Context-sensitive flow analyses: a hierarchy of model reductions

Ferdinanda Camporesi[2,3], Jérôme Feret[2], and Jonathan Hayman[1,2]

[1] Computer Laboratory, University of Cambridge, UK
[2] DIENS (INRIA/ÉNS/CNRS), Paris, France
[3] Dipartimento di Scienze dell'Informazione, Università di Bologna, Italy

Abstract. Rule-based modelling allows very compact descriptions of protein-protein interaction networks. However, combinatorial complexity increases again when one attempts to describe formally the behaviour of the networks, which motivates the use of abstractions to make these models more coarse-grained.
Context-insensitive abstractions of the intrinsic flow of information among the sites of chemical complexes through the rules have been proposed to infer sound coarse-graining, providing an efficient way to find macro-variables and the corresponding reduced models. In this paper, we propose a framework to allow the tuning of the context-sensitivity of the information flow analyses and show how these finer analyses can be used to find fewer macro-variables and smaller reduced differential models.

1 Introduction

Modellers of molecular signalling networks must cope with the combinatorial explosion of protein states generated by post-translational modifications and complex formations. Rule-based models provide a powerful alternative to approaches that require an explicit enumeration of all possible molecular species of a system [17,1]. Such models consist of formal rules stipulating the (partial) contexts for specific protein-protein interactions to occur. The behaviour of the models can be formally described by stochastic or differential semantics. Yet, the naive computation of these semantics does not scale to large systems, because it does not exploit the lower resolution at which rules specify interactions.

Rules explicitly describe the intrinsic flow of information between sites of complexes. Indeed, rules are contextual, and this context defines the state of which sites has an influence on the kinetic of corresponding interaction. This information can be used to detect some correlations that can be safely ignored. Thus, we can cut molecular complexes into molecular patterns, called fragments, and derive a system coarse-grained which describes exactly the concentration (or population) evolution of these fragments. This method never requires the execution of the concrete rule-based model and the approach is proved exact by abstract interpretation [12].

The so-obtained coarse-graining crucially depends on the accuracy of the analysis of the intrinsic flow of information. In this paper, we introduce a framework to tune the context-sensitivity of the analyses of the flow of information and derive the induced coarse-graining. This way, our analysis can zoom-in or zoom-out to increase or decrease the accuracy of the description of the flow of information which pass through each site, according to some conditions about the states of the other sites around this site. It bridges the gap between fully insensitive analyses (where the information about the sites are summarized according to their type only) [18,16,6] and fully context-sensitive analyses (that are computed in the concrete on molecular species) [21] that have been proposed so far, and provide a whole hierarchy of trade-off between accuracy and efficiency.

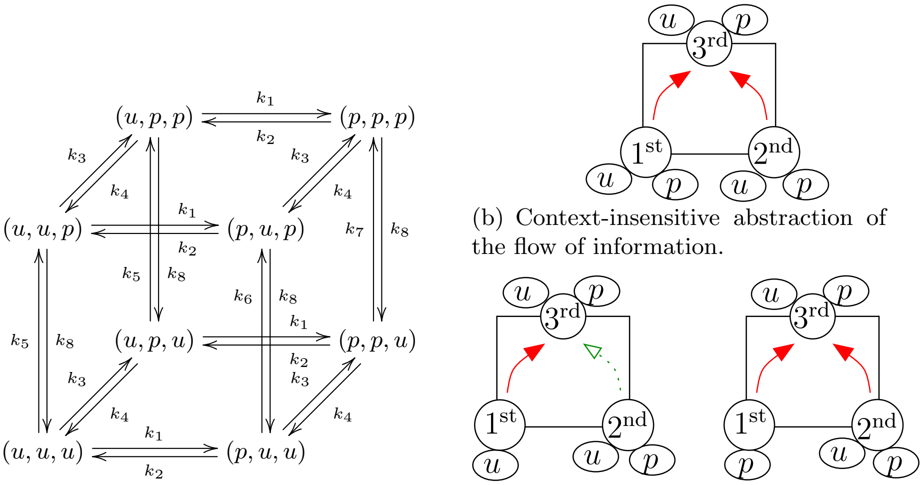
Related works. Dependencies between sites and reactions have been used in systematic methods for (hand-)writing coarse-grained models [10,4,8,9]. In [3], these approaches have been automatised for the models that are written in BNGL [1]. These informal methods do not provide exact coarse-graining in several cases, as whenever a site is activated through a binding or in the case of homo/hetero dimerizations. In [18,16,6], we have introduced a formal framework which copes with the full Kappa [17] and which ensures a formal relation between the initial differential semantics and the reduced one, by the means of abstract interpretation [12,11]. A similar framework has been proposed for lumping the stochastic semantics (which is defined as a continuous time Markov chain) [20,19]. Symmetries can also be used to reduce the combinatorial complexity of models [7].

These methods are context-insensitive: for each kind of agents, all the information about the agents of this kind is summarised into the single node of a graph, called the contact map. In [21] a fully context-sensitive framework is proposed in which the abstraction of the information flow is done in the concrete, thanks to a direct iteration on the molecular species. In comparison, the framework that is proposed in this paper allows the user to select any trade-off between fully context-insensitive and fully context-sensitive abstractions of the flow of information.

Context-sensitive approximations of graph-structures have been deeply studied in the field of memory analysis, where complex invariants [23] about recursive data-structures have to be inferred [24,2]. Our framework is a kind of partitioning [5], a generic method for refining abstractions.

2 Case study

Before describing the framework formally, we introduce a case study. We consider one kind of protein P with three numbered phosphorylation sites. Each site can be phosphorylated, or not, thus each instance of P can take 8 configurations. We consider that any site can get phosphorylated or lose its phosphorylation, thus there are 24 chemical reactions (3 per configurations). Configurations and reactions are summarised in Fig. 1(a). The configuration of a protein P is denoted as a triple (s_1, s_2, s_3) of symbols among u and p. In general, the rate of phosphorylation (resp. dephosphorylation) can depend on the state of the other sites. Here we make the assumption that only the phosphorylation rate of the third



(a) Configurations and reactions.

(b) Context-insensitive abstraction of the flow of information.

(c) Context-sensitive abstraction of the flow of information. The dotted arrow represents a dependency that can be dropped if the rate k_1 is zero.

Fig. 1. Case study

site depends on the state of the two other sites, but we assume that the phosphorylation rate of the third site is the same in the configurations (u,u,u) and (u,p,u). Thus, our system is parameterised by 8 kinetic rates (e.g.. see Fig.1(a)).

So as to model the behaviour of our system, we assume 1) that the system satisfies the well stirred assumption of mass action law, and 2) that the population of proteins is large. Under these assumptions, the behaviour of the system can be formalised by the means of the following system of differential equations, which describes the derivatives of the concentrations of each configuration of P as an expression of the concentrations of the configurations of P:

$$\begin{array}{l}
[(u,u,u)]' = k_2[(p,u,u)] + k_4[(u,p,u)] + k_8[(u,u,p)] - (k_1+k_3+k_5)[(u,u,u)]\\
[(u,u,p)]' = k_2[(p,u,p)] + k_4[(u,p,p)] + k_5[(u,u,u)] - (k_1+k_3+k_8)[(u,u,p)]\\
[(u,p,p)]' = k_2[(p,p,p)] + k_3[(u,u,p)] + k_5[(u,p,u)] - (k_1+k_4+k_8)[(u,p,p)]\\
[(u,p,u)]' = k_2[(p,p,u)] + k_3[(u,u,u)] + k_8[(u,p,p)] - (k_1+k_4+k_5)[(u,p,u)]\\
[(p,p,u)]' = k_1[(u,p,u)] + k_3[(p,u,u)] + k_8[(p,p,p)] - (k_2+k_4+k_7)[(p,p,u)]\\
[(p,p,p)]' = k_1[(u,p,p)] + k_3[(p,u,p)] + k_7[(p,p,u)] - (k_2+k_4+k_8)[(p,p,p)]\\
[(p,u,p)]' = k_1[(u,u,p)] + k_4[(p,p,p)] + k_6[(p,u,u)] - (k_2+k_3+k_8)[(p,u,p)]\\
[(p,u,u)]' = k_1[(u,u,u)] + k_4[(p,p,u)] + k_8[(p,u,p)] - (k_2+k_3+k_6)[(p,u,u)].
\end{array}$$

Providing an initial state mapping each configuration to their initial concentration, this system has a unique smooth solution over the time interval $\mathbb{R}^+$.

Now, we wonder whether or not our model can be coarse-grained. Thus we are looking for a set macro-variables which are defined as a linear combination of the variables of the initial systems (so called micro-variables) that are self-consistent. That is to say that the derivatives of the macro-variables must be expressed as functions of only the macro-variables. In previous works [18,16,6], we have introduced frameworks for detecting self-consistent coarse-graining thanks to an over-approximation of the flow of information between the states of the sites of proteins. Indeed the flow of information can be summarised by annotating a contact map (which describes the different kinds of proteins, their sites, their potential phosphorylation states and their potential binding) with an oriented relation over the sites. This relation summarises how each site may influence the other ones: an arrow from a site s_1 to a site s_2 means that the capability of modifying the state of the site s_2 may change according to the state of the site s_1. The annotated contact map for our case study is given in Fig. 1(b). This is a context-insensitive approximation since all the information about the sites of P is summarised in a single node, regardless the states of its sites. The arrow from the 1$^{\text{st}}$ (resp. 2$^{\text{nd}}$) site to the 3$^{\text{rd}}$ one comes from the fact that the rate k_6 may not be equal to the rate k_5 (resp. k_7). But no other arrows are required, since the phosphorylation/dephosphorylation rates of both the 1$^{\text{st}}$ and 2$^{\text{nd}}$ sites do not depend on the states of other sites (indeed, we can check on Fig. 1(a) that the rates of the corresponding reactions are the same four by four). As a result, since the behaviour of the 3$^{\text{rd}}$ site depends on the state of all the other sites, no coarse-graining can be found in this way.

Indeed, without further assumptions, the model cannot be coarse-grained by any means. But interestingly, if we knockout the phosphorylation reactions of the 1$^{\text{st}}$ site (that is to say we set the rate k_1 equal to 0), the model can be coarse-grained by abstracting away the relation between the state of the 2$^{\text{nd}}$ site and the 3$^{\text{rd}}$ site in the case when the 1$^{\text{st}}$ site is activated. This is formalised by the following differential equations:

$$\begin{aligned}
&[(?,u,?)]' = k_4[(?,p,?)] - k_3[(?,u,?)]\\
&[(?,p,?)]' = k_3[(?,u,?)] - k_4[(?,p,?)]\\
&[(u,?,p)]' = k_2([(p,u,p)] + [(p,p,p)]) + k_5[(u,?,u)] - k_8[(u,?,p)]\\
&[(u,?,u)]' = k_2([(p,u,u)] + [(p,p,u)]) + k_8[(u,?,p)] - k_5[(u,?,u)]\\
&[(p,p,u)]' = k_3[(p,u,u)] + k_8[(p,p,p)] - (k_2 + k_4 + k_7)[(p,p,u)]\\
&[(p,p,p)]' = k_3[(p,u,p)] + k_7[(p,p,u)] - (k_2 + k_4 + k_8)[(p,p,p)]\\
&[(p,u,p)]' = k_4[(p,p,p)] + k_6[(p,u,u)] - (k_2 + k_3 + k_8)[(p,u,p)]\\
&[(p,u,u)]' = k_4[(p,p,u)] + k_8[(p,u,p)] - (k_2 + k_3 + k_6)[(p,u,u)],
\end{aligned}$$

where the macro-variables are intentionally defined as fragments or portions of configurations (the question mark symbol denotes sites which have been cut away), and extentionally as linear combinations of the configurations which contain these fragments:

$$\begin{aligned}
[(?,u,?)] &= [(u,u,u)] + [(u,u,p)] + [(p,u,u)] + [(p,u,p)]\\
[(?,p,?)] &= [(u,p,u)] + [(u,p,p)] + [(p,p,u)] + [(p,p,p)]\\
[(u,?,u)] &= [(u,u,u)] + [(u,p,u)]\\
[(u,?,p)] &= [(u,u,p)] + [(u,p,p)].
\end{aligned}$$

This coarse-graining can be discovered by tuning the context-sensitivity of the information flow analysis. Indeed, the behaviour of the protein P can be partitioned into two distinct modes. Whenever the 1$^{\text{st}}$ site is phosphorylated, the evolution of the state of the 3rd site is controlled by the state of both the 1st and the 2nd sites. But whenever the 1st site is not phosphorylated, the evolution of the state of the 3rd site is not controlled by the state of the 2nd site any more (this can be checked in Fig. 1(a) where the phosphorylation (resp. unphosphorylation) rate of the 3rd site is the same whatever the protein is in the state (u, u, u) or (u, p, u) (resp. (u, u, p) or (u, p, p))). This accurate approximation of the flow of information is out of the reach of context-insensitive analysis. Thus we propose to use arbitrary Σ-graphs where different annotations can be written according to the state of well chosen sites, unlike the contact map. An example Σ-graph is given in Fig. 1(c). We notice that two nodes are used to describe the protein P, according to whether or not the 1st site is phosphorylated. The notion of Σ-graph will be formally defined in Sect. 3. Then, we can annotate our Σ-graph with context-sensitive information about the flow of information and obtain the plain arrows in Fig. 1(c). Interestingly, in the left connected component, there is no flow of information from any site into the 2nd site and no flow of information from the 2nd site into the 3rd site. As a consequence, the fragments of proteins that contain the 1st and the 3rd site and the ones that only contain the 2nd site are good candidates as macro-variables. Yet, since in the right connected component there is a potential flow of information from the 1st and the 2nd sites into the 3rd site, any micro-variable where the 1st site is phosphorylated has to be preserved. Thus, we find again the set of macro-variables $\{[(?, u, ?)], [(?, p, ?)], [(u, ?, u)], [(u, ?, p)], [(p, p, u)], [(p, p, p)], [(p, u, p)], [(p, u, u)]\}$, which is self-consistent, as we have shown previously.

Then we may wonder why this coarse-graining is not self-consistent when the phosphorylation reaction of the 1st site is not knocked out. This is because configurations of the form $(u, ?, ?)$ can now be transformed into configurations of the form $(p, ?, ?)$. Then, so as to express the concentration of the configurations of the form $(p, ?, ?)$ which are produced this way, we need to express the configurations of the form $(u, ?, ?)$ with at least the same fine-grained level of description. This is captured by the right-gluing construction in [21]. In the present framework, it is necessary to duplicate the arrow between the 2nd site and the 3rd one, from the right connected component into the left one. The resulting arrow, drawn in dotted in Fig. 1(c), prevents any coarse-graining.

The rest of the paper is organised as follows. In Sec. 3, we remind the readers the notion of Σ-graphs and use them to abstract relations between sites in chemical complexes. In Sec. 4, we give an abstract syntax and a formal differential semantics for Kappa. In Sec. 5, we define our generic flow analysis and its induced model reduction.

3 Σ-graphs

Graphs with a given signature, Σ-graphs, play a central role in the semantics of Kappa. In this section, we remind the definition of Σ-graphs [14] and show how to annotate them with a relation over their sites.

Definition 1. *A signature is a tuple $\Sigma = (\Sigma_{ag}, \Sigma_{st}, \Sigma_{int}, \Sigma^{int}_{ag\text{-}st}, \Sigma^{lnk}_{ag\text{-}st})$ where Σ_{ag} is a finite set of agent types, Σ_{st} is a finite set of site identifiers, Σ_{int} is a finite set of internal state identifiers, $\Sigma^{lnk}_{ag\text{-}st} : \Sigma_{ag} \to \wp(\Sigma_{st})$ and $\Sigma^{int}_{ag\text{-}st} : \Sigma_{ag} \to \wp(\Sigma_{st})$ are site maps.*

Agent types in Σ_{ag} denote agents of interest, as kinds of proteins for instance. A site identifier in Σ_{st} represents an identified locus for capability of interactions between agents. Internal state identifiers in Σ_{int} are special attributes which encode potential state configurations, as the phosphorylation state for instance. Each agent type A is associated with a set of sites which can bear an internal state $\Sigma^{int}_{ag\text{-}st}(A)$ and a set of sites which can be linked $\Sigma^{lnk}_{ag\text{-}st}(A)$. We assume without reducing the expressive power of the framework that $\Sigma^{lnk}_{ag\text{-}st}(A) \cap \Sigma^{int}_{ag\text{-}st}(A) = \emptyset$, for any $A \in \Sigma_{ag}$ and we write $\Sigma_{ag\text{-}st}(A)$ for the set $\Sigma^{lnk}_{ag\text{-}st}(A) \uplus \Sigma^{int}_{ag\text{-}st}(A)$. In our case study, $\Sigma_{ag} = \{P\}$, $\Sigma_{st} = \{1^{\text{st}}, 2^{\text{nd}}, 3^{\text{rd}}\}$, $\Sigma_{int} = \{u, p\}$, $\Sigma^{int}_{ag\text{-}st}(P) = \Sigma_{st}$, and $\Sigma^{lnk}_{ag\text{-}st}(P) = \emptyset$.

Σ-graphs describe both patterns and chemical species. Their nodes are typed agents with some sites which can bear internal states and linking states. We introduce the set $\mathtt{Ext}$ as $\{\dashv, -\} \cup \{(A, s) \mid A \in \Sigma_{ag},\ s \in \Sigma^{lnk}_{ag\text{-}st}(A)\}$ for describing some linking states.

Definition 2. *A Σ-graph is a tuple $G = (\mathcal{A}, type, \mathcal{S}, \mathcal{L}, p\kappa)$ where $\mathcal{A}$ is a set of agents, $type : \mathcal{A} \to \Sigma_{ag}$ is a function mapping each agent to its type, $\mathcal{S}$ is a set of sites satisfying $\mathcal{S} \subseteq \{(n, s) \mid n \in \mathcal{A}, s \in \Sigma_{ag\text{-}st}(type(n))\}$, $\mathcal{L}$ is a symmetric relation such that $\mathcal{L} \subseteq (\{(n, i) \in \mathcal{S} \mid i \in \Sigma^{lnk}_{ag\text{-}st}(type(n))\} \cup \mathtt{Ext})^2 \setminus \mathtt{Ext}^2$, and $p\kappa$ maps each site $(n, i) \in \mathcal{S}$ such that $i \in \Sigma^{int}_{ag\text{-}st}(type(n))$ to a set of internal states $p\kappa(n, i) \in \wp(\Sigma_{int})$.*

A site $(n, i) \in \mathcal{S}$ such that $i \in \Sigma^{int}_{ag\text{-}st}(type(n))$ is called a property site, whereas a site $(n, i) \in \mathcal{S}$ such that $i \in \Sigma^{lnk}_{ag\text{-}st}(type(n))$ is called a binding site. Whenever $((n, i), \dashv) \in \mathcal{L}$, the binding site (n, i) may be free. Various levels of information can be given about the sites that can be bound. Whenever $((n, i), -) \in \mathcal{L}$, then the binding site (a, i) may be bound to any other site. Whenever $((n, i), (A', i')) \in \mathcal{L}$ for a given agent type $A' \in \Sigma_{ag}$ and a given site identifier $i' \in \Sigma^{lnk}_{ag\text{-}st}(A')$, then the binding site can be bound to the site i' of an agent of type A'. Whenever $((n, i), s) \in \mathcal{L}$ with $s \in \mathcal{S}$ then the binding site (n, i) may be bound to the binding site s. We introduce a sub-typing relation $\leq_G$ over binding states, that is defined as the least reflexive relation such that $- \leq_G (type(n), i) \leq_G (n, i)$, for any $n \in \mathcal{A}$ and $i \in \Sigma^{lnk}_{ag\text{-}st}(type(n))$.

For a Σ-graph G, we write as $\mathcal{A}_G$ for its set of agents, $type_G$ for its typing function, $\mathcal{S}_G$ its set of sites, $\mathcal{L}_G$ its set of links, and $p\kappa_G$ for the internal states map.

Two Σ-graphs can be related by structure-preserving functions, which are called homomorphisms.

Definition 3. *A homomorphism $h : G \to H$ between two Σ-graphs G and H is a function of agents $h : \mathcal{A}_G \to \mathcal{A}_H$ satisfying:*

1. $type_G(n) = type_H(h(n))$ for all $n \in \mathcal{A}_G$;

2. $(h(n), i) \in \mathcal{S}_H$ *for all* $(n, i) \in \mathcal{S}_G$*;*
3. $p\kappa_G(n, i) \subseteq p\kappa_H(h(n), i)$ *for all* $(n, i) \in \mathcal{S}_G$ *such that* $i \in \Sigma^{int}_{ag\text{-}st}(type_G(n))$*;*
4. $((h(n), i), (h(n'), i') \in \mathcal{L}_H$ *for all* $((n, i), (n', i')) \in \mathcal{L}_G \cap \mathcal{S}_G^2$*;*
5. *there exists* $y \in \mathcal{S}_H \cup \texttt{Ext}$ *such that* $((h(n), i), y) \in \mathcal{L}_H$ *and* $x \leq_H y$ *for all* $((n, i), x) \in \mathcal{L}_G$ *such that* $x \in \texttt{Ext}$*.*

Whenever h is an injection, then h is called an embedding. The number of embeddings between two Σ-graphs G and H is denoted as $[G, H]$. Whenever $G = H$ and h is a bijection, then h is called an automorphism. We notice that the identity function is always an automorphism. Homomorphisms $f : G \to H$ and $g : H \to K$ compose in the usual way. Moreover, whenever two homomorphisms $f : G \to H$ and $g : H \to G$ are such that $g \circ f$ is the identity homomorphism over G and $f \circ g$ is the identity homomorphism over H, then f and g are called isomorphisms and G and H are said to be isomorphic which is written $G \approx H$. All the constructions in this paper are defined up to isomorphisms. Isomorphisms (and automorphisms) are all embeddings.

Now we want to annotate a Σ-graph with a binary relation over its sites, so as to, for instance, abstract the flow of information among its sites. Two sites can be in relation when they belong to the same agent or when they are linked together.

Definition 4. *An annotated* Σ*-graph* $G^{\mathrm{a}} = (G, \rightsquigarrow_{G^{\mathrm{a}}})$ *is a pair where* G *is a* Σ*-graph and* $\rightsquigarrow_{G^{\mathrm{a}}}$ *is a subset of* $\{(n, i), (n, i') \mid n \in \mathcal{A}_G, (n, i), (n, i') \in \mathcal{S}_G\} \uplus (\mathcal{L}_G \cap \mathcal{S}_G^2)$*.*

Ordered pairs of sites in $\{(n, i), (n, i') \mid n \in \mathcal{A}_G, (n, i), (n, i') \in \mathcal{S}_G\}$ are called internal edges and are denoted as $(n, i) \overset{\vee}{\rightsquigarrow}_{G^{\mathrm{a}}} (n, i')$, whereas ordered pairs in $\mathcal{L}_G \cap \mathcal{S}_G^2$ are called external edges and are denoted as $(n, i) \overset{\wedge}{\rightsquigarrow}_{G^{\mathrm{a}}} (n', i')$. An ordered pair of sites can be connected by both an internal edge and an external edge. We omit the symbols $\vee$ and $\wedge$ when they are not important.

Given an annotated Σ-graph G^{a}, we write as G the Σ-graph and $\rightsquigarrow_{G^{\mathrm{a}}}$ the binary relation over its sites.

The set of annotations of a Σ-graph G forms a Boolean lattice isomorphic to the set of the parts of $\{((n, i), (n, i')) \mid n \in \mathcal{A}_G, (n, i), (n, i') \in \mathcal{S}_G\} \uplus (\mathcal{L}_G \cap \mathcal{S}_G^2)$. The least element is the empty annotation and is denoted as $G^{\emptyset} = (G, \emptyset)$, and the top element relates each pair of sites such that they either belong to the same agent or are linked together and is denoted as $G^{\top} = (G, \rightsquigarrow_{G^{\top}})$.

4 Differential semantics

Mixtures, representing the states to which rules are applied, and site graphs, representing patterns, are forms of Σ-graph. In particular, site graphs are finite, they have no links that immediately loop back to the same site and have at most one link from any site, moreover their sites can bear at most one internal state. Mixtures additionally specify the link state and the internal state of all sites and have no external links.

Definition 5. *A site graph G is a Σ-graph such that: 1) the set $\mathcal{A}_G$ is finite; 2) its link relation $\mathcal{L}_G$ is irreflexive; 3) for any binding site $(n,i) \in \mathcal{S}_G$, $((n,i),x) \in \mathcal{L}_G$ and $((n,i),y) \in \mathcal{L}_G$ implies $x = y$; and 4) for any state site $(n,i) \in \mathcal{S}_G$, $p\kappa(n,i)$ contains at most one element.*

In a site graph G, the states of some sites in $\mathcal{S}_G$ are specified, while others are not. The state of a binding site $(n,i) \in \mathcal{S}_G$ is specified if there exists $x \in \mathcal{S}_G \cup \mathtt{Ext}$ such that $((n,i),x) \in \mathcal{L}_G$, whereas the state of a state site $(n,i) \in \mathcal{S}_G$ is specified if $p\kappa_G(n,i)$ is a singleton.

Definition 6. *A Σ-graph G is said to be fully specified if the three following properties hold: 1) $\mathcal{S} = \{(n,s) \mid n \in \mathcal{A}, s \in \Sigma_{ag\text{-}st}(type(n))\}$; 2) $\mathcal{L} \subseteq (\mathcal{S} \cup \{\dashv\})^2$; and 3) the state of each site in $\mathcal{S}_G$ is specified.*

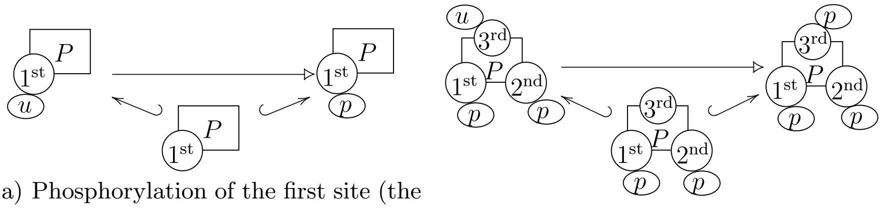
Definition 7. *A site graph G is said to be connected if for any pair of distinct agents $n_1, n_2 \in \mathcal{A}$, there exists two sites $i_1, i_2 \in \Sigma_{st}$ such that $(n_1,i_1), (n_2,i_2) \in \mathcal{S}$ and $(n_1,i_1) \rightsquigarrow^{\star}_{G^{\top}} (n_2,i_2)$.*

A site graph can be decomposed in a set of connected graphs, called its connected components. A mixture is a fully specified site graph. The variables of the differential semantics are the concentrations of the chemical species, which are defined as isomorphism classes of connected mixtures. Thus we introduce $\mathcal{C}$ as a set of connected mixtures such that for any connected mixture c there exists a unique connected mixture $c' \in \mathcal{C}$ such that $c \approx c'$. We assume that $\mathcal{C}$ is finite (that is to say that there is no polymerisation).

Transformations between site graphs are described by rules. A rule is a transformation between two site graphs, a left hand side (*lhs*) L and a right hand side (*rhs*) R. In a rule, some agents and some sites are preserved. This is specified by a site graph D which is embedded both into L and into R and which describes anything that is preserved. Not all transformations are allowed: one can remove and add agents, create links between free sites, free pairs of sites that are connected and change the internal state of some sites. The agents that are created have to fully define the state of their sites. We also make extra assumptions to simplify the definitions which are involved in the approximation of the flow of information: we assume that only the bonds that are shared between two sites can be removed, and that the agents that are removed have to fully define the state of their sites has well. The framework can be easily tuned to relax these two assumptions. Our requirements are formalised in the following definition:

Definition 8. *A rule is a span such that: $L \xhookleftarrow{f} D \xhookrightarrow{g} R$ such that :*

1. *for any span $L \xhookleftarrow{f'} D' \xhookrightarrow{g'} R$ and any embedding $D \xhookrightarrow{h} D'$ such that $f = f'h$ and $g = g'h$, then h is an isomorphism;*
2. *for any $x \in \mathtt{Ext} \setminus \{\dashv\}$ and any site $(n,i) \in \mathcal{S}_L$, if $((n,i),x) \in \mathcal{L}_L$ then there exists $m \in \mathcal{A}_D$ such that $n = f(m)$, $(m,i) \in \mathcal{S}_D$, and $((m,i),x) \in \mathcal{L}_D$;*
3. *for any $x \in \mathtt{Ext} \setminus \{\dashv\}$ and any site $(n,i) \in \mathcal{S}_R$, if $((n,i),x) \in \mathcal{L}_R$ then there exists $m \in \mathcal{A}_D$ such that $n = g(m)$, $(m,i) \in \mathcal{S}_D$, and $((m,i),x) \in \mathcal{L}_D$;*



(a) Phosphorylation of the first site (the rate depends on the states neither of the 2^{nd} site, nor the 3^{rd} site).

(b) Phosphorylation of the 3^{rd} site when the two other sites are phosphorylated.

Fig. 2. Examples of rules.

4. *if $m \in \mathcal{A}_D$, then for any $i \in \Sigma_{ag\text{-}st}(type_D(m))$, $(m, i) \in \mathcal{S}_D$ iff $(f(m), i) \in \mathcal{S}_L$ iff $((g(m), i)) \in \mathcal{S}_R$ and, in such a case, the state of the site $(f(m), i)$ is specified in the site graph L iff the state of the site $(g(m), i)$ is specified in the site graph R;*
5. *if $m \in \mathcal{A}_L$ and $m \notin image(f)$, then, for any $i \in \Sigma_{ag\text{-}st}(type_L(m))$, $(f(m), i) \in \mathcal{S}_L$ and the state of the site $(f(m), i)$ is specified in the site graph L;*
6. *if $m \in \mathcal{A}_R$ and $m \notin image(g)$, then, for any $i \in \Sigma_{ag\text{-}st}(type_R(m))$, $(g(m), i) \in \mathcal{S}_R$ and the state of the site $(g(m), i)$ is specified in the site graph R.*

The first property ensures that D is a local greatest upper bound.

A rule $L \hookleftarrow D \hookrightarrow R$ is usually denoted as $L \rightarrowtriangle R$ (leaving the two embeddings and the common region implicit).

$$\begin{tikzcd}
L' \arrow[rr, -Triangle] & & R' \\
& D' \arrow[ul, hook] \arrow[ur, hook] & \\
L \arrow[uu, hook, "h_L"] \arrow[rr, -Triangle] & & R \arrow[uu, hook, "h_R"'] \\
& D \arrow[uu, hook, "h_D"] \arrow[ul, hook] \arrow[ur, hook] &
\end{tikzcd}$$

Rules can be more or less refined [15,22], by adding more or less information about the context in which they can be applied. A rule $L' \hookleftarrow D' \hookrightarrow R'$ is said to be a refinement of the rule $L \hookleftarrow D \hookrightarrow R$ is and only if there exist three embeddings h_L, h_D, h_R which make the diagram on the right commute. In such a case, the two action maps and the embeddings h_L and h_R form a pushout (e.g. see [14]). Moreover, whenever L' is a mixture, then R' is a mixture as well. Given L' (resp. R') a site graph and an embedding f between L and L' (resp. between R and R') there exists a unique (up to isomorphisms) refinement that is defined by a triple of embeddings $(h_X)_{X \in \{L,D,R\}}$ such that $h_L = f$ (resp. $h_R = f$), this refinement is called the left-refinement (resp. right-refinement) of r by the embedding f. The unicity of the right-refinement does not hold in full Kappa, but follows from our simplifying assumptions. Yet, in general there exists a unique least refined refinement such that $h_L = f$.

Each rule comes with a kinetic rate k which is denoted as $r : L \rightarrowtriangle R \ @k$. The rule r can be seen as a symbolic representation of a set of reactions among chemical complexes, that is obtained as a left refinement of r by a join of embeddings mapping each connected component of L into a chemical complex $c \in \mathcal{C}$. For any such refinement, both the lhs and the rhs are mixtures and are respectively isomorphic to the disjoint union of a tuple of reactants $r_1, \ldots, r_m \in \mathcal{C}^\star$ and a tuple of products $p_1, \ldots, p_n \in \mathcal{C}^\star$. Each such refinement is associated with the following contribution to the system of differential equations, for any integer

s such that $1 \leq s \leq m$ and any integer t such that $1 \leq t \leq n$:

$$x_{r_s}(t)' \overset{+}{=} -k \cdot \frac{\prod_{1\leq j\leq m} x_{r_j}(t)}{[L,L]} \qquad \text{and} \qquad x_{p_t}(t)' \overset{+}{=} k \cdot \frac{\prod_{1\leq j\leq m} x_{r_j}(t)}{[L,L]},$$

which models reactants consumption and products creation (the expression $[L, L]$ is the number of automorphisms (i.e. symmetries) of L).

The differential semantics associated to a set of rules maps each initial state $\mathit{init} \in (\mathbb{R}^+)^{\mathcal{C}}$ to the unique solution $x \in ([0, T) \to \mathbb{R})^{\mathcal{C}}$ of the so obtained system of equations such that $x_c(0) = \mathit{init}_c$ for any $c \in \mathcal{C}$ and T is maximal. By construction, this solution satisfies the following property [16]: for any $t \in [0, T)$ and any $c \in \mathcal{C}$, $x_c(t) \geq 0$.

5 Context-sensitive model reduction

The annotation of a Σ-graph can be viewed as a symbolic representation of a set of patterns, called prefragments. More precisely, given an annotated Σ-graph G^{a}, a site graph P is a prefragment if we get a directed relation over its sites when we annotate it by the meet of the inverse image of the annotation of G^{a} by any homomorphism between P and G^{a}. This is formalised as follows:

Definition 9. *Given an annotated Σ-graph G^{a} and a Σ-graph H, we define the canonical annotation of H by the annotated Σ-graph G^{a} as follows: for any $(a, i), (a', i') \in \mathcal{S}_H$ and any $w \in \{\vee, \wedge\}$, $(a, i) \overset{w}{\rightsquigarrow}_{H,G^{\mathrm{a}}} (a', i')$ if and only if for all homomorphisms $\phi : G \to H$, $(\phi(a), i) \overset{w}{\rightsquigarrow}_{G^{\mathrm{a}}} (\phi(a'), i')$.*

Definition 10. *Given an annotated Σ-graph G^{a}, we say that a site graph P is a prefragment for G^{a} if and only if the set of sites $\mathcal{S}_P$ and the transitive and reflexive closure of the relation $\rightsquigarrow_{P,G^{\mathrm{a}}}$ form a directed set (i.e. for any $s_1, s_2 \in \mathcal{S}_P$, there exists $s \in \mathcal{S}_P$ such that $s_1 \rightsquigarrow^{\star}_{P,G^{\mathrm{a}}} s$ and $s_2 \rightsquigarrow^{\star}_{P,G^{\mathrm{a}}} s$).*

We notice that prefragments are connected and that, since the set $\mathcal{S}_P$ is finite, a site graph P is a prefragment if and only if there exists $s^{\bullet} \in \mathcal{S}_P$ such that for any site $s \in \mathcal{S}_P$, $s \rightsquigarrow_{P,G^{\star}_{\mathrm{a}}} s^{\bullet}$. In such a case, we call the site $s^{\bullet}$ (which may not be unique) a root of the prefragment.

A connected site graph P can also be seen extensionally as set of embeddings between itself and any reachable species in $\mathcal{C}$.

Definition 11. *For any connected site graph P, we define y_P as $\sum_{v\in\mathcal{C},\ \phi\,:\,P\to v} x_v$.*

The function x_v gives the concentration of v at time t. Thus the set of prefragments define a linear change of variables. Now we wonder how to annotate, given a set of rules, a Σ-graph such that this change of variables is self-consistent.

For this purpose, we will use a special kind of Σ-graphs, the summary graphs. Roughly speaking, summary graphs are used to abstract information about the potential overlaps between the left and right hand sides of rules and connected site graphs such that the common region contains sites that are modified by the rules. This will allows us to express the proper consumption and the proper production of these site graphs.

We now formalise the notions of summary graphs:

Definition 12. *A Σ-graph G is summary graph if the three following properties hold: 1) $\mathcal{L} \subseteq (\mathcal{S} \cup \{\dashv\})^2$; 2) for any chemical complex $V \in \mathcal{C}$, there exists a homomorphism $h : V \to G$; 3) for any homomorphism $h : P \to G$ between a connected site graph P and G, there exists a chemical complex $V \in \mathcal{C}$, an embedding $\phi : P \hookrightarrow V$ and a homomorphism $h' : V \to G$ such that $h = h'\phi$.*

The set of summary graphs is exactly the smallest set of Σ-graphs that contains the disjoint union of the family of chemical complexes and that is stable by disjoint union and folding of agent nodes (folding of agent nodes can be formalised as the application of a regular epimorphism). It follows that the disjoint union of all chemical complexes is a summary graph (the most concrete one), and the contact map is also a summary graph (the most abstract one).

An overlap between two site graphs is defined by a common region, which identifies some nodes in the two site graphs and a merged site graph, which ensures that the two site graphs are compatible. The common region can be chosen as a local lower-bound and the merged site graph as a local upper-bound, which ensures that each overlap is defined non-ambiguously:

Definition 13. *An overlap between two site graphs P and C is defined as a pair of a span $P \hookleftarrow X \hookrightarrow C$ and a cospan $P \hookrightarrow Y \hookleftarrow C$ of embeddings where X and Y are non-empty site graphs which make the square commute (see the diagram on the right) and such that for any other such pair of a span $P \hookleftarrow X' \hookrightarrow C$ and a cospan $P \hookrightarrow Y' \hookleftarrow C$ where X', Y' are site graphs, there exists a unique pair of embeddings $X' \hookrightarrow X$ and $Y \hookrightarrow Y'$ which makes the diagram on the right commute.*

$$\begin{tikzcd}
 & & & Y' \\
 & P \arrow[r, hook] \arrow[rru, hook, bend left] & Y \arrow[ru, hook, dashed, "!"] & \\
 & X \arrow[u, hook] \arrow[r, hook] \arrow[ru, phantom, "\lrcorner"] & C \arrow[u, hook] \arrow[ruu, hook, bend right] & \\
X' \arrow[ru, hook, dashed, "!"] \arrow[ruu, hook, bend left] \arrow[rru, hook, bend right] & & &
\end{tikzcd}$$

Flow of information abstracts the relation between the sites that are tested and the sites that are modified. A site is tested in a rule if it occurs in the lhs of this rule. A site $(n, i) \in \mathcal{S}_L$ is modified in the lhs L of a rule $L \xhookleftarrow{f} D \xhookrightarrow{g} R$ iff either $n \notin \mathit{image}(f)$, or (m, i) is not specified in D (where m is the unique agent $m \in \mathcal{A}_D$ such that $f(m) = n$). We define the same way the sites that are modified in the rhs of a rule. For the sake of simplicity, we assume that any connected component in the lhs of a rule that contains at least one site has at least one site that is modified (if not, we take an arbitrary site and continue as if it were modified). Thus, we consider that each connected component in the lhs of a rule has either no site, or has a site that is modified by the rule.

We call a path in a site graph P a sequence $(n_0, i_0) \overset{w_1}{\rightsquigarrow}_{P^\top} \dots \overset{w_k}{\rightsquigarrow}_{P^\top} (n_k, i_k)$ of steps in $\mathcal{S}_P$. The path is said alternating if, moreover, for any integer j between 1 and $k-1$, $w_j = \vee$ if and only if $w_{j+1} = \wedge$.

Some rules induce no direct flow of information. We say that a rule is trivial if it releases a bond between two sites in distinct agents without testing anything (except that the two sites are bound together). Thus a trivial rule is of the form $L \twoheadrightarrow R$ with $\mathcal{A}_L = \mathcal{A}_R = (\{n_A, n_B\}$, $\mathcal{S}_L = \mathcal{S}_R = \{(n_A, i_A), (n_B, i_B)\}$, $p\kappa_L = p\kappa_R = [\emptyset]$, $\mathcal{L}_L = \{((n_A, i_A), (n_B, i_B)), ((n_B, i_B), (n_A, i_A))\}$, and $\mathcal{L}_R = \{((n_A, i_A), \dashv), ((n_B, i_B), \dashv), (\dashv, (n_A, i_A)), (\dashv, (n_B, i_B))\}$ with $n_A, n_B \in \Sigma_{ag}$, $n_A \neq n_B$, $i_A \in \Sigma^{lnk}_{ag\text{-}st}(type(A))$ and $i_B \in \Sigma^{lnk}_{ag\text{-}st}(type(B))$.

We are now ready to give the constraints that has to be satisfied by an annotated summary graph so as to ensure that its induced change of variables is self-consistent.

Definition 14. *Suppose given a rule $r : L \overset{f}{\hookleftarrow} D \overset{g}{\hookrightarrow} R$, an annotated summary graph G^{a}. We say that G^{a} is compatible with the rule r if and only if the three following constraints are satisfied:*

1. **direct flow.** *if r is a non-trivial rule, for any homomorphism $\psi : L \to G$, any alternating path $p = (a_0, i_0) \overset{w_1}{\rightsquigarrow}_{L^\top} \dots \overset{w_k}{\rightsquigarrow}_{L^\top} (a_k, i_k)$ in the lhs L of the rule r such that the site (a_k, i_k) is modified by the rule r, and any integer j such that $0 \leq j < k$, $(\psi(a_j), i_j) \overset{w_j}{\rightsquigarrow}_{G^{\mathrm{a}}} (\psi(a_{j+1}), i_{j+1})$;*
2. **backward compatiblity.** *whatever r is trivial or not, for any site graph P, for any overlap $(S, \phi_1, \phi_2, \psi_1, h, X)$ between the site graph P and the site graph D, for any ground refinement $R_L \hookleftarrow R_D \hookrightarrow R_R$ of r defined by a triple of embeddings of the form $(h_l, h\psi_2, h_r)$, for any a homomorphism ϕ between R_R and G, for any homomorphism ψ between R_L and G, and for any site $(n, i) \in \mathcal{S}_S$ such that the state of the site $(\phi_2(n), i)$ is not specified in D (ie. the site $(f(\phi_2(n)), i)$ (resp. $(g(\phi_2(n)), i)$) is modified in L (resp. R), for any sequence $w_1, \dots w_j \in \{\vee, \wedge\}^k$ and any two integers j_0, j_1 such that:*
   - *$0 \leq j_0 \leq j_1 < k$,*
   - *$(\psi_1(n_{j_0}), i_{j_0}) = (\psi_2\phi_2(n), i)$,*
   - *$([\phi g' h \psi_1](n_j), i_j) \overset{w_{j+1}}{\rightsquigarrow}_{G^{\mathrm{a}}} ([\phi g' h \psi_1](n_{j+1}), i_{j+1})$ for any j such that $0 \leq j < j_1$,*
   - *$([\phi g' h \psi_1](n_{j+1}), i_{j+1}) \overset{w_{j+1}}{\rightsquigarrow}_{G^{\mathrm{a}}} ([\phi g' h \psi_1](n_j), i_j)$ for any j such that $j_1 \leq j \leq k$;*

   we have that:
   - *$([\psi f' h \psi_1](n_j), i_j) \overset{w_{j+1}}{\rightsquigarrow}_{G^{\mathrm{a}}} ([\psi f' h \psi_1](n_{j+1}), i_{j+1})$ for any j such that $0 \leq j < j_1$,*
   - *$([\psi f' h \psi_1](n_{j+1}), i_{j+1}) \overset{w_{j+1}}{\rightsquigarrow}_{G^{\mathrm{a}}} ([\psi f' h \psi_1](n_j), i_j)$ for any j such that $j_1 \leq j \leq k$.*

$$\begin{tikzcd}
 & & G & & \\
R_L \arrow[urr, bend left, "\psi"] \arrow[rrrr, -{Triangle[open]}] \arrow[dr, phantom, "\lrcorner", very near start] & & & & R_R \arrow[ull, bend right, "\phi"'] \arrow[dl, phantom, "\llcorner", very near start] \\
 & & R_D \arrow[ull, hook, "f'"'] \arrow[urr, hook', "g'"] & & \\
 & & X \arrow[u, hook, "h"] & & \\
L \arrow[uuu, hook] \arrow[rrrr, -{Triangle[open]}] & & & & R \arrow[uuu, hook'] \\
P \arrow[uurr, hook, "\psi_1"] & & D \arrow[uu, hook, "\psi_2"'] \arrow[ull, hook', "f"] \arrow[urr, hook, "g"'] & & \\
 & S \arrow[ul, hook', "\phi_1"] \arrow[ur, hook, "\phi_2"'] & & &
\end{tikzcd}$$

3. **cycle.** *Let $A, B \in \mathcal{A}$ be two agent types and $i_A \in \Sigma^{lnk}_{ag\text{-}st}(A)$ and $i_B \in \Sigma^{lnk}_{ag\text{-}st}(B)$ be two site identifiers. For any trivial rule r that removes bounds between the sites i_A in agents of type A and the sites i_B in agents of type B, if there exists a site $s \in \mathcal{S}_G$ and two agents n_A,$n_B \in \mathcal{A}_G$ such that : $type(n_A) = A$, $type(n_B) = B$ and two distinct paths $p = (n_A, i_A) \rightsquigarrow^*_{G^{\mathrm{a}}} s$ and $p' = (n_B, i_B) \rightsquigarrow^*_{G^{\mathrm{a}}} s$, then the rule r is considered to be not trivial, and the* direct flow *constraints (see Def. 14.(1)) must be applied also with it.*

The set of the annotations of a summary graph that are compatible with a set of rules forms a Moore family. Thus, it has a least element. Seeing each constraint instantiation as an upper closure operator, this least element is also the image of $G_\emptyset$ by the least upper bound in the lattice of the upper closure operators of this set of closure operators [25] and can be computed, whenever the summary graph G is finite, using asynchronous iterations [25,13] (the order in which the constraints are iterated do not matter). Yet, the least solution of the set of constraints depends on the choice of the sites which are chosen arbitrarily to be viewed as modified in the unmodified connected components. In practice, we first iterate the constraints in which these choices are not involved, and tune the choices afterwards, so as to minimise the annotation.

Now we assume that the annotation of the summary graph G is the least solution of the constraints in Def. 14 and that prefragments are defined as in Def. 10. Let us explain the constraints in Def. 14. Direct flow is obtained by taking any homomorphism between the lhs of a rule and the summary graph G and annotating the image of any alternating path between a site that is tested and a site that is modified. For instance, in the Σ-graph that is given in Fig. 1(c) the annotation of the right connected component describes the direct flow that is due to the rule that is given in Fig. 2(b). The existence of the homomorphism ensures the context-sensitivity of the analysis, since only the parts of G that match the lhs are annotated. As a consequence, we report no direct flow for the rule in Fig. 2(b) in the left connected component. Considering paths allows us to deal with distant control, that is to say agents which test the state of some sites, without modifying the states of any sites. Moreover, only considering alternating paths avoids spurious flows within agents. Formally, the direct flow constraints (see Def. 14.1) ensure the following property:

Property 1. For any overlap $(X, \psi_1, \psi_2, \phi_1, \phi_2, Y)$ between a connected component in the lhs of a non trivial rule and a prefragment such that there exists a site of the form $(\psi_1(n), i)$ that is modified in the rule, then the site graph Y is a prefragment as well.

In particular, since any connected component in the lhs of a rule contains a site that is modified, any connected component in the lhs of a rule is a prefragment.

Backward compatibility ensures that prefragments that overlap with the lhs of rules are always more refined than the ones that overlap with the rhs. This comes from the fact that a prefragment that overlaps with a rhs of a rule on a modified site $s^\bullet = (\phi_1(n_{j_0}), i_{j_0})$, contains at least one root r. Then any site s of the prefragment is reachable through the annotation by starting form the site $s^\bullet$, taking a path forward from $s^\bullet$ to the root r, and then a path backward from r to the site s. Thus we copy these paths at any place in G which matches with a potential antecedent of the pattern by the rule, which is ensured by the existence of the ground refinement $R_L \hookleftarrow R_D \hookrightarrow R_R$. For instance, the annotation of the right connected component in Fig. 1(c) has to be reported into the left connected component, due to the rule that is given in Fig. 2(a) and the ground refinement when last two sites are unphosphorylated (the relation between the 1st and the 3rd site can also be justified as a direct flow). Backward compatibility (see Def. 14.(2)) ensures the following property:

Property 2. For any overlap $(X, \psi_1, \psi_2, \phi_1, \phi_2, Y)$ between the rhs of a non trivial rule and a prefragment such that there exists a site of the form $(\psi_1(n), i)$ that is modified in the rule, if the connected component of the lhs of the right refinement of the rule by the embedding ϕ_1 is such that the number of connected components in the lhs is preserved, then any connected component in the lhs of the refined rule is a prefragment.

Backward compatibility is subject to abstraction. Instead of a ground refinement, one may look for a k-depth context around the site graph X. There will be more constraints to satisfy, but they will be easier to find.

When a trivial rule that breaks a bond between two sites is applied to a prefragment, it is crucial to express the concentration of this prefragment in which the two sites are actually bound together (since the rule performs two modifications simultaneously when it is applied with them). This is the purpose of constraints in Def. 14.(3).

Property 3. Suppose that there exists a trivial rule which breaks a bond between the site i_A of agents of type A and the site i_B of agents of type B. Then for any prefragment $\mathit{pf} = (\mathcal{A}, \mathit{type}, \mathcal{S}, \mathcal{L}, p\kappa)$ that contains two agents n_A and n_B such that $(n_A, i_A) \in \mathcal{S}$, $(n_B, i_B) \in \mathcal{S}$, $((n_A, i_A), (B, i_B)) \in \mathcal{L}$, $((n_B, i_B), (A, i_A)) \in \mathcal{L}$, and $(n_A, i_A) \neq (n_B, i_B)$ and for any agent $n^\bullet$ such that either $n^\bullet \in \mathcal{A}$ and $\mathit{type}(n^\bullet) = B$, or $n^\bullet \notin \mathcal{A}$, the site graph $(\mathcal{A} \cup \{n^\bullet\}, \mathit{type}[n^\bullet \to B], \mathcal{S} \cup \{(n^\bullet, i_B)\}, (\mathcal{L} \cup \mathcal{L}_+) \backslash \mathcal{L}_-, p\kappa)$ with $\mathcal{L}_+ = \{((n_A, i_A), (n_\bullet, i_B)), ((n_\bullet, i_B), (n_A, i_A))\}$, $\mathcal{L}_- = \{((n_\bullet, i_B), (A, i_A)), ((n_A, i_A), (B, i_B)), ((A, i_A), (n_\bullet, i_B)), ((B, i_B), (n_A, i_A))\}$, is a prefragment as well.

We may notice that the *cycle* constraints can be relaxed while still ensuring Prop. 3. But these are technical details that we skip for the sake of simplicity.

Properties 1,2, and 3 are enough to describe the evolution of the concentration of the prefragments by system of differential equations. We consider a set $\hat{\mathcal{F}}$ of prefragments, such that for any prefragment *pf* there exists a unique prefragment $\mathit{pf}' \in \hat{\mathcal{F}}$ such that $\mathit{pf} \approx \mathit{pf}'$.

Definition 15 (Consumption). *For any rule $L \multimap R$ @k, with L decomposed into connected components $c_1, \ldots, c_n$, and any overlap $(X, \psi_1, \psi_2, \phi_1, \phi_2, Y)$ between a connected component c_i and a prefragment $\mathit{pf} \in \hat{\mathcal{F}}$ such that $\psi_1(X)$ contains a site that is modified by the rule, then, the proper consumption term for pf due to this overlap can be expressed as follows:*

$$y'_{\mathit{pf}} \stackrel{\pm}{=} -k \cdot \frac{y_{\mathit{pf}} \cdot \prod_{1 \leq j \leq n, j \neq i} y_{c_j}}{[L, L]}.$$

Definition 16 (Production). *For any rule $r = L \multimap R$ @k and any overlap $(X, \psi_1, \psi_2, \phi_1, \phi_2, Y)$ between the right hand side R of the rule r and a prefragment $\mathit{pf} \in \hat{\mathcal{F}}$ such that $\psi_1(X)$ contains a site that is modified by the rule, we consider $L' \multimap R'$ @k the right refinement of r by the embedding ϕ_1. Then, the contribution is 0 whenever L and L' have not the same number of connected components, and is given as follows:*

$$y'_{\mathit{pf}} \stackrel{\pm}{=} k \cdot \frac{\prod_{1 \leq j \leq n} y_{c'_j}}{[L, L]}.$$

otherwise, where L' is decomposed into connected components $c'_1, \ldots, c'_n$.

We have skipped some technical details about the handling of trivial rules. Indeed, the overlap between a prefragment and the lhs of a trivial rule may not be a prefragment. Yet, in such a case, the obtained site graph is equivalent to a prefragment by removing a site and replacing a bound with the corresponding binding type. The same remark holds for production. We give more details in Appendix E.

The following theorem formalises the relation between the initial and the reduced system of differential equations:

Theorem 1. *We consider $x \in ([0,T) \to \mathbb{R})^{\mathcal{C}}$ the solution of the initial differential system with a given initial state $init \in (\mathbb{R}^{+})^{\mathcal{C}}$ and such that T is maximal and $y \in ([0,T') \to \mathbb{R})^{\hat{\mathcal{F}}}$ the solution of the reduced system with the initial state $init^{\sharp}$ that is defined as $init^{\sharp}_{\hat{f}} = \sum_{c \in \mathcal{C}} [\hat{f}, c] \cdot init_{\hat{f}}$ for any $\hat{f} \in \hat{\mathcal{F}}$ and such that T' is maximal. Then, $T = T'$, and for any prefragment $\hat{f} \in \hat{\mathcal{F}}$, at any time $t \in [0,T)$, $y_{\hat{f}}(t) = \sum_{c \in \mathcal{C}} [\hat{f}, c] \cdot x_{\hat{f}}(t)$.*

Thm.1 follows from the proof that can be found in [16] and which only requires the Properties 1, 2 and 3 to hold.

Some prefragments can be neutrally refined into a set of prefragments, which gives rises some numerical invariants. We call fragments the prefragments which cannot be refined this way. For any prefragment *pf* that is not a fragment, the variable y_{pf} can be eliminated of the system of equations, by replacing it a the corresponding linear combination of fragments.

6 Conclusion

We have introduced a parametric framework for coarse-graining the differential semantics of rule-based models. A summary graph is used to define which contexts are distinguished and allows us to tune the accuracy of our approximation of the flow of information between the sites of chemical complexes. The result of this analysis is used to detect useless correlations between the states of sites, which defines formally our coarse-graining.

As usual with partitioning techniques, the choice of the summary graph can be driven thanks to appropriate strategies. For instance, a transition system can be computed for each kind of site to abstract each qualitative behaviour. Then, we can choose to zoom in the accuracy of the analysis by distinguishing contexts according to the states of the sites the transition system of which is not strongly connected.

Our information flow analysis is based on the rules and can be degraded by neutral refinements [15,22]. In future works, we plan to study extrinsic notions of flow of information, based on the differential trajectories only, and relate them formally to our flow analyses. Besides, our framework is highly generic and we have focused on the formal foundations so far. In future works, we will address more practical issues: for instance we will define subsets of summary graphs, which make the computation of the set of coarse-grained variables easier.

References

1. M.L. Blinov, J.R. Faeder, B. Goldstein, and W. S. Hlavacek. Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291, 2004.
2. E. C. Bor-Yuh and X. Rival. Relational inductive shape analysis. In G. C. Necula and P. Wadler, editors, *POPL*, pages 247–260. ACM, 2008.
3. N. M. Borisov, A. S. Chistopolsky, J. R. Faeder, and B. N. Kholodenko. Domain-oriented reduction of rule-based network models. *IET Syst. Biol.*, 2, 2008.
4. N. M. Borisov, N. I. Markevich, B. N. Kholodenko, and E. Dieter Gilles. Signaling through receptors and scaffolds: Independent interactions reduce combinatorial complexity. *Biophysical Journal*, 89, 2005.
5. F. Bourdoncle. Abstract interpretation by dynamic partitioning. *J. Funct. Program.*, 2(4):407–423, 1992.
6. F. Camporesi and J. Feret. Formal reduction for rule-based models. In M. Mislove and J. Ouaknine, editors, *MFPS*, volume 276 of *ENTCS*, pages 29–59, Pittsburgh, USA, September 2011. Elsevier.
7. F. Camporesi, J. Feret, H. Koeppl, and T. Petrov. Combining Model Reductions. In M. Mislove and P. Selinger, editors, *MFPS*, volume 265 of *ENTCS*, pages 73–96, Ottawa, Canada, September 2010. Elsevier.
8. H. Conzelmann. *Mathematical Modeling of Cellular Signal Transduction Pathways — A Domain-Oriented Approach to Reduce Combinatorial Complexity.* PhD thesis, Institut für Systemdynamik des Universität Stuttgart, 2008.
9. H. Conzelmann, D. Fey, and E. D. Gilles. Exact model reduction of combinatorial reaction networks. *BMC Systems Biology*, 2, 2008.
10. H. Conzelmann, 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.
11. P. Cousot. *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, 21 March 1978.
12. P. Cousot and R. Cousot. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In R. M. Graham, M. A. Harrison, and R. Sethi, editors, *POPL*, pages 238–252. ACM, 1977.
13. P. Cousot and R. Cousot. Constructive versions of Tarski's fixed point theorems. *Pacific Journal of Mathematics*, 81(1):43–57, 1979.
14. V. Danos, J. Feret, W. Fontana, R. Harmer, J. Hayman, J. Krivine, C. Thompson-Walsh, and G. Winskel. Graphs, Rewriting and Pathway Reconstruction for Rule-Based Models. In D. D'Souza, J. Radhakrishnan, and K. Telikepalli, editors, *FSTTCS*, volume 18, Hyderabad, India, 2012. IARCS, LIPIcs.
15. V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling, symmetries, refinements. In J. Fisher, editor, *the 1st International Workshop, Formal Methods in Systems Biology - FMSB 2008*, volume 5054 of *LNCS*, pages 103–122, Cambridge, Royaume-Uni, 2008. Springer.
16. 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*, IEEE Computer Society, pages 362–381, Edinburgh, GB, 2010.
17. V. Danos and C. Laneve. Formal molecular biology. *TCS*, 325(1):69–110, 2004.

18. J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana. Internal coarse-graining of molecular systems. *PNAS*, 106(16), April 2009.
19. J. Feret, T. Henzinger, H. Koeppl, and T. Petrov. Lumpability Abstractions of Rule-based Systems. *TCS*, 431:137–164, 2012.
20. J. Feret, H. Koeppl, and T. Petrov. Stochastic fragments: A framework for the exact reduction of the stochastic semantics of rule-based models. *IJSI*. to appear.
21. R. Harmer, V. Danos, J. Feret, J. Krivine, and W. Fontana. Intrinsic Information carriers in combinatorial dynamical systems. *Chaos*, 20(3):037108, 2010.
22. E. Murphy, V. Danos, J. Feret, J. Krivine, and R. Harmer. Rule Based Modeling and Model Refinement. In H. Lodhi and S. Muggleton, editors, *Elements of Computational Systems Biology*, Wiley Book Series on Bioinformatics, pages 83–114. J. Wiley & Sons, 2010.
23. J. C. Reynolds. Separation logic: A logic for shared mutable data structures. In *LICS*, pages 55–74. IEEE Computer Society, 2002.
24. S. Sagiv, T.W. Reps, and R. Wilhelm. Parametric shape analysis via 3-valued logic. In A.W. Appel and A. Aiken, editors, *POPL*, pages 105–118. ACM, 1999.
25. M. Ward. The closure operators of a lattice. *Annals Math.*, 42:191–196, 1942.

We give the proofs for the main results of the paper.

A Notation

In the proofs, we use the following notation:

Definition 17. *Given a homorphism ϕ between two Σ-graphs G and H, and $(n,i) \in \mathcal{S}_G$ a site of G, we denote by $\phi(n,i)$ the site $(\phi(n),i)$ of H.*

B Proof for Prop. 1

Lemma 1. *Let G^{a} be an annotated summary graph. Let $\phi : P \to P'$ be a homomorphism between two Σ-graphs P and P'. Let $(n_1,i_1),(n_2,i_2) \in \mathcal{S}_P$ be two sites such that $(n_1,i_1) \rightsquigarrow^{\star}_{P,G^{\mathrm{a}}} (n_2,i_2)$ then $(\phi(n_1),i_2) \rightsquigarrow^{\star}_{P',G^{\mathrm{a}}} (\phi(n_2),i_2)$.*

Proof (Lemma 1). Let $(n,i),(n',i') \in \mathcal{S}_P$ be two sites in P and $w \in \{\vee,\wedge\}$ such that $(n,i) \overset{w}{\rightsquigarrow}_{P,G^{\mathrm{a}}} (n',i')$.

Let ψ be a homorphism between P' and G. By composition, $\psi\phi$ is a homorphism between P and G. Since $(n,i) \overset{w}{\rightsquigarrow}_{P,G^{\mathrm{a}}} (n',i')$, it follows, by Def. 9, that $(\psi(\phi(n)),i) \overset{w}{\rightsquigarrow}_{G^{\mathrm{a}}} (\psi(\phi(n')),i')$. Thus, by Def. 9, $(\phi(n),i) \overset{w}{\rightsquigarrow}_{P,G^{\mathrm{a}}} (\phi(n'),i')$.

It follows, by induction, that : $(\phi(n_1),i_2) \rightsquigarrow^{\star}_{P',G^{\mathrm{a}}} (\phi(n_2),i_2)$. □

Lemma 2. *Let G^{a} be an annotated summary graph. Let c_i be a connected component in the lhs of a rule r. We assume that the annotated summary graph is compatible with the rule r. Let $source_{c_i}, mod_{c_i} \in \mathcal{S}_{c_i}$ be two sites in c_i such that the site mod_{c_i} is modified in the rule r. Then, $source_{c_i} \rightsquigarrow^{\star}_{c_i,G^{\mathrm{a}}} mod_{c_i}$.*

Proof (Lemma 2). Since the sites source_{c_i} and mod_{c_i} are in the same connected components, there is an alternating path $(n_0, i_0)\overset{w_1}{\rightsquigarrow}_{c_i^\top} \dots \overset{w_k}{\rightsquigarrow}_{c_i^\top}(n_k, i_k)$ in c_i such that $(n_0, i_0) = \mathit{source}_{c_i}$ and $(n_k, i_k) = \mathit{mod}_{c_i}$.

Let ϕ be a homomorphism between c_i and G. Since, G^{a} is compatible with the rule r, by Def. 14.(1), it follows that $(\phi(n_0), i_0)\overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \dots \overset{w_k}{\rightsquigarrow}_{G^{\mathrm{a}}}(\phi(n_k), i_k)$.

Thus, by Def. 9, it follows that: $(n_0, i_0)\overset{w_1}{\rightsquigarrow}_{c_i,G^{\mathrm{a}}} \dots \overset{w_k}{\rightsquigarrow}_{c_i,G^{\mathrm{a}}}(n_k, i_k)$, which proves that: $\mathit{source}_{c_i} \rightsquigarrow^{\star}_{c_i,G^{\mathrm{a}}} \mathit{mod}_{c_i}$. □

Proof (Prop. 1). Let G^{a} be an annotated summary graph which is compatible with a non trivial rule $r \;:\; L \mathrel{-\!\!\triangleright} R$. Let $(X, \psi_1, \psi_2, \phi_1, \phi_2, Y)$ be an overlap between a connected component c_i in L and a prefragment F such that there exists a site $\mathit{mod}_X \in \mathcal{S}_X$ such that $\psi_1(\mathit{mod}_X) \in \mathcal{S}_{c_i}$ is modified in the rule r.

Let us prove that Y is a prefragment.

F is a prefragment. So, by Def. 10, there exists a site $\mathit{root}_F \in \mathcal{S}_F$ such that for any other site $\mathit{source}_F \in \mathcal{S}_F$, $\mathit{source}_F \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$.

$$\begin{tikzcd}
& G & \\
& Y \arrow[u, "\phi"] \arrow[d, phantom, "\vee"] & \\
c_i \arrow[ur, hook, "\phi_1"] & & F \arrow[ul, hook', "\phi_2"'] \\
& X \arrow[ul, hook, "\psi_1"] \arrow[ur, hook', "\psi_2"'] &
\end{tikzcd}$$

Let us show that for any site $\mathit{source}_Y \in \mathcal{S}_Y$, we have: $\mathit{source}_Y \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$.

Let $\mathit{source}_Y \in \mathcal{S}_Y$.

- If there exists a site $\mathit{source}_F \in \mathcal{S}_F$ in F such that $\phi_2(\mathit{source}_F) = \mathit{source}_Y$, we can conclude, by Def. 10, that $\mathit{source}_F \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$. By Lemma 1, it follows that: $\phi_2(\mathit{source}_F) \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$. Thus, since $\mathit{source}_Y = \phi_2(\mathit{source}_F)$, $\mathit{source}_Y \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$.

  In particular, $\phi_2(\psi_2(\mathit{mod}_X)) \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$.

- Otherwise, we know that the cospan $c_i \overset{\phi_1}{\hookrightarrow} Y \overset{\phi_2}{\hookleftarrow} F$ is a pushout of the span $c_i \overset{\psi_1}{\hookleftarrow} X \overset{\psi_2}{\hookrightarrow} F$, so there exists a site $\mathit{source}_{c_i} \in \mathcal{S}_{c_i}$ such that $\phi_1(\mathit{source}_{c_i}) = \mathit{source}_Y$. By Lemma 2, since the site $\psi_1(\mathit{mod}_X)$ is modified by the rule r, it follows that: $\mathit{source}_{c_i} \rightsquigarrow^{\star}_{c_i,G^{\mathrm{a}}} \psi_1(\mathit{mod}_X)$. By Lemma 1, we get that: $\phi_1(\mathit{source}_{c_i}) \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_1(\psi_1(\mathit{mod}_X))$. Since $\phi_1\psi_1 = \phi_2\psi_2$ and $\phi_1(\mathit{source}_{c_i}) = \mathit{source}_Y$, we get that: $\mathit{source}_Y \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\psi_2(\mathit{mod}_X))$. Yet, we have already proved, in the previous case, that: $\phi_2(\psi_2(\mathit{mod}_X)) \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$. By composition, it follows that: $\mathit{source}_Y \rightsquigarrow^{\star}_{Y,G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$.

So Y is a prefragment. □

C Proof for Prop. 2

Let r be a non trivial rule $L \overset{f}{\hookleftarrow} D \overset{g}{\hookrightarrow} R$ and let $(X, \psi_1, \psi_2, h_r, \phi_2, R')$ be an overlap between the right hand side R of the rule r and a prefragment $\mathit{pf} \in \hat{\mathcal{F}}$. We assume that there exists a site $\mathit{mod}_X \in \mathcal{S}_X$ such that $\psi_1(\mathit{mod}_X)$ is modified in the rule r.

We consider $r' = L' \overset{f'}{\hookleftarrow} D' \overset{g'}{\hookrightarrow} R'$ the right refinement of r by the embedding h_r and we denote by h_l the embedding between L and L' in this refinement

$$\begin{tikzcd}
 & G & & & \\
R_L \arrow[ur, "\phi_{R_L}"] \arrow[rrr, -{Triangle[open]}] & & & R_R \arrow[ull, bend right, "\psi"'] \arrow[dl, phantom, "\llcorner", very near start] & \\
 & R_D \arrow[ul, hook'] \arrow[urr, hook] & & & \\
L' \arrow[uuur, bend left, "\phi_{L'}"] \arrow[uu, hook, "h'_l"] \arrow[rrr, -{Triangle[open]}] & & & R' \arrow[uu, hook', "h'_r"'] \arrow[dl, phantom, "\llcorner", very near start] \arrow[dr, phantom, "\lrcorner", very near start] & F \arrow[l, hook', "\phi_2"] \\
 & D' \arrow[uu, hook, "h'_c"] \arrow[ul, hook', "f'"] \arrow[urr, hook, "g'"'] & & & \\
 & Y \arrow[u, hook, "h''_c"'] \arrow[dr, phantom, "\lrcorner", very near start] & P \arrow[l, hook', "\phi_4"] & & \\
L \arrow[uuuuuur, bend left=60, "\phi_L"] \arrow[uuu, hook, "h_l"] \arrow[rrr, -{Triangle[open]}] & & & R \arrow[uuu, hook', "h_r"'] & X \arrow[l, hook', "\psi_1"'] \arrow[uuu, hook', "\psi_2"] \arrow[ul, phantom, "\ulcorner", very near start] \\
 & D \arrow[ul, hook', "f"] \arrow[urr, hook, "g"'] \arrow[uuu, hook, "h_c"] \arrow[uu, hook', "\phi_3"'] & S \arrow[l, hook', "\psi_3"] \arrow[uu, hook', "\psi_4"'] \arrow[urr, hook] & &
\end{tikzcd}$$

Fig. 3. Commutating diagram for Appendix C.

and by h_c the embedding between D and D'. We assumes that L and L' have the same number of connected components.

The site graph F is a prefragment. We denote by $root_F \in \mathcal{S}_F$ a site such that for any site $source_F \in \mathcal{S}_F$, $source_F \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} root_[F]$.

Lemma 3. *Let $mod'_{L'}, mod''_{L'} \in \mathcal{S}_{L'}$ be two sites of the site graph L' that are modified by the rule r' and that belong to the same connected component in L'. Then, $mod'_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} mod''_{L'}$.*

Proof. Since the sites $mod'_{L'}$ and $mod''_{L'}$ are modified in the rule r' and that a refinement only add tested sites. There exists sites $mod'_L \in \mathcal{S}_L$ and $mod''_L \in \mathcal{S}_L$ such that $h_l(mod'_L) = mod'_{L'}$, $h_l(mod''_L) = mod''_{L'}$ and that both sites mod'_L and mod''_L are modified by the rule r. Since L and L' have the same number of connected components and h_l is an embedding between L and L', then necessarily, the sites mod'_L and mod''_L belong to the same connected component in L. It follows that there exists an alternating path $(n_0, i_0)\overset{w_1}{\rightsquigarrow}_{L^{\top}} \ldots \overset{w_k}{\rightsquigarrow}_{L^{\top}}(n_k, i_k)$ in L such that $(n_0, i_0) = mod'_L$ and $(n_k, i_k) = mod''_L$.

Let ϕ_L be a homomorphism between L and G. By Def. 14.(1), we have: $(\phi_L(n_0), i_0)\overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{G^{\mathrm{a}}}(\phi_L(n_k), i_k)$.

Thus, by Def. 9, we have: $(n_0, i_0)\overset{w_1}{\rightsquigarrow}_{L,G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{L,G^{\mathrm{a}}}(n_k, i_k)$. It follows that: $mod'_L \rightsquigarrow^{\star}_{L,G^{\mathrm{a}}} mod''_L$. We know that h_L is a homomorphism between L and L', thus, by lemma 1, it follows that $h_L(mod'_L) \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} h_L(mod''_L)$. Since $h_l(mod'_L) = mod'_{L'}$ and $h_l(mod''_L) = mod''_{L'}$, we get that: $mod'_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} mod''_{L'}$. □

Lemma 4. *If there exists a site $root_{D'} \in \mathcal{S}_{D'}$ such that $g'(root_{D'}) = \phi_2(root_F)$. Then, there exists a site $mod'_L \in \mathcal{S}_L$ that is modified by the rule r and such that $h_l(mod'_L) \rightsquigarrow^{\star}_{L,G^{\mathrm{a}}} f'(root_{D'})$.*

Proof. We assume that there exists a site $root_{D'} \in \mathcal{S}_{D'}$ in the site graph D' such that $g'(root_{D'}) = \phi_2(root_F)$.

By assumption the site $mod_X \in \mathcal{S}_X$ is such that the site $\psi_1(mod_X)$ is modified in the rule r. Thus, since the rule r' is a refinement of the rule r, the site $h_r(\psi_1(mod_X))$ is modified in the rule r'.

Since F is a fragment, there exists a path $p' = \psi_2(mod_X) \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} root_F$. We consider the longuest suffix $s'_0 \overset{w_1}{\rightsquigarrow}_{F^{\top}} \dots \overset{w_k}{\rightsquigarrow}_{F^{\top}} s'_j$ of p' such that for any l between 0 and j, there exists $s_l \in \mathcal{S}_{D'}$, such that for any integer l such that $0 \leq l \leq j$, $g'(s_l) = \phi_2(s'_l)$ and such that for any integer l such that $0 \leq l < j$ and $w_{l+1} = \wedge$, we have: $(s_l, s_{l+1}) \in \mathcal{L}_{D'}$. We denote $s_l = (n_l, i_l)$. for any integer l such that $0 \leq l \leq j$. Necessarily, the sites $g'(s_0)$ is modified in the rule r'. We denote $mod'_{R'}$ the site $g'(s_0)$. Since the rule r' is a refinement of r, there exists a site $mod'_R \in \mathcal{S}_R$ such that $h_r(mod'_R) = mod'_{R'}$ and that the site mod'_R is modified in the rule r.

Moreover, since the span $R \xhookleftarrow{\psi_1} X \xhookrightarrow{\psi_2} F$ is a pullback and $h_r(mod_{'}[R]) = \phi_2(s_0)$, there exists a site $mod'_X \in \mathcal{S}_X$ such that $\psi_1(mod'_X) = mod'_R$ and $\psi_2(mod'_X) = s_0$.

We consider the site graphs S and P which are defined as follows:

$$S = (\{n_0\}, [n_0 \to type_{D'}(n_0)], \{s_0\}, \emptyset, [\emptyset])$$

$$P = (\mathcal{A}_P, type_P, \mathcal{S}_P, \mathcal{L}_P, p\kappa_P)$$

with:
- $\mathcal{A}_P = \{n_l \mid 0 \leq l \leq j\}$;
- $type_P = [n_l \mapsto type_{D'}(n_l)]$;
- $\mathcal{S}_P = \{s_l \mid 0 \leq l \leq j\}$;
- $\mathcal{L}_P = \left\{ (s_l, s_{l+\varepsilon}) \;\middle|\; \begin{array}{l} 0 \leq l \leq j, \varepsilon \in \{-1, +1\}, \\ 0 \leq l + \varepsilon \leq j, w''_{\max(l,l+\varepsilon)} = \wedge \end{array} \right\}$;
- $p\kappa_P = [\emptyset]$.

By construction, the span $D \xhookleftarrow{\psi_3} S \xhookrightarrow{\psi_4} P$ such that $g(\psi_3(s_0)) = mod'_R$ and $\psi_4(s_0) = s_0$ has a relative pushout $D \xhookleftarrow{\phi_3} Y \xhookrightarrow{\phi_4} P$. Moreover, there exists an embedding h''_c between Y and D' such that: $h_c = h''_c \phi_3$.

Let us prove that $f'(s_0) \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(root_{D'})$.

Let $\phi_{L'}$ be a homomorphism between L' and G. Since G is a summary graph, by Def. 2, there exists an embedding h'_l between L' and a tuple R_L of chemical complexes in $\mathcal{C}$, and an homorphism ϕ_{R_L} between R_L and G such that $\phi_{L'} = \phi_{R_L} h'_l$. We denote by $R_L \hookleftarrow R_D \hookrightarrow R_R$ the left-refinement of r' through the embedding h'_l and by h'_R the corresponding embedding between R' and R_R.

By construction, $h''_c(\phi_4(s_0)) = s_0$. By rigidity, it follows that for any l such that $0 \leq l \leq j$, $h''_c(\phi_4(s_l) = s_l$. We already know that: $s'_0 \overset{w_1}{\rightsquigarrow}_{F,G^{\mathrm{a}}} \ldots \overset{w_j}{\rightsquigarrow}_{F,G^{\mathrm{a}}} s'_j$. Thus, by Lemma 1, it follows that: $\phi_2(s'_0) \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_j}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \phi_2(s'_j)$. By assumption, for any l such that $1 \leq l \leq j$, we have: $g'(s_l) = \phi_2(s'_l)$. It follows that: $[g'h''_c\phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_j}{\rightsquigarrow}_{R',G^{\mathrm{a}}} [g'h''_c\phi_4](s_j)$.

Since G is a summary graph, by Def. 2, there exists an homomorphism ψ between R_R and G. By Def. 9, since $\psi h'_r$ is a homomorphism between R' and G, it follows that $[\psi h'_r g' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_j}{\rightsquigarrow}_{G^{\mathrm{a}}} [\psi h'_r g' h''_c \phi_4](s_j)$.

Thus, since the site $\psi_1(\mathit{mod}'_X)$ is modified in the rule r and that, $\psi_1(\mathit{mod}'_X) = g(\psi_3(s_0))$, Def. 14.(2) applies (see Fig. 3).

It follows that: $[\phi_{R_L} h'_l f' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}}, \ldots \overset{w_j}{\rightsquigarrow}_{G^{\mathrm{a}}}, [\phi_{R_L} h'_l f' h''_c \phi_4](s_j)$.

Thus, since $\phi_{L'} = \phi_{R_L} h'_l$, $[\phi_{L'} f' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}}, \ldots \overset{w_j}{\rightsquigarrow}_{G^{\mathrm{a}}}, [\phi_{L'} f' h''_c \phi_4](s_j)$.

So $[f'h''_c\phi_4](s_0) \rightsquigarrow_{L',G^{\mathrm{a}}} w_1 \ldots \overset{w_j}{\rightsquigarrow}_{L',G^{\mathrm{a}}} [\phi_{L'} f' h''_c \phi_4](s_j)$, with $g'(h''_c(\phi_4(s_j))) = \phi_2(\mathit{root}_F)$ and $g'(h''_c(\phi_4(s_0))) = h_r(\mathit{mod}'_R)$ with mod'_R a site that is modified in the rule r. □

Lemma 5. *Let $\mathit{source}_{L'} \in \mathcal{S}_{L'}$ be a site of L'.*

At least one of the following properties holds:

1. *there exists a site $\mathit{mod}'_{L'} \in \mathcal{S}_{L'}$ such that the site $\mathit{mod}'_{L'}$ is modified by the rule r', $\mathit{source}_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}'_{L'}$;*
2. *there exists a site $\mathit{root}_D \in \mathcal{S}_D$ such that $\phi_2(\mathit{root}_F) = g'(h_c(\mathit{root}_D))$ and $\mathit{source}_{L'} \rightsquigarrow^{\star}_{F',G^{\mathrm{a}}} f'(h_c(\mathit{root}_D))$.*

Proof.

- If there exists a site $\mathit{source}_L \in \mathcal{S}_L$ such that $h_l(\mathit{source}_L) = \mathit{source}_{L'}$, then, let $\mathit{mod}'_L \in \mathcal{S}_L$ be a site that is modified by the rule r, and that belongs to the same connected component of L as source_L. By Lemma 2, $\mathit{source}_L \rightsquigarrow^{\star}_{L,G^{\mathrm{a}}} \mathit{mod}'_L$. Since the site mod'_L is modified in the rule r, the site $h_l(\mathit{mod}'_L)$ is modified in the rule r', and moreover, since $\mathit{source}_L \rightsquigarrow^{\star}_{L,G^{\mathrm{a}}} \mathit{mod}'_L$, we get, by Lemma 1, that: $h_l(\mathit{source}_L) \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} h_l(\mathit{mod}'_L)$.
- Otherwise, there exist two sites $\mathit{source}_{D'} \in \mathcal{S}_{D'}$ and $\mathit{source}_F \in \mathcal{S}_F$ such that $f'(\mathit{source}_{D'}) = \mathit{source}_L$ and $\phi_2(\mathit{source}_F) = g'(\mathit{source}_{D'})$.
  By Def. 10, $\mathit{source}_F \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$.
  By Lemma 1, it follows that $\phi_2(\mathit{source}_F) \rightsquigarrow^{\star}_{R',G^{\mathrm{a}}} \phi_2(\mathit{root}_F)$.

  Let us take a path $p_{R'} = s'_0 \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{R',G^{\mathrm{a}}} s'_k$ such that $s_0 = \phi_2(\mathit{source}_F)$ and $s_k = \phi_2(\mathit{root}_F)$. We consider the set J of the integers j between 1 and k such that $w_k = \wedge$ and such that for any pair of sites $s_A, s_B \in \mathcal{S}_{D'}$ $g'(s_A) \neq s'_{j-1}$, or $g'(s_B) \neq s'_j$, or $(s_A, s_B) \notin \mathcal{L}_{D'}$ (intuitively, $j \in J$ iff the $j - th$ step of $p_{R'}$ passes by a bond that have been created by the rule r').
  - If the set J is not empty, we consider j its minimum element. By construction, there exists a tuple of sites $(s_l)_{0 \leq l < j} \in \mathcal{S}^j_{D'}$ such that for any integer l which satisfies $0 \leq l < j$, $g'(s_l) = s'_l$ and such that the path $p_{D'} = s_0 \overset{w_1}{\rightsquigarrow}_{D'^{\top}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{D'^{\top}} s_{j-1}$ is well-defined, and such that the site $g'(s_{j-1})$ is modified in the rule r'. Let us write $\mathit{mod}'_{D'} = s_{j-1}$. Thus,

there exists a site mod'_D such that $h_c(\mathit{mod}'_D) = \mathit{mod}'_{D'}$ and the site $g(\mathit{mod}'_D)$ is modified in the rule r.
For any l such that $0 \leq l < j$, we write $(m_l, i_l) = s_l$. We consider the site graphs S and P which are defined as follows:

$$S = (\{m_{j-1}\}, [m_{j-1} \rightarrow \mathit{type}_{D'}(m_{j-1})], \{\mathit{mod}'_{D'}\}, \emptyset, [\emptyset])$$

$$P = (\{m_l \mid 0 \leq l < j\}, [m_l \rightarrow \mathit{type}_{D'}(m_l)], \{s_l \mid 0 \leq l < j\}, \mathcal{L}_P, [\emptyset])$$

with $\mathcal{L}_P = \left\{ (s_l, s_{l+\varepsilon}) \;\middle|\; \begin{array}{l} 0 \leq l < j, \varepsilon \in \{-1, +1\}, \\ 0 \leq l + \varepsilon < j, w_{\max(l,l+\varepsilon)} = \text{‘} \wedge \text{‘} \end{array} \right\}$.

By construction, the span $D \overset{\psi_3}{\hookleftarrow} S \overset{\psi_4}{\hookrightarrow} P$ such that $\psi_3(\mathit{mod}'_{D'}) = \mathit{mod}'_D$ and $\psi_4(\mathit{mod}'_{D'}) = s_{j-1}$, has a relative pushout $D \overset{\phi_3}{\hookleftarrow} Y \overset{\phi_4}{\hookrightarrow} P$. Moreover, there exists an embedding h''_c between Y and D' such that: $h_c = h''_c \phi_3$.
Let us show that $\mathit{source}_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} [f' h''_c \phi_4](s_{j-1})$.
Let $\phi_{L'}$ be a homomorphism between L' and G. Since G is a summary graph, by Def. 2, there exists an embedding h'_l between L' and a tuple R_L of chemical complexes in $\mathcal{C}$, and an homomorphism ϕ_{R_L} between R_L and G such that $\phi_{L'} = \phi_{R_L} h'_l$. We denote by $R_L \hookleftarrow R_D \hookrightarrow R_R$ the left-refinement of r' through the embedding h'_l and by h'_r the corresponding embedding between R' and R_R.
By construction, $g' h''_c \phi_4(s_{j-1}) = s'_{j-1}$. Thus by rigidity, for any l such that $0 \leq l < j$, $[g' h''_c \phi_4](s_l) = s'_l$. Since G is a summary graph, by Def. 2, there exists an homomorphism ψ between R_R and G. We know that: $s'_0 \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{R',G^{\mathrm{a}}} s'_{j-1}$. By Def. 9, since $\psi h'_r$ is a homomorphism between R' and G, it follows that $\psi(h'_r(s'_0)) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}} \psi(h'_r(s'_{j-1}))$.
Thus, $[\psi h'_r g' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}} [\psi h'_r g' h''_c \phi_4](s_{j-1})$.
Moreover, the site $g(\phi_3(s_{j-1}))$ is modified in the rule r. By Def. 14.(2) (see Fig. 3), $[\phi_{R_L} h'_l f' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}} [\phi_{R_L} h'_l f' h''_c \phi_4](s_{j-1})$.
Thus, $[\phi_{L'} f' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}} [\phi_{L'} f' h''_c \phi_4](s_{j-1})$.
So $[f' h''_c \phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{L',G^{\mathrm{a}}} \ldots \overset{w_{j-1}}{\rightsquigarrow}_{L',G^{\mathrm{a}}} [\phi_{L'} f' h''_c \phi_4](s_{j-1})$.
Then, since $f'(h''_c(\phi_4(s_0))) = \mathit{source}_{L'}$ and $h''_c(\phi_4(s_{j-1})) = \mathit{mod}'_{D'}$, we can conclude that: $\mathit{source}_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(\mathit{mod}'_{D'})$.

- Otherwise.
  By construction, there exists $(s_l)_{0 \leq l \leq k} \in \mathcal{S}^j_{D'}$ such that for any integer l which satisfies $0 \leq l \leq k$, $g'(s_l) = s'_l$ and such that the path $p_{D'} = s_0 \overset{w_1}{\rightsquigarrow}_{D'^{\top}} \ldots \overset{w_k}{\rightsquigarrow}_{D'^{\top}} s_k$ is well-defined, and such that $f'(s_0) = \mathit{source}_{L'}$ and $g'(s_k) = \phi_2(\mathit{root}_F)$.
  By assumption the site $\mathit{mod}_X \in \mathcal{S}_X$ is such that the site $\psi_1(\mathit{mod}_X)$ is modified in the rule r. Since F is a fragment, there exists a path $p' = \psi_2(\mathit{mod}_X) \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$.

  Let us consider the longuest suffix $s'''_0 \overset{w''_1}{\rightsquigarrow}_{F^{\top}} \ldots \overset{w''_k}{\rightsquigarrow}_{F^{\top}} s'''_j$ of p' such that for any l between 0 and j, there exists $s''_l \in \mathcal{S}_{D'}$, such that for any integer

l such that $0 \leq l \leq j$, $g'(s''_l) = \phi_2(s'''_l)$, and such that for any integer l such that $0 \leq l < j$, $(s''_l, s''_{l+1}) \in \mathcal{L}_{D'}$. Necessarily, the site $g'(s''_0)$ is modified in the rule r'. We write $(n''_l, i''_l) = s''_l$ for any integer l such that $0 \leq l \leq j$ and $(n_l, i_l) = s_l$ for any integer l such that $0 \leq l \leq k$.
We consider the site graphs S and P which are defined as follows:

$$S = (\{n''_0\}, [n''_0 \to \mathit{type}_{D'}(n''_0)], \{s''_0\}, \emptyset, [\emptyset])$$

$$P = (\mathcal{A}_P, \mathit{type}_P, \mathcal{S}_P, \mathcal{L}_P, p\kappa_P)$$

with:

$*$ $\mathcal{A}_P = \{n_l \mid 0 \leq l \leq k\} \cup \{n''_l \mid 0 \leq l \leq j\}$;

$*$ $\mathit{type}_P = [n \in \mathcal{A}_P \mapsto \mathit{type}_{D'}(n)]$;

$*$ $\mathcal{S}_P = \{s_l \mid 0 \leq l \leq k\} \cup \{s''_l \mid 0 \leq l \leq j\}$;

$*$ $\mathcal{L}_P = \begin{array}{l} \left\{ (s_l, s_{l+\varepsilon}) \;\middle|\; \begin{array}{l} 0 \leq l \leq k, \varepsilon \in \{-1,+1\}, \\ 0 \leq l+\varepsilon \leq k, w_{\max(l,l+\varepsilon)} = \text{`} \wedge \text{`} \end{array} \right\} \\ \cup \left\{ (s''_l, s''_{l+\varepsilon}) \;\middle|\; \begin{array}{l} 0 \leq l \leq j, \varepsilon \in \{-1,+1\}, \\ 0 \leq l+\varepsilon \leq j, w''_{\max(l,l+\varepsilon)} = \text{`} \wedge \text{`} \end{array} \right\} \end{array}$;

$*$ $p\kappa_P = [\emptyset]$.

By construction, the span $D \overset{\psi_3}{\hookleftarrow} S \overset{\psi_4}{\hookrightarrow} P$ such that $\psi_3(s''_0) = s''_j$ and $\psi_4(s''_0) = s''_j$ has a relative pushout $D \overset{\phi_3}{\hookleftarrow} Y \overset{\phi_4}{\hookrightarrow} P$. Moreover, there exists an embedding h''_c between Y and D' such that: $h_c = h''_c\phi_3$.
Let us show that $\mathit{source}_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(h''_c(\phi_4(s''_j)))$.
Let $\phi_{L'}$ be a homomorphism between L' and G. Since G is a summary graph, by Def. 2, there exists an embedding h'_l between L' and a tuple R_L of chemical complexes in $\mathcal{C}$, and an homorphism ϕ_{R_L} between R_L and G such that $\phi_{L'} = \phi_{R_L}h'_l$. We denote by $R_L \hookleftarrow R_D \hookrightarrow R_R$ the left-refinement of r' through the embedding h'_l and by h'_R the corresponding embedding between R' and R_R.
By assumption, for any l such that $0 \leq l \leq k$, $[g'h''_c\phi_4](s_l) = s'_l$. We have already proven that $s'_0 \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{R',G^{\mathrm{a}}} s'_k$. It follows that: $[g'h''_c\phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{R',G^{\mathrm{a}}} [g'h''_c\phi_4](s_k)$.
By construction, for any l such that $0 \leq l \leq j$, $[h''_c\phi_4](s_l) = s_l$. Moreover, we know hat $s''_0 \overset{w''_1}{\rightsquigarrow}_{D',G^{\mathrm{a}}} \ldots \overset{w''_j}{\rightsquigarrow}_{D',G^{\mathrm{a}}} s''_j$. Thus, by Lemma 1, since g' is a homorphism, it follows that: $[g'h''_c\phi_4](s''_0) \overset{w''_1}{\rightsquigarrow}_{R',G^{\mathrm{a}}} \ldots \overset{w''_j}{\rightsquigarrow}_{R',G^{\mathrm{a}}} [g'h''_c\phi_4](s''_j)$.
Since G is a summary graph, by Def. 2, there exists an homomorphism ψ between R_R and G. By Def. 9, since $\psi h'_r$ is a homomorphism between R' and G, it follows that $[\psi h'_r g' h''_c\phi_4](s_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_k}{\rightsquigarrow}_{G^{\mathrm{a}}} [\psi h'_r g' h''_c\phi_4](s_k)$ and that $[\psi h'_r g' h''_c\phi_4](s'') \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}} \ldots \overset{w_j}{\rightsquigarrow}_{G^{\mathrm{a}}} [\psi h'_r g' h''_c\phi_4](s''_j)$.
Moreover, $[\psi h'_r g' h''_c\phi_4](s_k) = [\psi h'_r\phi_2](\mathit{root}_F) = [\psi h'_r g' h''_c\phi_4](s''_j)$ and the site $[g\phi_3](s''_0)$ is modified in r (since r' is a refinement and $[g'h_c\psi_3](s''_0)$ is modified in r'.
Thus, we can apply Def. 14.(2) (see Fig. 3). We can conclude that: $[\phi_{R_L}h'_l f' h''_c\phi_4](s''_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}}, \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}}, [\phi_{R_L}h'_l f' h''_c\phi_4](s''_j)$.

Thus, $[\phi_{L'} f' h''_c \phi_4](s''_0) \overset{w_1}{\rightsquigarrow}_{G^{\mathrm{a}}}, \ldots \overset{w_{j-1}}{\rightsquigarrow}_{G^{\mathrm{a}}}, [\phi_{L'} f' h''_c \phi_4](s''_j)$.
So $[f' h''_c \phi_4](s''_0) \rightsquigarrow_{L',G^{\mathrm{a}}} w_1 \ldots \overset{w_{j-1}}{\rightsquigarrow}_{L',G^{\mathrm{a}}} [\phi_{L'} f' h''_c \phi_4](s''_j)$.
Then, since $f'(h''_c(\phi_4(s''_0))) = \mathit{source}_{L'}$ and $h''_c(\phi_4(s''_j)) = s''_j$ we can conclude that: $\mathit{source}_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(s''_j)$. We already know that $g'(s''_j) = \phi_2(\mathit{root}_F)$.

□

Proof (Prop. 2). Let us consider c'_i a connected component in L'.

- If there exists a site $\mathit{root}_{D'} \in \mathcal{S}_{D'}$ such that $g'(\mathit{root}_{D'}) = \mathit{root}_F$ and $f'(\mathit{root}_{D'}) \in \mathcal{S}_{c'_i}$.
  Let $\mathit{source}_{c'_i}$ be a site in $\mathcal{S}_{c'_i}$.
  Let us assume that $\mathit{source}_{c'_i} \not\rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(\mathit{root}_{D'})$. By Lemma 5, there exists a site $\mathit{mod}'_{L'} \in \mathcal{S}_{L'}$ that is modified by r' and such that $\mathit{source}_{c'_i} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}'_{L'}$. By Lemma 4, there exists a site $\mathit{mod}''_{L'} \in \mathcal{S}_{L'}$ that is modified in L' and such that $\mathit{mod}''_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(\mathit{root}_{D'})$. Since $\mathit{mod}'_{L'}$ and $\mathit{mod}''_{L'}$ are two sites of c'_i that are modified by the rule r', thus by Lemma 2, $\mathit{mod}'_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}''_{L'}$. By composition, we get that $\mathit{source}_{c'_i} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} f'(\mathit{root}_{D'})$, which is absurd.
  Thus, $(n,i) \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} (f'(n'), i')$.
  So C'_i is a prefragment.
- Otherwise, let $\mathit{mod}'_{L'} in \mathcal{S}_{L'}$ be a site which is modified by the rule r'.
  Let $\mathit{source}_{c'_i}$ be a site in $\mathcal{S}_{c'_i}$. By Lemma 5, there exists a site $\mathit{mod}''_{L'} \in \mathcal{S}_{L'}$ that is modified by r' and such that $\mathit{source}_{c'_i} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}''_{L'}$. Since $\mathit{mod}'_{L'}$ and $\mathit{mod}''_{L'}$ are two sites of c'_i that are modified by the rule r', by Lemma 2, $\mathit{mod}''_{L'} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}'_{L'}$. By composition, we get that $\mathit{source}_{c'_i} \rightsquigarrow^{\star}_{L',G^{\mathrm{a}}} \mathit{mod}'_{L'}$.
  So c'_i is a prefragment.

In both cases, c'_i is a prefragment. □

D Proof for Prop. 3

Proof. Suppose that there is a trivial rule which breaks a bond between the sites i_A of agents of type A and the sites i_B of agents of type B. Let $F = (\mathcal{A}_F, \mathit{type}_F, \mathcal{S}_F, \mathcal{L}_F, p\kappa_F)$ be a prefragment that contains two agents n_A and n_B such that we have: $(n_A, i_A) \in \mathcal{S}_F$, $(n_B, i_B) \in \mathcal{S}_F$, $((n_A, i_A), (B, i_B)) \in \mathcal{L}_F$, $((n_B, i_B), (A, i_A)) \in \mathcal{L}_F$, and $(n_A, i_A) \neq (n_B, i_B)$. Let n' be an agent such that either $n' \in \mathcal{A}_F$ and $\mathit{type}_F(n') = B$, or $n' \notin \mathcal{A}_F$. The site graph P that is defined as $(\mathcal{A}_F \cup \{n'\}, \mathcal{S}_F \cup \{(n', i_B)\}, (\mathcal{L}_F \cup \mathcal{L}_+) \setminus \mathcal{L}_-, p\kappa_F)$, where the set $\mathcal{L}_-$ is equal to $\{((n', i_B), (A, i_A)), ((n_A, i_A), (B, i_B)), ((A, i_A), (n', i_B)), ((B, i_B), (n_A, i_A))\}$ and the set $\mathcal{L}_+$ is equal to $\{((n_A, i_A), (n', i_B)), ((n', i_B), (n_A, i_A))\}$.

Let $\mathit{root}_F \in \mathcal{S}_F$ be a site such that for any $\mathit{source}_F \in \mathcal{S}_F$, $\mathit{source}_F \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$. Let us show that for any site $\mathit{source}_P \in \mathcal{S}_P$, $\mathit{source}_P \rightsquigarrow^{\star}_{P,G^{\mathrm{a}}} \mathit{root}_F$.

Let $\mathit{source}_P \in \mathcal{S}_P$.

1. If $\mathit{source}_P \in \mathcal{S}_F$.
   Let ϕ be the embedding between F and P mapping each agent $n \in \mathcal{A}_F$ into the agent $n \in \mathcal{A}_P$. By Def. 10, $\mathit{source}_P \rightsquigarrow^{\star}_{F,G^{\mathrm{a}}} \mathit{root}_F$. By Lemma 1, $\phi(\mathit{source}_P) \rightsquigarrow^{\star}_{P,G^{\mathrm{a}}} \phi(\mathit{root}_F)$.

Thus, $\mathit{source}_P \rightsquigarrow^{\star}_{P,G^{\mathsf{a}}} \mathit{root}_F$.

In particular, $(n_A, i_A) \rightsquigarrow^{\star}_{P,G^{\mathsf{a}}} \mathit{root}_F$.

2. Otherwise, $\mathit{source}_P = (n_B, i_B)$ and $((n_A, i_A), (n_B, i_B)) \in \mathcal{L}_P$.
We have already proved that: $(n_A, i_A) \rightsquigarrow^{\star}_{P,G^{\mathsf{a}}} \mathit{root}_F$.
Moreover, by Def. 14.(3), for any homorphism ϕ between P and G, we have that: $(\phi(n_B), i_B) \overset{\wedge}{\rightsquigarrow}_{G^{\mathsf{a}}} (\phi(n_A), i_A)$. Thus, $(n_B, i_B) \overset{\wedge}{\rightsquigarrow}_{P,G^{\mathsf{a}}} (n_A, i_A)$.
By transitivity, it follows that: $\mathit{source}_P \rightsquigarrow^{\star}_{P,G^{\mathsf{a}}} \mathit{root}_F$.

Thus, in any case, $\mathit{source}_P \rightsquigarrow^{\star}_{P,G^{\mathsf{a}}} \mathit{root}_F$. □

E Proof for Thm. 1

Thm. 1 follows from Prop. 1, Prop. 2, Prop. 3 by using the generic proof of [16, Sect. VII].

In the paper, we have skipped some technical details about the handling of trivial rules. The overlap between a prefragment and the lhs of a trivial rule may not be a prefragment. Moreover, it may happen that there exists an overlap between a prefragment and the rhs of trivial rule, such that the unique connected component of the corresponding refinement is not a prefragment. But in such a case, the consumption and production terms can be expressed easily by an other mean.

Let us consider a trivial rule $r : L \to R$ @k which releases bonds between the sites i_A of agents of type A and the sites i_B of agents of type B.

Let us consider a prefragment $F \in \hat{\mathcal{F}}$ which overlaps with L (both sites of L are modified by the rule). We assume that the join of the prefragment and L is not a prefragment. Then, because of Def. 14.(3), there cannot be two agents n_A and n_B such that $\mathit{type}_F(n_A) = A$, $\mathit{type}_F(n_B) = B$, $(n_A, i_A) \neq (n_B, i_B)$, $\{((n_A, i_A), (B, i_B)), ((n_B, i_B), (A, i_A))\} \subseteq \mathcal{L}_F$.

The consumption of this prefragment can then be expressed as follows:

$$y'_F \overset{+}{=} -\frac{k \cdot l \cdot y_F}{[L, L]}.$$

where l is the sum between the number of agents $n_A \in \mathcal{A}_F$ such that the site $(n_A, i_A) \in \mathcal{S}_F$ and $((n_A, i_A), (B, i_B)) \in \mathcal{L}_F$ and the number of agents $n_B \in \mathcal{A}_F$ such that the site $(n_B, i_B) \in \mathcal{S}_F$ and $((n_B, i_B), (A, i_A)) \in \mathcal{L}_F$.

Let us consider a prefragment $F \in \hat{\mathcal{F}}$ which overlaps with R. We assume that the connected component in the lhs of the corresponding refinement of the rule r is not a prefragment. Then necessarily, because of Def. 14.(2) and Def. 14.(3), there cannot be two agents n_A and n_B such that $\mathit{type}_F(n_A) = A$, $\mathit{type}_F(n_B) = B$, $(n_A, i_A) \neq (n_B, i_B)$, $\{((n_A, i_A), \dashv), ((n_B, i_B), \dashv)\} \subseteq \mathcal{L}_F$. We introduce X_A as the set of the agents $n \in F$ such that $\mathit{type}_F(n) = A$, $(n, i_A) \in \mathcal{S}_F$, and $((n, i_A), \dashv) \in \mathcal{L}_F$ and the set X_B as the set of the agents $n \in \mathcal{A}_F$ such that $\mathit{type}_F(n) = B$, $(n, i_B) \in \mathcal{S}_F$, and $((n, i_B), \dashv) \in \mathcal{L}_F$.

For any $n \in X_A$, we define the site graph $F^{-1}_{(n,A)}$ as $(\mathcal{A}_F, type_F, \mathcal{S}_F, \mathcal{L}', p\kappa_F)$ with $\mathcal{L}' = (\mathcal{L}_F \setminus \{((n, i_A), \dashv), (\dashv, (n, i_A))\}) \cup \{((n, i_A), (B, i_B)), ((n_B, i_B), (A, i_A))\}$. The site graph F^{-1}_n is indeed a prefragment. This is a consequence of the fact that F is a prefragment which overlaps with the rule r on a modified site, and that the G^{a} is compatible the rule r. This can be proved by using the same proof as the one that we have written in Appendix C (by using Def. 14.(2)).

For any $n \in X_B$, we define the prefragment $F^{-1}_{(n,B)}$ in the same way.

Then the production term for the fragment F can be expressed as follows:

$$y'_F \overset{+}{=} \frac{k}{[L, L]} \cdot \left(\sum_{n \in X_A} y_{F^{-1}_{(n,A)}} + \sum_{n \in X_B} y_{F^{-1}_{(n,B)}} \right)$$