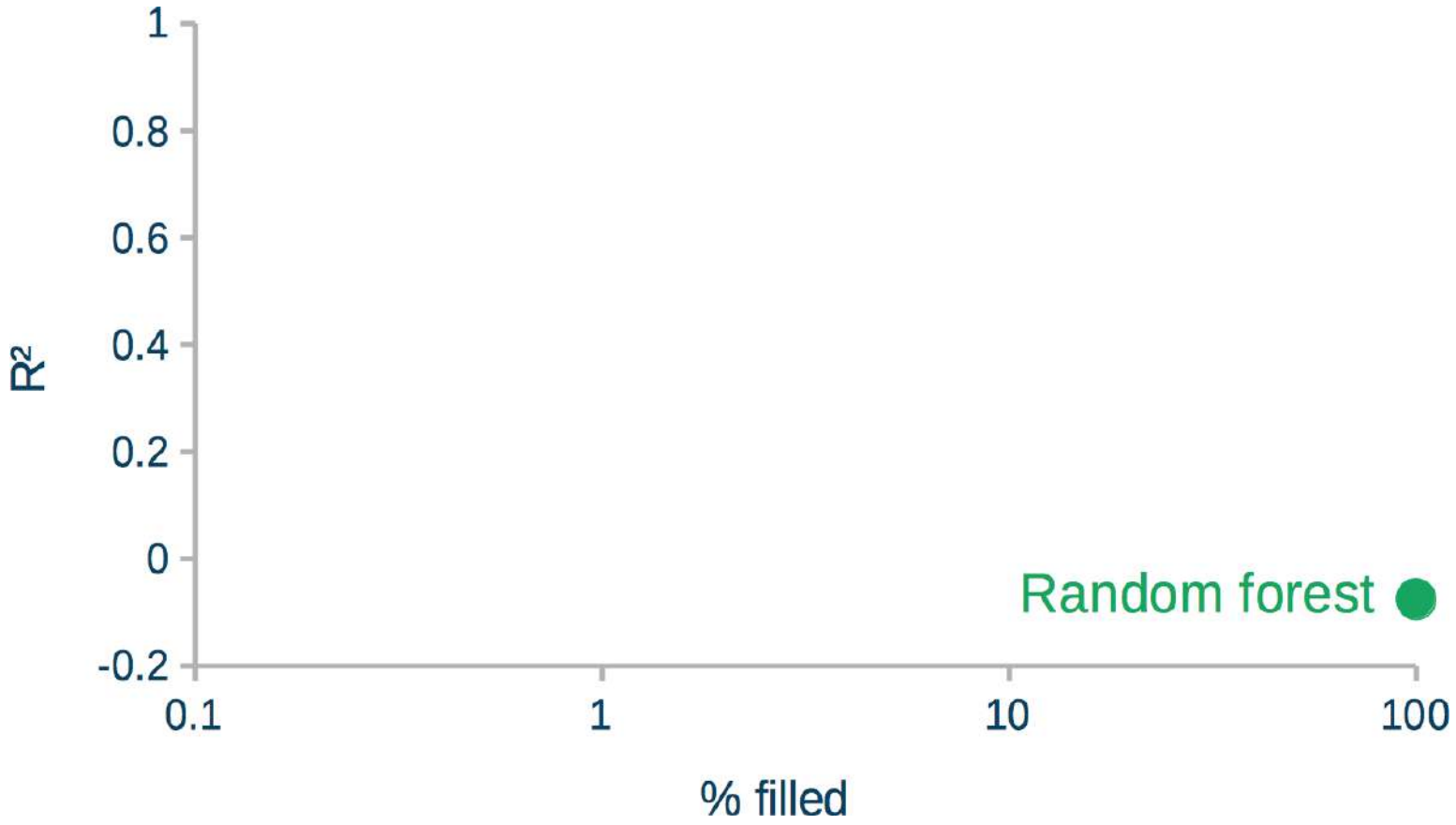
QSAR: quantitative structure-activity relationships





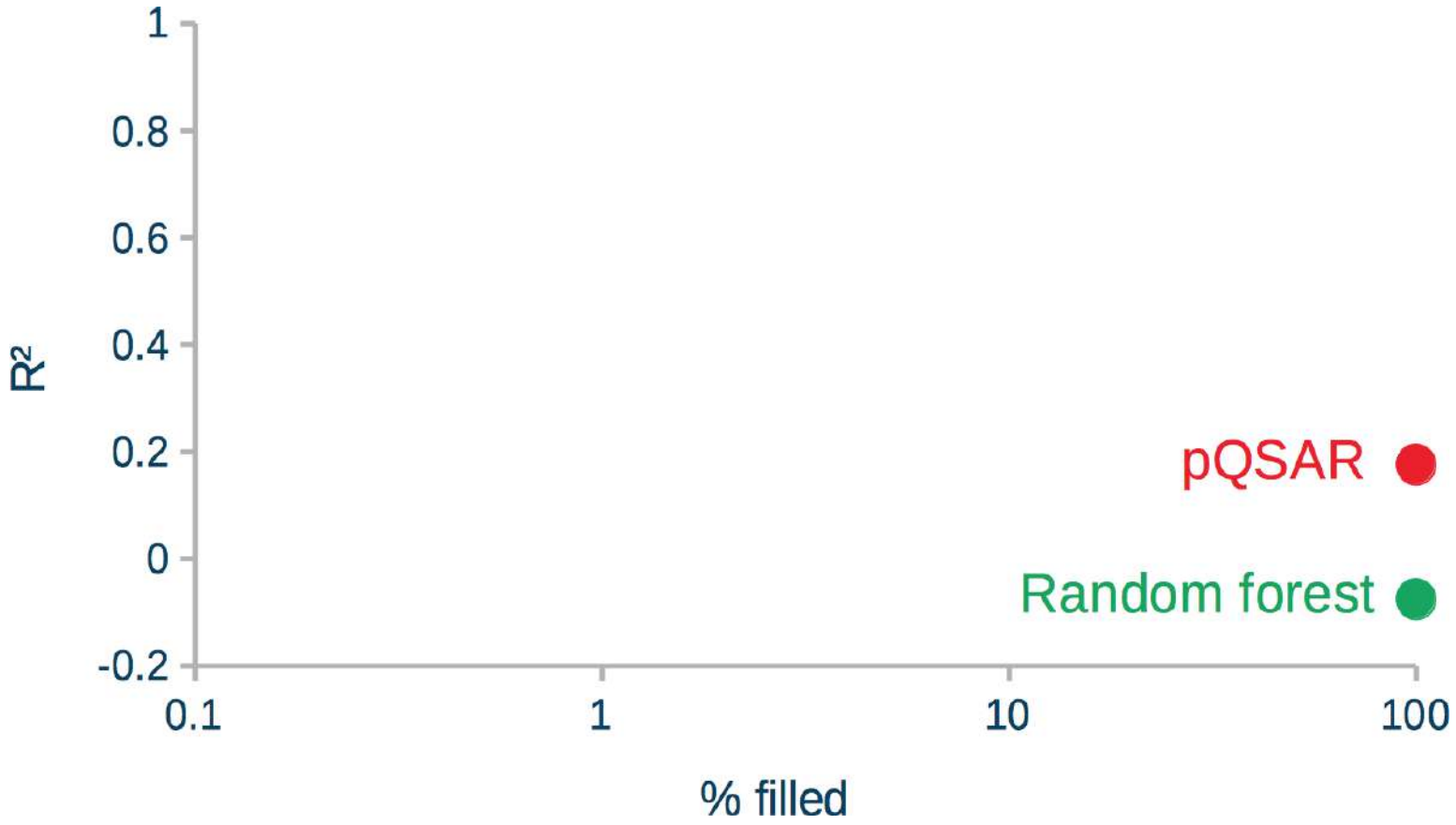
optibrium

Random forest



optibrium

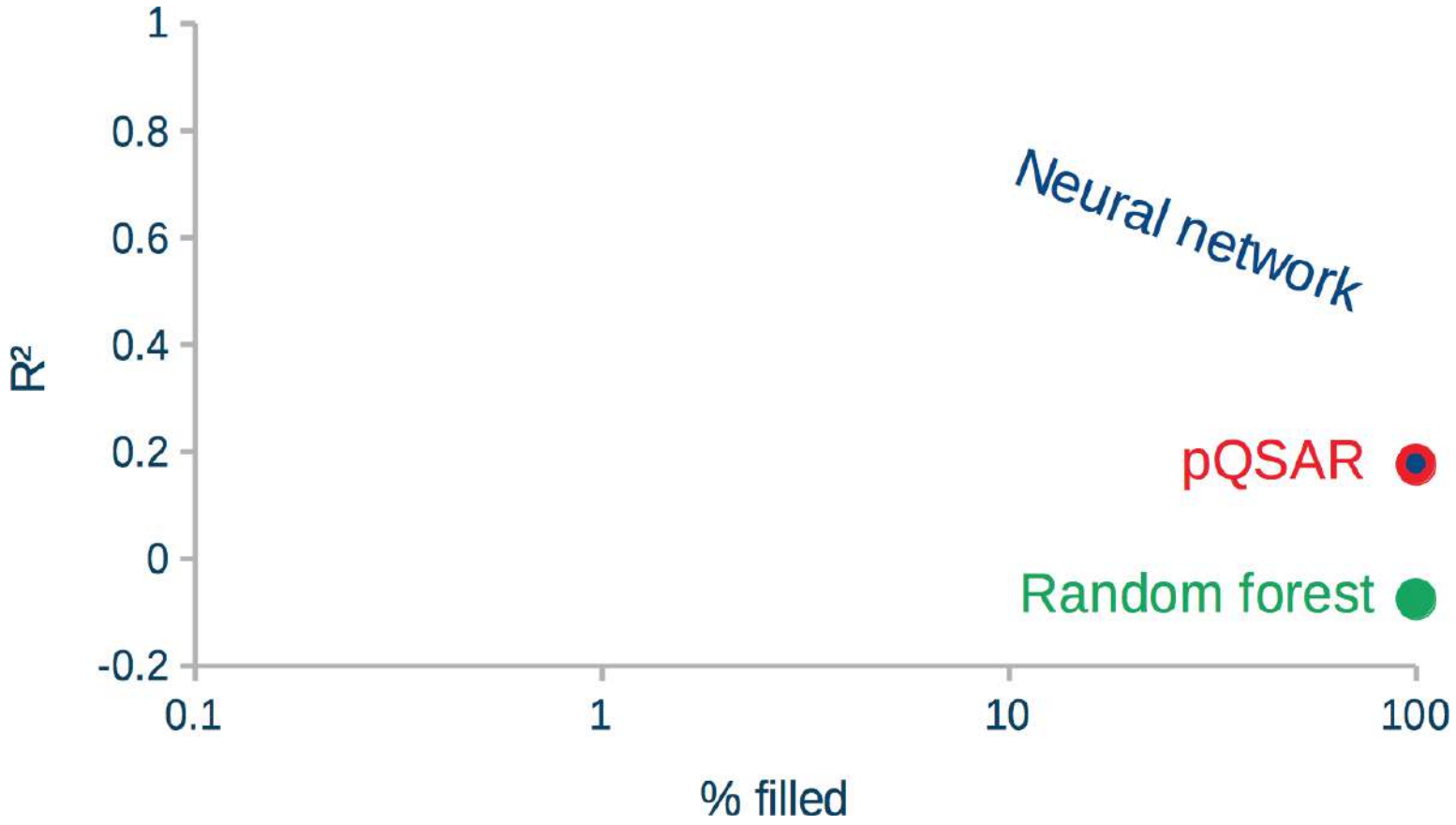
pQSAR



Martin, Valery, Polyakov, Tian, and Perez
J. Chem. Inf. Model. **57**, 2077 (2017)

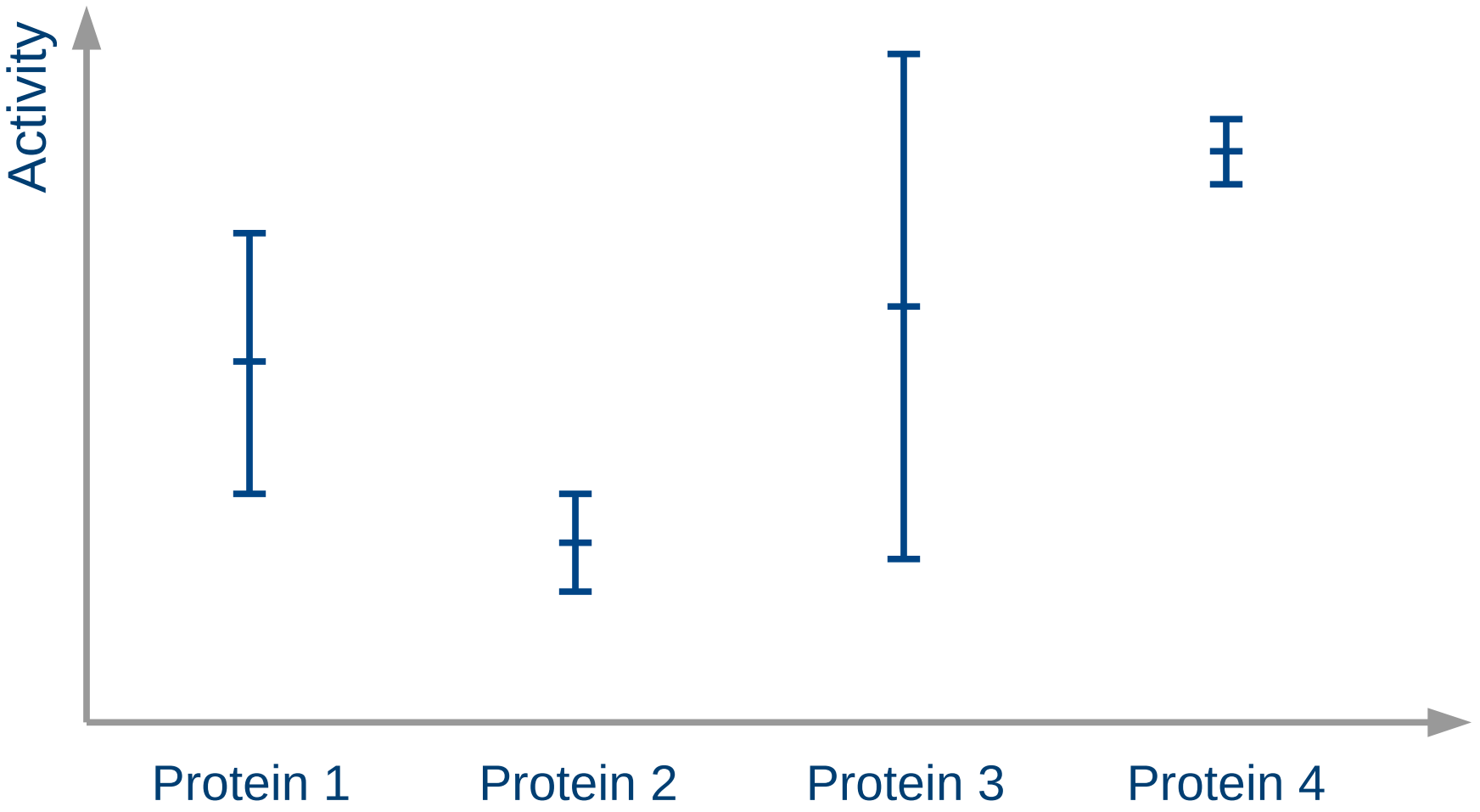
optibrium

Neural network without uncertainties



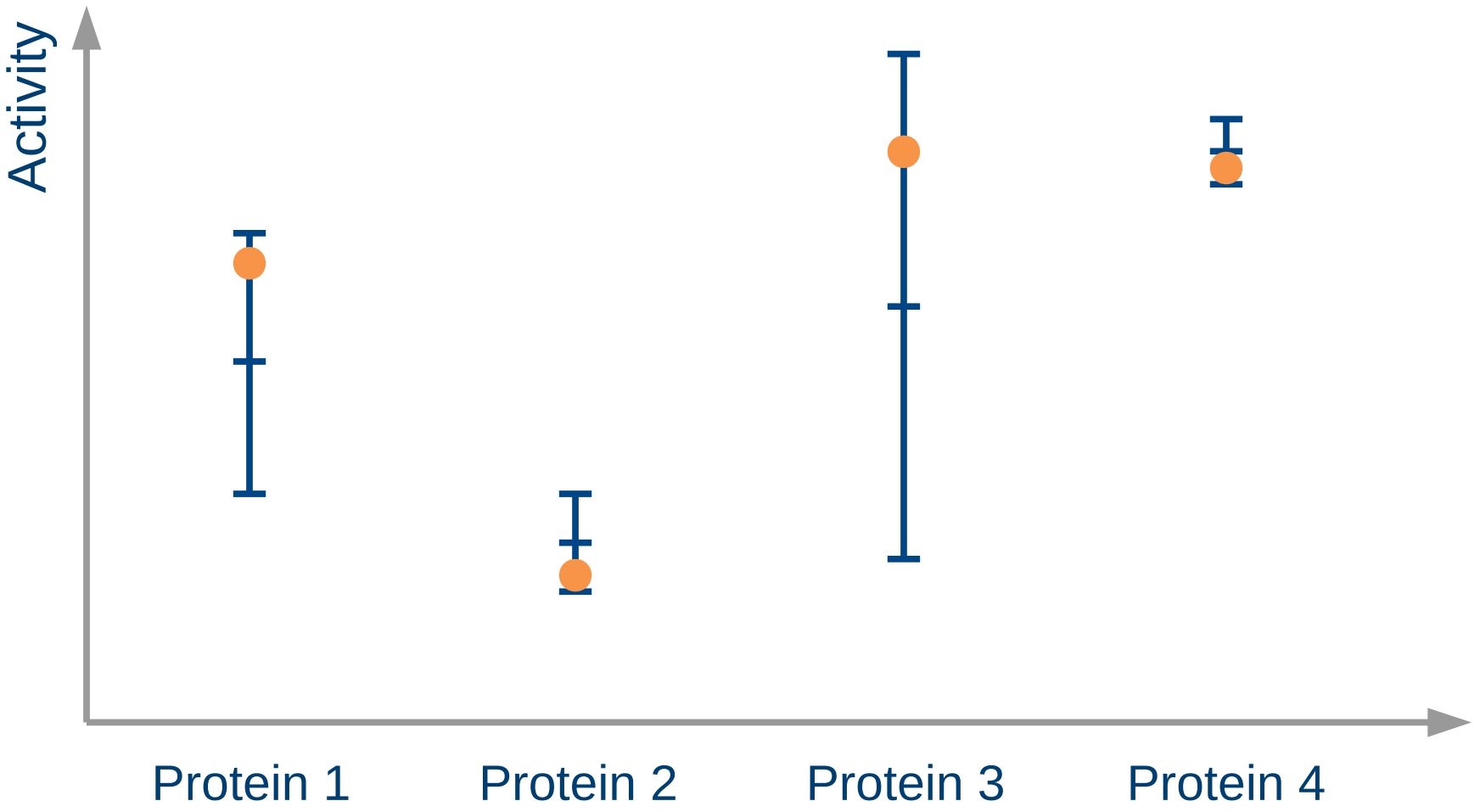
optibrium

Predicted activities have an uncertainty



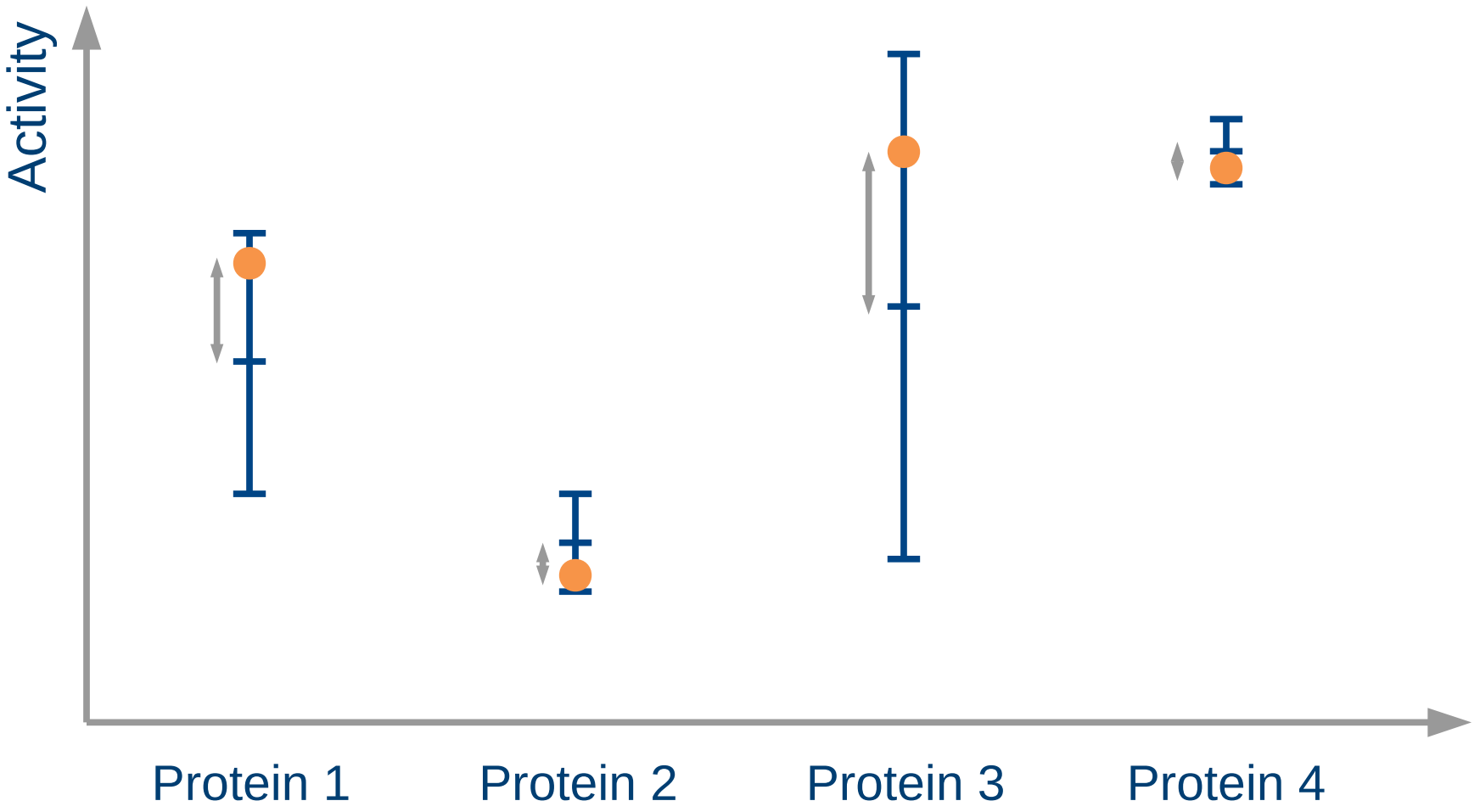
optibrium

Validation data should be within one standard error



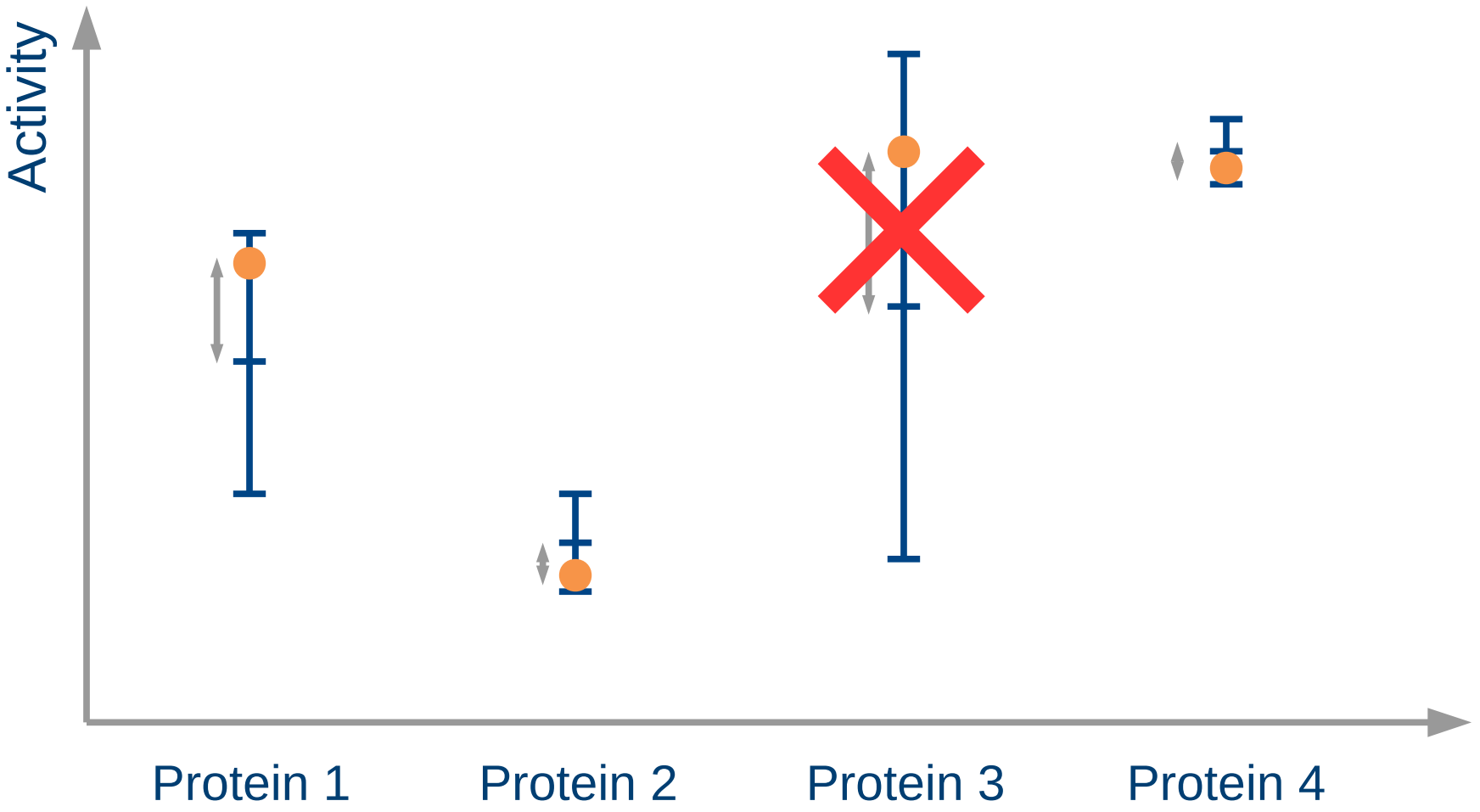
optibrium

R^2 metric calculated with difference from mean



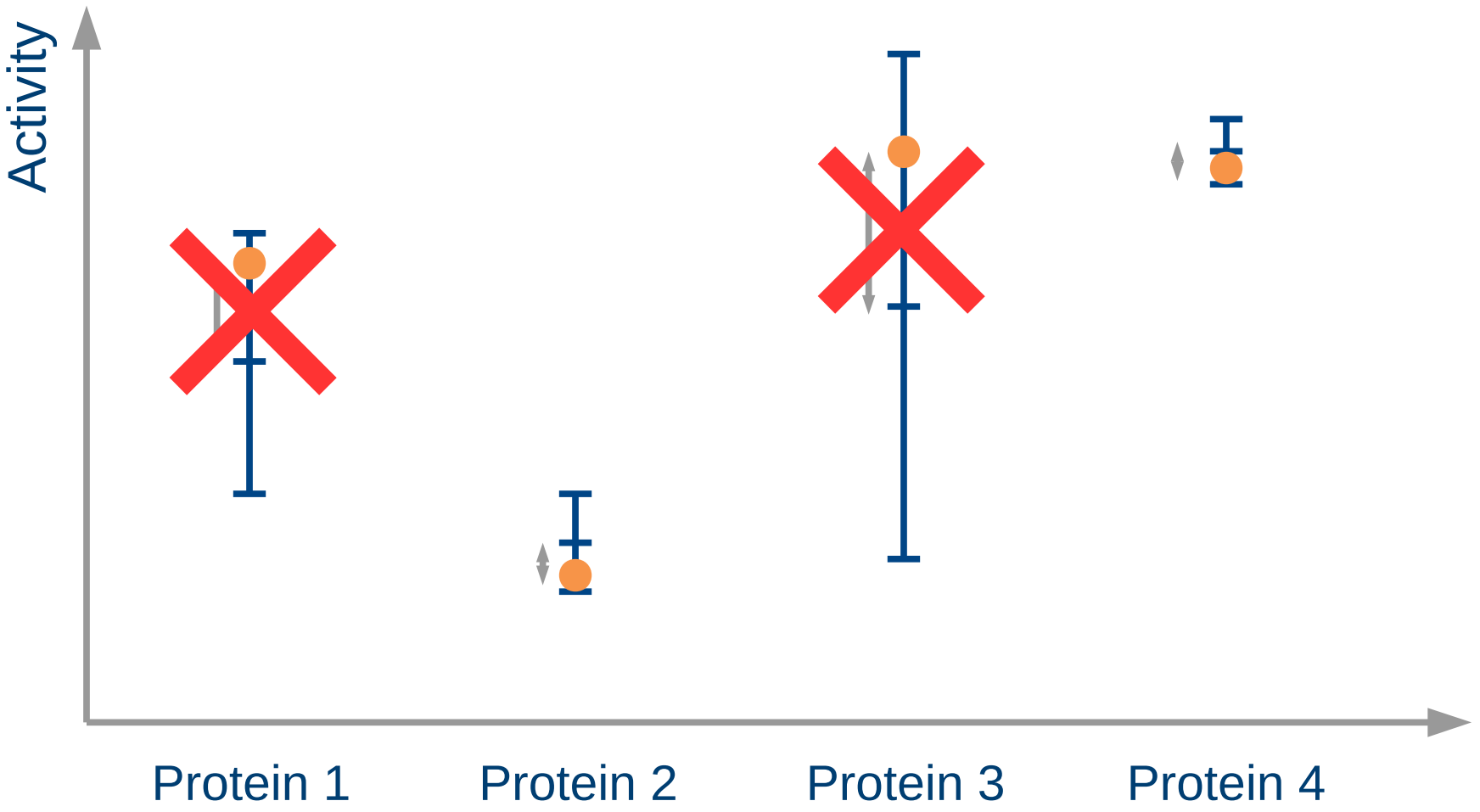
optibrium

Impute 75% of data with smallest uncertainty



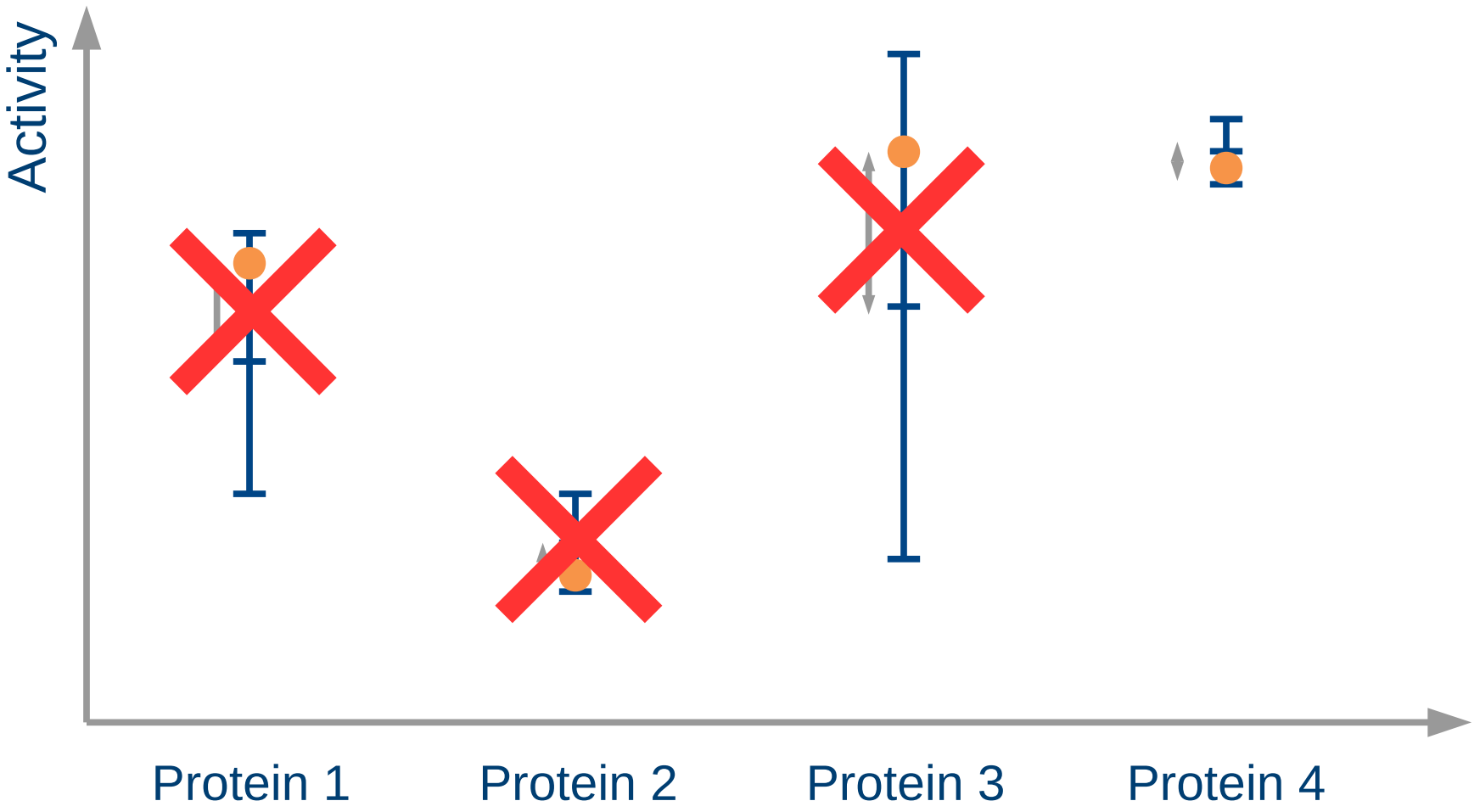
optibrium

Impute 50% of data with smallest uncertainty



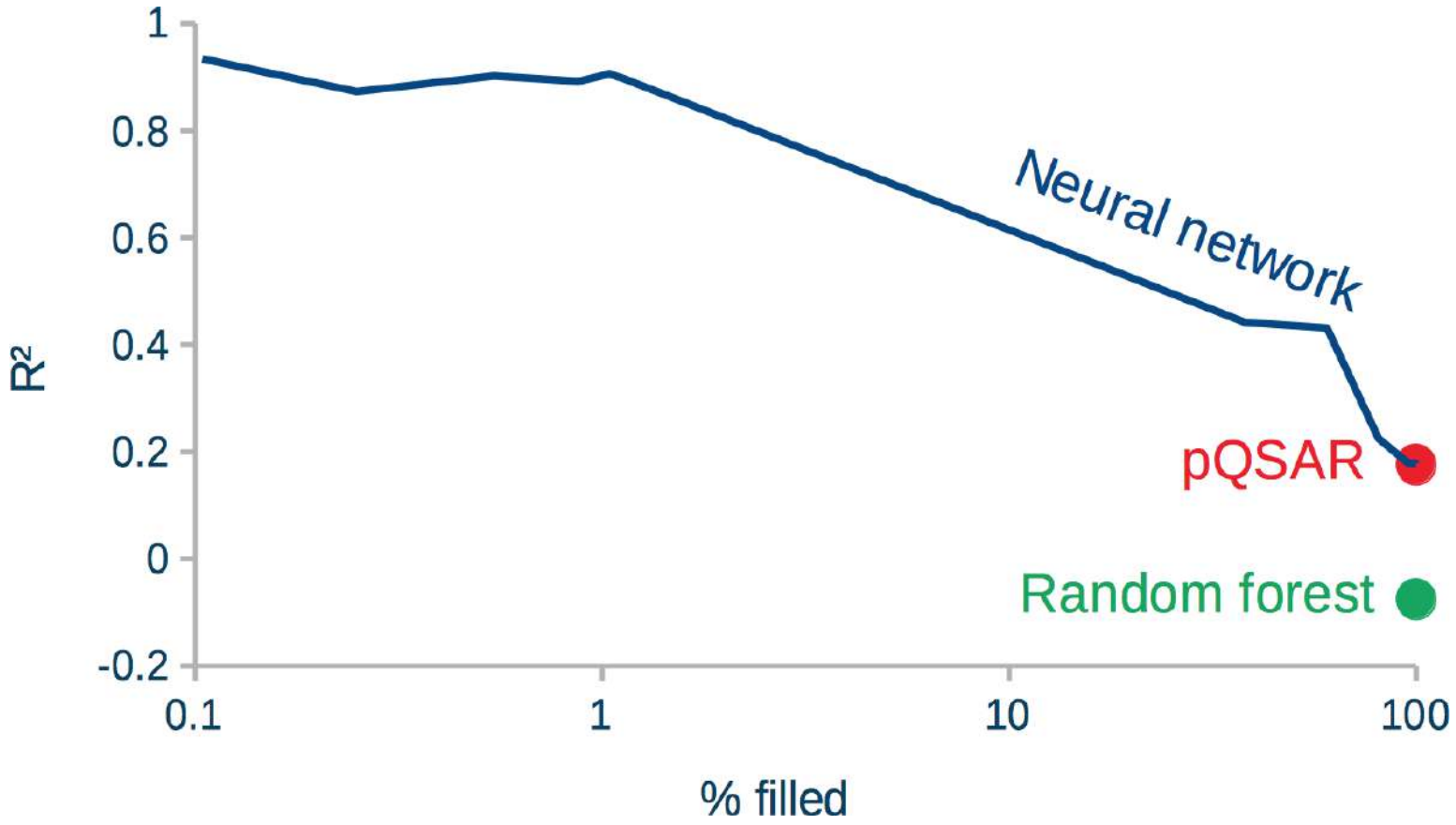
optibrium

Impute 25% of data with smallest uncertainty



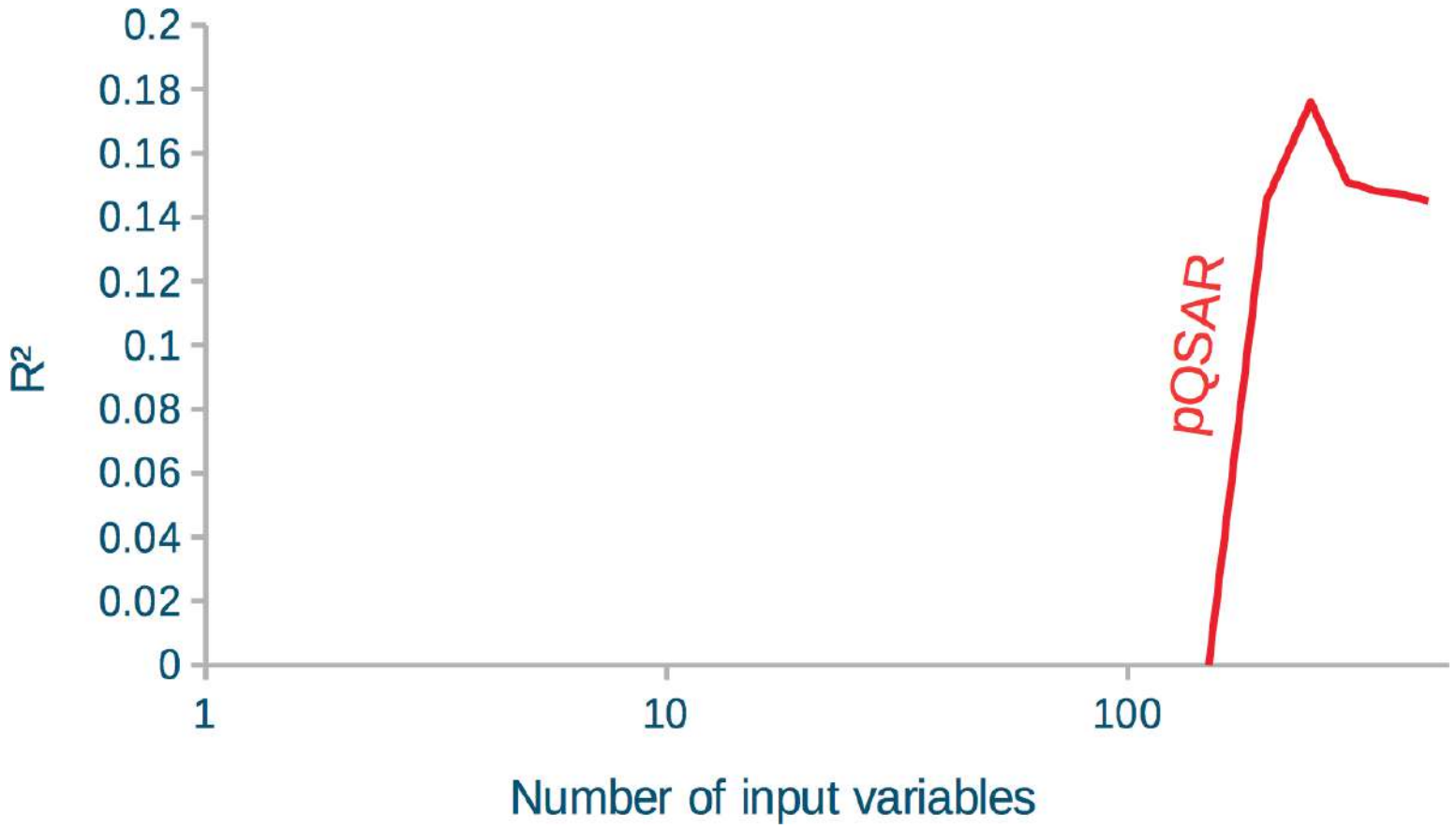
optibrium

Improved performance by exploiting uncertainties



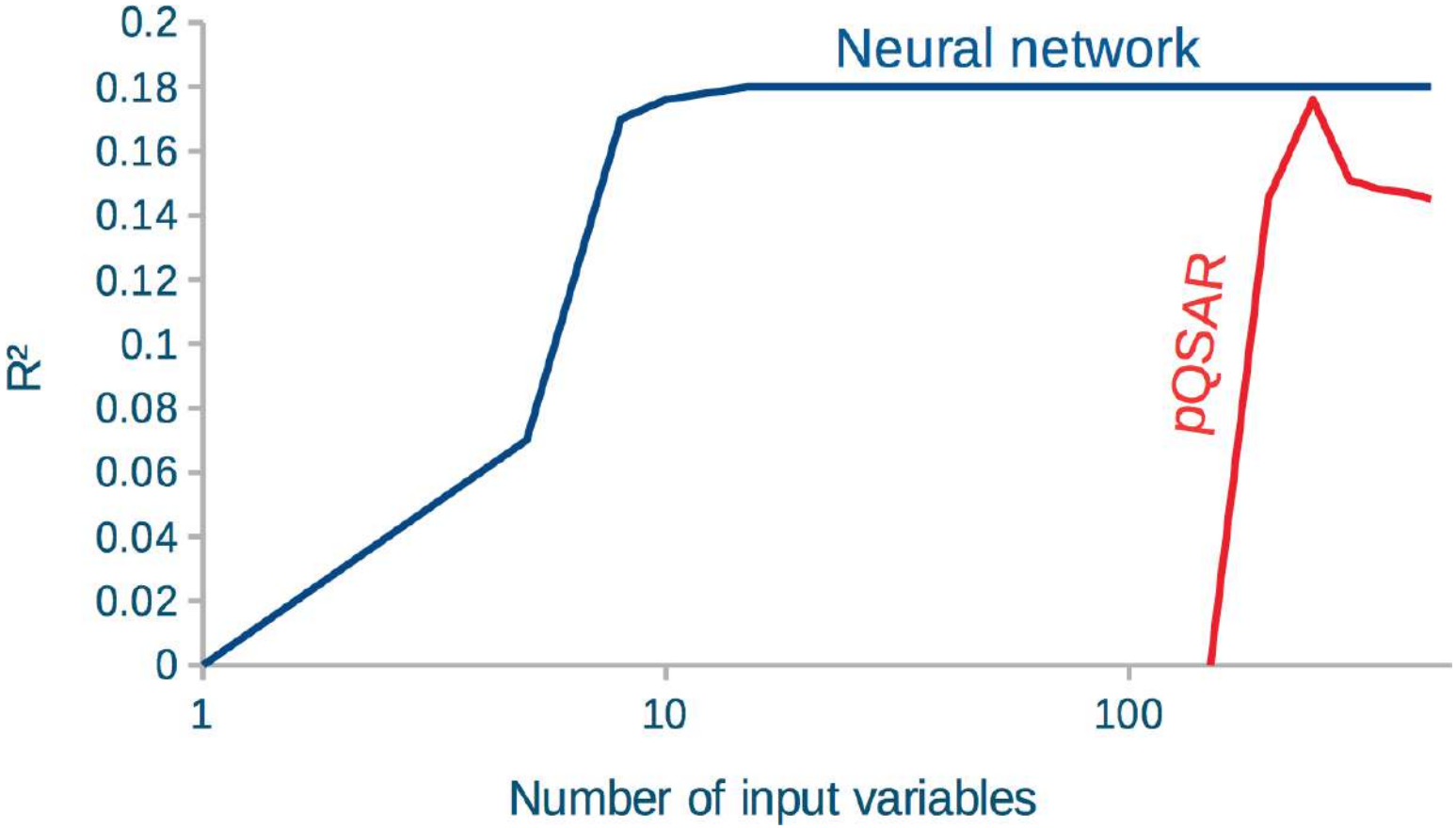
optibrium

Number of descriptors required for pQSAR



optibrium

Number of descriptors required for the neural network



optibrium

Experimental validation

Validated by another company who identified new proteins and experimentally measured the protein activity levels

Achieved 82% level of predictability, in concurrence with the level of filling selected

Summary

Merge different experimental quantities and computer simulations into a holistic design tool

Designed and experimentally verified alloy for direct laser deposition

Drug discovery by predicting protein activities with uncertainties