**Conclusion**

One-third (7/21, 33.3%) of TBBPA metabolites fall outside the Tox21 95th percentile applicability domain, with Coupling/Dimerization class showing 100% out-of-domain rate (p=0.0058), driven primarily by excessive molecular weight (71% of violations) and TPSA (43% of violations).

**Methods**

The analysis used Python (pandas, numpy, scipy.stats, matplotlib) to systematically assess applicability domain compliance. First, the molecular descriptors dataset (prediction_results_with_smiles.csv) was merged with transformation class assignments (tbbpa_transformation_classes.csv) using SMILES as the key. Tox21 reference ranges representing 95th percentile boundaries were defined based on standard ToxCast/Tox21 library characterization: MW (100-600 Da), XLogP (-2 to 6), and TPSA (0-140 Ų). For each of the 21 molecules, descriptor values (Weight, XLogP, TPSA) were compared against these thresholds using a custom function that flagged violations and calculated deviation magnitudes. Molecules exceeding any threshold in either direction were classified as out-of-domain. Class-level statistics were computed by grouping molecules by transformation class and calculating out-of-domain frequencies. Statistical significance was assessed using Fisher's exact test (appropriate for small sample sizes) comparing: (1) target classes (Glycosylation, Sulfation, Coupling/Dimerization) versus others, and (2) Coupling/Dimerization specifically versus all other classes. A comprehensive report table was generated containing SMILES, Molecule_ID, Transformation_Class, descriptor values, out-of-domain flag, and specific violation reasons with deviation magnitudes. Results were visualized in a horizontal stacked bar chart showing within-domain versus out-of-domain molecules by transformation class.

**Results**

Of 21 TBBPA metabolites, 7 (33.3%) were classified as outside the Tox21 applicability domain. Out-of-domain molecules by class: Coupling/Dimerization 4/4 (100%), Methylation 1/1 (100%), Sulfation 2/5 (40%), Debromination 0/5 (0%), Glycosylation 0/5 (0%), Debromination+Methylation 0/1 (0%).

Descriptor-specific violations: MW exceeded upper limit in 5/7 molecules (71.4%), TPSA exceeded in 3/7 (42.9%), XLogP exceeded in 2/7 (28.6%). Three molecules (42.9%) had multiple violations. Specific out-of-domain molecules: M03 (MW=690.89 Da, +90.89; TPSA=140.33 Ų, +0.33), M05 (MW=696.21 Da, +96.21; TPSA=145.98 Ų, +5.98), M06 (MW=637.30 Da, +37.30), M07 (MW=618.56 Da, +18.56), M11 (MW=683.98 Da, +83.98; TPSA=141.62 Ų, +1.62), M12 (XLogP=6.09, +0.09), M13 (XLogP=6.97, +0.97).

Statistical testing: Coupling/Dimerization class showed significantly higher out-of-domain rate versus all other classes (Fisher's exact test: odds ratio = infinity, p=0.0058). Combined polar conjugate classes (Glycosylation, Sulfation, Coupling/Dimerization) contained 6/7 (85.7%) of all out-of-domain molecules but did not show statistically significant enrichment versus other classes (Fisher's exact: OR=4.50, p=0.3371).

Class-specific descriptor profiles for hypothesized polar conjugates: Glycosylation (n=5, 0% out, mean MW=576.7 Da, mean TPSA=96.2 Ų), Sulfation (n=5, 40% out, mean MW=555.8 Da, mean TPSA=88.2 Ų), Coupling/Dimerization (n=4, 100% out, mean MW=655.0 Da, mean TPSA=120.0 Ų).

**Challenges**

The primary challenge was the absence of explicitly provided Tox21 reference ranges in the task

prompt. I applied literature-based values for Tox21/ToxCast chemical space (MW: 100-600 Da, XLogP: -2 to 6, TPSA: 0-140 Ų) representing 95th percentile boundaries, which is a discretionary decision that affects specific out-of-domain classifications. Small sample sizes (n=21 total, some classes n=1) limit statistical power for inter-class comparisons. The hypothesis initially stated "95th percentile" without clarifying whether this meant upper boundary only or both boundaries; I interpreted this as the typical applicability domain definition (molecules outside the central 95% range). Fisher's exact test returned infinite odds ratio for Coupling/Dimerization analysis due to zero count in one cell (4 out, 0 within domain), though p-value remains valid and highly significant. The Glycosylation class paradoxically showed 0% out-of-domain despite being hypothesized as "large polar conjugates," suggesting size/polarity thresholds distinguish this class from Coupling/Dimerization.

**Discussion**

The results partially support the research hypothesis but reveal important nuances. While a significant fraction (33.3%) of metabolites fall outside Tox21 applicability domain, the pattern differs from initial predictions. Coupling/Dimerization metabolites universally exceed domain boundaries (100%, highly significant p=0.0058), consistent with their formation of large dimeric structures. However, Glycosylation metabolites surprisingly remain entirely within domain despite being "large polar conjugates," suggesting the glycoside moiety addition does not push molecules beyond Tox21 space boundaries. This may reflect that glycosylation increases both MW and TPSA but not to the extent that dimerization does. Sulfation shows intermediate behavior (40% out-of-domain), with only the heaviest members (M06, M07) exceeding MW thresholds.

Molecular weight is the dominant driver of out-of-domain classification (71% of violations), consistent with Coupling/Dimerization creating molecules that are literally twice the size of monomers. TPSA violations (43%) occur primarily in coupling products where dual phenolic structures create extensive polar surface area. XLogP violations (29%) are less common and appear in methylated/coupling products where extended hydrophobic regions push lipophilicity above Tox21 ranges.

The finding that 85.7% of out-of-domain molecules come from the three hypothesized polar conjugate classes supports the general principle but the lack of statistical significance (p=0.34) may reflect the small Debromination+Methylation and Methylation class sizes. The highly significant result for Coupling/Dimerization specifically (p=0.0058) provides strong evidence that this transformation class systematically produces metabolites outside the training domain of Tox21-based toxicity models.

These findings have important implications for toxicity prediction reliability. The 7 out-of-domain molecules (M03, M05, M06, M07, M11, M12, M13) should be flagged as having uncertain toxicity predictions, as model performance typically degrades when applied to molecules with physicochemical properties outside training data distributions. Risk assessors should exercise caution when interpreting toxicity predictions for these compounds, particularly the coupling/dimerization products.

**Proposed Next Hypotheses**

1. Predicted toxicity endpoints for out-of-domain molecules (particularly Coupling/Dimerization products) will show higher model uncertainty or conflicting predictions across different toxicity assays compared to within-domain metabolites.
2. The electronic descriptors (HOMO, LUMO, HOMO-LUMO gap) will show distinct

patterns between within-domain and out-of-domain molecules, potentially revealing whether electronic properties contribute to applicability domain violations beyond simple physicochemical parameters.

**Artifacts**

**Artifact 1:**

**File name:** tbbpa_applicability_domain_report.csv

**Artifact description:** Comprehensive applicability domain report for 21 TBBPA metabolites containing SMILES, Molecule_ID, Transformation_Class, physicochemical descriptors (Weight, XLogP, TPSA), binary out-of-domain classification flag, and detailed violation reasons with deviation magnitudes from Tox21 reference ranges. Created by merging molecular descriptors with transformation classes and systematically comparing each molecule's MW, XLogP, and TPSA values against Tox21 95th percentile boundaries (MW: 100-600 Da, XLogP: -2 to 6, TPSA: 0-140 Ų), flagging violations and calculating exceedance amounts.