Differentiating the yield of chemical reactions using parameters in first-order kinetic equations to identify elementary steps that control the reactivity from complicated reaction path networks

Yu Harabuchi,[a,b] Tomohiko Yokoyama,[c,†] Wataru Matsuoka,[b,d] Taihei Oki,[*,b,c] Satoru Iwata,[*,a,b,c] and Satoshi Maeda[*,a,b,d,e]

[†]equally contributed as the first author

[a]*Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21, Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan.*

[b]*JST, ERATO Maeda Artificial Intelligence in Chemical Reaction Design and Discovery Project, Kita 10, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan.*

[c]*Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan.*

[d]*Department of Chemistry, Faculty of Science, Hokkaido University, Kita 10, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan.*

[e]*Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki 305-0044, Japan.*

*Corresponding author:
E-mail: oki@mist.i.u-tokyo.ac.jp, iwata@mist.i.u-tokyo.ac.jp, smaeda@eis.hokudai.ac.jp

Abstract:

The yield of a chemical reaction is obtained by solving the rate equation. This study introduces an approach for differentiating the yields using the parameters of the rate equation, which is expressed as a first-order linear differential equation. The yield derivative for a specific pair of reactant and product is derived by mathematically expressing the rate constant matrix contraction method, which is a simple kinetic analysis method. The parameters of the rate equation are the Gibbs energies of the intermediates and transition states in the reaction path network used to formulate the rate equation. Thus, the differentiating yield allows the numerical evaluation of the contribution of energy variation to the yield for each intermediate and transition

state in the reaction path network. In other words, a comparison of these values automatically extracts the factors affecting the yield from a complicated reaction path network consisting of numerous reaction paths and intermediates. This study verifies the behavior of the proposed approach through numerical experiments on the reaction path networks of a model system and the Rh-catalyzed hydroformylation reaction. Moreover, the possibility of using this approach for designing organometallic catalyst ligands is discussed.

1. Introduction

Chemical reactions are used in the manufacture of various products, such as medicines, clothing, and devices, and are a vital technology for human society. Much research has been conducted to develop reactions and improve their efficiency. In the research of chemical reactions, one of the critical challenges is to obtain the desired product selectively in high yields. Reaction yield is determined not only by the ease of the target reaction, but also by the competition between the target and side reactions. Therefore, chemists attempt to increase the reaction yield of the target product by using catalysts, modifying substrates, and optimizing reaction conditions. Most *in silico* reaction design focuses on the elementary steps that determine the yield of the desired product, the so-called rate- and selectivity-determining steps, and the modulation of the reaction barriers of the corresponding steps. However, a single reaction may have more than one or many such elementary steps (*vide infra*), and identifying all of them is generally not trivial. The *in silico* reaction design can go wrong if any of these steps are overlooked and not taken into account. It is generally challenging to systematically analyze all the possible side reactions experimentally. In contrast, quantum chemical calculations can be used to systematically search for paths of side reactions, which is expected to be helpful in identifying all rate- and selectivity-determining steps.

Quantum chemical calculations play a critical role in the analysis of chemical reactions.[1–6] Recently, it has become possible to systematically explore equilibrium (EQ) and transition state (TS) structures on potential energy surfaces.[7–10] An elementary step (also called

reaction path in this paper) represents a transition from an EQ to the other EQ via a TS as shown in **Figure 1a**. A graph representing the obtained EQs as nodes and reaction paths as edges is called a reaction path network (**Figure 1b**). The transition state theory[11,12] estimates the rate constants corresponding to each reaction path in the obtained reaction path network. Using these rate constants, kinetic simulations[12-16] are performed starting from the given reactants at a specific temperature and time to predict the reaction yield of each EQ. On-the-fly kinetic simulation during the reaction path search has improved the efficiency of the reaction path search itself[17] and has recently been deployed in quantum chemistry-aided retrosynthetic analysis (QCaRA).[18,19] QCaRA provides possible reactant candidates that give a particular product by following the multistep reaction processes in the inverse way starting from the product. The accumulation of QCaRA reaction path network data has been proposed,[20] and its data platform has been developed to enhance chemical reaction analysis and design.[21]
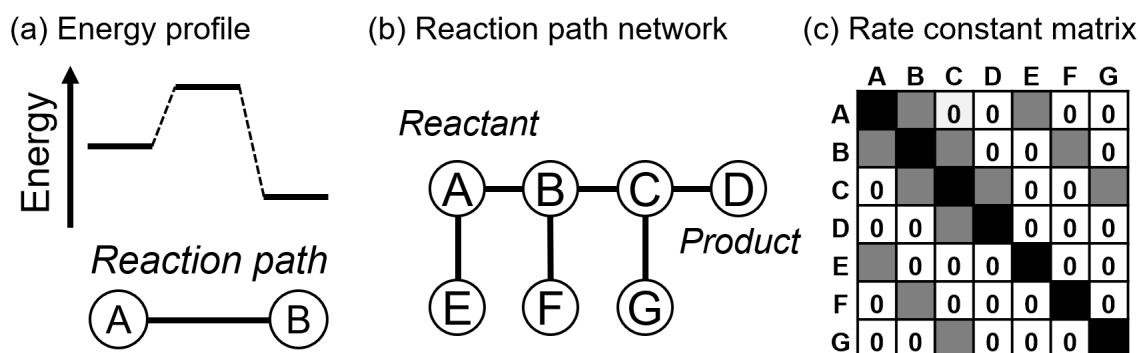
Figure 1. Schematic picture explaining the relationship between the reaction path network and yield derivative. (a) Energy profile along a single reaction path. Nodes represent the EQ, and the edge represents the reaction path via the TS. (b) A reaction path network where reaction paths connect EQs. (c) Rate constant matrix, K, corresponding to (b). Gray elements indicate positive values, and black elements indicate negative values.

In this study, the chemical kinetics is assumed to follow the first-order rate equation:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = Kx(t), \tag{1}$$

where $x(t)$ denotes the n-dimensional vector whose ith component $x_i(t)$ is the reaction yield of

EQ$_i$ at time t, and K is an $n \times n$ rate constant matrix. **Figure 1c** shows a conceptual matrix K. Higher-order rate equations are beyond the scope of this study. The rate constant matrix K has positive values (indicated by gray in **Figure 1c**) in off-diagonal entries corresponding to the edges of the reaction path network shown in **Figure 1a**, and negative values in diagonal entries (indicated by black in **Figure 1c**). TS$_{i-j}$ corresponds to the reaction between EQ$_i$ and EQ$_j$. The reaction path network shown in **Figure 1a** has six reaction paths that provide the six corresponding elements of K (**Figure 1c**). The (i,j) off-diagonal entry K_{ij} of K is a rate constant of a reaction path from EQ$_j$ to EQ$_i$, and the diagonal entry K_{jj} satisfies $K_{jj} = -\sum_{i \neq j} K_{ij}$. In the canonical ensemble, the rate constant K_{ij} is given by

$$K_{ij} = \Gamma \frac{k_B T}{h} \exp\left(-\frac{G_{ij} - G_j}{RT}\right), \tag{2}$$

where G_j and G_{ij} are the Gibbs energies of the EQ$_j$ and TS$_{i-j}$, respectively, k_B is the Boltzmann constant, h is the Planck constant, R is the gas constant, T is temperature, and Γ is the transmission coefficient.

The rate constant matrix contraction (RCMC) method[22,23] achieved an efficient kinetic simulation. In the RCMC method, the size of K is reduced by an operation called contraction, and the yield is obtained through the recursive contraction process (details are provided in the methodology section). Recently, the accuracy of the RCMC method has been discussed,[24] and the mathematical relationship between the Gibbs energy of each EQ and TS in the reaction path network and the reaction yield has been clarified.

This study introduces the concept of differentiating the reaction yield of the target product using the TS and EQ Gibbs energies on a given reaction path network. Our approach differentiating the reaction yield is based on the mathematical formulation of the RCMC method.[24] To understand the information that the yield derivative provides, we conduct numerical experiments on two reaction path networks of a model system and Rh-catalyzed hydroformylation.[25–28] The numerical experiments show that the elementary processes with yield

derivatives of large magnitude correspond to chemically essential ones to control the reaction yields in the network, i.e., rate- and selectivity-determining steps. In other words, the present approach differentiating the reaction yield provides a numerical prediction of the extent to which EQ and TS stabilities affect the yield of a particular product.

2. Methodology

2.1 Mathematics of Rate Constant Matrix Contraction (RCMC) Method

The RCMC method was originally introduced by Sumiya et al.[22] to reduce the number of variables in the kinetics equation (1) for the canonical ensemble and was later applied by Sumiya et al. to compute an approximate solution of (1) for the canonical and microcanonical ensembles. Below, we provide an overview of the method in the mathematically sophisticated formulation given in previous work.[24] Readers are referred to Iwata et al. [24] for more details.

Let t be the reaction time. The RCMC method first determines the set S of EQs, which are regarded as "steady" at time t, by executing the following greedy procedure repeatedly. Let T be the complement of S. Initially, we set S as the empty set and T as the entire set of EQs. We seek $i \in T$ that maximizes the diagonal K_{ii}. If $\frac{1}{|K_{ii}|} \geq t$, then the procedure halts. Otherwise, i is moved from T to S, and the rate constant matrix is updated as $K \leftarrow K_{TT} - \frac{K_{Ti}K_{iT}}{K_{ii}}$. Here, K_{TT} denotes the principal submatrix of K whose rows and columns are indexed by T. Similarly, K_{Ti} and K_{iT} are the ith column (resp. row) vectors of K_{TT}. We then return to determine the maximum diagonal.

Let S be the set of steady EQs found by the greedy algorithm above and T be its complement. By permuting the index of the EQs, we assume that $S = \{1, \ldots, k\}$ and $T = \{k + 1, \ldots, n\}$, where $k = |S|$. Then, the RCMC method computes the reaction yield q at time t by

$$q = Vp, \tag{3}$$

where p is a vector of the initial yields, and V is an $n \times n$ matrix given by

$$V = \begin{pmatrix} K_{SS}^{-1} K_{ST} V_{TT} K_{TS} K_{SS}^{-1} & -K_{SS}^{-1} K_{ST} V_{TT} \\ -V_{TT} K_{TS} K_{SS}^{-1} & V_{TT} \end{pmatrix}. \tag{4}$$

Here, K_{SS}, K_{ST}, and K_{TS} are the submatrices of K whose rows are indexed by S, S, and T, respectively, and whose columns are indexed by S, T, and S, respectively. The matrix V_{TT} is defined by $V_{TT} = \mathrm{diag}(\mathbf{1}^{\top} M)$, where $\mathbf{1}$ is the column vector whose components are all one, $M = I + K_{TS} K_{SS}^{-2} K_{ST}$ (I is the identity matrix), and $\mathrm{diag}(v)$ for a vector v denotes the diagonal matrix obtained by arranging the components of v to diagonals. Equations (3) and (4) give an explicit formulation of how the RCMC method computes the reaction yield q from the rate constant matrix K, provided the set S of steady EQs and the initial yield p. This nature of the RCMC method enables the analytical computation of the derivative of the yield q_i of EQ$_i$ with respect to the rate constants K or potential energies of the EQs and TSs.

2.2 Derivatives of reaction yields based on the RCMC method

This section describes the derivatives of the yield q_i of EQ$_i$ by the TS and EQ energies based on the mathematical understanding of the RCMC method. Using the chain rule, we first derive the derivatives of the yield by the rate constants, and then convert them to that by energies by multiplying the derivative of rate constants by energies to them.

We review a convension of differentiation of a scalar by a matrix. Let y be a scalar variable depending on an $m \times n$ matrix $X = (X_{ij})_{ij}$. Here, we do not assume any structure in X; that is, every entry in X is independent each other. The derivative of y by X is the following $n \times m$ matrix

$$\frac{\mathrm{d}y}{\mathrm{d}X} = \begin{pmatrix} \dfrac{\partial y}{\partial X_{11}} & \dfrac{\partial y}{\partial X_{21}} & \cdots & \dfrac{\partial y}{\partial X_{m1}} \\ \dfrac{\partial y}{\partial X_{12}} & \dfrac{\partial y}{\partial X_{22}} & \cdots & \dfrac{\partial y}{\partial X_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y}{\partial X_{1n}} & \dfrac{\partial y}{\partial X_{2n}} & \cdots & \dfrac{\partial y}{\partial X_{mn}} \end{pmatrix}.$$

Note that the (i, j) entry in $\frac{dy}{dX}$ is the derivative of y by the (j, i) entry in X.

Let us define a matrix $\Omega^* = \begin{pmatrix} -K_{SS}^{-1} K_{ST} \\ I_T \end{pmatrix}$ and a vector $w_i = V_{\{TT\}} \Omega^{*\top} e_i \circ q_T$. Here, for two vectors $a = (a_i)_i$ and $b = (b_i)_i$ of the same dimensions, we let $a \circ b$ denote the component-wise multiplication of a and b, i.e., $a \circ b = (a_i b_i)_i$. The derivative of q_i by the rate constant matrix K is given by

$$
\begin{pmatrix} \frac{\partial q_i}{\partial K_{SS}} & \frac{\partial q_i}{\partial K_{TS}} \\ \frac{\partial q_i}{\partial K_{ST}} & \frac{\partial q_i}{\partial K_{TT}} \end{pmatrix} = \begin{pmatrix} -K_{SS}^{-1} K_{ST} \frac{\partial q_i}{\partial K_{ST}} - \frac{\partial q_i}{\partial K_{TS}} K_{TS} K_{SS}^{-1} & -K_{SS}^{-1} p_S e_i^\top \Omega^* V_{TT} \\ -q_T e_i^\top K_{SS}^{-1} + w_i \mathbf{1}_S^\top K_{SS}^{-1} & 0 \end{pmatrix}
$$

if $i \in S$ and

$$
\begin{pmatrix} \frac{\partial q_i}{\partial K_{SS}} & \frac{\partial q_i}{\partial K_{TS}} \\ \frac{\partial q_i}{\partial K_{ST}} & \frac{\partial q_i}{\partial K_{TT}} \end{pmatrix} = \begin{pmatrix} -K_{SS}^{-1} K_{ST} \frac{\partial q_i}{\partial K_{ST}} - \frac{\partial q_i}{\partial K_{TS}} K_{TS} K_{SS}^{-1} & -K_{SS}^{-1} p_S e_i^\top \Omega^* V_{TT} \\ w_i \mathbf{1}_S^\top K_{SS}^{-1} & 0 \end{pmatrix}
$$

if $i \in T$.

Finally, using the equation (2), the yield derivative by the energy E_j of EQ$_j$ is calculated as

$$
\frac{\partial q_i}{\partial E_j} = \sum_{i',j'=1}^{n} \frac{\partial q_i}{\partial K_{i'j'}} \frac{\partial K_{i'j'}}{\partial E_j} = \sum_{i' \neq j} \frac{K_{i'j}}{RT} \frac{\partial q_i}{\partial K_{i'j}} - \sum_{i' \neq j} \frac{K_{i'j}}{RT} \frac{\partial q_i}{\partial K_{jj}}.
$$

Similarly, the yield derivative by the energy E_{jk} of TS$_{j\text{-}k}$ is given by

$$
\frac{\partial q_i}{\partial E_{jk}} = \sum_{i',j'=1}^{n} \frac{\partial q_i}{\partial K_{i'j'}} \frac{\partial K_{i'j'}}{\partial E_{jk}} = -\frac{K_{jk}}{RT} \frac{\partial q_i}{\partial K_{jk}} - \frac{K_{kj}}{RT} \frac{\partial q_i}{\partial K_{kj}} + \frac{K_{kj}}{RT} \frac{\partial q_i}{\partial K_{jj}} + \frac{K_{jk}}{RT} \frac{\partial q_i}{\partial K_{kk}}.
$$

3. Results and Discussion

3. 1 Application to a model reaction path network

Figure 2 depicts the model network consisting of twelve EQs and sixteen TSs. Their Gibbs energy values relative to the most stable **EQ11** are given near their location in kJ/mol. **Table 1** also presents the EQ and TS energies in kJ/mol. **Table 1a** shows the yields Φ of these

EQs after the chemical reaction at 300 K starting from **EQ0** and terminating at 86400 seconds. As shown in **Table 1a**, the reaction gives the major product P_1 (**EQ7**) with $\Phi = 98.2$ %, the side-product P_2 (**EQ9**) with $\Phi = 1.8$ %, and another very minor side-product P_3 (**EQ11**) with $\Phi = 0.000006$ %. In addition, **Table 1** lists the yield derivatives of the yields of the three products (**EQ7**, **EQ9**, and **EQ11**) when a reaction starts with reactant **EQ0 (R)**. The positive and negative yield derivatives indicate that destabilization and stabilization of the corresponding EQs and TSs, respectively, increase their yields.
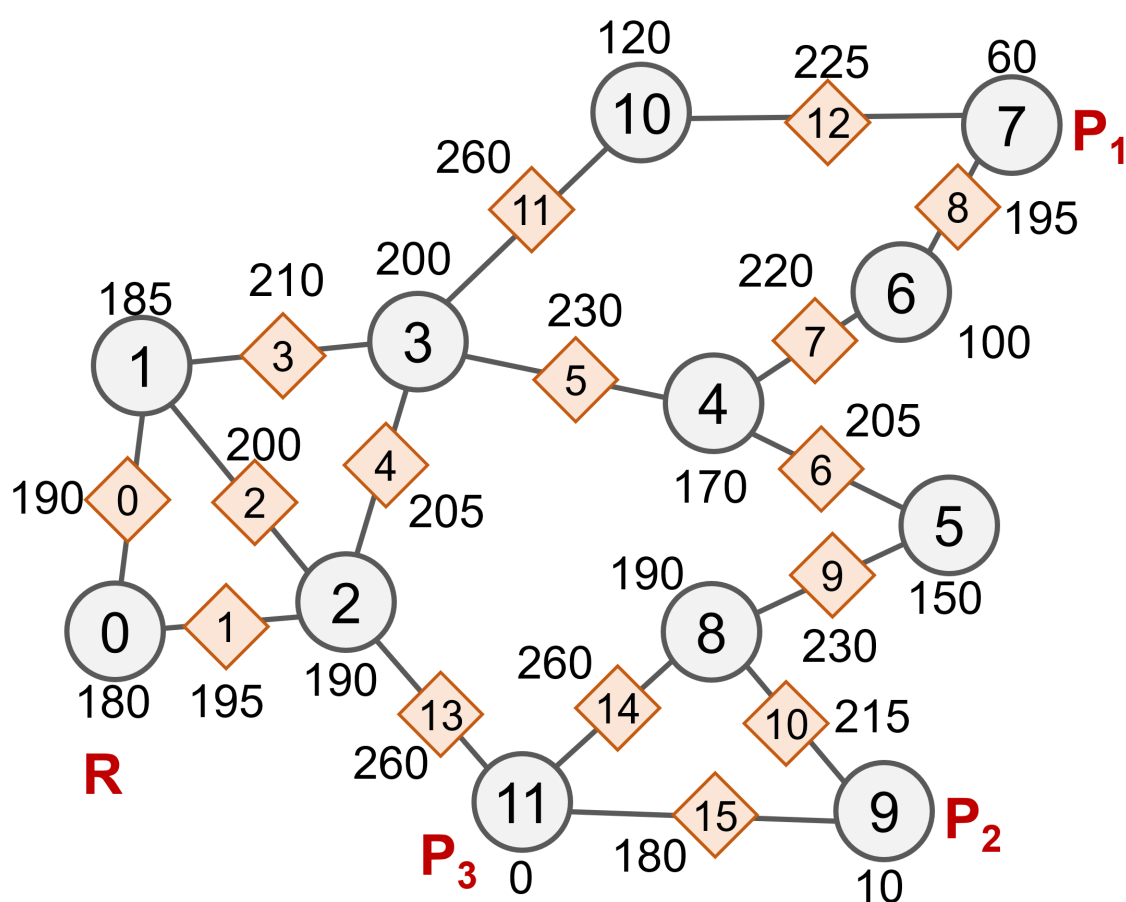
Figure 2. Model reaction path network. Circles and diamond shapes indicate EQs and TSs, respectively. The Gibbs energies for each EQ and TS are shown in kJ/mol relative to the most stable EQ, **EQ 11**. A reaction starting with reactant **EQ0 (R)** and yielding products **EQ7** (P_1), **EQ9** (P_2), and **EQ11** (P_3) with non-negligible yields is discussed.

Table 1: Reaction yields and yield derivatives (a) for EQs and (b) for TSs on the reaction path network shown in **Figure 2**. The reaction yields are computed with 300 K and 86400 s, and the

reactant is set to **EQ0** (**R**). The yield derivatives are for the yields of **EQ7**, **EQ9**, and **EQ11**. The yield derivative is shown in $(kJ/mol)^{-1}$.

(a) EQ

	Gibbs energy (kJ/mol)	Yields	Derivative of Yield of EQ7	Derivative of Yield of EQ9	Derivative of Yield of EQ11
EQ0	180		4.94×10^{-22}	3.19×10^{-34}	6.79×10^{-43}
EQ1	185	1.66×10^{-22}	6.66×10^{-23}	4.30×10^{-35}	9.14×10^{-44}
EQ2	190	2.24×10^{-23}	8.97×10^{-24}	5.79×10^{-36}	1.23×10^{-44}
EQ3	200	4.06×10^{-25}	1.63×10^{-25}	1.05×10^{-37}	2.24×10^{-46}
EQ4	170	6.79×10^{-20}	2.72×10^{-20}	1.76×10^{-32}	6.56×10^{-43}
EQ5	150	2.06×10^{-16}	8.27×10^{-17}	5.35×10^{-29}	1.99×10^{-39}
EQ6	100	1.07×10^{-7}	4.27×10^{-8}	1.20×10^{-24}	4.49×10^{-35}
EQ7	60	9.82×10^{-1}	-4.27×10^{-8}	0.00	0.00
EQ8	190	5.46×10^{-26}	2.19×10^{-26}	3.25×10^{-34}	3.00×10^{-47}
EQ9	10	1.78×10^{-2}	0.00	-1.20×10^{-24}	0.00
EQ10	120	3.51×10^{-11}	1.41×10^{-11}	7.19×10^{-30}	1.53×10^{-38}
EQ11	0	6.09×10^{-06}	0.00	0.00	-4.49×10^{-35}

(b) TS

	Gibbs energy (kJ/mol)	Derivative of Yield of EQ7	Derivative of Yield of EQ9	Derivative of Yield of EQ11
TS0	190	-1.93×10^{-11}	5.91×10^{-13}	5.06×10^{-15}
TS1	195	2.34×10^{-11}	-6.21×10^{-13}	-1.54×10^{-13}
TS2	200	2.51×10^{-13}	1.36×10^{-13}	-2.28×10^{-14}
TS3	210	-1.11×10^{-11}	-1.98×10^{-13}	1.13×10^{-11}
TS4	205	-8.14×10^{-11}	-1.45×10^{-12}	8.25×10^{-11}
TS5	230	-2.31×10^{-6}	-8.53×10^{-8}	2.40×10^{-6}
TS6	205	3.10×10^{-7}	-3.10×10^{-7}	2.91×10^{-14}
TS7	220	-7.00×10^{-3}	7.00×10^{-3}	4.21×10^{-8}
TS8	195	-3.11×10^{-7}	3.11×10^{-7}	1.87×10^{-12}
TS9	230	6.99×10^{-3}	-6.99×10^{-3}	6.56×10^{-10}
TS10	215	1.71×10^{-5}	-1.71×10^{-5}	1.06×10^{-10}
TS11	260	-4.34×10^{-8}	4.34×10^{-8}	1.48×10^{-11}
TS12	225	-3.50×10^{-14}	3.50×10^{-14}	1.20×10^{-17}
TS13	260	2.40×10^{-6}	4.34×10^{-8}	-2.44×10^{-6}
TS14	260	2.50×10^{-13}	1.04×10^{-10}	-1.04×10^{-10}
TS15	180	0.00	0.00	0.00

According to the network in **Figure 2**, the reaction first proceeds from **EQ0** to the intermediate **EQ5**, following the sequence of **EQ0-EQ2-EQ3-EQ4-EQ5**, in both paths yielding **EQ7** (**P$_1$**) and **EQ9** (**P$_2$**). After reaching **EQ5**, the reaction proceeds from **EQ5** to **EQ7** (**P$_1$**) and **EQ9** (**P$_2$**), following the sequence of **EQ5-EQ4-EQ6-EQ7** and **EQ5-EQ8-EQ9**, respectively.

Comparing the yield derivatives for the yields of P_1 and P_2, those of **TS6-TS10** are opposite in sign and are similar in magnitude. This is because the paths yielding P_1 and P_2 are in competition, and increasing or decreasing the yield of one has the opposite effect on the other.

Let us consider the situation of identifying TSs that determine the P_1 / P_2 selectivity based on these yield derivatives. The highest TSs along the paths from **EQ5** to **EQ7** (P_1) and from **EQ5** to **EQ9** (P_2) are **TS7** (230 kJ/mol) and **TS9** (220 kJ/mol), respectively. Thus, the difference in reaction rates of crossing **TS7** and **TS9** towards P_1 and P_2, respectively, significantly affect their yields. For the yields of P_1 and P_2, the yield derivatives are exceptionally large in magnitude for **TS7** and **TS9**, indicating that these two are the selectivity-determining TSs. In other words, the yield derivatives suggest that stabilizing **TS7** or destabilizing **TS9** improves the P_1 selectivity.

On the other hand, the yield derivatives for the yield of P_3 tell that the P_3 / (P_1 + P_2) selectivity is determined by the stabilities of **TS5** and **TS13**. Their values are opposite in sign and similar in magnitude. On the network, P_3 is thermodynamically the most stable. However, P_3 is kinetically inaccessible from P_1 and P_2 due to high barriers. Therefore, P_3 must be accessed from the reactant region including **EQ0**–**EQ3** through **TS13** under the reaction conditions (300 K, 86400 s). In other words, the P_3 selectivity is determined through the competition between the rates crossing **TS5** and **TS13** from the reactant region.

In contrast, most of the other TSs in the network have negligible effects on the yields of the three products. For example, **TS0**–**TS4** separates EQs in the reactant region **EQ0**–**EQ3**. Since these EQs quickly reach equilibrium under the reaction conditions, they show only small yield derivatives in magnitude. **TS11**, **TS12**, **TS14**, and **TS15** are other examples of yield derivatives of small magnitude. These TSs serve as barriers along minor (less feasible) routes to each product and thus have a negligible effect on kinetics. In addition, the yield derivatives of EQs are all small in magnitude, indicating that the reaction is kinetically controlled under the reaction conditions.

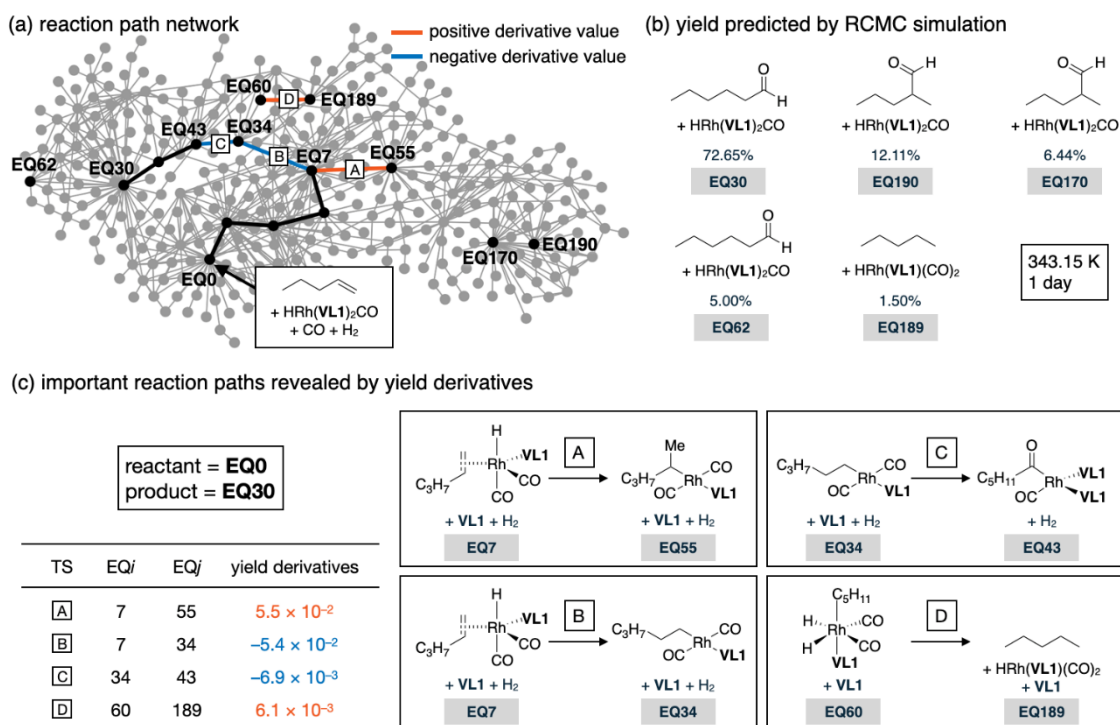## 3.2 Application to a reaction path network of Rh-catalyzed hydroformylation

The proposed approach was applied to the reaction path network of the Rh-catalyzed hydroformylation of 1-pentene. The network was previously reported by Matsuoka et al.[29] The energies were computed at the ωB97X-D/Def2-SVP level, and the solvent effect of the dichloroethane was incorporated by the C-PCM method.[30,31] The reaction path network was constructed using the single component-artificial force induced reaction (SC-AFIR) method[32] with the virtual ligand (VL) method,[29,33] where the electronic and steric parameters of VL were tuned to reproduce those of PPh₃. The details of the reaction path network calculation procedure have been described in the previous paper.[29]

The reaction path network was preprocessed by a structure-based clustering. A group of similar local minima with the same bond connectivity pattern was considered an EQ, assuming that these local minima immediately reach equilibrium before moving on to any other local minima. That is, the rate constants between EQs were determined by the generalized pre-equilibrium approximation (GPA) assuming pre-equilibrium among local minima in each EQ.[34,35] Based on GPA, the Boltzmann distribution of an EQ required for RCMC calculations was obtained by the sum of the Boltzmann distributions of the local minima belonging to the EQ. In the original network, each pair of EQs is connected by multiple paths, and the rate constant for the transition from one EQ to the other was calculated as the sum of the rate constants under the GPA for all paths connecting them. Once the rate constant matrix K was defined for the EQs, the reaction yield of each EQ was estimated using the RCMC method. Consequently, the preprocessed reaction path network includes 357 EQs, as shown in **Figure 4a**.

The kinetic simulations and the approach differentiating the yield were performed under the conditions of 343.15 K (70 °C) and 1 day (86400 s). Under these conditions, hydroformylation simulated using the reaction path network starting from **EQ0** yielded products shown in **Figure 4b**, including linear and branched aldehydes. Within the reaction path network shown in **Figure 4a**, **EQ30** and **EQ190** had the lowest energies among the EQs corresponding to the linear and branched products, respectively. The paths and EQs along the lowest energy path from the initial

structure (**EQ0**) to the linear product (**EQ30**) are highlighted with bold lines in the network. The linear aldehydes (**EQ30** and **EQ62**) and branched aldehydes (**EQ190** and **EQ170**) were predicted to be formed in about 78% and 19% of the total, respectively.

Here, by differentiating yields, this study attempted to identify critical reaction paths that control the yield of linear products. Therefore, yield derivative values were calculated using **EQ0** and **EQ30** as the reactant and product, respectively. The results are summarized in **Figure 4c**. This figure shows the four TSs with the highest absolute derivative values among all TSs. The reaction paths with the highest and second highest absolute derivative values (**TS$_A$** and **TS$_B$**) connected **EQ7** to **EQ55** and **EQ34**, respectively. These paths correspond to the insertion of 1-pentene into the Rh-hydride complex, where **TS$_A$** leads to the branched product, and **TS$_B$** gives the linear product. **EQ7** is the branching point of these routes, and the fact that **TS$_A$** and **TS$_B$** have the greatest influence on the yield of the linear product is reasonable, considering that the branched aldehyde is the second major product in this reaction (**Figure 4b**). The reaction paths with the third and fourth highest absolute derivative values (**TS$_C$** and **TS$_D$**) connect **EQ34** to **EQ43** and **EQ60** to **EQ189**, respectively. **TS$_C$** is the transition state for CO insertion from the rhodium alkyl complex (**EQ34**), and **TS$_D$** corresponds to the reductive elimination of pentane from the rhodium dihydride species (**EQ60**). Notably, **TS$_D$**, which is isolated from the minimum energy route, has a greater influence on product yield than most reaction paths that are included or connected to the minimum energy route. This result is not apparent, and significant effort would be required to identify this elementary step by manually analyzing the complicated reaction path network. Therefore, this result highlights the utility of the yield derivative method as it enables rapid and automatic analysis of the reaction path network.

Figure 4. (a) Reaction path network for Rh-catalyzed hydroformylation of 1-pentene. The nodes represent EQs, and the edges represent reaction paths. (b) The reaction yields are calculated by the RCMC method. The conditions for simulation were set to be 343.15 K and 1 day. The initial population was assigned to **EQ0**. (c) Critical reaction paths that affect the yield of linear product (**EQ30**). The four yield derivatives with the largest absolute values are listed in the table, and the corresponding chemical transformations are shown in the boxes.

As shown above, the present approach automatically and systematically extracts kinetically important EQs and TSs from a reaction path network. Such EQs and TSs, involved in rate- and selectivity-determining processes, have been used to understand and predict reactivity and selectivity for a variety of chemical reactions, including organo and organometallic catalysis.[1–6] In addition, the automation of their conformation screening has been a long-standing topic in the field of computational catalyst modeling.[36–40] On the other hand, *in silico* reaction design and discovery based on complex reaction path networks is one of the emerging topics. [41–

[43] However, there has been no systematic approach to finding all rate- and selectivity-determining EQs and TSs from a given reaction path network. The above examples show that the present approach works well for this purpose. This approach would be particularly useful in combination with the VLA screening approach,[33,29] which describes the ligands of organometallic catalyst by a few parameters and performs the ligand feature screening for a set of given TSs in a parameter space. Previously, we demonstrated automated ligand feature screening using the VLA screening for the above Rh-catalyzed hydroformylation of 1-pentene, focusing on TS_A and TS_B identified as selectivity-determining TSs based on human intuition.[33] Reaction path network calculation is also fully automated by our combined AFIR and RCMC methods. In addition, this study demonstrated full automation of the process of finding the rate- and selectivity-determining EQs and TSs from a given reaction path network. Therefore, the next challenge would be to fully automate a catalyst design starting from reaction path network exploration for a chemical transformation of interest, through automated extraction of rate- and selectivity-determining EQs and TSs, to the final VLA screening to identify the optimal ligand features.

Conclusions

This study proposed an approach to obtain numerical insight into the control of reaction yields in complicated chemical reactions by differentiating the yields of chemical reactions using EQ and TS Gibbs energies within reaction path networks. A mathematical understanding of the RCMC method[22] enabled us to differentiate the yield in a complicated reaction path network. Differentiating the yields mathematically provided the extent of the contribution of each EQ or TS Gibbs energy to the reaction yields. A numerical experiment using a model reaction path network clarified the relationship between the derivative value and its effect on the reaction yield. Furthermore, application to a reaction path network of Rh-catalyzed hydroformylation revealed that the present approach automatically extracts the paths that influence the selectivity of the reaction from a complicated network.

The proposed approach will be particularly useful for systems lacking prior knowledge of the rate- and selectivity-determining processes. In other words, its applications to unexplored reactivities would be an interesting future direction. In particular, an application of the integrated AFIR-RCMC reaction path network exploration, the present approach differentiating the yields, and VLA screening approaches to organometallic catalyst discovery will be a significant future challenge.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

References:

(1) Schlegel, H. B. Exploring Potential Energy Surfaces for Chemical Reactions: An Overview of Some Practical Methods. *J. Comput. Chem.* **2003**, *24*, 1514–1527. https://doi.org/10.1002/jcc.10231.

(2) Houk, K. N.; Cheong, P. H.-Y. Computational Prediction of Small-Molecule Catalysts. *Nature* **2008**, *455*, 309–313. https://doi.org/10.1038/nature07368.

(3) Thiel, W. Computational Catalysis-Past, Present, and Future. *Angew. Chem. Int. Ed.* **2014**, *53*, 8605–8613. https://doi.org/10.1002/anie.201402118.

(4) Sameera, W. M. C.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **2016**, *49*, 763–773. https://doi.org/10.1021/acs.accounts.6b00023.

(5) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539–543. https://doi.org/10.1021/acs.accounts.6b00532.

(6) Ahn, S.; Hong, M.; Sundararajan, M.; Ess, D. H.; Baik, M.-H. Design and Optimization of Catalysts Based on Mechanistic Insights Derived from Quantum Chemical Reaction Modeling. *Chem. Rev.* **2019**, *119*, 6509–6560. https://doi.org/10.1021/acs.chemrev.9b00073.

(7) Maeda, S.; Ohno, K.; Morokuma, K. Systematic Exploration of the Mechanism of Chemical

Reactions: The Global Reaction Route Mapping (GRRM) Strategy Using the ADDF and AFIR Methods. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683–3701. https://doi.org/10.1039/c3cp44063j.

(8)   Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems: Exploring Reaction Space in Molecular Systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1354. https://doi.org/10.1002/wcms.1354.

(9)   Simm, G. N.; Vaucher, A. C.; Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *J. Phys. Chem. A* **2019**, *123*, 385–399. https://doi.org/10.1021/acs.jpca.8b10007.

(10) Maeda, S.; Harabuchi, Y. Exploring Paths of Chemical Transformations in Molecular and Periodic Systems: An Approach Utilizing Force. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1538. https://doi.org/10.1002/wcms.1538.

(11) Fernández-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. Modeling the Kinetics of Bimolecular Reactions. *Chem. Rev.* **2006**, *106*, 4518–4584. https://doi.org/10.1021/cr050205w.

(12) Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J. Am. Chem. Soc.* **2014**, *136*, 528–546. https://doi.org/10.1021/ja408723a.

(13) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C,Chapter 11, 2nd Ed.; Cambridge Univ. Press: Cambridge, 1992; Pp. 456–495.*

(14) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical Kinetics and Dynamics, 2nd Ed.; Prentice Hall: Upper Saddle River, NJ, 1999.*

(15) Kozuch, S.; Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **2011**, *44*, 101–110. https://doi.org/10.1021/ar1000956.

(16) Rush, L. E.; Pringle, P. G.; Harvey, J. N. Computational Kinetics of Cobalt-Catalyzed Alkene Hydroformylation. *Angew. Chem. Int. Ed.* **2014**, *53*, 8672–8676. https://doi.org/10.1002/anie.201402115.

(17) Sumiya, Y.; Maeda, S. A Reaction Path Network for Wöhler's Urea Synthesis. *Chem. Lett.* **2019**, *48*, 47–50. https://doi.org/10.1246/cl.180850.

(18) Sumiya, Y.; Harabuchi, Y.; Nagata, Y.; Maeda, S. Quantum Chemical Calculations to Trace Back Reaction Paths for the Prediction of Reactants. *JACS Au* **2022**, *2*, 1181–1188. https://doi.org/10.1021/jacsau.2c00157.

(19) Mita, T.; Takano, H.; Hayashi, H.; Kanna, W.; Harabuchi, Y.; Houk, K. N.; Maeda, S. Prediction of High-Yielding Single-Step or Cascade Pericyclic Reactions for the Synthesis of Complex Synthetic Targets. *J. Am. Chem. Soc.* **2022**, *144*, 22985–23000. https://doi.org/10.1021/jacs.2c09830.

(20) Harabuchi, Y.; Maeda, S. Theoretical Chemical Reaction Database Construction Based on Quantum Chemistry-Aided Retrosynthetic Analysis. *ChemRxiv* **2022**.

https://doi.org/10.26434/chemrxiv-2022-tl4vj.

(21) Kuwahara, M.; Harabuchi, Y.; Maeda, S.; Fujima, J.; Takahashi, K. Searching Chemical Action and Network (SCAN): An Interactive Chemical Reaction Path Network Platform. *Digit. Discov.* **2023**, *2*, 1104–1111. https://doi.org/10.1039/D3DD00026E.

(22) Sumiya, Y.; Nagahata, Y.; Komatsuzaki, T.; Taketsugu, T.; Maeda, S. Kinetic Analysis for the Multistep Profiles of Organic Reactions: Significance of the Conformational Entropy on the Rate Constants of the Claisen Rearrangement. *J. Phys. Chem. A* **2015**, *119*, 11641–11649. https://doi.org/10.1021/acs.jpca.5b09447.

(23) Sumiya, Y.; Maeda, S. Rate Constant Matrix Contraction Method for Systematic Analysis of Reaction Path Networks. *Chem. Lett.* **2020**, *49*, 553–564. https://doi.org/10.1246/cl.200092.

(24) Iwata, S.; Oki, T.; Sakaue, S. Rate Constant Matrix Contraction Method for Stiff Master Equations with Detailed Balance. *arXiv* **2023**, 2312.05470 [math.NA]. https://doi.org/10.48550/arXiv.2312.05470.

(25) Beller, M.; Cornils, B.; Frohning, C. D.; Kohlpaintner, C. W. Progress in Hydroformylation and Carbonylation. *J. Mol. Catal. Chem.* **1995**, *104*, 17–85. https://doi.org/10.1016/1381-1169(95)00130-1.

(26) Breit, B.; Seiche, W. Recent Advances on Chemo-, Regio- and Stereoselective Hydroformylation. *Synthesis* **2001**, *2001*, 0001–0036. https://doi.org/10.1055/s-2001-9739.

(27) Franke, R.; Selent, D.; Börner, A. Applied Hydroformylation. *Chem. Rev.* **2012**, *112*, 5675–5732. https://doi.org/10.1021/cr3001803.

(28) Chakrabortty, S.; Almasalma, A. A.; De Vries, J. G. Recent Developments in Asymmetric Hydroformylation. *Catal. Sci. Technol.* **2021**, *11*, 5388–5411. https://doi.org/10.1039/D1CY00737H.

(29) Matsuoka, W.; Harabuchi, Y.; Maeda, S. Virtual Ligand Strategy in Transition Metal Catalysis Toward Highly Efficient Elucidation of Reaction Mechanisms and Computational Catalyst Design. *ACS Catal.* **2023**, *13*, 5697–5711. https://doi.org/10.1021/acscatal.3c00576.

(30) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001. https://doi.org/10.1021/jp9716997.

(31) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669–681. https://doi.org/10.1002/jcc.10189.

(32) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring Transition State Structures for Intramolecular Pathways by the Artificial Force Induced Reaction Method. *J. Comput. Chem.* **2014**, *35*, 166–173. https://doi.org/10.1002/jcc.23481.

(33) Matsuoka, W.; Harabuchi, Y.; Maeda, S. Virtual Ligand-Assisted Screening Strategy to Discover

Enabling Ligands for Transition Metal Catalysis. *ACS Catal.* **2022**, *12*, 3752–3766. https://doi.org/10.1021/acscatal.2c00267.

(34) Shaffer, J. S.; Chakraborty, A. K. Dynamics of Poly(Methyl Methacrylate) Chains Adsorbed on Aluminum Surfaces. *Macromolecules* **1993**, *26*, 1120–1136. https://doi.org/10.1021/ma00057a036.

(35) Rae, M.; Berberan-Santos, M. N. A Generalized Pre-Equilibrium Approximation in Chemical and Photophysical Kinetics. *J. Chem. Educ.* **2004**, *81*, 436. https://doi.org/10.1021/ed081p436.

(36) Maeda, S.; Ohno, K. Lowest Transition State for the Chirality-Determining Step in Ru((*R*)-BINAP)-Catalyzed Asymmetric Hydrogenation of Methyl-3-Oxobutanoate. *J. Am. Chem. Soc.* **2008**, *130*, 17228–17229. https://doi.org/10.1021/ja8065732.

(37) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Prediction of Enantioselectivity in Rhodium Catalyzed Hydrogenations. *J. Am. Chem. Soc.* **2009**, *131*, 410–411. https://doi.org/10.1021/ja806246h.

(38) Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O. Prediction of Stereochemistry Using Q2MM. *Acc. Chem. Res.* **2016**, *49*, 996–1005. https://doi.org/10.1021/acs.accounts.6b00037.

(39) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, *14*, 5249–5261. https://doi.org/10.1021/acs.jctc.8b00578.

(40) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O. Rapid Virtual Screening of Enantioselective Catalysts Using CatVS. *Nat. Catal.* **2018**, *2*, 41–45. https://doi.org/10.1038/s41929-018-0193-3.

(41) Mita, T.; Harabuchi, Y.; Maeda, S. Discovery of a Synthesis Method for a Difluoroglycine Derivative Based on a Path Generated by Quantum Chemical Calculations. *Chem. Sci.* **2020**, *11*, 7569–7577. https://doi.org/10.1039/D0SC02089C.

(42) Hayashi, H.; Katsuyama, H.; Takano, H.; Harabuchi, Y.; Maeda, S.; Mita, T. In Silico Reaction Screening with Difluorocarbene for N-Difluoroalkylative Dearomatization of Pyridines. *Nat. Synth.* **2022**. https://doi.org/10.1038/s44160-022-00128-y.

(43) Hayashi, H.; Maeda, S.; Mita, T. Quantum Chemical Calculations for Reaction Prediction in the Development of Synthetic Methodologies. *Chem. Sci.* **2023**, *14*, 11601–11616. https://doi.org/10.1039/D3SC03319H.