

## UvA-DARE (Digital Academic Repository)

### Graph-theoretical approach to algorithmic construction of complex reaction networks

Orlova, Y.

**Publication date**

2020

**Document Version**

Other version

**License**

Other

[Link to publication](#)

**Citation for published version (APA):**

Orlova, Y. (2020). *Graph-theoretical approach to algorithmic construction of complex reaction networks*.

**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.

# 6

## Closing remarks and Outlook

The thesis presents a computational methodology for modeling complex polymerization processes, which are characterized by the diversity of their monomer constituents. The proposed method overcomes limitations, which arise due to the infinite growth of polymers and wide variety of monomers. The method builds up on the idea that the chemical notation can be conceptualized with a formal grammar operating on graphs, hence molecular graphs play the role of "words" and chemical reactions of "grammatical rules". Such a grammar-based model is then used to reconstruct the dynamical system for the reaction kinetics on a generic basis. This thesis highlights the concept of grammar in chemical reaction engineering and strengthens the connection between two seemingly unrelated fields. This section offers the reflections on grammar-based modeling and its potential to tackle open problems in the area of bio-based polymers.

For small chemical systems with known reaction steps there is no need to automatize the construction of the reaction network. The problem arises when a molecule has more than one reactive site, and each reactive site undergoes several subsequent transformations. The reconstruction of such a reaction network becomes a difficult problem, which can be treated in an algorithmic manner. The desired algorithm requires input of basic knowledge of chemical reactions taking place on the level of functional groups, the reaction rules. The latter dependency introduces a human factor in the modelling process. Indeed, the automated extraction of reactions from chemical databases<sup>1</sup> could not be addressed in this thesis and is a challenge for future studies. Alternatively, as a promis-

ing future trend in computational chemistry, the reaction rules may be algorithmically discovered using quantum chemical calculations<sup>2</sup>.

One may also bring the creativity to the representation of the molecules. With this objective, we suggested the 'monomer approach' to represent a polymer as a finite set of its unique constituents. Moreover, we anticipate transfer of this concept to other modeling algorithms. Software packages that are dealing with large molecular graphs might benefit from monomer approach and ensure efficient storage of the information about molecular species. The concept is not restricted to oil chemistry alone and can be extended to any atomistic or coarse-grained representations of molecules.

After mapping large or even intractable molecular species on a smaller and finite set of their unique constituents, one already makes a step towards reducing the size of the underlying kinetic model, which in its essence is a large dynamical system. However, integrating such a model might still suffer from stiffness due to large differences in the magnitude of the rate parameters. Kinetic models describing the concentrations of reactive functional groups, instead of molecular species formed by the numerous combinations of functional groups, are smaller and easier to solve and analyze and they allow to be used in state-of-the-art parameter estimation procedures. When functional groups do not interfere with each other's reactivity, they undergo independent reaction pathways, and the system is fully characterized by relative changes in concentrations of these functional groups. An example of such a system in oil chemistry is triolein, which consists of three oleic fatty esters. Each oleic fatty ester contains at most one reactive site, which undergoes an autoxidation reaction pathway. However, the same does not hold for linoleic and linolenic fatty acids. Both fatty acids may simultaneously contain several functional groups, which may influence each other's reactivity and form new functional group with additional reactivity that undergoes its own reaction - in the thesis we have called this a 'complex pattern'. Consequently, our modeling methodology did have to reconstruct all explicit molecular structures, resulting in a kinetic model of the order of thousands of species. As we have not found any probabilistic way of predicting the appearance of complex patterns in a molecule, model reduction of systems with complex patterns still offers an intriguing research challenge. However, it is worthwhile to find a way to express the complex chemistry like that involving linoleic and linolenic acids in terms of functional groups as it forms a great opportunity for effective model reduction. This model reduction will open possibilities to perform effective sensitivity analysis and parameter optimization with experimental data.

Throughout this thesis we have fruitfully employed the direct connection between the reaction networks generated by the ARNG procedure and the kinetic models inferred

from it. However, as an alternative to generating and solving kinetic models consisting of large sets of ordinary differential equations (ODEs), it is possible to employ a stochastic technique: kinetic Monte Carlo (kMC) simulations (<sup>3</sup>). Classical kMC deals with predefined molecular species, typically a very large number:  $10^9$  to  $10^{11}$ . However, the technique allows to be extended to account for the automated creation of new species using predefined functional groups or 'patterns' in much the same way as in our ARNG-methodology, using isomorphism checks, etc. Reaction events between monomeric units can only happen if they are present in sufficient amounts, so improbable reactions are automatically discarded, which considerably reduces the number of species modeled. Combining kMC with automated generation of reaction schemes thus is expected to provide a reduced set of species and reactions as a basis for further modeling, analysis and parameter estimation. At present, in an ongoing collaboration between UvA and Northwestern University the described idea has been started as a PhD project.

Being a much used binding medium for oil paint, linseed oil is valued for its drying properties: slow transition from liquid state to a solid state allowing the artist to introduce changes to a painting. However, the change from a relatively low-viscous oil to a solid paint will significantly slow down the diffusion of reacting species. In other words, for some of the reactions diffusion will become rate-determining. In this thesis, we did not address this issue, but it forms an important factor to take into account while constructing the kinetic model. A crucial issue for diffusion is the accumulation of the insoluble fraction, or gel. It would be interesting to follow the development of the largest connected component in the system, its size and structure, and relate that to the restricted mobility of the diffusing molecules. A fraction of reactive functional groups, like radicals, may become trapped inside the polymer network. Therefore, as another future challenge we consider the development of models predicting visco-elastic and diffusion properties in relation to network structure.

The grammar-based modeling approach is generic and applicable to model other oils, their mixtures and further bio based materials. The work in this thesis has been performed with the ambitious goal: to predict network topology of linseed oil polymer. As this goal is computationally challenging, the methodology has been demonstrated on one fatty acid ester, EL, and has been extended to model triglycerides of triolein by introducing an extra hierarchical level in the molecules' structure.

Consequently, the next step is to model all three common fatty acids, oleic, linoleic and linolenic acid and assemble them into a triglyceride of linseed oil. Having such a model will open the possibilities to investigate the behaviour of various triglycerides, such as linseed oil, soybean oil, poppy seed oil and other oils relevant to food, cosmetics

and paint industries. In the field of cultural heritage science, knowing the molecular structures of low molecular weight degradation products of oils may help to answer questions about the mixtures of oils used by a certain artist and about the age of the painting. Such information is important for authenticity investigations.

As the binding medium of oil paint, linseed oil forms a dense polymer network enveloping pigment particles and fixing them on the surface of the canvas. It is not surprising that pigment particles are prone to chemical reactions themselves and interfere with the drying process by speeding up or slowing down autoxidation reactions. With the developed modeling methodology, adding interactions between linseed oil and pigments narrows down to including extra reaction rules acting on functional groups of oil molecules and pigment particles. Although for numerous pigments these reaction rules are not well studied yet, the demand of such information may inspire future research, which can be investigated using the advances of experimental as well as modeling techniques.

## References

- [1] P. P. Plehiers, G. B. Marin, C. V. Stevens, and K. M. Van Geem, *Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics*, *Journal of Cheminformatics* **10** (2018), 10.1186/s13321-018-0269-8.
- [2] T. Bereau and K. Kremer, *Automated parametrization of the coarse-grained martini force field for small organic molecules*, *Journal of chemical theory and computation* **11**, 2783 (2015).
- [3] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, *Journal of Computational Physics* **22**, 403 (1976).