ARTICLE

Visualization of reaction route map and dynamical trajectory in reduced dimension

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Takuro Tsutsumi,[a] Yuriko Ono,[b] and Tetsuya Taketsugu[a,b] *

In the quantum chemical approach, chemical reaction mechanisms are investigated based on a potential energy surface (PES). Automated reaction path search methods enable us to construct a global reaction route map containing multiple reaction paths corresponding to a series of elementary reaction processes. On-the-fly molecular dynamics (MD) method provides a classical trajectory exploring the full-dimensional PES based on electronic structure calculations. We have developed two reaction analysis methods, the on-the-fly trajectory mapping method and the reaction space projector (ReSPer) method, by introducing a structural similarity to a pair of geometric structures and revealed dynamic aspects affecting chemical reaction mechanisms. In this review, we will present the details of these analysis methods and discuss the dynamics effects of reaction path curvature and reaction path bifurcation with applications to the $CH_3OH + OH^-$ collision reaction and the Au_5 cluster branching and isomerization reactions.

1. INTRODUCTION

Under the Born–Oppenheimer approximation, chemical reactions are described as a series of elementary reaction processes on a potential energy surface (PES). In quantum chemical calculations, the intrinsic reaction coordinate (IRC)[1] is used as a reaction path to describe elementary reaction processes, which connects reactants and products through transition state (TS) structures with the steepest descent path in mass-weighted Cartesian coordinates. In the last 15 years, automated reaction-path search methods called anharmonic downward distortion following (ADDF)[2,3] and artificial force induced reaction (AFIR)[4–6] have been developed, and it has become possible to construct a global reaction route map consisting of a large number of equilibrium structures, TS structures, and IRCs connecting them. In addition, attempts have been made to calculate temporal changes in the concentrations of reaction intermediates and products under given experimental conditions by applying transition state theory to each elementary reaction process that constitutes the global reaction route map, deriving coupled reaction rate equations, and performing kinetic simulations.[7–9]

The IRC is a static reaction path defined by the shape of the PES, and in the actual chemical reaction process, the molecular system proceeds with vibrations in the vicinity of the IRC caused by atomic momentum. Miller, Handy, and Adams formulated a reaction path Hamiltonian in terms of a reaction coordinate and $3N - 7$ normal coordinates for the N atomic system.[10] When the IRC is highly curved in the coordinate space, the molecule system feels the centrifugal force in the negative direction of the curvature vector of the reaction path, and the vibration is greatly excited, which may cause the system to deviate from the IRC.[11,12] In addition, the shape of the PES in the direction of the vibrational degrees of freedom orthogonal to the IRC changes from a valley to a ridge (valley-ridge transition point) in some cases, and the reaction path may bifurcate into multiple valleys.[13,14] Such a large curvature of the IRC and the appearance of the valley-ridge transition point suggest a breakdown of the IRC picture due to dynamic effects.

On-the-fly molecular dynamics (MD) simulations based on the forces acting on atoms obtained from quantum chemical calculations can be used to elucidate the chemical reaction mechanism including dynamic effects with all degrees of freedom. Although the on-the-fly MD method is computationally expensive, it has been used to interpret experimental results and to elucidate chemical reaction mechanisms and their driving forces. This method has recently been applied to reaction processes in excited states.[15–18] The classical trajectory obtained by on-the-fly MD is a time series of molecular coordinates and velocities, and it shows the dynamic path of the molecular system moving on the PES for a given initial condition. Traditionally, on-the-fly trajectories have been analyzed by tracking temporal changes in molecular geometries such as internal coordinates selected by chemical intuition,[15–18] but unexpected reactions may be overlooked. In the dynamic process of an actual reaction, multiple elementary reaction processes may coincide, and thus, it is desirable to extend the reaction dynamics analysis method to chemical reactions involving multiple IRCs.

We have recently proposed an *on-the-fly trajectory mapping* method based on a global reaction route map. In this

[a] *Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan*
[b] *Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo 001-0021, Japan*

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

method, dynamic paths can be studied by projecting *on-the-fly* trajectories onto a global reaction route map consisting of multiple IRCs.[19] However, because the coordinate space describing the geometric structure of molecular systems is multidimensional, it is generally difficult to visualize the reaction route map and dynamic paths in an appropriate reduced-dimensionality coordinate space. Therefore, we have developed *Reaction Space Projector* (ReSPer) as a general program that extracts chemical reaction spaces from global reaction route maps and projects on-the-fly trajectories onto the reaction spaces using a dimensionality reduction method that mathematically embeds multidimensional data into a low-dimensional space.[20,21] In this review, we summarize our on-the-fly trajectory mapping method and ReSPer strategy and present the results of their applications to the OH^- + CH_3F → CH_3OH + F^- collision reaction and the isomerization reaction of Au_5 clusters.[19–21]

2. Methodology

2.1 Structure Similarity

First, we consider how to evaluate the similarity between two structures of a given molecular system. Although several useful molecular similarities have been proposed in biochemistry and chemoinformatics, such as the Tanimoto coefficient[22] and the generalized root mean square deviation (G-RMSD),[23] we adopt the simple definition of Euclidean distance (linear distance) in mass-weighted Cartesian coordinates. Let $\mathbf{x}^{(i)} = \left(x_1^{(i)}, y_1^{(i)}, z_1^{(i)}, \ldots, z_N^{(i)}\right)$ be the 3*N*-dimensional Cartesian coordinates of structure *i* of an *N*-atom molecule, and $\boldsymbol{\xi}^{(i)} = \left(\xi_1^{(i)}, \ldots, \xi_{3N}^{(i)}\right) = \left(\sqrt{m_1}x_1^{(i)}, \ldots, \sqrt{m_N}z_N^{(i)}\right)$ be the mass-weighted Cartesian coordinates, where m_α is the mass of the αth atom. The mass-weighted linear distance d_{ij} between structures *i* and *j* is defined as

$$d_{ij} = \sqrt{\sum_k^{3N}\left(\xi_k^{(i)} - \xi_k^{(j)}\right)^2} = \left|\boldsymbol{\xi}^{(i)} - \boldsymbol{\xi}^{(j)}\right|. \qquad (1)$$

In order to obtain the optimal distance between the structures, it is necessary to remove the displacements caused by the translational and rotational degrees of freedom of the two structures. The former can be removed exactly by setting each center of mass to the origin of the coordinate system, and the latter can be removed by determining the optimal orientation of the x-y-z axes to minimize d_{ij} using the Kabsch algorithm.[24]

Another issue is a nuclear permutation-inversion (NPI) isomer, which is generated from transpositions of identical nuclei and spatial inversion operation. According to the definition in eq. (1), the linear distance between the original structure and the NPI isomer is not zero because the atomic coordinates are different due to the different atomic numbering. Therefore, NPI isomers need to be handled appropriately depending on the chemical reaction process to be analyzed. To determine the inter-structure distance d_{ij}, we have implemented a *merged-NPI* option that selects the minimum mass-weighted linear distance between structure *i* and each NPI isomer of structure *j*. This allows merging the NPI isomers into one representative structure. The recently proposed G-RMSD is based on a bipartite matching algorithm, which can provide structure similarity between two structures with different molecular sizes without any inherent atomic mapping.[23]

2.2 On-the-fly Trajectory Mapping Method

The key idea of the on-the-fly trajectory mapping method is to assign the most similar reference structure to the structure on the trajectory using a distance function. For the structure on the trajectory at time *t*, $\boldsymbol{\xi}^{(\mathrm{trj})}(t)$, the distance function defined as the mass-weighted linear distance in eq (1), $d(t) = \left(d_1(t), \ldots, d_n(t)\right)$, are calculated for *n* reference structures, and the closest reference structure that gives the smallest distance is determined. **Figure 1a** shows a schematic picture of the on-the-fly trajectory mapping method where a trajectory runs in the reaction route map. The distance functions from the reference structures, MIN1, MIN2, TS1, and TS2, are shown as a function of time along the trajectory in Figure 1b. Focusing on the closest reference structures along the simulation time, one can realize the dynamic reaction route as TS1 → TS2 → MIN2.

Since the proposed method characterizes the dynamic reaction routes based on the subspace of the PES determined by the reference structures, the choice of the reference structures is important to adequately describe the chemical reaction of interest. For example, if one selects only equilibrium structures from the global reaction route map as reference structures, one can study the behavior of hopping from one minimum to another. If molecular structures along multiple IRC paths are selected, one can observe dynamic reaction routes associated with multiple chemical elementary processes during structural transformation.

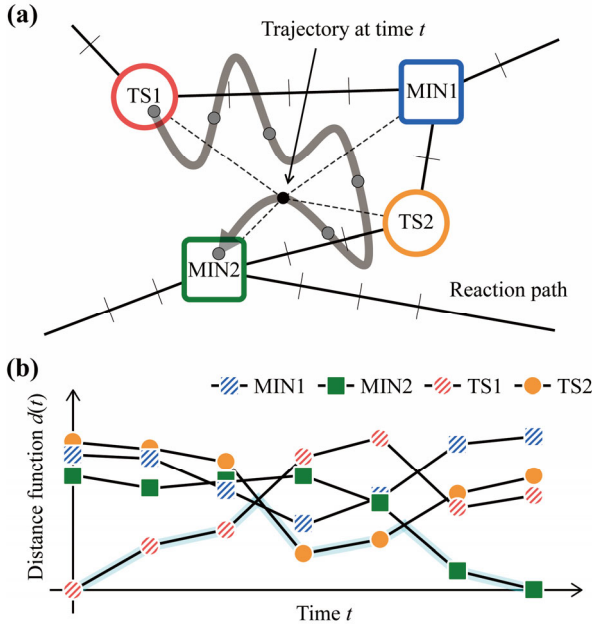

Figure 1. (a) Schematic picture of the on-the-fly trajectory mapping method where a trajectory runs in the reaction route map. (b) Changes in the distance functions from the four reference structures (MIN1, MIN2, TS1, and TS2) as a function of time *t*. The time series of the closest reference structure is highlighted with a sky blue line.

Here, we would like to mention trajectory-based approaches to elucidate the mechanism of chemical reactions. Czakó *et al.* proposed an Eckart-transformation-based approach to assign trajectory structures to stationary points on the PES and clarified the role of stationary points in various reaction paths in collision reactions.[25,26] Such a trajectory mapping method has been extended to elucidate chemical reaction mechanisms, and Martínez-Núñez *et al.* have developed the *Transition State Search using Chemical Dynamics Simulations* (TSSCDS) method that performs bond cleavage/formation searches to molecular structures along trajectories generated by high-energy dynamics simulations.[27] Furthermore, by combining TSSCDS with statistical rate theory, the chemical reaction mechanism under experimental conditions was also elucidated.[28,29]

2.3 Reaction Space Projection by Classical Multidimensional Scaling

Dimensionality reduction techniques to characterize and visualize multidimensional data in lower-dimensional space, such as principal component analysis (PCA),[30–32] (classical) multidimensional scaling (MDS),[20,21,33,34] locally linear embedding (LLE),[35] isometric feature mapping (ISOMAP),[34,35] and diffusion map[36] have been recently applied to the field of theoretical chemistry. The classical MDS (CMDS) approach embeds multidimensional data into a reduced-dimensional space using pairwise Euclidean distances to reflect the multidimensional data structure as much as possible.[37,38] In molecular terms, in low-dimensional space, similar molecular structures are close together, and dissimilar ones are far apart. These features correspond to the nature of PES, in which chemical reactions occur with continuous structural changes. Therefore, we developed Reaction Space Projector (ReSPer) to construct a reduced-dimensionality chemical reaction space using the CMDS method.[20]

The concept of CMDS is to convert the Euclidean distance matrix between a pair of *n* reference structures into a configuration matrix that determines the position of the structure in chemical reaction space. Here, we prepare a distance matrix **D** that represents the structure similarity between structures *i* and *j* using the linear distance d_{ij} given by eq (1). The distance matrix is, of course, a real symmetric matrix. CMDS visualizes *n* reference structures, such as molecular structures on an IRC path or reaction route map, in a *p*-dimensional subspace and is performed in the following steps:[20]

1) Prepare a distance matrix **D** where *ij*th element is $d_{ij}(i,j=1,\dots,n)$ in eq (1).
2) Convert a squared distance matrix $\left(\mathbf{D}^{(2)}\right)_{ij}=\left(d_{ij}\right)^2$ into an inner product matrix **Q** by the Young-Householder transformation (double centering)[39]:
$$\mathbf{Q}=-\frac{1}{2}\left(\mathbf{E}-\frac{1}{n}\mathbf{1}\right)\mathbf{D}^{(2)}\left(\mathbf{E}-\frac{1}{n}\mathbf{1}\right)^{T}, \quad (2)$$
where **E** and **1** are a unit matrix and a square matrix with all elements as 1, respectively.
3) Diagonalize **Q** and order eigenvalues $\{\lambda_1,\dots,\lambda_n\}$ and corresponding eigenvectors $\{\mathbf{L}_1,\dots,\mathbf{L}_n\}$ in a descending way of $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$.
4) Select *p* largest eigenvalues $\{\lambda_1,\dots,\lambda_p\}$ with the corresponding eigenvectors $\{\mathbf{L}_1,\dots,\mathbf{L}_p\}$ and determine principal coordinates (PCos) $\mathbf{Y}_a=\mathbf{L}_a\sqrt{\lambda_a}\ (1\leq a\leq p)$.

Strictly speaking, the maximum number of positive eigenvalues is 3*N*-6, corresponding to the dimension of PES, but it actually is 3*N*-3 because three degrees of freedom of rotation remain. The contribution of the selected principal coordinates to reproducing the relative position of the given data is estimated by the ratio Λ_a defined as[38]
$$\Lambda_a=\frac{\lambda_a}{\sum_c^b \lambda_c}, \quad (3)$$
where *b* is the number of non-negative eigenvalues of λ_c. The representability in the reduced-dimensionality subspace spanned by $\{\mathbf{Y}_1,\dots,\mathbf{Y}_p\}$ can be evaluated by the cumulated proportion Λ,[38]
$$\Lambda=\frac{\sum_c^p \lambda_c}{\sum_c^b \lambda_c}. \quad (4)$$
If Λ is close to 1, one can conclude that the positions of all reference structures in the full-dimensional space are almost reproduced in the *p*-dimensional principal coordinate space.

After constructing the reaction space by obtaining the principal coordinates from *n* reference structures on the PES, *m* structures along the trajectory, which are out-of-sample data, can be projected into the reaction space by out-of-sample extension.[40,41] Let $\mathbf{D}_{n\times n}$, $\mathbf{D}_{n\times m}$, and $\mathbf{D}_{m\times m}$ be the distance matrices between the reference structures, between the reference and out-of-sample structures, and between the out-of-sample structures, and the subscripts denote the sizes of the respective distance matrices. Subsequent to the CMDS method, out-of-sample extension is performed in the following steps:[21]

5) Make a squared distance matrix $\mathbf{A}^{(2)}$ defined as
$$\mathbf{A}^{(2)}=\begin{pmatrix}\mathbf{D}_{n\times n}^{(2)} & \mathbf{D}_{n\times m}^{(2)}\\ \mathbf{D}_{m\times n}^{(2)} & \mathbf{D}_{m\times m}^{(2)}\end{pmatrix} \quad (5)$$
where $\mathbf{D}_{n\times n}^{(2)}$, $\mathbf{D}_{n\times m}^{(2)}$, and $\mathbf{D}_{m\times m}^{(2)}$ are distance matrices with all elements of $\mathbf{D}_{n\times n}$, $\mathbf{D}_{n\times m}$, and $\mathbf{D}_{m\times m}$ squared and $\mathbf{D}_{m\times n}^{(2)}$ is a transpose matrix of $\mathbf{D}_{n\times m}^{(2)}$.
6) Apply the double centering function to $\mathbf{A}^{(2)}$ according to Trosset formula[41]:
$$\mathbf{B}=\tau_w(\mathbf{A})=-\frac{1}{2}\left(\mathbf{E}-\frac{\mathbf{e}\cdot\mathbf{w}^T}{n}\right)\mathbf{A}^{(2)}\left(\mathbf{E}-\frac{\mathbf{e}\cdot\mathbf{w}^T}{n}\right)^{T} \quad (6)$$
where **E**, **e**, and **w** are the unit matrix, an *n* + *m* dimensional column vector with all elements set to 1, and an *n* + *m* dimensional column vector with 1 (the first *n* components) and 0 (the remaining *m* components). For the part of $\mathbf{D}_{n\times n}^{(2)}$, eq (6) is consistent with the Young-Householder transformation shown in step 2).
7) Give initial *p*-dimensional coordinates $\mathbf{Z}_{m\times p}$ for *m* out-of-sample data, and make initial *p*-dimensional coordinates $\mathbf{X}_{(n+m)\times p}$ for *n* + *m* data:
$$\mathbf{X}_{(n+m)\times p}\equiv\begin{pmatrix}\mathbf{Y}_{n\times p}\\ \mathbf{Z}_{m\times p}\end{pmatrix}, \quad (7)$$
where $\mathbf{Y}_{n\times p}$ is the principal coordinate predefined for the *n* reference structures in step 4).
8) Optimize the following function[41] with respect to $\mathbf{Z}_{m\times p}$ to finally obtain the optimal $\mathbf{Z}_{m\times p}$.

$$\min\|\mathbf{B}-\mathbf{X}\mathbf{X}^T\|^2 = \min\left\|\begin{pmatrix}\mathbf{B}_{n\times n}-\mathbf{Y}\mathbf{Y}^T & \mathbf{B}_{n\times m}-\mathbf{Y}\mathbf{Z}^T\\ \mathbf{B}_{m\times n}-\mathbf{Z}\mathbf{Y}^T & \mathbf{B}_{m\times m}-\mathbf{Z}\mathbf{Z}^T\end{pmatrix}\right\|^2, \quad (8)$$

where $\mathbf{B}_{n\times n}$, $\mathbf{B}_{n\times m}$, $\mathbf{B}_{m\times n}$, and $\mathbf{B}_{m\times m}$ are parts of the **B** matrix given in step 6).

Through the above eight steps, the ReSPer program constructs a *p*-dimensional reaction space $\mathbf{Y}_{n\times p}$ determined by the principal coordinates of the *n* reference structures and projects the *m* structures included in the on-the-fly trajectory into the reaction space as $\mathbf{Z}_{m\times p}$. At the moment, our program uses the modified Powell algorithm of the SciPy software to optimizes eq (8) and projects the out-of-sample structures one by one into a fixed *p*-dimensional reaction space. In the following, the principal coordinates of all molecular structures are denoted by PCo*i* ($0 \le i \le p$).

Recently, other dimensionality reduction techniques have been applied to determine the principal component axes for a set of reference structures and to investigate dynamic reaction mechanisms. Hare *et al.* analyzed chemical transformations using principal component analysis (PCA) and visualized several IRC paths and MD trajectories.[31] They also published the PCA-based *PathReducer* program.[31] The MDS approach can also be used to construct potential energy landscapes of protein folding and conformational transitions.[33,42] For example, Oliveira *et al.* developed the *energy landscape visualization method* (*ELViM*) and very recently succeeded in exploring the energy landscape of an essentially disordered protein that is a conformationally polymorphic ensemble.[43]

3. Results and Discussion

3.1 Collision Reaction of OH^- +CH_3F.

We first apply the on-the-fly trajectory mapping method to the collision reaction OH^- + CH_3F, which is a seven-atom system with 15 internal degrees of freedom, as a demonstration of our proposed analysis method, and then perform the reduced-dimensionality analysis using the ReSPer program.[20,21] For this reaction, Hase and coworkers[44] have previously performed IRC analysis and *ab initio* direct dynamics simulations at the MP2/6-31+G* level and found that even though the product connected by the IRC is the hydrogen-bonded complex $[CH_3OH\cdots F]^-$, approximately 90 % of the trajectories starting from the TS proceed to the direct dissociation path leading to the dissociation product CH_3OH + F^- of the S_N2 reaction. This study emphasized the importance of "*non-IRC*" dynamic pathways in the actual reaction process and finally concluded that inefficient intramolecular vibrational energy redistribution (IVR) around the curved region on the IRC results in preventing the trajectories of the molecular system from following the IRC.

To analyze the trajectory of this collision reaction, the IRC was calculated, and on-the-fly MD simulations were performed at the MP2/6-31+G* level. In addition, to evaluate the curvature of the IRC, the scalar curvature $\kappa(s)$, defined as the norm of the curvature vector obtained by numerically differentiating the tangent vector along the reaction coordinates *s*, was calculated.[12] On-the-fly MD simulations were performed using initial conditions generated in the vicinity of the TS structure,

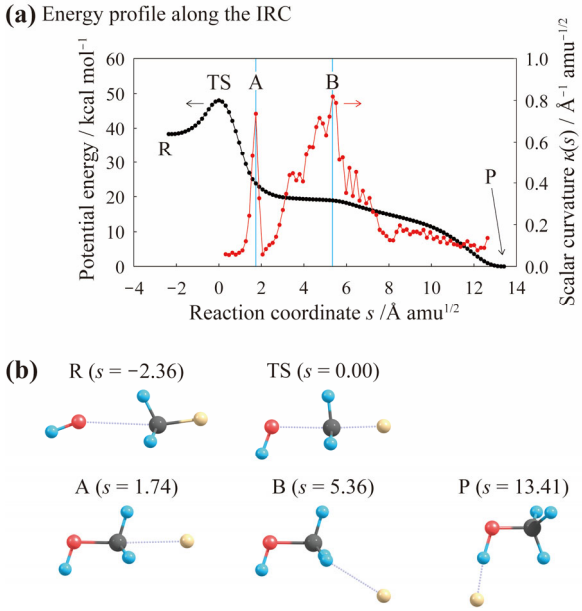

Figure 2. (a) Energy profile (in kcal/mol) along the IRC of OH^- + CH_3F → $[CH_3OH\cdots F]^-$ and the scalar curvature $\kappa(s)$ (red dots: $Å^{-1}amu^{-1/2}$) from TS to the product structure. (b) Five representative structures, reactant (R), TS, two maxima of the scalar curvature (A and B), and product (P), are also shown.

with the initial molecular coordinates and velocities following a Boltzmann distribution at 300 K. In this analysis, 50 trajectories were run in the forward direction and allowed to explore up to 1 ps on the PES with a time step of 0.1 fs until the C-F interatomic distance associated with F^- dissociation was greater than 8 Å or the distance function between the molecular structure along the trajectory and the product structure was less than 5 Å $amu^{1/2}$. Electronic structure and IRC calculations were performed with the GAMESS program,[45] and on-the-fly MD simulations were performed with the SPPR program.[46]

Figure 2 shows (a) the energy change along the IRC consisting of 109 structures, including the reactant (R), the transition state (TS), and the product (P), and (b) representative structures. The structural change along R → TS → A corresponds to the intuitive S_N2 reaction process: OH^- attacks C of CH_3F^- from behind F^- (R → TS), then F^- leaves with CH_3 reversed, and OH is bound to CH_3 (TS → A). However, in the structural change along the IRC, F^- bound to C is loosely bound to H on the CH_3 side (A → B), and finally F^- goes behind CH_3OH and binds to H of OH (B → P). Figure 2a also shows the scalar curvature (red), which is maximum at A (s = 1.74 Å $amu^{1/2}$) and B (s = 5.36 Å $amu^{1/2}$). Since the IRC around B is more extensively curved than around A, the on-the-fly trajectory is considered to deviate from the IRC from around B. In order to discuss the excitation of vibrational modes by the reaction path curvature in more detail, it has been proposed to project the curvature vector onto vibrational modes orthogonal to the IRC,[11] or to transform it into an internal coordinate representation based on the unified reaction valley approach (URVA).[47]

We now turn to a discussion of the dynamic aspects based on the on-the-fly trajectories: a preliminary classification of the 50 trajectories shows that 11 trajectories pass through the

vicinity of the IRC to reach the P region, and 39 trajectories proceed directly to the dissociation path, which reflects the branching ratio trend of the previous study.[44] Among the 39 trajectories leading to $CH_3OH + F^-$, in two trajectories, the dissociated F^- and CH_3OH moieties recombined to form $[CH_3OH \cdots F]^-$ near the A and B regions. To investigate the dynamic reaction process based on a single IRC path, the distance function $d(t)$ involving the five reference structures (R, TS, A, B, and P) was followed according to the procedure of the on-the-fly trajectory mapping method. The merged-NPI option was not used in this analysis.

Figure 3a shows a typical trajectory reaching the P region, where the closest reference structures are TS, A, B, and P in that order, indicating that the trajectory descends near the IRC from TS to A to B to P. After 100 fs, the trajectory temporarily leaves the B region, but eventually comes within 5 Å $amu^{1/2}$ of P, satisfying the termination condition. Figure 3b shows the non-IRC path where the trajectory deviates from the IRC due to centrifugal force caused by the sharply curved region on the reaction path. After the distance function from A is minimized, the distance functions from all reference structures increase sharply, indicating that the molecular system approaches structure A and then proceeds on the non-IRC path leading to the dissociation of F^- until the termination condition is met. In the case of Figure 3c, the initial behavior is the same as Figure 3b, but the system returns to the vicinity of the IRC after 600 fs, and the distance function from A shows a minimum at 740.0 fs. This behavior corresponds to a reaction process in which the electrostatic interaction between the dipole moment of CH_3OH and the charge of F^- causes the fragment to recombine once it has almost dissociated.[21]

Next, we applied ReSPer analysis to visualize the reaction space determined by the 109 structures located along the IRC. **Figure 4** shows the IRC profile in a two-dimensional reaction space spanned by two principal coordinates, PCo1 and PCo2, obtained in step 4) of section 2.3. Interestingly, in the ReSPer analysis, the single IRC can be visualized as a continuous curve in the reduced-dimensionality coordinate space. The proportions occupied by each principal coordinate are Λ_1 = 0.819 and Λ_2 = 0.134, respectively, and the cumulated proportion Λ of these two dimensions is 0.953. This cumulated proportion close to 1 indicates that the two principal coordinates can almost perfectly reproduce configurations of the molecular structures along the IRC in the full-dimensional coordinate space. In fact, the visualized reaction paths are much more curved around B than around A. This property reflects that the scalar curvature around B increases over a larger region than A, as shown in Fig. 2a. It should be noted here that the definition of the principal coordinates depends on the set of reference structures, and in the current case, the reference structures include more B-curve region structures than A-curve region structures. Thus, the ReSPer analysis suggests that the centrifugal force generated by the curvature near A causes vibrational excitation of CH_3 inversion but this motion does not appear in this two-dimensional subspace, while the centrifugal force generated by the curvature near B causes non-IRC behavior, effectively pushing the molecular system away from the IRC and into a direct dissociation path.[20]

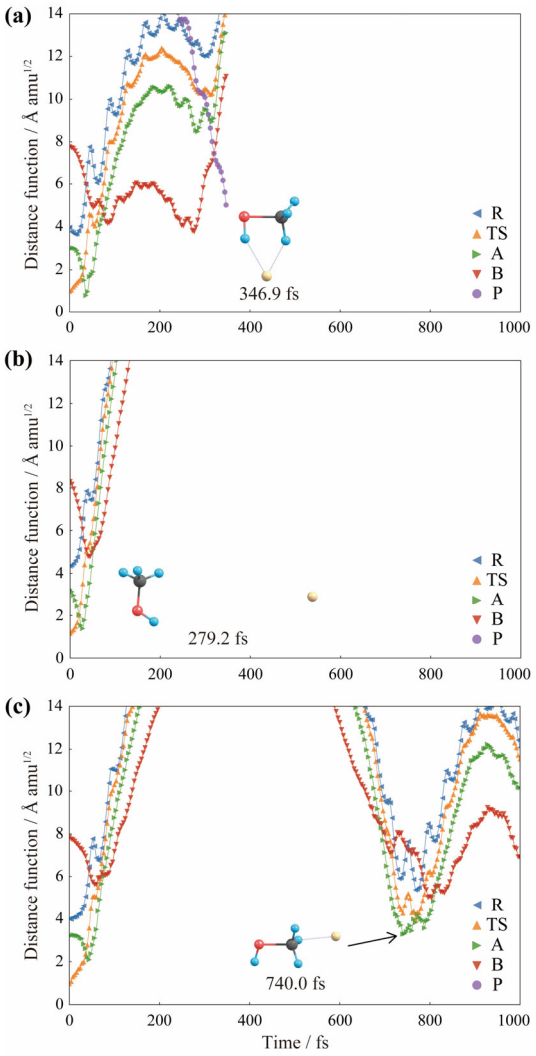

Figure 3. Variations of the distance functions from five reference structures, R, TS, A, B, and P, for (a) typical trajectory reaching the P region of $[CH_3OH \cdots F]^-$, (b) typical trajectory proceeding to the direct dissociation channel of $CH_3OH + F^-$, and (c) typical trajectory leading to the recombination of the dissociated F^- and the CH_3OH moiety. Blue-left triangles, orange-up triangles, green-right triangles, red-bottom triangles, and purple circles denote R, TS, A, B, and P. Terminal structures on (a-b) and the structure closest to A after leaving the IRC are shown. Reprinted with permission from ref.[21]: Tsutsumi, T.; Ono, Y.; Arai, Z.; Taketsugu, T. *J. Chem. Theory Comput.* **2020**, 16, 4029. Copyright 2021 American Chemical Society.

The physical interpretation of the principal coordinates can be obtained from the structural changes along each coordinate. In Fig. 4, PCo1 decreases gradually from R to P, while PCo2 changes mainly between TS and B with a small displacement of PCo1. Focusing on the changes in the internal coordinate along the IRC, the O-C-F bond angle decreases mainly in the order of R, B, and P, which is interpreted to be positively correlated with PCo1. In addition, the interatomic distance between C and F is

found to increase mainly during TS → B, which is interpreted to be positively correlated with PCo2. After passing through B, both PCo1 and PCo2 decrease linearly. This is because the O-C-F bond angle and the C-F interatomic distance decreases simultaneously due to F^- migration.[20]

To examine the dynamic paths, we performed the ReSPer analysis and projected the three on-the-fly trajectories already analyzed in Fig. 3 onto the reduced-dimensionality reaction space determined by the IRC. **Figures 5a, 5c, and 5d** shows the projection of the three trajectories considered in Fig. 3 onto the two-dimensional reaction space generated by the 109 reference structures along the IRC, as shown in Fig. 4. The three trajectories in Figs. 5a, 5c, and 5d correspond to the trajectories in Figs. 3a, 3b, and 3c, respectively, while Fig. 5b shows the potential energy landscape with the trajectory in Fig. 5a projected onto the two-dimensional subspace of PCo1 and PCo2. In Fig. 5b, the initial structure of the projected trajectory is located near TS, reflecting the fact that this simulation was initiated near TS. Plotting all the structures along the trajectory would be too much data to understand the behavior of the dynamic path, so we only plot the points where the potential energy shows a minimum feature along the trajectory.

Fig. 5a shows that the trajectory starts near TS and once leaves the IRC in the region where the reaction path is curved sharply, but it is attracted back to the IRC and enters the P region. The resulting behavior of the trajectories suggests that even if the molecular system reaches the terminal of the IRC, it does not always follow the IRC, but it may leave the IRC subject to the curvature of the reaction path. The characteristics of the PES that cause such dynamic behavior are clearly shown in the potential energy landscape including the IRC shown in Fig. 5b, where the potential energy decreases rapidly from TS to A, but it becomes relatively flat from A to B, which is the region before the direction of the IRC turns significantly. Since energy is conserved in MD calculations, the trajectories descend the PES, converting potential energy into kinetic energy, and the molecular system gains momentum and proceeds to the dissociation channel.

Figure 5c shows the case where the trajectory cannot resist the centrifugal force due to the curvature of the reaction path around B and leaves the IRC, resulting in a non-IRC direct dissociation path to $CH_3OH + F^-$. The ReSPer analysis[21] shows that the trajectories following the non-IRC path follow the PCo2 direction, which is correlated with the C-F interatomic distance, rather than the PCo1 direction, which is correlated with the O-C-F bond angle, confirming the inefficiency of IVR from the relative translation of $CH_3OH + F^-$ to vibrational degrees of freedom, such as O-C···F^- bending vibration, as proposed by Hase and coworkers[44]. Fig. 5d shows the trajectory where F^- and CH_3OH moiety are recombined, corresponding to the case where the trajectory leaves the IRC and then returns to the A region on the IRC. The difference between Fig. 5c and Fig. 5d can be attributed to the energy distribution of the molecular system at the sharply curved region of the IRC. In fact, in the trajectory of Fig 5c, F^- dissociates with a roaming motion around the CH_3OH moiety, which is appeared as a decrease in PCo1. On the other hand, in the trajectory of Fig. 5d, F^- dissociates straight from the CH_3OH moiety with the internal rotation of CH_3 in the CH_3OH part excited, which has little effect on PCo1. These results show that ReSPer analysis can elucidate dynamic reaction paths and their driving forces, such as centrifugal forces due to a highly-curvature of the reaction path, by projecting on-the-fly trajectories into the reduced-dimensionality reaction space.[21]

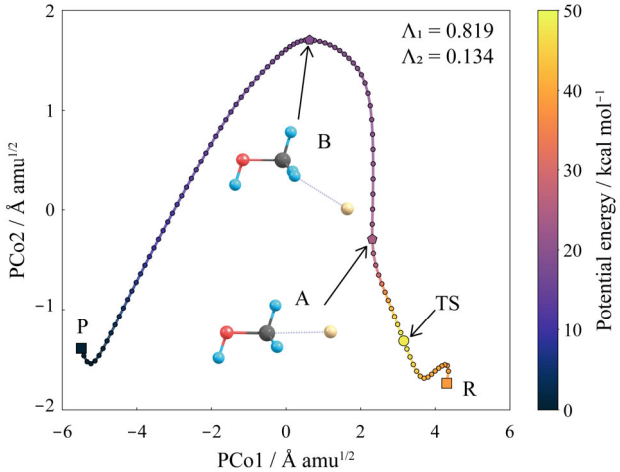

Figure 4. IRC profile of $OH^- + CH_3F \rightarrow [CH_3OH\cdots F]^-$ projected onto the two-dimensional reaction space. Two minima (R and P), two points with curvature maxima (A and B), one TS, and IRC points are denoted as squares, pentangles, a circle, and small circles with lines, respectively. The colormap shows the potential energy relative to P. Reprinted with permission from ref.[21]: Tsutsumi, T.; Ono, Y.; Arai, Z.; Taketsugu, T. *J. Chem. Theory Comput.* **2020**, 16, 4029. Copyright 2021 American Chemical Society.

3.2 Bifurcation and Isomerization Reactions of Au_5 Cluster.

While bulk gold is an inert material, gold nanoparticles exhibit catalytic activity that increases dramatically with decreasing radius.[48] For this reason, gold nanoparticles and gold clusters have attracted much attention as a research target.[49–53] Harabuchi *et al.* calculated a global reaction route map of the Au_5 cluster and investigated the bifurcation of the reaction path by searching for the valley-ridge transition (VRT) point where the PES orthogonal to the IRC changes from a valley to a ridge.[53] In this review, we use the on-the-fly trajectory mapping and the ReSPer method[19–21] to analyze the reaction mechanisms of isomerization and branching reactions in the Au_5 cluster.

We reused the global reaction route map for Au_5 that we calculated earlier.[53] This map was generated by density functional theory (DFT) calculations using the Perdew-Burke-Ernzerhof (PBE) functionals and the LanL2DZ basis set, using a combination of GRRM11[54] and Gaussian09.[55] **Figure 6** shows the two-dimensional global reaction route map where five minima (MIN*i*, *i* = 1~5) are connected *via* 14 TSs (TS*i-j* that connects MIN*i* and MIN*j*). Since the Au_5 cluster has five identical Au atoms, there are 240 (= 5! × 2) nuclear permutation-inversion (NPI) isomers.[19] Among the 14 TSs, seven TSs are

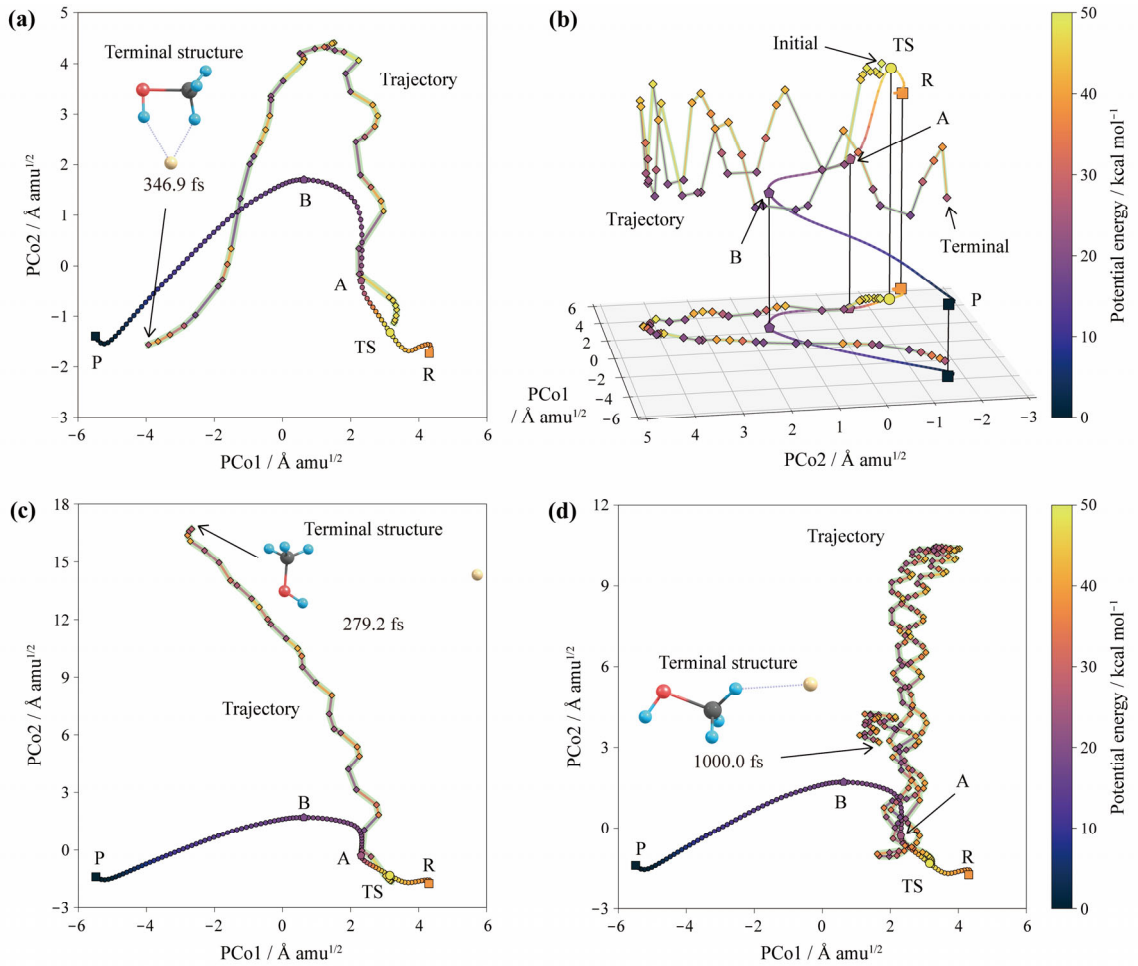

Figure 5. Projected trajectories in the two-dimensional reaction space as determined by IRC: (a) Trajectory reaching the P region ($[CH_3OH\cdots F]^-$), (b) Potential energy landscape with the trajectory shown in (a), (c) Trajectory leading to the direct dissociation path of $CH_3OH + F^-$, and (d) Trajectory leading to the recombination of dissociated F^- and CH_3OH moieties. Two minima (R and P), two curvature maxima (A and B), one TS, points on the IRC, and points on the trajectory are indicated by squares, pentagons, a circle, circles with lines, and diamonds with lines, respectively. The color map shows the relative value of potential energy to P.

classified as TS*i*-*i* connecting two NPI isomers of MIN*i* and are depicted with double solid black lines reaching to MIN*i*. Multiple TS*i*-*i* connecting the same MIN*i* are distinguished by letters of the alphabet (a, b, c, d) in the order of decreasing energy. The number of NPI isomers decreases when the molecule has a spatial symmetry: 120 NPI isomers for C_s and C_2 structures (MIN3, MIN5, TS1-1d, TS1-3a, TS1-3b, TS1-5, TS2-5, TS3-3, TS3-5, TS5-5), 60 NPI isomers for C_{2v} structures (MIN1, MIN4, TS1-1a, TS1-1b, TS1-1c, TS1-2, TS4-4), and 30 NPI isomers for D_{2h} structure (MIN2). The others for C_1 structures (TS1-4 and VRT) have 240 NPI isomers. Most of the structures are planar due to the relativistic effect of the gold atoms,[49] but only MIN4, TS1-1d, TS1-4, TS4-4, TS5-5, and VRT have non-planar geometry. Although in the previous study,[53] five VRTs were found, this study focus on the bifurcation of reaction path related to only one VRT that locates on the IRC starting from TS1-1d.

As described above, TS1-1a, TS1-1b, TS1-1c, and TS1-1d are the transition states that connect the 60 NPI isomers of MIN1 to each other. Each NPI isomer undergoes a conformational change to the other NPI isomers by crossing the activation barrier of each transition state (TS1-1a: 3.8 kcal/mol, TS1-1b: 13.7 kcal/mol, TS1-1c: 16.9 kcal/mol, TS1-1d: 20.8 kcal/mol). Therefore, the NPI isomers produced in the isomerization reaction are determined by the height of the activation barrier. **Figure 7** shows the set of NPI isomers linked by each TS1-1*x* (*x* = a, b, c, d). Here, the set of NPI isomers linked by each TS1-1*x* is called *"a closed island."* Figure 7a shows that the five NPI isomers of MIN1 are linked by structural transformation *via* TS1-1a, and the 60 NPI isomers of MIN1 are divided into 12 closed islands (60/5 = 12). Similarly, four, six, and two NPI isomers are linked through TS1-1b (Fig. 7b), TS1-1c (Fig. 7c), and TS1-1d (Fig. 7d), respectively, and there are 15, 10, and 30 closed islands by TS1-1b. TS1-1c, and TS1-1d, respectively. Note that the six NPI isomers linked by TS1-1c include the two NPI isomers linked by TS1-1d. As a result, the 60 NPI isomers of MIN1 will be fully linked by four sets of TS: (TS1-1a and TS1-1b), (TS1-1a and TS1-1c), (TS1-1a and TS1-1d), and (TS1-1b and TS1-1c). In addition, only 12 NPI isomers are linked by the set of (TS1-1b and TS1-1d).[19]

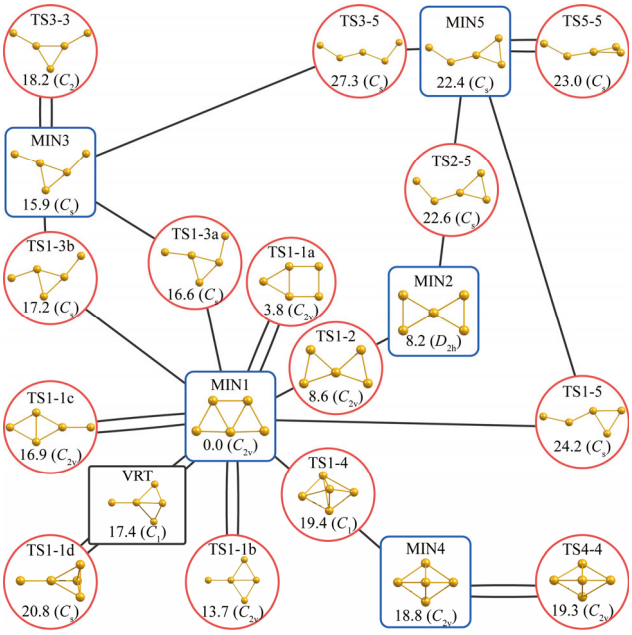

Figure 6. Two-dimensional global reaction route map showing the linkage between the five minima (blue squares) and the 14 transition state structures (red circles) of Au_5 with a valley-ridge transition (VRT) point on the IRC connecting TS1-1d and MIN1 indicated by a black square.[53] The point group and the relative energy (in kcal/mol) to MIN1 are also shown.

Next, on-the-fly MD simulations were performed for the process of descending from TS1-1d, which has the highest potential energy among nine TSs originating from MIN1, in the direction toward MIN1 to investigate the dynamics of Au_5 as it undergoes structural changes in the global reaction route map. Two hundred initial conditions were generated by randomly adding kinetic energy of 5 kcal/mol to nine normal modes of vibration, including an imaginary frequency mode at the TS1-1d structure. The time step was set to 5 fs, which was confirmed beforehand to preserve the energy conservation law, and 200 trajectories were run up to 3 ps. All quantum chemical calculations were performed at the PBE/LanL2DZ level using Gaussian09,[55] and on-the-fly MD simulations were performed with the SPPR program[46]. The ReSPer program was used to analyze the reduced-dimensionality reaction route map and trajectory for Au_5.

3.2.1 Bifurcation Reaction starting from TS1-1d. First, we focus on the bifurcation reaction caused by the presence of VRT on the IRC connecting TS1-1d to MIN1. TS1-1d consists of an Au atom and a pyramidal structure of Au_4. The base of TS1-1d is an isosceles triangle and belongs to C_s symmetry instead of C_{3v}. Therefore, in the configuration space, the three NPI isomers of TS1-1d exist in close proximity to each other, and there is an equilateral triangular pseudo-TS1-1d with C_{3v} symmetry at the center of these three structures. This C_{3v} symmetric pseudo-TS1-1d is the conical intersection between the ground state and the first excited state. Each TS1-1d connects two NPI isomers of MIN1, as shown in Fig. 7d; therefore, the three NPI isomers of TS1-1d connect six different NPI isomers of MIN1. **Figure 8** shows the schematic bifurcation reaction network associated

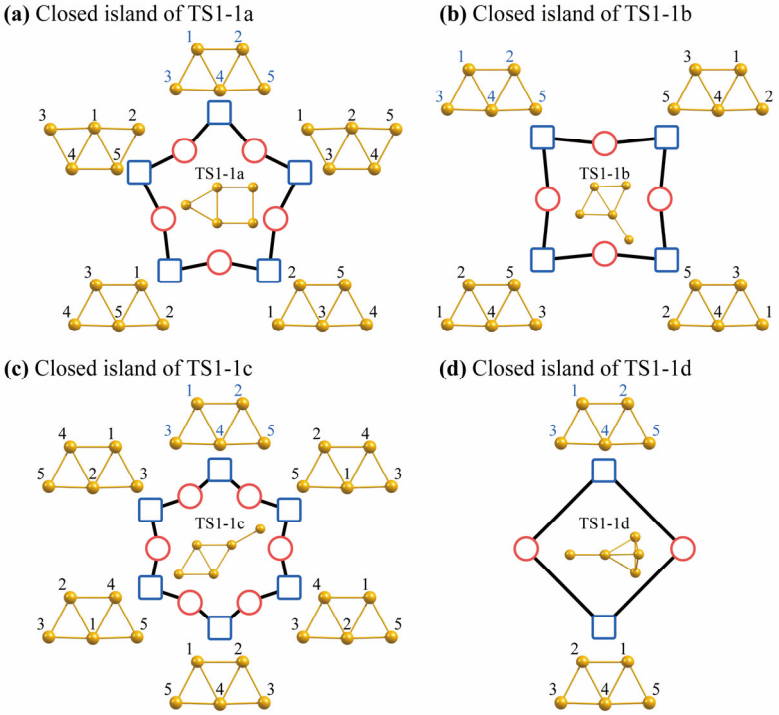

Figure 7. Closed islands in which the NPI isomers of MIN1 of Au_5 are connected by their transition states: (a) TS1-1a (activation barrier 3.8 kcal/mol), (b) TS1-1b (13.7 kcal/mol), (c) TS1-1c (16.9 kcal/mol), and (d) TS1-1d (20.8 kcal/mol). Within each closed island, the NPI isomers of MIN1 (indicated by blue squares) are connected *via* the NPI isomers of TS1-1*x* (indicated by red circles; *x* = a, b, c, d). The molecular structures of NPI isomers of MIN1 and TS1-1*x* are also given.

with the three TS1-1d and the six MIN1.[19,53] By carefully comparing Fig. 7b and Fig. 8, it is clear that its six minima are connected by three different NPI isomers of TS1-1b: MIN1 and MIN1'', MIN1' and MIN1''', and MIN1'''' and MIN1''''' are connected by TS1-1b, TS1-1b', and TS1-1b'', respectively. As reported in the previous study,[53] the three TS1-1b are located near the six VRT in the coordinate space and connect between the IRC-terminated product and branching product structures.

Here, we applied the on-the-fly trajectory mapping method to investigate the reaction process of the branching reaction. In this analysis, we chose 2130 NPI isomers in 20 structures (5 minima, 14 TSs, and VRT) as reference structures and calculated the distance function $d(t)$ for 200 trajectories starting from TS1-1d without using the merged-NPI option.[19] Since the changes in these distance functions are too complex, we summarized the dynamic reaction paths to reach the VRT region based on the schematic reaction network in Fig.8. By the trajectory mapping analysis, the 200 trajectories were classified into five types according to the number of initial hopping from TS1-1d to other NPI isomers of TS1-1d: (A) no hopping (96 trajectories), (B) one hopping (74 trajectories), (C) two hopping (23 trajectories), (D) three hopping (6 trajectories), and (E) four hopping (1 trajectory). We found that almost half of the trajectories (96/200) descended along the original IRC without hopping, while the others (104/200) remained in the TS1-1d region before descending along the IRC.[19]

To simplify the discussion, we define three types of VRTs: VRT_{IRC} is directly linked to a particular TS1-1d that stays after hopping, VRT_{bifur} is located in an IRC next to VRT_{IRC}, and VRT_{other} is a VRT excluding VRT_{IRC} and VRT_{bifur}. **Figure 9** shows the initial dynamical processes in (A) and (B). In the case of (A), most trajectories proceed directly to VRT_{IRC} (A1: 86/96), and after

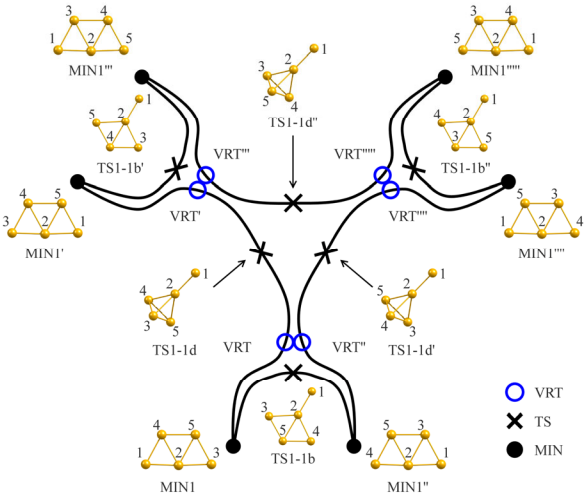

Figure 8. Schematic bifurcation reaction network associated with the three IRCs from the three NPI isomers of TS1-1d (cross marks) to the six NPI isomers of MIN1 (black circles). The three related NPI isomers of TS1-1b (cross marks) and the six NPI isomers of VRT1 (blue circles) are also shown.

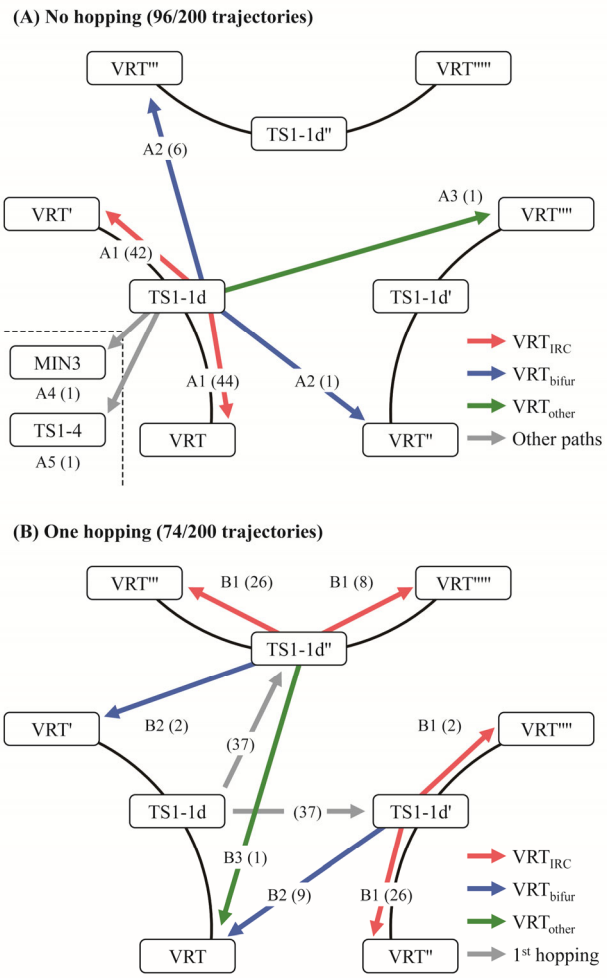

Figure 9. Schematic diagram of the initial dynamical process from TS1-1d to VRT: (A) without hopping (96/200) and (B) with one hopping (74/200). Here, three TS1-1d and six VRT are arranged as in Fig. 8. The red, blue, and green arrows indicate the reaction paths from the respective TS1-1d to VRT_{IRC}, VRT_{bifur}, and VRT_{other}. The number of trajectories is shown in parentheses.

passing through VRT_{IRC} (VRT and VRT'), these trajectories should undergo a bifurcation reaction. Such a process is a well-known bifurcation mechanism. However, some trajectories proceed to VRT_{bifur} (A2: 7/96) and VRT_{other} (A3: 1/96) after leaving TS1-1d. These results suggest that the molecule jumps from the TS1-1d region of the original IRC to the VRT region of the next IRC as it descends on the PES, which is called the *"IRC-jump"* process. Exceptionally, the other two trajectories proceed directly to MIN3 and TS1-4 without going through the VRT (A4: 1/96 and A5: 1/96). In case (B), the trajectory initially undergoes an IRC-jump from TS1-1d to TS1-1d' (37/74) or TS1-1d'' (37/74); after hopping to TS1-1d', most trajectories descend toward VRT'' (B1: 26/37) instead of VRT'''' (B1: 2/37). Similarly, after hopping to TS1-1d'', more trajectories descend towards VRT''' (B1: 26/37), and the trajectories towards VRT''''' becomes minor (B1: 8/37). This bias in B1 can be attributed to the fact that these trajectories already descend towards VRT and VRT' in the first

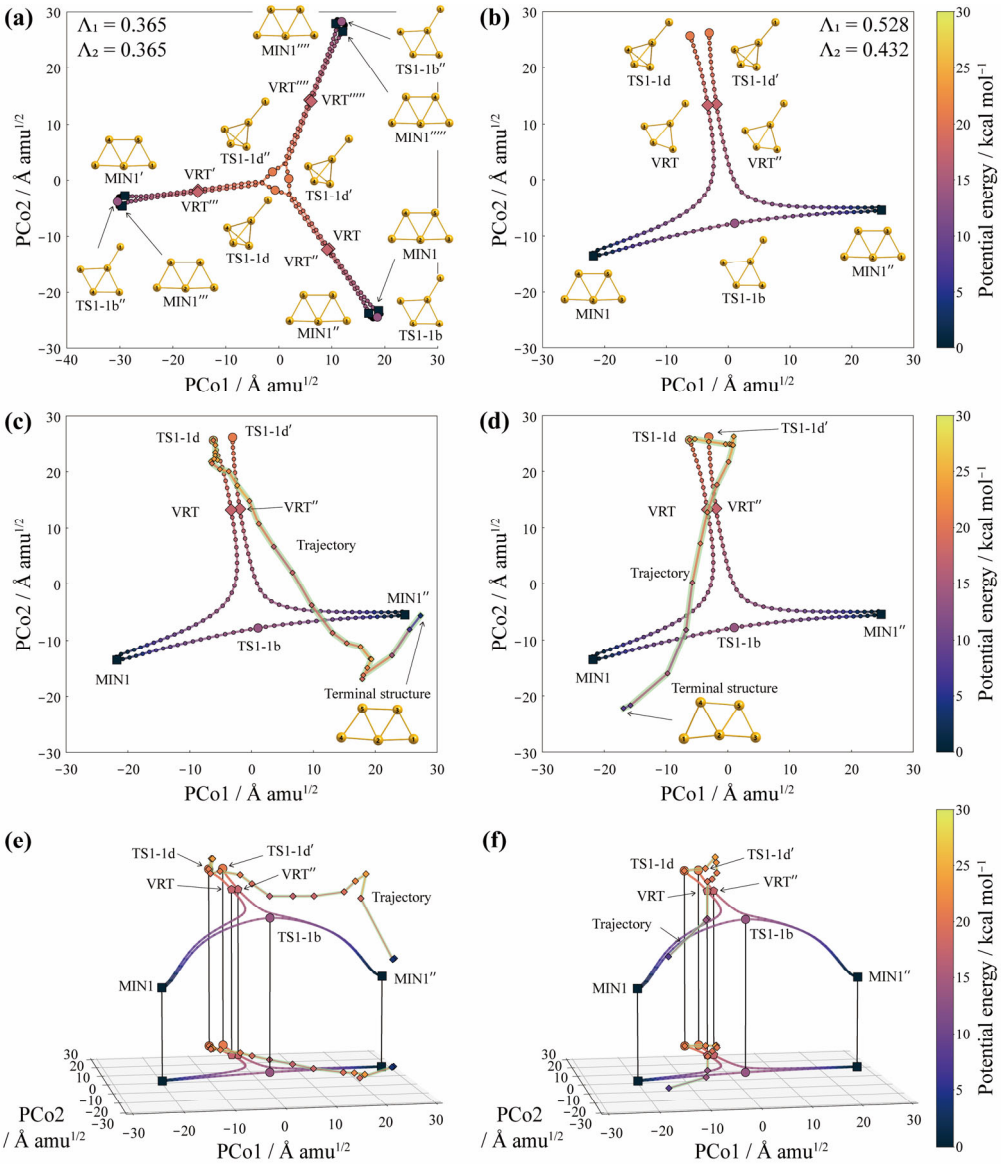

Figure 10. Two-dimensional reaction space determined from (a) six IRCs starting from TS1-1d, TS1-1d', TS1-1d'', TS1-1b, TS1-1b', and TS1-1b'' and (b) two parts of the IRCs from TS1-1d (TS1-1d') to MIN1 (MIN1''), and one IRC connecting MIN1 and MIN1'' *via* TS1-1b. (c-d) On-the-fly trajectories projected onto the two-dimensional reaction space in (b): (c) the trajectory going to the bifurcated product region, (d) the trajectory going to the original product region. (e) and (f) show the potential energy landscapes of (c) and (d), respectively. Circles, squares, and diamonds indicate TS, minima, and VRT. Smaller circles and smaller diamonds indicate structures on the IRC and along the trajectory (highlighted in light green). The color map shows the potential energy relative to MIN1.

hopping. As a result, 62 trajectories go to VRT_{IRC} (B1: 62/74), and a few trajectories go to VRT_{bifur} (B2: 11/74) and VRT_{other} (B3: 1/74). Among the remain 30 trajectories, 20 trajectories proceed to VRT_{IRC} after (C) two hopping, and three trajectories proceed to VRT_{bifur}. The remaining seven trajectories proceed to VRT_{IRC} after (D) three and (E) four hopping. In summary, the on-the-fly trajectory mapping method can reveal not only the conventional branching process that directly reaches VRT and VRT'' (A1) but also the non-trivial branching process associated with the IRC-jump before reaching the VRT region (A2-3 and B1-3).[19]

Next, the ReSPer program was used to project multiple IRCs into the two-dimensional reaction space.[21] As reference structures, 387 molecular structures were considered along six IRCs starting from TS1-1d, TS1-1d', TS1-1d'', TS1-1b, TS1-1b', and TS1-1b'' as shown in Fig. 8. **Figure 10a** shows the six IRCs projected onto the two-dimensional principal coordinates determined by the 387 reference structures. The proportion of each principal coordinate has the same value, 0.365, indicating that these two principal coordinates belong to the doubly degenerate symmetric representation. Although the cumulative proportion of the two principal coordinates, PCo1 and PCo2, is 0.730, the three IRCs associated with TS1-1b are

not clearly shown. This is because the chosen principal coordinates contribute significantly to the distribution of the three IRCs starting from TS1-1d. To reconstruct a suitable partial reaction space for discussing the bifurcation, we redefined the reference structures: two parts of the IRC from TS1-1d (TS1-1d') to MIN1 (MIN1'') and the IRC connecting MIN1 and MIN1'' *via* TS1-1b, for a total of 130 structures. Fig. 10b shows the reconstructed two-dimensional reaction space determined by the three selected reaction paths; the proportions of PCo1 and PCo2 are $\Lambda_1 = 0.528$ and $\Lambda_2 = 0.432$, with a cumulated proportion of 0.960. As expected, all the IRCs associated with the bifurcation reactions are well distributed in the reconstructed two-dimensional reaction space, which is an enlarged subset of Fig. 10a.

Finally, typical on-the-fly trajectories were projected onto the constructed two-dimensional reaction space.[21] Fig. 10c shows the case where the trajectory goes through an IRC-jump process and finally reaches MIN1'' corresponding to the branching product. Figure 10e shows the potential energy landscape, corresponding to the two-dimensional diagram in Figure 10c with the potential energy axis added as the third axis. The IRC-jump process is caused by the fact that the geometry of the PES orthogonal to the reaction coordinate becomes close to a flat near the VRT region. Interestingly, the projected trajectory jumps from one IRC to another before reaching the VRT region. This indicates that the close proximity of the two IRCs lowers the potential barrier for the IRC-jump to occur and promotes the bifurcation reactions. Fig. 10d and 10f show that the trajectory moves to the vicinity of TS1-1d' due to the IRC-jump near TS1-1d, but it returns to the original IRC *via* VRT'' and finally reaches MIN1. In summary, the ReSPer analysis revealed that dynamical reaction processes, such as IRC-jump, determine the final product in the branching reaction of Au_5.

3.2.2 Isomerization Reaction Analysis based on Global Reaction Route Map. In this section, we investigate the mechanism of structural change in Au_5 by using the on-the-fly trajectory mapping method based on the global reaction route map. First, the distance functions were calculated for 19 reference structures corresponding to 5 minima and 14 TSs using the merged-NPI option, and the optimal distance functions were obtained even if the on-the-fly trajectories belong to different atomic mappings of the reference structures.[19]

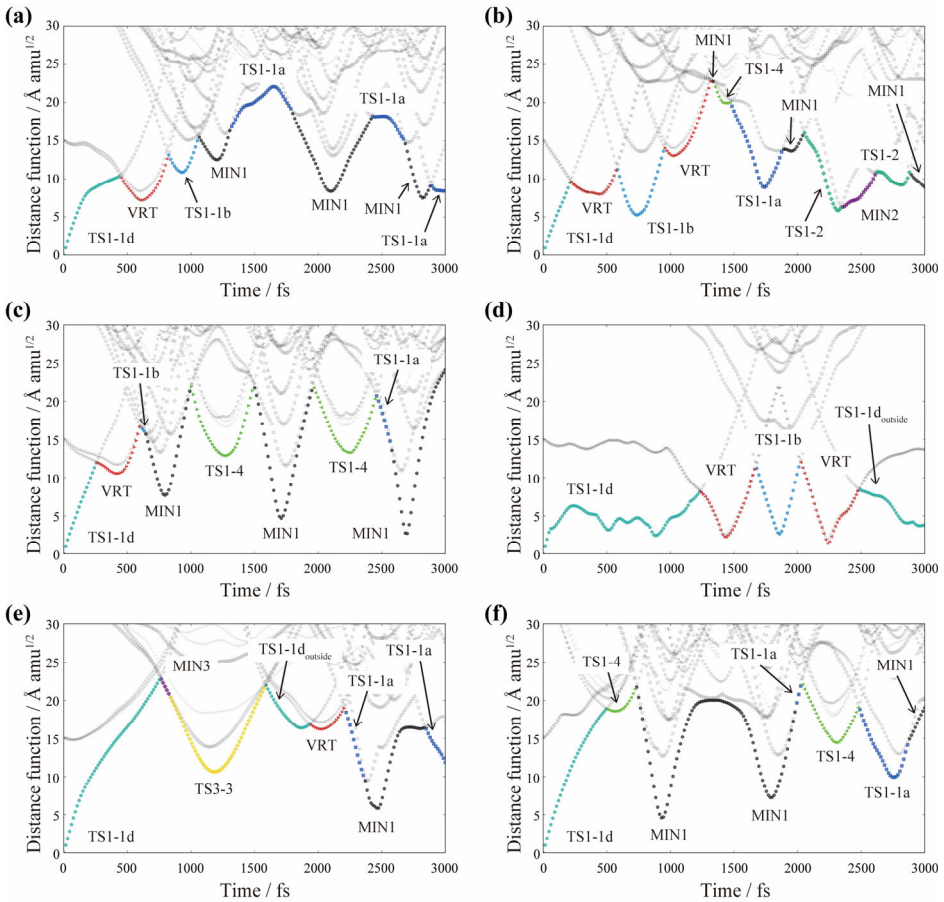

Figure 11. Variations of the distance function from the 19 reference structures of the Au_5 cluster (corresponding to 5 minima and 14 TSs). (a) Trajectory directly along the IRC to MIN1, (b) trajectory through TS1-1b and finally approaching MIN1, (c) trajectory approaching the non-planar structure of TS1-4, (d) trajectory that travels through the TS1-1d region in four hops and reaches a specific NPI isomer of TS1-1d except for TS1-1d, TS1-1d', and TS1-1d'' (called TS1-1d$_{outside}$), and (e-f) rare trajectory that directly approaches MIN3 and TS1-4 (corresponding to A4 and A5 in Fig. 9).

Figure 11 shows the evolution of the distance function along the six on-the-fly trajectories. Since all MD simulations start from TS1-1d, the distance function from TS1-1d is exactly zero at 0 fs. Fig. 11a shows that the trajectory starts from TS1-1d, descends on the IRC while passing the VRT, approaches TS1-1b located near the VRT, and then reaches MIN1 at 1200 fs. This is the most typical process, and 86 trajectories have been assigned. Fig. 11b shows that after the trajectory leaves TS1-1d, it approaches TS1-1b, which is a planar structure, and then proceeds to change its structure from planar to non-planar (1000-1500 fs) with the same momentum, eventually reaching around MIN1. In Fig. 11c, after reaching MIN1, the reaction proceeds to the non-planar structure as in Fig. 11b and tries to cross TS1-4, but it cannot reach MIN4 due to the relatively high activation barrier of TS1-4 (19.4 kcal/mol). The reaction process to regenerate the non-planar structure is thought to be accelerated by the momentum associated with the structural change from the non-planar structure (TS1-1d) to the planar structure (e.g., MIN1). In Fig. 11d, as described in the previous subsection, the trajectory hops four times through the three NPI isomers of TS1-1d before crossing the TS1-1d region. Then, it reaches the specific NPI isomers of TS1-1d except for TS1-1d, TS1-1d', and TS1-1d'' (called TS1-1d$_{outside}$). Figs. 11e and 11f show the very rare trajectories that approach MIN3 and TS1-4 directly without going through VRT/TS1-1b, corresponding to A4 and A5 in Fig. 9, respectively.

Figure 12 shows the dynamic reaction path obtained by the on-the-fly trajectory mapping method and schematically summarizes the reaction mechanism from TS1-1d to the first arrival MIN1, except for two trajectories that do not approach VRT/TS1-1b (Fig. 11e and f). After passing through VRT/TS1-1b, 198 trajectories approach seven different reference structures: (I) MIN1 (86/198), (II) TS1-3a (62/198), (III) TS1-1d$_{outside}$ (22/198), (IV) TS1-1a (16/198), (V) TS1-2 (7/198), (VI) TS1-4 (4/198), and (VII) TS1-3b (1/198). Note that eight trajectories, (III) (7/22) and (VII) (1/1), did not reach MIN1 during the 3 ps period. Surprisingly, more than half of the trajectories (114/200) did not descend near the IRC connecting TS1-1d and MIN1 and approached other TSs before reaching MIN1, especially 62 trajectories approached TS1-3a. This implies that the dynamic reaction path from TS1-1d to MIN1 should be discussed on the basis of multiple elementary reaction paths rather than a single IRC, supporting the existence of an IRC-jump process due to closely connected reaction paths in configuration space. More detailed results are summarized in ref [19].

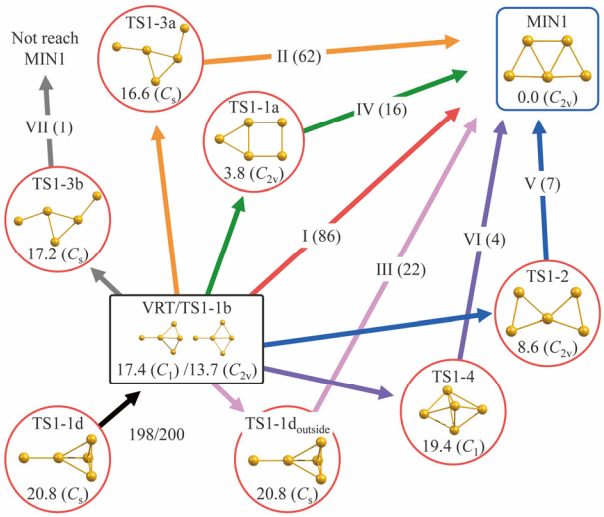

Figure 12. Schematic diagram of the dynamic reaction process of 198 on-the-fly trajectories from TS1-1d to MIN1 *via* VRT/TS1-1b. Seven different routes from I to VII are shown with the number of trajectories assigned. Note that TS1-1d$_{outside}$ is a specific NPI isomer of TS1-1d, except TS1-1d, TS1-1d', and TS1-1d'', as shown in Fig. 11.

In order to validate the dynamic reaction process based on the reduced-dimensionality reaction space, the ReSPer analysis was applied to 19 structures (5 minima and 14 TSs) using the merged-NPI option.[20] **Figure 13** shows the dispersion proportions of the 12 principal coordinates with positive eigenvalues obtained from the ReSPer analysis, the number of which corresponds to the number of degrees of freedom for rotation and vibration of Au_5 ($3N - 3 = 12$). The top three proportions of variance are 0.466, 0.200, and 0.143, and then the cumulated proportions of the first two and three principal coordinates are 0.666 and 0.809. Therefore, the configuration of the minima and TSs arranged in the reduced two- or three-dimensional subspaces may be useful for intuitively understanding the interrelationships of each structure.

Figure 14 shows the distributions along each principal coordinate PCo*i*, (a) PCo1, (b) PCo2, and (c) PCo3, for 19 reference structures (5 minima and 14 TSs). Five reference structures are also shown to gain insight into the physical meaning of each principal coordinate. According to Fig. 14a, the number of Au-Au bonds increases with increasing PCo1 (TS3-5: four, MIN1: seven, TS4-4: nine), and the molecular shape varies as quasi-linear (TS3-5), two- dimensional planar (TS1-5, TS1-3b, and MIN1), and three-dimensional non-planar structures (MIN4, TS1-4, and TS4-4). This result is in good agreement with the tendency of small gold clusters to prefer a two-dimensional planar structure due to relativistic effects.[49] Fig. 14b shows the structural change

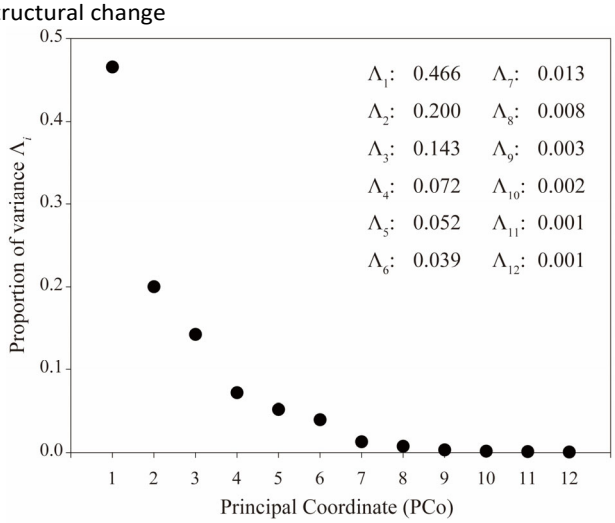

Figure 13. Dispersion proportions of the 19 structures on the global reaction route map of Au_5 obtained by the ReSPer analysis.

along PCo2 with blue arrows: the top-left and bottom-left gold atoms rotate clockwise when TS3-3 → TS1-3a → MIN1, and the center gold atom descends when MIN1 → TS1-2 → MIN2. Fig. 14c shows the change in molecular structure along PCo3. MIN1, TS1-1b, TS1-3b, TS1-5, and TS1-1d are composed of a triangle of Au_3 (shown by a blue triangle) and two Au atoms. As PCo3 increases, the two Au atoms move as shown by the blue arrows, eventually forming the non-planar structure of TS1-1d.

Figure 15a shows a two-dimensional global reaction route map of Au_5 with principal coordinates determined based on the ReSPer method, containing 5 minima and 14 TSs. Fig. 15b shows the same reaction route map, together with the molecular structure of each reference structure. The reduced-dimensionality global reaction route map of Au_5 here was automatically constructed based on the pairwise distance matrix of all reference structures using the ReSPer method with the merged-NPI option.[19,53] In the two-dimensional reaction space shown in Fig. 15a, the global minimum, MIN1, is located approximately at the center, and the other four minima are distributed around MIN1. Furthermore, TS*i-i* is located near MIN*i*, and TS*i-j* is located between MIN*i* and MIN*j*, which is consistent with chemical intuition about the nature of

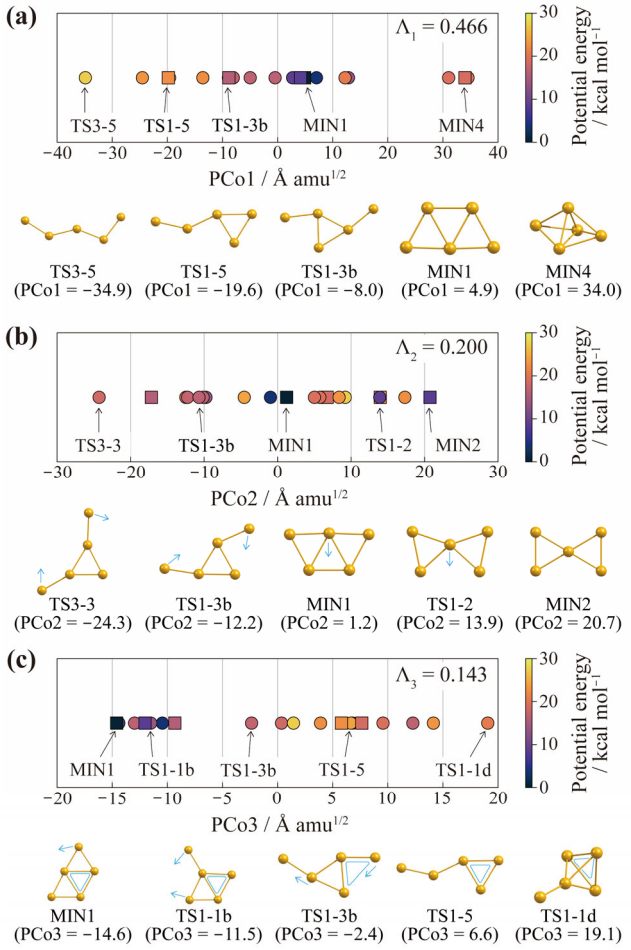

Figure 14. Distribution along the principal coordinates of 19 structures of Au_5 consisting of 5 minima and 14 TSs in the global reaction route map. (a) PCo1, (b) PCo2, and (c) PCo3.

multidimensional potential energy surface. This map shows that the initial structure of the on-the-fly MD simulation, TS1-1d, is relatively close to TS1-3a, which explains why many trajectories show a dynamical behavior approaching TS1-3a from TS1-1d, as described in Fig. 12. Similarly, based on the two-dimensional reaction space, we can explain why two trajectories approach directly from TS1-1d to MIN3 and TS1-4 (Figs. 11e and f). The configuration and linkage relationships of the reference structures constructed by the ReSPer analysis provide deep insight into the isomerization mechanism of Au_5.[20]

In order to analyze the dynamic reaction paths based on a reduced-dimensionality reaction space containing molecular structures along the IRC paths, 336 reference structures along 14 IRCs were selected, and the two-dimensional reaction space was constructed using the ReSPer program.[21] The merged-NPI option was turned on, and the length of each IRC was thinned down to approximately 10 steps. The proportion of variance was (Λ_1, Λ_2) = (0.436, 0.206), and the cumulated proportion was 0.642. **Figure 16a** shows the two-dimensional global reaction route map obtained by the ReSPer method based on 336 reference structures. Despite the dramatic increase in the number of reference structures, the configuration of each stationary point and the proportion of variance are almost the

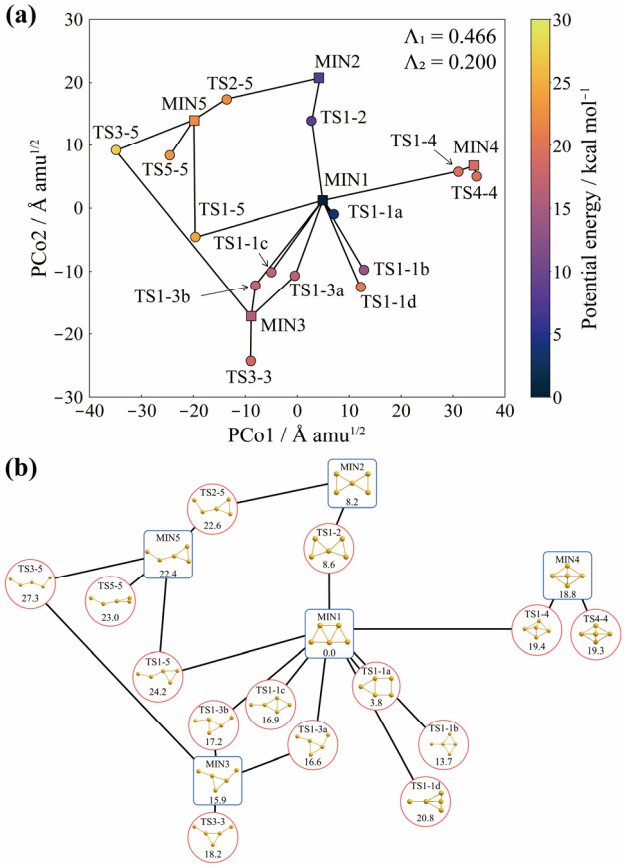

Figure 15. (a) Two-dimensional global reaction route map of Au_5 constructed by the ReSPer program, with 5 minima (blue squares) and 14 TSs (red circles) connected by black lines. (b) The same configuration of the global reaction route map in (a), with the addition of molecular structures. The relative energies (in kcal/mol) and the point groups for each structure are also shown.

same as in Fig. 15a. With the merged-NPI option, the forward and backward IRC parts starting from TS*i*-*i* that connect the different NPI isomers of MIN*i* are merged. For example, both sides of the IRC starting from TS3-3, which connects the two NPI isomers of MIN3 in the configurational space, are merged and appear as one path in this figure.

In Fig. 16a, the profile of IRCs between MIN2, MIN3, MIN4, and MIN5 is well described, but some IRCs, such as those involving TS1-1a, TS1-1b, TS1-1c, and TS1-1d, which connect the NPI isomers of MIN1, overlap with other IRCs. In such a situation, the dynamic reaction path originating from TS1-1d would not be adequately represented in this reaction space. Therefore, we performed the ReSPer analysis again for 260 reference structures selected from seven IRCs connected to MIN1 *via* TS1-1a, TS1-1b, TS1-1c, TS1-1d, TS1-2, TS1-3a, and TS1-3b in the relatively low-energy region around MIN1.[21]

Figs. 16b and 16c show the reconstructed two- and three-dimensional reaction spaces, where the proportions of variances are (Λ_1, Λ_2, Λ_3) = (0.385, 0.269, 0.153) and the cumulated proportions are 0.654 and 0.807 in the two- and three-dimensional reaction spaces. The two-dimensional reaction space in Fig. 16b reproduces the network structure of the IRC better than Fig. 16a, but the IRCs involved in TS1-1a, TS1-1b, TS1-1c, TS1-3a, and TS1-3b still overlap with others. On the other hand, Fig. 16c with the addition of PCo3 solves this problem and describes well the network structure of the entire IRC in the low-energy region. In fact, in the three-dimensional reaction space of Fig. 16c, the two IRCs connecting MIN1 and MIN3 are located far from TS1-1d, which is evident by comparing their respective structures in Fig. 15b. This result suggests that PCo3 has information that can describe the low-energy region of Au_5 in more detail than PCo1 and PCo2 can. The dynamic reaction paths were investigated by projecting the on-the-fly trajectory onto the reduced-dimensionality reaction space of Au_5.[21] Figs. 16b and 16c show the second most common trajectory, which corresponds to (II) in Fig. 12, departing from TS1-1d and approaching TS1-3a directly. These figures show that the trajectory deviated from the IRC after approaching the VRT vicinity and proceeded to the TS1-3a region. It then seems to stay near TS1-3a and finally reach MIN1. This behavior indicates that the molecular system is in the process of jumping from the IRC of TS1-1d to the IRC of TS1-3a, suggesting that the IRC tends to become dense near the minimum.

Figure 17 shows the two-dimensional potential energy landscape, which gives an overview of the PES in the reduced two-dimension, indicating an intuitive understandable dynamic reaction path descending through the PES in the order TS1-1d, TS1-3a, and MIN1. The projected trajectory leaves the IRC originating from TS1-1d and approaches the IRC originating from TS1-3a. In view of Figs. 16c and 17, it has been verified that the excitation of vibrational modes orthogonal to the reaction coordinate facilitates the IRC-jump process from the valley of the IRC starting from TS1-1d to another IRC associated with TS1-3a. However, it eventually returns to the original IRC by a re-IRC-jump before being captured by the IRC starting from TS1-3a.

The ReSPer analysis reveals the behavior of on-the-fly trajectories and their driving forces based on a reduced-dimensionality reaction space determined by properly designed reference structures. This powerful tool provides a more intuitive understanding of chemical reaction mechanisms by

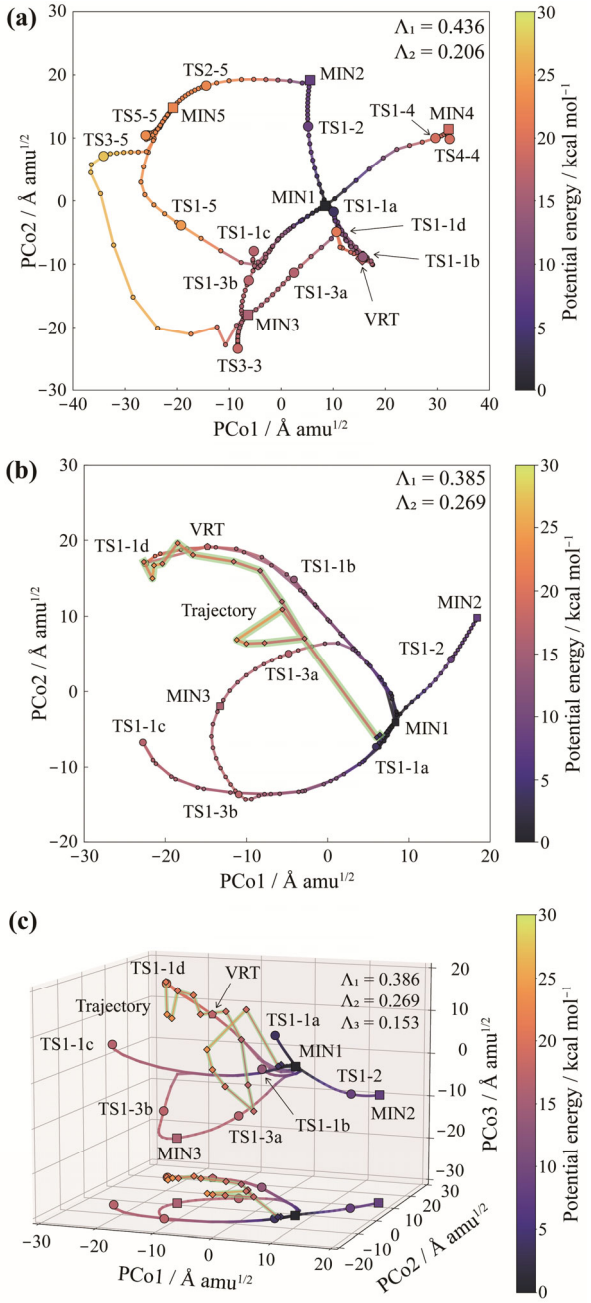

Figure 16. Reduced-dimensionality reaction space of Au_5 as determined by the ReSPer program: (a) two-dimensional reaction space determined by 336 reference structures along all IRCs; (b) two-dimensional reaction space and (c) three-dimensional reaction space determined by 260 reference structures along seven IRCs around MIN1. Projected trajectory is also shown in (b-c). Squares, circles, and diamonds indicate minima, TS, and VRT. The small circles and small diamonds indicate the structures along the IRC and trajectory (highlighted in light green). The color map shows the potential energy relative to MIN1.

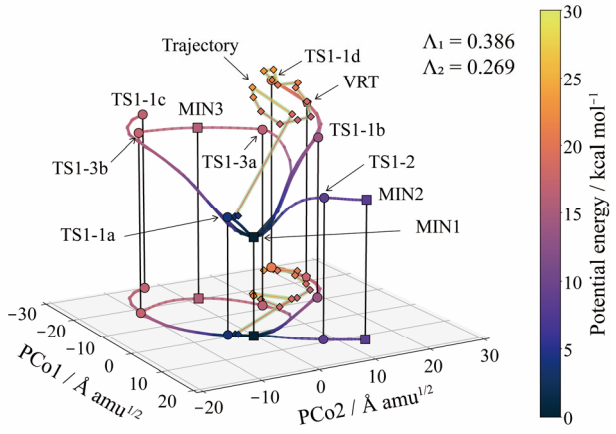

Figure 17. Reduced-dimensionality potential energy landscape of Au_5 as determined by the ReSPer program. Projected trajectory (the same as Fig. 16) is also shown. Squares, circles, and diamonds indicate minima, TS, and VRT. The small circles and small diamonds indicate the structures along the IRC and trajectory (highlighted in light green). The color map shows the potential energy relative to MIN1.

exploring the PES in full dimension and tracking the dynamic reaction process.[21]

4. Conclusion

Under the Born–Oppenheimer approximation, elementary chemical reaction processes have been treated separately in electronic structure theory and reaction dynamics, and reaction pathways defined on potential energy surfaces have played an important role as a concept that connects the electronic structure theory and reaction dynamics. IRC is a typical reaction path defined for elementary reactions utilized in studies based on static quantum chemical calculations. Reaction dynamics, on the other hand, is directly concerned with the process of motion of molecular systems on a potential energy surface. The on-the-fly MD method, which solves Newton's equation of motion numerically by calculating the forces acting on each atom from electronic structure calculations, is a very effective research method because it can describe not only the elementary reactions assumed in advance but also the motion of molecular systems determined by potential energy surfaces without any restrictions. We have developed an analytical method that combines static reaction path analysis and dynamic on-the-fly MD method to further understand chemical reactions. In this paper, we review our *on-the-fly trajectory mapping* method and *Reaction Space Projector* (ReSPer) method.

The key concept of our strategy is to describe the similarity of molecular structures as the linear distance in mass-weighted Cartesian coordinates between a pair of molecular structures. In order to better evaluate the pairwise distances, it is necessary to 1) properly align the Cartesian coordinate axes of a pair of molecules and 2) if necessary, consider nuclear permutation-inversion (NPI) isomers to compare molecular structures with different atomic numberings. The distance function is defined as the set of linear distances between each molecular structure in the time series data and the reference structure. Therefore, by following the distance function between the on-the-fly trajectory and the representative structures on the PES, we can intuitively recognize how the molecule is moving through the set of reference structures on the PES. This analysis method is called "*on-the-fly trajectory mapping*."

This method provides information on which reference structure on the 3*N*-6 dimensional (*N* is the number of atoms) PES the molecular system is moving toward, but it does not provide an intuitive picture of how the system moves through the reference structure network. Therefore, we developed the "*Reaction Space Projector (ReSPer)*" method by applying the classical multidimensional scaling method, which is a dimensionality reduction method that can reduce data in multidimensional space to a small number of important dimensions. In this method, a distance matrix reflecting the similarity between structures is first defined for a set of reference structures determined by static reaction path calculations and then transformed into principal coordinates. Furthermore, the dynamic trajectory data is projected into this principal coordinate space using the out-of-sample extension method. By this analysis, we successfully project the on-the-fly trajectories into the low-dimensional reaction space determined for the global reaction route map consisting of IRCs.

To demonstrate the usefulness of these new methods, we applied them to the analysis of chemical reaction mechanisms in the OH^- + CH_3F collision reaction and the bifurcation and isomerization reaction of Au_5. In the former collision reaction, the on-the-fly trajectory mapping method revealed that the trajectories along the IRC path prefer to leave the highly curved reaction path. In addition, the potential energy landscape generated by ReSPer shows that the IRC profile plateaus near the high curvature region and the molecular momentum pushes the trajectory out of the plateau region. In the latter application to Au_5, we succeed in capturing dynamic reaction processes such as IRC jumping behavior, in which a molecule moves from the vicinity of one IRC to another, which cannot be explained by conventional IRC-based reaction analysis.

Under high energy conditions, such as combustion and photoinactivation reactions, the IRC network loses its role as a reference reaction route map. However, our strategy provides a reasonable dynamic reaction route as long as the reduced-dimensionality reaction space defines the critical region for the reaction. Therefore, the proposed method can work as a robust tool to elucidate various kinds of realistic chemical reaction mechanisms with dynamic effects based on reaction route maps. In the future, we hope to apply these methods to more general chemical reactions, such as organic reactions and photoexcited reactions, which will lead to the elucidation of chemical reaction mechanisms and the discovery of novel chemical reactions.

Author Contributions

We strongly encourage authors to include author contributions and recommend using CRediT for standardised contribution descriptions. Please refer to our general author guidelines for more information about authorship.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors are sincerely grateful to Prof. Z. Arai (Chubu Univ.) for collaborated works on the Reaction Space Projector method. We also thank Prof. S. Maeda (Hokkaido Univ.) and Prof. Harabuchi (Hokkaido Univ.) for collaborated works on the on-the-fly trajectory mapping method. This work was partly supported by the Elements Strategy Initiative of MEXT (JPMXP0112101003), the Photo-excitonix Project in Hokkaido University, and JST CREST Grant Number JPMJCR1902, Japan. A part of calculations was performed using the Research Center for Computational Science, Okazaki, Japan.

Notes and references

1 K. Fukui, *J. Phys. Chem.*, 1970, **74**, 4161–4163.
2 S. Maeda, T. Taketsugu, K. Morokuma and K. Ohno, *Bull. Chem. Soc. Jpn.*, 2014, **87**, 1315–1334.
3 K. Ohno and S. Maeda, *Chem. Phys. Lett.*, 2004, **384**, 277–282.
4 S. Maeda, Y. Harabuchi, M. Takagi, K. Saita, K. Suzuki, T. Ichino, Y. Sumiya, K. Sugiyama and Y. Ono, *J. Comput. Chem.*, 2018, **39**, 233–250.
5 S. Maeda and K. Morokuma, *J. Chem. Theory Comput.*, 2011, **7**, 2335–2345.
6 S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu and K. Morokuma, *Chem. Rec.*, 2016, **16**, 2232–2248.
7 Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T. Taketsugu and S. Maeda, *J. Phys. Chem. A*, 2015, **119**, 11641–11649.
8 T. Mita, Y. Harabuchi and S. Maeda, *Chem. Sci.*, 2020, **11**, 7569–7577.
9 Y. Sumiya and S. Maeda, *Chem. Lett.*, 2020, **49**, 553–564.
10 W. H. Miller, N. C. Handy and J. E. Adams, *J. Chem. Phys.*, 1980, **72**, 99–112.
11 S. Kato and K. Morokuma, *J. Chem. Phys.*, 1980, **73**, 3900–3914.
12 T. Taketsugu and M. S. Gordon, *J. Chem. Phys.*, 1995, **103**, 10042–10049.
13 T. Taketsugu and T. Hirano, *J. Chem. Phys.*, 1993, **99**, 9806–9814.
14 S. R. Hare, A. Li and D. J. Tantillo, *Chem. Sci.*, 2018, **9**, 8937–8945.
15 Y. Ootani, K. Satoh, A. Nakayama, T. Noro and T. Taketsugu, *J. Chem. Phys.*, 2009, **131**, 194306.
16 Y. Harabuchi, K. Keipert, F. Zahariev, T. Taketsugu and M. S. Gordon, *J. Phys. Chem. A*, 2014, **118**, 11987–11998.
17 Y. Harabuchi, R. Yamamoto, S. Maeda, S. Takeuchi, T. Tahara and T. Taketsugu, *J. Phys. Chem. A*, 2016, **120**, 8804–8812.
18 T. Tsutsumi, Y. Harabuchi, R. Yamamoto, S. Maeda and T. Taketsugu, *Chem. Phys.*, 2018, **515**, 564–571.
19 T. Tsutsumi, Y. Harabuchi, Y. Ono, S. Maeda and T. Taketsugu, *Phys. Chem. Chem. Phys.*, 2018, **20**, 1364–1372.
20 T. Tsutsumi, Y. Ono, Z. Arai and T. Taketsugu, *J. Chem. Theory Comput.*, 2018, **14**, 4263–4270.
21 T. Tsutsumi, Y. Ono, Z. Arai and T. Taketsugu, *J. Chem. Theory Comput.*, 2020, **16**, 4029–4037.
22 T. T. Tanimoto, *An elementary mathematical theory of classification and prediction*, New York, NY, 1958.
23 T. Fukutani, K. Miyazawa, S. Iwata and H. Satoh, *Bull. Chem. Soc. Jpn.*, 2020, bcsj.20200258.
24 W. Kabsch, *Acta Crystallogr. Sect. A*, 1976, **32**, 922–923.
25 B. Olasz and G. Czakó, *Phys. Chem. Chem. Phys.*, 2019, **21**, 1578–1586.
26 G. Czakó, T. Győri, B. Olasz, D. Papp, I. Szabó, V. Tajti and D. A. Tasi, *Phys. Chem. Chem. Phys.*, 2020, **22**, 4298–4312.
27 E. Martínez-Núñez, *J. Comput. Chem.*, 2015, **36**, 222–234.
28 E. Martínez-Núñez, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14912–14921.
29 S. Vázquez, X. Otero and E. Martinez-Nunez, *Molecules*, 2018, **23**, 3156.
30 T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G. J. Rylance, R. L. Johnston and D. J. Wales, *J. Chem. Phys.*, 2005, **122**, 084714.
31 S. R. Hare, L. A. Bratholm, D. R. Glowacki and B. K. Carpenter, *Chem. Sci.*, 2019, **10**, 9954–9968.
32 J. Peng, Y. Xie, D. Hu and Z. Lan, *J. Chem. Phys.*, 2021, **154**, 094122.
33 P. Pisani, F. Caporuscio, L. Carlino and G. Rastelli, *PLoS One*, 2016, **11**, 1–23.
34 X. Li, Y. Xie, D. Hu and Z. Lan, *J. Chem. Theory Comput.*, 2017, **13**, 4611–4623.
35 W. Shi, T. Jia and A. Li, *Phys. Chem. Chem. Phys.*, 2020, **22**, 17460–17471.
36 A. M. Virshup, J. Chen and T. J. Martínez, *J. Chem. Phys.*, 2012, **137**, 22A519.
37 W. S. Torgerson, *Psychometrica*, 1952, **17**, 401–419.
38 W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 3rd Edit., 2015.
39 G. Young and A. S. Householder, *Psychometrika*, 1938, **3**, 19–22.
40 Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux and M. Ouimet, in *In Proceedings of the 16th International Conference on Neural Information Processing Systems (NIPS'03)*, MIT Press, Cambridge, MA, USA, 2003, pp. 177–184.
41 M. W. Trosset and C. E. Priebe, *Comput. Stat. Data Anal.*, 2008, **52**, 4635–4642.
42 A. B. Oliveira, F. M. Fatore, F. V. Paulovich, O. N. Oliveira and V. B. P. Leite, *PLoS One*, 2014, **9**, e100861.

43 A. B. Oliveira Junior, X. Lin, P. Kulkarni, J. N. Onuchic, S. Roy and V. B. P. Leite, *J. Chem. Theory Comput.*, 2021, acs.jctc.1c00027.

44 L. Sun, K. Song and W. L. Hase, *Science*, 2002, **296**, 875–878.

45 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.

46 Y. Harabuchi, M. Okai, R. Yamamoto, T. Tsutsumi, Y. Ono and T. Taketsugu, 2020.

47 W. Zou, T. Sexton, E. Kraka, M. Freindorf and D. Cremer, *J. Chem. Theory Comput.*, 2016, **12**, 650–663.

48 M. Haruta, *Nature*, 2005, **437**, 1098–1099.

49 P. Pyykkö, *Angew. Chemie Int. Ed.*, 2004, **43**, 4412–4456.

50 M. Gao, A. Lyalin, M. Takagi, S. Maeda and T. Taketsugu, *J. Phys. Chem. C*, 2015, **119**, 11120–11130.

51 M. Sugiuchi, J. Maeba, N. Okubo, M. Iwamura, K. Nozaki and K. Konishi, *J. Am. Chem. Soc.*, , DOI:10.1021/jacs.7b10201.

52 S. Xu, Y. Zhang, Y. Wei and G. Tian, *Comput. Theor. Chem.*, 2020, **1190**, 113007.

53 Y. Harabuchi, Y. Ono, S. Maeda and T. Taketsugu, *J. Chem. Phys.*, 2015, **143**, 014301.

54 K. . O. K. . Maeda, S.; Osada, Y.; Morokuma, 2011.

55 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, 2016.