Author for correspondence:
Mike Steel
e-mail: mike.steel@canterbury.ac.nz

THE ROYAL SOCIETY
PUBLISHING

The structure of autocatalytic networks, with application to early biochemistry

Mike Steel[1], Joana C. Xavier[2] and Daniel H. Huson[3]

[1]Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand
[2]Institute for Molecular Evolution, Heinrich Heine Universität, Dusseldorf, Germany
[3]Center for Bioinformatics, University of Tübingen, Tubingen, Germany

MS, 0000-0001-7015-4644; JCX, 0000-0001-9242-8968

Metabolism across all known living systems combines two key features. First, all of the molecules that are required are either available in the environment or can be built up from available resources via other reactions within the system. Second, the reactions proceed in a fast and synchronized fashion via catalysts that are also produced within the system. Building on early work by Stuart Kauffman, a precise mathematical model for describing such self-sustaining autocatalytic systems (RAF theory) has been developed to explore the origins and organization of living systems within a general formal framework. In this paper, we develop this theory further by establishing new relationships between classes of RAFs and related classes of networks, and developing new algorithms to investigate and visualize RAF structures in detail. We illustrate our results by showing how it reveals further details into the structure of archaeal and bacterial metabolism near the origin of life, and provide techniques to study and visualize the core aspects of primitive biochemistry.

1. Introduction

The process by which life arose from abiotic chemistry on Earth more than 4 billion years ago remains an outstanding scientific question [1]. Although the precise details of the origin of life may be difficult or impossible to know with any certainty, a more realistic goal is how life *might* have begun; in other words, what is a scientifically plausible scenario? Although a number of specific proposals have been put forward, such as the hydrothermal vent scenario of Martin & Russell [2], there is currently no general agreement on the processes that led to life. A complicating issue is that the emergence of life requires several steps to occur, including the establishment of metabolism, containment (e.g. the formation of a protocell), encoding and replication via a rudimentary information-processing system, and natural selection. Nevertheless, comparing the metabolism of organisms across the tree (or network) of life provides some clues into the nature of early metabolism. In particular, the metabolic networks of bacteria and archaea share certain structural features and thereby provide insights into the nature of early metabolism prior to the separation of these two domains. A recent study [3] used mathematical and computational techniques (developed further in this paper) to help identify and investigate these ancestral features, based on two prokaryotes thought to be close to early life (justification for their assumed ancestry is detailed therein). Here, we focus on the metabolism of one of those prokaryotes, the methanogenic archaea *Methanococcus maripaludis*. Eukaryotes, being more recent (derived from a symbiogenic event between archaea and bacteria) were excluded from these analyses.

A ubiquitous feature of all life on earth is the ability for an organism's metabolism to be simultaneously self-sustaining and collectively autocatalytic. Systems that combine these two general properties have been studied within a formal framework, sometimes referred to as RAF theory [4] (RAF refers to 'Reflexively Autocatalytic and *F*–generated', defined in §1.2). This approach traces back to Stuart Kauffman's pioneering work on 'collectively autocatalytic networks' [5–7] in polymer models of early life, which was subsequently

developed further mathematically (see [4,8] and the references there-in). RAF theory also overlaps with other graph-theoretic approaches in which the emergence of directed cycles in reaction graphs plays a key role (see e.g. [9–11]). RAF theory has also been applied in other fields, including ecology [12] and cognition [13]. In this paper, we extend RAF theory further to provide new techniques for exploring and visualizing the structure of RAFs and related concepts, and apply them to large metabolic networks that are close to early branches of the tree of life. We have implemented these methods in a new interactive program called *CatlyNet* [14].

The structure of this paper is as follows. In the next section, we summarize the key definitions and results in RAF theory, and illustrate these concepts on primitive metabolic network data, from a recent study. We also discuss and clarify some technical issues concerning bidirectional reactions and catalysis options. In §3, we investigate the finer structure of RAFs and related entities that can provide further insight into complex metabolic networks. Our emphasis is on approaches that can be efficiently carried out algorithmically (i.e. in polynomial time, rather than being NP-hard). In §4, we show how one can readily identify a unique minimal (core) RAF, if it exists, and identify the reactions that must have proceeded uncatalysed initially (though catalysed once the RAF is fully formed). We also discuss additional complications that arise when molecule types not only catalyse reactions but can also inhibit them. We end with some brief concluding comments.

1.1. Preliminary background and definitions

The following definitions are phrased in the language of chemistry; however, the same formalism and concepts have been applied in other areas (e.g. ecology, cognition, economics) by interpreting 'molecule type', 'reaction', 'catalysis' and 'food set' in a different setting (for details, see [15]).

A *catalytic reaction system* (with food set), abbreviated CRS, is a quadruple $\mathcal{Q} = (X, R, C, F)$, where X is a set of *molecule types*, R is a set of *reactions* (defined shortly), C is a subset of $X \times R$ called the *catalysis assignment* (which has the interpretation that if $(x, r) \in C$ then molecule type x acts as a catalyst for reaction r) and F is a subset of X, regarded as the set of molecule types that are freely available in the environment.

For the set R, we regard a 'reaction' as an ordered pair (A, B) of sets, with the elements in A and B being subsets of X. The interpretation here is that the molecule types in A combine to produce the molecule types in B. The sets A and B are referred to, respectively, as the *reactants* of r (denoted $\rho(r)$) and *products* of r (denoted $\pi(r)$). We will let $\pi(R') = \bigcup_{r \in R'} \pi(r)$ denote the set of molecule types that are a product of at least one reaction from R'. Throughout this paper, we will often denote reactions by using arrows, and catalysis with square brackets. For example, if reaction r combines molecule types a and b to generate x, and r is catalysed by either y or z we write $r : a + b\ [y, z] \longrightarrow x$.

Note that chemical reactions also typically involve stoichiometric considerations, where more than one molecule may be required as a reactant (or produced as a product). For example, the reaction $r : 2a + b \longrightarrow x + 3y$ leads to the sets $\rho(r) = \{a, b\}$, $\pi(r) = \{x, y\}$ and thereby ignores multiplicities. In RAF theory, treating the reactants and products as sets (rather than multisets) simplifies the theory and statement of the results, generally leads to no substantive differences

than if stoichiometry had been modelled explicitly. Similarly, reactions that are bidirectional can also be handled within the existing (directional) framework, as we describe in §2.3. We also point out here that RAF theory does not explicitly use kinetic features of reaction systems; in other words, it is based on a 'minimalistic' and 'high-level' description of a CRS based on a discrete notion of 'catalysis' (and later also 'inhibition'). This has two advantages: (i) it allows for a high level of generality, and the development of a variety of theorems and fast algorithms, and (ii) it can be applied to large and complex systems for which detailed kinetic information (reactions rates etc.) is not available. Nevertheless, the inclusion of rate information in RAF theory has been investigated in recent work [15], and is a topic that would benefit from further work in the future.

1.2. RAFs, CAFs and pRAFs

Given a CRS $\mathcal{Q} = (X, R, C, F)$, subset X' of X and a subset R' of R, we say that X' is *closed relative to* R' if X' satisfies the following property: for all $r \in R'$, $\rho(r) \subseteq X' \Rightarrow \pi(r) \subseteq X'$. In other words, for each reaction r in R' that has all its reactants present in X', every product of r is also in X'.

Given a subset R' of R, $cl_{R'}(F)$ is the intersection of all subsets X' of X that contain F and are closed relative to R'. This is well defined, since the full set of molecule types X is closed relative to any subset of R. The set $cl_{R'}(F)$ has a simple interpretation: it is the set of molecule types in F together with any other molecule types x in X for which x can be generated from F by some sequence of reactions from R' where each reaction in this sequence has each of its reactants present in F or is a product of an earlier reaction in the sequence. Moreover, $cl_{R'}(F)$ can be computed quickly (in polynomial time in the size of $\mathcal{Q}$) [16].

If R' has the property that each reaction in R' has its reactants in $cl_{R'}(F)$, then R' is said to be *F-generated*. In other words, R' is F-generated if all reactants required for any reaction in R' can be built up starting from the food set by applying only reactions in R'. By lemma 3.1 of ([17]), R' is *F-generated* if and only if the following condition holds:

— R' can be ordered $r_1, r_2, \ldots, r_k$ so that, for each $i \geq 1$, each reactant of r_i is contained in the food set and/or is a product of an earlier reaction in the sequence.

Given a CRS $\mathcal{Q}$, a subset R' of R is a *RAF* (respectively, *CAF* or *pRAF*) for $\mathcal{Q}$ if $R' \neq \emptyset$ and the following conditions hold, respectively:

[RAF] Each reaction in R' has at least one catalyst present in $F \cup \pi(R')$ and R' is F-generated.

[CAF] R' can be ordered $r_1, r_2, \ldots, r_k$ so that, for each $i \geq 1$, each reactant of r_i and at least one catalyst of r_i is contained in the food set and/or is a product of an earlier reaction in the sequence).

[pRAF] For each $r \in R'$, each of the reactants of r and at least one catalyst of r are contained in $F \cup \pi(R')$.

It is clear from these definitions that every CAF is a RAF and every RAF is a pRAF. Notice that a pRAF is a RAF if and only if the pRAF is also F-generated (the extra condition that a pRAF requires in order to be F-generated can be stated in terms of a certain graph on the set of reactions having no directed cycle[1]).

Table 1. Abbreviations used throughout this paper (with section numbers).

abbreviation	name
CRS	catalytic reaction system, (§1.1)
RAF	reflexively autocatalytic and *F*-generated set (§1.2)
CAF	constructively autocatalytic and *F*-generated set (§1.2)
pRAF	pseudo-RAF set (may not be *F*-generated) (§1.2)
subRAF	RAF set that is a strict subset of the maxRAF (§1.3)
maxRAF	maximal RAF (unique) (§1.3)
maxCAF	maximal CAF (unique) and maximal pRAF (unique) (§1.3)
max-pRAF	
irrRAF	irreducible (minimal) RAF set (§1.3)
uRAF	uninhibited RAF set (defined later in §4.4)

The abbreviation 'RAF' comes from the two conditions in the definition (RA=Reflexively Autocatalytic; F=*F*-generated). An equivalent definition of a RAF is as a non-empty subset R' of R for which every reaction has all its reactants and at least one catalyst present in $\mathrm{cl}_{R'}(F)$. Similarly, CAF refers to constructively autocatalytic and F-generated, and pRAF refers to pseudo-RAF (it need not be *F*-generated).

One may also consider extra conditions to avoid trivialities in the above definition of a RAF. For example, a RAF, CAF or pRAF R' may be required to contain at least one reaction that generates a molecule type that is not found in the food set. Such conditions can usually be handled by simply modifying the CRS (e.g. in the case described, removing each reaction that has no product outside the food set from R).

1.3. Maximal and minimal sets

Since the union of any collections of RAFs (respectively, CAFs or pRAFs) for $\mathcal{Q}$ is a RAF (respectively, CAF or pRAF) for $\mathcal{Q}$, it immediately follows that when a RAF, CAF or pRAF exists, there is a unique maximal one called the *maxRAF*, *maxCAF* and *max-pRAF*, respectively. Moreover, these can be found by a fast (polynomial-time) algorithm, which also correctly reports whether a RAF, CAF or pRAF is present [16].

A RAF that is a strict subset of the maxRAF is sometimes called a *subRAF*, and we say that R' is an irreducible RAF (*irrRAF*) for $\mathcal{Q}$ if R' has no subRAF. Clearly, any RAF of size 1 is an irrRAF; a CAF is an irrRAF if and only if it has size 1. When a CRS has a RAF, finding an irrRAF is easy (being polynomial time in the size of $\mathcal{Q}$) but finding the smallest RAF for $\mathcal{Q}$ (which is necessarily an irrRAF) is NP-hard [17].

As a number of acronyms and abbreviations are used throughout this paper, we summarize these in table 1, which also indicates the section where each abbreviation is defined.

1.4. Examples

Consider the system (from [18]) involving three catalysed reactions with $X = \{s, t, u, st, su, stu\}$ and with food set $F = \{s, t, u\}$:
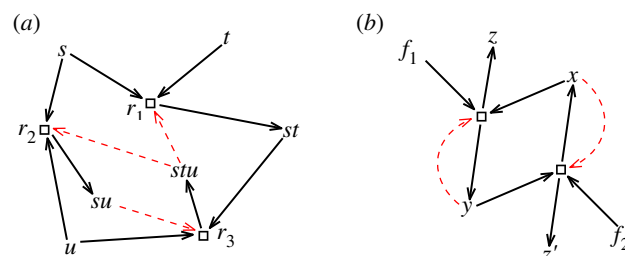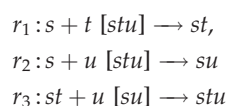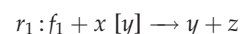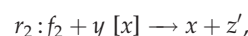
$$r_1 : s + t \ [stu] \longrightarrow st,$$
$$r_2 : s + u \ [stu] \longrightarrow su$$

and

$$r_3 : st + u \ [su] \longrightarrow stu$$

Figure 1. (*a*) A CRS system (from [18]) with food set $F = \{s, t, u\}$, which forms a RAF. This is also an irrRAF but is not a CAF (and does not contain a CAF). (*b*) A simple example of a pRAF that is not a RAF.

The set $\{r_1, r_2, r_3\}$ forms a RAF (but not a CAF) for this CRS; moreover, this RAF is also an irrRAF, since all three reactions are required for a RAF to be present in any RAF. An example of a pRAF that is not a RAF is given by the following system:
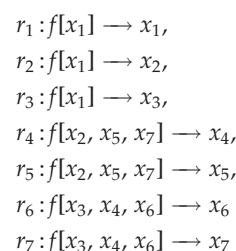
$$r_1 : f_1 + x \ [y] \longrightarrow y + z$$

and

$$r_2 : f_2 + y \ [x] \longrightarrow x + z',$$

where $\{f_1, f_2\}$ denotes the food set. These two systems are shown in figure 1.
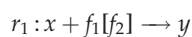
A second, more complex example is the laboratory-based autocatalytic ribozyme system from [19] analysed in [20] consisting of seven reactions that constitute a RAF. This maxRAF is also not a CAF (nor does it contain a CAF); however it contains 66 other RAFs as subsets. Formally, this system has food set $F = \{f\}$ and seven reactions

$$r_1 : f[x_1] \longrightarrow x_1,$$
$$r_2 : f[x_1] \longrightarrow x_2,$$
$$r_3 : f[x_1] \longrightarrow x_3,$$
$$r_4 : f[x_2, x_5, x_7] \longrightarrow x_4,$$
$$r_5 : f[x_2, x_5, x_7] \longrightarrow x_5,$$
$$r_6 : f[x_3, x_4, x_6] \longrightarrow x_6$$
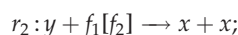
and

$$r_7 : f[x_3, x_4, x_6] \longrightarrow x_7$$

1.5. Relationships between RAFs, CAFs, pRAFs and related notions

We now describe the main differences among the notions of CAF, RAF and pRAF more informally. A CAF is a system of reactions that can start with the food set and build itself up in such a way that when a reaction occurs not only are all of its reactants available (either from the food set or as products of other earlier reactions) but at least one catalyst is as well. By contrast, a RAF can initially allow one or more reactions to proceed uncatalysed (and hence slowly); the requirement is simply that, eventually, all reactions in the RAF must be catalysed. A pRAF is slightly different; it is a subset of reactions where all the reactants and catalysts required are either in the food set or produced by some reaction. However, the system may not be able to build itself up from scratch (even slowly without catalysis) by just starting from the food set. In other words, once a pRAF exists, it can persist, but it may not be able to form in the first place, starting from just the food set, because some reactants for a reaction may not be available (as they can only be produced by subsequent reactions). Nevertheless, pRAFs may still play a role in the origin of life-like systems.

For example, consider a chemical reaction system comprising the reactions $\{r_1, r_2\}$, where:

$$r_1 : x + f_1[f_2] \longrightarrow y$$

and

$$r_2 : y + f_1[f_2] \longrightarrow x + x;$$

together with a food set $F = \{f_1, f_2\}$. This system is a pRAF, and so it cannot form just from this food set (this conclusion does not depend on the stoichiometry in r_2 where two copies of x are produced rather than one, however the following argument does depend on this stoichiometry). Now suppose that x becomes available from some environmental source (and so is added to the food set F) and then this source disappears again (perhaps due to some transient biochemical or geological event). In that case, the pair of reactions $\{r_1, r_2\}$ is able to generate an increasing supply of molecule type x for this pRAF to continue indefinitely. We plan to test for this type of scenario in future work.

The distinction between CAFs and RAFs may seem subtle but it has significant impacts: firstly, CAFs typically require much higher rates of catalysis and/or a much richer food set to form in model systems [21]. Furthermore, systems can have a very large number of subRAFs, whereas the number of CAFs is typically small, so the population of RAFs in a system is generally much richer than that of CAFs. Accordingly, we often focus more on RAFs, particularly for questions involving origins, where catalysts may have initially been rare (precluding CAFs) and where no other existing life was present to kick-start metabolism (precluding pRAFs).

RAFs are related to Robert Rosen's (M,R) systems (as described in [22]; see also [23]), and have a connection to 'chemical organization theory' [24]. This second connection arises because if R' is F-generated (e.g. a RAF or a CAF) and if we assume that the food molecules are being freely made available from some source, the closed set of molecules $\mathrm{cl}_{R'}(F)$ has an assignment of strictly positive reactions rates $\mathbf{v}$ so that $S\mathbf{v} \geq \mathbf{0}$, where S is the stoichiometric matrix for the system (i.e. each non-food molecule is generated at least as fast as it is used up), by lemma 4.1 of [17]. Other dynamical aspects of RAFs concerning reaction rates have been explored in [15,20], and related dynamics have been explored recently by [25,26].

2. Applications and observations

2.1. Application
In [3], the metabolism of two ancient prokaryotic species was explored for the presence of maxRAFs. Here, we use the network of one of those species, *Methanococcus maripaludis*, to explore the differences between the maxRAF, maxCAF and max-pRAF. The food set is FS2 detailed in table S1 of [3], including inorganic molecules, inorganic catalysts and abiotic organic carbon — with the addition of nicotinamide adenine dinucleotide (NAD) as the sole organic catalyst (44 molecule types in total). We chose NAD as it was the organic catalyst with the highest impact on the maxRAF size. The network is the same as reported for *M. maripaludis* in [3] with 965 reactions including pooling reactions (operational reactions that equate synonymous cofactors). The sizes for maxRAF, maxCAF and max-pRAF reported next (produced by the software CatlyNet [14]) also include pooling reactions.

With this new food set, which tests the network in an abiotic setting where NAD was generated by the environment, we obtain a maxCAF with 12 reactions using 20 food molecules, and a maxRAF with 84 reactions using 34 food molecules (electronic supplementary material, Part 2). Note the large maxRAF expansion that NAD alone allows for (one order of magnitude), as with the same food set without NAD, the maxRAF for *M. maripaludis* consisted of eight reactions only (see table S1 in [3]). We note that in this experiment, the maxCAF allows for the production of two amino acids, L-alanine and L-cysteine, and the maxRAF allows for the production of those and also L-aspartate. It is also important to note that the max-pRAF (which has size 540 reactions using 38 food molecules) is far from the size of the full network.

2.2. Visualisation and exploration
One way to visualize a RAF is as a graph in which the nodes represent reactions and there is an edge from reaction r_1 to r_2 if r_2 'depends' on r_1. Here, there are two ways to define 'depend':

(i) r_2 requires a product from r_1 as an input reactant or input catalyst, or

(ii) r_2 requires a product from r_1 as an input reactant.

We call (i) a *dependency graph* and (ii) a *reactant dependency graph*. Figure 2 shows these graphs for the archaeal dataset described in the previous example (a maxRAF of size 84, containing a maxCAF of size 12, with a food set of size 44).

2.3. Bidirectional reactions and catalysis options
One can easily extend the definition of a CRS and RAFs (and CAFs) to allow some reactions in R to be bidirectional (i.e. reversible) by regarding such reactions to be a pair of (ordinary) directed reactions, a feature common in metabolic networks [3]. For example, a reaction r such as $r : a + b \ [x, y] \longleftrightarrow c + d$ can be regarded as the pair of reactions $\{r_+, r_-\}$, where r_+ is the forward reaction $a + b \ [x, y] \longrightarrow c + d$ and r_- is the backward reaction $c + d \ [x, y] \longrightarrow a + b$.

Given a CRS $\mathcal{Q} = (X, R, C, F)$ where R includes one or more bidirectional reactions, let $R^{\pm}$ be the set of (ordinary, directed) reactions in which each bidirectional reaction $r \in R$ is replaced by r_+ and r_-. For any subset R_1 of R, define $R_1^{\pm}$ similarly. Let $\mathcal{Q}^{\pm} = (X, R^{\pm}, C^{\pm}, F)$ be the corresponding CRS, where $C^{\pm}$ is the catalysis assignment in which a catalyst x of r is a catalyst of r_+ and r_-.

If $\{r_+\}$ or $\{r_-\}$ is a RAF for $\mathcal{Q}^{\pm}$, then it can readily be checked that $\{r_+, r_-\}$ is a RAF for $\mathcal{Q}^{\pm}$. However, if $\{r_+, r_-\}$ is a RAF for $\mathcal{Q}^{\pm}$, it is possible that either $\{r_+\}$ or $\{r_-\}$ may fail to be a RAF for $\mathcal{Q}^{\pm}$. To see this, consider the following forward reaction: $r_+ : f_1 + f_2, \ [x] \longrightarrow g$, where $f_1, f_2 \in F$, $g \in F$ and x is either in F or equal to g. In this case, $\{r_+\}$ and $\{r_+, r_-\}$ are both RAFs but $\{r_-\}$ is not. Nevertheless, it is easily seen that at least one of these two singleton sets must be a RAF for $\mathcal{Q}^{\pm}$. The following proposition generalizes this observation; its proof is provided in the electronic supplementary material.

Proposition 2.1. *Let* $\mathcal{Q} = (X, R, C, F)$, *where* R *includes one or more bidirectional reactions and let* R_1 *be a subset of* R. *Then* $R_1^{\pm}$ *is a RAF (respectively, CAF) for* $\mathcal{Q}^{\pm}$ *if and only if there exists a subset* R_1' *of* $R_1^{\pm}$ *that is a RAF (respectively, CAF) for* $\mathcal{Q}^{\pm}$ *and satisfies the following condition:*

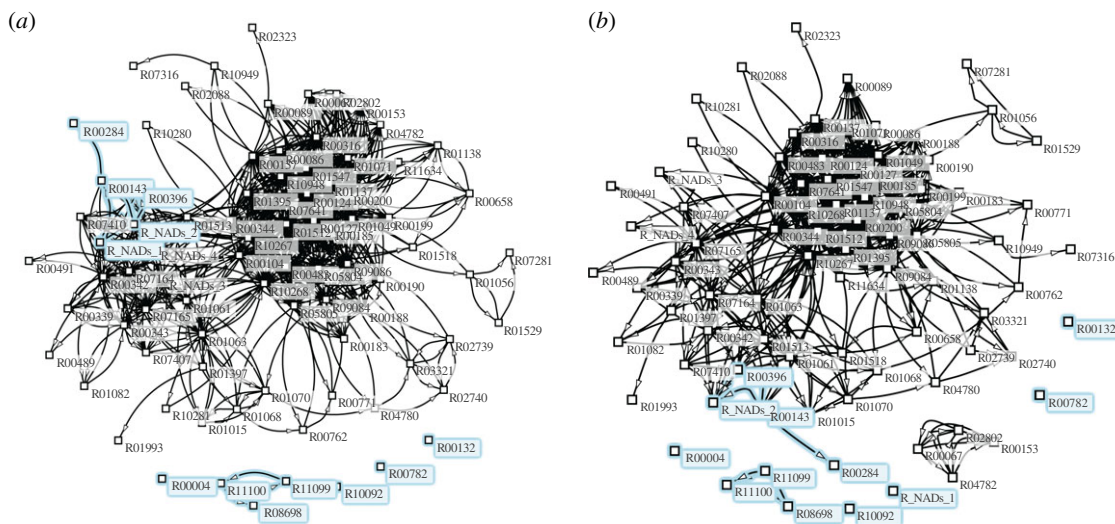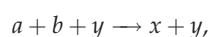$|R_1' \cap \{r_+, r_-\}| = 1$ *for all bidirectional reactions* $r \in R_1$.

(a)

(b)

Figure 2. The dependency graph (a) and the reactant dependency graph (b) for the maxRAF of size 84 of the archaea dataset described in §2.1 (which involved 965 reactions). In both cases, the graphs have a vertex set equal to the reactions in the maxRAF, and the edges are directed with an arc from r to r' whenever a product of r is either a reactant or a catalyst of r' (for the dependency graph) or when a product of r is a reactant of r' (for the reactant dependency graph). Reactions also present in the maxCAF are highlighted in blue. Figures were generated using *CatlyNet*.

Moreover, if R_2 is nested between R'_1 and R_1 (i.e. $R'_1 \subseteq R_2 \subset R_1$), then R_2 is a RAF (respectively, CAF) for $Q^{\pm}$.
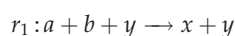
We can now extend the definition of RAFs (and irrRAFs) to CRS systems in which none, some or all the reactions are bidirectional. We say that a non-empty subset R_1 of R is a *RAF for Q* if $R_1^{\pm}$ is a RAF for $Q^{\pm}$ (which, by proposition 2.1, is equivalent to the condition that each bidirected reaction in R_1 can be replaced by either the forward or backward reaction to give a RAF for $Q^{\pm}$). Furthermore, we say that such a RAF R_1 is an *irrRAF for Q* if no strict subset of R_1 is a RAF. Note that a single bidirectional reaction that is a RAF is necessarily an irrRAF (even though it may be the case that one of the directed versions of the reaction is a RAF). Analogous definitions apply for CAFs.

Note that in counting the number of reactions in a RAF, CAF or pRAF, bidirectional reactions are counted as a single reaction (rather than two). We end this section with a remark. A main reason that RAF theory is based on the notion of directed rather than bidirectional reactions is to allow greater generality in theory; for example, to certain RAF settings outside chemistry, and to situations in biochemistry where either the rate laws are absent or unknown, or the reaction proceeds overwhelmingly in one direction. A second reason is that including bidirectional reactions can be easily described within this directional framework, as we have seen, if one were to take bidirectional reactions as a primitive notion to found RAFs on, then it is problematic to describe how RAF theory would handle directed reactions without explicitly invoking further additional notions and information such as reaction rates.
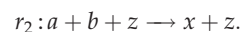
Distinctions between catalysts and reactants: often, a catalyst is represented by a molecule type that appears as both a reactant and a product. For example, the catalysed reaction $r: a + b\ [y] \longrightarrow x$ could be viewed as
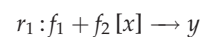
$$a + b + y \longrightarrow x + y,$$

whereas the catalysed reaction $r: a + b\ [y, z] \longrightarrow x$ could be represented by a pair of reactions

$$r_1: a + b + y \longrightarrow x + y$$
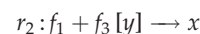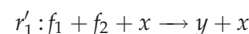
and

$$r_2: a + b + z \longrightarrow x + z.$$

By doing this expansion, we can easily see that a CRS Q contains a CAF if and only if the expanded reaction system contains a (non-empty) F-generated subset. However, the same is not true for RAFs. For example, consider the pair of reactions:
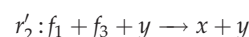
$$r_1: f_1 + f_2\ [x] \longrightarrow y$$

and

$$r_2: f_1 + f_3\ [y] \longrightarrow x$$

with food set $F = \{f_1, f_2, f_3\}$. Then $\{r_1, r_2\}$ is a RAF however, for the expanded version

$$r'_1: f_1 + f_2 + x \longrightarrow y + x$$

and

$$r'_2: f_1 + f_3 + y \longrightarrow x + y$$

the pair $\{r'_1, r'_2\}$ fails to be (or contain) an F-generated set. However, if one now adds the two additional (slow) reactions $r''_1: f_1 + f_2 \longrightarrow y$, and $r''_2: f_1 + f_3 \longrightarrow x$ to r'_1 and r'_2, then the resulting system of four reactions now becomes an F–generated set.

Allowing some reactions to not require catalysis, and complex catalysis rules: the requirement in a RAF/CAF/pRAF that *all* reactions be catalysed is sometimes overly severe (including in metabolism [3]). However, it is no problem to extend the definition of RAFs, CAFs and pRAFs to allow certain pre-specified reactions to not require catalysis. Formally, the easiest way to do this within the existing framework (and so that all the results, statements and theorems derived for RAFs, CAFs and pRAFs remain true) is simply to replace each pre-specified reaction r that does not require catalysis by adding (x, r) to the catalysis assignment C, for some molecule type x, which could be chosen from $\rho(r)$ or from the existing food set; alternatively, as in [3], x can be added as a formal catalyst for r to the food set.

Complex catalysis rules allow for reactions to not only be catalysed by molecule types, but also combinations of two or more molecule types (provided they are all present). This can be incorporated into the standard setting (of simple catalysis) by introducing additional auxillary reactions (for details, see [15]).

The role of the food set: given a CRS $\mathcal{Q} = (X, R, C, F)$, suppose we replace all of the food elements by a single element (call it f), and replace the set of food reactants and food catalysts of each reaction by $\{f\}$, while leaving catalysis unchanged otherwise. This simplified CRS $\mathcal{Q}'$ has RAFs, CAFs and pRAFs that correspond exactly to those of $\mathcal{Q}$. While this simplification can be useful, there are certain questions for which the details of the food set and its role in reactions becomes important. For example, two topical questions in early metabolism are the following: (i) given a CRS $\mathcal{Q} = (X, R, C, F)$, what is a largest number of elements of the food set that one can remove so that $\mathcal{Q}$ still has a RAF? (ii) Which elements of the food set have the greatest impact on the size of the RAF obtained? It can be shown that question (i) is NP-hard (by a reduction from the NP-complete problem SET COVER). The related question of how much of the food set can be removed so as to not alter the maxRAF was considered in [27], where it was shown to be also NP-hard.

3. Structural properties of RAFs, CAFs and pRAFs

We begin this section with a definition and some preliminary observations.

Given a CRS $\mathcal{Q} = (X, R, C, F)$, let 2^R denote the power set (the collection of all subsets) of R and let $\varphi_{\mathrm{RAF}} : 2^R \to 2^R$ be the function defined as follows: for any subset R' of R, $\varphi_{\mathrm{RAF}}(R')$ is the (unique) maximal RAF of $\mathcal{Q}$ that is entirely contained within R', provided that R' contains a RAF for $\mathcal{Q}$; otherwise, $\varphi_{\mathrm{RAF}}(R') = \emptyset$. Formally stated

$$\varphi_{\mathrm{RAF}}(R') = \bigcup \{R'' \subseteq R' : \forall r \in R'', \rho(r) \subseteq \mathrm{cl}_{R''}(F) \text{ and } \exists x \in \mathrm{cl}_{R''}(F) : (x, r) \in C\}.$$

Note that the function $\varphi = \varphi_{\mathrm{RAF}}$ satisfies the following three 'interior operator'[2] properties: For all $R' \subseteq \varphi(R)$

(I1): $\varphi(R') \subseteq R'$,
(I2): $R' \subseteq R'' \Rightarrow \varphi(R') \subseteq \varphi(R'')$ and
(I3): $\varphi(\varphi(R')) = \varphi(R')$.

An immediate consequence of (I_1) and (I_2) is that $\varphi_{\mathrm{RAF}}(R')$ is contained in the intersection of R' and the maxRAF of R. One can define φ_{CAF} and φ_{pRAF} in a similar manner, and these also satisfy the three interior properties.

3.1. Intersection systems involving RAFs, CAFs and pRAFs

A question that arose in the analysis of [3] is: when is the maxRAF of the intersection of two sets of reactions the same as the intersection of the two maxRAFs? We explore this question formally, beginning with a further definition. Recall that for a CRS $\mathcal{Q} = (X, R, C, F)$, and a subset R' of R, $\varphi_{\mathrm{RAF}}(R')$ denotes the subset of R consisting of the maxRAF of $\mathcal{Q}$ when R is replaced by R', provided that this maxRAF exists; otherwise $\varphi_{\mathrm{RAF}}(R') = \emptyset$. Note that $\varphi_{\mathrm{RAF}}(R')$ is always

a subset of R'. The proof of the following theorem is given in the electronic supplementary material.

Theorem 3.1. *Let $\mathcal{Q} = (X, R, C, F)$ be a CRS that has a RAF.*

(a) *For any two subsets R_1 and R_2 of R (which are not necessarily RAFs), the following identity holds:*

$$\varphi_{\mathrm{RAF}}(R_1 \cap R_2) = \varphi_{\mathrm{RAF}}(\varphi_{\mathrm{RAF}}(R_1) \cap \varphi_{\mathrm{RAF}}(R_2)).$$

(b) *The maxRAF of the intersection of two sets of reactions is a subset of the intersection of the two maxRAFs; moreover, the inclusion is strict if and only if the intersection of the two max-RAFs is not a RAF.*

Application: let R_A and R_B be the subsets of reactions in the metabolic network of ancient archaea and bacteria for the CRS $\mathcal{Q} = (X, R, C, F)$ investigated in [3]. As mentioned in §1.1, reactions common to both archaea and bacteria provide candidates for reactions that were present closer to the origin of life before these two domains separated. Here, $|R_A| = 965$ and $|R_B| = 1238$ and the food set has size 68 (including all organic catalysts). In this case, $\varphi(R_A)$ (the maxRAF of the CRS with the archaea reaction set R_A) has size 221, and $\varphi(R_B)$ (the maxRAF for the CRS with the bacterial reaction set R_B) has size 411. Both these maxRAFs are also maxCAFs. Moreover, the intersection of these two maxRAFs (i.e. $\varphi(R_A) \cap \varphi(R_B)$) has size 184, and the maxRAF for this set of reactions (i.e. $\varphi(\varphi(R_A) \cap \varphi(R_B))$) contains a maxRAF of size 131. By theorem 3.1(a), this maxRAF of size 131 is also the maxRAF of $R_A \cap R_B$.

The question now arises as to why this maxRAF is more than 50 reactions smaller than the intersection of the two maxRAFs $\varphi(R_A) \cap \varphi(R_B)$. Here, theorem 3.1(b) is relevant. The max-pRAF of $R_A \cap R_B$ has size 178, which is just six ($= 184–178$) reactions fewer than the size of the intersection of the two maxRAFs (so most reactions in the intersection are catalysed by either a product of intersection or an element of the food set). This suggests that the failure of the F-generated condition may cause the greatest reduction. In other words, there are different reactions in the archaeal and bacterial networks that produce essential reactants for reactions in the intersection, reactants which are not in the food set. This result may reflect divergent evolution from a common ancestor or a redundancy already present at the origins of metabolism, and would benefit from further investigation.

3.2. Refining the notions of RAFs, CAFs and pRAFs

So far, RAF theory has paid minimal attention to thermodynamic and kinetic considerations. One way to extend the theory in this direction is the following, which generalizes an approach in [15]. Suppose that R' is a RAF for $\mathcal{Q}$ (a directly analogous treatment is possible also for CAF and pRAFs). For each reaction $r \in R'$, let $\nu(r, R')$ be an associated non-negative score. For example, if different catalysts for a reaction lead to different rates, then we could let $\nu(r, R')$ be the maximal catalysis rate of r across all catalysts that are produced by R' (or present in the food set). Note that such a scoring function satisfies the following monotonicity property:

$$r \in R' \subseteq R'' \Rightarrow \nu(r, R') \leq \nu(r, R''). \tag{3.1}$$

For $t \geq 0$, let $\mathrm{RAF}_{\nu,t}(\mathcal{Q})$ be the set of RAFs R' for $\mathcal{Q}$ for which:

$$\nu(r, R') \geq t \text{ for all } r \in R'. \qquad (3.2)$$

Thus for this interpretation of $\nu(r, R')$ in terms of rates, a RAF in $\mathrm{RAF}_{\nu,t}(\mathcal{Q})$ is one for which all reactions in R' are able to proceed at a rate of at least t. However, one could also consider other types of functions ν that satisfy the monotonicity property (3.1).

The following theorem provides an immediate algorithm that is polynomial-time (in the size of $\mathcal{Q}$) that determines whether or not $\mathcal{Q}$ has a RAF that satisfies condition (3.2) and, if so, constructing a unique maximal one, provided that ν satisfies condition (3.1). The proof is provided in the electronic supplementary material.

Theorem 3.2. *Let $\mathcal{Q} = (X, R, C, F)$ be a CRS with a RAF, and consider any scoring function ν that satisfies condition (3.1). If $\mathrm{RAF}_{\nu,t}(\mathcal{Q}) \neq \emptyset$ then $\mathrm{RAF}_{\nu,t}(\mathcal{Q})$ has a unique maximal element $R_{\nu,t}(\mathcal{Q})$ and this set is the terminal set R_k of the following nested decreasing sequence of subsets: $R = R_0 \supset R_1 \supset \cdots \supset R_k(=R_{k+1})$, where*

$$R_{i+1} = \varphi_{\mathrm{RAF}}(\{r \in R_i : \nu(r, R_i)) \geq t\}),$$

for $i = 0, \ldots, k-1$. On the other hand, if $\mathrm{RAF}_{\nu,t}(\mathcal{Q}) = \emptyset$, then the nested decreasing sequence R_i terminates with $R_k = \emptyset$.

3.3. A generalization, applying theorems 3.1 and 3.2

Given a CRS $\mathcal{Q}$, suppose that we have a collection of functions $(\varphi_\alpha : 2^R \to 2^R; \alpha \in \mathcal{A})$ that satisfy the interior operator properties (I_1)–(I_3) described earlier. Consider the following partial order $\preceq$ on $\mathcal{A}$, defined by:

$$\alpha \preceq \alpha' \Leftrightarrow \text{ for all } R' \subseteq R \text{ one has } \varphi_\alpha(R') \subseteq \varphi_{\alpha'}(R'). \qquad (3.3)$$

For example, we could take $\mathcal{A} = \{0, \mathrm{RAF}, \mathrm{CAF}, \mathrm{pRAF}, 1\}$, where 0 and 1 refer to the functions $0(R') = \emptyset$ and $1(R') = R'$ for all $R' \in R$. For this choice of $\mathcal{A}$, we obtain a total ordering

$$0 \preceq \mathrm{CAF} \preceq \mathrm{RAF} \preceq \mathrm{pRAF} \preceq 1.$$

A larger class is obtained by considering $\mathrm{RAF}_{\nu,t}(\mathcal{Q})$ as described in §3.2, where ν satisfies condition (3.1) and with t fixed. In that case, consider the function $\varphi_{\mathrm{RAF}_{\nu,t}}$ that maps each subset R' of R to the unique maximal RAF contained in R' that satisfies condition (3.2) (or to the empty-set if no such RAF exists). Such a function is well defined by theorem 3.2 and it satisfies properties (I_1)–(I_3). Similarly, we can consider $\mathrm{CAF}_{\nu,t}(\mathcal{Q})$ and $\mathrm{pRAF}_{\nu,t}(\mathcal{Q})$, where ν again satisfies condition (3.1),

We then have: $\mathrm{RAF}_{\nu,t'} \preceq \mathrm{RAF}_{\nu,t} \preceq \mathrm{RAF}_{\nu,0} = \mathrm{RAF}$ for all $0 \leq t \leq t'$, and similarly for CAFs and pRAFs. The collection $\mathcal{A} = \{0, 1\} \cup \bigcup_{t \geq 0} \{\mathrm{RAF}_{\nu,t}, \mathrm{CAF}_{\nu,t}, \mathrm{pRAF}_{\nu,t}, 1\}$ is then a partially ordered set that includes CAF, RAF and pRAF, and with $\mathrm{RAF}_{\nu,t} \preceq \mathrm{CAF}_{\nu,t} \preceq \mathrm{pRAF}_{\nu,t}$ for each given t.

We can now state and establish a generalization of theorem 3.1(a), the proof of which is provided in the electronic supplementary material.

Theorem 3.3. *Given a CRS $\mathcal{Q} = (X, R, C, F)$, let $(\varphi_\alpha, \alpha \in \mathcal{A})$ be any collection of RAF, CAF or pRAF operators (as described above) that satisfy conditions (I_1)–(I_3), and where $\mathcal{A}$ is partially ordered by (3.3). Let R_1 and R_2 be any two subsets of the reaction set R and let $\alpha, \beta, \beta' \in \mathcal{A}$ satisfy $\alpha \leq \beta$ and $\alpha \leq \beta'$. Then:*

(i) $\varphi_\alpha(\varphi_\beta(R_1)) = \varphi_\beta(\varphi_\alpha(R_1)) = \varphi_\alpha(R_1).$
(ii) $\varphi_\alpha(\varphi_\beta(R_1) \cap \varphi_{\beta'}(R_2)) = \varphi_\alpha(R_1 \cap R_2).$
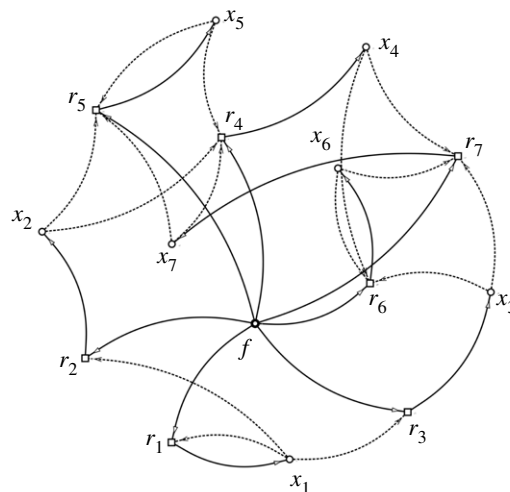

Figure 3. A seven-reaction system that comprises a RAF (from [20] based on [19]). As with figure 1, this RAF is a maxRAF (by default, since it is the entire set of reactions, though in general this is not necessary), and this maxRAF contains 65 RAF subsets (subRAFs), but no CAF. Dashed arrows indicate catalysis. Figure produced by *CatlyNet*.

Remark. Theorem 3.1(a) corresponds to the case $\alpha = \beta = \beta' = \mathrm{RAF}$ in part (ii) of theorem 3.3. As another (typical, but randomly chosen) application of theorem 3.3(ii), the maxCAF of the intersection of the R_1 and the pRAF of R_2 is always the same as the maxCAF of the intersection of the maxRAF of R_1 and the maxRAF of R_2 (both are equal to the maxCAF of $R_1 \cap R_2$).

3.4. Closed RAFs, and quotient RAFs

Let $\mathcal{Q} = (X, R, C, F)$ be a CRS and let R' be a non-empty subset of R. We say that R' is a *closed* subset of R if it has following property: for each reaction r in R (the full reaction set), if all of the reactants of r and at least one catalyst of r are present in $F \cup \pi(R')$ then r is present in R'. The *closure* of a R', denoted $\overline{R'}$, is the intersection of all closed subsets of R containing R'. Since the full reaction set R is (trivially) closed, the definition of closure is well defined and $\overline{R'}$ is closed (indeed, $\overline{R'} = R'$ if and only if R' is closed).

Let $\mathcal{C}[\mathcal{Q}]$ denote the set of closed non-empty subsets of reactions in $\mathcal{Q}$ and let $\mathcal{C}_{\mathrm{RAF}}[\mathcal{Q}]$ denote the set of all closed RAFs for $\mathcal{Q}$. Define $\mathcal{C}_{\mathrm{CAF}}[\mathcal{Q}]$ and $\mathcal{C}_{\mathrm{pRAF}}[\mathcal{Q}]$ similarly. Closed RAFs correspond to a particular type of 'chemical organization' within the framework of 'chemical organization theory' [28] and also play a central role in the recent semigroup-based approach of [29]. For the experimental system shown in figure 3, this RAF contains one other closed RAF (namely $\{r_4, r_5, r_6, r_7\}$) and 65 other RAFs that are not closed.

When $\mathcal{Q}$ has a RAF (respectively CAF or RAF), then the maxRAF (respectively, maxCAF or max-pRAF) is closed. The maxRAF may contain many closed RAFs as strict subsets, however, determining whether the maxRAF contains a closed RAF as a strict subset has recently been shown to be NP-hard [30]. By contrast, the maxCAF contains no other closed CAF. Indeed, $\mathcal{Q}$ has a CAF if and only if $\mathcal{C}[\mathcal{Q}] \neq \emptyset$, in which case $\mathcal{C}[\mathcal{Q}] = \mathcal{C}_{\mathrm{CAF}}[\mathcal{Q}]$, which consists just of the maxCAF for $\mathcal{Q}$. It can be shown that a closed RAF R' for a CRS $\mathcal{Q}$ is fully determined by $\mathcal{Q}$ and $\rho(R) \cup \pi(R')$ and can be reconstructed from this set ([31], lemma 1).

A *minimal closed RAF* is a closed RAF that contains no closed RAF as a strict subset. A particular example of a minimal closed RAF is a closed irrRAF.

Note that a CRS that has a RAF contains at least one minimal closed RAF (since the maxRAF is a closed RAF); however, the CRS may not contain a closed irrRAF. Note also that the closure of an irrRAF for Q is not necessarily a minimal closed RAF for Q. Consider, for example, the CRS consisting of $r_1: f_1 + f_2 [x] \longrightarrow x$ and $r_2: f_2 + f_3 [f_2] \longrightarrow y$ (with $F = \{f_1, f_2, f_3\}$) and the irrRAF $\{r_1\}$, the closure of which is $\{r_1, r_2\}$; this is not a minimal closed RAF, since it contains a smaller closed RAF, namely $\{r_2\}$. However, the opposite containment holds as we now state (the short proof is in the electronic supplementary material).

Proposition 3.4. *Let R' be a RAF for $Q = (X, R, C, F)$. If R' is a minimal closed RAF for Q, then R' is the closure of an irrRAF of Q. Moreover, R' equals the closure of any irrRAF contained in R'.*

Note that every subset RAF R' is contained in a unique minimal closed RAF, called the closure of R', denoted $\overline{R'}$ and defined by

$$\overline{R'} = \bigcap \{R'' \in \mathcal{C}_{RAF}[Q] : R' \subseteq R''\}.$$

This definition extends to the setting discussed earlier where the CRS Q contains bidirectional reactions. In that case, we say that a subset R_1 of reactions is a *closed RAF* for Q if $R_1^{\pm}$ is a closed RAF for $Q^{\pm}$. It can be shown that if R_1 is a closed RAF for a CRS Q that contains bidirectional equations, then: $|R_1^{\pm} \cap \{r_+, r_-\}| = 2$ for all bidirectional reactions $r \in R_1$.

3.5. Quotient RAFs

This idea of looking at quotient structures in RAF theory is motivated in part by [32], where techniques from algebra were suggested as an approach for deriving a coarse-grain description of complex biochemical reaction networks. However, the construction of a quotient here is somewhat different. Given a CRS $Q = (X, R, C, F)$ and a subset R' of R, let Q/R' be the CRS obtained from Q by deleting R' from R (i.e. replacing R by $R \setminus R'$) and adding all products of all reactions from R' into F. The proof of the following result is given in the electronic supplementary material.

Theorem 3.5. *Let $Q = (X, R, C, F)$ be a CRS that has a RAF, and let $R' \subsetneq R$.*

(a)

(i) *If R'' is a RAF for Q with $R' \subsetneq R''$, then $R'' \setminus R'$ is a RAF for Q/R'.*

(ii) *If R' is a RAF for Q and R^* is a RAF for Q/R', then $R' \cup R^*$ is a RAF for Q.*

(iii) *If R' is closed, then Q/R' has no CAF.*

(b)

If R' is a RAF for Q, then $\mathrm{maxRAF}(Q/R') = \mathrm{maxRAF}(Q) \setminus R'$.

We refer to a RAF $R'' \setminus R'$ in theorem 3.5(a–i) as a quotient RAF for the (quotient) CRS Q/R'. Note that $R'' \setminus R'$ is not necessarily a RAF for Q; instead, it is a set of reactions that can be added to a RAF to create a larger RAF (called a 'co-RAF' in [17] or 'periphery' in [33]).

A particular case of interest is where R'' is the maxRAF for Q and R' is the maxCAF for Q. In that case, part (b) states that the maxRAF of $Q/\mathrm{maxCAF}(Q)$ is obtained from the maxRAF of Q by deleting the maxCAF of Q (moreover, $Q/\mathrm{maxCAF}(Q)$ has no CAF by part (a–iii)). For the global anaerobic prokaryotic metabolism dataset (consisting of 6029 reactions, from the study in [3]), this has a maxRAF of size 580 and a maxCAF of size 239, so the quotient RAF (taking R' to be the maxCAF and R'' to be the maxRAF) has size $580 - 239 = 314$.

4. Special types of RAFs and reactions

In this section, we introduce and explore two new notions in RAF theory.

4.1. Core RAFs

For any CRS Q, the set of its RAFs forms a partially ordered set (poset) under subset inclusion, with a unique maximal element, namely the maxRAF. The minimal elements of this poset are the irrRAFs; in general, there may be (exponentially) many of these. A natural question is: when does Q have a unique minimal RAF (i.e. a RAF R' that a subset of every other RAF for Q)? We call such a RAF, when it exists, a *core RAF* (this is different from the notion of 'core' in [33] which is closer to irrRAF). Clearly, if Q has a core RAF, then it has only one. Furthermore, a core RAF for Q will be the unique smallest RAF for Q, so a first approach might be to develop an algorithm to find the smallest RAF in a CRS. However, this problem turns out to be NP-hard in general [17], so we need an alternative strategy. Note also that a core RAF for Q exists if and only if Q has exactly one irrRAF; however, there is no efficient algorithm known for counting the number of irrRAFs. Nevertheless, the following result provides a fast way to determine whether or not Q has a core RAF and, if so, constructing it. One simply computes the maxRAF of the set of those reactions r that are essential for any RAF to exist.[3] The proof of the theorem 4.1 is provided in the electronic supplementary material.

Theorem 4.1. *Let $Q = (X, R, C, F)$ be a CRS with a RAF, and let*

$$R^- = \{r \in \varphi_{RAF}(R) : \varphi_{RAF}(R \setminus \{r\}) = \emptyset\}.$$

Then Q has a core RAF if and only if R^- is a RAF for Q, in which case, R^- is the core RAF for Q. In particular, determining whether or not Q has a core RAF, and if so finding it, can be performed in polynomial time in the size of Q.

Applying theorem 4.1 to the archaea dataset described earlier (figure 2), reveals that no core RAF is present.

4.2. Detecting spontaneous reactions in RAFs

A fundamental observation is that a reaction cannot proceed until all its reactants are available, whereas if a catalyst for the reaction is absent, the reaction may still proceed, albeit slowly, and later speed up when a catalyst becomes available. We formalize this as follows. Consider any CRS $Q = (X, R, C, F)$. Let R' be a RAF for Q (e.g. the maxRAF, or some sub-RAF). An admissible ordering for R' is a linear ordering o of the

reactions of R', $o = (r_1, r_2, \ldots, r_k)$, for which (i) all the reactants of r_1 are present in the food set, and (ii) for each $i \geq 2$, each reactant of r_i is either present in the food set or is produced as the product of at least one earlier reaction. In other words, if we let $R_i = \{r_1, \ldots, r_i\}$, for $1 \leq i \leq k$, then $\rho(r_i) \subseteq \mathrm{cl}_{R_{i-1}}(F)$ for $i = 1, \ldots, k$ where $\mathrm{cl}_{R_0}(F) = F$ and where, for $j > 1$, $\mathrm{cl}_{R_j}(F)$ is the union of the set $\mathrm{cl}_{R_{j-1}}(F)$ and the set of products of reaction r_j that have all their reactants in $\mathrm{cl}_{R_{j-1}}(F)$. A basic result concerning admissible orderings is the following result (from lemma 3.1 of [17]).

Proposition 4.2. *For any CRS $\mathcal{Q} = (X, R, C, F)$, a subset R' of R has an admissible ordering if and only if R' is F-generated. In particular, every RAF has an admissible ordering.*

Given an admissible ordering $o = (r_1, r_2, \ldots, r_k)$ for R', we say that r_i *starts uncatalysed* in o if none of the catalysts for reaction r_i is present in $\mathrm{cl}_{\{r_1, \ldots, r_{i-1}\}}(F)$; otherwise, we say that r_i *starts catalysed*. For example, consider the following two systems with food set $F = \{a, b, c\}$, both of which are themselves RAFs.

system 1	system 2
$r : a + b \; [x] \longrightarrow y$	$r : a + b \; [x] \longrightarrow x$
$r' : a + c \; [y] \longrightarrow x$	$r' : a + c \; [x] \longrightarrow y$

Both systems admit the two possible admissible orderings (r, r') and (r', r). For system 1, r starts uncatalysed for the ordering (r, r') but not for (r', r); for system 2, r starts uncatalysed under both orderings.

Given an arbitrary CRS $\mathcal{Q}$, a RAF R' for $\mathcal{Q}$ and variable integer k, a natural question asks whether R' has an admissible ordering in which the number of reactions that start uncatalysed is at most k. This problem is called k–*CAF RAF* (since $k = 0$ is the condition for R' to form a CAF) and this problem was shown to be NP-complete in [34].

We now describe a polynomial-time algorithm to solve the k–CAF RAF problem when it is restricted to RAFs that have the property that each reaction has all its reactants in the food set (the so-called elementary setting of [15]). To describe this, we need to introduce some further notation. Given a CRS $\mathcal{Q}$ and a RAF R' for $\mathcal{Q}$ that has all its reactants present in the food set, let (R', A) be the directed graph on vertex set R' where there is an arc from r to r' if $r \neq r'$ and at least one product of r catalyses r'. Let R'_F be the set of reactions in R' that have an element of the food set as a catalyst. Remove from R' all reactions that are reachable by a directed path from a vertex in R'_F and let $\mathcal{G}(R')$ be the resulting digraph, and $\chi[\mathcal{G}(R')]$ its associated condensation digraph (whose elements are the strongly connected components of $\mathcal{G}(R')$). Note that $\chi[\mathcal{G}(R')]$ is acyclic, and can be computed in polynomial time in the size of $\mathcal{Q}$. The proof of the following proposition is given in the electronic supplementary material.

Proposition 4.3. *Let $\mathcal{Q}$ be a CRS, and R' is RAF for $\mathcal{Q}$ in which every reaction has all its reactants in the food set. The smallest value of k for which R' is a k-CAT RAF is equal to the number of vertices of $\chi[\mathcal{G}(R')]$ of in-degree equal to 0. Moreover, it is necessary and sufficient that (any) one reaction in each such strongly connected component of $\chi[(\mathcal{G}(R')]$ starts uncatalysed.*

Application. For the seven-reaction RAF of the experimental system discussed earlier in §1.4 (figure 3), there is no reaction
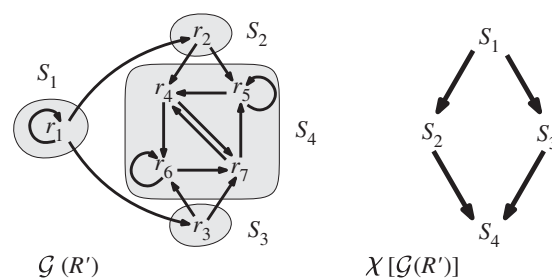

Figure 4. Left: The graph $\mathcal{G}(R')$ for the seven-reaction RAF R' of the experimental system from §1.4, with its four strongly connected components shaded. Right: The associated condensation digraph $\chi[\mathcal{G}(R')]$.

that is catalysed by the (single) element of the food set, and the associated graph $\mathcal{G}(R')$ described above consists of four strongly connected components ($S_1 - S_4$), as shown in figure 4. The associated condensation digraph $\chi[\mathcal{G}(R')]$ has just a single vertex of in-degree 0 (namely S_1) and so, by proposition 4.3, exactly one reaction (but no more) needs to be spontaneous in forming the original RAF. Since S_i consists of a single reaction, this means that r_1 must start uncatalysed.

For the remainder of this section, we consider the following simpler variation on k–CAF RAF.

— Given a RAF R' for $\mathcal{Q}$ and a reaction $r \in R'$, does every possible admissible ordering for R' require r to start uncatalysed?

It turns out there is a fast (polynomial-time) algorithm to solve this problem, which we now describe. We first introduce a further definition and some further notation. Given a CRS $\mathcal{Q} = (X, R, C, F)$, a RAF R' for $\mathcal{Q}$ and $r \in R'$, we say that r is *spontaneous* in R' if r starts uncatalysed for every admissible ordering for R'.

Given the pair (R', r), where R' is a RAF for $\mathcal{Q}$ and $r \in R'$, let $C(r, R')$ be the set of catalysts of r that are present in $\mathrm{cl}_{R'}(F)$. For $x \in C(r, R')$, let $r[x]$ be the reaction obtained from r by adding x as an additional reactant to r, without altering the products or catalysts of r. For example, the reaction $r : a + b \; [x, y] \longrightarrow z + w$ has $C(r) = \{x, y\}$, and $r[x]$ is the reaction: $r[x] : a + b + x \; [x, y] \longrightarrow z + w$. Finally, let $R'[r, x]$ denote the set of reactions obtained from R' by replacing the reaction r in R' by $r[x]$ (i.e. $R'[r, x] := (R \setminus \{r\}) \cup \{r[x]\}$). The main result of this section is the following (the proof is provided in the electronic supplementary material).

Theorem 4.4. *Let $\mathcal{Q} = (X, R, C, F)$ be a CRS with that RAF.*

(a) *Given any RAF R' for $\mathcal{Q}$ and reaction $r \in R'$, the following are equivalent:*

  (i) *r is spontaneous in R';*
  (ii) *$R'[r, x]$ is not F-generated for any $x \in C(r, R')$;*
  (iii) *The maxRAF of $(X, R'[r, x], C, F)$ is not equal to $R'[r, x]$ for any $x \in C(r, R')$.*

(b) *A reaction r is spontaneous in every RAF of $\mathcal{Q}$ containing r if and only if r is spontaneous in the maxRAF of $\mathcal{Q}$.*

Theorem 4.4 provides an algorithm to identify the spontaneous reactions. First, compute the maxRAF for $\mathcal{Q}$. Then for each reaction r in this maxRAF, test if r has a catalyst in the food set. If so, then r is not spontaneous. Otherwise, go through the remaining catalysts of r that are a product of a

reaction in maxRAF($\mathcal{Q}$) and for each such catalyst — say x — compute the maxRAF of $(X, \text{maxRAF}(\mathcal{Q})[r, x], C, F)$. If this equals maxRAF($\mathcal{Q}$)$[r, x]$ for any such x, then r is not spontaneous. Otherwise, r is spontaneous.
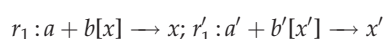
4.3. Application

We examined the archaea dataset described above with a maxRAF of size 84 and a maxCAF of size 12. Since the maxCAF is (considerably) smaller than the maxRAF, at least one spontaneous reaction is required. Applying the above algorithm, we find that a single reaction in this system is spontaneous; moreover, this single spontaneous reaction suffices to transform the maxCAF into the maxRAF. The reaction concerned converts NAD into another important organic catalyst, ATP. In cellular metabolism, this reaction is carried reversibly by the enzyme nicotinamide mononucleotide adenylyltransferase.[4] This is significant, as it points to a possible route for large catalytic expansion at the origins of metabolism and deserves further experimental investigation.
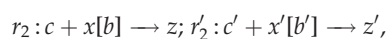
4.4. The impact of inhibition on RAF formation

So far, a CRS allows molecule types to catalyse a reaction. However, molecule types can also inhibit reactions. As with catalysis, inhibition can be regarded as a subset I of $X \times R$, where $(x, r) \in I$ indicates that molecule type x inhibits reaction r. Thus we describe an CRS with inhibition using a five-tuple (X, R, C, I, F). Following [21], given an arbitrary CRS with inhibition $\mathcal{Q} = (X, R, C, I, F)$, an *uninhibited RAF* (uRAF) is a RAF R' for $\mathcal{Q}$ that also has the property that no reaction in R' is inhibited by any product of R' or by any element of the food set. Note that if an uRAF exists, then there may be more than one maximal uRAF (in contrast to RAFs, where the maxRAF is the unique maximal RAF).

Although a polynomial-time maxRAF algorithm exists for determining whether or not a CRS has a RAF, the task of determining whether or not an arbitrary CRS with inhibition has an uRAF is NP-hard [21]. More recently, it has been shown [30] that the NP-hardness of determining the existence of uRAF also holds even when every reaction in R has all its reactants in the food set (i.e. the 'elementary' CRS setting [15]). Nevertheless, the following procedure provides a way to search for an uRAF: let R' be the set of reactions in the maxRAF of $\mathcal{Q}$ that are inhibited by either a product of the maxRAF or an element of the food set. Then compute the maxRAF of the resulting CRS obtained from $\mathcal{Q}$ by replacing R by R'. If this second maxRAF exists, it is a uRAF (such an approach was used to find uRAFs in [35]). In addition, when the number of inhibiting molecule types is bounded, there is a polynomial-time algorithm for determining whether or not uRAFs exist.

However, it seems natural to require that any uRAF R' should also be closed, since if some molecule x can be generated from a reaction r (not part of the uRAF) and the reactants and catalyst (but no inhibitor) of r is present in $F \cup \pi(R')$ then we should expect x to be generated and this molecule x might then inhibit some reaction within R' (thereby destroying it). We now describe a simple example to illustrate this, and discuss its consequences. Consider the following system of four reactions:

$$r_1 : a + b[x] \longrightarrow x; \; r'_1 : a' + b'[x'] \longrightarrow x'$$

and

$$r_2 : c + x[b] \longrightarrow z; \; r'_2 : c' + x'[b'] \longrightarrow z',$$

where $F = \{a, b, c, a', b', c'\}$. Suppose that z inhibits r'_1 and z' inhibits r_1. In that case, the set $\{r_1, r_2, r'_1, r'_2\}$ contains precisely two closed uRAFs (namely $\{r_1, r_2\}$ and $\{r'_1, r'_2\}$) however, their union fails to be a uRAF. Also, $\{r_1, r_1'\}$ is a uRAF but its closure $\{r_1, r_2, r'_1, r'_2\}$ fails to be a uRAF. This simple example highlights some key differences in the structure of uRAFs versus RAFs (where the union or closures of RAFs remain RAFs).

With the focus on closed uRAFs, there is a possible way to search heuristically for closed uRAFs by using irrRAFs. As mentioned earlier, irrRAFs can be sampled in polynomial time, and for each sampled irrRAF one can further compute its closure and test for inhibition in polynomial time. In this way, a closed irrRAF can be discovered, provided that the number of irrRAFs is not too large. The justification of the approach is based on the following result, established in the electronic supplementary material.

Proposition 4.5. *A CRS $\mathcal{Q}$ has a closed uRAF if and only if $\mathcal{Q}$ has an irrRAF R' for which no reaction in $\overline{R'}$ (the closure of R') is inhibited by any element of $F \cup \pi(\overline{R'})$.*

5. Concluding comments

In this paper, we have derived and described a number of new results concerning the structure of RAF sets and outlined methods for identifying these structures and other characteristics present in metabolic networks of interest in early biochemistry. Our emphasis is on describing properties and algorithms that can be efficiently implemented (i.e. they have polynomial rather than exponential running time) so that they can be applied to large databases, as well as techniques to visualize and simplify complex networks (such as the notion of quotient RAFs). Most techniques described have been implemented in open-source public software [14]. In future work, we plan to investigate the detailed structure of primitive archaea metabolic networks further, and explore the impact of inhibition on RAF formation, including variations on the strong form of inhibition described above, by allowing inhibitors to only partially nullify catalysis.

Endnotes

[1] Theorem 1(b) of [17].
[2] This terminology comes from topology.
[3] These are the reactions with an 'importance' index of 1.0, in *CatlyNet*.
[4] https://doi.org/10.1107/S1399004715015497.

References

1. Javaux EJ. 2019 Challenges in evidencing the earliest traces of life. *Nature* **572**, 451–460. (doi:10. 1038/s41586-019-1436-4)

2. 2007 On the origin of biochemistry at an alkaline hydrothermal vent. *Phil. Trans. R. Soc. B* **362**, 1887–1926. (doi:10.1098/rstb.2006.1881)

3. Xavier JC, Hordijk W, Kauffman S, Steel M, Martin WF. 2020 Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. B* **287**, 2019–2377. (doi:10. 1098/rspb.2019.2377)

4. Hordijk W, Steel M. 2017 Chasing the tail: the emergence of autocatalytic networks. *Biosystems* **152**, 1–10. (doi:10.1016/j.biosystems.2016.12.002)

5. Kauffman S. 1971 Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *J. Cybern.* **1**, 71–96. (doi:10.1080/01969727108545830)

6. Kauffman S. 1986 Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1–24. (doi:10.1016/S0022-5193(86)80047-9)

7. Kauffman SA. 1993 *The origins of order*. Oxford, UK: Oxford University Press.

8. Hordijk W. 2019 A history of autocatalytic sets. *Biol. Theory* **14**, 224–246. (doi:10.1007/s13752-019-00330-w)

9. Bollobás B, Rasmussen S. 1989 First cycles in random directed graph processes. *Discr. Math.* **75**, 55–68. (doi:10.1016/0012-365X(89)90078-2)

10. Jain S, Krishna S. 1998 Autocatalytic sets and the growth of complexity in an evolutionary model. *Phys. Rev. Lett.* **81**, 5684–5687. (doi:10.1103/PhysRevLett.81.5684)

11. Jain S, Krishna S. 2001 A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proc. Natl Acad. Sci. USA* **98**, 543–547. (doi:10.1073/pnas.98.2.543)

12. Cazzolla Gatti R, Kauffman S, Ulanowicz R. 2018 Niche emergence as an autocatalytic process in the evolution of ecosystems. *J. Theor. Biol.* **454**, 110–117. (doi:10.1016/j.jtbi.2018.05.038)

13. Gabora L, Steel M. 2017 Autocatalytic networks in cognition and the origin of culture. *J. Theor. Biol.* **431**, 87–95. (doi:10.1016/j.jtbi.2017.07.022)

14. Huson D, Steel M. 2020 *CatlyNet*. https://github.com/husonlab/catlynet.

15. Steel M, Hordijk W, Xavier JC. 2018 Autocatalytic networks in biology: structural theory and algorithms. *J. R. Soc. Interface* **16**, 20180808. (doi:10.1098/rsif.2018.0808)

16. Hordijk W, Steel M. 2004 Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* **227**, 451–461. (doi:10.1016/j.jtbi.2003.11.020)

17. Steel M, Hordijk W, Smith J. 2013 Minimal autocatalytic networks. *J. Theor. Biol.* **332**, 96–107. (doi:10.1016/j.jtbi.2013.04.032)

18. Letelier J-C, Soto-Andrade J, Abarzua FG, Cornish-Bowden A, Cárdenas ML. 2006 Organizational invariance and metabolic closure: analysis in terms of (M,R) systems. *J. Theor. Biol.* **238**, 949–961. (doi:10.1016/j.jtbi.2005.07.007)

19. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. 2012 Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72–77. (doi:10.1038/nature11549)

20. Hordijk W, Steel M. 2013 A formal model of autocatalytic sets emerging in an RNA replicator system. *J. Syst. Chem.* **4**, 3. (doi:10.1186/1759-2208-4-3)

21. Mossel E, Steel M. 2005 Random biochemical networks: the probability of self-sustaining autocatalysis. *J. Theor. Biol.* **233**, 327–336. (doi:10.1016/j.jtbi.2004.10.011)

22. Letelier JC et al. 2010 (MR) systems and RAF sets: common ideas, tools and projections. In *Proc. of the Alife XII Conference*. Odense, Denmark, pp. 94–100.

23. Cornish-Bowden A, Cárdenas ML. 2007 Organizational invariance in (M,R)-systems. *Chem. Biodiv.* **4**, 2396–2406. (doi:10.1002/cbdv.200790195)

24. Dittrich P, Speroni di Fenizio P. 2007 Chemical organisation theory. *Bull. Math. Biol.* **69**, 1199–1231. (doi:10.1007/s11538-006-9130-8)

25. Liu Y, Sumpter JT. 2018 Mathematical modeling reveals spontaneous emergence of self-replication in chemical reaction systems. *J. Biol. Chem.* **293**, 18854–18863. (doi:10.1074/jbc.RA118.003795)

26. Peng Z, Plum A, Gagrani P, Baum DA. 2020 An ecological framework for the analysis of prebiotic chemical reaction networks. *J. Theor. Biol.* **507**, 110451. (in press). (doi:10.1016/j.jtbi.2020.110451)

27. Sousa FL, Hordijk W, Steel M, Martin W. 2015 Autocatalytic sets in *E coli.* metabolism. *J. Syst. Chem.* **6**, 4.

28. Hordijk W, Steel M, Dittrich P. 2018 Autocatalytic sets and chemical organizations: modeling self-sustaining reaction networks at the origin of life. *New J. Phys.* **20**, 015011. (doi:10.1088/1367-2630/aa9fcd)

29. Loutchko D. 2019 Semigroup models for biochemical reaction networks. (http://arxiv.org/abs/1908.04642).

30. Weller-Davies O, Steel M, Hein J. 2020 Complexity results for autocatalytic network models. *Math. Biosci.* **325**, 108365. (doi:10.1016/j.mbs.2020.108365)

31. Smith J, Steel M, Hordijk W. 2014 Autocatalytic sets in a partitioned biochemical network. *J. Syst. Chem.* **5**, 2. (doi:10.1186/1759-2208-5-2)

32. Loutchko D. 2019 Algebraic coarse-graining of biochemical reaction networks. (http://arxiv.org/abs/1908.05483).

33. Vasas V, Fernando C, Santos M, Kauffman S, Szathmáry E. 2012 Evolution before genes. *Biol. Direct* **7**, 1. (doi:10.1186/1745-6150-7-1)

34. Hordijk W, Wills P, Steel M. 2014 Autocatalytic sets and biological specificity. *Bull. Math. Biol.* **76**, 201–224. (doi:10.1007/s11538-013-9916-4)

35. Hordijk W, Steel M. 2016 Autocatalytic sets in polymer networks with variable catalysis distributions. *J. Math. Chem.* **54**, 1997–2021. (doi:10.1007/s10910-016-0666-z)