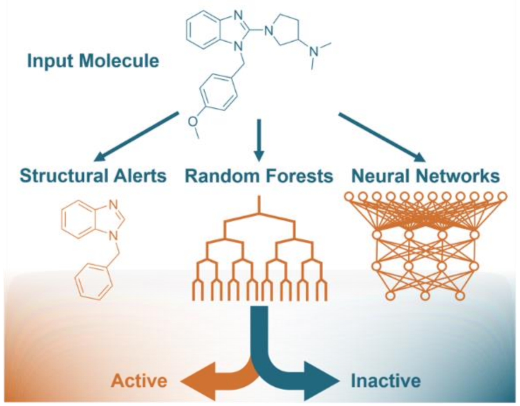
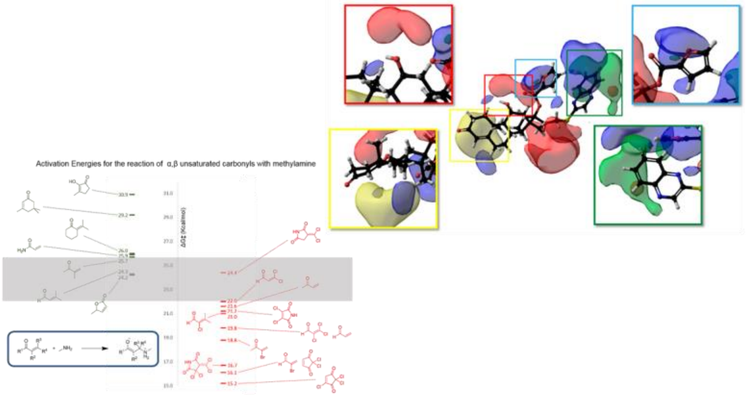
In Silico Approaches to Link Adverse Outcomes to Molecular Initiating Events through AOPs

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

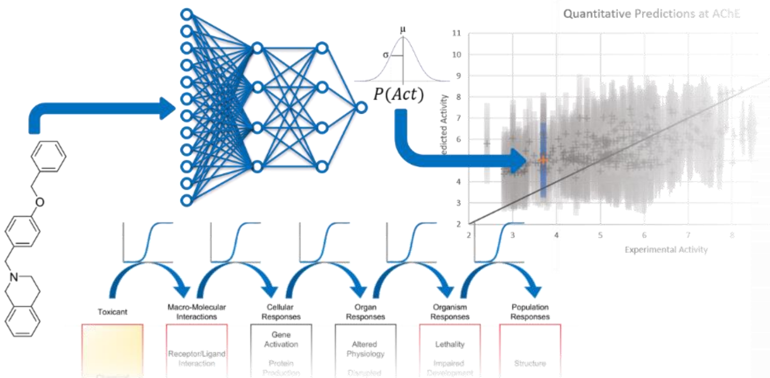
WC11 – 26th August 2021



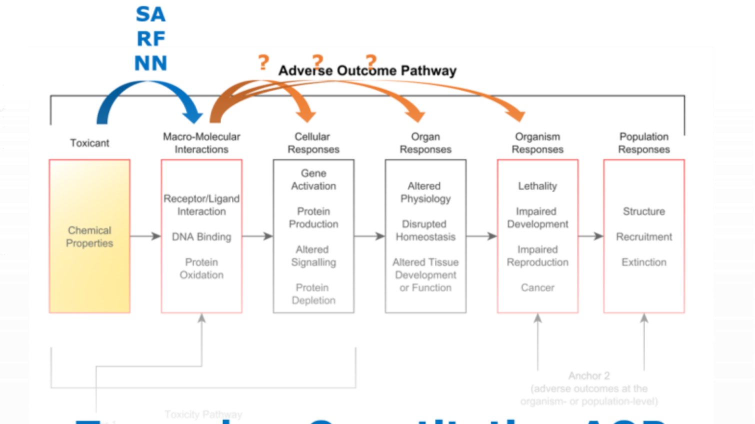
Mechanistic Toxicity Predictions



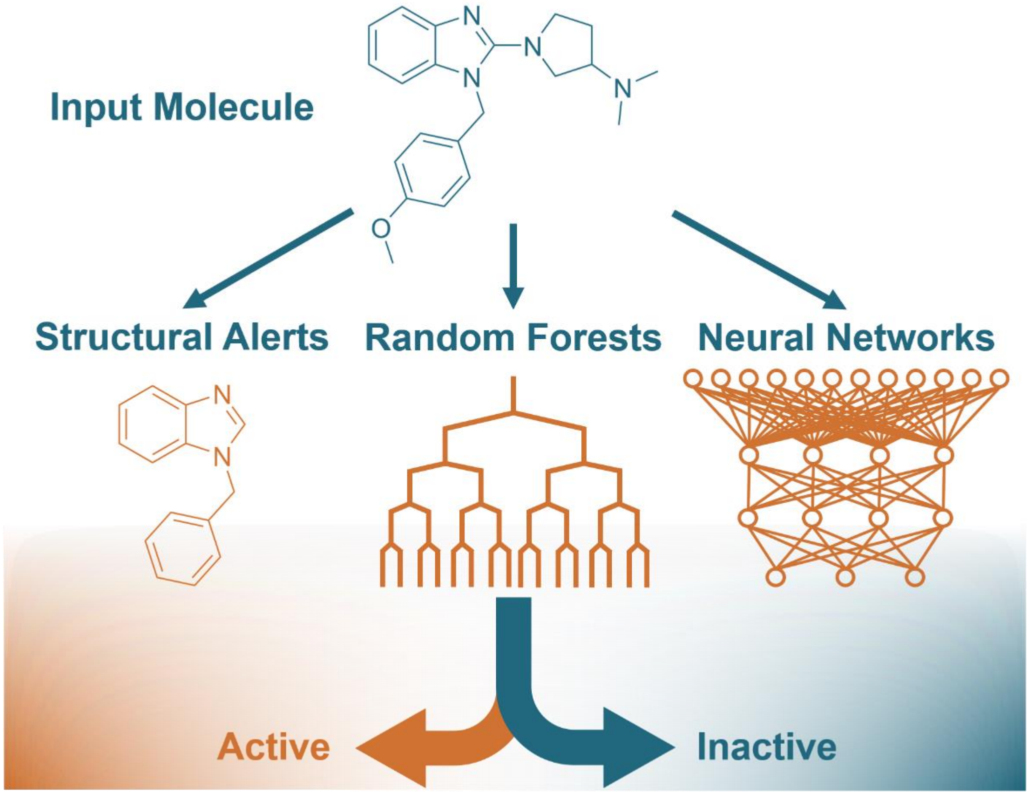
Other Computational Methods



Bayesian Neural Networks



Towards a Quantitative AOP



Input Molecule

Structural Alerts

Random Forests

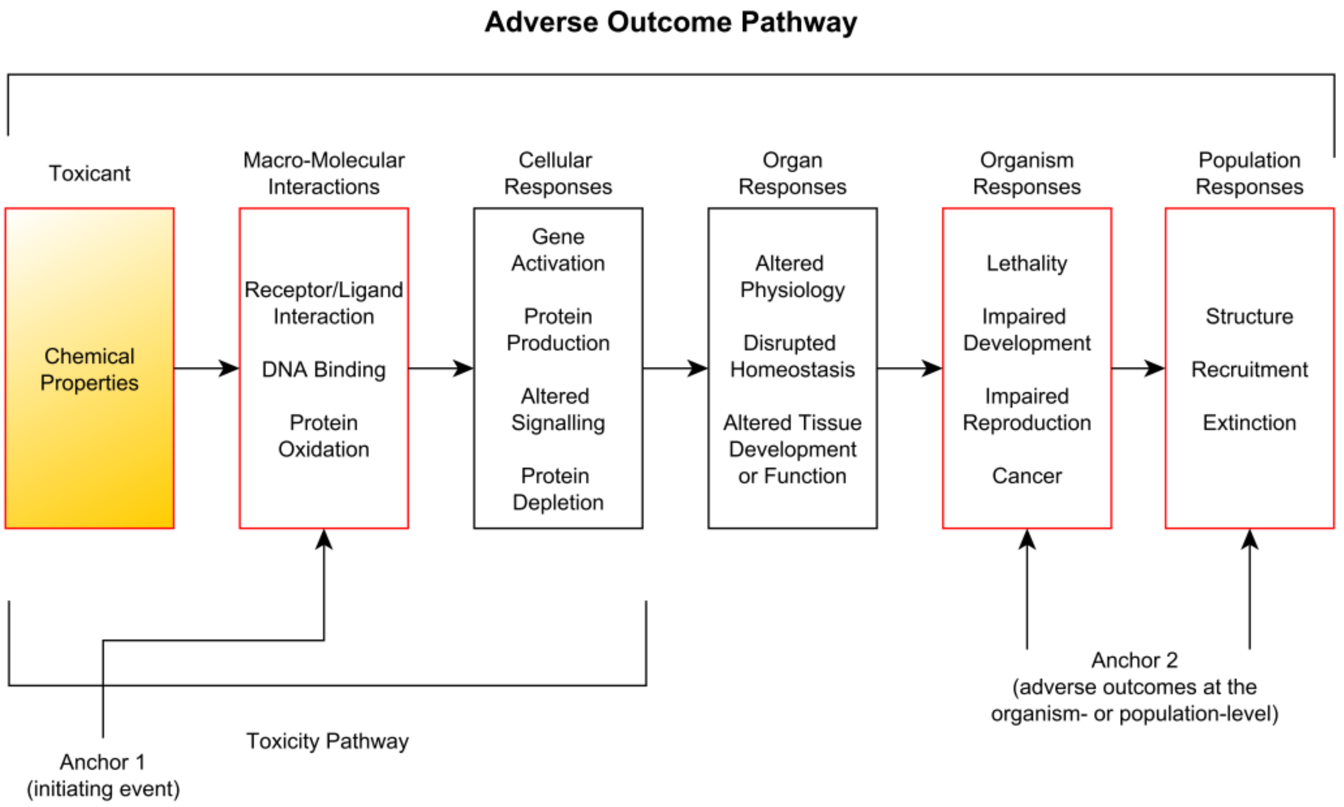
Neural Networks

Active

Inactive

Mechanistic Toxicity Predictions

Adverse Outcome Pathway



Adverse Outcome Pathway

Toxicant

Chemical Properties

Macro-Molecular Interactions

Receptor/Ligand Interaction

DNA Binding

Protein Oxidation

Cellular Responses

Gene Activation

Protein Production

Altered Signalling

Protein Depletion

Organ Responses

Altered Physiology

Disrupted Homeostasis

Altered Tissue Development or Function

Organism Responses

Lethality

Impaired Development

Impaired Reproduction

Cancer

Population Responses

Structure

Recruitment

Extinction

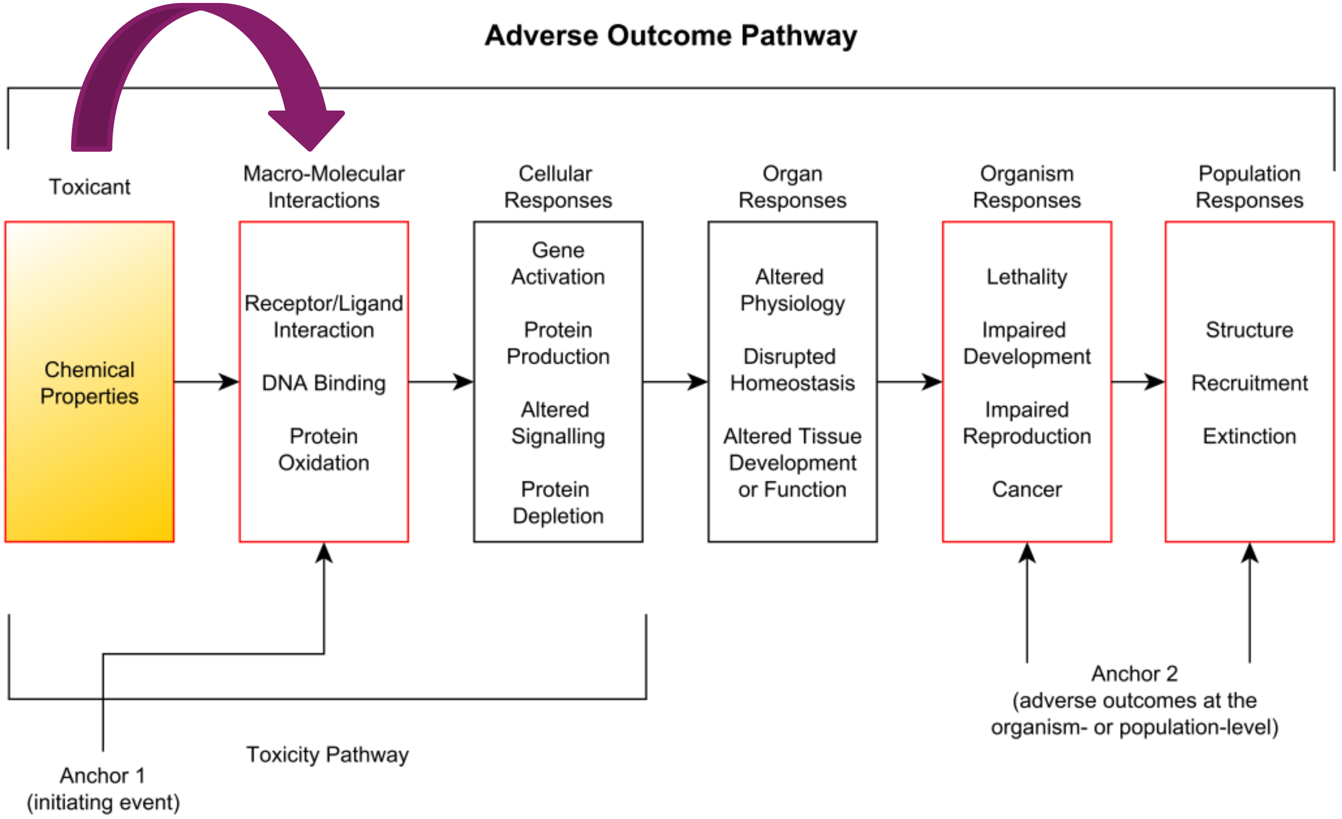
Anchor 1 (initiating event)

Toxicity Pathway

Anchor 2 (adverse outcomes at the organism- or population-level)

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

Adverse Outcome Pathway



Adverse Outcome Pathway

Toxicant

Chemical Properties

Macro-Molecular Interactions

Receptor/Ligand Interaction

DNA Binding

Protein Oxidation

Cellular Responses

Gene Activation

Protein Production

Altered Signalling

Protein Depletion

Organ Responses

Altered Physiology

Disrupted Homeostasis

Altered Tissue Development or Function

Organism Responses

Lethality

Impaired Development

Impaired Reproduction

Cancer

Population Responses

Structure

Recruitment

Extinction

Anchor 1 (initiating event)

Toxicity Pathway

Anchor 2 (adverse outcomes at the organism- or population-level)

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

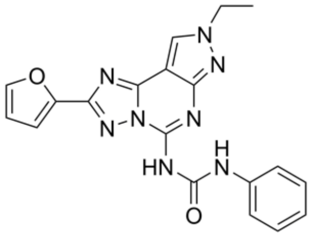
Structural Alerts

ChEMBL

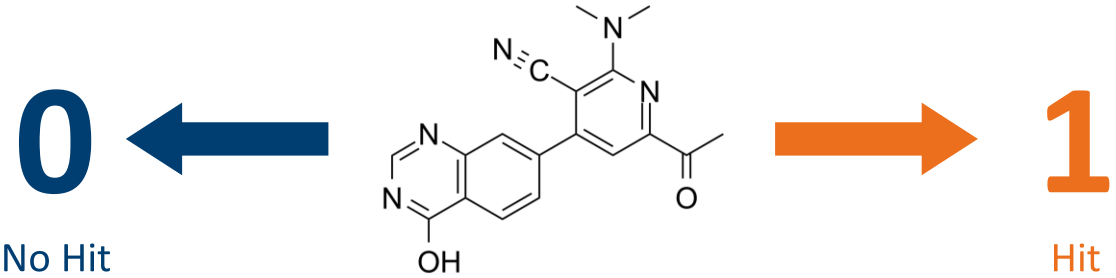
Bowes 2012
Targets

Computational Tools

2D Fragments
SMILES



Performance Data



0

No Hit

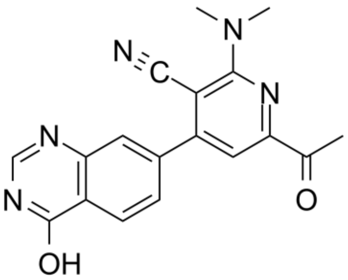
1

Hit

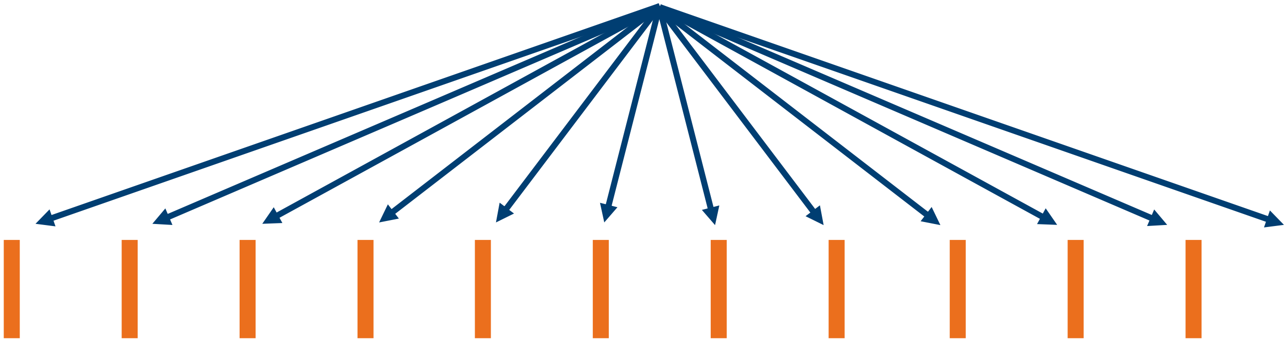
Novel Compound

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

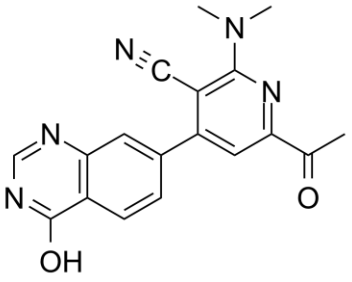


Yes Yes No Yes No Yes Yes No No Yes Yes Yes

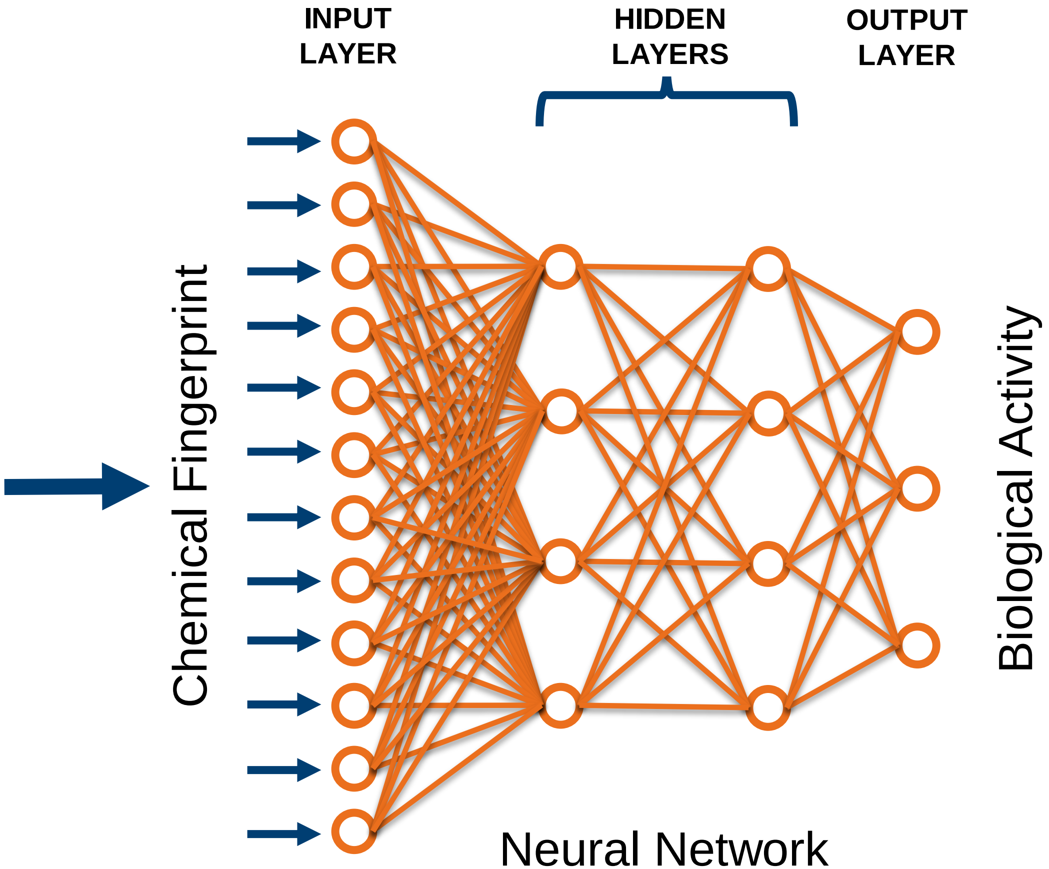
8/ 12 Trees predict Yes
Therefore overall prediction is Yes

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

Neural Network



Novel Compound



Neural Network

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

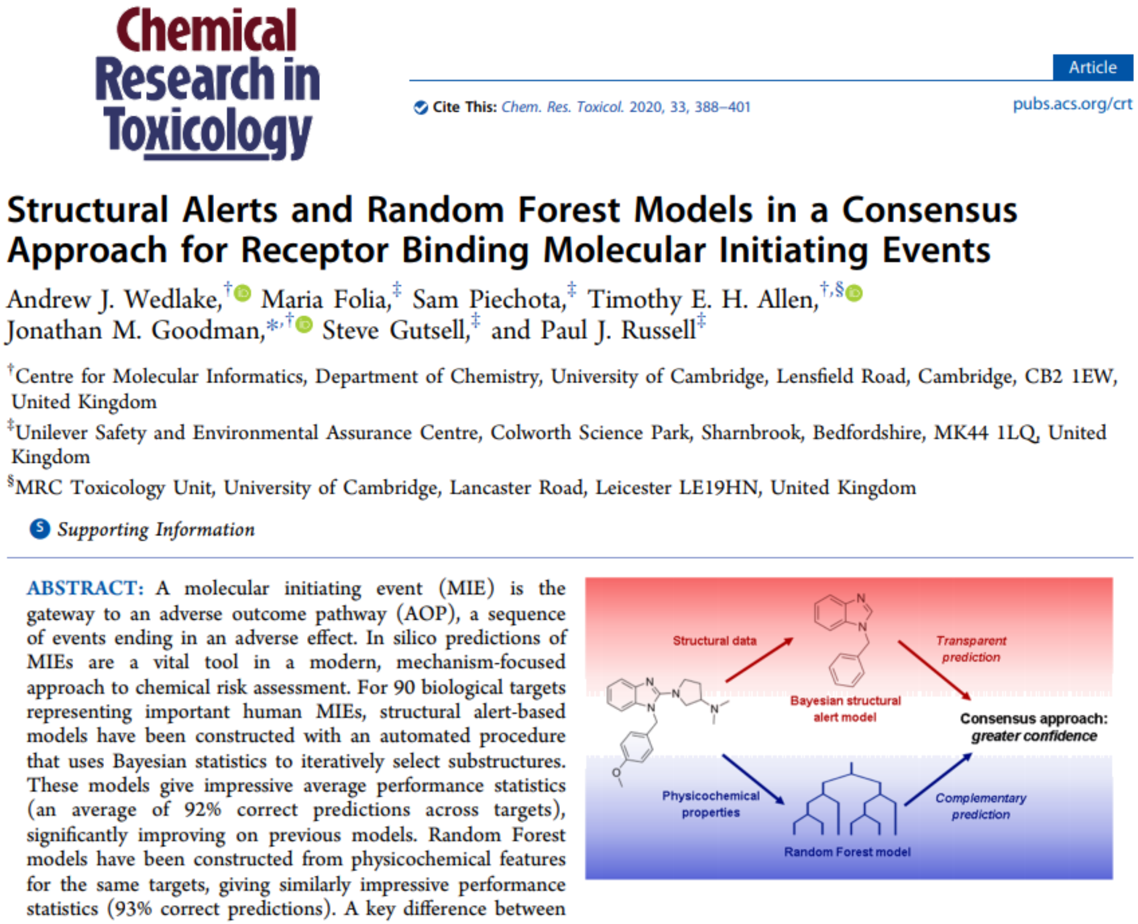
Model Performance

		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



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

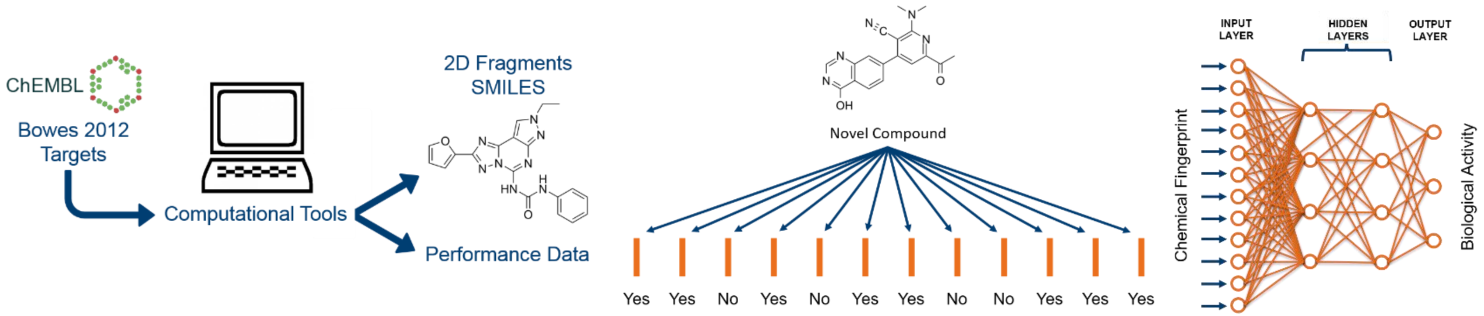
[§]MRC Toxicology Unit, University of Cambridge, Lancaster Road, Leicester LE19HN, United Kingdom

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



ChEMBL

Bowes 2012 Targets

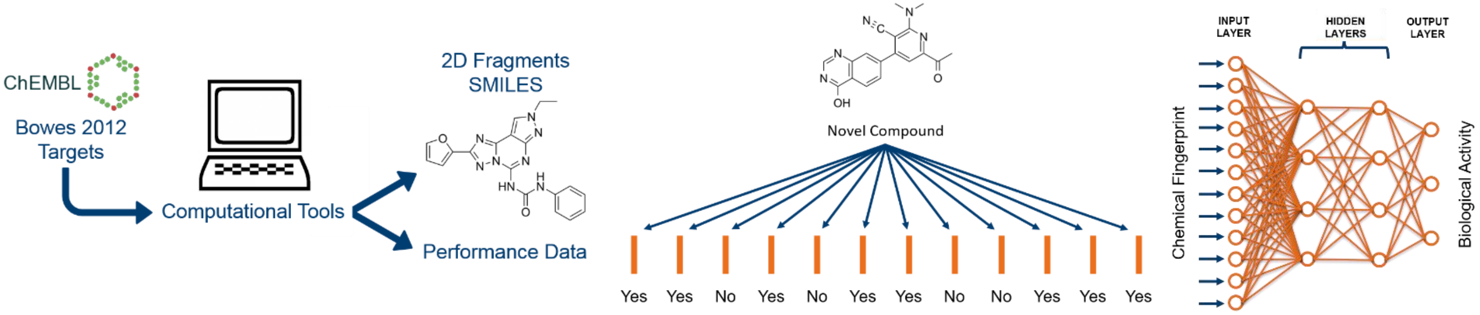
Computational Tools

2D Fragments SMILES

Performance Data

Novel Compound

Combined Model

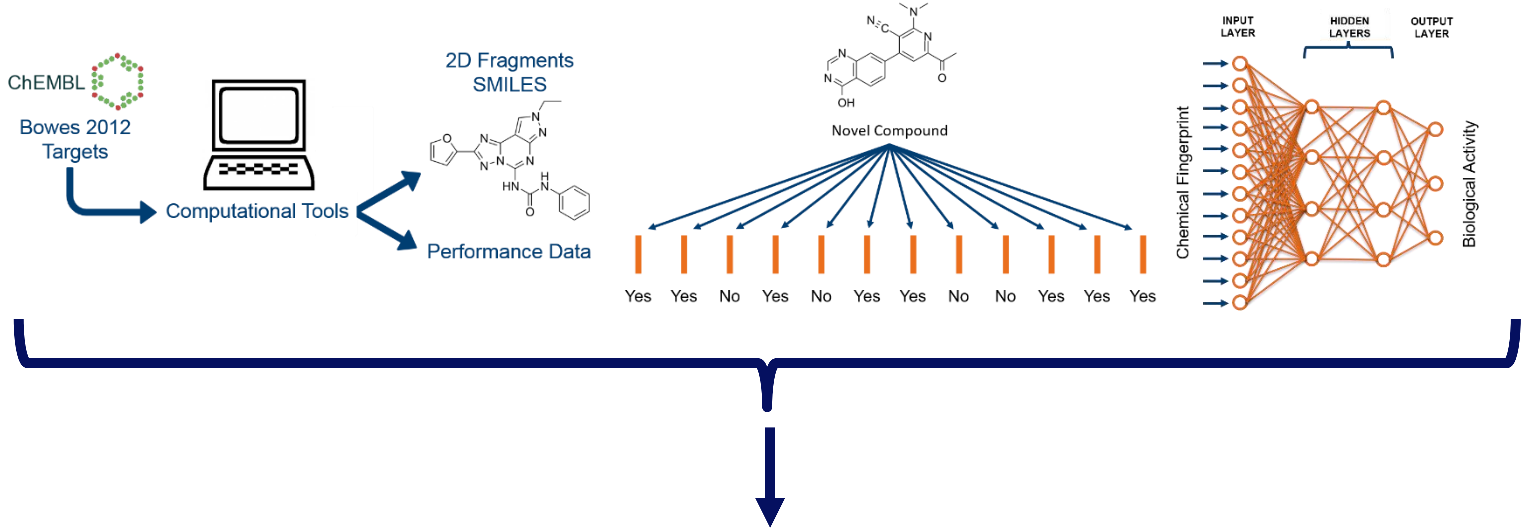


Transparent Prediction Algorithm

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	

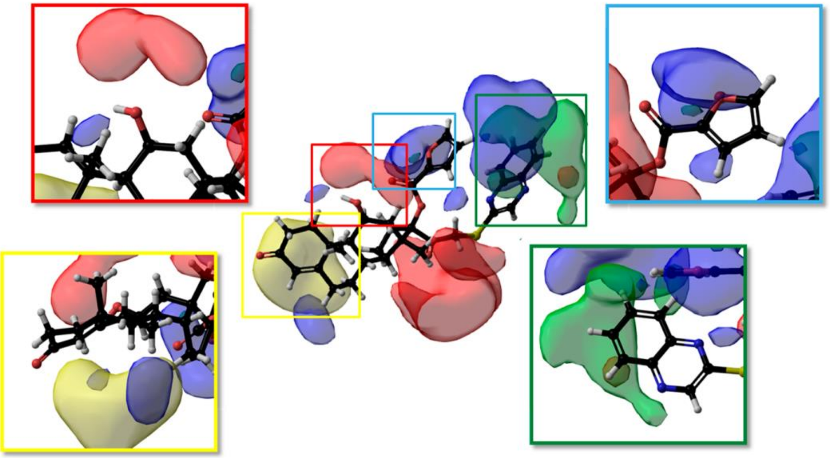
Wedlake, A.J. *et al.* (**2019**) *Chem. Res. Toxicol.*, 33; 388.
Allen, T.E.H. *et al.* (**2020**) *Chem. Sci.*, 11; 7335.

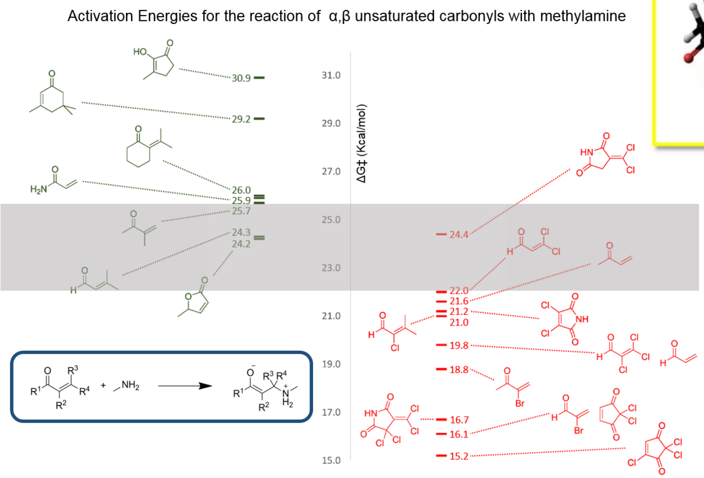
Conclusions

- 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

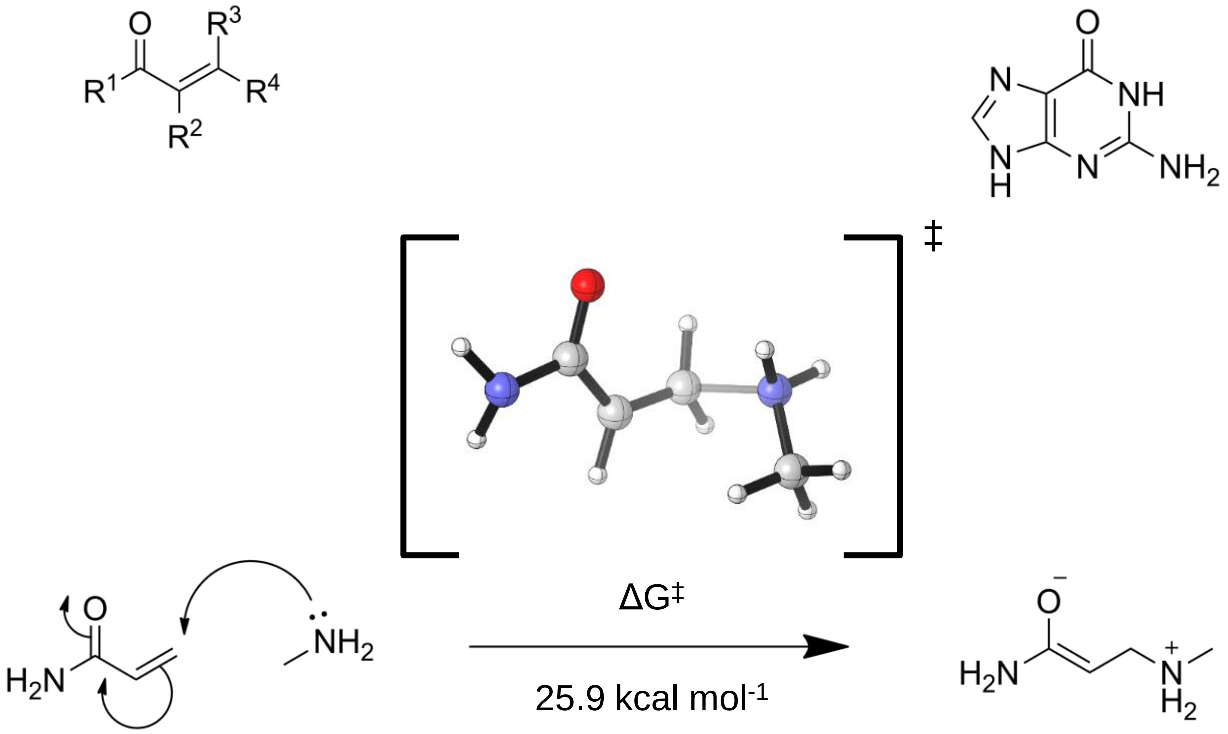
- Using these algorithms together increases their performance and the confidence we can have in their predictions





Other Computational Methods

DFT for Mutagenicity



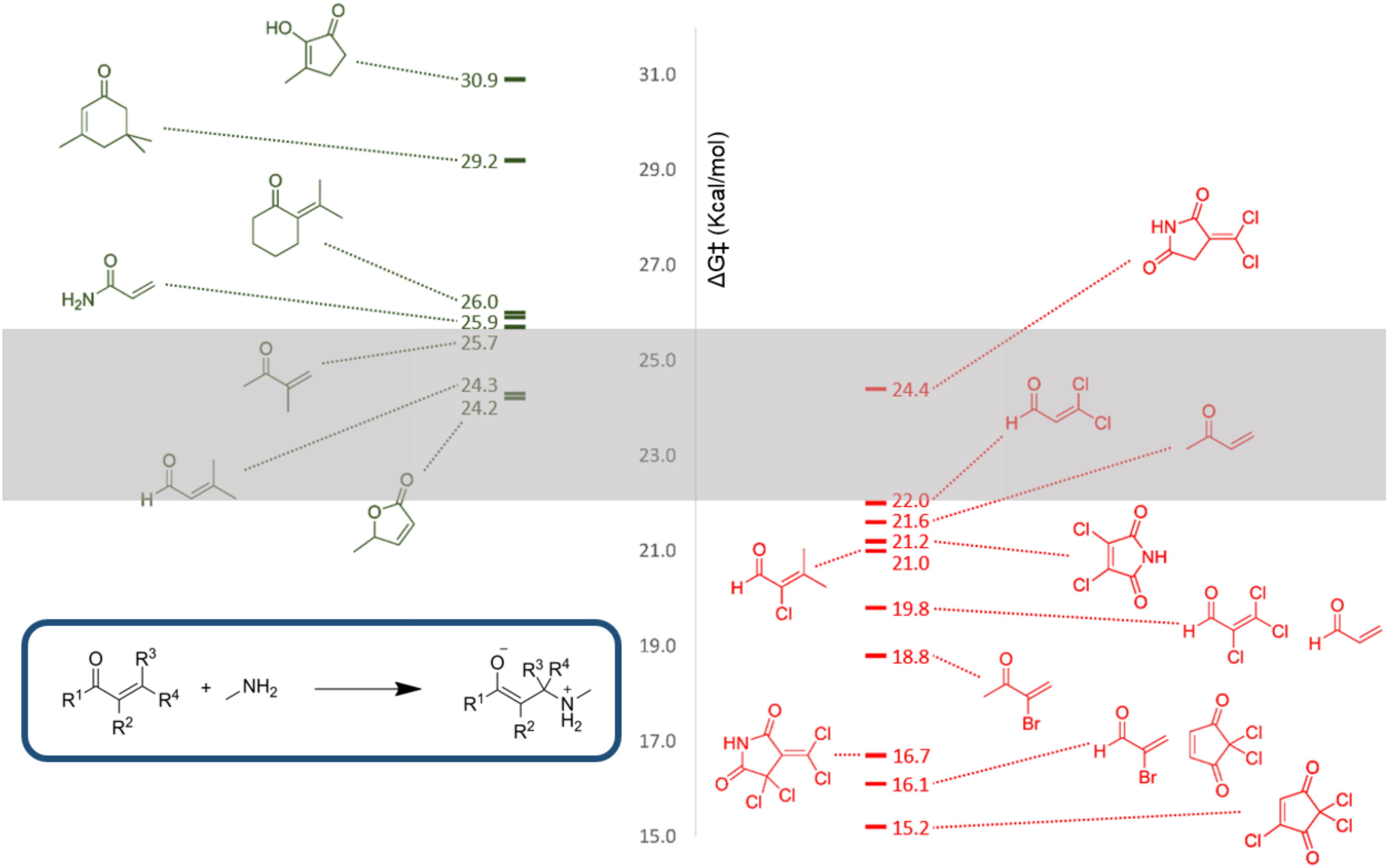
$\Delta G^{\ddagger}$

25.9 kcal mol^{-1}

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

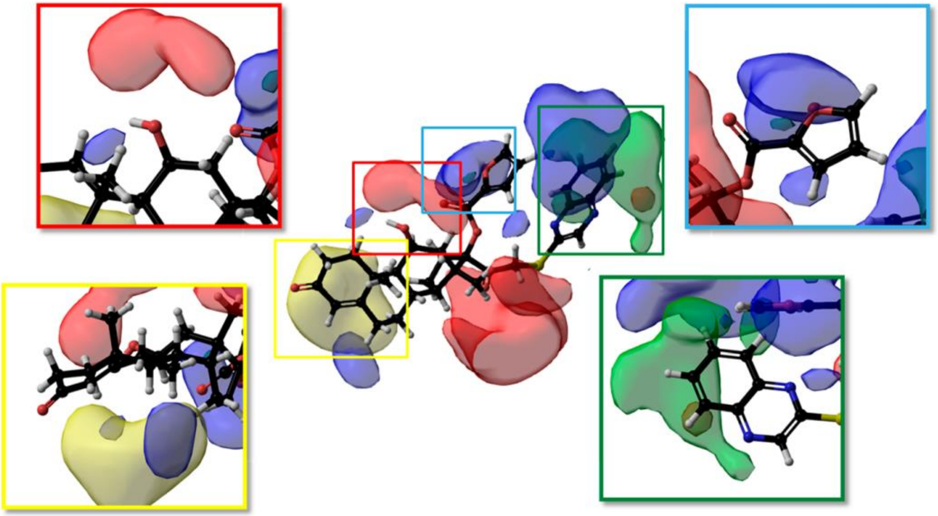
DFT for Mutagenicity

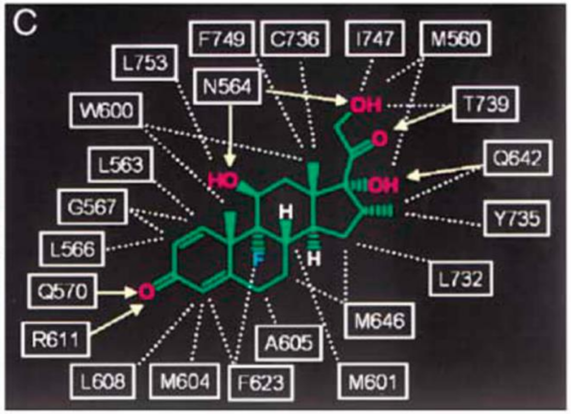
Activation Energies for the reaction of α,β unsaturated carbonyls with methylamine

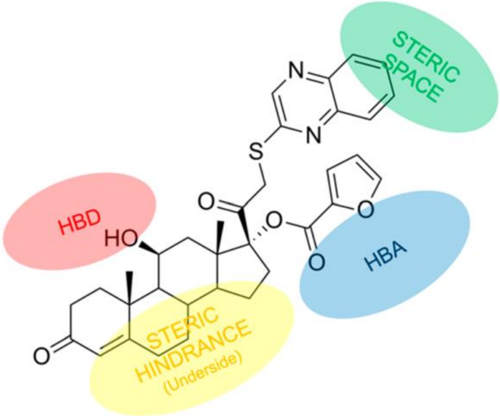


Allen, T. E. H. *et al*. (**2018**) *J. Chem. Inf. Mod.*, 58; 1266.
DFT - Optimization: B3LYP, 6-31+G(d), iefpcm; SPE: M062X, def2tzvpp, iefpcm

3DQSAR Studies

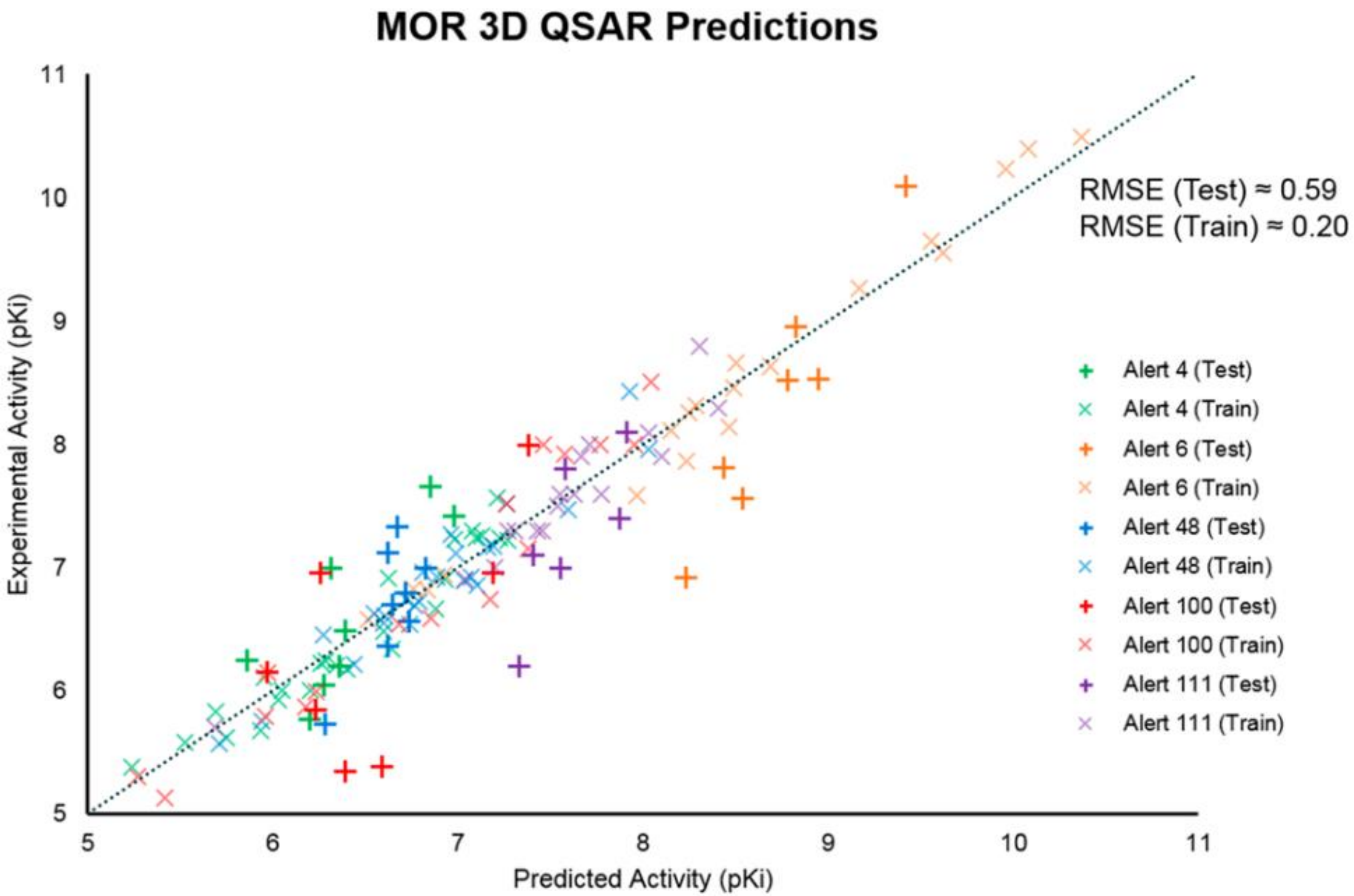






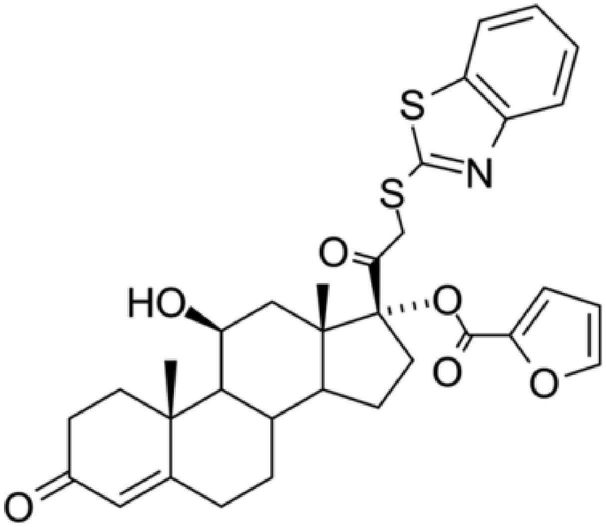
Allen, T. E. H. *et al*. (**2020**) *Chem. Res. Toxicol.*, 33; 324.
Bledsoe, R. K. *et al*. (**2002**) *Cell*, 110; 93.

3DQSAR Studies

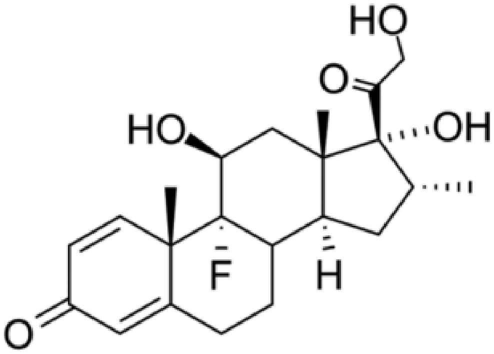


Allen, T. E. H. *et al*. (**2020**) *Chem. Res. Toxicol.*, 33; 324.

3DQSAR Studies



Target Chemical



Dexamethasone

Predicted pIC50 = 8.46

Experimental pIC50 = 8.70

Experimental pIC50 = 8.53

Developmental effects?

Developmental effects

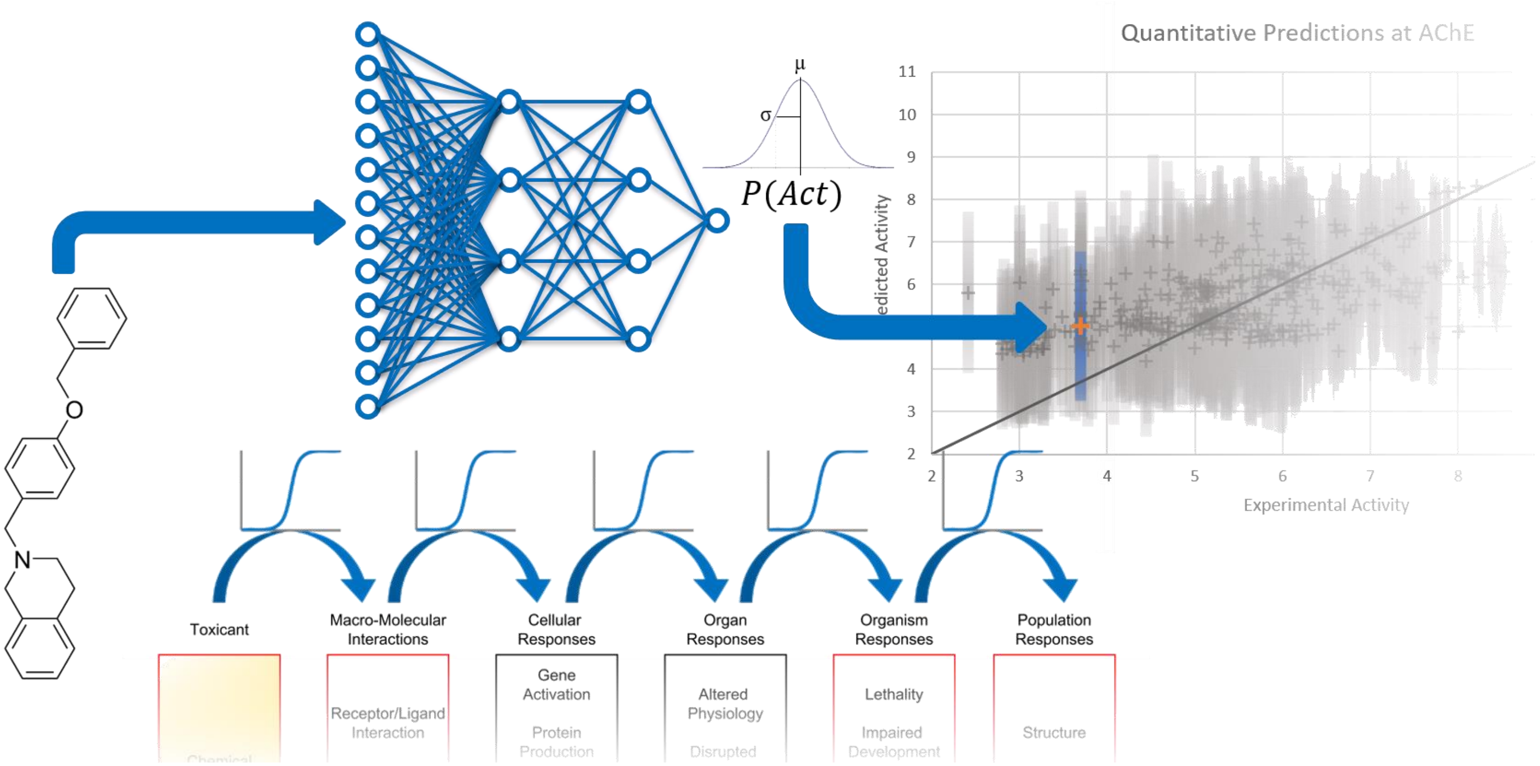
Allen, T. E. H. *et al*. (**2020**) *Chem. Res. Toxicol.*, 33; 324.

Conclusions

- DFT calculations provide a window into the molecular interactions driving some Ames positive results

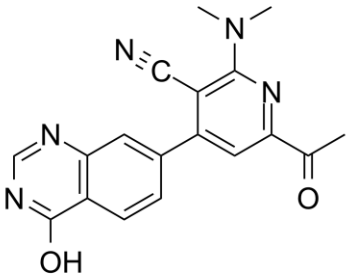
- CoMFA 3D QSAR calculations can provide analogous image of receptor binding interactions

- Both of these approaches increase understanding in how and why specific molecules activate MIEs

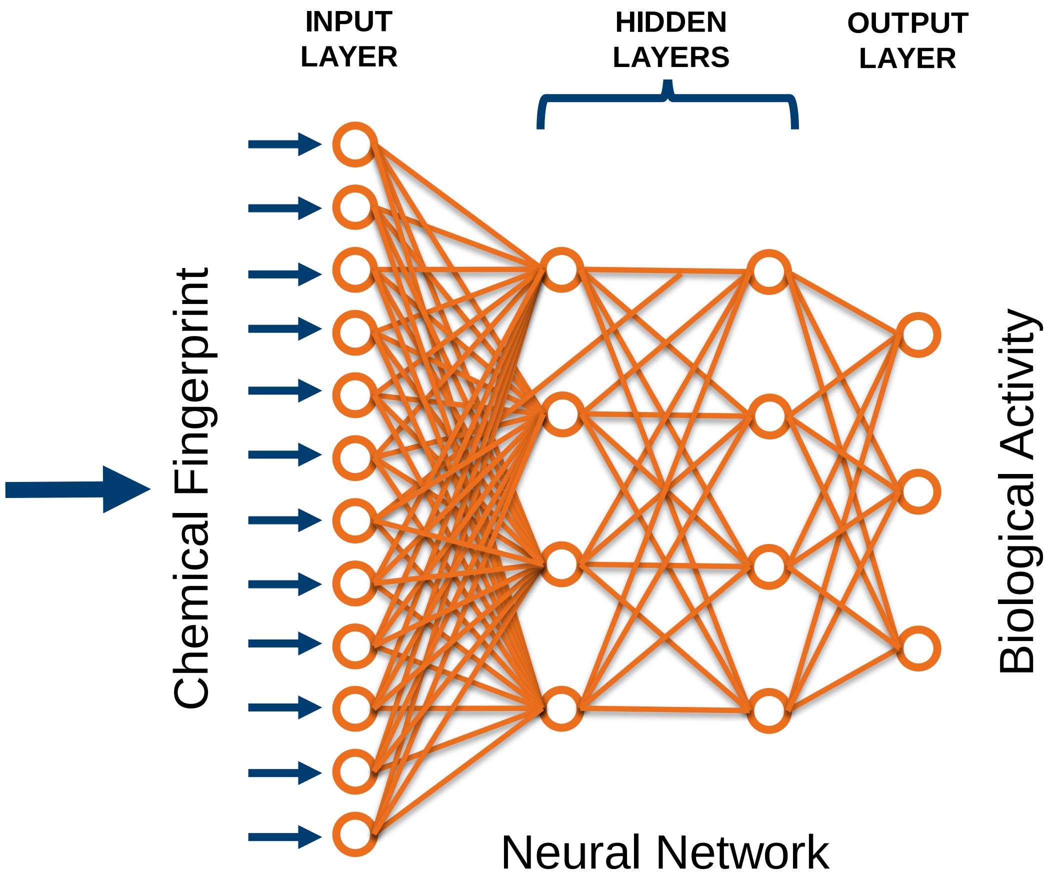


Bayesian Neural Networks

Quantitative Neural Network?



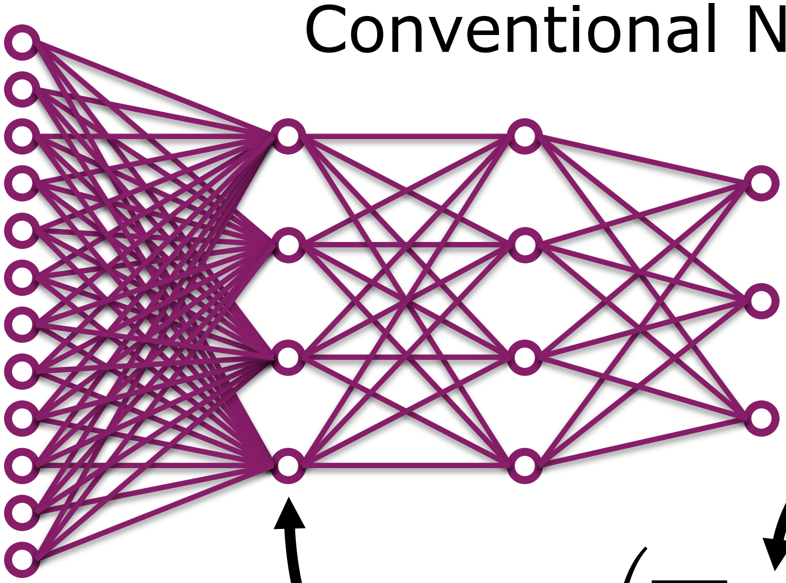
Novel Compound



Neural Network

Conventional Learning

Conventional Neural Network



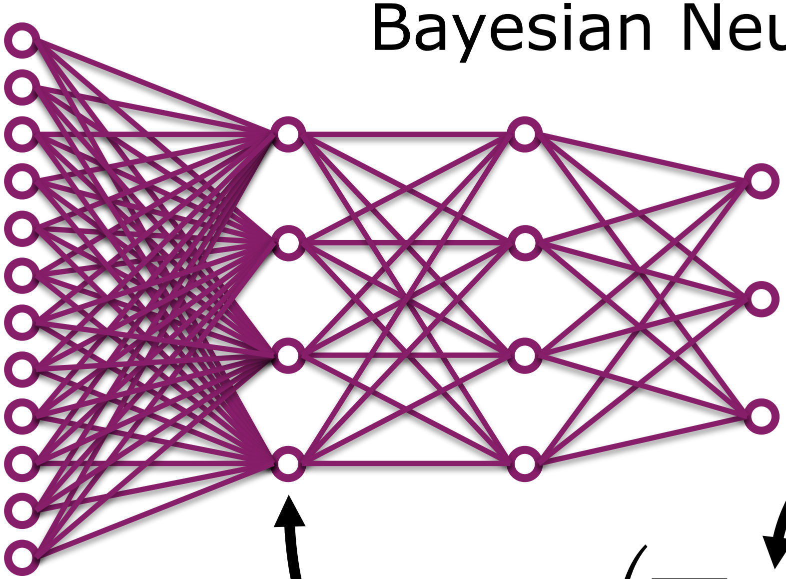
$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{K}(\left(\sum_{\boldsymbol{i}} \boldsymbol{w}_{\boldsymbol{i}}(\boldsymbol{x}_{\boldsymbol{i}})\right) + \boldsymbol{b})$$

Weights and biases
as fixed values

Therefore predictions
are fixed values

Bayesian Learning

Bayesian Neural Network



Weights and biases defined as probability distributions

Gives predictions meaningful errors

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{K}\left(\left(\sum_{\boldsymbol{i}} \boldsymbol{w}_{\boldsymbol{i}}(\boldsymbol{x}_{\boldsymbol{i}})\right) + \boldsymbol{b}\right)$$

Best Models

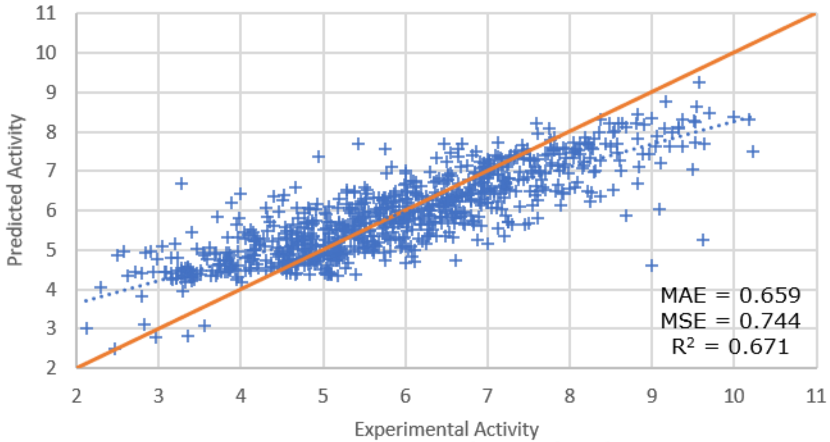
	Train MAE	Valid. MAE	Test MAE	Ext. Val. MAE
AVERAGE	0.487	0.613	0.621	0.943
SD	0.030	0.051	0.051	0.223

	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

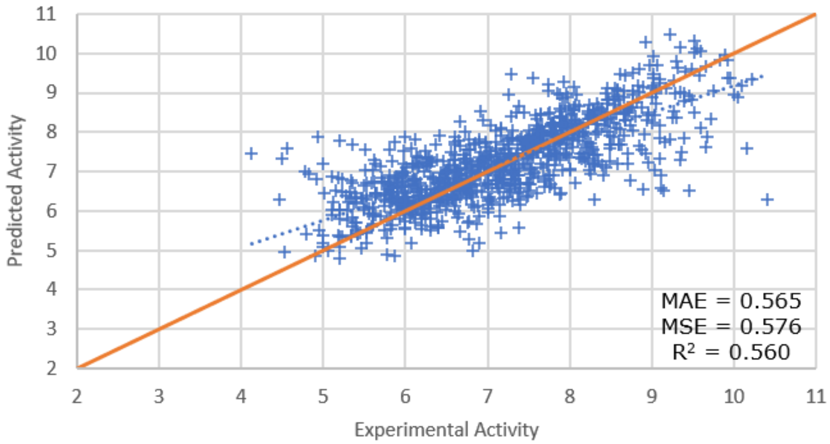
MAE *= Mean Absolute Error*
SD *= Standard Deviation*
R^2 *= Coefficient of Determination for Linear Correlation*

Some Models (Test Set)

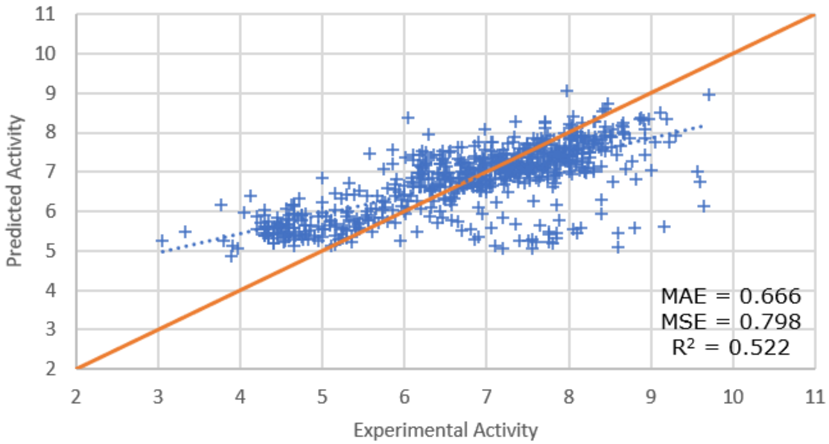
AChE Quantitative Predictions (Test)



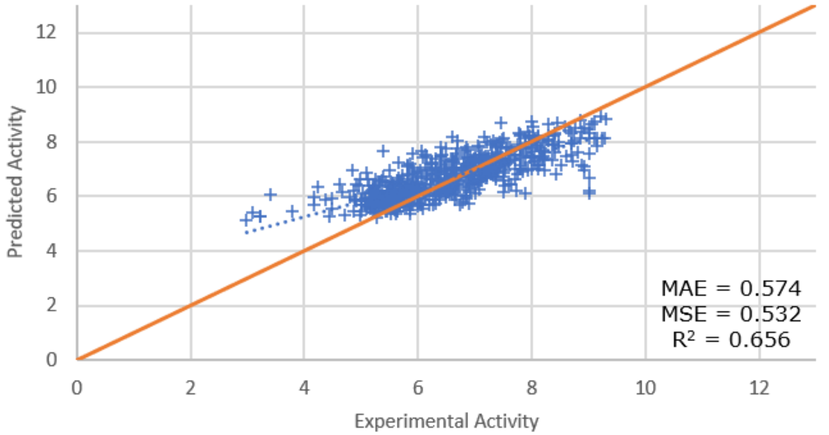
HTR2A Quantitative Predictions (Test)



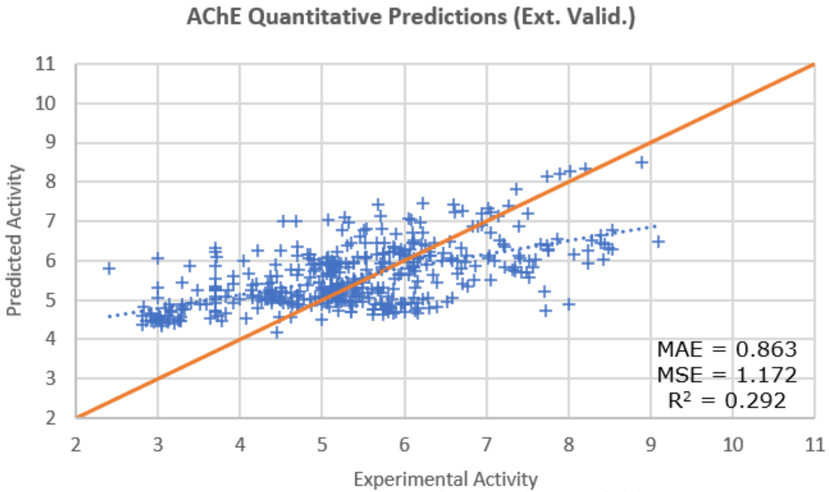
NR3C1 Quantitative Predictions (Test)

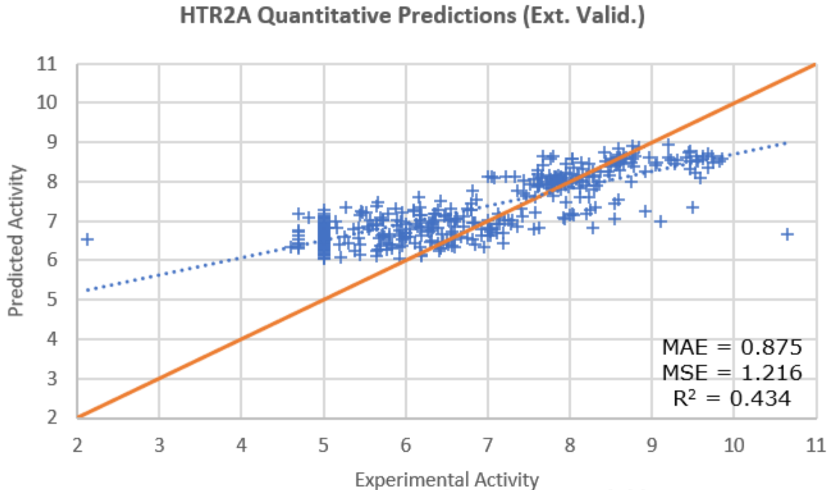


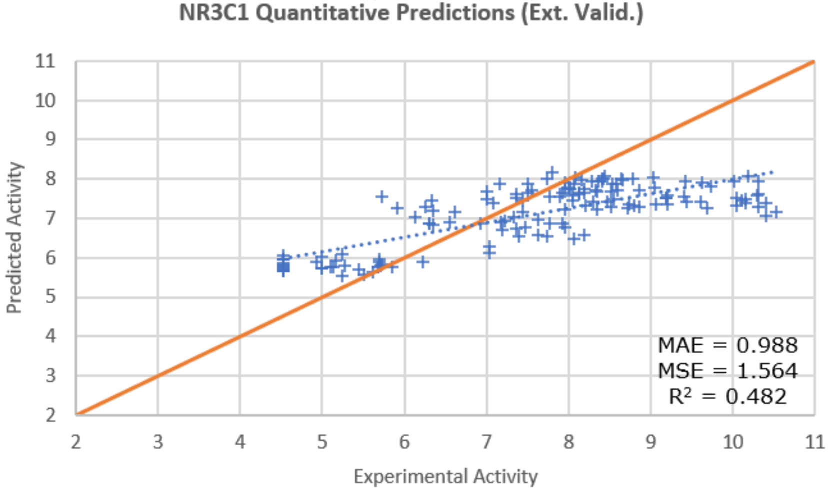
SLC6A3 Quantitative Predictions (Test)

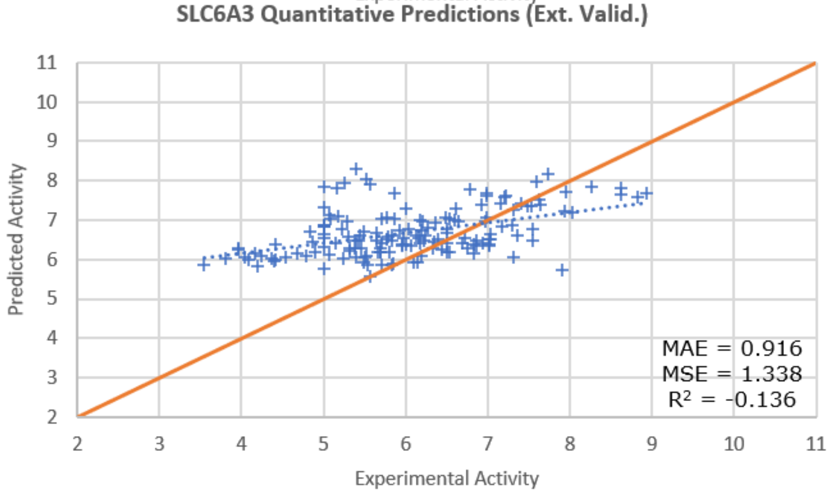


Some Models (Ext. Valid. Set)

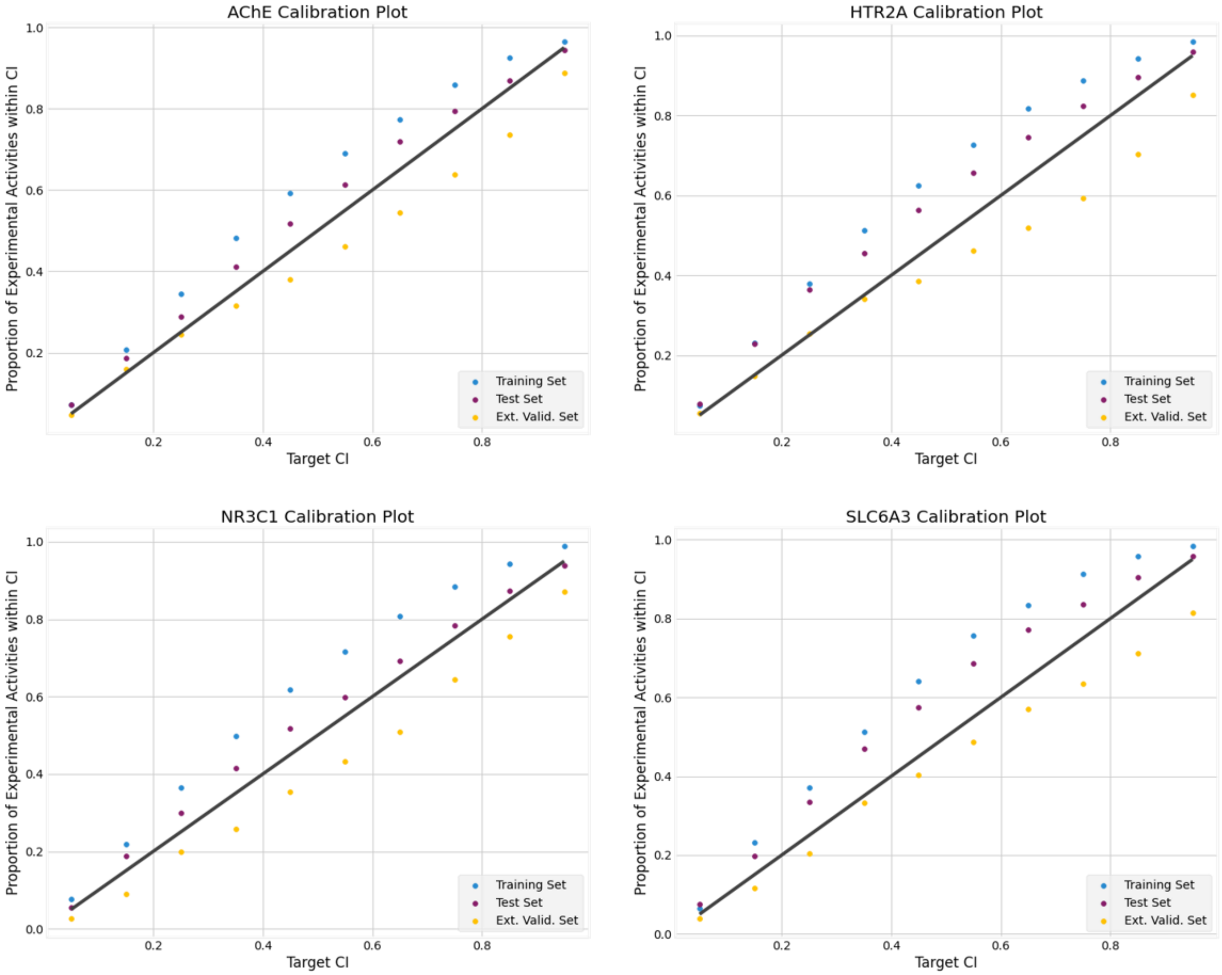




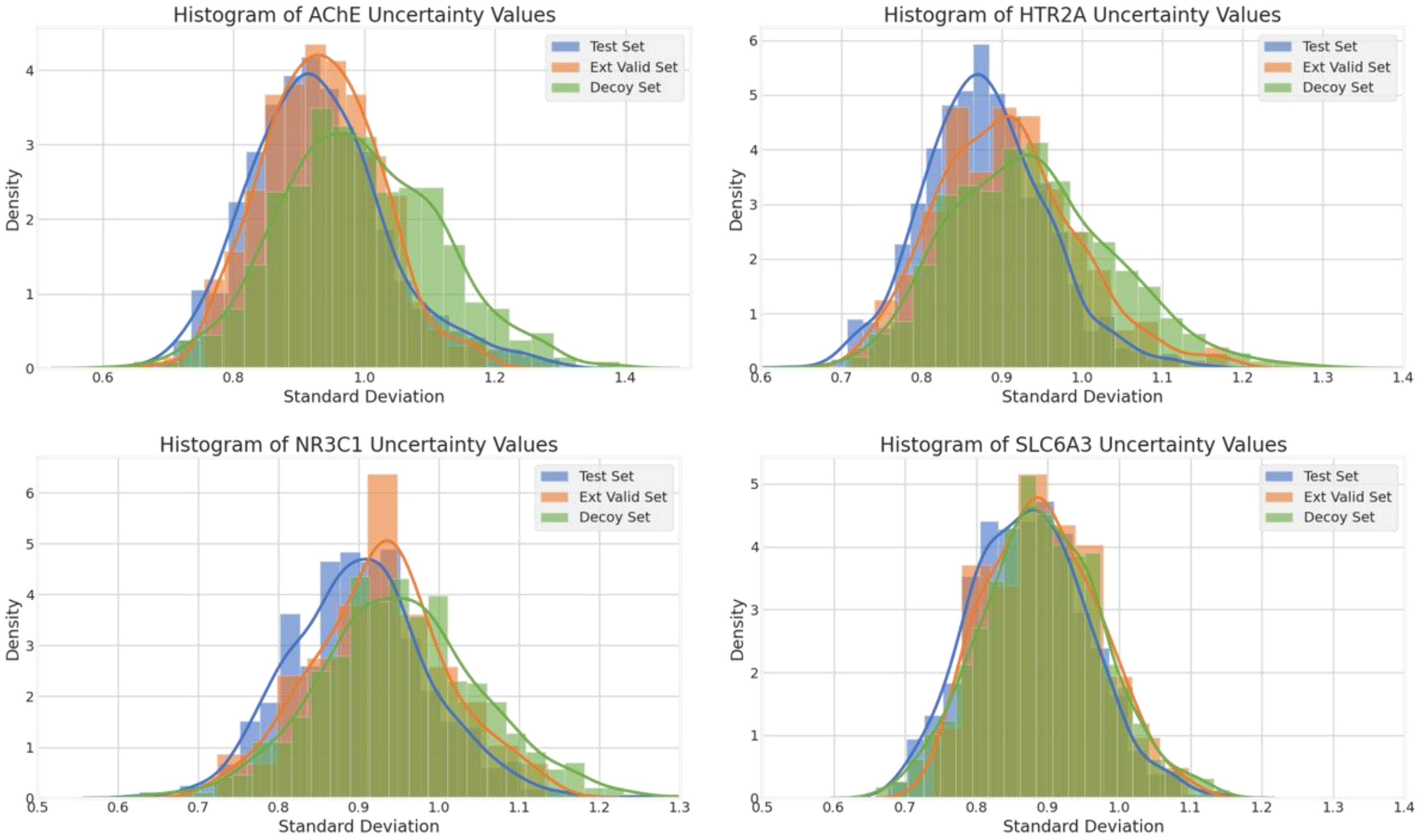




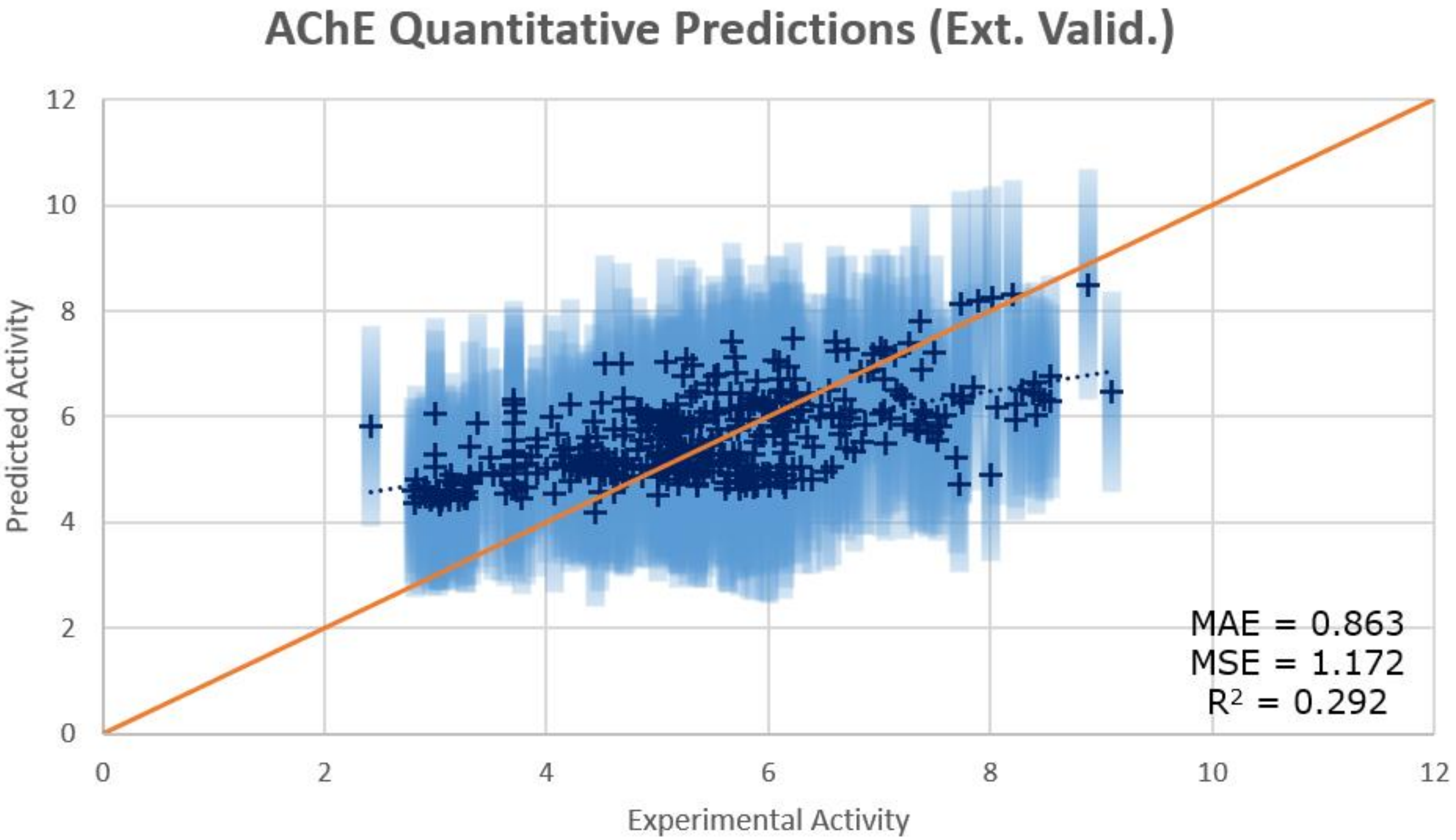
Calibration of Uncertainties



Distribution of Uncertainties



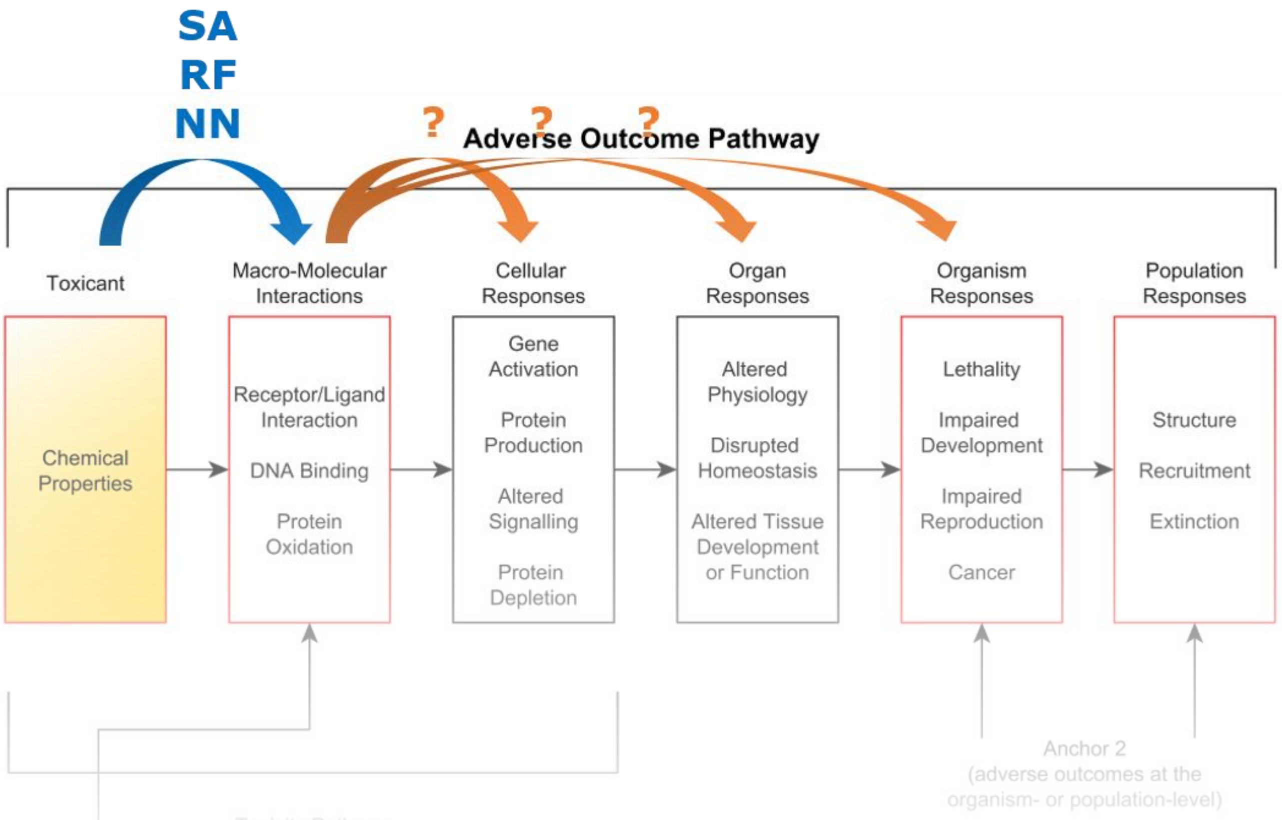
Predictions with Uncertainties



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



SA
RF
NN

? ? ?

Adverse Outcome Pathway

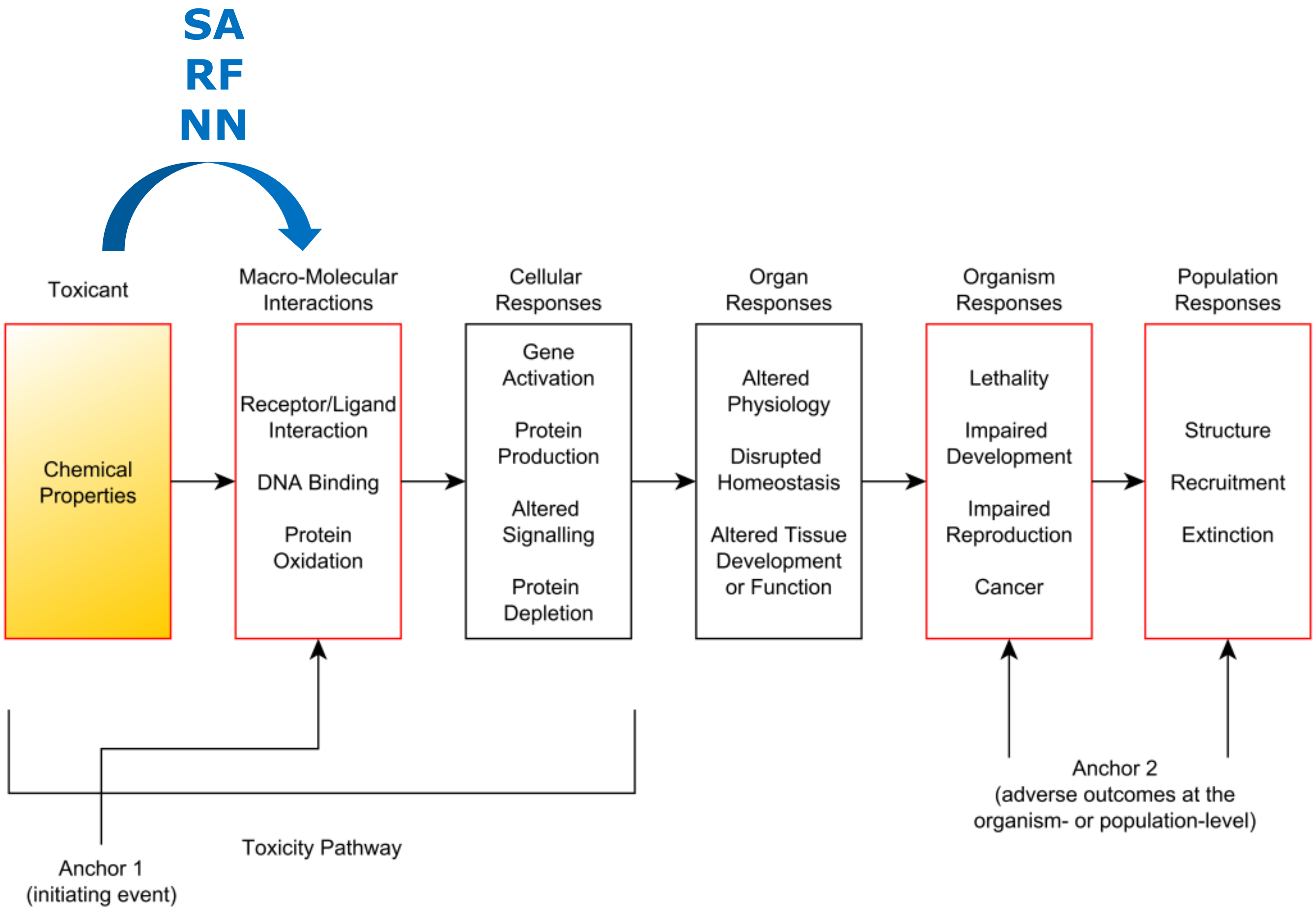
Toxicant	Macro-Molecular Interactions	Cellular Responses	Organ Responses	Organism Responses	Population Responses
Chemical Properties	Receptor/Ligand Interaction DNA Binding Protein Oxidation	Gene Activation Protein Production Altered Signalling Protein Depletion	Altered Physiology Disrupted Homeostasis Altered Tissue Development or Function	Lethality Impaired Development Impaired Reproduction Cancer	Structure Recruitment Extinction

Anchor 2
(adverse outcomes at the
organism- or population-level)

Toxicity Pathway

Towards a Quantitative AOP

AOP Modelling



SA
RF
NN

Toxicant
Chemical Properties

Macro-Molecular Interactions
Receptor/Ligand Interaction
DNA Binding
Protein Oxidation

Cellular Responses
Gene Activation
Protein Production
Altered Signalling
Protein Depletion

Organ Responses
Altered Physiology
Disrupted Homeostasis
Altered Tissue Development or Function

Organism Responses
Lethality
Impaired Development
Impaired Reproduction
Cancer

Population Responses
Structure
Recruitment
Extinction

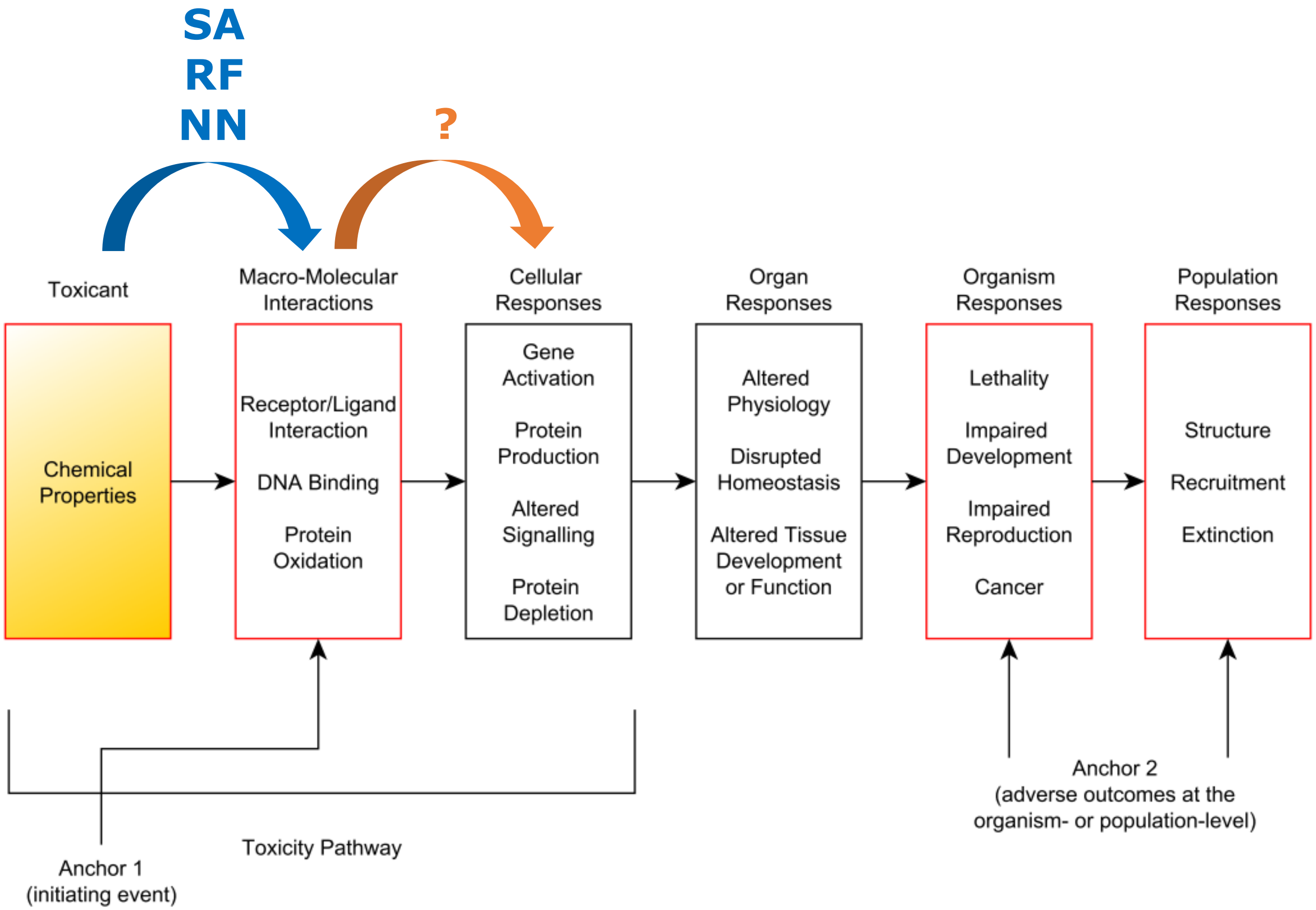
Anchor 1
(initiating event)

Toxicity Pathway

Anchor 2
(adverse outcomes at the organism- or population-level)

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

AOP Modelling



SA
RF
NN

?

Toxicant

Chemical Properties

Macro-Molecular Interactions

Receptor/Ligand Interaction

DNA Binding

Protein Oxidation

Cellular Responses

Gene Activation

Protein Production

Altered Signalling

Protein Depletion

Organ Responses

Altered Physiology

Disrupted Homeostasis

Altered Tissue Development or Function

Organism Responses

Lethality

Impaired Development

Impaired Reproduction

Cancer

Population Responses

Structure

Recruitment

Extinction

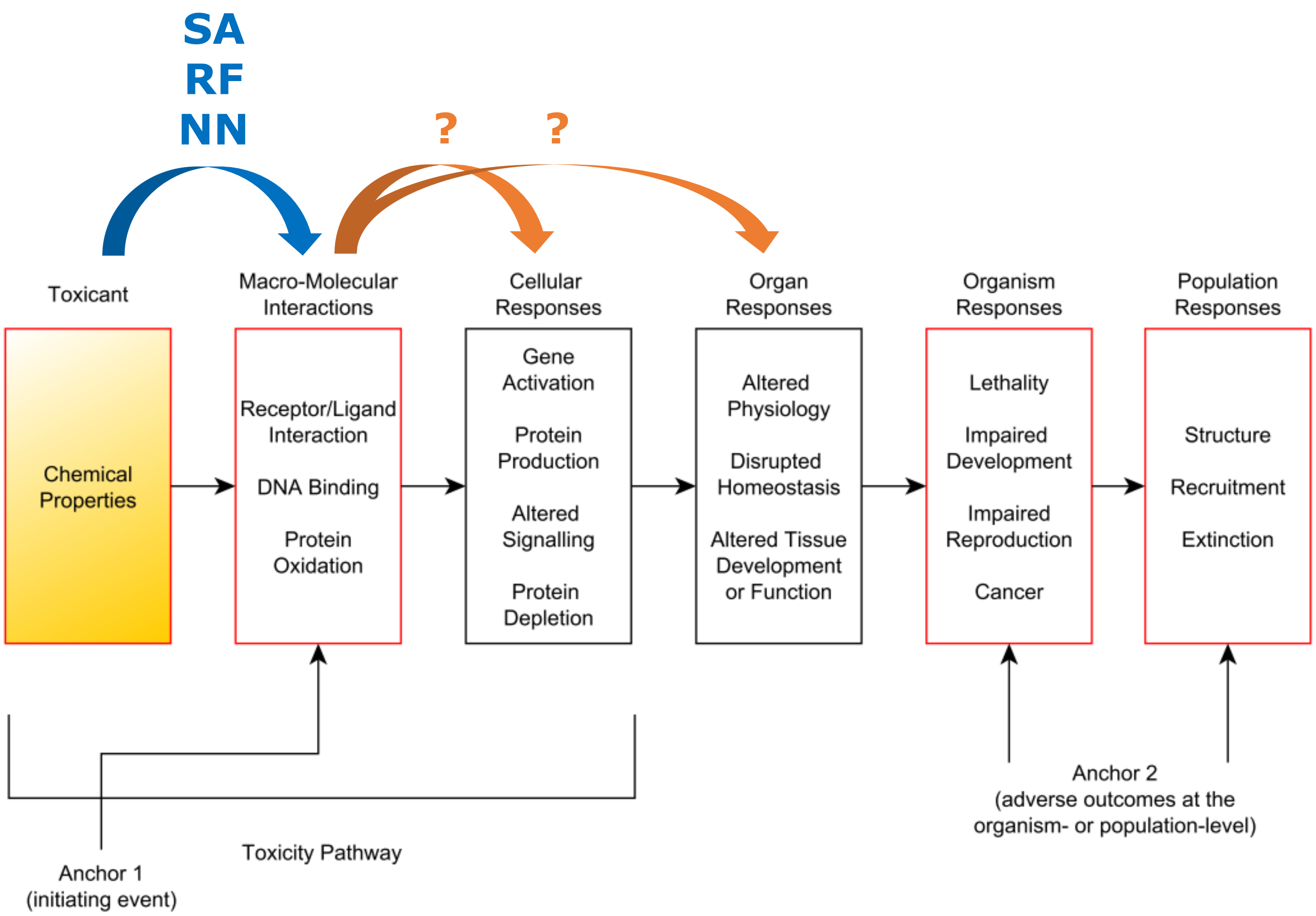
Toxicity Pathway

Anchor 1
(initiating event)

Anchor 2
(adverse outcomes at the organism- or population-level)

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

AOP Modelling



SA
RF
NN

? **?**

Toxicant
Chemical Properties

Macro-Molecular Interactions
Receptor/Ligand Interaction
DNA Binding
Protein Oxidation

Cellular Responses
Gene Activation
Protein Production
Altered Signalling
Protein Depletion

Organ Responses
Altered Physiology
Disrupted Homeostasis
Altered Tissue Development or Function

Organism Responses
Lethality
Impaired Development
Impaired Reproduction
Cancer

Population Responses
Structure
Recruitment
Extinction

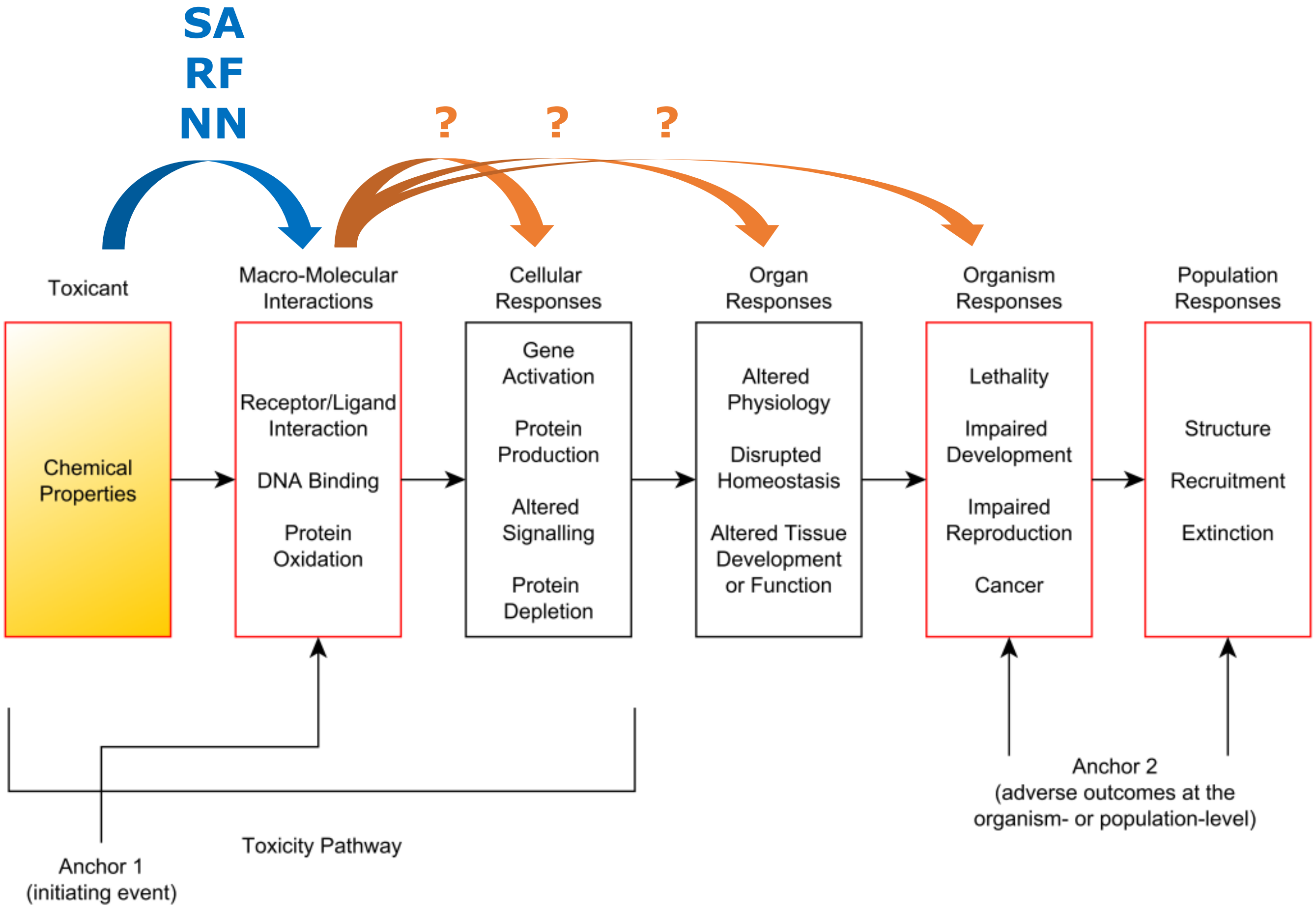
Anchor 1
(initiating event)

Toxicity Pathway

Anchor 2
(adverse outcomes at the organism- or population-level)

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

AOP Modelling

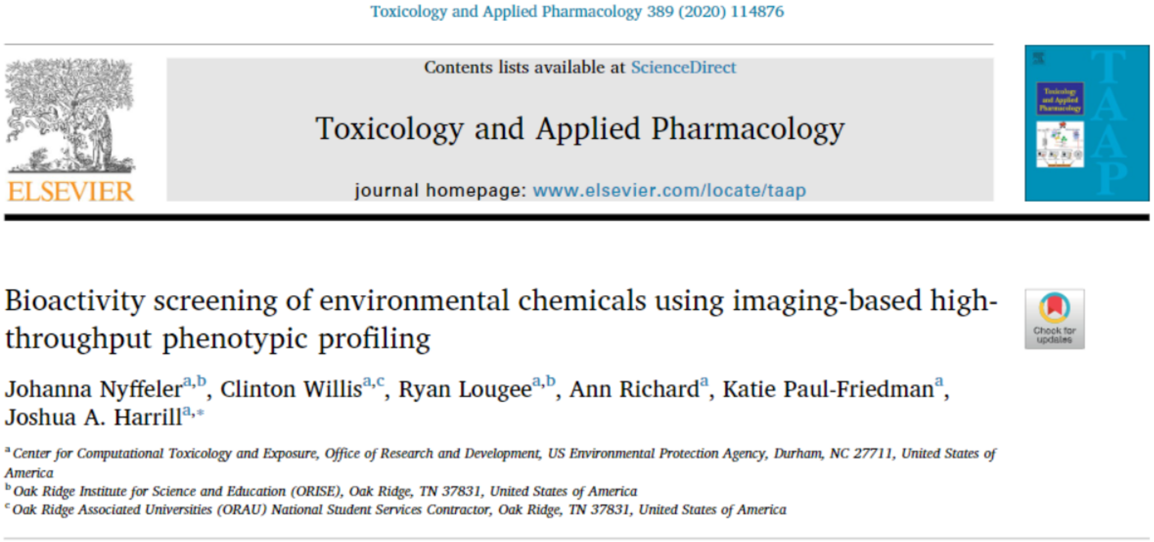


SA
RF
NN

? ? ?

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

Nyffeler Phenotypic Profiling Study



Toxicology and Applied Pharmacology 389 (2020) 114876

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/taap

ELSEVIER

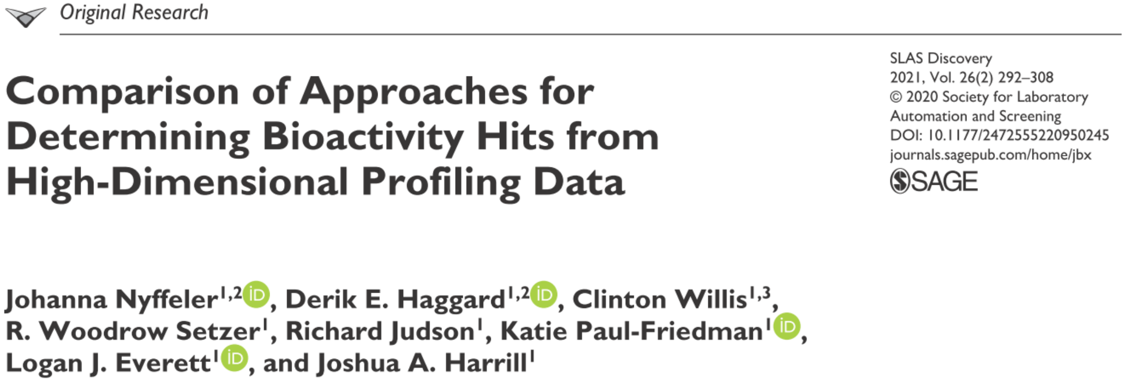
Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling

Johanna Nyffeler[a,b], Clinton Willis[a,c], Ryan Lougee[a,b], Ann Richard[a], Katie Paul-Friedman[a], Joshua A. Harrill[a,*]

[a] Center for Computational Toxicology and Exposure, Office of Research and Development, US Environmental Protection Agency, Durham, NC 27711, United States of America

[b] Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN 37831, United States of America

[c] Oak Ridge Associated Universities (ORAU) National Student Services Contractor, Oak Ridge, TN 37831, United States of America



Original Research

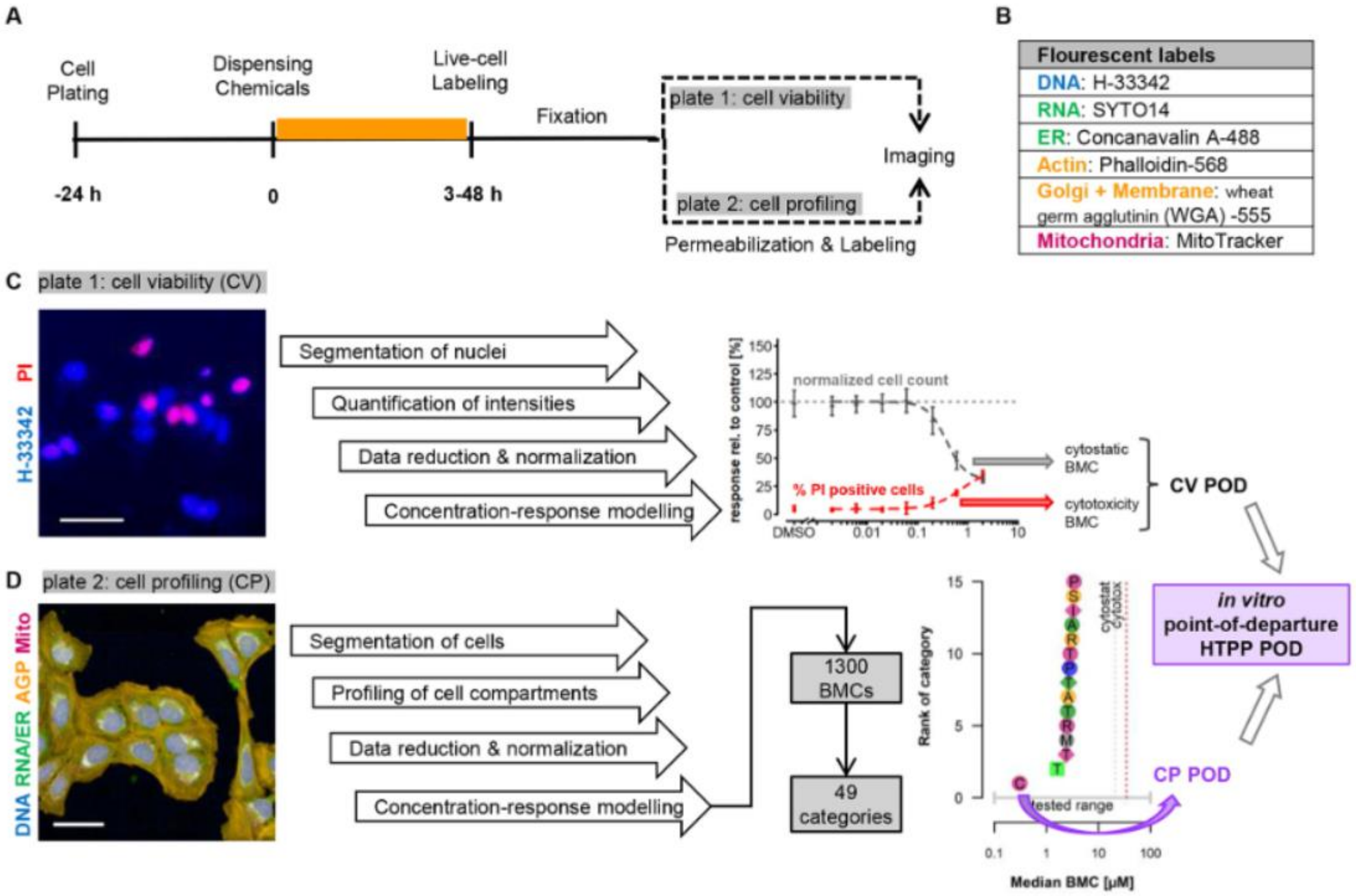
SLAS Discovery
2021, Vol. 26(2) 292–308
© 2020 Society for Laboratory Automation and Screening
DOI: 10.1177/2472555220950245
journals.sagepub.com/home/jbx
SAGE

Comparison of Approaches for Determining Bioactivity Hits from High-Dimensional Profiling Data

Johanna Nyffeler[1,2], Derik E. Haggard[1,2], Clinton Willis[1,3], R. Woodrow Setzer[1], Richard Judson[1], Katie Paul-Friedman[1], Logan J. Everett[1], and Joshua A. Harrill[1]

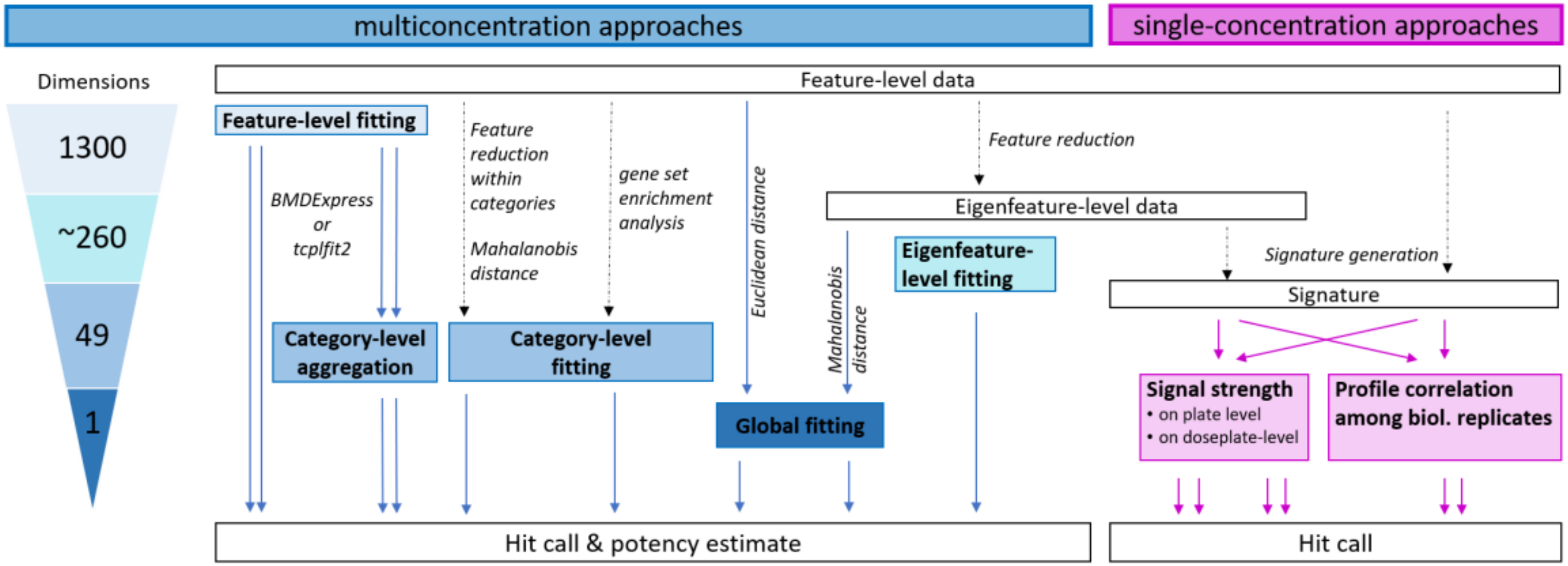
Nyffeler, J., *et al*. (**2020**) *Toxicol. Appl. Pharmacol.*, 389; 114876.
Nyffeler, J., *et al*. (**2021**) *SLAS Discov.*, 26; 292.

Overall Procedure



Nyffeler, J., *et al*. (**2020**) *Toxicol. Appl. Pharmacol.*, 389; 114876.

Data Processing



multiconcentration approaches

single-concentration approaches

Dimensions

1300

~260

49

1

Feature-level data

Feature-level fitting

BMDExpress or tcplfit2

Feature reduction within categories

Mahalanobis distance

gene set enrichment analysis

Euclidean distance

Feature reduction

Eigenfeature-level data

Mahalanobis distance

Eigenfeature-level fitting

Signature generation

Signature

Category-level aggregation

Category-level fitting

Global fitting

Signal strength
- on plate level
- on doseplate-level

Profile correlation among biol. replicates

Hit call & potency estimate

Hit call

Nyffeler, J., *et al*. (**2021**) *SLAS Discov.*, 26; 292.

BMC Data Distribution

Column	Category	% Positive	Column	Category	% Positive
1	AGP_Axial_Cells	35.9	26	ER_Profile_Cytoplasm	35.5
2	AGP_Compactness_Cells	30.7	27	ER_Radial_Cells	38.1
3	AGP_Intensity_Cytoplasm	33.6	28	ER_Symmetry_Cells	29.6
4	AGP_Intensity_Membrane	17.6	29	ER_Texture_Cytoplasm	38.0
5	AGP_Intensity_Ring	31.3	30	ER_Texture_Ring	36.3
6	AGP_Profile_Cytoplasm	30.9	31	Mito_Axial_Cells	33.5
7	AGP_Profile_Nuclei	26.3	32	Mito_Compactness_Cells	44.3
8	AGP_Radial_Cells	34.9	33	Mito_Intensity_Cytoplasm	37.7
9	AGP_Symmetry_Cells	23.5	34	Mito_Intensity_Ring	42.5
10	AGP_Texture_Cytoplasm	31.2	35	Mito_Profile_Cytoplasm	37.8
11	AGP_Texture_Membrane	20.1	36	Mito_Profile_Nuclei	15.5
12	AGP_Texture_Ring	32.8	37	Mito_Radial_Cells	42.5
13	DNA_Axial_Nuclei	40.1	38	Mito_Symmetry_Cells	32.9
14	DNA_Compactness_Nuclei	43.0	39	Mito_Texture_Cytoplasm	42.6
15	DNA_Intensity_Nuclei	24.8	40	Mito_Texture_Ring	44.8
16	DNA_Profile_Cytoplasm	40.7	41	Position	34.7
17	DNA_Profile_Nuclei	39.3	42	RNA_Axial_Nuclei	32.7
18	DNA_Radial_Cells	44.6	43	RNA_Compactness_Nuclei	33.7
19	DNA_Radial_Nuclei	39.2	44	RNA_Intensity_Nuclei	7.4
20	DNA_Symmetry_Nuclei	35.7	45	RNA_Profile_Nuclei	31.2
21	DNA_Texture_Nuclei	43.0	46	RNA_Radial_Nuclei	40.7
22	ER_Axial_Cells	32.2	47	RNA_Symmetry_Nuclei	28.4
23	ER_Compactness_Cells	41.0	48	RNA_Texture_Nuclei	26.5
24	ER_Intensity_Cytoplasm	30.8	49	Shape	42.5
25	ER_Intensity_Ring	29.6			

BMC Classification Modelling

- NN and RF models were trained using three molecular representations (ECFP, MACCS keys and PhysChem descriptors). Results are shown below:

		Validation Set				Test Set			
Model	**Architecture**	SE	SP	ACC	MCC	SE	SP	ACC	MCC
NN-ECFP	**100100**	26.5	91.9	67.4	0.249	30.0	91.0	66.9	0.272
RF-ECFP	**100**	97.7	11.0	43.5	0.158	94.8	12.9	45.3	0.127
NN-MACCS	**100100**	49.8	87.4	73.3	0.408	50.4	78.0	67.1	0.295
RF-MACCS	**10**	95.9	19.3	48.0	0.214	93.1	17.3	47.3	0.150
NN-PhysChem	**100**	40.3	84.6	68.0	0.279	38.3	84.8	66.3	0.262
RF-PhysChem	**10**	92.6	20.8	47.6	0.177	90.8	18.6	47.2	0.129

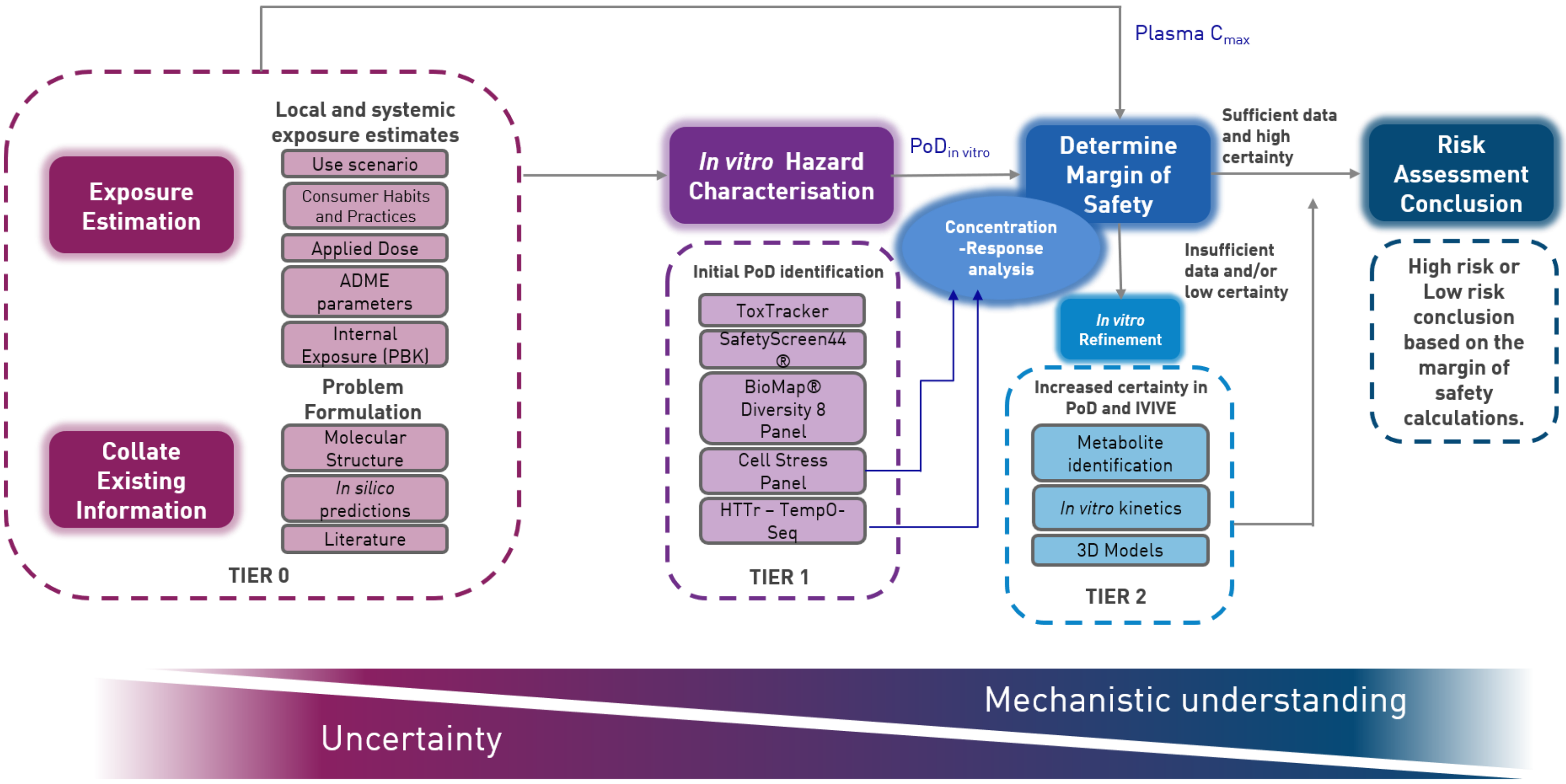
Conclusions

- Further exploration of biological responses through AOPs remains a challenge qualitatively and quantitatively

- Data availability makes this a challenge for *in silico* modelling

- Computational prediction of phenotypic cellular responses may be able to feed into a weight of evidence response

Ab Initio NGRA Framework



Plasma C_{max}

Local and systemic exposure estimates

Exposure Estimation

- Use scenario
- Consumer Habits and Practices
- Applied Dose
- ADME parameters
- Internal Exposure (PBK)

Problem Formulation

Collate Existing Information

- Molecular Structure
- *In silico* predictions
- Literature

TIER 0

***In vitro* Hazard Characterisation**

$PoD_{in\ vitro}$

Initial PoD identification

- ToxTracker
- SafetyScreen44®
- BioMap® Diversity 8 Panel
- Cell Stress Panel
- HTTr – TempO-Seq

TIER 1

Concentration-Response analysis

Determine Margin of Safety

Sufficient data and high certainty

Insufficient data and/or low certainty

***In vitro* Refinement**

Increased certainty in PoD and IVIVE

- Metabolite identification
- *In vitro* kinetics
- 3D Models

TIER 2

Risk Assessment Conclusion

High risk or Low risk conclusion based on the margin of safety calculations.

Uncertainty

Mechanistic understanding

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

In Silico Approaches to Link Adverse Outcomes to Molecular Initiating Events through AOPs

Compound → Molecular Initiating Event → Adverse Outcome Pathway

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

WC11 – 26th August 2021

MRC | Medical Research Council

MRC Toxicology Unit

UNIVERSITY OF CAMBRIDGE