Water Research 294 (2026) 125499

Contents lists available at ScienceDirect

Water Research

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

Screening toxic transformation products of emerging pollutants in advanced oxidation processes with 3D deep learning and in vitro assays

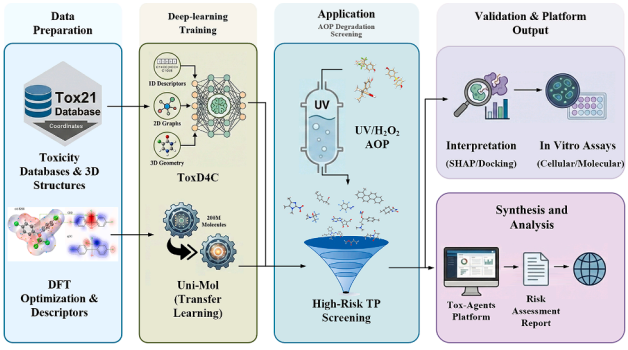
Fulin Shao [a], Weiying Li [a,b,*], Zhiwei Liang [a], Yu Zhou [a], Dawei Zhang [a], Yu Chang [a]

[a] College of Environmental Science and Engineering, Tongji University, Shanghai, 200092, PR China
[b] College of Environmental Science and Engineering, Shanghai Institute of Pollution Control and Ecological Security, State Key Laboratory of Pollution Control and Resource Reuse, Tongji University, Shanghai, 200092 PR China

HIGHLIGHTS

- 3D-aware multi-modal deep learning predicts 31 molecular toxicity endpoints.
- SE(3)-equivariant models outperform descriptor-based QSAR across benchmarks.
- Tox-Agents platform automates toxicity reporting via RAG and LLMs.
- UV/H2O2 case study reveals high-risk transformation products in water.
- In vitro JC-1, CCK-8 and NR/SR assays validate model-predicted toxicities.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Advanced oxidation processes
Emerging pollutants
Transformation products
Structure–toxicity relationships
Deep learning

ABSTRACT

The rapid and precise toxicity assessment of chemical pollutants and their byproducts formed during water treatment and in aquatic environments remains a significant environmental challenge, as the predictive power of conventional quantitative structure–activity relationship (QSAR) models is limited by their reliance on simplified molecular descriptors. To address this, ToxD4C, a novel multi-modal deep learning framework, was developed to simultaneously classify and regress 31 toxicity endpoints, covering nuclear receptor and enzyme panels, stress response assays, mutagenicity, carcinogenicity, cardiopulmonary toxicity, and various environmental toxicities of concern for water quality management. ToxD4C uniquely integrates three-dimensional molecular geometries, graph attention networks, and SE(3)-equivariant Transformer architectures, effectively capturing complex stereochemical and electronic molecular features. In parallel, a pretrained Uni-Mol model was fine-tuned via transfer learning on Density Functional Theory (DFT)-optimized structures, independently generating normalized toxicity predictions with enhanced reliability and generalization. Both approaches outperformed traditional descriptor-based models across validation tests. Feature-attribution analysis (SHAP) highlighted key physicochemical drivers of predicted toxicity, and receptor docking offered mechanistic context for selected receptor-mediated endpoints. Applied to realistic UV/H_2O_2 advanced oxidation scenarios in a real water matrix, this approach efficiently identified high-risk transformation products, and their predicted toxicity was further validated in vitro using JC-1 mitochondrial membrane potential, CCK-8 cell viability, and nuclear receptor/stress-

* Corresponding author at: College of Environmental Science and Engineering, Tongji University, Shanghai, 200092, PR China.
E-mail address: 123lwyktz@tongji.edu.cn (W. Li).

https://doi.org/10.1016/j.watres.2026.125499
Received 17 November 2025; Received in revised form 14 January 2026; Accepted 2 February 2026
Available online 3 February 2026
0043-1354/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

response reporter assays. These tools are integrated within the open-source Tox-Agents platform, enabling rapid and interpretable decision-making for water treatment and environmental risk assessment.

1. Introduction

Advanced Oxidation Processes (AOPs) are widely employed to eliminate emerging organic contaminants from water systems; however, incomplete mineralization often generates a spectrum of transformation products (TPs) whose environmental risks remain largely uncharacterized (Castellano-Hinojosa et al., 2023; Ferrando-Climent et al., 2017). Recent large-scale chemical release incidents highlight the urgent need for rapid, mechanism-aware screening not only of parent chemicals but also of TPs formed during emergency treatment and remediation (Zhang et al., 2025). Unlike stable parent compounds, these TPs are formed rapidly, often possess altered stereochemistry or functional group distributions that traditional monitoring overlooks, and lack commercially available standards for experimental testing (Liu et al., 2021; Roy and Mitra, 2012). A widening gap exists between chemical proliferation and regulation: of the 200+ million registered chemicals, less than 0.04 % are covered by regulatory inventories like the EPA's, and conventional analytical methods apply to fewer than 5 % of known compounds (LaKind et al., 2016; Wang et al., 2020). Although high-resolution mass spectrometry aids identification, assessing TP toxicity remains a bottleneck due to rapid kinetics and structural complexity (Bletsou et al., 2015). AOP-induced transformations—such as stereoselective hydroxylation or ring cleavage—create byproducts with 3D geometries and electronic distributions distinct from their parents. These features often elude capture by traditional 2D molecular descriptors, highlighting the need for 3D-aware deep learning models that can infer toxicity directly from the theoretical geometries of these complex intermediates (McCord et al., 2022). Consequently, given the prohibitive cost and slow pace of experimental validation, rapid computational assessment is essential to prevent the inadvertent formation of byproducts more toxic than their precursors (Mansouri et al., 2021; Raies and Bajic, 2016; Tran et al., 2023; Verma et al., 2010).

Most widely used toxicity prediction platforms today rely on traditional quantitative structure–activity relationship (QSAR) approaches, with the EPA's T.E.S.T. software serving as a representative example. These models typically employ two-dimensional (2D) molecular descriptors—such as SMILES strings, fingerprints, and topological indices—with algorithms like hierarchical clustering, linear regression, group contribution, nearest-neighbor analysis, and consensus methods. While effective in certain applications, 2D descriptors primarily encode atom connectivity and local neighborhoods; they often omit global planarity/dihedral constraints, absolute stereochemistry, and 3D electrostatic fields that govern receptor engagement and reactivity (Ertl et al., 2022). The limitations of 2D descriptors are exemplified by concrete toxicity cases: for dioxins and PCBs, biological potency depends on substitution pattern and planarity—only 2,3,7,8-substituted PCDDs/PCDFs and coplanar (non-/mono-ortho) PCBs engage the aryl hydrocarbon receptor (AhR) with high potency, whereas ortho substitution twists the biaryl, disrupts coplanarity, and markedly lowers AhR engagement—features encoded by 3D dihedral and electrostatic fields rather than by position-agnostic 2D counts (Couture et al., 1990; Gu et al., 2019; Van Den Berg et al., 1998). Likewise, stereochemistry governs hazard in widely used pesticides: cis and 1R configurations of pyrethroids are typically more potent than their trans/1S counterparts, and mecoprop-P (the R-enantiomer) is the herbicidally active form (Cohen et al., 1993). This highlights a critical limitation of 2D-based QSAR for stereochemically or conformationally sensitive toxicants and motivates geometry- and physics-aware molecular representations.

The fundamental limitation of 2D descriptors has prompted a paradigm shift towards 3D molecular representations (Subbiah and Chinnappan, 2020; Zubrod et al., 2024). While recent machine learning advances have enhanced predictive accuracy, a critical gap remains: many commonly used toxicity models compress molecular structure into simplified feature spaces—typically 2D descriptors or single-conformer summaries—that erase absolute/axial chirality, collapse conformational flexibility and dihedral/planarity constraints, and omit spatially resolved electrostatics (e.g., partial charges, ESP/multipoles, and hydrogen-bond directionality) that are critical for recognition and reactivity (Zhong et al., 2021a, 2022). Consider stereoisomers—molecules with identical connectivity but different 3D arrangements—which can exhibit dramatically different toxicities despite identical 2D representations. In contrast to alignment-dependent 3D-QSAR methods (e.g., CoMFA), SE(3)-equivariant networks operate directly on 3D coordinates and are equivariant to global rotations and translations, reducing the need for manual alignment and extensive rotational data augmentation. However, many commonly used E(3)/O(3)-equivariant or invariant formulations based solely on reflection-invariant scalars (e.g., distances/angles) are chirality-blind and cannot distinguish enantiomers unless chirality-sensitive (parity-odd) features are introduced or reflection symmetry is explicitly broken (Dumitrescu et al., 2025; Fuchs et al., 2020). This disconnect between molecular reality and computational abstraction has driven the field towards spatially-aware modeling approaches that preserve the full complexity of molecular structure (Zhang et al., 2024).

Driven by progress in deep learning, 3D molecular representation models have recently gained traction for their ability to learn spatial and electronic features with greater physical and chemical fidelity (Zhong et al., 2021b; Zhong and Guan, 2023). Among these, GraphMVP (Liu et al., 2022), GeminiMol (Wang et al., 2025), and Uni-Mol (Zhou et al., 2023) have shown excellent performance in tasks such as molecular property prediction, docking, and conformational sampling (Basak et al., 2007; Cecily Mary Glory et al., 2017). Large-scale 3D pretraining (e.g., Uni-Mol) provides symmetry-aware representations that transfer effectively to small-molecule property prediction (Wang et al., 2024; Zhao et al., 2020). However, most existing 3D representation models have been developed and benchmarked primarily in drug-discovery settings, where training objectives and evaluation tasks emphasize drug-like chemical space and relatively stable, energy-minimized conformers. In the context of AOP-derived transformation products, this assumption can be limiting because many intermediates are conformationally flexible, short-lived, and may undergo functional-group edits that alter both local reactivity and global shape. Motivated by this gap, ToxD4C was developed for environmental risk assessment of transformation products. The framework integrates graph-based information with 3D conformational features by using multi-conformer sampling and a fusion architecture, allowing toxicity screening to account for both substructure-level cues and overall molecular geometry across structurally diverse transformation products.

This study introduces an integrated deep learning framework to address key challenges in environmental toxicity modeling. The primary contribution is a hybrid Graph Neural Network (GNN)-Transformer model that uniquely merges 2D graph attention networks with 3D SE(3)-equivariant Transformers to capture comprehensive molecular features (Veličković et al., 2018). To enhance predictive power, transfer learning was leveraged by fine tuning the large scale pretrained Uni-Mol model on a dataset of Density Functional Theory (DFT) optimized structures. DFT refinement supplies physically consistent geometries and electronic descriptors (e.g., electrostatic potential minima, HOMO-LUMO gap, quadrupole moments) that are linked to AhR activation and hERG blockade, thereby enriching the model with physics-informed features that improve the biological faithfulness of toxicity predictions. Ablation studies were conducted to validate the model's architecture, confirming

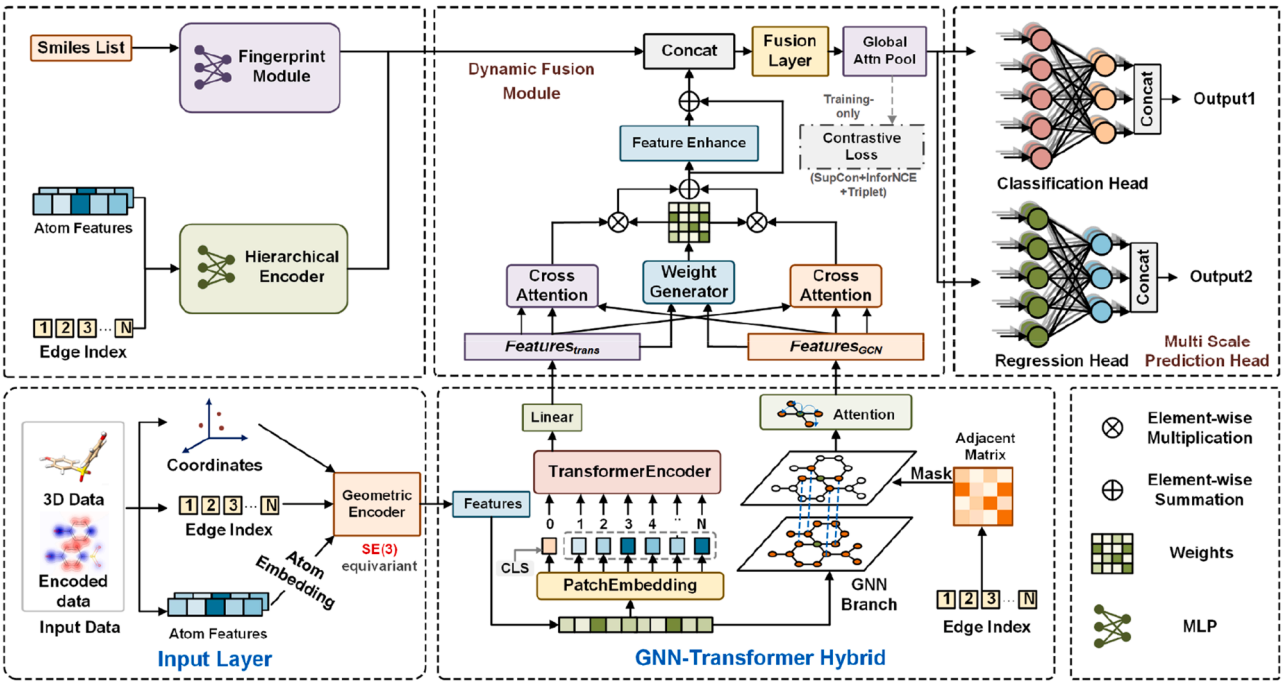


Fig. 1. Architecture of the ToxD4C framework integrating fingerprint, graph, and 3D geometric representations via a GNN–Transformer hybrid for multi-task molecular toxicity prediction.

the significant contribution of each key component to the final predictive performance. An optional report-generation layer accompanies the workflow, formatting model outputs into a structured summary and integrating user-supplied local documents via retrieval to support transparent reporting of screening outcomes (Dong et al., 2024).

2. Methods

2.1. The ToxD4C framework

In this work, a multi-modal, multi-task deep-learning framework, ToxD4C, was developed for comprehensive molecular-toxicity prediction. The benchmark dataset comprised 31 endpoints—26 classification and five regression tasks (He et al., 2024). Crucially, learning targets were defined at the individual endpoint level using experimental ground-truth labels. Detailed dataset statistics (sample sizes, class-imbalance ratios, and missing-label patterns) are reported in Tables S4 (classification) and S5 (regression). During multi-task training, a mask-aware loss function was employed to ensure that optimization and evaluation were driven solely by observed labels, preventing bias from missing entries. These endpoints were chosen to reflect commonly used evidence streams in environmental chemical assessment, spanning apical hazards/organ toxicities and mechanistic screening panels (NR/SR and CYP isoforms), together with key regression endpoints relevant to environmental fate and acute toxicity (e.g., BCF, aquatic LC50/IGC50, and LD50) (Bailer et al., 2000; Burkhard, 2021; Shi et al., 2023). Each molecule was represented simultaneously in fingerprint, graph, and geometric space. Three complementary fingerprints—2048-bit Extended Connectivity Fingerprint (ECFP, radius=2), 167-bit MACCS keys, and 2048-bit RDKit—along with 15 physicochemical descriptors were concatenated to encode both local substructures and global properties (Chen and Reynolds, 2002; Nicholls and Grant, 2005). Molecular graphs derived from SMILES strings included 119-dimensional atomic features. Molecular geometries were represented as a multi-conformer ensemble generated with RDKit ETKDG and MMFF94 energy minimization (Halgren, 1996). During training, 9 conformers were randomly sampled in each epoch, and their Cartesian coordinates were fed to the SE(3)-equivariant encoder for geometric reasoning (Fuchs et al., 2020). A six-layer hybrid core combined a three-layer graph attention branch for local chemical neighborhoods with a parallel three-layer Transformer for long range geometric context. The Dynamic Fusion Module integrates the two feature streams using bidirectional cross attention and a learnable gate, and the fused embedding is passed to the multi task prediction heads (Fig. 1) (He et al., 2016; Veličković et al., 2018). For deployment-oriented confidence reporting, a similarity-based reliability score was computed as the maximum Tanimoto similarity between a query compound and the training set using ECFP fingerprints. Implementation specifics (optimizers, schedules, regularization settings, and ablations) are detailed in SI §S2.

2.2. Transfer learning toxicity prediction with pre-trained models

The study was based on 7330 curated molecules from the Tox21 database (Helma, 2005; Richard et al., 2021). To evaluate the trade-off between screening sensitivity and mechanistic rigor, two distinct labeling regimes were established for the Uni-Mol transfer learning experiments. First, a baseline positive-proportion label (Pc) was defined as the unweighted fraction of active assays derived from the Tox21SL list (spanning 78–978 assays per compound), serving as a broad heuristic for general bioactivity. Second, to mitigate assay artifacts and systematic biases arising from uneven assay inventories, a refined mechanism-balanced score (Sc) was constructed using a ToxCast Pipeline (tcpl)-compliant pipeline. As detailed in Section S12, this pipeline filters non-specific cytotoxicity bursts ($\Delta \geq 3\log10$) and technical artifacts. Crucially, hit probabilities are aggregated hierarchically—first within mechanistic families and then across families—using statistical regularization techniques (Jeffreys prior shrinkage) to stabilize sparse signals, ensuring the score reflects biological diversity rather than assay count (Feshuk et al., 2023).

To complement these labels, three-dimensional structures were generated and energy-minimized at the ωB97X-D/6–31G(d,p) level with a PCM water model (Gaussian 16), from which 51 quantum-chemical and 13 PubChem descriptors were derived. As a geometry-sensitivity control, we repeated Uni-Mol fine-tuning using RDKit ETKDG/

MMFF94 conformers in place of DFT-optimized coordinates; the PCA shift between paired representations and the resulting screening AUC are summarized in Figure S32. While these physicochemical properties functioned as input features for conventional machine-learning models such as XGBoost, they were utilized within the Uni-Mol framework as auxiliary regression targets to regularize the SE(3)-equivariant encoder. This multi-task setup encourages the model to encode electronic and steric information that is relevant to molecular interactions. Ablations (Table S20) showed that adding lower-reliability auxiliary targets (e.g., ALIE_min) did not harm downstream toxicity prediction, suggesting the multi-task objective avoids negative transfer. Separately, the SE(3)-equivariant Uni-Mol model was fine-tuned on the DFT-optimized coordinates via transfer learning under stratified five-fold cross-validation with a scaffold-aware assignment. At evaluation only, test-time augmentation (five stochastic passes with small coordinate perturbations) summarized predictive dispersion, and temperature scaling was applied to produce calibrated probabilities. Performance was summarized with an imbalance-aware panel (ROC-AUC, PR-AUC, F1; additional metrics and calibration are reported in the Supporting Information). Post-hoc interpretation used TreeSHAP on the descriptor-based head; SHAP did not participate in training or model selection (Mangalathu et al., 2020). To provide mechanistic context, docking against estrogen receptor β (ERβ; PDB 1 × 7B), estrogen receptor α (ERα), androgen receptor (AR), and peroxisome proliferator-activated receptor gamma (PPARγ) was performed, and short molecular-dynamics (MD) simulations were used to assess pose stability (Allouche, 2012). Detailed protocols—label construction, split definitions, calibration/Test Time Augmentation (TTA), and docking/MD—are provided in the Supporting Information.

2.3. Predicting degradation pathways via enhanced-sampling molecular dynamics

Metadynamics-enhanced molecular dynamics was carried out in a spherical nanoreactor whose harmonic wall oscillated from $R_{max} + 1.0$ Å to $R_{max} + 2.5$ Å, where R_{max} is the current maximum distance of any atom from the molecular geometric center; this adaptive confinement promotes frequent barrier crossing without over-compressing the system (Tribello et al., 2014). A history-dependent Gaussian bias acted on an adaptive MULTI_RMSD (root-mean-square deviation) collective variable, refreshed from on-the-fly averaged configurations, enabling rapid escape from explored minima (Trizio et al., 2024). Forces and energies were evaluated with either a DeePMD-kit (LAMMPS), the semi-empirical GFN2-xTB Hamiltonian, or any DFT level available in CP2K (Bannwarth et al., 2019; Zeng et al., 2023). This design ensured that the various computational engines were readily interchangeable, allowing the desired method to be selected through a single input keyword. Reactive events were automatically identified and clustered, with representative structures subsequently re-optimized at the selected quantum chemistry level, ensuring a self-consistent degradation network.

2.4. Retrieval-Augmented toxicity reporting via the tox-agents platform

An open-source platform, Tox-Agents (https://github.com/PhelanShao/tox-agents), was developed to automate and standardize toxicity screening workflows. The platform integrates the predictive models developed in this study with a lightweight information-retrieval and templated reporting layer to generate structured screening summaries for downstream interpretation. Within Tox-Agents, Uni-Mol is used for rapid first-pass toxicity triage and bulk descriptor/property profiling, whereas ToxD4C provides the primary endpoint-resolved predictions across the 31-endpoint benchmark; the report layer summarizes both outputs side-by-side for prioritization. This reporting component is optional and does not affect model training, benchmarking, or the numerical prediction results; implementation details and example outputs are provided in the Supporting Information (Dong et al., 2024). To demonstrate environmental applicability, the workflow was applied to transformation products generated from tetrabromobisphenol A-diallyl ether (TBBPA-DAE), bisphenol A (BPA), and dicofol during UV/H_2O_2 advanced oxidation in a real water matrix. Transformation products

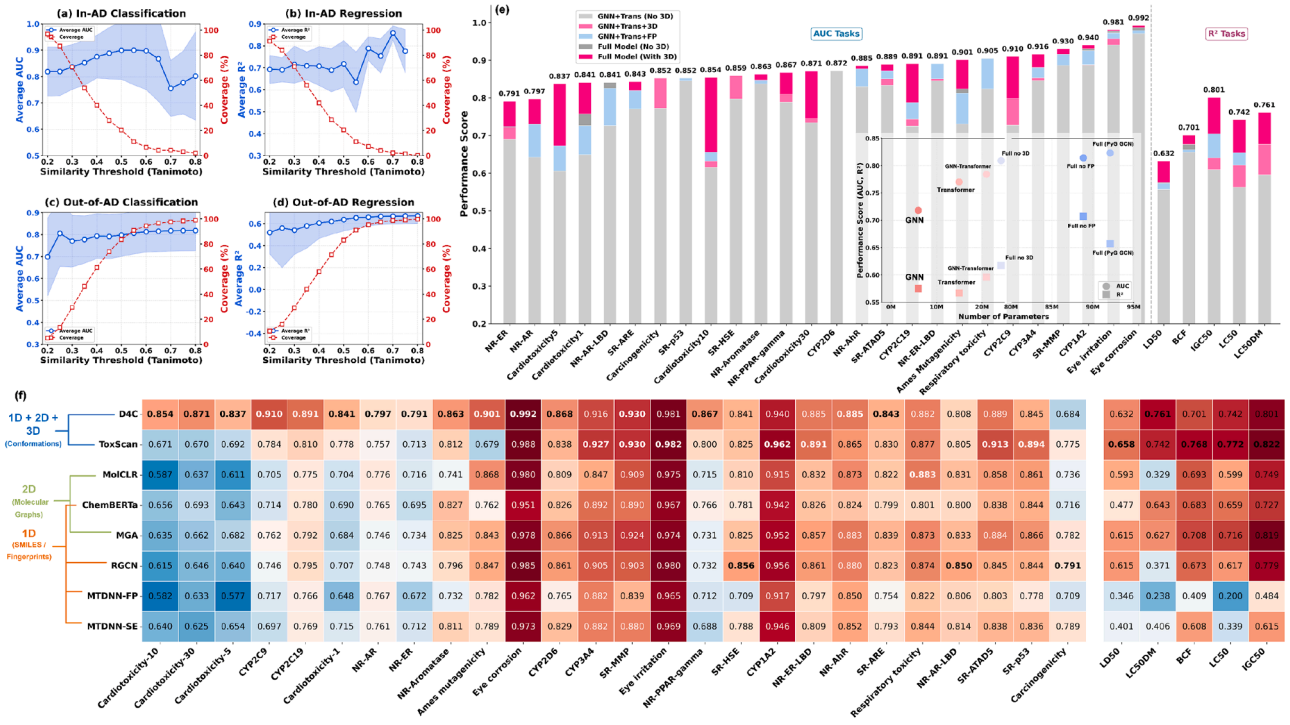


Fig. 2. Applicability-domain (AD) behavior and benchmark performance of ToxD4C across 31 toxicity endpoints. (a-d) Average AUC/R^2 and coverage as a function of Tanimoto similarity threshold for in-AD and out-of-AD settings in classification and regression. (e) Endpoint-wise performance of architectural variants and the model size–performance trade-off. (f) Heatmap comparison of ToxD4C with representative baseline models on classification (AUC) and regression (R^2) tasks.

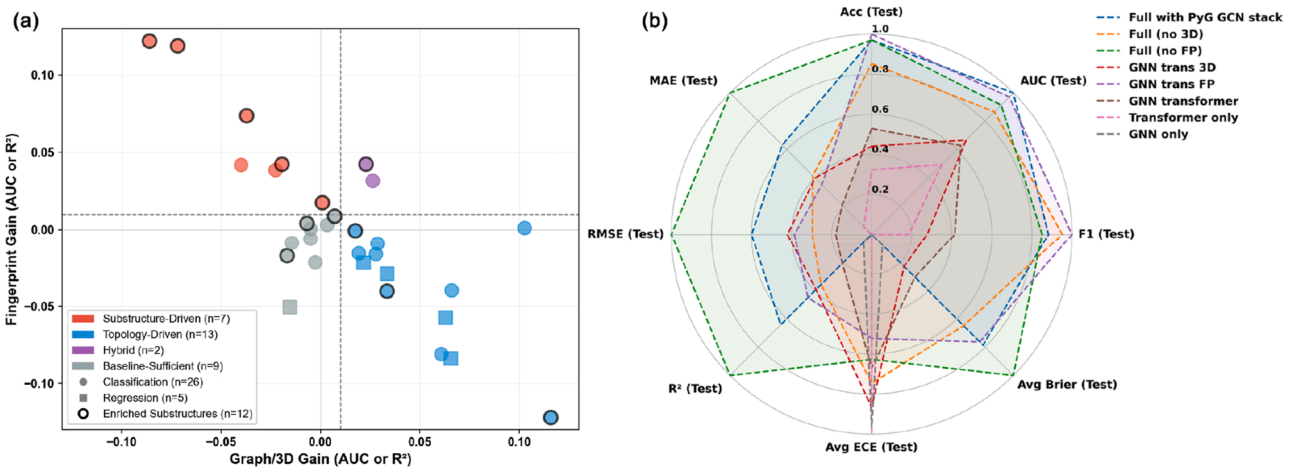


Fig. 3. Gain-based mechanistic classification of toxicity endpoints (a). The scatter plot positions 31 toxicity endpoints based on the performance gain from Graph/3D features (x-axis) versus Fingerprint features (y-axis). Circles represent classification tasks; squares represent regression tasks. (b) Ablation (Test) radar plot of normalized overall metrics across configurations.

were identified by LC–MS, and model-predicted endpoints were used to support hazard screening and prioritization of products for further evaluation.

3. Results and discussion

3.1. Performance evaluation and ablation analysis

The ToxD4C framework demonstrated competitive performance across a comprehensive molecular toxicity benchmark, with different architectural variants achieving the best results in distinct task subsets. As presented in the performance heatmap (Fig. 2f), the full model achieved an average AUC of 0.845 across 26 classification tasks, outperforming established methods including ToxScan (0.827), MolCLR (0.798), and ChemBERTa (0.796). Notably, the model exhibited superior accuracy in predicting key environmental fate and aquatic toxicity endpoints—critical for ecological risk assessment—yielding R^2 values of 0.701 for Bioconcentration Factor (BCF) and 0.801 for Tetrahymena pyriformis toxicity (IGC50). Further examination of the applicability domain in Fig. 2(a-d) delineates a divergence in generalization capabilities. While classification models maintained robustness against structural shifts—retaining out-of-domain AUC scores exceeding 0.7—regression precision exhibited heightened sensitivity to the chemical similarity between query and training samples. This suggests that reliable quantitative predictions for continuous environmental parameters necessitate stricter similarity thresholds compared to qualitative classification. Despite these aggregate successes, the optimal architectural configuration was not uniform across all endpoints. The variant lacking the fingerprint module yielded superior results in 9 tasks (29.0 %), whereas the GNN-Transformer enhanced with 3D features outperformed others in 8 tasks (25.8 %). This observed variability counters the assumption that maximizing architectural complexity universally enhances performance and implies that toxicity prediction requires architectures tailored to specific endpoint characteristics (Seal et al., 2025).

To interpret this task dependence mechanistically, a gain-based endpoint map was constructed to contrast the marginal contributions of Graph/3D features against the fingerprint branch (Fig. 3). Using a threshold of 0.01 (AUC or R^2), endpoints were partitioned into four mechanistic regimes: Topology-Driven (Graph/3D gain $>$ 0.01 and Fingerprint gain $<$ 0.01), Substructure-Driven (Fingerprint gain $>$ 0.01 and Graph/3D gain $<$ 0.01), Hybrid (both gains $>$ 0.01), and Baseline-Sufficient (otherwise). This taxonomy reveals a strong negative correlation (r = -0.7937) between the two modality gains, suggesting a competitive trade-off rather than simple additivity. The efficacy of the fingerprint-free variant on 9 endpoints is directly explained by this taxonomy. Notably, 80 % (4/5) of the regression endpoints (e.g., BCF, IGC50) fall into the Topology-Driven category. In these continuous-value contexts, toxicity is predominantly governed by smooth global physicochemical constraints (e.g., hydrophobicity and molecular size); consequently, high-dimensional hashed fingerprints may introduce quantization noise rather than informative signals, failing to capture the continuous manifold of molecular properties (Mayr et al., 2016). Across the entire benchmark, the most pronounced Graph/3D gains were observed for specific receptor-mediated classification tasks, such as CYP3A4 (gain of 0.1025) and NR-PPAR-γ (0.1157), consistent with activities governed by global shape complementarity and steric–electrostatic balance.

Conversely, the necessity of the fingerprint module for Substructure-Driven tasks is validated by independent enrichment analysis. Representative endpoints such as Carcinogenicity and Cardiotoxicity exhibit statistically significant enrichment of specific MACCS keys in the toxic class (FDR $<$ 0.05), often with large effect sizes (e.g., Carcinogenicity MACCS-52, Odds Ratio = 4.46). This confirms that explicit substructure encodings provide the primary discriminative cues for motif-dependent toxicities. Integrating these insights, the ablation study supports a "parsimony with alignment" design strategy: 3D geometry (SE(3)-equivariant encoding) and supervised contrastive regularization serve as effective defaults, yielding improvements ranging from 2.7 to 21.3 percentage points in classification AUC and 10.4 to 47.0 percentage points in regression R^2 (Sharma et al., 2023). Meanwhile, the fingerprint branch is best managed via cross-attention fusion—which outperformed simple concatenation by 0.047 AUC—allowing the model to dynamically suppress noisy modalities in topology-dominated tasks while retaining informative motifs where mechanistically required.

3.2. Transfer learning with Uni-Mol for enhanced 3D-aware toxicity scoring

The dominant paradigm in artificial intelligence—pretraining on vast, heterogeneous corpora to build task-agnostic priors, followed by task-specific fine-tuning—has been adapted for molecular toxicity. In an approach analogous to that of large language and vision models, the Uni-Mol model is first pretrained on a vast chemical space of over 200 million molecular conformations. This process equips the encoder with a transferable understanding of common substructures, stereochemistry, and energetically favorable 3D arrangements (Huang et al., 2016; Idakwo et al., 2020). This rich chemical prior enables the learning of

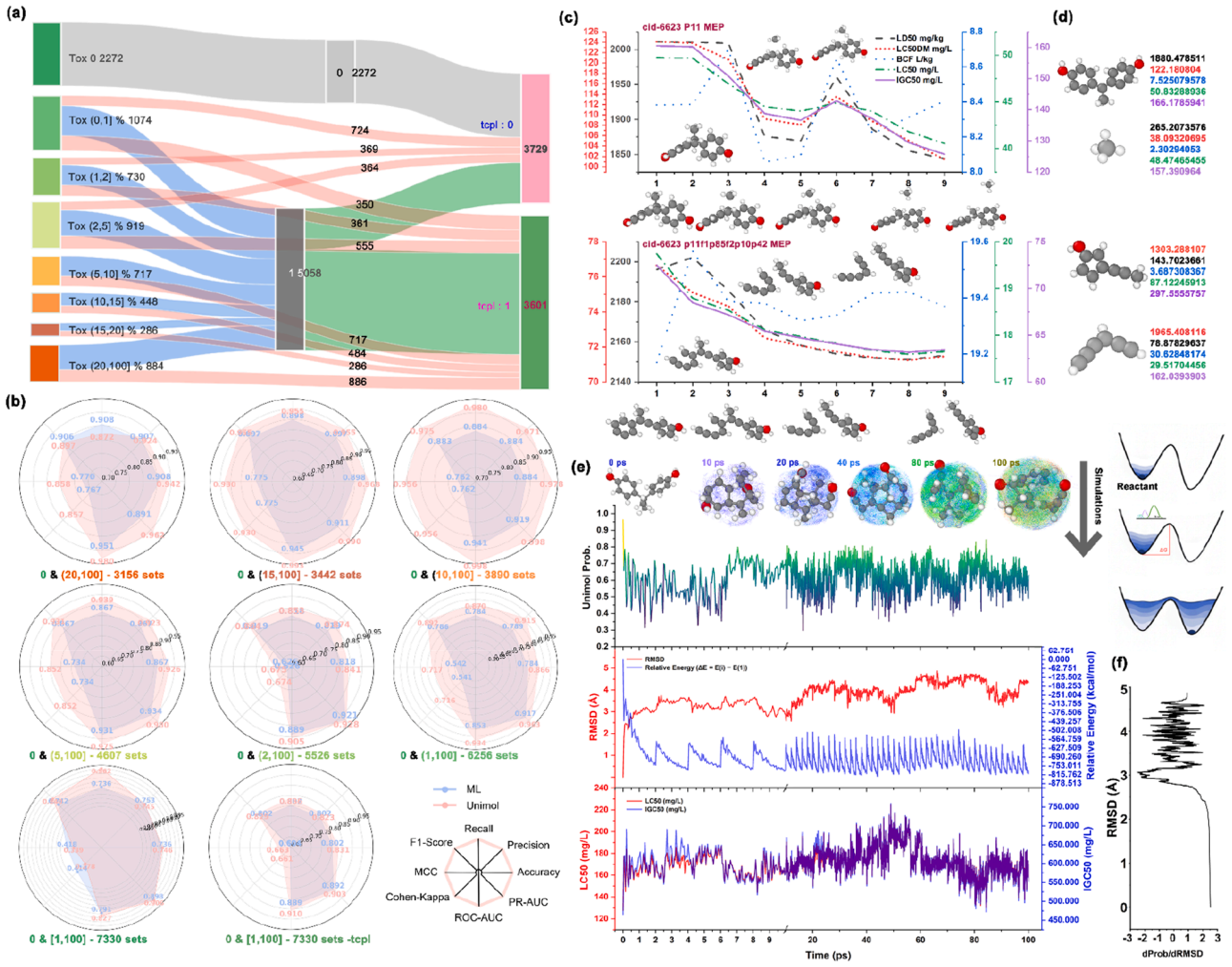


Fig. 4. Distribution of assay-positivity rates in the Tox21–10k set and the mapping from a direct-toxicity label to tcpl (ToxCast pipeline) active/inactive calls (a); Performance of conventional ML vs. Uni-Mol transfer learning under different dataset partitioning schemes; radar plots summarize F1-score, precision, recall, accuracy, MCC, Cohen's κ, ROC-AUC, and PR-AUC (b); ToxD4C regression predictions for multiple endpoints (e.g., LD_{50}, LC_{50}, BCF, IGC_{50}) along the minimum-energy paths (MEPs) of two representative bisphenol A (BPA) degradation reactions (c); Predicted endpoint values for the corresponding reaction products (d); Metadynamics "nanoreactor" simulation of BPA over 100 ps: structural snapshots (top) with time series of predicted Uni-Mol probability, RMSD and relative electronic energy, and concurrent LC_{50}/IGC_{50} predictions (right axes), illustrating stabilization as products form (e); Sensitivity analysis of the transfer-learned toxicity model: RMSD versus the derivative of the predicted probability with respect to RMSD (dProb/dRMSD) (f).

more fine-grained and physically grounded structure-toxicity relationships than is possible with traditional, handcrafted 2D descriptors alone. Fine-tuning was performed on DFT-optimized coordinates under stratified cross-validation with a scaffold-aware assignment to reflect distribution shift. Two complementary labeling regimes were used. A permissive screening label marked a molecule as toxic whenever at least one of its available Tox21 assays was positive among the seventy-eight to nine-hundred seventy-eight assays per compound. In parallel, tcpl procedures produced dose–response-vetted per-endpoint hits with artifact and cytotoxicity control that were aggregated within and across mechanistic families into a mechanism-balanced chemical-level score S_c (Feshuk et al., 2023; Yang et al., 2019). With either regime, the SE (3)-equivariant Uni-Mol branch surpassed conventional machine-learning baselines across the imbalance-aware metric panel. Aggregated multi-endpoint training provided additional inductive bias beyond within-type pooling while single-endpoint heads preserved endpoint specificity (Fig. 4a and b).

The advantage of a 3D-aware representation is further illustrated by stress tests along chemistry rather than solely across static datasets. Along minimum-energy reaction paths for representative bisphenol-A degradations, regression heads track systematic shifts in LD50, LC50, BCF, and IGC50 as the reactant geometry evolves toward products, and the predicted values for the isolated products align with the trends observed along the paths (Fig. 4c and d). These behaviors are expected when toxicity is coupled to changes in charge distribution, local accessibility, and conjugation along the reaction coordinate. A metadynamics "nanoreactor" experiment provides a dynamic complement in which a biasing potential is gradually deposited on selected collective variables to accelerate barrier crossing; during the one-hundred-picosecond trajectory the Uni-Mol probabilities stabilize as product wells are reached, the RMSD plateaus, and dProb/dRMSD decreases (Fig. 4e-f). The joint evidence indicates that the encoder maps small, spatially localized rearrangements into coherent changes in toxicity predictions, whereas two-dimensional fingerprints would collapse stereoisomers and flatten physically meaningful variations along reaction pathways. Together, large-scale pretraining combined with SE(3)-equivariant fine-tuning yielded a toxicity scorer that is accurate and mechanistically interpretable, leveraging broad chemical priors, preserving three-dimensional symmetry constraints, and using task-aware fine-tuning and calibration to remain robust under distributional shift in environmental assessment.

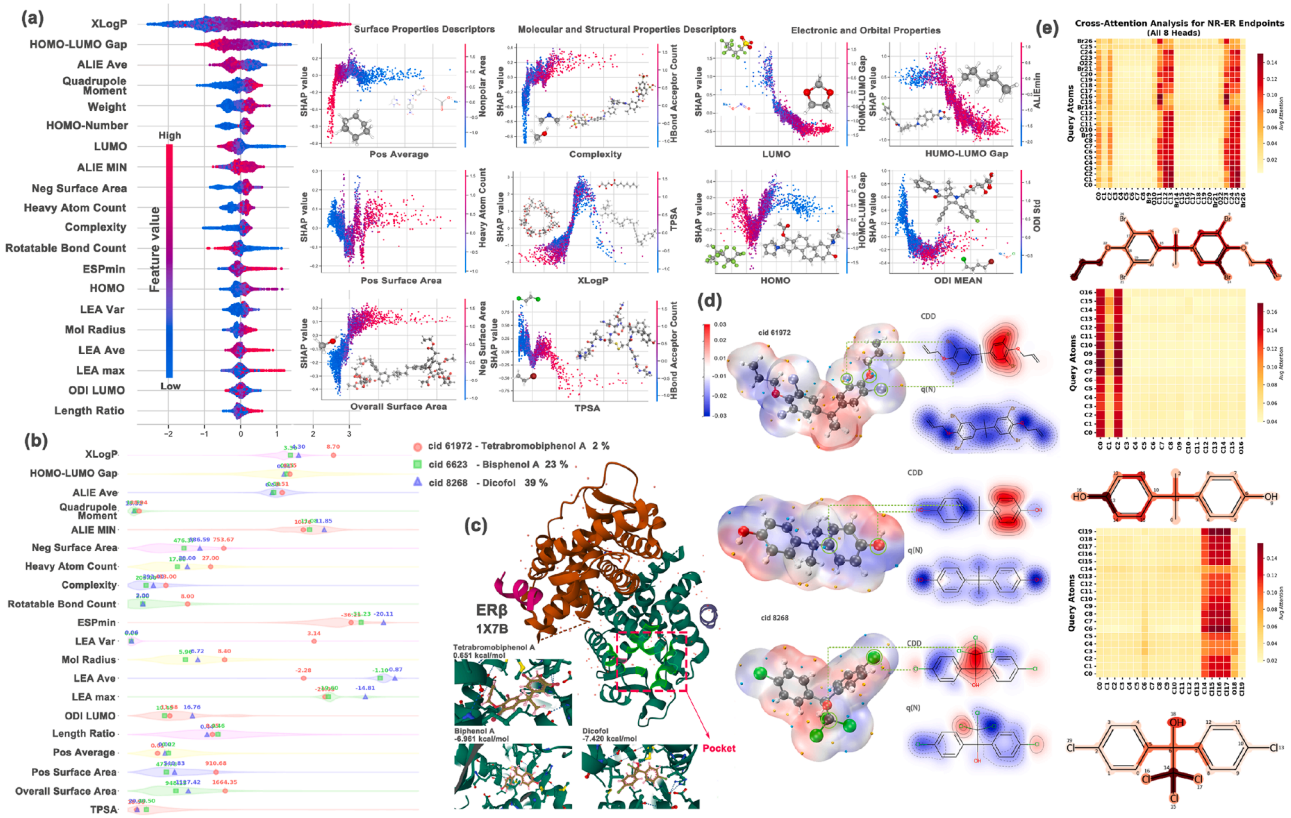


Fig. 5. Mechanistic triangulation of ERβ-mediated toxicity aligning Deep Learning Attention with Physical Simulation. (a) Global SHAP feature importance; (b) Dataset feature distributions; (c) ERβ binding pocket (PDB 1 × 7B); (d) Electronic reactivity maps (ESP/CDD); (e) Convergence of ToxD4C Cross-Attention Weights (Heatmap) with MD-derived Interaction Fingerprints, demonstrating that the model specifically 'attends' to the same phenolic motifs (red regions) that drive physical receptor binding.

3.3. Interpretable and mechanistic insights into toxicity drivers

To leverage the post-hoc interpretability of conventional machine learning, attribution analysis was grounded in the descriptor-based XGBoost model—identified as the best-performing ML configuration—and TreeSHAP was applied (Figure S3, Table 1). These analyses identified key molecular features driving predicted toxicity, which are interpreted within the adverse outcome pathway (toxicological AOP) framework (Fig. 5). Notably, XLogP (lipophilicity) serves as a key proxy for bioavailability and toxicokinetic (TK) amplification, acting as the dominant driver for aquatic regression endpoints (e.g., IGC50, BCF) governed by membrane partitioning (Lambert et al., 2022). Beyond bulk properties, electronic descriptors such as the HOMO–LUMO gap and Average Local Ionization Energy (ALIE) provide a plausible chemical basis for Key Events (KEs) involving electrophilic attack and oxidative stress, chemically consistent with Stress Response (SR) endpoints (Hillebrand et al., 2024; LoPachin and Gavin, 2014). Crucially for specificity, steric-electrostatic features like the quadrupole moment appear critical for receptor-mediated Molecular Initiating Events (MIEs), capturing the asymmetric charge distributions required for specific ligand-receptor binding (e.g., NR-ER) even when net dipoles are weak (Neel et al., 2017).

To bridge the gap between data-driven predictions and physical reality, molecular docking and short molecular dynamics (MD) simulations were employed as orthogonal interpretability checks, focusing specifically on the Estrogen Receptor β (ERβ, PDB 1 × 7B) as a direct mechanistic link to the NR-ER endpoints. This combined evidence—model attention, electronic reactivity indicators, and receptor interaction profiles—is referred to as a mechanism-oriented triangulation to support interpretation. Visualized in Fig. 5e, the ToxD4C Cross-Attention mechanism consistently assigns its highest weights to phenolic hydroxyls and halogenated rings across the TBBPA → BPA → Dicofol series. Conceptual DFT reactivity maps provide an electronic counterpart to these structural motifs highlighted by model attention. In Fig. 5d, electrostatic potential (ESP) and condensed dual descriptor (CDD) maps consistently localize electronically active regions on phenolic oxygens and activated ring positions, which are expected to

Table 1
Key Molecular Descriptor Thresholds for Toxicity Prediction Derived from SHAP Analysis.

Feature	Threshold Value	CI Lower	CI Upper	Reliability
XLogP	3.05893	2.97430	3.150712	0.979287
HOMO-LUMO Gap (a.u.)	0.33105	0.328424	0.333599	0.999068
ALIE Ave (a.u.)	0.50461	0.50018	0.508898	0.999361
Quadrupole Moment (a.u.)	21.1766	21.0696	21.27608	0.999801
Weight (Da)	246.334	244.8681	247.7464	0.991997
LUMO (a.u.)	-0.00517	-0.008458	-0.001918	0.986424
$ALIE_{min}$ (eV)	11.2949	10.92579	11.78723	0.489448
Negative Surface Area ($Bohr^2$)	359.924	358.417	361.266	0.998766
Heavy Atom Count	14.3053	14.13596	14.48571	0.989859
Complexity	184.588	184.1447	185.056	0.999999
Rotatable Bond Count	2.52924	2.514552	2.549708	0.999932
ESP_{min} (kcal/mol)	-36.8484	-37.28863	-36.33131	0.958385
HOMO (a.u.)	-0.29269	-0.295569	-0.289849	0.99852
LEA Var (eV)	0.06576	0.062453	0.068951	0.999583
Mol Radius (angstroms)	6.30992	6.285233	6.336594	0.996349
LEA Ave (a.u.)	-0.97949	-0.988604	-0.966204	0.814490
LEA Max (eV)	-17.6020	-17.87164	-17.28808	0.644671
ODI LUMO	21.2497	18.41094	25.89896	0.491487
Length Ratio	0.545714	0.527464	0.56729	0.524082

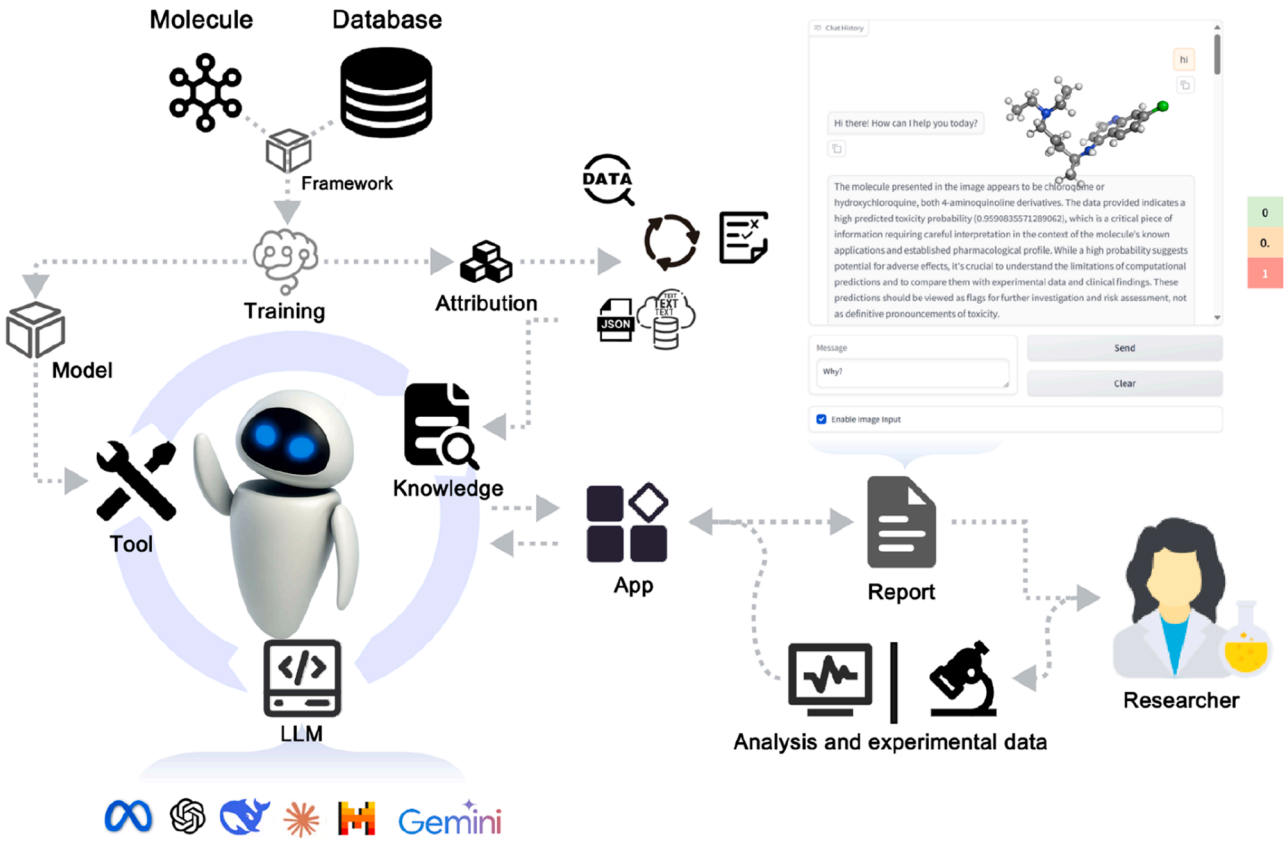


Fig. 6. Architecture and Deployment of the Toxicity Analysis Intelligent Agent.

govern hydrogen-bonding propensity and noncovalent recognition in estrogen-receptor binding. Strikingly, these same atoms coincide with persistent contact hotspots (e.g., Glu305/Arg346) in the ERβ pocket revealed by short MD interaction fingerprints. This co-localization links the learned attention pattern to chemically plausible pharmacophoric features, strengthening an endpoint-anchored (NR-ER/NR-ER-LBD) mechanistic interpretation rather than a purely statistical attribution (Shanle and Xu, 2011).

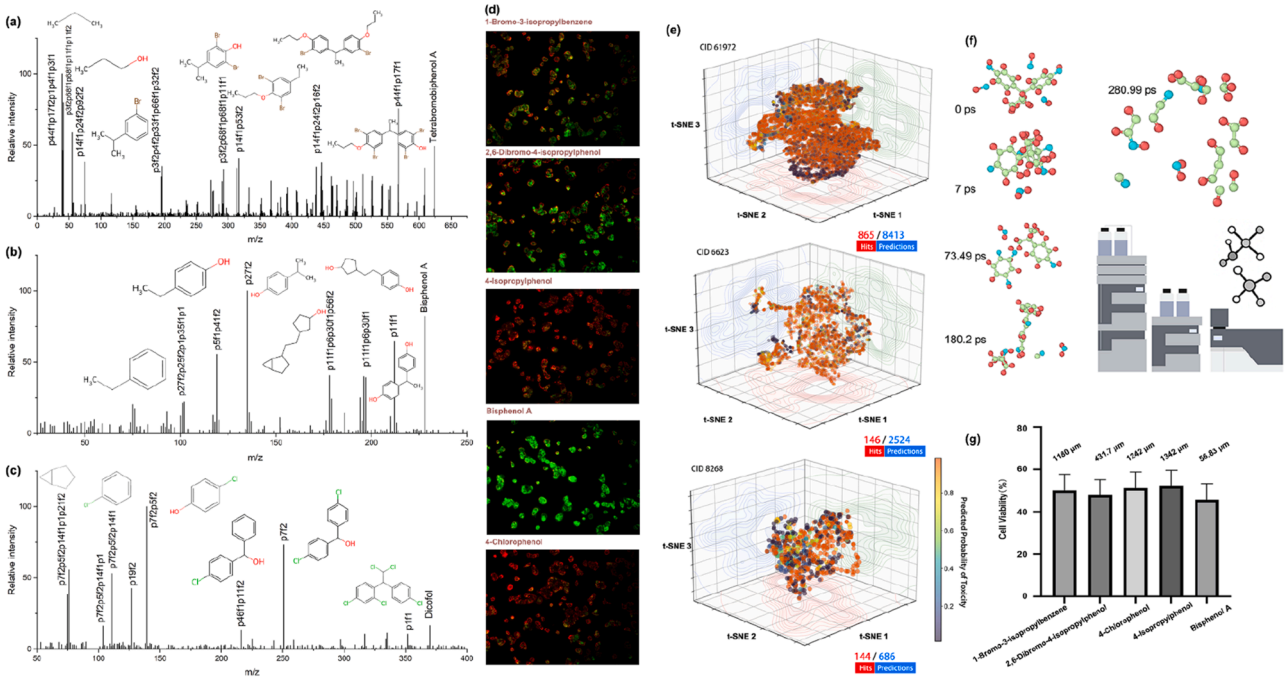


Fig. 7. Distribution and matching analysis of experimental MS data for degradation products (a)(b)(c); JC-1 staining of mitochondrial membrane potential in HepG2 cells exposed to selected degradation products (d); t-SNE visualization of degradation product molecular structures with mapped predicted toxicity (e); simulation process using a molecular dynamics-based nanoreactor for BPA (f); and CCK-8–based cell viability of HepG2 cells treated with selected degradation products (g).

Crucially, docking and short MD simulations serve as orthogonal support specifically for receptor-mediated MIE endpoints (e.g., NR-ER, NR-ER-LBD), whereas apical regression endpoints (e.g., LC50) represent complex outcomes used for screening prioritization rather than mechanistic proof. This rationale guided the selection of the five representative degradation products for experimental validation, ensuring coverage across these distinct mechanistic tiers—from MIEs and KEs to Apical outcomes—thereby rigorously testing the framework's generalizability. Comprehensive docking profiles and detailed interaction analyses for broader receptor panels are provided in Figure S33. Notably, the model successfully discriminated between high-affinity binders (e.g., BPA) and low-toxicity derivatives (e.g., diallyl-ether analogs), a distinction attributed to specific steric hindrance and electronic incompatibility within the receptor pocket rather than generic structural similarity. This confirms the framework's ability to identify subtle structural modifications critical for toxicological outcomes.

3.4. 'Tox-Agents' platform for automated risk assessment of environmental transformation products

Tox-Agents is an open-source reporting workflow designed to operationalize high-dimensional predictive toxicology for environmental transformation products. In practical screening settings, a study may yield dozens to hundreds of candidate molecules, and each molecule is associated with a dense output bundle: calibrated predictions for 31 toxicity endpoints, 51 physicochemical and electronic descriptors, SHAP-based attribution summaries, and applicability-domain (AD) reliability flags. Interpreting these outputs molecule-by-molecule is time-consuming and error-prone, and it becomes difficult to maintain consistent endpoint-level reasoning across a product ensemble. Tox-Agents addresses this bottleneck by using a structured, retrieval-assisted, templated analysis layer to organize the multi-endpoint profiles, highlight coherent mechanism-relevant patterns (e.g., concordant NR/SR signals versus isolated flags), and generate concise, evidence-traceable screening reports with visual summaries (Fig. 6). The resulting reports are intended as decision-support artifacts for prioritization—helping users compare products, identify recurring hazardous motifs, and select candidates for targeted confirmation—without altering any underlying numerical predictions.

To benchmark screening outputs against biological measurements, five representative degradation products (1-bromo-3-isopropylbenzene, 2,6-dibromo-4-isopropylphenol, 4-chlorophenol, 4-isopropylphenol, and bisphenol A) were selected based on detection along the transformation pathway and the availability of analytical standards (Table S18). The panel spans structural motifs relevant to endocrine signaling (NR-ER) and cytotoxicity-related endpoints. In HepG2 assays, JC-1 staining (Fig. 7d) showed pronounced mitochondrial depolarization for 4-chlorophenol, 4-isopropylphenol, and bisphenol A, whereas 1-bromo-3-isopropylbenzene retained predominantly red fluorescence, consistent with SR-MMP calls and corresponding scores (Table S17). CCK-8 assays (Fig. 7g) yielded IC_{50} values spanning more than one order of magnitude (12.97–234.9 mg L^{-1}), with bisphenol A being the most cytotoxic and 1-bromo-3-isopropylbenzene the least, matching their separation in the t-SNE embedding (Fig. 7e). The CCK-8 IC_{50} values reported here represent acute in vitro cytotoxicity thresholds used to benchmark model discrimination and to support relative potency ranking among prioritized products. The in vitro dosing design (1–50 μM, 24 h in HepG2; SI Section S25.2) was chosen to obtain full dose–response curves for IC_{50} estimation rather than to mimic environmentally realistic exposure scenarios. Bridging to environmental risk will require coupling these hazard estimates with measured or modeled exposure concentrations, together with toxicokinetic considerations, chronic exposure, and mixture effects. In addition, 4-isopropylphenol and bisphenol A were NR-ER positive and showed the highest NR-ER prediction probabilities, whereas the remaining products were ER-inactive with low scores. Overall, the qualitative agreement between predicted NR/SR profiles and orthogonal cellular readouts is consistent with the use of the reporting workflow to prioritize transformation products for follow-up testing.

By translating high-dimensional model outputs into a standardized, evidence-traceable summary, the reporting workflow enables rapid triage and prioritization of large transformation-product libraries and supports targeted confirmation of candidates with coherent multi-endpoint signals and higher predicted reliability. The framework is applicable across diverse water matrices and multiple AOP configurations, where matrix effects can obscure risk-relevant byproducts. Practical deployment highlights a trade-off between model capacity and inference speed; thus, future work should prioritize model distillation for lightweight deployment and integrate online analytical signals to enable dynamic, near-real-time toxicity screening.

4. Conclusions

This study introduces an integrated computational framework (ToxD4C) that combines 3D deep learning with enhanced sampling molecular dynamics and in vitro validation, improving upon traditional QSAR methods constrained to 2D representations. An optional report generation utility is provided to summarize model outputs and user supplied evidence for transparent screening and prioritization. By capturing subtle stereochemical and electronic features essential for toxicity, the framework enables rapid hazard screening of transformation products and prioritizes candidates for targeted confirmation. Importantly, AOP treatment can generate products with higher intrinsic hazard potential than their parents, underscoring the need for follow-up quantification and chronic/mixture evaluation before risk characterization. Consequently, the developed open-source platform (Tox-Agents) enables rapid, interpretable, and actionable toxicity assessments, empowering stakeholders to proactively manage chemical risks. Ultimately, this approach facilitates informed design of safer chemicals, optimizes environmental remediation technologies, and represents a pivotal step toward predictive, proactive environmental protection.

Data availability

The source code is available at https://github.com/PhelanShao/tox-agents.

CRediT authorship contribution statement

Fulin Shao: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Weiying Li:** Writing – original draft, Project administration, Funding acquisition, Formal analysis. **Zhiwei Liang:** Validation, Supervision, Resources, Data curation. **Yu Zhou:** Visualization, Validation, Formal analysis. **Dawei Zhang:** Writing – original draft, Methodology, Funding acquisition, Data curation. **Yu Chang:** Writing – review & editing, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was financially supported by the National Key R & D Program of China during the 14th Five-Year Plan (Grant No. 2024YFC3810901-03) and the National Natural Science Foundation of China (Grant No. 52470012).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2026.125499.

References

Allouche, A., 2012. Software news and updates Gabedit — A graphical user interface for computational chemistry softwares. J. Comput. Chem. 32, 174–182. https://doi.org/10.1002/jcc.

Bailer, A.J., Hughes, M.R., Denton, D.L., Oris, J.T., 2000. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. Environ. Toxicol. Chem. 19, 141–150. https://doi.org/10.1002/etc.5620190117.

Bannwarth, C., Ehlert, S., Grimme, S., 2019. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. J. Chem. Theory Comput. 15, 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

Basak, S.C., Natarajan, R., Nowak, W., Miszta, P., Klun, J.A., 2007. Three dimensional structure-activity relationships (3D-QSAR) for insect repellency of diastereoisomeric compounds: A hierarchical molecular overlay approach. SAR QSAR Environ. Res. 18, 237–250. https://doi.org/10.1080/10629360701303784.

Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. TrAC Trends Anal. Chem. 66, 32–44. https://doi.org/10.1016/j.trac.2014.11.009.

Burkhard, L.P., 2021. Evaluation of published bioconcentration factor (BCF) and bioaccumulation factor (BAF) data for per-and polyfluoroalkyl substances across aquatic species. Environ. Toxicol. Chem. 40, 1530–1543.

Castellano-Hinojosa, A., Gallardo-Altamirano, M.J., González-López, J., González-Martínez, A., 2023. Anticancer drugs in wastewater and natural environments: A review on their occurrence, environmental persistence, treatment, and ecological risks. J. Hazard. Mater. 447, 130818. https://doi.org/10.1016/j.jhazmat.2023.130818.

Cecily Mary Glory, D., Madivanane, R., Sambathkumar, K., 2017. Electronic structure investigations of 1-amino-2,6-dimethylpiperidine by NMR spectral studies by ab initio and DFT calculations. Indian J. Pure Appl. Phys. 55, 638–648.

Chen, X., Reynolds, C.H., 2002. Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. J. Chem. Inf. Comput. Sci. 42, 1407–1414. https://doi.org/10.1021/ci025531g.

Cohen, M.D., Kargacin, B., Klein, C.B., Costa, M., 1993. Mechanisms of chromium carcinogenicity and toxicity. Crit. Rev. Toxicol. 23, 255–281. https://doi.org/10.3109/10408449309105012.

Couture, L.A., Abbott, B.D., Birnbaum, L.S., 1990. A critical review of the developmental toxicity and teratogenicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin: recent advances toward understanding the mechanism. Teratology 42, 619–627. https://doi.org/10.1002/tera.1420420606.

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Liu, T., Chang, B., Sun, X., Li, L., Sui, Z., 2024. A survey on In-context learning. https://doi.org/10.48550/arXiv.2301.00234.

Dumitrescu, A., Korpela, D., Heinonen, M., Verma, Y., Iakovlev, V., Garg, V., Lähdesmäki, H., 2025. E(3)-equivariant models cannot learn chirality: field-based molecular generation. https://doi.org/10.48550/arXiv.2402.15864.

Ertl, P., Gerebtzoff, G., Lewis, R., Muenkler, H., Schneider, N., Sirockin, F., Stiefl, N., Tosco, P., 2022. Chemical reactivity prediction: current methods and different application areas. Mol. Inf. 41, 1–11. https://doi.org/10.1002/minf.202100277.

Ferrando-Climent, L., Gonzalez-Olmos, R., Anfruns, A., Aymerich, I., Corominas, L., Barceló, D., Rodriguez-Mozaz, S., 2017. Elimination study of the chemotherapy drug tamoxifen by different advanced oxidation processes: transformation products and toxicity assessment. Chemosphere 168, 284–292. https://doi.org/10.1016/j.chemosphere.2016.10.057.

Feshuk, M., Kolaczkowski, L., Dunham, K., Davidson-Fritz, S.E., Carstens, K.E., Brown, J., Judson, R.S., Paul Friedman, K., 2023. The ToxCast pipeline: updates to curve-fitting approaches and database structure. Front. Toxicol. 5, 1275980. https://doi.org/10.3389/ftox.2023.1275980.

Fuchs, F., Worrall, D., Fischer, V., Welling, M., 2020. SE(3)-transformers: 3D Roto-translation equivariant attention networks. Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 1970–1981

Gu, W., Zhao, Y., Li, Q., Li, Y., 2019. Environmentally friendly polychlorinated naphthalenes (PCNs) derivatives designed using 3D-QSAR and screened using molecular docking, density functional theory and health-based risk assessment. J. Hazard. Mater. 363, 316–327. https://doi.org/10.1016/j.jhazmat.2018.09.060.

Halgren, T.A., 1996. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J. Comput. Chem. 17, 553–586. https://doi.org/10.1002/(sici)1096-987x(199604)17:5/6%253C553::aid-jcc3%253E3.0.co;2-t.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, S., Wu, L., Wang, F., Zhang, Y., Li, R., Zhao, Y., Xu, H., Gao, Z., Bo, X., 2024. A multiscale-information-embedded universal toxicity prediction framework. https://doi.org/10.21203/rs.3.rs-4926613/v1.

Helma, C., 2005. Predictive toxicology. Predict. Toxicol. 1–522. https://doi.org/10.1016/b978-0-12-387815-1.00033-2.

Hillebrand, L., Liang, X.J., Serafim, R.A.M., Gehringer, M., 2024. Emerging and re-emerging warheads for targeted covalent inhibitors: an update. J. Med. Chem. 67, 7668–7758. https://doi.org/10.1021/acs.jmedchem.3c01825.

Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S.A., Attene-Ramos, M., Zhao, T., Austin, C.P., Simeonov, A., 2016. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat. Commun. 7, 10425. https://doi.org/10.1038/ncomms10425.

Idakwo, G., Thangapandian, S., Luttrell, J., Li, Y., Wang, N., Zhou, Z., Hong, H., Yang, B., Zhang, C., Gong, P., 2020. Structure–activity relationship-based chemical classification of highly imbalanced Tox21 datasets. J. Cheminform. 12, 66. https://doi.org/10.1186/s13321-020-00468-x.

LaKind, J.S., Overpeck, J., Breysse, P.N., Backer, L., Richardson, S.D., Sobus, J., Sapkota, A., Upperman, C.R., Jiang, C., Beard, C.B., Brunkard, J.M., Bell, J.E., Harris, R., Chretien, J.-P., Peltier, R.E., Chew, G.L., Blount, B.C., 2016. Exposure science in an age of rapidly changing climate: challenges and opportunities. J. Expo. Sci. Environ. Epidemiol. 26, 529–538. https://doi.org/10.1038/jes.2016.35.

Lambert, F.N., Vivian, D.N., Raimondo, S., Tebes-Stevens, C.T., Barron, M.G., 2022. Relationships between aquatic toxicity, chemical hydrophobicity, and mode of action: Log Kow revisited. Arch. Environ. Contam. Toxicol. 83, 326–338. https://doi.org/10.1007/s00244-022-00944-5.

Liu, S., Wang, H., Liu, W., Lasenby, J., Guo, H., Tang, J., 2022. Pre-training molecular graph representation with 3D geometry. https://doi.org/10.48550/arXiv.2110.07728.

Liu, X., Yang, L., Wang, M., Zheng, M., Li, C., Qin, L., Liu, G., 2021. Insights into the formation and profile of chlorinated polycyclic aromatic hydrocarbons during chlorobenzene and chloroethylene manufacturing processes. Environ. Sci. Technol. 55, 15929–15939. https://doi.org/10.1021/acs.est.1c05688.

LoPachin, R.M., Gavin, T., 2014. Molecular mechanisms of aldehyde toxicity: A chemical perspective. Chem. Res. Toxicol. 27, 1081–1091. https://doi.org/10.1021/tx5001046.

Mangalathu, S., Hwang, S.H., Jeon, J.S., 2020. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Eng. Struct. 219, 110927. https://doi.org/10.1016/j.engstruct.2020.110927.

Mansouri, K., Karmaus, A., 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., Rallo, R., Roncaglioni, A., Rowlands, C., Ruiz, P., Russo, D., 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., 2021. Erratum: Catmos: collaborative acute toxicity modeling suite. Environ. Health Perspect 129, 1–18. https://doi.org/10.1289/EHP9883.

Mayr, A., Klambauer, G., Unterthiner, T., Hochreiter, S., 2016. DeepTox: toxicity prediction using Deep Learning. Front. Environ. Sci. 3. https://doi.org/10.3389/fenvs.2015.00080.

McCord, J.P., Groff, L.C., Sobus, J.R., 2022. Quantitative non-targeted analysis: bridging the gap between contaminant discovery and risk characterization. Environ. Int. 158, 107011. https://doi.org/10.1016/j.envint.2021.107011.

Neel, A.J., Hilton, M.J., Sigman, M.S., Toste, F.D., 2017. Exploiting non-covalent π interactions for catalyst design. Nature 543, 637–646. https://doi.org/10.1038/nature21701.

Nicholls, A., Grant, J.A., 2005. Molecular shape and electrostatics in the encoding of relevant chemical information. J. Comput. Aided. Mol. Des. 19, 661–686. https://doi.org/10.1007/s10822-005-9019-x.

Raies, A.B., Bajic, V.B., 2016. In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 6, 147–172. https://doi.org/10.1002/wcms.1240.

Richard, A.M., Huang, R., Waidyanatha, S., Shinn, P., Collins, B.J., Thillainadarajah, I., Grulke, C.M., Williams, A.J., Lougee, R.R., Judson, R.S., Houck, K.A., Shobair, M., Yang, C., Rathman, J.F., Yasgar, A., Fitzpatrick, S.C., Simeonov, A., Thomas, R.S., Crofton, K.M., Paules, R.S., Bucher, J.R., Austin, C.P., Kavlock, R.J., Tice, R.R., 2021. The Tox21 10K compound Library: Collaborative Chemistry advancing toxicology. Chem. Res. Toxicol 34, 189–216. https://doi.org/10.1021/acs.chemrestox.0c00264.

Roy, K., Mitra, I., 2012. Electrotopological State atom (E-State) index in drug design, QSAR, property prediction and toxicity assessment. Curr. Comput. Aided-Drug Des. 8, 135–158. https://doi.org/10.2174/157340912800492366.

Seal, S., Mahale, M., García-Ortegón, M., Joshi, C.K., Hosseini-Gerami, L., Beatson, A., Greenig, M., Shekhar, M., Patra, A., Weis, C., Mehrjou, A., Badré, A., Paisley, B., Lowe, R., Singh, S., Shah, F., Johannesson, B., Williams, D., Rouquie, D., Clevert, D.-A., Schwab, P., Richmond, N., Nicolaou, C.A., Gonzalez, R.J., Naven, R., Schramm, C., Vidler, L.R., Mansouri, K., Walters, W.P., Wilk, D.D., Spjuth, O., Carpenter, A.E., Bender, A., 2025. Machine learning for toxicity prediction using chemical structures: pillars for success in the real world. Chem. Res. Toxicol. 38, 759–807. https://doi.org/10.1021/acs.chemrestox.5c00033.

Shanle, E.K., Xu, W., 2011. Endocrine disrupting chemicals targeting estrogen receptor signaling: identification and mechanisms of action. Chem. Res. Toxicol. 24, 6–19. https://doi.org/10.1021/tx100231n.

Sharma, B., Chenthamarakshan, V., Dhurandhar, A., Pereira, S., Hendler, J.A., Dordick, J.S., Das, P., 2023. Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. Sci. Rep. 13, 4908. https://doi.org/10.1038/s41598-023-31169-8.

Shi, X.-X., Wang, Z.-Z., Sun, X.-L., Wang, Y.-L., Liu, H.-X., Wang, F., Hao, G.-F., Yang, G.-F., 2023. Toxicological data bank bridges the gap between environmental risk assessment and green organic chemical design in one health world. Green Chem. 25, 2170–2219. https://doi.org/10.1039/D2GC03973G.

Subbiah, S.S., Chinnappan, J., 2020. A review of short term load forecasting using deep learning. Int. J. Emerg. Technol. 11, 378–384.

Tran, T.T.Van, Surya Wibowo, A., Tayara, H., Chong, K.T., 2023. Artificial intelligence in drug toxicity prediction: recent advances, challenges, and future perspectives. J. Chem. Inf. Model. 63, 2628–2643. https://doi.org/10.1021/acs.jcim.3c00200.

Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G., 2014. PLUMED 2: new feathers for an old bird. Comput. Phys. Commun. 185, 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.

Trizio, E., Rizzi, A., Piaggi, P.M., Invernizzi, M., Bonati, L., 2024. Advanced simulations with PLUMED: OPES and machine learning collective variables. https://doi.org/10.48550/arXiv.2410.18019.

Van Den Berg, M., Birnbaum, L., Bosveld, A.T., Brunström, B., Cook, P., Feeley, M., Giesy, J.P., Hanberg, A., Hasegawa, R., Kennedy, S.W., Kubiak, T., Larsen, J.C., Van Leeuwen, F.X., Liem, A.K., Nolt, C., Peterson, R.E., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F., Zacharewski, T., 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. Environ. Health Perspect 106, 775–792. https://doi.org/10.1289/ehp.98106775.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph attention networks. https://doi.org/10.48550/arXiv.1710.10903.

Verma, J., Khedkar, V., Coutinho, E., 2010. 3D-QSAR in drug design - a review. Curr. Top. Med. Chem. 10, 95–115. https://doi.org/10.2174/156802610790232260.

Wang, L., Wang, S., Yang, H., Li, S., Wang, X., Zhou, Y., Tian, S., Liu, L., Bai, F., 2024. Conformational space profiling enhances generic molecular representation for AI-powered ligand-based drug discovery. Adv. Sci. 2403998, 1–14. https://doi.org/10.1002/advs.202403998.

Wang, S., Zeng, Y., Yang, H., Tian, S., Zhou, Y., Wang, L., Chen, X., Wang, H., Gao, Z., Bai, F., 2025. Discovery of novel GluN1/GluN3A NMDA receptor inhibitors using a deep learning-based method. Acta Pharmacol. Sin. 1–8. https://doi.org/10.1038/s41401-025-01571-1.

Wang, Z., Walker, G.W., Muir, D.C.G., Nagatani-Yoshida, K., 2020. Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. Environ. Sci. Technol. 54, 2575–2584. https://doi.org/10.1021/acs.est.9b06379.

Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., Barzilay, R., 2019. Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model 59, 3370–3388. https://doi.org/10.1021/acs.jcim.9b00237.

Zeng, J., Zhang, D., Lu, D., Mo, P., Li, Zeyu, Chen, Y., Rynik, M., Huang, L., Li, Ziyao, Shi, S., Wang, Yingze, Ye, H., Tuo, P., Yang, J., Ding, Y., Li, Y., Tisi, D., Zeng, Q., Bao, H., Xia, Y., Huang, J., Muraoka, K., Wang, Yibo, Chang, J., Yuan, F., Bore, S.L., Cai, C., Lin, Y., Wang, B., Xu, J., Zhu, J.-X., Luo, C., Zhang, Y., Goodall, R.E.A., Liang, W., Singh, A.K., Yao, S., Zhang, J., Wentzcovitch, R., Han, J., Liu, J., Jia, W., York, D.M., E, W., Car, R., Zhang, L., Wang, H., 2023. DeePMD-kit v2: A software package for deep potential models. J. Chem. Phys. 159, 054801. https://doi.org/10.1063/5.0155600.

Zhang, H., Hou, X., Liu, J., Su, J., Li, S., Chen, W., Guo, F., Jiang, G., 2025. Diverse in vitro liver models reveal comprehensive biotransformation pathways of tetrabromobisphenol A. Environ. Int. 109848. https://doi.org/10.1016/j.envint.2025.109848.

Zhang, S., Zhao, D., Cui, Q., 2024. Gap-δenergy, a new metric of the bond energy State, assisting to predict molecular toxicity. ACS. Omega 9, 17839–17847. https://doi.org/10.1021/acsomega.3c07682.

Zhao, Z., Qin, J., Gou, Z., Zhang, Y., Yang, Y., 2020. Multi-task learning models for predicting active compounds. J. Biomed. Inf. 108, 103484. https://doi.org/10.1016/j.jbi.2020.103484.

Zhong, S., Guan, X., 2023. Count-based Morgan fingerprint: A more efficient and interpretable molecular representation in developing machine learning-based predictive regression models for water contaminants' Activities and properties. Environ. Sci. Technol. 57, 18193–18202. https://doi.org/10.1021/acs.est.3c02198.

Zhong, S., Hu, J., Yu, X., Zhang, H., 2021a. Molecular image-convolutional neural network (CNN) assisted QSAR models for predicting contaminant reactivity toward OH radicals: transfer learning, data augmentation and model interpretation. Chem. Eng. J. 408, 127998. https://doi.org/10.1016/j.cej.2020.127998.

Zhong, S., Zhang, K., Wang, D., Zhang, H., 2021b. Shedding light on "Black Box" machine learning models for predicting the reactivity of HO[rad] radicals toward organic compounds. Chem. Eng. J. 405, 126627. https://doi.org/10.1016/j.cej.2020.126627.

Zhong, S., Zhang, Y., Zhang, H., 2022. Machine learning-assisted QSAR models on contaminant reactivity toward four oxidants: combining small data sets and knowledge transfer. Environ. Sci. Technol. 56, 681–692. https://doi.org/10.1021/acs.est.1c04883.

Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., Ke, G., 2023. Uni-Mol: A universal 3D molecular representation learning framework. https://doi.org/10.26434/chemrxiv-2022-jjm0j-v4.

Zubrod, J.P., Galic, N., Vaugeois, M., Dreier, D.A., 2024. Bio-QSARs 2.0: unlocking a new level of predictive power for machine learning-based ecotoxicity predictions by exploiting chemical and biological information. Environ. Int. 186, 108607. https://doi.org/10.1016/j.envint.2024.108607.