

Generative Algorithm for Molecular Graphs Uncovers Products of Oil Oxidation

Yuliia Orlova, Alessa A. Gambardella, Ivan Kryven,* Katrien Keune, and Piet D. Iedema



Cite This: *J. Chem. Inf. Model.* 2021, 61, 1457–1469



Read Online

ACCESS |



Metrics & More

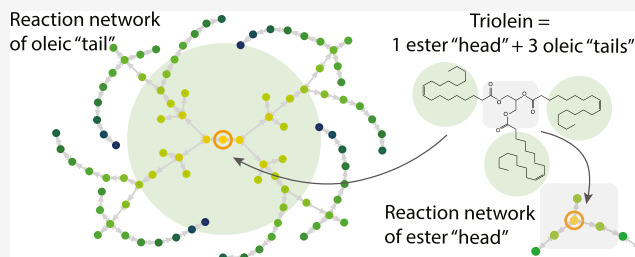


Article Recommendations



Supporting Information

ABSTRACT: The autoxidation of triglyceride (or triacylglycerol, TAG) is a poorly understood complex system. It is known from mass spectrometry measurements that, although initiated by a single molecule, this system involves an abundance of intermediate species and a complex network of reactions. For this reason, the attribution of the mass peaks to exact molecular structures is difficult without additional information about the system. We provide such information using a graph theory-based algorithm. Our algorithm performs an automatic discovery of the chemical reaction network that is responsible for the complexity of the mass spectra in drying oils. This knowledge is then applied to match experimentally measured mass spectra with computationally predicted molecular graphs. We demonstrate this methodology on the autoxidation of triolein as measured by electrospray ionization-mass spectrometry (ESI-MS). Our protocol can be readily applied to investigate other oils and their mixtures.



experimentally measured mass spectra with computationally predicted molecular graphs. We demonstrate this methodology on the autoxidation of triolein as measured by electrospray ionization-mass spectrometry (ESI-MS). Our protocol can be readily applied to investigate other oils and their mixtures.

INTRODUCTION

Modern tools in the field of analytical chemistry are progressively increasing their resolution and are able to generate a manifold of physical and chemical data for the characterization of functional materials.^{1,2} The interpretation of such data is becoming a challenge, wherein computational modeling plays a role of growing importance. Hence, the models are becoming of considerable complexity themselves, which challenges their implementation and provides an opportunity for their automated construction. Consequently, the emergent models exploit the benefits of experiments coupled with computational models to understand complex chemical systems. One significant application in this context is the autoxidation of triglyceride, which is relevant in food chemistry,^{3,4} cosmetics,⁵ biofuels,⁶ and cultural heritage.^{7–10}

Triglycerides are formed from one glycerol molecule and three fatty acids (palmitic, oleic, linoleic, linolenic) in various ratios. Fatty acid chains may differ by the number of unsaturations, which are responsible for a chain of autoxidation reactions that lead to the formation of irregular macromolecular structures. The resulting macromolecules are composed of a wide range of monomer units bearing multiple types and numbers of cross-links. Meanwhile, the esters between the glycerol and fatty acids undergo hydrolysis releasing free fatty acids into the system. Hydrolysis, along with oxidative scission and peroxide decomposition, is responsible for the eventual breakdown of macromolecules.

Several experimental studies of oil autoxidation analyze the changes of relative amounts of functional groups in oil samples. Muizebelt et al.¹¹ reported the evolution of various functional groups, such as bis-allylic carbons, conjugated double bonds,

epoxides, acids, and cross-links, which characterize the drying process of ethyl linoleate. Work by Oakley et al.¹² and Fjällström et al.¹³ reported volatile aldehydes emission during the autoxidation of linoleic acid. Drier influence on autoxidation of oils was studied by Oyman et al.,^{14,15} van Gorkum et al.,¹⁶ Bouwman et al.,¹⁷ Mallégol et al.,¹⁸ and Spier et al.¹⁹ On the theoretical side, the experimental studies are supported by the detailed exploration of reaction mechanisms utilizing quantum mechanics calculations.^{12,20} In addition, studies by Oakley et al. and Iedema et al.^{21–23} present detailed mathematical models, which describe the concentration profiles of functional groups characteristic to the drying of ethyl linoleate. Both interpretation of experimental measurements and computational modeling techniques will benefit from knowledge of the explicit reaction mechanism describing all intermediate and product molecular species.

Mass spectrometry (MS) provides more detailed information about a measured sample than spectroscopic techniques, such as Fourier-transform infrared spectroscopy (FTIR). Generating a wealth of information about the molecular species present in samples, direct MS measurements used in autoxidation of oils are often difficult to interpret due to: (1) fragmentation of the molecules that occurs during the measurement and (2) the

Received: October 6, 2020

Published: February 22, 2021



possibility for multiple species of equal or very similar molecular masses to contribute to the same peak.^{24,25} These two points increase the complexity of peak assignments, although the fragmentation of the molecules may be mitigated in some instances with soft ionization [e.g., electrospray ionization-mass spectrometry (ESI-MS)].^{26,27} Although the peak assignment task has been assisted with tandem MS^{26,27} or application of the Kendrick mass defect technique,^{28,29} high time consumption, methodological challenges, and unanswered questions still remain. Therefore, direct MS interpretation can benefit from the assistance of computational modeling. Grimme et al.^{30,31} describe how to handle fragmentation of a single small molecule by simulating the conditions inside the instrument using quantum-chemical calculations. However, this approach is not yet applicable in the case of large or complex molecules, as the spectrum of triglyceride simultaneously features peaks from both intermediate and final products (e.g., of the autoxidation reaction mechanism).

Understanding the aging of oil, both computationally and experimentally, requires choosing an appropriate model system that is simpler than linseed oil, while it also allows for the complexities to be added later. We choose triolein (i.e., glycerol trioleate), a triacylglycerol (TAG) that is formed from three symmetrical oleic esters and a glycerol molecule. The shape of the triolein molecule resembles the triglyceride structure of linseed oil, while monounsaturated fatty esters make triolein an ideal model system of olive oil. The reaction scheme of triolein has a large number of intermediate species but only a small number of reaction types. Holding significance for food lipid oxidation studies, triolein thus will serve as a practical benchmark for the model development. The reaction mechanism therefore still produces a significant number of possible extractable products that can be measured with direct MS. With such systems, a global picture is obtained of the reaction processes and changes that happen in the measured sample.

To identify a relevant set of possible molecular structures corresponding to peaks in mass spectra following triolein autoxidation, we turn to the automated reaction network generation (ARNG).^{32–34} The advent of the ARNG inspired numerous graph theory-based approaches to study reaction pathways of chemical mechanisms^{35–38} and retrosynthesis.³⁹ The ARNG method recursively applies all chemically relevant transformations to all molecular species, which are represented as molecular graphs. However, using this algorithmic discovery of reaction steps on large molecules becomes prohibitive because the algorithm is NP-hard.⁴⁰ To overcome this problem, we have formulated ARNG for random graphs in Orlova et al.,³⁴ and, in this paper, extended this method further to many cross-link types. Because of our approach, a large irregular macromolecule is viewed as an ensemble of randomly interconnected monomers, which have precise structures and are represented as molecular graphs. In the case of triolein, the macromolecule is broken up into smaller pieces: glycerol with adjacent esters and three oleic chains. Therefore, the computational complexity of the ARNG method is considerably reduced to a large but tractable size. For the purpose of this paper, the result of this ARNG application consists of all chemically distinguishable structures of glycerol and derivatives of oleic esters that result from hydrolysis, oxidation, and β scission reactions. As triolein is not a drying oil, this study focuses solely on oxidative reactions, which are predominant in triolein. Thus, in this paper, the cross-linking reactions are not part of ARNG implementation.

On the experimental side, masses following the gently accelerated drying of triolein in the presence of inert titanium dioxide (rutile form) are measured with the use of electrospray ionization-mass spectrometry (ESI-MS). In contrast to previous reports on the MS of oils, where peak assignments were performed manually,⁴¹ this work matches the products generated by the adapted ARNG to the complex mass spectra. This paper is based on Chapter 4 of PhD Thesis.⁴²

METHODS

Experimental Setup. Paint Model Preparation and Aging. Paint models were prepared using inorganically coated titanium dioxide (rutile form) pigment (Tronox CR-826), which was chosen for its chemical inertness toward linseed oil binding medium compared to other pigments.⁴³ On an automatic muller comprising two glass plates, triolein (Sigma-Aldrich, glyceryl trioleate $\geq 97.0\%$ (TLC), CAS No. 122-32-7) was added to the pigment (37.6% w/w in pigment), and both were gently mixed together with a palette knife. The paint was then mixed between the two glass plates by rotating the top plate 25 turns (under a weight of 10 kg) and, at completion, consolidated at the center of the bottom plate using the palette knife; this step was then repeated once. Mixed paint was deposited on a Melinex polyester wrap (thickness, 250 μm , #M026 from ref 44) and drawn into a thin strip with a drawdown bar (height, 25 μm). The paint film was set to dry flat overnight under lab conditions for ca. 24 h. A snipping was removed and placed in the dark in an acid-free archive box (type C, blue gray, Verenigde Bedrijven, Jansen-Wijsmuller & Beuns B.V.) and left under lab conditions until analysis. The remainder of the film was heated in an oven (preheated) at 70 °C without light for 5 days and, at removal, also added to an acid-free archive box until analysis. In all instances, air was free to circulate and light was only present briefly (< 5 min per sample) during sample handling. All analyses were performed less than a week after completion of heating.

Electrospray Ionization-Mass Spectrometry (ESI-MS). Direct-injection ESI-MS measurements were performed in both positive and negative modes on solvent extractions from the paint films using a Micromass Q-tof-2 equipped with a nanospray and ESI source. For all solutions and cleaning, ethanol (Fisher Scientific, CHROMASOLV, absolute, for HPLC) was used. Extractions were prepared using 0.1–0.5 mg scrapings from the paint film (through the entire thickness) dissolved in ethanol (50 μL) for 1 h, mixed via a vortex for 20 s, and centrifuged (8 G) for 7 min. Following centrifugation to precipitate pigment particles, the supernatant liquid was removed and mixed 1:1 with 20 mM ammonium acetate in ethanol.^{41,45} All solutions were handled using a glass syringe and rinsed 10 x with ethanol between samples. The samples were delivered to the Micromass instrument via a Micromass CapLC system using 10 mM ammonium acetate in ethanol as an eluent after calibrating with 0.1% phosphoric acid solution (50:50, DI water/acetonitrile).

Data Processing, Peak Picking, and Matching. Data were collected and processed with MassLynx 4.0 software (Waters). Using standard scientific packages in Python (3.6.5),^{46–49} experimental peak positions were then determined applying a maximum filter to the mass spectra (smoothed via a Gaussian filter). Experimental peak positions were compared to exact masses for the most abundant isotope of each computationally determined structure adjusted by the exact mass of each possible ion. The calculations were performed using the exact mass

calculator of the MS Online Tools of Scientific Instrument Services (<https://www.sisweb.com/referenc/tools/exactmass.htm>)⁵⁰ with addition of ammonium, hydrogen, or sodium for positive mode and addition of acetate or subtraction of hydrogen for negative mode. A match between experimental peak positions and computational masses was accepted for pairs of peaks within a tolerance of 0.15 m/z , smaller than half the distance between two isotopic peaks.

ARNG. ARNG is an algorithm that automatically discovers reaction pathways. The algorithm starts with an initial set of molecules defined in terms of molecular graphs and a predefined list of reaction templates. The first step is to find the reactive functional group in the initial molecule and apply a corresponding transformation. According to the reaction templates, the functional group of the product comes in the place of the reactive functional group. In this way, a new configuration of each molecule is created. Such new molecules may undergo further transformations. The algorithm stops when no new structures can be produced. The main steps are illustrated in Figure 1.

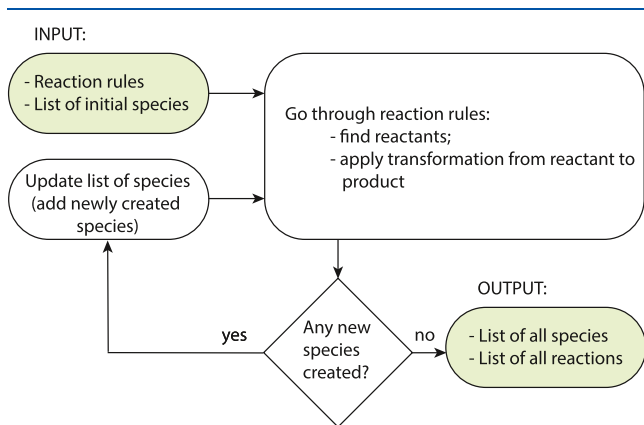


Figure 1. Block diagram representing the main steps of the ARNG.

RESULTS AND DISCUSSION

Representation of a Macromolecule. Since molecular structures derived from triolein can bear several functional groups, manually writing down all of the reaction steps and intermediate products is infeasible due to the large number of combinations in which multifunctional molecules may interact. To set up an automated exploration process yielding such reaction steps,³⁴ we will first formalize the structures of irregular macromolecules so that they become palatable to a computer algorithm. As illustrated in Figure 2a, we distinguish two hierarchical levels: the macromolecule level (Level 1) and the fragment level (Level 2). When applied to the triolein system, the meaning of this description is illustrated in Figure 2b. Namely:

- Level 1: The triolein molecule is represented as a random graph^{51–53} with different types of nodes (three oleic chains and glycerol unit) and ester bonds as edges (connections between oleic chains and glycerol unit). The random graph representation enables the analysis of the properties of large molecules by studying their fragments, which, in turn, obey the connectivity statistics. At this level, “head” and “tail” atoms are introduced, indicating that the oleic chains and glycerol unit are studied

separately, even though they are connected via the head–tail pair.

- Level 2: This level describes molecular graphs of the fatty acid chains and the glycerol unit with adjacent esters molecules. On this level, each node of the molecular graph represents an atom. Precise knowledge of the molecular graphs is necessary for the ARNG to decide which chemical compounds can react and what the products of such reactions could be.

The above-introduced levels are the key concepts for adapting the ARNG to handle large molecules, such as triolein. The algorithm acts separately on molecular graphs of the molecular substructure formed by glycerol with adjacent esters and the substructure formed by oleic chains (Level 2). It reconstructs all possible transformations of these two substructures as they follow hydrolysis and autoxidation reactions, respectively. Then, all possible triolein-derived molecules (Level 1) are reconstructed by joining atoms corresponding to “heads” and “tails” between all configurations of the glycerol part of the molecule after hydrolysis and all states of the three oleic chains after autoxidation reactions. Cross-linking reactions between oleic chains are not considered in this work and will additionally increase the complexity of the resulting macromolecules.

Automatic Discovery of Reactions for Triolein. The organic chemistry of the complete triolein molecule is manually intractable, but at the level of functional groups, the chemistry allows easy manual treatment. We singled out 5 small molecules (oxygen, hydroxide, water, initiator, and initiator radical) and 18 functional groups that are the minimum requirement to describe the autoxidation of the whole system (see Figure 2b). We further describe the interactions between the functional groups and small molecules by 22 reaction templates, as discussed below. A reaction template is a transformation that maps a functional group (subgraph) of a reactant molecule to a functional group of a product molecule. The functional groups and the reaction templates (see Supporting Information Figures S1 and S2 for the complete list) are carefully formulated using known reaction steps from oil autoxidation literature^{54–56} (see Figure 3), which are summarized in the paragraph below. The validity of implemented reaction steps is assessed by the ability of the algorithm to form all experimentally reported functional groups, which are typical for the oil oxidation. The reaction templates are formulated using functional groups from Figure 2b and several auxiliary molecular substructures (see Supporting Information Figure S3) that were defined to efficiently represent some reaction products in the ARNG setup. In the current work, the reaction template formulation is done manually; however, there exist methods for the automatic extraction of reaction templates from chemical databases in the context of synthesis planning.^{57–59} This can be considered as a promising enrichment of the ARNG methodology in future.

The diversity of the intermediate and product species derived from triolein is caused by three main reaction pathways: hydrolysis, oxidation of unsaturated double bond, and β scission of the hydrocarbon chains.^{60,61} Under the influence of water, hydrolysis of ester bonds of triglyceride (triacylglycerol, TAG) results in the formation of free fatty acids, diglycerides (diacylglycerol, DAGs), and monoglycerides (monoacylglycerol, MAGs) (see Supporting Information Figure S4). The autoxidation process is initiated at the double bonds of the unsaturated carbon chains. A hydrogen is abstracted from an allylic carbon on the oleic acid tail, forming an alkyl radical that

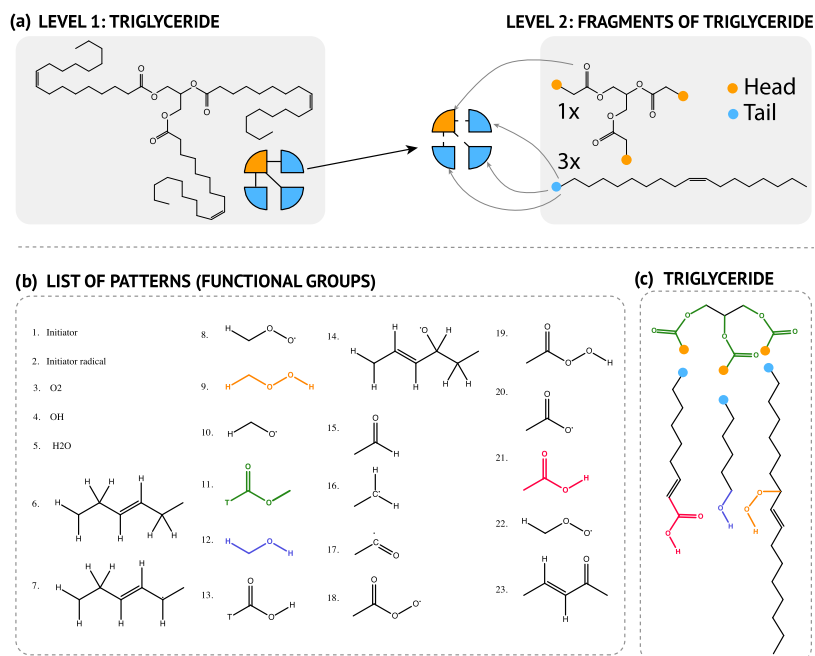


Figure 2. (a) Fragmented representation of the triolein molecule in two levels: Level 1, a large molecule (triolein) is represented as a random graph with edges being ester bonds; Level 2, molecular graphs of glycerol with adjacent esters and three oleic chains, edges are covalent bonds. (b) Functional groups necessary to define the reaction mechanism of triolein autoxidation. (c) Example of a triolein-derivative molecule composed of the following fragments: glycerol with adjacent esters and three carbon chains (previously oleic chains), having carboxyl, hydroxy, and hydroperoxy functional groups. Functional groups present in the molecule are highlighted in the same colors as in (b).

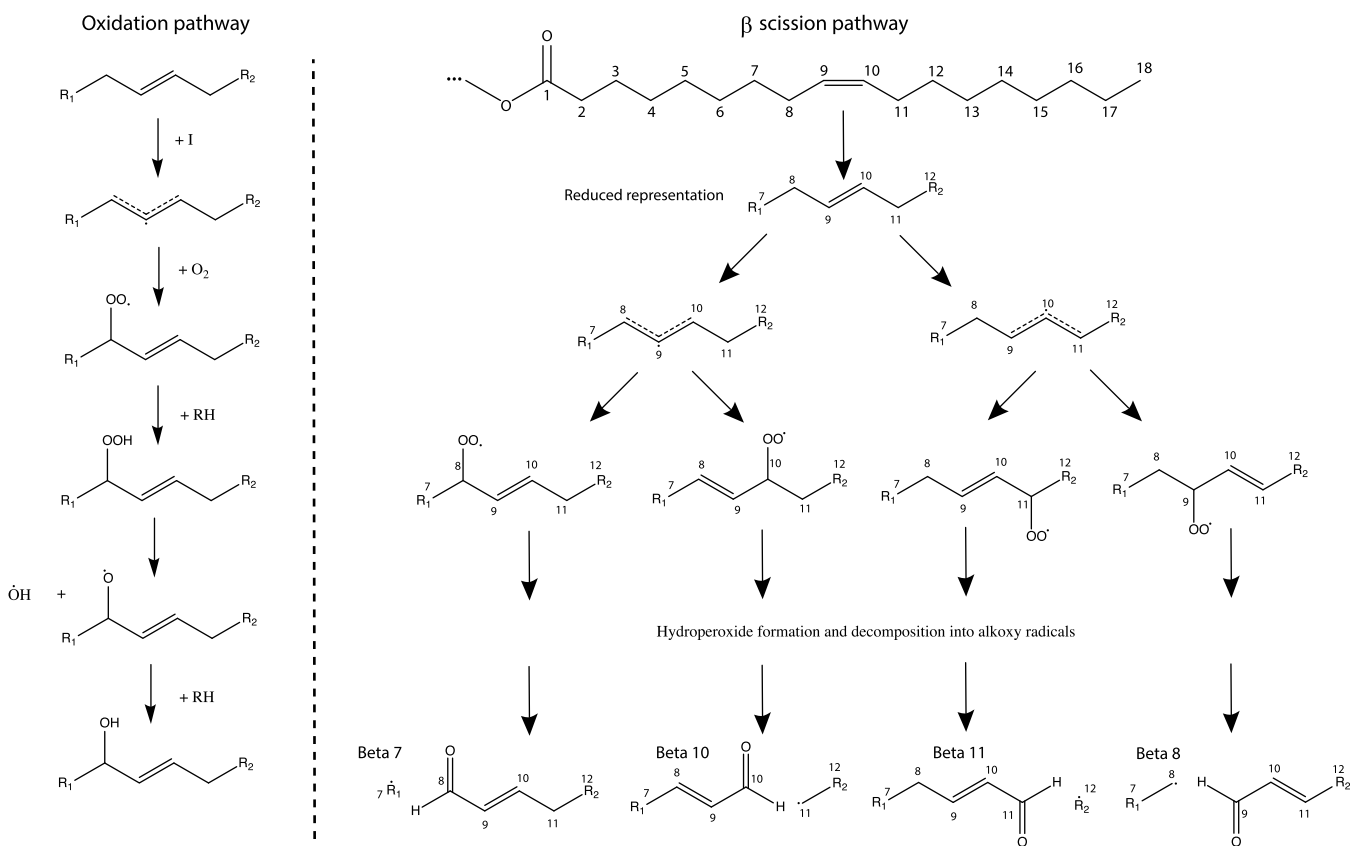


Figure 3. Reaction pathways of unsaturated double bonds leading to the formation of a wide range of monomers. (Left) Oxidation pathway and (right) β scission pathway leading to four different products.

very rapidly reacts with oxygen to form a peroxy radical. These radicals abstract other allylic hydrogens, thus forming hydro-

peroxides. Decomposition of hydroperoxides results in alkoxy and hydroxy radicals. This reaction happens under the influence

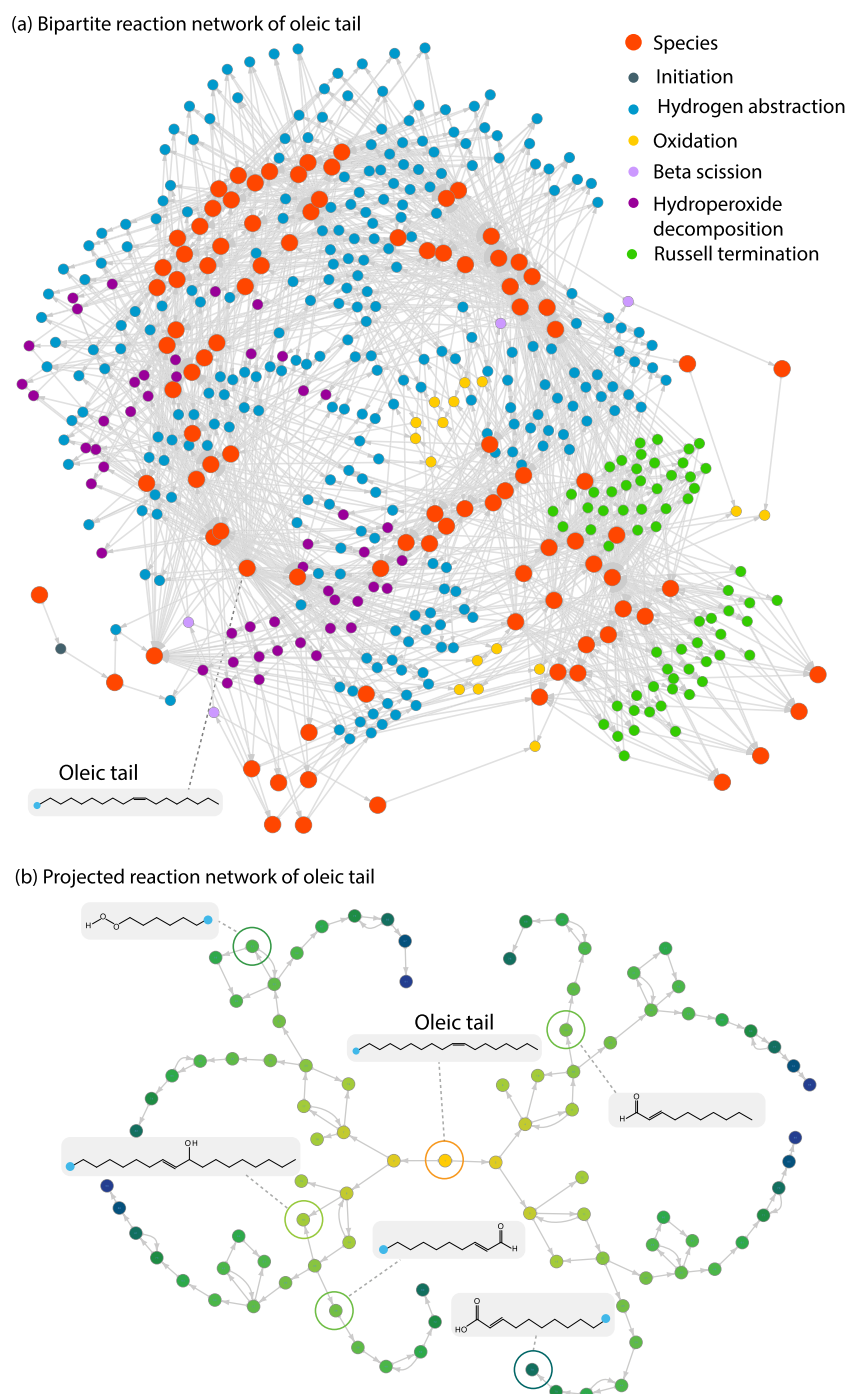


Figure 4. (a) Bipartite reaction network of transformations of oleic chain with two types of nodes: species and reactions. Species correspond to large red nodes, and reactions correspond to smaller nodes of different colors. (b) Projected monopartite reaction network of transformations of oleic chain with one type of nodes that corresponds to species. The edges point from a reactant to a product. The root node in the center (circled) is the initial state of the oleic tail. Color intensity indicates the distance from the root. Each node represents a distinct derivative of oleic chain formed via predefined oxidation scheme. See Supporting Information Figure S5 for an extended version of this figure with all molecular graphs.

of light and chemicals containing transition metals, like cobalt, that are used as drying agents for drying oils such as linseed oil. Two reaction pathways are then possible from the alkoxy radical state: (1) another allylic hydrogen abstraction by the alkoxy radical to form an alcohol or (2) β scission of C–C bond next to the alkoxy radical. If the scission occurs on the side closer to the glycerol ester, the disconnected fragment remains in the system. However, if the β scission happens on the side closer to the end of the hydrocarbon chain, low-molecular-weight components

(aldehydes, ketones) volatilize.⁶² Aldehydes that remain in the system undergo hydrogen abstraction, forming an alkyl radical, which again undergoes the oxidation pathway. This pathway leads to the formation of carboxylic acids, stable degradation products. Furthermore, all of the radicals that appear during the autoxidation mechanism may terminate via recombination reactions and form oligomers. A special type of termination, Russell termination, consumes peroxy radicals and produces ketones, alcohols, and oxygen. As the present work focuses on

matching masses of monomers, termination via radical recombination leading to oligomers is not included in the set of reaction templates.

With respect to our fragmentation scheme, the input for the ARNG consists of molecular graphs representing the initial species at Level 2 and the above-described reaction mechanism represented as reaction templates. The algorithm then discovers all derivative products by recursively applying the reaction templates. Two ways of visualization of the reaction network for triolein oxidation mechanism are illustrated in Figure 4a,b. Figure 4a is a visualization of full reaction network. It is a bipartite network that includes both species and reactions. Species are connected to each other through reactions. An edge coming from a reactant species points toward a reaction node. An edge from a reaction node points toward a product species. Species nodes correspond to big red circles, while the reaction nodes are smaller circles of different colors. Reactions are categorized in initiation, hydrogen abstraction, oxidation, β scission, hydroperoxide decomposition, and Russell termination. Figure 4b shows a bipartite reaction network from Figure 4a projected on a monopartite one, where nodes correspond to the species and the edges correspond to the transformation from a reactant to a product. Colors correspond to the distance from the initial state of oleic chain. The essential features of such transformations may also be understood by drawing a parallel with “phylogenetic” trees (see Supporting Information Figure S5). These trees specify all derivatives that could be obtained from a single input species by following the shortest route of transformations. Since these species are only the transformed fragments, the transformed complete triglycerides are obtained by connecting all derivatives of heads and tails to obtain molecular graphs of the intermediate and product species generated by the reaction scheme of triolein implemented in ARNG.

The complete reconstruction generates 14 045 unique molecular graphs. This number includes numerous repetitions of molecular structures, which are all distinguishable by graph isomorphism search.⁶³ The ARNG performs this search every time a potentially new molecular graph is generated. The isomorphism search distinguishes all differences in the structure of a molecular graph representing the isomers, while these isomers are identical in the ESI-MS analysis. This is illustrated by the following example. The molecule of triolein is symmetric in terms of its fatty ester composition. Moreover, the reactive site of each fatty acid tail, the double bond, is also symmetric. When triolein undergoes a reaction and is transformed into a product, a functional group may be located on the allylic position on either side of the double bond. Although this leads to the formation of two chemically different molecules, their chemical reactivity is the same according to our assumptions concerning chemical reactivity. Furthermore, as this type of isomerism may happen to each of the three identical fatty esters, six very similar structural isomers are resulting with identical mass and chemical functionality. The ARNG methodology generates these six configurations of molecular graphs as different graphs. This example illustrates that the majority of 14 045 molecular graphs may be structural isomers, which differ only by the location of a functional group with respect to the double bond and a fatty acid ester group. To reduce this set, the molecules that are characterized having the same mass, number, and type of functional groups, as well as length of fatty esters are grouped together. From each group, a representative molecule (the first molecule generated by the ARNG within each group) is added

to the reduced set of molecular graphs. Thus, the set of molecular graphs generated by the ARNG is reduced to 1483 molecules, which are further used for the identification of ESI-MS measured peaks.

Importantly, along with this procedure, all masses of the reconstructed molecules may be computed for easy comparison to ESI-MS spectral products of triolein autoxidation, hydrolysis, and β scission. This calculated set of masses is the direct link to the interpretation of the measured MS spectrum.

Matching Masses of ESI-MS and ARNG. The results from ESI-MS provide information about the masses and relative abundance of various molecules present in the measured sample, including arising from variations in isotopic compositions of a given molecule. Although direct MS formally measures mass-to-charge ratio, it is assumed, under the experimental conditions applied, that the absolute charge is equal to 1 (see ref 25). Mass spectra from both positive and negative modes were considered for matching, as each mode has been previously reported to show different sensitivities to oil-based products.⁶⁴

For this first proof of concept, consideration of reaction products was limited to those that are most isotopically abundant. Thus, exact masses were calculated for each computationally derived structure (i.e., generated by the ARNG algorithm) assuming the mass of only the most abundant isotope for each element (see the Methods section). To match the peaks in each mass spectrum, the calculated masses were further corrected to match their corresponding ions (with the most abundant isotopes) formed in the ESI-MS: adding masses of H^+ , Na^+ , or NH_4^+ for positive mode and subtracting mass of H^- or adding mass of CH_3COO^- for negative mode. Then, the experimentally measured masses were matched to the calculated ones allowing 0.15 m/z difference between them (see Supporting Information Figure S6). Molecular structures that are generated by the ARNG algorithm and matched to mass spectrum peaks are given in Supporting Information Table S2. See also Supplementary data for raw files.

Matching Masses of ESI-MS and ARNG Including Oxidation Pathway. We will now illustrate the peak matching procedure aided by the ARNG model. Figure 5 shows the MS measurement on the sample at the early stage of drying process of triolein measured in positive mode together with the products matched by the ARNG model. Matching is done with the molecular structures generated by the algorithm accounting only for the oxidation pathway that occurs after the allylic hydrogen

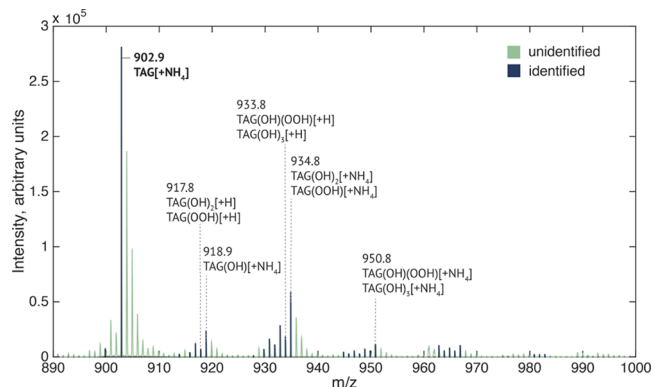


Figure 5. Spectrum of extracts from triolein with titanium dioxide at early drying stage measured in positive mode showing products of oxidation pathways. Bold dark blue lines highlight the peaks that were matched with oxidation products for a given ionizing ion.

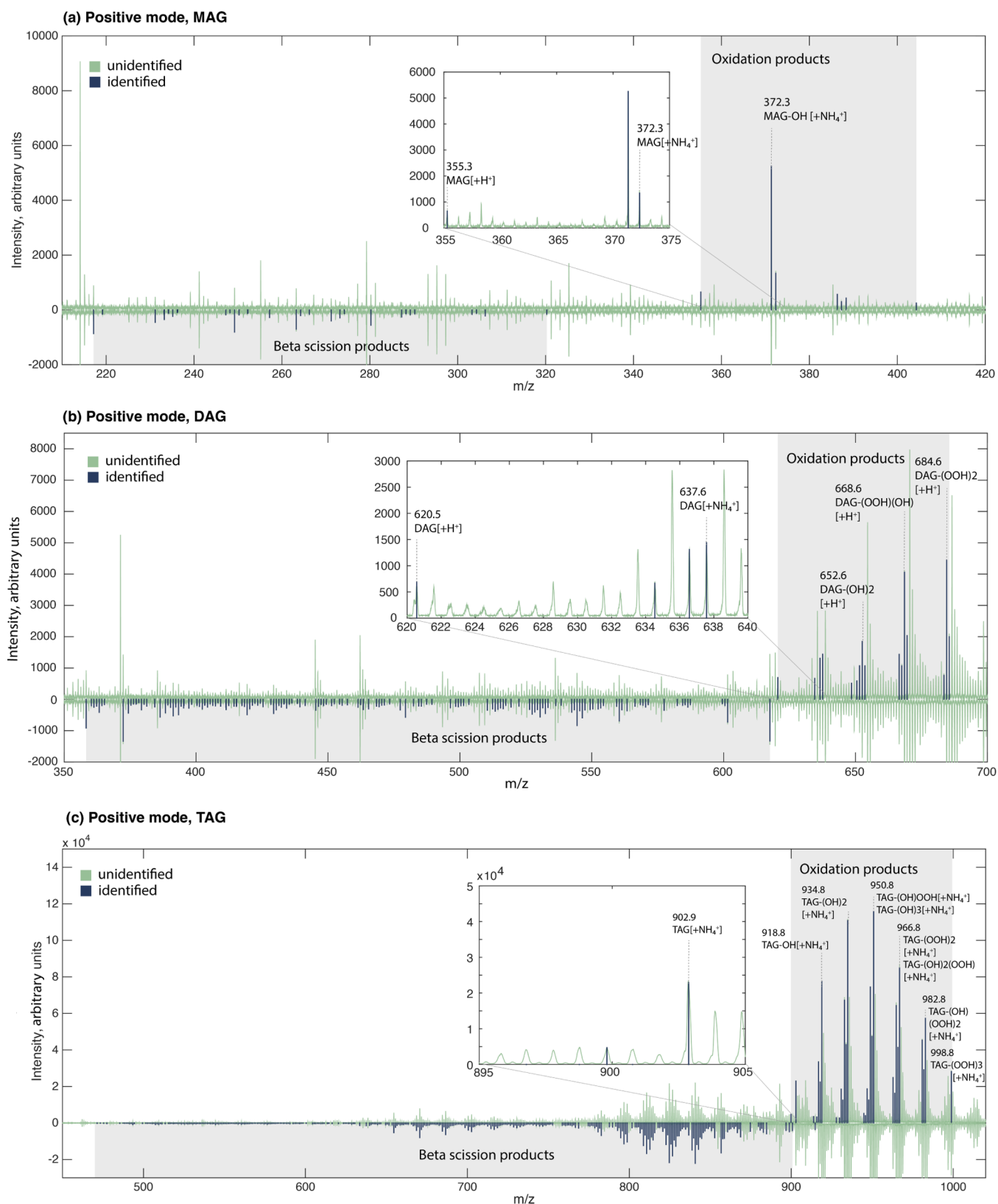


Figure 6. Spectra of extracts from triolein aged with titanium dioxide measured in positive mode highlighting products of hydrolysis, oxidation, and β scission. Products of oxidation and β scission pathways of MAG (a), DAG (b), and TAG (c). The dark vertical lines in the upper part of each spectrum highlight the peaks that were matched with oxidation products, and the dark vertical lines on the lower mirrored spectrum highlight the peaks that were matched to the products formed after β scission reaction. The light gray areas in the upper part of each spectrum correspond to the approximate regions for oxidation products, and the light gray areas in the lower part of each spectrum correspond to the approximate regions for β scission products.

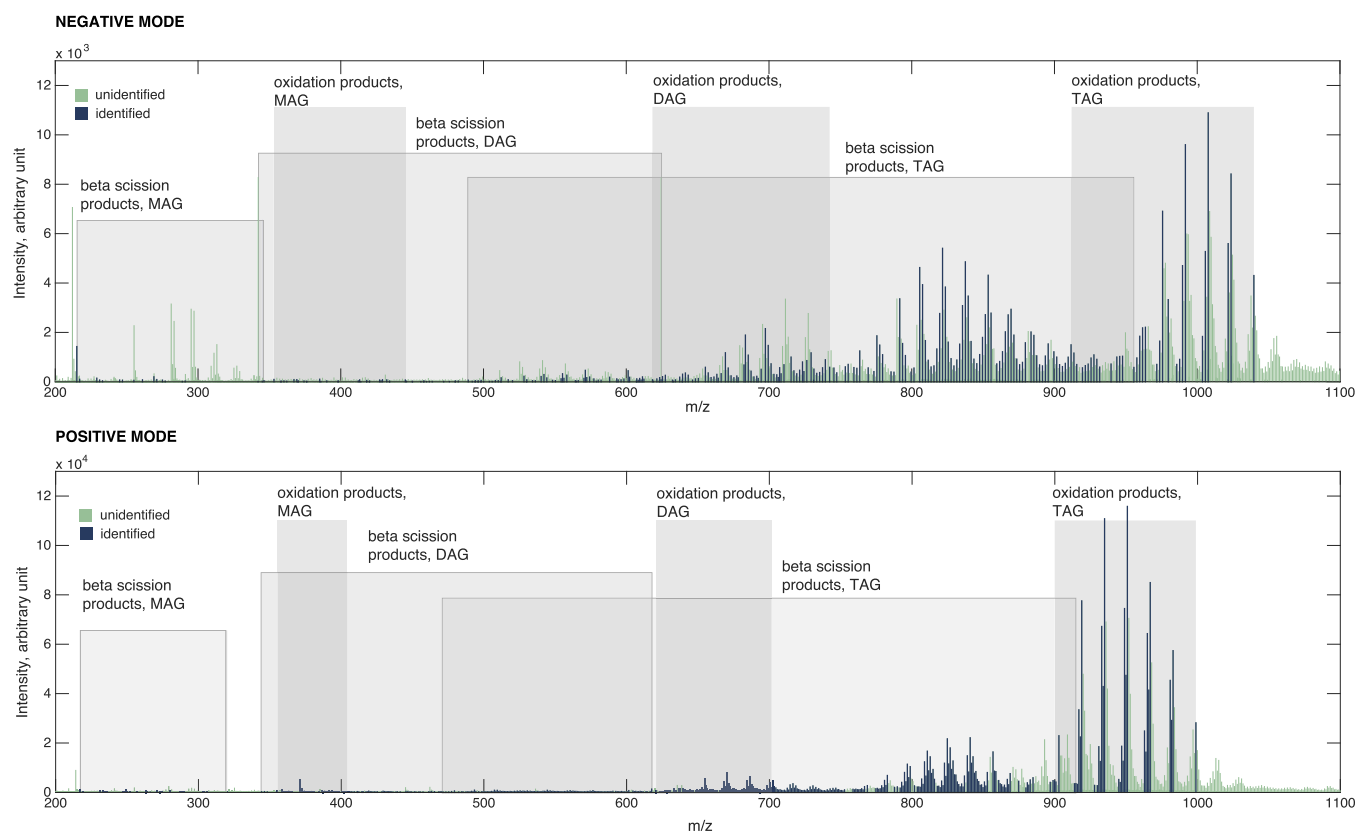


Figure 7. Regions of mass spectra (both negative and positive modes) highlighting the anticipated products of oxidation and β scission on triolein (TAG) and its products after hydrolysis (MAG and DAG). Peaks matched to the molecular structures generated by the ARNG (dark blue) in the measured mass spectra (light green) of extracts from triolein aged with titanium dioxide. Heights for matched peaks are set to the height of the corresponding measured peaks.

abstraction (see Figure 3). These peaks can be easily assigned in a manual manner and serve as a proof of concept for our methodology. The beginning of the drying process of triolein is characterized by the oxidation pathway, where the expected reaction products are hydroxides and hydroperoxides, which are known to be stable products under MS measurement conditions.⁶⁵ By including only the oxidation pathway in the ARNG methodology, the reaction scheme consists of triolein and all its derivatives containing different number and combinations of [OH] and [OOH] groups on their oleic chains.

Further, the methodology is applied to the artificially aged triolein sample that demonstrates a wide variety of measured MS peaks. We will proceed with smaller, cutdown parts of the model that describe only a limited amount of reactions and/or species and therefore only match part of the measured MS spectrum. Then, we will gradually increase the number of species and reactions taken into account in the model and observe how the matching with the measured MS spectrum is improved. The result of this model-supported matching exercise is depicted in Figure 6 (positive mode) and Supporting Information Figure S7 (negative mode). The model is first cut down to only the oxidation reactions of MAG, where no more than four products are detectable under the ionization conditions in both modes: pure MAG and MAG containing hydroperoxide, hydroxyl, or carbonyl group. In the spectrum measured in the positive mode, see the upper part of spectrum in Figure 6a, there are seven peaks that correspond to four molecular species with the aforementioned functional groups accounting for two different ions, H^+ and CH_3COO^- . In the negative mode, the intensity of these

peaks is low relative to other peaks, indicating that MAGs are minor products in the detected sample. More pronounced intensity in the region of the oxidation of MAG is seen in positive mode, indicative of the variable sensitivities of the two modes to different products.⁶⁴ Subsequently, we increase the number of species by including DAGs and TAGs as well. In the upper parts of the spectra in Figure 6b,c, we highlight the peaks corresponding to the oxidation of DAGs and TAGs. We conclude that the intensity of DAG oxidation products is also rather small, while the oxidation products of TAGs are more pronounced. The abundance of TAGs also implies that this species did not undergo substantial hydrolysis. A similar trend concerning oxidation versus hydrolysis products is observed when regarding the measured and matched peaks in the negative mode (see Supporting Information Figure S7).

Matching Masses of ESI-MS and ARNG Including β Scission Pathway. Next, we extend the reaction scheme of the ARNG model by including the β scission reaction in the set of reaction templates, and the procedure described above is repeated by successively matching peaks for TAG, DAG, and MAG molecules. Peaks that could be matched after introducing β scission are shown in the lower parts of spectra in Figure 6a–c. One observes that this reaction gives rise to a large variety of masses of intermediate and product molecular species. Comparing the matched parts of the spectra of oxidation only with the spectra of matched peaks after β scission in Figure 7 indicates that oxidation of MAG overlaps with β scission products of DAG. In addition, we see that oxidation products of DAG strongly overlap with the β scission products of TAG in the

spectra. It is obvious that this matching analysis reveals boundaries when attributing products to the various possible reaction pathways. The boundaries for oxidation and β scission products of MAG, DAG, and TAG are summarized in Table 1.

Table 1. Regions of Spectra in Positive and Negative Modes Containing the Anticipated Products of Oxidation and β Scission on Triolein-Derived Species and Their Hydrolysis Products (MAG and DAG)

	oxidation	β scission
	Positive Mode	
MAG	355.3–404.3 m/z	217.2–320.3 m/z
DAG	620.6–701.6 m/z	344.3–617.6 m/z
TAG	899.8–998.8 m/z	417.6–914.8 m/z
	Negative Mode	
MAG	353.2–445.2 m/z	229.1–345.0 m/z
DAG	618.4–742.5 m/z	342.1–624.4 m/z
TAG	911.6–1039.7 m/z	489.3–955.6 m/z

Matching Masses of ESI-MS and ARNG Including All Reaction Pathways. The result of the mass matching for the triolein molecule is shown in Figure 7 (for mass matching per ion in positive and negative modes, see Supporting Information Figure S8). For the sake of clarity, peak matching is shown in the range between 200 and 1100 m/z only. Although this range does not include very small products, the plot yet demonstrates the variety of molecular species present in the material after aging (with the highest possible mass of the monomer of 998.8 m/z in positive mode and 1039.7 m/z in negative mode). The measured peaks that matched to the calculated molecular structures are highlighted in dark blue, while the remaining

measured peaks are shown in light green. Note that the height of the peak corresponds to the measured intensity. The methodology was able to match 56 out of 151 high-intensity peaks (higher than 10% of intensity of the highest peak in the measured mode) in negative mode and 35 out of 67 in positive mode. Out of 1483 computationally generated molecular graphs, 665 were found in the negative-mode spectrum and 1330 were found in the positive-mode spectrum of artificially aged sample. High-resolution spectra with matched peaks can be seen in Supporting Information Figures S11 and S12. Although numerous peaks in the measured spectra are matched, there are peaks in positive as well as in negative modes, namely, regions of ca. 700–800 m/z and beyond 1000 m/z that do not match any molecular structure generated by ARNG. The unidentified regions on the spectrum might correspond to dimers formed from relatively low-molecular-mass species or other reaction pathways that are not included in this study. We did not analyze this in further detail as this work focuses on identifying monomeric reaction products only.

Automated Identification of Functional Groups. The reconstructed molecular graph of each molecular species contains one or more functional groups from Figure 2b. An example of a triolein-derivative molecule having carboxy, hydroxy, and hydroperoxy functional groups is shown in Figure 2c. Functional groups present in the molecule are highlighted in the same colors as in Figure 2b. This allows grouping of molecules present in the mass spectrum according to their functional groups to access the overall state of the measured sample. The overall distribution of these groups is given in Figure 8 with three pie charts constructed from matching with ESI results from both negative and positive modes. Three charts represent the relative amount of (1) chemical classes: acids,

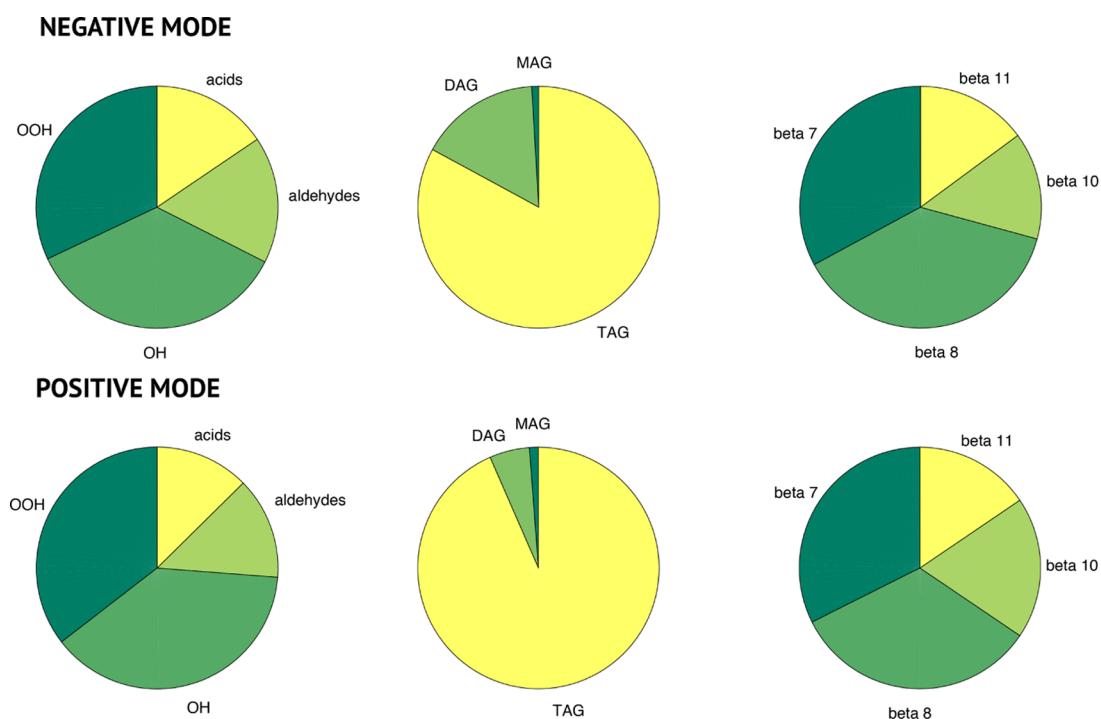


Figure 8. Pie charts illustrating relative ratios of various products present in the sample for two different measurement modes: positive and negative accounting for all ionizing ions. Pie charts on the left demonstrate relative ratios of oxidation products contained in the sample: hydroperoxides, alcohols, carboxylic acids, and aldehydes. Pie charts in the middle demonstrate relative ratios of hydrolysis products: triglycerides (triacylglycerol, TAG), diglycerides (DAG), and monoglycerides (MAG). Pie charts on the right illustrate relative ratios of β scission products: β 7, β 8, β 10, and β 11 corresponding to the number of carbons left on the oleic chain (counting from the ester bond).

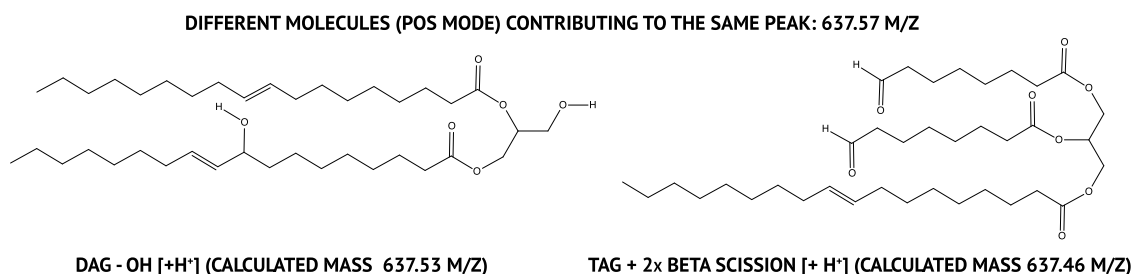


Figure 9. Example of two different molecules both contributing to the mass of 637.57 m/z measured in positive mode. Calculated masses account for ion H^+ . The molecule on the left is a diglyceride with hydroxyl on one of its oleic chains. The molecule on the right is a triglyceride after β scission occurring on two oleic chains. For the complete list of matched molecular structures, see Supporting Information Tables S1 and S2.

aldehydes, alcohols, and hydroperoxides; (2) hydrolysis products: triglycerides, diglycerides, and monoglycerides; and (3) β scission products, namely, the relative amounts of carbon chains (lengths of 7, 8, 10, and 11 carbons) that remain connected to the glycerol. This overview illustrates the different outcomes of the analysis with negative and positive modes, complementing earlier observations^{41,64} and offering a detailed quantification of the complex variety of oil-based products. In the same manner, functional groups may be analyzed in spectra; see Supporting Information Figure S9, where the highlighted peaks correspond to the molecular structures containing alcohols and aldehydes. This approach additionally provides the possibility to extract information about any functional group present in the system, to compare samples changing in time, or to explore the behavior of material in the presence of various additives, serving as only one of the many possible applications of the ARNG-MS combination.

Identification of Species Contributing to the Same Measured Mass. One final feature of the ARNG-based matching to be mentioned is the ability to reveal various species contributing to the same measured mass. Experimentally, this would require additional analytical techniques and method development, such as chromatographic separation or tandem MS. Figure 7c,d demonstrates the complexity of the measured sample showing significant overlaps between the products of different reaction pathways. This implies that some peak intensities may get contribution from more than one possible structure. Histograms on Supporting Information Figure S13 show the number of identified peaks that are matched to the same number of molecular graphs. One can see that the majority of the peaks are matched to one or two molecular graphs, while the highest number of molecular graphs per peak being 13 in negative mode and 27 in positive mode.

As structural isomers have the same molecular formula, their masses are exactly the same. However, the reaction pathways leading to such structural isomers can vary significantly. For example, in the case of triolein, the addition of one hydroperoxy group, or [OOH], may occur on either side of a double bond. Slightly more complex, triolein following the addition of one hydroperoxy group ([OOH]) is a structural isomer to triolein following the addition of two hydroxy groups (2[OH]); as each group requires a hydrogen abstraction, the products have the same mass yet were achieved through slightly different pathways. Structural isomers may also be formed following different β scission pathways. For example, triolein following β scission on two chains may result in a pair of chains of length 7 and 11 carbons or length 8 and 10 carbons, both of which produce the same mass (see Supporting Information Figure S10).

However, species of very similar mass but more distinct structures can also occur. In Figure 9, two states of triolein are shown, which both may contribute to the same mass of 637.57 m/z measured in positive mode, as their exact masses fall in the tolerance interval for peak assignment in our methodology and, more importantly, are not resolvable in the mass spectra (see Supporting Information Figure S6). These molecules have a distinctly different structure: the molecule on the left has one fatty acid hydrolyzed and an added hydroxy group on one of the remaining acids, while the molecule on the right is a result of two β scission reactions occurring on two of its fatty acids resulting in two chains of eight carbons ending in aldehydes (i.e., “ β 8” in Figure 8). These differences further illustrate how the ARNG-MS approach helps to identify not only relevant products but also the chemical pathways occurring in the oxidation of triolein.

As it currently stands, the algorithm may identify several molecular structures contributing to the same peak; however, the algorithm does not assign the probabilities to these molecular structures. Such a problem can be partially relieved using kinetic modeling, which brings the information about the concentration of various species at different points of time. Another possibility is to rank the molecular structures corresponding to a single peak by estimating their energies of formation with quantum-chemical calculations. These ideas are worth investigating in future and are out of the scope of our current paper.

DISCUSSION AND CONCLUSIONS

This paper addresses interpretation of complex data from MS using automated reaction discovery. Combining ARNG with the results from ESI-MS allows one to depart from manual peak assignment and enrich the output of this experimental technique by matching molecular structures to the measured masses. This approach was applied to study the aged sample of triolein with monounsaturated oleic chains. Although triolein is considered to be one of the simpler cases of triglyceride autoxidation, it leads to an intertwined scheme of reactions that involves a large number of intermediate and product species, as is illustrated by the wide range of masses measured by ESI-MS. The reaction products from ARNG show good coverage of the area of the identified spectrum. The methodology matched 56 out of 151 abundant species in negative mode and 35 out of 67 in positive mode. Unidentified peaks from ca. 700–800 m/z and beyond 1000 m/z that can be seen in Figure 7 might correspond to dimers (formed by lower-molecular-mass species) or products of other autoxidative reactions, for which we did not account in this work. The methodology can be extended to model polymers by introducing an additional hierarchical level in the representation of a molecule. This implies that one should be

able to infer information about the connectivity of the whole polymer network from the fragments given, and thus, reconstruct dimers, trimers, higher oligomers, etc. In the future, we will exploit this property of our algorithm to compute macroscopic polymer properties, as, for instance, the average size or the size distribution.

The method can identify products of particular reaction pathways by including and excluding various reactions from the set of reaction templates defined in the ARNG, as was demonstrated with the hydrolysis, oxidation, and β scission reactions. We identified the regions of the mass spectrum corresponding to the anticipated products of particular reaction pathways. Reconstructing explicit molecular graphs using the ARNG provides access to a detailed description of the functional groups attributed to the MS peaks. This information is presented in the shape of pie charts that indicate relative amounts of characteristic functional groups present in the matched molecules. Identification of reaction pathways and pie charts representation of various functional groups provide global information about a sample as a whole. The model is furthermore able to distinguish between different species with the same or similar molecular mass, thereby assisting in the interpretation of peaks in the mass spectrum with contributions from more than one molecule.

This work is a pilot study, which demonstrates possible tandem between automated reaction network generation methodology and mass spectrometry. On the experimental side, the availability of ultrahigh-resolution devices (e.g., orbitrap, FT-ICR) would help to validate our peak matching methodology and improve its precision. The coupling of experiments and computational methodology demonstrated in this paper has potential to complement tandem MS analyses and be used for instances when high-resolution measurements are not feasible.

The results of this paper have a qualitative nature and can be considered as a first step toward large-scale studies of chemical systems with similar complexity to triglycerides. The ARNG specifies all possible species in the course of reactions but does not quantify their concentrations. Such a task would require translating the output of the algorithm into a set of ordinary differential equations, a kinetic model having quantitative predictive power. The relative ratios of the concentrations of the individual molecular masses can then be related to the relative ratios of the intensities of the peaks of the corresponding masses. Such information can be used to deduce the (relative) speed of various reactions and ultimately estimate the kinetic rate constants. However, the estimation of the kinetic parameters for the numerous reactions is a formidable task. This has to rely on such concepts as kinetic similarity within “families of reactions”^{22,23,56,66} and will further require developing model reduction techniques, which will enable one to carry out the ultimate quantitative validation step with experimental data from MS and other sources.

Finally, the presented two-level hierarchy approach for disassembling a molecule into smaller substructures reduces the amount of computational time needed for the ARNG to reconstruct all possible intermediate and product species. This approach may be used to study various configurations of more complex triglycerides, in which case the algorithm may be applied separately to resolve reaction schemes evolving from different fatty acids: oleic, linoleic, and linolenic. Subsequently, making use of a similar hierarchical fragmentation procedure may allow for the study and reconstruction of the complete set of

triglycerides. Thus, many larger systems involving various configurations of fatty acids are within the reach of the current modeling—experiment paradigm. This will ultimately allow one to study the effects of various additives on aging of oil (e.g., metal-containing pigments in oil-based paints).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c01163>.

Structural graphs of identified species within top 90% of measured intensity (Table S1) (PDF)

Reaction rules necessary to describe triolein autoxidation and hydrolysis (Figures S1 and S2); auxiliary patterns (Figure S3); schematic representation of all possible configurations of triglyceride molecule after hydrolysis (Figure S4); history tree indicating the shortest paths of transformations from the oleic fragment of triolein (Figure S5); example of a peak with more than one contribution (Figure S6); spectra of extracts from triolein aged with titanium dioxide measured in negative mode (Figure S7); matched peaks with different ions (Figure S8); matched peaks per functional group (Figure S9); isomers formed via β scission (Figure S10); high-resolution spectra of extracts from triolein aged with titanium dioxide measured in positive mode (Figure S11); high-resolution spectra of extracts from triolein aged with titanium dioxide measured in negative mode (Figure S12); and histograms showing frequency of identified peaks (Figure S13) (PDF)
Supplementary data (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Ivan Kryven – *Mathematical Institute, Utrecht University, Utrecht 3584 CD, The Netherlands; Centre for Complex Systems Studies, Utrecht 3584 CE, The Netherlands;*
 orcid.org/0000-0002-3964-2196; Email: i.kryven@uu.nl

Authors

Yuliia Orlova – *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam 1098 XH, The Netherlands*

Alessa A. Gambardella – *Rijksmuseum, Amsterdam 1071 ZC, The Netherlands*

Katrien Keune – *Rijksmuseum, Amsterdam 1071 ZC, The Netherlands*

Piet D. Iedema – *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam 1098 XH, The Netherlands*

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.0c01163>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Y.O. gratefully acknowledges financial support for PREDAGIO project from The Netherlands Organisation for Scientific Research (NWO). Andrea Gargano is acknowledged for his constructive criticism on the manuscript. A.A.G. thanks AkzoNobel (The Netherlands) for funding; Rob Erdmann of

the Rijksmuseum for support in Python; The Cultural Heritage Agency of the Netherlands (RCE) for providing the ESI-MS and the automatic muller (on permanent loan from Old Holland); and Art Ness Proaño Gaibor, Klaas Jan van den Berg, Federica Parlanti, and Fabiana Di Gianvincenzo of the RCE for assistance during experiments.

REFERENCES

- (1) Tetko, I. V.; Engkvist, O.; Koch, U.; Reymond, J.-L.; Chen, H. BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry. *Mol. Inf.* **2016**, *35*, 615–621.
- (2) May, J. C.; McLean, J. A. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annu. Rev. Anal. Chem.* **2016**, *9*, 387–409.
- (3) Gunstone, F. *Oils and Fats in the Food Industry*; Wiley, 2009.
- (4) Gunstone, F. *Vegetable Oils in Food Technology: Composition, Properties and Uses*; Wiley, 2011.
- (5) Lisa, M.; Holčapek, M. Triacylglycerols Profiling in Plant Oils Important in Food Industry, Dietetics and Cosmetics Using High-Performance Liquid Chromatography-Atmospheric Pressure Chemical Ionization Mass Spectrometry. *J. Chromatogr. A* **2008**, *1198–1199*, 115–130.
- (6) Zhou, X.; Nolte, M. W.; Mayes, H. B.; Brent, H.; Shanks, L. J. B. Experimental and Mechanistic Modeling of Fast Pyrolysis of Neat Glucose-Based Carbohydrates. 1. Experiments and Development of a Detailed Mechanistic Model. *Ind. Eng. Chem. Res.* **2014**, *53*, 13274–13289.
- (7) Colombini, M. P.; Modugno, F.; Giacomelli, M.; Francesconi, S. Characterisation of Proteinaceous Binders and Drying Oils in Wall Painting Samples by Gas Chromatography-Mass Spectrometry. *J. Chromatogr. A* **1999**, *846*, 113–124.
- (8) Bonaduce, I.; Carlyle, L.; Colombini, M. P.; Duce, C.; Ferrari, C.; Ribechini, E.; Selleri, P.; Tiné, M. R. New Insights into the Ageing of Linseed Oil Paint Binder: A Qualitative and Quantitative Analytical Study. *PLoS One* **2012**, *7*, No. e49333.
- (9) Neff, W. E.; Byrdwell, W. Characterization of Model Triacylglycerol (Triolein, Trilinolein and Trilinolenin) Autoxidation Products via High-Performance Liquid Chromatography Coupled with Atmospheric Pressure Chemical Ionization Mass Spectrometry. *J. Chromatogr. A* **1998**, *818*, 169–186.
- (10) Baij, L.; Astefanei, A.; Hermans, J.; Brinkhuis, F.; Groenewegen, H.; Chassouant, L.; Johansson, S.; Corthals, G.; Tokarski, C.; Iedema, P.; Keune, K. Solvent-Mediated Extraction of Fatty Acids in Bilayer Oil Paint Models: a Comparative Analysis of Solvent Application Methods. *Heritage Sci.* **2019**, *7*, 1–8.
- (11) Muizebelt, W. J.; Donkerbroek, J. J.; Nielen, M. W. F.; Hussem, J. B.; Biemond, M. E. F.; Klaasen, R. P.; Zabel, K. H. Oxidative Crosslinking of Alkyd Resins Studied with Mass Spectrometry and NMR Using Model Compounds. *J. Coat. Technol.* **1998**, *70*, 83–93.
- (12) Oakley, L. H.; Casadio, F.; Shull, K. R.; Broadbelt, L. J. Theoretical Study of Epoxidation Reactions Relevant to Hydrocarbon Oxidation. *Ind. Eng. Chem. Res.* **2017**, *56*, 7454–7461.
- (13) Fjällström, P.; Andersson, B.; Nilsson, C.; Andersson, K. Drying of Linseed Oil Paints: a Laboratory Study of Aldehyde Emissions. *Ind. Crops Prod.* **2002**, *16*, 173–184.
- (14) Oyman, Z. O.; Ming, W.; van der Linde, R. Oxidation of Model Compound Emulsions for Alkyd Paints Under the Influence of Cobalt Drier. *Prog. Org. Coat.* **2003**, *48*, 80–91.
- (15) Oyman, Z.; Ming, W.; van der Linde, R.; van Gorkum, R.; Bouwman, E. Effect of [Mn(acac)₃] and its Combination with 2,2'-bipyridine on the Autoxidation and Oligomerisation of Ethyl Linoleate. *Polymer* **2005**, *46*, 1731–1738.
- (16) van Gorkum, R.; Bouwman, E. The Oxidative Drying of Alkyd Paint Catalysed by Metal Complexes. *Coord. Chem. Rev.* **2005**, *249*, 1709–1728.
- (17) Bouwman, E.; van Gorkum, R. A Study of New Manganese Complexes as Potential Driers for Alkyd Paints. *J. Coat. Technol. Res.* **2007**, *4*, 491–503.
- (18) Mallégol, J.; Gardette, J.-L.; Lemaire, J. Long-Term Behavior of Oil-Based Varnishes and Paints. Photo- and Thermooxidation of Cured Linseed Oil. *J. Am. Oil. Chem. Soc.* **2000**, *77*, 257–263.
- (19) Spier, E.; Neuenschwander, U.; Hermans, I. Insights into the Cobalt(II)-Catalyzed Decomposition of Peroxide. *Angew. Chem., Int. Ed.* **2013**, *52*, 1581–1585.
- (20) Oakley, L. H.; Casadio, F.; Shull, K. R.; Broadbelt, L. J. Examination of Mechanisms for Formation of Volatile Aldehydes from Oxidation of Oil-Based Systems. *Ind. Eng. Chem. Res.* **2018**, *57*, 139–149.
- (21) Iedema, P.; Hermans, J.; Keune, K.; van Loon, A.; Stols-Witlox, M. In *Mathematical Modeling of Mature Oil Paint Networks*, ICOM-CC 17th Triennial Conference Preprints, Melbourne, ISBN 978-92-9012-410-8; Bridgland, J., Eds.; International Council of Museums: Paris, 2014; art. 1604, p 8.
- (22) Oakley, L. H.; Casadio, F.; Shull, K. R.; Broadbelt, L. J. Microkinetic Modeling of the Autoxidative Curing of an Alkyd and Oil-Based Paint Model System. *Appl. Phys. A* **2015**, *121*, 869–878.
- (23) Oakley, L. H.; Casadio, F.; Shull, P. K. R.; Broadbelt, P. L. J. Modeling the Evolution of Crosslinked and Extractable Material in an Oil-Based Paint Model System. *Angew. Chem., Int. Ed.* **2018**, *57*, 7413–7417.
- (24) Han, X.; Gross, R. W. Quantitative Analysis and Molecular Species Fingerprinting of Triacylglyceride Molecular Species Directly from Lipid Extracts of Biological Samples by Electrospray Ionization Tandem Mass Spectrometry. *Anal. Biochem.* **2001**, *295*, 88–100.
- (25) Li, M.; Butka, E.; Wang, X. Comprehensive Quantification of Triacylglycerols in Soybean Seeds by Electrospray Ionization Mass Spectrometry with Multiple Neutral Loss Scans. *Sci. Rep.* **2014**, *4*, No. 6581.
- (26) El-Anead, A.; Cohen, A.; Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **2009**, *44*, 210–230.
- (27) Glish, G. L.; Vachet, R. W. The Basics of Mass Spectrometry in the Twenty-First Century. *Nat. Rev. Drug Discovery* **2003**, *2*, 140–150.
- (28) Sleno, L. The Use of Mass Defect in Modern Mass Spectrometry. *J. Mass Spectrom.* **2012**, *47*, 226–236.
- (29) Lerno, L. A.; German, J. B.; Lebrilla, C. B. Method for the Identification of Lipid Classes Based on Referenced Kendrick Mass Analysis. *Anal. Chem.* **2010**, *82*, 4236–4245.
- (30) Grimme, S. Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306–6312.
- (31) Bauer, C. A.; Grimme, S. First Principles Calculation of Electron Ionization Mass Spectra for Selected Organic Drug Molecules. *Org. Biomol. Chem.* **2014**, *12*, 8737–8744.
- (32) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Ind. Eng. Chem. Res.* **1994**, *33*, 790–799.
- (33) Broadbelt, L. J.; Pfaendtner, J. Lexicography of Kinetic Modeling of Complex Reaction Networks. *AIChE J.* **2005**, *51*, 2112–2121.
- (34) Orlova, Y.; Kryven, I.; Iedema, P. D. Automated Reaction Generation for Polymer Networks. *Comput. Chem. Eng.* **2018**, *112*, 37–47.
- (35) Lee, J.-u.; Kim, Y.; Kim, W. Y.; Oh, H. B. Graph Theory-Based Reaction Pathway Searches and DFT Calculations for the Mechanism Studies of Free Radical-Initiated Peptide Sequencing Mass Spectrometry (FRIPS MS): a Model Gas-Phase Reaction of GGR tri-peptide. *Phys. Chem. Chem. Phys.* **2020**, *22*, 5057–5069.
- (36) Song, J. Building Robust Chemical Reaction Mechanisms: Next Generation of Automatic Model Construction Software. Ph.D. Thesis, Massachusetts Institute of Technology, 2004.
- (37) Battin-Leclerc, F. Development of Kinetic Models for the Formation and Degradation of Unsaturated Hydrocarbons at High Temperature. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2072–2078.
- (38) Vandewiele, N. M.; Geem, K. M. V.; Reyniers, M.-F.; Marin, G. B. Genesys: Kinetic Model Construction Using Chemo-Informatics. *Chem. Eng. J.* **2012**, *207–208*, 526–538.

- (39) Badowski, T.; Molga, K.; Grzybowski, B. A. Selection of Cost-Effective Yet Chemically Diverse Pathways From the Networks of Computer-Generated Retrosynthetic Plans. *Chem. Sci.* **2019**, *10*, 4640–4651.
- (40) Garey, M. R.; Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*; Freeman: New-York, 1979.
- (41) Lee, J.; Bonaduce, I.; Modugno, F.; Nasa, J. L.; Ormsby, B.; van den Berg, K. J. Scientific Investigation Into the Water Sensitivity of Twentieth Century Oil Paints. *Microchem. J.* **2018**, *138*, 282–295.
- (42) Orlova, Y. Graph-theoretical approach to algorithmic construction of complex reaction networks. Ph. D. Thesis, University of Amsterdam, 2020 (<https://hdl.handle.net/11245.1/e263a9d3-cca7-4185-9042-23718957269c>).
- (43) van Driel, B. A.; van den Berg, K. J.; Smout, M.; Dekker, N.; Kooyman, P. J.; Dik, J. Investigating the Effect of Artists' Paint Formulation on Degradation Rates of TiO₂-Based Oil Paints. *Heritage Sci.* **2018**, *6*, 21.
- (44) Labshop—Chemicals, Restoration and Art Supplies, 2021. www.labshop.nl.
- (45) Modugno, F.; Gianvincenzo, F. D.; Degano, I.; van der Werf, I. D.; Bonaduce, I.; van den Berg, K. J. On the Influence of Relative Humidity on the Oxidation and Hydrolysis of Fresh and Aged Oil Paints. *Sci. Rep.* **2019**, *9*, No. 5533.
- (46) Oliphant, T. E. *A Guide to NumPy*; Trelgol Publishing: USA, 2006; Vol. 1.
- (47) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (48) Virtanen, P. et al. SciPy 1.0-Fundamental Algorithms for Scientific Computing in Python. 1907, arXiv:physics/1907.10121. arXiv.org e-Print archive. <http://arxiv.org/abs/1907.10121>.
- (49) McKinney, W. In *Data Structures for Statistical Computing in Python*, Proceedings of the 9th Python in Science Conference, 2010; pp 51–56.
- (50) *Exact Mass Calculator, Single Isotope Version*, 2021. <https://www.sisweb.com/referenc/tools/exactmass.htm>.
- (51) Kryven, I.; Duivenvoorden, J.; Hermans, J.; Iedema, P. D. Random Graph Approach to Multifunctional Molecular Networks. *Macromol. Theory Simul.* **2016**, *25*, 449–465.
- (52) Kryven, I. Analytic Results on The Polymerisation Random Graph Model. *J. Math. Chem.* **2018**, *56*, 140–157.
- (53) Kryven, I. Bond Percolation in Coloured and Multiplex Networks. *Nat. Commun.* **2019**, *10*, No. 404.
- (54) Porter, N. A.; Caldwell, S. E.; Mills, K. A. Mechanisms of Free Radical Oxidation of Unsaturated. *Lipids* **1995**, *30*, 277–290.
- (55) Soucek, M.; Khattab, T.; Wu, J. Review of Autoxidation and Driers. *Prog. Org. Coat.* **2012**, *73*, 435–454.
- (56) Orlova, Y.; Harmon, R. E.; Broadbelt, L. J.; Iedema, P. D. Review of the Kinetics and Simulations of Linseed Oil Autoxidation. *Prog. Org. Coat.* **2021**, *151*, No. 106041.
- (57) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289.
- (58) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem., Int. Ed.* **2016**, *55*, 5904–5937.
- (59) Jaworski, W.; Szymkuć, S.; Mikulak-Klucznik, B.; Piecuch, K.; Klucznik, T.; Kaźmierowski, M.; Rydzewski, J.; Gambin, A.; Grzybowski, B. A. Automatic Mapping of Atoms Across Both Simple and Complex Chemical Reactions. *Nat. Commun.* **2019**, *10*, No. 1434.
- (60) Frankel, E. *Lipid Oxidation*; Oily Press Lipid Library Series; Elsevier Science, 2014.
- (61) Sikorski, Z.; Kolakowska, A. *Chemical and Functional Properties of Lipids; Chemical & Functional Properties of Components*; CRC Press, 2010.
- (62) Lazzari, M.; Chiantore, O. Drying and Oxidative Degradation of Linseed Oil. *Polym. Degrad. Stab.* **1999**, *65*, 303–313.
- (63) Ullmann, J. An Algorithm for Subgraph Isomorphism. *J. Assoc. Comput. Mach.* **1976**, *23*, 31–42.
- (64) van Dam, E. P.; van den Berg, K. J.; Gaibor, A. N. P.; van Bommel, M. Analysis of Triglyceride Degradation Products in Drying Oils and Oil Paints Using LC-ESI-MS. *Int. J. Mass Spectrom.* **2017**, *413*, 33–42. SI: Nico Nibbering Issue.
- (65) Hui, S.-P.; Sakurai, T.; Ohkawa, F.; Furumaki, H.; Jin, S.; Fuda, H.; Takeda, S.; Kurosawa, T.; Chiba, H. Detection and Characterization of Cholesteryl ester Hydroperoxides in Oxidized LDL and Oxidized HDL by Use of an Orbitrap Mass Spectrometre. *Anal. Bioanal. Chem.* **2012**, *404*, 101–112.
- (66) Orlova, Y.; Gambardella, A. A.; Harmon, R. E.; Kryven, I.; Iedema, P. D. Finite Representation of Reaction Kinetics in Unbounded Biopolymer Structures. *Chem. Eng. J.* **2021**, *405*, No. 126485.