Physicochemical Prediction of Metabolite Fragmentation in Electrospray Tandem Mass Spectrometry

TANAKA WATARU

Doctor of Philosophy

Department of Genetics
School of Life Science
SOKENDAI (The Graduate University for Advanced Studies)

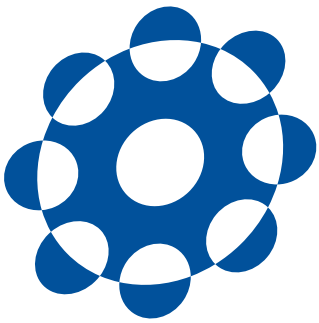
Physicochemical Prediction of Metabolite Fragmentation in Electrospray Tandem Mass Spectrometry

by

Tanaka, Wataru

Dissertation for

Doctor of Philosophy



Department of Genetics

School of Life Science

SOKENDAI (The Graduate University for Advanced Studies)

January 2018

Acknowledgments

First of all, I would like to thank Prof. Masanori Arita for supervising me during my doctoral course as well as giving me a wonderful opportunity to begin my metabolomics studies. His insightful suggestion and discussion strongly promoted my studies. I also express my gratitude to Dr. Hiroshi Tsugawa in RIKEN Center for Sustainable Resource Science (CSRS) for providing me an occasion to cooperate on the excellent studies. I learned much about metabolomics from the cooperative works with him. Members in Laboratory of Biological Networks also gave me daily discussion and support for my works.

I also would like to thank Prof. Kazuhiro Maeshima, Prof. Shin-ya Miyagishima, Prof. Yasukazu Nakamura, Prof. Akatsuki Kimura, and Associate Professor Yukiko Yamazaki for giving me beneficial advice as SOKENDAI Progress Committee members; even though my studies are out of their own fields. My appreciation (in advance) is also toward constructive discussion on my dissertation suggested by the above four professors and Dr. Masami Hirai in RIKEN CSRS.

I received aid from the Short-Stay Study Abroad Program of SOKENDAI to visit University of California, Davis, in 2016. Thanks to that, I could join International Summer Session in Metabolomics, which was held for two weeks, and learn from basics to the forefront of metabolomics with researchers from around the world. It was a precious experience that encouraged me in metabolomics works.

I appreciate a lot of leading I have received in my research career. In my master's course and undergraduate period, I acquired computational skills in Computational Biophysics Group in University of Tsukuba with directed by Dr. Mitsuo Shoji and Prof. Yasuteru Shigeta. The indispensable base of this dissertation was built there. My research career began in National Institute of Technology, Numazu College. Associate

Professor Chikahiro Egami, presently in Tokyo University of Agriculture, demonstrated the first instance of doing science to me. My fundamental research interest was obtained from the study of mathematical biology I worked on under him.

Finally, I deeply thank my family for supporting me for everything throughout my life including the doctoral course.

Abstract

The word ‘metabolome’ means the whole set of metabolites existing in an organism. To diagnose disease or examine influence of drugs and diets, metabolome in human urine, exhaled air, blood, etc. have been measured in various studies. Functional genomics also often utilizes metabolome as the endpoint of cellular regulation which responds to genotype and environment.

One of the major techniques to measure metabolome is mass spectrometry (MS). MS distinguishes metabolites according to their mass in high sensitivity. To obtain detailed structural information of metabolites, especially to classify structural isomers, tandem mass spectrometry (MS/MS) is utilized. Each metabolite is decomposed into fragments which correspond to its substructures. Mass and abundance of each fragment is described as an MS/MS spectrum. As the fragmentation pattern depends on the metabolite structure, matching of the measured MS/MS spectrum with a standard one, which is stored in databases, is an important criterion to identify metabolites.

Current bottleneck of metabolite identification is insufficiency of MS/MS spectral databases. In a metabolome measurement, usually only up to $\sim$ 20% of detected metabolites are identified because of absence of matching standard MS/MS spectra. Since the standard spectra are measured from purified metabolites, there are limitations to obtain MS/MS spectra of all existing metabolites by experimental measurements.

To overcome that, many research groups have tried to build a theoretical MS/MS spectral library independent of the measurements. Several groups, including us, have achieved that, but mechanisms of metabolite fragmentation are not fully clarified. Without mechanistic insights, prediction of general metabolite fragmentation will not be realized. To reach the mechanistic prediction of metabolite fragments, I will discuss utilization of physicochemical calculation. The aim of the present study is to construct

a physicochemical strategy to predict metabolite fragmentation in MS/MS.

At first, I examine ability of physicochemical calculation to determine fragment structures. The subjects to calculate are putative substructures annotated on peaks of MS/MS spectra with matching compositional formulae. The compositional formulae count in movement of hydrogen atoms along the fragmentation, which is formalized as hydrogen rearrangement (HR) rules. Comparison of enthalpy calculated with computational chemistry between bound and fragmented state provides evidence that the putative fragments indeed arise in experimental measurements with feasibly low activation energy. Furthermore, several putative fragments which have unusual electronic structure, such as radicals, are proved sufficiently stabilized by broadly distributed electronic structures confirmed by the calculation. The physicochemical calculation reveals energetic stability of fragments and the origin of the stability by elucidating electronic structures on metabolites.

As an extension of the energetic analyses, I will present metabolite identification without standard MS/MS spectra. The metabolite to identify is a lipid, β-hydroxyl ceramide, measured with negative-ion mode MS/MS. It has structural isomers such as an α-hydroxyl ceramide and a ceramide with phytosphingosine. Since the β-hydroxyl ceramide has no authentic standard, its standard MS/MS spectrum is not available in any database. Chromatographic separation and positive-ion mode MS/MS discriminate the β-hydroxyl ceramide from most of its structural isomers. However, another isomer, γ-hydroxyl ceramide, remains unseparated. The problem is whether β- and γ-hydroxyl ceramides can produce the diagnostic fragment experimentally measured in negative-ion mode MS/MS. I simulate the entire fragmentation processes of the ceramides by computational chemistry. The simulation reveals that the β-hydroxyl ceramide smoothly produces the measured fragment, while the γ-hydroxyl ceramide hardly produces it because of the high energy barrier. From that, we identify metabolite structure of the measured MS/MS spectrum as the β-hydroxyl ceramide without standard spectra.

The investigation stated above uncovers the energetic influence of chemical groups for bond cleavage. A chemical group next to the cleaved bond fairly lowers the energy barrier of the fragmentation.

Finally, I integrate the effects of chemical groups into spectral prediction as bonding pattern assignment. Each chemical bond composing a metabolite is assigned a bonding

pattern which consists of two bound atoms and neighboring chemical groups. Precise cleavage activation energy of the bonding pattern is calculated with physicochemical simulation of its fragmentation process. By tracing low-energy cleavage steps, a theoretical MS/MS spectrum is predicted. On a dipeptide molecule, its experimental standard MS/MS spectrum is successfully reproduced by this strategy. Physicochemical calculation is proved promising to predict MS/MS spectra only from metabolite structures. Details of the prediction as well as several improvements suggested by another molecule will be discussed in this dissertation.

To build a theoretical MS/MS spectral library which can cope with the enormous variety of metabolites, activation energy of all bonding patterns need to be pre-computed and stored in a form of library. With the library, each bond composing a metabolite can be assigned its cleavage activation energy with its bonding pattern recognized. Then fragmentation of the metabolite is predicted by tracing low-energy cleavage to produce a theoretical MS/MS spectrum. Remaining problems to realize that will be discussed at the end of this dissertation.

Publication notes

- Chapter 2 is based on the paper [1].
- Chapter 3 is based on the paper [2].
- We have submitted the work in Chapter 4.

Contents

List of Figures

List of Tables

1 General Introduction

1.1 The beginnings of metabolite profiling

Metabolic profiles began to be measured in the late 1940s. However, the measurement was not tangible and needed tremendous effort until chromatographic technologies advanced sufficiently in the late 1960s [3]. In 1971, Pauling *et al.* quantified about 250 substances from samples of breath and about 280 substances from urine vapor by using gas chromatography (GC) [4]. This is one of the earliest study on quantitative metabolite profiling. A few years later, he and his coworkers identified 42 metabolites from urine vapor with mass spectrometry (MS) coupled with gas chromatography (GC-MS) [5].

In the purpose of developing diagnosis procedures of diseases, a number of groups began to quantify and identify metabolites from biological samples by using GC-MS from the late 1970s [6, 7]. Around 2000, Phillips *et al.* discovered difference in presence of volatile metabolites included in breath between schizophrenia patients and normal people [8]. Poli *et al.* also analyzed volatile organic compounds (VOC) in exhaled air of

patients of lung cancer, healthy smokers, and non-smokers. They successfully classified patients with their VOC profiles [9]. They succeeded in distinguishing the patients with pattern recognition of the metabolite profiles. In the latest time, technologies in chromatography, MS, and statistical analyses have made significant progress. An instance of their practical use is development of a diagnosis kit for depression based on reduced ethanolamine in depressed patients [10].

1.2 Metabolomics as biochemical phenotyping

Metabolites are considered as the final products resulting from responses of biological systems against genetic or environmental changes [11]. The metabolite profiles provided broader insights on a function of a gene. For example, productions of some kinds of lipids were altered by mutation on a gene involved in stomatal development [12]. Metabolome analyses has provided quantitative and wide knowledge on metabolism as the endpoint of cellular regulation depending on genotype and environment.

From around 2000, quantitative changes of metabolites has been utilized in functional genomics. Oliver *et al.* suggested that changes in the concentrations of metabolites were required to elucidate function of a gene [13, 14]. With metabolome analyses, Raamsdonk *et al.* elucidated function of a gene that shows no apparent phenotypes when deleted [15]. They showed quantitative variations in metabolite profiles caused by differences on genotype and cultural environment. Tweeddale *et al.* called a total metabolite pool "metabolome" and quantified its changes in *Escherichia coli* as response to an environmental change [16]. Fiehn *et al.* showed that metabolite profiles obtained by GC-MS represented variations in genotypes of *Arabidopsis thaliana*. They detected and quantified over 300 metabolites [12]. Half of them were identified with their structures by comparing data obtained from commercially available standard compounds.

1.3 Techniques for metabolite identification

There are two main techniques to quantify and identify metabolites: one is nuclear magnetic resonance (NMR), and the other is MS. Each technique has specific advantages

and disadvantages. For example, MS is usually more sensitive than NMR, but NMR does not need prior separation of metabolites needed by MS; NMR can be performed directly on biological systems. In this section, principles of the metabolite identification by ^{1}H NMR [17] and MS [18] are summarized.

1.3.1 Proton nuclear magnetic resonance

The proton (^{1}H) NMR method detects magnetization of hydrogen nuclei composing compounds. When an external magnetic field having particular frequency is supplied to a nucleus, it is significantly magnetized; that is resonance. The frequency that causes the resonance is peculiar to species of the nucleus, but it shifts depending on a chemical condition where the nucleus is; for example, hydrogen nuclei in a methyl group (RCH_3) and in a methylene group (R_2CH_2) show resonance in different frequency. They are called chemical shifts and important clues to infer the structure of the compound. Chemical shifts of a broad range of chemical structures have been thoroughly analyzed and collected. Thanks to that, we can guess what compounds are included in a sample by ^{1}H NMR. There are also NMR methods targeting other species of nuclei like ^{13}C.

The major advantage of NMR compared to MS is ability to measure intact tissues. NMR can be applied to biological fluids without any separation or derivatization. Owing to that, NMR-based study can analyze metabolite profiles altered by physiological stimuli in living systems [19, 20]. In particular, this sort of studies focusing on dynamical changes of metabolite profiles in living systems is called “metabonomics.”

1.3.2 Mass spectrometry

In a mass spectrometer, compounds included in a biological sample are ionized and separated by their mass-to-charge ratios (m/z). The unit of mass m is Dalton (Da); 1 Da means 1/12 of ^{12}C mass. The charge z is measured in units of a charge of an electron (1.602×10^{-19} C). The values of m/z are measured by observing motions of the ions in an electromagnetic field. The number of ions having each value of m/z is counted separately and described as intensities in a mass spectrum.

There are several variations of ionization instruments. Electron ionization (EI) is one of widely used techniques. Accelerated electrons collide with compounds and an electron is dispelled out of the compound. That leads to odd-electron radical cations.

As the cations are provided excess energy by the collision, they are fragmented into substructures; EI is called "hard" ionization owing to this feature. Therefore, a mass spectrum obtained with EI contains several peaks corresponding to an entire molecule and fragments. The spectra including structural information are useful to identify the compounds. Theoretical prediction of fragmentation in EI-MS has been challenged for a long time [21, 22, 23]. Another common technique is electrospray ionization (ESI). In contrast to long-established EI, ESI has attracted significant attention since around 1990. A solution including compounds and some ions, e.g. H^+, K^+, or CH_3COO^-, is electrically sprayed from a capillary. When sprayed, the compounds catch one or more ions to form even-electron ions. Great difference from EI is that ESI can produce negative ions by bonding anions to compounds. Because of "soft" ionization, ESI maintains the original structures of fragile compounds. Thus, ESI is suitable for polysaccharides, peptides, and proteins. Production of polyvalent ions also helps with measuring macromolecules by reducing m/z values. There are other options such as chemical ionization, field desorption, fast atom bombardment, and matrix assisted laser desorption ionization; their details are omitted here.

There are also some kinds of mass analyzers. The first one is a magnetic sector. When accelerated ions fly through a magnetic field, their paths are curved by magnetic force. The curvatures are different depending on the mass of ions; a heavier ion draws the arc with a larger radius owing to strong centrifugal force. Because the path to the ion detector is fixed by the equipment, by scanning the magnetic field strength, ions with the specific mass are collected. Another is a quadrupole. Four rod-shaped electrodes surrounding the path of ions are supplied voltage with high frequency. The voltage causes oscillation to the ions, and ions having specific mass can pass through the analyzer to the detector. The quadrupole is suitable to ESI, where arising ions have relatively low velocity. A time of flight (TOF) analyzer has high sensitivity and resolving power. Ions are accelerated by a certain voltage and drift through a tube to the detector. Supplied with the same energy, light ions fly faster than heavy ones. By measuring the time till arrival, mass of ions are estimated. There are in principle no loss of ions; thus TOF analyzers demonstrate high sensitivity. Its resolving power can be increased by elongating the drift tube. Ion traps and Fourier transform-ion cyclotron resonance are other examples, but I will skip them.

Since there are isobaric metabolites, having identical mass, m/z is sometimes not

enough for the separation. In most cases, mass spectrometers are used in combination with chromatographic techniques, e.g. GC and liquid chromatography (LC). MS has a good affinity to the chromatographic techniques. As EI ionizes compounds in a gas phase, it is highly compatible with GC. LC can be directly combined with ESI where compounds are ionized from a solution. These techniques provide additional separation of compounds. Another way to distinguish isobaric compounds is tandem mass spectrometry (MS/MS). Details of MS/MS are explained in the next subsection.

1.3.3 Tandem mass spectrometry

In tandem mass spectrometry (MS/MS or MS^2), one peak in a mass spectrum is picked, and corresponding precursor ions are transferred into second mass analyzing with fragmented into product ions. One way to implement that is serial connection of two mass analyzers. For example, it can be executed by triple quadrupoles. Ions are separated in the first quadrupole, and then ions with one specific mass value are transferred into the second quadrupole. In the second quadrupole, the picked ions are fragmented by collision induced dissociation (CID), of which details are described below. Product ions from the second quadrupole are analyzed in the third quadrupole. Mass and intensities of product ions are described as a tandem mass or MS/MS spectrum.

CID is a technique to activate precursor ions and cause fragmentation. It is effective to soft ionization like ESI which produces low-energy ions [24]. Precursor ions go through a chamber filled with neutral gas such as He, N_2, or Ar. The ions collide with the molecules of the gas and receive energy. The energy is immediately converted into internal vibration of the ions and induces dissociation of chemical bonds resulting in fragment ions. There are two types of CID depending on amount of kinetic energy given to the ions. Low-energy collision accelerates ions with 1-200 electron volt (eV), while high-energy collision employs energy in order of keV.

1.4 Metabolite identification with tandem mass spectrometry

1.4.1 Importance and insufficiency of standard spectral library

The chemical analysis working group in The Metabolomics Standard Initiative (MSI) defined four levels of metabolite identification:

- level 1: identified metabolites
- level 2: putatively annotated compounds
- level 3: putatively characterized compound classes
- level 4: unknown metabolites

Level 1 needs that results from experimental measurement are compared with results from purified authentic standards measured in the same laboratory with the same analytical methods. Level 2 can be achieved without measuring authentic standards by guessing the name of the compound from comparison between experimental results and public or commercial spectral libraries. If some features characteristic of a certain compound class, but not characteristic to a single chemical species, the analysis is on level 3. Level 4 means only that the obtained peak is not a noise. Many of MS-based studies refer to spectral libraries instead of data of authentic standards obtained in own laboratories. Therefore, they achieve annotation of level 2, not identification of level 1, according to the definition. In this dissertation, "identification" means the annotation of level 2.

In metabolome analyses, tandem mass spectrometry (MS/MS) is a powerful tool to identify various metabolites from a biological sample at once. Nowadays most of mass spectrometric analyses are accompanied by GC or LC, but coelution of some metabolites makes the separation unclear. MS/MS provides additional information to distinguish metabolites, especially isobaric ones. Metabolites are decomposed into fragments by CID in a tandem mass spectrometer. Mass and intensity of each fragments are described as an MS/MS spectrum. Since fragmentation patterns of metabolites are different depending on their molecular structure, MS/MS spectra are characteristic of metabolites and considered as key clues for identification.

Although metabolites are identified by searching standard MS/MS spectra stored in

public, commercial, or private repository, insufficiency of current spectral libraries impedes comprehensive identification of metabolites. If a standard spectrum similar to a measured one is found, the corresponding metabolite is identified; otherwise, the metabolite remains unidentified and the measured spectrum is discarded. Several groups have endeavored to create substantial spectral libraries, NIST Tandem Mass Spectral Libraries [25], MassBank [26], METLIN [27], GNPS [28], and Human Metabolome Database (HMDB) [29] for example. Nevertheless chemical space to discover is so vast that the libraries have covered just a tiny part. A standard spectrum is created by measuring a purified metabolite; that is a complicated and costly process. In PubChem database, about 92 million compounds including synthetic ones are registered. Against such numerous compounds, measuring standard spectra one by one is supposed impossible. For those reasons, current metabolomics studies have not reached "omics" in a true sense. The majority of entire metabolites existing in organisms are out of our sight; we are focusing on their extremely limited part.

As I mentioned in Subsection 1.3.2, mass spectra obtained by EI-MS also include fragment information. Since species of ions produced by EI-MS and MS/MS with CID are different, they result in different spectra. Standard spectra from EI-MS are stored in spectral libraries separately from ones from MS/MS and also used for the identification. In this dissertation, I focus on metabolite identification by ESI-MS/MS with CID fragmentation.

1.4.2 Challenges for metabolite identification independent of standard spectral library

To overcome the insufficiency of MS/MS spectral libraries, many research groups have challenged to theoretically identify metabolites without measured standard spectra. One approach is guessing metabolite structures directly from fragmentation patterns appearing on MS/MS spectra. CSI:FingerID [30, 31] and CFM-ID [32, 33] are successful examples in this approach. Another approach is to build databases of MS/MS spectra derived by theoretical prediction. LipidBlast [34] successfully produced theoretical MS/MS spectra of a wide range of lipids.

A fragmentation process starting from one metabolite structure is described as a tree data structure called fragment tree [35, 36]. An example of fragmentation tree is

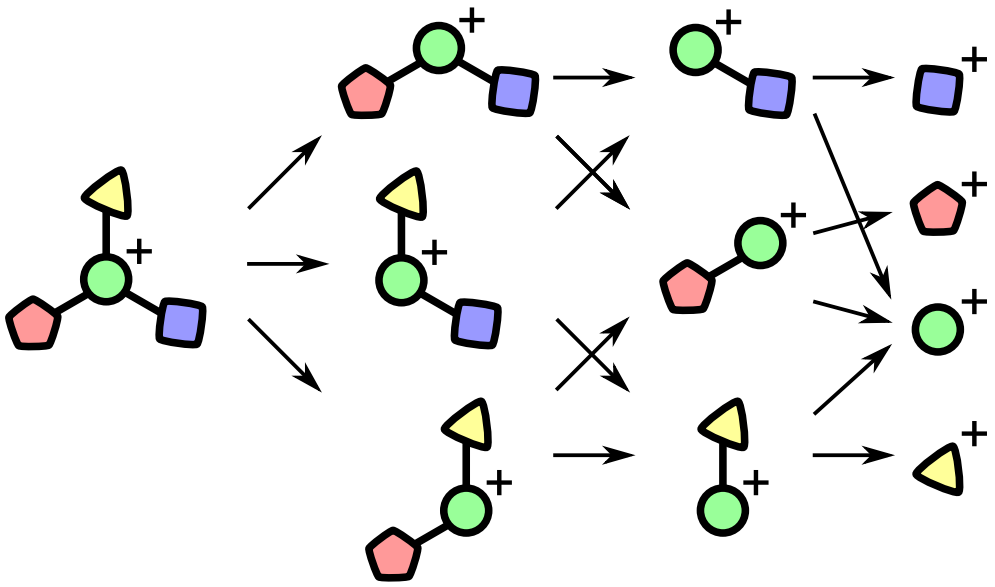


Figure 1.1: An example structure of fragment tree. Adapted from Fig. 2 in Ref [33].

illustrated in Figure 1.1. Each node corresponds to fragments of the metabolite, and each edge corresponds to one fragmentation step. From accurate mass represented by each peak of MS/MS spectra, the atomic composition of the fragment can be estimated. By analyzing an MS/MS spectrum, putative fragment trees representing the MS/MS spectrum can be constructed. Each fragment has several fragmentation pathways to reach itself which consist of fragmentation steps. For example, the fragment represented by a single triangle has two pathways in Figure 1.1. If experimental fragments can be reached without missing nodes or edges, the reliability of the fragment tree becomes high. That can be used as a criterion to annotate a precursor structure of the MS/MS spectrum [35].

CSI:FingerID [30, 31] utilizes fragmentation tree calculated from measured MS/MS spectra. Here, the fragmentation tree only has compositional information; connection among substructures, which are illustrated by a single shape, is not considered. Means of machine learning predicts presence of characteristic substructures, called fingerprints, represented by a measured spectrum. Then a compound database is searched for metabolite structures composed of the fingerprints predicted to present in the measured spectra. This method allows spectrum-unknown metabolites to be

identified.

CFM-ID [32, 33] is a machine learning-based strategy. It predicts MS/MS spectra only from metabolite structures. Probability of bond cleavage depending on structural environment in the vicinity of the bond is learned from experimental standard spectra stored in databases. On the basis of the probability, metabolite fragmentation is simulated as a stochastic process and MS/MS spectra are theoretically generated. To identify metabolites without standard MS/MS spectra, candidate compounds based on their mass are sought from compound databases; mass of unknown metabolites to be identified can be obtained from the first step of MS. By predicting fragmentation of the candidates by CFM-ID, candidates are scored and ranked on the basis of similarity between standard and predicted MS/MS spectra [33].

Building purely computational mass spectral library is attracting attention in recent years [37]. LipidBlast [34] extrapolates fragmentation of lipids to create theoretical spectra. A lipid molecule is regarded as combination of three moieties. As the majority of fragmentation of lipids is decomposition into each of the moieties, theoretical MS/MS spectra of lipids are comprehensively generated by changing their combination.

These methods are successful but provide few insights on mechanisms of fragmentation in MS/MS. Machine learning is a black-box approach; why the fragmentation occurs is unclear. LipidBlast is clearer approach, but it can be applied to lipid molecules naturally. As metabolite fragmentation is unimolecular reaction, physicochemical theories such as quantum chemistry and statistical dynamics can supply meaningful prediction. In EI-MS, where metabolites decompose into odd-electron ions, quantum mechanics and molecular dynamics have been utilized to predict mass spectra [22, 23]. On the other hand, ionization and fragmentation processes in ESI-MS/MS with CID, where even-electron ions are dominant, are different from ones in EI-MS, and the physicochemical theories have not been applied to ESI-MS/MS to reveal mechanisms and predict fragmentation.

1.5 Aim and organization of the dissertation

In this dissertation, I discuss application of physicochemical theories to metabolite fragmentation in ESI-MS/MS. Computational chemistry was utilized to validate theoretically expected fragmentation, to distinguish isobaric metabolites, and to predict

fragmentation pathways in ESI-MS/MS. I confirmed applicability of computational chemistry for determining fragment structures and built a procedure to predict fragments from a metabolite structure.

In Chapter 2, theoretical validation of "hydrogen rearrangement rules" [1] is presented. They are simple rules based on the classic even-electron rule but powerful to predict fragment mass including hydrogen movement. I performed enthalpy calculation on putative fragmentation schemes to confirm existence of the fragments. All of the putative fragments were proved to result from properly low enthalpy changes which could be compensated by chemical collision in MS/MS. Computational chemistry was shown to be a useful tool for validation of fragment structures not clearly measured.

In Chapter 3, an identification result of a metabolite without standard spectra conducted with calculating energy change along its fragmentation pathway is shown. In comprehensive identification of sphingolipids [2], one of the detected metabolites had two candidates of isobaric molecular structures. Since standard spectra of both has not been obtained, some theoretical approaches were needed to choose a correct structure. I energetically analyzed fragmentation pathways of two candidate structures and provided evidence to decide which candidate is correct. From the calculation, an important concept of adjacent group effect was gained.

From physicochemical analyses described in Chapter 2 and 3, I constructed physicochemical prediction of metabolite fragmentation in MS/MS. Its detailed procedures and results are shown in Chapter 4. The core techniques are assignment of bonding patterns and activation energy calculation. Bonding pattern makes the method applicable to any molecules independent of their size. By calculating activation energy, bond cleavage tendency can be quantitatively estimated.

Conclusion remarks and discussion on future works are given in Chapter 5. My prediction method is currently ad hoc; energy calculations are performed after assigning bonding patterns to a target molecule. For original purpose, results of energy calculations should be stored so that the prediction is done immediately a target molecular structure is given. Remaining problems to incarnate that are described as future works.

2

Validation of hydrogen rearrangement rules

2.1 Introduction

2.1.1 Necessity of mechanistic elucidation of metabolite fragmentation

To identify metabolites from biological samples by LC-MS/MS, four criteria are mainly considered: retention time, precursor m/z, isotopic ratio, and MS/MS spectra [38]. MS/MS spectra have much information about structures of detected metabolites, but currently many of them are unavailable because of insufficient standard libraries as stated in Subsection 1.4.1; the coverage is estimated at 5 % of all known metabolites [1]. By computationally predicting MS/MS spectra, many researchers have been trying to fill the huge gap. Especially, rigorous relation between precursor structures and product

fragments is demanded. LipidBlast [34] succeeded for several lipid compound classes. However, fragmentation of small molecules caused by CID is broadly unclear. In order to facilitate metabolite identification with MS/MS spectra, mechanistic association of product ions with precursor structures is necessary.

2.1.2 Hydrogen rearrangement in MS/MS

One key to predict fragment structures is hydrogen rearrangement (HR). According to the "even-electron rule," even-electron cations, i.e. positive ions without unpaired electron, will not usually lose a radical to create odd-electron ion [39]. That means an even-electron ion will decompose into a smaller even-electron ion and a neutral molecule. The even-electron rule is based on the Octet rule, where a stable atom should have eight electrons around it. With satisfying the Octet rule, positive and negative charges in even-electron ions are basically carried by protonated and deprotonated molecules, respectively. To acquire a charge, an atom adjacent to the cleaved bond frequently absorb a H atom from the opposite part. That causes mass of ions to shift by 1, 2, or 3 Da from a substructure of its precursor ion [40]. Thus, prediction of the exact number of rearranged H atoms is needed to estimate accurate fragment mass.

Examples of HR are illustrated in Figure 2.1. It is assumed that one proton is attached to the nitrogen atom to ionize the molecule. When the middle covalent bond between a nitrogen atom and a carbon atom is cleaved, two kinds of ions are supposed to be produced. One is produced from the upper pathway in Figure 2.1. The nitrogen atom initially has one H atom. In the ionization step, the nitrogen atom is protonated, and then it absorbs a H atom from the cleaved opposite side to acquire a positive charge. The total number of rearranged H atoms is +2. The other fragment ion is produced from the lower pathway in Figure 2.1. On this pathway, we focus on the carbon atom connected to nitrogen, oxygen, and another carbon atoms. In both ionization and cleavage steps, the carbon atom obtains no H atoms. After the cleavage, the carbon atom becomes an unstable structure having less than eight electrons. The connected oxygen atom can share a lone pair with the carbon atom to make a triple bond. In that way, the carbon atom can be stable without absorbing H atoms.

In Figure 2.2, HR rules derived from statistics of fragment peaks in ESI-MS/MS spectra recorded in MassBank are summarized. Bold characters indicate dominant HR

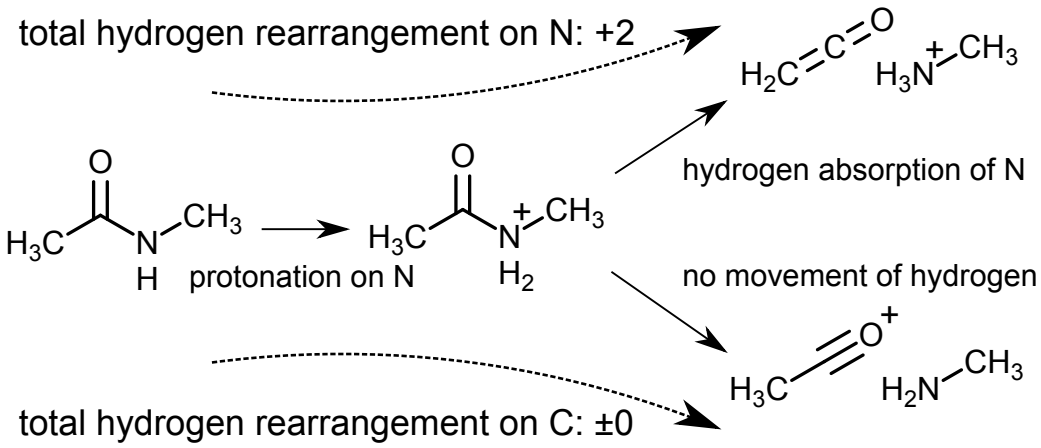


Figure 2.1: Examples of hydrogen rearrangement.

behaviors. Since CID employed by ESI-MS/MS measurement often causes multistep fragmentation, HR rules are classified to "first" and "second and later" bond cleavage. In first bond cleavage, fragmentation follows ionization of molecules. On the other hand, second and later bond cleavage occurs to fragment ions which already have a charge. Thus, the number of rearranged H atoms is different in each of the cases. At first bond cleavage in positive ion mode, a C atom takes no H atoms, and N and O atoms take two atoms. P and S atoms behave like both C and N atoms; the major part is C-like behavior. At first bond cleavage in negative ion mode, all five atoms can undergo the cleavage without HR. C and P atoms dominantly lose two H atoms. Losing one H atom, which violates even-electron rule, is the unique behavior by a S atom. Second and later bond cleavage is somewhat complicated; all atoms can undergo two HR patterns in both positive and negative ion mode.

By considering the HR rules, MS-FINDER program attempts to elucidate a metabolite structure from MS and MS/MS spectra [1]. Considered with soft ionization, an obtained MS spectrum has no fragment peaks; only peaks corresponds to an entire molecule appear. Even though the spectrum does not include fragment ions, there are several peaks because of isotopes like ^{13}C. Isotopic ratio calculated from the MS spectrum is also an important clue. The program generates chemical formula from accurate precursor mass, isotopic ratio, and some other criteria. After that, chemical structures coinciding with the generated formula are searched for from structural databases. The

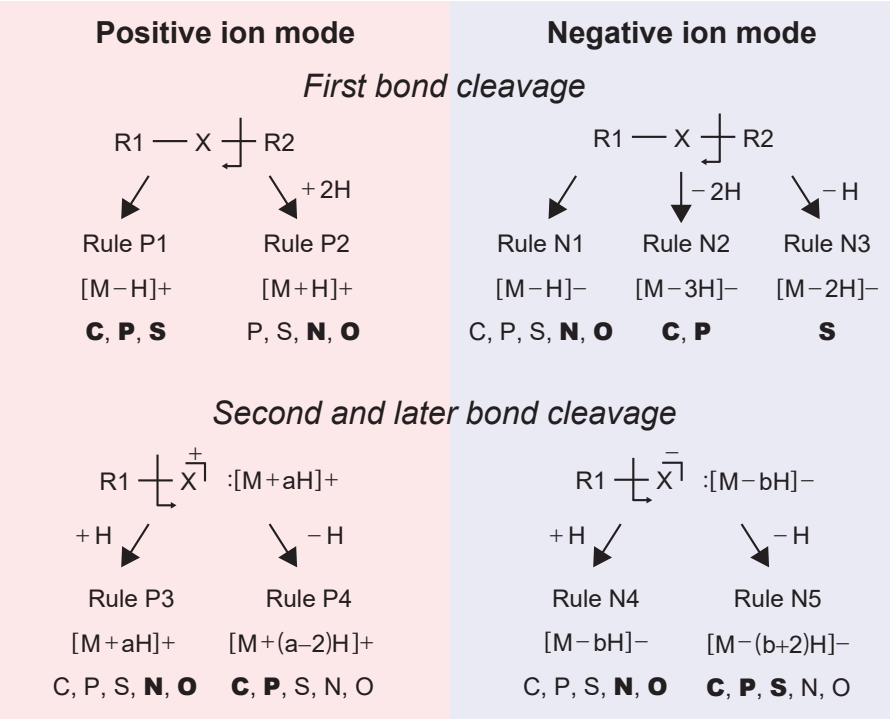


Figure 2.2: Summary of hydrogen rearrangement rules. Reproduced from Ref. [1].

candidate structures are computationally fragmented, and mass of the fragments are checked against the measured MS/MS spectrum. In this comparison, fragment mass is scored on the basis of the HR rules. The program ranks the candidates by the score and suggests metabolite structures likely to be measured. That is the purpose to derive the HR rules.

The HR rules were verified both statistically and theoretically. Statistics from MS/MS spectra recorded in MassBank revealed that the major part of carbon, nitrogen, and oxygen atoms follows the even-electron rule, while sulfur and phosphorus atoms showed different behavior [1]. There are also some exceptions in C, N, and O atoms. I investigated the causes of the rule-violating behavior of S and P atoms and the exceptions of C, N, and O atoms by using computational chemistry. In addition, I calculated enthalpy changes in fragmentations predicted by the MS-FINDER program to confirm the fragmentation is likely to occur. Since the structural prediction was applied on spectrum-unknown metabolites, we needed a basis for rightness of the prediction. Computational chemistry revealed that predicted fragmentations can be

sufficiently activated by energy supplied from low-energy CID.

2.2 Enthalpy calculation in putative fragmentation pathways

2.2.1 Hypothesizing fragmentation pathways

Fragmentation pathways which led to predicted fragment ions were hypothesized manually. The MS-FINDER program annotated fragment structures onto peaks of MS/MS spectra measured from several metabolites on the basis of HR rules. Fragmentation trees like an example illustrated in Figure 1.1 were then constructed. Fragmentation pathways to produce the annotated fragments were investigated by tracing electron transfer within the molecule.

At first, a position of protonation or deprotonation which facilitate subsequent fragmentation was assumed. A proton can absorb an electron pair from heteroatoms, i.e. other than C and H atoms, or double- or triple-bonded C atoms. Heteroatoms provides a lone pair to the proton, and π-electron composing the multiple bonds can catch the proton. The protonation causes a positive charge on the site. Acidic sites such as phosphate and sulfate were mainly selected as deprotonation sites, which carry a negative charge. Generally, moieties around a charge are relatively unstable. Thus, the charged sites were the starting points for the fragmentation pathways.

A charge can be stabilized by heteroatoms or charge distribution. Heteroatoms can maintain a charge with sutsifying the Octet rule. When a positively charged atom cuts one of the bonds connected to itself, the opponent of the cleaved bond receives a charge. In Figure 2.3 a, a charge is initially on the N atom and moves to the middle C atom after the cleavage. The charged C atom has only six electrons (a single bond with a C atom and a double bond with an O atom) and is therefore in a unstable state. The initially charged N atom can absorb a proton from the opposite part to remove a charge (the upper part of Figure 2.3 a). Otherwise, the O atom adjacent to the C atom which received a charge can provide a electron pair (the lower part of Figure 2.3 b). Since the N and O atoms have eight electrons (four single bonds and a triple bond and a lone pair, respectively), the conclusive ions are stable. Even if there are only C atoms,

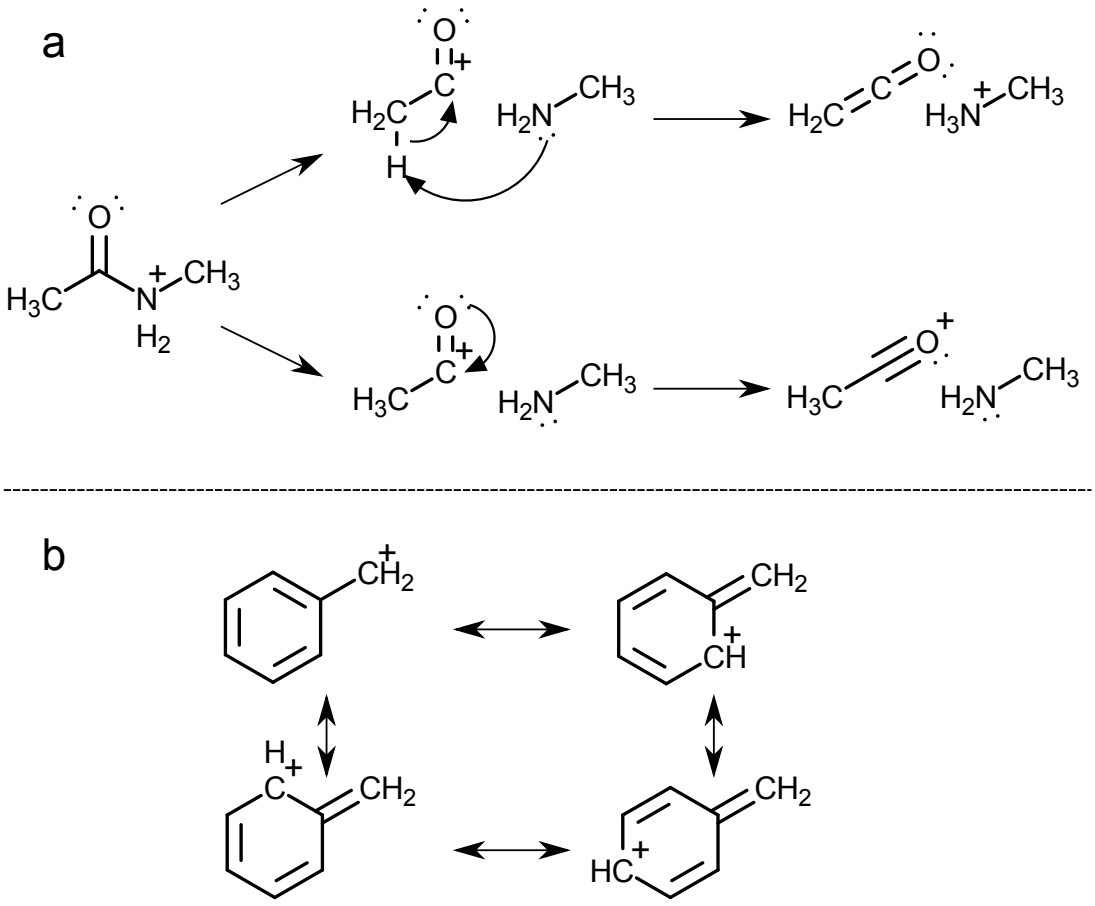


Figure 2.3: Typical fragmentation schemes of positive ions. (a) Two possible schemes to stabilize the charge which arose on the middle C atom. Two dots indicate paired electrons, and curved arrows indicate transfer of paired electrons. (b) Charge distribution within a phenyl group.

a charge can be stabilized by distribution (see Figure 2.3 b). In a toluene structure, the position of the charge can be changed without rearranging the atoms. The true structure of the ion is an average structure of the four structures. Thus, magnitude of the charge in each structure is regarded as 1/4; this is called “partial charge” because, in reality, an electron cannot be divided. By decreasing the partial charge, an ion can stabilize its structure.

Metabolites should undergo fragmentation pathways with stable structures. By tracing stable structures stated above from ionized metabolites to their fragments, I

generated putative fragmentation pathways. To examine reliability of the putative pathways, enthalpy changes along them were calculated with computational chemistry.

2.2.2 A semi-empirical method for enthalpy calculation

Molecular orbital (MO) method is a computational method to determine electron distribution over molecular orbitals held by a molecule. Calculation of potential energy and geometrical optimization can be performed by the calculation. Potential energy depends on arrangement of atoms such as bond length, bond angle, and bond torsion. Geometrical optimization is executed by repeating slight changes of the arrangement which cause energy decrease. If the energy decrease becomes smaller than a certain threshold, an optimal structure of the molecule is achieved.

In this study, a calculation method called "semi-empirical" in contrast with "first principle" was used. In first principle calculation, potential energy of a molecule is fully derived by solving quantum mechanic equations. As the equations cannot be solved analytically, they are solved by repeating numerical integrations. That causes massive computational cost. Semi-empirical methods are developed in order to decrease the computational cost. In first principle calculation, there are numerous integrations having less effects than the core integrations. Semi-empirical methods neglect these less influential integrations [41]. To compensate for the approximation, they introduce parameters instead of the integrations; the values originally derived from the integrations are replaced by provided parameters based on experimental and computational data. Semi-empirical methods are classified by how to approximate and parameterize.

All enthalpy calculations in this chapter were performed by using MOPAC2012 (Molecular Orbital PACkage) program [42]. Approximation method was modified neglect of diatomic overlap (MNDO) [43], no additional approximation other than above, and parameterization method was parametric method number 7 (PM7) [44, 45]. In this method, enthalpy of the molecule is derived from four parameters: electronic energy, nuclear-nuclear repulsion energy, energy to eliminate all the valence electrons, and total enthalpy of atomization of all atoms. Former two were calculated by numerical integration and latter two were assigned empirically to each element.

Enthalpy changes in fragmentation pathways were calculated as energy difference

between the precursor ion and its fragments. The precursor and fragment ion structures are built in a graphical user interface of MOPAC, WebMO [46]. Heats of formation of the precursor ion (ΔH_f^{prec}), the fragment ion ($\Delta H_f^{frag,ion}$), and the neutral fragment ($\Delta H_f^{frag,neut}$) were calculated with MOPAC2012. Enthalpy change along the fragmentation pathway was calculated as $\Delta H_f^{frag,ion} + \Delta H_f^{frag,neut} - \Delta H_f^{prec}$. If the enthalpy change is lower than 0 kcal mol^{-1}, fragments are more stable than their precursor ion. That means the fragmentation easy to occur with mild activation by CID. Calculation of each H_f took several seconds with a laptop equipped with 4 GB memory.

2.3 Results of presuming fragmentation pathways and enthalpy calculation on them

2.3.1 Behaviors of P and S atoms

Despite identical electron placement in the outermost shell between N and P atoms, having three valence electrons, HR behaviors of them were different according to the statistic (see Figure 2.2). The same can be said for O and S atoms.

A manual inspection found that P atoms exist majorly as phosphates in biological environments; that is the reason why they behave differently from N atoms. MS/MS peak annotation for 2′-deoxycytidine 5′-diphosphate is shown in Figure 2.4 a. Of these, fragment a, d, and f result from bond cleavage on P atoms. Fragmentation scheme for fragment a is in Figure 2.4 b. This is dehydration from the β phosphate group. The whole fragmentation pathways annotated to all fragments in Figure 2.4 are illustrated in Figure 2.5 and 2.6. Fragment d and f are also explained by similar schemes where H atoms in the phosphate groups move to other O atoms. A negative charge is stably maintained by charge distribution over the phosphate group. The fragmentation occurs independently of the charged site and needs low energy. Enthalpy changes along the schemes for fragment a, d, and f were calculated as −59.3, −47.9, and −48.0 kcal mol^{-1}, respectively (Table 2.1). The changes were as low as those in production of fragment b or e, which followed the HR rules for C or O atoms, respectively. The low-energy cleavage of P–O bonds is characteristic of the phosphate group structures. Therefore, the HR behavior of P atoms is different from N atoms.

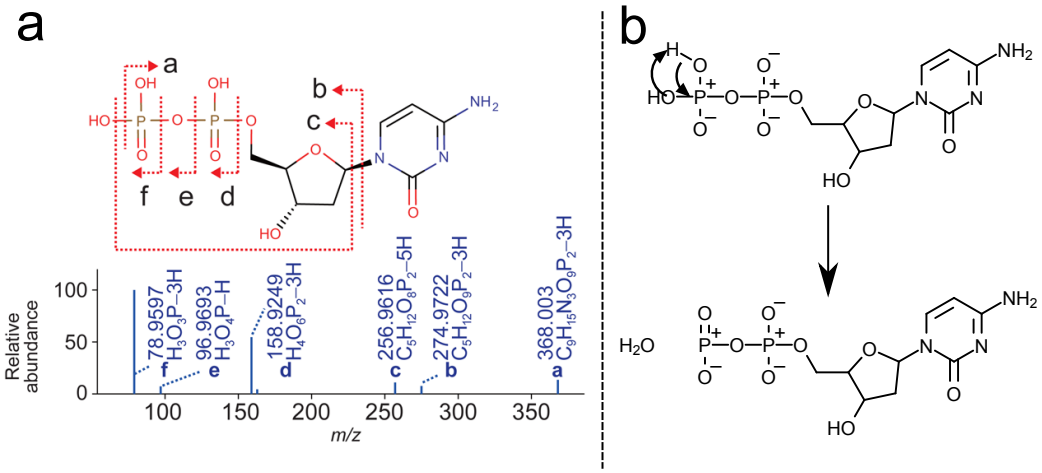


Figure 2.4: (a) Annotation for MS/MS spectral peaks of 2′-deoxycytidine 5′-diphosphate measured in negative ion mode and (b) a fragmentation scheme for fragment a. Reproduced from Ref. [1].

Table 2.1: Enthalpy changes in fragmentation of 2′-deoxycytidine 5′-diphosphate.

fragment	enthalpy change (kcal mol^{-1})
a	−59.3
b	−49.8
c	−30.0
d	−47.9
e	−31.3
f	−48.0

S atoms showed particular behavior which violate the even-electron rule. Fragmentation of 3-indoxyl sulfate in negative ion mode indicated an example to produce an odd-electron ion (Figure 2.7). The odd-electron ion results from “homolysis,” where an electron pair composing a bond is divided into single electrons upon cleavage of the bond. If an electron pair moves together, the bond cleavage is called “heterolysis” and results in even-electron ion. Since odd-electron ions produced by homolysis are generally less stable than even-electron ions, odd-electron ions are considered to hardly arise in CID-activated fragmentation. However, the sulfate structure was able to stabilize the unpaired electron by distributing it over the structure. Calculated enthalpy

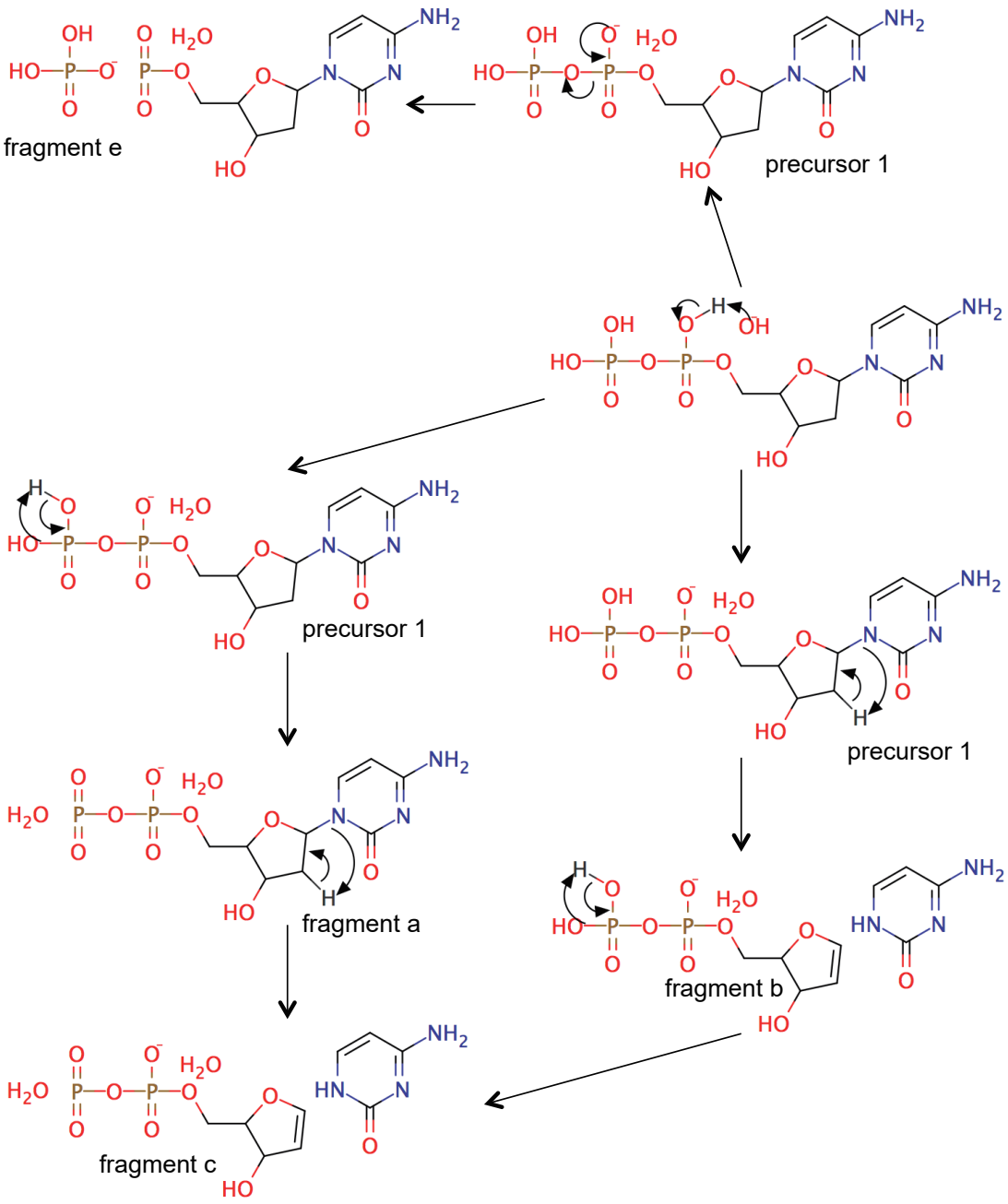


Figure 2.5: Putative fragmentation pathways of 2′-deoxycytidine 5′-diphosphate to produce fragments a, b, c, and e in Figure 2.4. Reproduced from Ref. [1].

value of the odd-electron ion (fragment b' in Figure 2.7) relative to its precursor ion, −31.3 kcal mol^{-1}, was as low as that of the even-electron ion (fragment b), −37.4

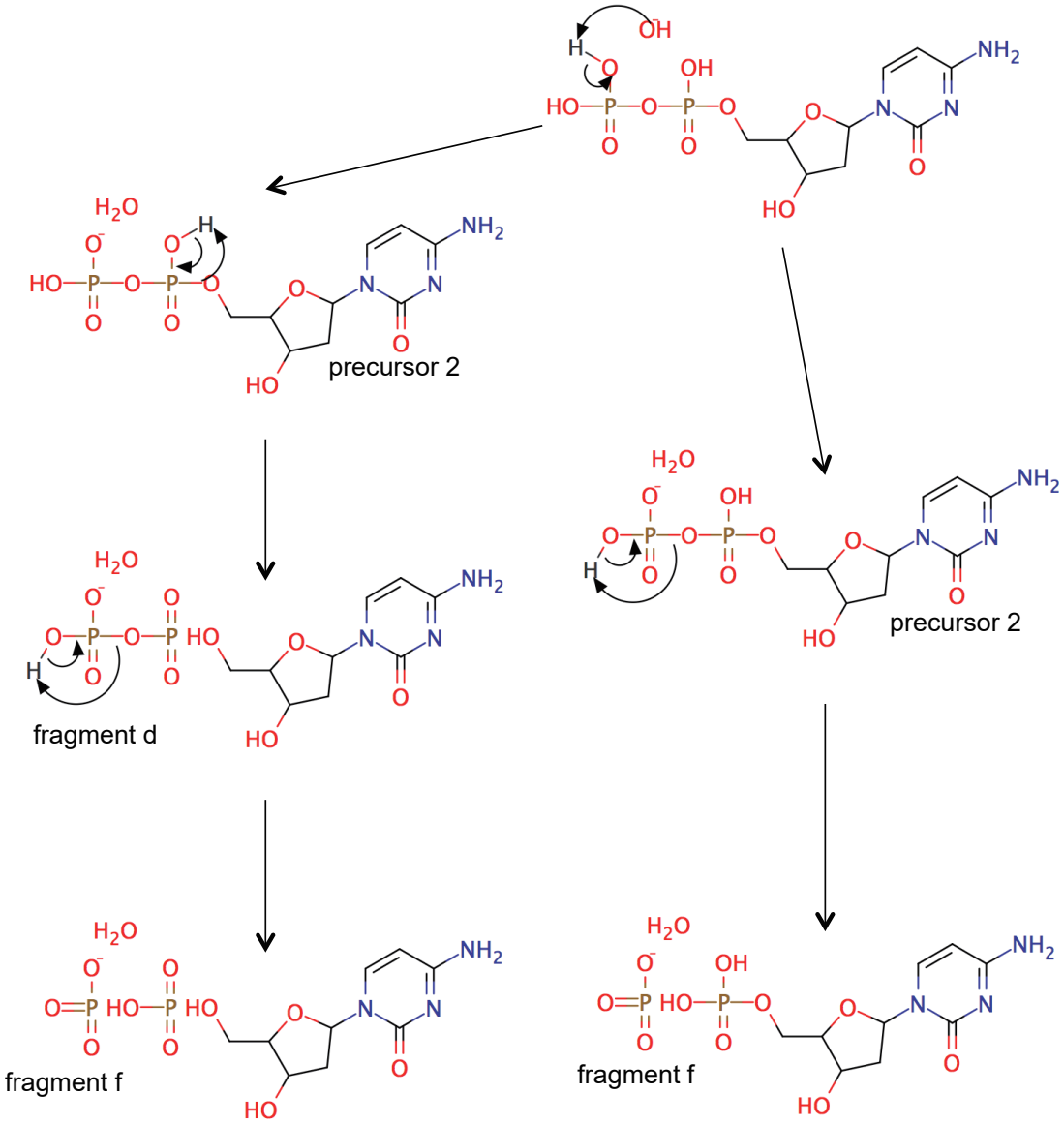


Figure 2.6: Putative fragmentation pathways of 2′-deoxycytidine 5′-diphosphate to produce fragments d and f in Figure 2.4. Reproduced from Ref. [1].

kcal mol^{-1}. The value was as low as that of fragment a (see Figure 2.8 and Table 2.2). Actually, fragment b' was more abundant than fragment b; nevertheless fragment b' has higher enthalpy. The possible reason is that fragment b needs to absorb a H atom from the distant N atom to compensate its valence. The movement of the H atom requires proper conformation of the molecule, and that decreases reaction rate of the

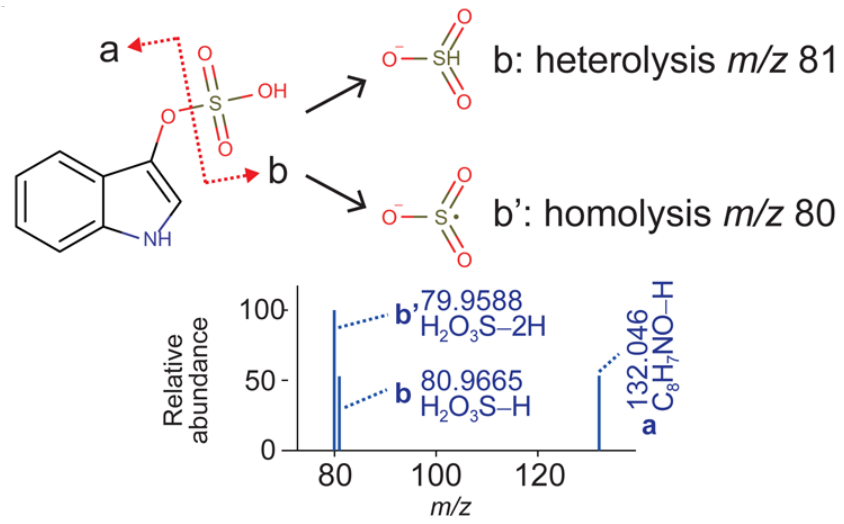


Figure 2.7: Annotation for MS/MS spectral peaks of 3-indoxyl sulfate measured in negative ion mode. Reproduced from Ref. [1].

Table 2.2: Enthalpy changes in fragmentation of 3-indoxyl sulfate.

fragment	enthalpy change (kcal mol^{-1})
a	−32.8
b	−37.4
b'	−31.3

heterolysis fragmentation. The homolysis requires no particular conformation. Thus, fragment b' is produced faster than fragment b.

2.3.2 Exceptions of C, N, and O atoms to HR rules

Exceptional fragments to the HR rules are exemplified in Figures 2.9, 2.10, and 2.11. They are classified into two patterns: radical ions and charge-transferred ions. I computed their stable electronic structures and enthalpy changes along the fragmentation.

Example 1. Phosphocholine. The loss of three H atoms along cleavage on N-C and C-O bonds in phosphocholine (Figure 2.9) is explained by a radical cation. Again, the odd-electron cation is unstable. The unpaired electron is distributed over the conjugated bond indicated by a blue circle. Calculated enthalpy change along the

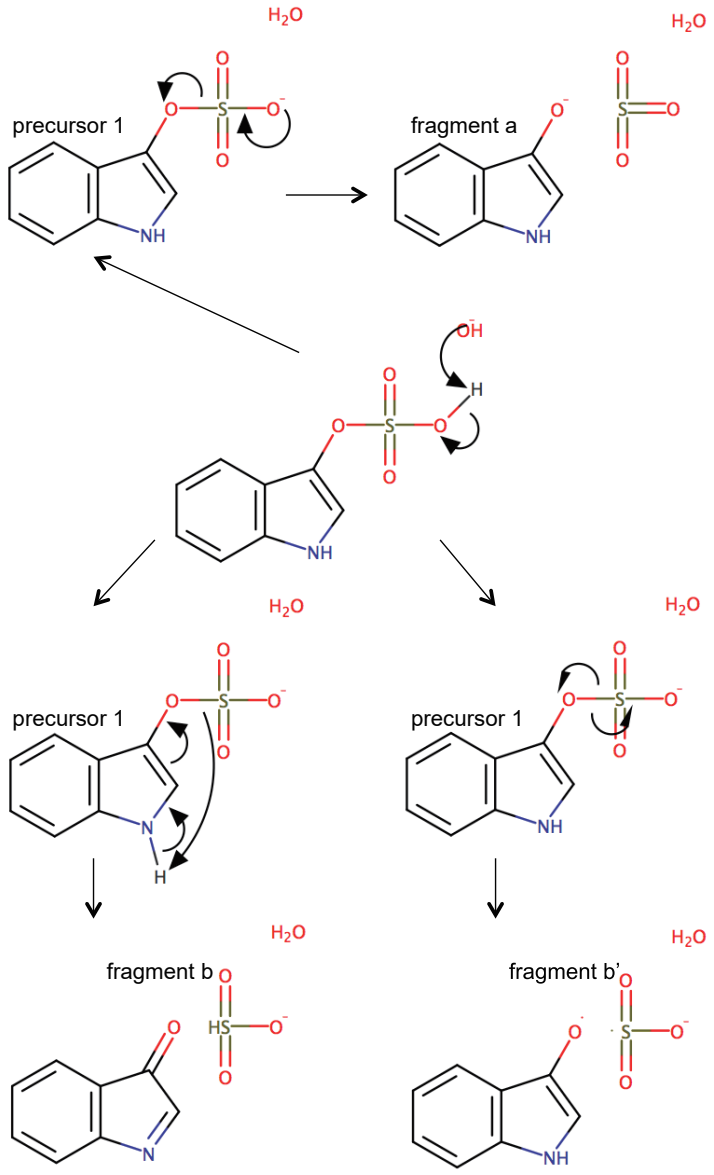


Figure 2.8: Putative fragmentation pathways of 3-indoxyl sulfate to produce fragments a, b, and b' in Figure 2.7. Reproduced from Ref. [1].

cleavage was 65.02 kcal mol^{-1} ~ 2.8 eV. As low-energy CID provides energy ~ 10 eV, this fragmentation can be activated sufficiently. In addition, the computational chemistry method can estimate spin population, which infers localization of an unpaired electron, on each atom. If the spin population of an atom is 1 or −1, the unpaired electron is completely localized on the atom; that means the radical is unstable. The spin

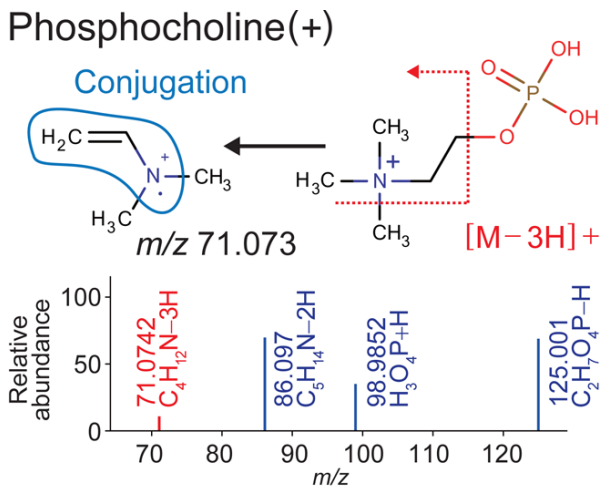


Figure 2.9: Exceptional fragment ions observed in MS/MS spectra of phosphocholine measured in positive-ion mode ESI-MS/MS. The red peak is an exceptional fragment to HR rules. The unpaired electron was distributed over conjugated bonds in the blue circle to gain a stable electronic structure. Reproduced from Ref. [1].

populations on the C and N atoms in the blue circle in Figure 2.9 were calculated as 0.67 (~ 2/3), −0.33 (~ −1/3), and 0.68 (~ 2/3) from the left top. Sum of the three populations nearly equals to 1; i.e., one unpaired electron is distributed over the three atoms. These results mean the unpaired electron is distributed over the three atoms to stabilize the radical cation.

Example 2. Kaempferide. In Figure 2.10, another exception caused by a radical anion is shown. Here, the range of conjugation is much broader because of the aromatic structure. Spin population on each atom involved in the conjugation is summarized in Table 2.3. The densest population is 0.26 ~ 1/4 on atom 5 and 10; that is sparser than population on phosphocholine with the smaller conjugation moiety. The broad conjugation moiety contributes to lower relative enthalpy of the radical anion as 34.71 kcal mol^{-1} ~ 1.5 eV.

If there are some conjugated bonds adjacent to a cleavage position where an unpaired electron arises, the radical ion is stabilized and easily produced even from an even-electron ion. That is the origin of odd-electron-related exceptions to the HR rules.

Example 3. Isoproturon. Example in Figure 2.11 does not originate from radical ions. Charge transfer caused the exception to the HR rule for N–C bond cleavage.

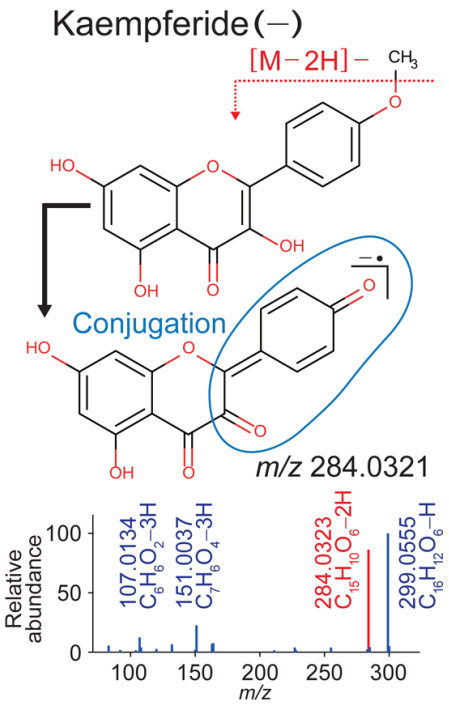
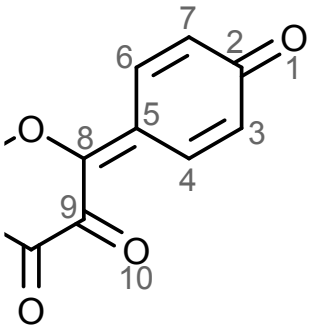


Figure 2.10: Exceptional fragment ions observed in MS/MS spectra of kaempferide measured in negative-ion mode ESI-MS/MS. The red peak is an exceptional fragment to HR rules. The unpaired electron was distributed over conjugated bonds in the blue circle to gain a stable electronic structure. Reproduced from Ref. [1].

According to the HR rules, a charged N atom side which results from N–C bond cleavage captures two H atoms. That HR is needed to raise a charge on the N atom. In the case of isoproturon, the N atom side gained a charge without movement of a H atom. The charge arose on a C atom composing the benzene moiety instead of the N atom. Bond rearrangement in the benzene moiety occurred immediately after the cleavage, and the N atom no longer needed to capture two H atoms to gain a charge. Usually, a charge prefers being on N and O atoms rather than C atoms in order to satisfy the Octet rule. Here, the charge was however stabilized on C atoms by distribution. Partial charges on each moiety (C, N atoms and bound H atoms) is summarized in Table 2.4. As there were three charged sites (moiety 4, 6, and 9), resonance structures to distribute the charge were formulated as Figure 2.12. Because of the distribution, relative energy was sufficiently low as 37.0 kcal mol^{-1} ~ 1.6 eV. From the MO calculations, distribution of a positive charge facilitating the exceptional

Table 2.3: Spin population on each atom included in conjugated bonds of a fragment radical anion produced from kaempferide.



atom No.	spin population
1	0.14
2	−0.03
3	0.21
4	−0.12
5	0.26
6	−0.13
7	0.23
8	0.12
9	0.03
10	0.26

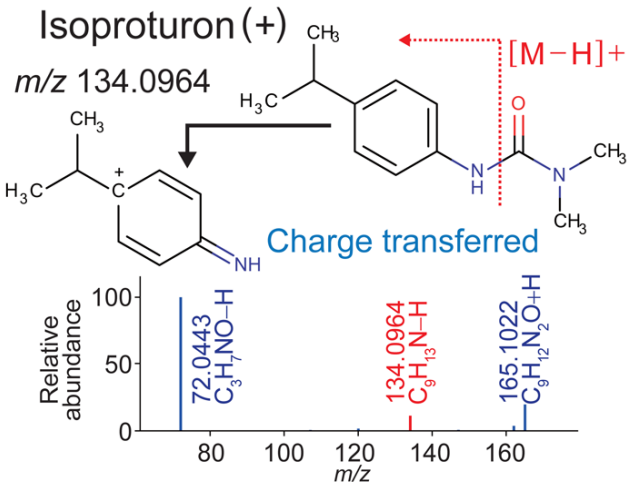


Figure 2.11: Exceptional fragment ions observed in MS/MS spectra of isoproturon measured in positive-ion mode ESI-MS/MS. The red peak is an exceptional fragment to HR rules. A positive charge caused by the cleavage immediately moves to the benzene structure to gain a stable electronic structure. Reproduced from Ref. [1].

fragmentation was revealed.

Table 2.4: Calculated partial charges on each moiety of a fragment cation produced from isoproturon. A moiety includes one C or N atom and bound H atoms; e.g., moiety 1 includes one C and three H atoms.

moiety No.	partial charge
1	0.08
2	0.03
3	0.07
4	0.37
5	−0.05
6	0.19
7	0.07
8	0.03
9	0.33
10	−0.12

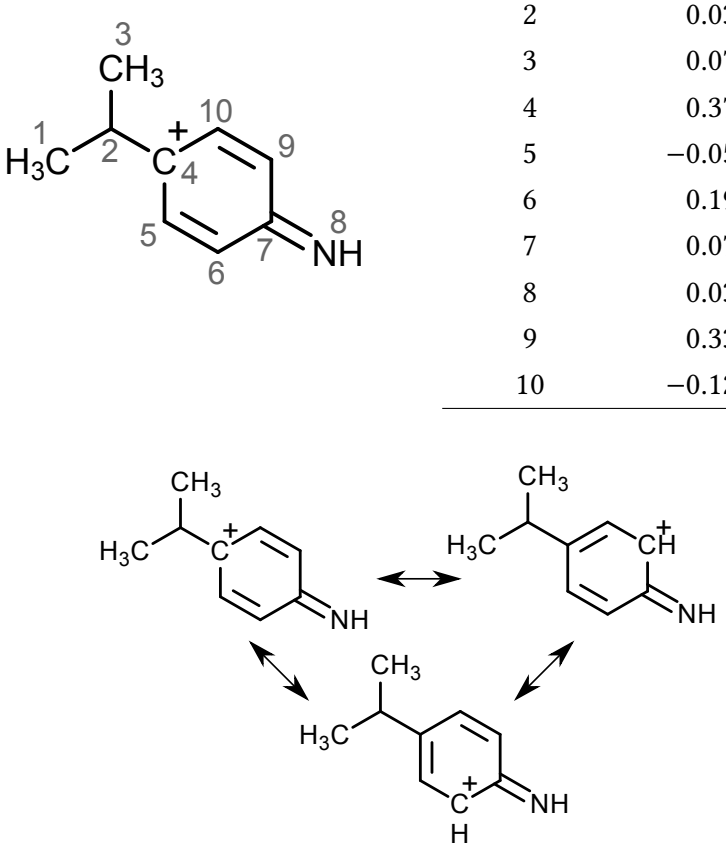


Figure 2.12: Resonance structures of fragment radical cation of isoproturon.

2.4 Conclusion

HR is an important key to elucidate exact fragment structures from MS/MS spectra as well as reveal detailed fragmentation mechanisms which occur in tandem mass spectrometer. Typical HR behaviors were formulated as the HR rules, and they were utilized for metabolite identification with MS/MS. Semi-empirical MO calculation quantified relative enthalpy of fragment ions and electron distribution over them. The calculation revealed how stable the fragment ions were and how they stabilized

their charged structures to ensure their existence. Especially, the origins of several exceptions to the HR rules were clarified by comparing enthalpy and analyzing electron distribution. That consolidated reliability of the HR rules. Through this study, computational chemistry was proved helpful to investigate mechanisms of CID-activated fragmentation in MS/MS.

3 Comprehensive identification of sphingolipids

3.1 Introduction

3.1.1 Building theoretical MS/MS spectral libraries for sphingolipid classes to identify them comprehensively

LipidBlast is a theoretical MS/MS spectral library covering 26 lipid classes and has ability to expand its coverage [47, 48, 38]. LipidBlast has succeeded in comprehensively constructing *in silico* MS/MS spectral libraries for glycerolipids and sphingolipids [34]. The construction is based on extrapolating measured fragmentation of lipid molecules. Variations of lipids are computationally generated by changing combination of functional groups (fatty acids), scaffolds, and header groups. A heuristic method utilizing experimentally measured MS/MS spectra produces theoretical MS/MS spectra

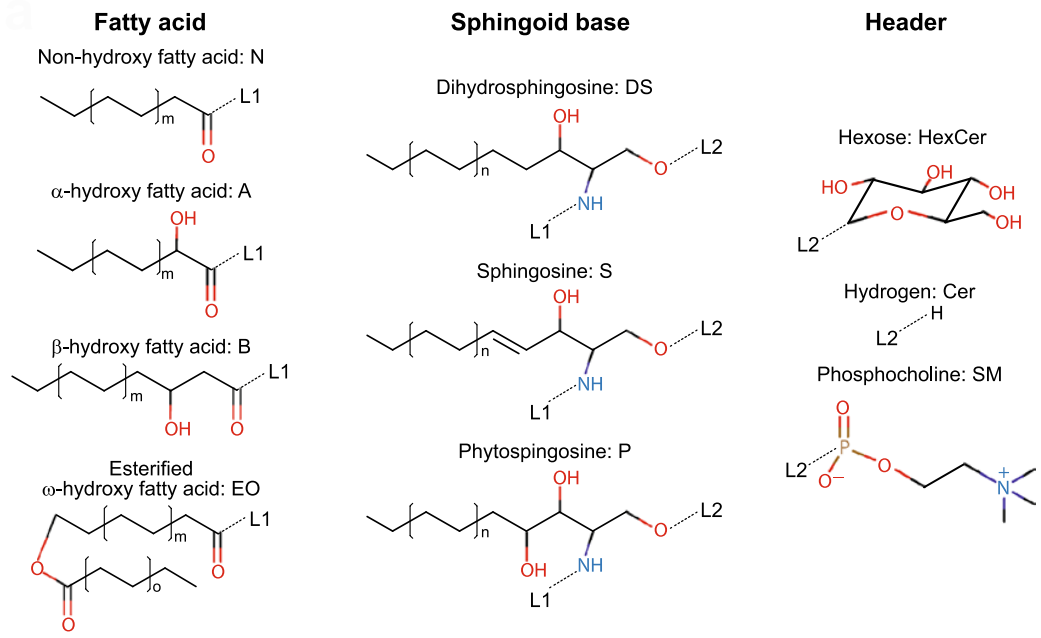


Figure 3.1: Three moieties composing sphingolipid structures. Reproduced from Ref. [2].

for computationally generated lipid molecules to cover all varied structures in a class. As the developers has been providing LipidBlast templates for novel lipid classes, theoretical libraries for lipids not originally covered in LipidBlast can also be generated [47].

Theoretical MS/MS spectral libraries for 21 sphingolipid classes in negative ion mode ESI-MS/MS were newly constructed [2]. The classes included eight human ceramide classes [49], one murine ceramide class [50], their monoglycosides, and sphingomyelin. Sphingolipid structures were regarded as combination of three moieties: fatty acid, sphingoid base, and header (Figure 3.1). The abbreviated names of the classes were originally used by Masukawa *et al.* [49]; for example, a sphingolipid consisting of an α-hydroxyl fatty acid (A), a sphingosine base (S), and a hydrogen atom header (Cer) is denoted as "Cer [AS]". The 21 sphingolipid classes were represented by variations of the combination.

By using *in silico* MS/MS spectral libraries, sphingolipids included in several biological samples were comprehensively identified. By extrapolating typical fragmentation

of sphingolipid classes measured from authentic standards, comprehensive MS/MS spectral libraries were built. With the libraries, sphingolipids in mouse ear tissue, mouse liver tissue, and two types of human cells (HeLa and HEK) were throughly identified. Around 150-230 sphingolipids were identified on each of the samples [2].

3.1.2 Detailed annotation of diagnostic fragments produced from each sphingolipid class

In order to elucidate typical fragmentation patterns of sphingolipid classes, MS-FINDER program with HR rules [1] was utilized to annotate substructures to diagnostic fragment ions. Sphingolipid molecules representative of their class were measured with negative ion mode ESI-MS/MS. Substructures of the sphingolipid were assigned to product ion peaks in the measured MS/MS spectrum counting in hydrogen movements according to the HR rules with manual curation.

Diagnostic fragments produced from Cer [NS] are shown in Figure 3.2. Fragment a, corresponding to loss of 30 (CH_2O) or 32 (CH_3OH) Da, can determine the header; loss of hexose or phosphocholine headers must result in larger decrease in mass. Fragments b, c, and d are specific to fatty acids. On the fatty acid moiety, length of its carbon chain and nonexistence of a hydroxy group ([A] or [B]) or an ester bond ([EO]) are confirmed from mass of these fragments. From fragments e and f, the structure of the sphingoid base is determined. They reflect length of the carbon chain and existence of the double bond in sphingosine [S]. For dihydrosphingosine [DS] or phytosphingosine [P], mass of 2 Da (two H atoms) or 18 Da (two H and one O atoms) will be added there, respectively. Fragments a and b showed dehydration which resulted in loss of 16 Da.

Ceramide having an α-hydroxyl fatty acid, Cer [AS], showed additional diagnostic fragments (Figure 3.3). Fragments a, b, d, and e of Cer [AS] corresponds to addition of a hydroxy group to fragments a, b, c, and d of Cer [NS], respectively. The identical one to fragment f of Cer [AS] was observed in Cer [NS] as fragment f. Here, a recombined structure of fragment d, described as −NH+O, was detected. The recombination frequently occurs when measuring ceramides in negative-ion mode ESI-MS/MS [51]. Actually, Cer [NS] also displayed the recombination, but its abundance was too low to be detected. Fragment c was characteristic of Cer [AS] as well as its subsequent fragment h. Fragment g resulting from dissociation of the α carbon was also newly

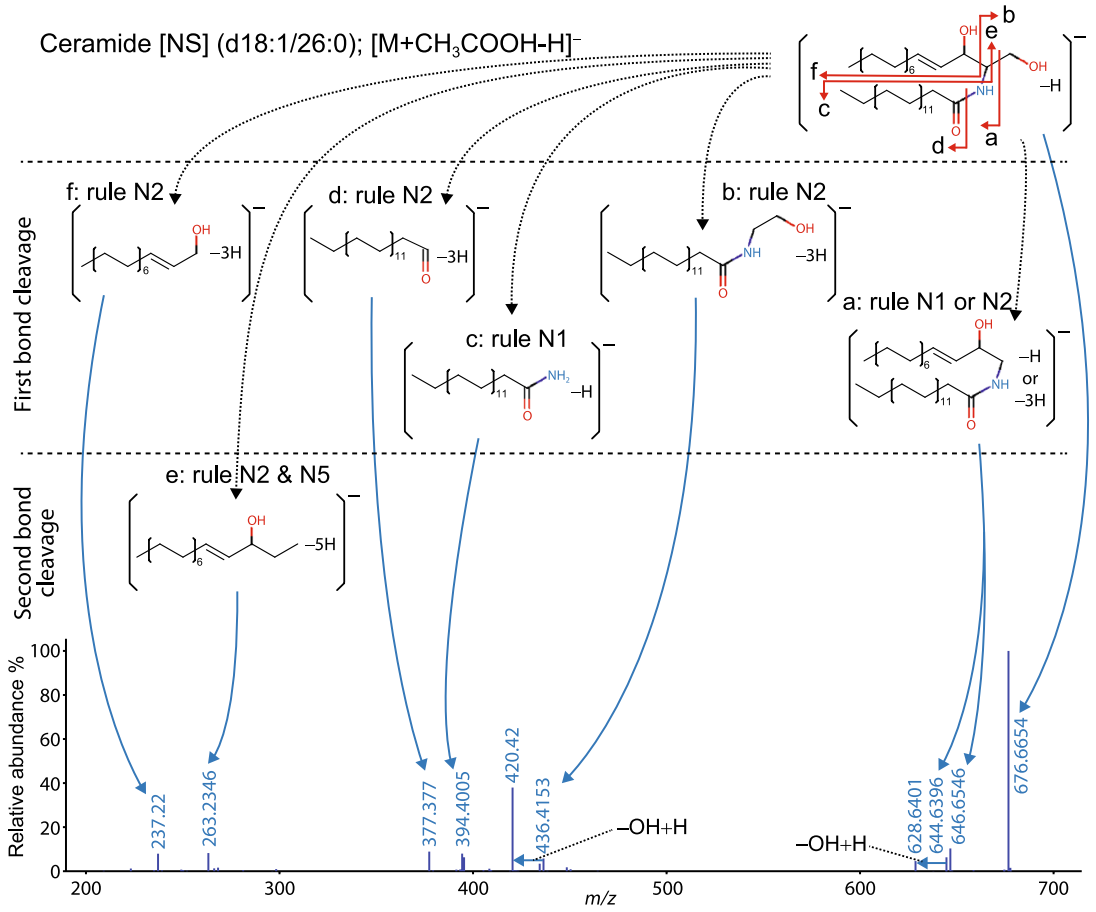


Figure 3.2: Diagnostic fragments of a Cer [NS] molecule. The substructures are described in neutral form with the number of rearranged hydrogen atoms. Reproduced from Ref. [2].

observed.

Representative molecules of other sphingolipid classes were also measured in negative-ion mode ESI-MS/MS, and their diagnostic fragments were annotated to measured MS/MS spectra. Based on the annotated diagnostic fragments, theoretical MS/MS spectral libraries for sphingolipid classes were constructed.

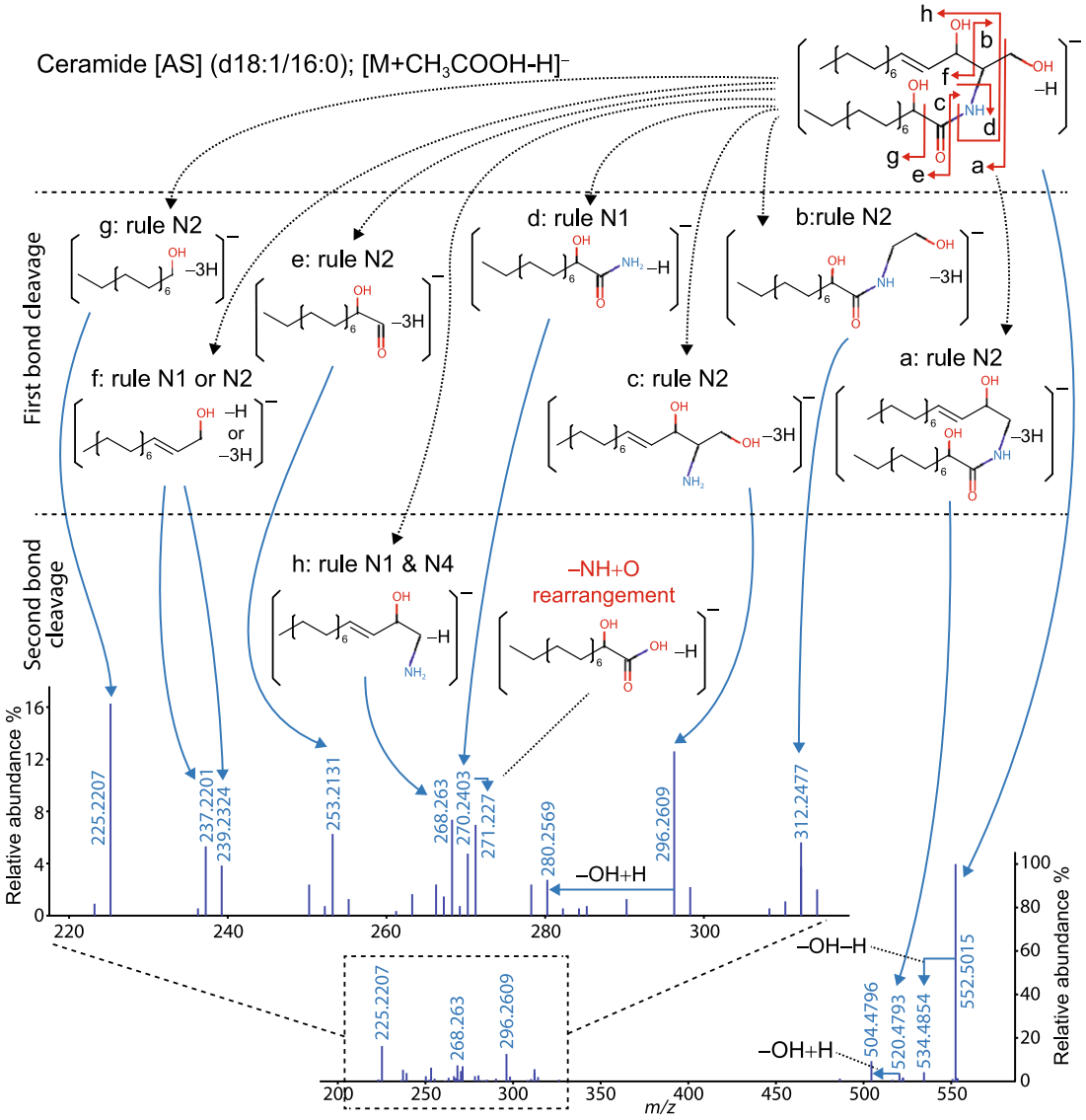


Figure 3.3: Diagnostic fragments of a Cer [AS] molecule. The substructures are described in neutral form with the number of rearranged hydrogen atoms. Reproduced from Ref. [2].

3.1.3 Absence of authentic standards for ceramides with a β-hydroxyl fatty acid

Madison *et al.* suggested that Cer [BS] existed in mouse skin by the chromatographic study [50]. There were some signs that Cer [BS] was likely to exist also in the

LC-MS/MS measurement of the murine samples in Ref. [2]. Positional difference of a hydroxy group in mass-identical pairs, such as Cer [BS] and Cer [AS] or Cer [BDS] and Cer [NP] (see Figure 3.1), cannot be distinguished by MS because the mass is identical. To identify ceramides with β-hydroxyl fatty acids, theoretical evidences which can be obtained without measuring authentic standards were required.

One method is predicting retention time (RT) of ceramides in LC. To predict the RT, Log *P* value, which is an index of hydrophobicity and known to correlate with the RT [52], was estimated with XLogP algorithm [53]. In addition to Log *P* value, PaDEL program [54] was utilized to estimate other properties. Several properties as well as Log *P* were selected to build a regression model to predict RT of ceramides. From the prediction using the regression model, RT of Cer [BS] was earlier than Cer [AS] [2].

Another is using MS/MS spectra. As any authentic standards of ceramides with a β-hydroxyl fatty acid were unavailable, a theoretical library of the ceramide classes cannot be built from MS/MS spectra of authentic standards. Then, I performed energetic analysis on fragmentation of Cer [BS] by computational chemistry. The computation revealed that a fragment unique to a ceramide likely to be Cer [BS] was truly produced from the Cer [BS] structure. Comparison with calculation results on another possible structure ensured the existence of Cer [BS]. Details of my computation will be described in the next section.

3.2 Identifying the unknown molecule by using computational chemistry

3.2.1 Computational methods

The computational method I used here was also semi-empirical MO calculation, whose details are described in Subsection 2.2.2. The software was updated to MOPAC2016 [55]; however, the computational method was same (PM7 parameterization [44, 45]).

To obtain the maximum enthalpy change, or "activation enthalpy," along a fragmentation process, one-dimensional grid calculation was conducted (Figure 3.4). One bond in a molecule is selected and repeatedly elongated by a small step size, e.g. 0.05 Å. On each grid point, the molecular structure is optimized, and its enthalpy is calculated.

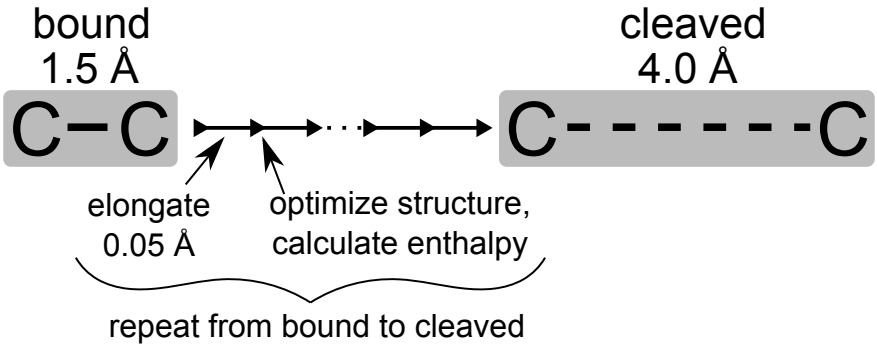


Figure 3.4: Conceptual diagram of one-dimensional grid calculation on cleavage of the C–C bond.

The elongation and calculation are repeated until the structure reaches a destinated structure. Here, the initial structure is a bound molecule and the final structure is dissociated two molecules. To simulate fragmentation processes of ceramides, bond elongation from 1.5 Å to 4.0 Å with step size of 0.05 Å, i.e. 51 grid points, was computed.

Since the grid calculation requires many repeated computations, stronger computational resource was needed. NIG Supercomputer System was utilized for the grid calculation. It took around two hours to finish the grid calculation on one ceramide with 32 GB memory (4 GB × 8 CPUs). Two long carbon chains, a fatty acid and a sphingoid base, increased the computational time of each structural optimization.

3.2.2 Energetic comparison of fragment processes to annotate its precursor structure

The candidate molecule for Cer [BS] produced a characteristic fragment with $m/z = 340$ (fragment a in Figure 3.5). It is produced by α-elimination, where α carbon of hydroxy group is dissociated, from ceramides with a β-hydroxyl fatty acid. The α-elimination is common in fragmentation of sphingolipids [51].

Another candidate which might be recognized as the precursor of the fragment a in Figure 3.5 was a γ-hydroxyl ceramide. It has a hydroxy group on the C atom next to the one that a β-hydroxy group is attached to; this is an imaginary structure. By predicting RT and checking MS/MS spectra in positive-ion mode, other isobaric ceramides such as Cer [AS] and Cer [NP], which have a hydroxy group on different

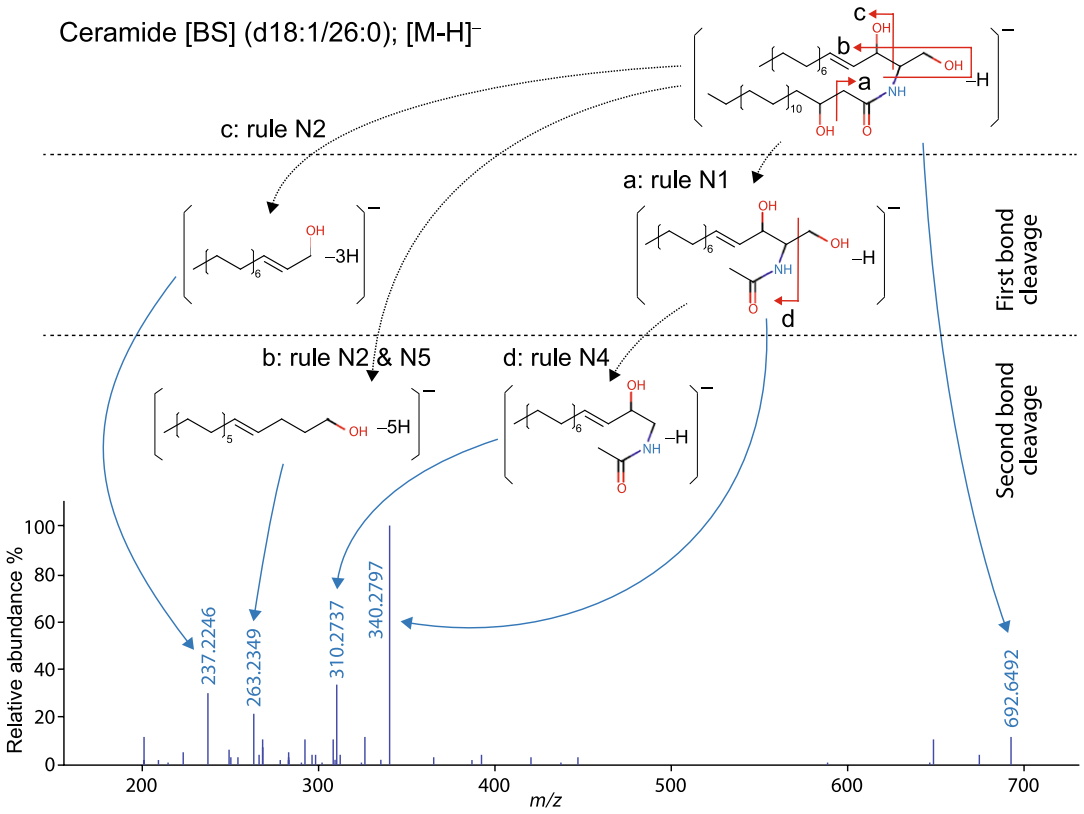


Figure 3.5: The experimental MS/MS spectrum of Cer [BS] in negative-ion mode ESI-MS/MS. Fragment a is most abundant and characteristic to the β-hydroxy group. Reproduced from Ref. [2].

positions, were excluded from the candidates [2]. Then, the problem was to confirm that the characteristic fragment with $m/z = 340$ was correctly produced from the β-hydroxyl ceremide, not the γ-hydroxyl ceramide.

Possible fragmentation schemes of β- and γ-hydroxyl ceramides are summarized in Figure 3.6. Scheme A is α-elimination on the β-hydroxyl ceramide resulting in the diagnostic fragment a in Figure 3.5. The negatively charged hydroxy group was transformed into a formyl group by the α-elimination. Then the negative charge transferred to the dissociated C atom and spread to the O atom of the carbonyl group. Both the neutral fragment and the product ion were considered structurally stable. To produce the same fragment, the γ-hydroxyl ceramide needs to undergo Scheme B in Figure 3.6. Here, the structure of neutral fragment was different from one produced in

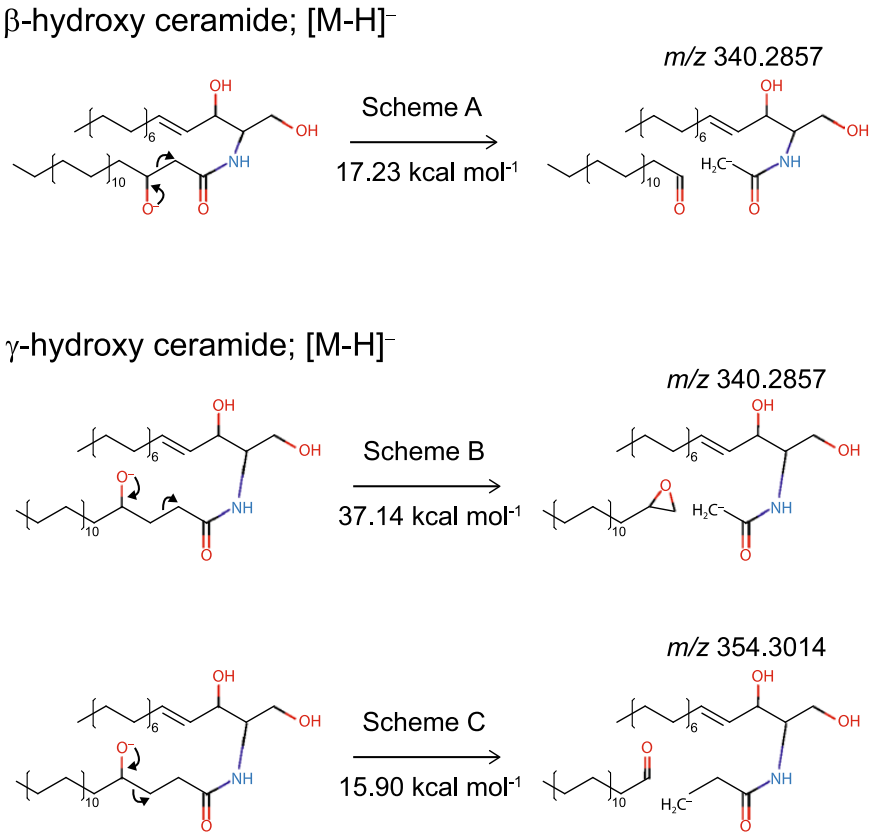


Figure 3.6: Fragmentation schemes of β- and γ-hydroxyl ceramides. Schemes A and B result in the characteristic fragment a in Figure 3.5, while Scheme C produces another fragment. Reproduced from Ref. [2]

Scheme A in contrast to the same structure of the product ion. The epoxy structure, having a small triangle, caused large distorsion of the angles between the bonds; the triangle had angles around 60 degrees, which were much smaller than the stable angle of 109.5 degrees. Scheme C is α-elimination on the γ-hydroxyl ceramide, which leads to the different structure of the product ion. In this case, neutral fragment formed a stable formyl group instead of the unstable triangle structure.

Activation enthalpy for each scheme is written under the arrow in Figure 3.6. Whole enthalpy changes along the schemes are plotted in Figure 3.7. The graph shows enthalpy changes along the bond elongation from 1.5 to 3.0 Å; there were no significant enthalpy changes after the bond length became 3.0 Å. Both Scheme A and C were α-elimination and showed similarly low activation enthalpy, 17.23 and

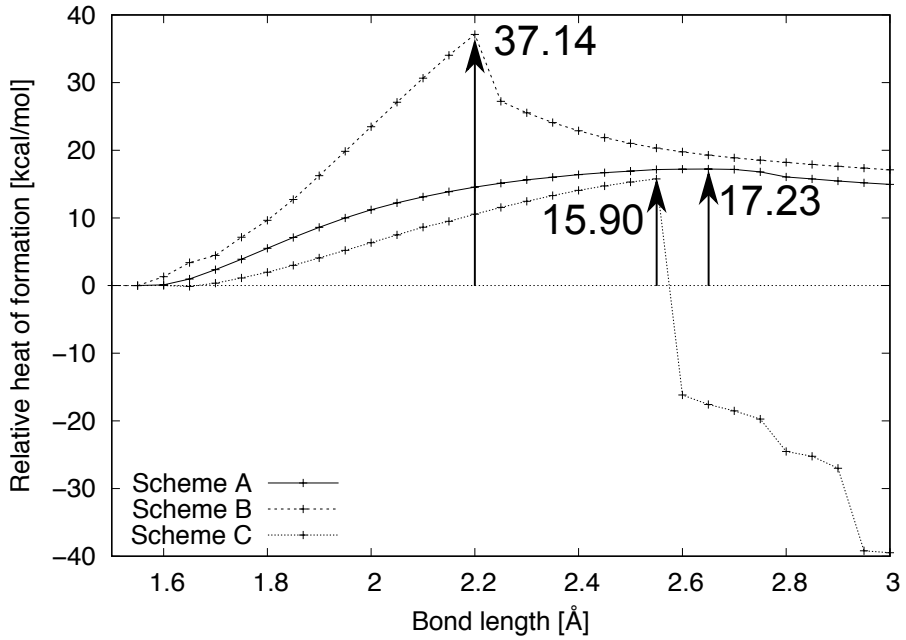


Figure 3.7: Enthalpy changes along the bond elongation in Schemes A, B, and C illustrated in Figure 3.6. Arrows indicate the activation enthalpy, which is the maximum change of enthalpy.

15.90 kcal mol^{-1}, respectively. On the other hand, Scheme B showed much higher activation enthalpy, 37.14 kcal mol^{-1}. With lower activation enthalpy, fragmentation proceeds exponentially faster. This result indicated two facts: the diagnostic fragment with $m/z = 340$ was easily produced from the β-hydroxyl ceramide (Scheme A), and production of the fragment from the γ-hydroxyl ceramide (Scheme B) would be inhibited by much faster fragmentation where the fragment with $m/z = 354$ was produced (Scheme C). Therefore, we concluded that the MS/MS spectrum having the fragment with $m/z = 340$ (Figure 3.5) was truly measured from the Cer [BS].

Large decrease of enthalpy in Scheme C in Figure 3.7 was caused by Van der Waals interaction between the fatty acid and the sphingoid base; it was a kind of artificial parts of the computation. In the initial state, carbon chains of the fatty acid and the sphingoid base were separated. After the bond cleavage, the two chains contacted with each other in parallel as they lost a structural hindrance in their connected site. The contact along their carbon chains reach broadly and decreased enthalpy drastically with the intermolecular forces. In practical MS/MS measurement, provided energy

is so large that fragments split immediately. Therefore, the enthalpy drop after the cleavage should be ignored here. That is the reason why the maximum enthalpy was employed to discuss existence of the fragments.

Since the MS/MS spectrum in Figure 3.5 was identified as Cer [BS], theoretical MS/MS spectral libraries for β-hydroxyl ceramides were constructed with extrapolation of the MS/MS spectrum. By using them, a total of 415 lipids was identified from four kinds of samples: murine ear, murine liver, HeLa cell, and HEK cell. Ceramides with a β-hydroxy group were specific to murine ear; four structures of Cer [BDS] and six structures of Cer [BS] were identified. Human cells and murine liver did not contain any Cer [BDS] nor Cer [BS]. Means of computational chemistry assisted with identifying sphingolipid classes whose authentic standards were unavailable.

3.3 Conclusion

Theoretical MS/MS spectral libraries were constructed by extrapolating typical fragmentation of their representative molecules in order to comprehensively identify lipids contained by the biological samples. With authentic standards, experimental MS/MS spectra was obtained from the measurements, and theoretical libraries were constructed on the basis of them. The problem was about ceramides with a β-hydroxy group; no authentic standards of them were available. Against that, I performed computational precursor annotation by computing activation enthalpy to confirm existence of the fragment. My computation excluded the wrong candidate, a γ-hydroxyl ceramide, and proved that the obtained experimental spectrum was surely gained from the β-hydorxy ceramide. The examined spectrum was utilized to construct the theoretical libraries, and inclusion of β-hydorxy ceramides of murine ear was confirmed by using them.

From these analyses, two concepts were verified to be useful to analyze fragmentation pathways: activation enthalpy and adjacent chemical groups. Activation enthalpy pointed easiness of bond cleavage out more properly than enthalpy difference between precursor and product structures. Product structures are sometimes, especially when the molecules is large, stabilized in surplus by unintended interactions. By considering activation enthalpy, we can exclude the excess stabilization. The lowered activation enthalpy of Scheme C compared with Scheme B in Figure 3.6 originated from the

hydroxy group adjacent to the cleaved bond. That implied chemical groups were possible to decrease activation enthalpy by altering electron structures around them. In the next chapter, I will discuss computational fragment prediction with the two concepts.

4

Physicochemical prediction of metabolite fragmentation in MS/MS

4.1 Introduction

4.1.1 Limitation of metabolite identification using standard spectral libraries

In metabolome analyses, MS/MS is a powerful tool to identify metabolites from biological samples. In MS/MS, metabolites are fragmented into their substructures and the fragmentation pattern at each time slice is recorded as an MS/MS spectrum. Since MS/MS spectra are different for each structure, they are used as fingerprints to identify metabolites, especially to distinguish structurally isomeric metabolites. For the purpose of metabolite identification, many databases, either commercial or public domain, provide standard MS/MS spectra[25, 26, 27, 28, 29]. The matching of a

standard spectrum with an experimentally observed one is considered a prerequisite of the metabolite identification.

An apparent drawback of the library approach is that no database is comprehensive in a true sense; we cannot prepare spectra for all possible metabolites. In the PubChem database for example, over 94 million structures, including synthetic ones, are registered. To obtain a standard MS/MS spectrum, we need to measure a pure, i.e. commercially extracted, metabolite with a MS/MS platform one by one, and the measurement is a complicated and costly process. It is therefore technically difficult to prepare a spectral library of more than thousands of standard compounds.

4.1.2 Current prediction methods to build theoretical standard libraries

To overcome this intrinsic limitation, many research groups have challenged to predict MS/MS spectra only from molecular structures. For phosphoglycerolipids, whose fragmentation pattern is well known, a theoretical library has been constructed by extrapolating their fragment mass values and their intensities for possible structural variants combinatorially [47]. This approach should be applicable for other structure classes. The library construction, however, seems much harder for small metabolites, especially for those containing heteroatoms such as oxygen and nitrogen. Actually, FravonoidSearch, which is a throughly curated theoretical MS/MS spectral library for flavonoids, required several years to be built because of much complex variations of flavonoid structures [56].

One successful spectrum prediction was achieved by machine learning systems [32, 33], which learn fragmentation patterns from existing spectra in databases. Other groups systematically classified fragmentation patterns in order to understand the mechanism of fragmentation in MS/MS [57, 58, 59]. Nevertheless, how fragmentation processes proceed in MS/MS remains largely unclear and theoretical clues for more accurate prediction have been awaited.

4.1.3 Incarnating mechanistic fragment prediction by physicochemical theories

Here, I focus on physicochemical properties of metabolites. As stated in Chapter 2 and 3, I have calculated bond cleavage energy and electron distributions of some metabolite structures by means of computational chemistry and clarified mechanisms of their fragmentation pathways in ESI-MS/MS.

I extend this physicochemical approach to predict fundamental fragmentation in ESI-MS/MS. Each bond is assigned a "bonding pattern" (BP) including its neighborhood, i.e. two atoms and chemical groups adjacent to the bond. Then activation energy of each BP is calculated with computational chemistry. Prediction of fragmentation is performed by finding the lowest activation energy and its comparison with experimentally measured spectra are explained. Since any molecular structure can be regarded as a composition of BPs, this method is applicable to a variety of metabolites in principle.

4.2 Prediction methods

4.2.1 Assignment of bonding pattern and cleavage

BPs consist of two bound atoms and their adjacent chemical groups. They are assigned for all single bonds between a carbon atom and a heteroatom, which are principal candidates for cleavage. Single bonds between two carbon atoms or multiple bonds are excluded from the cleavage candidates.

For each BP, two types of bond cleavage, "direct" and "rearranged," are computed. When a BP is cleaved, the heteroatom adjacent to the bond is assumed to be protonated. Since our study presuppose fragmentation in positive-ion mode ESI-MS/MS, only positive even-electron ions are considered [39]; although there were some example of odd-electron ions in Chapter 2, they were rare cases. Direct cleavage is not accompanied with hydrogen movement; a positive charge moves to the carbon atom. Rearranged cleavage is associated with hydrogen movement from the carbon atom side to the heteroatom. If small molecules such as water, ammonia, ethylene, and carbon monoxide are dissociated, hydrogen movement to such small molecules is ignored.

All BPs analyzed in the present study are summarized in Table 4.1. They are chosen

Table 4.1: Bonding patterns analyzed in the present study.

No.	bound atoms	neighboring groups		activation energy (kJ mol^{-1}) direct	rearranged
1	C–N	C($-CONH_2$)	N(none)	203	–
2	C–N	C($-CH_3$, =O)	N($-CH_3$)	124	95
3	C–N	C($-CH_3$, =O)	N($=CH_2$)	100	68
4	C–N	C($-CH_3$, –COOH)	N(none)	236	–
5	C–N	C($-CH_3$, –COOH)	N($-COCH_3$)	197	161
6	C–O	C($-CH_3$, =O)	O(none)	25	–
7	C–C	C($-NH_2$)	C(=O)	−99	–
8	C–C	C($-CH_3$, $-NHCH_3$)	C(=O)	−153	–
9	C–N	C($-CH_3$)	N(none)	245	–
10	C–N	C($-CH_3$)	N($-CNHNH_2$)	238	132
11	C–N	C($-CH_3$)	N(=C=NH)	206	–
12	C–N	C(=NH)	N($-CH_3$)	403	–
13	C–N	C($-NH_2$, =NH)	N($-CH_3$)	105	79
14	C–N	C($-NHCH_3$, =NH)	N(none)	51	–
15	C–O	C($-CH_3$)	O($-COCH_3$)	–	105
16	C–O	C($-CH_3$, =O)	O($-CH_3$)	28	91
17	C–C	C($-CH_3$, $-NH_2$)	C(=O)	−129	–

from two example molecules: leucylglycine and ethyl argininate (Figure 4.1(a) and (b), respectively). The numbers in Figure 4.1 correspond to the number of BPs described in Table 4.1. Cleavage activation energy is an indicator of cleavage easiness. By assigning BPs and their cleavage activation energy on chemical bonds composing a metabolite structure, fragmentation pathways constituted by several bond cleavage steps are predicted.

To make the computation simpler, substructures distant from the bond are replaced with hydrogen atoms. For example, in BP 5, a precise group next to the bond is isobutyl ($-CH_2CH(CH_3)_2$). However, the group is replaced with a methyl group ($-CH_3$, see Figure 4.1(a) and Table 4.1). This simplification is applied for all BPs.

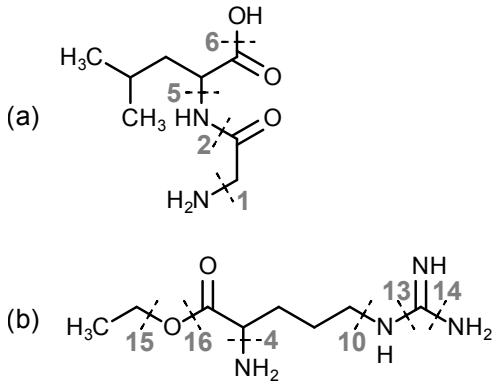


Figure 4.1: Assignment of BPs on (a) leucylglycine and (b) ethyl argininate. Bonds crossed by dashed lines are assigned BPs. Only single bonds between a carbon atom and a heteroatom, which are likely to be cleaved, are considered.

4.2.2 Calculation of activation energy

Activation energy required to cleave each BP is calculated with computational chemistry software. Here, stability of molecular structures are evaluated with "Gibbs free energy," which includes an entropy term in addition to enthalpy. First, a model molecular structure of each BP is built computationally. Then a reaction pathway is simulated from the initial connected structure to the cleaved structure, and the corresponding change of Gibbs free energy along the dissociation is calculated by Reaction plus [60]. If the dissociation process exhibits maximal free energy between the initial and the cleaved state, transition state optimization starting from the structure with maximal free energy is performed by Gaussian 09 [61]. Activation energy is defined as the energy difference between the initial state and the transition state. If the initial or cleaved state has the maximal energy, activation energy is defined as the free energy difference from the initial state to the cleaved state. This means that the activation energy is negative when the initial state is high-energy.

High-precision Gibbs free energy of the initial state, transition state and fragmented state (ions and neutrals) is calculated by the CBS-QB3 method [62], where a typical error is around 1 kcal mol^{-1}[63]. Structural optimization of all molecules is performed with the density functional theory (DFT) method. Functional and basis set are CAM-B3LYP including the long-range correction and 6-31+G(d,p) including polarization

and diffusion of atomic orbital, respectively. The long-range correction is important for bond-breaking and making reactions [64]. All calculations here are done with Gaussian 09.

4.2.3 Computational time

The computation time depends on the structural environment, especially the number of atoms, of each BP. It typically takes around three hours to calculate activation energy of one BP, which includes reaction pathway optimization and high-precision energy calculation. In a few cases, computational time exceeded ten hours. For example, computation for BP 5 composed of 18 atoms required 19 hours to obtain activation energy.

Computational environments are as follows. For reaction pathway optimization, 246.4 GB memory (7.7 GB × 32 cores, Intel Xeon E5-2690) was utilized. For structural optimization and high-precision energy calculation of each molecule (initial, transition, and fragmented states), 61.6 GB memory (7.7 GB × 8 cores, Intel Xeon E5-2690) was utilized. All the computations were executed with supercomputer systems in Research Center for Computational Science, Okazaki, Japan.

4.3 Prediction results for two example molecules

4.3.1 Strategy of the prediction using bonding patterns

The fragment prediction is based on assignment of BPs, which represent the vicinity of cleaved bond, with their cleavage activation energy. All bonds composing a metabolite are scanned, and candidates for bond cleavage, i.e. single bonds formed by a carbon atom and a heteroatom, are transcribed as BPs which include two atoms forming the bond and its neighboring chemical groups. A list of activation energy of BPs is referred to, and values of activation energy are assigned on the candidate bonds to be cleaved. By cleaving the candidate bonds with low activation energy, fragment structures are predicted. Assignment of BPs on the predicted fragment structures predicts their further fragmentation. Collecting m/z values of the predicted fragments yields a theoretical MS/MS spectrum. Here, I will present example processes of fragment

prediction on leucylglycine and ethyl argininate.

Activation energy for cleavage of each BP can be calculated irrespective of metabolite structures and stored in the form of Table 4.1, though here only BPs included in leucylglycine and ethyl argininate are listed to explain examples of the prediction. By expanding the table of calculated activation energy for BPs, the coverage of our prediction method becomes broader. Rough estimation of the number of BPs possibly existing in all metabolites is presented in Section 4.4. How to reduce the BPs in order to cover the huge metabolite space is also discussed.

4.3.2 Leucylglycine

Activation energy for cleavage of each BP in leucylglycine was assigned from the upper part of Table 4.1. The lowest activation energy was obtained as direct cleavage of BP 6 (25 kJ mol^{-1}), i.e. the C-O bond was the easiest to cleave. This cleavage corresponded to water elimination from a carboxyl group, which is common fragmentation in MS/MS [59]. Rearranged cleavage of the BP 2 showed relatively low activation energy (95 kJ mol^{-1}). This was an amide cleavage and its product ion was equivalent to the y-ion of a general peptide bond cleavage [65]. Through this analysis, we could confirm that the activation energy was indeed lower for experimentally obtained fragmentations.

Next, activation energy for further fragmentation was calculated for leucylglycine (Figure 4.2). In the second fragmentations, C–C bonds might become cleavage candidates because a carbon atom might be charged after the first fragmentation. A positively charged carbonyl group by the cleavage of BP 2(direct) or 6(direct) may be dissociated as a carbon monoxide. In this case, the C–C bond cleavage becomes energetically favorable; otherwise the C–C cleavage requires much higher energy and not regarded as a candidate for fragmentation. Some of the second fragmentation showed negative activation energies, i.e. the product structure is more stable than the initial structure. In particular, cleavage of BP 6 is followed by the cleavage of BP 8 immediately (activation energy = -153 kJ mol^{-1}), and the product ions may undergo further rearranged cleavages of BP 3 with relatively low activation energy (68 kJ mol^{-1}).

Fragmentation pathways of leucylglycine were predicted from the activation energies. The pathway with the lowest energy, i.e. the most likely to occur in

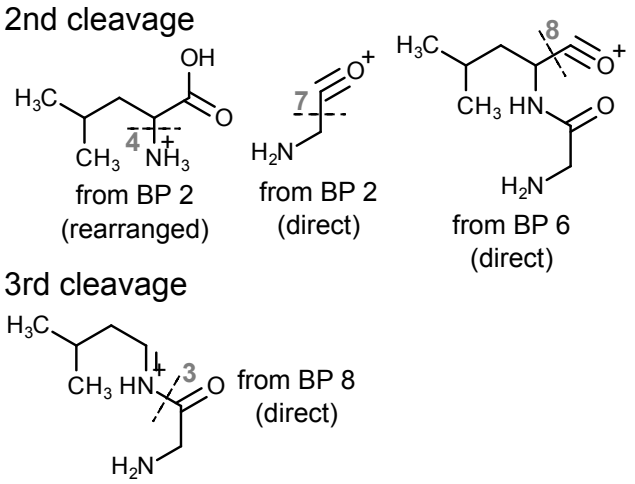


Figure 4.2: Second and third fragmentation of product ions from leucylglycine. The numerals correspond to the number of BPs in Table 4.1. Cleavage of the BP leading to each fragment is shown on its bottom or right.

experiments, was the consecutive cleavage of BPs 6(direct) → 8(direct) → 3(rearranged), where activation energies were 25 → −153 → 68 kJ mol^{-1}, respectively. Rearranged cleavage of BP 2 (activation energy = 95 kJ mol^{-1}) was also energetically favorable, but the cleavage would not proceed to the next cleavage of BP 4 due to its high activation energy (253 kJ mol^{-1}). The predicted fragmentation pathways are illustrated in Figure 4.3. When compared with a measured standard spectrum from MassBank database (MassBank ID: KO003025) [26] in Figure 4.4, the prediction successfully reproduced the major fragmentations.

4.3.3 Ethyl Argininate

Activation energy for cleavage of each BP in ethyl argininate was assigned from the lower part of Table 4.1 (see Figure 4.1 (b)). Here, BP 4, which appeared in the leucylglycine molecule, was assigned again on the middle amino group. The lowest energy was calculated as the direct cleavage of BP 15 (28 kJ mol^{-1}), which corresponded to the dissociation of an alcohol from the ester bond moiety. It is noteworthy that this cleavage is similar to water elimination from the carboxyl group with the lowest activation energy (25 kJ mol^{-1}) in BPs on leucylglycine. This fact implies that effects on bond cleavage of a methoxy group is not much different from that of a hydroxy

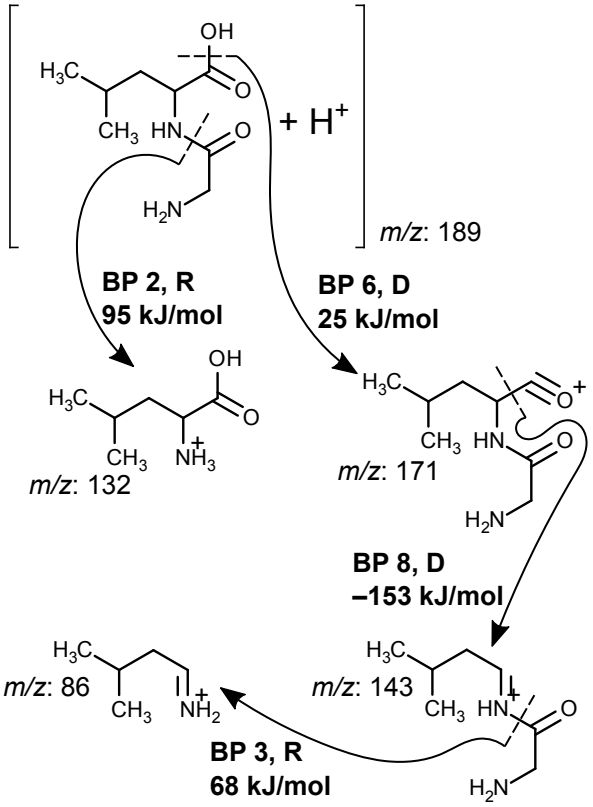


Figure 4.3: Predicted fragmentation pathways of leucylglycine. The cleaved BP with D (direct) or R (rearranged) and its activation energy are described. Initial direction of arrows indicates the side of the protonated substructures.

group. Rearranged cleavage of BP 13 (dissociation of carbodiimide from the guanidine moiety) and direct cleavage of BP 14 (ammonia elimination) also showed relatively low activation energy (79 and 51 kJ mol^{-1}, respectively).

For ethyl argininate, no third-step cleavage was considered stable and only second-step cleavage was investigated (see Figure 4.5). The fragmentation pathway with the lowest activation energy was the cleavage of BPs 14(direct) → 17(direct) (28 → −129 kJ mol^{-1}), which resulted in the fragment of $m/z = 129$. The other second fragmentation required more than 200 kJ mol^{-1} to be activated.

By selecting relatively low activation energy from the calculation results, m/z values of fragments predicted to be produced are 161 (rearranged cleavage of BP 13 activated by 79 kJ mol^{-1}), 186 (direct cleavage of BP 14 activated by 51 kJ mol^{-1}), 157

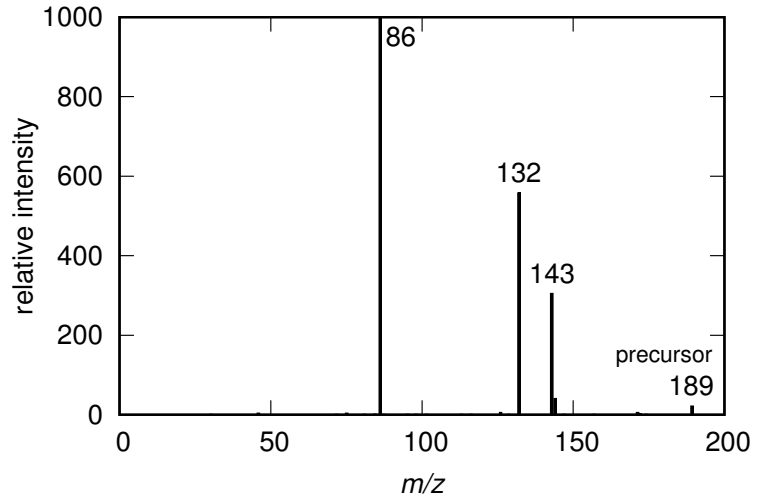


Figure 4.4: A standard MS/MS spectrum of leucylglycine (MassBank ID: KO003025). Numerals on top of major peaks indicate m/z. The spectrum was experimentally obtained by positive-ion mode LC-ESI-QQ MS/MS with collision energy = 20 eV.

(direct cleavage of BP 16 activated by 28 kJ mol^{-1}), and 129 (consecutive direct cleavage of BPs 16 and 17 activated by 28 and −129 kJ mol^{-1}, respectively) (see Figure 4.6). Rearranged cleavage of BP 16 can be activated by relatively low 91 kJ mol^{-1} of energy; whereas, direct cleavage of BP 16 is activated by much lower energy (28 kJ mol^{-1}) and may suppress the rearranged cleavage of BP 16 as its competitor. This prediction was coincided with a measured standard spectrum from MassBank (MassBank ID: KO002249) in Figure 4.7 only at one major peak with m/z = 186.

Two of the major peaks (m/z = 144 and 60) in the standard spectrum were predicted to require high activation energy (238 kJ $mol^{-1}$for direct cleavage of BP 10 and 132 kJ $mol^{-1}$for rearranged cleavage of BP 10, respectively). The fragment with m/z = 144 could also be produced from a two-step cleavage of BPs 14(direct) → 11(direct), which required 51 → 206 kJ mol^{-1}, respectively, or BPs 13(rearranged) → 9(direct), which required 79 → 245 kJ mol^{-1}, respectively. The former was the lower-energy pathway to produce the fragment of m/z = 144 than the latter; however, the maximum activation energy of 206 kJ $mol^{-1}$is still high. On ethyl argininate, intensities of the product ions relative to the precursor ion were lower in comparison with leucylglycine. This implies that the precursor ion of ethyl argininate is harder to decompose than that of leucylglycine. That is consistent with the observed relatively high activation energies.

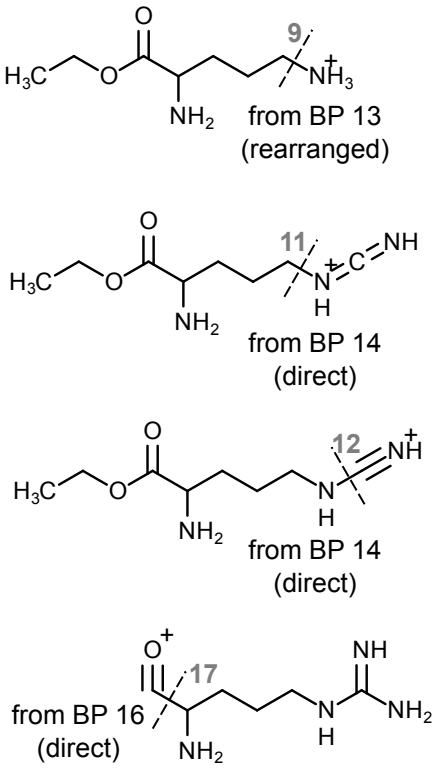


Figure 4.5: Second fragmentation of product ions from ethyl argininate. Two fragmentation pathways can be considered for resonance structures of the product ion from direct cleavage of bonding pattern 14.

The fragment of m/z = 70 never appeared in the prediction. This missing fragment most likely came from a charge remote fragmentation (see Fig. 4.6). It is produced when the fragment produced with the lowest activation energy, whose m/z is 129, loses 59 Da. That loss of mass is regarded as elimination of guanidine moiety on the right side of ethyl argininate. This fragmentation is called charge remote fragmentation, where a bond distant from the charged site is cleaved. In the present prediction, no charge remote fragmentation was considered. Energy released by the cleavage of BP 17 (129 kJ mol^{-1}) might help the further fragmentation.

Another inconsistency was the absence of m/z 161 in the standard spectrum despite its relatively low activation energy (79 kJ mol^{-1}). This is explained by further loss of ammonia resulting in m/z = 144; however, activation energy of the cleavage is high (245 kJ mol^{-1}). Such high activation energy should come from an unstable structure of

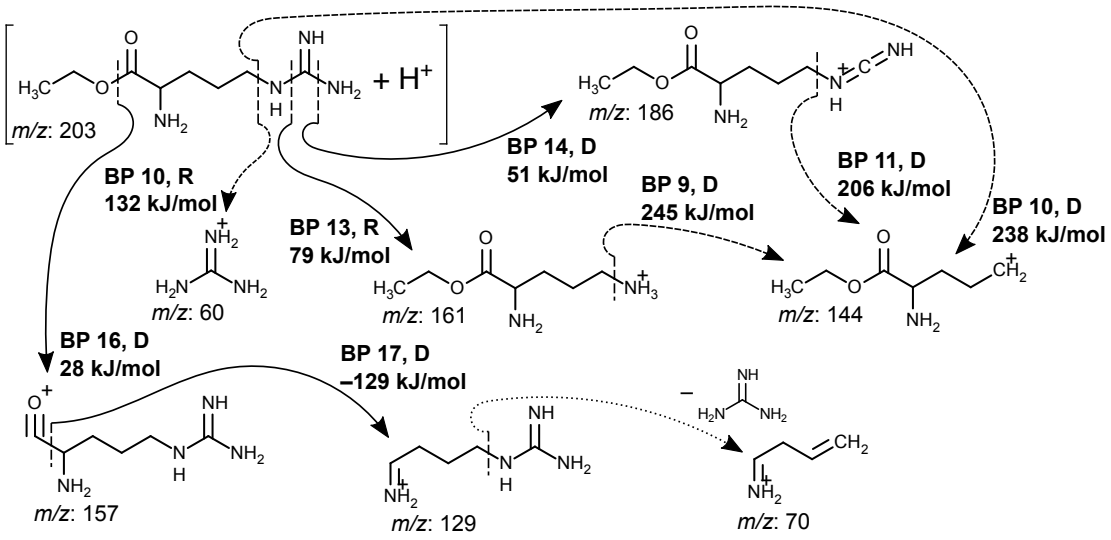


Figure 4.6: Tentative fragmentation pathways of ethyl argininate. Solid and dashed arrows indicate pathways with low activation energy and high activation energy (> 100 kJ mol^{-1}), respectively. The cleaved BP with D (direct) or R (rearranged) and its activation energy are described. The pathway with a dotted arrow is not predicted but explaining the standard spectrum. Initial direction of the arrows indicates which side of the cleaved bond is detected as a product ion.

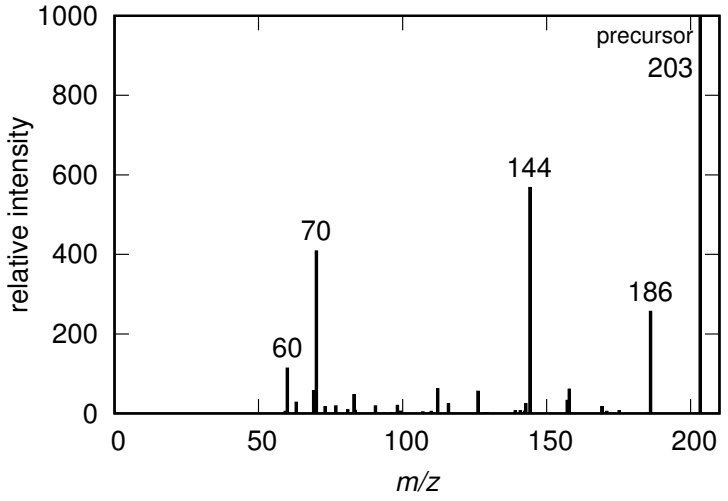


Figure 4.7: A standard MS/MS spectrum of ethyl argininate (MassBank ID: KO002249). Numerals on top of peaks indicate m/z values. The spectrum was obtained by positive-ion mode LC-ESI-QQ MS/MS with collision energy = 20 eV.

the fragment ion having a charge on the carbon atom connected by two hydrogen atoms. The highly electrophilic structure could absorb electron lone pairs on the amino group or the carbonyl group to cause a ring structure, and the rearrangement could accelerate the ammonia elimination. In the introduction of our BPs, remote structures were ignored and caused the discrepancy between computation and actual fragmentation.

4.4 The number of bonding patterns which exist in all metabolite structures

To apply the prediction to any metabolites, we need to include all BPs which appear in metabolite structures. The question is how many BPs exist. When considering nearest-neighbor four kinds of atoms (hydrogen, carbon, nitrogen, and oxygen) to a cleaved bond, the number of BPs becomes as follows:

$$({}_4\mathrm{H}_3 + 3 \times 4 + 2) \times ({}_4\mathrm{H}_2 + 3 + 4) = 34 \times 17 = 578 \tag{4.1}$$

where H is "repeated combination," which is computed as ${}_n\mathrm{H}_r = {}_{n+r-1}\mathrm{C}_r$. The left and right parentheses indicate patterns of carbon atom side and heteroatom side, respectively (see Figure 4.8). If the patterns are extended to a broader range which is reached by two bonds from the cleaved bond, the number of patterns becomes roughly 70,000.

I am considering some strategies to reduce the number of BPs. First of all, the estimation above includes unnatural patterns like a nitrogen atom connected by two nitrogen atoms. Such patterns can be excluded by searching metabolite database.

Superposition of effects from multiple chemical groups should also be investigated. If an effect from a set of three methyl groups can be represented by superposing an effect from one methyl group, the number of BPs decreases drastically. In the fragment prediction of leucylglycine and ethyl argininate, there were some sets of similar BPs. For example, BP 6 and 16 were different only about the oxygen atom connected by a hydrogen atom and a methyl group, respectively, and their activation energies were almost same. BPs 7, 8, and 17 correspond to elimination of carbon monoxide. With

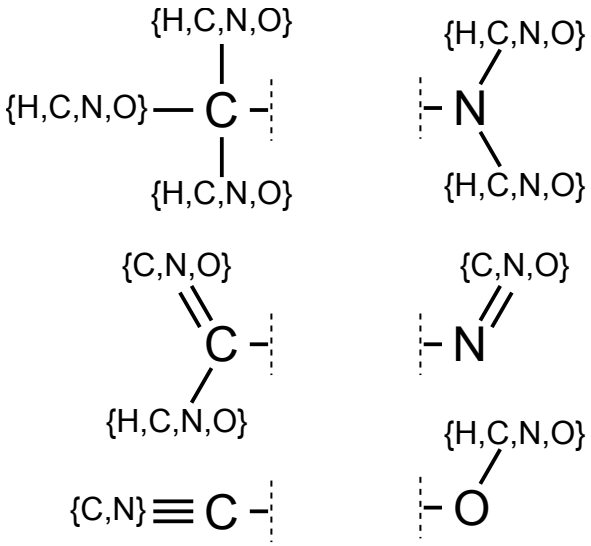


Figure 4.8: Possible patterns of atoms adjacent to a cleaved bond. The total number of the patterns is calculated in Equation 4.1.

more methyl groups added, their activation energies became lower (BP 7 > 17 > 8, see Table 4.1). These results suggest that addition of methyl groups on a carbon atom side reduces activation energy in contrast to addition on an oxygen atom side. Additional calculation of BPs with more methyl groups will contribute to elucidate superposed effects from multiple chemical groups.

Comparison of similar chemical groups is also important to simplify structures included in BPs. BP 9 is the most basic one, which includes only a methyl group as its neighboring group. BP 1 has a neighboring group founded on a carbon atom ($-CONH_2$), but the group is different from a methyl group and lowers activation energy by 42 kJ mol^{-1}. On the other hand, BP 4 including a methyl group and a carboxyl group, which are founded on carbon atoms, has similar activation energy to the basic BP 9. Such comparison will clarify whether chemical groups on a carbon atom, like $-CONH_2$ and a carboxyl group, can be treated as a methyl group or not. We are now investigating it by computing BPs composed of several carbon modifications.

4.5 Conclusion

Current standard MS/MS spectrum library is not sufficient for comprehensive identification of metabolites. I have presented MS/MS spectrum prediction by means of computational chemistry, and the prediction well matched with the standard spectrum of a dipeptide, leucylglycine. For another molecule, ethyl argininate, additional fragmentation pathways with high activation energy and recombination were required to reproduce its standard spectrum.

The examples show benefits and limitation of the approach: assignment of bonding patterns (BPs) and activation energy calculation. By focusing on a local structure, BPs are independent of size of metabolites and can be applied widely to most metabolites. Energetical analyses can provide quantitative results on fragmentation mechanism. However, two improvements are required to extend applicability of the prediction method. One is the introduction of charge remote fragmentation, which can be implemented by assigning BPs on bonds without ionization. The other is the consideration of energetic dependency applied to accelerate fragmentation in experimental systems. Activation energy can be converted into reaction rate of the fragmentation and its dependency on applied energy or temperature can be computed. The computation, however, will become combinatorially complex. From quantitatively detailed energetics, I plan to estimate plausible threshold of activation energy to judge whether a fragment ion is produced.

To extend the coverage of the prediction, I need to calculate more BPs and store their activation energy. As a simple estimation suggests enormous number of BPs to be calculated, we need to decrease the number by merging chemical groups having similar energetical effects. To achieve that, I currently work on additional computations in order to obtain correct simplification of chemical groups included in BPs.

5 Conclusion

5.1 Summary

In metabolome analyses, insufficiency of standard MS/MS spectral libraries is a bottleneck for comprehensive metabolite identification. The metabolite identification depends on spectral libraries, but costly processes to obtain a standard MS/MS spectrum impede expansion of the libraries. Theoretical approaches such as machine learning and empirical extrapolation have been developed; nevertheless, mechanisms of metabolite fragmentation in MS/MS measurements are widely unclear.

In this dissertation, I have discussed applicability of computational chemistry to mechanistic prediction of metabolite fragmentation in MS/MS. Molecular enthalpy derived from physicochemical calculation on metabolite structures was utilized to confirm existence of the fragments. Free energy extended from the enthalpy functioned as a criterion to predict proper fragment structures feasible to be produced.

In Chapter 2, enthalpy calculation provided evidence that fragments predicted by HR rules really existed. Physicochemical stability highly depends on arrangement

of electrons on the molecule. Derived from the even-electron rule, where behavior of electrons is regulated, the HR rules provide physicochemically stable structures of fragments. Enthalpy calculation proved that metabolite fragments observed in practical measurements were surely energetically stable.

Contribution of computational chemistry in Chapter 2 was not only quantitative evaluation of stability but also elucidation of the mechanism to stabilize the fragment structures. There were some odd-electron fragment ions, which violate the even-electron rule, in the statistics of experimental standard MS/MS spectral libraries, though they generally require much higher energy to dissociate. The MO calculation revealed the electron distribution over conjugated bonds of the odd-electron ions. Delocalization of the unpaired electron was quantitatively obtained as well as potential energy. The unstable isolated electron was shared by several atoms to decrease potential energy of the radical ions. This result emphasized that fragmentation tendency highly depended on electronic structures around the cleaved bond.

A concrete application of computational chemistry to identify metabolites without standard MS/MS spectra was presented in Chapter 3. An unknown experimental MS/MS spectrum was finally identified as a β-hydroxy ceramide, whose authentic standard is unavailable. Activation enthalpy calculation distinguished the proper fragmentation pathway to produce the observed fragment from other pathways unlikely to occur in MS/MS measurement. The fragmentation processes were fully simulated by computational chemistry with grid calculation. With that, undesired effects in the computation like intermolecular forces acting between fragments were eliminated. Clear division between proper and improper fragmentation pathway was achieved by computational chemistry.

An important concept arose in Chapter 3 was influence of neighboring groups. Provision of an electron pair from the neighboring chemical group fairly lowered the activation enthalpy required by the bond cleavage. The ability to provide an electron pair depends on electronic structure of the chemical group, such as lone electron pairs and π electrons. The results from the pathway simulation implied that, by investigating the electronic structures represented by the chemical groups in the vicinity of cleaved bonds, breakage tendency of the bonds would be estimated.

In Chapter 4, the ideas gained as above were integrated in prediction of MS/MS spectra only from metabolite structures. A metabolite structure was decomposed into

the set of bonding patterns. Activation energy to cleave each bonding pattern was precisely calculated by pathway calculation and transition state optimization with computational chemistry. The prediction succeeded in reproducing fragmentation of a dipeptide molecule experimentally measured with MS/MS. For another molecule having a long chain structure, ethyl argininate, the prediction failed to reproduce some of experimental fragments. From detailed inspection of failure prediction, necessity of charge-remote fragmentation and a quantitative formulation of correlation between activation energy and intensity of fragments was indicated.

This prediction method can be applied to any metabolite structures in principle, because bonding patterns are independent of structural size. Mechanisms to determine fragmentation pathways were theoretically and quantitatively explained in detail by energetic influence of chemical groups adjacent to the cleaved bond. The physicochemical strategy was proved to be able to predict plausible fragments produced from metabolites.

5.2 Future works

Currently, the fragment prediction I suggested is "ad hoc." Activation energy is calculated on demand about bonding patterns which appear in a selected molecule. In order to apply the prediction to the vast metabolite space, cleavage activation energy of possible bonding patterns should be computed beforehand. I am planning to compute and store activation energy of all bonding patterns.

To execute that, the problem is how many bonding patterns need to be considered. As Equation 4.1, possible patterns are roughly 600. However, when the bounds included in bonding patterns are extended, the number of patterns increases explosively. The proper simplification of bonding patterns requires to be elucidated. For example, whether energetic influence of a formyl group (-CHO) is different from an acetyl group ($-COCH_3$) is a question to solve. Another problem is superposition. If effect of three methyl groups is represented by adding that of one methyl group up, the number of bonding patterns to compute decreases drastically.

At present, energetic difference among similar chemical groups like formyl and acetyl is investigated. Activation energy to cleave a bonding pattern including each of the similar chemical groups is computed and compared with each other. From

the comparison, chemical groups which do not lower the activation energy are excluded from the candidate bonding patterns to be computed and stored. A trend of energetically influential groups has begun to appear gradually. I am also planning to investigate influence from superposition of chemical groups. By collecting the results, a set of possible bonding patterns with proper size will be derived.

The ultimate goal of this study is to build a theoretical MS/MS spectral library adapted to any metabolite structures. For that purpose, physicochemical computation and assignment of bonding patterns are utilized to mechanistically predict metabolite fragmentation in MS/MS measurement. To cope with the vast metabolite space, proper approximation of metabolite structures into bonding patterns needs to be obtained. The ongoing comparison between cleavage activation energy of similar bonding patterns must be continued to elucidate that.

Bibliography

[1] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, and M. Arita, "Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software," *Analytical Chemistry*, vol. 88, pp. 7946–7958, 2016.

[2] H. Tsugawa, K. Ikeda, W. Tanaka, Y. Senoo, M. Arita, and M. Arita, "Comprehensive identification of sphingolipid species by *in silico* retention time and tandem mass spectral library," *Journal of Cheminformatics*, vol. 9, p. 19, 2017.

[3] S. C. Gates and C. C. Sweeley, "Quantitative metabolic profiling based on gas chromatography," *Clinical Chemistry*, vol. 24, pp. 1663–1673, 1978.

[4] L. Pauling, A. B. Robinson, R. Teranishi, and P. Cary, "Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, pp. 2374–2376, 1971.

[5] K. Matsumoto, D. Partridge, A. Robinson, L. Pauling, and R. Flath, "The identification of volatile compounds in human urine," *Journal of Chromatography*, vol. 85, pp. 31–34, 1973.

[6] B. A. Knights, M. Legendre, J. L. Laseter, and J. S. Storer, "Use of high-resolution open tubular glass capillary columns to separate acidic metabolites in urine," *Clinical Chemistry*, vol. 21, pp. 888–891, 1975.

[7] S. C. Gates, C. C. Sweeley, W. Krivit, D. DeWitt, and B. E. Blaisdell, "Automated metabolic profiling of organic acids in human urine. II. analysis of urine samples

from "healthy" adults, sick children, and children with neuroblastoma." *Clinical Chemistry*, vol. 24, pp. 1680–1689, 1978.

[8] M. Phillips, G. A. Erickson, M. Sabas, J. P. Smith, and J. Greenberg, "Volatile organic compounds in the breath of patients with schizophrenia." *Journal of Clinical Pathology*, vol. 48, pp. 466–469, 1995.

[9] D. Poli, P. Carbognani, M. Corradi, M. Goldoni, O. Acampa, B. Balbi, L. Bianchi, M. Rusca, and A. Mutti, "Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study." *Respiratory Research*, vol. 6, p. 71, 2005.

[10] S. Ogawa, K. Hattori, D. Sasayama, Y. Yokota, R. Matsumura, J. Matsuo, M. Ota, H. Hori, T. Teraishi, S. Yoshida, T. Noda, Y. Ohashi, H. Sato, T. Higuchi, N. Motohashi, and H. Kunugi, "Reduced cerebrospinal fluid ethanolamine concentration in major depressive disorder." *Scientific Reports*, vol. 5, p. 7796, 2015.

[11] O. Fiehn, "Metabolomics–the link between genotypes and phenotypes." *Plant Molecular Biology*, vol. 48, pp. 155–171, 2002.

[12] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey, and L. Willmitzer, "Metabolite profiling for plant functional genomics," *Nature Biotechnology*, vol. 18, pp. 1157–1161, 2000.

[13] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," *Trends in Biotechnology*, vol. 16, pp. 373–378, 1998.

[14] S. G. Oliver, "Functional genomics: lessons from yeast," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 357, pp. 17–23, 2002.

[15] L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver, "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations," *Nature Biotechnology*, vol. 19, pp. 45–50, 2001.

[16] H. Tweeddale, L. Notley-McRobb, and T. Ferenci, "Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis." *Journal of bacteriology*, vol. 180, pp. 5109–16, 1998.

[17] R. M. Silverstein, F. X. Webstar, and D. J. Kiemle, *Spectrometric Identification of Organic Compounds*, 7th ed. John Wiley & Sons, Inc., 2005, ch. 3.

[18] R. M. Silverstein, F. X. Webstar, and D. J. Kiemle, *Spectrometric Identification of Organic Compounds*, 7th ed. John Wiley & Sons, Inc., 2005, ch. 1.

[19] J. K. Nicholson and J. C. Lindon, "Systems biology: Metabonomics," *Nature*, vol. 455, pp. 1054–1056, 2008.

[20] J. C. Lindon, E. Holmes, and J. K. Nicholson, "Peer reviewed: So what's the deal with metabonomics?" *Analytical Chemistry*, vol. 75, pp. 384 A–391 A, 2003.

[21] H. M. Rosenstock, M. B. Wallenstein, a. L. Wahrhaftig, and H. Eyring, "Absolute rate theory for isolated systems and the mass spectra of polyatomic molecules," *Proceedings of the National Academy of Sciences*, vol. 38, pp. 667–678, 1952.

[22] J. Cautereels, M. Claeys, D. Geldof, and F. Blockhuys, "Quantum chemical mass spectrometry: *ab initio* prediction of electron ionization mass spectra and identification of new fragmentation pathways," *Journal of Mass Spectrometry*, vol. 51, pp. 602–614, 2016.

[23] S. Grimme, "Towards first principles calculation of electron impact mass spectra of molecules," *Angewandte Chemie International Edition*, vol. 52, pp. 6306–6312, 2013.

[24] J. H. Gross, *Mass Spectrometry*. Springer-Verlag Berlin Heidelberg, 2004, ch. 2.

[25] X. Yang, P. Neta, and S. E. Stein, "Quality control for building libraries from electrospray ionization tandem mass spectra," *Analytical Chemistry*, vol. 86, pp. 6393–6400, 2014.

[26] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda,

Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka, "MassBank: a public repository for sharing mass spectral data for life sciences," *Journal of Mass Spectrometry*, vol. 45, pp. 703–714, 2010.

[27] C. A. Smith, G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, and G. Siuzdak, "METLIN: a metabolite mass spectral database." *Therapeutic Drug Monitoring*, vol. 27, pp. 747–751, 2005.

[28] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, and N. Bandeira, "Sharing and community curation of mass spectrometry data with global natural products social molecular networking," *Nature Biotechnology*, vol. 34, pp. 828–837, 2016.

[29] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. MacInnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel, and L. Querengesser, "HMDB: the human metabolome database," *Nucleic Acids Research*, vol. 35, pp. D521–D526, 2007.

[30] K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker, "Searching molecular structure databases with tandem mass spectra using CSI:FingerID," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 12 580–12 585, 2015.

[31] H. Shen, K. Dührkop, S. Böcker, and J. Rousu, "Metabolite identification through multiple kernel learning on fragmentation trees," *Bioinformatics*, vol. 30, pp. i157–i164, 2014.

[32] F. Allen, A. Pon, M. Wilson, R. Greiner, and D. Wishart, "CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra," *Nucleic Acids Research*, vol. 42, pp. W94–W99, 2014.

[33] F. Allen, R. Greiner, and D. Wishart, "Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification," *Metabolomics*, vol. 11, pp. 98–110, 2015.

[34] T. Kind, K.-H. Liu, D. Y. Lee, B. DeFelice, J. K. Meissen, and O. Fiehn, "LipidBlast *in silico* tandem mass spectrometry database for lipid identification," *Nature Methods*, vol. 10, pp. 755–758, 2013.

[35] S. Böcker and F. Rasche, "Towards *de novo* identification of metabolites by analyzing tandem mass spectra," *Bioinformatics*, vol. 24, pp. i49–i55, 2008.

[36] A. Vaniya and O. Fiehn, "Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics," *TrAC Trends in Analytical Chemistry*, vol. 69, pp. 52–61, 2015.

[37] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita, and O. Fiehn, "Identification of small molecules using accurate mass MS/MS search," *Mass Spectrometry Reviews*, 2017.

[38] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita, "MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis," *Nature Methods*, vol. 12, pp. 523–526, 2015.

[39] M. Karni and A. Mandelbaum, "The 'even-electron rule'," *Organic Mass Spectrometry*, vol. 15, pp. 53–64, 1980.

[40] H. Nakata, "A rule to account for mass shifts in fragmentations of even-electron organic ions in mass spectrometry." *Journal of the Mass Spectrometry Society of Japan*, vol. 50, pp. 173–188, 2002.

[41] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. John Wiley & Sons, Inc., 2007, ch. 3.

[42] J. J. P. Stewart, *MOPAC2012 Version 15.314W*, Stewart Computational Chemistry. [Online]. Available: http://openmopac.net

[43] M. J. S. Dewar and W. Thiel, "Ground states of molecules. 38. the MNDO method. approximations and parameters," *Journal of the American Chemical Society*, vol. 99, pp. 4899–4907, 1977.

[44] J. J. P. Stewart, "Optimization of parameters for semiempirical methods I. method," *Journal of Computational Chemistry*, vol. 10, pp. 209–220, 1989.

[45] J. J. P. Stewart, "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters," *Journal of Molecular Modeling*, vol. 19, pp. 1–32, 2013.

[46] J. R. Schmidt and W. F. Polik, *WebMO Version 15.0.003*, WebMO LLC, accessed in June 2015. [Online]. Available: http://www.webmo.net

[47] T. Kind, Y. Okazaki, K. Saito, and O. Fiehn, "LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries." *Analytical chemistry*, vol. 86, pp. 11 024–11 027, 2014.

[48] Y. Ma, T. Kind, A. Vaniya, I. Gennity, J. F. Fahrmann, and O. Fiehn, "An *in silico* MS/MS library for automatic annotation of novel FAHFA lipids," *Journal of Cheminformatics*, vol. 7, p. 53, 2015.

[49] Y. Masukawa, H. Narita, E. Shimizu, N. Kondo, Y. Sugai, T. Oba, R. Homma, J. Ishikawa, Y. Takagi, T. Kitahara, Y. Takema, and K. Kita, "Characterization of overall ceramide species in human *stratum corneum*." *Journal of lipid research*, vol. 49, pp. 1466–1476, 2008.

[50] K. C. Madison, D. C. Swartzendruber, P. W. Wertz, and D. T. Downing, "Sphingolipid metabolism in organotypic mouse keratinocyte cultures," *Journal of Investigative Dermatology*, vol. 95, pp. 657–664, 1990.

[51] F.-F. Hsu and J. Turk, "Characterization of ceramides by low energy collisional-activated dissociation tandem mass spectrometry with negative-ion electrospray ionization," *Journal of the American Society for Mass Spectrometry*, vol. 13, pp. 558–570, 2002.

[52] C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, and S. Neumann, "MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation," *Journal of Cheminformatics*, vol. 8, p. 3, 2016.

[53] R. Wang, Y. Fu, and L. Lai, "A new atom-additive method for calculating partition coefficients," *Journal of Chemical Information and Computer Sciences*, vol. 37, pp. 615–621, 1997.

[54] C. W. Yap, "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints," *Journal of Computational Chemistry*, vol. 32, pp. 1466–1474, 2011.

[55] J. J. P. Stewart, *MOPAC2016 Version 16.175L*, Stewart Computational Chemistry. [Online]. Available: http://openmopac.net

[56] N. Akimoto, T. Ara, D. Nakajima, K. Suda, C. Ikeda, S. Takahashi, R. Muneto, M. Yamada, H. Suzuki, D. Shibata, and N. Sakurai, "FlavonoidSearch: A system for comprehensive flavonoid annotation by mass spectrometry," *Scientific Reports*, vol. 7, p. 1243, 2017.

[57] D. P. Demarque, A. E. M. Crotti, R. Vessecchi, J. L. C. Lopes, and N. P. Lopes, "Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products," *Natural Product Reports*, vol. 33, pp. 432–455, 2016.

[58] Z. Yang, R. Nakabayashi, Y. Okazaki, T. Mori, S. Takamatsu, S. Kitanaka, J. Kikuchi, and K. Saito, "Toward better annotation in plant metabolomics: isolation and structure elucidation of 36 specialized metabolites from *Oryza sativa* (rice) by using MS/MS and nmr analyses," *Metabolomics*, vol. 10, pp. 543–555, 2014.

[59] K. Levsen, H.-M. Schiebel, J. K. Terlouw, K. J. Jobst, M. Elend, A. Preiß, H. Thiele, and A. Ingendoh, "Even-electron ions: a systematic study of the neutral species lost in the dissociation of quasi-molecular ions," *Journal of Mass Spectrometry*, vol. 42, pp. 1024–1044, 2007.

[60] *Software to optimize reaction paths along the user's expected ones*, HPC Systems Inc., written in Japanese. [Online]. Available: http://www.hpc.co.jp/chem/react1.html

[61] 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. M. Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, 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 E.09*, Gaussian, Inc., Wallingford, CT, U.S.

[62] J. A. Montgomery, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. VI. use of density functional geometries and frequencies," *The Journal of Chemical Physics*, vol. 110, pp. 2822–2827, 1999.

[63] A. Karton, "A computational chemist's guide to accurate thermochemistry for organic molecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 6, pp. 292–310, 2016.

[64] J.-W. Song, T. Tsuneda, T. Sato, and K. Hirao, "An examination of density functional theories on isomerization energy calculations of organic molecules," *Theoretical Chemistry Accounts*, vol. 130, pp. 851–857, 2011.

[65] B. Paizs and S. Suhai, "Fragmentation pathways of protonated peptides," *Mass Spectrometry Reviews*, vol. 24, pp. 508–548, 2005.