



Title	Characterizing Reaction Route Map of Realistic Molecular Reactions Based on Weight Rank Clique Filtration of Persistent Homology
Author(s)	Murayama, Burai; Kobayashi, Masato; Aoki, Masamitsu et al.
Citation	Journal of chemical theory and computation, 19(15), 5007-5023 https://doi.org/10.1021/acs.jctc.2c01204
Issue Date	2023-08-08
Doc URL	https://hdl.handle.net/2115/93056
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 © 2023 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-KXKGGZZBWTV65FK8F4AK
Type	journal article
File Information	Murayama_GRRMplusPD_v3.pdf



Characterizing Reaction Route Map of Realistic Molecular Reactions based on Weight Rank Clique Filtration of Persistent Homology

Burai Murayama,[†] Masato Kobayashi,^{*,†,‡} Masamitsu Aoki,[¶] Suguru Ishibashi,[¶]
Takuya Saito,[¶] Takenobu Nakamura,[§] Hiroshi Teramoto,^{||} and Tetsuya
Taketsugu^{†,‡}

[†]*Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060-0810,
Japan*

[‡]*WPI-ICReDD, Hokkaido University, Sapporo 060-0810, Japan*

[¶]*Department of Mathematics, Faculty of Science, Hokkaido University, Sapporo 060-0810,
Japan*

[§]*National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8568,
Japan*

^{||}*Faculty of Engineering Science, Kansai University, Suita 564-8680, Japan*

E-mail: k-masato@sci.hokudai.ac.jp

Abstract

A reaction route map (RRM) constructed using the GRRM program is a collection of elementary reaction pathways, each of which comprises two equilibrium (EQ) geometries and one transition state (TS) geometry connected by an intrinsic reaction coordinate (IRC). An RRM can be mathematically represented by a graph with weights

assigned to both vertices, corresponding to EQs, and edges, corresponding to TSs, representing the corresponding energies. In this study, we propose a method to extract topological descriptors of a weighted graph representing an RRM based on persistent homology (PH). The work of Mirth *et al.* [*J. Chem. Phys.* **2021**, *154*, 114114], in which PH analysis was applied to the $(3N - 6)$ -dimensional potential energy surface of an N atomic system, is related to the present method, but our method is practically applicable to realistic molecular reactions. Numerical assessments revealed that our method can extract the same information as the method proposed by Mirth *et al.* for the 0-th and 1-st PHs, except for the death of the 1-st PH. In addition, the information obtained from the 0-th PH corresponds to the analysis using the disconnectivity graph. The results of this study suggest that the descriptors obtained using the proposed method accurately reflect the characteristics of the chemical reactions and/or physicochemical properties of the system.

1 Introduction

Reaction pathways are fundamental for expressing chemical reactions in theoretical studies. Reaction pathways are defined by the potential energy surface (PES), which is a scalar function on $(3N - 6)$ -variables in the case of an N atomic system.^{1,2} The first-order saddle point and local minimum point of the PES represent the transition state (TS) geometry and equilibrium (EQ) geometry, respectively. The steepest descent curve connecting two EQ geometries via a TS is known as the intrinsic reaction coordinate (IRC),³ which represents an elementary reaction pathway. As the $(3N - 6)$ variables are determined based on the positions of the atoms, the reaction pathways aid in the expression of the configurational changes that occur through the chemical reaction from one EQ to another and facilitate the overall understanding of the mechanism of chemical reactions.

A reaction route map (RRM) is a collection of elementary reaction pathways that characterizes chemical reactions in N atomic chemical systems.^{4,5} Mathematically, an RRM is

expressed as a weighted graph with assigned weights corresponding to the energies of the EQs or TSs. Weighted graphs representing RRM can describe the physical properties of chemical reactions, such as the reaction rate or lifetime. Although visualization of the PES is difficult owing to its high dimensionality, RRM can be visualized on a plane. Therefore, RRM represents a reduction in information from the PES, transforming an object that cannot be visualized into one that can.^{6,7} Therefore, RRM is widely used as a standard representation method for chemical reactions of N atomic systems.⁴

Several algorithms for TS search from an EQ structure have been proposed so far. The eigenvector following (EF) method, originally proposed by Cerjan and Miller,⁸ follows the gentlest ascent path along the eigenvector of the Hessian with the smallest eigenvalue at the EQ to locate a TS. Jørgensen *et al.*⁹ used the gradient extremal path, where the gradient is an eigenvector of the Hessian, to connect stationary points. Quapp *et al.*¹⁰ proposed the reduced gradient following (RGF) method by modifying the gradient extremal following method. Zimmerman¹¹ proposed a single-ended growing string method to locate an approximate TS from a single EQ structure. Further, several algorithms have been proposed for the automatic and extensive construction of RRM for complicated molecular systems.¹² The global reaction route mapping (GRRM) strategy of Maeda *et al.*¹³ is a typical algorithm. The GRRM program comprises two methodologies: anharmonic downward distortion following (ADDF)^{14,15} and artificial force-induced reaction (AFIR)¹⁶⁻¹⁸ methods. The ADDF method is based on the downward distortion of the one-dimensional potential energy curve, along with isomerization or dissociative reaction from a potential minimum, with respect to the harmonic potential around the minimum. The AFIR method is a more intuitive method that applies an artificial force between two atoms or atomic groups to induce a reaction forcibly. By applying the ADDF or AFIR method, the GRRM program automatically identifies the EQ and TS geometries sequentially. Owing to these features, the GRRM program has been used in nanocluster catalysts¹⁹ and organic synthesis design (Ref. 20, with a maximum of 300 EQs and 3964 TSs), and surface chemistry (Ref. 21, with 133 EQs and 298

TSs). In addition, RRM data can be constructed using other programs such as Chemoton,^{22,23} KinBot,²⁴⁻²⁶ AARON,²⁷⁻²⁹ ChemTraYzer,^{30,31} and others.³²⁻³⁴ RRM data obtained via these programs are often more complicated than those obtained manually owing to the inclusion of multiple chemically equivalent reaction paths with different conformations. Thus, massive and complicated RRM data are obtained and used for predictive material design.

Utilization of the GRRM strategy introduces a problem in the representation and feature extraction of RRM data. In general, the layout of the visualization of graphs is arbitrary.^{35,36} In contrast, the GRRM strategy generates a large graph that contains multiple elementary reaction pathways. In such large graphs, arbitrariness can lead to the visualization of the same graph in different ways, making it difficult to capture the intrinsic features of the graph. It is difficult to determine whether two different visualizations represent the same graph, particularly for those with a high number of pathways. Similar problems exist in the visualization and feature extraction of RRM data obtained using the other algorithms. Although RRM data is already a well-reduced representation of PES, further reduction is required for complicated systems to overcome this problem and extract rich information from RRM data.

Several methods have been proposed to address this issue. Tsutsumi *et al.*³⁷⁻³⁹ proposed a reaction-space projector (ReSPer) method to generate reduced dimensional coordinates automatically based on the structural similarity evaluated via the classical multidimensional scaling method. In addition, the rate-constant matrix contraction (RCMC) method proposed by Sumiya *et al.*,⁴⁰ which was subsequently used to reduce the exploration space in GRRM,⁴¹ can be regarded as a method that summarizes the given RRM data. In particular, the disconnectivity graph (DG)^{42,43} and persistent homology (PH)^{44,45} are generic methods and promising candidates. DG displays the hierarchy of stable states on a PES using a tree structure. It is currently the most popular visualization method and has been applied to a large amount of simulation data, irrespective of whether the basis is a realistic molecular system or a toy model.¹ However, based on its original definition, DG can only represent the lowest energy transition barrier between two stable states that are not necessarily directly

connected. Hence, it cannot express cyclic pathways, which are frequently found in RRM, though a related extension has been developed subsequently.⁴⁶ Recently, PH was used to extract the topological features of a PES by Mirth *et al.*⁴⁵ PH is a recent data analysis technique proposed in the field of applied mathematics. Thus, fewer studies have been conducted on it compared to the case of DG. As described later, PH can represent cyclic pathways. However, the formulation by Mirth *et al.* requires all stationary points of the PES to be located analytically. Thus, although their work can be regarded as a starting point for understanding the relationship between the reaction pathways and PH, their method cannot be applied directly to RRM obtained via GRRM.

Herein, we propose a method for applying PH to RRM, which is applicable to realistic materials obtained via GRRM. First, we confirmed that the proposed method can extract information consistent with the results of Mirth *et al.* by applying it to the RRM of *n*-pentane, which was investigated in Ref. 45 based on the model potential. Subsequently, the method was applied to RRM of metal nanoclusters and organic molecular systems, revealing its capability of extracting an RRM’s characteristic features, including not only those extracted by DG but also those newly identified by the proposed scheme, such as cyclic reaction pathways.

The rest of this paper is organized as follows. In Section 2, the computational and mathematical methods used are discussed, including a brief introduction to the GRRM program, the application of PH to a weighted graph, and a demonstration of the proposed scheme on a toy model RRM. The actual applications to the RRM of metal nanoclusters and organic molecular systems and their interpretations are presented in Section 3, followed by concluding remarks in Section 4.

2 Methods

2.1 RRMs and ADDF/AFIR Methods

Here, an RRM is defined as a collection of EQ structures of a given atomic constitution along with TS structures (i.e., first-order saddle points) connecting pairs of EQs along IRCs, together with their corresponding energies. Because the EQs and TSs on each RRM share a certain atomic constitution, the dissociation channels were not considered. Therefore, an RRM is represented by an undirected graph $G = (V, E)$, with energy weights assigned to both vertices ($w_V(v_i)$, corresponding to the energy of i -th EQ) and edges ($w_E(e_i)$, corresponding to the energy of i -th TS). G may contain loops (i.e., elementary identity reactions) and/or multiple edges (i.e., different elementary reactions with identical reactants and products). In practice, we constructed the RRM of a certain system using either the ADDF or AFIR method implemented in the GRRM program.^{47,48}

The ADDF method¹⁵ follows the anharmonic downward distortion (ADD) of the PES, which is evaluated as the difference between the harmonic potential approximated by the EQ structure and actual potential energy. An efficient method for searching for ADD is to introduce an EQ-centered hypersphere drawn with normal coordinates scaled by 1/2 power of the force constant, over which the energy at the harmonic approximation level is constant. Because each minimum on a scaled hypersphere surface corresponds to the ADD, the reaction route can be followed along the ADD by increasing the hypersphere radius. Although the ADDF method is conceptually well-defined, it requires Hessian computation to construct the scaled hypersphere surface at each EQ, limiting its practical application. Moreover, not all reaction paths can be searched by tracking the ADD paths because of the bifurcation of anharmonic distortion stationary paths.⁴⁹

In contrast, the AFIR method¹⁸ is a more intuitive and practical procedure for constructing RRMs. In this method, an artificial force is introduced between the system fragments to induce a reaction without an energy barrier. The reaction path obtained by minimizing the

potential energy augmented by the artificial force is called the AFIR path. Subsequently, the artificial force term is removed, the maximum energy point is adopted as the approximated TS structure, and the actual TS structure is optimized. In the AFIR method, the applied force is scaled using the parameter γ , which corresponds to the model collision energy. In general, TSs with reaction barriers below γ can be identified using the AFIR method.

The GRRM program was designed to efficiently search for undiscovered reaction paths (EQs and TSs). When the program identifies a candidate EQ or TS structure, it determines whether it is chemically equivalent to a previously identified EQ or TS structure. Thus, it does not distinguish between nuclear permutation isomers and nuclear inversion isomers, i.e., only one of the two chiral isomers (enantiomers) is included in the list of EQs or TSs. This restriction can significantly reduce the cost for the exploration of reaction paths.⁵⁰ However, the unmodified RRM network obtained via GRRM does not correspond to the network on $(3N - 6)$ -dimensional PES—it is contracted using nuclear permutation-inversion symmetry. Note that there is a method for discriminating between chiral isomers (e.g., see Ref. 51).

2.2 Outline of Persistent Homology for PES

This section intends to briefly explain the relationship between PH and PES, as in the previous work by Mirth *et al.*⁴⁵ The explanation here is neither devoted to our method nor mathematically rigorous. However, it provides an intuitive sketch of the PH for the chemical reaction. The rigorous introduction to the proposed method is presented in the next section.

Let x denote a configuration of $3N - 6$ component coordinate, and $\Phi(x)$ denote the potential energy. Suppose the region whose energy is below a i.e. $L(a) = \{x | \Phi(x) \leq a\}$. In other words, for a given topography $\Phi(x)$, $L(a)$ represents the shape of the sea surface with water level a . Figure 1 shows a simplified $\Phi(x)$ example introduced for the explanation. As can be seen in the figure, the shape of the sea surface $L(a)$ itself expands as a increases but retains its topology, except when a crosses a stationary point.

First, when a is below the global minimum of $\Phi(x)$, nothing exists. By increasing a and

crossing the minimum, the first basin appears (first row in Fig. 1). Let the value of energy a at this time be the birth b_0 of this basin with the label "0". As a increases, two basins with labels "1", "2" appear when a crosses the local minima (second and third rows in Fig. 1). These minima are recorded as b_1 and b_2 . Each time a crosses a saddle, the basin with a large birth merges with that with a small birth. In the fourth row in Fig. 1, basins "0" and "2" are merged. In this case, basin "2" has a larger birth b_2 than the basin "0" b_0 . Therefore, we consider that basin "2" merged into basin "0", and basin "0" survived. Thus, the death of basin "2" d_2 has been introduced. In the same way, the death of basins, d_1 will be introduced (fifth row in Fig. 1). The basin "0" survives forever; thus, we define $d_0 = \infty$. The collection $\{(b_i, d_i)\}$ is a so-called persistence diagram. The persistence diagram of the basin is especially called the 0-th persistence diagram.

The merged basins sometimes form a closed path with impassable space inside (sixth row in Fig. 1). Such a closed path is another object of homology. The sixth row in Fig. 1 describes the energy where the path labeled "3" appears. We define energy a as the birth of path b_3 . Finally, the impassable space vanishes when 3 crosses a second-order saddle point (the last row in Fig. 1). We define energy a as the death of path d_3 . Many closed paths can exist in a more complicated PES than the example shown in Fig. 1. The collection $\{(b_i, d_i)\}$ for closed paths is called the 1-st persistence diagram.

However, the description of chemical reactions does not include higher-order PES saddle points within the transition state theory. The GRRM program extracts only the zeroth and first-order saddles. Thus, death in the 1-st persistence diagram cannot be determined. To apply PH to RRM, topological features associated with higher-order saddles should properly be discarded. In other words, it is necessary to develop a new method that targets only RRM and not PES. In the next section, we present a theoretical treatment to construct a PH applicable to RRM data. Here, we introduce a parameter ε to preserve information on detours (e.g., stepwise and concerted pathways) expressed in 1-st persistence diagrams. This is because these pathways, e.g., contribute entropically to the reaction rate.

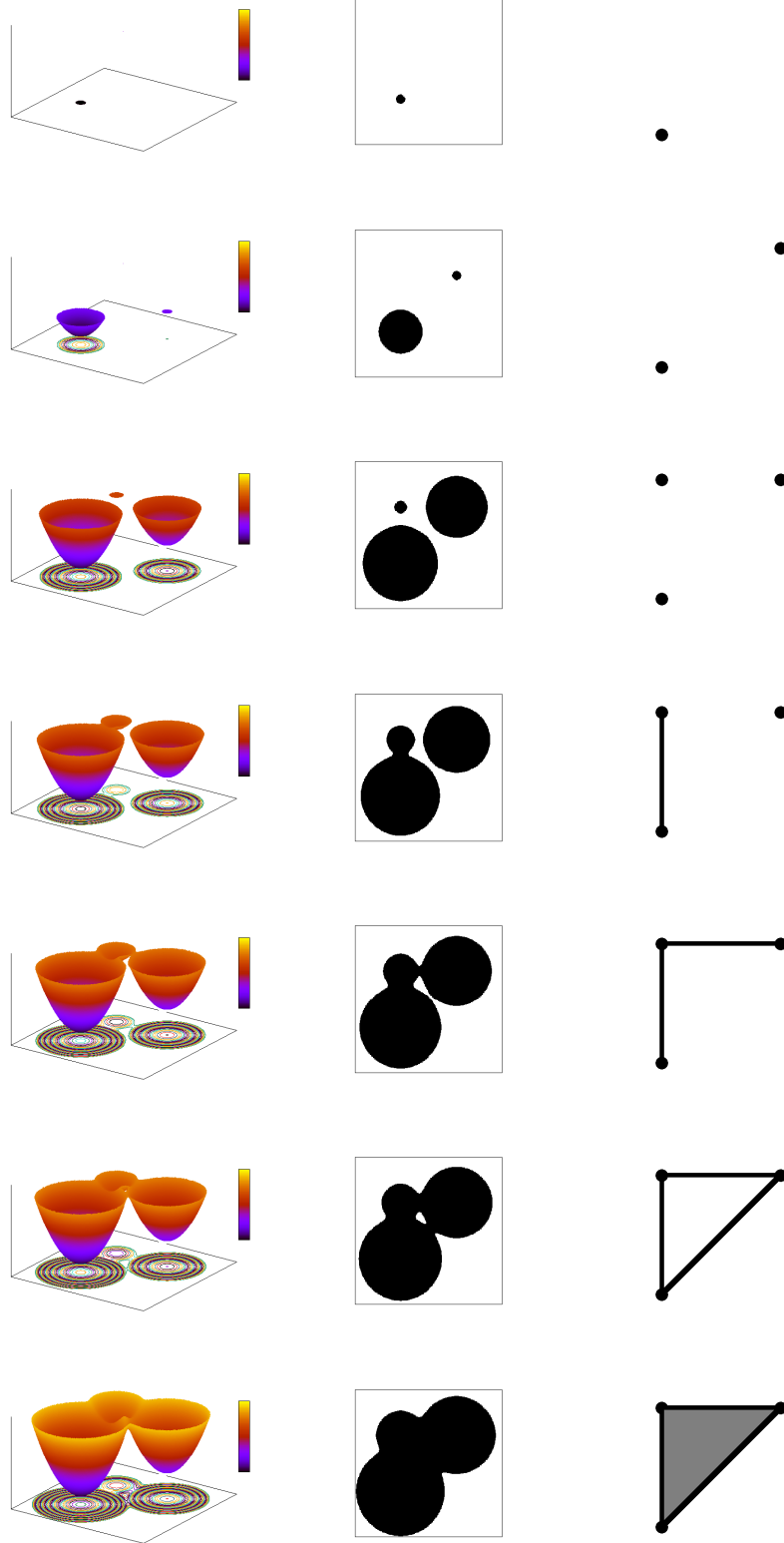


Figure 1: Schematic for an explanation of PH for PES. The lower figure corresponds to greater energy a . The PES below a (surfaces in the left column) with the contour map (bases there), $L(a)$ (the center column), and its topological representation (known as simplicial complex) (the right column) are depicted.

2.3 Rigorous Mathematical Method

In this section, we review the theoretical background of PH for the filtered clique complex of a weighted undirected simple graph, which is a mathematical expression of a simplified RRM. PH is a mathematical tool that extracts topological features from discrete data. In our application such input data are weighted graphs. We constructed a topological space from a weighted graph called a *clique complex*. This is the set of *cliques* and allows us to extract higher-dimensional information from a graph. A nested structure called a *filtration* is required to construct a PH. The clique complex can be equipped with a filtration determined by a weighting of the graph. The information obtained by PH can vary depending on the way the filtration is designed. Our filtration is adjusted using a small constant ε to capture when cliques appear, although there is a natural filtration ($\varepsilon = 0$ case) by a weighting. PH is reviewed in Refs. 52 and 53.

2.3.1 Simple graphs and clique complexes

Before explaining the clique complex, we introduce the terminology of graph theory.^{54,55} A (undirected) simple graph $G = (V, E)$ is a pair of sets: a set of *vertices*, V , and a set of *edges*, $E \subset \{\{u, v\} \subset V \mid u \neq v\}$. By definition, simple graphs do not contain loops or multiple edges. In an RRM, V and E correspond to the set of EQs and TSs, respectively. We denote the edge with endpoints, u and v , as $\overline{uv}$. We assume that the vertex set V is finite (therefore, the edge set E is finite). A pair $G' = (V', E')$ is a *subgraph* of G if $V' \subset V$ and $E' \subset \{\overline{uv} \in E \mid u, v \in V'\}$. We write $G' \subset G$ if G' is a subgraph of G .

Herein, we introduce the concepts of cliques and clique complexes. For an integer $p \geq 2$, a p -*clique* in a graph $G = (V, E)$ is a subset S of V such that S contains p vertices and each distinct pair $u, v \in S$ is connected by an edge in E . A 1-clique is a singleton set $\{v\}$ that contains vertex $v \in V$. The collection of all cliques in G is called the *clique complex* of G , denoted by $X(G)$.

For example, consider the graph $G = (V, E)$ in Fig. 2 (a). The set $\{v_0, v_1, v_2, v_3\}$ in

Fig. 2 (b) is a 4-clique, and $\{v_0, v_5, v_6\}$ in Fig. 2 (c) is a 3-clique because each distinct pair of vertices is connected by an edge (represented by a dashed line). Note that all non-empty subsets of a clique are cliques; e.g., $\{v_0, v_1, v_2\}$ is a 3-clique. In addition, each singleton set $\{v\}$ of vertex $v \in V$ is a 1-clique, and each edge in E is a 2-clique. However, $\{v_0, v_4, v_5, v_6\}$ in Fig. 2 (d) is not a clique since, for instance, v_0 and v_4 are not directly connected.

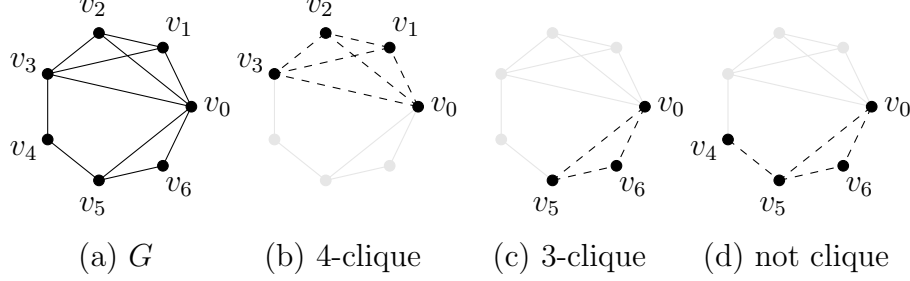


Figure 2: (a) An example of graph G ; in (b), the set of vertices $\{v_0, v_1, v_2, v_3\}$ forms a 4-clique in the graph G since each two vertices are connected by an edge (indicated by a dashed line). Similarly, (c) is a 3-clique. However, (d) is not a clique since v_0 and v_4 are not connected.

2.3.2 Weighting and filtration

To express the energy values in graphs and clique complexes, we introduce *weighting* on a graph, which is defined as a pair of real-valued functions $w_V: V \rightarrow \mathbf{R}$ and $w_E: E \rightarrow \mathbf{R}$. We denote both weight functions by w if it is clear from the context whether w refers to w_V or w_E . For our purposes, we consider only the weight function $w: G \rightarrow \mathbf{R}$ that satisfies the condition

$$w_V(u), w_V(v) \leq w_E(\overline{uv}) \text{ for } \overline{uv} \in E. \quad (1)$$

Subsequently, we define a weighting $w^G: X(G) \rightarrow \mathbf{R}$ on the clique complex as follows: for a p -clique $\sigma = \{v_{i_1}, \dots, v_{i_p}\}$, $(i_1 < \dots < i_p)$,

$$w^G(\sigma; \varepsilon) = \begin{cases} w_V(v_{i_1}) & (p = 1) \\ w_E(\overline{v_{i_1}v_{i_2}}) & (p = 2) \\ \max\{w^G(\tau; \varepsilon) \mid \tau \subsetneq \sigma\} + \varepsilon & (p \geq 3), \end{cases} \quad (2)$$

or equivalently,

$$w^G(\sigma; \varepsilon) = \begin{cases} w_V(v_{i_1}) & (p = 1) \\ \max\{w_E(\overline{v_{i_j}v_{i_k}}) \mid 1 \leq j < k \leq p\} + (p - 2)\varepsilon & (p \geq 2), \end{cases} \quad (3)$$

where ε denotes a parameter with a non-negative real number. Henceforce, $w^G(\sigma; \varepsilon)$ is simply referred to as $w^G(\sigma)$. By construction, the weighting w^G satisfies the following condition:

$$w^G(\tau) \leq w^G(\sigma) \text{ for cliques } \tau \subset \sigma. \quad (4)$$

In this setting, we can construct an increasing sequence corresponding to the clique complex. First, we define

$$X(G)(a) := \{\sigma \in X(G) \mid w^G(\sigma) \leq a\} \quad (5)$$

for a real number $a \in \mathbf{R}$. By the condition of Eq. (4), this is a subset of cliques of G each of whose weights is less than or equal to a .

By construction, if $a < a'$, $X(G)(a) \subset X(G)(a')$. Thus, corresponding to an increasing sequence $a_0 < a_1 < \dots < a_l$ of values of w^G , we obtain an increasing sequence

$$X(G)(a_0) \subset X(G)(a_1) \subset \dots \subset X(G)(a_l). \quad (6)$$

This sequence is called (ε) -*adjusted weight rank clique filtration* on the clique complex $X(G)$ of a weighted graph (G, w) with respect to $\varepsilon \geq 0$. For $\varepsilon = 0$, this filtration coincides with the weight rank clique filtration proposed by Petri *et al.*,⁴⁴ which treats graphs with real-value weights on the edges and all vertices appearing from the beginning. As explained in Sec. 2.5, the introduction of non-zero ε can enrich the information derived from the PH analysis.

Figure 3 depicts an example of a weighting w on graph G depicted in Fig. 2. In the three

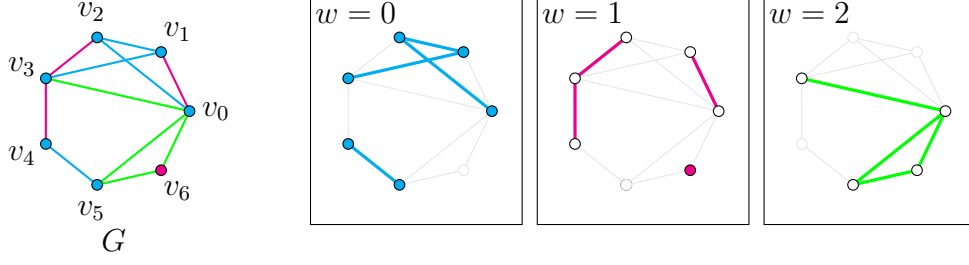


Figure 3: An example of a weighting w on G . Vertices $v_0, v_1, \dots, v_5$ and edges $\overline{v_0v_2}$, $\overline{v_1v_2}$, $\overline{v_1v_3}$, $\overline{v_4v_5}$ have a weight value 0. Similarly a vertex v_6 and edges $\overline{v_0v_1}$, $\overline{v_2v_3}$, $\overline{v_3v_4}$ have a weight 1 (white vertices have already appeared and just indicate the endpoints). Remained edges have 2.

figures on the right, the filled-in vertices and thickened edges have corresponding weights $w = 0, 1$, and 2 , whereas the gray edges and unfilled vertices have different weights. This weighting satisfies the condition (1), i.e., the weight of an edge is equal to or greater than the weight of each of its endpoints. The resulting clique complex of G is a 3-dimensional polyhedron, consisting of a solid tetrahedron (3-dimensional), triangle (2-dimensional), edge (1-dimensional), and vertex (0-dimensional). The weighting of the clique complex $X(G)$ induced by w is determined as follows: Let $\varepsilon > 0$ be sufficiently small, i.e., at least $\varepsilon < 1$. The 1- and 2-cliques have the same weights as the corresponding vertices and edges. For 3-cliques, the weights are as follows:

$$w^G(\sigma) = \begin{cases} 1 + \varepsilon & \text{if } \sigma = \{v_0, v_1, v_2\}, \{v_1, v_2, v_3\} \\ 2 + \varepsilon & \text{if } \sigma = \{v_0, v_1, v_3\}, \{v_0, v_2, v_3\}, \{v_0, v_5, v_6\}. \end{cases}$$

Because the maximal weight of the cliques contained in a 4-clique $\{v_0, v_1, v_2, v_3\}$ is $2 + \varepsilon$, the weight of $\{v_0, v_1, v_2, v_3\}$ is $2 + 2\varepsilon (> 2 + \varepsilon)$. Based on the weighting, w^G , the adjusted weight rank clique filtration on $X(G)$ is of the following form:

$$X(G)(0) \subset X(G)(1) \subset X(G)(1 + \varepsilon) \subset X(G)(2) \subset X(G)(2 + \varepsilon) \subset X(G)(2 + 2\varepsilon) = X(G).$$

In Fig. 4, $X(G)(0)$ comprises two connected components. Vertex v_6 and edges $\overline{v_0v_1}$, $\overline{v_2v_3}$,

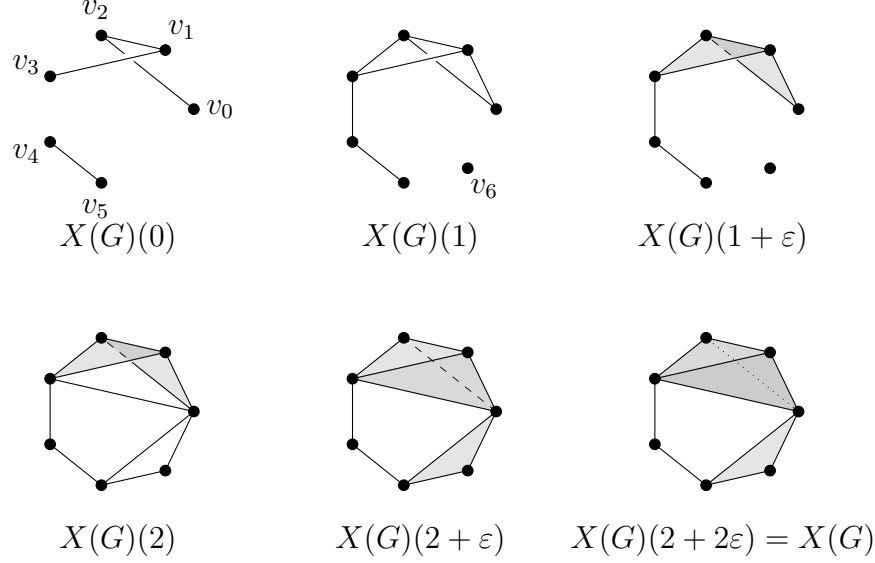


Figure 4: Increasing sequence of $X(G)(a)$ defined in Eq. (5) in the adjusted weight rank clique filtration on $X(G)$ determined by the weighting presented in Fig. 3 when the positive ε parameter is adopted. This is analogous to the sequence of the topological representations (simplicial complexes) in Fig. 1.

and $\overline{v_3v_4}$ appear in $X(G)(1)$. Note that, e.g., 2-cliques $\{v_0, v_1\}$, $\{v_0, v_2\}$ and $\{v_1, v_2\}$ are contained in $X(G)(1)$, while the 3-clique $\{v_0, v_1, v_2\}$ is not since it has a weight greater than 1. The shaded triangles following $X(G)(1 + \varepsilon)$ indicate the corresponding 3-cliques that consist of the vertices of the triangle. $X(G)(2 + \varepsilon)$ does not contain a filling (a 4-clique) in the tetrahedron $v_0v_1v_2v_3$ whereas $X(G)(2 + 2\varepsilon)$ does. Note that $X(G)(1)$ cannot be presented as a clique complex of certain graphs because though the existence of the edges (2-cliques) $\overline{v_0v_1}$, $\overline{v_0v_2}$, and $\overline{v_1v_2}$ implies the existence of the 3-clique $\{v_0, v_1, v_2\}$, $X(G)(1)$ does not contain $\{v_0, v_1, v_2\}$. Similarly, $X(G)(2)$ and $X(G)(2 + \varepsilon)$ cannot be realized as clique complexes. In general, $X(G)(a)$ in the adjusted weight rank clique filtration is not necessarily a clique complex, whereas the weight rank clique filtration ($\varepsilon = 0$) consists of the clique complexes of subgraphs of G .

In particular, the weight rank clique filtration ($\varepsilon = 0$) can be defined by the sequence of subgraphs of graph G . For $a \in \mathbf{R}$, we define

$$V(a) = \{v \in V \mid w(v) \leq a\} \quad \text{and} \quad E(a) = \{e \in E \mid w(e) \leq a\}. \quad (7)$$

By Eq. (4), $G(a) = (V(a), E(a))$ forms a graph. If $a < a'$, then $G(a) \subset G(a')$. For a weighted graph (G, w) and an increasing sequence $a_0 < a_1 < \dots < a_l$, we have the following sequence of subgraphs:

$$G(a_0) \subset G(a_1) \subset \dots \subset G(a_l).$$

By considering the clique complex for each graph $G(a_k)$, we obtain the following sequence of clique complexes:

$$X(G(a_0)) \subset X(G(a_1)) \subset \dots \subset X(G(a_l)).$$

This sequence is equivalent to that in Eq. (6) for $\varepsilon = 0$.

For example, we have a sequence of subgraphs above G , as shown in Fig. 5, using the weighting defined in Fig. 3. The adjusted weight rank clique filtration on $X(G)$ is shown in Fig. 4 and the weight rank clique filtration corresponds to the case, $\varepsilon = 0$, which consists of $X(G(0)) = X(G)(0)$, $X(G(1)) = X(G)(1) \cup X(G)(1 + \varepsilon)$, and $X(G(2)) = X(G)(2) \cup X(G)(2 + \varepsilon) \cup X(G)(2 + 2\varepsilon)$.

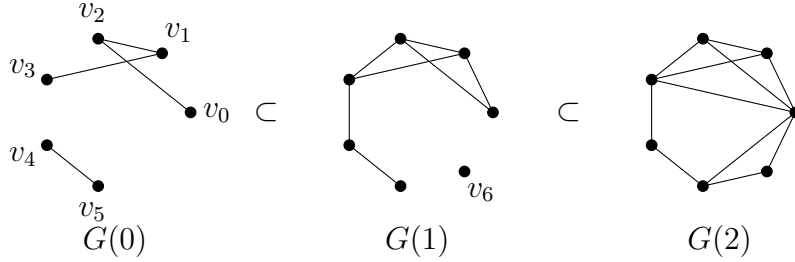


Figure 5: Increasing sequence of subgraphs $G(a) = (V(a), E(a))$ defined by Eq. (7) determined by the weighting presented in Fig. 3.

2.3.3 Persistent homology and its visualization

A clique complex has a *simplicial complex* structure. Given a filtration on a simplicial complex, PH can be defined in general. Here, we briefly summarize the terminology used in PH theory. Further details of this theory can be found in Refs. 52 and 53.

We denote the p -th homology group of a clique complex X with coefficients in $\mathbf{Z}/2\mathbf{Z} =$

$\{0, 1\}$ ($\mathbf{Z}$ is the ring of integers)⁵⁶ by $H_p(X)$. For a filtered clique complex X with filtration

$$X(0) \subset X(1) \subset \cdots \subset X(l),$$

we denote the homomorphism induced by the inclusion $X(i) \subset X(j)$ by $\iota_i^j: H_*(X(i)) \rightarrow H_*(X(j))$. The p -th persistent homology of the filtered clique complex X is the pair, $\{H_p(X(i))\}_i$ and $\{\iota_i^j\}_{i \leq j}$. An element $\alpha \in H_p(X(i))$ is *born* in $X(i)$ ($\alpha \neq 0$) if there is no $\beta \in H_p(X(i-1))$ such that $\iota_{i-1}^i(\beta) = \alpha$, and it *dies* in $X(j)$ ($i < j$) if j is the smallest index for which $\iota_i^j(\alpha) = 0$. It is always possible to define the set of generators of $H_p(X(i))$ following the *elder rule* (like as the first-in-last-out stack, the element born last dies first, i.e., the elements born earlier (*elder* elements) survive), which is simply referred to as the set of *generators* in this paper. If α is born in $X(i)$ but never dies, we say that α *lives forever*.

Each generator α of $H_p(X(i))$ may be associated with the point (a_i, a_j) in the Euclidean plane $\mathbf{R}^2$ if α is born in $H_p(X(i))$ and dies in $H_p(X(j))$. This information can be summarized using persistence diagrams and persistence barcodes. Henceforth, we refer to $b_\alpha = a_i$ as the *birth* and to $d_\alpha = a_j$ as the *death* of the generator α . If α is born in $H_p(X(i))$ and lives forever, we associate α with the point $(a_i, \infty) \in (\mathbf{R} \cup \{\infty\})^2$. The p -th persistence diagram of X , $\mathcal{D}_p(X)$, is the multiset of points $(a_i, a_j) \in (\mathbf{R} \cup \{\infty\})^2$ corresponding to generators that are born in $H_p(X(i))$ and either die in $H_p(X(j))$ or live forever. The p -th persistence barcode of X , $\mathcal{B}_p(X)$, is the multiset of half-open intervals $[a_i, a_j) \subset \mathbf{R} \cup \{\infty\}$ corresponding to generators that are born in $H_p(X(i))$ and either die in $H_p(X(j))$ or live forever.

In the case of the filtered clique complex constructed from the sequence of subgraphs in Fig. 5, 0-dimensional homology classes correspond to the connected components in $G(0)$. These are born in $H_0(X(0))$, two of them die in $H_0(X(1))$, and the class corresponding to v_6 is born in $H_0(X(1))$ and dies in $H_0(X(2))$. The 1-dimensional homology class corresponding to rectangle $v_0v_3v_4v_5$ is born in $H_1(X(2))$ and lives forever. Persistence diagrams and barcodes were obtained for this example (Fig. 6). The length of the bar corresponding to point

(a_i, a_j) in the persistence diagram is $a_j - a_i$. This coincides with the length of the line segment between (a_i, a_j) and the intersection of the diagonal and vertical lines from (a_i, a_j) . Here, the elder rule is used to draw persistence diagrams. According to this rule, when two components are merged, the previously born component is retained.

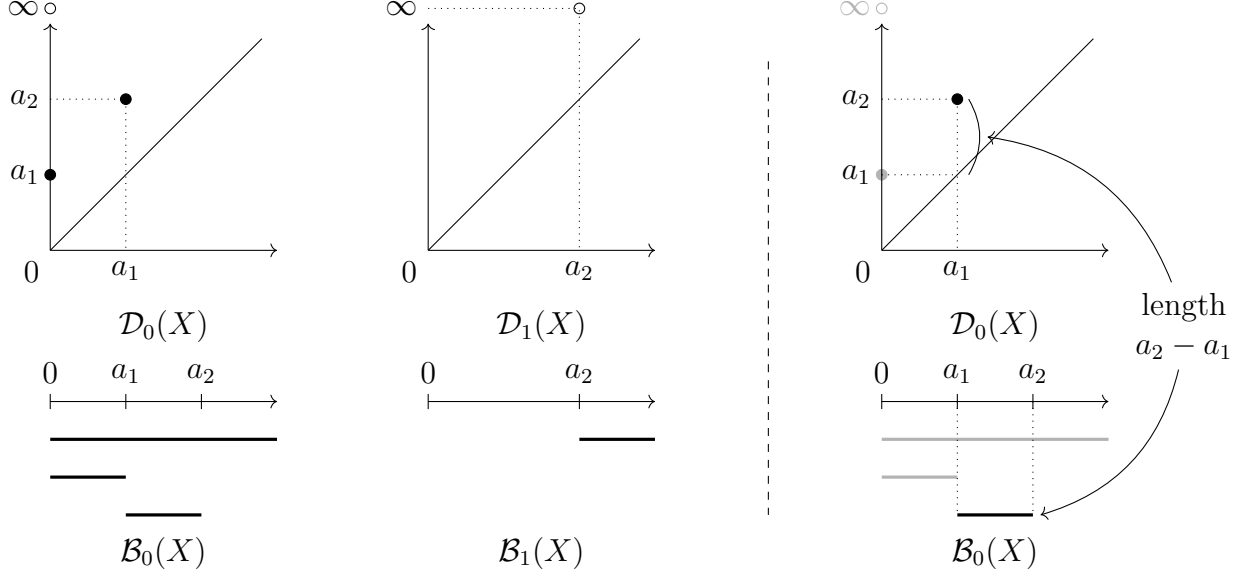


Figure 6: 0-th and 1-st persistence diagrams and barcodes of the clique complex $X = X(G)$ constructed from a weighted graph (G, w) in Fig. 3

In the following argument, we focus on the 0-th and 1-st PH and their diagrams. To make the figures concise, we depict the 0-th and 1-st persistence diagrams together—the upper half region presents the 0-th persistence diagram (horizontal axis represents birth and vertical axis represents death) and the lower half region presents the 1-st persistence diagram (vertical axis represents birth and horizontal axis represents death) in Fig. 7.

2.4 Approximation of the entire RRM graph using a simple graph

As the entire RRM graph, G^{entire} may include loops and/or multiple edges for most systems, the adjusted weight rank clique filtration cannot be applied to G^{entire} directly. However, as an approximate treatment, the filtration can be applied to a simple subgraph G associated with the entire RRM graph G^{entire} . This was realized by eliminating loop edges and multiple

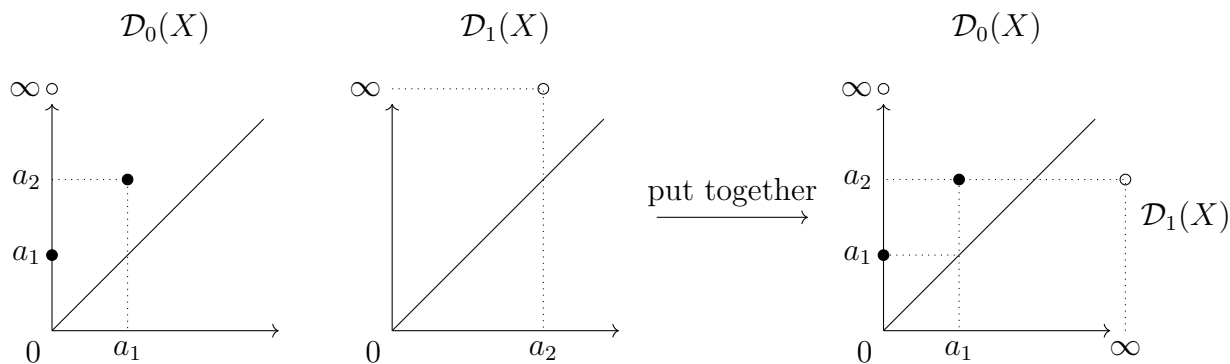


Figure 7: Concise representation of the 0-th and 1-st PD in Fig. 6.

edges other than those with the lowest energy weights (Fig. 8).

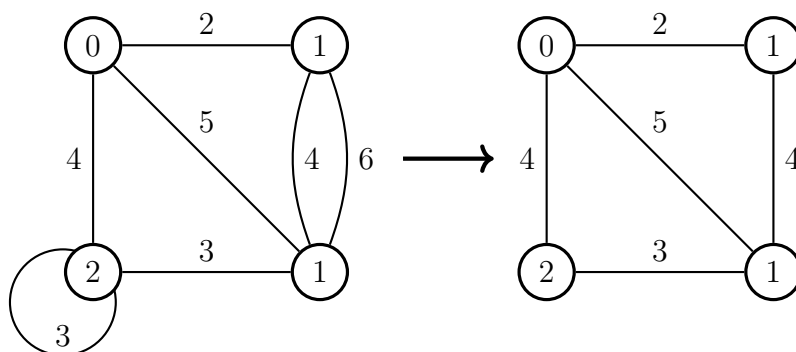


Figure 8: (left) RRM containing loops and multiple edges. The number associated with each vertex or edge expresses its energy (or equivalently, weight). (right) Its simplified RRM.

When the weight on each vertex or edge represents the corresponding potential energies, the requirement of Eq. (1) is satisfied because the potential energy at the TS is higher than that at EQs connected by the corresponding IRC. Note that this requirement may not be satisfied when the corrected energy, such as the Gibbs free energy, is represented by the weight value.

2.5 Demonstration of the proposed method on the RRM of a toy model

To illustrate the information extracted by the proposed method, its application to the RRM of a toy model was considered (Fig. 9(a)). The vertices and edges represent the EQ and

TS structures, respectively. For the RRM, $G = (V, E)$, $V = \{v_1, v_2, v_3, v_4\}$, and $E = \{\overline{v_1 v_2}, \overline{v_1 v_3}, \overline{v_1 v_4}, \overline{v_2 v_3}, \overline{v_3 v_4}\}$. The weights attributed to the vertices and edges correspond to the relative energies of the corresponding structures with respect to the global minimum energy and are referred to as $w_V(v_1) = 0$ and $w_E(\overline{v_1 v_2}) = 2$. The first generator of 0-th PH is born at $a = 0$, where $V(0) = \{v_1\}$ and $E(0) = \{\}$. Fig. 9(b) depicts the sequence for $G(a)$. Because this first connected component does not vanish, even when $a \rightarrow \infty$, this generator lives forever. This is plotted as $(b, d) = (0, \infty)$ in $\mathcal{D}_0$ (Fig. 9(c)), whereas it is expressed as the half-line starting from $a = 0$ in $\mathcal{B}_0$ (Fig. 9(d)). Note that the blue (or red) color indicates that the generator is of the 0-th (or 1-st) PH in these representations. The second and third generators are born simultaneously at $a = 1$, where $V(1) = \{v_1, v_2, v_3\}$ and $E(1) = \{\}$. At $a = 2$, one of these generators dies because edge $\overline{v_1 v_2}$ is included in $E(2)$. This is plotted as $(b, d) = (1, 2)$ in $\mathcal{D}_0$ and is expressed as a line segment between $a = 1$ and 2 in $\mathcal{B}_0$. Note that the dying generator is not born at $a = 0$ because the elder rule is adopted. Simultaneously, another generator derived from vertex v_4 is born. This dies at $a = 3$ by edge $\overline{v_3 v_4}$ and is plotted as $(b, d) = (2, 3)$ in $\mathcal{D}_0$. The other generator born in $a = 1$ dies at $a = 4$ by the edge $\overline{v_2 v_3}$ (or $\overline{v_1 v_4}$). Simultaneously, the generator of the 1-st PH is born because $E(4) = \{\overline{v_1 v_2}, \overline{v_1 v_4}, \overline{v_2 v_3}, \overline{v_3 v_4}\}$ constitutes a ring. At $a = 5$, this ring is split into two rings by edge $\overline{v_1 v_3}$ and another generator of the 1-st PH is born. Note that, if ε in Eq. (2) is not zero, the 3-clique, $\{v_1, v_2, v_3\}$, e.g., is not included in $X(G)(5)$. Then, these generators of the 1-st PH die in $a = 5 + \varepsilon$ where 3-cliques $\{v_1, v_2, v_3\}$ and $\{v_1, v_3, v_4\}$ are included in $X(G)(5 + \varepsilon)$. These are plotted as $(b, d) = (4, 5 + \varepsilon)$ and $(b, d) = (5, 5 + \varepsilon)$ in $\mathcal{D}_1$. The introduction of non-zero ε enriches the information extracted from PH analysis because when $\varepsilon = 0$, only $(b, d) = (4, 5)$ is detected.

As the 0-th PH represents the number of connected components corresponding to a certain weight, a , an obvious hierarchy exists in the inclusion of vertices. For example, vertex v_4 is unified into vertex v_3 for $a = 3$. In the persistence barcode representation, this relation corresponds to a short bar and a longer bar that includes the short bar entirely. For example,

bar $(b, d) = (2, 3)$ is unified into bar $(b, d) = (1, 4)$. Therefore, a long bar corresponds to a *superbasin*,^{1,42,57} defined as a collection of EQs connected by low barriers. In the persistence diagram representation, this relationship is plotted on the rising left shoulder line.

Fig. 10 summarizes the flowchart for obtaining the persistence barcode (or equivalently, the persistence diagram) from a list of EQs and TSs. The code used to obtain the persistence barcode and diagram from the GRRM program output is available on GitHub.⁵⁸

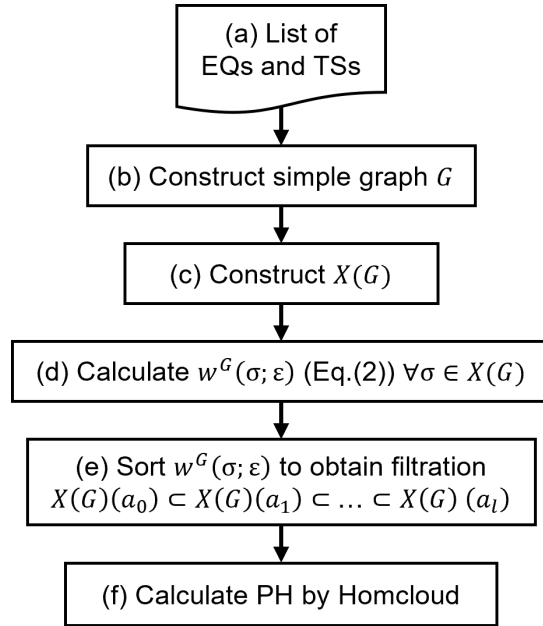


Figure 10: Flowchart to obtain the persistence barcode.

2.6 Difference from the sublevel set PH of PES

In Ref. 45, Mirth *et al.* applied sublevel set PH to the two-dimensional model potential of n -pentane. In their sublevel-set PH, a sequence of subspaces of the $(3N - 6)$ -dimensional PES, in which the energy is smaller than the threshold a , is constructed. Therefore, all stationary points (not only EQs and TSs but also those with a higher index) in the PES must be located. Their results revealed that the sublevel-set PH analysis accurately reflects the shape of the PES and produces better descriptors than DG. In this section, we apply the adjusted weight rank clique filtration to the partial RRM of n -pentane, which is constructed

based on the GRRM-GDSP database by excluding all bond rearrangement pathways, and compare it with the result of Mirth *et al.* For comparison, a partial RRM was adopted. The unmodified partial RRM obtained from GRRM-GDSP⁵ is shown in Fig. 11(a). Although the network size is very compact in this case, it only spans a portion of the full network owing to the nuclear inversion symmetry. Reconstructing the full network yielded the RRM depicted in Fig. 11(b). The proposed method was applied to this full network. In this simple case, the full network was reconstructed manually. A general method for reconstructing a full network by considering the symmetry of the PES is currently under construction and will be reported in a subsequent paper.⁵⁹

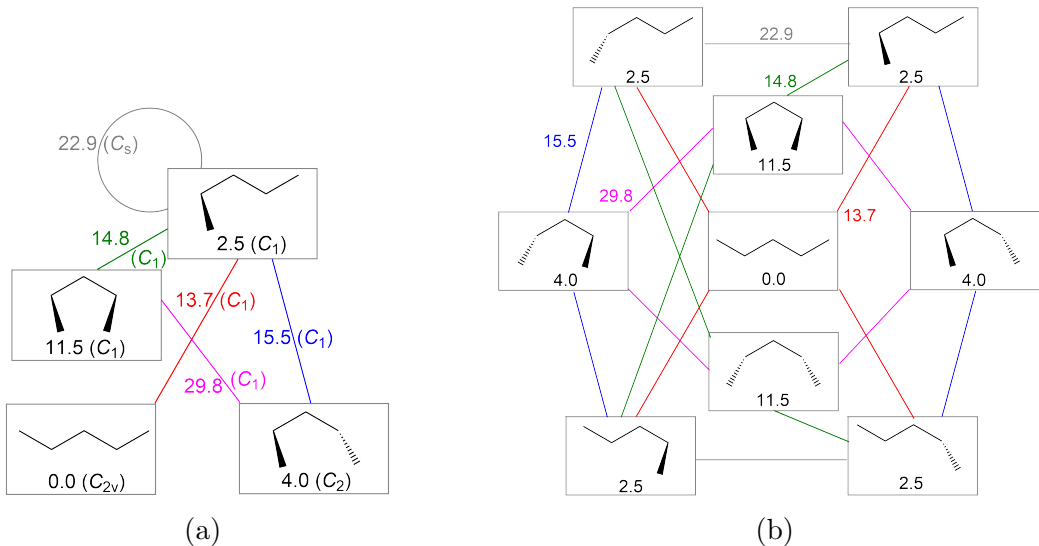


Figure 11: (a) Partial RRM of *n*-pentane obtained from GRRM-GDSP database and (b) its reconstructed network considering nuclear inversion symmetry.

Figure 12(a) depicts the persistence barcode obtained from the adjusted weight rank clique filtration of the RRM (Fig. 11(b)). Figure 12(b) is the same figure, but for the model potential adopted in Ref. 45. For comparison, the persistence barcode obtained by filtration based on the sublevel sets by Mirth *et al.* was reprinted (Fig. 12(c)). Here, we denote the vertex and edge with weights w as $v(w)$ and $e(w)$, respectively (when expressed as $V(w)$ and $E(w)$ with a capital letter, they represent the set collected up to the threshold w). For the 0-th persistence barcode ($\mathcal{B}_0$), the bars in Figs. 12(a) and (b) show qualitative agreement,

although the bar lengths were different. However, the bars are identical for Figs. 12(b) and (c). This clearly shows that the current filtration can extract the same information as the sublevel-set filtration if the adopted potential is the same. Specifically, for the first four bars other than the top semi-infinite bar, the 0-th persistence barcode corresponds to $v(2.5)$ and $e(13.7)$ (red edges) in Fig. 11(b). In the case of the model potential conducted by Mirth *et al.*, the energies of $v(11.5)$ and $e(14.8)$ (green edges) coincide with those of $v(4.0)$ and $e(15.5)$ (blue edges), respectively, owing to the lack of steric repulsion. Therefore, the next four bars in $\mathcal{B}_0$ correspond to the green edges and their associated vertices. By accounting for steric repulsion, which was not considered in the model potential, the degeneracy of the bars was partially resolved using the proposed method. As for the 1-st persistence barcode ($\mathcal{B}_1$), the number of bars is identical among Figs. 12(a), (b), and (c), while their births in Fig. 12(a) differ from those shown in Figs. 12(b) and (c) owing to the difference in the adopted potential. For $\mathcal{B}_1$ in Fig. 12(a), the first two bars and the next two bars correspond to cycles $(v(0.0), v(2.5), v(11.5), v(2.5))$ and $(v(0.0), v(2.5), v(4.0), v(2.5))$, respectively, which also degenerate in Fig. 12(b). In Fig. 12(c), these bars die because of the second-order saddle point, although they do not die in Figs. 12(a) and (b) owing to the lack of a direct pathway between $v(0.0)$ and $v(11.5)/v(4.0)$ or $v(2.5)$ and another $v(2.5)$. The next two bars correspond to cycles $(v(0.0), v(2.5), v(2.5))$. By Figs. 12(a) and (b), the length of these bars is ε because they are 3-cliques. In contrast, in Fig. 12(c), the bars do not die because the configuration space of the model potential adopted by Mirth *et al.* is a two-dimensional torus.⁴⁵ For the final four bars corresponding to cycles $(v(2.5), v(4.0), v(11.5))$, those depicted in Fig. 12(c) die by the second-order saddle point, whereas the lengths in Figs. 12(a) and (b) are ε because they are 3-cliques. Note that in the low-temperature limit, or in the range where the transition state theory is valid, because the second-order saddle point does not affect the kinetics, the difference between Figs. 12(b) and (c) does not affect the interpretation of the chemical kinetics. The death of the bar in the 1-st persistence barcode is just a descriptor defined as a quantity characterizing the PES or RRM.

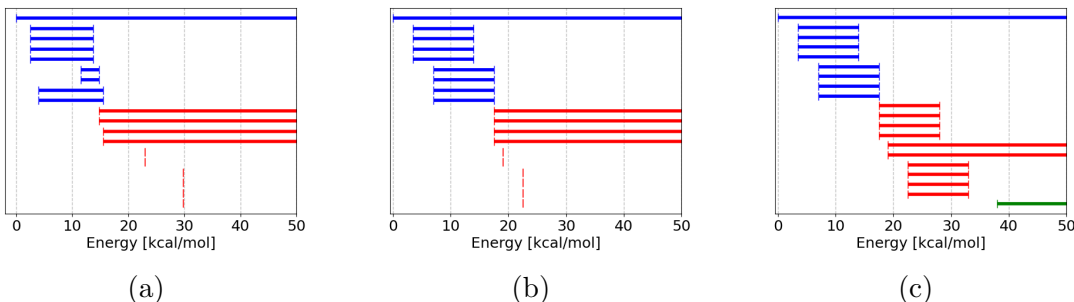


Figure 12: Persistence barcodes of *n*-pentane obtained using (a,b) adjusted weight rank clique filtration for (a) DFT level energy and (b) model potential and (c) sublevel set PH for model potential (from Ref. 45, green bar corresponds to the 2-nd persistence barcode)

2.7 Computational Details

RRMs of the metal nanoclusters were constructed using the SC-AFIR method implemented in the GRRM program. The force parameter $\gamma = 300 \text{ kJ mol}^{-1}$ is adopted during the search for the reaction path. Electronic structure calculations were performed using the Gaussian16 program⁶⁰ with the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional⁶¹ and the LanL2DZ basis set associated with the effective core potential.⁶² The RRM of the organic molecules were obtained from Prof. K. Ohno at the Institute for Quantum Chemical Exploration (IQCE) and are available on the GRRM-GDSP website.⁵ The level of the electronic structure calculations was designated as B3LYP/6-31G(d). The RRM for the Claisen rearrangement reported in Ref. 63, where the geometry optimizations were performed using the M06-2X exchange-correlation functional⁶⁴ with the 6-311+G(2d,p) basis set^{65,66} and the single-point energies were recalculated using the CCSD(T) method alongside the jun-cc-pVTZ basis set,^{67,68} was provided by Prof. Maeda and Dr. Nagahata. A program to reproduce our methods is available in Ref. 58.

3 Results and Discussion

In this section, we describe the chemical interpretation of the information obtained from the persistence barcode (or equivalently, the persistence diagram).

3.1 Persistence barcode/diagram for RRM of metal nanoclusters

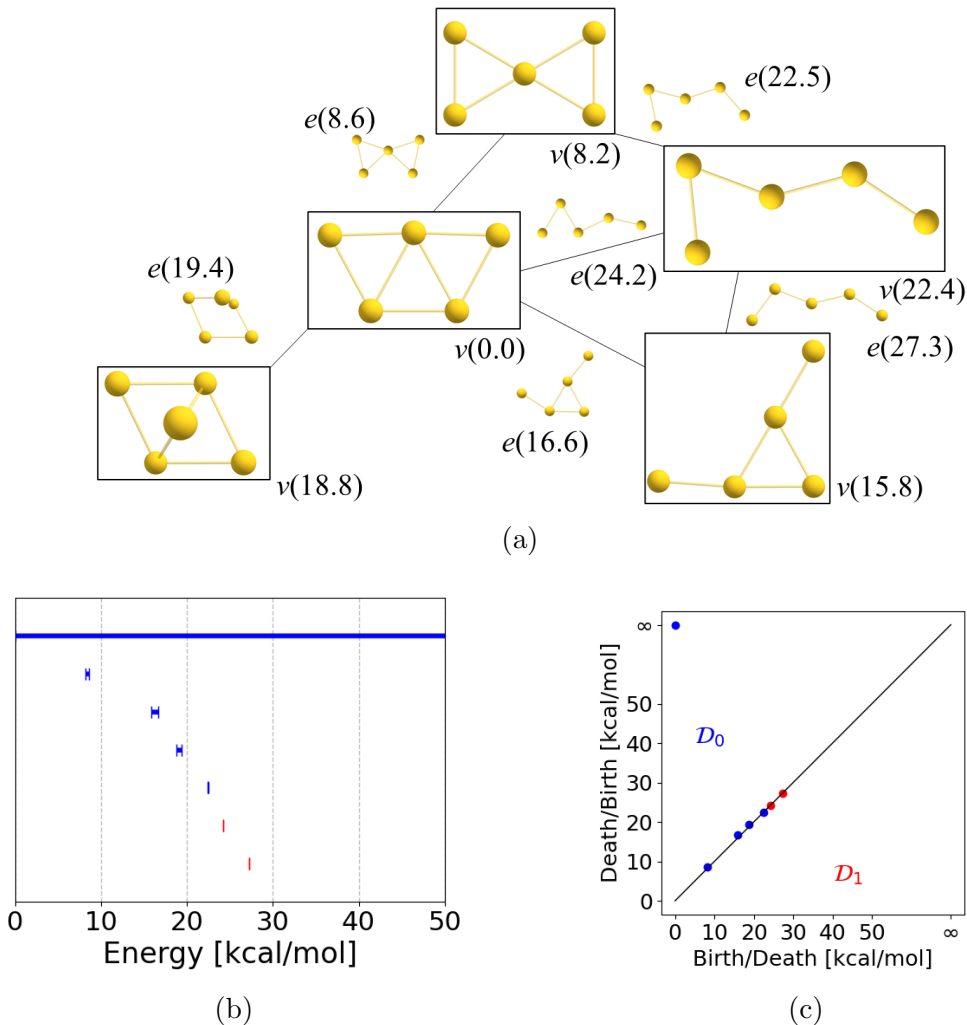


Figure 13: (a) RRM of Au_5 with corresponding structures, (b) its persistence barcode, and (c) its persistence diagram.

First, we focus on the RRM of the Au_5 cluster, which was investigated in detail in Ref. 69, and is illustrated in Fig. 13 (a). The persistence barcode constructed using the proposed method is shown in Fig. 13 (b). The number attributed to each vertex or edge is the

relative energy in kcal mol⁻¹. In this section, each vertex, v , or edge, e , is designated by its energy value; for instance, $v(0.0)$ refers to the vertex with an energy value of 0.0 kcal mol⁻¹ corresponding to the W-shaped C_{2v} structure. In this RRM, $v(0.0)$ is the global minimum and is connected to all other vertices. At $a = 8.2$ (henceforth, the unit of energy is omitted), $v(8.2)$ corresponding to the X-shaped D_{2h} structure appears, but is disconnected from the former vertex, $v(0.0)$. They are merged at $a = 8.6$, leading to a generator of the 0-th PH with $(b, d) = (8.2, 8.6)$. This generator of the 0-th PH indicates that a potential basin with a minimum energy of 8.2 kcal mol⁻¹ is incorporated within another potential basin (in this case, $v(0.0)$) if the energy parameter a is raised to 8.6 kcal mol⁻¹. Similarly, the generators of the 0-th PH with $(b, d) = (15.8, 16.6)$, $(18.8, 19.4)$, and $(22.4, 22.5)$ were obtained from this RRM. The lengths of the bars of these generators (i.e., $d - b$) were significantly small. This implies that the basins corresponding to these EQs [$v(8.2)$, $v(15.8)$, $v(18.8)$, and $v(22.4)$] are very shallow. It should be noted that this persistence barcode (and persistence diagram) cannot directly provide information for identifying the basin to which the merged basin is incorporated—this can be expressed in the disconnectivity graph. At $a = 24.2$, $v(0.0)$ is directly connected to $v(22.4)$, whereas they are already connected via $v(8.2)$ at $a = 22.5$. The emergence of two different routes (i.e., stepwise and concerted pathways) causes the birth of a generator of the 1-st PH. Because these three vertices share a complete connection at $a = 24.2$, the generator dies immediately at $a = 24.2 + \varepsilon$. The case of the generator with $(b, d) = (27.3, 27.3 + \varepsilon)$, which originates from vertices $v(0.0)$, $v(15.8)$, $v(22.4)$, is similar. For an intuitive understanding, we provide a one-to-one correspondence between the element of the RRM and the bar, as depicted in Fig. S1 in the Supporting Information.

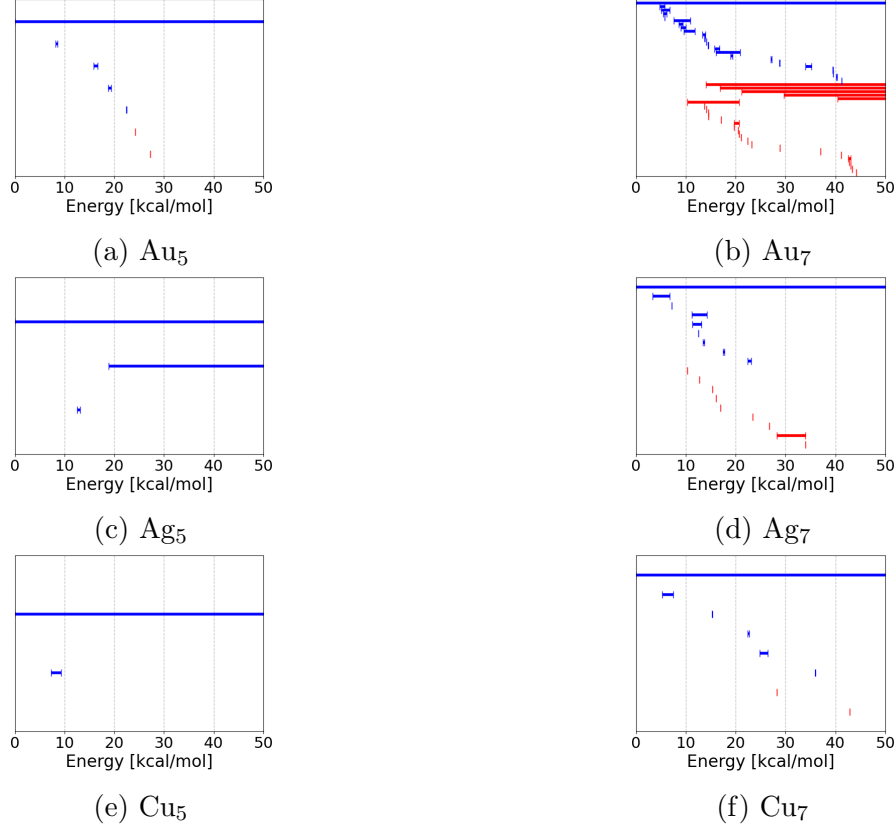
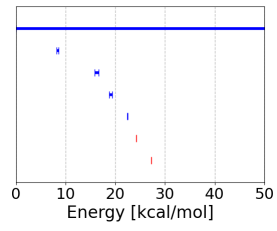


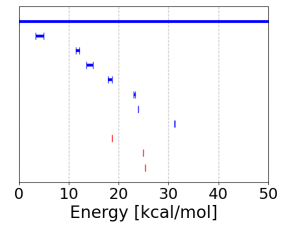
Figure 14: Persistence barcodes for RRM of (a) Au₅, (b) Au₇, (c) Ag₅, (d) Ag₇, (e) Cu₅, and (f) Cu₇ systems.

Figure 14 depicts the persistence barcodes for the RRM of single-element coinage metal clusters with five and seven atoms. If the number of atoms is fixed, the number of generators of PH increases towards the bottom of the periodic table (the number of bars and the average and standard deviation (SD) of bar lengths for each barcode are summarized in Table S1 in the Supporting Information). As the number of generators of the 0-th PH is related to the number of EQ geometries, this tendency can be interpreted to be indicative of the hardness of the element, i.e., softer elements have more generators. For the same element, the number of generators increases as the number of atoms increases. Also, the number of long bars in the persistence barcode tends to increase with an increase in the number of atoms. This implies that the number of metastable cluster structures that can persist over a lengthy period increases as the cluster size increases.

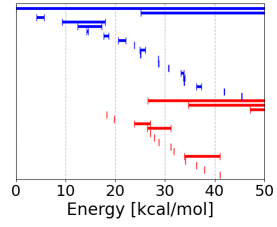
Figure 15 depicts the persistence barcodes for the RRM of the five-atomic alloy clusters



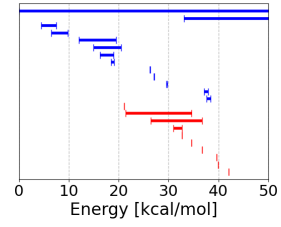
(a) Au_5



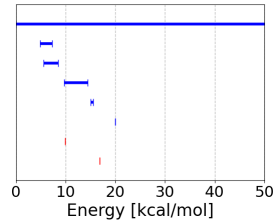
(b) Au_4Cu



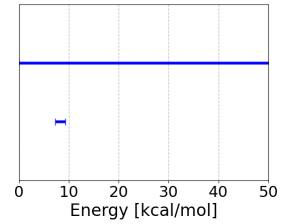
(c) Au_3Cu_2



(d) Au_2Cu_3



(e) AuCu_4



(f) Cu_5

Figure 15: Persistence barcodes for RRM of (a) Au_5 , (b) Au_4Cu , (c) Au_3Cu_2 , (d) Au_2Cu_3 , (e) AuCu_4 , and (f) Cu_5 .

of Au and Cu. The results corresponding to all possible mixing patterns (i.e., Au_5 , Au_4Cu , Au_3Cu_2 , Au_2Cu_3 , AuCu_4 , and Cu_5) are presented. As the mixing ratio becomes more uniform (i.e., $X_5 \rightarrow X_4Y \rightarrow X_3Y_2$), the number of generators of the PH increases. This is related to the number of permutation isomers, which increases as the mixing ratio becomes more uniform. Interestingly, the average length of the bars in the barcode increases as the mixing ratio becomes more uniform, indicating an increase in the number of metastable cluster structures persisting over long intervals.

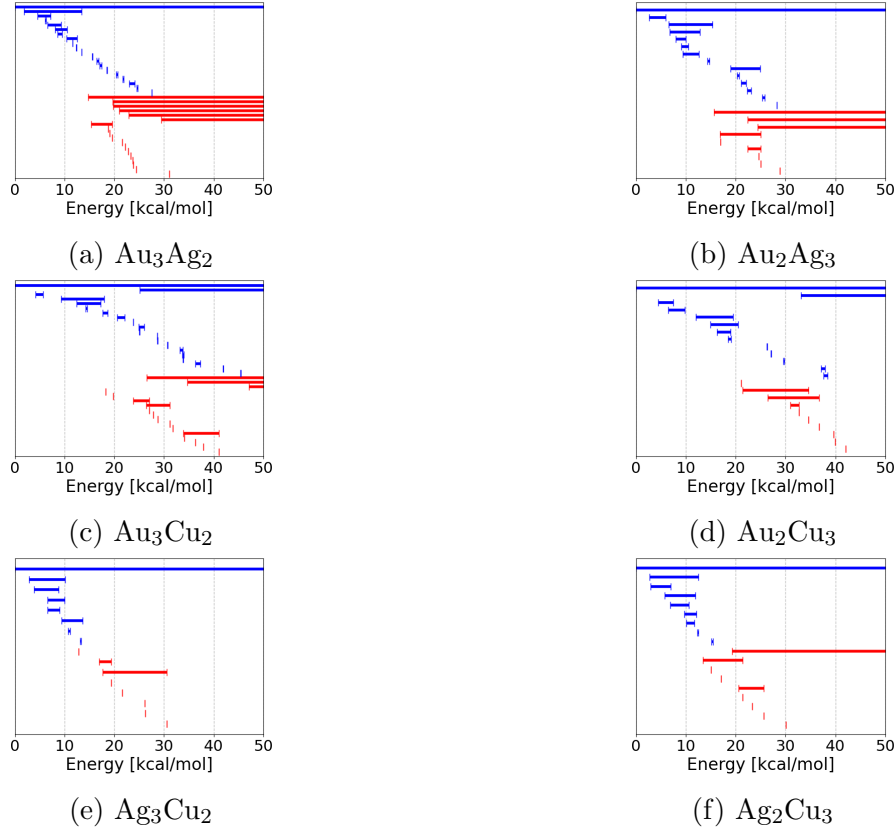


Figure 16: Persistence barcodes for RRM of (a) Au_3Ag_2 , (b) Au_3Cu_2 , (c) Ag_3Cu_2 , (d) Au_2Ag_3 , (e) Au_2Cu_3 , and (f) Ag_2Cu_3

Figure 16 compares the persistence barcodes for the RRM of X_3Y_2 ($X, Y = \text{Cu}, \text{Ag}, \text{or Au}$) corresponding to different elements. For the same combination of elements (i.e., X_3Y_2 and X_2Y_3), the persistence barcodes tend to exhibit similar trends. For the combination of Au and Cu, the generators of PH appear over a wide energy range, extending up to 50 kcal

mol^{-1} , while for other combinations, they mostly appear at up to 30 kcal mol^{-1} . The number of generators is comparatively lower for the combination of Ag and Cu than for the others. On the contrary, for the combination of Au and Ag, the number of generators, especially those of the 1-st PH living forever, is higher than those of the others. This suggests that the analysis can be used to extract the affinity of two metal elements or the properties of metal alloys.

The RRM for Ag_3Cu_2 is depicted in Fig. 17. Vertices $v(2.9)$ and $v(3.9)$ appear at $a = 2.9$ and $a = 3.9$, respectively. Each of $v(0.0)$, $v(2.9)$, and $v(3.9)$ belongs to a different connected component from the others up to $a = 8.8$. Correspondingly, the generators of the 0-th PH are born at $a = 2.9$ and $a = 3.9$. At $a = 8.8$, $e(8.8)$ connects $v(2.9)$ and $v(3.9)$ and the generator born at $a = 3.9$ dies based on the elder rule. In this case, unlike in the case of Au_5 , the persistence barcode does not uniquely determine whether $v(3.9)$ is connected to $v(0.0)$ or $v(2.9)$. At $a = 12.9$, a loop consisting of three vertices ($v(0.0)$, $v(2.9)$, and $v(10.7)$) appears, and a generator of the 1-st PH is born. Because this loop is a 3-clique, this generator dies at $a = 12.9 + \varepsilon$. At $a = 17.0$, a loop consisting of four vertices ($v(0.0)$, $v(2.9)$, $v(3.9)$, and $v(13.2)$) appears, and a generator of the 1-st PH is born. The generator does not die immediately. At $a = 19.4$, $v(0.0)$ is directly connected to $v(3.9)$ via $e(19.4)$ and $v(0.0), v(2.9), v(3.9)$ and $v(0.0), v(3.9), v(13.2)$ constitute two 3-cliques. Then, another generator of the 1-st PH is born at $a = 19.4$ and the generators of the 1-st PH die at $a = 19.4 + \varepsilon$. As discussed above, the generator of the 1-st PH provides information regarding the detour in the reaction pathway.

For comparison, the DG of the RRM of Ag_3Cu_2 (Fig. 17), which was illustrated using PyConnect,⁷⁰ is depicted in Fig. 18. DG complements the persistence barcode of the 0-th PH by providing inclusion relationships between the two potential basins. For example, it is evident from the DG that vertex $v(3.9)$ is connected to $v(2.9)$ at $a = 8.8$. However, as mentioned previously, DG cannot provide higher-order information about connections, such as loop structures. Both DG and persistent homology methods provide descriptors regarding the topology of the RRM.

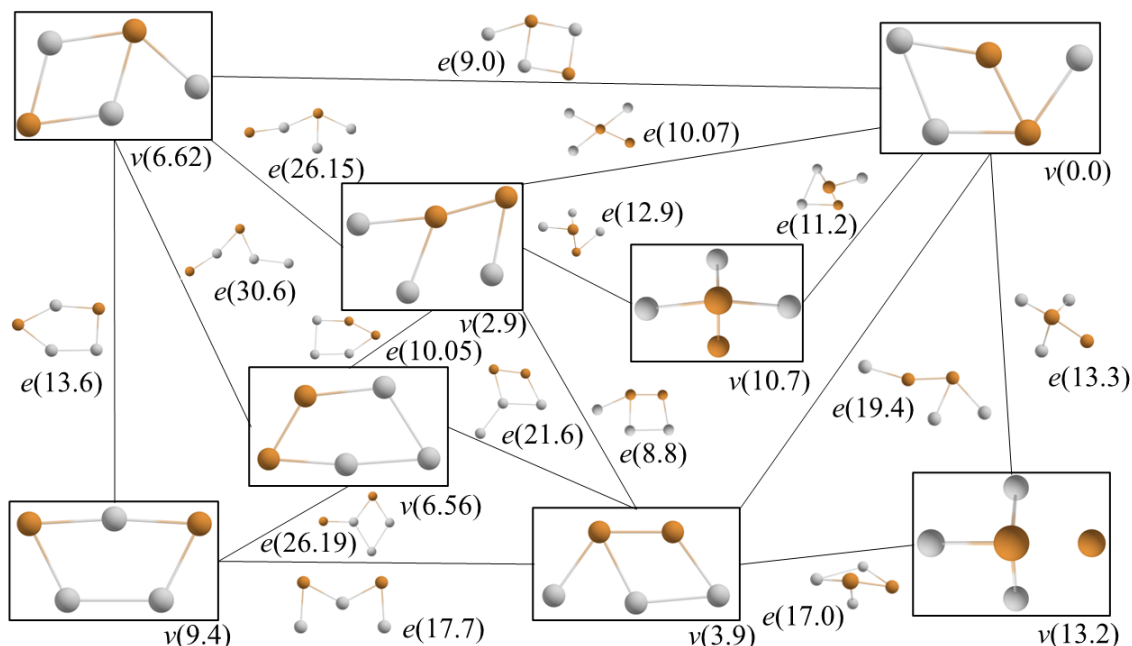


Figure 17: RRM of Ag_3Cu_2 with their structures. This corresponds to Fig. 16 (e)

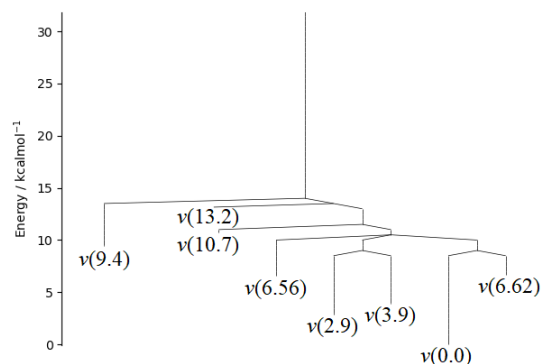


Figure 18: DG for the RRM of Ag_3Cu_2 . DG provides inclusion relationships between the two potential basins. However, DG cannot provide information on loop structures.

3.2 Persistence diagram for RRM of organic molecules

In this subsection, the proposed method is applied to RRM of organic molecules. The diagram representation of PH is used because the number of generators in this case can be significantly larger than that of the metal clusters. Again, note that the bar length in $\mathcal{B}_0$ (or $\mathcal{B}_1$) corresponds to the vertical (or horizontal) distance from the diagonal line in $\mathcal{D}_0$ (or $\mathcal{D}_1$).

First, the size dependencies of the persistence diagrams are compared corresponding to identical chemical element ratios. Fig. 19 depicts persistence diagrams for the RRM of H_2CO and $\text{H}_4\text{C}_2\text{O}_2$. The number of generators of both the 0-th and 1-st PH significantly increases as the number of atoms is doubled. However, the plots appearing in the persistence diagram of H_2CO can be found at positions like those of $\text{H}_4\text{C}_2\text{O}_2$. In addition, the number of plots, apart from the diagonal line, increases as the system size increases. This trend was also confirmed for the metal nanocluster (Fig. 14).

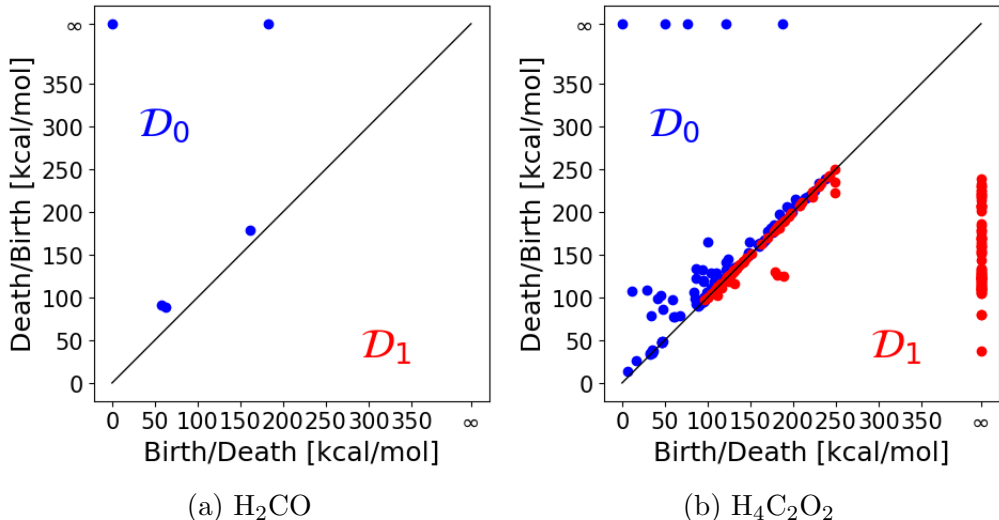


Figure 19: Persistence diagrams for RRM of (a) H_2CO and (b) $\text{H}_4\text{C}_2\text{O}_2$

Next, the persistence diagrams for RRM of H_4C_2 , $\text{H}_2\text{C}_2\text{N}_2$, and HC_2NO_2 are compared, where identical numbers of atoms but different numbers of chemical elements are considered (Fig. 20). For H_4C_2 , there are only 2 EQs, namely, ethylene ($\text{H}_2\text{C}=\text{CH}_2$) and HCCH_3 —the energy of the latter is $76.1 \text{ kcal mol}^{-1}$ higher than that of ethylene. Therefore, in this case, there exists only one generator of the 0-th PH (except for the obvious one with $(b, d) = (0, \infty)$) corresponding to the TS between these two EQs and there is no generator of the 1-st PH. As the number of elements increases, the range of the plots gradually expands. Simultaneously, the numbers of generators of both the 0-th and 1-st PH increase significantly. In particular, the number of plots far from the diagonal increases, which can be confirmed from the average bar length in $\mathcal{B}_1$) provided in Table S1 in the Supporting Information, while

the trend is not obvious in $\mathcal{D}_0$ for $\text{H}_2\text{C}_2\text{N}_2$ and HC_2NO_2 . This indicates that the number of stable isomers increases as the number of elements increases, even when the number of atoms is maintained constant. However, it should be noted that nuclear permutation and inversion symmetry is not considered in the current application. As discussed in Sec. 2.6, the persistence diagram changes when symmetrically degenerated vertices and edges are resolved. In addition, in $\mathcal{D}_1$, the plot with $d = \infty$ increases significantly as the number of elements increases. This may indicate that the two different reaction pathways connecting a pair of EQs do not primarily consist of direct (concerted) and two-step pathways in RRM of organic molecular systems compared to those of metal clusters.

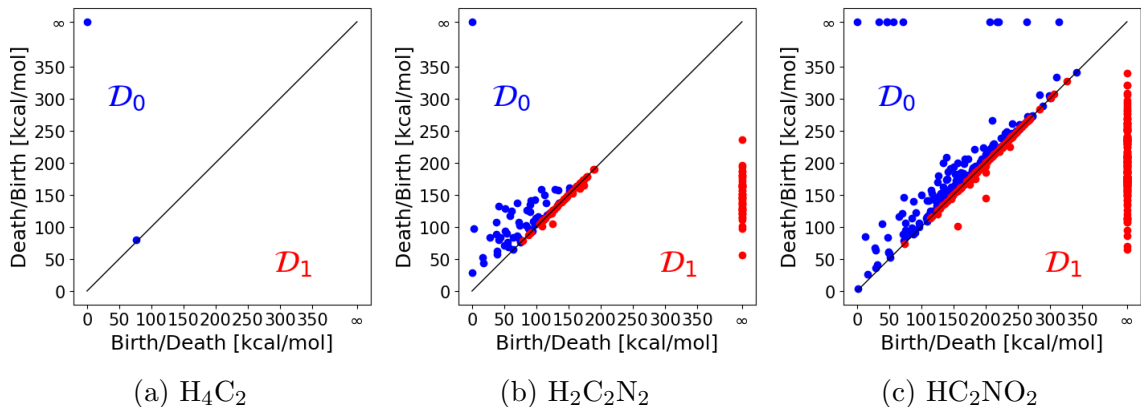
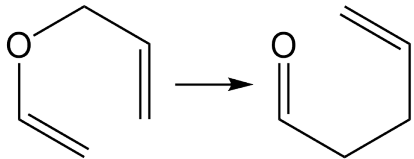


Figure 20: Persistence diagrams for RRM of (a) H_4C_2 , (b) $\text{H}_2\text{C}_2\text{N}_2$, and (c) HC_2NO_2

Finally, the proposed method is applied to partial RRM for the Claisen rearrangement ([3,3]-sigmatropic rearrangement) of allyl vinyl ether, $\text{C}_5\text{H}_8\text{O}$:

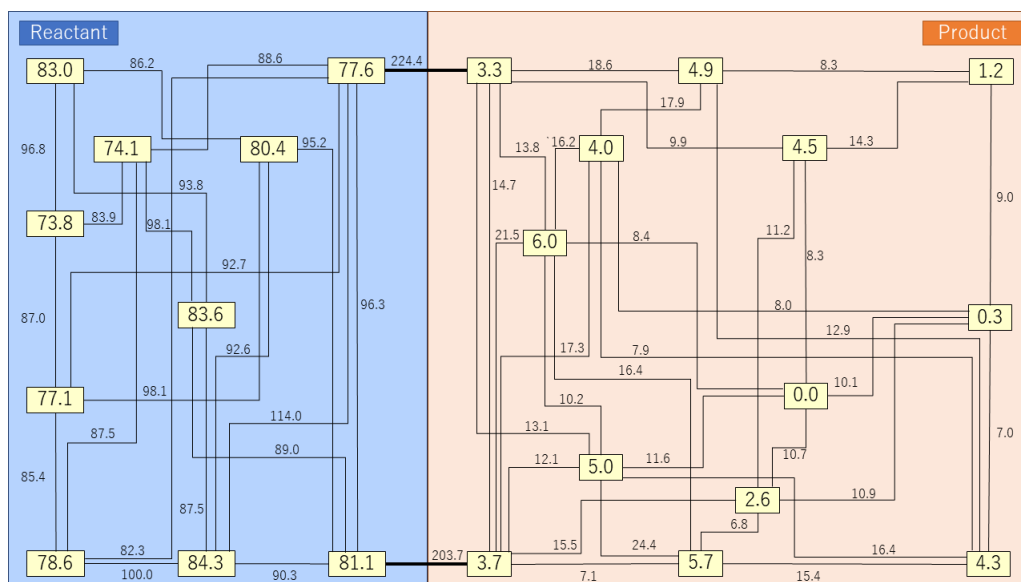


which was reported in Ref. 63 and shown in Fig. 21(a). In this partial RRM, only [3,3]-sigmatropic rearrangements and conformational changes are included. Figure 21(b) depicts the persistence diagram for this partial RRM. In Fig. 21(a), the RRM can be roughly divided into two regions (superbasins), namely, the left half, with high energy, corresponding to the

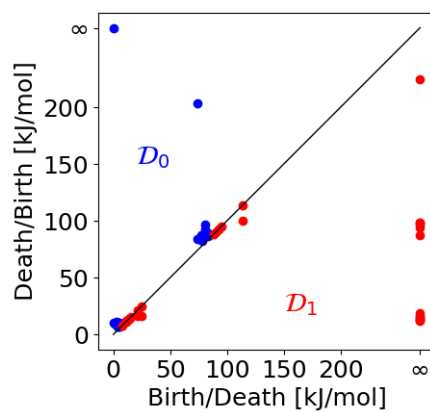
reactant allyl vinyl ether structure and the right half, with low energy, corresponding to the product γ, δ -unsaturated carbonyl structure. Each region contains several EQ structures connected by TSs with low activation energies. These two regions are connected by two elementary reaction paths with high activation energies. In Fig. 21(b), bundles of reactant and product regions appear around a diagonal line. Namely, the product region in the RRM is represented by $0 \leq b \leq 25$ and $0 \leq d \leq 25$ in $\mathcal{D}_0$ and $\mathcal{D}_1$, as well as by $10 \leq b \leq 20$ and $d = \infty$ in $\mathcal{D}_1$. However, the reactant region is represented by $70 \leq b \leq 115$ and $70 \leq b \leq 115$ in $\mathcal{D}_0$ and $\mathcal{D}_1$, as well as by $85 \leq b \leq 100$ and $d = \infty$ in $\mathcal{D}_1$. The plot at $(b, d) = (73.8, 203.7)$ in $\mathcal{D}_0$ corresponds to the lowest-energy TS of the Claisen rearrangement. While this TS connects the vertices with weights of 3.7 and 81.1, the reactant region constitutes the superbasin with the lowest weight of 73.8, leading to $b = 73.8$. The plot at $(b, d) = (224.4, \infty)$ in $\mathcal{D}_1$ indicates the existence of a second pathway connecting the reactant and product regions. In summary, superbasins were displayed as plots around the diagonal line, as well as plots with similar b but $d = \infty$ in $\mathcal{D}_1$, while the reactions connecting two superbasins were displayed as plots far away from the diagonal line. As demonstrated here, the persistence diagram successfully captures the characteristics of the reaction pathways of the Claisen rearrangement.

4 Concluding Remarks

The GRRM program, which enables the automated search for reaction paths of chemical systems, extends the efficiency and applicability of quantum chemical calculations. The program outputs RRMs, which consist of minima (EQs), first-order saddle points (TSs), and their connections (IRCs), to the PES of a chemical system. Most previous GRRM-based studies on the mechanisms of certain reactions have considered only a small portion of TSs with small activation energies and EQs connected by them. As the RRM of an N atomic system contains all EQs and TSs, which are significantly important geometries among the $(3N - 6)$ -dimensional PESs from the perspective of chemistry, discarding information falling



(a)



(b)

Figure 21: (a) Partial RRM of allyl vinyl ether C_5H_8O and (b) its persistence diagram

outside a small energy barrier is too wasteful. Here, a computational method is proposed to extract the characteristic features of an RRM based on the PH applied to a weighted graph representing an RRM. The proposed method provides a unique feature value called a persistence diagram or barcode for an RRM, which is independent of the depiction of the RRM. In this method, the 0-th PH gives information about the connection between the EQ and TS, which corresponds to the analysis by the DG, while the 1-st PH gives information about the circumference of the reaction path. Numerical demonstrations establish that these feature values reflect chemical information about the PES of the system, e.g., chemical hardness, element-pair affinity, and character of the reaction pathways.

If one is interested only in a particular reaction, a detailed analysis of the RRM of the system is sufficient. However, the feature values extracted from the current scheme are expected to be useful for comparing or categorizing multiple RRM. For example, metal nanoclusters exhibit non-scalable catalytic activity depending on their size and composition. Computational prediction of the catalytic activity of metal nanoclusters based on straightforward *ab initio* calculations is often difficult because of their dependence on various external factors, such as temperature, support, and co-catalysts. However, clusters with RRM like those already known to be good catalysts can be expected to exhibit good catalytic activity. Therefore, the utilization of the feature values of the persistence diagram as a descriptor of metal nanoclusters and further categorization of metal nanoclusters based on the similarity of RRM is expected to be beneficial. To achieve this, a method is required to measure the similarity between two persistence diagrams quantitatively. Examples of such methods include the Wasserstein distance⁷¹ and bottleneck distance.⁷² However, the suitability of these metrics for this purpose remains unclear. In addition, as discussed in Sec. 2.6, a general method for constructing a full network corresponding to $(3N - 6)$ -dimensional PES is expected to facilitate a deeper understanding of the system. Another limitation of the current study is the absence of dissociation channels in RRM. Dissociation reactions represented as $A \rightarrow B + C$ are partially included in RRM because molecules B and C often form so-

called product complex. However, molecular systems B and C also constitute independent RRMs at their dissociation limits. Dissociation channels can be categorized into upward and downward channels, which can be searched optionally using the GRRM program. The consideration of dissociation channels in RRMs and the hierarchical structure of RRMs in the current scheme is expected to broaden the applicability of the method and deepen the understanding extracted from it.

However, the adjusted weight rank clique filtration proposed here is applicable not only to RRMs but also to any type of weighted graph. Because of the rapid increase in computer performance as well as the increased availability of network information, obtaining a vast network has become much easier than in the past. The application of the proposed filtration to graphs other than those representing RRMs is expected to yield interesting results.

Acknowledgement

This work was supported in part by the Institute for Quantum Chemical Exploration (IQCE), JSPS KAKENHI for Transformative Research Areas “Hyper-ordered Structures Science” (Grant Numbers: JP21H05544 and JP23H04093 to M.K.), for Scientific Research (Grant Number: JP23H01915) and the Photo-excitonix Project of Hokkaido University. We thank Prof. Ohno (IQCE) for providing RRMs of organic molecules and Prof. Maeda (Hokkaido Univ.) and Dr. Nagahata (Hokkaido Univ.) for providing the RRM of the Claisen rearrangement. Some of the reported calculations were performed using computer facilities at the Research Center for Computational Science, Okazaki (Projects: 21-IMS-C018 and 22-IMS-C019), and at the Research Institute for Information Technology, Kyushu University, Japan. The Institute for Chemical Reaction Design and Discovery (ICReDD) was established by the World Premier International Research Initiative (WPI) of MEXT, Japan. We would like to thank Editage (www.editage.com) for English language editing.

Supporting Information Available

The Supporting Information is available free of charge via the Internet at <http://pubs.acs.org>.

- One-to-one correspondences between the persistence barcode and RRM in Au₅
- Number of bars and statistics of bar lengths in the persistence barcodes

References

- (1) Wales, D. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press, 2003.
- (2) Mezey, P. G. *Potential energy hypersurfaces*; Elsevier: Amsterdam, 1987.
- (3) Fukui, K. Formulation of the reaction coordinate. *The Journal of Physical Chemistry* **1970**, *74*, 4161–4163.
- (4) Satoh, H.; Oda, T.; Nakakoji, K.; Uno, T.; Iwata, S.; Ohno, K. RMapDB: chemical reaction route map data for quantum mechanical-based data chemistry. *Materials Cloud Archive* **2020**, *2020.138*.
- (5) GRRM-GDSP (Institute for Quantum Chemical Exploration). https://iqce.jp/SRPS/GRRM-GDSP_DEMO_e_new.HTM.
- (6) Satoh, H.; Oda, T.; Nakakoji, K.; Uno, T.; Iwata, S.; Ohno, K. "Maizo"-chemistry Project: toward Molecular- and Reaction Discovery from Quantum Mechanical Global Reaction Route Mappings. *Journal of Computer Chemistry, Japan* **2015**, *14*, 77–79.
- (7) Satoh, H.; Oda, T.; Nakakoji, K.; Uno, T.; Tanaka, H.; Iwata, S.; Ohno, K. Potential Energy Surface-Based Automatic Deduction of Conformational Transition Networks and Its Application on Quantum Mechanical Landscapes of D-Glucose Conformers. *Journal of Chemical Theory and Computation* **2016**, *12*, 5293–5308.

- (8) Cerjan, C. J.; Miller, W. H. On finding transition states. *The Journal of Chemical Physics* **1981**, *75*, 2800–2806.
- (9) Jørgensen, P.; Jensen, H. J. A.; Helgaker, T. A gradient extremal walking algorithm. *Theoretica Chimica Acta* **1988**, *73*, 55–65.
- (10) Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. Searching for saddle points of potential energy surfaces by following a reduced gradient. *Journal of Computational Chemistry* **1998**, *19*, 1087–1100.
- (11) Zimmerman, P. M. Single-ended transition state finding with the growing string method. *Journal of Computational Chemistry* **2015**, *36*, 601–611.
- (12) Doye, J.; Wales, D. Surveying a potential energy surface by eigenvector-following Applications to global optimisation and the structural transformations of clusters. *Zeitschrift für Physik D Atoms, Molecules and Clusters* **1997**, *40*, 194–197.
- (13) Maeda, S.; Ohno, K.; Morokuma, K. Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Physical Chemistry Chemical Physics* **2013**, *15*, 3683–3701.
- (14) Ohno, K.; Maeda, S. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chemical Physics Letters* **2004**, *384*, 277–282.
- (15) Maeda, S.; Taketsugu, T.; Morokuma, K.; Ohno, K. Anharmonic downward distortion following for automated exploration of quantum chemical potential energy surfaces. *Bulletin of the Chemical Society of Japan* **2014**, *87*, 1315–1334.
- (16) Maeda, S.; Morokuma, K. A systematic method for locating transition structures of $A+B\rightarrow X$ type reactions. *The Journal of Chemical Physics* **2010**, *132*, 241102.
- (17) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring transition state structures for

- intramolecular pathways by the artificial force induced reaction method. *Journal of Computational Chemistry* **2014**, *35*, 166–173.
- (18) Maeda, S.; Harabuchi, Y.; Takagi, M.; Taketsugu, T.; Morokuma, K. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *Chemical Record* **2016**, *16*, 2232–2248.
- (19) Iwasa, T.; Sato, T.; Takagi, M.; Gao, M.; Lyalin, A.; Kobayashi, M.; Shimizu, K.-i.; Maeda, S.; Taketsugu, T. Combined Automated Reaction Pathway Searches and Sparse Modeling Analysis for Catalytic Properties of Lowest Energy Twins of Cu₁₃. *The Journal of Physical Chemistry A* **2019**, *123*, 210–217.
- (20) 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*, 804–814.
- (21) Sugiyama, K.; Sumiya, Y.; Takagi, M.; Saita, K.; Maeda, S. Understanding CO oxidation on the Pt(111) surface based on a reaction route network. *Physical Chemistry Chemical Physics* **2019**, *21*, 14366–14375.
- (22) Unsleber, J. P.; Grimmel, S. A.; Reiher, M. Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks. *Journal of Chemical Theory and Computation* **2022**, *18*, 5393–5409.
- (23) <https://github.com/qcscine/chemoton>.
- (24) Van de Vijver, R.; Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Computer Physics Communications* **2020**, *248*, 106947.
- (25) Zádor, J.; Martí, C.; Van de Vijver, R.; Johansen, S. L.; Yang, Y.; Michelsen, H. A.; Najm, H. N. Automated Reaction Kinetics of Gas-Phase Organic Species over Multiwell Potential Energy Surfaces. *The Journal of Physical Chemistry A* **2023**, *127*, 565–588.

- (26) <https://github.com/zadorlab/KinBot>.
- (27) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: an automated reaction optimizer for new catalysts. *Journal of Chemical Theory and Computation* **2018**, *14*, 5249–5261.
- (28) Ingman, V. M.; Schaefer, A. J.; Andreola, L. R.; Wheeler, S. E. QChASM: Quantum chemistry automation and structure manipulation. *WIREs Computational Molecular Science* **2021**, *11*, e1510.
- (29) <https://github.com/QChASM/Aaron>.
- (30) Döntgen, M.; Przybylski-Freund, M.-D.; Kröger, L. C.; Kopp, W. A.; Ismail, A. E.; Leonhard, K. Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations. *Journal of Chemical Theory and Computation* **2015**, *11*, 2517–2524.
- (31) <https://www.scm.com/doc/AMS/Utilities/ChemTraYzer.html>.
- (32) Türtcher, P. L.; Reiher, M. Pathfinder– Navigating and Analyzing Chemical Reaction Networks with an Efficient Graph-Based Approach. *Journal of Chemical Information and Modeling* **2022**, *63*, 147–160.
- (33) Ismail, I.; Chantreau Majerus, R.; Habershon, S. Graph-Driven Reaction Discovery: Progress, Challenges, and Future Opportunities. *The Journal of Physical Chemistry A* **2022**, *126*, 7051–7069.
- (34) Unsleber, J. P.; Reiher, M. The exploration of chemical reaction networks. *Annual Review of Physical Chemistry* **2020**, *71*, 121–142.
- (35) Battista, G. D.; Eades, P.; Tamassia, R.; Tollis, I. G. Algorithms for drawing graphs: an annotated bibliography. *Computational Geometry* **1994**, *4*, 235–282.

- (36) Herman, I.; Melancon, G.; Marshall, M. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* **2000**, *6*, 24–43.
- (37) Tsutsumi, T.; Ono, Y.; Arai, Z.; Taketsugu, T. Visualization of the Intrinsic Reaction Coordinate and Global Reaction Route Map by Classical Multidimensional Scaling. *Journal of Chemical Theory and Computation* **2018**, *14*, 4263–4270.
- (38) Tsutsumi, T.; Ono, Y.; Taketsugu, T. Visualization of reaction route map and dynamical trajectory in reduced dimension. *Chemical Communications* **2021**, *57*, 11734–11750.
- (39) Tsutsumi, T.; Ono, Y.; Taketsugu, T. Reaction Space Projector (ReSPer) for Visualizing Dynamic Reaction Routes Based on Reduced-Dimension Space. *Topics in Current Chemistry* **2022**, *380*, 19.
- (40) 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. *The Journal of Physical Chemistry A* **2015**, *119*, 11641–11649.
- (41) Sumiya, Y.; Maeda, S. A Reaction Path Network for Wöhler’s Urea Synthesis. *Chemistry Letters* **2019**, *48*, 47–50.
- (42) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics* **1997**, *106*, 1495–1517.
- (43) Wales, D. J. The energy landscape as a unifying theme in molecular science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2005**, *363*, 357–377.

- (44) Petri, G.; Scolamiero, M.; Donato, I.; Vaccarino, F. Topological Strata of Weighted Complex Networks. *PLoS ONE* **2013**, *8*, e66506.
- (45) Mirth, J.; Zhai, Y.; Bush, J.; Alvarado, E. G.; Jordan, H.; Heim, M.; Krishnamoorthy, B.; Pflaum, M.; Clark, A.; Z, Y.; Adams, H. Representations of energy landscapes by sublevelset persistent homology: An example with n-alkanes. *The Journal of Chemical Physics* **2021**, *154*, 114114.
- (46) Okushima, T.; Niiyama, T.; Ikeda, K. S.; Shimizu, Y. Connectivity graph: Multiple connectivity on potential energy surface does affect the dynamics. *Physical Review E* **2007**, *76*, 036109.
- (47) 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. *Journal of Computational Chemistry* **2018**, *39*, 233–251.
- (48) Maeda, S.; Harabuchi, Y. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *WIREs Computational Molecular Science* **2021**, *11*, e1538.
- (49) Ebisawa, S.; Tsutsumi, T.; Taketsugu, T. Geometric analysis of anharmonic downward distortion following paths. *Journal of Computational Chemistry* **2021**, *42*, 27–39.
- (50) Ohno, K.; Satoh, H. *Exploration on Quantum Chemical Potential Energy Surfaces: Towards the Discovery of New Chemistry*; Royal Society of Chemistry: London, 2022.
- (51) Sobez, J.-G.; Reiher, M. Molassembler: Molecular graph construction, modification, and conformer generation for inorganic and organic molecules. *Journal of Chemical Information and Modeling* **2020**, *60*, 3884–3900.

- (52) Edelsbrunner, H.; Harer, J. *Computational Topology*; American Mathematical Society: Providence, Rhode Island, 2009.
- (53) Otter, N.; Porter, M. A.; Tillmann, U.; Grindrod, P.; Harrington, H. A. A roadmap for the computation of persistent homology. *EPJ Data Science* **2017**, *6*, 17.
- (54) Diestel, R. *Graph Theory*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2017; Vol. 173.
- (55) Kozlov, D. *Combinatorial Algebraic Topology*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; Vol. 21.
- (56) Wallace, D. A. R. *Groups, Rings and Fields*; Springer: London, 1998.
- (57) Saito, S.; Matsumoto, M.; Ohmine, I. In *Advances in Classical Trajectory Methods*; Hase, W. L., Ed.; JAI Press, 1999; Vol. 4; pp 105–151.
- (58) <https://github.com/mochimochi-brian/GRRM-PD>.
- (59) Teramoto, H.; Saito, T.; Aoki, M.; Murayama, B.; Kobayashi, M.; Nakamura, T.; Taketsugu, T. 2023, DOI: 10.48550/arXiv.2305.08072, arXiv:2305.08072.
- (60) 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 Jr., J. A.; 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. 2016.
- (61) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.
- (62) Hay, P. J.; Wadt, W. R. *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *The Journal of Chemical Physics* **1985**, *82*, 299–310.
- (63) Nagahata, Y.; Maeda, S.; Teramoto, H.; Horiyama, T.; Taketsugu, T.; Komatsuzaki, T. Deciphering Time Scale Hierarchy in Reaction Networks. *The Journal of Physical Chemistry B* **2016**, *120*, 1961–1971.
- (64) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.
- (65) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980**, *72*, 650–654.
- (66) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry* **1983**, *4*, 294–301.
- (67) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

- (68) Papajak, E.; Truhlar, D. G. Convergent partially augmented basis sets for post-hartree-fock calculations of molecular properties and reaction barrier heights. *Journal of Chemical Theory and Computation* **2011**, *7*, 10–18.
- (69) 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. *The Journal of Chemical Physics* **2015**, *143*, 014301.
- (70) Smeeton, L. C.; Oakley, M. T.; Johnston, R. L. Visualizing energy landscapes with metric disconnectivity graphs. *Journal of Computational Chemistry* **2014**, *35*, 1481–1490.
- (71) Turner, K.; Mileyko, Y.; Mukherjee, S.; Harer, J. Fréchet Means for Distributions of Persistence Diagrams. *Discrete & Computational Geometry* **2014**, *52*, 44–70.
- (72) Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J. Stability of Persistence Diagrams. *Discrete & Computational Geometry* **2007**, *37*, 103–120.