Self-generating autocatalytic networks: structural results, algorithms, and their relevance to evolutionary processes

Daniel Huson[1], Joana C. Xavier[2,3], Allen Rodrigo[4], Mike Steel[5]

[1]Institute for Bioinformatics and Medical Informatics, Tübingen University, Tübingen, Germany.
[2]Dayhoff Labs, London, UK.
[3]Department of Chemistry, Imperial College, London, UK.
[4]School of Biological Sciences, University of Auckland, Auckland, New Zealand.
[5]Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand.

Abstract

The concept of an autocatalytic network of reactions that can form and persist, starting from just an available food source, has been formalised by the notion of a Reflexively-Autocatalytic and Food generated (RAF) set. The theory and algorithmic results concerning RAFs have been applied to a range of settings, from metabolic questions arising at the origin of life, to ecological networks, and cognitive models in cultural evolution. In this paper, we present new structural and algorithmic results concerning RAF sets, the identification of irreducible RAFs, and determining when certain reactions must occur before other reactions in a RAF (or, more generally, any F-generated set). We also describe how RAF theory extends to allow certain reactions to not require catalysis, and/or to require only catalysts that are not present in the food set, as well as to find minimal sets that are both a RAF and produce a given subset of elements. In all cases, the theory leads to efficient (polynomial-time) algorithms. We also briefly illustrate applications of RAF theory to two evolutionary processes in biology, namely, early metabolism and modelling the evolution of holobionts.

Keywords: autocatalytic network, RAFs, algorithms, metabolic network, holobionts

1 Introduction

A central property of the chemistry of living systems is that they combine two basic features: (i) the ability to survive on an ambient food source and (ii) each biochemical reaction in the system requires only reactants and a catalyst that are provided by other reactions in the system (or are present in the food set). The notion of a self-sustaining 'collectively autocatalytic set' captures these basic features, and their study was pioneered by Stuart Kauffman [22, 23]. By investigating a simple binary polymer model, Kauffman showed that collectively autocatalytic sets invariably emerge once the network of polymers becomes sufficiently large.

The notion of a collectively autocatalytic set was subsequently formalized more precisely as a 'reflexively autocatalytic and food-generated' (RAF) set (defined shortly). RAF sets (RAFs) are related to, but somewhat different from Robert Rosen's (M, R) systems [21], 'organizations' in chemical organization theory [6, 18], and recent models of autocatalytic networks applied to the setting of the origin-of-life [3, 34].

The application of RAFs has expanded beyond polymer models to the analysis of real living systems (e.g., the metabolic network of prebiotic systems [45, 46]), as well as simple autocatalytic networks that have been constructed in laboratory studies, either from RNA molecules [15] or from peptides [14]. They have also been discussed further in modelling the origin of life (see e.g., [7, 43]).

The generality of RAF theory also means that a 'reaction' need not refer specifically to a chemical reaction, but to any process in which certain items are combined and transformed into new items, and where similar items facilitate (or catalyse) the process without being used up in the process. This has led to the application of RAF theory to processes beyond biochemistry, such as cognitive modelling in cultural evolution [8, 9], ecology [10, 12], and economics [12].

Two important features of the RAF approach are the degree of generality RAFs allow, and the fact that very large systems can be analysed precisely by fast algorithms.

In this paper, we describe further extensions and applications of RAF theory, as follow:

- An extension of the RAF approach that provides a unified handling of complex catalysis, leading to new mathematical results [Sections 2.2, 3];
- A method to identify the extent to which catalysts within a RAF are generated as products of reactions of the RAF [Section 3.2];
- Techniques to help identify all the irreducible (minimal) RAFs within a system [Section 4] and to describe a RAF in terms of irreducible RAFs [Section 4.2];
- Algorithms to determine which reactions need to come 'early' in a RAF [Section 5];
- Applications of RAF theory to model the evolution of holobionts [Section 6] and to investigate primitive metabolic networks.

2 Catalytic reaction systems

2.1 Reaction Systems

A *reaction system* is a pair (X, R) consisting of a finite nonempty set X of *elements* (e.g., molecule types) and a finite set R of *reactions*. Here a *reaction* $r \in R$ refers to an ordered pair (A, B) where A and B are multisets of elements from X. We will write $r : a_1 + \cdots + a_k \to b_1 + \cdots + b_l$ to denote the reaction that has reactants $\{a_1, \ldots, a_k\}$, and products $\{b_1, \ldots, b_l\}$[1]. We let $\rho(r)$ denote the set corresponding to A (i.e., ignoring multiplicities), and $\pi(r)$ denote the set corresponding to B (ignoring multiplicities); it is assumed implicitly that $\rho(r), \pi(r) \neq \emptyset$. For a subset R' of R, it is convenient to let $\pi(R') = \bigcup_{r \in R'} \pi(r)$ denote the set of the products of the reactions in R'.

Next, consider a reaction system (X, R) together with a particular subset F of X. The set F can be interpreted as a set of elements that are freely available to the system; accordingly, F is referred to as a *food set*. A subset R' is *F-generated* if the reactions in R' can be placed in some linear order, say $r_1, r_2, \ldots, r_k$, such that the following property holds: for $\rho(r_1) \subseteq F$ and for all values of j between 2 and k, we have $\rho(r_j) \subseteq F \cup \pi(\{r_1, \ldots, r_{j-1}\})$. In other words, the reactions in $\mathcal{R}'$ are F-generated if they can proceed in some order so that the reactant(s) of each reaction are available by the time they are first required. We call such an ordered sequence of R' an *admissible ordering*. Since there are $k!$ ways to order k reactions, it may not be immediately obvious that the F-generated condition can be verified in polynomial time; however, there is a simple way to do so, as we now describe.

We first recall some further terminology. Given a subset R' of reactions R, a subset W of X is said to be *R'-closed* precisely when each reaction $r \in R'$ that has all its reactant(s) in W also has all its the product(s) in W (i.e. $r \in R', \rho(r) \subseteq W \Rightarrow \pi(r) \subseteq W$). The union of two closed sets need not be closed; nevertheless, given any nonempty subset W_0 of X there is a unique minimal R'-closed set containing W_0, denoted $\mathrm{cl}_{R'}(W_0)$. This can be computed in polynomial time in the size of the system by constructing a nested increasing sequence of subsets of the elements $W_0 \subset W_1, \ldots \subset W_k$ where:

$$W_{i+1} = W_i \cup \{x \in X : \exists r \in R' : \rho(r) \subseteq W_i, x \in \pi(r)\},$$

for $i \geq 0$, and by letting W_k denote the terminal set in this sequence (i.e. k is the first value of i for which $W_i = W_{i+1}$). In the lemma that follows (from [39]) we will take $W_0 = F$ and consider $\mathrm{cl}_{R'}(F)$.

Lemma 1. *In any reaction system (X, R) with a food set $F \subseteq X$, and a nonempty subset R' of R, the following three conditions are equivalent:*

(i) R' is F-generated (i.e., R' has an admissible ordering);
(ii) $\rho(r) \subseteq \mathrm{cl}_{R'}(F)$ for all $r \in R'$;
(iii) $\mathrm{cl}_{R'}(F) = F \cup \pi(R')$.

[1] In applications, a bidirectional reaction is generally regarded as a pair of reactions (forward and backward) with the same catalysis assignment.

2.2 Catalytic reaction systems (allowing complex catalysis)

A *catalytic reaction system* (CRS) is a reaction system with a food set (X, R, F) together with a subset χ of $2^X \times R$ where $(U, r) \in \chi$. Thus χ is a collection of pairs (U, r) where $U \subseteq X$ and $r \in R$. For $(U, r) \in \chi$ we refer to U as a *catalyst set* for r. This is a generalisation of earlier treatments in which χ consisted of a subset of $X \times R$ (i.e. simple catalysis by single elements). Our extension here to this more general catalysis framework allows for complex (i.e. conjunctive) catalysis rules, where catalysts of a reaction may require the presence of two or more elements of X (e.g., cofactors of enzymes), as described in Section 4.1 of [40]. However, this earlier treatment of complex catalysis rules required the introduction of fictitious new reactions and elements to the original CRS. Here, our more direct approach allows both simple and complex catalysis rules that require no additional reactions or elements to be introduced. It also permits the further option that particular uncatalysed reactions can appear in an autocatalytic system (since the definition of χ allows $(\emptyset, r) \in \chi$), thereby addressing a recent concern discussed in Section 3.1.

We denote a CRS by writing $\mathcal{Q} = (X, R, \chi, F)$, and we let $|\mathcal{Q}| = |X| + |R| + |\chi|$ denote the *size* of $\mathcal{Q}$ and we write

$$r : a_1 + \cdots + a_k[U_1, \ldots, U_m] \to b_1 + \cdots + b_l$$

to denote the reaction that has reactants $\{a_1, \ldots, a_k\}$, products $\{b_1, \ldots, b_l\}$ and catalyst sets $\{U_1, \ldots, U_m\}$, where $(U_i, r) \in \chi$ for all i. When U_i is a singleton set (say $\{c_i\}$), we will often write c_i in place of $\{c_i\}$ in our example systems.

2.3 RAFs

Given a CRS $\mathcal{Q} = (X, R, \chi, F)$, a subset R' of R is said to be *Reflexively-Autocatalytic and Food generated* (more briefly a *RAF*) if R' is nonempty and if for each $r \in R'$, the reactants of r and at least one catalyst set U for r (as specified by χ) are subsets of $\mathrm{cl}_{R'}(F)$.

An equivalent definition for a nonempty set R' to be a RAF for $\mathcal{Q}$ is that R' is F-generated, and each reaction $r \in R'$ has a catalyst set U that is a subset of $F \cup \pi(R')$. A further equivalent definition is the following:

> R' can be ordered $r_1, r_2, \ldots, r_k$ so that for each $i \geq 1$, the reactants of r_i are present in X_i, where $X_1 = F$ and $X_i = F \cup \pi(\{r_1, \ldots, r_{i-1}\})$, and at least one catalyst set U of r_i is a subset of X_k.

If a CRS $\mathcal{Q}$ has a RAF, then it has a unique maximal RAF (which is the union of all the RAFs for $\mathcal{Q}$), which is denoted maxRAF($\mathcal{Q}$). A RAF R for $\mathcal{Q}$ is said to be an iRAF (irreducible RAF) if $R \setminus \{r\}$ is not a RAF for $\mathcal{Q}$ and contains no RAF for $\mathcal{Q}$. A stronger notion than a RAF is a *CAF* (constructively-autocatalytic and F-generated set) where the third equivalent definition of a RAF (above) is strengthened to 'and at least catalyst set U of r_i is a subset of X_i' (rather than 'of X_k'). In other words, R' is a CAF if it has an admissible ordering in which at least one catalyst set has each of its elements already present in the food set or produced by an earlier reaction in the ordering. Every CAF is also a RAF, but the converse containment does

not hold. Although RAFs and CAFs appear to very similar concepts, they exhibit quite different properties. For example, if a CRS $\mathcal{Q}$ has a CAF, then this CAF must contain a reaction r for which all the reactants of r and at least one catalyst of r lie in F, in which case $\{r\}$ is itself a CAF of size 1. By contrast, a large RAF need not contain any 'small' RAF within it. Moreover, theoretical and simulation studies on polymer systems reveal that the level of catalysis required for a CAF to be present is exponentially higher than that required for a RAF [30], and in real biochemical systems that have been studied (e.g., [37, 45]), the maxRAF is generally not a CAF. Thus, in this paper, we focus on the more general notion of a RAF.

Examples: A simple example to illustrate these notions is provided by the following CRS where $X = \{a, b, c, d, e\}$, $F = \{a, b\}$ and $R = \{r_1, r_2, r_3, r_4\}$ and χ are as follows:

$$r_1 : a + a\ [\{c, d\}, e] \to c \qquad r_3 : b + b\ [\ ] \to e$$
$$r_2 : b + c\ [\emptyset] \to d + d \qquad r_4 : a + e\ [a] \to b$$

This system has $\{r_1, r_2\}$ as its maxRAF, which is also the unique iRAF (thus no other subset of reactions forms a RAF). No CAF is present. Note that since $(\emptyset, r_2) \in \chi$ the reaction r_2 is permitted to be part of a RAF, whereas for r_3, χ contains no subset of X (including $\emptyset$) and so r_3 cannot be part of any RAF.

A second example of a RAF arising in an experimental system is provided in Fig. 1

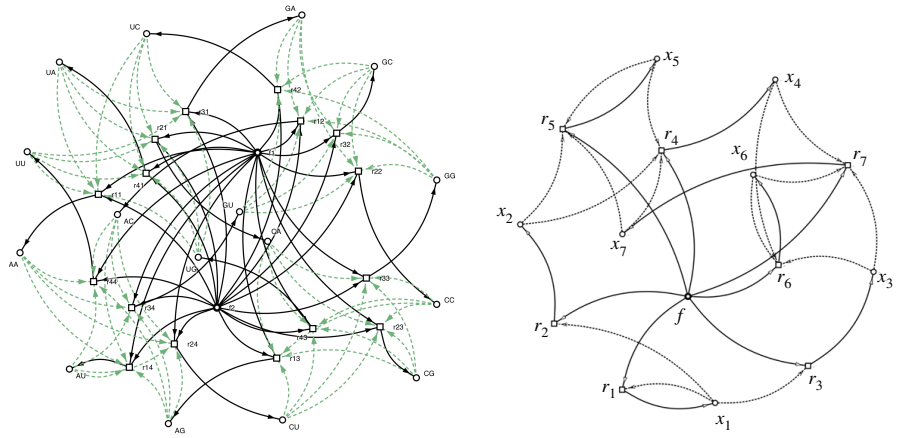

Fig. 1 *Left:* An experimental RNA system described in [42] involving 16 reactions, and 18 molecule types (including a food set of size 2, which form the reactants of each reaction). The reactions are denoted by squares, elements (in this case, RNA replicators) are denoted by circles, reaction pathways are indicated by solid arrows, and catalysis is indicated by dotted arrows. This system forms a RAF [15]. *Right:* A subset of seven of the reactions from the full set (with the two food elements combined into a single element f), which forms a RAF. This RAF, analysed in [15] contains within it 67 other subsets that form RAFs including four irreducible RAFs (iRAFs). This RAF is not a CAF (nor does it contain one). Figures produced by *CatlyNet* [20].

3 Mathematical aspects of RAFs

We now recall the concept of the maxRAF operator φ (from [41], Section 3). For any subset R' of R, let $\mathcal{Q}|R'$ be the CRS (X, R', χ', F) where χ' is the restriction of χ to $2^X \times R'$, and let $\varphi : 2^R \to 2^R$ be the function defined as follows:

$$\varphi(R') = \begin{cases} \text{maxRAF}(\mathcal{Q}|R'), & \text{if } \mathcal{Q}|R' \text{ has a RAF;} \\ \emptyset, & \text{otherwise.} \end{cases} \tag{1}$$

The function $\varphi(R')$ can be determined in a computationally efficient way, as follows:

$$\varphi(R') = \bigcap_{i \geq 1} R_i,$$

where $R_1 = R'$ and for $i > 1$,

$$R_i = \{r \in R_{i-1} : \rho(r) \subseteq \text{cl}_{R_{i-1}}(F) \text{ and } (U, r) \in \chi \text{ for some } U \subseteq \text{cl}_{R_{i-1}}(F)\}. \tag{2}$$

Notice that the sets R_i defined in Eqn. (2) form a decreasing nested sequence, and so $\varphi(R')$ is precisely the set R_j for the first value of $j \geq 1$ for which $R_{j+1} = R_j$. In particular, a nonempty subset R' of R is a RAF if and only if $\varphi(R') = R'$. Moreover, $\varphi(R)$ is precisely the maxRAF of $\mathcal{Q}$ (if it exists), or is the emptyset otherwise.

A number of interesting algebraic (semigroup) properties of the map φ have been explored recently in [27] (see also [28], which considers a a more general notion than a RAF, which corresponds to what have been called 'pseudo-RAFs' in the RAF literature).

The map φ is also an example of the more general notion of an interior operator (on 2^R), and several basic results in RAF theory can be derived by this property alone. To explain this further, an arbitrary function $\psi : 2^Y \to 2^Y$ is said to be an *interior operator*[2] if it satisfies the following three properties (nesting, monotonicity, and idempotence) for all subsets A, A' of Y:

- $\psi(A) \subseteq A$,
- $A \subseteq A' \Rightarrow \psi(A) \subseteq \psi(A')$, and
- $\psi(\psi(A)) = \psi(A)$.

It was recently shown that if Y is a finite set of size at least 12, then there exist interior operator on Y that cannot be represented as a maxRAF operator of a suitably chosen CRS (X, R_Y, χ, F) where R_Y is a set of reactions bijectively associated with Y and where $\chi \subseteq X \times R_Y$ (i.e. the original CRS setting which does not allow complex catalysis rules) [38]. Thus some generic properties of RAFs in this setting cannot be established by using the interior operator properties alone. However, if one allows complex catalysis rules, we have the following contrasting result (further details and a proof are provided in the Appendix).

[2] The term 'interior operator' comes from topology, since the function that assigns to any subspace S of a topological space the topological interior of S satisfies the three properties described.

Proposition 1. *For any finite set Y, any interior operator ψ on 2^Y can be represented as the maxRAF operator φ for a suitably chosen CRS $\mathcal{Q} = (X, R_Y, \chi, F)$, where R_Y is bijectively associated with Y and $\chi \subseteq 2^X \times R_Y$ permits complex catalysis rules. Moreover, R_Y can be chosen so that each of its reactions has a single reactant f which comprises the food set F and a single distinct product.*

Proposition 1 has an interesting consequence for the structure of RAFs in any CRS. Specifically, it implies that it is possible to simplify the reactions and the food set, while preserving the number, sizes, and containment structure of the RAFs (albeit at the price of possibly increasing substantially the number and complexity of catalysts for the reactions). More precisely, we have the following result (obtained from Proposition 1 by taking ψ to be the maxRAF operator of $\mathcal{Q}$).

Corollary 1. *For any CRS $\mathcal{Q} = (X, R, \chi, F)$, there is a matching CRS $\mathcal{Q}' = (X', R', \chi', F')$, where (i) each of the reactions in R' has a single reactant f which comprises the food set F, and a single distinct product and (ii) there is a canonical bijective correspondence between R and R' that induces a bijection between the RAFs of $\mathcal{Q}$ and the RAFs of $\mathcal{Q}'$.*

A worked example to illustrate this result is provided in the Appendix.

3.1 The interpretation of catalysis and autocatalysis in RAFs

Note that if $(\emptyset, r) \in \chi$, then r does not require any element of X for its catalysis, and so RAFs under our more general definition can contain uncatalysed reactions in settings where this is appropriate. This is particularly relevant, as a number of papers (e.g., [1, 13, 26, 25]) have pointed to the restrictive nature of RAFs in requiring that all reactions in the RAF must be catalysed. For example, the authors of [1] stated:

> "Here, the assumption that all reactions are catalysed appears very unrealistic. Unfortunately, the algorithms for recognizing RAFs [...] do not seem to generalise to arbitrary networks composed of non-catalysed reactions."

and in [13] the author states:

> "The requirement that every reaction be catalyzed by another molecule seems too strong when we are dealing with simple metabolisms in the earliest forms of life. In modern organisms, almost every reaction of small-molecule metabolites is catalyzed by enzymes. However, prior to the existence of RNA and proteins, there were no enzymes, so we need a theory that deals with the reactions of small molecules without insisting that the reactions be catalyzed."

In fact, in previous applications of RAF theory to the origins of metabolism [45, 46] a fictional catalyst 'Spontaneous' was assigned to reactions known to occur uncatalysed. This catalyst was added to the food set in all simulations. Also, prior to the advent of genetic coding and enzymes, catalysis must have existed in the small-molecule world as well. Small molecules are increasingly being shown to catalyse multiple reactions in the absence of enzymes [5, 31].

RAF theory also treats 'catalysis' in a general way and this allows for efficient graph-theoretic algorithms, which apply independently of any detailed kinetic (or even

stoichiometric) considerations. This generality also allows for applications in a variety of areas outside chemistry. Essentially, we regard a catalyst as any element that facilitates, speeds-up, or synchronises a reaction without taking part in the reaction itself (as a reactant).

For example, in economic applications, a factory facilitates (i.e. catalyses) the production of items from the incoming raw materials, but is not itself consumed by that process. In cognitive modelling [8], a reaction that combines ideas to form a new idea could be enhanced (i.e. catalysed) by a need, thought, memory, or stimuli. In ecology [10], a catalyst is a species that enables some other interaction in an ecological network. In biochemistry, it can also be helpful to treat catalysis in a quite general way; for instance, the formation of a lipid boundary to encompass a primitive metabolism can be viewed as a sequence of reactions which, once complete, forms a structural element (a complete lipid membrane) that catalyses all the reactions within the newly formed protocell (since the system within it no longer disperses) [16].

Moreover, in biochemistry, food elements can be direct catalysts for multiple reactions, as is the case with universally-essential metal ions. The generality of RAF theory allows for this, in contrast to some other models of network autocatalysis.

It is tempting to treat a catalyst x of a reaction r as simply an additional reactant and product of r (i.e., adding x to both sides of the reaction, as is sometimes used in chemical notation), thereby reducing the notion of a RAF to simply an F-generated subset of the resulting (modified) reaction system. In other words, the catalysed reaction $x + y[c] \to z$ (where c is a catalyst) might be viewed as $x + y + c \to z + c$. However, this misses an important distinction. Namely, in RAF theory, it is assumed that a reaction r may proceed (at a slow rate intially) provided that all its reactants are available even if no catalyst is initially available; however r will subsequently be part of a RAF set, provided that a catalyst of r is (eventually) produced by at least one other reaction in the RAF. Indeed, it is entirely possible for a CRS to contain a RAF yet the modified reaction system (adding catalysts to both sides of a reaction) might have no F-generated subset (indeed, this has been observed in real biochemical networks [15, 37, 45]).

On another note, alternative catalysts exist for the same reaction in multiple systems, and particularly in biochemistry [45]. If explicitly added to both sides, these alternatives would require duplicates of the same reaction to be created. This would make the CRS unrealistically complex and much more difficult to analyse. Misunderstandings concerning the related notion of 'autocatalysis' have been discussed in [17] and [35].

3.2 'Strictly autocatalytic' RAFs and minimal RAFs that generate a given element

If a CRS $\mathcal{Q}$ has the property that each reaction is catalysed by at least one element of the food set, then the RAF sets for $\mathcal{Q}$ coincide precisely with the F-generated sets. More generally, for any CRS, it is possible for a RAF set to have the property that some (or all) of its reactions are catalysed by elements of the food set, which renders the notion of 'autocatalytic' less applicable (though not entirely, since each reaction in the RAF might still also be catalysed by at least one product of the other reactions in

the RAF). To formalise this notion further, we introduce a new definition and describe a result that characterises the condition which captures a more focused notion of an 'autocatalytic' RAF.

Given a CRS $Q = (X, R, \chi, F)$, we say that a subset R' of R is *strictly autocatalytic* if each reaction in R' has at least one catalyst that involves one or or more products of reactions of R' and is not a subset of the food set. We say R' is a strictly autocatalytic RAF for $\mathcal{Q}$ if R' is a RAF and is strictly autocatalytic.

Example: Consider the following catalytic reaction system:

$$\begin{aligned} r_1 &: a + b\ [\{a, d\}] \to e \\ r_2 &: b + c\ [\{a, b\}, e] \to d \\ r_3 &: d\ [\{a, b\}] \to c \end{aligned}$$

where $X = \{a, b, c, d, e\}$ and $F = \{a, b, c\}$. Then, $\{r_2\}, \{r_1, r_2\}, \{r_2, r_3\}$ and $\{r_1, r_2, r_3\}$ are RAFs, however $\{r_1, r_2\}$ is the only strictly autocatalytic RAF. □

Let $\hat{R}_i, i \geq 1$ be the nested decreasing sequence of subsets of R defined as follows: $\hat{R}_1 = R$, and for $i > 1$, consider the following modification of Eqn. (2):

$$\hat{R}_i = \{r \in \hat{R}_{i-1} : \rho(r) \subseteq \mathrm{cl}_{\hat{R}_{i-1}}(F) \text{ and } (U, r) \in \chi \text{ for some } U \subseteq \mathrm{cl}_{\hat{R}_{i-1}}(F) \text{ and } U \not\subseteq F\}. \tag{3}$$

The following result shows that there is a polynomial-time algorithm (in $|\mathcal{Q}|$) to determine whether or not $\mathcal{Q}$ has a strictly autocatalytic RAF and, if so, to construct a maximal one. The proof is provided in the Appendix.

Proposition 2. *Let $\mathcal{Q} = (X, R, \chi, F)$ be a CRS. Then $\mathcal{Q}$ has a strictly autocatalytic RAF if and only if $\bigcap_{i \geq 1} \hat{R}_i \neq \emptyset$, in which case, $\bigcap_{i \geq 1} \hat{R}_i$ is the unique maximal strictly autocatalytic RAF for $\mathcal{Q}$.*

Application: We investigated the large primitive metabolic data-set used in [45] based on 6039 reactions and a food set of size 68, which resulted in a maxRAF of size 1357. Applying the strictly autocatalytic RAF algorithm determined that no such RAF exists. This, in turn, implies that in any RAF for this system, one or more elements in the food set play a pivotal role in catalysing some reaction(s). This is consistent with the aforementioned essentiality of metals in biocatalysis of central metabolism. At the same time, this result reveals that there are unknown prebiotic routes (or unknown non-enzymatic catalysts) to the production of organic cofactors which must be investigated in the laboratory.

3.3 A minimal RAF that generates given element(s)

Next, suppose that a CRS $\mathcal{Q} = (X, R, \chi, F)$ and that an element $x \in X \setminus F$ is produced by some reaction in maxRAF($\mathcal{Q}$). A relevant question in certain applications is to find a minimal subset R' of reactions within maxRAF($\mathcal{Q}$) that is both a RAF and produces one or more particular elements of interest (e.g., it plays a key role in a metabolic network, such as ATP or NAD, the universal currencies of energy and redox power in metabolism, respectively). In other words, R' is a RAF that produces x, and every proper subset of R' either fails to be a RAF or fails to produce all of the specified elements. It turns out that finding such a minimal set has a polynomial-time solution, as we now show (the proof is provided in the Appendix).

Proposition 3. *Suppose that one or more particular elements $x_1, \ldots, x_k \in X \setminus F$ are produced by some reaction in* maxRAF($\mathcal{Q}$)*. Let $\mathcal{Q}^{+\{x_1,\ldots,x_k\}}$ be the CRS obtained from $\mathcal{Q}$ by replacing each catalyst $(U, r) \in \chi$ with $(U \cup \{x_1, \ldots, x_k\}, r)$. Then the collection of minimal subsets of* maxRAF($\mathcal{Q}$) *that are simultaneously RAFs of $\mathcal{Q}$ and produce each of the elements $x_1, \ldots, x_k$ are precisely the iRAFs of $\mathcal{Q}^{+\{x_1,\ldots,x_k\}}$.*

As an example with the 16-reaction CRS shown in Fig. 1, for each element x of $X \setminus F$ the minimal subset that is both a RAF and generates x has size either 1 or 2.

4 Identifying all minimal RAFs (iRAFs) in a CRS

Let $\mathcal{Q} = (X, R, \chi, F)$ be a fixed CRS that contains a RAF. Recall that a RAF R for $\mathcal{Q}$ is said to be an iRAF (irreducible RAF) if $R \setminus \{r\}$ is not a RAF for $\mathcal{Q}$ and contains no RAF for $\mathcal{Q}$. Every RAF of a CRS contains an iRAF, and there is a simple (polynomial-time in $|\mathcal{Q}|$) algorithm for finding one or more iRAFs within any given RAF; however, the number of iRAFs for $\mathcal{Q}$ can grow exponentially with $|\mathcal{Q}|$ (Theorem 1 of [19]). Moreover, determining the size of a *smallest* iRAF is known to be NP-hard [39]. Nevertheless, there is a simple (and polynomial-time) algorithm to determine whether $\mathcal{Q}$ has just one iRAF; we simply ask whether the set

$$Core(\mathcal{Q}) = \{r \in R : \varphi(R \setminus \{r\}) = \emptyset\}$$

is a RAF. Then it can be shown (see [41] Theorem 4.1) that this set is a RAF if and only if $\mathcal{Q}$ has a unique iRAF. In that case, $Core(\mathcal{Q})$ is the unique iRAF for $\mathcal{Q}$.[3].

Moreover, it can easily be seen that there is a simple way to test whether or not an iRAF R_1 is the only iRAF for $\mathcal{Q}$; this holds provided that:

$$\varphi(\varphi(R) \setminus \{r\}) = \emptyset, \text{ for all } r \in R_1.$$

This is computationally less-intensive than computing $Core(\mathcal{Q})$, since it involves searching over the reactions that are just in R_1 rather than all of R.

In a similar way, it can be shown that if $\mathcal{Q}$ has (at least) two iRAFs, R_1 and R_2 then these are the only iRAFs for $\mathcal{Q}$ if and only if for all $r \in R_1 \setminus R_2$, we have $\varphi(\varphi(R) \setminus \{r\}) = R_2$, and for all $r \in R_2 \setminus R_1$, we have $\varphi(\varphi(R) \setminus \{r\}) = R_1$.

[3] The set $Core(\mathcal{Q})$ is the set of *essential* reactions in R (eg. in CatlyNet).

A direct extension of this last result to three or more iRAFs is problematic, since although two iRAFs cannot be nested (i.e. neither can be a subset of the other) in the case of three iRAFs, it is possible for each one to be a subset of the union of the two others. An example is provided by the CRS $\mathcal{Q} = (X, R, C, F)$, where $F = \{f\}$, $R = \{r_1, r_2, r_3\}$ and

$$\begin{aligned} r_1 &: f\ [x_1] \to x_2 + x_3 \\ r_2 &: f\ [x_2] \to x_1 + x_3 \\ r_3 &: f\ [x_3] \to x_1 + x_2 \end{aligned}$$

This system has three iRAFs, namely, $\{r_1, r_2\}, \{r_1, r_3\}$, and $\{r_2, r_3\}$, each of which contained in the union of the other two.

However, the following result shows that testing a small number (say, k) of iRAFs is feasible in polynomial time in $|\mathcal{Q}|$ (provided that k is fixed). The result is a slight strengthening of Theorem 2 of [39]; its significance is that it allows one to determine whether a given set of k iRAFs is the set of *all* iRAFs for $\mathcal{Q}$ in polynomial time in $|\mathcal{Q}|$, provided that k is fixed (however, the algorithm is exponential in k). The proof of the following result is given in the Appendix.

Proposition 4. *A given collection $R_1, \ldots, R_k$ of iRAFs for $\mathcal{Q}$ is the set of* all *iRAFs for $\mathcal{Q}$ if and only if the following condition holds: For all choices $(r_1, \ldots, r_k)$ where $r_i \in R_i$, we have $\varphi(\varphi(R) \setminus \{r_1, \ldots, r_k\}) = \emptyset$.*

4.1 An algorithm for finding all iRAFs in a CRS

Proposition 4 assumes that the iRAFs are given in advance (e.g., found by heuristic search); however, the approach can be extended to identify the complete set of iRAFs of $\mathcal{Q}$ even if they are not given in advance.

Thus, when the number of iRAF is small (say, less than some value k), the following algorithm is polynomial-time in $|\mathcal{Q}|$ (but is exponential in k). The algorithm proceeds as follows and when it terminates, it produces the complete list of iRAFs of $\mathcal{Q}$.

- Given $\mathcal{Q} = (X, R, F, C)$, determine the maxRAF; providing that it exists, compute an iRAF, denoted R_1.
- For $i \geq 1$, construct a sequence of iRAFs starting with R_1, as follows.

(i) For each integer j with $1 \leq j \leq i$ and each choice of i reactions $\{r_1, r_2, \ldots, r_i\}$ where $r_j \in R_j$ for each j, compute $\varphi(\varphi(R) \setminus \{r_1, \ldots, r_i\})$. If this set is the empty set for all such choices of $\{r_1, \ldots, r_i\}$, then $R_1, \ldots, R_i$ is the complete set of iRAFs for $\mathcal{Q}$, so the algorithm terminates.

(ii) Otherwise, if $\varphi(\varphi(R) \setminus \{r_1, \ldots, r_i\}) \neq \emptyset$ for some choice of $\{r_1, \ldots, r_i\}$, then compute an iRAF of $\varphi(\varphi(R) \setminus \{r_1, \ldots, r_i\})$ and set R_{i+1} equal to this iRAF, and proceed to step (iii).

(iii) Repeat step (i) and, if necessary, step (ii).

4.2 Describing a RAF in terms of a composition sequence involving iRAFs

Any RAF contains the union of its iRAFs; however, it may be strictly larger than this union. A simple example is provided by the simple catalytic reaction system:

$$r_1 : f\ [y] \to y$$
$$r_2 : y\ [z] \to z$$

where $F = \{f\}$ and $X = \{f, y, z\}$. This system is a RAF but its only iRAF is $\{r_1\}$. In general, a RAF R' is the union of its iRAFs if and only if each reaction in R' is contained in an iRAF.

Nevertheless, any RAF R' in a CRS $\mathcal{Q} = (X, R, \chi, F)$ can be described by a sequence of iRAFs and associated catalytic reaction systems. This is loosely analogous to the description of finite groups in abstract algebra via a 'composition series', in which a group is reduced to the trivial group via quotients that are simple groups (here, the analogue of a group (respectively, a simple group) is a subset of reactions (respectively, an iRAF), and the analogue of a quotient group is the complement of a iRAF in a set of reactions).

To describe this in our setting, suppose we have a CRS, $\mathcal{Q} = (X, R, \chi, F)$ and any nonempty subset $R' \subseteq R$. A *composition sequence* for R' is a nested decreasing sequence of subsets of R', $R_1, \ldots,$ with $R_1 = R'$ and with $R_{i+1} = R_i \setminus \hat{R}_i$ where $\hat{R}_i$ is an iRAF of $\mathcal{Q}_i = (X, R_i, \chi, F \cup \pi(\hat{R}_{i-1}))$ for each $i \geq 1$ for which $\mathcal{Q}_i$ has a (nonempty) RAF. Since the sets R_i form a nested decreasing sequence of sets, we refer to the final distinct set (i.e. $\bigcap_{i \geq 1} R_i$) as the *terminal set* of the sequence.

It is clear that every subset R' of R has a composition sequence, and that this can be constructed in polynomial time in $|\mathcal{Q}|$. Moreover, by definition, the iRAF sets $\hat{R}_i$ are disjoint. In the simple 2-reaction example above, the unique composition sequence for $R' = \{r_1, r_2\}$ is: $\{r_1, r_2\}, \{r_2\}, \emptyset$. The proof of the following result is provided in the Appendix.

Proposition 5. *Let $\mathcal{Q} = (X, R, \chi, F)$ be a CRS, let R' be a nonempty subset of R, and let $R_1, R_2, \ldots$ be a composition sequence for R'. The following are equivalent:*

(i) R' is a RAF for $\mathcal{Q}$;
(ii) $R_1, R_2, \ldots$ has terminal set $\emptyset$.
(iii) The associated iRAF sets $(\hat{R}_i, i \geq 1)$ partition R'.

Note that the iRAF sets $\hat{R}_i$ in (iii) are not, in general, iRAFs of the original CRS $\mathcal{Q}$, and since the sets $\hat{R}_i$ are (by definition) disjoint, the partitioning condition in (iii) is equivalent to the condition that the union of the sets $\hat{R}_i$ is equal to R'.

5 Ordering and stratifying reactions and elements in F-generated sets

In this section, we describe some structural and algorithmic results that allow for the identification of necessarily 'early' (or 'late') reactions (and elements) in any F-generated set.

Consider a reaction system with food set (X, R, F), and an F-generated subset R' of R. It is possible that a given reaction in R' must always occur before (one or several) other given reactions in R' under any admissible ordering of the reaction set. We describe a mathematical result that allows this ordering information to be determined by a polynomial-time algorithm.

First, we define a partial order on R' by writing $r \preceq r'$ if either $r = r'$ or if $r \neq r'$ and for every admissible ordering of R' the reaction r occurs before r' (in this latter case, we write $r \prec r'$).

Example: Consider the following reaction system:

$$\begin{aligned} r_1 &: f_1 + f_2 \to x \\ r_2 &: f_2 + f_3 \to y \\ r_3 &: f_4 + x \to z \end{aligned}$$

where $X = \{f_1, f_2, f_3, f_4, x, y, z\}$ and $F = \{f_1, f_2, f_3, f_4\}$. Then, $r_1 \prec r_3$. □

Let $\mathcal{P}_{R'}$ denote the resulting poset $(R', \preceq)$. Note that every admissible ordering of $\mathcal{R}'$ it is a linear extension of the poset $\mathcal{P}_{R'}$ (but not conversely), and an F-generated set R' has a *unique* admissible ordering if and only if $\mathcal{P}_{R'}$ is a total order.

To determine whether or not $r \prec r'$, let ξ be a new element that is not present in X. Given a reaction r, let ${}_{+\xi}r$ be the reaction obtained from r by adding ξ as an additional reactant of r. Similarly, given a reaction r', let $r'_{+\xi}$ be the reaction obtained from r' by adding ξ as an additional product of r'. Let $R'[{}_{+\xi}r, r'_{+\xi}]$ be the resulting set of reactions obtained from R' by replacing r with ${}_{+\xi}r$ and r' with $r'_{+\xi}$, respectively.

In the following result, $\mathcal{G}(R')$ denotes the directed graph that has vertex set R' and there is an arc from r to r' if at least one reactant x of r' that is a product of r and, in addition, $x \notin \mathrm{cl}_{R' \setminus \{r\}}(F)$.

Proposition 6. *Suppose that R' is an F-generated set, and consider any two distinct reactions r, r' in R'. The following are then equivalent:*

(i) $r \prec r'$.
(ii) $R'[{}_{+\xi}r, r'_{+\xi}]$ *is not F-generated.*
(iii) There is a directed path in $\mathcal{G}(R')$ from r to r'.

Consequently, $\mathcal{P}_{R'}$ and its associated Hasse diagram can be determined in polynomial time in the size of (X, R', F).

The ordering described can be extended to include any bidirectional reaction r, which has an associated forward reaction r^+ and a backward reaction r^-. In that case, we write $r_1 \prec r_2$ if $r_1^\alpha \prec r_2^\beta$ for all $\alpha, \beta \in \{+, -\}$.

5.1 Stratifying the reactions and elements in an F-generated set

Given a reaction system with food set (X, R, F) there is a natural way to partition an F–generated set R' into 'levels'. Specifically, let $R_0 = \{r \in R' : \rho(r) \subseteq F\}$, and for $i \geq 1$, let:

$$R_i = \left\{ r \in R' : \rho(r) \subseteq F \cup \pi\left(\bigcup_{0 \leq j < i} R_j \right) \right\}.$$

By definition, these sets are 'nested' (i.e. $R_i \subseteq R_{i+1}$ for all i). Since R' is F–generated, it follows that $R_i = R'$ for some value of i (to see this just consider an admissible ordering of R').

For each $r \in R'$, we can then define an associated ranking $\lambda(r)$ to be the smallest value of j for which $r \in R_j$. Thus the subsets of reactions from R' that have fixed lambda values (i.e. $\lambda(r) = 0, 1, 2 \ldots$) partitions R' into disjoint subsets. Suppose that R' is an F-generated set for (X, R, F), and $r, r' \in R'$. Then:

$$r \prec r' \Longrightarrow \lambda(r) < \lambda(r'). \tag{4}$$

To justify Eqn (4), we establish the contrapositive. Suppose that $\lambda(r') \leq \lambda(r)$. Then there is an admissible ordering of R' in which r' comes before r. Thus, not every admissible ordering of R' has r before r', so $r \prec r'$ does not hold.

Note also that if R' is an F-generated set for (X, R, F), and we write $r < r'$ if $r \neq r$ and at least one reactant of r' is a product of r but is not a product of any other reaction in R' (e.g., in Example 5, we have $r_1 < r_3$). Then the condition $r < r'$ implies that $r \prec r'$.

In a similar manner, we can also obtain a stratification of $\Pi(R')$ for any F–generated set R', as follows. Let $X_0 = F$, and for each $i \geq 1$, let:

$$X_i = X_{i-1} \cup \left(\bigcup_{r \in R' : \rho(r) \subseteq X_{i-1} \cup F} \pi(r) \right).$$

Thus X_i is X_{i-1} together with all the products of all the reactions iin R' that have all their reactants in $X_{i-1} \cup F$. The resulting collection of sets of elements from X provides a nested increasing sequence of sets $X_0 \subset \cdots \subset X_K = X_{K+1}$, where $X_K = \mathrm{cl}_{R'}(F)$. Next, define $\lambda : \Pi(R') \to \{1, 2, \ldots\}$ as follows:

$$\lambda(x) = \min\{i : x \in X_i\}.$$

In this way, the elements of $\Pi(R')$ are partitioned by their λ values. Moreover, the function λ has the following interpretation: $\lambda(x)$ is the length of the shortest admissible ordering of reactions (from R') that ends in a reaction that has x as a product. Thus if $\lambda(x) < \lambda(y)$, the element x can arise in fewer admissible 'steps' from the food set F than element y can.

6 Modelling the evolution of holobionts (and other ecological ensembles) via RAFs

In this section, we provide a novel application of RAF theory to model the evolution of holobionts (the application of RAFs to other systems in ecology has recently been studied by [10, 11]). Briefly, a holobiont is an ensemble of organisms of different species that have a symbiotic relationship with each other and a 'host'.

Well-known examples of holobionts include lichens as fungal–algal assemblages [24, 29], termites and the microbes in their digestive tracts [2], and plants for which bacteria assist with nitrogen-fixation [44, 33, 47]. These well-studied holobionts demonstrate the significant interspecific co-dependence of function that instantiates conceptual models of holobionts [36].

Given a host H, let $\mathcal{M}$ be the set of microbial OTUs (operational taxonomic units) that could potentially be part of a 'holobiont' with H, and let n denote the size of $\mathcal{M}$. Thus, a *holobiont for* H is a pair $h = (H, S)$ for a nonempty subset S of $\mathcal{M}$, and we refer to S as the microbiome set for h. In biology, only certain subsets (choices of M) are *viable*, and we consider a simple model for the evolution of these viable holobionts. Here we model this within RAF theory. Although we have focussed on the holobiont setting, the approach we describe may also be relevant for other settings where evolutionary processes involve sequential gain or loss of items from a set. Consider the following two types of elementary 'reactions'[4]. :

(R_+) $\quad r : S \to S \cup \{m\}$, where $m \in \mathcal{M} \setminus S$.
(R_-) $\quad r : S \to S \setminus \{m\}$, where $m \in S$.

In (R_+), a microbial OTU m is added to the microbiome of the host H, while in (R_-), a microbial OTU is lost from the microbiome. More briefly, we write $r : S \to S'$ to denote either of these reactions.

RAF theory provides a way to model a simple form of a holobiont's evolution by considering the catalytic reaction system (CRS) $\mathcal{Q} = (X, R, \chi, F)$, where:

- X is the collection of all possible subsets $\mathcal{M}$ (thus X has size 2^n)
- R is the set of reactions (of the three types) described above;
- F is the 'food set' is F consisting of the individual microbial OTUs (i.e. $\{\{m\} : m \in \mathcal{M}\}$);
- the catalysis assignment χ is simply the condition that whenever a reaction $r : S \to S'$ has the property that S' is (at least) as viable as S as a holobiont set for host H then $(S', r) \in \chi$.

Note that r_- and r_+ can both be catalysed, but only if the respective microbiome sets S and S' are equally viable. It is possible to form cycles of catalysis even when there are no ties in viability[5] as shown in Fig. 2.

In this way, we can obtain a catalytic reaction system (CRS) $\mathcal{Q} = (X, R, \chi, F)$ and thereby investigate RAFs, minimal RAFs, etc. The interpretation of a RAF for

[4]These reactions are written in a different way to the original description, however, it should be clear how to convert between them (e.g. $\{m_1, m_2\} \to \{m_1, m_2, m_3\}$ becomes $m_1 + m_2 \to m_1 + m_2 + m_3$).

[5]If viability is measured by an additive fitness function (i.e. the fitness of S is $\sum_{s \in S} f(s)$ for some function f), then a cycle of catalysed reactions requires all of the holobiont sets to have equal fitness values.

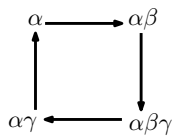

Fig. 2 An example of a possible 4-cycle of reactions. Any cycles of reactions must necessarily be of an even length, since each reaction changes the (even–odd) parity of the size of the holobiont set.

$\mathcal{Q}$ is that it is a viable holobiont set, that evolved via the elementary steps described above (starting from just the host) so that at each stage, the holobiont set is viable. In a laboratory-based study of colonization and competitive dominance of bioluminescent bacteria and their natural hosts, sepiolid squid, [32] showed that bacterial species native to specific squid species are able to competitively exclude other non-native bacteria when both colonise the same species. The authors concluded that their experiments provided evidence for processes that might lead to parallel, but equally functional, evolution of host–microbe relationships.

The evolution of RAFs in this model can also be modelled by a continuous-time stochastic process, which predicts the eventual distribution of holobiont sizes. In the simplest such model, one microbiome species in $\mathcal{M}$ (a set of size n) randomly joins a given holobiont of size i (via reaction R_+) at rate $\lambda \cdot (n-i)$ or is randomly lost from the holobiont (via reaction R_-) at rate $\mu \cdot i$. Standard Markov chain theory then provides the stationary distribution for the expected proportion π_k of holobionts of size $k \geq 0$, given by:

$$\pi_k = \binom{n}{k} (\lambda/\mu)^k (1+\lambda/\mu)^{-n}.$$

The special case where $\lambda = \mu$ is just the classic Ehrenfest model from statistical physics.

6.1 The RAF boundary of $\mathcal{Q}$

This simple model of a holobiont motivates a new notion in RAF theory, namely, that of the *RAF boundary* of a CRS set. Given a CRS $\mathcal{Q} = (X, R, C, F)$, let $\partial(\mathcal{Q})$ be the subset of X consisting of all those elements $x \in X$ for which (i) x is product of at least one reaction in the maxRAF of $\mathcal{Q}$, and (ii) x is not the reactant of any other reaction in the maxRAF of $\mathcal{Q}$. In the case where $\mathcal{Q}$ has no RAF, we set $\partial(\mathcal{Q}) = \emptyset$. The *RAF interior* of a CRS is the subset of the maxRAF of $\mathcal{Q}$ that is not in the RAF boundary. In short, $\partial(\mathcal{Q})$ can be viewed as the holobiont sets one would observe at the end of an evolutionary process, while the interior describes the routes by which these holobiont sets could have arisen under the model above.

An example of these concepts is illustrated in Fig. 3, where $\mathcal{M} = \{\alpha, \beta, \gamma, \delta\}$, and we will write x and xy in place of $\{x\}$, and $\{x, y\}$, etc, to simplify the notation. Note that $\alpha\beta\delta$ and $\beta\gamma\delta$ are not in the maxRAF (and so not in the RAF boundary) even though there is a blue arrow leading to them, since the source holobiont at the other end of this edge was not able to form under the evolutionary model described (the fact that the arrow is blue says that *if* the source holobiont were to exist (e.g., be made in a laboratory) then it could evolve into the more viable form).

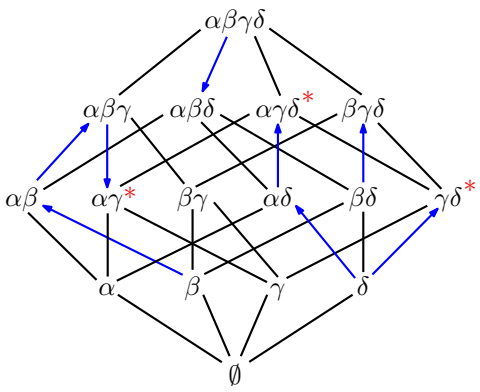

Fig. 3 A graph showing pathways for the formation of possible holobionts involving four microbial OTUs. The edges show the possible reactions (in either direction) involving the gain or loss of a single microbial OTU. The blue edges denote reactions that are catalysed by their product, with the edge oriented in the direction of the favoured holobiont. For this CRS $\mathcal{Q}$, the maxRAF consists of the six reactions $\delta \to \gamma\delta$, $\beta \to \alpha\beta \to \alpha\beta\gamma \to \alpha\gamma$, and $\delta \to \alpha\delta \to \alpha\gamma\delta$, and the RAF boundary of $\mathcal{Q}$ is the three microbiome sets $\alpha\gamma, \alpha\gamma\delta$ and $\gamma\delta$, which are indicated by red stars.

Leaving aside this particular example, several questions arise for the general setting including: What does $\partial(\mathcal{Q})$ tell us about the interior? and given $\partial(\mathcal{Q})$ (or some element x of it), what is a most parsimonious path within the maxRAF (i.e. evolutionary scenario) for arriving at x? A further question is the following: Given an element x in the RAF interior of $\mathcal{Q}$ and an element y in the RAF boundary, is there a viable (i.e. catalysed) path within the maxRAF from x to y (this problem can be easily solved by standard graph theoretic techniques, even though an 'evolutionary' ordering of an arbitrary collection of finite sets under a less constrained model is NP-complete [4]). Alternatively, one could ask, for two elements x, y in the RAF boundary, what common ancestor(s) in the RAF interior they could have had. We do not explore these questions further here.

7 Concluding comments

Autocatalytic networks have provided a formal tool to investigate processes in early biochemistry and certain other settings that involve the formation and evolution of complex structures. The discrete nature of the model provides a way to develop and implement efficient algorithms that can be applied to large data sets, and to elucidate their structural properties, such as their building blocks in terms of iRAFs or the ordering of reactions.

In this paper, we have developed further techniques that open the door to more detailed investigations into the structural properties of RAFs and iRAFs, their extensions to more general catalytic scenarios, and the investigation of the constraints on the order of reactions and the appearance of particular elements of interest. Several of the techniques described here have been implemented in the open-source software package *CatlyNet* [20], which we plan to apply to investigate further aspects of early metabolism and related evolutionary questions.

8 Acknowledgements

We thank the Royal Society Te Apārangi (New Zealand) for funding under the Catalyst Leader programme (agreement No. ILF-UOC1901).

References

[1] J. L. Andersen, C. Flamm, D. Merkle, and P. F. Stadler, *Defining autocatalysis in chemical reaction networks*. Arxiv: 2107.03086, 2021.

[2] J. Arora, Y. Kinjo, J. Šobotník, and et al., *The functional evolution of termite gut microbiota*, Microbiome, 10 (2022), p. 78, https://doi.org/10.1186/s40168-022-01258-3.

[3] A. Blokhuis, D. Lacoste, and P. Nghe, *Universal motifs and the diversity of autocatalytic systems*, Proceedings of the National Academy of Sciences, 117 (2020), pp. 25230–25236, https://doi.org/10.1073/pnas.2013527117.

[4] L. Bulteau, G. Sacomoto, and B. Sinaimeri, *Computing an evolutionary ordering is hard*, Electronic Notes in Discrete Mathematics, 50 (2015), pp. 255–260.

[5] Q. Dherbassy, R. J. Mayer, K. B. Muchowska, and J. Moran, *Metal-pyridoxal cooperativity in nonenzymatic transamination*, J. Am. Chem. Soc., 145 (2023), pp. 13357–13370.

[6] P. Dittrich and P. Speroni di Fenizio, *Chemical organization theory*, Bulletin of Mathematical Biology, 69 (2007), pp. 1199–1231.

[7] A. Filisetti, M. Villani, C. Damiani, A. Graudenzi, A. Roli, W. Hordijk, and R. Serra, *On RAF sets and autocatalytic cycles in random reaction networks*, in Advances in Artificial Life and Evolutionary Computation, C. Pizzuti and G. Spezzano, eds., Cham, 2014, Springer International Publishing, pp. 113–126.

[8] L. Gabora and M. Steel, *Autocatalytic networks in cognition and the origin of culture*, J. Theor. Biol., 431 (2017), pp. 87–95.

[9] L. Gabora and M. Steel, *A model of the transition to behavioral and cognitive modernity using reflexively autocatalytic networks*, J. R. Soc. Interface, 17 (no. 20200545) (2020), pp. 1–14, https://doi.org/10.1098/rsif.2020.0545.

[10] R. C. Gatti, B. Fath, W. Hordijk, S. Kauffman, and R. Ulanowicz, *Niche emergence as an autocatalytic process in the evolution of ecosystems*, Journal of Theoretical Biology, 454 (2018), pp. 110–117.

[11] R. C. Gatti, W. Hordijk, and S. Kauffman, *Biodiversity is autocatalytic*, Ecological Modelling, 346 (2017), pp. 70–76.

[12] R. C. Gatti, R. Koppl, B. D. Fath, S. Kauffman, W. Hordijk, and R. E. Ulamnowicz, *On the emergence of ecological and economic niches*, Journal of Bioeconomics, 22 (2020), pp. 99–127.

[13] P. G. Higgs, *When is a reaction network a metabolism? criteria for simple metabolisms that support growth and division of protocells*, Life, 11 (2021), p. 966, https://doi.org/10.3390/life11090966.

[14] W. Hordijk, S. Shichor, and G. Ashkenasy, *The influence of modularity, seeding, and product inhibition on peptide autocatalytic network dynamics*,

ChemPhysChem, 19 (2018), pp. 2437–2444.
[15] W. Hordijk and M. Steel, *A formal model of autocatalytic sets emerging in an RNA replicator system*, J. Syst. Chem., 4 (2013), pp. 1–9.
[16] W. Hordijk and M. Steel, *Autocatalytic sets and boundaries*, Journal of Systems Chemistry, 6 (2015), pp. 1–5, https://doi.org/10.1186/s13322-014-0006-2.
[17] W. Hordijk and M. Steel, *Chasing the tail: The emergence of autocatalytic networks*, Biosystems, 152 (2017), pp. 1–10.
[18] W. Hordijk, M. Steel, and P. Dittrich, *Autocatalytic sets and chemical organizations: Modeling self-sustaining reaction networks at the origin of life*, New J. Phys., 20 (no. 015011) (2018), pp. 2–14.
[19] W. Hordijk, M. Steel, and S. Kauffman, *The structure of autocatalytic sets: Evolvability, enablement, and emergence*, Acta Biotheor., 60 (2012), pp. 379–392.
[20] D. H. Huson and M. Steel, *Catlynet*, 2023, https://github.com/husonlab/catlynet.
[21] S. Jaramillo and et al, *(M,R) systems and RAF sets: common ideas, tools and projections*, in in Proc. ALIFEXII Conf., Odense, Denmark, August 2010, 17, MIT Press, Cambridge, MA, 2010, pp. 94–100.
[22] S. A. Kauffman, *Autocatalytic sets of proteins*, J. Theor. Biol., 119 (1986), pp. 1–24, https://doi.org/10.1016/S0022-5193(86)80047-9.
[23] S. A. Kauffman, *The Origins of Order*, Oxford University Press, New York, NY, USA, 1993.
[24] V. L., L. Muggia, O. Peksa, T. Ridka, and P. Skaloud, *The complexity of symbiotic interactions influences the ecological amplitude of the host: A case study in Stereocaulon (lichenized Ascomycota)*, Mol. Ecol., 27 (2018), pp. 3016–3033.
[25] D. Lancet, R. Zidovetzki, and O. Markovitch, *Systems protobiology: origin of life in lipid catalytic networks*, J. Roy. Soc. Interface, 15 (no. 20180159 (2018), pp. 1–36, https://doi.org/10.1098/rsif.2018.0159.
[26] Y. Liu, *On the definition of a self-sustaining chemical reaction system and its role in heredity*, Biology Direct, 15 (2020), p. 15, https://doi.org/10.1186/s13062-020-00269-0.
[27] D. Loutchko, *An algebraic characterization of self-generating chemical reaction networks using semigroup models*, J. Math. Biol., 86 (2023), pp. 1–36, https://doi.org/10.1007/s00285-023-01899-4.
[28] D. Loutchko, *Semigroup models for biochemical reaction networks*, J. Math. Biol., 86 (2023), p. 78, https://doi.org/10.1007/s00285-023-01898-5.
[29] S. M., I. Schmitt, D. Weber, and F. Dal Grande, *Fungal host affects photosynthesis in a lichen holobiont*, J. Fungi (Basel), 8 (no. 1267) (2022), pp. 1–10.
[30] E. Mossel and M. Steel, *Random biochemical networks: the probability of self-sustaining autocatalysis*, J. Theor. Biol., 233 (2005), pp. 327–336.
[31] K. B. Muchowska, S. J. Varma, and J. Moran, *Nonenzymatic metabolic reactions and life's origins*, Chem. Rev., 120 (2020), pp. 7708–7744.
[32] M. K. Nishiguchi, E. G. Ruby, and M. J. McFall-Ngai, *Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in Sepiolid squid-vibrio symbioses*, Appl. Environ.

Microbiol., 64 (1998), pp. 3209–3213.

[33] S. L. Ottinger, C. F. Miniat, and N. Wurzburger, *Nitrogen and light regulate symbiotic nitrogen fixation by a temperate forest tree*, Oecologia, 201 (2023), pp. 565–574.

[34] Z. Peng, J. Linderoth, and D. A. Baum, *The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life*, PLoS Comput Biol, 18 (no. e1010498) (2022), pp. 1–28.

[35] Z. Peng, K. Paschek, and J. C. Xavier, *What Wilhelm Ostwald meant by "autokatalyse" and its significance to origins-of-life research*, BioEssays, 44 (no. 2200098) (2022).

[36] E. Rosenberg and I. Zilber-Rosenberg, *The hologenome concept of evolution after 10 years*, Microbiome, 6 (no. 78) (2018), pp. 1–14.

[37] F. L. Sousa, W. Hordijk, M. Steel, and W. F. Martin, *Autocatalytic sets in E. coli metabolism*, Journal of Systems Chemistry, 6 (no. 4) (2015), pp. 1–21.

[38] M. Steel, *Interior operators and their relationship to autocatalytic networks*, Acts Biotheoretica, (accepted). arXiv:2305.16714 (2023).

[39] M. Steel, W. Hordijk, and J. Smith, *Minimal autocatalytic networks*, Journal of Theoretical Biology, 332 (2013), pp. 96–107.

[40] M. Steel, W. Hordijk, and J. C. Xavier, *Autocatalytic networks in biology: structural theory and algorithms*, J. Royal Society Interface, 16 (no. 20180808) (2019), pp. 1–10.

[41] M. Steel, J. C. Xavier, and D. H. Huson, *The structure of autocatalytic networks, with application to early biochemistry*, J. Royal Society Interface, 17 (n. 20200488) (2020), pp. 1–11.

[42] N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, and N. Lehman, *Spontaneous network formation among cooperative RNA replicators*, Nature, 491 (2012), pp. 72–77.

[43] V. Vasas, C. Fernando, M. Santos, S. Kauffman, and E. Szathmáry, *Evolution before genes*, Biology Direct, 7 (no. 1) (2012), pp. 1–14, https://doi.org/10.1186/1745-6150-7-1.

[44] C. E. Wendlandt, K. A. Gano-Cohen, P. J. N. Stokes, B. N. R. Jonnala, A. J. Zomorrodian, and et al, *Wild legumes maintain beneficial soil rhizobia populations despite decades of nitrogen deposition*, Oecologia, 198 (2022), pp. 419–430.

[45] J. C. Xavier, W. Hordijk, S. Kauffman, M. Steel, and W. F. Martin, *Autocatalytic chemical networks at the origin of metabolism*, Proc. Roy. Soc. B., 287 (no. 20192377) (2020), pp. 1–10, https://doi.org/10.1098/rspb.2019.2377.

[46] J. C. Xavier and S. Kauffman, *Small-molecule autocatalytic networks are universal metabolic fossils*, Phil. Trans. Roy. Soc. A, 380 (no. 20210244) (2022).

[47] J. Yun, C. Wang, F. Zhang, L. Chen, Z. Sun, and et al, *A nitrogen fixing symbiosis-specific pathway required for legume flowering*, Sci. Adv., 9 (no. ade1150) (2023), pp. 1–14.

9 Appendix: Mathematical details

Proof of Proposition 1: Let ψ be an arbitrary interior operator on 2^Y. We construct a CRS $\mathcal{Q}_\psi = (X, R_Y, \chi, F)$ that permits complex catalysis rules, as follows. Let $F = \{f\}$, let $X = \{x_y : y \in Y\} \cup \{f\}$, and let $R_Y = \{r_y : y \in Y\}$ where r_y is the reaction $f \to x_y$. It remains to describe, for each reaction $r_y \in R_Y$, its associated catalyst sets (i.e. the subsets U of X for which $(U, r_y) \in \chi$). Let

$$C(y) = \{A \subseteq Y : y \in A, \psi(A) = A\}, \tag{5}$$

the fixed sets of ψ that contain y. Then we let (U, r_y) be an element of χ if and only if $U = \{x_a : a \in A\}$ for some set $A \in C(y)$.

Let $b : Y \to R_Y$ denote the bijection $y \mapsto r_y$.

Claim 1: A nonempty set A of Y is a fixed set of ψ if and only if $b(A) = \{r_y : y \in A\}$ is a RAF for $\mathcal{Q}_\psi$.

To establish this claim, suppose first that A is a fixed set of ψ (i.e. $\psi(A) = A$). Since any nonempty subset of R_Y is F-generated (because each reaction in R_Y has only the single reactant f which lies in the food set), it suffices to show that each reaction r_y in $b(A)$ has an associated catalyst set U with the property that each of its elements is produced by at least one reaction in $b(A)$. Suppose that $r_y \in b(A)$. Then $y \in A$, and, by assumption, A is a fixed set of ψ, so $A \in C(y)$; moreover, the elements $\{x_a : a \in A\}$ are products of reactions in $b(A)$. Thus, taking $U = \{x_a : a \in A\}$ we have $(U, r_y) \in \chi$; in other words, r_y is catalysed by products of reactions in $b(A)$. Since this holds for all $r_y \in b(A)$, it follows that $b(A)$ is a RAF.

Conversely, suppose that $b(W)$ is a RAF for $\mathcal{Q}_\psi$ for some subset W of Y. Then for each $w \in W$, the reaction $r_w : f \to x_w$ has a catalyst set of the form $\{x_a : a \in A_w\}$ where A_w is a nonempty subset of W satisfying $w \in A_w$ and $\psi(A_w) = A_w$. It remains to show that $\psi(W) = W$. Since $\psi(W) \subseteq W$ it suffices to determine the converse containment. We have:

$$W \subseteq \bigcup_{w \in W} A_w = \bigcup_{w \in W} \psi(A_w) = \psi\left(\bigcup_{w \in W} A_w\right) \subseteq \psi(W),$$

as required to establish Claim 1.

Next observe that the RAFs for $\mathcal{Q}_\psi$ are simply the nonempty fixed sets of φ (the maxRAF operator on R_Y for the CRS $\mathcal{Q}_\psi$), and so ψ and $b^{-1} \circ \varphi \circ b$ are two interior operators on 2^Y, and these two functions the same fixed sets by Claim 1 (noting also that φ and ψ also trivially fix $\emptyset$). However, any two interior operators on 2^Y that have the same collection of fixed sets are identical functions. This is because any interior map ν on a set 2^Y is completely determined by its fixed sets since for any set $Y' \subseteq Y$ we have:

$$\nu(Y') = \bigcup \{U \subseteq Y' : \nu(U) = U\}.$$

Thus $\psi = b^{-1} \circ \varphi \circ b$, completing the proof. □

Remark: It can be shown that Claim 1 also holds if we restrict the catalyst sets $U = \{x_a : a \in A\}$ to just those that correspond to the *minimal* sets $A \in C(y)$ (where minimal refers to set inclusion).

An example to illustrate Corollary 1: Consider the following CRS (based on Kauffman's binary polymer model with food set $F = \{0, 1, 00, 01, 10, 11\}$ from Fig. 1 of [40]):

$$\begin{aligned}
&r_1 : 10 + 0\ [01100] \to 100\\
&r_2 : 01 + 100\ [0] \to 01100\\
&r_3 : 10 + 1\ [0] \to 101\\
&r_4 : 11 + 10\ [101] \to 1110\\
&r_5 : 1110 + 0\ [101] \to 11100
\end{aligned}$$

This system is itself a RAF and it contains six other RAFs as subsets, namely $\{r_1, r_2\}, \{r_3\}, \{r_1, r_2, r_3\}, \{r_3, r_4\}, \{r_3, r_4, r_5\}, \{r_1, r_2, r_3, r_4\}$. Following the procedure described in the proof, we associate with r_i the reaction $r'_i : f[U_i] \to x_i$ where the set of (complex) catalysts is $U_i = \{\{x_j : r_j \in A\} : A$ is a RAF containing $r_i\}$. For example, for r_1 we have:

$$U_1 = \{\{x_1, x_2\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4, x_5\}\}.$$

We can further simply U_1 in line with remark following the proof of Proposition 1 above. For U_1 we only need to keep the single set $\{x_1, x_2\}$ since all other sets in U_1 contain this set. Applying this simplification to the other sets U_i leads to the resulting reduced system:

$$\begin{aligned}
&r'_1 : f\ [\{x_1, x_2\}] \to x_1\\
&r'_2 : f\ [\{x_1, x_2\}] \to x_2\\
&r'_3 : f\ [\{x_3\}] \to x_3\\
&r'_4 : f\ [\{x_3, x_4\}] \to x_4\\
&r'_5 : f\ [\{x_3, x_4, x_5\}] \to x_5
\end{aligned}$$

The RAFs of this system (with $F = \{f\}$) then correspond bijectively to the RAFs of the original system (under the bijection $r_i \leftrightarrow r'_i$ for each i).

Proof of Proposition 2: Suppose that $R_{\text{int}} := \bigcap_{i \geq 1} \hat{R}_i \neq \emptyset$. Then R_{int} is a RAF, and every reaction of R_{int} has a catalyst U_i that is not a subset of F, and so R_{int} is a strictly autocatalytic RAF. On the other hand, if R' is a strictly autocatalytic RAF, then since $R' \subseteq R$ we have (by induction on $i \geq 1$) $R' \subseteq \hat{R}_i$ for each i. Thus, since $R_{\text{int}} = \hat{R}_k$ for some value of k we have $R' \subseteq R_{\text{int}}$ and so R_{int} is the unique maximal strictly autocatalytic RAF for $\mathcal{Q}$. □

Proof of Proposition 3: Suppose that R' is a minimal subset of maxRAF($\mathcal{Q}$) that is a RAF and produces all of the elements $x_1, \ldots, x_k$. Then R' is a RAF of $\mathcal{Q}^{+\{x_1, x_2, \ldots, x_k\}}$.

Moreover, if R'' is a strict subset of R' that is a both a RAF and that produces $x_1, \ldots, x_k$, then R'' is also a RAF of $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$ and so R' is not a minimal set with the joint properties of being a RAF and producing $x_1, \ldots, x_k$. Thus R' is an iRAF of $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$. Conversely, suppose that R' is an iRAF of $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$. Then R' is a RAF of $\mathcal{Q}$ and R' must produce all of the elements $x_1, \ldots, x_k$ (since all reactions in $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$ require each of the elements $x_1, \ldots, x_k$ to be produced because none of these elements lie in the food set F (by assumption)), and thus R' is a RAF of $\mathcal{Q}$ that produces $x_1, \ldots, x_k$. Moreover, R' is a minimal subset with this property, since otherwise R' contains a strict subset R'' with these properties and this set would be a RAF of $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$, which contradicts the assumption that R' is an iRAF of $\mathcal{Q}^{+\{x_1,x_2,\ldots,x_k\}}$. □

Proof of Proposition 4: The argument just used the interior operator properties of φ. If $\varphi(\varphi(R) \setminus \{r_1, \ldots, r_k\}) \neq \emptyset$ then $\varphi(R) \setminus \{r_1, \ldots, r_k\}$ contains an iRAF, and this iRAF cannot be one of $R_1, \ldots, R_k$, since for each i, r_i was chosen from R_i yet $r_i \notin \varphi(\varphi(R) \setminus \{r_1, \ldots, r_k\})$ by definition. Conversely, if there is an iRAF R' of R that is different from $R_1, \ldots, R_k$, then selecting r_i in $R_i \setminus R'$ for each i (this is possible since R_i is not a subset of R'), R' is an iRAF of $\varphi(R) \setminus \{r_1, \ldots, r_k\}$ and so $R' = \varphi(R') \subseteq \varphi(\varphi(R) \setminus \{r_1, \ldots, r_k\})$, thus this set is nonempty. □

Proof of Proposition 5: We begin with a lemma.

Lemma 2. *Suppose that R' and $R'' \subsetneq R'$ are both RAFs for $\mathcal{Q} = (X, R, \chi, F)$. Then $R' \setminus R''$ is a RAF for $(X, R, \chi, F \cup \pi(R''))$.*

To see why this holds, simply observe that if o' is an admissible ordering for R' and o^- is the induced ordering of $R' \setminus R''$ then o^- is an admissible ordering of $R' \setminus R''$ for the CRS with expanded food set $(X, R, \chi, F \cup \pi(R''))$.

Returning to the proof of Proposition 5, observe that (ii) and (iii) equivalent (by the comments following the lemma) so it suffices to show that (iii) implies (i), and, conversely, that (i) implies (iii). Suppose first that (iii) holds. Since R' equals the (disjoint) union of the sets $\hat{R}_i$ we can select an admissible ordering of $\hat{R}_i$ for each i, and extend this to an ordering of the reactions of R' in which all reactions from $\hat{R}_i$ come before those of $\hat{R}_{i+1}$ for each i. This ordering is an admissible ordering of R', and each reaction in R' is either catalysed by an element of the food set or by the product of another reaction from R'. Thus R' is a RAF for $\mathcal{Q}$.

The proof of the converse direction ((i) implies (iii)) follows from Lemma 2, by induction on the length of the sequence $R_1, R_2, \ldots$ and the fact that every RAF contains an iRAF. □

Proof of Proposition 6: We begin by stating a lemma. Recall the definition of $\mathcal{G}(R')$ (the directed graph that has vertex set R' and there is an arc from r to r' if at least one reactant x of r' that is a product of r and, in addition, $x \notin \mathrm{cl}_{R' \setminus \{r\}}(F)$).

Lemma 3. *[39] In any reaction system (X, R) with a food set $F \subseteq X$, a nonempty subset R' of R is F-generated if and only if the following condition holds: For each $r \in R'$, (i) $\rho(r) \subseteq \pi(R') \cup F$ and (ii) the graph $\mathcal{G}(R')$ has no directed cycles (including loops).*

We now establish the required implications.

(i) $\Rightarrow$ (ii): [We show that the negation of (ii) implies (i)] Suppose that $R'[_{+\xi}r, r'_{+\xi}]$ is F-generated and let o' be an admissible ordering. Then $r'_{+\xi}$ comes before $_{+\xi}r$ in o' (since $r'_{+\xi}$ is the only reaction that produces the reactant ξ required for $_{+\xi}r$). Moreover, the ordering o obtained from o' by replacing $_{+\xi}r$ and $r'_{+\xi}$ with r and r' respectively is an admissible ordering for R' in which r' comes before r and so $r \prec r'$ does not hold.

(ii) $\Rightarrow$ (iii): Suppose that $R'[_{+\xi}r, r'_{+\xi}]$ is not F-generated. Then $R'[_{+\xi}r, r'_{+\xi}]$ violates one of the two conditions stated Lemma 3 (with R' replaced by $R'[_{+\xi}r, r'_{+\xi}]$). Since condition (i) clearly holds for $R'[_{+\xi}r, r'_{+\xi}]$, the second condition must fail - that is $\mathcal{G}(R'[_{+\xi}r, r'_{+\xi}])$ contains a directed cycle. However, R' is F-generated, and so (again by Lemma 3) the graph $\mathcal{G}(R')$ has no directed cycle. Moreover, identifying the vertices of these two graphs (i.e. where r is associated with $_{+\xi}r$, and r' with $r'_{+\xi}$) the only difference between the arc sets of these graphs is that $\mathcal{G}(R'[_{+\xi}r, r'_{+\xi}])$ contains an arc from $r'_{+\xi}$ to $_{+\xi}r$, while $\mathcal{G}(R')$ need not have an arc from r' to r. Since the former graph has a directed cycle but the latter does not, there is a directed path in $\mathcal{G}(R')$ from r to r', as claimed.

(iii) $\Rightarrow$ (i): Suppose that (r_i, r_j) is an arc in $\mathcal{G}(R')$. Then in any admissible ordering o for $\mathcal{G}(R')$ reaction r_i comes before r_j. There is thus a reactant x of r_j that is a product of r_i and, in addition, x is not in the set $\mathrm{cl}_{R'\setminus\{r_i\}}(F)$. Thus, in any admissible ordering for R', r_i must come before r_j. Therefore, by transitivity, if there is a directed path from r to r' in $\mathcal{G}(R')$, then r comes before r' in any admissible ordering of R', and therefore $r \prec r'$. $\square$