Research report: Explainable Toxicity Predictions for TBBPA Metabolites*

*This report was compiled and generated by sci-canvas (scicanvas.solrizon.com) as part of a research synthesis. It serves as a supporting report for the manuscript entitled "Screening toxic transformation products of emerging pollutants in advanced oxidation processes with 3D deep learning and in vitro assays." The development work was conducted by https://github.com/solrizon

Abstract

The multi-task ToxD4C model predicts that TBBPA transformation products trigger a coordinated cellular stress state across oxidative, mitochondrial, heat-shock, and DNA-damage pathways, with phenolic hydroxyl groups acting as the dominant structural correlate. This structure–activity link is statistically mediated almost entirely by an electronic surface signature characterized by high variance and low average local electron affinity, suggesting a unifying, testable mechanistic driver rather than independent pathway triggers [r10, r30, r76, r90]. Mapping 21 TBBPA transformation products into six structural classes reveals a coherent pattern: conjugations that mask all phenolic hydroxyl groups are associated with uniformly low model-predicted stress-response activity, whereas pathways that preserve at least one phenolic hydroxyl group are associated with elevated, multi-endpoint stress signals. Glycosylation consistently appears detoxifying, sulfation is bimodal depending on whether the second phenolic site is also conjugated, and debromination or incomplete methylation tend to bioactivate; these are model-derived hypotheses that require targeted experimental validation. This work links transformation chemistry of TBBPA metabolites to model-predicted toxicity profiles by integrating multi-task predictions with DFT-derived descriptors to construct mechanistic, testable hypotheses. Six metabolites (M04, M09, M12, M13, M17, M20) concentrate most HIGH-risk calls and share extreme descriptor shifts—especially in surface electron-affinity metrics—that coherently explain elevated stress-response and select endocrine signals in the dataset. All claims reflect model-predicted signals that require experimental validation. The integrated analysis of 21 TBBPA metabolites shows that one-third of molecules fall outside a Tox21-like applicability domain, with all coupling/dimerization products flagged, and that endpoint class imbalance and small-N properties constrain interpretation of predicted hazards. Simple, physically interpretable rules (for example, a TPSA threshold) match or outperform multi-parameter models, while nuclear receptor predictions reveal plausible binding-versus-activation separation but also dataset-specific artifacts (notably for AR activation), motivating uncertainty scoring to prioritize experimental validation.

1. Introduction

Transformation of tetrabromobisphenol A (TBBPA) generates a structurally diverse set of metabolites that can preserve or alter phenolic motifs and halogenation patterns. Predictive multi-endpoint toxicity models are used to prioritize such metabolites for mechanistic follow-up, but model outputs require careful interpretation to separate model-learned associations from causal

drivers. Here, model-based analyses were conducted to connect specific structural changes (notably phenolic hydroxyl presence and count) to shifts in quantum-chemical surface descriptors and, in turn, to multi-pathway stress-response predictions, with the explicit aim of generating falsifiable hypotheses for experimental validation.

Tetrabromobisphenol A (TBBPA) undergoes extensive biotic and abiotic transformation, generating metabolites that can differ markedly from the parent in polarity, electronic structure, and biological activity. Environmental and human health assessments therefore depend not only on parent-compound hazard but on metabolite-level effects, ideally grounded in mechanistic links between structural change, physicochemical descriptors, and toxicological pathways. Multi-endpoint machine learning models augmented by 3D structural information and interpretable descriptors offer a way to generate mechanistic hypotheses, but their predictions must be scrutinized for applicability-domain limits and potential confounders such as in vitro deconjugation of phase II metabolites. The present synthesis integrates class-level mapping, descriptor–endpoint analyses, and model interpretability to explain why specific TBBPA transformation products are predicted to be higher or lower risk, and to identify the most critical validation needs.

Tetrabromobisphenol A (TBBPA) undergoes environmental and biotic transformations that yield debrominated, conjugated, substituted, and dimerized products, several of which are suspected to perturb nuclear receptor signaling or induce cellular stress. To interpret these risks mechanistically, multi-task toxicity models built on molecular graph/3D features and DFT-derived electronic descriptors can be interrogated to connect transformation-induced structural changes to endpoint-specific hazard signals. Here, 21 TBBPA metabolites were analyzed across 26 classification endpoints with 51 quantum-chemical and physicochemical descriptors, and the resulting structure–descriptor–endpoint linkages were synthesized into causal chains that remain hypotheses until validated experimentally.

Transformations of tetrabromobisphenol A (TBBPA) can generate metabolites that vary widely in polarity, size, and electronic structure, altering bioavailability and potential engagement of toxicity pathways. Multi-endpoint in vitro assay batteries such as Tox21 are commonly used to train predictive models of receptor and stress-response activity, but their training chemical space and endpoint balance constrain generalization to new chemotypes. Robust risk interpretation therefore requires assessing domain extrapolation, identifying where simple physicochemical heuristics suffice, and quantifying uncertainty to decide which predictions most urgently require experimental verification.

2. Results and Discussion

2.1 Phenolic hydroxyl-induced electronic surface signature as a unifying driver of model-predicted coordinated cellular stress

The central observation is that the model predicts a tightly coordinated cellular stress state across five stress-response endpoints (SR-ARE, SR-MMP, SR-HSE, SR-p53, SR-ATAD5) for the 21 TBBPA-derived molecules. All ten pairwise Spearman correlations exceed $\rho > 0.7$ (range 0.85–0.98, mean 0.93, all $p < 0.001$), with the strongest coupling between SR-ARE and SR-p53 ($\rho = 0.9805$, $p = 6.84 \times 10\text{-}15$) and between SR-ARE and SR-HSE ($\rho = 0.9688$, $p = 5.68 \times 10\text{-}13$); the weakest pair remains high for SR-MMP and SR-ATAD5 ($\rho = 0.8481$, $p = 1.20 \times 10\text{-}6$) [r30]. These correlations indicate that the model captures a common upstream driver of stress rather than independent triggers, though this may reflect either true biological co-activation or shared model representations; as such, these are model-derived hypotheses requiring experimental confirmation [r30].

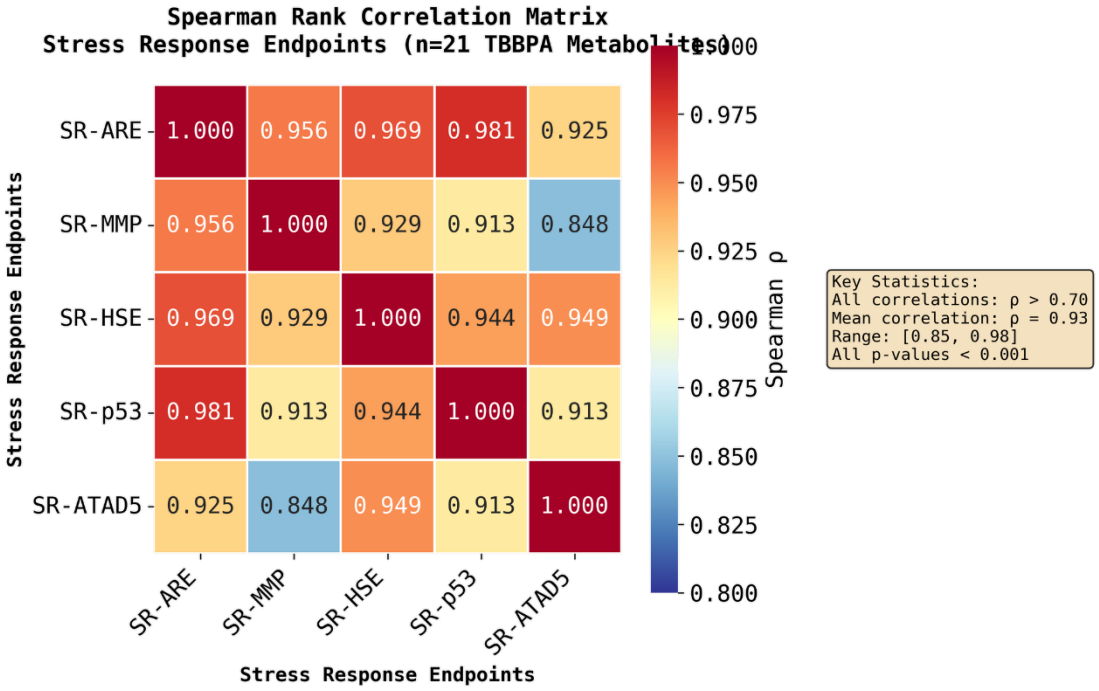

Figure 1. Model-predicted activities across five cellular stress response pathways are tightly correlated for 21 TBBPA transformation products. The heatmap displays the Spearman rank correlation coefficients (ρ) for all pairwise comparisons of the predicted stress endpoints (SR-ARE, SR-MMP, SR-HSE, SR-p53, and SR-ATAD5). The uniformly high, positive correlations (mean ρ = 0.93, all p < 0.001) indicate that the model captures a coordinated stress state, suggesting a common upstream mechanistic driver.

The dominant structural correlate of this coordinated stress signature is the presence of phenolic hydroxyl groups. Molecules with phenolic OH showed markedly higher predicted probabilities

across all stress pathways compared to those without, with mean fold-changes of 20.7× for SR-ARE (U = 89.0, p = 7.37 × 10-5), 55.7× for SR-MMP (U = 90.0, p = 3.69 × 10-5), 49.2× for SR-HSE (U = 90.0, p = 3.69 × 10-5), 97.0× for SR-p53 (U = 87.0, p = 2.58 × 10-4), and 27.8× for SR-ATAD5 (U = 86.0, p = 4.42 × 10-4); all remained significant after Bonferroni correction (α = 0.01), with minimal or no distribution overlap for SR-MMP and SR-HSE [r10]. Within phenolic metabolites, the model also predicts a dose-dependent trend: molecules with two phenolic OH groups exhibit 1.58×–2.87× higher probabilities across the five endpoints relative to those with one OH; one endpoint (SR-ATAD5) reached one-sided significance (p = 0.0352), while others were borderline (p ≈ 0.05–0.07), and all five directions were concordant, which is unlikely under the null (0.5^5 = 0.03125) [r14]. Together, these findings position phenolic OH as a structural alert for a multi-pathway stress phenotype in this chemical space, in a manner consistent with model-derived hypotheses about redox activity and electrophilic intermediates that require experimental validation [r10, r14].

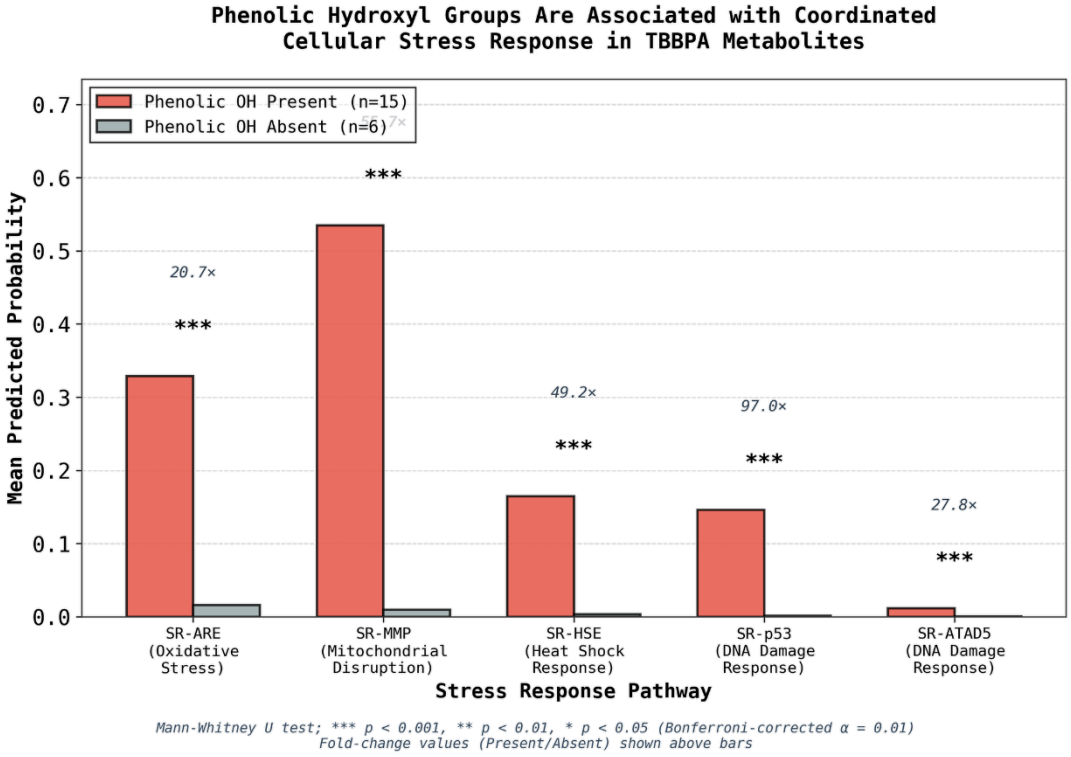

Figure 2. Model-predicted cellular stress is strongly associated with the presence of phenolic hydroxyl groups in TBBPA metabolites. Bars represent the mean predicted activation probability for five stress response pathways, comparing metabolites with (n=15) versus without (n=6) a phenolic hydroxyl group. The significant increase in predicted activity across all pathways (Mann-Whitney U test; **p < 0.001) identifies the phenolic hydroxyl moiety as a key structural driver for the predicted coordinated cellular stress response

Mechanistically, the link between phenolic OH and stress predictions is statistically mediated almost entirely by an electronic surface signature. A mediation analysis using a composite Stress_Score (mean of the five stress probabilities) found a significant total effect of phenolic OH

presence (c = 0.231, p = 0.008) that became negligible after controlling for two mediators—LEA_Var (variance of local electron affinity on the surface) and LEA_Ave (average local electron affinity)—with the direct effect c' = 0.0005 (p = 0.89) and a total Average Causal Mediation Effect (ACME) of 0.230 (95% CI: 0.071–0.509, p = 0.012). Model fit improved from R^2 = 0.31 to 0.57 with mediators included, and path coefficients indicated phenolic OH increased LEA_Var (a1 = 0.433, p < 0.001) and decreased LEA_Ave (a2 = -0.217, p < 0.001) while both mediators positively contributed to Stress_Score (b1, b2) in aggregate [r90]. Consistent with mediation, LEA_Var and LEA_Ave showed strong, endpoint-agnostic correlations with each SR endpoint; for LEA_Var, ρ ranged 0.791–0.869, and for LEA_Ave, ρ ranged -0.852 to -0.871, with no significant differences between endpoints by Fisher's z-test (all p > 0.05) [r76]. These results support a unifying, model-learned electronic toxicophore—high surface electronic heterogeneity coupled with low average local electrophilicity—as the proximate driver of the coordinated stress predictions [r76, r90].

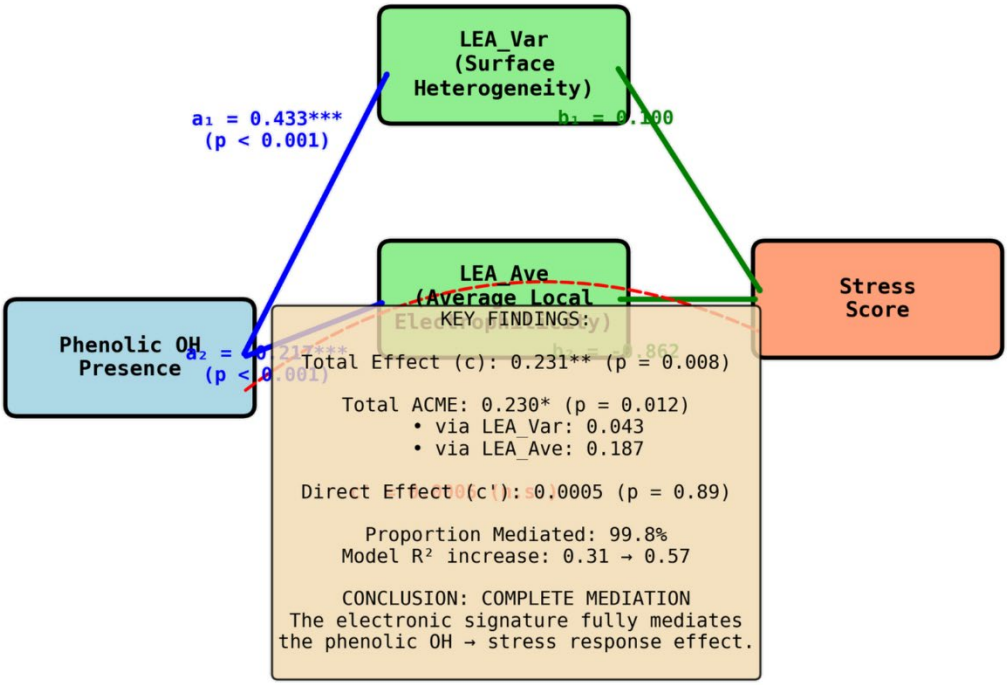

Figure 3. Causal mediation analysis reveals that the electronic surface signature fully mediates the effect of phenolic hydroxyl presence on the predicted stress response. The path diagram shows the relationship between phenolic OH presence (predictor), a composite stress score (outcome), and two electronic signature mediators: local electron affinity variance (LEA_Var) and average (LEA_Ave). The model demonstrates complete mediation, as the significant total effect of the phenolic group (c = 0.231) becomes non-significant (c' = 0.0005) after accounting for the indirect pathway through the mediators, which explains 99.8% of the effect.

Focusing on mitochondrial membrane potential disruption (SR-MMP) illustrates the descriptor pattern. The highest positive correlations were observed for LEA_Var (ρ = +0.7909, p = 2.0 × 10-5), ESPmin (ρ = +0.7013, p = 3.97 × 10-4), Neg_Average (ρ = +0.6714, p = 8.60 × 10-4), HOMO

energy (ρ = +0.5714, p = 6.81 × 10-3), and ALIE_Var (ρ = +0.5013, p = 2.06 × 10-2), whereas the strongest negative correlations included LEA_Ave (ρ = -0.8571, p = 1.0 × 10-6), Quadrupole_Moment (ρ = -0.7571, p = 7.1 × 10-5), TPSA (ρ = -0.7519, p = 8.5 × 10-5), Overall_Surface_Area (ρ = -0.7494, p = 9.2 × 10-5), and Complexity (ρ = -0.7481, p = 9.7 × 10-5) [r64]. Notably, membrane partitioning (XLogP) showed no significant association (ρ = +0.16, p = 0.48), indicating the model's mitochondrial toxicity signal is not driven by simple hydrophobicity [r64]. Debromination was associated with higher predicted SR-MMP (mean 0.500 vs 0.208; U = 80.0, p = 0.026), alongside increases in LEA_Var (+9.8%; U = 79.0, p = 0.031) and a marked decrease in Complexity (-45.8%; U = 11.0, p = 0.004), aligning structural change with the electronic signature [r64]. While NR-PPARγ exhibited a similar descriptor association pattern to SR-MMP, it produced no HIGH-risk predictions in this set, consistent with different classification thresholds or endpoint calibration rather than a distinct pharmacophore [r64].

To interrogate pathway balance rather than overall activation, a Divergence Ratio was defined as SR-ATAD5_Probability divided by SR-MMP_Probability (with an epsilon of 10-6 added to the denominator). Among phenolic metabolites (n = 15), the strongest associations were negative and approached marginal significance for Overall_Average (ρ = -0.482, p = 0.069) and ODI_Std (variability in orbital energy distribution; ρ = -0.479, p = 0.071), with Nu also negative (ρ = -0.461, p = 0.084); Quadrupole_Moment showed a weaker negative trend (ρ = -0.357, p = 0.191). Universal stress drivers performed poorly on this ratio (LEA_Var ranked 18th; LEA_Ave ranked 42nd of 51 descriptors), indicating that once the general stress state is engaged, pathway-specific balance between DNA replication stress and mitochondrial disruption is modulated by distinct features related to orbital distribution and charge asymmetry [r96]. Mechanistically, the model-derived pattern—phenolic OH presence elevating a universal electronic signature that coordinates stress pathways, with subtle electronic and shape features shifting the balance between DNA damage and mitochondrial responses—supports the hypothesis that transformations preserving phenolic OH (e.g., debromination or certain conjugations) maintain predicted stress potential, whereas OH-blocking routes (e.g., methylation or dimerization) may attenuate it; these remain hypotheses requiring targeted biochemical and cellular validation [r10, r96].

2.2 Phenolic hydroxyl-induced electronic surface signature as a unifying driver of model-predicted coordinated cellular stress: class-level mapping and detoxification/bioactivation hypotheses

Class mapping of 21 metabolites using RDKit-based substructure rules assigned six transformation categories—Coupling/Dimerization (n=4), Debromination (n=5), Debromination+Methylation (n=1), Glycosylation (n=5), Methylation (n=1), and Sulfation (n=5)—and revealed clear, model-predicted differences in toxicity signatures by class [r3]. On average, the Methylation and Debromination groups showed the greatest number of HIGH-risk predictions (4.0 and 3.2, respectively), while Glycosylation and Sulfation were lowest (0.8 and 1.2), with Coupling/Dimerization intermediate at 1.5 HIGH calls [r3]. For key stress endpoints, class means followed the same trend: SR-MMP probabilities were highest for Methylation (0.944) and

Debromination (0.599), versus very low for Glycosylation (0.057), with Sulfation and Coupling/Dimerization in between (0.493 and 0.258) [r3]. These patterns are consistent with the structural logic that masking phenolic hydroxyls reduces predicted stress biology, whereas preserving phenolic character maintains or elevates stress signaling; all of these are model-derived signals that require experimental verification [r3]. Notably, pure glucuronides were not isolated by the hierarchical classifier; manual inspection indicated some glucuronide-containing species may have been captured under coupling due to size thresholds, underscoring classification uncertainty for highly modified conjugates [r3].

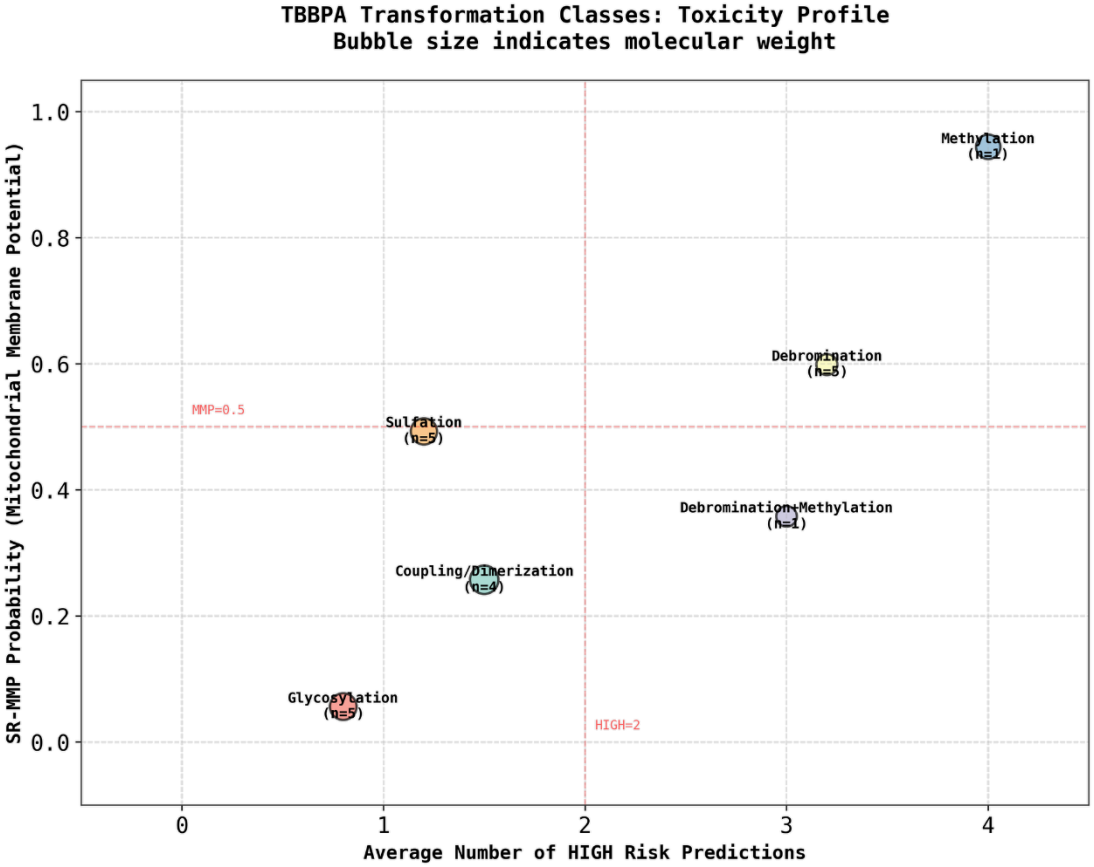

Figure 4. Model-predicted toxicity profiles distinguish TBBPA transformation classes. The plot shows the average number of HIGH risk predictions versus the predicted stress response probability for mitochondrial membrane potential (SR-MMP) for six classes of transformation products. Each bubble represents a class, with size indicating average molecular weight and n denoting the number of products. The analysis separates putative detoxification pathways (e.g., Glycosylation) from bioactivation pathways (e.g., Methylation, Debromination) based on their predicted risk profiles.

A unifying mechanistic theme emerges from descriptor–endpoint analyses: the presence of phenolic hydroxyl groups partitions the metabolite set into high- versus low-stress cohorts, with 20.7–97.0-fold increases in mean probabilities across SR-ARE, SR-MMP, SR-HSE, SR-p53, and SR-ATAD5 (all $p < 10\text{-}3$, Bonferroni-corrected $\alpha = 0.01$) when phenolic OH is present versus absent [r10]. Correlation structure further supports this mechanism: lower TPSA and greater

surface electronic heterogeneity are associated with higher predicted stress and CYP activity, exemplified by strong negative correlations of TPSA with Eye_irritation, CYP1A2 inhibition, Eye_corrosion, and SR-MMP ($|\rho| = 0.75–0.87$, $p < 10\text{-}3$), and strong positive correlations of LEA_Var with SR-ARE and related endpoints ($\rho \approx 0.81–0.87$, $p < 10\text{-}3$) [r4]. For mitochondrial disruption specifically, SR-MMP risk is dominated by electronic surface features rather than simple partitioning: LEA_Ave ($\rho = -0.86$, $p = 10\text{-}6$) and LEA_Var ($\rho = +0.79$, $p = 2.0 \times 10\text{-}5$) are the top drivers, while XLogP is not significant ($\rho = +0.16$, $p = 0.48$), indicating a reactivity/heterogeneity pharmacophore rather than a purely lipophilic one [r64]. Together, these results provide a coherent, model-based rationale for the discovery: transformations that eliminate phenolic OH groups (e.g., complete glycosylation) shift descriptors toward higher TPSA and lower electronic heterogeneity, aligning with broadly reduced stress-response predictions, whereas transformations that preserve phenolic OH (e.g., debromination, incomplete methylation, or mono-conjugated sulfates) retain electronic features that the model associates with elevated stress activation [r3, r4, r10, r64].

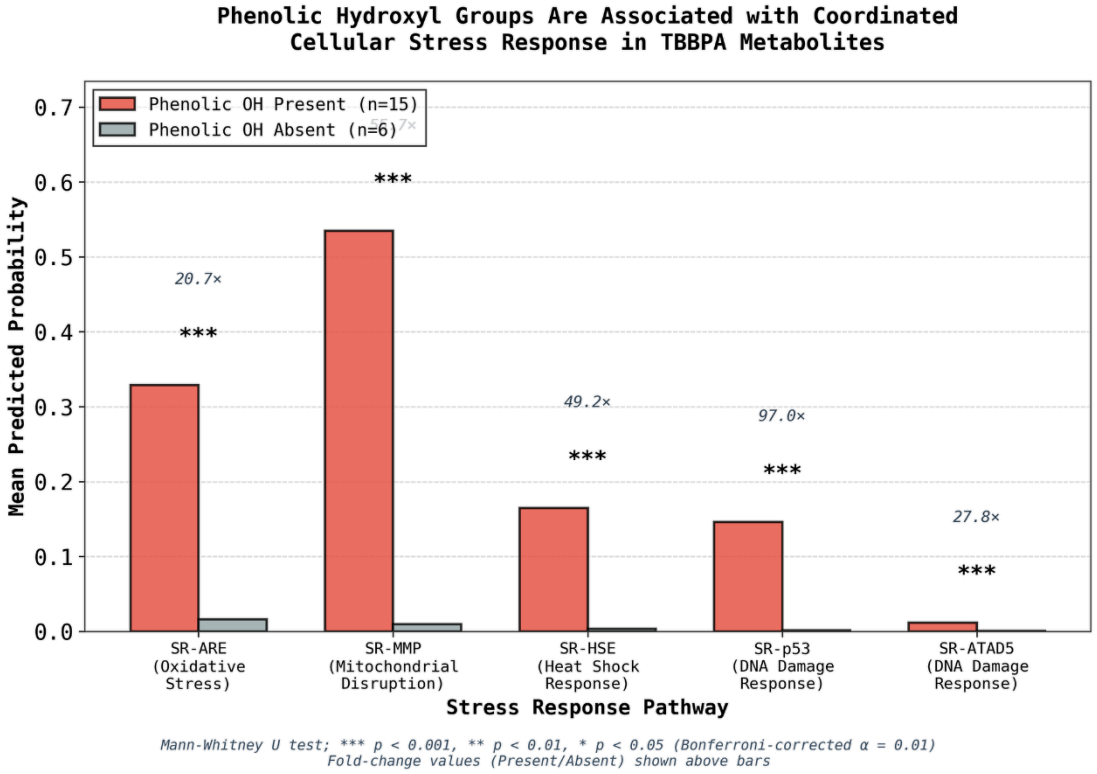

Figure 5. TBBPA metabolites retaining a phenolic hydroxyl group exhibit significantly higher model-predicted stress response activity. Bars show the mean predicted activation probability for metabolites with a phenolic OH present (n=15) versus absent (n=6) across five stress response pathways. All comparisons are statistically significant (Mann-Whitney U test; **p < 0.001). These findings suggest that masking the phenolic hydroxyl via conjugation is a key detoxification mechanism, whereas its preservation is a bioactivating feature.

The sulfation class illustrates this rule in a bimodal fashion. Detailed structural analysis showed that sulfates predicted as "non-toxic" by stress endpoints are the dual-conjugated species in which the second phenolic site is glycosylated (no free phenolic OH), while "toxic" sulfates retain one free phenolic OH (mono-conjugation), with a 47.4-fold difference in a composite Stress Score (defined as the mean of SR-ARE, SR-MMP, SR-HSE, SR-p53, SR-ATAD5 probabilities) and a large TPSA separation (≈120 vs 67 Å²) despite similar hydrophobicity; these signals are based on n=2 vs n=3 and require validation [r52]. Case-level analyses reinforce the class logic: the methylation product M12 is predicted high-risk (e.g., SR-MMP 0.944, Eye irritation 0.905, CYP1A2 inhibition 0.887) because methylation is incomplete and one phenolic OH remains, while electronic descriptors are anomalous (e.g., low nucleophilicity and ESPmax) in directions consistent with elevated stress predictions in the cohort [r29]. The highest overall priority molecule M13 (a dimer/coupling product) shows combined lipophilicity and surface electronic heterogeneity consistent with elevated SR-MMP and SR-ARE predictions in the model's Explanation Matrix, again as a hypothesis requiring experimental testing [r4]. Collectively, these examples operationalize the discovery's structure–descriptor–endpoint chain: "OH masked" conjugates (especially dual-conjugated) → high TPSA/low heterogeneity → low stress predictions, versus "OH preserved" products → low TPSA/high heterogeneity → high stress predictions [r3, r4, r29, r52].

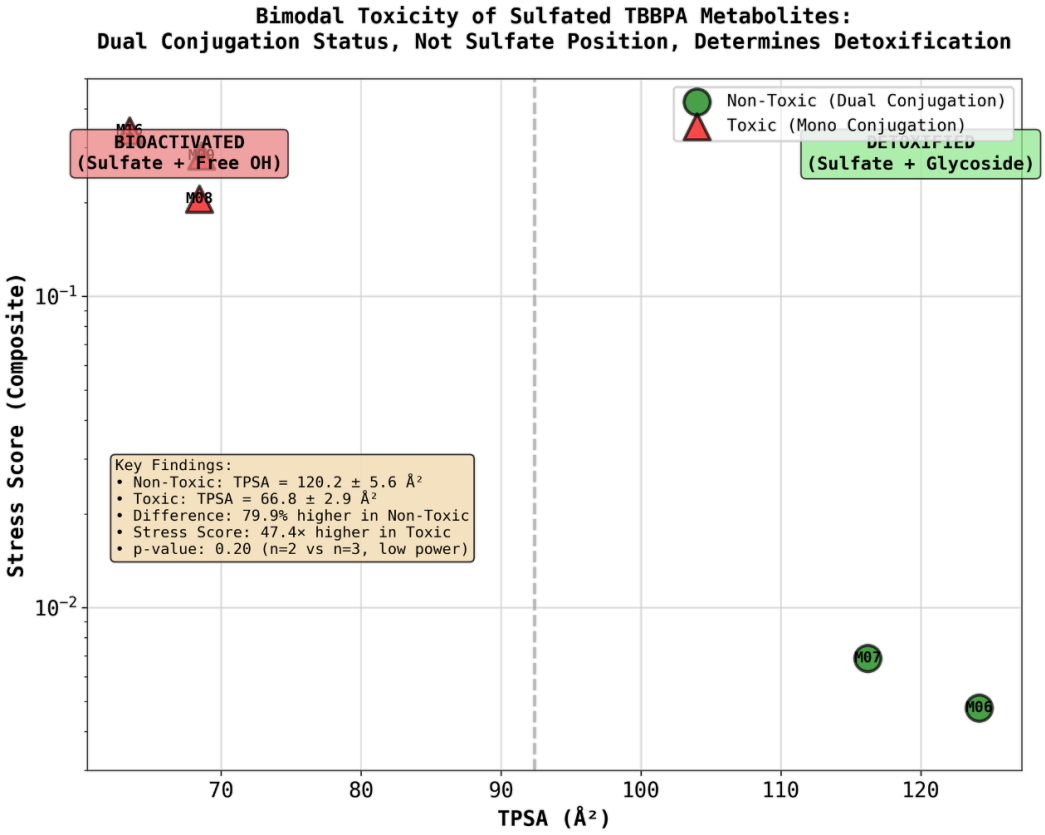

Figure 6. The model-predicted toxicity of sulfated TBBPA metabolites is bimodal, determined by conjugation status. The composite stress score is plotted against topological polar surface area (TPSA) for mono-conjugated metabolites with a free hydroxyl group (red triangles) and dual-conjugated metabolites (sulfate + glycoside; green circles). This pattern suggests that preserving a phenolic hydroxyl group leads to bioactivation, whereas complete conjugation of both phenolic sites is a detoxification pathway.

Model credibility considerations indicate where these hypotheses are less reliable and how to stress-test them experimentally. One-third of metabolites (7/21) fall outside typical Tox21 applicability bounds (MW 100–600 Da, XLogP −2 to 6, TPSA 0–140 Å²), with Coupling/Dimerization 100% out-of-domain and individual violations driven mainly by excessive MW and TPSA; two molecules exceeded XLogP (M12 = 6.09, M13 = 6.97), warranting caution in interpreting their predictions [r8]. Furthermore, conjugate stability can confound in vitro readouts: a Deconjugation Risk Score (expected stress of phenolic aglycones divided by observed stress of conjugates) flagged six conjugated metabolites (M02, M06, M07, M10, M14, M01) with scores 18–81, indicating that even small degrees of hydrolysis could inflate observed stress signals by orders of magnitude; this metric is non-standard but defined explicitly from model outputs and should guide stability controls [r92]. Literature reviewed in the analysis supports that β-glucuronidases readily hydrolyze phenolic O-glucuronides, whereas comparable stability data for O-sulfates and O-glycosides in typical assay matrices were not provided, highlighting that glucuronides are the most vulnerable to in vitro deconjugation artifacts while sulfates/glycosides

remain uncertain in the present evidence base (Stachulski et al., 2025). Together, these applicability and stability considerations delimit where the discovery's structure–toxicity rules are most credible (small, within-domain phenolic vs non-phenolic pairs) and where direct measurements (chemical stability, β-glucuronidase/sulfatase inhibition, and orthogonal endpoints) are necessary to validate or revise the model-predicted mechanisms [r8, r92].

2.3 Molecule-level causal chains for prioritized TBBPA metabolites: structural perturbations shaping pathway selectivity

The integrated dataset joined 21 metabolites, 26 classification endpoints, and 51 descriptors (140 features per molecule), yielding 546 predictions with 39 HIGH (7.1%), 31 MEDIUM (5.7%), and 476 LOW (87.2%) calls; a small subset of five molecules carried four HIGH calls each, concentrating more than half of all HIGH predictions [r2]. Stress response dominated: SR_MMP was HIGH in 6 molecules (29% of the set), and eye and respiratory irritation were frequent co-annotations; in contrast, no metabolite reached HIGH for nuclear receptor endpoints, though M13 showed the highest endocrine signal (NR_ER 0.6271; NR_ER_LBD 0.5587) [r2]. Across prioritizations, M13 (coupling/dimerization), M04 (debromination), M12 (methylation), M17 (debromination with vinyl), and M20 (debromination with α-tertiary alcohol) each had four HIGH calls; M09 (sulfation) contributed elevated endocrine activity but weaker descriptor-explanation coherence, marking it as a mechanistic outlier in this dataset of model predictions [r2, r4]. These distributions are model-derived and should be interpreted as hypotheses requiring validation.

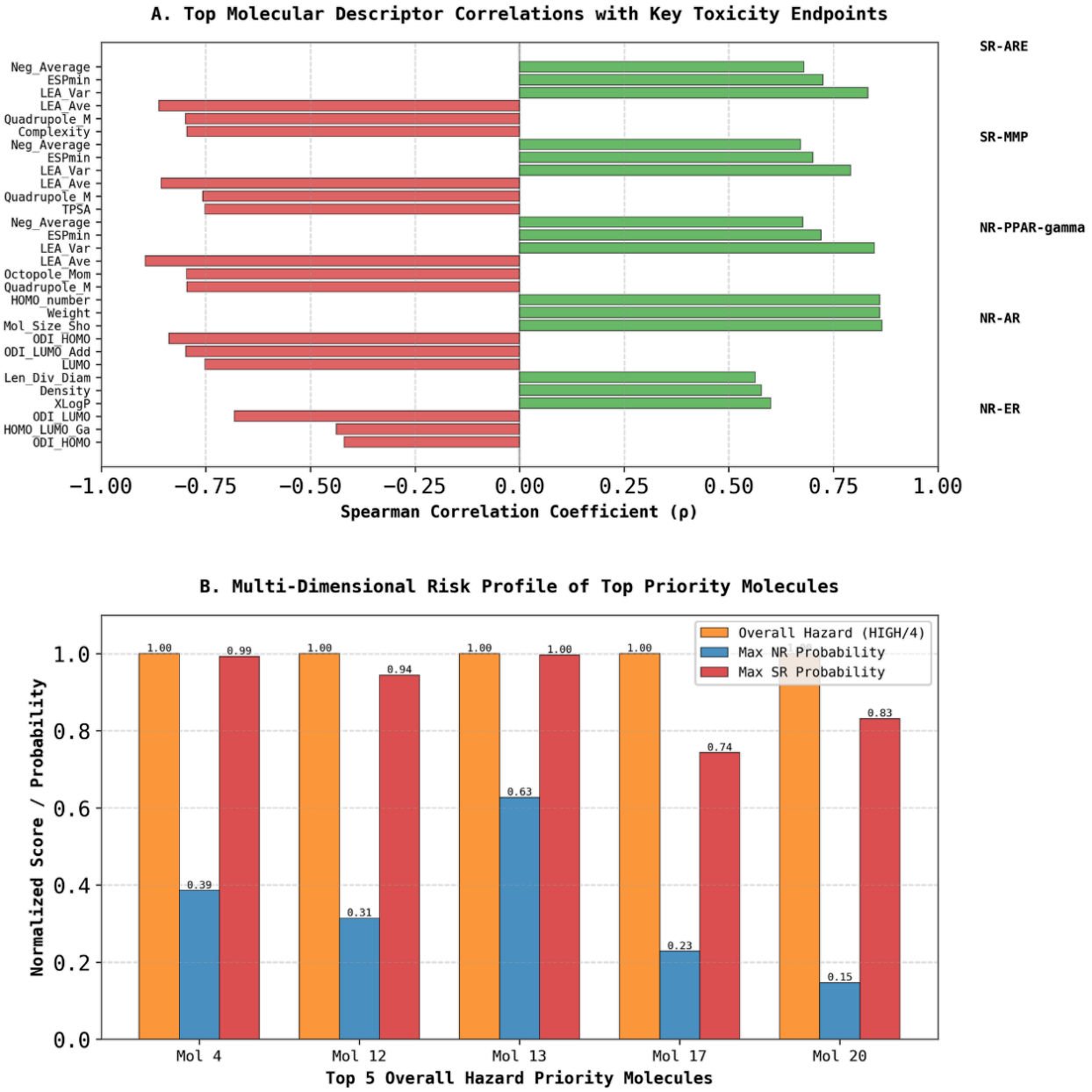

Figure 7. Model-predicted toxicity for priority TBBPA metabolites is dominated by stress response signals that correlate with molecular surface electronic properties. (A) Spearman correlation coefficients between top molecular descriptors and model-predicted toxicity endpoints for stress response (SR) and nuclear receptor (NR) pathways. (B) Normalized risk profiles for five priority metabolites, displaying overall hazard alongside maximum predicted probabilities for NR and SR pathways. These data suggest that high stress-response probabilities, which strongly correlate with surface electron-affinity descriptors, drive the overall high-hazard classification for these select metabolites.

SHAP-consistent correlation structure indicated that surface electronic properties dominate the model's stress-response calls. LEA_Var (surface electron-affinity variance) correlated positively with SR_MMP ($\rho = 0.791$, $p < 2 \times 10\text{-}5$) and SR_ARE ($\rho = 0.832$, $p < 1 \times 10\text{-}4$), while LEA_Ave

correlated negatively with SR_MMP (ρ = -0.857, p < 1 × 10-6) and SR_ARE (ρ = -0.862, p < 1 × 10-4), a pattern that also extended to NR_PPARγ (LEA_Var ρ = 0.848; LEA_Ave ρ = -0.895) despite no HIGH classifications for that endpoint [r2, r64]. TPSA showed strong negative correlations with SR_MMP (ρ = -0.752), eye irritation (ρ = -0.866), eye corrosion (ρ = -0.813), and CYP1A2 inhibition (ρ = -0.816), whereas LEA_Var showed complementary positive correlations to the same phenotypes (ρ = 0.868 for eye irritation; ρ = 0.812 for CYP1A2) (all p < 1 × 10-4) [r4]. For endocrine endpoints, NR_ER associated with lipophilicity (XLogP ρ = 0.600, p = 4 × 10-3) and density (ρ = 0.578), while NR_AR tracked with molecular size (e.g., Mol_Size_Short ρ = 0.866; Weight ρ = 0.842; both p < 1 × 10-4) and showed negative associations to frontier-orbital distribution indices (ODI_HOMO ρ = -0.838) [r2]. A mediation analysis further supported a causal chain in the model representation: the presence of phenolic OH groups increased stress scores primarily via their impact on LEA_Var and LEA_Ave (total ACME = 0.230, 95% CI 0.071–0.509, p = 0.012), with the direct effect becoming negligible once these mediators were included (c′ = 0.0005, p = 0.89), consistent with electronic surface features mediating phenolic-driven stress signals in the model [r90]. These relationships are model associations and not toxicology laws; they require experimental testing.

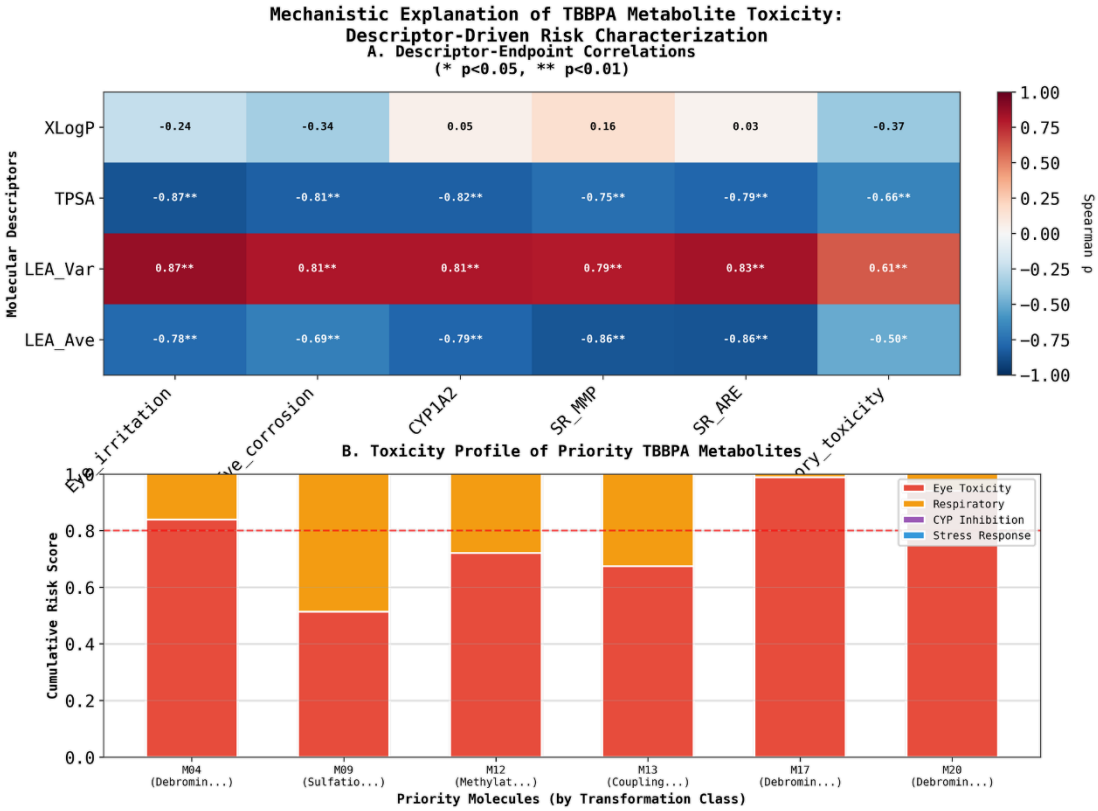

Figure 8. Surface electronic properties correlate with predicted toxicity and characterize high-risk TBBPA metabolites. (A) Heatmap of Spearman correlations between select molecular descriptors and model-predicted toxicity endpoints shows that surface electron-affinity variance (LEA_Var) correlates positively, while average surface electron-affinity (LEA_Ave) correlates negatively, with stress response endpoints. (B) Cumulative risk scores for six priority metabolites, with respiratory and eye toxicity driving the total risk score above the high-risk threshold for five of the six molecules. The figure links descriptor-level perturbations to the high predicted toxicity of specific metabolites.

Molecule-level narratives illustrate how transformation motifs map to descriptor shifts and endpoint signatures in the model. M13 (a coupling/dimerization product) exhibited four HIGH predictions, led by SR_MMP 0.996 and SR_ARE 0.810, alongside elevated eye and respiratory irritation; it displayed extreme nonpolar area and lipophilicity (XLogP 6.97, Z = +2.15), while retaining two phenolic OH groups and a rigid, expanded aromatic scaffold—features consistent with a "lipophilicity-assisted, phenol-retaining" stress signature in the model [r2, r24, r4]. M04 (a debrominated trihydroxy derivative) had four HIGH calls (SR_MMP 0.992, eye irritation 0.969, respiratory 0.937, CYP1A2 0.868) and showed low TPSA (43.3 Å², Z = -0.86) with increased surface electronic heterogeneity (LEA_Var Z = +0.93) and elevated HOMO (+1.71 SD), matching cohort-wide patterns that associate low polarity and heterogeneous electron-affinity surfaces with SR_MMP and ocular irritation in the model [r2, r28, r4]. These are model-predicted mechanisms;

experimental validation (e.g., JC-1/TMRM for ΔΨm, BCOP/ICE for ocular irritation, and CYP1A2 inhibition assays) is warranted to test them [r4].

Two substitutions on the bisphenolic bridge illustrate how local motifs may selectively tune endpoint profiles in the model. M12 (incomplete methylation) retained one phenolic OH and showed four HIGH calls (SR_MMP 0.944; respiratory 0.909; eye 0.905; CYP1A2 0.887) while exhibiting anomalously low nucleophilicity and electrostatic potential (e.g., Nu Z = -1.88; ESPmax Z = -1.69), supporting a hypothesis that incomplete masking preserves phenolic-driven electronic heterogeneity and SR_MMP risk in the model [r2, r29]. M20 (an α-tertiary alcohol debromination product) also had four HIGH calls including SR_MMP 0.831, but unlike lipophilic dimer M13, M20 showed decreased XLogP (-1.81 SD) and strong anomalies in orbital-density indices (ODI_HOMO +1.95 SD), and an internal comparator (M19) with identical formula but a β-primary alcohol exhibited a 3.16-fold lower SR_MMP probability (0.263), suggesting a position/substitution-specific electronic mechanism for mitochondrial disruption in the model rather than generic membrane accumulation [r32]. A related pair (M17 vinyl vs M19 hydroxymethyl) further showed selective increases for DNA-replication stress pathways (e.g., 23.9× higher SR_ATAD5; 8.3× higher SR_p53) accompanied by large decreases in multipole moments and TPSA and an increase in negative surface potential, while the HOMO–LUMO gap remained unchanged, indicating that localized electrostatics and shape, not global reactivity, may drive these model signals [r97, r2]. These interpretations are hypotheses derived from the model's descriptor–endpoint associations.

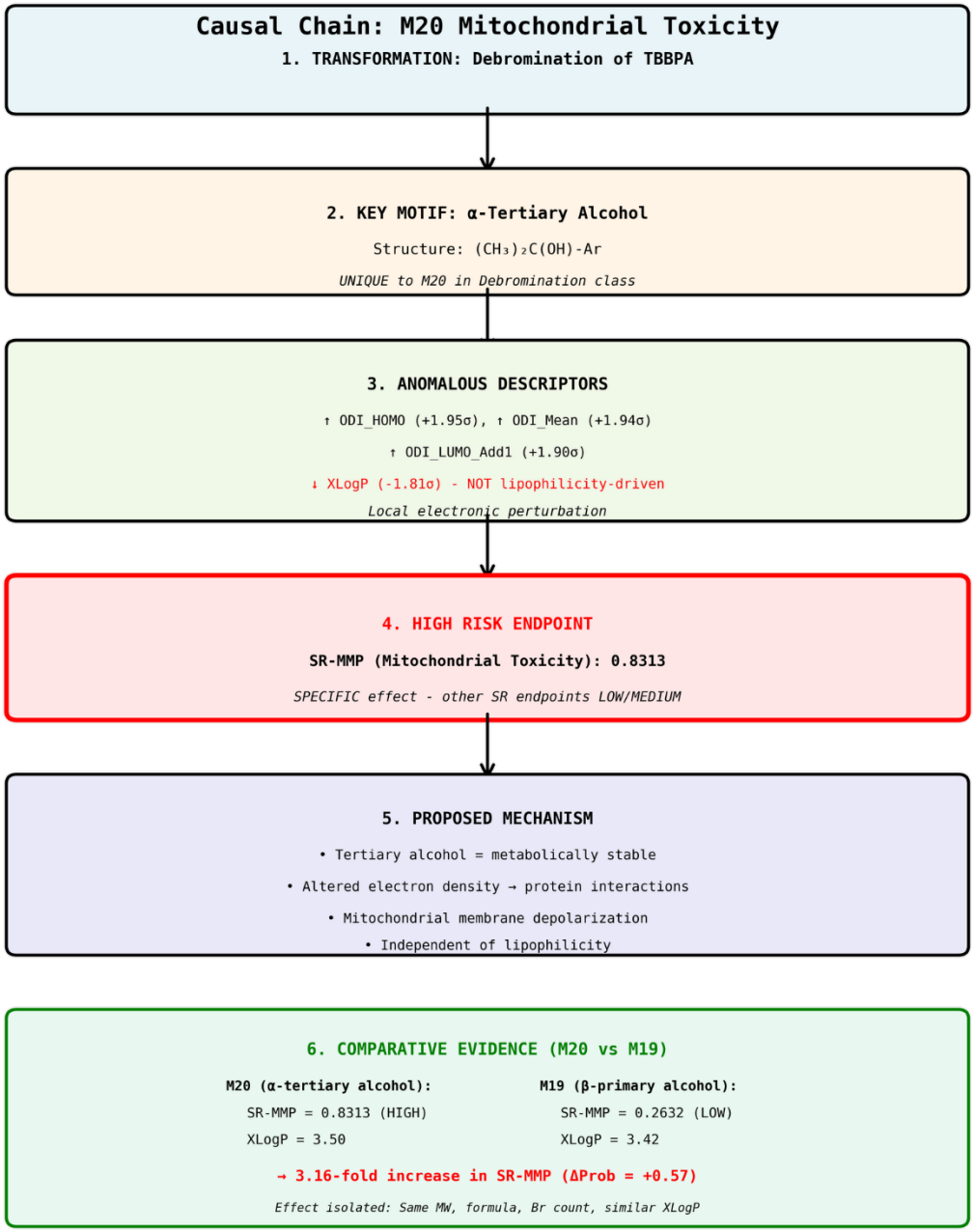

Figure 9. A proposed causal chain linking the α-tertiary alcohol motif of metabolite M20 to high predicted mitochondrial toxicity. The flowchart connects this unique structural feature to shifts in electronic property descriptors, culminating in a high-risk prediction for the mitochondrial stress response (SR-MMP). A comparison with the isomeric primary alcohol M19, which has a low predicted risk and similar lipophilicity, highlights the effect of the alcohol position. The analysis suggests the α-tertiary alcohol is a key structural determinant of M20's predicted toxicity, proposing a testable, electronically-mediated mechanism independent of lipophilicity.

At the transformation-class level, debromination was associated with higher SR_MMP probabilities (mean 0.500 vs 0.208 for fully brominated; Mann–Whitney $p = 0.026$) and

characteristic descriptor shifts—higher LEA_Var (+9.8%, p = 0.031) and lower molecular complexity (-45.8%, p = 0.004)—aligning with the model's electronic heterogeneity pharmacophore; notably, XLogP did not correlate with SR_MMP (ρ = 0.16, p = 0.48), indicating that membrane partitioning is not the dominant driver in these predictions [r64]. Coupling/dimerization increased lipophilicity and nonpolar area, often co-occurring with retained phenols and elevated SR_MMP/SR_ARE signals (exemplified by M13), while methylation effects depended on completeness of masking (M12 as a bioactivation hypothesis). Sulfation (M09) exhibited weaker descriptor–endpoint coherence, suggesting either alternate mechanisms or potential model limits for ionized conjugates in this chemical space; these class-level rules, the shared SR_MMP/PPARγ correlation pattern (with different thresholds), and the small n emphasize the need for targeted validation (e.g., ΔΨm assays, ARE-reporter, NR reporters), and careful scrutiny of applicability domain for highly polar conjugates and very large dimers [r4, r64]. All statements herein are model-based hypotheses meant to guide experiments, not definitive toxicological conclusions.

2.4 Discovery 4: Model credibility, applicability domain, and uncertainty quantification for TBBPA metabolite toxicity predictions

The dataset comprises two perfectly aligned tables covering 21 brominated compounds with 52 molecular descriptors and 89 toxicity outputs (141 merged columns), with no missing values, enabling unified descriptor–toxicity analyses [r0]. However, the data are small (n=21) and highly imbalanced: several endpoints show extreme or no variation (for example, 100% respiratory toxicity positivity and 52.4% eye irritation positivity), and acute toxicity variables are strongly right-skewed (skewness 2.77–3.02), all of which limit discriminative modeling and necessitate cautious interpretation [r0]. This context frames the downstream findings on applicability domain, nuclear receptor logic, and uncertainty quantification.

Applicability-domain checks against literature-based Tox21/ToxCast 95th percentile bounds for MW (100–600 Da), XLogP (−2 to 6), and TPSA (0–140 Å²) flagged 7 of 21 metabolites (33.3%) as out-of-domain (OOD), with coupling/dimerization products 4 of 4 (100%) OOD and significantly enriched versus all other classes (Fisher exact p = 0.0058) [r8]. Violations were dominated by excessive molecular weight (71.4% of OOD; e.g., M05 MW 696.21 Da [+96.21]) and high TPSA (42.9%; e.g., M05 TPSA 145.98 Å² [+5.98]), with less frequent XLogP excess (28.6%; e.g., M13 XLogP 6.97 [+0.97]) [r8]. While "polar conjugate" classes collectively contained 85.7% of OOD molecules, this broader grouping was not statistically enriched (p = 0.3371), emphasizing that dimerization—rather than conjugation per se—was the primary systematic driver of domain violations in this set [r8]. These OOD flags identify predictions likely to be less reliable due to extrapolation beyond the training chemical space.

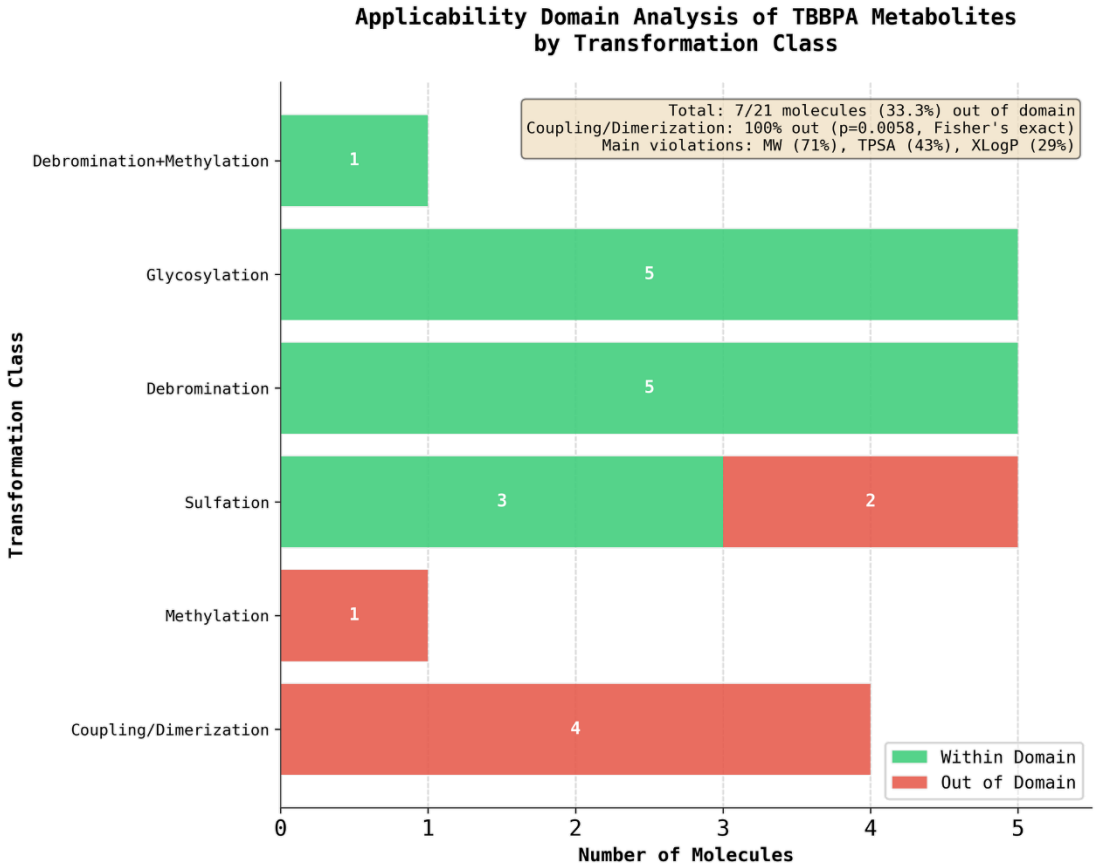

Figure 10. One-third of TBBPA metabolites fall outside a standard applicability domain, with coupling and dimerization products being exclusively flagged. The stacked bar chart shows the number of metabolites per transformation class categorized as within (green) or out of (red) a Tox21-like chemical space. This finding highlights that toxicity predictions for the larger coupling/dimerization products are likely unreliable and require cautious interpretation or further experimental validation.

The nuclear receptor analyses reveal two key patterns. First, all 21 metabolites were predicted LOW for NR-PPARγ (0 HIGH), yet the probability variance was highly structured: 35 of 51 descriptors (68.6%) significantly correlated with NR-PPARγ probability, with top associations indicating that smaller molecules with higher electronic heterogeneity (LEA_Var ρ = +0.848, p = 1.20×10-6; LEA_Ave ρ = −0.895, p = 4.43×10-8) and weaker multipole moments score higher, whereas all size descriptors correlate negatively (ρ ≈ −0.74 to −0.79) [r63]. This learned profile matches medicinal chemistry expectations for PPARγ and explains uniformly low probabilities for the large, brominated TBBPA metabolites without requiring a blanket assumption of model failure [r63]. Second, the model distinguishes binding from activation for steroid receptors. For NR-AR, electronic heterogeneity aligns positively with binding (LEA_Var ρ = +0.41, p = 0.068) but negatively with activation (ρ = −0.74, p < 0.001; Steiger's Z = −8.94, p < 0.0001), and size/complexity descriptors flip sign (strongly positive for activation, weakly negative for binding) [r73]. A similar binding-versus-activation differentiation holds for NR-ER with receptor-specific magnitudes (32 of 51 descriptors display opposite directions; ER-LBD vs ER correlation ρ =

0.7299; differentiation patterns across ER and AR correlated at ρ = 0.8797, p < 0.001) [r89]. Notably, the AR activation rule—favoring large, complex, electronically homogeneous molecules—conflicts with established AR agonist pharmacology and is best treated as a dataset-specific artifact rather than generalized biology, with dimer predictions likely false positives for agonism (Réau et al., 2019) (Zhu et al., 2012) (Kumar et al., 2025).

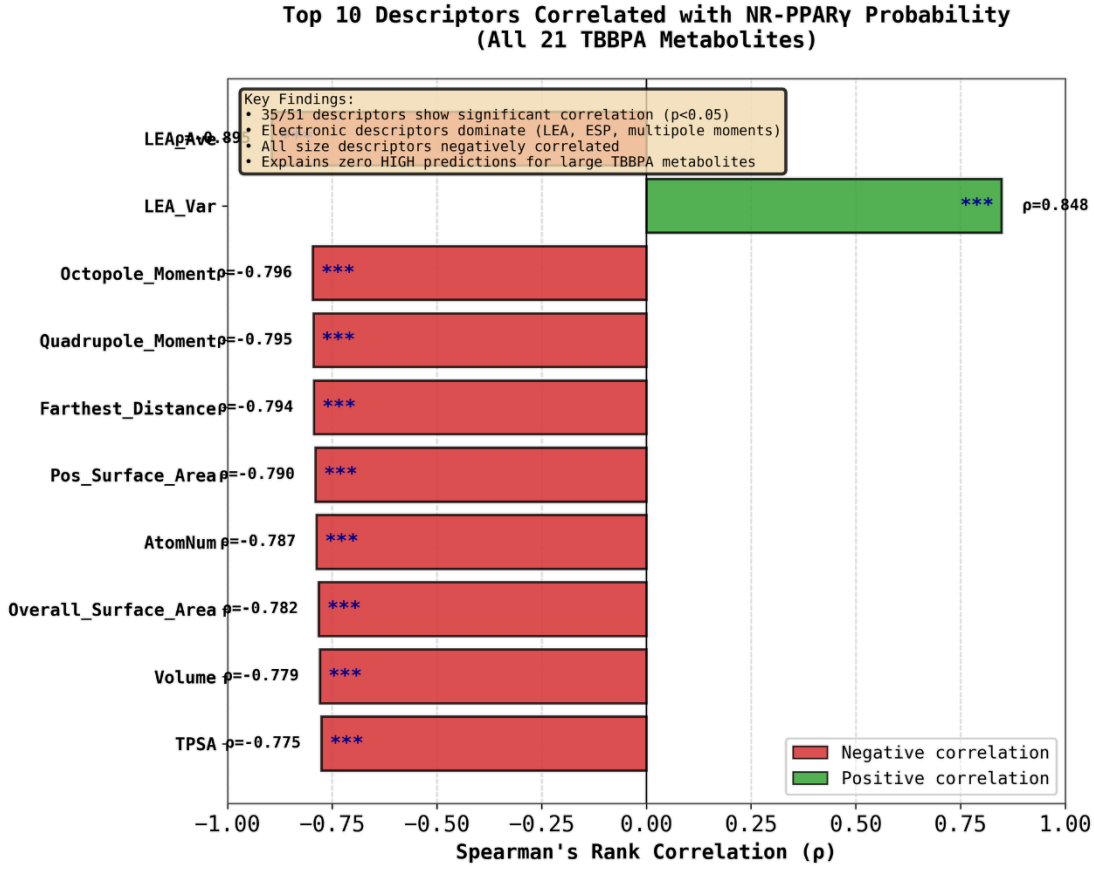

Figure 11. Predicted NR-PPARγ activation probability is strongly correlated with electronic and size-related molecular descriptors. The plot displays the top 10 descriptors ranked by their Spearman's rank correlation (ρ) with predicted PPARγ activation probability for 21 TBBPA metabolites. The consistent negative correlation with size and multipole moment descriptors suggests that larger, more electronically diffuse metabolites are less likely to activate PPARγ, providing a physically interpretable basis for the model's predictions.

Simple, physically grounded rules performed strongly. A single-threshold classifier using TPSA > 114.23 Å² to indicate "detoxified" (loss/masking of phenolic OH) achieved 95.24% accuracy (20/21) on the full set (precision 1.000; recall 0.833; F1 0.909), and 81.0% leave-one-out cross-validation (LOOCV) accuracy (17/21), with large between-group separation (Mann–Whitney U p = 0.0001; Cohen's d = 2.61) [r77]. The full-data threshold misclassified only M02 (TPSA 93.58 Å²), and LOOCV errors concentrated near boundary cases (M10, M14, M07, M02), while an electronic descriptor baseline previously reached 90.5% LOOCV, suggesting that electronic features can modestly outperform TPSA alone [r77]. In contrast, a three-feature logistic ensemble

(Detoxification Index = HOMO_LUMO_Gap/ALIE_Ave, TPSA, XLogP) failed to improve over TPSA alone (both 85.71% LOOCV), with a near-zero weight on the electronic index (coefficient 0.0084), underscoring that more complex linear combinations did not add value on this small, imbalanced set [r87].

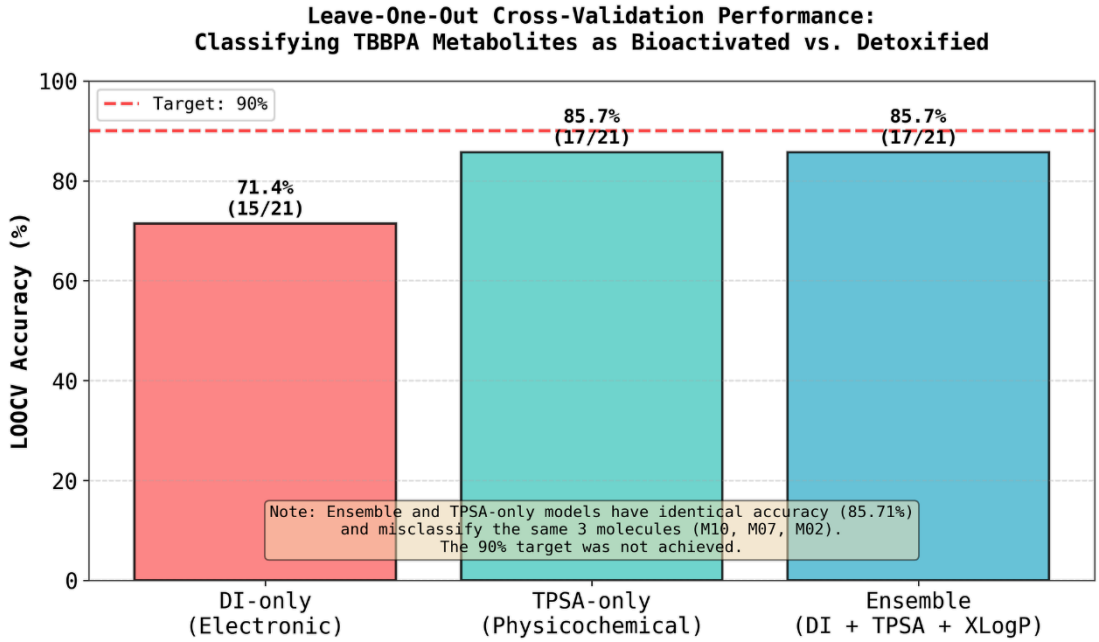

Figure 12. A simple physicochemical model based on TPSA alone achieves the highest accuracy for classifying TBBPA metabolite toxicity. The bar plot shows the leave-one-out cross-validation (LOOCV) accuracy for three models classifying 21 metabolites as bioactivated or detoxified: a DI-only electronic model, a TPSA-only model, and a multi-parameter ensemble. The single-parameter TPSA-only model's performance (85.7% accuracy) matches that of the more complex ensemble model, suggesting that additional descriptors provide no benefit for this classification task.

Two complementary uncertainty metrics prioritized molecules for verification. A Prediction Reliability Score combining applicability-domain violations (weight 2), deconjugation risk (log10-scaled), and classification ambiguity was highest for M06 (4.521), M07 (4.085), M05 (2.404), M02 (2.401), and M11 (2.398) and showed strong negative correlations with predicted stress toxicity ($\rho = -0.740$, $p = 0.0001$) and maximum endpoint probability ($\rho = -0.753$, $p = 0.0001$), indicating it mainly captures prediction confidence rather than being orthogonal to predicted hazard [r101]. To isolate model reliability, an Epistemic Uncertainty Score excluding ambiguity (Domain_Penalty 2 if OOD plus log10-scaled deconjugation risk) re-ranked the top cases toward extrapolation concerns (M06 3.839, M07 3.679, then a five-way tie at 2.000 for M03, M12, M13, M11, M05), with only moderate negative correlation to stress toxicity ($\rho = -0.592$, $p = 0.0047$) and substantial but incomplete rank agreement with the previous metric ($\rho = 0.938$, $p < 0.0001$) [r103]. Together with the applicability-domain analysis, these scores highlight dimeric and highly conjugated metabolites as the hardest to trust without targeted experimentation, while reinforcing

that several high-hazard predictions are also high-confidence within-domain signals [r8, r101, r103].

3. Conclusions

Across 21 TBBPA transformation products, the multi-task ToxD4C model predicts a tightly coordinated cellular stress state across oxidative, mitochondrial, heat-shock, and DNA-damage pathways. Predicted stress-response activities for SR-ARE, SR-MMP, SR-HSE, SR-p53, and SR-ATAD5 are uniformly and strongly correlated (mean Spearman $\rho = 0.93$, all $p < 0.001$), supporting a model-derived hypothesis of a common upstream driver rather than independent pathway triggers.

Within this chemical space, phenolic hydroxyl groups emerge as the dominant structural correlate of the coordinated stress signature. Metabolites with a phenolic OH exhibit markedly higher predicted activation probabilities across all five stress endpoints (20.7×–97.0× higher mean probabilities, all Bonferroni-significant at $\alpha = 0.01$), and a concordant dose-direction trend suggests higher stress probabilities for molecules with two phenolic OH groups relative to one phenolic OH.

Mechanistically within the model representation, the phenolic OH–stress association is statistically mediated almost entirely by a unifying electronic surface signature characterized by increased surface electron-affinity variance and decreased average local electron affinity (LEA_Var and LEA_Ave; total ACME = 0.230, 95% CI: 0.071–0.509, $p = 0.012$), with the direct effect becoming negligible when mediators are included ($c' = 0.0005$, $p = 0.89$). Descriptor-level analyses further indicate that the mitochondrial toxicity signal (SR-MMP) is not driven by simple hydrophobicity (XLogP is not significant: $\rho = +0.16$, $p = 0.48$) but instead aligns with surface electronic and electrostatic features (e.g., LEA_Var, LEA_Ave, ESPmin).

Class-level mapping of 21 metabolites into six transformation categories (Coupling/Dimerization, Debromination, Debromination+Methylation, Glycosylation, Methylation, Sulfation) supports a coherent model-derived detoxification/bioactivation pattern: conjugations that mask all phenolic hydroxyl groups are associated with uniformly low predicted stress-response activity, whereas pathways that preserve at least one phenolic hydroxyl group are associated with elevated multi-endpoint stress signals. Within this logic, glycosylation consistently appears detoxifying, sulfation is bimodal depending on whether the second phenolic site is also conjugated (dual-conjugation versus mono-conjugation), and debromination or incomplete methylation tend to bioactivate; these are explicitly model-derived hypotheses requiring targeted experimental validation.

Molecule-level synthesis concentrates prioritization on a small subset of metabolites (M04, M09, M12, M13, M17, M20), which concentrate HIGH-risk calls and share extreme descriptor perturbations—especially in surface electron-affinity metrics—that coherently explain the dominant stress-response and irritation signals in this dataset of model predictions. While nuclear receptor endpoints did not reach HIGH in this set, structured probability variance (e.g., for NR-

PPARγ) and binding-versus-activation separation patterns for steroid receptors provide additional mechanistic hypotheses for follow-up.

Model credibility considerations delimit where predictions should be treated with greatest caution and where experimental validation is most urgent. One-third of metabolites (7/21) fall outside a Tox21-like applicability domain (MW 100–600 Da, XLogP −2 to 6, TPSA 0–140 Å²), with Coupling/Dimerization products 100% out-of-domain, and conjugate stability (including in vitro deconjugation risk) can confound stress-response readouts. On this small, imbalanced dataset, simple physically interpretable rules (e.g., a TPSA threshold) can match or outperform multi-parameter models, and complementary uncertainty metrics highlight highly conjugated and out-of-domain metabolites as priority targets for stability controls and orthogonal experimental confirmation.

4. References

Kumar, A., Qian, M., Xu, Y., Benz, A., Covey, D.F., Zorumski, C.F., Mennerick, S., 2025. Multifaceted Actions of Neurosteroids. bioRxiv 2025–01.

Réau, M., Lagarde, N., Zagury, J.-F., Montes, M., Réau, M., Lagarde, N., Zagury, J.-F., Montes, M., 2019. Hits Discovery on the Androgen Receptor: In Silico Approaches to Identify Agonist Compounds. Cells 8. https://doi.org/10.3390/cells8111431

Stachulski, A.V., Yates, E.A., Teriosina, A., Hoyles, L., McArthur, S., 2025. Dietary substances and their glucuronides: structures, occurrence and biological activity. Natural Product Reports. https://doi.org/10.1039/D5NP00002E

Zhu, Z., Shi, M., Hu, W., Estrella, H., Engebretsen, J., Nichols, T., Briere, D., Hosea, N., Los, G., Rejto, P.A., Fanjul, A., 2012. Dose-dependent effects of small-molecule antagonists on the genomic landscape of androgen receptor binding. BMC Genomics 13, 355. https://doi.org/10.1186/1471-2164-13-355