Kinetic Analysis for Complex Networks of Elementary Steps in Chemical Reactions

(化学反応素過程の複雑ネットワークに対する速度解析)

Yosuke Sumiya

Graduate School of Chemical Sciences and Engineering

Hokkaido University

Contents

Chapter 1 General Introduction

Advances in quantum chemical calculation have enabled accurate and efficient theoretical elucidations of chemical reaction mechanisms and kinetics[1]. Quantum chemical calculation has been used by many researchers.

Chemical reaction mechanisms can be studied by exploring equilibrium (EQ) and transition state (TS) structure on the potential energy surface (PES). These structures have been investigated by geometry optimization. The most calculated reaction path is the intrinsic reaction coordinate (IRC)[2,3], the mass-weighted steepest decent path starting from TS. Geometry optimization requires a good guess about the EQ or TS. When there are multiple reaction paths, all of them should be examined and the most kinetically preferable pathway should be determined. However, making initial guess for all reaction paths is very difficult. Theoretical elucidation of unknown reaction mechanisms thus is difficult.

In order to overcome this problem, automated reaction path search method is indispensable. Recently, global reaction route mapping (GRRM) program[4-14] and other reaction path search methods[15-27] have been developed. The GRRM/artificial force induced reaction (AFIR) method has been mainly applied to organic reactions[28-41] and various reactions[42-48] since 2010.

Even if a reaction system is simple, there are a great number of structures and reaction paths. For example, there are 23 EQs and 66 TSs in Claisen rearrangement of allyl vinyl ether, which is one of the simplest organic reactions. EQs and reaction paths via TS form complex network[49,50]. This network includes conformers of reactant and product. Moreover, the network including byproducts generated via high energy TSs becomes very large[11]. In such networks, conformational rearrangement proceeds at

about 10^{-13} second and bond rearrangement at about 10^{3} second. Then, it is difficult to achieve a time evolution covering experimental timescale (10^{2}–10^{4} s). This problem has been known as stiff problem[51] in numerical analysis of differential equations.

The author developed the method for kinetic analysis of complex network composed of many elementary steps in chemical reactions. This method is called rate constant matric contraction (RCMC) and contracts fast processes and allows for estimating overall rate constants between the obtained groups.

On the other hand, when the reaction systems are complicated or the number of atoms in the system increases, the reaction path network becomes very large[52-54]. Then, obtaining the network becomes difficult even if automated reaction path search is used. A new algorithm for construction of the network is thus required. In the context of this background, the author also proposed a method for reducing the computational cost of construction of reaction path network using RCMC.

This thesis consists of eight chapters including this chapter. Chapter 2 to 4 are described about kinetic analysis, and Chapter 5 to 7 are described about automated reaction path search. The outline of each chapter is as follows.

In Chapter 2, the author describes the background of kinetic analysis. Chemical reactions are represented as a reaction path network composed of many elementary steps. The network has complex structure where many processes with a wide range of different rate constants coexist. Various methods for kinetic analysis have been proposed to analyze such reaction path network. In this chapter, the author introduce the transition state theory (TST) and methods for kinetic analysis which is commonly used.

In Chapter 3, the author proposes a new method for calculating overall rate constant from a given reaction path network, which is called rate constant matrix

contraction (RCMC). Recently, advances in quantum chemical calculation and algorithm have enabled systematical search of reaction paths on PES. Nowadays, a method for analyzing reaction path networks is required. The RCMC contracts a rate constant matrix recursively and allows to estimate overall rate constant from a reactant to a product. In this chapter, the author applies the RCMC to two kinds of Claisen rearrangement.

In Chapter 4, the author proposes *f*-RCMC as the extension of the RCMC. Unimolecular decomposition plays an important role in gas phase reactions, and a method for calculating the branching ratio is required. In principle, branching ratio can be obtained by solving all rate equations. However, there is a problem in terms of numerical analysis. When the elementary steps have a wide range of different rate constants, it is difficult to achieve a time evolution covering all of the timescales buried in an entire reaction path network due to the numerical stability and the computational cost. This problem is called stiff problem. The *f*-RCMC can be an alternative approach for overcoming this problem. This method allows to obtain the branching ratio without performing kinetic simulation. In this chapter, this method is applied to two kinds of unimolecular decompositions, C_3H_5 decomposition and C_4H_5 decompositions.

In Chapter 5, the author introduces methods for reaction path search. Quantum chemical calculation has been a powerful tool for chemical research. However, there is a problem in geometry optimization. Namely, a good guess is required for finding reaction paths. It is thus difficult to make theoretical elucidation of unknown reaction mechanisms. A systematical approach, which can search important structures automatically, needs to be developed. In this chapter, the author introduces the history of the development of automated reaction path search.

In Chapter 6, the author proposes a way to construct reaction path networks based on AFIR and RCMC and reduce the computational cost. Total number of reaction paths increases exponentially according to the increase of the number of atoms in the system. Global search of reaction paths thus is difficult even in systems with a few atoms. Computational cost has to be reduced for further applications. The algorithm introduced in this chapter can extract paths behaving as reaction bottlenecks, and TS optimization and IRC calculation are applied to these paths. The other paths are calculated by the locally updated plane method. In this chapter, this approach is applied to the reaction of H_2O on Cu(111) surface.

In Chapter 7, the author proposes an algorithm based on AFIR and RCMC. This algorithm allows to reduce computational time for reaction path search by limiting search area. Since it takes computational time to search all reaction paths around an EQ, the overall computational time would be reduced by improving efficiency of this search. The algorithm introduced in this chapter allows to construct a reaction path network efficiently by automated reaction path search and kinetic analysis. Then, many elementary steps existing irrespective of the actual reaction are not searched. In this chapter, the author applies this algorithm to Passerini reaction.

In Chapter 8, the author presents general conclusion.

References

[1] F. Jensen, Introduction to Computational Chemistry, 2nd Edition, Wiley, Chichester, **2007**.

[2] K. Fukui, *Acc. Chem. Res.* **1981**, *14*, 363.

[3] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int. J. Quant. Chem.* **2015**, *115*, 258.

[4] K. Ohno, S. Maeda, *Chem. Phys. Lett.* **2004**, *384*, 277.

[5] S. Maeda, K. Ohno, *J. Phys. Chem. A* **2005**, *109*, 5742.

[6] K. Ohno, S. Maeda, *J. Phys. Chem. A* **2006**, *110*, 8933.

[7] K. Ohno, S. Maeda, *Phys. Scr.* **2008**, *78*, 058122.

[8] S. Maeda, K. Morokuma, *J. Chem. Phys.* **2010**, *132*, 241102.

[9] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2011**, *7*, 2335.

[10] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683.

[11] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, *35*, 166.

[12] S. Maeda, T. Taketsugu, K. Morokuma, K. Ohno, *Bull. Chem. Soc. Jpn*. **2014**, *87*, 1315.

[13] W. M. C. Sameera, S. Maeda, K. Morokuma, *Acc. Chem. Res.* **2016**, *49*, 763.

[14] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec.* **2016**, *16*, 2232.

[15] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, T. R. Walsh, *Adv. Chem. Phys.* **2000**, *115*, 1.

[16] M. Dallos, H. Lischka, E. V. D. Monte, M. Hirsch, W. Quapp, *J. Comput. Chem.* **2002**, *23*, 576.

[17] J. Baker, K. Wolinski, *J. Comput. Chem.* **2011**, *32*, 43.

[18] P. M. Zimmerman, *J. Comput. Chem.* **2013**, *34*, 1385.

[19] D. Rappoport, C. J. Galvin, D. Yu. Zubarev, Alán Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, *10*, 897.

[20] B. Schaefer, S. Mohr, M. Amsler, S. Goedecker, *J. Chem. Phys.* **2014**, *140*, 214102.

[21] D. J. Wales, *J. Chem. Phys.* **2015**, *142*, 130901.

[22] S. Habershon, *J. Chem. Phys.* **2015**, *143*, 094106.

[23] E. Martínez-Núñez, *J. Comput. Chem.* **2015**, *36*, 222.

[24] M. Bergeler, G. N. Simm, J. Proppe, M. Reiher, *J. Chem. Theory Comput.* **2015**, *11*, 5712.

[25] X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757.

[26] L.-P. Wang, R. T. McGibbon, V. S. Pande, T. J. Martinez, *J. Chem. Theory Comput.* **2016**, *12*, 638.

[27] M. Yang, J. Zou, G. Wang, S. Li, *J. Phys. Chem. A* **2017**, *121*, 1351.

[28] S. Maeda, S. Komagawa, M. Uchiyama, K. Morokuma, *Angew. Chem. Int. Ed.* **2011**, *50*, 644.

[29] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2012**, *8*, 380.

[30] M. Hatanaka, S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2013**, *9*, 2882.

[31] M. Hatanaka, K. Morokuma, *J. Am. Chem. Soc.* **2013**, *135*, 13972.

[32] R. Uematsu, S. Maeda, T. Taketsugu, *Chem. Asian J.* **2014**, *9*, 305.

[33] G. Zeng, S. Maeda, T. Taketsugu, S. Sakaki, *Angew. Chem. Int. Ed.* **2014**, *126*, 4721.

[34] G. P. Petrova, H.-B. Li, K. Maruoka, K. Morokuma, *J. Phys. Chem. B* **2014**, *118*, 5154.

[35] R. Uematsu, E. Yamamoto, S. Maeda, H. Ito, T. Taketsugu, *J. Am. Chem. Soc.* **2015**, *137*, 4090.

[36] S. A. Moteki, H. Maruyama, K. Nakayama, H.-B. Li, G. Petrova, S. Maeda, K. Morokuma, K. Maruoka, *Chem. Asian J.* **2015**, *10*, 2112.

[37] M. Hatanaka, K. Morokuma, *ACS Catal.* **2015**, *5*, 3731.

[38] R. Ramozzi, K. Morokuma, *J. Org. Chem.* **2015**, *80*, 5652.

[39] M. Puripat, R. Ramozzi, M. Hatanaka, W. Parasuk, V. Parasuk, K. Morokuma, *J. Org. Chem.* **2015**, *80*, 6959.

[40] W. M. C. Sameera, M. Hatanaka, T. Kitanosono, S. Kobayashi, K. Morokuma, *J. Am. Chem. Soc.* **2015**, *137*, 11085.

[41] R. Uematsu, C. Saka, Y. Sumiya, T. Ichino, T. Taketsugu, S. Maeda, *Chem. Comm*. **2017**, *53*, 7302.

[42] S. Maeda, R. Saito, K. Morokuma, *J. Phys. Chem. Lett.* **2011**, *2*, 852.

[43] Y. Matsuda, K. Hoki, S. Maeda, K.-i. Hanaue, K. Ohta, K. Morokuma, N. Mikami, A. Fujii, *Phys. Chem. Chem. Phys.* **2012**, *14*, 712.

[44] P. Farahani, S. Maeda, J. S. Francisco, M. Lundberg, *ChemPhysChem* **2015**, *16*, 181.

[45] S. Maeda, E. Abe, M. Hatanaka, T. Taketsugu, K. Morokuma, *J. Chem. Theory Comput.* **2012**, *8*, 5058.

[46] M. Isegawa, S. Maeda, D. J. Tantillo, K. Morokuma, *Chem. Sci.* **2014**, *5*, 1555.

[47] M. Gao, A. Lyalin, S. Maeda, T. Taketsugu, *J. Chem. Theory Comput.* **2014**, *10*, 1623.

[48] M. Gao, A. Lyalin, M. Takagi, S. Maeda, T. Taketsugu, *J. Phys. Chem. C* **2015**, *119*, 11120.

[49] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T. Taketsugu, S. Maeda, *J. Phys. Chem. A* **2015**, *119*, 11641.

[50] Y. Sumiya, T. Taketsugu, S. Maeda, *J. Comput. Chem*. **2017**, *38*, 101.

[51] R. J. Gelinas, *J. Comput. Phys.* **1972**, *9*, 222.

[52] J. P. K. Doye, D. J. Wales, *J. Chem. Phys*. **2002**, *116*, 3777.

[53] H. Tokoyama, H. Yamakado, S. Maeda, K. Ohno, *Bull. Chem. Soc. Jpn*., **2015**, *88*, 1284.

[54] H. Tokoyama, H. Yamakado, S. Maeda, K. Ohno, *Chem. Lett*., **2014**, *43*, 702.

Chapter 2 Background of Kinetic Analysis

2.1 Introduction

Chemical substances are generated by bond rearrangement in chemical reactions. Processes of conformational change such as internal rotation of methyl group could be also included in a broad sense. Chemical kinetics deals with reaction rate of these processes and is one of the foundations of basic chemistry. Modern chemical kinetics started with a paper submitted by Arrhenius in 1889[1]. In this paper, Arrhenius revealed that rate constant k is given by the following equation.

$$k = A\exp\left(\frac{-E_a}{RT}\right) \qquad (2.1)$$

where T is temperature; A is pre-exponential factor; E_a is activation energy; R is gas constant. Through research on eq. (2.1), H. Eyring[2], M. Evans, and M. Polanyi[3] developed transition state theory (TST) in 1935 and the theory became the basis of reaction kinetics.

Chemical kinetics plays an important role in various fields such as organic chemistry, interface chemistry, cosmochemistry, and biochemistry. For example, it connects these fields and gives a common academic basis to help understanding at the molecular level. Chemical reactions in various fields proceed on highly different timescales. Slow reactions take more than a few hours. It is well known that organic reactions often take a few days; while if one interests the details of reactions based on individual motions of atoms and/or molecules, the timescale becomes extremely short. For example, molecular vibration reaction is about 10^{-14} s.

In recent years, a method to investigate landscape on potential energy surface (PES) calculated by quantum chemical calculation has been developed and has allowed

to construct reaction path network based on transition between energy minima[4]. The results obtained by this method showed that the reaction path network contains hundreds of molecular structures (energy minima) and elementary steps even in a small system with only about five atoms. The timescales of the elementary steps are often highly different, where the timescale of bond rearrangement is 10^4 s whereas local bond rotation is 10^{-12} s. Namely, different timescales can be buried in a network. A systematic method for analyzing such a network is desired.

In this chapter, the author introduces the equations used in TST, and reviews the methods for kinetic analysis of reaction path network. Concretely, the author reviews (1) rate determining elementary step (RDES) model; (2) lowest reactant to transition state (LRTS) model; (3) quasi steady state approximation (QSSA); (4) generalized pre-equilibrium approximation (GPEA); (5) TS sampling. (1) The RDES model only takes the rate determining elementary steps into account; (2) in LRTS model, it is assumed that the most stable reactant and transition states of the rate determining elementary steps are directly connected; (3) in QSSA, lifetime of intermediates is set to zero; (4) in GPEA, the thermal equilibrium in the reactant region is assumed; (5) TS sampling is convenient and can estimate selectivity with high accuracy.

2.2 Rate Constant based on Transition state theory

In transition state theory (TST), the rate constant can be estimated from information on the molecular structure of the first-order saddle point on the potential energy surface (PES). A minimum point on PES corresponds to an equilibrium structure (EQ) and the first-order saddle point corresponds to the transition state (TS). The mass-weighted steepest decent path leading from TS to EQ is called the intrinsic reaction coordinate

(IRC) path[5]. When EQ i and j are connected by IRC path through single TS, the rate constant $k_{i\to j}$ of the elementary step is estimated by the following equation[6],

$$k_{i\to j} = \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{TS} - \Delta G_i}{RT}\right) \qquad (2.2)$$

where ΔG_{TS} and ΔG_i are Gibbs free energies at the TS and EQ i, respectively, k_B is the Boltzmann constant, h is the Planck constant, R is gas constant, and T is temperature. The transmission coefficient Γ is usually set to unity or the following value[7].

$$\Gamma = 1 + \frac{1}{24}\left(\frac{h\nu^{\ddagger}}{k_B T}\right)^2 \qquad (2.3)$$

In eq. (2.3), $\nu^{\ddagger}$ is magnitude of imaginary frequency at the TS. These equations are used when the system is described by the canonical ensemble. In case the system follows the microcanonical ensemble, the TST rate constant is as follows[6].

$$k_{i\to j} = \frac{1}{h}\frac{G_{TS}}{N_i} \qquad (2.4)$$

In eq. 2.4, G_{TS} and N_i are the sum of states at $TS_{i\text{-}j}$ and the density of states at EQ i, respectively. G_{TS} and N_i were evaluated by Whitten-Rabinovitch equations [8].

When there are two or more elementary steps directly connecting EQs i and j, $k_{i\to j}$ represents the sum of the rate constants for these elementary steps. The rate constant for an elementary step in which the TS has the symmetry different from the reactant was scaled by the factor ρ_X / ρ_{TS}, where ρ_X and ρ_{TS} are the numbers of permutational isomers of the EQ X and TS, respectively[9]. When there is no elementary step directly connecting EQs i and j, $k_{i\to j}$ is zero. Rate constants between identical EQs, $k_{i\to i}$, are also zero.

With the above rate constant, first-order rate equations are represented as the product of the rate constant matrix and the population vector,

$$\frac{d}{dt}\begin{bmatrix}[A]\\[B]\\\vdots\end{bmatrix} = \begin{bmatrix}-\sum_{X\neq A} k_{A\to X} & k_{B\to A} & \cdots\\ k_{A\to B} & -\sum_{X\neq B} k_{B\to X} & \cdots\\ \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}[A]\\[B]\\\vdots\end{bmatrix} \qquad (2.5)$$

where [i] denotes the population of EQ i.

2.3 Rate Determining Elementary Step Model

The concept of a rate determining step is central to the kinetics community[12]. The most typical definition is the slowest step, and is also regarded as the bottleneck. Here, consider Fig. 2.1 to illustrate the rate determining step. The EQ 1 and 2 are reactants; the EQ 3 is a product. The EQ 1 and 2 are connected by low energy barrier, and the EQ 2 and 3 are connected by high energy barrier. Then, the rate determining step is the elementary step from EQ 2 to EQ 3.

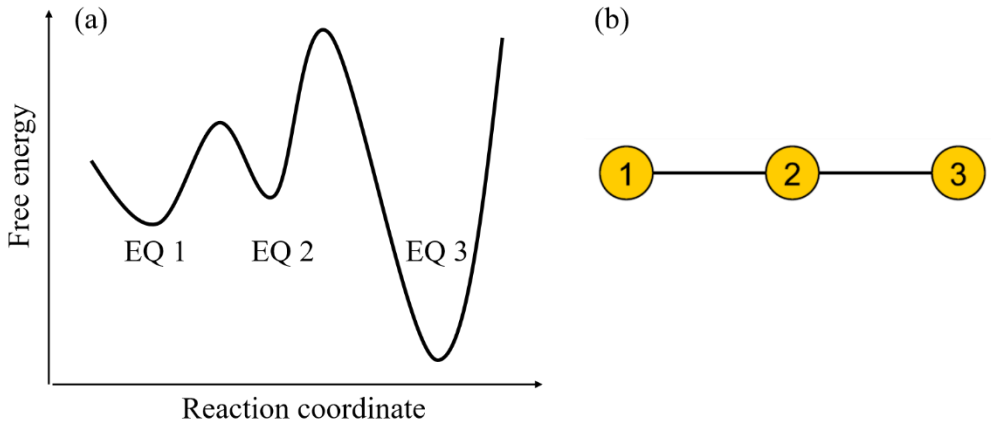

Fig. 2.1 (a) Model reaction profile and (b) the reaction path network of elementary steps. Nodes and edges correspond to EQs and reaction paths via TS. The EQ 1 and EQ 2 are reactants, and EQ 3 is product.

In kinetic analysis of reaction mechanisms, reactivity is often discussed from only important elementary steps such as rate determining steps. In the case of Fig. 2.1, the value of the rate constant $k_{2\to3}$ of the rate determining step is particularly noticed. The

value of $k_{2\to3}$ is calculated by substituting the free energy of $TS_{2\text{-}3}$ and EQ 2 into eq. (2.2).

$$k_{2\to3} = \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{2-3} - \Delta G_2}{RT}\right) \quad (2.6)$$

This rate constant does not include the effect of the most stable EQ in reactant region. This thus does not correspond to the overall rate constant from the reactant region to the product region. This estimation of rate constant is called the rate determining elementary step (RDES) model.

The reaction profile can also be represented by the graph as shown in Fig. 2.1(b), where node and edge are EQ and reaction pathway via TS, respectively. This graph equivalent to the reaction profile is called a reaction path network composed of elementary steps. This representation is convenient for visualization of complex reaction profiles and it helps to grasp the entire reaction.

2.4 Lowest Reactant to Transition State Model

The lowest reactant to transition state (LRTS) model is often used to estimate rate constants of chemical reactions. In this model, if the most stable EQ (the most stable conformer of the reactant or the most stable intermediate) in the reactant region is known, the overall rate constant can be estimated by substituting the free energy of the most stable EQ *X* and the TS at the rate determining step into eq. (2.2).

$$k_{X\to P_n} = \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{i-j} - \Delta G_X}{RT}\right) \quad (2.7)$$

Here, the rate determining step connects with EQ *i* in the reactant region and EQ *j* in the product P_n region. When there are two or more elementary steps directly connecting EQs *i* and *j*, $k_{X\to P_n}$ represents the sum of each rate constant. In the case of Fig. 2.1, the overall rate constant is as follow:

$$k_{1\to3} = \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{2-3} - \Delta G_1}{RT}\right) \qquad (2.8)$$

This estimation of the rate constant assumes that the most stable EQ and the TS of the rate determining step are directly connected and it is often used as better approximation than RDES model. It should be noted that this rate constant does not include the effect of EQ 2.

2.5 Quasi Steady State Approximation

Quasi steady state approximation (QSSA) is called the Bodenstein principle[13]. At first step of QSSA, species that can be regarded as quasi steady state is selected. Such a species is called quasi-steady-state (QSS) species. In QSSA, intermediates with high reactivity and low population are chosen as QSS species. In Fig. 2.1, EQ 2 is the QSS-species connecting reactant EQ 1 and product EQ 3. When QSSA is applied to EQ 2, following equation is assumed:

$$\frac{d[2]}{dt} = 0 \qquad (2.9)$$

where [2] is the concentration of EQ 2. Then, the following equation holds from the rate equation for EQ 2:

$$k_{1\to2}[1] + k_{3\to2}[3] - k_{2\to1}[2] - k_{2\to3}[2] = 0 \qquad (2.10)$$

Thus,

$$[2] = \frac{k_{1\to2}[1] + k_{3\to2}[3]}{k_{2\to1} + k_{2\to3}} \qquad (2.11)$$

By substituting the above equation into the rate equation for EQ 3,

$$\frac{d[3]}{dt} = -\frac{k_{3\to2}k_{2\to1}}{k_{2\to1} + k_{2\to3}}[3] + \frac{k_{1\to2}k_{2\to3}}{k_{2\to1} + k_{2\to3}}[1] = -k_{3\to1}^{QSSA}[3] + k_{1\to3}^{QSSA}[1] \qquad (2.12)$$

Here,

$$k_{1\to3}^{QSSA} = \frac{k_{1\to2}k_{2\to3}}{k_{2\to1}+k_{2\to3}} \qquad (2.13)$$

$k_{3\to1}^{QSSA}$ is calculated in the same way. Therefore, the reaction in Fig. 1 is replaced by the below reaction:

$$\text{EQ 1} \rightarrow \text{EQ 3} \quad (2.14)$$

where $k_{1\to3}^{QSSA}$ can be regarded as the overall rate constant of this reaction.

The QSSA has been used to reduce the stiffness of rate equations, and is also written in physical chemistry textbook[14]. However, the error induced by the QSSA should be noted. Many researches on the error induced by the QSSA have been conducted[15,16]. It should also be noted that the overall rate constant obtained by the QSSA does not include the entropic effect of EQ 2.

2.6 Generalized Pre-Equilibrium Approximation

The generalized pre-equilibrium approximation assumes that the reactant region is equilibrated[17,18]. In this approximation, the Boltzmann distribution of the EQs that are the initial state of rate determining elementary steps is calculated. Next, by multiplying it by the rate constant of the rate determining elementary step, one can obtain the overall rate constant from the reactant region *I* to the product region *J*. This rate constant is represented by the following equation:

$$k_{I\to J}^{GPEA} = \frac{exp\left(-\frac{\Delta G_i}{RT}\right)}{\sum_{p\in I} exp\left(-\frac{\Delta G_p}{RT}\right)} \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{TS_{i-j}} - \Delta G_i}{RT}\right) \qquad (2.15)$$

where the initial state of the rate determining elementary step in reactant region *I* is denoted by EQ *i*, and the sum of denominator of the Boltzmann distribution is taken for the set ***I*** composed of all EQs in reactant region.

In the case of Fig. 2.1, Boltzmann distributions of EQ 1 and 2 are first calculated.

Next, one can obtain the overall rate constant by multiplying the obtained Boltzmann distribution of EQ 2 by the rate constant from EQ 2 to EQ 3. That is, the overall rate constant is represented as below.

$$k^{GPEA}_{\text{reactant}\rightarrow\text{product}} = \frac{\exp\left(-\frac{\Delta G_2}{RT}\right)}{\sum_{p\in\text{reactant}}\exp\left(-\frac{\Delta G_p}{RT}\right)} \Gamma \frac{k_B T}{h} \exp\left(-\frac{\Delta G_{TS_{2-3}} - \Delta G_2}{RT}\right) \qquad (2.16)$$

Based on the GPEA, it is possible to estimate the overall rate constant taking the entropic effects of all EQs in reactant region into account, and the rate constant is more accurate than the model introduced in Section 2.3-2.5. Here, it should be noted that GPEA is effective only when it is possible to assume thermal equilibrium in reactant region. For this reason, more general models are required.

2.7 Transition State Sampling

Improving selectivity is an important issue in designing chemical reactions. Selectivity can be obtained by comparing the rate constants from the reactant to each product which are estimated by the methods introduced in Section 2.3-2.6. On the other hand, it is also possible to evaluate quantitative selectivity more conveniently. At first, it is necessary to explore all TSs leading to each product. Any method may be used for TS sampling. For example, anharmonic downward distortion following (ADDF) and artificial force induced reaction (AFIR) are effective[19]. Then, selectivity can be estimated by calculating the Boltzmann distribution for all of the obtained TSs and taking the sum of the corresponding products. Although this approach seems simple, this gives the same accuracy as comparison of the rate constants based on GPEA by assuming no back reaction from product regions. The reason is as follows.

First, using eq. (2.15), the overall rate constant from a reactant region *I* to a

product region J_n is given by the following equation:

$$k_{I\to J_n}^{GPEA} = \sum_{i\in I}\sum_{j\in J_n}\left[\frac{\exp\left(-\frac{\Delta G_i}{RT}\right)}{\sum_{p\in I}\exp\left(-\frac{\Delta G_p}{RT}\right)}\Gamma\frac{k_B T}{h}\exp\left(-\frac{\Delta G_{TS_{i-j}}-\Delta G_i}{RT}\right)\right] \quad (2.17)$$

where $\Delta G_{TS_{i-j}} = \infty$ for EQ pairs not directly connected by a TS. Assuming that there is no return from the product J_n, the population of the product at time t is:

$$[J_n](t) = k_{I\to J_n}^{GPEA}\int_0^\infty [I](t)\,dt \quad (2.18)$$

Then, the ratio of $[J_n](t)$ to the population of all products is given by the following equation.

$$\frac{[J_n]}{\sum_x [J_x]} = \frac{k_{I\to J_n}^{GPEA}}{\sum_x k_{I\to J_x}^{GPEA}} = \frac{\sum_{i\in I}\sum_{j\in J_n}\exp\left(-\Delta G_{TS_{i-j}}/RT\right)}{\sum_x\sum_{i\in I}\sum_{j\in J_x}\exp\left(-\Delta G_{TS_{i-j}}/RT\right)} \quad (2.19)$$

The above equation is the Boltzmann distribution of all of the TSs at the boundary between I and J_n. That is, assuming no reverse reaction from product regions and GPEA, the selectivity of the reaction can be conveniently estimated by the Boltzmann distribution of TSs.

With this method, quantitative selectivity was reported for the stereoselectivity of water-based Mukaiyama aldol reactions[20]. In this example, 165 TSs with different conformation were obtained for the C-C bond formation step, and the branching ratio of the syn product and the anti product was estimated by the Boltzmann distribution of these TSs. As a result, the obtained diastereomeric ratio (syn/anti) was 75:25, where the experimental ratio is 73:27[21]. This result indicates that TS sampling is effective for estimating quantitative selectivity. However, it should be noted that this approach assumes thermal equilibrium in reactant region. In addition, it is also assumed that there is no back reaction from products.

2.8 Conclusion

In this chapter, the author introduced the basic formula used in the modern chemical kinetics. In recent years, development of algorithms and computers has allowed to construct reaction path networks based on potential energy surface/landscape. These networks involve elementary steps that proceed over a wide range of different timescales. In Section 2.3-2.6, the approaches for kinetic analysis dealing with such complicated reaction networks were introduced. In Section 2.7, the author introduced TS sampling which allows to estimate selectivity. The above approaches assume various approximations. We therefore should be careful about its scope. More general approaches for kinetic analysis are also required.

References

[1] S. Arrhenius, *Z. Phys. Chem*. **1889**, *4*, 226.

[2] H. Eyring, *J. Chem. Phys*., **1935**, *3*, 107.

[3] M. G. Evans, M. Polanyi, *Trans. Faraday Soc*. **1935**, *31*, 875.

[4] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec*. **2016**, *16*, 2232.

[5] K. Fukui, *J. Phys. Chem*. **1970**, *74*, 4161.

[6] S. J. Klippenstein, V. S. Pande, D. G. Truhlar, *J. Am. Chem. Soc.* **2014**, *136*, 528-546.

[7] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, D. G. Truhlar, *Chem. Rev.* **2006**, *106*, 4518.

[8] G. Z. Whitten, B. S. Rabinovitch, *J. Chem. Phys.* **1963**, *38*, 2466.

[9] A. J. Karas, R. G. Gilbert, M. A. Collins, *Chem. Phys. Lett.* **1992**, *193*, 181.

[10] J. I. Steinfeld, J. S. Francisco, W. L. Hase, Chemical Kinetics and Dynamics, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, **1999**.

[11] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Chap. 11, 2nd ed., Cambridge Univ. Press, Cambridge **1992**; pp. 456-495.

[12] S. Kozuch, J. M. L. Martin, *ChemPhysChem* **2011**, *12*, 1413.

[13] T. Turanyi, A. S. Tomlin: Analysis of Kinetic Reaction Mechanism; Springer, Berlin, **2014**; pp. 30-32.

[14] P. Atkins, J. de Paula: Atkins' Physical Chemistry 8th ed.; OUP Oxford, 2014; pp. 812-814.

[15] T. Turanyi, A. S. Tomlin, M. J. Piling, *J. Phys. Chem*. **1993**, *97*, 163.

[16] T. Turanyi, A. S. Tomlin: Analysis of Kinetic Reaction Mechanism; Springer, Berlin, **2014**; pp. 231-242.

[17] M. Rae, M. N. Berberan-Santos, *J. Chem. Educ.* **2004**, *81*, 436.

[18] J. S. Shaffer, A. K. Chakraborty, *Macromolecules* **1993**, *26*, 1120.

[19] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683.

[20] M. Hatanaka, S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2013**, *9*, 2882.

[21] S. Kobayashi, *Synlett.* **1994**, 689.

Chapter 3 New method for kinetic analysis: Rate Constant Matrix Contraction

3.1 Introduction

In recent years, theoretical studies on the mechanism of organic reactions have been conducted extensively[1-3]. Many of these previous studies have discussed reactivity and selectivity in terms of the energy barriers of important reaction steps, i.e., rate- and selectively-determining elementary steps. Furthermore, the rate constants for these important steps have generally been calculated based on the transition state theory (TST)[4,5]. However, this approach is sometimes considered to be simplistic, since reactivity and selectivity can be affected by the whole reaction profile. Therefore, quantitative discussion would require *overall* rate constants that take into account the entire reaction network.

In principle, an overall rate constant can be evaluated by calculating rate constants for all elementary steps and solving the resulting rate equations in a numerical manner for all local states. However, it can be difficult to achieve a time evolution covering all of the timescales buried in an entire reaction network when the elementary steps have a wide range of different rate constants. This situation occurs in the reaction network of most organic reactions. For example, the rate constants of local bond rotations are around 10^{12} s^{-1}, whereas those of bond reorganization steps are around 10^{-4} s^{-1}. Hence, approaches that simplify the rate equations by adopting the quasi-steady state approximation (QSSA) and/or the partial equilibrium assumption, have been employed in recent studies on the kinetics of organic reactions[6,7].

Similar problems to this have also been studied in other fields, including

combustion chemistry, where the lumping approach[8,9] and its extensions[10-12] have been used to solve these problems. The lumping approach is capable of the reduction of a complicated reaction network by eliminating unimportant reaction channels and states that rapidly decay to give more stable states. Approaches involving the clustering of rapidly interconverting states have also been developed and used to evaluate global conformational rearrangements in peptides, as well as phase transition processes in clusters[13,14]. The overall rate constants between the resulting superbasins can then be estimated using other methods such as the generalized pre-equilibrium approximation (GPEA)[15,16]. The kinetic Monte Carlo simulation is another powerful approach that provides efficient access to the overall rate constant of a chemical reaction[17,18]. This approach effectively samples states that would otherwise require a long time simulation to be reached by a Monte Carlo approach.

In this study, the author has proposed a recursive method for calculating the overall rate constants. This method is called “Rate Constant Matrix Contraction (RCMC)”[34]. The aim of this study is to develop a simple and robust numerical algorithm, and investigate its application to the reaction networks of organic reactions. The new method, RCMC, only requires a reaction network and a single parameter, namely k_{MAX}, which represents the upper limit of rate constants in the resulting rate equations. A reaction network should be prepared prior to the analysis, while the k_{MAX} value can be determined easily by adjusting the value based on results of the recursive contraction, as demonstrated below. In this study, reaction networks were obtained by the single-component artificial force induced reaction (SC-AFIR) method[19]. Approaches to construct reaction networks are not the main focus in this chapter and the such reviews are provided in Chapter 5.

3.2 Algorithm of Rate Constant Matrix Contraction in Canonical Ensemble

In the following, the set of all states obtained at the *n*-th loop is denoted by $N^{(n)}$, states *i*, *j*, *k*, *l*, and *m* are by the components of $N^{(n)}$ ($i, j, k, l, m \in N^{(n)}$), the rate constant for the elementary step from the state *i* to the state *j* is by $k_{i\to j}^{(n)}$, and the population of the state *i* at $t = \infty$, that is, the Boltzmann distribution of the state *i*, is denoted by $P_i^{(n)}$. It follows that $N^{(0)}$, $k_{i\to j}^{(0)}$, and $P_i^{(0)}$ correspond to the original set of all states, rate constants, and populations at $t = \infty$, respectively.

The algorithm of RCMC consists of the following twelve steps.

1. Input a reaction profile and the parameter k_{MAX}, and initialize *n* ($n = 0$).
2. Compute $k_{i\to j}^{(0)}$ for all elementary steps on the reaction profile by eq. (2.2), and obtain the rate equations. When there are two or more elementary steps that connect states *i* and *j* directly, $k_{i\to j}^{(0)}$ is the sum of rate constants for these elementary steps. When there is no elementary step that connects states *i* and *j* directly, $k_{i\to j}^{(0)}$ is zero. The rate constant between the identical states, $k_{i\to i}^{(0)}$, is also zero.
3. Compute population at $t = \infty$, $P_i^{(0)}$, for all states based on the Boltzmann distribution.
4. Identify a pair of states *i* and *j* which has the maximum rate constant $k_{i\to j}^{(n)}$.
5. If $k_{i\to j}^{(n)} < k_{\mathrm{MAX}}$, then exit from the recursive contraction loop.
6. Apply the QSSA to the state *i*. In this step, all rate constants $k_{k\to l}^{(n)}$ are updated to $k_{k\to l}'^{(n+1)}$ by the following equations (3.1) and (3.2).

$$k_{k\to l}'^{(n+1)} = k_{k\to l}^{(n)} + k_{k\to i}^{(n)} k_{i\to l}^{(n)} \sigma_i^{(n)} \qquad (3.1)$$

$$\sigma_i^{(n)} = \frac{1}{\sum_{m\in N^{(n)}} k_{i\to m}^{(n)}} \qquad (3.2)$$

7. Rate constants that involve the state *i* are updated to zero ($k_{i\to m}'^{(n+1)} = k_{m\to i}'^{(n+1)} = 0$).
8. Set the population at $t = \infty$ of state *i* to zero, $P_i^{(n+1)} = 0$; in the step 7, the reaction flux entering the state *i* was set to zero, and $P_i^{(n+1)}$ should thus be zero.
9. Update $P_k^{(n+1)}$. The $P_i^{(n)}$, which was updated to zero in the step 8, is distributed to

the neighboring states. In contrast, states were simply grouped or deleted in the previous approaches[12-14]. Instead, in this algorithm, by referring the reaction flux per unit time from the state i to the state k at $t = \infty$, i.e., $k_{i\to k}^{(n)} P_i^{(n)}$, a part of $P_i^{(n)}$ defined by the reaction flux per unit time multiplied by the normalization constant $\sigma_i^{(n)}$ is added to $P_k^{(n)}$. Thus, $P_k^{(n+1)}$ is obtained by the following equation:

$$P_k^{(n+1)} = P_k^{(n)} + k_{i\to k}^{(n)} \sigma_i^{(n)} P_i^{(n)} \qquad (3.3)$$

10. In general, the solution of the rate equations using $k_{k\to l}^{\prime(n+1)}$ at $t = \infty$ does not reproduce $P_k^{(n+1)}$. Thus, $k_{k\to l}^{\prime(n+1)}$ is further updated to $k_{k\to l}^{(n+1)}$ by the following eq. (3.4). Derivation of eq. (3.4) is shown in Section 3.4. The solution of the rate equations at $t = \infty$ using $k_{k\to l}^{(n+1)}$ reproduces $P_k^{(n+1)}$[34].

$$k_{k\to l}^{(n+1)} = \frac{1}{1+\sigma_i^{(n)} k_{k\to i}^{(n)}} k_{k\to l}^{\prime(n+1)} \qquad (3.4)$$

11. Define $N^{(n+1)}$ as the set of all states except for the contracted state i.
12. Increase n by 1 ($n \to n + 1$), and return to the step 4.

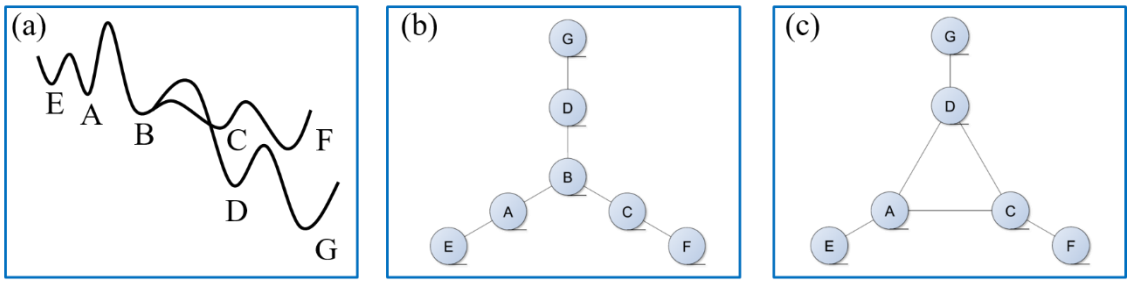

Fig. 3.1 (a) A model reaction profile. (b) The reaction path network. (c) The contracted reaction path network.

Here, the above procedure is illustrated with a model reaction profile shown in Fig. 3.1(a). The model reaction profile is represented as the reaction path network in Fig. 3.1(b). In Fig. 3.1, the elementary step from B to C has the maximum rate constant. Therefore, in the step 6, the QSSA is applied to the state B. For example, $k_{\mathrm{A}\to\mathrm{C}}^{(0)}$, which is zero, is updated to $k_{\mathrm{A}\to\mathrm{C}}^{\prime(1)}$, by the following equations.

$$k'^{(1)}_{A\to C} = k^{(0)}_{A\to C} + k^{(0)}_{A\to B} k^{(0)}_{B\to C} \sigma^{(0)}_{B} \qquad (3.5)$$

$$\sigma^{(0)}_{B} = \frac{1}{k^{(0)}_{B\to A} + k^{(0)}_{B\to C} + k^{(0)}_{B\to D}} \qquad (3.6)$$

Similarly, all the rate constants concerning states that are directly connected to the state B, $k'^{(1)}_{A\to C}$, $k'^{(1)}_{C\to A}$, $k'^{(1)}_{A\to D}$, $k'^{(1)}_{D\to A}$, $k'^{(1)}_{C\to D}$ and $k'^{(1)}_{D\to C}$, are updated. The other rate constants are just copied, because the second term in eq. (3.1) is zero. In the step 7, the rate constants related to the state B, $k'^{(1)}_{A\to B}$, $k'^{(1)}_{B\to A}$, $k'^{(1)}_{B\to C}$, $k'^{(1)}_{C\to B}$, $k'^{(1)}_{B\to D}$ and $k'^{(1)}_{D\to B}$, are set to zero. In the step 8, $P^{(1)}_{B}$ is also set to zero. Then, $P^{(0)}_{B}$ is distributed to the neighboring states A, C, and D. For example, $P^{(1)}_{A}$ is obtained as follows:

$$P^{(1)}_{A} = P^{(0)}_{A} + k^{(0)}_{B\to A} \sigma^{(0)}_{B} P^{(0)}_{B} \qquad (3.7)$$

This equation indicates that a part of $P^{(0)}_{B}$ is added to $P^{(0)}_{A}$ with the weight of the rate constant $k^{(0)}_{B\to A}$ multiplied by the normalization constant $\sigma^{(0)}_{B}$. Similarly, $P^{(1)}_{C}$ and $P^{(1)}_{D}$ are updated. Populations at $t = \infty$ of the states E, F, and G do not change because the second term of eq. (3.3) is zero. In the step 10, rate constants are further updated so that the solution of the rate equations at $t = \infty$ reproduces the populations at $t = \infty$ obtained in the step 9. For example, $k^{(1)}_{A\to E}$ is obtained by the following equation.

$$k^{(1)}_{A\to E} = \frac{1}{1 + \sigma^{(0)}_{B} k^{(0)}_{A\to B}} k'^{(1)}_{A\to E} \qquad (3.8)$$

Here, let us explain this correction briefly. In the step 9, the population at $t = \infty$ (Boltzmann distribution) of the state A increases. This indicates that the relative free energy of the state A decreases. As a result, the free energy difference between the state A and the TS increases and the corresponding rate constant decreases. In eq. (3.8), the factor multiplied to $k'^{(1)}_{A\to E}$ represents the decrease of the rate constant. In the step 10, the other rate constants, $k^{(1)}_{A\to C}$, $k^{(1)}_{A\to D}$, $k^{(1)}_{C\to A}$, $k^{(1)}_{C\to D}$, $k^{(1)}_{C\to F}$, $k^{(1)}_{D\to A}$, $k^{(1)}_{D\to C}$, and $k^{(1)}_{D\to G}$, are also updated. Finally, a new state set $N^{(1)}$ is obtained by removing the contracted state B.

3.3 Application1: Claisen rearrangement of CH_2-CHO-CH_2-CH-CH_2

The Claisen rearrangement of allyl vinyl ether CH_2-CHO-CH_2-CH-CH_2 is one of the

simplest of all of the known organic rearrangement reactions, and the mechanism of this reaction has been fully elucidated[24]. Based on theoretical calculations, it has been shown that the C-C bond formation and C-O bond cleavage steps in this rearrangement occur in a concerted manner with a single TS[25]. The results of a gas-phase kinetic study showed that the rate constant of the Claisen rearrangement of allyl vinyl ether follows the Arrhenius equation[26]. In this study, the author used a reaction network consisting of 10 and 13 conformers for the reactant region and the product region, respectively. The reaction network is shown in Fig. 3.2. Here, the terms “reactant” and “product” shall be used to mean sets of 10 and 13 conformers in the corresponding regions, respectively.

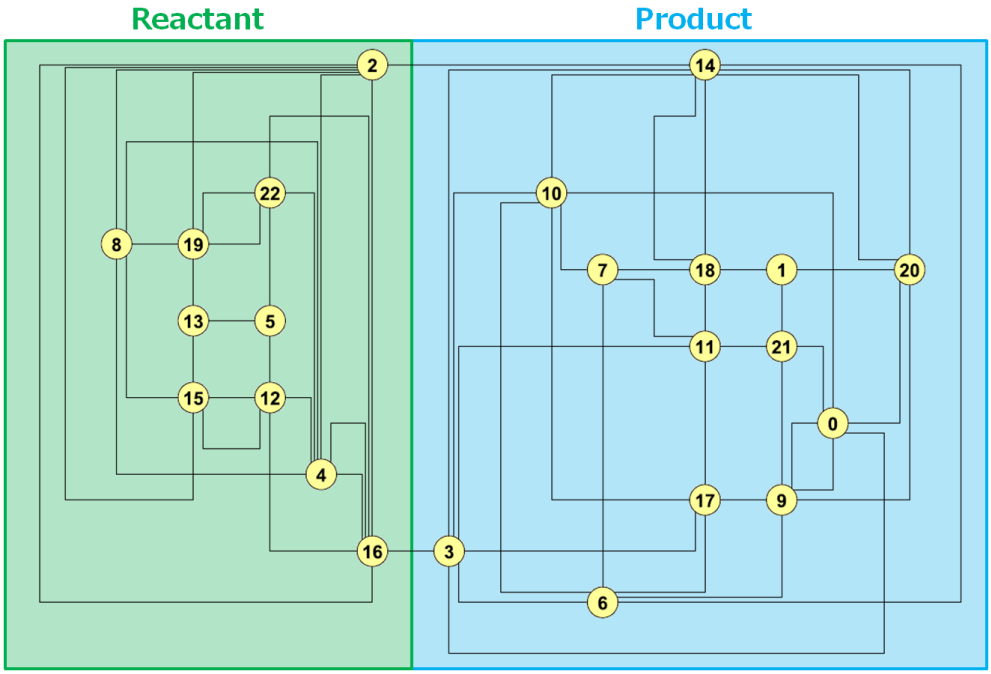

Fig. 3.2 The reaction path network of Claisen rearrangement of allyl vinyl ether. Circle (node) and line (edge) correspond to equilibrium structure (EQ) and reaction pathway via transition state (TS). Left side and right side are the reactant and the product conformers, respectively.

The reaction path network was obtained using density functional theory at the B3LYP/6-31G level. All of the local minima (EQs) and transition states (TSs) were subsequently reoptimized at the M062X/6-311+G(2d,p) level[27]. Finally, I obtained 23 EQs and 66 TSs at the M062X level. The IRC paths were computed at the M062X level starting from all 66 of the different TSs. Among the 66 IRC paths generated in this way, two corresponded to pathways directly connecting the reactant and the product. These two IRC paths also corresponded to pathways that proceed through chair- and boat-type TSs[25]. The other 64 IRC paths corresponded to pathways associated with local conformational changes. To achieve high levels of chemical accuracy, single point CCSD(T)-F12a/jul-cc-pVTZ calculations[28,29] were performed for all of the EQs and TSs obtained in the current study. Finally, the Gibbs free energies were estimated at the experimental temperature (469.1 K) using the electronic energy values at the CCSD(T) level, a rigid rotor approximation based on structures at the M062X level, harmonic vibrational frequencies at the M062X level and the ideal gas approximation. The rate constants for all elementary steps were estimated using eq. (2.2).

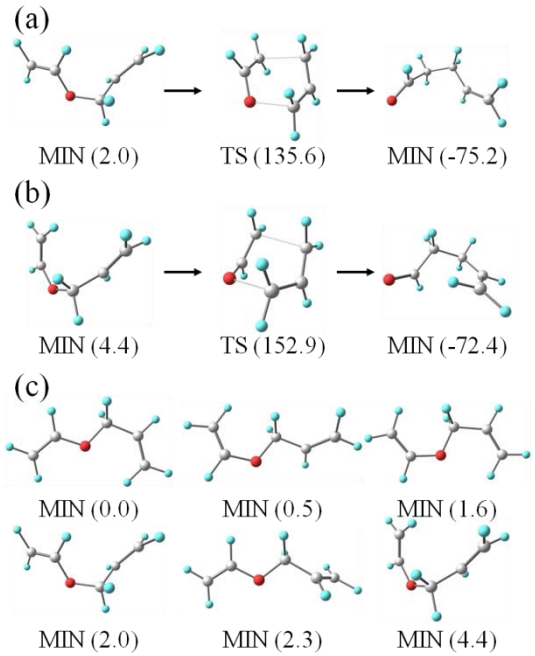

Fig. 3.3 Elementary steps via the chair-type TS (a) and the boat-type TS (b), and the six most stable conformers in the reactant region (c) for the Claisen rearrangement of allyl vinyl ether. The relative Gibbs free energy values at 469.1 K are shown in parentheses in kJ mol^{-1}.

The two elementary steps that directly connects one conformer of the reactant with one conformer of the product are shown in Fig. 3.3 (a) and (b). The six most stable conformers in the reactant region are also shown in Fig. 3.3 (c). The relative Gibbs free energy values of the different conformers at 469.1 K are shown in parentheses in kJ mol^{-1}. The rate constants for the two elementary steps in Fig. 3.3 (a) and (b) were determined to be 1.454×10^{-2} and 3.209×10^{-4} s^{-1}, respectively. When these two direct paths were the only pathways considered for the Claisen rearrangement, the total rate constant was estimated as the sum of the rate constants of these two steps. In this study, the author has referred to this approach as the rate-determining elementary step (RDES) model (see

Section 2.2), because this model only takes the rate-determining elementary steps into consideration. The rate constant of the RDES model was determined to be 1.486×10^{-2} s^{-1}, which was very different from that of the experimental value, 2.875×10^{-3} s^{-1}[26]. It was envisaged that the rate constant could be improved by considering the most stable conformer in the reactant region. With this in mind, the author considered the use of the lowest conformer to transition state (LCTS) model (same as lowest reactant to transition state model in Section 2.3). For the LCTS model, the most stable conformer in the reactant region shown in Fig. 3.3 (c) was the initial state of the two elementary steps in Fig. 3.3 (a) and (b). In other words, the rate constant was calculated using eq. (2.2) based on the Gibbs free energy gaps between the TSs in Fig. 3.3 (a) and (b) and the lowest energy conformer in Fig. 3.3 (c). This model is typically used for the analysis of organic reactions when the most stable (or likely most stable) conformer in the reactant region is available. The rate constant determined using the LCTS model, which is the sum of the two LCTS rate constants for the two elementary steps in Fig. 3.3 (a) and (b), was 8.881×10^{-3} s^{-1}. The value obtained by the LCTS model was therefore closer to the experimental value than that of the RDES model.

Table 3.1 Rate constants k and the corresponding activation free energies $\Delta G^{\ddagger}$.

	k / s^{-1}	$\Delta G^{\ddagger}$ / kJ mol^{-1}
RDES model [a]	1.486×10^{-2}	133.1 (133.5) [c]
LCTS model [b]	8.881×10^{-3}	135.1 (135.5) [c]
RCMC (this work)	1.794×10^{-3}	141.3 (141.8) [c]
Experiment	2.875×10^{-3}	139.5

[a] The rate-determining elementary step model (see text).

[b] The lowest conformer to transition state model (see text).

[c] Values in parentheses are obtained by eq. (2.2) with $\Gamma = 1$.

A detailed discussion of the rate constant obtained by the RCMC is provided below. This rate constant accounts for all of the conformational rearrangement pathways in the reactant region. The rate constant obtained using the RCMC was 1.794×10^{-3} s^{-1}. Notably, this value was in very good agreement with the experimental value. In this study, the k_{MAX} value was set as 6.931×10^{-1} s^{-1}; in an irreversible reaction with a k value of 6.931×10^{-1} s^{-1}, the population of the reactant would decrease to 50% of its original value in one sec with the total number of states being two. This k_{MAX} value was reasonable because only two superstates remained after the recursive contraction and these two corresponded to those for the reactant and product. It is noteworthy that the overall rate constant of the current method did not change when the k_{MAX} value was varied from 1.794×10^{-3} to 6.061×10^{10} s^{-1}. The consistency of the rate constant was attributed to the fact that the number of final superstates did not change for any of these variations in the k_{MAX} value for this reaction network. This result therefore indicated that the reaction network was not sensitive to the k_{MAX} value. The k_{MAX} value can therefore be decided

easily based on the final superstates obtained with a few different k_{MAX} value.

The rate constants k obtained by the different models are shown in Table 1, together with the corresponding activation free energies $\Delta G^{\ddagger}$, which were estimated using eq. (3.9).

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{hk}{\Gamma k_B T}\right) \qquad (3.9)$$

The activation free energies based on the rate constants obtained using eq. (2.2) with an Γ value of eq. (2.3), and those calculated with an Γ value of 1, were very close to each other. This result indicated that the one-dimensional tunneling correction in eq. (2.3) was having very little impact on the rate constant. The rate constants were systematically improved by considering the most stable conformer (LCTS model vs. RDES model), as well as all of the conformational rearrangement paths (i.e., the results of the RCMC vs. those of the LCTS model). In terms of the activation free energy, the RCMC reproduced the experimental value with an error of only 1.8 kJ mol^{-1}. The difference between the activation free energies of the RCMC and the RDES model was 8.2 kJ mol^{-1}, which was attributed to the conformational entropy in the reactant region.

3.4 Application2: Claisen rearrangement of Ph-CH-CHO-CH$_2$-CH-CH-Me

I subsequently applied the RCMC to the Claisen rearrangement of the substituted allyl vinyl ether Ph-CH-CHO-CH$_2$-CH-CH-Me, where a terminal H atom on the allyl and vinyl groups was substituted with Ph and Me, respectively. The reaction network for this rearrangement was calculated at the M062X/6-311+G(2d,p) level[27]. I did not perform CCSD(T) single point calculations in this case because an experimental rate constant was not available. Furthermore, the purpose of this part of the study was to demonstrate the

applicability of the RCMC to a reaction network that also results in the formation of a byproduct. This reaction gives two stereoisomers, including the *syn*- and *anti*-products. Using the SC-AFIR, I obtained 17, 35 and 32 conformers for the reactant, *syn*-product and *anti*-product, respectively, as well as 278 TSs. Among the 278 TSs, two corresponded to the Claisen rearrangement, whilst the remaining 276 corresponded to conformational rearrangements. This result was confirmed by computing all of the IRC paths starting from the 278 different TSs. Fig. 3.4 (a) and (b) shows the two elementary steps responsible for the formation of the *syn*- and *anti*-products via the chair- and boat-type TSs, respectively. The six most stable conformers in the reactant region are also shown in Fig. 3.4 (c). The relative free energy values at 469.1 K are shown in kJ mol^{-1} in parentheses in Fig. 3.4. These results indicated that the preferred path to the *syn*-product proceeded via a chair-type TS, and were therefore in agreement with general understanding of the Claisen rearrangement.

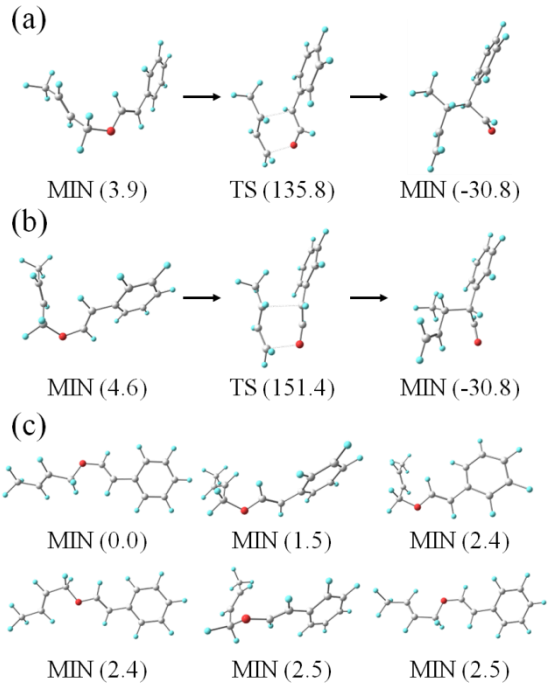

Fig. 3.4 Elementary steps via the chair-type TS (a) and the boat-type TS (b), and the six most stable conformers in the well of the reactant (c) for the Claisen rearrangement of a substituted allyl vinyl ether. The relative Gibbs free energy values at 469.1 K are shown in parentheses in kJ mol^{-1}

Table 3.2 Rate constants k and the corresponding activation free energies $\Delta G^{\ddagger}$.

Method	k [c] / s^{-1}	$\Delta G^{\ddagger}$ [c] / kJ mol^{-1}	k [d] / s^{-1}	$\Delta G^{\ddagger}$ [d] / kJ mol^{-1}
RDES model [a]	2.231×10^{-2}	131.5 (131.9) [e]	4.848×10^{-4}	146.4 (146.8) [e]
LCTS model [b]	8.327×10^{-3}	135.3 (135.8) [e]	1.526×10^{-4}	150.9 (151.4) [e]
RCMC (his work)	1.434×10^{-3}	142.2 (142.6) [e]	2.601×10^{-5}	157.8 (158.2) [e]

[a] The rate-determining elementary step model (see text).

[b] The lowest conformer to transition state model (see text).

[c] Values for the *syn*-product.

[d] Values for the *anti*-product.

[e] Values in parentheses are obtained by eq. (2.3) with $\Gamma = 1$.

The rate constants from the reactant region to the two products were estimated by the three different models described above, and the resulting values are shown in Table 2 for comparison. The corresponding activation free energies were estimated using eq. (3.9), and the resulting values are also shown in Table 3.2. The k_{MAX} parameter for these calculations was once again set to 6.931×10^{-1} s^{-1}. This k_{MAX} value was reasonable also in this system because three superstates remained and these three corresponded to those for the reactant, product (*syn*-product) and byproduct (*anti*-product). The general trends

observed in Table 3.2 were similar to those observed in Table 3.1. The rate constants obtained by the RDES model were the largest, followed by those of the LCTS model and the present method, which were the smallest. The results with $\Gamma = 1$ indicated that the one-dimensional tunneling correction was making a negligible contribution to the calculations. It is noteworthy, that the activation free energy determined using the RDES model was significantly lower than the value obtained by the RCMC, with a difference of 10.7 kJ mol^{-1}. This difference in the two values was attributed to differences in the conformational entropy in the reactant region. The impact of this difference on the overall rate constant was accounted for in the current method by considering all of the conformers in the reactant region, as well as their interconversion pathways. The importance of sampling the different molecular conformations of the molecules involved in organic reactions has recently been highlighted in the literature[5,30]. The importance of different molecular conformation has also been recognized in enzymatic reactions[31,32]. The results demonstrated that conformational effects can also have a significant impact on the activation free energy of simple organic molecules.

To let the discussion be more accurate, the author estimated the $\Delta G^{\ddagger}$ values after the CCSD(T)-F12a single-point calculations by applying a very simple correction to the profile obtained by the M062X calculations. For this purpose, the relative energy of the two key TSs shown in Fig. 3.4 (a) and (b) were corrected by the difference between the relative energy obtained for the analogous TSs shown in Fig. 3.3 (a) and (b) with the M062X and CCSD(T)-F12a levels. This treatment relies on an assumption that level dependences of computational errors in these bond reorganization TSs are similar between the two different systems. The $\Delta G^{\ddagger}$ values were then estimated by the RCMC or the RDES model using the reaction network of the M062X calculations with this

correction. The correction reduced the $\Delta G^{\ddagger}$ values for reactions giving the *syn*- and *anti*-product by 6.3 and 8.9 kJ mol^{-1}, respectively, regardless of models (the RCMC or the RDES model), and the best estimates of the $\Delta G^{\ddagger}$ values for these two channels were therefore 135.9 and 148.9 kJ mol^{-1}, respectively.

The RCMC therefore reproduced the thermodynamic equilibrium despite the fact that the QSSA is used. The correction in eq. (3.4), which was introduced to impose this condition, reduced the errors caused by the QSSA. A numerical illustration of this improvement is provided in Section 3.5 for a simple profile. Further investigation towards developing a deeper understanding of the importance of this correction is currently underway.

The accuracy of any method of quantum chemical calculation should be considered as an important performance parameter because poor accuracy can lead to significant errors in the calculations. The approximations used in the Gibbs free energy evaluations, including the harmonic, rigid rotor and ideal gas approximations, can also be significant error sources. In this study, the author assumed that the IRC was the reaction path and that the two endpoints of the IRC were the initial and final states of each elementary step. An assumption of this type can lead to large errors in molecular systems that strongly deviate from the IRC in terms of their dynamics or systems with dynamic bifurcation pathways[22]. The use of eqs. (2.2) and (2.3) to determine the rate constant of a specific transformation, the Born-Oppenheimer approximation, as well as the assumption that the system follows the Markov property, can also result in errors. Whilst I appreciate that there are many different potential sources of error in these calculation, the author feels that a detailed study of these different error sources and their impact on the calculations conducted in this study is beyond the scope of this study.

The RCMC gave a few superstates that were connected by a slow process. This result is similar to those of several previous methods where the original states are grouped together to give only a few superstates[13,14]. The results of the current study were also similar in some ways to those of the reduction methods, where the rapidly converting channels and states are eliminated[10-12]. The RCMC can therefore be considered as an extension of these methods. However, the definition of a superstate used in the RCMC differed from the definitions used in previous methods, in that the superstates were expressed as the weighted sums of the original states (EQs). In other words, all original states in the current study made non-zero contributions to all the superstates. This representation of the superstate would therefore provide a reasonable chemical picture for intermediate states that provide an equal contributing to two superstates.

One advantage of the RCMC is that the overall rate constants can be obtained by simple inputs, i.e., a rough reaction timescale (e.g., $1/k_{MAX}$) and a reaction profile. As shown in this example, the RCMC allowed for several superstates to remain and provided a platform for discussing the selectivity. The use of a timescale based on $1/k_{MAX}$ scans therefore allows for the hierarchical organization of the different states with respect to time. The hierarchical organization of the individual steps buried within these reaction networks with respect to time is shown elsewhere in a separate paper together with new definitions for the TSs throughout the reaction profiles/networks[33].

The computational cost of the present algorithm was found to be proportional to $<s^3$, where s is the number of original states (EQs). During the first contraction loop, the s^2 elements in the rate constant matrix were updated. The contraction loop was subsequently repeated no more than $s - 2$ times. Given that the rate constant matrix is sparse and the size of the matrix reduces in size with each contraction loop, it was

envisaged that the total cost would scale as $<s^3$. This cost would be acceptable for the networks of organic reactions with less than one thousand states. In fact, the calculations conducted in the current study for profiles containing 23 and 84 states rapidly reached completion. Furthermore, it is apparent from eqs. (3.1)–(3.4) that it would not be possible for the magnitude of the denominator to become much smaller than that of numerator in these equations, and that the present algorithm is therefore numerically stable. Whilst it is envisaged that the method disclosed in the current study could be applied to a variety of different reactions, the author considers the key focus of this work to be its application to organic reactions. In this study, the author has only applied this new method to first-order reactions and its extension to second- or higher-order reactions will form the basis of a future publication.

3.4 Derivation of eq. (3.4)

In the (n + 1)-th loop, a part of population at $t = \infty$ of the state i which is contracted in this loop is added to the state m by eq. (3.3). In this study, this correction of population at $t = \infty$ is reflected to the rate constant by the following procedure. In the following, this procedure is explained using four quantities: population at $t = \infty$ of state $P_m^{(n)}$, rate constant $k_{l\to m}^{(n)}$, relative free energy of state $\Delta G_m^{(n)}$, and relative free energy of TS $\Delta G_{\mathrm{TS}_{l-m}}^{(n)}$. I further introduce the following three assumptions.

I. $P_m^{(n)}$ follows the Boltzmann distribution in any n.

II. $\Delta G_m^{(n)}$ changes only when $P_m^{(n)}$ changes.

III. $\Delta G_{\mathrm{TS}_{l-m}}^{(n)}$ changes only when the QSSA is applied to the state i ($l \neq i$, $m \neq i$).

Below, using eq. (3.3) and these three assumptions, eq. (3.4) is derived. From the assumption I, $P_m^{(n)}$ is expressed by eq. (3.10).

$$P_m^{(n)} = \frac{e^{-\Delta G_m^{(n)}/RT}}{\sum_{j \in N^{(n)}} e^{-\Delta G_j^{(n)}/RT}} \qquad (3.10)$$

Let us consider a state l which is not adjacent to the state i. For such a state l, $P_l^{(n)}$ does not change in the (n + 1)-th loop ($P_l^{(n+1)} = P_l^{(n)}$), because $k_{i \to l}^{(n)} = 0$ and the second term of the right-hand side of eq. (3.3) is zero. Following the assumption II, $\Delta G_l^{(n)}$ remains unchanged as well ($\Delta G_l^{(n+1)} = \Delta G_l^{(n)}$). These two conditions, that is, $P_l^{(n+1)} = P_l^{(n)}$ and $\Delta G_l^{(n+1)} = \Delta G_l^{(n)}$, are satisfied simultaneously only if the following relation is satisfied.

$$\sum_{j \in N^{(n+1)}} e^{-\Delta G_j^{(n+1)}/RT} = \sum_{j \in N^{(n)}} e^{-\Delta G_j^{(n)}/RT} \qquad (3.11)$$

When the assumptions I and II are adopted, eq. (3.11) holds in any n. In other words, the denominator of the right-hand side of eq. (3.10) does not change depending on n. Thus, I replace the denominator by a constant W, and rewrite eq. (3.10) as follows.

$$P_m^{(n)} = \frac{e^{-\Delta G_m^{(n)}/RT}}{W} \qquad (3.12)$$

Based on eq. (3.12), the relative free energy of the state k, $\Delta G_k^{(n+1)}$, is given by the following equation.

$$\Delta G_k^{(n+1)} = -RT \ln\left(P_k^{(n+1)} W\right) \qquad (3.13)$$

Herein, using eq. (2.2), $k_{k \to l}^{(n+1)}$ is expressed by the following equation.

$$k_{k \to l}^{(n+1)} = \Gamma \frac{k_B T}{h} e^{-\left(\Delta G_{\mathrm{TS}_{k-l}}^{(n+1)} - \Delta G_k^{(n+1)}\right)/RT} \qquad (3.14)$$

On the other hand, the rate constant obtained by the QSSA, $k'^{(n+1)}_{k \to l}$, is expressed by the following equation.

$$k'^{(n+1)}_{k \to l} = \Gamma \frac{k_B T}{h} e^{-\left(\Delta G_{\mathrm{TS}_{k-l}}^{(n+1)} - \Delta G_k^{(n)}\right)/RT} \qquad (3.15)$$

Note in eq. (3.15) that $\Delta G_k^{(n+1)}$ in eq. (3.14) is replaced by $\Delta G_k^{(n)}$. This is because $P_k^{(n)}$ has not yet been updated when the QSSA is applied in the (n + 1)-th loop; following the

assumption II, neither $\Delta G_k^{(n)}$ nor $P_k^{(n)}$ has yet been updated. In contrast, following the assumption III, $\Delta G_{\mathrm{TS}_{k-l}}^{(n+1)}$ is used in both eq. (3.14) and eq. (3.15), because both $k_{k\to l}^{(n+1)}$ and $k'^{(n+1)}_{k\to l}$ are quantities that have been obtained after application of the QSSA to the state i in the $(n + 1)$-th loop.

Let us consider the following apparent relation.

$$k_{k\to l}^{(n+1)} = \frac{k_{k\to l}^{(n+1)} P_k^{(n+1)}}{P_k^{(n+1)}} \qquad (3.16)$$

Then, by substituting eq. (3.3) into eq. (3.16), the following equation is obtained.

$$k_{k\to l}^{(n+1)} = \frac{k_{k\to l}^{(n+1)} P_k^{(n+1)} / P_k^{(n)}}{1+\sigma_i^{(n)} k_{i\to k}^{(n)} P_i^{(n)} / P_k^{(n)}} \qquad (3.17)$$

At thermal equilibrium, the following relation concerning the detailed balance is satisfied.

$$k_{i\to k}^{(n)} P_i^{(n)} = k_{k\to i}^{(n)} P_k^{(n)} \qquad (3.18)$$

By comparing eqs. (3.14) and (3.15), the following relation is obtained.

$$k'^{(n+1)}_{k\to l} = k_{k\to l}^{(n+1)} P_k^{(n+1)} / P_k^{(n)} \qquad (3.19)$$

Finally, eq. (3.4) can be obtained by substituting eqs. (3.18) and (3.19) into eq. (3.17).

3.5 Impact of Rate Constant Matrix Contraction

Using a simple network containing three states (similar to Fig. 2.1), impact of the correction of eq. (3.4) is examined. Here, consider a profile containing states **A**, **B**, and **C**. The states **A** and **B** are connected by **TS$_{AB}$**, and **B** and **C** by **TS$_{BC}$**. Relative free energy values of **A**, **B**, **C**, **TS$_{AB}$**, and **TS$_{BC}$** are set to 0.0, 10.0, −∞, 50.0, and 75.0 kJ/mol, respectively. For simplicity, the reaction flux from **C** to **B** was set to zero by setting the relative free energy of state **C** to −∞.

Time evolution of population of state **C** was simulated starting from the initial condition in which state **A** is populated 100%, by the fourth-order Runge-Kutta method with the time step of 1.150×10^{-6} s. The simulation was applied to three profiles: the original three state network, a two state network simplified by the QSSA, and a two state network simplified by the QSSA with the correction of eq. (3.4). Fig. 3.5 shows time evolutions of population of state **C** obtained by the simulations on the three different profiles. In Fig. 3.5, the green solid curve, the blue solid curve, and the red dotted curve, respectively, correspond to the time evolution of state **C** on the original three state network (labeled as "reference"), on the network simplified by the QSSA (labeled as "QSSA"), and on the network simplified by the QSSA with the correction of eq. (3.4) (labeled as "this work"). Curves of both "QSSA" and "this work" reproduced the reference curve well; the three curves nearly overlap in Figure 3.5. The curve of "this work" still overlap with the reference curve even in the enlarged figure shown in inset. This demonstrates that the correction of eq. (3.4) reduces the error of the QSSA.

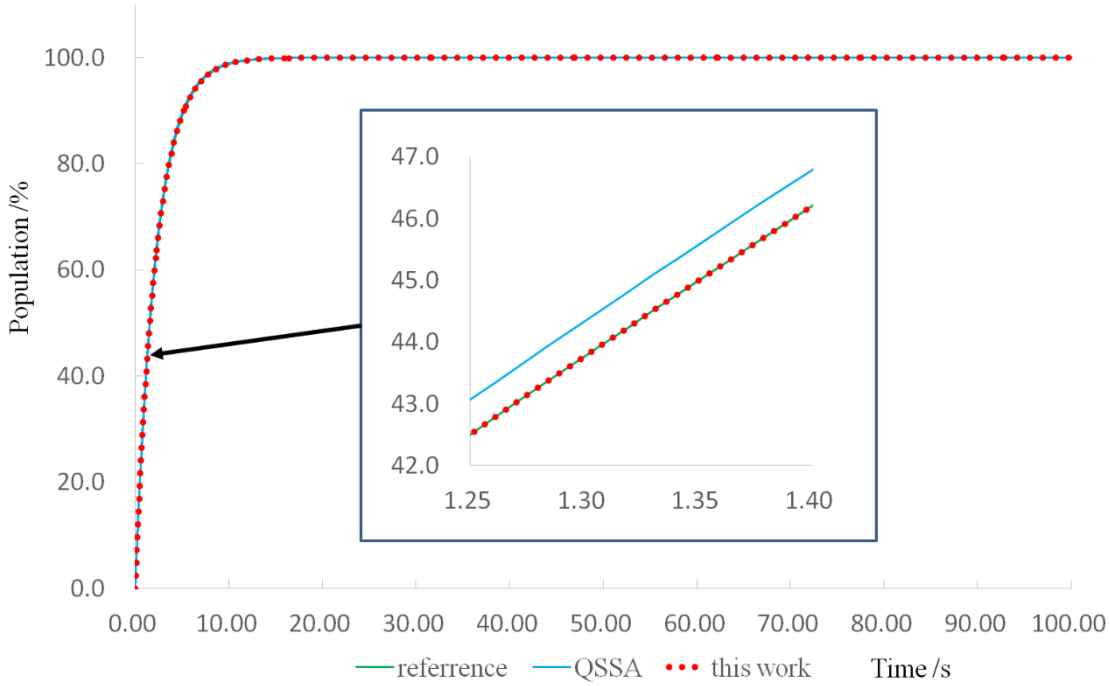

Fig. 3.5 The time dependency of population of state **C**. The curve obtained by solving the rate equations for the original three state network numerically is shown with the green solid curve and is denoted “reference”. The curve for the network simplified by the QSSA is shown with the blue solid curve and denoted “QSSA”, and the curve for the network simplified by the QSSA with the correction of eq. (3.4) is shown with the red dotted curve and denoted “this work”. An enlarged figure is shown in inset to emphasize the impact of the correction of eq. (3.4).

3.6 Conclusion

In this work, the author proposed a new approach for estimating the overall rate constant of a given reaction profile “Rate Constant Matrix Contraction (RCMC)”. A new algorithm was developed and applied to the Claisen rearrangement reactions of allyl vinyl ether and a substituted allyl vinyl ether. The overall rate constant obtained for the Claisen rearrangement of allyl vinyl ether was compared with the experimental rate constant in the gas-phase. The results revealed that the computed rate constant reproduced the experimental value very closely. It was also found that the conformational entropy in the reactant region made a substantial contribution to the overall rate constant even in the Claisen rearrangement of small chain molecules. It is therefore envisaged that the current method could be used as a powerful tool for estimating the overall rate constants of different organic reactions.

References

[1] N. Koga, K. Morokuma, *Chem. Rev.* **1991**, *91*, 823.

[2] S. Niu, M. B. Hall, *Chem. Rev.* **2000**, *100*, 353.

[3] K. N. Houk, P. H.-Y. Cheong, *Nature* **2008**, *455*, 309.

[4] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, D. G. Truhlar, *Chem. Rev.* **2006**, *106*, 4518.

[5] S. J. Klippenstein, V. S. Pande, D. G. Truhlar, *J. Am. Chem. Soc.* **2014**, *136*, 528.

[6] S. Kozuch, S. Shaik, *Acc. Chem. Res.* **2010**, *44*, 101.

[7] L. E. Rush, P. G. Pringle, J. N. Harvey, *Angew. Chem., Int. Ed.* **2014**, *53*, 8672.

[8] J. Wei, J. C. Kuo, *Ind. Eng. Chem. Fundam.* **1969**, *8*, 114.

[9] J. C. Kuo, J. Wei, *Ind. Eng. Chem. Fundam.* **1969**, *8*, 124.

[10] J. C. Keck, *Prog. Energy Combust. Sci.* **1990**, *16*, 125.

[11] T. Lu, K. L. Chung, *Proc. Combust. Inst.* **2005**, *30*, 1333.

[12] T. Turanyi, A. S. Tomlin, Analysis of Kinetic Reaction Mechanisms; Springer, Berlin, **2014**; pp 210-231.

[13] D. A. Evans, D. J. Wales, *J. Chem. Phys.* **2003**, *118*, 3891.

[14] S. V. Krivov, M. Karplus, *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 14766.

[15] J. S. Shaffer, A. K. Chakraborty, *Macromolecules* **1993**, *26*, 1120.

[16] M. Rae, M. N. Berberan-Santos, *J. Chem. Educ.* **2004**, *81*, 436.

[17] D. T. Gillespie, *Ann. Rev. Phys. Chem.* **2007**, *58*, 35.

[18] A. Chatterjee, D. G. Vlachos, *J. Comput. Aided Mater. Des.* **2007**, *14*, 253.

[19] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, *35*, 166.

[20] H. B. Schlegel, *J. Comput. Chem.* **2003**, *24*, 1514.

[21] D. J. Wales, *Int. Rev. Phys. Chem.* **2006**, *25*, 237.

[22] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int. J. Quantum Chem.* **2015**, *115*, 258.

[23] K. Fukui, *J. Phys. Chem*. **1970**, *74*, 4161.

[24] L. Kürti, B. Czakó, Claisen rearrangement. In Strategic applications of named reactions in organic synthesis; Elsevier B.V.: Amsterdam, The Netherlands, 2005; pp 88-89.

[25] R. L. Vance, N. G. Rondan, K. N. Houk, F. Jensen, W. T. Borden, A. Komornicki, E. Wimmer, *J. Am. Chem. Soc.* **1988**, *110*, 2314.

[26] F. W. Schuler, W. M. George, *J. Am. Chem. Soc.* **1950**, *72*, 3155.

[27] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[28] T. B. Adler, K. Gerald, H.-J. Werner, *J. Chem. Phys.* **2007**, *127*, 221106/1-4.

[29] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, D. G. Truhlar, *J. Chem. Theory Comput.* **2011**, *7*, 3027.

[30] P. Seal, D. G. Truhlar, *J. Am. Chem. Soc.* **2014**, *136*, 2786.

[31] B. P. English, W. Min, A. M. Van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, X. S. Xie, *Nat. Chem. Biol.* **2006**, *2*, 87.

[32] M. B. J. Roeffaers, G. D. Cremer, H. Uji-I, B. Muls, B. F. Sels, P. A. Jacobs, C. D. F. Schryver, D. E. D. Vos, Hofkens, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12603.

[33] Y. Nagahata, S. Maeda, H. Teramoto, T. Horiyama, T. Taketsugu, T. Komatsuzaki, *J. Phys. Chem. B*, **2016**, *120*, 1961.

[34] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T. Taketsugu, S. Maeda, *J. Phys. Chem. A*, **2015**, *119*, 11641.

Chapter 4 Rate Constant Matrix Contraction in Microcanonical Ensemble and Full Rate Constant Matrix Contraction for Obtaining Branching Ratio of Unimolecular Decomposition

4.1 Introduction

Unimolecular decomposition of small molecules is one of important processes in various reactions such as combustion reaction, atmospheric reaction, and interstellar reaction [1-6]. Hence, there have been a number of experimental and theoretical studies. In such studies, theoretical calculation has been employed as a powerful tool.

There are two main approaches to theoretically obtain the branching ratio. One of them is the ab initio molecular dynamical (AIMD) simulation [7-14]. The branching ratio can be computed by running a huge number of trajectories and taking their statistics. This approach does not use any assumption except for that the nuclei move on the adiabatic potential energy surface (PES) following classical mechanics, and therefore can be quite accurate when it is combined with a reliable quantum chemical calculation method. However, in general it is computationally demanding to calculate pathways that occur in timescales longer than 1 nanosecond by AIMD simulations.

The other is the kinetic simulation based on the transition state theory (TST) [15-20], and it allows for seeing events that occur in timescales of nanosecond or longer. This approach requires a reaction path network including all relevant transition state (TS) structures and equilibrium (EQ) structures. Although obtaining all TSs and EQs has been a hard task, recent advances in algorithms of automated reaction path search have reduced the difficulty considerably[21]. However, there remains another difficulty. It can be

difficult to achieve a time evolution covering all of the time scales buried in an entire reaction network when the elementary steps have a wide range of different rate constants due to the numerical stability and the computational cost[22]. This problem has been called as stiff problem in numerical analysis of differential equations. This makes it difficult to obtain fully converged populations of dissociated products even in small systems. On the other hand, the first-order rate equations are linear differential equations, for which a general solution is available[23]. The general solution requires eigenvalues of the rate constant matrix. The stiffness in the rate equations again causes a serious problem. Numerical diagonalization of asymmetric matrices that include elements with a wide range of different values gives severe numerical errors[24]. This again prevents one from obtaining branching ratios using this approach.

In this chapter the author proposes an alternative approach based on the rate constant matrix contraction (RCMC) method [25] introduced in Chapter 3. This method also requires the rate equations but does neither kinetic simulation nor AIMD simulation. The RCMC has originally been developed to reduce the size of the rate equations and to accelerate the kinetic simulation. In this chapter, the author introduces an additional procedure to redistribute initial populations. Furthermore, the contraction is applied to all EQs, where it is applied only to kinetically unstable EQs in the original RCMC scheme. Such a full contraction (*f*-RCMC) gives the branching ratio for all the dissociated products without solving the rate equations.

This *f*-RCMC method assumes that molecules once dissociated do not return. Therefore, the *f*-RCMC would be suitable to analysis of unimolecular decomposition in gas-phase. In this chapter, formulas of RCMC that had been originally derived for the canonical ensemble was reformulated for the micro-canonical ensemble. Fortunately,

identical formulas were obtained for both canonical and micro-canonical ensembles. It follows that the *f*-RCMC procedure can be applied to both canonical and micro-canonical ensembles without any change.

The *f*-RCMC method was applied to gas-phase unimolecular decomposition of C_3H_5 and C_4H_5. For these radical species, reaction path networks including 30 and 234 elementary steps, respectively, were obtained by the single-component artificial force induced reaction (SC-AFIR) [26]. With the resulting rate constant matrices, dissociation branching ratios were computed by two different approaches, i.e., the present *f*-RCMC method and the conventional kinetic simulation. Values obtained by these two approaches agreed rigorously up to eight decimal places. This demonstrated that concerning accuracy the *f*-RCMC could be a perfect alternative to the conventional kinetic simulation approach. Furthermore, for C_4H_5, the conventional kinetic simulation needed time evolution until 2.10×10^{-7} sec. to give converged branching ratios, and this simulation took about 95 hours. In contrast, the *f*-RCMC gave the values quickly in less than 1 second. The *f*-RCMC method would thus be a highly efficient alternative of the conventional kinetic simulation and be promising in evaluating unimolecular decomposition branching ratios in large systems.

4.2 Algorithm of Rate Constant Matrix Contraction in Microcanonical Ensemble

In this chapter, the RCMC is described again before introducing the present *f*-RCMC approach. The RCMC reduces the size of rate constant matrix from $N\times N$ to $(N-M)\times(N-M)$ through M time contractions. This provides $N-M$ superstates that are given as a linear combination of the N original states, where an original state corresponds to EQ. The superstate could be compared to the superbasin [29]; the superstate is

represented as a weighted sum of all original states in contrast to the superbasin that is a group of original states. Moreover, off-diagonal elements of the resulting $(N-M)\times(N-M)$ rate constant matrix directly correspond to overall rate constants between superstates.

To determine coefficients of the linear combination of original states, the author introduced a procedure, i.e., redistribution of Boltzmann distribution P_i of each original state i to the resulting superstates [25]. This procedure guaranteed that a simulation with the contracted rate constant matrix converges to the one with the original rate constant matrix at a sufficiently long timescale. As the author had done in my previous study, P_i was redistributed in systems described by the canonical ensemble. In contrast, the density of states N_i rather than P_i is redistributed in systems described by the microcanonical ensemble. As shown in section 4.6, redistribution of N_i gave identical formula that was obtained previously for redistribution of P_i. The following procedure as well as formulas can thus be applicable to both canonical and microcanonical ensembles.

A procedure to contract a $(N-n)\times(N-n)$ rate constant matrix into a $(N-n-1)\times(N-n-1)$ one is below. In the following algorithm, the set of states (either original states at $n = 0$ or superstates at $n \geq 1$) obtained at the nth contraction is denoted by $S^{(n)}$; states i, j, k, l, and m correspond to the components of $S^{(n)}$ ($i, j, k, l, m \in S^{(n)}$); the rate constant for the elementary step from state i to state j is denoted by $k_{i\to j}^{(n)}$; and the density of states for state i is denoted by $N_i^{(n)}$. It follows that $S^{(0)}$, $k_{i\to j}^{(0)}$, and $N_i^{(0)}$ correspond to the set of original state (EQ), the original rate constant, and the original density of states, respectively.

The contraction is made with the following seven steps.

1. Identify the state pair i and j with the maximum rate constant $k_{i\to j}^{(n)}$.

2. Apply the quasi-steady state approximation to the state *i* and update $k_{i\to j}^{(n)}$ to $k_{k\to l}'^{(n+1)}$ using eqs. (4.1) and (4.2).

$$k_{k\to l}'^{(n+1)} = k_{k\to l}^{(n)} + k_{k\to i}^{(n)} k_{i\to l}^{(n)} \sigma_i^{(n)} \qquad (4.1)$$

$$\sigma_i^{(n)} = \frac{1}{\sum_{m\in S^{(n)}} k_{i\to m}^{(n)}} \qquad (4.2)$$

3. Update all rate constants related to the state *i* to zero ($k_{i\to m}'^{(n+1)} = k_{m\to i}'^{(n+1)} = 0$).

4. Update density of states $N_i^{(n+1)}$ using eq. (4.3).

$$N_k^{(n+1)} = N_k^{(n)} + k_{i\to k}^{(n)} \sigma_i^{(n)} N_i^{(n)} \qquad (4.3)$$

In this step, redistribution of $N_i^{(n)}$ to the neighboring states is made. By referring the ratio of rate constants $k_{i\to k}^{(n)} \sigma_i^{(n)}$ from the state *i* to the other states, a part of $N_i^{(n)}$, i.e., $k_{i\to k}^{(n)} \sigma_i^{(n)} N_i^{(n)}$, is added to $N_k^{(n)}$.

5. Update $N_i^{(n+1)}$ to zero.

6. Update $k_{k\to l}'^{(n+1)}$ to $k_{k\to l}^{(n+1)}$ using eq. (4.4).

$$k_{k\to l}^{(n+1)} = \frac{1}{1+\sigma_i^{(n)} k_{k\to i}^{(n)}} k_{k\to l}'^{(n+1)} \qquad (4.5)$$

At $t = \infty$, a kinetic simulation with rate constants in this form reproduces the ratio of densities of states of the original states, i.e., the correct distribution at $t = \infty$ [30, 31]. Note that eq. (4.5) is identical to the one derived for the canonical ensemble shown in Section 3.2 [25].

7. Define $S^{(n+1)}$ as the set of all superstates except the contracted state *i*.

The contraction is repeated recursively until the maximum rate constant in the

contracted rate constant matrix becomes smaller than a certain threshold. This threshold is denoted by k_{MAX} and given by users. The value of k_{MAX} has been estimated using a timescale τ of target reaction as $k_{MAX} = \tau^{-1}$. After the contraction, a few superstates that are connected by slow steps with small rate constants $k < k_{MAX}$ are obtained. This enables a sufficiently long time-evolution in the kinetic simulation. It is noted that N_i is used in this study for the microcanonical ensemble, and N_i seen in the above contraction steps is just replaced by P_i when the canonical ensemble is considered [25] (Section 3.2). Although previous studies have discussed the error of quasi steady state approximation [32, 33], redistribution of density of states (or Boltzmann distribution) in step 4 of the present algorithm provides a reasonable correction, as shown in Chapter 3.

4.3 Full Rate Constant Matrix Contraction

Some IRC paths correspond to dissociation channels (DCs). In this work, the DC which in general consists of two or more species is considered as one state. For example, a reaction $H_2C{=}C\text{-}CH_3 \rightarrow H_2C{=}C + CH_3$ is regarded as a transition from state i ($H_2C{=}C\text{-}CH_3$) to state j ($H_2C{=}C + CH_3$). Furthermore, the rate constant for the elementary step from a state of DC to a state of EQ is set to zero. This treatment assumes that the secondary reaction between dissociated products does not occur and thus is applicable only to unimolecular decomposition in the gas-phase.

In the *f*-RCMC algorithm, the RCMC introduced in the last section is applied to the rate constant matrix including DCs as original states. Its application is repeated until all rate constants become zero, and therefore this procedure is termed *f*-RCMC. If DC is not included, the *f*-RCMC gives only one superstate. When there are x DCs, the *f*-RCMC gives x superstates, where each superstate is expressed by a weighted sum of all EQs and

one DC. In other words, the *f*-RCMC distributes all EQs to each DC with appropriate weights. These superstates thus are called dissociative superstate (DSS).

In the RCMC procedure, redistribution of either N_i or P_i is made. In the *f*-RCMC, redistribution of the initial population Q_i is additionally introduced. To be precise, in the step 4 of the RCMC algorithm, $Q_k^{(n+1)}$ is updated by the following equation.

$$Q_k^{(n+1)} = Q_k^{(n)} + k_{i\to k}^{(n)} \sigma_i^{(n)} Q_i^{(n)} \qquad (4.6)$$

The original (initial) population $Q_k^{(0)}$ is defined by users considering experimental conditions. For example, among isomers (EQs) of a system, the state k (kth EQ) has population 1.0 ($Q_k^{(0)} = 1.0$) and all other states have population 0.0 ($Q_{l\neq k}^{(0)} = 0.0$) when the kth EQ is the reactant. By the full contraction, $Q_k^{(0)}$ is distributed to DSSs through M time contractions. Then, $Q_i^{(M)}$ for the ith DSS directly gives the branching ratio of the corresponding DC. Note that the initial population can be given in any ways as shown in the results section. Furthermore, I demonstrated numerically that the present *f*-RCMC method rigidly reproduce the converged branching ratios obtained by solving the original rate equations.

Reliability of the *f*-RCMC method as well as the ordinary kinetic simulation depends on the accuracy of the rate constant matrix. The standard TST which is adopted in this study assumes ideal gas, rigid rotor, and harmonic oscillator approximations, and its reliability varies depending on several factors. However, further discussion on the accuracy of the rate constant matrix is beyond the scope of this study. The present *f*-RCMC can be applied to any rate constant matrices no matter how it is obtained. On the other hand, the *f*-RCMC is not applicable to cases where elements of the rate constant

matrix vary depending on time. Second- or higher-order reactions are also cases to which the *f*-RCMC is not applicable.

4.4 Application1: Decomposition of C_3H_5

Hydrocarbon radicals play important roles in combustion chemistry. Their unimolecular decomposition has been widely studied [34-44]. In this study, I at first applied the present *f*-RCMC method to C_3H_5. The reaction network was obtained using SC-AFIR method, where the model collision energy parameter γ in the AFIR was set to 1000 kJ/mol [26]. The initial search was performed at the B3LYP/6-31G level, and all of the obtained EQs and TSs were subsequently reoptimized at the B3LYP/def2-SV(P) level. Then, single point UCCSD(T)/def2-TZVP calculations were done for all EQs and TSs obtained at the B3LYP/def2-SV(P) level. The zero point energy (ZPE), the sum of states, and the density of states were evaluated using harmonic frequencies obtained at the B3LYP/def2-SV(P) level. The computational level adopted in this study is UCCSD(T)/def2-TZVP//B3LYP/def2-SV(P). All these calculations were done using a developmental version of the GRRM program [45] combined with the Gaussian 09 electronic structure calculation program [46].

The author obtained nine EQs by the search. Among them, the five most stable EQs, denoted by EQ_a1-5, are shown in Fig. 4.1, where their relative energies (including ZPE correction) are shown in kJ/mol below each structure. The search also found seven DCs (DC1-7) as listed in Table 4.1. Dissociated fragments (DF*n*) and their structures are presented in Figure 1. Weak complexes between two DFs were regarded to be dissociated. The rate constants for all obtained elementary steps were estimated using eq (2.4).

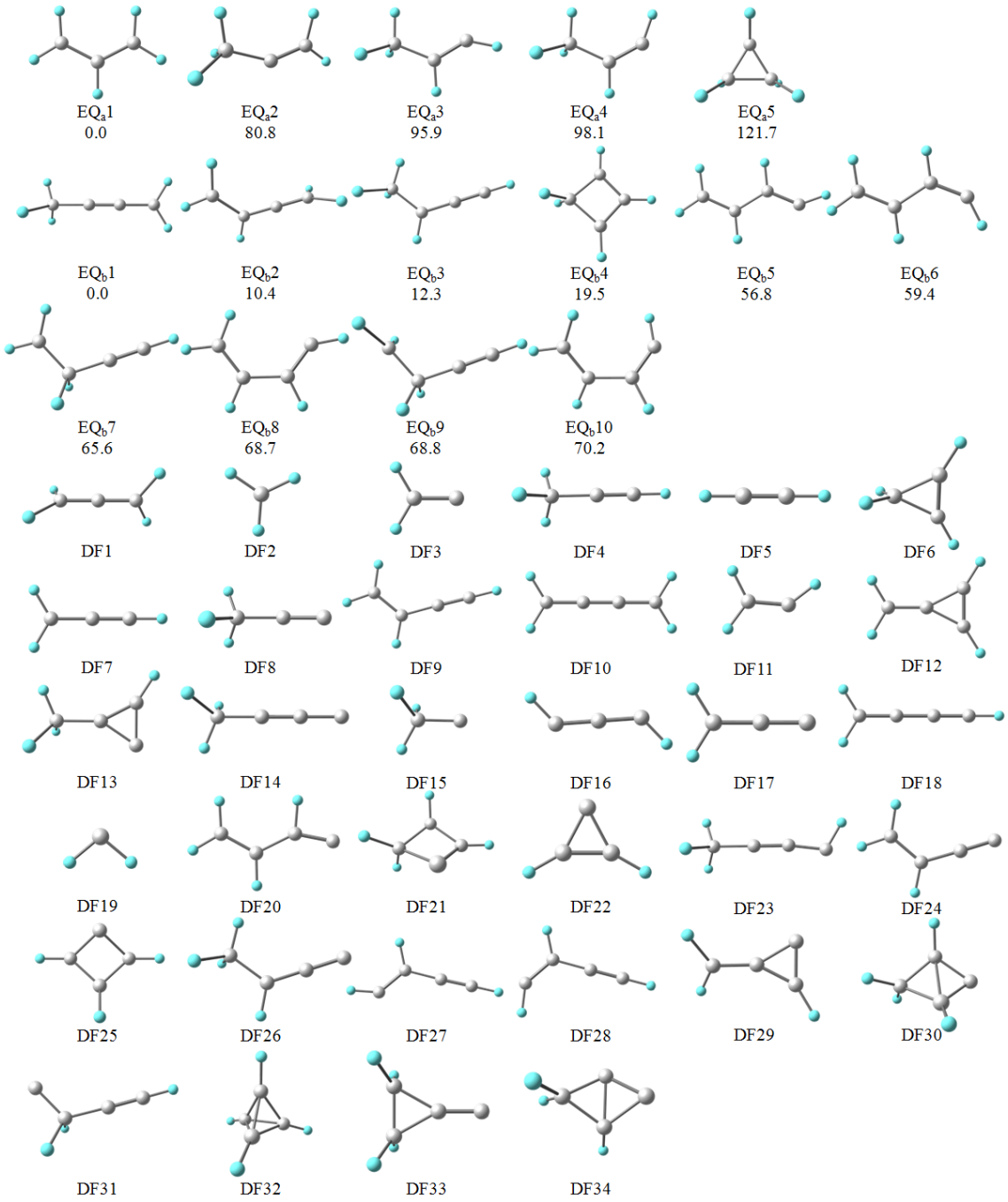

Fig. 4.1 List of equilibrium structures (EQs) and dissociated fragments (DFs) at B3LYP/def2-SV(P) level for C_3H_5 and C_4H_5 systems. EQ_a1-5 correspond to the five most stable structures for C_3H_5, and EQ_b1-10 are the ten most stable structures for C_4H_5. Relative electronic energy values including zero-point energy corrections are shown in kJ/mol. DF1-34 are dissociated fragments that are produced by unimolecular

decomposition of C_3H_5 and/or C_4H_5.

Table 4.1 A list of dissociation channels; DC1-7 for C_3H_5 and DC8-34 for C_4H_5.[a]

DC	Combination of DF	DC	Combination of DF
1	DF1 + H	18	DF18 + H_2
2	DF3 + DF2	19	DF19 + DF7
3	DF4 + H	20	DF20 + H
4	DF5 + DF2	21	DF21 + H
5	DF6 + H	22	DF22 + DF2
6	DF7 + H_2	23	DF23 + H
7	DF8 + H_2	24	DF24 + H_2
8	DF9 + H	25	DF25 + H_2
9	DF10 + H	26	DF26 + H
10	DF11 + DF5	27	DF27 + H_2
11	DF11 + DF3	28	DF28 + H_2
12	DF12 + H	29	DF29 + H_2
13	DF13 + H	30	DF30 + H
14	DF14 + H_2	31	DF31 + H_2
15	DF15 + DF5	32	DF32 + H
16	DF16 + DF2	33	DF33 + H
17	DF17 + DF2	34	DF34 + H_2

[a] See Fig. 4.1 for structures of DF*n*.

Let us consider the case where the total energy was set to 500 kJ/mol relative to EQ_a1 and the initial population 1.0 was given to EQ_a1. After the *f*-RCMC, the initial population was redistributed to each DSS. The branching ratio obtained as the ratio of $Q_i^{(M)}$ of DSSs is listed in the second column of Table 4.2. Tthe author separately performed a kinetic simulation using the original rate constant matrix for comparison, where the differential equations were solved by the 4th-order Runge-Kutta (RK4) method with the implementation described in ref. [47]. The RK4 step size was set to $0.1/k_{\text{largest}}$ s

considering the largest rate constant k_{largest} in the original rate constant matrix, and the time evolution was made until the sum of populations of all EQs become less than 10^{-20}. The branching ratios obtained by the two different approaches agreed rigorously up to thirteen decimal places. Because of the nearly perfect agreement in the branching ratios obtained by the two different approaches, only a single value is shown in each entry of Table 2. The tiny difference would be due to numerical errors arisen from the RK4 method with the finite-sized time step.

Table 4.2 The branching ratios of DC1-7 in unimolecular decomposition of C_3H_5 with the total energy 500.0 kJ/mol relative to EQ_a1.[a]

DCs	EQ_a1[b]	EQ_a2[b]	EQ_a3[b]	EQ_a4[b]	EQ_a5[b]
(DC1) DF1 + H	8.89×10^{-1}	5.75×10^{-2}	3.03×10^{-3}	1.74×10^{-3}	8.37×10^{-1}
(DC2) DF2 + DF3	5.50×10^{-2}	6.92×10^{-1}	3.34×10^{-3}	1.93×10^{-3}	5.23×10^{-2}
(DC3) DF4 + H	2.30×10^{-2}	2.50×10^{-1}	1.47×10^{-1}	8.48×10^{-2}	2.20×10^{-2}
(DC4) DF2 + DF5	1.86×10^{-2}	9.17×10^{-4}	8.47×10^{-1}	9.11×10^{-1}	1.83×10^{-2}
(DC5) DF6 + H	1.46×10^{-2}	3.74×10^{-5}	5.37×10^{-5}	3.11×10^{-5}	7.02×10^{-2}
(DC6) DF7 + H_2	1.64×10^{-4}	4.15×10^{-7}	5.10×10^{-7}	2.94×10^{-7}	1.54×10^{-4}
(DC7) DF8 + H_2	1.22×10^{-9}	5.98×10^{-11}	5.52×10^{-8}	3.19×10^{-8}	1.19×10^{-9}

[a] The total energy was set to 500.0 kJ/mol relative to EQ_a1.

[b] Initial population 1.0 was given to the corresponding EQ.

To further compare the results obtained by the two different approaches, the four additional cases where the initial population 1.0 was given to EQ_a2, EQ_a3, EQ_a4, or EQ_a5 were considered. The setup of the RK4 method was identical to that adopted in the last case where the RK4 time step was set to $0.1/k_{\text{largest}}$ s and the time evolution was performed until the sum of populations of all EQs becomes less than 10^{-20}. The obtained branching ratios are listed in Table 4.2. Values obtained by the two different approaches again agreed

rigorously up to thirteen decimal places, and these results further supported reliability of the *f*-RCMC method.

The author further considered the approach that uses the general solution of linear differential equations. The procedure is based on matrix algebra[23]. The first-order rate equation is represented as eq. (4.7). This equation can be rewritten as,

$$\frac{d\mathbf{A}}{dt} = \mathbf{K}\mathbf{A} \qquad (4.7)$$

where **A** is the population vector, and **K** is the rate constant matrix. When **A** contains *n* elements, **K** is an *n*×*n* matrix. In many cases, **K** is the asymmetric matrix. If **K** is diagonalizable and there are *n* eigenvectors, the solution of eq. (4.7) can be expressed as eq. (4.8) with eigenvalue λ_n and eigenvector-matrix **P**.

$$A(t) = \mathbf{P}\exp\left(t \cdot diag\left(\lambda_1 \cdots \lambda_n\right)\right)\mathbf{P}^{-1}\mathbf{A}(0) \qquad (4.8)$$

Thus, the branching ratio of unimolecular decomposition, in principle, can be obtained by taking the limit of *t* to infinity. However, the rate constant matrix of C_3H_5 has elements with a wide range of different values (10^7–10^{13} s^{-1}). Moreover, in many chemical reactions, the range becomes even larger [25]. The eigenvalues of an asymmetric matrix are very sensitive to the difference in matrix elements and suffers seriously from the rounding error [24]. Thus, it is not easy to obtain eigenvalues of **K** accurately with a finite-precision arithmetic algorithm. In fact, the author applied the subroutine "DGEEV" in LAPACK, a standard open-source library, to the rate constant matrix of C_3H_5. However, the solution of eq. (4.8) converged to zero as *t* went to infinity. This was because there was no zero-eigenvalue, where one zero-eigenvalue should appear when the rate constant matrix is diagonalized.

As seen in Table 2, the branching ratios varied depending on the initial condition.

When the initial population was given to EQ_a1, DC1 which gave H + $H_2C{=}C{=}CH_2$ was the major channel. The major channel changed to DC2 (H_2CC + CH_3) when the initial population was given to EQ_a2. This was because the direct dissociation rather than isomerization to EQ_a1 was preferred from EQ_a2. Similarly, the direct dissociation DC4 was the major channel when the population was given to EQ_a3 or EQ_a4. In the case where the population was given to EQ_a5, DC1 was the major channel. This was because DC1 from EQ_a5 also corresponded to a direct dissociation in which dissociation of the H atom and the ring-opening occurred through a single IRC path.

The author subsequently applied the current method to the same reaction at a different total energy. The total energy was set to 293.4 kJ/mol, where this value is 100 kJ/mol higher than the total energy of DF2 and DF5. The branching ratios were calculated for the above five cases where initial population was given to one of the five EQs. The setup of the RK4 method was determined in the same way adopted in the above examples. The obtained branching ratios are shown in Table 3, where the results obtained by the *f*-RCMC reproduced those by the kinetic simulations precisely up to eight decimal places. When the total energy was set to the small value, the branching ratios of DC2, 5, 6, and 7 were zero. This was because TSs leading to these DCs were higher than the total energy and were not accessible. These results confirmed that the *f*-RCMC was highly accurate at this total energy.

Table 4.3 The branching ratios of DC1-7 in unimolecular decomposition of C_3H_5 with the total energy 293.4 kJ/mol relative to EQ_a1.[a]

DCs	EQ_a1[b]	EQ_a2[b]	EQ_a3[b]	EQ_a4[b]	EQ_a5[b]
(DC1) DF1 + H	9.68×10^{-1}	1.61×10^{-1}	2.26×10^{-4}	2.25×10^{-4}	9.68×10^{-1}
(DC2) DF2 + DF3	0.0	0.0	0.0	0.0	0.0
(DC3) DF4 + H	2.42×10^{-2}	8.39×10^{-1}	4.75×10^{-2}	4.73×10^{-2}	2.42×10^{-2}
(DC4) DF2 + DF5	8.02×10^{-3}	1.41×10^{-5}	9.52×10^{-1}	9.52×10^{-1}	8.02×10^{-3}
(DC5) DF6 + H	0.0	0.0	0.0	0.0	0.0
(DC6) DF7 + H_2	0.0	0.0	0.0	0.0	0.0
(DC7) DF8 + H_2	0.0	0.0	0.0	0.0	0.0

[a] The total energy was set to 293.4 kJ/mol relative to EQ_a1.

[b] Initial population 1.0 was given to the corresponding EQ.

4.5 Application2: Decomposition of C_4H_5

The author further applied the *f*-RCMC method to unimolecular decomposition of C_4H_5. The reaction profile at the UCCSD(T)/def2-TZVP//B3LYP/def2-SV(P) level was obtained using the same way adopted in the C_3H_5 case. The search found 69 EQs, and the 10 most stable EQs (EQ_b1-10) are listed in Fig. 4.1. The relative energies including ZPE corrections (in kJ/mol) are also shown in Figure 1. In addition, 27 DCs (DC8-34) as listed in Table 1 were obtained. DFs for these DCs are found in Fig. 4.1. The rate constants for all obtained elementary steps were estimated using eq. (2.4).

Table 4.4 shows the unimolecular decomposition branching ratios for cases where the total energy was set to 500 kJ/mol relative to EQ_b1 and the initial population 1.0 was given to one of the ten most stable EQs. The branching ratios of top four DCs are listed in Table 4.4. The branching ratios obtained by the *f*-RCMC method agreed precisely up to nine decimal places with those obtained by the ordinary kinetic simulation, where the setup of the RK4 simulation was identical to that adopted in the application to C_3H_5. Table 4.4 shows that DC8 and DC10 giving CH_2CHCCH + H or CH_2CH + HCCH,

respectively, are preferred, and their preference changes depending on the initial population. Note that the kinetic simulation in which the initial population was given to EQ_b1 took 95 hours. In contrast, the *f*-RCMC gave the branching ratios in less than 1 second. On the other hand, the approach that uses the general solution of linear differential equations again did not give reasonable results.

Table 4.4 The branching ratios of top four DCs in unimolecular decomposition of C_4H_5 with the total energy 500.0 kJ/mol relative to EQ_b1.[a]

EQ_b1[b]	EQ_b2[b]	EQ_b3[b]	EQ_b4[b]	EQ_b5[b]
(DC8) 9.15×10^{-1}	(DC8) 9.16×10^{-1}	(DC8) 9.89×10^{-1}	(DC10) 8.16×10^{-1}	(DC10) 8.23×10^{-1}
(DC9) 6.28×10^{-2}	(DC9) 6.30×10^{-2}	(DC10) 9.79×10^{-3}	(DC8) 1.81×10^{-1}	(DC8) 1.77×10^{-1}
(DC10) 1.40×10^{-2}	(DC10) 1.36×10^{-2}	(DC16) 1.29×10^{-3}	(DC21) 1.97×10^{-3}	(DC9) 6.54×10^{-4}
(DC11) 6.47×10^{-3}	(DC11) 6.49×10^{-3}	(DC17) 8.89×10^{-5}	(DC9) 8.17×10^{-4}	(DC11) 6.92×10^{-5}
EQ_b6[b]	EQ_b7[b]	EQ_b8[b]	EQ_b9[b]	EQ_b10[b]
(DC10) 8.78×10^{-1}	(DC8) 6.37×10^{-1}	(DC10) 8.04×10^{-1}	(DC8) 6.26×10^{-1}	(DC10) 8.63×10^{-1}
(DC8) 1.22×10^{-1}	(DC10) 3.61×10^{-1}	(DC8) 1.95×10^{-1}	(DC10) 3.72×10^{-1}	(DC8) 1.36×10^{-1}
(DC9) 5.39×10^{-4}	(DC19) 1.13×10^{-3}	(DC9) 7.95×10^{-4}	(DC19) 1.11×10^{-3}	(DC9) 8.16×10^{-4}
(DC11) 5.67×10^{-5}	(DC9) 3.61×10^{-4}	(DC11) 1.72×10^{-5}	(DC9) 3.63×10^{-4}	(DC11) 8.51×10^{-5}

[a] The total energy was set to 500.0 kJ/mol relative to EQ_b1.

[b] Initial population 1.0 was given to the corresponding EQ.

The author further presents a case with a small total energy, where the value was set to the sum of energies of DF5 and DF11 plus 100 kJ/mol (309.1 kJ/mol). The branching ratios obtained for this condition are shown in Table 4.5. In this case, unfortunately, I was not able to conduct the ordinary kinetic simulation with the criteria adopted in the above cases. A time evolution until 4.86×10^{-7} s by the RK4 method with the step size $1/k_{\text{largest}}$ (4.86×10^{-15} s in this case) took ~4 days in the case where the initial population 1.0 was given to EQ_b1, and the population of EQ_b1 did not become less than

0.2 even after such a long simulation. the author therefore abandoned to obtain converged branching ratios with the kinetic simulation. In contrast, the *f*-RCMC gave the branching ratios in less than 1 second. This clearly shows the advantage of the *f*-RCMC method over the ordinary kinetic simulation approach.

Table 4.5. The branching ratios of top four DCs in unimolecular decomposition of C_4H_5 with the total energy 309.1 kJ/mol relative to EQ_b1.[a]

EQ_b1[b]	EQ_b2[b]	EQ_b3[b]	EQ_b4[b]	EQ_b5[b]
(DC8) 9.96×10^{-1}	(DC8) 9.96×10^{-1}	(DC8) 1.00×10^{0}	(DC10) 7.31×10^{-1}	(DC10) 7.31×10^{-1}
(DC9) 2.52×10^{-3}	(DC9) 2.52×10^{-3}	(DC10) 3.31×10^{-7}	(DC8) 2.69×10^{-1}	(DC8) 2.69×10^{-1}
(DC10) 1.26×10^{-3}	(DC10) 1.26×10^{-3}	(DC9) 2.55×10^{-7}	(DC9) 1.39×10^{-5}	(DC9) 1.39×10^{-5}
(DC12) 1.68×10^{-7}	(DC12) 1.68×10^{-7}	(DC12) 1.70×10^{-11}	(DC12) 9.26×10^{-10}	(DC12) 9.25×10^{-10}
EQ_b6[b]	EQ_b7[b]	EQ_b8[b]	EQ_b9[b]	EQ_b10[b]
(DC10) 7.31×10^{-1}	(DC8) 1.00×10^{0}	(DC10) 7.31×10^{-1}	(DC8) 1.00×10^{0}	(DC10) 7.31×10^{-1}
(DC8) 2.69×10^{-1}	(DC9) 1.31×10^{-5}	(DC8) 2.69×10^{-1}	(DC9) 1.31×10^{-5}	(DC8) 2.69×10^{-1}
(DC9) 1.39×10^{-5}	(DC10) 1.18×10^{-5}	(DC9) 1.39×10^{-5}	(DC10) 1.18×10^{-5}	(DC9) 1.39×10^{-5}
(DC12) 9.25×10^{-10}	(DC12) 8.71×10^{-10}	(DC12) 9.26×10^{-10}	(DC12) 8.71×10^{-10}	(DC12) 9.26×10^{-10}

[a] The total energy was set to 309.1 kJ/mol relative to EQ_b1.

[b] Initial population 1.0 was given to the corresponding EQ.

4.6 Derivation of eq. (4.5)

To derive eq. (4.5), the author further introduces the following two assumptions.

A. $G_{l-m}^{(n)}$ changes only when the QSSA is applied to EQ i ($l \neq i$, $m \neq i$) in the step 2, where $G_{l-m}^{(n)}$ is the sum of states at the TS between state l and state m.

B. $N_m^{(n)}$ changes only when redistribution of $N_i^{(n)}$ is made in the step 4.

With the assumptions A and B, $k'^{(n+1)}_{k\to l}$ of eq. (4.1) can be expressed as follows.

$$k'^{(n+1)}_{k\to l} = \frac{G^{(n+1)}_{k-l}}{hN^{(n)}_k} \qquad (4.9)$$

After application of the procedure in the step 2, G has already been changed by the QSSA, but redistribution of N has not yet been made. On the other hand, $k^{(n+1)}_{k\to l}$ of eq. (4.5) can be given as,

$$k^{(n+1)}_{k\to l} = \frac{G^{(n+1)}_{k-l}}{hN^{(n+1)}_k} \qquad (4.10).$$

After application of the procedure in the step 6, both G and N have already been changed in the step 2 and in the step 4, respectively.

Then, by substituting eq. (2.4) into eq. (4.1), the following relation is obtained.

$$k'^{(n+1)}_{k\to l} = k^{(n)}_{k\to l} + k^{(n)}_{k\to i}k^{(n)}_{i\to l}\sigma^{(n)}_i = \frac{1}{hN^{(n)}_k}\left(G^{(n)}_{k-l} + G^{(n)}_{k-i}k^{(n)}_{i\to l}\sigma^{(n)}_i\right) \qquad (4.11)$$

By comparing eqs. (4.9) and (4.11), the following relation is obtained.

$$G^{(n+1)}_{k-l} = G^{(n)}_{k-l} + G^{(n)}_{k-i}k^{(n)}_{i\to l}\sigma^{(n)}_i \qquad (4.12)$$

The following equation is obtained by substituting eq. (4.12) into eq. (4.10).

$$k^{(n+1)}_{k\to l} = \frac{G^{(n)}_{k-l}}{hN^{(n+1)}_k} + \frac{G^{(n)}_{k-i}}{hN^{(n+1)}_k}k^{(n)}_{i\to l}\sigma^{(n)}_i \qquad (4.13)$$

By substituting eq. (4.3) into the first term in the right-hand side of eq. (4.13), the following relation is obtained.

$$\frac{G^{(n)}_{k-l}}{hN^{(n+1)}_k} = \frac{G^{(n)}_{k-l}}{hN^{(n)}_k + hk^{(n)}_{i\to k}\sigma^{(n)}_i N^{(n)}_i} \qquad (4.14)$$

By replacing $k^{(n)}_{i\to k}$ in eq. (4.14) by the expression of eq. (2.4), eq. (4.15) is obtained.

$$\frac{G^{(n)}_{k-l}}{hN^{(n+1)}_k} = \frac{1}{\frac{hN^{(n)}_k}{G^{(n)}_{k-l}} + \frac{G^{(n)}_{i-k}}{G^{(n)}_{k-l}}\sigma^{(n)}_i} \qquad (4.15)$$

Further rearrangements of eq. (4.15) gives,

$$\frac{G_{k-l}^{(n)}}{hN_k^{(n+1)}} = \frac{1}{\frac{hN_k^{(n)}}{G_{k-l}^{(n)}} + \frac{hN_k^{(n)}}{G_{k-l}^{(n)}} \frac{G_{i-k}^{(n)}}{hN_k^{(n)}} \sigma_i^{(n)}} \qquad (4.16).$$

By replacing G and N in the right-hand side of eq. (4.16) with the expression of rate constant in eq. (2.4), the following relation is obtained.

$$\frac{G_{k-l}^{(n)}}{hN_k^{(n+1)}} = \frac{1}{\frac{1}{k_{k\to l}^{(n)}} + \frac{k_{k\to i}^{(n)}}{k_{k\to l}^{(n)}} \sigma_i^{(n)}} \qquad (4.17)$$

eq. (4.17) is further rearranged to,

$$\frac{G_{k-l}^{(n)}}{hN_k^{(n+1)}} = \frac{k_{k\to l}^{(n)}}{1 + k_{k\to i}^{(n)} \sigma_i^{(n)}} \qquad (4.18).$$

By substituting eq. (4.3) into the second term in the right-hand side of eq. (4.13), the following relation is obtained.

$$\frac{G_{k-i}^{(n)}}{hN_k^{(n+1)}} k_{i\to l}^{(n)} \sigma_i^{(n)} = \frac{G_{k-i}^{(n)}}{hN_k^{(n)} + hk_{i\to k}^{(n)} \sigma_i^{(n)} N_i^{(n)}} k_{i\to l}^{(n)} \sigma_i^{(n)} \qquad (4.19)$$

By replacing $k_{i\to k}^{(n)}$ in eq. (4.19) by the expression of eq. (2.4), eq. (4.20) is obtained.

$$\frac{G_{k-i}^{(n)}}{hN_k^{(n+1)}} k_{i\to l}^{(n)} \sigma_i^{(n)} = \frac{1}{\frac{hN_k^{(n)}}{G_{k-i}^{(n)}} + \sigma_i^{(n)}} k_{i\to l}^{(n)} \sigma_i^{(n)} \qquad (4.20)$$

By replacing G and N in the right-hand side of eq. (4.20) with the expression of rate constant in eq. (2.4), the following relation is obtained.

$$\frac{G_{k-i}^{(n)}}{hN_k^{(n+1)}} k_{i\to l}^{(n)} \sigma_i^{(n)} = \frac{k_{k\to i}^{(n)}}{1 + k_{k\to i}^{(n)} \sigma_i^{(n)}} k_{i\to l}^{(n)} \sigma_i^{(n)} \qquad (4.21)$$

By substituting eqs. (4.18) and (4.21) into eq. (4.13), the following equation is obtained.

$$k_{k\to l}^{(n+1)} = \frac{1}{1 + k_{k\to i}^{(n)} \sigma_i^{(n)}} \left(k_{k\to l}^{(n)} + k_{k\to i}^{(n)} k_{i\to l}^{(n)} \sigma_i^{(n)} \right) \qquad (4.22)$$

This eq. (4.22) is identical to eq. (4.5).

4.7 Conclusion

In Chapter 4, the author introduced the *f*-RCMC method as an extension of the RCMC method. This method allows for obtaining the branching ratio of unimolecular decomposition without performing the kinetic simulation. This method assumes that the dissociated fragments do not undergo secondary reactions, and is thus effective only for unimolecular decomposition in the gas-phase. The method was numerically tested with unimolecular decomposition of C_3H_5 and C_4H_5, where their reaction profiles were generated at UCCSD(T)/def2-TZVP//B3LYP/def2-SV(P) level using the SC-AFIR method. The branching ratios obtained by the *f*-RCMC method agreed precisely up to eight decimal places with those obtained by the conventional kinetic simulation. The *f*-RCMC method requires only the rate constant matrix as input and gives the branching ratios immediately without solving the rate equations. It is therefore expected that the *f*-RCMC method will be a highly efficient alternative to the ordinary kinetic simulation in predicting the branching ratios of unimolecular decomposition.

References

[1] R. Liu, K. Morokuma, A. M. Mebel, M. C. Lin, *J. Phys. Chem.* **1996**, *100*, 9314.

[2] D. Chakraborty, R. P. Muller, S. Dasgupta, W. A. Goddard, *J. Phys. Chem. A* **2000**, *104*, 2261.

[3] A. M. Mebel, M. C. Lin, D. Chakraborty, J. Park, S. H. Lin, Y. L. Lee, *J. Chem. Phys.* **2001**, *114*, 8421.

[4] J. L. Miller, M. J. Krisch, L. J. Butler, J. Shu, *J. Phys. Chem. A* **2005**, *109*, 4038.

[5] A. Lamdera, S. P. Krishtal, V. V. Kislov, A. M. Mebel, R. I. Kaiser, *J. Chem. Phys.* **2008**, *128*, 214301.

[6] X. Gu, R. I. Kaiser, A. M. Mebel, *Chem. Phys. Chem.* **2008**, *9*, 350.

[7] W. Chen, W. L. Hase, H. B. Schlegel, *Chem. Phys. Lett.* **1994**, *228*, 436.

[8] X. Li, J. M. Millam, H. B. Schlegel, *J. Chem. Phys.* **2000**, *113*, 10062.

[9] W. L. Yim, Z. F. Liu, *J. Am. Chem. Soc.* **2001**, *123*, 2243.

[10] M. A. Collins, *Theor. Chem. Acc*. **2002**, *108*, 313.

[11] O. Isayev, L. Gorb, M. Qasim, J. Leszczynski, *J. Phys. Chem. B* **2008**, *112*, 11005.

[12] J. M. Bowman, B. C. Shepler, *Ann. Rev. Phys. Chem.* **2011**, *62*, 531.

[13] S. Grimme, *Angew. Chem. Int. Ed*. **2013**, *52*, 6306.

[14] M. Paranjothy, R. Sun, Y. Zhuang, W. L. Hase, *WIREs Comput. Mol. Sci.* **2013**, *3*, 296.

[15] L. Drahos, K. Vékey, *J. Mass Spectrom.* **2001**, *36*, 237.

[16] H. Y. Lee, V. V. Kislov, S. H. Lin, A. M. Mebel, D. M. Neumark, *Chem. Eur. J.* **2003**, *9*, 726.

[17] J. A. Miller, S. J. Klippenstein, *J. Phys. Chem*. A **2003**, *107*, 2680.

[18] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, D. G. Truhlar, *Chem. Rev.* **2006**,

106, 4518.

[19] J. A. Miller, S. J. Klippenstein, *J. Phys. Chem. A* **2006**, *110*, 10528.

[20] D. S. N. Parker, S. Maity, B. B. Dangi, R. I. Kaiser, A. Landera, A. M. Mebel, *Phys. Chem. Chem. Phys.* **2014**, *16*, 12150.

[21] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int. J. Quantum Chem.* **2015**, *115*, 258.

[22] R. J. Gelinas, *J. Comput. Phys.* **1972**, *9*, 222.

[23] J. I. Steinfeld, J. S. Francisco, W. L. Hase, Chemical Kinetics and Dynamics, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, **1999**.

[24] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Chap. 11, 2nd ed., Cambridge Univ. Press, Cambridge **1992**; pp. 456-495.

[25] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T, Taketsugu, S. Maeda, *J. Phys. Chem. A* **2015**, *119*, 11641.

[26] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec*. **2016**, *16*, 2232.

[27] G. Z. Whitten, B. S. Rabinovitch, *J. Chem. Phys.* **1963**, *38*, 2466.

[28] A. J. Karas, R. G. Gilbert, M. A. Collins, *Chem. Phys. Lett.* **1992**, *193*, 181.

[29] O. M. Becker, M. Karplus. *J. Chem. Phys.* **1997**, *106*, 1495.

[30] J. P. Doye, D. J. Wales, R. S. Berry, *J. Chem. Phys.* **1995**, *103*, 4234.

[31] D. J. Wales, Energy Landscapes: Applications to Clusters, Biomolecules and Glasses; Cambridge University Press: Cambridge, UK, **2003**.

[32] T. Turanyi, A. S. Tomlin, M. J. Piling, *J. Phys. Chem*. **1993**, *97*, 163.

[33] T. Turanyi, A. S. Tomlin: Analysis of Kinetic Reaction Mechanism; Springer, Berlin, **2014**; pp. 231-242.

[34] H. M. Frey, R. Walsh, *Chem. Rev.* **1969**, *69*, 103.

[35] W. Tsang, *J. Am. Chem. Soc.* **1985**, *107*, 2872.

[36] V. D. Knyazev, I. R. Slagle, *J. Phys. Chem.* **1996**, *100*, 5318.

[37] J. Yu, S. Eser, *Ind. Eng. Chem. Res.* **1997**, *36*, 585.

[38] P. E. Savage, *J. Anal. Appl. Pyrolysis*, **2000**, *54*, 109.

[39] A. Miyoshi, J. Widjaja, N. Yamauchi, M. Koshi, H. Matsui, *Proc. Combust. Inst.* **2002**, *29*, 1285.

[40] J. B. A. Mitchell, C. Rebrion-Rowe, J. L. Le Garrec, G. Angelova, H. Bluhme, K. Seiersen, L.H. Andersen *J. Mass Spectrom.* **2003**, *227*, 273.

[41] G. Angelova, O. Novotny, J. B. A. Mitchell, C. Rebrion-Rowe, J. L. Le Garrec, H. Bluhme, A. Svendsen, L. H. Andersen, *J. Mass Spectrom.* **2004**, *235*, 7.

[42] P. Zámostný, Z. Bělohlav, L. Starkbaumová, J. Patera, *J. Anal. Appl. Pyrolysis* **2010**, *87*, 207.

[43] A. C. Davis, J. S. Francisco, *J. Phys. Chem. A* **2011**, *115*, 2966.

[44] M. N. Ryazantsev, A. Jamal, S. Maeda, K. Morokuma, *Phys. Chem. Chem. Phys.* **2015**, *17*, 27789.

[45] S. Maeda, Y. Harabuchi, Y. Sumiya, M. Takagi, M. Hatanaka, Y. Osada, T. Taketsugu, K. Morokuma, K. Ohno, GRRM program package, a developmental version ed., Hokkaido University, 2016, see http://grrm.chem.tohoku.ac.jp/GRRM/index_e.html.

[46] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., 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, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09, Revision D.01, Gaussian Inc., Wallingford CT, 2013.

[47] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Chap. 16, 2nd ed., Cambridge Univ. Press, Cambridge **1992**; pp. 710-714.

Chapter 5 Background of Automated Reaction Path Search

5.1 Introduction

Programs such as density functional theory have been developed and quantum chemical calculation is becoming a powerful tool in chemical research. To predict chemical reaction mechanisms, systematical search for equilibrium structure (EQs) and transition states (TSs) on potential energy surface (PES) is required. Searching on PES is not easy since PES is a multidimensional space. Thus, the search area must be limited to the periphery of the reaction paths in order to reduce computational time.

Searching for single EQ can always be done by geometry optimization starting from any initial structure. However, all EQs cannot be obtained by repeating this method. The same EQ is often obtained even if starting from different initial structures. This calculation spends a lot of computational time for nothing.

In the case of TS search, finding even only single one is not easy. Although there are several methods for searching TS near initial structures[1,2], the convergence radius is generally not large. TS cannot be obtained from an arbitrary point. When a reactant and a product are known, methods for finding TS by searching in the direction in which the energy of the highest point of the path connecting them is lowered have also been developed[3-5]. However, it is difficult to search a reaction path connected with unknown EQs by such methods.

In order to overcome these problems, new methods have been developed based on the characteristics of reaction pathway connecting EQ and TS. In addition to mathematical features such as minimum point and saddle point, EQ and TS are connected by a reaction pathway. If one can trace reaction pathways, reaction path networks can be obtained efficiently without unnecessary calculation. The history of the development of

systematic search method for reaction pathways is the history of exploring the characteristics of reaction pathways, and many methods based on various features were developed (Section 5.2). These are introduced in the next section. Specifically, the author briefly introduces eigenvector following (EF) method[6], sphere optimization (SO) method[7], gradient extremal following (GEF) method[6], reduced gradient following (RGF) method[8], and finally summarize anharmonic downward distortion following (ADDF)[9-11] and artificial force induced reaction (AFIR)[12-15].

5.2 History of Automated Reaction Path Search

In this section, the author introduces various methods for reaction path search using the features of reaction path. As mathematical properties, there are important differences between EQ and TS. All the eigenvalues of Hessian in EQ are positive whereas only one eigenvalue of Hessian in TS is negative and the others are positive. The eigenvector of negative eigenvalue in TS represents the direction of reaction path, along which the IRC path grows.

Eigenvector following (EF) method and sphere optimization (SO) technique use these features. The EF method searches a stationary point by tracing toward eigenvector direction of normal mode analyses[6]. The SO technique involves a sequence of constrained optimizations on hypersurfaces with increasingly larger radii[7]. Although these methods worked well and sometimes found TS starting from an EQ, they often searched for pathways far apart from the actual reaction pathways.

Next, gradient extremal following (GEF) method was proposed. The GEF method tracing the gradient extremal point on the constant energy constraint from an EQ is another surface-walking method[6]. The GEF has been applied to the ab initio PES of

formaldehyde (HCHO) at the HF/STO-3G level and the first systematic global analysis of ab initio PES has been made for a system of more than three atoms in 1996[16]. In the search, adjacent EQ and TS however may not be connected directly, sometimes passing TS or EQ different from the actual reaction paths. Since the GEF method also searches high order stationary points in addition to EQ and TS, unnecessary search is performed. For these reasons, there is no application to global search of a system with five atoms or more. In 1998, a reduced gradient following (RGF) technique was applied to HCHO at the level of HF/STO-3G[8]. However, since the RGF technique also searches unnecessary paths as in the GEF method, global search of a system with five atoms or more was not reported.

The features used in the above methods are gradient and Hessian, which are first derivative and second derivative on PES, respectively. It is difficult to apply gradient to search because gradient is all zero in every direction at EQ. Hessian also represents the restoring force to bring back to the stationary point, while low vibration modes hardly show the directions of reaction pathways. Therefore, in order to search reaction paths using mathematical features around EQs, the anharmonicity of PES involving third or higher terms should be noted[9-11].

In such a situation, anharmonic downward distortion following (ADDF) method was developed[9-11]. The ADDF utilizes the fact that real potential is distorted below harmonic potentials. The ADDF was developed based on the hypothesis that there are reaction paths in the direction of large anharmonic downward distortion (ADD). This hypothesis captures the characteristics of chemical reactions well, and isomerization reaction paths (A → X type) and dissociation reaction paths (A → X + Y type) could be obtained systematically from any initial structure. In 2006, the global reaction network of

HCOOH, which is a reaction system with five atoms, was reported using the ADDF[11]. In subsequent applications, the reaction networks of various systems has been obtained[14,17].

In further development, artificial force induced reaction (AFIR) method was reported in 2011. While the types of reaction pathways obtained by the ADDF is the isomerization (A → X type) and the dissociation (A → X + Y type), the AFIR can also obtain convergent synthetic reaction pathways (A + B → X type) and recombination pathways (A + B → X + Y type). The AFIR has been applied most extensively to organic reaction and organometallic catalysis, in combination with quantum chemical calculation[18-27]. The other applications are: gas phase reactions[28], enzymatic catalysis[29]; domino reactions[30]; and metal cluster catalysis[31,32]. In Chapter 3, 4, 6, 7, and 8, the AFIR is used as an automated reaction path search method. The details of the AFIR are introduced in the next section.

5.3 Artificial Force Induced Reaction Method

The AFIR method induces structural deformations in a molecule by applying artificial force between fragments and finds the pathway for the corresponding reaction. The idea of AFIR is simple; just push fragments A and B together or pull them apart. This procedure corresponds to minimize the following AFIR function[33]:

$$F(\mathbf{Q}) = E(\mathbf{Q}) + \rho\alpha \frac{\sum_{i\epsilon A}\sum_{j\epsilon B}\omega_{ij}r_{ij}}{\sum_{i\epsilon A}\sum_{j\epsilon B}\omega_{ij}} \qquad (5.1)$$

This function consists of two terms, i.e., the Born-Oppenheimer potential energy surface $E(\mathbf{Q})$ of geometrical parameters $\mathbf{Q}$ and the artificial force term. The parameter in the artificial term α determines the strength of the force. The coefficient ρ is either 1 to

push together, or -1 to pull them apart. The force term is given as a weighted sum of the distance r_{ij} between atom i in the fragment A and atom j in the fragment B, and the weight function ω_{ij} is:

$$\omega_{ij} = \left(\frac{(R_i+R_j)}{r_{ij}}\right)^p \qquad (5.2)$$

This weight function assigns a stronger force to the closer atom pairs and a weaker force to the more distant pairs. In eq. (5.2), the inverse distance $1/r_{ij}$ is scaled by $R_i + R_j$, the sum of covalent radii of atom i and j, to treat all elements equivalently. It was confirmed that results did not strongly depend on the choice of p, and p is usually set to 6.0.

For convenience, the parameter α is determined by the following equation:

$$\alpha = \frac{\gamma}{\left(2^{-1/6} - \left(1+\sqrt{1+\gamma/\varepsilon}\right)^{-1/6}\right)R_0} \qquad (5.3)$$

This α corresponds to the mean force that acts on two atoms in their direct collision on the Lennard-Jones (LJ) potential with collision energy γ, in the area from the minimum to the turning point. The standard Ar-Ar parameters of the LJ potential, i.e., R_0 = 3.8164 Å and ε = 1.0061 kJ/mol, were employed. The model collision energy parameter γ defines an approximate upper limit of the barrier height that can be eliminated by the force term. The γ parameter can be chosen by users, depending approximately on the highest TS energies searched. By applying the AFIR to obtained EQs one after another, a reaction path network can be obtained.

In AFIR, there are two kinds of algorithms. The first is a multicomponent algorithm that defines two or more reactants as fragments and reacts them from various directions. When third reactant C exists, two additional terms between A and C and between B and C are added to eq. 5.1. The same is true in the case of four or more. The second is single-component (SC) algorithm that defines various fragments systematically

in a single molecule and applies AFIR to all combinations of them. The SC-AFIR is applicable to reactions involving multiple reactants by treating the reactant complex as a single element.

The artificial force term in eq. 5.1 is designed not to change the reactivity of PES. This term adds an isotropic artificial force to each atom. Fig. 5.1 (a) shows the interaction potential contour map between H atom and CO_2 molecule at the B3LYP/cc-pVTZ level, where the positions of the C and O atoms in CO_2 were fixed at those of optimized CO_2[13]. In Fig. 5.1 (a), all directions are almost completely repulsive. However, the dents in the directions indicated by arrow are seen. In these directions, H atom can approach closer to CO_2 molecule.

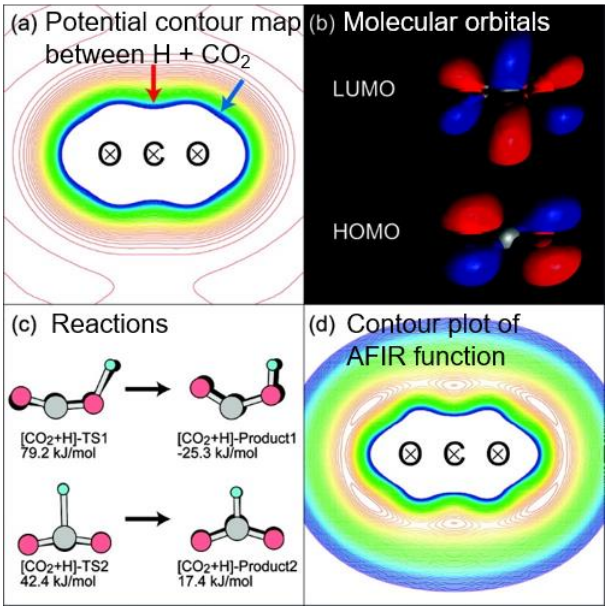

Fig. 5.1 (a) the interaction potential contour map between H atom and CO_2 molecule. (b) the interactions between 1s orbital of the H atom and LUMOs of CO_2; 1s orbital of the H atom and HOMOs of CO_2. (c) these reactions taking place when the H atom collides toward the dents of (a). (d) A Contour plot of the AFIR function with γ = 200 kJ/mol.

The perpendicular direction should be one reactive site, where 1s orbital of the H atom and the lowest unoccupied molecular orbitals (LUMOs) of CO_2 in Fig. 5.1(b) interacts. The diagonal direction should be another reactive site because of interaction between 1s orbital of the H atom and the highest occupied molecular orbitals (HOMOs) of CO_2 in Fig. 5.1(b). The reaction has been found to take place in Fig. 5.1(c) when the H atom collides toward either of these dents. If these directions can be detected, one can obtain reaction paths. The reactive directions can be found easily on the AFIR function (eq. 5.1). That is, if the AFIR function is minimized from any points, two dents are obtained. This is because the second term of eq. 5.1 highlights sighs of reactivity appearing in PES, that is, dents of contour. The AFIR function has a landscape shown in Fig. 5.1(d), which is a contour plot of the AFIR function with γ = 200kJ/mol (the second term of eq. 5.1 is added to the plot of Fig. 5.1(a)). This allows to obtain paths leading to products without deviating considerably from the reaction paths despite performing a rough manipulation that adds artificial force.

Fig. 5.1(c) compares approximate AFIR structures with true TSs and true products, where AFIR structures obtained by γ = 200kJ/mol are shown in black behind the true stationary structures. The AFIR structures, especially the approximate TS structures, are quite similar to the true ones.

The development of the reaction path search algorithm enables us to obtain complicated reaction networks. However, computational complexity for searching all reaction paths increases sharply as the number of atoms in the system increases. In the next section, it is explained how reaction network increases.

5.4 Explosion of the Size of Reaction Path Network

Although recent advances in reaction path search algorithms have made it possible to obtain various reaction path networks, the computational complexity of the networks increases sharply as the number of atoms in the system increases. For example, molecules in which N different atoms are arranged in a chain have $N!/2$ EQs. When the target molecule contains same atoms with n, m, and k pairs, the number of chain structures becomes $N!/(n!m!k!)$[34]. Then, although the number is considerably less than the original number, the number of EQs becomes very large if N increases. The number of TSs is even larger than EQs.

Wales et al. also reported that the number of EQ and TS increases exponentially according to the increase of N[35]. In the paper by Wales et al., the number of EQ and TS were estimated using Lennard-Jones (LJ) cluster (Table 5.1). In ordinary chemical reactions, the number of EQ and TS for each N would be larger than the values in Table 5.1. Thus, obtaining all structures is not easy and it is necessary to reduce the computational cost for quantum chemical calculation.

Table 5.1 The number of EQ and TS in LJ_N clusters[35].

	N									
	4	5	6	7	8	9	10	11	12	13
EQ	1	1	2	4	8	21	64	170	513	1505
TS	1	2	3	12	42	165	635	2423	7868	25653

In addition, the following feature can be utilized: Paths of reaction bottlenecks have a great influence on the kinetics, while other paths have a small influence. For

example, the computational cost can be greatly reduced if the accurate optimization calculation is applied to the TSs and paths of reaction bottlenecks and other paths are supplemented by a fast method. Moreover, limiting the search area according to required information would also be effective. Some EQs and TSs obtained by global search has extremely high energy. Such structures are not important under normal experimental conditions. In Chapter 6 and 7, such approaches are introduced, respectively.

5.5 Conclusion

In this chapter, the author introduced the difficulty for searching EQ and TS on PES. Reaction path search methods have been developed using the characteristics of reaction pathways. In Section 5.2, the history of the development of reaction path search methods was explained with some examples. The development of automated reaction path search method had made significant progress in 2006, and the global reaction route map of HCOOH using ADDF was reported. After that, AFIR had been developed and applied to more complex and diverse chemical reactions. In Section 5.3, the nature of AFIR was explained with the PES of H atom and CO_2 molecule. In Section 5.4, the author introduced the problem that reaction path networks become very large as the number of atoms in the system increases. The efficiency of reaction path search can be improved by using fast quantum calculation method or by parallel calculation. To calculate bottlenecks of reaction with high accuracy is also effective. To search only important parts under experimental conditions by limiting the search area is also effective.

References

[1] A. Banerjee, N. Adams, J. Simons, R. Shepard, *J. Phys. Chem*. **1985**, *89*, 52.

[2] P. Csazar, P. Pulay, *J. Mol. Struct*. **1984**, *114*, 31.

[3] T. A. Hlgren, W. N. Lipscomb, *Chem. Phys. Lett*. **1977**, *49*, 225.

[4] K. Muller, L. D. Brown, *Theor. Chim. Acta*. **1979**, *53*, 75.

[5] C. Choi. R. Elder, *J. Chem. Phys*. **1991**, *94*, 751.

[6] J. Pancir, *Collect. Czech. Chem. Commun*. **1975**, *40*, 1112.

[7] Y. Abashkin, N. Russo, *J. Chem. Phys*. **1994**, *100*, 4477.

[8] W. Quapp, M. Hirsch, O. Imig, D. Heidrich, *J. Comput. Chem*. **1998**, *19*, 1087.

[9] K. Ohno, S. Maeda, *Chem. Phys. Lett*. **2004**, *384*, 277.

[10] S. Maeda, K. Ohno, *J. Phys. Chem*. A, **2005**, *109*, 5742.

[11] K. Ohno, S. Maeda, *J. Phys. Chem*. A, **2006**, *110*, 8933.

[12] S. Maeda, K. Morokuma, *J. Chem. Phys*. **2010**, *132*, 241102.

[13] S. Maeda, K. Morokuma, *J. Chem. Theory Comput*. **2011**, *7*, 2335.

[14] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys*. **2013**, *15*, 3683.

[15] S. Maeda, T.Taketsugu, K. Morokuma, *J. Comput. Chem*. **2014**, *35*, 166.

[16] K. Bondensgård, F. Jensen, *J. Chem. Phys*. **1996**, *104*, 8025.

[17] S. Maeda, T.Taketsugu, K. Morokuma, K. Ohno, *Bull. Chem. Soc. Jpn*., **2014**, *87*, 1315.

[18] S. Maeda, S. Komagawa, M. Uchiyama, K. Morokuma, *Angew. Chem. Int. Ed*. **2011**, *50*, 644.

[19] S. Maeda, K. Morokuma, *J. Chem. Theory Comput*. **2012**, *8*, 380.

[20] M. Hatanaka, S. Maeda, K. Morokuma, *J. Chem. Theory Comput*. **2013**, *9*, 2882.

[21] M. Hatanaka, K. Morokuma, *J. Am. Chem. Soc*. **2013**, *135*, 13972.

[22] R. Uematsu, S. Maeda, T. Taketsugu, *Chem. Asian J*. **2014**, *9*, 305.

[23] G. Zeng, S. Maeda, T. Taketsugu, S. Sakaki, *Angew. Chem. Int. Ed*. **2014**, *126*, 4721.

[24] R. Uematsu, E. Yamamoto, S. Maeda, H. Ito, T. Taketsugu, *J. Am. Chem. Soc*. **2015**, *137*, 4090.

[25] M. Hatanaka, K. Morokuma, *ACS Catal*. **2015**, *5*, 3731.

[26] R. Ramozzi, K. Morokuma, *J. Org. Chem*. **2015**, *80*, 5652.

[27] W. M. C. Sameera, M. Hatanaka, T. Kitanosono, S. Kobayashi, K. Morokuma, *J. Am. Chem. Soc*. **2015**, *137*, 11085.

[28] Y. Matsuda, K. Hoki, S. Maeda, K.-i. Hanaue, K. Ohta, K. Morokuma, N. Mikami, A. Fujii, *Phys. Chem. Chem. Phys*. **2012**, *14*, 712.

[29] S. Maeda, E. Abe, M. Hatanaka, T. Taketsugu, K. Morokuma, *J. Chem. Theory Comput*. **2012**, *8*, 5058.

[30] M. Isegawa, S. Maeda, D. J. Tantillo, K. Morokuma, *Chem. Sci*. **2014**, *5*, 1555.

[31] M. Gao, A. Lyalin, S. Maeda, T. Taketsugu, *J. Chem. Theory Comput*. **2014**, *10*, 1623.

[32] M. Gao, A. Lyalin, M. Takagi, S. Maeda, T. Taketsugu, *J. Phys. Chem*. C **2015**, *119*, 11120.

[33] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec*., **2016**, *16*, 2232.

[34] K. Ohno, S. Maeda, *Mol. Sci*. **2011**, *5*, A0042.

[35] J. P. K. Doye, D. J. Wales, *J. Chem. Phys*. **2002**, *116*, 3777.

Chapter 6 For Improving the Efficiency of Reaction Network Generation: A Case Study for H_2O on Cu(111) Surface

6.1 Introduction

Finding reaction paths is one of the most important task in the theoretical study on the chemical reaction mechanisms. Reaction paths thus have been calculated in many theoretical studies[1-5]. In many studies, equilibrium structure (EQ) and transition state (TS) structure have been calculated by geometry optimization.

However, there is a problem in geometry optimization. Namely, it requires a good guess or previous knowledge of the reaction mechanism. It is thus not easy to calculate unknown reaction mechanisms. Moreover, in complex reactions, the reaction paths become highly complicated multistep and form network structure. Calculating and constructing such complex reaction path networks is a hard task.

For example, surface reaction is severe complicated. In surface reactions, various elementary processes, such as adsorption / desorption of molecules, migration of surface adsorbed species, bond formation / breaking in them, etc., are involved. For construction of such entire reaction networks, automated reaction path search would be promising.

Methods for automated reaction path search in molecular systems have been developed[6-29]. On the other hand, applications of automated reaction path search to surface reactions have been limited[30-32]. Moreover, there have been no concrete examples in which the entire reaction network including paths of migration processes as well as those in high energy regions was constructed with slab model.

For highly complicated reaction systems, reaction path networks may contain hundreds or more EQs and thousands or more TSs. To construct such reaction path

networks, a large computational cost is required. It is therefore necessary to develop a method for reducing the computational cost.

In this chapter, the author tackles these subject using the artificial force induced reaction (AFIR) method[16]. In addition, the author proposes a way to construct a reaction path network efficiently applying kinetic analysis. As an illustrative example, the author calculates the reaction of H_2O absorbed on Cu(111) surface.

6.2 Computational details

In this study, SIESTA 3.2 program was used to compute energy and gradient at each structure[33,34]. The pseudopotentials were prepared with the parameters in the GGA pseudopotential database[35]. The PBE functional and the DZP basis set were used. The electronic temperature was set to 10000.0 K, where the occupation function proposed by Methfessel and Paxton was applied[36]. The Monkhorst-Pack grid for the *k*-point sampling was changed depending on the search stages as described below. The Cu(111) surface was prepared starting from the structure of Cu crystal. At first, the lattice constant was optimized; the optimized value 3.708 Å is slightly larger than the experimental value 3.615 Å. A slab model composed of two layers containing 16 Cu atoms in each layer was adopted. In this study, all atoms in the Cu(111) surface were fixed at the initial positions throughout.

For systematical reaction path search, the AFIR method was used. The search was initiated from a single H_2O adsorbed structure shown in Fig. 6.1. To avoid large separation of the surface adsorbed species from the center of the surface during the automated search, a weak force was applied between the central Cu dimer and each atom in H_2O using the AFIR function with the collision energy parameter γ = 25 kJ/mol, where

this weak bias was finally eliminated from the results as explained below. The central Cu dimer and atoms in H_2O were considered as "target atoms" in the automated fragment generation procedure of the SC-AFIR algorithm; it was thus assumed that chemical bond rearrangements occur among these five atoms. The γ value adopted in the SC-AFIR was 500 kJ/mol. The artificial force was applied not only to atoms in adsorbed species but also to surface atoms that were kept fixed at the initial positions. AFIR path for all fragment pairs defined at each local minimum were calculated strting from all of the obtained EQs, where this criteria to generate fragment pairs at each EQ are described in detail in reference[37]. The search therefore is not stochastic path sampling but a completely deterministic path search.

In the initial search, the Monkhorst-Pack grid and the mesh cutoff value were set to 1×1×1 and 150.0 Ry, respectively.

6.3 Idea and Strategy for Improving the Efficiency of Reaction Network Generation

During the search by the AFIR method, AFIR paths obtained by minimization of the AFIR function were reoptimized by the locally updated plane (LUP) method[38,39], and a network composed of LUP paths were obtained. The LUP method optimizes discrete path points toward the direction perpendicular to path tangent and gives a LUP path which can be a good approximation of the corresponding intrinsic reaction coordinate (IRC) path[40,41].

The LUP path network is usually enough to discuss qualitative overviews of the reaction mechanisms. To further discuss quantitative reaction kinetics, actual TSs for paths that behaved as the reaction bottlenecks were optimized. To automatically extract such paths, the rate constant matrix contraction (RCMC) method was utilized[42,43].

RCMC recursively contracts an original $p \times p$ rate constant matrix for the path network including p local equilibrium structures (EQs), giving a $q \times q$ rate constant matrix describing transitions between q superstates ($p > q$), where each superstates are expressed as a weighted sum of all the original states, i.e., EQs on the path network. In other words, by using the RCMC, EQs that convert to another EQ within a certain timescale t are regarded to belong to the same superstate, and a few superstates that interconvert to each other in a longer timescale than t is obtained. Users can input t. In this study, $t = 10^{-3}$ second was adopted.

When EQs are contracted to superstates, the grade of membership is calculated for all EQs. The grade of membership ω_{Ax} represents the degree to which EQ x is attributed to superstate A. In the context of fuzzy clustering, A superstate corresponds to a fuzzy set, and a grade of membership is the value of membership function.

Paths connecting two EQs that belong to two different superstates behave bottlenecks of interconversion between the two superstates and are extracted as follows. At first, the membership grade ω_{Ai} is considered, where state i and j are in the set of original state, $S^{(0)}$, superstate A and B are in the set of superstate in the n-th loop, $S^{(n)}$. The sum of ω_{Ai} over all superstates is unity.

$$\sum_{A \in S^{(n)}} \omega_{Ai} = 1 \qquad (7.1)$$

Paths connecting different original states i and j and satisfying the following three conditions simultaneously are regarded to be bottlenecks between different superstates A and B.

A. $\omega_{Ai}\,\omega_{Bj} > 0.25$ (7.2)

B. $\omega_{Ai}\,\omega_{Bj} > \omega_{Ai}\,\omega_{Aj}$ (7.3)

C. $\omega_{Ai}\,\omega_{Bj} > \omega_{Bi}\,\omega_{Bj}$ (7.4)

6.4 Reaction Network of H_2O on Cu(111) Surface

The initial search found 42 local minima and 266 LUP paths automatically. Since the LUP path network still includes the weak bias applied to prevent the adsorbed species going away from the surface center, all the LUP paths were further optimized by the LUP method on the PES without the bias. In this calculation, the Monkhorst-Pack grid was set to 2×2×2. Through this calculation, a LUP path network which does not include the bias and consists of 26 local minima and 291 LUP paths were obtained. During this calculation, normal mode analysis was done at all local minima and the highest energy points along the LUP paths, in order to analyze the obtained network based on kinetics.

In kinetic analysis based on the RCMC, $t = 10^{-3}$ second was adopted. RCMC was done with three different temperatures T = 300, 500, and 1000 K, and the highest point along paths selected as a bottleneck at one of the three temperatures were further optimized to an actual TS. In total, 13 actual TSs were obtained starting from the highest energy points along the selected LUP paths. IRC paths were then calculated starting from these 13 TSs.

Finally, a path network consisting of 291 LUP paths and 13 IRC paths was obtained. In total, 97829 gradient calculations were required in the above three calculations, i.e., initial search, LUP path optimizations without the bias, and the final TS optimization and IRC calculation. This number includes gradient calculations in normal mode analysis. The final network was obtained automatically just by running these three calculations sequentially. Although IRC paths for short timescale events are approximated by LUP paths, influence of errors in their barrier heights on the kinetics in longer timescale than 1 second should be negligible, as discussed in Section 6.6.

The obtained path network is depicted in Fig. 6.1, where IRC and LUP path connections are indicated by thick and thin lines, respectively. In the network, there exist four adsorption states, i.e., H_2O, H + OH, H + H + O, and H_2 + O. The path network is much more complicated than expected for a triatomic system, because of surface migration paths.

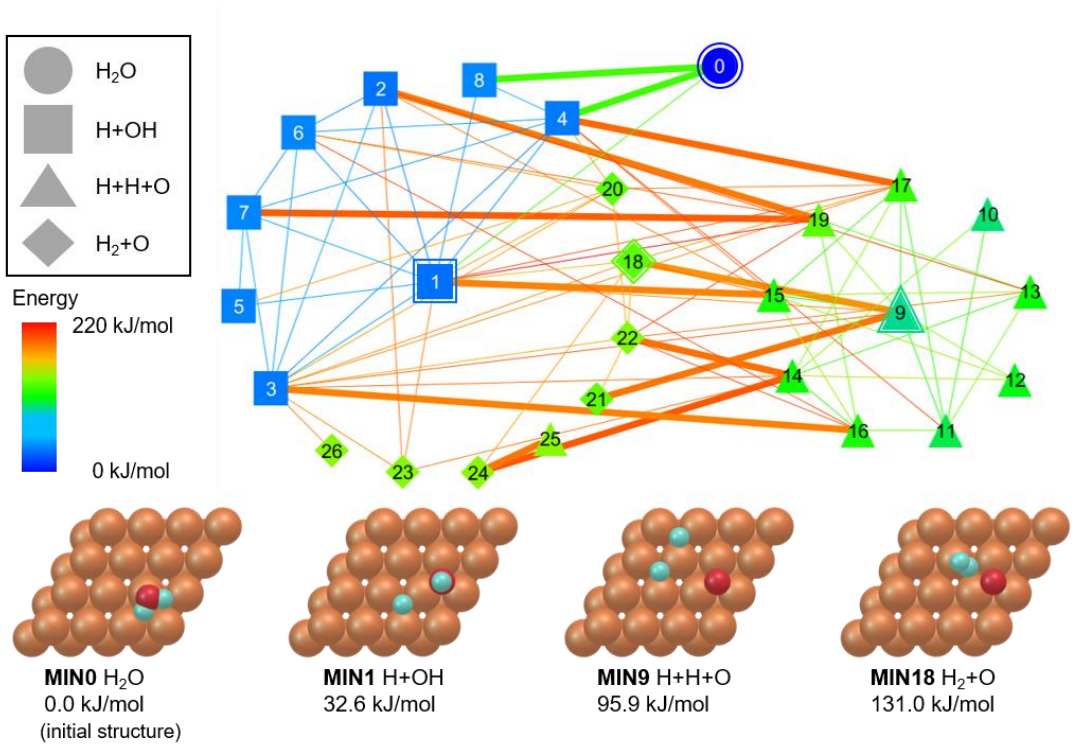

Fig. 6.1 The reaction path network of H_2O on Cu(111) surface.

6.5 Kinetics of H_2O on Cu(111) Surface

Fig. 6.2 shows the most stable structures of each adsorption state, i.e., H_2O, H + OH, H + H + O, and H_2 + O. RCMC with $t = 10^{-3}$ second and T = 300 K gave three superstates, i.e., SS0, SS1, and SS2 in Fig. 6.2. SS0 contains nearly the most stable local minimum MIN0 (H_2O). SS1 mainly includes H + OH (MIN1-8) and H_2 + O (MIN18, 20-26). This

indicates that H_2 + O is not kinetically stable and interconverts to H + OH within 10^{-3} second. The remaining SS2 corresponds to H + H + O (MIN9-17, 19). Activation free energy values for transitions among these three superstates are also shown in Fig. 6.2, where these values were obtained by taking logarithm of the overall rate constants obtained as off-diagonal elements of the final 3×3 rate constant matrix obtained by the RCMC.

RCMC with t = 1 second and T = 300 K gave two superstates, where SS2 obtained by RCMC with $t = 10^{-3}$ second was contracted to SS1 because the activation free energy for the transition from SS2 to SS1 was small enough with t = 1 second. Finally, RCMC with $t = 10^{3}$ second and T = 300 K gave only one superstate which includes all the original states. RCMC with different t revealed a timescale hierarchy of the present reaction path network[44].

The present procedure gave TSs for transitions among the three superstates automatically. All surface migration processes occur in shorter timescales than 10^{-3} second, and only bond reorganization paths behave as bottlenecks in this system. The situation may change in systems in which adsorption energy is much larger than the present system, and in such a case RCMC will automatically identify migration processes as bottlenecks. The present procedure based on RCMC does not involve any heuristic path selection such as a selection based on bond connectivity pattern and thus is fully systematic; only criterion to choose important paths is kinetics.

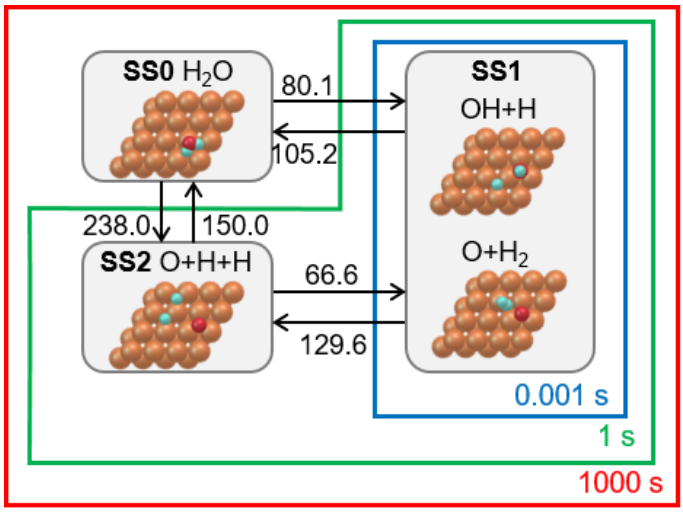

Fig. 6.2 Time hierarchy of the reaction of H_2O on Cu(111) and the contracted network obtained by the RCMC with different timescale. SS is superstate. The values near the arrows are activation free energies in kJ/mol. The blue, green, and red line corresponds to the region of superstate obtained by the RCMC with $t = 10^{-3}$, 1 and 10^{-3}, respectively.

6.6 Influence of Errors in Barrier Height

Influence of errors in barrier heights in paths of short timescale events on the longer timescale kinetics is examined here. Let us consider a model network consisting of ten local minima (EQ1-10) shown in Fig. 6.3. Among these ten, nine (EQ1-9) have the same energy and are separated by barriers of 56.3 kJ/mol, where its rate constant is 1.00×10^{3} s^{-1} at $T = 300$ K. The remaining one (EQ 10) is connected to only one of the other nine (EQ 9) through a path with barrier from EQ 9 to EQ 10 of 73.5 kJ/mol and from EQ 10 to EQ 9 of 226.5 kJ/mol. The rate constants at $T = 300$ K for the forward (EQ 9 to EQ 10) and backward (EQ 10 to EQ 9) reactions are 1.00 and 2.29×10^{-27} s^{-1}, respectively.

RCMC with $t = 1.00\times10^{-3}$ second and $T = 300.0$ K gave two superstates: one mainly including EQ1-9 and the other dominated by EQ10. The rate constant from the

former superstate to the latter was 1.111×10^{-1} s^{-1}, where the small decrease of the rate constant from 1.0 is due to conformational entropy in the superstate including the nine local minima[42,43]. To see the influence of the barrier heights in the superstate of EQ1-9, the author assumes an error ±10 kJ/mol in the barrier between these EQs. When the barrier is increased from 56.3 kJ/mol to 66.3 kJ/mol, the rate constant from the superstate of EQ1-9 to the other superstate was 1.111×10^{-1} s^{-1}. On the other hand, when the barrier is decreased from 56.3 kJ/mol to 46.3 kJ/mol, the rate constant from the superstate of EQ1-9 to the other superstate was 1.111×10^{-1} s^{-1}. These numerical data showed that influence of errors in the barrier height in the superstate of EQ1-9 was negligible in the overall kinetics of this system. This result justifies the reaction path network consisting of LUP paths and IRC paths as adopted in the present procedure.

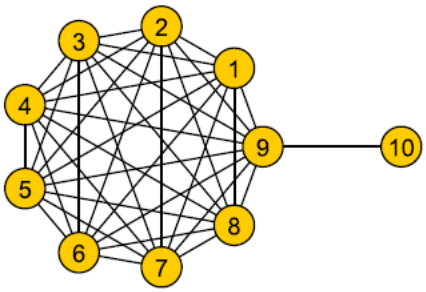

Fig. 6.3 Model reaction network. Node and edge correspond to EQ and reaction pathway. The free energies of EQ 1 to 9 are 0.0 kJ/mol, and the one of EQ 10 is −226.5 kJ/mol.

6.7 Conclusion

In this chapter, the author proposed a cost-effective way to construct reaction path networks based on SC-AFIR and RCMC. The initial search is done by SC-AFIR to obtain approximate paths of the LUP path optimization method. Then, paths behaving as reaction bottlenecks are extracted by RCMC, and TS optimization and IRC calculation are applied to these paths. The resulting path network consists of IRC paths representing its bottlenecks (chemical bond rearrangements in the present system) and LUP paths for fast processes (surface migrations in the present system). Such a hybrid network is enough to discuss kinetics of long timescales. The search does not rely on any heuristic path selection and thus is full systematic. Its performance was examined for the reaction of H_2O on Cu(111) surface. The present scheme would be promising for unbiased construction of many kinds of reaction path networks such as surface reactions and organic reactions.

References

[1] N. Koga and K. Morokuma, *Chem. Rev.*, **1991**, *91*, 823.

[2] S. Niu and M. B. Hall, *Chem. Rev.*, **2000**, *100*, 353.

[3] M. Torrent, M. Solà and G. Frenking, *Chem. Rev.*, **2000**, *100*, 439.

[4] T. Ziegler and J. Autschbach, *Chem. Rev.*, **2005**, *105*, 2695.

[5] K. N. Houk and P. H.-Y. Cheong, *Nature*, **2008**, *455*, 309.

[6] K. Ohno, S. Maeda, *Chem. Phys. Lett.* **2004**, *384*, 277.

[7] S. Maeda, K. Ohno, *J. Phys. Chem. A* **2005**, *109*, 5742.

[8] K. Ohno, S. Maeda, *J. Phys. Chem. A* **2006**, *110*, 8933.

[9] K. Ohno, S. Maeda, *Phys. Scr.* **2008**, *78*, 058122.

[10] S. Maeda, K. Morokuma, *J. Chem. Phys.* **2010**, *132*, 241102.

[11] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2011**, *7*, 2335.

[12] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683.

[13] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, *35*, 166.

[14] S. Maeda, T. Taketsugu, K. Morokuma, K. Ohno, *Bull. Chem. Soc. Jpn*. **2014**, *87*, 1315.

[15] W. M. C. Sameera, S. Maeda, K. Morokuma, *Acc. Chem. Res.* **2016**, *49*, 763.

[16] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec.* **2016**, *16*, 2232.

[17] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, T. R. Walsh, *Adv. Chem. Phys.* **2000**, *115*, 1.

[18] M. Dallos, H. Lischka, E. V. D. Monte, M. Hirsch, W. Quapp, *J. Comput. Chem.* **2002**, *23*, 576.

[19] J. Baker, K. Wolinski, *J. Comput. Chem.* **2011**, *32*, 43.

[20] P. M. Zimmerman, *J. Comput. Chem.* **2013**, *34*, 1385.

[21] D. Rappoport, C. J. Galvin, D. Yu. Zubarev, Alán Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, *10*, 897.

[22] B. Schaefer, S. Mohr, M. Amsler, S. Goedecker, *J. Chem. Phys.* **2014**, *140*, 214102.

[23] D. J. Wales, *J. Chem. Phys.* **2015**, *142*, 130901.

[24] S. Habershon, *J. Chem. Phys.* **2015**, *143*, 094106.

[25] E. Martínez-Núñez, *J. Comput. Chem.* **2015**, *36*, 222.

[26] M. Bergeler, G. N. Simm, J. Proppe, M. Reiher, *J. Chem. Theory Comput.* **2015**, *11*, 5712.

[27] X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757.

[28] L.-P. Wang, R. T. McGibbon, V. S. Pande, T. J. Martinez, *J. Chem. Theory Comput.* **2016**, *12*, 638.

[29] M. Yang, J. Zou, G. Wang, S. Li, *J. Phys. Chem. A* **2017**, *121*, 1351.

[30] S. Ohno, K. Shudo, M. Tanaka, S. Maeda, K. Ohno, *J. Phys. Chem. C* **2010**, *114*, 15671.

[31] X.-J. Zhang, C. Shang, Z.-P. Liu, *J. Chem. Phys.* **2017**, *147*, 152706.

[32] Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, *Nat. Commun.* **2017**, *8*, 14621.

[33] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, J. Phys. Condens. Matter 2002, 14, 2745

[34] E. Artacho, J. M. Cela, J. D. Gale, A. García, J. Junquera, R. M. Martin, P. Ordejón, D. Sánchez-Portal, J. M. Soler, SIESTA 3.2 (Patch Level 4), Fundación General Universidad Autónoma de Madrid, Spain, 2013, see http://www.uam.es/siesta .

[35] Translation of Abinit's GGA pseudo database to Siesta format:

https://departments.icmab.es/leem/siesta/Databases/Pseudopotentials/periodictable-gga-abinit.html .

[36] M. Methfessel, A. T. Paxton, Phys. Rev. B 1989, 40, 3616.

[37] S. Maeda, Y. Harabuchi, M. Takagi, K. Saita, K. Suzuki, T. Ichino, Y. Sumiya, K. Sugiyama, Y. Ono, *J. Comput. Chem.* **2018***, 39,* 233.

[38] C. Choi, R. Elber, *J. Chem. Phys.* **1991**, *94*, 751.

[39] P. Y. Ayala, H. B. Schlegel, *J. Chem. Phys.* **1997**, *107*, 375.

[40] K. Fukui, *Acc. Chem. Res.* **1981**, *14*, 363.

[41] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int. J. Quant. Chem.* **2015**, *115*, 258.

[42] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T. Taketsugu, S. Maeda, *J. Phys. Chem. A* **2015**, *119*, 11641.

[43] Y. Sumiya, T. Taketsugu, S. Maeda, *J. Comput. Chem.* **2017**, *38*, 101.

[44] Y. Nagahata, S. Maeda, H. Teramoto, T. Horiyama, T. Taketsugu, T. Komatsuzaki, *J. Phys. Chem. B* **2016**, *120*, 1961.

Chapter 7 New Algorithm for Automated Reaction Path Search: Application of Kinetic Analysis

7.1 Introduction

Theoretical studies have elucidated various organic reaction mechanisms[1-3]. Density functional theory has been used in many studies. In such studies, equilibrium structure (EQ) and transition state (TS) structure have been calculated by geometry optimization. Each EQ is connected by intrinsic reaction coordinate (IRC) paths via TS[4].

Recently, automated reaction path search methods have been developed[5-21] and have allowed to construct reaction path networks of various reactions such as organic reactions. In reaction path network, EQ and TS correspond to node and edge. A network includes various elementary steps such as bond-rearrangement, conformational rearrangement, and formation of complex. As the number of atoms in a reaction system increases, the total number of node and edge increases exponentially[22]. Thus, obtaining all EQs and TSs is difficult.

Chemical reactions under normal experimental conditions proceed using thermal energy. In most cases, reaction pathways with extremely high activation energy are unimportant[23]. Using this character, the author proposed an efficient algorithm to explore reaction paths on PES. This algorithm uses two methods, namely the AFIR and the rate constant matric contraction (RCMC)[24,25].

This algorithm was applied to an organic multicomponent reaction. Reactions of more than two reactant molecules are generally referred to as multicomponent reactions[26]. Organic multicomponent reactions provide a powerful tool towards the one-pot synthesis of complex compounds. In the reactions, many elementary steps

compete and the reaction mechanism becomes complex. In this study, the author applied the present algorithm to Passerini reaction which is one of the multicomponent reactions, and obtained the reaction network with a reasonable computational cost.

7.2 New Algorithm of Automated Reaction Path Search Based on Kinetic Analysis

The present algorithm consists of automated path search (search program) and kinetic analysis (navigation system). In this work, the GRRM/artificial force induced reaction (AFIR)[27] was adopted as a search program. Kinetic analysis was used as a navigation system. The search program and the navigation system run in parallel. The search program selectively follows kinetically preferable reaction pathways from a given structure under input temperature and timescale.

In the present algorithm, kinetic analysis is applied when the search program finds a new TS. The RCMC is used in kinetic analysis. RCMC recursively contracts an original $N\times N$ rate constant matrix for the path network including N local EQs, giving a $(N-M)\times(N-M)$ rate constant matrix describing transitions between $N-M$ superstates ($N > M$), where each superstates are expressed as a weighted sum of all the original states, i.e., EQs on the reaction path network. Traffic volumes of all EQs are computed by the RCMC. Traffic volume represents accessibility of population. EQs with non-zero traffic volume are recorded in search list. The search program reads the search list,and select an EQ to which the subsequent path search is applied. The list is updated when the RCMC is applied. When the search is applied to all EQs in the search list, path search is terminated. The procedure of this algorithm is as follows.

Step1 Input an initial structure for search, temperature T, and timescale τ. When the input

is completed, the structure is recorded in the search list and the search program is executed.

Step2 When the search program finds a new TS, the rate constant matrix is computed from the obtained network and the RCMC is applied. Then, the traffic volume for the EQs is calculated.

Step3 EQs with non-zero traffic volume are written in the search list. The search program reads the updated list.

Step4 The search program determines an EQ to which the subsequent path search is applied from the search list. Then, return to Step 2. When the search program is applied to all EQs in the search list, search is terminated. An initial EQ for the subsequent search is selected in the order of R defined by the following equation.

$$R = \frac{\xi_1}{1+\{-\theta(\xi_2)\log_{10}(\Lambda)\}} \qquad (7.1)$$

Here, ξ_1, ξ_2 are a random number from 0 to 1, and $\theta(\xi_2)$ is the Heaviside function defined by the following equation:

$$\theta(\xi_2) = \begin{cases} 0\ (0 \leq \xi_2 < 0.5) \\ 1\ (0.5 \leq \xi_2 \leq 1) \end{cases} \qquad (7.2)$$

The traffic volume Λ is defined to be a value from 0 to 1. According to the above equation, EQ with high log-scaled traffic volume is easy to select. In addition, EQ in the search list is select randomly with 50% probability because of ξ_1 and $\theta(\xi_2)$.

The above algorithm allows to omit search from EQs that cannot be kinetically reached and to enlarge the reaction path network in the direction in which the reaction proceeds. The calculation of traffic volume in Step 2 is introduced in the next section.

7.3 Calculation of Traffic Volume by Rate Constant Matrix Contraction

In this section, the procedure for calculation of traffic volume using RCMC is explained.

In the conventional RCMC, contraction was applied to an initial state with maximum rate constant, whereas in this paper contraction is applied to an initial state having maximum reaction rate and the initial population is redistributed to adjacent states. The redistribution of initial population is performed according to eq. (4.6) introduced in Chapter 4. Here, the initial population does not is redistributed to the state where it is connected by process with a rate constant smaller than the inverse of the input timescale τ.

In calculation for traffic volume Λ_j, the membership grade of EQ j for superstate I ($\omega_{Ij}^{(n)}$); the original Boltzmann distribution of EQ j ($P_j^{(0)}$); the population of superstate I ($Q_I^{(n)}$); the ratio of the Boltzmann distribution of EQ j in superstate I to all Boltzmann distribution in superstate I ($P_{Ij}^{(n)}$); population of EQ j that is assigned by returning initial population of all superstates to original EQs ($T_j^{(n)}$) are used, where $\omega_{jj}^{(0)} = 1.0$. In addition, the set of states (either original states at $n = 0$ or superstates at $n \geq 1$) obtained at the nth contraction is denoted by $S^{(n)}$; superstates I, J, K, L, and M correspond to the components of $S^{(n)}$ ($I, J, K, L, M \in S^{(n)}$); the rate constant from superstate I to superstate J is denoted by $k_{I\to J}^{(n)}$; and the Boltzmann distribution for superstate I is denoted by $P_I^{(n)}$. It follows that $S^{(0)}$, $k_{x\to y}^{(0)}$, and $P_x^{(0)}$ correspond to the set of original state (EQ), the original rate constant of the elementary step from EQ x to EQ y, and the original Boltzmann distribution of EQ x, respectively. Here, EQ x and y are the components of $S^{(0)}$($\mathrm{x}, \mathrm{y} \in S^{(0)}$).

Let us explain how to calculate traffic volume through a procedure to contract a $(N-n)\times(N-n)$ rate constant matrix into a $(N-n-1)\times(N-n-1)$ one.

Step1 Identify the superstate *I* with the maximum of outflow reaction rate $V_{I\to\cdot}^{(n)}$. $V_{I\to\cdot}^{(n)}$ is calculated by the following equation.

$$V_{I\to\cdot}^{(n)} = \sum_{J\in S^{(n)}} k_{I\to J}^{(n)} Q_I^{(n)} \quad (7.3)$$

If $V_{I\to\cdot}^{(n)} < 1/\tau$, then Step 10. Otherwise, the operations from Step 2 are sequentially performed.

Step2 Apply the quasi-steady state approximation to the superstate *I* and update $k_{K\to L}^{(n)}$ to $k_{K\to L}^{\prime(n+1)}$ using eqs. (7.4) and (7.5).

$$k_{K\to L}^{\prime(n+1)} = k_{K\to L}^{(n)} + k_{K\to I}^{(n)} k_{I\to L}^{(n)} \sigma_I^{(n)} \quad (7.4)$$

$$\sigma_I^{(n)} = \frac{1}{\sum_{M\in S^{(n)}} k_{I\to M}^{(n)}} \quad (7.5)$$

Step3 Update all rate constants involving the superstate *I* to zero ($k_{I\to M}^{\prime(n+1)} = k_{M\to I}^{\prime(n+1)} = 0$).

Step4 Update Boltzmann distribution and initial population using eqs. (7.6) and (7.7). Eq. (7.7) is the same as eq. (4.6) proposed in Chapter 4. The original initial population $Q_k^{(0)}$ may be set by users in consideration of experimental condition.

$$P_K^{(n+1)} = P_K^{(n)} + k_{I\to K}^{(n)} \sigma_I^{(n)} P_I^{(n)} \quad (7.6)$$

$$Q_K^{(n+1)} = Q_K^{(n)} + k_{I\to K}^{(n)} \sigma_I^{(n)} Q_I^{(n)} \quad (7.7)$$

Step5 Update $\omega_{Kj}^{(n+1)}$ using following equation:

$$\omega_{Kj}^{(n+1)} = \omega_{Kj}^{(n)} + k_{I\to K}^{(n)} \sigma_I^{(n)} \omega_{Ij}^{(n)} \quad (7.8)$$

where, the degree to which EQ *j* belongs to superstate *I* is distributed to the adjacent superstates. That is, by referring the ratio of rate constants $k_{I\to K}^{(n)} \sigma_I^{(n)}$ from the superstate *I* to the other states, a part of membership grade of EQ *j* for superstate *I* ($k_{I\to K}^{(n)} \sigma_I^{(n)} \omega_{Ij}^{(n)}$) is added to $\omega_{Kj}^{(n)}$.

Step6 Compute $P_{Mj}^{(n)}$, $T_j^{(n)}$, $P_{Mj}^{(n+1)}$, $T_j^{(n+1)}$ using the following equations.

$$P_{Mj}^{(n)} = \sum_{M \in S^{(n)}} \frac{P_j^{(0)} \omega_{Mj}^{(n)}}{\sum_{j \in S^{(0)}} P_j^{(0)} \omega_{Mj}^{(n)}} \qquad (7.9)$$

$$T_j^{(n)} = \sum_{M \in S^{(n)}} Q_M^{(n)} P_{Mj}^{(n)} \quad (7.10)$$

Step7 Update $P_i^{(n+1)}$ to zero $(P_i^{(n+1)} = 0)$.

Step8 Update $k'^{(n+1)}_{k \to l}$ to $k_{k \to l}^{(n+1)}$ using eq. (7.11).

$$k_{k \to l}^{(n+1)} = \frac{1}{1 + \sigma_i^{(n)} k_{k \to i}^{(n)}} k'^{(n+1)}_{k \to l} \qquad (7.11)$$

Step9 Define $S^{(n+1)}$ as the set of all superstates except the contracted superstate *I* and return to Step 1.

Step10 Compute traffic volume Λ_j using the following equation.

$$\Lambda_j = \sum_{n=1}^{loop} \left| T_j^{(n)} - T_j^{(n-1)} \right| \qquad (7.12)$$

If $\Lambda_j > 1.0$, then $\Lambda_j = 1.0$.

7.4 Application to Passerini Reaction: Reaction Path Network

Passerini reaction is known as a ternary reaction to produce a-acyloxycarboxamide from carboxylic acids, aldehydes (or ketones) and isocyanides[26]. Recent theoretical study on the Passerini reaction showed that this reaction is a four-component reaction in which a second acid (carboxylic acid) acts as an organic catalyst[28]. In the previous study, the reaction between a formaldehyde, a complex of dimer of formic acid, and a methyl isocyanide was calculated using GRRM/AFIR. Then, the option "*NoBondRearrange*" was used to terminate search when chemical bond rearranges[27].

In this work, the present algorithm does not rely on any heuristic path selection such as bond connectivity and is applied to the same Passerini reaction. The present

algorithm allows to search kinetically preferable reaction pathways. At first, 44 structures in random arrangement between formaldehyde, dimer of formic acid, and methyl isocyanide were optimized as initial structures.

Electronic structure calculations were done by the B3LYP-D3/D95V method. At all obtained EQs and TSs, normal mode analysis was made. Rate constant of the elementary steps was estimated by eqs. (2.2) and (2.3). By the present algorithm, 704 EQs and 5787 TSs were obtained (Fig. 7.1). These structures were reoptimized at B3LYP-D3/D95V(d) level. This reaction path network contains multiple side reaction paths, intermediates, and resting states. Here, note that there were more than 100 000 EQs of the LJ_{20} cluster[22], where this Passerini reaction contains 20 atoms. Since the number of EQs obtained by the present algorithm is 704, the search efficiency was improved at least about 200 times.

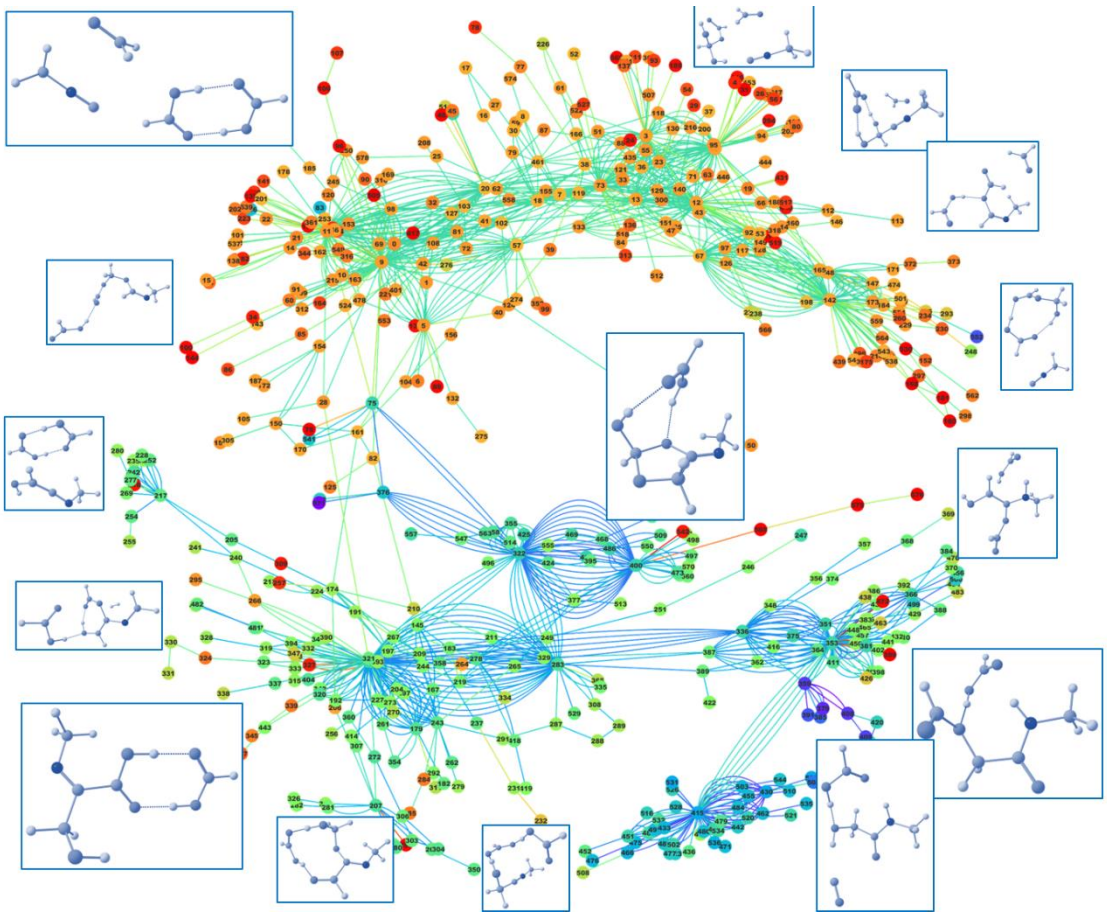

Fig. 7.1 The reaction path network of the Passerini reaction obtained by the present algorithm. Reaction path search was executed at B3LYP-D3/D95V, and reoptimization was executed at B3LYP-D3/D95V(d). Node and edge correspond to equilibrium structure and reaction pathway via TS.

7.5 Application to Passerini Reaction: Kinetics

The time hierarchical structure of this reaction was obtained by applying the RCMC again to the rate constant matrix of the obtained reaction path network in Fig. 7.1. With timescale in the range from 1.5×10^{-6} to 3.0 second, a reaction mechanism consisting of the following four superstates was obtained (Fig. 7.2).

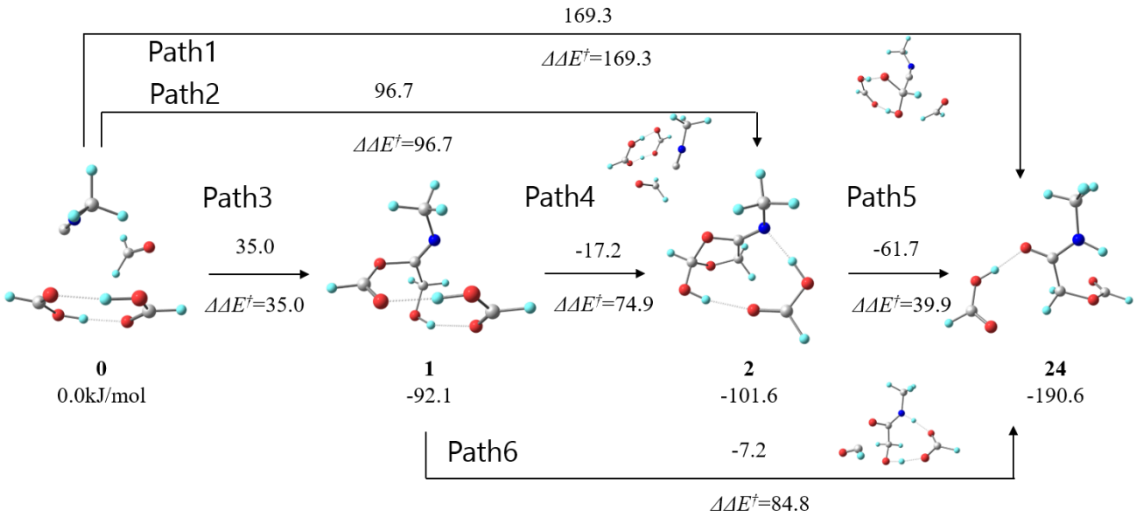

Fig. 7.2 Reaction mechanism of the Passerini reaction catalyzed by formic acid molecule. Each structure is the most stable structure in terms of energy among the EQs in the superstates.

In the above reaction mechanism, the activation energy of the reaction pathway via intermediates 1 and 2 is low. There are also reaction pathways not passing through intermediates 1 and 2. By applying the RCMC with less than 1.5×10^{-6} second, changes between conformers can also be extracted. In addition, by applying the RCMC with timescale set within the range of 1.5×10^{-6} to 3s, two superstates are formed, which correspond to reactant and product, respectively. The structures 1 and 2 are attributed to the superstate corresponding to reactant; the structures 3 and 4 are to the one corresponding to product. The rate constant between these superstates corresponds to the overall rate constant of this Passerini reaction, was 0.38 s^{-1}. The overall rate constant includes the effects of resting state[29] and conformational entropy[24,25,30,31]. If these effects are neglected, the rate constant becomes 1.32 s^{-1}. This difference can also be

understood as a multi-structural effect which importance is pointed out in the calculation of the rate constant in recent years[32].

There are 8 side reactions from the reactant 0, and 7 side reactions from the intermediate 1. An important reaction pathway in these side reactions is a path leading to the structure 8. Since there is no pathway from the intermediate 8 to another structure and it is stable in terms of energy, the intermediate 8 is a resting state which reduces reaction rate. The intermediate 8 is formed as a formic acid passes hydrogen to nitrogen and the remaining part of the formic acid is bonded to carbon. The overall rate constant changed from 1.32 to 0.38 due to these resting states and the effect of the conformer of the intermediate 1.

Moreover, when the RCMC is applied with setting timescale to 3 seconds or longer, the reactant 0, the intermediate 1 and 2 are contracted to the product 24. This indicates that 3 seconds corresponds to the timescale for the product in this Passerini reaction to be sufficiently obtained.

The above results indicate that the present algorithm is effective for generation of reaction networks and kinetic analysis.

7.6 Conclusion

The author proposed an algorithm for construction of reaction path networks, based on SC-AFIR and RCMC. In this algorithm, SC-AFIR and RCMC act as search program and navigation system and run in parallel. The present algorithm can search only kinetically accessible area under the input initial structure, temperature, and reaction time. When search program finds a new TS, navigation system applies kinetic analysis based on RCMC to the obtained reaction path network, and creates search list. In the search list, candidates of accessible EQs are written and the search program reads this list. The search dose not rely on any heuristic path selection such as bond connectivity pattern and thus is fully systematic: only criterion to choose EQs is kinetics. Its performance was examined for Passerini reaction, which is one of multicomponent reactions. The present algorithm would be promising for unbiased construction of reaction path networks.

References

[1] N. Koga, K. Morokuma, *Chem. Rev.* **1991**, *91*, 823.

[2] S. Niu, M. B. Hall, *Chem. Rev.* **2000**, *100*, 353.

[3] K. N. Houk, P. H.-Y. Cheong, *Nature* **2008**, *455*, 309.

[4] K. Fukui, *J. Phys. Chem*. **1970**, *74*, 4161.

[5] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, T. R. Walsh, *Adv. Chem. Phys.* **2000**, *115*, 1.

[6] M. Dallos, H. Lischka, E. V. D. Monte, M. Hirsch, W. Quapp, *J. Comput. Chem.* **2002**, *23*, 576.

[7] J. Baker, K. Wolinski, *J. Comput. Chem.* **2011**, *32*, 43.

[8] P. M. Zimmerman, *J. Comput. Chem.* **2013**, *34*, 1385.

[9] D. Rappoport, C. J. Galvin, D. Yu. Zubarev, Alán Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, *10*, 897.

[10] B. Schaefer, S. Mohr, M. Amsler, S. Goedecker, *J. Chem. Phys.* **2014**, *140*, 214102.

[11] D. J. Wales, *J. Chem. Phys.* **2015**, *142*, 130901.

[12] S. Habershon, *J. Chem. Phys.* **2015**, *143*, 094106.

[13] E. Martínez-Núñez, *J. Comput. Chem.* **2015**, *36*, 222.

[14] M. Bergeler, G. N. Simm, J. Proppe, M. Reiher, *J. Chem. Theory Comput.* **2015**, *11*, 5712.

[15] X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757.

[16] L.-P. Wang, R. T. McGibbon, V. S. Pande, T. J. Martinez, *J. Chem. Theory Comput.* **2016**, *12*, 638.

[17] M. Yang, J. Zou, G. Wang, S. Li, *J. Phys. Chem. A* **2017**, *121*, 1351.

[18] K. Ohno, S. Maeda, *Chem. Phys. Lett*. **2004**, *384*, 277.

[19] S. Maeda, K. Ohno, *J. Phys. Chem*. A, **2005**, *109*, 5742.

[20] K. Ohno, S. Maeda, *J. Phys. Chem*. A, **2006**, *110*, 8933.

[21] S. Maeda, K. Morokuma, *J. Chem. Phys*. **2010**, *132*, 241102.

[22] J. P. K. Doye, D. J. Wales, *J. Chem. Phys*. **2002**, *116*, 3777.

[23] K. Ohno, S. Maeda, *Mol. Sci*. **2011**, *5*, A0042.

[24] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T, Taketsugu, S. Maeda, *J. Phys. Chem. A* **2015**, *119*, 11641.

[25] Y. Sumiya, T. Taketsugu, S. Maeda, *J. Comput. Chem*. **2017**, *38*, 101.

[26] J. Zhu, Q. Wang, M.-X. Wang, Multicomponent Reactions in Organic Synthesis, Wiley-VCH, 2015.

[27] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, *35*, 166.

[28] S. Maeda, S. Komagawa, M. Uchiyama, K. Morokuma, *Angew. Chem. Int. Ed*. **2011**, *50*, 644.

[29] S. Kozuch, J. M. L. Martin, ChemPhysChem **2011**, *12*, 1413.

[30] B. P. English, W. Min, A. M. Van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, X. S. Xie, *Nat. Chem. Biol.* **2006**, *2*, 87.

[31] M. B. J. Roeffaers, G. D. Cremer, H. Uji-I, B. Muls, B. F. Sels, P. A. Jacobs, C. D. F. Schryver, D. E. D. Vos, Hofkens, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12603.

[32] P. Seal, D. G. Truhlar, *J. Am. Chem. Soc.* **2014**, *136*, 2786.

Chapter 8 General Conclusion

In the present thesis, The author developed a new method for kinetic analysis, rate constant matrix contraction (RCMC), for analysis of complex networks composed of elementary steps based on quantum chemical calculation in Chapter 2 to 4. From Chapter 5, changing viewpoint to reaction path search, The author developed algorithms for reducing computational cost of reaction path search with the RCMC.

In Chapter 2, The author summarized the background of kinetic analysis and introduced the methods.

In Chapter 3, The author proposed the method for calculating overall rate constant from a given reaction path network, RCMC. The RCMC is applied to Claisen rearrangement and it reproduced the experimental rate constant. The result indicates that conformational entropy in reactant region plays important role in reactivity even in the Claisen rearrangement of small chain molecules. It is therefore envisaged that the present method could be used as a powerful tool for estimating the overall rate constants of different organic reactions.

In Chapter 4, The author introduced the *f*-RCMC as an extension of the RCMC. This method allows for obtaining the branching ratio of unimolecular decomposition without performing the kinetic simulation. This method assumes that the dissociated fragments do not undergo secondary reactions, and is thus effective only for unimolecular decomposition in the gas-phase. The method was numerically tested with unimolecular decomposition of C_3H_5 and C_4H_5, where their reaction profiles were constructed using the SC-AFIR method. The branching ratios obtained by the *f*-RCMC method agreed precisely up to eight decimal places with those obtained by the conventional kinetic simulation. The *f*-RCMC method requires only the rate constant matrix as input and gives

the branching ratios immediately without solving the rate equations. It is therefore expected that the *f*-RCMC method will be a highly efficient alternative to the ordinary kinetic simulation in predicting the branching ratios of unimolecular decomposition.

In Chapter 5, the author introduced methods and history of automated reaction path search. Next, the author introduced that the number of EQ and TS increases exponentially according to the increase of the number of atoms in the system.

In Chapter 6, based on Chapter 5, the author proposed a way to reduce computational cost for construction of reaction path networks. This way is based on kinetic analysis with RCMC and automated reaction path search with GRRM/AFIR. In this scheme, (1) initial search is done by SC-AFIR to obtain approximate paths of LUP path optimization, and (2) paths that behave as reaction bottlenecks are extracted by the RCMC, and TS optimization and IRC calculation are applied to these paths. This allows to reduce the computational cost for TS optimization and IRC calculation. The obtained hybrid network composed of LUP paths and IRC paths is enough to discuss kinetics of long timescales. This scheme was applied to the reaction of H_2O on Cu(111) surface. The present scheme would be promising for unbiased construction of reaction path networks.

In Chapter 7, based on Chapter 5, the author proposed an algorithm reducing computational cost for construction of reaction path networks. This algorithm is also based on kinetic analysis with RCMC and automated reaction path search with GRRM/AFIR. The present algorithm can search only kinetically accessible area under the input initial structure, temperature, and reaction time. The author applied this algorithm to Passerini reaction, and the reaction path network including multistep paths from reactant to product and intermediates and byproducts that can be produced was obtained. The present scheme would also be promising for unbiased construction of

reaction path networks.

In the present thesis, the author presented the method for kinetic analysis of complex network composed of chemical reaction elementary steps, RCMC. The present method allows to estimate the reaction overall rate constant from a given network and reduce the stiffness of the rate equations. In addition, it was found that conformational entropy can make a substantial contribution to overall rate constant, namely reactivity. On the other hand, in the case of complex reaction systems, reaction path networks become more complicated, and obtaining the network becomes difficult even if automated reaction path search is used. The author thus proposed two ways to improving the efficiency of automated reaction path search using the RCMC in order to deal with the problem.

The author hopes that the results in this thesis contribute to elucidation of reaction path network and mechanism in complex systems, and to gain theoretical insight of various reactions.

Acknowledgement

本研究は北海道大学大学院理学研究院、前田理教授のご指導のもと行われました。6 年間にわたり私にご指導いただきました。ここに篤くお礼申し上げます。

武次徹也教授、小林正人講師、小松崎民樹教授のお三方には、研究遂行にあたり、多くのご助言を賜りました。深く感謝いたします。