Internal coarse-graining of molecular systems

Jérôme Feret [*], **Vincent Danos** [†], **Jean Krivine** [*], **Russ Harmer** [‡], **and Walter Fontana** [*]

[*]Harvard Medical School, Boston, USA, [†]University of Edinburgh, Edinburgh, United Kingdom, and [‡]CNRS & Paris Diderot, Paris, France

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Modelers of molecular signaling networks must cope with the combinatorial explosion of protein states generated by post-translational modifications and complex formation. Rule-based models provide a powerful alternative to approaches that require an explicit enumeration of all possible molecular states of a system. Such models consist of formal rules stipulating the contexts in which specific protein-protein interactions occur. However, the exploration of rule-based dynamics requires a stochastic simulation, which can become very costly. We address this problem with an automatic procedure for converting a rule-based model into a system of differential equations of vastly reduced dimension. This is achieved by constructing a self-consistent set of coarse-grained variables that are endogenously determined by the dynamics of the system, as implied by the rules. These variables refer to patterns of system components that can be viewed as the carriers of actionable information. The method rests entirely on the static analysis of a rule-based model, is exact, does not depend on the values of the rate constants appearing in the rules, and typically generates a dramatic reduction in the dimension of the system, making it amenable to traditional methods of numerical integration.

coarse-graining | rule-based models | molecular systems biology

Molecular biology is spectacularly successful in disassembling cellular systems and anchoring cell-biological behaviors of staggering complexity in chemistry. This raises the challenge of reconstituting molecular systems formally, in pursuit of principles that would make their behavior more intelligible and their control more deliberate. This pursuit is as much driven by the practical need to cure disease as it reflects a want for a theoretical perspective needed to understand the complexity of cellular phenotypes.

Two broad problems stand out on the theoretical frontier. First, we must be able to represent and analyze molecular interaction systems of combinatorial complexity. While ubiquitous, such systems are perhaps most notorious in the context of cellular signaling. The post-translational modification of proteins and their non-covalent association into transient complexes generate an astronomic number of possible molecular species that can relay signals [1]. The question then becomes how to reason about system dynamics if we cannot possibly consider a differential equation (whether stochastic or deterministic) for each chemical species that can appear in a system. It is seductive to chase a heuristic dynamical model that captures the "essence" of a complex hairball of interactions. Yet, to attain such desirable simplifications in an unbiased fashion (or find out whether they even exist), we may first have to study the behavior of a system at the level of its apparent empirical complexity.

Second, understanding *systems* requires resisting the temptation of adopting the view of an outside observer. Such a view is indeed appropriate for the chemical analysis of systems. When dissecting the composition and interactional structure of a network, the experimenter interacts with that network to create measurable distinctions. Yet, the network, as a dynamical system, may not be capable of these same observations. For example, an experimental technique might differentiate between SOS recruited to the membrane via GRB2 bound to SHC bound to the EGF receptor and SOS recruited via GRB2 bound to the EGF receptor directly. However, from the perspective of the EGF signaling system such a distinction might not be observable for lack of an endogenous interaction through which such a difference could become consequential. The endogenous units of the dynamics may differ from the exogenous units of the analysis.

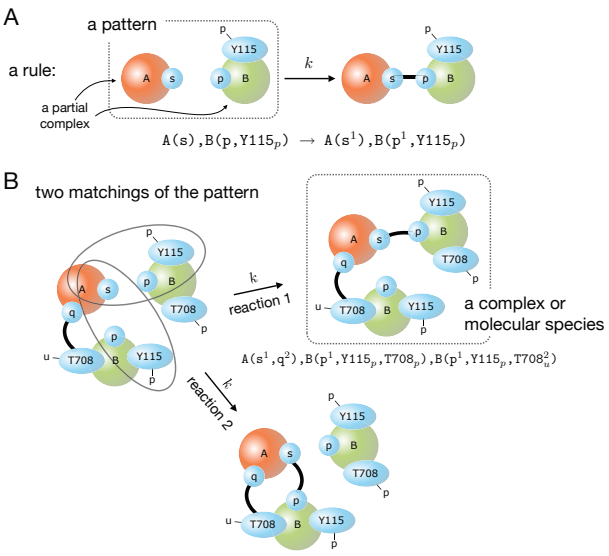


Fig. 1. Rules and reactions in Kappa. **A:** A rule captures a high-level mechanistic statement (empirical or hypothetical) about a protein-protein interaction in terms of a rewrite directive plus a rate constant k. The left hand side (lhs) of the rule is a pattern of partially specified agents, and represents the contextual information necessary for identifying all reaction instances that proceed according to the rule. The right hand side (rhs) expresses the actions that occur when the conditions specified on the lhs are met in a reaction mixture of Kappa agents. A connected subgraph on the lhs of a rule is called a component or partial complex. **B:** The rule in **A** matches a combination of agents in two distinct ways giving rise to two possible reactions with different outcomes. In a simulation, the choice is made probabilistically. In a textual representation, agents are names followed by an interface of sites delimited by parentheses. Bonds are labelled by superscripts and internal states at a site by subscripts. In the graphical rendition, internal states are indicated as labeled barbs. See SI and text for more details.

How do we discover the former? The present contribution attempts an answer.

Empirical analysis deals with actual systems; formal understanding deals with models of actual systems informed by empirical analysis. Discovering endogenously coarse-grained variables thus hinges on a representation that can cope with the combinatorial complexity of molecular systems. We address the latter challenge by employing a language, Kappa [2], that forms the basis of a substantive modeling framework [3, 4, 5, 6] within which we cast our approach to endogenous variables. This framework [7] will be detailed elsewhere, but the next section provides a brief outline of concepts that we shall need.

Reserved for Publication Footnotes

Kappa: A language for molecular biology

Kappa [2] is a formal language for defining agents (typically meant to represent proteins) as sets of sites that constitute abstract resources for interaction, as illustrated in Figure 1 and detailed in section 1 of the Supplementary Information (SI). Sites can hold internal state, as generated through post-translational modifications, and engage in binding relations with sites of other agents. In this language, a possibly temporary association of proteins is a connected graph, called a complex (of agents), as shown in the box of Figure 1B. The nodes of the graph are agents, but the endpoints of edges are sites, which belong to agents. While an agent can bear many connections, a site can bear only one.

Kappa enables the expression of *rules* of protein-protein interaction at a simple yet meaningful level of *tunable resolution*. The idea of a rule, Figure 1A, is to specify only the molecular context required for an interaction, along with a rate constant. Therein lies the power of rules to control combinatorial complexity. The left hand side (lhs) of a rule defines a pattern consisting of agents for which some sites and/or states have been omitted (see SI, section 1.2). The right hand side (rhs) exhibits the modifications that occur when the lhs pattern is matched in a mixture of agents. The difference between rhs and lhs is called the *action* of the rule. Sites mentioned on the lhs are said to be *tested* by the rule. Sites that are tested but not modified are also referred to as the *context* of a rule. Rules expand by pattern matching into potentially numerous reaction instances (whose rate constants are related to the rate constant of the rule) that involve particular combinations of molecular species, each of which realizes the context required for the rule to apply, see Figure 1B and Figure S4.

Kappa-rules stand in analogy to reaction rules in organic chemistry, where aspects of molecules that are irrelevant to a chemical rearrangement are designated as "remainder" groups. In Kappa, they are simply not mentioned. While chemistry has a theoretical foundation for rationalizing rules of reaction, Kappa-rules only codify observations, not why these observations might make sense to a structural biologist or biochemist. We formally define the grammar of Kappa and the concepts of rule, pattern, matching, and replacement in the SI. The reader should be able to follow our reasoning with an intuitive grasp of Kappa-rules, following Figure 1.

Kappa-rules are not only descriptive, they also are executable, and induce a stochastic dynamics on a mixture of agents (implemented by a suitably generalized Gillespie algorithm [3]). A Kappa model of a biological system is a collection of rules (with rate constants) and an initial mixture of agents on which these rules act. This makes a Kappa model akin to a concurrent computer program whose instructions are rules that asynchronously change the state of a store represented by the reaction mixture. Programs are formal objects that can be analyzed statically. *Static analysis* is about discovering behavioral properties of a program without running it (hence the name), much like a system of differential equations can be studied without simulating it. This involves inspection of the causal dependencies among rules and approximation of the molecular states that are reachable from a given initial condition. It is the central role of static analysis that sets our framework apart from other rule-based approaches [8, 9], whose goal is the automated assembly of kinetic differential equations by an iterative fixed-point construction of all reaction instances from a set of rules. The combinatorial explosion inherent in molecular signaling makes such goals impractical and often impossible. In a pilot study of EGF signaling, we collated 66 rules representing mechanistic observations of pertinent protein-protein interactions. These rules would produce 10^{19} molecular species. Our current EGF model has grown to about 350 rules. It thus appears more useful to forgo the expansion into an inscrutably large system of equations, and instead apply static analysis techniques directly to the rule collection. To complement the static analysis, rules can also drive a stochastic simulation [3, 10].

A rule-based model affords the inclusion of a wealth of information about protein-protein interactions, without the need to prematurely hypothesize which molecular states are or are not relevant to the dynamics of a system of interest. Still, it would be desirable to express a Kappa-model as a system of differential equations for the purpose of rapid numerical integration and exploration. In this contribution, we show how a Kappa-model can be transformed into a closed system of differential equations cast in terms of internally coarse-grained variables that refer to linear combinations of molecular species, which we call fragments.

Independence and self-consistency

Before detailing the fragmentation procedure, we discuss two simple examples aimed at clarifying the notions of independence and self-consistency. The ideas developed here are related to the elegant work of Borisov [11] and Conzelmann [12, 13], although we differ in our framing and objectives.

Example 1 (tileability of a system). Consider a scaffold protein C with a specific binding site for A and B (Figure S6). This system has 6 possible molecular species, $A, B, C, A.C, C.B, A.C.B$ (the dot indicates a bond), related by four association and dissociation reactions. To assert that the binding sites are independent is to assert that the rate constant for the binding of A to C is the same as the rate constant for the binding of A to $C.B$. Analogous equalities hold for the interactions between C and B. In such a system we can conceptually split the centerpiece C into two fragments (Figure S6A), one containing only the A-binding site, the other only the B-binding site. Let us denote the former fragment with $C*$ and the latter with $*C$, the asterisk indicating that we don't care about the corresponding binding site. The system then splits into two independent subsystems, self-consistently described by $\{A, C*, A.C*\}$ and $\{B, *C, *C.B\}$. Self-consistent means that each set of variables is closed with regard to its own dynamics. Independent means that we can reconstruct the microscopic dynamics from the description of both subsystems, as outlined next.

Suppose we pick a C at random from the reaction mixture and observe it to be bound to a B. The conditional probability that the same C is also bound to an A is formally written as $P(A.C*|*C.B)$. If, on the other hand, we choose not to observe the B-binding site of the C that we picked, that probability is $P(A.C*|*C*)$. Clearly, independence means that the two conditional probabilities are equal:

$$P(A.C*|*C.B) = P(A.C*|*C*).$$

By definition of a conditional probability, we obtain:

$$P(A.C.B) = \frac{P(A.C*)P(*C.B)}{P(*C*)}. \qquad [1]$$

These relationships are reflected by the corresponding (time-dependent) concentrations $[A.C*]$, $[*C.B]$, and $[*C*]$. Thus, equation (1) asserts

$$[A.C.B] = \frac{([A.C] + [A.C.B])([C.B] + [A.C.B])}{[C] + [A.C] + [C.B] + [A.C.B]}, \qquad [2]$$

from which it follows that

$$X \doteqdot [A.C.B][C] - [A.C][C.B] = 0. \qquad [3]$$

It is straightforward to verify that $X(t) = 0$ is an invariant of motion, provided the concentrations satisfied $X(0) = 0$:

$$\frac{dX}{dt} = -X(k_1^{A,C}[A] + k_{-1}^{AC} + k_1^{C,B}[B] + k_{-1}^{CB}), \qquad [4]$$

where $k_1^{A,C}, k_1^{C,B}$ denote association constants and k_{-1}^{AC}, k_{-1}^{CB} dissociation constants. Note that $X(t)$ will decay exponentially, if $X(0) \neq 0$. Equation **[1]** is a manifestation of independence and can be generalized to *define* a class of systems whose fragments (coarse-grained elements) behave like tiles in the following sense. A fragment is the partial specification of a molecular species. A fragment

may omit some agents, sites, and states. (We shall define fragments more precisely later.) In our example, the element $A.C*$ is such a fragment. It specifies a set of molecular species by ignoring the B-binding site of C. Two fragments $\mathcal{F}_1$ and $\mathcal{F}_2$ can be "snapped" together (possibly in more than one way), $\mathcal{F}_1 \sharp \mathcal{F}_2$, if they have a non-empty overlap $\mathcal{F}_1 \cap \mathcal{F}_2$ on which they agree with regard to the names and states both specify. In our example, the fragments $\mathcal{F}_1 = A.C*$ and $\mathcal{F}_2 = *C.B$ can be snapped together, since they overlap in the agent name C and don't conflict in the states of the sites they mention. (The first fragment specifies the state of the A-binding site, which the second fragment ignores, and the second fragment specifies the state of C's B-binding site, which the first fragment ignores.) The overlap $\mathcal{F}_1 \cap \mathcal{F}_2$ is $*C*$.

If a self-consistent set of fragments $\mathfrak{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_n\}$ obeys the independence equation **[1]**, we can extend a fragment $\mathcal{F}_i$ into a fragment $\mathcal{F}_i \sharp \mathcal{F}_j$ whose concentration is given by the product of the concentrations of $\mathcal{F}_i$ and $\mathcal{F}_j$ divided by the concentration of the snapping region, the overlap $\mathcal{F}_i \cap \mathcal{F}_j$. By extending fragments, we can reconstruct any molecular species that can possibly occur in the system, while in the process computing its concentration via **[1]**. The dynamics in $\mathfrak{F}$ will be exact, provided the initial condition satisfied the independence relations of the form **[3]**, otherwise it will approach the exact dynamics exponentially according to **[4]**. Typical situations are more subtle, as illustrated in the next example.

Example 2 (stealth correlation). Imagine our scaffold agent C with three binding sites, denoted as a (the A-binding site), c (a controller site), and b (the B-binding site). We simplify the description and only consider whether a site is occupied (state o) or not (state u). Denote the binding state of agent C by a triplet (acb) and its corresponding concentration by $[acb]$. To make our point expeditiously, it suffices to only consider (pseudo) unimolecular binding reactions (thus assuming that ligands are in excess and disregarding unbinding events). In contrast to Example 1, we require site c to be occupied for binding to occur at sites a and b. Like in Example 1, the binding process at site a is still independent of the binding state of site b, and vice versa. Binding at a and b proceeds with (pseudo) rate constants k_a and k_b, respectively, while binding at c proceeds with rate constant k_c. Again, as in Example 1, the system inherently splits into two self-consistent sets of coarse-grained variables $\{(uu*), (uo*), (oo*)\}$ and $\{(*uu), (*ou), (*oo)\}$. Each fragment now includes the central binding site c, since it determines whether a or b can undergo binding. (Thus, by construction, $[uu*] = [*uu] = [uuu]$.) However, unlike in Example 1, we cannot recombine the subsystems to reconstruct the original set of variables via equation **[1]**. In analogy to equation **[3]**, we define $X \doteqdot [ooo][uou] - [oou][uoo]$ as a measure of independence and obtain

$$\frac{dX}{dt} = -X(k_a + k_b) + k_c[ooo][uuu], \qquad \textbf{[5]}$$

indicating that the two subsystems remain correlated. While the coarse grained dynamics is still exact, the coarse-grained variables can no longer be used to reconstruct the complete original system by tiling; some information pertaining to correlations has been lost contrary to Example 1. As outside observers, we can reason over the reaction equations globally and notice that we could measure the state of a, to give us information about the state of c, from which we could infer the state of b. But from the system's internal point of view, the correlation between a and b is not observable. This is, in fact, the general situation.

Determining internally coarse-grained fragments

We next explain the general process for obtaining endogenously coarse-grained variables from a rule-based Kappa model. Our running example is adapted from early events in epidermal growth factor (EGF) signaling [14]. These events include the binding of EGF (agent `E`) to the receptor (agent `R`), the subsequent dimerization of the receptor, and the eventual recruitment of SOS (agent `O`). The model consists of 39 rules r01-r39, listed in section 3 of the SI. The names of agent sites were chosen fairly arbitrarily. We usually write separate rules for binding and unbinding actions, because unbinding often occurs under less restrictive contexts than binding. Together, these rules imply 356 possible distinct molecular species. We shall see, however, that based on these rules of interaction the system can only make 38 internal distinctions. Differential equations in these 38 variables self-consistently describe the dynamics of the system.

By modifying patterns, rules consume and produce fragments. The problem is to determine the mass action terms that each rule contributes to the kinetics of any given fragment. For the system of fragments to be consistent, each particular kinetic term must relate to only one fragment. The requirement for consistency, thus, constrains the relationship between pattern components and fragments. Specifically, any fragment that properly intersects a component on the left hand side (lhs) of a rule, and whose intersection contains a site that is modified by the action, must contain that component. Hence, each fragment either has no intersection with or completely contains (i.e. extends) a modified pattern component. This requirement translates into syntactical criteria that we shall discuss below. These criteria amount to an analysis of how dependencies between sites, specified by rules, propagate within and across agents.

The role of rate constants. The variables are determined by static analysis, wherein the system is never executed. Thus, there is neither a need to list rate constants, nor to provide any particular initial condition (other than assuming that all five types of agents are present). Nonetheless, kinetic distinctions are implicit in the system of rules. For example, empirical evidence might demand that two sets of reaction instances that could have been subsumed under a single rule actually need to be distinguished on kinetic grounds. In that case, the model would contain two rules that specify the same action (on the same agents), while differing in the context required to separate the reaction instances associated with a particular rate constant. For example, the actions specified by the rules r24: $\texttt{R(Y48}_p\texttt{),S(c,Y7}_u\texttt{)} \rightarrow \texttt{R(Y48}^1_p\texttt{),S(c}^1\texttt{,Y7}_u\texttt{)}$ and r28: $\texttt{R(Y48}_p\texttt{),S(c,Y7}^1_p\texttt{),G(a}^1\texttt{,b)} \rightarrow \texttt{R(Y48}^2_p\texttt{),S(c}^2\texttt{,Y7}^1_p\texttt{),G(a}^1\texttt{,b)}$ are the same: `S` (at its site named `c`) binds the receptor `R` (at its site `Y48` in a phosphorylated state). However, rule r24 specifies that this occurs when site `Y7` of `S` is unphosphorylated and free, while r28 specifies that this happens when site `Y7` of `S` is phosphorylated and bound to G. If kinetic distinctions were to require a ultimate refinement into molecular species, rules and fragments would degenerate into reactions and molecular species, respectively. However, once a rule collection is given, quantitative differences in reaction velocities are of no further consequence for fragmentation. Naturally, the values of rate constants matter for the fragment dynamics, but not for determining the fragments themselves.

Compression and reachable states. Coarse-graining proceeds by inspecting the structure of rules. A rule should be concise, in the sense of avoiding redundant contextual conditions on its lhs. Yet, what classifies as redundant depends on the remaining rules in the model. Recall that rules are both formal representations of empirical information and executable model elements. In their former role, they reflect the necessarily parochial view of an empirical observation. Consider, for example, rule r02 expressing the binding of ligand to receptor: $\texttt{R(l,r),E(r)} \rightarrow \texttt{R(l}^1\texttt{,r),E(r}^1\texttt{)}$. The rule mentions two sites, `l` and `r`, of the receptor `R`. Site `l` is the ligand (EGF) binding site, whose state is modified by the action of r02, while `r` is the site at which the receptor dimerizes (as described in r03). Rule r02 asserts that binding of `E` (EGF) to `R` requires not just a free `l`-site, but also a free `r`-site. This may be a legitimate empirical observation and should therefore be recorded accordingly. However, in the frame of reference provided by the other rules of this particular model, there is no *reachable state* of the reaction mixture in which `R` could possibly

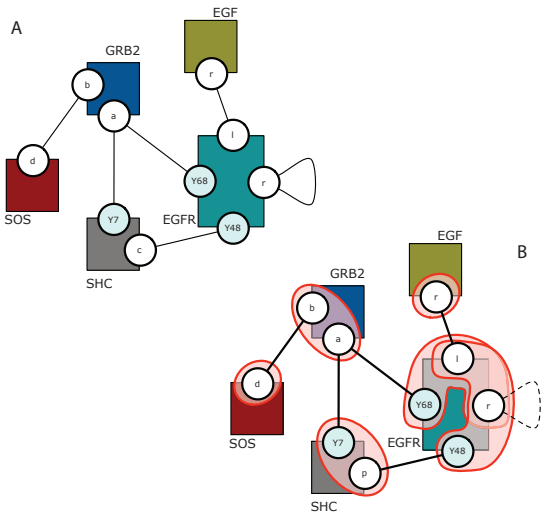


Fig. 2. The contact map. **A:** The contact map is a graph whose nodes are the agents in the model and whose edges are reachable binding interactions between sites. Filled circles indicate sites with reachable modifications of state. The contact map is a refined version of what is known as a protein-protein interaction (PPI) map, in that its edges end in sites at agents, not just agents. **B:** The annotated contact map (ACM) after decoration induced by the directives Cov1-Cov4 and Edg1. See text for details.

dimerize before binding E. Hence, in the particular context of the remaining 38 rules, asking for site r to be free is a redundant condition for the firing of rule r02, as a free l-site implies a free r-site. Without removing such redundancies, fragments would carry fictitious dependencies. We developed a technique for the automatic removal of unnecessary contextual specifications in a given set of rules.

The key difficulty in implementing this compression is the need to determine the set of molecular states that are potentially reachable by the combined action of all rules in the model. Clearly, an exact enumeration would defeat the rule-based approach. Our static analysis prevents this situation by computing an overapproximation of the reachable states by recasting rules and agents in terms of *local views*, wherein an agent has information only about its own state and the names of sites and agents *directly* bound to it. The local view of an agent ignores information about the internal states of sites not owned by the agent, or the names of agents more than one bond removed from it. The reachable local views are always finite. When combined with one another, they generate an overapproximation of the accessible states in the original model. The technique may therefore generate false positives, yet it *never* will construct false negatives. This abstract interpretation [15] of a rule set permits an extremely rapid approximation of reachable states, and, importantly, a symbolic representation of that set, obviating an explicit expansion into its (possibly infinitely many) members. A detailed exposition of abstract interpretation within the Kappa framework is described in [6]; a more accessible description with a discussion of rule compression is forthcoming. In ref. [6] we provide a characterization of systems for which this technique is exact. These conditions are fulfilled in the present EGF example. In section 3 of the SI, we list the 39 compressed rules cr01-cr39.

The contact map. A very useful concept in the construction of fragments is the *contact map* (CM), Figure 2A. The CM summarizes possible interactions as determined by our reachability analysis sketched in the previous section, and is generated automatically from a set of rules. The CM is a graph whose nodes are the agents that appear in the model. Recall that agents are sets of sites. These sites are the endpoints of edges representing possible binding interactions. Certain sites are colored to indicate that their state can be modified.

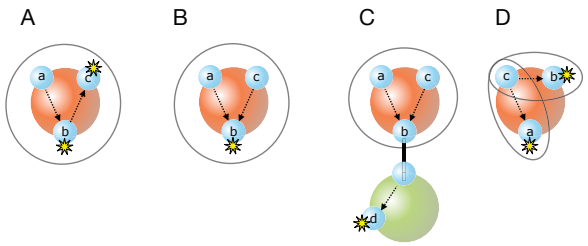


Fig. 3. Examples illustrating the syntactical criteria Cov1 and Cov2 for determining classes in the covering of an agent. See text for further details.

Annotating the contact map. We next formulate syntactical criteria that ensure the proper relationships of containment between fragments and pattern components on the lhs of rules, as stipulated before. Another way of rationalizing the syntactical criteria is to consider how dependencies between sites propagate within and across agents. These dependencies result from the contextual specifications in the rules, and determine which fragments the system can distinguish. Fragments are held together by "pathways" of control between sites.

The CM is a convenient canvas for summarizing these dependencies. For this purpose, the CM will be annotated with two types of information. (1) For each agent type A, we define a covering $\mathcal{C}(\mathtt{A})$ of the set of its sites. A covering $\mathcal{C}$ of a set S is a set of subsets of S, called classes, such that (i) no class is empty, (ii) no class is a subset of another class, and (iii) the union of all classes yields S. A covering differs from a partition in that the elements of a covering need not be pairwise disjoint. (2) For each edge in the CM, we define its type as either "solid" (a conduit of dependencies) or "dotted" (not a conduit).

We first list the syntactical criteria to determine a covering for an agent, and follow up with some explanatory remarks.

Cov1 (backward closure): If a rule tests a site a in an agent A and modifies a site b in the same agent, any class in $\mathcal{C}(\mathtt{A})$ that contains b must also contain a.

Cov2 (control portal): If a rule tests a site a in an agent A, and A is connected through a site b to an agent that is modified, any class in $\mathcal{C}(\mathtt{A})$ that contains b must also contain a.

Cov3 (remote test): For each agent in an unmodified pattern component on the lhs of a rule, there must be a class in the agent's covering that contains the sites tested by the rule.

When constructing fragments (in the next section), sites will be included as classes. Classes summarize control flows *within* an agent and contribute to defining the shape of a fragment. Final fragments are derived by distinguishing all possible state combinations of sites within a class. Thus, large classes will increase the number of fragments combinatorially. Imagine starting out with each site in its own class (akin to a null hypothesis of independence). Next, imagine a rule that tests the state of site a and modifies the state of site b of agent Red in Figure 3B. The very rule postulates that this control (and the ensuing correlation between a and b) is observable by the system, thus forcing the two sites into the same class by Cov1. Imagine a further rule involving agent Red, which asserts that its site c also controls b. To track the state of b, the system must now distinguish simultaneously between the states at both controlling sites a and c. Hence, there must be fragments containing agent Red in all possible states at sites a, b, and c.

Recall Example 2, as recast in Figure 3D, where site c, controls binding events at sites a and b. This inverts the control flow, compared to Figure 3B, resulting in two separate classes for the sites of Red (and giving rise to the fragments that figured in our discussion of Example 2). The separation of a and b into distinct classes is equivalent to saying that no correlation between the state at a and the state at b is observable by the system. For it to become observable, a rule

would have to, say, test b and modify a. But such a rule would trigger Cov1, merging the two classes into one, as in Figure 3B.

Figure 3C considers the case in which a rule tests the state of site a of agent Red and modifies site d in agent Green that is bound to Red through b. Although the controlling influence of a on d is direct, the simultaneous test for a link between Red and Green correlates the state of b in Red with that of a. One might say that the controling influence propagates through the bond. Site b is a "portal", and if another rule correlates another site in Red with b, the same argument as in Cov1 applies. Cov2 is thus really a version of Cov1.

When a rule jointly tests sites of an agent A in a lhs pattern that is disconnected from the target of a rule's action, correlations between the test sites at A become observable, forcing these sites into one class. That is the content of Cov3.

Control not only correlates sites within an agent, but also propagates across bonds. The edge annotations in the contact map serve the purpose of classifying whether a bond has, or transmits, control information. Let us call a dissociation rule trivial, if it only tests whether the bond to be broken exists. Clearly, such a rule does not convey controling power to the bond. Hence:

Edg1: We only look at bonds on the left hand side of rules. Whenever a bond is mentioned on the left hand side of only trivial dissociation rules, its corresponding edge in the annotated contact map must be dotted, otherwise it is solid.

Another way of phrasing Edg1 is to require that a bond appearing on the left hand side of a rule testing anything beyond the bond itself must be represented with a solid edge in the annotated contact map. In case the bond is a semi-link (see SI), all edges into which the semi-link expands must be solid. For example, if the bond in Figure 3C appears on the lhs of a rule, it would be painted solid in the ACM.

Fragments. Using the ACM, we construct *reachable fragments*. A fragment $\mathcal{F}$ is a partially specified complex of agents, such that:

Frag1: For any agent A in $\mathcal{F}$, the sites of A mentioned in $\mathcal{F}$ constitute a class in $\mathcal{C}$(A). Each such site is assigned a possible valuation of internal states and binding states.

Frag2: If a site a of an agent A is connected to a site b of an agent B through a dotted link in the ACM, the bound state of a in $\mathcal{F}$ is indicated as a superscript identifying the bond type, without including agent B in $\mathcal{F}$. The bond type is labeled *partner@site*, where *partner* is the name of the binding partner and *site* the name of its binding site. Thus, the partial complex $\mathtt{A(a^{B@b})}$ identifies binding site a of agent A as bound to agent B at site b.

Frag3: An agent A and an agent B, both in $\mathcal{F}$, can occur bound to one another only if the corresponding edge in the ACM is solid.

Given an ACM, all fragments can be generated systematically following the directives Frag1-Frag3. Consider, for example, agent R in our rules. According to the ACM in Figure 2B, we have a choice between two classes. Let us choose class {l,r,Y48}. Next, we assign a possible state to each site in that class. For example, all sites are free, and Y48 is unphosphorylated. This yields fragment $\mathtt{R(Y48}_u\mathtt{,l,r)}$, which is $\mathcal{F}34$ in the complete list of fragments for our example (section 2.3 of the SI). Alternatively, we might choose Y48 to be phosophorylated (fragment $\mathcal{F}15$). However, if we also choose Y48 to be bound, then the solid link in the ACM forces agent S into the same fragment, along with its site p as the link's endpoint. In turn, p forces inclusion of the class to which it belongs, {p,Y7}. Now we need to assign states to p and Y7 in agent S. For example, $\mathtt{S(Y7}_p\mathtt{,c}^1\mathtt{),R(Y48}_p^1\mathtt{,l,r)}$, which is fragment $\mathcal{F}04$. A further fragment is obtained by considering site r in agent R to be bound. Site r can bind to another R-agent, but the link is dotted. A dotted link at r does not force another instance of agent R into the fragment. Instead, the bound state is only indicated with its type: $\mathtt{S(Y7}_p\mathtt{,c}^1\mathtt{),R(Y48}_p^1\mathtt{,l,r}^{\mathtt{R@r}}\mathtt{)}$. This fragment, however, does not show up in our list. The state in which R is dimerized at site r cannot occur if the ligand binding site l is empty. Such a fragment is automatically eliminated from the list because the reachable state analysis recognizes it as inaccessible. The elimination of inaccessible fragments is exact, since it is based on local views only (rather than assembling local views into an overapproximation of reachable states).

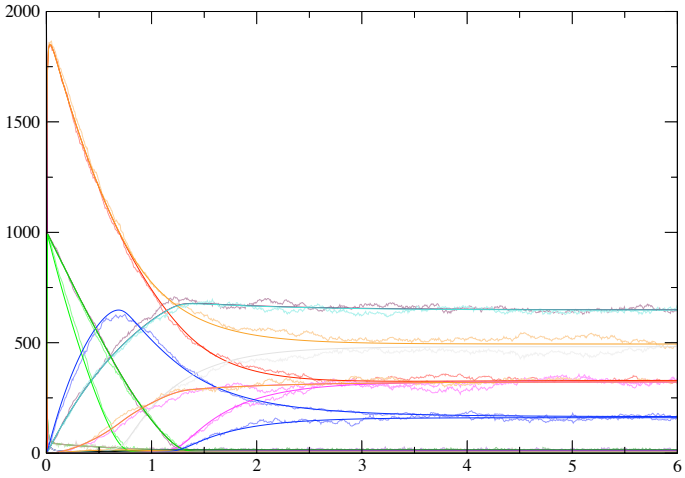


Fig. 4. A comparison between microscopic and fragment dynamics. Wiggly curves: The Kappa-model of the early EGFR example is executed with a Gillespie simulation (microscopic dynamics [3]), while reporting the coarse-grained fragment concentrations. Steady curves: The output of the deterministic fragment dynamics.

Assembling a dynamical system for fragments. Directives Cov1-Cov3 ensure that for any lhs of a rule and every occurrence of an agent A there always is a class in the covering of A that contains the sites mentioned by the lhs. The fragment construction directives Frag2-Frag3 then ensure that every partial complex on the left hand side of a rule is contained in some fragment. We need this to compute the concentration of partial complexes (pattern components) on the lhs of rules in terms of fragments (SI, section 4). The concentration of a pattern, such as a lhs component or a fragment, is the number of matchings it has in a reaction mixture. If a pattern is progressively refined into the specification of a molecular species, its concentration becomes the number of instances of that molecular species in the mixture. Because fragments are refinements of partial complexes on the lhs, the concentration of any such partial complex can be expressed as a linear combination of fragment concentrations, provided, however, that these fragments do not to overlap in their matching instances or we would overcount. We have segregated the characterization of such "orthogonal" fragments into the SI (section 4.2), where the reader can also find the formal expression (equation **[12]**) connecting the concentrations of partial complexes to the concentrations of fragments.

To express the evolution of fragment concentrations, we require that each time a fragment contains a site modified by a rule, it then contains the entire affected component in the lhs of the rule. This, too, is guaranteed by our directives. The dynamical system for fragments is then constructed by deriving mass action terms for the consumption and production of fragments from compressed rules. Consider, for example, a rule of the form $Z, Z' \rightarrow Z\text{–}Z'$, which binds two partial complexes Z and Z'. Based on this rule, the differential equation $d[\mathcal{F}_i]/dt$ for each fragment $\mathcal{F}_i$ that extends Z gains a consumption term $-\gamma[\mathcal{F}_i][Z']$, where $[Z']$ is expressed as a sum of fragment concentrations, using equation **[12]** of the SI. γ is a rate constant related in a simple way to the constant of the rule. Such a term must be multiplied by the number of ways in which Z matches a particular $\mathcal{F}_i$. On the production side, the kinetic terms depend on

the link type in the ACM. Consider, for example, a bond of the solid type. A kinetic term $\gamma[\mathcal{F}_i][\mathcal{F}_j]$ is generated for the differential equation $d[\mathcal{F}_k]/dt$ of every fragment $\mathcal{F}_k$ that matches Z–Z', where $\mathcal{F}_i$ and $\mathcal{F}_j$ are fragments matching Z and Z', respectively, subject to the constraint that the match of $\mathcal{F}_k$ is the disjoint sum of the matches of $\mathcal{F}_i$ and $\mathcal{F}_j$ to their respective fragments. On the other hand, a dotted bond means that the formation of a partial complex Z–Z' represents the formation of two fragments obtained by replacing the bond with a binding type label, such as $Z^{B@b}$ and $Z'^{A@a}$, if the bond is between site a of $A \in Z$ and site b of $B \in Z'$. There is no information in $Z'^{A@a}$ needed to compute the concentration of $Z^{B@b}$. Thus, every fragment $\mathcal{F}_k$ matched by $Z^{B@b}$ gains a production term $\gamma[\mathcal{F}_i][Z']$, where $\mathcal{F}_i$ is the fragment matched by Z when $\mathcal{F}_k$ matches $Z^{B@b}$. A similar argument goes for fragments matched by $Z'^{A@a}$. A detailed accounting for all rule types can be found in section 4.4 of the SI. Figure 4 was obtained by running a microscopic stochastic simulation of the early EGF system, driven by rules r01-r39, while reporting the concentrations of fragments $\mathcal{F}_{01}$-$\mathcal{F}_{38}$. We then overlayed the integration of the deterministic dynamical system for fragments.

Conclusions: Esse est percipi

Rule-based representations are necessary for coping with combinatorial systems, but they require a stochastic simulation to generate the full dynamics, and this can be computationally very expensive. It would be highly useful, therefore, to construct a deterministic projection of rule-based dynamics. However, an expansion into the most fine-grained level of molecular species is precisely the practical impossibility that makes a rule-based approach necessary in the first place. Moreover, many biologists share the intuition that not all molecular species that might occur in a pathway possess biological significance. What are, then, the relevant variables?

Model reduction techniques are based either on approximations whose validity cannot always be guaranteed throughout the dynamical evolution of a system, or on projections onto variables largely defined *ad hoc* by the modeler. In this paper, we have taken the stance that the relevant variables are those that represent the distinctions the system itself is capable of making, based on the dynamics induced by the rules that define it. We call this "internal coarse-graining". We have presented (and implemented) an automatic procedure for the construction of a self-consistent set of internally coarse-grained variables whose deterministic dynamics is exact, in the sense of ending up in the same coarse-grained state that would be obtained by first running the dynamics at the level of molecular species and then coarse-graining. The system variables refer to so-called fragments, that is, patterns (linear combinations) of molecular species. Our procedure yields 38 such fragments for a model of early events in EGFR signaling [14] consisting of 39 rules that would expand into 356 molecular species. A pilot study of a larger portion of the EGFR system, comprising 66 rules expanding into 10^{19} molecular species, yields about 180,000 fragments - still a lot, but a lot better than practical infinity.

In sum, our contribution is twofold: First, we identify those patterns of a molecular system that carry the information that the system can actually access. To an outside observer, there may be (and there typically is) much more "information" in the system, but that information cannot be read by the system and is thus not actionable from within. Second, we bring many rule-based models of combinatorially complex systems within reach of established numerical integration techniques, thereby enabling a rapid exploration of their dynamics, without forgoing the benefits of the rule-based representation. While the dynamics at the level of molecular species cannot be reconstructed from the fragment dynamics (unless the system is tileable), the character of the former is still reflected in the latter.

The fragment equations contain all the rate constants of the rule-based representation, which enables, in principle, the deployment of parameter estimation techniques that require rapid integration. However, observational data for parameter fits must refer to fragments. Such data may be difficult to obtain directly, as they require detection techniques for specific patterns of molecular species. Yet, if measurements cover the temporal dynamics of a wide enough range of molecular species, such data could be aggregated into an empirical fragment dynamics.

Fragments provide a view of the informational architecture of a system [16], as implied by its rule-based representation. It might be insightful to attempt a sensitivity analysis of the fragmentation process, to determine which rules, when changed, have the biggest impact on the nature and number of fragments. Can highly consequential rules be guessed from the annotated contact map? How do such maps compare for a given molecular system across different circumstances, such as diseases, tissues, species, and across developmental and evolutionary time?

ACKNOWLEDGMENTS. We thank Holger Conzelmann for graciously making available his PhD thesis work. We are grateful to Eric Deeds and Peter Sorger for numerous insightful discussions.

1. Hlavacek, W. S, Faeder, J. R, Blinov, M. L, Posner, R. G, Hucka, M, & Fontana, W. *Rules for Modeling Signal-Transduction Systems* (2006) *Science STKE* 344, re6.
2. Danos, V & Laneve, C. *Formal Molecular Biology* (2004) *Theoretical Computer Science* 325, 69–110.
3. Danos, V, Feret, J, Fontana, W, & Krivine, J. (2007) *Scalable simulation of cellular signalling networks*, Lecture Notes in Computer Science. (Springer), Vol. 4807, pp. 139–157.
4. Danos, V, Feret, J, Fontana, W, Harmer, R, & Krivine, J. (2007) *Rule-based modelling of cellular signalling*, Lecture Notes in Computer Science. (Springer, Lisboa, Portugal), Vol. 4703, pp. 17–41.
5. Danos, V, Feret, J, Fontana, W, Harmer, R, & Krivine, J. (2008) *Rule-based modelling, symmetries, refinements*, Lecture Notes in Bioinformatics. (Springer, Cambridge, UK), Vol. 5054, pp. 103–122.
6. Danos, V, Feret, J, Fontana, W, & Krivine, J. (2008) *Abstract interpretation of cellular signalling networks*, Lecture Notes in Computer Science. (Springer), Vol. 4905, pp. 83–97.
7. Plectix BioSystems, Inc. (http://www.plectix.com). Contact support@plectix.com for instructions on how to access the Kappa modeling platform.
8. Blinov, M. L, Faeder, J. R, & Hlavacek, W. S. *BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains* (2004) *Bioinformatics* 20, 3289–3292.
9. Mallavarapu, A, Thomson, M, Ullian, B, & Gunawardena, J. *Programming with models: modularity and abstraction provide powerful capabilities for systems biology* (2008) *J. Roy. Soc. Interface* doi:10.1098/rsif.2008.0205.
10. Yang, J, Monine, M. I, Faeder, J. R, & Hlavacek, W. S. *Kinetic Monte Carlo method for rule-based modeling of biochemical networks* (2008) *Phys. Rev. E* 78, 031910 (7 pages).
11. Borisov, N. M, Markevich, N. I, Kholodenko, B. N, & Gilles, E. D. *Signaling through Receptors and Scaffolds: Independent Interactions Reduce Combinatorial Complexity* (2005) *Biophysical Journal* 89, 951–966.
12. Conzelmann, H, Saez-Rodriguez, J, Sauter, T, Kholodenko, B. N, & Gilles, E. D. *A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks* (2006) *BMC Bioinformatics* 7, 34.
13. Conzelmann, H. (2008) Ph.D. thesis (Universität Stuttgart).
14. Blinov, M. L, Faeder, J. R, Goldstein, B, & Hlavacek, W. S. *A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity* (2006) *BioSystems* 83, 136–151.
15. Cousot, P & Cousot, R. (1977) *Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints*. (ACM Press, New York, NY, Los Angeles, California), pp. 238–252.
16. Shalizi, C. R & Crutchfield, J. P. *Computational Mechanics: Pattern and Prediction, Structure and Simplicity* (2001) *J. Stat. Phys.* 104, 817–879.

Supporting Information

Jérôme Feret[*],**Vincent Danos**[†],**Jean Krivine**[*] ,**Russ Harmer**[‡] **and Walter Fontana**[*]

[*]Harvard Medical School, Boston, USA,[†]University of Edinburgh, Edinburgh, United Kingdom, and [‡]CNRS & Paris Diderot, Paris, France

1 Kappa

Kappa is a formal language for representing molecular objects as agents [1, 2]. These agents are decorated with sites that carry modifiable state and/or bind other agents to form complexes. Empirical observations about the states of proteins that permit them to interact in specific ways are expressed as rules in analogy to the rules of organic chemistry, see Figure 1. Kappa is a way of expressing observations at a level of resolution that is attuned to how molecular biologists have come to think about networks of protein-protein interactions. In its current implementation, Kappa is phenomenological, in the sense that the language is not based on a theory of why certain interactions occur or how they occur physically. In this section, we provide a formal syntax of the language, illustrating it with an equivalent graphical rendering, as used in Figure 1.

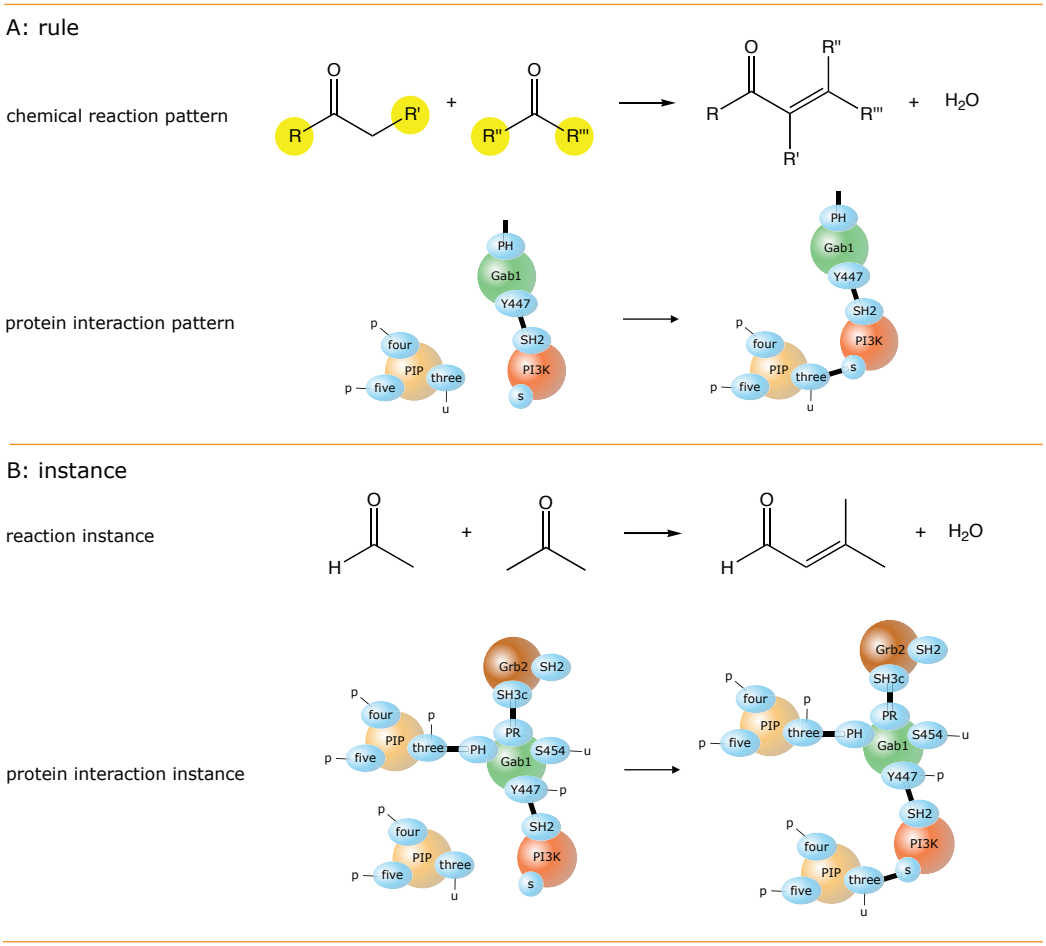


Fig. 1. Rule-based languages: the analogy between chemistry and Kappa. A: An aldol condensation is used as an example to illustrate the concept of a reaction rule in chemistry. The rule details only those molecular parts that are relevant to a particular scheme of reaction. Because the transformation of these parts is described independently of their contexts R, R', R", and R"' (highlighted), we refer to such a rule as a decontextualized reaction. Sometimes a generic scheme, such as in (A), requires refinement into special classes defined by different sets of contexts R-R"'. B: Upon full specification of the contexts $R = H$, $R' = H$, $R'' = CH_3$, and $R''' = CH_3$, the rule (A) becomes a reaction instance (B). Kappa proceeds in complete analogy, though at a far more coarse-grained level of description. A rule (A) describes the context required for a local interaction to occur. Panel B shows an instance that complies with the rule depicted in A.

Reserved for Publication Footnotes

1.1 Agents, complexes, mixtures. We first provide a formal definition of the context-free grammar of Kappa [3, 4], followed by a few explanations for readers unfamiliar with Backus-Naur notation.

In the following, let $\mathcal{A}$ be a set of agent names, $\mathcal{S}$ a set of site names (and let $\wp(\mathcal{S})$ denote the powerset of $\mathcal{S}$), $\mathbb{V}$ a set of internal states, and $\mathbb{N}$ a set of labels. Further, let $\psi : \mathcal{A} \mapsto \wp(\mathcal{S})$ be a map that associates an agent name to a set of sites, called the agent's interface.

Definition 1.1 (Agents)**.**

(i)	agent	a	::=	$N(\sigma)$	
(ii)	agent name	N	::=	$A \in \mathcal{A}$	
(iii)	interface	σ	::=	$\varepsilon \mid s, \sigma$	
(iv)	site	s	::=	n_ι^λ	
(v)	site name	n	::=	$x \in \mathcal{S}$	
(vi)	internal state	ι	::=	ϵ	(*any state*)
				$\mid m \in \mathbb{V}$	
(vii)	binding state	λ	::=	ϵ	(*free*)
				$\mid\ -$	(*semi-link: "bound to something"*)
				$\mid$?	(*unspecified: free or bound*)
				$\mid\ i \in \mathbb{N}$	(*bond label*)

Definition 1.2 (Expressions)**.**

(viii)	expression	E	::=	$\varepsilon \mid a, E$

Definition 1.3 (Well-formedness)**.**

(ix)	unique interface	each site name in the scope of an agent named A must be in $\psi(A)$
(x)	agent scope	a site name can occur only once in a given interface
(xi)	binary binding	a binding state $i \in \mathbb{N}$ occurs exactly twice, if it occurs at all

Definition 1.4 (Structural equivalence)**.**

(xii)	interface	$E, A(\sigma, s', s, \sigma), E'$	$\equiv$	$E, A(\sigma, s, s', \sigma), E'$	*(site permutation)*
(xiii)	mixture	E, a, a', E'	$\equiv$	E, a', a, E'	*(agent permutation)*
(xiv)	edge labels	$i, j \in \mathbb{N} \wedge i$ not in E	$\Rightarrow$	$E[i/j] \equiv E$	*(relabeling)*

The grammatical rules (i)-(xi) define well-formed expressions in Kappa. We shall define the syntax of Kappa-rules in section 1.3. The notion of "rule-based" models refers to rules expressing actions, not to the grammatical rules defining the terms of the language.

The grammatical rule (i) defines the overall syntax of an agent as consisting of a name N, taken from the set $\mathcal{A}$ (rule ii), and an interface σ. For example, we may call an agent `ErbB1`. Rule (iii) defines the interface of an agent as a finite set $\sigma = \{s_1, s_2, \ldots, s_n\}$ of *sites*. The vertical bar (|) in (iii) indicates a choice in the recursive application of the grammar when constructing agents. The rule is recursive because σ appears on both sides of the definition: a set of sites consists of a site s and a set of sites. Each time we iterate over (iii), we instantiate a different site s. The s in (iii) refers to the syntactical category "site" defined in (iv). The construction of an interface terminates by choosing the empty interface ε. The sites of an agent control the interactions it participates in. These interactions are defined by Kappa-rules, section 1.3. As indicated in (v), a site s is referred to by an arbitrary name in $\mathcal{S}$, much like an agent. According to (iv), a site carries two types of information, notated as a superscript and subscript to the site name. The subscript ι (iota) of a site refers to its *internal state*, which may be an arbitrary value, as specified in (vi). In most biological interpretations, the value of an internal state indicates a post-translational modification, such as "phosphorylated", "unphosphorylated", "methylated". The superscript λ of a site refers to its *binding state*, defined in (vii) . Agents may be bound to other agents at sites that belong to them. To indicate that site `l` of agent `ErbB1` is bound to site `r` of agent `EGF`, we deploy the same superscript at both sites. For example, the expression $\texttt{ErbB1(l}^2\texttt{),EGF(r}^2\texttt{)}$ indicates an agent `ErbB1` that is bound to an agent `EGF` at the sites indicated. A superscript uniquely labels a bond between two agents, as laid out in rule (xi). The superscript ϵ means that the site is unbound (free), while a subscript ϵ indicates an unspecified state (like a wild card). We typically do not write the value ϵ. For example, $A(s_\epsilon^\epsilon) \equiv A(s)$.

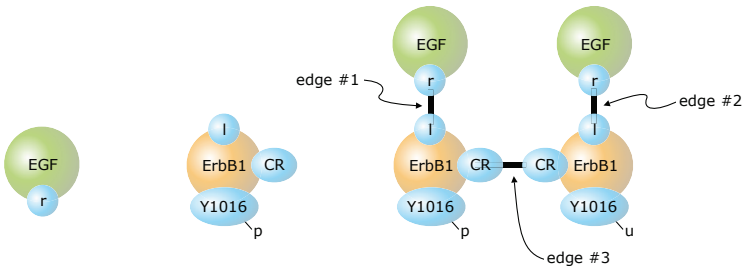
The object $\texttt{ErbB1(l}^2\texttt{),EGF(r}^2\texttt{)}$ is not itself an agent, because an agent has only one name by virtue of (i). In fact, `ErbB1` bound to agent `EGF` is a *complex*, which belongs to the syntactical category of *expression*, Definition 1.2. In the grammar rule (viii) for forming expressions, the symbol a refers to agents, as defined in (i)-(vii). An expression is simply a set of comma-separated agents. The syntactical category of expression thus includes the notion of a complex. For example, the expression

$$\texttt{EGF(r}^1\texttt{)} \ ,\ \texttt{ErbB1(l}^1\texttt{,CR}^3\texttt{,Y1016}_p\texttt{)} \ ,\ \texttt{EGF(r}^2\texttt{)} \ ,\ \texttt{ErbB1(l}^2\texttt{,CR}^3\texttt{,Y1016}_u\texttt{)} \qquad \textbf{[1]}$$

denotes a complex in which two `ErbB1` agents, each bound to an `EGF` agent, have dimerized on their sites named `CR` (Figure 2).

An agent is an atomic entity, in the sense of not being decomposable into further agents. A complex is a connected graph of agents. (In chemistry, an atom would correspond to an agent in our sense, and a molecule to a complex.) An expression is more general than a complex, since Definition 1.2 does not require any bindings between agents in an expression. Figure 2 illustrates an expression (and a graphical presentation) consisting of an agent $\texttt{EGF(r)}$, an agent $\texttt{ErbB1(l,CR,Y1016}_p\texttt{)}$, and the complex represented in **[1]**. As defined in (viii), an expression is a graph over agents whose connected components are complexes.

An agent should be thought of as being associated with a unique interface (by virtue of the mapping ψ). As we shall see later, agents in an expression are oftentimes mentioned with only a subset of their sites. Rule (ix) ensures that these sites are elements of the agent's interface.



EGF(r), ErbB1(l,CR,Y1016$_p$),EGF(r^1),ErbB1(l^1,CR^3,Y1016$_p$),EGF(r^2),ErbB1(l^2,CR^3,Y1016$_u$)

Fig. 2. A Kappa expression. The textual representation of a small reaction mixture containing 6 agents that are divided into three complexes (underlined) is shown at the bottom. The two complexes on the left are simple agents, while the complex on the right is made of 4 agents hanging together as shown. An equivalent graphical rendition is depicted above the textual expression, exhibiting the complexes in the same order (left to right) as in the expression below. Names of agents and sites are written inside their corresponding nodes, while internal states of sites, such as the phosphorylated state (p) of Y1016 at ErbB1, are indicated by a labeled barb coming off the corresponding site node.

We would like an expression to represent the contents of a well-stirred mixture or chemical solution. To formalize this intent, we define structural equivalences between expressions, Definition 1.4. This is a standard procedure in computer science to undo the distortion in literal meaning arising from the constraints of linear text. The first two equivalences, (xii) and (xiii), erase any notion of space in the Kappa language. This is important to keep in mind, since textual (and graphical) renditions have a tendency to fool us. In particular, rule (xii) states that an interface is a set, not an ordered sequence of sites. Hence, the placement of sites in a graphical representation, such as Figure 2, has no significance. Rule (xiii) states that an expression has no spatial meaning. Every agent or complex is "equidistant" from any other, since all shuffles of an expression are equivalent. An expression, therefore, represents a well-mixed solution of molecular objects. Rule (xiv) states that we can relabel edges (bonds) as we please, provided the labels remain unique. Thus, if j is an edge label in an expression E and i is not, then we can substitute i for j in E (denoted by $E[i/j]$) without changing the meaning of E.

1.2 Patterns. An expression representing the contents of a reaction mixture typically contains agents with fully specified interface. It is useful, however, to consider agents with only a partially specified interface. Recall that chemical rules, such as the one in Figure 1A, refer to partially specified molecules for the purpose of isolating a transformation that occurs across many reaction instances consisting of different fully specified molecules.

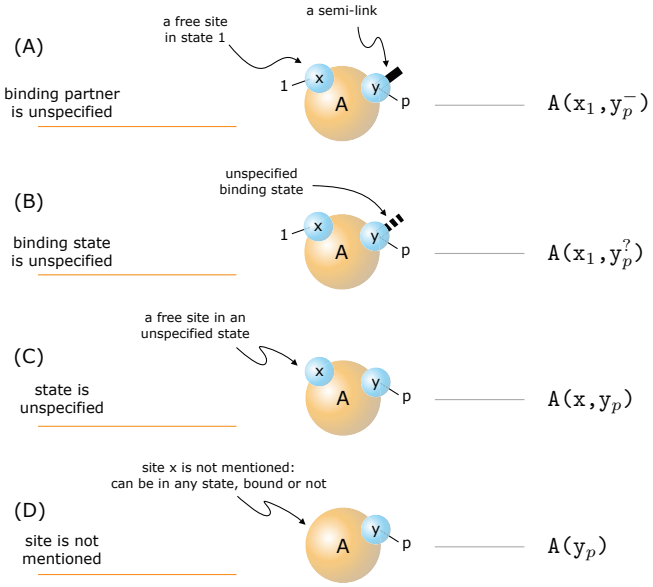


Fig. 3. Basic Kappa patterns. A pattern is a partially specified agent (or set of agents). Various ways of omitting information are shown. A: the binding partner of a site is left unspecified. B: the binding state of a site is left unspecified. C: the internal state of a site is unspecified. D: both internal and binding states of a site are left unspecified by not mentioning the site at all. See text for details.

A *pattern* is an expression of partially specified agents. Figure 3 depicts the basic types of patterns exemplified by an agent A(x,y) with two sites:

1. *Unspecified binding partner.* The expression $\mathtt{A(x_1,y_p^-)}$, Figure 3A, specifies an agent in state 1 at site x and in state p at site y. In addition, site y is bound, but we don't specify to whom. We call this a *semi-link* and indicate it by a hyphen (−) instead of an edge label.
2. *Unspecified binding state.* The agent expression $\mathtt{A(x_1,y_p^?)}$, Figure 3B, is similar to the previous one, except that we do not care whether site y is bound. We indicate this by a question mark (?) in the bond superscript. Note that by *not* mentioning any binding state for site x we assert that this site is free (unbound).
3. *Unspecified internal state.* In $\mathtt{A(x,y}_p\mathtt{)}$, Figure 3C, we do not care about the internal state of site x, because we omit its subscript. However, we do care that the site be free (as in all previous cases). Site y is in state p and free.
4. *Omitted site.* In $\mathtt{A(y}_p\mathtt{)}$, Figure 3D, we omit site x entirely, asserting that we don't care about its internal state nor its binding state. Site y is in state p and free.

1.3 Rules. The main use of patterns is in the definition of Kappa rules. In analogy to chemical reaction rules, a rule is a pair of expressions that are typically patterns:

$$E_{\text{left}} \longrightarrow E_{\text{right}}.$$

The pattern E_{left} defines conditions on internal states and binding states of agents that have to be satisfied for the rule to apply. Rules are applied to a mixture, that is, an expression S representing the contents of a reaction system at a given time. The basic idea is illustrated in Figure 4 for the rule

$$\mathtt{Gab1(PH^-,Y447^1),PI3K(SH2^1,s),PIP(three}_u\mathtt{,four}_p\mathtt{,five}_p\mathtt{)} \quad [2]$$

$$\downarrow$$

$$\mathtt{Gab1(PH^-,Y447^1),PI3K(SH2^1,s^2),PIP(three}_u^2\mathtt{,four}_p\mathtt{,five}_p\mathtt{)}$$

(which we write vertically for ease of placement on the page). Below the rule in Figure 4, we have sketched a hypothetical mixture. We want to identify *a* configuration of (fully specified) agents in the reaction mixture S that satisfies the pattern of reactants on the left hand side (lhs), E_{left}, of the rule. When such a configuration has been located, it is replaced by the configuration specified on the right hand side (rhs), E_{right}, of the rule. Replacement consists in updating the internal states and the binding states that are changed by the rule. The operational meaning of a match and a replacement are formalized in section 1.4 and 1.5, respectively.

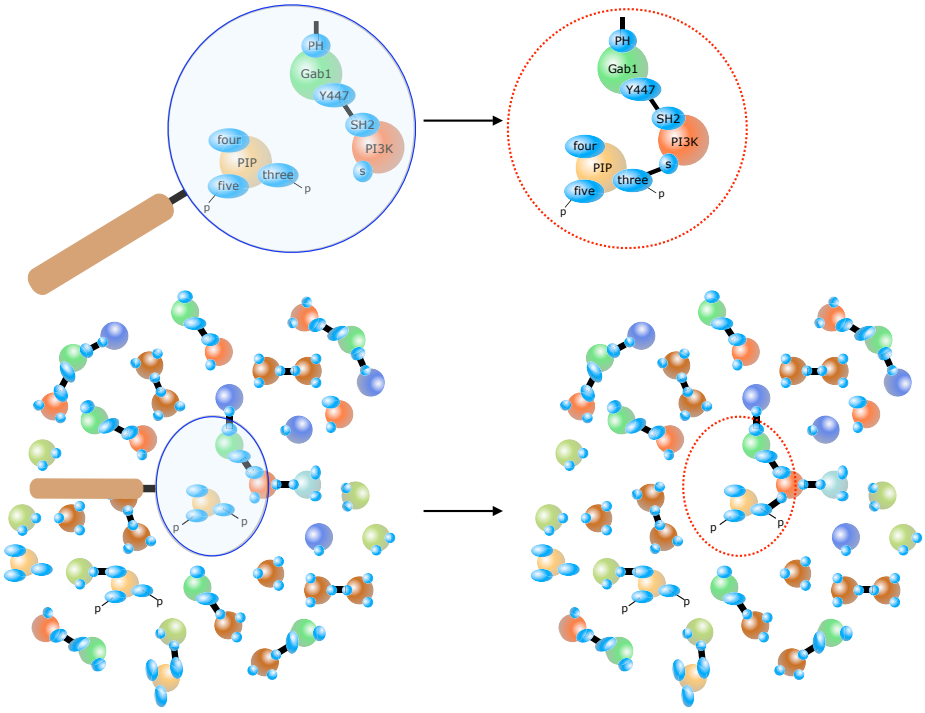


Fig. 4. Rule application in Kappa. First, a match between the pattern on the left hand side of a rule (blue lens) and the mixture (bottom) is identified. The action specified by the rule is then applied to the matching configuration, resulting in a new configuration according to the rule's right hand side (red circle). Many matchings may be possible for any given rule and many different rules may be applicable at any given moment. Rules and matchings are chosen for execution in a way that generates probabilistically correct sequences of events, following a generalization of Gillespie's algorithm for stochastic chemical kinetics [3].

We can think of a rule as an *action* that is applied to a configuration in the mixture. The action is the difference between the right hand side (rhs) and lhs of a rule. The differences may be many, such as changing several internal states and binding states at once, but they all boil down to a handful of *elementary actions* that cannot be further decomposed within the present definition of Kappa: binding, unbinding, and the change of an internal state. Kappa also allows for the creation and the removal of an agent.

Rules must obey certain constraints to be sound. Obviously, expressions E_{left} and E_{right} must be well-formed, that is, in compliance with Definitions 1.1, 1.2, and 1.3. The interpretation of a rule, however, requires a mapping of agent identities across the arrow. We must know

which agents on the right of a (textual) rule correspond to which agents on its left. There are several ways of defining such a mapping. We opted for a simple convention: both sides of a rule, E_{left} and E_{right}, are compared with one another proceeding from the left of each expression. The comparison only checks agent names and interfaces, but is blind to the states of the sites mentioned. It ends at the first difference. This procedure identifies a longest left-anchored substring – a prefix – common to both expressions. (It may be empty.) The prefix now establishes a sequential correspondence between agents on the left and right hand sides of a rule. Anything after the common prefix is interpreted in terms of deletions and introductions of agents, depending on whether an agent is missing on the right or left hand side, respectively. Subtleties of the mapping rise to the user's attention only when using textual input. A graphical interface hides them completely.

An example may help.

$$\texttt{A(x}^1\texttt{),B(x}^1\texttt{,y}_u\texttt{)} \rightarrow \texttt{A(x}^1\texttt{),B(x}^1\texttt{,y}_p\texttt{)} \qquad \{\ \text{change state of B} \qquad [3]$$

The common prefix in rule 3 establishes a correspondence between the agents mentioned on the left and the right. This rule states that if agent A is bound at site x to B at site x and B is unphosphorylated at site y (more precisely, "site y is in state u"), B will be phosphorylated at y – a common situation in signaling.

Let us now replace $\texttt{A(x}^1\texttt{),B(x}^1\texttt{,y}_p\texttt{)}$ with $\texttt{B(x}^1\texttt{,y}_p\texttt{),A(x}^1\texttt{)}$. By themselves, these expressions denote the same graph or complex. However, in the context of a rule, where a correspondence between agents on both sides has to be established to represent a set of actions, the structural equivalence, Definition 1.4, is suspended. The left and the right hand side of the rule have no common prefix, which triggers the addition and deletion actions:

$$\texttt{A(x}^1\texttt{),B(x}^1\texttt{,y}_u\texttt{)} \rightarrow \texttt{B(x}^1\texttt{,y}_p\texttt{),A(x}^1\texttt{)} \qquad \begin{cases} \text{delete the A referenced on the left} \\ \text{delete the B referenced on the left} \\ \text{add a B(x,y}_p\text{)} \\ \text{add an A(x)} \\ \text{bind B(x,y}_p\text{) at x to A(x) at x} \end{cases} \qquad [4]$$

1.4 Pattern matching. Matching is a process that establishes whether a more detailed expression E' conforms to a less detailed expression E. To gain some intuition, consider agents first. A specification A' of an agent *conforms* to a specification A, if

(i) A' and A coincide in agent name and all site names that A mentions, and
(ii) the state values ($\iota \in \mathbb{V} \mid \epsilon$) and binding values ($\lambda \in \mathbb{N} \mid - \mid ?$) of each site mentioned in A, are either equal or less specific than those mentioned in A'. With regard to binding state, '?' is less specific than ϵ or '$-$', and '$-$' is less specific than a label $i \in \mathbb{N}$. With regard to internal state, ϵ is less specific than a value $\iota \in \mathbb{V}$.

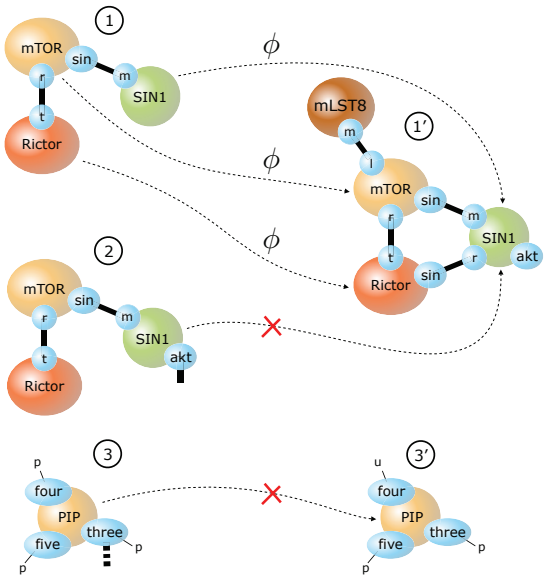


Fig. 5. Pattern matching. The embedding (fitting) of less detailed graphs on the left into more detailed graphs on the right. Graph 1 on the left embeds into complex 1' (a signaling assembly known as mTORC2, consisting of mTOR, Rictor, SIN1, and mLST8). Graph 1 agrees in all the names and states it mentions with graph 1'. Since graph 1 omits the site l from agent mTOR, l's state in graph 1' is irrelevant. The label ϕ indicates the injective mapping from agents in 1 into agents in 1'; we say graph 1 embeds into graph 1', ϕ, $G_1 \lhd_\phi G_1'$. The graphs G_1' and G_1 can be rendered in terms of their respective string expressions E_1' and E_1. Upon arranging the strings according to ϕ, the criteria in Definition 1.5 establish that E_1' conforms to E_1, $E_1' \vDash E_1$. Graph 2, in contrast, does not fit complex 1', as the former demands that SIN1 be bound to something at its site akt, but 1' specifies site akt to be free. In graph 3, agent PIP on the left does not match PIP in complex 3' on the right, as the latter has site four in an unphosphorylated state, while the former requests a phosphorylated state. There is no disagreement on site three, as graph 3 does not care about its binding state (the dotted line stands for a '?' in the textual representation, indicating "bound or unbound").

The concept of a match can be extended to expressions (mixtures) E' and E, by saying that E' conforms to E, written as $E' \vDash E$, if every agent in E' conforms to a distinct agent in E. In particular, anything conforms to an empty expression. Usually, E' is a reaction mixture,

and E is the pattern on the lhs of a rule. We next formalize the notion of "being conformant" as a satisfaction relation $\vDash$. Symbols refer to the corresponding syntactical categories as in the agent Definition 1.1. The specificity ranking of binding states is such that '?' (unknown) subsumes 'ϵ' (free) and '$-$' (bound), and '$-$' (bound) subsumes a binding label indicating a specific bond to an agent identified in the expression. Likewise, the specificity ranking of internal states is such that 'ϵ' (unspecified) subsumes any specified state. In symbols:

$$\text{binding state:} \quad ? \nearrow \epsilon \quad ? \searrow - \longrightarrow \lambda \in \mathbb{N}, \qquad \text{internal state:} \quad \epsilon \longrightarrow \iota \in \mathbb{V}$$

where the arrow means "is a superset of" (or, equivalently "is less specific than"): $x \to y \equiv x \supseteq y$. Equality applies between two λs that are identical in value. Of course, we have $\epsilon = \epsilon$ and $? = ?$ (question marks). In the following, a fraction denotes an inference from the precondition (in the numerator) to the postcondition (in the denominator), i.e., $\frac{A}{B}$ means "if A then B".

Definition 1.5 (Conforming, Matching)**.** To establish whether E' conforms to (matches) E, $E' \vDash E$, apply the following criteria:

(i)	site match	$n_{\iota'}^{\lambda'} \vDash n_{\iota}^{\lambda}$, if $\lambda' \subseteq \lambda$ and $\iota' \subseteq \iota$
(ii)	empty interface	$\sigma' \vDash \emptyset$
(iii)	interface	$\dfrac{s' \vDash s \quad \sigma' \vDash \sigma}{s', \sigma' \vDash s, \sigma}$
(iv)	agent name	$\dfrac{\sigma' \vDash \sigma}{N(\sigma') \vDash N(\sigma)}$
(v)	empty expression	$E' \vDash \varepsilon$
(vi)	expression	$\dfrac{a' \vDash a \quad E_l \vDash E_r}{a', E_l \vDash a, E_r}$

Definition 1.5 is understood as a relation between literal expressions (strings of text), established by stepping through the strings E' and E from left to right. However, finding a match of E' to E may necessitate the inspection of several structural equivalences of E', generated by reordering agents, interfaces, and relabeling bonds using Definition 1.4. It is not part of Definition 1.5 to produce such a reordering; rather, the reordering is an implicit input through the literal form of E'.

Embedding a graph into another graph. An expression E' can be represented by a graph $G' = \llcorner E' \lrcorner$, in which nodes are agents identified by their sequential position in the expression. The structural equivalence between two expressions E_1 and E', $E_1 \equiv E'$ (Definition 1.4) then induces an isomorphism *iso* between the corresponding graphs $G_1 = \llcorner E_1 \lrcorner$ and G'. A match $E' \vDash E$ asserts that E is a subsequence of E' at the level of agents (Definition 1.5 (i) also requires that the sites mentioned in an agent of E' have an internal state and binding state at least as specific as those of the sites in the corresponding agent of E). We can think of a match $E' \vDash E$ as a projection *proj* that removes agents in $G' = \llcorner E' \lrcorner$. Thus, $G_1 \xrightarrow{iso} G' \xrightarrow{proj} G$ defines an embedding of G into G_1. There may be several isomorphisms *iso* that generate distinct matches (and corresponding graph embeddings).

We go from graphs to expressions by uniquely labeling agent nodes and bonds with natural numbers. A labeled graph G then represents an expression $E = \ulcorner G \urcorner$, in which agents are written out in the order of their labels. An embedding between two graphs is given by two functions, one (f_1) for agents, the other (f_2) for bonds. f_1 is the composition of an isomorphism *iso* and an injection *inj* that maps nodes of a smaller graph into nodes of a larger graph, while preserving agent identities: $G_2 \xrightarrow{iso} G \xrightarrow{inj} G'$. There may be several isomorphisms *iso*, each of which generates an embedding. The pair (iso, f_2) constitutes a proof of structural equivalence between $E_2 = \ulcorner G_2 \urcorner$ and E, while $G \xrightarrow{inj} G'$ corresponds to the fact that $E' \vDash E$. We denote a particular embedding $\phi = iso \circ inj$ of G_2 into G' by $G_2 \lhd_\phi G'$.

Caution: we think of the less detailed graph G as being embedded ("fitted") into the more detailed graph G', while a more detailed expression E' matches a less detailed pattern E. Figure 5 provides a few graphical examples to fix the concepts. The formal definitions appear a bit involved because agents have a type (e.g. an agent of type *EGF*) as well as an identity (e.g. the agent #1 of type *EGF*, as distinct from the agent #2 of type *EGF*, in an activated EGF-receptor dimer). Intuitively, an embedding of one graph G_2 into another G' (the first not larger than the second) is a process whereby we move G_2 over G', trying to overlay G_2 on G' such that agent types match (as well as the states of the sites mentioned by the agents of G_2). Several overlays may be possible because either graph may contain multiple agents of the same type.

1.5 Replacing a pattern. The execution of a rule $E_{\text{left}} \to E_{\text{right}}$ consists in testing whether an expression S conforms to E_{left}, $S \vDash E_{\text{left}}$, as defined in section 1.4, and then overwriting (updating) the matching region in S with E_{right}. Typically, the expression S represents the contents of a reaction mixture. Here we formalize what it means to overwrite an expression E_l with another expression E_r, $E_l[E_r]$. The definition of replacement below makes use of a "null"-agent $\emptyset$ for the purpose of describing agent deletion and addition. However, we have not defined a null-agent in Definition 1.1. Instead, we shall use the following convention. Let *prefix* be the longest common left-anchored substring between the lhs and the rhs in the rule *lhs* $\to$ *rhs*, as described in section 1.3. Let L (R) be the remainder of *lhs* (*rhs*) after the *prefix*, thus, *lhs* = *prefix*,L and *rhs* = *prefix*,R. For the replacement rules to add the agents in R and delete those in L, we pad the rule with appropriately placed null-agents, $|R|$ null agents on the left and $|L|$ on the right:

$$prefix, L, \underbrace{\emptyset, \ldots, \emptyset}_{|R| \text{ times}} \longrightarrow prefix, \underbrace{\emptyset, \ldots, \emptyset}_{|L| \text{ times}}, R.$$

Proper execution of replacement must avoid capturing (i.e. duplicating) bond labels that exist elsewhere in E_l. Our implementations automatically avoid capture by relabeling using Definition 1.4.

Definition 1.6 (Replacement)**.**

(i)	overwrite state	$n_{\iota_l}^{\lambda_l}[n_{\iota_r}^{\lambda_r}] = n_{\iota_r}^{\lambda_r}$
(ii)	internal state unchanged	$n_{\iota_l}^{\lambda_l}[n^{\lambda_r}] = n_{\iota_l}^{\lambda_r}$
(iii)	interface unchanged	$\sigma[\emptyset] = \sigma$
(iv)	overwrite interface	$(s,\sigma)[s_r,\sigma_r] = s[s_r],\sigma[\sigma_r]$
(v)	overwrite agent	$N(\sigma)[N(\sigma_r)] = N(\sigma[\sigma_r])$
(vi)	agent deletion	$N(\sigma)[\emptyset] = \emptyset$
(vii)	agent introduction	$\emptyset[N(\sigma_r)] = N(\sigma_r)$
(viii)	expression unchanged	$E[\varepsilon] = E$
(ix)	overwrite expression	$(a,E)[a_r,E_r] = a[a_r],E[E_r]$

2 Self-consistency and independence

2.1 Example 1. Example 1 is depicted in Figure 6A, and consists of three agents `A(a)`, `B(b)`, and `C(a,b)` that participate in binding (and unbinding) events according to the 4 rules shown on the right (forward and backward arrows count as distinct rules):

$$\left.\begin{array}{rcl} A + C & \underset{k_{-1}^{AC}}{\overset{k_1^{A,C}}{\rightleftharpoons}} & A.C \\ A + C.B & \underset{k_{-1}^{AC}}{\overset{k_1^{A,C}}{\rightleftharpoons}} & A.C.B \end{array}\right\} \quad \texttt{A(a),C(a)} \longleftrightarrow \texttt{A(a}^{\texttt{1}}\texttt{),C(a}^{\texttt{1}}\texttt{)}$$

$$\left.\begin{array}{rcl} B + C & \underset{k_{-1}^{CB}}{\overset{k_1^{B,C}}{\rightleftharpoons}} & C.B \\ B + A.C & \underset{k_{-1}^{CB}}{\overset{k_1^{B,C}}{\rightleftharpoons}} & A.C.B \end{array}\right\} \quad \texttt{B(b),C(b)} \longleftrightarrow \texttt{B(b}^{\texttt{1}}\texttt{),C(b}^{\texttt{1}}\texttt{)} \qquad [5]$$

The four Kappa-rules on the right translate into 8 reactions shown on the left. The mutual independence of the interactions between `A` and the scaffold `C` on the one hand and `B` and `C` on the other is expressed by omitting site `a` and `b`, repectively, from the specification of these interaction rules in Kappa. As a consequence, the binding reaction between `A` and `C` (right arrow in the first rule) expands into two microscopic reactions on the left with exactly the same rate constants, k_1^{AC}. The full dynamical system is given by:

$$\begin{aligned}
\frac{d[A]}{dt} &= k_{-1}^{AC}\left([A.C] + [A.C.B]\right) - k_1^{A,C}[A]\left([C] + [C.B]\right) \\
\frac{d[B]}{dt} &= k_{-1}^{CB}\left([C.B] + [A.C.B]\right) - k_1^{B,C}[B]\left([C] + [A.C]\right) \\
\frac{d[C]}{dt} &= k_{-1}^{AC}[A.C] + k_{-1}^{CB}[C.B] - [C]\left([A]\,k_1^{A,C} + [B]\,k_1^{B,C}\right) \\
\frac{d[A.C]}{dt} &= [A]\,k_1^{A,C}[C] + k_{-1}^{CB}[A.C.B] - [A.C]\left(k_{-1}^{AC} + [B]\,k_1^{B,C}\right) \\
\frac{d[C.B]}{dt} &= k_{-1}^{AC}[A.C.B] + k_1^{B,C}[B][C] - [C.B]\left(k_{-1}^{CB} + [A]\,k_1^{A,C}\right) \\
\frac{d[A.C.B]}{dt} &= k_1^{A,C}[A][C.B] + k_1^{B,C}[B][A.C] - [A.C.B]\left(k_{-1}^{AC} + k_{-1}^{CB}\right)
\end{aligned} \qquad [6]$$

This system can be coarse-grained by defining a new set of variables that capture the fact that `A` and `B` cannot know about each other despite their interactions with a shared `C`.

$$\begin{aligned}
[A] & \\
[C*] &\doteq [C] + [C.B] \\
[A.C*] &\doteq [A.C] + [A.C.B] \\
[B] & \\
[*C] &\doteq [C] + [A.C] \\
[*C.B] &\doteq [C.B] + [A.C.B]
\end{aligned}$$

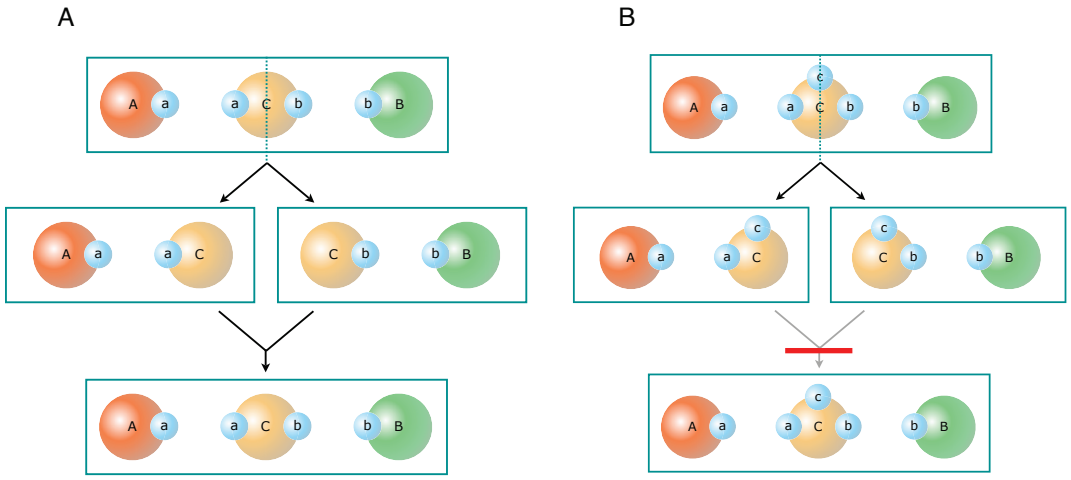


Fig. 6. Independence and self-consistency. The Figure depicts schematically Examples 1 (A) and 2 (B) described in the text. In both cases, the binding interactions on the "left" and the "right" of agent C do not influence one another. However, in case (B), agent C internally synchronizes the sites a and b through a dependency on the state of site c, but this correlation is not "readable" by the interactions that define the system. As in case (A), case (B) can be split into two self-consistently described subsystems, but they are no longer independent. The coarse-grained variables that enable the separation into subsystems cannot be used to reconstitute the microscopic dynamics.

The first group of three variables forms a self-consistent subsystem:

$$\begin{aligned}\frac{d[A]}{dt} &= k_{-1}^{AC}[A.C*] - k_{1}^{A,C}[A][C*]\\ \frac{d[C*]}{dt} &= k_{-1}^{AC}[A.C*] - k_{1}^{A,C}[A][C*]\\ \frac{d[A.C*]}{dt} &= k_{1}^{A,C}[A][C*] - k_{-1}^{AC}[A.C*]\end{aligned} \qquad [7]$$

The second group of three also forms a self-consistent subsystem:

$$\begin{aligned}\frac{d[B]}{dt} &= k_{-1}^{BC}[*C.B] - k_{1}^{B,C}[B][*C]\\ \frac{d[*C]}{dt} &= k_{-1}^{BC}[*C.B] - k_{1}^{B,C}[B][*C]\\ \frac{d[*C.B]}{dt} &= k_{1}^{B,C}[B][*C] - k_{-1}^{BC}[*C.B]\end{aligned} \qquad [8]$$

C does not propagate any information between A and B, and it does not correlate them either. The independence between the two subsystems can simply be expressed by observing that in equilibrium we must have $[A.C.B]/[C.B] = [A.C][C]$. This motivates the quantity $X \doteqdot [A.C.B][C] - [A.C][C.B]$, whose dynamics is easily obtained from the equations **[6]** as:

$$\frac{dX}{dt} = -X(k_{1}^{A,C}[A] + k_{-1}^{AC} + k_{1}^{C,B}[B] + k_{-1}^{CB}). \qquad [9]$$

The independence property can be used to invert the coarse-graining and reconstruct the original microscopic variables. In particular, when $X(0) = 0$, the reconstruction follows the simple tiling process sketched in the main text (equation **[1]**). Note that if $X(0) = 0$, then $X(t) = 0$ for all t.

2.2 Example 2. Example 2, Figure 6B, has a similar setup as Example 1, but the central scaffold C(a,c,b) now has three binding sites. The purpose of site c is to control whether sites a and b are available for binding interactions. Assume that site c of agent C has to be bound (by something) to turn on the binding capability of the other two sites. To make the point expeditiously, assume that all binding interactions are pseudo first-order because of excess A, B, and a fourth agent that binds the controller site of C, Let us also assume that all interactions are irreversible. This enables us to just focus on how agent C approaches full occupancy. Define a binding state of C as a triplet (acb) indicating the status of each site as either occupied, o, or unoccupied, u. As in Example 1, the rate constants express that binding events at site a are not influenced by binding events at site b. The system is then described by the 5 reactions shown below on the left and whose dynamics is detailed

on the right:

$$\begin{array}{l} (uuu) \xrightarrow{k^c} (uou) \\ (uou) \xrightarrow{k^l} (oou) \\ (uoo) \xrightarrow{k^l} (ooo) \\ (uou) \xrightarrow{k^r} (uoo) \\ (oou) \xrightarrow{k^r} (ooo) \end{array} \quad \text{with} \quad \begin{array}{l} \frac{d[uuu]}{dt} = -k^c[uuu] \\ \frac{d[uou]}{dt} = -(k^l + k^r)[uou] + k^c[uuu] \\ \frac{d[uoo]}{dt} = -k^l[uoo] + k^r[uou] \\ \frac{d[oou]}{dt} = k^l[uou] - k^r[oou] \\ \frac{d[ooo]}{dt} = k^l[uoo] + k^r[oou] \end{array} \qquad [10]$$

Considering the absence of any information propagation between sites a and b, we can define new coarse-grained variables:

$$\begin{array}{l} [uuu] \\ [uo*] \doteq [uou] + [uoo] \\ [oo*] \doteq [oou] + [ooo] \\ [*ou] \doteq [uou] + [oou] \\ [*oo] \doteq [uoo] + [ooo] \end{array}$$

These variables separate the system into two self-consistently described subsystems:

$$\begin{array}{l} \frac{d[uuu]}{dt} = -k^c[uuu] \\ \frac{d[uo*]}{dt} = -k^l[uo*] + k^c[uuu] \\ \frac{d[oo*]}{dt} = k^l[uo*] \end{array} \quad \text{and} \quad \begin{array}{l} \frac{d[uuu]}{dt} = -k^c[uuu] \\ \frac{d[*ou]}{dt} = -k^r[*ou] + k^c[uuu] \\ \frac{d[*oo]}{dt} = k^r[*ou] \end{array} \qquad [11]$$

Yet, because the state of the controller site c correlates the states of a and b,

$$X \doteq [ooo][uou] - [oou][uoo] = 0$$

is no longer an invariant of the dynamics:

$$\frac{dX}{dt} = -X(k^l + k^r) + k^c[ooo][uuu].$$

As indicated in the main text, the coarse-graining, which enabled us to view the system as two self-consistent subsystems, throws away correlation information, preventing reconstitution of the orignal microscopic description. From a viewpoint internal to the system, this is no loss, as the correlation cannot be observed from within the system, and a microscopic description is therefore irrelevant.

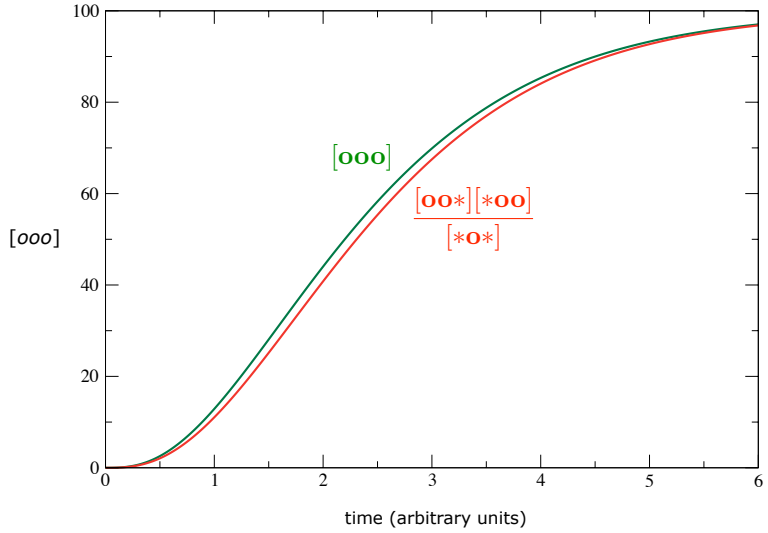


Fig. 7. Coarse-graining error. The Figure depicts the concentration dynamics of the fully occupied form of c, $[ooo]$, according to the equations **[10]** (green). The red curve shows the dynamics of $[ooo]$ when it is computed using the independence equation **[1]** in the main text, $[ooo] = [oo*][*oo]/[*o*]$. Although this relation is violated in Example 2, the utter simplicity of this scenario makes the tiling approximation still appear reasonable. This would not be the case in more complex situations.

3 The early-EGF model

We use a canonical cell signaling pathway in mammalian cells, the epidermal growth factor (EGF or ErbB) pathway, to illustrate our coarse-graining procedure. This pathway involves four receptor tyrosine kinases that interact with several extracellular ligands and each other. More than 100 proteins are implicated in the production and the processing of intracellular signals induced by EGF receptor activity. The EGF system has a controling influence on cell division, cell fate and cell morphology.

It is not our objective here to provide a useful model of the EGF pathway. In fact, the rule system below describes only a tiny set of early events that occur in EGF signaling. Our main objective, rather, is to use this small snippet of pathway to elucidate the logic of our approach to coarse-graining in rule-based representations of complex molecular interaction systems.

In essence, a signal arrives at the cell membrane in the form of a ligand, EGF (E in our rules), which binds to the extra-cellular portion of a receptor tyrosine kinase, EGFR (named R in our system), that reaches across the membrane. This binding process is the content of rule r02 (rule r01 describes the reverse process). Upon binding E, an R becomes capable of binding to a neighbouring R, also bound to a ligand (rule r03). Receptor pairs can cross-activate one another, meaning that they mutually phosphorylate certain of their intra-cellular residues (rules r05, r07). These phosphorylated residues now serve as binding sites for a variety of proteins in the cytoplasm, such as GRB2 (named G in our system) and SHC (here named S). G can bind the phosphorylated site $\mathtt{Y68}_p$ of R, as expressed in rules r12-r15. Concurrently, G can bind SOS (agent O), while G is bound to R (r16), or bound to S (r20, r22), or standalone (r18). Likewise, S can engage with R, while being free (r24, r26) or bound to G (r28, r30). The contexts of G, R, and S compound combinatorially and occasionally interfere with one another. In fact, there is a competition between two "mini-pathways" for recruiting O to the membrane receptor R: via G alone or via G bound to S. The competition derives from S and R interacting with the same binding site at G, forcing a choice for any individual G, as can be clearly seen in the contact map of the main text. These pathways can be tracked automatically with procedures that will be detailed in forthcoming manuscripts (but see [5]). All rule actions are reversible, and several actions come in contextual variants (so-called refinements) reflecting differences in rate constants.

The textual exposition of the rules below can be rendered (and edited) without loss of information in a graphical format along the lines of Figure 4A.

3.1 Rules.

r01: $\mathtt{E}(\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^1,\mathtt{r}) \longrightarrow \mathtt{E}(\mathtt{r})\,,\mathtt{R}(\mathtt{l},\mathtt{r})$
r02: $\mathtt{E}(\mathtt{r})\,,\mathtt{R}(\mathtt{l},\mathtt{r}) \longrightarrow \mathtt{E}(\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^1,\mathtt{r})$
r03: $\mathtt{E}(\mathtt{r}^2)\,,\mathtt{E}(\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^2,\mathtt{r})\,,\mathtt{R}(\mathtt{l}^1,\mathtt{r}) \longrightarrow \mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^2,\mathtt{r}^1)$
r04: $\mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^2,\mathtt{r}^1) \longrightarrow \mathtt{E}(\mathtt{r}^2)\,,\mathtt{E}(\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^2,\mathtt{r})\,,\mathtt{R}(\mathtt{l}^1,\mathtt{r})$
r05: $\mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{Y68}_u,\mathtt{l}^2,\mathtt{r}^1) \longrightarrow \mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{Y68}_p,\mathtt{l}^2,\mathtt{r}^1)$
r06: $\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{R}(\mathtt{Y68}_u)$
r07: $\mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{Y48}_u,\mathtt{l}^2,\mathtt{r}^1) \longrightarrow \mathtt{E}(\mathtt{r}^3)\,,\mathtt{E}(\mathtt{r}^2)\,,\mathtt{R}(\mathtt{l}^3,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{Y48}_p,\mathtt{l}^2,\mathtt{r}^1)$
r08: $\mathtt{R}(\mathtt{Y48}_p) \longrightarrow \mathtt{R}(\mathtt{Y48}_u)$
r09: $\mathtt{R}(\mathtt{Y48}_p^1,\mathtt{r}^-)\,,\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^1) \longrightarrow \mathtt{R}(\mathtt{Y48}_p^1,\mathtt{r}^-)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)$
r10: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^-) \longrightarrow \mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^-)$
r11: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{S}(\mathtt{Y7}_u,\mathtt{c})$
r12: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p^1)$
r13: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p)$
r14: $\mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}_p^1)$
r15: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}_p^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}_p)$
r16: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p^1)\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{R}(\mathtt{Y68}_p^2)\,,\mathtt{O}(\mathtt{d}^1)$
r17: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{R}(\mathtt{Y68}_p^2)\,,\mathtt{O}(\mathtt{d}^1) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p^1)\,,\mathtt{O}(\mathtt{d})$
r18: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^1)\,,\mathtt{O}(\mathtt{d}^1)$
r19: $\mathtt{G}(\mathtt{a},\mathtt{b}^1)\,,\mathtt{O}(\mathtt{d}^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{O}(\mathtt{d})$
r20: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c})\,,\mathtt{O}(\mathtt{d}^1)$
r21: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c})\,,\mathtt{O}(\mathtt{d}^1) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})\,,\mathtt{O}(\mathtt{d})$
r22: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d}^1)$
r23: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d}^1) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d})$
r24: $\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}) \longrightarrow \mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^1)$
r25: $\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^1) \longrightarrow \mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_u,\mathtt{c})$
r26: $\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)$
r27: $\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1) \longrightarrow \mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c})$
r28: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^1)$
r29: $\mathtt{G}(\mathtt{a}^2,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^1) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})$
r30: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c})\,,\mathtt{O}(\mathtt{d}^1) \longrightarrow \mathtt{G}(\mathtt{a}^3,\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2)$
r31: $\mathtt{G}(\mathtt{a}^3,\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c})\,,\mathtt{O}(\mathtt{d}^1)$
r32: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^1)$
r33: $\mathtt{G}(\mathtt{a}^2,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p^1)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)$
r34: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})$
r35: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c})$
r36: $\mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})$

r37: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1_p,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c})$
r38: $\mathtt{G}(\mathtt{a},\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2) \longrightarrow \mathtt{G}(\mathtt{a}^3,\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}^3_p,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2)$
r39: $\mathtt{G}(\mathtt{a}^3,\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}^3_p,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^2)\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)\,,\mathtt{O}(\mathtt{d}^2)$

3.2 Compressed rules.

Each of the rules r01-r39 is subject to a compression procedure, as outlined in the main text. The compression of r04 (expressing the dissociation of dimerized R) into cr04 has lost a symmetry. Rule r04 has a symmetry that generates two equivalent embeddings into any concrete receptor dimer present in the reaction mixture. This symmetry is detected by the simulation algorithm [3], and causes the rate constant of r04 to be adjusted by a factor of 1/2. However, the symmetry is lost upon compression. $\mathtt{R}(\mathtt{r}^-)$ on the left hand side of cr04 can still be matched in two equivalent ways by any receptor dimer in the mixture, but the simulation algorithm would forgo a division by 2, because of no detectable symmetry in the structure of cr04. This loss of symmetry is recognized by the compression process, which automatically compensates by adjusting the rate constant of cr04 to be half that of r04 (which we have arbitrarily set to 1). This is of no consequence for constructing the set of coarse-grained variables, but is required for preserving the quantitative kinetics of the rule system upon compression.

cr01: $\mathtt{R}(\mathtt{l}^-,\mathtt{r}) \longrightarrow \mathtt{R}(\mathtt{l},\mathtt{r})$
cr02: $\mathtt{E}(\mathtt{r})\,,\mathtt{R}(\mathtt{l}) \longrightarrow \mathtt{E}(\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^1)$
cr03: $\mathtt{R}(\mathtt{l}^-,\mathtt{r})\,,\mathtt{R}(\mathtt{l}^-,\mathtt{r}) \longrightarrow \mathtt{R}(\mathtt{l}^-,\mathtt{r}^1)\,,\mathtt{R}(\mathtt{l}^-,\mathtt{r}^1)$
cr04: $\mathtt{R}(\mathtt{r}^-) \longrightarrow \mathtt{R}(\mathtt{r})$ @ 0.5
cr05: $\mathtt{R}(\mathtt{Y68}^?_u,\mathtt{r}^-) \longrightarrow \mathtt{R}(\mathtt{Y68}^?_p,\mathtt{r}^-)$
cr06: $\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{R}(\mathtt{Y68}_u)$
cr07: $\mathtt{R}(\mathtt{Y48}^?_u,\mathtt{r}^-) \longrightarrow \mathtt{R}(\mathtt{Y48}^?_p,\mathtt{r}^-)$
cr08: $\mathtt{R}(\mathtt{Y48}_p) \longrightarrow \mathtt{R}(\mathtt{Y48}_u)$
cr09: $\mathtt{R}(\mathtt{Y48}^1,\mathtt{r}^-)\,,\mathtt{S}(\mathtt{Y7}^?_u,\mathtt{c}^1) \longrightarrow \mathtt{R}(\mathtt{Y48}^1,\mathtt{r}^-)\,,\mathtt{S}(\mathtt{Y7}^?_p,\mathtt{c}^1)$
cr10: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^-) \longrightarrow \mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^-)$
cr11: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{S}(\mathtt{Y7}_u,\mathtt{c})$
cr12: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}^1_p)$
cr13: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{R}(\mathtt{Y68})$
cr14: $\mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}_p) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}^1_p)$
cr15: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}^1) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68})$
cr16: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}^1)\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{R}(\mathtt{Y68}^2)\,,\mathtt{O}(\mathtt{d}^1)$
cr17: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y68}^1) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y68}^1)$
cr18: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^1)\,,\mathtt{O}(\mathtt{d}^1)$
cr19: $\mathtt{G}(\mathtt{a},\mathtt{b}^-) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})$
cr20: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c})\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}^2,\mathtt{c})\,,\mathtt{O}(\mathtt{d}^1)$
cr21: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c})$
cr22: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)\,,\mathtt{S}(\mathtt{Y7}^2,\mathtt{c}^-)\,,\mathtt{O}(\mathtt{d}^1)$
cr23: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-)$
cr24: $\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}^?_u,\mathtt{c}) \longrightarrow \mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}^?_u,\mathtt{c}^1)$
cr25: $\mathtt{S}(\mathtt{Y7}^?_u,\mathtt{c}^-) \longrightarrow \mathtt{S}(\mathtt{Y7}^?_u,\mathtt{c})$
cr26: $\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)$
cr27: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^-) \longrightarrow \mathtt{S}(\mathtt{Y7}_p,\mathtt{c})$
cr28: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b})\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}^2,\mathtt{c}^1)$
cr29: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c})$
cr30: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y48}_p)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^2,\mathtt{b}^-)\,,\mathtt{R}(\mathtt{Y48}^1_p)\,,\mathtt{S}(\mathtt{Y7}^2,\mathtt{c}^1)$
cr31: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c})$
cr32: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1_p,\mathtt{c}^-)$
cr33: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7},\mathtt{c}^-)$
cr34: $\mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1_p,\mathtt{c})$
cr35: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b})\,,\mathtt{S}(\mathtt{Y7},\mathtt{c})$
cr36: $\mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1_p,\mathtt{c})$
cr37: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7},\mathtt{c})$
cr38: $\mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1_p,\mathtt{c}^-)$
cr39: $\mathtt{G}(\mathtt{a}^1,\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7}^1,\mathtt{c}^-) \longrightarrow \mathtt{G}(\mathtt{a},\mathtt{b}^-)\,,\mathtt{S}(\mathtt{Y7},\mathtt{c}^-)$

3.3 Fragments.

In this section we list the 38 self-consistent coarse-grained variables generated by the automatic procedure outlined in the main text and in section 4. These 38 variables form a dynamical system (shown in section 5 whose state at any time t is identical to the state attained by the microscopic dynamics (involving 356 variables) and subsequent coarse-graining.

$\mathcal{F}$01: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^2)$, $\mathtt{G}(\mathtt{a}^3,\mathtt{b}^4)$, $\mathtt{O}(\mathtt{d}^4)$
$\mathcal{F}$02: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^2)$, $\mathtt{G}(\mathtt{a}^3,\mathtt{b}^4)$, $\mathtt{O}(\mathtt{d}^4)$
$\mathcal{F}$03: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)$, $\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c}^3)$, $\mathtt{R}(\mathtt{Y48}_p^3,\mathtt{l},\mathtt{r})$, $\mathtt{O}(\mathtt{d}^1)$
$\mathcal{F}$04: $\mathtt{R}(\mathtt{Y48}_p^1,\mathtt{l},\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^1)$
$\mathcal{F}$05: $\mathtt{G}(\mathtt{a},\mathtt{b}^1)$, $\mathtt{O}(\mathtt{d}^1)$
$\mathcal{F}$06: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^2)$
$\mathcal{F}$07: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c}^2)$
$\mathcal{F}$08: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)$, $\mathtt{S}(\mathtt{Y7}_p^2,\mathtt{c})$, $\mathtt{O}(\mathtt{d}^1)$
$\mathcal{F}$09: $\mathtt{S}(\mathtt{Y7}_p,\mathtt{c})$
$\mathcal{F}$10: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})$, $\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c})$
$\mathcal{F}$11: $\mathtt{G}(\mathtt{a},\mathtt{b})$
$\mathcal{F}$12: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^2)$, $\mathtt{G}(\mathtt{a}^3,\mathtt{b})$
$\mathcal{F}$13: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_p^3,\mathtt{c}^2)$, $\mathtt{G}(\mathtt{a}^3,\mathtt{b})$
$\mathcal{F}$14: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})$, $\mathtt{S}(\mathtt{Y7}_p^1,\mathtt{c}^2)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l},\mathtt{r})$
$\mathcal{F}$15: $\mathtt{R}(\mathtt{Y48}_p,\mathtt{l},\mathtt{r})$
$\mathcal{F}$16: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p,\mathtt{l}^1,\mathtt{r})$
$\mathcal{F}$17: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$
$\mathcal{F}$18: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^2)$
$\mathcal{F}$19: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^2)$
$\mathcal{F}$20: $\mathtt{R}(\mathtt{Y48}_p^1,\mathtt{l},\mathtt{r})$, $\mathtt{S}(\mathtt{Y7}_u,\mathtt{c}^1)$
$\mathcal{F}$21: $\mathtt{S}(\mathtt{Y7}_u,\mathtt{c})$
$\mathcal{F}$22: $\mathtt{O}(\mathtt{d})$
$\mathcal{F}$23: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^3)$, $\mathtt{O}(\mathtt{d}^3)$
$\mathcal{F}$24: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^3)$, $\mathtt{O}(\mathtt{d}^3)$
$\mathcal{F}$25: $\mathtt{G}(\mathtt{a}^2,\mathtt{b}^1)$, $\mathtt{R}(\mathtt{Y68}_p^2,\mathtt{l},\mathtt{r})$, $\mathtt{O}(\mathtt{d}^1)$
$\mathcal{F}$26: $\mathtt{G}(\mathtt{a}^1,\mathtt{b})$, $\mathtt{R}(\mathtt{Y68}_p^1,\mathtt{l},\mathtt{r})$
$\mathcal{F}$27: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p^2,\mathtt{l}^1,\mathtt{r})$, $\mathtt{G}(\mathtt{a}^2,\mathtt{b})$
$\mathcal{F}$28: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p^2,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$, $\mathtt{G}(\mathtt{a}^2,\mathtt{b})$
$\mathcal{F}$29: $\mathtt{R}(\mathtt{Y68}_p,\mathtt{l},\mathtt{r})$
$\mathcal{F}$30: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p,\mathtt{l}^1,\mathtt{r})$
$\mathcal{F}$31: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_p,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$
$\mathcal{F}$32: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_u,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$
$\mathcal{F}$33: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y48}_u,\mathtt{l}^1,\mathtt{r})$
$\mathcal{F}$34: $\mathtt{R}(\mathtt{Y48}_u,\mathtt{l},\mathtt{r})$
$\mathcal{F}$35: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_u,\mathtt{l}^1,\mathtt{r}^{\mathtt{R@r}})$
$\mathcal{F}$36: $\mathtt{E}(\mathtt{r}^1)$, $\mathtt{R}(\mathtt{Y68}_u,\mathtt{l}^1,\mathtt{r})$
$\mathcal{F}$37: $\mathtt{R}(\mathtt{Y68}_u,\mathtt{l},\mathtt{r})$
$\mathcal{F}$38: $\mathtt{E}(\mathtt{r})$

4 Translating rules into a dynamical system for fragments

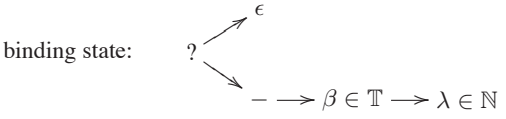
4.1 Partial Complex. A partial complex is a connected graph (a component) that occurs on either side of a rule. In our static analysis, semi-links in partial complexes are internally expanded into all possible binding partners, and labeled with a bond type of the form *partner@site*. For example, $\mathtt{R(Y48}_p^1\mathtt{),S(Y7}^2\mathtt{,c}^1\mathtt{),G(a}^2\mathtt{,b}^{O@d}\mathtt{)}$ is a partial complex (it is the right hand side of cr30). Thus:

Definition 4.1 (Partial complex)**.**

A partially specified complex (or partial complex for short) is a connected expression of Kappa agents, such that

1. the set of sites shown for agent A is a subset of the interface of A
2. the internal state of a site may be omitted
3. the binding state of a site must be any of (i) free, (ii) bound, or (iii) labeled with a bond type indicating the names of the bound agent and its binding site

We extend the concept of a match, $E' \vDash E$, Definition 1.5, or the concept of an embedding $G \lhd_\phi G'$ (see end of section 1.4), to expressions containing a bond type by simply extending the specificity ranking of binding states (section 1.4). Let $\mathbb{T}$ denote the set of binding type labels, and let β be a binding type variable:



Everything else then follows the axiomatization in section 1.4.

Fragments are defined in the main text by the instructions Frag1-Frag3. Fragments are partial complexes, too, but whose shape is constrained by the annotated contact map (ACM). Our goal in this section is to sketch the construction of the kinetic system of differential equations describing the concentration dynamics for fragments. To this end, we must evaluate how each rule in a model contributes to the production and consumption of fragments. For a rule to be translatable into a set of reactions between fragments, we must ensure that any fragment that properly intersects a component on the lhs of a rule, and whose intersection contains a site that is modified by the action, must contain that component. This is achieved by the syntactical criteria, Cov1-Cov3 and Edg1, in conjunction with the fragment construction directives Frag1-Frag3, as explained in the main text.

4.2 Every partial complex occurring in a rule is contained in some fragment. By construction, fragments are more refined than partial complexes that occur on the left hand side of rules, except for trivial dissociation rules. Directives Cov1-Cov3 ensure that there always is a class in the covering of an agent that contains at least as many sites as any occurrence of that agent on the left of a rule. The fragment construction directives Frag2-Frag3 then guarantee that fragments extend at least as much as the partial complexes that constitute the pattern components on the left hand side of (non-trivial) rules. By construction, fragments do not contain bonds that are dotted in the ACM, and thus do not extend partial complexes with such bonds. However, when a partial complex is cut at a dotted bond, and the bond is replaced with two bond type labels acting as stubs, each piece can be embedded in a fragment. This is reflective of the fact that there are no correlations between these pieces that the system of rules could detect. The construction of fragments also ensures that when a partial complex Z contains a site that is modified, Z is contained in each fragment exhibiting that site.

4.3 Expressing the concentration of a partial complex in terms of fragment concentrations. We always can use fragment concentrations to express the concentration of a partial complex embedded in a given fragment. However, we cannot, in general, extend a fragment and express its concentration using other fragments, as this woud require the independence conditions, equation **[1]** in the main text, to hold. As we saw in Example 2, these conditions do not hold in general.

To compute the concentration of a partial compex, we need to use fragments that extend it. But we must be careful about which fragments we use. A partial complex and a fragment, each identify a set of fully specified molecular species into which they embed. The concentration of a partial complex is the sum total of the concentrations of these species. To express the concentration of a partial complex as a combination of fragment concentrations, we must ensure that the fragments used in the combination partition the set of molecular species into which the partial complex expands, or we would overcount. A set of fragments that complies with this requirement is called *orthogonal*.

Orthogonal fragments. Let Z be a partial complex, $\mathfrak{F}$ the set of fragments, and $\mathcal{F}_1, \mathcal{F}_2 \in \mathfrak{F}$ two fragments that contain Z: $Z \lhd_\phi \mathcal{F}_1$ and $Z \lhd_{\phi'} \mathcal{F}_2$. (See section 1.4, paragraph labeled "embedding", for a definition of the embedding relation $\lhd_\phi$.) We define two fragments $\mathcal{F}_1, \mathcal{F}_2$ that contain Z as "orthogonal", when the agents on which $\mathcal{F}_1$ and $\mathcal{F}_2$ agree exhibit the same sites. Orthogonal fragments differ with regard to internal states and binding states at sites of agents they have in common, and, thus, constitute a set of patterns whose matching instances in a reaction mixture do *not* overlap. This is important for computing the concentration of a partial complex Z in terms of fragments, since we must avoid counting matching instances of Z in the reaction mixture more than once. The set of fragments $\{\mathcal{F}_1, \ldots, \mathcal{F}_n\}$ from which we compute the concentration of Z should constitute a refinement of Z, in the sense of partitioning the matching instances of Z.

The formal definition of orthogonality makes use of the concept of a formal path. A formal path p is a set of symbolic instructions for navigating through a graph representing a Kappa complex. Starting at node (agent) `A` and folowing the directives provided by p will lead us to a unique target node `T`, which we denote by $\mathtt{A}.p$. p could be expressed as a sequence of bonds to travel. For example, $\mathtt{A}.p$ might result in $\mathtt{A.a.s}_1\mathtt{.B}_1\mathtt{.s}_2 \ldots \mathtt{s}_{2i-1}\mathtt{.B}_i\mathtt{.s}_{2i} \ldots \mathtt{.t.T}$, which is a path that goes from the originating agent `A` to agent $\mathtt{B}_1$ over a link between site `a` of `A` and site $\mathtt{s}_1$ of agent $\mathtt{B}_1$, and from there to agent $\mathtt{B}_2$ over a link between site $\mathtt{s}_2$ of agent $\mathtt{B}_1$ and site $\mathtt{s}_3$ of agent $\mathtt{B}_2$, and so on, to enter the target agent `T` through its site `t`. (This specification of a path is somewhat redundant, since we don't need to mention the agents, as the bound site at which we leave an agent uniquely determines the agent we end up in.) An empty path means that we stay at the originating agent `A`.

Definition 4.2 (Orthogonal fragments)**.**

Let Z be a partial complex. Let $(\mathcal{F}_1, \phi)$ and $(\mathcal{F}_2, \phi')$ be two fragments and respective matching injections, such that $Z \lhd_\phi \mathcal{F}_1$ and $Z \lhd_{\phi'} \mathcal{F}_2$. $(\mathcal{F}_1, \phi)$ and $(\mathcal{F}_2, \phi')$ are orthogonal with respect to Z, written as $(\mathcal{F}_1, \phi) \bowtie_Z (\mathcal{F}_2, \phi')$, if and only if for any agent A of Z and any path p at least one of these statements is true:

1. $\phi(A).p$ is not defined in $\mathcal{F}_1$.
2. $\phi'(A).p$ is not defined in $\mathcal{F}_2$.
3. $\phi(A).p$ and $\phi'(A).p$ are both defined in $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and have the same set of sites.

A fragment $\mathcal{F}_i$ may be matched in more than one way by a partial complex Z, yielding more than one injection map ϕ. Let $\phi_{i,k}$ be the k-th embedding of Z in $\mathcal{F}_i$. Let us collect all fragments that contain Z and all ways in which they contain Z. Define $C(Z)$ as a largest set of such fragments (it need not be unique) that are mutually (Z-)orthogonal. Thus, $C(Z) = \{(\mathcal{F}_1, \phi_{1,1}), \ldots, (\mathcal{F}_1, \phi_{1,n_1}), \ldots, (\mathcal{F}_m, \phi_{m,1}), \ldots, (\mathcal{F}_m, \phi_{m,n_m})\}$, such that $(\mathcal{F}_i, \phi_{i,k}) \bowtie_Z (\mathcal{F}_j, \phi_{j,l})$ for all $i, j = 1, \ldots, m$, $k = 1, n_i$, and $l = 1, n_j$. Denote the set of fragments in $C(Z)$ by $C_{\mathcal{F}}(Z) = \{\mathcal{F} \mid \exists \phi \text{ s.t. } (\mathcal{F}, \phi) \in C(Z)\}$. The concentration of Z can now be expressed in terms of fragment concentrations as:

$$[Z] = \frac{1}{|\{\phi | Z \lhd_\phi Z\}|} \sum_{\mathcal{F}_i \in C_{\mathcal{F}}(Z)} [\mathcal{F}_i]\, n_i \qquad \textbf{[12]}$$

Equation **[12]** formalizes our intuition that the concentration of a partial complex Z is the sum of the concentrations of the fragments that contain it, times a multiplicity counting the number of ways in which Z matches a given fragment. The only complexity comes from chosing the fragments over which the sum runs in such a way that no two fragments overlap in the set of molecular instances they match. That is what the orthogonality criterion is meant to ensure. Finally, we divide by the number of automorphisms of Z to compensate for any symmetries in Z.

4.4 Assembling the dynamical system for fragments. We assemble the system of differential equations for the fragments by determining the mass action terms that each rule type contributes. For the sake of simplifying exposition, we shall only be concerned with rules consisting of at most two components on the lhs, and whose action modifies a single internal state or a single binding state. It is straightforward to generalize this to n components and to multiple actions within the same rule. The k after the @-sign refers to the rate constant of the rule. Since we shall build a term for the fragment dynamics from each component on the lhs of a rule separately, the rate constant γ that enters the fragment dynamics must compensate for the number of automorphisms, *auto(lhs)*, of the lhs: $\gamma = k/auto(lhs)$. Whenever we refer to the concentration of a partial complex, $[Z]$, we mean its expansion into fragment concentrations according to equation **[12]**. To indicate that we are building up differential equations sequentially, we shall use the symbol $\stackrel{+}{=}$ as meaning "add this term to the previous ones for this equation".

1. $\mathbf{Z}, \mathbf{Z}' \longrightarrow \mathbf{Z}^*, \mathbf{Z}' \,@\, \mathbf{k}$

This type of rule modifies the partial complex Z.

Consumption terms. The kinetic equation of each fragment that contains Z gains a consumption term

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma [\mathcal{F}_i][Z'].$$

The universal quantifier over ϕ means that the rate at which $\mathcal{F}_i$ is consumed depends on the number of ways that Z can be emdedded in $\mathcal{F}_i$.

Production terms. The kinetic equation of each fragment containing Z^* gains a production term

$$\forall \mathcal{F}_k \; \forall \phi \text{ s.t. } Z^* \lhd_\phi \mathcal{F}_k \text{ and } \mathcal{F}_i \text{ s.t. } Z \lhd_{f(\phi)} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^* \; : \quad \frac{d[\mathcal{F}_k]}{dt} \stackrel{+}{=} \gamma [\mathcal{F}_i][Z'].$$

Clearly, the fragments $\mathcal{F}_k$ and $\mathcal{F}_i$ must be related by the rule action; $\mathcal{F}_k = \mathcal{F}_i^*$ means that the fragment $\mathcal{F}_k$ is obtained by applying the rule action to fragment $\mathcal{F}_i$. The notation $f(\phi)$ indexing the embedding of Z into $\mathcal{F}_i$ is meant to specify that this embedding is related to the embedding ϕ of Z^* into $\mathcal{F}_k$, since the relatedness of Z to its modified form Z^* forces not only a relatedness of $\mathcal{F}_i$ to $\mathcal{F}_k$, but also of the way these fragments extend the corresponding partial complexes Z and Z^*.

2. $\mathbf{Z}, \mathbf{Z}' \longrightarrow \mathbf{Z}$–$\mathbf{Z}' \,@\, \mathbf{k}$

This type of rule binds the partial complexes Z and Z'.

Consumption terms. The kinetic equation of each fragment that contains Z gains a consumption term

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma [\mathcal{F}_i][Z'].$$

Likewise for Z':

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z' \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma [\mathcal{F}_i][Z].$$

Production terms. On the production side, we must distinguish between the case that the link in the ACM is of type "solid" and the case that it is of type "dotted".

solid link

$$\forall \mathcal{F}_k \; \forall \phi_1, \phi_2 \text{ s.t. } Z\text{–}Z' \lhd_{\phi_1 \uplus \phi_2} \mathcal{F}_k \text{ and } \mathcal{F}_i \text{ s.t. } Z \lhd_{f(\phi_1)} \mathcal{F}_i \text{ and } \mathcal{F}_j \text{ s.t. } Z' \lhd_{f(\phi_2)} \mathcal{F}_j \text{ and } \mathcal{F}_k = \mathcal{F}_i\text{–}\mathcal{F}_j \; : \quad \frac{d[\mathcal{F}_k]}{dt} \stackrel{+}{=} \gamma [\mathcal{F}_i][\mathcal{F}_j].$$

Again, we must express that the embeddings of Z–Z', Z, and Z' into $\mathcal{F}_k$, $\mathcal{F}_i$, and $\mathcal{F}_j$, respectively, are related. The notation $\phi_1 \uplus \phi_2$ denotes the disjoint sum of ϕ_1 and ϕ_2: The domains of ϕ_1, $dom(\phi_1)$, and ϕ_2, $dom(\phi_2)$, have an empty intersection, and $(\phi_1 \uplus \phi_2)(x) = \phi_1$ if $x \in dom(\phi_1)$ or $(\phi_1 \uplus \phi_2)(x) = \phi_2(x)$ if $x \in dom(\phi_2)$.

dotted link

Assume the bond to be between site a of $A \in Z$ and site b of $B \in Z'$, and let $Z^{B@b}$ and $Z'^{A@a}$ denote the partial complexes obtained from severing the bond in Z–Z' and replacing it with a binding-type label.

$$\forall \mathcal{F}_k \; \forall \phi \text{ s.t. } Z^{B@b} \lhd_\phi \mathcal{F}_k \text{ and } \mathcal{F}_i \text{ s.t. } Z \lhd_{f(\phi)} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{B@b} \; : \quad \frac{d[\mathcal{F}_k]}{dt} \stackrel{+}{=} \gamma [\mathcal{F}_i][Z'].$$

Likewise for Z':

$$\forall \mathcal{F}_k \; \forall \phi \text{ s.t. } Z'^{A@a} \lhd_\phi \mathcal{F}_k \text{ and } \mathcal{F}_i \text{ s.t. } Z' \lhd_{f(\phi)} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{A@a} \; : \quad \frac{d[\mathcal{F}_k]}{dt} \stackrel{+}{=} \gamma [\mathcal{F}_i][Z].$$

3. $\mathbf{Z}$–$\mathbf{Z}' \longrightarrow \mathbf{Z}, \mathbf{Z}' \,@\, \mathbf{k}$

This type of rule dissociates the partial complex Z–Z'.

Consumption terms. On the consumption side, we must distinguish between the case that the link in the ACM is of type "solid" and the case that it is of type "dotted".

solid link

The kinetic equation of each fragment that contains Z–Z' gains a consumption term

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z\text{–}Z' \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma [\mathcal{F}_i].$$

dotted link (By definition, this is a "trivial dissociation rule".)

As above, assume the bond to be between site a of $A \in Z$ and site b of $B \in Z'$, and let $Z^{B@b}$ and $Z'^{A@a}$ denote the partial complexes obtained from severing the bond in Z–Z' and replacing it with a bond type label.

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z^{B@b} \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma[\mathcal{F}_i].$$

Likewise for Z':

$$\forall \mathcal{F}_i \; \forall \phi \text{ s.t. } Z'^{A@a} \lhd_\phi \mathcal{F}_i \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} -\gamma[\mathcal{F}_i].$$

Production terms. Here, too, we must distinguish between the case that the link in the ACM is of type "solid" and the case that it is of type "dotted".

solid link

$$\forall \mathcal{F}_i \; \forall \phi_Z \text{ s.t. } Z \lhd_{\phi_Z} \mathcal{F}_i \text{ and } \forall \mathcal{F}_k \; \forall \phi_{Z'} \text{ s.t. } \mathcal{F}_i\text{–}Z' \lhd_{f(\phi_Z) \uplus \phi_{Z'}} \mathcal{F}_k \text{ and } (\mathcal{F}_k, f(\phi_Z) \uplus \phi_{Z'}) \in C(\mathcal{F}_i\text{–}Z') \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} \gamma[\mathcal{F}_k].$$

The above might benefit from a verbal expansion. In the reaction type Z–$Z' \rightarrow Z, Z'$, pick a fragment $\mathcal{F}_i$ that extends the partial complex Z in a particular way (that's an instance of the first two universal quantifiers). The production rate of $\mathcal{F}_i$ will be first order in a fragment $\mathcal{F}_k$ that extends the single partial complex on the left hand side of the rule, Z–Z'. Yet, the $\mathcal{F}_k$ in question cannot extend any old Z–Z', but must extend an instance that contains the $\mathcal{F}_i$ that will emerge after the bond is split. Hence the conditon that $\mathcal{F}_i$–$Z' \lhd_{f(\phi_Z) \uplus \phi_{Z'}} \mathcal{F}_k$. The injection map associated with this embedding must be constrained by how we chose $\mathcal{F}_i$ to extend Z. Finally, all $\mathcal{F}_k$ that contribute to the production of a given $\mathcal{F}_i$ must be mutually orthogonal with respect to $\mathcal{F}_i$–Z', as defined in section 4.3, to avoid multiple-counting the molecular species into which the $\mathcal{F}_k$ expand.

Analogous production terms arise for fragments that extend Z' on the right hand side of the rule.

dotted link

As above, assume the bond to be between site a of $A \in Z$ and site b of $B \in Z'$, and let $Z^{B@b}$ and $Z'^{A@a}$ denote the partial complexes obtained from severing the bond in Z–Z' and replacing it with a bond type label.

$$\forall \mathcal{F}_k \; \forall \phi \text{ s.t. } Z^{B@b} \lhd_\phi \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z \lhd_{f(\phi)} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{B@b} \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} \gamma[\mathcal{F}_k].$$

Likewise for Z':

$$\forall \mathcal{F}_k \; \forall \phi \text{ s.t. } Z'^{A@a} \lhd_\phi \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z' \lhd_{f(\phi)} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{A@a} \; : \quad \frac{d[\mathcal{F}_i]}{dt} \stackrel{+}{=} \gamma[\mathcal{F}_k].$$

Our implementation is actually much more straightforward than this section suggests, because the algorithm makes direct use of the embeddings ϕ, which remain abstract in a notation that does not exploit the structure of expressions.

5 The dynamical system for the early EGF model

This section shows the output generated by our automatic procedure for the rule system r01-r39 listed in section 3.1. The results have been obtained entirely by static analysis, as detailed in the main text and section 4 of this Supporting Information. The dynamical system for fragments constitutes an endogenously coarse-grained and self-consistent system that is sound with respect to the microscopic kinetics. Sound means that the outcome is identical whether one first executes the deterministic microscopic kinetics with subsequent coarse-graining or first coarse-grains with subsequent execution of the fragment dynamics. Note that the microscopic system was *never* represented explicitly (and thus never executed). It was only represented implicitly by the system of rules (which were not executed either). Because of the ability to bypass an explicit representation, the causal analysis of microscopic systems involving astronomic numbers of distinct microscopic states (molecular species) becomes possible.

The entire fragmentation of the early EGF example, beginning with the reachability analysis, followed by rule compression, fragmentation, and dynamical system generation took less than 0.2s on a 2GHz Intel Centrino Duo with 1Gb RAM. The mass action terms for the fragment dynamics resulting from each rule, as well as the fully assembled dynamical system, are listed in the next section.

5.1 List of kinetic terms generated from each rule for each fragment. We report the kinetic production and consumption terms for each fragment as generated by analysis of the compressed rule system of section 3.2. $\mathcal{R}^{\text{rule}}_{\text{frag}}$ denotes the kinetic terms pertinent to the dynamical equation for the fragment indicated in the subscript, and which result from the rule identified in the superscript. Thus $\mathcal{R}^{39}_7 = \mathcal{F}_1$ means that our static analysis generates from compressed rule cr39 one unimolecular production term (involving fragment $\mathcal{F}_1$) for fragment $\mathcal{F}_7$. For the sake of a less cluttered presentation, we have set all rate constants to 1. (The right hand side of each $\mathcal{R}^{\text{rule}}_{\text{frag}}$ equation should be multiplied by the rate constant associated with the rule indicated in the superscript.)

Kinetic terms generated from rule cr39:

$\mathcal{R}^{39}_7 = \mathcal{F}_1$
$\mathcal{R}^{39}_6 = \mathcal{F}_2$
$\mathcal{R}^{39}_5 = \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3$
$\mathcal{R}^{39}_4 = \mathcal{F}_3$
$\mathcal{R}^{39}_3 = -\mathcal{F}_3$

$\mathcal{R}_2^{39} = -\mathcal{F}_2$
$\mathcal{R}_1^{39} = -\mathcal{F}_1$

Kinetic terms generated from rule cr38:

$\mathcal{R}_7^{38} = -\mathcal{F}_5 \cdot \mathcal{F}_7$
$\mathcal{R}_6^{38} = -\mathcal{F}_5 \cdot \mathcal{F}_6$
$\mathcal{R}_5^{38} = -\mathcal{F}_5 \cdot (\mathcal{F}_4 + \mathcal{F}_6 + \mathcal{F}_7)$
$\mathcal{R}_4^{38} = -\mathcal{F}_4 \cdot \mathcal{F}_5$
$\mathcal{R}_3^{38} = \mathcal{F}_4 \cdot \mathcal{F}_5$
$\mathcal{R}_2^{38} = \mathcal{F}_5 \cdot \mathcal{F}_6$
$\mathcal{R}_1^{38} = \mathcal{F}_5 \cdot \mathcal{F}_7$

Kinetic terms generated from rule cr37:

$\mathcal{R}_9^{37} = \mathcal{F}_8$
$\mathcal{R}_8^{37} = -\mathcal{F}_8$
$\mathcal{R}_5^{37} = \mathcal{F}_8$

Kinetic terms generated from rule cr36:

$\mathcal{R}_9^{36} = -\mathcal{F}_5 \cdot \mathcal{F}_9$
$\mathcal{R}_8^{36} = \mathcal{F}_5 \cdot \mathcal{F}_9$
$\mathcal{R}_5^{36} = -\mathcal{F}_5 \cdot \mathcal{F}_9$

Kinetic terms generated from rule cr35:

$\mathcal{R}_{11}^{35} = \mathcal{F}_{10}$
$\mathcal{R}_{10}^{35} = -\mathcal{F}_{10}$
$\mathcal{R}_9^{35} = \mathcal{F}_{10}$

Kinetic terms generated from rule cr34:

$\mathcal{R}_{11}^{34} = -\mathcal{F}_9 \cdot \mathcal{F}_{11}$
$\mathcal{R}_{10}^{34} = \mathcal{F}_9 \cdot \mathcal{F}_{11}$
$\mathcal{R}_9^{34} = -\mathcal{F}_9 \cdot \mathcal{F}_{11}$

Kinetic terms generated from rule cr33:

$\mathcal{R}_{14}^{33} = -\mathcal{F}_{14}$
$\mathcal{R}_{13}^{33} = -\mathcal{F}_{13}$
$\mathcal{R}_{12}^{33} = -\mathcal{F}_{12}$
$\mathcal{R}_{11}^{33} = \mathcal{F}_{12} + \mathcal{F}_{13} + \mathcal{F}_{14}$
$\mathcal{R}_7^{33} = \mathcal{F}_{12}$
$\mathcal{R}_6^{33} = \mathcal{F}_{13}$
$\mathcal{R}_4^{33} = \mathcal{F}_{14}$

Kinetic terms generated from rule cr32:

$\mathcal{R}_{14}^{32} = \mathcal{F}_4 \cdot \mathcal{F}_{11}$
$\mathcal{R}_{13}^{32} = \mathcal{F}_6 \cdot \mathcal{F}_{11}$
$\mathcal{R}_{12}^{32} = \mathcal{F}_7 \cdot \mathcal{F}_{11}$
$\mathcal{R}_{11}^{32} = -\mathcal{F}_{11} \cdot (\mathcal{F}_4 + \mathcal{F}_6 + \mathcal{F}_7)$
$\mathcal{R}_7^{32} = -\mathcal{F}_7 \cdot \mathcal{F}_{11}$
$\mathcal{R}_6^{32} = -\mathcal{F}_6 \cdot \mathcal{F}_{11}$
$\mathcal{R}_4^{32} = -\mathcal{F}_4 \cdot \mathcal{F}_{11}$

Kinetic terms generated from rule cr31:

$\mathcal{R}_{17}^{31} = \mathcal{F}_1$
$\mathcal{R}_{16}^{31} = \mathcal{F}_2$

$\mathcal{R}_{15}^{31} = \mathcal{F}_3$
$\mathcal{R}_{8}^{31} = \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3$
$\mathcal{R}_{3}^{31} = -\mathcal{F}_3$
$\mathcal{R}_{2}^{31} = -\mathcal{F}_2$
$\mathcal{R}_{1}^{31} = -\mathcal{F}_1$

Kinetic terms generated from rule cr30:

$\mathcal{R}_{17}^{30} = -\mathcal{F}_8 \cdot \mathcal{F}_{17}$
$\mathcal{R}_{16}^{30} = -\mathcal{F}_8 \cdot \mathcal{F}_{16}$
$\mathcal{R}_{15}^{30} = -\mathcal{F}_8 \cdot \mathcal{F}_{15}$
$\mathcal{R}_{8}^{30} = -\mathcal{F}_8 \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17})$
$\mathcal{R}_{3}^{30} = \mathcal{F}_8 \cdot \mathcal{F}_{15}$
$\mathcal{R}_{2}^{30} = \mathcal{F}_8 \cdot \mathcal{F}_{16}$
$\mathcal{R}_{1}^{30} = \mathcal{F}_8 \cdot \mathcal{F}_{17}$

Kinetic terms generated from rule cr29:

$\mathcal{R}_{17}^{29} = \mathcal{F}_{12}$
$\mathcal{R}_{16}^{29} = \mathcal{F}_{13}$
$\mathcal{R}_{15}^{29} = \mathcal{F}_{14}$
$\mathcal{R}_{14}^{29} = -\mathcal{F}_{14}$
$\mathcal{R}_{13}^{29} = -\mathcal{F}_{13}$
$\mathcal{R}_{12}^{29} = -\mathcal{F}_{12}$
$\mathcal{R}_{10}^{29} = \mathcal{F}_{12} + \mathcal{F}_{13} + \mathcal{F}_{14}$

Kinetic terms generated from rule cr28:

$\mathcal{R}_{17}^{28} = -\mathcal{F}_{10} \cdot \mathcal{F}_{17}$
$\mathcal{R}_{16}^{28} = -\mathcal{F}_{10} \cdot \mathcal{F}_{16}$
$\mathcal{R}_{15}^{28} = -\mathcal{F}_{10} \cdot \mathcal{F}_{15}$
$\mathcal{R}_{14}^{28} = \mathcal{F}_{10} \cdot \mathcal{F}_{15}$
$\mathcal{R}_{13}^{28} = \mathcal{F}_{10} \cdot \mathcal{F}_{16}$
$\mathcal{R}_{12}^{28} = \mathcal{F}_{10} \cdot \mathcal{F}_{17}$
$\mathcal{R}_{10}^{28} = -\mathcal{F}_{10} \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17})$

Kinetic terms generated from rule cr27:

$\mathcal{R}_{17}^{27} = \mathcal{F}_7$
$\mathcal{R}_{16}^{27} = \mathcal{F}_6$
$\mathcal{R}_{15}^{27} = \mathcal{F}_4$
$\mathcal{R}_{9}^{27} = \mathcal{F}_4 + \mathcal{F}_6 + \mathcal{F}_7$
$\mathcal{R}_{7}^{27} = -\mathcal{F}_7$
$\mathcal{R}_{6}^{27} = -\mathcal{F}_6$
$\mathcal{R}_{4}^{27} = -\mathcal{F}_4$

Kinetic terms generated from rule cr26:

$\mathcal{R}_{17}^{26} = -\mathcal{F}_9 \cdot \mathcal{F}_{17}$
$\mathcal{R}_{16}^{26} = -\mathcal{F}_9 \cdot \mathcal{F}_{16}$
$\mathcal{R}_{15}^{26} = -\mathcal{F}_9 \cdot \mathcal{F}_{15}$
$\mathcal{R}_{9}^{26} = -\mathcal{F}_9 \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17})$
$\mathcal{R}_{7}^{26} = \mathcal{F}_9 \cdot \mathcal{F}_{17}$
$\mathcal{R}_{6}^{26} = \mathcal{F}_9 \cdot \mathcal{F}_{16}$
$\mathcal{R}_{4}^{26} = \mathcal{F}_9 \cdot \mathcal{F}_{15}$

Kinetic terms generated from rule cr25:

$\mathcal{R}_{21}^{25} = \mathcal{F}_{18} + \mathcal{F}_{19} + \mathcal{F}_{20}$

$\mathcal{R}_{20}^{25} = -\mathcal{F}_{20}$
$\mathcal{R}_{19}^{25} = -\mathcal{F}_{19}$
$\mathcal{R}_{18}^{25} = -\mathcal{F}_{18}$
$\mathcal{R}_{17}^{25} = \mathcal{F}_{18}$
$\mathcal{R}_{16}^{25} = \mathcal{F}_{19}$
$\mathcal{R}_{15}^{25} = \mathcal{F}_{20}$

Kinetic terms generated from rule cr24:

$\mathcal{R}_{21}^{24} = -\mathcal{F}_{21} \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17})$
$\mathcal{R}_{20}^{24} = \mathcal{F}_{15} \cdot \mathcal{F}_{21}$
$\mathcal{R}_{19}^{24} = \mathcal{F}_{16} \cdot \mathcal{F}_{21}$
$\mathcal{R}_{18}^{24} = \mathcal{F}_{17} \cdot \mathcal{F}_{21}$
$\mathcal{R}_{17}^{24} = -\mathcal{F}_{17} \cdot \mathcal{F}_{21}$
$\mathcal{R}_{16}^{24} = -\mathcal{F}_{16} \cdot \mathcal{F}_{21}$
$\mathcal{R}_{15}^{24} = -\mathcal{F}_{15} \cdot \mathcal{F}_{21}$

Kinetic terms generated from rule cr23:

$\mathcal{R}_{22}^{23} = \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3$
$\mathcal{R}_{14}^{23} = \mathcal{F}_3$
$\mathcal{R}_{13}^{23} = \mathcal{F}_2$
$\mathcal{R}_{12}^{23} = \mathcal{F}_1$
$\mathcal{R}_{3}^{23} = -\mathcal{F}_3$
$\mathcal{R}_{2}^{23} = -\mathcal{F}_2$
$\mathcal{R}_{1}^{23} = -\mathcal{F}_1$

Kinetic terms generated from rule cr22:

$\mathcal{R}_{22}^{22} = -\mathcal{F}_{22} \cdot (\mathcal{F}_{12} + \mathcal{F}_{13} + \mathcal{F}_{14})$
$\mathcal{R}_{14}^{22} = -\mathcal{F}_{14} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{13}^{22} = -\mathcal{F}_{13} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{12}^{22} = -\mathcal{F}_{12} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{3}^{22} = \mathcal{F}_{14} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{2}^{22} = \mathcal{F}_{13} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{1}^{22} = \mathcal{F}_{12} \cdot \mathcal{F}_{22}$

Kinetic terms generated from rule cr21:

$\mathcal{R}_{22}^{21} = \mathcal{F}_8$
$\mathcal{R}_{10}^{21} = \mathcal{F}_8$
$\mathcal{R}_{8}^{21} = -\mathcal{F}_8$

Kinetic terms generated from rule cr20:

$\mathcal{R}_{22}^{20} = -\mathcal{F}_{10} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{10}^{20} = -\mathcal{F}_{10} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{8}^{20} = \mathcal{F}_{10} \cdot \mathcal{F}_{22}$

Kinetic terms generated from rule cr19:

$\mathcal{R}_{22}^{19} = \mathcal{F}_5$
$\mathcal{R}_{11}^{19} = \mathcal{F}_5$
$\mathcal{R}_{5}^{19} = -\mathcal{F}_5$

Kinetic terms generated from rule cr18:

$\mathcal{R}_{22}^{18} = -\mathcal{F}_{11} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{11}^{18} = -\mathcal{F}_{11} \cdot \mathcal{F}_{22}$
$\mathcal{R}_{5}^{18} = \mathcal{F}_{11} \cdot \mathcal{F}_{22}$

Kinetic terms generated from rule cr17:

$\mathcal{R}_{28}^{17} = \mathcal{F}_{23}$
$\mathcal{R}_{27}^{17} = \mathcal{F}_{24}$
$\mathcal{R}_{26}^{17} = \mathcal{F}_{25}$
$\mathcal{R}_{25}^{17} = -\mathcal{F}_{25}$
$\mathcal{R}_{24}^{17} = -\mathcal{F}_{24}$
$\mathcal{R}_{23}^{17} = -\mathcal{F}_{23}$
$\mathcal{R}_{22}^{17} = \mathcal{F}_{23} + \mathcal{F}_{24} + \mathcal{F}_{25}$

Kinetic terms generated from rule cr16:

$\mathcal{R}_{28}^{16} = -\mathcal{F}_{22} \cdot \mathcal{F}_{28}$
$\mathcal{R}_{27}^{16} = -\mathcal{F}_{22} \cdot \mathcal{F}_{27}$
$\mathcal{R}_{26}^{16} = -\mathcal{F}_{22} \cdot \mathcal{F}_{26}$
$\mathcal{R}_{25}^{16} = \mathcal{F}_{22} \cdot \mathcal{F}_{26}$
$\mathcal{R}_{24}^{16} = \mathcal{F}_{22} \cdot \mathcal{F}_{27}$
$\mathcal{R}_{23}^{16} = \mathcal{F}_{22} \cdot \mathcal{F}_{28}$
$\mathcal{R}_{22}^{16} = -\mathcal{F}_{22} \cdot (\mathcal{F}_{26} + \mathcal{F}_{27} + \mathcal{F}_{28})$

Kinetic terms generated from rule cr15:

$\mathcal{R}_{31}^{15} = \mathcal{F}_{23}$
$\mathcal{R}_{30}^{15} = \mathcal{F}_{24}$
$\mathcal{R}_{29}^{15} = \mathcal{F}_{25}$
$\mathcal{R}_{25}^{15} = -\mathcal{F}_{25}$
$\mathcal{R}_{24}^{15} = -\mathcal{F}_{24}$
$\mathcal{R}_{23}^{15} = -\mathcal{F}_{23}$
$\mathcal{R}_{5}^{15} = \mathcal{F}_{23} + \mathcal{F}_{24} + \mathcal{F}_{25}$

Kinetic terms generated from rule cr14:

$\mathcal{R}_{31}^{14} = -\mathcal{F}_{5} \cdot \mathcal{F}_{31}$
$\mathcal{R}_{30}^{14} = -\mathcal{F}_{5} \cdot \mathcal{F}_{30}$
$\mathcal{R}_{29}^{14} = -\mathcal{F}_{5} \cdot \mathcal{F}_{29}$
$\mathcal{R}_{25}^{14} = \mathcal{F}_{5} \cdot \mathcal{F}_{29}$
$\mathcal{R}_{24}^{14} = \mathcal{F}_{5} \cdot \mathcal{F}_{30}$
$\mathcal{R}_{23}^{14} = \mathcal{F}_{5} \cdot \mathcal{F}_{31}$
$\mathcal{R}_{5}^{14} = -\mathcal{F}_{5} \cdot (\mathcal{F}_{29} + \mathcal{F}_{30} + \mathcal{F}_{31})$

Kinetic terms generated from rule cr13:

$\mathcal{R}_{31}^{13} = \mathcal{F}_{28}$
$\mathcal{R}_{30}^{13} = \mathcal{F}_{27}$
$\mathcal{R}_{29}^{13} = \mathcal{F}_{26}$
$\mathcal{R}_{28}^{13} = -\mathcal{F}_{28}$
$\mathcal{R}_{27}^{13} = -\mathcal{F}_{27}$
$\mathcal{R}_{26}^{13} = -\mathcal{F}_{26}$
$\mathcal{R}_{11}^{13} = \mathcal{F}_{26} + \mathcal{F}_{27} + \mathcal{F}_{28}$

Kinetic terms generated from rule cr12:

$\mathcal{R}_{31}^{12} = -\mathcal{F}_{11} \cdot \mathcal{F}_{31}$
$\mathcal{R}_{30}^{12} = -\mathcal{F}_{11} \cdot \mathcal{F}_{30}$
$\mathcal{R}_{29}^{12} = -\mathcal{F}_{11} \cdot \mathcal{F}_{29}$
$\mathcal{R}_{28}^{12} = \mathcal{F}_{11} \cdot \mathcal{F}_{31}$
$\mathcal{R}_{27}^{12} = \mathcal{F}_{11} \cdot \mathcal{F}_{30}$
$\mathcal{R}_{26}^{12} = \mathcal{F}_{11} \cdot \mathcal{F}_{29}$

$\mathcal{R}_{11}^{12} = -\mathcal{F}_{11} \cdot (\mathcal{F}_{29} + \mathcal{F}_{30} + \mathcal{F}_{31})$

Kinetic terms generated from rule cr11:

$\mathcal{R}_{21}^{11} = \mathcal{F}_9$
$\mathcal{R}_{9}^{11} = -\mathcal{F}_9$

Kinetic terms generated from rule cr10:

$\mathcal{R}_{20}^{10} = \mathcal{F}_4$
$\mathcal{R}_{19}^{10} = \mathcal{F}_6$
$\mathcal{R}_{18}^{10} = \mathcal{F}_7$
$\mathcal{R}_{7}^{10} = -\mathcal{F}_7$
$\mathcal{R}_{6}^{10} = -\mathcal{F}_6$
$\mathcal{R}_{4}^{10} = -\mathcal{F}_4$

Kinetic terms generated from rule cr9:

$\mathcal{R}_{18}^{9} = -\mathcal{F}_{18}$
$\mathcal{R}_{7}^{9} = \mathcal{F}_{18}$

Kinetic terms generated from rule cr8:

$\mathcal{R}_{34}^{8} = \mathcal{F}_{15}$
$\mathcal{R}_{33}^{8} = \mathcal{F}_{16}$
$\mathcal{R}_{32}^{8} = \mathcal{F}_{17}$
$\mathcal{R}_{17}^{8} = -\mathcal{F}_{17}$
$\mathcal{R}_{16}^{8} = -\mathcal{F}_{16}$
$\mathcal{R}_{15}^{8} = -\mathcal{F}_{15}$

Kinetic terms generated from rule cr7:

$\mathcal{R}_{32}^{7} = -\mathcal{F}_{32}$
$\mathcal{R}_{17}^{7} = \mathcal{F}_{32}$

Kinetic terms generated from rule cr6:

$\mathcal{R}_{37}^{6} = \mathcal{F}_{29}$
$\mathcal{R}_{36}^{6} = \mathcal{F}_{30}$
$\mathcal{R}_{35}^{6} = \mathcal{F}_{31}$
$\mathcal{R}_{31}^{6} = -\mathcal{F}_{31}$
$\mathcal{R}_{30}^{6} = -\mathcal{F}_{30}$
$\mathcal{R}_{29}^{6} = -\mathcal{F}_{29}$

Kinetic terms generated from rule cr5:

$\mathcal{R}_{35}^{5} = -\mathcal{F}_{35}$
$\mathcal{R}_{31}^{5} = \mathcal{F}_{35}$

Kinetic terms generated from rule cr4:

$\mathcal{R}_{36}^{4} = \mathcal{F}_{35}$
$\mathcal{R}_{35}^{4} = -\mathcal{F}_{35}$
$\mathcal{R}_{33}^{4} = \mathcal{F}_{32}$
$\mathcal{R}_{32}^{4} = -\mathcal{F}_{32}$
$\mathcal{R}_{31}^{4} = -\mathcal{F}_{31}$
$\mathcal{R}_{30}^{4} = \mathcal{F}_{31}$
$\mathcal{R}_{28}^{4} = -\mathcal{F}_{28}$
$\mathcal{R}_{27}^{4} = \mathcal{F}_{28}$
$\mathcal{R}_{24}^{4} = \mathcal{F}_{23}$
$\mathcal{R}_{23}^{4} = -\mathcal{F}_{23}$

$\mathcal{R}^4_{19} = \mathcal{F}_{18}$
$\mathcal{R}^4_{18} = -\mathcal{F}_{18}$
$\mathcal{R}^4_{17} = -\mathcal{F}_{17}$
$\mathcal{R}^4_{16} = \mathcal{F}_{17}$
$\mathcal{R}^4_{13} = \mathcal{F}_{12}$
$\mathcal{R}^4_{12} = -\mathcal{F}_{12}$
$\mathcal{R}^4_{7} = -\mathcal{F}_{7}$
$\mathcal{R}^4_{6} = \mathcal{F}_{7}$
$\mathcal{R}^4_{2} = \mathcal{F}_{1}$
$\mathcal{R}^4_{1} = -\mathcal{F}_{1}$

Kinetic terms generated from rule cr3:

$\mathcal{R}^3_{36} = -\mathcal{F}_{36} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{35} = \mathcal{F}_{36} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{33} = -\mathcal{F}_{33} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{32} = \mathcal{F}_{33} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{31} = \mathcal{F}_{30} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{30} = -\mathcal{F}_{30} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{28} = \mathcal{F}_{27} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{27} = -\mathcal{F}_{27} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{24} = -\mathcal{F}_{24} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{23} = \mathcal{F}_{24} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{19} = -\mathcal{F}_{19} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{18} = \mathcal{F}_{19} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{17} = \mathcal{F}_{16} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{16} = -\mathcal{F}_{16} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{13} = -\mathcal{F}_{13} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{12} = \mathcal{F}_{13} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{7} = \mathcal{F}_{6} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{6} = -\mathcal{F}_{6} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{2} = -\mathcal{F}_{2} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$
$\mathcal{R}^3_{1} = \mathcal{F}_{2} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$

Kinetic terms generated from rule cr2:

$\mathcal{R}^2_{38} = -\mathcal{F}_{38} \cdot (\mathcal{F}_{25} + \mathcal{F}_{26} + \mathcal{F}_{29} + \mathcal{F}_{37})$
$\mathcal{R}^2_{37} = -\mathcal{F}_{37} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{36} = \mathcal{F}_{37} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{34} = -\mathcal{F}_{34} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{33} = \mathcal{F}_{34} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{30} = \mathcal{F}_{29} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{29} = -\mathcal{F}_{29} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{27} = \mathcal{F}_{26} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{26} = -\mathcal{F}_{26} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{25} = -\mathcal{F}_{25} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{24} = \mathcal{F}_{25} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{20} = -\mathcal{F}_{20} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{19} = \mathcal{F}_{20} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{16} = \mathcal{F}_{15} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{15} = -\mathcal{F}_{15} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{14} = -\mathcal{F}_{14} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{13} = \mathcal{F}_{14} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{6} = \mathcal{F}_{4} \cdot \mathcal{F}_{38}$
$\mathcal{R}^2_{4} = -\mathcal{F}_{4} \cdot \mathcal{F}_{38}$

$\mathcal{R}_3^2 = -\mathcal{F}_3 \cdot \mathcal{F}_{38}$
$\mathcal{R}_2^2 = \mathcal{F}_3 \cdot \mathcal{F}_{38}$

Kinetic terms generated from rule cr1:

$\mathcal{R}_{38}^1 = \mathcal{F}_2 + \mathcal{F}_6 + \mathcal{F}_{13} + \mathcal{F}_{16} + \mathcal{F}_{19} + \mathcal{F}_{33}$
$\mathcal{R}_{37}^1 = \mathcal{F}_{36}$
$\mathcal{R}_{36}^1 = -\mathcal{F}_{36}$
$\mathcal{R}_{34}^1 = \mathcal{F}_{33}$
$\mathcal{R}_{33}^1 = -\mathcal{F}_{33}$
$\mathcal{R}_{30}^1 = -\mathcal{F}_{30}$
$\mathcal{R}_{29}^1 = \mathcal{F}_{30}$
$\mathcal{R}_{27}^1 = -\mathcal{F}_{27}$
$\mathcal{R}_{26}^1 = \mathcal{F}_{27}$
$\mathcal{R}_{25}^1 = \mathcal{F}_{24}$
$\mathcal{R}_{24}^1 = -\mathcal{F}_{24}$
$\mathcal{R}_{20}^1 = \mathcal{F}_{19}$
$\mathcal{R}_{19}^1 = -\mathcal{F}_{19}$
$\mathcal{R}_{16}^1 = -\mathcal{F}_{16}$
$\mathcal{R}_{15}^1 = \mathcal{F}_{16}$
$\mathcal{R}_{14}^1 = \mathcal{F}_{13}$
$\mathcal{R}_{13}^1 = -\mathcal{F}_{13}$
$\mathcal{R}_6^1 = -\mathcal{F}_6$
$\mathcal{R}_4^1 = \mathcal{F}_6$
$\mathcal{R}_3^1 = \mathcal{F}_2$
$\mathcal{R}_2^1 = -\mathcal{F}_2$

5.2 The dynamical system for fragments.

$\dot{\mathcal{F}}_1 = \mathcal{R}_1^3 + \mathcal{R}_1^4 + \mathcal{R}_1^{22} + \mathcal{R}_1^{23} + \mathcal{R}_1^{30} + \mathcal{R}_1^{31} + \mathcal{R}_1^{38} + \mathcal{R}_1^{39}$
$\dot{\mathcal{F}}_2 = \mathcal{R}_2^1 + \mathcal{R}_2^2 + \mathcal{R}_2^3 + \mathcal{R}_2^4 + \mathcal{R}_2^{22} + \mathcal{R}_2^{23} + \mathcal{R}_2^{30} + \mathcal{R}_2^{31} + \mathcal{R}_2^{38} + \mathcal{R}_2^{39}$
$\dot{\mathcal{F}}_3 = \mathcal{R}_3^1 + \mathcal{R}_3^2 + \mathcal{R}_3^{22} + \mathcal{R}_3^{23} + \mathcal{R}_3^{30} + \mathcal{R}_3^{31} + \mathcal{R}_3^{38} + \mathcal{R}_3^{39}$
$\dot{\mathcal{F}}_4 = \mathcal{R}_4^1 + \mathcal{R}_4^2 + \mathcal{R}_4^{10} + \mathcal{R}_4^{26} + \mathcal{R}_4^{27} + \mathcal{R}_4^{32} + \mathcal{R}_4^{33} + \mathcal{R}_4^{38} + \mathcal{R}_4^{39}$
$\dot{\mathcal{F}}_5 = \mathcal{R}_5^{14} + \mathcal{R}_5^{15} + \mathcal{R}_5^{18} + \mathcal{R}_5^{19} + \mathcal{R}_5^{36} + \mathcal{R}_5^{37} + \mathcal{R}_5^{38} + \mathcal{R}_5^{39}$
$\dot{\mathcal{F}}_6 = \mathcal{R}_6^1 + \mathcal{R}_6^2 + \mathcal{R}_6^3 + \mathcal{R}_6^4 + \mathcal{R}_6^{10} + \mathcal{R}_6^{26} + \mathcal{R}_6^{27} + \mathcal{R}_6^{32} + \mathcal{R}_6^{33} + \mathcal{R}_6^{38} + \mathcal{R}_6^{39}$
$\dot{\mathcal{F}}_7 = \mathcal{R}_7^3 + \mathcal{R}_7^4 + \mathcal{R}_7^9 + \mathcal{R}_7^{10} + \mathcal{R}_7^{26} + \mathcal{R}_7^{27} + \mathcal{R}_7^{32} + \mathcal{R}_7^{33} + \mathcal{R}_7^{38} + \mathcal{R}_7^{39}$
$\dot{\mathcal{F}}_8 = \mathcal{R}_8^{20} + \mathcal{R}_8^{21} + \mathcal{R}_8^{30} + \mathcal{R}_8^{31} + \mathcal{R}_8^{36} + \mathcal{R}_8^{37}$
$\dot{\mathcal{F}}_9 = \mathcal{R}_9^{11} + \mathcal{R}_9^{26} + \mathcal{R}_9^{27} + \mathcal{R}_9^{34} + \mathcal{R}_9^{35} + \mathcal{R}_9^{36} + \mathcal{R}_9^{37}$
$\dot{\mathcal{F}}_{10} = \mathcal{R}_{10}^{20} + \mathcal{R}_{10}^{21} + \mathcal{R}_{10}^{28} + \mathcal{R}_{10}^{29} + \mathcal{R}_{10}^{34} + \mathcal{R}_{10}^{35}$
$\dot{\mathcal{F}}_{11} = \mathcal{R}_{11}^{12} + \mathcal{R}_{11}^{13} + \mathcal{R}_{11}^{18} + \mathcal{R}_{11}^{19} + \mathcal{R}_{11}^{32} + \mathcal{R}_{11}^{33} + \mathcal{R}_{11}^{34} + \mathcal{R}_{11}^{35}$
$\dot{\mathcal{F}}_{12} = \mathcal{R}_{12}^3 + \mathcal{R}_{12}^4 + \mathcal{R}_{12}^{22} + \mathcal{R}_{12}^{23} + \mathcal{R}_{12}^{28} + \mathcal{R}_{12}^{29} + \mathcal{R}_{12}^{32} + \mathcal{R}_{12}^{33}$
$\dot{\mathcal{F}}_{13} = \mathcal{R}_{13}^1 + \mathcal{R}_{13}^2 + \mathcal{R}_{13}^3 + \mathcal{R}_{13}^4 + \mathcal{R}_{13}^{22} + \mathcal{R}_{13}^{23} + \mathcal{R}_{13}^{28} + \mathcal{R}_{13}^{29} + \mathcal{R}_{13}^{32} + \mathcal{R}_{13}^{33}$
$\dot{\mathcal{F}}_{14} = \mathcal{R}_{14}^1 + \mathcal{R}_{14}^2 + \mathcal{R}_{14}^{22} + \mathcal{R}_{14}^{23} + \mathcal{R}_{14}^{28} + \mathcal{R}_{14}^{29} + \mathcal{R}_{14}^{32} + \mathcal{R}_{14}^{33}$
$\dot{\mathcal{F}}_{15} = \mathcal{R}_{15}^1 + \mathcal{R}_{15}^2 + \mathcal{R}_{15}^8 + \mathcal{R}_{15}^{24} + \mathcal{R}_{15}^{25} + \mathcal{R}_{15}^{26} + \mathcal{R}_{15}^{27} + \mathcal{R}_{15}^{28} + \mathcal{R}_{15}^{29} + \mathcal{R}_{15}^{30} + \mathcal{R}_{15}^{31}$
$\dot{\mathcal{F}}_{16} = \mathcal{R}_{16}^1 + \mathcal{R}_{16}^2 + \mathcal{R}_{16}^3 + \mathcal{R}_{16}^4 + \mathcal{R}_{16}^8 + \mathcal{R}_{16}^{24} + \mathcal{R}_{16}^{25} + \mathcal{R}_{16}^{26} + \mathcal{R}_{16}^{27} + \mathcal{R}_{16}^{28} + \mathcal{R}_{16}^{29} + \mathcal{R}_{16}^{30} + \mathcal{R}_{16}^{31}$
$\dot{\mathcal{F}}_{17} = \mathcal{R}_{17}^3 + \mathcal{R}_{17}^4 + \mathcal{R}_{17}^7 + \mathcal{R}_{17}^8 + \mathcal{R}_{17}^{24} + \mathcal{R}_{17}^{25} + \mathcal{R}_{17}^{26} + \mathcal{R}_{17}^{27} + \mathcal{R}_{17}^{28} + \mathcal{R}_{17}^{29} + \mathcal{R}_{17}^{30} + \mathcal{R}_{17}^{31}$
$\dot{\mathcal{F}}_{18} = \mathcal{R}_{18}^3 + \mathcal{R}_{18}^4 + \mathcal{R}_{18}^9 + \mathcal{R}_{18}^{10} + \mathcal{R}_{18}^{24} + \mathcal{R}_{18}^{25}$
$\dot{\mathcal{F}}_{19} = \mathcal{R}_{19}^1 + \mathcal{R}_{19}^2 + \mathcal{R}_{19}^3 + \mathcal{R}_{19}^4 + \mathcal{R}_{19}^{10} + \mathcal{R}_{19}^{24} + \mathcal{R}_{19}^{25}$
$\dot{\mathcal{F}}_{20} = \mathcal{R}_{20}^1 + \mathcal{R}_{20}^2 + \mathcal{R}_{20}^{10} + \mathcal{R}_{20}^{24} + \mathcal{R}_{20}^{25}$
$\dot{\mathcal{F}}_{21} = \mathcal{R}_{21}^{11} + \mathcal{R}_{21}^{24} + \mathcal{R}_{21}^{25}$

$$\dot{\mathcal{F}}_{22} = \mathcal{R}_{22}^{16} + \mathcal{R}_{22}^{17} + \mathcal{R}_{22}^{18} + \mathcal{R}_{22}^{19} + \mathcal{R}_{22}^{20} + \mathcal{R}_{22}^{21} + \mathcal{R}_{22}^{22} + \mathcal{R}_{22}^{23}$$
$$\dot{\mathcal{F}}_{23} = \mathcal{R}_{23}^{3} + \mathcal{R}_{23}^{4} + \mathcal{R}_{23}^{14} + \mathcal{R}_{23}^{15} + \mathcal{R}_{23}^{16} + \mathcal{R}_{23}^{17}$$
$$\dot{\mathcal{F}}_{24} = \mathcal{R}_{24}^{1} + \mathcal{R}_{24}^{2} + \mathcal{R}_{24}^{3} + \mathcal{R}_{24}^{4} + \mathcal{R}_{24}^{14} + \mathcal{R}_{24}^{15} + \mathcal{R}_{24}^{16} + \mathcal{R}_{24}^{17}$$
$$\dot{\mathcal{F}}_{25} = \mathcal{R}_{25}^{1} + \mathcal{R}_{25}^{2} + \mathcal{R}_{25}^{14} + \mathcal{R}_{25}^{15} + \mathcal{R}_{25}^{16} + \mathcal{R}_{25}^{17}$$
$$\dot{\mathcal{F}}_{26} = \mathcal{R}_{26}^{1} + \mathcal{R}_{26}^{2} + \mathcal{R}_{26}^{12} + \mathcal{R}_{26}^{13} + \mathcal{R}_{26}^{16} + \mathcal{R}_{26}^{17}$$
$$\dot{\mathcal{F}}_{27} = \mathcal{R}_{27}^{1} + \mathcal{R}_{27}^{2} + \mathcal{R}_{27}^{3} + \mathcal{R}_{27}^{4} + \mathcal{R}_{27}^{12} + \mathcal{R}_{27}^{13} + \mathcal{R}_{27}^{16} + \mathcal{R}_{27}^{17}$$
$$\dot{\mathcal{F}}_{28} = \mathcal{R}_{28}^{3} + \mathcal{R}_{28}^{4} + \mathcal{R}_{28}^{12} + \mathcal{R}_{28}^{13} + \mathcal{R}_{28}^{16} + \mathcal{R}_{28}^{17}$$
$$\dot{\mathcal{F}}_{29} = \mathcal{R}_{29}^{1} + \mathcal{R}_{29}^{2} + \mathcal{R}_{29}^{6} + \mathcal{R}_{29}^{12} + \mathcal{R}_{29}^{13} + \mathcal{R}_{29}^{14} + \mathcal{R}_{29}^{15}$$
$$\dot{\mathcal{F}}_{30} = \mathcal{R}_{30}^{1} + \mathcal{R}_{30}^{2} + \mathcal{R}_{30}^{3} + \mathcal{R}_{30}^{4} + \mathcal{R}_{30}^{6} + \mathcal{R}_{30}^{12} + \mathcal{R}_{30}^{13} + \mathcal{R}_{30}^{14} + \mathcal{R}_{30}^{15}$$
$$\dot{\mathcal{F}}_{31} = \mathcal{R}_{31}^{3} + \mathcal{R}_{31}^{4} + \mathcal{R}_{31}^{5} + \mathcal{R}_{31}^{6} + \mathcal{R}_{31}^{12} + \mathcal{R}_{31}^{13} + \mathcal{R}_{31}^{14} + \mathcal{R}_{31}^{15}$$
$$\dot{\mathcal{F}}_{32} = \mathcal{R}_{32}^{3} + \mathcal{R}_{32}^{4} + \mathcal{R}_{32}^{7} + \mathcal{R}_{32}^{8}$$
$$\dot{\mathcal{F}}_{33} = \mathcal{R}_{33}^{1} + \mathcal{R}_{33}^{2} + \mathcal{R}_{33}^{3} + \mathcal{R}_{33}^{4} + \mathcal{R}_{33}^{8}$$
$$\dot{\mathcal{F}}_{34} = \mathcal{R}_{34}^{1} + \mathcal{R}_{34}^{2} + \mathcal{R}_{34}^{8}$$
$$\dot{\mathcal{F}}_{35} = \mathcal{R}_{35}^{3} + \mathcal{R}_{35}^{4} + \mathcal{R}_{35}^{5} + \mathcal{R}_{35}^{6}$$
$$\dot{\mathcal{F}}_{36} = \mathcal{R}_{36}^{1} + \mathcal{R}_{36}^{2} + \mathcal{R}_{36}^{3} + \mathcal{R}_{36}^{4} + \mathcal{R}_{36}^{6}$$
$$\dot{\mathcal{F}}_{37} = \mathcal{R}_{37}^{1} + \mathcal{R}_{37}^{2} + \mathcal{R}_{37}^{6}$$
$$\dot{\mathcal{F}}_{38} = \mathcal{R}_{38}^{1} + \mathcal{R}_{38}^{2}$$

1. **Danos, V & Laneve, C. *Formal Molecular Biology* (2004) *Theoretical Computer Science* 325, 69–110.**
2. **Plectix BioSystems, Inc. (http://www.plectix.com). Contact support@plectix.com for instructions on how to access the Kappa modeling platform.**
3. **Danos, V, Feret, J, Fontana, W, & Krivine, J. (2007) *Scalable simulation of cellular signalling networks*, Lecture Notes in Computer Science. (Springer), Vol. 4807, pp. 139–157.**
4. **Danos, V, Feret, J, Fontana, W, & Krivine, J. (2008) *Abstract interpretation of cellular signalling networks*, Lecture Notes in Computer Science. (Springer), Vol. 4905, pp. 83–97.**
5. **Danos, V, Feret, J, Fontana, W, Harmer, R, & Krivine, J. (2007) *Rule-based modelling of cellular signalling*, Lecture Notes in Computer Science. (Springer, Lisboa, Portugal), Vol. 4703, pp. 17–41.**