

Supporting Information

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

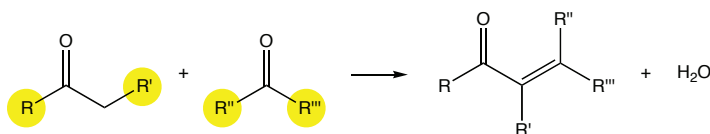
*Harvard Medical School, Boston, USA, [†]University of Edinburgh, Edinburgh, United Kingdom, and [‡]CNRS & Paris Diderot, Paris, France

1 Kappa

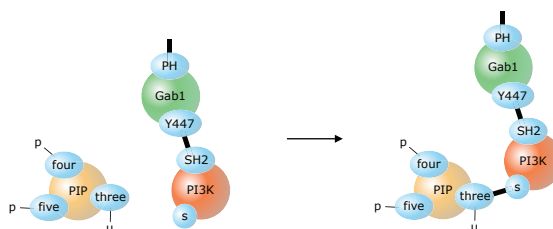
Kappa is a formal language for representing molecular objects as agents [1, 2]. These agents are decorated with sites that carry modifiable state and/or bind sites of other agents to form complexes. Empirical observations about the states of proteins that permit them to interact in specific ways are expressed as rules at a level of abstraction consistent with how molecular biologists approach networks of protein-protein interactions. 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, Figure 1. In Kappa, irrelevant context is 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. In this section, we provide a formal syntax of the language and an equivalent graphical rendering, Figure 1.

A: rule

chemical reaction pattern

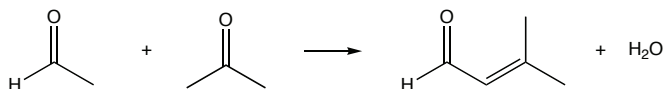


protein interaction pattern



B: instance

reaction instance



protein interaction instance

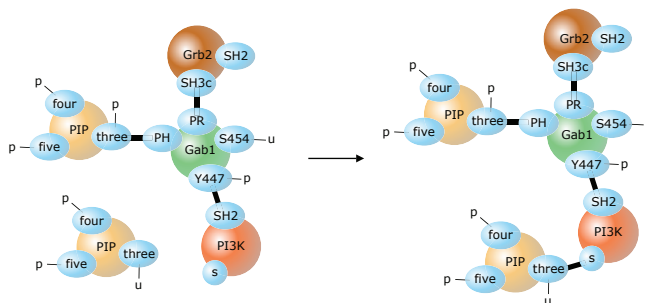


Fig. 1. Rule-based languages: the analogy between chemistry and Kappa. A: An aldol condensation is used as an example to illustrate the concept of a reaction rule in chemistry. The rule details only those molecular parts that are relevant to a particular scheme of reaction, designating unspecified context in terms of remainder groups R , R' , R'' , and R''' (highlighted). Sometimes a generic scheme, such as in (A), requires refinement into special classes defined by different sets of contexts R to 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. 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.

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	$\sigma ::= \varepsilon \mid s, \sigma$
(iv)	site	$s ::= n_i^\lambda$
(v)	site name	$n ::= x \in \mathcal{S}$
(vi)	internal state	$\iota ::= \epsilon \quad (\text{any state})$
(vii)	binding state	$\lambda ::= \epsilon \quad (\text{free})$
		$\mid - \quad (\text{semi-link: "bound to something"})$
		$\mid ? \quad (\text{unspecified: free or bound})$
		$\mid i \in \mathbb{N} \quad (\text{bond label})$

Definition 1.2 (Expressions).

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

Definition 1.3 (Well-formedness).

(ix)	unique interface	the sites form a set and 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	$E, A(\sigma, s', s, \sigma), E' \equiv E, A(\sigma, s, s', \sigma), E' \quad (\text{site permutation})$
(xiii)	mixture	$E, a, a', E' \equiv E, a', a, E' \quad (\text{agent permutation})$
(xiv)	edge labels	$i, j \in \mathbb{N} \wedge i \text{ not in } E \Rightarrow E[i/j] \equiv E \quad (\text{relabeling})$

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) and (ix) define the interface of an agent as a finite set $\sigma = \{s_1, s_2, \dots, s_n\}$ 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 either assumes some definite value or is left unspecified (ϵ), as declared 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 1 of agent ErbB1 is bound to site r of agent EGF, we deploy the same superscript at both sites. For example, the expression $\text{ErbB1}(1^2), \text{EGF}(r^2)$ 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, $A(s_\epsilon) \equiv A(s)$.

The object $\text{ErbB1}(1^2), \text{EGF}(r^2)$ 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

$$\text{EGF}(r^1) \text{ , ErbB1}(1^1, \text{CR}^3, \text{Y1016}_p) \text{ , EGF}(r^2) \text{ , ErbB1}(1^2, \text{CR}^3, \text{Y1016}_u) \quad [\mathbf{1}]$$

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 $\text{EGF}(r)$, an agent $\text{ErbB1}(1, \text{CR}, \text{Y1016}_p)$, and the complex represented in [1]. As defined in (viii), an expression is a graph over agents whose connected components are complexes.

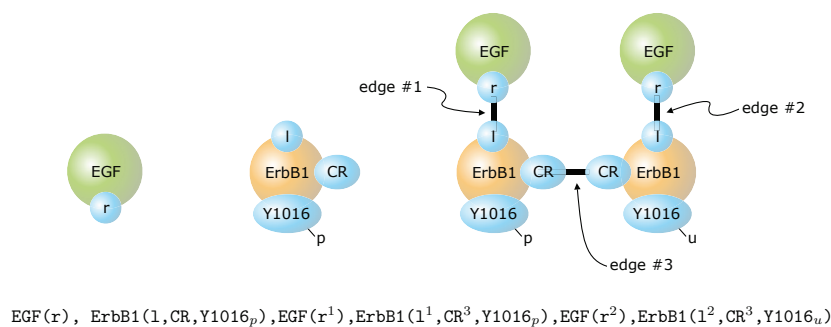


Fig. 2. A Kappa expression. The textual representation of a small reaction mixture containing 6 agents that are divided into three complexes (underlined) is shown at the bottom. The two complexes on the left are simple agents, while the complex on the right is made of 4 agents hanging together as shown. 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.

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.

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 complexes made of agents with a completely specified interface. In the main text we refer to these as molecular species or ground-level objects. However, it is useful 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 an action that occurs across many reaction instances consisting of different fully specified molecules.

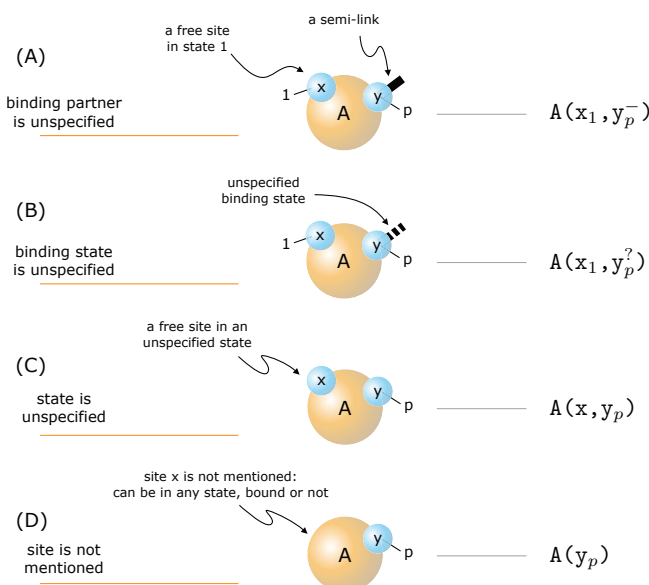


Fig. 3. Basic Kappa patterns. A pattern consists of a partially specified agent or set of agents. Information can be omitted as follows. A: the binding partner of a site is left unspecified. B: the binding state of a site is left unspecified. C: the internal state of a site is unspecified. D: both internal 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 consisting of partially specified agents. Figure 3 depicts the basic patterns exemplified by an agent $A(x, y)$ with two sites:

1. *Unspecified binding partner.* The expression $A(x_1, y_p^-)$, 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 $A(x_1, y_p^?)$, 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 $A(x, y_p)$, 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 $A(y_p)$, 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

$$\begin{array}{c} \text{Gab1}(\text{PH}^-, \text{Y447}^1), \text{PI3K}(\text{SH2}^1, s), \text{PIP}(\text{three}_u, \text{four}_p, \text{five}_p) \\ \downarrow \\ \text{Gab1}(\text{PH}^-, \text{Y447}^1), \text{PI3K}(\text{SH2}^1, s^2), \text{PIP}(\text{three}_u^2, \text{four}_p, \text{five}_p) \end{array} \quad [2]$$

(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.

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. An example may help.

$$A(x^1), B(x^1, y_u) \rightarrow A(x^1), B(x^1, y_p) \quad \{ \text{change state of B} \} \quad [3]$$

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 $A(x^1), B(x^1, y_p)$ with $B(x^1, y_p), A(x^1)$. 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:

$$A(x^1), B(x^1, y_u) \rightarrow B(x^1, y_p), A(x^1) \quad \left\{ \begin{array}{l} \text{delete the A referenced on the left} \\ \text{delete the B referenced on the left} \\ \text{add a B}(x, y_p) \\ \text{add an A}(x) \\ \text{bind B}(x, y_p) \text{ at } x \text{ to A}(x) \text{ at } x \end{array} \right. \quad [4]$$

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 $\iota \in \mathbb{V}$.

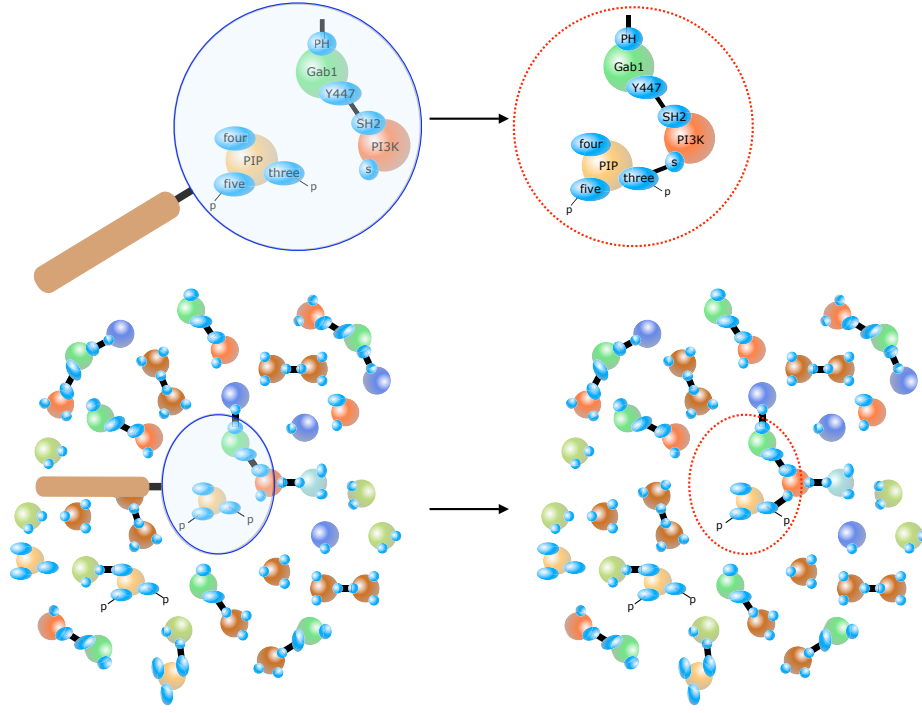


Fig. 4. Rule application in Kappa. First, a match between the pattern on the left hand side of a rule (blue lens) and the mixture (bottom) is identified. The action specified by the rule is then applied to the matching configuration, resulting in a new configuration according to the rule's right hand 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 [3] of the Doob-Gillespie algorithm [5, 6] for stochastic chemical kinetics.

The concept of a match can be extended to expressions (mixtures) E' and E , by saying that E' conforms to E , written as $E' \models E$, 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 \models . 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:

$$\begin{array}{ll} \text{binding state:} & ? \begin{cases} \nearrow \epsilon \\ \searrow - \end{cases} \begin{array}{l} \longrightarrow \lambda \in \mathbb{N}, \\ \longrightarrow \lambda \in \mathbb{N}, \end{array} \\ \text{internal state:} & \epsilon \longrightarrow \iota \in \mathbb{V} \end{array} \quad [5]$$

where the arrow means "is a superset of" (or, equivalently "is less specific than"): $x \rightarrow y \equiv x \supseteq y$. Equality applies between two λ s 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 , $E' \models E$, apply the following criteria:

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

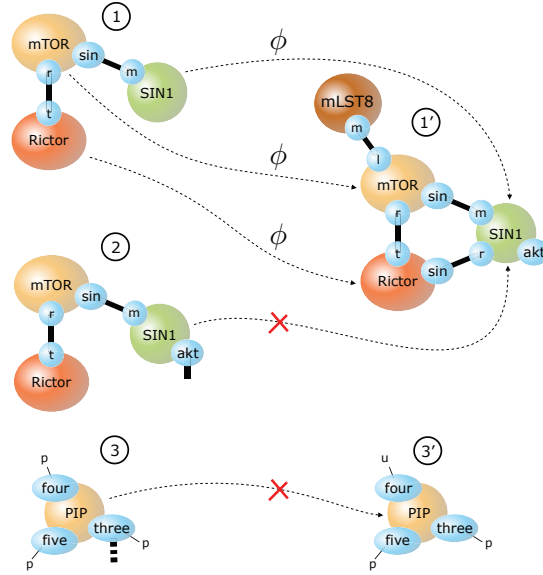


Fig. 5. Pattern matching. The embedding (fitting) of less detailed graphs on the left into more detailed graphs on the right. Graph 1 on the left embeds into complex 1' (a signaling assembly known as mTORC2, consisting of mTOR, Rictor, SIN1, and mLST8). Graph 1 agrees in all the names and states it mentions with graph 1'. Since graph 1 omits the site *l* from agent mTOR, 1'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', $\phi, G_1 \triangleleft_\phi G'_1$. 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 , $E'_1 \models E_1$. 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").

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 site graph G' , in which nodes are agents identified by their sequential position in the expression (see Figure 2) and edges correspond to bonds between sites as indicated in the expression. The concept of matching an expression E' to a less specific expression E , $E' \models E$, has a natural extension in the notion of embedding the graph G (corresponding to E) into the graph G' (corresponding to E'). An embedding of G into G' is equivalent to first finding an expression E'' that is structurally the same as E' , $E'' \equiv E'$, which amounts to a graph isomorphism between G' and G'' (corresponding to E''), and then a matching of E' to E , which amounts to a graph inclusion from G into G' . In symbols: $E' \equiv E'' \models E$ corresponds to $G'' \equiv G' \xleftarrow{\text{incl}} G$. Keep in mind that we think of a smaller (less detailed) graph G as being embedded into a larger (more detailed) graph G' , while a more detailed expression E' matches a less detailed expression E (the "pattern"). When dealing with graph embeddings it is more natural to simply think of the smaller graph G being relabelled (isomorphism) to "fit" the larger one. Intuitively, an embedding of G into G' (the first not larger than the second) is a process whereby we move G over G' , trying to overlay G on G' such that agent types match (as well as the states of the sites mentioned in G). Several overlays may be possible, because either graph may contain multiple agents of the same type connected in the same way. Each overlay generates a distinct embedding. For example, let G be the obvious graph of $A_1(x^1), A_2(y^1)$ and G' the graph of $A_1(x^1), A_2(y^1, x^2), A_3(y^2)$. (Since isomorphisms play a role, we attach an identifier to each agent.) There are two embeddings of G into G' : (1) the inclusion of G into G' and (2) a nontrivial isomorphism of G , i.e. the graph of $A_2(x^1), A_3(y^1)$, that also embeds into G' . Figure 5 provides a few graphical examples to fix the concepts. It is much more convenient to reason formally in terms of graphs than expressions, which is what we do in the main text and in subsequent sections. We notate the embedding of G into G' as $G \triangleleft_\phi G'$, where the subscript ϕ indicates a particular embedding.

1.5 Replacing a pattern. The execution of a rule $E_{\text{left}} \rightarrow E_{\text{right}}$ consists in testing whether an expression S conforms to E_{left} , $S \models E_{\text{left}}$, 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 $lhs \rightarrow rhs$, as described in section 1.3. Let L (R) be the remainder of *lhs* (*rhs*) after the *prefix*, thus, $lhs = \text{prefix}, L$ and $rhs = \text{prefix}, R$. 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:

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

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. Furthermore, to apply a rule with an empty lhs (production of agents) or with an empty

rhs (deletion of agents), we need to extend the structural equivalences, Definition 1.4, and the matching relation, Definition 1.5, with a dummy "empty agent", \emptyset , that matches an empty lhs expression and that is an empty rhs that can be used to overwrite the deleted agent(s). Thus, Definition 1.4 is extended with $E \equiv E, \emptyset$, and Definition 1.5 with $\emptyset \models \emptyset$.

Definition 1.6 (Replacement).

(i)	overwrite binding state	$\lambda_l[\lambda_r] = \begin{cases} \lambda_l & \text{if } \lambda_r \rightarrow \lambda_l \text{ in } [\mathbf{5}], \\ \lambda_r & \text{otherwise.} \end{cases}$
(ii)	overwrite internal state	$\iota_l[\iota_r] = \begin{cases} \iota_l & \text{if } \iota_r \rightarrow \iota_l \text{ in } [\mathbf{5}], \\ \iota_r & \text{otherwise.} \end{cases}$
(iii)	overwrite site	$n_{\iota_l}^{\lambda_l}[n_{\iota_r}^{\lambda_r}] = n_{\iota_l[\iota_r]}^{\lambda_l[\lambda_r]}$
(iv)	interface unchanged	$\sigma[\emptyset] = \sigma$
(v)	overwrite interface	$(s, \sigma)[s_r, \sigma_r] = s[s_r], \sigma[\sigma_r]$
(vi)	overwrite agent	$N(\sigma)[N(\sigma_r)] = N(\sigma[\sigma_r])$
(vii)	agent deletion	$N(\sigma)[\emptyset] = \emptyset$
(viii)	agent introduction	$\emptyset[N(\sigma_r)] = N(\sigma_r)$
(ix)	expression unchanged	$E[\varepsilon] = E$
(x)	overwrite expression	$(a, E)[a_r, E_r] = a[a_r], E[E_r]$

2 Glue-on-overlap procedure for fragments

Figure 6 illustrates an aspect of fragment construction. Consider whether pattern B can be made into a fragment with regard to rule r from Figure 2 in the main text, $r: A(a_u, b^1), B(c^1) \rightarrow A(a_p, b^1), B(c^1)$. Recall from the discussion of Figure 2 (main text) that we need $B^\diamond \subset r_{\text{lhs}}^\diamond$. While this is not the case for the B in Figure 6, B might be further specialized (refined) into B' by adding more context, after which it might qualify as a fragment for r . Refining B into B' , such that $B'^\diamond \subset r_{\text{lhs}}^\diamond$, amounts to glueing together B and r_{lhs} . To do this, we first identify a glueing region. The glueing region of B and r_{lhs} is the set of agents and sites that both mention, along with a mutually compatible state (following the specificity ranking given in [5]). If there is no such state, the glueing region is empty. In the middle of the diamond in Figure 6 we see that B and r_{lhs} both mention agent A with site a in state u . Thus, $A(a_u)$ is the glueing region (shown at the top of the diamond). If B had left the internal state of site a unspecified, i.e. $B := C(a^1), A(a, c^1, d)$, then the glueing region with r_{lhs} would be $A(a)$, since an unspecified internal state at site a is compatible with any specific state at that site, in particular u . In contrast, the pattern $B := C(a^1), A(a_p, c^1, d)$ would result in an empty glueing region.

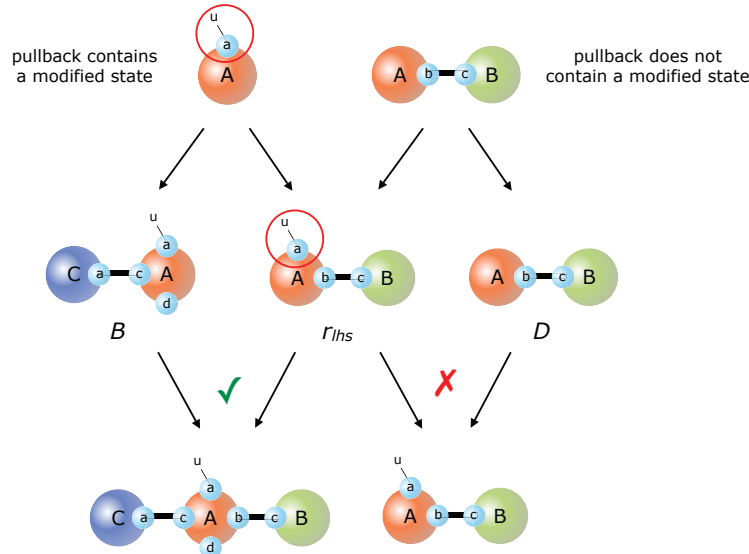


Fig. 6. Fragment construction by glue-on-overlap. Pattern B and the lhs component, r_{lhs} , of rule r (Figure 2 of main text) have a glueing region, known in category theory as a pullback. The glueing region acts as an instruction for joining both patterns into a new one (the so-called pushout), shown at the bottom of the diamond. Patterns D and r_{lhs} are not joined (red cross), because their pullback does not contain a site modified by r . See text for details.

We now can join B and r_{lhs} by overlaying them on the glueing region, $A(a_u)$. This yields a new pattern that qualifies as a fragment as far as rule r is concerned. This new pattern is subject to a similar procedure with the next rule in line to be checked.

Such a refinement is not needed, however, unless a condition is met. The red circle in Figure 6 indicates the site and state modified by r , which the glueing region must contain if a pattern is to be refined into a fragment candidate. For example, pattern D in Figure 6 does have a glueing region with r_{lhs} , which does not contain the site and state modified by r . Rule r still acts on instances of D , since any ground-level instance that matches r_{lhs} also matches D . However, since D does not care about the action of r , any instance of D transformed by r results in another instance of D , with no net effect on the concentration of D . This only means that there is no point in further refining such a pattern with regard to r . Nonetheless, a rule s might give rise to a fragment whose overlap with rule r does not contain sites and states modified by r . Naturally, that fragment's ODE will not receive any kinetic terms from r .

3 Illustration of syntactical criteria

To illustrate syntactical criteria, such as Cov1, we consider flows of control within and across agents. Consider Figure 4B in the main text and imagine two rules r_1 and r_2 both of which modify the state at site b of agent A, but r_1 conditions the modification of b on the state of site a, while r_2 conditions it on the state of c. To fix ideas, let $r_1: A(a_0, b_0) \rightarrow A(a_0, b_1) @ k_1$ and $r_2: A(c_1, b_1) \rightarrow A(c_1, b_0) @ k_2$, as shown in Figure 7. Let a ground-level species be denoted by a triplet abc reporting the state (0 or 1) at each site and let $*$ be a wildcard for expressing patterns. The lhs of r_1 , $A(a_0, b_0)$, which is $00*$, is not a unit of the dynamics, as can be easily seen: $d[00*]/dt = d[000]/dt + d[001]/dt = -k_1[000] - k_1[001] + k_2[011] = -k_1[00*] + k_2[011]$. The reason is a configuration in $00*$ - specifically 001 - that satisfies the lhs of r_1 and the rhs of r_2 . As a consequence, the rate equation of $[00*]$ receives a term from the action of r_2 (on 011). Thus, the concentration of 011 matters for describing the dynamics of the system, which means that the state at both controlling sites a and c must be known at any time. This forces both sites into the same covering class as b.

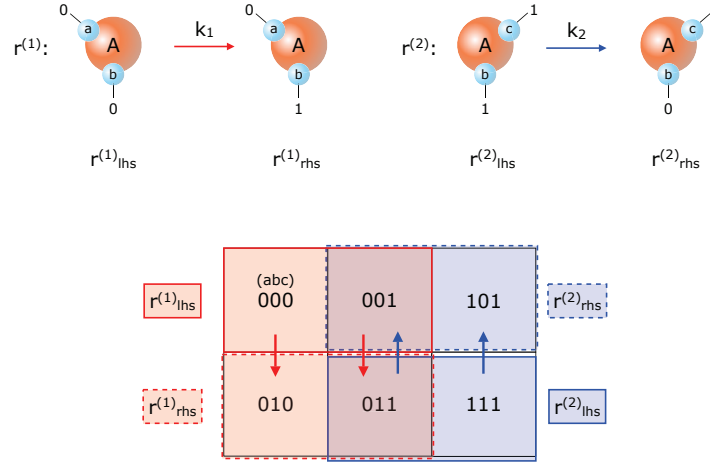


Fig. 7. Example illustrating directive Cov1 for the case shown in Figure 4B of the main text. Two rules, as described in the text, are shown at the top. The bottom illustrates the action of the rules on microconfigurations (ground-level objects). The solid (dotted) rectangles cover the configurations that match the pattern on the lhs (rhs) of each rule. See text for details.

The case of Figure 4D in the main text consists in two rules r_1 and r_2 that modify sites b and c, respectively, while both testing a condition on site a. For example, let $r_1: A(a_0, b_0) \rightarrow A(a_0, b_1) @ k_1$ and $r_2: A(a_0, c_0) \rightarrow A(a_0, c_1) @ k_2$, as shown in Figure 8. Here, too, there is a ground-level configuration, 000, whose concentration is affected by both rules, because it matches the lhs of r_1 and r_2 . However, r_2 transforms 000 into 001, which still matches the lhs of r_1 . In fact, this is an example of a situation in which the glueing region (section 2) between the lhs of r_1 and the lhs of r_2 does *not* contain either modified site. Hence, r_2 maps one subset of the extension of $00*$ into another without affecting the concentration of $00*$. Indeed, $00*$ is a fragment, as we can easily check: $d[00*]/dt = d[000]/dt + d[001]/dt = -k_1[000] - k_2[000] - k_1[001] + k_2[000] = -k_1[00*]$. Agent A has therefore two covering classes, $\{a, b\}$ and $\{a, c\}$, for a total of four fragments: $00*$, $01*$, $0*0$, and $0*1$. In general, a covering class contains the backward closure of all sites that control a particular locus of modification.

4 Independence and self-consistency

We discuss two simple examples to clarify the notions of independence [7, 8, 9] and self-consistency.

4.1 Example 1: Independence and tileable systems. Consider a scaffold protein $C(a, b)$ with a specific binding site for $A(a)$ and $B(b)$, as shown in Figure 9A. This system comprises six possible molecular species, $A, B, C, A.C, C.B, A.C.B$, where the dot indicates a bond. In this section, we use mathematical font (slanted) when referring to molecular species or patterns and their concentration variables as they appear in

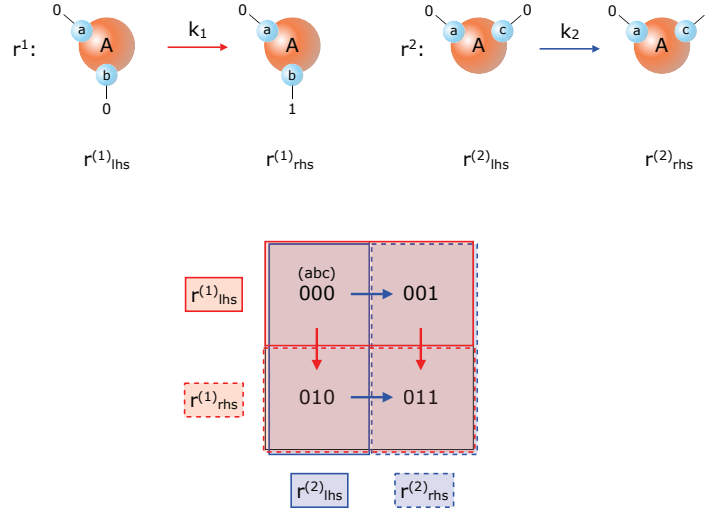
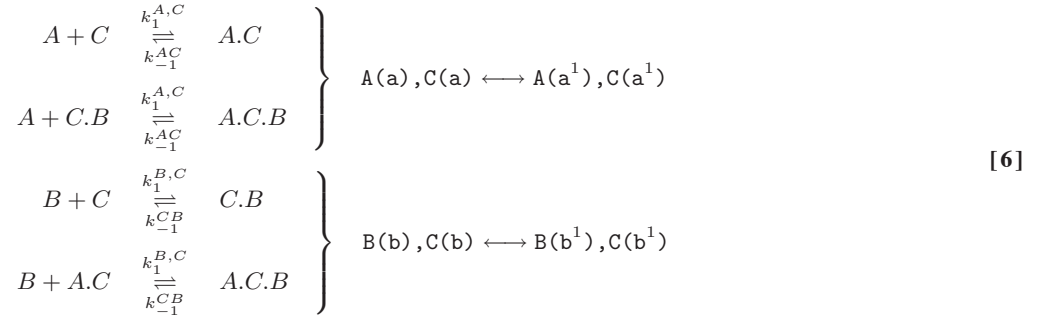


Fig. 8. Example illustrating directive Cov1 for the case shown in Figure 4D of the main text. Two rules, as described in the text, are shown at the top. The bottom illustrates the action of the rules on microconfigurations (ground-level objects). The solid (dotted) rectangles cover the configurations that match the pattern on the lhs (rhs) of each rule. See text for details.

deterministic kinetic equations. The six species are related by four association and dissociation reactions, as shown on the left:



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 $C.B$, and likewise for the interactions between C and B . This kinetic indistinguishability means that only four Kappa-rules, shown on the right of [6], are needed to express the eight reactions 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 , respectively, from the corresponding interaction rules. 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 identical rate constants, $k_1^{A,C}$.

The full dynamical system is given by:

$$\begin{aligned}
 \frac{d[A]}{dt} &= k_{-1}^{AC} ([A.C] + [A.C.B]) - k_1^{A,C} [A] ([C] + [C.B]) \\
 \frac{d[B]}{dt} &= k_{-1}^{CB} ([C.B] + [A.C.B]) - k_1^{B,C} [B] ([C] + [A.C]) \\
 \frac{d[C]}{dt} &= k_{-1}^{AC} [A.C] + k_{-1}^{CB} [C.B] - [C] ([A] k_1^{A,C} + [B] k_1^{B,C}) \\
 \frac{d[A.C]}{dt} &= k_1^{A,C} [A][C] + k_{-1}^{CB} [A.C.B] - [A.C] (k_{-1}^{AC} + [B] k_1^{B,C}) \\
 \frac{d[C.B]}{dt} &= k_{-1}^{AC} [A.C.B] + k_1^{B,C} [B][C] - [C.B] (k_{-1}^{CB} + [A] k_1^{A,C}) \\
 \frac{d[A.C.B]}{dt} &= k_1^{A,C} [A][C.B] + k_1^{B,C} [B][A.C] - [A.C.B] (k_{-1}^{AC} + k_{-1}^{CB})
 \end{aligned}
 \quad [7]$$

This system can be coarse-grained by conceptually splitting the centerpiece C into two fragments (Figure 9A), one containing only the A -binding site, the other only the B -binding site. This captures the fact that A and B cannot know about each other despite their interactions with a shared C . Let us denote the former fragment with C^* and the latter with $*C$, the asterisk indicating that we don't care about the corresponding

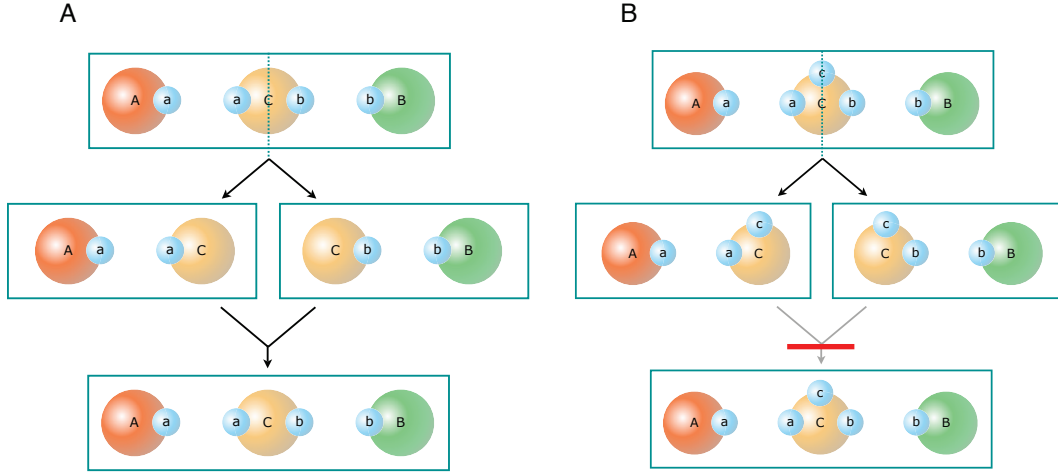


Fig. 9. Independence and self-consistency. The Figure depicts schematically Examples 1 (A) and 2 (B) described in the text. In both cases, the binding interactions on the "left" and the "right" of agent C do not influence one another. However, in case (B), agent C internally synchronizes the sites a 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.

binding site. The system then splits into two independent subsystems, self-consistently described by $\{A, C^*, A.C^*\}$ and $\{B, *C, *C.B\}$.

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

Self-consistent means that each set of variables is closed with regard to its own dynamics:

$$\begin{aligned}
 \frac{d[A]}{dt} &= k_{-1}^{AC} [A.C^*] - k_1^{A,C} [A][C^*] \\
 \frac{d[C^*]}{dt} &= k_{-1}^{AC} [A.C^*] - k_1^{A,C} [A][C^*] \\
 \frac{d[A.C^*]}{dt} &= k_1^{A,C} [A][C^*] - k_{-1}^{AC} [A.C^*]
 \end{aligned} \tag{8}$$

and

$$\begin{aligned}
 \frac{d[B]}{dt} &= k_{-1}^{BC} [*C.B] - k_1^{B,C} [B][*C] \\
 \frac{d[*C]}{dt} &= k_{-1}^{BC} [*C.B] - k_1^{B,C} [B][*C] \\
 \frac{d[*C.B]}{dt} &= k_1^{B,C} [B][*C] - k_{-1}^{BC} [*C.B]
 \end{aligned} \tag{9}$$

C does not propagate any information between A and B , and it does not correlate them either. Independent then 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 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:

$$P(A.C.B) = \frac{P(A.C^*)P(*C.B)}{P(*C^*)}. \tag{10}$$

These relationships are reflected by the corresponding (time-dependent) concentrations $[A.C^*]$, $[*C.B]$, and $[*C^*]$. Thus, equation (10) asserts

$$[A.C.B] = \frac{([A.C] + [A.C.B])([C.B] + [A.C.B])}{[C] + [A.C] + [C.B] + [A.C.B]}, \tag{11}$$

from which it follows that

$$X \doteq [A.C.B][C] - [A.C][C.B] = 0. \quad [12]$$

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

$$\frac{dX}{dt} = -X(k_1^{A,C}[A] + k_{-1}^{AC} + k_1^{C,B}[B] + k_{-1}^{CB}), \quad [13]$$

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 [10] is a manifestation of independence and can be generalized to *define* a class of systems whose fragments behave like tiles in the following sense. Two fragments \mathcal{F}_1 and \mathcal{F}_2 can be "snapped" together (possibly in more than one way), $\mathcal{F}_1 \cup \mathcal{F}_2$, if they have a non-conflicting valuation on the sites of agents both mention, $\mathcal{F}_1 \cap \mathcal{F}_2$, precisely as the glue-on-overlap in section 2. 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 $\mathfrak{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_n\}$ obeys the independence equation [10], we can extend a fragment \mathcal{F}_i into a fragment $\mathcal{F}_i \cup \mathcal{F}_j$ 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 in this way, we can invert the coarse-graining. That is, we can reconstruct any molecular species that can possibly occur in the system, while in the process computing its concentration via the tiling equation [10]. The microscopic dynamics expressed in terms of \mathfrak{F} will be exact, provided the initial condition satisfied the independence relations of the form [12], otherwise it will approach the exact dynamics exponentially according to [13]. Typical situations, however, are more subtle, as illustrated in the next example.

4.2 Example 2: Stealth correlation. Example 2, Figure 9B, 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. (After appropriate name changes, this corresponds to the control flow depicted in Figure 4D of the main text.) 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 binding process at site a is independent of the binding state of site b , and vice versa. The system is then described by the 5 reactions shown below on the left and whose dynamics is detailed on the right:

$$\begin{array}{ll} (uuu) \xrightarrow{k^c} (uou) & \frac{d[uuu]}{dt} = -k^c[uuu] \\ (uou) \xrightarrow{k^l} (oou) & \frac{d[uou]}{dt} = -(k^l + k^r)[uou] + k^c[uuu] \\ (uoo) \xrightarrow{k^l} (ooo) & \text{with } \frac{d[uoo]}{dt} = -k^l[uoo] + k^r[uou] \\ (uou) \xrightarrow{k^r} (uoo) & \frac{d[uou]}{dt} = k^l[uou] - k^r[oou] \\ (oou) \xrightarrow{k^r} (ooo) & \frac{d[oou]}{dt} = k^l[uoo] + k^r[oou] \end{array} \quad [14]$$

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

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

Again, as in Example 1, the system 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, $[uu*] = [*uu] = [uuu]$.)

$$\begin{array}{ll} \frac{d[uuu]}{dt} = -k^c[uuu] & \frac{d[uuu]}{dt} = -k^c[uuu] \\ \frac{d[uo*]}{dt} = -k^l[uo*] + k^c[uuu] & \text{and } \frac{d[*ou]}{dt} = -k^r[*ou] + k^c[uuu] \\ \frac{d[oo*]}{dt} = k^l[uo*] & \frac{d[*oo]}{dt} = k^r[*ou] \end{array} \quad [15]$$

However, unlike in Example 1, we cannot recombine the subsystems to reconstruct the microscopic dynamics via equation [10]. In analogy to equation [12], we define $X \doteq [ooo][uou] - [oou][uoo]$ as a measure of independence and obtain

$$\frac{dX}{dt} = -X(k_a + k_b) + k_c[ooo][uuu], \quad [16]$$

indicating that the two subsystems remain correlated, because the state of the controller site c correlates the states of a and b . While the coarse grained dynamics is still exact, it can no longer be inverted by tiling. As can be seen from Figure 10, the concentration of the fully occupied

form $[ooo]$ is underestimated by assuming independence (tiling equation [10]), because the injection of new instances of (uou) , by virtue of the first reaction in [14], keeps introducing correlations between sites a and b, thus maintaining $[ooo]$ above what an observer would expect by measuring $[oo*]$, $[*oo]$, and $[*o*]$ and assuming independence. This deviation from independence ceases once the system has exhausted its reservoir of (uuu) (having converted it all to (ooo) , see Figure 10), since, after all, binding at site a is independent from binding at b. The coarse-graining, which enabled us to describe the dynamics of the system in terms of two self-consistent subsystems, throws away correlation information, preventing reconstitution of the original microscopic description. From a viewpoint internal to the system, this is no loss, as the correlation cannot be observed from within the system (unless additional specific interactions are posited), and a microscopic description is therefore irrelevant. As outside observers, however, 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.

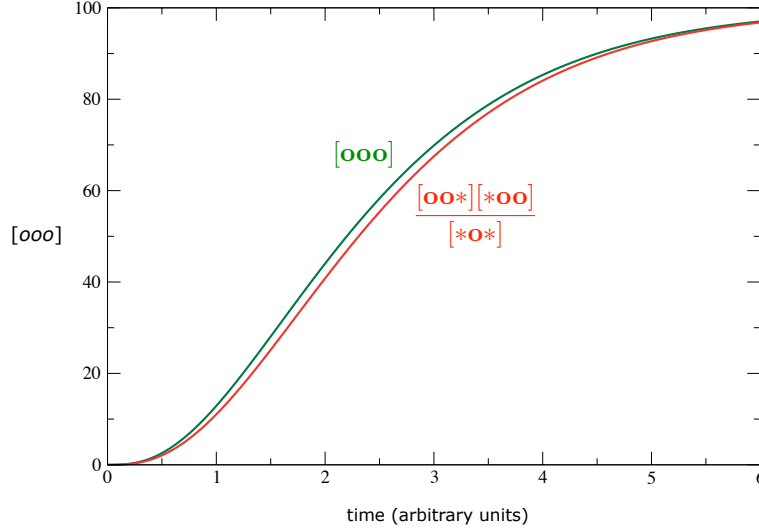


Fig. 10. Error from independence assumption. The Figure depicts the concentration dynamics of the fully occupied form of C, $[ooo]$, according to the equations [14] (green). The red curve shows the dynamics of $[ooo]$ when it is computed using the independence equation [10], $[ooo] = [oo*][*oo]/[*o*]$. 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.

5 The early-EGF model

We use a canonical cell signaling pathway in mammalian cells, the epidermal growth factor receptor (EGFR 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 involved in the production and the processing of intracellular signals induced by EGF receptor activity. The EGF system has a controlling influence on cell division, cell fate and cell morphology.

The rule system below describes only a tiny set of early events (that are by no means agreed-upon) in EGFR signaling. In the present context, our objective is not to provide an accurate model of the EGFR pathway, but rather to use this relatively small example to illustrate 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 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). Using different rules for the same action is a means for differentiating between (potentially) different kinetics (or simply rate constants) depending on context. 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 [10]). 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.

5.1 Rules.

$$\begin{aligned}
 \text{r01: } E(r^1), R(1^1, r) &\longrightarrow E(r), R(1, r) \\
 \text{r02: } E(r), R(1, r) &\longrightarrow E(r^1), R(1^1, r) \\
 \text{r03: } E(r^2), E(r^1), R(1^2, r), R(1^1, r) &\longrightarrow E(r^3), E(r^2), R(1^3, r^1), R(1^2, r^1)
 \end{aligned}$$

r04: $E(r^3), E(r^2), R(1^3, r^1), R(1^2, r^1) \longrightarrow E(r^2), E(r^1), R(1^2, r), R(1^1, r)$
 r05: $E(r^3), E(r^2), R(1^3, r^1), R(Y68_u, 1^2, r^1) \longrightarrow E(r^3), E(r^2), R(1^3, r^1), R(Y68_p, 1^2, r^1)$
 r06: $R(Y68_p) \longrightarrow R(Y68_u)$
 r07: $E(r^3), E(r^2), R(1^3, r^1), R(Y48_u, 1^2, r^1) \longrightarrow E(r^3), E(r^2), R(1^3, r^1), R(Y48_p, 1^2, r^1)$
 r08: $R(Y48_p) \longrightarrow R(Y48_u)$
 r09: $R(Y48_p^1, r^-), S(Y7_u, c^1) \longrightarrow R(Y48_p^1, r^-), S(Y7_p, c^1)$
 r10: $S(Y7_p, c^-) \longrightarrow S(Y7_u, c^-)$
 r11: $S(Y7_p, c) \longrightarrow S(Y7_u, c)$
 r12: $G(a, b), R(Y68_p) \longrightarrow G(a^1, b), R(Y68_p^1)$
 r13: $G(a^1, b), R(Y68_p^1) \longrightarrow G(a, b), R(Y68_p)$
 r14: $G(a, b^-), R(Y68_p) \longrightarrow G(a^1, b^-), R(Y68_p^1)$
 r15: $G(a^1, b^-), R(Y68_p^1) \longrightarrow G(a, b^-), R(Y68_p)$
 r16: $G(a^1, b), R(Y68_p^1), O(d) \longrightarrow G(a^2, b^1), R(Y68_p^2), O(d^1)$
 r17: $G(a^2, b^1), R(Y68_p^2), O(d^1) \longrightarrow G(a^1, b), R(Y68_p^1), O(d)$
 r18: $G(a, b), O(d) \longrightarrow G(a, b^1), O(d^1)$
 r19: $G(a, b^1), O(d^1) \longrightarrow G(a, b), O(d)$
 r20: $G(a^1, b), S(Y7_p^1, c), O(d) \longrightarrow G(a^2, b^1), S(Y7_p^2, c), O(d^1)$
 r21: $G(a^2, b^1), S(Y7_p^2, c), O(d^1) \longrightarrow G(a^1, b), S(Y7_p^1, c), O(d)$
 r22: $G(a^1, b), S(Y7_p^1, c^-), O(d) \longrightarrow G(a^2, b^1), S(Y7_p^2, c^-), O(d^1)$
 r23: $G(a^2, b^1), S(Y7_p^2, c^-), O(d^1) \longrightarrow G(a^1, b), S(Y7_p^1, c^-), O(d)$
 r24: $R(Y48_p), S(Y7_u, c) \longrightarrow R(Y48_p^1), S(Y7_u, c^1)$
 r25: $R(Y48_p^1), S(Y7_u, c^1) \longrightarrow R(Y48_p), S(Y7_u, c)$
 r26: $R(Y48_p), S(Y7_p, c) \longrightarrow R(Y48_p^1), S(Y7_p, c^1)$
 r27: $R(Y48_p^1), S(Y7_p, c^1) \longrightarrow R(Y48_p), S(Y7_p, c)$
 r28: $G(a^1, b), R(Y48_p), S(Y7_p^1, c) \longrightarrow G(a^2, b), R(Y48_p^1), S(Y7_p^2, c^1)$
 r29: $G(a^2, b), R(Y48_p^1), S(Y7_p^2, c^1) \longrightarrow G(a^1, b), R(Y48_p), S(Y7_p^1, c)$
 r30: $G(a^2, b^1), R(Y48_p), S(Y7_p^2, c), O(d^1) \longrightarrow G(a^3, b^2), R(Y48_p^1), S(Y7_p^3, c^1), O(d^2)$
 r31: $G(a^3, b^2), R(Y48_p^1), S(Y7_p^3, c^1), O(d^2) \longrightarrow G(a^2, b^1), R(Y48_p), S(Y7_p^2, c), O(d^1)$
 r32: $G(a, b), R(Y48_p^1), S(Y7_p, c^1) \longrightarrow G(a^2, b), R(Y48_p^1), S(Y7_p^2, c^1)$
 r33: $G(a^2, b), R(Y48_p^1), S(Y7_p^2, c^1) \longrightarrow G(a, b), R(Y48_p^1), S(Y7_p, c^1)$
 r34: $G(a, b), S(Y7_p, c) \longrightarrow G(a^1, b), S(Y7_p^1, c)$
 r35: $G(a^1, b), S(Y7_p^1, c) \longrightarrow G(a, b), S(Y7_p, c)$
 r36: $G(a, b^-), S(Y7_p, c) \longrightarrow G(a^1, b^-), S(Y7_p^1, c)$
 r37: $G(a^1, b^-), S(Y7_p^1, c) \longrightarrow G(a, b^-), S(Y7_p, c)$
 r38: $G(a, b^2), R(Y48_p^1), S(Y7_p, c^1), O(d^2) \longrightarrow G(a^3, b^2), R(Y48_p^1), S(Y7_p^3, c^1), O(d^2)$
 r39: $G(a^3, b^2), R(Y48_p^1), S(Y7_p^3, c^1), O(d^2) \longrightarrow G(a, b^2), R(Y48_p^1), S(Y7_p, c^1), O(d^2)$

5.2 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. $R(r^-)$ 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: $R(1^-, r) \longrightarrow R(1, r)$
 cr02: $E(r), R(1) \longrightarrow E(r^1), R(1^1)$
 cr03: $R(1^-, r), R(1^-, r) \longrightarrow R(1^-, r^1), R(1^-, r^1)$
 cr04: $R(r^-) \longrightarrow R(r) @ 0.5$
 cr05: $R(Y68_u^2, r^-) \longrightarrow R(Y68_p^2, r^-)$
 cr06: $R(Y68_p) \longrightarrow R(Y68_u)$
 cr07: $R(Y48_u^2, r^-) \longrightarrow R(Y48_p^2, r^-)$
 cr08: $R(Y48_p) \longrightarrow R(Y48_u)$
 cr09: $R(Y48_p^1, r^-), S(Y7_u^2, c^1) \longrightarrow R(Y48_p^1, r^-), S(Y7_p^2, c^1)$
 cr10: $S(Y7_p, c^-) \longrightarrow S(Y7_u, c^-)$
 cr11: $S(Y7_p, c) \longrightarrow S(Y7_u, c)$

cr12: $G(a, b), R(Y68_p) \longrightarrow G(a^1, b), R(Y68_p^1)$
 cr13: $G(a^1, b), R(Y68^1) \longrightarrow G(a, b), R(Y68)$
 cr14: $G(a, b^-), R(Y68_p) \longrightarrow G(a^1, b^-), R(Y68_p^1)$
 cr15: $G(a^1, b^-), R(Y68^1) \longrightarrow G(a, b^-), R(Y68)$
 cr16: $G(a^1, b), R(Y68^1), 0(d) \longrightarrow G(a^2, b^1), R(Y68^2), 0(d^1)$
 cr17: $G(a^1, b^-), R(Y68^1) \longrightarrow G(a^1, b), R(Y68^1)$
 cr18: $G(a, b), 0(d) \longrightarrow G(a, b^1), 0(d^1)$
 cr19: $G(a, b^-) \longrightarrow G(a, b)$
 cr20: $G(a^1, b), S(Y7^1, c), 0(d) \longrightarrow G(a^2, b^1), S(Y7^2, c), 0(d^1)$
 cr21: $G(a^1, b^-), S(Y7^1, c) \longrightarrow G(a^1, b), S(Y7^1, c)$
 cr22: $G(a^1, b), S(Y7^1, c^-), 0(d) \longrightarrow G(a^2, b^1), S(Y7^2, c^-), 0(d^1)$
 cr23: $G(a^1, b^-), S(Y7^1, c^-) \longrightarrow G(a^1, b), S(Y7^1, c^-)$
 cr24: $R(Y48_p), S(Y7_u^?, c) \longrightarrow R(Y48_p^1), S(Y7_u^?, c^1)$
 cr25: $S(Y7_u^?, c^-) \longrightarrow S(Y7_u^?, c)$
 cr26: $R(Y48_p), S(Y7_p, c) \longrightarrow R(Y48_p^1), S(Y7_p, c^1)$
 cr27: $S(Y7_p, c^-) \longrightarrow S(Y7_p, c)$
 cr28: $G(a^1, b), R(Y48_p), S(Y7^1, c) \longrightarrow G(a^2, b), R(Y48_p^1), S(Y7^2, c^1)$
 cr29: $G(a^1, b), S(Y7^1, c^-) \longrightarrow G(a^1, b), S(Y7^1, c)$
 cr30: $G(a^1, b^-), R(Y48_p), S(Y7^1, c) \longrightarrow G(a^2, b^-), R(Y48_p^1), S(Y7^2, c^1)$
 cr31: $G(a^1, b^-), S(Y7^1, c^-) \longrightarrow G(a^1, b^-), S(Y7^1, c)$
 cr32: $G(a, b), S(Y7_p, c^-) \longrightarrow G(a^1, b), S(Y7_p^1, c^-)$
 cr33: $G(a^1, b), S(Y7^1, c^-) \longrightarrow G(a, b), S(Y7, c^-)$
 cr34: $G(a, b), S(Y7_p, c) \longrightarrow G(a^1, b), S(Y7_p^1, c)$
 cr35: $G(a^1, b), S(Y7^1, c) \longrightarrow G(a, b), S(Y7, c)$
 cr36: $G(a, b^-), S(Y7_p, c) \longrightarrow G(a^1, b^-), S(Y7_p^1, c)$
 cr37: $G(a^1, b^-), S(Y7^1, c) \longrightarrow G(a, b^-), S(Y7, c)$
 cr38: $G(a, b^-), S(Y7_p, c^-) \longrightarrow G(a^1, b^-), S(Y7_p^1, c^-)$
 cr39: $G(a^1, b^-), S(Y7^1, c^-) \longrightarrow G(a, b^-), S(Y7, c^-)$

5.3 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 6. These 38 variables form a dynamical system (shown in section 7) whose state at any time t is identical to the state attained by the microscopic dynamics (involving 356 variables) and subsequent coarse-graining.

$\mathcal{F}_1: E(r^1), R(Y48_p^2, l^1, r^{\text{R@r}}), S(Y7_p^3, c^2), G(a^3, b^4), 0(d^4)$
 $\mathcal{F}_2: E(r^1), R(Y48_p^2, l^1, r), S(Y7_p^3, c^2), G(a^3, b^4), 0(d^4)$
 $\mathcal{F}_3: G(a^2, b^1), S(Y7_p^2, c^3), R(Y48_p^3, l, r), 0(d^1)$
 $\mathcal{F}_4: R(Y48_p^1, l, r), S(Y7_p, c^1)$
 $\mathcal{F}_5: G(a, b^1), 0(d^1)$
 $\mathcal{F}_6: E(r^1), R(Y48_p^2, l^1, r), S(Y7_p, c^2)$
 $\mathcal{F}_7: E(r^1), R(Y48_p^2, l^1, r^{\text{R@r}}), S(Y7_p, c^2)$
 $\mathcal{F}_8: G(a^2, b^1), S(Y7_p^2, c), 0(d^1)$
 $\mathcal{F}_9: S(Y7_p, c)$
 $\mathcal{F}_{10}: G(a^1, b), S(Y7_p^1, c)$
 $\mathcal{F}_{11}: G(a, b)$
 $\mathcal{F}_{12}: E(r^1), R(Y48_p^2, l^1, r^{\text{R@r}}), S(Y7_p^3, c^2), G(a^3, b)$
 $\mathcal{F}_{13}: E(r^1), R(Y48_p^2, l^1, r), S(Y7_p^3, c^2), G(a^3, b)$
 $\mathcal{F}_{14}: G(a^1, b), S(Y7_p^1, c^2), R(Y48_p^2, l, r)$
 $\mathcal{F}_{15}: R(Y48_p, l, r)$
 $\mathcal{F}_{16}: E(r^1), R(Y48_p, l^1, r)$
 $\mathcal{F}_{17}: E(r^1), R(Y48_p, l^1, r^{\text{R@r}})$
 $\mathcal{F}_{18}: E(r^1), R(Y48_p^2, l^1, r^{\text{R@r}}), S(Y7_u, c^2)$
 $\mathcal{F}_{19}: E(r^1), R(Y48_p^2, l^1, r), S(Y7_u, c^2)$
 $\mathcal{F}_{20}: R(Y48_p^1, l, r), S(Y7_u, c^1)$
 $\mathcal{F}_{21}: S(Y7_u, c)$
 $\mathcal{F}_{22}: 0(d)$
 $\mathcal{F}_{23}: E(r^1), R(Y68_p^2, l^1, r^{\text{R@r}}), G(a^2, b^3), 0(d^3)$
 $\mathcal{F}_{24}: E(r^1), R(Y68_p^2, l^1, r), G(a^2, b^3), 0(d^3)$

$\mathcal{F}_{25}: G(a^2, b^1), R(Y68_p^2, l, r), O(d^1)$
 $\mathcal{F}_{26}: G(a^1, b), R(Y68_p^1, l, r)$
 $\mathcal{F}_{27}: E(r^1), R(Y68_p^2, l^1, r), G(a^2, b)$
 $\mathcal{F}_{28}: E(r^1), R(Y68_p^2, l^1, r^{R@r}), G(a^2, b)$
 $\mathcal{F}_{29}: R(Y68_p, l, r)$
 $\mathcal{F}_{30}: E(r^1), R(Y68_p, l^1, r)$
 $\mathcal{F}_{31}: E(r^1), R(Y68_p, l^1, r^{R@r})$
 $\mathcal{F}_{32}: E(r^1), R(Y48_u, l^1, r^{R@r})$
 $\mathcal{F}_{33}: E(r^1), R(Y48_u, l^1, r)$
 $\mathcal{F}_{34}: R(Y48_u, l, r)$
 $\mathcal{F}_{35}: E(r^1), R(Y68_u, l^1, r^{R@r})$
 $\mathcal{F}_{36}: E(r^1), R(Y68_u, l^1, r)$
 $\mathcal{F}_{37}: R(Y68_u, l, r)$
 $\mathcal{F}_{38}: E(r)$

6 Translating rules into a dynamical system for fragments

6.1 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, $R(Y48_p^1), S(Y7^2, c^1), G(a^2, b^{O@d})$ is a partial complex (it is the right hand side of cr30). Thus:

Definition 6.1 (Partial complex).

A partially specified complex (or partial complex for short) is a connected expression, 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) a bond stub indicating the names of the bound agent and its binding site

We extend the concept of a match, $E' \models E$, Definition 1.5, or the concept of an embedding $G \triangleleft_\phi G'$ (see end of section 1.4), to expressions containing a stub by simply extending the specificity ranking of binding states (section 1.4) in the obvious way.

Fragments, as defined in the main text, 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, as explained in the main text.

6.2 Every lhs component of a rule is contained in some fragment. 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 growth process (see main text, subsection "Fragment assembly") then guarantees that fragments extend at least as much as any lhs component in a rule, except for pure dissociation rules. By construction, fragments do not contain bonds that are soft in the ACM, and thus do not extend components with such bonds. However, when such a component is cut at a soft bond, and the bond is replaced with two stubs, each piece can be embedded in a fragment. This reflects the fact that the system of rules cannot detect any correlations between such pieces (or the bond would be solid). The construction of fragments also ensures that when a lhs component Z contains a site that is modified, Z is contained in each fragment exhibiting that site.

6.3 Expressing the concentration of a subfragment in terms of fragment concentrations. Before proceeding we need some clarification on the notion of "concentration" in a formal rule-based approach that deals with patterns (partial complexes). What is the concentration of a partial complex in a mixture of fully specified complexes (species)?

Counting embeddings. In a stochastic setting, in which a reaction mixture x is a multi-set of molecular species, one should distinguish between two quantities for each partial complex Z : the number of embeddings ϕ (section 1.4) of Z into x , written $|\{\phi \mid Z \triangleleft_\phi x\}|$, and the number of embeddings corrected by the number of automorphisms of Z , $auto(Z) = |\{\phi' \mid Z \triangleleft_{\phi'} Z\}|$ (since Z may have symmetries):

$$[Z] := |\{\phi \mid Z \triangleleft_\phi x\}| / auto(Z)$$

In the case where the partial complex Z is actually a (fully specified) species, the corrected number of embeddings is the number of *occurrences* of Z in x , which yields the concentration, when divided by the volume, as in ODEs.

Given two partial complexes Z and Z' such that Z embeds into Z' , the set $\{\phi \mid Z \triangleleft_\phi Z'\}$ of embeddings can be quotiented by the equivalence relation \sim , relating any pair of embeddings ϕ and ϕ' such that ϕ' can be written as $\phi \circ \sigma$ for σ an automorphism of Z' . Again, we have two choices; either we count embeddings, or embeddings up to \sim . We write $Z \trianglelefteq_\phi Z'$ when ϕ is an embedding equivalence class and we have therefore:

$$|\{\phi \mid Z \triangleleft_\phi Z'\}| = |\{Z \trianglelefteq_\phi Z'\}| \times auto(Z')$$

Orthogonal fragments. A subfragment is a partial complex (subsection 6.1) that embeds in a fragment. It is an important "technical object" in our method. It shows up when we compute production rates for fragments whose concentration is affected by the dissociation of a solid bond $Z-Z'$. Such a dissociation will give rise to a piece Z (and also Z') that might embed into a fragment \mathcal{F} . As the lhs component of a rule, $Z-Z'$

embeds into a fragment and so will Z by virtue of the fragment growth process (see main text). To determine the contribution of the dissociation rule to the production rate of \mathcal{F} , we need the concentration of $\mathcal{F}-Z'$. Yet, $\mathcal{F}-Z'$ is not itself a fragment, but rather a subfragment. For our method to result in a closed system of equations, we must be able to express the concentration of this subfragment in terms of fragments. It turns out that this is indeed the case for any subfragment, as we show next. (We cannot, in general, extend a fragment and express its concentration using other fragments, as this would require the independence conditions, equation [10], to hold. As we saw in Example 2, these conditions do not hold in general.)

To compute the concentration of a subfragment, we need to use fragments that extend it. But we must be careful about which fragments we use. A subfragment and a fragment, each identifies a set of fully specified molecular species into which they embed (i.e. their extension, as explained in the main text, section “From rules to ODEs”). The concentration of a subfragment is the sum total of the concentrations of these species (weighted by appropriate symmetry related constants). If we are to use a combination of fragment concentrations, we must ensure that the fragments used indeed partition the set of molecular species into which the subfragment expands, or we would overcount. A set of fragments that complies with this requirement is called *orthogonal*.

Let Z be a subfragment, \mathfrak{F} the set of fragments, and $\mathcal{F}_1, \mathcal{F}_2 \in \mathfrak{F}$ two fragments that contain Z : $Z \triangleleft_{\phi} \mathcal{F}_1$ and $Z \triangleleft_{\phi'} \mathcal{F}_2$. (See section 1.4, paragraph labeled “embedding”, for a definition of the embedding relation \triangleleft_{ϕ} .) 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 expressing the concentration of the subfragment Z in terms of fragments, since we must avoid double counting matching instances of Z in the reaction mixture. The set of fragments $\{\mathcal{F}_1, \dots, \mathcal{F}_n\}$ 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 following the directives provided by p will lead us to a unique target node T , which we denote by $A.p$. The path p is expressed as a sequence of bonds to travel. For example, $A.p$ might result in $A.a.s_1.B_1.s_2 \dots s_{2i-1}.B_i.s_{2i} \dots t.T$ (with $s_{2i-1} \neq s_{2i}$), which is a path that goes from the originating agent A to agent B_1 over a link between site a of A and site s_1 of agent B_1 , and from there to agent B_2 over a link between site s_2 of agent B_1 and site s_3 of agent B_2 , and so on, to enter the target agent T through its site t . If any step some site is missing, then $A.p$ is not defined in the complex in question. An empty path means that we stay at the originating agent A .

Definition 6.2 (Orthogonal fragments).

Let Z be a subfragment. Let \mathcal{F}_1 and \mathcal{F}_2 be two fragments, and ϕ, ϕ' be embedding classes such that $Z \trianglelefteq_{\phi} \mathcal{F}_1$ and $Z \trianglelefteq_{\phi'} \mathcal{F}_2$. We say (\mathcal{F}_1, ϕ) and (\mathcal{F}_2, ϕ') are Z -orthogonal, 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 subfragment Z , yielding more than one embedding class ϕ . Let us collect all embedding classes mapping Z to some fragment, and define $C(Z)$ as a largest set of such embedding classes (it need not be unique) that are mutually Z -orthogonal. Write n_i for the number of embedding classes ϕ in $C(Z)$ with target \mathcal{F}_i , ie such that $Z \trianglelefteq_{\phi} \mathcal{F}_i$.

The concentration of Z can now be expressed as:

$$[Z] = \frac{1}{|\{\phi | Z \triangleleft_{\phi} Z\}|} \sum_i n_i [\mathcal{F}_i] \text{auto}(\mathcal{F}_i) \quad [17]$$

Equation [17] 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 choosing 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 .

6.4 Assembling the dynamical system for fragments – Version A. We assemble the system of differential equations for the fragments by determining the mass action terms that each rule type contributes. The description offered in this subsection has a deliberately “algorithmic” flavor. In the next subsection, we offer a version B that offers a more concise and abstract presentation which the reader might also find useful.

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 multiple 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, $\text{auto}(\text{lhs})$, of the lhs: $\gamma = k / \text{auto}(\text{lhs})$. Thanks to [17] we can refer to the concentration of any subfragment, $[Z]$, and be sure we can replace with an expansion into fragment concentrations. To indicate that we are building up differential equations sequentially, we shall use the symbol \pm as meaning “add this term to the previous ones for this equation”.

$Z, Z' \longrightarrow Z^*, Z' @ k$

This type of rule modifies the partial complex Z .

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

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm -\gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z'] \text{auto}(Z')).$$

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 embedded in \mathcal{F}_i . The said quantification, here as well as in the subsequent cases, is over embedding classes, not over plain embeddings.

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

$$\forall \mathcal{F}_k \forall \phi \text{ s.t. } Z^* \trianglelefteq_{\phi} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z \trianglelefteq_{\phi^*} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^* : \quad \frac{d[\mathcal{F}_k]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z'] \text{auto}(Z')).$$

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 ϕ^* 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^* .

Z, Z' \longrightarrow Z-Z' @ 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

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z'] \text{auto}(Z')).$$

Likewise for Z' :

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z' \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z] \text{auto}(Z)).$$

Production terms. On the production side, we must distinguish between solid and soft links in the ACM.

Z-Z' solid link

$$\forall \mathcal{F}_k \forall \phi_1, \phi_2 \text{ s.t. } Z-Z' \trianglelefteq_{\phi_1 \uplus \phi_2} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z \trianglelefteq_{\phi_1^*} \mathcal{F}_i \text{ and } \forall \mathcal{F}_j \text{ s.t. } Z' \trianglelefteq_{\phi_2^*} \mathcal{F}_j \text{ and } \mathcal{F}_k = \mathcal{F}_i - \mathcal{F}_j :$$

$$\frac{d[\mathcal{F}_k]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([\mathcal{F}_j] \text{auto}(\mathcal{F}_j)).$$

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 , $\text{dom}(\phi_1)$, and ϕ_2 , $\text{dom}(\phi_2)$, have an empty intersection, and $(\phi_1 \uplus \phi_2)(x) = \phi_1(x)$ if $x \in \text{dom}(\phi_1)$ or $(\phi_1 \uplus \phi_2)(x) = \phi_2(x)$ if $x \in \text{dom}(\phi_2)$.

Z-Z' soft 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.

$$\forall \mathcal{F}_k \forall \phi \text{ s.t. } Z^{B@b} \trianglelefteq_{\phi} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z \trianglelefteq_{\phi^*} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{B@b} : \quad \frac{d[\mathcal{F}_k]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z'] \text{auto}(Z')).$$

Likewise for Z' :

$$\forall \mathcal{F}_k \forall \phi \text{ s.t. } Z'^{A@a} \trianglelefteq_{\phi} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z' \trianglelefteq_{\phi^*} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{A@a} : \quad \frac{d[\mathcal{F}_k]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i))([Z] \text{auto}(Z)).$$

Z-Z' \longrightarrow Z, Z' @ k

This type of rule dissociates the partial complex $Z-Z'$.

Consumption terms. On the consumption side, we must distinguish between a solid and a soft link in the ACM.

Z-Z' solid link

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

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z-Z' \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i)).$$

Z-Z' soft link (By definition, this is a “pure 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.

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z^{B@b} \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i)).$$

Likewise for Z' :

$$\forall \mathcal{F}_i \forall \phi \text{ s.t. } Z'^{A@a} \trianglelefteq_{\phi} \mathcal{F}_i : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_i] \text{auto}(\mathcal{F}_i)).$$

Production terms. Here, too, we must distinguish between solid and soft links.

Z-Z' solid link

$$\forall \mathcal{F}_i \forall \phi_Z \text{ s.t. } Z \leq_{\phi_Z} \mathcal{F}_i \text{ and } \forall \mathcal{F}_k \forall \phi_{Z'} \text{ s.t. } \mathcal{F}_i - Z' \leq_{\phi_Z^* \uplus \phi_{Z'}} \mathcal{F}_k \text{ and } (\mathcal{F}_k, \phi_Z^* \uplus \phi_{Z'}) \in C(\mathcal{F}_i - Z') :$$

$$\frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_k] \text{auto}(\mathcal{F}_k)).$$

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 condition that $\mathcal{F}_i - Z' \leq_{\phi_Z^* \uplus \phi_{Z'}} \mathcal{F}_k$. 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 6.3, 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.

Z-Z' soft 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.

$$\forall \mathcal{F}_k \forall \phi \text{ s.t. } Z^{B@b} \leq_{\phi} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z \leq_{\phi^*} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{B@b} : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_k] \text{auto}(\mathcal{F}_k)).$$

Likewise for Z' :

$$\forall \mathcal{F}_k \forall \phi \text{ s.t. } Z'^{A@a} \leq_{\phi} \mathcal{F}_k \text{ and } \forall \mathcal{F}_i \text{ s.t. } Z' \leq_{\phi^*} \mathcal{F}_i \text{ and } \mathcal{F}_k = \mathcal{F}_i^{A@a} : \quad \frac{d[\mathcal{F}_i]}{dt} \pm \gamma([\mathcal{F}_k] \text{auto}(\mathcal{F}_k)).$$

Our implementation is actually more straightforward and uniform than this enumeration 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. Version B to which we turn now does present things in a more uniform way.

6.5 Assembling the dynamical system for fragments – Version B (not for the faint of heart). In this section we show more abstractly how to write the system of ODEs governing the time evolution of ACM-based fragments. To simplify things and expedite the presentation, we suppose that all bonds are solid; we also suppose that rule actions involve no agent deletion or creation, i.e. all actions are reversible, and we write α^{-1} for the action inverse to α . We will also use a reinforced version of property Q1 (defined in the main text); see below. We only aim at giving a sense of the general construction, not its full development and justification, which will be detailed elsewhere.

Given a global state of the system x , a rule r , and a fragment \mathcal{F} , we want to express how r affects the concentration of \mathcal{F} . Specifically, we want to express the negative (consumption) and positive (production) terms coming from r as functions of fragment concentrations (self-consistency). The overall differential equation for \mathcal{F} is then obtained by summing the contributions of all rules to \mathcal{F} (see below for a complete example).

In this section it is crucial to keep in mind the distinction made in the subsection above on “Counting embeddings”. As a reminder, there are two possibilities for defining the “concentration” of \mathcal{F} (or any complex) in a mixture x : the number of embeddings of \mathcal{F} in the reaction mixture, which we denote by $|\mathcal{F}; x|$, and the *discounted concentration* $|\mathcal{F}; x|/|\mathcal{F}; \mathcal{F}|$ familiar from deterministic chemical kinetics. In this section we shall use exclusively the number of embeddings and denote it with $[\mathcal{F}; x]$ (henceforth omitting the $|\cdot|$ for cardinals), being aware that we are abusing notation (and we shall do so even more below).

We next list a few conventions about embeddings, and rules.

- Given s, x two Kappa expressions – that is: mixtures (fully or partially specified), complexes (fully or partially specified), rule components, etc. – $[s; x]$ stands for the set of embeddings of s in x ;
- We will *not* use the traditional discounted concentrations, $[A; x]/[A; A]$, but rather the number of embeddings $[A; x]$, which we will sometimes write as $[A]$ when x is clear from the context.
- We will use a more convenient notation for rules where $r = s, \alpha, k$ consists of a left hand side s , an action which can be any combination of (internal state) modification, binding and unbinding, and a rate $k > 0$. This formulation is equivalent to the more intuitive $lhs \rightarrow rhs @ k$ notation that we have used so far.
- Given a rule $r = s, \alpha, k$, a mixture x , the set of events (i.e. rule applications) associated with r in x is by definition in bijective correspondence with $[s; x]$ (we distinguish here between an embedding of s into x and the event that it determines). Given $f \in [s; x]$, we write $f(\alpha) \cdot x$ for the outcome of the r -event associated with f .
- We write $s_1 \subseteq_* s$ when s_1 is a tuple of components of s that is modified by α .
- The rule *activity* is plainly $k[s; x]$ (ie we do not divide the rate by $[s; s]$ in this context which avoids the γ of the first presentation above). This rule activity is the expected number of applications of r per time unit, i.e. the flux of r in the ODE limit.

In general the rate at which a rule $r = s, \alpha, k$ consumes/produces embeddings of \mathcal{F} is:

$$\delta_r(\mathcal{F}) = k \sum_{f \in [s; x]} ([\mathcal{F}; f(\alpha) \cdot x] - [\mathcal{F}; x])$$

Indeed, the activity of r can be written $k \sum_{f \in [s; x]} 1$ and is the expected number of applications of r per time unit. The above formula expresses the expected change in the number of embeddings of \mathcal{F} due to the r -events occurring (synchronously) per time unit. Note that a rule can both consume and produce \mathcal{F} in the same event. We will evaluate both contributions of r to $[\mathcal{F}; x]$, written as $\delta_r^-(\mathcal{F})$ and $\delta_r^+(\mathcal{F})$, separately.

To evaluate the activity of a rule, we will assume, as is customary in deterministic chemical kinetics, a negligible occurrence of situations in which the embeddings of lhs components are not jointly injective. In other words, if Z and Z' are lhs rule components, we will overestimate $[Z, Z'; x]$ as $[Z; x][Z'; x]$.

Consumption terms

Choose an $f \in [s; x]$. For f to consume \mathcal{F} , there must be a modified lhs component s_1 whose image under f intersects an occurrence of \mathcal{F} in x on modified sites. By Q1 (see main text) f factorizes as $f = \gamma(\phi + I)$, with I the identity on $s \setminus s_1$, $\phi \in [s_1; \mathcal{F}]$, and $\gamma \in [\mathcal{F}, s \setminus s_1; x]$.

We can summarise this factorization as follows:

$$s_1, s \setminus s_1 \xrightarrow{\phi + I} \mathcal{F}, s \setminus s_1 \xrightarrow{\gamma} x$$

If we overapproximate $[\mathcal{F}, s \setminus s_1; x]$ as $[\mathcal{F}; x][s \setminus s_1; x]$, we get the following bijective enumeration of the \mathcal{F} -consuming events associated with r :

$$\forall s_1 \subseteq_\star s, \forall \phi \in [s_1; \mathcal{F}] : \delta_{r, \phi}^-(\mathcal{F}) = k[\mathcal{F}, s \setminus s_1] = k[\mathcal{F}][s \setminus s_1] \quad [18]$$

By Q2-3 (see main text), each contribution can be expressed as a function of fragment concentrations.

Given f , the ϕ part of the factorisation, and therefore also the γ part, is only determined up to a symmetry $\sigma \in [\mathcal{F}; \mathcal{F}]$. Indeed, if $\gamma\phi = f$, then $\gamma\sigma\sigma^{-1}\phi = f$. To avoid a redundant count, it appears that one should divide the above terms by $[\mathcal{F}; \mathcal{F}]$. However, each factorisation consumes $[\mathcal{F}; \mathcal{F}]$ embeddings of \mathcal{F} in x , and both cancel each other out. Note that we use here a stronger version of Q1, namely that the γ, ϕ decomposition is unique up to symmetries of \mathcal{F} -this is only to make things simpler and is not a requirement of our algorithm.

Production terms

Choose an $f \in [s; x]$, and write $s' := \alpha \cdot s$ for the rule rhs, $x' := f(\alpha) \cdot x$ for the state obtained by triggering the r -event associated with f , and $f' \in [s'; x']$ for the unique post-event embedding corresponding to f .

For f to produce an \mathcal{F} there must be a modified rhs component $s'_1 \subseteq s'$ that factorises f' as $f' = \gamma'(\phi' + I)$ with $\phi' \in [s'_1; \mathcal{F}]$, for some γ' uniquely up to $[\mathcal{F}; \mathcal{F}]$. We can summarize the situation as follows:

$$\begin{array}{ccccc} s' = s'_1, s' \setminus s'_1 & \xrightarrow{\phi' + I} & \mathcal{F}, s' \setminus s'_1 & \xrightarrow{\gamma'} & x' \\ \uparrow \alpha & & \downarrow (\phi' + I)(\alpha^{-1}) & & \uparrow f(\alpha) \\ s & \xrightarrow{\phi + I} & (\phi' + I)(\alpha^{-1})(\mathcal{F}, s' \setminus s'_1) & \xrightarrow{\gamma} & x \end{array}$$

where the choice of ϕ' uniquely determines ϕ , and therefore γ . Hence, \mathcal{F} -producing f s are in bijection with $s'_1 \subseteq_\star \alpha \cdot s$, $\phi' \in [s'_1; \mathcal{F}]$ up to $[\mathcal{F}; \mathcal{F}]$, and $\gamma \in [(\phi' + I)(\alpha^{-1}) \cdot (\mathcal{F}, s' \setminus s'_1); x]$.

Putting everything together we get a bijective enumeration of the \mathcal{F} -producing events associated with r :

$$\forall s'_1 \subseteq_\star \alpha \cdot s, \forall \phi' \in [s'_1; \mathcal{F}] : \delta_{r, \phi'}^+(\mathcal{F}) = k[(\phi' + I)(\alpha^{-1}) \cdot (\mathcal{F}, \alpha \cdot s \setminus s'_1)] \quad [19]$$

As in the consumption case, given f , the ϕ' part of the factorization is only unique up to $[\mathcal{F}; \mathcal{F}]$, but since each choice creates $[\mathcal{F}; \mathcal{F}]$ new embeddings of \mathcal{F} , the terms are correct.

From the definition of fragments (see main text) it easily follows that the components of $(\phi' + I)(\alpha^{-1}) \cdot (\mathcal{F}, \alpha \cdot s \setminus s'_1)$ are subfragments. By Q2-3, this is enough to guarantee that each contribution is expressible solely in terms of fragment concentrations.

A concrete dissociation example

We can illustrate the construction of the production terms with a concrete case of a dissociation rule.

Set $s = A(a^1), B(b^1)$, with $\alpha \cdot s = s' = A(a), B(b)$, and:

$$\begin{aligned} \mathcal{F} &= A_1(a, x^1), A_2(x^1, a) \\ x &= N * (A_1(a, x^1), A_2(x^1, a^2), B(b^2)), \end{aligned}$$

where the latter expression means that the mixture x consists of N copies of the indicated expression. The indices on the A s distinguish the two occurrences in \mathcal{F} . The r -events are of the form $f_i = (A, B \mapsto A_{2i}, B)$ and each produces \mathcal{F} (twice). Specifically, one has $[\mathcal{F}; x] = 0$, and $[\mathcal{F}; f_i(\alpha) \cdot x] = 2$. The general formula mentions a s'_1 which can only be $A(a)$ in this case, so that $\alpha \cdot s \setminus s'_1 = B(b)$, and a ϕ' which can be either $\phi'_1 := (A \mapsto A_1)$ or $\phi'_2 := (A \mapsto A_2)$. By inverting the rule along both extensions $\phi'_i + I$, we get isomorphic partial complexes:

$$\begin{aligned} (\phi'_1 + I)(\alpha^{-1}) \cdot (\mathcal{F}, B(b)) &= A_1(a^2, x^1), A_2(x^1, a), B(b^2) \\ (\phi'_2 + I)(\alpha^{-1}) \cdot (\mathcal{F}, B(b)) &= A_1(a, x^1), A_2(x^1, a^2), B(b^2) \end{aligned}$$

with equal contributions $\delta_{r, \phi'_1}^+(F) = \delta_{r, \phi'_2}^+(F) = kN$, and we find that the total rate at which embeddings of \mathcal{F} are produced is $2kN$, as it should.

7 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 5.1. The results have been obtained entirely by static analysis, as detailed in the main text and section 6 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. (It took 0.42s, if we include automatic LaTeX report generation.) 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.

7.1 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 5.2. $\mathcal{R}_{\text{frag}}^{\text{rule}}$ 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 $\mathcal{R}_7^{39} = \mathcal{F}_1$ 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 $\mathcal{R}_{\text{frag}}^{\text{rule}}$ equation should be multiplied by the rate constant associated with the rule indicated in the superscript.)

Kinetic terms generated from rule cr39:

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

Kinetic terms generated from rule cr38:

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

Kinetic terms generated from rule cr37:

$$\begin{aligned}\mathcal{R}_9^{37} &= \mathcal{F}_8 \\ \mathcal{R}_8^{37} &= -\mathcal{F}_8 \\ \mathcal{R}_5^{37} &= \mathcal{F}_8\end{aligned}$$

Kinetic terms generated from rule cr36:

$$\begin{aligned}\mathcal{R}_9^{36} &= -\mathcal{F}_5 \cdot \mathcal{F}_9 \\ \mathcal{R}_8^{36} &= \mathcal{F}_5 \cdot \mathcal{F}_9 \\ \mathcal{R}_5^{36} &= -\mathcal{F}_5 \cdot \mathcal{F}_9\end{aligned}$$

Kinetic terms generated from rule cr35:

$$\begin{aligned}\mathcal{R}_{11}^{35} &= \mathcal{F}_{10} \\ \mathcal{R}_{10}^{35} &= -\mathcal{F}_{10} \\ \mathcal{R}_9^{35} &= \mathcal{F}_{10}\end{aligned}$$

Kinetic terms generated from rule cr34:

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

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

Kinetic terms generated from rule cr33:

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

Kinetic terms generated from rule cr32:

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

Kinetic terms generated from rule cr31:

$$\begin{aligned}\mathcal{R}_{17}^{31} &= \mathcal{F}_1 \\ \mathcal{R}_{16}^{31} &= \mathcal{F}_2 \\ \mathcal{R}_{15}^{31} &= \mathcal{F}_3 \\ \mathcal{R}_8^{31} &= \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3 \\ \mathcal{R}_3^{31} &= -\mathcal{F}_3 \\ \mathcal{R}_2^{31} &= -\mathcal{F}_2 \\ \mathcal{R}_1^{31} &= -\mathcal{F}_1\end{aligned}$$

Kinetic terms generated from rule cr30:

$$\begin{aligned}\mathcal{R}_{17}^{30} &= -\mathcal{F}_8 \cdot \mathcal{F}_{17} \\ \mathcal{R}_{16}^{30} &= -\mathcal{F}_8 \cdot \mathcal{F}_{16} \\ \mathcal{R}_{15}^{30} &= -\mathcal{F}_8 \cdot \mathcal{F}_{15} \\ \mathcal{R}_8^{30} &= -\mathcal{F}_8 \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17}) \\ \mathcal{R}_3^{30} &= \mathcal{F}_8 \cdot \mathcal{F}_{15} \\ \mathcal{R}_2^{30} &= \mathcal{F}_8 \cdot \mathcal{F}_{16} \\ \mathcal{R}_1^{30} &= \mathcal{F}_8 \cdot \mathcal{F}_{17}\end{aligned}$$

Kinetic terms generated from rule cr29:

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

Kinetic terms generated from rule cr28:

$$\begin{aligned}\mathcal{R}_{17}^{28} &= -\mathcal{F}_{10} \cdot \mathcal{F}_{17} \\ \mathcal{R}_{16}^{28} &= -\mathcal{F}_{10} \cdot \mathcal{F}_{16} \\ \mathcal{R}_{15}^{28} &= -\mathcal{F}_{10} \cdot \mathcal{F}_{15} \\ \mathcal{R}_{14}^{28} &= \mathcal{F}_{10} \cdot \mathcal{F}_{15}\end{aligned}$$

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

Kinetic terms generated from rule cr27:

$$\begin{aligned}
\mathcal{R}_{17}^{27} &= \mathcal{F}_7 \\
\mathcal{R}_{16}^{27} &= \mathcal{F}_6 \\
\mathcal{R}_{15}^{27} &= \mathcal{F}_4 \\
\mathcal{R}_9^{27} &= \mathcal{F}_4 + \mathcal{F}_6 + \mathcal{F}_7 \\
\mathcal{R}_7^{27} &= -\mathcal{F}_7 \\
\mathcal{R}_6^{27} &= -\mathcal{F}_6 \\
\mathcal{R}_4^{27} &= -\mathcal{F}_4
\end{aligned}$$

Kinetic terms generated from rule cr26:

$$\begin{aligned}
\mathcal{R}_{17}^{26} &= -\mathcal{F}_9 \cdot \mathcal{F}_{17} \\
\mathcal{R}_{16}^{26} &= -\mathcal{F}_9 \cdot \mathcal{F}_{16} \\
\mathcal{R}_{15}^{26} &= -\mathcal{F}_9 \cdot \mathcal{F}_{15} \\
\mathcal{R}_9^{26} &= -\mathcal{F}_9 \cdot (\mathcal{F}_{15} + \mathcal{F}_{16} + \mathcal{F}_{17}) \\
\mathcal{R}_7^{26} &= \mathcal{F}_9 \cdot \mathcal{F}_{17} \\
\mathcal{R}_6^{26} &= \mathcal{F}_9 \cdot \mathcal{F}_{16} \\
\mathcal{R}_4^{26} &= \mathcal{F}_9 \cdot \mathcal{F}_{15}
\end{aligned}$$

Kinetic terms generated from rule cr25:

$$\begin{aligned}
\mathcal{R}_{21}^{25} &= \mathcal{F}_{18} + \mathcal{F}_{19} + \mathcal{F}_{20} \\
\mathcal{R}_{20}^{25} &= -\mathcal{F}_{20} \\
\mathcal{R}_{19}^{25} &= -\mathcal{F}_{19} \\
\mathcal{R}_{18}^{25} &= -\mathcal{F}_{18} \\
\mathcal{R}_{17}^{25} &= \mathcal{F}_{18} \\
\mathcal{R}_{16}^{25} &= \mathcal{F}_{19} \\
\mathcal{R}_{15}^{25} &= \mathcal{F}_{20}
\end{aligned}$$

Kinetic terms generated from rule cr24:

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

Kinetic terms generated from rule cr23:

$$\begin{aligned}
\mathcal{R}_{22}^{23} &= \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3 \\
\mathcal{R}_{14}^{23} &= \mathcal{F}_3 \\
\mathcal{R}_{13}^{23} &= \mathcal{F}_2 \\
\mathcal{R}_{12}^{23} &= \mathcal{F}_1 \\
\mathcal{R}_3^{23} &= -\mathcal{F}_3 \\
\mathcal{R}_2^{23} &= -\mathcal{F}_2 \\
\mathcal{R}_1^{23} &= -\mathcal{F}_1
\end{aligned}$$

Kinetic terms generated from rule cr22:

$$\begin{aligned}
\mathcal{R}_{22}^{22} &= -\mathcal{F}_{22} \cdot (\mathcal{F}_{12} + \mathcal{F}_{13} + \mathcal{F}_{14}) \\
\mathcal{R}_{14}^{22} &= -\mathcal{F}_{14} \cdot \mathcal{F}_{22} \\
\mathcal{R}_{13}^{22} &= -\mathcal{F}_{13} \cdot \mathcal{F}_{22}
\end{aligned}$$

$$\begin{aligned}\mathcal{R}_{12}^{22} &= -\mathcal{F}_{12} \cdot \mathcal{F}_{22} \\ \mathcal{R}_3^{22} &= \mathcal{F}_{14} \cdot \mathcal{F}_{22} \\ \mathcal{R}_2^{22} &= \mathcal{F}_{13} \cdot \mathcal{F}_{22} \\ \mathcal{R}_1^{22} &= \mathcal{F}_{12} \cdot \mathcal{F}_{22}\end{aligned}$$

Kinetic terms generated from rule cr21:

$$\begin{aligned}\mathcal{R}_{22}^{21} &= \mathcal{F}_8 \\ \mathcal{R}_{10}^{21} &= \mathcal{F}_8 \\ \mathcal{R}_8^{21} &= -\mathcal{F}_8\end{aligned}$$

Kinetic terms generated from rule cr20:

$$\begin{aligned}\mathcal{R}_{22}^{20} &= -\mathcal{F}_{10} \cdot \mathcal{F}_{22} \\ \mathcal{R}_{10}^{20} &= -\mathcal{F}_{10} \cdot \mathcal{F}_{22} \\ \mathcal{R}_8^{20} &= \mathcal{F}_{10} \cdot \mathcal{F}_{22}\end{aligned}$$

Kinetic terms generated from rule cr19:

$$\begin{aligned}\mathcal{R}_{22}^{19} &= \mathcal{F}_5 \\ \mathcal{R}_{11}^{19} &= \mathcal{F}_5 \\ \mathcal{R}_5^{19} &= -\mathcal{F}_5\end{aligned}$$

Kinetic terms generated from rule cr18:

$$\begin{aligned}\mathcal{R}_{22}^{18} &= -\mathcal{F}_{11} \cdot \mathcal{F}_{22} \\ \mathcal{R}_{11}^{18} &= -\mathcal{F}_{11} \cdot \mathcal{F}_{22} \\ \mathcal{R}_5^{18} &= \mathcal{F}_{11} \cdot \mathcal{F}_{22}\end{aligned}$$

Kinetic terms generated from rule cr17:

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

Kinetic terms generated from rule cr16:

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

Kinetic terms generated from rule cr15:

$$\begin{aligned}\mathcal{R}_{31}^{15} &= \mathcal{F}_{23} \\ \mathcal{R}_{30}^{15} &= \mathcal{F}_{24} \\ \mathcal{R}_{29}^{15} &= \mathcal{F}_{25} \\ \mathcal{R}_{25}^{15} &= -\mathcal{F}_{25} \\ \mathcal{R}_{24}^{15} &= -\mathcal{F}_{24} \\ \mathcal{R}_{23}^{15} &= -\mathcal{F}_{23} \\ \mathcal{R}_5^{15} &= \mathcal{F}_{23} + \mathcal{F}_{24} + \mathcal{F}_{25}\end{aligned}$$

Kinetic terms generated from rule cr14:

$$\begin{aligned}
\mathcal{R}_{31}^{14} &= -\mathcal{F}_5 \cdot \mathcal{F}_{31} \\
\mathcal{R}_{30}^{14} &= -\mathcal{F}_5 \cdot \mathcal{F}_{30} \\
\mathcal{R}_{29}^{14} &= -\mathcal{F}_5 \cdot \mathcal{F}_{29} \\
\mathcal{R}_{25}^{14} &= \mathcal{F}_5 \cdot \mathcal{F}_{29} \\
\mathcal{R}_{24}^{14} &= \mathcal{F}_5 \cdot \mathcal{F}_{30} \\
\mathcal{R}_{23}^{14} &= \mathcal{F}_5 \cdot \mathcal{F}_{31} \\
\mathcal{R}_5^{14} &= -\mathcal{F}_5 \cdot (\mathcal{F}_{29} + \mathcal{F}_{30} + \mathcal{F}_{31})
\end{aligned}$$

Kinetic terms generated from rule cr13:

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

Kinetic terms generated from rule cr12:

$$\begin{aligned}
\mathcal{R}_{31}^{12} &= -\mathcal{F}_{11} \cdot \mathcal{F}_{31} \\
\mathcal{R}_{30}^{12} &= -\mathcal{F}_{11} \cdot \mathcal{F}_{30} \\
\mathcal{R}_{29}^{12} &= -\mathcal{F}_{11} \cdot \mathcal{F}_{29} \\
\mathcal{R}_{28}^{12} &= \mathcal{F}_{11} \cdot \mathcal{F}_{31} \\
\mathcal{R}_{27}^{12} &= \mathcal{F}_{11} \cdot \mathcal{F}_{30} \\
\mathcal{R}_{26}^{12} &= \mathcal{F}_{11} \cdot \mathcal{F}_{29} \\
\mathcal{R}_{11}^{12} &= -\mathcal{F}_{11} \cdot (\mathcal{F}_{29} + \mathcal{F}_{30} + \mathcal{F}_{31})
\end{aligned}$$

Kinetic terms generated from rule cr11:

$$\begin{aligned}
\mathcal{R}_{21}^{11} &= \mathcal{F}_9 \\
\mathcal{R}_9^{11} &= -\mathcal{F}_9
\end{aligned}$$

Kinetic terms generated from rule cr10:

$$\begin{aligned}
\mathcal{R}_{20}^{10} &= \mathcal{F}_4 \\
\mathcal{R}_{19}^{10} &= \mathcal{F}_6 \\
\mathcal{R}_{18}^{10} &= \mathcal{F}_7 \\
\mathcal{R}_7^{10} &= -\mathcal{F}_7 \\
\mathcal{R}_6^{10} &= -\mathcal{F}_6 \\
\mathcal{R}_4^{10} &= -\mathcal{F}_4
\end{aligned}$$

Kinetic terms generated from rule cr9:

$$\begin{aligned}
\mathcal{R}_{18}^9 &= -\mathcal{F}_{18} \\
\mathcal{R}_7^9 &= \mathcal{F}_{18}
\end{aligned}$$

Kinetic terms generated from rule cr8:

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

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:

$$\mathcal{R}_{36}^4 = \mathcal{F}_{35}$$

$$\mathcal{R}_{35}^4 = -\mathcal{F}_{35}$$

$$\mathcal{R}_{33}^4 = \mathcal{F}_{32}$$

$$\mathcal{R}_{32}^4 = -\mathcal{F}_{32}$$

$$\mathcal{R}_{31}^4 = -\mathcal{F}_{31}$$

$$\mathcal{R}_{30}^4 = \mathcal{F}_{31}$$

$$\mathcal{R}_{28}^4 = -\mathcal{F}_{28}$$

$$\mathcal{R}_{27}^4 = \mathcal{F}_{28}$$

$$\mathcal{R}_{24}^4 = \mathcal{F}_{23}$$

$$\mathcal{R}_{23}^4 = -\mathcal{F}_{23}$$

$$\mathcal{R}_{19}^4 = \mathcal{F}_{18}$$

$$\mathcal{R}_{18}^4 = -\mathcal{F}_{18}$$

$$\mathcal{R}_{17}^4 = -\mathcal{F}_{17}$$

$$\mathcal{R}_{16}^4 = \mathcal{F}_{17}$$

$$\mathcal{R}_{13}^4 = \mathcal{F}_{12}$$

$$\mathcal{R}_{12}^4 = -\mathcal{F}_{12}$$

$$\mathcal{R}_7^4 = -\mathcal{F}_7$$

$$\mathcal{R}_6^4 = \mathcal{F}_7$$

$$\mathcal{R}_2^4 = \mathcal{F}_1$$

$$\mathcal{R}_1^4 = -\mathcal{F}_1$$

Kinetic terms generated from rule cr3:

$$\mathcal{R}_{36}^3 = -\mathcal{F}_{36} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{35}^3 = \mathcal{F}_{36} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{33}^3 = -\mathcal{F}_{33} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{32}^3 = \mathcal{F}_{33} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{31}^3 = \mathcal{F}_{30} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{30}^3 = -\mathcal{F}_{30} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{28}^3 = \mathcal{F}_{27} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{27}^3 = -\mathcal{F}_{27} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{24}^3 = -\mathcal{F}_{24} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{23}^3 = \mathcal{F}_{24} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{19}^3 = -\mathcal{F}_{19} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{18}^3 = \mathcal{F}_{19} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{17}^3 = \mathcal{F}_{16} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{16}^3 = -\mathcal{F}_{16} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{13}^3 = -\mathcal{F}_{13} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\mathcal{R}_{12}^3 = \mathcal{F}_{13} \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})$$

$$\begin{aligned}
\mathcal{R}_7^3 &= \mathcal{F}_6 \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36}) \\
\mathcal{R}_6^3 &= -\mathcal{F}_6 \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36}) \\
\mathcal{R}_2^3 &= -\mathcal{F}_2 \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36}) \\
\mathcal{R}_1^3 &= \mathcal{F}_2 \cdot (\mathcal{F}_{24} + \mathcal{F}_{27} + \mathcal{F}_{30} + \mathcal{F}_{36})
\end{aligned}$$

Kinetic terms generated from rule cr2:

$$\begin{aligned}
\mathcal{R}_{38}^2 &= -\mathcal{F}_{38} \cdot (\mathcal{F}_{25} + \mathcal{F}_{26} + \mathcal{F}_{29} + \mathcal{F}_{37}) \\
\mathcal{R}_{37}^2 &= -\mathcal{F}_{37} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{36}^2 &= \mathcal{F}_{37} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{34}^2 &= -\mathcal{F}_{34} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{33}^2 &= \mathcal{F}_{34} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{30}^2 &= \mathcal{F}_{29} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{29}^2 &= -\mathcal{F}_{29} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{27}^2 &= \mathcal{F}_{26} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{26}^2 &= -\mathcal{F}_{26} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{25}^2 &= -\mathcal{F}_{25} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{24}^2 &= \mathcal{F}_{25} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{20}^2 &= -\mathcal{F}_{20} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{19}^2 &= \mathcal{F}_{20} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{16}^2 &= \mathcal{F}_{15} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{15}^2 &= -\mathcal{F}_{15} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{14}^2 &= -\mathcal{F}_{14} \cdot \mathcal{F}_{38} \\
\mathcal{R}_{13}^2 &= \mathcal{F}_{14} \cdot \mathcal{F}_{38} \\
\mathcal{R}_6^2 &= \mathcal{F}_4 \cdot \mathcal{F}_{38} \\
\mathcal{R}_4^2 &= -\mathcal{F}_4 \cdot \mathcal{F}_{38} \\
\mathcal{R}_3^2 &= -\mathcal{F}_3 \cdot \mathcal{F}_{38} \\
\mathcal{R}_2^2 &= \mathcal{F}_3 \cdot \mathcal{F}_{38}
\end{aligned}$$

Kinetic terms generated from rule cr1:

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

7.2 The dynamical system for fragments.

$$\begin{aligned}
\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
\end{aligned}$$

8 A comparison

For the purpose of comparison, and as a further test case, we applied our procedure to a model of crosstalk between EGF and insulin receptors, treated by Conzelmann et al. in [11] (the "CFG model"). We also applied our procedure to a simplification of the CFG-model that consists in removing contextual specifications on the lhs of dissociation rules (turning them into "pure dissociation rules" per our terminology in the main paper, section on "Syntactical criteria for annotating the contact map".) The automatically generated reports are available as separate additional information, as they comprise 86 and 39 pages, respectively.

The CFG model (table 7 of [11]) consists of 76 rules giving rise to 2899 molecular species. Conzelmann et al. report their system to comprise 5182 species. The discrepancy is a consequence of operating without a formal agent-based language. The authors of [11] have therefore no way of accounting for symmetries that might be present in molecular species. Because of extensive dimer formation, symmetries are rampant, shrinking the number of distinguishable species by 44%. The method for reducing CFG as described in [11] yields 391 coarse-grained variables. Our automatic procedure yields 208 fragments. (In [11], the 391 variables are subsequently reduced to 87 by applying a strategy for detecting systemic modules, which is an altogether different method than coarse-graining as we understand it. This method could be applied to our ODE system as well.)

Context-dependency of dissociation causes soft bonds to become solid (see directive Edg1 in the main paper). This, in turn, gives rise to larger fragments. Larger fragments often entail more numerous fragments, since a family of fragments is obtained by generating all possible state valuations on a chosen class of sites. This effect is illustrated by removing the context-dependency of dissociation in the CFG model. The resulting simplified model still consists of 76 rules and preserves the 2899 possible species, but our procedure now generates only 88 fragments.

1. Danos, V & Laneve, C. *Formal Molecular Biology* (2004) *Theoretical Computer Science* 325, 69–110.
2. Access to the Kappa modeling platform is provided at www.cellucide.com.
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. Doob, J. L. *Markoff chains D denumerable case* (1945) *Trans. Amer. Math. Soc.* 58, 455D–473.
6. Gillespie, D. T. *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions* (1976) *Journal of Computational Physics* 22, 403–434.
7. Borisov, N. M, Markevich, N. I, Hoek, J. B, & Kholodenko, B. N. *Signaling through Receptors and Scaffolds: Independent Interactions Reduce Combinatorial Complexity* (2005) *Biophysical Journal* 89, 951–966.
8. 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.
9. Conzelmann, H. (2008) Ph.D. thesis (Institut für Systemdynamik der Universität Stuttgart).
10. 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.
11. Conzelmann, H, Fey, D, & Gilles, E. D. *Exact model reduction of combinatorial reaction networks* (2008) *BMC Systems Biology* 2, 78.