# Rule-Based Gillespie Simulation of Chemical Systems

Jakob L. Andersen[1,2], Christoph Flamm[2],
Daniel Merkle[4], and Peter F. Stadler[2,3]

[1] Research group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Wien A-1090, Austria
[2] Institute for Theoretical Chemistry, University of Vienna, Wien A-1090, Austria
[3] Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig D-04107, Germany
[4] Department of Mathematics and Computer Science, University of Southern Denmark, Odense M DK-5230, Denmark

We implement a modular stochastic simulation engine for a rule-based chemical models where atoms are explicitly represented. Using a strategy framework originally developed for generating reaction networks, it is possible to tune the degree of network-freeness of the simulation.

The clear separation of specification and implementation applied to the design of nano-scale molecular devices has made it possible for Computer Science to make significant contributions to biotechnology — examples include DNA-based "walkers", "cranes", or "gears" [4]. Formal models such as rule-based calculi have been successfully applied to the modelling and analysis of biological and chemical networks [5]. Most of these engineering and verification approaches are employed on a level of abstraction where the most elementary entities are at least macromolecules such as DNA strands or proteins. Any approaches that encode a specification on an atomic level inherently requires the elementary entities as well as the analytic tools to be on an atomic level as well. Examples include the design and analysis of stable isotope experiments and the detailed investigation of the mechanism of how medically important secondary metabolites are constructed in plant cells. In both examples time-dependent distribution of individual atoms must be modelled.

In this contribution we combine in a modular way a rigorous modelling approach for chemistry based on graph rewriting with established stochastic simulation techniques in order to pave the way towards engineering and verification on the level of abstraction where atoms are explicitly represented. In order to balance accuracy and the high computational effort of the intrinsic underlying combinatorics, we borrow approaches from the area of algorithmic engineering.

*Network-Free Gillespie Simulation* The original formulation of the stochastic simulation algorithm by Gillespie [9] assumes that a reaction network is given at the outset. However, the simulation algorithm also be used as an "explorative" tool to find and study novel chemical behaviour in implicitly defined systems. This mode of operation is of particular importance for reactive chemical systems, where either the exact conditions that constrain the state space are unknown or

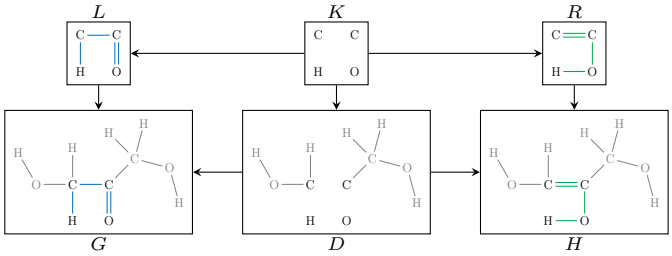


Fig. 1: Application of a rule ($L \leftarrow K \rightarrow R$) for keto-enol tautomerism to a graph $G$ representing dihydroxyacetone, to obtain the product $H$. Note that $L$ matches $G$ in 4 ways: 2 different hydrogens can be matched on two different carbon atoms.

the state space itself is combinatorially large or even unbound. Chemical cracking, signal transduction networks, or combustion chemistry are good examples of this type of reaction systems. This has led to the development of so-called network-free Gillespie simulators, where a rule-based calculus is used to generate reactions as needed for the stochastic simulation. The two most prominent examples are Kappa [5] and BioNetGen [10], which specifically target modules on a macromolecular level and their interactions. A recent review [11] describes the challenges in implementing these kind of simulation engines.

*A Rule-Based Calculus for General Chemistry* Both Kappa [5] and BioNetGen [7] are based on a calculus with so-called site-graphs. While this model is appropriate for abstracted entities, such as macro-molecules, it is less suitable for modelling the underlying chemical entities. Instead we base our simulation on a more classical graph model of chemistry, with molecules encoded as general labelled undirected simple graphs [2]. Vertices model individual atoms and edges model bonds. Labels then encode specific bond types and combinations of chemical elements and charges. An extension to this model allows for representation of local geometric features [3], which for example is needed for a faithful encoding of stereo-information. Several graph transformation formalisms has been developed, each with their own semantics. We use the graph transformation system implemented in the software package MØD [2], which is based on the Double Pushout (DPO) formalism. As opposed to the Single Pushout (SPO) formalism, used in Kappa [6], the DPO formalism has no side-effects of rule application, thus making them invertible. This is more suitable for low-level chemistry where individual atoms are tracked, and where chemical reactions in principle are reversible. A rule application is illustrated in Fig. 1.

*A Modular Simulation Engine* Enumerating all potential rule applications in each simulation step can be computationally expensive and it is thus desirable to be able to vary the degree of "network freeness", depending on the system under investigation and available memory. The strategy framework [1] in MØD was originally developed as a domain-specific programming language for guiding the generation of reaction networks. Using this framework we can formulate the stochastic simulation as a special case of network generation, driven by the

actions of the simulation. Already derived parts of the implicitly defined reaction network can therefore be cached, thus speeding up the simulation. Seperating the concerns of the graph transformation from the actual stochastic simulation addtionally makes it easier to customize the simulation for particular scenarios, e.g., flow reaction systems.

Network-free simulations require computational prediction of reaction rates. This can be as simple as giving the same rate to all reactions instantiated from the same rule, as for example in Kappa [5]. An example of a more complex scheme is given in [8], where the Wiener index of molecules is being used, along with experimentally corroborated constants, to assign reaction rates. Using the graph transformation system in MØD it is possible to get access to the complete graph structure of each molecule, and it is thereby trivial to let users define custom reaction rate algorithms. Symmetries of both rules and molecules may additionally influence the reaction rate [11]. For example, in Fig. 1 there are 4 matches of the rule onto the educt molecule, but they are all equivalent. With extensive access to the graph matching process in MØD it will be possible to implement fine-grained descisions of how predicted rates should be affected by the number of matches.

## References

1. Andersen, J.L., Flamm, C., Merkle, D., Stadler, P.F.: Generic strategies for chemical space exploration. Int. J. Comp. Bio. and Drug Design (2014).
2. Andersen, J.L., Flamm, C., Merkle, D., Stadler, P.F.: A software package for chemically inspired graph transformation. In: Graph Transformation: 9th International Conference, ICGT (2016).
3. Andersen, J.L., Flamm, C., Merkle, D., Stadler, P.F.: Chemical graph transformation with stereo-information. In: Graph Transformation - 10th International Conference, ICGT (2017).
4. Credi, A., Silvi, S., Venturi, M.: Molecular machines and motors: Recent advances and perspectives. Springer (2014)
5. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Rule-based modelling of cellular signalling. In: Proc 8th Int Conf Concurrency Theory (CONCUR) (2007)
6. Danos, V., Feret, J., Fontana, W., Harmer, R., Hayman, J., Krivine, J., Thompson-Walsh, C., Winskel, G.: Graphs, Rewriting and Pathway Reconstruction for Rule-Based Models. In: IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS) (2012).
7. Faeder, J.R., Blinov, M.L., Hlavacek, W.S.: Rule-based modeling of biochemical systems with bionetgen. In: Systems biology. Springer (2009)
8. Faulon, J.L., Sault, A.G.: Stochastic generator of chemical structure. 3. reaction network generation. J. Chem. Inf. Comput. Sci. (2001).
9. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. (1976).
10. Harris, L.A., Hogg, J.S., Tapia, J.J., Sekar, J.A.P., Sanjana, G., Korsunsky, I., Arora, A., Barua, D., Sheehan, R.P., Faeder, J.R.: BioNetGen 2.2: Advances in rule-based modeling. Bioinformatics (2016)
11. Suderman, R., Mitra, E.D., Lin, Y.T., Erickson, K.E., Feng, S., Hlavacek, W.S.: Generalizing gillespie's direct method to enable network-free simulations. Bull. Math. Biol. (2018).