

Supporting Information

Modelling *in vitro* mutagenicity using multi-task deep learning and REACH data

PanagiotisG. Karamertzanis^{†*}, Mike Rasenberg[‡], Imran Shah[‡], Grace Patlewicz[‡]

[†] European Chemicals Agency (ECHA), Telakkakatu 6, 00150 Helsinki, Finland

[‡] Center for Computational Toxicology and Exposure (CCTE), US Environmental Protection Agency,
109 TW Alexander Dr, Research Triangle Park, NC 27711, USA

*Corresponding author:

Email address: panagiotis.karamertzanis@echa.europa.eu

The supporting information contains the molecular structure curation procedure, additional information on the external (hold out) test set validation and applicability domain considerations.

Molecular structure curation

The quality of the training data sets an upper bound to the predictive performance of a model. There may be errors in the measured quantity, but also in the structural representation of the input structure. A further complexity arises from the fact that there is no unique representation of molecular structures. Two structurally similar molecules may elicit similar toxicological responses, but the model may not be able to associate the toxicity with the common structural moieties, unless the two structures have been consistently standardised prior to the descriptor calculation or the representation of the molecules as graphs. Complications may, for example, arise by the different representation of nitro groups, tautomers and salts.

Building on the seminal review by Fourches et al.¹ on the importance of structure curation in cheminformatics and modelling, numerous studies and publicly available tools have emerged in recent years^{2–7} for both structure checking and curation. In this work, experimental data were retrieved from the QSAR Toolbox and EPA's DSSTox⁸, both of which have applied some structure checking and standardisation. Hence, a minimal workflow was used. Structures were first checked for covalently bonded alkali metals and had such metals split into their own fragments. The structures were then passed through the RDKit^{3,9} standardiser, followed by the removal of toxicologically insignificant fragments (see repository¹⁰). The remaining fragments were subsequently transformed using a set of tautomerism reactions taken from the QSAR-ready^{6,11} workflow that has been used extensively in international collaborative projects^{12–14}. The reactions were applied recursively until no further transformations were possible. Following tautomerism, structures were cleaned from all stereoisomerism information, including ring cis/trans isomerism, and were finally neutralised, e.g. to convert carboxylates to their neutral acid equivalent. Finally, a set of checks has been applied to ensure that only structures suitable for modelling are kept. In particular, we only allowed atoms C, O, N, Cl, S, F, Br, P, B, Si, I, and H, imposed a minimum of one carbon atom, a minimum of one bond, only one fragment, only single, double, triple and aromatic bond types, and a maximum molecular weight of 1000. Atom hybridisations could only be unspecified, sp², sp³ or sp, and molecules could not have an overall charge. The latter filter removed a small number of structures for which the RDKit^{3,9} neutralisation algorithm failed to remove the charge. The structure cleaning and standardisation workflow was applied equally to both the training and external (hold-out) test sets.

External (hold-out) test set validation and model comparison

This section contains the detailed external (hold-out) test set validation results for the FFNN, GCN and GAT single and multi-task models at the aggregated endpoint level. The main manuscript contains the results for AttentiveFP (Table 7).

Table S1 External (hold-out) test set validation for the FFNN single- and multi-task models; the metrics in bold correspond to 200 or more samples and can be indicative of the model generalisation error

FFNN model	Task	Source	Number of structures	Sensitivity	Specificity	Balanced accuracy	AUC
Ames (ST)	Ames	ISSTY ¹	666 (310 positives)	0.666 ± 0.029	0.771 ± 0.014	0.718 ± 0.008	0.773 ± 0.004
GM (ST)	GM	ToxValDB ²	665 (131 positives)	0.723 ± 0.004	0.812 ± 0.009	0.767 ± 0.002	0.853 ± 0.003
GM (ST)	GM	NTP ³	95 (77 positives)	0.688 ± 0.013	0.833 ± 0.000	0.761 ± 0.006	0.823 ± 0.003
CA (ST)	CA	ToxValDB	889 (323 positives)	0.612 ± 0.034	0.751 ± 0.032	0.681 ± 0.012	0.748 ± 0.009
MN (ST)	MN	ToxValDB ²	453 (162 positives)	0.533 ± 0.029	0.685 ± 0.011	0.609 ± 0.009	0.650 ± 0.003
MN (ST)	MN	Baderna 2020 ⁴	144 (65 positives)	0.528 ± 0.009	0.823 ± 0.022	0.674 ± 0.014	0.757 ± 0.003
Ames-GM-CA-MN (MT)	Ames	ISSTY ¹	539 (272 positives)	0.719 ± 0.004	0.727 ± 0.013	0.723 ± 0.004	0.783 ± 0.004
	GM	ToxValDB ²	223 (36 positives)	0.713 ± 0.016	0.818 ± 0.005	0.766 ± 0.008	0.848 ± 0.001
	GM	NTP ³	6 (5 positives)	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
	CA	ToxValDB ²	311 (98 positives)	0.656 ± 0.016	0.803 ± 0.014	0.730 ± 0.002	0.811 ± 0.004
	MN	ToxValDB ²	57 (11 positives)	0.636 ± 0.000	0.928 ± 0.033	0.782 ± 0.017	0.818 ± 0.002
	MN	Baderna 2020 ⁴	42 (25 positives)	0.400 ± 0.000	0.863 ± 0.034	0.631 ± 0.017	0.644 ± 0.019

¹ data taken from the ISSTY database¹⁵ extracted from the QSAR Toolbox

² data taken from ToxValDB¹⁶

³ data taken from the mammalian cell mutagenicity data collection of the National Toxicology Program¹⁷

⁴ Baderna et al. dataset¹⁸

⁵ the reported uncertainty is the standard deviation from the predictions obtained from the three models fitted to the whole training set with the optimal hyperparameters

Table S2 External (hold-out) test set validation for the GAT single- and multi-task models; the metrics in bold correspond to 200 or more samples and can be indicative of the model generalisation error

GAT model	Task	Source	Number of structures (positives)	Sensitivity	Specificity	Balanced accuracy	AUC
Ames (ST)	Ames	ISSTY ¹	666 (310 positives)	0.745 ± 0.022	0.723 ± 0.036	0.734 ± 0.013	0.792 ± 0.011
GM (ST)	GM	ToxValDB ²	665 (131 positives)	0.743 ± 0.074	0.777 ± 0.098	0.760 ± 0.017	0.830 ± 0.013
GM (ST)	GM	NTP ³	95 (77 positives)	0.745 ± 0.059	0.759 ± 0.085	0.752 ± 0.014	0.804 ± 0.020
CA (ST)	CA	ToxValDB	889 (323 positives)	0.596 ± 0.067	0.724 ± 0.060	0.660 ± 0.005	0.732 ± 0.006
MN (ST)	MN	ToxValDB ²	453 (162 positives)	0.551 ± 0.085	0.643 ± 0.077	0.597 ± 0.005	0.638 ± 0.005
MN (ST)	MN	Baderna 2020 ⁴	144 (65 positives)	0.523 ± 0.148	0.755 ± 0.082	0.639 ± 0.037	0.741 ± 0.004
Ames-GM-CA-MN (MT)	Ames	ISSTY ¹	539 (272 positives)	0.755 ± 0.022	0.738 ± 0.021	0.746 ± 0.018	0.804 ± 0.008
	GM	ToxValDB ²	223 (36 positives)	0.713 ± 0.032	0.802 ± 0.009	0.758 ± 0.019	0.791 ± 0.022

	GM	NTP ³	6 (5 positives)	0.933 ± 0.115	1.000 ± 0.000	0.967 ± 0.058	1.000 ± 0.000
	CA	ToxValDB ²	311 (98 positives)	0.633 ± 0.084	0.792 ± 0.063	0.712 ± 0.017	0.783 ± 0.003
	MN	ToxValDB ²	57 (11 positives)	0.576 ± 0.139	0.877 ± 0.055	0.726 ± 0.080	0.765 ± 0.036
	MN	Baderna 2020 ⁴	42 (25 positives)	0.587 ± 0.061	0.725 ± 0.068	0.656 ± 0.064	0.662 ± 0.025

¹ data taken from the ISSTY database¹⁵ extracted from the QSAR Toolbox
² data taken from ToxValDB¹⁶
³ data taken from the mammalian cell mutagenicity data collection of the National Toxicology Program¹⁷
⁴ Baderna et al. dataset¹⁸
⁵ the reported uncertainty is the standard deviation from the predictions obtained from the three models fitted to the whole training set with the optimal hyperparameters

Table S3 External (hold-out) test set validation for the GCN single- and multi-task models; the metrics in bold correspond to 200 or more samples and can be indicative of the model generalisation error

GCN model	Task	Source	Number of structures (positives)	Sensitivity	Specificity	Balanced accuracy	AUC
Ames (ST)	Ames	ISSTY ¹	666 (310 positives)	0.782 ± 0.013	0.689 ± 0.041	0.735 ± 0.015	0.799 ± 0.004
GM (ST)	GM	ToxValDB ²	665 (131 positives)	0.761 ± 0.034	0.775 ± 0.040	0.768 ± 0.015	0.837 ± 0.018
GM (ST)	GM	NTP ³	95 (77 positives)	0.749 ± 0.020	0.667 ± 0.056	0.708 ± 0.032	0.774 ± 0.010
CA (ST)	CA	ToxValDB	889 (323 positives)	0.492 ± 0.067	0.796 ± 0.041	0.644 ± 0.013	0.714 ± 0.002
MN (ST)	MN	ToxValDB ²	453 (162 positives)	0.597 ± 0.036	0.611 ± 0.040	0.604 ± 0.002	0.624 ± 0.005
MN (ST)	MN	Baderna 2020 ⁴	144 (65 positives)	0.533 ± 0.018	0.814 ± 0.019	0.674 ± 0.007	0.760 ± 0.035
Ames-GM-CA-MN (MT)	Ames	ISSTY ¹	539 (272 positives)	0.700 ± 0.113	0.767 ± 0.067	0.733 ± 0.023	0.801 ± 0.011
	GM	ToxValDB ²	223 (36 positives)	0.694 ± 0.028	0.863 ± 0.013	0.779 ± 0.007	0.828 ± 0.004
	GM	NTP ³	6 (5 positives)	0.867 ± 0.231	1.000 ± 0.000	0.933 ± 0.115	1.000 ± 0.000
	CA	ToxValDB ²	311 (98 positives)	0.602 ± 0.037	0.798 ± 0.033	0.700 ± 0.026	0.749 ± 0.032
	MN	ToxValDB ²	57 (11 positives)	0.545 ± 0.182	0.870 ± 0.058	0.708 ± 0.090	0.801 ± 0.050
	MN	Baderna 2020 ⁴	42 (25 positives)	0.547 ± 0.061	0.608 ± 0.090	0.577 ± 0.066	0.643 ± 0.046

¹ data taken from the ISSTY database¹⁵ extracted from the QSAR Toolbox
² data taken from ToxValDB¹⁶
³ data taken from the mammalian cell mutagenicity data collection of the National Toxicology Program¹⁷
⁴ Baderna et al. dataset¹⁸
⁵ the reported uncertainty is the standard deviation from the predictions obtained from the three models fitted to the whole training set with the optimal hyperparameters

The training set for our chromosome aberration models had 2774 structures compared to 11510 for Ames, and hence the poorer performance can be attributed to the smaller training dataset. However, it is likely that the quality of the dataset is also lower. Honda et al.^{19,20} demonstrated that re-evaluation of the chromosome aberration test results using improved cytotoxicity indices reduces the false-positive rate. The authors re-evaluated 129 substances out of which 39 were positive in the in vitro chromosome aberration test, and found that 11 need to be re-categorized as negative without the occurrence of an equal number of false negatives. It is not known if the re-categorization of 20-25% of positives would apply uniformly in the whole training set, as some of the tests are more recent, but

it could partially explain the fact that from all of our models the chromosome aberration had the lowest cross-validation sensitivity as shown in **Error! Reference source not found.** All of the 11 substances (acenaphthene, 2-(4-morpholinyldithio)benzothiazole, 4,4'-sulfonyldiphenol, 4-methylbenzoic acid, 2,3,6-trimethylphenol, 1,3-bis(2-methylphenyl)guanidine, 2,4-di-tert-butylphenol, o-acetoacetotoluidine, 2,2,6,6-tetramethyl-4-hydroxypiperidine, 3-methylbenzoic acid and thymol) are present in our training set as positive in chromosome aberration, whilst the AttentiveFP 4-task model consistently predicts them as genotoxicants for this assay. Hence, our models may have been affected by the presence of false positives in the experimental data. A retrospective analysis of chromosome aberration assay data could enhance the development of predictive methods and can be facilitated by the availability of robust study summaries under REACH, with models assisting in identifying discrepancies.

Applicability domain considerations

The external test set performance metrics are comparable with the state-of-the-art models but lower than the cross-validation ones, partially because a number of structures in the external (hold-out) test set do not have structural analogues in the training set. If we were to apply an applicability domain then a number of structures would not be predicted, but the performance metrics for the ones remaining would improve. As an illustration, we took the gene mutation in bacteria predictions of the AttentiveFP 4-task model and compared them with the ISSTY external (hold-out) test set. For each structure in ISSTY that is not present in the training set, we computed the maximum Tanimoto similarity to all structures in the training set. Essentially, this provides a rudimentary measure of the availability of structural analogues in the training set for each predicted structure, without accounting for the local data density and local performance accuracy. The Tanimoto similarity was computed using binary Morgan fingerprints with radius 2 and length 2048 and the performance metrics on the external (hold-out) test set were calculated for various similarity thresholds. As the similarity threshold increases fewer and fewer from the 539 structures (272 positives) are considered within the applicability domain. Even with a modest similarity threshold of 0.5 approximately 40% of the structures in external (hold-out) test set would be considered as structurally different. On the other hand, all performance metrics increase, with the balanced accuracy being roughly equal to the cross-validation one for a similarity threshold of 0.7. No metrics were computed for 0.9 similarity threshold because of the low number of structures that are all negative. These results suggest that the ISSTY database would ideally have been included in the training set as it contains structural moieties not generally represented in the training set we used. This would be advisable for fitting the most accurate and generalisable model. It would not be meaningful to repeat this investigation for the other in vitro

assays because the size of the external (hold-out) test sets is not sufficiently large. Reducing the number of predicted structures would simply increase further the variance of the metrics.

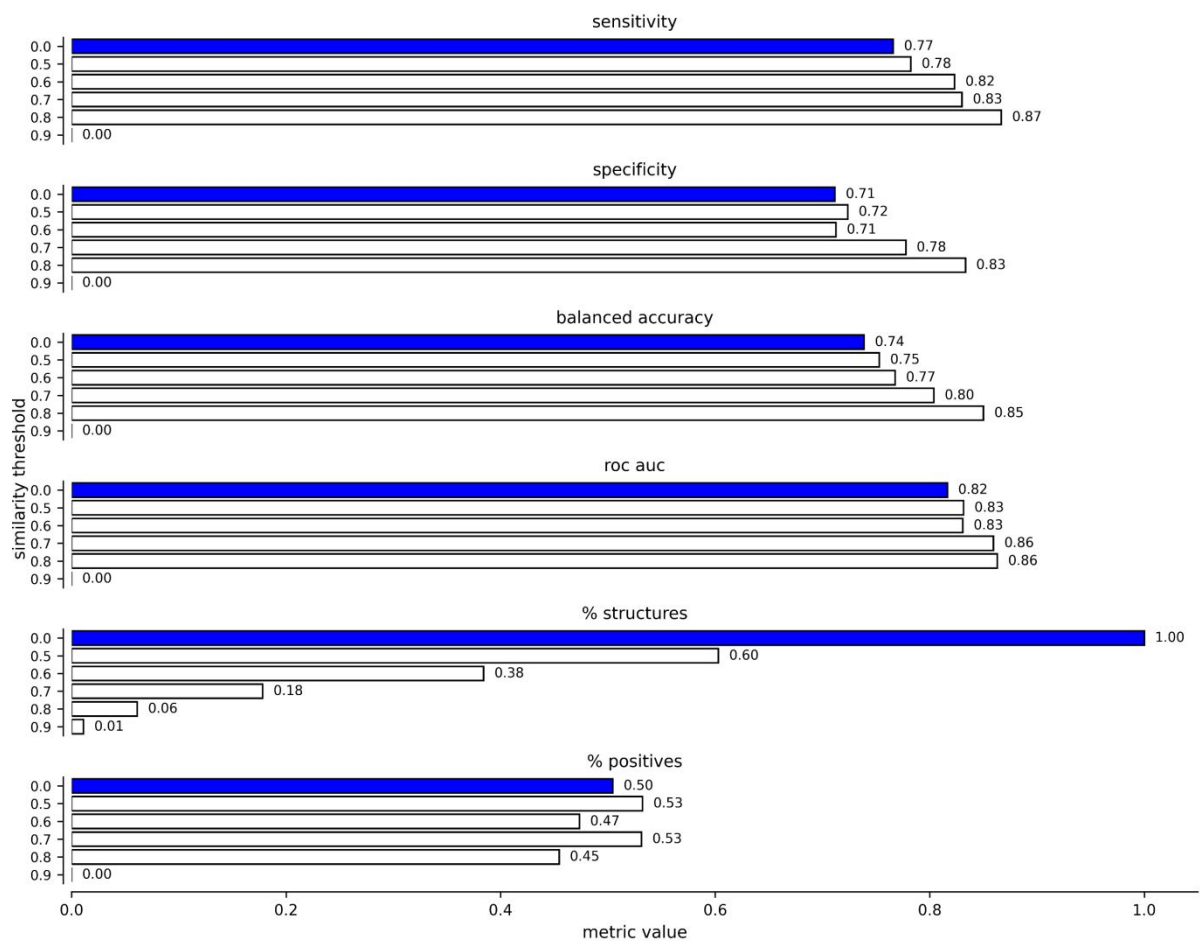


Figure S1 Effect of the similarity threshold on the performance metrics for bacterial gene mutation based on the ISSTY external (hold-out) test set and the AttentiveFP 4-task model. The number of structures ranges from 539 (no similarity threshold) to 6 all of which are negative that does not allow metrics to be computed (0.9 similarity threshold). The bar in blue indicates the pessimistic performance metrics reported in the paper.

Supporting information datasets

In addition to this document, the supporting information contains the following material:

- 109 • S0_S0_GNN_mutagenicity_models_SI_revised.docx (this document)
- 110 Supporting information with the methodology to curate molecular structures, additional
- 111 external (hold-out) test set validation and model comparison results, and applicability domain
- 112 considerations.
- 113 • S1_training_set_4_task_public.xlsx
- 114 Training set for the aggregated (endpoint level) models for gene mutation in bacteria, and
- 115 gene mutation, chromosome aberration and micronucleus in mammalian cells
- 116 • S2_training_set_10_task_public.xlsx
- 117 Training set for the AMES models for individual strain/metabolic activation combinations
- 118 • S3_structural_alerts_coverage_AMES_agg_GM_CA_MN_early_stopping_public.xlsx
- 119 Structural alerts for the structures in the training set of the aggregated (endpoint level) models
- 120 for gene mutation in bacteria, and gene mutation, chromosome aberration and micronucleus
- 121 in mammalian cells
- 122 • S4_RF.zip
- 123 Detailed cross-validation results for the classical machine learning models
- 124 • S5_deep_learning
- 125 Detailed hyperparameter tuning data for the deep learning models
- 126 • S6_compare_predictive_performance_AMES_agg_GM_CA_MN.xlsx
- 127 Model comparison based on cross-validation and external (hold-out) test set validation of the
- 128 aggregated (endpoint level) models for gene mutation in bacteria, and gene mutation,
- 129 chromosome aberration and micronucleus in mammalian cells
- 130 • S7_compare_predictive_performance_AMES_5_strains.xlsx
- 131 Model comparison based on cross-validation and external (hold-out) test set validation of the
- 132 AMES models for individual strain/metabolic activation combinations
- 133 • S8_test_set_AMES_ISSTY.xlsx
- 134 Test set for the AMES models for individual strain/metabolic activation combinations
- 135 originating from the ISSTY database
- 136 • S9_test_set_AMES_aggregated_ISSTY.xlsx
- 137 Test set for the aggregated (endpoint level) models for gene mutation in bacteria originating
- 138 from the ISSTY database
- 139 • S10_test_set_ToxValDB.xlsx
- 140 Test set for the aggregated (endpoint level) models for gene mutation, chromosome
- 141 aberration and micronucleus in mammalian cells originating from the ToxValDB database
- 142 • S11_test_set_NTP.xlsx
- 143 Test set for the aggregated (endpoint level) models for gene mutation in mammalian cells
- 144 originating from the NTP database
- 145 • S12_test_set_Baderna.xlsx
- 146 Test set for the aggregated (endpoint level) models for micronucleus in mammalian cells
- 147 originating from the Baderna et al. dataset
- 148
- 149

References

- (1) Fourches, D.; Muratov, E.; Tropsha, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J Chem Inf Model* 2010, 50 (7), 1189–1204. <https://doi.org/10.1021/ci100176x>.
- (2) Dolciami, D.; Villasclaras-Fernandez, E.; Kannas, C.; Meniconi, M.; Al-Lazikani, B.; Antolin, A. A. CanSAR Chemistry Registration and Standardization Pipeline. *J Cheminform* 2022, 14 (1), 28. <https://doi.org/10.1186/s13321-022-00606-7>.
- (3) Bento, A. P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An Open Source Chemical Structure Curation Pipeline Using RDKit. *J Cheminform* 2020, 12 (1), 51. <https://doi.org/10.1186/s13321-020-00456-1>.
- (4) Karapetyan, K.; Batchelor, C.; Sharpe, D.; Tkachenko, V.; Williams, A. J. The Chemical Validation and Standardization Platform (CVSP): Large-Scale Automated Validation of Chemical Structure Datasets. *J Cheminform* 2015, 7 (1), 30. <https://doi.org/10.1186/s13321-015-0072-8>.
- (5) Hähnke, V. D.; Kim, S.; Bolton, E. E. PubChem Chemical Structure Standardization. *J Cheminform* 2018, 10 (1), 36. <https://doi.org/10.1186/s13321-018-0293-8>.
- (6) Mansouri, K.; Moreira-Filho, J. T.; Lowe, C. N.; Charest, N.; Martin, T.; Tkachenko, V.; Judson, R.; Conway, M.; Kleinstreuer, N. C.; Williams, A. J. Free and Open-Source QSAR-Ready Workflow for Automated Standardization of Chemical Structures in Support of QSAR Modeling. *J Cheminform* 2024, 16 (1), 19. <https://doi.org/10.1186/s13321-024-00814-3>.
- (7) *mcs07/MolVS: Molecule Validation and Standardization*. <https://github.com/mcs07/MolVS> (accessed 2024-08-12).
- (8) Grulke, C. M.; Williams, A. J.; Thillanadarajah, I.; Richard, A. M. EPA's DSSTox Database: History of Development of a Curated Chemistry Resource Supporting Computational Toxicology Research. *Computational Toxicology* 2019, 12, 100096. <https://doi.org/10.1016/j.comtox.2019.100096>.
- (9) Landrum, G.; Tosco, P.; Kelley, B.; Ric; sriniker; gedeck; Cosgrove, D.; Vianello, R.; NadineSchneider; Kawashima, E.; N, D.; Dalke, A.; Jones, G.; Cole, B.; Swain, M.; Turk, S.; AlexanderSavelyev; Vaucher, A.; Wójcikowski, M.; Take, I.; Probst, D.; Scalfani, V. F.; Ujihara, K.; godin, guillaume; Pahl, A.; Berenger, F.; JLVarjo; jasondbiggs; strets123; JP. Rdkit/Rdkit. February 23, 2024. <https://zenodo.org/records/10633624> (accessed 2024-02-08).
- (10) *pkaramertzanis/GNN*. <https://github.com/pkaramertzanis/GNN> (accessed 2024-09-07).
- (11) *Standardization workflow for QSAR-ready chemical structures pretreatment. – KNIME Community Hub*. https://hub.knime.com/kmansouri/spaces/Public/QSAR-ready_2.5.8~5TRvnGfMJsgTkcZu/current-state (accessed 2024-08-13).
- (12) Mansouri, K.; Abdelaziz, A.; Rybacka, A.; Roncaglioni, A.; Tropsha, A.; Varnek, A.; Zakharov, A.; Worth, A.; Richard, A. M.; Grulke, C. M.; Trisciuzzi, D.; Fourches, D.; Horvath, D.; Benfenati, E.; Muratov, E.; Wedebye, E. B.; Grisoni, F.; Mangiatordi, G. F.; Incisivo, G. M.; Hong, H.; Ng, H. W.; Tetko, I. V.; Balabin, I.; Kancherla, J.; Shen, J.; Burton, J.; Nicklaus, M.; Cassotti, M.; Nikolov, N. G.; Nicolotti, O.; Andersson, P. L.; Zang, Q.; Politi, R.; Begger, R. D.; Todeschini, R.; Huang, R.; Farag, S.; Rosenberg, S. A.; Slavov, S.; Hu, X.; Judson, R. S. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. *Environ Health Perspect* 2016, 124 (7), 1023–1033. <https://doi.org/10.1289/ehp.1510267>.

- (13) Kleinstreuer, N. C.; Ceger, P.; Watt, E. D.; Martin, M.; Houck, K.; Browne, P.; Thomas, R. S.; Casey, W. M.; Dix, D. J.; Allen, D.; Sakamuru, S.; Xia, M.; Huang, R.; Judson, R. Development and Validation of a Computational Model for Androgen Receptor Activity. *Chem Res Toxicol* 2017, 30 (4), 946–964. <https://doi.org/10.1021/acs.chemrestox.6b00347>.
- (14) Mansouri, K.; Karmaus, A. L.; Fitzpatrick, J.; Patlewicz, G.; Pradeep, P.; Alberga, D.; Alepee, N.; Allen, T. E. H.; Allen, D.; Alves, V. M.; Andrade, C. H.; Auernhammer, T. R.; Ballabio, D.; Bell, S.; Benfenati, E.; Bhattacharya, S.; Bastos, J. V.; Boyd, S.; Brown, J. B.; Capuzzi, S. J.; Chushak, Y.; Ciallella, H.; Clark, A. M.; Consonni, V.; Daga, P. R.; Ekins, S.; Farag, S.; Fedorov, M.; Fourches, D.; Gadaleta, D.; Gao, F.; Gearhart, J. M.; Goh, G.; Goodman, J. M.; Grisoni, F.; Grulke, C. M.; Hartung, T.; Hirn, M.; Karpov, P.; Korotcov, A.; Lavado, G. J.; Lawless, M.; Li, X.; Luechtefeld, T.; Lunghini, F.; Mangiatordi, G. F.; Marcou, G.; Marsh, D.; Martin, T.; Mauri, A.; Muratov, E. N.; Myatt, G. J.; Nguyen, D.-T.; Nicolotti, O.; Note, R.; Pande, P.; Parks, A. K.; Peryea, T.; Polash, A. H.; Rallo, R.; Roncaglioni, A.; Rowlands, C.; Ruiz, P.; Russo, D. P.; Sayed, A.; Sayre, R.; Sheils, T.; Siegel, C.; Silva, A. C.; Simeonov, A.; Sosnin, S.; Southall, N.; Strickland, J.; Tang, Y.; Teppen, B.; Tetko, I. V.; Thomas, D.; Tkachenko, V.; Todeschini, R.; Toma, C.; Tripodi, I.; Trisciuzzi, D.; Tropsha, A.; Varnek, A.; Vukovic, K.; Wang, Z.; Wang, L.; Waters, K. M.; Wedlake, A. J.; Wijeyesakere, S. J.; Wilson, D.; Xiao, Z.; Yang, H.; Zahoranszky-Kohalmi, G.; Zakharov, A. V.; Zhang, F. F.; Zhang, Z.; Zhao, T.; Zhu, H.; Zorn, K. M.; Casey, W.; Kleinstreuer, N. C. CATMoS: Collaborative Acute Toxicity Modeling Suite. *Environ Health Perspect* 2021, 129 (4), 47013. <https://doi.org/10.1289/EHP8495>.
- (15) Benigni, R.; Battistelli, C. L.; Bossa, C.; Tcheremenskaia, O.; Crettaz, P. New Perspectives in Toxicological Information Management, and the Role of ISSTOX Databases in Assessing Chemical Mutagenicity and Carcinogenicity. *Mutagenesis* 2013, 28 (4), 401–409. <https://doi.org/10.1093/mutage/get016>.
- (16) *ToxValDB* in *CompTox*. <https://clowder.edap-cluster.com/datasets/61147fefe4b0856fdc65639b#folderId=66853bf7e4b0a7c65d177bc3&page=0> (accessed 2025-01-02).
- (17) *Mammalian Cell Mutagenicity | DTT Data Collections Guided Search*. <https://doi.org/10.22427/NTP-DATA-022-00002-0009-000-5>.
- (18) Baderna, D.; Gadaleta, D.; Lostaglio, E.; Selvestrel, G.; Raitano, G.; Golbamaki, A.; Lombardo, A.; Benfenati, E. New in Silico Models to Predict in Vitro Micronucleus Induction as Marker of Genotoxicity. *J Hazard Mater* 2020, 385, 121638. <https://doi.org/10.1016/j.jhazmat.2019.121638>.
- (19) Honda; Fujita, Y.; Kasamatsu, T.; Fuchs, A.; Fautz, R.; Morita, O. Necessity for Retrospective Evaluation of Past-Positive Chemicals in in Vitro Chromosomal Aberration Tests Using Recommended Cytotoxicity Indices. *Genes and Environment* 2018, 40 (1), 2. <https://doi.org/10.1186/S41021-017-0091-Y>.
- (20) Fujita, Y.; Ito, Y.; Morita, O.; Honda, H. Validation of Retrospective Evaluation Method for False Genotoxic Chemicals with Strong Cytotoxicity: Re-Evaluation Using in Vitro Micronucleus Test. *Fundam Toxicol Sci* 2016, 3 (6), 251–256. <https://doi.org/10.2131/FTS.3.251>.

