

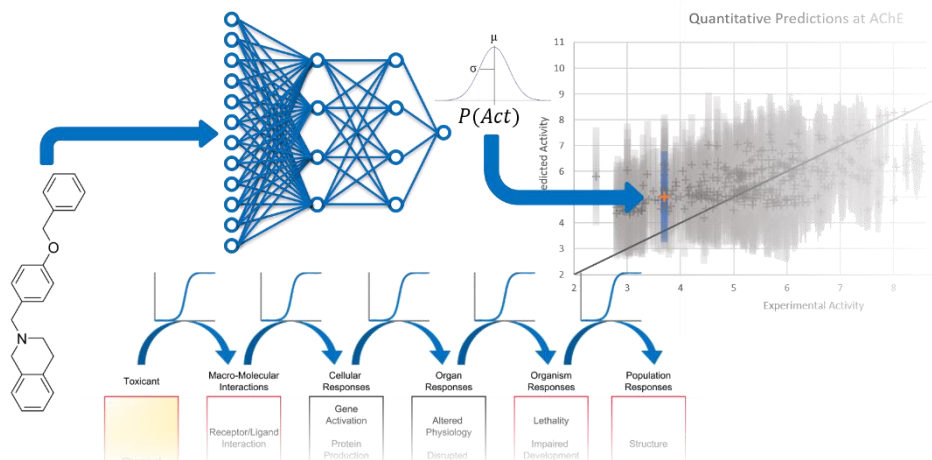
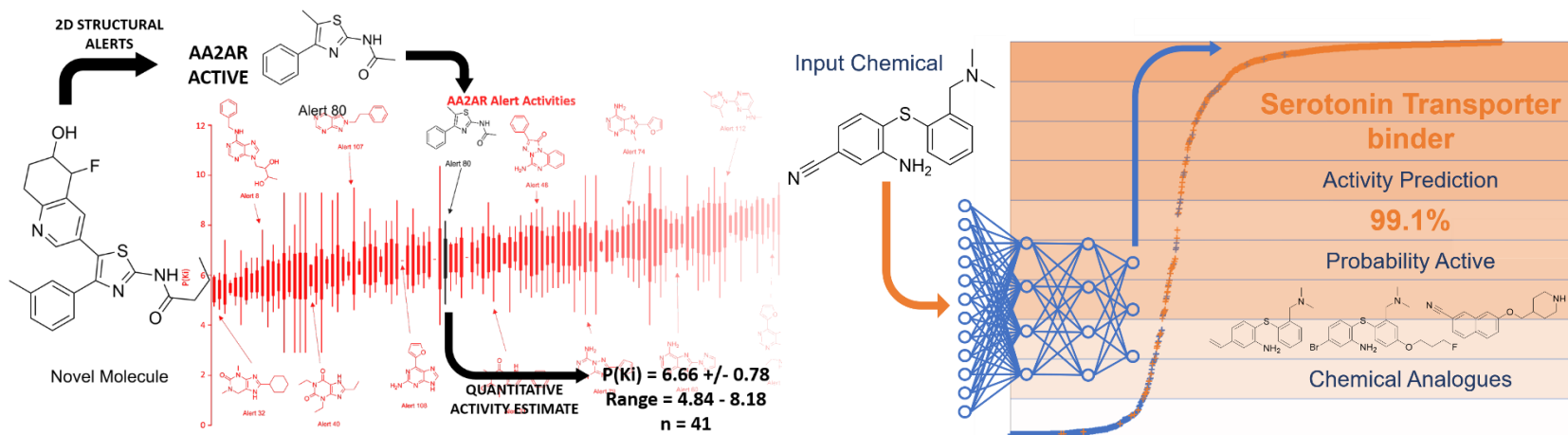
In Silico Models to Predict Human Molecular Initiating Events

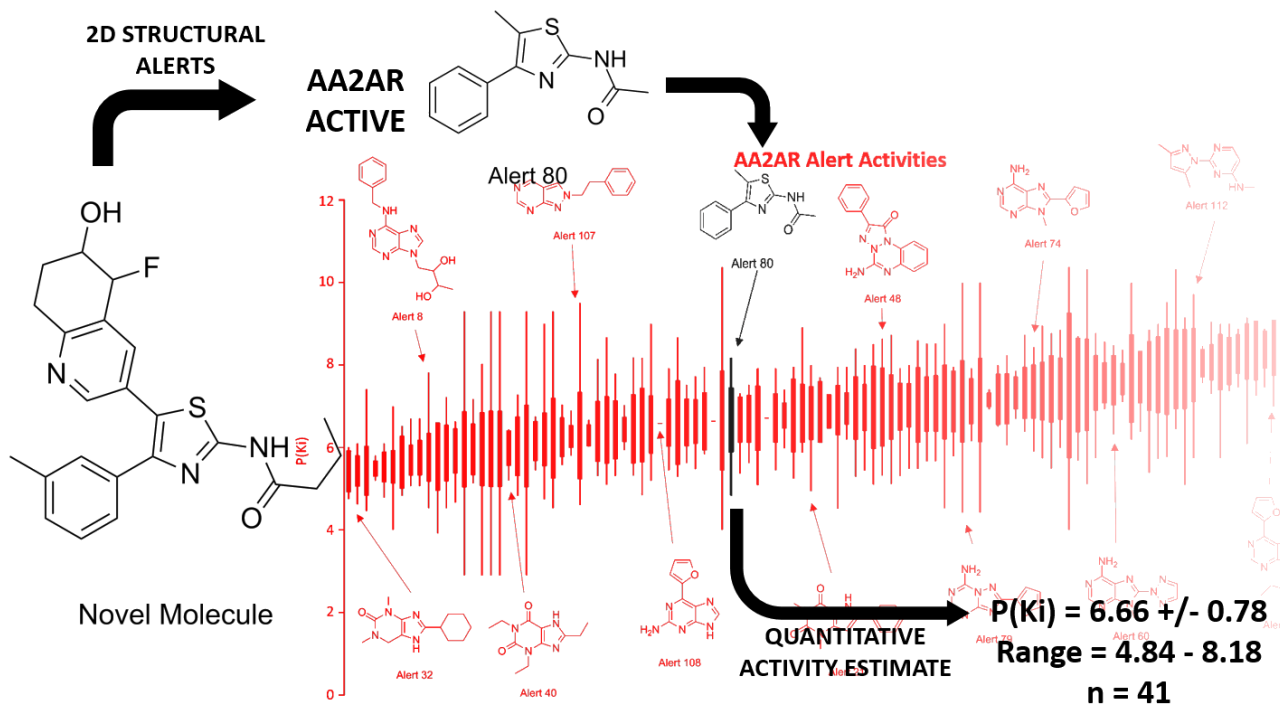
Timothy E H Allen

AJ Wedlake, E Gelžinytė, AM Middleton,
JM Goodman, S Gutsell, P Kukic, PJ Russell

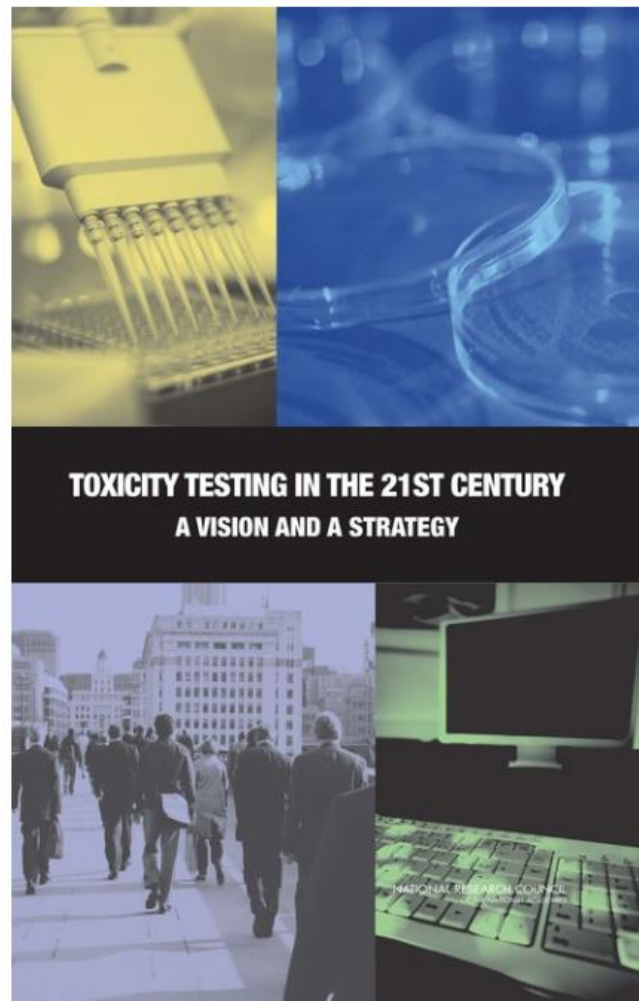
Lush Prize Conference

11th November 2020

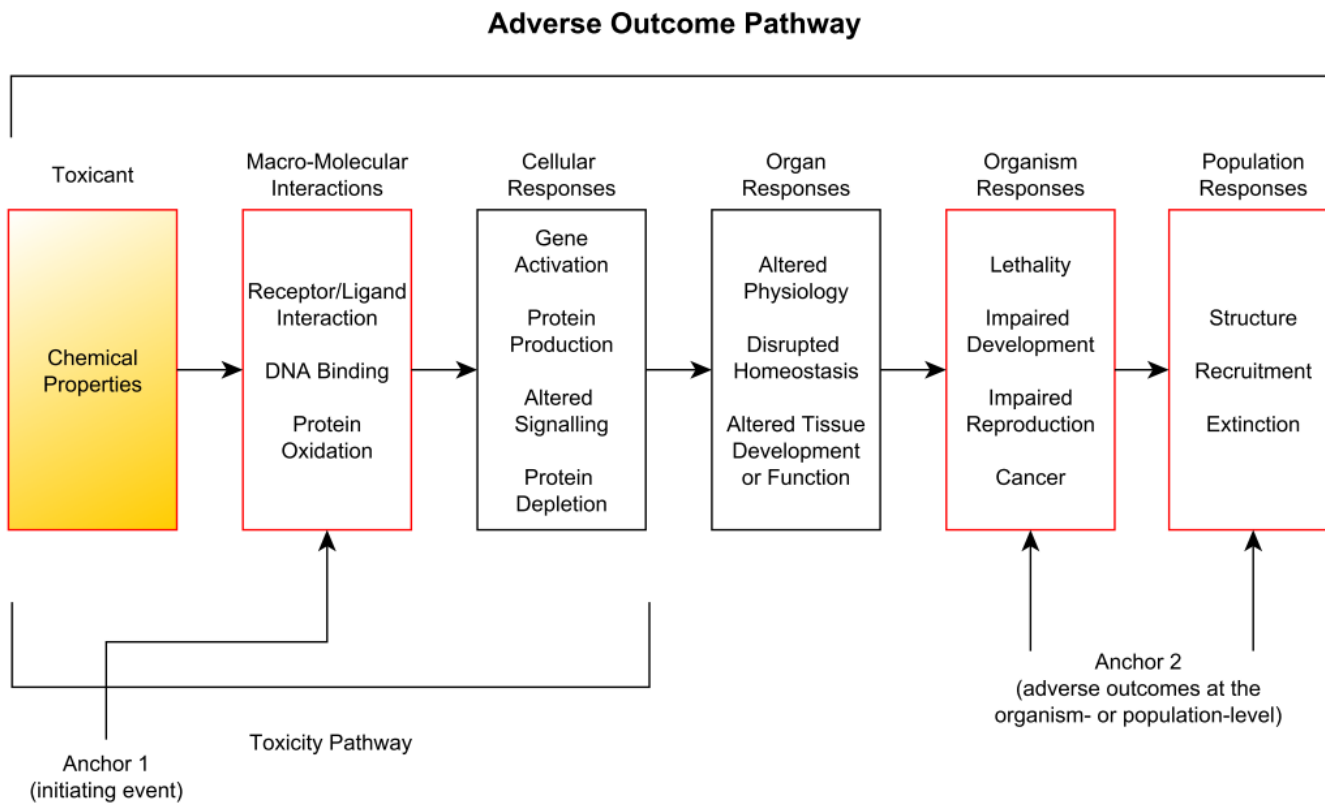




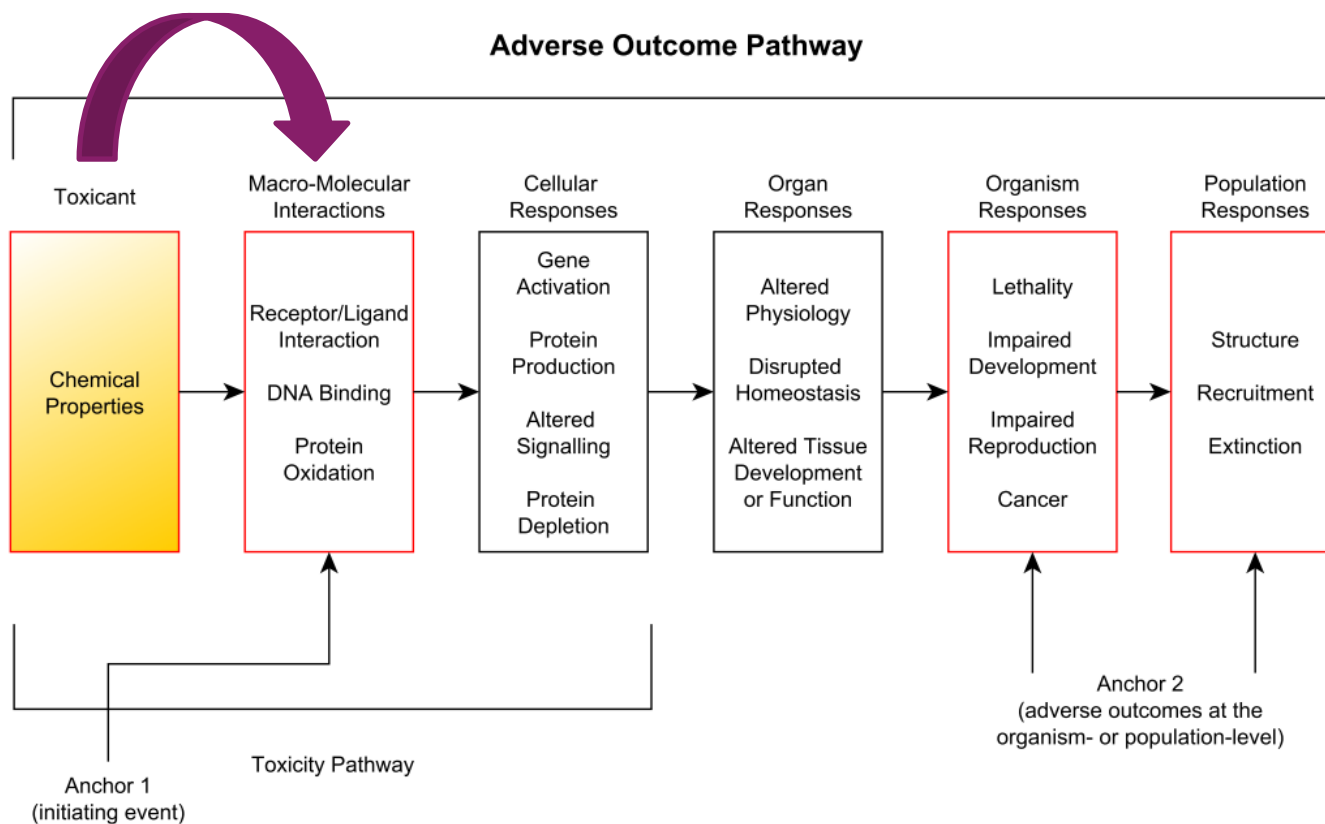
Structural Alerts



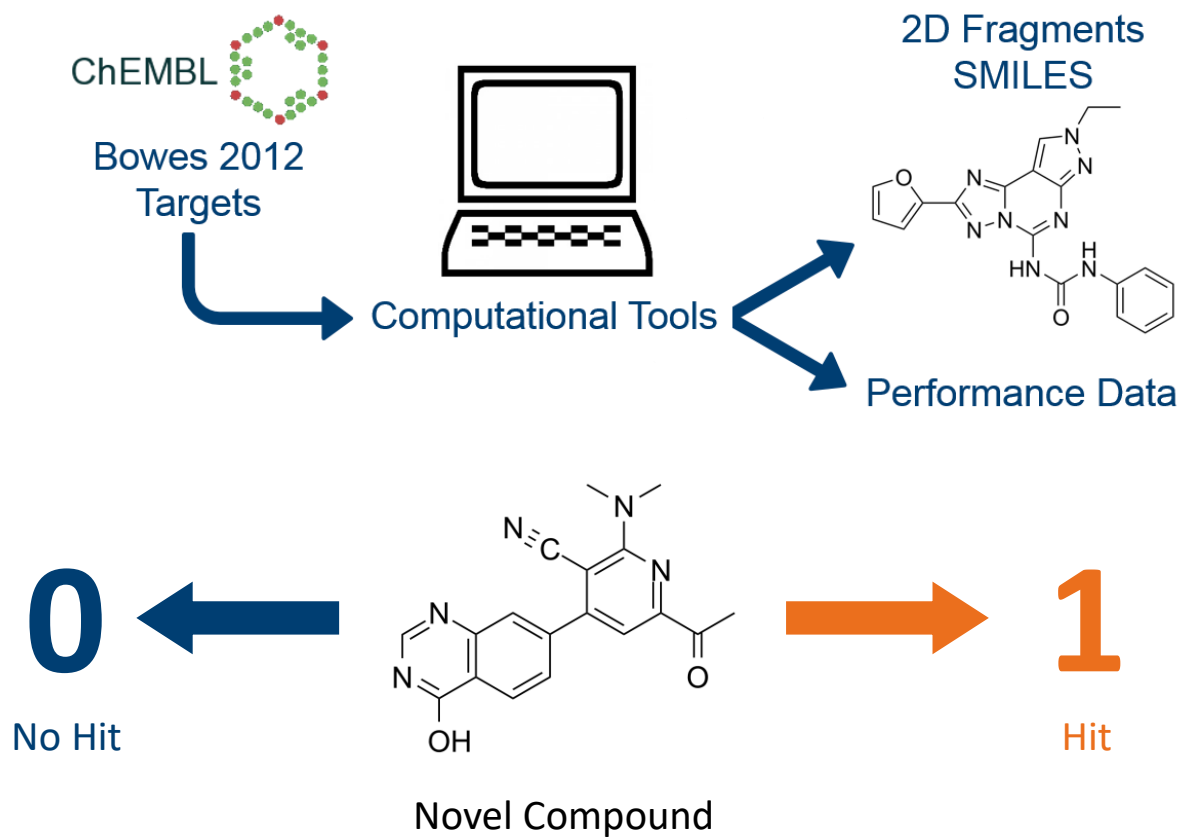
Adverse Outcome Pathway



Adverse Outcome Pathway



Model Construction

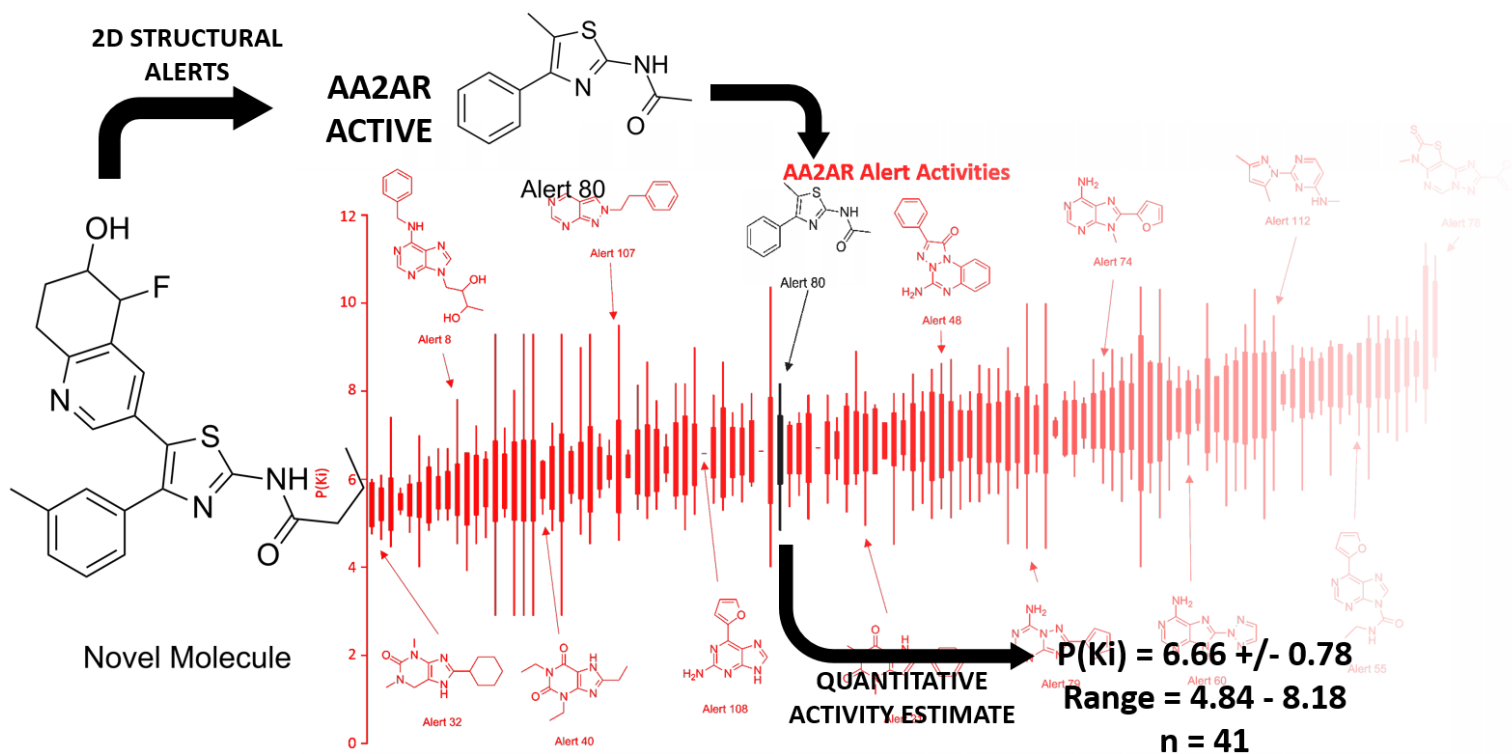


Model Performance

	Test Data			
	SE	SP	ACC	MCC
AVERAGE	66.3	98.6	97.8	0.666
SD	19.0	1.9	2.1	0.148

SE: Sensitivity, SP: Specificity, ACC: Accuracy, MCC: Matthews Correlation Coefficient,
SD: Standard Deviation

In Silico Assessment Procedure



Structural Alerts and Random Forest Models in a Consensus Approach for Receptor Binding Molecular Initiating Events

Andrew J. Wedlake,[†] Maria Folia,[‡] Sam Piechota,[‡] Timothy E. H. Allen,^{†,§} Jonathan M. Goodman,^{*,†} Steve Gutsell,[‡] and Paul J. Russell[‡]

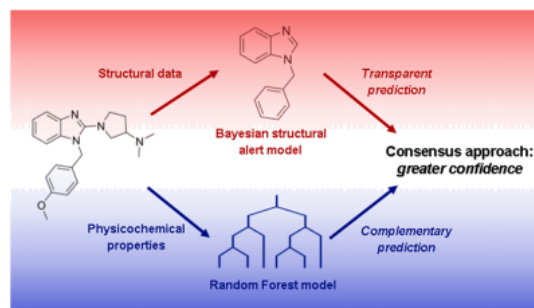
[†]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

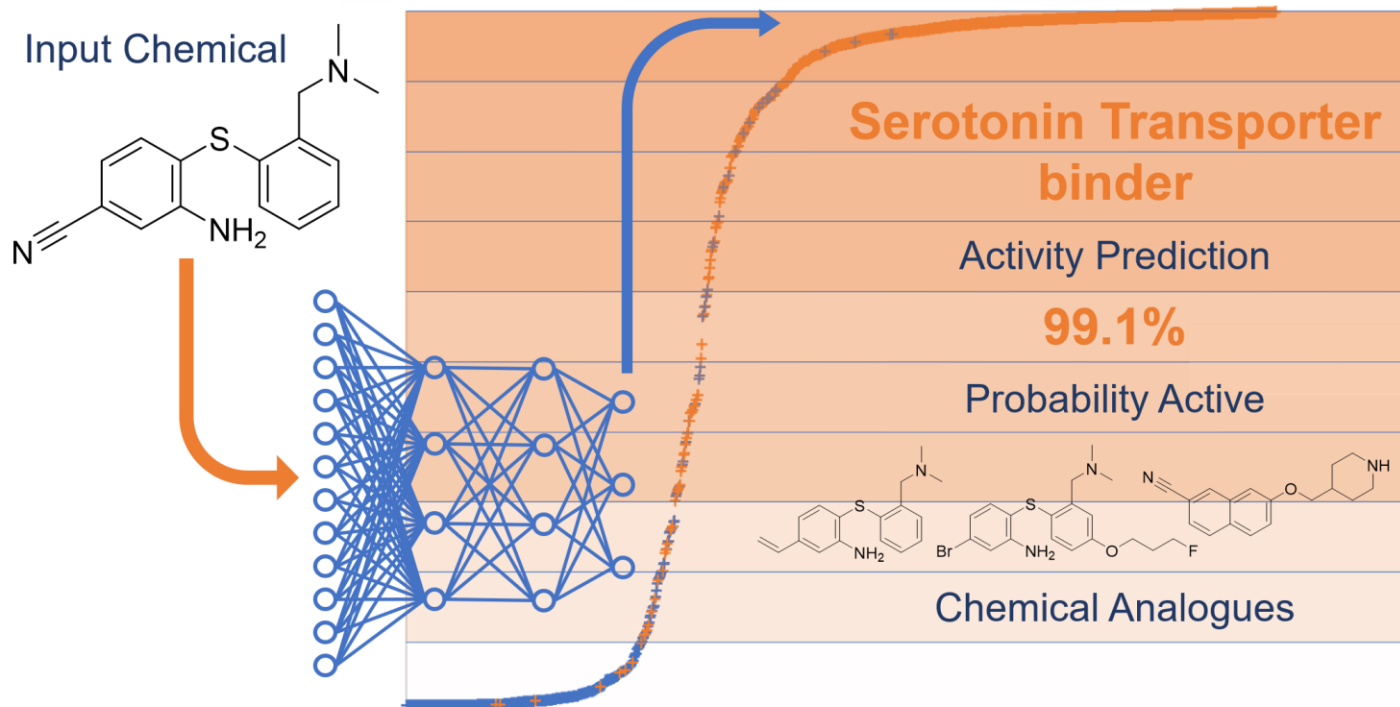
[‡]Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire, MK44 1LQ, United Kingdom

[§]MRC Toxicology Unit, University of Cambridge, Lancaster Road, Leicester LE19HN, United Kingdom

Supporting Information

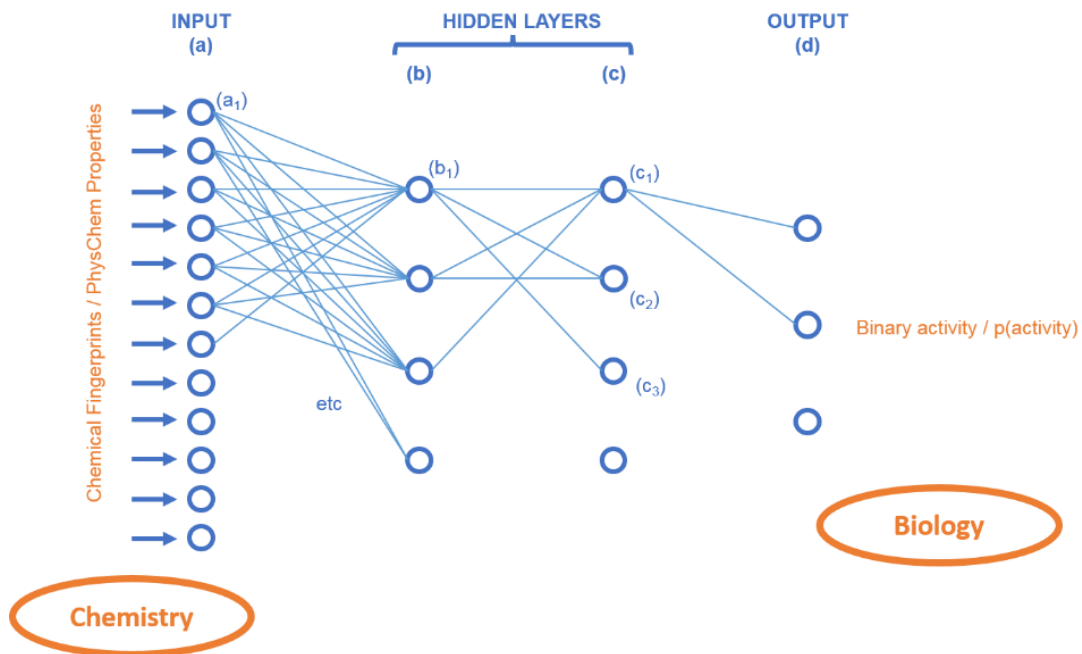
ABSTRACT: A molecular initiating event (MIE) is the gateway to an adverse outcome pathway (AOP), a sequence of events ending in an adverse effect. In silico predictions of MIEs are a vital tool in a modern, mechanism-focused approach to chemical risk assessment. For 90 biological targets representing important human MIEs, structural alert-based models have been constructed with an automated procedure that uses Bayesian statistics to iteratively select substructures. These models give impressive average performance statistics (an average of 92% correct predictions across targets), significantly improving on previous models. Random Forest models have been constructed from physicochemical features for the same targets, giving similarly impressive performance statistics (93% correct predictions). A key difference between





Machine Learning

Neural Networks

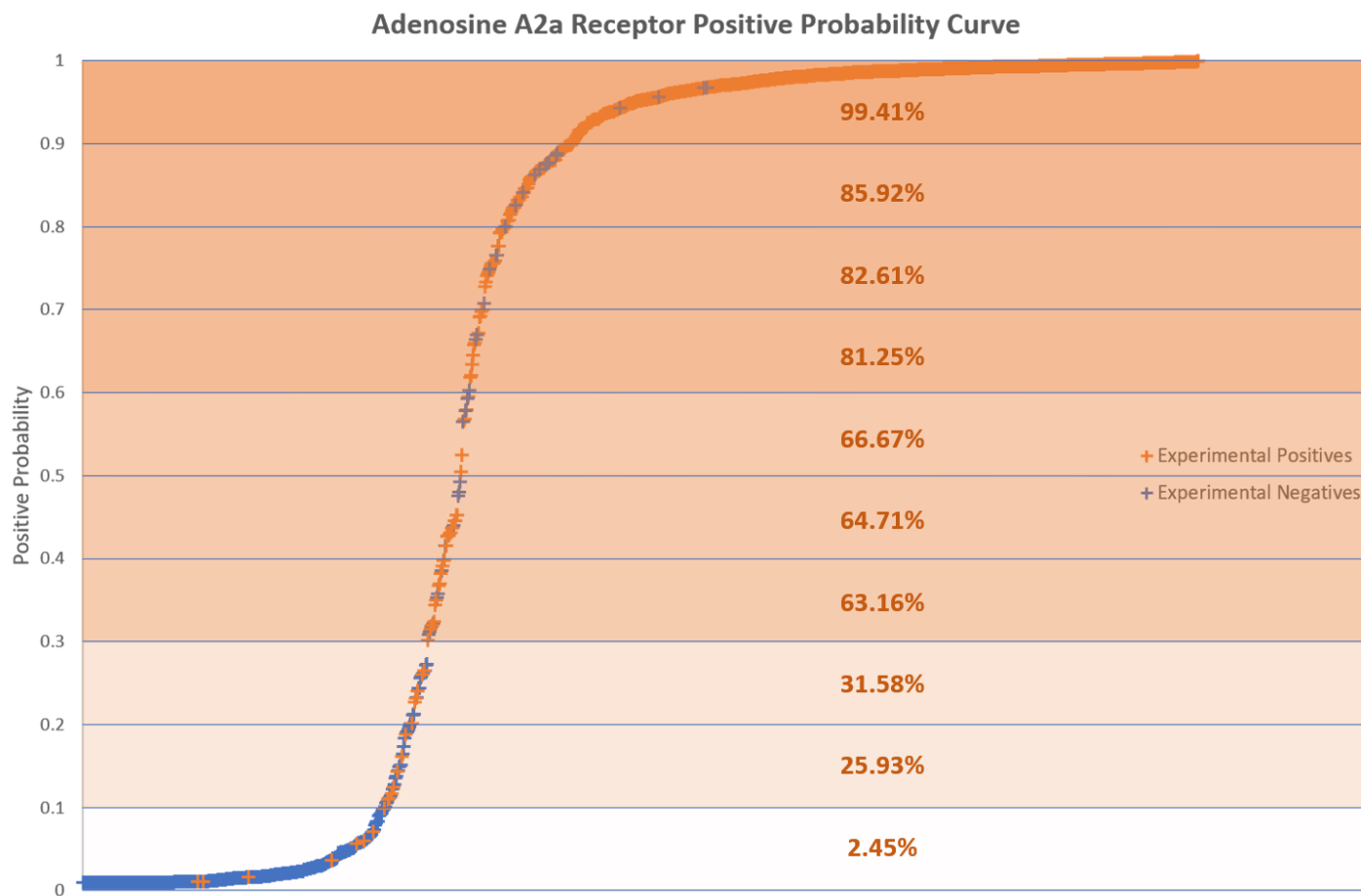


Average Model Performance

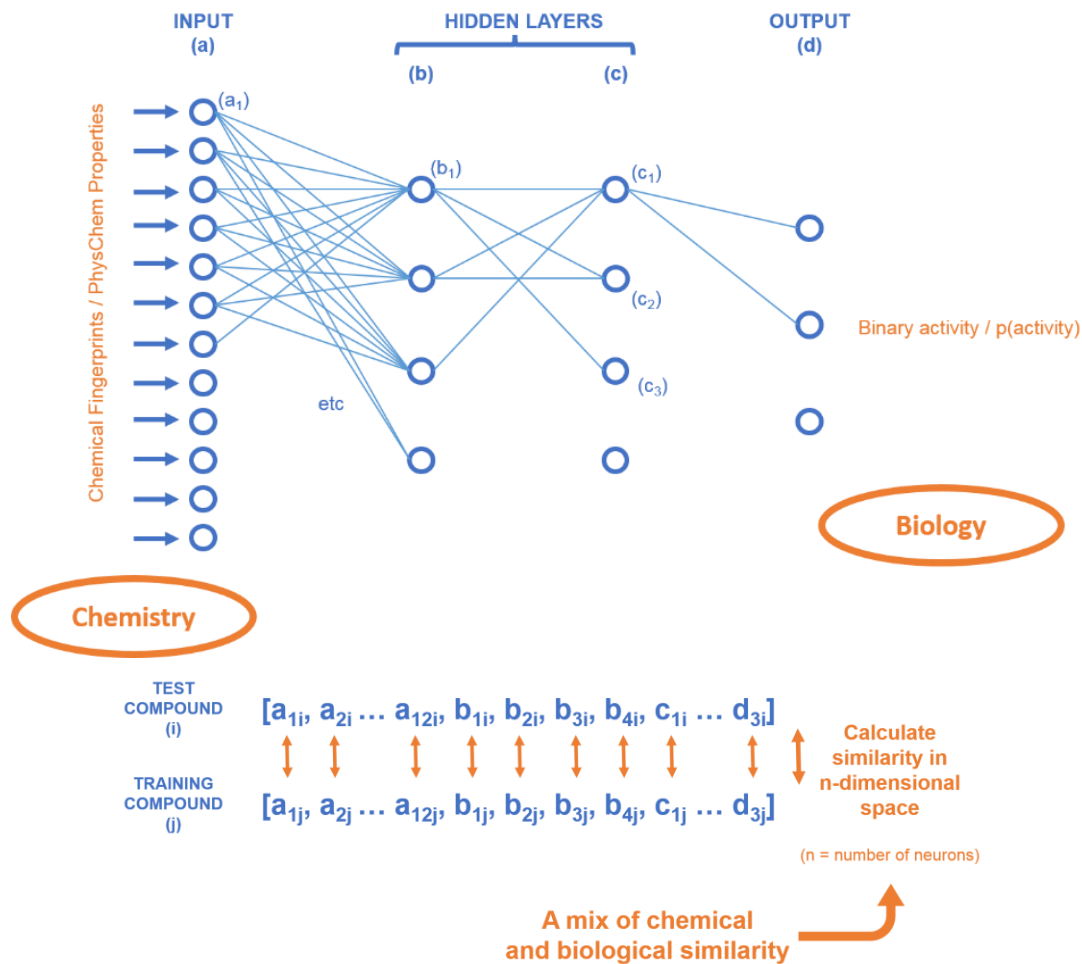
	Training Data					Validation Data					Test Data				
	SE	SP	ACC	MCC	ROC-AUC	SE	SP	ACC	MCC	ROC-AUC	SE	SP	ACC	MCC	ROC-AUC
AVERAGE	92.1	96.5	95.8	0.901	0.99	86.9	93.2	92.5	0.822	0.96	86.2	92.9	92.2	0.814	0.96
SD	8.8	4.2	3.1	0.069	0.02	11.7	5.9	4.1	0.091	0.04	12.1	6.5	4.2	0.093	0.04

SE: Sensitivity, SP: Specificity, ACC: Accuracy, MCC: Matthews Correlation Coefficient, ROC-AUC: Area Under Receiver Operating Characteristic Curve, SD: Standard Deviation

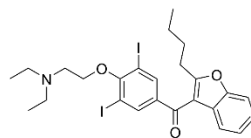
Positive Probability Curve



Network Activation Similarity

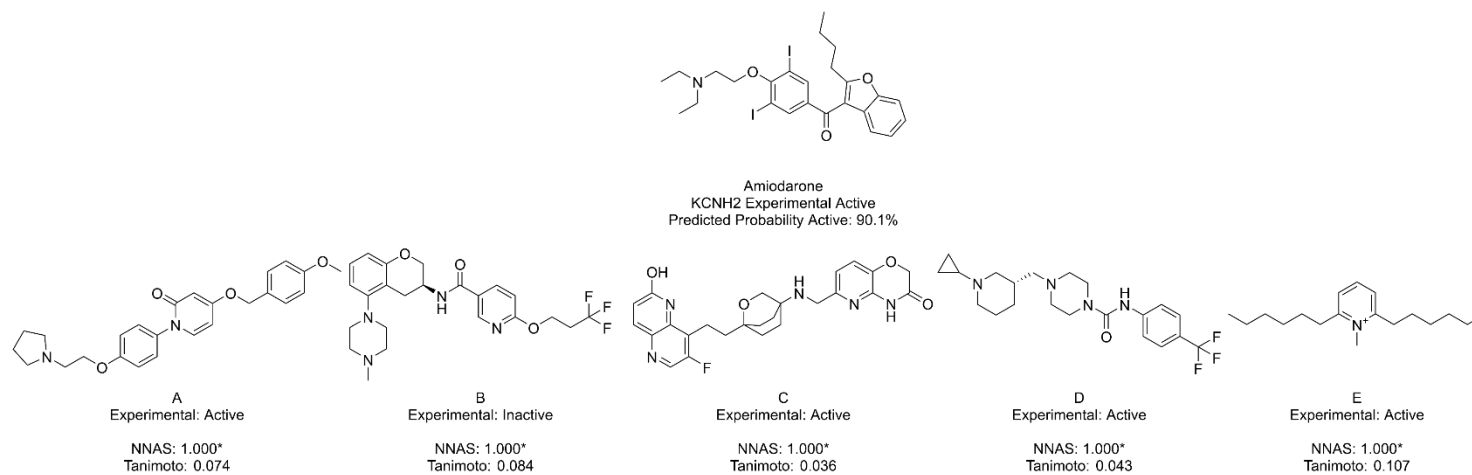


Amiodarone (hERG)

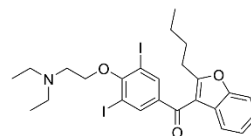


Amiodarone
KCNH2 Experimental Active
Predicted Probability Active: 90.1%

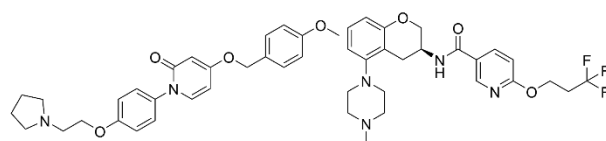
Amiodarone (hERG)



Amiodarone (hERG)

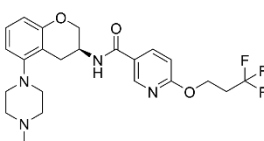


Amiodarone
KCNH2 Experimental Active
Predicted Probability Active: 90.1%



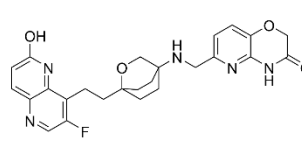
A
Experimental: Active

NNAS: 1.000*
Tanimoto: 0.074



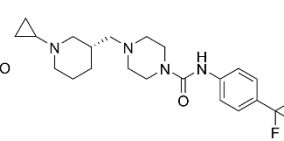
B
Experimental: Inactive

NNAS: 1.000*
Tanimoto: 0.084



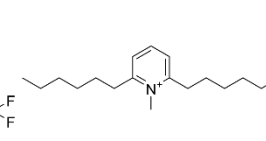
C
Experimental: Active

NNAS: 1.000*
Tanimoto: 0.036



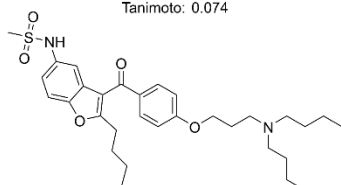
D
Experimental: Active

NNAS: 1.000*
Tanimoto: 0.043



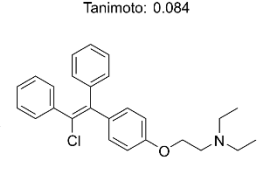
E
Experimental: Active

NNAS: 1.000*
Tanimoto: 0.107



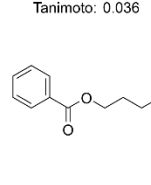
F
Experimental: Active

NNAS: 0.996
Tanimoto: 0.275



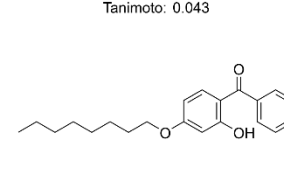
G
Experimental: Active

NNAS: 0.997
Tanimoto: 0.180



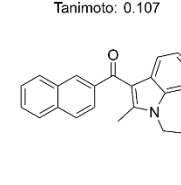
H
Experimental: Inactive

NNAS: 0.909
Tanimoto: 0.157



I
Experimental: Inactive

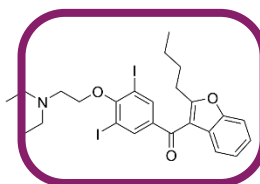
NNAS: 0.915
Tanimoto: 0.154



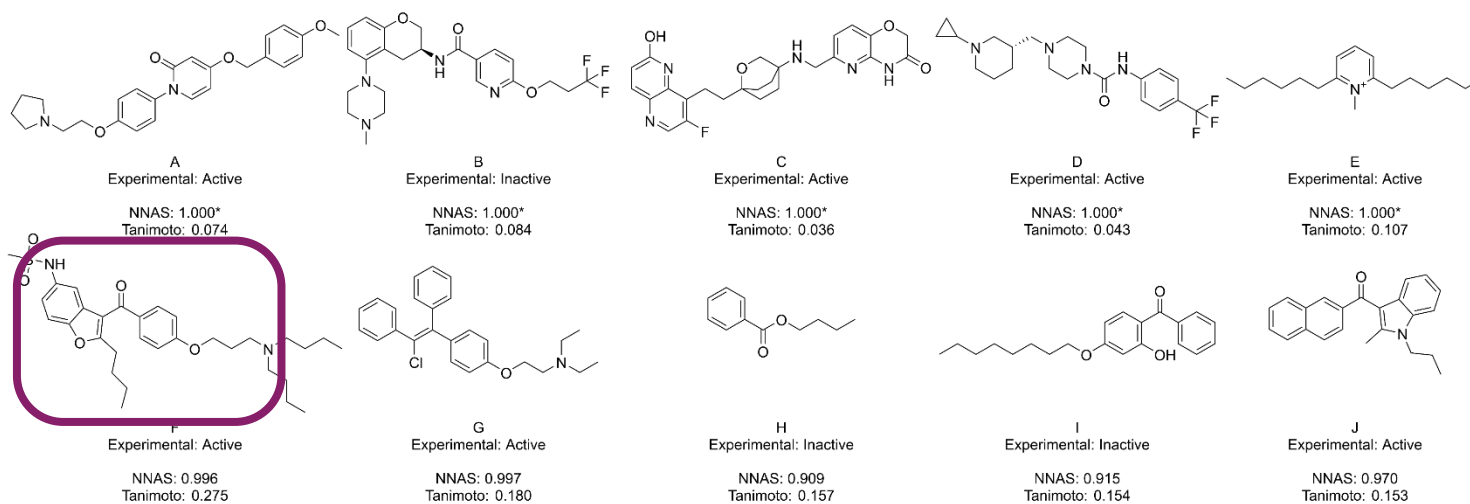
J
Experimental: Active

NNAS: 0.970
Tanimoto: 0.153

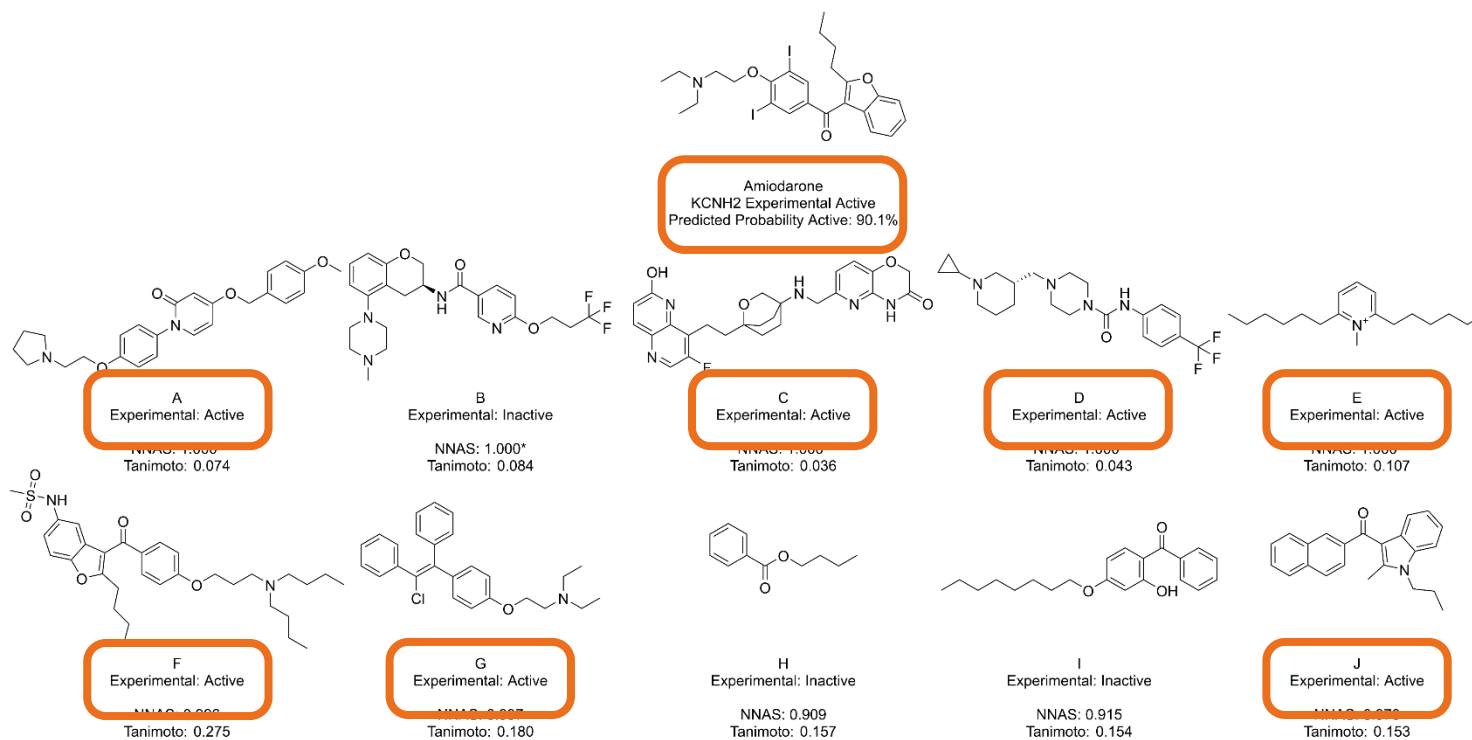
Amiodarone (hERG)



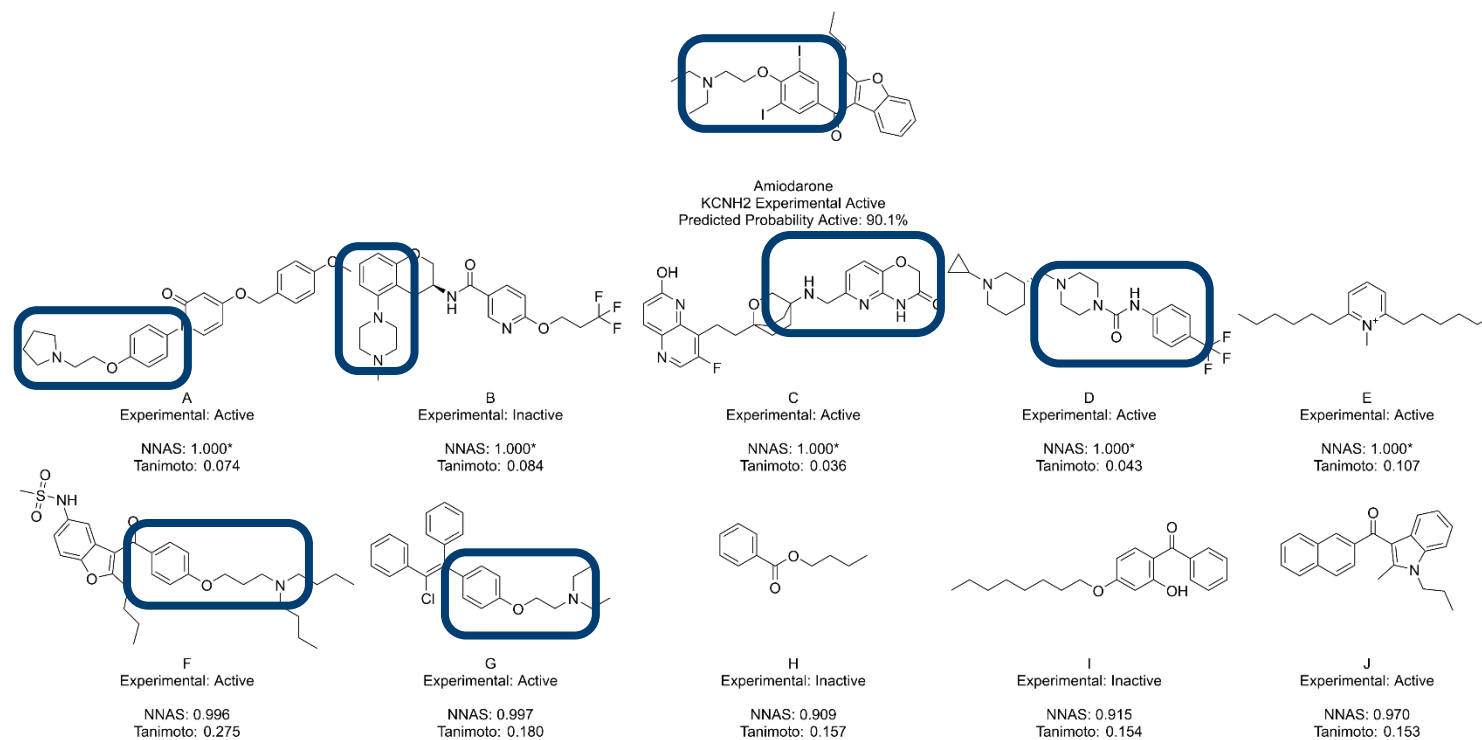
Amiodarone
KCNH2 Experimental Active
Predicted Probability Active: 90.1%

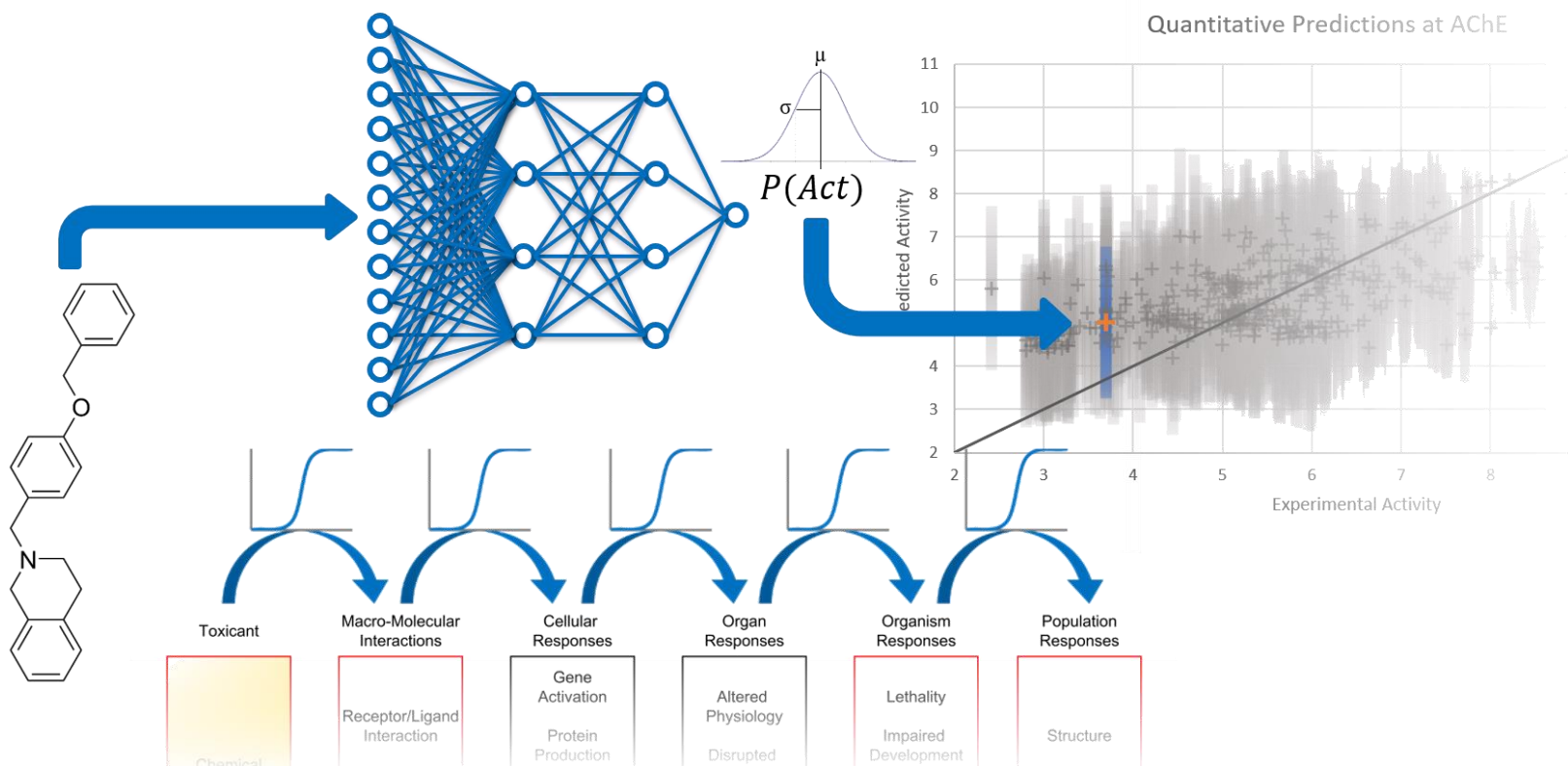


Amiodarone (hERG)



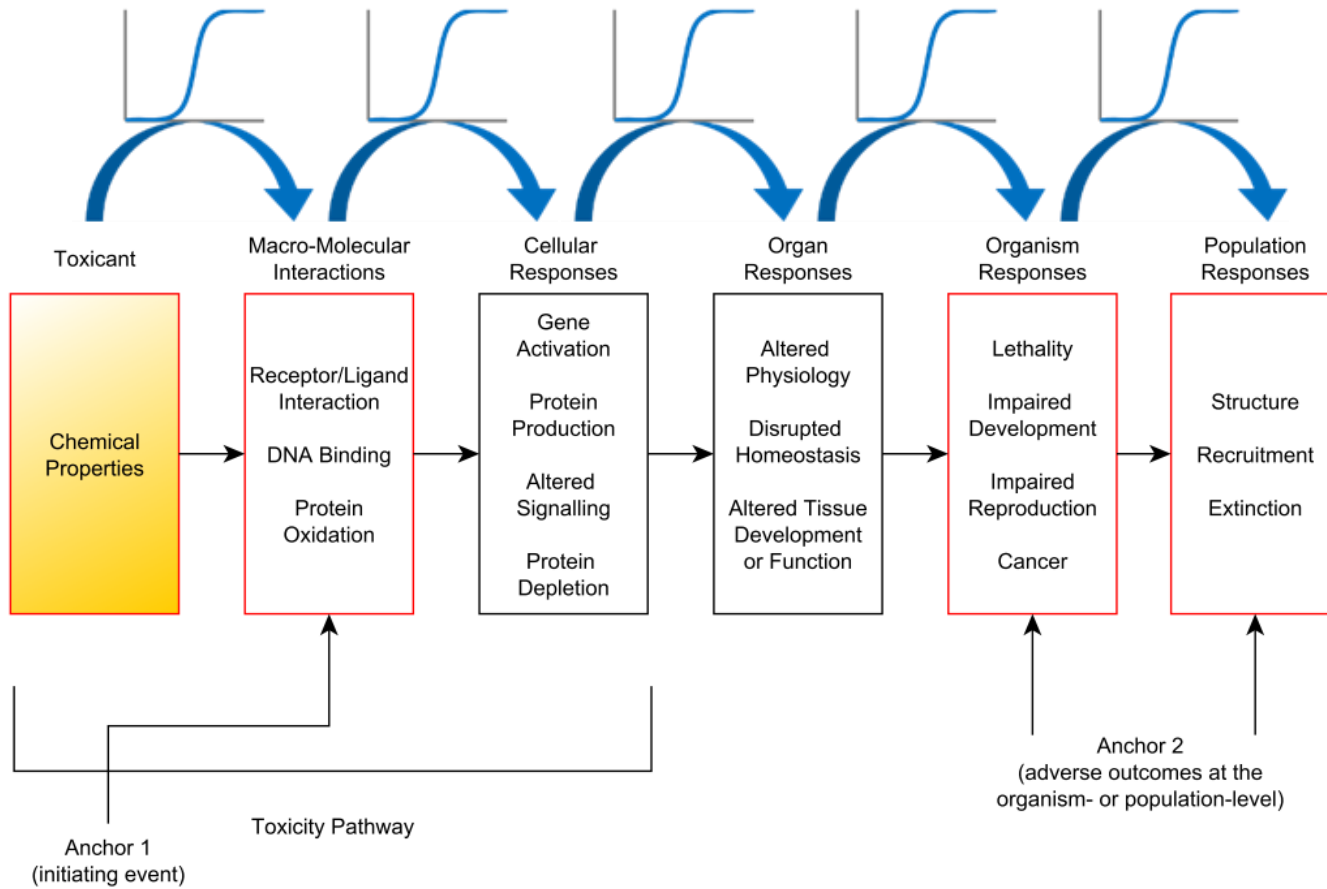
Amiodarone (hERG)





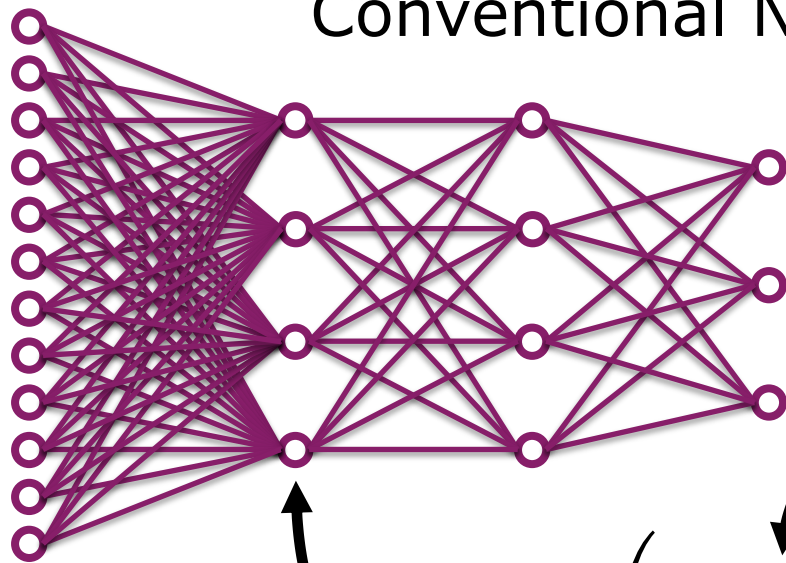
Quantitative Predictions

Quantitative Modelling



Conventional Learning

Conventional Neural Network

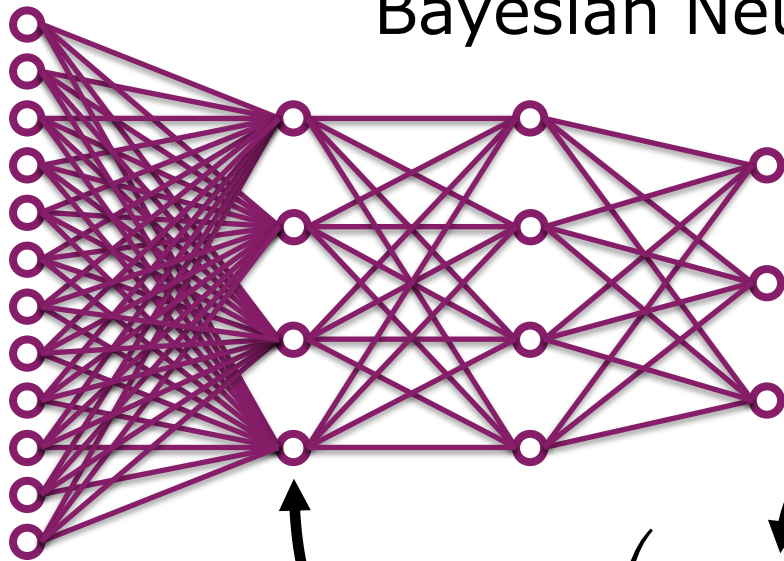


Weights and biases as fixed values

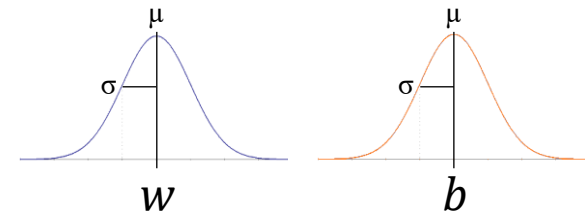
Therefore predictions are fixed values

$$f(\mathbf{x}) = K\left(\sum_i w_i(x_i)\right) + \mathbf{b}$$

Bayesian Neural Network



Weights and biases defined as probability distributions



Gives predictions meaningful errors

$$f(\mathbf{x}) = K\left(\sum_i w_i(x_i)\right) + \mathbf{b}$$

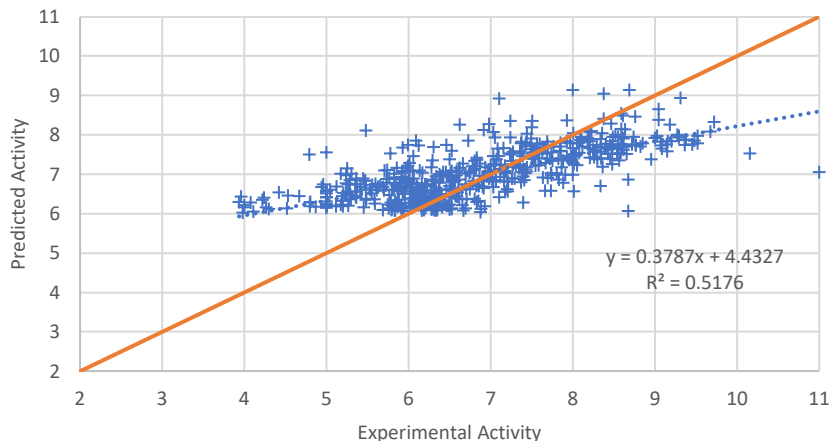
Best Models

	Train MAE	Valid. MAE	Test MAE	Ext. Val. MAE
AVERAGE	0.4870	0.6143	0.6202	0.9163
SD	0.0359	0.0515	0.0535	0.1968

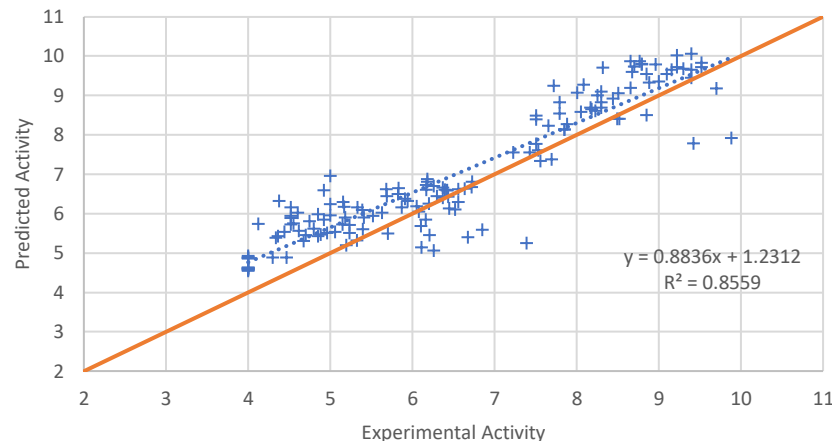
MAE: Mean Absolute Error, SD: Standard Deviation

Some Models (Ext. Valid. Set)

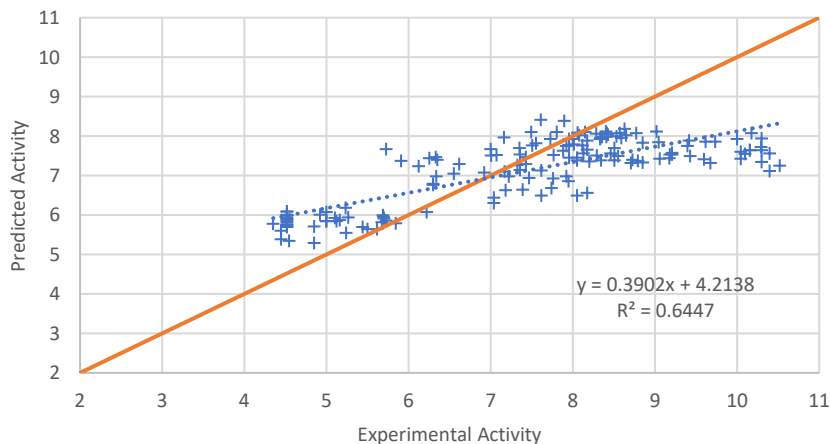
Quantitative Predictions at DD2R



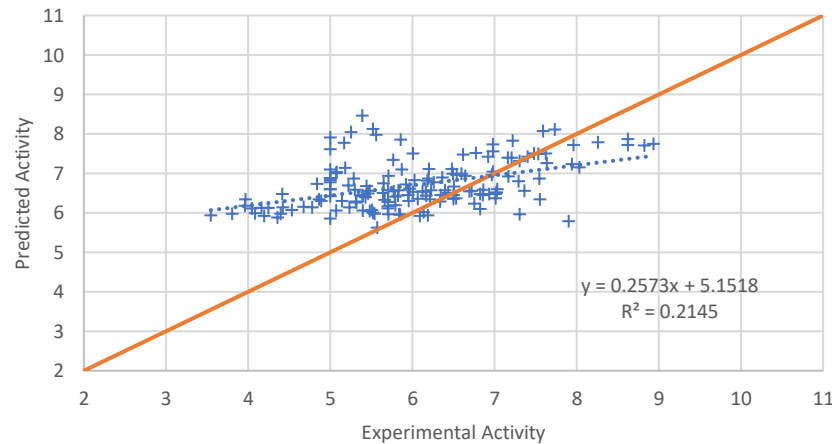
Quantitative Predictions at CHRM3



Quantitative Predictions at NR3C1

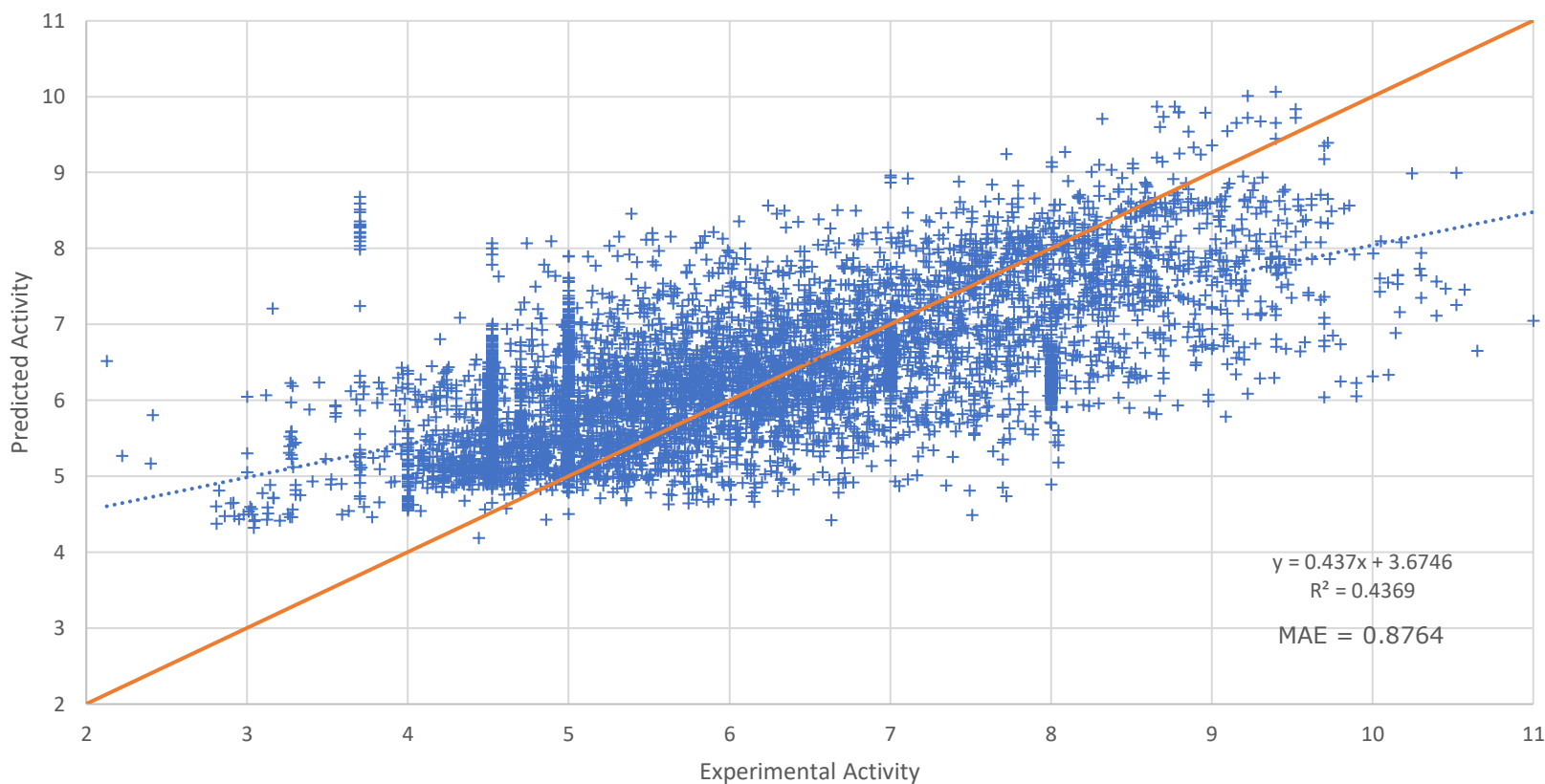


Quantitative Predictions at SLC6A3



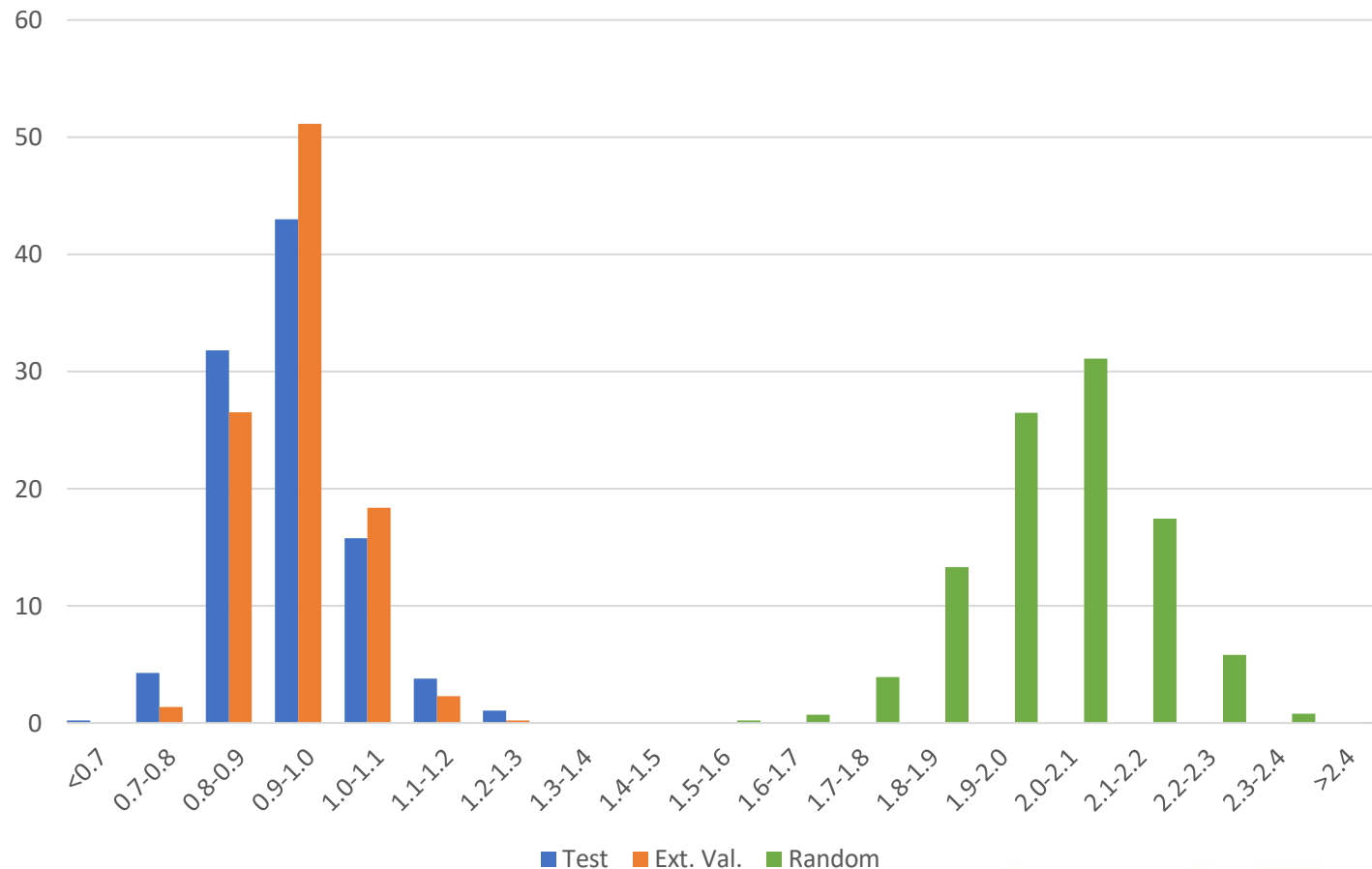
All Predictions (Ext. Valid. Set)

Quantitative Predictions at All Targets



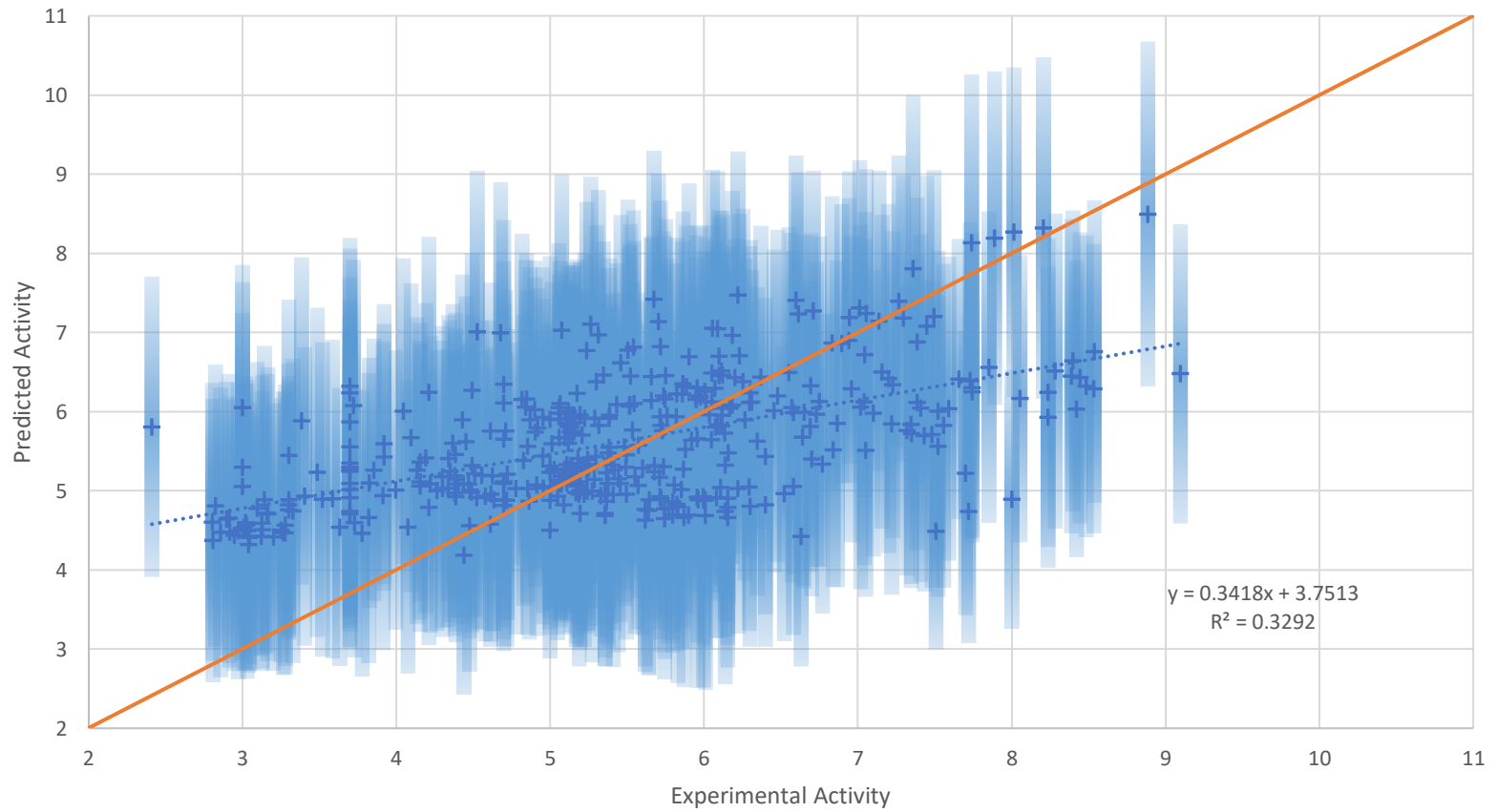
Distribution of Uncertainties

Acetylcholinesterase Uncertainty Histogram by Proportion



Predictions with Uncertainties

Quantitative Predictions at Acetylcholinesterase (Ext. Valid.) with 95% CI

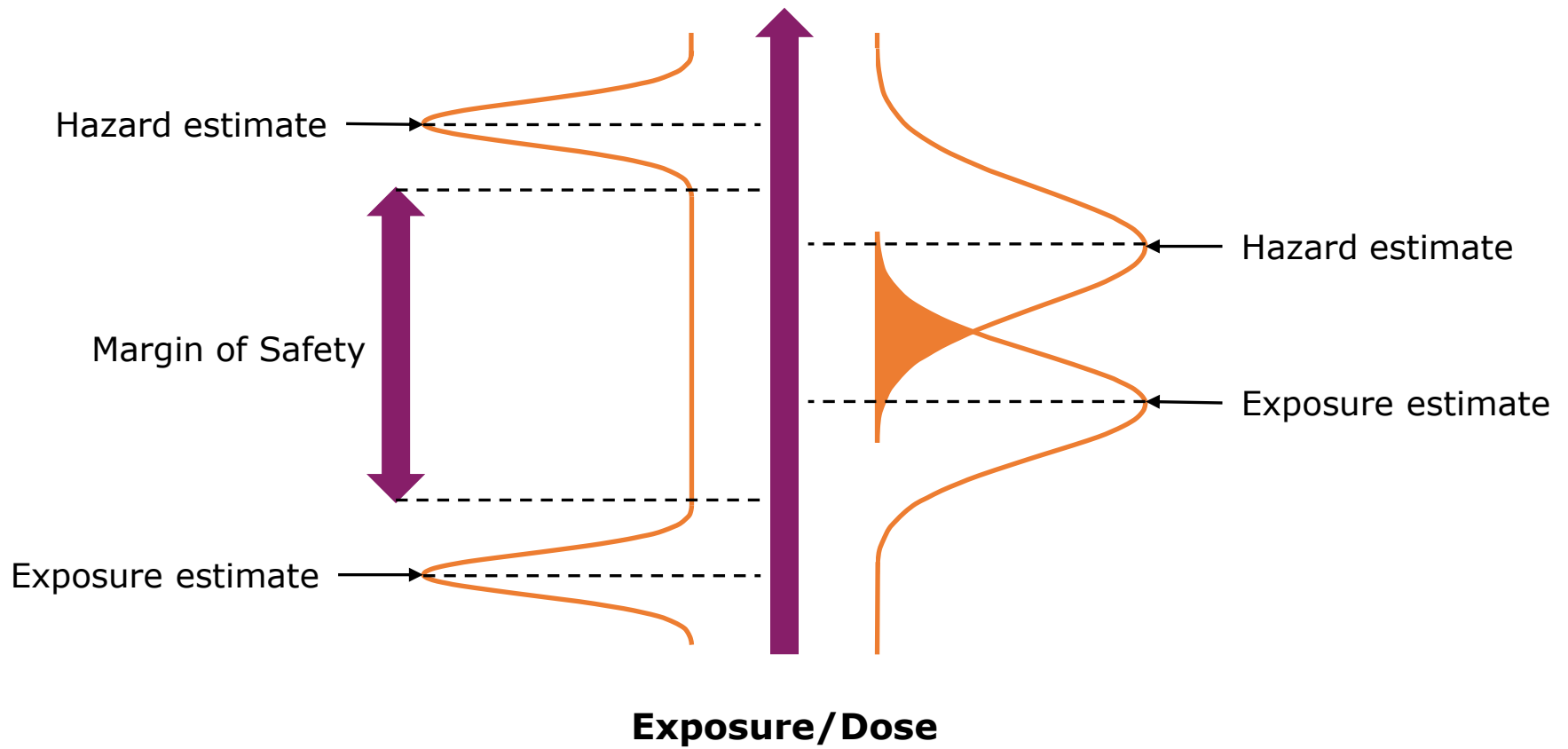


Safety Decision Making



Exposure/Dose

Safety Decision Making



Conclusions

- **Molecular initiating events** are good targets for computational toxicology modelling
- We have applied a number of modelling approaches to predict these important interactions
- In binary classification, **a combination of structural alerts, random forests and neural networks** give the highest statistical performance and most confidence
- **These models have been implemented at SEAC Unilever** into their NGRA safety evaluation procedure
- Machine learning can also be used for regression modelling, and **Bayesian learning offers opportunities to understand the uncertainty in predictions**

So...

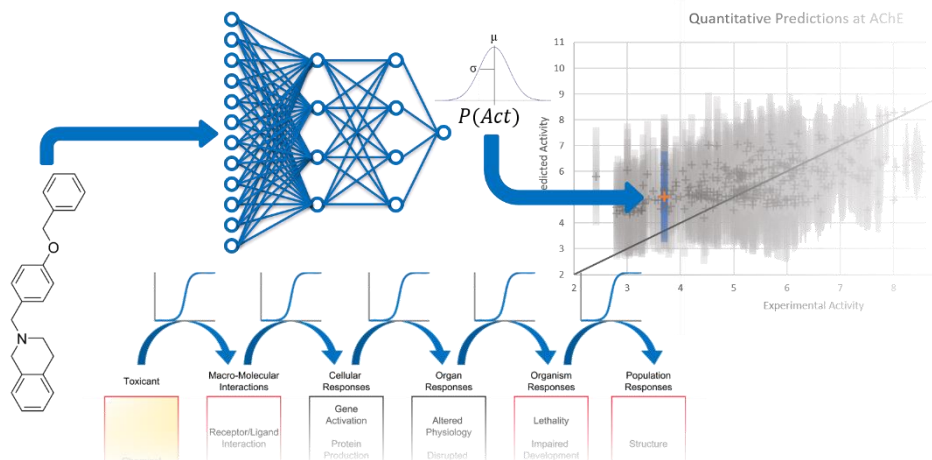
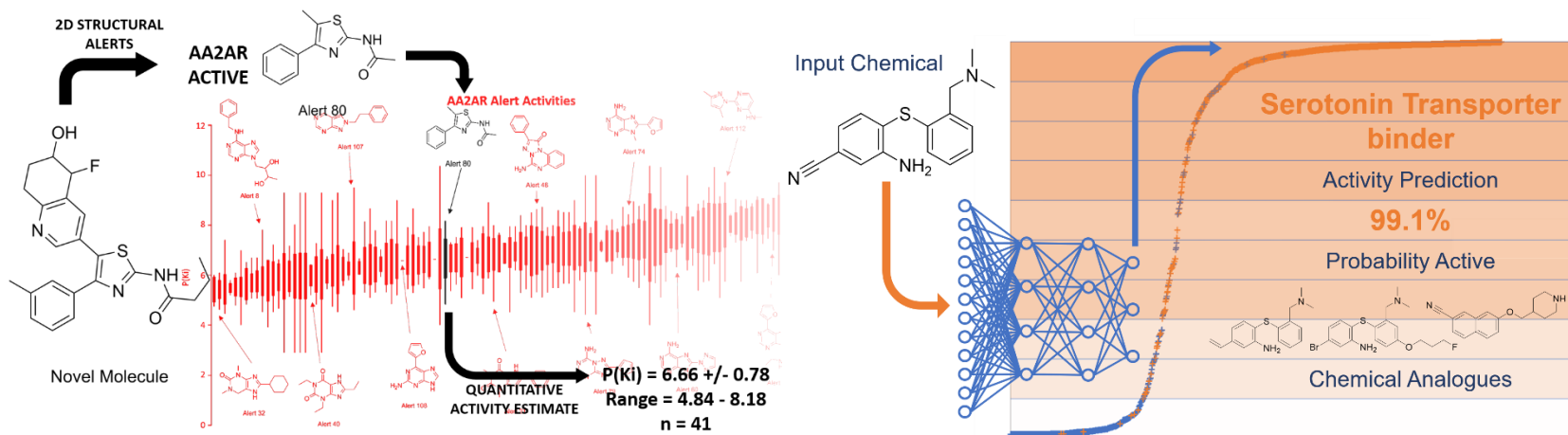
MRC

Toxicology
Unit



Can **BIG**
DATA

Replace Animal Testing?



Acknowledgements



- Professor Jonathan Goodman
- Professor Anne Willis
- Unilever
- Dr Paul Russell, Dr Predrag Kukic, Dr Steve Gutsell
& colleagues at SEAC, Unilever
- Dr Andrew Wedlake
- Elena Gelžinytė
- Dr Alistair Middleton
- The Centre for Molecular Informatics
- The MRC Toxicology Unit
- St. John's College



