

# Weight Rank Clique Filtration を用いたパーシステント・ホモロジーによる実在分子の反応経路地図の可視化

北海道大学 大学院総合化学院 総合化学専攻  
村山武来 (Burai MURAYAMA)

## 概要

近年、化学の世界ではこれまでの実験による反応の探索ではなく、コンピューターを用いた計算による反応の探索が推し進められている。その探索のための基盤となるツールがポテンシャルエネルギー曲面 (PES) である。PES は  $3N - 6$  次元 ( $N$ : 原子数) の関数値であり、化学反応は PES 上の動態として考えることができる。PES からの情報抽出の方法として、GRRM プログラムが前田らによって開発された。このプログラムは PES 上の 2 つの平衡状態 (EQ) と 1 つの遷移状態 (TS) からなる素反応経路を固有反応経路 (IRC) で接続したネットワークである反応経路地図 (RRM) を指定された系について自動的に探索する。しかしながら、得られる RRM は非常に巨大であり、横断的に比較することが難しい。そこで本研究では、RRM を表す重み付きグラフのトポロジー的な記述子を、パーシステントホモロジー (PH) に基づき抽出する手法を提案する。本研究の結果、提案手法を用いて得られた記述子は、物質の化学反応特性や物理化学的特性を正確に反映していることが示唆された。本稿は小林正人准教授、中村壮伸博士、寺本央准教授、青木雅允氏、石橋卓氏、齋藤琢弥氏との共同研究に基づく。

## 1 背景

反応経路は、理論研究において化学反応を表現するための基礎となるものである。反応経路はポテンシャルエネルギー曲面 (PES) 上で定義され、これは  $N$  原子系の場合、 $(3N - 6)$  変数上のスカラー関数である [1, 2]。PES の一次鞍点と局所極小点はそれぞれ遷移状態 (TS) 構造と平衡構造 (EQ) を表している。TS を経由して 2 つの EQ 形状を結ぶ最急降下曲線は固有反応座標 (IRC) [3] と呼ばれ、素反応経路を表している。反応経路は原子の位置から  $(3N - 6)$  変数が決定されるため、化学反応による EQ から EQ への配置変化を表現しやすく、化学反応機構全般の理解に役立つと考えられている。

反応経路地図 (RRM) は、 $N$  原子化学系の化学反応を特徴付ける素反応経路の集合体である [4, 5]。RRM は、EQ や TS のエネルギーに対応した重みを持つ重み付きグラフで表現される。RRM を表す重み付きグラフは、反応速度や寿命など、化学反応の物理的特性を表現することができる。PES は次元が高いため可視化が困難だが、RRM は平面上に可視化することができる。したがって、RRM は PES の情報を縮小したものであり、可視化できない対象を可視化できるものに変換したものである。そのため、RRM は  $N$  原子系の化学反応の標準的な表現方法として広く使われている [6, 7]。

近年、グローバル反応ルートマッピング (GRRM) 戦略により、複雑な分子系の RRM を自動的かつ広範囲に構築することが可能になった [8]。GRRM プログラムは 2 つの方法論—anharmonic

downward distortion following (ADDF) 法 [9, 10] と artificial force-induced reaction (AFIR) 法 [11, 12, 13] から構成されている。ADDF 法は、ポテンシャル極小からの異性化反応や解離反応に伴い、一次元ポテンシャルエネルギー曲線が極小付近の調和ポテンシャルに対して下方に歪むことを利用したものである。AFIR 法は、2つの原子または原子団の間に人工的な力を加え、強制的に反応を誘導する、より直感的な方法である。ADDF 法や AFIR 法を適用することで、GRRM プログラムは自動的に EQ と TS の構造を順次特定することができる。このような特徴から、GRRM プログラムはナノクラスター触媒 [14] や有機合成設計 [15]、表面化学 [16] など、様々な分野で利用されている。しかしながら、GRRM 戦略の活用は、RRM の表現と特徴抽出の点で問題がある。一般に、グラフの可視化のレイアウトは任意である [17, 18]。これに対し、GRRM 戦略では、複数の素反応経路を含む大きなグラフが生成される。このような大きなグラフでは、同じグラフが異なる方法で恣意性をもって可視化され、グラフの本質的な特徴を捉えることが難しくなる。特に経路数が多いグラフでは、2つの異なる可視化が同じグラフを表しているかどうかを判断することは困難である。RRM は既に PES を十分に縮小した表現であるが、複雑なシステムでは前述の問題を克服し、RRM から豊富な情報を抽出するために、さらなる縮小が必要である。

この問題を解決するために、いくつかの方法が提案されている。Disconnectivity graph (DG) [19, 20] と Persistent Homology (PH) [21, 22] は汎用性のある手法で有力な候補である。

DG は、PES 上の安定状態の階層を木構造で表示する、現在最も普及している可視化手法であり、現実的な分子系やトイモデルを問わず多くのシミュレーションデータに適用されている [1]。しかし、本来の定義では、DG は必ずしも直結していない2つの安定状態間の最低エネルギー遷移障壁しか表すことしかできない。そのため、RRM によく見られる循環的な経路を表現することができず、関連する拡張がその後開発されたものの、DG は RRM によく見られる循環的な経路を表現することができない [23]。

また、最近 Mirth らによって PES のトポロジー的特徴を抽出するために PH が利用された [22]。PH は周期的な経路を表現することが可能であるが、Mirth らの論文の定式化では、PES の全定常点を解析的に位置づける必要がある。したがって、彼らの研究は反応経路と PH の関係を理解するための出発点として捉えることができるが、彼らの方法を GRRM で得られた RRM にそのまま適用することはできない。

そこで、私たちは GRRM によって得られた実在物質の RRM に PH を適用する方法を提案する [24]。この手法を金属ナノクラスターや有機分子系の RRM に適用し、本手法の RRM の特徴を抽出する能力を明らかにした。

## 2 手法

### 2.1 AFIR/ADDF 法

本研究では、RRM を、与えられた原子構造の EQ 構造と、EQ の対を IRC に沿って結ぶ TS 構造 (すなわち、一次鞍点) の集まりと、それに対応するエネルギーとで定義する。したがって、RRM は無向グラフ  $G = (V, E)$  で表され、頂点 ( $w_V(v_i)$ ,  $i$  番目の EQ のエネルギーに対応) と辺 ( $w_E(e_i)$ ,  $i$  番目の TS のエネルギーに対応) にエネルギーの重みが設定されている。実際には、ある系の RRM

は GRRM プログラムで実装されている ADDF 法または AFIR 法を用いて構築することができる [25, 26]。

ADDF 法 [10] は、EQ 構造で近似した調和ポテンシャルと実際のポテンシャルエネルギーとの差として評価される PES の非調和下降歪み (ADD) に従う方法である。ADD を探索する効率的な方法は、力定数の 1/2 乗でスケールされた法線座標で描かれた EQ 中心の超球を導入し、その上で調和近似レベルのエネルギーが一定であるようにすることである。スケールされた超球面上の各最小値は ADD に対応するので、超球の半径を大きくすれば ADD に沿って反応経路をたどることができる。ADDF 法は概念的にはよく定義されているが、各 EQ でスケールされた超球面を構築するためにヘシアン計算が必要であり、実用的な応用は限定的である。また、非調和歪み定常経路の分岐があるため、ADD 経路の追跡で必ずしもすべての反応経路を探索できるわけでもない [27]。

これに対し、RRM を構成する手順として、より直感的で実用的なのが AFIR 法 [13] である。この方法では、対象系のフラグメント間に人工的な力を導入し、エネルギー障壁のない反応を誘発させる。人工的な力によって増大したポテンシャルエネルギーを最小化することで得られる反応経路を AFIR 経路と呼ぶ。その後、人工力項を除去し、最大エネルギー点を近似的な TS 構造として採用し、実際の TS 構造を最適化する。AFIR 法では、モデル衝突エネルギーに相当するパラメータ  $\gamma$  を用いて、加える力をスケールする。大雑把に言うと、 $\gamma$  以下の反応障壁を持つ TS は AFIR 法で同定することができる。

## 2.2 Adjusted Weight Rank Clique Filtration

この節では、重み付き無向単純グラフのフィルター付きクリーク複体に対する PH について簡単に述べる。

まず単純グラフについて次のように書く [28, 29]。グラフ  $G = (V, E)$  は有限、無向、単純なものを考える。RRM では、 $V$  と  $E$  がそれぞれ EQ と TS の集合に相当する。辺は  $\overline{uv}$  と表す。頂点集合  $V$  は有限であるとする (従って、辺集合  $E$  も有限である)。 $G'$  が  $G$  の部分グラフであることを  $G' \subset G$  と書く。

次に、クリークとクリーク複体を導入する。グラフ  $G = (V, E)$  における  $p$ -clique とは、整数  $p \geq 2$  に対して、 $S$  が  $p$  個の頂点を含み、任意の  $u, v \in S$  が  $G$  の辺によって接続されているような  $V$  の部分集合のことである。1-clique とは、 $v \in V$  を含む集合  $\{v\}$  のことである。 $G$  のすべてのクリークの集まりを  $G$  の *clique complex* といい、 $X(G)$  と表記する。

RRM 上の構造エネルギーをグラフやクリーク複体で表現するために、グラフ上の実数値関数  $w_V: V \rightarrow \mathbf{R}$  と  $w_E: E \rightarrow \mathbf{R}$  の対で重みを導入する。文脈から  $w$  が  $w_V$  を指すのか  $w_E$  を指すのかが明らかな場合は、両方の重み関数を  $w$  と表記する。ここでは、以下の条件を満たす重み関数  $w: G \rightarrow \mathbf{R}$  のみを考える。

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

続いて、クリーク複体に対する重み付け  $w^G: X(G) \rightarrow \mathbf{R}$  を次のように定義する。 $p$ -clique

$\sigma = \{v_{i_1}, \dots, v_{i_p}\}$ ,  $(i_1 < \dots < i_p)$ , に対して

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

また同様に、

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

ここで、 $\varepsilon$  は負の数でない実数のパラメータである。したがって、 $w^G(\sigma; \varepsilon)$  を単に  $w^G(\sigma)$  と表す。構成上、重み付け  $w^G$  は次の条件を満たす。

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

このように定めると、クリーク複体に対する増大列を構成することができる。

まず、

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

を実数  $a \in \mathbf{R}$  に対して定義する。式 (1) の条件により、これは  $G$  のクリークの部分集合のうち重みが  $a$  以下である部分集合である。構成から、 $a < a'$  のとき、 $X(G)(a) \subset X(G)(a')$  となる。したがって、 $w^G$  の値の増加列  $a_0 < a_1 < \dots < a_l$  に対応する増加列がある。

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

この増大列を、 $\varepsilon \geq 0$  の重み付きグラフ  $(G, w)$  のクリーク複体  $X(G)$  において、 $(\varepsilon)$ -adjusted weight rank clique filtration と呼ぶ。また、 $\varepsilon = 0$  の場合、Petri らが提案した weight rank clique filtration と一致する [21]。

## 2.3 Persistet Homology

クリーク錯体は単体複体の構造を持っている。単体複体上のフィルトレーションが与えられれば、一般に PH を定義することができる。ここでは結果を参照するのに必要な情報のみまとめる。理論の詳細については [30] を参照せよ。

本研究結果を出力するのにあたって我々は Persistence Diagram (PD) あるいは Persistence Barcode (PB) を用いる。PD とは、PH の各生成元について生成時刻  $a_1$  から消滅時刻  $a_2$  までを点  $(a_1, a_2)$  として表したものであり、クリーク  $X$  に対して  $i$  次元の PH を  $\mathcal{D}_i(X)$  と表す。一方で PB は、PH の各生成元について生成時刻  $a_1$  から消滅時刻  $a_2$  までを区間  $(a_1, a_2)$  として表し、同様に  $\mathcal{B}_i(X)$  と表す。それぞれ図 1 のように表現される。

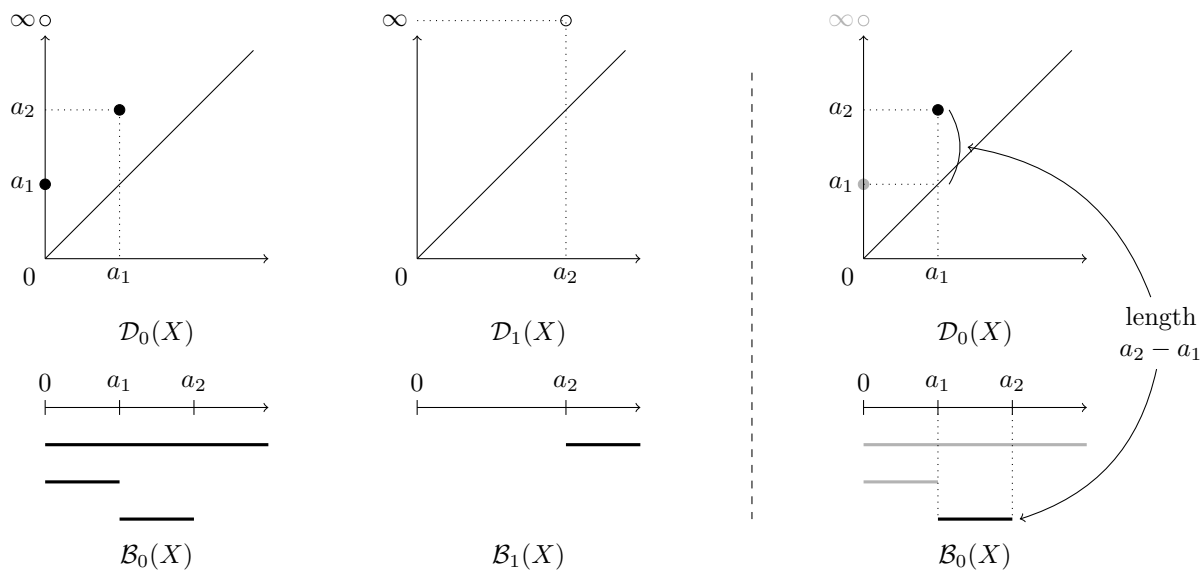


図 1: 0, 1 次 of PD と PB

本稿中では 0, 1 次 of PH までしか取り扱わないため、PD をまとめて表現するために図 2 のように、 $\mathcal{D}_1(X)$  を対角線を軸にして反転させ、まとめて表示する。

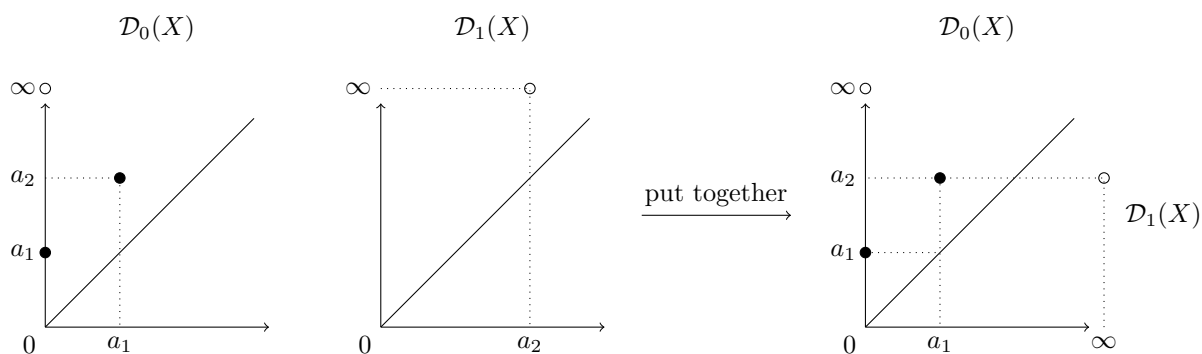
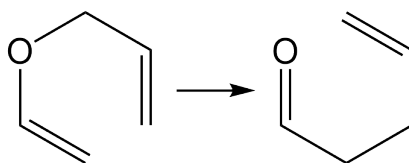


図 2: 本稿中での PD の表示

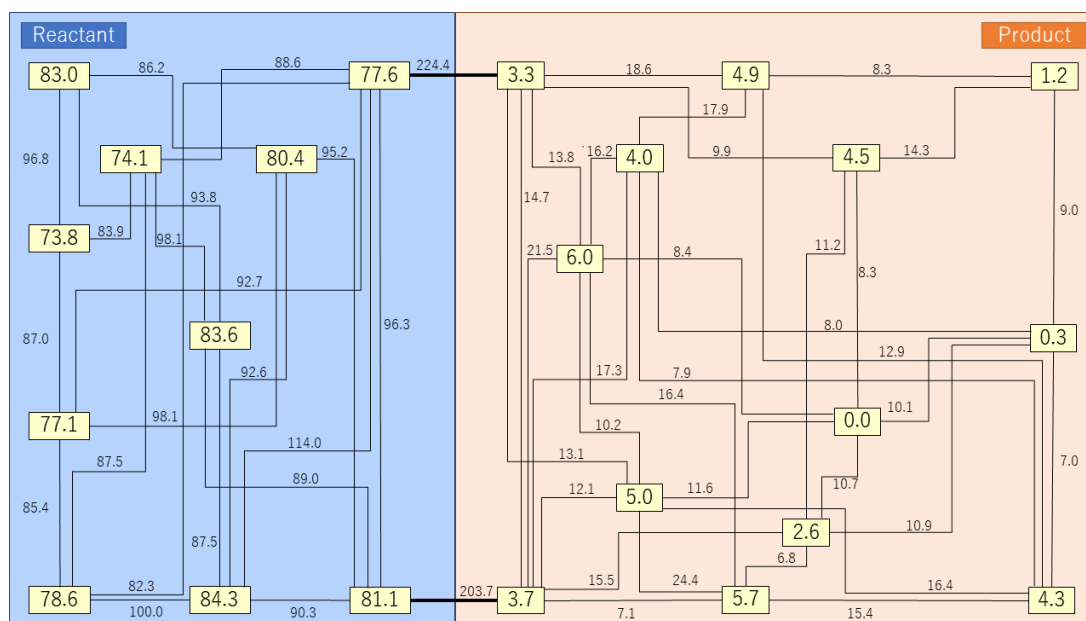
### 3 計算結果の紹介

貨幣金属ナノクラスター 5/7 原子 1/2 元系及びアリルビニルエーテルの RRM に対してパーシステントホモロジーを計算した。主な結果は次の 2 つである。

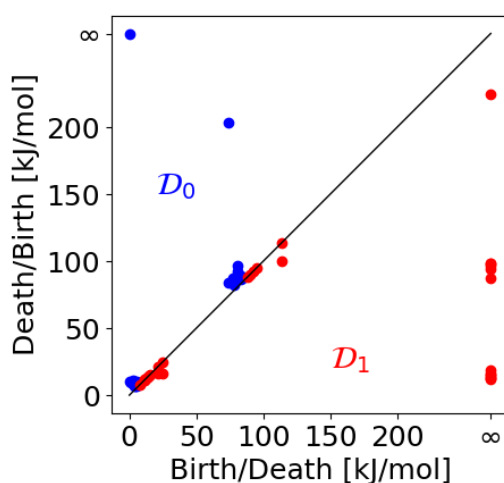
まずアリルビニルエーテル  $C_5H_8O$  のクライゼン転移反応：



に対して適用した。図 3(a) は [3,3]-シグマトロピック転位とコンフォメーション変化のみが含まれた部分的な RRM である。図 3(b) は、この部分 RRM のパーシステンスバーコードである。図中、エネルギーの高い左半分が反応物であるアリルビニルエーテル構造、エネルギーの低い右半分が生成物である  $\gamma, \delta$ -不飽和カルボニル構造に対応している。 $D_0$  では、これらのプロットは対角線の周りに現れている。(b, d) = (80, 200) 付近のプロットは、Claisen 転位の最低エネルギー TS に対応する。また、(b, d) = (200,  $\infty$ ) のプロットは、反応領域と生成領域が 2 つの経路によってつながっていることを示している。このように、本手法を適用したパーシステントバーコードはクライゼン転位の反応経路の特徴をよく反映している。



(a)



(b)

図 3: (a) アリルビニルエーテル  $C_5H_8O$  の部分 RRM と (b) その PD

次に合金ナノクラスターに対する適用を見てみる。 $X_3Y_2$  ( $X, Y = \text{Cu, Ag, or Au}$ ) の RRM について、異なる元素に対応するパーシスタンスバーコードを比較したのが図 4 である。同じ元素の組み合わせ（つまり、 $X_3Y_2$  と  $X_2Y_3$ ）では、持続性バーコードは似たような傾向を示す傾向がある。特に、Au と Cu の組み合わせでは、PH の発生源は 50 kcal/mol までの広いエネルギー範囲で現れるが、他の組み合わせでは、ほとんどが 30 kcal/mol までのエネルギー範囲で現れる。Ag と Cu の組み合わせでは、他の組合せに比べて生成元の数が比較的少なくなっている。逆に、Au と Ag の組み合わせでは、特に 1 次元の消滅が無限大の生成元の数比其他のものよりも多くなっている。このことは、前述の解析が 2 つの金属元素の親和性や金属合金の性質を抽出するために利用できることを示唆している。

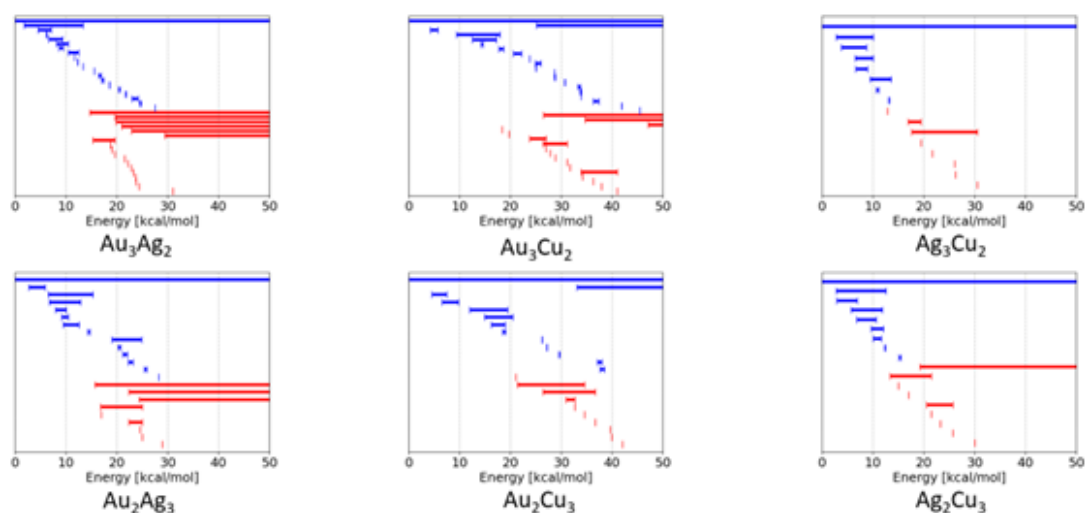


図 4:  $X_3Y_2$  ( $X, Y = \text{Cu, Ag, or Au}$ ) についての PB

以上のように本手法は RRM の特徴を抽出し、金属の親和性など化学的性質を捉えられることを示唆する結果を得た。当日は他の計算結果についても紹介する。

## 参考文献

- [1] David Wales. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press, 2003.
- [2] Paul G. Mezey. *Potential energy hypersurfaces*. Elsevier, Amsterdam, 1987.
- [3] Kenichi Fukui. Formulation of the reaction coordinate. *The Journal of Physical Chemistry*, Vol. 74, No. 23, pp. 4161–4163, 11 1970.
- [4] Hiroko Satoh, Tomohiro Oda, Kumiyo Nakakoji, Takeaki Uno, Satoru Iwata, and Koichi Ohno. RMapDB: chemical reaction route map data for quantum mechanical-based data chemistry, 2020.
- [5] GRRM-GDSP (Institute for Quantum Chemical Exploration).
- [6] Hiroko Satoh, Tomohiro Oda, Kumiyo Nakakoji, Takeaki Uno, Satoru Iwata, and Koichi

- Ohno. "Maizo"-chemistry Project: toward Molecular- and Reaction Discovery from Quantum Mechanical Global Reaction Route Mappings. *Journal of Computer Chemistry, Japan*, Vol. 14, No. 3, pp. 77–79, 2015.
- [7] Hiroko Satoh, Tomohiro Oda, Kumiyo Nakakoji, Takeaki Uno, Hiroaki Tanaka, Satoru Iwata, and Koichi Ohno. Potential Energy Surface-Based Automatic Deduction of Conformational Transition Networks and Its Application on Quantum Mechanical Landscapes of  $\beta$ -Glucose Conformers. *Journal of Chemical Theory and Computation*, Vol. 12, No. 11, pp. 5293–5308, 11 2016.
- [8] Satoshi Maeda, Koichi Ohno, and Keiji Morokuma. 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*, Vol. 15, No. 11, p. 3683, 2013.
- [9] Koichi Ohno and Satoshi Maeda. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chemical Physics Letters*, Vol. 384, No. 4-6, pp. 277–282, 1 2004.
- [10] Satoshi Maeda, Tetsuya Taketsugu, Keiji Morokuma, and Koichi Ohno. Anharmonic downward distortion following for automated exploration of quantum chemical potential energy surfaces. *Bulletin of the Chemical Society of Japan*, Vol. 87, No. 12, 2014.
- [11] Satoshi Maeda and Keiji Morokuma. A systematic method for locating transition structures of  $A+B \rightarrow X$  type reactions. *The Journal of Chemical Physics*, Vol. 132, No. 24, p. 241102, 6 2010.
- [12] Satoshi Maeda, Tetsuya Taketsugu, and Keiji Morokuma. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *Journal of Computational Chemistry*, Vol. 35, No. 2, pp. 166–173, 1 2014.
- [13] Satoshi Maeda, Yu Harabuchi, Makito Takagi, Tetsuya Taketsugu, and Keiji Morokuma. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *Chemical Record*, Vol. 16, No. 5, 2016.
- [14] Takeshi Iwasa, Takaaki Sato, Makito Takagi, Min Gao, Andrey Lyalin, Masato Kobayashi, Ken-ichi Shimizu, Satoshi Maeda, and Tetsuya Taketsugu. Combined Automated Reaction Pathway Searches and Sparse Modeling Analysis for Catalytic Properties of Lowest Energy Twins of Cu $_{13}$ . *The Journal of Physical Chemistry A*, Vol. 123, No. 1, pp. 210–217, 1 2019.
- [15] Hiroki Hayashi, Hitomi Katsuyama, Hideaki Takano, Yu Harabuchi, Satoshi Maeda, and Tsuyoshi Mita. In silico reaction screening with difluorocarbene for N-difluoroalkylative dearomatization of pyridines. *Nature Synthesis*, Vol. 1, pp. 804–814, 8 2022.
- [16] Kanami Sugiyama, Yosuke Sumiya, Makito Takagi, Kenichiro Saita, and Satoshi Maeda. Understanding CO oxidation on the Pt(111) surface based on a reaction route network. *Physical Chemistry Chemical Physics*, Vol. 21, No. 26, pp. 14366–14375, 2019.



- [17] Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G Tollis. Algorithms for drawing graphs: an annotated bibliography. *Computational Geometry*, Vol. 4, No. 5, pp. 235–282, 10 1994.
- [18] I. Herman, G. Melancon, and M.S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 6, No. 1, pp. 24–43, 2000.
- [19] Oren M. Becker and Martin Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*, Vol. 106, No. 4, pp. 1495–1517, 1 1997.
- [20] David J. Wales. The energy landscape as a unifying theme in molecular science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 363, No. 1827, pp. 357–377, 2 2005.
- [21] Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino. Topological Strata of Weighted Complex Networks. *PLoS ONE*, Vol. 8, No. 6, p. e66506, 6 2013.
- [22] Joshua Mirth, Yanqin Zhai, Johnathan Bush, Enrique G. Alvarado, Howie Jordan, Mark Heim, Bala Krishnamoorthy, Markus Pflaum, Aurora Clark, Y Z, and Henry Adams. Representations of energy landscapes by sublevelset persistent homology: An example with n-alkanes. *The Journal of Chemical Physics*, Vol. 154, No. 11, p. 114114, 3 2021.
- [23] T. Okushima, T. Niiyama, K. S. Ikeda, and Y. Shimizu. Connectivity graph: Multiple connectivity on potential energy surface does affect the dynamics. *Physical Review E*, Vol. 76, No. 3, p. 036109, 9 2007.
- [24] Burai Murayama, Masato Kobayashi, Masamitsu Aoki, Suguru Ishibashi, Takuya Saito, Takenobu Nakamura, Hiroshi Teramoto, and Tetsuya Taketsugu. Characterizing Reaction Route Map of Realistic Molecular Reactions based on Weight Rank Clique Filtration of Persistent Homology. *arXiv e-prints*, p. arXiv:2211.15067, November 2022.
- [25] Satoshi Maeda, Yu Harabuchi, Makito Takagi, Kenichiro Saita, Kimichi Suzuki, Tomoya Ichino, Yosuke Sumiya, Kanami Sugiyama, and Yuriko Ono. Implementation and performance of the artificial force induced reaction method in the GRRM17 program. *Journal of Computational Chemistry*, Vol. 39, No. 4, pp. 233–251, 2 2018.
- [26] Satoshi Maeda and Yu Harabuchi. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *WIREs Computational Molecular Science*, Vol. 11, No. 6, 11 2021.
- [27] Shuichi Ebisawa, Takuro Tsutsumi, and Tetsuya Taketsugu. Geometric analysis of anharmonic downward distortion following paths. *Journal of Computational Chemistry*, Vol. 42, No. 1, pp. 27–39, 1 2021.
- [28] Reinhard Diestel. *Graph Theory*, Vol. 173. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017.
- [29] Dmitry Kozlov. *Combinatorial Algebraic Topology*, Vol. 21. Springer Berlin Heidelberg,

Berlin, Heidelberg, 2008.

- [30] Hiraoka Yasuaki. *Protein Structure and Topology: Introduction to Persistent Homology*. Kyoritsu Shuppan, 2013.