Complex Functional Rates in Rule-Based Languages for Biochemistry

Cristian Versari, Gianluigi Zavattaro

Complex Functional Rates in Rule-based Languages for Biochemistry

Cristian Versari
BioComputing, LIFL, University of Lille 1, France

Gianluigi Zavattaro
Dep. of Computer Science, University of Bologna, Italy

Abstract

Rule-based languages (like, for example, Kappa, BioNetGen, and BioCham) have emerged as successful models for the representation, analysis, and simulation of bio-chemical systems. In particular Kappa, although based on reactions, differs from traditional chemistry as it allows for a graph-like representation of complexes. It follows the *"don't care, don't write"* approach: a rule contains the description of only those parts of the complexes that are actually involved in a reaction. Hence, given any possible combination of complexes that contain the reactants, such complexes can give rise to the reaction. In this paper we address the problem of extending the *"don't care, don't write"* approach to cases in which the actual structure of the complexes involved in the reaction could affect it (for instance, the mass of the complexes could influence the rate). The solutions that we propose is κ_F, an extension of the Kappa-calculus in which rates are defined as functions of the actually involved complexes.

1 Introduction

Rule-based languages like Kappa [13, 9], BioNetGen [17] and BioCham [18] (see the review [21] for a more complete list) have recently emerged as successful models for the representation, analysis, and simulation of bio-chemical systems. In particular, Kappa [13, 9] has been proposed as a formally defined modeling language for biological systems. It allows for the representation of systems composed of molecules with an internal state and an interface used to allow them to bind and unbind. Namely, each molecule has an associated interface composed of sites. Sites represent the possibility for the molecule to bind with another one. Molecule bindings connect two sites of two distinct molecules. The evolution of the system is represented by means of reactions indicating under which conditions the molecules change their internal state, new bindings can be established, or old bindings can vanish. In this way, complexes are represented as groups of molecules connected through bindings.

The distinct feature of Kappa is the *"don't care, don't write"* approach: in a reaction, the reactants are not mandatorily fully described, but they can be identified by a pattern, i.e. an abstract description that can be matched by several different concrete molecules. In this way, a rule contains the description of only those parts of the complexes that are actually involved in a reaction. Hence, a Kappa rule typically gives rise to a combinatorially large number of concrete reactions. This is a great advantage with respect to traditional models based on concrete reactions (or on ordinary differential equations), that are difficult to write, and even more difficult to modify in case the initially written model does not faithfully represent the intended biological system. The rule-based approach revealed particularly appropriate for the modeling of biological signaling networks, as discussed in [21].

The downside of the *"don't care, don't write"* approach as realized in Kappa is that only the "local" properties that characterize the dynamics of the system can be described: indeed, the kinetics of reactions must depend only on the part of reactants matched by the corresponding patterns. In this way, it is not

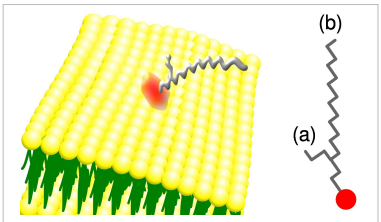

Figure 1: On the left, three-dimensional representation of a non-linear polymer attached to a membrane protein. On the right, schematic representation of the same polymer with labeling of its two free ends (a) and (b).

possible to take into account properties that still influence the kinetics of the system but regard molecular complexes in their entirety. One common example of dynamics affected by such non-local properties is the reactivity of complexes as influenced by their masses: generally, the higher the mass of a complex, the lower the reactivity of its components, because of the slower velocity at which the complex drifts by brownian motion.

To take into account this kind of non-local effects, we propose *Kappa with complex functional rates* (κ_F for short), an extension of the Kappa-calculus which has the possibility to express the rate of a reaction as a function of the complexes in which the reacting molecules actually reside. The *"don't care, don't write"* approach is still used to abstractly specify via patterns the reactants, but the rate can now depend on any property that emerges from the whole complexes and the localization of the actual reacting molecules inside such complexes.

In order to exemplify this calculus, we consider as a case study the effect of mass on the phenomenon of *polymerization*, characterized by the (reversible) binding of simple molecules (monomers) to each other, so to create long chains of variable length: polymers. This kind of system is easily representable in Kappa: there is only one basic molecule, the *monomer* having two sites in its interface, and there are only two reactions, one for the generation of a new binding between two free sites of two distinct monomers, and one for the elimination of a previously established binding. Each of these reactions has its own rate, thus the model considers two constant rates: the *binding* and the *unbinding* rates. As previously discussed, in real systems rates are influenced by the mass of the polymers. In fact, the higher the mass of a polymer, the lower is its motility, which is one of the factors contributing to its reaction propensity. According to this observation, the binding rate for short polymers should be higher than the binding rate for long polymers, so that a faithful model of polymerization should not consider only one constant binding rate.

With κ_F it is possible to take into account this effect, as any other effect at the level of molecular complex, for example by expressing the binding rate as a function of the length of the two polymers to which the two reacting monomers belong (see Section 4.1 for a detailed description).

In order to be able to ease the representation of any kind of non-local effect at the level of molecular complexes, κ_F reactions are enriched with *colors*: the basic idea is to use a color to identify each of the reacting molecules inside their molecular complexes. In the cited case of polymerization, this allows for example the expression of rates which depend on the position of the reacting monomers inside the polymer chain. Therefore, in the system depicted in Figure 1 it would be possible to express different rates for the two free ends (a) and (b) of the polymer — despite them having the same chemical composition — for example as (likely non-linear) functions of the lengths of their respective sub-chains as a consequence of their different mobility in space. According to this intuition, reaction rates are expressed as functions of colored complexes. More precisely, once the reacting molecules are detected in the solution, they are colored and the relevant complexes are obtained by transitive closure following their bindings. The functional rate is then applied to such colored complexes to compute the actual rate of the reaction.

Besides the polymerization case study, we also report the representation of a rather sophisticated nano device, a supramolecular system behaving like a nanoscale elevator [2]. This system, like most of the nano

devices [3], is obtained by integration of several structural and functional molecular subunits. The Kappa modeling approach is particularly appropriate for such systems: each subunit and its features can be modeled in isolation following the "don't care, don't write" approach. Nevertheless, a typical phenomenon observed on these systems is that, upon aggregation, the kinetics of each of the subunits is affected by the other subunits in the same device. We show that κ_F, thanks to the functional rates, allows also for a natural modeling of these phenomena.

It is worth noting that the simulator for the Kappa language `KaSim` [1] already includes the possibility to associate to a reaction a pair of rates, the first one to be used when the reactants are freely floating and the second one for the case in which they are part of the same complex. This mechanism allows for the modeling of interesting systems like, for instance, linear polymers that do not form rings due to their rigid structure. This can be obtained by associating to the binding reaction a pair of rates in which the second one is set to 0 to indicate that the binding reaction is disallowed when the reactants belong to the same complex. This approach is anyway less expressive than the functional rates that we propose in this paper. For instance, it could be the case that the polymers can form rings only when they are sufficiently long. In κ_F it is possible to model such systems by indicating that the reaction rate is equal to 0 only if the complex to which the reactants belong does not include enough monomers. But in κ_F it is possible to model even more complex situations: any physical or chemical effect causing, for example, different reactivities for the two ends of the polymer in Figure 1 could be taken into account for any molecular complex, independently of the number of its components, or the complexity of its structure.

It is also worth to mention the variant of Kappa presented in [10], where the so called *rule refinement* approach is presented. The idea of rule refinement is to replace a rule with a set of rules, each one strengthening the conditions under which the initial rule can be applied. Our approach is different because we do not add rules to a Kappa model, but we simply allow for the definition of the reaction rate as a function of the complexes in which the reactants actually reside. We consider our approach more appropriate for the modeling of systems in which some specific physico-chemical properties of the complexes hosting the reactants have an impact on the system kinetics. On the contrary, the rule refinement approach revealed appropriate to study the distribution of already known rates from a Kappa rule to its refinements. In fact, one of the main contribution of [10] is the definition of a mechanism for inferring the rates of the refined rules in such a way that the kinetics of the initial non refined system is preserved.

From a syntactic point of view, κ_F is a slight modification of Kappa: the constant reaction rate is replaced by a functional rate. From a semantic point of view, on the contrary, the modifications are relevant (see Section 2 for the details). Due to this significant difference, it is not trivial to modify the Kappa simulation and analysis tools to use them on κ_F. For this reason, and in order to obtain results that could experimentally justify the introduction of the new calculus, we have studied a translation from κ_F to standard chemical reaction networks. This can be done by associating to each κ_F complex a chemical species, and then by considering for each κ_F reaction rule all its possible instantiations on those species. For all the cases where the so-called "combinatorial explosion" (arising from the many internal states and the many ways in which molecules could bind to each other) is not prohibitive, the translation to chemical reaction networks is a reasonable and easy to use way to simulate and analyze biochemical systems, as it allows the modeler to exploit all the tools already available for traditional chemistry.

Structure of the paper

In Section 2 we present the syntax and the semantics of κ_F defined in terms of Continuous Time Markov Chains (CTMCs). In Section 3 we show how to translate a κ_F model into a stochastic chemical reaction network (CRN). The correctness of the translation is proved by showing that the CTMCs of the κ_F model and of the corresponding CRN are isomorphic. In Section 4 we exploit the translation to traditional chemistry to perform simulations on κ_F models. We first analyze a case study involving linear polymerization to highlight some of the discrepancies between the modeling approaches of other rule-based languages and κ_F. We then discuss the nanoscale elevator case study: this example is used to present one promising field of application for κ_F, namely the modeling and simulation of supramolecular systems and nano devices. At the end of Section 4 we comment the performances and scalability of the exploited technique to realize the simulations.

Finally, in Section 5 we discuss the related literature and draw some conclusions.

2 The κ_F calculus

As in Kappa, the basic component of the κ_F calculus is the *molecule*. Each molecule belongs to a *species*. Each species is characterized by a *species name*, a predefined number of *species fields*, and a predefined number of *species sites*. The fields are used to model the internal state of a molecule, while the sites are used to represent the bonds between molecules. A molecule of one species is modeled by a *state* specified in terms of the evaluation of its fields, and by an *interface* that specifies, for each of its sites, whether it is unlinked or linked to the site of another molecule by a specific bond. A *solution* is a consistent multiset of molecules: a multiset of molecules is consistent when each bond is connected to exactly two sites of two distinct molecules. A *complex* is a connected solution, i.e. a maximal group of connected molecules.

The dynamics of solutions is specified in terms of *reactions*. A reaction specifies under which conditions a group of distinct molecules can interact, and how their internal state and/or interface are modified as effect of the interaction. The interaction can also remove or generate molecules. Each reaction has an associated *rate* constant that quantifies its speed. Differently from Kappa, in the κ_F calculus it is possible to specify rates as functions of the complexes in which the reacting molecules are hosted. This apparently minimal difference with the Kappa-calculus, requires a significant modification of the formal definition of the semantics.

2.1 Syntax

We consider a countable set of *species* names ranged over by A. Species are sorted according to the number of *fields* and *sites* they possess. Let $\mathfrak{s}_f(\cdot)$ and $\mathfrak{s}_s(\cdot)$ be two functions from A to natural numbers; the integers 1, 2, ..., $\mathfrak{s}_f(\mathrm{A})$ and 1, 2, ..., $\mathfrak{s}_s(\mathrm{A})$ are respectively the fields and the sites of A (in particular, $\mathfrak{s}_f(\mathrm{A}) = 0$ means there is no field, $\mathfrak{s}_s(\mathrm{A}) = 0$ means there is no site).

We consider a countable set of *bond* identifiers ranged over by x, y, ... Sites may be either *bound* to other sites or *unbound*, i.e. not connected to other sites. The state of sites are defined by injective maps, called *interfaces* and ranged over by σ, ρ, $\cdots$. Given a species A, its interfaces are partial functions from $\{1, \cdots, \mathfrak{s}_s(\mathrm{A})\}$ to the set of bond names or a special empty value ε. A site a is bound with bond x in σ if $\sigma(a) = x$; it is unbound if $\sigma(a) = \varepsilon$. For instance, if A is a species with three sites, $(2 \mapsto x, 3 \mapsto \varepsilon)$ is one of its interfaces. In order to ease the reading, we write this map as $2^x + 3$ (the empty value is always omitted). This interface σ does not define the state of the site 1, which may be bound or not. In the following, when we write $\sigma + \sigma'$ we assume that the domains of σ and σ' are disjoint. We require interfaces to be injective in order to ensure that two sites belonging to the same molecule cannot be bound: this reflects the impossibility for single molecules to form self-complexes. In other words, we impose that the endpoints of a bond cannot belong to the same molecule.

Fields represent the internal state of a species. The values of fields are defined by maps, called *evaluations*, and ranged over by u, v, $\cdots$. For instance, if A is a species with three fields, $[1 \mapsto 5, 2 \mapsto 0, 3 \mapsto 4]$ is an evaluation of its fields. As before, we write this map as $1^5 + 2^0 + 3^4$. We assume there are finitely many internal states, that is every field h is mapped into values in $\{0, \cdots, n_h\}$. In the following, we use partial evaluations and, when we write the union of evaluations $u + v$, we implicitly assume that the domains of u and v are disjoint.

We are now ready to define the syntax for κ_F solutions.

Definition 1 (Solutions)**.** *The syntax of κ_F solutions is defined by the following grammar:*

$$\begin{array}{rcl} \mathrm{S} & ::= & \textsc{Mol} \quad | \quad \mathrm{S}, \mathrm{S} \\ \textsc{Mol} & ::= & \mathrm{A}[u](\sigma) \end{array}$$

with "," associative (but not commutative). We write $\textsc{Mol} \in \mathrm{S}$ *if* $\mathrm{S} = \mathrm{S}_1, \textsc{Mol}, \mathrm{S}_2$ *for some (possibly empty) solutions* $\mathrm{S}_1, \mathrm{S}_2$.

Notice that, according to the previously introduced notation, $A[u](\sigma)$ denotes a molecule of species A, with evaluation u and interface σ. Moreover, notice that we do not assume commutativity of "," because the order is relevant when a color is associated to a solution (see Definition 3).

In the remainder of the paper we will use the following notation:

- $S, S', S_1, \ldots$ denote *solutions* (i.e. each field and site of each molecule is specified and each bond identifier appears exactly twice);
- $P, P', P_1, \ldots$ denote *pre-solutions* (i.e. each field and site of each molecule is specified but bond identifiers may appear once or twice);
- $M, M', M_1, \ldots$ denote solution *patterns* (i.e. molecules fields and sites may be omitted and bond identifiers may appear once or twice);
- we use b_S to denote the bond identifiers occurring in a solution S.

Notice that the notion of solution allows us to easily formalize the notion of *complex*: a complex is a solution that does not strictly include another solution.

In order to denote the reacting molecules inside one complex, we introduce colors. Intuitively, colors are vectors of identifiers that will be associated to solutions in order to have an identification mechanism for single molecules inside a solution.

Definition 2 (Color)**.** *Let $\mathcal{C}$ be a denumerable set of color identifiers, with $\epsilon \in \mathcal{C}$ denoting the empty color. A color is a tuple $\tilde{c} = (c_1, \ldots, c_n)$ of color identifiers $c_i \in \mathcal{C}$, such that an identifier different from ϵ can appear only once, namely, if $c_i \neq \epsilon$ then $c_i \neq c_j \; \forall i \neq j$, with $1 \leq i, j \leq n$.*

If $c_i \neq \epsilon \; \forall i = 1, \ldots, n$ then $\tilde{c}$ is said to be saturated*.*

Given two colors $\tilde{c}_1 = (c_1^1, \ldots, c_{n_1}^1)$ and $\tilde{c}_2 = (c_1^2, \ldots, c_{n_2}^2)$ then, for i ranging over $1, \ldots, n_1$ and j over $1, \ldots, n_2$:

- *if for every i s.t. $c_i^1 \neq \epsilon$ we have that $c_i^1 \neq c_j^2$ for every j, then $\tilde{c}_1, \tilde{c}_2$ are said distinct;*
- *if $\tilde{c}_1, \tilde{c}_2$ are distinct, then $\tilde{c}_1 \frown \tilde{c}_2$ denotes the color $\tilde{c} = (c_1^1, \ldots, c_{n_1}^1, c_1^2, \ldots, c_{n_2}^2)$;*
- *$\tilde{c}_1 \subseteq \tilde{c}_2$ if and only if for every i s.t. $c_i^1 \neq \epsilon$ then $c_i^1 = c_j^2$ for some j.*

We are now ready to introduce the notion of solution enriched with a color allowing for the identification of the single molecules inside the solution. The identification of single molecules in real chemical solutions is usually impossible, but we introduce the notion of colored solution as a mathematical object that will allow us to specify the functional rates we are interested in.

Definition 3 (Colored solution)**.**
A colored solution $S^{\tilde{c}}$ *is a pair $(S, \tilde{c})$ where S is a solution $S = \text{MOL}_1, \ldots, \text{MOL}_n$ and $\tilde{c} = (c_1, \ldots, c_n)$ is a color.*

With $S_1^{\tilde{c}_1}, S_2^{\tilde{c}_2}$ we denote the colored solution $S^{\tilde{c}} = (S, \tilde{c})$ where $S = S_1, S_2$ and $\tilde{c} = \tilde{c}_1 \frown \tilde{c}_2$.

We write that $\text{MOL}^c \in S^{\tilde{c}}$ if $S^{\tilde{c}} = S_1^{\tilde{c}_1}, \text{MOL}^c, S_2^{\tilde{c}_2}$ for some (possibly empty) colored solutions $S_1^{\tilde{c}_1}$ and $S_2^{\tilde{c}_2}$.

The above definitions are also naturally extended to pre-solutions and patterns.

We now introduce structural congruence for solutions, which allows for the reordering of molecules inside the solution and for the renaming of bond identifiers. This is used when it is necessary to avoid to distinguish between two syntactically different κ_F systems that represents the same bio-chemical solution.

Definition 4 (Structural congruence)**.**
$\equiv$ is the least congruence over the set of (colored) solutions satisfying the following two rules:

- *renaming of bonds:*
  $S_1 \equiv S_2$ *($S_1^{\tilde{c}} \equiv S_2^{\tilde{c}}$) if there is an injective renaming $\mathcal{I}$ of bonds in S_1 such that $\mathcal{I}(S_1) = S_2$;*

- *permutation of (colored) solutions:*
  $\mathrm{S}_1, \mathrm{S}_2 \equiv \mathrm{S}_2, \mathrm{S}_1$ *(*$\mathrm{S}_1^{\tilde{c}_1}, \mathrm{S}_2^{\tilde{c}_2} \equiv \mathrm{S}_2^{\tilde{c}_2}, \mathrm{S}_1^{\tilde{c}_1}$*).*

With $[\mathrm{S}]_\equiv$ *we denote the congruence class of the solution* S*.*

Structural congruence is extended naturally to pre-solutions and patterns, as well as to their colored variants.

We are finally ready to define the κ_F reactions, used to specify the dynamic evolution of κ_F systems.

Definition 5 (Reactions)**.** *A* κ_F *reaction* ρ *is a triple* $(\mathrm{M}_1^{\tilde{c}_1}, f_\rho, \mathrm{M}_2^{\tilde{c}_2})$ *usually written as* $\rho = \mathrm{M}_1^{\tilde{c}_1} \xrightarrow{f_\rho} \mathrm{M}_2^{\tilde{c}_2}$ *where* $\mathrm{M}_1^{\tilde{c}_1}$ *and* $\mathrm{M}_2^{\tilde{c}_2}$ *are patterns specifying the possible* reactants *and the corresponding* products*, respectively, and* f_ρ *is a function from colored solutions to non-3negative numbers (representing actual rates) that preserves structural congruence, i.e. if* $\mathrm{S}^{\tilde{c}} \equiv \mathrm{S}'^{\tilde{c}'}$ *then* $f_\rho(\mathrm{S}^{\tilde{c}}) = f_\rho(\mathrm{S}'^{\tilde{c}'})$*.*

Notice that the definition of reaction is essentially the same as in Kappa, with the unique difference that a functional rate is considered. Intuitively, the functional rate f_ρ is responsible for checking the complexes in which the reacting molecules reside, and according to their structure, a corresponding rate is computed. For simplicity, we have considered as domain of f_ρ the entire set of colored solutions, but in practice only the solutions simply composed by the complexes in which the reactants are hosted are relevant. This is made clear in Table 1 where f_ρ is applied only to the complexes directly involved in the reaction. Moreover, the colors $\tilde{c}_1$ and $\tilde{c}_2$ are used in the reaction to keep track of the identity of the reactants: the colors of the molecules that are removed occur only in $\tilde{c}_1$, those that are generated are colored only in $\tilde{c}_2$, while the other reactants occur in both with the same color. This will be formalized in Definition 7.

Example 1. *We now formalize in* κ_F *the example of linear polymerization informally described in the Introduction. To ease the notation, we consider linear polymers that do not form rings, characterized by a binding rate* λ *and an unbinding rate* λ'*. The more elaborate case of polymerization with binding rate depending on the length of the reacting polymers is a trivial modification of this example, and will be discussed in details in Section 4.1.*

We consider only one species A *representing the monomers. Monomers have no fields (so we omit the evaluation) and have an interface with two sites. We consider the following binding and unbinding rules:*

$$\mathrm{A}(1), \mathrm{A}(2) \xrightarrow{f} \mathrm{A}(1^x), \mathrm{A}(2^x)$$
$$\mathrm{A}(1^x), \mathrm{A}(2^x) \xrightarrow{f'} \mathrm{A}(1), \mathrm{A}(2)$$

where we have omitted colors as they are not relevant. The functions f *and* f' *are defined as follows:*

$$f(\mathrm{S}^{\tilde{c}}) \;=\; \textit{if } \mathrm{S} \textit{ contains two complexes then } \lambda \textit{ else } 0$$
$$f'(\mathrm{S}^{\tilde{c}}) \;=\; \lambda'$$

Notice that the impossibility to form rings is here obtained simply by defining the binding rate as 0 in case the two reacting monomers belong to the same polymer.

We now consider a more sophisticated modeling of monomer dissociation, in which the unbinding rate depends on the position of the detaching monomers inside the polymer. For instance, in the polymer represented in Figure 1 the monomers close to the extremity could have a greater detaching rate due to their higher mobility in space. To express this phenomenon we can associate to the unbinding reaction a functional rate that requires the exploitation of colors to identify the position of the detaching monomers inside the polymer. Consider, for instance, the following reaction rule:

$$\mathrm{A}(1^x)^{c_1}, \mathrm{A}(2^x)^{c_2} \xrightarrow{f''} \mathrm{A}(1)^{c_1}, \mathrm{A}(2)^{c_2}$$

with

$$f''(\mathrm{S}^{\tilde{c}}) \;=\; \frac{\lambda}{min\{dist(\mathrm{S}^{\tilde{c}}, c_1, end_2), dist(\mathrm{S}^{\tilde{c}}, c_2, end_1)\}}$$

where $dist(\mathrm{S}^{\tilde{c}}, c_i, end_j)$ *is the distance in the colored polymer* $\mathrm{S}^{\tilde{c}}$ *between the monomer colored with* c_i *and the extremity having the site* j *free. Notice that the smaller the distance between the detaching polymers and a polymer extremity, the higher the rate.*

2.2 Semantics

In κ_F, once the reactants are identified, it is necessary to identify the complexes in which such reactants are hosted. Those complexes are represented by a minimal solution that includes the reactants. If P is the pre-solution composed of the reactants only, and P, P' is such a minimal solution, we denote this with $\textsc{minsol}(\mathrm{P}, \mathrm{P}')$.

Definition 6. $\textsc{minsol}(\mathrm{P}, \mathrm{P}')$ *if and only if*

- P, P' *is a solution, i.e. every bond identifier appears exactly twice;*
- *if* $\mathrm{P}' \equiv \mathrm{P}'', \mathrm{P}'''$, *with* P'' *not empty pre-solution, then* $\mathrm{P}, \mathrm{P}'''$ *is not a solution.*

The last notation that we need is used to formalize the matching between a pattern and an actual (pre)solution. A pattern is essentially a partial description of a group of molecules: by adding the remaining information we can achieve the description of an actual instantiation of the pattern. We use the notation $\mathrm{M}_1^{\tilde{c}_1} \triangleleft \mathrm{M}_2^{\tilde{c}_2}$ to denote the extension of the colored pattern $\mathrm{M}_1^{\tilde{c}_1}$ with the colored pattern $\mathrm{M}_2^{\tilde{c}_2}$. Colors are used to relate the single molecules inside the two patterns.

Definition 7. *Let* $\mathrm{M}_1^{\tilde{c}_1}$ *and* $\mathrm{M}_2^{\tilde{c}_2}$ *be two colored patterns, such that*

- $\tilde{c}_1, \tilde{c}_2$ *are saturated;*
- *for every* $\textsc{Mol}_j^c \in \mathrm{M}_2^{\tilde{c}_2}$, *with* $\textsc{Mol}_j = \mathrm{A}[u](\sigma)$, *there exists* $\textsc{Mol}_i^c \in \mathrm{M}_1^{\tilde{c}_1}$ *such that* $\textsc{Mol}_i = \mathrm{A}[u'](\sigma')$ *and* u, u' *are disjoint, as well as* σ *and* σ'.

Then, $\mathrm{M}_1^{\tilde{c}_1} \triangleleft \mathrm{M}_2^{\tilde{c}_2} = \mathrm{M}^{\tilde{c}_1}$, *where for every* $\textsc{Mol}_k^c \in \mathrm{M}^{\tilde{c}_1}$:

- *if* $c \notin \tilde{c}_2$, *then* $\textsc{Mol}_k^c = \textsc{Mol}_i^c$, *with* $\textsc{Mol}_i^c \in \mathrm{M}_1^{\tilde{c}_1}$ *for some* i;
- *if* $c \in \tilde{c}_2$, *then* $\textsc{Mol}_k^c = \mathrm{A}[u_1 + u_2](\sigma_1 + \sigma_2)$, *with* $\textsc{Mol}_i^c \in \mathrm{M}_1^{\tilde{c}_1}$ *for some* i, $\textsc{Mol}_j^c \in \mathrm{M}_2^{\tilde{c}_2}$ *for some* j, *and* $\textsc{Mol}_i = \mathrm{A}[u_1](\sigma_1)$, $\textsc{Mol}_j = \mathrm{A}[u_2](\sigma_2)$.

We are finally ready to define the operational semantics of a κ_F system.

Definition 8 (κ_F semantics)**.** *Given a set of reactions* R *and an initial solution* S_0, *we denote with* $\mathrm{LTS}(\mathrm{S}_0, R)$ *its operational semantics, defined as the minimal labeled transition system whose states are congruence classes of solutions and the labels are non-negative numbers (denoting rates) that contains* $[\mathrm{S}_0]_\equiv$ *and the transitions* $[\mathrm{S}]_\equiv \xrightarrow{\lambda} [\mathrm{S}']_\equiv$ *that can be inferred by using the rules in Table 1.*

We first observe that the operational semantics is well defined as the choice of S and S′ taken as representatives of the congruence classes in the last rules is not important: given a solution S, each of its structurally congruent solutions has the same outgoing transitions thanks to the premise $\mathrm{S}^{\tilde{c}} \equiv \mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}, \mathrm{S}_2^{\epsilon}$ of the second rule in Table 1.

We now comment the rules in Table 1. The first rule is used to instantiate the patterns in a reaction, in order to fully specify the reactants and the products. Formally, the two patterns $\mathrm{M}_1^{\tilde{c}_1}$ and $\mathrm{M}_2^{\tilde{c}_2}$ are both extended with $\mathrm{M}^{\tilde{c}}$ in order to obtain the reactants $\mathrm{P}_1^{\tilde{c}_1}$ and the products $\mathrm{P}_2^{\tilde{c}_2}$ (notice that the colors $\tilde{c}_1$ and $\tilde{c}_2$ are those used in the definition of the considered reaction rule ρ). A similar rule is sufficient to specify the traditional Kappa semantics, according to which a rate is a constant associated to a rule. In κ_F, on the contrary, the rate is a function of the complexes in which the reactants are hosted. So it is necessary to consider another rule, the second one, that lifts the transitions inferred by the first rule to an entire solution $\mathrm{S}^{\tilde{c}}$ that contains the reactants $\mathrm{P}_1^{\tilde{c}_1}$, the other molecules P_3^{ϵ} hosted in the complexes of the reactants, and additional molecules S_2^{ϵ} not involved in the reaction. The solution $\mathrm{S}^{\tilde{c}}$ is colored in order to identify the actual reactants inside the solution. The rate of the reaction can be computed applying the functional rate to the (sub)solution $\mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}$ composed of the complexes in which the reactants are hosted. In this rule we also add two conditions on the bond identifiers: the first one ensures that the new bonds generated by the reaction are denoted by fresh identifiers; the second one guarantees that if a bond is removed by the reaction, both of

Table 1: Operational semantics of κ_F

$$\frac{\mathrm{M}_1^{\tilde{c}_1} \xrightarrow{f_\rho} \mathrm{M}_2^{\tilde{c}_2} \in R \qquad \mathrm{M}_1^{\tilde{c}_1} \triangleleft \mathrm{M}^{\tilde{c}} = \mathrm{P}_1^{\tilde{c}_1} \qquad \mathrm{M}_2^{\tilde{c}_2} \triangleleft \mathrm{M}^{\tilde{c}} = \mathrm{P}_2^{\tilde{c}_2}}{\mathrm{P}_1^{\tilde{c}_1} \xrightarrow{f_\rho}_\rho \mathrm{P}_2^{\tilde{c}_2}}$$

$$\frac{\begin{array}{c}\mathrm{S}^{\tilde{c}} \equiv \mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}, \mathrm{S}_2^{\epsilon} \qquad \mathrm{P}_1^{\tilde{c}_1} \xrightarrow{f}_\rho \mathrm{P}_2^{\tilde{c}_2} \qquad \textsc{minsol}(\mathrm{P}_1, \mathrm{P}_3) \\ f(\mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}) = \lambda \qquad (b_{\mathrm{P}_2} \setminus b_{\mathrm{P}_1}) \cap b_{\mathrm{P}_3, \mathrm{S}_2} = \emptyset \qquad (b_{\mathrm{P}_1} \setminus b_{\mathrm{P}_2}) \cap b_{\mathrm{P}_3} = \emptyset\end{array}}{\mathrm{S} \xrightarrow{\lambda}_{\tilde{c},\rho} \mathrm{P}_2, \mathrm{P}_3, \mathrm{S}_2}$$

$$\frac{\lambda = \sum_{\{\tilde{c},\rho,\lambda' \,:\, \mathrm{S} \xrightarrow{\lambda'}_{\tilde{c},\rho} \mathrm{S}'' \text{ (with } \mathrm{S}'' \equiv \mathrm{S}')\}} \lambda'}{[\mathrm{S}]_\equiv \xrightarrow{\lambda} [\mathrm{S}']_\equiv}$$

its ends are specified as reactants. We add this last condition because we want to disallow the specification of reactions that leave dangling bond identifiers.

The last rule is used to count how many different transitions have the same effect of transforming a solution structurally congruent to S, to a solution structurally congruent to S′: all the rates of the distinct transitions are summed and one unique transition is considered from $[\mathrm{S}]_\equiv$ to $[\mathrm{S}']_\equiv$ labeled with the obtained sum.

Example 2. *We continue Example 1 by considering a solution* S *composed of one polymer of length* l*, namely, a solution composed of* l *molecules of species* A *composing only one complex, with one molecule with the site* 1 *free and one molecule with the site* 2 *free. We consider the last detaching reaction rule in the Example 1*

$$\mathrm{A}(1^x)^{c_1}, \mathrm{A}(2^x)^{c_2} \xrightarrow{f''} \mathrm{A}(1)^{c_1}, \mathrm{A}(2)^{c_2}$$

and we identify it with ρ*. The first rule in Table 1 guarantees that*

$$\mathrm{A}(1^x, 2^y)^{c_1}, \mathrm{A}(1^z, 2^x)^{c_2} \xrightarrow{f''}_\rho \mathrm{A}(1, 2^y)^{c_1}, \mathrm{A}(1^z, 2)^{c_2}$$

for every y, z *bond names or* ε*. The second rule ensures that*

$$\mathrm{S} \xrightarrow{f''(\mathrm{S}^{\tilde{c}})}_{\tilde{c},\rho} \mathrm{S}'$$

where $\tilde{c}$ *is a coloring for* S *that associates colors (namely* c_1 *and* c_2*) only to the two detaching monomers. The solution* S′ *is a solution composed of two polymers of length* i *and* j *such that* $i + j = l$*. We now focus on one of these pairs* i, j*, assuming* $i \neq j$*. It is interesting to notice that there exists two distinct colorings for* S *that can generate such two polymers: a coloring such that the molecule colored* c_1 *will be part of the polymer of length* i *after the reaction, and the coloring in which such molecule will be part of the polymer of length* j*. If we name* $\tilde{c}'$ *and* $\tilde{c}''$ *these two colors, we have the two following transitions*

$$\mathrm{S} \xrightarrow{f''(\mathrm{S}^{\tilde{c}'})}_{\tilde{c}',\rho} \mathrm{S}' \qquad \mathrm{S} \xrightarrow{f''(\mathrm{S}^{\tilde{c}''})}_{\tilde{c}'',\rho} \mathrm{S}''$$

Notice that $\mathrm{S}' \equiv \mathrm{S}''$ *and that there are no other transitions leading to a solution* $\mathrm{S}''' \equiv \mathrm{S}'$*. Notice also that according to the definition of* f'' *in the Example 1, the rates of the two transitions coincide. Let* λ *be such rate. We can conclude the example by observing that by application of the last rule of Table 1 we obtain*

$$[\mathrm{S}]_\equiv \xrightarrow{2\times\lambda} [\mathrm{S}']_\equiv$$

We conclude this section by observing that the operational semantics is very close to a Continuous Time Markov Chain (CTMC), with the difference that κ_F allows for self-transitions (i.e. transitions with the same source and target states). To obtain a CTMC we proceed as follows: given a set of reactions R and an initial solution S_0, we denote with $\mathrm{CTMC}(\mathrm{S}_0, R)$ the transition system obtained by removing from $\mathrm{LTS}(\mathrm{S}_0, R)$ the self-transitions and the transitions labeled with 0.

3 The Chemistry of κ_F

In the previous sections we have presented the κ_F language for the specification of bio-chemical systems, and its semantics defined in terms of a labeled transition system and a corresponding CTMC. We now show how to define a translation from κ_F systems to traditional chemistry. The existence of such a translation allows us to exploit on κ_F all those tools available for the so-called Chemical Reaction Networks (CRN) like, for instance, the discrete- and continuous-state simulation algorithms respectively based on the semantics of chemical reactions defined in terms of CTMCs or Ordinary Differential Equations (ODEs). The correctness of the translation is proved by showing that given a κ_F system, and the corresponding CRN, the two associated CTMCs are isomorphic.

Also in CRNs the basic component is the *molecule*, where each molecule belong to a *species*. Differently from Kappa and κ_F, chemical species have no structure but only a chemical species name. We will use X, Y, Z, ... to range over chemical species names.

Definition 9 (Chemical Reaction Network)**.**
A Chemical Reaction Network (CRN) is a set of chemical reactions of the form:

$$r \ : \ X_1 + \ldots + X_n \xrightarrow{\lambda} Y_1 + \ldots + Y_m$$

where r is a reaction identifier (we assume that reaction identifiers are pairwise different), with $n > 0$ and $m \geq 0$ ($m = 0$ means that the right hand side is empty), and such that $Y_1 + \ldots + Y_m$ cannot be obtained as a re-ordering of $X_1 + \ldots + X_n$. Moreover, λ is a strictly positive number representing the rate constant of the chemical reaction.

We now report the definition of the discrete-state semantics for CRNs, which is expressed in terms of a CTMC. Solutions are represented as multisets of species names, indicating the number of occurrences of molecules for every species. Given a solution $\mathcal{S}$, we denote with $\mathcal{S}(X)$ the number of molecules of species X in $\mathcal{S}$. Given a chemical reaction r

$$r \ : \ X_1 + \ldots + X_n \xrightarrow{\lambda} Y_1 + \ldots + Y_m$$

we respectively denote with $react(r) = \{X_1, \ldots, X_n\}$ and $prod(r) = \{Y_1, \ldots, Y_m\}$ the multiset of reactants —the left hand side of the reaction— and of products —the right hand side. With $setReact(r)$ we represent the set of species occurring among the reactants of r. For each $X \in setReact(r)$, with $r(X)$ we denote the number of occurrences of X in $react(r)$. Consider now a solution $\mathcal{S}$. We say that the chemical reaction r can occur if $react(r) \subseteq \mathcal{S}$. If the reaction occurs, the solution becomes $\mathcal{S}' = (\mathcal{S} \setminus react(r)) \uplus prod(r)$. The actual rate of the application of the reactions depends on the possible combinations of the reactants in the solution $\mathcal{S}$. Namely, if the rate constant of the reaction r is λ, the overall rate is:

$$rate(r, \mathcal{S}) \ = \ \lambda \times \prod_{X \in setReact(r)} \binom{\mathcal{S}(X)}{r(X)}$$

where $\binom{n}{k}$ is the binomial coefficient that computes the combinations of k elements taken among n available elements. We denote the possibility to perform such a chemical reaction with the notation $\mathcal{S} \longrightarrow_r \mathcal{S}'$.

We are now ready to define the CTMC associated to a CRN with initial solution.

Table 2: Chemical reactions for κ_F

$$\frac{\rho \in R \quad \mathrm{S} \xrightarrow{\lambda}_{\tilde{c},\rho} \mathrm{S}' \quad \mathrm{S}^{\tilde{c}} \equiv \mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon} \quad \textsc{minsol}(\mathrm{P}_1, \mathrm{P}_3) \quad \mathfrak{sol}(\mathrm{S}) \neq \mathfrak{sol}(\mathrm{S}') \quad \lambda \neq 0}{(\mathrm{S}, \tilde{c}, \rho) \ : \ \mathfrak{denot}(\mathfrak{sol}(\mathrm{S})) \xrightarrow{\lambda} \mathfrak{denot}(\mathfrak{sol}(\mathrm{S}'))}$$

Definition 10 (Discrete-state Semantics)**.**
Given an initial solution $\mathcal{S}_0$ and a set of reactions R, its discrete-state semantics is defined by the CTMC on chemical solutions, denoted with $DSS(\mathcal{S}_0, R)$, obtained as the minimal labeled transition system containing as initial state the initial solution $\mathcal{S}_0$ and the transitions $\mathcal{S} \xrightarrow{\lambda} \mathcal{S}'$ obtained as instantiations of the following rule:

$$\frac{\lambda = \sum_{(r \ : \ \mathcal{S} \longrightarrow_r \mathcal{S}')} rate(r, \mathcal{S}) \qquad \lambda > 0}{\mathcal{S} \xrightarrow{\lambda} \mathcal{S}'}$$

We now describe how to translate a κ_F system to a CRN. Intuitively, we associate to each complex a chemical species, and then we consider all the combinations of complexes that host the reactants of one of the κ_F reactions, thus triggering that reaction.

We assume the existence of a function $\mathfrak{species}(\_)$ that, given a κ_F complex, returns the corresponding species name in the CRN. Such a function satisfies the following property: given two complexes S and S$'$ we have that $\mathfrak{species}(\mathrm{S}) = \mathfrak{species}(\mathrm{S}')$ if and only if S$\equiv$S$'$. Given a κ_F solution S, we denote with $\mathfrak{sol}(\mathrm{S})$ the corresponding solution in the CRN. This function is defined as follows: if S$\equiv$S$_1, \ldots, S_n$ with S$_1, \ldots, S_n$ complexes, then $\mathfrak{sol}(\mathrm{S}) = \biguplus_{i=1\ldots n} \mathfrak{species}(\mathrm{S}_i)$.

In order to write chemical reactions in the form $X_1 + \ldots + X_n \xrightarrow{\lambda} Y_1 + \ldots + Y_m$ starting from a multiset of reactants and a multiset of products, we describe a (deterministic) way for denoting chemical solutions as $Z_1 + \ldots + Z_k$. We assume the existence of a total ordering relation $\preceq$ on species names. Given the chemical solution $\mathcal{S} = \{Z_1, \ldots, Z_k\}$ with $Z_i \preceq Z_j$ for every $0 \leq i < j \leq k$, we denote with $\mathfrak{denot}(\mathcal{S})$ the notation $Z_1 + \ldots + Z_k$.

We are now ready to define the CRN associated to a κ_F system. Technically speaking we proceed as follows. We use the transitions of the form S $\xrightarrow{\lambda}_{\tilde{c},\rho}$ S$'$, defined in Table 1, to generate chemical reactions identified by the triple $(\mathrm{S}, \tilde{c}, \rho)$. Then we select only a subset of these chemical reactions by taking one representative for each congruence class of κ_F solutions. Also in this case we have that the choice of the representative is not relevant as structurally congruent κ_F solutions have the same outgoing transitions (the unique observable effect is in the identifier of the chemical reactions that will include the selected representative S and its color $\tilde{c}$).

Definition 11. *Given a set of κ_F reactions R, we denote with $CRN(R)$ a maximal set of chemical reactions that can be inferred using the rule in Table 2 satisfying the following property: given two chemical reactions $(\mathrm{S}, \tilde{c}, \rho), (\mathrm{S}', \tilde{c}', \rho') \in CRN(R)$, if* S$\equiv$S$'$ *then* S $=$ S$'$.

Notice that for systems in which infinitely many distinct complexes can be generated, the corresponding CRN could be infinite as well. We will discuss in the Conclusions that, despite in Kappa the problem of checking finiteness of the generable complexes is undecidable, there are interesting fragments of Kappa in which this problem is decidable [16] and abstract interpretation techniques that over-approximate the set of all possible complexes [12].

Example 3. *We now discuss the translation into a CRN of the κ_F system defined in the Example 1, where we have formalized linear polymerization with binding rate λ and unbinding rate λ' (we consider the case of detaching rule where the rate is fixed), under the assumption that polymers do not form rings. We consider the chemical species P_i, with $i > 0$, to denote polymers of length i (notice that P_1 denotes a monomer free*

on both of its sites). Given a pair of polymers P_i and P_j, they can engage a binding reaction that produces P_{i+j}. It is interesting to observe that given a solution $\mathrm{S} = P_i, P_j$ composed of two such polymers, there are two distinct colorings that can be considered in the instantiation of the binding reaction pattern. The two colorings capture the two possible ways in which P_i and P_j can bind: the binding between the monomer of P_i with the first site free and the monomer of P_j with the second site free, and the binding between the monomer of P_i with the second site free and the monomer of P_j with the first site free. These two distinct colorings for the same solution $\mathrm{S} = P_i, P_j$, generates two instances of the following binding chemical reaction:

$$P_i + P_j \xrightarrow{\lambda} P_{i+j} \qquad \textit{(for every } i \textit{ and } j\textit{)}$$

We omit the reaction identifiers for simplicity. A similar discussion applies also to the unbinding reaction (see Example 2). Given a polymer P_l, it can generate two polymers P_i and P_j with $i + j = l$. But, given a solution composed of one polymer of length l, there are two possible colorings for each pair i, j (with $i \neq j$): one corresponding to the case in which the first produced polymer P_i comes from the first i monomers in P_l, and one corresponding to the case in which it comes form the last monomers. On the contrary, if $i = j$, we have that there is only one possible coloring because the unbinding monomers are uniquely identifiable as those in the middle of the polymer. So the CRN will include also two distinct instances for each of the following unbinding chemical reactions:

$$P_l \xrightarrow{\lambda'} P_i + P_j \quad \textit{(for every } l, i, j > 0 \textit{ s.t. } i \neq j \textit{ and } l = i + j\textit{)}$$

plus one instance for each of the following additional unbinding chemical reactions:

$$P_l \xrightarrow{\lambda'} P_i + P_i \qquad \textit{(for every } l, i > 0 \textit{ s.t. } l = i + i\textit{)}$$

Also in this case we omit the reaction identifiers.

We are finally ready to prove the correctness of the translation.

Theorem 1. *Given a set of κ_F reactions R, a corresponding chemical reaction network $CRN(R)$, and an initial κ_F solution S_0, we have that*

$$[\mathrm{S}]_{\equiv} \xrightarrow{\lambda} [\mathrm{S}']_{\equiv} \in \mathrm{CTMC}(\mathrm{S}_0, R)$$

if and only if

$$\mathfrak{sol}(\mathrm{S}) \xrightarrow{\lambda} \mathfrak{sol}(\mathrm{S}') \in DSS(\mathfrak{sol}(\mathrm{S}_0), CRN(R))$$

Proof. (Sketch) The κ_F transitions and the reactions of the corresponding CRN are defined similarly with a unique significant difference. Consider a solution S and a reaction ρ with a left hand side that can be instantiated in different ways in S: in the κ_F system the overall rate is computed by counting the different colors of the instantiations, while in the CRN there is a mixture of this technique and the traditional technique used in chemistry to count (through binomial coefficients) the number of combinations for selecting the reacting complexes inside S.

The unique interesting cases are when there are different colorings for S that give rise to the same instantiation of the functional rate, i.e. there are at least two colors $\tilde{c}$ and $\tilde{c}'$ such that $\mathrm{S}^{\tilde{c}} \equiv \mathrm{S}^{\tilde{c}'} \equiv \mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}, \mathrm{S}_2^{\epsilon}$ with $\mathrm{P}_1^{\tilde{c}_1} \xrightarrow{f_\rho}_{\rho} \mathrm{P}_2^{\tilde{c}_2}$ and $\textsc{minsol}(\mathrm{P}_1, \mathrm{P}_3)$ (so the functional rate is applied to $\mathrm{P}_1^{\tilde{c}_1}, \mathrm{P}_3^{\epsilon}$ in both cases). The rate of the corresponding transition in the CTMC multiplies the functional rate for the number of possible colorings. This latter depends on two factors: different ways for coloring the same complex, see Example 2, and multiple instances of the same complexes. Following the Example 2 consider, for instance, a solution S composed of three polymers of length l and consider a transition breaking one of them in two polymers of length i and j, with $i \neq j$. It is easy to see that there are 6 possible colorings, two for each polymer.

In the CTMC obtained from the CRN we will have the same rate for the corresponding transition. In fact, it is computed as follows. Consider first the number of instances of the corresponding chemical

reaction. This can be computed by considering the number of possible colorings $\tilde{c}''$ for transitions of the form $\mathrm{S}' \xrightarrow{\lambda}_{\tilde{c}'',\rho} \mathrm{P}_2, \mathrm{P}_3$ where S' is a subsolution of S containing only the complexes of the reactants. In the example of polymers above this number is 2. After, it is necessary to multiply such number for the number of possible ways in which the reacting complexes can be selected in the current solution (this multiplication is done by the function $rate(\_,\_)$). In the example, this multiplying factor is 3. The obtained number corresponds to the number of colorings considered by the κ_F semantics. □

4 Case studies

Example 1 constitutes a simple case that highlights the expressiveness of κ_F, despite the very conservative modeling approach with respect to the Kappa-calculus: in such example, functional rates are useful in order to forbid unwanted reactions, that would result difficult to avoid otherwise within the standard compositional modeling approach typical of Kappa. The main feature exploited there is the capability of reasoning about the number of complexes that are actually involved in the reaction.

In the following we are going to show that functional rates joined with the expressive power of Kappa can be exploited even further, in order to make reactions depend on almost any kind of physical or chemical property of the reacting complexes. We present two case studies highlighting the benefits of functional rates, the first one about the effect of mass on linear polymerization, the second one about the modelling of an artificial nano device.

4.1 Linear Polymerization

As a simple property to study, but relevant in particular for biochemical systems where sophisticated complexations take place —i.e. exactly those systems that can be suitably modeled in Kappa— we chose to consider the effect of mass on the kinetic rates of reacting complexes.

We can easily denote the mass of molecules in Kappa by adding a field to each species, with values $m \in \{0, \ldots, n_m\}$, where n_m is the number of distinct values for the masses of molecules considered in the system. The actual mass of each molecule is then obtained by a function $\mathfrak{m}(m)$, that can be exploited in κ_F to adjust reaction rates.

Let us consider again Example 1: if we add such information to each monomer of the species A, we obtain that each molecule $\mathrm{A}[u](\sigma)$ is denoted by one field storing the (index for $\mathfrak{m}(\cdot)$ of the) mass of the molecule, and two binding sites. For example, the polymer of length two would be denoted by $\mathrm{A}[1^{m_\mathrm{A}}](1 + 2^x)$, $\mathrm{A}[1^{m_\mathrm{A}}](1^x + 2)$, with $\mathfrak{m}(m_\mathrm{A})$ corresponding to the mass of each monomer of species A. The shape of reaction rules is exactly the same as before:

$$\mathrm{A}(1), \mathrm{A}(2) \xrightarrow{f_m} \mathrm{A}(1^x), \mathrm{A}(2^x)$$
$$\mathrm{A}(1^x), \mathrm{A}(2^x) \xrightarrow{f'_m} \mathrm{A}(1), \mathrm{A}(2)$$

On the contrary, the associated rate functions f_m, f'_m are modified to take into account the mass function $\mathfrak{m}(\cdot)$. The generality of the κ_F approach allows the modeler to express any kind of mass-dependent kinetics for the binding reaction: here we consider a simple relation based, according to [20], on the inverse dependence of the rate and the square root of the masses of the attaching polymers. The unbinding rate is considered

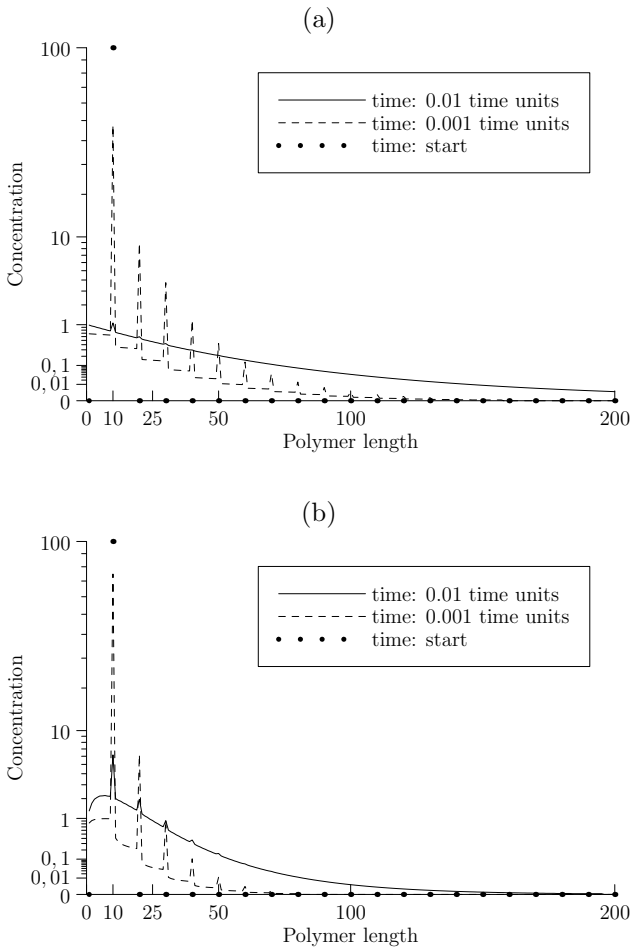

Figure 2: (a) Snapshots at three different time points (respectively $\mathbf{0}, \mathbf{10^{-2}}, \mathbf{10^{-3}}$ time units) of the average time-course simulation of the system described in Example 1, with initial solution of $\mathbf{10^2}$ polymers of length 10 and parameters $\lambda = \lambda' = 1$, $\mathfrak{m}(m_{\mathrm{A}}) = 1$. (b) Identical system depicted at the same time points, but with mass-dependent kinetics.

independent of the mass.

$$
\begin{array}{rcl}
f_m(\mathrm{S}^{\tilde{c}}) & = & \text{if S contains two complexes then} \\
 & & \quad \text{let } \mathrm{S}_1, \mathrm{S}_2 \text{ be two complexes in S} \quad \text{in} \\
 & & \quad \text{let } M_1 = \sum_{\mathrm{A}[1^m](\sigma) \in \mathrm{S}_1} \mathfrak{m}(m) \quad \text{in} \\
 & & \quad \text{let } M_2 = \sum_{\mathrm{A}[1^m](\sigma) \in \mathrm{S}_2} \mathfrak{m}(m) \quad \text{in} \\
 & & \quad \text{let } M_{12} = \frac{M_1 \cdot M_2}{M_1 + M_2} \quad \text{in} \\
 & & \qquad \lambda_m \cdot \sqrt{\frac{2}{M_{12}}} \\
 & & \text{else } 0 \\
f'_m(\mathrm{S}^{\tilde{c}}) & = & \lambda'
\end{array}
$$

f_m is designed so that λ_m represents the binding rate for monomers. Consequently, when $\lambda = \lambda_m$ then

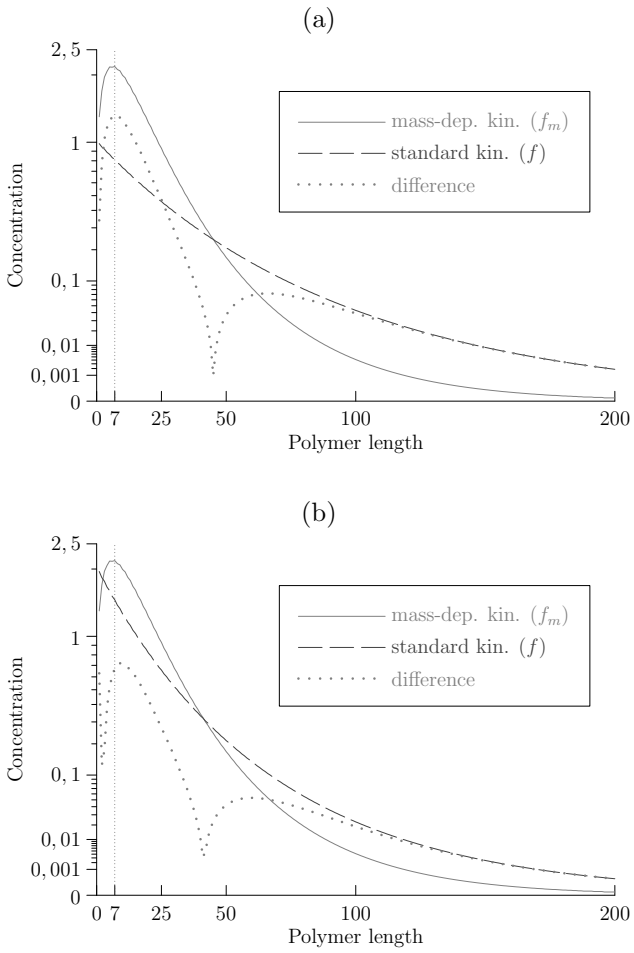

Figure 3: (a) Snapshot at steady state of the systems in Figure 2(a) and Figure 2(b), and their difference. (b) Snapshot at steady state of the systems in Figure 2(a) and Figure 2(b), and their difference, but with corrected binding rate constant $\lambda = \mathbf{0.5}$ for the system described in (a).

$f_m(\mathrm{S}^{\tilde{c}}) = f(\mathrm{S}^{\tilde{c}})$ only for $\mathrm{S}^{\tilde{c}}$ containing two totally disconnected monomers, while $f'_m = f'$ for any input solution.

It is possible to observe the average behavior of the two different kinetics represented by f and f_m in Figure 2, where time course simulations have been produced by translating the two corresponding κ_F systems to CRNs. In Figure 2(a), the system described in Example 1 is simulated with an initial solution composed of 10^2 polymers of length 10, with monomer mass $\mathfrak{m}(m_{\mathrm{A}}) = 1$, while reaction parameters are assigned the value $f = f' = 1$. Each graph corresponds to a snapshot of the system, taken at different times. The transient sawtooth-like shape of the middle graph is due to the binding of the polymers of length 10 composing the initial solution, that gives rise to temporary high concentrations of polymers with lengths multiple of 10. Near to steady state, the concentration of polymers as a function of their lengths approaches a decreasing exponential distribution. Figure 2(b) depicts the behavior of the system with same initial solution, values for parameters and time points for snapshots, but with the kinetics represented by function f_m. It is possible to observe a similar transitory sawtooth shape, although steady state is reached at a slightly later time.

More surprisingly, at steady state the shapes of the curves corresponding to the two kinetics are considerably different, as shown in Figure 3(a). It is quite reasonable to expect a lower concentration of long

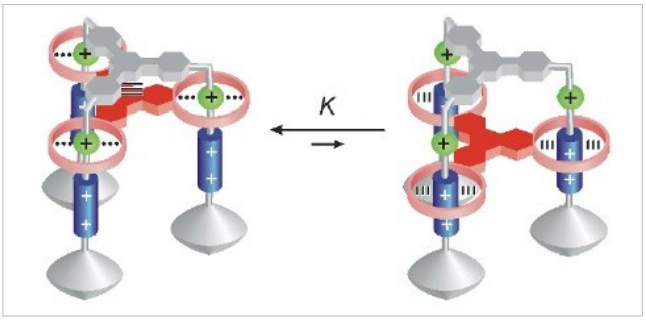

Figure 4: Schematic representation of the nanoscale elevator in [2]. The platform (the red structure) is connected to the rings of three 2-stations rotaxanes (the vertical legs) which are fused together on their top parts (the grey structure). The platform is moved by the stimuli that shift the rotaxane rings from one station to the other one.

polymers, due to the lower binding rates at greater lengths, but surprisingly the curve produced by mass-dependent kinetics is characterized by a maximum around lengths 5-7, instead of length 1 as for standard kinetics. The different shape and the lower (on average) binding rate for polymers gives rise to a difference of concentrations of the same order of magnitude of the two curves. Of course a partial correction for such difference could be easily introduced by lowering the binding rate for standard kinetics: Figure 3(b) depicts the same systems with parameters $\lambda = 0.5$ and $\lambda_m = 1$. Despite the approaching of the curves, their different shape makes it impossible to obtain a reasonable super-imposition.

4.2 The Nanoscale Elevator

As a second case study we consider the molecular machine behaving like a nanoscale elevator presented in [2]. This nano device (schematically depicted in Figure 4) is obtained by integration of several structural and functional molecular subunits. The subunits used to move the elevator platform are three bistable rotaxanes. Rotaxanes are systems composed of a molecular axle surrounded by a ring-type (macrocyclic) molecule. Bulky chemical moieties ("stoppers") are placed at the extremities of the axle to prevent the disassembly of the system. In rotaxanes containing two different recognition sites on the axle ("stations"), it is possible to switch the position of the macrocyclic ring between the two stations by an external energy input. In particular, the rotaxanes used in [2] have two stations, an ammonium/amine molecule (*Nh* in the following) green colored in Figure 4 and a bipyridinium molecule (*Bipy* in the following) colored in blue. The *Nh* molecule can be protonated and deprotonated by adding acid or base to the solution: when it is protonated the stable position for the ring is on the *Nh* station (as depicted in Figure 4), while it is on the *Bipy* station when it is deprotonated.

The behavior of such rotaxane has been modeled in [8] by using a Kappa-like language extended with instantaneous reactions. These latter were used to immediately communicate to all the molecules belonging to the same rotaxane the occurred (de)protonation of the *Nh*. This is no longer needed in κ_F as functional rates allow the modeler to express the influence of the internal state of the *Nh* molecule on the behavior of the entire rotaxane.

We model the rotaxane by considering three distinct molecules for representing the *Nh*, the *Bipy* and the ring, respectively. The two stations are connected by a permanent bond, while the ring has a switchable binding to one of them (such bond indicates the current location of the ring).

We complete the model of the nanoscale elevator by considering how the three rotaxanes are connected together. Following the structure depicted in Figure 4 we add to each of the three rotaxanes a *Top* molecule connected to the *Nh* station, and we bind together the three *Top* molecules. Also the three rings are

connected together to represent the platform. The complete representation of the elevator is then as follows:

$$
\begin{aligned}
&\mathit{Top}(l^{t_1} + \mathit{nh}^{s_1} + r^{t_2}),\ \mathit{Top}(l^{t_2} + \mathit{nh}^{s_2} + r^{t_3}),\ \mathit{Top}(l^{t_3} + \mathit{nh}^{s_3} + r^{t_1}),\\
&\mathit{Nh}[h^1](\mathit{top}^{s_1} + \mathit{bipy}^{r_1} + \mathit{ring}^{x_1}),\ \mathit{Ring}(l^{p_1} + \mathit{link}^{x_1} + r^{p_2}),\ \mathit{Bipy}(\mathit{nh}^{r_1} + \mathit{ring}),\\
&\mathit{Nh}[h^1](\mathit{top}^{s_2} + \mathit{bipy}^{r_2} + \mathit{ring}^{x_2}),\ \mathit{Ring}(l^{p_2} + \mathit{link}^{x_2} + r^{p_3}),\ \mathit{Bipy}(\mathit{nh}^{r_2} + \mathit{ring}),\\
&\mathit{Nh}[h^1](\mathit{top}^{s_3} + \mathit{bipy}^{r_3} + \mathit{ring}^{x_3}),\ \mathit{Ring}(l^{p_3} + \mathit{link}^{x_3} + r^{p_1}),\ \mathit{Bipy}(\mathit{nh}^{r_3} + \mathit{ring})
\end{aligned}
$$

where we use mnemonic names to represent sites and fields. In the first line we present the three *Top* molecules each one connected to a left and a right *Top* molecule. Moreover, each *Top* is connected to the *Nh* molecule of one rotaxane. The three rotaxanes are represented in the subsequent three lines. Notice that we assume that the *Ring* molecules are connected to the *Nh* station, and that each *Ring* is connected to a left and a right *Ring* molecule. The *Nh* molecules have one field h: the field holds 0 when the *Nh* is deprotonated, it holds 1 otherwise. We assume the *Nh* molecules initially protonated.

We now move the representation of the dynamics of the system. Two kinds of reactions are used: those for protonation/deprotonation between the *Nh* and an acid-base molecule, and those for switching the bond between the ring and the two stations. The rate of the ring movement from one station to the other one depends on the protonated/deprotonated state of the *Nh*. We model this dependency by using a functional rate. The two ring movement reactions are as follows:

$$
\begin{aligned}
&\mathit{Nh}(\mathit{bipy}^{r_1} + \mathit{ring}^{x_1})^{c_1},\ \mathit{Ring}(\mathit{link}^{x_1})^{c_2},\ \mathit{Bipy}(\mathit{nh}^{r_1} + \mathit{ring})^{c_3} \overset{f_{\text{mov}}}{\longleftrightarrow}\\
&\mathit{Nh}(\mathit{bipy}^{r_1} + \mathit{ring})^{c_1},\ \mathit{Ring}(\mathit{link}^{x_1})^{c_2},\ \mathit{Bipy}(\mathit{nh}^{r_1} + \mathit{ring}^{x_1})^{c_3}
\end{aligned}
$$

$$
\begin{aligned}
f_{\text{mov}}(\mathrm{S}^{\tilde{c}}) \ = \quad & \text{let } \mathit{Nh}[h^x](\mathit{ring}^y + \sigma)^{c_1} \in \mathrm{S}^{\tilde{c}} \qquad \text{in}\\
& \quad \text{if } y = \varepsilon \text{ then}\\
& \quad\quad \text{if } x = 0 \text{ then } \lambda^1_{\text{mov}} \text{ else } \lambda^2_{\text{mov}}\\
& \quad \text{else}\\
& \quad\quad \text{if } x = 0 \text{ then } \lambda^3_{\text{mov}} \text{ else } \lambda^4_{\text{mov}}
\end{aligned}
$$

According to the κ_F semantics, the solution $\mathrm{S}^{\tilde{c}}$ that is passed to the functional rate f_{mov}, will be the part of the current solution composed of the elevator to which the reacting rotaxane belongs. We use the color c_1 to identify the *Nh* molecule of the reacting rotaxane. The functional rate returns one of four possible rates λ^i_{mov}, depending on the combination of two distinct factors: whether the *Nh* is protonated or not, and whether the *Ring* is moving from the *Nh* to the *Bipy* or vice versa.

We now consider the second kind of reactions that are concerned with the proton exchange between the *Nh* and the acid-base molecules. The rate of these reactions are influenced by an interesting phenomenon observed on the behavior of the nanoscale elevator. The (de)protonation of the three *Nh* molecules of an elevator follows three distinct processes. Upon addition of acid-base to the solution, the (de)protonation effect do not distribute homogeneously among the *Nh* molecules, but among the elevators. Namely, the *"first equivalent of base does not lead to a statistical mixture of differently protonated species but rather causes the first deprotonation process to occur"*. One likely cause of this phenomenon is that the (de)protonation rate of the *Nh* is influenced by the current (de)protonated state of the other two *Nh* molecules in the same elevator. According to this interpretation, the protonation/deprotonation reactions can be modeled as follows:

$$
\begin{aligned}
&\mathit{Nh}[h^1], \mathit{Base}[h^0] \quad \underset{\xleftarrow{f_{\text{b-prot}}}}{\xrightarrow{f_{\text{b-deprot}}}} \quad \mathit{Nh}[h^0], \mathit{Base}[h^1]\\
&\mathit{Nh}[h^0], \mathit{Acid}[h^1] \quad \underset{\xleftarrow{f_{\text{a-prot}}}}{\xrightarrow{f_{\text{a-deprot}}}} \quad \mathit{Nh}[h^1], \mathit{Acid}[h^0]
\end{aligned}
$$

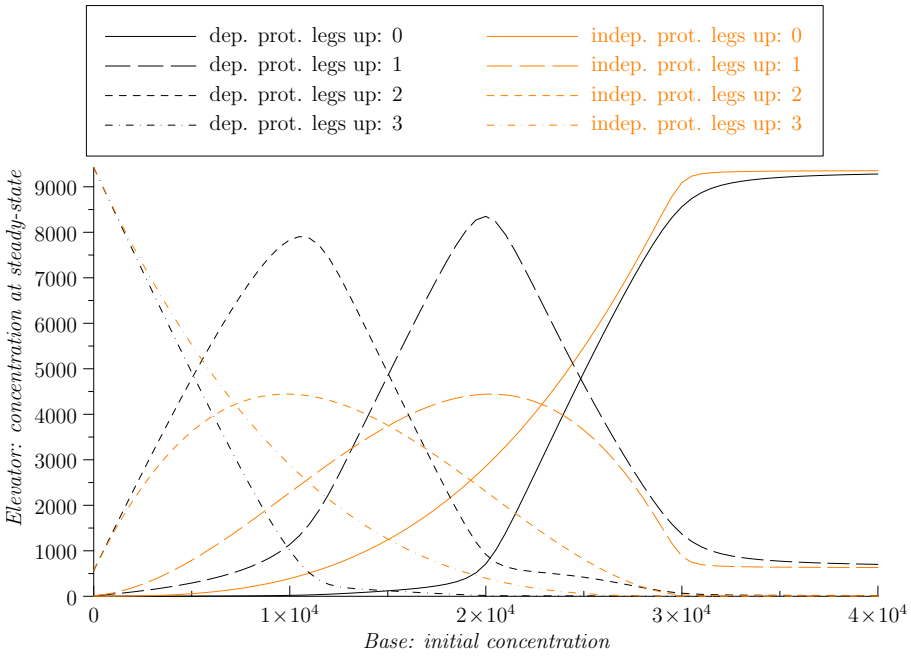

Figure 5: Comparison of possible behaviors of the elevator model at steady state in relation to the assumption of dependence or independence of the protonation/deprotonation process on the total number of already protonated *Nh* molecules. On the horizontal axis the initial number of $Base[h^0]$ molecules in the solution: each point of the graph represents the result of a run with different initial number of base molecules. On the vertical axis the corresponding number of molecules at steady state for different configurations and functional rates of the elevator, starting with an initial concentration of 10^4 elevators each with all the 3 *Nh* molecules protonated. The four species listed on the left of the legend represent the number of elevators with respectively 0 to 3 "legs up" (i.e. number of rotaxanes whose ring is bound to the *Nh* molecule) in the case of movement dependent on the number of protonated *Nh* molecules as described in (1), with $\lambda_{\text{b-deprot}} = s_{\text{b-deprot}} = 1, \lambda_{\text{b-prot}} = 10^{-2}, s_{\text{b-prot}} = -1$. The four species on the right represent the same elevator states but in the case of movement independent of the total number of protonated *Nh* molecules, that is with $s_{\text{b-deprot}} = s_{\text{b-prot}} = 0$ and $\lambda_{\text{b-deprot}} = 10^2, \lambda_{\text{b-prot}} = 10^{-2}$. For both the dependent and the independent cases, we have $\lambda^1_{\text{mov}} = \lambda^4_{\text{mov}} = 20, \lambda^2_{\text{mov}} = \lambda^3_{\text{mov}} = 10^3$.

$$f_{\text{k}}(\text{S}^{\tilde{c}}) \;=\; \text{let } P = \textstyle\sum_{Nh[h^x](\sigma)\in \text{S}^{\tilde{c}}} x \quad \text{in} \quad (\lambda_{\text{k}}) \cdot 10^{s_{\text{k}} \cdot P} \qquad \text{for k} \in \{\text{b-deprot}, \text{b-prot}, \text{a-deprot}, \text{a-prot}\} \tag{1}$$

where we use *Base* and *Acid* molecules with a field h which holds 0 or 1 to denote whether the molecule is ready to receive or donate a proton, respectively. In this case, the functional rate modifies a base rate λ_{k} according to the number P of *Nh* protonated in the same elevator.

The translation to CRNs of the elevator model generated a network with 26 chemical species (2 for the protonated/deprotonated acid or base, 24 for all the possible states of each elevator) and 144 chemical rules, which was then easily analyzed by means of deterministic simulation. Figure 5 shows a comparison of the different behavior at steady state of such model in relation to the assumption of dependence or independence of the (de)protonation process on the total number of protonated *Nh* molecules in each elevator.

The results of our simulations are reported in Figure 5. It is worth noticing that the curves for the case with (de)protonation rate dependent on the state of the other *Nh* molecules in the same elevator highlight (with their sharpness) the existence of three distinct (de)protonation phases, which are not observable if

such rates are independent (see the smooth curves instead).

4.3 Performance and Scalability

The translation of κ_F models to chemical reaction networks allowed us to exploit already existing tools for the analysis of chemical systems (such as Copasi [22], and in general any SBML-compatible tool [23]). In particular we were able to adopt deterministic simulation, that was the most suitable technique for the case studies presented here both in terms of efficiency and of the kind of information that we needed to show. To this purpose, the translation of the κ_F system had to be applied before the simulation, so that the CRN could be provided as input to the used simulation tool.

It is worth remarking that in this way implementation efforts are minimized, but important drawbacks may come from the size of the generated CRN, which constitutes the main bottleneck of the presented approach. When the number of corresponding chemical species generated during the translation is very high (or infinite) one is usually forced to manually set an upper bound and truncate the translation at an arbitrary point. Two are the main disadvantages then. First, the truncation might have negative effects on the reliability of the simulation, if relevant chemical species were not included. Second, the computational cost of the translation may make this approach inefficient, because several species (and reactions) that on the contrary are not relevant for the simulation are generated anyway. In practice it is often the case that manual truncation can be safely applied, since reasonable upper bounds can be established with few attempts.

A more clever approach, that however does not allow the exploitation of existing simulation tools, consists in considering the generation on the fly of chemical species and reactions at simulation time. In this way it is possible to minimize the cost of the translation as well as of the simulation, and even to exclude errors due to improper truncation: the only requirement is that at any time during the simulation the number of species with non zero concentration is small enough to fit the maximal available memory. This approach can be applied in general for many simulation techniques, including stochastic but, remarkably, also deterministic simulation. In fact, although under the deterministic assumption at any time $t > 0$ every producible chemical species is theoretically present with positive concentration, the numerical approximation introduced by the finite representation of real numbers considerably reduces the set of species with numerically detectable non zero concentration. Therefore, even in the case of deterministic simulation it is possible to deal with systems denoted by unbounded number of species without introducing any approximation error coming from the truncation of the CRN. One simple example that could be handled in this way is represented by (reversible) linear polymerization without a priori bound on the maximal length of polymers.

Since in the case studies presented here the number of chemical species of the corresponding CRNs was bounded, the generation on the fly of the CRN was not needed. So we chose to implement preliminary translation with manual truncation, that in our case did not entail any loss of precision or inefficiency. A naive Prolog implementation allowed us to generate the CRNs in a reasonable time: fractions of a second for the molecular elevator, while for the polymerization example the quite high number of corresponding chemical reactions (250 thousand for an initial solution with 1 thousand monomers) required eight minutes with a standard desktop computer (single thread implementation, 3 GHz CPU). For these models, translation times were marginal with respect to simulation times, consisting in about 3 minutes for each run of the elevator model and 30 minutes for each run of the linear polymerization model.

Without particular implementation efforts or dedicated hardware, practical upper bounds to the size of CRNs can be currently placed around few tens of millions roughly in terms of sum of number of species and reactions. Beyond this limit, two are the main issues to be solved: the time required for the generation of the CRN, and the memory needed. The first issue can be mitigated by parallelization of the translation, that can be straighforwardly implemented by proper splitting of rules and generated complexes among different threads, with some overhead coming from unavoidable synchronization of the parallel instances. The second issue defines instead quite sharply, in terms of the maximal available computer memory, the limit of applicability of the approach based on translation to CRNs.

5 Conclusion

The *"don't care, don't write"* approach adopted in the Kappa-calculus, as well as in other rule-based languages like BioNetGen [4], opened the way for introducing compositional modeling in rule-based process calculi, and provides very compact and readable descriptions of biochemical systems in the presence of sophisticated molecule bindings. While compositional modeling represents in general a desirable advantage in the hands of the modeler, it becomes a limit when important properties of the system cannot be described in a compositional calculus because of their intrinsic non-compositionality.

In this paper we applied to the Kappa-calculus a technique of general applicability for the extension of process calculi for biochemical modeling. The resulting extended calculus, that we called κ_F, allows us to take into account non-compositional properties (physical, chemical, etc.) of the modeled systems without losing the advantage of a compositional description. The extension technique consists in the introduction of functional rates for biochemical rules, which are calculated as functions not only of the reactants of the rule but also of the whole set of molecules linked to them. Thanks to the wide applicability of this approach, similar results may be obtained also for many other process calculi with binding capabilities (e.g. [18, 31, 32, 33]).

In the Introduction we have already commented the increase of expressiveness of κ_F with respect to Kappa and the possibility in the Kappa simulator `KaSim` to associate to reactions a pair of rates, the first one to be used when the reactants do not belong to the same complex, the second case otherwise. This latter mechanism is useful to resolve the ambiguity of Kappa rules among two reactants A and B that could be applied in a context where A and B are sometimes already connected and sometimes disconnected. Indeed, this would lead to an inconsistency in the definition of the kinetic rate which should have a volume dependency in the former case and no volume dependency in the latter. Nevertheless, the introduction of physical or chemical properties that influence reaction rates as functions of the whole involved complexes makes unfeasible any attempt of modeling in Kappa. On the contrary, κ_F turned out to be suitable, as shown for the model of linear polymerization in the presence of mass-dependent kinetics (see Section 4.1).

The stochastic simulator NFsim [34], based on an extension of the BioNetGen language, allows the expression of rate functions which can depend on properties either global (at the level of the system) or "local" (at the level of the molecular complexes involved in the reaction). While the first kind of properties is not directly included in κ_F and should be encoded manually by the modeler, the latter kind makes NFsim capabilities closer to κ_F. However, the adoption of *colors* in κ_F semantics allows the modeler to take into account more sophisticated properties which depend not only on the number of molecules of any kind that form each molecular complex, but also on the way they are arranged to form the complex. In other words, only in κ_F rate functions can exploit the information pertaining the graph-like structure of each complex involved in the reaction and the position of reacting molecules inside them.

Despite the expressiveness of κ_F, we provided its formal translation in traditional chemistry and proved the correctness. If we think of chemical reaction networks as stochastic Petri nets [28], then our approach is a generalization to the stochastic context of the idea applied in [29] to map the π-calculus to standard Petri nets.

Thanks to this translation it is possible to apply to (some classes of) κ_F models the efficient verification techniques (such as simulation by ordinary differential equations, as well as by efficient stochastic algorithms [19, 5]) and reuse, at least in principle, the existing software tools developed for traditional chemistry (e.g [7], but in general any tool supporting languages comparable to traditional chemistry, like SBML [24]).

The behavior of the model of linear polymerization was indeed analyzed by means of deterministic simulation after automated translation to chemistry, which allowed us to observe, in the presence of mass-dependent effects (particularly relevant for complexes with high number of components), the inconsistency of usual modeling approaches.

However, the superior expressiveness of process calculi like Kappa with respect to traditional chemistry can lead to translations with an infinite number of chemical species and rules. In general, as the Kappa-calculus is Turing complete, the problem of checking whether a Kappa model can generate only finitely many complexes is undecidable. Nevertheless, there are fragments of Kappa for which this problem turns out to be decidable [16], and also techniques —based on abstract interpretation— which are capable of computing

an over-approximation of the set of reachable complexes [12] that can be used to prove, in some cases, that this set is finite.

Future work directions point at several aims. First, it must be investigated to what extent the introduction of functional rates in bio-oriented process calculi semantics can be pushed, in particular in those calculi equipped with high-level structural rearrangement primitives (for example, calculi with nested compartments [6, 33]). Moreover, it is still unclear how this approach can be adapted to cope with more sophisticated calculi like [26, 25], where functional rates have been already introduced but with a different technique, based on communication constraints. A special case is the React(C) language [27], that can be regarded as an extension of Kappa as well as of κ_F, but differently from κ_F the functional rate takes under consideration the entire system (i.e. the entire solution). We found the κ_F approach more appropriate when one wants to specify models which are modular at least at the level of complexes, even if not modular at the level of the single molecules. It must be investigated how and to what extent the technique used for the translation of κ_F to traditional chemistry can be generalized in order to be applied to React(C) as well, thus giving the possibility to introduce also in React(C) a notion of complex. More sophisticated mappings to traditional chemistry (e.g. along the line of [30]) may be helpful in this regard.

Last, more efficient translation techniques may allow us to widen the class of models manageable in practice. For example, abstract interpretation has been already applied to Kappa [11] to reduce, under some circumstances, the number of chemical species and reactions resulting from a translation from Kappa to standard chemistry similar to the one we have presented in this paper. In order to apply such techniques in κ_F, it is necessary to check their applicability in the presence of the functional rate.

References

[1] KaSim: kappa language simulator. http://www.kappalanguage.org.

[2] J. D. Badjić, V. Balzani, A. Credi, S. Silvi, and J. F. Stoddart. A molecular elevator. *Science*, 303(5665):1845–1849, 2004.

[3] V. Balzani, A. Credi, and M. Venturi. *Molecular devices and machines - Concepts and perspectives for the nano world, 2nd Edition*. Wiley-VCH, Weinheim, 2008.

[4] M. Blinov, J. Faeder, B. Goldstein, and W. Hlavacek. Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289, 2004.

[5] Y. Cao, H. Li, and L. Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys*, 121(9):4059–4067, September 2004.

[6] L. Cardelli. Brane Calculi. In Danos and Schächter [14], pages 257–278.

[7] F. Ciocchetta, A. Duguid, S. Gilmore, M. L. Guerriero, and J. Hillston. The bio-pepa tool suite. In *QEST*, pages 309–310. IEEE Computer Society, 2009.

[8] A. Credi, M. Garavelli, C. Laneve, S. Pradalier, S. Silvi, and G. Zavattaro. nanok: A calculus for the modeling and simulation of nano devices. *Theor. Comput. Sci.*, 408(1):17–30, 2008.

[9] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling of cellular signalling. In *CONCUR*, volume 4703 of *Lecture Notes in Computer Science*, pages 17–41. Springer, 2007.

[10] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling, symmetries, refinements. In *FMSB*, volume 5054 of *Lecture Notes in Computer Science*, pages 103–122. Springer, 2008.

[11] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Abstracting the differential semantics of rule-based models: Exact and automated model reduction. In *LICS*, pages 362–381. IEEE Computer Society, 2010.

[12] V. Danos, J. Feret, W. Fontana, and J. Krivine. Abstract interpretation of cellular signalling networks. In *VMCAI*, volume 4905 of *Lecture Notes in Computer Science*, pages 83–97, 2008.

[13] V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110, 2004.

[14] V. Danos and V. Schächter, editors. *Computational Methods in Systems Biology, International Conference CMSB 2004, Paris, France, May 26-28, 2004, Revised Selected Papers*, volume 3082 of *Lecture Notes in Computer Science*. Springer, 2005.

[15] P. Degano and R. Gorrieri, editors. *Computational Methods in Systems Biology, 7th International Conference, CMSB 2009, Bologna, Italy, August 31-September 1, 2009. Proceedings*, volume 5688 of *Lecture Notes in Computer Science*. Springer, 2009.

[16] G. Delzanno, C. D. Giusto, M. Gabbrielli, C. Laneve, and G. Zavattaro. The *kappa*-lattice: Decidability boundaries for qualitative analysis in biological languages. In Degano and Gorrieri [15], pages 158–172.

[17] J. R. Faeder, M. L. Blinov, and W. S. Hlavacek. Rule-based modeling of biochemical systems with bionetgen. *Methods in Molecular Biology*, 500:113167, 2009.

[18] F. Fages and S. Soliman. Formal cell biology in biocham. In *SFM*, volume 5016 of *Lecture Notes in Computer Science*, pages 54–80, 2008.

[19] M. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[20] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[21] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka, and W. Fontana. Rules for modeling signal-transduction systems. *Science Signaling*, 2006(344), 2006.

[22] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. Copasi – a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[23] M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, H. Kitano, A. Arkin, B. Bornstein, D. Bray, A. Cornish-Bowden, et al. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[24] M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, H. Kitano, A. Arkin, B. Bornstein, D. Bray, A. Cornish-Bowden, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, 2003.

[25] M. John, C. Lhoussaine, and J. Niehren. Dynamic compartments in the imperative *pi*-calculus. In Degano and Gorrieri [15], pages 235–250.

[26] M. John, C. Lhoussaine, J. Niehren, and A. M. Uhrmacher. The attributed pi calculus. In M. Heiner and A. M. Uhrmacher, editors, *CMSB*, volume 5307 of *Lecture Notes in Computer Science*, pages 83–102. Springer, 2008.

[27] M. John, C. Lhoussaine, J. Niehren, and C. Versari. Biochemical reaction rules with constraints. In G. Barthe, editor, *ESOP*, volume 6602 of *Lecture Notes in Computer Science*, pages 338–357. Springer, 2011.

[28] M. A. Marsan. Stochastic Petri nets: an elementary introduction. In G. Rozenberg, editor, *European Workshop on Applications and Theory in Petri Nets*, volume 424 of *Lecture Notes in Computer Science*, pages 1–29. Springer, 1988.

[29] R. Meyer. A theory of structural stationarity in the π-calculus. *Acta Inf.*, 46(2):87–137, 2009.

[30] R. Meyer and R. Gorrieri. On the relationship between π-calculus and finite place/transition petri nets. In M. Bravetti and G. Zavattaro, editors, *CONCUR*, volume 5710 of *Lecture Notes in Computer Science*, pages 463–480. Springer, 2009.

[31] A. Phillips and L. Cardelli. Efficient, correct simulation of biological processes in the stochastic pi-calculus. In *CMSB*, volume 4695 of *LNCS*, pages 184–199, 2007.

[32] C. Priami and P. Quaglia. Beta binders for biological interactions. In Danos and Schächter [14], pages 20–33.

[33] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Y. Shapiro. BioAmbients: an abstraction for biological compartments. *Theor. Comput. Sci.*, 325(1):141–167, 2004.

[34] M. W. Sneddon, J. R. Faeder, and T. Emonet. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nature Methods*, 8:177183, 2011.