



Title	Acceleration of Reaction Space Projector Analysis Using Combinatorial Optimization : Application to Organic Chemical Reactions
Author(s)	Qu, Lihao; Tsutsumi, Takuro; Ono, Yuriko et al.
Citation	Journal of chemical theory and computation, 20(24), 10931-10941 https://doi.org/10.1021/acs.jctc.4c01072
Issue Date	2024-12-24
Doc URL	https://hdl.handle.net/2115/97336
Rights	This document is the Accepted Manuscript version of a Published Work that appeared in final form in Journal of chemical theory and computation, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://pubs.acs.org/articlesonrequest/AOR-N8URDSR8HDEDP8HSUDTC .
Type	journal article
File Information	ReSPer-A0_MS_final2_clean.pdf



Acceleration of Reaction Space Projector Analysis Using Combinatorial Optimization: Application to Organic Chemical Reactions

Lihao Qu¹, Takuro Tsutsumi^{2*}, Yuriko Ono³, and Tetsuya Taketsugu^{2,3*}

¹ Graduate School of Chemical Sciences and Engineering, Hokkaido University, Sapporo 060-0810, Japan

² Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan

³ Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo 001-0021, Japan

Abstract

In recent years, automated reaction path search methods have established the concept of a reaction route network. The Reaction Space Projector (ReSPer) visualizes the potential energy hypersurface into a lower-dimensional subspace using principal coordinates. The main time-consuming process in ReSPer is calculating the structural distance matrix, making it impractical for complex organic reaction route networks. We implemented the Alternate Optimization (AO) algorithm, one of the combinatorial optimizations, in ReSPer to reduce computational costs. Evaluations using gold clusters and the Au₅ reaction route network showed that ReSPer-AO accurately computes distances with lower computational costs.

Applying ReSPer-AO to the C₅H₈O reaction route network clarified dynamic conformation changes in its potential energy landscape. The ReSPer-AO method enables analysis of chemical reactions and dynamic conformations in a low-dimensional reaction space that accurately represents hydrocarbon reaction route networks.

Keywords: Combinatorial Optimization, Dimensionality Reduction, Reaction Route Network, Claisen Rearrangement Reaction

*Corresponding authors: Takuro Tsutsumi (t.tsutsumi@sci.hokudai.ac.jp) and
Tetsuya Taketsugu (take@sci.hokudai.ac.jp)

1. Introduction

Chemical reactions can be described by a series of elementary reaction processes that form a network of chemical reaction paths on the potential energy surface (PES). The intrinsic reaction coordinate (IRC)^{1,2} is the minimum energy path starting from a transition state structure (TS) in mass-weighted Cartesian coordinates and represents the structural changes along an elementary reaction process. Over the past two decades, several automated reaction path search methods have been developed.^{3–8} The artificial force induced reaction (AFIR)^{8,9} method has allowed for the systematic exploration of IRC paths, establishing the concept of a reaction route network made up of many reaction paths. The AFIR method has clarified reaction mechanisms in both organic chemistry^{10,11} and more complex areas such as surface chemistry.^{12,13} Recently, by combining AFIR with reaction kinetics, the quantum chemical retrosynthesis analysis (QCaRA) has been proposed.¹⁴ The reaction route network is also used as a reference for analyzing *ab initio* molecular dynamics (AIMD) simulations.¹⁵

Since the reaction route network exists in the high-dimensional manifold of the PES, visualizing the network structure in lower-dimensional space and extracting features that characterize chemical reactions are challenging tasks. The connectivity of the reaction route network can be visualized using versatile tools such as Gephi¹⁶ and Cytoscape¹⁷; however, it is still challenging to characterize chemical reactions in the high-dimensional manifold. Recently, the mathematical tool of persistent homology analysis was applied to the graph structure of the reaction route network to extract feature information from chemical reaction networks.¹⁸ Dimensionality reduction is another practical approach for extracting features from the high-dimensional manifold. Recently, various dimensionality reduction methods

have been applied to chemical reaction analyses, including principal component analysis (PCA),^{19–21} multidimensional scaling (MDS),^{22–25} locally linear embedding (LLE),²⁶ isometric feature mapping (Isomap)^{22,26,27}, and uniform manifold approximation and projection (UMAP).²¹

We have developed the reaction space projector (ReSPer) method^{23,25,28,29} based on the classical MDS (CMDS) method, which is one of the linear dimensionality reduction techniques.³⁰ The ReSPer method allows us to construct a low-dimensional reaction space that preserves the structural distance relationships in the full-dimensional PES as accurately as possible.²³ ReSPer can also project out-of-sample data, such as trajectories obtained from molecular dynamics simulation, into the pre-constructed reaction space.²⁵ By applying the ReSPer method to isomerization reactions in organic compounds and small gold clusters, we have successfully gained chemical insights into reaction mechanisms and established guidelines for controlling chemical reactions.^{23,25,29} Very recently, the ReSPer method has been extended to the analysis of photochemistry.³¹

However, the computational cost of ReSPer analysis has limited its application to practical organic reactions with a large number of reaction pathways. The ReSPer method requires inputting an Euclidean distance matrix, where each non-diagonal element represents the Euclidean distance between two molecular structures in the reference set. To accurately evaluate these distances, it is necessary to eliminate contributions from translation, rotation, and atom mapping for each pair of structures. Atom mapping refers to the correspondence of atom indices between two molecular structures, resulting in nuclear permutation inversion (NPI) isomers, which are generated through permutations of identical nuclei and spatial

inversion operations. While the contributions of translation and rotation can be efficiently eliminated using the Kabsch algorithm,^{32,33} optimizing atom mapping is time-consuming as it is a combinatorial optimization problem. To address the atom mapping challenge, we previously developed the NPI-search method. This method exhaustively searches for the minimum Euclidean distance between a reference structure and all possible NPI isomers of a target structure.¹⁵ Although the NPI-search method provides an accurate Euclidean distance between two molecular structures, it is computationally expensive due to its brute-force approach. This is particularly challenging for hydrocarbon compounds with a large number of NPI isomers.

Efficient algorithms for solving the atom mapping problem in calculating Euclidean distances in chemistry have been proposed, including iterative approaches,^{34–40} branch-and-bound methods,^{37,40} and Monte Carlo simulations.⁴¹ Recently, Fukutani *et al.* developed the generalized root mean square deviation (G-RMSD) method to efficiently align the orientations of two molecular structures using the alternate optimization (AO) technique, which iteratively optimizes rotational degrees of freedom and atom mappings between structures.³⁸ Very recently, the G-RMSD method was improved to find exact solutions for molecular similarity.⁴⁰ Wales *et al.* also proposed the optimal alignment method called PERMDIST, which successfully explored physically realistic interpolation paths connecting a pair of double-end structures.³⁶ This approach has many similarities with G-RMSD related to the alternative optimization algorithm. In 2017, the globally optimal PERMDIST (Go-PERMDIST) that deterministically provides the global minimum RMSD was proposed.³⁷

In this study, we implement the AO-based Euclidean distance calculation method in ReSPer and validate the correctness and efficiency of the ReSPer-AO method compared to the NPI-search-based ReSPer method. Additionally, we analyze the large-scale IRC network of C₅H₈O, involving the Claisen rearrangement reaction of allyl vinyl ether,¹⁰ using the ReSPer-AO method.

2. Methodology

2.1. Reaction Space Projector: ReSPer

Here, we introduce the ReSPer procedure. ReSPer utilizes the CMDS method for dimensionality reduction to capture the relative structural distance relationships of reference structures in high-dimensional configurational space.^{23,25,28,29} Firstly, the distance matrix **D** for all pairs of reference structures needs to be computed. The element (**D**)_{ab} is defined as the Euclidean distance between molecular structures *a* and *b* in mass-weighted Cartesian coordinates, as follows:

$$(\mathbf{D})_{ab} = d_{ab} = \sqrt{\sum_{k=1}^N \|\xi_k^{(a)} - \xi_k^{(b)}\|^2} \quad (1)$$

where *N* is the number of atoms, and $\xi_k^{(a)}$ is the *k*th atomic coordinate of molecule structure *a* in the mass-weighted Cartesian coordinates. To obtain the minimum Euclidean distance, it is necessary to align the orientations of the two molecular structures as closely as possible. This alignment procedure will be explained in the next subsection.

Secondly, the CMDS method is applied to the Euclidean distance matrix. The steps of CMDS are as follows: (i) Apply the Young-Householder transformation⁴² to the squared

distance matrix $\mathbf{D}^{(2)}$, which is composed of matrix elements of squared distance d_{ab}^2 , to obtain the inner product matrix $\mathbf{Q}$. (ii) Diagonalize $\mathbf{Q}$ to obtain eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and eigenvectors $\{\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_n\}$, where n is the number of reference structures. (iii) Select the first p eigenvalues and corresponding eigenvectors in descending order, and define the principal coordinates (PCos) as follows: $\mathbf{Y}_i = \mathbf{L}_i \sqrt{\lambda_i}$ ($1 \leq i \leq p$). Note that the maximum number of p is $3N - 3$, including vibrational and rotational degrees of freedom. These PCos can be used to construct a p -dimensional reaction subspace. The eigenvalue λ_i corresponds to the representability of $\mathbf{Y}_i$ determining the low-dimensional reaction space.³⁰ Thus, the proportion of variance Λ_i and the cumulated proportion Λ are defined as follows:

$$\Lambda_i = \frac{\lambda_i}{\sum_j^m \lambda_j} \quad (2)$$

$$\Lambda = \sum_i^p \Lambda_i = \frac{\sum_j^p \lambda_j}{\sum_j^m \lambda_j} \quad (3)$$

where m represents the number of positive eigenvalues. If necessary, ReSPer can project out-of-sample data, such as a trajectory, onto the predefined low-dimensional reaction space defined by PCos using the out-of-sample extended CMDS method. The detailed procedure is summarized in the literature.^{23,25,28,29}

So far, ReSPer has successfully established an approach for considering dynamical reaction processes in organic compounds and small gold clusters by referring to the IRC path and the IRC network. This approach has revealed IRC-jump processes, where a molecule jumps from one IRC to another without crossing a transition state structure. The applications of ReSPer have been reported in our previous papers.^{23,25,28,29,31}

2.2. Alternate Optimization Method

To evaluate the minimum Euclidean distance, it is necessary to align the coordinate axes of two molecular structures as closely as possible. For a given pair of structures, this involves eliminating translational and rotational contributions and optimizing the atom mapping. The translational contribution can be easily eliminated by moving the center of mass of the molecular structures to the coordinate origin. The rotational contribution, when considered independently, can be optimized using the Kabsch method.^{32,33} However, since the atom mapping problem is dependent on rotational optimization, both should be optimized iteratively.

So far, we have developed the NPI-search method, a brute-force approach to obtain the minimum Euclidean distance by exhaustively searching through a pair of a reference structure and all NPI isomers of a target structure.¹⁵ In the NPI-search method, rotational optimization is performed after the atom mapping optimization, and then the Euclidean distance is computed. The computational complexity of the NPI-search method is approximately $O(\prod_k^M N_k!)$, where M is the number of identical nuclei, and N_k is the number of the k th identical nuclei. This high computational complexity makes applying the NPI-search method to hydrocarbon compounds challenging. In contrast, the AO method is a combinatorial optimization technique that iteratively optimizes the rotational contribution and the atom mapping problem. This approach greatly reduces computational complexity compared to the brute-force NPI-search method.

Here, the procedure of the AO method is summarized. Let $\xi^{(a)}$ be the mass-weighted Cartesian coordinate of a molecular structure a , with the center of mass already moved to the

coordinate origin. The Euclidean distance in the mass-weighted Cartesian coordinate considering the rotational and atom mapping treatments can be rewritten as follows:

$$d_{ab} = \sqrt{\sum_{k=1}^N \left\| \xi_k^{(a)} - \mathbf{R} \xi_{\sigma(k)}^{(b)} \right\|^2} \quad (4)$$

where N is the number of atoms, $\mathbf{R}$ is a rotation matrix, and $\sigma(k)$ is a permutation function that replaces the k th atom with another identical atom. To minimize d_{ab} , it is necessary to find appropriate $\mathbf{R}$ and $\sigma(k)$. Expanding the Euclidean distance, d_{ab} , we get:

$$\begin{aligned} d_{ab}^2 &= \sum_{k=1}^N \left\| \xi_k^{(a)} - \mathbf{R} \xi_{\sigma(k)}^{(b)} \right\|^2 \\ &= \sum_{k=1}^N \left(\xi_k^{(a)T} \xi_k^{(a)} - \xi_k^{(a)T} \mathbf{R} \xi_{\sigma(k)}^{(b)} - \xi_k^{(b)T} \mathbf{R}^T \xi_k^{(a)} + \xi_{\sigma(k)}^{(b)T} \xi_{\sigma(k)}^{(b)} \right) \\ &= \sum_{k=1}^N \left\| \xi_k^{(a)} \right\|^2 + \left\| \xi_{\sigma(k)}^{(b)} \right\|^2 - 2 \text{trace} \left(\mathbf{R} \sum_{k=1}^N \xi_k^{(a)T} \xi_{\sigma(k)}^{(b)} \right) \\ &= \sum_{k=1}^N \left\| \xi_k^{(a)} \right\|^2 + \left\| \xi_{\sigma(k)}^{(b)} \right\|^2 - 2 \text{trace}(\mathbf{R} \mathbf{B} \mathbf{P} \mathbf{A}^T). \end{aligned} \quad (5)$$

Here, $\mathbf{A}$ and $\mathbf{B}$ are the mass-weighted Cartesian coordinate matrices of reference structures a and b , respectively, and $\mathbf{P}$ is the permutation matrix for replacing identical atoms. The minimization problem can be transformed into the maximization problem of $\text{trace}(\mathbf{R} \mathbf{B} \mathbf{P} \mathbf{A}^T)$. In the AO procedure, the minimum d_{ab} is found by iteratively optimizing the rotational matrix $\mathbf{R}$ and the permutation matrix $\mathbf{P}$. First, $\mathbf{R}$ is optimized against a fixed $\mathbf{P}$, and then $\mathbf{P}$ is optimized against a fixed optimal $\mathbf{R}$.

2.3. ReSPer-AO Method for Calculating Mass-weighted Euclidean Distance

The G-RMSD method was proposed to calculate the optimal root-mean-square deviation (RMSD) between two molecular structures by introducing the AO method.³⁸ The

concept of AO involves the iterative optimization of rotation and atom mapping (permutation) contributions to achieve the optimal geometric orientation. In the G-RMSD method, rotational and permutation optimizations are conducted using the Kabsch method,^{32,33} which is also used in the NPI-search method, and the Hungarian algorithm^{43,44} a discrete optimization method, respectively. However, the results of the AO method strongly depend on the initial orientation of the molecule pair because AO is a local optimization technique. To address this dependency on initial orientations, spatially uniform initial rotation matrices were introduced in G-RMSD.³⁸ In this study, we employed the strategy of using initial rotation matrices to determine the mass-weighted Euclidean distance between two molecular structures based on the combination of ReSPer and AO methods.

The initial rotation matrices are generated with reference to the rotational symmetry of the icosahedron, a polyhedron with 60 rotation axes, the most numerous in three-dimensional space.³⁸ However, when considering the inversion operation related to chiral isomers, 60 initial rotation matrices may be insufficient. In the ReSPer-AO method, for chiral isomers, 60 rotation-inversion matrices are generated by applying the inversion operation to 60 rotation matrices, and in total, 120 initial rotation (-inversion) matrices will be used. Additionally, the Kabsch algorithm can consider the inversion operation during rotational optimization, which is optionally conducted in our ReSPer program. Therefore, we implemented an option to select whether to consider the inversion operation depending on the purpose of the analysis in the ReSPer-AO method.

Figure 1 illustrates the scheme for calculating the mass-weighted Euclidean distance using the ReSPer-AO method, which is inspired by the G-RMSD method.³⁸ Here, the inner

loop (highlighted in green) corresponds to the AO scheme, while the outer loop (highlighted in blue) scans initial rotation matrices $\{\mathbf{R}_i^0\}$ that provide different initial orientations of two molecular structures. The challenge for ReSPer-AO is to efficiently determine the optimal rotation matrix $\mathbf{R}$ and the optimal permutation matrix $\mathbf{P}$ such that the objective function $\text{trace}(\mathbf{R}\mathbf{B}\mathbf{P}\mathbf{A}^T)$ is maximized. Let j be set to 1 and the initial rotation matrices be set to $\{\mathbf{R}_i^{j-1=0}\} = (\mathbf{R}_1^0, \mathbf{R}_2^0, \dots, \mathbf{R}_{i=60 \text{ or } 120}^0)$. Notes that i and j correspond to the outer loop and the inner loop in **Fig. 1**, respectively. The ReSPer-AO method is first given by mass-weighted Cartesian coordinate matrices corresponding to two molecular structures ($\mathbf{A}$, $\mathbf{B}$) and initial rotation matrices $\{\mathbf{R}_i^{j-1=0}\}$, then explores the optimal mass-weighted Euclidean distance as follows:

- 1) Set i and j to 1 and F_{best} to be 0.
- 2) Get an initial rotation matrix $\mathbf{R}_i^{j-1}$.
- 3) Optimize the permutation matrix $\mathbf{P}_i^j$ under the fixed rotation matrix $\mathbf{R}_i^{j-1}$ to maximize the objective function, $\text{trace}(\mathbf{R}_i^{j-1}\mathbf{B}\mathbf{P}_i^j\mathbf{A}^T)$.
- 4) Optimize the rotation matrix $\mathbf{R}_i^j$ under the fixed permutation matrix $\mathbf{P}_i^j$ to maximize $\text{trace}(\mathbf{R}_i^j\mathbf{B}\mathbf{P}_i^j\mathbf{A}^T)$.
- 5) If $\mathbf{P}_i^j = \mathbf{P}_i^{j-1}$, save $\mathbf{P}_i^j$ and $\text{trace}(\mathbf{R}_i^j\mathbf{B}\mathbf{P}_i^j\mathbf{A}^T)$ as $\mathbf{P}_i$ and F_i , respectively, and go to step 6. Otherwise, increase j by 1 and return to step 3.
- 6) Only if $F_i > F_{\text{best}}$, set $\mathbf{P}_i$ and F_i to be $\mathbf{P}_{\text{best}}$ and F_{best} , respectively. Then, increase i by 1.
- 7) If $i > 60$ or 120 , calculate d_{ab} by using $\mathbf{P}_{\text{best}}$. Otherwise, return to step 2.

In this study, rotation optimization was performed using the Kabsch method,^{32,33} which is

similar to the NPI-search method, and permutation optimization was performed using the Jonsker-Volgenant algorithm,⁴⁴ a discrete optimization method. The inversion operation for molecular coordinates during rotation optimization can optionally be set in the ReSPer procedure. After searching through all initial rotation matrices, lists of local solutions for $\mathbf{P}_i$ and F_i are obtained. Therefore, the best permutation matrix $\mathbf{P}_{\text{best}}$, giving the optimal atom mapping, can be determined by searching for the maximum objective functions. Finally, after overlapping molecular structure $\mathbf{A}$ with the permuted structure $\mathbf{BP}_{\text{best}}$ using the Kabsch method, the mass-weighted Euclidean distance between the two mass-weighted molecular structures is obtained.

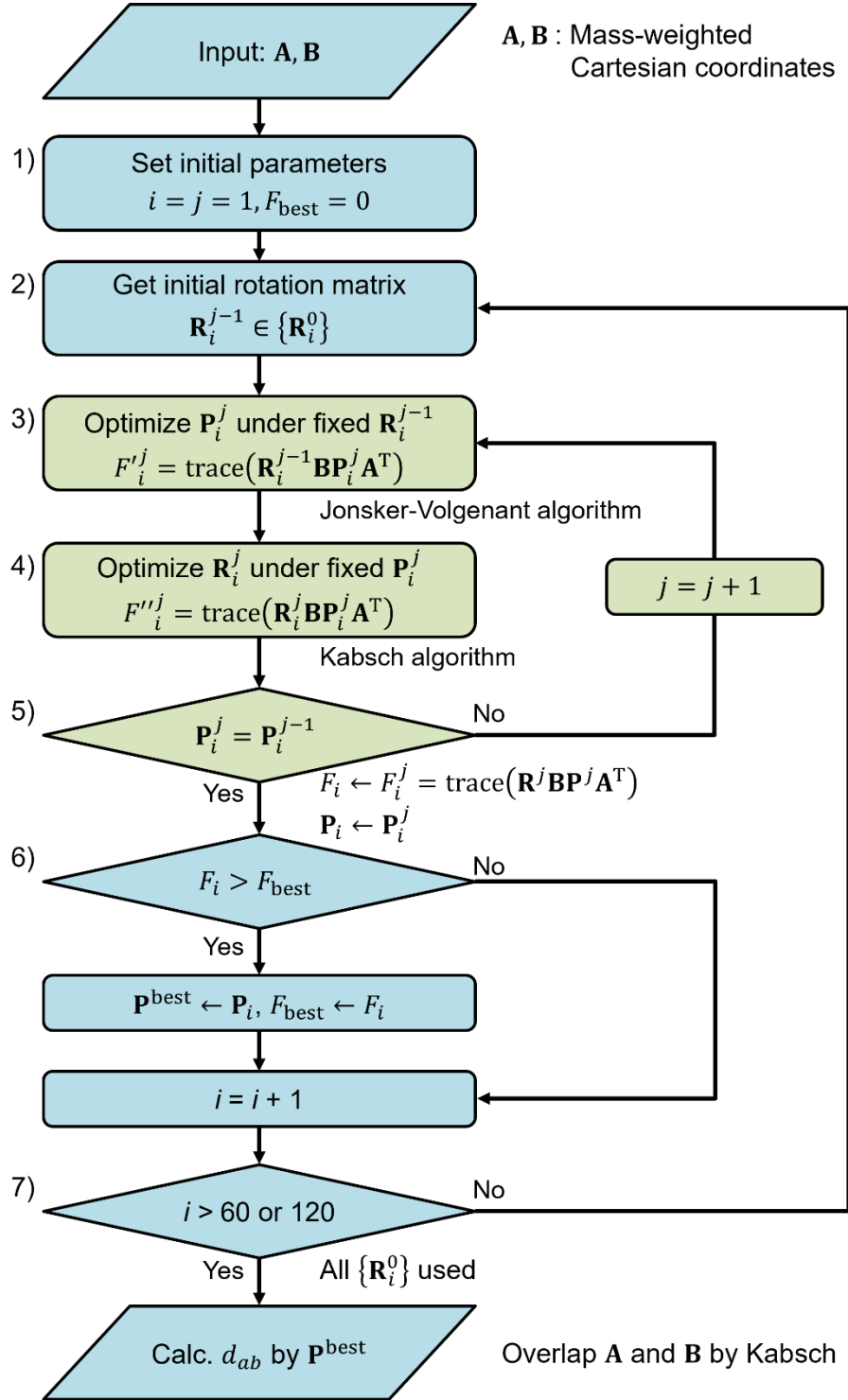


Figure 1. Scheme of the mass-weighted Euclidean distance calculation by the ReSPer-AO method.

3. Results and Discussion

3.1. Correctness of ReSPer-AO method

To compare the correctness and computational cost of the ReSPer-AO method and the NPI-search method, we conducted numerical experiments on randomly generated virtual gold clusters Au_n ($n = 3, 4, \dots, 10, 20, \dots, 100$). The virtual clusters were generated as follows.

(i) Ten seed structures $\{\mathbf{S}_i^n\}$ ($i = 1, 2, \dots, 10$) for n -atomic clusters were randomly generated using Molclus.⁴⁵ Here, $\mathbf{S}_i^n$ represents the i th molecular coordinates for an n -atomic cluster. (ii) Rotation matrices $\{\mathbf{R}_j^n\}$ and permutation matrices $\{\mathbf{P}_j^n\}$ were randomly generated for $j = 1, 2, \dots, 50$, using the Scipy⁴⁶ and Numpy⁴⁷ packages in Python, respectively. (iii) Fifty virtual structures $\mathbf{V}_{i,j}^n = \mathbf{R}_j^n \mathbf{S}_i^n \mathbf{P}_j^n$ were generated by applying $\{\mathbf{R}_j^n\}$ and $\{\mathbf{P}_j^n\}$ to each seed structure $\mathbf{S}_i^n$. These processes generate ten groups containing 50 virtual NPI structures for each n -mer.

Firstly, ten groups of structural distance matrices for n -atomic clusters were computed, and their correctness was examined. Since the distance matrix is symmetric, we only consider the upper triangular matrix without diagonal elements, totaling 1,225 elements for each n . In the NPI-search method, which provides the exact solution, the Euclidean distances between NPI isomers generated from the same seed structure (in the same group) should all be zero because these isomers have identical geometry except for rotation and permutation contributions. Therefore, the sum of 1,225 elements in the distance matrix can be used to evaluate the correctness of the ReSPer-AO method.

Table 1 shows the correctness rates of Euclidean distance calculations for virtual structures $\{\mathbf{V}_{i,j}^n\}$, averaging the rates for all n -atomic clusters. The detailed correctness rates for each n -atomic cluster are summarized in the Supporting Information (**Tables S1-4** and

Figure S1). This result indicates that the ReSPer-AO method can compute highly accurate Euclidean distances using only 60 initial rotation matrices and the Kabsch algorithm without inversion operations. Note that as the number of atoms increases, the randomly generated seed structures tend to become more spherically symmetric, making it easier to fall into local minima during permutation optimization. This phenomenon is reflected in the fact that the correctness rate will not be exactly 100%. However, given that molecular structures within complex reaction route networks, such as organic molecules, seldom exhibit spherical symmetry, the ReSPer-AO method can effectively visualize a complex reaction route network of organic compounds.

Table 1. Correctness rates for calculating Euclidean distances between generated virtual isomers with and without considering the inversion operation.

Correct answers / %	# of initial rotation matrices	
	60	120
Kabsch w/ inversion	99.9	99.9
Kabsch w/o inversion	99.9	99.9

Furthermore, we examine two scenarios related to inversion symmetry. Suppose a molecular structure **B** is generated by applying only the inversion operation on a molecular structure **A**. In the first scenario (I), the original structure **A** and the inverted structure **B** are not identified as the same ($d_{ab} \geq 0$), whereas in the second scenario (II), **A** and **B** are identified as the same ($d_{ab} = 0$). Essentially, scenario (I) can distinguish chiral isomers, while scenario (II) cannot. As mentioned in Section 2.3, the ReSPer-AO method controls

inversion symmetry during the generation of initial rotation matrices, resulting in either 60 or 120 rotation matrices. Additionally, the Kabsch algorithm can control inversion symmetry by performing rotation optimization with or without considering the inversion operation. This study also compared the correctness of the ReSPer-AO method in both scenarios (I) and (II).

To discuss inversion symmetry, we additionally generated inverted NPI isomers by applying the inversion matrix, $\text{diag}(1,1,-1)$, to the preliminary generated 50 NPI isomers for each n -atomic cluster. **Table 2** shows the correctness rate of Euclidean distance calculations between original and inverted virtual structures. This result demonstrates that accurate Euclidean distance computation between the original and inverted virtual structures requires 120 rotation matrices and consideration of the inversion operation in the Kabsch method. In some cases, 60 rotation matrices and consideration of inversion in the Kabsch method can be sufficient, but these are limited to smaller n -atomic clusters. Detailed correctness rates for each n -atomic cluster are summarized in Supporting Information (**Tables S5-8** and **Figure S1**).

For the analysis of scenario (I), the correctness rates should be zero. Thus, **Tables 1** and **2** indicate that, for scenario (I), 60 initial rotation matrices and the Kabsch algorithm without considering inversion operations are appropriate in terms of reducing computational complexity. On the other hand, it is clear that 120 initial rotation matrices and the Kabsch algorithm considering inversion operations are needed for scenario (II). Henceforth, the former condition is defined as condition (I), while the latter is defined as condition (II).

Table 2. Correctness rates for calculating Euclidean distances between generated virtual isomers, including inverted isomers, with and without considering the inversion operation.

Correct answers / %	# of initial rotation matrices	
	60	120
Kabsch w/ inversion	42.1	99.8
Kabsch w/o inversion	0.0	0.0

3.2. Computational cost of ReSPer-AO method

Next, we compared the elapsed time required to calculate the structural distance between a single pair of n -atomic gold clusters using the ReSPer-AO and the NPI-search methods. As n increases, the number of NPI isomers for n -atomic gold clusters increases rapidly, following the formula $n! \times 2$. Therefore, in this comparison, the NPI-search method, which exhaustively searches all possible NPI isomers, was only used for $n = 3 - 10$. All elapsed times for the ReSPer-AO method were estimated by averaging multiple runs. These tests were performed on a workstation with an Intel Xeon Gold 6326 2.90 GHz processor using only one core. The memory capacity was 256 GB.

Figure 2 shows the elapsed time for the NPI-search and ReSPer-AO methods. The elapsed time for the NPI-search method increases exponentially as n increases, with the computation time theoretically predicted to be $O(n!)$. For $n = 10$, the number of NPI isomers reaches 7,257,600. In this case, the NPI-search method takes approximately one day to calculate one structural distance. Therefore, it is confirmed that the NPI-search method becomes impractical for calculating the structural distance matrix as the number of NPI

isomers increases. It should be emphasized that the enormous number of NPI isomers in organic compounds makes it difficult to use the NPI-search method based on the brute-force strategy, similar to the challenge faced with n -atomic gold clusters.

On the other hand, the ReSPer-AO method is considerably faster than the NPI-search method, enabling the computation of a single distance in less than 0.3 seconds even for $n = 100$. The computational order of the ReSPer-AO method was evaluated as $O(n^{1.98})$. Technically, the bottleneck in the distance calculation for the ReSPer-AO method is the permutation optimization, which is known as a computational order of $O(n^3)$.⁴⁴ Hence, while the computational order of the ReSPer-AO method increases with n , it is expected to converge at $O(n^3)$. From these results, it is evident that the ReSPer-AO method can dramatically reduce computational costs while maintaining correctness comparable to the NPI-search method. This advantage allows the ReSPer-AO method to visualize complex reaction pathway networks of hydrocarbon compounds with an enormous number of NPI isomers.

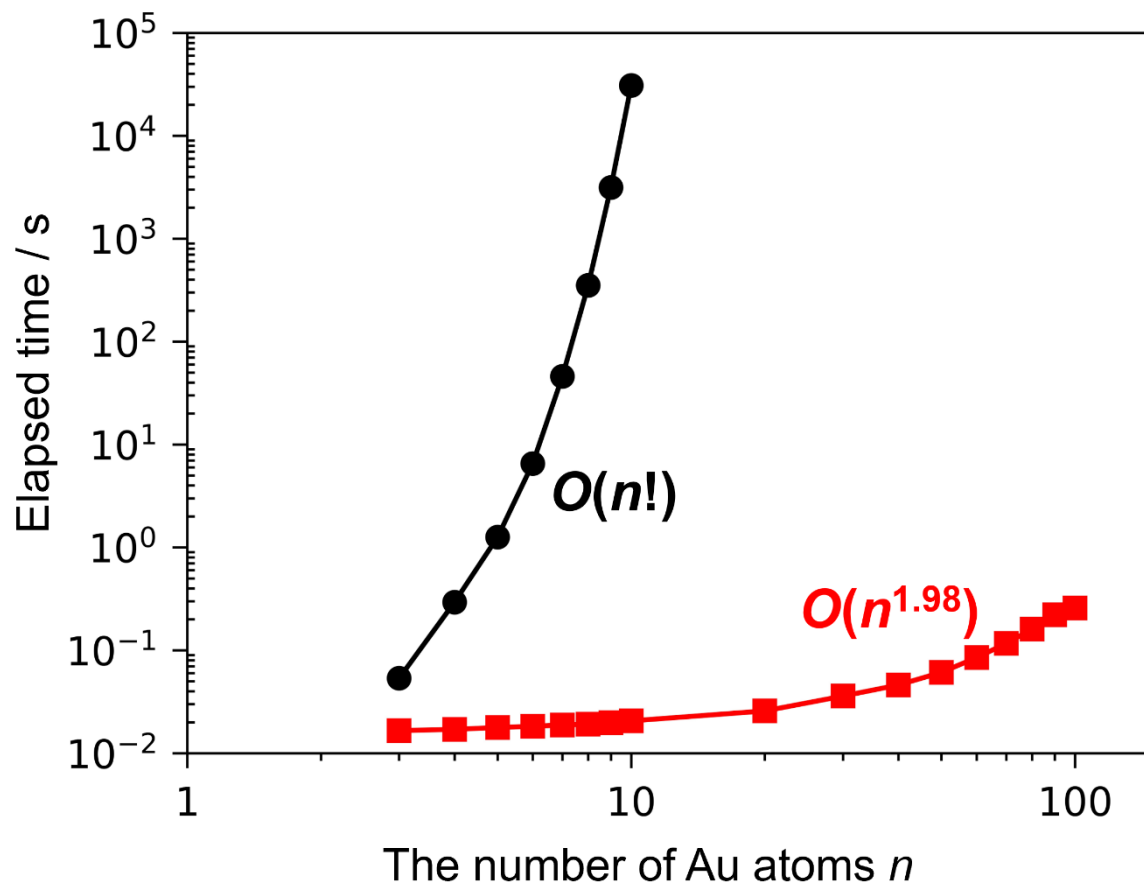


Figure 2. Elapsed time for calculating a single structural distance between two structures using the NPI-search method (black circles) and the ReSPer-AO method (red squares). The elapsed time for the ReSPer-AO method was computed by averaging multiple runs. Due to the high computational cost, the NPI-search method was only performed for $n = 3, 4, \dots, 10$. The ReSPer-AO method, however, was performed for $n = 3, 4, \dots, 10, 20, \dots, 100$.

3.3. Application to reaction route networks

To demonstrate the ReSPer-AO method, we applied both the ReSPer-AO and NPI-search methods to three reaction route networks: Au₅, Ag₃Cu₂, and C₄H₆. The isomerization

reaction of Au₅ has been investigated through TS exploration and IRC calculations, with a comprehensively searched reaction route network that includes all IRCs on the PES reported previously.⁴⁸ This network has recently served as a model for visualizing low-dimensional reaction spaces with ReSPer,²³ as well as for analyzing mathematical tools that provide deeper insights into reaction networks.^{18,49} Similarly, the exhaustive reaction route network of Ag₃Cu₂ has been reported and characterized using persistent homology.¹⁸ The reaction route network of C₄H₆ was previously reported;⁹ however, no attempts have been made to visualize it in low-dimensional coordinate space. In this section, we reused the two exhaustively searched reaction route networks of Au₅⁴⁸ and Ag₃Cu₂,¹⁸ which were obtained using the PBE/LanL2DZ level of theory. For C₄H₆, we searched for the reaction route network around 1,3-butadiene and identified 173 equilibrium structures (EQs) and 663 TSs. The reaction path search was conducted using the single-component AFIR (SC-AFIR) method at the ω B97XD/6-31G level with $\gamma = 1000$ kJ/mol, and all stationary points and paths obtained by SC-AFIR were refined (RePath) at the ω B97XD/6-31G* level with DownDC = 15. All electronic structure calculations and reaction path searches were performed using Gaussian16⁵⁰ and the development version of GRRM⁵¹, respectively.

To compare the ReSPer-AO and NPI-search methods, we analyzed the elapsed time for determining PCos on a workstation with an Intel Xeon Gold 6326 2.90 GHz processor using 32 cores, as well as the correctness rate of the Euclidean distance, and the mean and standard deviation of the cumulative error for Euclidean distances across all structure pairs. Here, the correctness rate of the ReSPer-AO method was computed by assuming that the Euclidean distance obtained via the NPI-search method was correct. In this section, all

Euclidean distances were calculated under condition (II).

First, we visualized the exhaustive reaction route network of Au₅, which includes five EQs, 14 TSs, and 14 IRCs. To create a two-dimensional network visualization, we selected 336 reference structures from the IRC network, including all EQs and TSs. The elapsed time for the ReSPer-AO method was 1 minute 50 seconds, about 38 times faster than the 1 hour 9 minutes by the NPI-search method. **Figure 3a** shows the two-dimensional reaction space determined by the ReSPer-AO method. The correctness rate of the ReSPer-AO method is 99.8%, with the remaining 0.2% converging to local solutions due to the highly symmetrical molecular structure, such as EQ2. The mean and standard deviation of the cumulative error for the Euclidean distance are $1.8 \times 10^{-3} \text{ \AA amu}^{1/2}$ and $1.3 \times 10^{-3} \text{ \AA amu}^{1/2}$, respectively. These results confirm that the positions of all reference structures in the two-dimensional spaces constructed by the ReSPer-AO and NPI-search methods are consistent, as shown in Supporting Information (**Figure S2**).

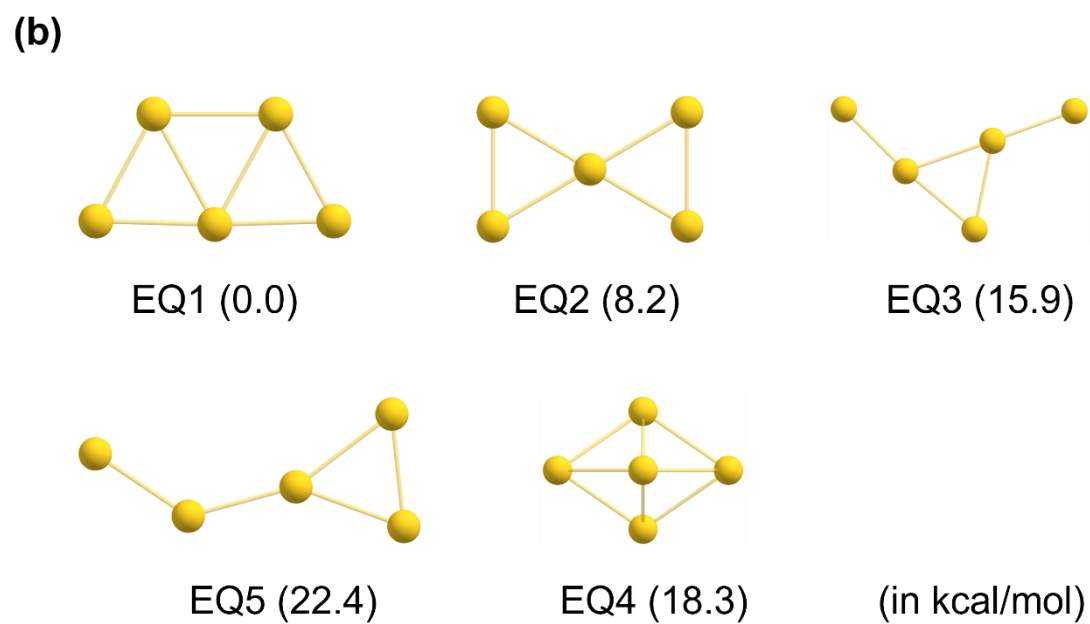
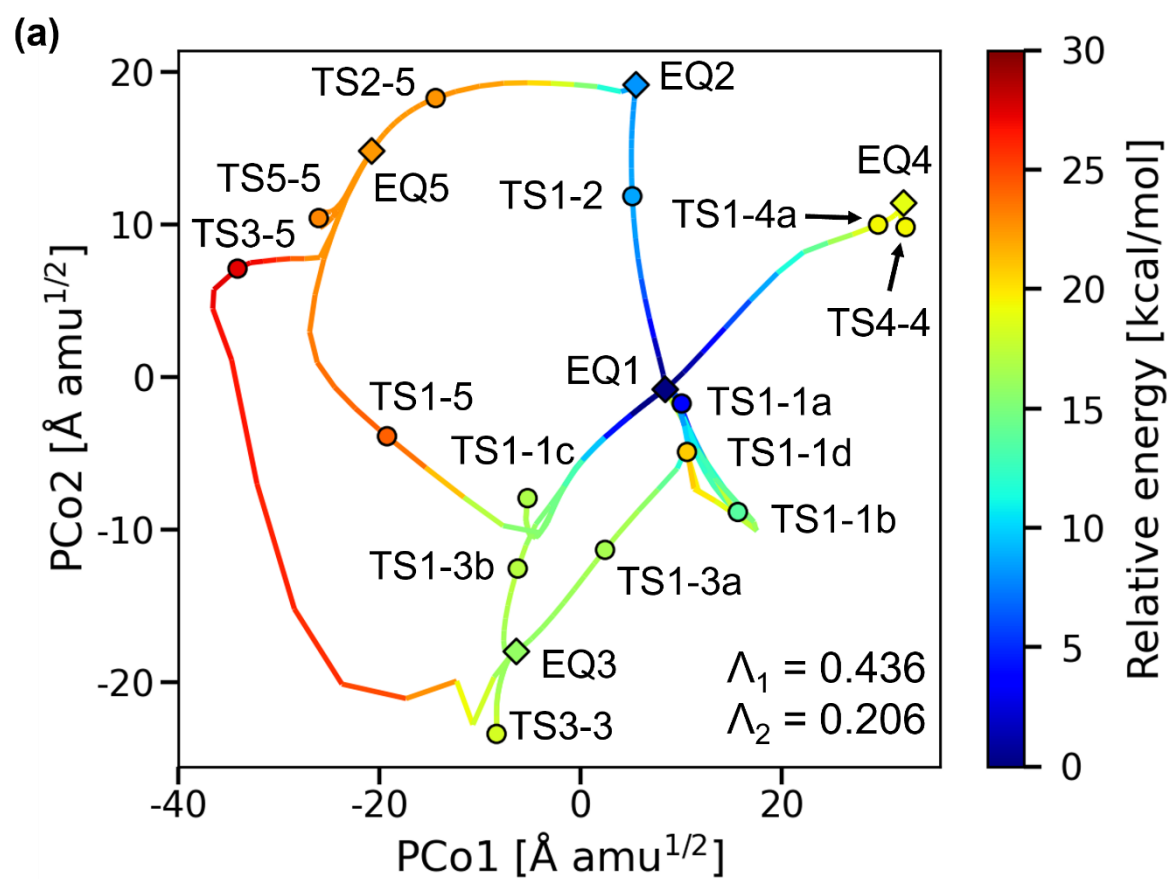


Figure 3. (a) Two-dimensional reaction space of Au₅ as defined by the ReSPer-AO method. EQs, TSs, and IRCs are represented by diamonds, circles, and curved lines, respectively. Reference structures along the IRC paths, excluding EQs and TSs, are omitted. Colors indicate the potential energy relative to the global minimum. (b) All EQs with their relative energies (in kcal/mol) are shown.

Next, we applied both methods to the exhaustive reaction route network of Ag₃Cu₂, which includes eight EQs, 37 TSs, and 37 IRCs. We selected 2,075 reference structures from the IRC network to visualize the two-dimensional network. Notably, the total number of NPI isomers for Ag₃Cu₂ ($3! \times 2! \times 2 = 24$) is smaller than that for Au₅ ($5! \times 2 = 240$). The ReSPer-AO method was completed in 51 minutes, approximately 4.7 times faster than the 4 hours required by the NPI-search method. **Figure 4a** shows the two-dimensional reaction space determined by the ReSPer-AO method. The ReSPer-AO method fully provides correct solutions for the atom-mapping problem; however, due to a numerical error on the order of 10^{-9} , the correctness rate is 99.998%. Therefore, we conclude that the ReSPer-AO method accurately reproduces the two-dimensional reaction space constructed by the NPI-search method, as shown in Supporting Information (**Figure S3**).

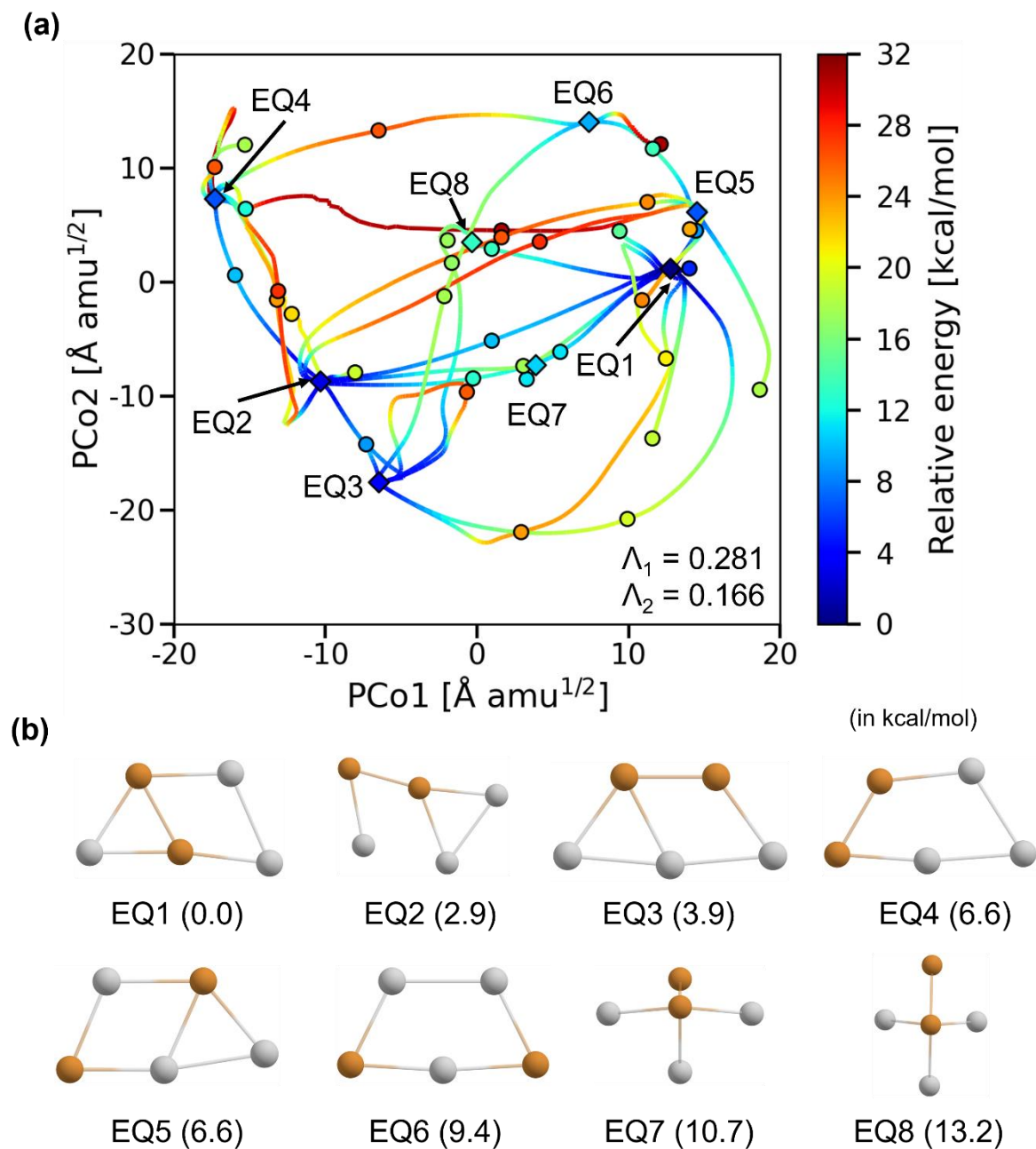


Figure 4. (a) Two-dimensional reaction space of Ag_3Cu_2 as defined by the ReSPer-AO method. EQs, TSs, and IRCs are represented by diamonds, circles, and curved lines, respectively. Reference structures along the IRC paths, excluding EQs and TSs, are omitted. Colors indicate the potential energy relative to the global minimum. (b) All EQs with their

relative energies (in kcal/mol) are shown.

Finally, we demonstrated the visualization of the two-dimensional reaction space of C_4H_6 using the ReSPer-AO and the NPI-search methods. Due to the computational cost of the NPI-search method, it was impractical to visualize all the explored IRCs. To reduce computational demands, we defined a partial network consisting of EQs linked by a single IRC originating from 1,3-butadiene, excluding TSs with relative energies greater than 120 kcal/mol along with their associated IRCs and EQs. This partial network includes eight EQs, 17 TSs, and 17 IRCs. **Figure 5a** shows the two-dimensional reaction space, containing 154 reference structures, constructed by the ReSPer-AO method. The elapsed time for the ReSPer-AO method was 31 seconds, approximately 4,400 times faster than the 38 hours by the NPI-search method. Similar to the Au_5 cluster, 0.6% of the Euclidean distance calculations converged on local solutions due to molecular symmetry. The mean and standard deviation of the cumulative error in the Euclidean distance are $9.8 \times 10^{-3} \text{ \AA amu}^{1/2}$ and $1.4 \times 10^{-2} \text{ \AA amu}^{1/2}$, respectively, indicating that the two-dimensional spaces constructed using ReSPer-AO and NPI-search are nearly identical, as shown in Supporting Information (**Figure S4**).

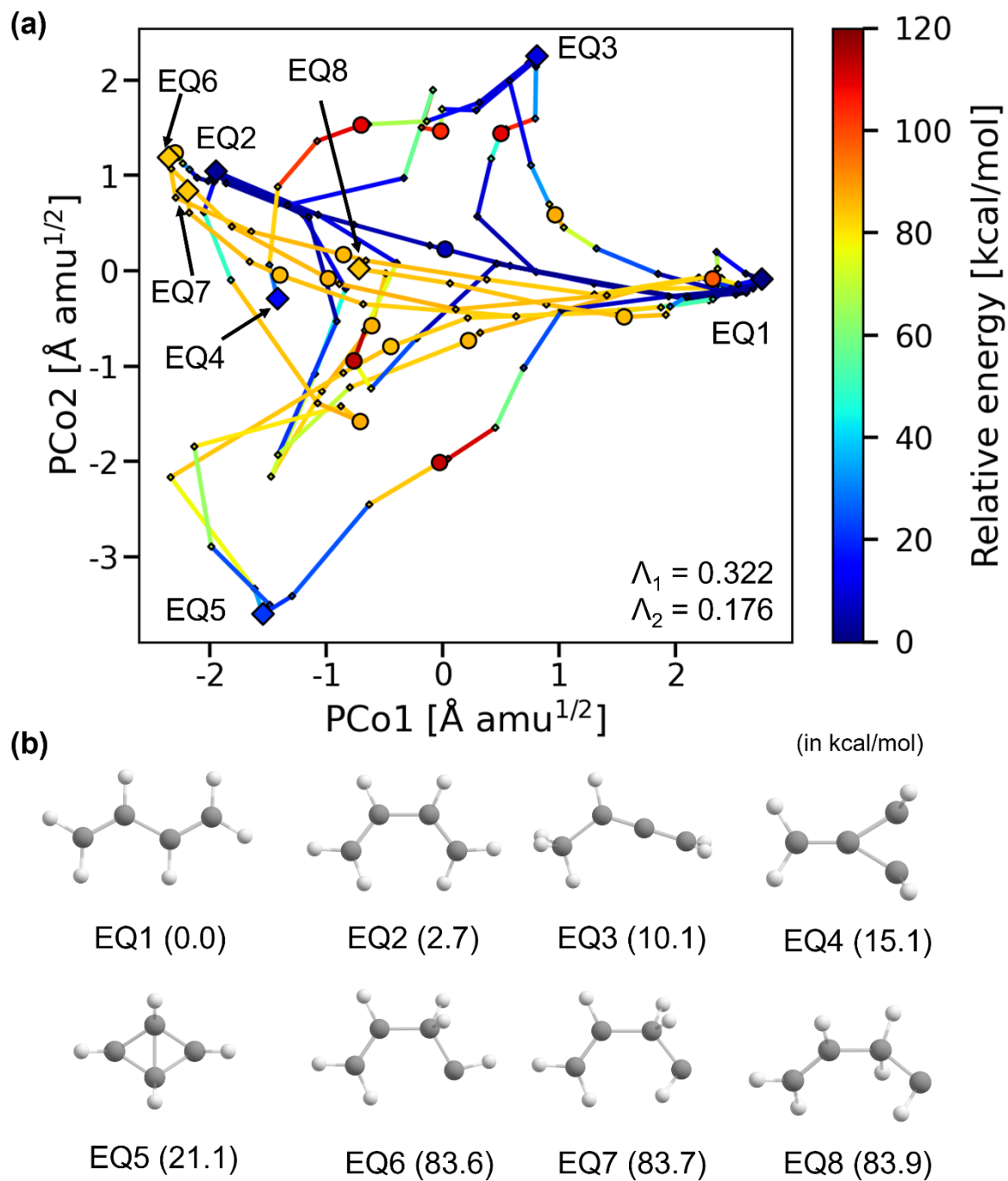


Figure 5. (a) Two-dimensional reaction space of C_4H_6 as defined by the ReSPer-AO method. EQs, TSs, and IRCs are represented by diamonds, circles, and curved lines, respectively.

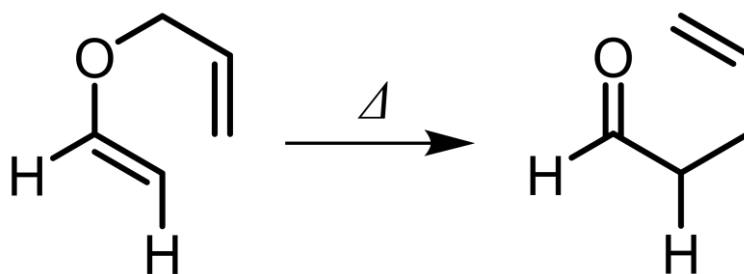
Reference structures along the IRC paths, excluding EQs and TSs, are omitted. Colors indicate the potential energy relative to the global minimum. (b) All EQs with their relative energies (in kcal/mol) are shown.

From these three demonstrations, we concluded that the ReSPer-AO method significantly reduced the computational cost compared to the NPI-search method while maintaining the accuracy of the Euclidean distance, successfully visualizing the two-dimensional reaction space. Although distance calculations in the ReSPer-AO method occasionally converge to local solutions, particularly for highly symmetric molecular structures, the resulting error in the Euclidean distance is very small (on the order of 10^{-3} Å amu^{1/2}) and has minimal impact on the coordinates in the two-dimensional reaction space.

3.4. Application to complicated reaction route network of C₅H₈O

Here, we analyze a complicated reaction route network involving the Claisen rearrangement reaction of allyl vinyl ether (C₅H₈O),¹⁰ as shown in **Scheme 1**. The number of NPI isomers of C₅H₈O is 9,676,800, making it infeasible to analyze using the NPI-search method. First, we searched for the IRC paths related to the Claisen rearrangement reaction using the AFIR method.⁸ To obtain the IRC network, we initially searched for the IRC paths at the GFN2-xTB⁵² level with $\gamma = 200.0$ kJ/mol, identifying 32 EQs and 82 TSs. No options were added to restrict the search. The obtained TSs were subsequently reoptimized at the

M062X/6-311+G(2d,p) level of theory. Finally, the IRC paths were computed at the M062X/6-311+G(2d,p) level starting from all reoptimized TSs using the RePath method in AFIR. As a result, we found 24 EQs and 69 TSs, excluding eight TSs with relative energies greater than 70 kcal/mol, along with the corresponding four EQs. All EQs and TSs were validated by normal mode analysis. Among the 69 IRCs, two corresponded to Claisen rearrangement reactions of allyl vinyl ether, directly linking the reactant and product regions. In these regions, ten and 14 EQs were identified, respectively. Electronic structure calculations for the reoptimization and RePath were performed using Gaussian16.⁵⁰



Scheme 1. Claisen rearrangement reaction for C₅H₈O.

Figure 6a shows the two-dimensional reaction route network of C₅H₈O defined by the ReSPer-AO method. Here, 2,307 structures along 69 IRCs, including 24 EQs and 69 TSs, were selected as reference structures, with the number of molecular structures reduced to approximately 30 for each IRC. The contribution ratios of PCo1 and PCo2 (Λ_1 , Λ_2) are (0.404, 0.142), giving a cumulated contribution of 0.644. This indicates that 64% of the information in the full-dimensional reaction space is captured in the two-dimensional reaction subspace. Representative molecular structures are shown in **Figure 6b**, and their molecular coordinates

are shown in Supporting Information.

Figure 6a clearly shows the existence of two TSs connecting the regions for reactant and product conformers, R-region and P-region, related to the Claisen rearrangement reactions, named the chair-type TS (TS_{chair}) and boat-type TS (TS_{boat}). In the previous study, this reaction system was analyzed by kinetic simulation, which suggested that chemical reaction processes were separated into reactant and product regions by TS_{chair} and TS_{boat} with high reaction barriers.¹⁰ This result is consistent with the landscape of the two-dimensional reaction network determined by the ReSPer-AO method.

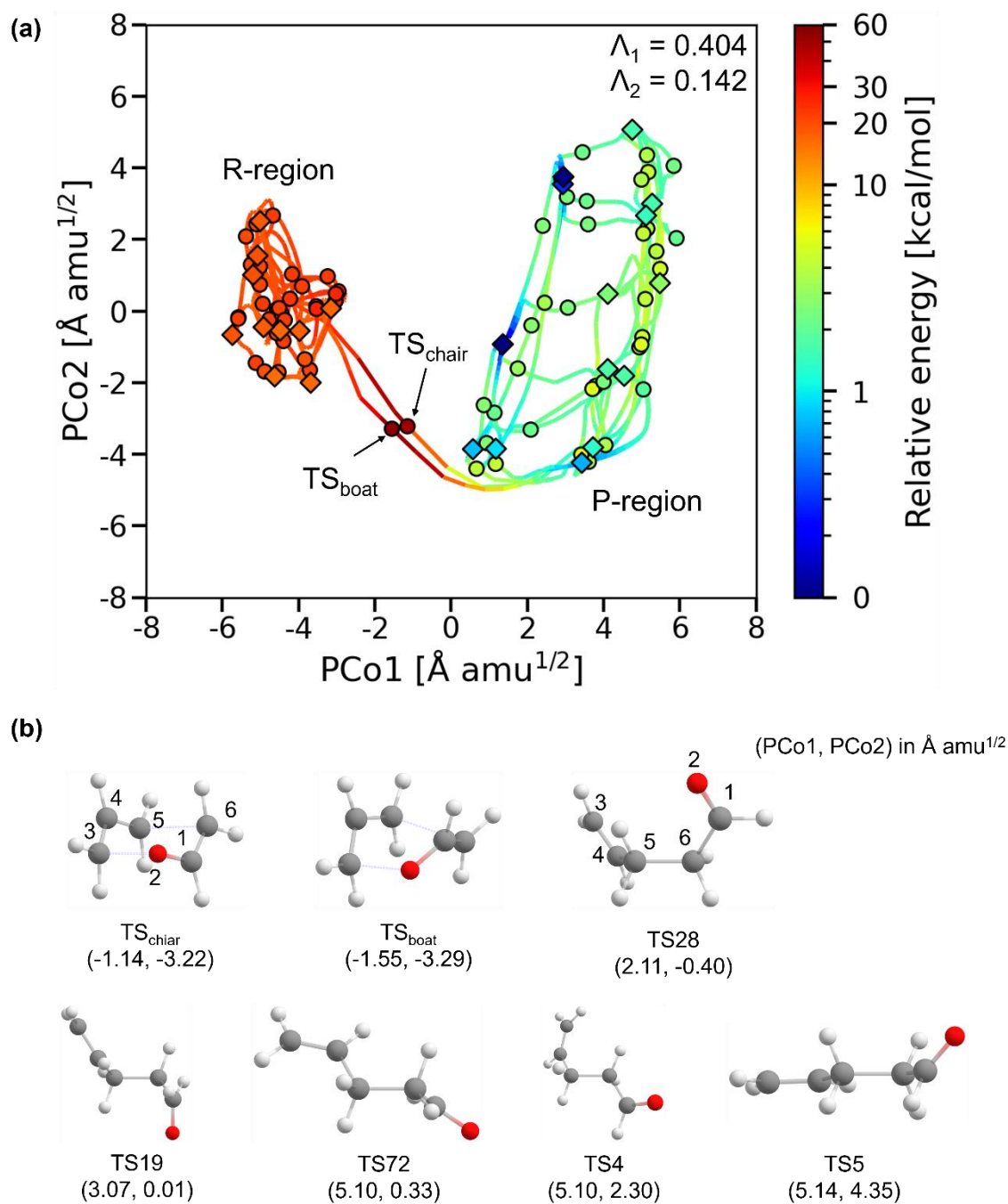


Figure 6. (a) Two-dimensional reaction route network of $\text{C}_5\text{H}_8\text{O}$, including the Claisen rearrangement reactions of allyl vinyl ether, constructed by the ReSPer-AO method. EQs, TSs, and IRCs are denoted as diamonds, circles, and curved lines, respectively. The color

indicates the potential energy relative to the global minimum. (b) Representative structures in the product region and two TS structures (TS_{boat} and TS_{chair}) connecting reactant and product regions.

The ReSPer analysis enables us to understand the physical meanings of PCos by examining structural changes along the extracted principal coordinate axes. From a global perspective, it is clear that PCo1 represents the Claisen rearrangement reaction, connecting R- and P-regions. To further clarify the physical meaning of PCo-*i*, we examined the structural changes of reference structures along PCo-*j* ($j \neq i$) while keeping PCo-*i* constant. Here, we focused on the product region where PCo1 ranges from 0.0 to 7.0 Å amu^{1/2} and discussed two dihedral angles, $d_{(O2C1C6C5)}$ and $d_{(C1C6C5C4)}$, corresponding to the conformation between O2 and C5 and the torsion of the main chain, respectively. These dihedral angles are listed in **Tables 3** and **4**. As PCo1 increases with PCo2 near 0.0, $d_{(O2C1C6C5)}$ also increases, as shown in Table 3. Indeed, molecular structures in the product region change from the eclipsed conformation to the staggered conformation as PCo1 increases, which is consistent with structural changes along IRCs parallel to PCo1 in the product region. Similarly, as PCo2 increases with PCo1 near 5.0, $d_{(O2C1C6C5)}$ increases, as shown in Table 4. This corresponds to the fact that molecular structures change from arch-shaped (6-membered ring) conformers of the alkyl chain (C1C6C5C4) to linear conformers as PCo2 increases, as indicated by the representative molecular structures in **Figure 6b**. From these results, we concluded that the ReSPer-AO method has successfully extracted structural features of the Claisen rearrangement reaction and visualized the complicated reaction route network of C₅H₈O in

the two-dimensional reaction space.

Table 3. PCos (in Å amu^{1/2}) and corresponding torsion angles $d_{(O2C1C6C5)}$ and $d_{(C1C6C5C4)}$ (in degrees) for the representative structures in the product region. TSs are sorted in descending order of PCo1.

	PCo1	PCo2	$d_{(O2C1C6C5)}$	$d_{(C1C6C5C4)}$
TS28	2.11	-0.40	1.3	122.4
TS19	3.07	0.01	75.2	61.7
TS72	5.10	0.33	142.9	70.21

Table 4. PCos (in Å amu^{1/2}) and corresponding torsion angles $d_{(O2C1C6C5)}$ and $d_{(C1C6C5C4)}$ (in degrees) for the representative structures in the product region. TSs are sorted in descending order of PCo2.

	PCo1	PCo2	$d_{(O2C1C6C5)}$	$d_{(C1C6C5C4)}$
TS72	5.10	0.33	142.9	70.2
TS4	5.15	2.30	175.8	120.1
TS5	5.14	4.35	176.1	176.1

Finally, dynamical reaction processes simulated by the AIMD method were projected into the preconstructed two-dimensional reaction space using the ReSPer-AO method. AIMD is a molecular dynamics technique that uses potential gradients computed from electronic structure calculations to provide a classical trajectory, which is a time-series data of atomic coordinates and velocities. In this study, AIMD simulations were started from TS_{chair}, which has lower energy than TS_{boat}. Atomic coordinates and velocities were generated

using Boltzmann sampling, where initial energy to each normal mode is added according to the Boltzmann distribution at 469.1 K. We selected 100 sets of initial coordinates and velocities directed toward the product region and ran the simulation until the simulation time reached 80 fs with a time step of 0.2 fs. The AIMD simulations were performed by SPPR,⁵³ and electronic structure calculations were performed using Gaussian16⁵⁰ at the M062X/6-311+G(2d,p) level, consistent with the IRC search calculations.

Figure 7 shows representative AIMD trajectories projected into the two-dimensional reaction space depicted in Figure 5a. These trajectories were classified according to their different dynamical behaviors in the variation of $d_{(\text{O2C1C6C5})}$ using the ReSPer-AO method. Figure 6a shows three trajectories that follow the IRC after leaving TS_{chair} and reach around EQ4 at $(\text{PCo1}, \text{PCo2}) = (3.73, -3.81)$, which is linked to TS_{chair} by the IRC. This behavior is dominant; 50 out of 100 trajectories were classified this way. These trajectories show that $d_{(\text{O2C1C6C5})}$ changes from approximately 60° to 160° , consistent with one of the IRCs. Figure 6b shows trajectories that deviate from the IRC before reaching EQ4, with 44 trajectories classified in this manner. Further analysis revealed that $d_{(\text{O2C1C6C5})}$ for these trajectories is mainly within the range of 60° to 120° , which is narrower compared to the trajectories in Figure 6a. Figure 6c shows trajectories that depart from the IRC immediately after leaving TS_{chair} , with six trajectories classified this way. The distribution of $d_{(\text{O2C1C6C5})}$ for these trajectories is mainly between 60° and 80° , narrower than those in Figures 5a and 5b. By analyzing the dynamical reaction behaviors using the ReSPer-AO method, we successfully classified the fractional conformation changes after passing through TS_{chair} . This analysis of AIMD trajectories in the two-dimensional reaction space, which

accurately reproduces the potential energy surface containing various elementary reaction processes, allows for detailed understanding without relying on chemical intuition.

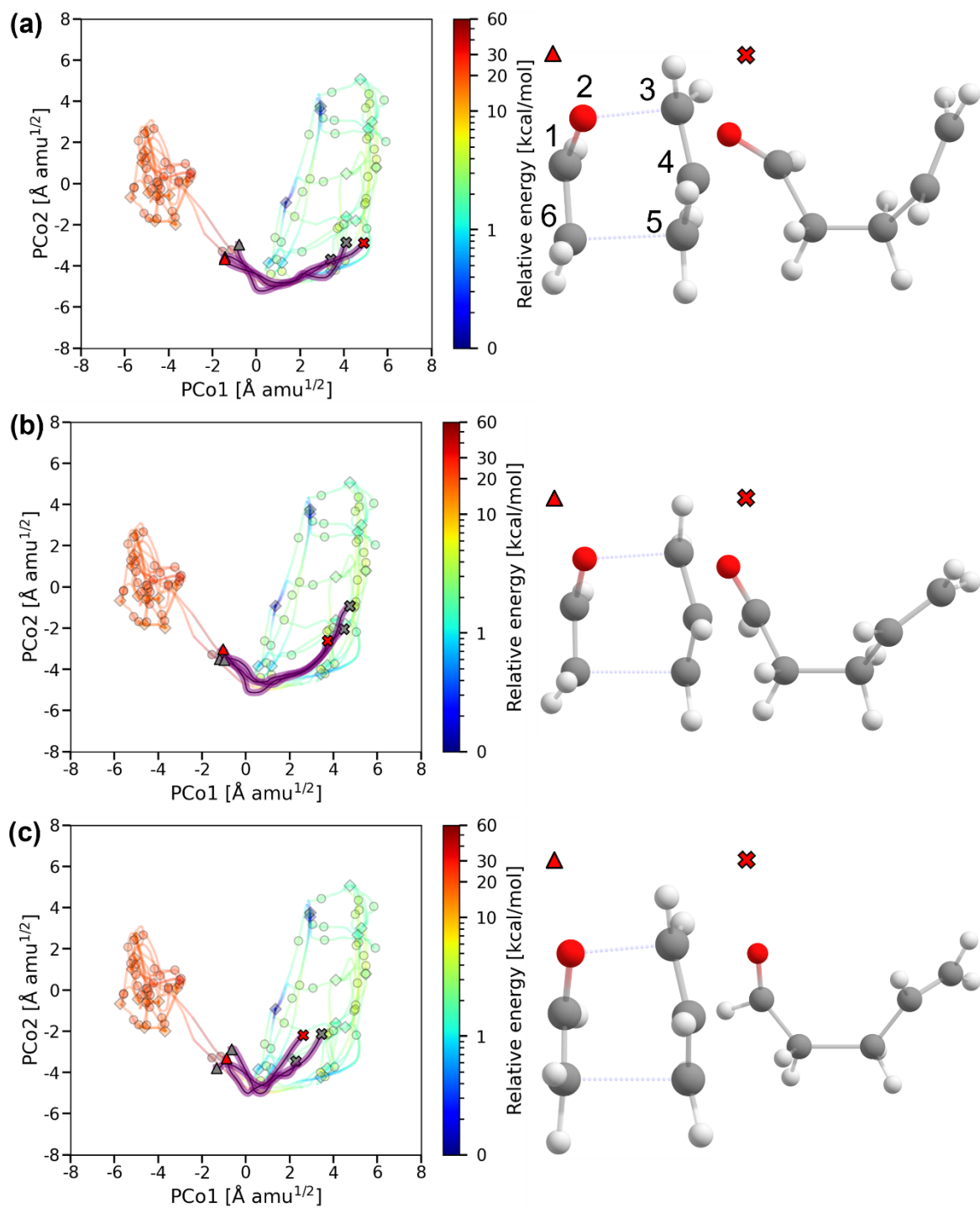


Figure 7. Projected AIMD trajectories onto the two-dimensional reaction space defined by the ReSPer-AO method: (a) Three trajectories following the IRC after leaving TS_{chair} ; (b)

three trajectories departing from the IRC midway; (c) three trajectories departing from the IRC immediately after leaving TS_{chair} . Initial and terminal structures in AIMD trajectories are denoted as triangles and cross marks, respectively. The molecular structures corresponding to red triangles and red crosses in the two-dimensional reaction space are also shown.

4. Conclusion

The Reaction space projector (ReSPer) method is designed to visualize the potential energy hypersurface in a lower-dimensional subspace by extracting principal coordinates that characterize inter-structural dissimilarity from given reference structures. The most time-consuming part of the ReSPer procedure is calculating the structural distance matrix composed of Euclidean distances between all possible structure pairs because contributions of molecular axes and atom mapping must be optimized to obtain optimal distances. While the molecular alignment problem can be efficiently solved using the Kabsch algorithm, the atom mapping problem is challenging due to the combinatorial explosion of nuclear permutation inversion (NPI) isomers. To address this problem, we previously developed the NPI-search method, which exhaustively explores atom mappings between a reference structure and all possible NPI isomers of a target structure. Although the ReSPer analysis based on the NPI-search method has successfully elucidated chemical reaction mechanisms for relatively small molecules, its application to complex reaction route networks of organic compounds is impractical due to the significantly high computational cost.

To accelerate atom mapping optimization, we implemented the alternate optimization (AO) algorithm,³⁸ a combinatorial optimization technique, in the ReSPer

program and validated its correctness and computational cost compared to the NPI-search method. The computational complexity of calculating an Euclidean distance in the ReSPer-AO method was estimated as $O(n^{1.98})$, compared to $O(n!)$ in the NPI-search method. Consequently, we demonstrated that the ReSPer-AO method can significantly reduce computational time without losing correctness in Euclidean distance calculations. Indeed, the ReSPer-AO method successfully reproduced the two-dimensional reaction space of Au_5 constructed by the NPI-search method.

We then applied the ReSPer-AO method to visualize the complex reaction route network of $\text{C}_5\text{H}_8\text{O}$, which includes the Claisen rearrangement reactions of allyl vinyl ether and has approximately 10 million NPI isomers. Using the ReSPer-AO method, we successfully visualized the reaction route network consisting of 69 IRCs within 200 minutes, gaining insights into the physical meanings of the extracted PCos. Additionally, the ReSPer-AO method clarified differences in dynamical conformation changes by projecting AIMD trajectories onto the predefined two-dimensional reaction space representing the potential energy landscape of $\text{C}_5\text{H}_8\text{O}$.

The ReSPer-AO method facilitates the analysis of chemical reaction mechanisms based on a low-dimensional reaction space that accurately reproduces a reaction route network of a hydrocarbon compound characterized by a large number of NPI isomers and complex IRC network structures. In the future, we plan to expand the ReSPer-AO method to practical organic chemical reactions and biological reactions to extract reactive degrees of freedom and elucidate dynamical reaction mechanisms without relying on the experience and intuition of chemists.

Acknowledgments

We appreciate Prof. Masato Kobayashi and Mr. Burai Murayama for providing the reaction route network of Ag_3Cu_2 . This work was supported by JST SPRING, Grant Number JPMJSP2119 to L. Qu; and JST, ACT-X Grant Number JPMJAX23DE, Japan and JSPS KAKENHI grant number JP22K14640 to T. Tsutsumi. This work was partly supported by the Photo-excitonix Project in Hokkaido University, JST CREST grant number JPMJCR1902, and MEXT Program (Data Creation and Utilization-Type Material Research and Development Project Grant Number JPMXP1122712807). Part of the calculations were performed using the Research Center for Computational Science, Okazaki, Japan (Project: 22-IMS-C019).

Author Contributions

L. Qu: Conceptualization, Methodology, Software, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

T. Tsutsumi: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing.

Y. Ono: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing.

T. Taketsugu: Conceptualization, Methodology, Writing – original draft, Writing – review

& editing, Resources, Funding acquisition.

Competing Interests

The authors declare no competing interests.

ORCID

Takuro Tsutsumi: 0000-0001-5086-112

Tetsuya Taketsugu: 0000-0002-1337-6694

References

- (1) Fukui, K. Formulation of the Reaction Coordinate. *J Phys Chem* **1970**, *74* (23), 4161–4163. <https://doi.org/10.1021/j100717a029>.
- (2) Maeda, S.; Harabuchi, Y.; Ono, Y.; Taketsugu, T.; Morokuma, K. Intrinsic Reaction Coordinate: Calculation, Bifurcation, and Automated Search. *Int J Quantum Chem* **2015**, *115* (5), 258–269. <https://doi.org/10.1002/qua.24757>.
- (3) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry. *J Chem Theory Comput* **2014**, *10* (3), 897–907. <https://doi.org/10.1021/ct401004r>.
- (4) Martínez-Núñez, E. An Automated Method to Find Transition States Using Chemical Dynamics Simulations. *J Comput Chem* **2015**, *36* (4), 222–234. <https://doi.org/10.1002/jcc.23790>.
- (5) Yang, M.; Zou, J.; Wang, G.; Li, S. Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method. *Journal of Physical Chemistry A* **2017**, *121* (6), 1351–1361. <https://doi.org/10.1021/acs.jpca.6b12195>.
- (6) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. Blackwell Publishing Inc. March 1, 2018. <https://doi.org/10.1002/wcms.1354>.
- (7) Field-Theodore, T. E.; Taylor, P. R. ALTRUISM: A Higher Calling. *J Chem Theory Comput* **2020**, *16* (7), 4388–4398. <https://doi.org/10.1021/acs.jctc.0c00388>.
- (8) Maeda, S.; Harabuchi, Y. Exploring Paths of Chemical Transformations in Molecular and Periodic Systems: An Approach Utilizing Force. *WIREs Computational Molecular Science* **2021**, *11* (6). <https://doi.org/10.1002/wcms.1538>.
- (9) Maeda, S.; Harabuchi, Y.; Takagi, M.; Saita, K.; Suzuki, K.; Ichino, T.; Sumiya, Y.; Sugiyama, K.; Ono, Y. Implementation and Performance of the Artificial Force Induced Reaction Method in the GRRM17 Program. *J Comput Chem* **2018**, *39* (4), 233–251. <https://doi.org/10.1002/jcc.25106>.
- (10) Sumiya, Y.; Nagahata, Y.; Komatsuzaki, T.; Taketsugu, T.; Maeda, S. Kinetic Analysis for the Multistep Profiles of Organic Reactions: Significance of the Conformational Entropy on the Rate Constants of the Claisen Rearrangement. *Journal of Physical Chemistry A* **2015**, *119* (48), 11641–11649. <https://doi.org/10.1021/acs.jpca.5b09447>.
- (11) Hayashi, H.; Katsuyama, H.; Takano, H.; Harabuchi, Y.; Maeda, S.; Mita, T. In Silico

Reaction Screening with Difluorocarbene for N-Difluoroalkylative Dearomatization of Pyridines. *Nature Synthesis* **2022**, *1* (10), 804–814. <https://doi.org/10.1038/s44160-022-00128-y>.

- (12) Sugiyama, K.; Saita, K.; Maeda, S. A Reaction Route Network for Methanol Decomposition on a Pt(111) Surface. *J Comput Chem* **2021**, *42* (30), 2163–2169. <https://doi.org/10.1002/jcc.26746>.
- (13) Nabata, H.; Maeda, S. Systematic Search for Thermal Decomposition Pathways of Formic Acid on Anatase TiO₂ (101) Surface. *ChemCatChem* **2023**, *15* (19). <https://doi.org/10.1002/cctc.202300752>.
- (14) Mita, T.; Takano, H.; Hayashi, H.; Kanna, W.; Harabuchi, Y.; Houk, K. N.; Maeda, S. Prediction of High-Yielding Single-Step or Cascade Pericyclic Reactions for the Synthesis of Complex Synthetic Targets. *J Am Chem Soc* **2022**, *144* (50), 22985–23000. <https://doi.org/10.1021/jacs.2c09830>.
- (15) Tsutsumi, T.; Harabuchi, Y.; Ono, Y.; Maeda, S.; Taketsugu, T. Analyses of Trajectory On-the-Fly Based on the Global Reaction Route Map. *Physical Chemistry Chemical Physics* **2017**, *2018* (20), 1364–1372. <https://doi.org/https://doi.org/10.1002/wcms.1354>.
- (16) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media* **2009**, *3* (1), 361–362. <https://doi.org/10.1609/icwsm.v3i1.13937>.
- (17) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- (18) Murayama, B.; Kobayashi, M.; Aoki, M.; Ishibashi, S.; Saito, T.; Nakamura, T.; Teramoto, H.; Taketsugu, T. Characterizing Reaction Route Map of Realistic Molecular Reactions Based on Weight Rank Clique Filtration of Persistent Homology. *J Chem Theory Comput* **2023**, *19* (15), 5007–5023. <https://doi.org/10.1021/acs.jctc.2c01204>.
- (19) Karamzadeh, R.; Karimi-Jafari, M. H.; Sharifi-Zarchi, A.; Chitsaz, H.; Salekdeh, G. H.; Moosavi-Movahedi, A. A. Machine Learning and Network Analysis of Molecular Dynamics Trajectories Reveal Two Chains of Red/Ox-Specific Residue Interactions in Human Protein Disulfide Isomerase. *Sci Rep* **2017**, *7* (1). <https://doi.org/10.1038/s41598-017-03966-5>.

- (20) Hare, S. R.; Bratholm, L. A.; Glowacki, D. R.; Carpenter, B. K. Low Dimensional Representations along Intrinsic Reaction Coordinates and Molecular Dynamics Trajectories Using Interatomic Distance Matrices. *Chem Sci* **2019**, *10*(43), 9954–9968. <https://doi.org/10.1039/c9sc02742d>.
- (21) Le, H. M.; Kumar, S.; May, N.; Martinez-Baez, E.; Sundararaman, R.; Krishnamoorthy, B.; Clark, A. E. Behavior of Linear and Nonlinear Dimensionality Reduction for Collective Variable Identification of Small Molecule Solution-Phase Reactions. *J Chem Theory Comput* **2022**, *18* (3), 1286–1296. <https://doi.org/10.1021/acs.jctc.1c00983>.
- (22) Li, X.; Xie, Y.; Hu, D.; Lan, Z. Analysis of the Geometrical Evolution in On-the-Fly Surface-Hopping Nonadiabatic Dynamics with Machine Learning Dimensionality Reduction Approaches: Classical Multidimensional Scaling and Isometric Feature Mapping. *J Chem Theory Comput* **2017**, *13* (10), 4611–4623. <https://doi.org/10.1021/acs.jctc.7b00394>.
- (23) Tsutsumi, T.; Ono, Y.; Arai, Z.; Taketsugu, T. Visualization of the Intrinsic Reaction Coordinate and Global Reaction Route Map by Classical Multidimensional Scaling. *J Chem Theory Comput* **2018**, *14* (8), 4263–4270. <https://doi.org/10.1021/acs.jctc.8b00176>.
- (24) Oliveira, A. B.; Yang, H.; Whitford, P. C.; Leite, V. B. P. Distinguishing Biomolecular Pathways and Metastable States. *J Chem Theory Comput* **2019**, *15* (11), 6482–6490. <https://doi.org/10.1021/acs.jctc.9b00704>.
- (25) Tsutsumi, T.; Ono, Y.; Arai, Z.; Taketsugu, T. Visualization of the Dynamics Effect: Projection of on-the-Fly Trajectories to the Subspace Spanned by the Static Reaction Path Network. *J Chem Theory Comput* **2020**, *16* (7), 4029–4037. <https://doi.org/10.1021/acs.jctc.0c00018>.
- (26) Shi, W.; Jia, T.; Li, A. Quasi-Classical Trajectory Analysis with Isometric Feature Mapping and Locally Linear Embedding: Deep Insights into the Multichannel Reaction on an $\text{NH}_3^+(4\text{A})$ Potential Energy Surface. *Physical Chemistry Chemical Physics* **2020**, *22* (31), 17460–17471. <https://doi.org/10.1039/d0cp01941k>.
- (27) Shi, W.; Wang, K.; Zhang, P.; Yu, L.; Li, A. Mode-Specific Dynamics in Multichannel Reaction $\text{NH}_3^+ + \text{H}_2$. *Physical Chemistry Chemical Physics* **2021**, *23* (36), 20352–20358. <https://doi.org/10.1039/D1CP02603H>.
- (28) Tsutsumi, T.; Ono, Y.; Taketsugu, T. Visualization of Reaction Route Map and

- Dynamical Trajectory in Reduced Dimension. *Chemical Communications* **2021**, 57 (89), 11734–11750. <https://doi.org/10.1039/D1CC04667E>.
- (29) Tsutsumi, T.; Ono, Y.; Taketsugu, T. Reaction Space Projector (ReSPer) for Visualizing Dynamic Reaction Routes Based on Reduced-Dimension Space. *Top Curr Chem* **2022**, 380 (3), 19. <https://doi.org/10.1007/s41061-022-00377-7>.
- (30) Härdle, W. K.; Simar, L. *Applied Multivariate Statistical Analysis*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2015. <https://doi.org/10.1007/978-3-662-45171-7>.
- (31) Tsutsumi, T.; Ono, Y.; Taketsugu, T. Multi-State Energy Landscape for Photoreaction of Stilbene and Dimethyl-Stilbene. *J Chem Theory Comput* **2022**, 18 (12), 7483–7495. <https://doi.org/10.1021/acs.jctc.2c00560>.
- (32) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica Section A* **1976**, 32 (5), 922–923. <https://doi.org/10.1107/S0567739476001873>.
- (33) Umeyama, S. Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Trans Pattern Anal Mach Intell* **1991**, 13 (4), 376–380. <https://doi.org/10.1109/34.88573>.
- (34) Vásquez-Pérez, J. M.; Martínez, G. U. G.; Köster, A. M.; Calaminici, P. The Discovery of Unexpected Isomers in Sodium Heptamers by Born–Oppenheimer Molecular Dynamics. *J Chem Phys* **2009**, 131 (12). <https://doi.org/10.1063/1.3231134>.
- (35) Marques, J. M. C.; Llanio-Trujillo, J. L.; Abreu, P. E.; Pereira, F. B. How Different Are Two Chemical Structures? *J Chem Inf Model* **2010**, 50 (12), 2129–2140. <https://doi.org/10.1021/ci100219f>.
- (36) Wales, D. J.; Carr, J. M. Quasi-Continuous Interpolation Scheme for Pathways between Distant Configurations. *J Chem Theory Comput* **2012**, 8 (12), 5020–5034. <https://doi.org/10.1021/ct3004832>.
- (37) Griffiths, M.; Niblett, S. P.; Wales, D. J. Optimal Alignment of Structures for Finite and Periodic Systems. *J Chem Theory Comput* **2017**, 13 (10), 4914–4931. <https://doi.org/10.1021/acs.jctc.7b00543>.
- (38) Fukutani, T.; Miyazawa, K.; Iwata, S.; Satoh, H. G-RMSD: Root Mean Square Deviation Based Method for Three-Dimensional Molecular Similarity Determination. *Bull Chem Soc Jpn* **2021**, 94 (2), 655–665. <https://doi.org/10.1246/BCSJ.20200258>.
- (39) Vásquez-Pérez, J. M.; Zárate-Hernández, L. Á.; Gómez-Castro, C. Z.; Nolasco-Hernández, U. A. A Practical Algorithm to Solve the Near-Congruence Problem for

- Rigid Molecules and Clusters. *J Chem Inf Model* **2023**, *63* (4), 1157–1165. <https://doi.org/10.1021/acs.jcim.2c01187>.
- (40) Nabika, T.; Iwata, S.; Satoh, H. Exact Solution for Minimization of Root Mean Square Deviation with G-RMSD to Determine Molecular Similarity. *Bull Chem Soc Jpn* **2024**, *97*(4). <https://doi.org/10.1093/bulcsj/uoae037>.
- (41) Sadeghi, A.; Ghasemi, S. A.; Schaefer, B.; Mohr, S.; Lill, M. A.; Goedecker, S. Metrics for Measuring Distances in Configuration Spaces. *Journal of Chemical Physics* **2013**, *139*(18). <https://doi.org/10.1063/1.4828704>.
- (42) Young, G.; Householder, A. S. Discussion of a Set of Points in Terms of Their Mutual Distances. *Psychometrika* **1938**, *3* (1), 19–22. <https://doi.org/10.1007/BF02287916>.
- (43) Kuhn, H. W. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly* **1955**, *2* (1–2), 83–97. <https://doi.org/10.1002/nav.3800020109>.
- (44) Crouse, D. F. On Implementing 2D Rectangular Assignment Algorithms. *IEEE Trans Aerosp Electron Syst* **2016**, *52* (4), 1679–1696. <https://doi.org/10.1109/TAES.2016.140952>.
- (45) Tian Lu. *Molclus program Version 1.9.9.9*. <http://www.keinsci.com/research/molclus.html>(accessed 1,13, 2024) (accessed 2024-01-13).
- (46) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. Pietro; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.;

- Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* **2020**, *17* (3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- (47) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- (48) Harabuchi, Y.; Ono, Y.; Maeda, S.; Taketsugu, T. Analyses of Bifurcation of Reaction Pathways on a Global Reaction Route Map: A Case Study of Gold Cluster Au₅. *J Chem Phys* **2015**, *143*(1). <https://doi.org/10.1063/1.4923163>.
- (49) Teramoto, H.; Saito, T.; Aoki, M.; Murayama, B.; Kobayashi, M.; Nakamura, T.; Taketsugu, T. Reproducing the Reaction Route Map on the Shape Space from Its Quotient by the Complete Nuclear Permutation-Inversion Group. *J Chem Theory Comput* **2023**, *19*(17), 5886–5896. <https://doi.org/10.1021/acs.jctc.3c00500>.
- (50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16, Revision C.01*; Gaussian Inc. Wallingford CT, 2016.
- (51) Maeda, S.; Harabuchi, Y.; Sumiya, Y.; Takagi, M.; Hatanaka, M.; Osada, Y.; Taketsugu,

- T.; Morokuma, K.; Ohno, K. *GRRM, A Development Version of April 9, 2020*.
- (52) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J Chem Theory Comput* **2019**, *15* (3), 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>.
- (53) Tsutsumi, T.; Oda, K.; Qu, L.; Sakugawa, J.; Minegishi, Y.; Fujita, A.; Iwata, K.; Kambara, R.; Mataka, K.; Harabuchi, Y.; Ono, Y.; Taketsugu, T. *SPPR (a Developmental Version); Hokkaido University: Sapporo, Japan., 2022*.

TOC graphic

