Conclusion

The ToxD4C model's NR-PPARγ predictions are not random; variance in predicted probabilities across 21 TBBPA metabolites shows highly significant correlations with 35 of 51 molecular descriptors (68.6%, all $p<0.05$), revealing that the model associates PPARγ activity with smaller molecules having heterogeneous but weaker electronic features—characteristics absent in the large, brominated TBBPA metabolites, explaining zero HIGH-risk predictions.

Methods

The analysis merged two CSV files (toxicity predictions and molecular descriptors) using SMILES identifiers to create a unified dataset of 21 TBBPA metabolites with 51 descriptors plus NR-PPARγ probability values. For each of the 51 molecular descriptors (electronic properties like HOMO/LUMO, DFT-based descriptors like local electron affinity and electrostatic potential, and physicochemical properties like molecular weight and TPSA), Spearman's rank correlation coefficient (ρ) was calculated against the continuous NR-PPARγ probability values using scipy.stats.spearmanr. Spearman correlation was chosen over Pearson due to the small sample size (n=21), potential non-normality, and the presence of outliers. All correlations were tested for statistical significance at $\alpha=0.05$. Results were sorted by absolute correlation magnitude to identify the top 3 positive and top 3 negative correlations. A final visualization was created showing the top 10 correlated descriptors with their ρ values, p-values, and significance indicators. The analysis used Python 3 with pandas (data manipulation), numpy (numerical operations), scipy.stats (correlation analysis), and matplotlib/seaborn (visualization).

Results

Dataset Characteristics:

- All 21 molecules classified as LOW risk for NR-PPARγ (0 HIGH predictions)
- NR-PPARγ probability range: 0.000528 to 0.228398 (mean: 0.025051 ± 0.049784)

Correlation Analysis:

- 35 of 51 descriptors (68.6%) showed statistically significant correlation ($p<0.05$)
- 25 descriptors showed highly significant correlation ($p<0.001$)

Top 3 Positively Correlated Descriptors:

1. **LEA_Var** (Local Electron Affinity Variance): $\rho = +0.848$, $p = 1.20\times10^{-6}$
2. **ESPmin** (Minimum Electrostatic Potential): $\rho = +0.721$, $p = 2.28\times10^{-4}$
3. **Neg_Average** (Average Negative ESP): $\rho = +0.677$, $p = 7.57\times10^{-4}$

Top 3 Negatively Correlated Descriptors:

1. **LEA_Ave** (Average Local Electron Affinity): $\rho = -0.895$, $p = 4.43\times10^{-8}$
2. **Octopole_Moment**: $\rho = -0.796$, $p = 1.59\times10^{-5}$
3. **Quadrupole_Moment**: $\rho = -0.795$, $p = 1.67\times10^{-5}$

Molecular Size Correlations: All size descriptors showed strong negative correlations (ρ ranging from -0.74 to -0.79, all $p<0.0001$): AtomNum (-0.787), Weight (-0.762), Volume (-0.779), TPSA (-0.775), Complexity (-0.769), Overall_Surface_Area (-0.782), Farthest_Distance (-0.794).

Highest vs Lowest Probability Molecules:

- Highest probability (0.228): Simple debrominated TBBPA (LEA_Ave=-2.434, LEA_Var=3.019)
- Lowest probability (0.000528): Large glucuronide conjugate (LEA_Ave=-2.085, LEA_Var=2.320)

Challenges

The primary challenge was the small sample size (n=21), which limited statistical power and required careful interpretation of effect sizes alongside p-values. While 35 correlations reached statistical significance, this should be interpreted cautiously given the 51 comparisons performed (though no multiple testing correction was applied per the exploratory nature of the analysis). The dataset description indicated 52 descriptors, but only 51 non-SMILES columns were present in the descriptor file. The predicted probabilities were highly right-skewed (mean=0.025, max=0.228), with most values clustered near zero, which could affect correlation robustness. However, Spearman correlation is rank-based and thus robust to this distribution. The analysis is purely correlational and cannot establish causation or definitively reveal the model's internal mechanisms, only identify associations between descriptor values and model outputs.

Discussion

The analysis successfully refutes the null hypothesis that the model's NR-PPARγ predictions are random. The discovery that 68.6% of descriptors show significant correlations provides compelling evidence of systematic decision-making by the ToxD4C model, despite its failure to predict any HIGH-risk molecules in this dataset.

Implicit Pharmacophore Revealed: The model associates higher PPARγ activity probability with:

1. **Electronic heterogeneity over uniformity**: Higher variance but lower average electron affinity (LEA_Var positive, LEA_Ave negative)
2. **Weaker electrostatic features**: Less negative ESP minimum and average (positive correlations)
3. **Simpler charge distributions**: Lower multipole moments (Octopole and Quadrupole negative)
4. **Smaller molecular size**: All size descriptors negatively correlated

This pharmacophore profile describes small molecules with diverse but moderate electronic properties—NOT large, complex molecules with strong electrostatic features. Known PPARγ agonists (thiazolidinediones, glitazones) are typically 250-450 Da with balanced hydrophobic/hydrophilic regions, while TBBPA metabolites range from 450-900+ Da and possess strong electron-withdrawing bromine atoms and bulky conjugates (glucuronides, sulfates).

Why Zero HIGH Predictions: The TBBPA metabolites fall systematically outside the model's learned PPARγ active chemical space. They are too large (high molecular weight, surface area, atom count), too electronically uniform (low LEA_Var relative to training data expectations), and possess too-strong electrostatic features (very negative ESP values). The model correctly identifies them as structurally dissimilar to its training set of PPARγ actives, resulting in uniformly low probabilities. This is actually appropriate model behavior—cautious predictions when compounds fall outside the applicability domain.

Literature Consistency: This finding aligns with known PPARγ structure-activity relationships. PPARγ ligand binding domain accommodates moderate-sized ligands with balanced electronic properties. Large, highly halogenated compounds like brominated flame retardants are not typical PPARγ modulators. The model's implicit learning of size constraints and electronic preferences mirrors established medicinal chemistry knowledge.

Model Limitation vs. Model Insight: Rather than a pure "blind spot," the zero HIGH predictions may reflect appropriate uncertainty handling. The strong correlations reveal the model IS extracting meaningful physicochemical information and making decisions based on learned patterns, but those patterns indicate these molecules are poor PPARγ candidates. This distinction is crucial for risk

assessment: low predictions for structurally dissimilar compounds are more reliable than extrapolated predictions.

Proposed Next Hypotheses

1. **Hypothesis: Electronic heterogeneity (LEA_Var) serves as a universal descriptor across multiple nuclear receptor endpoints** - If LEA_Var correlates with NR-PPARγ predictions, it may also correlate with predictions for other nuclear receptor endpoints (NR-ER, NR-AR, NR-Aromatase) in the dataset, suggesting the model uses electronic surface diversity as a general indicator of ligand-receptor binding promiscuity across the nuclear receptor superfamily.
2. **Hypothesis: The ratio of molecular size to electronic heterogeneity predicts applicability domain boundaries** - Creating a composite metric (e.g., AtomNum/LEA_Var or Volume/LEA_Var) will show that molecules with high size-to-heterogeneity ratios consistently receive lower probabilities across multiple toxicity endpoints, and this ratio will correlate with previously identified out-of-domain (OOD) flags, potentially serving as a simple applicability domain filter for the ToxD4C model.