Available online at www.sciencedirect.com
ScienceDirect
Electronic Notes in Theoretical Computer Science
ELSEVIER
Electronic Notes in Theoretical Computer Science 268 (2010) 77–96
www.elsevier.com/locate/entcs

Fragments-based Model Reduction: Some Case Studies

Jérôme Feret[1,2]

Laboratoire d'informatique de l'École normale supérieure
(INRIA/ENS/CNRS)
Paris, France

Abstract

Molecular biological models usually suffer from a dramatic combinatorial blow up. Indeed, proteins form complexes and can modify each others, which leads to the formation of a huge number of distinct chemical species (ie non-isomorphic connected components of proteins). Combinatorial complexity forbids an explicit description of the quantitative semantics (stochastic or differential), since the set of states is usually a vector space the dimension of which is the number of distinct chemical species. Model reduction aims at reducing this complexity by providing another grain of observation. Fragments-based reduction consists in computing a set (hopefully smaller than the set of chemical species) of pieces of chemical species, such that the evolution of the number (or concentration) of these pieces can be soundly described in self-consistent abstract quantitative semantics. In this paper, we provide several intuitive examples so as to give some intuition about why this approach may work; and why stochastic semantics are more difficult to abstract than differential semantics.

Keywords: rules-based modeling, model reduction, differential semantics, stochastic semantics, abstract interpretation.

1 Introduction

Signaling pathways are made of several kinds of proteins which may interact with each other via complexation and posttranslational modification (such as phosphorilation). These interactions enable the communication between cells, and within each cell: thanks to these interactions, a cell can receive signals, propagate and integrate these signals so that a specific cellular response (cell death, cell proliferation, cell differentiation, and so on and so forth) can be triggered. Signaling pathways usually suffer for a combinatorial blow-up in the number of chemical species (that is the number of non isomorphic connected components which can occur at run-time),

[1] We would like to thank all the people who have collaborated with us on these topics, especially Vincent Danos, Walter Fontana, Russell Harmer, and Jean Krivine for their contribution in the ODE-fragments framework, and, Heinz Koeppl and Tatjana Petrov for their contribution in the stochastic-fragments framework. This work was partially supported by the AbstractCell ANR-Chair of Excellence.

[2] Email: feret@ens.fr

doi:10.1016/j.entcs.2010.12.007

which makes hard both their specification and their modeling. Typically, Even a simple model of the EGF receptor signaling network can generate more than 10^{23} non-isomorphic species [14],

Rules-based modeling [18,2] offers an elegant and compact solution for describing signaling pathways (and other molecular biological systems as well). The main principle is that potential interactions can be described without specifying all the potential context of application in which this interaction is enabled. This way, a rule can be seen as a symbolic description of a (potentially infinite) set of chemical reactions.

Yet, the combinatorial complexity raises again when one wants to define and compute the behavior of a model. Two kinds of quantitative semantics are usually considered: the stochastic semantics and the differential semantics. In stochastic semantics [21], a continuous-time Markov chain or a weighted labeled transition system can be associated to a model, and used so as to define the state occupancy (density) distribution, or the trace density distribution (which is more fine-grained). Another choice has to be made about the description of the state of the system, which can be denoted by either a graph with identified agents (individual-based semantics), or a multi-set of chemical species (population-based semantics). Since the computation of all these distributions is usually to complex, the trace density distribution is usually sampled by using Gillespie's simulation method [5,23,22]. In differential semantics [20,16], the state of the system is described as a vector of species concentrations, and a differential system of equations gives the evolution of these concentrations.

Both species-based numerical stochastic simulation (such as in [24,19]) and differential semantics numerical integration require enumerating species (and reactions) either beforehand, or on-the-fly. Thus, these approaches suffer for the combinatorial number of potential chemical species. The use of individual-based stochastic simulation methods (such as in [17]) avoids the explicit enumeration of species and reactions, but they require an explicit description of each protein of the system in memory. Thus, individual-based approaches do not scale up when they are too many instances of proteins. It follows that the number of instances of each protein and the number of potential distinct chemical species are important parameters. They define what we call *the two combinatorial walls* beyond which it is no longer possible to compute the quantitative properties of a model.

Model reduction [4,11,10,3,9,20,16,21] aims at reducing the number of variables in quantitative semantics by providing another grain of observation. Fragments-based reduction consists in computing a set (hopefully smaller than the set of chemical species) of pieces of chemical species, such as the evolution of the number (or concentration) of these pieces can be soundly described in self-consistent abstract quantitative semantics. In [20,16,21], these reduced systems are automatically extracted from the rules-based description of models, without ever enumerating neither the set of reactions, nor the set of chemical species. The relation between the so obtained reduced semantics and the initial ones are formalized by abstract interpretation [13,12]: the solution of the reduced differential system is the exact

projection of the solution of the initial system and the density distribution of the reduced stochastic semantics is also the exact projection of the density distribution of the initial stochastic semantics.

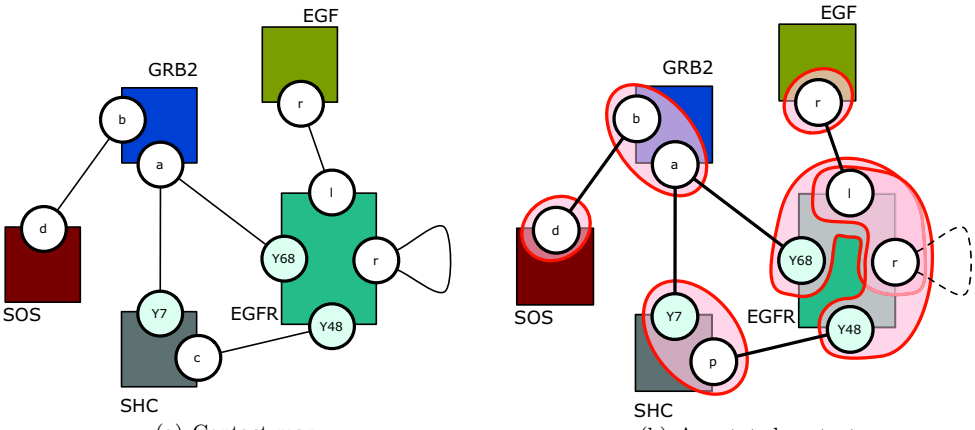
The reduction of stochastic semantics is intrinsically more difficult than the reduction of differential semantics. The framework in [20,16] for reducing differential semantics is based on the fact that rules cannot observe the correlation between specific parts of some chemical species. Thus these chemical species can easily be cut into fragments. It turns out that stochastic semantics can observe much more correlations, which makes the approach which is proposed in [20,16] inefficient for reducing stochastic semantics. In [21] backward bisimulations [8] are used in order to ensure that rules cannot enforce correlations between the state of some identified parts of chemical species, so as to show that the stochastic system is weakly lumpable [7,26]. Nevertheless, the induced reduction is correct only if there is no correlation between the state of fragments at initial time.

In this paper, we provide several intuitive examples so as to give some intuition about why fragments-based reduction may work; and why it is more difficult to abstract stochastic semantics than differential semantics.

In Sect. 2, we explain on a model of the early events of the EGFR pathway, why it should be possible to split chemical species into smaller fragments so as to reduce quantitative semantics of rules-based models. In Sect. 3, we detail an example of a model which can be split into two independent subsystems (or modules), and we use this property to reduce both the differential and the stochastic semantics. In Sect. 4, we show an example of coupled semi-reactions in which a given reaction application operates on two fragments simultaneously. We show that it raises no issue when reducing the differential semantics, whereas it forbids some reduction in the case of the stochastic semantics. In Sect. 5, we show that, in the stochastic semantics, one protein can control the behavior of another one even if they are not in the same connected components in the left hand side of a reaction. In Sect. 6, we report the dimension of the state space of some pathways and the dimension of the state space of the reduced systems (both for the differential semantics and the stochastic semantics) for several examples, and we conclude.

2 A breach in the combinatorial walls

In this Section, we show on the example of the early events of the *EGF* pathway [1], why it should be possible to reduce the dimension of quantitative semantics by considering fragments of chemical species. In the early events of the *EGF* pathway, some membranal receptors *EGFR* are activated by some ligands *EGF*, which initiates a cascade of interactions, which *in fine* leads to the recruitment by the receptors *EGFR* of some proteins *SOS*. The proteins *SOS* are carried by some transport molecules *GRB2*. These transport molecules can be recruited by the receptors according to two different ways. The set of proteins and the potential bonds between these proteins are summarized in a contact map, which is given in Fig. 1(a). The interactions between proteins can be described in Kappa [18] by the



(a) Contact map. (b) Annotated contact map.

Fig. 1. Maps for the early EGF model.

39 rules (25 unidirectional rules and 7 bidirectional rules) which are given in Fig. 2. We do not provide kinetic rates, but we make no particular assumption on these rates. This way, when two rules describe the same interaction but with different contexts of application (for instance rules (r12) and (r13)), it means that the kinetic of these interactions may be context-dependent.

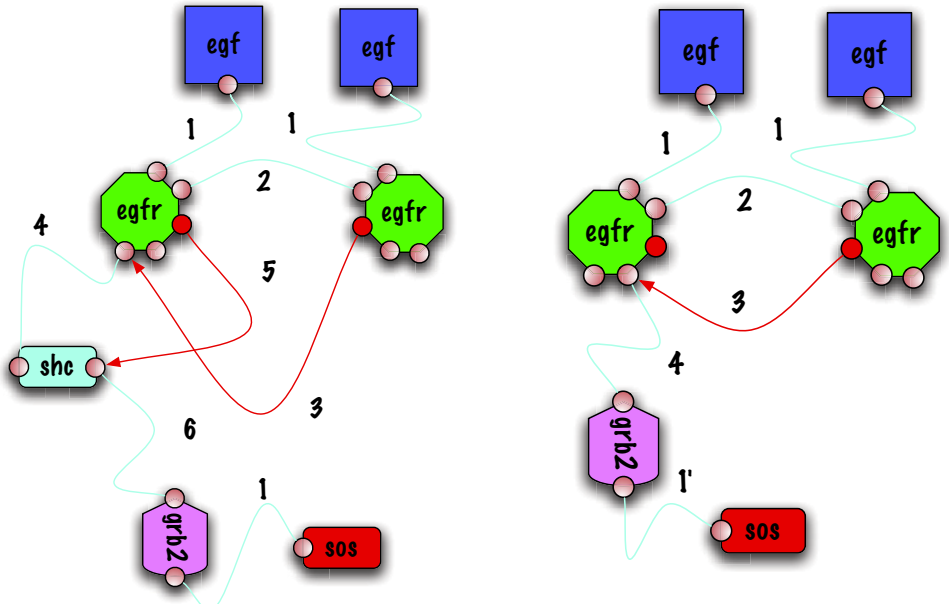
Firstly, we sketch the first scenario (for the recruitment of a protein *SOS* by a receptor *EGFR*). This scenario is depicted in Fig. 3(a). A receptor *EGFR* can recruit a ligand *EGF* (r01 — step 1), and bind another activated *EGFR* so as to form a dimer (r03 — step 2). Whenever a receptor *EGFR* is bound to another receptor, the site *Y48* can be phosphorilated (r05 — step 3). Then, *EGFR* can recruit an adapter molecule called *SHC* (r07, r08, r09, or r10 — step 4) (at a rate which depends on the state of *SHC*). Then, *EGFR* can phosphorilate *SHC* (r11 — step 5). *SHC* can then recruit a transport molecule *GRB2* (r14, r15, r16, or r17 — step 6).

Yet, each receptor has a shorter way to recruit a transport molecule. This second scenario is depicted in Fig. 3(b) and is sketched as follows. The site *Y68* of *EGFR* can be phosphorilated (r20 — step 3), and then recruit *GRB2* directly (r22, r23, or r24 — step 4). Last, the transport molecule *GRB2* can bind a protein *SOS* (r25 — step 1′) independently of the other interactions. Moreover, all interactions are reversible (sometimes in particular context).

We would like to track the number of proteins *SOS* which are bound (indirectly) to a receptor. One possibility would be to count the number of occurrences (in stochastic semantics) or the concentration (in differential semantics) of each chemical species, weighted by the number of proteins *SOS* which are bound to a receptor in this species. There are indeed 356 chemical species. Yet, one can observe that there are four sites in each dimer, which can recruit a protein *SOS*. Moreover, these sites do not operate any control over each others. Thus, intuitively, it should be possible to abstract the correlation between the states of these four sites among each dimer. Indeed, the fact that a given dimer has recruited several *SOS* is not

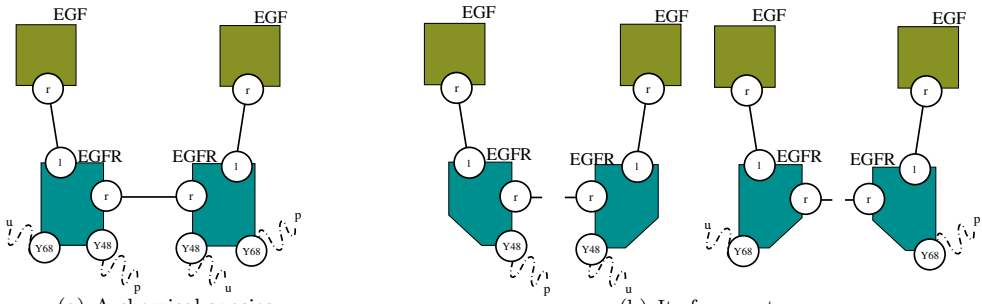
r01: $EGF(r)\ ,\ EGFR(l) \longrightarrow EGF(r^1)\ ,\ EGFR(l^1)$
r02: $EGFR(l^-,r) \longrightarrow EGFR(l,r)$
r03: $EGFR(l^-,r)\ ,\ EGFR(l^-,r) \longrightarrow EGFR(l^-,r^1)\ ,\ EGFR(l^-,r^1)$
r04: $EGFR(r^-) \longrightarrow EGFR(r)$
r05: $EGFR(Y48_u,r^-) \longrightarrow EGFR(Y48_p,r^-)$
r06: $EGFR(Y48_p) \longrightarrow EGFR(Y48_u)$
r07: $EGFR(Y48_p)\ ,\ SHC(Y7_u,pi) \longleftrightarrow EGFR(Y48^1_p)\ ,\ SHC(Y7_u,pi^1)$
r08: $EGFR(Y48_p)\ ,\ SHC(Y7_p,pi) \longleftrightarrow EGFR(Y48^1_p)\ ,\ SHC(Y7_p,pi^1)$
r09: $EGFR(Y48_p)\ ,\ GRB2(a^1,b)\ ,\ SHC(Y7^1,pi) \longleftrightarrow EGFR(Y48^2_p)\ ,\ GRB2(a^1,b)\ ,\ SHC(Y7^1,pi^2)$
r10: $EGFR(Y48_p)\ ,\ GRB2(a^1,b^-)\ ,\ SHC(Y7^1,pi) \longleftrightarrow EGFR(Y48^2_p)\ ,\ GRB2(a^1,b^-)\ ,\ SHC(Y7^1,pi^2)$
r11: $EGFR(Y48^1,r^-)\ ,\ SHC(Y7_u,pi^1) \longrightarrow EGFR(Y48^1,r^-)\ ,\ SHC(Y7_p,pi^1)$
r12: $SHC(Y7_p,pi^-) \longrightarrow SHC(Y7_u,pi^-)$
r13: $SHC(Y7_p,pi) \longrightarrow SHC(Y7_u,pi)$
r14: $GRB2(a,b)\ ,\ SHC(Y7_p,pi^-) \longleftrightarrow GRB2(a^1,b)\ ,\ SHC(Y7^1_p,pi^-)$
r15: $GRB2(a,b)\ ,\ SHC(Y7_p,pi) \longleftrightarrow GRB2(a^1,b)\ ,\ SHC(Y7^1_p,pi)$
r16: $GRB2(a,b^-)\ ,\ SHC(Y7_p,pi) \longleftrightarrow GRB2(a^1,b^-)\ ,\ SHC(Y7^1_p,pi)$
r17: $GRB2(a,b^-)\ ,\ SHC(Y7_p,pi^-) \longleftrightarrow GRB2(a^1,b^-)\ ,\ SHC(Y7^1_p,pi^-)$
r18: $GRB2(a^1,b)\ ,\ SHC(Y7^1,pi)\ ,\ SOS(d) \longleftrightarrow GRB2(a^2,b^1)\ ,\ SHC(Y7^2,pi)\ ,\ SOS(d^1)$
r19: $GRB2(a^1,b)\ ,\ SHC(Y7^1,pi^-)\ ,\ SOS(d) \longleftrightarrow GRB2(a^2,b^1)\ ,\ SHC(Y7^2,pi^-)\ ,\ SOS(d^1)$
r20: $EGFR(Y68_u,r^-) \longrightarrow EGFR(Y68_p,r^-)$
r21: $EGFR(Y68_p) \longrightarrow EGFR(Y68_u)$
r22: $EGFR(Y68_p)\ ,\ GRB2(a,b) \longleftrightarrow EGFR(Y68^1_p)\ ,\ GRB2(a^1,b)$
r23: $EGFR(Y68_p)\ ,\ GRB2(a,b^-) \longleftrightarrow EGFR(Y68^1_p)\ ,\ GRB2(a^1,b^-)$
r24: $EGFR(Y68^1)\ ,\ GRB2(a^1,b)\ ,\ SOS(d) \longleftrightarrow EGFR(Y68^2)\ ,\ GRB2(a^2,b^1)\ ,\ SOS(d^1)$
r25: $GRB2(a,b)\ ,\ SOS(d) \longleftrightarrow GRB2(a,b^1)\ ,\ SOS(d^1)$

Fig. 2. A model for the early EGF pathway in Kappa.



(a) Long arm. (b) Short arm.

Fig. 3. Stories for the early EGF model.



(a) A chemical species. (b) Its fragments.

Fig. 4. A chemical species and its abstraction as a (multi) set of fragments.

important, if we only care about the overall number of *SOS* that are recruited by receptors. This abstraction comes down to consider a dimer as four *independent* parts (or fragments). Thus, instead of counting the number of occurrences of each species, we count the number of fragments of species obtained by cutting dimers into four parts.

We have shown in [20,16] that this abstraction is sound and efficient in the differential semantics. Indeed, the concentration of the proteins *SOS* can be obtained from a differential system of 38 fragments. The set of fragments and the differential system can be computed automatically from the set of rules, and without ever explicitly generating neither the set of species, nor the set of reactions. For that purpose, we consider an annotated contact map (eg. see Fig. 1(b)) which is computed automatically from the set of rules. This annotated contact map contains the directives to explain how to cut chemical species into fragments. In this map, each agent is fitted with a covering of its set of sites, moreover some edges are dotted. In Fig. 4, we illustrate through an example how a chemical soup (Fig. 4(a)) can be decomposed as a (multi) set of fragments (Fig. 4(b)). Intuitively, each solid edge in the annotated contact map is preserved, whereas each dotted edge is cut. In the latter case, we keep only the information that the site is bound (as for the site r of receptors for instance). Then for each protein, we only keep a set of sites which matches with a covering class in the annotated contact map (this way, for each receptor, we keep either the set of sites $\{l, r, \mathit{Y48}\}$ or the set of sites $\{l, r, \mathit{Y68}\}$), and we consider any such combination. We notice that since we use coverings instead of partitions, fragments may overlap. This property is essential so as to achieve an efficient reduction.

Since the sites *Y48* and *Y68* do not belong to a same covering class and because the edge between the site r in protein *EGFR* and itself is dotted in the annotated contact map, the correlation between what is attached to the four phosphorilable sites in each dimer is abstracted away. This leads to a good reduction factor. Indeed the number of variables for the dimers in the initial semantics is of the form $mn(mn+1)/2$, whereas the number of fragments for the dimers in the reduced semantics is of the form $m+n$ (where m stands for the number of fragments which contains a receptor with the site r bound and which documents the site *Y48*; and n stands for the number of fragments which contains a receptor with the site r bound

and which documents the site *Y68*).[3] At last, we can notice that this reduction is possible although the short and the long arms share some common resources. They both uses the protein *GRB2* for instances. This illustrates the fact that concurrency is not an issue for our reduction framework: we can reduce the semantics of some systems without identifying fully independent modules.

Nevertheless, the same reduction cannot be achieved in the case of the stochastic semantics, because the semi-reactions that operate over the fragment of receptors are coupled (eg see Sect. 4): For instance, when a bond between two receptors is released, in the stochastic semantics, two receptors are modified. It follows that it is impossible to abstract the state of the system by a multi-set of fragments, while preserving the density distribution of traces.

3 An example of two independent modules

In this section, we consider a model which can be split into two independent subsystems (or modules) and we show that this property can be used so as to reduce both the stochastic and the differential semantics of this model.

Consider a protein B with two independent sites: one to bind protein A and one to bind protein C. Association/dissociation between A and B, and between B and C, are described by the following reactions:

$$A + B \underset{k_d^{AB}}{\overset{k^{AB}}{\rightleftharpoons}} AB \qquad B + C \underset{k_d^{BC}}{\overset{k^{BC}}{\rightleftharpoons}} BC$$

$$A + BC \underset{k_d^{AB}}{\overset{k^{AB}}{\rightleftharpoons}} ABC \qquad AB + C \underset{k_d^{BC}}{\overset{k^{BC}}{\rightleftharpoons}} ABC$$

Importantly, we have assumed that the association and dissociation rates, k^{AB} and k_d^{AB}, are the same whether or not B is bound to C; and similarly for the association/dissociation rates k^{BC}, k_d^{BC} of B and C.

The differential semantics of this model is defined by the following system of differential equations:

$$\begin{aligned}
[A]' &= k_d^{AB}\left([AB] + [ABC]\right) - k^{AB}\left([B] + [BC]\right)[A] \\
[C]' &= k_d^{BC}\left([BC] + [ABC]\right) - k^{BC}\left([B] + [AB]\right)[C] \\
[B]' &= k_d^{AB}[AB] + k_d^{BC}[BC] - \left(k^{AB}[A] + k^{BC}[C]\right)[B] \\
[AB]' &= k^{AB}[A][B] + k_d^{BC}[ABC] - \left(k_d^{AB} + k^{BC}[C]\right)[AB] \\
[BC]' &= k_d^{AB}[ABC] + k^{BC}[B][C] - \left(k_d^{BC} + k^{AB}[A]\right)[BC] \\
[ABC]' &= k^{AB}[A][BC] + k^{BC}[AB][C] - \left(k_d^{AB} + k_d^{BC}\right)[ABC].
\end{aligned}$$

[3] In forthcoming examples, there may seem to be no reduction. This is due to the small size of the examples. In these examples, the number of variables is of the form n^2, whereas the number of fragments is of the form $2 \times n$, which gives no reduction for $n = 2$.

We notice that we can abstract away the correlation between the state of the two binding sites of each protein B. Indeed, we can check that the following equations:

$$[A]' = k_d^{AB}[AB?] - k^{AB}[A][B?] \qquad [C]' = k_d^{BC}[?BC] - k^{BC}[?B][C]$$
$$[B?]' = k_d^{AB}[AB?] - k^{AB}[A][B?] \qquad [?B]' = k_d^{BC}[?BC] - k^{BC}[?B][C]$$
$$[AB?]' = k^{AB}[A][B?] - k_d^{AB}[AB?] \qquad [?BC]' = k^{BC}[?B][C] - k_d^{BC}[?BC]$$

where[4] $[AB?] \triangleq [AB] + [ABC]$, $[B?] \triangleq [B] + [BC]$, $[?BC] \triangleq [BC] + [ABC]$, and $[?B] \triangleq [B] + [AB]$, are satisfied.

Note that, in this particular case, the states of the two binding sites of each protein B are independent, which can be checked analytically. Let us introduce $X \triangleq [ABC][?B?] - [AB?][?BC]$. The expression X measures the degree of independence between B's two binding sites. It turns out that: $X' = -X(k^{AB}[A] + k^{BC}[C] + k_d^{AB} + k_d^{BC})$; and, as a consequence, the property $X = 0$ is an invariant of the system. It follows that, provided that the two binding states are independent at time $t = 0$, one can express the concentration of any species at a given time t by using only the concentrations of the part of chemical species A, C, $B?$, $AB?$, $?B$, and $?BC$ at time t.

Now we focus on the stochastic semantics, in particular we study the state occupancy distribution of our model. In this semantics, a chemical soup is denoted by a 6-tuple $\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle$ of natural numbers (in $\mathbb{N}$), where n_X is the number of instances of X in the chemical soup, for any $X \in \{A, B, C, AB, BC, ABC\}$. The probability $P_t(\sigma)$ that the system is in a given state σ at time t is given by the following master equation:

$$\begin{aligned}
&P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)' = \\
&\quad k^{AB}(n_A+1)(n_B+1)P_t(\langle n_A{+}1, n_B{+}1, n_C, n_{AB}{-}1, n_{BC}, n_{ABC}\rangle) \\
&+\ k_d^{AB}(n_{AB}+1)P_t(\langle n_A{-}1, n_B{-}1, n_C, n_{AB}{+}1, n_{BC}, n_{ABC}\rangle) \\
&+\ k^{AB}(n_A+1)(n_{BC}+1)P_t(\langle n_A{+}1, n_B, n_C, n_{AB}, n_{BC}{+}1, n_{ABC}{-}1\rangle) \\
&+\ k_d^{AB}(n_{ABC}+1)P_t(\langle n_A{-}1, n_B, n_C, n_{AB}, n_{BC}{-}1, n_{ABC}{+}1\rangle) \\
&+\ k^{BC}(n_B+1)(n_C+1)P_t(\langle n_A, n_B{+}1, n_C{+}1, n_{AB}, n_{BC}{-}1, n_{ABC}\rangle) \\
&+\ k_d^{BC}(n_{BC}+1)P_t(\langle n_A, n_B{-}1, n_C{-}1, n_{AB}, n_{BC}{+}1, n_{ABC}\rangle) \\
&+\ k^{BC}(n_{AB}+1)(n_C+1)P_t(\langle n_A, n_B, n_C{+}1, n_{AB}{+}1, n_{BC}, n_{ABC}{-}1\rangle) \\
&+\ k_d^{BC}(n_{ABC}+1)P_t(\langle n_A, n_B, n_C{-}1, n_{AB}{-}1, n_{BC}, n_{ABC}{+}1\rangle) \\
&-\ k^{AB}n_A(n_B+n_{BC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)
\end{aligned}$$

[4] A question mark '?' on the left (resp. right) of B stands for "whatever the protein B is bound to a protein A (resp. C), or not".

$$- k_d^{AB}(n_{AB} + n_{ABC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)$$
$$- k^{BC}(n_B + n_{AB})n_C P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)$$
$$- k_d^{BC}(n_{BC} + n_{ABC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)$$

In this particular example, we can abstract away the correlation between the state of the two binding sites of each protein B: Given a state $\sigma = \langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle$, we denote by $\beta^A(\sigma)$ the triple $\langle n_A, n_B + n_{BC}, n_{AB} + n_{ABC}\rangle$, and by $\beta^C(\sigma)$ the triple $\langle n_C, n_B + n_{AB}, n_{BC} + n_{ABC}\rangle$. The probability $P_t^A(\sigma^A)$ that the system is in a state σ such that $\beta^A(\sigma) = \sigma^A$ at time t, and the probability $P_t^C(\sigma^C)$ that the system is in a state σ such that $\beta^C(\sigma) = \sigma^C$ at time t, satisfy the following master equations:

$$\begin{aligned}
&P_t^A(\langle n_A, n_{B?}, n_{AB?}\rangle)' = \\
&\quad k^{AB}(n_A + 1)(n_{B?} + 1)P_t^A(\langle n_A{+}1, n_{B?}{+}1, n_{AB?}{-}1\rangle) \\
&+ k_d^{AB}(n_{AB?} + 1)P_t^A(\langle n_A{-}1, n_{B?}{-}1, n_{AB?}{+}1\rangle) \\
&- (k^{AB}n_A n_{B?} + k_d^{AB}n_{AB?})P_t^A(\langle n_A, n_{B?}, n_{AB?}\rangle)
\end{aligned}$$

$$\begin{aligned}
&P_t^C(\langle n_C, n_{?B}, n_{?BC}\rangle)' = \\
&\quad k^{BC}(n_{?B} + 1)(n_C + 1)P_t^C(\langle n_C{+}1, n_{?B}{+}1, n_{?BC}{-}1\rangle) \\
&+ k_d^{BC}(n_{?BC} + 1)P_t^C(\langle n_C{-}1, n_{?B}{-}1, n_{?BC}{+}1\rangle) \\
&- (k^{BC}n_{?B}n_C + k_d^{AB}n_{?BC})P_t^C(\langle n_C, n_{?B}, n_{?BC}\rangle).
\end{aligned}$$

Statistical independence can be revisited in the context of stochastic semantics as follows. We say that the two binding states of the protein B are statistically independent, if and only if, $P_t(\sigma_1) = P_t(\sigma_2)$, for any states σ_1 and σ_2 such that $\beta^A(\sigma_1) = \beta^A(\sigma_2)$ and $\beta^C(\sigma_1) = \beta^C(\sigma_2)$. We can check analytically that in this particular example, the two binding states of the protein B are statistically independent.

4 An example of coupled semi-reactions

It hardly ever happens that two modules are fully independent in a model (as it was the case in the example in the previous section). In this section, we show an example with coupled semi-reactions, that is a reaction that operates of two fragments of a species simultaneously. We will see that such coupled semi-reactions do not prevent the reduction of the differential semantics, but they forbid the reduction of the stochastic semantics.

We consider two kinds of proteins, A and B. Each protein can be unphosphorilated, or phosphorilated. Moreover, a protein A and a protein B may form a complex

AB. We use the symbol $\star$ as a superscript when a protein is phosphorilated. This way, a fully phosphorilated complex is denoted by $A^\star B^\star$.

The behavior of a chemical soup can be described by the following set of reactions:

$$A \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} A^\star \qquad B \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} B^\star \qquad A + B \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} AB$$

$$AB \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} A^\star B \qquad AB \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} AB^\star \qquad A^\star + B \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} A^\star B$$

$$AB^\star \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} A^\star B^\star \qquad A^\star B \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} A^\star B^\star \qquad A + B^\star \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} AB^\star$$

$$A^\star + B^\star \underset{k_{A..B}}{\overset{k_{AB\star}}{\rightleftharpoons}} A^\star B^\star$$

We have assumed that the likelihood that two proteins form a complex may be different when both proteins are phosphorilated (if we take $k_{AB} \neq k_{AB\star}$) (see third column, direct way). We have also assumed that all the other reactions are purely local. That is to say that the kinetic of phosphorilation and dephosphorilation of both the protein A (see first column) and the protein B (see second column) depends neither on the fact that the protein is in a complex, or not, nor (if it is in a complex) on the phosphorilation state of the other protein in the complex. Moreover, the kinetic of complex dissociation does not depend on the phosphorilation state of the two proteins in a given complex (see third column, converse way).

In this example, we would like to abstract the correlation between the phosphorilation state of proteins in complex. This could be achieved, by splitting each complex into two parts, and by abstracting away which parts are connected together. With such an abstraction, a dissociation reaction can be seen as two semi-reactions: one to unbind a (bound) protein A (phosphorilated, or not) and one to unbind a (bound) protein B (phosphorilated, or not). These two semi-reactions are coupled. Indeed the choice of the phosphorilation state of the protein A and the choice of the phosphorilation state of the protein B are entangled by the correlation between these two phosphorilation states. We will see in this section that it raises no issue in the abstraction of the differential semantics, but that this correlation cannot be abstracted away in the stochastic semantics.

The differential semantics of this model is defined by the following system of differential equations:

$$\begin{aligned}
[A]' &= k_{A-}[A^\star] + k_{A..B}([AB] + [AB^\star]) - (k_{A+} + k_{AB}([B] + [B^\star]))[A] \\
[A^\star]' &= k_{A+}[A] + k_{A..B}([A^\star B] + [A^\star B^\star]) - (k_{A-} + k_{AB}[B] + k_{AB\star}[B^\star])[A^\star] \\
[B]' &= k_{B-}[B^\star] + k_{A..B}([AB] + [A^\star B]) - (k_{B+} + k_{AB}([A] + [A^\star]))[B] \\
[B^\star]' &= k_{B+}[B] + k_{A..B}([AB^\star] + [A^\star B^\star]) - (k_{B-} + k_{AB}[A] + k_{AB\star}[A^\star])[B^\star]
\end{aligned}$$

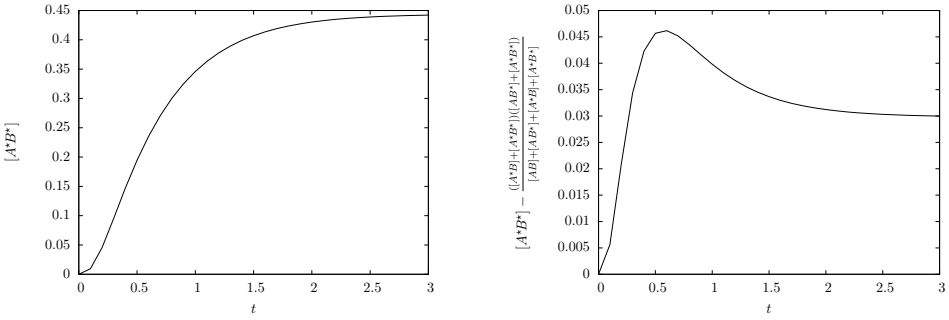


Fig. 5. On the left, the evolution of the concentration of the protein $A^\star B^\star$. On the right, the evolution of the difference between the concentration of the protein $A^\star B^\star$ and the expression $([A^\star B] + [A^\star B^\star])([AB^\star] + [A^\star B^\star])/([AB] + [A^\star B] + [AB^\star] + [A^\star B^\star])$. All rates are set to 1, except $k_{AB\star}$ which is set to 10. At time 0, the concentration of A is set to 2, the concentration of B as well, and the concentration of any other chemical species is set to 0.

$$
\begin{aligned}
[AB]' &= k_{A-}[A^\star B] + k_{B-}[AB^\star] + k_{AB}[A][B] - (k_{A+} + k_{B+} + k_{A..B})[AB] \\
[A^\star B]' &= k_{A+}[AB] + k_{B-}[A^\star B^\star] + k_{AB}[A^\star][B] - (k_{A-} + k_{B+} + k_{A..B})[A^\star B] \\
[AB^\star]' &= k_{A-}[A^\star B^\star] + k_{B+}[AB] + k_{AB}[A][B^\star] - (k_{A+} + k_{B-} + k_{A..B})[AB^\star] \\
[A^\star B^\star]' &= k_{A+}[AB^\star] + k_{B+}[A^\star B] + k_{AB\star}[A^\star][B^\star] - (k_{A-} + k_{B-} + k_{A..B})[A^\star B^\star].
\end{aligned}
$$

We can also notice, that unlike the example in Sect. 3, the phosphorilation state of the two proteins in complex is in general correlated. We show in Fig. 5(a) an example of trajectory for the concentration of the protein $A^\star B^\star$ along the time, and in Fig. 5(b) the difference between this concentration and the value of the following expression:

$$\frac{([AB^\star] + [A^\star B^\star])([A^\star B] + [A^\star B^\star])}{[AB] + [AB^\star] + [A^\star B] + [A^\star B^\star]}.$$

Nevertheless, even if the phosphorilation state of proteins in complex is correlated, this correlation can be abstracted away. Indeed, we can notice that the following equations:

$$
\begin{aligned}
[A]' &= k_{A-}[A^\star] + k_{A..B}[AB^\diamond] - (k_{A+} + k_{AB}([B] + [B^\star]))[A] \\
[A^\star]' &= k_{A+}[A] + k_{A..B}[A^\star B^\diamond] - (k_{A-} + k_{AB}[B] + k_{AB\star}[B^\star])[A^\star] \\
[B]' &= k_{B-}[B^\star] + k_{A..B}[A^\diamond B] - (k_{B+} + k_{AB}([A] + [A^\star]))[B] \\
[B^\star]' &= k_{B+}[B] + k_{A..B}[A^\diamond B^\star] - (k_{B-} + k_{AB}[A] + k_{AB\star}[A^\star])[B^\star] \\
[AB^\diamond]' &= k_{A-}[A^\star B^\diamond] + k_{AB}[A]([B] + [B^\star]) - (k_{A+} + k_{A..B})[AB^\diamond] \\
[A^\star B^\diamond]' &= k_{A+}[AB^\diamond] + k_{AB}[A^\star][B] + k_{AB\star}[A^\star][B^\star] - (k_{A-} + k_{A..B})[A^\star B^\diamond] \\
[A^\diamond B]' &= k_{B-}[A^\diamond B^\star] + k_{AB}[B]([A] + [A^\star]) - (k_{B+} + k_{A..B})[A^\diamond B] \\
[A^\diamond B^\star]' &= k_{B+}[A^\diamond B] + k_{AB}[A][B^\star] + k_{AB\star}[A^\star][B^\star] - (k_{B-} + k_{A..B})[A^\diamond B^\star],
\end{aligned}
$$

where [5] $[AB^\diamond] \triangleq [AB]+[AB^\star]$, $[A^\star B^\diamond] \triangleq [A^\star B]+[A^\star B^\star]$, $[A^\diamond B] \triangleq [AB]+[A^\star B]$, and $[A^\diamond B^\star] \triangleq [AB^\star]+[A^\star B^\star]$, are satisfied.

Yet, this correlation forbids the reduction of the stochastic semantics. Let us explain why. In the stochastic semantics, a chemical soup can be denoted by a 8-tuple $\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle$ of natural numbers, where n_X is the number of instance of X in the chemical soup, for any $X \in \{A, A^\star, B, B^\star, AB, A^\star B, AB^\star, A^\star B^\star\}$. The probability $P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)$ that the system is in a given state $\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle$ at time t is given by the following master equation:

$$
\begin{aligned}
&P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)' = \\
&\quad k_{A+}(n_A+1)P_t(\langle n_A{+}1, n_{A^\star}{-}1, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A+}(n_{AB}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{+}1, n_{A^\star B}{-}1, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A+}(n_{AB^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}{+}1, n_{A^\star B^\star}{-}1\rangle)\\
&+\ k_{A-}(n_{A^\star}+1)P_t(\langle n_A{-}1, n_{A^\star}{+}1, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A-}(n_{A^\star B}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{-}1, n_{A^\star B}{+}1, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A-}(n_{A^\star B^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}{-}1, n_{A^\star B^\star}{+}1\rangle)\\
&+\ k_{B+}(n_B+1)P_t(\langle n_A, n_{A^\star}, n_B{+}1, n_{B^\star}{-}1, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{B+}(n_{AB}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{+}1, n_{A^\star B}, n_{AB^\star}{-}1, n_{A^\star B^\star}\rangle)\\
&+\ k_{B+}(n_{A^\star B}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}{+}1, n_{AB^\star}, n_{A^\star B^\star}{-}1\rangle)\\
&+\ k_{B-}(n_{B^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B{-}1, n_{B^\star}{+}1, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{B-}(n_{AB^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{-}1, n_{A^\star B}, n_{AB^\star}{+}1, n_{A^\star B^\star}\rangle)\\
&+\ k_{B-}(n_{A^\star B^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}{-}1, n_{AB^\star}, n_{A^\star B^\star}{+}1\rangle)\\
&+\ k_{AB}(n_A+1)(n_B+1)P_t(\langle n_A{+}1, n_{A^\star}, n_B{+}1, n_{B^\star}, n_{AB}{-}1, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{AB}(n_A+1)(n_{B^\star}+1)P_t(\langle n_A{+}1, n_{A^\star}, n_B, n_{B^\star}{+}1, n_{AB}, n_{A^\star B}, n_{AB^\star}{-}1, n_{A^\star B^\star}\rangle)\\
&+\ k_{AB}((n_{A^\star}+1)(n_B+1))P_t(\langle n_A, n_{A^\star}{+}1, n_B{+}1, n_{B^\star}, n_{AB}, n_{A^\star B}{-}1, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{AB\star}((n_{A^\star}+1)(n_{B^\star}+1))P_t(\langle n_A, n_{A^\star}{+}1, n_B, n_{B^\star}{+}1, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}{-}1\rangle)\\
&+\ k_{A..B}(n_{AB}-1)P_t(\langle n_A{-}1, n_{A^\star}{-}1, n_B, n_{B^\star}, n_{AB}{+}1, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A..B}(n_{AB^\star}-1)P_t(\langle n_A{-}1, n_{A^\star}, n_B, n_{B^\star}{-}1, n_{AB}, n_{A^\star B}, n_{AB^\star}{+}1, n_{A^\star B^\star}\rangle)\\
&+\ k_{A..B}(n_{A^\star B}-1)P_t(\langle n_A, n_{A^\star}{-}1, n_B{-}1, n_{B^\star}, n_{AB}, n_{A^\star B}{+}1, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&+\ k_{A..B}(n_{A^\star B^\star}-1)P_t(\langle n_A, n_{A^\star}{-}1, n_B, n_{B^\star}{-}1, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}{+}1\rangle)\\
&-\ k_{A+}(n_A+n_{AB}+n_{AB^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{A-}(n_{A^\star}+n_{A^\star B}+n_{A^\star B^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{B+}(n_B+n_{AB}+n_{A^\star B})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{B-}(n_{B^\star}+n_{AB^\star}+n_{A^\star B^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{AB}((n_A+n_{A^\star})(n_B+n_{B^\star})-n_{A^\star}n_{B^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{AB\star}n_{A^\star}n_{B^\star}P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)\\
&-\ k_{A..B}(n_{AB}+n_{AB^\star}+n_{A^\star B}+n_{A^\star B^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle)
\end{aligned}
$$

As in the example of Sect. 3, we would like to abstract away the correlation between the phosphorilation state of the proteins A and the phosphorilation state of the proteins B which belong to the same complex. Given a state $\sigma = \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}\rangle$, we denote by $\beta(\sigma)$ the 8-tuple $\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle$ (such a tuple is called an abstract state). The probability $P_t^\sharp(\sigma^\sharp)$ that the system is in a state σ such that $\beta(\sigma) = \sigma^\sharp$

[5] The superscript $\diamond$ stands for "whatever the phosphorilation state is".

at time t, satisfies the following equation:

$$
\begin{aligned}
&P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)' = \\
&\quad k_{A+}(n_A+1)P_t^\sharp(\langle n_A{+}1, n_{A^\star}{-}1, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{A+}(n_{AB^\diamond}+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}{+}1, n_{A^\star B^\diamond}{-}1, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{A-}(n_{A^\star}+1)P_t^\sharp(\langle n_A{-}1, n_{A^\star}{+}1, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{A-}(n_{A^\star B^\diamond}+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}{-}1, n_{A^\star B^\diamond}{+}1, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{B+}(n_B+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B{+}1, n_{B^\star}{-}1, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{B+}(n_{A^\diamond B}+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}{+}1, n_{A^\diamond B^\star}{-}1\rangle)\\
&+\ k_{B-}(n_{B^\star}+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B{-}1, n_{B^\star}{+}1, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{B-}(n_{A^\diamond B^\star}+1)P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}{-}1, n_{A^\diamond B^\star}{+}1\rangle)\\
&+\ k_{AB}(n_A+1)(n_B+1)P_t^\sharp(\langle n_A{+}1, n_{A^\star}, n_B{+}1, n_{B^\star}, n_{AB^\diamond}{-}1, n_{A^\star B^\diamond}, n_{A^\diamond B}{-}1, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{AB}(n_A+1)(n_{B^\star}+1)P_t^\sharp(\langle n_A{+}1, n_{A^\star}, n_B, n_{B^\star}{+}1, n_{AB^\diamond}{-}1, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}{-}1\rangle)\\
&+\ k_{AB}((n_{A^\star}+1)(n_B+1))P_t^\sharp(\langle n_A, n_{A^\star}{+}1, n_B{+}1, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}{-}1, n_{A^\diamond B}{-}1, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{AB\star}((n_{A^\star}+1)(n_{B^\star}+1))P_t^\sharp(\langle n_A, n_{A^\star}{+}1, n_B, n_{B^\star}{+}1, n_{AB^\diamond}, n_{A^\star B^\diamond}{-}1, n_{A^\diamond B}, n_{A^\diamond B^\star}{-}1\rangle)\\
&+\ k_{A..B}\tilde{E}_t(n_{AB} \mid \langle n_A{-}1, n_{A^\star}, n_B{-}1, n_{B^\star}, n_{AB^\diamond}{+}1, n_{A^\star B^\diamond}, n_{A^\diamond B}{+}1, n_{A^\diamond B^\star}\rangle)\\
&+\ k_{A..B}\tilde{E}_t(n_{AB^\star} \mid \langle n_A{-}1, n_{A^\star}, n_B, n_{B^\star}{-}1, n_{AB}{+}1, n_{A^\star B}, n_{AB^\star}, n_{A^\star B^\star}{+}1\rangle)\\
&+\ k_{A..B}\tilde{E}_t(n_{A^\star B} \mid \langle n_A, n_{A^\star}{-}1, n_B{-}1, n_{B^\star}, n_{AB}, n_{A^\star B}{+}1, n_{AB^\star}{+}1, n_{A^\star B^\star}\rangle)\\
&+\ k_{A..B}\tilde{E}_t(n_{A^\star B^\star} \mid \langle n_A, n_{A^\star}{-}1, n_B, n_{B^\star}{-}1, n_{AB}, n_{A^\star B}{+}1, n_{AB^\star}, n_{A^\star B^\star}{+}1\rangle)\\
&-\ k_{A+}(n_A+n_{AB^\diamond})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{A-}(n_{A^\star}+n_{A^\star B^\diamond})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{B+}(n_B+n_{A^\diamond B})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{B-}(n_{B^\star}+n_{A^\diamond B^\star})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{AB}((n_A+n_{A^\star})(n_B+n_{B^\star})-n_{A^\star}n_{B^\star})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{AB\star}(n_{A^\star}n_{B^\star})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle)\\
&-\ k_{A..B}(n_{A^\star B^\diamond}+n_{AB^\diamond})P_t^\sharp(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle),
\end{aligned}
$$

where for any expression $X(\sigma)$ and any (abstract) state $\sigma^\sharp$, the expression $\tilde{E}_t(X(\sigma) \mid \sigma^\sharp)$ denotes the product between the conditional expectation $E_t(X(\sigma) \mid \sigma^\sharp)$ of the expression $X(\sigma)$ knowing that $\beta(\sigma) = \sigma^\sharp$, and the probability $P_t^\sharp(\sigma^\sharp)$ of being in a state σ such that $\beta(\sigma) = \sigma^\sharp$.

Whenever $k_{AB} = k_{AB\star}$, we can check that the fact that $P_t(\sigma) = P_t(\sigma')$ for any pair of states σ,σ' such that $\beta(\sigma) = \beta(\sigma')$ is an invariant. Thus, provided that $k_{AB} = k_{AB\star}$ and that there is no correlation between the phosphorilation state of the proteins A and B which are bound together at time $t = 0$, one can use the following properties:

$$
\begin{aligned}
E_t(n_{AB} \mid \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle) &= \frac{n_{AB^\diamond} n_{A^\diamond B}}{AB^\diamond + A^\star B^\diamond}\\
E_t(n_{AB^\star} \mid \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle) &= \frac{n_{AB^\diamond} n_{A^\diamond B^\star}}{AB^\diamond + A^\star B^\diamond}\\
E_t(n_{A^\star B} \mid \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle) &= \frac{n_{A^\star B^\diamond} n_{A^\diamond B}}{AB^\diamond + A^\star B^\diamond}\\
E_t(n_{A^\star B^\star} \mid \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB^\diamond}, n_{A^\star B^\diamond}, n_{A^\diamond B}, n_{A^\diamond B^\star}\rangle) &= \frac{n_{A^\star B^\diamond} n_{A^\diamond B^\star}}{AB^\diamond + A^\star B^\diamond},
\end{aligned}
$$

so as to write conditional expectations of n_{AB}, $n_{AB^\star}$, $n_{A^\star B}$, and $n_{A^\star B^\star}$ as time-independent expressions of $n_{AB^\diamond}$, $n_{A^\star B^\diamond}$, $n_{A^\diamond B}$, and $n_{A^\diamond B^\star}$.

Whenever $k_{AB} \neq k_{AB\star}$, these conditional expectations may be time-dependent. We show in Fig. 6(a) that the ratio between the probability of being in the state

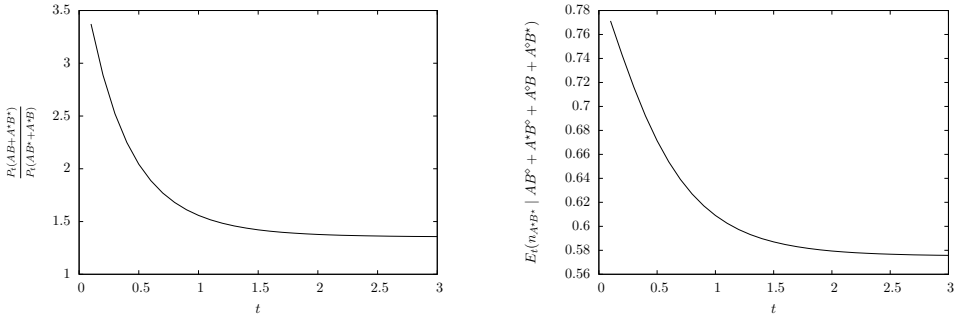


Fig. 6. On the left, quotient between the probability of being in the state $AB + A^\star B^\star$ and the probability of being in the state $AB^\star + A^\star B$. On the right, conditional expectation of the number of fully phosphorilated complexes $A^\star B^\star$ knowing that all proteins are bound, and that there is exactly one phosphorilated protein A and exactly one phosphorilated B. All rates are set to 1, except $k_{AB\star}$ which is set to 10. At time 0, the chemical soup is made of two proteins A and two proteins B, none of these proteins being phosphorilated or bound.

$AB + A^\star B^\star$ and the probability of being in the state $AB^\star + A^\star B$ is time-dependent. Moreover, we show in Fig. 6(b) that the conditional expectation of $n_{A^\star B^\star}$ knowing that we are in the (abstract) state $AB^\diamond + A^\star B^\diamond + A^\diamond B + A^\diamond B^\star$ is time-dependent as well, which forbids doing the same simplification as in the differential semantics.

We have seen through this example that some reactions may operate simultaneously over two fragments. This leads to coupled semi-reactions. We have noticed that coupled semi-reactions raise no issue when reducing the differential semantics. We say that the application of semi-rules is fair in the differential semantics, since the proportion of the concentration of a given fragment that is consumed by a semi-reaction does not depend on the correlation between the states of the two fragments. This is not the case in the stochastic semantics: we have noticed that the stochastic semantics can be reduced only if the state of the two fragments are not correlated, otherwise the choice of the fragments on which coupled semi-reactions operate is entangled, which forbids the reduction. In other words, we say that in the differential semantics, we can abstract away the correlations which are not observed by rules, whereas in the stochastic semantics, we have to prove that the rules cannot enforce correlations between the state of some fragments and we use this property so as to reduce the dimension of the state space of the system. In the later case, the reduction is only valid when there is no correlation at time $t = 0$.

5 An example of distant control

In this section, we show that, in the stochastic semantics, one protein can control the behavior of another one even if they are not in the same connected component in the left hand side of a reaction.

We consider a kind of proteins, A, which bears two phosphorilation sites. Each phosphorilation site can be unphosphorilated, or phosphorilated. We use the symbol $\star$ to denote phosphorilated sites. The phosphorilation state of the first site is written as a superscript, whereas the phosphorilation state of the second site is written as a

subscript. This way, a protein A having the first site phosphorilated and the second site unphosphorilated is denoted by $A^\star$.

The behavior of a chemical soup can be described by the following set of reactions:

$$\begin{array}{lll} & A + A^\star \xrightarrow{k_+} A_\star + A^\star & \\ A \underset{k^-}{\overset{k^+}{\rightleftharpoons}} A^\star & A^\star + A^\star \xrightarrow{k_+} A^\star_\star + A^\star & A^\star_\star \xrightarrow{k_-} A^\star \\ A_\star \underset{k^-}{\overset{k^+}{\rightleftharpoons}} A^\star_\star & A + A^\star_\star \xrightarrow{k_+} A_\star + A^\star_\star & A_\star \xrightarrow{k_-} A \\ & A^\star + A^\star_\star \xrightarrow{k_+} A^\star_\star + A^\star_\star & \end{array}$$

We have assumed (see second column) that the kinetic of the phosphorilation of the second site of a protein depends on the number of the other proteins that are phosphorilated on their first site — that is to say that the proteins that are phosphorilated on their first site catalyzes the phosphorilation of the second site in the other proteins. We have also assumed that other reactions are purely local, that is to say that the kinetic of phosphorilation and dephosphorilation on the first site does not depend on the phosphorilation state of the second site (neither of the protein being phosphorilated, nor of the other proteins) (see first column), and that the kinetic of dephosphorilation of the second site does not depend on the phosphorilation state of the first site of the proteins in the soup (see third column).

In this example, we would like to abstract the correlation between the phosphorilation state of the two sites of each protein. This could be achieved, by splitting each complex into two parts, and by abstracting away which parts are connected together. It raises an issue for reducing the stochastic semantics. Indeed, one can notice that the reaction which activates the second site of protein favors the phosphorilation of the second site of the protein in the state A. For instance, if we assume that both the number of instances of the protein in state A and the number of instances of the protein in the state $A^\star$ is equal to m, and that the number of instances of the protein in the state $A^\star_\star$ is equal to n. Then, the cumulative activity of the following two reactions:

$$A + A^\star \xrightarrow{k_+} A_\star + A^\star \qquad A + A^\star_\star \xrightarrow{k_+} A_\star + A^\star_\star$$

is equal to $n(n+m)$, whereas the cumulative activity of the following two reactions:

$$A^\star + A^\star \xrightarrow{k_+} A^\star_\star + A^\star \qquad A^\star + A^\star_\star \xrightarrow{k_+} A^\star_\star + A^\star_\star$$

is equal to $n(n+m-1)$ (the subtraction by 1 comes from the fact that each reactant must be mapped to distinct instances of chemical species). Nevertheless, it does not forbid the reduction of the differential semantics: intuitively, the term 1 vanishes because we consider an infinite number of instances, within an infinite volume.

Let us check formally that the differential semantics of this model can be reduced and explain why we do not know how to abstract its stochastic semantics. This

differential semantics is defined by the following system of differential equations:

$$
\begin{aligned}
[\mathrm{A}]' &= k^-[\mathrm{A}^\star] + k_-[\mathrm{A}_\star] - (k^+ + k_+([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]))[\mathrm{A}] \\
[\mathrm{A}^\star]' &= k^+[\mathrm{A}] + k_-[\mathrm{A}^\star_\star] - (k^- + k_+([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]))[\mathrm{A}^\star] \\
[\mathrm{A}_\star]' &= k^-[\mathrm{A}^\star_\star] + k_+[\mathrm{A}]([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]) - (k^+ + k_-)[\mathrm{A}_\star] \\
[\mathrm{A}^\star_\star]' &= k^+[\mathrm{A}_\star] + k_+[\mathrm{A}^\star]([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]) - (k^- + k_-)[\mathrm{A}^\star_\star].
\end{aligned}
$$

We notice that the correlation between the two sites can be abstracted away. Indeed, we notice that the following equations:

$$
\begin{aligned}
[\mathrm{A}_\diamond]' &= k^-[\mathrm{A}^\star_\diamond] - k^+[\mathrm{A}_\diamond] \\
[\mathrm{A}^\star_\diamond]' &= k^+[\mathrm{A}_\diamond] - k^-[\mathrm{A}^\star_\diamond] \\
[\mathrm{A}^\diamond]' &= k_-[\mathrm{A}^\diamond_\star] - k_+[\mathrm{A}^\diamond][\mathrm{A}^\star_\diamond] \\
[\mathrm{A}^\diamond_\star]' &= k_+[\mathrm{A}^\diamond][\mathrm{A}^\star_\diamond] - k_-[\mathrm{A}^\diamond_\star],
\end{aligned}
$$

where $[\mathrm{A}_\diamond] \triangleq [\mathrm{A}] + [\mathrm{A}_\star]$, $[\mathrm{A}^\star_\diamond] \triangleq [\mathrm{A}^\star] + [\mathrm{A}^\star_\star]$, $[\mathrm{A}^\diamond] \triangleq [\mathrm{A}] + [\mathrm{A}^\star]$, and $[\mathrm{A}^\diamond_\star] \triangleq [\mathrm{A}_\star] + [\mathrm{A}^\star_\star]$, are satisfied.

We now wonder whether the same reduction can be used in the case of the stochastic semantics. In the stochastic semantics, a chemical soup can be denoted by a 4-tuple $\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle$ of natural numbers, where n_X is the number of instance of X in the chemical soup, for any $X \in \{\mathrm{A}, \mathrm{A}_\star, \mathrm{A}^\star, \mathrm{A}^\star_\star\}$. The probability $P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle)$ that the system is in a given state $\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle$ at time t is given by the following master equation:

$$
\begin{aligned}
&P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle)' = \\
&\quad k^+(n_\mathrm{A}+1)P_t(\langle n_\mathrm{A}{+}1, n_{\mathrm{A}^\star}{-}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
&+\ k^+(n_{\mathrm{A}_\star}+1)P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{+}1, n_{\mathrm{A}^\star_\star}{-}1\rangle) \\
&+\ k^-(n_{\mathrm{A}^\star}+1)P_t(\langle n_\mathrm{A}{-}1, n_{\mathrm{A}^\star}{+}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
&+\ k^-(n_{\mathrm{A}^\star_\star}+1)P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{-}1, n_{\mathrm{A}^\star_\star}{+}1\rangle) \\
&+\ k_+(n_\mathrm{A}+1)(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star})P_t(\langle n_\mathrm{A}{+}1, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{-}1, n_{\mathrm{A}^\star_\star}\rangle) \\
&+\ k_+(n_{\mathrm{A}^\star}+1)(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star}-1)P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}{+}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}{-}1\rangle) \\
&+\ k_-(n_{\mathrm{A}_\star}+1)P_t(\langle n_\mathrm{A}{-}1, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{+}1, n_{\mathrm{A}^\star_\star}\rangle) \\
&+\ k_-(n_{\mathrm{A}^\star_\star}+1)P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}{-}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}{+}1\rangle)
\end{aligned}
$$

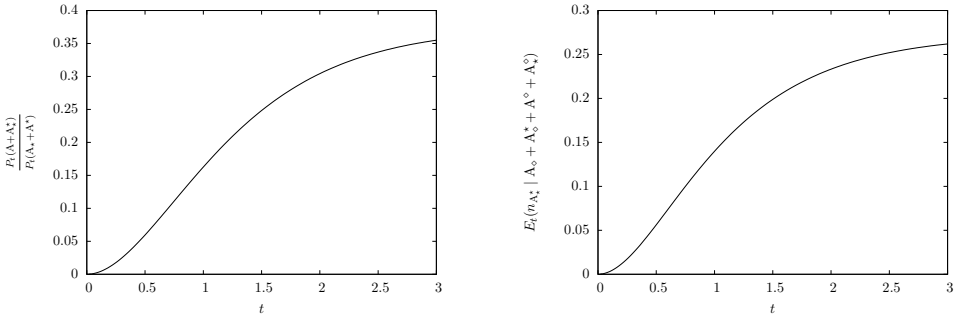


Fig. 7. On the left, quotient between the probability of being in the state $A + A^\star_\star$ and the probability of being in the state $A_\star + A^\star$. On the right, conditional expectation of the number of protein in the state $A^\star$ knowing that (i) there are two proteins, (ii) exactly one protein is phosphorilated at its the first site, and (iii) exactly one protein (potentially the same) is phosphorilated at its second site. All rates are set to 1. At time 0, the chemical soup is made of two proteins A fully unphosphorilated.

$$
\begin{aligned}
&- k^+(n_A + n_{A_\star})P_t(\langle n_A, n_{A^\star}, n_{A_\star}, n_{A^\star_\star}\rangle)\\
&- k^-(n_{A^\star} + n_{A^\star_\star})P_t(\langle n_A, n_{A^\star}, n_{A_\star}, n_{A^\star_\star}\rangle)\\
&- k_+(n_A(n_{A^\star} + n_{A^\star_\star}) + n_{A^\star}(n_{A^\star} + n_{A^\star_\star} - 1))P_t(\langle n_A, n_{A^\star}, n_{A_\star}, n_{A^\star_\star}\rangle)\\
&- k_-(n_{A_\star} + n_{A^\star_\star})P_t(\langle n_A, n_{A^\star}, n_{A_\star}, n_{A^\star_\star}\rangle).
\end{aligned}
$$

Given a state $\sigma = \langle n_A, n_{A^\star}, n_{A_\star}, n_{A^\star_\star}\rangle$, we denote by $\beta(\sigma)$ the 4-tuple $\langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}, n_{A^\diamond_\star}\rangle$ (such a tuple is called an abstract state). The probability $P^\sharp_t(\sigma^\sharp)$ that the system is in a state σ such that $\beta(\sigma) = \sigma^\sharp$ at time t, satisfies the following equation:

$$
\begin{aligned}
&P^\sharp_t(\langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}, n_{A^\diamond_\star}\rangle)' =\\
&\quad k_+(n_{A_\diamond} + 1)P^\sharp_t(\langle n_{A_\diamond}+1, n_{A^\star_\diamond}-1, n_{A^\diamond}, n_{A^\diamond_\star}\rangle)\\
&+ k^-(n_{A^\star_\diamond} + 1)P^\sharp_t(\langle n_{A_\diamond}-1, n_{A^\star_\diamond}+1, n_{A^\diamond}, n_{A^\diamond_\star}\rangle)\\
&+ k_+(n_{A^\diamond} + 1)n_{A^\star_\diamond}P^\sharp_t(\langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}+1, n_{A^\diamond_\star}-1\rangle)\\
&+ k_-(n_{A^\diamond_\star} + 1)P^\sharp_t(\langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}-1, n_{A^\diamond_\star}+1\rangle)\\
&- (k^+n_{A_\diamond} + k^-n_{A^\star_\diamond} + k_+n_{A^\diamond}n_{A^\star_\diamond} + k_-n_{A^\diamond_\star})P^\sharp_t(\langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}, n_{A^\diamond_\star}\rangle)\\
&- k_+\tilde{E}_t(n_{A^\star} \mid \langle n_{A_\diamond}, n_{A^\star_\diamond}+1, n_{A^\diamond}, n_{A^\diamond_\star}-1\rangle)\\
&+ k_+\tilde{E}_t(n_{A^\star} \mid \langle n_{A_\diamond}, n_{A^\star_\diamond}, n_{A^\diamond}, n_{A^\diamond_\star}\rangle),
\end{aligned}
$$

where for any expression $X(\sigma)$ and any (abstract) state $\sigma^\sharp$, the expression $\tilde{E}_t(X(\sigma) \mid \sigma^\sharp)$ denotes the product between the conditional expectation $E_t(X(\sigma) \mid \sigma^\sharp)$ of the expression $X(\sigma)$ knowing that $\beta(\sigma) = \sigma^\sharp$ and the probability $P^\sharp_t(\sigma^\sharp)$ of being in a state σ such that $\beta(\sigma) = \sigma^\sharp$.

Model	early EGF	EGF/Insulin cross talk	SFB
Species	356	2899	$\sim 2.10^{19}$
ODE fragments	38	208	$\sim 2.10^{5}$
Stochastic fragments	356	618	$\sim 2.10^{19}$

Fig. 8. Reduction factors for differential fragments [20,16] and stochastic fragments. We try these reduction methods on three models. The first one is the model of the early events of the EGF pathway (see Sect. 2); the second one, taken from [10, table 7], describes the cross-talk between another model of the early events of the EGF pathway and the insulin receptor; whereas the third one is a version of a pilot study on a larger section of the EGF pathway [15,1,25,6].

In general, the conditional properties of the number of instances of proteins in the form $A^\star$ having fixed a given abstract state, is time-dependent. We show in Fig. 7(a) that the ratio between the probability of being in the state $A + A^\star_\star$ and the probability of being in the state $A_\star + A^\star$ is time-dependent. Moreover, we show in Fig. 7(b) that the conditional expectation of $n_{A^\star}$ knowing that we are in the (abstract) state $A_\diamond + A^\star_\diamond + A^\diamond + A^\diamond_\star$ is time-dependent, which forbids doing the same simplication as in the differential semantics.

We have seen through this example that, because a given instance of chemical species can only be used once as a reactant when applying a given chemical reaction, some corrective terms as $+1$ or -1 may appear in master equations. These corrective terms may forbid the reduction of stochastic semantics. Nevertheless, this is not an issue when reducing differential semantics, since these corrective terms vanish when we consider an infinite number of instances of proteins (within an infinite volume).

6 Conclusion

In this paper, we have illustrated through small examples why it is more difficult to reduce the dimension of the state space of stochastic semantics than the one of differential semantics. In the case of the differential semantics, it is possible to abstract away some correlations between the state of some fragments of chemical species, because these correlations are not observed by the (groups of) reactions. This is not so easy in the case of stochastic semantics, because a given reaction application may operate on several fragments simultaneously, in such a case the choice for the state of fragments on which semi-reactions are applied is driven by the correlation between the state of these two fragments (see Sect. 4). Moreover, stochastic semantics counts individuals which leads to some constant corrective terms (such as increment or decrement by 1) which also forbids exact reduction (see Sect. 5).

In Fig. 8, we give the number of chemical species, the number of differential fragments, and the number of stochastic fragments for three bigger models. The reduction factor for the differential semantics is very interesting, whereas there is almost no reduction in the stochastic case. A careful look into the models would show that this is due to coupled semi-reactions. Moreover, the reduction that arises in the second model is due to a protein which has two fully independent parts (as

in Sect. 3).

This emphasizes how interesting the stochastic semantics is: the stochastic semantics does not only describe a limit behavior, but also showes the variability of a system and how robust a system is to stochastic variations. The counterpart is that it is very difficult to handle with (as a formal object) and to simplify.

References

[1] Blinov, M. L., J. R. Faeder, B. Goldstein and W. S. Hlavacek, *A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity*, BioSystems **83** (2006), pp. 136–151.

[2] Blinov, M. L., J. R. Faeder and W. S. Hlavacek, *BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains*, Bioinformatics **20** (2004), pp. 3289–3292.

[3] Borisov, N. M., A. S. Chistopolsky, J. R. Faeder and B. N. Kholodenko, *Domain-oriented reduction of rule-based network models*, IET Syst. Biol. **2** (2008), pp. 342–351.

[4] Borisov, N. M., N. I. Markevich, B. N. Kholodenko and E. D. Gilles, *Signaling through receptors and scaffolds: Independent interactions reduce combinatorial complexity*, Biophysical Journal **89** (2005), pp. 951–966.

[5] Bortz, A. B., M. H. Kalos and J. L. Lebowitz, *A new algorithm for monte carlo simulation of ising spin systems*, Journal of Computational Physics **17** (1975), pp. 10–18.

[6] Brightman, F. A. and D. A. Fell, *Differential feedback regulation of the mapk cascade underlies the quantitative differences in egf and ngf signalling in pc12 cells*, FEBS Letters **482** (2000), pp. 169–174.

[7] Buchholz, P., *Exact and ordinary lumpability in finite markov chains*, Journal of Applied Probability **31** (1994), pp. 59–75.

[8] Buchholz, P., *Bisimulation relations for weighted automata*, Theoretical Computer Science **393** (2008), pp. 109–123.

[9] Conzelmann, H., "Mathematical Modeling of Cellular Signal Transduction Pathways — A Domain-Oriented Approach to Reduce Combinatorial Complexity," Ph.D. thesis, Institut für Systemdynamik des Universität Stuttgart (2008).

[10] Conzelmann, H., D. Fey and E. D. Gilles, *Exact model reduction of combinatorial reaction networks*, BMC Systems Biology **2** (2008), p. 78.

[11] Conzelmann, H., J. Saez-Rodriguez, T. Sauter, B. N. Kholodenko and E. D. Gilles, *A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks*, BMC Bioinformatics **7** (2006), p. 34.

[12] Cousot, P., "Méthodes itératives de construction et d'approximation de points fixes d'opérateurs monotones sur un treillis, analyse sémantique de programmes (in French)," Thèse d'État ès sciences mathématiques, Université Joseph Fourier, Grenoble, France (1978).

[13] Cousot, P. and R. Cousot, *Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints*, in: *Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 1977, pp. 238–252.

[14] Danos, V., J. Feret, W. Fontana, R. Harmer and J. Krivine, *Rule-based modelling of cellular signalling*, in: *Proceedings of the 18th International Conference on Concurrency Theory (CONCUR'07)*, LNCS **4703** (2007), pp. 17–41.

[15] Danos, V., J. Feret, W. Fontana, R. Harmer and J. Krivine, *Rule-based modelling of cellular signalling, invited paper*, in: *Proceedings of the Eighteenth International Conference on Concurrency Theory, CONCUR '2007, Lisbon, Portugal*, Lecture Notes in Computer Science **4703** (2007), pp. 17–41.

[16] Danos, V., J. Feret, W. Fontana, R. Harmer and J. Krivine, *Abstracting the differential semantics of rule-based models: exact and automated model reduction*, in: *Proceedings of the Twenty-Fifth Annual IEEE Symposium on Logic in Computer Science, LICS '2010*, Edinburgh, UK, 2010, to appear.

[17] Danos, V., J. Feret, W. Fontana and J. Krivine, *Scalable simulation of cellular signaling networks*, in: Z. Shao, editor, *APLAS*, Lecture Notes in Computer Science **4807** (2007), pp. 139–157.

[18] Danos, V. and C. Laneve, *Core formal molecular biology*, Theoretical Computer Science **325** (2003), pp. 69–110.

[19] Degano, P., D. Prandi, C. Priami and P. Quaglia, *Beta-binders for biological quantitative experiments*, in: *Proceedings of QAPL 2006*, ENTCS **164**, 2006, pp. 101–117.

[20] Feret, J., V. Danos, J. Krivine, R. Harmer and W. Fontana, *Internal coarse-graining of molecular systems*, Proceedings of the National Academy of Sciences **106** (2009), pp. 6453–6458.

[21] Feret, J., H. Koeppl and T. Petrov, *Stochastic fragments: A framework for the exact reduction of the stochastic semantics of rule-based models*, International Journal of Software and Informatics To appear.

[22] Gillespie, D. T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comp. Phys. **22** (1976), pp. 403–434.

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

[24] Phillips, A. and L. Cardelli, *Efficient, correct simulation of biological processes in the stochastic pi-calculus*, in: *Proceedings of CMSB'07*, 2007, to appear.

[25] Schoeberl, B., C. Eichler-Jonsson, E. D. Gilles and G. Müller, *Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors.*, Nat Biotechnol **20** (2002), pp. 370–375.

[26] Sokolova, A. and E. d. Vink, *On relational properties of lumpability*, in: *Proc. PROGRESS Workshop 2003* (2003), p. 6pp.