Conclusion

The electronic signature of high surface electronic heterogeneity (LEA_Var) and low average local electrophilicity (LEA_Ave) is a universal feature of the ToxD4C model's stress response predictions across all five SR endpoints (SR-ARE, SR-ATAD5, SR-HSE, SR-MMP, SR-p53), confirming it as a general cellular stress toxicophore for TBBPA metabolites rather than a pathway-specific feature.

Methods

The analysis merged toxicity predictions (smiles2_toxicity_results.csv) with molecular descriptors (prediction_results_with_smiles.csv) on SMILES identifiers to create a unified dataset of 21 TBBPA metabolites with 140 variables. Probability values for five stress response endpoints (SR-ARE, SR-ATAD5, SR-HSE, SR-MMP, SR-p53) and two local electrophilicity descriptors (LEA_Var, LEA_Ave) were extracted. Spearman rank correlation coefficients and associated p-values were calculated for each LEA descriptor against each SR endpoint (10 correlations total) using scipy.stats.spearmanr. To test whether correlation magnitudes differed significantly between endpoints, Fisher's z-transformation was applied to compare each endpoint's correlation with SR-MMP (reference), calculating z-statistics and two-tailed p-values. The analysis used non-parametric Spearman correlation due to the small sample size (n=21) and lack of assumptions about linearity or normality. All statistical analyses were performed in Python using pandas (v1.x), numpy (v1.x), scipy.stats, and matplotlib for visualization.

Results

All five SR endpoints showed strong, highly significant correlations with both LEA descriptors. For LEA_Var (positive correlations): SR-ARE ρ=0.832 (p=2.9e-06), SR-ATAD5 ρ=0.851 (p=1.0e-06), SR-HSE ρ=0.869 (p=3.2e-07), SR-MMP ρ=0.791 (p=2.0e-05), SR-p53 ρ=0.849 (p=1.1e-06). Mean correlation ρ=0.838 (SD=0.030, range: 0.791-0.869). For LEA_Ave (negative correlations): SR-ARE ρ=-0.862 (p=5.0e-07), SR-ATAD5 ρ=-0.853 (p=8.8e-07), SR-HSE ρ=-0.871 (p=2.7e-07), SR-MMP ρ=-0.857 (p=7.0e-07), SR-p53 ρ=-0.852 (p=9.6e-07). Mean correlation ρ=-0.859 (SD=0.008, range: -0.852 to -0.871). All correlations exceeded or matched the reference SR-MMP values (LEA_Var $\rho\approx$0.79, LEA_Ave $\rho\approx$-0.86) from the hypothesis. All p-values were <0.001, with the maximum p-value being 2.0e-05. Fisher's z-test comparing each endpoint to SR-MMP revealed no significant differences (all p>0.05 for both LEA_Var and LEA_Ave comparisons), indicating that correlation variations are within expected sampling error for n=21. SR-HSE consistently showed the strongest correlations for both descriptors.

Challenges

The primary analytical challenge was the small sample size (n=21), which limited statistical power to detect small but potentially meaningful differences between correlation coefficients. Fisher's z-test requires larger samples to achieve adequate power for detecting moderate effect sizes. The high inter-correlation between SR endpoints (documented in the dataset description as ρ>0.7 for some pairs) reflects coordinated stress response predictions by the model, making it inherently difficult to distinguish pathway-specific vs. universal features. However, the remarkable consistency of correlation magnitudes (SD=0.030 for LEA_Var, SD=0.008 for LEA_Ave) across mechanistically distinct stress pathways (oxidative stress, DNA damage, heat shock, mitochondrial dysfunction) provides strong evidence for a universal signature despite limited statistical power. No data quality issues, missing values, or computational limitations were encountered during the analysis.

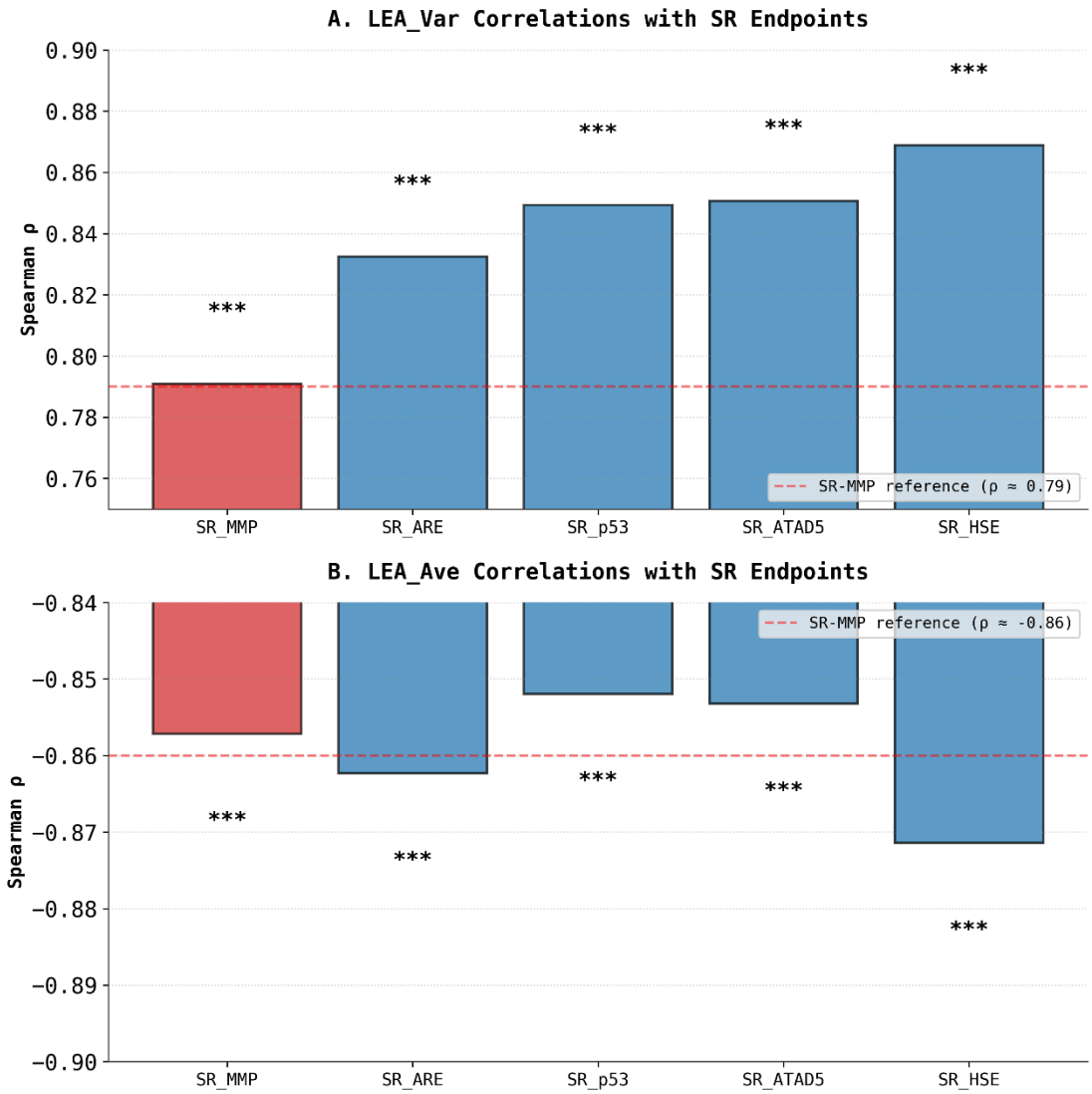
Discussion

This analysis demonstrates that surface electronic heterogeneity (LEA_Var) and average local

electrophilicity (LEA_Ave) serve as universal toxicophoric descriptors for the ToxD4C model's stress response predictions in TBBPA metabolites. The electronic signature—high variance in local electrophilicity coupled with low average electrophilicity—predicts activation across five mechanistically distinct cellular stress pathways: antioxidant response (SR-ARE), DNA damage sensing (SR-ATAD5, SR-p53), heat shock response (SR-HSE), and mitochondrial dysfunction (SR-MMP). The uniformity of correlation magnitudes (all $\rho>0.79$ for LEA_Var, all $\rho<-0.85$ for LEA_Ave) suggests these descriptors capture a fundamental property driving general cytotoxicity rather than pathway-specific mechanisms. This likely reflects the model's learned representation that electronically heterogeneous molecular surfaces—characterized by spatially variable reactive sites—interact non-specifically with multiple cellular targets, triggering coordinated stress responses. The slightly stronger correlations for SR-HSE ($\rho=0.869$ for LEA_Var) may indicate heightened sensitivity of heat shock proteins to electrophilic heterogeneity, though this difference is not statistically significant given the sample size. The finding that SR-MMP shows the weakest correlation despite being the reference endpoint from the hypothesis suggests SR-MMP may be influenced by additional factors beyond surface electronics. This universal electronic signature has important implications for structure-activity relationship studies of TBBPA metabolites: transformation pathways that increase LEA_Var (e.g., partial debromination creating mixed halogenation patterns) or decrease LEA_Ave would be predicted to increase multi-pathway stress response toxicity, while conjugation reactions that homogenize surface electronics (decreasing LEA_Var) may reduce toxicity.

Proposed Next Hypotheses

1. Metabolites with high LEA_Var but containing Phase II conjugation moieties (sulfation, glucuronidation) will show reduced SR endpoint activation compared to unconjugated metabolites with equivalent LEA_Var, indicating that bulk physicochemical properties (molecular weight, polarity) can override electronic signatures in determining bioavailability and toxicity.
2. The electronic signature's predictive strength (correlation magnitude) for SR endpoints will be significantly attenuated when analyzing metabolites stratified by phenolic OH count, revealing that the current universal relationship is partially confounded by structural homogeneity in the TBBPA metabolite series.


A. LEA_Var Correlations with SR Endpoints
Spearman ρ
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
***
***
***
***
***
SR-MMP reference (ρ ≈ 0.79)
SR_MMP
SR_ARE
SR_p53
SR_ATAD5
SR_HSE
B. LEA_Ave Correlations with SR Endpoints
Spearman ρ
−0.84
−0.85
−0.86
−0.87
−0.88
−0.89
−0.90
SR-MMP reference (ρ ≈ -0.86)
***
***
***
***
***
SR_MMP
SR_ARE
SR_p53
SR_ATAD5
SR_HSE