UNIVERSITY OF CAMBRIDGE

Machine learning for data-driven design of materials and molecules

Gareth Conduit

Model sparse datasets by exploiting property-property relationships

Merge data, computer simulations, and physical laws

Exploit uncertainty to focus on most robust designs

Reduce costly experiments to accelerate discovery

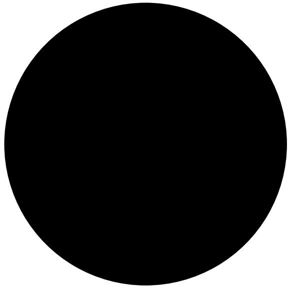
Commercialized as Alchemite™ by Intellegens

Challenge: *What other physical information could be extracted from a dataset?*

Black box machine learning for materials design

Properties

Composition



Defects

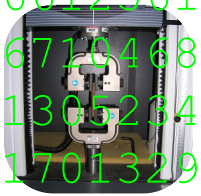
Fatigue

Strength

Train the machine learning

3870454990176143
6412046921823707
Composition
6488783419689686
1181558158737756
4102468240322648
3464176636980663
7857681349204530
2240819727856471
9839630878154322
1166912246415911

6488348704023749
Properties
8012754834215237
2272034535605225
8174681581845845
Defects
6224043080254884
1132084110885053
4535407535613613
8023131978697758
Fatigue
5488943723909634
8211320709698126
2677601866125614
0045818967104685
Strength
6583375313052341
2052413617013298
5560923129005660



Machine learning predicts material properties

Composition

Properties

Defects

Fatigue

Strength

Nickel superalloys with Rolls Royce University Technology Centre

Dr Vadegadde Duggappa

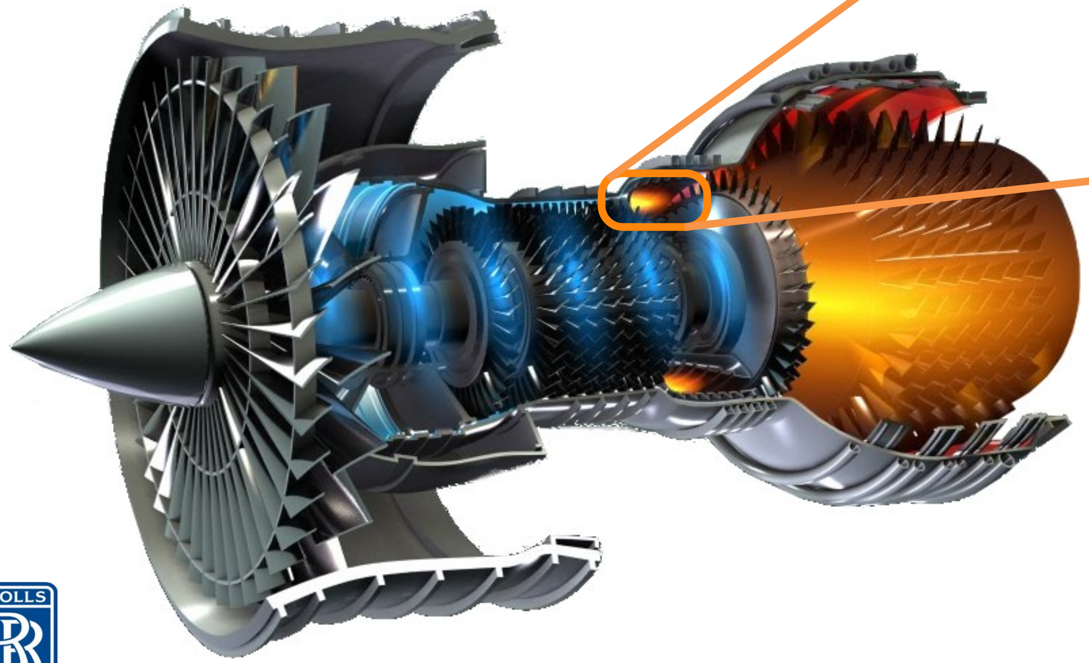
Dr Bryce Conduit

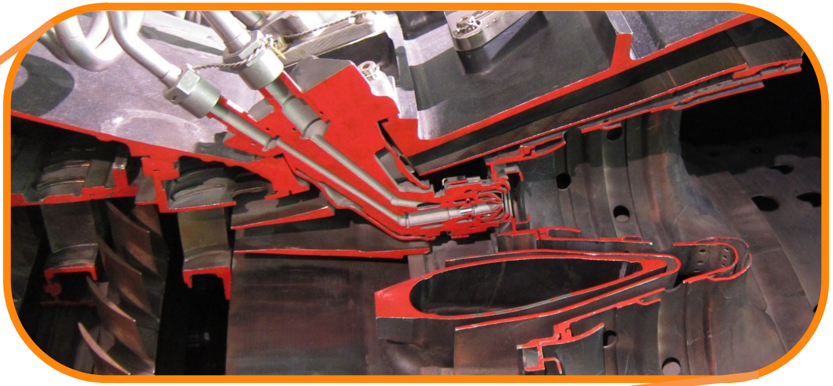
Professor Howard Stone

Dr Gareth Conduit

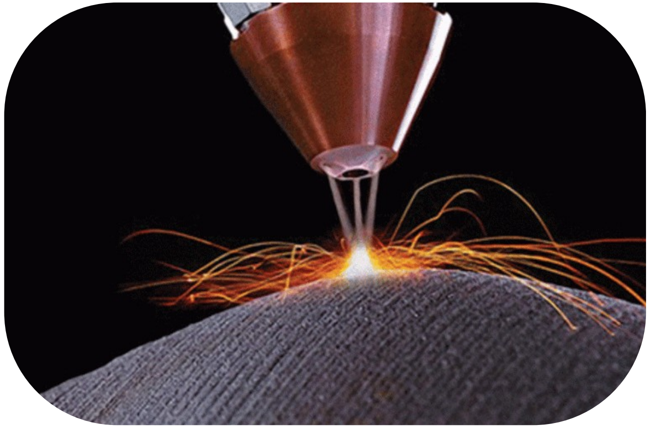
Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design **168**, 107644 (2019)

Combustor in a jet engine



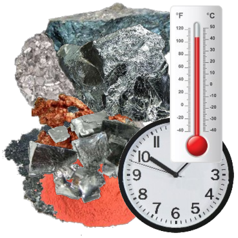


Defects form during printing



Laser

Data available to model defect density



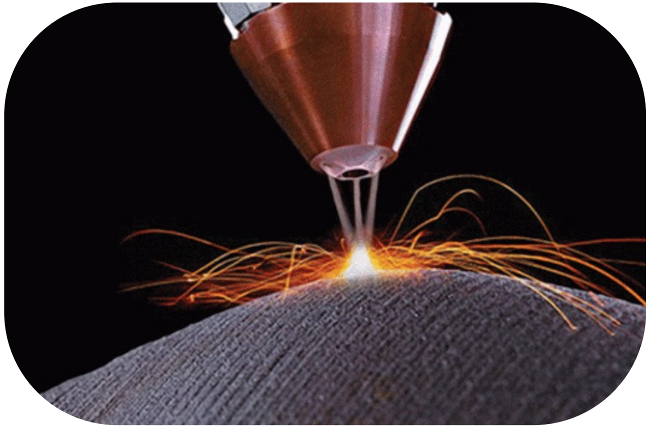
10 entries

Composition and heat treatment space 30 dimensions

Requires 31 points to fit a hyperplane

Just 10 data entries available to model defect density

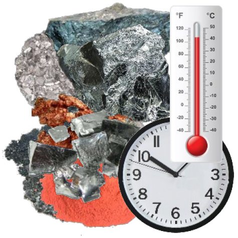
Ability for printing and welding are strongly correlated



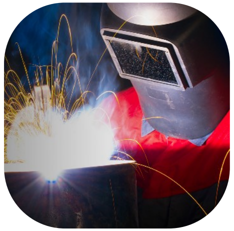
Laser

Electricity

First predict weldability



1000 entries

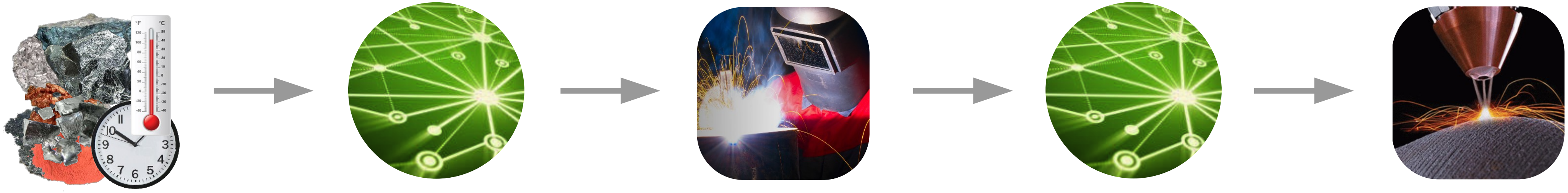


Use 1000 weldability entries to understand complex composition → weldability model

Use weldability to predict defects formed

1000 entries

10 entries



Use 1000 weldability entries to understand complex composition → weldability model

10 defects entries capture the simple weldability → defect relationship

Two interpolations give composition → defects extrapolation

Target properties

Elemental cost	$< 25\ \$\mathrm{kg}^{-1}$
Density	$< 8500\ \mathrm{kgm}^{-3}$
γ’ content	< 25 wt%
Oxidation resistance	$< 0.3\ \mathrm{mgcm}^{-2}$
Defects	< 0.15% defects
Phase stability	> 99.0 wt%
γ’ solvus	> 1000˚C
Thermal resistance	$> 0.04\ \mathrm{K\Omega^{-1}m^{-3}}$
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 and processing variables

Cr 19%

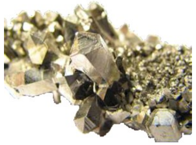
Co 4%

Mo 4.9%

W 1.2%

Zr 0.05%

Nb 3%



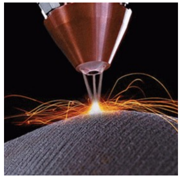
Al 2.9%

C 0.04%

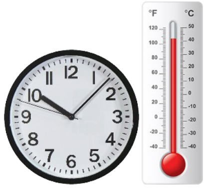
B 0.01%

Ni

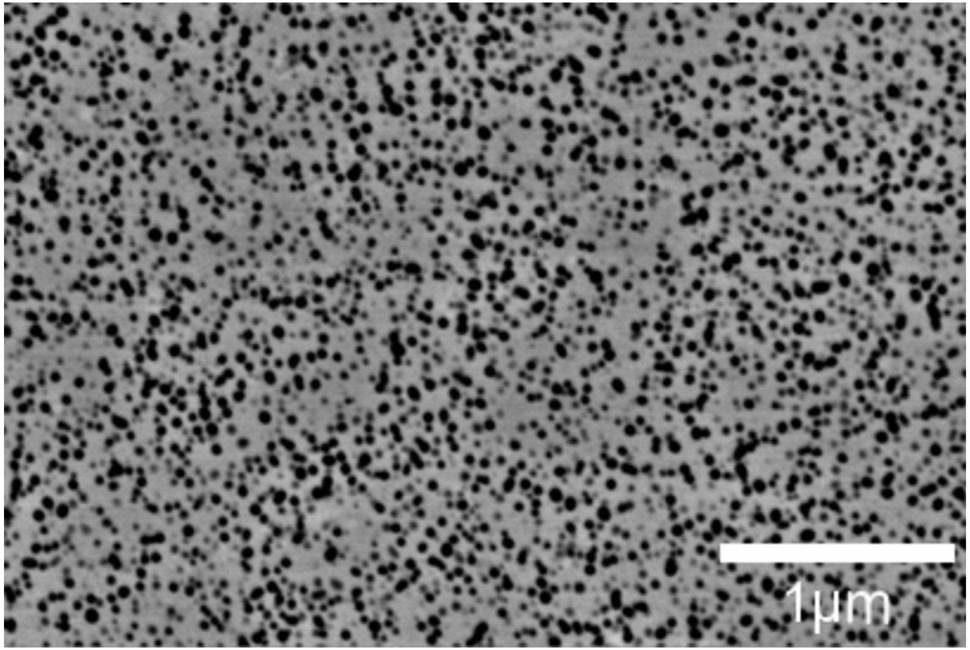
Expose 0.8



T_{HT} 1300ºC

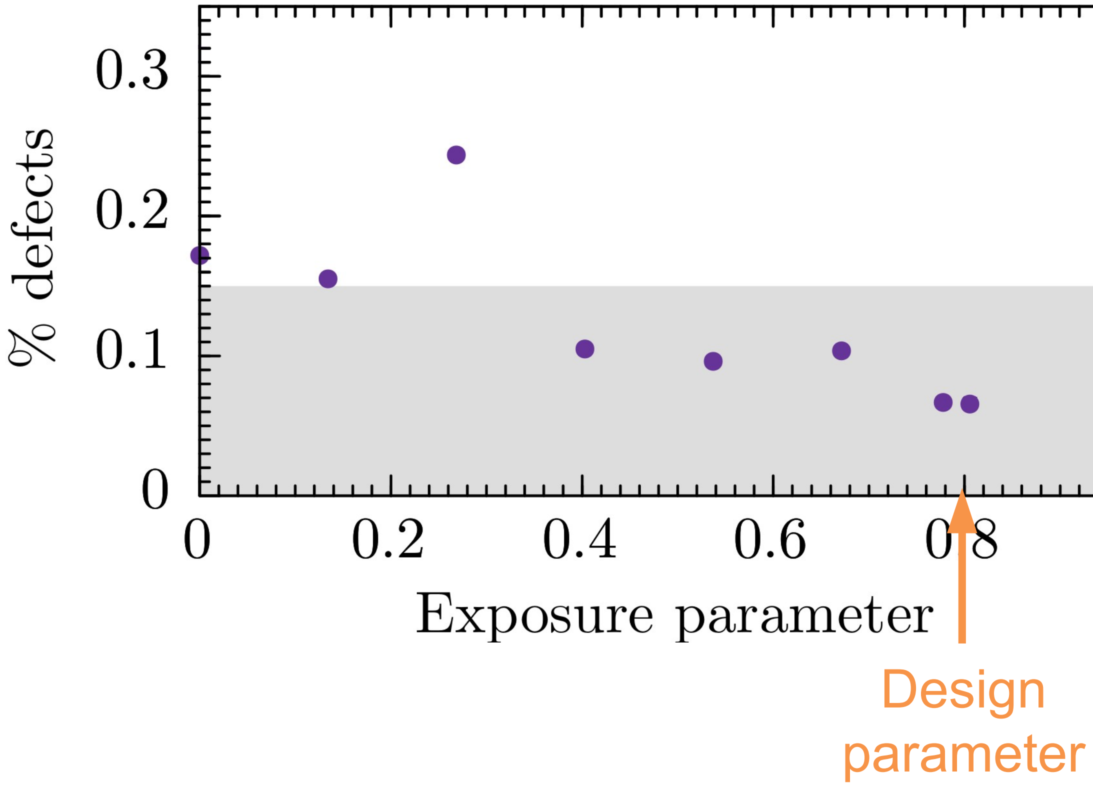


Microstructure



Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design **168**, 107644 (2019)

Testing the defect density



Probabilistic neural network identification of an alloy for direct laser deposition
Materials & Design **168**, 107644 (2019)

Development of methodology

2013

Multiple properties for Rolls Royce engines

Development of methodology

2013

2014

Multiple properties for Rolls Royce engines

Property-property correlations with Rolls Royce and BP

Development of methodology

Concurrent materials design

2013	2014	2015
Multiple properties for Rolls Royce engines	Property-property correlations with Rolls Royce and BP	Royal Society University Research Fellowship

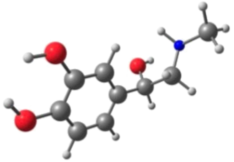
Development of methodology

Concurrent materials design

2013	2014	2015	2016
Multiple properties for Rolls Royce engines	Property-property correlations with Rolls Royce and BP	Royal Society University Research Fellowship	Experiment-simulation correlations with Samsung

Development of methodology

Concurrent materials design



2013	2014	2015	2016	2017
Multiple properties for Rolls Royce engines	Property-property correlations with Rolls Royce and BP	Royal Society University Research Fellowship	Experiment-simulation correlations with Samsung	Drug discovery study with etherapeutics

Development of methodology

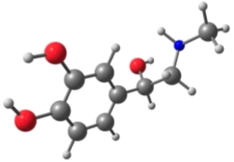
THE ROYAL SOCIETY

Concurrent materials design

UNIVERSITY OF CAMBRIDGE

SAMSUNG

therapeutics



2013	2014	2015	2016	2017	2018
Multiple properties for Rolls Royce engines	Property-property correlations with Rolls Royce and BP	Royal Society University Research Fellowship	Experiment-simulation correlations with Samsung	Drug discovery study with etherapeutics	Founding of Intellegens

Exploit uncertainty to design a drug with Optibrium

Dr Tom Whitehead

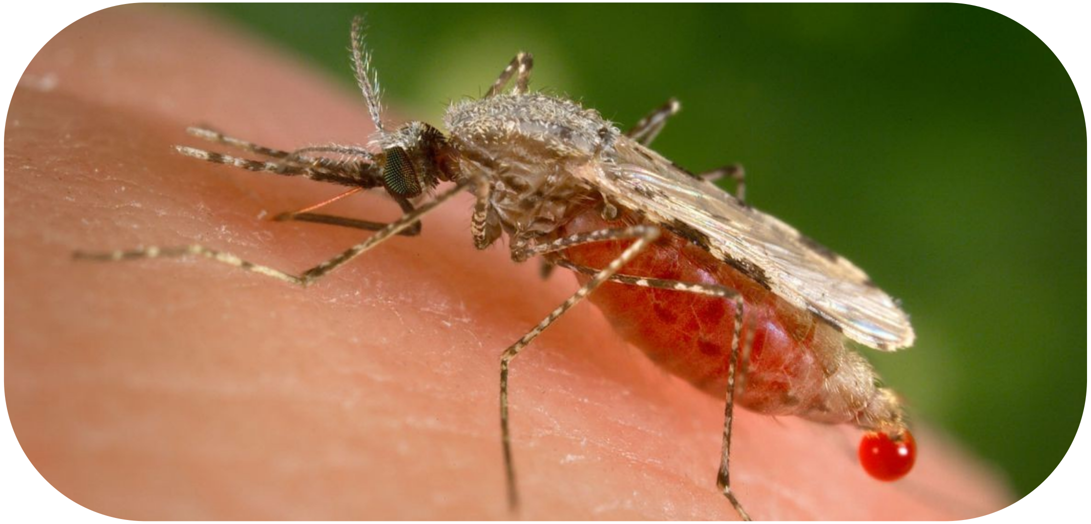
Dr Ben Irwin

Dr Matt Segall

Dr Gareth Conduit

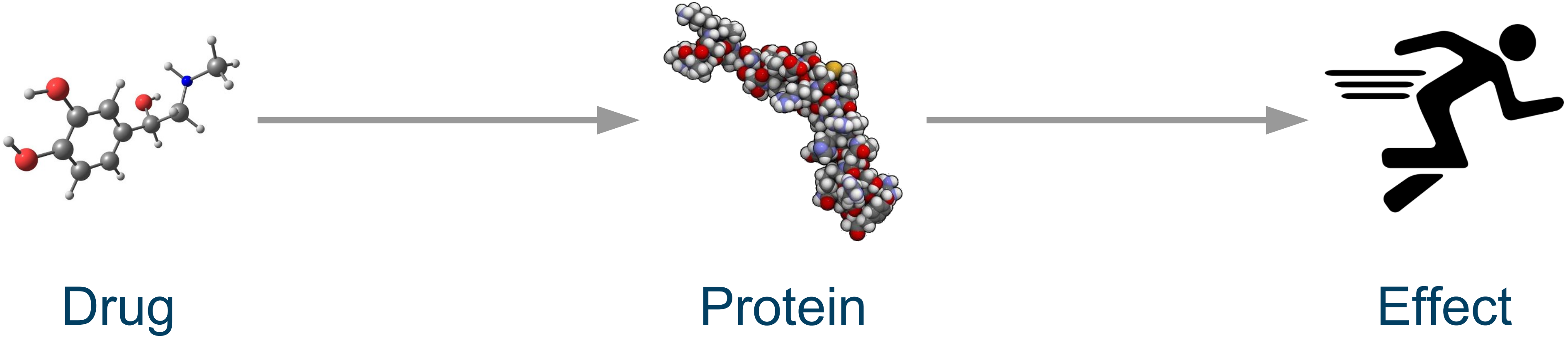
An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials
Journal of Medicinal Chemistry 64, 16450 (2021)
Imputation of Assay Bioactivity Data using Deep Learning
Journal of Chemical Information and Modeling, 59, 1197 (2019)

Open Source Malaria contest

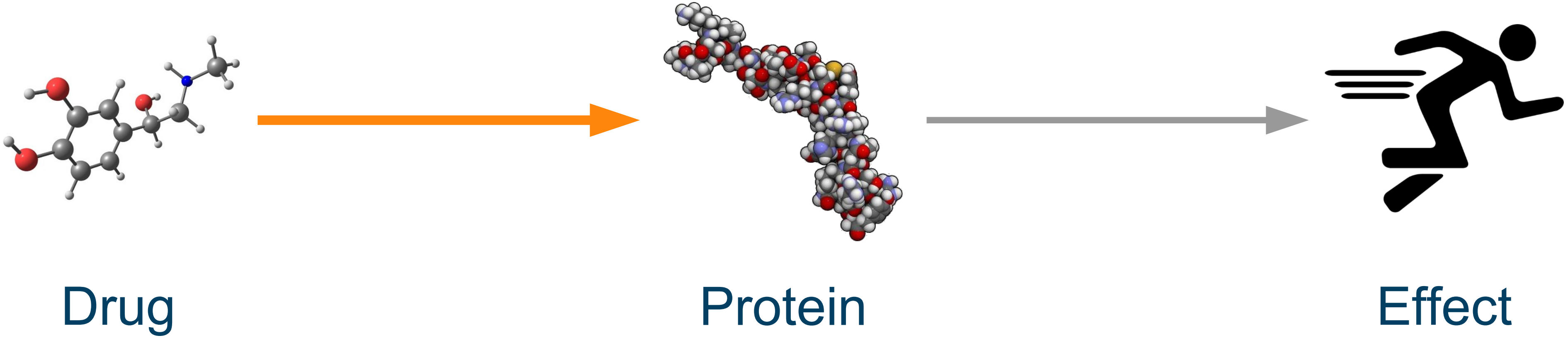


OPEN SOURCE MALARIA
Looking for New Medicines

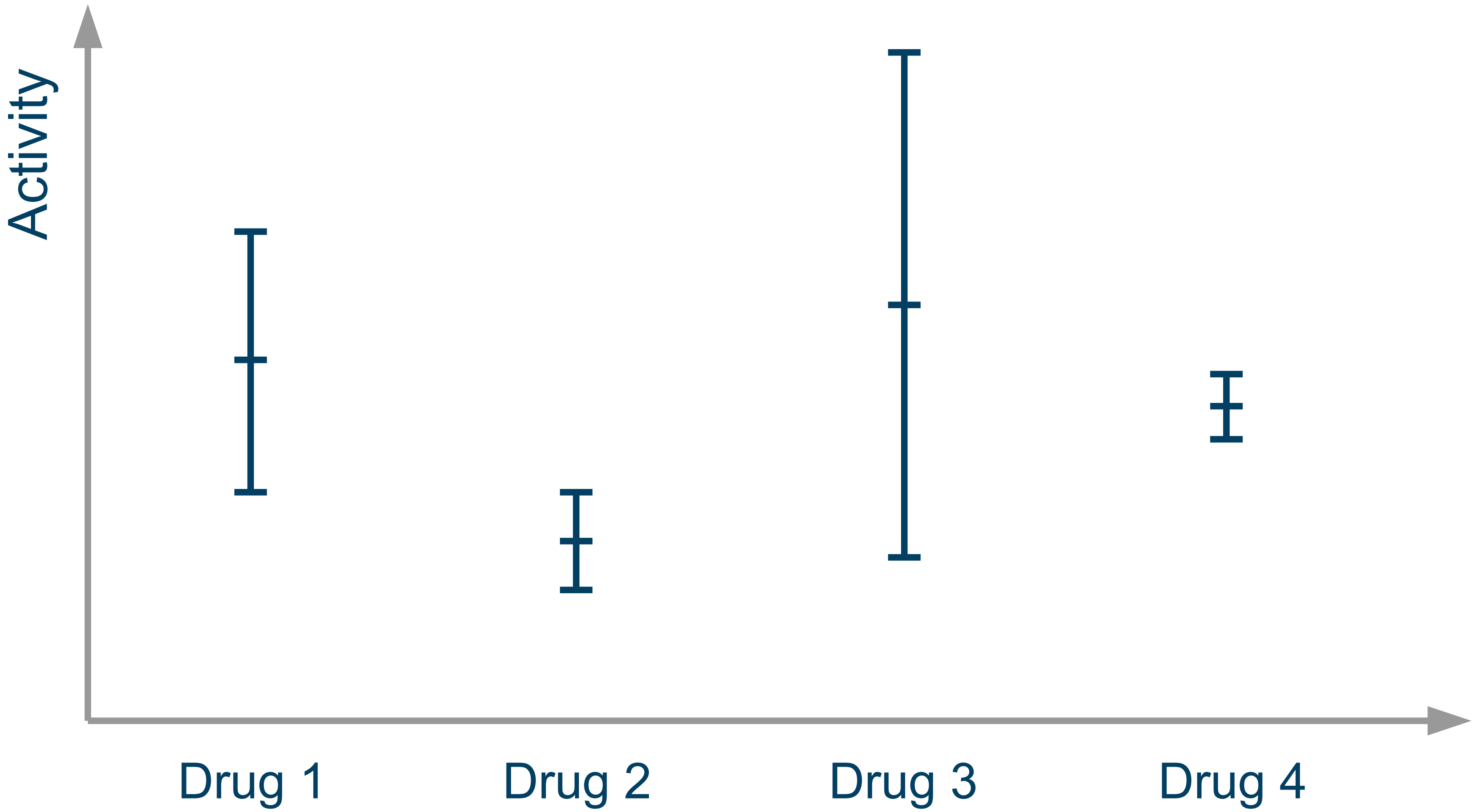
Action of a drug



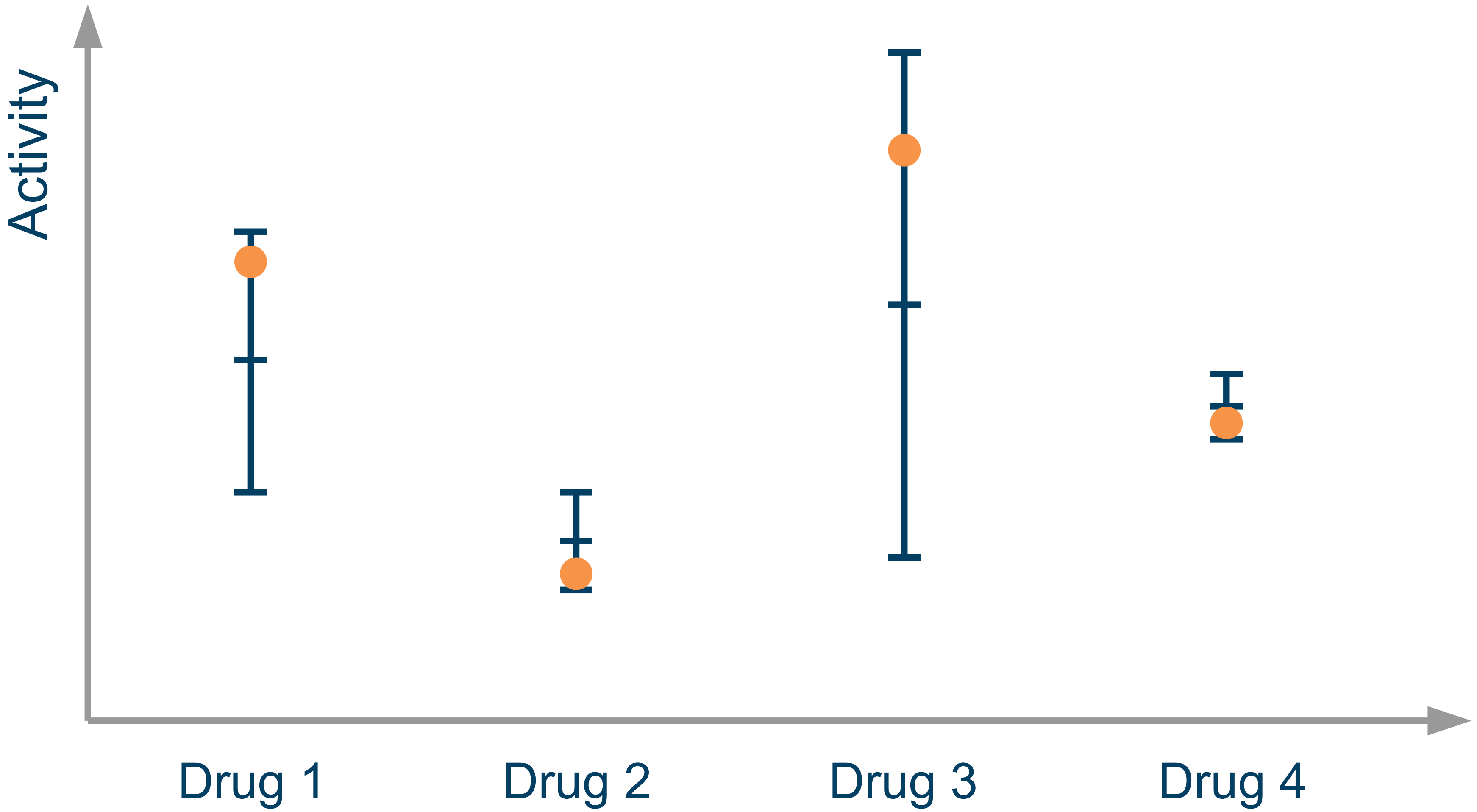
Action of a drug



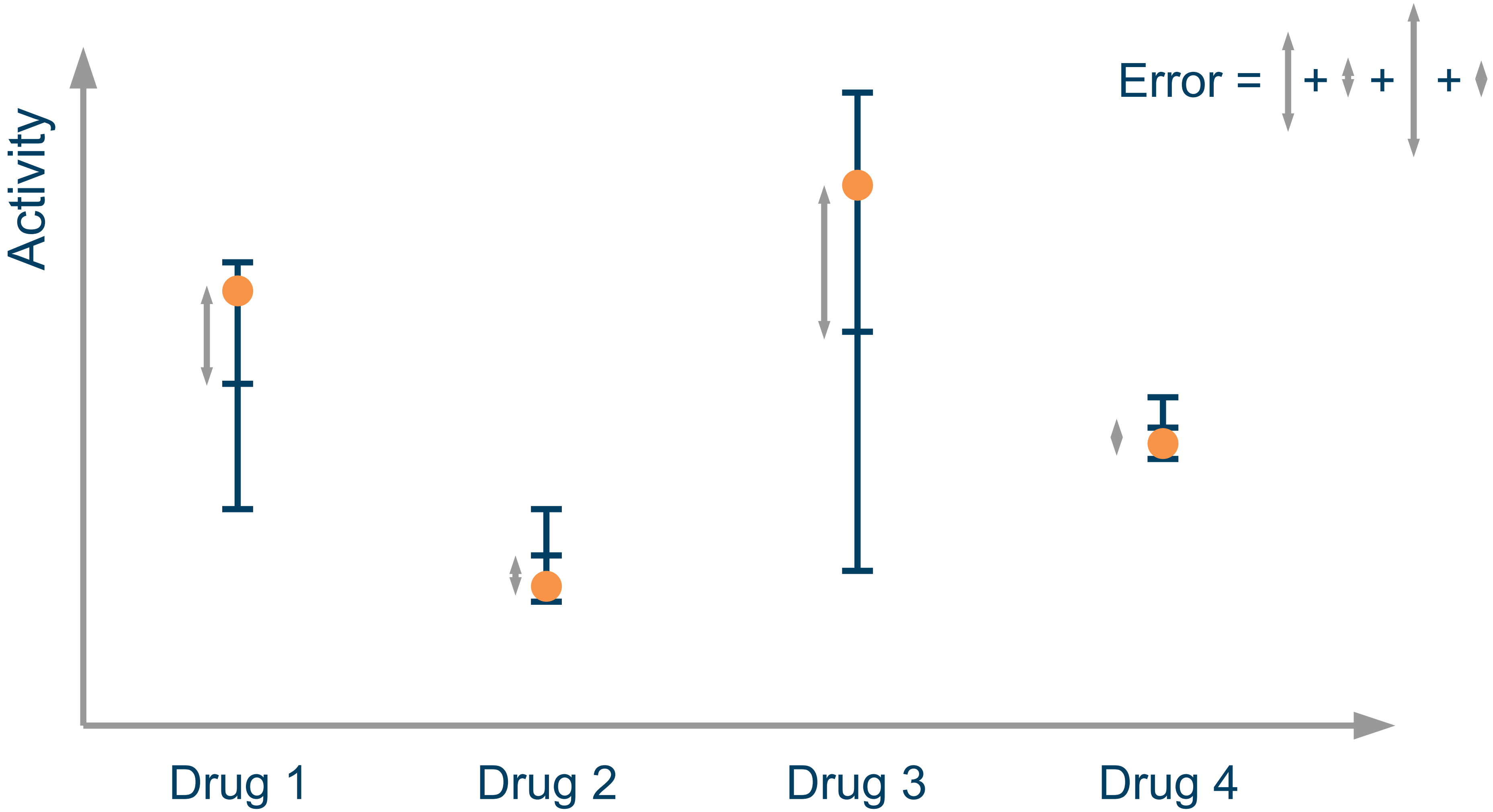
Predictions have an uncertainty



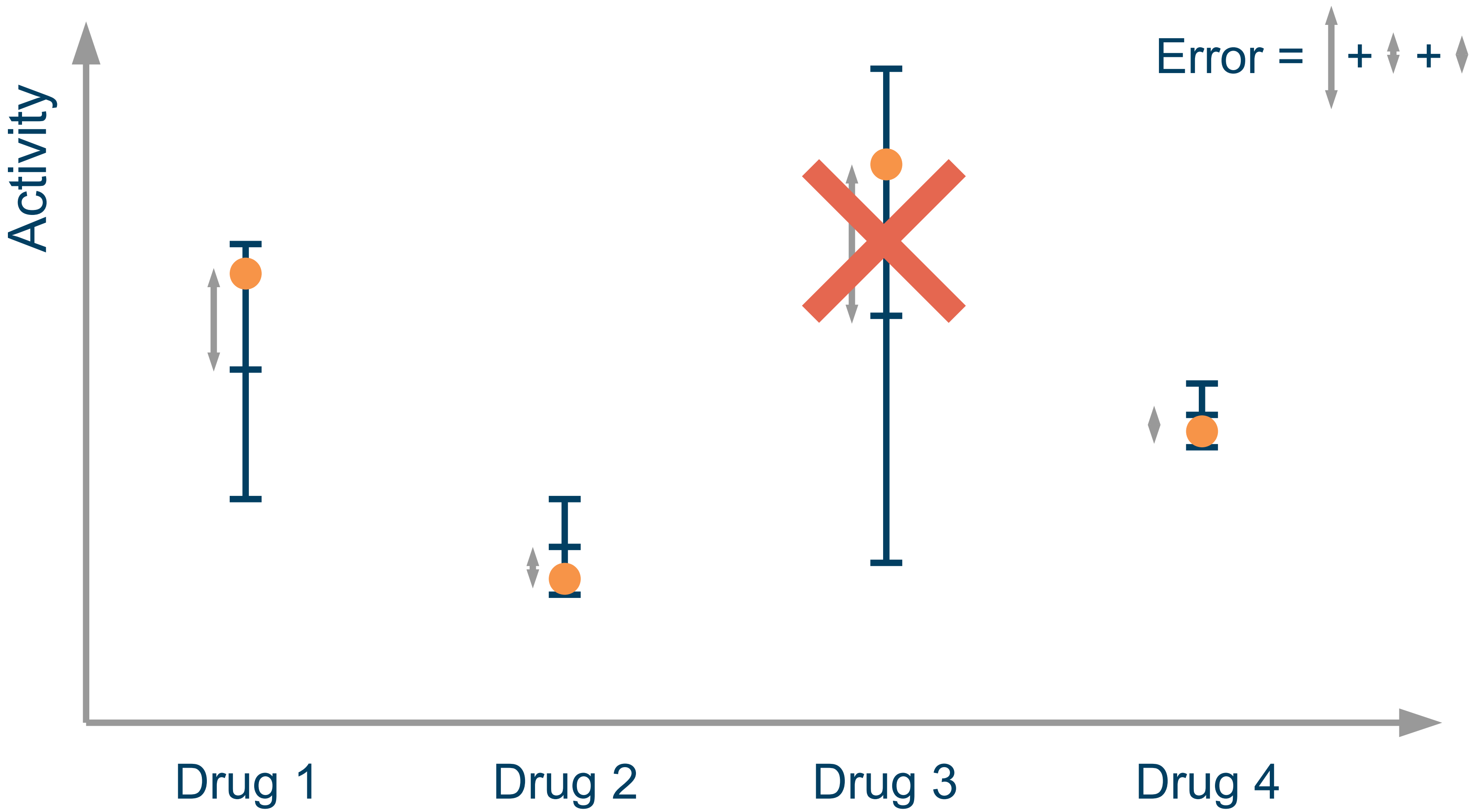
Validation data typically within one standard deviation



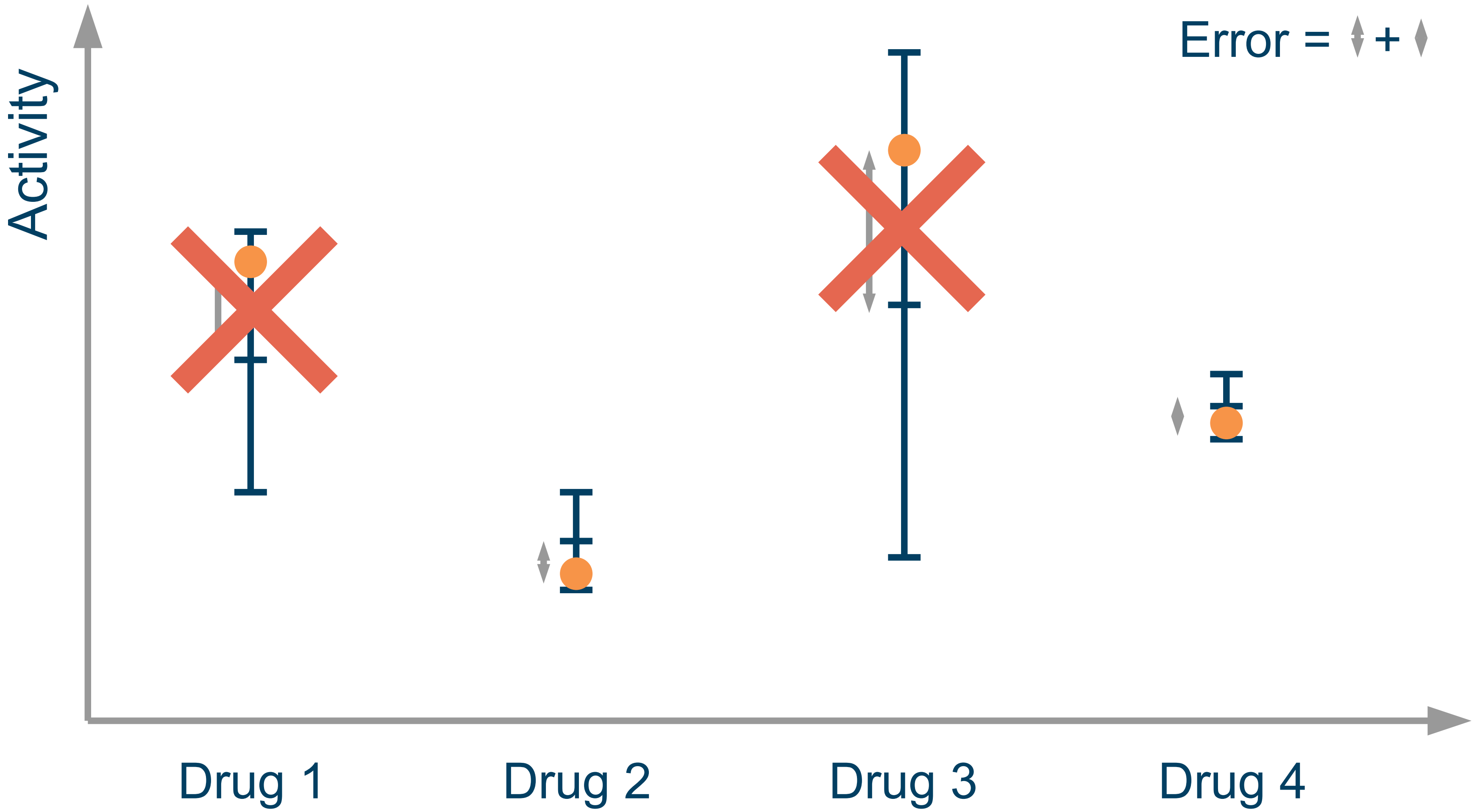
Accuracy R^2 metric calculated with difference from mean



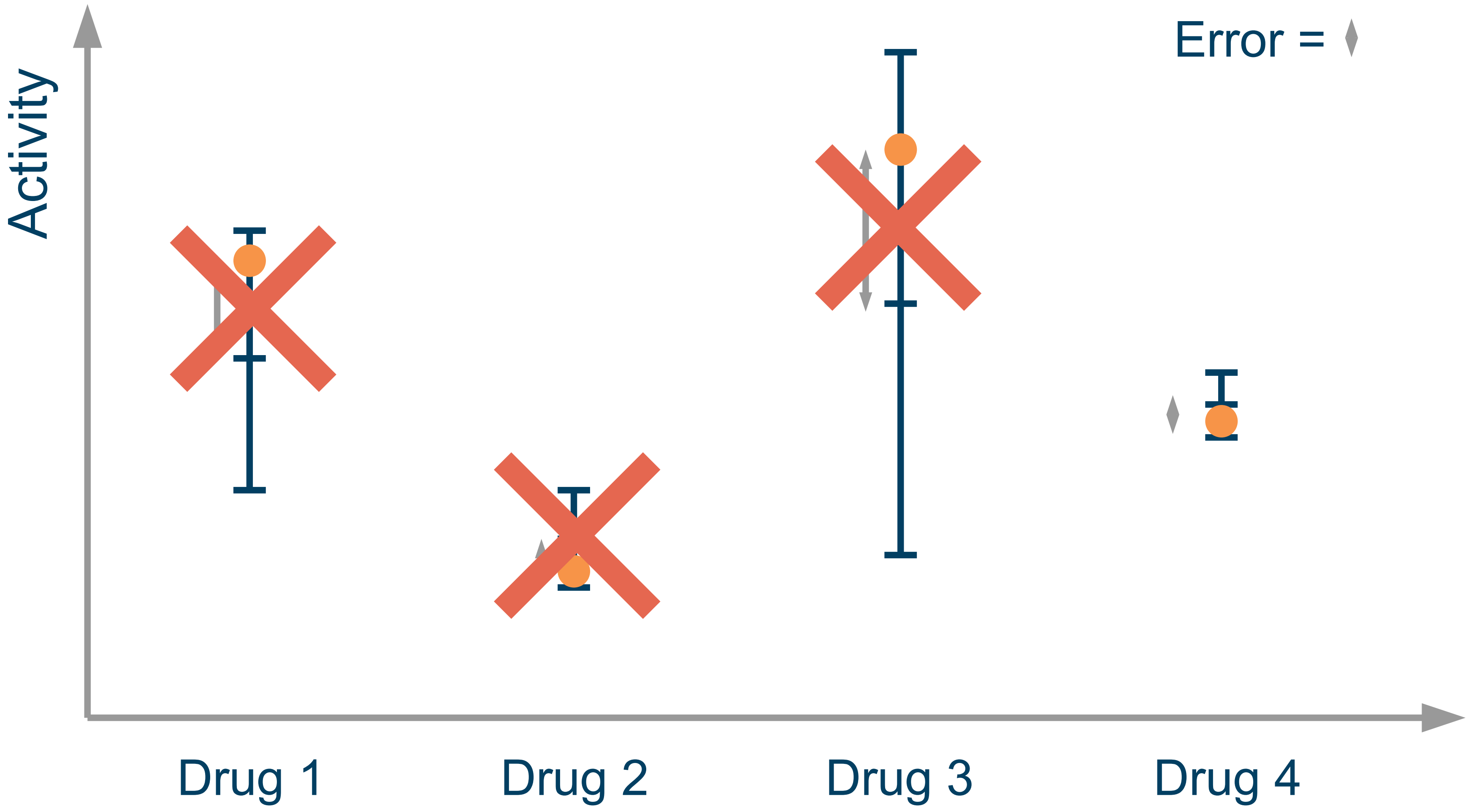
Impute 75% of data with smallest uncertainty



Impute 50% of data with smallest uncertainty



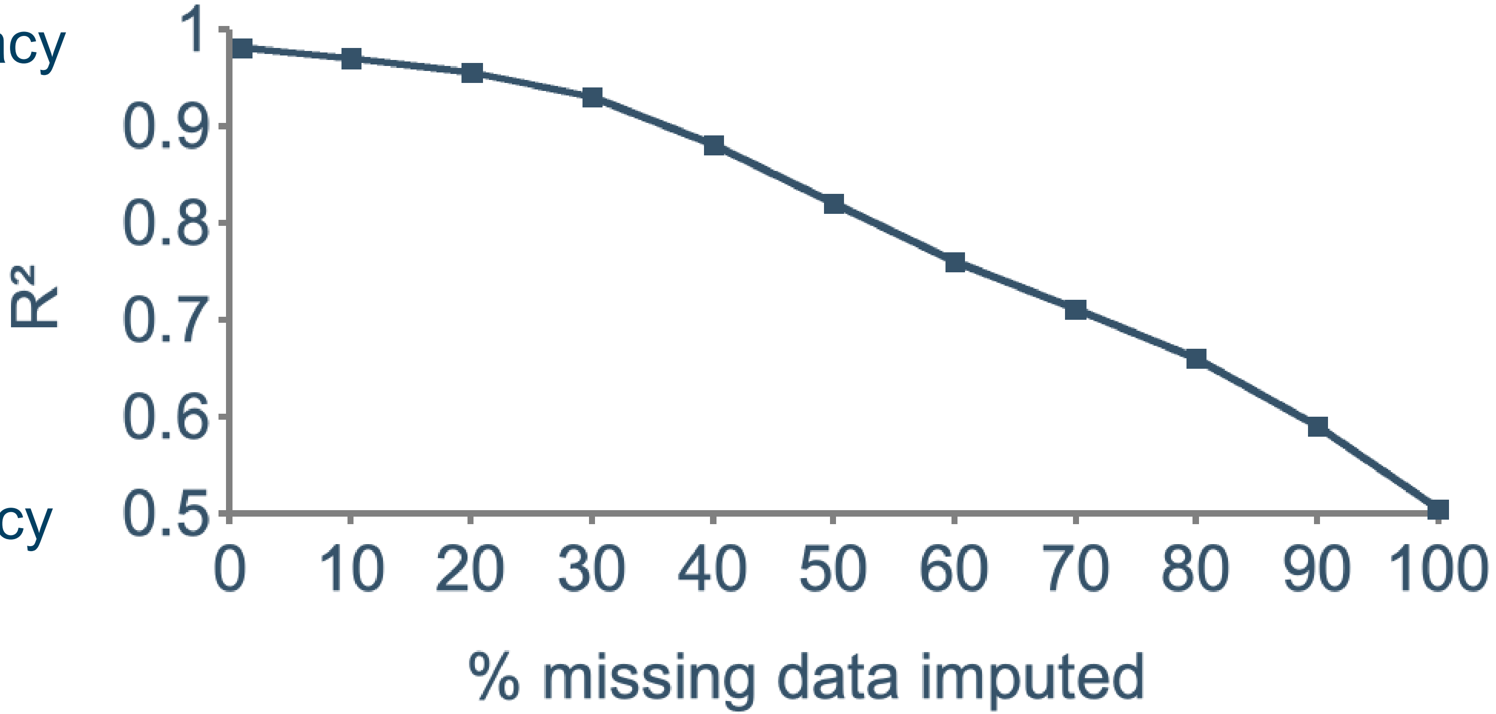
Impute 25% of data with smallest uncertainty



Improved performance by exploiting uncertainty

High accuracy

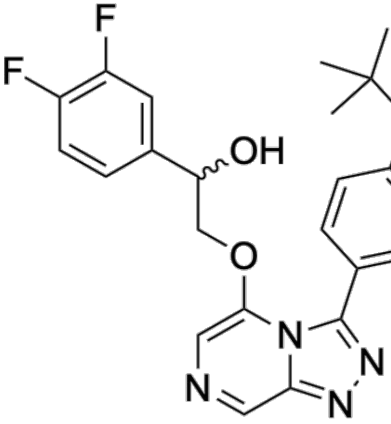
Low accuracy



Different drugs can treat the same ailment

Focus on compounds with low uncertainty

Open Source Malaria experimental validation

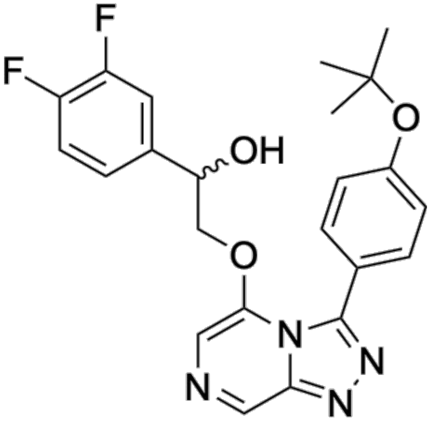


Optibrium & Intellegens

0.647 µM

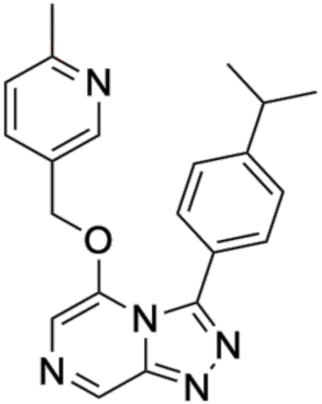
Journal of Medicinal Chemistry **64**, 16450 (2021)

Open Source Malaria other compounds



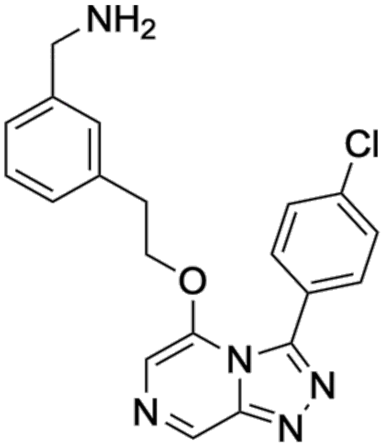
Optibrium & Intellegens

0.647 μM



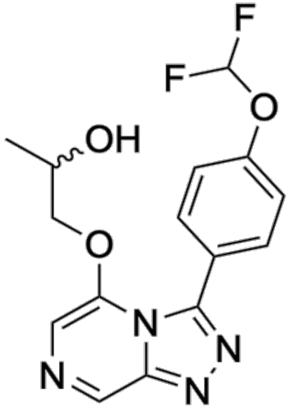
Davy Guan

>25 μM



Exscientia

10.9 μM



Molomics

>25 μM

Journal of Medicinal Chemistry **64**, 16450 (2021)

Commercialization

e therapeutics

bp

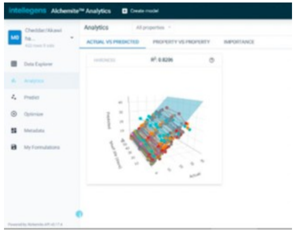
2018

Transfer contracts from University

Commercialization

2018	2019
Transfer contracts from University	Consultancy work

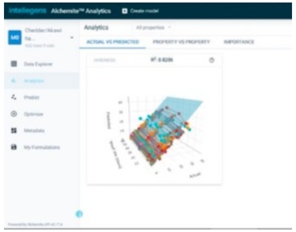
Commercialization



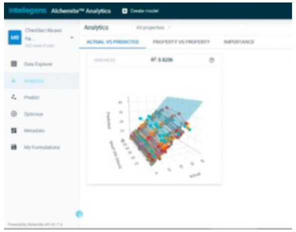
2018	2019	2020
Transfer contracts from University	Consultancy work	Launch Alchemite Analytics™ product

Commercialization

2018	2019	2020	2021
Transfer contracts from University	Consultancy work	Launch Alchemite Analytics™ product	Launch Cerella™ product with Optibrium

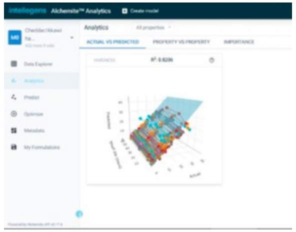


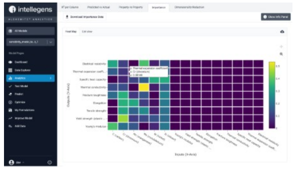
Commercialization



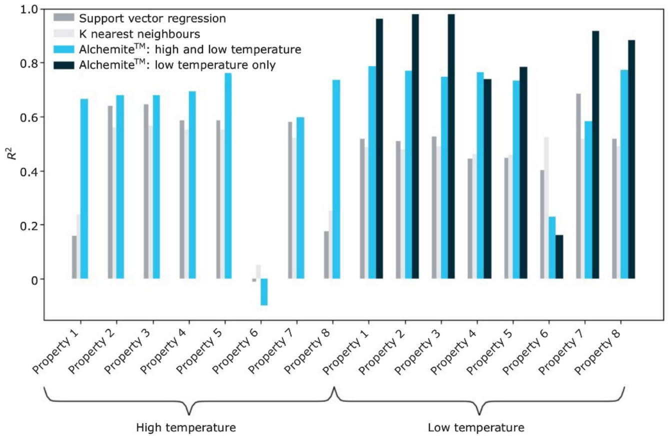
2018	2019	2020	2021	2022
Transfer contracts from University	Consultancy work	Launch Alchemite Analytics™ product	Launch Cerella™ product with Optibrium	Launch product with ANSYS Granta

Commercialization

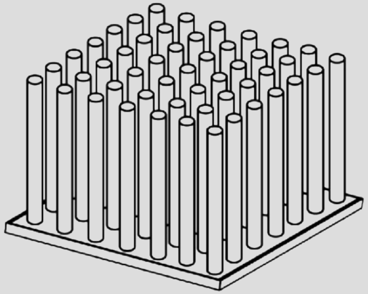


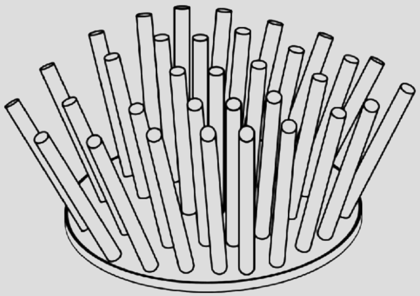


2018	2019	2020	2021	2022	2023
Transfer contracts from University	Consultancy work	Launch Alchemite Analytics™ product	Launch Cerella™ product with Optibrium	Launch product with ANSYS Granta	Enterprise licenses & healthcare market

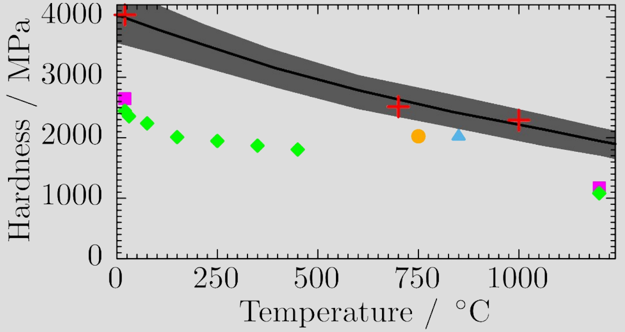


Johnson Matthey Technology Review
66, 130 (2022)





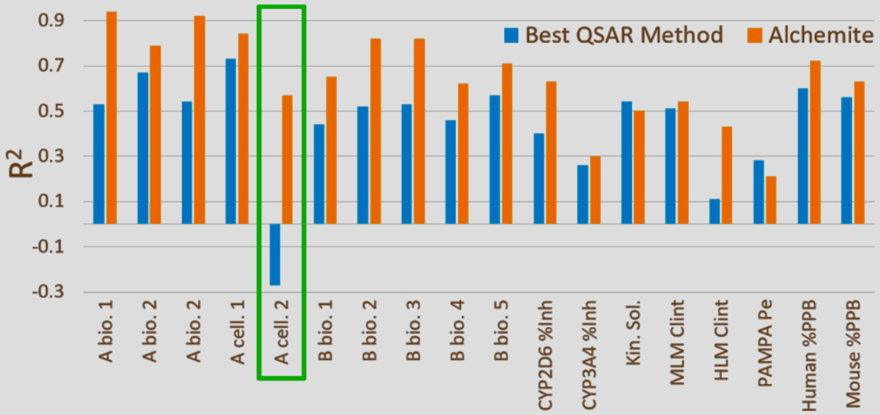
NASA Technical Memorandum
20220008637



Materials & Design **131**, 358 (2017)
Scripta Materialia **146**, 82 (2018)
Data Centric Engineering **3**, e30 (2022)

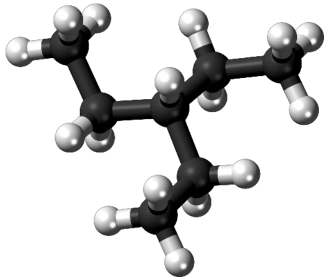
Alloy	Source	ANN	Δ_σ	Actual
Steel AISI 301L	193	269	5	238[23]
Steel AISI 301	193	267	5	221[23]
Al 1080 H18	51	124	5	120[23]
Al 5083 wrought	117	191	14	300,190[4, 23]
Al 5086 wrought	110	172	11	269,131[4, 23]
Al 5454 wrought	102	149	14	124[23]
Al 5456 wrought	130	201	11	165[23]
INCONEL600	223	278	10	≥550[23]

Computational Materials
Science **147**, 176 (2018)

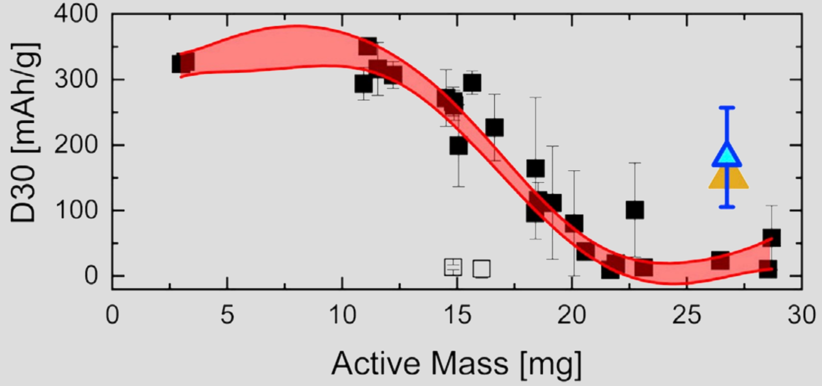


J. of Chem. Info. & Model. **60**, 2848 (2020)
Applied AI Letters **2**, e31 (2021)
Molecular Pharmaceutics **19**, 1488 (2022)

Journal of Computer-Aided Molecular Design **35**, 112501140 (2021)



Fluid Phase Equilibria **501**, 112259 (2019)
Journal of Chemical Physics **153**, 014102 (2020)



Nature Machine Intelligence **2**, 161 (2020)
Cell Reports Physical Science **2**, 100683 (2021)

Model sparse datasets by exploiting property-property relationships

Merge data, computer simulations, and physical laws

Exploit uncertainty to focus on most robust designs

Reduce costly experiments to accelerate discovery

Commercialized as Alchemite™ by Intellegens

Challenge: *What other physical information could be extracted from a dataset?*