Conclusion

The vinyl group in M17, compared to the hydroxymethyl group in M19, induces selective DNA replication stress pathway activation (23.9× fold change in SR-ATAD5, 8.3× in SR-p53) through five specific molecular descriptor changes—TPSA (-46.1%), Octopole Moment (-33.8%), Quadrupole Moment (-27.1%), Dipole Moment (-23.8%), and Neg_Average (+20.2%)—all significantly correlated with SR-ATAD5 ($|\rho| > 0.67$, $p < 0.006$), while global reactivity descriptors like HOMO-LUMO gap remain unchanged (-2.6%, non-significant).

Methods

This analysis employed a systematic descriptor comparison approach to identify the molecular determinants of vinyl group selectivity in TBBPA metabolite M17. First, molecular descriptor data (prediction_results_with_smiles.csv, 51 descriptors) and toxicity predictions (smiles2_toxicity_results.csv, 31 endpoints) were merged on SMILES identifiers. M17 (SMILES: CC(C1=CC(Br)=C(O)C(Br)=C1)=C) and M19 (SMILES: CC(CO)C1=CC(Br)=C(O)C(Br)=C1) were isolated, and absolute and percent differences were calculated for all 51 descriptors. The phenolic metabolite subset (n=15) was identified using the derived artifact phenolic_oh_sr_are_analysis.csv. Spearman correlations between each descriptor and SR-ATAD5 probability were calculated for the phenolic subset using scipy.stats.spearmanr. Descriptors meeting dual criteria (>20% absolute difference between M17/M19 AND significant correlation with SR-ATAD5 at $p<0.05$) were identified. The top 5 descriptors were further analyzed for selectivity by calculating correlations with all five stress response endpoints (SR-ATAD5, SR-p53, SR-ARE, SR-HSE, SR-MMP) and computing DNA damage selectivity ratios (average $|\rho|$ for ATAD5/p53 divided by average $|\rho|$ for ARE/HSE/MMP). HOMO-LUMO gap was specifically examined to test the hypothesis about global reactivity descriptors. All statistical analyses were performed in Python using pandas, numpy, and scipy.

Results

Seven descriptors met both criteria (>20% difference AND $p<0.05$ correlation with SR-ATAD5). The top 5, ranked by absolute Spearman correlation strength, were:

1. **Quadrupole_Moment**: M17=13.19, M19=18.09, Δ=-27.1%, ρ=-0.696, p=0.0039
2. **Octopole_Moment**: M17=96.66, M19=146.01, Δ=-33.8%, ρ=-0.693, p=0.0042
3. **TPSA**: M17=20.29, M19=37.65, Δ=-46.1%, ρ=-0.689, p=0.0045
4. **Neg_Average**: M17=-0.0164, M19=-0.0205, Δ=+20.2%, ρ=+0.675, p=0.0058
5. **Dipole_Moment**: M17=3.28, M19=4.31, Δ=-23.8%, ρ=-0.671, p=0.0061

All five descriptors showed DNA damage pathway selectivity ratios >1.2 (range: 1.20-1.35), indicating stronger correlations with SR-ATAD5/SR-p53 than with other stress endpoints (SR-ARE, SR-HSE, SR-MMP). All five showed stronger absolute correlation with SR-ATAD5 than with SR-ARE (oxidative stress).

The hypothesis regarding global reactivity descriptors was confirmed: HOMO-LUMO gap showed only -2.57% difference (M17=0.332 eV, M19=0.341 eV), with non-significant correlation to SR-ATAD5 (ρ=+0.186, p=0.5075). HOMO and LUMO individually also showed minimal changes (+1.92% and -11.0%, respectively).

M17 exhibited 23.88× higher SR-ATAD5 probability (0.0689 vs 0.0029) and 8.33× higher SR-p53 probability (0.447 vs 0.054) compared to M19, with smaller fold changes for other stress endpoints (SR-ARE: 4.62×, SR-HSE: 8.61×, SR-MMP: 2.33×).

Challenges

The primary analytical challenge was the small sample size (n=15 phenolic metabolites), which limits statistical power for detecting correlations. However, the strong effect sizes ($|\rho| > 0.67$) and low p-values (all <0.007 for top 5) provide confidence in the findings despite the limited sample. The dual-criteria filtering strategy (>20% difference AND significant correlation) was essential to avoid identifying descriptors that either differed between M17/M19 but were unrelated to toxicity, or correlated with toxicity but didn't explain the M17-M19 difference. The analysis successfully confirmed the hypothesis about global vs local descriptor selectivity, with multipole moments (quadrupole, octopole) and surface electrostatic properties (TPSA, Neg_Average) showing strong effects while frontier orbital energies (HOMO-LUMO gap) showed negligible changes. One limitation is that the descriptors available in the dataset do not include truly localized metrics near the vinyl/hydroxymethyl substitution site, so we relied on whole-molecule descriptors that reflect the structural difference indirectly.

Discussion

This analysis provides mechanistic insight into why the vinyl group in M17 selectively increases predicted DNA replication stress pathway activation compared to the hydroxymethyl analog M19. The key finding is that this selectivity operates through changes in molecular electrostatic properties and shape (multipole moments, topological polar surface area, surface electrostatic potential) rather than through global electronic reactivity (HOMO-LUMO gap unchanged).

The 46.1% reduction in TPSA when replacing hydroxymethyl with vinyl suggests reduced hydrogen bonding capacity and altered membrane permeability, potentially affecting cellular distribution and DNA interaction. The substantial decreases in quadrupole (-27.1%) and octopole (-33.8%) moments indicate altered charge distribution asymmetry, which may affect specific protein-ligand interactions in DNA damage response pathways. The 20.2% increase in Neg_Average (average negative electrostatic potential on the molecular surface) suggests enhanced electron-rich regions that could facilitate binding to positively charged DNA-binding proteins or DNA itself.

Critically, all five descriptor changes showed selectivity ratios >1.2, meaning they correlate more strongly with DNA damage pathways (SR-ATAD5, SR-p53) than with other stress responses (oxidative stress, heat shock, mitochondrial membrane potential). This mechanistic specificity supports the hypothesis that the vinyl group induces unique structural changes beyond general toxicity enhancement.

The negligible change in HOMO-LUMO gap (-2.6%) confirms that the vinyl and hydroxymethyl substituents do not significantly alter global frontier orbital energies or general electrophilic/nucleophilic reactivity. This is consistent with both substituents being non-conjugated with the aromatic ring system and suggests the toxicity difference arises from conformational, steric, or localized electrostatic effects rather than redox chemistry.

These findings add critical mechanistic detail to the causal chain: vinyl group → reduced TPSA + altered multipole moments + enhanced negative surface potential → selective DNA damage pathway activation. This represents a distinct mechanism from the phenolic OH-driven stress response previously identified in the dataset.

Proposed Next Hypotheses

1. The reduced TPSA and altered multipole moments in M17 compared to M19 lead to differential subcellular localization, with M17 showing enhanced nuclear accumulation and direct DNA interaction, testable through molecular docking studies with DNA-binding domains of ATAD5 or p53 proteins.

2. The 20.2% increase in Neg_Average (negative surface electrostatic potential) in M17 compared to M19 enhances binding affinity to the positively charged DNA minor groove, which can be validated experimentally using surface plasmon resonance or isothermal titration calorimetry with DNA oligonucleotides.

Artifacts

Artifact 1:

File name: m17_m19_descriptor_analysis.csv

Artifact description: A comprehensive summary table containing the top 5 molecular descriptors that drive the vinyl group (M17) selectivity for DNA replication stress pathway activation. For each descriptor, the table includes M17 and M19 values, percent difference, Spearman correlation with SR-ATAD5 (rho and p-value), correlation with SR-p53, and DNA damage selectivity ratio. This artifact was created by merging descriptor difference calculations with Spearman correlation analysis across the phenolic metabolite subset (n=15), followed by dual-criteria filtering (>20% difference AND $p<0.05$) and selectivity ratio computation.