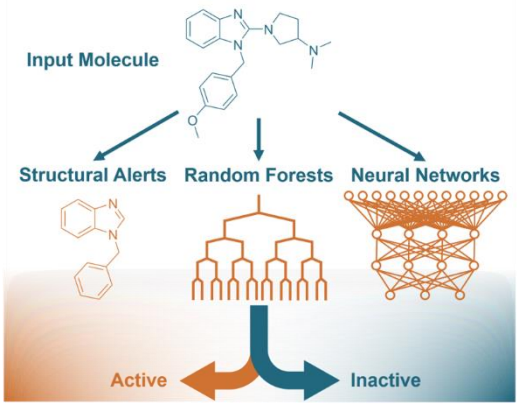
Use of chemical informatics, quantum chemistry modelling and artificial intelligence algorithms to predict molecular initiating events

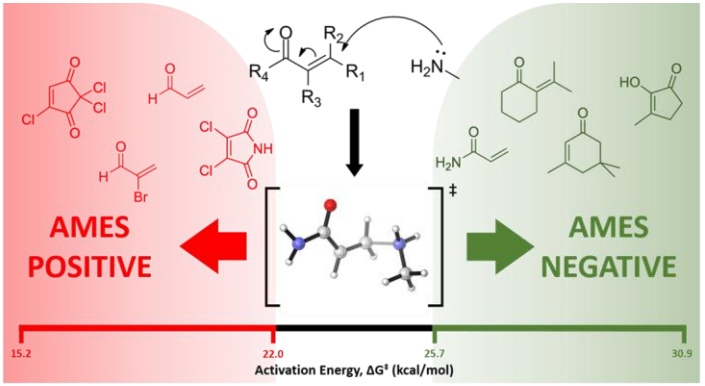
Timothy E H Allen – teha2@cam.ac.uk

AJ Wedlake, MN Grayson, AM Middleton, M Folia, M Baltazar, P Piechota, E Gelžinytė, JM Goodman, PJ Russell, P Kukic, S Gutsell

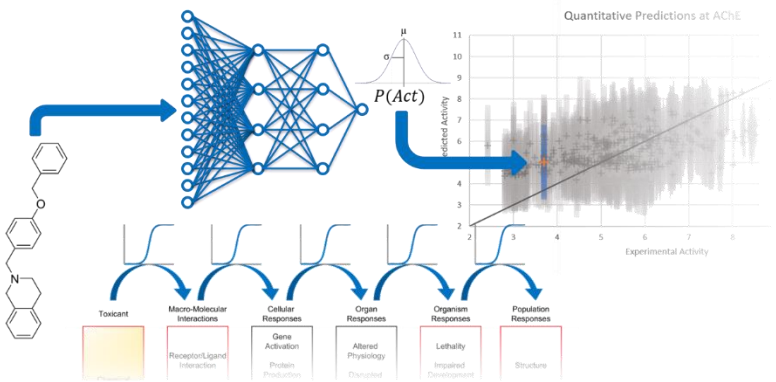
EUROTOX – 27th September 2021



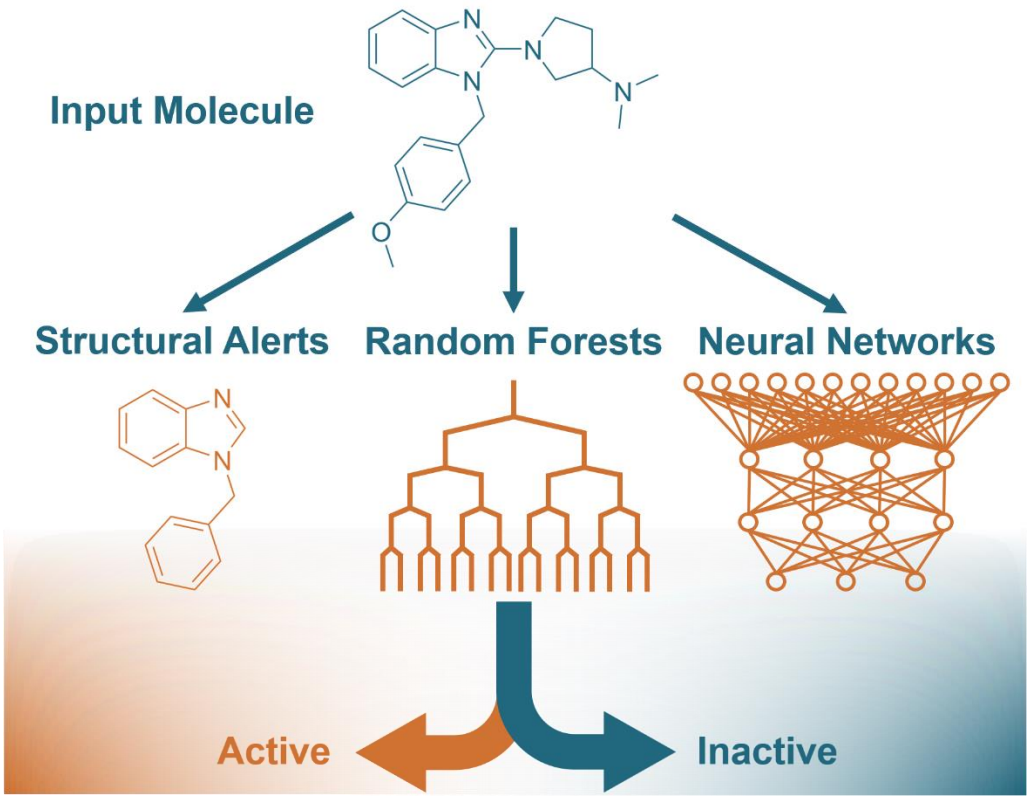
Mechanistic Toxicity Predictions



Quantum Chemistry Predictions

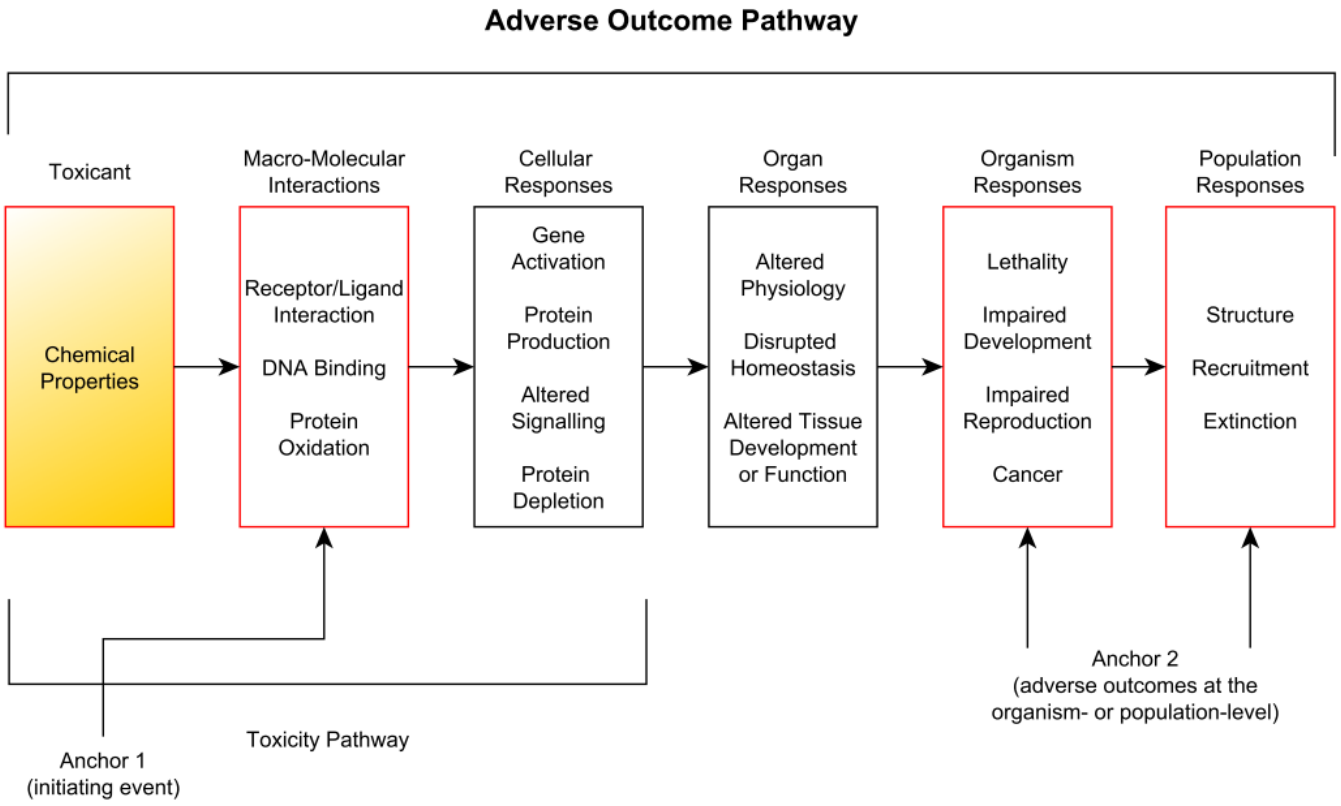


Bayesian Neural Networks



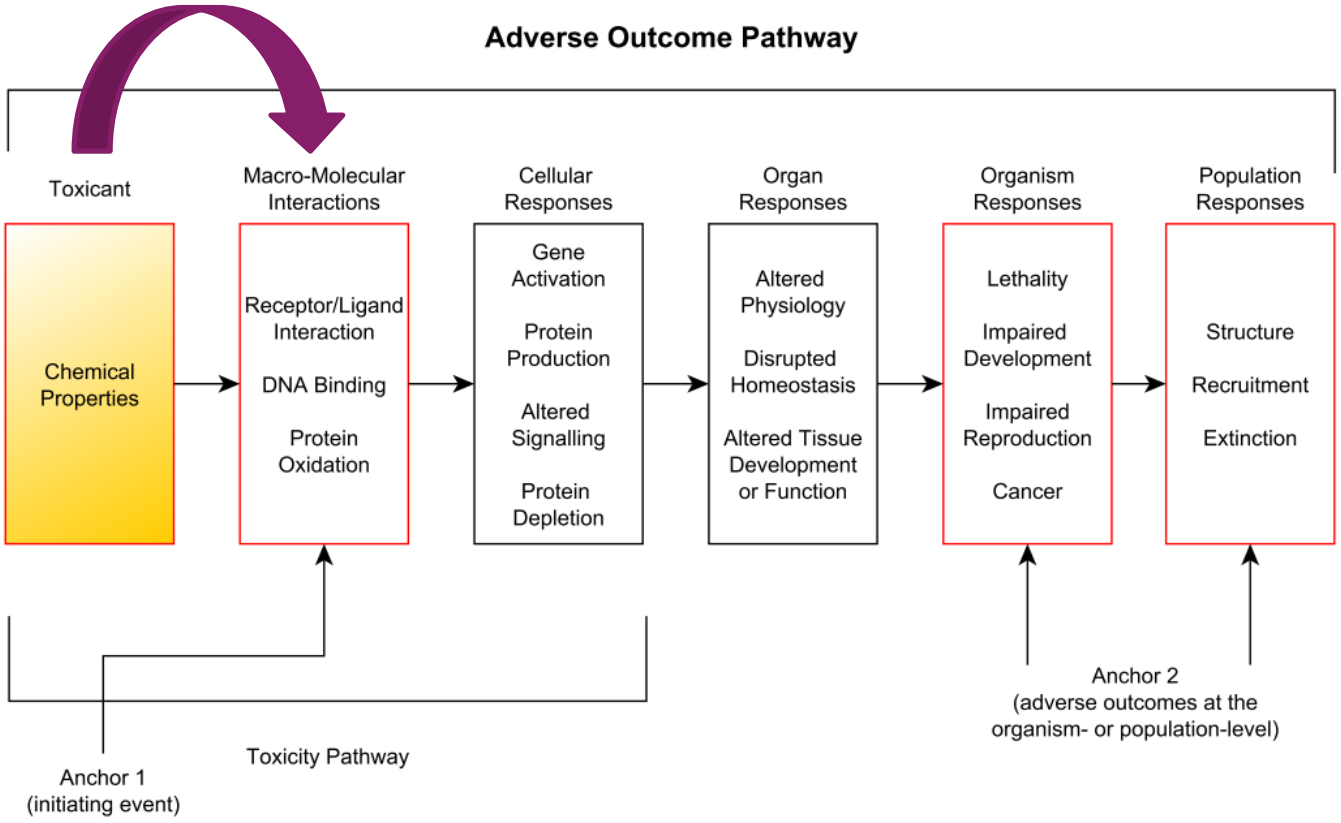
Mechanistic Toxicity Predictions

Adverse Outcome Pathway



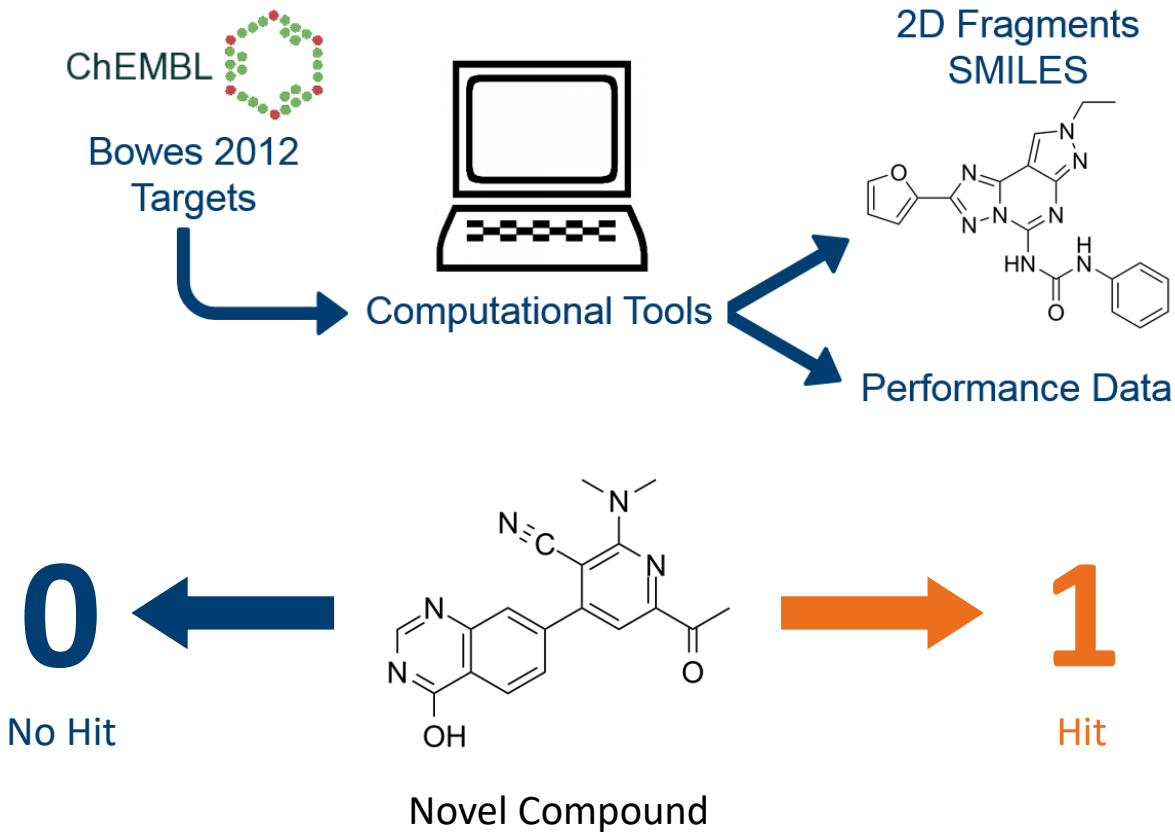
Ankley, G.T., *et al.* (**2010**) *Environ. Toxicol. Chem.*, 29; 730.

Adverse Outcome Pathway



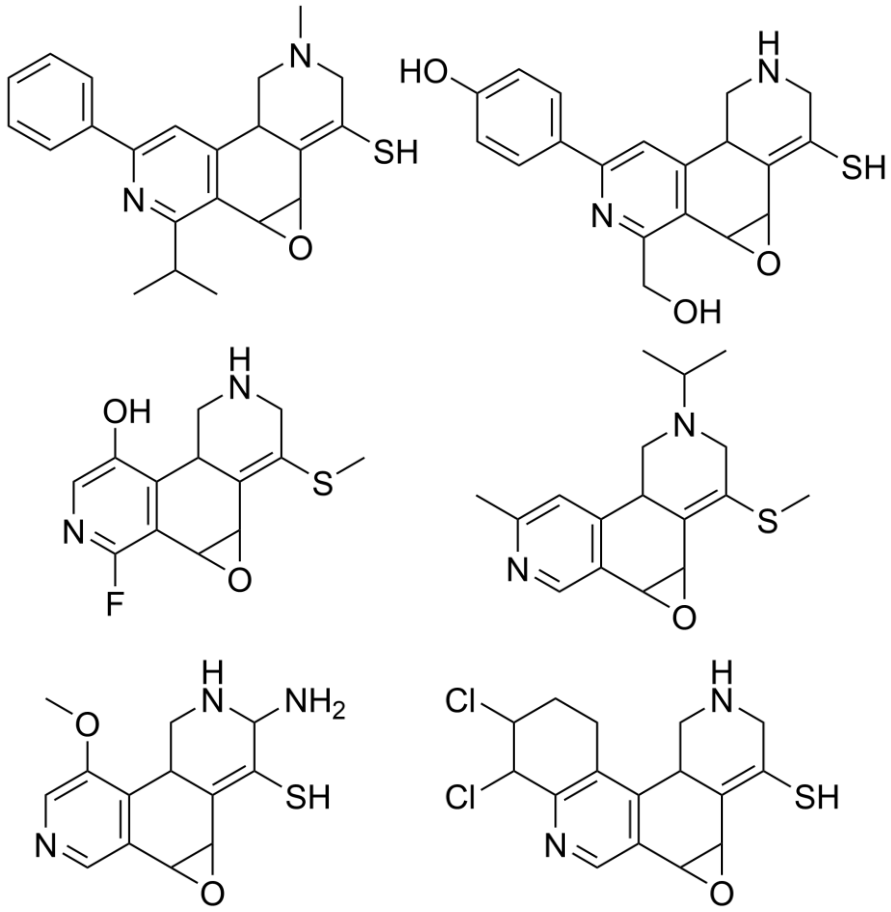
Ankley, G.T., *et al.* (**2010**) *Environ. Toxicol. Chem.*, 29; 730.

Structural Alerts

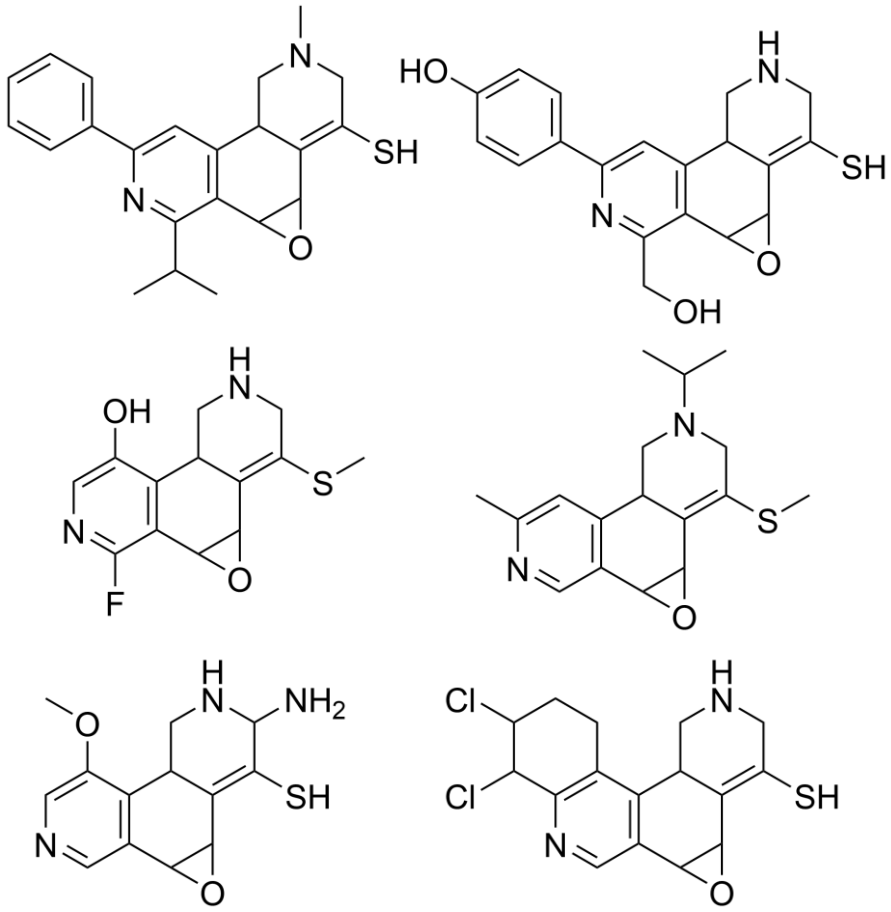


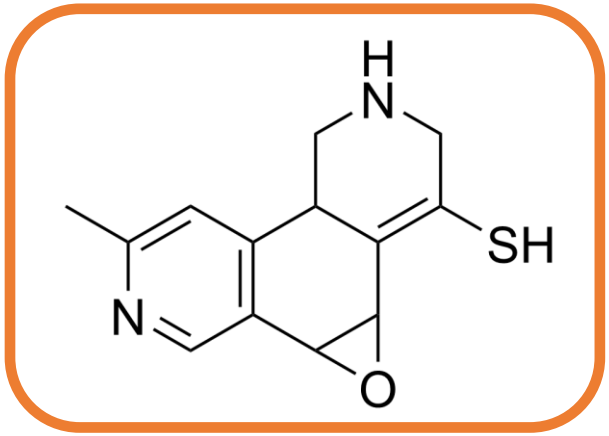
Allen, T.E.H. *et al.* (**2018**) *Toxicol. Sci.*, 165; 213.
Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.

Maximal Common Substructure



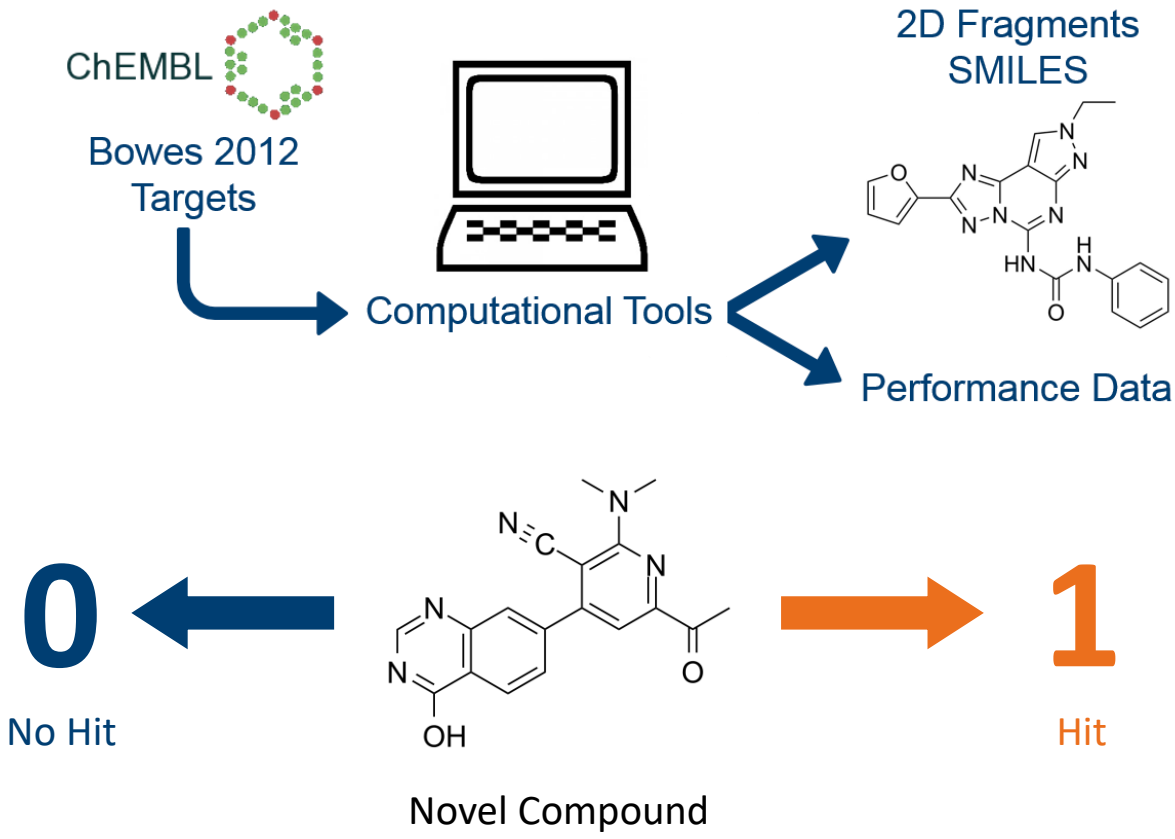
Maximal Common Substructure





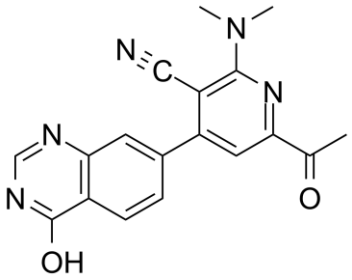
Present in 6 of 6 chemicals

Structural Alerts

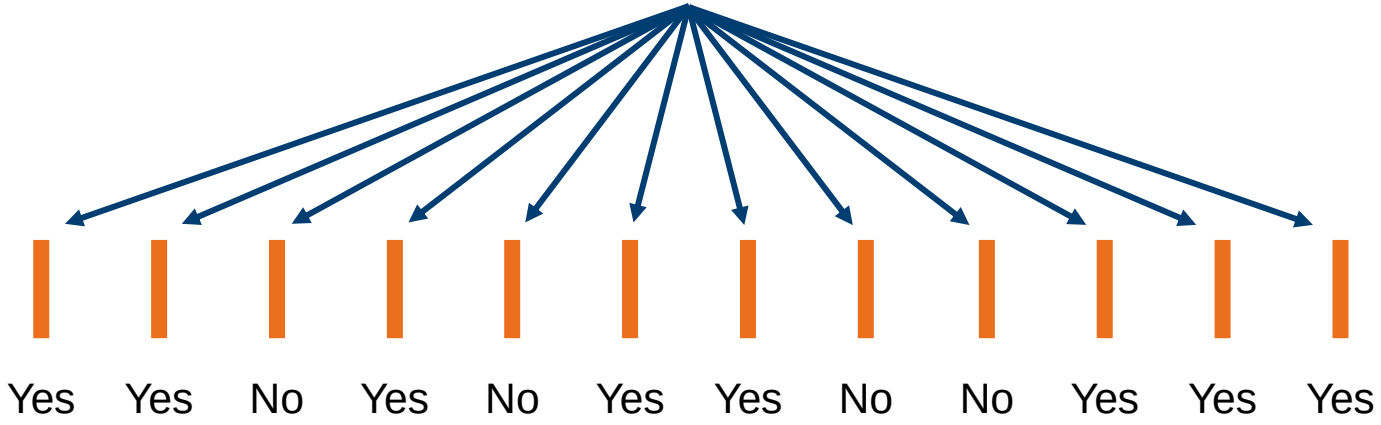


Allen, T.E.H. *et al.* (**2018**) *Toxicol. Sci.*, 165; 213.
Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.

Random Forest



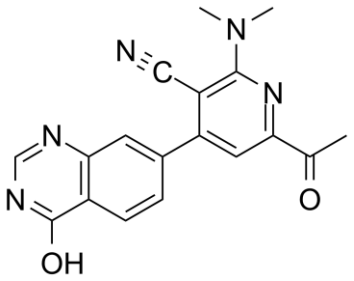
Novel Compound



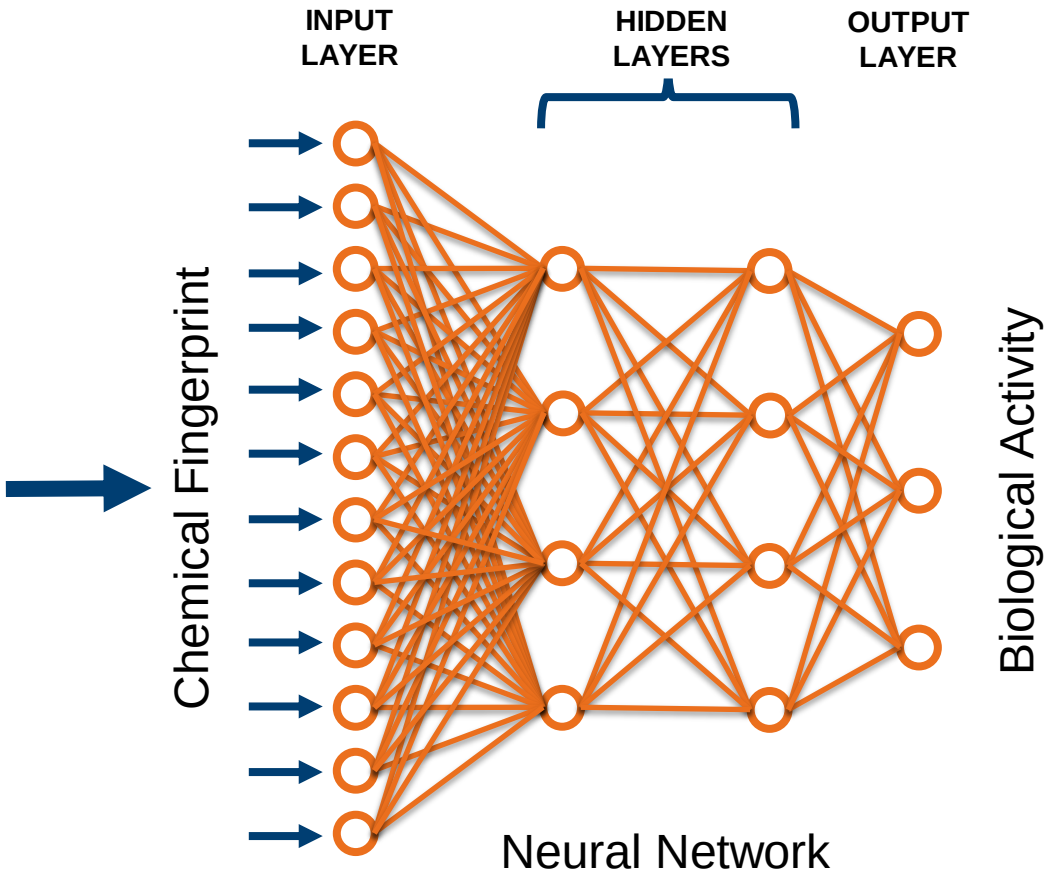
8/ 12 Trees predict Yes
Therefore overall prediction is Yes

Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.

Neural Network

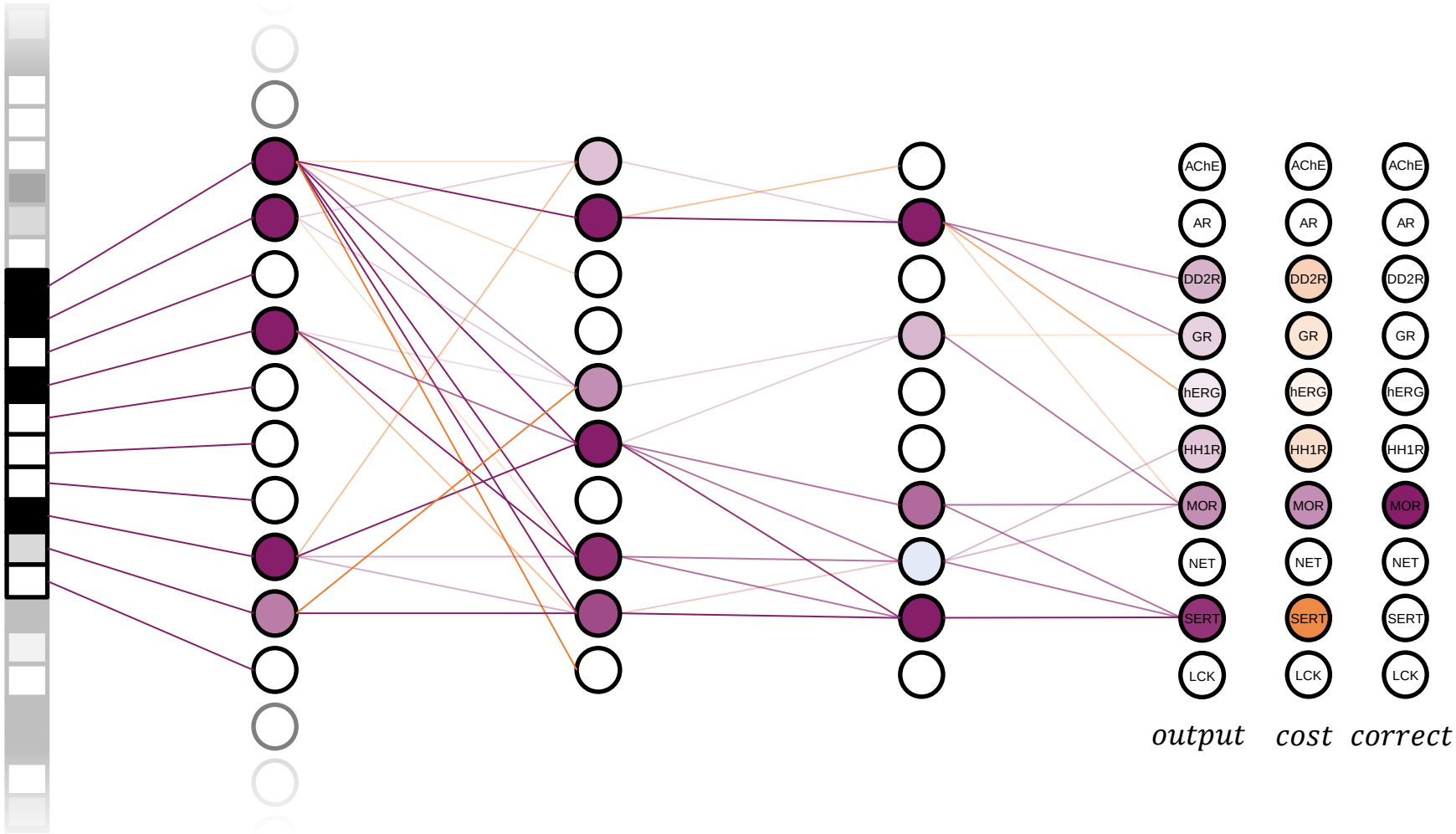


Novel Compound

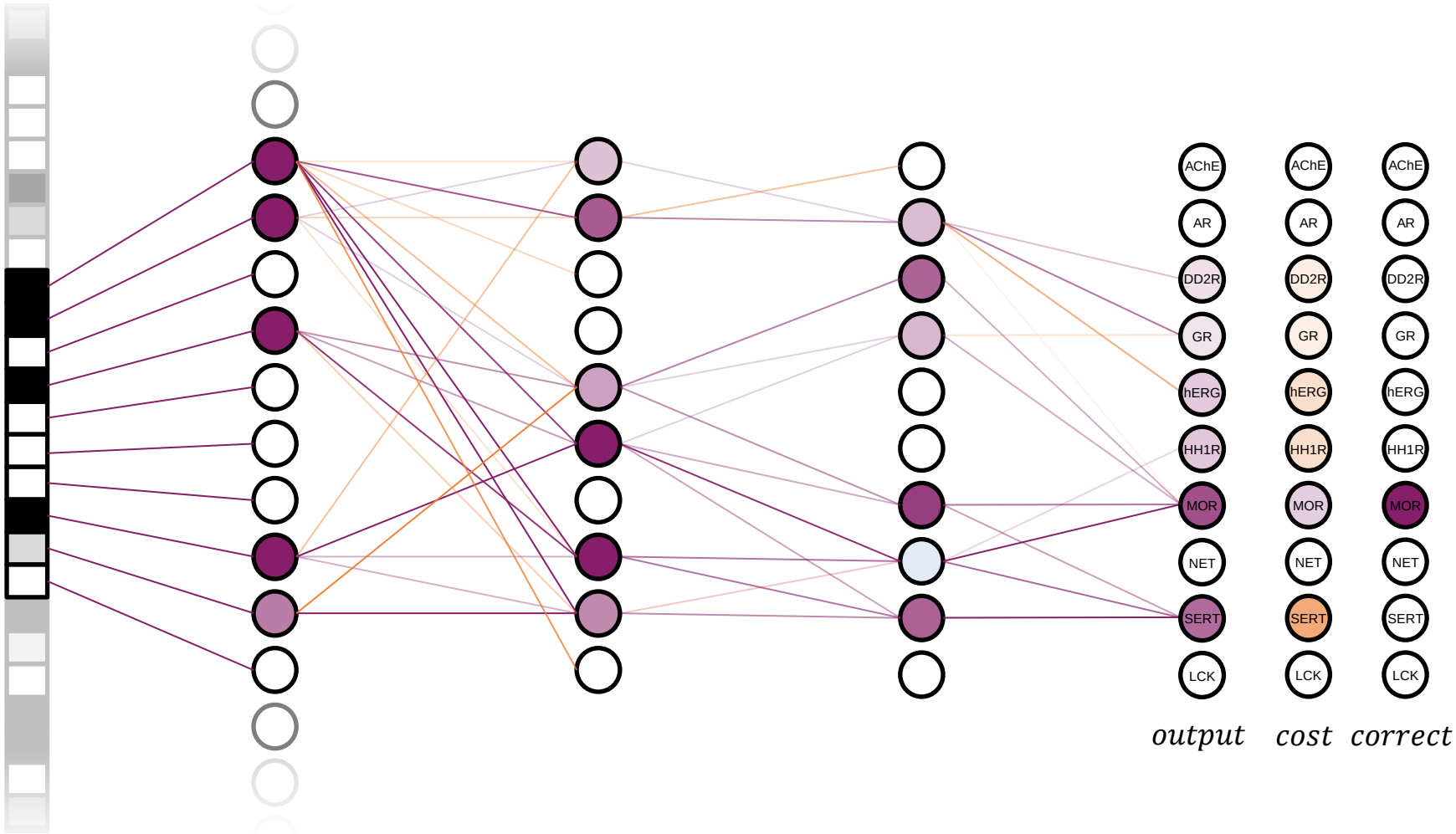


Allen, T.E.H. *et al.* (**2020**) *Chem. Sci.*, 11; 7335.

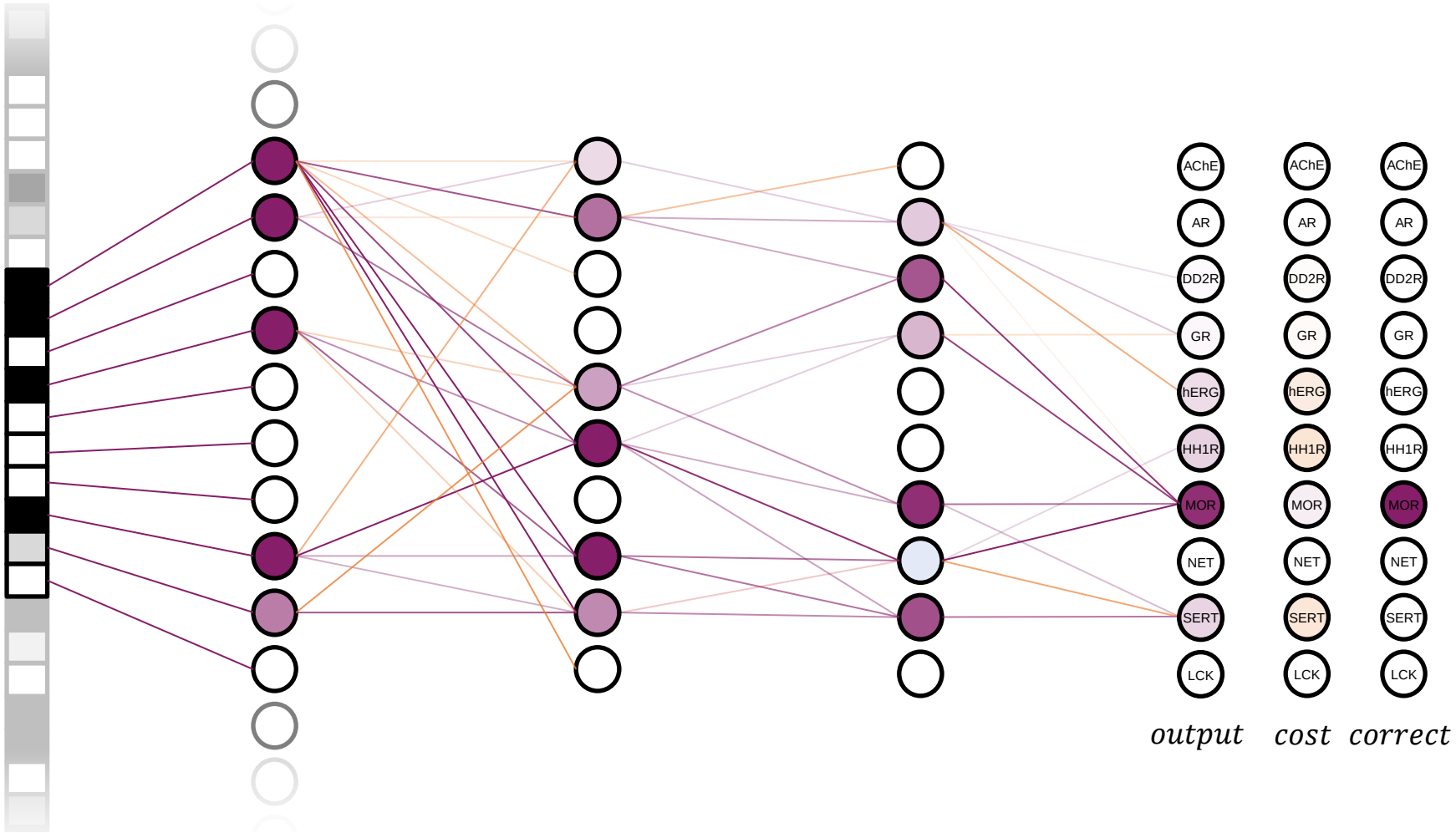
Neural Network



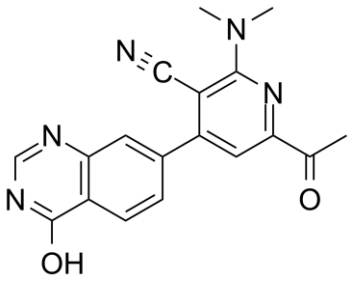
Neural Network



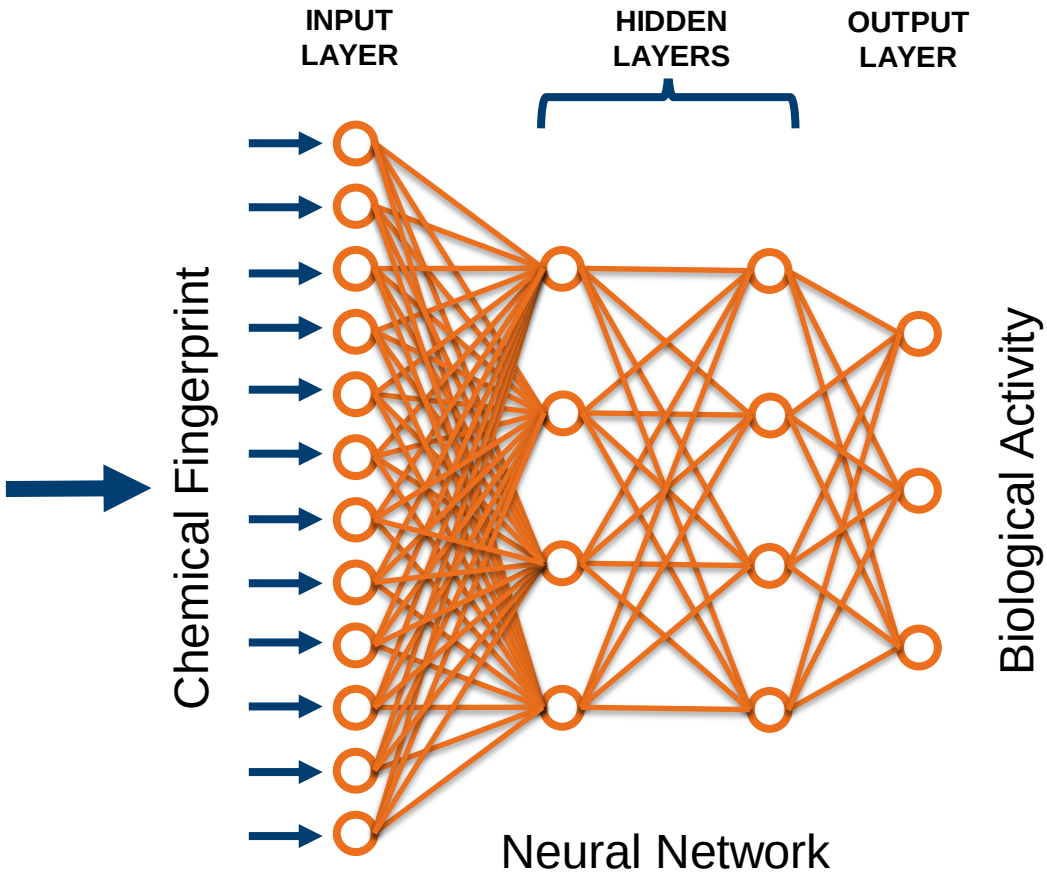
Neural Network



Neural Network



Novel Compound



Allen, T.E.H. *et al.* (**2020**) *Chem. Sci.*, 11; 7335.

Model Performance

		Held Out Test Data			
		SE	**SP**	**ACC**	**MCC**
Structural Alerts	**Average**	84.1	93.5	91.1	0.790
	SD	11.6	4.6	4.2	0.096
Random Forests	**Average**	89.0	90.4	92.2	0.815
	SD	11.6	8.1	4.0	0.091
Neural Networks	**Average**	87.8	93.6	92.8	0.832
	SD	10.4	5.9	4.0	0.089

SE *= sensitivity (percentage active chemicals correctly assigned)*
SP *= specificity (percentage negative chemicals correctly assigned)*
ACC *= overall quality (percentage of chemicals correctly assigned)*
MCC *= Matthews correlation coefficient (score from -1 to 1 with a higher score indicating a better model. scores account for imbalance in dataset)*

Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.
Allen, T.E.H. *et al.* (**2020**) *Chem. Sci.*, 11; 7335.

Consensus Modelling

MRC Toxicology Unit

Chemical Research in Toxicology

Article

Cite This: *Chem. Res. Toxicol.* 2020, 33, 388–401

pubs.acs.org/crt

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[‡]

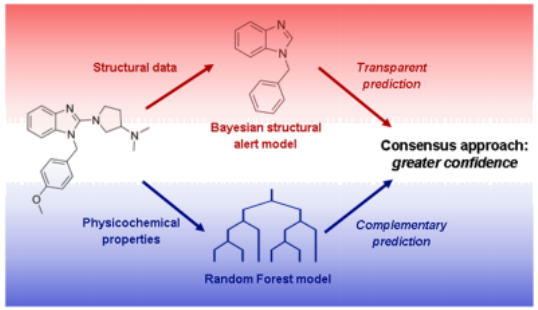
[†]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

[‡]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

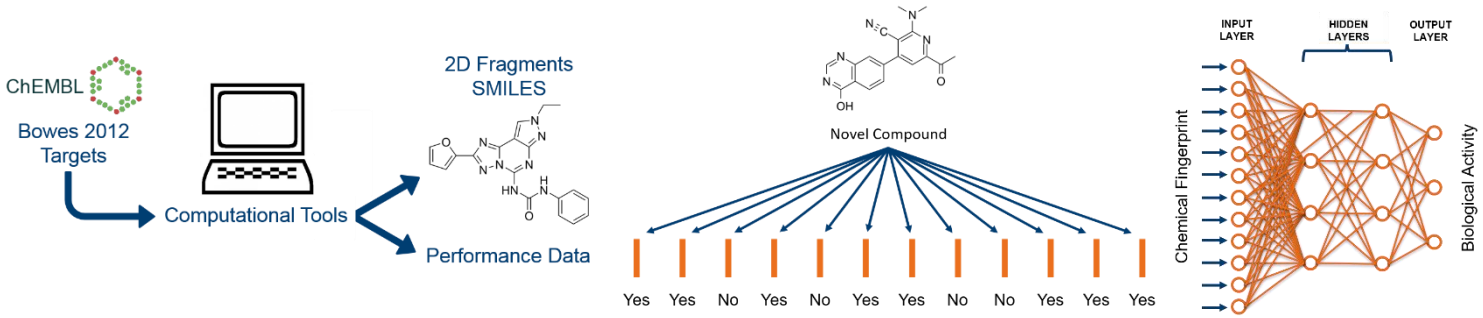
Supporting Information

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

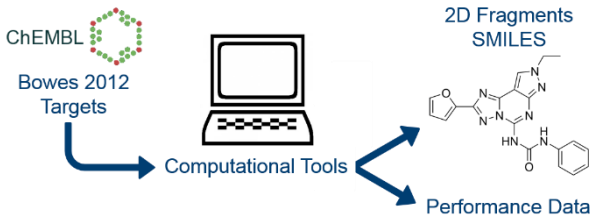


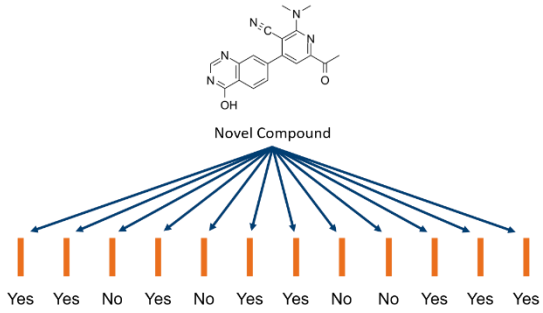
Wedlake, A.J., *et al*. (**2020**) *Chem. Res. Toxicol.*, 33; 388.

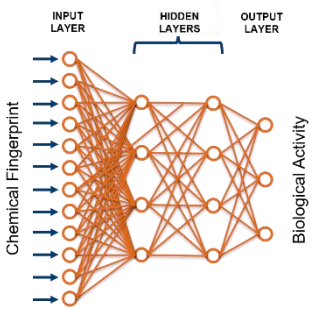
Combined Model



Combined Model





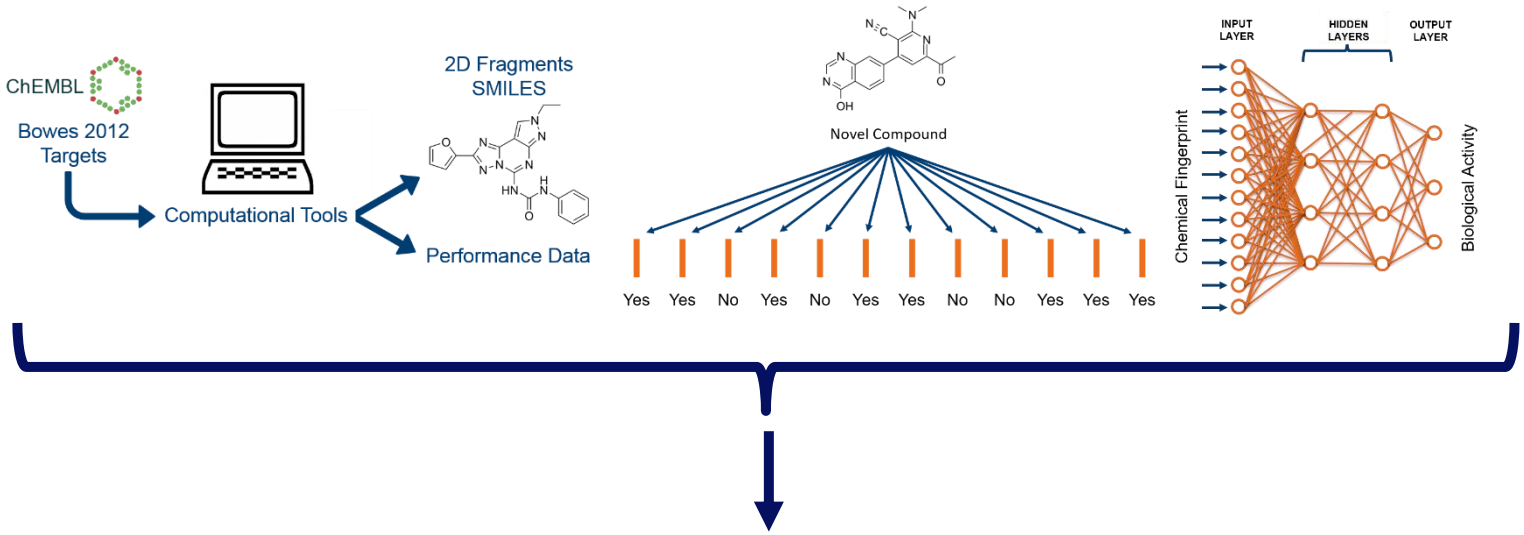


Transparent Prediction Algorithm

Use of Physicochemical Descriptors

Highest Quality Predictions

Combined Model



Consensus approach;
Increased confidence

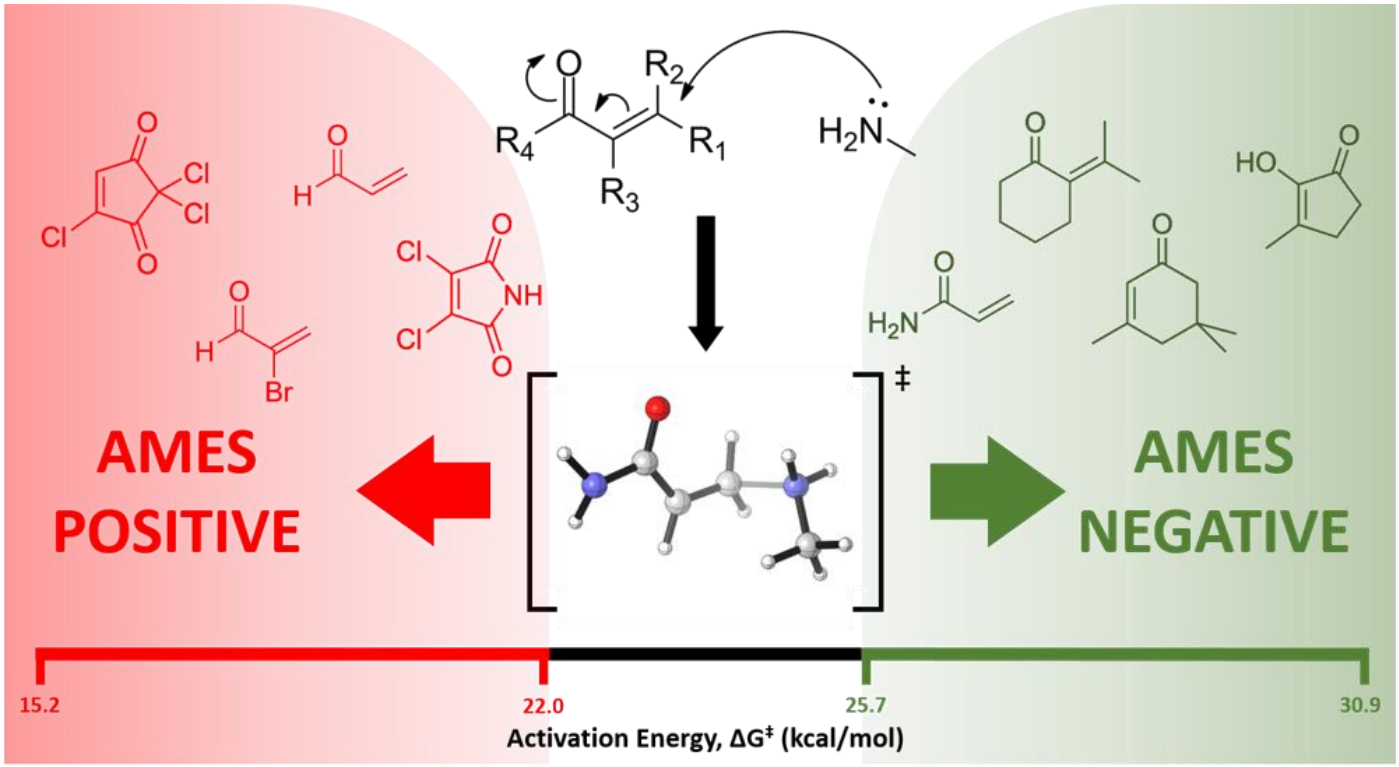
Combined Model

Method	SE	SP	ACC	MCC	% Unassigned
Majority Vote	90.7	90.5	92.3	0.827	
ΔSA	4.1	-0.4	2.1	0.045	
ΔRF	-0.8	3.7	1.0	0.023	
ΔNN	0.4	0.4	0.4	0.009	
Unanimous	92.8	93.8	94.9	0.882	9.6
ΔSA	6.2	2.9	4.7	0.100	
ΔRF	1.3	7.0	3.6	0.078	
ΔNN	2.5	3.7	3.0	0.064	

Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.
Allen, T.E.H. *et al.* (**2020**) *Chem. Sci.*, 11; 7335.

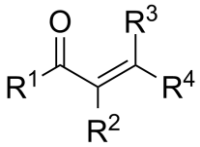
Conclusions

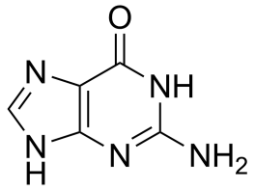
- MIEs make great targets for *in silico* toxicity predictions based on chemistry
- Structural alerts, random forests and neural networks have all been developed to make these predictions
- Using these algorithms together increases their performance and the confidence we can have in their predictions
- Combining algorithms helps us overcome their specific flaws

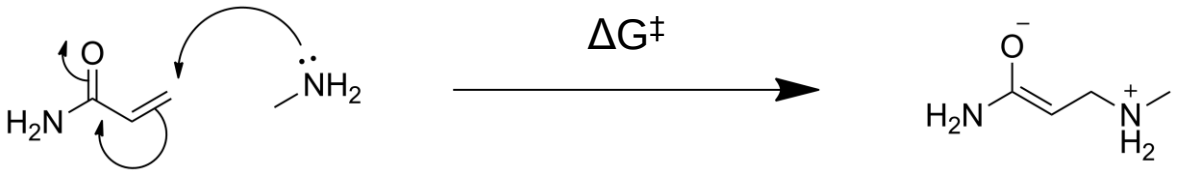


Quantum Chemistry Predictions

Modelling the Transition State

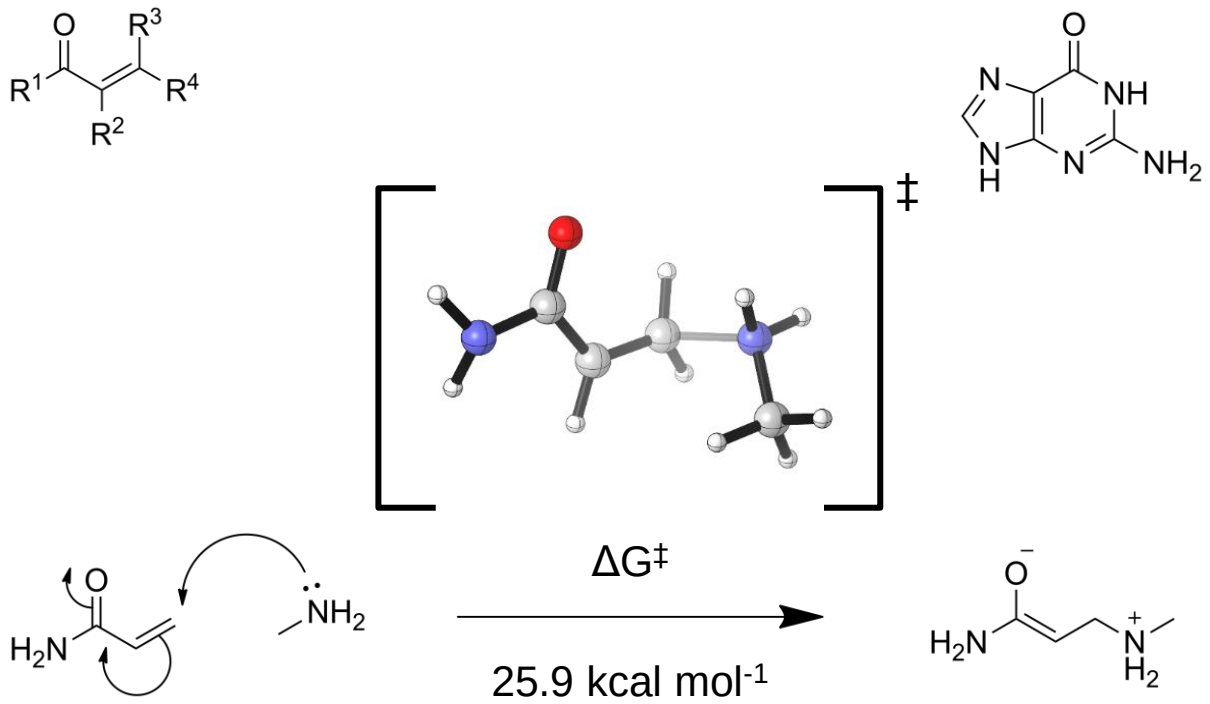






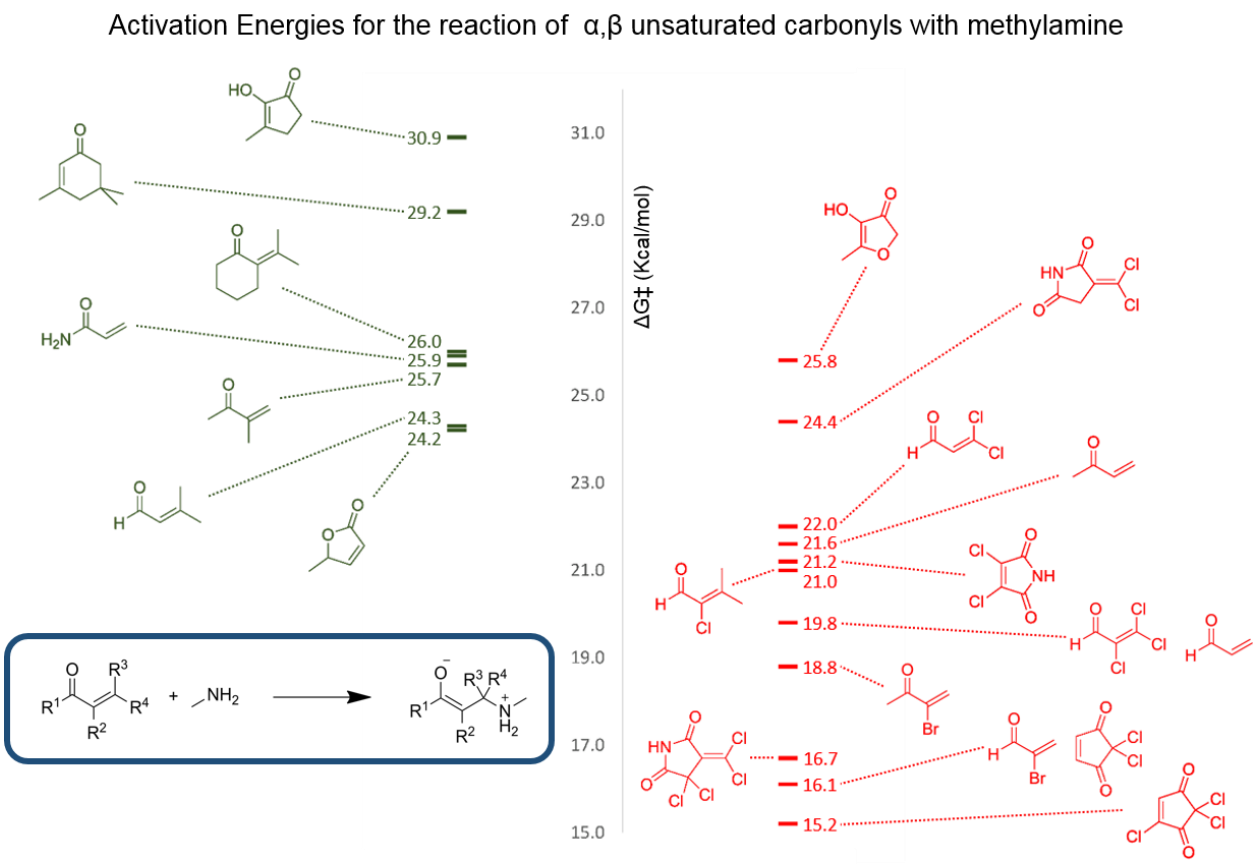
Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.,* 58; 1266.

Modelling the Transition State



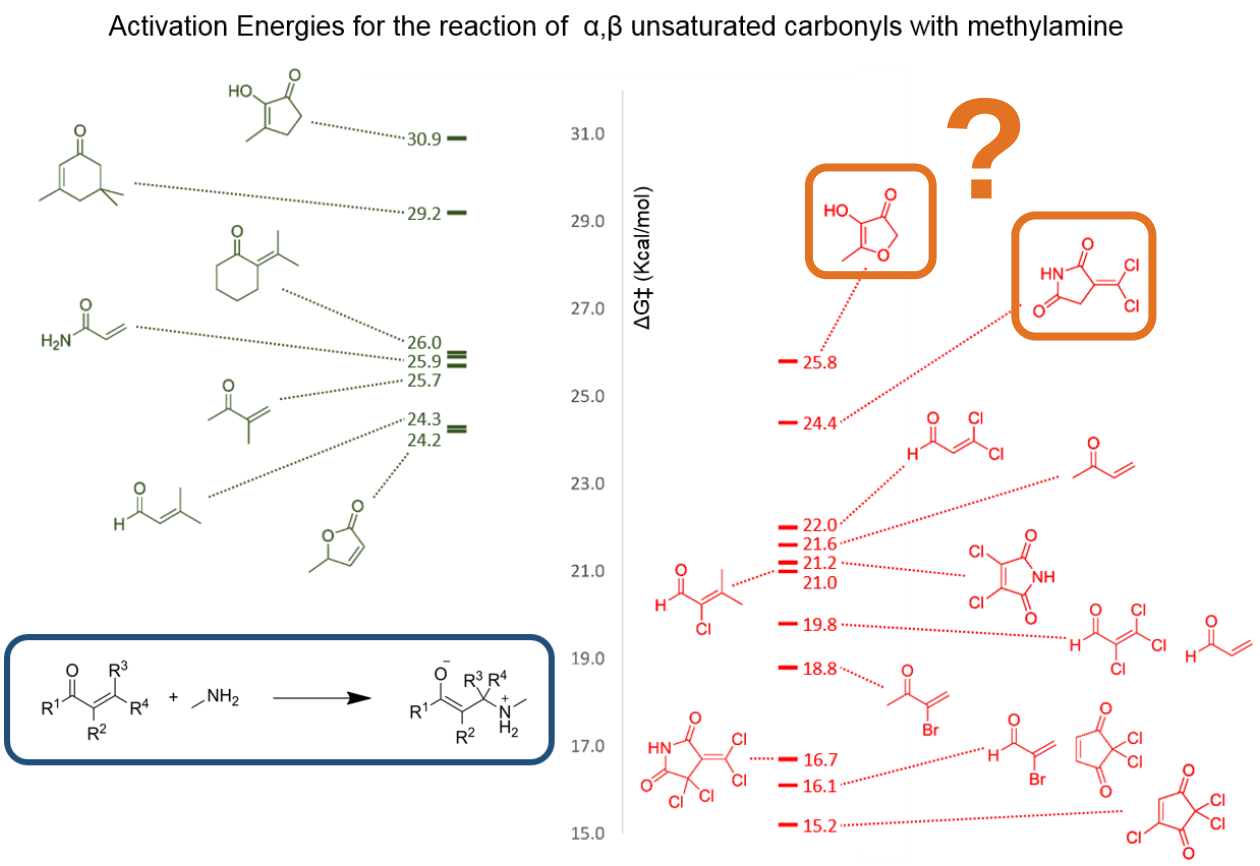
Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.,* 58; 1266.
DFT - Optimization: B3LYP, 6-31+G(d), iefpcm; SPE: M062X, def2tzvpp, iefpcm

Activation Free Energies



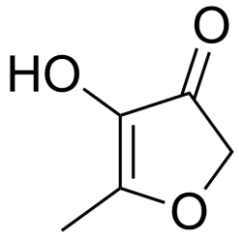
Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
DFT - Optimization: B3LYP, 6-31+G(d), iefpcm; SPE: M062X, def2tzvpp, iefpcm

Activation Free Energies



Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
DFT - Optimization: B3LYP, 6-31+G(d), iefpcm; SPE: M062X, def2tzvpp, iefpcm

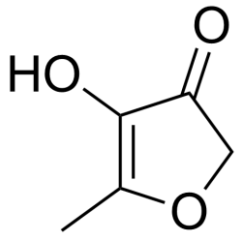
4-Hydroxy-5-methyl-3-furanone



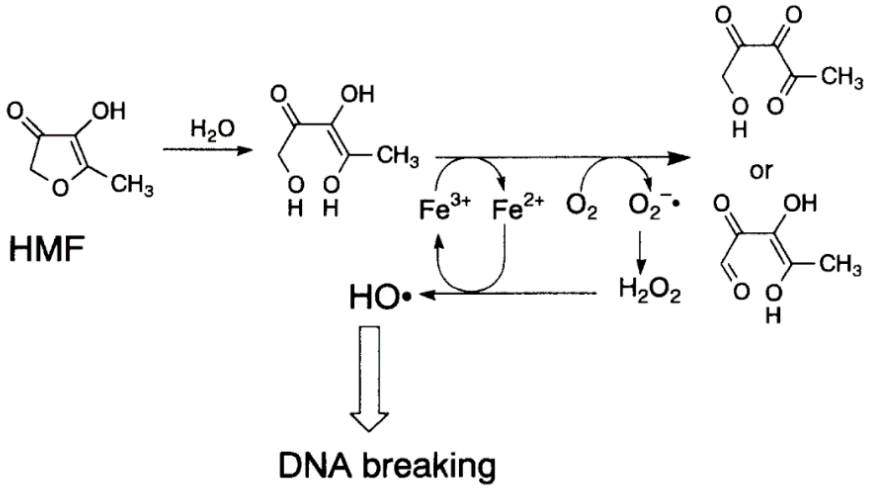
$\Delta G^{\ddagger} = 25.8$ kcal mol^{-1}

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.,* 58; 1266.

4-Hydroxy-5-methyl-3-furanone



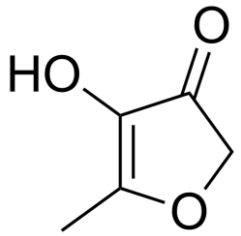
$\Delta G^{\ddagger} = 25.8$ kcal mol^{-1}



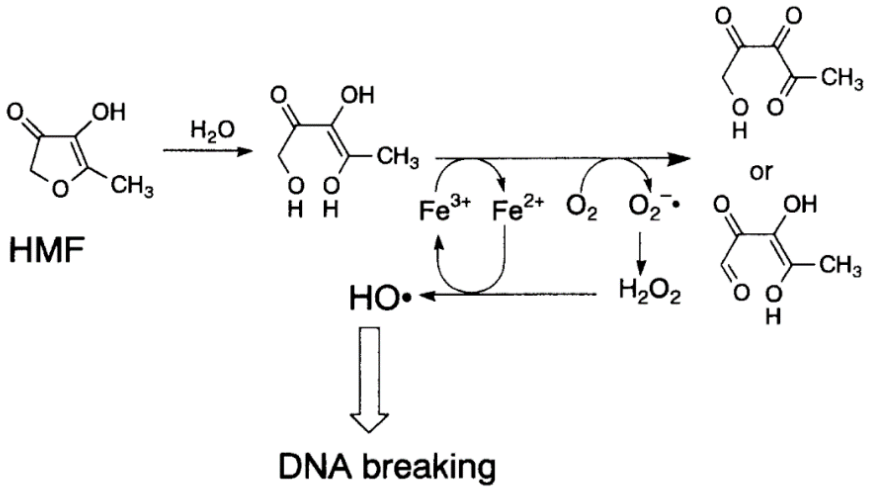
Scheme 2. Possible mechanisms for DNA breaking by HMF.

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
Hiramoto, K., *et al*. (**1996**) *Mut. Res.*, 359; 119.

4-Hydroxy-5-methyl-3-furanone



$\Delta G^{\ddagger} = 25.8$ kcal mol^{-1}

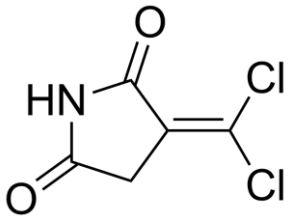


Scheme 2. Possible mechanisms for DNA breaking by HMF.

"HMF found to be mutagenic probably due to generation of active oxygen radicals"

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
Hiramoto, K., *et al*. (**1996**) *Mut. Res.*, 359; 119.

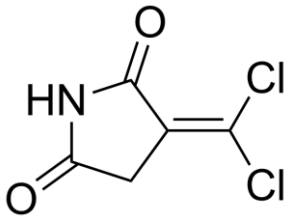
3-(Dichloromethylene)-2,5-pyrrolidinedione



$\Delta G^{\ddagger} = 24.4$ kcal mol^{-1}

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.,* 58; 1266.

3-(Dichloromethylene)-2,5-pyrrolidinedione



$\Delta G^{\ddagger} = 24.4$ kcal mol^{-1}

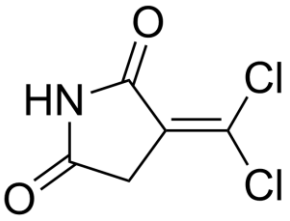
Table 2. Mutagenicity of Synthesized Chlorinated Imides in *S. typhimurium* TA100[a]

compound	dose, μg/plate (nmol/plate)	revertants/plate	molar mutagenicity [slope (rev/nmol)]	r^b
DCMI[c]	1 (6)	351; 365	28	0.975
	0.5 (3)	244; 254		
	0.125 (0.75)	233; 201		
solvent control	(0)	179; 178		
B[d]	125 (694)	280; 318; 325	0.24	0.971
	50 (278)	190; 230; 218		
	10 (56)	142; 151; 128		
D	10 (47)	451; 533; 513	7.7	0.987
	5 (23)	337; 344; 347		
	1 (4.7)	188; 153; 165		
E	0.1 (0.4)	687; 590; 623	1450[e]	0.95[e]
	0.025 (0.1)	260; 295; 278		
solvent control	(0)	152; 148; 148		
E[f]	0.1 (0.4)	806; 875; 920		
	0.05 (0.2)	451; 514; 515		
	0.025 (0.1)	383; 333; 361		
solvent control	(0)	193; 221; 158		

[a] The positive control, methyl methanesulfonate, gave >1200 revertants/plate in all assays at 0.4 μL/plate. [b] Correlation coefficient from linear regression. [c] Assay 1. [d] Assay 2. [e] Data combined from assays 2 and 3. [f] Assay 3.

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
Haddon, W.F., et al. (1996) J. Agric. Food Chem., 44; 256.

3-(Dichloromethylene)-2,5-pyrrolidinedione



$\Delta G^{\ddagger} = 24.4$ kcal mol^{-1}

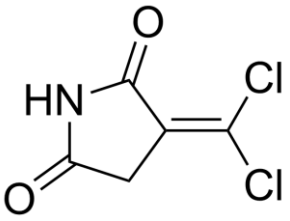
Table 2. Mutagenicity of Synthesized Chlorinated Imides in *S. typhimurium* TA100[a]

compound	dose, μg/plate (nmol/plate)	revertants/plate	molar mutagenicity [slope (rev/nmol)]	r^b
DCMI[c]	1 (6) 0.5 (3) 0.125 (0.75)	351; 365 244; 254 233; 201	28	0.975
solvent control	(0)	179; 178		
B[d]	125 (694) 50 (278) 10 (56)	280; 318; 325 190; 230; 218 142; 151; 128	0.24	0.971
D	10 (47) 5 (23) 1 (4.7)	451; 533; 513 337; 344; 347 188; 153; 165	7.7	0.987
E	0.1 (0.4) 0.025 (0.1)	687; 590; 623 260; 295; 278	1450[e]	0.95[e]
solvent control	(0)	152; 148; 148		
E[f]	0.1 (0.4) 0.05 (0.2) 0.025 (0.1)	806; 875; 920 451; 514; 515 383; 333; 361		
solvent control	(0)	193; 221; 158		

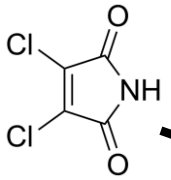
[a] The positive control, methyl methanesulfonate, gave >1200 revertants/plate in all assays at 0.4 μL/plate. [b] Correlation coefficient from linear regression. [c] Assay 1. [d] Assay 2. [e] Data combined from assays 2 and 3. [f] Assay 3.

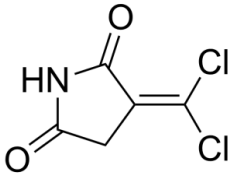
Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
Haddon, W.F., et al. (1996) J. Agric. Food Chem., 44; 256.

3-(Dichloromethylene)-2,5-pyrrolidinedione



$\Delta G^{\ddagger} = 24.4$ kcal mol^{-1}





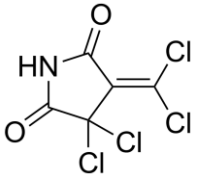


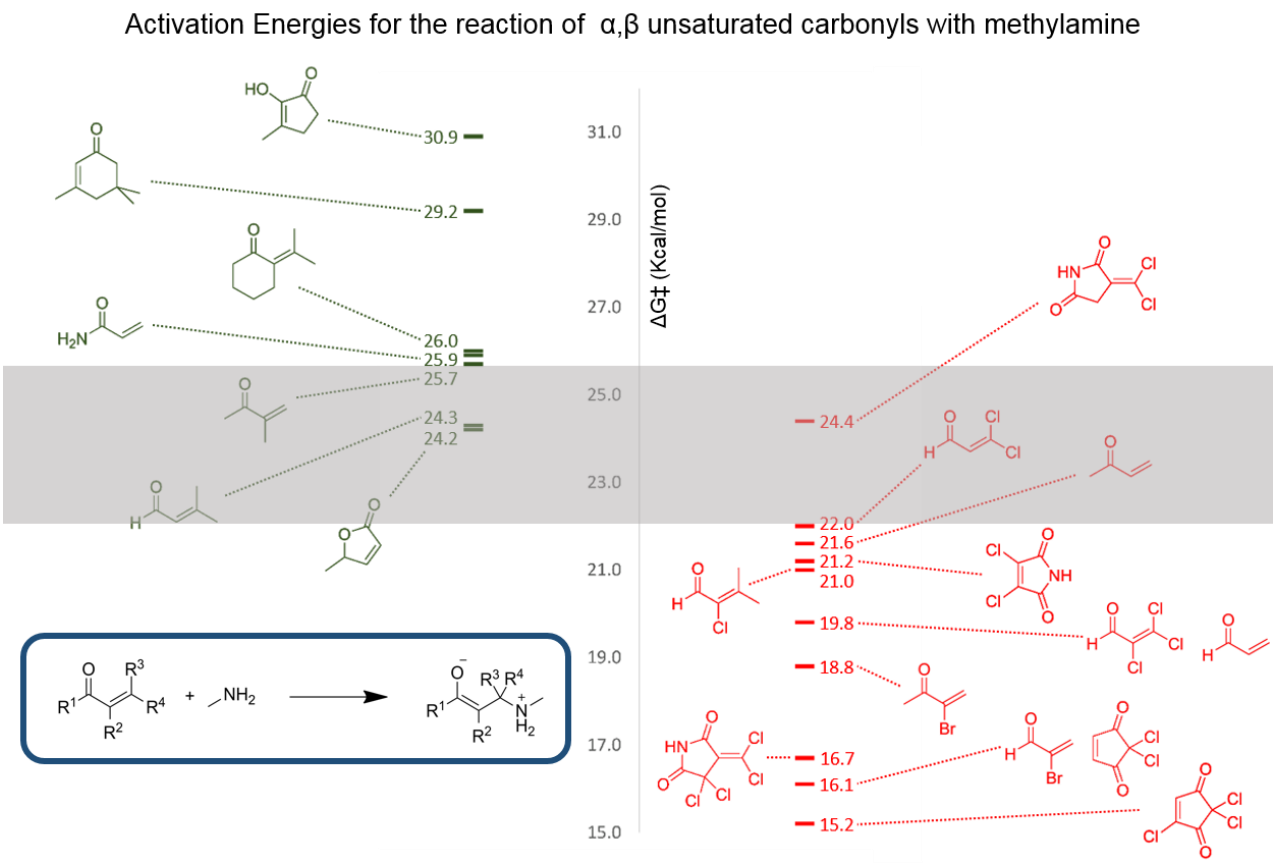
Table 2. Mutagenicity of Synthesized Chlorinated Imides in *S. typhimurium* TA100[a]

compound	dose, μg/plate (nmol/plate)	revertants/plate	molar mutagenicity [slope (rev/nmol)]	r^b	$\Delta G^{\ddagger}$ (kcal mol^{-1})
DCMI[c]	1 (6)	351; 365	28	0.975	21.2
	0.5 (3)	244; 254			
	0.125 (0.75)	233; 201			
solvent control	(0)	179; 178			
B[d]	125 (694)	280; 318; 325	0.24	0.971	24.4
	50 (278)	190; 230; 218			
	10 (56)	142; 151; 128			
D	10 (47)	451; 533; 513	7.7	0.987	
	5 (23)	337; 344; 347			
	1 (4.7)	188; 153; 165			
E	0.1 (0.4)	687; 590; 623	1450[e]	0.95[e]	16.1
	0.025 (0.1)	260; 295; 278			
solvent control	(0)	152; 148; 148			
E[f]	0.1 (0.4)	806; 875; 920			
	0.05 (0.2)	451; 514; 515			
	0.025 (0.1)	383; 333; 361			
solvent control	(0)	193; 221; 158			

[a] The positive control, methyl methanesulfonate, gave >1200 revertants/plate in all assays at 0.4 μL/plate. [b] Correlation coefficient from linear regression. [c] Assay 1. [d] Assay 2. [e] Data combined from assays 2 and 3. [f] Assay 3.

Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
Haddon, W.F., et al. (1996) J. Agric. Food Chem., 44; 256.

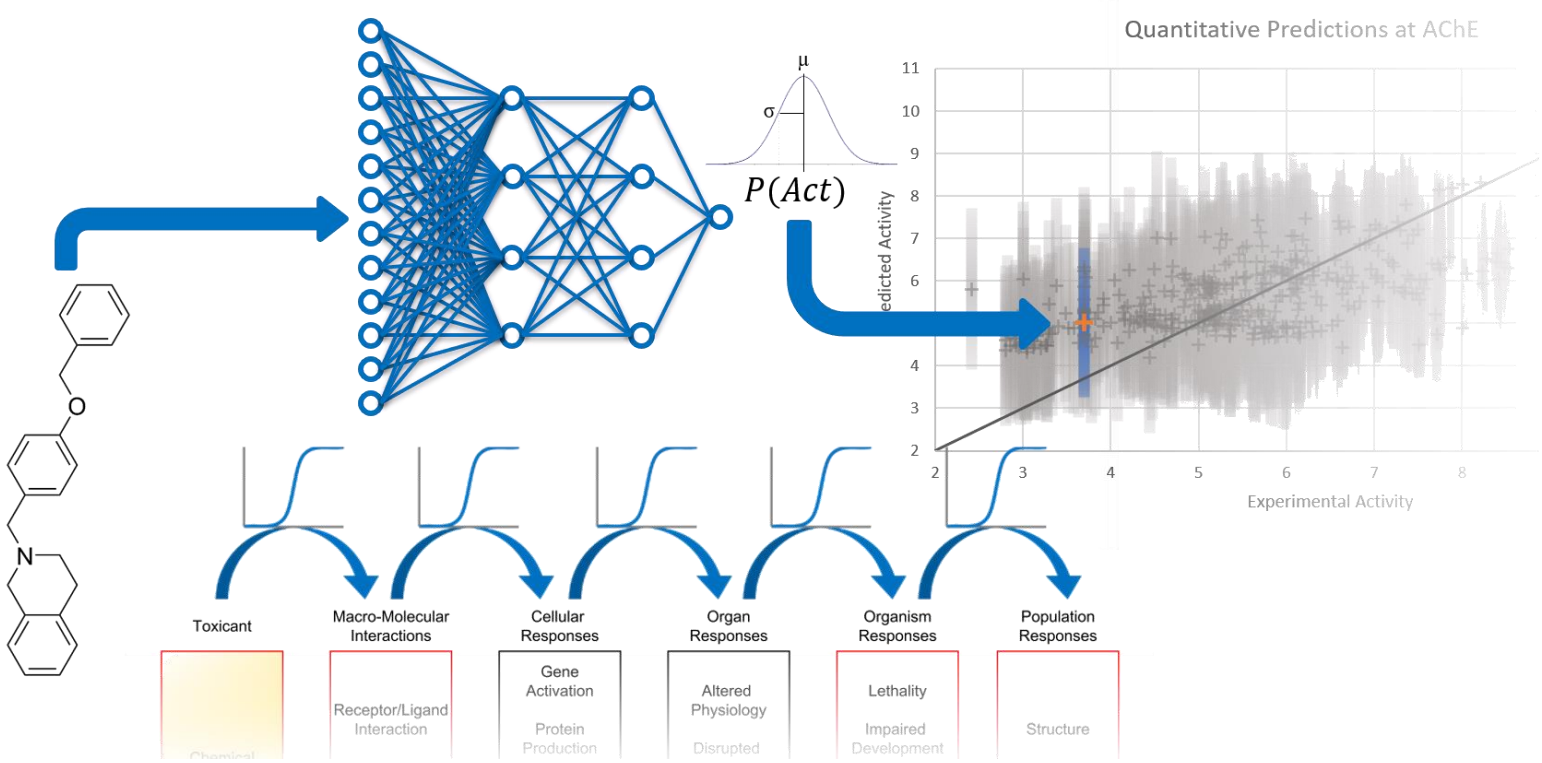
The Final Picture



Allen, T.E.H. *et al*. (**2018**) *J. Chem. Inf. Model.*, 58; 1266.
DFT - Optimization: B3LYP, 6-31+G(d), iefpcm; SPE: M062X, def2tzvpp, iefpcm

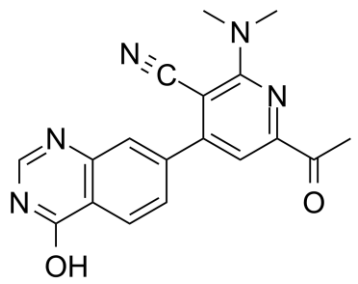
Conclusions

- DFT calculations provide a window into the molecular interactions driving some Ames positive results
- DFT allows direct modelling of the MIE in this case, and may also be useful for other covalent bond forming MIEs
- This approach helps increase understanding in how and why specific molecules covalently modify DNA

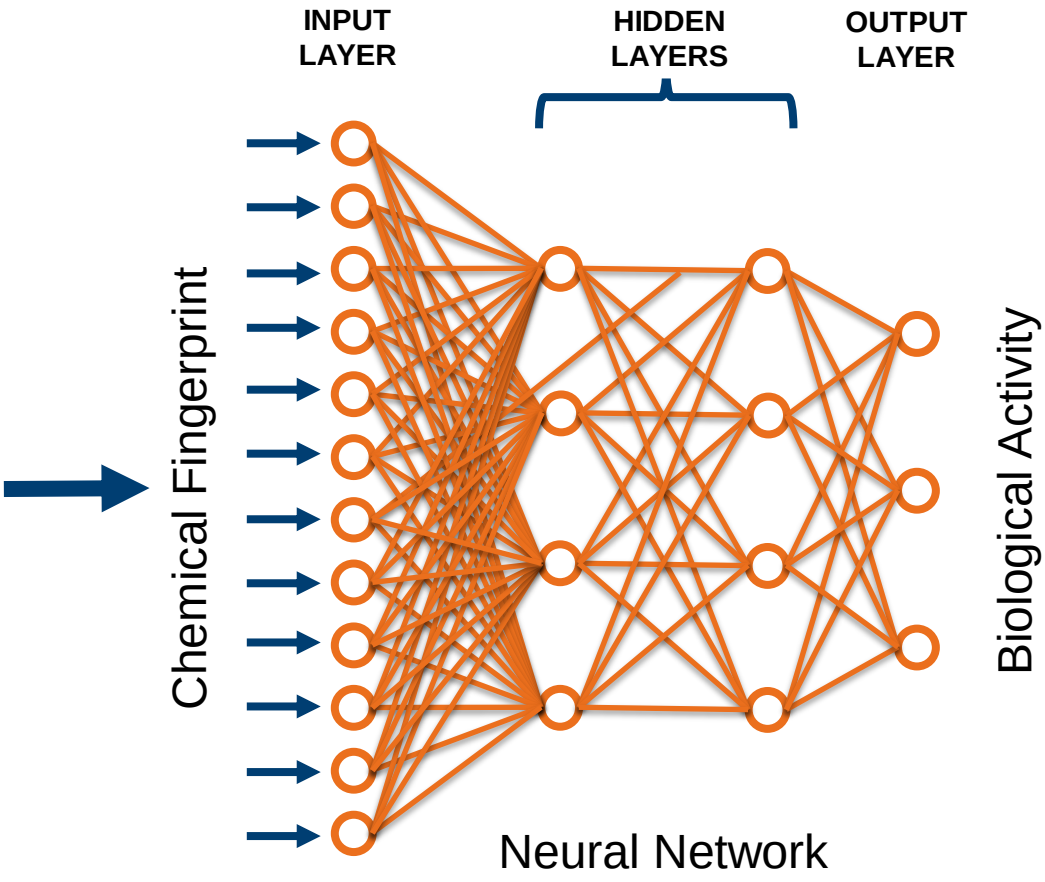


Bayesian Neural Networks

Quantitative Neural Network?

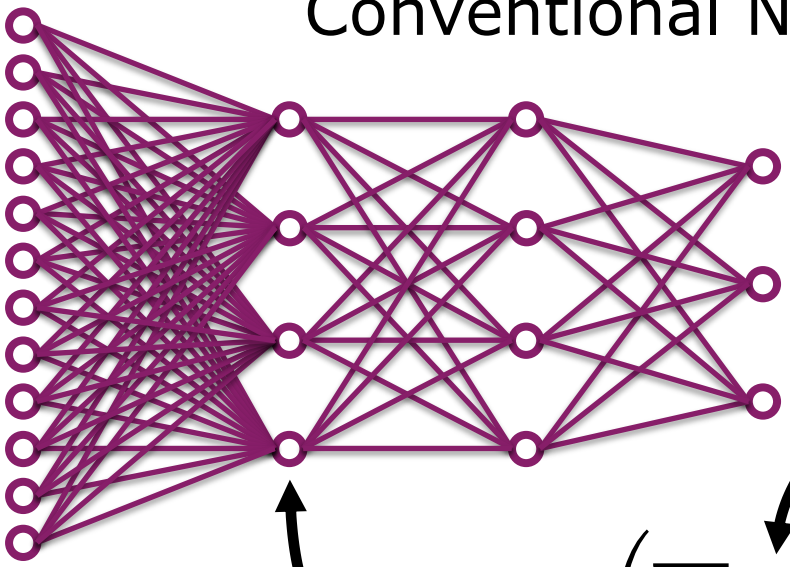


Novel Compound



Conventional Learning

Conventional Neural Network



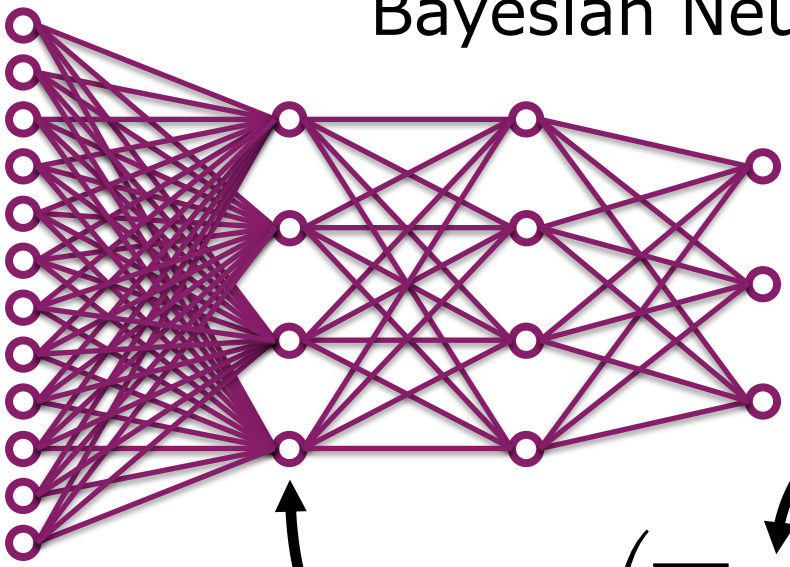
Weights and biases as fixed values

Therefore predictions are fixed values

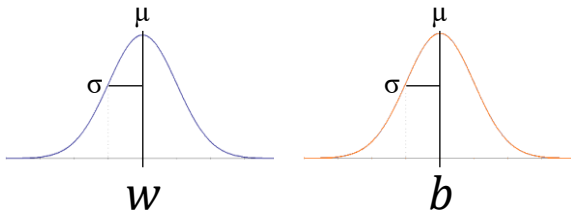
$$f(x) = K\left(\left(\sum_i w_i(x_i)\right) + b\right)$$

Bayesian Learning

Bayesian Neural Network



Weights and biases defined as probability distributions



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

Gives predictions meaningful errors

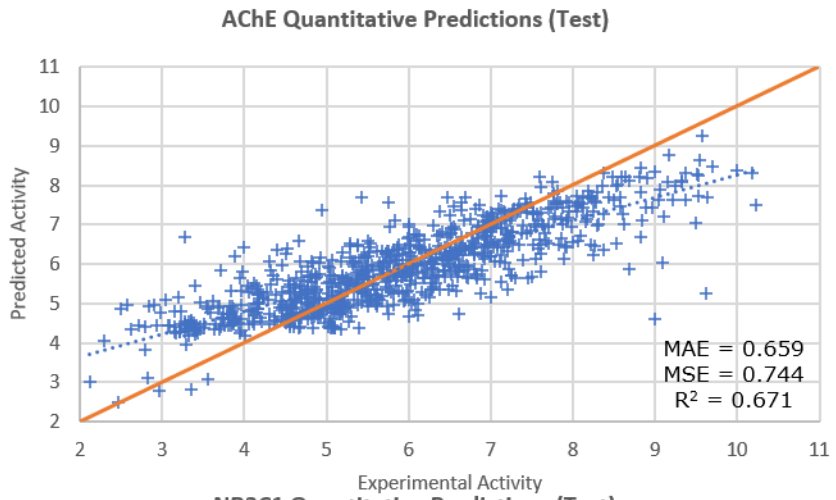
Best Models

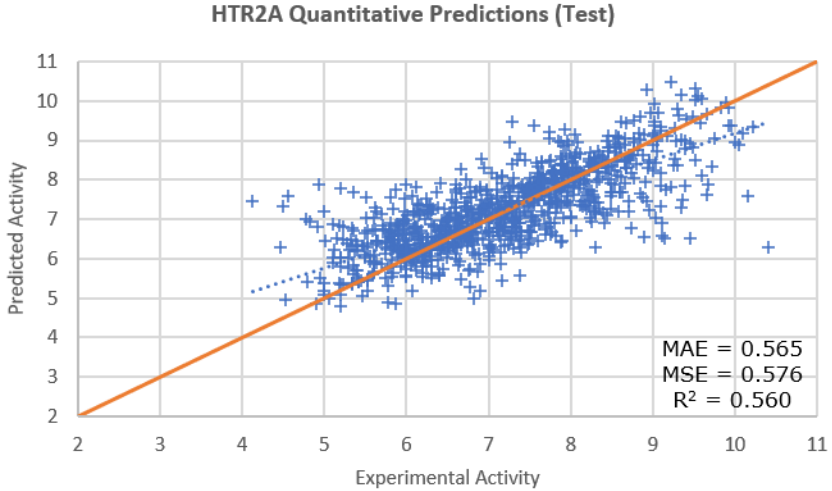
	Train MAE	Valid. MAE	Test MAE	Ext. Val. MAE
AVERAGE	0.487	0.613	0.621	0.943
SD	0.030	0.051	0.051	0.223

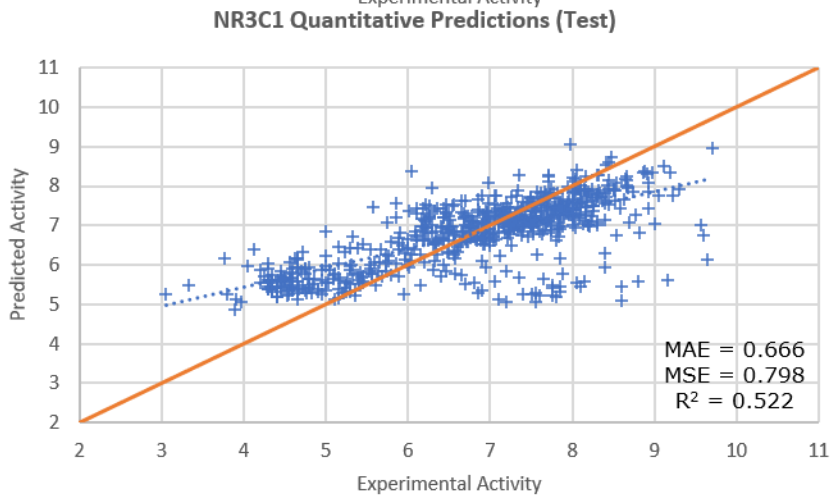
	Train R^2	Valid. R^2	Test R^2	Ext. Val. R^2
AVERAGE	0.743	0.586	0.572	0.128
SD	0.067	0.089	0.094	0.437

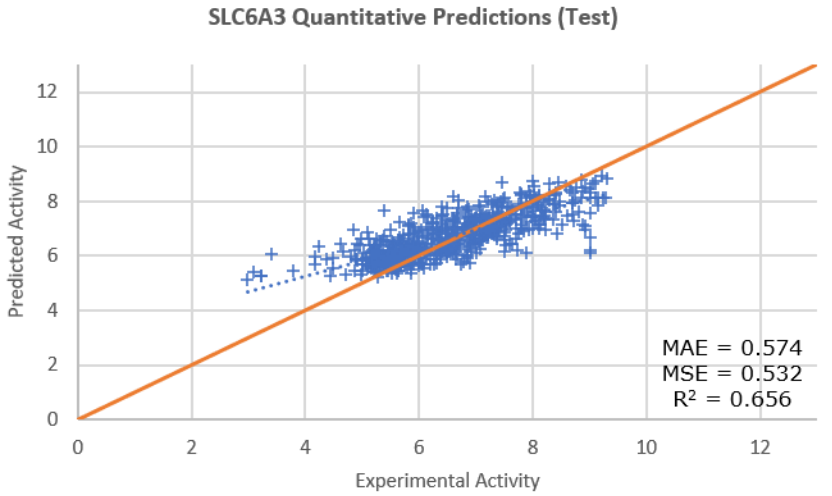
MAE *= Mean Absolute Error*
SD *= Standard Deviation*
R^2 *= Coefficient of Determination for Linear Correlation*

Some Models (Test Set)

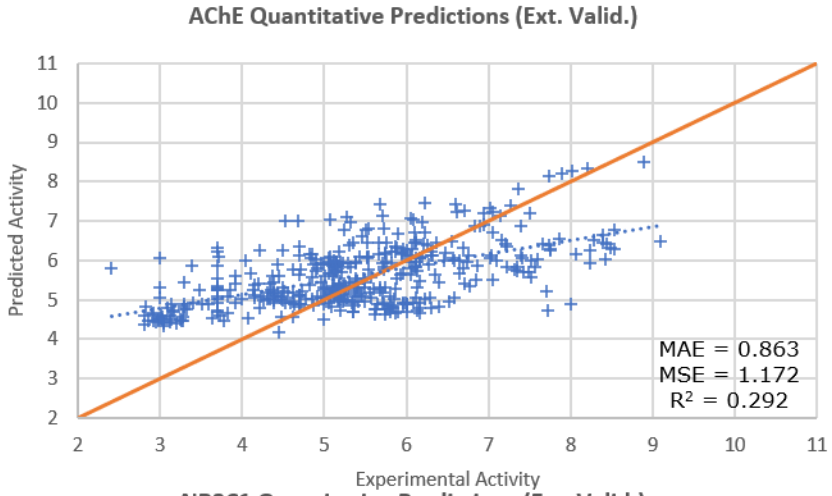


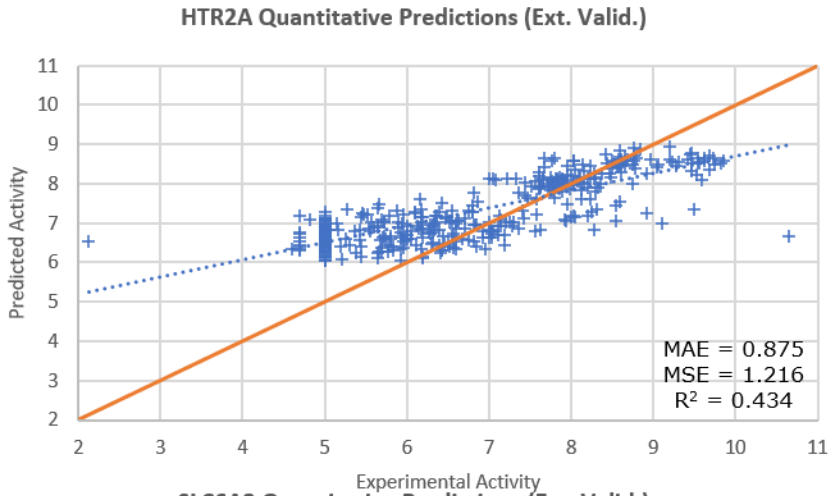


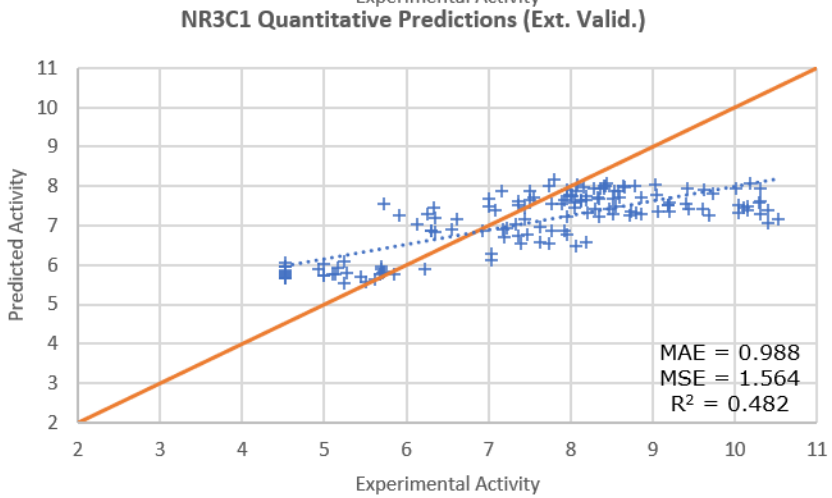


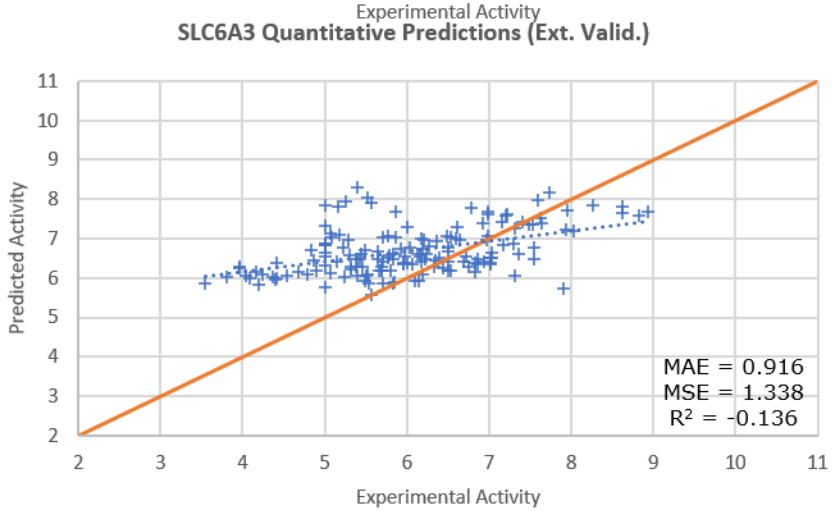


Some Models (Ext. Valid. Set)

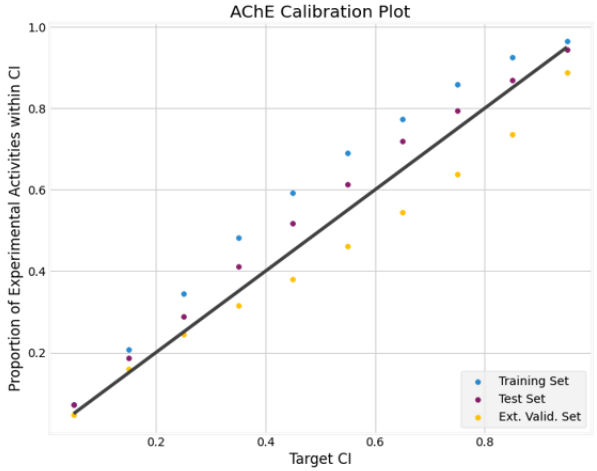


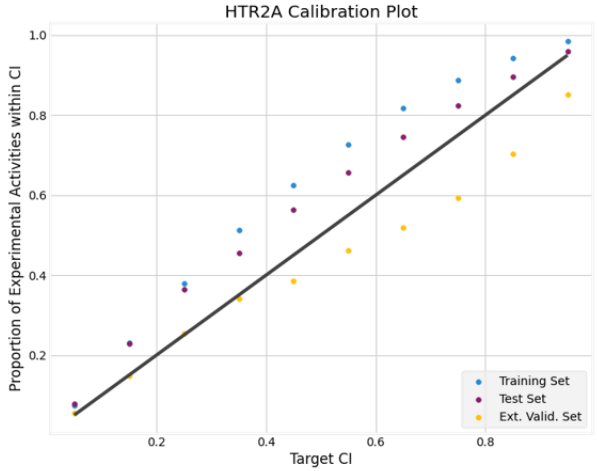


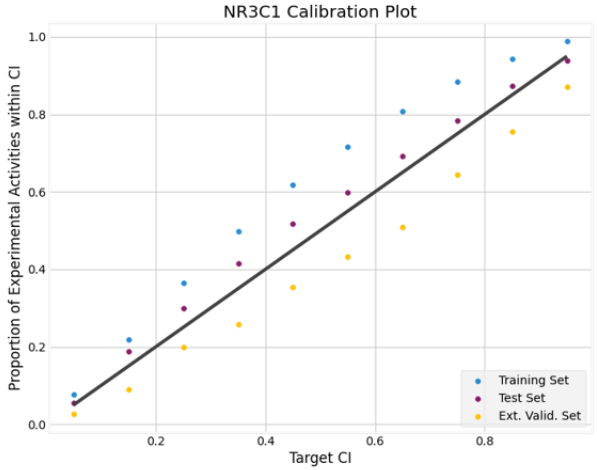


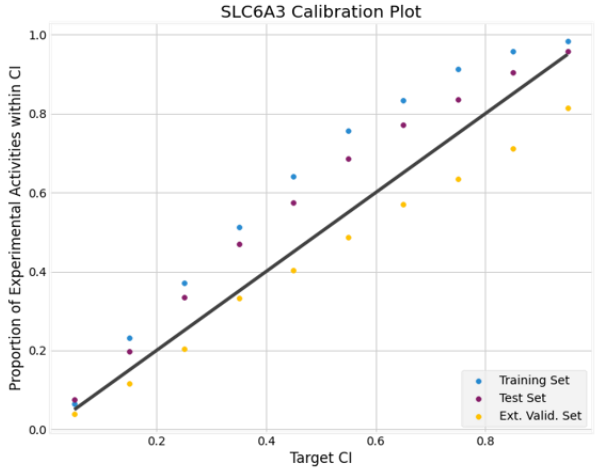


Calibration of Uncertainties

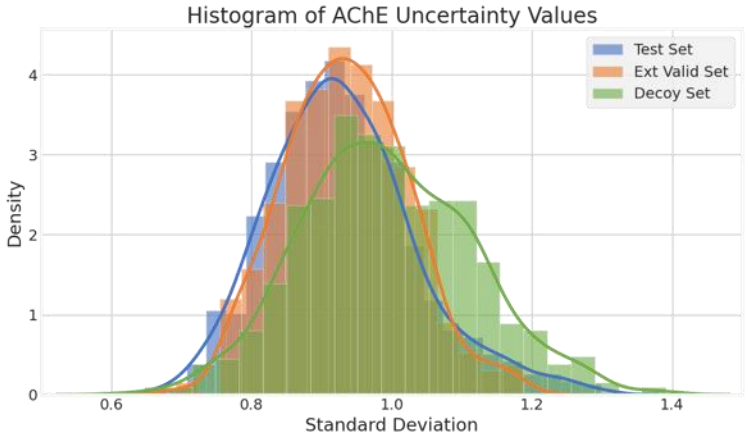


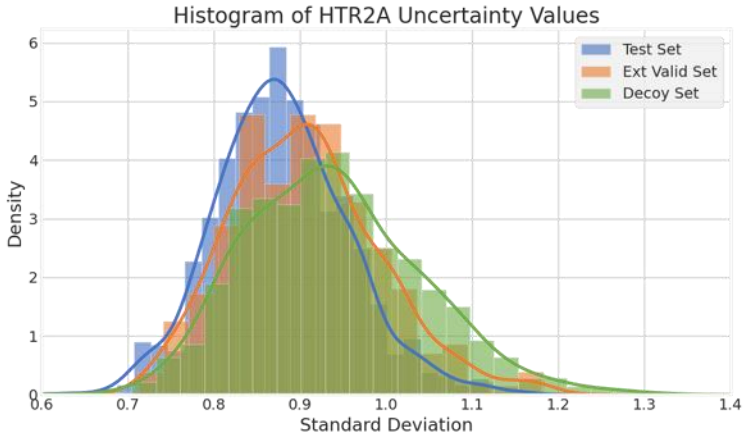


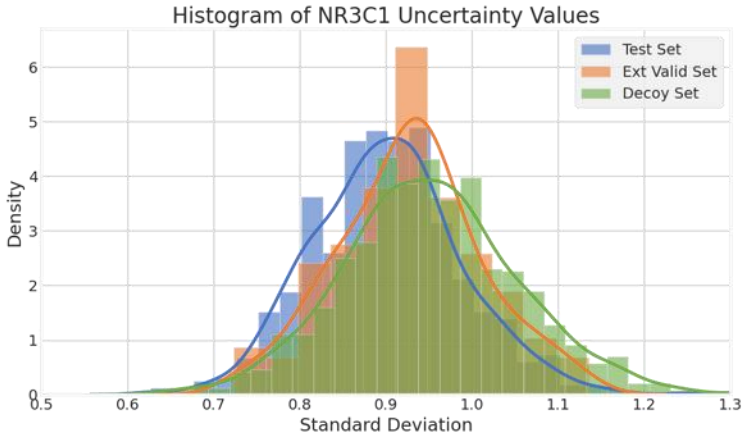


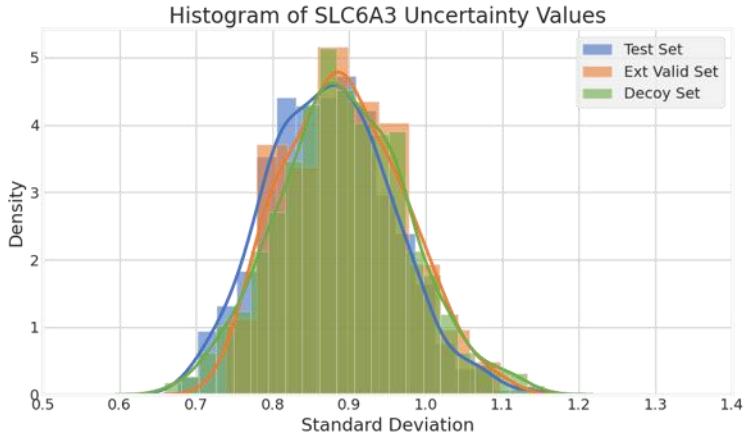


Distribution of Uncertainties

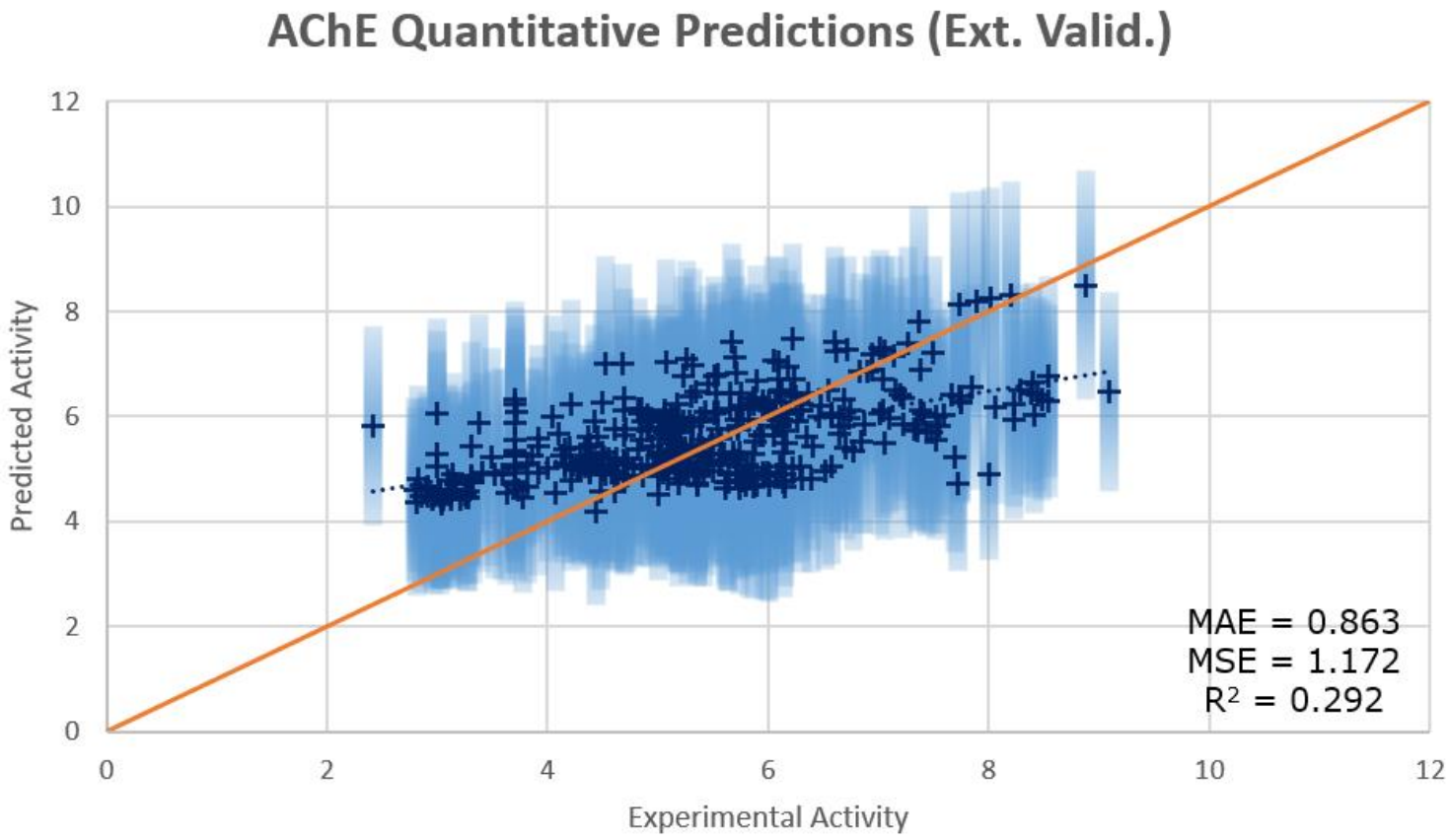




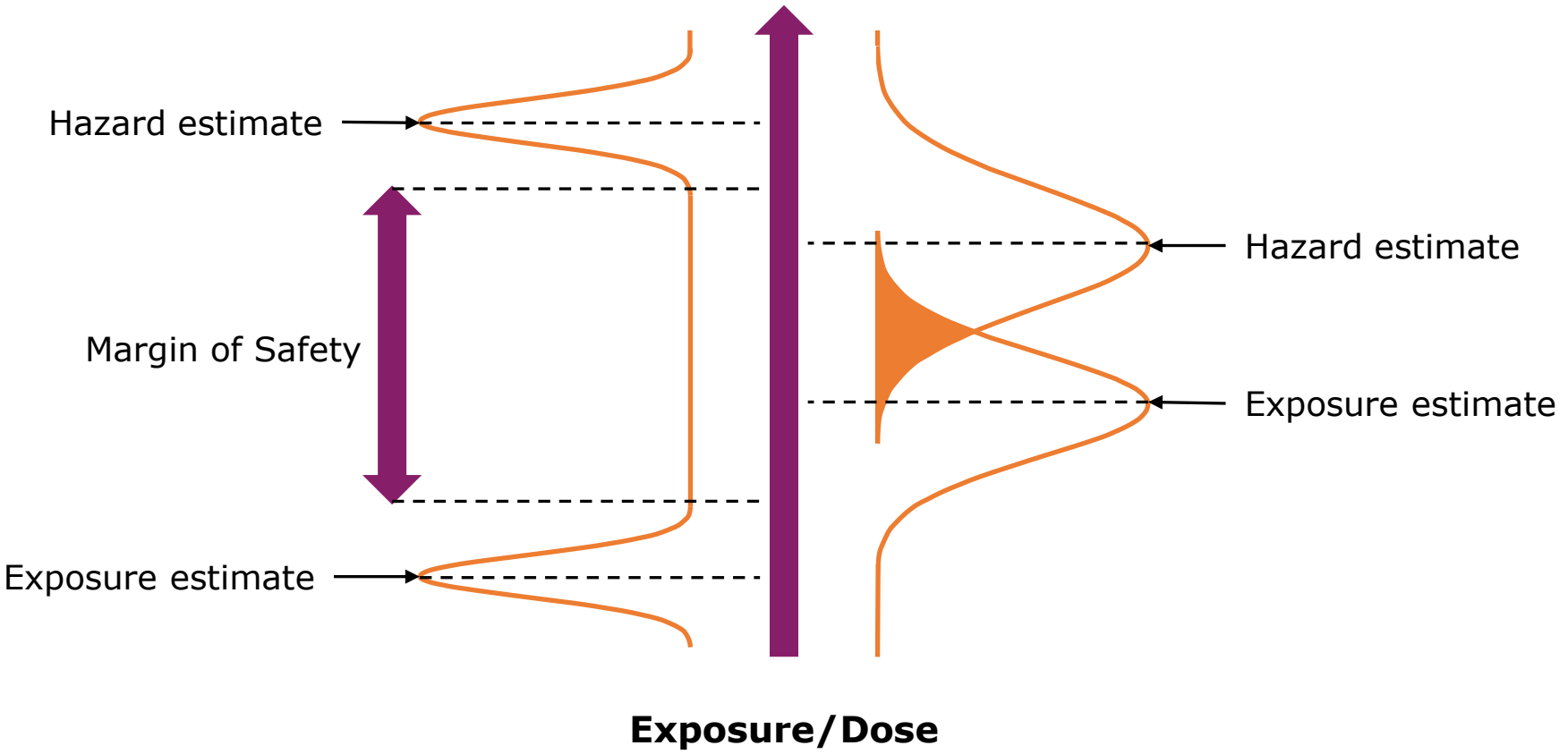




Predictions with Uncertainties



Safety Decision Making

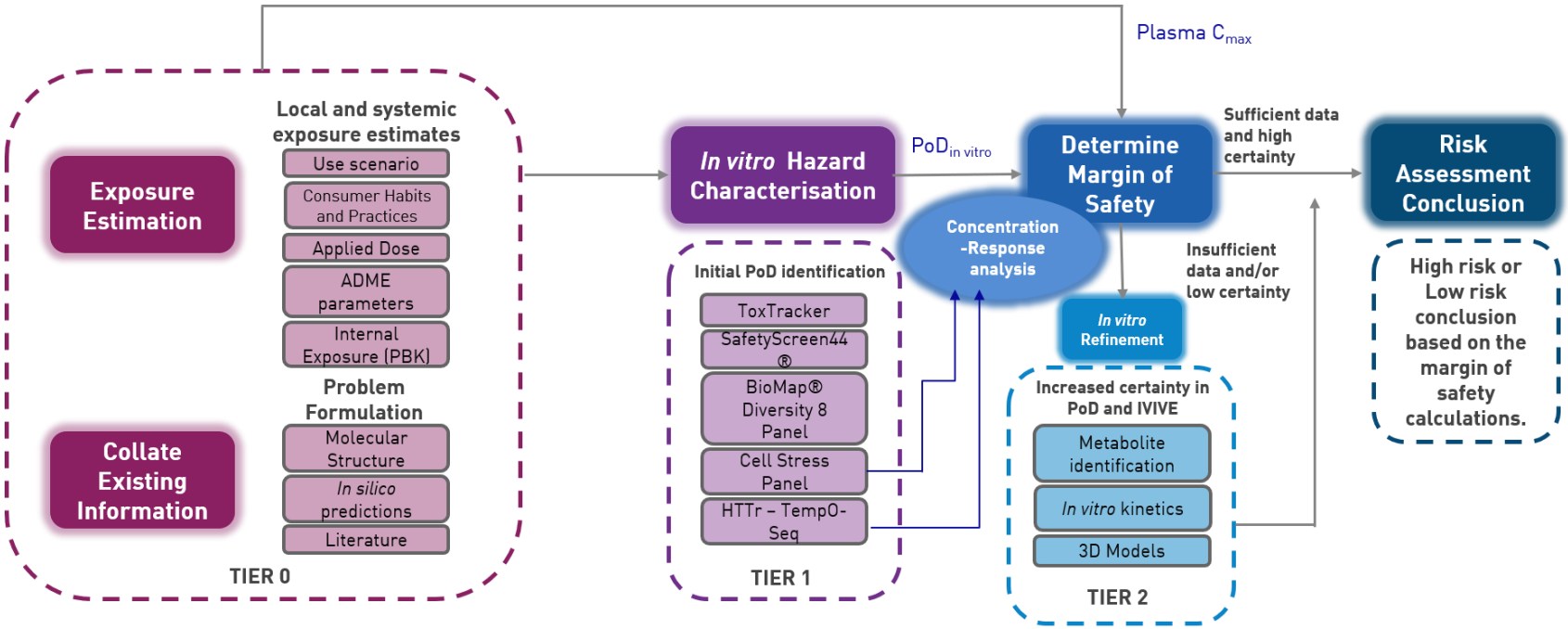


Conclusions

- Bayesian learning regression neural networks provide the ability to both make quantitative predictions and understand the uncertainty in those predictions

- These algorithms have been shown to be useful in the prediction of molecular activity at human MIEs

Ab Initio NGRA Framework



Baltazar, M.T., *et al.* (**2020**) *Toxicol. Sci.*, 176; 236.

Acknowledgements

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

References

- **Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment.** Allen, T.E.H.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *Chem. Res. Toxicol.* **2014**, 27 (12), 2100-2112. DOI: 10.1021/tx500345j
- **Using Transition State Modeling to Predict Mutagenicity for Michael Acceptors.** Allen, T.E.H.; Grayson, M.N.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *J. Chem. Inf. Model.* **2018**, 58 (6), pp 1266-1271. DOI: 10.1021/acs.jcim.8b00130
- **Using 2D Structural Alerts to Define Chemical Categories for Molecular Initiating Events.** Allen, T.E.H.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *Toxicol. Sci.* **2018**, 165 (1), 213-223. DOI: 10.1093/toxsci/kfy144
- **Quantitative Predictions for Molecular Initiating Events using 3D Quantitative Structure-Activity Relationships.** Allen, T.E.H.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *Chem. Res. Toxicol.* **2020**, 33(2), pp 324-332. DOI: 10.1021/acs.chemrestox.9b00136
- **Structural Alerts and Random Forest Models in a Consensus Approach for Receptor Binding Molecular Initiating Events.** Wedlake, A.J., Folia, M., Piechota, S., Allen, T.E.H.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *Chem. Res. Toxicol.* **2020**, 33(2), 388-401. DOI: 10.1021/acs.chemrestox.9b00325
- **Neural Network Activation Similarity: A New Measure to Assist Decision Making in Chemical Toxicology.** Allen, T.E.H.; Wedlake, A.J.; Gelžinytė, E.; Gong, C.; Goodman, J.M.; Gutsell, S.; Russell, P.J. *Chem. Sci.* **2020**, 11, 7335-7348. DOI: 10.1039/d0sc01637c
- **Confidence in Inactive and Active Predictions from Structural Alerts.** Wedlake, A.J.; Allen, T.E.H.; Goodman, J.M.; Kukic, P.; Gutsell, S.; Russell, P.J. *Chem. Res. Toxicol.* **2020**, 33(12), 3010-3022. DOI: 10.1021/acs.chemrestox.0c00332

Use of chemical informatics, quantum chemistry modelling and artificial intelligence algorithms to predict molecular initiating events

Timothy E H Allen – teha2@cam.ac.uk

AJ Wedlake, MN Grayson, AM Middleton, M Folia, M Baltazar, P Piechota, E Gelžinytė, JM Goodman, PJ Russell, P Kukic, S Gutsell

EUROTOX – 27th September 2021