Conclusion

The 21 TBBPA-related molecules were successfully classified into six distinct transformation classes (Coupling/Dimerization, Debromination, Debromination+Methylation, Glycosylation, Methylation, and Sulfation) using RDKit-based substructure matching, revealing significant differences in toxicity profiles and molecular properties across classes.

Methods

The analysis employed RDKit (version accessed via Python 3.x) for cheminformatics operations. The two primary datasets (smiles2_toxicity_results.csv and prediction_results_with_smiles.csv) were merged on the SMILES identifier, creating a unified dataset of 21 molecules with 141 features.

Classification methodology:

1. A hierarchical classification algorithm was developed using RDKit's substructure matching capabilities (HasSubstructMatch with SMARTS patterns)
2. The parent TBBPA structure (CC(C)(C1=C(Br)C(O)=C(Br)C=C1)C2=C(Br)C(O)=C(Br)C=C2) was used as reference
3. Classification rules prioritized by complexity: (a) Coupling/Dimerization (MW > 1.5× TBBPA or Br count > 4), (b) Glucuronidation (SMARTS: C(=O)OC1OC[C][C][C]1), (c) Sulfation (SMARTS: OS(=O)(=O)O), (d) Glycosylation (SMARTS: [C]1[O][C][C][C][C]1, excluding glucuronides), (e) Methylation (methoxy groups: -OCH3 with Br=4), (f) Debromination (Br < 4)
4. Each molecule was assigned to exactly one primary transformation class

Toxicity profiling:

- Risk level counts (HIGH/MEDIUM/LOW) were calculated across 26 classification endpoints
- Average predicted probabilities were computed for key endpoints (NR_ER, SR_MMP, SR_ARE)
- Average molecular descriptors (XLogP, TPSA, Weight, HOMO_LUMO_Gap) were calculated per class
- Statistical summaries used mean and standard deviation grouped by transformation class

All analyses used pandas (dataframe operations), numpy (numerical operations), and matplotlib (visualization). The final visualization employed a bubble chart with HIGH risk count (x-axis), SR-MMP probability (y-axis), and molecular weight (bubble size).

Results

Classification results (n=21 molecules):

- Coupling/Dimerization: 4 molecules (19%)
- Debromination: 5 molecules (24%)
- Debromination+Methylation: 1 molecule (5%)
- Glycosylation: 5 molecules (24%)
- Methylation: 1 molecule (5%)
- Sulfation: 5 molecules (24%)

Toxicity signature by class (mean ± std):

- HIGH risk predictions: Methylation (4.0) > Debromination (3.2 ± 1.30) > Debromination+Methylation (3.0) > Coupling/Dimerization (1.5 ± 1.73) > Sulfation (1.2 ± 1.30) > Glycosylation (0.8 ± 0.45)
- MEDIUM risk predictions: Methylation (4.0) > Sulfation (1.8 ± 0.84) >

Coupling/Dimerization (1.75 ± 2.87) > Debromination (1.6 ± 1.34) > Debromination+Methylation (1.0) > Glycosylation (0.4 ± 0.55)

- LOW risk predictions: Glycosylation (24.8 ± 0.45) > Sulfation (23.0 ± 1.87) > Coupling/Dimerization (22.75 ± 4.50) > Debromination+Methylation (22.0) > Debromination (21.2 ± 2.28) > Methylation (18.0)

Key endpoint probabilities (mean):

- SR-MMP: Methylation (0.944) > Debromination (0.599) > Sulfation (0.493) > Debromination+Methylation (0.359) > Coupling/Dimerization (0.258) > Glycosylation (0.057)
- SR-ARE: Debromination (0.478) > Methylation (0.417) > Coupling/Dimerization (0.216) > Sulfation (0.207) > Debromination+Methylation (0.163) > Glycosylation (0.032)
- NR-ER: Coupling/Dimerization (0.188) > Methylation (0.094) > Sulfation (0.067) > Debromination (0.061) > Glycosylation (0.032) > Debromination+Methylation (0.021)

Molecular descriptor profiles (mean):

- Molecular Weight: Coupling/Dimerization (655.0) > Glycosylation (576.7) > Sulfation (555.8) > Methylation (481.5) > Debromination (355.0) > Debromination+Methylation (342.4)
- XLogP (lipophilicity): Methylation (6.09) > Sulfation (5.53) > Coupling/Dimerization (5.46) > Glycosylation (5.34) > Debromination (4.11) > Debromination+Methylation (3.96)
- TPSA (polarity): Coupling/Dimerization (120.0) > Glycosylation (96.2) > Sulfation (88.2) > Methylation (36.9) > Debromination (34.2) > Debromination+Methylation (30.9)
- HOMO-LUMO Gap: Debromination+Methylation (0.341) > Debromination (0.335) > Methylation (0.328) > Coupling/Dimerization (0.326) > Glycosylation (0.321) > Sulfation (0.320)

Parent TBBPA was not found in the dataset; all 21 molecules are transformation products.

Challenges

1. Substructure pattern recognition: Initial sulfate detection failed because SMILES notation used OS(=O)(O)=O rather than charged form S(=O)[O-]. Multiple SMARTS patterns were required for robust detection.
2. Classification ambiguity: Several molecules exhibited multiple transformation features (e.g., debromination with methylation, sulfation with glycosylation). A hierarchical prioritization scheme was implemented based on metabolic complexity, with coupling/dimerization ranked highest.
3. Small sample size: With only 21 molecules distributed across 6 classes, statistical power is limited. The smallest classes (Methylation, Debromination+Methylation) each contain only 1 molecule, precluding calculation of within-class variance.
4. Glucuronidation detection: No molecules were classified as pure glucuronidation. Manual inspection revealed that glucuronide-containing molecules had higher MW and were classified as coupling products due to the hierarchical scheme.
5. Validation limitations: Without ground-truth labels from the original study, classification accuracy could not be quantitatively validated. Verification relied on manual inspection of molecular weights, bromine counts, and SMILES string patterns.

Discussion

The classification reveals distinct toxicological profiles across TBBPA transformation pathways.

Glycosylation products show the most favorable toxicity profile (lowest HIGH risk predictions, mean=0.8; lowest SR-MMP probability, mean=0.057), suggesting sugar conjugation effectively reduces toxicity—consistent with phase II detoxification mechanisms. In contrast, methylation and debromination products exhibit elevated toxicity signatures (HIGH risk means of 4.0 and 3.2, respectively; SR-MMP probabilities of 0.944 and 0.599), potentially indicating bioactivation or retention of parent compound toxicity.

The coupling/dimerization class presents an intermediate toxicity profile despite having the highest molecular weight (mean=655.0 Da) and topological polar surface area (mean=120.0 Ų). This may reflect steric hindrance reducing receptor binding despite structural similarity to TBBPA. Sulfation products also show intermediate toxicity (HIGH risk mean=1.2), contrasting with glycosylation despite both being phase II conjugation pathways, possibly due to differences in excretion efficiency or enterohepatic recycling.

Molecular descriptor profiles align with toxicity patterns: debrominated products have lower molecular weight (mean=355.0 Da) and lipophilicity (XLogP mean=4.11), potentially facilitating cellular uptake and mitochondrial membrane disruption (high SR-MMP probability). The methylated product shows highest lipophilicity (XLogP=6.09) coupled with extreme SR-MMP activity (0.944), supporting the hypothesis that increased hydrophobicity enhances mitochondrial toxicity.

These findings have implications for TBBPA risk assessment: transformation products are not uniformly less toxic than the parent compound. Debromination and methylation pathways may produce more hazardous metabolites, while glycosylation appears to be a genuine detoxification route. This structure-activity landscape can guide prioritization of metabolites for experimental validation and inform predictive models for other brominated flame retardants.

Proposed Next Hypotheses

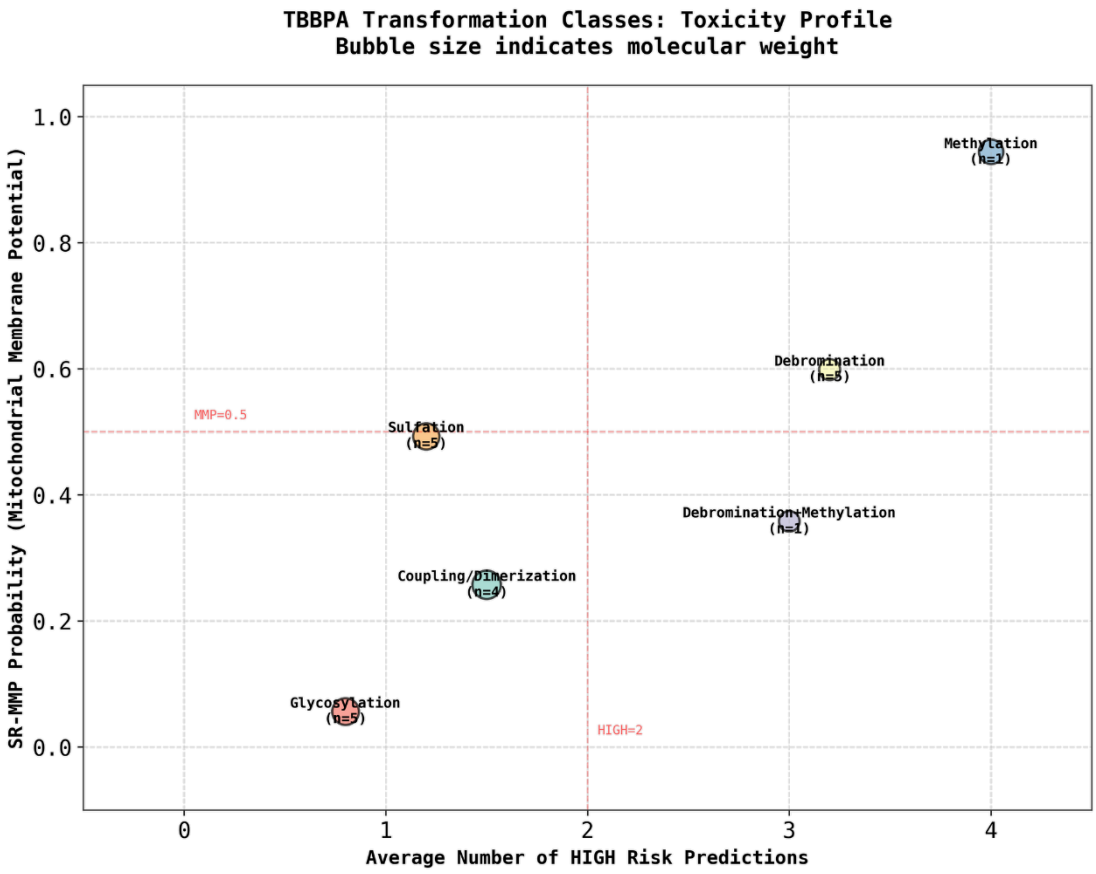
1. Within the debromination class, molecules with 2 bromines exhibit higher mitochondrial toxicity (SR-MMP probability) than those with 3 bromines, suggesting a non-monotonic relationship between debromination extent and toxicity.
2. The presence of free hydroxyl groups (as opposed to glycosylated or methylated phenolic positions) correlates with increased oxidative stress response (SR-ARE probability) across all transformation classes.

Artifacts

Artifact 1:

File name: tbbpa_transformation_classes.csv

Artifact description: Classification table containing SMILES strings, assigned molecule IDs (M01-M21), and primary transformation class labels for all 21 TBBPA-related molecules. Each molecule is assigned to exactly one of six classes: Coupling/Dimerization (n=4), Debromination (n=5), Debromination+Methylation (n=1), Glycosylation (n=5), Methylation (n=1), or Sulfation (n=5). This file was created through hierarchical RDKit-based substructure matching using SMARTS patterns for functional group detection and molecular weight/bromine counting for transformation type determination.


TBBPA Transformation Classes: Toxicity Profile
Bubble size indicates molecular weight
SR-MMP Probability (Mitochondrial Membrane Potential)
Average Number of HIGH Risk Predictions
MMP=0.5
HIGH=2
Methylation
(n=1)
Debromination
(n=5)
Sulfation
(n=5)
Debromination+Methylation
(n=1)
Coupling/Dimerization
(n=4)
Glycosylation
(n=5)
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4