
MSU Graduate Theses


Spring 2024

Assessing Interatomic Potentials for Molecular Dynamics Simulation of Soybean Oil Pyrolysis

Tanner Garrett Rust
Missouri State University, Rust001@live.missouristate.edu

As with any intellectual project, the content and views expressed in this thesis may be considered objectionable by some readers. However, this student-scholar's work has been judged to have academic value by the student's thesis committee members trained in the discipline. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

Follow this and additional works at: <https://bearworks.missouristate.edu/theses>

 Part of the [Computational Chemistry Commons](#), [Organic Chemistry Commons](#), and the [Physical Chemistry Commons](#)

Recommended Citation

Rust, Tanner Garrett, "Assessing Interatomic Potentials for Molecular Dynamics Simulation of Soybean Oil Pyrolysis" (2024). *MSU Graduate Theses*. 3954.
<https://bearworks.missouristate.edu/theses/3954>

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.

For more information, please contact bearworks@missouristate.edu.

**ASSESSING INTERATOMIC POTENTIALS FOR MOLECULAR
DYNAMICS SIMULATION OF SOYBEAN OIL PYROLYSIS**

A Master's Thesis

Presented to

The Graduate College of
Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree
Master of Science, Chemistry

By

Tanner Rust

May 2024

ASSESSING INTERATOMIC POTENTIALS FOR MOLECULAR DYNAMICS SIMULATION OF SOYBEAN OIL PYROLYSIS

Chemistry and Biochemistry

Missouri State University, May 2024

Master of Science

Tanner Rust

ABSTRACT

The world today relies on hydrocarbon combustion for many reasons, including its high energy density that provides ease of transportation. However, hydrocarbons sourced from fossil fuels are not expected to last forever. Biodiesel, a renewable alternative, has many attractive benefits but comes with other downsides. Biodiesel can gel in cold environments and may leave residue in an engine. Pyrolysis of biodiesel has shown promise in addressing these common detriments. Inducing pyrolysis on biodiesel feedstock (commonly soybean oil in the USA) would be an attractive option presuming it continues to produce fossil fuel analogs similar to biodiesel pyrolysis. Herein, Langevin molecular dynamics were employed to simulate the pyrolysis of 6400 soybean oil-based triglycerides (SOBTs). One hundred runs containing 64 triglycerides each were performed at 2000K for 10 picoseconds with 1 femtosecond timesteps. ANI-2x, a machine-learned interatomic potential, was used as the energy calculator. Bond breaking and forming events in each run were observed and analyzed. The results matched expectations from bond dissociation energy (BDE) values for oleic and linoleic acids (those with BDE data available).

KEYWORDS: renewable fuel, pyrolysis, thermal cracking, triglycerides, soybean oil, molecular dynamics, Langevin dynamics, machine learning, interatomic potential, computational chemistry, python

**ASSESSING INTERATOMIC POTENTIALS FOR MOLECULAR
DYNAMICS SIMULATION OF SOYBEAN OIL PYROLYSIS**

By

Tanner Rust

A Master's Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Chemistry

May 2024

Approved:

Matthew R. Siebert, Ph.D., Thesis Committee Chair

Gautam Bhattacharyya, Ph.D., Committee Member

Gary A. Meints, Ph.D., Committee Member

Ridwan Sakidja, Ph.D., Committee Member

Julie Masterson, Ph.D., Dean of the Graduate College

In the interest of academic freedom and the principle of free speech, approval of this thesis indicates the format is acceptable and meets the academic criteria for the discipline as determined by the faculty that constitute the thesis committee. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

ACKNOWLEDGEMENTS

I would like to acknowledge all the faculty and students that I know in the Missouri State University Department of Chemistry and Biochemistry. I am extremely fortunate to have found myself in a group of people who nurture success and allow me to feel at home.

I would like to especially thank Dr. Siebert, Dr. Bhattacharyya, and Dr. Meints (in no order). These three professors have provided me with excellent mentorship on my journey through the chemistry program.

TABLE OF CONTENTS

Chapter 1. Background of Fuels	1
1.1 Fossil Fuels	1
1.2 Electric Vehicles	3
1.3 Biodiesel	4
1.4 Biodiesel Properties	6
1.5 Biodiesel Production	8
1.6 Pyrolysis	9
Chapter 2. Background of Computation	13
2.1 Molecular Dynamics	13
2.2 Interatomic Potentials in Classical Molecular Dynamics	16
2.3 ANI Potential	18
2.4 ReaxFF	20
2.5 Python	20
Purpose	21
Chapter 3. Methods	22
3.1 Creation of Samples	22
3.2 Running Dynamics and Extracting Data	23
3.3 Bond Dissociation Energy Calculations	23
Chapter 4. Results and Discussion	24
4.1 The Variable Identity of the Triglyceride	24
4.2 The Indexing of Atoms	27
4.3 The Identification of Statistically Significant Events	31
4.4 The Assessment of Fatty Acid Bond Breakage and Formations in Nanoreactor Simulations	31
4.5 Bond Dissociation Energy Analysis	67
Chapter 5. Conclusion	74
References	76
Appendix: Scatter Plots of Initial Breakage Count versus BDE	80

LIST OF TABLES

Table 1. Percentages of fatty acid residues in soybean oil and canola oil.	27
Table 2. Order of presentation of fatty acid residues.	35
Table 3. Trendline equations and correlation coefficients for initial breakage counts versus BDE plot.	65
Table 4. Bond dissociation energies for homolytic C-H bond cleavage of common alkanes.	68
Table 5. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes forming vinylic radicals.	69
Table 6. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes forming vinylic radicals.	69
Table 7. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes forming vinylic radicals.	70
Table 8. Bond dissociation energies for homolytic C-C bond cleavage of common alkanes.	70
Table 9. Bond dissociation energies for homolytic C-C bond cleavage of common alkenes.	71
Table 10. Bond dissociation energies for homolytic bond cleavage of other relevant compounds.	71
Table 11. Mean unsigned deviation for each potential.	72
Table 12. Mean signed deviation for each potential.	72
Table 13. Root-mean-squared deviation for each potential.	73
Table 14. Percent error values for each potential.	73

LIST OF FIGURES

Figure 1.1 Fossil fuel reserves for groups of countries and European countries.	1
Figure 1.2 U.S. Daily Petroleum Usage in thousand barrels per day.	2
Figure 1.3 Gasoline spot price at New York Harbor normalized with 1986 CPI.	3
Figure 1.4 Structural representations of linoleic acid, glycerol, and trilinolein.	5
Figure 1.5 Pie chart representing the main feedstock sources for biodiesel production worldwide by weight percent in 2007.	6
Figure 1.6 Change in particulate matter and NO _x emissions for soy methyl ester as well as rapeseed (canola) methyl/ethyl ester blends	8
Figure 1.7 Simplified representation of fractions produced by fractional distillation of crude oil.	11
Figure 1.8 Primary reactions involved in the general mechanism of triglyceride cracking.	12
Figure 2.1 Representative graph of Equation 4 with tunable parameters equal to 1.	17
Figure 2.2 Potential energy surface for Water.	18
Figure 3.1 Demonstration of SMILES code being converted into three-dimensional triglycerides.	22
Figure 4.1 Structural representations of the glycerol backbone and the fatty acid residues.	25
Figure 4.2 Three-dimensional view of a numbered triglyceride.	28
Figure 4.3 The order of the atomic index assignments for atoms in a triglyceride.	29
Figure 4.4 General atomic indices of heavy atoms in linoleic acid.	30

Figure 4.5 Example of the usage of directed graphs (digraphs).	34
Figure 4.6 Two-dimensional bond breakage and formation diagrams for palmitic acid.	36
Figure 4.7 Digraph for palmitic acid.	37
Figure 4.8 Pathway 1, 1 for palmitic acid.	38
Figure 4.9 Pathway 1, 2 for palmitic acid.	38
Figure 4.10 Pathway 1, 3 for palmitic acid.	39
Figure 4.11 Pathway 2, 1, 1 and 4, 1, 1 for palmitic acid.	39
Figure 4.12 Pathway 3, 1, 1 for palmitic acid.	40
Figure 4.13 Pathway 5, 1 for palmitic acid.	40
Figure 4.14 Two-dimensional bond breakage and formation diagrams for stearic acid.	41
Figure 4.15 Digraph for stearic acid.	42
Figure 4.16 Pathway 1, 1 for stearic acid.	43
Figure 4.17 Initial breakages in oleic acid versus oleic acid BDEs for heavy-heavy bonds in a bar chart.	44
Figure 4.18 Initial breakages in oleic acid versus oleic acid BDEs for hydrogen-heavy bonds in a bar chart.	45
Figure 4.19 Two-dimensional bond breakage and formation diagrams for oleic acid.	46
Figure 4.20 Digraph for oleic acid.	47
Figure 4.21 Pathway 1, 1 for oleic acid.	48
Figure 4.22 Pathway 1, 2 for oleic acid.	48
Figure 4.23 Pathway 1, 3 for oleic acid.	49

Figure 4.24 Pathway 2, 1, 1 for oleic acid.	49
Figure 4.25 Pathway 3, 1, 1 for oleic acid.	50
Figure 4.26 Pathway 4, 1, 1 for oleic acid.	51
Figure 4.27 Initial breakages in linoleic acid versus linoleic acid BDEs for heavy-heavy bonds in a bar chart.	52
Figure 4.28 Initial breakages in linoleic acid versus linoleic acid BDEs for hydrogen-heavy bonds in a bar chart.	53
Figure 4.29 Two-dimensional bond breakage and formation diagrams for linoleic acid.	54
Figure 4.30 Digraph for linoleic acid.	55
Figure 4.31 Pathway 1, 1 for linoleic acid.	56
Figure 4.32 Pathway 1, 2 for linoleic acid.	56
Figure 4.33 Pathway 2, 1, 1 and 3, 1, 1 for linoleic acid.	57
Figure 4.34 Pathway 4, 1 for linoleic acid.	58
Figure 4.35 Pathway 5, 1, 1 for linoleic acid.	59
Figure 4.36 Pathway 6, 1 for linoleic acid.	60
Figure 4.37 Two-dimensional bond breakage and formation diagrams for linolenic acid.	61
Figure 4.38 Digraph for linolenic acid.	62
Figure 4.39 Pathway 1, 1 for linolenic acid.	63
Figure 4.40 Pathway 2, 1 for linolenic acid.	63
Figure 4.41 Flowchart demonstrating the data pipeline for molecular dynamics of triglycerides.	66

BACKGROUND OF FUELS

1.1 Fossil Fuels

Fossil fuels are currently the main energy source for most countries and are expected to continue to providing a majority of energy through 2030.^{1,2} The dependency on fossil fuel can be attributed to the attractive high energy density (less oxidized than other fuel sources such as wood) and large initial supply. The issue lies in that fossil fuels are effectively nonrenewable, and many developed countries do not have long-term fossil fuel reserves, which consist of oil, fuel, and natural gas (Figure 1.1).¹ Moving away from fossil fuels may prove difficult, as the widespread usage implies large infrastructure put in place to use fossil fuels.

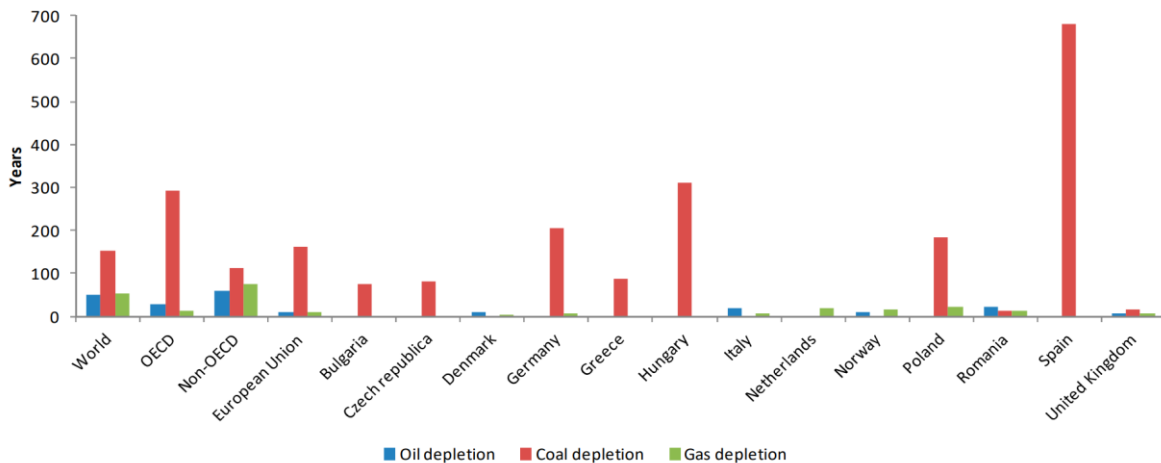


Figure 1.1 Fossil fuel reserves for groups of countries and European countries. Each bar represents the number of years until full depletion of the respective resource at the usage rate in 2019 of each country. The Organisation for Economic Co-operation and Development (OECD) is made up of thirty-eight countries and includes the United States. Reproduced with permission under Creative Common CC BY license.¹

While supplies diminish, the usage of fossil fuels continues to rise at an expected rate of approximately 1.1% per year.² Petroleum usage in both the transportation sector and industrial

sector in the United States have followed this trend, with the only notable instances counter to the trend being the 2008 financial crisis and the 2020 pandemic (Figure 1.2).³

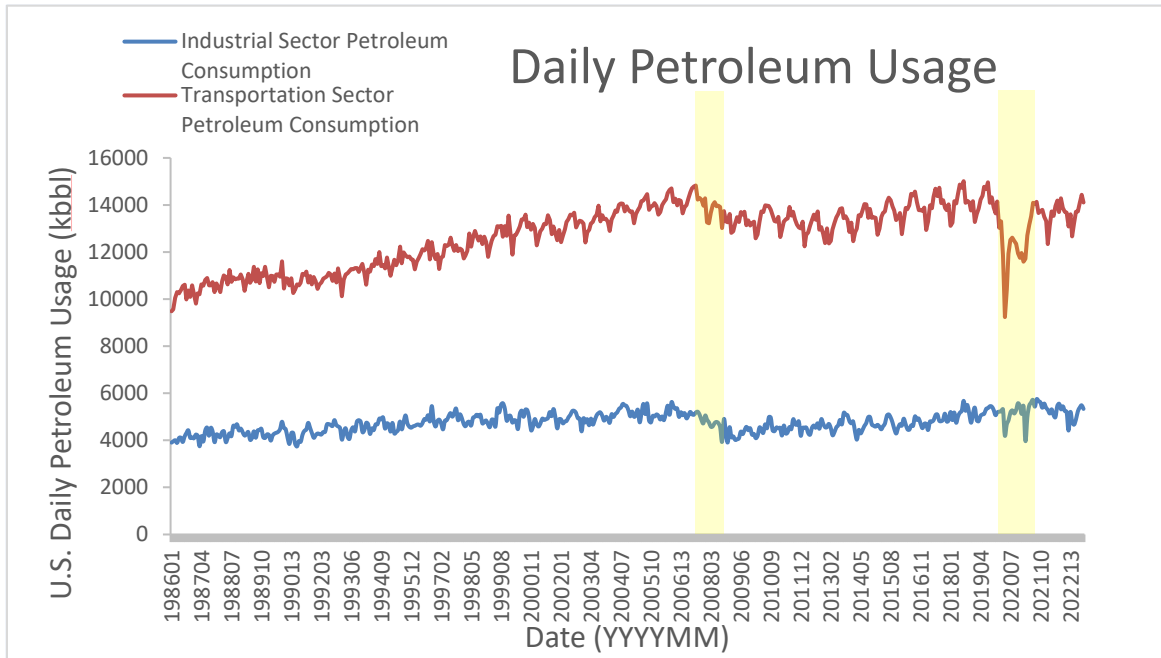


Figure 1.2 U.S. Daily Petroleum Usage in thousand barrels per day. Notable periods of time can be seen in highlighted areas. The left highlight represents the 2007-2008 financial crisis. The right highlight represents the beginning of the COVID-19 pandemic. Data from U.S. Energy Information Administration.³

Following supply and demand, fossil-fuel derived commodities (such as gasoline) have shown a trend of increasing in price (Figure 1.3).³ The large fluctuations in the price seen in Figure 1.3 are consistent with the notion that fossil fuels are subject to fluctuation much more than their renewable counterparts.² The price has been normalized for inflation since 1986.⁴

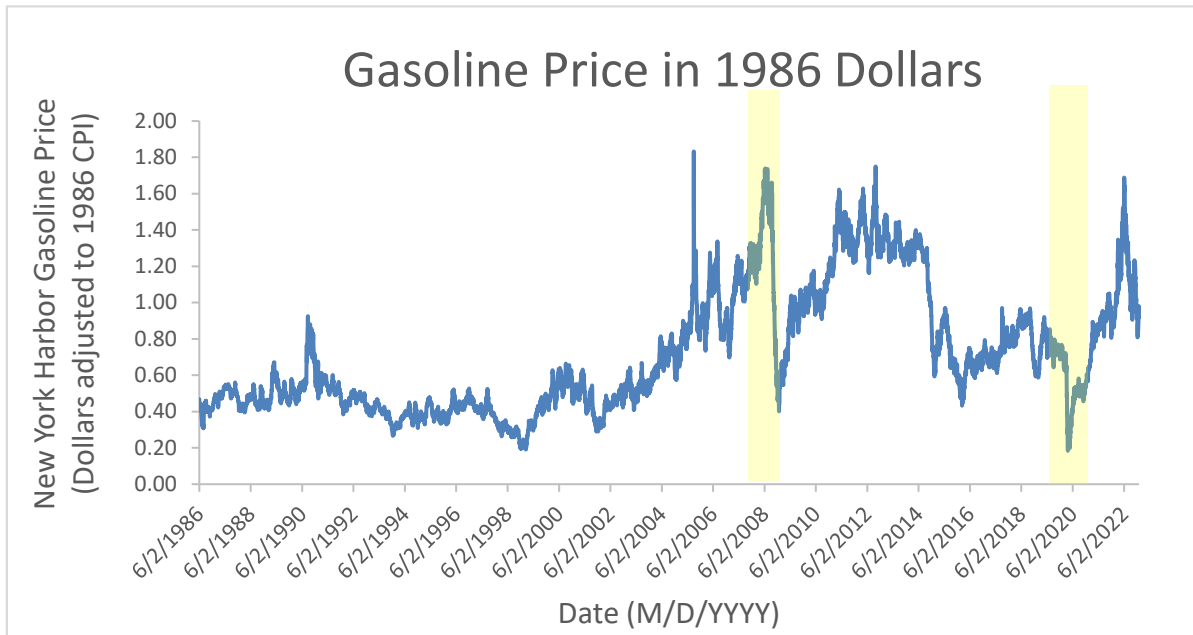


Figure 1.3. Gasoline spot price at New York Harbor, normalized with 1986 CPI. Notable periods of time can be seen in highlighted areas. The left highlight represents the 2007-2008 financial crisis. The right highlight represents the beginning of the COVID-19 pandemic. Price data from U.S. Energy Information Administration and CPI data from U.S. Bureau of Labor Statistics.^{3,4}

1.2 Electric Vehicles

In pursuit of an alternative for petroleum fueled transport, the sale of electric vehicles has risen dramatically since the turn of the 20th century.⁵ As electric vehicles have become more of a viable option, many countries have set goals to increase their electric auto fleet through tax incentives.⁵ Electric vehicles do not depend on petroleum for transportation, but rather the source of power for the electrical grid.^{6,7} Approximately 22% of total electricity generation in the United States in 2022 was comprised of renewable resources.⁸ The majority of electricity produced in the United States in 2021 originated from the usage of coal and natural gas, making up shares of 21.9% and 38.4% of the total electricity produced, respectively.⁹ It is estimated that the widespread adoption of electrical vehicles and the

electrification of buildings would lead to a 20%-38% increase in total electricity consumption by 2050.⁷

Similar to conventional vehicles with oil, electric vehicles encounter another supply chain issue when it comes to batteries.^{6,10} The raw materials needed for the production of batteries for electric vehicles are limited to certain geographical regions, leading to similar trade policy issues that fossil fuels face.^{10,11} Following the current growth pattern, it is estimated that by 2040, the raw material requirements for the production of electric vehicles will require up to 8 times the production rate of raw materials in 2023.¹²

1.3 Biodiesel

Biodiesel, by definition, is a fatty acid alkyl ester (FAAE) that can be used in place of conventional diesel (mixture of aliphatic and aromatic hydrocarbons with between 10 and 18 C atoms) in engines.¹³⁻¹⁶ This alternative to conventional diesel, or just “diesel”, is commonly produced by transesterification of a fatty acid (FA) ester or triglyceride (TG, shown in Figure 1.4) derived from a renewable vegetable oil or animal fat (biomass).^{14,17,18} The transesterification of vegetable oils is not a new process, and was originally performed in the mid-1800s.¹⁹ The production of biodiesel has shown great success in commercial applications despite having some incompatibilities with diesel engines and a change in emissions.^{15,20-22}

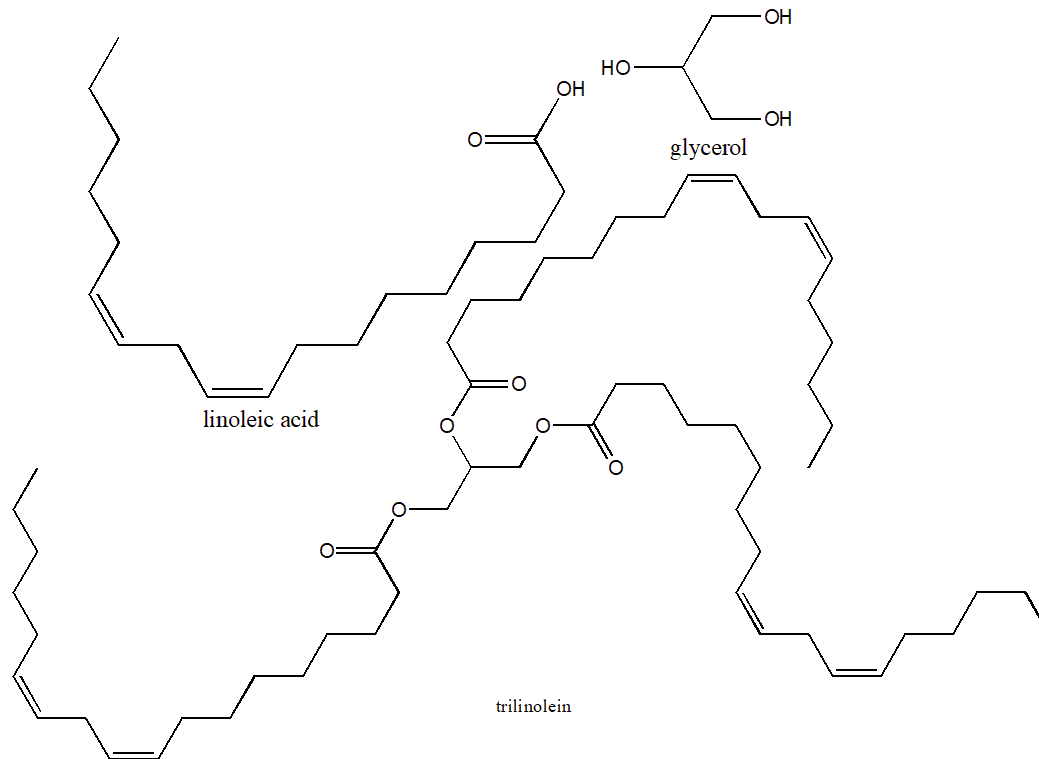


Figure 1.4 Structural representations of linoleic acid, glycerol, and trilinolein (a triglyceride containing 3 fatty acid residues of linoleic acid).

The transesterification process involves a reaction of a TG with a short-chain monohydric (one hydroxyl group) alcohol in the presence of a catalyst, ideally a base. The first transesterification reaction yields a diacylglycerol or diglyceride (DG) and a FFAE.¹⁴ Another monohydric alcohol will ideally react with a DG to produce a FFAE and a monoacylglycerol or monoglyceride (MG).¹⁴ Ideally, a similar reaction will then take place, with the final products identified as three FFAEs and one glycerol.¹⁴

The TGs used in the production of biodiesel can be sourced from a variety of readily available, renewable oils.^{14,23} They are an attractive alternative to crude oil as a source of hydrocarbons for fuel, due to their abundance in nature.²³ Worldwide in 2007, rapeseed

(canola) oil was the predominant feedstock in biodiesel production, making up 48% of the production with 4.6 million metric tons of oil.¹⁴ Rapeseed oil was followed by soybean oil at 22%, and palm at 11% with lesser used oils making up the remaining 19% (Figure 1.5).¹⁴

Biodiesel Production Feedstock Sources by Weight Percent

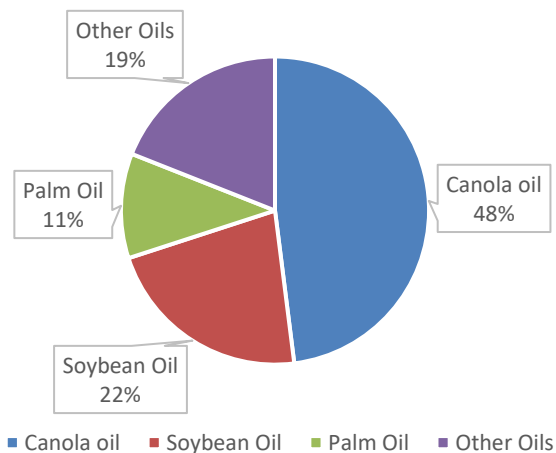


Figure 1.5. Pie chart representing the main feedstock sources for biodiesel production worldwide by weight percent in 2007.¹⁴

Though rapeseed is the most used by over double worldwide, it should be noted that soybean oil is the most used vegetable oil in the United States.^{14,17} Though vegetable oils are the most common source of TGs, animal fats and microorganisms are also a source of TGs.²³

1.4 Biodiesel Properties

Though biodiesel has shown success in fueling vehicles in the same manner as conventional diesel, they are not chemically identical. Biodiesel containing fuels (biodiesel and conventional diesel blend) are known to have higher lubricity than their conventional counterpart, and can form deposits.¹⁵ As well as deposits, fuel filters can be clogged or

degraded, which scales directly with the glycerol content of the fuel (glycerol is a byproduct of triglyceride transesterification that is not completely removed).¹⁵ Though literature is sparse regarding the long term effects of biodiesel on automotive parts, some studies have shown a higher failure rate in biodiesel-exposed auto parts compared to solely conventional diesel.^{21,24} Fuel injector and fuel pump failure, as well as increased deposits on fuel injectors and fuel pumps, have been shown to be the largest contributors to this notion.²¹

In addition to biodiesel's effect on engines, biodiesel also poses changes in emissions from combustion.^{15,22} Nitric oxides, denoted as NO_x , have been shown to be produced in greater quantities from biodiesel-containing blends as opposed to conventional diesel.^{15,22,25} However, the NO_x production has been shown to be both feedstock-dependent and engine-dependent and not greater than conventional diesel in every case.^{15,25} Particulate matter (PM) is also a metric for emission comparison, with biodiesel nearly undisputedly showing lower particulate emission as well as clearer smoke opacity.^{15,22,25} The differences between the emissions of conventional diesel and biodiesel become more apparent with an increase of oxygen in the fuel (Figure 1.6).²⁵

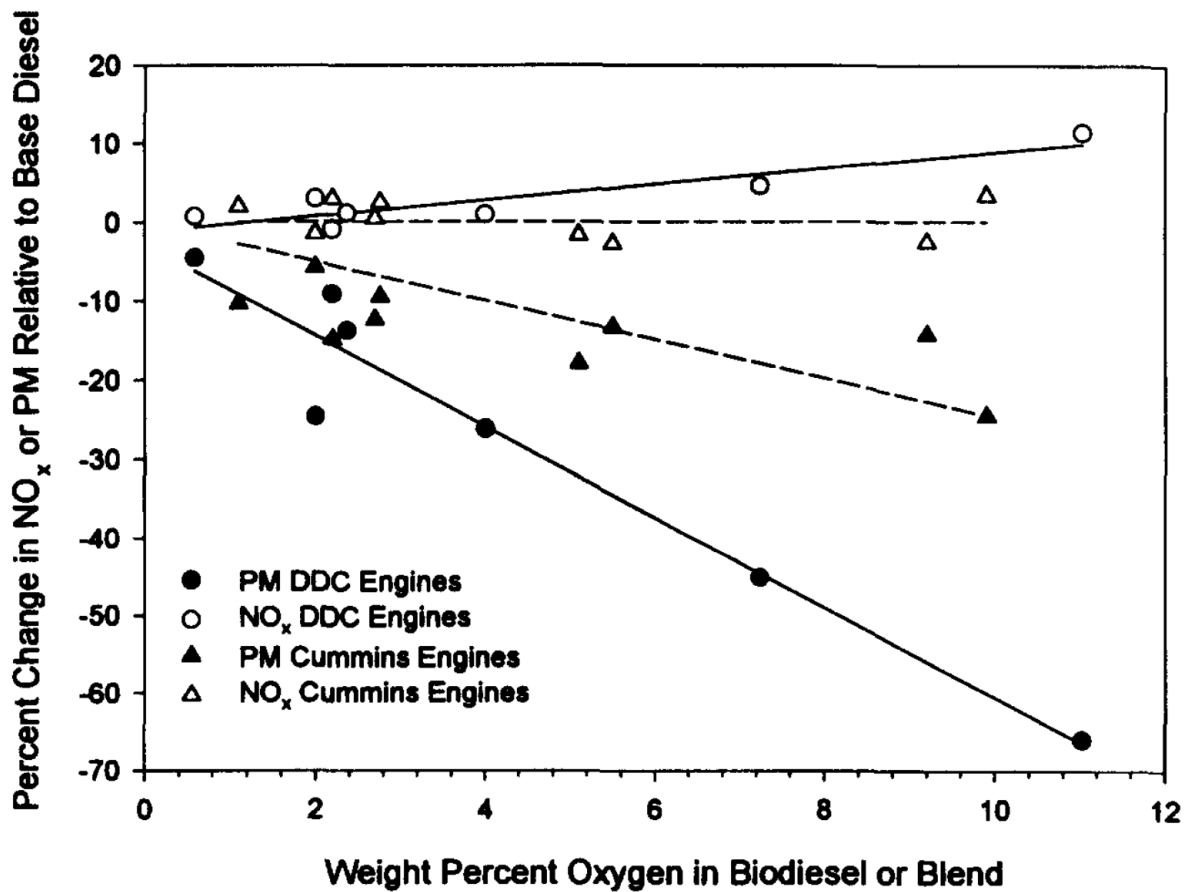


Figure 1.6 Change in PM and NOx emissions for soy methyl ester as well as rapeseed (canola) methyl and ethyl ester blends. Stock, four-stroke engines produced by manufacturers Detroit Diesel Corporation (DDC) and Cummins Incorporated are used. Reproduced with written permission from Elsevier.²⁵

1.5 Biodiesel Production

While a base catalyst is commonly used in the transesterification of high-quality feedstocks, a consideration that must be made is the quality of the feedstock used. If the feedstock is particularly low quality, perhaps with a free fatty acid (FFA) content of more than 3% by weight, base catalyzed transesterification is no longer an effective option.¹⁴ The base catalyst will react with the FFAs to form salts of fatty acids, also known as saponification

(formation of soap). This reaction would irreversibly quench the catalyst and result in a mixture of FFA, soap, TGs, DGs, MGs, biodiesel, glycerol, water, and unused monohydric alcohol.¹⁴ This mixture is unusable as a fuel as is, and any water in the mixture with base catalyst will hydrolyze fatty acid arms on TGs, DGs, and MGs to form more FFAs.^{14,26} An alternative, albeit a slower and much more costly method of addressing high FFA content, is acid pretreatment of the feedstock before transesterification with homogeneous base catalysts. This method shows consistent results of products derived from feedstock with lower FFA content.^{14,26}

When more steps are taken in production, the cost of production will inevitably increase. Biodiesel is relatively expensive to make compared to the final cost of petroleum diesel.²⁷ This creates a problem concerning incentive of production of biodiesel, considering the cost of raw materials used to create biodiesel is greater than the cost of the finished product of its conventional counterpart. Due to this, efforts have been formed to find less expensive sources of reactants for the creation of biodiesel.

1.6 Pyrolysis

In pursuit of alternative fuel at a lower cost, other methods have been explored as a source of renewable energy. An economical way of producing a renewable fuel from biomass is pyrolysis.^{18,28} Pyrolysis is a broad term that can describe high temperature thermal cracking, which leads to a wide variety of products, or lower temperature processes that can lead exclusively to the decarboxylation of fatty acids and TGs.^{28,29} It comes with the upside that it is cheap and simple to construct, and has been fairly well studied. Pyrolysis of vegetable oils is particularly beneficial in areas containing abundant hydro processing industry (hydro cracking

or hydro treating using hydrogen), as the technology required for pyrolysis is similar to that of the refining of conventional petroleum.²⁸

Applying pyrolysis to biomass can yield biofuel fractions of each phase - solid, liquid, and gas. The amount of each phase and the types of compounds contained depend strongly on the conditions of the pyrolysis (heat and duration). The products of the process would be most beneficial in the form of liquid due to ease of transportation and refining to become an economical product. It has been shown that a pyrolytic process of less than a few seconds at a temperature of approximately 500°C is optimal to maximize the yield of liquid phase products from dry feedstock.^{28,30} These processes can be defined as “fast” pyrolysis and have been shown to yield up to around 80% liquid product by weight. This can be compared to a lower temperature pyrolysis. Under catalytic treatment at 40 bar and 300°C with activated carbon supported palladium, stearic acid and the three-armed TG of stearate (found in vegetable oils) can be decarboxylated.²⁹ N-heptadecane and their corresponding byproducts can be produced in near 100% yield.²⁹ When vegetable oils are subjected to pyrolysis, they are converted to hydrocarbon-rich fuels.²⁸ Using these oils as a feedstock comes with other benefits including easy transportation (liquid), similar heat content to diesel, neutral pH, and little-to-no water content.²⁸

The pathways regarding pyrolysis of TGs are a current area of research. The cleavage of carbon-carbon bonds in pyrolysis can create lower molecular weight (MW) organic products commonly used as fuel, such as gasoline.²³ These lighter products can be more representative of the lighter fractions of crude oil fractional distillation, compared to the long chain fatty acid methyl esters that are produced by simple transesterification.^{23,31} This is not a new aspect, as it

is used in the production of gasoline (naphtha).³² Some of the heavier fractions of crude oil are pyrolyzed (converted) to maximize the gasoline fraction.³² Figure 1.7 shows a simplified representation of the fractional distillation process of crude oil.

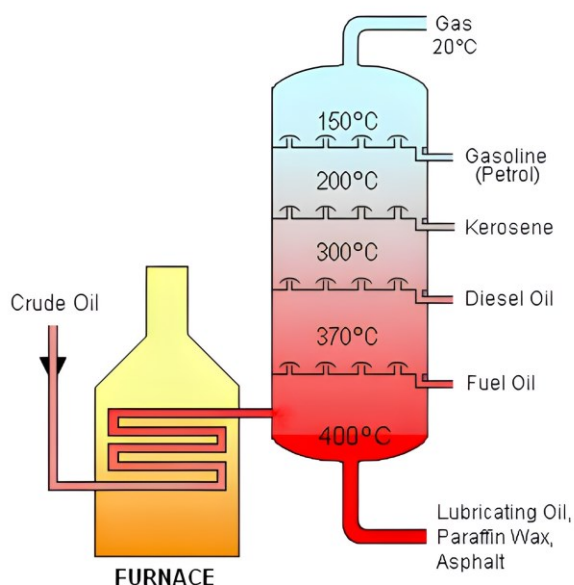


Figure 1.7 Simplified representation of fractions produced by fractional distillation of crude oil. Reproduced with permission under Creative Commons Attribution-Share Alike 3.0 Unported license. Upscaled 400%.³³

Pyrolysis in TGs is either initialized by the breakage of the carbonyl carbon to ester oxygen bond, or the ester oxygen to glycerol carbon bond.²³ In the first case, the glycerol has a possibility of forming high MW products such as coke and tar due to a formation of an unstable intermediate such as acrolein that can begin a polymerization reaction.²³ Specific reaction paths are more well known of the pyrolysis processes that use catalysts.²³ Proposed pathways for non-catalyzed pyrolysis of TGs can be seen in Figure 1.8.

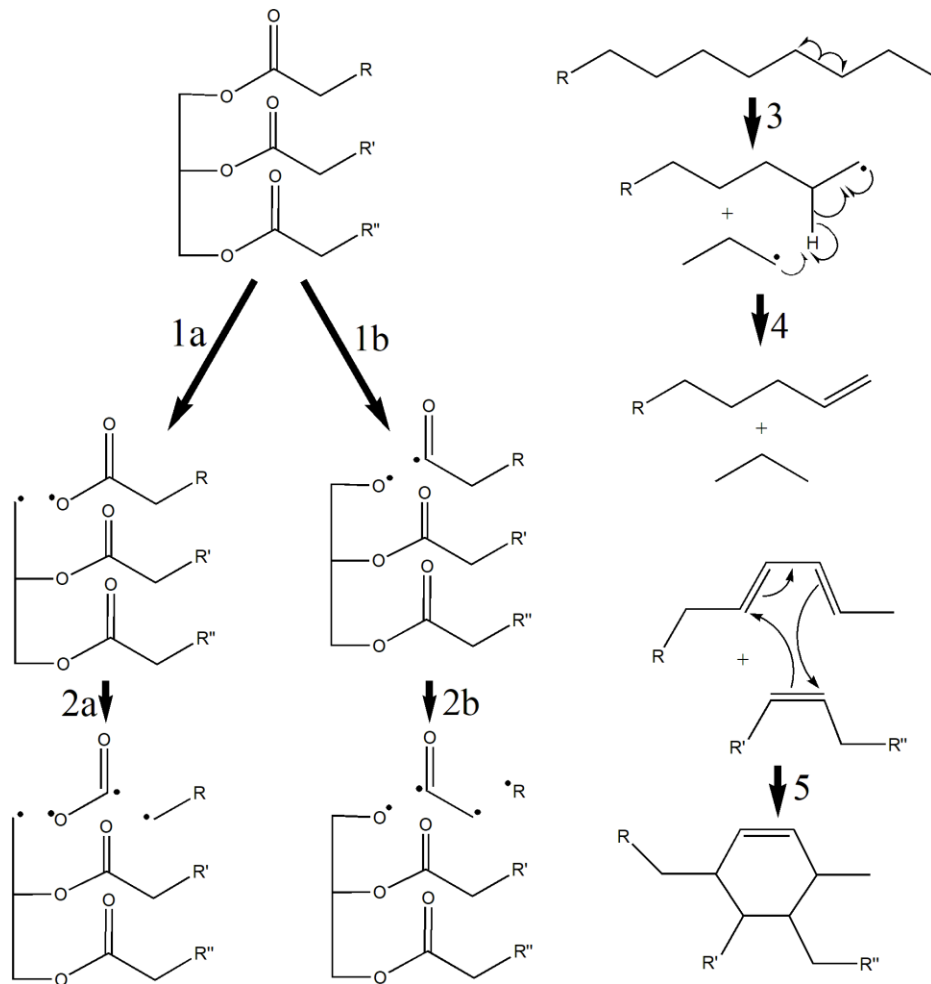


Figure 1.8 Primary reactions involved in the general mechanism of TG cracking.²³ Reactions 1 and 2 represent the expected initial pathway of TG cracking.²³ Reactions 3 and 4 represent the common mechanism for cracking of the alkyl chain, followed by radical quenching through hydrogen radical formation and exchange.²³ Reaction 5 represents a Diels-Alder reaction that may form cyclized products in the cracking process.²³

Non-catalyzed pyrolysis has been studied less; however, the studies that have been conducted show product patterns. As decarboxylation is not catalyzed, it has been shown that long chain FAs can be commonly found as a stable product. In some cases, however, a ketene is formed which would lead to a mixture of alkanes, alkenes, and aromatics in place of the long chain FA.²³ Allowing high MW FAs to accumulate as the ultimate product is not ideal, as it does

not add “fuel value” since this mixture would suffer from undesirable fuel properties (low fuel density, high freezing point, and acidity).²³ The production of large amounts of FAs is the least attractive part of non-catalyzed pyrolysis methods.²³

BACKGROUND OF COMPUTATION

2.1 Molecular Dynamics

Molecular dynamics are simulations of motion of individual molecules that are used to model solids, liquids, and gases.³⁴ In a real experiment, a given quantity is measured over a certain time interval using an instrument such as a thermometer for the property of temperature.³⁵ If the measurement is prone to statistical noise, the measurement will become more accurate the longer the measurement is averaged.³⁵ The same idea can be applied to molecular dynamics simulations, with an observable dependent on the positions and momenta of particles in a system.³⁵ Once again using temperature as an example, the instantaneous temperature for a finite system can be obtained using Equation 1.

$$T(t) = \sum_{i=1}^N \frac{m_i v_i^2(t)}{k_B N_f} \quad (1)$$

With T representing temperature, t representing time, N representing the number of atoms in the simulation, m_i representing the mass of atom i , v_i representing the velocity of atom i , k_B representing the Boltzmann constant, and N_f representing the degrees of freedom in the simulation.

Molecular dynamics simulations come in different forms. These include classical molecular dynamics (moving bodies based on classical physics equations), *ab initio* molecular dynamics (moving bodies based on quantum physics), and semi-empirical molecular dynamics (using some real-world data with a blend of other calculation methods). The most accessible

form of molecular dynamics is classical molecular dynamics as consumer grade computers can run the calculations required of a simulation.

Classical molecular dynamics (CMD) revolve around the idea that given initial conditions and forces, the output of a process can be predicted.³⁴ In other words, given the knowledge of the placement of atoms and how they interact, the whole reaction can be extrapolated. CMD is performed through the calculation of forces exerted on each atom at various intervals of the simulation, called timesteps.³⁵ Using the calculated forces for each timestep, Newton's second law is integrated, providing velocities for each atom.³⁵

Ab initio molecular dynamics (AIMD), with *ab initio* meaning 'from the beginning', differs from traditional molecular dynamics by relying on the quantum Schrödinger equation instead of Newton's second law.³⁶ This is typically more accurate yet more computationally intensive approach to simulating a molecular system.³⁶ AIMD is typically used in realistic simulation of complex systems, which can replace the need for a practical experiment that is infeasible due to either a financial or logistical reason.³⁶

An alternative method to purely *ab initio* calculations is semi-empirical (SE) methods. These types of methods implement empirical data found in relevant real-world scenarios to replace some of the computationally intensive integrals, effectively giving the calculations a head start.^{37,38} SE methods come with the implied downside of lower accuracy due to imposed approximations.^{37,38}

Due to time and computing power limitations, CMD simulations can cut computing costs compared to AIMD by using more approximate methods that are much faster.^{37,38} However, CMD comes with much less accuracy compared to AIMD. The golden standard of energy

computation would be matching the accuracy of *ab initio* methods whilst improving the time of computation.^{37,38}

An ideal implementation of machine learning could provide large amounts of accuracy while also providing a quick calculation.^{37,38} Recently, machine learning has become a lot more popular in modeling atomistic simulations due to the speed of computation being up to five orders of magnitude higher than *ab initio* calculations.^{37,38} Machine learning cannot necessarily reason like a human how a new type of calculation (i.e. different chemical environment) would be different from a calculation on it was trained.^{37,38} Ideally, the machine learning model would be trained on data that would be analogous to any calculations that it would encounter in experiments.^{37,38}

Langevin dynamics, a specific model for classical molecular dynamics, is a modification of Brownian (random) motion that additionally considers the effects of molecules surrounding a body.^{39,40} The simplest continuous form of the Langevin equation is shown in Equation 2.⁴¹

$$ma(t) = -\nabla E(x(t)) - \gamma m(v(t)) + R(t) \quad (2)$$

with m representing mass, $a(t)$ representing acceleration of the particle with respect to time, x representing the position of a particle, γ representing the friction coefficient (in reciprocal time units), $v(t)$ representing velocity with respect to time, $R(t)$ representing a zero-mean stationary gaussian process with respect to time, satisfies the following expressions in Equation 3.⁴¹

$$\langle R(t) \rangle = 0, \quad \langle R(t) \cdot R(\hat{t})^T \rangle = 2\gamma k_B T m \delta(t - \hat{t}) \quad (3)$$

With δ representing the Dirac delta, and \hat{t} representing the Fourier transform of t , T representing the temperature of the system, and k_B representing the Boltzmann constant.⁴¹

Equation 2's terms can be broken down into separate contributions to the resulting force ($ma(t)$). The first term, $\nabla E(x(t))$, is the contribution of pairwise interactions between particles. The second term, $\gamma m(v(t))$, is the contribution of friction because of the system being treated as a fluid and can be tuned with the friction coefficient γ . The third term, $R(t)$, is the contribution of random movement, which can be tuned with the temperature T .⁴¹

2.2 Interatomic Potentials in Classical Molecular Dynamics

The key to how atoms interact (attraction and repulsion) with each other is an interatomic potential, or potential energy surface. Some of the more well known, and simple interatomic potentials include the Morse potential and Lennard-Jones potential.⁴² The general form for the Morse potential, which models diatomic molecules similar to the harmonic oscillator while accounting for anharmonicity, is shown in Equation 4.^{42,43}

$$V_{morse} = \epsilon [2e^{-2\alpha r} - e^{-\alpha r}] \quad (4)$$

Equation 4 represents energy (V_{morse}) as a function of interatomic distance (r). ϵ and α correspond to tunable parameters. The Lennard-Jones potential, which models van der Waals interactions between electronically neutral atoms or molecules, is similar with the same number of tunable parameters, and can be seen in Equation 5.^{42,44}

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \quad (5)$$

These equations allow for a pairwise interaction between two atoms, which can make up the first term, $\nabla E(x(t))$, of Equation 2. The graph of Equation 4 can be seen in Figure 2.1

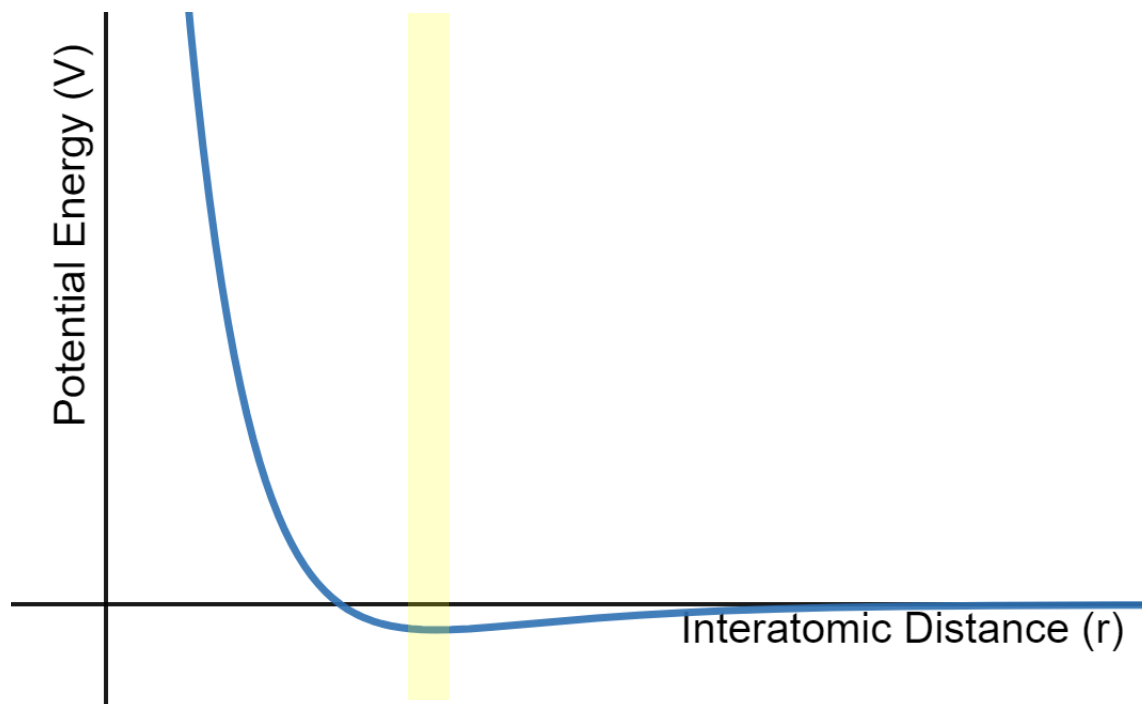


Figure 2.1 Representative graph of Equation 4 with tunable parameters equal to 1. The highlighted region of the graph represents equilibrium bond distance.

Figure 2.1 is a representative graph of the Morse potential. The graph contains three regions of interest. The first region is the global minimum, which is representative of the interatomic distance to minimize potential energy, the optimal interatomic distance in a representative molecule. This region is more commonly known as the equilibrium bond distance. The second region is the rapid increase to the left of the minimum (as r approaches 0), as this is representative of interatomic repulsion as two atoms get too close together. The third region is to the right of the minimum (as r approaches infinity), which is representative of two atoms with a large interatomic distance, and therefore having minimal interaction.

Potential energy surfaces (PESs) are effectively the greater-than-two-dimension versions of a potential energy curve/energy profile.⁴⁵ Three-dimensional PESs are commonly used to visually describe potential energy as a function of two molecular properties such as bond length

and bond angle.^{45,46} An example of this can be seen in the PES of water, shown in Figure 2.2 with potential energy on the z-axis, H-O distance on the x-axis, and the H-O-H angle on the y-axis.⁴⁶

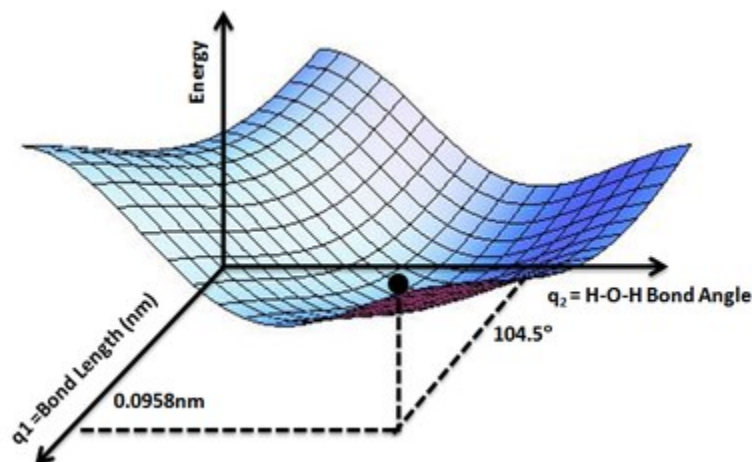


Figure 2.2 Potential energy surface for Water (H_2O). Reproduced with permission from Creative Commons Attribution-Share Alike 3.0 Unported.⁴⁶

2.3 ANI Potential

ANI, short for ANAKIN-ME: Accurate Neural network engine for Molecular Energies, is a neural-network-derived potential that aims to allow for accurate and quick molecular calculations. Previously, neural network applications to molecular dynamics simulations had the inability to allow for different numbers of bodies in the simulation as the optimized weights were designated for a fixed number of input nodes.⁴⁷ There has been interest in the creation of a transferable machine learning model, which does not require fixed parameters and can span large use cases.³⁸ The transferable model, in theory, obtains large amounts of quantum mechanical training data with the goal of being able to extrapolate to new systems from the

vast examples of systems in the training data.³⁸ The downside of this approach is that any training data that is not catered to a given case can decrease the accuracy of the model's calculations on that given case.³⁸

ANI-1 contains 57,951 molecules in the training data set, consisting of every possible molecule up to 8 atoms for carbon, nitrogen, and oxygen.³⁷ The potential energy surfaces that were used to train the model were found by Normal Mode Sampling (NMS).³⁷ NMS consists of calculating the normal mode of the molecule, and then perturbing the structure in equilibrium along the normal modes to maximum energy.³⁷ The point of this is to plot data of the various energies between the minimum and maximum values.³⁷ From this method, the data set boasts 17.2 million scenarios.³⁷

The ANI potential would use a modified version of the previous neural network models such that they can be more transferable than the originals.^{37,47} The novel method to achieve this is a modification to the original neural network in which atomic environmental vectors (AEVs) are created as molecular representations.^{37,38} These AEVs are built for each atom, taking in a radial part based on the atomic number, and an angular part based on the atomic number pairs.^{37,38} This contrasts with the original neural networks that would not differentiate with respect to atomic numbers.^{37,47}

ANI-1 shows higher accuracy in selected test cases than DFT with semi-empirical DFTB and PM6, two common methods.³⁷ The speed of these computations was extremely quick for ANI-1, with the amount of time being comparable to the time it would take to perform classical force field method calculations.³⁷

The latest release of ANI, ANI-2x, aims to expand on a common application of computational chemistry, namely drug prediction.³⁸ ANI-2x is an expansion of ANI-1x that adds in S, F, and Cl due to their prevalence in drug compounds.³⁸ As well as the addition of these elements, ANI-2x also improved on previous version's ability to better predict molecular torsion profiles, which is significantly applicable to drug development.³⁸

2.4 ReaxFF

Mentioned previously, large-scale atomistic simulations can benefit from using classical force fields and applying them as pairwise functions between atoms using an interatomic potential.^{37,38} Reactive force field (ReaxFF) uses 2 relationships between atoms to achieve this: between bond distance and bond order (number of bonding pairs of electrons between two atoms), as well as between bond order and bond energy.⁴⁸ ReaxFF utilizes this to create an interatomic potential that leads to proper dissociation of bonds to separated atoms.⁴⁸ This model shows similar accuracy to a popular semi-empirical quantum calculation method, PM3, while being about 100 times faster than PM3 and about 10,000 times faster than *ab initio* calculations.

2.5 Python

Python is a programming language that is used for general purposes, which is oriented to enable the construction of clear, logical applications to both large and small tasks.⁴⁹ Python has integrated well with computational chemistry, with libraries dedicated to the subject such as Atomic Simulation Environment (ASE).⁴⁰ ASE focuses on setting up, manipulating, running, visualizing, and analyzing chemical simulations.⁴⁰ Python is also supported by Jupyter Notebook which is a free, open-source, interactive web-browser based computational notebook which is

well-utilized by data scientists and other researchers alike.⁵⁰ The format of the notebook encourages the creation of easy-to-read code which can also function as a manual.⁵⁰ This interactive mode of working with large amounts of chemical data meshes well with pandas, a python package that is designed to allow ease of working with relational or labeled data.⁵¹

PURPOSE

This project aims to create computational infrastructure that can be used to better understand the pyrolysis reaction pathway in the case of triglycerides. The ANI potential (version ANI-2x) will be evaluated in its ability to model known pathways of pyrolysis of triglycerides. These aims will be achieved through monitoring of bonds breaking and forming in a pyrolysis simulation and comparing ANI calculated bond dissociation energies of common triglyceride pyrolysis products to experimental literature values. Optimally, the ANI potential will be proved accurate, and the simulation results will allow the identification of reaction conditions that will optimize the hydrocarbon content in the products of pyrolyzing soybean oil. If this is achieved, the next step could be subjecting the reaction conditions to experimental studies.

METHODS

3.1 Creation of Samples

Simplified molecular-input line-entry system (SMILES) was used to textually assemble TGs, as it allows molecules to be expressed in the form of plain text in a condensed form without the explicit representation of implicit hydrogens (e. g. CCCC is the SMILES representation for butane, C₄H₁₀).⁵² Python code was written to create the SMILES text for TGs and this text was, ultimately, used to create an ensemble of 6400 molecules (100 runs of 64 TGs).

For each TG string, now part of the ensemble, the python code used Open Babel, a cheminformatics conversion tool, to create three dimensional geometry files (.xyz format) of each TG from SMILES format.⁵³ The geometry files used Open Babel's fast forcefield optimization tool to generate a stable conformer.⁵³ ASE was used to read each geometry file and assemble them into a nanoreactor containing 64 TGs (Figure 3.1).⁴⁰ This procedure was repeated 100 times to produce the ensemble of 6400 individual TGs.

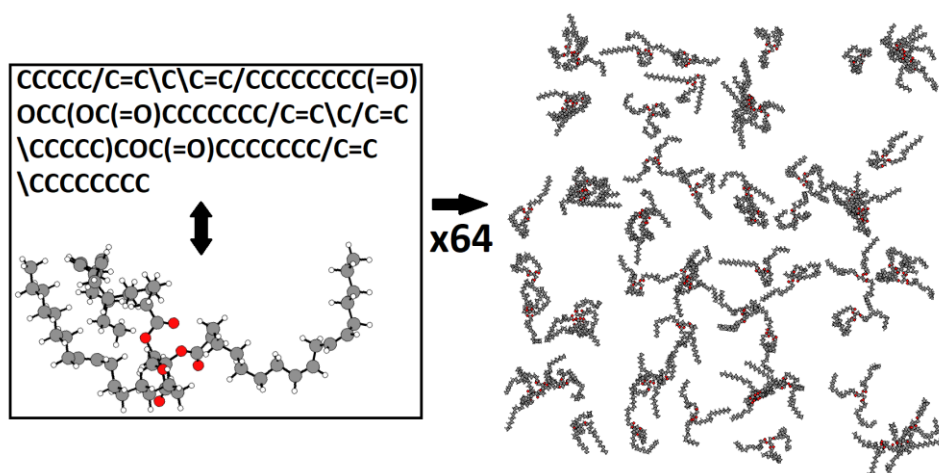


Figure 3.1 Demonstration of SMILES code being converted into three-dimensional triglycerides. Repeated for 64 triglycerides and combined into a master cell (right).

3.2 Running Dynamics and Extracting Data

The 100 master geometry files were then subjected to Langevin dynamics at 2,000K (temperature to see bond breakages or formations in roughly two-thirds of triglycerides) for 10,000 steps with a timestep of 1 femtosecond and an ASE friction coefficient of 0.019645 fs^{-1} (near the center of the typical range of friction coefficients) using ANI-2x as the potential. The script then created a trajectory file, a movie of 10,000 geometry files.

After molecular dynamics was completed for each nanoreactor (64 TGs), the script split the total trajectory into its component trajectories for each TG within. This was performed for ease of analysis. Each frame of every TG trajectory was scanned by the script for changes in bonding, either a breakage or a formation of a bond. All results were compiled into a master spreadsheet containing four pieces of information for each change in bonding (event), whether it is a break or formation, the two atoms involved, and the timestep at which the change took place.

3.3 Bond Dissociation Energy Calculations

Bond dissociation energies (BDEs) were calculated four times, once per version available of the ANI potential (ANI-1x, ANI-1ccx, and ANI-2x), and once using ReaxFF as the potential. The calculated BDEs were compared with the literature values for the BDEs. The signed and unsigned differences between the potential calculated BDEs and the literature BDEs for each general category of reaction (C-H, C-C, C-X) were used to calculate the mean unsigned deviation (MUSD), mean signed deviation (MSD), and the root-mean-squared deviation (RMSD).

RESULTS AND DISCUSSION

4.1 The Variable Identity of the Triglyceride

Oils and fats are complex mixtures of many TGs, each composed of a glycerol backbone and three fatty acids. The fatty acids can vary in identity. That is, TGs in these mixtures can be composed of various fatty acid residues and can mix or match the residues depending on the source. For example, soybean oil contains a higher proportion of saturated fatty acid residues (i.e. palmitic and stearic acid) than canola oil. The general form of a TG can be seen in Figure 4.1 along with the R, R', and R'' groups representing the fatty acid residues (shown below the glycerol backbone). Palmitic acid and stearic acid's R groups, the first fatty acid residues shown, are made up of 15 sp^3 hybridized carbons and 17 sp^3 hybridized carbons, respectively. The oleic, linoleic, and linolenic acid residues all contain 17 carbons, however, each of them contains at least one unsaturation. Oleic acid, linoleic acid, and linolenic acid residues contain 2 sp^2 hybridized, 4 sp^2 hybridized, and 6 sp^2 hybridized carbons, respectively. Below the previously mentioned residues, the fatty acid residues of cis-11-eicosenoic acid and erucic acid are shown which make up a minor component of canola oil. The cis-11-eicosenoic acid residue contains 19 carbons, with 2 of them being sp^2 hybridized. The erucic acid residue contains two more carbons than cis-11-eicosenoic acid, and the same amount of sp^2 hybridized carbons.

In a study that obtained refined soybean oil from Ag Processing Inc (AGP®), they found that their soybean oil contained, by weight, 11.5% palmitic acid chains, 4% stearic acid chains, 24.5% oleic acid chains, 53% linoleic acid chains, and 7% linolenic acid chains.²³

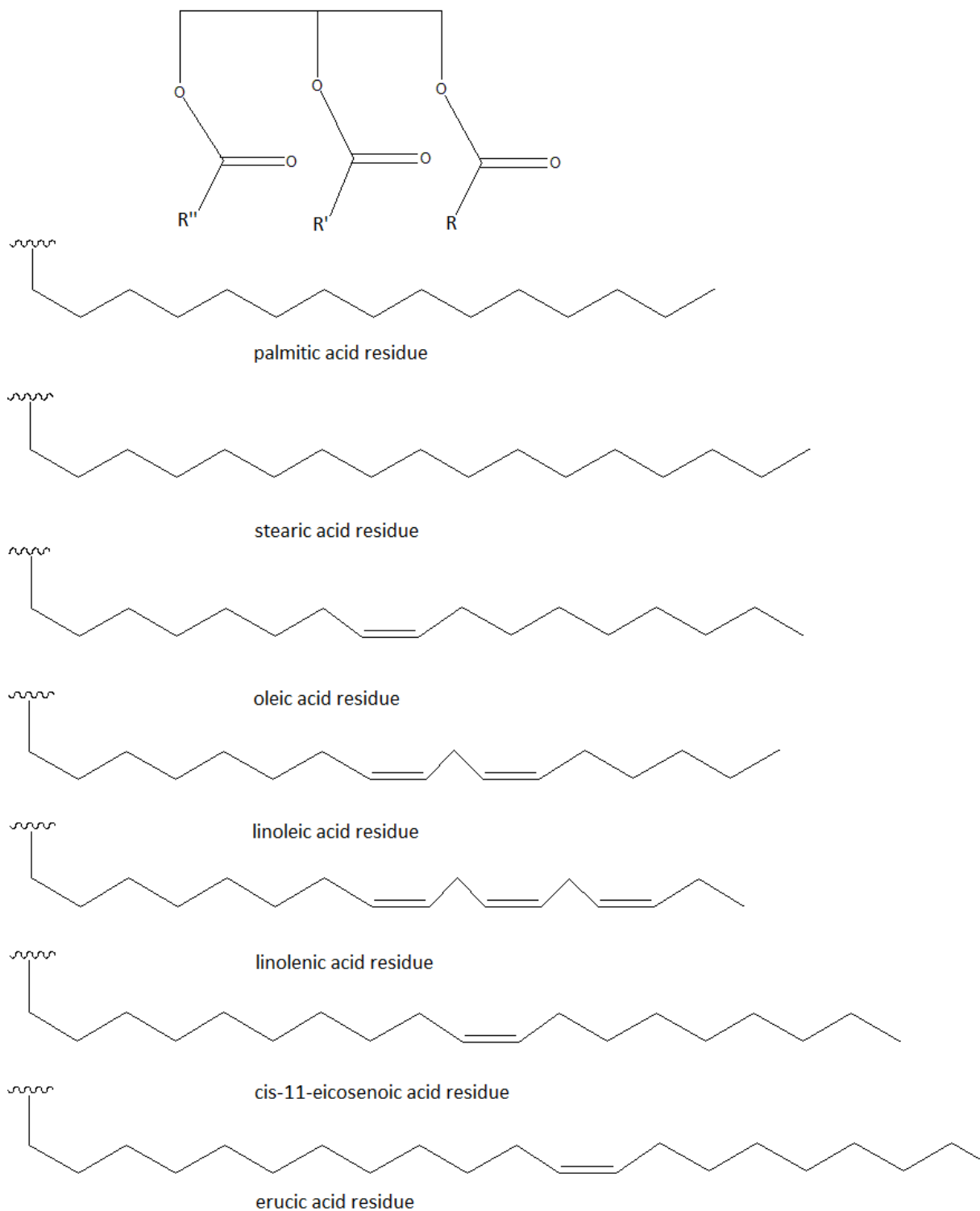


Figure 4.1 Structural representations of the glycerol backbone (top) and the fatty acid residues that can be found on the R, R', and R'' groups of the glycerol in soybean oil (top five shown fatty acid residues) and canola oil (all shown fatty acid residues).

This contrasts with the same study's data on canola oil, with the makeup by weight being 4% palmitic acid chains, 2% stearic acid chains, 60% oleic acid chains, 20% linoleic acid chains, and 10% linolenic acid chains. Additionally, canola oil contains residues of two fatty acids that are negligible or 0% in their soybean oil, cis-11-eicosenoic acid and erucic acid.²³ These fatty acid chains made up their weight of canola oil at 2% and 0.2%, respectively.²³ The weight percentages for all fatty acids in soybean oil were converted to their equivalent percentages by number, which is more applicable to nanoscale computation. The percentages for residues of palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid in soybean oil were converted to approximately 12.5%, 3.9%, 24.1%, 52.5%, and 7%, respectively.

The creation of a representative fraction of soybean oil was the first hurdle of this study. To assemble an ensemble of triglycerides to reflect the experimental makeup of soybean oil, we needed to find an approach that would be easy to automate. During this step of the research, we designated SMILES as the simplest way to connect fatty acid arms to glycerol backbones before the SMILES codes were converted into a three-dimensional file. Initially, when creating the first ensemble, 640 TGs was deemed a large enough data set to begin data analysis. We continued to add data to this ensemble, ultimately incorporating 6400 TGs. The fatty acid makeup of the 6400 TGs is shown in Table 1.

Table 1. Percentages of fatty acid residues in soybean oil and canola oil. Both weight percentages were taken from reference 22. Percentages of fatty acid residues by absolute number in soybean oil was converted from mass percent using molar mass.

Fatty Acid Residue	Fatty Acid Molar Mass (g/mol)	Referenced Soybean Oil % Weight (Approx)	Referenced Canola Oil % Weight (Approx)	Referenced Soybean Oil Converted to % Number	This Study's Sample of Soybean Oil % Number (6400 TGs)
Palmitic	256.43	11.5	4	12.5	12.0
Stearic	284.48	4	2	3.9	3.0
Oleic	282.46	24.5	60	24.1	22.7
Linoleic	280.45	53	20	52.5	56.4
Linolenic	278.43	7	10	7	5.8
Cis-11-eicosenoic	310.51	0	2	0	0
Erucic	338.57	0	0.2	0	0

4.2 The Indexing of Atoms

The Jupyter Notebook that previously analyzed each frame of the molecular dynamics to find events used atomic index to denote which atoms were involved in an event. The atomic indices were defined when the master geometry files were split into individual geometry files and loaded into ASE.⁴⁰ A numbered (or indexed) TG containing a linolenic acid residue for arm 1, a linoleic acid residue for arm 2, and an oleic acid residue for arm 3 can be seen in Figure 4.2. The red spheres refer to oxygen atoms, the dark grey spheres carbon atoms, and the light grey spheres hydrogen atoms. It should be noted that the bends in the hydrocarbon portions of the triglyceride are due to the cis saturations.

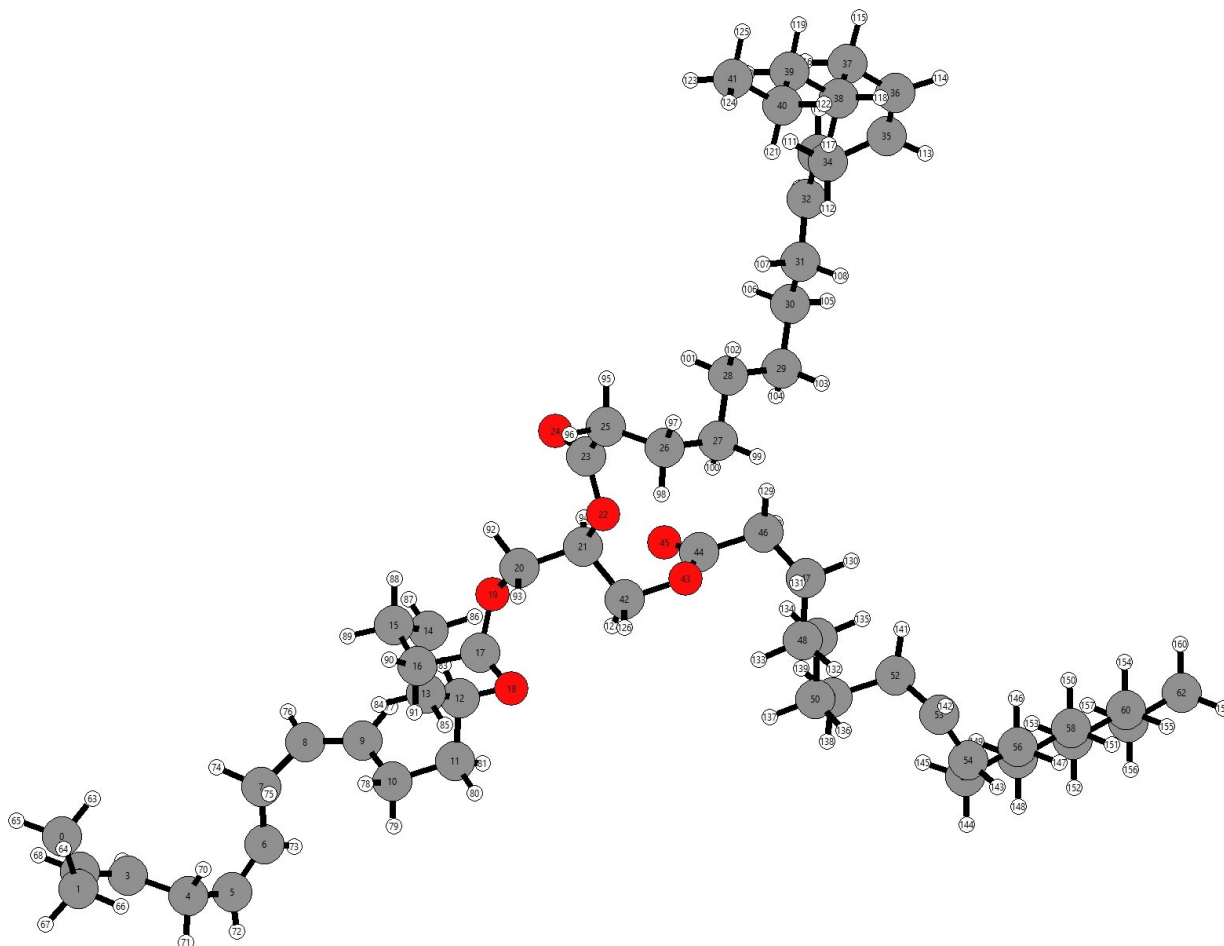


Figure 4.2 Three-dimensional view of a numbered triglyceride with a linolenic acid residue for arm 1, a linoleic acid residue for arm 2, and an oleic acid residue for arm 3. The atomic indices can be seen in the center of each atom representation. Atomic Simulation Environment's graphical user interface (ASE GUI) is used to view this molecule.⁴⁰

Since TGs possess different numbers of atoms, two atoms with the same index from two different triglycerides were not necessarily in the same position relative to the glycerol backbone. In other words, any arm identity change would lead to a domino effect in which every atomic index assignment after the change would also include a shift of index number. Thus, we sought a general list of atomic indices to compare between locations. The

metaphorical line of dominos can be seen in Figure 4.3, with the highlighted boxes representing indices that are directly assigned based on the fatty acid identity.

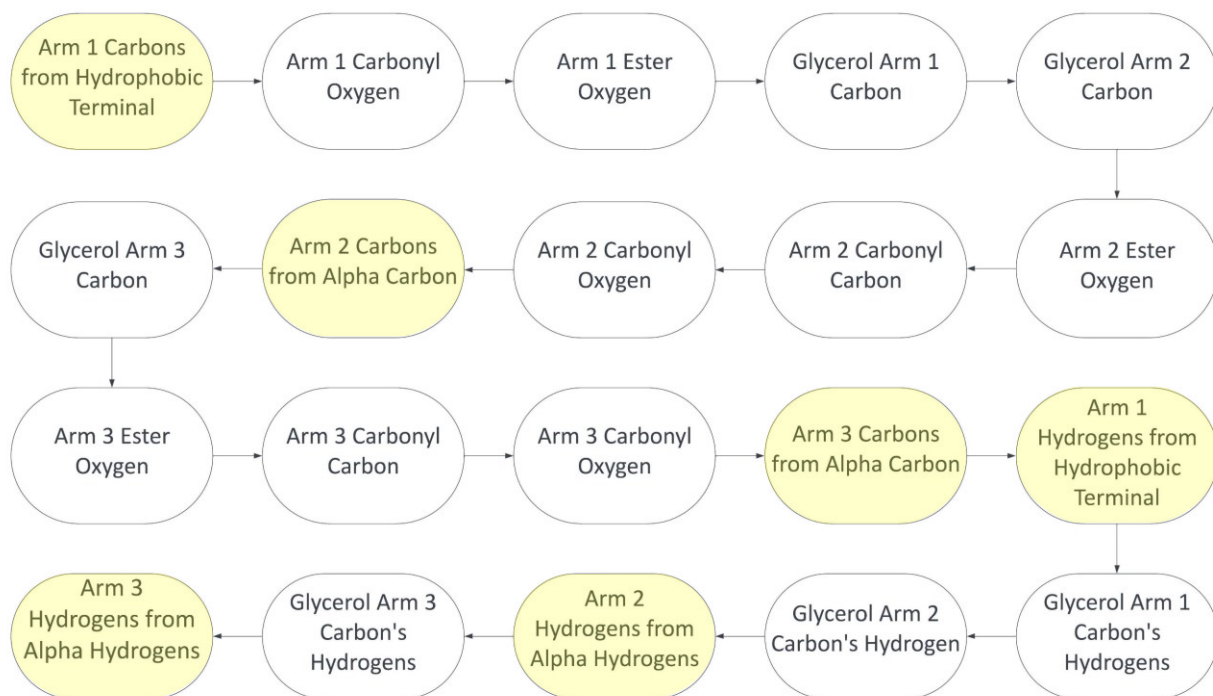


Figure 4.3 The order of the atomic index assignments for atoms in a TG with those areas that vary directly due to FA identity highlighted. Any change to the number of atoms at any point has a domino effect, shifting all numbers after.

An example of this is the atomic indices of a fatty acid residue in the arm 3 position being translated into atomic indices as if it were in the arm 1 position, would require incorporating the identity of both arm 1 and arm 2 as they would both shift the indices of arm 3 by a predictable amount. The atomic indices of heavy atoms in linoleic acid can be seen in Figure 4.4.

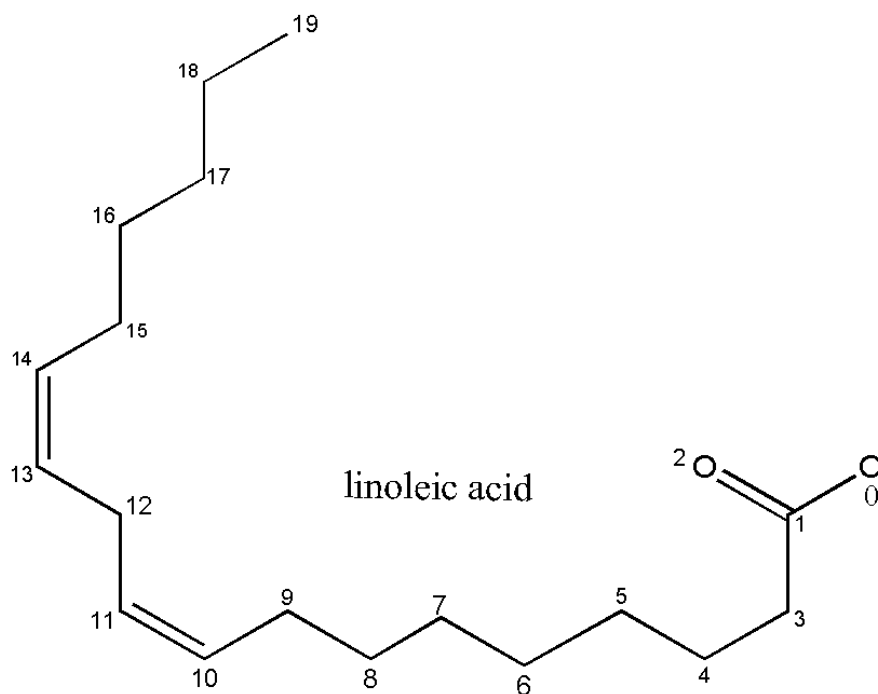


Figure 4.4 General atomic indices of heavy atoms in linoleic acid.

Allowing variability in TGs also led to issues in quantifying events that had taken place in a run. This meant that the data created by a run and analysis of a run needed extra manipulation to become interpretable data. The fatty acid residues were treated as independent of the other two arms of the triglyceride since other unpublished work by Siebert and Saquer suggested that intermolecular reaction was limited on the timescale of simulation found in this work. The implication of this is that any analysis that is compared to literature data of fatty acids alone, or a simple fatty acid ester (i.e. fatty acid methyl esters found in biodiesel) will not consider inter-arm events or interactions. Inter-arm events are not commonplace in these pyrolysis simulations but are nonzero so should be noted.

4.3 The Identification of Statistically Significant Events

After each event's atomic indices were translated into the general atomic indices form of their corresponding fatty acid residue, raw counts of events of each pairing of atomic indices were tallied. Each fatty acid residue held a tally for formation of bonds between two atoms and another tally for the breakage of bonds between two atoms. This information was consolidated into matrices using the master atomic indices. The raw counts for each bond were then converted to a proportion of an event of the specific type being between the two atoms in question ($P(i)$), using Equation 6.

$$P(i) = \frac{n_i}{n_{obs}} \cdot 100 \quad (6)$$

With n_i representing the occurrences of event i taking place, and n_{obs} representing occurrences of all events of that type (i.e. breakages in linoleic acid residues). The percentages were tested for statistical significance, by determining the variance using Equation 7.⁵⁴

$$\sigma = 1.96 \sqrt{\frac{P(i) \cdot P(\neq i)}{n_{obs}}} \quad (7)$$

With σ representing variance, $P(\neq i)$ representing the proportion of all events that are not event i , and 1.96 being the constant for the 95% confidence interval. The events with $P(i)$ greater than σ were considered statistically significant and included in further analysis.⁵⁴

4.4 The Assessment of Fatty Acid Bond Breakage and Formations in Nanoreactor Simulations

We will begin the assessment with a few different viewpoints of what took place between timestep 0 and timestep 10000. This assessment will include analysis of initial breakages (closest analog to BDE evaluation), overall breakages and formations found in the

simulations, and sequences of events that will represent pathways of reactions. The different types of data for a given fatty acid will all be combined into one section, totaling five sections.

Covalent bonds containing only carbon and/or oxygen will be referred to as heavy bonds and covalent bonds containing one carbon/oxygen and one hydrogen will be referred to as hydrogen-heavy bonds. It should be noted that every hydrogen-heavy bond that will combine counts of chemically equivalent hydrogen atoms. In other words, the two hydrogens (atom 18 and atom 19) on the alpha carbon of palmitic acid residues are combined in data as just atom 18. Bonds in these diagrams will be referred to in shorthand such that the bond between atom 1 and atom 3 is bond 1-3.

While it is equally important to assess the results of every fatty acid residue, the strongest assessment will be toward the linoleic acid and oleic acid as they occur the greatest number of times in the dataset. Using conventional wisdom when addressing bond breakages, the bonds with the highest breakage frequency, without external influences, should be those with the lowest BDEs. This can be equated to molecular dynamics most accurately by only considering breakages that occur before any other events. Bar graphs containing raw counts of initial bond breakage as well as literature calculated BDEs can be seen in the sections for oleic and linoleic acid.⁵⁵

Each fatty acid includes a two-dimensional diagram with arrows pointing to statistically significant breakages and a two-dimensional diagram for bond formations in the same format. As well as displaying the percentage of the event near the arrow pointing to the bond, the variance can also be seen as a \pm percent and was calculated at the 95% confidence interval (coefficient of 1.96) using Equation 7. The hydrogen-heavy formation diagrams specifically have

unique labels on arrows pointing to formed bonds with hydrogens, with the labels following the format “3 - x.x% ± x.x%”, which represents the hydrogen that has broken off from atom 3 and formed the bond at which the arrow points. An important aspect when reading these graphs is to consider that the statistically significant events are aggregate totals throughout molecular dynamics simulation. This is important because events are dependent on one another and can create sequences of events that would not be otherwise seen in a bond dissociation energy-oriented dataset.

The two-dimensional diagrams are followed by digraphs. Digraphs, short for directed graphs, are effectively forms of flow charts made up of nodes (events) connected by edges (arrows between events). Each node indicates an event that has happened in their respective type (breaking or forming, denoted by b or f, respectively) and the previous events. This can be interpreted as instructions for a mechanistic pathway. In the center of the digraph lies a “start” node, indicating the original structure of the fatty acid residue. Every node besides the start indicates the atomic indices of an event, with the edge between the nodes having a counter in the middle edge that indicates the number of times that edge has been traversed. Unlike bond breakage diagrams, chemically equivalent hydrogens are not combined in digraphs. An example of a digraph applied to the competition between S_N1 , S_N2 , E1 and E2 is seen in Figure 4.5. The reaction pathways can be seen at the top, and the digraph corresponding to those reaction pathways can be seen below them.

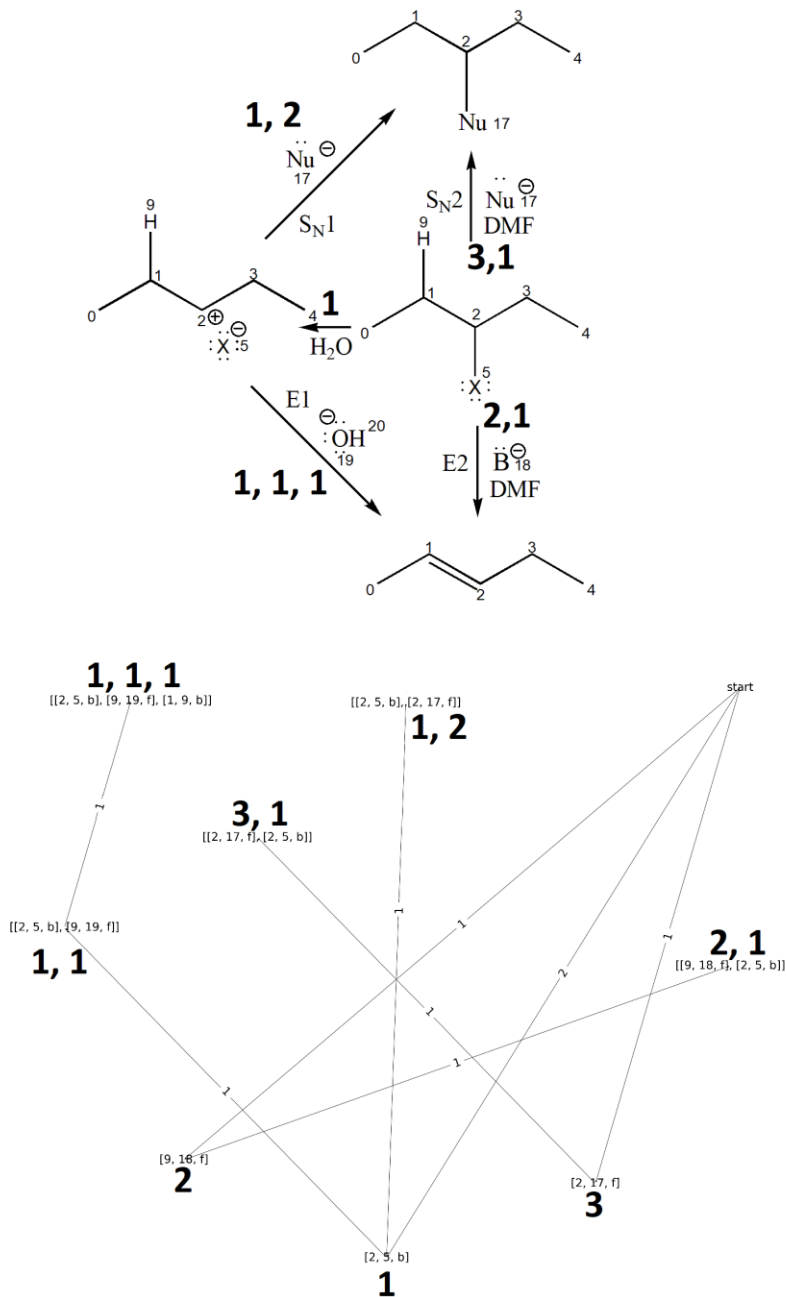


Figure 4.5 Example of the usage of directed graphs (digraphs). Four types of example reactions are included above, S_N1, S_N2, E1 and E2. The digraph for the breakage of each reaction is seen in the bottom left. The numbers in the format (1, 1, 1) refer to the specific pathway, ordered by the most common pathway. In this case, only pathway 1 is more common than any other pathway.

The digraph can be interpreted as follows: The start node in the top right of the breakage digraph is connected to three nodes, [2, 5, b], [9, 18, f], and [2, 17, f]. This indicates

the breakage of the bond between bond 2-5 (halogen leaving), the formation of the bond 9-18 (proton transfer), and the formation of the bond 2-17 (S_N2 backside attack). The nodes branching from the “start” node are connected to secondary nodes, indicating subsequent events. The notation of 1, 1 refers to the first subsequent event from the first event connected to the “start” node. An example of this is pathway 1, 1, 1 which is the breakage of 2-5 (halogen leaving, “1”), followed by the formation of 9-19 (formation of water, “1, 1”), nearly simultaneously followed by the water leaving via the breakage of bond 1-9 (“1, 1, 1”).

Each statistically significant event was then created as a node in a digraph for each fatty acid, allowing for sequences of events to describe reaction pathways. An exception for this is the digraph of linoleic acid, which required the pruning of nodes under 16 events due to the large volume of data. Nodes were only displayed for those that were a part of a reaction pathway.

The order of the fatty acids presented are ordered identically to how they were mentioned when demonstrating the percentage makeup of soybean oil, which follows an ‘increasing in complexity’ pattern, seen in Table 2.

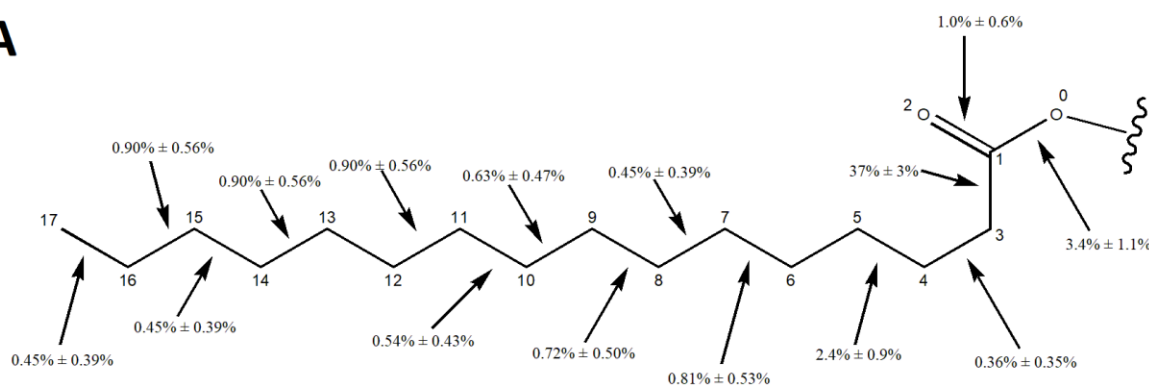
Table 2. Ordering of presentation of fatty acid residues, demonstrating the increase in complexity due to increase in carbon number and carbon-carbon double bonds.

Order of Presentation	Fatty Acid Residue	Number of Carbon Atoms	Carbon-Carbon Double Bonds
1	Palmitic	16	0
2	Stearic	18	0
3	Oleic	18	1
4	Linoleic	18	2
5	Linolenic	18	3

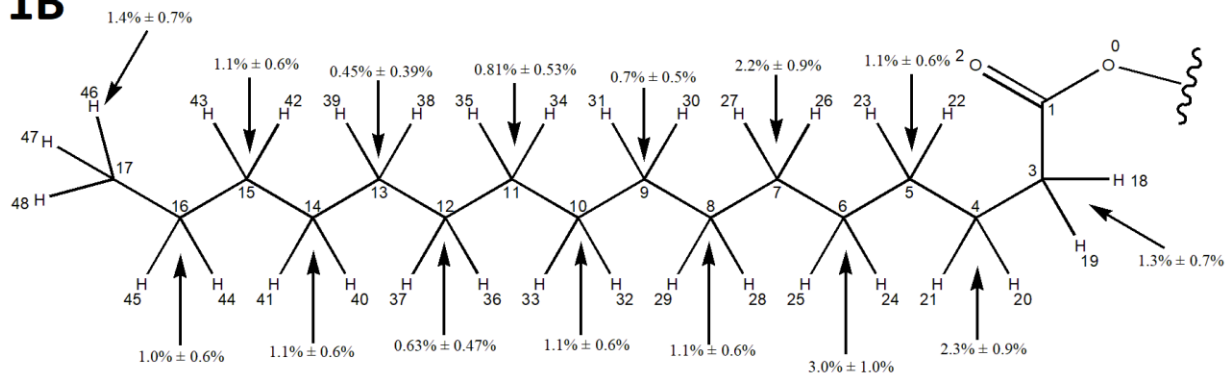
Palmitic

The bond event diagrams and digraph for palmitic acid can be seen in Figure 4.6 and Figure 4.7, respectively. The digraph mechanisms can be seen in be seen in Figure 4.8, Figure 4.9, Figure 4.10, Figure 4.11, Figure 4.12, and Figure 4.13.

1A



1B



1C

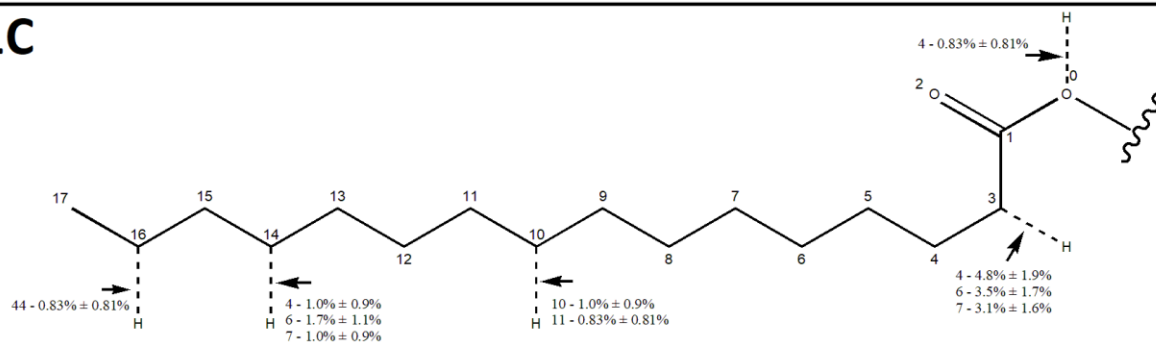


Figure 4.6 Two-dimensional bond breakage and formation diagrams for palmitic acid. Structure 1A represents the breakage of heavy-heavy bonds, structure 1B represents the breakage of hydrogen-heavy bonds, and 1C represents the formation of hydrogen-heavy bonds.

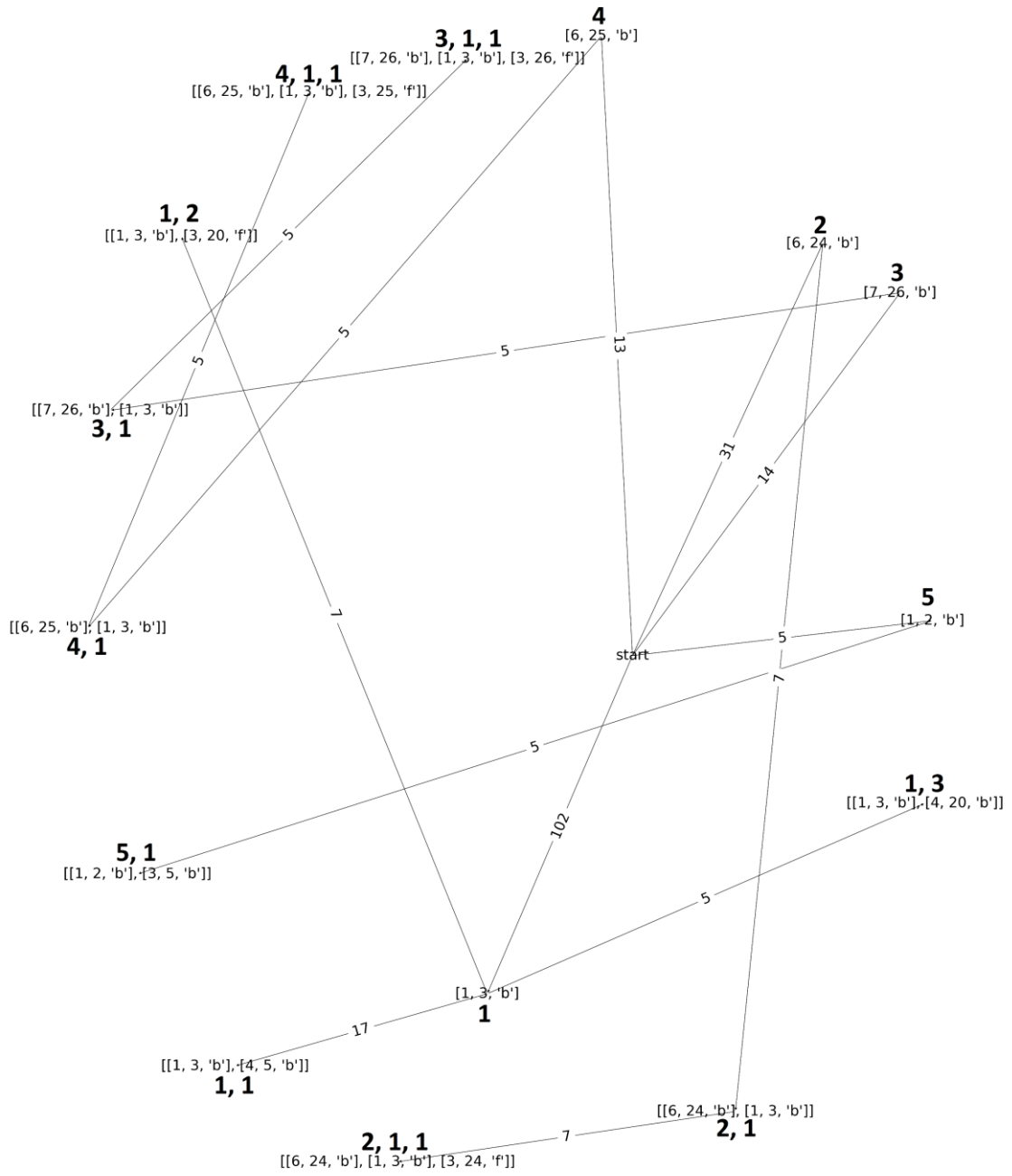


Figure 4.7 Digraph for palmitic acid.

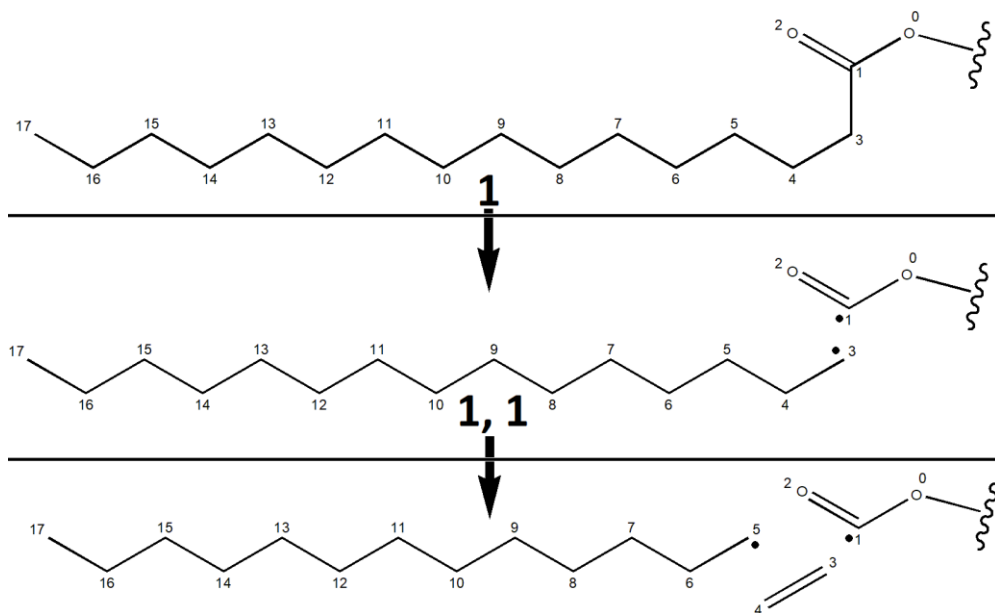


Figure 4.8 Pathway 1, 1 for palmitic acid. The breakage of 1-3 occurs followed by the breakage of 4-5.

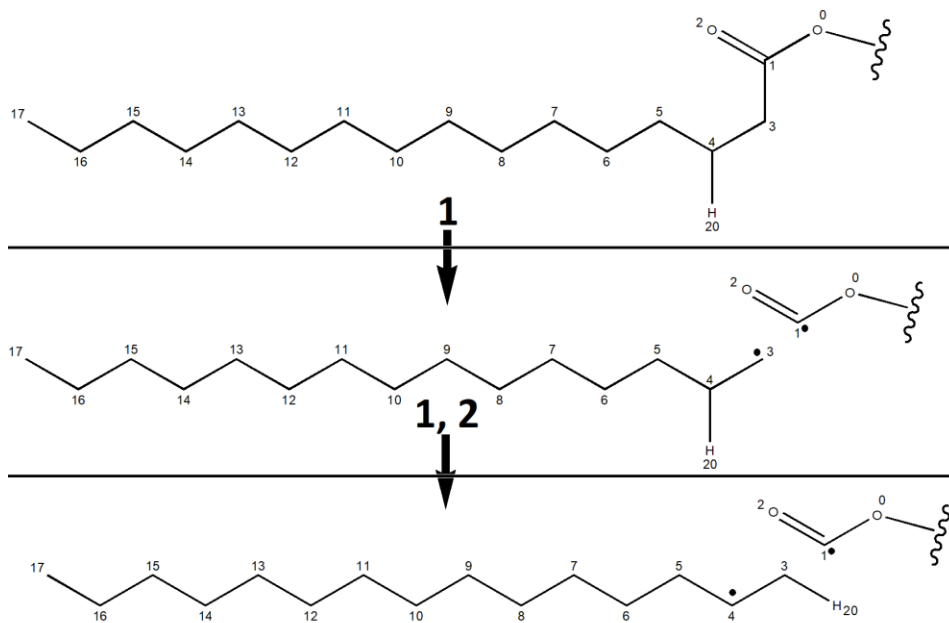


Figure 4.9 Pathway 1, 2 for palmitic acid. The breakage of 1-3 occurs followed by the formation of 3-20. The breakage of 4-20 is implied.

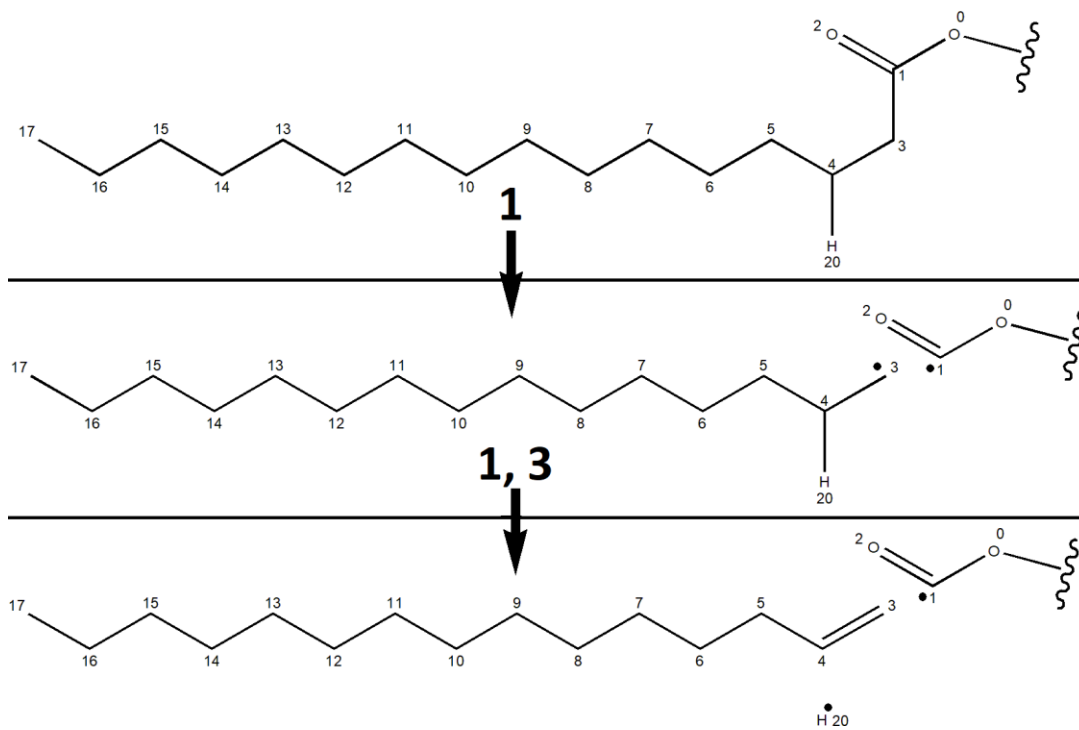


Figure 4.10 Pathway 1, 3 for palmitic acid. The breakage of 1-3 occurs followed by the breakage of 4-20.

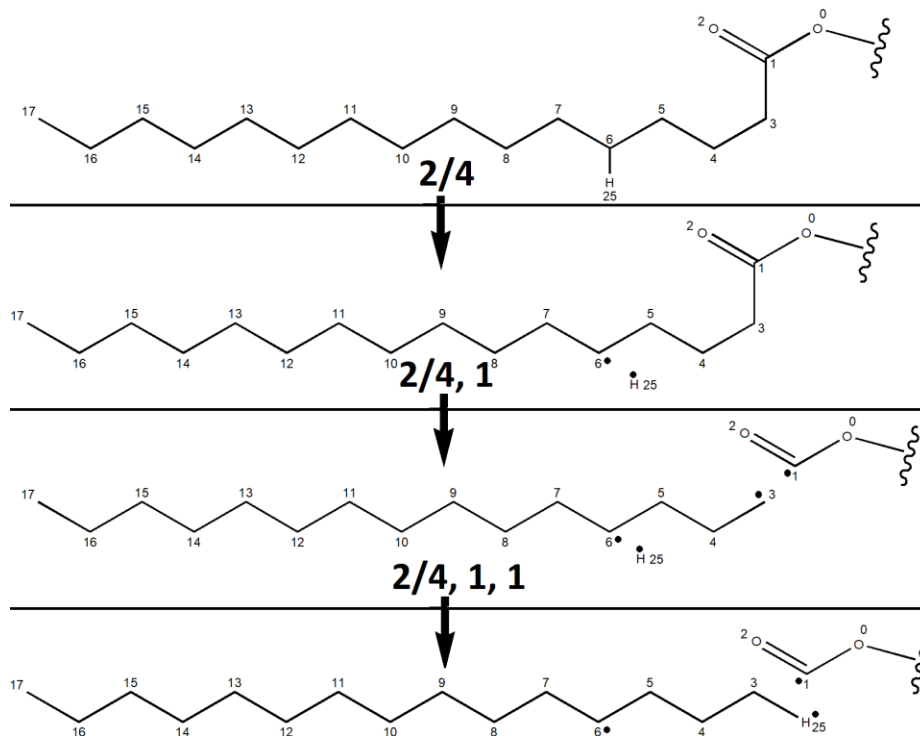


Figure 4.11 Pathway 2, 1, 1 and 4, 1, 1 for palmitic acid. The two pathways are identical, indicated by the notation 2/4, 1, 1. The breakage of 6-24/6-25 occurs followed by the breakage of 1-3 and the formation of 3-24/3-25.

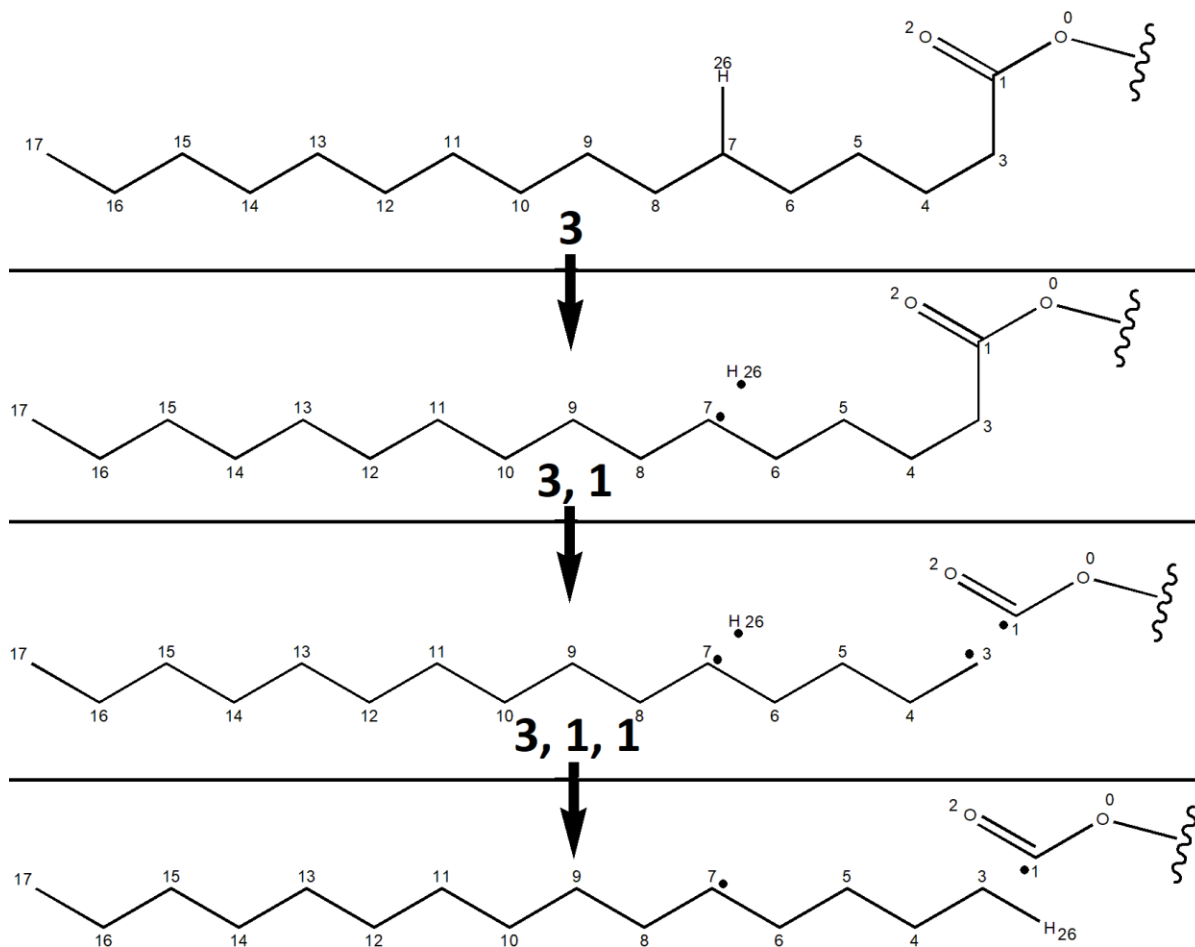


Figure 4.12 Pathway 3, 1, 1 for palmitic acid. The breakage of 7-26 occurs followed by the breakage of 1-3 and the formation of 3-26.

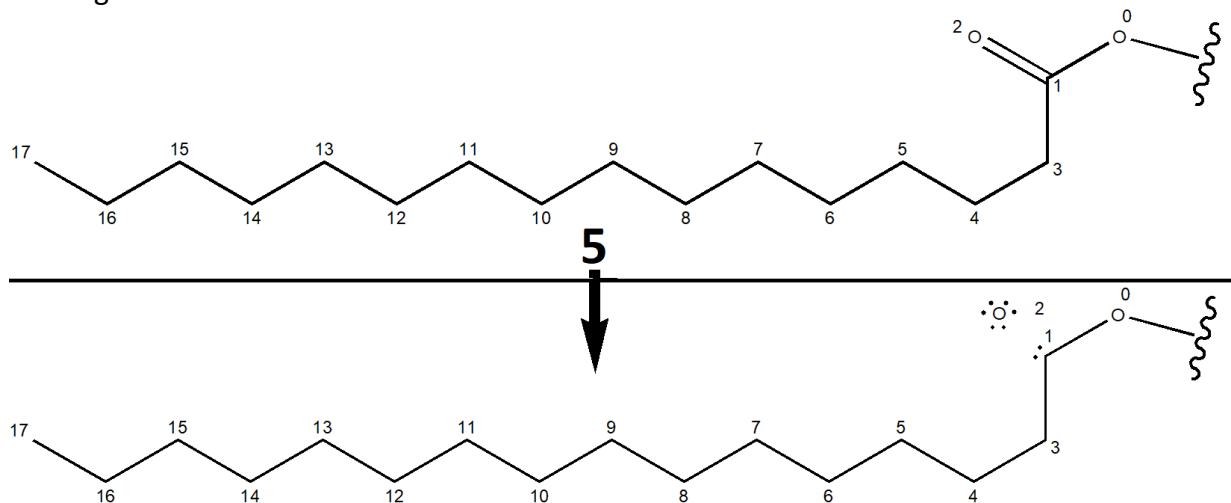
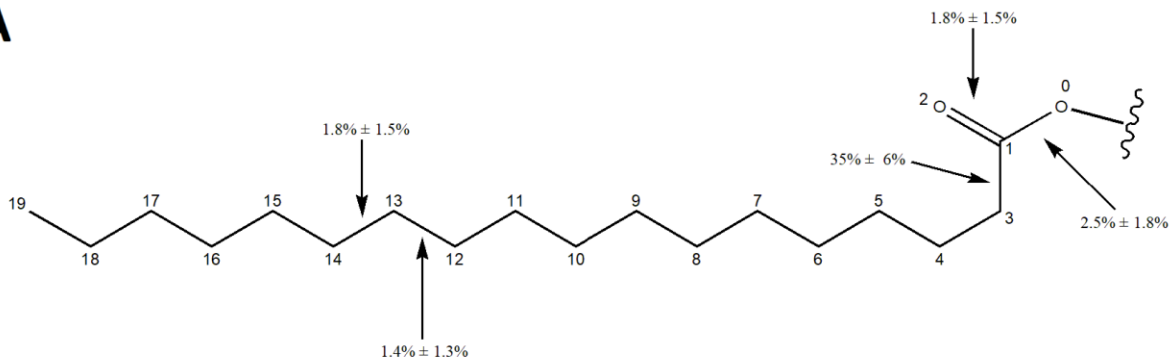


Figure 4.13 Pathway 5, 1 for palmitic acid. The breakage of 1-2 occurs followed by the breakage of 3-5. This, however, cannot be displayed due to 3-5 not existing in the initial fatty acid.

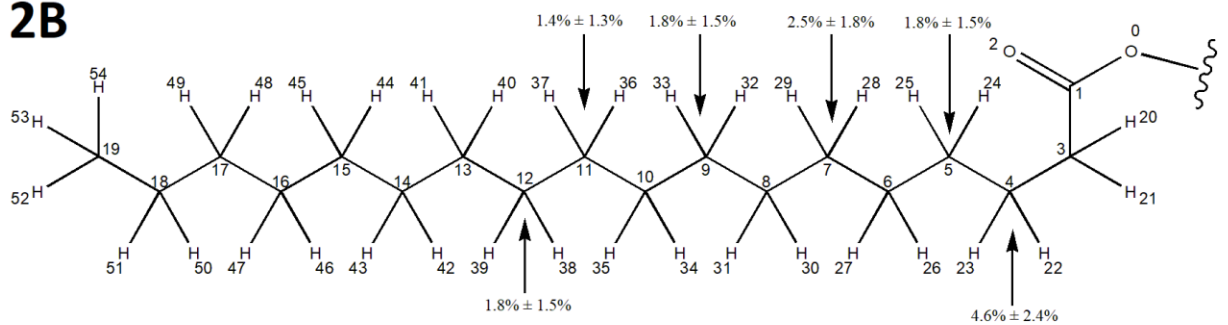
Stearic

The bond event diagrams and digraph for stearic acid can be seen in Figure 4.14 and Figure 4.15, respectively. The digraph mechanism can be seen in be seen in Figure 4.16.

2A



2B



2C



Figure 4.14 Two-dimensional bond breakage and formation diagrams for stearic acid. Structure 2A represents the breakage of heavy-heavy bonds, structure 2B represents the breakage of hydrogen-heavy bonds, and 2C represents the formation of hydrogen-heavy bonds.

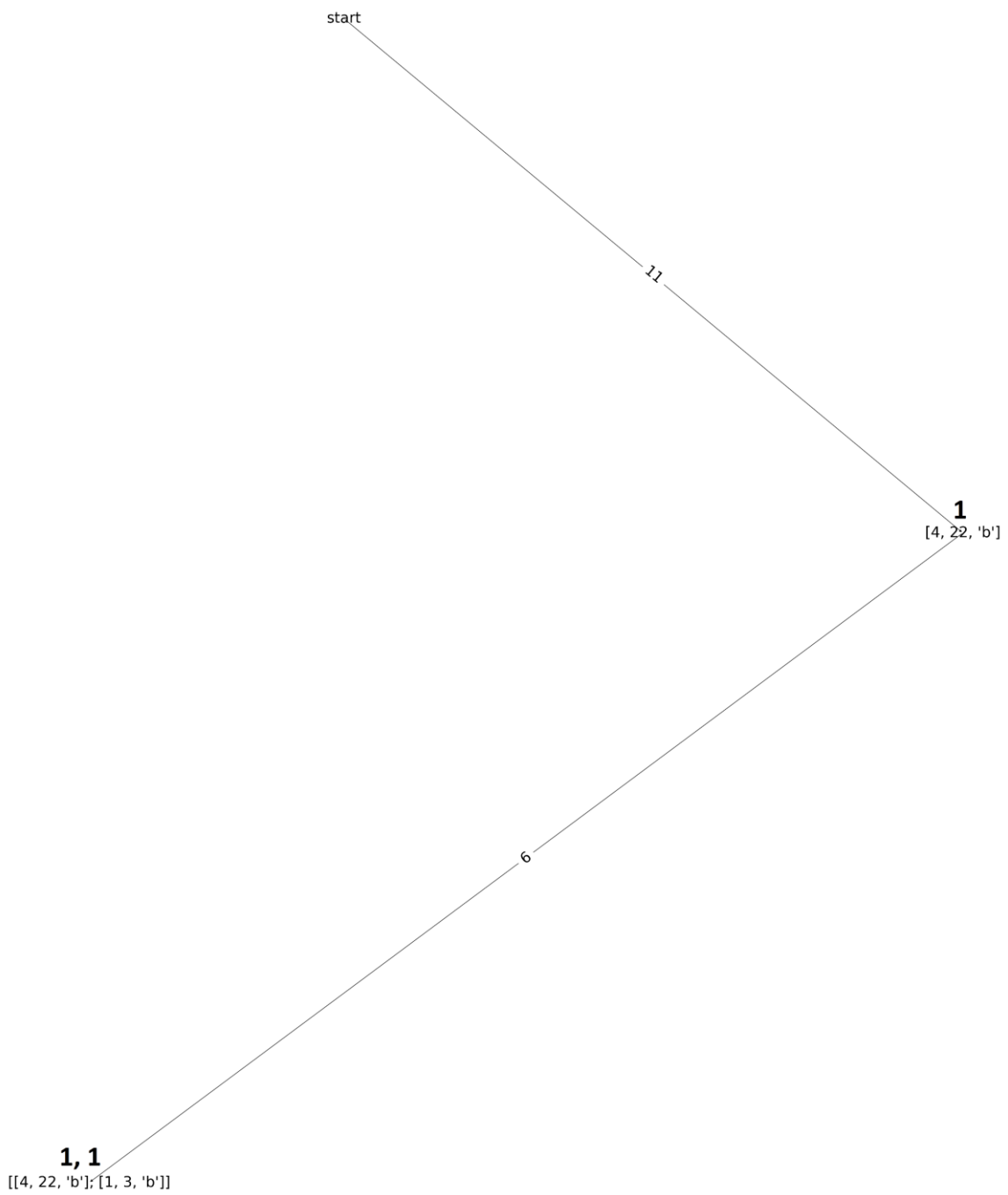


Figure 4.15 Digraph for stearic acid.

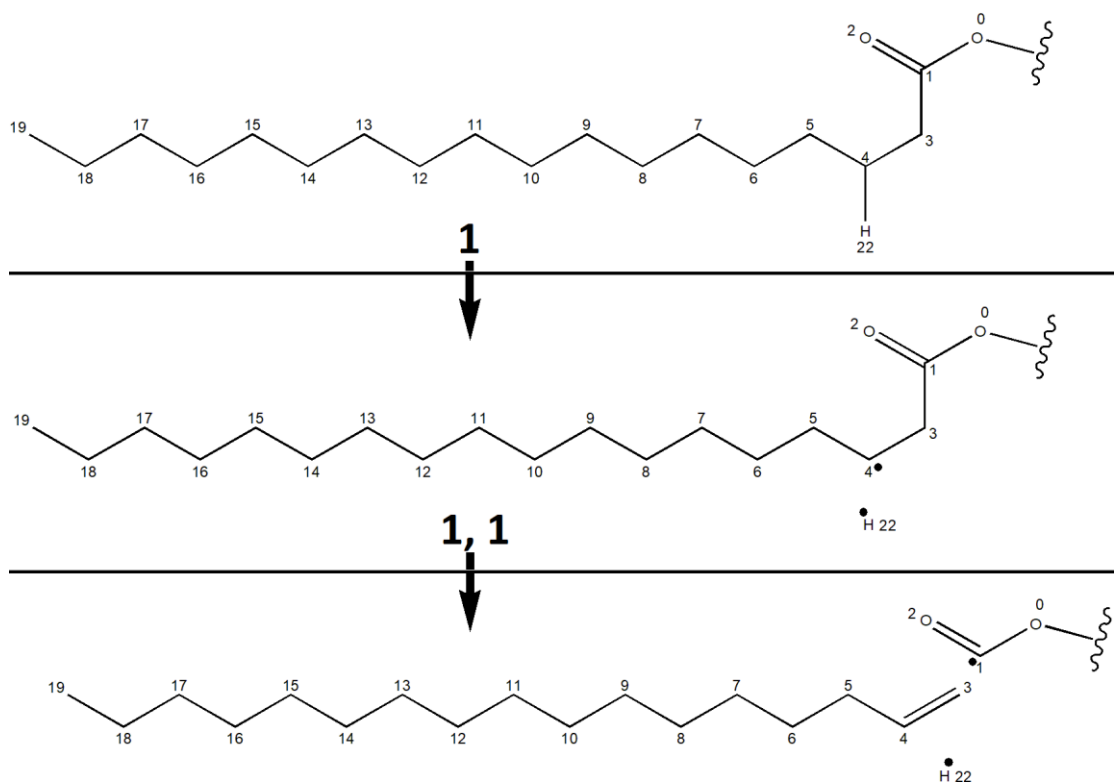


Figure 4.16 Pathway 1, 1 for stearic acid. The breakage of 4-22 occurs and is followed by the breakage of 1-3.

Oleic

The initial breakage versus bond dissociation energy diagrams for oleic acid can be seen in Figure 4.17 and 4.18. The bond event diagrams and digraph for oleic acid can be seen in Figure 4.19 and Figure 4.20, respectively. The digraph mechanisms can be seen in Figure 4.21, Figure 4.22, Figure 4.23, Figure 4.24, Figure 4.25, and Figure 4.26.

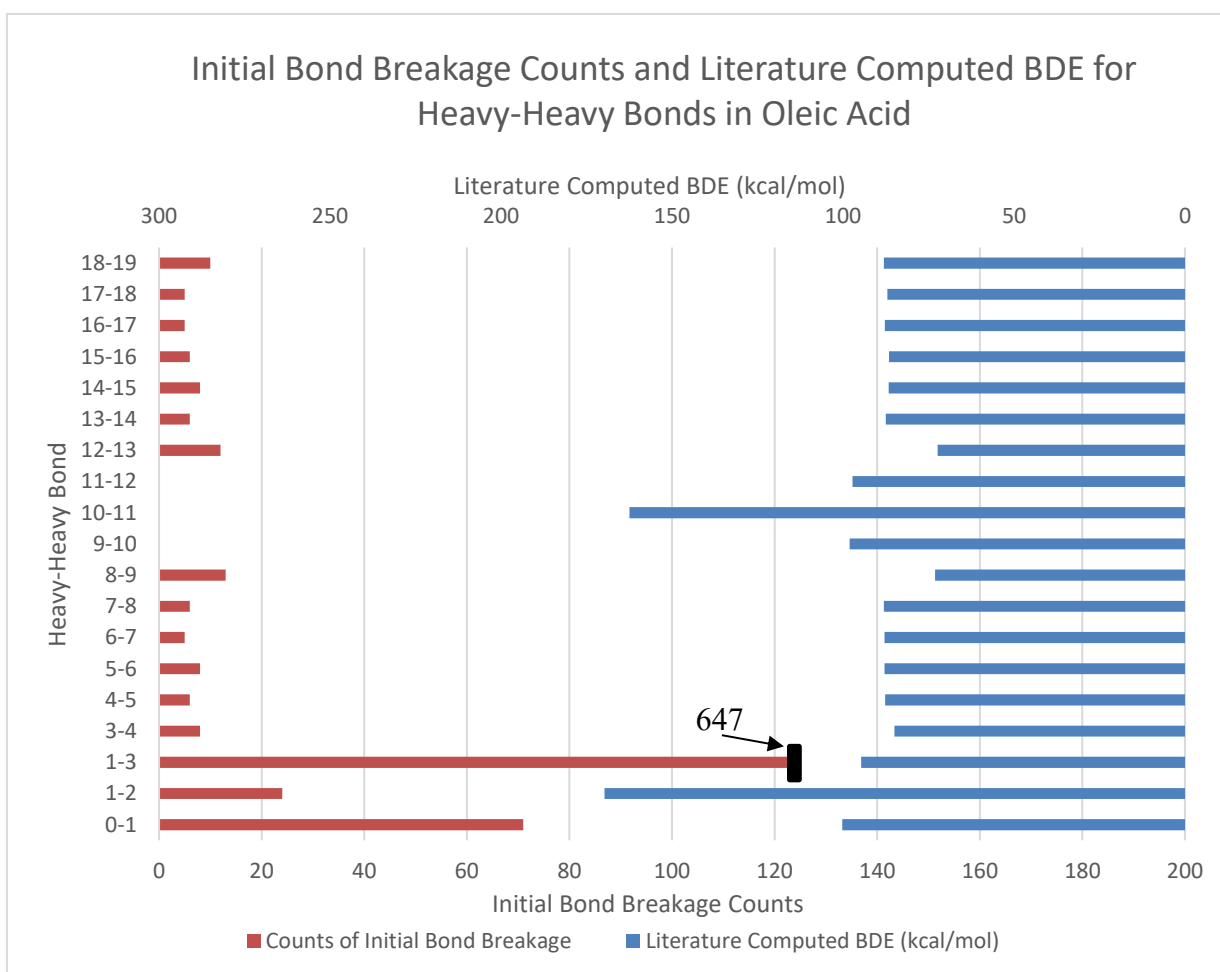


Figure 4.17 Initial breakages in oleic acid versus oleic acid BDEs for heavy-heavy bonds in a bar chart. The left side represents the raw counts of the heavy-heavy bond being the first breakage of a given oleic acid in the molecular dynamics run. The right side represents the literature calculated values from density functional theory (M06-2X/6-31+G(d,p)) in kcal/mol.^{55,56}

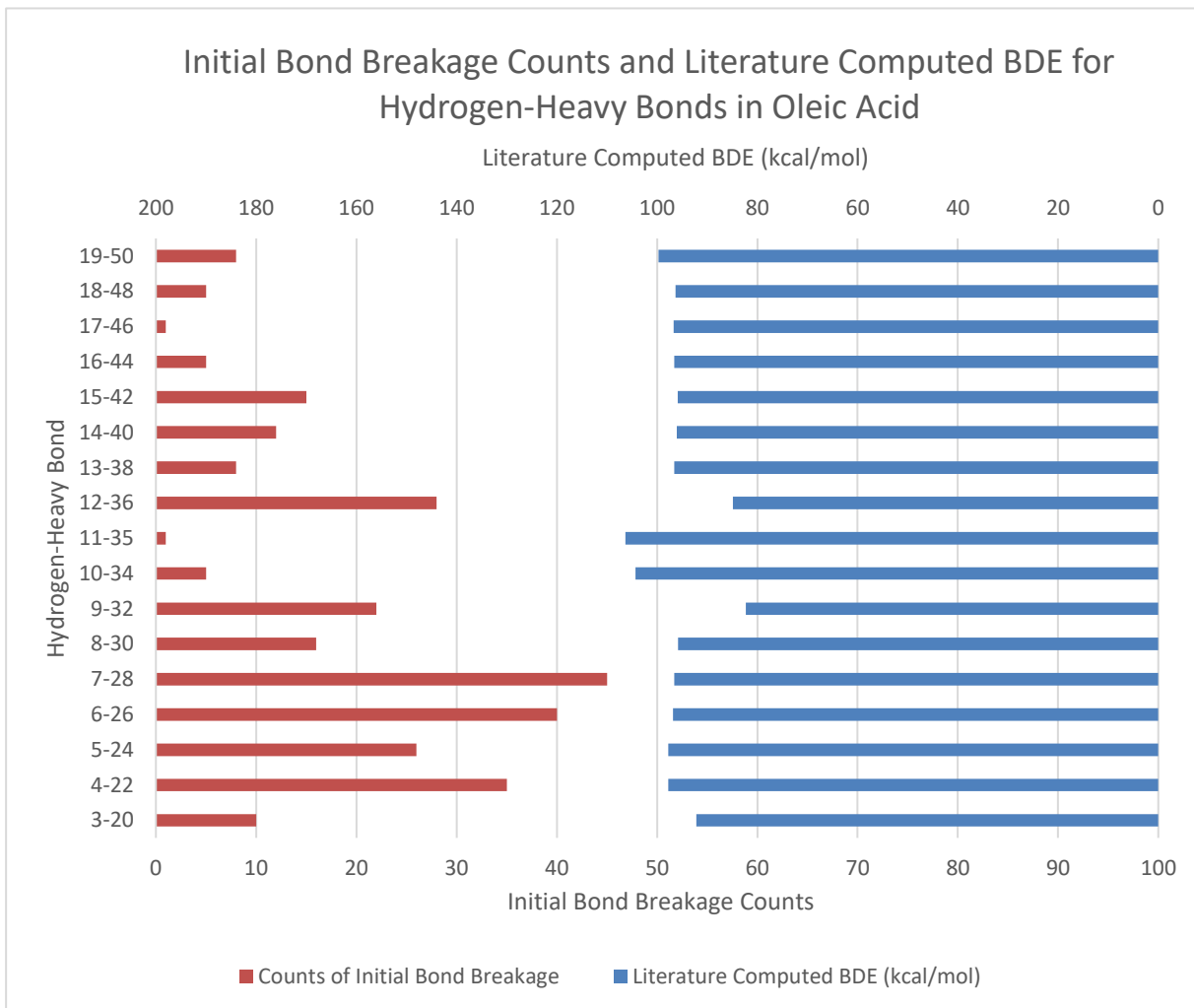


Figure 4.18 Initial breakages in oleic acid versus oleic acid BDEs for hydrogen-heavy bonds in a bar chart. The left side represents the raw counts of the hydrogen-heavy bond being the first breakage of a given oleic acid in the molecular dynamics run. The right side represents the literature calculated values from density functional theory (M06-2X/6-31+G(d,p)) in kcal/mol.^{55,56}

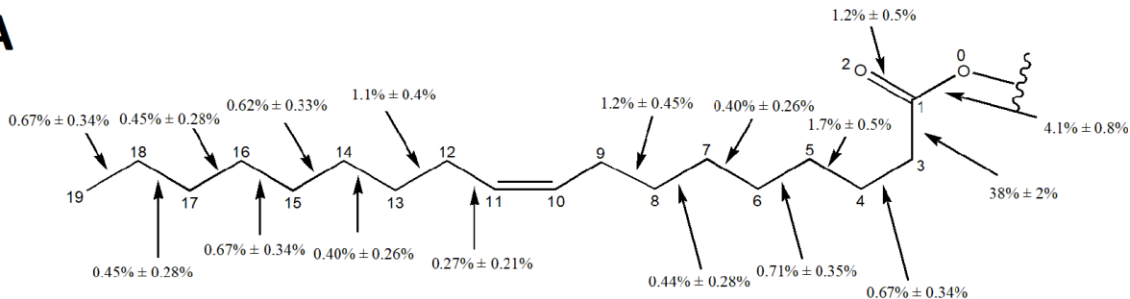
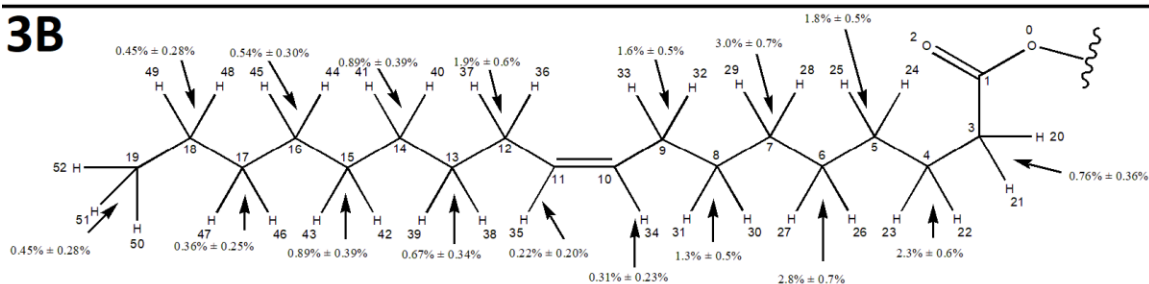
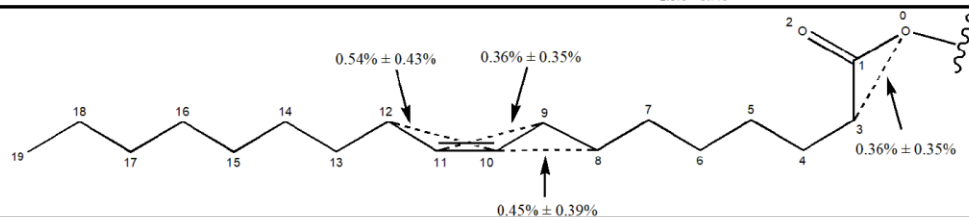
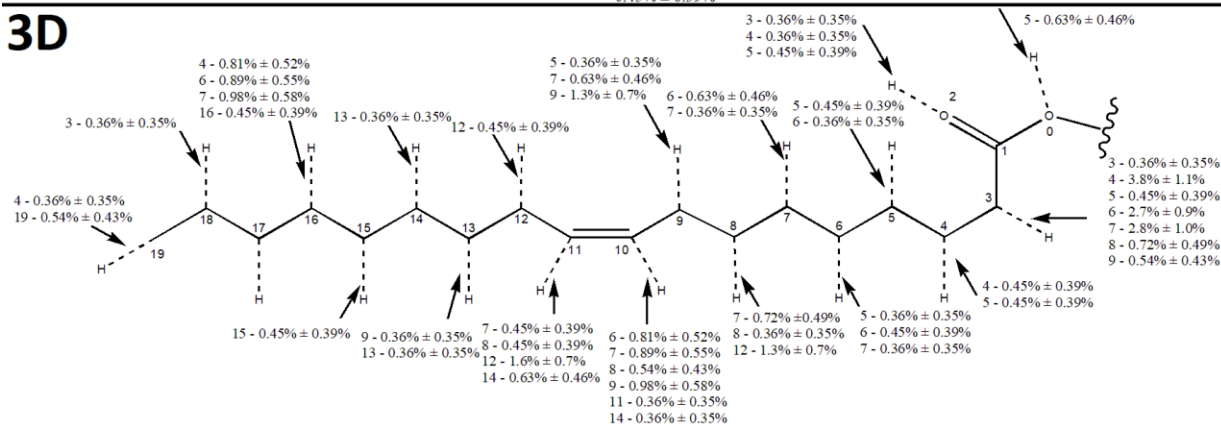
3A**3B****3C****3D**

Figure 4.19 Two-dimensional bond breakage and formation diagrams for oleic acid. Structure 3A represents the breakage of heavy-heavy bonds, structure 3B represents the breakage of hydrogen-heavy bonds, 3C represents the formation of heavy-heavy bonds, and 3D represents the formation of hydrogen-heavy bonds.

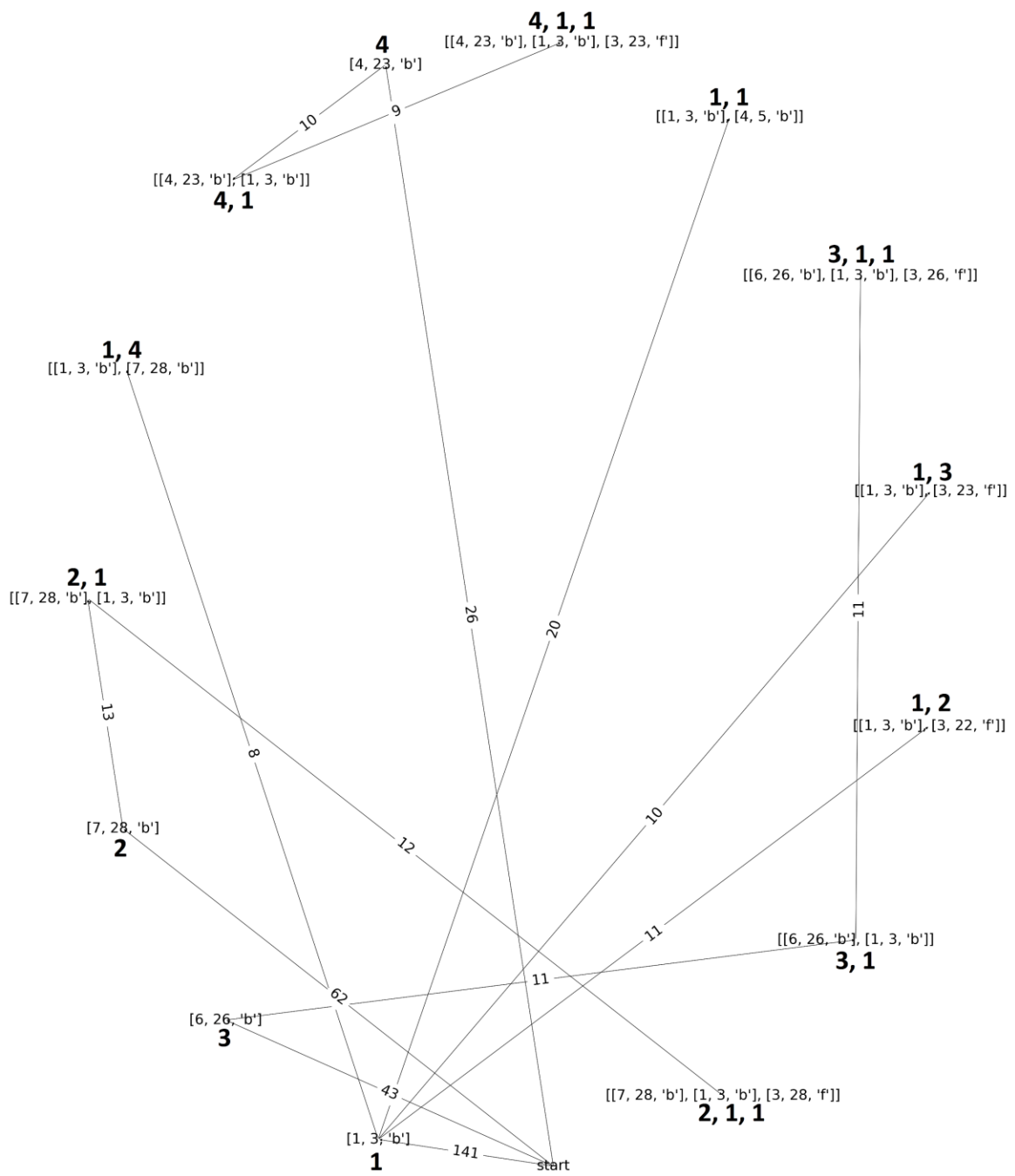


Figure 4.20 Digraph for oleic acid.

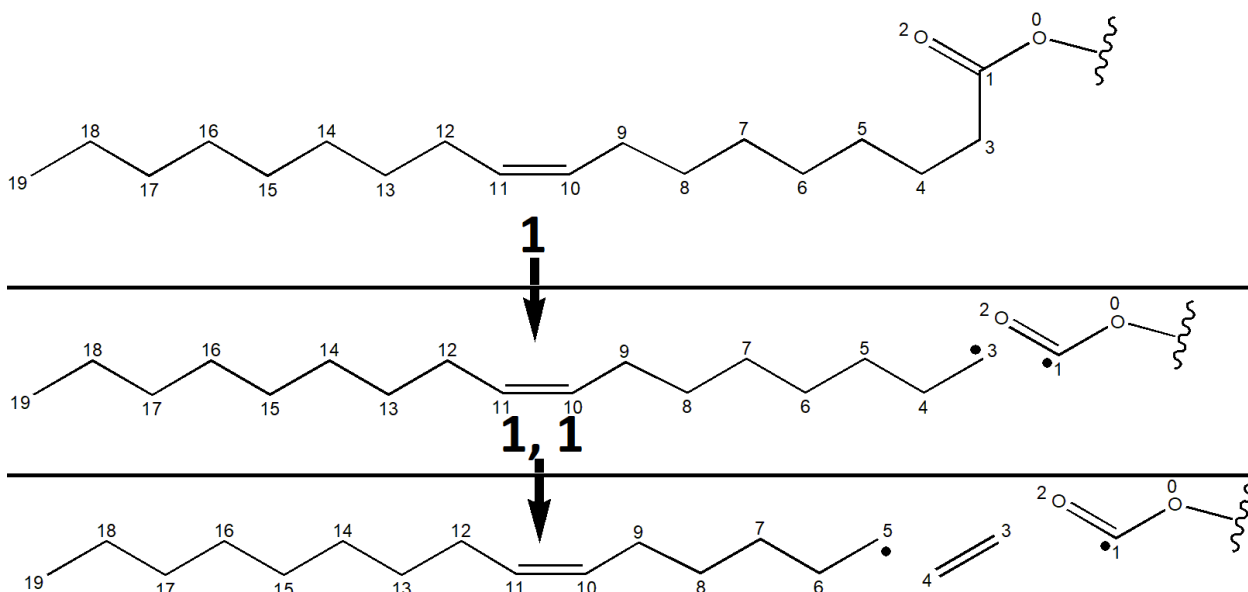


Figure 4.21 Pathway 1, 1 for oleic acid. The breakage of 1-3 occurs and is followed by the breakage of 4-5.

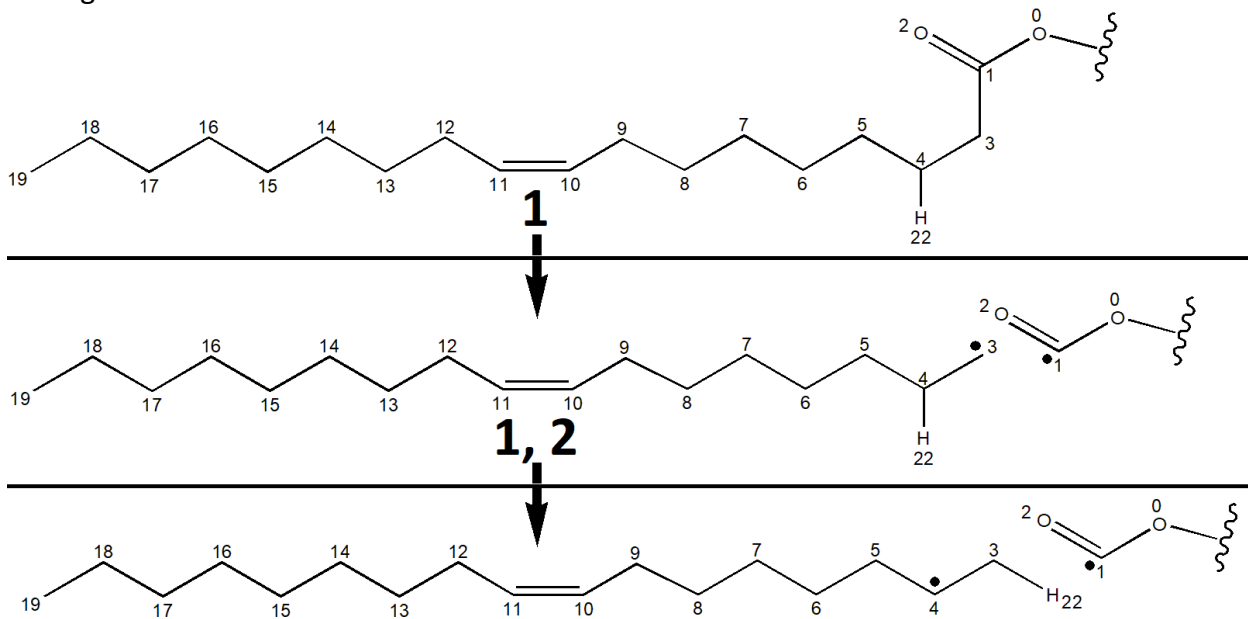


Figure 4.22 Pathway 1, 2 for oleic acid. The breakage of 1-3 occurs and is followed by the formation of 3-22. The breakage of 4-22 is implied.

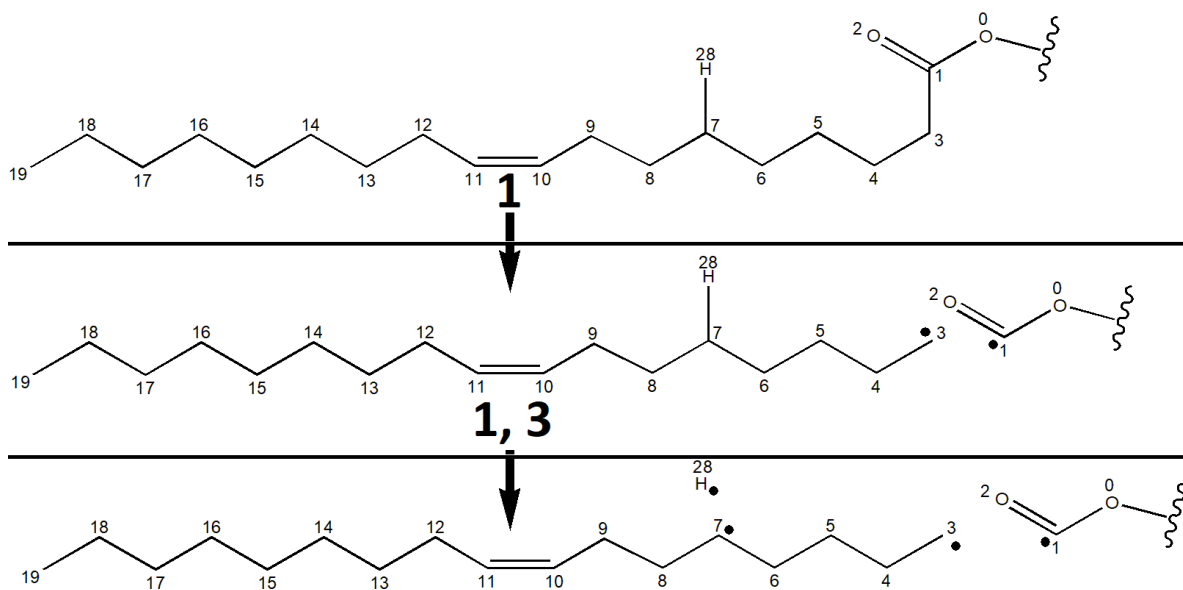


Figure 4.23 Pathway 1, 3 for oleic acid. The breakage of 1-3 occurs and is followed by the breakage of 7-28.

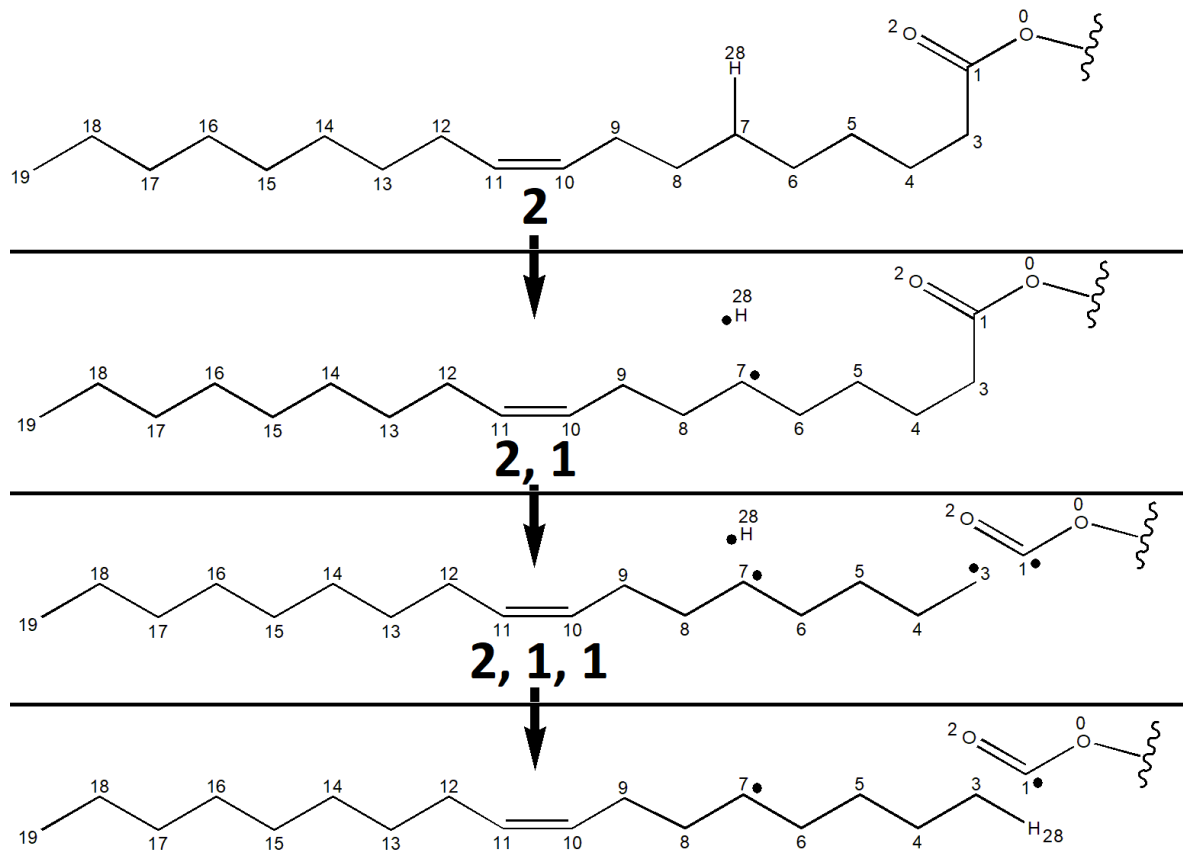


Figure 4.24 Pathway 2, 1, 1 for oleic acid. The breakage of 7-28 occurs and is followed by the breakage of 1-3 before the formation of 3-28.

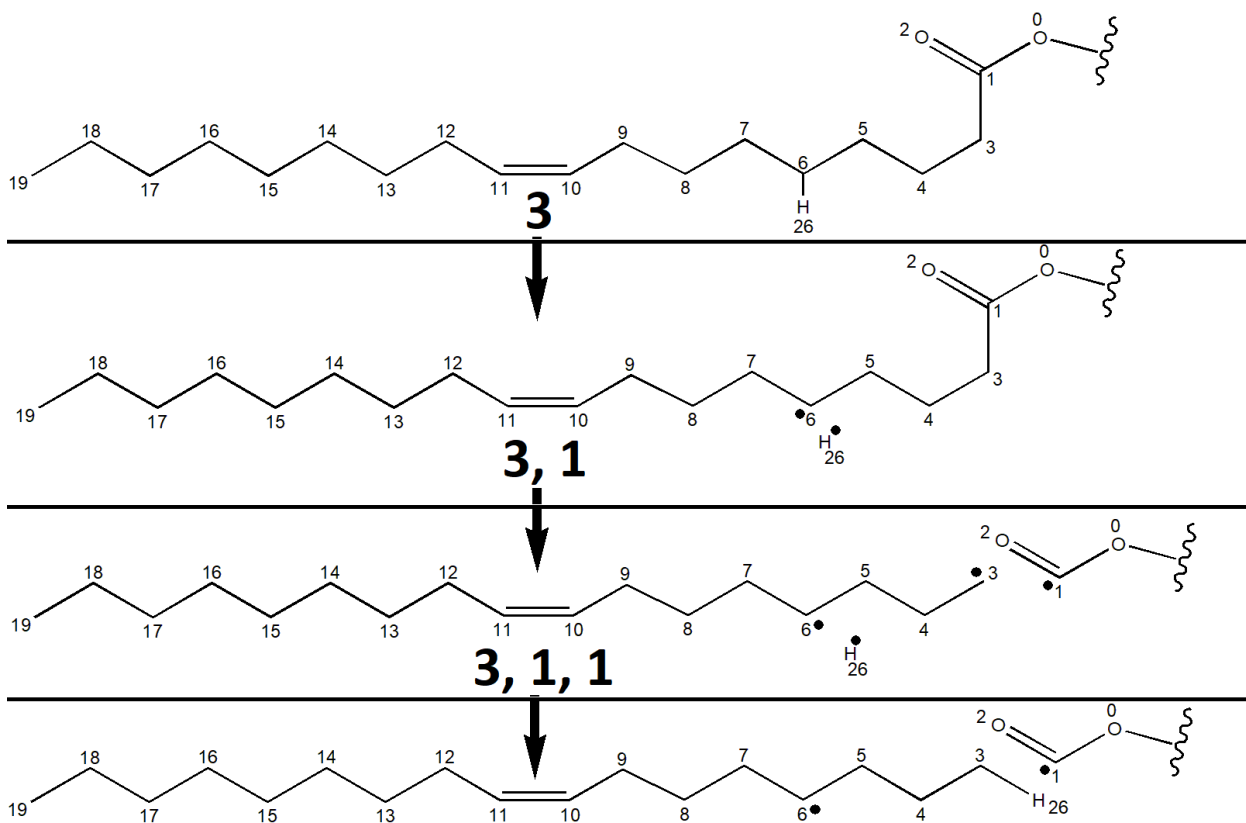


Figure 4.25 Pathway 3, 1, 1 for oleic acid. The breakage of 6-26 occurs and is followed by the breakage of 1-3 before the formation of 3-28.

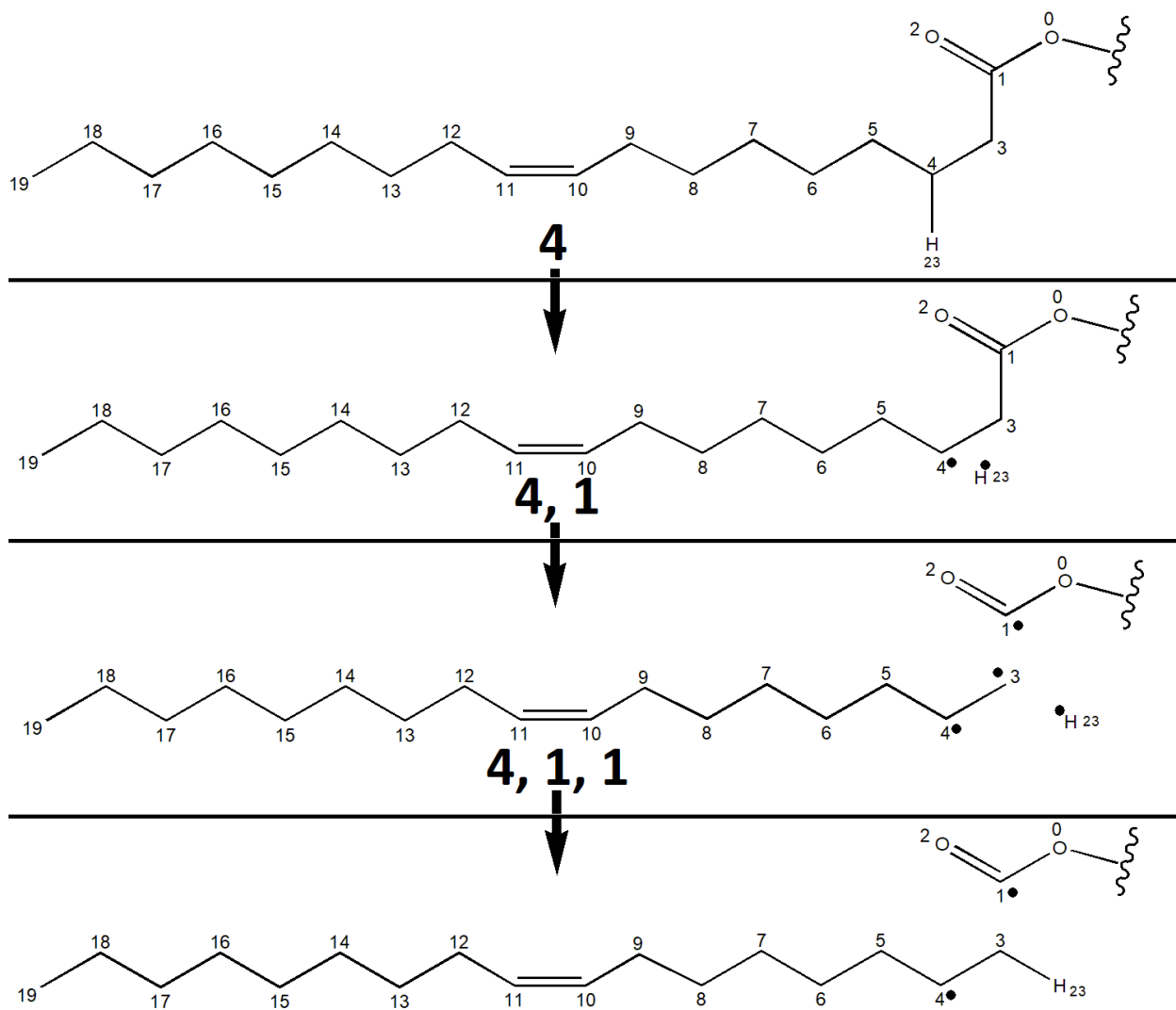


Figure 4.26 Pathway 4, 1, 1 for oleic acid. The breakage of 4-23 occurs and is followed by the breakage of 1-3 before the formation of 3-23.

Linoleic

The initial breakage versus bond dissociation energy diagrams for linoleic acid can be seen in Figure 4.27 and 4.28. The bond event diagrams and digraph for oleic acid can be seen in Figure 4.29 and Figure 4.30, respectively. The digraph mechanisms can be seen in be seen in Figure 4.31, Figure 4.32, Figure 4.33, Figure 4.34, Figure 4.35, and Figure 4.36.

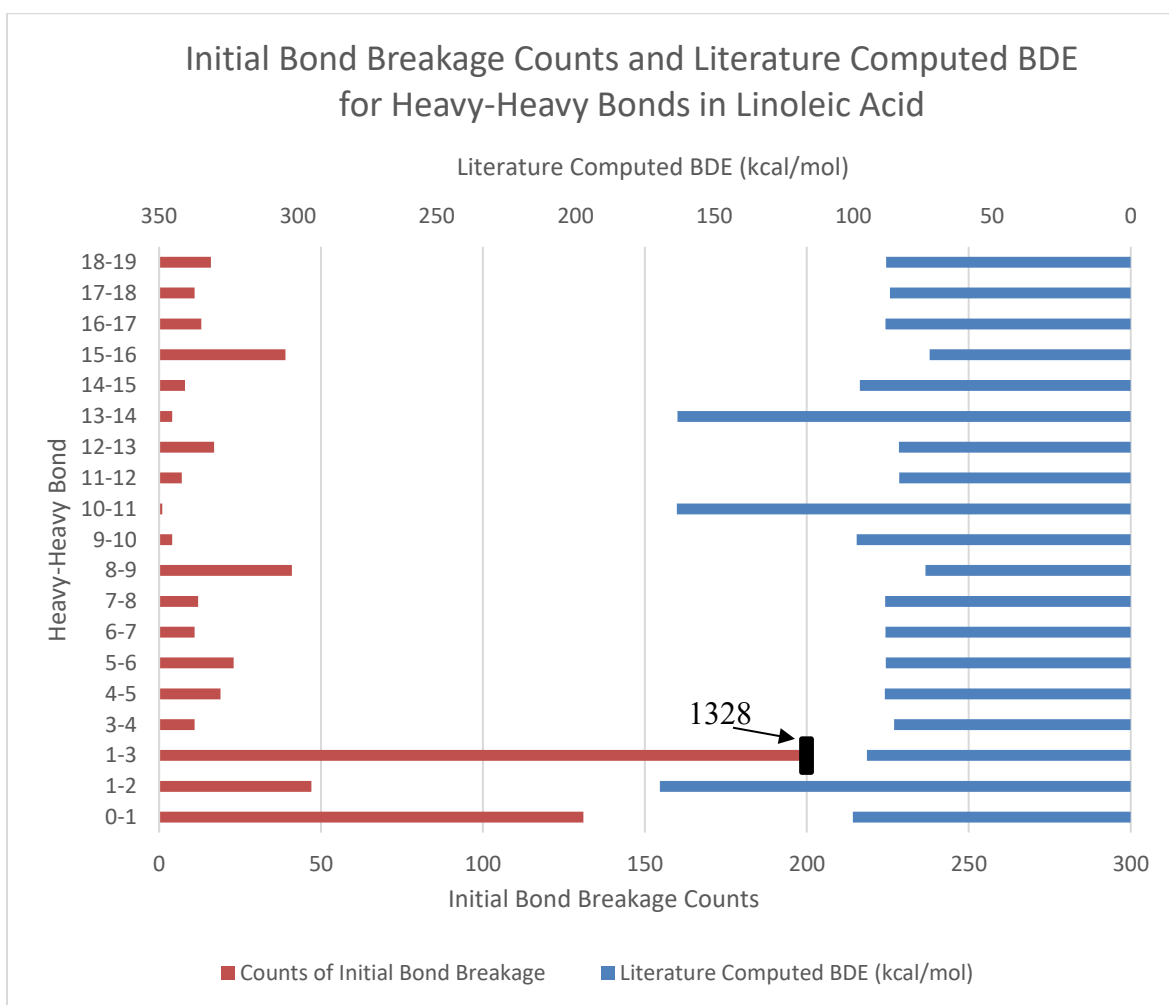


Figure 4.27 Initial breakages in linoleic acid versus linoleic acid BDEs for heavy-heavy bonds in a bar chart. The left side represents the raw counts of the heavy-heavy bond being the first breakage of a given linoleic acid in the molecular dynamics run. The right side represents the literature calculated values from density functional theory (M06-2X/6-31+G(d,p)) in kcal/mol.^{55,56}

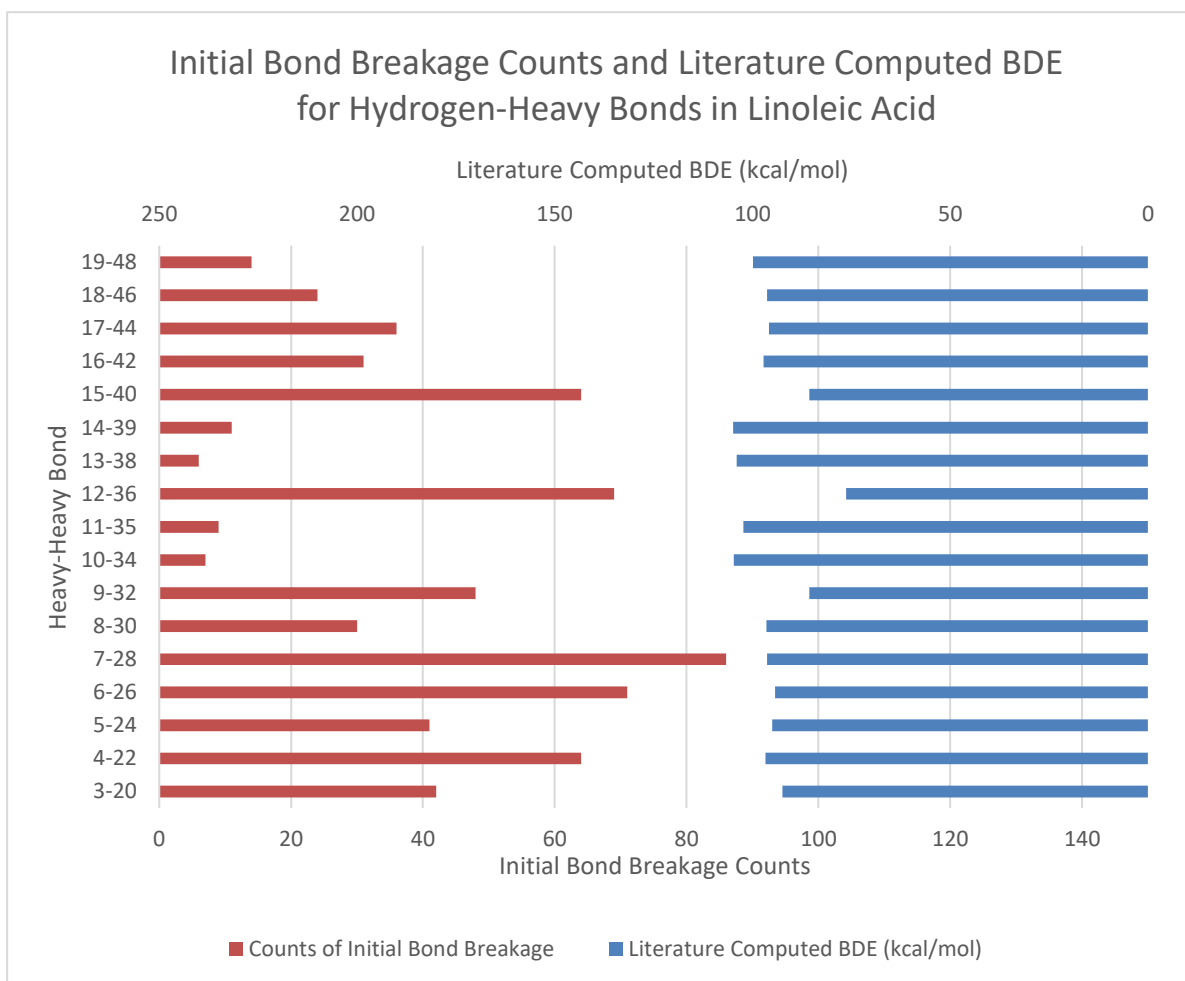
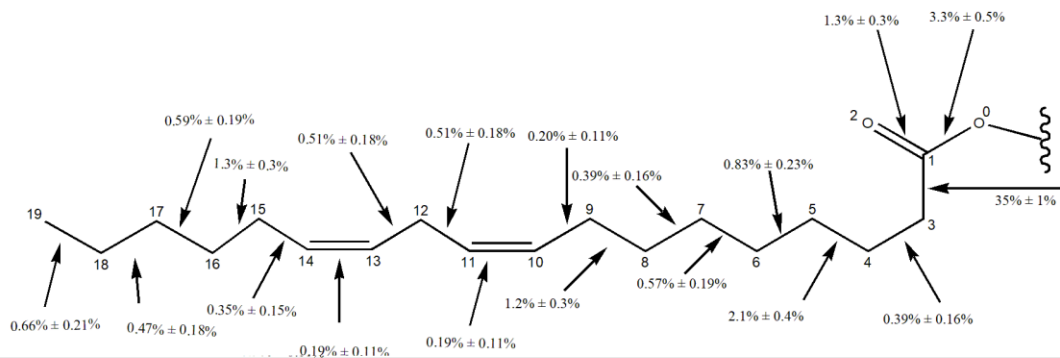
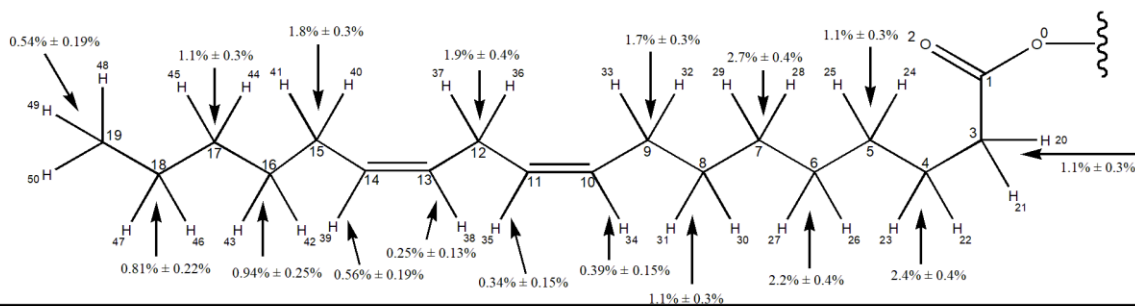


Figure 4.28 Initial breakages in linoleic acid versus linoleic acid BDEs for hydrogen-heavy bonds in a bar chart. The left side represents the raw counts of the hydrogen-heavy bond being the first breakage of a given linoleic acid in the molecular dynamics run. The right side represents the literature calculated values from density functional theory (M06-2X/6-31+G(d,p)) in kcal/mol.^{55,56}

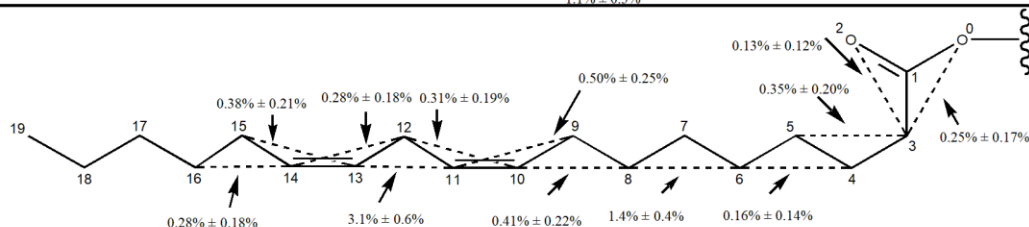
4A



4B



4C



4D

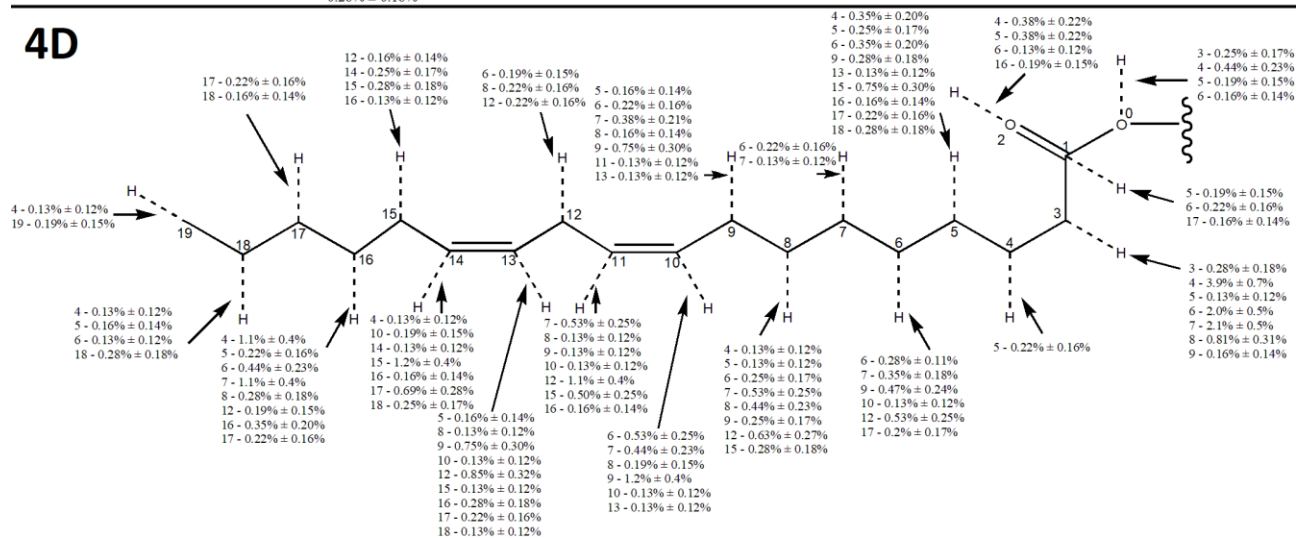


Figure 4.29 Two-dimensional bond breakage and formation diagrams for linoleic acid. 4A represents the breakage of heavy-heavy bonds, structure 4B represents the breakage of hydrogen-heavy bonds, 4C represents the formation of heavy-heavy bonds, and 4D represents the formation of hydrogen-heavy bonds.

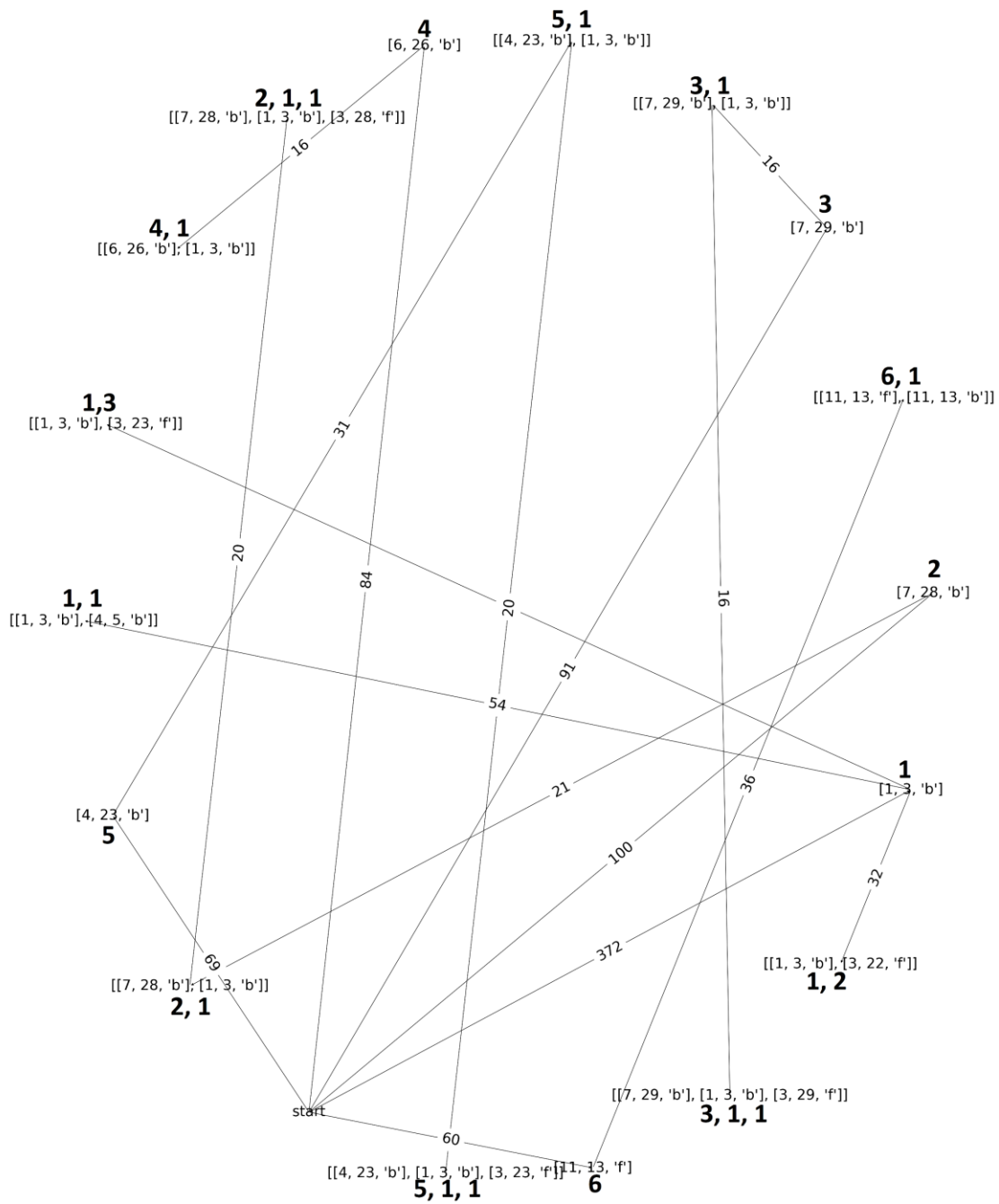


Figure 4.30 Digraph for linoleic acid.

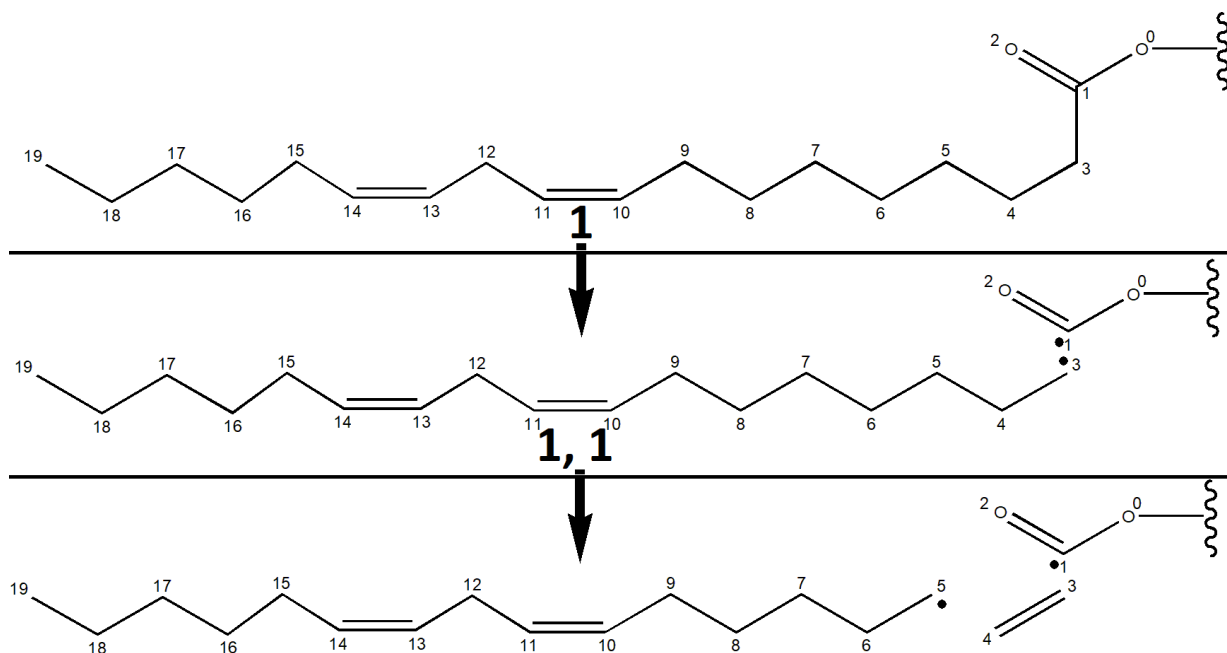


Figure 4.31 Pathway 1, 1 for linoleic acid. The breakage of 1-3 occurs before the breakage of 4-5.

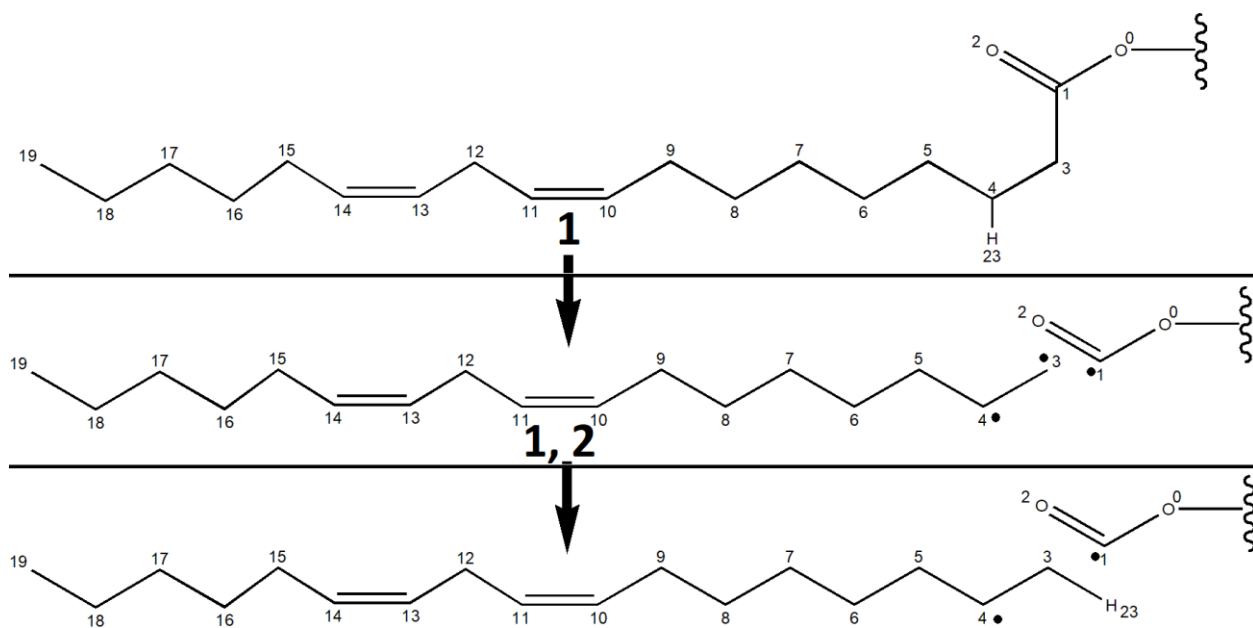


Figure 4.32 Pathway 1, 2 for linoleic acid. The breakage of 1-3 occurs before the formation of 3-23. The breakage of 4-23 is implied.

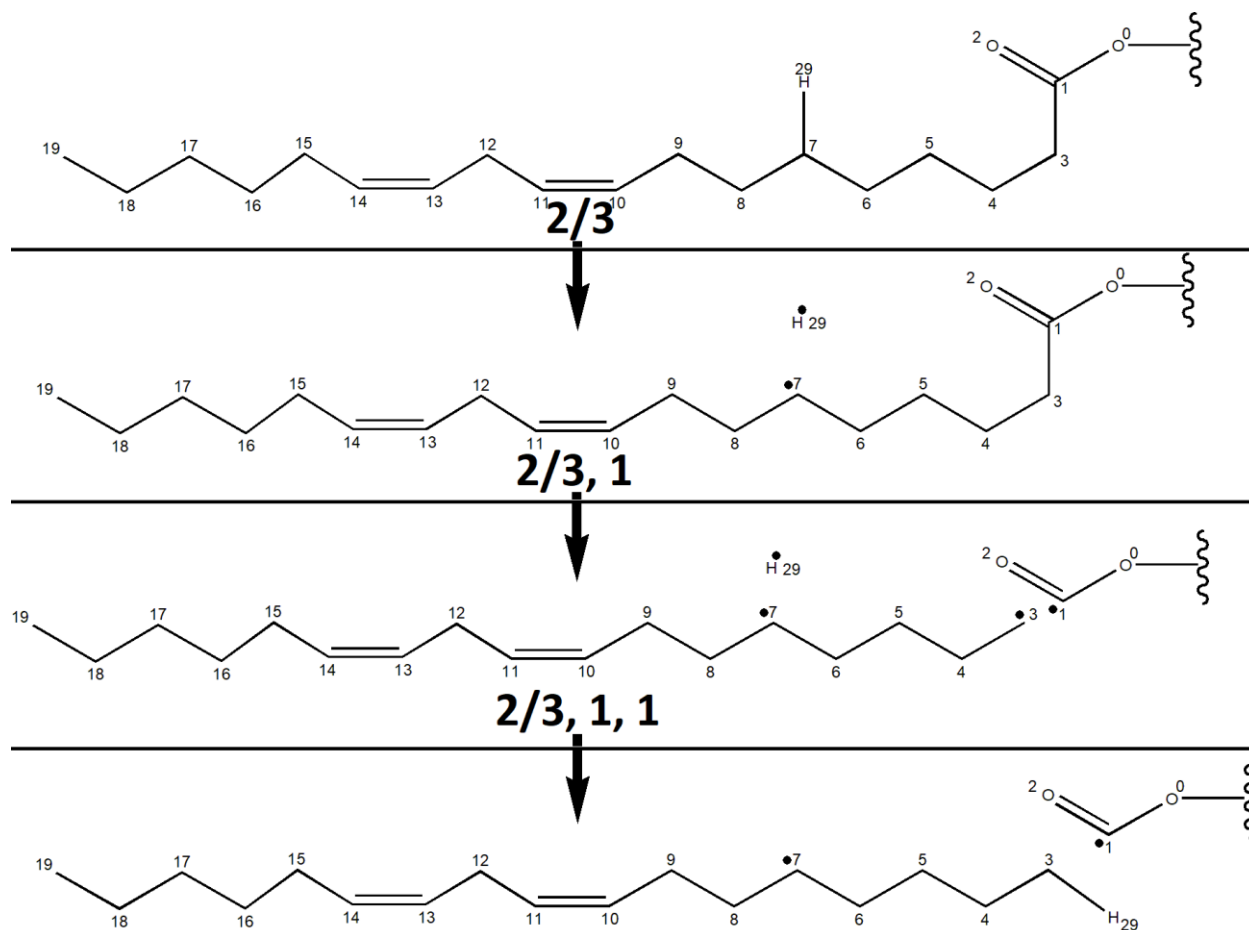


Figure 4.33 Pathway 2, 1, 1 and 3, 1, 1 for linoleic acid. The two pathways are identical, indicated by the notation 2/3, 1, 1. The breakage of 7-29/7-28 occurs before the breakage of 1-3, followed by the formation of 3-28/3-29.

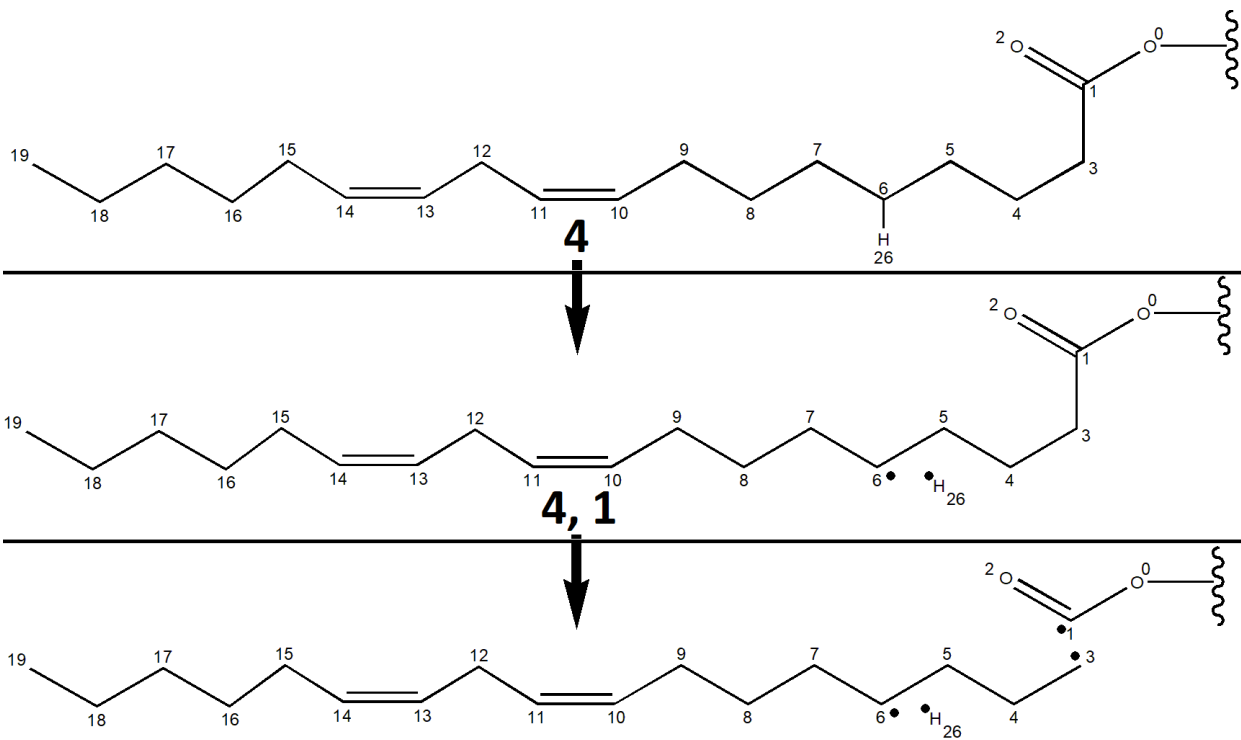


Figure 4.34 Pathway 4, 1 for linoleic acid. The breakage of 6-26 occurs before the breakage of 1-3.

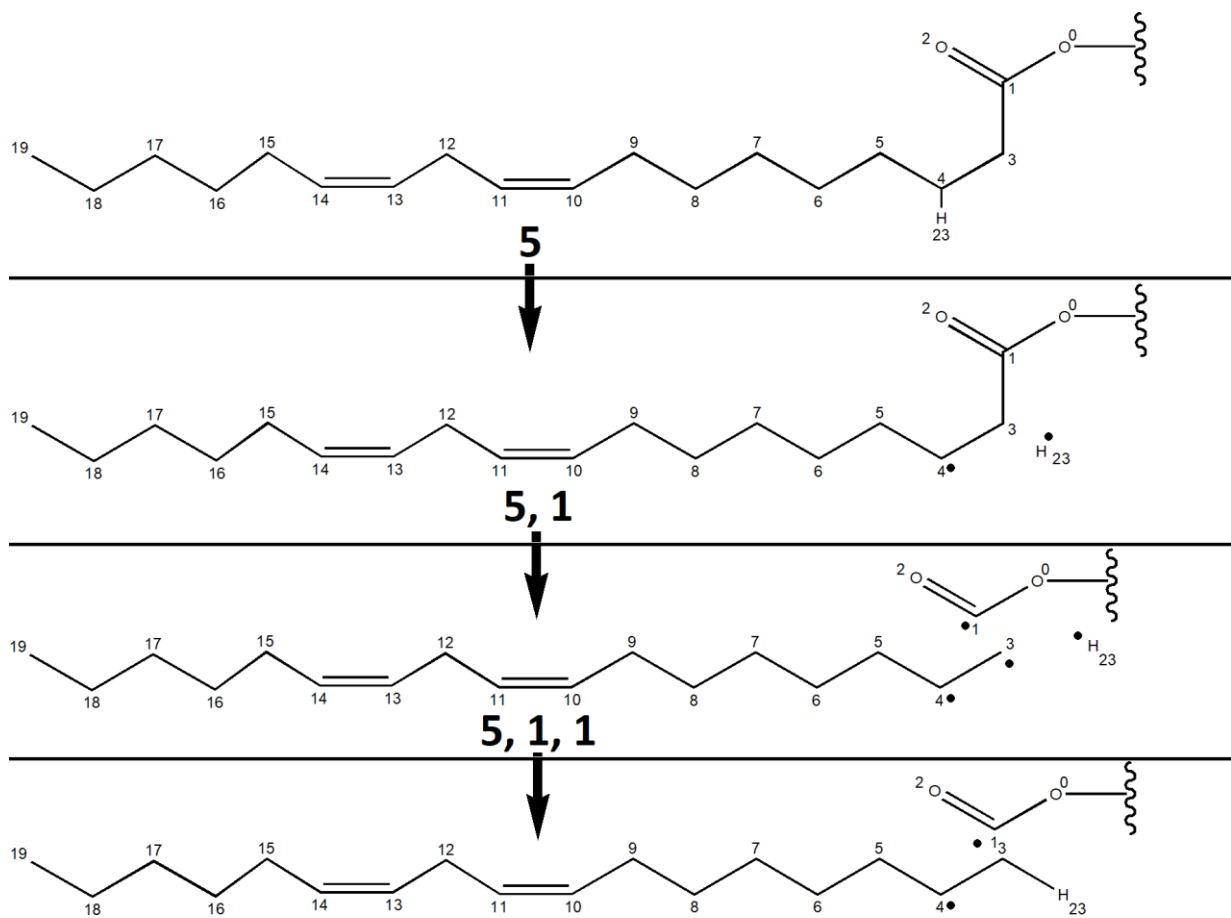


Figure 4.35 Pathway 5, 1, 1 for linoleic acid. The breakage of 4-23 occurs before the breakage of 1-3, followed by the formation of 3-23.

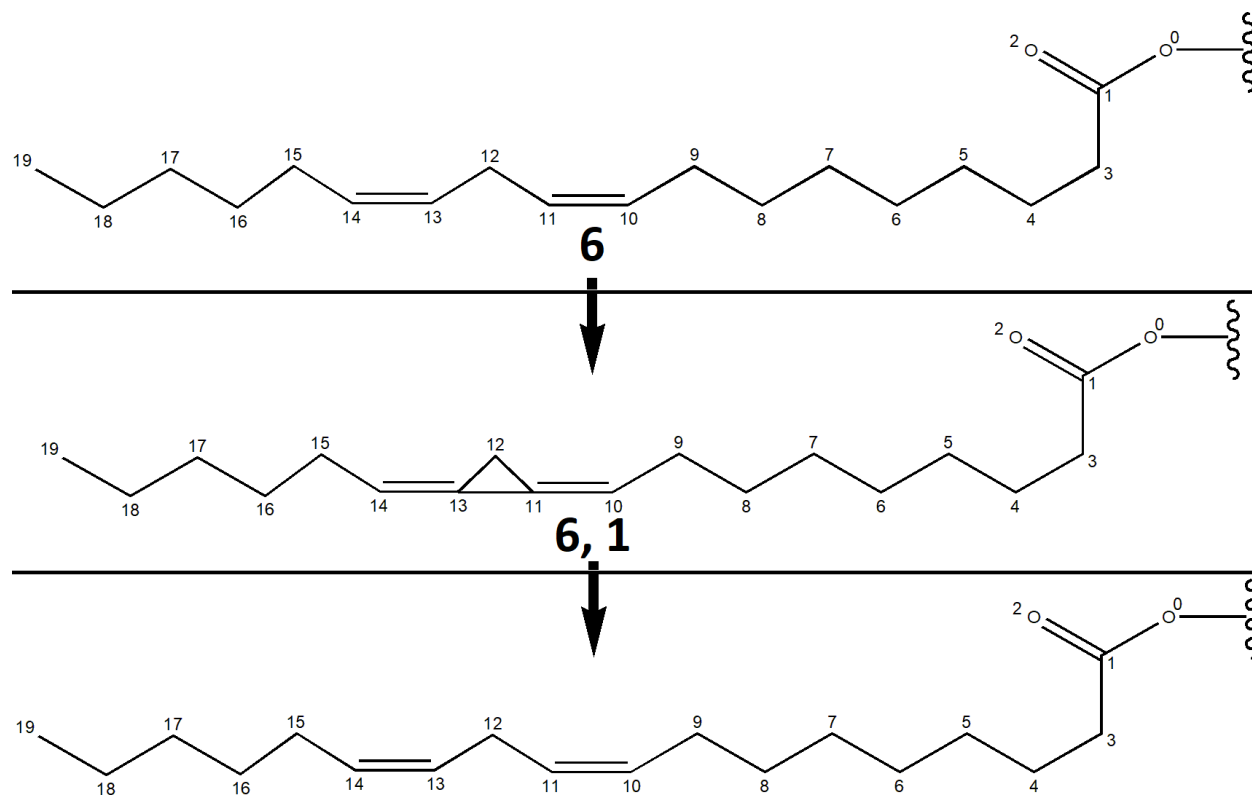
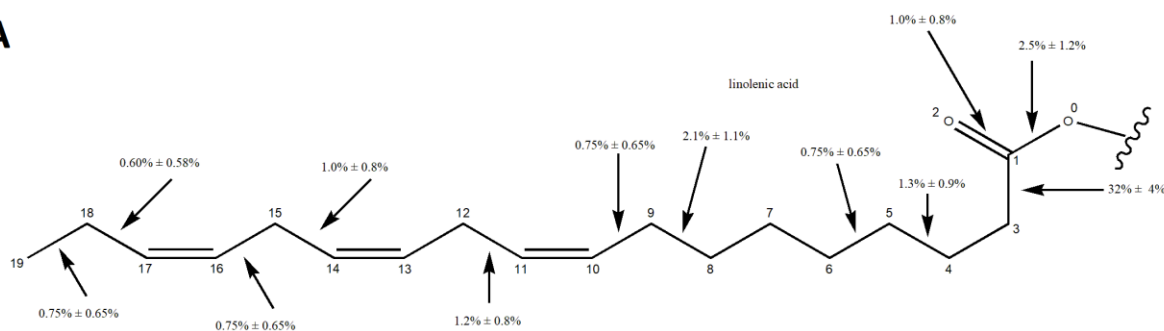


Figure 4.36 Pathway 6, 1 for linoleic acid. The formation of 11-13 occurs before the breakage of 11-13. It should be noted that other events may have occurred before the formation of 11-13, so an exceeded valence shell may not actually be observed.

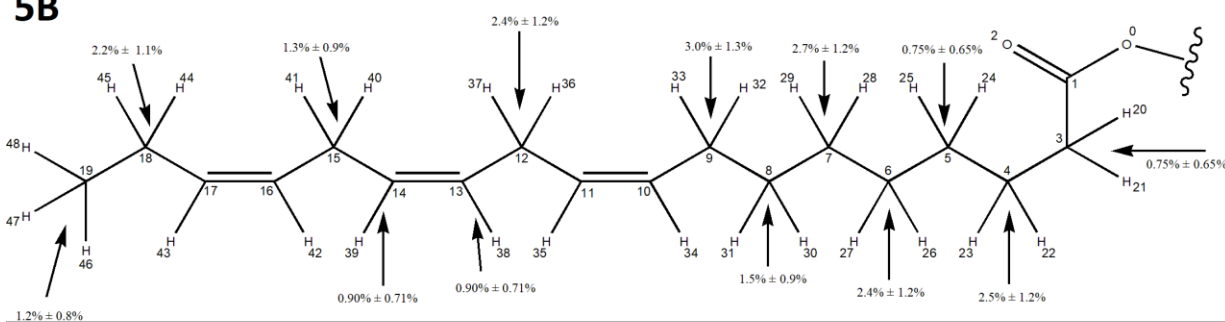
Linolenic

The bond event diagrams and digraph for oleic acid can be seen in Figure 4.37 and Figure 4.38, respectively. The digraph mechanisms can be seen in be seen in Figure 4.39 and Figure 4.40.

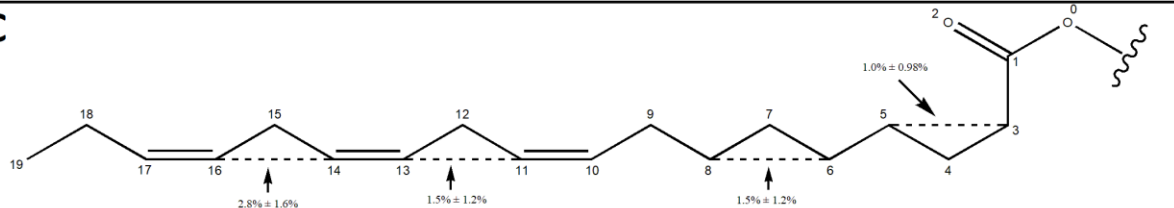
5A



5B



5C



5D

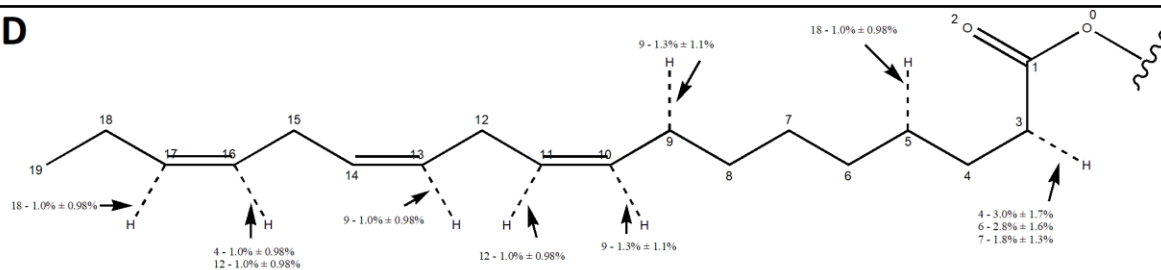


Figure 4.37 Two-dimensional bond breakage and formation diagrams for linolenic acid. 5A represents the breakage of heavy-heavy bonds, structure 5B represents the breakage of hydrogen-heavy bonds, 5C represents the formation of heavy-heavy bonds, and 5D represents the formation of hydrogen-heavy bonds.

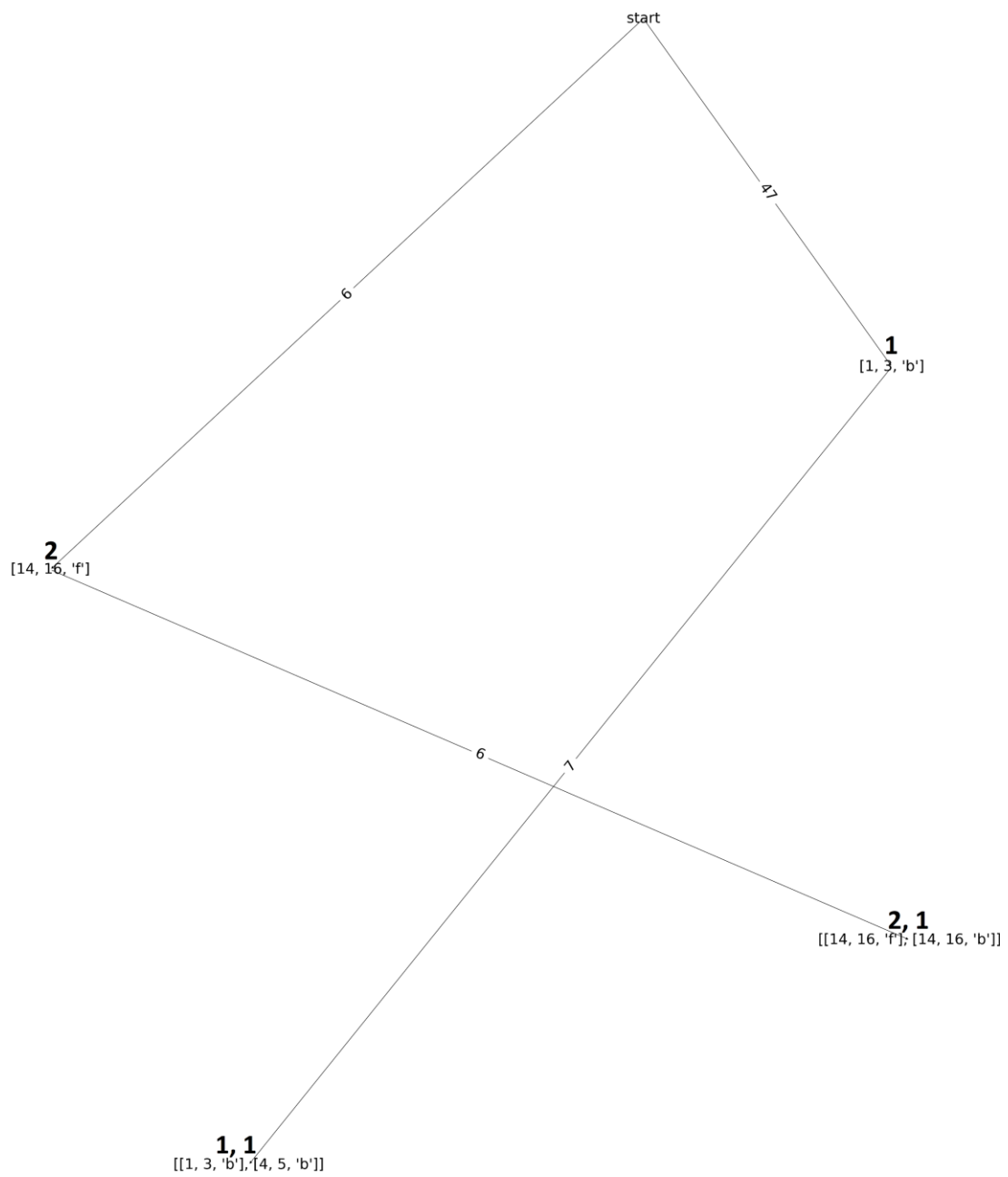


Figure 4.38 Digraph for linolenic acid.

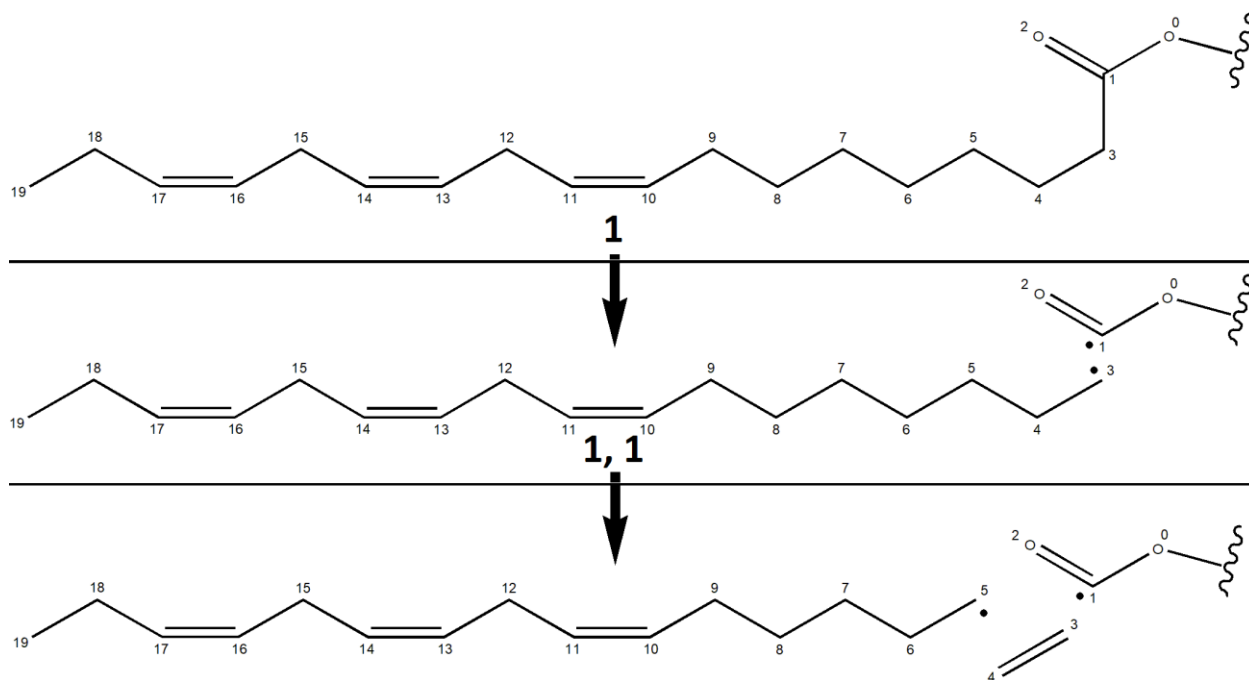


Figure 4.39 Pathway 1, 1 for linolenic acid. The breakage of 1-3 occurs before the breakage of 4-5.

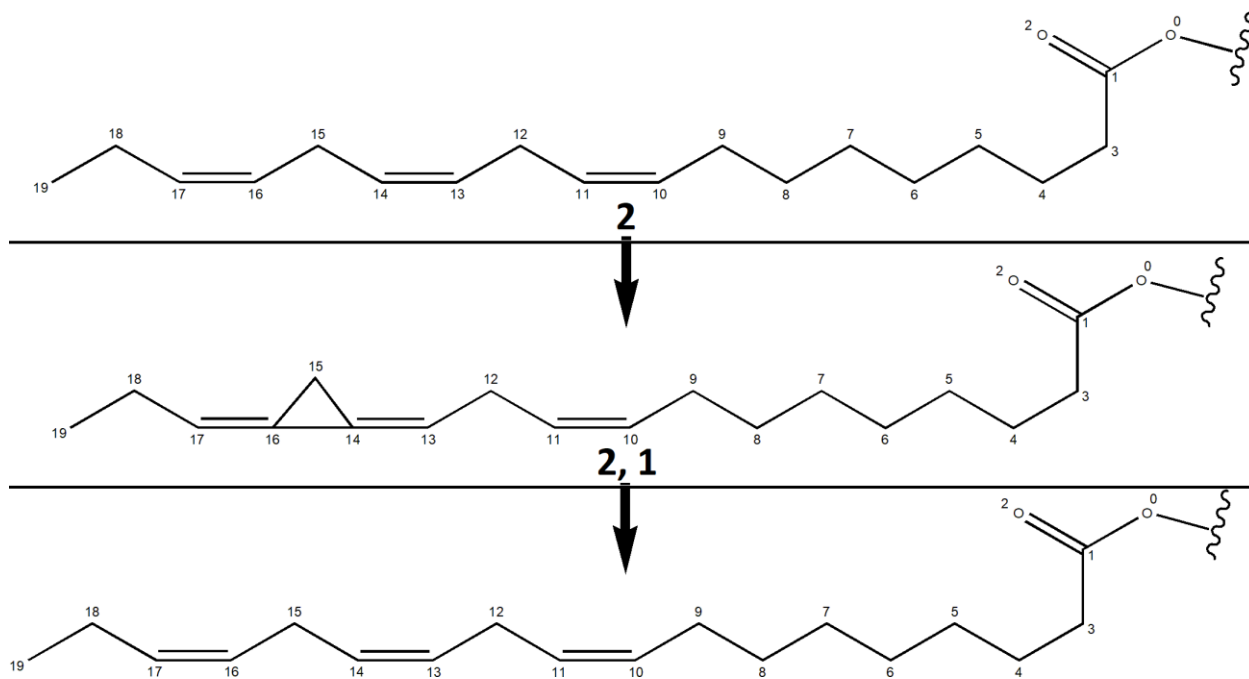


Figure 4.40 Pathway 2, 1 for linolenic acid. The formation of 14-16 occurs before the breakage of 14-16. It should be noted that other events may have occurred before the formation of 11-13, so an exceeded valence shell may not actually be observed.

The breakage of 1-3 has been shown to be the most common breakage by nearly one magnitude, which can be seen in structure A of Figures 4.6, 4.14, 4.19, 4.29, and 4.37. This may be due to the unrecorded events in the glycerol backbone which, in tandem with the breakage of 1-3, may form CO₂ in pathways seen in Figure 1.8. Pathways 1a shows the breakage of a glycerol carbon (no atomic index assigned) bond with the ester oxygen (atom 0) and 2a shows the breakage of 1-3. This explanation would also apply to other increased breakages near the glycerol backbone, such as 0-1, which breaks relatively frequently compared to other carbon-carbon bonds despite being sp² hybridized.

Though only including initial breakages in Figures 4.17, 4.18, 4.27, and 4.28 were intended to circumvent the issue of recording bond breakages that are dependent on other events, the issue still exists with the breakage of bonds near the glycerol backbone as the glycerol backbone events were not recorded in these initial counts. As bonds get farther away from the glycerol backbone, the expected results (from BDE values) are more frequently observed. This is contrary to the bonds near the glycerol backbone (mostly 0-1, 1-2, and 1-3) with much more breakages than expected given relatively high BDE values. The linear trendline and R² values for each initial breakage versus BDE plot can be seen in Table 3.

For heavy-hydrogen bonds, bond 7-28 for both oleic and linoleic acid is the most broken bond. This is, perhaps, due to the same reason as previously mentioned, events in the glycerol backbone. As can be seen in Figures 4.23 and 4.33, the breakage of 7-28 occurs frequently in tandem with the breakage of 1-3, in order to form 3-28. Due to these circumstances,

information of events in the glycerol backbone would help a great deal in explaining the most common events observed in this work.

Table 3. Trendline equations and correlation coefficients for initial breakage counts versus BDE plot. See Appendix for plots.

Fatty acid	Initial breakage (y) versus BDE (x) of...	Linear trendline equation	Linear trendline R ²
Oleic acid	Heavy-heavy bonds	$y = -0.0132x + 45.48$	0.000005
	Hydrogen-heavy bonds	$y = -0.7749x + 91.01$	0.103050
Linoleic acid	Heavy-heavy bonds	$y = -0.4532x + 137.13$	0.002031
	Hydrogen-heavy bonds	$y = -2.3292x + 260.89$	0.471888

Figure 4.41 demonstrates a flowchart can be seen that demonstrates the data pipeline for the molecular dynamics simulations. The bond dissociation energy data was not included in this pipeline as it was not required to create digraphs or bond event diagrams. This was created for ease of understanding the process of changing one triglyceride into interpretable mechanistic pathways. The bubbles in the chart are representative of the current form of the data, and the arrows in the chart are representative of the processing applied to that data. Note that the digraph bubble ultimately turns into a large number of chemdraw sketches, in order to make the nodes more easily interpretable to the audience.

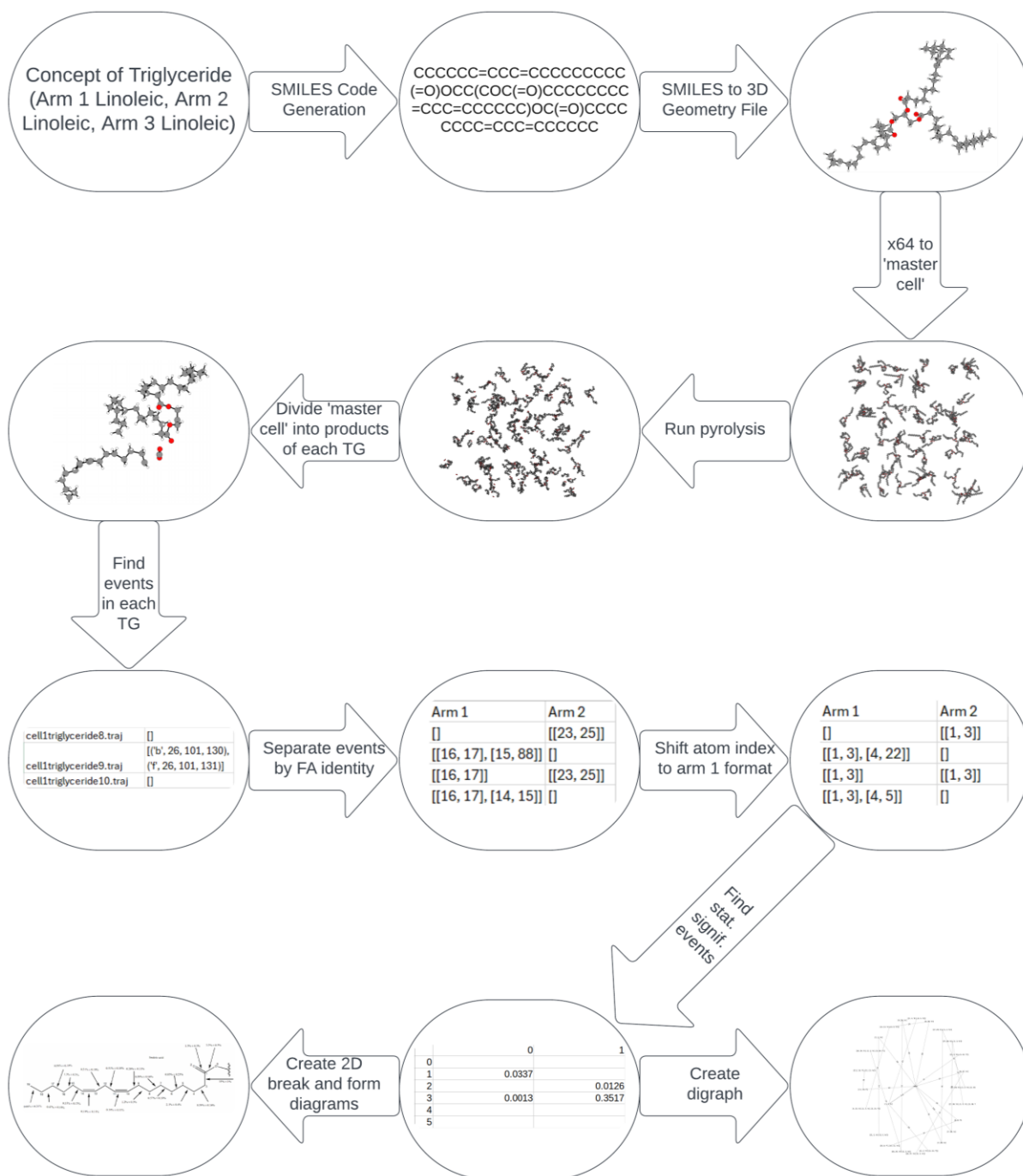


Figure 4.41 Flowchart demonstrating the data pipeline for molecular dynamics of triglycerides.

4.5 Bond Dissociation Energy Analysis

BDEs represent the strength of a chemical bond and can be defined by the difference between the enthalpies of the unbonded and the bonded fragment radicals. This energy can be represented by graphs such as Figure 2.1, of which the BDE (in the gas phase) is equal to the difference between the minimum of the graph (energy of the bonded pair) and $\lim_{r \rightarrow \infty} f(r)$ (energy of the unbonded pair). BDEs were then calculated for homolytic bond cleavage of common hydrocarbons and other common molecules that would be encountered in biodiesel pyrolysis simulations, with a geometry file for each molecule taken from a paper by Missouri State alumnus Zachary Wilson.^{55,56} This was performed by calculating potential energy for the geometry optimized products ($U_{products}$) and the potential energy for the geometry optimized reactant ($U_{reactant}$), and subtracting the energy of the reactant from the energy of the products (Equation 6).

$$BDE = U_{products} - U_{reactant} \quad (6)$$

The mean-signed deviation (MSD, Equation 7), mean-unsigned deviation (MUSD, Equation 8), and root-mean-squared deviation (RMSD, Equation 9) were calculated for each table of reactions. n represents the number of reactions in the table, \widehat{BDE}_i represents the literature experimental bond dissociation energy of reaction i , and BDE_i represents the calculated bond dissociation energy of reaction i .^{55,57,58}

$$MSD = \frac{1}{n} \sum_{i=1}^n \widehat{BDE}_i - BDE_i \quad (7)$$

$$MUSD = \frac{1}{n} \sum_{i=1}^n |\widehat{BDE}_i - BDE_i| \quad (8)$$

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{BDE}_i - BDE_i)^2} \quad (9)$$

The BDE values for homolytic C-H bond cleavage of alkanes can be seen in Table 4, alkenes to form vinylic radicals in Table 5, and alkenes to form allylic radicals in Table 6. Additionally, Table 7 shows the BDE values of C-H bonds in alkenes to form homoallylic radicals.

Table 4. Bond dissociation energies for homolytic C-H bond cleavage of common alkanes.^{55,57,58}

Reactant	Product		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
Methane CH₄	CH₃ 	H [•]	104.9	11.1	12.5	176.9	107.4
Ethane 		H [•]	100.8	32.9	29.0	88.0	117.7
Propane 		H [•]	101.3	43.3	36.4	98.5	67.3
Propane 		H [•]	98.4	43.4	30.5	80.9	113.7
Butane 		H [•]	100.91	43.8	37.4	97.7	71.6
Butane 		H [•]	98.69	37.9	27.3	83.5	91.1
Pentane 		H [•]	97.98	36.9	32.3	92.8	77.6
Pentane 		H [•]	94.48	40.8	28.4	83.2	84.2
Hexane 		H [•]	95	39.7	28.0	83.1	57.1
Heptane 		H [•]	94.92	39.3	27.9	83.0	91.8

Table 5. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes leading to the formation of vinylic radicals.^{55,57,58}


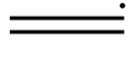

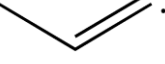
Reactant	Product		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
Ethylene 		H [•]	110.53	43.4	42.4	97.1	100.8
Propene 		H [•]	111.18	68.7	52.1	117.1	100.1

Table 6. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes leading to the formation of allylic radicals.^{55,57,58}

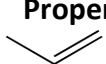
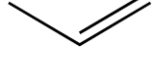
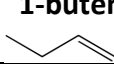
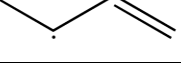
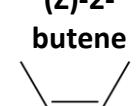
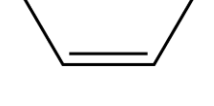
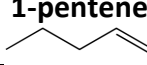

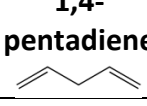
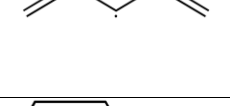
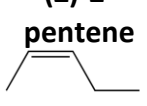
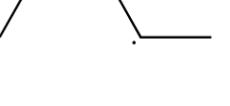
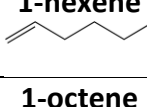

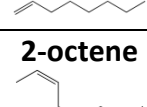
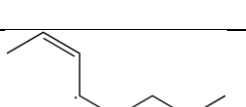




Reactant	Product		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
Propene 		H [•]	87.44	17.3	14.3	75.4	32.7
1-butene 		H [•]	82.74	11.6	10.5	70.2	54.9
(Z)-2-butene 		H [•]	87.35	16.9	14.8	73.4	26.1
1-pentene 		H [•]	82.95	12.1	11.0	71.4	30.0
1,4-pentadiene 		H [•]	77.28	4.3	3.8	63.7	-6.9
(Z)-2-pentene 		H [•]	81.6	15.4	14.1	70.5	36.7
1-hexene 		H [•]	83.6	12.1	11.2	71.6	6.4
1-octene 		H [•]	83.4	12.2	11.3	71.7	31.5
2-octene 		H [•]	81.9	16.6	15.5	71.6	28.8

Table 7. Bond dissociation energies for homolytic C-H bond cleavage of common alkenes leading to the formation of homoallylic radicals.^{55,57,58}

Reactant	Product		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
1-butene 		H•	98.1	44.1	37.2	98.3	75.7

The BDE values for homolytic C-C bond cleavage of alkanes can be seen in Table 8, values for a wide variety of C-C bond cleavage in alkenes in Table 9, and values for other relevant compounds (ethers) in Table 10.

Table 8. Bond dissociation energies for homolytic C-C bond cleavage of common alkanes.^{55,57,58}


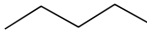
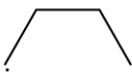
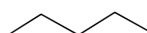
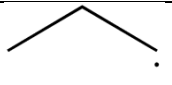


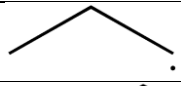
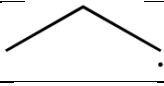
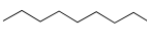
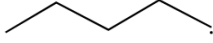
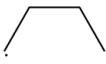
Reactant	Product		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
Ethane 	$\cdot\text{CH}_3$	$\cdot\text{CH}_3$	90.0	30.0	27.5	245.0	54.7
Pentane 		$\cdot\text{CH}_3$	88.7	64.5	54.4	167.0	-15.1
Pentane 			87.7	87.5	71.9	80.3	-16.3
Hexane 			88.1	97.8	79.5	90.5	-62.0
Nonane 			84.2	92.4	76.7	84.3	-6.6

Table 9. Bond dissociation energies for homolytic C-C bond cleavage of common alkenes.^{55,57,58}

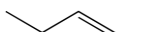








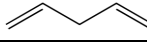


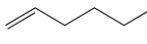


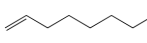





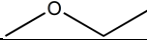


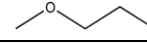


Reactant	Products		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
1-butene 			99.6	90.8	79.9	82.3	47.7
(Z)-2-butene 			100.6	91.3	70.1	188.1	32.7
(Z)-2-butene 			100.2	114.1	87.6	99.9	42.6
1,4-pentadiene 			85.7	74.5	64.9	69.1	-1.6
1-hexene 			74.3	71.4	56.9	67.2	-61.9
1-octene 			70.8	65.2	52.8	61.5	1.6

Table 10. Bond dissociation energies for homolytic bond cleavage of other relevant compounds.^{55,57,58}

Reactant	Products		BDE (kcal/mol)				
	1	2	Literature	ANI-1x	ANI-1ccx	ANI-2x	ReaxFF
Dimethyl Ether 			84.1	40.8	35.2	150.2	52.3
Methoxypropane 			86.1	77.7	63.6	76.0	12.6
Methoxybutane 			84.4	41.5	35.7	152.2	45.6

The mean unsigned deviation, mean signed deviation, root-mean squared deviation, and percent error from the previous tables can be seen in Table 11, Table 12, Table 13, and Table 14, respectively.

Table 11. Mean unsigned deviation for each potential.^{55,57,58}

MUSD (kcal/mol)								
Potential	Overall	Alkane C-H	Vinylic =C-H	Allylic C-H	Homo-allylic C-H	Alkane C-C	Mixed Alkene C-C	C-X
ANI-1x	46.1	61.8	54.8	70.0	54.0	20.5	8.6	31.5
ANI-1ccx	52.6	69.8	63.5	71.3	60.8	25.7	19.8	40.0
ANI-2x	22.7	16.4	9.67	12.1	0.175	48.7	23.0	48.0
ReaxFF	50.7	17.7	10.4	56.4	22.4	96.8	78.3	48.0

Table 12. Mean signed deviation for each potential.^{55,57,58}

MSD (kcal/mol)								
Potential	Overall	Alkane C-H	Vinylic =C-H	Allylic C-H	Homo-allylic C-H	Alkane C-C	Mixed Alkene C-C	C-X
ANI-1x	-44.3	-61.8	-54.8	-70.0	-54.0	-13.3	-4.0	-31.5
ANI-1ccx	-52.6	-69.8	-63.5	-71.3	-60.8	-25.7	-19.8	-40.0
ANI-2x	7.0	-2.0	-3.73	-12.07	0.175	45.7	6.2	41.3
ReaxFF	-48.8	-10.8	-10.42	-56.4	-22.4	-96.8	-78.3	-48.0

Table 13. Root-mean-squared deviation for each potential.^{55,57,58}

RMSD (kcal/mol)								
Potential	Overall	Alkane C-H	Vinylic =C-H	Allylic C-H	Homo- allylic C-H	Alkane C-C	Mixed Alkene C-C	C-X
ANI-1x	106.2	125.7	112.3	140.0	108.0	59.0	18.7	71.0
ANI-1ccx	115.5	140.4	127.4	142.7	121.7	66.0	40.1	83.8
ANI-2x	78.6	50.4	20.7	24.2	0.351	155.5	74.7	110.0
ReaxFF	124.4	42.7	20.9	117.2	44.8	207.1	166.4	102.8

Table 14. Percent error values for each potential.^{55,57,58}

Percent Error (%)								
Potential	Overall	Alkane C-H	Vinylic =C-H	Allylic C-H	Homo- allylic C-H	Alkane C-C	Mixed Alkene C-C	C-X
ANI-1x	50.6	62.4	49.5	84.3	55.0	23.0	9.5	37.3
ANI-1ccx	57.7	70.6	57.3	85.9	62.0	28.9	22.6	47.3
ANI-2x	25.0	16.3	8.7	14.5	0.2	54.4	24.4	56.9
ReaxFF	58.8	18.0	9.4	68.2	22.8	110.6	93.3	56.4

CONCLUSION

In an effort to develop tools to help us better understand the pathways of the pyrolysis of triglycerides derived from soybean oil, molecular dynamics was employed in the form of Langevin dynamics. ANAKIN-ME: Accurate Neural network engine for Molecular Energies (ANI-2x) was used as the interatomic potential. To be able to validate the results from the molecular dynamics runs, various data representation techniques were employed to be contrasted with BDEs of bonds in oleic and linoleic acid determined by DFT calculations in previous work. Literature BDE values determined by experiment from various bonds in hydrocarbons were also contrasted with BDEs calculated using various potentials including ANI-2x.⁵⁵ The molecular dynamics results were also contrasted with the proposed pathways, found by mass spectrometry in experimental work.²³ ANI has been recognized as a step toward a “universal” transferable potential. However, ANI was trained to predict absolute electronic energies, not bond dissociation energies (as required herein). This work shows that ANI somewhat agrees with the expected outcome (R^2 value ranging from near 0 to 0.4712 in Table 3).^{TABLE 3} The figures depicting the relationship between initial breakage counts and BDEs for oleic and linoleic acid (Figure 4.17, Figure 4.18, Figure 4.27, and Figure 4.28) show inverse correlation with the negative slopes seen in Table 3. However, bonds near the carbonyl (0-1, 1-2, and 1-3) deviate substantially from this trend. This could be caused by breakages near the glycerol backbone being influenced by events within glycerol that happened before the “initial events” that are presented in this work. For example, reaction 1a and reaction 2a in Figure 1.8 suggest that CO₂ formation is one of the two initial pathways. Due to this, an assumption can be made that the formation of CO₂ causes the breakage of 1-3 to be a much more favorable breakage if the ester-

glycerol bond was broken, versus a completely intact triglyceride or monoacylglycerol. With this limit acknowledged, ANI-2x appears to incorporate the gross expectations determined by BDE calculations and thermodynamic arguments.

Due to the focus on FA reaction pathways, glycerol pathways were ignored. In future work, events should either consider the carbon in which the fatty acid is connected by an ester linkage or the glycerol backbone altogether. Following this, it would be beneficial to quantify final products produced categorized by fatty acid residue. Other future work to be done includes incorporating interatomic potentials to compare against ANI-2x. One ReaxFF version, initially used for combustion of an Illinois No. 6 coal char, may be a promising interatomic potential in the pyrolysis of triglycerides.^{48,59} The usage of LAMMPS (Large-scale Atomic-Molecular Massively Parallel Simulator) instead of ASE in the future may be beneficial for ease of use with other interatomic potentials such as the previously mentioned ReaxFF potential.⁶⁰

REFERENCES

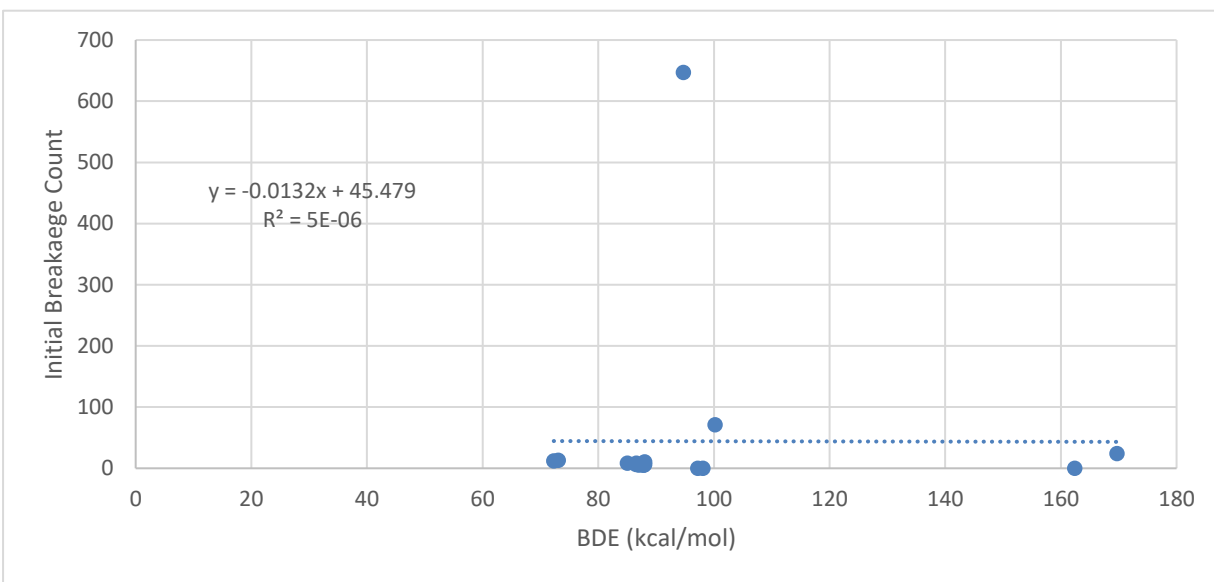
- (1) Martins, F.; Felgueiras, C.; Smitkova, M.; Caetano, N. Analysis of Fossil Fuel Energy Consumption and Environmental Impacts in European Countries. *Energies* **2019**, *12* (6), 964. <https://doi.org/10.3390/en12060964>.
- (2) Shafiee, S.; Topal, E. When Will Fossil Fuel Reserves Be Diminished? *Energy Policy* **2009**, *37* (1), 181–189. <https://doi.org/10.1016/j.enpol.2008.08.016>.
- (3) *Total Energy Annual Data - U.S. Energy Information Administration (EIA)*. <https://www.eia.gov/totalenergy/data/annual/index.php>.
- (4) *Consumer Price Index (CPI) Databases : U.S. Bureau of Labor Statistics*. <https://www.bls.gov/cpi/data.htm> (accessed 2023-11-08).
- (5) Kapustin, N. O.; Grushevenko, D. A. Long-Term Electric Vehicles Outlook and Their Potential Impact on Electric Grid. *Energy Policy* **2020**, *137*, 111103. <https://doi.org/10.1016/j.enpol.2019.111103>.
- (6) Crabtree, G. The Coming Electric Vehicle Transformation. *Science* **2019**, *366* (6464), 422–424. <https://doi.org/10.1126/science.aax0704>.
- (7) Mai, T. T.; Jadun, P.; Logan, J. S.; McMillan, C. A.; Muratori, M.; Steinberg, D. C.; Vimmerstedt, L. J.; Haley, B.; Jones, R.; Nelson, B. *Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States*; NREL/TP--6A20-71500, 1459351; 2018; p NREL/TP--6A20-71500, 1459351. <https://doi.org/10.2172/1459351>.
- (8) *Electricity in the U.S. - U.S. Energy Information Administration (EIA)*. <https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php> (accessed 2023-12-06).
- (9) US EPA, O. *Electric Power Sector Basics*. <https://www.epa.gov/power-sector/electric-power-sector-basics> (accessed 2024-02-02).
- (10) Mayyas, A.; Steward, D.; Mann, M. The Case for Recycling: Overview and Challenges in the Material Supply Chain for Automotive Li-Ion Batteries. *Sustain. Mater. Technol.* **2019**, *19*, e00087. <https://doi.org/10.1016/j.susmat.2018.e00087>.
- (11) Barbé, A. Emerging International Trade Issues for Fossil Fuel. *U. S. Int. Trade Comm.* **2015**.
- (12) Maisel, F.; Neef, C.; Marscheider-Weidemann, F.; Nissen, N. F. A Forecast on Future Raw Material Demand and Recycling Potential of Lithium-Ion Batteries in Electric Vehicles. *Resour. Conserv. Recycl.* **2023**, *192*, 106920. <https://doi.org/10.1016/j.resconrec.2023.106920>.
- (13) McCormick, R.; Moriarty, K. *Biodiesel Handling and Use Guide: Sixth Edition*; NREL/TP-4A00-86939, 2001221, MainId:87714; 2023; p NREL/TP-4A00-86939, 2001221, MainId:87714. <https://doi.org/10.2172/2001221>.
- (14) Moser, B. R. Biodiesel Production, Properties, and Feedstocks. In *Biofuels*; Tomes, D., Lakshmanan, P., Songstad, D., Eds.; Springer New York: New York, NY, 2011; pp 285–347. https://doi.org/10.1007/978-1-4419-7145-6_15.
- (15) Lapuerta, M.; Armas, O.; Rodriguezfernandez, J. Effect of Biodiesel Fuels on Diesel Engine Emissions. *Prog. Energy Combust. Sci.* **2008**, *34* (2), 198–223. <https://doi.org/10.1016/j.pecs.2007.07.001>.

- (16)Huth, M.; Heilos, A. Fuel Flexibility in Gas Turbine Systems: Impact on Burner Design and Performance. In *Modern Gas Turbine Systems*; Elsevier, 2013; pp 635–684.
<https://doi.org/10.1533/9780857096067.3.635>.
- (17)Demirbas, A. Recent Developments in Biodiesel Fuels. *Int. J. Green Energy* **2007**, *4* (1), 15–26. <https://doi.org/10.1080/15435070601015395>.
- (18)Karaosmanog, F. Biofuel Production Using Slow Pyrolysis of the Straw and Stalk of the Rapeseed Plant. **1999**.
- (19)Duffy, P. XXV.—On the Constitution of Stearine. *Q. J. Chem. Soc. Lond.* **1853**, *5* (4), 303–316.
<https://doi.org/10.1039/QJ8530500303>.
- (20)Alsultan, A. G.; Asikin Mijan, N.; Mansir, N.; Razali, S. Z.; Yunus, R.; Taufiq-Yap, Y. H. Combustion and Emission Performance of CO/NO_x/SO_x for Green Diesel Blends in a Swirl Burner. *ACS Omega* **2021**, *6* (1), 408–415. <https://doi.org/10.1021/acsomega.0c04800>.
- (21)National Renewable Energy Laboratory (NREL), Golden, CO. *Characterization of Biodiesel Oxidation and Oxidation Products*; NREL/TP-540-39096, 909194; 2005; p NREL/TP-540-39096, 909194. <https://doi.org/10.2172/909194>.
- (22)Abed, K. A.; Gad, M. S.; El Morsi, A. K.; Sayed, M. M.; Elyazeed, S. A. Effect of Biodiesel Fuels on Diesel Engine Emissions. *Egypt. J. Pet.* **2019**, *28* (2), 183–188.
<https://doi.org/10.1016/j.ejpe.2019.03.001>.
- (23)Kubátová, A.; Luo, Y.; Šťávová, J.; Sadrameli, S. M.; Aulich, T.; Kozliak, E.; Seames, W. New Path in the Thermal Cracking of Triacylglycerols (Canola and Soybean Oil). *Fuel* **2011**, *90* (8), 2598–2608. <https://doi.org/10.1016/j.fuel.2011.04.022>.
- (24)Matbouei, M.; Weston, D. P.; Liang, X.; Hainsworth, S. V. An Investigation of the Effect of Temperature on the Oxidation Processes of Metallic Diesel Engine Fuel System Materials and B100 Biodiesel from Used Cooking Oil in Exposure Testing. *Fuel Lond. Engl.* **2021**, *285*, 119063.
<https://doi.org/10.1016/j.fuel.2020.119063>.
- (25)Graboski, M. S.; McCormick, R. L. Combustion of Fat and Vegetable Oil Derived Fuels in Diesel Engines. *Prog. Energy Combust. Sci.* **1998**, *24* (2), 125–164.
- (26)Lotero, E.; Liu, Y.; Lopez, D. E.; Suwannakarn, K.; Bruce, D. A.; Goodwin, J. G. Synthesis of Biodiesel via Acid Catalysis. *Ind. Eng. Chem. Res.* **2005**, *44* (14), 5353–5363.
<https://doi.org/10.1021/ie049157g>.
- (27)Haas, M. J. Improving the Economics of Biodiesel Production through the Use of Low Value Lipids as Feedstocks: Vegetable Oil Soapstock. *Fuel Process. Technol.* **2005**, *86* (10), 1087–1096.
<https://doi.org/10.1016/j.fuproc.2004.11.004>.
- (28)Maher, K. D.; Bressler, D. C. Pyrolysis of Triglyceride Materials for the Production of Renewable Fuels and Chemicals. *Bioresour. Technol.* **2007**, *98* (12), 2351–2368.
<https://doi.org/10.1016/j.biortech.2006.10.025>.
- (29)Kubičková, I.; Snáře, M.; Eränen, K.; Mäki-Arvela, P.; Murzin, D. Yu. Hydrocarbons for Diesel Fuel via Decarboxylation of Vegetable Oils. *Catal. Today* **2005**, *106* (1), 197–200.
<https://doi.org/10.1016/j.cattod.2005.07.188>.
- (30)Bridgwater, A. V.; Peacocke, G. V. C. Fast Pyrolysis Processes for Biomass. *Renew. Sustain. Energy Rev.* **2000**, *4* (1), 1–73. [https://doi.org/10.1016/S1364-0321\(99\)00007-6](https://doi.org/10.1016/S1364-0321(99)00007-6).
- (31)*Crude oil distillation and the definition of refinery capacity*.
<https://www.eia.gov/todayinenergy/detail.php?id=6970> (accessed 2023-11-07).

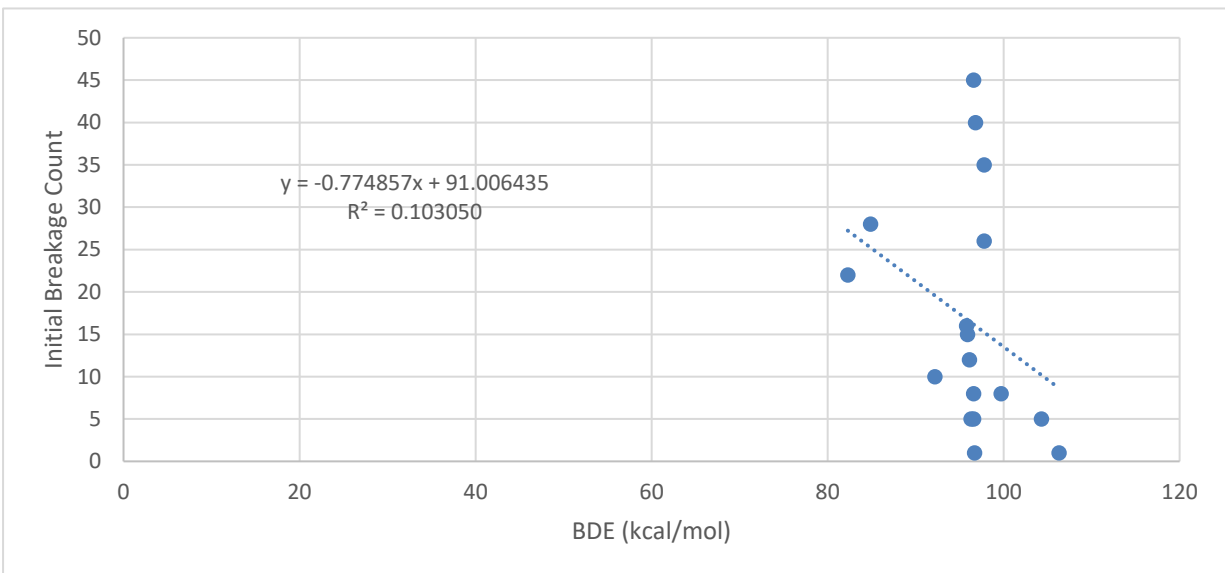
- (32) Crynes, B. L.; Albright, L. F.; Tan, L.-F. Thermal Cracking. In *Encyclopedia of Physical Science and Technology*; Elsevier, 2003; pp 613–626. <https://doi.org/10.1016/B0-12-227410-5/00153-8>.
- (33) *File:Crude Oil Distillation.png - Wikipedia*.
https://commons.wikimedia.org/wiki/File:Crude_Oil_Distillation.png (accessed 2023-11-07).
- (34) Haile, J. M. *Molecular Dynamics Simulation Elementary Methods*; John Wiley & Sons, Inc, 1997.
- (35) Frenkel, D.; Smit, B. Molecular Dynamics Simulations Chapter 4. In *Understanding Molecular Simulation*; Elsevier, 2023; pp 97–124. <https://doi.org/10.1016/B978-0-32-390292-2.00012-X>.
- (36) Paquet, E.; Viktor, H. L. Computational Methods for Ab Initio Molecular Dynamics. *Adv. Chem.* **2018**, *2018*, 1–14. <https://doi.org/10.1155/2018/9839641>.
- (37) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203. <https://doi.org/10.1039/C6SC05720A>.
- (38) *Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens*. <https://doi.org/10.1021/acs.jctc.0c00121>.
- (39) Rudin, A.; Choi, P. Diffusion in Polymers. In *The Elements of Polymer Science & Engineering*; Elsevier, 2013; pp 275–304. <https://doi.org/10.1016/B978-0-12-382178-2.00006-7>.
- (40) *Atomic Simulation Environment — ASE documentation*. <https://wiki.fysik.dtu.dk/ase/#> (accessed 2024-01-01).
- (41) Schlick, T. *Molecular Modeling and Simulation : An Interdisciplinary Guide*; New York : Springer, 2002.
- (42) Ackland, G. J.; Bonny, G. Interatomic Potential Development. In *Comprehensive Nuclear Materials*; Elsevier, 2020; pp 544–572. <https://doi.org/10.1016/B978-0-12-803581-8.11687-X>.
- (43) Morse, P. M. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Phys. Rev.* **1929**, *34* (1), 57–64. <https://doi.org/10.1103/PhysRev.34.57>.
- (44) Lennard-Jones, J. E. Cohesion. *Proc. Phys. Soc.* **1931**, *43* (5), 461–482. <https://doi.org/10.1088/0959-5309/43/5/301>.
- (45) Chemistry (IUPAC), T. I. U. of P. and A. *IUPAC - potential-energy (reaction) surface (P04780)*. <https://doi.org/10.1351/goldbook.P04780>.
- (46) Potential Energy Surface. *Wikipedia*; 2023.
- (47) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401. <https://doi.org/10.1103/PhysRevLett.98.146401>.
- (48) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105* (41), 9396–9409. <https://doi.org/10.1021/jp004368u>.
- (49) Kuhlman, D. A Python Book: Beginning Python, Advanced Python, and Python Exercises.
- (50) Perkel, J. M. Why Jupyter Is Data Scientists' Computational Notebook of Choice. *Nature* **2018**, *563* (7729), 145–146. <https://doi.org/10.1038/d41586-018-07196-1>.
- (51) *Package overview — pandas 2.1.4 documentation*. https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html (accessed 2024-01-01).
- (52) James, C. A. OpenSMILES Specification, 2016. www.opensmiles.org (accessed 2024-01-02).

- (53) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3* (1), 33. <https://doi.org/10.1186/1758-2946-3-33>.
- (54) *Poisson and Binomial Distributions*. https://www.uvm.edu/~statdhtx/StatPages/More_Stuff/PoissonBinomial/PoissonBinom.html (accessed 2024-03-29).
- (55) Wilson, Z. R.; Siebert, M. R. Methyl Linoleate and Methyl Oleate Bond Dissociation Energies: Electronic Structure Fishing for Wise Crack Products. *Energy Fuels* **2018**, *32* (2), 1779–1787. <https://doi.org/10.1021/acs.energyfuels.7b02798>.
- (56) Wilson, Z. R. Ab Initio Methyl Linoleate Bond Dissociation Energies: First Principles Fishing for Wise Crack Products.
- (57) *Active Thermochemical Tables version 1.130*. <https://atct.anl.gov/Thermochemical%20Data/version%201.130/index.php> (accessed 2024-02-06).
- (58) Lossing, F. P.; Maccoll, A. Free radicals by mass spectrometry, XLVII. Ionization potentials and ionic heats of formation of C₅–C₇ alkyl radicals. *Can. J. Chem.* **1976**, *54* (6), 990–992. <https://doi.org/10.1139/v76-142>.
- (59) Castro-Marcano, F.; Kamat, A. M.; Russo, M. F.; Van Duin, A. C. T.; Mathews, J. P. Combustion of an Illinois No. 6 Coal Char Simulated Using an Atomistic Char Representation and the ReaxFF Reactive Force Field. *Combust. Flame* **2012**, *159* (3), 1272–1285. <https://doi.org/10.1016/j.combustflame.2011.10.022>.
- (60) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In 'T Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171. <https://doi.org/10.1016/j.cpc.2021.108171>.

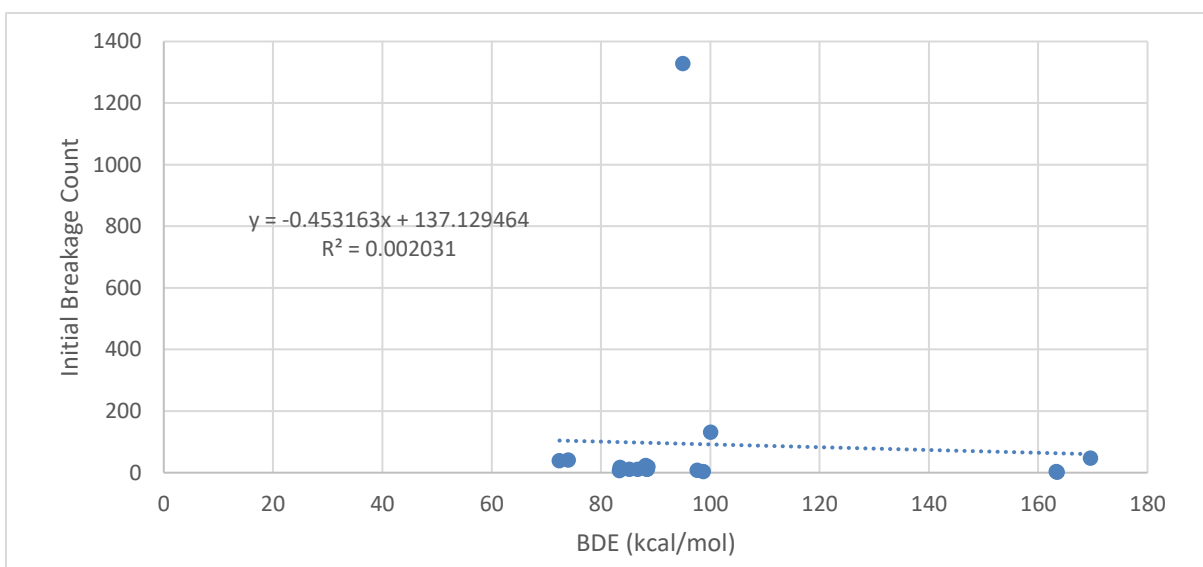
APPENDIX: SCATTER PLOTS OF INITIAL BREAKAGE COUNT VERSUS BDE



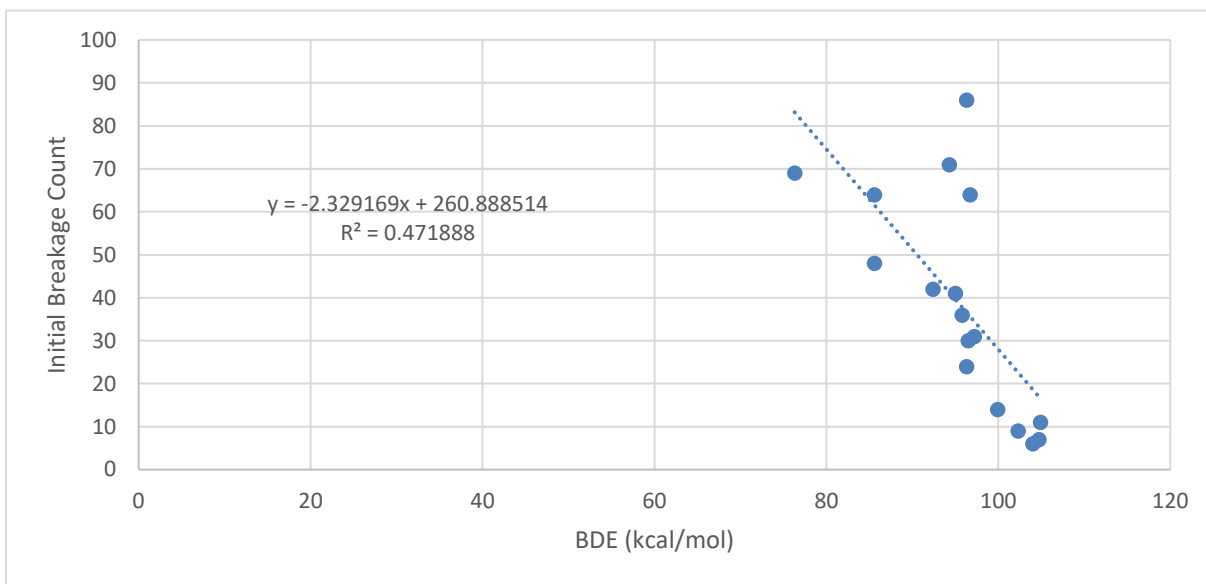
Scatter plot representing initial breakage counts versus BDE (kcal/mol) of heavy-heavy bonds seen in oleic acid, with equation shown in Table 3.



Scatter plot representing initial breakage counts versus BDE (kcal/mol) of hydrogen-heavy bonds seen in oleic acid, with equation shown in Table 3.



Scatter plot representing initial breakage counts versus BDE (kcal/mol) of heavy-heavy bonds seen in linoleic acid, with equation shown in Table 3.



Scatter plot representing initial breakage counts versus BDE (kcal/mol) of hydrogen-heavy bonds seen in linoleic acid, with equation shown in Table 3.