Autocatalysis in chemical networks: unifications and extensions

Alex Blokhuis,[1,2] David Lacoste,[1] and Philippe Nghe[2]

[1)] *Gulliver Laboratory, UMR CNRS 7083, PSL University, 10 rue Vauquelin, Paris F-75231, France*

[2)] *Laboratoire de Biochimie, Chimie Biologie et Innovation, ESPCI Paris, PSL University,10 rue Vauquelin, Paris F-75231, France*

(Dated: July 1, 2020)

Autocatalysis is essential for the origin of life and chemical evolution. However, the lack of unified framework so far prevents a systematic study of autocatalysis. Here, we derive general stoichiometric conditions for catalysis and autocatalysis in chemical reaction networks from basic principles. This allows for a classification of minimal autocatalytic motifs. While all known autocatalytic systems indeed contain minimal motifs, the classification also reveals hitherto unidentified motifs. We further examine conditions for kinetic viability of such networks, which depends on the autocatalytic motifs they contain and is notably increased by internal catalytic cycles. Finally, we show how this framework extends the range of conceivable autocatalytic systems, by applying our stoichiometric and kinetic analysis to autocatalysis emerging from coupled compartments. The unified approach to autocatalysis presented in this work lays a foundation towards the building of a systems-level theory of chemical evolution.

PACS numbers: 05.40.-a 82.65.+r 82.20.-w

INTRODUCTION

The capacity of living systems to replicate themselves is rooted in a chemistry that makes more of itself, i.e. an autocatalytic system. Autocatalysis appears to be ubiquitous in living systems from molecules to ecosystems[1]. It is also likely to have been continually present since the beginning of life and is invoked as a key element in prebiotic scenarios[2–5]. Surprisingly, autocatalysis is considered to be a rarity in chemistry[6]. Developments in systems chemistry are changing this view, with an increasing number of autocatalytic systems synthesized de novo[7–9]. Chemical replicators have been endowed with biomimetic properties such as protein-like folding[10] and parasitism[11]. Autocatalysis has also found technological applications, e.g. enantiomer enrichment and acid amplification[12–14].

Understanding autocatalysis represents a primary challenge for theory. Models based on autocatalysis were first built to explain a diversity of dynamical behaviors in so called dissipative structures, such as bistable reactions[15], oscillating reactions, and chemical waves[16]. Autocatalysis then became a central topic in the study of self-replication dynamics in biological and prebiotic systems[3,17–19] (see[20–22] for recent reviews).

Despite this history, a unified theory of autocatalysis is still lacking. Such a theory is needed to understand the origins, diversity and plausibility of autocatalysis. It would also provide design principles for artificial autocatalytic systems. Here, we present a framework that unifies the different descriptions of autocatalysis and is based on reaction network stoichiometry[23–27].

Let us start from basic definitions in chemistry as established by IUPAC[28] (see SI Sec. I for full definitions), where autocatalysis is a particular form of catalysis: *A substance that increases the rate of a reaction without modifying the overall standard Gibbs energy change (ΔG°) in the reaction; the process is called catalysis. The catalyst is both a reactant and product of the reaction. Catalysis brought about by one of the products of a (net) reaction is called autocatalysis.*

From this definition, we derive conditions to determine whether a subnetwork embedded in a larger chemical network, can be catalytic or autocatalytic. These conditions provide a mathematical basis to identify minimal motifs, called autocatalytic cores. We found that cores have five fundamental categories of motifs. They allow classification of all previously described forms of autocatalysis, and also reveal hitherto unidentified autocatalytic schemes. We then study the kinetic conditions, which we call viability conditions, under which autocatalytic networks can appear and be maintained on long times. We find that networks have different viabilities depending on their core structure, and notably that viability is increased by internal catalytic cycles. Finally, we expand the repertoire of autocatalytic systems, by demonstrating a general mechanism for its emergence in multicompartment systems (e.g. porous media, vesicles, multiphasic systems). This mechanism strongly relaxes chemical requirements for autocatalysis, making the phenomenon much more diverse than previously thought.

EXAMPLES, DEFINITIONS AND CONVENTIONS

Catalysis and autocatalysis

The following reactions have the same net mass balance but a different status regarding catalysis:

$$\mathrm{A} \overset{\text{(I)}}{\rightleftharpoons} \mathrm{B}, \quad \mathrm{A}+\mathrm{E} \overset{\text{(II)}}{\rightleftharpoons} \mathrm{B}+\mathrm{E}, \quad \mathrm{A}+\mathrm{B} \overset{\text{(III)}}{\rightleftharpoons} 2\mathrm{B}. \tag{1}$$

Since no species is both a reactant and product in reaction (I), it should be regarded as uncatalyzed. Reactions (II) and (III) instead contain species which are both a reactant and a product, species E in reaction (I) and species B in reaction (III) and following the definition above, these species can be considered as catalysts. In reaction (II), the amount of species E remains unchanged, in contrast to the case of reaction (III), where the species B experiences a net production. For this reason, reaction (III) represents genuine autocatalysis. Although reaction (II) is usually referred to as simply catalyzed in the chemistry literature, we propose to call it an example of allocatalysis to contrast it with the case of autocatalysis, catalysis being common to both.

We emphasize that stoichiometric considerations are necessary but not sufficient to characterize catalysis, which according to the definition should also accelerate the rate of the net reaction. In the following, we will first generalize the stoichiometric conditions, then examine kinetic ones.

Stoichiometric matrix and reaction vectors

Reaction networks are represented as a stoichiometric matrix $\boldsymbol{\nu}$[23,26], in which columns correspond to reactions and rows to species. The entries in a column are the stoichiometric coefficients of the species participating in that reaction, the coefficient is negative for every species consumed and positive for every species produced. A reaction vector $\boldsymbol{g} = [g_1, .., g_r]^T$ results in a change of species numbers $\Delta \boldsymbol{n} = \boldsymbol{\nu} \cdot \boldsymbol{g}$. The support of $\boldsymbol{g}$, denoted $\mathrm{supp}(\boldsymbol{g})$, is the set of its non-zero coordinates. A reaction cycle is a non-zero reaction vector $\boldsymbol{c}$ such that no net species number change occurs : $\boldsymbol{\nu} \cdot \boldsymbol{c} = \boldsymbol{0}$, or equivalently, $\mathbf{c}$ belongs to the right null space of $\boldsymbol{\nu}$. Vectors $\boldsymbol{b}^T$ belonging to the left null space of $\boldsymbol{\nu}$ induce conservation laws, because in that case $\boldsymbol{b} \cdot \boldsymbol{n}$ represents a conserved quantity. The case of all coefficients b_k nonnegative is referred to as a mass-like conservation law. For example in Fig.1a, conserved quantities are $n_{\mathrm{E}} + n_{\mathrm{EA}}$ (catalysts) and $n_{\mathrm{A}} + n_{\mathrm{EA}} + n_{\mathrm{B}}$ (total compounds).

Lastly, catalyzed reactions may not always be distinguished from uncatalyzed one in the stoichiometric matrix. For instance, in reactions (II-III), catalysts cancel on each side leading to the same column vector as for (I). This is avoided by describing catalysis through a sequence of reactions steps from which it emerges, so that a participating species is either a reactant or a product:

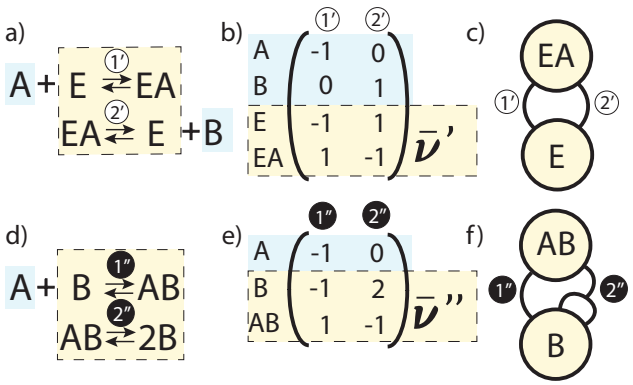

Figure 1: Different representations for allocatalysis (a,b,c) and autocatalysis (d,e,f). a) Combining reactions (1')+(2') affords an allocatalytic cycle that converts A to B. b) stoichiometric matrix of a), the dashed square encloses the allocatalytic submatrix $\bar{\boldsymbol{\nu}}'$ for network b). c) Graph representation of the allocatalytic subnetwork. d) Combining (1")+(2") affords an autocatalytic cycle converting A to B. e) stoichiometric matrix of d), the dashed square encloses the autocatalytic submatrix $\bar{\boldsymbol{\nu}}''$ for network e). f) a graph representation of the autocatalytic subnetwork.

$$\mathrm{A}+\mathrm{E} \overset{\text{IIa}}{\rightleftharpoons} \mathrm{EA} \overset{\text{IIb}}{\rightleftharpoons} \mathrm{E}+\mathrm{B}, \tag{2}$$

$$\mathrm{A}+\mathrm{B} \overset{\text{IIIa}}{\rightleftharpoons} \mathrm{AB} \overset{\text{IIIb}}{\rightleftharpoons} 2\mathrm{B} \quad . \tag{3}$$

We call this convention *non-ambiguity* and assume henceforth that it is respected.

CATALYSIS AND AUTOCATALYSIS IN STOICHIOMETRIC MATRICES

In this section, we will consider any possible submatrix $\bar{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}$, the stoichiometric matrix of a reaction network, and ask whether the stoichiometry of the corresponding subnetwork, called a motif, is compatible with the definitions of allocatalysis or autocatalysis. Note that such identification neither makes a priori assumptions on the values and signs of reaction vector coefficients, nor on kinetics, or on which species are catalytic or not. A matrix $\bar{\boldsymbol{\nu}}$ is a restriction of $\boldsymbol{\nu}$ to certain rows and columns, which respectively correspond to the species and reactions of the motif under consideration.

The restriction of the rows means that the species of $\boldsymbol{\nu}$ are separated into internal species of the motif (rows of $\bar{\boldsymbol{\nu}}$) and external species (remaining rows of $\boldsymbol{\nu}$). These external species could be, in some cases,

chemostatted[26], and represent feedstock compounds, also called the food set[29], and waste from the point of view of internal species of the motif. In Fig.1, external species have been colored in blue, while stoichiometric submatrices have been boxed in yellow. Fig 1a and 1d represent examples of allocatalysis and autocatalysis, respectively, with their respective submatrices $\bar{\boldsymbol{\nu}}'$ and $\bar{\boldsymbol{\nu}}''$, and hypergraph representations Fig 1c and Fig 1f.

Restriction of columns separates reactions which are part of the motif and those which occur outside of it. A motif such that each of its reactions has at least one reactant and at least one product is called *autonomous*. This means that every column of $\bar{\boldsymbol{\nu}}$ contains a positive and a negative coefficient. Below, we pose autonomy as a condition for catalysis. Indeed, it ensures that the production of any species of the motif is conditional on the presence of other chemical species of the motif. Otherwise, rate acceleration would be allowed unconditional on an already present substance, in opposition to the definition of catalysis. Autonomy is less restrictive than former conditions for autocatalysis[24], and is similar to the siphon concept in Petri Nets[30], but without assumption on reaction signs (see SI Sec. I). Note that it does not forbid that reactions outside of the motif produce species of the motif.

Criterion for allocatalysis

By definition, allocatalysis is an ensemble of reactions by which a set of species remain conserved in number (the catalysts) while other external species undergo a turnover which changes their numbers. This leads to the following conditions:

There exists a set of species $\boldsymbol{S}$, a submatrix $\bar{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}$ restricted to $\boldsymbol{S}$, and a non-zero reaction vector $\boldsymbol{c}$ such that: i) $\bar{\boldsymbol{\nu}}$ is autonomous; ii) $\mathrm{supp}(\boldsymbol{c})$ is included in the columns of $\bar{\boldsymbol{\nu}}$; iii) $\boldsymbol{c}$ is a reaction cycle of $\bar{\boldsymbol{\nu}}$ ($\bar{\boldsymbol{\nu}} \cdot \boldsymbol{c} = \boldsymbol{0}$), and; iv) $\boldsymbol{\nu} \cdot \boldsymbol{c} \neq 0$. The members of $\boldsymbol{S}$ which participate in $\boldsymbol{c}$ (i.e. that are consumed and produced) are called allocatalysts, $\boldsymbol{c}$ an allocatalytic cycle and $\bar{\boldsymbol{\nu}}$ an allocatalytic matrix.

Condition (i) has been discussed above. Condition (ii) expresses the involvement of the catalysts in the reactions $\boldsymbol{c}$, where all columns of $\bar{\boldsymbol{\nu}}$ are non-zero due to (i), so that all reactions of $\boldsymbol{c}$ involve catalysts. Condition (iii) expresses the conservation of catalysts and (iv) the net reaction. Since the reaction cycle $\boldsymbol{c}$ is a cycle of the reduced matrix but not of the original matrix, some authors have qualified it as emergent and shown that it can establish a non-equilibrium steady state driven by the turnover of the external species[26]. Note that being allocatalytic is not a property of the sub-matrix $\bar{\boldsymbol{\nu}}$ alone but involves the larger matrix $\boldsymbol{\nu}$ as imposed by condition (iv).

Criterion for autocatalysis

By definition, autocatalysis is the process by which a combination of reactions involves a set of species which all increase in number conditional on species in the set itself (the autocatalysts), while other species undergo a turnover. This leads to the following conditions:

There exists a set of species $\boldsymbol{S}$, a submatrix $\bar{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}$ restricted to $\boldsymbol{S}$, and a reaction vector $\boldsymbol{g}$ such that: i) $\bar{\boldsymbol{\nu}}$ is autonomous, ii) all coordinates of $\Delta \boldsymbol{n} = \bar{\boldsymbol{\nu}} \cdot \boldsymbol{g}$ are strictly positive, or equivalently, $\bar{\boldsymbol{\nu}}$ has no mass-like conservation laws. The members of $\boldsymbol{S}$ consumed (and produced) by $\boldsymbol{g}$ are called autocatalysts, $\boldsymbol{g}$ an autocatalytic mode and $\bar{\boldsymbol{\nu}}$ an autocatalytic matrix.

Condition (i) ensures the conditionality of the reactions on autocatalysts, as it forbids cases where species of $\boldsymbol{S}$ are produced from external reactants only, thus playing the role of conditions (i) and (ii) in the definition of allocatalysis. Condition (ii) expresses the increase in autocatalyst number. The equivalence between the two formulations of condition (ii) is an immediate consequence of Gordan's theorem[31]. Importantly, the second formulation of (ii) does not involve an autocatalytic mode $\boldsymbol{g}$, so that (i) and (ii) can be expressed as properties of a matrix itself, in contrast with allocatalysis. This allows us to look for minimal autocatalytic motifs, which we do next. Note that external species must feed the autocatalytic system in order to guarantee the net mass increase imposed by condition (ii).

Autocatalytic cores

An autocatalytic core is an autocatalytic motif which is minimal because it does not contain any smaller autocatalytic motif. Consequently, an autocatalytic system is either a core, or it contains one or several cores. The stoichiometric conditions show that characterizing cores is equivalent to finding all autonomous matrices whose image contains vectors with only strictly positive components. This well-posed formulation allowed us to show that the stoichiometric matrix $\bar{\boldsymbol{\nu}}$ of an autocatalytic core must verify a number of non-obvious properties reported below and demonstrated in the SI Sections II andIII.

First, $\bar{\boldsymbol{\nu}}$ must be square (the number of species equals the number of reactions) and invertible. The inverse has a chemical interpretation. By definition of the inverse, the k-th column of $\bar{\boldsymbol{\nu}}^{-1}$ is a reaction vector such that species k increases by one unit, making it an elementary mode of production. Likewise, the reaction vector obtained by summing the columns of $\bar{\boldsymbol{\nu}}^{-1}$ leads to a net increase by one unit of every autocatalyst, which thus represents an elementary mode of autocatalysis. This shows how stoichiometry informs on fundamental

modes of autocatalysis[27].

Second, every forward reaction of a core involves only one core species as a reactant. While this excludes reactions between two different core species, a single core species may react with itself. As $\bar{\boldsymbol{\nu}}$ is square, this also implies that every species of a core is consumed (none is only produced), thus is an autocatalyst. Furthermore, every species is the reactant of a single reaction. Overall, every species is uniquely associated with a reaction as being its reactant, so that $\bar{\boldsymbol{\nu}}$ admits a representation with a negative diagonal and zero or positive coefficients elsewhere, at least one coefficient of each column being strictly positive to ensure autonomy.

These properties are constraining enough to allow an exhaustive enumeration of reaction graphs that are cores. Autocatalytic cores are found to belong to five categories, denoted as Type I to Type V. Fig. 2a represents typical members of each category as reaction hypergraphs (see SI Fig. S1 for general cases). As can be seen in these graphs, all minimal motifs contain a fork, which ends either in the same compound (or node) for Type I or in different compounds for Types II to V. The presence of this fork is consistent with the intuition that autocatalysis requires reaction steps that amplify the amount of autocatalysts. The orange square on the links between the nodes indicate that these links could contain further nodes and reactions in series, provided certain rules on cycles below.

The five types differ in their number of graph cycles[32] and the way these cycles overlap. Type I consists of a single graph cycle that is weight-asymmetric, defined as the product of the stoichiometric coefficients of its reaction products being different than that its reactants. Types II-V can be described as two overlapping graph cycles, where any such graph cycle involving a strict subset of the core species must be an allocatalytic cycle, i.e. weight-symmetric (it would otherwise be of Type I, contradicting minimality).

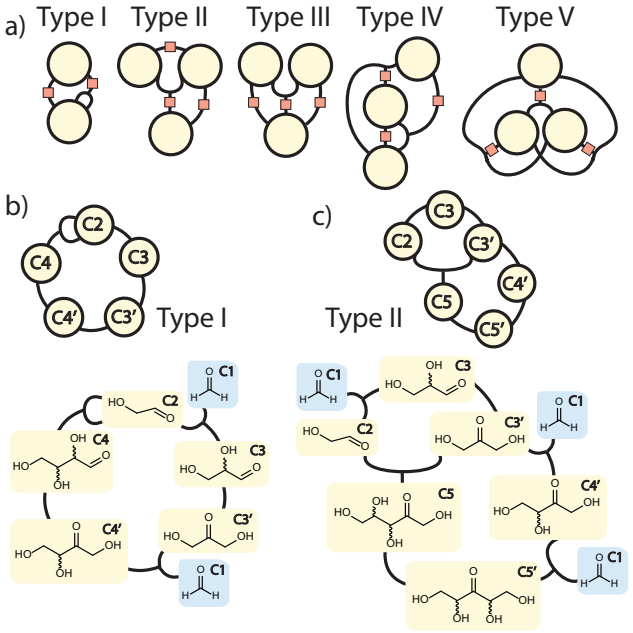

Figure 2: a) Five minimal motifs. Orange squares indicate where further nodes and reactions may be added, provided this preserves the motif type (I,II,III,IV,V) and minimality. b+c) Examples of chemical networks, along with their autocatalytic cores. blue: external species, yellow: autocatalysts. b) Type I: Breslow's 1959 mechanism for the formose reaction[33] c) Type II: Another autocatalytic cycle in the formose reaction. Species denoted as Cx inside the nodes refer to molecules containing x carbon atoms, which are shown below in standard chemical representation.

Unification of autocatalytic schemes

The stoichiometric characterization of autocatalysis provides a unified approach to autocatalytic networks reported in the literature. The examples below are further detailed in SI Sec. III. The formose reaction is a classic example of autocatalysis known to contain many autocatalytic cycles[34]. Fig.2b and c show Type I and III cores both found in the formose reaction. Similarly, autocatalytic cores of Type I and III can be found in the Calvin cycle and reverse Krebs cycle (SI Fig. S4). Some reaction steps Fig.2b may be catalyzed externally (e.g. by enzymes, base, ions), but external catalysis in general does not alter the core. By the same token, proposed examples of auto-induction introduced in[21,35] contain Type I and III cores (SI Fig. S3).

In the GARD (Graded Autocatalysis Replication Domain) model for self-enhancing growth of amphiphile assemblies[4,5], all underlying autocatalysis is described (SI Fig. S6) by Type I cycles with one fork and Type II cycles built up from sequential nonoverlapping allocatalytic cycles (cross-incorporation, such as N_3 in Fig. 3). More generally, when such catalytic cycles are compactly written as single reactions as in (1), they can be treated in the RAF (Reflexively Autocatalytic and Food-generated) framework[29], where they form irreducible RAF-sets[36]. This formally establishes the recently suggested link[5,37] between these models.

Another reported form of autocatalysis is 'chemical amplification' due to cavitands[38]. The mechanism involves a reactive compound in a molecular cage, whose free counterpart can react to form two species that exchange with the caged species, thus amplifying its release. We find that this process can be described within our framework and corresponds to a Type III core (SI Fig. S5).

Overall, previously described autocatalytic schemes comprise Types I, II and III. We have not yet found examples of Types IV and V.

VIABILITY OF AUTOCATALYTIC NETWORKS

Stoichiometric conditions do not guarantee that autocatalysts within motifs amplify. Whether an initial autocatalyst amplifies or degrades depends on kinetic considerations. To address this so-called fixation problem[17,22], we examined the probability P_{ex} of extinction (or $1-P_{ex}$ of fixation) of species within autocatalytic motifs, as a function of transition probabilities of reaction steps.

Considering a homogeneous system with a steady supply of reactants, several authors have noted that in the highly dilute autocatalyst regime, appreciable rates require first-order autocatalysis[17,22,39], i.e. each forward reaction step only involves one autocatalyst. Among first-order order networks, fixation models have so far focused on Type I networks (e.g. Fig.2b), which have a single graph cycle containing n species. In a transition step, a given species may either proceed irreversibly to the next species or disappear as a result of degradation. King found that if every reaction step k among n steps of the cycle has a success probability Π_k^+ ($1-\Pi_k^+$ being the degradation probability), fixation is possible for a doubling probability $p_2 = \prod_{k=1}^{n} \Pi_k^+ \geq 1/2$[39]. This minimum value of p_2 above which fixation is possible is called the decay threshold[19,40]. Bagley et al.[17] used birth-death processes to derive P_{ex} for an autocatalytic loop containing one species ($n = 1$). Schuster reported detailed time-dependent statistics for such networks in various contexts[22].

Here, we extend the treatment of the fixation problem so as to include reversible reactions and networks beyond Type I using the theory of branching processes[41]. In these stochastic processes, an autocatalytic species $\mathrm{X_s}$ is, after a sequence of reaction steps in the network, replaced by k copies. Reaction sequences yielding k copies happen with a probability p_k, such that

$$\mathrm{X_s} \xrightarrow{p_0} \emptyset, \quad \mathrm{X_s} \xrightarrow{p_1} \mathrm{X_s}, \quad ... \quad \mathrm{X_s} \xrightarrow{p_k} k\mathrm{X_s}, \ ... \ . \quad (4)$$

The probability P_{ex} that $\mathrm{X_s}$ goes extinct is then the probability that its k descendants independently go extinct:

$$P_{ex} = p_0 + p_1 P_{ex} + p_2 P_{ex}^2 + ... = \sum_{k=0}^{\infty} p_k P_{ex}^k. \quad (5)$$

The main difficulty here is to derive p_k from transition probabilities Π_k. A procedure for this is given in SI Sec. IV, where branching processes are constructed from reaction networks. Below, we exemplify this method by generalizing known results for Type I networks, solutions for other networks being detailed in the SI Sec. IV. We then apply it to compare the P_{ex} of autocatalytic motifs which differ in their core structures.

Reversible Type I cycles

Consider a Type I cycle consisting of n reaction steps, such as N_1 in Fig. 3b, and let us start at the first step with species $\mathrm{X_1}$ (marked node). Ultimately, $\mathrm{X_1}$ will either be successfully converted and yield $2\mathrm{X_1}$ or be degraded prematurely, which simplifies (4) to

$$\emptyset \xleftarrow{p_0} \mathrm{X_1} \xrightarrow{p_2} 2\mathrm{X_1}, \quad (6)$$

with $p_0 + p_2 = 1$. The overall outcome described by (6) corresponds to the simplest type of branching process: a birth-death process. (5) then becomes a quadratic equation that yields

$$P_{ex} = \begin{cases} \frac{1}{p_2} - 1, & p_2 \geq \frac{1}{2}, \\ 1, & p_2 < \frac{1}{2}. \end{cases} \quad (7)$$

This generalizes Bagley et al's observation for Type I networks to $n > 1$ and reversible reactions. For reversible reactions, p_2 is found by considering all possible sequences of forward and backward reactions along the cycle. From $\mathrm{X_k}$, let Π_k^- be the transition probability to revert to $\mathrm{X_{k-1}}$, and Π_k^+ to convert to $\mathrm{X_{k+1}}$. We have

$$p_2 = \prod_{k=1}^{n} \Pi_k^+ \Gamma_k, \quad (8)$$

$$\Gamma_{k+1} = \sum_{s=0}^{\infty} (\Pi_{k+1}^- \Gamma_k \Pi_k^+)^s = \frac{1}{1 - \Pi_{k+1}^- \Gamma_k \Pi_k^+}, \quad (9)$$

where Γ_k recursively ($\Gamma_1 = 1$) counts the statistical weight of all back-and-forth trajectories from $\mathrm{X_k}$ to itself, in terms of Π_k^- and Π_k^+. In the irreversible reaction limit $\Pi_k^- \to 0, \Gamma_k \to 1$ King's expression for p_2[39] is recovered.

Viability of autocatalytic cores

To investigate how autocatalytic motif structure affects survival, we calculated P_{ex} for five different cores[42] (N_1 to N_5, Fig. 3): they are of equal size (6 reaction steps, 6 species), all reactions proceed irreversibly with the same success probability ζ, which plays a similar role as the transition probability Π_k^+ in the example above and is sometimes called specificity[19,39,40].

Fig. 3 highlights how P_{ex} depends on ζ for each core structure. The highest ζ for extinction ($P_{ex} = 1$) is observed for the Type I cycle N_1, and progressively lower values are found for N_2 to N_4, which are all of Type II. Type V network N_5 tolerates the lowest specificity ζ before extinction, sustaining almost three times higher failure rates $1-\zeta$ than N_1. These differences can be qualitatively understood by counting the minimum number of steps needed to produce more autocatalysts. In respective order, networks N_1 to N_5 in Fig. 3b do so

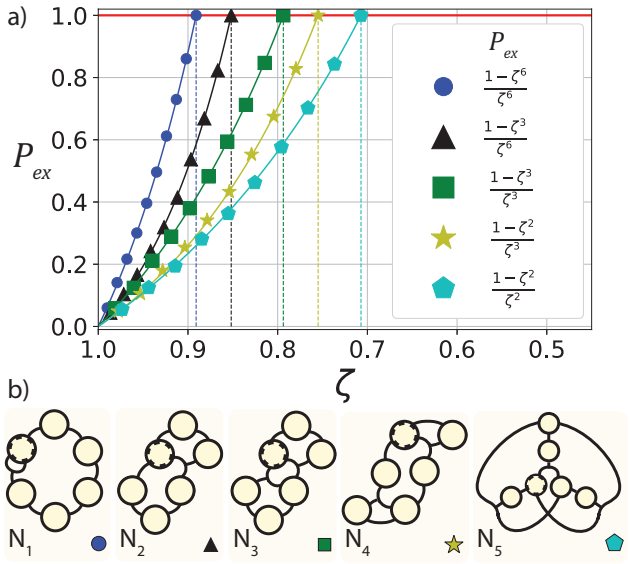

Figure 3: a) P_{ex} as function of ζ (legend: $P_{ex}(\zeta)$ for $P_{ex} < 1$) for b) 5 autocatalytic networks of similar size, starting at the dashed node. N_1: Type I cycle. N_2: Type II with one fork. N_3: Type II, two nonoverlapping allocatalytic cycles, a common motif in GARD with a 1st order RAF representation. N_4: Type II: allocatalytic cycles connected by intermediate steps. N_5: Type V. Symbols: P_{ex} after 1000 simulated trials, detailed in SI Sec. VII, lines: exact solution, derived in SI Sec. V

in six, four, three, three and two steps. In particular, given their symmetries, the P_{ex} of N_3 and N_5 have the same dependence on ζ as a 3 and 2-membered Type I cycle, respectively.

It has been suggested that large networks are disfavored in general[39]. The examples of Fig. 3 indicate that this can be counterbalanced by the presence of more allocatalytic cycles in the network. This is in particular the case for autocatalytic sets, as every net reaction must participate in an allocatalytic cycle.

EXTENSIONS: MULTICOMPARTMENT AUTOCATALYSIS

We finally show how stoichiometric criteria allow the identification of autocatalysis that emerges from compartments coupled via selective exchange, as found in systems comprising vesicles, pores, emulsions and complex coacervates[43]. The reaction network in Fig. 4a is incapable of autocatalysis as it does not contain any autocatalytic core. However, when we place this network in two compartments α and β coupled by a membrane permeable to A and A_2B, a Type II core emerges (Fig. 4b).

The core identification indicates a possible setting for autocatalysis: U and V are chemostatted in α,

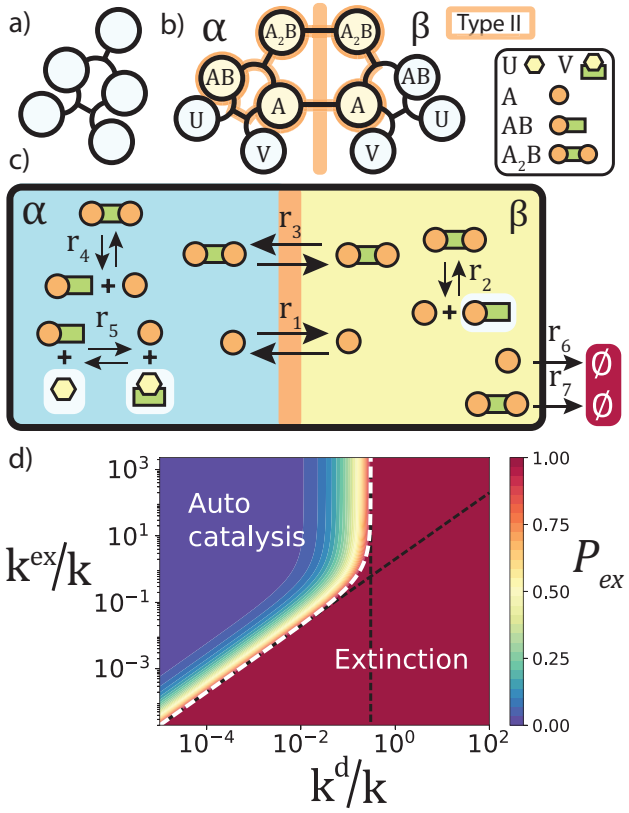

Figure 4: Multicompartment autocatalysis. a) Reaction network with two reactions and five species in a single compartment. The network does not contain any autocatalytic core, thus cannot perform autocatalysis. b) Same reaction network as in (a), but duplicated in two compartments α and β, coupled by the selective exchange of species A and A_2B. A Type II core, highlighted in orange, emerges. c) Open reactor with two compartments, a semi-permeable membrane, degradation and exchange reactions. Chemostatted species have a lighter background: U and V in α and AB in β. d) Extinction probability P_{ex} for multicompartment autocatalysis in (c), starting from a single A_α, as a function of exchange rate k^{ex} and degradation rate k^d, relative to other relevant reaction rates fixed at k. Slanting asymptote: exchange-limited survival $k_{ex} = 2k^d$. Vertical asymptote: reaction-limited survival $k^d/k = \frac{\sqrt{13}-3}{2}$. Dashed white line: transition between extinction and potential fixation ($P_{ex} < 1$). Expressions for P_{ex} and asymptotes are derived in SI Sec. VI.

and AB is chemostatted in β (Fig. 4c). The reaction involving U and V may in principle also take place in β, but it is not required for autocatalysis as it is not part of the type II core. In the present example, U and V are absent in β.

We now apply our viability analysis to this autocatalytic network in the presence of degradation reactions (r_6 and r_7 in Fig. 4c). Let us introduce a characteristic rate k for reactions r_2, r_4 and r_5, a degradation rate

k^d for reactions r_6 and r_7 and an exchange rate k^{ex} for reactions r_1 and r_3. Figure 4d shows the extinction probability as a function of the degradation rate k^d and exchange rate k^{ex}, both normalized by k. To overcome the degradation threshold ($P_{ex} < 1$), the ratio k/k^d must lie above a certain threshold (vertical black dotted line in Fig. 4d), and the rate of exchange k^{ex} should outpace the rate of degradation (black slanting dotted line in Fig. 4d). Stochastic simulations (SI Sec. VII and Fig. S9) confirm autocatalytic growth in these conditions.

In this example of coupled compartments, compounds are no longer restricted to one role: AB is an autocatalyst in α and a feedstock in β. Chemical reactions are no longer restricted to one direction: The reaction used for reproduction in α is reused in β to provide the missing step to close the cycle. Such multicompartment autocatalysis is however more general. For instance, a single reaction $\mathrm{A} \rightleftharpoons \mathrm{B} + \mathrm{C}$ can give rise to Type III motifs, given three compartments coupled by selective exchange as detailed in SI Sec. VIII and Fig. S9.

DISCUSSION

We presented a theoretical framework for autocatalysis based on stoichiometry, which allows a precise identification of the different forms of autocatalysis. Starting with a large stoichiometric matrix, we provide criteria for reaction network motifs that allow allocatalysis and autocatalysis. A detailed analysis of the graph structure contained in these reduced stoichiometric matrices reveals that they contain only five possible recurrent motifs, which are minimal in the sense that they do not contain smaller motifs. Fundamental modes of production of minimal autocatalytic cores are encoded in the column vectors of the inverse of the autocatalytic core submatrix. Autocatalytic cores are found to have a single reactant species for each reaction. This means that autocatalytic networks require the availability of certain chemical species in their cores to operate properly, but also implies that the proper functioning of an autocatalytic network will guarantee the stable supply of certain products, a definitive advantage when these products are key enzymes or metabolites.

We identified these minimal motifs in known examples of autocatalysis such as the formose reaction, central metabolic cycles, the GARD model and RAF sets. Autocatalytic cores also provide a basis for algorithms to identify these recurring autocatalytic motifs in large chemical networks[44,45], as has been done for gene regulatory networks[46]. In this way, we may be able to break the complexity of large chemical networks into smaller, more manageable structures[47]. Additionally, autocatalytic cores are the building block of evolution in prebiotic chemistries[36], thus their identification paves to the way of a systematic exploration of the possible modes chemical evolution[48].

Autocatalytic motifs provide different degrees of robustness, which we evaluated using the notion of viability. Viability can be computed as a survival probability in an appropriately defined branching process. This approach is generally applicable to autocatalytic models upon identification of their cores, highlighting the interest of a unified framework. Viability results from a competition between reactions that produce autocatalysts and side-reactions such as degradation. This is intimately related to the 'paradox of specificity'[19,40]: autocatalytic motifs are more likely to be found in large networks with many different chemical components engaging in many different reactions, but putting many components together favors side-reactions, leading to extinction.

Multicompartment autocatalysis introduced here offers a way around this problem: coupled compartments effectively enlarge the number of species without requiring new reactions. In multicompartment autocatalysis, cycles rely on the environmental coupling of reaction networks, which allows access to conditions unattainable in a single compartment. In this way, autocatalysis can emerge from reaction schemes as simple as a bimolecular reaction, provided certain semi-permeability conditions are met for the exchange of compounds between compartments. In the example shown here (Fig. 4), this allowed us to reuse the compounds and reactions to complete autocatalytic cycles. The principle is more general, however: autocatalysis may also emerge from coupling phases with physical-chemical conditions conducive to different reactions, as observed in liquid-solid[49], solid-gas[50] interfaces. Liquid-liquid interfaces in cellular organization and multiphase coacervates[43] are promising places to further explore such principles.

Overall, our framework shows that autocatalysis comes in a diversity of forms and can emerge in unexpected ways, indicating that autocatalysis in chemistry must be more widespread than previously thought. This invites to search for further extensions of autocatalysis, which provides new vistas for understanding how chemistry may complexify towards life[51].

MATERIALS AND METHODS

Theoretical methods and derivation of results are detailed in the Supplementary Appendix comprising the following sections: 1) Terminology and definitions, 2) derivation of autocatalytic cores from Graph theory, 3) their chemical interpretation and 4) application to formose, autoinduction, metabolic cycles, chemical amplification, RAF sets, GARD 5) branching process derivation and determination of P_{ex}. 6) determination of P_{ex} for Fig. 3. 7) determination of P_{ex} for Fig. 4d.

8) stochastic simulations. 9) autocatalysis from one bimolecular reaction and 3 compartments.

ACKNOWLEDGEMENTS

AB acknowledges stimulating discussions with D. van de Weem and Z. Zeravcic. AB and DL acknowledge support from Agence Nationale de la Recherche (ANR-10-IDEX-0001-02) and PSL IRIS OCAV). PN acknowledges s. Krishna and C. Flamm for discussions. AB, DL and PN acknowledge Andrew Griffiths for careful reading of the manuscript.

REFERENCES

[1] W. Hordijk, BioScience **63**, 877 (2013).
[2] A. I. Oparin, *Origin of Life* (Dover, 1952).
[3] S. A. Kauffman, J. Theor. Biol. **119**, 1 (1986).
[4] D. Segré, D. Ben-eli, D. W. Deamer, and D. Lancet, Orig. Life Evol. Biosph. **31**, 119 (2001).
[5] D. Lancet, R. Zidovetzki, and O. Markovitch, J. R. Soc. Interface **15**, 20180159 (2018).
[6] L. E. Orgel, Plos Biol. **6**, 5 (2008).
[7] A. Bissette, B. Odell, and S. Fletcher, Nat Commun **5**, 4607 (2014).
[8] S. N. Semenov, L. J. Kraft, A. Ainla, M. Zhao, M. Baghbanzadeh, V. E. Campbell, K. Kang, J. M. Fox, and G. M. Whitesides, Nature **537**, 656 (2016).
[9] H. N. Miras, C. Mathis, W. Xuan, D.-L. Long, R. Pow, and L. Cronin, Proc. Natl. Acad. Sci. U. S. A. (2020), 10.1073/pnas.1921536117.
[10] B. Liu, C. G. Pappas, E. Zangrando, N. Demitri, P. J. Chmielewski, and S. Otto, J. Am. Chem. Soc. **141**, 1685 (2019).
[11] M. Altay, Y. Altay, and S. Otto, Angew. Chem. Int. Ed. **57**, 10564 (2018).
[12] C. Viedma, Phys. Rev. Lett. **94**, 065504 (2005).
[13] I. Baglai, M. Leeman, M. Kellogg, and W. Noorduin, Org. Biomol. Chem. **17**, 35 (2019).
[14] K. Ichimura, K. Arimitsu, and K. Kudo, Chemistry Lett. **24**, 551 (1995).
[15] F. Schlögl, Z. Phys. **253**, 147 (1972).
[16] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (Wiley, New York, 1977).
[17] R. J. Bagley, D. J. Farmer, and W. Fontana, in *Artif. Life II* (1991) pp. 141–158.
[18] D. Segré, D. Lancet, O. Kedem, and Y. Pilpel, Orig. Life Evol. Biosph. **28**, 501 (1998).
[19] E. Szathmáry, Philos. Trans. R. Soc. Lond. B. Biol. Sci. **361**, 1761 (2006).
[20] W. Hordijk and M. Steel, Journal of Theoretical Biology **227**, 451 (2004).
[21] D. G. Blackmond, Angew. Chem. Int. Ed. **48**, 386 (2009).
[22] P. Schuster, Monatsh. Chem. **150**, 763 (2019).
[23] M. Feinberg, *Foundations of Chemical Reaction Network Theory* (Springer, New York, 2019).
[24] U. Barenholz, D. Davidi, E. Reznik, Y. Bar-on, N. Antonovsky, E. Noor, and R. Milo, eLife **6**, e20667 (2017).
[25] A. Deshpande, M. Gopalkrishnan, and C. Science, Bull. Math. Biol. **76**, 2570 (2014).
[26] M. Polettini and M. Esposito, J. Chem. Phys. **141**, 024117 (2014).
[27] S. Schuster, C. Hilgetag, J. H. Woods, and D. A. Fell, "Elementary modes of functioning in biochemical replicators," in *Computation in Cellular and Molecular Biological Systems* (Singapore: World Scientific, 1996) pp. 151–165.
[28] A. D. McNaught and A. Wilkinson, *IUPAC Compendium of Chemical Terminology, 2nd ed. ("the Gold Book")* (Blackwell Scientific Publications, Oxford, 1997).
[29] W. Hordijk, J. Theor. Biol. **435**, 22 (2017).
[30] M. Gopalkrishnan, Bull. Math. Biol. **73**, 2962 (2011).
[31] Borwein, Jonathan and Lewis, Adrian S., *Convex Analysis and Nonlinear Optimization* (Springer-Verlag New York, 2006).
[32] *Graph cycles* are closed paths in the reaction hypergraph and should not be confounded with *reaction cycles* which are right null vectors of the stoichiometric matrix.
[33] R. Breslow, Tetrahedron Lett. **21**, 22 (1959).
[34] A. Ricardo, F. Frye, M. A. Carrigan, J. D. Tipton, D. H. Powell, and S. A. Benner, J. Org. Chem. **71**, 9503 (2006).
[35] A. J. P. Teunissen, T. F. E. Paffen, I. A. W. Filot, M. D. Lanting, R. J. C. van der Haas, T. F. A. de Greef, and E. W. Meijer, Chem. Sci. **10**, 9115 (2019).
[36] V. Vasas, C. Fernando, M. Santos, S. Kauffman, and E. Szathmáry, Biol. Direct **7**, 1 (2012).
[37] S. A. Kauffman, *A world beyond physics: the emergence & evolution of life* (Oxford University Press, 2019).
[38] J. Chen, S. Körner, S. L. Craig, S. Lin, D. M. Rudkevich, and J. Rebek, Proc. Natl. Acad. Sci. U. S. A. **99**, 2593 (2002).
[39] G. King, BioSystems **15**, 89 (1982).
[40] E. Szathmáry, Philos. Trans. R. Soc. B Biol. Sci. **355**, 1669 (2000).
[41] S. Karlin and H. M. Taylor, *A first course in stochastic processes* (Academic Press, New York, 1975).
[42] Note that in general, the calculation of P_{ex} may involve reactions that are not in the core. Here, we consider cases where this is not necessary.
[43] T. Lu and E. Spruijt, Journal of the American Chemical Society **142**, 2905 (2020).
[44] á. Kun, B. Papp, and E. Szathmáry, Genome Biol **9**, R51 (2008).
[45] Andersen, J. L. and Flamm, C. and Merkle, D. and Stadler, P. F., IEEE/ACM Trans. Comput. Biol. Bioinform **16**, 510 (2017).
[46] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[47] P. Dittrich and P. Speroni di Fenizio, Bull. Math. Biol. **69**, 1199 (2007).
[48] Z. Peng, A. Plum, P. Gagrani, and D. A. Baum, arXiv (2020), arXiv:2001.02533, arXiv:arXiv:2001.02533.
[49] Halpern, J, J. Electrochem. Soc. **100**, 421 (1953).
[50] P. Grosfils, P. Gaspard, and T. V. de Bocarmé, The Journal of Chemical Physics **143**, 064705 (2015).
[51] R. Krishnamurthy, Chem. Eur. J. **24**, 16708 (2018).
[52] K. Andersson, G. Ketteler, H. Bluhm, S. Yamamoto, H. Ogasawara, L. G. M. Pettersson, M. Salmeron, and A. Nilsson, J. Am. Chem. Soc. **130**, 2793 (2008).
[53] D. Angeli, P. D. Leenheer, and E. D. Sontag, Math. Biosci. **210**, 598 (2007).
[54] J. Chen, S. Körner, S. L. Craig, D. M. Rudkevich, and J. Rebek, Nature **415**, 385 (2002).
[55] G. King, J. Theor. Biol. **123**, 493 (1986).

SUPPLEMENTARY INFORMATION

I. TERMINOLOGY

A. IUPAC Definitions

To promote the consistent use of terminology, IUPAC committees establish recommendations which serve as a basis for the Compendium of Chemical Terminology (*The Gold Book*), of which some relevant entries are reproduced here. Comments in italic have been added for clarity.

Chemical Reaction: A process that results in the interconversion of chemical species. Chemical reactions may be elementary reactions or stepwise reactions. (It should be noted that this definition includes experimentally observable interconversions of conformers.) Detectable chemical reactions normally involve sets of molecular entities as indicated by this definition, but it is often conceptually convenient to use the term also for changes involving single molecular entities (i.e. 'microscopic chemical events').

Catalyst: A substance that increases the rate of a reaction without modifying the overall standard Gibbs energy change in the (*net*) reaction; the process is called catalysis. The catalyst is both a reactant and product of the (*catalyzed*) reaction. The words catalyst and catalysis should not be used when the added substance reduces the rate of reaction (see inhibitor). Catalysis can be classified as homogeneous catalysis, in which only one phase is involved, and heterogeneous catalysis, in which the reaction occurs at or near an interface between phases. Catalysis brought about by one of the products of a (*net*) reaction is called autocatalysis. Catalysis brought about by a group on a reactant molecule itself is called intramolecular catalysis. The term catalysis is also often used when the substance is consumed in the (*net*) reaction (for example: base-catalysed hydrolysis of esters). Strictly, such a substance should be called an activator.

Autocatalytic Reaction: A (*net*) chemical reaction in which a product (or a reaction intermediate) also functions as a catalyst. In such a reaction the observed rate of reaction is often found to increase with time from its initial value.

B. Allocatalysis

We refer to allocatalysis as the form of catalysis in which, at the end of a catalytic cycle, the catalyst(s) have not changed in number. By their equal participation in either direction, allocatalysts will thus drop out of the net reaction. Some authors refer to autocatalysis as homocatalysis and allocatalysis as heterocatalysis, which is a lexicologically consistent choice of terms that express an opposition (same vs different). This opposition between same and different is e.g. found in the IUPAC terminology for a homogeneous catalysis (occurring in the same phase) and heterogeneous catalysis. For the IUPAC recommended terminology 'autocatalysis', a consistent choice that expresses this opposition is 'allocatalysis' (self vs other).

C. Remarks

1. *Directionality of autocatalysis*

In an allocatalytic reaction, catalysts are produced and consumed in equal amounts, and the term does not distinguish between the direction in which the reaction proceeds (catalysis equally increases the rates of 'forward' and 'backward' reactions by the thermodynamic criterion). In contrast, the term autocatalysis only applies in one direction, due to the requirement of having catalysts be the product of the net reaction. Consequently, in writing the simplest reaction balance

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \rightleftharpoons 2\mathrm{B}, \tag{10}$$

the definition applies when the net reaction $\mathrm{A} \rightleftharpoons \mathrm{B}$ exhibits acceleration, owing to B being a catalyst and product of the net reaction. In the opposite sense, $\mathrm{B} \rightleftharpoons \mathrm{A}$, B is no longer a product, but it can still accelerate the reaction. This case is typically referred to as 'reverse autocatalysis'.

A simple experimental example is the catalytic disproportionation of water adsorbed on a copper surface $\mathrm{H_2O}(ads)$[52].

$$2\mathrm{H_2O}(ads) \overset{1}{\rightleftharpoons} \mathrm{H_2O \cdot OH}(ads) + \mathrm{H}(ads), \tag{11}$$

$$\mathrm{H_2O \cdot OH}(ads) \overset{2}{\rightleftharpoons} \mathrm{H_2O}(ads) + \mathrm{OH}(ads). \tag{12}$$

Performing r_1 and r_2 from left to right, the catalyst $\mathrm{H_2O}$ is consumed, yielding reverse autocatalysis for the net reaction:

$$\mathrm{H_2O}(ads) \rightleftharpoons \mathrm{H}(ads) + \mathrm{OH}(ads). \tag{13}$$

In the opposite sense, the stoichiometric conditions for autocatalysis are obtained:

$$\mathrm{H}(ads) + \mathrm{OH}(ads) + \mathrm{H_2O}(ads) \rightleftharpoons 2\mathrm{H_2O}(ads). \tag{14}$$

2. *Inhibition*

A compound is a catalyst in the context of a particular experimental condition where it accelerates the rate of a reaction. A change of those conditions may change this label, e.g. the compound may become an

inhibitor instead. Consider

$$\mathrm{A} + \mathrm{B} \overset{1}{\rightleftharpoons} \mathrm{AB}, \tag{15}$$

$$\mathrm{A} + \mathrm{E} \overset{2}{\rightleftharpoons} \mathrm{AE}, \tag{16}$$

$$\mathrm{AE} + \mathrm{B} \overset{3}{\rightleftharpoons} \mathrm{AEB}, \tag{17}$$

$$\mathrm{AEB} \overset{4}{\rightleftharpoons} \mathrm{E} + \mathrm{AB}. \tag{18}$$

For a rate acceleration to occur for the net reaction $\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}$, steps 2,3,4 must together proceed faster than reaction 1. Template-assisted ligation of DNA or RNA provides an example where this may not be the case: below a critical annealing temperature the detachment of the product (here, AB in step 4) becomes increasingly slower for longer strands, and the template becomes an inhibitor.

3. Autonomy and siphons

The concept of autonomy is closely related to the concept a siphon in Chemical Reaction Networks (CRN) theory[53]: *A Siphon Σ is a subset $\Sigma \subset \mathcal{S}$ of all species S, which for each reaction that has a species in Σ as a product, has at least one of its reactants in Σ.*

In this definition, a reaction can be irreversible in a mathematical sense: the reverse reaction does not exist. For a reversible CRN, a reverse reaction does exist, and the siphon definition must apply to the forward and backward direction, thus becomes equivalent to autonomy. A reaction must then imply both one or more products and one or more reactants from Σ (siphon reactions), or none at all (external reactions). Note that autonomy is less restrictive than the conditions posed in Barenholz et al.[24], where in addition species must be both reactant and product of a reaction. This last conditions is a proved as a consequence for minimal structures in our choice of formalism (see below).

II. MATHEMATICAL DERIVATION OF AUTOCATALYTIC CORES

A. Reaction graph definitions

Reaction graphs described below correspond to weighted directed hypergraphs without self-loops in the language of graph theory. In this section, the word *cycle* is used in the sense of graph theory (see below), not in the sense of *reaction cycle* used to denote right null-vectors of the stoichiometric matrix. In the following, letters used for scalars indicate positive numbers, and cycles and paths are understood as directed.

Definition 1. *A **reaction graph** $\mathcal{H}$ is a triplet (S, R, M) where $\{s_1, ..., s_n\}$ is the species set, $R = \{r_1, ..., r_m\}$ is reaction set, each r_j being an ordered pair (Xj, Yj) of non-empty and non-intersecting subsets of S respectively called reactants and products, and M is the stoichiometric matrix with coefficients m_{ij}. $m_{ij} = 0$ when the species s_i does not participate to the reaction r_j, $m_{ij} < 0$ when species s_i is a reactant of r_j, and $m_{ij} > 0$ when species s_i is a product of r_j.*

The stoichiometric matrix M contains all the information about the hypergraph, the column of M corresponding to reactions, and the rows to species.

Definition 2. *A **subgraph** of $\mathcal{H} = (S, R, M)$ is a triplet $\mathcal{H}' = (S', R', M')$ where S' is a subset of S, R' is a set of reactions which reactants and products are in S' and intersect the reactant set and product set of a reaction in R, with corresponding stoichiometric coefficients M'.*

Definition 3. *A reaction graph is **square** if it has the same number of reactions and species.*

Definition 4. *A **directed path** is a sequence of alternating species and reactions, all reactions and species being distinct, where species which precede and succeed a reaction are respectively a reactant and a product of it. A path is a **minimal path** if it is not possible to form a path starting and ending at the same species using a strict subset of its reactions. A path is **semi-open** if it either starts or ends with an edge.*

Definition 5. *In a directed path, an edge has a **back-branch** if one of its products is a species located upstream in the path, and this product is called a back-product. An edge has a **forward-branch** if one of its reactants is a species located downstream in the path, and this product is called a forward-reactant.*

Definition 6. *A **cycle** has an identical definition as a path, except that the first and last species are the same species. A cycle is minimal if it is not possible to form a cycle with a subset of its species and reactions.*

Definition 7. *A species S is the **solitary** reactant (product) of a reaction if S is the only reactant (resp. product) of this reaction, otherwise it is a **co-reactant** (resp. co-product).*

Definition 8. *A reaction is **simple** if has a single reactant and a single product.*

Definition 9. *Consider a simple reaction R with reactant x and product y, with respective stoichiometries $-a$ and b. The **contraction** of R consists of: (i) removing R; (ii) merging x and y into a single species z, and; (iii) multiplying by a (resp. b) the stoichiometric coefficients associated with z for all reactions formerly associated with x (resp. y).*

Definition 10. *In a square graph, a **perfect matching** is a bijection between species and reactions. It corresponds to all pairs $(i, \sigma(i))_{i=1...N}$, where i is the*

index of a species, σ is a permutation of $[1...N]$, and $\sigma(i)$ is the index of a reaction. A perfect matching of a graph, or a subgraph, $\mathcal{G}$ is called a $\mathcal{G}$-matching.

Remarks:

- Consider reactions $E = (\{a\}, \{b, c\})$ and $F = (\{b\}, \{c\})$; $a - E - b - F - c$ is a path, but it is not minimal because it contains $a - E - c$. Simple paths are necessarily minimal (Fig. S1a).
- Consider $E = (\{a\}, \{b\})$ and $F = (\{b\}, \{a, c\})$; $a - E - b - F - c$ is a minimal path and F has a back-branch (Fig. S1a). In particular, it contains a cycle $a - E - F - a$.
- Simple paths are exactly cycle-free minimal paths.
- More generally, a minimal path can have back-products and forward-reactants (Fig. S1a).
- A hypergraph cycle can have reactions connecting several of its species, but a minimal cycle cannot. Thus, a minimal cycle only contains simple paths (it is identical to a cycle in a regular graph).
- Cycles are square.

B. Relationship with linear algebra

$x \succ 0$ denotes a real vector with only strictly positive coordinates.

Definition 11. *A matrix M is **productive** if there exists a real vector γ such that $M.\ \gamma \succ 0$. Equivalently, M intersects the strictly positive orthant $\mathbb{R}^{n}_{>0}$.*

Definition 12. *A matrix is **autonomous** if its columns all contain a strictly negative and a strictly positive coefficient.*

Definition 13. *A minimal cycle is **weight-symmetric** if the product of the absolute values of the stoichiometric coefficients associated with its reactants equals the product of the stoichiometric coefficients associated with its products. Otherwise, the cycle is weight-asymmetric.*

Remarks:

- Autonomous matrices are stoichiometric matrices of reactions systems such that every reaction has at least one reactant and at least one product.
- The Leibniz formula for the determinant shows that the non-zero terms of $det(M)$ exactly correspond to the products of stoichiometric coefficients $m_{i,\sigma(i)}$ determined by each possible perfect matching of the graph.
- A minimal cycle $\mathcal{C}$ has exactly two perfect matchings, which correspond to the matching of its reactions with their solitary reactants and solitary products respectively. $det(\mathcal{C}) = 0$ if and only if $\mathcal{C}$ is weight-symmetric.
- Be $\mathcal{H}'$ the hypergraph obtained from $\mathcal{H}$ by contracting a simple reaction R. The cofactor expansion implies that $|det(\mathcal{H}')| = |det(\mathcal{H})|$. Additionally, any cycle $\mathcal{C}$ of $\mathcal{H}$ becomes a cycle $\mathcal{C}'$ in $\mathcal{H}'$ obtained by contracting R, and $|det(\mathcal{C}')| = |det(\mathcal{C})|$.

C. Autocatalytic cores

Definition 14. *A **core** is a minimal productive reaction graph, i.e. a reaction graph that does not contain any productive subgraph.*

Finding cores is equivalent to finding minimal matrices which are productive and autonomous. In this section, we denote $M \in \mathbb{R}^{n,m}$ the stoichiometric matrix of the graph $\mathcal{H}$, where n is the number of rows (species) and m the number of columns (reactions). If M is productive, we denote γ a vector such that $M.\gamma \succ 0$. Productivity of M is indifferent to the sign of its columns as the sign of the coefficients of γ is not constrained. Therefore, we choose the convention that any productive vector γ is positive, up to taking the opposite for some columns of M.

Remarks:

- M invertible implies M productive, as the image of M is then the full space $\mathbb{R}^{n,m}$, which contains the strictly positive orthant.
- Below, species or reactions 'can be removed' is understood as 'can be removed while preserving productivity and autonomy'. Being able to remove a row (a species) or a column (a reaction) contradicts the minimality of a core.
- Removing columns (reactions) preserves autonomy, but not necessarily productivity.
- Removing rows (species) preserves productivity, but not necessarily autonomy.
- A row corresponding to a species that is always a co-reactant or a co-product can be removed without affecting autonomy (every column still contains positive and negative coefficients) and productivity.

Proposition 1. *In a core, every species is both a reactant and a product.*

Proof. Every species must be produced, otherwise it would not be possible to find a positive γ verifying $M.\gamma \succ 0$. Now, suppose a species S is never a reactant

and only a product. All reactions such that S is their only product can be removed without affecting the productivity of other species. In the resulting graph, either S is not the product of any reaction anymore, or S is only a co-product of the remaining reactions. In both cases, S can be removed without affecting autonomy. □

Proposition 2. *A core is square, invertible, every species is the solitary reactant of a reaction, and is reactant for this reaction only.*

Proof. Consider $M \in \mathbb{R}^{n,m}$ a productive stoichiometric matrix with n species and m reactions, with rank k. Obviously $k \leq m, n$. We must also have $m = k$, otherwise a column could be removed while preserving the image of M, thus productivity. Additionally, every species must be the solitary reactant of at least one reaction, otherwise it could be removed. This implies $m \geq n$. Overall, $k = m \geq n \geq k$, so that $k = m = n$, meaning that M is square and invertible. As every species S is the solitary reactant of at least one reaction R and $m = n$, the species is a reactant for only R. □

Remark: Property 2 implies that M can be rearranged in such a way that it has a strictly negative diagonal, and only non-negative off-diagonal coefficients.

Proposition 3. *In a core, a square autonomous submatrix must have a product outside its set of species.*

Proof. Be A a square autonomous submatrix of M and write $M = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$. We need to show that $B > 0$. As A is autonomous, every A-reaction has a reactant in the set of A-species. By Property 2, reactants are always solitary, thus $B \geq 0$. Now suppose $B = 0$. Then $det(M) = det(A).det(D)$. Either $det(A) = 0$, implying $det(M) = 0$, contradicting M invertible, or $det(A) \neq 0$, then A is productive, contradicting the minimality of M. □

Proposition 4. *A core is strongly connected.*

Proof. Consider a species x_0 of M. Below, we recursively construct sets $D_k = \{x_1, ..., x_k\}$ of increasing cardinal, such that every x_i is downstream x_0, until $k = n - 1$, implying that for any species $y \neq x$, there exists a path from x to y. We denote $R(S)$ the only reaction with reactant S, which is well defined by Property 2. **Step 1**: We take x_1 as a product of $R(x_0)$. **Step k**: Suppose D_k exists, $k < n - 1$. Re-arrange M such the top left block A of size k corresponds to species set D_k and reaction set $R(D_k)$. As A is autonomous, by Property 3, there exists a species x_{k+1} outside of D_k which is downstream D_k, hence D_{k+1} exists. □

Proposition 5. *In a core, every species is involved in a cycle.*

Proof. Obvious from Property 4, considering the back and forth paths joining any two species. □

Proposition 6. *Consider a partition of a core into two species sets V and W. V cannot be upstream of W, i.e. reactions with products in W cannot have all their reactants in V.*

Proof. Suppose on the contrary that M can be written $M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ where A and B span V, C spans W, and $B \leq 0$. Given the latter inequality, Property 1 implies that A is non-empty and autonomous. Consider $\gamma > 0$ such that $M.\gamma \succ 0$. Be α and β the respective restrictions of γ to the reaction spaces of A and B. Then $A.\alpha + B.\beta \succ 0$. As $B.\beta \leq 0$, we have $A.\alpha \succ 0$, contradicting the minimality of M. □

The definitions below are generalizations of the notion of ear decomposition in regular graphs (Fig. S1b).

Definition 15. *A **hyper-ear** is a hypergraph comprising a minimal cycle $\mathcal{C}$, called the base cycle, such that $\mathcal{C}$ has a reaction with a product x outside $\mathcal{C}$, and a minimal path $\mathcal{P}$ starting at x, such that its last reaction, R, has a product in $\mathcal{C}$, R being the only $\mathcal{P}$-reaction to have a product in $\mathcal{C}$. A **proto-ear** has a similar definition, but where $\mathcal{C}$ is a simple path (called the base path) instead of a minimal cycle.*

Description of proto-ears and hyperears - In a hyper-ear or a proto-ear, any $\mathcal{C}$-reaction can have x as a product, and any $\mathcal{C}$-species can be the product of R, the last reaction of $\mathcal{P}$ (Fig. S1b). We denote by u a species which is the product of R, v any $\mathcal{C}$-species which is the reactant of a $\mathcal{C}$-reaction which produces x, '$-$' a simple path (including the empty path), o a simple path comprising at least one species, o being a subcase of the '$-$' category. By convention, a uv motif can correspond to a single same species which is both a R-product and a reactant for x production. Any proto-ear falls into a class described by a chain made of the symbols u, v, o, and '$-$', as soon as it contains at least a u and a v. Reciprocally, any such chain represents one or more proto-ears. Any hyper-ear is obtained by cyclic closure of the base path of a proto-ear. Such closure is denoted by the $'*'$ symbol at the beginning and the end of the chain.

Theorem 1. *Cores are of one of the following types:*

- *TYPE I: a weight-asymmetric minimal cycle;*
- *TYPE II: a cycle comprising all species as solitary reactants, and one or more weight-symmetric sub-cycles without intersection between them;*

- *TYPES III-IV-V: one of the three types of hyper-ears, any subcycle of which is weight-symmetric: (III) $*u-vo*$; (IV) $*u-u-v*$; (V) $*u-v-u-v*$.*

Proof. Core type are represented in Fig. S1c. $\mathcal{H}$ denotes the reaction graph and M its stoichiometric matrix.

SUFFICIENCY

Step 1: All types have a non-zero determinant.

TYPE I: Invertibility is a direct consequence of the remark on the determinant of minimal cycles.

TYPE II: Consider a minimal weight-symmetric subcycle $\mathcal{C}$ of $\mathcal{H}$. Any $\mathcal{H}$-matching consistent with a $\mathcal{C}$-matching can be associated to another $\mathcal{H}$-matching obtained by reversing the $\mathcal{C}$-matching. Thus, $det(\mathcal{H})$ has the form $det(\mathcal{C}).\alpha + \beta$, where the first term comprises all $\mathcal{C}$-consistent $\mathcal{H}$-matchings. As $det(\mathcal{C}) = 0$, it suffices to show that β is a single non-zero term. By Property 3, $\mathcal{C}$ must have a reaction R_1 with a product x_1 outside of $\mathcal{C}$. As by definition of TYPE II, subcycles are non-intersecting, R_1 must be the only $\mathcal{C}$-reaction with a product outside $\mathcal{C}$, and x_1 must be the only product of R_1 outside $\mathcal{C}$. All non-zero terms in β require matching R_1 to x_1, otherwise R_1 would be matched with a $\mathcal{C}$-species, which would impose a $\mathcal{C}$-matching ($\mathcal{C}$ being a minimal cycle), contradicting the definition of β. We now order the indexes k following the downstream order of the x_k species along $\mathcal{H}$, and show recursively that β corresponds to the $\mathcal{H}$-matching where every R_k matches x_k: (step 1) By construction, R_1 matches x_1. (step k) Suppose R_{k-1} matches x_{k-1}. Either R_k is a simple reaction, and, as its reactant x_{k-1} is already matched, R_k necessarily matches its only product x_k. Or R_k has a back-branch forming a minimal cycle. However, cycles are non-intersecting, so that the back-product x_i of R_k is necessarily downstream x_k and upstream x_{k-1}, thus x_i is already matched by the recursion hypothesis. Thus, R_k can only match x_k.

TYPES III-IV-V: Consider the stoichiometric matrix M with non-negative off-diagonal coefficients:

$$M = \begin{pmatrix} -1 & c & e \\ a & -1 & f \\ b & d & -1 \end{pmatrix} \tag{19}$$

We have $det(M) = -1 + df + ac + ade + bcf + be$. Strict subcycles must be weight-symmetric, as otherwise, the minimality of M would be contradicted. Consequently, when their factors are both non-zero, the products df, ac, and be must be equal to 1. By contracting all simple reactions and multiplying the columns of M if necessary so that solitary reactant stoichiometries are all normalized to -1, TYPE III corresponds to $d = f = 0$ and $a, b, c, e > 0$, TYPE IV to $f = 0$ and $a, b, c, d, e > 0$, and TYPE IV to all coefficients strictly positive. In all these cases, we have $det(M) > -1 + ac + be = 1$.

Step 2: All types are minimal

TYPE I: Removing any subset of species or reactions would result in an acyclic graph, contradicting Property 5.

TYPE II: Removing any subset of species or reactions either leads to a hypergraph where a subset of species is upstream the rest, contradicting Property 6, or to a non-invertible minimal cycle.

TYPE III-V: Removal of any set of reactions (and a fortiori of species, given that they all are solitary reactants) leaves at most one cycle in the graph, the latter being non-invertible.

NECESSITY

By Property 5, there exists a minimal cycle $\mathcal{C}$ in $\mathcal{H}$. Either $\mathcal{C}$ is weight-asymmetric, then $\mathcal{H} = \mathcal{C}$ is of TYPE I. Or $\mathcal{C}$ is weight-symmetric. Then, by Property 3, there is a $\mathcal{C}$-reaction with a product x outside $\mathcal{C}$. By Property 4, there exists a path $\mathcal{P}$ from x to any species in $\mathcal{C}$ (Fig. S1d). We can take $\mathcal{P}$ minimal and such that only its last reaction has a product in $\mathcal{C}$, thus forming a hyper-ear $*mu-vm*$ where m symbols stand for any other hyper-ear motif. Below, we show that hyper-ears are either one of the types II-V, or contain a TYPE II, III or IV core as a subgraph, overall demonstrating that $\mathcal{H}$ is necessarily a hyper-ear of TYPE II-V. □

Before continuing the proof of the theorem, we show an additional property based on the sufficiency of TYPE I and TYPE III.

Proposition 7. *In a core, a minimal path can have back-branches but no forward-branch, and the cycles formed by back-branches are non-intersecting.*

Proof. Consider a minimal path $\mathcal{P}$ in a core $\mathcal{H}$. Forward-branches require reactions to have multiple reactants, contradicting Property 1. Therefore, $\mathcal{P}$ is either a simple path or it has back-branches. Suppose there exists two back-products y and z of the respective reactions r_y and r_z, respectively forming intersecting cycles $\mathcal{C}_y$ and $\mathcal{C}_z$ (Fig. S1e). Without loss of generality, y and z can be taken closest, y upstream of z, and such that there is no other cycle nested within $\mathcal{C}_y$ and $\mathcal{C}_z$ than possibly themselves. $\mathcal{C}_z$ is necessarily weight-symmetric, as it would otherwise contradict the minimality of $\mathcal{H}$. There are two cases. Case 1: r_z is upstream r_y ($\mathcal{C}_z$ is nested within $\mathcal{C}_y$). Call $\mathcal{P}'$ the path from the product of r_z to r_y then y then z. Then $\mathcal{C}_z$ and $\mathcal{P}'$ form a TYPE III core. Case 2: r_y is upstream r_z ($\mathcal{C}_z$ and $\mathcal{C}_y$ are entangled). Call $\mathcal{P}'$ the path joining y to z. Then $\mathcal{C}_z$ and $\mathcal{P}'$ form a TYPE III core. Overall, we have shown that the existence intersecting cycles along a minimal contradicts the minimality of $\mathcal{H}$. □

We now go back to the proof of the theorem.

Proof. Consider the minimal path $\mathcal{P}$ of the hyper-ear, where $\mathcal{P}$ starts at x. By Property 7, $\mathcal{P}$ has no forward-branch.

Case A (TYPE II): Suppose $\mathcal{P}$ has one or more back-branch. Consider a reaction R' of $\mathcal{P}$ which has a back-branch, with back-product y' and forming a cycle $\mathcal{C}'$. $\mathcal{C}'$ is necessarily weight-symmetric by minimality of $\mathcal{H}$. Call x' the product of R' outside $\mathcal{C}'$. Consider the path $\mathcal{P}'$ starting at x', then going downstream of $\mathcal{P}$ until it reaches a shortest subpath $u-v$ of $\mathcal{C}$, then to x, and finally from x to y' in $\mathcal{P}$. By Property 7, all cycles along $\mathcal{P}$ are non-intersecting, and by construction, $\mathcal{C}$ is non-intersecting with the cycles of $\mathcal{P}$. Consequently, $\mathcal{P}'$ only has non-intersecting cycles, so that $\mathcal{C}'$ and $\mathcal{P}'$ form a TYPE II motif. Thus, the base cycle of the hyper-ear cannot have species outside $u-v$, as otherwise the core made of $\mathcal{C}'$ and $\mathcal{P}'$ would be a strict subset of $\mathcal{H}$, contradicting its minimality. This shows that $\mathcal{H}$ is a $*u-v*$ hyper-ear with a minimal path comprising non-intersecting cycles, the latter being necessarily weight-symmetric. Thus $\mathcal{H}$ is a TYPE II core.

Case B (TYPES III-V): Suppose $\mathcal{P}$ has no back-branch, in other words $\mathcal{P}$ is a simple path.

Subcase B.1 (TYPE III): Suppose $\mathcal{H}$ a hyper-ear which has only one u and one v symbol, thus a $*u-v-*$ hyper-ear. The case $*u-v*$ with a simple path is covered by the definition of TYPE II. If the $'-'$ symbol does not represent an empty path, then it comprises at least one species, so that $\mathcal{H}$ is a $*u-vo*$ hyper-ear with simple path $\mathcal{P}$, corresponding to TYPE III.

Subcase B.2 (TYPE IV): Suppose the hyper-ear comprises one u or v symbol in addition to the $u-v$ motif. We first note that $u-u-v$ and $u-v-v$ proto-ears are isomorphic to $*u-v-*$, which falls into the categories of TYPE II or TYPE III cores. Therefore, in a core, a hyper-ear containing two successive u or v symbols is necessarily of the form $*u-u-v*$ or $*v-v-u*$, or their cyclic permutations, as any additional symbol would allow to find a strict subgraph $u-u-v$ or $u-v-v$ forming a core, contradicting the minimality of $\mathcal{H}$. Furthermore, $*u-u-v*$ and $*v-v-u*$ hyper-ears are isomorphic. Indeed, the matrices of their reduced forms correspond to the matrix shown in the SUFFICIENCY section of the theorem, where exactly one coefficient is set to zero. Thus, these motifs as well as all their cyclic permutations, fall into the category of TYPE IV.

Subcase B.3 (TYPE V): Subcase B.2 imposes that any motif containing four or more u or v symbols must alternate u and v symbols. Any motifs with five or more u or v symbols contains, up to permutation, a $u-v-u-v$ proto-ear as a strict subgraph. However, the latter is isomorphic to TYPE IV cores, which implies that $*u-v-u-v*$ (TYPE V) is the only motif in this class which does not contradict the minimality of $\mathcal{H}$.

□

III. CHEMICAL INTERPRETATION OF AUTOCATALYTIC CORES

We remind that the notion of 'graph cycle' (closed successions of nodes and edges) differs from the notion 'reaction cycle' (right null vectors of the stoichiometric matrix). The latter name was historically chosen because the two notions overlap in the particular case of the simplest catalytic cycles, but there are counter-examples for both (we give one later). We employ 'reaction cycle' and 'graph cycle' to distinguish these notions.

By the minimality of autocatalytic cores, an autocatalytic motif is either a core, or it contains one or several cores. If P is a property of cores, then for any autocatalytic motif, either it verifies P, or it contains an autocatalytic motif verifying P. In particular,

- by Property 1, every autocatalytic motif contains an invertible autocatalytic motif;
- by Property 2, every autocatalytic motif contains an autocatalytic motif such that every product is also a catalyst of the reaction, or equivalently, such that every species appears on both side of the total chemical equation.

The five categories of cores are schematically represented in Fig. S1c. The convention of representation are as follows. The edges of the graph correspond to reactions and the yellow nodes to species. Reaction have two sides, the reactant side and product side, and each side of the edge representing the reaction connects to one or several nodes representing the species, with a stoichiometry for each connection. In the Type I core, the edge from the top node to the bottom node ends with a fork, the two ends of which connect to a single bottom node (S1c). This means that the bottom species is produced with a stoichiometry of 2 by this reaction. For simplicity, we have represented the connection between all other edges and nodes without fork. However, in all generality, any connection between a edge and a node could be a fork (e.g. could be of stoichiometry >1), as soon as the rules on graph cycle symmetry are respected, as explained below. Forks also appear in Types II-V, but where they connect to two distinct nodes, which are two distinct product species. The orange squares indicate that the reaction represented can be replaced by a chain of reactions with a single reactant species and a single product species.

The mathematical derivations are done without constraint on the number of reactants or products a reaction step can have. However, in a chemical system,

elementary reactions are in principle either unimolecular or bimolecular. This restricts the range of possible core graphs. This restriction operates only at the level of the stoichiometry of each single reaction, not at the level of the overall structure of cores. Indeed, even without invoking the restriction on bimolecularity, the mathematical derivation results in cores such that every reaction step involves at most a single species as reactant and at most two species as products, so that bimolecularity can always be respected, provided rules on the stoichiometry of each connection.

In a chemical autocatalytic core, an edge with a fork connecting to two distinct nodes (e.g. a reaction with two distinct product species) must have a stoichiometry 1 for each connection in order to respect bimolecularity. If a reaction has only one product species, then its stoichiometry can be 1 or 2, as soon as it respects rules on cycle stoichiometry, which we detail now.

Consider a simple graph cycle $\mathcal{C}$, simple meaning that every reaction as a single reactant species and a single product species. Note a_i (resp. b_i) the stoichiometry of the reactant (resp. product) of reaction i. If $\prod_i a_i \neq \prod_i b_i$, we say that $\mathcal{C}$ is weight-asymmetric, otherwise it is weight-symmetric. Weight-asymmetric simple graph cycles are an example of graph cycle which has no reaction cycle. Weight-symmetric graph cycle taken in isolation have a reaction cycle (their determinant is zero as shown in the derivation of cores above), thus they correspond to allocatalytic cycles in the context of a larger reaction graph where the reactions of the graph cycle consume and/or produce species outside of its own species set. The most classic example of an allocatalytic cycle is represented in Fig S1f, where reactant S is provided from the environment, binds to catalyst E to form complex ES converted into EP, finally dissociated into E which is recycled, and product P.

Type I cores are weight-asymmetric simple graph cycles, as $\mathcal{C}$ in S1c . Consequently, in types II-V, any simple graph cycle must be weight-symmetric (these graph cycles corresponding reaction cycles), otherwise the core would contain a Type I core, contradicting its minimality. For example in S1c: in Type II, $\mathcal{C}$ must be symmetric, but this does not apply to $\mathcal{C}'$ because it is not even a simple cycle; in types III-V, $\mathcal{C}$, $\mathcal{C}'$ and $\mathcal{C}''$ must be symmetric. Type II cores consist of a large graph cycle ($\mathcal{C}'$ in Fig S1c) comprising smaller graph cycles embedded within it (for example $\mathcal{C}$ in Fig S1c). Given the above, each of these smaller cycles must be weight-symmetric, and obeys the definition of an allocatalytic cycle. The Type II category includes circularly closed successions of such allocatalytic cycles, where the product of one allocatalytic cycles is the catalyst of the next allocatalytic cycle, which is a typical example of autocatalytic set. In addition however, Type II allows intermediate non catalyzed reaction steps.

Notably, every core follows the basic structure represented on Fig S1d, comprising a base cycle $\mathcal{C}$ and a minimal path (in the sense of reaction hypergraphs, see former section) starting from a reaction fork and joining back to a node of $\mathcal{C}$. This structure should enable a systematic algorithmic search for autocatalytic motifs in large stoichiometries.

We illustrate the concepts of autocatalytic submotif and the properties demonstrated above on a toy model for the Formose reaction, to which we have added one auxiliary reaction $\mathrm{C_3} \rightleftharpoons \mathrm{D_3}$, so that we obtain a stoichiometric matrix

$$\boldsymbol{\nu} = \begin{array}{c} \\ \mathrm{C_1} \\ \mathrm{C_2} \\ \mathrm{C_3} \\ \mathrm{D_3} \\ \mathrm{C_4} \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1 & -1 & 0 & 0 \\ -1 & 0 & 2 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix} \end{array} \quad \begin{array}{r} \mathrm{C_1 + C_2} \overset{1}{\rightleftharpoons} \mathrm{C_3} \\ \mathrm{C_1 + C_3} \overset{2}{\rightleftharpoons} \mathrm{C_4} \\ \mathrm{C_4} \overset{3}{\rightleftharpoons} \mathrm{2\,C_2} \\ \mathrm{C_3} \overset{4}{\rightleftharpoons} \mathrm{D_3} \end{array} \tag{20}$$

this full matrix has a left nullvector $\boldsymbol{l} = (1, 2, 3, 3, 4)$, i.e. we have a mass-like conservation law

$$L = n_{\mathrm{C_1}} + 2n_{\mathrm{C_2}} + 3n_{\mathrm{C_3}} + 3n_{\mathrm{D_3}} + 4n_{\mathrm{C_4}}. \tag{21}$$

Thus, this matrix does not correspond to an autocatalytic motif. Upon removing $\mathrm{C_1}$ (then considered as a feedstock molecule), we obtain an autocatalytic matrix $\boldsymbol{\nu}_*$

$$\boldsymbol{\nu}_* = \begin{array}{c} \\ \mathrm{C_2} \\ \mathrm{C_3} \\ \mathrm{D_3} \\ \mathrm{C_4} \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1 & 0 & 2 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix} \end{array} \quad \begin{array}{r} \mathrm{C_2} \overset{1}{\rightleftharpoons} \mathrm{C_3} \\ \mathrm{C_3} \overset{2}{\rightleftharpoons} \mathrm{C_4} \\ \mathrm{C_4} \overset{3}{\rightleftharpoons} \mathrm{2\,C_2} \\ \mathrm{C_3} \overset{4}{\rightleftharpoons} \mathrm{D_3} \end{array} \tag{22}$$

This matrix is autonomous and invertible with inverse:

$$\boldsymbol{\nu}_*^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 2 & 2 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} = (\boldsymbol{g}^{\mathrm{C_2}}, \boldsymbol{g}^{\mathrm{C_3}}, \boldsymbol{g}^{\mathrm{D_3}}, \boldsymbol{g}^{\mathrm{C_4}}). \tag{23}$$

The columns of the inverse are reaction vectors associated to a given species, of which they produce one extra unit. We will refer to them as elementary production modes.

Among the production modes, $\boldsymbol{g}^{\mathrm{D_3}} = (2, 1, 1, 1)^T$ is the sole vector which uses the auxiliary reaction, and $\mathrm{D_3}$ only ever occurs as a product

$$2\mathrm{C_2} + 2\mathrm{C_3} + \mathrm{C_4} \underset{-\boldsymbol{g}^{\mathrm{D_3}}}{\overset{\boldsymbol{g}^{\mathrm{D_3}}}{\rightleftharpoons}} 2\mathrm{C_2} + 2\mathrm{C_3} + \mathrm{C_4} + \mathrm{D_3}. \tag{24}$$

If we now consider $\boldsymbol{\Gamma} = \boldsymbol{g}^{\mathrm{C_2}} + \boldsymbol{g}^{\mathrm{C_3}} + \boldsymbol{g}^{\mathrm{D_3}} + \boldsymbol{g}^{\mathrm{C_4}}$, we

obtain a reaction balance such that every species has a net increase:

$$\mathbf{\Gamma} : 7\mathrm{C}_2 + 6\mathrm{C}_3 + 4\mathrm{C}_4 \longrightarrow 8\mathrm{C}_2 + 7\mathrm{C}_3 + 5\mathrm{C}_4 + \mathrm{D}_3. \quad (25)$$

$\boldsymbol{\nu}_*$ is not minimal as D_3 is only produced, and does not participate as a catalyst. Indeed, Property 2 implies the existence of an autocatalytic submotif without species which does not participate as a catalyst. An autocatalytic submotif is obtained by removing the reaction

$$\mathrm{C}_3 \rightleftharpoons \mathrm{D}_3, \quad (26)$$

and the species D_3, to obtain the autonomous submatrix $\bar{\boldsymbol{\nu}}$

$$\bar{\boldsymbol{\nu}} = \begin{array}{c} \\ \mathrm{C}_2 \\ \mathrm{C}_3 \\ \mathrm{C}_4 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \end{array} \quad \begin{array}{l} \mathrm{C}_2 \overset{1}{\rightleftharpoons} \mathrm{C}_3 \\ \mathrm{C}_3 \overset{2}{\rightleftharpoons} \mathrm{C}_4 \\ \mathrm{C}_4 \overset{3}{\rightleftharpoons} 2\,\mathrm{C}_2 \end{array} \quad (27)$$

The matrix $\bar{\boldsymbol{\nu}}$ is also invertible

$$\bar{\boldsymbol{\nu}}^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 2 \end{array} \\ \begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \end{array} = (\boldsymbol{g}^{\mathrm{C}_2}, \boldsymbol{g}^{\mathrm{C}_3}, \boldsymbol{g}^{\mathrm{C}_4}). \quad (28)$$

The columns of $\bar{\boldsymbol{\nu}}^{-1}$ correspond to reaction vectors, whose application yields one net copy of the corresponding molecule, e.g. $\Delta n_k = (\bar{\boldsymbol{\nu}} \cdot \boldsymbol{g}^{(k)})_k = 1$. This is illustrated for C_2 in Fig. S2.

The individual replication cycles are

$$\mathrm{C}_2 + \mathrm{C}_3 + \mathrm{C}_4 \underset{-\boldsymbol{g}^{\mathrm{C}_2}}{\overset{\boldsymbol{g}^{\mathrm{C}_2}}{\rightleftharpoons}} 2\mathrm{C}_2 + \mathrm{C}_3 + \mathrm{C}_4 \quad (29)$$

$$2\mathrm{C}_2 + \mathrm{C}_3 + \mathrm{C}_4 \underset{-\boldsymbol{g}^{\mathrm{C}_3}}{\overset{\boldsymbol{g}^{\mathrm{C}_3}}{\rightleftharpoons}} 2\mathrm{C}_2 + 2\mathrm{C}_3 + \mathrm{C}_4$$

$$2\mathrm{C}_2 + 2\mathrm{C}_3 + \mathrm{C}_4 \underset{-\boldsymbol{g}^{\mathrm{C}_4}}{\overset{\boldsymbol{g}^{\mathrm{C}_4}}{\rightleftharpoons}} 2\mathrm{C}_2 + 2\mathrm{C}_3 + 2\mathrm{C}_4$$

We can construct $\mathbf{\Gamma} = \boldsymbol{g}^{\mathrm{C}_2} + \boldsymbol{g}^{\mathrm{C}_3} + \boldsymbol{g}^{\mathrm{C}_4}$, which leads to the overall reaction

$$5\mathrm{C}_2 + 4\mathrm{C}_3 + 3\mathrm{C}_4 \underset{-\mathbf{\Gamma}}{\overset{\mathbf{\Gamma}}{\rightleftharpoons}} 6\mathrm{C}_2 + 5\mathrm{C}_3 + 4\mathrm{C}_4. \quad (30)$$

We see here that every species has a net production and is on both sides of the balance, thus participates as a autocatalyst. Furthermore, $\bar{\boldsymbol{\nu}}$ is a Type I core and consequently does not contain any smaller autocatalytic submotifs.

A. Autoinduction

The concept of autoinduction was put forward by D. Blackmond[21], to distinguish between autocatalysis that is mediated by external catalysts (i.e. not part of the autocatalysts that are reproduced) called 'autoinduction', and autocatalysis that functions without the aid of external catalysts (i.e. all catalysts are autocatalysts). We thus obtain a hybrid of pure allocatalysis and autocatalysis, for which the simplest example would be

$$\mathrm{A} + \mathrm{B} + \mathrm{E} \rightleftharpoons 2\mathrm{A} + \mathrm{E}. \quad (31)$$

The IUPAC definitions impose that autoinduction qualifies as autocatalysis. It follows then from our framework that we can find autocatalytic cores in autoinduction networks, which is indeed confirmed in Fig. S3 for the two types of autoinduction that have been proposed[21].

The concept of 'autoinduction' addresses a notion of self-sufficiency (also encountered in RAF sets, Sec.III D): external allocatalysts become essential to succesful autocatalysis, yet they are not reproduced. Depending on the context, they could be seen as part of the environment, in the same sense as essential feedstock species. Examples of autoinduction occur in autocatalytic metabolic networks (with locally allocatalytic enzymes) or e.g. the formose reaction which is often catalyzed by base and divalent metal ions.

The presence of external allocatalytic cycles does not add new cycles to the autocatalytic core. A practical consequence is that one can write catalyzed reactions very compactly for the core, while still maintaining nonambiguity, which we make use of in our analysis of metabolic cycles.

B. Metabolic cycles

At least two metabolic cycles are known to be autocatalytic. In our analysis of autoinduction, we pointed out that the core does not contain external allocatalysts (here: enzymes). Written purely in terms of autocatalysts, we find a Type II autocatalytic core for the reverse Krebbs cycle (Fig. S4a). For the Calvin cycle depicted in Fig. S4b, we identify three Type I cores (two structurally equivalent, differing only in the reaction chosen to link the same core species) and 4 of Type II.

C. Chemical amplification

Chen et al.[38,54] advanced a general strategy to achieve self-amplifying behavior, demonstrated by the encapsulation of the reactive compound DCC in a cavitand (indicated by an oval around the molecule in

EDCC, Fig. S5), which they referred to as chemical amplification. This phenomenon is consistent with the IUPAC definition of autocatalysis, which we will now illustrate by finding its autocatalytic core.

Let us first assess a reaction network formed by two reactions proposed by Chen et al, in absence of the cavitand complexes. In that case, applying our formalism readily reveals that the network cannot exhibit autocatalysis: there are no autocatalytic submatrices.

$$\boldsymbol{\nu} = \begin{array}{c} \\ \text{DCC} \\ \text{X} \\ \text{I1} \\ \text{Y} \\ \text{DCU} \\ \text{Z} \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} -1 & 0 \\ -1 & 0 \\ 1 & -1 \\ 0 & -1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array} \qquad \begin{aligned} \text{DCC} + \text{X} &\overset{1}{\rightleftharpoons} \text{I1}, \\ \text{I1} + \text{Y} &\overset{2}{\rightleftharpoons} \text{DCU} + \text{Z}. \end{aligned} \tag{32}$$

Now, let us add the species EDCC and two exchange reactions with DCU and Z that liberate DCC, i.e.

$$\text{DCC} + \text{X} \overset{1}{\rightleftharpoons} \text{I1}, \tag{33}$$

$$\text{I1} + \text{Y} \overset{2}{\rightleftharpoons} \text{DCU} + \text{Z}, \tag{34}$$

$$\text{EDCC} + \text{DCU} \overset{3}{\rightleftharpoons} \text{DCC} + \text{EDCU}, \tag{35}$$

$$\text{EDCC} + \text{Z} \overset{4}{\rightleftharpoons} \text{DCC} + \text{EZ}, \tag{36}$$

for which we have the matrix

$$\boldsymbol{\nu} = \begin{array}{c} \\ \text{DCC} \\ \text{X} \\ \text{I1} \\ \text{Y} \\ \text{DCU} \\ \text{Z} \\ \text{EDCC} \\ \text{EDCU} \\ \text{EZ} \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1 & 0 & 1 & 1 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{37}$$

This network admits an autocatalytic submatrix, and we obtain an autocatalytic core of Type III consisting of the species DCC, I1, DCU and Z, as can be seen in Fig. S5b. The new reagent EDCC serves as a feedstock compound, that allows to dispense new DCC, that can now serve as an autocatalyst. The generality of the mechanism follows from an exchange that could have been performed with different reactants than DCU and Z to yield an equivalent network, as shown in Refs[38,54].

D. Autocatalysis in RAF sets

A definition of Reflexively Autocatalytic Food-generated sets can be found in Ref.[20]: *a set of reactions $\mathcal{R}$ is RAF if every reaction is catalyzed by at least one molecule involved in a reaction in $\mathcal{R}$, and every reactant in $\mathcal{R}$ can be constructed from the food set f by successive applications of reactions from $\mathcal{R}$.* The philosophy behind the RAF set is that 'every reaction' in a RAF-set can be accelerated by the catalysts themselves. RAF sets do not perform autoinduction, except when members of the food set f within a RAF-set also serve as allocatalysts.

In the RAF-set formalism, reaction and catalysis are distinct mathematical objects. Graphically, RAF-sets are typically represented as bipartite graphs (S6a), with nodes (white squares) corresponding to reactions, which connect to nodes (colored circles) which serve as reactants (entering bold arrow) and products (leaving bold arrow) via directed edges. A dashed arrow connecting to a reaction indicates that a species is a catalyst for a reaction. Within the RAF framework, the following terms are employed as distinct: i) autocatalytic reaction (*"is a single chemical reaction for which one of the products also catalyzes the reaction"*), ii) autocatalytic cycle (*"is a sequence of reactions that, once completed, results in two (or more) copies of the molecule that was started with"*such as the toy formose reaction), iii) autocatalytic set (or RAF set).

It is important to note that the RAF-set formalism and the distinctions above depend on the chosen level of coarse-graining of the description. In that description (see Fig. S6a) allocatalysts in the same allocatalytic cycle are represented by a single species. The combination of reactions that form the allocatalytic cycle is represented by a single dashed line, connecting to an uncatalyzed reaction. Here, there exists a level of description where a RAF set appears as autocatalytic reaction involving catalytic cycles. By comparing Figs. S6a, we see that the level of description is critical in assessing whether a network is a RAF-set or not. From the definition of a RAF-set, the detailed version of the network in Fig.S6a would not be a RAF, but the less detailed description next to it would be.

Another example is the toy formose reaction, where two different choices of coarse-graining yield two different description in the RAF framework. Even when neglecting the role of catalytic base and metal ions, the toy formose reaction without coarse-graining is not a RAF-set[29], since its individual steps are not catalyzed:

$$\text{C}_1 + \text{C}_2 \rightleftharpoons \text{C}_3, \quad \text{C}_1 + \text{C}_3 \rightleftharpoons \text{C}_4, \quad \text{C}_4 \rightleftharpoons 2\text{C}_2. \tag{38}$$

Framed solely in terms of C_1 and one autocatalyst (e.g. C_2, C_4) one could propose a description in which we are oblivious of these steps, and write it as an autocatalytic reaction in the sense of RAF sets, such as

$$2\text{C}_1 + \text{C}_2 \rightleftharpoons 2\text{C}_2, \; 4\text{C}_1 + \text{C}_4 \rightleftharpoons 2\text{C}_4. \tag{39}$$

This would work as long as we consider them satisfactory levels of description, depending on the context

(e.g. further reactivity of intermediates, separation of timescales): either of these hypothetical cases satisfy the RAF criteria.

An interesting contrast occurs between RAF sets and our formalism: RAF-sets require a level of coarse-graining without intermediates to define catalysis in terms of single catalysts, which ensures a compact and unique description. Our formalism requires each catalytic cycle to have at least one intermediate to satisfy nonambiguity. Although less compact, the description is flexible: adding further steps is always possible and does not alter the conclusions, which must remain in accord with chemical definitions.

In describing catalysis in terms of uncatalyzed reactions, it becomes possible to formalize the underlying structure of catalysis in different models. We will now illustrate how this formally establishes that all autocatalysis in the GARD model qualifies as a RAF-set.

E. GARD model

GARD stands for graded autocatalysis replication domain[18], and is a model for the autocatalytic assembly of amphiphile assemblies (e.g. micelles). It describes micelles with a composition $\boldsymbol{n} = \{n_1, ..., n_s\}$, that follow an evolution equation

$$\frac{dn_i}{dt} = \left(k_i^+ \rho_i N - k_i^- n_i\right)\left(1 + \frac{1}{N}\sum_{j=1}^{s} \beta_{ij} n_j\right). \quad (40)$$

The surfaces are in contact with a reservoir, that contains species $\mathrm{Z_i}$ at concentration ρ_i and that can enter the surface, which has an area proportional to N. The incorporation happens with a base rate of k_i^+, but can be facilitated by other amphiphiles, for which the catalytic rate enhancement is characterized by β_{ij}. A special ingredient in GARD is the division process, which splits a mature aggregate in two new ones, after achieving a maximal size.

We examine here stoichiometric mechanisms leading to (40). Let us consider amphiphiles A and B, which can be in the micelle or reservoir, labelled I and II (see Fig. S6). To catalyze incorporation of the other, a complex is formed with a reservoir species, and subsequent dissociation takes place in the micelle

$$\mathrm{A^I} + \mathrm{B^{II}} \rightleftharpoons [\mathrm{AB}]^\mathrm{I} \rightleftharpoons \mathrm{A^I} + \mathrm{B^I}, \quad (41)$$
$$\mathrm{B^I} + \mathrm{A^{II}} \rightleftharpoons [\mathrm{BA}]^\mathrm{I} \rightleftharpoons \mathrm{A^I} + \mathrm{B^I}. \quad (42)$$

This network corresponds exactly to the network discussed in Fig. S6a, where the products of two allocatalytic cycles serve as a catalyst for each other, a Type II core in our stoichiometric formalism.

More generally, labelling amphiphiles as $\mathrm{A_k}$ ($k \in \{1, 2, .., s\}$, the entry β_{ij} encodes the contribution for the allocatalytic cycle

$$\mathrm{A^{II}_i} + \mathrm{A^I_j} \underset{-\boldsymbol{c}^*}{\overset{\boldsymbol{c}^*}{\rightleftharpoons}} \mathrm{A^I_i} + \mathrm{A^I_j} \quad (43)$$

where $\boldsymbol{c}^*$ denotes a reaction vector for the catalytic cycle, for any description that verifies nonambiguity. When $i = j$, this is a Type I autocatalytic cycle (Fig. S6c). When $i \neq j$ (cross-incorporation), Type II autocatalytic cycles are obtained, which are built up from n sequential allocatalytic incorporation steps and which end in the incorporation of the original amphiphile. Fig. S6 shows examples for $n = 2$ and $n = 3$. The importance of the allocatalysis step (43) is graded by the entry β_{ij}. For a given n-step autocatalytic motif to exist, we require

$$\prod_{k=1}^{n} \beta_{s_{k+1}s_k} > 0, \quad s_1 \neq s_2 \neq .. \neq s_n. \quad (44)$$

In practice, all $\beta_{ij} > 0$, so all motifs exist in principle.

The starting point of GARD is (40), i.e. a coarse-grained description in which allocatalysis can be described as a single step and a single allocatalyst. From Fig. S6, we see that we can obtain networks in GARD that would be RAF-sets in the RAF-framework by coarse-graining an incorporation cycle to convert it to catalysis in the RAF sense. The illustrated procedure extends to all autocatalysis in GARD, which is fully characterized by the continuation of the structures in Fig. S6 to their n-step Type II analogues. (44) guarantees that each reaction in GARD is catalyzed. It follows that all autocatalysis in the GARD model can formally be treated as a RAF-set.

Interestingly, the RAF-set formalism treats catalysis as pertaining to chemistry in single phases, with the environment supplying food locally through rapid exchange. In GARD, we instead have phase-transfer catalysis between a bulk medium (II) and an interface (I). A species in the bulk then serves as the food. Once the exact same species enters the interface, however, it may cease to be abundant or have a fixed concentration due to rapid exchange. It may thus no longer have the properties ascribed to the food set in the bulk and should ipso facto be treated as a different species as described in the final section of the main text.

IV. THE EXTINCTION PROBLEM AND FIXATION

In this section, we show that the extinction problem (finding the extinction probability at long times, P_{ex}) can be solved by a mapping to a branching process. We will first derive how in a system initially at steady-state perturbed with dilute autocatalysts, key statistical properties emerge: autocatalytic species can

be treated as independent, and their environment as fixed. Extinction becomes exponentially less likely as the population size continues to grow, which means P_{ex} can be determined in the dilute limit.

We then proceed by applying the framework to a variety of networks.

A. Context

We consider a reaction network, in a macroscopic system (letting N denote the amount of chemical species, let us say $N > 10^{23}$, or similarly, let the system volume V be large) in a steady-state. For simplicity, let us first consider a CSTR (single phase, ideally mixed), with a residence time τ, corresponding to a uniform degradation rate k_d.

Now, we perturb the steady-state with a handful ($O(1)$) of new (not yet present in the system) autocatalysts $\{\mathrm{X_k}\} = \{\mathrm{X_1}, \mathrm{X_2}, ..., \mathrm{X_s}\}$ that are part of the same autocatalytic core.

We consider that the population can grow due to catalysis by autocatalysts, and the population decays by degradation reactions and effective degradation. For example, outflow out of a CSTR is considered as a first-order degradation process:

$$\mathrm{X_k} \rightarrow \emptyset \tag{45}$$

The problem we wish to solve is the extinction problem: For an initial population of autocatalysts $\{\mathrm{N_{X_k}}\} = \{N_{\mathrm{X_1}}, .., N_{\mathrm{X_s}}\}$ what is the probability $P_{ex}(\{\mathrm{N_{X_k}}\})$, that, after a long time, the autocatalyst population goes extinct ($\forall k \;\; N_{\mathrm{X_k}} = 0$)?

B. Large system limit

Let us first note that we (deliberately) consider the initial stochastic kinetics in a large system, with a small number of autocatalysts, such that reactions among autocatalysts of the kind

$$\mathrm{X_k} + \mathrm{X_j} \longrightarrow \text{"}Product(s)\text{"} \tag{46}$$

are exceedingly rare and slow (the probability that a given $\mathrm{X_k}$ molecule encounters another $\mathrm{X_j}$ in a given time-frame scales with $N_{\mathrm{X_j}}/N$), where $N_{\mathrm{X_j}}$ is initially of the order 1. It follows survival of autocatalytic cycles requiring such reactions in the forward sens is hampered in large systems.

Reactions that are first order in terms of autocatalysts are not hampered

$$\mathrm{X_k} \rightleftharpoons \text{"}Product(s)\text{"} \tag{47}$$

$$\mathrm{X_k} + \mathrm{Y_j} \rightleftharpoons \text{"}Product(s)\text{"} \tag{48}$$

where $\mathrm{Y_j}$ is a feedstock compound that was already (abundantly) present in the system at a fixed molar fraction $x_{\mathrm{Y_j}}$. The probability for one $\mathrm{X_k}$ molecule to encounter a $\mathrm{Y_j}$ does not change with N, as $x_{\mathrm{Y_j}}$ remains fixed (and macroscopic).

Note furthermore that, when autocatalysts are rare, reactions producing more autocatalysts are 'irreversible'

$$\mathrm{X_k} + \mathrm{Y_j} \longrightarrow \mathrm{X_l} + \mathrm{X_m} \tag{49}$$

in the sense that the reverse reaction is exceedingly more rare than the forward reaction.

C. Constant composition, constant transition rates

The effect of rare autocatalysts on the steady-state composition (maintained by influx and degradation in a CSTR) is initially small: every reaction introduces changes in molar fraction of the order $1/N$ (or in concentration terms, $1/V \propto 1/N$). For large N, the alterations of the composition will be vanishingly small while the autocatalysts are rare.

Consequently, we can approximate the reactor composition in which an autocatalyst is placed, as the steady-state composition before perturbation. We then assume that the molar fractions of species $\mathrm{Y_k}$ consumed by autocatalysts are sufficiently abundant i.e. $x_{\mathrm{Y_k}} \gg 1/N$, which was also required for (48). For sufficient N, deviations from this approximation become vanishingly small.

A given $\mathrm{X_k}$ will therefore have a fixed transition rate $w_k^+ = k x_{\mathrm{Y_j}}$ to perform (48). Similarly, for (47) and CSTR degradation, a fixed transition rate $w = k$ is found, depending only on a rate constant.

D. Independence

In the rare-autocatalyst regime, all reaction steps we consider for autocatalysts are first-order, and they occur at fixed rates. It follows that autocatalysts do not influence each other, and they can each be treated independently. Thus, we can treat each autocatalyst type separately:

$$P_{ex}(\{\mathrm{N_{X_k}}\}) = P_{ex}(N_{\mathrm{X_1}}) P_{ex}(N_{\mathrm{X_2}}) .. P_{ex}(N_{\mathrm{X_s}}). \tag{50}$$

Also each individual autocatalyst can be treated as such:

$$P_{ex}(N_{\mathrm{X_k}}) = P_{ex}(\mathrm{X_k})^{N_{\mathrm{X_k}}} \tag{51}$$

Where $P_{ex}(\mathrm{X_k})$ denotes the extinction probability of a population initially composed of a single $\mathrm{X_k}$ species.

We thus need a method for finding $P_{ex}(\mathrm{X_k})$. This will be obtained by mapping the problem to a branching process.

E. A branching process

An attractive method we propose for finding $P_{ex}(\mathrm{X_k})$ (from here on simplified to P_{ex}), is by considering the distribution of 'descendants' $\mathrm{X_k}$ a species $\mathrm{X_k}$ will generate. From this, we construct a Branching Process, represented chemically as a single parent molecule $\mathrm{X_s}$ yielding k descendants:

$$\mathrm{X_s} \xrightarrow{\mathfrak{P}_k} k\mathrm{X_s}, \tag{52}$$

with p_k a distribution of the number of descendants which depends on the network topology.

Knowing p_k suffices to find P_{ex}, since the probability to go extinct is the probability that all descendants independently ((51)) go extinct:

$$P_{ex} = p_0 + p_1 P_{ex} + p_2 P_{ex}^2 + ... = \sum_{k=0}^{\infty} p_k P_{ex}^k, \tag{53}$$

We will now highlight some possible choices for Branching Processes and their associated p_k.

F. A Birth-Death Process for the Type I cycle

Consider a simple Type-I cycle such as in Fig S7a. Here, simple refers to there being a direct path of (effective) unimolecular steps between the starting compound ($\mathrm{B_1}$) and final compound ($\mathrm{B_2}$), followed by a single fragmentation step producing two $\mathrm{B_1}$ from one $\mathrm{B_2}$

$$\mathrm{B_2} \longrightarrow 2\mathrm{B_1}. \tag{54}$$

Starting from the marked node ($\mathrm{B_1}$), let p_2 be the probability of successfully forming $2\mathrm{B_1}$, i.e.

$$\mathrm{B_1} \xrightarrow{p_2} 2\mathrm{B_1}, \tag{55}$$

where p_2 contains the contribution of all possible trajectories (here: going back and forth between $\mathrm{B_1}$ and $\mathrm{B_2}$) that precede the irreversible fragmentation step, i.e.

$$\mathrm{B_1} \xrightarrow{\Pi_1^+} \mathrm{B_2} \xrightarrow{\Pi_2^-} \mathrm{B_1} ... \xrightarrow{\Pi_1^+} \mathrm{B_2} \xrightarrow{\Pi_2^+} 2\mathrm{B_1} \tag{56}$$

Where Π_1^+, Π_2^- and Π_2^+ are success probabilities for the single reaction steps. These transitions compete with irreversible degradation processes

$$\mathrm{B_1} \xrightarrow{\Pi_1^\emptyset} \emptyset, \quad \mathrm{B_2} \xrightarrow{\Pi_2^\emptyset} \emptyset. \tag{57}$$

Where $\Pi_1^\emptyset$, $\Pi_2^\emptyset$ are success probabilities for the degradation reaction. Due to total probability conservation, we have

$$\Pi_1^\emptyset + \Pi_1^+ = 1, \tag{58}$$

$$\Pi_2^\emptyset + \Pi_2^- + \Pi_2^+ = 1 \tag{59}$$

Ultimately, a $\mathrm{B_1}$ species will either be replaced by 2 new ones ($2\mathrm{B_1}$), or none ($\emptyset$):

$$\emptyset \xleftarrow{p_0} \mathrm{B_1} \xrightarrow{p_2} 2\mathrm{B_1}, \tag{60}$$

which is a chemical representation of the simplest type of branching process: a birth-death process. In (53), the only nonzero terms will come from 0 descendants (p_0) and 2 descendants (p_2):

$$P_{ex} = p_0 + p_2 P_{ex}^2 = 1 - p_2 + p_2 P_{ex}^2. \tag{61}$$

Solving the quadratic equation yields two solutions

$$P_{ex}^{\pm} = \frac{1 \pm (1 - 2p_2)}{2p_2}. \tag{62}$$

For our problem, the 'physical' solution is P_{ex}^+ while $p_2 \geq 1/2$. Beyond that point, $P_{ex}^+ > 1$, while we require $0 \leq P_{ex} \leq 1$, so $P_{ex}^- = 1$ becomes the only physical solution.

$$P_{ex} = \begin{cases} \frac{1}{p_2} - 1, & p_2 \geq \frac{1}{2}, \\ 1, & p_2 < \frac{1}{2}, \end{cases} \tag{63}$$

The average number of descendants is $2p_2$, which means that below $p_2 = 1/2$ (the decay threshold) $\mathrm{B_1}$ is on average replaced by less then one $\mathrm{B_1}$ species and extinction is guaranteed.

G. A Branching Process for the same type I cycle

To illustrate that there is a variety of choices for the stochastic process under study, we will here consider an alternative choice for the simple Type-I cycle, which is more generally applicable. Noting that a single cycle is successful with probability p_2, we now consider the number of successful cycles a single $\mathrm{B_1}$ provides, knowing that at some point degradation intervenes with probability $1 - p_2$. A succession of k cycles preceding a failure will thus lead to k successors

$$\mathrm{B_1} \xrightarrow{p_k} k\mathrm{B_1}. \tag{64}$$

The probability to spawn k descendants, p_k, is the probability of k successful Bernoulli trials followed by failure, and follows a geometric distribution

$$p_k = p_2^k (1 - p_2). \tag{65}$$

Injecting this distribution in (53), we find a geometric series

$$P_{ex} = \sum_{k=0}^{\infty}(1-p_2)(p_2 P_{ex})^k = \frac{1-p_2}{1-p_2 P_{ex}}. \quad (66)$$

Multiplying both sides by $1-p_2P_{ex}$ then yields the quadratic equation (61) previously found for the Birth-Death process. This is necessary, since we are calculating the same quantity P_{ex} for the same network. It highlights that we may construct a variety of branching processes to find P_{ex}.

H. Microscopic expressions for p_2

We can construct p_2 from microscopic details. In terms of transition probabilities, we find p_2 by summing over all trajectories in (56):

$$p_2 = \sum_{k=0}^{\infty}\Pi_1^+(\Pi_2^-\Pi_1^+)^k\Pi_2^+ = \frac{\Pi_1^+\Pi_2^+}{1-\Pi_2^-\Pi_1^+} \quad (67)$$

A more detailed description is possible when our description is Markovian (i.e. reactions are sufficiently elementary). Let w_k^+ denote a forward transition rate, to go from $\mathrm{X_k}$ to $\mathrm{X_{k+1}}$. Let w_k^- denote a backward transition rate from $\mathrm{X_k}$ to $\mathrm{X_{k-1}}$ and $w_k^\emptyset$ the degradation rate for $\mathrm{X_k}$. We may then write

$$\Pi_1^+ = \frac{w_1^+}{w_1^+ + w_1^\emptyset},\ \Pi_2^- = \frac{w_1^-}{w_1^- + w_2^\emptyset + w_2^+}, \quad (68)$$

$$\Pi_2^+ = \frac{w_2^+}{w_1^- + w_2^\emptyset + w_2^+}. \quad (69)$$

I. The irreversible reaction limit

Simple Type-I cycles have been studied in the limit where all reactions proceed irreversibly[17,19,40] ($\forall k\ w_k^- \to 0$). In this limit backward reactions are ignored ($\Pi_2^- = 0$), which for our example leads to

$$p_2 = \Pi_1^+\Pi_2^+ = \frac{w_1^+}{w_1^+ + w_1^\emptyset}\frac{w_2^+}{w_2^+ + w_2^\emptyset}. \quad (70)$$

In this limit, there is only one trajectory that contributes to p_c, and each step involves a competition between the forward reaction and degradation only.

In studies using simple Type-I cycles, the fraction

$$\zeta_k = \frac{w_k^+}{w_k^+ + w_k^\emptyset}, \quad (71)$$

has been referred to as the specificity of reaction step[19,39,40,55] k, which for irreversible reactions coincides with the transition probability Π_k^+

$$\lim_{w_{k-1}^-\to 0}\Pi_k^+ = \lim_{w_{k-1}^-\to 0}\frac{w_k^+}{w_{k-1}^- + w_k^+ + w_1^\emptyset} = \zeta_k \quad (72)$$

For simple Type-I networks with n reaction steps (n distinct edges), the irreversible limit ((70)) generalizes to[39,55]

$$p_2 = \prod_{k=1}^{n}\Pi_k^+ = \prod_{k=1}^{n}\frac{w_k^+}{w_k^+ + w_k^\emptyset}, \quad (73)$$

which we will show more formally in the next section.

J. The simple Type I cycle with n steps

To expand our discussion on simple Type I cycles, we will now derive a general expression for p_2, when there are n steps, of which the first $n-1$ are treated as reversible. The problem is illustrated in Fig. S7c.

Let us denote $P_{\mathrm{X_k}\to\mathrm{X_j}}$ the probability to reach $\mathrm{X_j}$, starting from $\mathrm{X_k}$. For $P_{\mathrm{X_1}\to\mathrm{X_2}}$, there is only one trajectory:

$$\mathrm{X_1} \xrightarrow{\Pi_1^+} \mathrm{X_2}, \quad (74)$$

and hence $P_{\mathrm{X_1}\to\mathrm{X_2}} = \Pi_1^+$. For $P_{\mathrm{X_2}\to\mathrm{X_3}}$, we need to consider that we can go back and forth between $\mathrm{X_2}$ and $\mathrm{X_1}$:

$$\mathrm{X_1} \underset{\Pi_2^-}{\overset{\Pi_1^+}{\rightleftharpoons}} \mathrm{X_2} \xrightarrow{\Pi_2^+} \mathrm{X_3}. \quad (75)$$

We can absorb the contribution of hopping back-and-forth in the factor Γ_2:

$$P_{\mathrm{X_2}\to\mathrm{X_3}} = \Gamma_2\Pi_2^+ \quad (76)$$

$$\Gamma_2 = \sum_{k=0}^{\infty}(\Pi_2^-\Pi_1^+)^k = \frac{1}{1-\Pi_2^-\Pi_1^+} \quad (77)$$

Now, let us consider this argument when we want to find $P_{\mathrm{X_3}\to\mathrm{X_4}}$ for

$$\mathrm{X_1} \underset{\Pi_2^-}{\overset{\Pi_1^+}{\rightleftharpoons}} \mathrm{X_2} \underset{\Pi_3^-}{\overset{\Pi_2^+}{\rightleftharpoons}} \mathrm{X_3} \xrightarrow{\Pi_3^+} \mathrm{X_4}. \quad (78)$$

Now, we need to consider that we can go back and forth between $\mathrm{X_3}$ and $\mathrm{X_2}$, and that at $\mathrm{X_2}$ we can go back and forth between $\mathrm{X_2}$ and $\mathrm{X_1}$ (captured by Γ_2) before moving to $\mathrm{X_3}$. We absorb all this in the factor Γ_3:

$$P_{\mathrm{X_3}\to\mathrm{X_4}} = \Gamma_3\Pi_3^+ \quad (79)$$

$$\Gamma_3 = \textstyle\sum_{k=0}^{\infty}(\Pi_3^-\Gamma_2\Pi_2^+)^k = \frac{1}{1-\Pi_3^-\Gamma_2\Pi_2^+} \quad (80)$$

We can repeat this argument, e.g. for the kth step

$$\mathrm{X}_1 \underset{\Pi_2^-}{\overset{\Pi_1^+}{\rightleftharpoons}} \mathrm{X}_2 \underset{\Pi_3^-}{\overset{\Pi_2^+}{\rightleftharpoons}} ... \underset{\Pi_k^-}{\overset{\Pi_{k-1}^+}{\rightleftharpoons}} \mathrm{X_k} \xrightarrow{\Pi_k^+} \mathrm{X_{k+1}}, \tag{81}$$

which then yields the following recursive expressions for $k \geq 1$

$$P_{\mathrm{X_k}\to\mathrm{X_{k+1}}} = \Gamma_k \Pi_k^+ \tag{82}$$

$$\Gamma_{k+1} = \textstyle\sum_{s=0}^{\infty} (\Pi_{k+1}^- \Gamma_k \Pi_k^+)^s = \frac{1}{1-\Pi_{k+1}^- \Gamma_k \Pi_k^+}. \tag{83}$$

where Γ_2 imposes that $\Gamma_1 = 1$.

The total probability p_2 to reach $\mathrm{X_n}$ from $\mathrm{X_1}$ and then perform the final irreversible fragmentation reaction r_n, is then

$$p_2 = \left(\prod_{k=1}^{n-1} P_{\mathrm{X_k}\to\mathrm{X_{k+1}}} \right) \Pi_n^+ \Gamma_n = \prod_{k=1}^{n} \Pi_k^+ \Gamma_k, \tag{84}$$

When backward transitions become negligible ($\forall k\ \Pi_k^- \to 0$), we have $\forall k \geq 1\ \Gamma_k = 1$, and p_2 then acquires its well-known limit expression described by (73).

K. A Branching Process for a Type II cycle

As an example of a simple Type II cycle, we consider the network in Fig. S7e, starting with the species $\mathrm{C_1}$. Our approach will be reminiscent of our branching process in Sec. IV G, but with a repeated branching step.

With a probability p_C, $\mathrm{C_1}$ will successfully perform the allocatalytic cycle $r_1 + r_2 + r_3$ (with some possible back-and-forths), yielding overall

$$\mathrm{C_1} \xrightarrow{p_\mathrm{C}} \mathrm{C_1} + \mathrm{D_1}. \tag{85}$$

The probability $\mathcal{P}_k^\mathrm{D}$ that k successful cycles occur before the first failure (i.e. degradation $\mathrm{C_k} \to \emptyset$) is

$$\mathcal{P}_k^\mathrm{D} = (1 - p_\mathrm{C}) p_\mathrm{C}^k. \tag{86}$$

Corresponding effectively to the overal reaction

$$\mathrm{C_1} \xrightarrow{\mathcal{P}_k^\mathrm{D}} k\mathrm{D_1}. \tag{87}$$

Now, let p_D be the probability that $\mathrm{D_1}$ succesfully completes a cycle $r_4 + r_2 + r_3$ (including back-and-forths):

$$\mathrm{D_1} \xrightarrow{p_\mathrm{D}} \mathrm{C_1} + \mathrm{D_1}. \tag{88}$$

The probability of k successful cycles before failure becomes

$$\mathcal{P}_k^\mathrm{C} = (1 - p_\mathrm{D}) p_\mathrm{D}^k. \tag{89}$$

Combined, a single $\mathrm{C_1}$ has been then replaced according to

$$\mathrm{C_1} \longrightarrow s\mathrm{D_1} \longrightarrow (n_1 + n_2 + ... + n_s)\mathrm{C_1}. \tag{90}$$

Let us denote $k = \sum_{l=1}^{s} n_l$ as the number of descendants. The distribution of the number of descendants p_k then becomes

$$p_k = \sum_{s=0}^{\infty} \mathcal{P}_s^\mathrm{P} \sum_{n_1,..,n_s}^{\infty} \prod_{r=0}^{s} \mathcal{P}_{n_r}^\mathrm{DCU} \delta_{n_1+...+n_s}^k, \tag{91}$$

which simplifies to

$$p_0 = 1 - \beta \frac{\alpha}{1-\alpha}, \qquad p_k = \beta \alpha^k,\ k \geq 1 \tag{92}$$

where

$$\alpha = \frac{p_\mathrm{D}}{1 - p_\mathrm{C}(1 - p_D)}, \quad \beta = \frac{p_\mathrm{C}(1-p_\mathrm{D})(1-p_\mathrm{C})}{1 - p_\mathrm{C}(1-p_\mathrm{D})}. \tag{93}$$

From (53), we then find P_{ex}. By rewriting the geometric series due to (93), we have

$$P_{ex} = 1 - \beta \frac{\alpha}{1-\alpha} - \beta \frac{\alpha P_{ex}}{1 - \alpha P_{ex}}, \tag{94}$$

which admits the solutions $P_{ex} = 1$ and

$$P_{ex} = \frac{\beta}{\alpha - 1} + \frac{1}{\alpha} = \frac{1 - p_\mathrm{C}}{p_\mathrm{D}}. \tag{95}$$

L. Microscopic expressions for p_C and p_D

Let us first consider how p_C can be constructed from smaller reaction steps. To do so, we observe that the first step must be

$$\mathrm{C_1} \xrightarrow{\Pi_1^+} \mathrm{C_2}, \tag{96}$$

with Π_1^+ the probability of success, competing with degradation

$$\mathrm{C_1} \xrightarrow{\Pi_1^\emptyset} \emptyset, \tag{97}$$

for which $\Pi_1^\emptyset = 1 - \Pi_1^+$.

Arrived at $\mathrm{C_2}$, going back-and forth reversibly

becomes possible for neighboring nodes C_1, C_3 and D_1:

$$C_1 \underset{\Pi_2^-}{\overset{\Pi_1^+}{\rightleftharpoons}} C_2 \underset{\Pi_3^-}{\overset{\Pi_2^+}{\rightleftharpoons}} C_3, \tag{98}$$

$$D_1 \underset{\Theta_1^-}{\overset{\Theta_1^+}{\rightleftharpoons}} C_2, \tag{99}$$

Where $\Pi_k^+, \Pi_k^-, \Theta_k^-, \Theta_k^+$ all denote success probabilities. Starting at C_1, a successful trajectory necessarily starts with reaction r_1 (Π_1^+) and ends with r_2+r_3 ($\Pi_2^+\Pi_3^+$), all in the forward sense. In between, we can go back-and-forth to C_1,C_3 and D_1 ($\Pi_2^-\Pi_1^+ + \Theta_2^-\Theta_1^+ + \Pi_2^+\Pi_3^-$) any number of times. Summing over all possible trajectories, p_C then becomes

$$p_C = \sum_{k=0}^{\infty} \Pi_1^+ (\Pi_2^-\Pi_1^+ + \Theta_2^-\Theta_1^+ + \Pi_2^+\Pi_3^-)^k \Pi_2^+\Pi_3^+ \tag{100}$$

which sums to

$$p_C = \frac{\Pi_1^+\Pi_2^+\Pi_3^+}{1-(\Pi_2^-\Pi_1^+ + \Theta_2^-\Theta_1^+ + \Pi_2^+\Pi_3^-)} \tag{101}$$

Starting at D_1, a successful trajectory necessarily starts with r_4 (Θ_1^+) in the forward sense, to form C_2. From there on, a successful trajectory follows the previous calculation, i.e. $p_D = \left(\Theta_1^+/\Pi_1^+\right) p_C$:

$$p_D = \frac{\Theta_1^+\Pi_2^+\Pi_3^+}{1-(\Pi_2^-\Pi_1^+ + \Theta_2^-\Theta_1^+ + \Pi_2^+\Pi_3^-)} \tag{102}$$

In the limit where all reactions proceed irreversibly forward (Sec. IV I), p_C and p_D only have a contributing from a single straight trajectory

$$p_C = \Pi_1^+\Pi_2^+\Pi_3^+, \quad p_D = \Theta_1^+\Pi_2^+\Pi_3^+. \tag{103}$$

M. Type III cycles with one fragmentation step

Let us now consider a Type III network composed of n species $\{X_1, .., X_n\}$ and reaction steps $\{r_1, .., r_n\}$, where the final fragmentation step r_n produces

$$X_n \longrightarrow X_1 + X_s, \quad 1 < s < n, \tag{104}$$

as shown in Fig. S7e

We may then introduce the success rates for the allocatalytic cycles for X_1 and X_s

$$X_1 \xrightarrow{p_{c,1}} X_1 + X_s, \tag{105}$$

$$X_s \xrightarrow{p_{c,s}} X_1 + X_s. \tag{106}$$

Which was exactly our starting point in Sec. IV K. Starting at X_s, we may thus directly use P_{ex}, upon replacing p_C with p_s and p_D with p_1, thus finding

$$P_{ex} = \frac{1-p_{c,s}}{p_{c,1}}. \tag{107}$$

We now turn to the problem of finding expressions for $p_{c,s}$ and $p_{c,1}$.

By structural analogy to simple Type I cycles (apart from the fragmentation step), we may again write

$$P_{X_k \to X_{k+1}} = \Pi_k^+\Gamma_k \tag{108}$$

$$\Gamma_{k+1} = \textstyle\sum_{s=0}^{\infty}(\Pi_k^-\Gamma_k\Pi_k^+)^s = \frac{1}{1-\Pi_{k+1}^-\Gamma_k\Pi_k^+}. \tag{109}$$

with $\Gamma_1 = 1$.

Since $p_{c,s} = \left(\prod_{k=s}^{n-1} P_{X_k \to X_{k+1}}\right) \Pi_n^+\Gamma_n$ we find

$$p_{c,s} = \left(\prod_{k=s}^{n} \Pi_k^+\Gamma_k\right) \tag{110}$$

N. Symmetric motifs

When the network motif is symmetric in structure, and if the transitions preserve this symmetry, this can be exploited to simplify calculations and gain insight in topological aspects of autocatalysis and robustness. Experimentally, this symmetry rarely applies for the transitions, but it can be made applicable for the purpose of our analysis, e.g. by setting transition probabilities to values that reflect the structural symmetry.

Consider a series of m linked allocatalytic cycles (Fig. S8b), all consisting of n nodes and n edges which are structurally equivalent:

$$X_{kn+1} \xrightarrow{\Pi_{kn+1}^+} .. \xrightarrow{\Pi_{(k+1)n}^+} X_{kn+1} + X_{(k+1)n+1}, \tag{111}$$

which loops back at the mth cycle:

$$X_{m-n+1} \xrightarrow{\Pi_{m-n+1}^+} .. \xrightarrow{\Pi_{mn}^+} X_1 + X_{m-n+1}, \tag{112}$$

As before, Π_k^+ denotes a forward transition probability, and we also introduce reverse reactions and degradation in the usual sense:

$$X_k \xrightarrow{\Pi_k^-} X_{k-1}, \tag{113}$$

$$X_k \xrightarrow{\Pi_k^\emptyset} \emptyset. \tag{114}$$

Now, let us choose transition probabilities such that they are equivalent among the equistructural allocatalytic cycles, i.e. periodic in n: $\Pi_k^+ = \Pi_{k+n}^+$ and idem for reverse steps ($\Pi_k^- = \Pi_{k+n}^-$) and by extension, degra-

dation ($\Pi_k^{\emptyset} = \Pi_{k+n}^{\emptyset}$), since $\Pi_k^+ + \Pi_k^- + \Pi_k^{\emptyset} = 1$.

Then, finding $P_{ex}(\mathrm{X_k})$ is no different [footnote: provided we choose $1 \leq k \leq m-n$] from finding $P_{ex}(\mathrm{X_{k+n}})$, which is seen readily in Fig. S8a by rotating the networks and interchanging the labels. Applying this symmetry to Fig.S8a for the six-membered cycle $(n = 2, m = 3)$, we can thus write

$$\begin{aligned} P_{ex}(\mathrm{X_1}) = P_{ex}(\mathrm{X_3}) = P_{ex}(\mathrm{X_5}). \\ P_{ex}(\mathrm{X_2}) = P_{ex}(\mathrm{X_4}) = P_{ex}(\mathrm{X_6}). \end{aligned} \tag{115}$$

We characterize the success of each equivalent allocatalytic cycle (n steps) by the same probability p_2. We can then express the extinction probability in terms of the reaction products of one allocatalytic cycle:

$$P_{ex}(\mathrm{X_1}) = 1 - p_2 + p_2 P_{ex}(\mathrm{X_1}) P_{ex}(\mathrm{X_3}) \tag{116}$$

$$= 1 - p_2 + p_2 P_{ex}(\mathrm{X_1})^2, \tag{117}$$

where we have used the symmetry in (116), which yields the exact second order equation we obtained for a simple Type I cycle of n steps. We can thus resolve the general problem using our previously derived solutions.

V. EXPRESSIONS FOR FIG 3, A SURVEY OF P_{ex} FOR VARIOUS STRUCTURES

In Fig. 3 in the main text, the behavior of P_{ex} is compared for a number of networks (N_1 to N_5), in the limit where all reaction steps proceed irreversibly (Sec. IV I), and where all reaction steps do so with a common success probability ζ (also known in the literature as specificity). Of course, this is an abstraction that is hard to realize experimentally, and its purpose is the following: by controlling for kinetics, we can systematically investigate and compare how survival is affected by network topology. In this section, we will derive the functional dependence of P_{ex} on ζ for the structures discussed in the main text.

A. N_1: a 6-membered Type I cycle

In Sec. IV J, we derived the general solution for n-membered Type I cycles in terms of the probability p_2 to reach X_n and perform r_n. Starting from $\mathrm{X_1}$, we can now recover the solution for $n = 6$

For $n = 6$, p_2 becomes

$$p_2 = \left(\prod_{k=1}^{5} P_{\mathrm{X_k} \to \mathrm{X_{k+1}}} \right) \Pi_6^+ \Gamma_6 = \prod_{k=1}^{6} \Pi_k^+ \Gamma_k \tag{118}$$

Moving to the irreversible reaction limit, and controlling the reaction specifity ($\forall k \geq 1\ \Pi_k^+ = \zeta,\ \Gamma_k = 1$), we obtain $p_2 = \zeta^6$. Upon injection in the solution $P_{ex} = (1 - p_2)/p_2$ (for $p_2 \geq 1/2$) ((63)), we then find

$$P_{ex} = \frac{1 - \zeta^6}{\zeta^6} \tag{119}$$

B. N_2: a 6-membered asymmetric Type III cycle

In Sec. IV M, the general solution for n-membered Type III cycles with one fragmentation reaction was derived. For a 6-membered cycle with the fragmentation reaction

$$\mathrm{X_6} \longrightarrow \mathrm{X_1} + \mathrm{X_4}, \tag{120}$$

which corresponds to network N_2 in the main text.

Having $\mathrm{X_4}$ as the starting species, we can express $P_{ex}(\mathrm{X_4})$ in terms of (110)

$$P_{ex} = \frac{1 - p_{c,4}}{p_{c,1}}. \tag{121}$$

In the irreversible reaction limit with fixed specificity ($\forall k \geq 1\ \Pi_k^+ = \zeta,\ \Gamma_k = 1$), P_{ex} becomes

$$P_{ex} = \frac{1 - \zeta^3}{\zeta^6} \tag{122}$$

C. N_3: 6-membered Type II network with RAF representation

The network N_3 consists of two nonoverlapping allocatalytic cycles, which produce each other's allocatalyst:

$$\mathrm{A_1} \underset{}{\overset{\Pi_1^+}{\rightleftharpoons}} \mathrm{A_2} \overset{\Pi_2^+}{\rightleftharpoons} \mathrm{A_3} \xrightarrow{\Pi_3^+} \mathrm{A_1} + \mathrm{B_1}, \tag{123}$$

$$\mathrm{B_1} \overset{\Theta_1^+}{\rightleftharpoons} \mathrm{B_2} \overset{\Theta_2^+}{\rightleftharpoons} \mathrm{B_3} \xrightarrow{\Theta_3^+} \mathrm{A_1} + \mathrm{B_1}. \tag{124}$$

Choosing our transition probabilities $\Pi_k^+ = \Theta_k^+$, we can exploit the network symmetry as outlined in Sec. IV N and $P_{ex}(\mathrm{A_1}) = P_{ex}(\mathrm{B_1})$. Let p_c be the probability that either $\mathrm{A_1}$ or $\mathrm{B_1}$ successfully finishes an allocatalytic cycle. We may then write

$$P_{ex}(\mathrm{A_1}) = 1 - p_c + p_c P_{ex}(\mathrm{A_1}) P_{ex}(\mathrm{B_1}) \tag{125}$$

$$= 1 - p_c + p_c P_{ex}(\mathrm{A_1})^2. \tag{126}$$

The success rate p_c can be expressed in the familiar way

$$p_c = \Pi_1^+ \Pi_2^+ \Pi_3^+ \Gamma_1 \Gamma_2 \Gamma_3. \tag{127}$$

In the irreversible limit ($\forall k \geq 1, \Gamma_k = 1$) with fixed specificity ζ, $p_c = \zeta^3$, and thus we find for $p_c \geq 1/2$

$$P_{ex} = \frac{1-\zeta^3}{\zeta^3}. \tag{128}$$

D. N_4: 6-membered symmetric Type III network

The network N_4 consists of two nonoverlapping allocatalytic cycles, which produce a precursor $(\mathrm{A}_0, \mathrm{B}_0)$ for each other's allocatalyst:

$$\mathrm{A}_0 \overset{\Pi_0^+}{\rightleftharpoons} \mathrm{A}_1 \overset{\Pi_1^+}{\rightleftharpoons} \mathrm{A}_2 \xrightarrow{\Pi_2^+} \mathrm{A}_1 + \mathrm{B}_0, \tag{129}$$

$$\mathrm{B}_0 \overset{\Theta_0^+}{\rightleftharpoons} \mathrm{B}_1 \overset{\Theta_1^+}{\rightleftharpoons} \mathrm{B}_2 \xrightarrow{\Theta_2^+} \mathrm{A}_0 + \mathrm{B}_1. \tag{130}$$

Choosing our transition probabilities to follow the symmetry of the network, we can write $\Pi_k^+ = \Theta_k^+$, and $P_{ex}(\mathrm{A_k}) = P_{ex}(\mathrm{B_k})$. Finally, we note that B_0 can either i) degrade with probability $\Theta_0^\emptyset = 1 - \Theta_0^+$, or ii) form B_1 with probability $\Theta_0^+ = \Pi_0^+$, such that

$$P_{ex}(\mathrm{B}_0) = 1 - \Pi_0^+ + \Pi_0^+ P_{ex}(\mathrm{B}_1) = 1 - \Pi_0^+ + \Pi_0^+ P_{ex}(\mathrm{A}_1). \tag{131}$$

Denoting p_c the probability that A_1 performs a successful allocatalytic cycle (yielding $\mathrm{A}_1 + \mathrm{B}_0$), we can write the extinction probability as

$$P_{ex}(\mathrm{A}_1) = 1 - p_c + p_c P_{ex}(\mathrm{A}_1) P_{ex}(\mathrm{B}_0), \tag{132}$$

which upon injecting (131) yields the following expression for $P_{ex}(\mathrm{A}_1)$

$$P_{ex} = 1 - p_c + p_c(1 - \Pi_0^+) P_{ex} + p_c \Pi_0^+ P_{ex}^2. \tag{133}$$

Solving the quadratic equation (133), we find

$$P_{ex}(\mathrm{A}_1) = \frac{1 - p_c(1-\Pi_0^+) \pm \sqrt{(p_c(1+\Pi_0^+) - 1)^2}}{2 p_c \Pi_0^+} \tag{134}$$

which yields $P_{ex} = 1$ and

$$P_{ex} = \frac{1-p_c}{p_c \Pi_0^+}. \tag{135}$$

We can write p_c in terms of back-and-forths starting at A_1, terminating with an irreversible fragmentation

$$p_c = \textstyle\sum_{k=0}^{\infty} (\Pi_1^- \Pi_0^+ + \Pi_1^+ \Pi_2^-)^k \Pi_1^+ \Pi_2^+ \tag{136}$$

$$= \frac{\Pi_1^+ \Pi_2^+}{1 - \Pi_1^- \Pi_0^+ + \Pi_1^+ \Pi_2^-}, \tag{137}$$

so that in the irreversible limit with fixed specificity ($\forall k \geq 1\ \Pi_{k+} = \zeta,\ \Pi_{k-} = 0$), we obtain

$$P_{ex} = \frac{1-\zeta^2}{\zeta^3}. \tag{138}$$

An alternative way of seeing this is that, by symmetry, P_{ex} is the same as that for a 3-membered Type III cycle with one fragmentation step (forming X_2), for which we can directly use the solution derived in Sec. IV M.

E. A trio of symmetric analogues

As derived in Sec. IV N, the interlinked allocatalytic cycles of size n behave, due to symmetry, as an n-membered simple Type-I cycle. Reproducing the solution for the 2-membered cycles (see Sec. IV F) in the irreversible limit with $\forall k \geq 1\ \Pi_k^+ = \zeta,\ \Pi_k^- = 0$, we thus find

$$P_{ex} = \frac{1-\zeta^2}{\zeta^2} \tag{139}$$

F. N_5: a type V core

In N_5 we have the reactions

$$\mathrm{A}_1 \rightleftharpoons \mathrm{A}_2 \longrightarrow \mathrm{B}_1 + \mathrm{C}_1, \tag{140}$$

$$\mathrm{B}_1 \rightleftharpoons \mathrm{B}_2 \longrightarrow \mathrm{A}_1 + \mathrm{C}_1, \tag{141}$$

$$\mathrm{C}_1 \rightleftharpoons \mathrm{C}_2 \longrightarrow \mathrm{A}_1 + \mathrm{B}_1. \tag{142}$$

If our transitions follow the symmetry of the network, we have $P_{ex}(\mathrm{A_k}) = P_{ex}(\mathrm{B_k}) = P_{ex}(\mathrm{C_k})$. Denoting p_c the success probability of the allocatalytic cycle, we can write P_{ex} in terms of (140)

$$P_{ex}(\mathrm{A}_1) = 1 - p_c + p_c P_{ex}(\mathrm{B}_1) P_{ex}(\mathrm{C}_1) \tag{143}$$

$$= 1 - p_c + p_c P_{ex}(\mathrm{A}_1)^2, \tag{144}$$

which is the solution found for the 2-membered cycle. Noting that

$$p_c = \Pi_1^+ \Pi_2^+ \Gamma_1 \Gamma_2 \tag{145}$$

the irreversible limit with fixed specificity ($\forall k \geq 1\ \Pi_k^+ = \zeta,\ \Gamma_k = 1$) yields, as in Sec. IV F,

$$P_{ex} = \frac{1-\zeta^2}{\zeta^2} \tag{146}$$

VI. EXPRESSIONS FOR FIG 4D, A PHASE DIAGRAM FOR MULTICOMPARTMENT AUTOCATALYSIS

Fig. 4d represents a case of a Type III core with one fragmentation reaction (Sec.IV M), consisting of 5

members:

$$\mathrm{AB}_\alpha \underset{}{\overset{r_1}{\rightleftharpoons}} \mathrm{A}_\alpha \overset{r_2}{\rightleftharpoons} \mathrm{A}_\beta \overset{r_3}{\rightleftharpoons} \mathrm{A_2B}_\beta \overset{r_4}{\rightleftharpoons} \mathrm{A_2B}_\alpha, \quad (147)$$

$$\mathrm{A_2B}_\alpha \xrightarrow{r_5} \mathrm{AB}_\alpha + \mathrm{A}_\alpha, \quad (148)$$

There are two degradation reactions (r_6 and r_7), both taking place in the compartment β:

$$\mathrm{A}_\beta \xrightarrow{r_6} \emptyset, \qquad \mathrm{A_2B}_\beta \xrightarrow{r_7} \emptyset. \quad (149)$$

Here r_k refers to the reaction label as found in Fig.4d, α, β to the compartments and $\mathrm{A}, \mathrm{AB}, \mathrm{A_2B}$ to species as defined in Fig.4d.

We are guaranteed that, starting from A_α or AB_α, eventually A_β will be formed with probability 1. We can thus simplify the problem by only considering reactions r_3 to r_5 and species $\mathrm{A}_\beta, \mathrm{A_2B}_\beta$ and $\mathrm{A_2B}_\alpha$, i.e. we can then write for the transition probabilities

$$\mathrm{A}_\beta \underset{\Pi_2^-}{\overset{\Pi_1^+}{\rightleftharpoons}} \mathrm{A_2B}_\beta \underset{\Pi_3^-}{\overset{\Pi_2^+}{\rightleftharpoons}} \mathrm{A_2B}_\alpha \xrightarrow{\Pi_3^+} 2\mathrm{A}_\beta, \quad (150)$$

$$\mathrm{A}_\beta \xrightarrow{\Pi_1^d} \emptyset, \qquad \mathrm{A_2B} \xrightarrow{\Pi_2^d} \emptyset. \quad (151)$$

We readily find that this effective description obeys the solution for a simple Type-I network with 3 members, with p_2 a success probability for a cycle

$$p_2 = \Pi_1^+ \Pi_2^+ \Pi_3^+ \Gamma_1 \Gamma_2 \Gamma_3 \quad (152)$$

$$P_{ex} = \frac{1 - p_2}{p_2} \quad (153)$$

Where we have used Γ_k for calculating back-and-forth trajectories, as derived in sec. IV J:

$$\Gamma_{k+1} = \sum_{s=0}^{\infty} (\Pi_{k+1}^- \Gamma_k \Pi_k^+)^s = \frac{1}{1 - \Pi_{k+1}^- \Gamma_k \Pi_k^+}. \quad (154)$$

with $\Gamma_1 = 1$. Let us now give a microscopic kinetic interpretation to the competing processes, by considering sufficiently elementary transitions on the level of a single species for which we can introduce rate constants

- $w_3^+ = k_3^+ x_{\mathrm{AB}_\beta}$, sequestration of AB_β (present with a fixed molar fraction x_{AB_β}), which plays the role of a feedstock species in compartment β in r_3 proceeding forward.
- $w_3^d = k_6^d$ degradation, reaction r_6.
- $w_4^- = k_3^-$ release of AB_β, r_3 proceeding backward.
- $w_4^+ = k_4^+$ exchange of $\mathrm{A_2B}$ from compartment β to α, when r_4 proceeds forward.
- $w_4^d = k_7^d$ degradation, reaction r_7.
- $w_5^- = k_4^-$ exchange, from compartment α to β, when r_4 proceeds backward.
- $w_5^+ = k_5^+$ release of $\mathrm{A_2B}_\alpha$ through locally irreversible reaction r_5.

We then obtain

$$\Pi_1^+ = \frac{w_3^+}{w_3^+ + w_3^d}, \quad \Pi_1^d = \frac{w_3^d}{w_3^+ + w_3^d}, \quad (155)$$

$$\Pi_2^+ = \frac{w_4^+}{w_4^+ + w_4^d + w_4^-}, \quad \Pi_2^- = \frac{w_4^-}{w_4^+ + w_4^d + w_4^-}, \quad (156)$$

$$\Pi_2^d = \frac{w_4^d}{w_4^+ + w_4^d + w_4^-}, \quad (157)$$

$$\Pi_3^+ = \frac{w_5^+}{w_5^+ + w_5^-}, \quad \Pi_3^- = \frac{w_5^-}{w_5^+ + w_5^-}. \quad (158)$$

Noting that $\Gamma_1 = 1$, the product $\Gamma_2 \Gamma_3$ simplifies to

$$\Gamma_2 \Gamma_3 = \frac{1}{1 - \Pi_2^- \Pi_1^+ - \Pi_3^- \Pi_2^+}. \quad (159)$$

This allows to fully express P_{ex} in terms of 8 microscopic coefficients. Reactions r_1 and r_2 would give 4 more rate constants and two more molar fractions (for chemostatted species). However, their values do not alter P_{ex}.

For the purpose of illustration, we will consider the competition between exchange, degradation, and other transitions. To do so, we choose one rate for exchange $k_4^+ = k_4^- = k^{ex}$ and one rate for degradation $k_6^d = k_7^d = k^d$. Furthermore, we let the sequestration-release steps be equally probable in compartment β, and match release in α: $w_3^+ = w_4^- = w_5^+ = k$. The transition success probabilities can then be expressed in terms of two ratios

$$\Delta = k^d/k, \qquad \Xi = k^{ex}/k \quad (160)$$

which upon substitution yields

$$\Pi_1^+ = \tfrac{1}{1+\Delta}, \quad \Pi_1^d = \tfrac{\Delta}{1+\Delta}, \quad \Pi_2^+ = \tfrac{\Xi}{1+\Delta+\Xi}, \quad (161)$$

$$\Pi_2^d = \tfrac{\Delta}{1+\Delta+\Xi}, \quad \Pi_2^- = \tfrac{1}{1+\Delta+\Xi}, \quad (162)$$

$$\Pi_3^+ = \tfrac{1}{1+\Xi}, \quad \Pi_3^- = \tfrac{\Xi}{1+\Xi}. \quad (163)$$

This permits to construct the phase diagram for P_{ex} in Fig. 4D in terms of the variables Δ and Ξ.

A. Phase boundaries and limits

Let us first find an expression for the boundary between autocatalysis and deterministic extinction, which occurs when $P_{ex} = 1$ and (153) coincide, i.e. when $p_c = 1/2$, which upon substitution of (159) becomes:

$$\Pi_1^+ \Pi_2^+ \Pi_3^+ = \frac{1}{2}(1 - \Pi_2^- \Pi_1^+ - \Pi_3^- \Pi_2^+) \quad (164)$$

In terms of Ξ and Δ, we have

$$2\frac{1}{1+\Delta}\frac{\Xi}{1+\Delta+\Xi}\frac{1}{1+\Xi} = 1 - \frac{1}{1+\Delta+\Xi}\frac{1}{1+\Delta} - \frac{\Xi}{1+\Xi}\frac{\Xi}{1+\Delta+\Xi} \quad (165)$$

Which rearranges to a linear dependence in Ξ

$$\Delta^2\Xi + \Delta^2 + 3\Delta\Xi + 2\Delta - \Xi = 0, \quad (166)$$

from which we obtain for the phase boundary

$$\Xi = \frac{\Delta(\Delta+2)}{1-3\Delta-\Delta^2}. \quad (167)$$

In the regime where reactions outpace degradation ($\Delta \ll 1$), extinction will be due to rate-limiting exchange ($\Xi \ll 1$). Taking (167), dividing by Δ and letting $\Delta \to 0$, the ratio Ξ/Δ tends to

$$\frac{\Xi}{\Delta} = 2, \quad (168)$$

as also seen in the phase diagram. When exchange is very rapid ($\Xi \to \infty$), it ceases to be rate-limiting, and degradation will only compete with other reactions. This occurs when we let the denominator of (167) become 0, i.e.

$$1 - 3\Delta - \Delta^2 = 0, \quad (169)$$

which has solutions

$$\Delta^{\pm} = \frac{-3 \pm \sqrt{13}}{2}, \quad (170)$$

of which $\Delta = \frac{-3+\sqrt{13}}{2} \approx 0.30$ is the only physical solution, as can also be seen in the phase diagram.

VII. STOCHASTIC SIMULATIONS

We can numerically sample the extinction probability $P_{ex}(\{N_X\}_0)$, with $\{N_X\}_0$ the initial autocatalyst population, by performing stochastic simulations of the kinetics using Gillespie's algorithm which start from the initial composition $\{N_X\}_0$, using transition rates as discussed previously. The compositions obtained from the algorithm are used to sample extinction or survival, as detailed below.

When the total autocatalyst population in a run reaches zero, this is sampled as an extinction event. When the population is not extinct at the end of a prescribed number of reactions n_r, the total number N_{ac} of autocatalysts is compared with a threshold population N_{tresh}. If $N_{ac} > N_{tresh}$, the run is sampled as survival (The error in this approximation scales as $P_{ex}^{N_{tresh}}$). Otherwise, the composition after n_r steps is used as an initial composition to repeat this protocol, again for n_r steps (and so forth if again the threshold is not reached). Once either the threshold is reached or extinction occurs, it is sampled accordingly and the run is terminated.

For networks N_1 to N_5, the symmetry of the problem is used to sample reactions: a random number χ between 0 and 1 is drawn, and if $\chi < \zeta$, a forward reaction step is performed for an autocatalyst picked at random. If $\chi > \zeta$, a degradation step is performed for an autocatalyst picked at random.

In Fig. S9a a short time interval is shown for 10 simulation runs performed for the multicompartment network using, for completeness, all 5 autocatalysts $\mathrm{AB}_\alpha, \mathrm{A}_\alpha, \mathrm{A}_\beta, \mathrm{A_2B}_\beta, \mathrm{A_2B}_\alpha$ (Since $\mathrm{AB}_\alpha, \mathrm{A}_\alpha$ return to A_β with probability 1, the exact solution for P_{ex} can also be found with the three-species network as shown in Sec. VI). We start with $N_{\mathrm{A}_\beta} = 1.0$ and all other autocatalysts at 0. Rate parameters for r_3 to r_7 follow the conventions outlined above, and are set at $k^{ex} = 1.0$,$k^d = 0.3$, $k_3^+ = 100$, $k = 10.0$. r_1 is treated as irreversible with an effective rate constant $k_1^+ = 10.0$ and exchange of A is performed with a rate k_{ex}. From 100000 simulation runs ($N_{tresh} = 40$) we obtain the numerical estimate of $P_{ex} = 0.698 \pm 0.004$. This is consistent with our exact solution (153), which yields $P_{ex} = 0.6999$.

Among the 10 short runs for the multicompartment network in Fig. S9a, 4 runs reach extinction within the simulation time, whereas the autocatalysis in one run (green) has reached an exponential growth regime that is close to deterministic with a total population of 40 autocatalysts.

VIII. MULTICOMPARTMENT AUTOCATALYSIS WITH THREE COMPARTMENTS

Let us consider a bimolecular reaction

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C}, \quad (171)$$

which can occur in three different compartments labeled α, β, γ, as shown in Fig. S9b. Let us couple these compartments through the following exchanges

$$\mathrm{A}_\alpha \rightleftharpoons \mathrm{A}_\beta, \qquad \mathrm{B}_\beta \rightleftharpoons \mathrm{B}_\gamma, \quad (172)$$

$$\mathrm{C}_\alpha \rightleftharpoons \mathrm{C}_\beta \rightleftharpoons \mathrm{C}_\gamma. \quad (173)$$

Removing A_γ, C_α then immediately yields the Type III autocatalytic core shown in Fig. S9c.

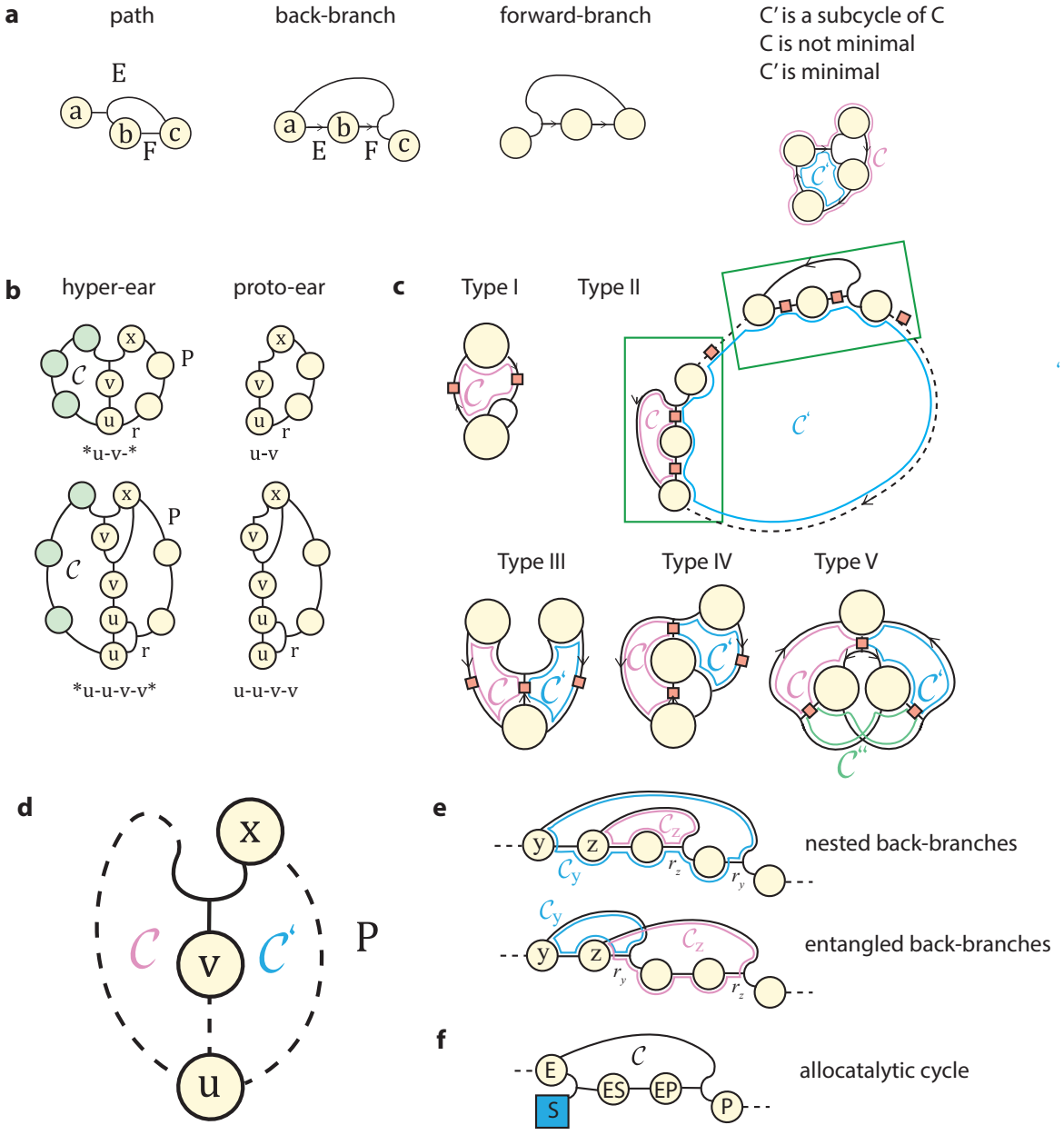

Figure S1: a) Hypergraph paths and cycle examples. b) Examples of hyper-ears (left) and associated proto-ears (right) obtained by removing green species. The syntax below graphs (see section II C) describes the relationships between non-P species and path P. Nodes 'u' are products of r, nodes 'v' are reactants of reactions producing x, '-' is any series of reactions with a single reactant and product, '*' denotes the closure of $\mathcal{C}$ by the green path. c) Autocatalytic cores. Edges are oriented consistently along cycles, so that reaction have a single reactant. Orange squares are chains of arbitrary length made of reactions with a single reactant species and product species. Edge-to-node connections are weighted by a stoichiometric coefficient, represented explicitly only for Type I by a fork (stoichiometry of 2). $\mathcal{C}$, $\mathcal{C}'$ and $\mathcal{C}'$ are cycles. In Type II, the dotted path may comprise multiple cycles similar to the green box. In main text Fig. 2, only the case of a single green box is represented for simplicity. In Type I, $det\mathcal{C} \neq 0$; in Type II, $det(\mathcal{C}) = 0$ and $det(\mathcal{C}')$ can be any value; in Types III-V, $det(\mathcal{C}) = det(\mathcal{C}') = det(\mathcal{C}'') = 0$. d) Generic hyper-ear structure of cores. e) Nested and entangled back-branches. f) Example of allocatalytic cycle.

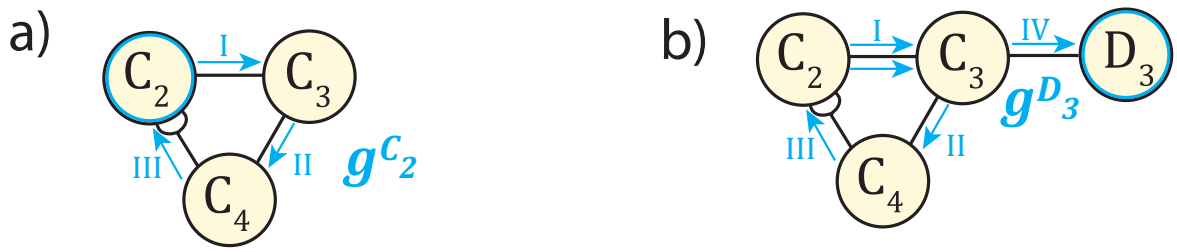

Figure S2: a) A decorated Toy Formose reaction given by the submatrix $\boldsymbol{\nu}_*$, obtained by removing C_1. The replication cycle $\boldsymbol{g}^{D_3}$ is illustrated in blue. In this network, only species C_2, C_3 and C_4 are autocatalysts. b) The minimal formose reaction in its autocatalytic subnetwork, an example of an SFA. Arrows illustrate the replication cycle $\boldsymbol{g}^{C_2}$.

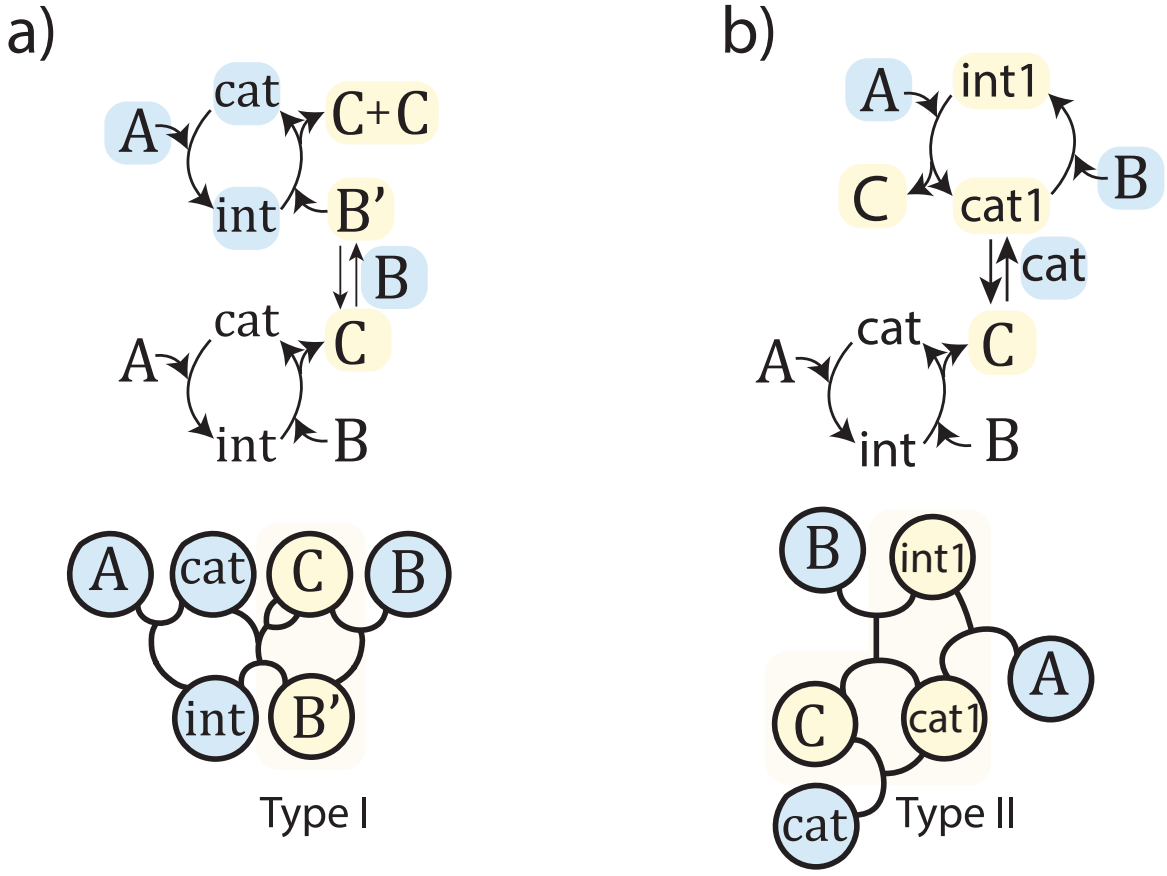

Figure S3: a) top: A product-enhanced cycle, Blackmond's first type of autoinduction. bottom: Hypergraph, containing a Type I autocatalytic core (yellow). External food and allocatalysts are marked in blue. b) top: A ligand-accelerated cycle, Blackmond's second type of autoinduction. bottom: Hypergraph, containing a Type II autocatalytic core (yellow) and supporting external food (blue). Note that, in both cases, external allocatalytic cycles are not part of the core. Allocatalysts are treated on equal footing with feedstock and waste in isolating a core.

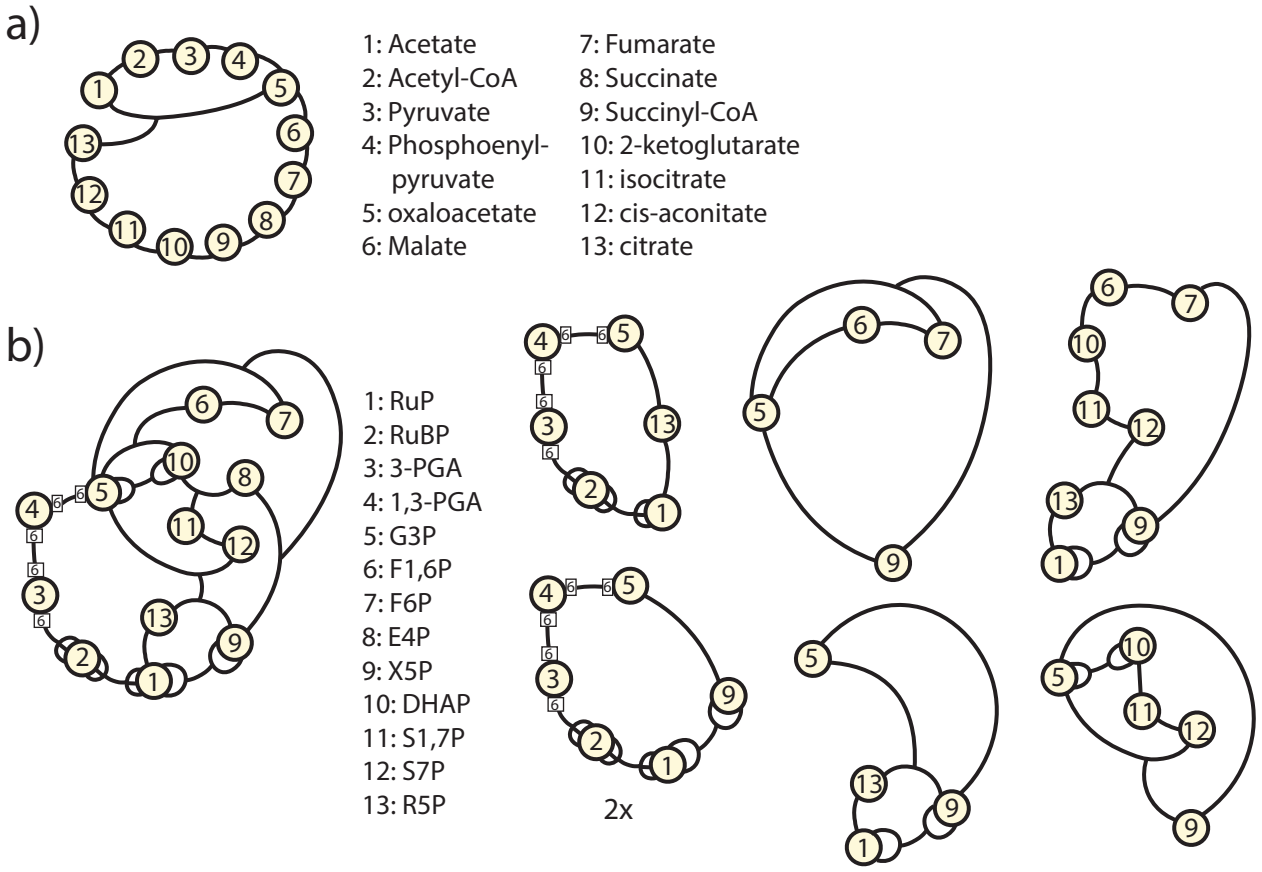

Figure S4: a) Autocatalysts in the Reverse Krebbs cycle, which yield a Type II core. b) Autocatalysts in the Calvin Cycle. We find 3 cores of Type I (2 equivalent, up to the choice of reaction to link 5 and 9), and 4 cores of Type II.

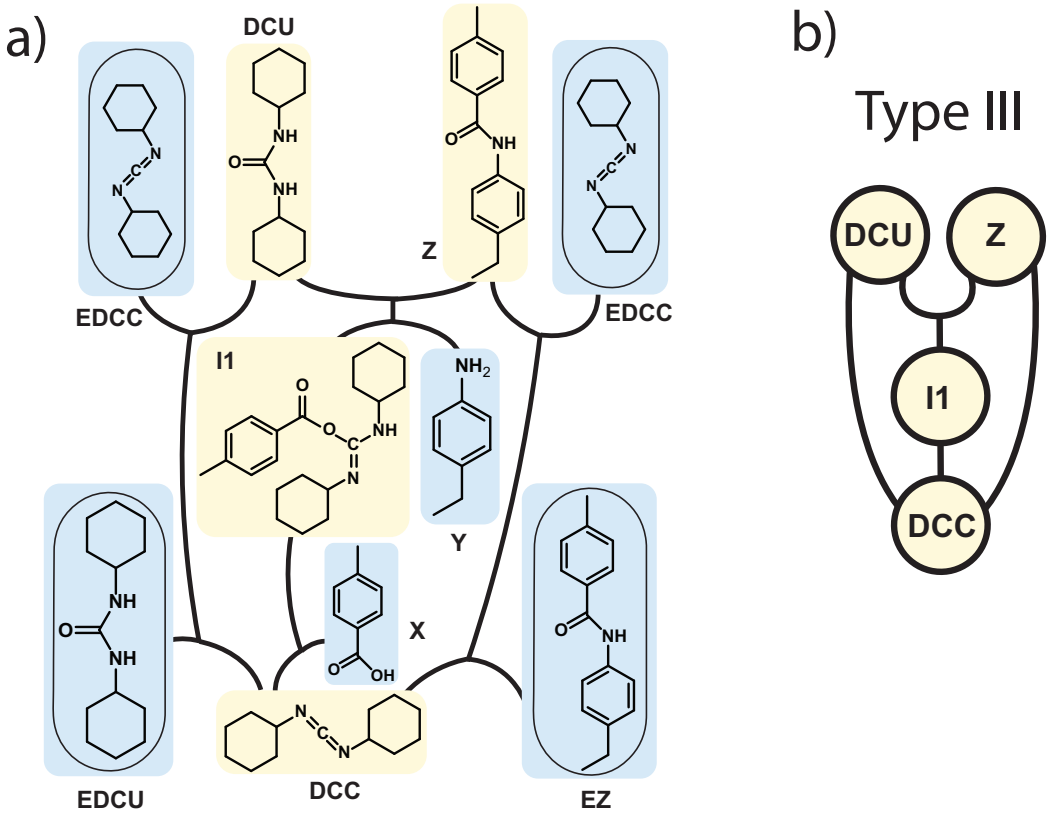

Figure S5: a) Chemical network for the cavitand-amplification network. blue: feedstock compounds and waste. beige: autocatalysts. b) Type III autocatalytic core for the cavitand-amplification network.

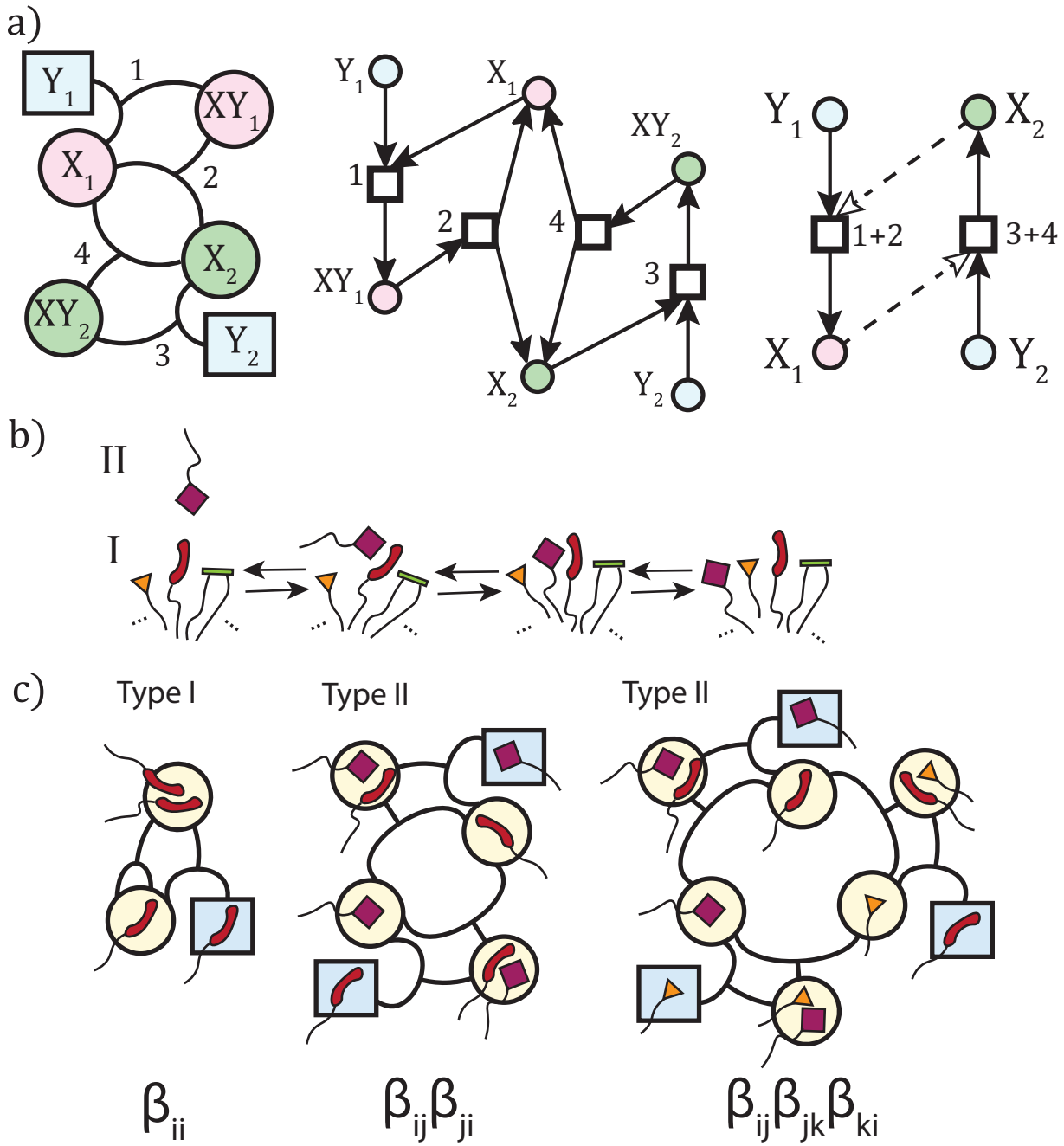

Figure S6: a) A Type II autocatalytic network with its feedstock compounds (or *Food set*), as encountered in GARD. Colored nodes highlight two distinct allocatalytic cycles that yield an autocatalytic cycle when combined. b) The same network, in a bipartite graph representation used in the RAF sets formalism. Specifying the mechanism in terms of uncatalyzed reaction steps removes the RAF property. c) A coarse-grained representation, where allocatalysts in the same allocatalytic cycle are represented by a single species, and each allocatalytic cycle has been replaced with a dashed line, to indicate the net reaction being catalyzed. b. A catalytic incorporation mechanism: the red (telephone shape) amphiphile forms a complex with the purple (square shape) amphiphile, which mediates its incorporation in a micelle or vesicle. c. Autocatalytic networks of co-assembling amphiphiles. Amphiphiles in square nodes are reservoir species, those in circular nodes represent amphiphiles in a micelle or membrane. Diagonal terms of the catalytic matrix encode Type I autocatalysis, i.e. direct self-incorporation. All other autocatalysis (cross-incorporation) is of Type II: sequences of nonoverlapping allocatalytic cycles.

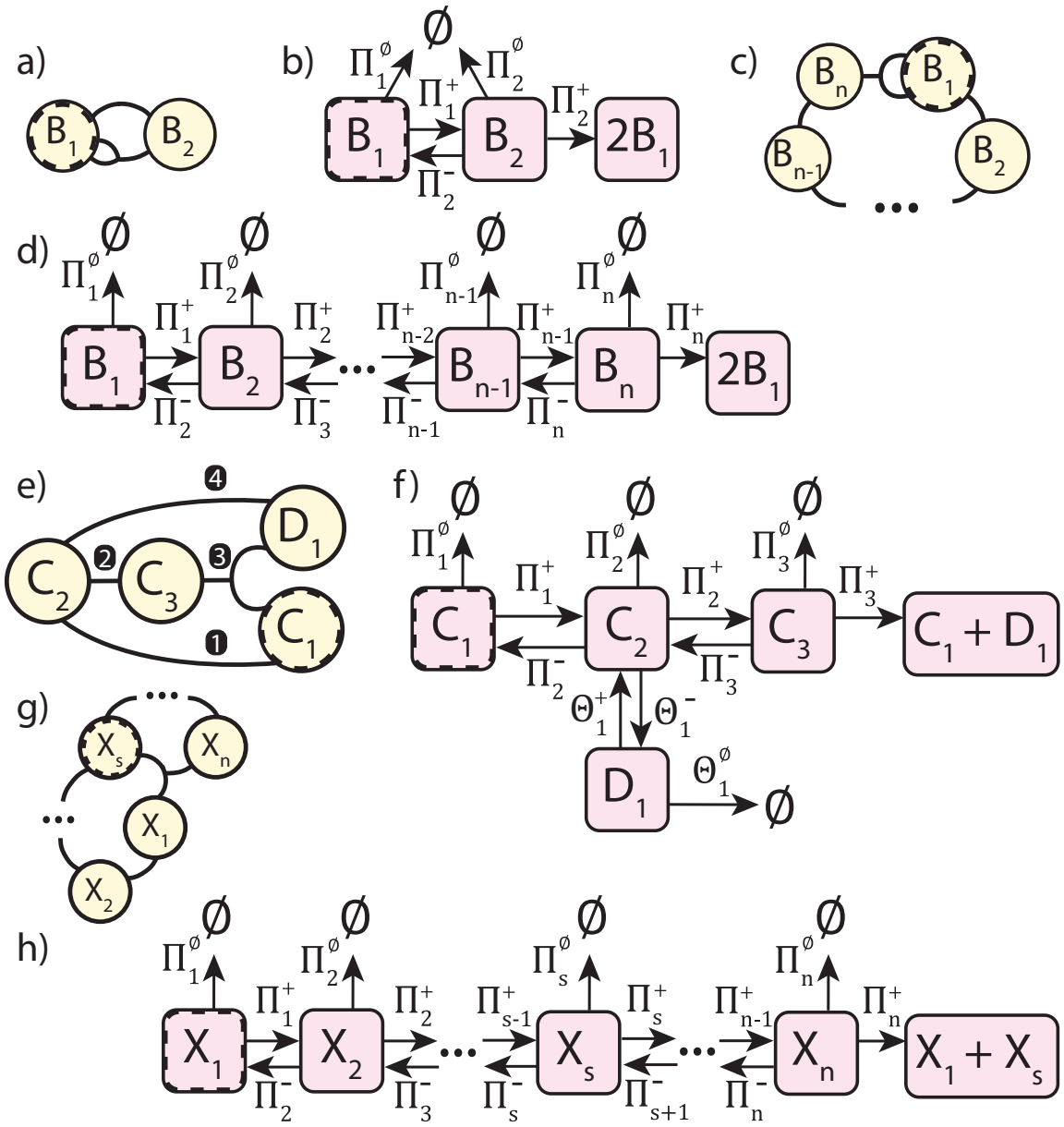

Figure S7: a) A Type I subnetwork. B_1 and B_2 reversibly interconvert, B_2 can also irreversibly form two B_1, marked by the forked edge. b) transition network, Π_k^+ denotes the probability that, starting at B_k, the next transition will be a step forward in the cycle, Π_k^- a step backward, and $\Pi_k^\emptyset$ a degradation. Fragmentation (yielding $2B_1$) and degradation ($\emptyset$) are absorbing states, attained with probabilities p_c and $1-p_c$ respectively. c) Autocatalytic core for a Type I cycle with n nodes. d) Transition network for c e) schematic for an autocatalytic core for a Type III cycle. f) Transition network for e) g) General schematic for an autocatalytic core for a Type II cycle with a single fragmentation step. h) Transition network for g.

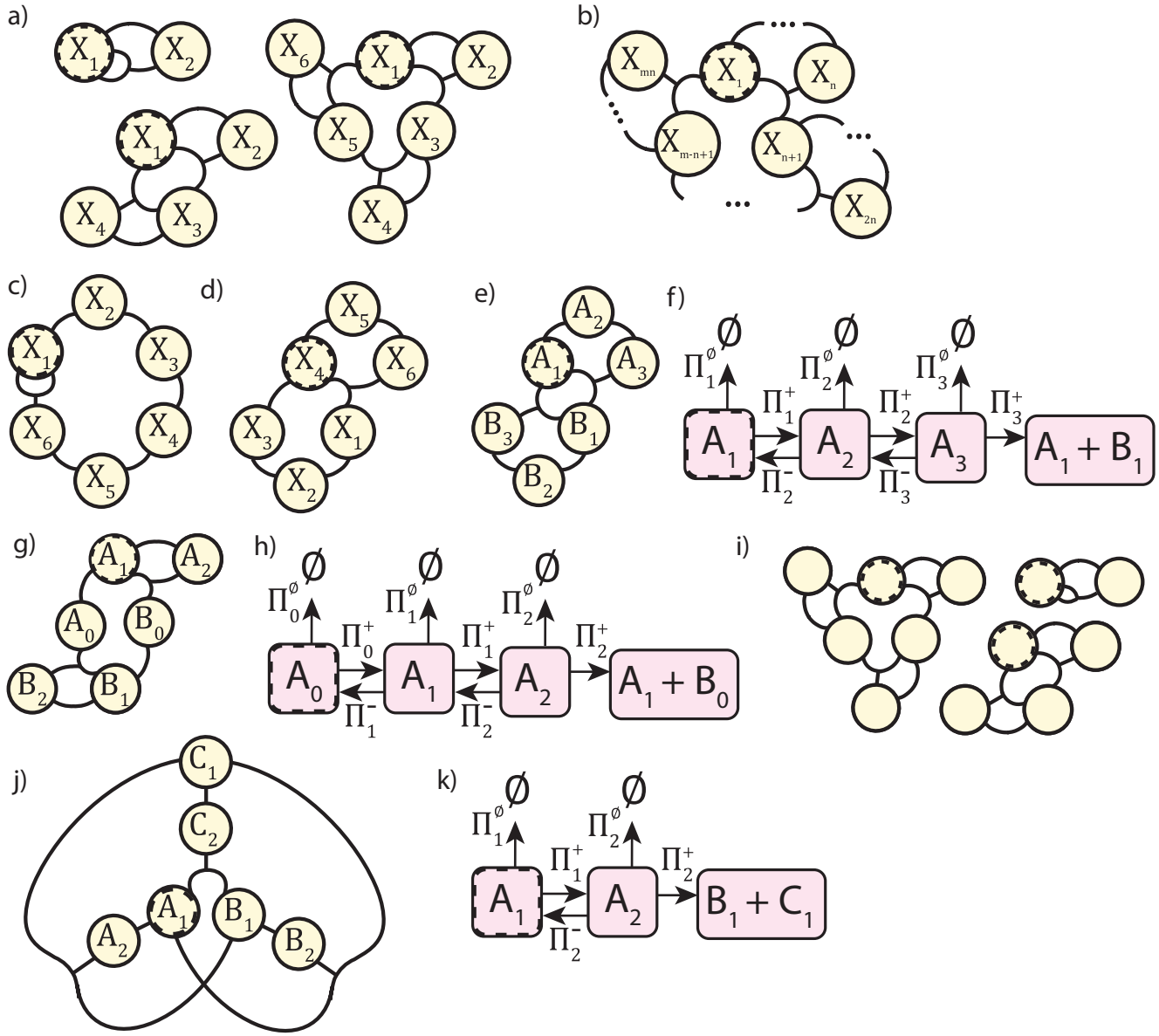

Figure S8: a) three motifs with the same $P_{ex}(\mathrm{A_k})$, when the transition network follows the same symmetry as the network structure. b) A more general case: a multiple of allocatalytic cycles of size n. c) 6-membered Type I cycle. d) a 6-membered Type II cycle. e) a 6-membered Type III cycle. f) Transition network for one of the two allocatalytic cycles. g) a 6-membered Type III cycle. The transition of a precursor to allocatalyst is not mediated by an allocatalytic cycle, and hence this is not a RAF. h) Transition network for g). i) a trio of symmetric analogues. j) a Type V autocatalytic core. k) Transition network for j.

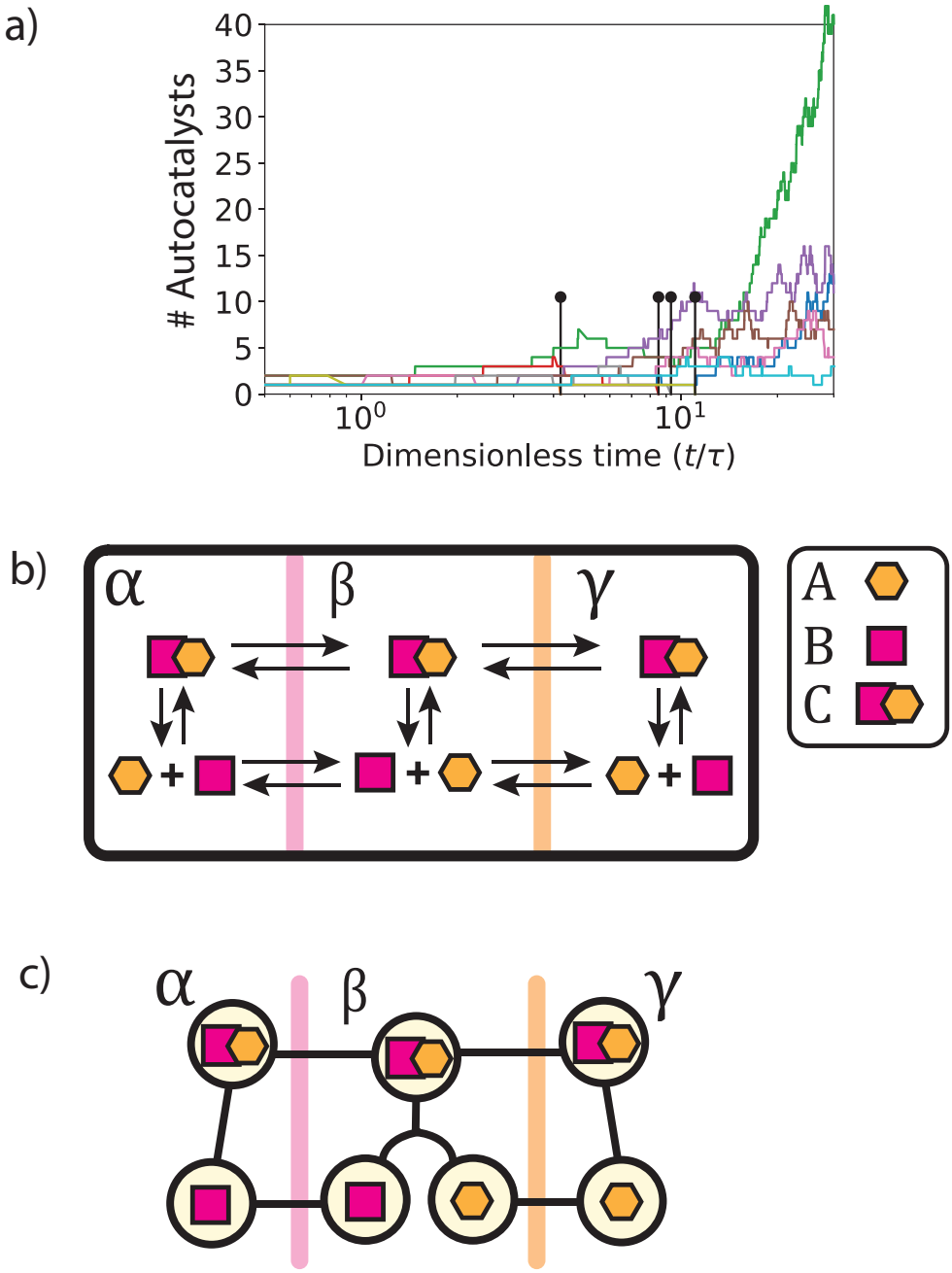

Figure S9: a) Simulation using Gillespie's Algorithm. Each of the 10 colored lines represents the sum of autocatalysts ($\sum_X N_X$) over time in a single run. Extinction events ($\sum_X N_X = 0$) are marked with a black pin at their corresponding timepoint. b) Three-compartment network with a single bimolecular chemical reaction A + B $\rightleftharpoons$ C. The pink wall separating α from β is permeable to B and C. The orange wall separating β from γ is permeable to A and C. c) Autocatalytic core. The use of three compartment allows to construct the reactions to complete the cycle (the ears) directly from the reproduction step.