Conclusion

The Divergence Ratio (SR-ATAD5/SR-MMP) shows strongest associations with ODI_Std (orbital energy distribution variability, ρ=-0.48, p=0.071) and Overall_Average (electronic surface property, ρ=-0.48, p=0.069), both approaching marginal significance, while universal stress drivers LEA_Var and LEA_Ave rank poorly (18th and 42nd out of 51 descriptors), partially supporting the hypothesis that pathway-specific molecular features control the DNA/mitochondrial stress balance.

Methods

Analysis was performed using Python 3 with pandas, numpy, scipy.stats, RDKit, and matplotlib. The two primary CSV files (smiles2_toxicity_results.csv containing 21 molecules × 89 columns of toxicity predictions, and prediction_results_with_smiles.csv containing 21 molecules × 52 columns of molecular descriptors) were merged on the SMILES identifier. Phenolic hydroxyl groups were identified using RDKit SMARTS pattern matching ([OH]c), filtering to 15 molecules with at least one phenolic OH group. The Divergence Ratio was calculated as SR-ATAD5_Probability / (SR-MMP_Probability + 1×10^{-6}) for each molecule. Spearman rank correlations were computed between the Divergence Ratio and all 51 molecular descriptors using scipy.stats.spearmanr. Statistical significance was assessed at α=0.05 (two-tailed), with the critical correlation coefficient (|ρ|=0.514) calculated from the t-distribution with df=13. Results were ranked by absolute correlation strength and categorized by p-value thresholds ($p<0.05$ for statistical significance, $p<0.10$ for marginal significance). Descriptors were classified into functional categories (electronic, orbital, charge asymmetry, structural) for mechanistic interpretation.

Results

Of 15 phenolic TBBPA metabolites analyzed, the Divergence Ratio (SR-ATAD5/SR-MMP) ranged from 0.003 to 0.112 (mean=0.021, SD=0.026), with SR-ATAD5 probabilities ranging 0.0004-0.0689 and SR-MMP probabilities ranging 0.024-0.996. No descriptors achieved statistical significance ($p<0.05$) due to limited sample size (n=15, requiring $|\rho|\geq0.514$). Three descriptors showed marginal significance ($p<0.10$): Overall_Average (ρ=-0.482, p=0.069), ODI_Std (ρ=-0.479, p=0.071), and Nu (ρ=-0.461, p=0.084). The hypothesized ODI_Std metric ranked 2nd of 51 descriptors, partially confirming the prediction that orbital energy distribution variability influences stress pathway divergence. Quadrupole_Moment showed moderate negative correlation (ρ=-0.357, p=0.191, rank=6/51), consistent with the hypothesis. Universal stress drivers performed poorly: LEA_Var ranked 18th (ρ=0.296, p=0.283) and LEA_Ave ranked 42nd (ρ=-0.204, p=0.467), confirming they are distinct from pathway-specific drivers. Among all six ODI metrics tested, ODI_Std showed the strongest association, while other ODI metrics (ODI_HOMO_1, ODI_HOMO, ODI_LUMO, ODI_LUMO_Add1, ODI_Mean) ranked 13th-48th. The top observed correlation (|ρ|=0.482) reached 93.8% of the required effect size for significance. Negative correlations dominated the top 10 (9 of 10), indicating that higher values of electronic and structural descriptors generally shift the balance toward mitochondrial stress (SR-MMP) relative to DNA damage stress (SR-ATAD5).

Challenges

The primary analytical challenge was severe statistical power limitation due to the small sample size (n=15 phenolic metabolites). With df=13, the critical correlation coefficient for significance at α=0.05 was $|\rho|\geq0.514$, representing a large effect size threshold. The top observed correlation (|ρ|=0.482) fell just short of significance despite being substantial. This created ambiguity in hypothesis testing—the analysis could not definitively confirm statistically significant correlations,

but the marginally significant results ($p<0.10$) and high effect sizes (93.8% of required threshold) suggest real associations that are underpowered rather than absent. The filtering step (selecting only phenolic metabolites) reduced the already-small dataset from 21 to 15 molecules, exacerbating power issues. Additionally, the Divergence Ratio exhibited substantial range (0.003-0.112, ~37-fold variation), suggesting potential outlier influence on rank-based correlations. The hypothesis specified "statistically significant correlations ($p<0.05$)" but none were found; the analysis had to rely on marginal significance and effect size interpretation to address the research question. Multiple testing correction (e.g., Bonferroni) was not applied given the exploratory nature and would have further increased the significance threshold to $|\rho|\geq0.68$ for 51 tests, making detection of any associations impossible with n=15.

Discussion

The analysis provides partial support for the hypothesis that pathway-specific molecular features control the balance between DNA damage stress (SR-ATAD5) and mitochondrial stress (SR-MMP). ODI_Std, representing variability in orbital energy distribution across the molecular structure, emerged as the second-strongest correlate ($\rho=-0.479$, $p=0.071$, rank 2/51), approaching marginal significance. This finding aligns with the hypothesis that orbital energy distribution metrics would associate with stress pathway divergence. The negative correlation indicates that molecules with more uniform orbital energy distribution (lower ODI_Std) tend to show higher Divergence Ratios, favoring DNA damage stress over mitochondrial stress. Quadrupole_Moment, representing charge asymmetry, also showed a negative correlation ($\rho=-0.357$, rank 6/51), consistent with predictions, though with weaker evidence.

Critically, the universal stress drivers LEA_Var and LEA_Ave—which previous analyses identified as strongly associated with general stress response activation—showed weak correlations with the Divergence Ratio (ranks 18 and 42 of 51 descriptors). This confirms the hypothesis that pathway-specific drivers are distinct from general toxicity triggers. While LEA descriptors predict whether stress pathways will be activated at all, ODI_Std and related electronic descriptors appear to modulate the relative balance between specific stress pathways once activation occurs.

The dominance of negative correlations among top descriptors (9 of top 10) reveals a general mechanistic trend: molecules with higher electronic heterogeneity, larger molecular size, and greater charge asymmetry preferentially activate mitochondrial stress pathways over DNA damage pathways. The Overall_Average descriptor (mean electrostatic potential on molecular surface, $\rho=-0.482$, rank 1/51) was the strongest correlate, suggesting that surface electronic properties may be key mediators of pathway-specific toxicity.

The limited statistical power (n=15, requiring $|\rho|\geq0.514$ for $p<0.05$) prevented definitive confirmation, but the consistency of findings across multiple related descriptors (ODI_Std, Overall_Average, Nu, Quadrupole_Moment all showing moderate-to-strong negative correlations) suggests robust underlying relationships. The top correlation reached 93.8% of the required effect size, indicating that the relationships are likely real but underpowered in this small dataset.

Proposed Next Hypotheses

1. ODI_Std above 1.8 preferentially activates mitochondrial membrane potential disruption (SR-MMP) over DNA damage response (SR-ATAD5), with this threshold discriminating between conjugated metabolites (which show high SR-MMP despite low ODI_Std) and debrominated metabolites (which show low SR-MMP with high ODI_Std).
2. The combination of low Overall_Average (<0.0060) and high Quadrupole_Moment (>30)

synergistically predicts extreme mitochondrial stress dominance (Divergence Ratio <0.01), creating a two-descriptor classifier that outperforms single-descriptor models for pathway-specific toxicity prediction.