Theoretical Study on Amorphous Oligomerization of Organic Molecules and Conformation Determination of Length-controlled Organic Oligomers

(有機分子の不定形多量化反応と長さが制御された有機多量体のコンフォメーション決定に関する理論的研究)

Keisuke Tashiro

Graduate School of Chemical Sciences and Engineering

Hokkaido University

March 2024

Contents

1. General introduction

1.1 Computational Chemistry for Organic Oligomers

The theoretical study of organic oligomers is important as a guideline for molecular design in efficient polymer synthesis. Additionally, in recent years, examples have been reported where organic oligomers themselves exhibit unique behavior due to their intrinsic molecular size and shape[A1]. As symbolized by Moore's Law, computational resources have exponentially increased over the past few decades. In parallel with hardware advancements, software development has also progressed. In the field of quantum chemistry, efficient algorithms has been developed year by year. This thesis focuses on two organic oligomer systems that have traditionally been computationally expensive. I performed quantum chemical calculations by leveraging modern computational methods and supplementing them with my own algorithms to address their shortcomings. The first topic is the calculation of reaction mechanisms for complex polymerization systems. Polymer synthesis using molecules with a single active site has been widely analyzed using quantum chemical calculations because it is easy to reduce to elementary reactions[A2]. However, tracking the reactions of irregularly quantified reactions involving multiple active sites has been a challenging task due to the exponential increase in intermediate structures with increasing polymerization. I analyzed the reaction mechanism of irregularly quantified reactions using the reaction path automatic search program GRRM method and identified the key elementary reactions from complex reaction pathway networks. The second topic is calculations on the conformational determination of spatially flexible organic oligomers.

Determining the conformation of compounds in various phases such as gas and solution phases is one of the most important challenges in chemical research, and this is no exception in organic oligomers. However, it has been difficult to determine the (absolute) conformation of organic oligomers, especially those without rigid chemical structures, based on experimental evidence. I determined the conformation of organic oligomer-cation complexes in the gas phase through a collaboration of experimental and computational approaches by correlating the collision cross-sections obtained from ion mobility-mass spectrometry with those calculated from snapshots obtained from molecular dynamics calculations.

1.2 Global Reaction Route Mapping Strategy

It is important to investigate intrinsic reaction coordinates for understanding chemical reactions because in a static reaction transition states and IRCs define by them are like a compass that determine how a chemical reaction will proceed. However, a comprehensive search of the entire potential energy surface had been difficult due to the exponential increase in degrees of freedom as the number of atoms in the system increases. The GRRM strategy led to a solution to this difficulty by using the characteristics of chemical space. In 2004, Ohno and Maeda proposed the anharmonic downward distribution following (ADDF) method[A3], one of the automated reaction path search methods, based on a scaled hyper-sphere search (SHS) method. This method allows the IRC to “climb” from EQ and can be combined with conventional IRC calculations to discover TS in the chain reaction. The artificial force induced reaction (AFIR) method[A4] was proposed in later years by Maeda and coworkers to explore more efficiently the reaction paths. Direct illustration of AFIR method is often given by the push-and-pull image of two fragments. These automatic methods have contributed to a drastic reduction in the time and effort required to search for reaction pathways, which previously had to be manually investigated.

1.3 *in silico* CCS Calculation method

The collision cross section（CCS） is a physical quantity that appears as the energy integral of the momentum transfer cross section. *in silico* CCS predictions based on computational molecular modeling has been proposed by several method and have proven highly valuable as well. Among the available methods to compute CCS, the trajectory method (TM) is the most accurate[A5,A6]. Although TM is accurate, TM require repeating the process of tracing the gas scattering trajectories while calculating the ionic potentials of analyte every time in short time steps so that its computational cost is expensive. The projection approximation(PA) method, using projected cross sections, and the exact hard sphere scattering(EHSS) method, employing a hard sphere approximation, are models that avoid this problem. The PA method does not consider non-contact interactions with gas molecules (e.g. Coulomb interactions) and instead collision events with projected cross sections, so cross sections tend to be underestimated. On the other hand, the EHSS method considers collision events as collisions between rigid bodies, so collisions that should have been moderated by intermolecular interactions occur directly, and the collision cross section tends to be estimated overestimated. There was also a considerable research to overcome the weaknesses of these methods, such as the PSA(projected superposition approximation) method[A7], which improves the evaluation of the impact surface of the PA method, and the DHSS(diffuse hard sphere scattering) method[A8], which attempts to improve the EHSS method by giving real-gas-like reflection and diffusion effects to the colliding gas molecules.

1.4 Density Functional based Tight Binding method

In quantum chemistry, the density functional theory(DFT) is now widely used because it can efficiently consider electron correlations in terms of electron densities. The fundamental concept of DFT is accelerating the calculation of many-body electron systems by considering the Hamiltonian operator as a potential functional expressed in terms of electron densities. In the early ages, it was based on the Thomas-Fermi theory[A9] using the kinetic and exchange energy functional, but now it is replaced by a method for solving the Schrodinger equation based on the Hohenberg-Kohn theorem. The Kohn-Sham method[A10], which is an exact formulation of Hohenberg-Kohn theorem [A11] adopts the idea of considering a matrix wavefunction instead of a kinetic energy functional. The energy functional is a solution of the Kohn-Sham equation as follows.

$$E[\rho] = V[\rho] + T_s[\rho] + J[\rho] + E_{xc}[\rho]$$

where $V[\rho], T_s[\rho], J[\rho],$ and $E_{xc}[\rho]$ is a electron-nucleus potential functional kinetic energy functional, Coulomb potential functional and exchange-correlation potential functional, respectively. A huge challenge facing DFT is that we don't know what the universal functional corresponding to the exact electron energy was generated.

The Density Functional Tight-Binding(DFTB) method [A12] is a semi-empirical method based on DFT. The total energy of the DFTB is obtained by Taylor expansion of the Kohn-Sham energy in terms of electric density. DFTB1 is defined without charge fluctuation terms, DFTB2 is defined with terms up to the second order, and DFTB3 is defined with terms up to the third order. DFTB3 is a popular method because of its computational level, capable of describing the electrification fluctuations of atoms and hydrogen bonding. The total energy expression of DFTB3 is as follows:

$$E_{DFTB3} = \sum_{A>B}^{\text{atom}} V_{AB}^{\text{rep}} + \sum_{\mu\nu}^{AO} P_{\mu\nu} H_{\mu\nu}^{0} + \sum_{AB}^{\text{atom}} \gamma_{AB} \Delta q_A \Delta q_B + \sum_{AB}^{\text{atom}} \Gamma_{AB} \Delta q_A^2 \Delta q_B$$

where $P_{\mu\nu}$, Δq_A is density matrix and Mulliken charge of atom A, respectively. Other variables are parametrized on the basis of DFT calculated values, and then all charges are determined self-consistently.

1.5 Overview of This Thesis

This thesis is composed of four chapters. Chapter 1 is general introduction and reflected on the analytical methods of chemical mechanisms used in this thesis. In Chapter 2, I show the application of chemical reaction path analysis to the amorphous oligomerization of organic molecules. I succeeded in discovering the key elementary reaction of the complex oligomerization reaction of organic molecular called HMF to apply a systematic search algorithm. In Chapter 3, I developed a theoretical calculation scheme of collision cross section that corresponds to ion mobility mass spectrometry, which is a spectroscopic method for measuring collision cross sections. I applied this scheme to IM-MS spectra of polyketone systems with multiple metastable structures and clarify the origin of the splitting of the IM- MS spectral peaks. Chapter 4 is a general conclusion, and I summarized this thesis.

1.7 Reference

[A1] M. Uesaka, Y. Saito, S. Yoshioka, Y. Domoto, M. Fujita, Y. Inokuma, Oligoacetylacetones as Shapable Carbon Chains and Their Transformation to Oligoimines for Construction of Metal-organic Architectures, *Commun. Chem.*, **2018**, 1, Article Number 23.

[A2] Konstantinov, *et al*, *Mol. Syst. Des. Eng*., **2018**, 3, 228-242

[A3] K. Ohno and S. Maeda, A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface, *Chem. Phys. Lett.*, **2004**, 384, 277

[A4] S. Maeda and K. Morokuma, A systematic method for locating transition structures of A+B-->X type reaction, *J. Chem. Phys.*, **2010,** 132, 241102.

[A5] M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys.Chem*. **1996**, 100, 16082−16086

[A6] C. Ieritano, J. Crouse, J. L. Campbell and W. S. Hopkins, A parallelized molecular collision cross section package with optimized accuracy and efficiency. *Analyst*, **2019**, 144, 1660−1670.

[A7] C. Bleiholder, T. Wyttenbach, and M. T. Bowers, *Int. J. Mass Spectrom.*, **2011** 308(1), 1.

[A8] C. Larriba and C. J. Hogan, *J. Phys. Chem. A*, **2013**, 117（19）, 3887.

[A9] E. Fermi. *Z. Phys.*, **1928**, 48:73-79.

[A10] W. Kohn and L. J. sham. *Phys. Rev. A,* **1965,** 140:1133-1138.

[A11] P. Hohenberg and W. Kohn. *Phys. Rev. B,* **1964** 136:864-871.

[A12] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B*, **1998**, *58*, 7260–7268.

2. Reaction Path Analysis of humin formation from 5-(hydroxymethyl)furfural under basic condition

2.1 Introduction

Efforts to reduce dependence on fossil-derived chemicals due to limited resources are a key aspect of sustainable development. Inedible biomass resources are being explored as alternatives to fossil fuels [B1]. Such challenges have become less imaginary with the development of recent years. The biomass market is growing, and the current bio-based chemical and polymer production is estimated at around 90 million tons[B2].There is also a keen interest in research. The number of publications on ethanol, which is a popular biobased chemicals, has increased 25 times in last decade[B3]. There are various types of chemicals obtained from biomass conversion, depending on the feedstock and the progress of conversion (Fig. B1). For example, CO2 platform mainly yields C1 compounds such as methanol, urea, and formic acid, while the syngas platform yields various carbohydrates including ethanol and isobutane. The sugar platform is by far the largest platform of all, and the platform that plays a mainstream role in biomass conversion. The reason is that a wide variety of C5-C6 compounds can be synthesized from the sugar platform, leading to the generation of various chemicals for polymer production, fuel synthesis, and other applications. Furan-2,5-dicarboxylic acid (FDCA) [B4] is particularly significant in biomass chemistry as it serves as a feedstock for polyethylene terephthalate (PET) substitutes from inedible biomass. Polymerizing FDCA produces polyethylene furanoate, which exhibits superior gas barrier properties compared to PET, replacing fossil-derived

chemicals with more valuable biomass chemicals. The oxidation of 5-(hydroxymethyl)furfural (HMF) to FDCA (Fig. B2) has garnered global interest and research [B5–B8]. However, the complex reaction network involved in this process often leads to the undesirable solid byproduct humin, reducing the efficiency of FDCA conversion and hindering large-scale industrial production [B9]. Understanding the reaction mechanism leading to humin formation through quantum chemical calculations is crucial for designing conditions (e.g., catalysts, temperature, solvent) for efficient FDCA production, yet this area remains largely unexplored due to the lack of methodologies for analyzing intricate reaction networks systematically.

The formation of humins from HMF has been experimentally studied. Horvat *et al.* propose that under acidic conditions, acid-catalyzed hydrolysis leads to a furan ring-opening reaction to form humin [B10,B11]. Patil and Lund demonstrate that polymerization to form humin occurs by the aldol addition of the ring-opened product (2,5-dioxo-6-hydroxy-hexanal) with HMF [B12]. The structure of humin has been analyzed by one-dimensional (1D) and two-dimensional (2D) solid-state nuclear magnetic resonance (NMR) [B11,B12]. Recently, biomass processing using ionic liquid solvents has attracted significant attention [B14]. However, HMF is degraded to humins even in ionic liquid solvents, and its structure has been revealed by analytical techniques, such as Fourier-transform infrared (FT-IR) spectroscopy, ^{13}C-NMR, and X-ray photoelectron spectroscopy (XPS) [B16]. Furthermore, it has been reported that humin is formed even in pure HMF [B17]. Nakajima et al. have reported that HMF–acetal, which is produced by the reaction of HMF with 1,3-propanediol (PD), exhibits high thermal stability against thermal degradation [B9]. The oxidation and oxidative esterification of HMF–acetal affords FDCA and its esters in excellent yield

and selectivity, even in concentrated solutions (> 10 wt%) [B18]. These reactions proceed under basic conditions because the acetal compounds undergo hydrolysis under acidic or neutral conditions. Additionally, a base is added to dissolve FDCA with low solubility as a neutralizing salt. Further clarification of the humin formation mechanism based on quantum chemical calculations is desirable for the effective use of HMF in the future.

In this study, I conduct a comprehensive reaction-path search for humin formation under the conditions of FDCA formation by the oxidation of enriched HMF using the artificial-force induced reaction (AFIR) method, which is a part of the global reaction route map (GRRM) strategy of Maeda *et al.* [B19–B23]. The multi-component (MC) AFIR method is applied to candidate bimolecular reactions in the oligomerization of HMF, which enables the systematic exploration of association reactions. Considering the basic reaction conditions, the following reactions are identified as candidates for the initial elementary reactions: HMF + HMF, HMF + H_2O, HMF + OH^-, and HMF + O_2. By repeatedly applying the MC-AFIR method to the reaction of HMF with the newly obtained chemical species, the formation process of HMF oligomers can be obtained. The results show that *humin* formation is unlikely to occur when only HMF and H_2O are present in the reaction system: the barrier for the reaction between HMF and HMF or HMF and H_2O is at least 186.5 kJ mol^{-1}. In contrast, the reaction between HMF and OH^- yields many intermediates with a low barrier. Thus far, I have obtained a trimer of HMF in the simulations under basic conditions. We conclude that these reactions are the initial steps in the formation of humin from HMF under basic conditions. Furthermore, the mechanism of the suppression of humin formation in concentrated solutions by acetal protection is computationally investigated.

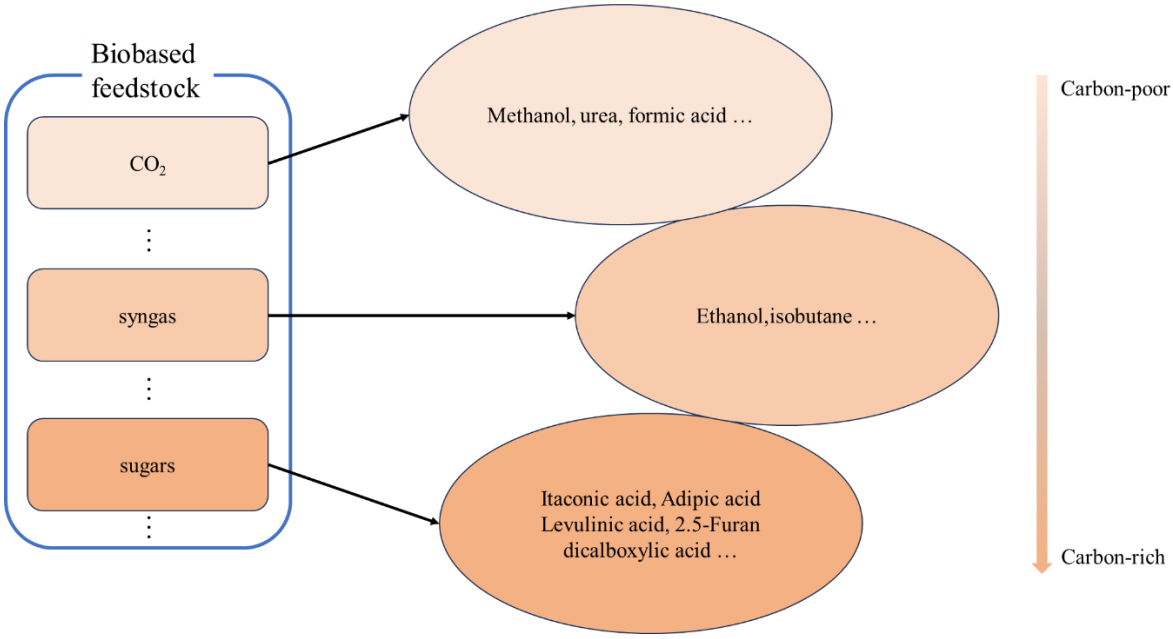

Fig. B1. Variety of biomass conversion platforms

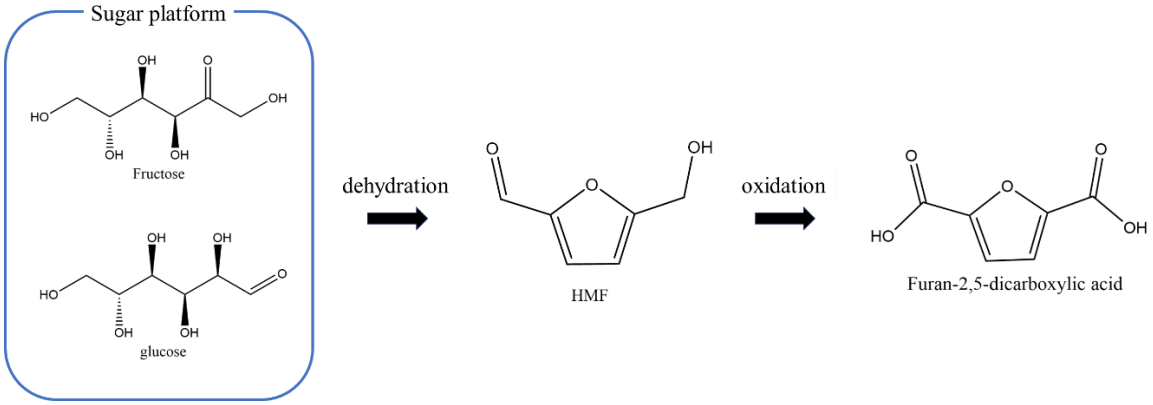

Fig. B2. Synthetic route of FDCA synthesis from sugars via HMF

2.2 Computational detail

The Gaussian16 software package [B24, B25] with Grimme's D3 dispersion correction [B24]. The 6-31+G(d) basis set [B27–B30] was used for all atoms. Water solvent effects were considered using the integral equation formalism variant of the polarizable continuum model (IEFPCM) [B31].The obtained Gibbs Energies are shown in Fig,B3. I also calibrated the electronic energy by single-point calculations with the wB97x-D/aug-cc-pVTZ level of theory. The obtained Gibbs Energies are shown in Fig. B2. While the energy values are different to some extent, the following discussion remains qualitatively the same for values in either Figure.

To systematically search for reaction paths, I employed the AFIR method implemented in the GRRM17 program [B21], which induced a reaction by applying an artificial force between two moieties. The following procedure was employed to analyze the multistep polymerization reactions because the AFIR method implemented in GRRM17 cannot directly handle multistep reactions with different participating molecules (Fig. B4). First, the reactions listed in Table B1 were selected as candidates for the initiation reactions in a basic aqueous solution. The intermediates derived from these reactions were obtained using the MC-AFIR method [B18]. Thereafter, the reactions of these intermediates with HMF were selected as candidates for new bimolecular reactions and examined using the MC-AFIR method. For some intermediates, such as ring-opening systems, further single-component (SC)-AFIR calculations were performed to identify intramolecular rearrangements and conformational changes. By repeating these steps, I could comprehensively investigate the process from the initial humin formation to the HMF trimer. Parameter γ, which represents the magnitude of the artificial force used in the AFIR method, was set in the range of 200–500 kJ mol^{-1}. The number of samples was

set to increase as the size of the system increased or as the system was considered more important. Specifically, the number of samples was set to 100–500 for the initial reaction, 300–600 for the dimerization reaction, and 500–2000 for the trimerization reaction. The reactions obtained by the MC-AFIR method were clustered into elementary chemical reactions, i.e., if the reactants and products of the two reactions were the same in the canonical simplified molecular-input line-entry system (SMILES) representation, these reactions were considered to be chemically identical. The conversion of molecular geometry into canonical SMILES was performed using the Open Babel program [B32,B33]

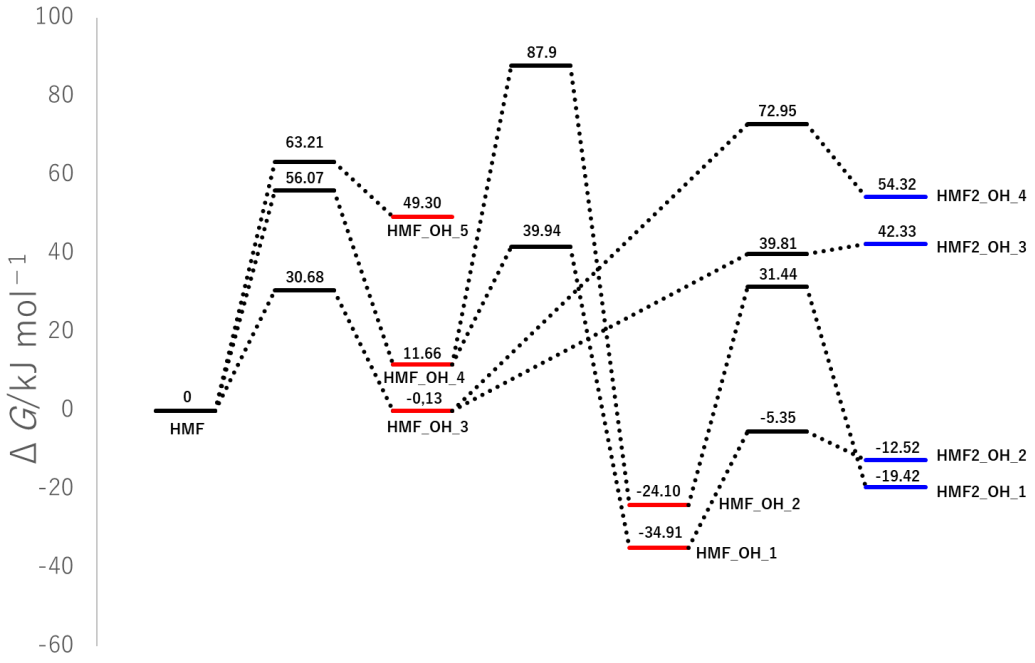

Fig. B3. Gibbs energy diagram for Fig. B6 up to dimer formation. Red and blue lines indicate the energies of $[HMF + OH]^-$ and $[2HMF + OH]^-$, respectively.

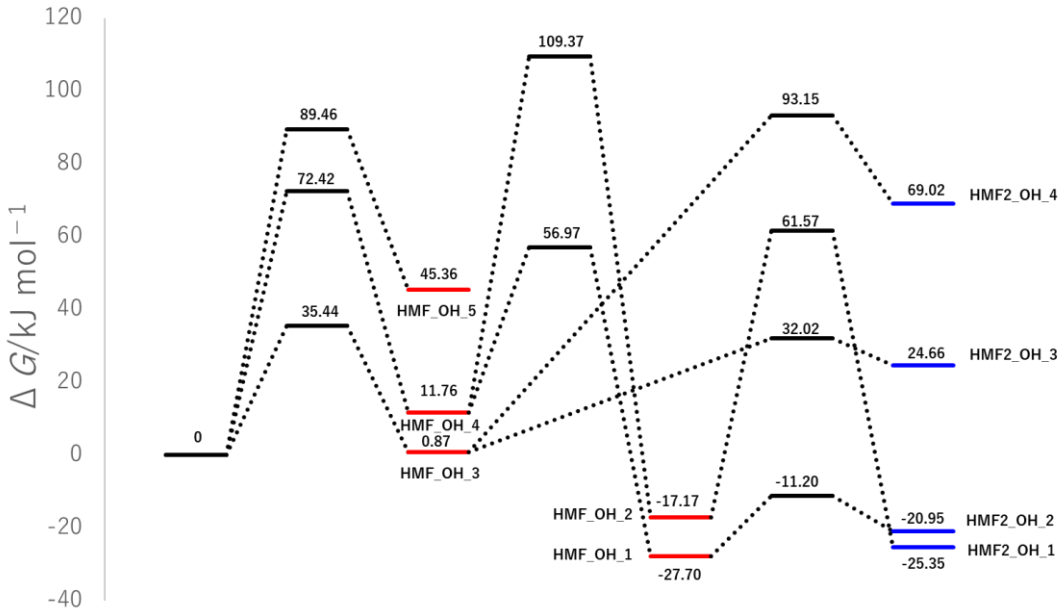

Fig. B4. Gibbs energy diagram for Fig. B7 up to dimer formation. Red and blue lines indicate the energies of $[HMF + OH]^-$ and $[2HMF + OH]^-$, respectively.

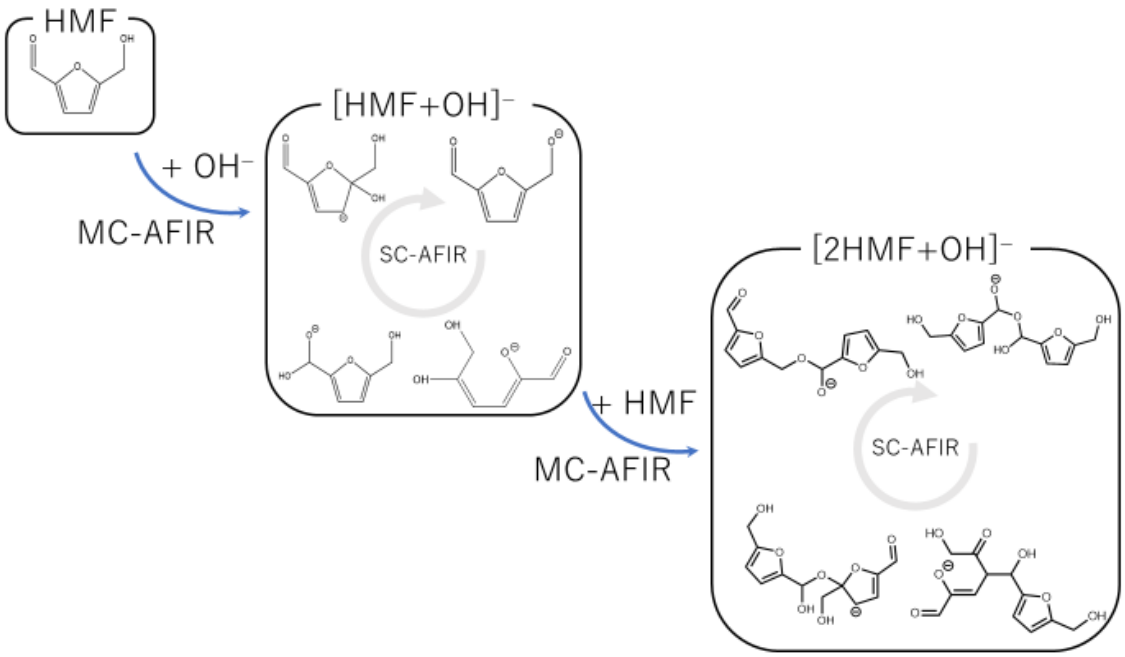

Figure B5. Reaction search scheme using multi-step MC-AFIR method.

2.3 Result and Discussion

The reaction paths obtained from the series of AFIR calculations are summarized in Fig. B6. Also reaction route map calibrated the electronic energy by single-point calculations with the wB97x-D/aug-cc-pVTZ level of theory show in Fig. B7. The values in parentheses next to the labels are Gibbs energies (in kJ mol^{-1}) relative to the reactants. Each arrow represents an elementary reaction, and the value above it is the Gibbs activation energy (in kJ mol^{-1}). The following subsections provide detailed discussions of each process.I focused HMF+OH^- in the following discussion because activation energies of the other reactions quite high. Furthermore, the qualitative discussion will remain the same so that I discussed based on Fig. B6.

2.3.1 Initial Step of Humin Formation

Previous experimental studies have demonstrated that HMF, ranging in purity from 97% to 99%, degrades to a dimer within approximately two weeks, even under room temperature conditions [B17]. At 473 K, this degradation process is nearly complete within 2 hours [B9]. Initially, the interactions between two HMF molecules were investigated using the MC-AFIR method, resulting in the identification of a total of 79 transition states (TSs), including three elementary bimolecular addition reactions. The energy profiles of these reactions are depicted in Fig. B8. Among these, the bimolecular reaction with the lowest activation energy was a Diels-Alder type reaction, requiring 186.5 kJ mol^{-1}, which is significantly high for room temperature conditions. Despite considering the influence of a water solvent in the calculations, the results remained consistent when varying the solvent to HMF, with dielectric constant (ε) and solvent radius (R_{solv}) parameters determined as ε = 26.38 and R_{solv} = 3.384 Å, respectively, based on molecular dynamics simulations. Therefore, the initial degradation of HMF is

attributed to a reaction with impurities. Among the initial reactions of HMF monomers as listed in Table B1, the reaction with OH^- exhibited the lowest activation energy of 30.7 kJ mol^{-1}, while reactions with H2O and O2 (triplet) had activation energies of 209.9 and 135.7 kJ mol^{-1}, respectively. The energy profiles of these reactions are shown in Fig. B9 and Fig. B10. MC-AFIR calculations for the reaction of HMF + OH^- revealed 30 TSs. For chemical reactions with activation energies below 100 kJ mol^{-1}, three hydroxylation reactions were observed at carbons 3 and 5 of the furan ring and the formyl carbon, yielding HMF_OH_5, HMF_OH_4, and HMF_OH_3, respectively (Fig. B6). This is consistent with chemical intuition, as a negative charge on the oxygen atom of the formyl group would induce a resonance structure with a positive charge on one of these three carbons. Additionally, a reaction involving the carbon at position 4 was identified in the MC-AFIR calculation; however, its activation energy was 121.6 kJ mol^{-1}, considerably higher than the aforementioned energies and thus not depicted in Fig. B6. Among these reactions, the formation of HMF_OH_3 was found to be favorable in terms of activation and reaction energies. HMF_OH_3 represents the anionic form of 5-hydroxymethyl-2-furancarboxylic acid, which leads to the production of FDCA. However, HMF_OH_4 induced ring-opening and dehydration reactions, resulting in the formation of more stable intermediates, HMF_OH_2 and HMF_OH_1, respectively, with relatively low activation energies. Consequently, the reaction of HMF with OH^- yielded five intermediates (i.e., HMF_OH_X, X = 1-5) with activation energies below 100 kJ mol^-. The labels "HMF_OH_X" are numbered based on the stability of the species, with HMF_OH_1, produced by the dehydration of HMF_OH_4, being the most stable.

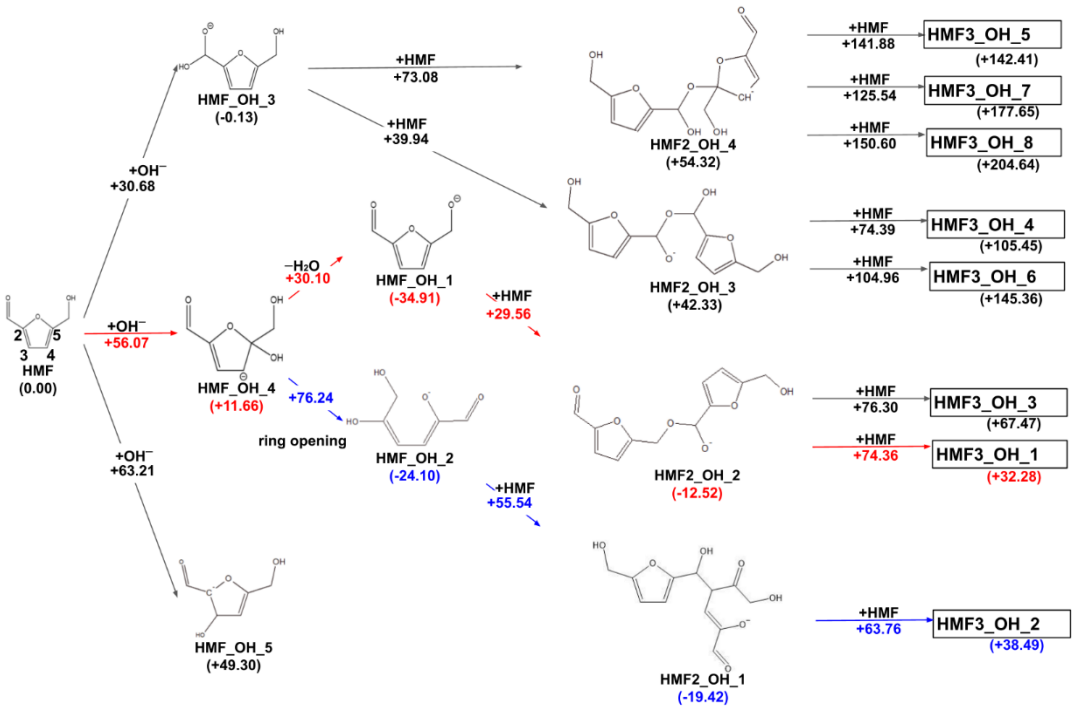

Figure B6. Reaction route map for the initial steps of humin formation under basic conditions; red arrows show the pathway leading to the most stable trimer, and blue arrows show the pathway leading to the most stable dimer.

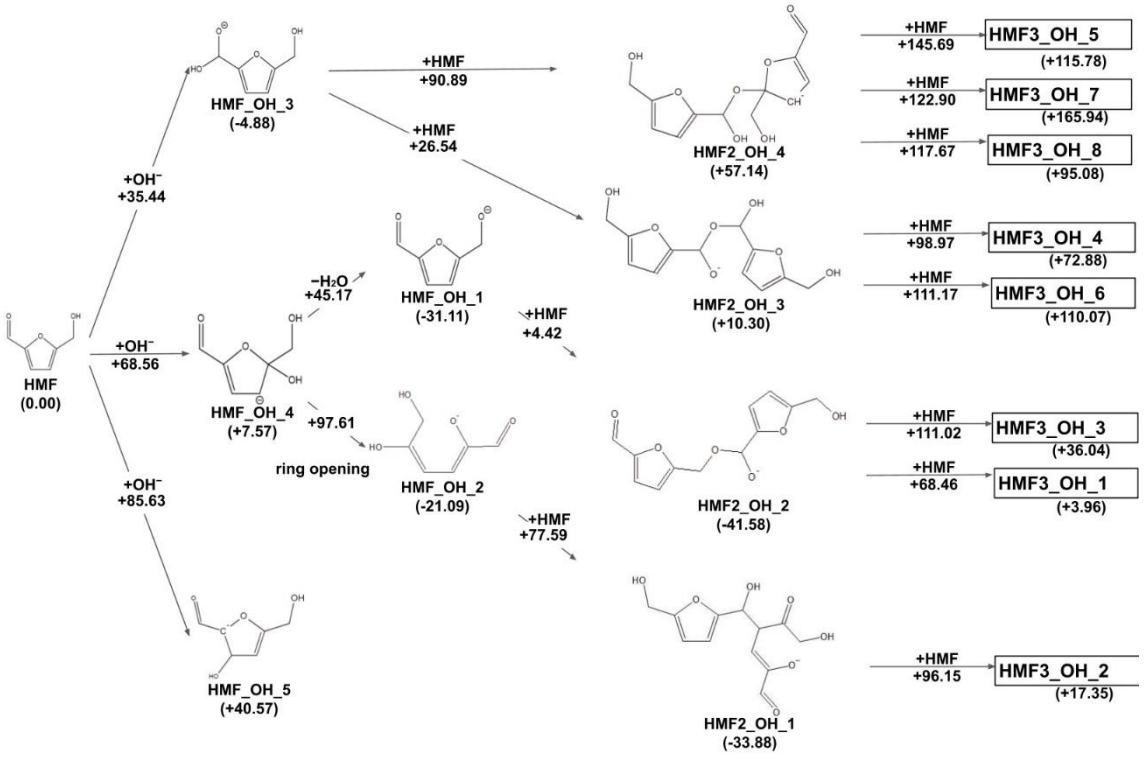

Figure B7. Reaction route map for the initial steps of humin formation under basic conditions calibrated the electronic energy by single-point calculations with the wB97x-D/aug-cc-pVTZ level of theory.

Table B1. Candidates of the initial reactions for which the reaction-path searches were conducted

Candidate	Lowest activation energy /kJ mol^{-1}
HMF + HMF	186.5
HMF + H_2O	209.9
HMF + OH^-	30.7
HMF + O_2 (singlet)	111.6
HMF + O_2 (triplet)	135.7

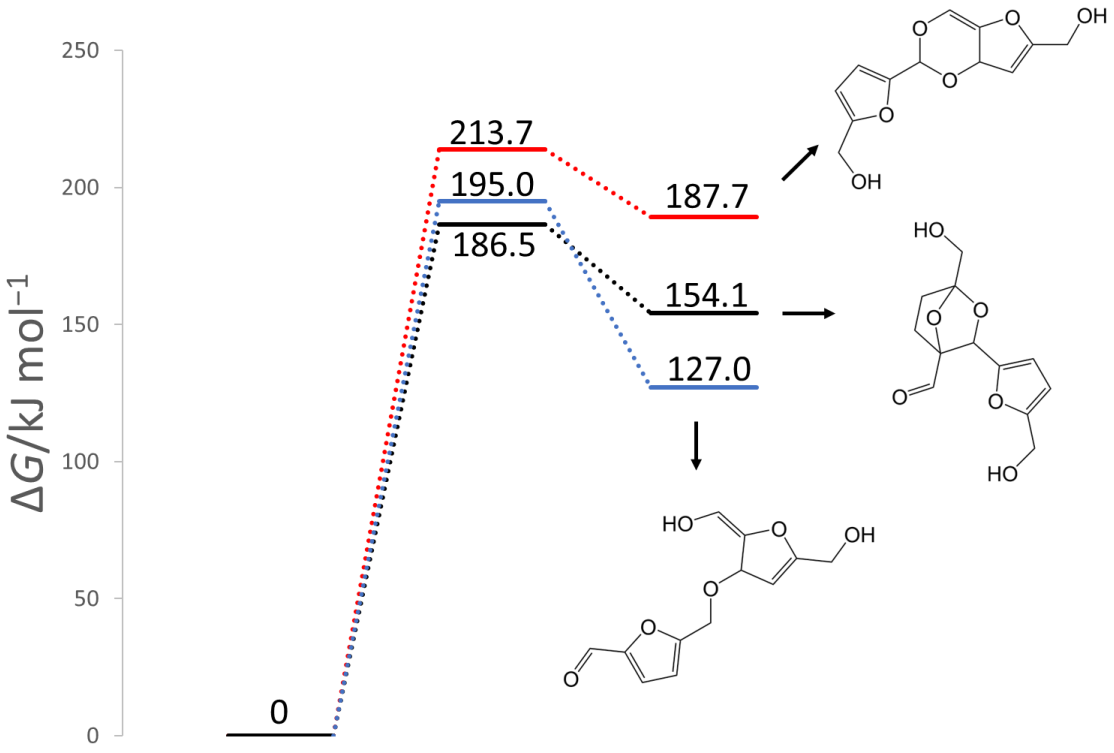

Figure B8. Gibbs energy diagram for the reactions of HMF + HMF under neutral conditions (298 K, 1 atm)

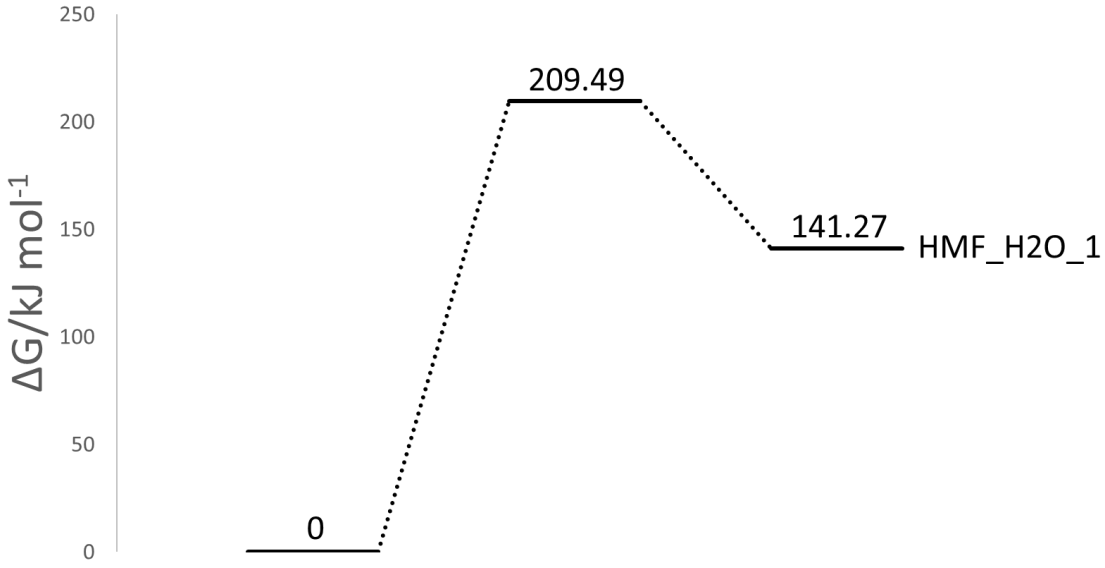

Fig. B9. Gibbs energy diagram or the reaction of HMF + H_2O under neutral conditions (298 K, 1 atm).

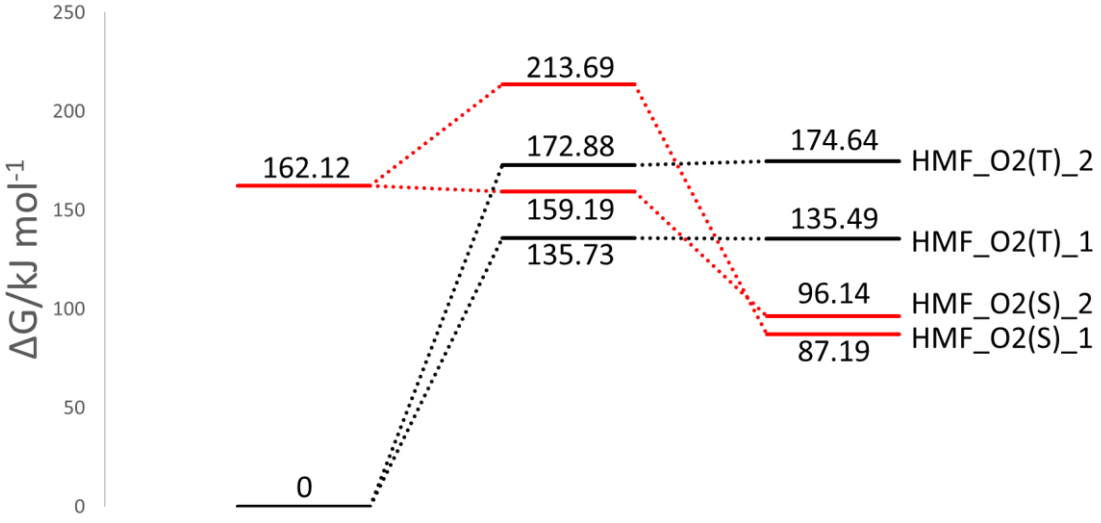

Fig. B10. Gibbs energy diagram for the reactions of HMF + O_2 under neutral conditions (298 K, 1 atm).

2.3.2 Global Reaction Network of Humin Formation

The dimerization process was observed for HMF_OH_1, HMF_OH_2, and HMF_OH_3. the dimerization of HMF_OH_5 was not investigated in this study due to its high minimum barrier height of 196.2 kJ mol^{-1}. The most stable dimer, HMF2_OH_1, was formed through an aldol-like reaction of HMF_OH_2 with HMF, with a barrier height of 55.5 kJ mol^{-1}. The dimerization reaction with the lowest barrier involved HMF_OH_1 + HMF, where O−attacked the formyl group of the new HMF molecule, yielding HMF2_OH_2 with a barrier of 29.6 kJ mol^{-1}. The trend of the lowest activation energy for dimerization at the formyl carbon was similar to that observed for the reaction of HMF + OH^-. These dimers exhibited negative relative energies and are likely important intermediates contributing to humin formation. For HMF_OH_3, two dimerization pathways were identified, leading to HMF2_OH_3 and HMF2_OH_4, with relative energies considerably higher than those of other dimers. Since HMF_OH_3 is a derivative of 5-hydroxymethyl-2-furancarboxylic acid (HMFCA), an intermediate in FDCA synthesis, its formation may not lead to humin formation. Additionally, eight pathways for trimer formation were identified based on the chemical structures shown in Fig. B11. The reaction barriers for trimer formation were all above 60 kJ mol^{-1}, indicating that trimer formation is slower than dimer formation under basic conditions.

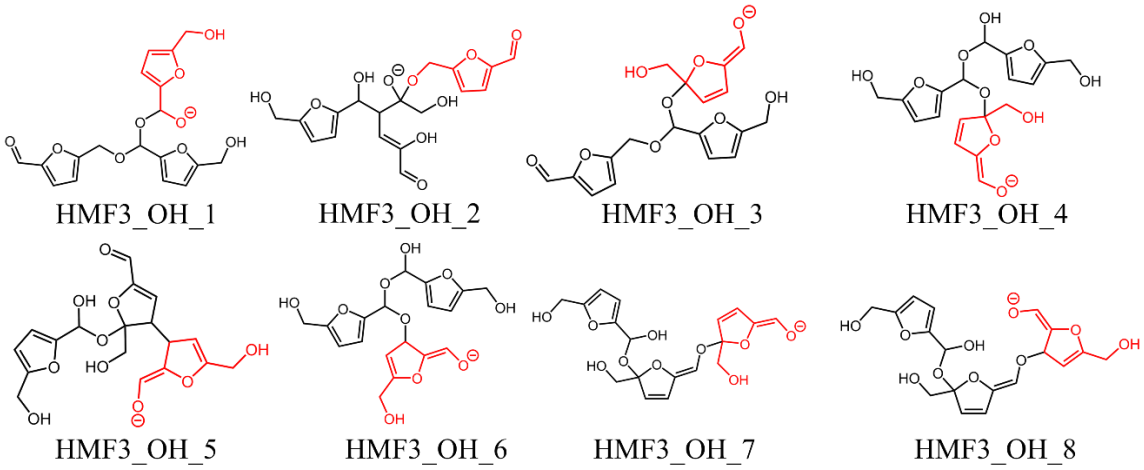

Figure B11. Trimer structures obtained in this study; the last attacking HMF is shown in red.

2.3.3 Effect of Acetal Protection

2.3.3A. Comparison between HMF Hydroxylation and PD-HMF Hydroxylation

In order to investigate the impact of acetal protection, the reaction of PD-HMF (HMF–acetal protected by PD) with OH^- was analyzed computationally. Figure B12 illustrates the hydroxylation pathways at carbons 3 and 5 of the furan ring of PD-HMF and the protected formyl carbon, leading to the formation of PD-HMF_OH_5, PD-HMF_OH_4, and PD-HMF_OH_3, respectively. The protection provided by the acetal group directly hinders hydroxylation at the formyl carbon, resulting in an activation energy of 175.2 kJ mol^{-1} to obtain PD-HMF_OH_3, approximately 145 kJ mol^{-1} higher than the activation energy required to obtain HMF_OH_3. Notably, the acetal protection increased the activation energies for obtaining PD-HMF_OH_4 and PD-HMF_OH_5 by approximately 70 and 100 kJ mol^{-1}, respectively. This increase can be attributed to the loss of π-conjugation with the formyl C=O group, which destabilizes the products. Additionally, this phenomenon can be explained in terms of electrophilicity. Figure B13 presents the results of the natural population analysis [B34,B35] for HMF and PD-HMF, computed at the B3LYP-D3/6-31+G(d) level, with the same IEFPCM solvation model used for the energy calculations. Acetal protection reduced the charge on the carbons at positions 3/5 of the furan ring from -0.159/0.324 to -0.311/0.274, indicating inhibition of the reaction with OH^- at these positions. Thus, acetal protection was theoretically demonstrated to stabilize HMF.

2.3.3B. Discussion on the FDCA formation from HMF and PD-HMF under basic condition

In the conversion of HMF to FDCA, the oxidation of the aldehyde represents the first step. Among the three initial pathways, the activation energy for the formation of HMF_OH_3 was

the lowest. Byproducts of the hydroxylation reaction, particularly HMF_OH_4 and the subsequent HMF_OH_1 and HMF_OH_2, may contribute to humin formation, a side reaction of FDCA formation. Notably, the activation energy for HMF_OH_4 formation was below 60 kJ mol^{-1}, suggesting that this side reaction could occur at room temperature, consistent with the observation that the side reaction becomes more prominent with increasing HMF concentration. Conversely, in the conversion of PD-HMF to FDCA, the initial reaction was not with OH^-, but with the oxidation of alcohol. This indicates that in the reaction involving PD-HMF, all reactions with OH^- observed for HMF were suppressed by acetal protection, leading to the guidance of the initial reaction by alcohol oxidation and avoiding humin formation.

Although I did not consider the catalyst or molecular oxygen on the metal support in this study, the results shown in this paper suggest a deeper understanding of humin formation in various basic solutions and the effect of acetal protection on HMF [B36–B38].

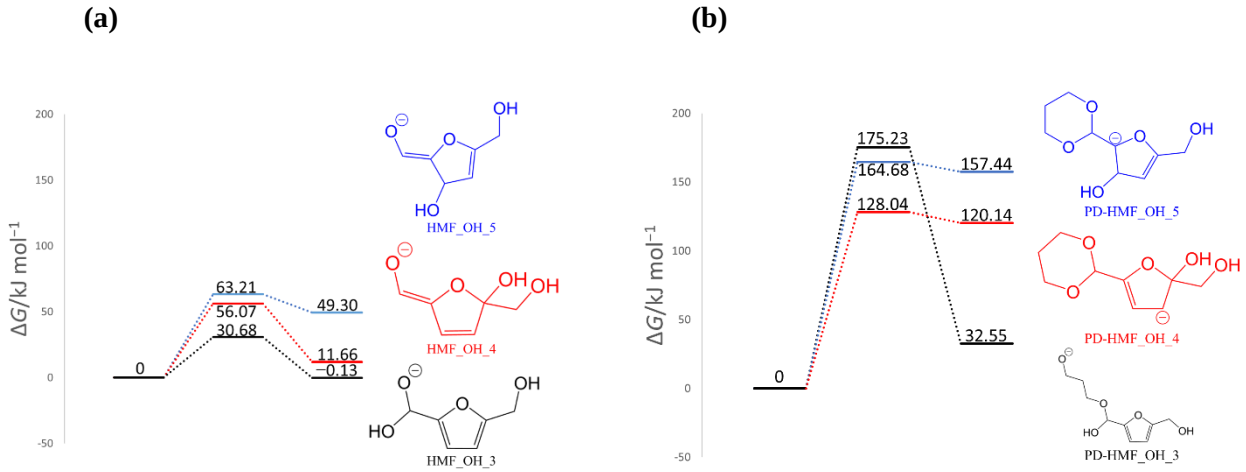

Figure B12. Gibbs energy diagrams of the reactions of (a) HMF + OH^- and (b) PD-HMF + OH^- (298 K, 1 atm)

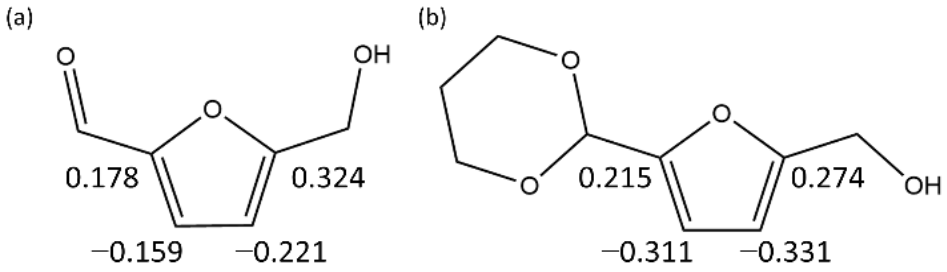

Figure B13. Natural electron populations on furan rings of (a) HMF and (b) PD-HMF

2.4 Conclusion

In this study, I conducted a comprehensive analysis of the reaction pathways leading to humin formation from HMF under basic conditions. Despite reports of humin formation even with pure HMF, the direct addition reaction path (HMF + HMF) was found to have a barrier of at least 186.5 kJ mol^{-1}, suggesting it is unlikely to occur at room temperature. In contrast, the reaction of HMF + OH^- revealed three pathways with barriers of less than 65 kJ mol^{-1}. Notably, HMF_OH_4 emerged as a critical intermediate leading to HMF dimers with a low reaction barrier, while trimerization was found to have a higher barrier than dimerization. Additionally, AFIR calculations were performed using PD-HMF instead of HMF to investigate the impact of acetal protection on reactivity. The results showed a significant increase in activation energies for reactions involving the protected formyl carbon and the other two initial reactions. This finding explains the experimental stability of PD-HMF under basic conditions and strengthens the validity of the initial reaction pathway for humin formation identified in this study. This study underscores the utility of the MC-AFIR method in computational chemistry for elucidating complex, multistep reactions.

2.5 Reference

[B1] de Jong, E.; Stichnothe, H.; Bell, G.; Jørgensen, H. *Bio-Based Chemicals, a 2020 Update*; IEA *Bioenergy*, 2020. 978-1-910154-69-4. Online: http://task42.ieabioenergy.com/wp-content/uploads/2020/02/Bio-based-chemicals-a-2020-update-final-200213.pdf (accessed Dec 17, 2023).

[B2] Burns, C., Higson, A., Hodgson, E. 2016. Five recommendations to kick-start bioeconomy innovation in the UK. . *Biofuels Bioprod. Bioref*. 10:12-16.

[B3] Rajeswari, S., Baskaran, D., Saravanan, P., Rajasimman, M., Rajamohan, N., & Vasseghian, Y. Production of ethanol from biomass–Recent research, scientometric review and future perspectives. 2022, *Fuel*, *317*, 123448.

[B4] de Jong, E.; Dam, M. A.; Sipos, L.; Gruter, G.-J. M. Furandicarboxylic Acid (FDCA), A Versatile Building Block for a Very Interesting Class of Polyesters. In *Biobased Monomers, Polymers and Materials*; Smith, P. B., Gross, R., Eds.; ACS Symposium Series; *American Chemical Society*, 2012; pp 1– 13.September 7, 2022 at 2:52 PM

[B5] M. Sajid, X. Zhao and D. Liu, Production of 2,5-furandicarboxylic acid (FDCA) from 5-hydroxymethylfurfural (HMF): recent progress focusing on the chemical-catalytic routes, Green Chem., 2018, 20, 5427 —5453

[B6] D. Zhao, T. Su, Y. Wang, R. Varma, C. Len, Recent advances in catalytic oxidation of 5-hydroxymethylfurfural, Mol Catal, 2020, 495, Article 111133

[B7] M. G. Davidson, S. Elgie, S. Parsons and T. J. Young, Production of HMF, FDCA and their derived products: a review of life cycle assessment (LCA) and techno-economic analysis (TEA) studies, Green Chem., 2021, 23, 3154 —3171

[B8] Albonetti, S.; Hu, C.; Saravanamurugan, S. (eds.), Green Conversion of HMF, Special Collections in ChemSusChem (2022) https://chemistry-europe.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1864-564X.Green-Conversion-of-HMF. (accessed Augest 10)

[B9] Kim, M; Su, Y.; Fukuoka, A; Hensen, E.J.M.; Nakajima, K. Aerobic Oxidation of 5-(Hydroxymethyl)furfural Cyclic Acetal Enables Selective Furan-2,5-dicarboxylic Acid Formation with CeO_2-Supported Gold Catalyst. *Angew. Chem. Int. Ed.* 2018, *130*, 8367−8371.

[B10] Horvat, J.; Klaic, B.; Metelko, B.; Sunjic, V. Mechanism of levulinic acid formation. *Tetrahedron Lett.* 1985, *26*, 2111–2114.

[B11] Horvat, J.; Klaic, B.; Metelko, B.; Sunjic, V. Mechanism of levulinic acid formation in acid catalyzed hydrolysis of 2-(hydroxymethyl)furan and 5-(hydroxymethyl)-2-furancarboxaldehyde. *Croat. Chem. Acta* 1986, *59*, 429–438.

[B12] Patil, S. K. R.; Lund, C. R. F. Formation and growth of humins via aldol addition and condensation during acid-catalyzed conversion of 5-hydroxymethylfurfural *Energy Fuels* 2011, *25*, 4745–4755.

[B13] Hoang, T. M. C.; van Eck, E. R. H.; Bula, W. P.; Gardeniers, J. G. E.; Lefferts, L.; Seshan, K. Humin based by-products from biomass processing as a potential carbonaceous source for synthesis gas production. *Green Chem.* 2015, *17*, 959–972.

[B14] van Zandvoort, I; Koers, E. J.; Weingarth, M.; Bruijnincx, P. C. A.; Baldus, M.; Weckhuysen, B. M. Structural characterization of ^{13}C-enriched

humins and alkali-treated ^{13}C humins by 2D solid-state NMR. *Green Chem.* 2015, *17*, 4383–4392.

[B15] Zakrzewska, M. E.; Bogel-łukasik, E.; Bogel-łukasik, R. Ionic Liquid-Mediated Formation of 5-Hydroxymethylfurfural - A Promising Biomass-Derived Building Block. *Chem. Rev.* 2011, *111*, 397–417.

[B16] Al Ghatta, A.; Zhou, X.; Casarano, G.; Wilton-Ely, J. D. E. T.; Hallett, J. P. Characterization and Valorization of Humins Produced by HMF Degradation in Ionic Liquids: A Valuable Carbonaceous Material for Antimony Removal. *ACS Sustainable Chem. Eng.* 2021, *9*, 2212–2223.

[B17] Galkin, K. I.; Krivodaeva, E. A.; Romashov, L. V.; Zalesskiy, S. S.; Kachala, V. V.; Burykina, J. V.; Ananikov, V. P.; Critical Influence of 5-Hydroxymethylfurfural Aging and Decomposition on the Utility of Biomass Conversion in Organic Synthesis. *Angew. Chem. Int. Ed.* 2016, *55*, 8338–8342.

[B18] Kim, M; Su, Y.; Aoshima, T; Fukuoka, A; Hensen, E.J.M.; Nakajima, K. Effective Strategy for High-Yield Furan Dicarboxylate Production for Biobased Polyester Applications. *ACS Catal* 2019, 9, 5, 4277-4285

[B19] Maeda, S.; Ohno, K.; Morokuma, K. Systematic Exploration of the Mechanism of Chemical Reactions: The Global Reaction Route Mapping (GRRM) Strategy Using the ADDF and AFIR Methods. *Phys. Chem. Chem. Phys.* 2013, *15*, 3683.

[B20] Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring Transition State Structures for Intramolecular Pathways by the Artificial Force Induced Reaction Method. *J. Comput. Chem.* 2014, *35*, 166−173.

[B21] Maeda, S.; Harabuchi, Y.; Takagi, M.; Taketsugu, T.; Morokuma, K. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *Chem. Rec.* 2016, *16*, 2232−2248.

[B22] Maeda, S.; Harabuchi, Y.; Takagi, M.; Saita, K.; Suzuki, K.; Ichino, T.; Sumiya, Y.; Sugiyama, K.; Ono, Y. Implementation and Performance of the Artificial Force Induced Reaction Method in the GRRM17 Program. *J. Comput. Chem.* 2018, *39*, 233−250.

[B23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian16 (Revision A.03), Gaussian, Inc., Wallingford, CT, 2016.

[B24] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 1993, *98*, 5648.

[B25] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* 1994, *98*, 11623.

[B26] Grimme, S.; Antony, J.; Ehrlich, S.; Kreig, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 2010, *132*, 154104.

[B27] Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* 1971, *54*, 724-728.

[B28] Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* 1972, *56*, 2257-2261.

[B29] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* 1973, *28*, 213-222.

[B30] Clark, T.; Chandrasekhar, J.; Spitznagel, Günther W.; Schleyer, P. v. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* 1983, *4*, 294-301.

[B31] Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* 1997, *101*, 10506-10517.

[B32] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* 2011, *3*, 33.

[B33] The Open Babel Package, version 2.3.2 http://openbabel.org (accessed June 2022)

[B34] A. E. Reed, R. B. Weinstock and F. Weinhold, Natural population analsys, *J. Chem. Phys.*, 1985, 83, 735.

[B35] A. E. Reed, L. A. Curtiss and F. Weinhold, Intermolecular interactions from a natural bond orbital, donor acceptor viewpoint, *Chem. Rev.*, 1988, 88, 899.

[B36] Ferdy J. A. G. Coumans, Zhanna Overchenko, Jan J. Wiesfeld, Nikolay Kosinov, Kiyotaka Nakajima, and Emiel J. M. Hensen, Protection strategies for the Conversion of Biobased Furanics to Chemical Building Blocks, *ACS Sustainable Chemistry & Engineering* 2022 10 (10), 3116-3130

[B37] T. Boonyakarn, J. J. Wiesfeld, M. Asakawa, L. Chen, A. Fukuoka, E. J. M. Hensen, K. Nakajima, Effective Oxidation of 5-Hydoroxymethylfurfural to 2,5-Diformylfuran by an Acetal Protection Strategy, *ChemSusChem* 2022, 15, e202200059;

[B38] J. J. Wiesfeld, M. Asakawa, T. Aoshima, A. Fukuoka, E. J. M. Hensen, K. Nakajima, A Catalytic Strategy for Selective Production of 5-Formylfuran-2-carboxylic Acid and Furan-2,5-dicarboxylic Acid, *ChemCatChem* 2022, 14, e202200191;

3. Conformation Determination of Polyketone Compounds Using Molecular Dynamics

3.1 Introduction

3.1.2 Length-controlled organic oligomers : potential of Aliphatic polyketones

Length-controlled organic oligomers, which were synthesized by Inokuma and co-workers[C1] are aliphatic polyketones with string-like flexibility. The origin of its flexibility comes from the alternating sequence of 1,3- and 1,4-diketones, which can achieve a wide range of conformations as it does not have a rigid planar structure like vicinal polyketones. An interesting behavior of these oligomers is that they show assembly with metal cations in gas phase. This is a rope-wrapping dynamics that cannot be confirmed by observation in a small molecule or polymers. Elucidating the detailed conformation of this assembly is an interesting topic that will guide the precise control of molecular dynamics of this molecule, which exhibits unique behavior. There are several methods for determining the three-dimensional structure of molecules in the solid and liquid phases, such as NMR and X-ray diffraction. Nevertheless, structural analysis in the gas phase is still limited. Ion mobility mass spectrometry (IM-MS)[C2-7] is a method to determine the three-dimensional structure of molecules in the gas phase using collision cross sections (CCS). Specifically, the collisions of molecules within the detector are caused, and the momentum change due to the collision is detected as the arrival time. Arrival times are converted to CCS by the Mason-Schamp equation[C8]. The traditional approach to determining structure from CCS obtained by IM-MS

consisted of preparing candidate structures, performing *in silico* CCS calculations and comparing them with the obtained CCS. This workflow is effective for IM-MS spectra with unimodal distribution. However, IM-MS spectra of polyketones with large number (>8) were found to have the multimodal distribution. In addition, the many potential of ion-carbonyl coordination configurations made it difficult to arbitrarily prepare structures. Thereafter, I need to develop specific workflow for this analysis.

3.1.2 Ion Mobility Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) is a spectroscopic method that combines mass spectrometry (MS) and ion mobility spectrometry (IMS). It is widely used to separate analyte ions in gas phase, especially mixtures of ions with the same molecular weight. Although structural and stereochemical isomers that show the same signal patterns in MS, they can be distinguished by difference in the ion mobility arrival times in IM-MS. Experimental arrival time can be treated as a structural feature by converting it to collision cross section (CCS). The Mason-Schamp equation is often used to calculate the analyte ion's CCS from arrival time, where Ω is the CCS, z is ion charge, e is elementary charge, k_b is Boltzmann's constant, T is drift region temperature, m_I is mass of an analyte ion, m_N is mass of buffer gas, L is length of ion mobility cell, P is pressure in cell, N is number of buffe gas molecules in a unit constant volume, and t_D is arrival time.

$$\Omega = \frac{\sqrt{18\pi}}{16}\frac{ze}{\sqrt{k_b T}}\sqrt{\frac{1}{m_I}+\frac{1}{m_N}}\frac{t_D E}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N}$$

To analyze experimental CCSs obtained here, it is necessary to compare them with accurate CCSs in the database or calculated CCSs.

3.1.3 *in silico* CCS Calculation

in silico CCS calculation method is essential to discuss the structure from the experimentally measured CCS. There are broadly categorized into three approaches.

First approach is straightforward method for cross sections that consider the projected area of the candidate structure. It called projection approximation (PA) method and can be performed faster once the structure has been determined. On the other hand, it tend to underestimate the CCS value relative to the measured one because of the lack of consideration for collision effects with gases. Second approach is simulation of collisions between ion and buffer gas. It usually performed by classical molecular dynamics calculations and fully evaluate ion and buffer gas interactions. It called trajectory method (TM) since MD trajectory directly leads to CCS value. The last approach is a recently appearing method using machine learning to make predictions. The first two approaches are implemented in common CCS calculation methods (e.g., Mobcal[C9] and IMoS[C10]) and widely used. A significant aspect of using these methods is that they compute the CCS of a single point structure, i.e., the result depends on the input structure. This means an appropriate initial structure is necessary to obtain reliable calculation results.

3.2 Computational Method

The computational workflow adopted in the present study consists of three steps: conformation sampling, CCS evaluation, and clustering. Details are as follows.

3.2.1 Conformation Sampling with DCDFTB program

As a first step, the author performed MD simulations using the density-functional tight-binding(DFTB) Hamiltonian[C11], a semi empirical Hamiltonian that allows rapid MD simulation with quantum chemical potential. The author used the DCDFTBMD program and employed the 3ob third-order parameter sets. MD simulations were performed at 398 K adopting Nosé-Hoover chain thermostat with a time step of 0.25 fs. Inital structures was created as follows. First. MD was performed with only aliphatic polyketone chains without a sodium ion. Next, 10 structures was randomly extracted from the trajectory, and a sodium ion was placed in rectangular volume enclosing the string. The author extracted snapshots every 200 fs from the trajectories, yielding a total of 6000 structures.

3.2.2 *in silico* CCS Calculation refined from PA Method

As mentioned in the previous section, the most accurate method for calculating collision cross sections in silico is the TM. However, in this study, we employed the PA method for calculating cross sections, considering the need to compute cross sections for tens of thousands of structures and the importance of qualitatively reproducing the spectra. Nevertheless, the naive projection area method showed a large deviation from experimental values. Therefore, I adopted a modified PA method that includes a correction for the volume of a simple collision gas outside the molecule (Fig.C1).

3.2.3 Clustering

To investigate the characteristics of the obtained structures, I performed clustering based on structural features. We utilized the CCS and distance between Na^+ ion and O atoms as descriptors and employed the Mean Shift method[C12]. The mean-shift clustering is similar to the *K*-means clustering, but it does not require the number of clusters to be specified *a priori*. This was necessary because the number of distinct structural species within the structural ensemble was unknown. The results of the clustering were visualized using Principal Component Analysis (PCA), and the validity of the clustering was discussed.

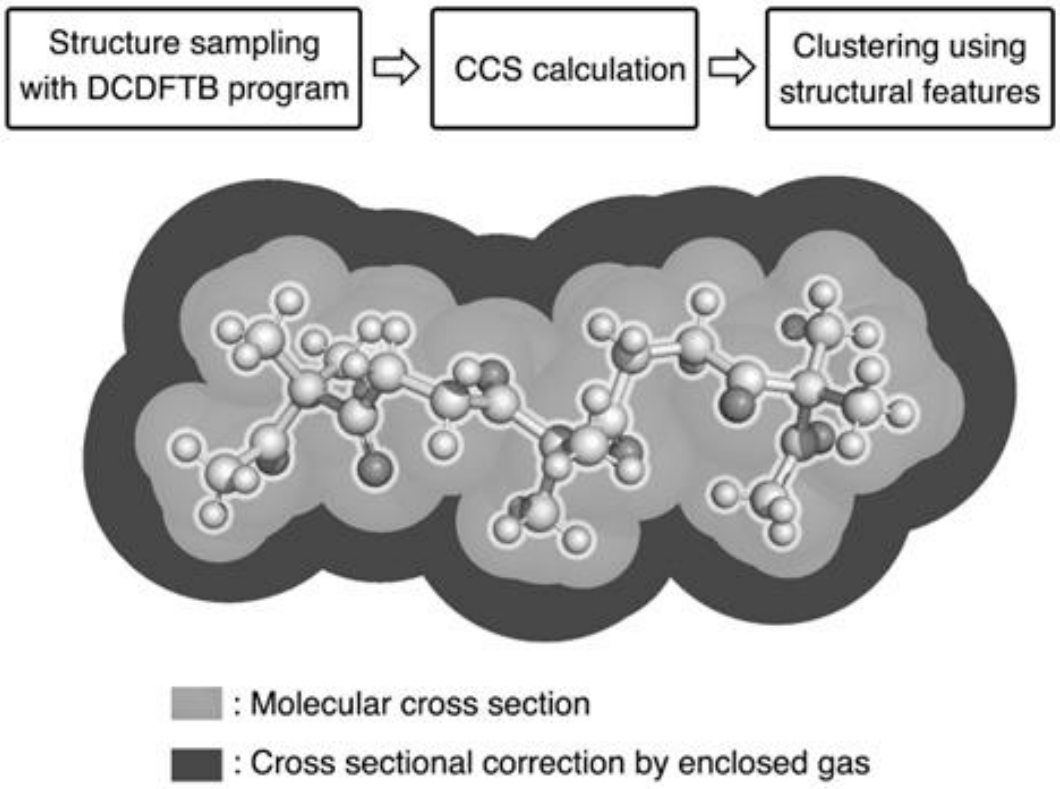

Fig. C1. Schmatic illustration of correction of CCS for buffer gas.

3.3 Result and Discussion

3.3.1 CCS Calculation for Polyketone hexamer

The MD simulations conducted for [1 + Na]+ yielded 11,000 conformations, with calculated CCS values ranging from 255 to 330 $Å^2$. While these values differ from those obtained through IM-MS, the MD calculations effectively replicated the experimental CCS distribution for [1 + Na]+. The CCS versus frequency plot exhibited an almost unimodal distribution (Figure C2.b). Most of the simulated conformations indicated that the Na+ ion was coordinated by 5 to 6 oxygen atoms, including the terminal carbonyl oxygen O1 (Figure C3). The other terminal portion of 1 did not contribute to cation coordination, suggesting that the relatively broad distribution of CCS for [1 + Na]+ arises from the conformational flexibility of the uncoordinated polyketone terminal.

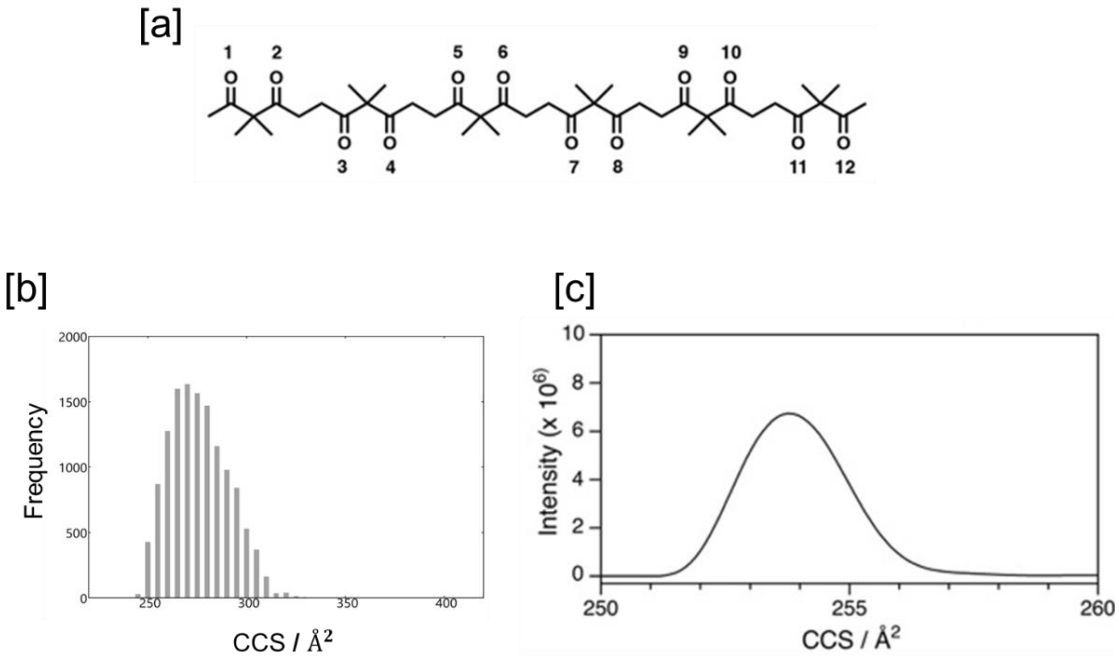

Fig. C2. a)Structural formula of a polyketone 6-mer b)Theoretical CCS distribution of a polyketone 6-mer c)The experimental spectrum of a polyketone 6-mer

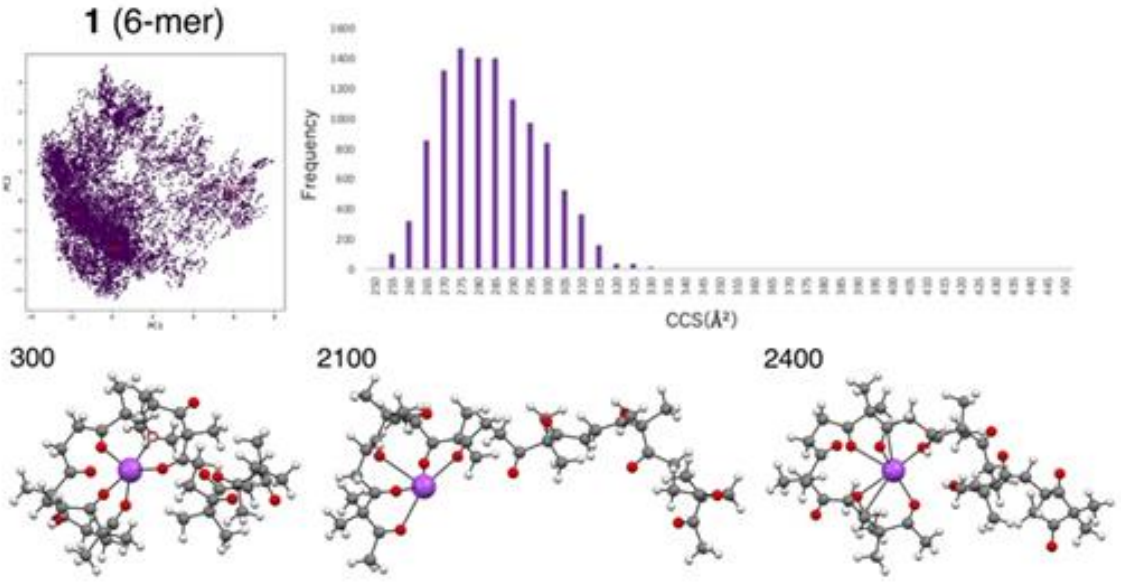

Fig. C3. PCA visualization of 11000 6-mers and typical structures

3.3.3 CCS Calculation for Polyketone octamer

The MD simulations for [**1** + Na]$^+$ generated 6000 conformations and their calculated CCS values varied in the range of 300–420 Å^2. Although there are shift in the range between calculation and experimental data, calculated CCS distribution reproduces the multimodal distribution. The PCA visualization of the octamer was shown in Fig.C4. Based on the coordination and CCS features, the MD-simulated structure for [**1** + Na]$^+$ cation complexes were classified into cluster 1-4. Complexes in the cluster 1 typically chelate the Na$^+$ ion with 5 oxygen atoms of **1** at the O3, O4, O5, O9, and O10 positions. The non-coordinating sites of the polyketone chain (O11–O16) were more likely to be folded by intramolecular hydrogen bonding interactions, thus showing relatively sharp peaks of calculated CCS mostly in the range of 310–350 Å^2.

Complexes in the cluster 2 chelated the Na$^+$ ion at the central sites of the polyketone chain using 4–6 oxygen atoms: *e.g.* O4, O5, O10, O12; O5, O6, O9, O10, O11; O5, O6, O7, O8, O11, O12; for tetra-, penta-, and hexacoordinated complexes, respectively. Because the terminal parts of the polyketone chain did not contribute to the coordination, various conformations were observed to show rather broad distribution of CCS in the range of 310–370 Å^2. Cluster 3 exhibited the narrowest distribution of CCS in the range of 305-330 Å^2. Complexes in this cluster adopted macrocyclic conformations characterized by the cooperative Na$^+$ ion coordination using terminal carbonyl oxygen atoms O1 and O16. The narrow distribution of CCS is attributed to the loss of degrees of freedom due to the formation of macrocyclic conformations. The coordination modes of the complexes in cluster 4 were almost similar to those in cluster 1. However, the

uncoordinated part of the polyketone chain was not well-folded, and different chain conformations were found, giving a wide distribution of CCS.

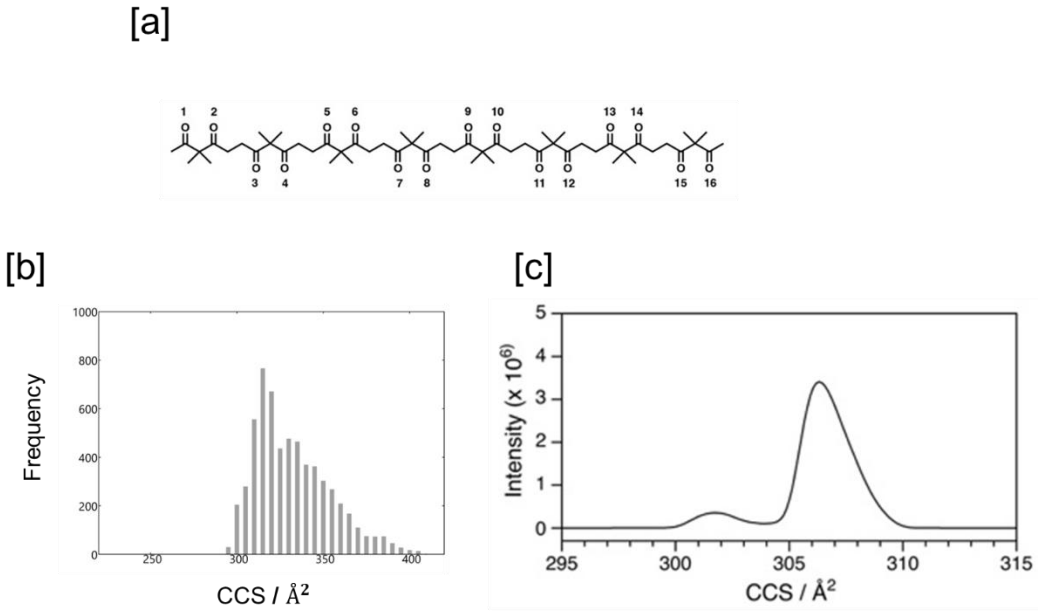

Fig. C4. a)Structural formula of a polyketone octamer b)Theoretical CCS distribution of a polyketone octamer c)The experimental spectrum of a polyketone octamer

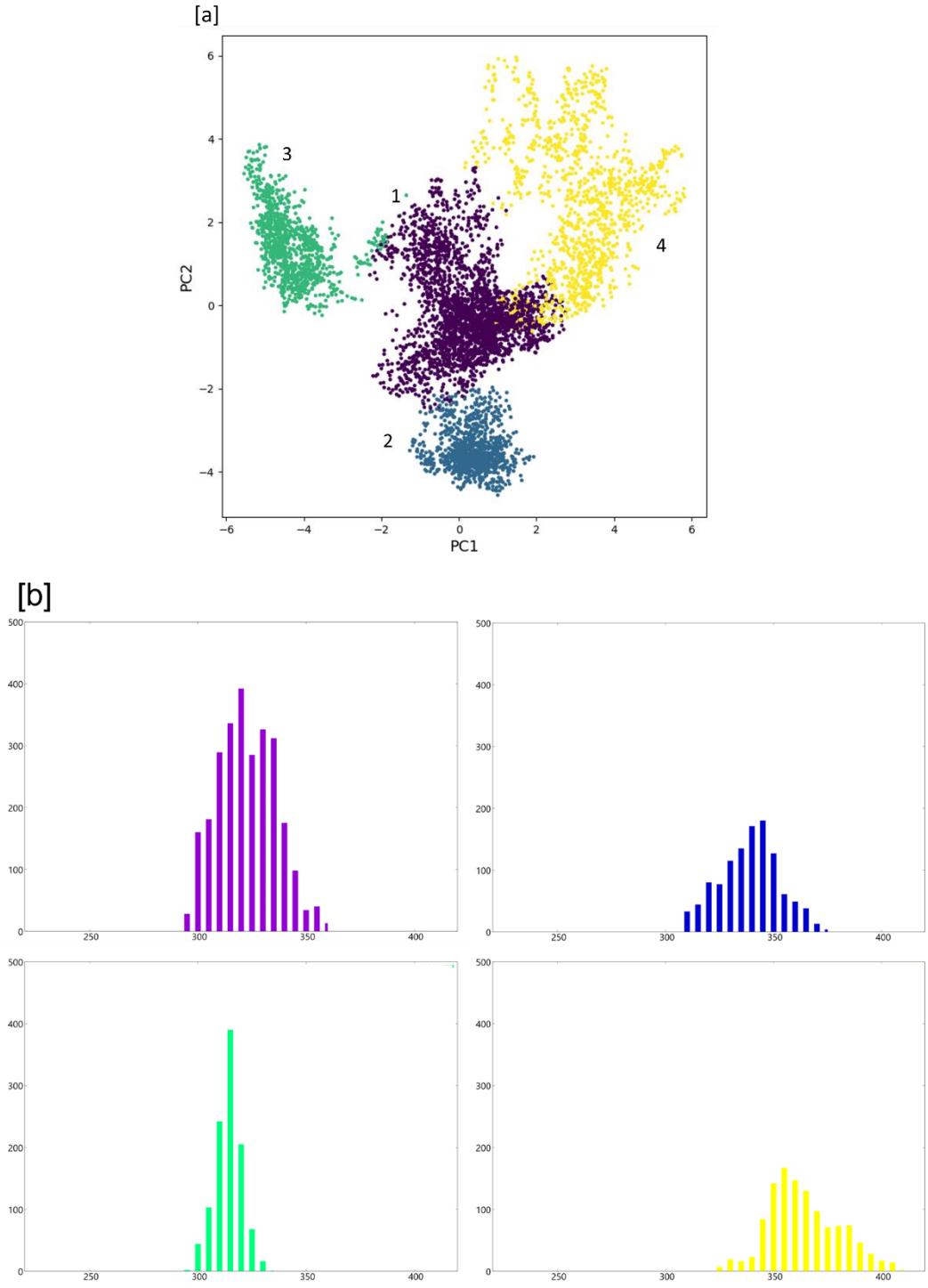

Fig. C5. a)PCA visualization of 6000 8-mers b) CCS distribution for each cluster

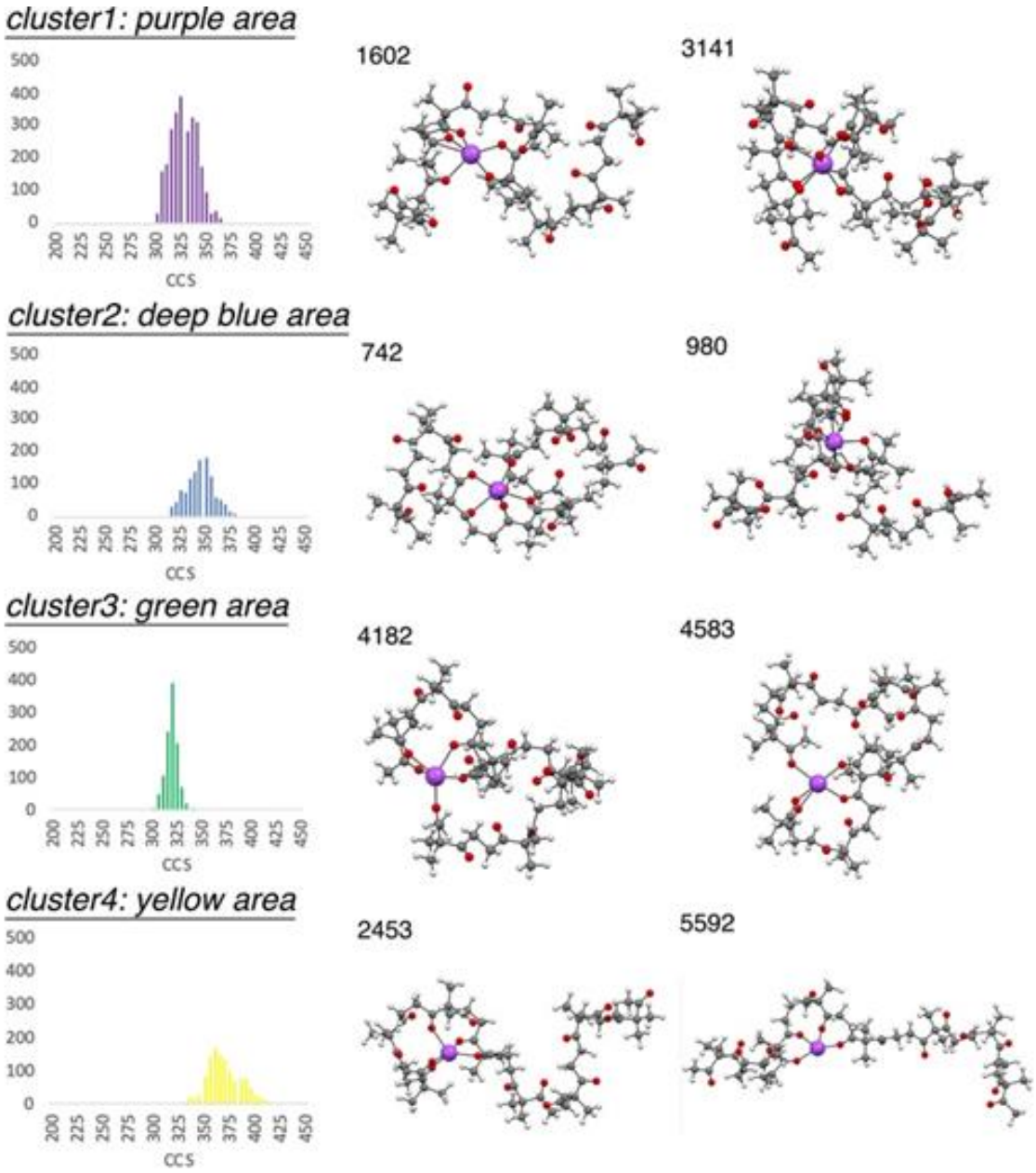

Fig. C6. Typical structure of each cluster

In ion mobility-mass spectrometry (IM-MS) experiments, detected ions traverse the drift tube over milliseconds, necessitating the averaging of observed collision cross-section (CCS) values over the flight time. To approximate this time-averaging effect in our simulations, we computed the average CCS values for 150 consecutive snapshot structures from each trajectory, corresponding to an average over 29.8 picoseconds. The resulting CCS distribution, compared with the non-averaged distribution in Fig. C7, demonstrates that time averaging serves to narrow the distribution of each peak. While our approach does not fully replicate the magnitude and shape of the experimental distribution, partially due to the limitations of the averaging time feasible in MD simulations, the CCS values for the snapshot structures were adjusted to qualitatively reproduce the distribution, including its bimodal structure.

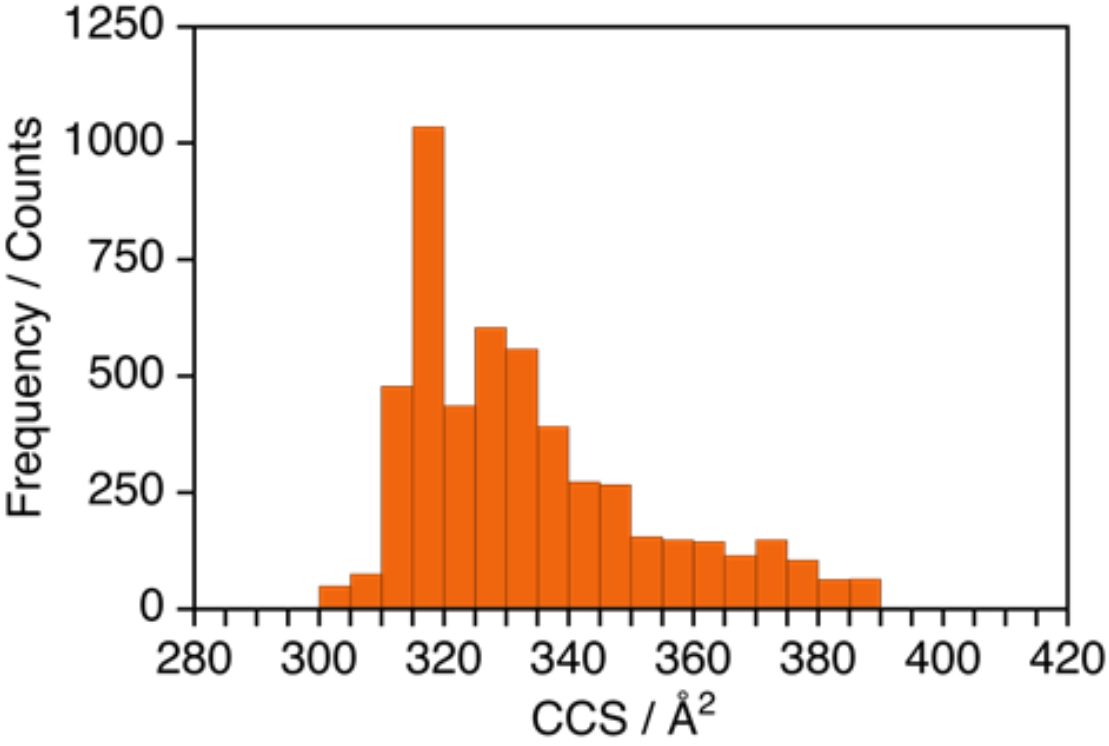

Fig. C7. Time-averaged CCS distribution of a polyketone octamer

3.4 Conclusion

In this study, the author developed a MD-based conformational simulation method that can reproduce CCS distribution of structurally flexible molecular ions. This method allowed analyzing the length-dependent sodium cation binding modes of discrete polyketones oligomer in gas phase. Even CCS distributions obtained by IM-MS are broad and multimodal, comprehensive conformational analysis could provide reasonable understanding of target ion conformations. While there are many analyzed examples and CCS simulation methods for proteins, this method would be arguably applicable to the overall spectrum shape of rather small and structurally flexible molecules. Therefore, the MD-simulation protocol would make IM-MS analysis more useful and user-friendly as a gas phase structural analysis.

3.5 Reference

[C1] M. Uesaka, Y. Saito, S. Yoshioka, Y. Domoto, M. Fujita, Y. Inokuma, Oligoacetylacetones as Shapable Carbon Chains and Their Transformation to Oligoimines for Construction of Metal-organic Architectures, *Commun. Chem.*, **2018**, 1, Article Number 23.

[C2] E. Christofi, P. Barran, *Chem. Rev.*, **2023**, *123*, 2902–2949.

[C3] L. Liu, Z. Wang, Q. Zhang, Y. Mei, L. Li, H. Liu, Z. Wang, L. Yang, *Anal. Chem.*, **2023**, *95*, 134–151.

[C4] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V. Robinson, *Nat. Protoc.*, **2008**, *3*, 1139–1152.

[C5] Z. Zhou, M. Luo, X. Chen, Y. Yin, X. Xiong, R. Wang, Z.-J. Zhu, *Nat. Commun.*, **2020**, *11*, 4334.

[C6] M. F. Mesleh et al., *J.Phys. Chem.* **1996**, 100, 16082.

[C7] C. Larriba, C. J. Hogan, Jr. Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A,* **2013,** 117, 3887−3901

[C8] E. A. Mason and H. W. Schamp, Jr., *Ann. Phys.*,**1958,** 4(3), 233 .

[C9] a) M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, Structural Information from Ion Mobility Measurements: Effects of the Long Range Potential, J. Phys. Chem. 1996, 100, 16082-16086; *Erratum, J. Phys. Chem.* A **1997**, 101, 968. b)A. A. Shvartsburg and M. F. Jarrold, An Exact Hard Spheres Scattering Model for the Mobilities of Polyatomic Ions, *Chem. Phys. Lett.* **1996**, 261, 86-91.

[C10] a) Larriba-Andaluz, C., Hogan, C.J.: Collision cross -ection calculations for polyatomic ions considering rotating diatomic/linear gas molecules. *J. Chem. Phys.*, **2014,** 141, 194107. b) Ouyang, H., Larriba-Andaluz, C., Oberreit, D.R., Hogan, C.J.: The collision cross-sections of iodide salt cluster ions in air via differential mobility analysis-mass spectrometry. *J. Am. Soc. Mass Spectrom.* 24, 1833–1847 (2013). c) Larriba-Andaluz, C., Hogan, C.: Novel interfaced approach to mobility calculations with diffuse scattering and Maxwell rotational distributions for diatomic gases in the free molecular regime. Abstract Paper *Am. Chem.,* **2013,** S 246. d) Larriba, C., Hogan, C.J.: Ion mobilities in diatomic gases: measurement versus prediction with non-specular scattering models. *J. Phys. Chem. A.* **2013,** 117, 3887–3901. e) Larriba, C., Hogan, C.J.: Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation. *J. Comput. Phys.* **2013,** 251, 344–363.

[C11] M. Gaus, Q. Cui and M. Elstner : *J. Chem. Theory Comput.,* **2011,** 7, 931

[C12] Y. Cheng, *IEEE Trans. Pattern Anal. and Machine Intell.,* **1995**, 17, 790–799.

4. General Conclusion

I did theoretical research on two organic molecules in this doctoral thesis. In Chapter 2, I employed the global reaction route map (GRRM) method for analyzing amorphous oligomerization of 5-(hydroxymethyl)furfural (HMF). The artificial force induced reaction (AFIR) method was applied step by step to the oligomerization system in which the number of atoms in the system increases incrementally to investigate the reaction, and key reactions in the reaction path network of HMF oligomerization was specified. Hydroxylation of the carboxyl group of the HMF did not lead to much oligomerization and hydroxylation of furan ring is important for generation of variety intermediate. Although some initial reactions were predicted based on organic chemical insights, this is the first example to demonstrate that the other competing reactions do not exist based on systematic analysis. In Chapter 3, I develop a theoretical calculation scheme of collision cross section to identify the conformation of aliphatic polyketones with multiple ion mobility mass spectrometry peaks. The scheme consists of fast molecular dynamics (MD) sampling using semi-empirical methods, CCS evaluation using my original Python program and clustering method to determine the number of clusters *a priori.* This scheme successfully reproduced both the one-peak polyketones hexamer spectrum and the two-peak polyketones octamer spectrum. This study implies that the conformation can be identified from MD sampling for systems with many metastable conformations and IM-MS peaks.

These studies are applications showing that the advance of computational chemistry had made it possible to analyze complex systems that were previously challenging, and also cases in which the combination of experimental chemistry, computational

chemistry, and informatics has produced new insights. I hope that the outcome of this thesis will contribute to the understanding of the related field, and look forward to further increasing the number of such a cross-field application.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to Professor Tetsuya Taketsugu. His generous support was indispensable for the progress of my study.

I would also like extend my gratitude to Associate Professor Masato Kobayashi. He provided me with insightful advice, even in challenging times, patiently waited for progress and guided me in the right direction.

Furthermore, I would like to express my gratitude to the all staffs of Quantum Chemistry Laboratory. They supported me to ensure that my research stayed on track. I am deeply thankful for their assistance.

I think that my research life was enriched by the kindness of the laboratory members. Thank you all for your kindness.

I would like to express my gratitude to Professor Kiyotaka Nakajima for providing precise advice from an experimental perspective on Chapter 2, and to Professor Yasuhide Inokuma and Dr. Kazuaki Ohara for their support from the experimental and spectroscopic perspectives on Chapter 3.

I am grateful to the Institute for Quantum Chemical Exploration for the IQCE Fellowships for Young Scientists. Also The financial support provided by the Hokkaido University DX Doctoral Fellowship of MEXT was indispensable for the smooth progress of my research.

Finally, I would like to thank my family for respecting and supporting my choice.