Supporting Information for: Beyond Minimum Energy Conical Intersections: A Data-Driven Reconstruction of the Accessible Intersection Seam

Conner J. Baucom,[a] Eleftherios Mainas[a] and Elisa Pieri*[a]

Contents

1 Preprocessing of Cartesian coordinates

Before dimensionality reduction and any other analysis, when using Cartesian coordinates every SP geometry must be brought into a common reference frame so that distances in Cartesian feature space reflect genuine structural differences. For our set of highly symmetrical test molecules, the necessary preprocessing alignment procedure consists of three layers: (i) the rigid-body superposition algorithm, (ii) the treatment of molecular permutation symmetry, and (iii) the strategy used to select reference structures for alignment.

1.1 Rigid-body alignment

To compare geometries in Cartesian space, rigid-body rotations and translations must be removed. We employ the Kabsch algorithm, [1,2] which we briefly summarize below. Given two geometries represented by coordinate matrices $\mathbf{P}$ and $\mathbf{Q}$, the alignment is performed by minimizing the root-mean-square deviation (RMSD), defined as

$$\mathrm{RMSD}(\mathbf{P},\mathbf{Q}) = \sqrt{\frac{\sum_{i=1}^{N} w_i \left\|\mathbf{p}_i - \mathbf{q}_i\right\|^2}{\sum_{i=1}^{N} w_i}}, \tag{1}$$

where N is the number of atoms, w_i are atomic weights (set to unity unless otherwise specified), and $\mathbf{p}_i$ and $\mathbf{q}_i$ are the Cartesian coordinates of atom i in the two geometries. Coordinates are treated as row vectors. The Kabsch algorithm determines the optimal rotation $\mathbf{R} \in \mathrm{SO}(3)$ and translation $\mathbf{t}$:

$$\mathrm{RMSD}^2(\mathbf{R},\mathbf{t}) = \frac{\sum_{i=1}^{N} w_i \left\|\mathbf{p}_i - (\mathbf{q}_i\mathbf{R} + \mathbf{t})\right\|^2}{\sum_{i=1}^{N} w_i}. \tag{2}$$

For fixed $\mathbf{R}$, the minimizing translation maps the weighted centroid of $\mathbf{Q}$ onto that of $\mathbf{P}$,

$$\mathbf{t}^* = \bar{\mathbf{p}} - \bar{\mathbf{q}}\mathbf{R}, \qquad \bar{\mathbf{p}} = \frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}, \qquad \bar{\mathbf{q}} = \frac{\sum_i w_i \mathbf{q}_i}{\sum_i w_i}. \tag{3}$$

After centering each structure by subtracting its weighted centroid, we obtain the matrices $\tilde{\mathbf{P}} = \mathbf{P} - \bar{\mathbf{p}}$ and $\tilde{\mathbf{Q}} = \mathbf{Q} - \bar{\mathbf{q}}$. The problem then reduces to finding the rotation that maximizes $\mathrm{Tr}(\mathbf{R}^\top \mathbf{H})$, where

$$\mathbf{H} = \tilde{\mathbf{P}}^\top \mathbf{W} \tilde{\mathbf{Q}}, \qquad \mathbf{W} = \mathrm{diag}(w_1, \ldots, w_N). \tag{4}$$

Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition. The optimal proper rotation is

$$\mathbf{R}^* = \mathbf{V} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mathrm{sign}(\det(\mathbf{V}\mathbf{U}^\top)) \end{pmatrix} \mathbf{U}^\top, \tag{5}$$

which enforces $\det(\mathbf{R}^*) = +1$ and thus preserves chirality. Allowing improper rotations ($\det(\mathbf{R}) = -1$) corresponds to minimizing over $\mathrm{O}(3)$ rather than $\mathrm{SO}(3)$ and effectively permits reflection.[3,4] For strictly achiral geometries this distinction is immaterial; however, for geometries related by enantiomeric pyramidalization or out-of-plane distortions, reflections can map enantiomeric configurations onto one another. In the present work we allow $\mathrm{O}(3)$ alignments so that enantiomeric distortions are treated as equivalent. This avoids artificially splitting the dataset into mirror-related clusters and ensures that the analysis reflects seam morphology rather than handedness.

1.2 Permutation symmetry

Rigid-body alignment methods such as the Kabsch algorithm assume a fixed correspondence between atoms in the two structures being compared. The RMSD objective is defined over paired atoms (p_i, q_i), implicitly assuming that row i of $\mathbf{P}$ corresponds to row i of $\mathbf{Q}$ in equation 3. However, in molecular dynamics simulations of symmetric molecules, the same local structural event may occur at any of several symmetry-equivalent atoms. For example, in benzene a pyramidalization can take place at any of the six equivalent carbon atoms, and in ethylene or butadiene a hydrogen transfer may involve either of the symmetry-related hydrogen atoms attached to a carbon. These processes generate geometries that are physically equivalent but differ by a permutation of atom indices. If the atom correspondence is not resolved, rigid-body alignment may compare non-equivalent atoms and yield an artificially inflated RMSD. Therefore, when evaluating structural similarity we must identify the permutation σ^* of symmetry-equivalent atoms that yields the correct correspondence between structures, selecting the atom ordering/relabeling that has minimum RMSD after the alignment. The alignment problem is therefore extended to a joint optimization over rotations, translations, and permutations:

$$\sigma^* = \underset{\sigma}{\mathrm{argmin}}\, \mathrm{RMSD}(\mathbf{P}, \mathbf{Q}_\sigma) \tag{6}$$

where $\mathbf{Q}_\sigma$ denotes the coordinate matrix obtained by permuting the atoms of $\mathbf{Q}$ according to the permutation σ. For systems containing multiple indistinguishable atoms, identifying σ^* is a combinatorial problem whose worst-case cost grows factorially with the number of equivalent atoms. The most straightforward approach is brute force: all permutations that exchange atoms of the same element are explicitly enumerated. If n_X denotes the number of atoms of element X, the total number of permutations considered is $\prod_X n_X!$. For example, this yields $2! \cdot 4! = 48$ permutations for ethylene, $6! \cdot 6! = 518{,}400$ for benzene, and $4! \cdot 6! = 17{,}280$ for butadiene, all to be tested on each dataset point. In the brute-force approach, a full Kabsch alignment is performed for each candidate permutation σ, and the corresponding RMSD is evaluated; the permutation yielding the global minimum defines σ^*. This strategy guarantees recovery of the optimal atom correspondence and therefore serves as a reference standard. However, its factorial scaling renders it computationally expensive for systems containing more than a few symmetry-equivalent atoms or for datasets comprising thousands of geometries.

Alternative approaches While in this work we employ brute-force to solve permutations, for larger datasets and higher numbers of permutationally-invariant atoms, alternative strategies may be used. For example, we tested through RDKit[5] the ability of automorphism[6] to reduce the factorial cost while retaining chemical fidelity by exploiting the molecular symmetry of both reactant (i.e. SP) and product (i.e. reference structure). For a given molecule, a graph automorphism is a reordering of atom indices that leaves every bond intact - for example, the six rotations and six reflections of benzene's ring each constitute a valid automorphism. Rather than exhaustively testing all element-constrained permutations, we restrict the search to the set obtained by composing the automorphism groups of the SP geometry and the reference structure for each connectivity family. The composition ensures that symmetry-equivalent orderings from both the reactant and product are explored. In practice, for benzene this reduces the search from $518{,}400$ permutations to 12, with no loss in alignment quality. The method does, however, have a known limitation: permutations that require simultaneously reassigning a hydrogen from one heavy atom to another (as can occur during hydrogen migration, where bond assignment may become ambiguous depending on the connectivity threshold) are not reachable by composing graph automorphisms from either endpoint, because automorphisms preserve molecular connectivity and therefore cannot generate permutations corresponding to a change in bonding pattern while keeping the heavy-atom framework fixed. For such geometries, the brute-force ground truth and the automorphism result diverge, and the brute-force permutation should be preferred.

A second alternative we explored is a two-stage procedure combining maximum common substructure (MCS) detection with optimal bipartite matching. In the first stage, the largest fragment of atoms and bonds shared between the reference and target structures is identified using RDKit's MCS routine;[5] this shared skeleton, which remains well-defined even when bonds have broken or formed elsewhere in the molecule, is used to compute an initial Kabsch alignment. In the second stage, the resulting rigid rotation is applied to all atoms, and the remaining atom-to-atom correspondence is resolved element-wise by the Hungarian algorithm,[7] which finds the minimum-cost assignment between reference and target atoms of the same type based on their now-approximately-superimposed positions. This approach handles bond-breaking events gracefully, since the MCS anchor contracts to the unambiguous part of the molecule, and avoids factorial enumeration entirely. However, the method does not guarantee recovery of the global minimum-RMSD permutation: the Hungarian assignment is optimal only conditional on the initial MCS-derived rotation, which may itself be one of several possible orientations. In practice, we found that for the near-symmetric distorted geometries typical of photodynamics trajectories, this approach can yield suboptimal permutations that brute-force resolves correctly, making it unsuitable as a replacement for the reference standard in this context.

To quantify the reliability of the alternative strategies, we benchmarked the automorphism and MCS+Hungarian methods against the brute-force reference across all five photodynamics datasets used in this work. Brute-force and MCS+Hungarian calculations were parallelized

across all available CPU cores using Python's `ProcessPoolExecutor`, whereas automorphism required no parallelization due to its negligible runtime. For each geometry, the optimal permutation σ^* was determined independently by all three methods using the same reference structure. The permutations returned by automorphism and MCS+Hungarian were then compared element-wise to the brute-force result; a geometry was considered correctly resolved if the returned permutation matched the brute-force optimum. Per-dataset statistics are reported in Table 1. The alternative methods provide substantial speedups, but at the cost of some loss in accuracy. This effect is most pronounced for ethylene, where connectivity changes between SP families. Overall, although the brute-force approach is computationally demanding, its cost remains modest compared to that of running nonadiabatic dynamics and subsequent analysis.

Table 1 Comparison of permutation-search methods against the brute-force reference across datasets. For each dataset, the reference structure used for alignment is a symmetry-broken MECI geometry. Accuracy is reported as percentage of SPs for which the method returned a permutation identical to the brute-force optimum. Wall-clock times are for a single machine with 12 CPU cores; brute-force and MCS+Hungarian were parallelized, automorphism was not due to its already remarkable speed.

Molecule	N	Brute-force		Automorphism		MCS + Hungarian	
		permutations	time (s)	accuracy	time (s)	accuracy	time (s)
ethylene S_1/S_0	2557	48	1.7	60%	1.1	76%	2.0
butadiene S_1/S_0	1246	17280	65.9	98%	0.4	96%	2.4
butadiene S_2/S_1	881	17280	46.6	99%	0.2	100%	2.1
benzene S_1/S_0	3657	518400	6122.9	100%	1.2	100%	11.0
benzene S_2/S_1	1717	518400	2756.1	100%	0.6	100%	6.0

1.3 Reference-Selection Strategies

The choice of reference structure has a profound effect on the quality of the alignment, regardless of which permutation-search strategy is employed. We illustrate this by considering the extreme case of using highly symmetrical structures as reference. In these cases (for example, when using the planar D_{2h} equilibrium geometry of ethylene or the D_{6h} ground state of benzene) the minimum-RMSD alignment is nearly degenerate: many permutation–rotation combinations yield essentially the same RMSD, and the algorithm is forced to break the tie arbitrarily. As a consequence, two SP geometries that are physically similar (e.g., both exhibiting a slight pyramidalization at the same carbon) may be assigned different permutations and rotated into entirely different orientations, simply because the flat reference offers no geometric anchor to distinguish them. This inconsistency is invisible at the level of individual RMSD values to the given alignment reference, which appear perfectly reasonable, but it corrupts any subsequent analysis that relies on comparing aligned coordinates across the dataset, such as dimensionality reduction or clustering. The root cause is that the degeneracy of the alignment problem is controlled by the symmetry of the *reference*, not the target: a symmetric reference maximizes the number of equivalent solutions and therefore maximizes the risk of incoherent alignments. A symmetry-broken reference (such as a MECI geometry, which by construction has lower symmetry than the Franck–Condon point) provides a geometric anchor that yields consistent alignments for nearby structures. In this work, when a single reference is required we select the *medoid* of the MECI set, defined as the MECI geometry whose optimally permuted and aligned RMSD to all other MECIs is minimal. This choice provides a representative reference that lies near the centre of the MECI ensemble in RMSD space. Nevertheless, the use of a single reference can still bias the alignment and influence subsequent analysis, which motivates the exploration of alternative reference-selection strategies.

In the multi-reference alignment strategy, every geometry is aligned against the full set of MECI structures, all expressed in a common coordinate frame obtained by aligning each MECI to the medoid via brute-force permutation search and Kabsch rotation. The alignment yielding the lowest RMSD is retained, and the corresponding MECI defines the family assignment. This approach removes the need to select a single reference *a priori*: geometries near different regions of the seam (e.g., twisted vs. hydrogen-migration CIs) are aligned to the MECI that best represents their local structure. Since all MECIs share the medoid's coordinate frame, geometries assigned to different families remain geometrically consistent, and distance-based molecular representations (such as inverse distance matrices or atom-centered descriptors) are directly comparable across families. The main limitation is that the quality of the family assignment depends on the completeness of the MECI set: if an important region of the photochemical landscape is not represented, geometries from that region will be assigned to the nearest available MECI by default, potentially distorting the family boundaries. Conversely, if the MECI set contains near-duplicate structures, spawns near both will be assigned to whichever yields marginally lower RMSD, introducing an arbitrary boundary. Unlike single-reference alignment, where the RMSD measures distortion from a fixed structure, the multi-reference RMSD reflects proximity to the nearest known MECI - a physically meaningful coordinate that captures how far a geometry has evolved from its crossing region. In the Results, we examine how well these two strategies preserve the local and global embedding structure of the dataset.

Evaluation of reference choice strategies To determine the most appropriate reference selection strategy for each molecule, we evaluated how closely each method reproduces the optimal pairwise distances between geometries obtained from global pairwise alignment. We constructed a reference pairwise distance matrix $D^{(\mathrm{opt})}$, where each element corresponds to the minimal RMSD obtained by solving the full permutation (brute force) and rotation optimization for each geometry pair. When complete enumeration was computationally infeasible, a representative subset of $D^{(\mathrm{opt})}$ was computed using a subset of randomly selected SPs (200 for benzene and butadiene). For each candidate alignment strategy m, we then constructed the corresponding pairwise distance matrix $D^{(m)}$ and quantified its agreement with $D^{(\mathrm{opt})}$ using complementary global and local metrics. Global agreement between the matrices was evaluated using both rank-based (Spearman) and linear (Pearson) correlation coefficients, as well as complementary error-based measures including the Frobenius norm of their difference (reported in both absolute and relative form) and the mean squared error (MSE). In addition, local structure preservation was assessed

by measuring the overlap between the k-nearest-neighbor sets induced by $D^{(m)}$ and $D^{(\mathrm{opt})}$ at multiple neighborhood sizes. The reference strategy yielding the closest agreement with $D^{(\mathrm{opt})}$ across these metrics was selected for each system. When metrics were not in unanimous agreement, priority was given to preserving local neighborhood structure.

1.3.1 Optimal reference choice strategy

We first evaluated the performance of single- and multi-reference alignment strategies across representative systems using the metrics defined in the Methods (Figure 1). For ethylene and both butadiene manifolds (S_0/S_1 and S_1/S_2), the multi-reference approach consistently yields a more accurate reconstruction of the optimal pairwise distance matrix $D^{(\mathrm{opt})}$. In ethylene, the improvement is particularly pronounced, with the multi-reference scheme providing higher rank agreement (Spearman correlation), reduced global error (Frobenius norm and MSE), and markedly improved preservation of local neighborhood structure, as reflected in the k-nearest-neighbor overlap across all neighborhood sizes considered. A similarly consistent trend is observed for butadiene S_0/S_1, where multi-reference alignment outperforms the single-reference approach across both global and local metrics, indicating a more faithful representation of the overall distance landscape and its local connectivity.

For butadiene S_1/S_2 sampling, the comparison is more nuanced but still favors the multi-reference approach overall. While the single-reference scheme yields slightly higher overlap for the smallest neighborhood sizes, these differences are modest and confined to the most local regime. In contrast, the multi-reference approach provides systematically better global agreement with $D^{(\mathrm{opt})}$, as evidenced by higher correlation coefficients and lower Frobenius and MSE errors, and also improves neighborhood preservation at larger k. Taken together, these results demonstrate that incorporating multiple references is essential for capturing the configurational heterogeneity of these systems, leading to a more accurate and robust representation of both global and local geometric relationships.

In contrast, for benzene the two strategies yield comparable agreement with $D^{(\mathrm{opt})}$, with each performing marginally better for different subsets of the metrics. Given the relatively limited configurational heterogeneity of the benzene seam space, a single-reference description is sufficient to capture the relevant geometric relationships while maintaining a consistent reference definition within the molecule. Accordingly, we adopt the single-reference strategy for both benzene manifolds. The markedly greater configurational heterogeneity of ethylene and butadiene, which encompasses substantial connectivity changes, by comparison makes the multi-reference approach clearly preferable for those systems.

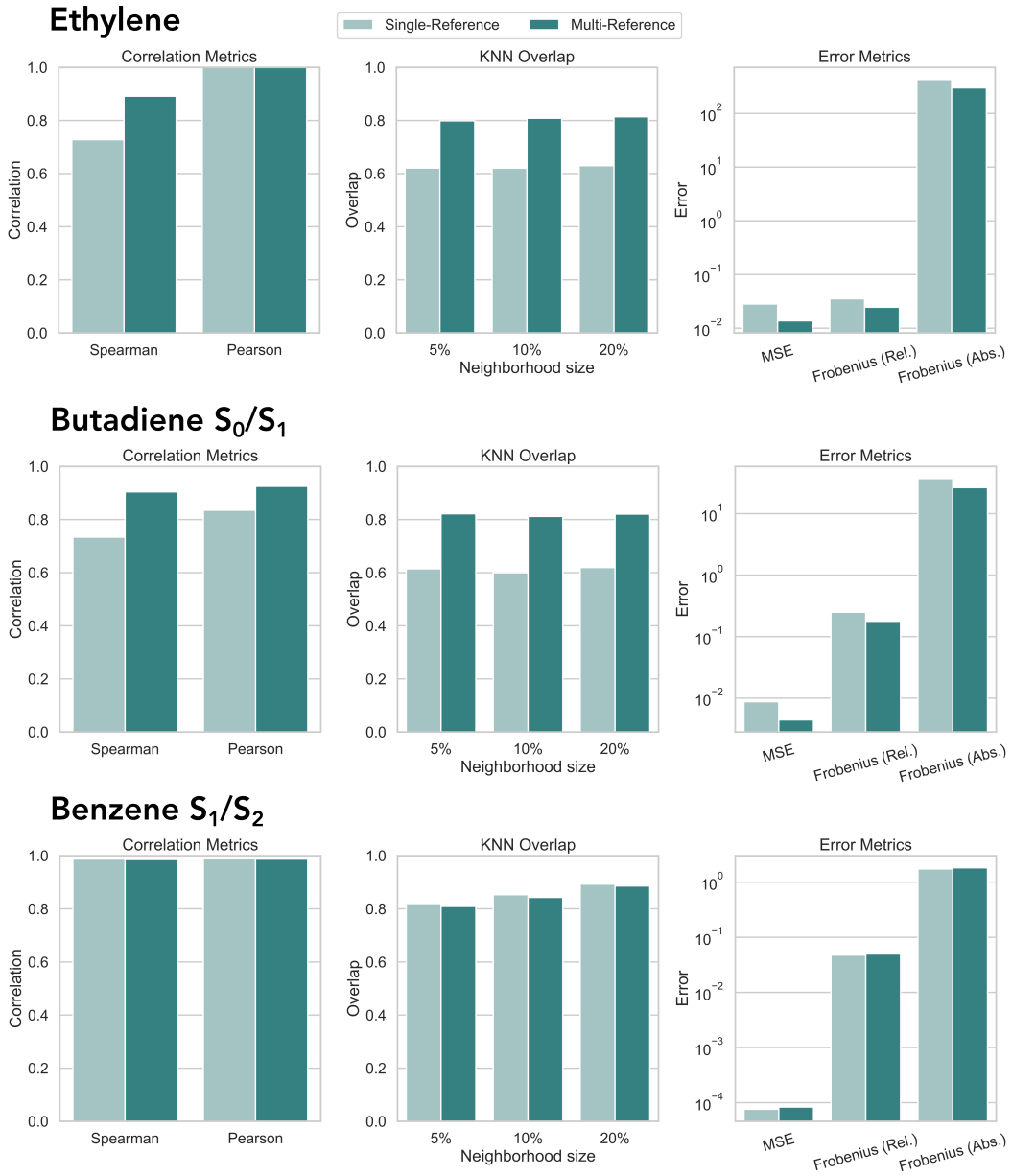

Figure 1 Comparison of single- and multi-reference alignment strategies across representative systems. Left panels: Global agreement with the optimal pairwise distance matrix $D^{(\mathrm{opt})}$, quantified by Spearman and Pearson correlation coefficients. Central panels: Local structure preservation, measured as the overlap between k-nearest-neighbor sets induced by $D^{(m)}$ and $D^{(\mathrm{opt})}$ at multiple neighborhood sizes. Right panels: Global error metrics, including the Frobenius norm (relative and absolute) and mean squared error (MSE) between $D^{(m)}$ and $D^{(\mathrm{opt})}$. Across ethylene and butadiene systems, the multi-reference strategy consistently improves both global agreement and local neighborhood preservation, whereas for benzene the two approaches yield comparable performance, with single-reference providing a sufficiently accurate description.

2 Butadiene S_0/S_1 MECIs

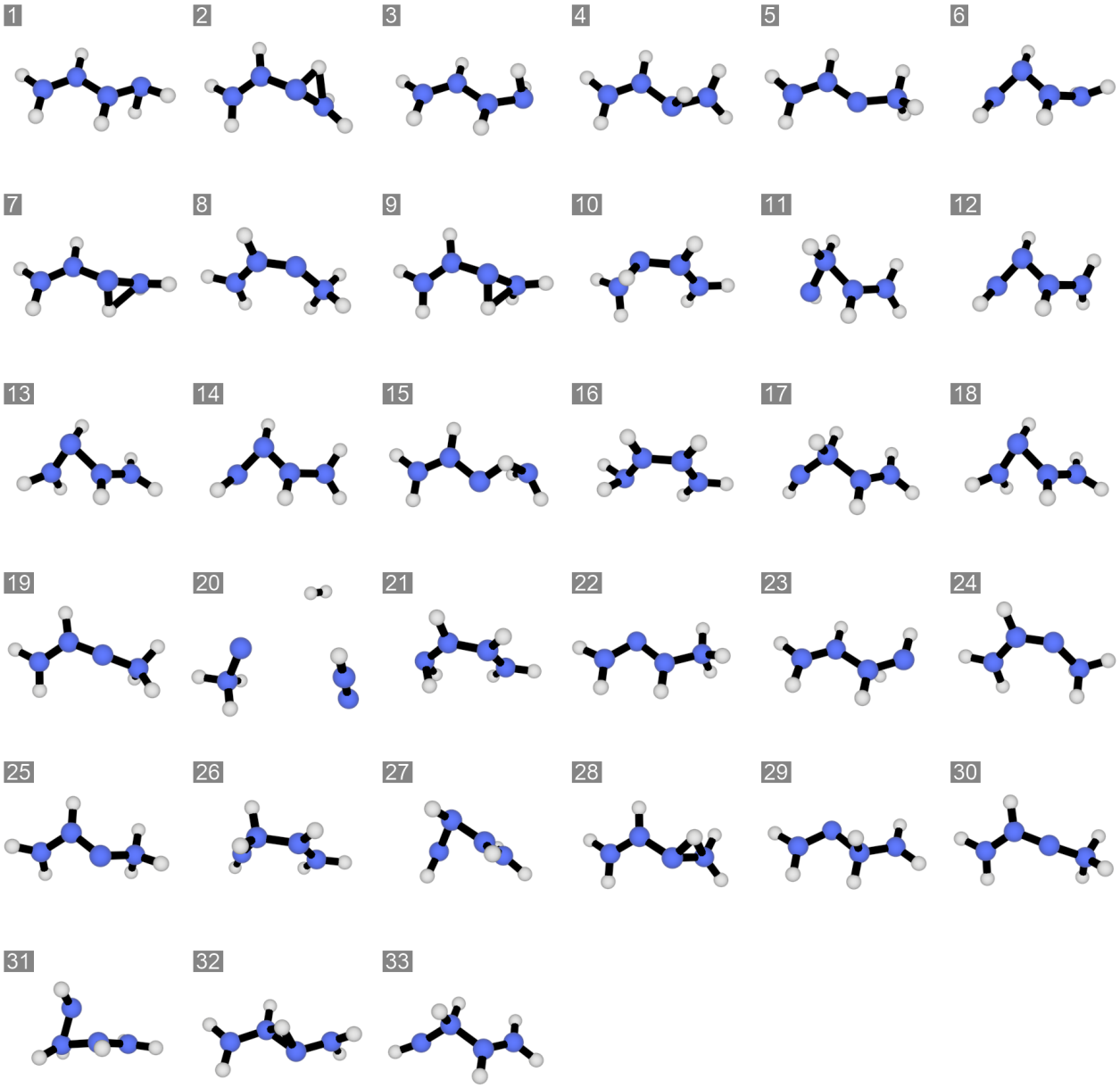

Figure 2 3D structure of the 33 MECIs identified from the butadiene S_0/S_1 seam sampling. Bonds are drawn based on the default distance cutoff implemented in RDKit.

MECI	ΔE	Pop.	MECI	ΔE	Pop.	MECI	ΔE	Pop.
1	-3.0	37.3%	12	-2.2	0.8%	23	-2.2	0.4%
2	-2.3	0.1%	13	-2.2	0.2%	24	-4.4	0.1%
3	-2.8	35.6%	14	-2.1	0.1%	25	-4.2	0.2%
4	-2.4	8.2%	15	-3.2	0.1%	26	-3.5	0.1%
5	-4.5	1.5%	16	-3.3	0.4%	27	-3.0	0.2%
6	-2.6	3.0%	17	-3.6	0.8%	28	-2.3	0.1%
7	-2.3	5.1%	18	-2.6	0.4%	29	-2.4	0.1%
8	-4.4	0.3%	19	-4.2	0.6%	30	-4.2	0.1%
9	-2.4	2.3%	20	1.6	0.1%	31	-3.5	0.4%
10	-2.4	0.6%	21	-3.0	0.6%	32	-1.6	0.1%
11	-3.0	0.5%	22	-4.6	0.1%	33	-3.6	0.1%

Table 2 Energy gaps relative to the Franck–Condon point and fractions of spawning points that converge to each MECI upon optimization, for the S_0/S_1 MECIs of butadiene.

3 Comprehensive embeddings results

3.1 Ethylene

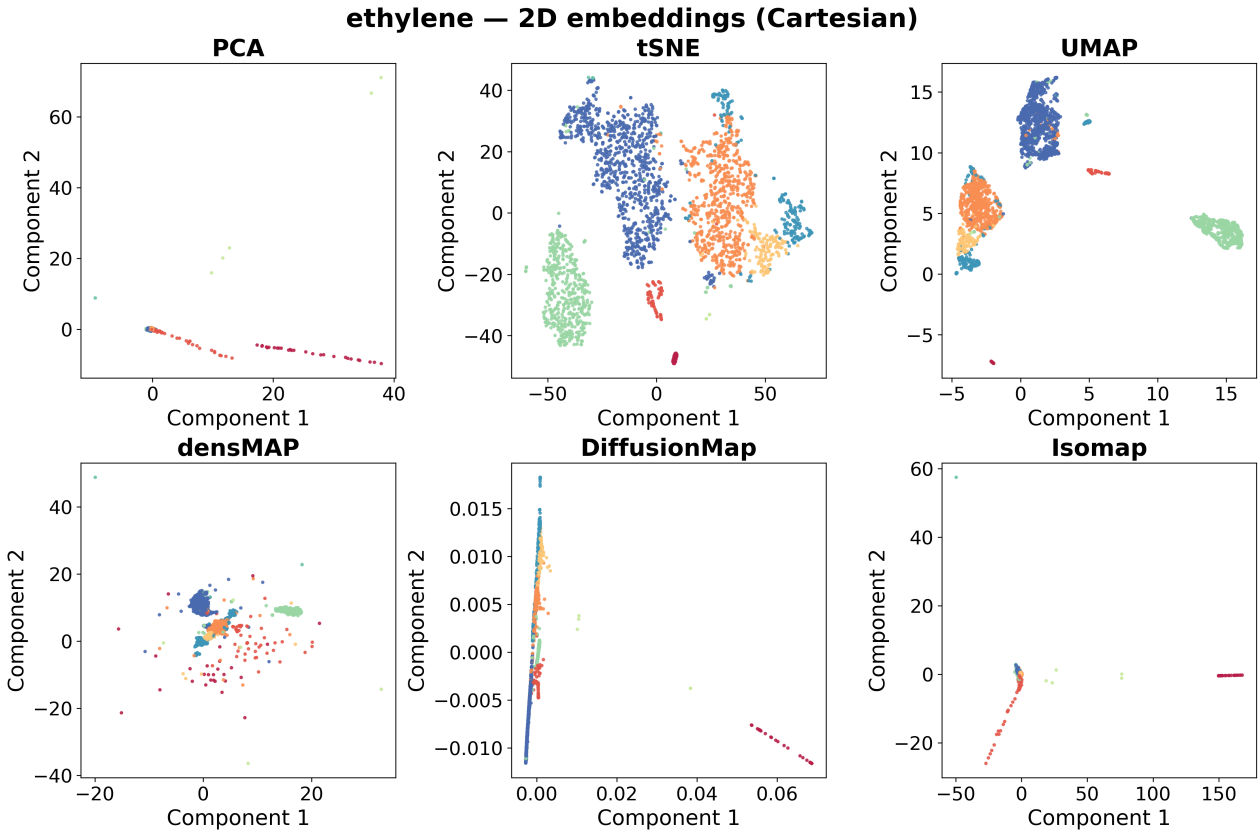

Figure 3 Comparison of dimensionality reduction methods for Cartesian representations of ethylene: two-dimensional embeddings obtained using PCA, t-SNE, UMAP, densMAP, isomap and diffusion maps, with points colored according to the MECI reached upon optimization of the corresponding SP.

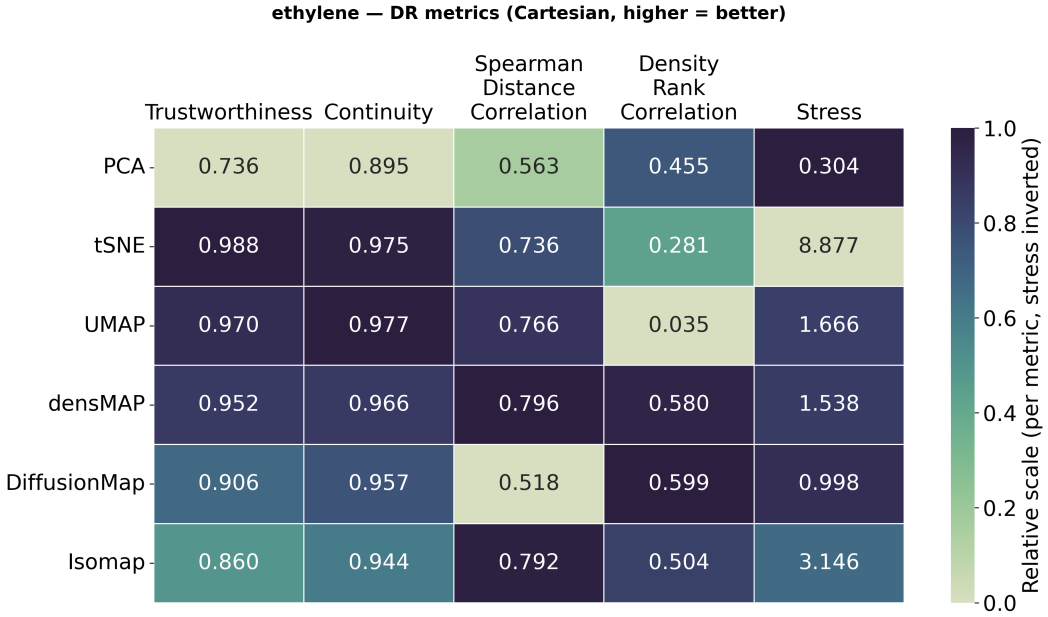

Figure 4 Comparison of dimensionality reduction methods for Cartesian representations of ethylene: quantitative metrics evaluating the preservation of local structure (trustworthiness, continuity), global geometry (Spearman distance correlation, stress), and sampling density (density rank correlation). Heatmaps are shown on a normalized scale from 0 (worst) to 1 (best) for all metrics except stress, which is inverted prior to normalization so that lower distortion corresponds to higher scores. Consequently, darker colors indicate better performance, while lighter colors indicate poorer preservation.

3.2 Butadiene

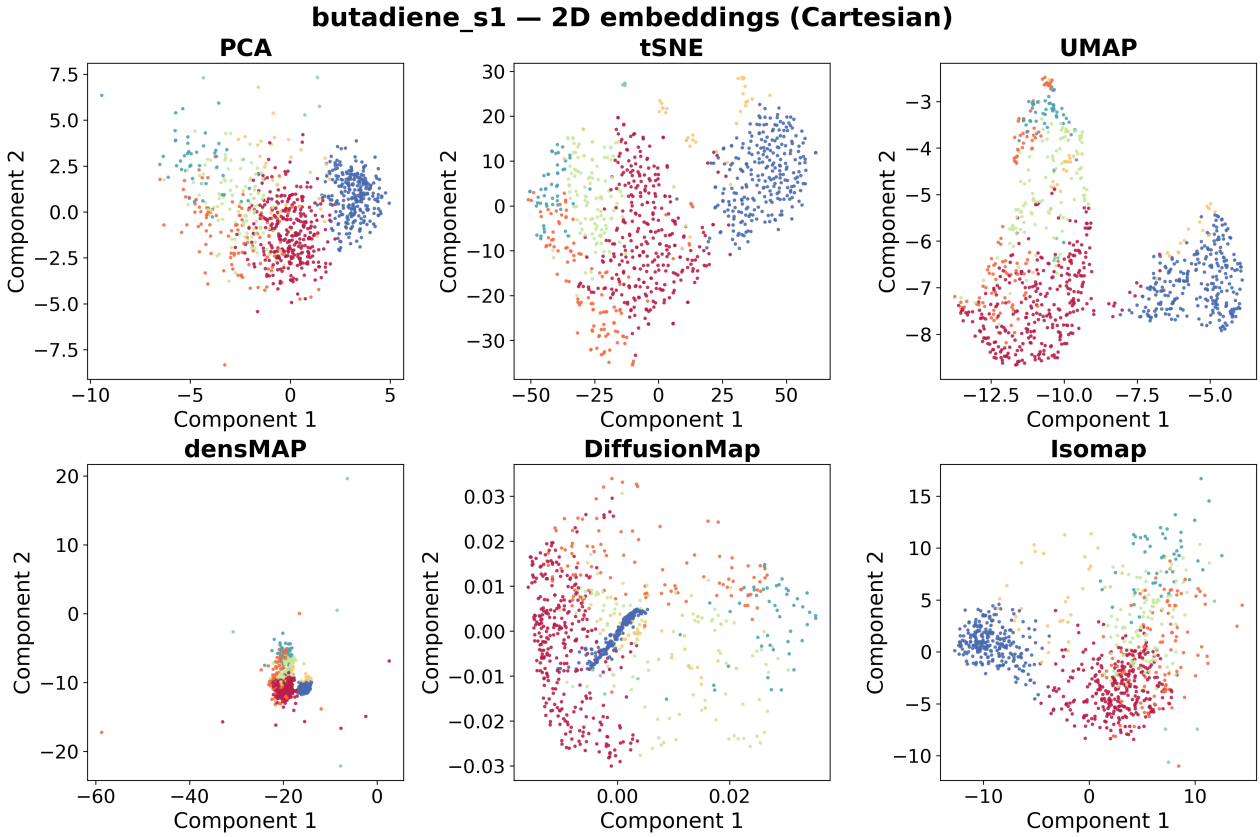

Figure 5 Comparison of dimensionality reduction methods for Cartesian representations of butadiene S_1/S_2: two-dimensional embeddings obtained using PCA, t-SNE, UMAP, densMAP, isomap and diffusion maps, with points colored according to the MECI reached upon optimization of the corresponding SP.

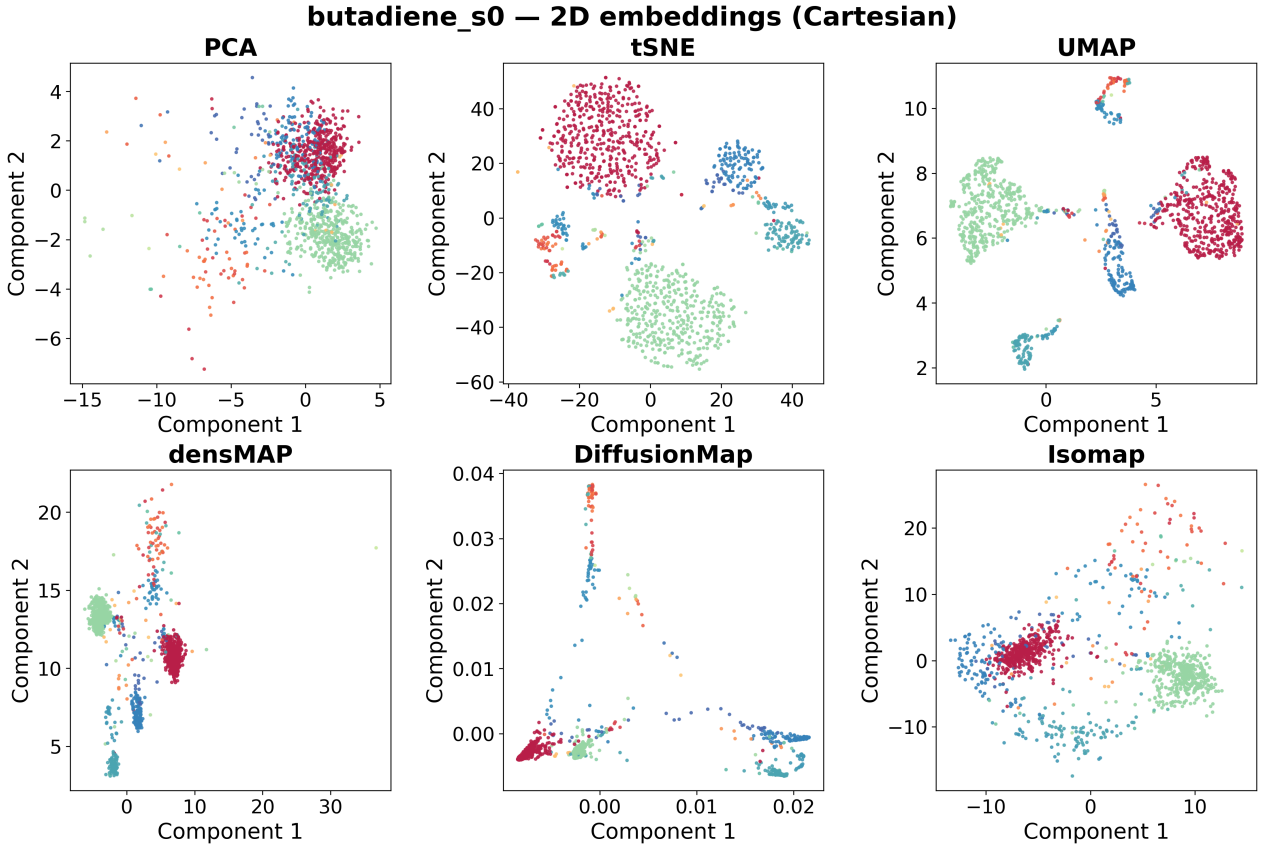

Figure 6 Comparison of dimensionality reduction methods for Cartesian representations of butadiene S_0/S_1: two-dimensional embeddings obtained using PCA, t-SNE, UMAP, densMAP, isomap and diffusion maps, with points colored according to the MECI reached upon optimization of the corresponding SP.

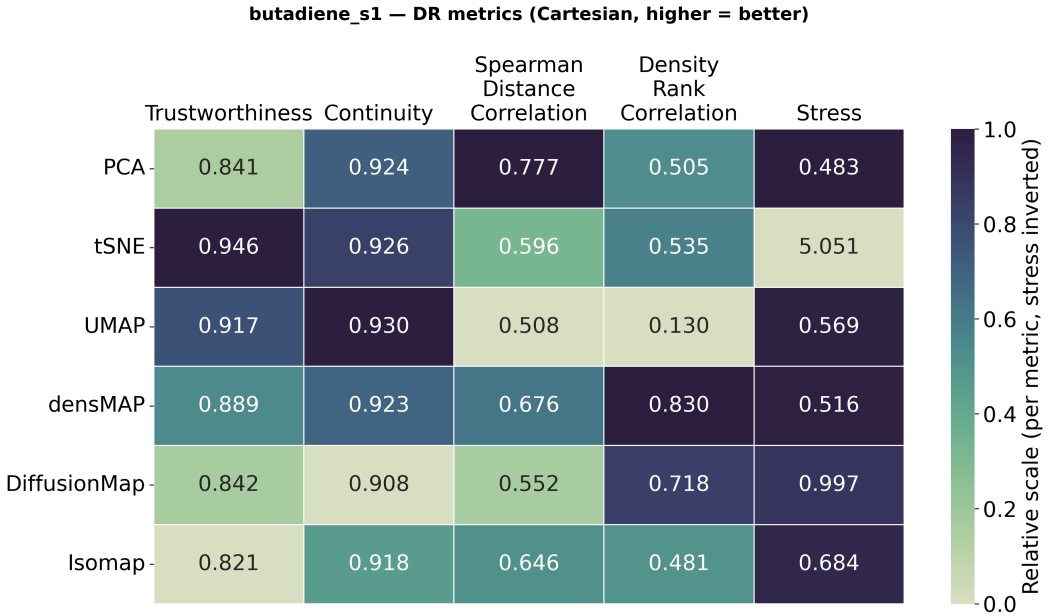

Figure 7 Comparison of dimensionality reduction methods for Cartesian representations of butadiene S_1/S_2: quantitative metrics evaluating the preservation of local structure (trustworthiness, continuity), global geometry (Spearman distance correlation, stress), and sampling density (density rank correlation). Heatmaps are shown on a normalized scale from 0 (worst) to 1 (best) for all metrics except stress, which is inverted prior to normalization so that lower distortion corresponds to higher scores. Consequently, darker colors indicate better performance, while lighter colors indicate poorer preservation.

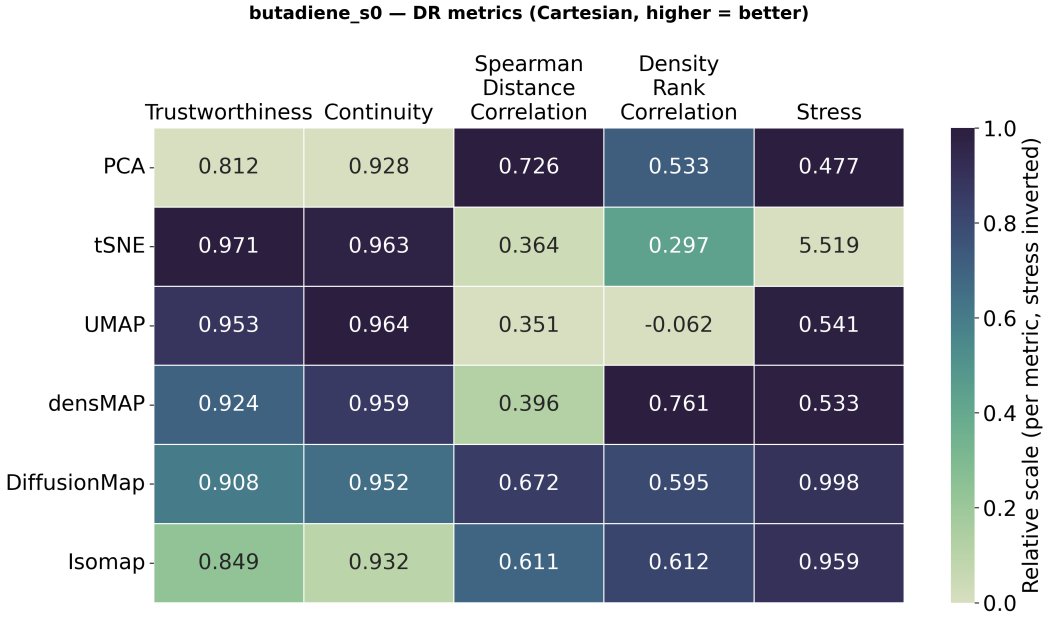

Figure 8 Comparison of dimensionality reduction methods for Cartesian representations of butadiene S_0/S_1: quantitative metrics evaluating the preservation of local structure (trustworthiness, continuity), global geometry (Spearman distance correlation, stress), and sampling density (density rank correlation). Heatmaps are shown on a normalized scale from 0 (worst) to 1 (best) for all metrics except stress, which is inverted prior to normalization so that lower distortion corresponds to higher scores. Consequently, darker colors indicate better performance, while lighter colors indicate poorer preservation.

3.3 Benzene

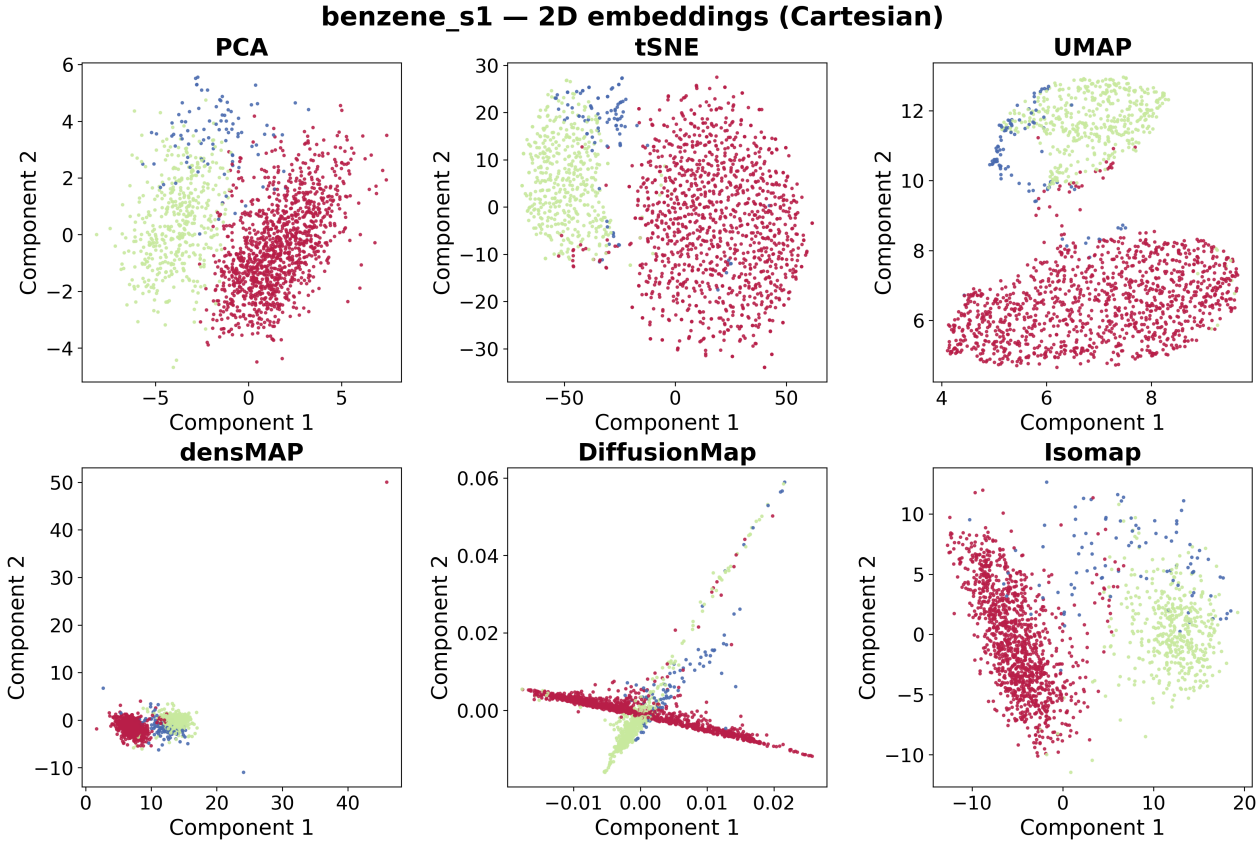

Figure 9 Comparison of dimensionality reduction methods for Cartesian representations of benzene S_1/S_2: two-dimensional embeddings obtained using PCA, t-SNE, UMAP, densMAP, isomap and diffusion maps, with points colored according to the MECI reached upon optimization of the corresponding SP.

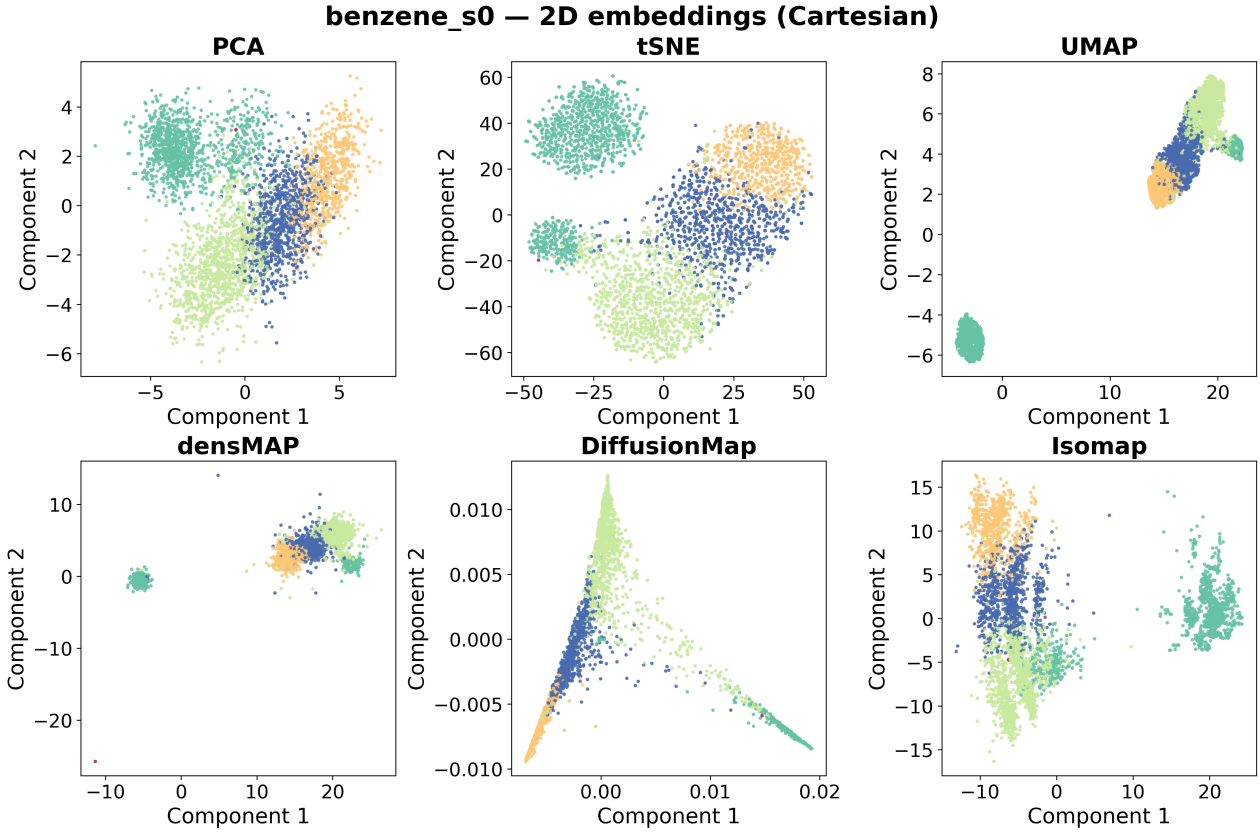

Figure 10 Comparison of dimensionality reduction methods for Cartesian representations of benzene S_0/S_1: two-dimensional embeddings obtained using PCA, t-SNE, UMAP, densMAP, isomap and diffusion maps, with points colored according to the MECI reached upon optimization of the corresponding SP.

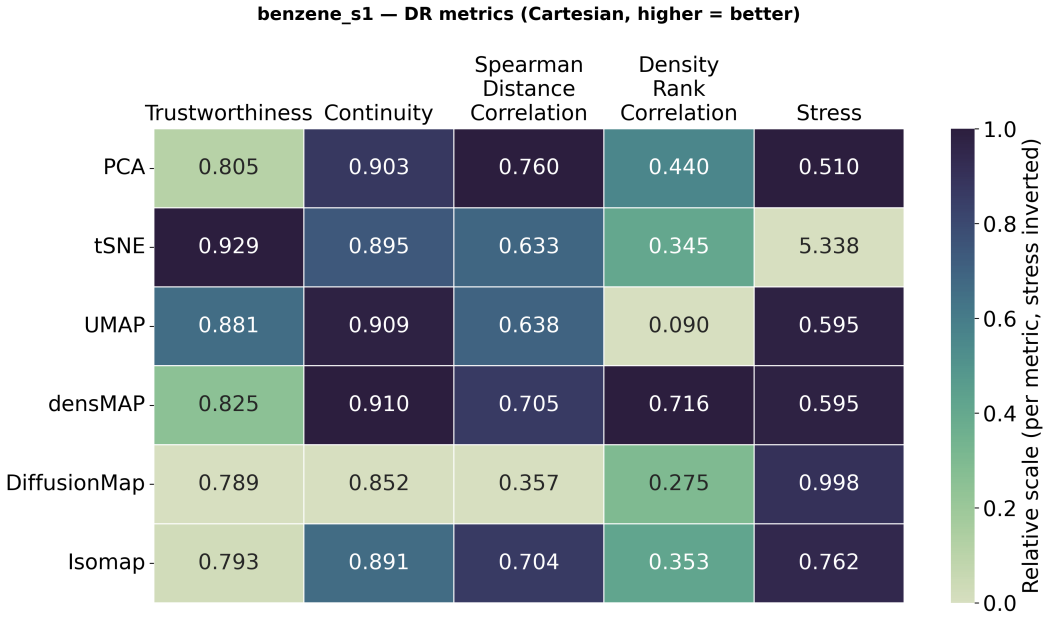

Figure 11 Comparison of dimensionality reduction methods for Cartesian representations of benzene S_1/S_2: quantitative metrics evaluating the preservation of local structure (trustworthiness, continuity), global geometry (Spearman distance correlation, stress), and sampling density (density rank correlation). Heatmaps are shown on a normalized scale from 0 (worst) to 1 (best) for all metrics except stress, which is inverted prior to normalization so that lower distortion corresponds to higher scores. Consequently, darker colors indicate better performance, while lighter colors indicate poorer preservation.

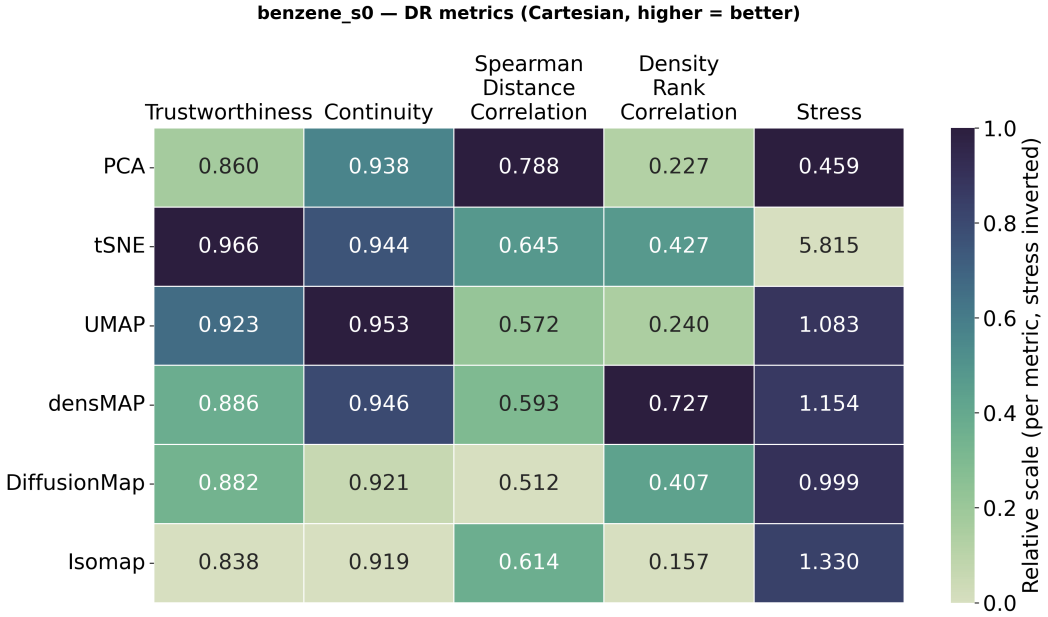

Figure 12 Comparison of dimensionality reduction methods for Cartesian representations of benzene S_0/S_1: quantitative metrics evaluating the preservation of local structure (trustworthiness, continuity), global geometry (Spearman distance correlation, stress), and sampling density (density rank correlation). Heatmaps are shown on a normalized scale from 0 (worst) to 1 (best) for all metrics except stress, which is inverted prior to normalization so that lower distortion corresponds to higher scores. Consequently, darker colors indicate better performance, while lighter colors indicate poorer preservation.

4 Detailed Basin Analysis

Basin ID	N_k	# MECIs in basin	$\langle d_m \rangle$
1	467	2	3.922
2	12	0	-
3	982	2	2.149
4	24	0	-
5	1039	4	4.032
6	33	1	1.499

Table 3 Basin analysis of ethylene ($\delta = 10$ and $\varepsilon = 0.005$).

Basin ID	N_k	MECIs in basin	$\langle d_m \rangle$
1	106	5	1.387
2	435	2	0.814
3	167	13	5.792
4	538	13	5.805

Table 4 Basin analysis of butadiene S_0 ($\delta = 5$, $\varepsilon = 0.1$).

Basin ID	N_k	# MECIs in basin	$\langle d_m \rangle$
1	315	4	18.042
2	566	4	2.370

Table 5 Basin analysis of butadiene S_1/S_2 ($\delta = 5$ and $\varepsilon = 0.01$).

Basin ID	N_k	# MECIs in basin	$\langle d_m \rangle$
1	807	1	0.163
2	347	0	-
3	1046	2	4.381
4	849	1	0.174
5	608	1	0.500

Table 6 Basin analysis of benzene S_0/S_1 ($\delta = 5$ and $\varepsilon = 0.005$).

Basin ID	N_k	# MECIs in basin	$\langle d_m \rangle$
1	535	2	1.962
2	1182	1	0.430

Table 7 Basin analysis of benzene S_1/S_2 ($\delta = 5$ and $\varepsilon = 0.005$).

5 Correspondence analysis

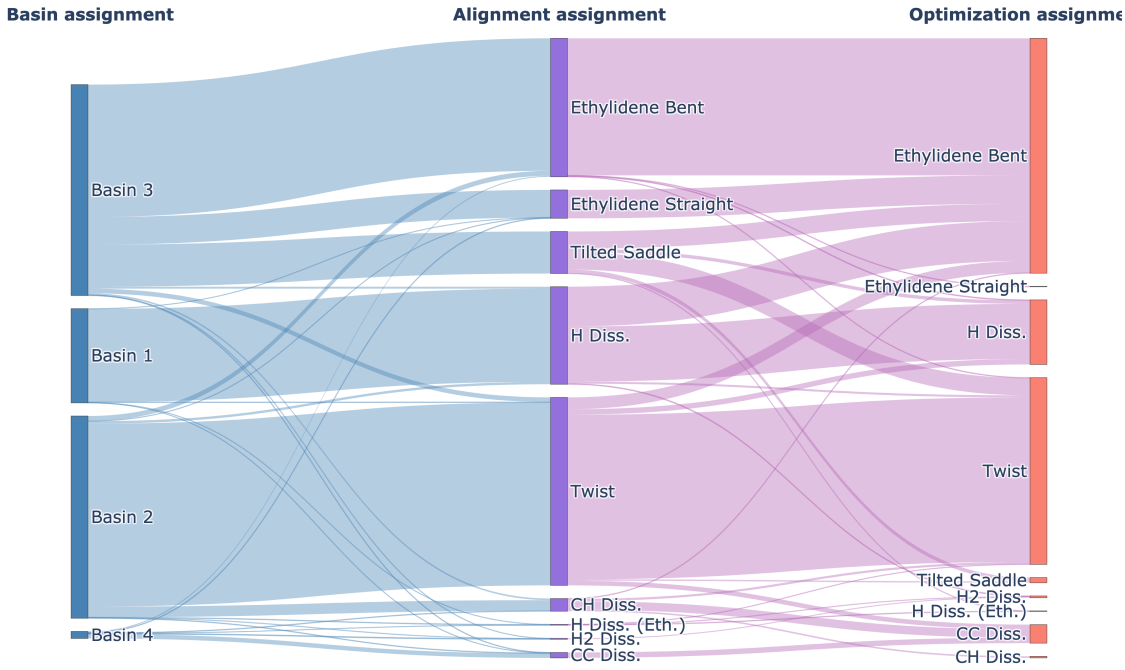

Figure 13 Three-way mapping between density-based basin assignments, alignment-based MECI labels, and optimization outcomes for ethylene. Flows connect density basins (left) to the MECI to which each SP geometry is closest in RMSD (center), and to the MECI reached upon optimization (right), with widths proportional to the number of SP geometries following each path.

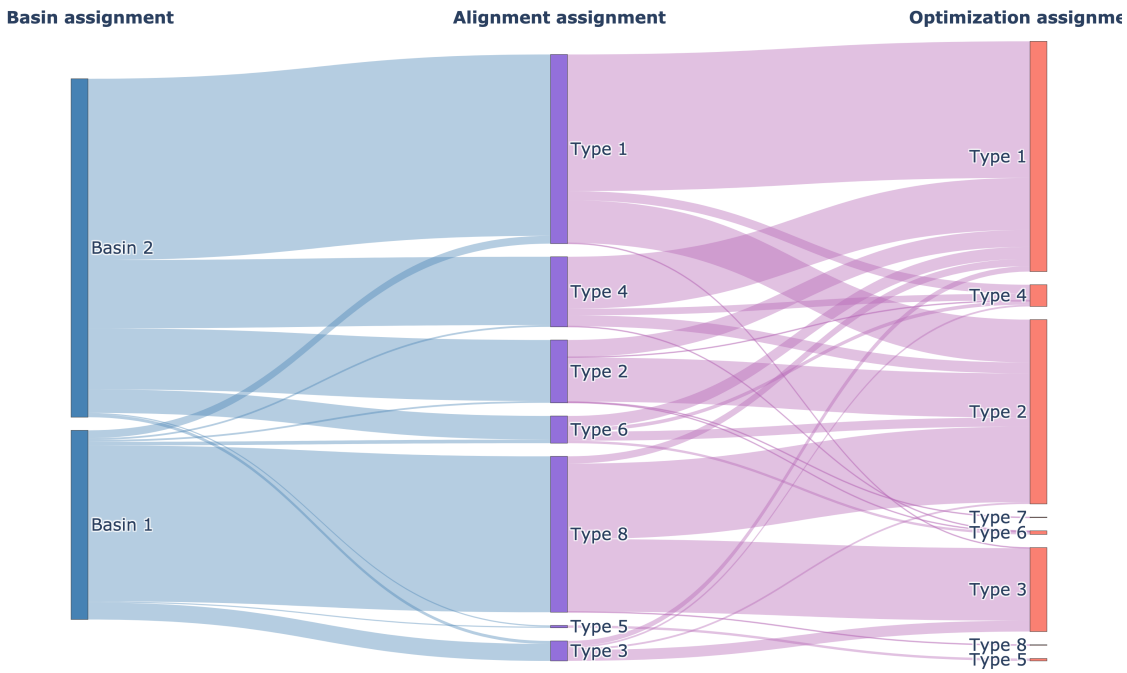

Figure 14 Three-way mapping between density-based basin assignments, alignment-based MECI labels, and optimization outcomes for butadiene S_1/S_2. Flows connect density basins (left) to the MECI to which each SP geometry is closest in RMSD (center), and to the MECI reached upon optimization (right), with widths proportional to the number of SP geometries following each path.

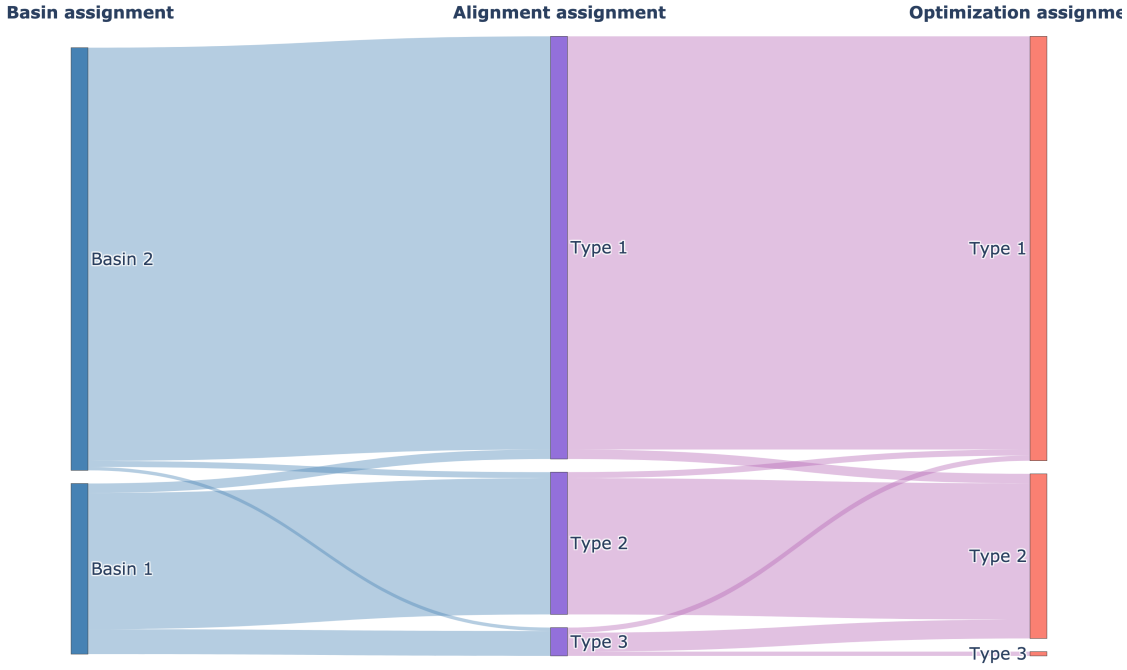

Figure 15 Three-way mapping between density-based basin assignments, alignment-based MECI labels, and optimization outcomes for benzene S_1/S_2. Flows connect density basins (left) to the MECI to which each SP geometry is closest in RMSD (center), and to the MECI reached upon optimization (right), with widths proportional to the number of SP geometries following each path.

References

[1] W. Kabsch, *Acta Crystallographica Section A*, 1976, **32**, 922–923.

[2] W. Kabsch, *Acta Crystallographica Section A*, 1978, **34**, 827–828.

[3] B. K. P. Horn, *J. Opt. Soc. Am. A*, 1987, **4**, 629–642.

[4] E. A. Coutsias, C. Seok and K. A. Dill, *Journal of Computational Chemistry*, 2004, **25**, 1849–1857.

[5] G. Landrum, P. Tosco, B. Kelley, R. Rodriguez, D. Cosgrove, R. Vianello, sriniker, P. Gedeck, G. Jones, E. Kawashima, N. Schneider, D. Nealschneider, tadhurst cdd, A. Dalke, M. Swain, B. Cole, S. Turk, A. Savelev, N. Maeder, R. Walker, A. Vaucher, M. Wójcikowski, H. Faara, I. Take, V. F. Scalfani, D. Probst, K. Ujihara, Y. Pechersky, J. Monat and J. Lehtivarjo, *rdkit/rdkit: 2025_09_6 (Q3 2025) Release (Release_2025_09_6)*, 2026.

[6] B. D. McKay and A. Piperno, *Journal of Symbolic Computation*, 2014, **60**, 94–112.

[7] H. W. Kuhn, *Naval Research Logistics Quarterly*, 1955, **2**, 83–97.