intellegens.co.uk
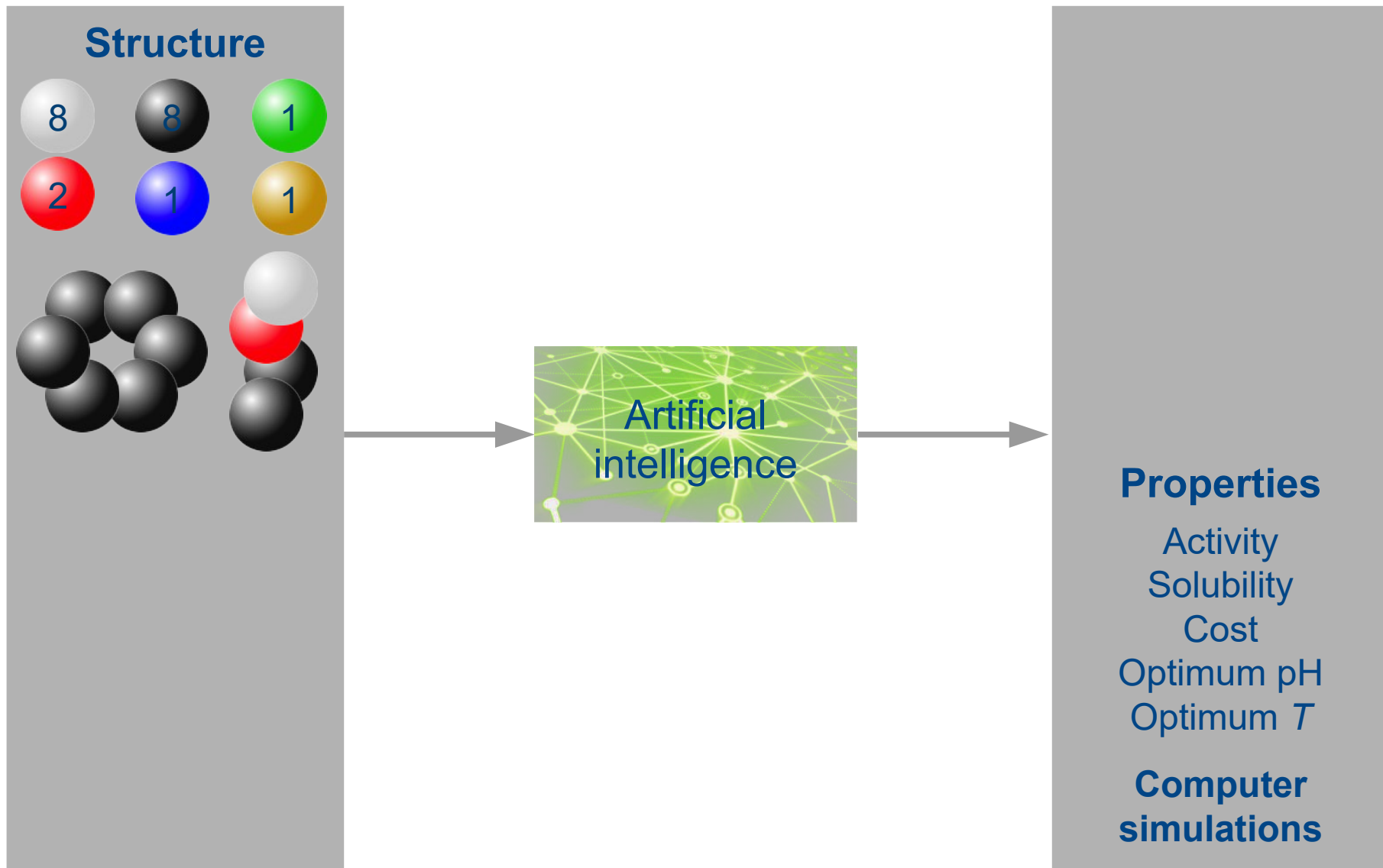
Neural networks for drug discovery

Gareth Conduit

Neural networks for drug design

Structure

8   8   1
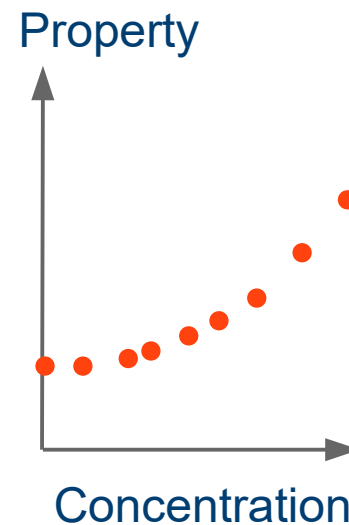
2   1   1

Artificial intelligence

Properties

Activity
Solubility
Cost
Optimum pH
Optimum T

Computer simulations

Neural networks for drug design
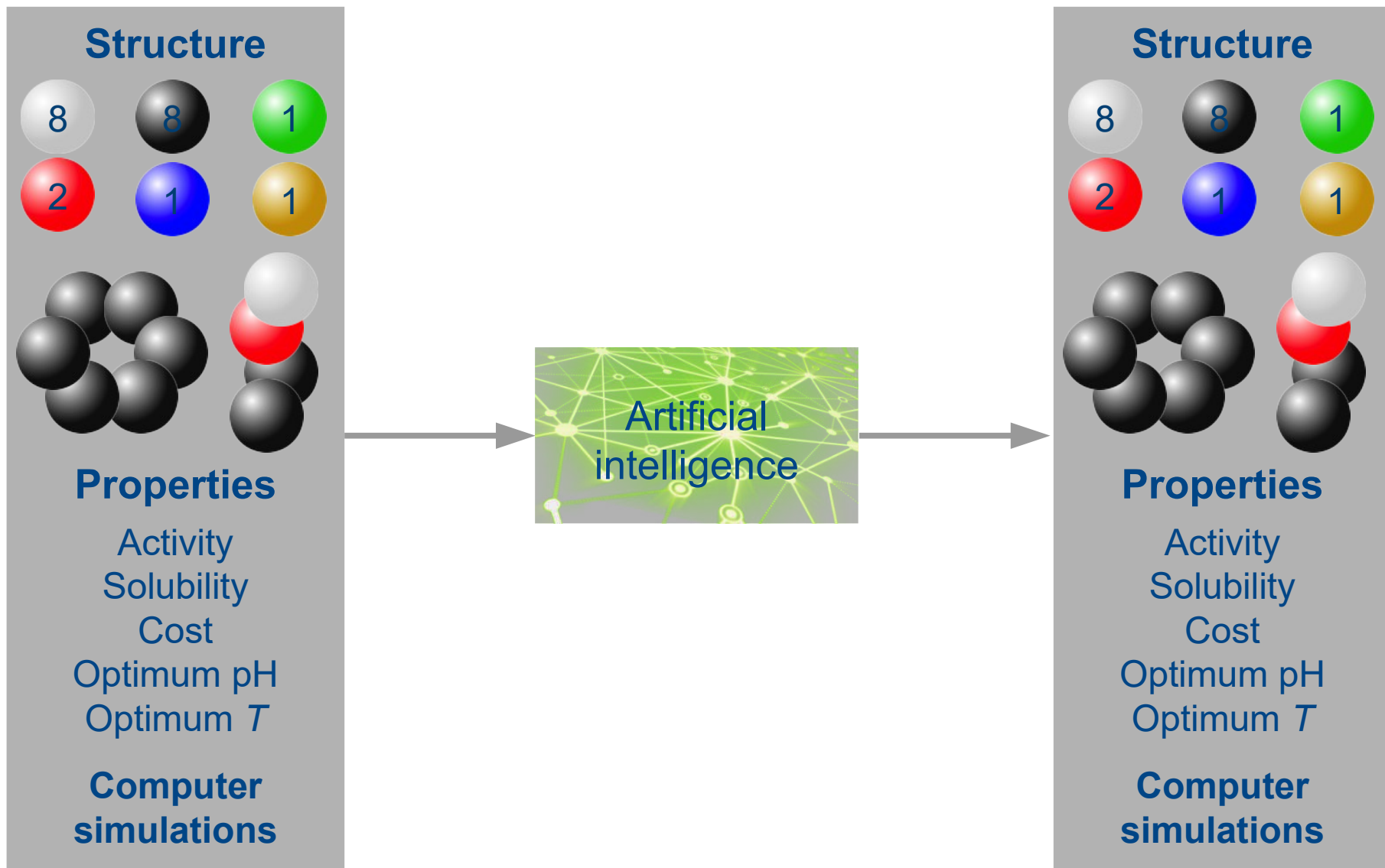
Neural networks for drug design: correlated data
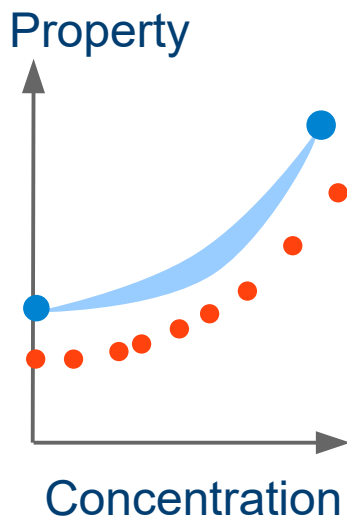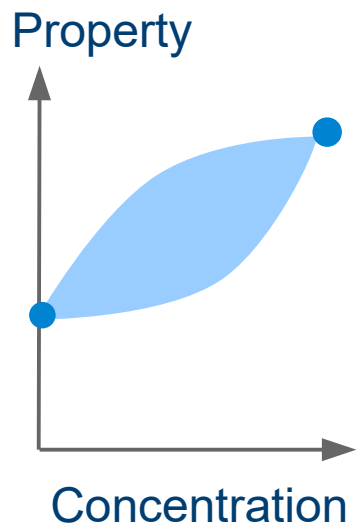
Two-dimensional fragmented data

x	y
✓	✓
✓	✓
✓	✗
✓	✗
✗	✓

Three-dimensional fragmented data

x	y	z
✓	✓	✓
✓	✓	✗
✓	✗	✓
✓	✗	✗
✗	✓	✗

Three requirements for the neural network tool

x	y	z
✓	✓	✓
✓	✓	✗
✓	✗	✓
✓	✗	✗
✗	✓	✗

Correlated data
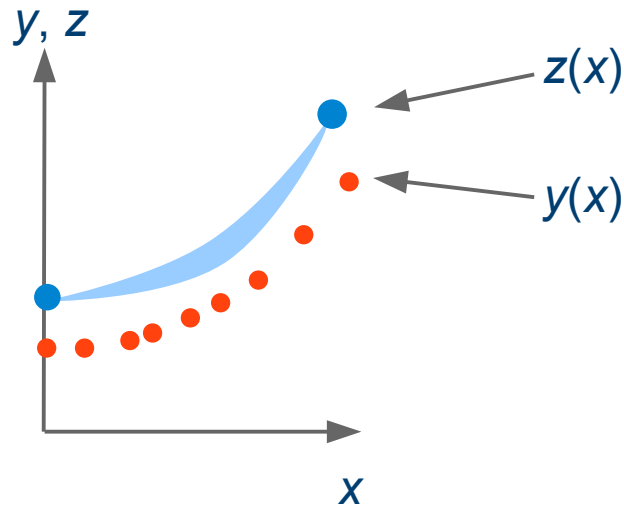
Uncorrelated data

Uncertainties
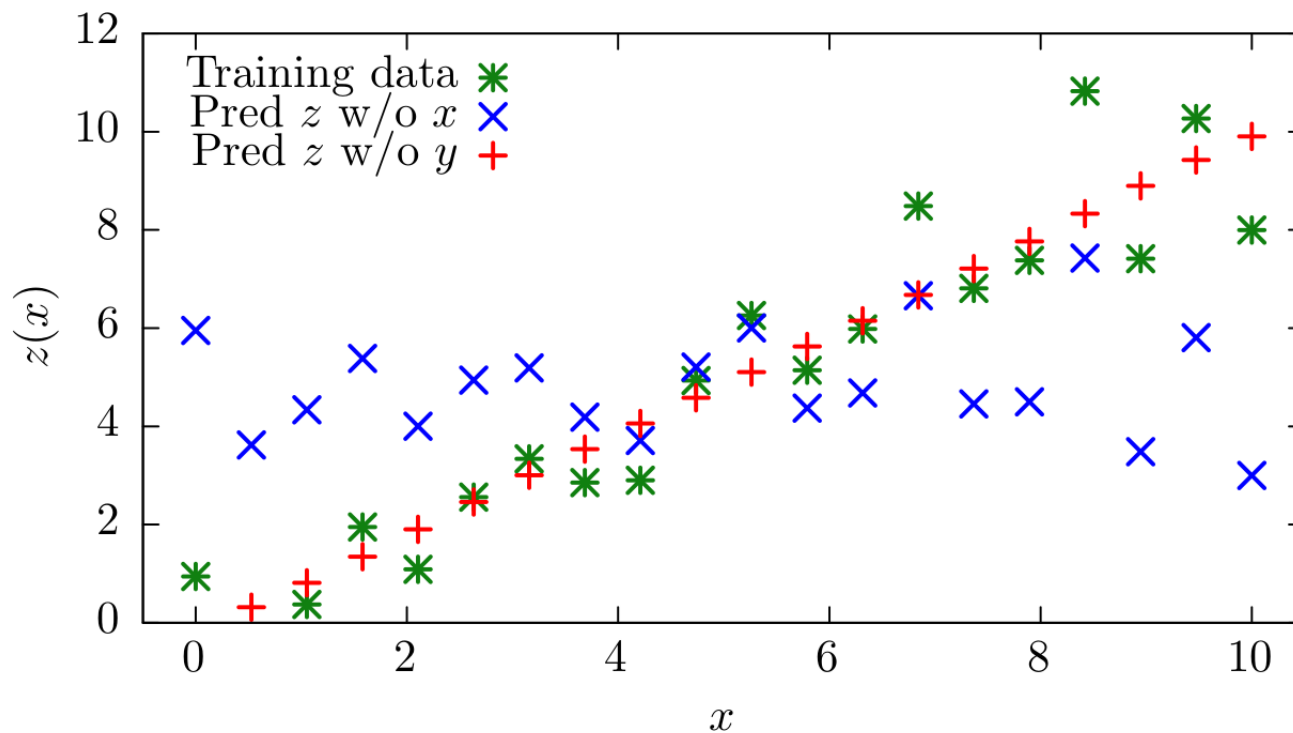
Correlated fragmented data

$z(x) = y(x) + c$ with $y = x^2$

if we know one of x, y, or z we can recover both missing quantities

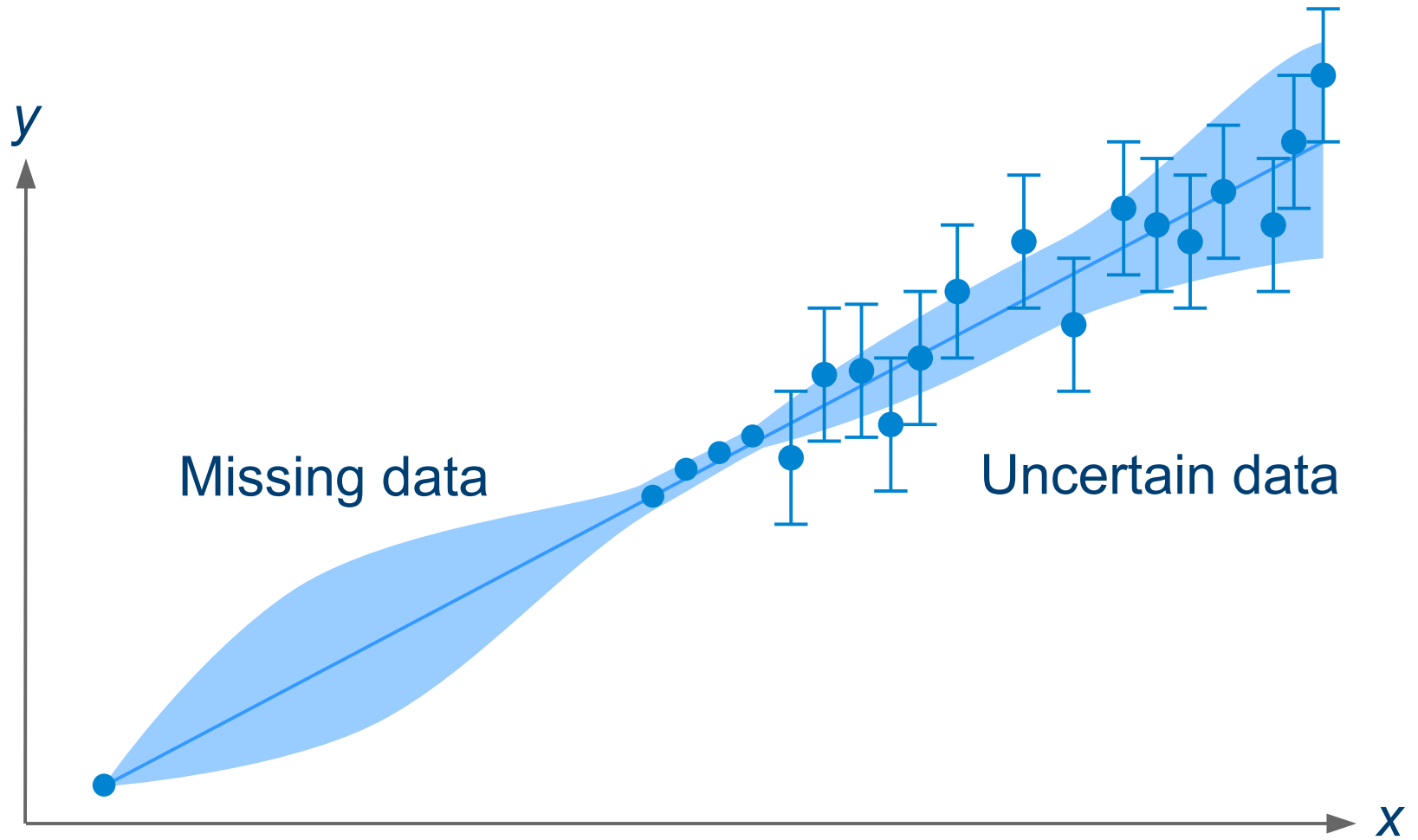# Uncorrelated fragmented data

$z(x) = y(x) + x$ with $y(x) = \mathrm{rand}(x)$

need to know both x and y to recover z

Uncertainty

Industrial applications of neural network tool

DFT and experimental

DFT and experimental
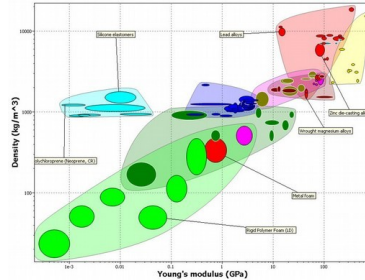

Quantum mechanics
and experimental

Industrial applications of neural network tool

Experimental

Experimental

Structural and experimental

Proposed alloy

Cr:15.8 Co: 20.0 Mo: 0.5 W: 0.5 Ta: 4.9 Nb: 1.1 Al: 2.4

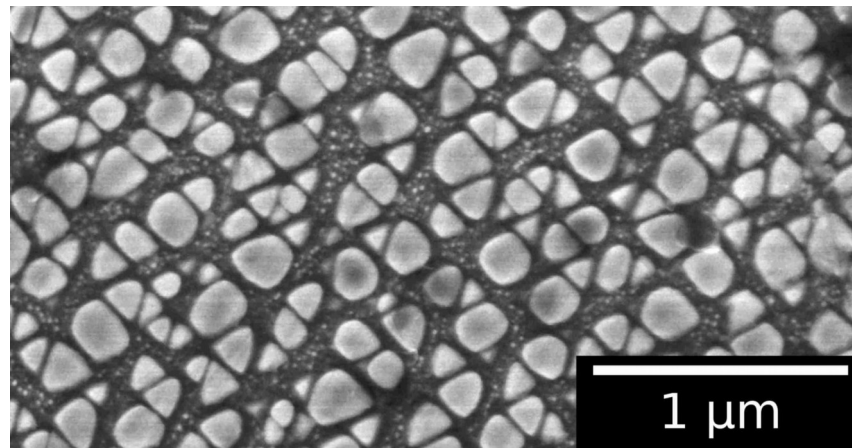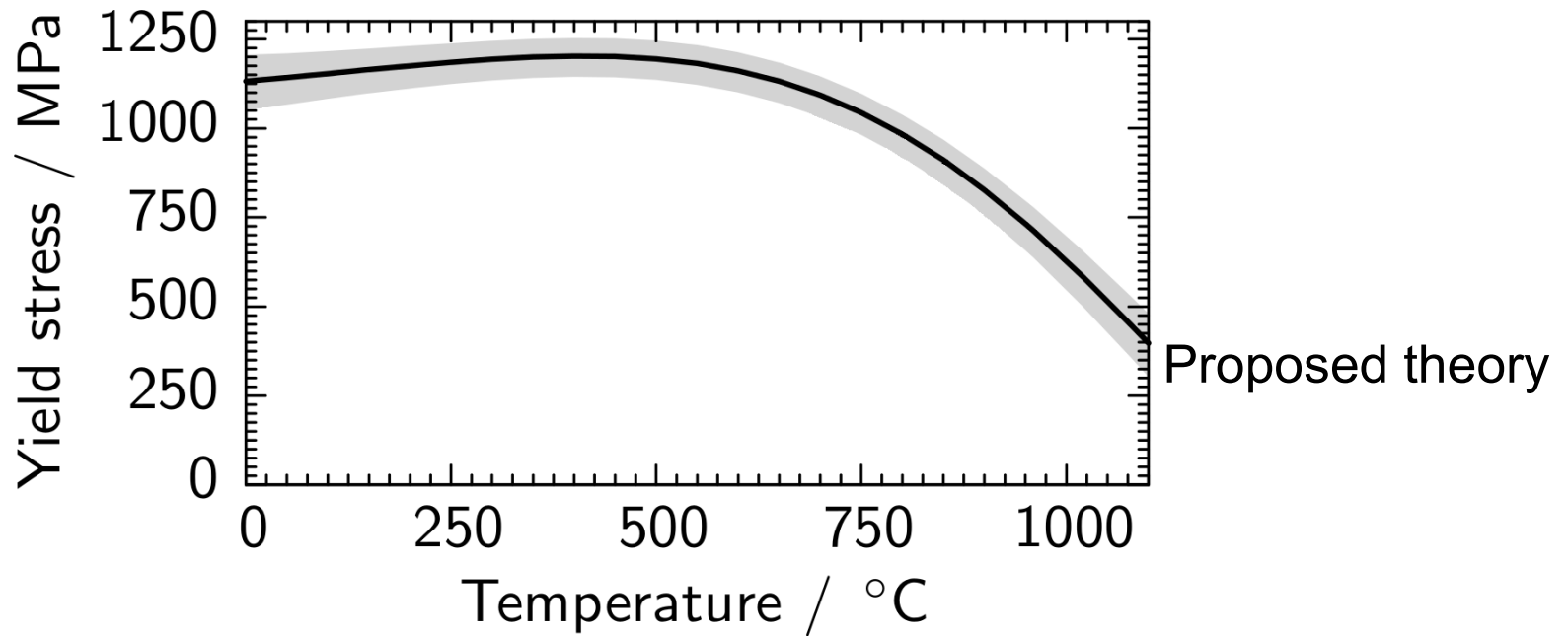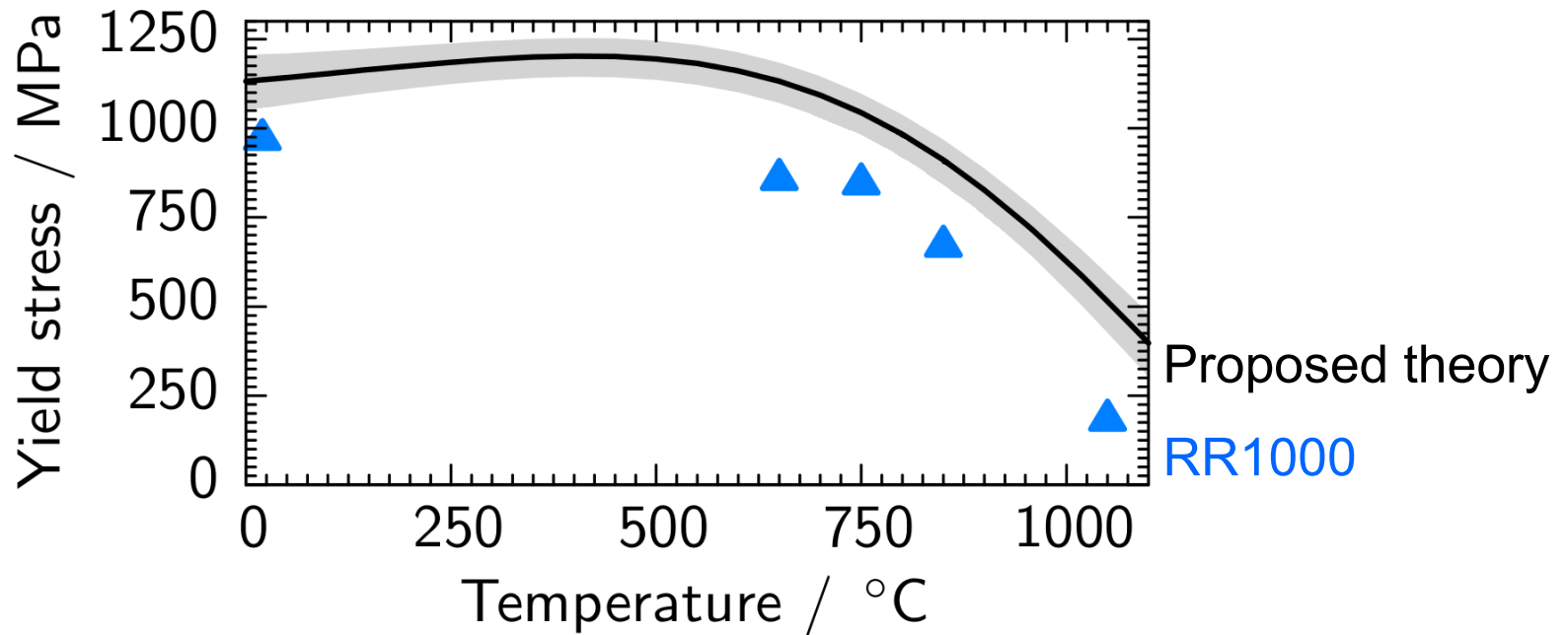Ti: 3.0 Fe: 3.9 Mn: 0.2 Si: 0.2 C: 0.02 B: 0.06 Zr: 0.18
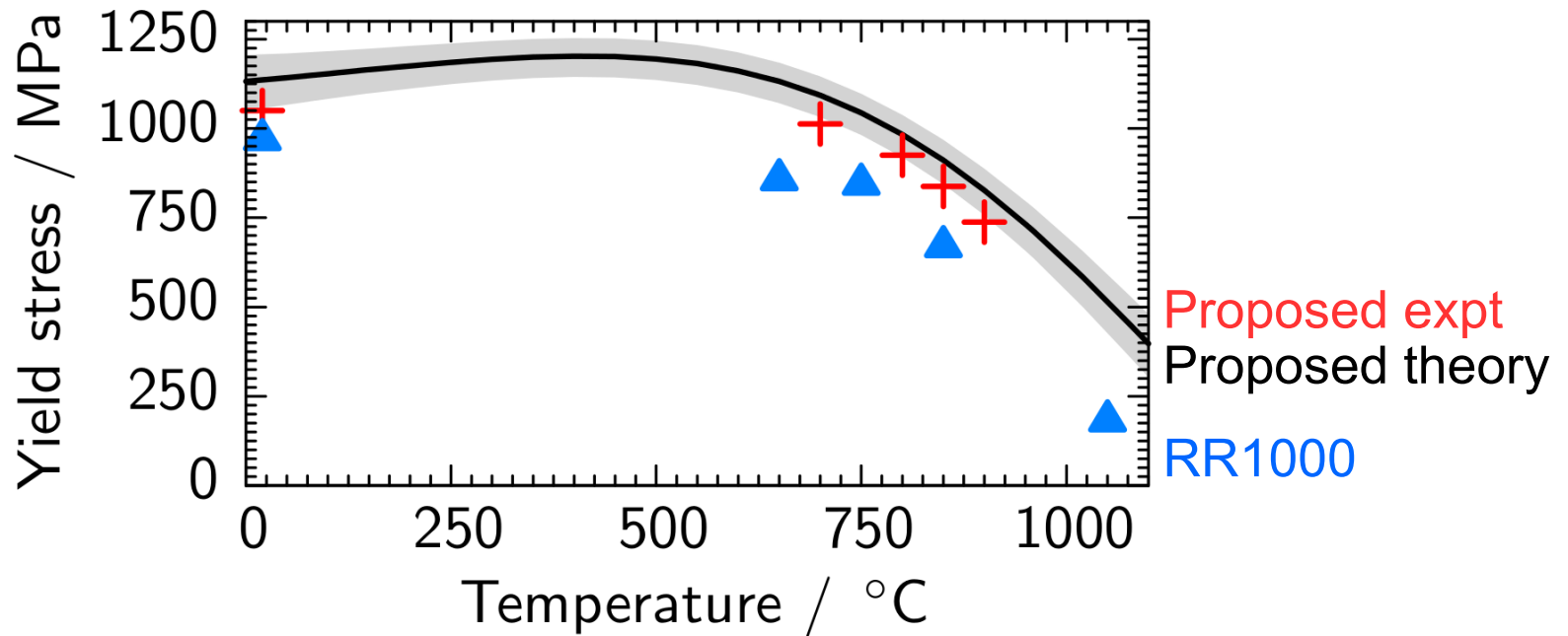
Ni: 47.2 900ºC 30 hours

Microstructure

1 μm

Testing the yield stress

Testing the yield stress

Testing the yield stress
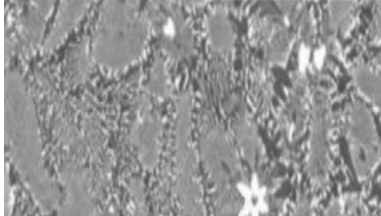
Alloys discovered

Cr-Cr$_2$Ta alloys
Intermetallics, 48, 62

Combustor alloy
GB1408536
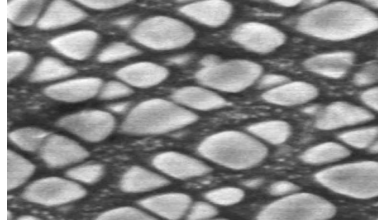
RR1000 grain growth
Acta Materialia, 61, 3378
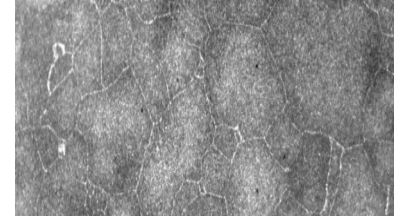
Discovery algorithm
EP14153898
US 2014/177578
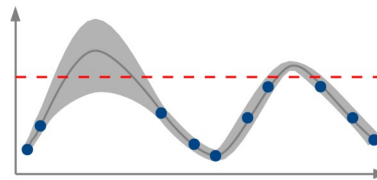
Ni disc alloy
EP14157622
US 2013/0052077 A2
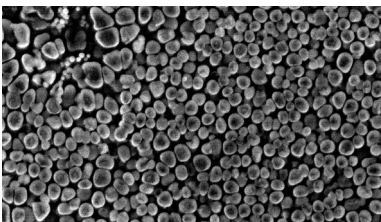
Mo-Hf forging alloy
EP14161255
US 2014/223465
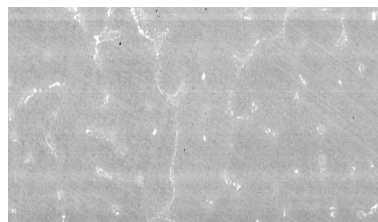
Mo-Nb forging alloy
EP14161529
US 2014/224885
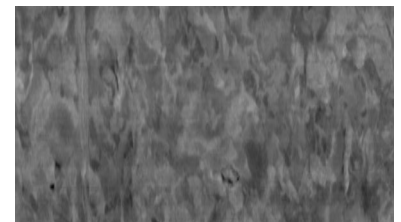
Database integrity
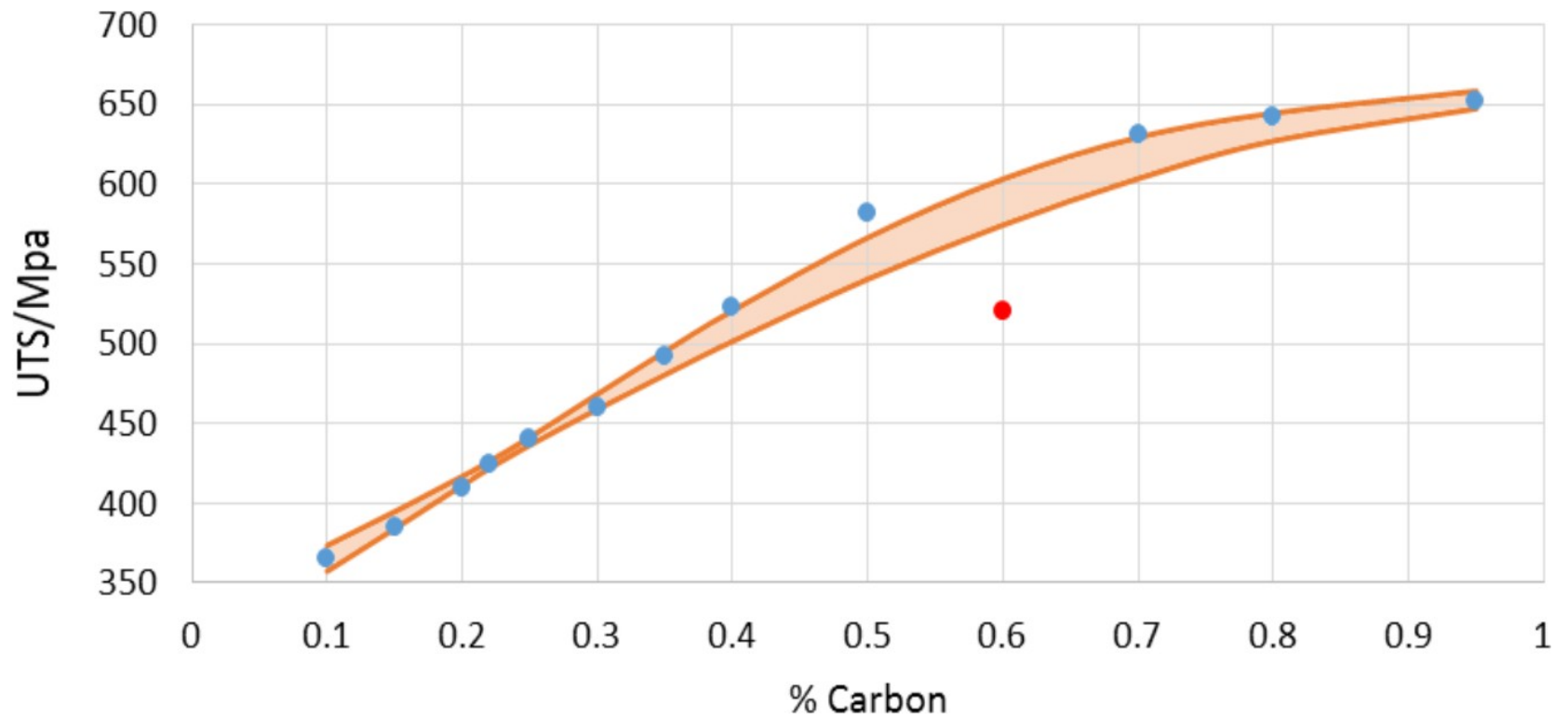
Database integrity
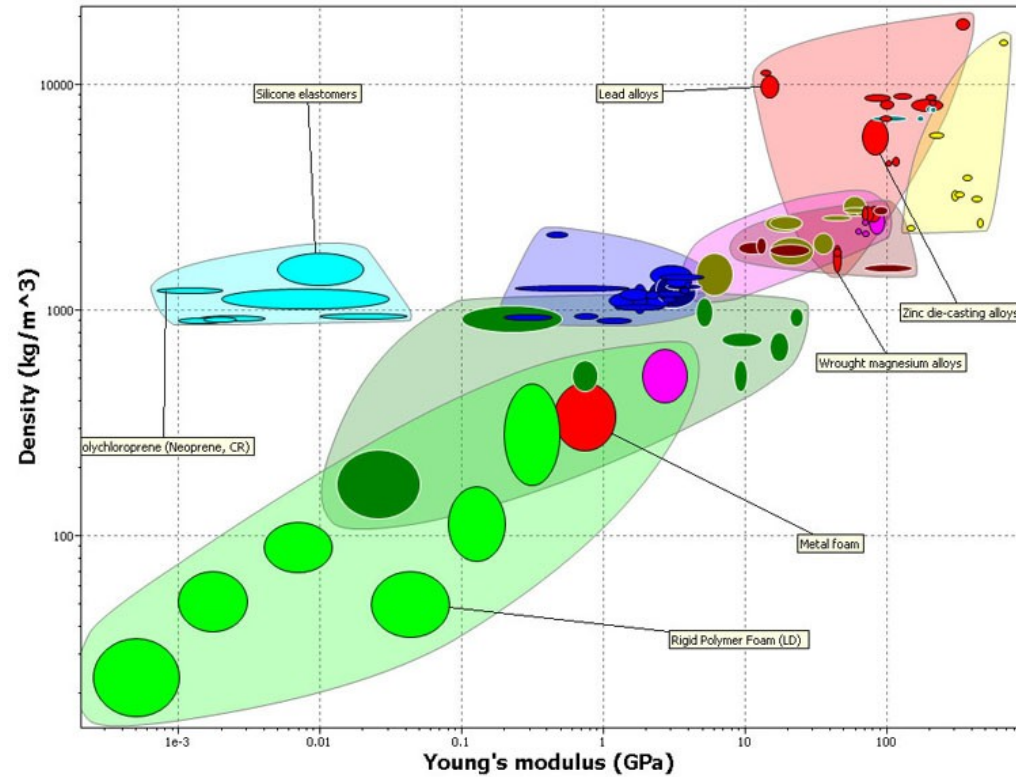
Found 792 erroneous points confirmed against primary sources

Protein activity database

Database contains 10,000 proteins and 2,000,000 compounds

Protein

Drug
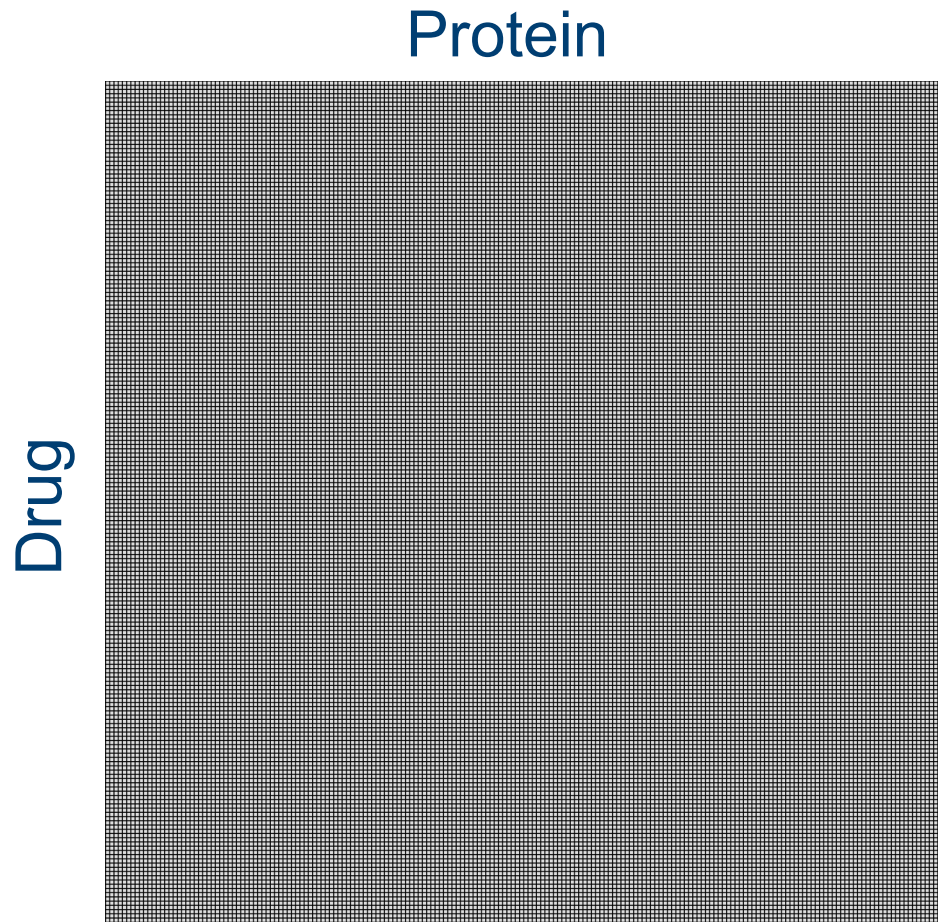
Protein activity data

Database has protein activity for 0.1% of entries

Protein

Drug
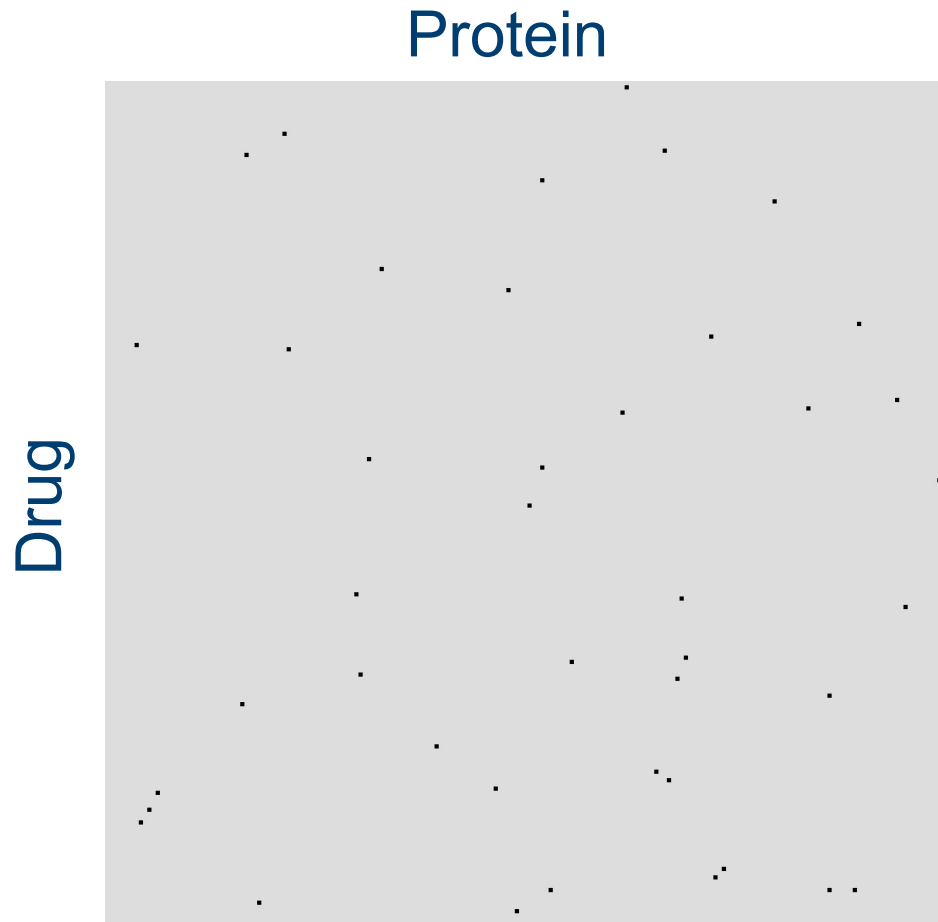
Protein activity data

Filled in 32% of the data points with 75% accuracy

Protein
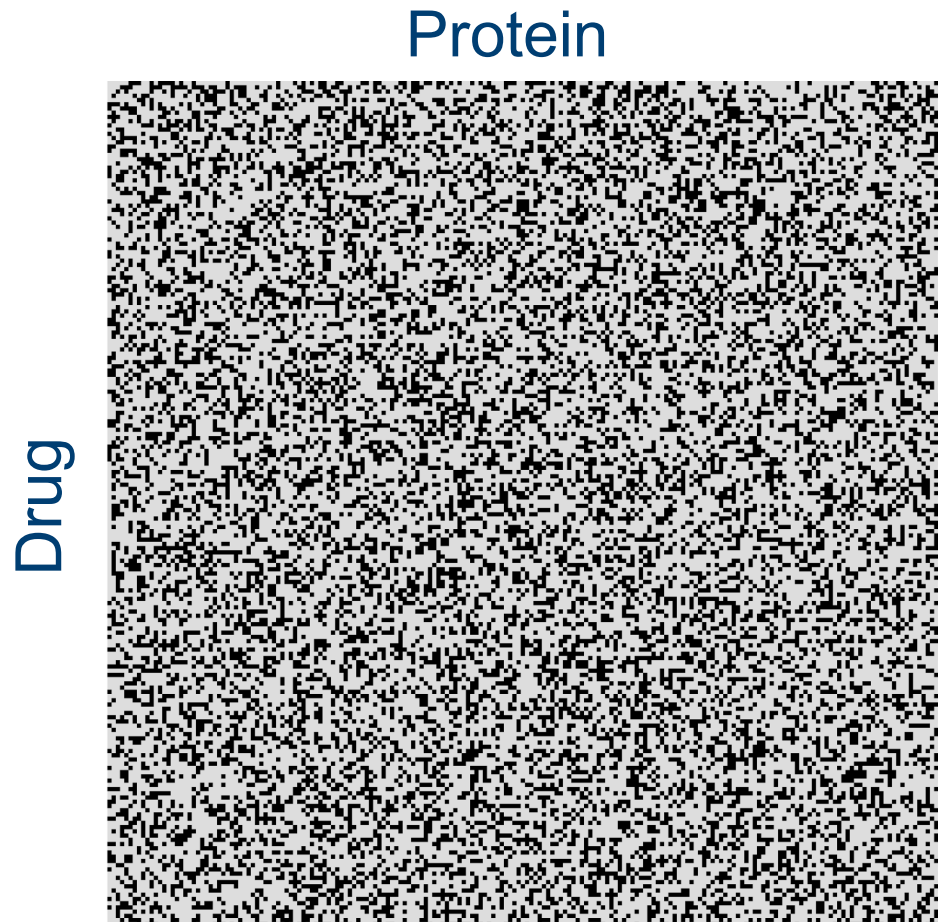
Drug

Cross-validation

Train from first half of the dataset
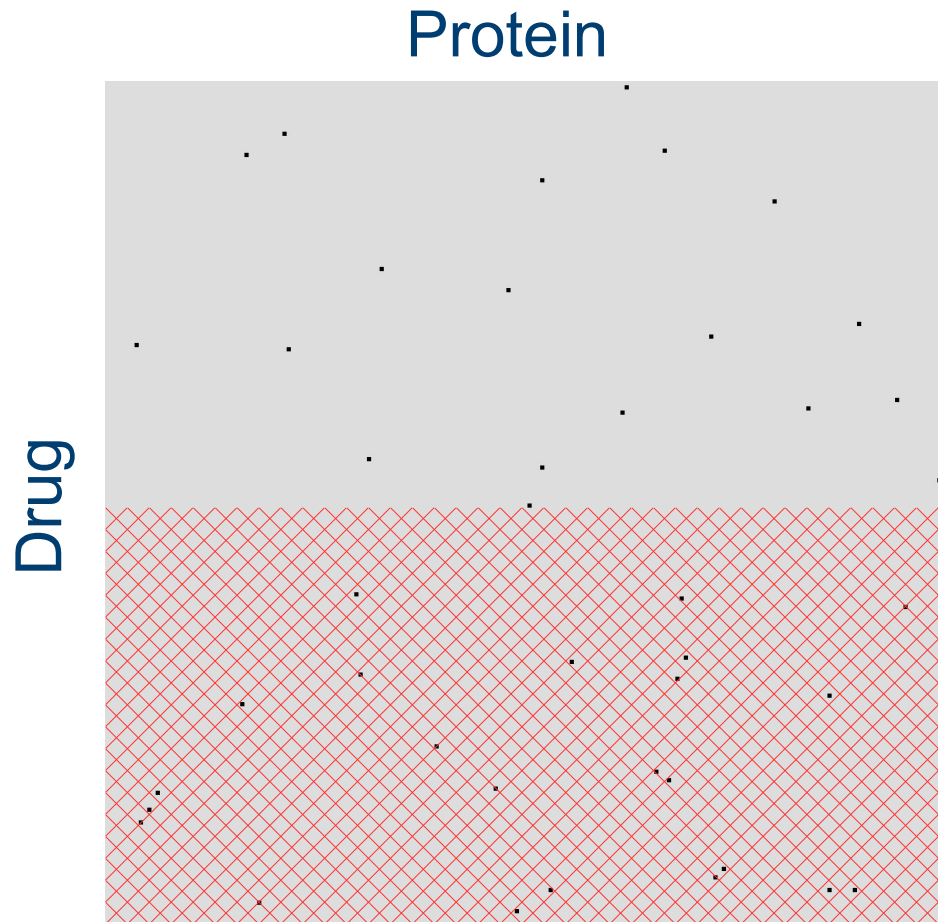
Cross-validation

Test accuracy against second half of the dataset
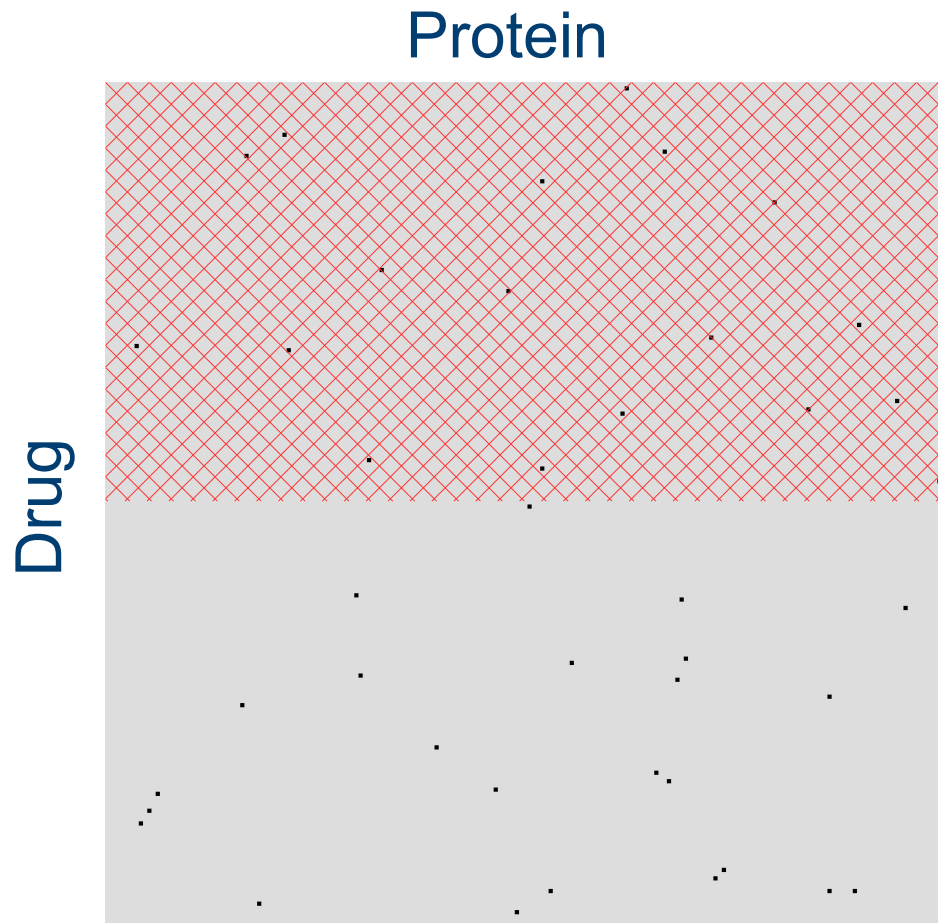
Statistics
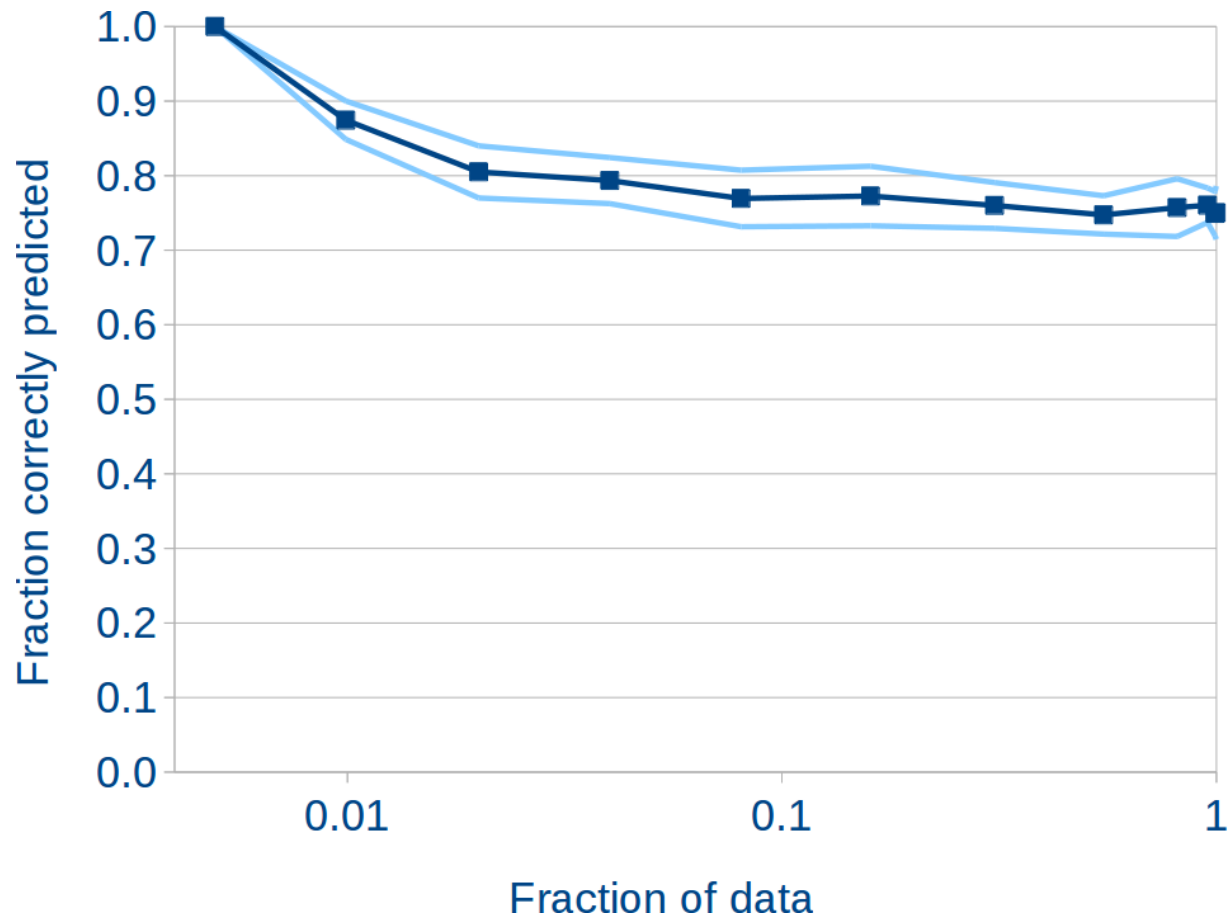
Introducing chemical knowledge of the drug

Exploit SMILES chemical structure to enhance predictions

CCCCN(CC)CCNC(=O)c1cc2c(nn(C)c2s1)-c1ccccc1F

to quantify chemical through 193 descriptors to capture

- Atoms present

- Functional groups

- Properties of the bonds

- Aromatic rings

Chemical knowledge in the database

Protein

Drug + structure
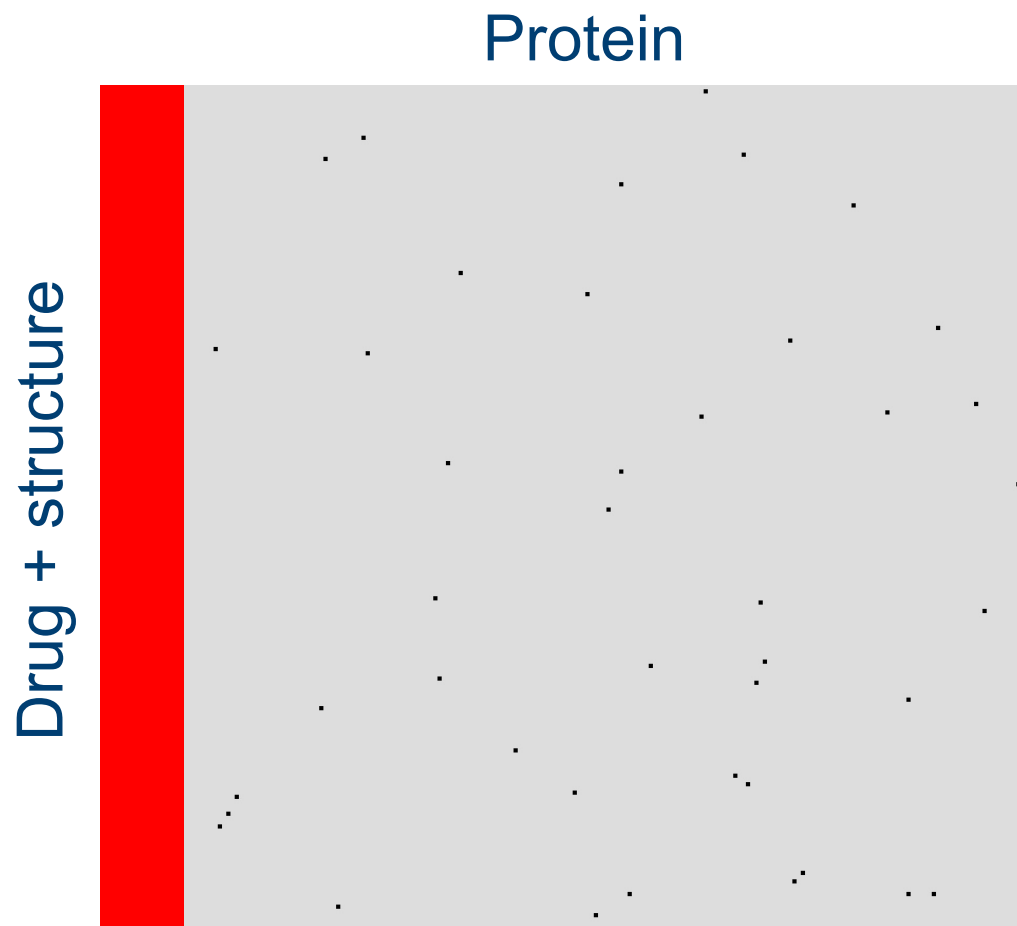
Chemical knowledge of the drug improves predictions

Filled in 61% of the data points with 75% accuracy

Protein
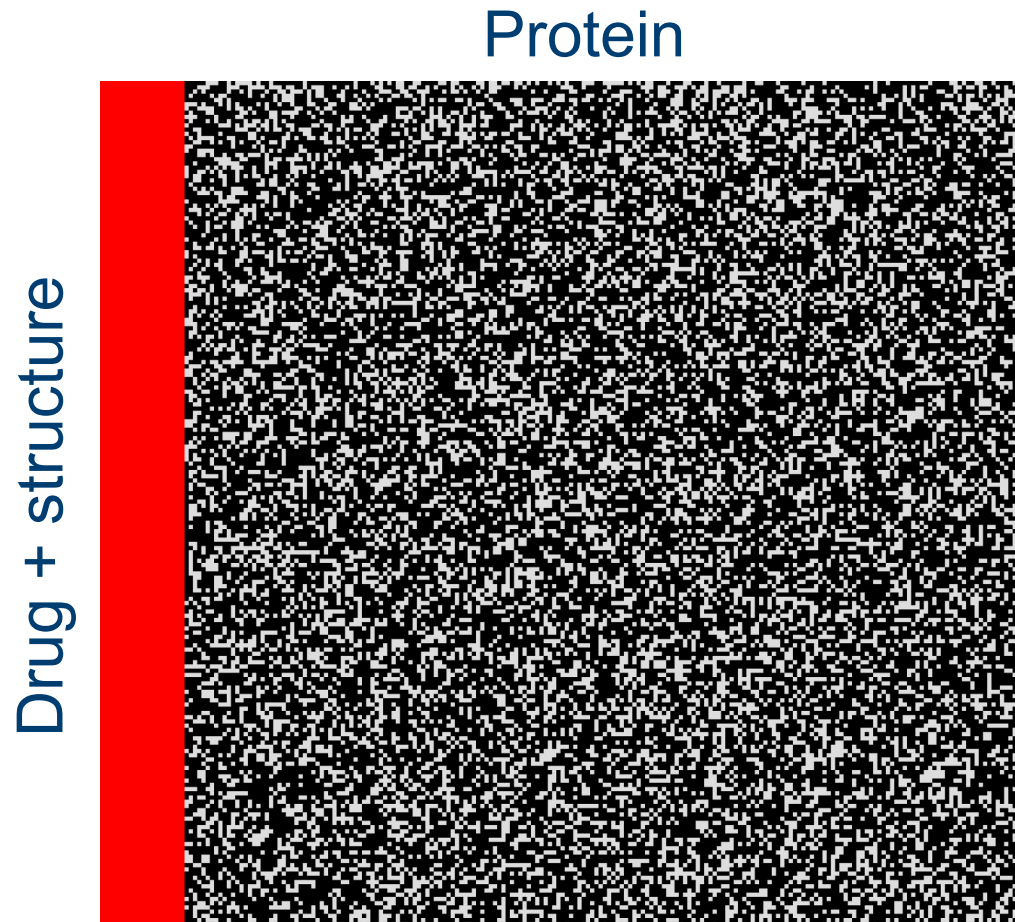
Drug + structure

Predict activity of a new drug

Filled in 76% of the data points with 75% accuracy

Protein + structure
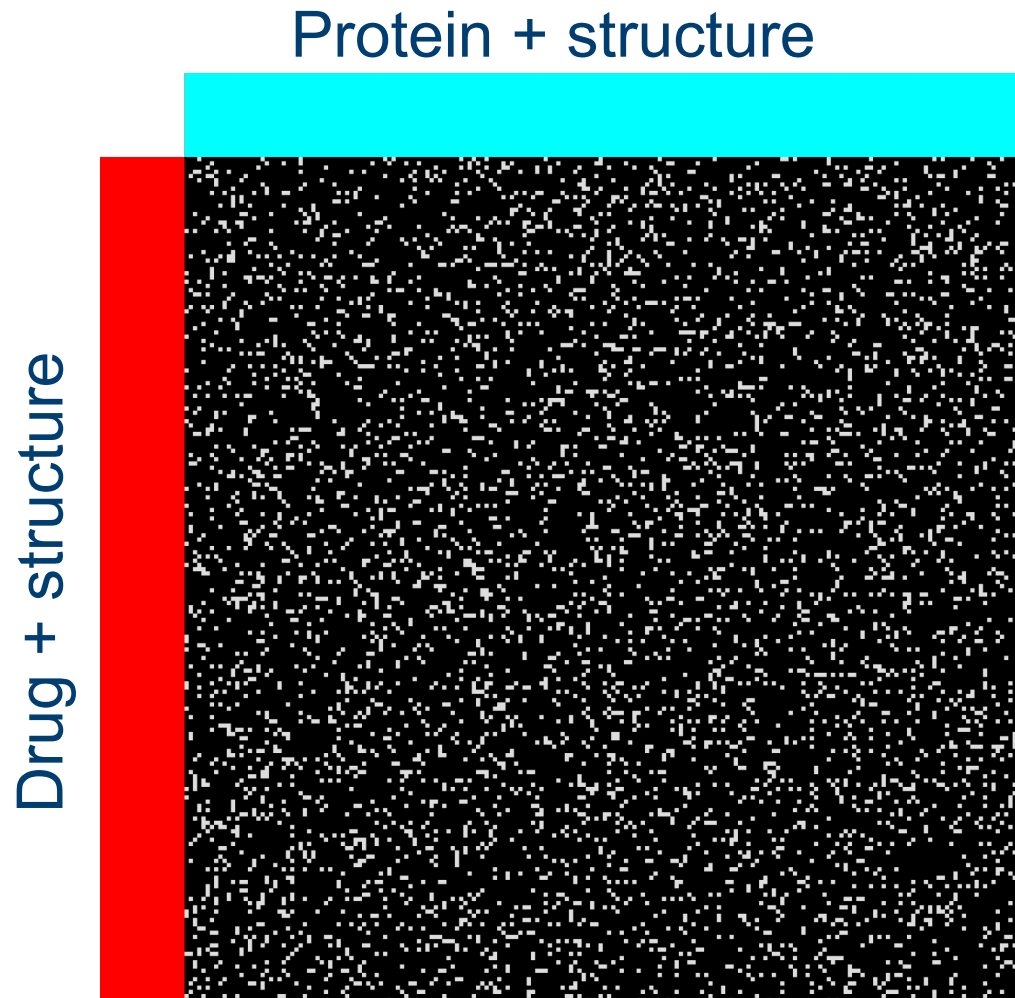
Drug + structure

Summary of drug discovery

Filled in the 0.1% complete protein activity database to:

32% using activity

61% using activity and drug structure

76% using activity, drug structure, and protein structure

Interface – create a network

Define parameters (scope of job / process)

Add multiple runs with different parameters

Where to process job, locally or cloud

Input data files: CSV or SQL

2635.76	2704.79	2746.66	2779.96
2633.45	2702.58	2743.62	2775.40
2631.34	2700.70	2740.99	2771.27
2629.60	2699.17	2738.77	2767.61
2628.58	2698.02	2736.99	2764.49
2628.88	2697.25	2735.66	2762.00
2629.91	2696.87	2734.79	2760.22
2631.32	2696.88	2734.37	2759.24
2632.93	2697.28	2734.42	2759.10
2634.64	2698.05	2734.91	2759.76
2636.35	2699.18	2735.85	2761.12
2638.00	2700.64	2737.22	2763.09

F INTELLEGENS | Pro ×   presentation - Goo ×   ×

← → C ⓘ localhost:3000/networks/new?project_id=3

Intellegens   Back to home page   Admin

network in project Material desig   perimnet With custom checbox   ion elements.

Input data   Run Location   **Paramaters**   Variants

Run type
- ☐ Create and write network
- ☐ Print out results for input data
- ☐ Search for erroneous points
- ☐ Print out specific results
- ☐ Search for optimal
- ☐ Search for duplicate points
- ☐ Recursive Network creation
- ☐ Time the run
- ☐ Print results for sparse data

Num Models   4

Num Hidden Nodes   4

Interface – analyze outputs

Tools to query, analyze, and export data

Key top level statistics on performance of network

Summary of job parameters

llagens   Back to home page   Administrator ▾

Network created by ben@ben.com on 2017-03-16 17:36:13 UTC

| Details | Files generated | DataIn/Out | Ouptut | Graph |

Project:	Test project1
Created by:	ben@ben.com
File:	
Version:	v1.4.2
Pair	1,2
Num models	4
Hidden nodes	4
Max epochs	1000
Task List	3
Fraction to remove	0.8
Num Cycles	1
Iteration loops	1
Mixing Fraction	1.0
Duplicate tolerance	0.1
Opimization cycles	1000000
Local search	.FALSE.
Sparse data	.FALSE.
Error tole	

Error rate	0.0147
Error over sigma	1.2933
Last Updated:	2017-05-02 12:1
Created:	2017-03-16 17:3

Summary

Used artificial intelligence to discover materials and drugs

Handle fragmented data

Merge experiments and simulations into holistic design tool

Worked with six different companies, formed startup intellegens