Mathematical Biosciences 325 (2020) 108365

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Complexity results for autocatalytic network models

Oliver Weller-Davies [a], Mike Steel [b,*], Jotun Hein [a]

[a] Department of Statistics, Oxford University, Oxford, UK
[b] Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Keywords:
Catalytic reactions system
Origin of metabolism
Polymer
Computational complexity

ABSTRACT

A key step in the origin of life is the emergence of a primitive metabolism. This requires the formation of a subset of chemical reactions that is both self-sustaining and collectively autocatalytic. A generic approach to study such processes ('RAF theory') has provided a precise and computationally effective way to address these questions, both on simulated data and in laboratory studies. In this paper, we solve some questions posed in more recent papers concerning the computational complexity of some key questions in RAF theory. In particular, although there is a fast algorithm to determine whether or not a catalytic reaction network contains a subset that is both self-sustaining and autocatalytic (and, if so, find one), determining whether or not sets exist that satisfy certain additional constraints turns out to be NP-hard.

1. Introduction

The origin of life remains an unsolved challenge in science. Once considered a problem 'beyond science', many researchers now believe a solution may not be far off [1,2]. This prospect has partly been fuelled by recent efforts to integrate formal models and mathematical techniques into the field, particularly for understanding a key step in the origin of life, namely the origin of metabolism from early chemistry on Earth [3]. Initially motivated by the emergent qualitative properties of discrete random networks (such as early works of Paul Erdös and Alfred Rényi [4]), one approach to formally capture 'life-like' emergence in a chemical system involves the notion of a collectively autocatalytic set, a concept pioneered for polymer systems by Stuart Kauffman [5]. This approach was later developed more rigorously with *Reflexively Auto-catalytic F-generated sets* (RAFs). Such sets couple together two basic requirements for any living system: reactions are catalysed by molecules types (e.g. enzymes, cofactors etc.) generated from within the system, and second, every reaction within the system requires just molecule types that can be constructed from an ambient food source using only reactions within the system (i.e. it is self-sustaining from the chemistry of the environment).

In a series of papers beginning in 2000 to the present [6–8], RAFs have been investigated in both simulated and laboratory-based systems of early metabolism [9,10] as discussed in [11] as well as in the analysis of metabolism in the bacterium *Escherichia coli* [12] and in a recent study into ancient metabolism revealed by analysing large biochemical databases [13]. RAF theory has also been applied in fields such as ecology [14], economics [15] and cultural evolution [16], and has recently been re-expressed within the language of semigroups [17].

In this paper we derive a number of new results in RAF theory, answering some outstanding complexity questions that have been posed in earlier papers. We begin in Section 2 by providing basic definitions and summarising some results and questions from RAF theory.

In Section 3 we investigate the computational complexity of two problems (posed in [18] and [11]) concerning RAFs that are 'closed' (a biochemically relevant condition we describe there). Namely, (i) does a RAF set contain a closed subRAF set? (ii) Is there a closed RAF set that is 'uninhibited' (i.e. where no molecule in the set inhibits any reaction) in the simple case when (just) a single molecule type inhibits (just) one reaction? We show that both questions are NP-complete (even though Question (i) is easily solved without the 'closure' restriction).

In Section 4, we investigate the simpler 'elementary' CRS framework (where the reactants of each reaction are present in the food source) and solve two complexity problems (posed in [11,12]): (i) Does an uninhibited RAF exist? (ii) What is the size of the largest irreducible RAF? We show that these questions are NP-hard in the elementary CRS setting. We end with some brief concluding comments.

Our complexity results not only answer open questions posed previously in the literature, they are also directly relevant to the study of early metabolism as they indicate how certain analytical questions need to be handled. In particular, the investigation of large metabolic databases (such as those considered in [12] and [13]) relied on the existence of known fast (polynomial-time) and exact algorithms for finding the maximal RAF when it exists. However, in these studies it was not clear how to undertake more detailed analysis (e.g. searching for closed subRAFs). Our complexity results suggest that fast exact algorithms do not exist, and approaches based on integer linear programming or heuristics will be required instead.

* Corresponding author.
E-mail addresses: mike.steel@canterbury.ac.nz (M. Steel), hein@stats.ox.ac.uk (J. Hein).

https://doi.org/10.1016/j.mbs.2020.108365
Received 28 October 2019; Received in revised form 16 April 2020; Accepted 17 April 2020
Available online 30 April 2020

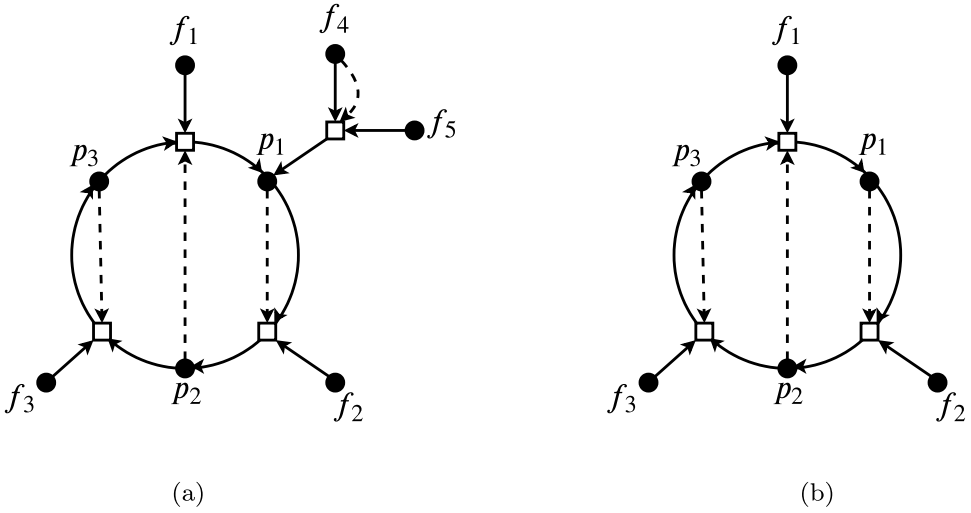


Fig. 1. The CRS in (a) contains four different RAFs, the largest of which (the maxRAF) being the set of all reactions; the CRS in (b) contains no RAFs as all reaction subsets in (b) fail to be F-generated. Here, food is indicated by f_*.

2. Definitions and mathematical preliminaries

RAFs are defined within the context of a *catalytic reaction system*. Formally, a catalytic reaction system (CRS) is a quadruple $Q = (X, R, C, F)$ where:

- X denotes a set of *molecule types*.
- R denotes a set of reactions between sets of molecule types. Each reaction r in R involves a subset A of X called the *reactants* of r; these reactants combine together (with various multiplicities) to produce a subset B of X called the set of *products* or r (with various multiplicities). In this paper, we need to distinguish reactions only up to the sets A, B (i.e. including multiplicities does not change any of the results) and so we write $r = (A, B)$ or more briefly $r = A \to B$. For example, the reaction $a + a \to aa$ would be written as $r = (\{a\}, \{aa\})$.
- $C \subseteq X \times R$ denotes a *catalyzation* assignment, where if $(x, r) \in C$, we say that the molecule type x *catalyses* the reaction r.
- $F \subseteq X$ denotes an ambient *food set* of molecule types, which are assumed to be freely available in the environment.

Often when drawing figures to represent CRS, we adopt the following convention: molecule types are represented by black circles, reactions by white squares, catalysis edges by dashed arrows and reactant/product edges by black arrows. Given a CRS $Q = (X, R, C, F)$, a subset of reactions $R' \subseteq R$ is a *RAF* set (or, more briefly, a RAF) for Q if R' is non-empty and both of the following conditions hold:

(RA) *Reflexively Auto-catalytic:* Every reaction $r \in R'$ is catalysed by a molecule type x that is either in the food set F or is the product of another reaction $r' \in R'$.

(F) *F-generated*: The reactions in R' can be written in a linear order $r_0, r_1, \ldots, r_n$ such that for every reaction $r_i = (A_i, B_i) \in R'$, each reactant $x \in A_i$ is either in the food set or is the product of another reaction occurring earlier in the ordering; that is, $\forall x \in A_i$, $x \in F$ or $x \in B_j$ for some $j < i$.

RAFs underlie the metabolism of both existing cellular life [12] and also arise in laboratory models of early life [9–11]. The concept of a RAF couples two features that seem to be essential in the earliest metabolism at the origin of life as well as in extant cellular life. Firstly, the reactions need to be catalysed by molecules present in the system (in modern metabolism, these catalysts are highly efficient, and are based on enzymes and cofactors, whereas in early metabolism it is likely that much simpler catalysts based on metals such as iron would be involved). Biochemical catalysts not only speed up reactions by many orders of magnitude but they also allow reactions to be synchronised [19]. Secondly, the system must be 'self-sustaining' from an available (external) food source; in other words, the reactants of each reaction in the system must either be in the food set or be derivable from the food set by a sequence of reactions within the system. These two features are combined into the two conditions (RA) and (F). The RAF concept is illustrated in Fig. 1(a), along with the weaker notion of a 'pseudo-RAF' (Part (b) of Fig. 1) that satisfies the (RA) condition but fails to be F-generated.

It is easily seen that the union of two or more RAFs is a RAF, and thus if a CRS has a RAF, it has a unique maximal one, called the maxRAF. Any given RAF R' may contain another RAF R'' as a strict subset, in which case we say that R'' is a *subRAF* of R'. Although it is not perhaps obvious from the above definition, it turns out that there is a (polynomial-time) algorithm for computing the maxRAF within a given CRS, if one exists at all [7]. We describe this maxRAF algorithm shortly.

2.1. The closure of a set of molecule types

Fix a CRS $Q = (X, R, C, F)$. Given a set of reactions $R' \subseteq R$ and a set of molecule types $X' \subseteq X$, the molecule closure $\mathrm{cl}_{R'}(X')$ is the unique minimal subset $W \subseteq X$ of molecule types satisfying the following two conditions: (i) $X' \subseteq W$; (ii) for every reaction $(A, B) \in R'$: $A \subseteq W \implies B \subseteq W$ (as defined in [7]).

In other words, $\mathrm{cl}_{R'}(X')$ denotes the set of molecule types arrived at if we were to continually apply reactions from R', wherever we could, ignoring catalysis constraints and starting only with molecule types from X'. Mostly we will be considering the case where X' is the food set F. To aid the mathematical analysis, we will use an equivalent but alternative definition of a RAF set (as defined and justified in [20]) that incorporates the molecular closure. Both definitions will be used in the proofs to follow.

Lemma 1. *Given a CRS $Q = (X, R, C, F)$, a subset of reactions $R' \subseteq R$ is a RAF for Q if and only if it is non-empty and both of the following conditions hold:*

(i) For every reaction $r \in R'$, there is at least one molecule type $x \in \mathrm{cl}_{R'}(F)$ with $(x, r) \in C$;

(ii) For every reaction $r = (A, B) \in R'$, $A \subseteq \mathrm{cl}_{R'}(F)$.

Condition (ii) is equivalent to the F-generated condition, but Condition (i) is stronger than the RA condition described earlier; however, the

combination of Conditions (i) and (ii) is equivalent to the combination of the earlier conditions of RA and F-generated.

Given a CRS $\mathcal{Q} = (X, R, C, F)$, the maxRAF algorithm computes a nested decreasing sequence of subsets of R, starting from R:

$$R = R_0 \supset R_1 \supset \cdots \supset R_k = R_{k+1}$$

where R_{j+1} is the subset of reactions in R_j that have all their reactants and at least one catalyst in $\mathrm{cl}_{R_j}(F)$. At the first value of k for which $R_k = R_{k+1}$, this set is either empty (in which case, $\mathcal{Q}$ has no RAF) or it is the unique maximal RAF set (maxRAF). For a proof of these assertions, see [7]. A rudimentary runtime analysis of the algorithm gives us $O(|R|^3|X|)$ time and $O(|X| + |R|)$ space [7], although under common circumstances (such as when simulating the Binary Polymer Model) it tends to run subquadratically in time with the number of reactions (see the simulations in [7]). Efforts have been made to improve this runtime and optimised implementations exist which perform significantly better over certain networks (see [21]).

3. Complexity results for closed RAFs

One important notion that ties the structural properties of RAFs to chemical realism is to require a RAF to be *closed*. Formally, given a CRS $\mathcal{Q} = (X, R, C, F)$, a RAF $R' \subseteq R$ is said to be *closed* if there is no reaction in $R - R'$ that has all its reactants and at least one catalyst in $\mathrm{cl}_{R'}(F)$. In other words, in a closed RAF, all reactions that *can* happen are included in the RAF. For example, in Fig. 1(a) there are four different RAFs, but only two of these four are closed; the closed RAFs in Fig. 1(a) are $\{\{f_4, f_5\} \to \{p_1\}, \{p_1, f_2\} \to \{p_2\}\}$ and the maxRAF (containing all reactions). Note that the maxRAF is always closed, if it exists.

Closed RAFs are of particular relevance to evolutionary theories of RAFs and early metabolic cycles: by finding the closed subRAFs of a particular RAF, it may be possible to trace back its 'ancestral' history; that is, a sequence of 'stable' states that could have initially lead to the production of the RAF (see [11,18]). In a recent paper [18], a direct, formal relationship between closed RAFs and a field known as Chemical Organisation Theory was established (adding to earlier links (see [20])). Using this connection, a new type of algorithm was developed to enumerate the set of all closed subRAFs existing within the maxRAF [18]. Although the new algorithm had reasonable performance for RAFs of size $\leq$ 200, it was not shown to run in polynomial-time [18]. It has remained an open question (posed in [18] and [11]) as to whether such a polynomial-time algorithm exists.

In this section, we present our complexity results surrounding closed RAFs. We solve the open problem posed in [11,18] by demonstrating the NP-completeness of finding closed subRAFs. We also we show that finding closed 'uninhibited' RAFs is not fixed-parameter tractable (FPT) in the number of inhibitions (unlike the case for non-closed RAFs [22]).

To motivate this setting, we consider first two questions without the 'closure' constraint. Given a CRS $\mathcal{Q} = (X, R, C, F)$ that has a RAF, let $\hat{x} \in X$ be any molecule type. There are then simple polynomial-time algorithms to determine answers to each of the following questions.

Does $\mathcal{Q}$ have:

(i) a RAF that does not produce $\hat{x}$?
(ii) a RAF that is a strict subset of maxRAF($\mathcal{Q}$)?

For Problem (i), let $R^*_{\hat{x}}$ be the set of reactions in maxRAF($\mathcal{Q}$) that do not produce $\hat{x}$. The answer to (i) is 'yes' if and only if $\hat{x} \notin F$ and the CRS $(X, R^*_{\hat{x}}, C_{|X\times R^*_{\hat{x}}}, F)$ has a RAF.

For Problem (ii), let $R^*_r = \mathrm{maxRAF}(\mathcal{Q}) - \{r\}$. The answer to (ii) is 'yes' if and only if the CRS $(X, R^*_r, C_{|X\times R^*_r}, F)$ has a RAF for some $r \in \mathrm{maxRAF}(\mathcal{Q})$.

However, if we modify these two questions so as to require the desired RAF to be closed, they become much more difficult, as we now explain.

We start with the following problem: Given a CRS $\mathcal{Q} = (X, R, C, F)$, does $\mathcal{Q}$ contain a closed RAF R' that does not produce a certain molecule type $\hat{x} \in X$? More formally stated:

PROBLEM: *Forbidden-molecule closed RAF*
INSTANCE: A CRS $\mathcal{Q} = (X, R, C, F)$ and a particular molecule type $\hat{x} \in X$.
QUESTION: Does $\mathcal{Q}$ contain a closed RAF $R' \subseteq R$ with $\hat{x} \notin \mathrm{cl}_{R'}(F)$?

Theorem 3.1. *The Forbidden-molecule closed RAF problem is NP-hard.*

Proof. Our proof will follow a reduction from 3SAT. Given a formula $\mathcal{H}$, we will construct in polynomial-time a CRS $\mathcal{Q}_\mathcal{H} = (X, R, C, F)$ with a distinguished molecule type $\hat{x} \in X$ that has a closed RAF R' with $\hat{x} \notin \mathrm{cl}_{R'}(F)$ if and only if $\mathcal{H}$ admits a satisfying assignment.

We start by describing the *food gadget*. Our CRS $\mathcal{Q}_\mathcal{H}$ will contain a single food molecule type $F = \{f\}$. We will include in $\mathcal{Q}_\mathcal{H}$ two distinguished molecule types x_{out} and x_{in} as well as two distinguished reactions r_{out} and r_{in}. The reaction r_{out} will be catalysed by the food molecule f (for simplicity, we will not include this catalyzation arrow in the diagrams), and the reaction r_{in} will be catalysed by the molecule type x_{out}. We present the food gadget in Fig. 2(a) (for now we also exclude the reactants of r_{out} and the reactions that x_{in} acts as a reactant for).

Since r_{in} is (and will be, in the overall construction) the only reaction with all food-molecule reactants, any F-generated set in $\mathcal{Q}_\mathcal{H}$ must contain r_{in}. As RAFs are by definition F-generated, the following condition can therefore be inferred:

$$R' \subseteq R \text{ is a RAF} \implies r_{in} \in R' \tag{1}$$

As x_{out} is the only molecule type catalysing r_{in} and r_{out} is the only reaction producing x_{out}, we must also have:

$$R' \subseteq R \text{ is a RAF} \implies r_{out} \in R'. \tag{2}$$

Next, we will begin to present the *clause gadget*, displayed in Fig. 2(b). From (2), we can infer that each reactant of r_{out} must also be in $\mathrm{cl}_{R'}(F)$. Let $A_{out} \subseteq X$ denote the set of reactants of r_{out}. It follows that if $R' \subseteq R$ is a RAF then $\bigwedge_{x\in A_{out}} x \in \mathrm{cl}_{R'}(F)$. Further, since none of $x \in A_{out}$ are (or will be) food molecule types, we must have at least one reaction producing x. Let $\rho : X \to 2^R$ denote the set of reactions that produce a molecule type $x \in X$. The following holds:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_{x\in A_{out}} \bigvee_{r\in\rho(x)} r \in R' \tag{3}$$

In Fig. 2(b), condition (3) forces at least one reaction from each 'block' (of three reactions) in the middle row to be in R' — for *any* RAF set $R' \subseteq R$.

We will now formally define our clause gadgets and illustrate how they integrate with the structure of the given formula $\mathcal{H}$. For added formality, we denote $\mathcal{H} = \bigwedge_i c_i = \bigwedge_i l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)}$. That is, let c_i represent the ith clause in $\mathcal{H}$ and let $l_{(i,j)}$ denote the jth literal in the ith clause of $\mathcal{H}$. Each $l_{(i,j)}$ literal can either be *positive* or *negative*.

For each clause $c_i = l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)} \in \mathcal{H}$, let c_i^X be a reactant molecule type to r_{out} (i.e. let $c_i^X \in A_{out}$) and let $l^R_{(i,1)}, l^R_{(i,2)}, l^R_{(i,3)}$ be reactions producing c_i^X (i.e. let $l^R_{(i,1)}, l^R_{(i,2)}, l^R_{(i,3)} \in \rho(c_i^X)$). For each reaction $l^R_{(i,j)}$, also let $l^X_{(i,j)}$ be its single reactant and let c_i^X be its single catalysing molecule type. Stated formally: for each clause $c_i \in \mathcal{H}$, we have $l_{(i,j)} \in c_i \implies (l^R_{(i,j)} = (\{l^X_{(i,j)}\}, \{c_i^X\}) \in R) \wedge ((c_i^X, l^R_{(i,j)}) \in C)$ for $1 \leq j \leq 3$. See Fig. 3 for an illustration. We can now rewrite condition (3) as:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_i (l^R_{(i,1)} \in R') \vee (l^R_{(i,2)} \in R') \vee (l^R_{(i,3)} \in R') \tag{4}$$

Since each $l^R_{(i,j)}$ reaction from (4) must contain its reactants we also gain the condition:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_i (l^X_{(i,1)} \in \mathrm{cl}_{R'}(F)) \vee (l^X_{(i,2)} \in \mathrm{cl}_{R'}(F)) \vee (l^X_{(i,3)} \in \mathrm{cl}_{R'}(F)) \tag{5}$$

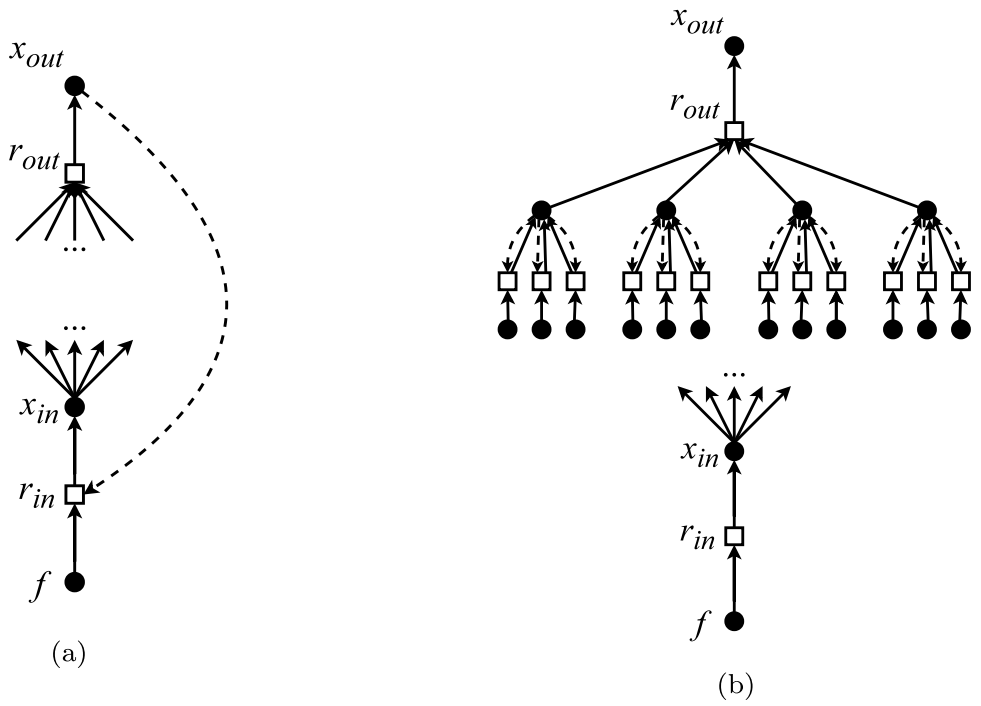


Fig. 2. (a) The food gadget: both r_{out} and r_{in} must necessarily be in any RAF set of $\mathcal{Q}_{\mathcal{H}}$. (b) The clause gadget: at least one reaction from each 'block' (of three) in the middle row must necessarily be in any RAF set $R' \subseteq R$. (In (b) we are hiding the catalyzation arrow from x_{out} to r_{in} for brevity).

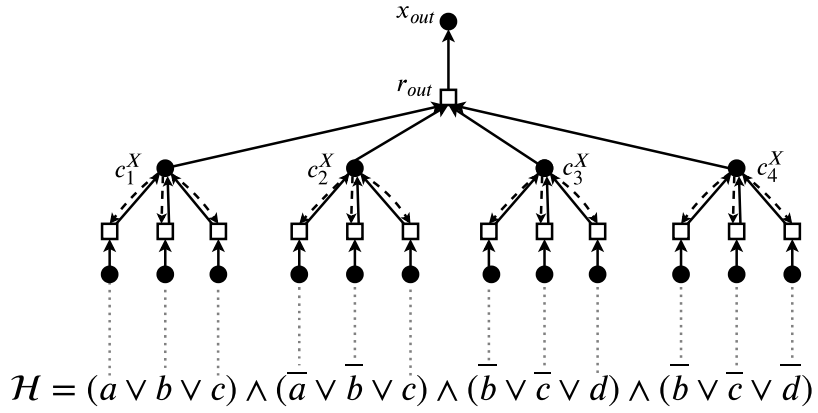


Fig. 3. Given a SAT formula e.g. $\mathcal{H} = (a \vee b \vee c) \wedge (\overline{a} \vee \overline{b} \vee c) \wedge (\overline{b} \vee \overline{c} \vee d) \wedge (\overline{b} \vee \overline{c} \vee \overline{d})$, we construct the clause gadget as above. For each literal occurrence, we have a single 'literal' reaction with a single 'literal' molecule reactant. (In the figure, this relationship is referenced by the faint dotted lines above each literal occurrence in $\mathcal{H}$.) Each literal reaction produces, and is catalysed by, a corresponding 'clause' molecule type. All clause molecules types are reactants to r_{out}.

As we proceed, we will refer to $l^R_{(i,j)} \in R$ as the 'literal' reaction and $l^X_{(i,j)}$ as the 'literal' molecule type for a literal occurrence $l_{(i,j)} \in \mathcal{H}$.

Next, we describe the *variable gadget*, displayed in Fig. 4. For each variable $v \in \mathcal{H}$, introduce a variable gadget to $\mathcal{Q}_{\mathcal{H}}$. For a variable v, the variable gadget consists of three reactions named v^R (highlighted in green), $\overline{v}^R$ (highlighted in pink) and v^R_{mid}. The central red molecule type is $\hat{x} \in X$, which is the forbidden molecule type. The sets of product molecule types L_V and $L_{\overline{V}}$ will be defined shortly. Formally, we have $v^R = (\{x_{in}\}, \{v^X_{cat}, v^X_{top}, v^X_{mid}\} \cup L_V)$, $\overline{v}^R = (\{x_{in}\}, \{\overline{v}^X_{cat}, v^X_{top}, \overline{v}^X_{mid}\} \cup L_{\overline{V}})$, $v^R_{mid} = (\{v^X_{mid}, \overline{v}^X_{mid}\}, \{\hat{x}\})$ for molecule types $v^X_{top}, v^X_{mid}, \overline{v}^X_{mid}, v^X_{cat}, \overline{v}^X_{cat}$ with $(v^X_{cat}, v^R), (\overline{v}^X_{cat}, \overline{v}^R), (x_{in}, v^R_{mid}) \in C$, and v^X_{top} a reactant to r_{out} (i.e. $v^X_{top} \in A_{out}$).

Since v^X_{top} is a reactant to r_{out} and as $v^R, \overline{v}^R$ are the only reactions producing r_{out}, by condition (3) it follows:

$$R' \subseteq R \text{ is a RAF } \implies (v^R \in R') \vee (\overline{v}^R \in R') \tag{6}$$

Further, for any closed RAF $R' \subseteq R$, if both $v^R, \overline{v}^R \in R'$, it follows $v^X_{mid}, \overline{v}^X_{mid}, x_{in} \in \mathrm{cl}_{R'}(F)$ and therefore $v^R_{mid} \in R'$ (by closure of R'). Since $v^R_{mid} \in R'$ produces our forbidden molecule type, we can infer (contrapositive) the following condition:

$$R' \subseteq R \text{ is a closed RAF } \wedge \hat{x} \notin \mathrm{cl}_{R'}(F) \implies (v^R \notin R') \vee (\overline{v}^R \notin R') \tag{7}$$

Combining (7) with (6) gives:

$$R' \subseteq R \text{ is a closed RAF } \wedge \hat{x} \notin \mathrm{cl}_{R'}(F) \implies (v^R \in R') \Leftrightarrow (\overline{v}^R \notin R') \tag{8}$$

Condition (8) means that any closed RAF in R' which does not produce the forbidden molecule type (i.e. $\hat{x} \notin cl_{R'}(F)$) must contain *exactly one* of the reactions v^R or $\overline{v}^R$ for every variable $v \in \mathcal{H}$. (Note the analogy between the inclusions of reactions $v^R, \overline{v}^R$ and truth assignments of a variable v in a SAT assignment.)

We will now consolidate our variable gadgets with our clause gadgets, and define the sets L_V and $L_{\overline{V}}$. For a variable gadget in $\mathcal{Q}_{\mathcal{H}}$ corresponding to a variable $v \in \mathcal{H}$, define L_V to be the set of all literal molecule types $l^X_{(i,j)}$ for $l_{(i,j)} \in \mathcal{H}$ being a *positive* occurrence of the variable v. Likewise, define $L_{\overline{V}}$ to be the set of all literal molecule types

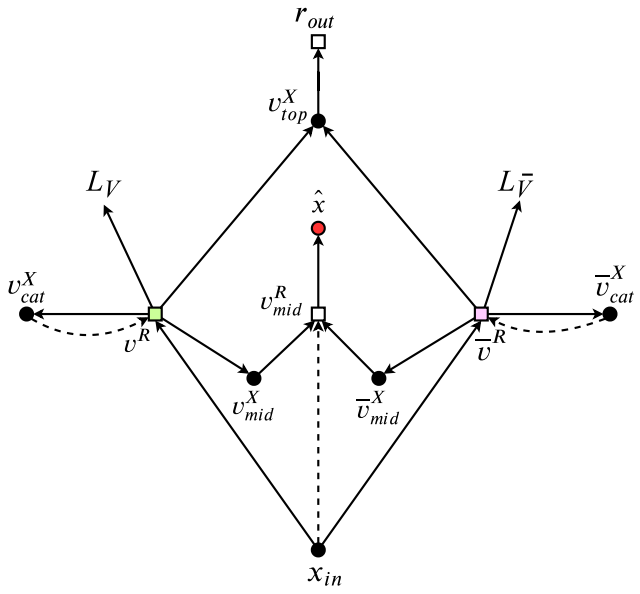


Fig. 4. The variable gadget for a variable $v \in \mathcal{H}$. In any closed RAF set $R' \subseteq R$ that does not produce $\hat{x}$, *exactly one* of $v^R, \overline{v}^R$ is included R'.

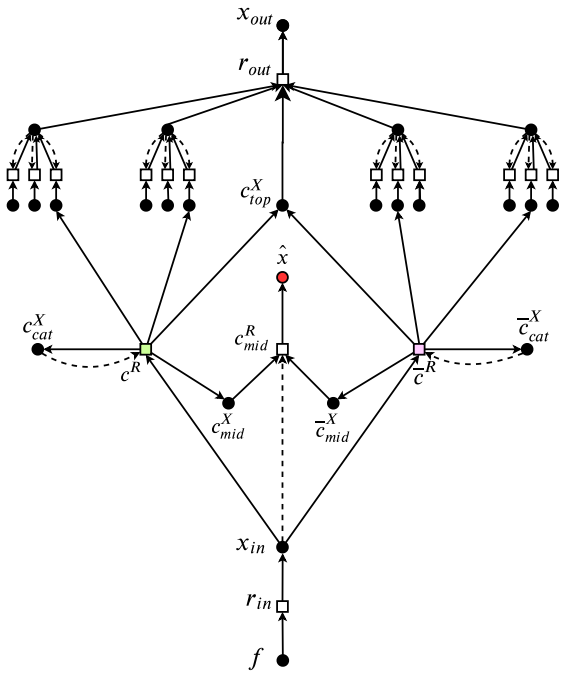


Fig. 5. Part of the construction of $\mathcal{Q}_{\mathcal{H}}$ for the formula $\mathcal{H} = (a \vee b \vee c) \wedge (\overline{a} \vee \overline{b} \vee c) \wedge (\overline{b} \vee \overline{c} \vee d) \wedge (\overline{b} \vee \overline{c} \vee \overline{d})$, only including the variable gadget for variable c and hiding the gadgets for variables a, b, d. We show how a single variable gadget fits in amidst the overall construction. The literal molecules and reactions are unlabelled for brevity. The catalyzation edges from f to r_{out} and from x_{out} to r_{in} are also not shown. As can be seen, the c^R and $\overline{c}^R$ reactions produce the literal molecule types corresponding to positive and negative occurrences of the variable c (i.e. c and $\overline{c}$) in $\mathcal{H}$.

$l^X_{(i,j)}$ for $l_{(i,j)} \in \mathcal{H}$ being a *negative* occurrence of the variable v. See Fig. 5 for an illustration.

This completes the construction definition. See Fig. 6 for an illustration of the overall construction. Since the size of each gadget is constant, our construction is linear in the size of the Formula $\mathcal{H}$ and therefore polynomial-time computable.

To complete the proof, we now establish the following claim: $\mathcal{H}$ admits a satisfying assignment if and only if $\mathcal{Q}_{\mathcal{H}}$ has a closed RAF R' with $\hat{x} \notin \mathrm{cl}_{R'}(F)$.

Proof. ($\Longrightarrow$) Suppose $\mathcal{Q}_{\mathcal{H}}$ contains a closed RAF R' with $\hat{x} \notin \mathrm{cl}_{R'}(F)$. We construct an assignment $\mathcal{A} \models \mathcal{H}$ as follows: for each variable $v \in \mathcal{H}$, let $\mathcal{A}(v) = 1$ if and only if $v^R \in R'$ (set $\mathcal{A}(v) = 0$ otherwise). By (8), $\mathcal{A}$ is well-defined. By (5), we must have $\bigwedge_i (l^X_{(i,1)} \in \mathrm{cl}_{R'}(F)) \vee (l^X_{(i,2)} \in \mathrm{cl}_{R'}(F)) \vee (l^X_{(i,3)} \in \mathrm{cl}_{R'}(F))$.

Now suppose WLOG that $l^X_{(i,1)} \in cl_{R'}(F)$ and that $l_{(i,1)}$ is a positive literal occurrence of the variable v in $\mathcal{H}$ (for some arbitrarily chosen clause at index i). By construction v^R will be the only reaction producing $l^X_{(i,1)}$, so $v^R \in R'$ and $\mathcal{A}(v) = 1$. Furthermore, as $l_{(i,1)}$ is a positive literal occurrence of the variable v, it follows $\mathcal{A}(l_{(i,1)}) = 1$ and therefore $\mathcal{A}(l_{(i,1)}) \vee \mathcal{A}(l_{(i,2)}) \vee \mathcal{A}(l_{(i,3)}) = 1$. The argument is symmetric for negative literal occurrences. Since the clause was chosen arbitrarily, all clauses must evaluate to true. We conclude that $\mathcal{A}$ must be a satisfying assignment.

($\Longleftarrow$) Suppose $\mathcal{H}$ admits a satisfying assignment $\mathcal{A}$. We construct a closed RAF R' with $\hat{x} \notin \mathrm{cl}_{R'}(F)$ in four simple steps: (i) Let $r_{in}, r_{out} \in R'$. (ii) For each variable $v \in \mathcal{H}$ with $\mathcal{A}(v) = 1$, let $v^R \in R'$. (iii) For each variable $v \in \mathcal{H}$ with $\mathcal{A}(v) = 0$, let $\overline{v}^R \in R'$. (iv) For each literal $l_{(i,j)} \in \mathcal{H}$ with $\mathcal{A}(l_{(i,j)}) = 1$, let $l^R_{(i,j)} \in R'$. We complete the proof by showing R' is a closed RAF that does not produce $\hat{x}$. We do this through four claims:

Claim 1: R' is F-generated.

Proof. We simply provide the linear ordering of reactions r_{in}, V, L, r_{out} for V and L the variable and literal reactions included in R', respectively. (The linear ordering of reactions within V or L does not matter.)

Claim 2: R' is autocatalytic.

Proof. All literal reactions $l^R_{(i,j)}$ produce the molecule types c^X_i which catalyse every $l^R_{(i,j)}$. The r_{out} reaction is catalysed by f. The r_{in} reaction is catalysed by x_{out} which is produced by r_{out}. Lastly, $v^R, \overline{v}^R$ reactions produce molecule types $v^X_{cat}, \overline{v}^X_{cat}$ which catalyse $v^R, \overline{v}^R$, respectively. This covers all cases, so R' is autocatalytic.

Claim 3: R' is closed.

Proof. We prove R' is closed by case analysis on the reactions outside of R'. Let $r \notin R'$, then we have three cases:

Case (i): $r = l^R_{(i,j)} \notin R'$ is a literal reaction. Suppose WLOG that $l_{(i,j)} \in \mathcal{H}$ is a positive occurrence of the variable $v \in \mathcal{H}$. Since $l^R_{(i,j)} \notin R'$ then $\mathcal{A}(l_{(i,j)}) = 0$, and as $l_{(i,j)}$ is a positive literal occurrence of v it follows $\mathcal{A}(v) = 0$ and so $v^R \notin R'$. As v^R is the only reaction producing $l^X_{(i,j)}$, we have $l^X_{(i,j)} \notin \mathrm{cl}_{R'}(F)$, so $l^R_{(i,j)}$ does not follow by closure. The argument is symmetric for negative literal occurrences in $\mathcal{H}$.

Case (ii): $r = v^R \notin R'$ or $r = \overline{v}^R \notin R'$ is a variable gadget reaction. The variable gadget reactions $v^R, \overline{v}^R$ are only catalysed by the molecule types they produce (which are v^X_{cat} and $\overline{v}^X_{cat}$, respectively). So they do not follow by closure.

Case (iii): $r = v^R_{mid}$. Since $\mathcal{A}$ is well defined, by construction we cannot have both $v^R, \overline{v}^R \in R'$, so the reactants $v^X_{mid}, \overline{v}^X_{mid}$ of v^R_{mid} cannot both be in $\mathrm{cl}_{R'}(F)$. It follows v^R_{mid} does not follow by closure.

Claim 4: R' does not produce $\hat{x}$ (i.e. $\hat{x} \notin cl_{R'}(F)$).

Proof. By construction of R', we did not include any v^R_{mid} reactions. Since they are the only reactions producing $\hat{x}$, it immediately follows that $\hat{x} \notin \mathrm{cl}_{R'}(F)$.

This completes the proof □

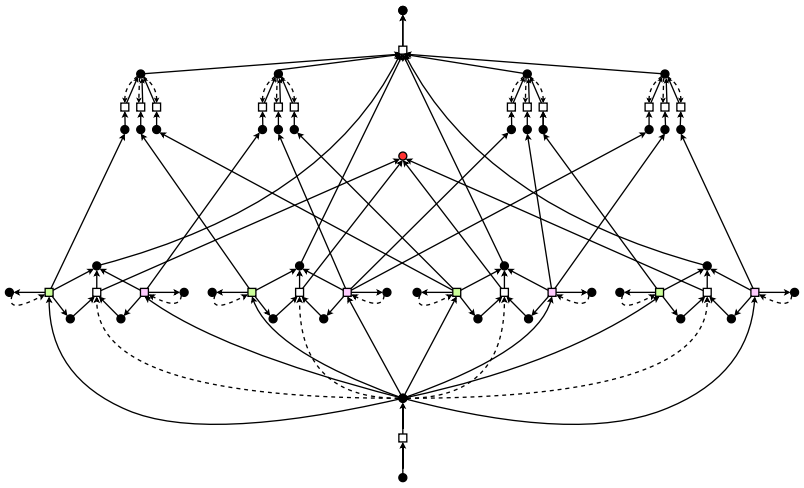


Fig. 6. The entire construction $\mathcal{Q}_{\mathcal{H}}$ for the formula $\mathcal{H} = (a \vee b \vee c) \wedge (\bar{a} \vee \bar{b} \vee c) \wedge (\bar{b} \vee \bar{c} \vee d) \wedge (\bar{b} \vee \bar{c} \vee \bar{d})$ is shown. All labels are hidden for brevity; again, the catalyzation edges from f to r_{out} and from x_{out} to r_{in} are also not shown.

3.1. The closed strict subRAF problem

Theorem 3.1 provides the tool for establishing the hardness of some questions posed in earlier papers. In particular, we can use it to solve the closed strict subRAF problem mentioned in [11,18].

PROBLEM: *Closed strict subRAF*
INSTANCE: A CRS $\mathcal{Q} = (X, R, C, F)$
QUESTION: Does $\mathcal{Q}$ contain a closed RAF $R' \subset R$ with $R' \subset \text{maxRAF}(\mathcal{Q})$?

Theorem 3.2. *The closed strict subRAF problem is NP-complete.*

Before we begin with the proof of Theorem 3.2, we will first introduce some new notation and establish a series of lemmas. Given two CRS $\mathcal{Q}_1 = (X_1, R_1, C_1, F_1)$, $\mathcal{Q}_2 = (X_2, R_2, C_2, F_2)$ and a relation $\sim$, we say $\mathcal{Q}_1 \sim \mathcal{Q}_2$ if and only if $X_1 \sim X_2$, $R_1 \sim R_2$, $C_1 \sim C_2$ and $F_1 \sim F_2$. In the proofs below, we often relate CRS by the relations $\subseteq$ and $\subset$ (e.g. $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ and $\mathcal{Q}_1 \subset \mathcal{Q}_2$).

Lemma 2. *Let $\mathcal{Q}_1, \mathcal{Q}_2$ be CRS with $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$. If R' is a RAF in $\mathcal{Q}_1$, then R' is a RAF in $\mathcal{Q}_2$.*

Proof. Let $\mathcal{Q}_1 = (X_1, R_1, C_1, F_1), \mathcal{Q}_2 = (X_2, R_2, C_2, F_2)$ be CRS with $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ and suppose that $R' \subseteq R_1$ is a RAF in $\mathcal{Q}_1$. Since $F_1 \subseteq F_2$ and $R' \subseteq R_1 \subseteq R_2$, it follows both F_1 and R' are well-defined in $\mathcal{Q}_2$, and so the closure $cl_{R'}(F_1)$ must also be well-defined in $\mathcal{Q}_2$. Furthermore, $cl_{R'}(F_1)$ in $\mathcal{Q}_2$ must equal $cl_{R'}(F_1)$ in $\mathcal{Q}_1$ as the reaction and food sets are identical. By the definition of a RAF (from Lemma 1), and since $C_1 \subseteq C_2$, it follows R' must be a RAF in $\mathcal{Q}_2$. □

Corollary 1 (*maxRAF Monotonicity*). *Let $\mathcal{Q}_1$, $\mathcal{Q}_2$ be CRS with $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$, then maxRAF$(\mathcal{Q}_1) \subseteq$ maxRAF$(\mathcal{Q}_2)$.*

Proof. The maxRAF of a CRS is the union of all RAFs in the CRS. The result therefore follows immediately from Lemma 2. □

Lemma 3. *Let $\mathcal{Q}_1$, $\mathcal{Q}_2$ be CRS with $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ and suppose there exists a RAF in $\mathcal{Q}_2$ that is not a subset of the maxRAF of $\mathcal{Q}_1$, then maxRAF$(\mathcal{Q}_1) \subset$ maxRAF$(\mathcal{Q}_2)$.*

Proof. Let R' denote a RAF in $\mathcal{Q}_2$ that is not a subset of maxRAF$(\mathcal{Q}_1)$. By Lemma 2, maxRAF$(\mathcal{Q}_1)$ is a RAF in $\mathcal{Q}_2$. As RAFs are closed under union, it follows that $R' \cup \text{maxRAF}(\mathcal{Q}_1)$ is a RAF in $\mathcal{Q}_2$, and as $R' \nsubseteq \text{maxRAF}(\mathcal{Q}_1)$, we have $R' \cup \text{maxRAF}(\mathcal{Q}_1) \supset \text{maxRAF}(\mathcal{Q}_1)$. Since all RAFs in a CRS are subsets of the maxRAF in that CRS, we have $\text{maxRAF}(\mathcal{Q}_1) \subset R' \cup \text{maxRAF}(\mathcal{Q}_1) \subseteq \text{maxRAF}(\mathcal{Q}_2)$. □

We now apply these results, as well as Theorem 3.1, to prove Theorem 3.2 (the closed strict subRAF problem is NP-complete).

Proof of Theorem 3.2. Given a set of reactions $R' \subseteq R$, one can check whether or not it is a closed RAF that is a strict subset of maxRAF$(\mathcal{Q})$ in polynomial-time, so the problem is in the class NP. To establish NP-completeness, we perform a polynomial-time reduction from the Forbidden-molecule closed RAF problem (which is NP-hard by Theorem 3.1). Given a CRS $\mathcal{Q} = (X, R, C, F)$ and a particular molecule type $\hat{x} \in X$, we construct a CRS $\hat{\mathcal{Q}} = (\hat{X}, \hat{R}, \hat{C}, \hat{F})$ such that $\mathcal{Q}$ has a closed RAF $R' \subseteq R$ with $\hat{x} \notin \text{cl}_{R'}(F)$ if and only if $\hat{\mathcal{Q}}$ has a closed subRAF $R'' \subset \hat{R}$ that is a strict subset of the maxRAF in $\hat{\mathcal{Q}}$.

The construction of $\hat{\mathcal{Q}}$ from $\mathcal{Q}$ is as follows (see Fig. 7 for a visualisation). For some newly introduced food molecule type $\hat{f}$, let $\hat{\mathcal{Q}} = (\hat{X}, \hat{R}, \hat{C}, \hat{F})$ where $\hat{X} = X \cup \{\hat{f}\}$, $\hat{R} = R \cup \{\{\hat{f}\} \to \{\hat{x}\}\}$, $\hat{C} = C \cup \{(\hat{x}, r) : r \in \hat{R}\}$ and $F = F \cup \{\hat{f}\}$. Essentially, we add a new food molecule type $\hat{f}$ to $\hat{\mathcal{Q}}$ and a reaction $\{\hat{f}\} \to \{\hat{x}\}$ that produces the forbidden molecule type $\hat{x}$; we then let $\hat{x}$ catalyse every reaction in $\hat{\mathcal{Q}}$, including the reaction $\{\hat{f}\} \to \{\hat{x}\}$. This construction is clearly polynomial-time computable in the size of $\mathcal{Q}$. We first note that as $\{\{\hat{f}\} \to \{\hat{x}\}\}$ is a RAF in $\hat{\mathcal{Q}}$ which is not a strict subset of the maxRAF in $\mathcal{Q}$ (it is nonexistent in $\mathcal{Q}$), by Lemma 3 we have maxRAF$(\mathcal{Q}) \subset$ maxRAF$(\hat{\mathcal{Q}})$.

We now establish the correctness of the construction (i.e. that $\mathcal{Q}$ contains a closed RAF $R' \subseteq R$ with $\hat{x} \notin \text{cl}_{R'}(F)$ if and only if $\hat{\mathcal{Q}}$ contains a closed RAF which is a strict subset of the maxRAF of $\hat{\mathcal{Q}}$). ($\Longrightarrow$) Suppose that $\mathcal{Q}$ contains a closed RAF $R' \subseteq R$ with $\hat{x} \notin \text{cl}_{R'}(F)$. We show that $\hat{\mathcal{Q}}$ contains a closed RAF which is a strict subset of the maxRAF of $\hat{\mathcal{Q}}$. By Lemma 2, R' must be a RAF in $\hat{\mathcal{Q}}$, and as $R' \subseteq \text{maxRAF}(\mathcal{Q}) \subset \text{maxRAF}(\hat{\mathcal{Q}})$ it follows R' is a RAF in $\hat{\mathcal{Q}}$ which is a strict subset of maxRAF$(\hat{\mathcal{Q}})$. We now aim to show that R' is closed in $\hat{\mathcal{Q}}$. We will do this by case analysis on the reactions $r \in \hat{R} - R'$ (i.e. reactions outside of R' in $\hat{\mathcal{Q}}$). There are two cases to consider: (i) $r = \{\hat{f}\} \to \{\hat{x}\}$ (i.e. r is our newly added reaction in $\hat{\mathcal{Q}}$) and (ii) $r \in R - R'$ (i.e. r is a reaction in $\mathcal{Q}$). In Case (i), since $\hat{x} \notin \text{cl}_{R'}(F)$ (in $\mathcal{Q}$) and as $\hat{f}$ does not catalyse anything, $\{\hat{f}\} \to \{\hat{x}\}$ will be uncatalyzed across $\text{cl}_{R'}(\hat{F})$ (in $\hat{\mathcal{Q}}$). In Case (ii), since $\{\hat{f}\} \to \{\hat{x}\}$ is uncatalyzed across $\text{cl}_{R'}(\hat{F})$ (from Case (i)), the rest of the construction $\hat{\mathcal{Q}}$ is identical to $\mathcal{Q}$. Therefore, as R' is closed in $\mathcal{Q}$ and so $r \in R - R'$ cannot proceed in $\mathcal{Q}$, it follows r cannot proceed in $\hat{\mathcal{Q}}$ either. We conclude that R' is closed in $\hat{\mathcal{Q}}$.

($\Longleftarrow$) Suppose that $R'' \subset \hat{R}$ is a closed RAF in $\hat{\mathcal{Q}}$ which is a strict subset of maxRAF$(\hat{\mathcal{Q}})$. We show that there exists a closed RAF in $\mathcal{Q}$ which does not produce the molecule type $\hat{x}$. Since $R'' \subset \text{maxRAF}(\hat{\mathcal{Q}})$, we must have $\hat{x} \notin \text{cl}_{R''}(\hat{F})$ (otherwise $\hat{x}$ would catalyse every reaction in $\hat{\mathcal{Q}}$, and so, by the closure of R'', it could no longer be a strict subset of

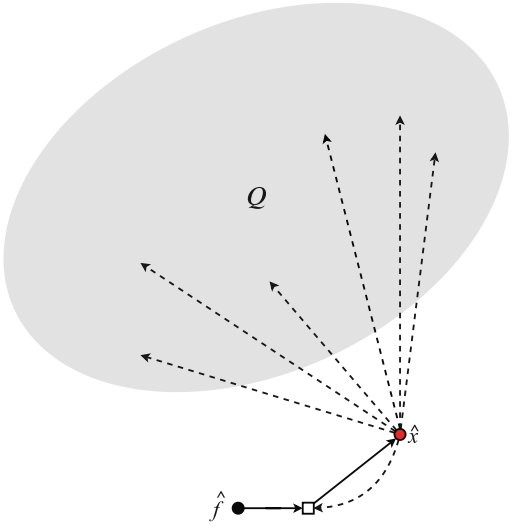


Fig. 7. The construction of the CRS $\hat{Q}$ (relative to the Q) is shown above. We define $\hat{Q}$ by adding to Q a new food molecule type $\hat{f}$ and a reaction $\{\hat{f}\} \to \{\hat{x}\}$ which produces the forbidden molecule type $\hat{x}$. We then let $\hat{x}$ catalyse $\{\hat{f}\} \to \{\hat{x}\}$ as well as every reaction in Q. The construction is such that maxRAF($\hat{Q}$) $\supset$ maxRAF(Q).

maxRAF($\hat{Q}$)). With $\hat{x} \notin \mathrm{cl}_{R''}(\hat{F})$, we must have $\{\hat{f}\} \to \{\hat{x}\} \notin R''$ (it is not catalysed), and therefore $R'' \subseteq R$ (i.e. R'' must be contained within the CRS Q). Since the food molecule $\hat{f}$ does not catalyse anything, without $\hat{x}$ or $\{\hat{f}\} \to \{\hat{x}\}$ the rest of the construction is identical to Q. It follows R'' must be a closed RAF in Q with $\hat{x} \notin cl_{R''}(F)$. □

3.2. Application to 'uninhibited' RAFs

So far, the CRS model only includes catalytic interactions between molecule types and reactions. This may be plausible in some circumstances, although in real biochemical systems it is very often the case that certain molecule types *inhibit* the presence of certain reactions. To describe this more formally, consider a CRS $Q = (X, R, C, F)$ together with an additional set $I \subseteq X \times R$. An *uninhibited RAF* (uRAF) is a RAF set R' for Q that satisfies the following additional property: for every reaction $r \in R'$, no molecule type in the food set or produced by another reaction in R' inhibits r. Formally: For each $r \in R'$, there is no molecule type $x \in \mathrm{cl}_{R'}(F)$ for which $(x, r) \in I$ [21]. Essentially, inhibition can be viewed as the exact opposite of catalyzation.

Following [23], it seems sensible to further restrict interest to uRAFs that are also closed, since an uninhibited *non-closed* RAF could provide the reactants and catalyst for one or more additional reactions (outside the RAF) to occur, which, in turn, generate a molecule which inhibits a reaction within the original RAF. Closed uRAFs, on the other hand, are truly free of inhibition from the products of such reactions. It was shown in [24] that determining the existence of a (non-closed) uRAF within a CRS Q is NP-hard, though it was shown in [22] to be fixed-parameter-tractable in the number of inhibiting molecule types. We prove the same result cannot be found for closed uRAFs (subject to P ≠ NP) by showing that the following question is NP-hard.

Theorem 3.3. *The following problem is NP-hard: Given a CRS $Q = (X, R, C, F)$ and a single inhibiting pair $(x, r) \in X \times R$, determine whether or not there exists a closed uRAF for Q with respect to the inhibiting set $I = \{(x, r)\}$.*

Proof. To establish NP-hardness, we again perform a polynomial-time reduction from the Forbidden-molecule closed RAF problem (*c.f.* Theorem 3.1). Given a CRS $Q = (X, R, C, F)$ and a particular molecule type $\hat{x} \in X$, we construct in polynomial-time a CRS Q' and an inhibiting singleton set I such that Q has a closed RAF $R' \subseteq R$ with $\hat{x} \notin \mathrm{cl}_{R'}(F)$ if and only if Q' has a closed, uninhibited RAF w.r.t I. The construction is simple: let $Q' = Q$ and define $I = \{(\hat{x}, r) : r \in R\}$ (i.e. make the forbidden molecule type $\hat{x}$ inhibit all reactions in the CRS Q'). It immediately follows that a closed RAF R' in Q with $\hat{x} \notin \mathrm{cl}_{R'}(F)$ will be a closed, uninhibited RAF in Q' w.r.t I as $\hat{x}$ is the only molecule type which is inhibiting and the rest of the construction is identical. Conversely, a closed uRAF R' for Q' w.r.t I must have $\hat{x} \notin \mathrm{cl}_{R'}(F)$ as otherwise $\hat{x}$ would inhibit all reactions and RAFs are *non-empty*. Again, since the rest of the construction is identical, we have R' a closed RAF in Q with $\hat{x} \notin \mathrm{cl}_{R'}(F)$. The construction is clearly polynomial-time computable, as required. □

4. Complexity results for elementary RAFs

To help avoid NP-complete problems arising in RAF theory, a simpler setting has recently been studied (see [11]). An *elementary* CRS is a special type of CRS where each reaction has all its reactants in the food set. Every reaction is therefore trivially F-generated and so the RAF condition reduces to just the RA catalysis condition (i.e. each reaction is catalysed by the product of some other reaction or by an element of the food set).

Although this setting seems quite restrictive, it is nevertheless pertinent in experimental systems [9,10] as well as in theoretical models [25]. Many problems that are known to NP-hard in the general setting are tractable in the elementary setting: for example, finding the minimum-size RAF and finding closed subRAFs of the maxRAF (proven to be NP-hard in Section 3) are both computable in polynomial time in the elementary setting [11]. We now present two questions concerning elementary CRSs that were posed recently in [11]:

Q1 Is there a polynomial-time algorithm to find a uRAF for an elementary CRS?

Q2 Is there a polynomial-time algorithm to find a *maximum*-sized irreducible RAF for an elementary CRS?

Here, we resolve both questions Q1 and Q2 by providing proofs of NP-hardness.

First recall that a uRAF is a RAF set $R' \subseteq R$ that satisfies the additional property: for each $r \in R'$, there is no molecule type $x \in \mathrm{cl}_{R'}(F)$ for which $(x, r) \in I$ [21].

Theorem 4.1. *Given an elementary CRS $Q = (X, R, C, F)$ and an inhibition assignment $I \subseteq X \times R$, determining whether or not a uRAF for Q exists is NP-complete.*

Proof. Checking whether a set of reactions $R' \subseteq R$ is both a RAF and uninhibited is polynomial-time computable, so the problem is in NP. We will establish NP-hardness via a reduction from 3SAT. Given a 3SAT formula $\mathcal{H}$ with m clauses, we construct in polynomial-time an elementary CRS $Q_{\mathcal{H}} = (X, R, F, C)$ and an inhibiting set $I \subseteq X \times R$ such that $Q_{\mathcal{H}}$ contains a uRAF w.r.t I if and only if $\mathcal{H}$ admits a satisfying assignment. For added formality, we denote $\mathcal{H} = \bigwedge_{i=1}^{m} c_i = \bigwedge_{i=1}^{m} l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)}$. That is, let c_i represent the ith clause in $\mathcal{H}$, and let $l_{(i,j)}$ denote the jth literal in the ith clause of $\mathcal{H}$. The construction of $Q_{\mathcal{H}} = (X, R, F, C)$ from $\mathcal{H}$ is as follows. First, we will have a single food molecule $f \in F$. Next, for each clause $c_i = l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)} \in \mathcal{H}$, let $l^R_{(i,1)}, l^R_{(i,2)}, l^R_{(i,3)} \in R$ be a set of 'literal' reactions in R and let $l^X_{(i,1)}, l^X_{(i,2)}, l^X_{(i,3)} \in X$ be a set of 'literal' molecule types in X. Each literal reaction $l^R_{(i,j)}$ will produce every literal molecule type in the subsequent clause (i.e. it will produce all the molecules types $l^X_{((i+1) mod m,1)}, l^X_{((i+1) mod m,2)}, l^X_{((i+1) mod m,3)}$), as well as an additional molecule type $I^X_{(i,j)}$ which we will refer to as the *inhibitor* for $l_{(i,j)}$. Each literal molecule $l^X_{(i,j)}$ will catalyse each literal molecule $l^R_{(i,j)}$. Formally, we have $(l^X_{(i,j)}, l^R_{(i,j)}) \in C_{\mathcal{H}}$ for each literal $l_{(i,j)} \in \mathcal{H}$, with $l^R_{(i,j)}$ defined below (see Fig. 8 for illustration).

$$l^R_{(i,j)} = (\{f\}, \{l^X_{((i+1) mod m,1)}, l^X_{((i+1) mod m,2)}, l^X_{((i+1) mod m,3)}, I^X_{(i,j)}\})$$

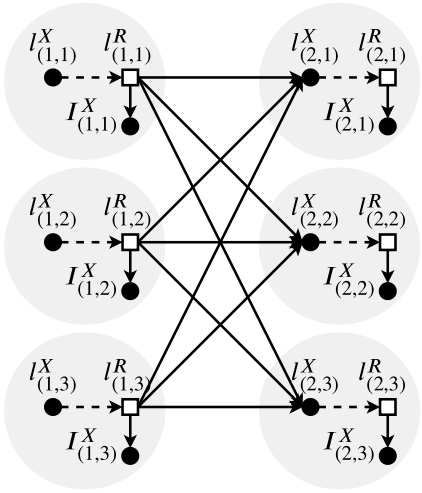


Fig. 8. Given a 3SAT formula with containing two clauses $c_1 = l_{(1,1)} \vee l_{(1,2)} \vee l_{(1,3)}$ and $c_2 = l_{(2,1)} \vee l_{(2,2)} \vee l_{(2,3)}$, we construct literal reactions and molecule types as above. (We hide the food molecule type for brevity; it is the sole reactant to all reactions.) Each reaction produces every literal molecule type for every literal in the succeeding clause; it is catalysed only by its corresponding literal molecule type. Each reaction also produces a corresponding 'literal inhibitor' molecule.

Finally, for each pair of literals $l_{(s,t)}, l_{(u,v)} \in \mathcal{H}$ which are opposite-parity occurrences of the same underlying variable (e.g. $l_{(s,t)} = v$ and $l_{(u,v)} = \overline{v}$), let $(I^X_{(s,t)}, l^R_{(u,v)}) \in I$. To better illustrate the proof we contrive a simple diagrammatic language, defined across Figs. 9(a) and 9(b). For an illustration of the overall construction expressed through this diagrammatic notation, see Figs. 10(a) and 10(b).

Lemma 4. *Let $\mathcal{H} = \bigwedge_{i=1}^m c_i = \bigwedge_{i=1}^m l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)}$ be a 3SAT formula. Any uRAF in the constructed CRS $Q_{\mathcal{H}}$ w.r.t the inhibiting set I cannot contain two literal reactions $l^R_{(s,t)}, l^R_{(u,v)}$ where $l_{(s,t)}, l_{(u,v)} \in \mathcal{H}$ are opposite-parity occurrences of the same underlying variable in $\mathcal{H}$ (e.g. $l_{(s,t)} = v$ and $l_{(u,v)} = \overline{v}$ in $\mathcal{H}$).*

Proof. By the construction of $Q_{\mathcal{H}} = (X, R, C, F)$ we have $(I^X_{(s,t)}, l^R_{(u,v)})$, $(I^X_{(u,v)}, l^R_{(s,t)}) \in I$ as $l_{(s,t)}, l_{(u,v)}$ are opposite-parity occurrences of the same underlying variable in $\mathcal{H}$. Let $R' \subseteq R$ be a uRAF in $Q_{\mathcal{H}}$ and suppose $l^R_{(s,t)} \in R'$. As $l^R_{(s,t)}$ produces $I^X_{(s,t)}$ we have $l^R_{(u,v)}$ inhibited, so if $l^R_{(s,t)} \in R'$ then $l^R_{(u,v)} \notin R'$. The argument is symmetric when supposing $l^R_{(u,v)} \in R'$. □

Lemma 5. *Let $\mathcal{H} = \bigwedge_{i=1}^m c_i = \bigwedge_{i=1}^m l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)}$ be a 3SAT formula. Any RAF in the constructed CRS $Q_{\mathcal{H}}$ must have at least one literal reaction $l^R_{(i,1)}, l^R_{(i,2)}, l^R_{(i,3)}$ included for every clause $c_i = l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)} \in \mathcal{H}$.*

Proof. Let $Q_{\mathcal{H}} = (X, R, C, F)$ denote the constructed CRS for the given 3SAT formula $\mathcal{H} = \bigwedge_{i=1}^m c_i = \bigwedge_{i=1}^m l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)}$ and let $R' \subseteq R$ be a RAF in $Q_{\mathcal{H}}$. Suppose WLOG that $l^R_{(i,j)} \in R'$. As R' is a RAF, $l^R_{(i,j)}$ must be catalysed. By construction, $l^R_{(i,j)}$ is only catalysed by the corresponding literal molecule type $l^X_{(i,j)}$, which is only produced by reactions in the preceding clause — i.e. $\rho(l^X_{(i,j)}) = \{l^R_{((i-1) \bmod m,1)}, l^R_{((i-1) \bmod m,2)}, l^R_{((i-1) \bmod m,3)}\}$. It follows that R' must also contain at least one of the literal reactions from the preceding clause. Inductively, we conclude R' must contain at least one literal reaction per clause in $\mathcal{H}$. □

We now prove that $Q_{\mathcal{H}}$ has a uRAF with respect to the inhibiting set I if and only if $\mathcal{H}$ admits a satisfying assignment.

Proof.

($\Longrightarrow$) Suppose $Q_{\mathcal{H}}$ has a uRAF $R' \subseteq R$ w.r.t I. We show the existence of a satisfying assignment $\mathcal{A} \vDash \mathcal{H}$. We define $\mathcal{A}$ formally below:

$$\mathcal{A}(v) = \begin{cases} 1, & \text{if } \exists l^R_{(i,j)} \in R' \text{ where } l_{(i,j)} \text{ is a } \textit{positive} \\ & \text{occurrence of } v \text{ in } \mathcal{H} \text{ (i.e. } l_{(i,j)} = v \text{ in } \mathcal{H}) \\ 0, & \text{if } \exists l^R_{(i,j)} \in R' \text{ where } l_{(i,j)} \text{ is a } \textit{negative} \\ & \text{occurrence of } v \text{ in } \mathcal{H} \text{ (i.e. } l_{(i,j)} = \overline{v} \text{ in } \mathcal{H}) \\ \textit{arbitrary}, & \text{otherwise} \end{cases}$$

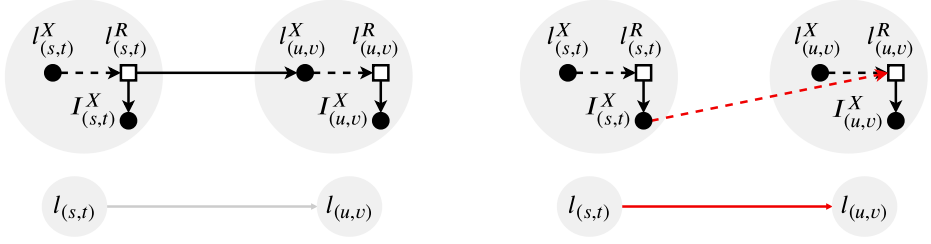
In other words, $\mathcal{A}$ assigns variables to make the literals $l_{(i,j)}$ in $\mathcal{H}$ positive-valued when the corresponding literal reaction $l^R_{(i,j)}$ is included in the uRAF R' (i.e. if $l^R_{(i,j)} \in R'$ then $\mathcal{A}(l_{(i,j)}) = 1$). By Lemma 4, $\mathcal{A}$ is well-defined. By Lemma 5, we have at least one literal reaction $l^R_{(i,1)}, l^R_{(i,2)}, l^R_{(i,3)}$ included for every clause $c_i = l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)} \in \mathcal{H}$. Each clause must therefore be satisfied under $\mathcal{A}$.

($\Longleftarrow$) Suppose that $\mathcal{A} \vDash \mathcal{H}$ is a satisfying assignment. We construct a uRAF R' for $Q_{\mathcal{H}}$ w.r.t I as follows: from each clause $c_i \in \mathcal{H}$, arbitrarily select one positive-valued literal $l_{(i,j)}$ under $\mathcal{A}$ and include the corresponding reaction $l^R_{(i,j)}$ in the reaction set R'. We now show that R' is a uRAF for $Q_{\mathcal{H}}$ w.r.t I with three claims.

Claim 1: R' is F-generated.

Proof. Each reaction has the food molecule $f \in F$ as its sole reactant (the CRS is elementary), so R' is trivially F-generated

Claim 2: R' is autocatalytic.



(a) Facilitating edge between literal triplets (b) Inhibiting edge between literal triplets

Fig. 9. For each literal reaction, literal molecule type and inhibitor for a literal $l_{(i,j)}$ (as grouped by the large grey circles), we denote a node named $l_{(s,t)}$ as the 'literal triplet' of $l_{(s,t)}$. Within our construction, there are two categories of edges that occur between literal triplets: Case (i) a literal reaction produces all the literal molecules in the subsequent clause; these are the types of edges displayed and discussed in Fig. 8. In (a) this is demonstrated by the reaction edge $l^R_{(s,t)} \longrightarrow l^X_{(u,v)}$ (solid black arrow). Between literal triplets, we denote this relationship by a grey directed edge and refer to it as a 'facilitating edge' between two literal triplets; this is also shown in (a). Case (ii) a literal inhibitor $I^X_{(s,t)}$ inhibits the literal reaction $l_{(u,v)}$ as $l_{(s,t)}$ and $l_{(u,v)}$ are opposite-parity occurrences of the same underlying variable in $\mathcal{H}$ (e.g. $l_{(s,t)} = v$ and $l_{(u,v)} = \overline{v}$). This is demonstrated in (b) through the inhibiting edge $I^X_{(s,t)} \longrightarrow l^R_{(u,v)}$ (dashed red arrow). Between literal triplets, we denote this relationship by a red directed edge and refer to it as an 'inhibiting edge'. This is also shown in (b). (Note that these edges will always occur in pairs; we have excluded the other direction in (b) for brevity.).

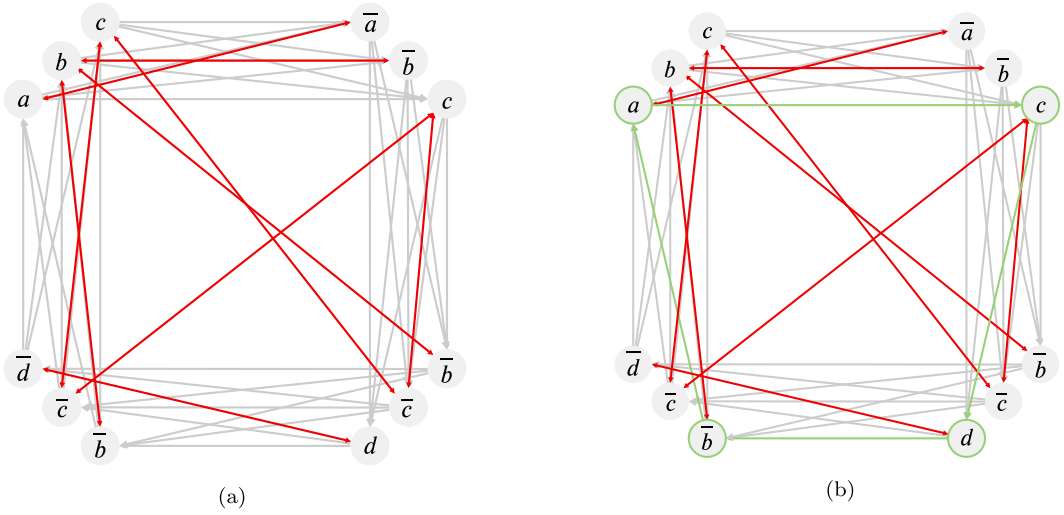


Fig. 10. Using our diagrammatic notation we show the overall construction corresponding to the formula $\mathcal{H} = (a \vee b \vee c) \wedge (\overline{a} \vee \overline{b} \vee c) \wedge (\overline{b} \vee \overline{c} \vee d) \wedge (\overline{b} \vee \overline{c} \vee \overline{d})$ in Fig (a). Every node represents a literal triple; grey and red arrows indicate facilitating and inhibiting edges, respectively (as described in Fig. 9). In Fig (b), we additionally highlight a cycle of reactions which form a uRAF in $\mathcal{Q}_{\mathcal{H}}$. Triplet nodes highlighted in green indicate that the corresponding literal reaction is included in the RAF (i.e. $l^R_{(i,j)}$ is included in the uRAF R' for the highlighted literal triplet $l_{(i,j)}$). The green highlighted arrows are still facilitating edges; they are simply intended to emphasise that a cycle of literal triplets has formed. Consequently, this corresponds to a satisfying assignment for $\mathcal{H}$, namely $\mathcal{A}(a) = 1, \mathcal{A}(c) = 1, \mathcal{A}(d) = 1, \mathcal{A}(b) = 0$.

Proof. Since one literal reaction $l^R_{(i,j)}$ is included in R' per clause, all literal molecules will be included in the food closure $\mathrm{cl}_{R'}(F)$. (Each literal reaction produces all literal molecules in the next clause.) Each reaction will therefore be catalysed.

Claim 3: R' is uninhibited.

Proof. Our inhibiting set I was defined so that for every pair of literals $l_{(s,t)}, l_{(u,v)} \in \mathcal{H}$ which correspond to opposite-parity occurrences of the same underlying variable in $\mathcal{H}$ (e.g. $l_{(s,t)} = v$ and $l_{(u,v)} = \overline{v}$), we have $(I^X_{(s,t)}, l^R_{(u,v)}) \in I$. Literals $l_{(s,t)}, l_{(u,v)}$ which are opposite-parity occurrences of the same underlying variable can never both be positive-valued under the same assignment; it follows no two such reactions $l^R_{(s,t)}, l^R_{(u,v)}$ will both be included in R' and so no reaction $l^R_{(u,v)}$ will be inhibited by an inhibitor $I^X_{(s,t)}$ (which is only produced by $l^R_{(s,t)}$). We conclude R' is uninhibited.

Finally, as we include only a polynomial number of reactions, molecule types, reactant/product edges and catalysis edges in $\mathcal{Q}_{\mathcal{H}}$ relative to the size of $\mathcal{H}$, we can construct $\mathcal{Q}_{\mathcal{H}}$ in polynomial-time. This completes the proof. □

We turn now to Question Q2. When a RAF set contains no subRAFs, it is said to be an *irreducible* RAF or, more briefly, an *irrRAF*. Note that a RAF is an irrRAF if it has the property that removing any single reaction from it results in a set that contains no RAF. Irreducible RAFs have been extensively used in structural analyses of RAFs in polymer models [20] and in extant metabolic systems [12], as well as to model 'coherent evolutionary units', an evolutionary analogue of cells (see [2]). Finding a single irrRAF within a RAF is easily shown to be computable in polynomial time [7]. However, finding the smallest irrRAF (or, equivalently, a minimum-size RAF) turns out to be NP-hard [20].

It has been shown earlier [20] that the problem of determining the size of a smallest RAF (which is necessarily an irrRAF) in a CRS is NP-hard. However, the problem of determining the size of a largest irrRAF in a CRS was previously of unknown complexity, and was a problem posed in [12]. Here, we show that this problem is not only NP-hard but it remains NP-hard when we restrict it to the setting of elementary CRS systems, which, in turn, answers a question posed in [11].

Theorem 4.2. *Determining the size of a largest irrRAF in an elementary CRS $\mathcal{Q}$ is NP-hard.*

To establish this result, we need to introduce the longest directed simple chordless cycle problem (see also the 'snake-in-the-box' problem [26]), which is known to be NP-hard [27]. For a digraph $G = (V, E)$, a simple chordless cycle is a sequence of distinct vertices $v_0, v_1, \ldots, v_{k-1}$ with $(v_i, v_{(i+1) \bmod k}) \in E$ for each $0 \leq i < k$, and where, for every pair of vertices v_i, v_j for $j \neq (i+1) \bmod k$, we have no edge $(v_i, v_j) \in E$. See Fig. 11 for an illustration.

Given a digraph $G = (V, E)$, we first describe a polynomial-time construction of a certain elementary CRS $\mathcal{Q}_G = (X, R, C, F)$ in three steps: (i) for each $v_i \in V$, let $v^R_i \in R$ be a reaction in $\mathcal{Q}_G$ (ii) for each $(v_i, v_j) \in E$, let $v^X_{(i,j)}$ be a molecule type in X which is produced by the reaction v^R_i and which catalyses the reaction v^R_j (by (i), v^R_i, v^R_j must exist in R) (iii) let $F = \{f\}$, with f being the sole reactant of all reactions ($\mathcal{Q}_G$ is elementary). Since we introduce a linear number of reactions, molecule types and edges with respect to $|G|$, our construction is computable in polynomial time. See Fig. 12 for an illustration of a single edge in the construction and Fig. 13 for a complete example.

Lemma 6. *Given a digraph $G = (V, E)$, let $\mathcal{Q}_G = (X, R, C, F)$ be the constructed CRS for G. A set $R' = \{v^R_0, v^R_1, \ldots, v^R_{k-1}\} \subseteq R$ is a RAF in $\mathcal{Q}_G$ if and only if for every vertex $v_j \in \{v_0, v_1, \ldots, v_{k-1}\} \subseteq V$ there exists another vertex $v_i \in \{v_0, v_1, \ldots, v_{k-1}\}$ with $(v_i, v_j) \in E$.*

Proof. Let $R' = \{v^R_0, v^R_1, \ldots, v^R_{k-1}\} \subseteq R$ be a subset of reactions in $\mathcal{Q}_G$ and accordingly let $\{v_0, v_1, \ldots, v_{k-1}\}$ be a subset of vertices from V. Since every reaction is trivially F-generated, a set $R' \subseteq R$ is a RAF in $\mathcal{Q}_G$ if and only if it is autocatalytic. Furthermore, the only catalysing molecule types in $\mathcal{Q}_G$ are those types $v^X_{(i,j)}$ for which $(v_i, v_j) \in E$. These molecule types are produced by v^R_i and catalyse v^R_j. A set of reactions $R' \subseteq R$ is therefore autocatalytic (and consequently a RAF) if and only if for each reaction $v^R_j \in R'$, there exists a reaction $v^R_i \in R'$ with $(v_i, v_j) \in E$. By construction, this happens if and only if for every vertex $v_j \in \{v_0, v_1, \ldots, v_{k-1}\}$, there exists another vertex $v_i \in \{v_0, v_1, \ldots, v_{k-1}\}$ with $(v_i, v_j) \in E$. □

Recall the definition of an irrRAF: a RAF set R' is said to be irreducible (an irrRAF) if R' contains no RAF as a strict subset. The following lemma links irrRAFs in $\mathcal{Q}_G$ to simple chordless cycles in G.

Lemma 7. *Given a digraph $G = (V, E)$, then $v_0, \ldots, v_{k-1}$ is a simple chordless cycle in G if and only if $\{v^R_0, \ldots, v^R_{k-1}\}$ is an irrRAF in $\mathcal{Q}_G$.*

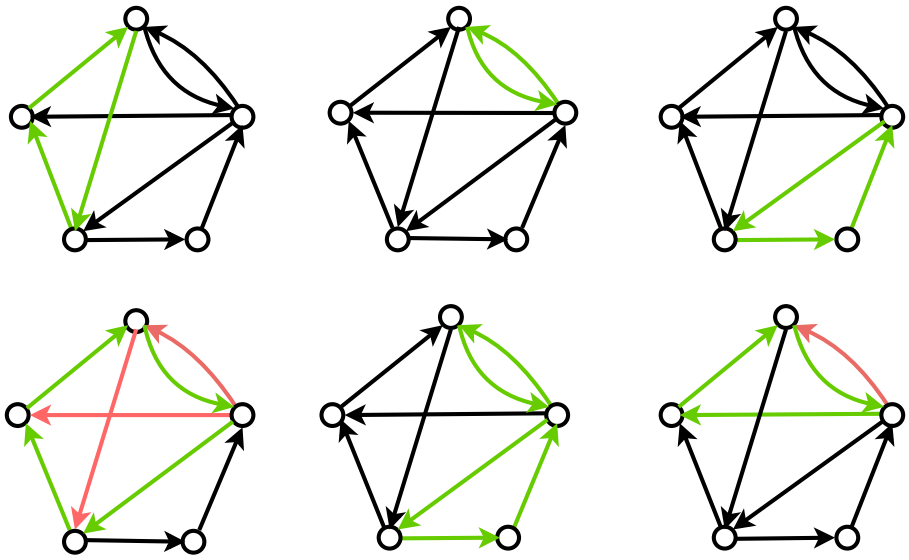


Fig. 11. The top three cycles (paths in green) are simple chordless cycles. The bottom three cycles (chords highlighted in red) are not simple chordless cycles (the bottom-middle one is not simple).

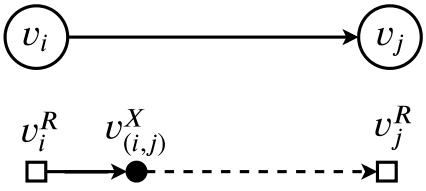


Fig. 12. For an edge $v_i \to v_j$ in a graph G (top), we construct a pair of reactions and a molecule type in $\mathcal{Q}_G$ (bottom). We hide the food molecule for brevity (it is the sole reactant to both reactions).

Proof. ($\Longrightarrow$) Suppose that $v_0, \dots, v_{k-1}$ is a simple chordless cycle in G. Since $v_0, \dots, v_{k-1}$ is simple, we can define a set $V' = \{v_0, \dots, v_{k-1}\}$ (simple implying no deduplication when translating to a set). As $v_0, \dots, v_{k-1}$ is a cycle, for every vertex $v_j \in V'$ there must exist another vertex $v_i \in V'$ with $(v_i, v_j) \in E$. By Lemma 6, the set of reactions $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\} \subseteq R$ must therefore be a RAF in $\mathcal{Q}_G$. We now show that R' is irreducible. Since $v_0, \dots, v_{k-1}$ is chordless in G, it follows that for all $0 \leq i < k$, the only vertex inbound to v_i from other vertices in V' is $v_{(i-1) mod k}$. It follows by construction that $v^X_{((i-1) \bmod k, i)}$ will be the only molecule type in $cl_{R'}(F)$ which catalyses v_i^R. We conclude that no strict subset of reactions $R'' \subset R'$ can be a RAF, otherwise, for some $0 \leq j < k$ we would have $v_j^R \in R''$ but with $v^R_{(j-1) \bmod k} \notin R''$ and therefore $v^X_{((i-1) \bmod k, i)} \notin cl_{R'}(F)$. Consequently, v_j^R would not be catalysed. We conclude that R' is an irrRAF in $\mathcal{Q}_G$.

($\Longleftarrow$) Suppose that $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$ is an irrRAF in $\mathcal{Q}_G$. Accordingly let $V' = \{v_0, v_1, \dots, v_{k-1}\}$ be a set of vertices from V. As R' is a RAF in $\mathcal{Q}_G$, by Lemma 6 it follows that for every vertex $v_j \in V'$ there exists another other vertex $v_i \in V'$ with $(v_i, v_j) \in E$. Now consider what happens if we remove a single vertex v from V'. In this case, there must exist a vertex v_j with no v_i such that $(v_i, v_j) \in E$ as otherwise by Lemma 6 we would see $V' - \{v\}$ corresponding to a subRAF $R' - \{v^R\} \subset R'$ (as R' is irreducible, this is a contradiction). It follows that for every vertex $v_j \in V'$, we have $(v_i, v_j) \in E$ for *exactly one* vertex $v_i \in V'$. Consequently, V' must either form a simple chordless cycle in G, or be comprised of the disjoint union of several simple chordless cycles in G. Suppose WLOG that V' comprises of two simple chordless cycles $v_0, v_1, \dots, v_{u-1}$ and $v_u, v_{u+1}, \dots, v_{k-1}$. By Lemma 6, both $v_0^R, v_1^R, \dots, v_{u-1}^R \subset R'$ and $v_u^R, v_{u+1}^R, \dots, v_{k-1}^R \subset R'$ must be RAFs in $\mathcal{Q}_G$. However, this is a contradiction as R' was assumed to be irreducible. We conclude that V' must contain the vertices of a single simple chordless cycle in G. □

Proof of Theorem 4.2. We apply Lemma 7 to the complexity questions surrounding RAF irreducibility. Recall that this lemma states that if we fix a digraph $G = (V, E)$, then $v_0, \dots, v_{k-1}$ is a simple chordless cycle in

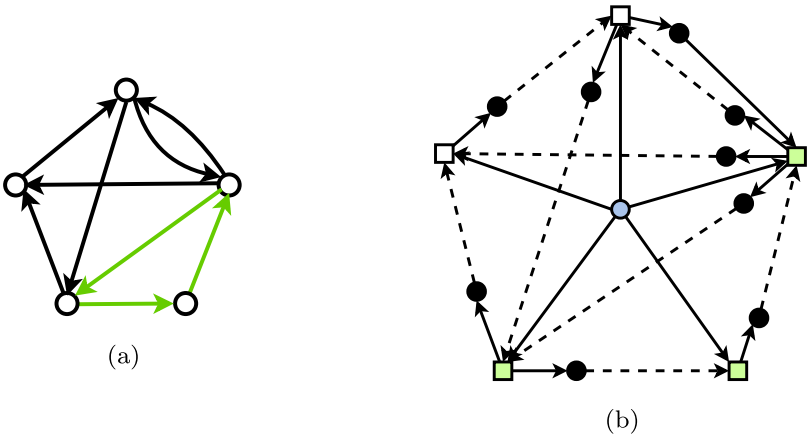


Fig. 13. A digraph G displayed on the left (a) with the constructed elementary CRS $\mathcal{Q}_G$ on the right (b) (the food molecule is at the centre, highlighted in blue). The simple chordless cycle in G of size 3 (edges highlighted in green) corresponds to an irrRAF in $\mathcal{Q}_G$ of size 3 (reactions highlighted in green).

G if and only if $\{v_0^R, \ldots, v_{k-1}^R\}$ is an irrRAF in Q_G. Finding the longest simple chordless cycle is a known NP-hard problem [27], so the proof follows immediately from Lemma 7. □

5. Concluding comments

In this paper, we have settled a number of computational complexity questions concerning the detection of subRAFs of particular types.

Our results also have relevance to other computational complexity questions arising in RAF theory. We will describe two examples (without details). First, by using Lemma 7, it can be shown that the question of whether or not a RAF can be partitioned into disjoint irrRAFs is NP-hard in general. Second, many intractable questions concerning RAFs have polynomial-time solutions when every reaction has all its reactants in the food set [11]. However, relaxing this condition so that each reaction has all its reactions within k reaction steps of the food set (a possible direction suggested in [11]) turns out to be NP-hard, even when $k = 2$ (by arguments similar to those used in the proof of Theorem 3.1).

As many of the interesting questions arising in RAF theory turn out to be NP-hard, it may be helpful to develop tailored SAT-solver or Linear Programming techniques for analysing CRS datasets (e.g. to determine uRAF and minimum RAF sizes), and to thus extend the computational results beyond what was previously tractable. We plan to investigate this in future work. Meantime, heuristic approaches can still provide a possible way to search for computationally difficult types of subRAFs, and such approaches have been implemented in RAF software.[1]

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W.F. Martin, M. Preiner, Origins of life, theories of, in: Reference Module in Life Sciences, Elsevier, 2017, http://dx.doi.org/10.1016/B978-0-12-809633-8.02403-1.
[2] V. Vasas, C. Fernando, M. Santos, S. Kauffman, E. Szathmáry, Evolution before genes, Biol. Direct 7 (1) (2012).
[3] M. Preiner, J.C. Xavier, A.d.N. Vieira, K. Kleinermanns, J.F. Allen, W.F. Martin, Catalysis, autocatalysis and the origin of metabolism, Interface Focus 9 (2019) 20190072.
[4] P. Erdös, A. Rényi, On the evolution of random graphs, in: Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, pp. 17–61.
[5] S.A. Kauffman, Autocatalytic sets of proteins, J. Theoret. Biol. 119 (1) (1986) 1–24.
[6] W. Hordijk, J. Hein, M. Steel, Autocatalytic sets and the origin of life, Entropy 12 (7) (2010) 1733–1742.
[7] W. Hordijk, M. Steel, Detecting autocatalytic, self-sustaining sets in chemical reaction systems, J. Theoret. Biol. 227 (4) (2004) 451–461.
[8] M. Steel, The emergence of a self-catalysing structure in abstract origin-of-life models, Appl. Math. Lett. 13 (3) (2000) 91–95.
[9] G. Ashkenasy, R. Jegasia, M. Yadav, M.R. Ghadiri, Design of a directed molecular network, Proc. Natl. Acad. Sci. USA 101 (10) (2004) 10872–10877.
[10] N. Vaidya, M.L. Manapat, Chen I.A., R. Xulvi-Brunet, E.J. Hayden, N. Lehman, Spontaneous network formation among cooperative RNA replicators, Nature 491 (2012) 72–77.
[11] M. Steel, W. Hordijk, J.C. Xavier, Autocatalytic networks in biology: structural theory and algorithms, J. Royal Soc. Interface 16 (2019) 20180808.
[12] F.L. Sousa, W. Hordijk, M. Steel, W.F. Martin, Autocatalytic sets in E. coli metabolism, J. Syst. Chem. 6 (1) (2015) 4.
[13] J.C. Xavier, W. Hordijk, S. Kauffman, Steel M., W.F. Martin, Autocatalytic chemical networks at the origin of metabolism, Proc. R. Soc. Lond. [Biol.] 287 (2020) 20192377.
[14] R. Cazzolla Gatti, B. Fath, W. Hordijk, S. Kauffman, R. Ulanowicz, Niche emergence as an autocatalytic process in the evolution of ecosystems, J. Theoret. Biol. 454 (2018) 110–117.
[15] W. Hordijk, M. Steel, Chasing the tail: The emergence of autocatalytic networks, Biosystems 152 (2017) 1–10.
[16] L. Gabora, M. Steel, Autocatalytic networks in cognition and the origin of culture, J. Theoret. Biol. 431 (2017) 87–95.
[17] D. Loutchko, Semigroup models for biochemical reaction networks, 2019, arXiv:1908.04642.
[18] W. Hordijk, M. Steel, P. Dittrich, Autocatalytic sets and chemical organizations: Modeling self-sustaining reaction networks at the origin of life, New J. Phys. 20 (2018) 015011.
[19] R. Wolfenden, M.J. Snider, The depth of chemical time and the power of enzymes as catalysts, Acc. Chem. Res. 34 (12) (2001) 938–945.
[20] M. Steel, W. Hordijk, J. Smith, Minimal autocatalytic networks, J. Theoret. Biol. 332 (2013) 96–107.
[21] W. Hordijk, J.I. Smith, M. Steel, Algorithms for detecting and analysing autocatalytic sets, Algorithms Mol. Biol. 10 (15) (2015).
[22] W. Hordijk, M. Steel, Autocatalytic sets extended: Dynamics, inhibition, and a generalization, J. Syst. Chem. 3 (5) (2012).
[23] J. Smith, M. Steel, W. Hordijk, Autocatalytic sets in a partitioned biochemical network, J. Syst. Chem. 5 (2) (2014).
[24] E. Mossel, M. Steel, Random biochemical networks: the probability of self-sustaining autocatalysis, J. Theoret. Biol. 233 (3) (2005) 327–336.
[25] S. Jain, S. Krishna, Autocatalytic sets and the growth of complexity in an evolutionary model, Phys. Rev. Lett. 81 (25) (1998) 5684–5687.
[26] W.H. Kautz, Unit-distance error-checking codes, IRE Trans. Electron. Comput. EC-7 (2) (1958).
[27] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman Co., New York, NY, USA, 1990.

[1] *CatlyNet* (https://github.com/husonlab/catlynet).