

# Combinatorial results for network-based models of metabolic origins

Oliver Weller-Davies<sup>a</sup>, Mike Steel<sup>b</sup> and Jotun Hein<sup>c</sup>

<sup>a</sup> *Department of Statistics, Oxford University, Oxford UK*  
Email: *oliver.weller-davies@keble.ox.ac.uk (corresponding author)*.

<sup>b</sup> *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand.* Email:  
*mike.steel@canterbury.ac.nz*

<sup>(c)</sup> *Department of Statistics, Oxford University, Oxford UK* Email: *hein@stats.ox.ac.uk*

---

## 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. One of the classic applications of this theory (arising from Stuart Kauffman's pioneering work in the 1980s) involves networks of polymers under cleavage and ligation reactions; in the first part of this paper, we provide the first exact description of the number of such reactions under various model assumptions. Conclusions from earlier studies relied on either approximations or asymptotic counting, and we show that the exact counts lead to similar (though not always identical) asymptotic results. In the second part of the 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 exist turns out to be NP-complete.

---

*Keywords:* Catalytic reactions system, origin of metabolism, polymer, computational complexity

## 1. Introduction

The origin of life remains an unsolved challenge in science. Once considered a problem ‘beyond science’, many researchers now believe its solution may not be far off [18, 28]. 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 [20]. Initially motivated by the emergent qualitative properties of discrete random networks (such as early works of Paul Erdős and Alfred Rényi [3]), one approach to formally capture ‘life-like’ emergence in a chemical system involves the notion of a collectively autocatalytic sets, a concept pioneered for polymer systems by Stuart Kauffman [13]. This approach was later developed more formally as the notion of *Reflexively Auto-catalytic F-generated sets* (RAFTs). Such sets couple together two basic requirements for any living system: reactions are catalysed by molecule 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 by using just 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, 24], RAFTs have been investigated in both simulated and laboratory-based systems of early metabolism ([1, 27] as discussed in [26]) as well as in the analysis of metabolism in the bacterium *Escherichia coli* [22] and a recent study into ancient metabolism revealed by analysing large biochemical databases [30]. RAFT theory has also been applied in other fields such as ecology [2], economics [10] and cultural evolution [4], and the theory has recently re-expressed within the language of semigroups [16].

In this paper, we derive a number of new results in RAFT theory, answering some outstanding complexity questions that have been posed in earlier papers. We also provide a more exact treatment of the enumeration of reactions in polymer models than has been given in any previous work. We begin in Section 2, by providing basic definitions and summarizing some results and questions from RAFT theory. In Section 3 we review existing polymer models (based on ‘oriented’ polymers), and then investigate the consequences of variations on this model that have been overlooked in earlier work, as well as derive precise enumerative and asymptotic formulae for the number of reactions in these cases. We then explore another polymer model (‘non-oriented’) that does not appear to have been considered before in the RAFT setting. We again derive exact and asymptotic formulae, and the consequences of these are also briefly discussed.

In Section 4, we investigate the computational complexity of two problems (posed in [11] and [26]) concerning RAFTs that are ‘closed’ (a biochemically relevant condition we describe there). Namely, (i) does a RAFT set contain a closed subRAFT set? (ii) Is there a closed RAFT set that is ‘ uninhibited’ (i.e., no molecule inhibits any reaction) in the simple case when (just) a single molecule 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 5, 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 [22], [26]): (i) What is the size of the largest irreducible RAFT? (ii) Does an uninhibited RAFT exist? We show that these questions are NP-hard in this elementary CRS setting. We end with some brief concluding comments.

## 2. Definitions and mathematical preliminaries

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

- $X$  denotes a set of *molecule types*.
- $R$  denotes a set of reactions between sets of molecule types. Specifically, each  $r = (A, B) \in R$  is a pair of *sets* of molecule types  $A, B \subseteq X$ . The molecule types in  $A$  are referred to as the *reactants* of the reaction; those in  $B$  as the *products* of the reaction.
- $C \subseteq X \times R$  denotes a *catalysation* assignment; where if  $(x, r) \in C$ , we say that the molecule  $x$  *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.

A simple example of a CRS is the binary polymer model introduced by [13]. Suppose, under appropriate chemical conditions, that we have a set of polymeric molecules. Here, ‘polymeric’ refers to macromolecules formed from the repeated concatenation of sub-molecules (monomers), such as how the polymer RNA is formed by repeated joins of the bases (monomers) A,U,C,G, or how proteins are formed from amino acids. Polymers can be represented as strings which leads naturally to two types of reaction. The first type, ligation reactions, denote the concatenation of two polymers (strings) to form a longer polymer. For example, in the binary (a,b) polymer setting, we have:



The second type of reaction (cleavage) is the reverse of ligation; it denotes the splitting up of a polymer into appropriate constituent sub-strings; for example:



We fix a constant  $n$  to denote the maximum length of polymers in the system. Within this setting, we also assume that all polymers of length  $\leq t$  are freely available and constantly replenished (a food set). Catalysis is typically modelled as a stochastic process – where each polymer catalyses a given ligation/cleavage reaction with a certain probability.

Any CRS (not only polymer models) can be represented as a directed graph with two types of vertices (molecule types and reactions) and two types of directed edges, one showing the flow of molecule types into and out of reactions, and the other type being catalysation of a reaction by a molecule type. When drawing these graphs, we typically follow the convention that molecule types are represented by circles, reactions by squares, reactant and product directed edges by solid-line arrows and catalysation directed edges by dashed-line arrows. See Fig. 1.

Given a CRS  $\mathcal{Q} = (X, R, C, F)$ , a subset of reactions  $R' \subseteq R$  is a *RAF* set (or, more briefly, a RAF) for  $\mathcal{Q}$  if  $R'$  is non-empty and both of the following conditions hold:

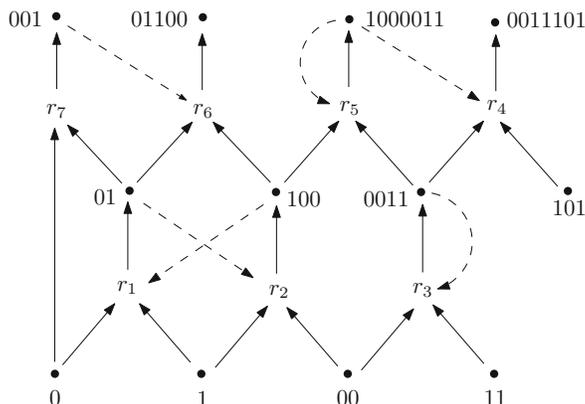


Figure 1: A simple CRS based on polymers over a two-letter alphabet (0,1), with a food set  $F = \{0, 1, 00, 11\}$  and seven reactions ( $r_1 - r_7$ ) (from [21]). Dashed arrows indicate catalysis; solid arrows show reactants entering a reaction and products leaving. In this CRS, there are exactly four RAFs (defined below), namely  $\{r_1, r_2\}$ ,  $\{r_3\}$ ,  $\{r_1, r_2, r_3\}$  and  $\{r_1, r_2, r_3, r_5\}$ .

(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  $r_0, r_1, \dots, r_n$  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  $j < i$ .

Roughly speaking, a RAF set is a subset of reactions that is both collectively autocatalytic (i.e. each reaction is catalysed by some molecule type involved in the reaction subset) and self-sustaining (i.e. each molecule type required for the reactions to occur can be built up starting from just the food set; using reactions only within the reaction subset). The RAF concept is illustrated in Fig. 2(a), along with the weaker notion of a ‘pseudo-RAF’ (Part (b) of Fig. 2) that satisfies the (RA) condition but fails to be  $F$ -generated.

RAFTs underly the metabolism of both existing cellular life [22] and also arise in laboratory models of early life [1, 26, 27]. 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 more likely be involved). Biochemical catalysts not only speed up reactions by many orders of magnitude but they also allow reactions to be synchronised [29]. 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 able to be built up from the food set by a sequence of reactions within the system. These two features are combined into the two conditions (RA) and (F).

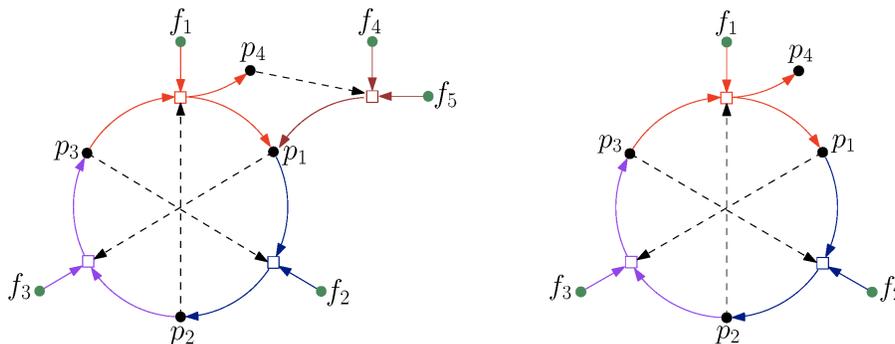


Figure 2: The set in (a) is an example of a RAF, while (b) is not a RAF (since it fails to be  $F$ -generated). Here, food is indicated by  $f_*$  (green vertices) and catalysis by dashed arrows.

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. 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 [8]. We describe this maxRAF algorithm shortly. Any given RAF  $R'$  may contain another RAF  $R''$  as a strict subset, in which case we say that  $R''$  is a *subRAF* of  $R'$ .

### 2.1. The closure of a set of molecule types

Fix a CRS  $\mathcal{Q} = (X, R, C, F)$ . Given a set of reactions  $R' \subseteq R$  and a set of molecule types  $X' \subseteq X$ , the molecule closure  $\text{cl}_{R'}(X')$  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 [8]).

In words,  $\text{cl}_{R'}(X')$  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 [25]) that incorporates the notion of molecule closure.

**Lemma 1.** *Given a CRS  $\mathcal{Q} = (X, R, C, F)$ , a subset of reactions  $R' \subseteq R$  is a RAF for  $\mathcal{Q}$  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  $x \in \text{cl}_{R'}(F)$  with  $(x, r) \in C$ ;*
- (ii) *For every reaction  $r = (A, B) \in R'$ ,  $A \subseteq \text{cl}_{R'}(F)$ .*

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) are 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  $R$ :

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

where  $R_{j+1}$  is the subset of reactions in  $R_j$  that have all their reactants and at least one catalyst in  $\text{cl}_{R_j}(F)$ . 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 [8]. A rudimentary runtime analysis of the gives us  $O(|R|^3|X|)$  time and  $O(|X| + |R|)$  space [8], although optimised implementations exist which tend to run sub-quadratically (in time) with the number of reactions [7].

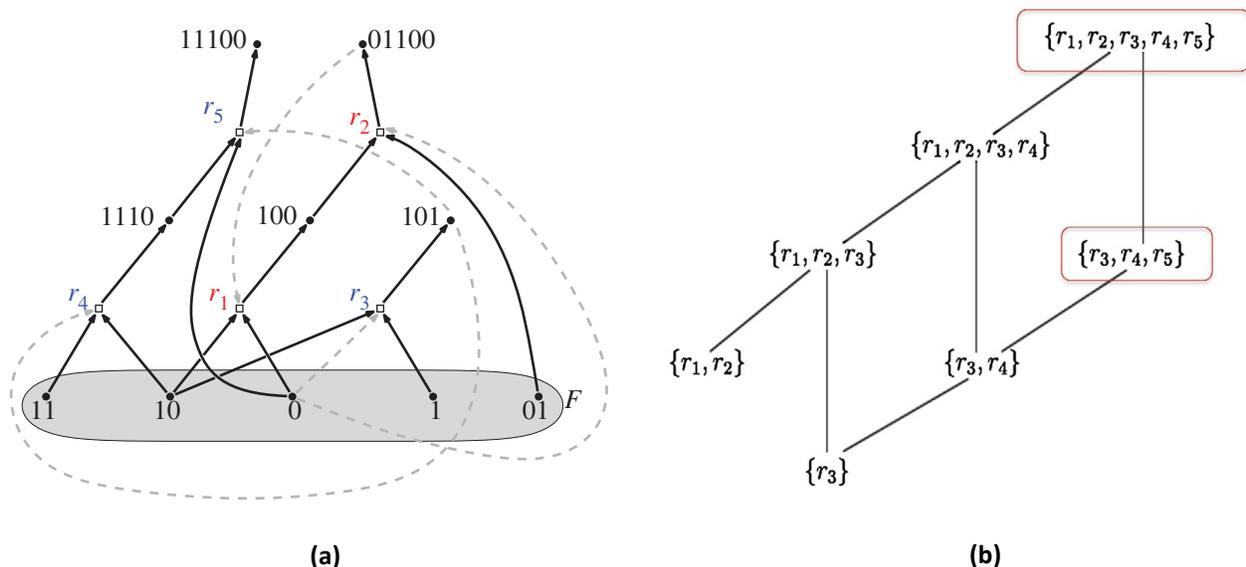


Figure 3: *Left*: The CRS above is a portion of the binary polymer model instance with food molecule types highlighted in the lower shaded section. The maxRAF of the CRS contains all reactions  $\{r_1, r_2, r_3, r_4, r_5\}$ . *Right*: The partially ordered set (poset) of RAFs for the CRS. The two circled RAFs are closed (a notion defined and studied in Section 4).

### 3. Polymer models: Combinatorial enumeration and its role in RAF emergence

We now describe polymer models more formally, following [19, 24], based on earlier work by [13]. In particular, we provide a mathematical analysis of some hitherto overlooked subtleties and variations concerning the classification and counting of molecule types and reactions within polymer models. The reason why such enumeration is important is that the key mathematical results concerning the emergence of RAFs in polymer models when the catalysis rates pass a certain threshold depend ultimately on the

number of reactions and molecule types, and the ratio of these two quantities. We describe this relevance in Section 3.3.

We first describe the set of molecule types in polymer systems. Consider an arbitrary finite alphabet  $\Sigma$  of size  $k \geq 2$  and let  $\Sigma^+$  denote the set of all finite-length (ordered) words formed from this alphabet. For  $w \in \Sigma^+$ , we will let  $|w|$  denote the length (number of characters) in  $w$ . We refer to  $w$  as an *oriented polymer* over the alphabet  $\Sigma$ .

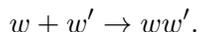
In the original polymer model from [13], developed further in [19, 24], the set  $X$  of molecule types consists of oriented polymers over the alphabet  $\Sigma$ . More precisely, for  $n \geq 1$ , let  $X_n$  be the set of words  $w \in \Sigma^+$  of length at most  $n$  (i.e.  $1 \leq |w| \leq n$ ). In this case, the number of different molecule types  $|X_n| = |\{w \in \Sigma^+ : |w| \leq n\}|$  is just the sum  $\sum_{i=1}^n k^i$  (all non-empty words of length  $\leq n$ ) and therefore:

$$|X_n| = \frac{k^{n+1} - k}{k - 1} \sim \frac{k^{n+1}}{k - 1} \quad (1)$$

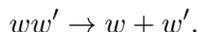
where here and below  $\sim$  denotes asymptotic equivalence as  $n$  grows (we regard  $k$  as fixed throughout). In other words,  $f(n) \sim g(n)$  if and only if  $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$  (for details, see [19]).

### 3.1. Reaction sets and their enumeration

In the polymer model, reactions consist of two complementary types. The first is a *ligation* reaction where two polymers are concatenated, formally:



The second is a *cleavage* reaction where a polymer is split in two:



Note that the cleavage and ligation reactions are reversals of each other and thus are in one-to-one correspondence. We will therefore mostly concentrate on enumerating ligation reactions.

We now highlight a subtle but important distinction in the following example. Consider the two cleavage reactions:



and



These two reactions have the same reactants and the same product, and the only distinction is the order in which the reactants appear on the left. Thus it is tempting to regard (2) and (3) as the same reaction; we call this *Convention A*, and it was the one that was assumed in [19, 24]. However, given that the polymers are oriented, one might regard the first reaction above as attaching  $ab$  to the left-hand end of the (oriented) polymer  $abab$  whereas the second reaction is attaching  $ab$  to the right-hand end of the polymer  $abab$ ; in this way, the reactions can be regarded as different. We call this *Convention B*, and it has also been

tacitly assumed in other papers on the studies involving the model, particularly with simulations. Note that this distinction between these two conventions vanishes in the non-oriented setting.

Let  $R_n^A$  and  $R_n^B$  denote the set of ligation reactions involving oriented polymers of size less or equal to  $n$  under Conventions A and B, respectively. The one-to-one correspondence between cleavage reactions and ligation reaction holds both for Convention A and for Convention B. Calculating the size of  $R_n^B$  is easy and was carried out in the earlier papers cited above. For  $i \geq 2$ , if we let  $S(i)$  denote the number of ordered triples of oriented polymers  $(u, v, w)$  where  $uv = w$  and  $|w| = i$ , then:

$$S(i) = (i - 1)k^i, \quad (4)$$

and we have:

$$|R_n^B| = \sum_{i=1}^n S(i) = \frac{nk^{n+1}}{k-1} - \frac{k^2(k^n-1)}{(k-1)^2} \sim \frac{nk^{n+1}}{k-1}. \quad (5)$$

For Convention A, the expression for  $|R_n^A|$  is more delicate. To describe this, fix an alphabet  $\Sigma$  of size  $k$ , and let  $U(i)$  denote the number of unique unordered sets  $\{u, v, z\}$  for  $u, v, z \in \Sigma^+$  with  $uv = z$  and  $|z| = i$ . We first state a lemma, which is a consequence of a more general result from Lyndon and Schützenberger [17].

**Lemma 2.** *If  $uv = vu$  for non-empty words  $u$  and  $v$ , then  $u$  and  $v$  are powers of a common word; that is, there exists a word  $w$  such that  $w^i = u$  and  $w^j = v$  for some  $i, j \geq 1$ . Therefore,  $uv = w^{i+j}$ . Here,  $w^i$  denotes  $i$  repeated copies of  $w$  joined together.*

□

This lemma allows us to characterise strings of the form  $uv = vu$  with  $u, v \in \Sigma^+$ . In particular, if  $z$  is a word of the form  $z = uv = vu$  for  $u, v \in \Sigma^+$ , then  $z$  is necessarily a word formed from a repeated sub-word. Words that can be decomposed in such a way are often referred to as *periodic*, where their *period* is taken to be the smallest sub-word that repeatedly joins to form the word; for instance, the word *abcabcabcabc* has period *abc*. The period of a word, if it exists, must be unique. Conversely, words that cannot be formed from repeated substrings in such a way are often referred to as *aperiodic* or *primitive* words (e.g. *abcdef* is aperiodic).

This leads to the following expression for  $|R_n^A|$  (details of the proof of Part (i) are in the Appendix), where  $\mu$  denotes the (classical) Möbius function for the partial ordered set of positive integers under division [23], defined by:

$$\mu(x) = \begin{cases} 1, & \text{if } x = 1; \\ 0, & \text{if } x \text{ is a square number;} \\ (-1)^r, & \text{if } x \text{ is square-free with } r \text{ distinct prime factors.} \end{cases} \quad (6)$$

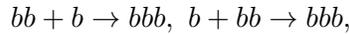
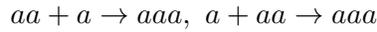
**Theorem 3.1.** *Fix an alphabet  $\Sigma$  of size  $k$ . Let  $|R_n^A|$  denote the set of cleavage and ligation reactions in the oriented polymer model under Convention A, with an alphabet of size  $k$  and polymer length of at most  $n$ . We then have the following:*

(i)  $|R_n^A| = \sum_{i=2}^n U(i)$ , where:

$$U(i) = (i-1)k^i - \sum_{d|i, d < i} \left\lfloor \frac{i/d-1}{2} \right\rfloor \sum_{d'|d} \mu\left(\frac{d}{d'}\right) k^{d'} \quad (7)$$

(ii)  $S(i) - U(i) = O(k^{i/3})$ , where  $S(i)$  is given in (4), and so  $|R_n^A| \sim |R_n^B|$  as  $n$  grows.

To illustrate Theorem 3.1, consider the binary polymer model (i.e.  $k = 2$ ). By Part (i) of Theorem 3.1 we have  $U(2) = S(2) = 4$  and  $U(3) = 2 \cdot 2^3 - 2 = 14$ , while  $S(3) = 16$ . Thus  $|R_3| = U(2) + U(3) = 18$ . In this case, for the two reaction pairs:



the reactions within each pair are counted separately by  $S(3)$  but only once by  $U(3)$ . The difference  $S(i) - U(i)$  quantifies the difference between Conventions A and B when the product of a cleavage reaction has size  $i$ . A graph of this difference is shown in Fig. 4.

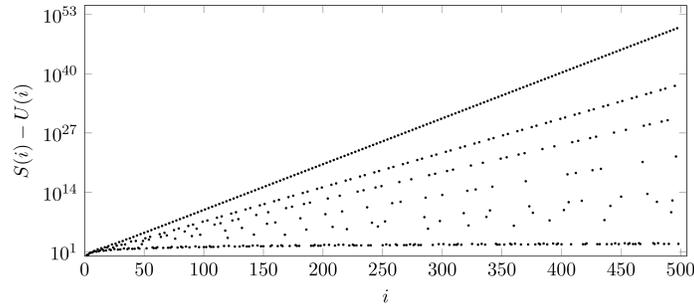


Figure 4: The difference between  $S(i)$  and  $U(i)$  as  $i$  increases from 1 to 500

*Outline Proof of Theorem 3.1 :* The proof of Part (i) is given in the Appendix. For Part (ii), Lemma 2 gives:  $S(i) - U(i)$  is the number of pairs of words of the form  $w^l, w^m$  where  $1 \leq l, m \leq i/|w|$  satisfies  $(l+m)|w| = i$  and  $l < m$ . Let  $r := l+m$  (so  $|w| = i/r$ ). Since  $l, m \geq 1$  and  $l \neq m$ , we have  $r \geq 3$ . Now if  $|w| = i/r$  there are  $k^{i/r}$  choices for  $|w|$  and so  $S(i) - U(i)$  is bounded above, as claimed, by:

$$\sum_{r=3}^i (r-1)k^{i/r} = \Theta(k^{i/3}).$$

The second claim in Part (ii) now follows by applying Part (i) and Eqn. (5) to give:

$$|R_n^A| = \sum_{i=2}^n U(i) \sim \sum_{i=2}^n S(i).$$

The asymptotic equivalence here holds because the sum of terms of order  $k^{i/3}$  from  $i = 1$  to  $i = n$  is of order  $k^{(n+1)/3}$  and thus is asymptotically negligible compared with  $|R_n^A|$ .  $\square$

### 3.2. Non-oriented polymers

The choice to consider polymers as oriented is not only computationally convenient; it also has some connection with bio-molecular sequence data, where (for example) DNA polymerase has an orientation (from the 5' to 3' end of the sequence). However, it may also be useful, particularly in the study of early biochemistry, to regard polymers as non-oriented; in other words, a polymer such as *accb* should be considered as identical to *bcca*, since one can rotate the first molecule in space to obtain the second. Thus, each oriented polymer  $w$  can be viewed as equivalent to its ‘reverse’ oriented polymer  $w^-$  obtained by reversing the order of the letters in the polymer. Note that a polymer is equal to its associated reverse polymer precisely if it is a palindromic polymer (e.g. *acca*). In this way, the set  $\Sigma^+$  is partitioned into pairs and singleton classes under the equivalence relation  $w \sim w'$  if  $w' = w$  or  $w' = w^-$ . We refer to these equivalence classes as *non-oriented polymers* and we will write  $\bar{w}$  for the equivalence class of  $w$ . Let  $\bar{X}_n = \{\bar{w} : w \in \Sigma, |w| \leq n\}$  denote the set of non-oriented polymers (equivalence classes) for words  $w \in \Sigma$  with  $1 \leq |w| \leq n$ .

*Example:*

For  $\Sigma = \{a, b\}$ , we have:  $|\bar{X}_2| = 5$ , since

$$\bar{X}_2 = \{\{a\}, \{b\}, \{aa\}, \{bb\}, \{ab, ba\}\}.$$

Similarly,  $|\bar{X}_3| = 11$ , since:

$$\bar{X}_3 = \{\{a\}, \{b\}, \{aa\}, \{bb\}, \{ab, ba\}, \{aaa\}, \{bbb\}, \{aba\}, \{bab\}, \{abb, bba\}, \{aab, baa\}\}.$$

By contrast,  $|X_2| = 6$  and  $|X_3| = 14$ .

#### Lemma 3.

$$|\bar{X}_n| = \begin{cases} \frac{k}{2(k-1)} (k^n + 2k^{n/2} - 3), & \text{if } n \text{ is even;} \\ \frac{k}{2(k-1)} (k^n + k^{(n+1)/2} - 1), & \text{if } n \text{ is odd.} \end{cases}$$

In particular,  $|\bar{X}_n| \sim \frac{1}{2}|X_n|$  as  $n$  grows.

*Proof:* Let  $p_i$  (respectively,  $n_i$ ) denote the number of palindromic (respectively non-palindromic) oriented polymers of length  $i$ . We have:

$$|\bar{X}_n| = \sum_{i=1}^n \frac{1}{2} n_i + \sum_{i=1}^n p_i = \frac{1}{2} \left( \sum_{i=1}^n k^i + \sum_{i=1}^n p_i \right),$$

where the second equality follows from the identity:  $n_i = k^i - p_i$ . Now,  $p_i$  is equal to  $k^{i/2}$  when  $i$  is even and is equal to  $k^{(i-1)/2} \times k = k^{(i+1)/2}$  when  $i$  is odd. Applying the geometric sequence identity

$\sum_{j=1}^m x^j = x(x^m - 1)/(x - 1)$  establishes the first claim. The second claim follows by combining this result with Eqn. (5).  $\square$

Next, we consider the enumeration of involving non-oriented polymers. Consider the following ligation reaction:

$$r : \bar{u} + \bar{v} \rightarrow \bar{z}, \quad (8)$$

where  $u, v$  and  $z$  are oriented polymers. Note that, for any two oriented polymers  $x, y$  we always have  $(xy)^- = y^-x^-$  and the reaction  $r$  in (8)) holds whenever  $z$  is one of the following:

$$z = uv \quad (= (v^-u^-)^-),$$

$$z = uv^- \quad (= (vu^-)^-),$$

$$z = u^-v \quad (= (v^-u)^-),$$

$$z = u^-v^- \quad (= vu)^-).$$

Thus  $\bar{u}$  and  $\bar{v}$  could ligate to form up to four different non-oriented polymers, depending on how many of these four possible  $z$  are equivalent to each other (allowing reversals). There are four cases to consider:

- (i)  $u$  and  $v$  are non-palindromic polymers (and  $u \neq v, v^-$ ). In this case, there are four distinct choices of  $z$  (no two of which are a reversal of each other), so we obtain four distinct reactions of the type in (8).
- (ii) If  $u, v$  are both non-palindromic and  $u = v$  or  $u = v^-$  we obtain three distinct choices of  $z$  and so two additional distinct reactions of the type in (8).
- (iii) If one of  $u$  or  $v$  is a palindromic polymer and the other is not, we obtain two distinct choices of  $z$  and so two distinct reactions of the type in (8).
- (iv) If both  $u$  and  $v$  are palindromic polymers, then all choices of  $z$  are equivalent and so there is just one reaction of the type in (8).

*Example:*

As an example of Case (i), consider the reaction of type (i):  $\bar{ab} + \bar{baa} \rightarrow \overline{abbaa}$ . The reactants on the left, give rise to three additional distinct reactions, namely:

$$\bar{ab} + \bar{baa} \rightarrow \overline{abaab}, \quad \bar{ab} + \bar{baa} \rightarrow \overline{babaa}, \quad \text{and} \quad \bar{ab} + \bar{baa} \rightarrow \overline{baaab}.$$

Next, consider the reaction of Type (ii):  $\bar{ab} + \bar{ba} \rightarrow \overline{abba}$ . The reactants on the left also give rise to:

$$\bar{ab} + \bar{ba} \rightarrow \overline{baba} \quad \text{and} \quad \bar{ab} + \bar{ba} \rightarrow \overline{baab}.$$

Next, consider the reaction of Type (iii):  $\bar{aa} + \bar{ab} \rightarrow \overline{aaab}$ . The reactants on the left give rise to one additional reaction, namely  $\bar{aa} + \bar{ab} \rightarrow \overline{aaba}$ .

Finally, the reaction of Type (iv)  $\bar{aa} + \bar{bab} \rightarrow \overline{aabaab}$  gives rise to no further reactions.

By applying these cases, once can, in principle, count the number  $N_{r,s}$  of cleavage reactions in which two non-oriented polymers of size  $r$  and  $s$  are combined to give a non-oriented polymer of length  $i = r + s$ . First, suppose that  $r < s$ , so that Case (ii) cannot arise. We then have:

$$N_{r,s} = (2n_r + p_r) \cdot (2n_s + p_s),$$

where, as before,  $n_m$  (respectively,  $p_m$ ) is the number of non-palindromic (respectively, palindromic) oriented polymers of length  $m$ .

The case where  $r = s$  (in which case,  $i = 2r$  is even) requires a separate description, since Case (ii) can now also arise. We have the following identity for the case  $r = s$ :

$$N_{r,r} = 4n_r(n_r - 1) + 3n_r + 2n_r p_r + p_r(p_r + 1)/2.$$

*Example:*

Consider the case  $r = s = 2$ . We have  $n_2 = 1, p_2 = 2$  and so  $N_{2,2} = 10$ . These correspond to the following reactions, classified according to the cases described above. For Case (i) we have  $4n_2(n_2 - 1) = 0$ , (i.e. no reactions are possible in this case).

For Case (ii),  $3n_r = 3$ :  $\overline{ab} + \overline{ab} \rightarrow \overline{abba}$ ,  $\overline{ab} + \overline{ab} \rightarrow \overline{abab}$ ,  $\overline{ab} + \overline{ab} \rightarrow \overline{baab}$ .

For Case (iii),  $2n_2 p_2 = 4$ :

$$\begin{aligned} \overline{aa} + \overline{ab} &\rightarrow \overline{aaab}, \quad \overline{aa} + \overline{ab} \rightarrow \overline{aaba}, \\ \overline{bb} + \overline{ab} &\rightarrow \overline{abbb}, \quad \overline{bb} + \overline{ab} \rightarrow \overline{babb} \end{aligned}$$

For Case (iv),  $p_2(p_2 + 1)/2 = 3$ :  $\overline{aa} + \overline{aa} \rightarrow \overline{aaaa}$ ,  $\overline{bb} + \overline{bb} \rightarrow \overline{bbbb}$ ,  $\overline{aa} + \overline{bb} \rightarrow \overline{aabb}$ .

The total number of ligation reactions in the non-oriented setting in which the product polymer has size exactly  $i$  (respectively, has size at most  $n$ ) is then  $\sum_{r=1}^{\lfloor i/2 \rfloor} N_{r,i-r}$  (respectively,  $\sum_{i=2}^n \sum_{r=1}^{\lfloor i/2 \rfloor} N_{r,i-r}$ ).

In principle, this provides an explicit (albeit complicated) expression for the total number of reactions; here, we just report the asymptotic behaviour as  $n$  grows. Since the number of cleavage reactions is equal to the number of ligation reactions in the non-oriented model, it suffices to consider just cleavage reactions. We have the following result: if we compare the non-oriented case to the oriented, the number of reactions is (asymptotically) double and the ratio of reactions to molecule types is increased by a factor of 4.

**Proposition 1.** *Let  $\overline{R}_n$  denote the set of ligation reactions involving non-oriented polymers up to size  $n$ . As  $n$  grows and with  $* = A, B$ , we have:  $|\overline{R}_n| \sim 2|R_n^*|$  and  $|\overline{R}_n|/|\overline{X}_n| \sim 4|R_n^*|/|X_n|$ .*

*Proof:* We have  $N_{r,i-r} = 4k^i - O(k^{i/2})$ , from which it follows that:

$$\sum_{i=2}^n \sum_{r=1}^{\lfloor i/2 \rfloor} N_{r,i-r} \sim \frac{2nk^{n+1}}{k-1}.$$

The second result then follows from Lemma 3.

### 3.3. Consequences for the emergence of RAFs in polymer systems

We now turn to the relevance of the enumeration of the various (related) classes of polymers and ligation-cleavage reactions to the degree of catalysis required for the emergence of RAFs. Recall that a CRS  $\mathcal{Q} = (X, R, C, F)$  consists not only of a molecule set  $X$  and the set of reactions  $R$ , but also an assignment of catalysis  $C$  and a food set  $F$ . To emphasise that  $\mathcal{Q}$  depends on the maximal polymer length  $n$ , we will often write this as  $\mathcal{Q}_n$ . In polymer models,  $F$  is typically taken to be all words of length at most  $t$  (where  $t$  is typically small and independent of  $n$  (e.g.  $t = 2$ )). As for catalysis, this is assigned randomly, and various models have been proposed. The simplest (dating back to Kauffman [13, 14]), assumes that each molecule catalyses each reaction with a fixed probability  $p = p_n$  (possibly dependent on  $n$ ) and that such events are independent across all pairs  $(x, r) \in X \times R$ .

Thus each molecule type catalyses an expected number of  $\mu_n = p_n |R|$  reactions. Notice that if  $p_n$  is independent of  $n$ , then  $\mu_n$  grows exponentially with  $n$ ; however, it turns out that  $\mu_n$  needs only to grow linearly with  $n$  for RAFs to arise with high probability. Moreover, there is a sharp transition here in the following sense (which following from results in [19]). In a polymer model with oriented polymers and either under Convention A or B above, we have (for any  $\epsilon > 0$ ):

$$\begin{aligned} \mu_n = n^{1-\epsilon} &\implies \lim_{n \rightarrow \infty} \mathbb{P}(\text{there exists a RAF for } \mathcal{Q}_n) = 0; \\ \mu_n = n^{1+\epsilon} &\implies \lim_{n \rightarrow \infty} \mathbb{P}(\text{there exists a RAF for } \mathcal{Q}_n) = 1. \end{aligned}$$

This linear transition has been observed in numerous simulation studies (first in [8]). A more fine-grained analysis (also from [19]) shows that if we write  $\mu = \lambda n$  then, for any fixed  $n$ :

$$\mathbb{P}(\text{there exists a RAF for } \mathcal{Q}_n) \rightarrow \begin{cases} 0, & \text{as } \lambda \rightarrow 0, \\ 1, & \text{as } \lambda \text{ grows.} \end{cases} \quad (9)$$

The linear dependence of catalysis rate on  $n$  in the transition from having no RAF to having a RAF in the oriented polymer setting is essentially because  $n$  is asymptotic to the ratio of the number of reactions divided by the number of molecule types (i.e.  $|R_n|/|X_n|$ ). For more general ‘polymer-like’ systems (including non-oriented polymers), there is an analogue of Eqn.(9) in Theorem 1 of [21], where again the ratio of reactions to molecule types plays a key role. This is the main reason why it is important to have an asymptotic measure of the size of these sets (as provided in Proposition 1).

## 4. Complexity results for closed RAFs

For the rest of the paper, we no longer restrict ourselves to the world of polymer CRS systems; instead, our results apply to any CRS.

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  $\text{cl}_{R'}(F)$ . In other words, a closed RAF is one in which every reaction that *can* happen *will* happen. For example, in Fig. 3, the

RAFTs  $\{r_3\}$ ,  $\{r_1, r_2\}$  and  $\{r_1, r_2, r_3, r_4, r_5\}$  (the maxRAF) are the only closed RAFTs present. Note that the maxRAF is always closed, if it exists.

Closed RAFTs are of particular relevance to evolutionary theories of RAFTs and early metabolic cycles: by finding the closed subRAFTs of a particular RAFT, 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 RAFT (see [11], [26])). In a recent paper [11], a direct, formal relationship between closed RAFTs and a field known as Chemical Organisation Theory was established (adding to earlier links (see [25])). Using this connection, a new type of algorithm was developed to enumerate the set of all closed subRAFTs existing within the maxRAF [11]. Although the new algorithm had reasonable performance for RAFTs of size  $\leq 200$ , it was not shown to run in polynomial time [11]. It has remained an open question (posed in [11] and [26]) as to whether such a polynomial-time algorithm exists.

In this section, we present our complexity results surrounding closed RAFTs. We solve the open problem posed in [11, 26] by demonstrating the NP-completeness of finding closed subRAFT. We also we show that finding closed ‘uninhibited’ RAFTs is not fixed-parameter tractable (FPT) in the number of inhibitions (unlike the case for non-closed RAFTs [9]).

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 RAFT, 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 RAFT  $R'$  that does not produce  $\hat{x}$ ?
- (ii) a RAFT that is a strict subset of  $\text{maxRAF}(\mathcal{Q})$ ?

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

For Problem (ii), let  $R_r^* = \text{maxRAF}(\mathcal{Q}) - \{r\}$ . The the answer to (ii) is ‘yes’ if and only if the CRS  $(X, R_r^*, C|_{X \times R_r^*}, F)$  has a RAFT for some  $r \in \text{maxRAF}(\mathcal{Q})$ .

However, if we modify these two questions so as to require the desired RAFT 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 RAFT  $R'$  that does not produce a certain molecule type  $\hat{x} \in X$ ? More formally stated:

**PROBLEM:** *Forbidden-molecule closed RAFT*

**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 RAFT  $R' \subseteq R$  with  $\hat{x} \notin \text{cl}_{R'}(F)$ ?

**Theorem 4.1.** *The Forbidden-molecule closed RAFT problem is NP-complete.*

The proof Theorem 4.1 involves a delicate reduction from 3SAT (3-satisfiability), and is presented in the Appendix.

Theorem 4.1 provides the tool for establishing the hardness of some questions posed in earlier papers [11, 26].

**PROBLEM:** *Closed strict subRAF*

**INSTANCE:** A CRS  $\mathcal{Q} = (X, R, C, F)$

**QUESTION:** Does  $\mathcal{Q}$  contain a closed RAF  $R' \subseteq R$  with  $R' \subsetneq \max\text{RAF}(\mathcal{Q})$ ?

**Theorem 4.2.** *The closed strict subRAF problem is NP-complete.*

*Proof:* Given a set of reactions  $R' \subseteq R$ , one can check whether or not it is both closed and is a strict subset of  $\max\text{RAF}(\mathcal{Q})$  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-complete by Theorem 4.1). 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  $R'' \subseteq \hat{R}$  that is a strict subset of the  $\max\text{RAF}$  in  $\hat{\mathcal{Q}}$ . The construction of  $\hat{\mathcal{Q}}$  from  $\mathcal{Q}$  is as follows. Let  $\hat{\mathcal{Q}} = (\hat{X}, \hat{R}, \hat{C}, \hat{F})$  where  $\hat{X} = X \cup \{\hat{f}\}$ ,  $\hat{R} = R \cup \{\{\hat{f}\} \rightarrow \{\hat{x}\}\}$ ,  $\hat{C} = C \cup \{(x, r) : r \in \hat{R}\}$  and  $\hat{F} = F \cup \{\hat{f}\}$ , for some newly introduced food molecule type  $\hat{f}$ . Essentially, we add a new food molecule type  $\hat{f}$  to  $\hat{\mathcal{Q}}$  and a reaction  $\{\hat{f}\} \rightarrow \{\hat{x}\}$  that produces the forbidden molecule  $\hat{x}$ ; we then let  $\hat{x}$  catalyse every reaction in  $\hat{\mathcal{Q}}$ . This construction is clearly polynomial-time computable in the size of  $\mathcal{Q}$ . We now note that since  $X \subseteq \hat{X}$ ,  $R \subseteq \hat{R}$ ,  $C \subseteq \hat{C}$  and  $F \subseteq \hat{F}$ , we have  $\max\text{RAF}(\mathcal{Q}) \subseteq \max\text{RAF}(\hat{\mathcal{Q}})$ ; furthermore, since  $\hat{\mathcal{Q}}$  contains the additional reaction  $\{\hat{f}\} \rightarrow \{\hat{x}\}$ , we have  $\max\text{RAF}(\mathcal{Q}) \subset \max\text{RAF}(\hat{\mathcal{Q}})$ .

We now establish the correctness of the construction.

( $\implies$ ) Suppose that  $\mathcal{Q}$  has a closed RAF  $R' \subseteq R$  with  $\hat{x} \notin \text{cl}_{R'}(F)$ . Since  $R'$  is a RAF in  $\mathcal{Q}$ , we have  $R' \subseteq \max\text{RAF}(\mathcal{Q})$ . Now, as  $\max\text{RAF}(\mathcal{Q}) \subset \max\text{RAF}(\hat{\mathcal{Q}})$ , we have  $R' \subset \max\text{RAF}(\hat{\mathcal{Q}})$ . In other words,  $R'$  is a strict subset of the  $\max\text{RAF}$  in  $\hat{\mathcal{Q}}$ . Furthermore, since  $X \subseteq \hat{X}$ ,  $R \subseteq \hat{R}$ ,  $C \subseteq \hat{C}$  and  $F \subseteq \hat{F}$ , it follows that  $R'$  is also a RAF in  $\hat{\mathcal{Q}}$ . We now complete the implication by showing that  $R'$  is closed in  $\hat{\mathcal{Q}}$ . To do this, we first note that  $\text{cl}_{R'}(F) = \text{cl}_{R'}(\hat{F})$  across  $\mathcal{Q}$  and  $\hat{\mathcal{Q}}$ , respectively, for any  $R' \subseteq R$ . This follows as for any  $R' \subseteq R$  we have  $\{\hat{f}\} \rightarrow \{\hat{x}\} \notin R'$ . Now suppose that  $r \in \hat{R} - R'$ . There are two cases to consider: (i):  $r \in R - R'$  and (ii)  $r = \{\hat{f}\} \rightarrow \{\hat{x}\}$ . In Case (i),  $\text{cl}_{R'}(F)$  is closed in  $\mathcal{Q}$ , and as  $\text{cl}_{R'}(F) = \text{cl}_{R'}(\hat{F})$ , we have  $R'$  closed in  $\hat{\mathcal{Q}}$ . In Case (ii), since  $R' \subseteq R$ , we have  $\{\hat{f}\} \rightarrow \{\hat{x}\} \notin R'$ . With this, and since  $\hat{x} \notin \text{cl}_{R'}(F)$  in  $\mathcal{Q}$ , it follows  $\hat{x} \notin \text{cl}_{R'}(\hat{F})$  in  $\hat{\mathcal{Q}}$  and so  $r = \{\hat{f}\} \rightarrow \{\hat{x}\}$  is uncatalysed across  $\text{cl}_{R'}(\hat{F})$ . Therefore  $R'$  is closed in  $\hat{\mathcal{Q}}$ , and thus  $R'$  is a closed, strict subRAF of the  $\max\text{RAF}$  of  $\hat{\mathcal{Q}}$ .

( $\impliedby$ ) Suppose that  $R'' \subseteq \hat{R}$  is a closed RAF in  $\hat{\mathcal{Q}}$  that is a strict subRAF of  $\max\text{RAF}(\hat{\mathcal{Q}})$ . Since  $R'' \subset \max\text{RAF}(\hat{\mathcal{Q}})$ , 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  $\max\text{RAF}(\hat{\mathcal{Q}})$ ). With  $\hat{x} \notin \text{cl}_{R''}(\hat{F})$ , we have  $\{\hat{f}\} \rightarrow \{\hat{x}\} \notin R''$ , and therefore  $R'' \subseteq R$  in  $\mathcal{Q}$ . Furthermore, as  $\text{cl}_{R'}(F) = \text{cl}_{R'}(\hat{F})$  across  $\mathcal{Q}$  and  $\hat{\mathcal{Q}}$  for any  $R' \subseteq R$  (as mentioned previously), and by construction,  $R''$  must also be a RAF in  $\mathcal{Q}$ . Since  $X \subseteq \hat{X}$ ,  $R \subseteq \hat{R}$ ,  $C \subseteq \hat{C}$  and  $F \subseteq \hat{F}$  (and as  $R''$  is closed in  $\hat{\mathcal{Q}}$ ),  $R''$  must also be closed in  $\mathcal{Q}$  with  $\hat{x} \notin \text{cl}_{R''}(F)$ . This completes the argument. □

#### 4.1. 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  $\mathcal{Q} = (X, R, C, F)$  together with an additional set  $I \subseteq X \times R$ . An *uninhibited RAF* (uRAF) is a RAF set  $R'$  for  $\mathcal{Q}$  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  $x \in \text{cl}_{R'}(F)$  for which  $(x, r) \in I$  [7]. Essentially, inhibition can be viewed as the exact opposite of catalysation. Note that a reaction that is inhibited can still be part of a RAF, providing that the only molecule types that inhibit the reaction are not present in  $\text{cl}_{R'}(F)$ .

Following [21]), it seems sensible to further restrict interest to uRAFs that are also closed RAFs, 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 that inhibits a reaction within the original RAF. Closed uRAFs, on the other hand, are truly free of inhibition from products of such reactions. It was shown in [19] that determining the existence of a (non-closed) uRAF within a CRS  $\mathcal{Q}$  is NP-hard, though it was shown in [9] to be fixed-parameter-tractable in the number of inhibiting molecule types.

For non-closed RAFs, an algorithm exists for finding uRAFs that is fixed-parameter-tractable in the number of inhibiting molecule types [9]. We prove the same result cannot be found for closed uRAFs (subject to  $P \neq NP$ ) by showing that the following question is NP-complete.

**Theorem 4.3.** *The following problem is NP-complete. Given a CRS  $\mathcal{Q} = (X, R, C, F)$  and a single inhibiting pair  $(x, r) \in X \times R$ , determine whether or not a closed uRAF for  $(\mathcal{Q}, I)$  exists.*

*Proof:* Checking that a set  $R'$  is a closed uRAF is decidable in polynomial-time, so the problem is in the class NP. To establish NP-completeness, we again perform a polynomial-time reduction from the Forbidden-molecule closed RAF problem (*c.f.* Theorem 4.1). Given a CRS  $\mathcal{Q} = (X, R, C, F)$  and a particular molecule type  $\hat{x} \in X$ , we construct, in polynomial time, a CRS  $\mathcal{Q}'$  and an inhibiting set  $I \in X \times R$  such that  $\mathcal{Q}$  has a closed RAF  $R' \subseteq R$  with  $\hat{x} \notin \text{cl}_{R'}(F)$  if and only if  $\mathcal{Q}'$  has a closed, uninhibited RAF. The construction is simple: let  $\mathcal{Q}' = \mathcal{Q}$  and define  $I = \{(x, r) : r \in R\}$  (*i.e.* make  $x$  inhibit all reactions in  $R$ ). It follows that a closed RAF  $R'$  in  $\mathcal{Q}$  with  $\hat{x} \notin \text{cl}_{R'}(F)$  will be a closed, uninhibited RAF in  $\mathcal{Q}'$ . Conversely, a closed uRAF  $R'$  for  $\mathcal{Q}'$  is a closed RAF for which  $\hat{x} \notin \text{cl}_{R'}(F)$ . Moreover, the construction is clearly polynomial-time computable, as required.  $\square$

## 5. Complexity results for elementary RAFs

To help avoid NP-complete problems arising in RAF theory, a simpler setting has recently been studied (see [26]). 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 ([27], [1]) as well as theoretical models ([12]). 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 4) are both computable in polynomial-time in

the elementary setting [26]. We now present two questions concerning elementary CRSs that were posed recently in [26]:

**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 reaction  $r \in R'$ , an inhibiting molecule type for  $r$  is not present in either the food set or as a product of another reaction in  $R'$ . [7].

**Theorem 5.1.** *Given an elementary CRS  $\mathcal{Q} = (X, R, C, F)$  and an inhibition assignment  $I \subseteq X \times R$ , determining whether or not a uRAF for  $\mathcal{Q}$  exists is NP-complete.*

The proof of this result is provided in the Appendix.

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. For example, in Fig. 3, the sets  $\{r_3\}$  and  $\{r_1, r_2\}$  are the only irrRAFs present. Irreducible RAFs have been extensively used in structural analyses of RAFs in polymer models [25] and in extant metabolic systems, [22] and to model ‘coherent evolutionary units’, an evolutionary analogue of cells (see [28]). Finding a single irrRAF within a RAF is easily shown to be computable in polynomial-time. However, finding a smallest irrRAF (or, equivalently, a minimum-size RAF) turns out to be NP-hard [25].

It has been shown earlier [25] 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 [22]. 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 [26].

**Theorem 5.2.** *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 [15]), which is known to be NP-hard [5]. For a digraph  $G = (V, E)$ , a simple chordless cycle is a sequence of distinct vertices  $v_0, v_1, \dots, v_{k-1}$  with  $(v_i, v_{(i+1) \bmod k}) \in E$  for each  $i < k$ , 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. 5 for an illustration.

Given a digraph  $G = (V, E)$ , we first describe a simple polynomial-time construction of a certain elementary CRS  $\mathcal{Q}_G = (X, R, C, F)$  as follows: for each  $v \in V$ , let  $v^R \in R$  be a reaction in  $\mathcal{Q}_G$ ; for each  $(u, v) \in E$ , let  $u_v^X \in X$  be a molecule produced by  $u^R$  that catalyses the reaction  $v^R$  in  $\mathcal{Q}_G$ ; let  $F = \{f\}$  with  $f$  being the sole reactant of all reactions. Since we introduce a linear number of reactions, molecule

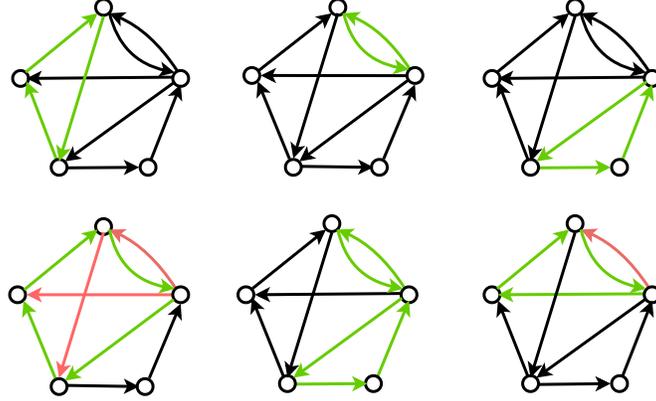


Figure 5: The top three cycles (paths in green) are simple chordless cycles. The bottom three cycles (chords highlighted in red) are not simple chordless cycles (the bottom-middle one is not simple; for the other two, chords are highlighted in red).

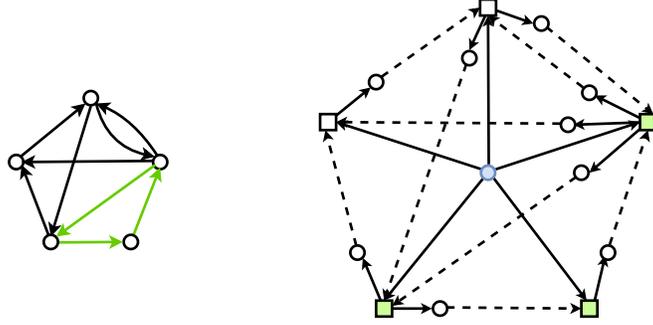


Figure 6: A digraph  $G$  displayed on the left, with the constructed elementary CRS  $\mathcal{Q}_G$  on the right (the food molecule is at the centre, highlighted in blue). The simple chordless cycle in  $G$  of size 3 (highlighted in green) corresponds to an irrRAF in  $\mathcal{Q}_G$  of size 3.

types and edges with respect to  $|G|$ , our construction is computable in polynomial-time. See Fig. 6 for an illustration.

Since every reaction is trivially F-generated, and since  $f$  does not catalyse any reactions, we can simplify the RAF definition in this setting: a set  $R' \subseteq R$  will be a RAF set in  $\mathcal{Q}_G$  if and only if each reaction  $r \in R'$  is catalysed by a molecule which is produced by another reaction  $r' \in R'$ . It follows that a set of reactions  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  is a RAF in  $\mathcal{Q}_G$  if and only if for each  $v_i^R \in R'$ , there is a reaction  $v_j^R \in R'$  that produces the molecule  $v_j^X \in \text{cl}_{R'}(F)$  with  $(v_j^X, v_i^R) \in C$ . By construction, this happens if and only if  $(v_j, v_i) \in E$ .

Furthermore, let us now consider a set  $V' = \{v_0, v_1, \dots, v_{k-1}\} \subseteq V$  of distinct vertices in  $G$  such that for each  $v_i \in V'$  there exists a  $v_j \in V'$  with  $\{v_j, v_i\}$  in  $E$ . By the construction of  $\mathcal{Q}_G$ , it follows that  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  will have every reaction  $v_i^R \in R'$  catalysed by a molecule  $v_j^X \in \text{cl}_{R'}(F)$  produced by  $v_j^R \in R'$ , and therefore  $R'$  is a RAF set.

Combining these together,  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  is a RAF in  $\mathcal{Q}_G$  if and only if for every vertex

$v_i \in V' = \{v_0, v_1, \dots, v_{k-1}\} \subseteq V$ , there exists another vertex  $v_j \in V'$  with  $(v_j, v_i) \in E$ . Formally, fix a set  $V' = \{v_0, v_1, \dots, v_{k-1}\} \subseteq V$ , then:

$$\forall v_i \in V', \exists v_j \in V' : (v_j, v_i) \in E \iff \{v_0^R, v_1^R, \dots, v_{k-1}^R\} \subseteq R \text{ is a RAF in } \mathcal{Q}_G. \quad (10)$$

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

**Lemma 4.** *Fix a digraph  $G = (V, E)$  and a number  $k \geq 1$ . Then  $v_0, \dots, v_{k-1}$  is a simple chordless cycle in  $G$  if and only if  $\{v_0^R, \dots, v_{k-1}^R\}$  is an irrRAF in  $\mathcal{Q}_G$ .*

*Proof:*

( $\implies$ ) Suppose that  $v_0, \dots, v_{k-1}$  is a simple chordless cycle in  $G$  for  $k \geq 2$ . Since  $v_0, \dots, v_{k-1}$  is simple, we can define a set  $V' = \{v_0, \dots, v_{k-1}\} \subseteq V$  (simple implying no deduplication when translating to a set). As  $v_0, \dots, v_{k-1}$  is a cycle, the LHS of (10) is true, so  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\} \subseteq R$  is a RAF in  $\mathcal{Q}_G$ . We now show that  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  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  $v_{(i-1) \bmod k}$ . By construction, we therefore see that for each reaction  $v_i^R \in R'$  we have  $v_{(i-1) \bmod k}^R$  the *only* reaction in  $R'$  producing a molecule that catalyses  $v_i^R \in R'$ . It follows that no strict subset of reactions  $R'' \subset R'$  is a RAF; otherwise, for some  $0 \leq j < k$  we would have  $v_{(j-1) \bmod k}^R \notin R''$  with  $v_j^R \in R''$  and  $v_j^R$  not being catalysed. Therefore  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  is an irrRAF in  $\mathcal{Q}_G$ .

( $\impliedby$ ) Suppose that  $R' = \{v_0^R, v_1^R, \dots, v_{k-1}^R\}$  is an irrRAF in  $\mathcal{Q}_G$ . Let  $V' = \{v_0, v_1, \dots, v_{k-1}\} \subseteq V$  be a distinct set of vertices from  $G$ . As  $R'$  is a RAF by (10) it follows that for every vertex  $v_i \in V'$  there exists another other vertex  $v_j \in V'$  with  $(v_j, v_i) \in E$  (i.e.  $\forall v_i \in V', \exists v_j \in V' : (v_j, v_i) \in E$ ). If we now consider any set  $V'' \subset V'$ , the same cannot apply else by (10) we would see  $V''$  corresponding to a subRAF  $R'' \subseteq R'$ . As  $R'$  is irreducible, this is a contradiction. It follows that for every vertex  $v_i \in V'$ , we have  $(v_j, v_i) \in E$  for exactly one vertex  $v_j \in V'$ . The set  $V'$  must form a simple chordless cycle.  $\square$

*Proof of Theorem 5.2:* We apply Lemma 4 to the complexity questions surrounding RAF irreducibility. Recall that this lemma states that if we fix a digraph  $G = (V, E)$  and a number  $k \geq 2$ , then  $v_0, \dots, v_{k-1}$  is a simple chordless cycle in  $G$  if and only if  $\{v_0^R, \dots, v_{k-1}^R\}$  is an irrRAF in  $\mathcal{Q}_G$ . Finding the longest simple chordless cycle is a known NP-hard problem [5], so the proof follows immediately from Lemma 4.  $\square$

## 6. Concluding comments

In this paper, we have provided the first exact formula for the number of reactions in polymer models, and established asymptotic results from them that verify earlier results that were based on heuristic approximations. We have also settled a number of computational complexity questions concerning the detection of subRAFs of particular types. As many of the interesting questions 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 these in future work.

## 7. References

- [1] G. ASHKENASY, R. JEGASIA, M. YADAV, AND M. GHADIRI, *Design of a directed molecular network*, Proc. Natl Acad. Sci. USA, 101 (2004), pp. 10872–10877.
- [2] R. CAZZOLLA GATTI, B. FATH, W. HORDIJK, S. KAUFFMAN, AND R. ULANOWICZ, *Niche emergence as an autocatalytic process in the evolution of ecosystems*, J. Theor. Biol., 454 (2018), pp. 110–117.
- [3] P. ERDÖS AND A. RÉNYI, *On the evolution of random graphs*, in Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, pp. 17–61.
- [4] L. GABORA AND M. STEEL, *Autocatalytic networks in cognition and the origin of culture*, J. Theor. Biol., 431 (2017), pp. 87–95.
- [5] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman Co., New York, NY, USA, 1990.
- [6] W. HORDIJK, J. HEIN, AND M. STEEL, *Autocatalytic sets and the origin of life*, Entropy, 12 (2010), pp. 1733–1742.
- [7] W. HORDIJK, J. I. SMITH, AND M. STEEL, *Algorithms for detecting and analysing autocatalytic sets*, Algorithms for Molecular Biology, 10 (2015), p. 15.
- [8] W. HORDIJK AND M. STEEL, *Detecting autocatalytic, self-sustaining sets in chemical reaction systems*, J. Theor. Biol., 227 (2004), pp. 451–461.
- [9] W. HORDIJK AND M. STEEL, *Autocatalytic sets extended: Dynamics, inhibition, and a generalization*, Journal of Systems Chemistry, 3 (2012), p. 5.
- [10] W. HORDIJK AND M. STEEL, *Chasing the tail: The emergence of autocatalytic networks*, Biosystems, 152 (2017), pp. 1–10.
- [11] W. HORDIJK, M. STEEL, AND P. DITTRICH, *Autocatalytic sets and chemical organizations: Modeling self-sustaining reaction networks at the origin of life*, New J. Phys., 20 (2018), p. 015011.
- [12] S. JAIN AND S. KRISHNA, *Autocatalytic sets and the growth of complexity in an evolutionary model*, Phys. Rev. Lett., 81 (1998), pp. 5684–5687.
- [13] S. A. KAUFFMAN, *Autocatalytic sets of proteins*, J. Theor. Biol., 119 (1986), pp. 1–24.
- [14] S. A. KAUFFMAN, *The Origins of Order*, Oxford University Press, New York, NY, USA, 1993.
- [15] W. H. KAUTZ, *Unit-distance error-checking codes*, IRE Transactions on Electronic Computers, EC-7 (1958).
- [16] D. LOUTCHKO, *Semigroup models for biochemical reaction networks*. arXiv:1908.04642, 2019.

- [17] R. C. LYNDON AND M. P. SCHÜTZENBERGER, *The equation  $a^m = b^n + c^p$  in a free group*, Michigan Math. J., 9 (1962), pp. 289–298.
- [18] W. F. MARTIN AND M. PREINER, *Origins of life, theories of*, in Reference Module in Life Sciences, Elsevier doi:10.1016/B978-0-12-809633-8.02403-1, 2017.
- [19] E. MOSSEL AND M. STEEL, *Random biochemical networks: the probability of self-sustaining autocatalysis*, J. Theor. Biol., 233 (2005), pp. 327–336.
- [20] M. PREINER, J. C. XAVIER, A. D. N. VIEIRA, K. KLEINERMANN, J. F. ALLEN, AND W. F. MARTIN, *Catalysis, autocatalysis and the origin of metabolism*, Interface Focus, 9 (2019), p. 20190072.
- [21] J. SMITH, M. STEEL, AND W. HORDIJK, *Autocatalytic sets in a partitioned biochemical network*, J. Syst. Chem., 5 (2014), p. 2.
- [22] F. L. SOUSA, W. HORDIJK, M. STEEL, AND W. F. MARTIN, *Autocatalytic sets in E. coli metabolism*, J. Syst. Chem., 6 (2015), p. 4.
- [23] R. P. STANLEY, *Enumerative combinatorics, Vol. 1*, Cambridge University Press, Cambridge, UK, 1995.
- [24] M. STEEL, *The emergence of a self-catalysing structure in abstract origin-of-life models*, Appl. Math. Lett., 13 (2000), pp. 91–95.
- [25] M. STEEL, W. HORDIJK, AND J. SMITH, *Minimal autocatalytic networks*, J. Theor. Biol., 332 (2013), pp. 96–107.
- [26] M. STEEL, W. HORDIJK, AND J. C. XAVIER, *Autocatalytic networks in biology: structural theory and algorithms*, J. Royal Society Interface, 16 (2019), p. 20180808.
- [27] N. VAIDYA, M. L. MANAPAT, C. I. A., R. XULVI-BRUNET, AND N. HAYDEN, E. J. LEHMAN, *Spontaneous network formation among cooperative RNA replicators*, Nature, 491 (2012), pp. 72–77.
- [28] V. VASAS, C. FERNANDO, M. SANTOS, S. KAUFFMAN, AND E. SZATHMÁRY, *Evolution before genes*, Biology Direct, 7 (2012), p. 1.
- [29] R. WOLFENDEN AND M. J. SNIDER, *The depth of chemical time and the power of enzymes as catalysts*, Acc. Chem. Res., 34 (2001), pp. 938–945.
- [30] J. C. XAVIER, W. HORDIJK, S. KAUFFMAN, S. M., AND W. F. MARTIN, *Autocatalytic chemical networks preceded proteins and RNA in evolution*, bioRxiv., <https://doi.org/10.1101/693879> (2019).

## 8. Appendix: Proofs of Theorems 3.1(i), 4.1 and 5.1.

### Proof of Theorem 3.1 Part (i):

*Part (i):* The proof consists of establishing a series of claims.

**Claim 1:** Using Convention B to count  $U(i)$  results in double-counts precisely whenever  $uv = vu = z, u \neq v$  and  $|z| = i$ .

To see this, observe that for  $u, v, z \in \Sigma^+$  with  $uv = z$ , we double-count using Convention B (instead of Convention A) precisely when there are two distinct triples  $(u, v, z)$  and  $(u', v', z')$  with  $uv = u'v' = z$  and the corresponding sets  $\{u, v, z\} = \{u', v', z\}$  equivalent. Since  $z$  is common to both sets, we have  $\{u, v, z\} = \{u', v', z\}$  if and only if  $\{u, v\} = \{u', v'\}$ ; however, since  $(u, v, z)$  and  $(u', v', z)$  are distinct, we also have  $u \neq u'$  or  $v \neq v'$ . This gives us  $\{u, v, z\} = \{u', v', z\}$  if and only if  $u = v'$  and  $v = u'$ . Rearranging gives  $uv = vu = z$ . This establishes Claim 1.

**Claim 2:** For each fixed  $i \geq 2$  there is a bijection between the set of periodic words  $w_i$  of length  $i$  and the set of aperiodic sub-words  $w_d$  of length  $d$  where  $d|i$  and  $d < i$ .

To see this, observe that every periodic word  $w_i$  must have a unique period, as the period is the *smallest* sub-word that repeatedly joins to form the word; further, that period is an aperiodic sub-word of length  $d < i$  with  $d|i$ . Next, fix an aperiodic word  $w_d$  of length  $d$  with  $d|i, d < i$ ; then  $w_d^{i/d} = w_i$ , so  $w_d$  uniquely corresponds (as a sub-word) to a periodic word  $w_i$  of length  $i$  with  $d|i, d < i$ . This establishes Claim 2.

With this bijection, we can easily derive a recursive formula to count the number of aperiodic words of length  $i$ .

**Claim 3:** The number of aperiodic words  $\bar{p}(i)$  of length  $i \geq 2$  is given exactly by:

$$\bar{p}(i) = k^i - \sum_{d|i, d < i} \bar{p}(d) \quad (11)$$

To see this, observe that the number of aperiodic words of length  $i$  is clearly  $k^i$  subtract the number of periodic words of length  $i$ . By Claim 2, we can count the number of periodic words of length  $i$  by counting the number of aperiodic words of length  $d < i, d|i$ . This gives Eqn.(11), and so establishes Claim 3.

We make a small point about the base case: technically,  $\bar{p}(1)$  is not defined above (a single character is aperiodic) though we still use the value of  $\bar{p}(1) = k$  as the base-case for  $i \geq 2$ . With this, we now achieve a non-recursive formula. Rearranging (11), we get:

$$k^i = \sum_{d|i} \bar{p}(d) \quad (12)$$

By applying the Möbius inversion formula we get:

$$\bar{p}(i) = \sum_{d|i} \mu\left(\frac{i}{d}\right) k^d \quad (13)$$

Recalling our original problem, where we double-count precisely whenever  $uv = vu = z$ ,  $u \neq v$  and  $|z| = i$  for any  $u, v, z \in \Sigma^+$ , it follows from Lemma 2 that  $z$  must therefore be periodic.

**Claim 4:** Using Convention B (rather than A), the number of double-counts per periodic word of length  $i$  with period length  $d$  is given exactly by  $\left\lfloor \frac{i/d-1}{2} \right\rfloor$ .

To see this, let  $w_i$  be a periodic word of length  $i$  and  $w_d$  its period of length  $d$ . It follows that  $w_d^j w_d^{i/d-j} = w_i$  for each  $1 \leq j < i/d$ . Using Convention B, we will double-count precisely whenever  $j < i/d - j$ , as  $j$  is symmetric for  $i/d - j$ . For an example, let  $w_i = abcabcabcabc$ . Now,  $w_i$  has a period  $abc$  and there are 3 ways to partition  $w_i$  with units  $abc$  using Convention B:

- (1)  $\{abc, abcabcabc, w_i\}$
- (2)  $\{abcabc, abcabc, w_i\}$
- (3)  $\{abcabcabc, abc, w_i\}$

However, (1) and (3) are the same sets — and therefore double-counted in the unordered version. Combining  $j < i/d - j$  with  $j \geq 1$  gives us  $\left\lfloor \frac{i/d-1}{2} \right\rfloor$  double-counts. This establishes Claim 4.

Next, recall that  $U(i)$  denote the number of unique sets  $\{u, v, z\}$  for  $u, v, z \in \Sigma^+$  such that  $uv = z$  and  $|z| = i$ .

**Claim 5:**  $U(i)$  is given by the expression in Eqn. (7).

To establish Claim 5 we use Convention B (which leads to Eqn. (4)) to give  $(i-1)k^i$  and then subtract the double-counts to give our desired Convention A. From Claim 4, each period word of length  $i$  and period  $d$  will incur exactly  $\left\lfloor \frac{i/d-1}{2} \right\rfloor$  double-counts; further, from Lemma 2 we double-count exclusively when this is the case. This gives us the formula:

$$U(i) = (i-1)k^n - \sum_{d|i, d < n} \left\lfloor \frac{i/d-1}{2} \right\rfloor \bar{p}(d)$$

Using the explicit formula for  $\bar{p}(d)$ , we arrive at Eqn. (7). This establishes Claim 5, and thereby Part (i).  $\square$

**Proof of Theorem 4.1:** Our proof will follow a reduction from 3SAT. Given a formula  $\mathcal{F}$ , we will construct in polynomial time a CRS  $\mathcal{Q}_{\mathcal{F}} = (X, R, C, F)$  with a distinguished molecule  $\hat{x} \in X$  that has a closed RAF  $R'$  with  $\hat{x} \notin \text{cl}_{R'}(F)$  if and only if  $\mathcal{F}$  admits a satisfying assignment.

We start by describing the *food gadget*. Our CRS  $\mathcal{Q}_{\mathcal{F}}$  will contain a single food molecule  $F = \{f\}$ . We will include in  $\mathcal{Q}_{\mathcal{F}}$  two distinguished molecule types  $x_{out}$  and  $x_{in}$  as well as two distinguished reactions

$r_{out}$  and  $r_{in}$ . We present the food gadget in Fig. 7(i) (for now we exclude the reactants of  $r_{out}$  and the reactions that  $x_{in}$  acts as a reactant for).

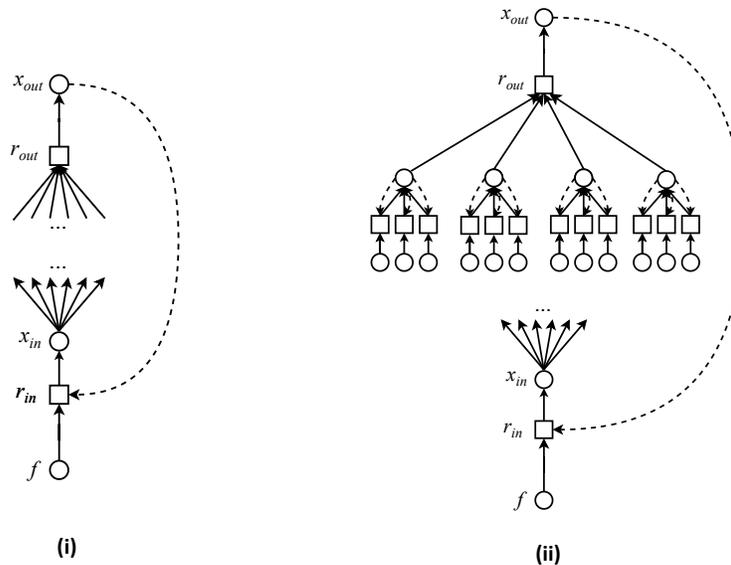


Figure 7: (i) The food gadget (incomplete): both  $r_{out}$  and  $r_{in}$  must necessarily be in any RAF set of  $\mathcal{Q}_{\mathcal{F}}$ . (ii) The clause gadget: at least one reaction from each ‘block’ (of three) in the middle row must be in any RAF set  $R' \subseteq R$ .

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

$$R' \subseteq R \text{ is a RAF} \implies r_{in} \in R' \quad (14)$$

As  $x_{out}$  is the only molecule catalyzing  $r_{in} \in R'$  and  $r_{out}$  is the only reaction producing  $x_{out}$ , we must also have:

$$R' \subseteq R \text{ is a RAF} \implies r_{out} \in R'. \quad (15)$$

Next, we present the *clause gadget*, displayed in Fig. 7(ii). From (15), we can infer that each reactant of  $r_{out}$  must also be in  $\text{cl}_{R'}(F)$ . Let  $A_{out} \subseteq X$  denote the set of reactants of  $r_{out}$ . It follows that if  $R' \subseteq R$  is a RAF then  $\bigwedge_{x \in A_{out}} x \in \text{cl}_{R'}(F)$ .

Further, since none of  $x \in A_{out}$  are food molecule types, we must have at least one reaction producing  $x$ . Let  $\rho : X \rightarrow 2^R$  denote the set of reactions that produce a molecule  $x \in X$ . The following holds:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_{x \in A_{out}} \bigvee_{r \in \rho(x)} r \in R' \quad (16)$$

In Fig. 7(ii), this condition 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 integrate our clause gadgets with the structure of the given formula  $\mathcal{F}$ . For each clause  $c_i = l_{i1} \wedge l_{i2} \wedge l_{i3} \in \mathcal{F} = \bigwedge_i c_i$ , let  $c_i^X$  be a reactant molecule to  $r_{out}$  (i.e. let  $c_i^X \in A_{out}$ ) and let  $l_{i1}^R, l_{i2}^R, l_{i3}^R$  be reactions producing  $c_i^X$  (i.e. let  $l_{i1}^R, l_{i2}^R, l_{i3}^R \in \rho(c_i^X)$ ). For each reaction  $l_{ij}^R$ , also let  $l_{ij}^X$  be its single reactant and let  $c_i^X$  be its single catalysing molecule. Stated formally: for each clause  $c_i \in \mathcal{F}$ , we have  $l_{ij} \in c_i \implies (l_{ij}^R = (\{l_{ij}^X\}, \{c_i^X\}) \in R) \wedge ((c_i^X, l_{ij}^R) \in C)$  for  $1 \leq j \leq 3$ .

We can now rewrite condition (16) as:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_i (l_{i1}^R \in R') \vee (l_{i2}^R \in R') \vee (l_{i3}^R \in R') \quad (17)$$

Since each  $l_{ij}^R$  reaction from 17 must contain its reactants we also gain the condition:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_i (l_{i1}^X \in \text{cl}_{R'}(F)) \vee (l_{i2}^X \in \text{cl}_{R'}(F)) \vee (l_{i3}^X \in \text{cl}_{R'}(F)) \quad (18)$$

As we proceed, will refer to  $l_{ij}^R \in R$  as the ‘literal’ reaction and  $l_{ij}^X$  as the ‘literal’ molecule for the literal occurrence  $l_{ij} \in \mathcal{F}$ .

Next, we describe the *variable gadget*, displayed in Fig. 8. For each variable  $v \in \mathcal{F}$ , introduce a variable

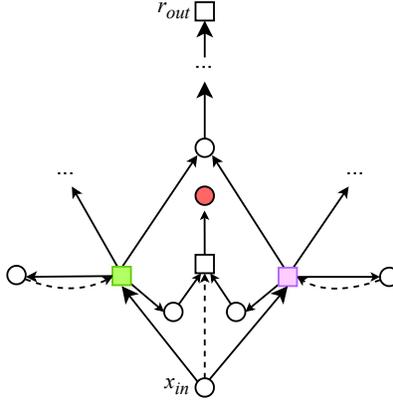


Figure 8: The variable gadget for a variable  $v \in \mathcal{F}$ : in any closed RAF set  $R' \subseteq R$  that does not produce  $\hat{x}$  we have exactly one of  $v^R, \bar{v}^R$  in  $R'$ .

gadget to  $\mathcal{Q}_{\mathcal{F}}$ . The variable gadget is parametrised by a variable  $v$  and consists of three reactions. It is displayed in Fig. 8. For a variable  $v$ , the variable gadget contains reactions named  $v^R$  (highlighted in green),  $\bar{v}^R$  (highlighted in pink) and  $v_{mid}^R$  (non highlighted: produces the red molecule). The central red molecule is  $\hat{x} \in X$ , which is the forbidden molecule. Formally, we have  $v^R = (\{x_{in}\}, \{v_{cat}^X, v_{top}^X, v_{mid}^X\})$ ,  $\bar{v}^R = (\{x_{in}\}, \{\bar{v}_{cat}^X, v_{top}^X, \bar{v}_{mid}^X\})$ ,  $v_{mid}^R = (\{v_{mid}^X, \bar{v}_{mid}^X\}, \{\hat{x}\})$  for molecule types  $v_{top}^X, v_{mid}^X, \bar{v}_{mid}^X, v_{cat}^X, \bar{v}_{cat}^X$  with

$(v_{cat}^X, v^R), (\bar{v}_{cat}^X, \bar{v}^R) \in C$ .

Since  $v_{top}^X$  is a reactant to  $r_{out}$ , by (15) it follows that  $v_{top}^X \in \text{cl}_{R'}(F)$  for any RAF set  $R' \subseteq R$ ; further, as  $v_{top}^X$  is only produced by  $v^R$  and  $\bar{v}^R$ , we can infer the following:

$$R' \subseteq R \text{ is a RAF} \implies (v^R \in R') \vee (\bar{v}^R \in R') \quad (19)$$

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

$$R' \subseteq R \text{ is a closed RAF} \wedge \hat{x} \notin \text{cl}_{R'}(F) \implies (v^R \notin R') \vee (\bar{v}^R \notin R') \quad (20)$$

Combining (20) with (19) gives:

$$R' \subseteq R \text{ is a closed RAF} \wedge \hat{x} \notin \text{cl}_{R'}(F) \implies (v^R \in R') \Leftrightarrow (\bar{v}^R \notin R') \quad (21)$$

We will now consolidate our variable gadgets with our clause gadgets.

For a variable gadget in  $\mathcal{Q}_{\mathcal{F}}$  corresponding to a variable  $v \in \mathcal{F}$ , let  $v^R$  produce all the literal molecule types  $l_{ij}^X$  for  $l_{ij} \in \mathcal{F}$  being a *positive* occurrence of the variable  $v$  — i.e. when  $l_{ij} = v \in c_i$ . Likewise, let  $\bar{v}^R$  produce all literal molecule types  $l_{ij}^X$  for  $l_{ij} \in \mathcal{F}$  being a *negative* occurrence of the variable  $v$  — i.e. when  $l_{ij} = \bar{v} \in c_i$ . Each literal molecule  $l_{ij}^X$  will be produced only by its corresponding variable gadget reaction; that is, either  $\rho(l_{ij}^X) = \{v^R\}$  or  $\rho(l_{ij}^X) = \{\bar{v}^R\}$  for  $l_{ij} = v$  or  $l_{ij} = \bar{v}$ , respectively. Since these are singleton sets, we define  $\rho_l$  to be the ‘unpacking’ of the set — i.e.  $\rho_l(l_{ij}^X) = v^R$  or  $\rho_l(l_{ij}^X) = \bar{v}^R$ . With this, we can rewrite (18) as:

$$R' \subseteq R \text{ is a RAF} \implies \bigwedge_i (\rho_l(l_{i1}^X) \in R') \vee (\rho_l(l_{i2}^X) \in R') \vee (\rho_l(l_{i3}^X) \in R') \quad (22)$$

This enables us to condition the structure of  $\mathcal{F}$  over the variable gadgets in  $\mathcal{Q}_{\mathcal{F}}$  (as opposed to the literal molecule types). See Fig. 9(i) for an illustration of how a single variable gadget fits in amidst the overall construction of  $\mathcal{Q}_{\mathcal{F}}$ , and Fig. 9(ii) for the entire construction.

This completes the construction description. Since the size of each gadget is constant, our construction is linear in the size of the Formula  $\mathcal{F}$  and therefore polynomial-time computable. We now prove the following claim:  $\mathcal{F}$  admits a satisfying assignment if and only if  $\mathcal{Q}_{\mathcal{F}}$  has a closed RAF  $R'$  with  $\hat{x} \notin \text{cl}_{R'}(F)$ .

*Proof.* ( $\implies$ ) Suppose  $\mathcal{Q}_{\mathcal{F}}$  has a closed RAF  $R'$  and  $\hat{x} \notin \text{cl}_{R'}(F)$ . Then we construct an assignment  $\mathcal{A}$  as follows: for each variable  $v \in \mathcal{F}$ , let  $\mathcal{A}(v) = 1$  if and only if  $v^R \in R'$  (set  $\mathcal{A}(v) = 0$  otherwise). By (21),  $\mathcal{A}$  is well-defined. By (22), we have  $\bigwedge_i \mathcal{A}(l_{i1}) \vee \mathcal{A}(l_{i2}) \vee \mathcal{A}(l_{i3})$ . It follows that  $\mathcal{A} \models \mathcal{F}$ .

( $\impliedby$ ) Suppose  $\mathcal{F}$  admits a satisfying assignment  $\mathcal{A}$ . We construct a closed RAF  $R'$  with  $\hat{x} \notin \text{cl}_{R'}(F)$  in four simple steps: (i) Let  $r_{in}, r_{out} \in R'$ . (ii) For each variable  $v \in \mathcal{F}$  with  $\mathcal{A}(v) = 1$ , let  $v^R \in R'$ . (iii) For each variable  $v \in \mathcal{F}$  with  $\mathcal{A}(v) = 0$ , let  $\bar{v}^R \in R'$ . (iv) For each positive-valued literal  $l_{ij} \in \mathcal{F}$  (with respect to  $\mathcal{A}$ ), let  $l_{ij}^R \in R'$ .

We now prove that  $R'$  is a closed RAF that does not produce  $\hat{x}$ . We first show that  $R'$  is F-generated.

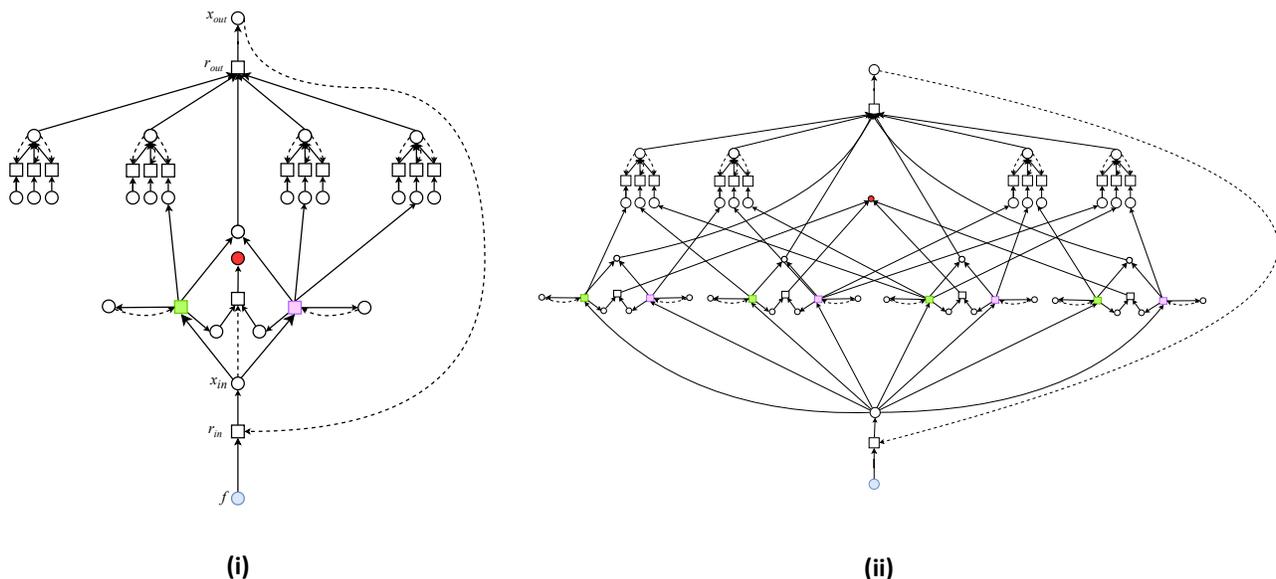


Figure 9: (i) Part of the construction for the Formula  $\mathcal{F} = (a \vee b \vee c) \wedge (a \vee \bar{b} \vee x) \wedge (a \vee \bar{x} \vee c) \wedge (a \vee \bar{x} \vee d)$ , only including the variable gadget for  $x$  (hiding the gadgets for variables  $a, b, c, d$ ). The food molecule is coloured blue. (ii) The entire construction  $\mathcal{Q}_{\mathcal{F}}$  for the formula  $\mathcal{F} = (a \vee b \vee c) \wedge (\bar{a} \vee \bar{b} \vee c) \wedge (\bar{b} \vee \bar{c} \vee d) \wedge (b \vee c \vee \bar{d})$ . The food molecule is coloured blue.

To do this, 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. Besides  $r_{in}$ , each reaction produces a molecule that catalyses it. Since  $r_{in}$  is catalysed by  $x_{out}$ , and  $x_{out}$  is produced by  $r_{out}$ , we have  $r_{in}$  catalysed, so  $R'$  is auto-catalytic. We now prove  $R'$  is closed by case analysis. Suppose  $r \in R - R'$ , then we have the following cases:

*Case (i):*  $r = l_{ij}^R \notin R'$  is a literal reaction. Suppose WLOG that  $l_{ij} \in \mathcal{F}$  is a positive occurrence of the variable  $v \in \mathcal{F}$ . Since  $l_{ij}^R \notin R'$ , by the construction of  $R'$  we must have  $l_{ij}$  negative-valued with respect to  $\mathcal{A}$ , so  $\mathcal{A}(v) = 0$ . With  $\mathcal{A}(v) = 0$  we necessarily have, by construction,  $v^R \notin R'$ . As  $v^R$  is the only reaction producing  $l_{ij}^X$ , we have  $l_{ij}^X \notin \text{cl}_{R'}(F)$ , so  $l_{ij}^R$  does not follow by closure. The argument is symmetric for negative literal occurrences in  $\mathcal{F}$ .

*Case (ii):*  $r = v^R \notin R'$  (or  $r = \bar{v}^R \notin R'$ ) is a variable gadget reaction. By the construction of the variable gadget, its (only) catalysing molecule  $v_{cat}^X$  (or  $\bar{v}_{cat}^X$ ) is not in the closure  $\text{cl}_{R'}(F)$ . So  $v^R$  ( $\bar{v}^R$ ) does not follow by closure.

*Case (iii):*  $r = v_{mid}^R$  (i.e. a variable gadget reaction producing  $\hat{x}$ ). Since  $\mathcal{A}$  is well defined, by construction we cannot have both  $v^R, \bar{v}^R \in R'$ , so the reactants  $v_{mid}^X, \bar{v}_{mid}^X$  of  $v_{mid}^R$  cannot both be in  $\text{cl}_{R'}(F)$ . It follows  $v_{mid}^R$  does not follow by closure. This enumerates all cases for reactions not included in  $R'$ , so  $R'$  is closed.

Finally, since  $v_{mid}^R \notin R'$  for all variables  $v \in \mathcal{F}$ , we clearly have  $\hat{x} \notin \text{cl}_{R'}(F)$ . This completes the proof.

□

**Proof of Theorem 5.1:** Our proof will follow a reduction from 3SAT. Given a 3SAT formula  $\mathcal{F}$  with  $m$  clauses and  $n$  variables, we construct in polynomial-time an elementary CRS  $\mathcal{Q}_{\mathcal{F}}^I$  and an inhibiting set  $I \subseteq X \times R$  that has a uRAF if and only if  $\mathcal{F}$  admits a satisfying assignment.

The construction of  $\mathcal{Q}_{\mathcal{F}}^I = (X, R, C, F)$  from  $\mathcal{F}$  is as follows. First, we will have a single food molecule  $F = \{f\}$ . Next, for each clause  $c_i = l_{(i,1)} \vee l_{(i,2)} \vee l_{(i,3)} \in \mathcal{F}$ , let  $l_{(i,1)}^R, l_{(i,2)}^R, l_{(i,3)}^R \in R$  be a set of ‘literal’ reactions in  $R$  and  $l_{(i,1)}^X, l_{(i,2)}^X, l_{(i,3)}^X \in X$  be a set of ‘literal’ molecule types in  $X$ . For each clause  $c_i \in \mathcal{F}$ , we will have every literal reaction producing the molecule types corresponding to the next clause  $c_{((i+1) \bmod m)}$  (forming a cycle). Formally, each reaction  $l_{(i,j)}^R$  for  $0 \leq i < n$  and  $1 \leq j \leq 3$  will be defined by:

$$l_{(i,j)}^R = (\{f\}, \{l_{((i+1) \bmod m,1)}^X, l_{((i+1) \bmod m,2)}^X, l_{((i+1) \bmod m,3)}^X\}) \text{ with } (l_{(i,j)}^X, l_{(i,j)}^R) \in C.$$

Finally, for every pair of literal reactions  $l_{(i,j)}^R, l_{(s,t)}^R \in R$  which are opposite-parity occurrences of the same underlying variable in  $\mathcal{F}$  (see previous proofs), let  $(l_{(i,j)}^X, l_{(s,t)}^R) \in I$ . See Fig. 10 for an abstract illustration.

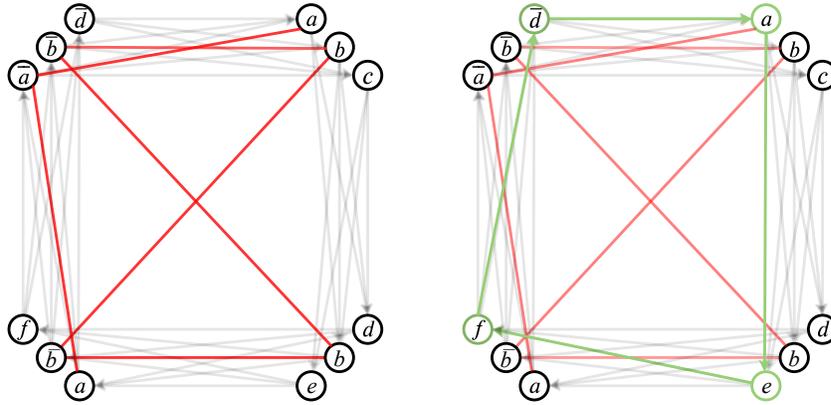


Figure 10: Given the 3SAT formula  $\mathcal{F} = (a \vee b \vee c) \wedge (d \vee b \vee e) \wedge (a \vee \bar{b} \vee f) \wedge (\bar{a} \vee \bar{b} \vee \bar{d})$ , we construct the CRS  $\mathcal{Q}_{\mathcal{F}}^I$ . The diagram above is simplified version of  $\mathcal{Q}_{\mathcal{F}}^I$ ; each node can be thought of as a literal reaction-molecule pair. The red edges denote the inhibiting relations between pairs of literal reactions / molecule types (enforcing a mutual exclusion) and the grey edges denote the catalytic relations between literal reactions in consecutive clauses. The RAF set (highlighted in green) corresponds to the satisfying assignment with variables  $a, e, f$  set true and  $\bar{d}$  false.

We now prove that  $G_{\mathcal{F}}^I$  has a uRAF if and only if  $\mathcal{F}$  admits a satisfying assignment:

( $\implies$ ) Suppose that  $G_{\mathcal{F}}^I$  has a uRAF  $R' \subseteq R$ . We show the existence of a satisfying assignment  $\mathcal{A} \models \mathcal{F}$ . We define  $\mathcal{A}$  to assign variables so that for each literal reaction  $l_{(i,j)}^R \in R'$ , the literal occurrence  $l_{i,j} \in \mathcal{F}$  is positive-valued under  $\mathcal{A}$  (i.e. if  $l_{i,j} = v$  then  $\mathcal{A}(v) = 1$ , and if  $l_{i,j} = \bar{v}$  then  $\mathcal{A}(v) = 0$  for  $l_{(i,j)}^R \in R'$ ). We will let  $\mathcal{A}$  assign arbitrarily for literals  $l_{(i,j)}$  with  $l_{(i,j)}^R \notin R'$ . By construction, since  $R'$  is

a uRAF, there cannot simultaneously exist two  $l_{(i,j)}^R, l_{(s,t)}^R \in R'$  for  $l_{(i,j)}, l_{(s,t)}$  opposite-parity occurrences of the same underlying variable in  $\mathcal{F}$  (else  $l_{(i,j)}^X \in \text{cl}_{R'}(F)$  with  $(l_{(i,j)}^X, l_{(s,t)}^R) \in I$  and  $R'$  would not be uninhibited). Consequently,  $\mathcal{A}$  is well-defined. Next, as  $R' \subseteq R$  is a RAF in  $G_{\mathcal{F}}^I$ , by construction, every reaction  $l_{i,j}^R \in R'$  must be catalysed by  $l_{i,j}^X \in \text{cl}_{R'}(F)$ . Since each  $l_{i,j}^X$  molecule is produced only by reactions  $l_{((i-1) \bmod m, j')}^R$  for  $1 \leq j' \leq 3$ , it follows that for all  $l_{(i,j)}^R \in R'$ :

$$l_{(i,j)}^R \in R' \implies l_{((i-1) \bmod m, 1)}^R \in R' \vee l_{((i-1) \bmod m, 2)}^R \in R' \vee l_{((i-1) \bmod m, 3)}^R \in R' \quad (23)$$

It follows  $R'$  must form a cycle with at least one literal reaction  $l_{(i,j)}^R$  included per clause in  $\mathcal{F}$ . By the construction of  $\mathcal{A}$  we have every clause positive-valued, so  $\mathcal{A} \models \mathcal{F}$ .

( $\Leftarrow$ ) Suppose that  $\mathcal{A} \models \mathcal{F}$  is a satisfying assignment. We construct a uRAF  $R'$  as follows: from each clause  $c_i \in \mathcal{F}$ , arbitrarily select exactly one positive-valued literal  $l_{(i,j)} \in c_i$  under  $\mathcal{A}$  and include the corresponding reaction  $l_{(i,j)}^R$  in the reaction set  $R'$ . Since exactly one literal reaction is chosen per clause, each literal reaction  $l_{(i,j)}^R \in R'$  is catalysed by  $l_{(i,j)}^X \in \text{cl}_{R'}(F)$  produced by a reaction  $l_{((i-1) \bmod m, j')}^R \in R'$  for some  $1 \leq j' \leq 3$ . Further, since we have no contradictory inclusions of literal reactions (as  $\mathcal{A}$  is well-defined), it follows by construction that no reactions are inhibited. Therefore  $R'$  is a uRAF.

Finally, as we include only a linear number of reactions, molecule types and catalysis edges in  $\mathcal{F}$ , we can construct  $G_{\mathcal{F}}^I$  in polynomial time. This completes the proof.  $\square$