

UvA-DARE (Digital Academic Repository)

Predicting the mass spectrum of polymerizing linoleates using weighted random graph modeling

van 't Hoff, T.E.; Orlova, Y.; Harmon, R.E.; Iedema, P.D.

DOI

[10.1016/j.cej.2023.145264](https://doi.org/10.1016/j.cej.2023.145264)

Publication date

2023

Document Version

Final published version

Published in

Chemical engineering journal

License

CC BY

[Link to publication](#)

Citation for published version (APA):

van 't Hoff, T. E., Orlova, Y., Harmon, R. E., & Iedema, P. D. (2023). Predicting the mass spectrum of polymerizing linoleates using weighted random graph modeling. *Chemical engineering journal*, 473, Article 145264. <https://doi.org/10.1016/j.cej.2023.145264>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Predicting the mass spectrum of polymerizing linoleates using weighted random graph modeling

Tamika E. van 't Hoff^{a,*}, Yuliia Orlova^a, Rebecca E. Harmon^{a,b,1}, Piet D. Iedema^a

^a University of Amsterdam, Van 't Hoff Institute for Molecular Sciences, Amsterdam 1098 XH, The Netherlands

^b Northwestern University, Department of Chemical and Biological Engineering, Evanston, IL 60208, USA

ARTICLE INFO

Keywords:

Biopolymer
Complex reaction network
Polymerization kinetics
Random graph theory
Molar mass distribution

ABSTRACT

Biopolymers and biopolymer networks that form via autoxidation, like in drying of oil paint or fat degradation in food components, contain a large variety of monomeric building blocks. While the monomer variety complicates the modeling itself, obtaining experimental validation of infinite polymer networks is inherently difficult as well. A new model is developed, where an automated reaction network generation (ARNG) procedure is used to automatically generate the monomer components, structures and masses, and their reactions. This methodology is combined with random graph (RG) modeling to predict global polymer properties: distributions of numbers of monomer units and molar masses, gel point and gel fraction. This computational framework is applied to two model systems for linseed oil paint binder: the polymerization of ethyl linoleate (EL) and methyl linoleate (ML). A novel method was constructed to deal with the variability of monomer masses that complicates inferring molar mass from monomer number distribution. By modeling the polymer as a weighted random graph where the nodes contain information about the monomer masses in the system, the total weight of the finite connected components is computed. The predicted mass spectrum of finite connected components is used for validation with experimental data. A size exclusion chromatography (SEC) trace of ML is employed, which after calibration using the proposed framework, proves consistency between model and SEC data. The model provides a practical approach to both characterize complex biopolymers as polymers in terms of molar mass distribution and gel point, while preserving the information down to the level of monomeric units.

1. Introduction

The varying sources of complexity in polymer modeling make the property prediction of non-linear polymers through their microstructure—molecular weight distribution, branching, tacticity, *etc.*—a challenging problem. In the present paper, the complexity of the modeling problem arises from both the consideration of a polymer network and the extensive size of the reaction network of a biopolymer that results in thousands of different monomeric building blocks [1]. Polymer networks have been the subject of significant modeling efforts since the days of Gordon [2] and Dušek [3,4], but many problems are left unresolved. Yet, in spite of the existence of various modeling methods [5–9], only few made an attempt of mathematically spanning the full development of polymer networks; from the early stages of more tractable finite molecules, through the gel point, ending up with the formation of one ‘giant component’ or a molecule of infinite size [10,11]. While properties of finite polymers are often experimentally

accessible, obtaining similar validation data for properties of polymer networks is much harder.

Additional problems arise due to the nature of biopolymers like oils and proteins. Whereas industrial copolymers generally consist of a few repeating monomers, biopolymers contain many more building blocks. Especially autoxidative polymerization processes are characterized by large and complex reaction schemes [12–15] producing many unique ‘monomeric’ units with different molecular formulae. This divergence from Flory's ‘repeated units’ idea [5] makes experimental validation more complicated with differing molar masses and molecular structures. Hence, the prediction of molecular weight distributions or mass spectra from models and their comparison with experimental data is extremely problematic. In the present paper, this problem is addressed.

Ultimately, our models should deal with both monomer variety and infinite polymer networks, and clarify the microstructure of biopolymers in terms of both chemical details, presence and distribution of functional groups, and the characterization of the polymer. The latter

* Corresponding author.

E-mail address: t.e.vanthoff@uva.nl (T.E. van 't Hoff).

¹ Present address: The Ohio State University, William G. Lowrie Department of Chemical and Biomolecular Engineering, Columbus, OH 43210, USA.

<https://doi.org/10.1016/j.cej.2023.145264>

Received 1 May 2023; Received in revised form 31 July 2023; Accepted 4 August 2023

Available online 8 August 2023

1385-8947/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature

ARNG	Automated reaction network generation
EL	Ethyl linoleate
ESI-MS	Electrospray ionization-mass spectrometry
LO	Linseed oil
ML	Methyl linoleate
ODE	Ordinary differential equation
RG	Random graph
SEC	Size exclusion chromatography
$c_{\mathbf{k},J}$	Concentration monomer species with degree vector \mathbf{k} and molar mass α_j
\mathbf{e}_i	Standard basis vector for half-edge type i
F_{max}	Functionality
\mathbf{k}	Degree vector
k_x	Degree half-edge type x
m	Connected component weight
m_{mon}	Average monomer mass
M	Number of unique molar masses
N	Number of half-edge types
N_x	Number of crosslinks
O_{add}	Ether crosslinks produced via addition reactions
O_{term}	Ether crosslinks produced via termination recombination reactions
OO_{add}	Peroxy crosslinks produced via addition reactions
OO_{term}	Peroxy crosslinks produced via termination recombination reactions
\mathbf{P}	Permutation matrix
s	Connected component size
$u(\mathbf{k})$	Probability degree distribution
$u_j(\mathbf{k})$	Mass specific probability degree distribution
$u(\mathbf{k}, J)$	Two-dimensional probability degree/mass distribution
$u_i(\mathbf{k}, J)$	Excess two-dimensional probability degree distribution of edge-type i
$U_J(\mathbf{z})$	Weight specific generating function of $u(\mathbf{k}, J)$
$U_{i,j}(\mathbf{z})$	Weight specific generating function of $u_i(\mathbf{k}, J)$
$w(m)$	Weight distribution of finite connected components
$w_{sa}(m)$	Weight distribution of finite connected components without size influences
$w(s, m)$	Size-weight distribution on finite connected components
$W(x, y)$	Generating function of size-weight distribution $w(s, m)$
$W_i(x, y)$	Generating function of the excess probability degree distribution of edge-type i
α	Mass vector
$\mathbb{E}[k_i]$	First partial moment of degree distribution $u(\mathbf{k}, J)$
$\sigma_{i,x}$	Pairing rule for directed, in-coming half-edge for crosslink type x
$\sigma_{o,x}$	Pairing rule for directed, out-going half-edge for crosslink type x
$\sigma_{u,x}$	Pairing rule for undirected half-edge for crosslink type x

pertains to not only typical microstructural properties such as molar mass distribution and branching topology but also to network structural aspects like cycles, dangling ends and glass transition temperature T_g . An important application of such a model is paint films. Recent research reveals that old oil paintings suffer from degradation due to chemical

reactions between metals present in pigment, mainly zinc and lead, and free fatty acids originating from the binding medium, or linseed oil (LO) [16]. The autoxidation process produces reactive carboxylic acid groups that turn out to be crucial in the degradation mechanism. Hence, the ability to predict the generation, consumption and distribution of these acid groups over long time scales in relation to network material properties would be helpful from the perspective of art preservation.

In the present paper, we model polymer networks as random graphs (RG), as initially introduced by Kryven [17] and later on applied to various systems [1,11,18–21]. By defining the *degree* of a monomer as the number of covalently connected monomers within the polymer network, the RG approach generates networks of interconnected monomers that assemble according to the ‘degree distribution’. Effectively, we assume that the connectivity of all monomeric species captures the essential chemical properties of the system to predict accurate material properties of the polymer.

The biopolymers we consider here are ethyl linoleate (EL) and methyl linoleate (ML), which are mono-glycerides bonded via ethyl- and methyl-ester groups to linoleic acid, a reactive fatty acid present in LO and shown in Fig. 2. In both cases, light and oxygen induce an autoxidative polymerization process by promoting the formation of free-radicals in the system that cause a cascade of reactions. Among the cascade are hydrogen abstraction of the reactive bis-allylic H-atoms, oxidation reactions, and finally crosslink formation by radical-radical recombination and addition to double bonds. There are various types of crosslinks that form during the polymerization of EL and ML, and some of them are directional in nature because of the reactions involved. With crosslink formation oligomers and finally polymer networks emerge [12]. The polymerization in both EL and ML solely involves the unsaturations of linoleic acid and the respective ethyl and methyl group do not influence the reactivity. Hence the involved chemistry is identical.

In previous work, Schamboeck and Kryven [11] were able to generate a ‘spectrum’ of polymer sizes for acrylates, counting the number of monomeric units in oligomers and polymers using the RG approach. Here, we expand the pre-existing RG approach to include information about the molecular mass of the monomers and present a method to compute the molecular weight distribution and mass spectrum at a given time, which opens the possibility to compare with measured data of EL and ML. To achieve this, the random graph is first transformed into a weighted graph where the nodes are weighted with the monomer mass. Second, two-dimensional auxiliary distributions supplement the equations used for polymer size distribution.

In addition to the expansion of the RG-methodology, this study presents also a novel modeling framework that combines automated reaction network generation (ARNG) with RG-modeling and applies it to biopolymerization. The article discusses the chemical system and main assumptions, along with an explanation of the weighted RG. In the results section, the effect of graph edge directionality and molar mass weighting assumptions are shown, and chemical interpretations of resulting patterns in the weight distribution are discussed. To validate the methodology, we compare the dimer concentrations produced with the ARNG-model and with the RG-model. The ARNG-model can predict the concentration of dimer species, but higher order oligomer concentration prediction requires the extension of the RG-model. Therefore, the RG-model is validated via comparison with the ARNG-model. Additionally, the article compares an experimental size exclusion chromatography-trace for ML with model predictions using model-supported calibration.

2. Modeling approach

Autoxidative polymerization involves thousands of components and millions reactions, yielding numerous unique monomer units. Consequently, the assembly of monomer units into a polymer structure

requires a new, dedicated modeling approach: a combination of automated reaction network generation (ARNG) and random graph (RG) modeling. The ARNG-model generates a full reaction network with a wide variety of monomer units from a starting conformation and the reactions that apply during the autoxidation. We have used a version of the ARNG-model, that was first published in [1,20,21] and later updated with data from [22]. The pre-existing ARNG-model is extended with RG-modeling to allow the prediction of the polymer material properties. RG-modeling requires input pertaining to the connectivity of all monomer species: concentrations through time of the unique monomeric species that form the output of the ARNG-model. Since we desire to compute the biopolymers' weight distribution, we need to include the variety of monomer units' molar masses that are produced by the autoxidation. This section discusses both models in depth.

2.1. Automated reaction network generating model

Prior to the RG-modeling, we need to obtain the concentration profiles of all monomeric species that are involved in the process of polymerization. For that we use the automated reaction network generation (ARNG) modeling methodology [23,24] that has been adapted for the case of complex polymerization processes [1,20–22]. The methodology is briefly and schematically represented in Fig. 1. In the ARNG, a monomer is represented as a molecular graph, and a reaction is represented as a transformation applied to reactive sites of a molecular graph called reactive patterns. The complete list of 49 reactive pattern structures is displayed in Fig. 1 of the SI. This modeling method uncovers reaction networks that are usually intractable by hand and involve thousands of molecular species and millions of reactions. Considering that by definition polymers are infinite in size, this methodology does not uncover explicit polymeric structures but rather the structures of their building blocks: monomeric species with their adjacent crosslinks. In this way, the methodology uncovers a finite-sized network of reactions that happen solely between the monomeric units that eventually constitute to infinite in size polymers. The ARNG approach adapted for polymers produces the concentration profiles of monomeric species with their adjacent crosslinks, or half-edges, and aligns perfectly with the proposed RG-modeling. Synergy between the ARNG- and the RG-modeling methodologies has been demonstrated in [1,20], and in this paper we show how including the bi-directional nature of crosslinks helps to strengthen the results.

In this paper we use EL and ML, see the structural formulas in Fig. 2, as two model systems of linseed oil. Since both molecules only differ by their head group and have exactly the same reactive sites, they share a library of reaction rules that capture their autoxidation pathways. While most of the pathways have been explored in previous papers [12,25–28], here we expand the library with the asymmetric character of the present addition reactions of both alkoxy radical and peroxy radical.

Starting with the initial structures of EL and ML monomers, the ARNG subsequently applies reaction rules to their reactive sites and generates a plethora of monomeric units that constitute resulting polymers. The result of this step is a reaction network, which is a graph with two types of nodes: monomeric species and reactions. By assigning the weights to these nodes: initial concentration to monomers and rate coefficients to reactions, the reaction network is then automatically translated into a system of ordinary differential equations (ODEs). The solution of the ODEs provides the concentration profiles of all monomeric species present in a reaction network. These concentrations are then lumped according to different types of crosslinks and functional groups. Lumped concentrations provide valuable information about the state of the system at every point of time. The model generates monomeric units as unconnected components that require assembly to describe polymer species.

Briefly summarizing some chemical features, both EL and ML polymerize around the unsaturations in the fatty acid tails [21]. Oxygen and

light combined with trace amounts of an initiator induce accelerated radical formation in the system that in turn initiate the polymerization by H-abstraction of the bis-allylic H-atom [25,29,30]. In practice, in oil-based paints, a dryer catalyzing the autoxidation reactions is often used to accelerate drying and enhance the initiation. In the present model we choose to dispense with a drying agent to avoid more complexity in the already complex chemistry of the autoxidation process. After initiation, the process proceeds by several elementary reactions that include oxidation, hydroperoxide decomposition and addition reactions [1,26,27,31]. Finally, the polymerization finishes with two types of termination reactions that remove radicals from the system by either the production of oxygen (disproportionation, or Russell termination) [28] or the creation of a crosslink (recombination) [32]. In experimental setups, it is common to add a cobalt drier to the unreacted system to accelerate the polymerization [1].

Within the polymerization of EL and ML, crosslinks form by termination as stated above, and by addition to monomers with conjugated double bonds. The crosslinks produced via addition reactions are *asymmetric* crosslinks, as opposed to the *symmetric* crosslinks that termination produces. The number of unsaturations and their position in the fatty acid tail determines the maximum number of crosslinks or degree a monomer forms and is defined as the *functionality* F_{max} [1]. For polymerizing ML and EL, $F_{max} = 3$.

The ARNG-model developed in our group [1] has been updated with kinetic rate parameters matching an EL polymerization experiment at 70 °C [22]. We modeled the system for 1000 h of drying, with a starting concentration EL of 3.71 mol/L. Although during polymerization the rates decrease as a result of solidification, we assume the rate parameters to be constant [1]. This assumption significantly reduces the complexity of the problem and allows to run the calculations in reasonable amount of time. The complexity of the algorithm is discussed in the results section. Further information about initial conditions and simulation details are discussed in Orlova et al. [21] and Harmon [22]. The reactions with their corresponding rate coefficients have been listed in the SI in Table 1.

To reduce computational effort in ARNG, it employs simplified structural representations of the starting monomers EL and ML, see Fig. 2 in SI, where the unreactive chain on one side of the reactive part and the unreactive ester on the other are not included in the ARNG graph representation. The full structural formulas of EL and ML are displayed in Fig. 2. They contain the bis-allylic H as the reactive part, which in the ARNG representation is reactive pattern number 7 in Fig. 1 of the SI. When computing the mass spectrum from ARNG, the masses of these two different tails have to be accounted for. This situation is explained in the SI in Fig. 2.

To further clarify the role of the reactive patterns in the ARNG, while discussing ARNG results we will refer to the relevant patterns that are part of the many intermediate and final species for the remainder of this paper.

2.2. Random graph model with molar mass weighted nodes

In the RG approach, a polymer is represented as a graph, where nodes are monomers and edges are crosslinks between monomers. The degree distribution of a polymer network describes the connectivity patterns between monomers [17]. Although the degree distribution only describes local interactions, the assembly of the polymer as a RG allows for the computation of global properties of the polymer network. A network may contain multiple types of edges, so an accurate description of the connectivity of the nodes requires a degree vector rather than a singular degree. Degree vector \mathbf{k} of a monomer is defined as the number of unique crosslinks by which it is bound to others. This makes \mathbf{k} a vector with $\mathbf{k} = (k_1, \dots, k_N)$ and N the number of unique half-edge types. Consequently, we define $u(\mathbf{k})$ as the probability distribution depending on degree vector \mathbf{k} . Fig. 3 shows on the LHS the translation of a dimer species into a connected component in the RG

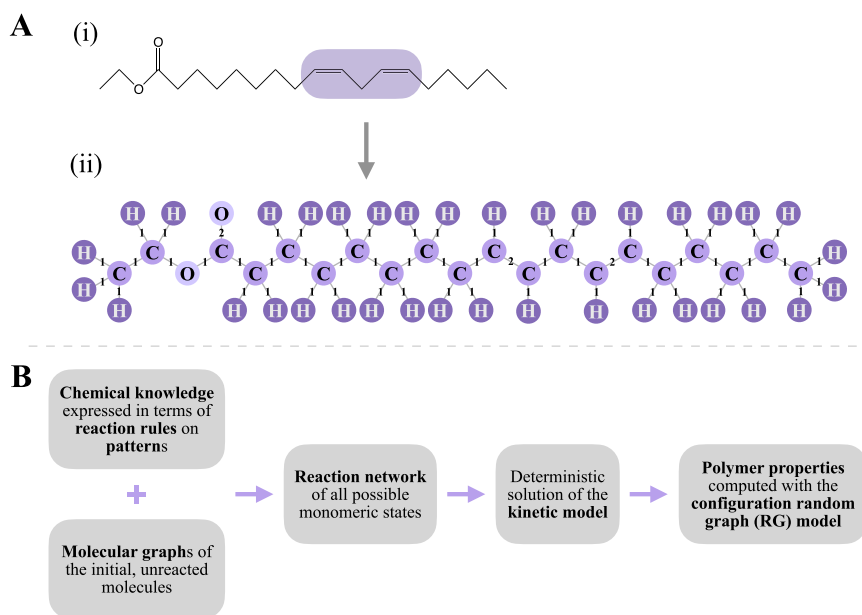


Fig. 1. A schematic representation of the automated reaction network generation (ARNG) methodology. A: The transformation of the chemical structure of Ethyl Linoleate (i), where the reactive site is highlighted (purple), into the molecular graph (ii) required for the ARNG. B: The key steps in ARNG [1]. All reactive patterns are listed in Fig. 1 of SI.

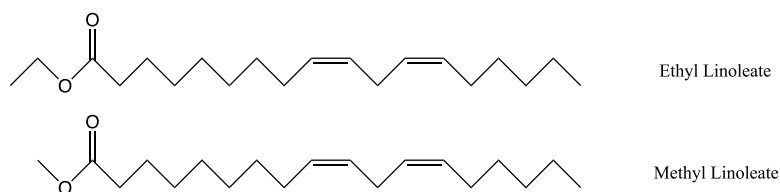


Fig. 2. The molecular structure of methyl linoleate and ethyl linoleate. The only reactive part of these monomers is the bis-allylic H, which is part of reactive pattern number 7, see Fig. 1 in the SI.

and an example node in a system with $N = 6$ dimensions and degree vector $\mathbf{k} = [1, 0, 0, 1, 0, 1]$ on the RHS. The probability $u(1, 0, 0, 1, 0, 1)$ represents the probability of finding a node with 1 half-edge of type 1, 1 half-edge of type 4, etc. A graph can have two edge-types: directed edges and undirected edges. The choice of considering a connection between two nodes generally depends on the relation between the depicted nodes.

In the case of polymerizing EL and ML, the ARNG-model [1] includes two types of crosslinking reactions, radical-radical termination and addition reactions, that produce three types of chemical crosslinks: peroxy, ether and alkyl, see Table 1 in the SI. Alkyl crosslinking happens to a negligible amount [1] and cross-termination between alkoxy and peroxy radical containing species is chemically improbable.

The nature of the crosslink reactions dictates which type of edge is needed for an accurate description. The crosslinks formed by two equally reactive reactants, termination reactions, are *symmetric* and can be represented by undirected edges, or two identical undirected half-edges. This is in contrast with the crosslinks produced via addition reactions, which are *asymmetric* and need to be depicted as directed edges. Since the crosslink originates from the radical containing species, the out-going half-edge is placed on the radical containing species and the in-coming half-edge on the conjugated double bond containing species, similarly to acrylate polymerization [11]. In the case of polymerizing EL/ML, both termination and addition produce ether and peroxy crosslinks, so for both crosslink types we require one undirected and two directed half-edges, resulting in $N = 6$ half-edge types. With $u(\mathbf{k})$ computed for every t , a graph constructed from $u(\mathbf{k}, t)$ allows to compute the properties of macromolecules at time t , for example the size distribution of polymers [11,18]. However, a direct comparison

between size distributions of biopolymers and experimental data is difficult [33].

To enable the computation of a molecular mass distribution with the RG-model, the graph needs to contain information pertaining molar mass distribution of the monomers. Therefore, we combine both the degree vector and the molar mass of a monomer into a two-dimensional degree distribution that produces a *weighted* graph after assembly. In contrast to a graph that is solely dependent on the connectivity of the nodes, in a weighted graph the nodes are weighted according to the mass of the represented monomers. With mass vector $\alpha = (\alpha_1, \dots, \alpha_M)$, where M denotes the number of unique molar masses present in the system, the two-dimensional degree distribution $u(\mathbf{k}, J)$ is given by:

$$u(\mathbf{k}, J) = \frac{c_{\mathbf{k}, J}}{\sum_{\mathbf{k}} \sum_J c_{\mathbf{k}, J}}, \quad (1)$$

where $c_{\mathbf{k}, J}$ denotes the concentration of monomer species with crosslinks matching degree vector \mathbf{k} and molar mass α_j . Similarly, $u(\mathbf{k}, J)$ provides the probability of a node in the graph with degree vector \mathbf{k} and weighted with mass α_j . Unlike the size of a monomer building block, the mass typically assumes non-integer values and is distributed in a non-equidistant manner, especially in biopolymers. Although it is possible to include the molar masses at true chemical resolution in mass vector α , it would result in a considerably long vector. Hence, we apply a coarse-graining technique by assuming integer molar mass values in an effort to reduce the computational cost.

A graph that accurately represents the polymer network is assembled by matching half edges. Undirected peroxy half-edges may only connect to other undirected peroxy half-edges and in-coming ether half-edges only to out-going ether half-edges. We follow the same

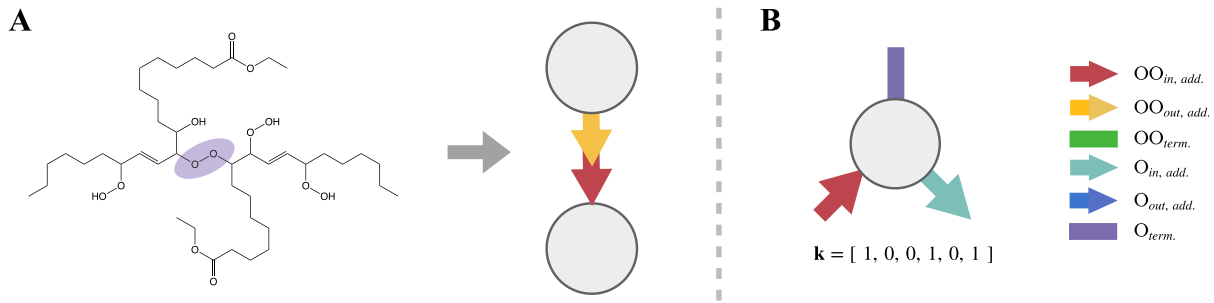


Fig. 3. A: The transformation from a dimer species connected via a peroxy crosslink as result of an addition reaction. Each monomer is represented as a node, and the asymmetric peroxy crosslink as a directed edge built from a peroxy-out half-edge and a peroxy-in half-edge. On the right hand side (B), a possible node in a system with $N = 6$, where N denotes the number of unique half-edge types and degree vector \mathbf{k} describes the connectivity of said node. Probability $u(1,0,0,1,0,1)$ gives the probability of finding nodes with connectivity \mathbf{k} , i.e. 1 peroxy in-edge, 1 ether out-edge, and 1 undirected ether half-edges. The six types of crosslinks correspond to the chemical structures of the ‘patterns’ displayed in Fig. 1 in the SI: in order of appearance: reactive pattern numbers 43, 45, 14, 44, 46, 15.

technique to pair edges of identical colors to each other as in colored graphs [19,34] and define permutation matrix \mathbf{P} as:

$$\mathbf{P} = \begin{pmatrix} \sigma_{i,OO_{add}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{o,OO_{add}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{u,OO_{term}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{i,O_{add}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{o,O_{add}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{u,O_{term}} \end{pmatrix}, \quad (2)$$

with $\mathbf{0}$ being a zero matrix, and $\sigma_{i,x}$, $\sigma_{o,x}$ and $\sigma_{u,x}$ denoting the pairing rules for directed (i , in-coming, o , out-going) and undirected (u) half-edges for crosslink type x ($x = OO_{add}, OO_{term}, O_{add}, O_{term}$). If \mathbf{P} denotes the pairing matrix, $\mathbf{P}_{i,j} = 1$ when half-edge types i and j form a bond, while when $\mathbf{P}_{i,j} = 0$ no bond is formed. The pairing rules for both types of edges are defined by:

$$\sigma_{i,x}, \sigma_{o,x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (3)$$

$$\sigma_{u,x} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4)$$

Similarly as with the computation of the connected component size distribution in colored graphs [34], we make use of the structure of the graph to compute the weight distribution of the components. A random half-edge (including undirected, in-coming and out-going half-edges) present in the graph is chosen and followed through the connected nodes. The distribution of the nodes reached by following a randomly selected edge i defines the excess degree distribution of edge-type i :

$$u_i(\mathbf{k}, J) = (k_i + 1) \frac{u(\mathbf{k} + \mathbf{e}_i, J)}{\mathbb{E}[k_i]}. \quad (5)$$

Here \mathbf{e}_i is the standard basis vector for half-edge type i and $\mathbb{E}[k_i]$ the first partial moment of the true degree distribution $u(\mathbf{k}, J)$. With the selection of edge-type i comes an increased probability to reach nodes with more i -edges, causing the necessity of multiplication factor $(k_i + 1)$ in the formulation of u_i . Random graph formalism makes use of the generating functions of the true and excess degree distributions that apply to a given system to compute its global properties. The computation of the weight distribution of the finite connected components requires the weight specific generating functions of $u(\mathbf{k}, J)$ and $u_i(\mathbf{k}, J)$, $U_J(\mathbf{z})$ and $U_{i,J}(\mathbf{z})$ respectively, formulated as:

$$U_J(\mathbf{z}) = \sum_{\mathbf{k} > \mathbf{0}} \mathbf{z}^{\mathbf{k}} u_J(\mathbf{k}), \quad J = 1, \dots, M, \quad (6)$$

$$U_{i,J}(\mathbf{z}) = \sum_{\mathbf{k} > \mathbf{0}} \mathbf{z}^{\mathbf{k}} u_{i,J}(\mathbf{k}), \quad J = 1, \dots, M \text{ and } i = 1, \dots, N, \quad (7)$$

where $\mathbf{z} = (z_1, \dots, z_N)$, $|z_i| < 1$, $z_i \in \mathbb{C}$ and $u_J(\mathbf{k}) = u(\mathbf{k}, J)$.

To compute the total weight of the finite components, we randomly select an edge, rather than an entire component. This method inherently introduces a bias depending on the size of the components: larger components contain more edges and are thus counted more frequently. In contrast to previous work, this counting method introduces problems for the computation of the weight distribution. To remove the influence of the component size, there needs to be some correlation between component size and mass to compute the true weight distribution of a polymer network. Unfortunately, this is generally not the case for non-industrial polymers, which contain a large variety of monomer building blocks and thus monomer masses. Particularly in the case of biopolymers where the polymerization process often includes scission reactions. Therefore, we propose the computation of a bi-variate size-weight distribution, $w(s, m)$, to accurately predict the weight distribution of the polymer network. Effectively, $w(s, m)$ provides the probability of the system containing finite connected components with size s and total weight m . Although we continue to count by selecting edges and therefore count to larger components more frequently, distribution $w(s, m)$ now stores both the size and total weight of the connected components. This allows us to determine the correlation between the size and weight of the components.

The numerical computation of size-weight distribution $w(s, m)$ requires weight specific generating functions of the biased weak components, $W_i(x, y)$, that arise from the edge selection. They form along with the generating function of $w(s, m)$, $W(x, y)$, a recursive set of algebraically solvable equations:

$$W(x, y) = \sum_{J=1}^M xy^{\alpha_J} U_J(\mathbf{P}\omega(x, y)), \quad (8)$$

$$W_i(x, y) = \sum_{J=1}^M xy^{\alpha_J} U_{i,J}(\mathbf{P}\omega(x, y)), \quad (9)$$

where $\omega(x, y) := (W_1(x, y), \dots, W_N(x, y))^T$ and both $W(x, y)$ and $W_i(x, y) : \mathbb{C} \rightarrow \mathbb{C}$. While progressing through the connected component, we count both the size and weight of the node, as incorporated in Eq. (8) as power values. Since the size of a node is by definition equal to 1, no power value is needed size wise. In contrast, the mass of a nodes is generally larger than 1 and thus must be counted α_J times when the weight of a particular node is $m = \alpha_J$.

Finally, the computation of a bi-variate distribution requires a two-step Cauchy integral to solve the generating functions in Eqs. (8) and (9):

$$w(s, m) = \frac{1}{2\pi i} \oint \frac{W(x, y)}{x^{s+1} y^{m+1}} dx dy, \quad (10)$$

where we integrate $W(x, y)$ subsequently over all component sizes and all component weights. Although the resulting $w(s, m)$ is still weighted

with component size, the distribution contains information pertaining the relation between component size and weight. The correct, component size-independent weight distribution is easily found by unweighting the values with their corresponding size, i.e. taking $w(s, m)/s$.

3. Results

For the given set of reaction rules for EL/ML, the reactive patterns (Fig. 1, SI) and the corresponding kinetic coefficients (Table 1, SI), the ARNG-model (Fig. 1) generates 4012 species undergoing over 1.5 million reactions and provides concentration profiles over time for all species [22]. This ARNG-output forms the input to compute the bivariate degree-mass distribution as input to the RG-model. In this section we will discuss the transition from concentration profiles to degree distribution. Note that this transition implies a loss of information. While the ARNG-results still contain and thus allow tracking of detailed chemical information of species like structural formulas, the degree distribution species are lumped together and only their molar mass, concentration and the type and number of crosslinks are retained. We will now first discuss the bivariate degree-weight distribution and then present the full molar mass distribution computed with the extended RG-model, allowing for directionality. These outcomes are validated in two ways: by comparing to direct results concerning monomers and dimers from ARNG and by comparing to size exclusion chromatography (SEC) data for ML oligomer formation.

3.1. Bivariate degree-weight degree distribution

The bivariate degree-weight distribution is computed from the ARNG concentration profiles. This implies the lumping of ARNG-generated species, which are close together in molar mass. Table 1 shows a few examples of ARNG-generated species with varying numbers of crosslinks possessing a given degree (see Fig. 3) and molar mass. The table displays structural formulas, weights and final concentrations of four monomer species that are dominant in both the ESI spectrum [22] and in the ARNG-model: a monomer without crosslinks and with a molar mass 200 g/mol - ethyl-9-oxononanoate - and the three most dominant species with one, two and three crosslinks determining their connectivity - peroxy crosslinks were the most abundant. Note that the columns 'degree' and 'molar mass' determine its place in the bivariate degree distribution. The last column of Table 1 also shows the total amount of species with corresponding connectivity N_x . Here we did not distinguish between molar mass and whether the species contained radicals.

By lumping the species according to their connectivity and unique molar mass, the species' concentration profiles transform into a normalized two-dimensional degree distribution $u(\mathbf{k}, J)$, with degree vector \mathbf{k} and mass vector α_j . The polymer system contains 16,192 unique $[\mathbf{k}, \alpha_j]$ vectors. Focusing solely on the connectivity of the monomers, the first dimension in $u(\mathbf{k}, J)$, we find 64 unique degree vectors \mathbf{k} whose probability values are displayed in Fig. 4. For visualization purposes, colors group the profiles according to the degree of ether crosslink formation via termination recombination reactions. Regarding the monomer weights present in the system, or the second dimension in $u(\mathbf{k}, J)$, a total of 253 unique monomer masses are present. Fig. 5 shows that the variation of masses at time point $t = 27.3$ h ranges from 18 g/mol to 406 g/mol. We chose this particular time point as the autoxidation progressed sufficiently to encompass scission reactions, resulting in a large variety of monomer species. Note that Figs. 4 and 5 combined form the full two-dimensional degree distribution $u(\mathbf{k}, J)$ and in fact show the partial summations $\frac{\sum_j c_{\mathbf{k}, \alpha_j}}{\sum_{\mathbf{k}} \sum_j c_{\mathbf{k}, \alpha_j}}$ and $\frac{\sum_{\mathbf{k}} c_{\mathbf{k}, \alpha_j}}{\sum_{\mathbf{k}} \sum_j c_{\mathbf{k}, \alpha_j}}$, respectively.

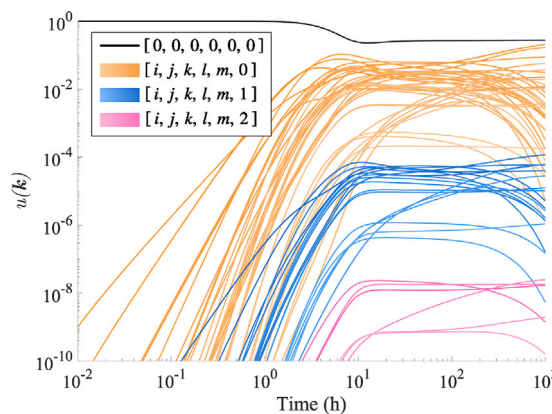


Fig. 4. Degree distribution $u(\mathbf{k})$ of ethyl linoleate, with $\mathbf{k} = [k_{i,OO}, k_{o,OO}, k_{u,OO}, k_{i,O}, k_{o,O}, k_{u,O}]$ and i , in-coming, o , out-going, u , undirected edge for crosslink type OO peroxy and O ether. The color-coding indicates profiles grouped according to the degree of ether crosslink formation via termination recombination reactions. Even without including molar mass as identifying characteristic, degree distribution $u(\mathbf{k})$ contains many degree vectors with significant probability values.

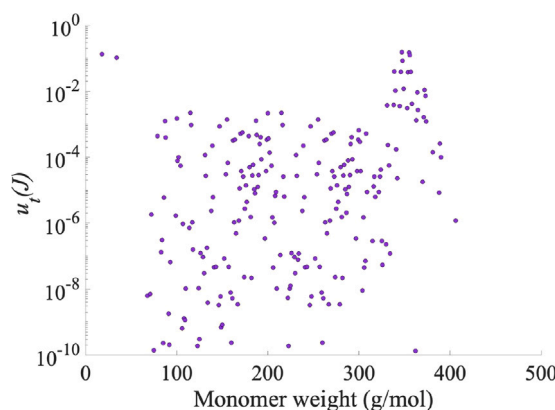


Fig. 5. The probability distribution $u(J)$ of the unique monomeric masses present in polymerizing ethyl linoleate at $t = 27.3$ h. The monomer connectivity is not taken into account. The figure shows large variation in prevalence of species with different mass. The monomer mass varies from 18 g/mol to 406 g/mol.

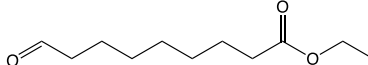
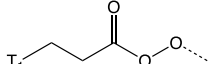
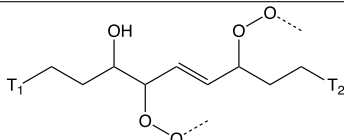
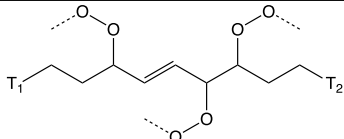
3.2. Molar mass distribution of polymerized EL

To compute the true molar mass distribution, the new RG-algorithm is used in combination with the unweighting procedure described above. To this end we had to compute the bi-variate size-weight distribution $w(s, m)$ to remove the influence of component size of the component mass spectrum from Eqs. (8), (9), (10) in Section 2.2. To indicate the effect of the size factor in the probabilities on long weight ranges, Fig. 6 shows the double mass weighted distribution for weights up to 10,000 g/mol. While $w(m)$ is the direct product of Eq. (10) after summing over all component sizes, $w_{su}(m)$ is the size-unweighted distribution. This upper limit corresponds to an average connected component size of 30, taking $m_{mon} = \mathbb{E}[m]$. We observe that the error between $w_{su}(m)$ and $w(m)$ increases with component weight. In the range from 0 to 400 g/mol the values of both distributions do not display significant differences, since these mass values relate solely to unconnected nodes and thus a size of 1. At higher mass ranges the probabilities diverge from each other with an approximate ratio of m/m_{mon} .

Looking at Fig. 6, we observe a repeating pattern in the mass distribution that corresponds to estimated component sizes obtainable by dividing the mass by the average mass of a monomer unit m_{mon} .

Table 1

An overview of the number of crosslinks N_x , degree k (see Fig. 3), molar mass, final concentration, and structural formulas of four species with highest concentrations from ARNG: a monomer (zeros crosslinks) and species with one, two and three crosslinks that are indicated with broken lines. For each of the categories peroxy crosslinks created via termination recombination – the undirected peroxy crosslink – are most abundant. The most important reactive patterns (Fig. 1, SI) present in the molecules are given in the next column. The columns ‘degree’ and ‘molar mass’ determine the position of these species in the bivariate degree distribution as input to RG-model. Note that one node of the bivariate degree input may contain several different species of similar weight, but entirely different structural formulas. The structural formulas contain Tails T_1 and T_2 that are defined in Fig. 2 of SI. The final column shows the total amount of species with connectivity N_x , without taking mass, radical and type of crosslinks into account.

Number of crosslinks, N_x	Degree k	Molar mass (g/mol)	Final concentration (mol/L)	Structural formula	Reactive patterns present	Number of species with N_x from ARNG
0	[0,0,0,0,0,0]	200	0.322		18	403
1	[0,0,1,0,0,0]	99	0.533		14, 27	1432
2	[0,0,2,0,0,0]	324	0.283		13, 14, 14	1610
3	[0,0,3,0,0,0]	307	0.635		14, 14, 14	567

Doing so, the yellow curves approximately coincide with the purple peaks. This demonstrates, in this particular case of EL polymerization, that although the masses within the degree distribution vary significantly (see Fig. 5), approximating the correct, size-unweighted mass distribution by division of $w(m)$ with m_{mon} still produces accurate values for $w_{su}(m)$. In other cases, where the monomer mass' range is greater so that oligomer masses overlap more, approximating $w_{su}(m)$ in a similar fashion will not produce sufficiently accurate values. For instance, in the case of linseed oil polymerization, many different types of fatty acid chains are present, one being linoleic acid, each with their own autoxidation processes and subsequent monomeric species. This will create a wealth of different masses of monomer units with greater overlap, which indeed would prevent to infer sizes from masses as for EL in Fig. 6.

To indicate the overlap of molar mass for polymer sizes, Fig. 7 shows $w_{su}(m)$ where we color code each component size. The molar mass range of any polymer size is between a clear upper and lower bound. Particularly when the upper limit is reached, distribution $w_{su}(m)$ falls sharply, while the minimal values show a slope that increases in length with component size. The peak pattern can be interpreted in chemical terms. The highest peaks are at regular distances, corresponding to EL autoxidation products, predominantly peroxidic species. The patterns of peaks with low molecular weights are created by combination of EL scission products. Even in smaller oligomers as dimers (red 154–778 g/mol, Fig. 7) and trimers (yellow 241–1150 g/mol, Fig. 7), a spread of peaks is discernible, rather than a few individual peaks. The spread of these clusters increases with the oligomer order. Moreover, at higher order oligomers, these clusters start to overlap, making it impossible to identify the peaks that are part of a specific oligomer size. This is directly related to the large mass variety and the combining of the monomer types into higher order structures. From literature, this agrees with observations made from experimental mass spectra, where hyphenated techniques are commonly used to provide extensive characterization [33,35].

The molar mass distribution computed here was based on the 6D formulation of the bivariate degree-weight distribution as shown in Fig. 3. If we would have ignored the distinction between directed and undirected edges, i.e. ignoring the difference between crosslinks made by termination crosslinking and addition crosslinking, we end up with

a simpler 2D formulation requiring less computational effort. We have compared the 2D with the 6D approach and found clear but not very large difference in the results. This issue is dealt with in greater detail in the SI with Fig. 3 in the SI showing both 6D and 2D concentrations.

3.3. Model validation: comparing RG with ARNG

As a first validation method we compare both the RG-model results and the ARNG-model results for monomer and dimer species of EL. Although the ARNG-model solely generates the monomer building blocks present in the polymer network, combining monomer species with just one crosslink results in all possible dimers. Even though there are only three different crosslinking reactions (OO-termination, OO- and O-addition), the existence of 1432 different species with just one crosslink already leads to a combinatorial explosion in the number of different dimers, amounting to over 70,000. Hence, obtaining even higher order oligomers from ARNG becomes an infeasible task. In contrast to the RG-model, the ARNG-model provides molar mass data with chemical resolution. Hence, the data from ARNG requires coarse-graining to the mass grid chosen for RG of 1 g/mol to enable the comparison between the two models. After coarse-graining we observe an exact agreement for monomers and dimers up to masses of 800 g/mol.

Reducing the resolution of the ARNG data involves arranging more than 70,000 dimer species on a discrete range between 400 and 800 g/mol with increments of 1 g/mol. This implies the overlap of mass peaks of dimers with identical mass, isomers, or masses close together per data point. Harmon compares the ARNG spectrum with a spectrum recorded with electrospray ionization-mass spectrometry (ESI-MS), which are given in the SI in Fig. 4. A direct quantitative comparison turned out to be impossible, in part due to the complex relationship between peak value and concentration of a species in ESI-MS. The ESI-MS signal of a species and its concentration are not always linearly proportional as consequence of the ionization method and the loss of volatile species that escape the system before detection. However, most of the monomer and dimer peaks that are dominant in ESI-MS were found to be dominant in the ARNG-model, and vice versa, those dominant in the model were also found in ESI-MS. To elaborate on the overlap of dimer masses, Fig. 8 provides a detailed part of the full spectrum from 690–720 g/mol and in particular the

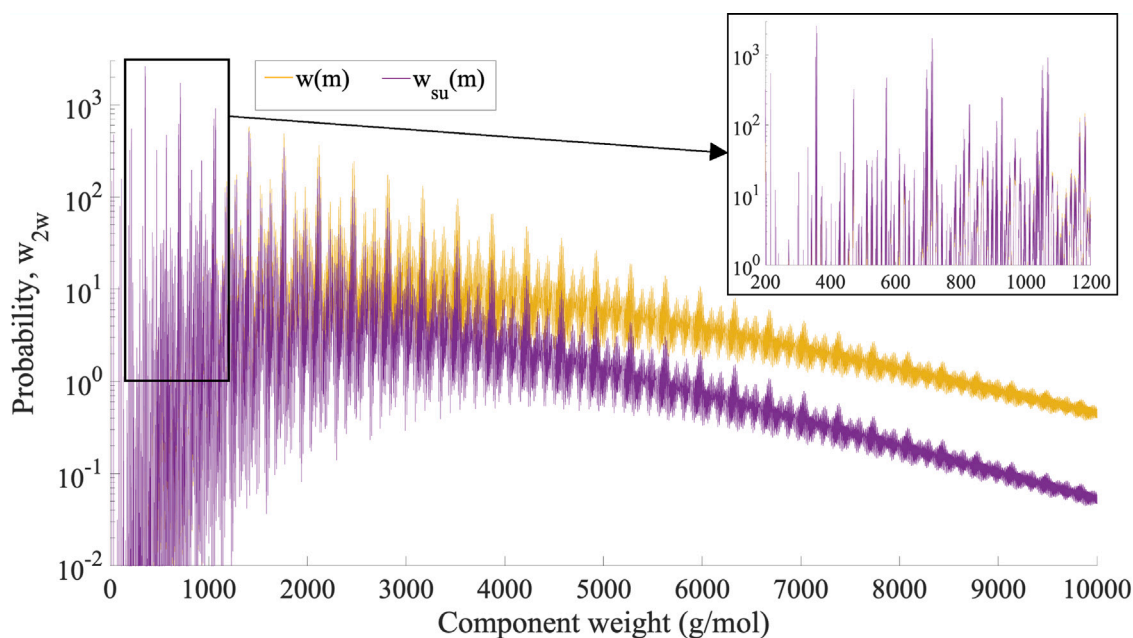


Fig. 6. Comparison of the weight distributions of ethyl linoleate at 70 °C after 144.2 h of polymerization, with the size-unweighted distribution $w_{su}(m)$ (purple, Eq. (8)) and the size-weighted distribution $w(m)$ (yellow). The difference between the two increases with m/s . The mass peaks are at regular distances, allowing to discern the associated sizes. The division of the yellow curves by $m/\mathbb{E}[m]$ coincides with the purple peaks. The highest peaks that appear with repeating mass intervals correspond to EL autoxidation products, predominantly peroxidic species. The zoomed subsection of the spectrum displays repeating patterns of two (groups of) peaks that result of the combination of EL scission products, see also Fig. 7.

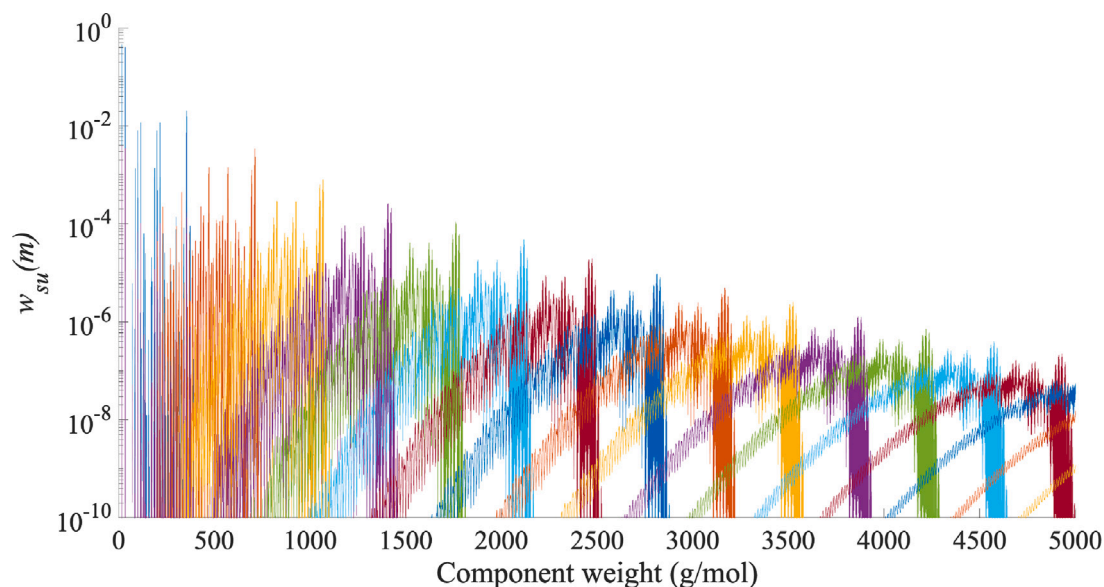


Fig. 7. A disassembled weight distribution for ethyl linoleate polymerization after 25.7 h, where each oligomer size plotted with a different color. Blue curves are monomers, red are dimers, yellow are trimers, etc. The mass peaks and the secondary peaks at lower molar mass (due to scission products) are now clearly visible with different colors (sizes). It is also clear that mass distributions become broader and featuring considerably more overlap between curves of different sizes with increasing mass. Overlapping regions are at very low concentrations (still above noise level) and do not interfere with the regular pattern of the highest peaks.

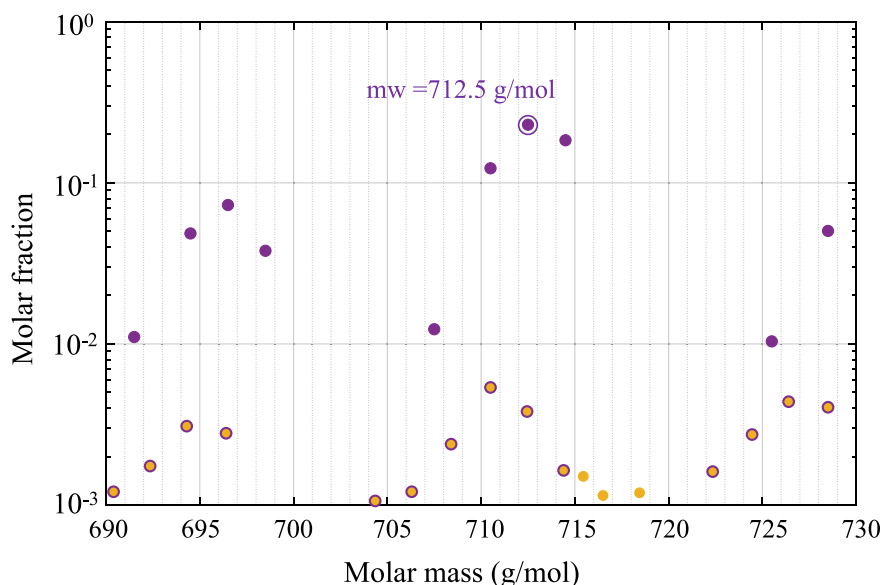


Fig. 8. Detail of electrospray ionization-mass spectrometry spectrum (ESI-MS, Fig. 4 in SI) between 690–730 g/mol with ESI-MS data (yellow) and ARNG/RG-identified points (purple circles) both coarse-grained to 1 g/mol. For comparison, ARNG/RG-identified points at model predicted molar fractions are inserted as purple dots. ESI-MS and model points differ in absolute value, but are the most dominant dimers in both ESI-MS and model. The point at 712.5 g/mol is unraveled in its hundreds of real chemical (semi-)isomers in the SI, Fig. 5 .

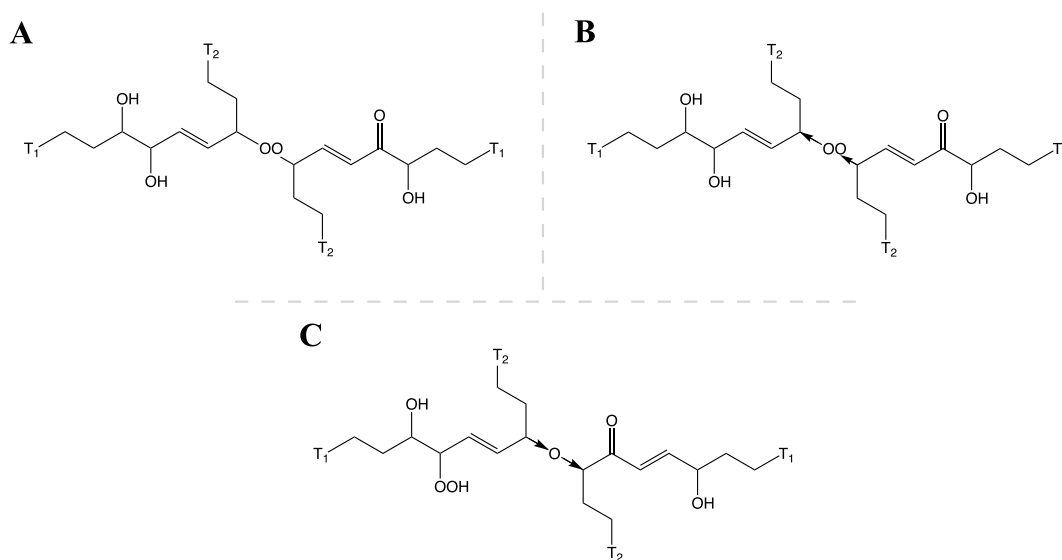


Fig. 9. Structural formulas of three dimer isomers with a molar mass of 712.5 g/mol. Each of the isomers are the most abundant species within their respective categories: dimers connected via peroxy crosslinks as a result of recombination termination (A, $4.64 \cdot 10^{-4}$ mol/L), a peroxy crosslinked dimer after addition reaction (B, $2.14 \cdot 10^{-4}$ mol/L), and dimers connected with an ether crosslink after addition reaction (C, $0.66 \cdot 10^{-4}$ mol/L). They appear in the RG bin of dimers with molar mass 712.5 g/mol together with hundreds of (semi-)isomers in the same mass bin, see Fig. 4 in the SI. The crosslinks shown correspond to different crosslink patterns, see Fig. 1 of SI. The undirected peroxy crosslink in dimer A corresponds to pattern 14, while the directed crosslink in dimer B requires an out-going and in-coming edges: pattern 43 and 45. Likewise, the directed ether crosslink has the complementary pair of in- and outgoing edges, patterns 44 and 46.

binned point with an average mass of 712.5 g/mol. This one point consists of hundreds of dimers, semi- or real isomers, with different types of crosslinks. Fig. 9 shows the structural formulas of the three most abundant dimer isomers with a mass of 712.5 g/mol that are either connected with a peroxy crosslink, created by both recombination termination and addition, or an ether crosslinks created by addition. The consequences of decreasing the mass resolution become relevant in practice when predicted mass spectra are compared with data from mass spectrometry as in Harmon [22]. Further elucidation of the unraveling of (semi-)isomers and the full spectrum are in the SI.

3.4. Model validation: comparison to SEC data

As the main comparison of RG to experimental data, we consider a size exclusion chromatogram (SEC) up to 2000 g/mol of drying methyl linoleate that was recorded by Marquez and Mourão [36,37]. Fig. 10 compares the values of the RG-predicted weight distribution (colored curves) and the experimental SEC of ML oligomerization after 3 h of exposure to air at 80 °C. As activation energies for most reactions are not yet available, it is not possible to run the ARNG model for the higher temperature, so we had to employ the model results

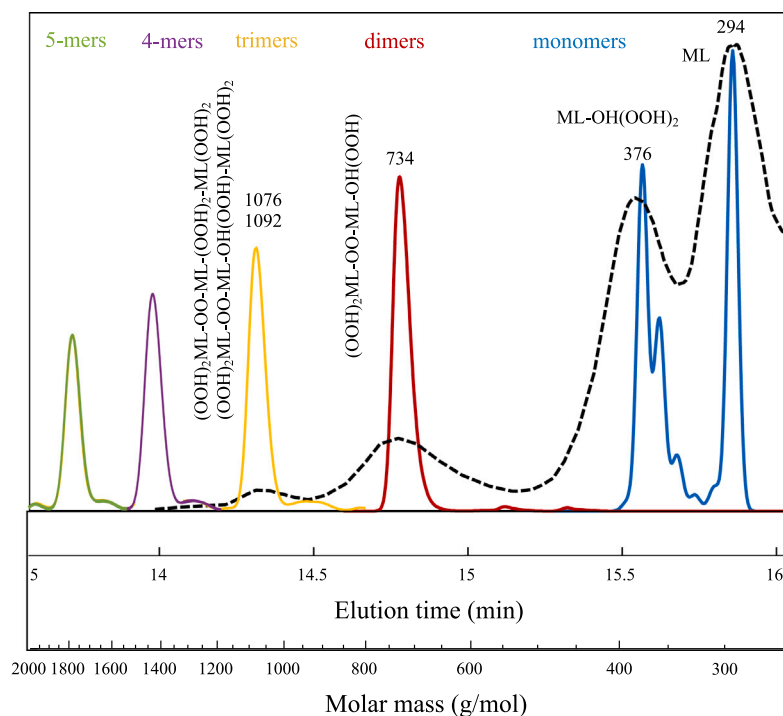


Fig. 10. Molar mass distribution of methyl linoleate from RG (colored curves for 1- to 5-mers after 16.5 h of simulated polymerization at 70 °C) and size exclusion chromatography (SEC) elution trace of ML polymerized at 80 °C for 3 h, dashed black curve [36]. Calibration of the experimental SEC-curve (elution time in minutes) to molar mass values executed using ARNG/RG-model. ML-peak corresponds to 98.5% conversion. Other peaks are 1- to 5-mers with high peroxide substitution. The shoulder at lower molar mass than unreacted ML (right) correspond to oligomers with ML-scission products, see also Figs. 6 and 7.

as is, validated with experiments at 70 °C. Since the experimental SEC trace shows significant amount of unreacted ML, we choose the model curve that matches the ML conversion of 98.5% after 16.5 h. The acceleration of the experimental ML oligomerization compared to the simulated polymerization is mainly due to a higher temperature than we use in the simulation, but also other experimental conditions such as ML film thickness. While the RG-predicted weight distribution produces a singular value for any mass weight, SEC produces a curve as signal. Therefore, for the sake of comparison, the computed peaks are smoothed using Gaussian distributions whose variance increase linearly with molar mass. However, this far from accurate approach of SEC peak broadening still generates narrower peaks for the higher oligomers than is seen in the real SEC trace. Also, the predicted oligomer peak height is consequently higher than the SEC peak heights. Apart from different peak broadening, this may be because the higher oligomers are generated at later stages during the polymerization. Then, oligomer formation may be slower due to diffusion limitations that are not accounted for in the model and thus result in lower peaks.

Marquez and Mourão [36,37] attribute the observed SEC peaks to monomeric and oligomeric peroxides, or monomer units that connect via peroxide crosslinks: unreacted ML (elution time 15.85 min), peroxidic monomer (15.54 min), peroxidic dimer (14.75 min), and peroxidic trimer (14.38 min). Now, the question arises as to exactly which peroxidic species, or exact weights, should these peaks be attributed to? In other words, how should the SEC be calibrated against molecular weights of the peroxidic species? To perform this calibration, we use the relative amounts of these species predicted for monomers and dimers by the ARNG-model. Although ARNG predicts a range of monomeric peroxides, it predominantly predicts ML-OH(OOH)₂ and ML-(OOH)₃ with the respective molar masses and concentrations: 376 g/mol and 0.033 mol/L, and 392 g/mol and $8.3 \cdot 10^{-4}$ mol/L. The notation represents a monomer molecule, ML, with additional side groups. The chemical pathway from ML to the monomer with 376 g/mol is displayed in Fig. 11. Note that this pathway differs considerably from that to the monomer of 392 g/mol, which is shown in the SI in Fig. 6.

Similarly, the ARNG-model predicts various peroxidic dimers, but mainly (HOO)₂ML-OO-ML(OH)OOH, 734 g/mol at 0.013 mol/L and (HOO)₂ML-OO-ML(OOH)₂, 750 g/mol at $2.6 \cdot 10^{-4}$ mol/L—the structural formulas are in the SI (Fig. 7). Interestingly, the most abundantly present monomers and dimers do not possess the maximum number of peroxide groups; one hydroxide group remains. The trimer peak used for calibration was obtained from the RG-model, as trimers and higher order oligomers are not attainable from ARNG. The RG-model predicts peaks of approximately equal concentration of 0.002 mol/L at respective molar masses of 1076 and 1092 g/mol, which correspond to two trimers: (HOO)₂ML-OO-ML(OH)OOH-OO-ML(OH)OOH and (HOO)₂ML-OO-ML(OOH)₂-OO-ML(OOH)₂. In contrast to the monomers and dimers, trimers do have in part the maximum substitution with peroxide groups. Thus considering ARNG/RG predicted monomers, dimers and trimers for the given chemistry and kinetics, we arrive at the following calibration points: unreacted ML 15.85 min/294 g/mol; peroxidic monomer 15.54 min/376 g/mol; peroxidic dimer 14.75 min/734 g/mol; and peroxidic trimer 14.38 min/1084 g/mol. These calibration points appear to be the most plausible and consistent in view of the mutual distance between peaks in the SEC-trace.

When using the set of calibration points, the model prediction and experimental SEC curve agree well. This illustrates the usefulness of the ARNG/RG-model for the validation of chemical mechanisms and kinetic parameters by SEC-data. Additionally, it shows the indispensability of the RG-model for the prediction of higher order oligomers and polymers.

4. Conclusions

The prediction of important characteristics of biopolymers such as the molar mass distribution using fundamental models is shown to be a challenging task. Where industrial linear polymers consist of one or few identical monomeric units, biopolymers contain many different monomers. This leads to a combinatorial problem for modeling that has, until recently, not been comprehensively treated in

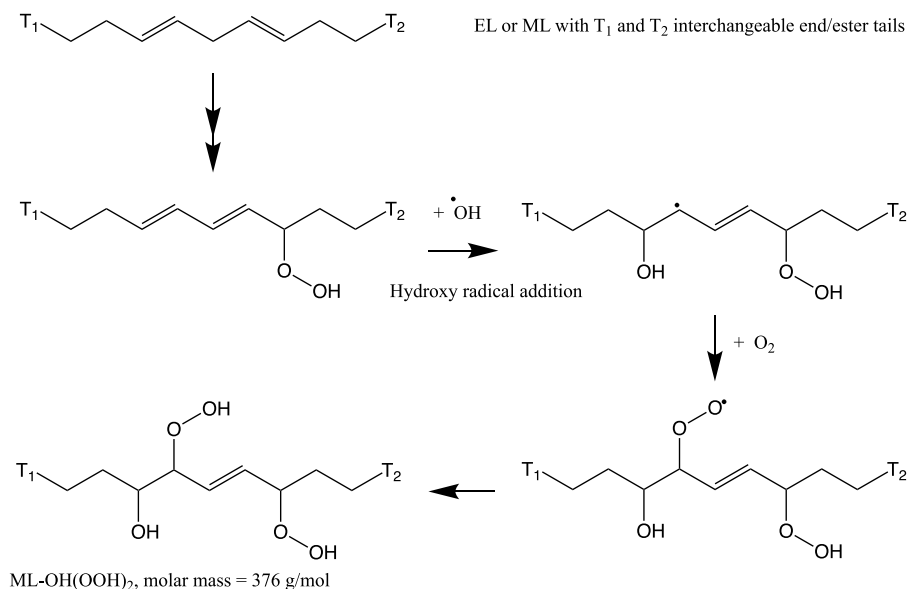


Fig. 11. Chemical pathway found in ARNG from ML to ML-OH(OOH)₂, molar mass = 376 g/mol. Only the reactive part is shown, T-groups referring to Fig. 2 in SI. The structure with conjugated double bonds and one peroxy group is an intermediate formed after a few steps from ML. The conjugated bonds react with hydroxy radical forming an intermediate radical with one double bond that reacts with O₂ and subsequently abstracts hydrogen, ultimately yielding ML-OH(OOH)₂.

polymer science. Our approach to the problem is to exploit an automated model generation technique that produces a refined and comprehensive kinetic model covering millions of reactions between thousands of different components coupled with random graph modeling to infer global polymer properties from the local connectivity of the present monomers. The incorporation of monomer functionality and molar mass enables rapid experimental model validation for complex polymerizing systems.

Model validation was successful in two ways. First, we saw exact agreement in monomer and dimer concentrations between ARNG-model and RG-model results. Second and even more important, we entered into the challenge of confronting modeling results with a renowned experimental technique: size exclusion chromatography. This required us to solve the decoupling of the molar mass and number of monomer units of a connected component. We successfully solved a bi-variate number of monomer units/molar mass distribution. The decoupling was less challenging than expected in view of the observed regular patterns in the molar mass distribution of EL and ML that were readily interpretable in terms of numbers of units per polymer. In future, when dealing with cases with even more widely different molar masses of the monomer units, we do expect more overlap, which makes the decoupling using a two-dimensional number of monomer units/molar mass distribution indispensable.

The experimental validation of our model using SEC on a double mass weighted concentration base resulted in plausible outcome. The model was required to calibrate the SEC trace against the molar mass on a logarithmic scale. The most consistent calibration indeed produced good agreement between model and experiments.

We conclude that the combination of ARNG with RG prove to be a valuable instrument to predict important properties of biopolymers. Comprehensive, integrated models sometimes disadvantageously require many input parameters with a diffuse impact on model outcomes. In contrast, our model proved to be transparent in this respect, despite its complexity. For instance, details of the chemical assumptions in the ARNG-part of the model, such as the degree of substitution by peroxide- and hydroxy-groups affects the predicted SEC-curve. The available SEC-data of different biopolymer systems—autoxidation of model esters in presence of driers, polymerization of linseed oil, etc.—could be

employed in combination with our model to investigate chemical mechanisms and kinetic rates. Subtle changes in the assumptions of the RG-part of the model, types of crosslinks present and crosslink directionality, also influence the results concerning experimentally accessible quantities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data and code can be made available on request.

Acknowledgments

YO: Financial support for PREDAGIO project from The Netherlands Organisation for Scientific Research (NWO) is gratefully acknowledged. RH: Financial support from the National Science Foundation (NSF) grant 1743748 is gratefully acknowledged. Dr. Ivan Kryven (Utrecht University) is acknowledged for proofreading the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cej.2023.145264>.

References

- [1] Y. Orlova, A.A. Gambardella, R.E. Harmon, I. Kryven, P.D. Iedema, Finite representation of reaction kinetics in unbounded biopolymer structures, *Chem. Eng. J.* 405 (2021) 126485, <http://dx.doi.org/10.1016/j.cej.2020.126485>.
- [2] M. Gordon, T.G. Parker, Xiii—the graph-like state of matter. i. statistical effects of correlations due to substitution effects, including steric hindrance, on polymer distributions, *Proc. R. Soc. Edinb. A* 69 (1971) 181–198.
- [3] K. Dušek, Build-up of polymer networks by initiated polyreactions, *Polymers* 13 (4) (1985) 321–328.
- [4] J. Mikes, K. Dusek, Simulation of polymer network formation by the monte carlo method, *Macromolecules* 15 (1) (1982) 93–99.

- [5] P.J. Flory, Molecular size distribution in three dimensional polymers, i. gelation I, *J. Am. Chem. Soc.* 63 (11) (1941) 3083–3090.
- [6] A.B. Scranton, N.A. Peppas, A statistical model of free-radical copolymerization/crosslinking reactions, *J. Polym. Sci. A* 28 (1) (1990) 39–57.
- [7] H. Tobita, Universality in branching frequencies and molecular dimensions during hyperbranched polymer formation: Step polymerization of ab₂ type monomer with equal reactivity, *Macromol. Theory Simul.* 25 (2) (2016) 116–122.
- [8] I. Kryven, P.D. Iedema, Topology evolution in polymer modification, *Macromol. Theory Simul.* 23 (1) (2014) 7–14.
- [9] I. Zapata-González, R.A. Hutchinson, K.A. Payne, E. Saldívar-Guerra, Mathematical modeling of the full molecular weight distribution in atp techniques, *AIChE J.* 62 (8) (2016) 2762–2777.
- [10] A. Torres-Knoop, V. Schamboeck, N. Govindarajan, P.D. Iedema, I. Kryven, Effect of different monomer precursors with identical functionality on the properties of the polymer network, *Commun. Mater.* 2 (1) (2021) 1–9.
- [11] V. Schamboeck, I. Kryven, P.D. Iedema, Acrylate network formation by free-radical polymerization modeled using random graphs, *Macromol. Theory Simul.* 26 (6) (2017) 1700047.
- [12] Y. Orlova, R.E. Harmon, L.J. Broadbelt, P.D. Iedema, Review of the kinetics and simulations of linseed oil autoxidation, *Prog. Org. Coat.* 151 (2021) 106041.
- [13] C.P. Broedersz, F.C. MacKintosh, Modeling semiflexible polymer networks, *Rev. Modern Phys.* 86 (3) (2014) 995.
- [14] B.A. Wolf, Unified thermodynamic modeling of polymer solutions: Polyelectrolytes, proteins, and chain molecules, *Ind. Eng. Chem. Res.* 52 (9) (2013) 3530–3536.
- [15] I. Mijangos, F. Navarro-Villoslada, A. Guerreiro, E. Piletska, I. Chianella, K. Karim, A. Turner, S. Piletsky, Influence of initiator and different polymerisation conditions on performance of molecularly imprinted polymers, *Biosens. Bioelectron.* 22 (3) (2006) 381–387.
- [16] J. Hermans, L. Zuidgeest, P. Iedema, S. Woutersen, K. Keune, The kinetics of metal soap crystallization in oil polymers, *Phys. Chem. Chem. Phys.* 23 (39) (2021) 22589–22600.
- [17] I. Kryven, Emergence of the giant weak component in directed random graphs with arbitrary degree distributions, *Phys. Rev. E* 94 (1) (2016) 012315.
- [18] I. Kryven, Analytic results on the polymerisation random graph model, *J. Math. Chem.* 56 (1) (2018) 140–157.
- [19] V. Schamboeck, P.D. Iedema, I. Kryven, Coloured random graphs explain the structure and dynamics of cross-linked polymer networks, *Sci. Rep.* 10 (1) (2020) 1–18.
- [20] Y. Orlova, I. Kryven, P.D. Iedema, Automated reaction generation for polymer networks, *Comput. Chem. Eng.* 112 (2018) 37–47.
- [21] Y. Orlova, A.A. Gambardella, I. Kryven, K. Keune, P.D. Iedema, Generative algorithm for molecular graphs uncovers products of oil oxidation, *J. Chem. Inf. Model.* 61 (3) (2021) 1457–1469.
- [22] R.E. Harmon, Computational Modeling of Polymerization and Degradation of Complex Polymer Systems (Ph.D. thesis), Universiteit van Amsterdam, 2022, ISBN 978–94–6421–941–8.
- [23] L.J. Broadbelt, S.M. Stark, M.T. Klein, Computer generated pyrolysis modeling: on-the-fly generation of species, reactions, and rates, *Ind. Eng. Chem. Res.* 33 (4) (1994) 790–799.
- [24] R.G. Susnow, A.M. Dean, W.H. Green, P. Peczak, L.J. Broadbelt, Rate-based construction of kinetic models for complex systems, *J. Phys. Chem. A* 101 (20) (1997) 3731–3740.
- [25] H. Wexler, Polymerization of drying oils, *Chem. Rev.* 64 (6) (1964) 591–611.
- [26] L.H. Oakley, F. Casadio, K.R. Shull, L.J. Broadbelt, Microkinetic modeling of the autoxidative curing of an alkyd and oil-based paint model system, *Appl. Phys. A: Mater. Sci. Process.* 121 (3) (2015) 869–878.
- [27] L.H. Oakley, F. Casadio, P.K.R. Shull, P.L.J. Broadbelt, Modeling the evolution of crosslinked and extractable material in an oil-based paint model system, *Angew. Chem. Int. Edn* 57 (25) (2018) 7413–7417, <http://dx.doi.org/10.1002/anie.201801332>.
- [28] L.H. Oakley, F. Casadio, K.R. Shull, L.J. Broadbelt, Theoretical study of epoxidation reactions relevant to hydrocarbon oxidation, *Ind. Eng. Chem. Res.* 56 (26) (2017) 7454–7461.
- [29] N.A. Porter, S.E. Caldwell, K.A. Mills, Mechanisms of free radical oxidation of unsaturated lipids, *Lipids* 30 (4) (1995) 277–290.
- [30] C.-W. Zhou, J.M. Simmie, K.P. Somers, C.F. Goldsmith, H.J. Curran, Chemical kinetics of hydrogen atom abstraction from allylic sites by 3o₂; implications for combustion modeling and simulation, *J. Phys. Chem. A* 121 (9) (2017) 1890–1899.
- [31] U. Jahn, J.-M. Galano, T. Durand, Beyond prostaglandins—chemistry and biology of cyclic oxygenated metabolites formed by free-radical pathways from polyunsaturated fatty acids, *Angew. Chem. Int. Edn* 47 (32) (2008) 5894–5955.
- [32] J. Pfaendtner, L.J. Broadbelt, Mechanistic modeling of lubricant degradation, 1. structure- reactivity relationships for free-radical oxidation, *Ind. Eng. Chem. Res.* 47 (9) (2008) 2886–2896.
- [33] C.A. Jackson, W.J. Simonsick Jr., Application of mass spectrometry to the characterization of polymers, *Curr. Opin. Solid State Mater. Sci.* 2 (6) (1997) 661–667.
- [34] I. Kryven, Bond percolation in coloured and multiplex networks, *Nature Commun.* 10 (1) (2019) 1–16.
- [35] S. Crotty, S. Gerişlioğlu, K.J. Endres, C. Wesdemiotis, U.S. Schubert, Polymer architectures via mass spectrometry and hyphenated techniques: A review, *Anal. Chim. Acta* 932 (2016) 1–21.
- [36] M.P.B. Mourão, Degradative Analytical Tools for Large Molecules Diagnosis of Tuberculosis and Aging of Oil Paintings (Ph.D. thesis), Universiteit van Amsterdam, 2018, ISBN 978–94–93019–17–1.
- [37] G. Márquez-Ruiz, F. Holgado, M. García-Martínez, M.C. Dobarganes, A direct and fast method to monitor lipid oxidation progress in model fatty acid methyl esters by high-performance size-exclusion chromatography, *J. Chromatogr. A* 1165 (1–2) (2007) 122–127.