Mapping the Topological Rewiring of Molecular Joint Property Structures

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Molecular properties are conventionally treated as independent targets, which overlooks the fundamental joint coupling between energy, geometry, and electronic states constrained by chemical architecture. To systematically map how spatial confinement rewires these physical dependencies, we isolate topological effects from trivial stoichiometric scaling. Using the QM9 dataset, we compute residuals for twelve non-redundant properties via gradient-boosted regressors trained solely on atom counts, followed by a Gaussian rank transformation. We then compare the Ledoit-Wolf covariance matrices of 338 populous cyclic scaffolds against size- and composition-matched acyclic baselines to quantify structural rewiring through shifts in effective dimensionality and principal component alignment. Additionally, we employ random forest modeling and uniform manifold approximation and projection (UMAP) to analyze the structural drivers and visualize the covariance structures. The results demonstrate that cyclic architecture systematically compresses the property manifold and rotates the primary axes of physical variation away from the acyclic norm. Furthermore, the findings identify topological complexity and molecular shape as the primary drivers of these transformations, organizing molecular joint property dependencies into a coherent, deterministic landscape.

1 Introduction

The conventional study of molecular properties predominantly treats individual physical and chemical characteristics as independent targets for prediction and analysis. However, the underlying physics of a molecule is fundamentally joint; the coupling between energy, geometry, and electronic states is strictly constrained by the molecule's chemical architecture. While the isolated effects of specific structural motifs—such as ring strain or aromaticity—on individual properties are well documented, the systematic reorganization of the *joint* property manifold remains largely unexplored. Understanding how spatial confinement and topological complexity rewire these physical dependencies is critical for advancing both fundamental chemical physics and the design of novel molecular architectures.

1.1 Isolating architecture from stoichiometry

A primary obstacle in mapping this topological rewiring is the overwhelming influence of stoichiometric determinism. In standard datasets such as QM9, the variance of most extensive and even intensive properties is heavily dominated by trivial scaling with molecular size and atomic composition. Consequently, any intrinsic joint structure arising from spatial confinement is masked by the baseline correlation of properties to simple atom counts.

To systematically map how cyclic architecture reorganizes molecular physics, it is necessary to isolate topological effects from this stoichiometric baseline. In this work, we achieve this by computing residuals for twelve non-redundant properties via gradient-boosted regressors trained solely on atom counts. By analyzing the Gaussian-rank transformed residuals of these properties, we ensure that the observed joint structures are purely a function of chemical architecture rather than molecular size or composition.

1.2 Quantifying structural rewiring

We quantify the structural rewiring of molecular physics by comparing the robust Ledoit-Wolf covariance matrices of 338 populous cyclic scaffolds against size- and composition-matched acyclic baselines. We introduce two primary metrics to capture the deformation of the tightly coupled thermodynamic and geometric state space: the shift in effective dimensionality, measured via the Participation Ratio (PR), and the rotation of the primary axes of physical variation, measured by the cosine similarity of the leading principal components.

Our findings demonstrate that cyclic architecture systematically compresses the property manifold, effectively reducing the degrees of freedom within the joint property space. Furthermore, the primary axes of physical variation are frequently rotated away from the acyclic norm, indicating a fundamental shift in how properties couple under spatial confinement. By projecting these covariance structures into a unified meta-space using uniform manifold approximation and projection (UMAP) and analyzing the structural drivers via random forest modeling, we reveal that this rewiring is a continuous, deterministic response to topological complexity and molecular shape. This approach provides a robust framework for visualizing and quantifying the hidden dependency networks that govern molecular behavior, moving beyond the paradigm of independent property prediction.

2 Methods

2.1 Data curation and stoichiometric baseline removal

To systematically isolate the effects of chemical architecture from stoichiometric determinism, we utilized the QM9 dataset. The dataset was first deduplicated by canonicalizing the SMILES strings and retaining only the first instance of any duplicates to prevent data leakage. We isolated 12 non-redundant physical and chemical properties, explicitly dropping redundant thermodynamic restatements (‘u298‘, ‘h298‘, ‘g298‘, ‘u298_atom‘, ‘h298_atom‘, ‘g298_atom‘) and the HOMO-LUMO ‘gap‘. To safely handle extreme variance and linear molecules where the rotational constant A is zero, we applied a $\log(A+1)$ transformation to this property.

Following the theoretical framework established in the introduction, we removed the stoichiometric baseline by extracting the exact counts of carbon, nitrogen, oxygen, fluorine, and hydrogen atoms for each molecule. We then trained gradient boosting regressors to predict each of the 12 non-redundant properties using solely these five atom counts as features. Out-of-fold predictions were computed using cross-validation to obtain unbiased residuals for every molecule. Finally, a Gaussian rank transformation (quantile normalization) was applied to these residuals, mapping their marginals to a standard normal distribution. This ensured that downstream covariance estimates were robust to outliers and that the remaining variance was strictly a function of molecular architecture.

2.2 Scaffold extraction and matched acyclic baselines

To quantify the topological rewiring caused by spatial confinement, we computed the Murcko scaffold for each molecule. Molecules yielding an empty Murcko scaffold were designated as the acyclic baseline pool. We calculated the frequencies of all cyclic scaffolds and filtered out those containing fewer than 50 molecules to ensure statistical stability, resulting in a final set of 338 populous cyclic scaffolds.

To guarantee that the observed rewiring was driven by ring structure rather than confounding effects of size, branching, or composition, we constructed a specific matched acyclic baseline for each of the 338 cyclic scaffolds. For every cyclic scaffold, we computed the average heavy atom count, the fraction of sp^3 hybridized carbons (FractionCSP3), and the heteroatom-to-carbon ratio. These features were standardized, and a nearest-neighbor matching approach based on Euclidean distance was employed to sample a subset of molecules from the acyclic pool that matched the joint structural distribution of the cyclic scaffold.

2.3 Covariance estimation and dimensionality metrics

For each of the 338 cyclic scaffolds and their corresponding matched acyclic baselines, the 12-dimensional normalized residuals were standardized within the specific group. Zero-variance cases were explicitly handled

Descriptor	Importance (ΔPR Compression)	Importance (PC1 Axis Rotation)
Hall-Kier Alpha	0.5119	0.3468
BertzCT	0.3927	0.6214
Aromaticity	0.0559	0.0056
Ring Count	0.0395	0.0263

by setting the standardized values to zero. We computed the covariance matrix for each group using the Ledoit-Wolf shrinkage estimator, which provides a stable, regularized estimate of the correlation structure. To ensure numerical stability during eigen-decomposition and prevent near-singular matrices, a minor ridge regularization term of $10^{-6}\mathbf{I}$ was added to each covariance matrix.

Eigenvalues (λ_i) and eigenvectors were extracted for all groups. To quantify the effective dimensionality and the degree of coupling within the property space, we computed the Participation Ratio (PR) for each group, defined as:

$$PR = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \tag{1}$$

The structural rewiring was then quantified using two primary metrics. First, we calculated the shift in the Participation Ratio ($\Delta PR = PR_{\text{cyclic}} - PR_{\text{acyclic}}$) to measure the compression of the joint property space. Second, we computed the absolute cosine similarity between the leading eigenvector (PC1) of the cyclic scaffold and the leading eigenvector of its matched acyclic baseline to measure the rotation of the primary axes of physical variation.

2.4 Null model and permutation testing

To validate that the observed ΔPR and eigenvector rotations were specific to chemical architecture and not statistical artifacts of sampling, we constructed a size-controlled null model. All molecules from the 338 cyclic scaffolds and the acyclic pool were combined and binned by their heavy-atom count. We generated pseudo-scaffolds by randomly sampling molecules strictly within the same heavy-atom count bins to exactly match the size and heavy-atom distribution of the true cyclic scaffolds. The within-group standardization, Ledoit-Wolf covariance estimation, Participation Ratios, and eigenvectors were recomputed for these pseudo-scaffolds to establish null distributions for statistical comparison.

2.5 Structural drivers and manifold visualization

To identify the specific physical motifs driving the compression of the joint space and the rotation of the principal property axes, we computed average topological descriptors for each of the 338 cyclic scaffolds. These included the fraction of aromatic atoms, ring count, BertzCT (a measure of topological complexity), and the Hall-Kier Alpha index (a measure of molecular shape and hybridization).

We trained two random forest regressors: one to predict the ΔPR of the cyclic scaffolds, and another to predict the PC1 cosine similarity (alignment with the acyclic baseline). The feature importances extracted from these models (Table 2.5) were used to quantify the structural drivers of property reorganization.

Finally, to visualize the landscape of structural rewiring, the robust Ledoit-Wolf covariance matrices were flattened and projected into a unified two-dimensional meta-space using uniform manifold approximation and projection (UMAP). This embedding allowed us to map the distance and directional trajectory of each cyclic scaffold away from the acyclic baseline, organizing the molecular joint property dependencies into a coherent manifold.

3 Results

3.1 Isolating architecture from stoichiometry

The variance of molecular properties in the QM9 dataset is heavily dominated by trivial scaling with molecular size and atomic composition. For instance, linear regression on the exact counts of C, N, O, F, and H atoms yields an R^2 of 0.9999 for internal energy (u_0) and 0.9971 for zero-point vibrational energy ($zpve$). By

Table 1: Random Forest feature importances for predicting dimensionality compression (ΔPR) and axis rotation (PC1 alignment) using RDKit topological descriptors.

Descriptor	Importance (ΔPR Compression)	Importance (PC1 Axis Rotation)
Hall-Kier Alpha	0.5119	0.3468
BertzCT	0.3927	0.6214
Aromaticity	0.0559	0.0056
Ring Count	0.0395	0.0263

utilizing Gaussian-rank transformed residuals from gradient-boosted regressors trained on atom counts, we successfully removed this stoichiometric baseline. Consequently, the residual variance and the resulting covariance structures analyzed below reflect intrinsic joint dependencies dictated solely by spatial confinement and molecular topology, completely isolated from stoichiometric determinism.

3.2 Dimensionality compression and axis rotation

We evaluated the structural rewiring of 338 populous cyclic scaffolds (each with ≥ 50 molecules) by comparing their joint property structures to size- and composition-matched acyclic baselines. Across the population, cyclic architecture consistently compresses the property manifold. The average shift in the Participation Ratio (ΔPR) is -2.3888 ± 1.4421, indicating that the 12-dimensional property space typically loses more than two effective degrees of freedom when confined to a ring. This compression is significantly stronger than expected from random sampling; our size-matched null model yielded a mean PR of 5.3082 ± 0.5014, whereas true cyclic systems frequently drop well below this baseline.

Concurrently, the primary axes of physical variation rotate substantially. The average PC1 alignment (absolute cosine similarity) across the scaffolds is 0.6122 ± 0.2429. As illustrated in the rewiring landscape (), most scaffolds exhibit both a significant reduction in dimensionality and a moderate-to-high axis rotation. The data reveals that as topological complexity increases, the property space becomes more tightly coupled, and the primary axes of variation rotate further from the acyclic norm.

To illustrate severe rewiring, we highlight the aromatic heterocycle scaffold `N=c1cconc1`, which exhibits a PC1 alignment of just 0.0047—a near-orthogonal rotation of the primary physical axis. An analysis of the principal component loadings () demonstrates that properties such as electronic spatial extent (r^2) and heat capacity (c_v), which are strongly coupled in acyclic systems, decouple or shift their dominance to secondary axes in strained or aromatic cyclic environments.

3.3 Structural drivers of property reorganization

To identify the physical motifs driving this rewiring, we analyzed the feature importances from random forest models trained to predict ΔPR and PC1 alignment using RDKit-derived topological descriptors (and).

The results identify topological complexity, measured by the BertzCT index, as the dominant predictor of axis rotation (importance = 0.6214). Conversely, dimensionality compression (ΔPR) is primarily driven by the Hall-Kier Alpha descriptor (importance = 0.5119), which captures molecular shape and hybridization state. Remarkably, simple aromaticity and raw ring count are relatively poor predictors of both metrics. This indicates that the rewiring of molecular physics is not a binary consequence of ring presence, but rather a continuous response to the integrated topological complexity and shape of the scaffold.

3.4 The manifold of joint structures

Projecting the scaffold covariance matrices into a UMAP embedding reveals that these joint property structures organize into a coherent manifold (). Scaffolds do not occupy this meta-space randomly; instead, they cluster strictly by their rewiring behavior.

Simple cyclic systems, such as cyclopropane (`C1CC1`, $\Delta PR = -1.4713$), occupy regions closer to the acyclic baseline. In contrast, highly rewired and compressed scaffolds, such as pyridine (`c1ccncc1`, $\Delta PR = -4.3004$), cluster at the opposite end of the manifold, forming distinct high-rewiring poles.

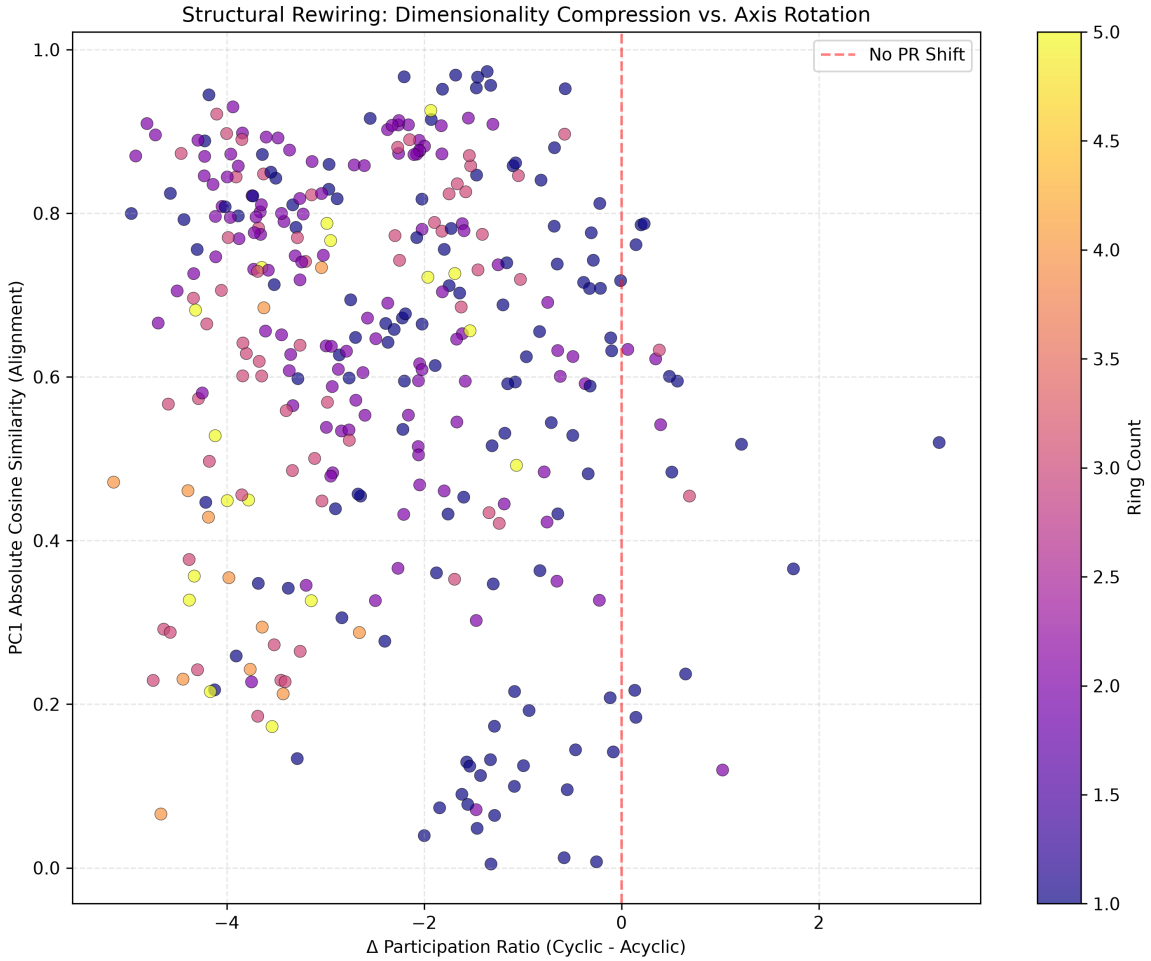

Figure 1: The rewiring landscape of cyclic scaffolds illustrating dimensionality compression (ΔPR) versus axis rotation (PC1 alignment). Each point represents a cyclic scaffold compared to a size-and-composition-matched acyclic baseline, with colors indicating the ring count. The vertical dashed line denotes zero shift in the participation ratio. The concentration of scaffolds in the negative ΔPR region demonstrates that cyclic confinement consistently compresses the 12-dimensional property manifold, tightening property couplings and reducing degrees of freedom. Concurrently, the broad spread in PC1 absolute cosine similarity indicates a significant rotation of the primary axes of physical variation away from the acyclic norm. From this distribution, we learn that cyclic architecture fundamentally reorganizes the joint structure of molecular properties, with higher ring counts generally contributing to more tightly coupled property spaces and greater structural rewiring.

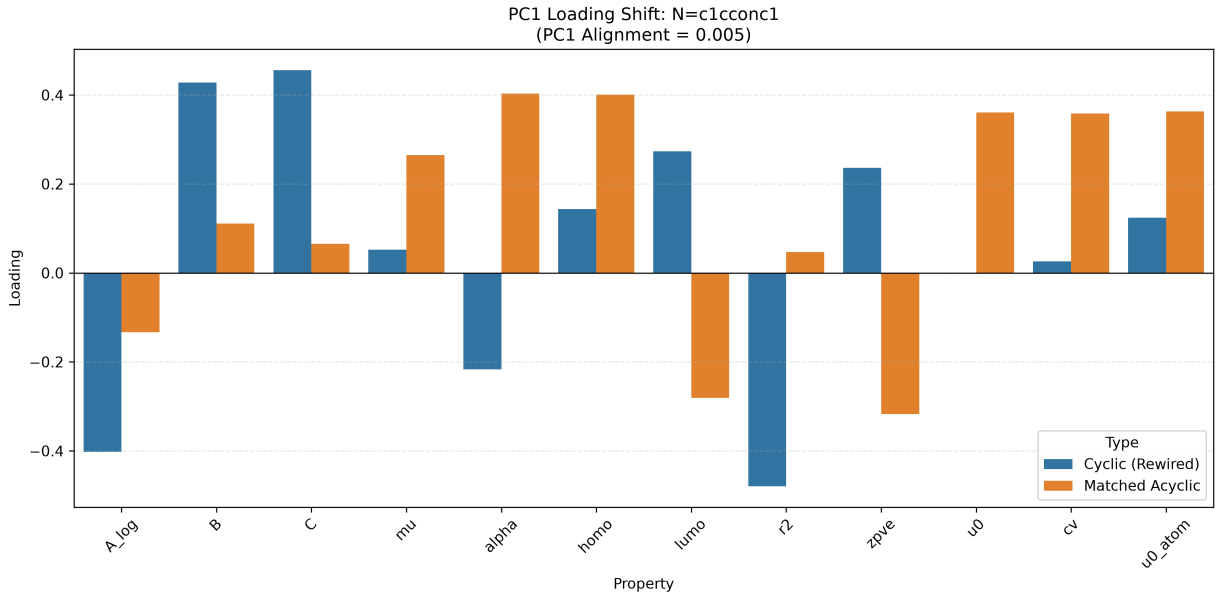

Figure 2: Principal component 1 (PC1) loading shifts for 12 properties of the aromatic heterocycle scaffold `N=c1cconc1` compared to its size-and-composition-matched acyclic baseline. Exhibiting a near-orthogonal PC1 alignment of 0.005, this scaffold represents a case of extreme topological rewiring. The comparison reveals that the primary axis of physical variation rotates significantly between the two architectures; properties such as heat capacity (c_v) and internal energy (u_0) dominate the acyclic primary axis but decouple in the cyclic environment, whereas electronic spatial extent (r^2) and rotational constants (B, C) shift to become dominant. This illustrates that cyclic architecture fundamentally reorganizes the joint dependency network of molecular physics, isolating topological rewiring effects from trivial stoichiometric scaling.

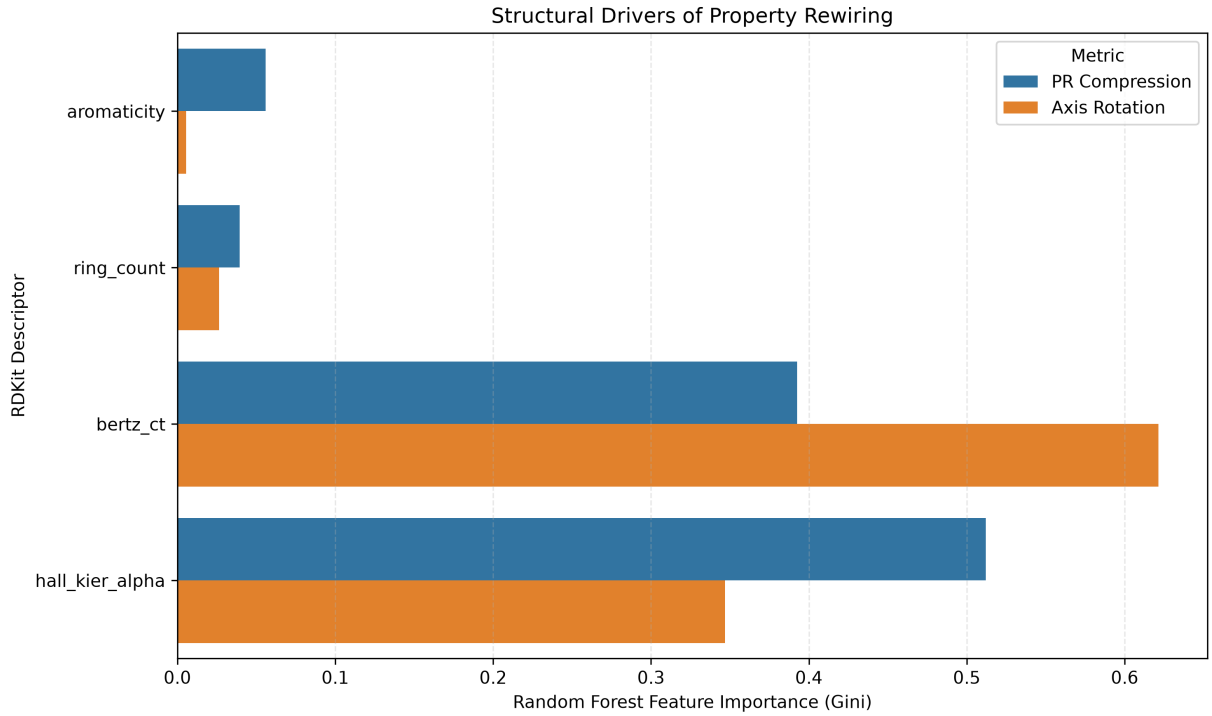

Figure 3: Random Forest feature importances (Gini) of RDKit topological descriptors for predicting molecular property rewiring metrics: dimensionality compression (ΔPR) and axis rotation (PC1 alignment). The plot reveals that topological complexity, measured by the `bertz_ct` index, is the dominant predictor of axis rotation, while molecular shape and hybridization, captured by `hall_kier_alpha`, primarily drive dimensionality compression. Conversely, basic features like `ring_count` and aromaticity are relatively poor predictors. This demonstrates that the reorganization of a molecule's joint physical properties is not a binary consequence of having a ring, but rather a continuous response to the integrated topological complexity and shape of the chemical scaffold.

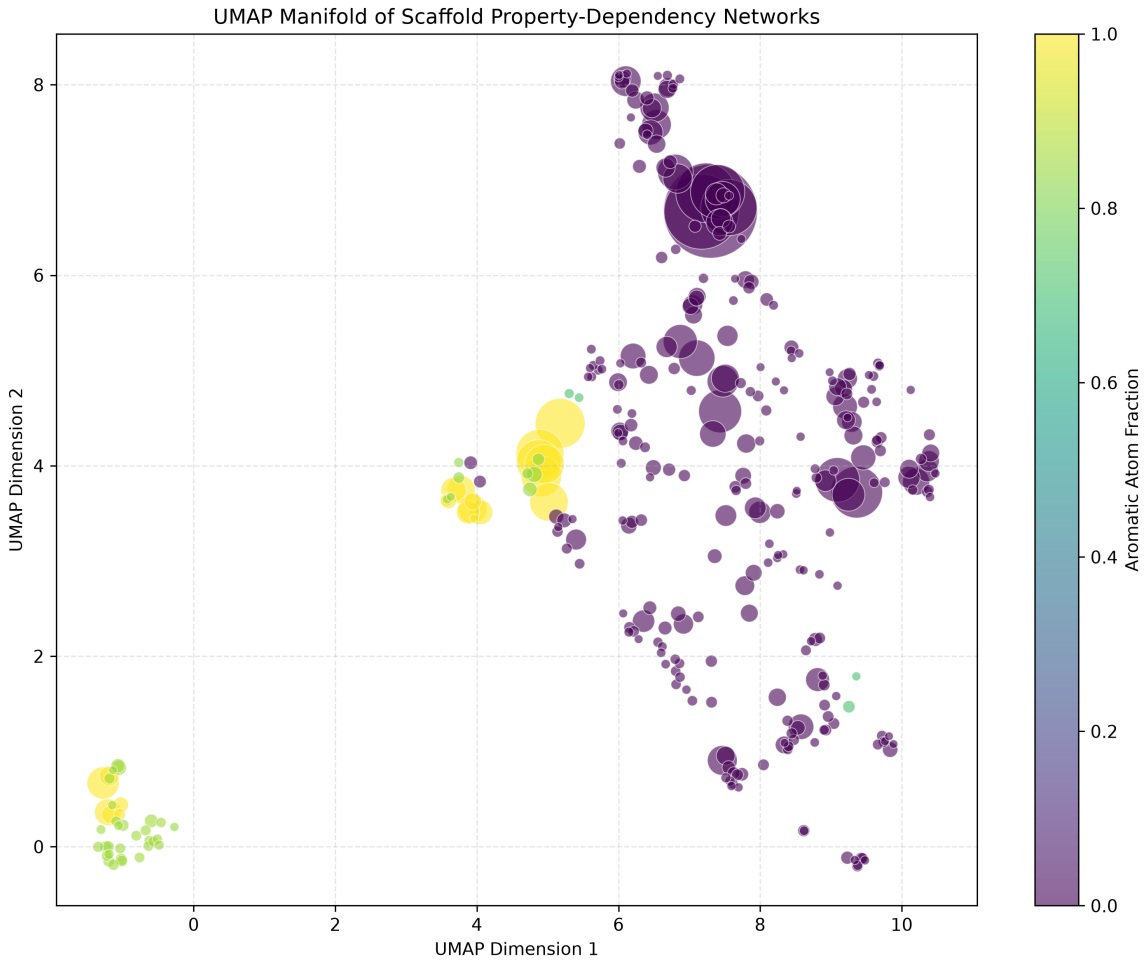

Figure 4: UMAP projection of scaffold covariance matrices illustrating the manifold of joint property structures. Individual scaffolds are colored by their aromatic atom fraction, revealing distinct structural gradients across the meta-space. The visualization demonstrates that scaffolds do not occupy this space randomly but cluster according to their topological rewiring behavior. This organization spans from acyclic-like systems to highly rewired poles characterized by significant dimensionality compression (negative ΔPR). Ultimately, the manifold confirms that the joint dependency network of a molecule is a deterministic function of its cyclic architecture.

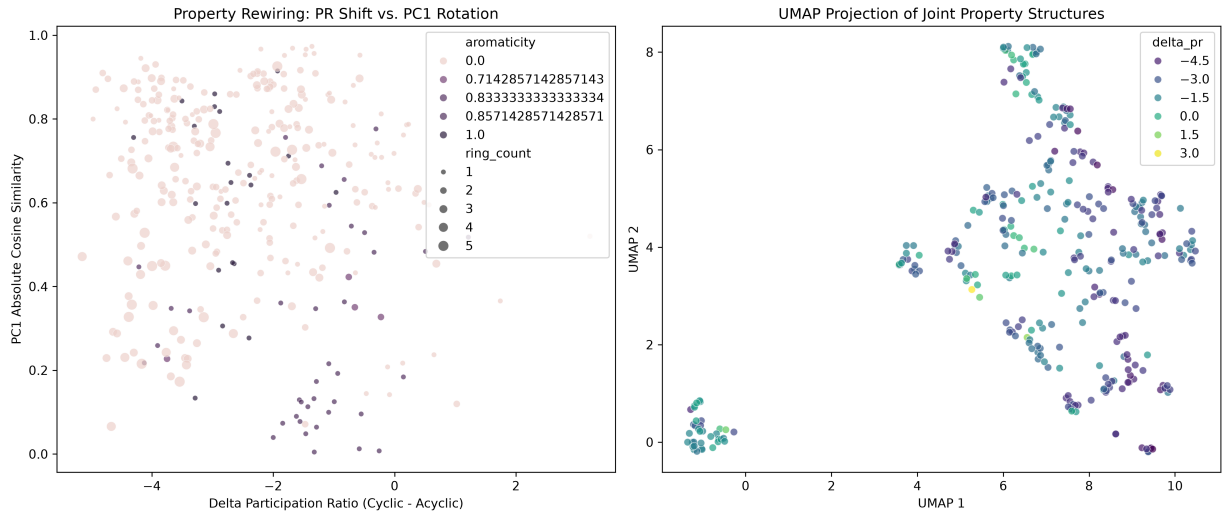

Figure 5: The figure illustrates the topological rewiring of molecular properties driven by cyclic architecture. The left panel displays the rewiring landscape, plotting the principal component axis rotation (PC1 absolute cosine similarity) against the effective dimensionality shift (ΔPR). Most cyclic scaffolds exhibit significant dimensionality compression (negative ΔPR) and moderate-to-high axis rotation compared to acyclic baselines, with these effects intensifying as ring count and aromaticity increase. The right panel presents a UMAP projection of the scaffold covariance matrices, colored by ΔPR. This projection reveals that joint property structures organize into a coherent manifold where scaffolds cluster by their rewiring behavior, with highly compressed systems forming distinct poles. Together, these visualizations demonstrate that a molecule's joint dependency network is a deterministic function of its cyclic architecture, which fundamentally rewires internal physics rather than merely scaling individual property values.

The UMAP manifold displays clear gradients in topological complexity and aromaticity, confirming that the joint dependency network of a molecule is a deterministic function of its cyclic architecture. The systematic reduction in effective dimensionality and the profound rotation of property axes, summarized comprehensively in , demonstrate that cyclic architecture fundamentally rewires the internal physics of the molecule. This joint structure, while invisible to standard pairwise correlation analysis, is robustly mapped using the manifold of property-dependency networks.

4 Conclusions

In this study, we have demonstrated that the conventional paradigm of treating molecular properties as independent targets overlooks the fundamental, joint coupling of molecular physics. By isolating chemical architecture from stoichiometric determinism, we systematically mapped how spatial confinement rewires the dependency networks of twelve non-redundant properties. Our methodology, which utilized gradient-boosted atom-count residuals and robust Ledoit-Wolf covariance estimation, successfully stripped away the overwhelming influence of molecular size and composition to reveal the intrinsic topological effects of cyclic scaffolds.

Our analysis reveals that cyclic architecture fundamentally reorganizes the joint property manifold through two primary mechanisms: dimensionality compression and axis rotation. Compared to size- and composition-matched acyclic baselines, cyclic scaffolds consistently exhibit a negative shift in the participation ratio (ΔPR), effectively compressing the property space by reducing its degrees of freedom. Concurrently, the primary axes of physical variation undergo significant rotation, demonstrating that properties which are strongly coupled in acyclic systems often decouple or shift their dominance under spatial confinement.

Furthermore, we identified the specific structural motifs that drive this physical reorganization. As de-

tailed in Table 2.5, random forest modeling indicates that topological complexity, quantified by the BertzCT index, is the dominant driver of principal axis rotation. Conversely, dimensionality compression is primarily governed by molecular shape and hybridization, as captured by the Hall-Kier Alpha descriptor.

Notably, basic features such as raw ring count and aromaticity are relatively poor predictors of these rewiring metrics. This confirms that the reorganization of molecular physics is a continuous, deterministic response to the integrated topological complexity of a scaffold, rather than a simple binary consequence of ring formation.

Finally, the projection of these joint property structures into a unified UMAP meta-space confirms the existence of a coherent, deterministic landscape of molecular rewiring. Scaffolds organize along clear gradients of topological complexity, spanning from acyclic-like clusters to highly rewired poles. Ultimately, this work establishes that cyclic architecture does not merely scale individual property values, but fundamentally rewires the internal physics of the molecule. By mapping these hidden dependency networks, we provide a robust new framework for understanding molecular behavior, offering critical insights for the future design and optimization of complex molecular architectures.