

## Autocatalytic Sets and Biological Specificity

Wim Hordijk · Peter R. Wills · Mike Steel

Received: date / Accepted: date

**Abstract** A universal feature of the biochemistry of any living system is that all the molecules and catalysts that are required for reactions of the system can be built up from an available food source by repeated application of reactions from within that system. RAF (reflexively autocatalytic and food-generated) theory provides a formal way to study such processes. Beginning with Kauffman's notion of "collectively autocatalytic sets", this theory has been further developed over the last decade with the discovery of efficient algorithms and new mathematical analysis. In this paper, we study how the behaviour of a simple binary polymer model can be extended to models where the pattern of catalysis more precisely reflects the ligation and cleavage reactions involved. We find that certain properties of these models are similar to, and can be accurately predicted from, the simple binary polymer model; however, other properties lead to slightly different estimates. We also establish a number of new results concerning the structure of RAFs in these systems.

**Keywords** Origin of life · autocatalytic sets · template-based catalysis · Wills–Henderson model

---

Wim Hordijk  
SmartAnalytiX.com  
Lausanne, Switzerland  
E-mail: wim@WorldWideWanderings.net

Peter R. Wills  
Universität Tübingen, Intergrative Transcriptomics  
Tübingen, Germany  
and  
University of Auckland, Dept. of Physics  
Auckland, New Zealand  
E-mail: p.wills@auckland.ac.nz

Mike Steel  
University of Canterbury, Allan Wilson Centre for Molecular Ecology and Evolution  
Christchurch, New Zealand  
E-mail: mike.steel@canterbury.ac.nz

## 1 Introduction

In its broadest sense, the term “autocatalysis” refers to a process whereby some entity facilitates the chemical construction of another instance of itself. Because this is also a molecular-level description of biological reproduction, the study of simply specified autocatalytic systems has proven to be a fruitful field for gaining insights into possible origins of life. While Eigen (1971) described the selective preservation of information in systems involving macromolecular sequences undergoing competitive reproduction, Kauffman (1971, 1986) drew attention to the coincidence of cooperative catalytic functionalities that could potentially create a self-sustaining system of polymers, irrespective of whether they resembled information-carrying genes, the hallmark of quasi-species and hypercycles (Eigen, 1971; Eigen and Schuster, 1979). The disparity between these approaches reflects differing views of how we account for the complex chemistry of biological systems. What features of molecular biological and biochemical processes characterise their integration at the origin of life? For Eigen (1971), the answer lies in the capacity of replicating polymers like RNA and DNA to evolve as a result of Darwinian selection, whereas Kauffman (1971, 1986) urges us to look at the possibility of self-amplifying networks generating themselves as a result of nothing more than natural coincidences of connectedness. It is not our intention to adjudicate the dispute implicit in these divergent points of view. Rather, we now wish to investigate the extent to which the theory of autocatalytic sets and its recent extensions (Kauffman, 1971, 1986, 1993; Steel, 2000; Hordijk and Steel, 2004; Mossel and Steel, 2005; Hordijk et al, 2011; Hordijk and Steel, 2012b; Hordijk et al, 2012; Hordijk and Steel, 2012a, 2013) can be applied to some of the core ideas of molecular biology and thereby contribute to the incremental refinement of the problem which is usually posed as the unanswerable question “What is Life?”

A significant objection raised against the idea of simple autocatalytic sets forming the nucleus of the original molecular processes that led to biology concerns the question of “biological specificity”. This was the term that the first molecular biologists (e.g., Crick (1958, 1970)) used to articulate the profound impression that variation in the structure of a single molecule (DNA) was responsible for the orderly variation in the corresponding particularities of organisms. The discovery and elucidation of the direct transfer of information from sequences of nucleotide triplets in DNA (or RNA) to sequences of amino acids in proteins – the genetic code – followed the prediction of Schrödinger (1944) that an organism, the biological phenotype, is constructed by use of a genotypic “codescript”. Schrödinger envisaged the codescript as information stored in an “aperiodic crystal”, the main structural features of which were found to be met by the one-dimensional sequence of heteropolymeric DNA (Watson and Crick, 1953).

However, it was soon realised that the mapping from polymer sequence information (especially genes) to the phenotypic properties of organisms was much more convoluted than the coded transfer of information from DNA to RNA to protein. The first step following execution of the genetic code, protein folding, results in the original sequence information becoming scrambled: the algorithm that maps it onto the catalytic functions of proteins is exceedingly complex. Nevertheless, though scrambled, this mapping is still orderly, rather than random; and if it were completely random there would be no path whereby success in the struggle for survival could be reliably reverse-coded in polymer sequences through

a gradual process of adaptation through natural selection. These considerations lead to the conclusion that if autocatalytic sets of polymers are to provide a plausible explanation for the origin of life, then their probable emergence must be demonstrated for systems of polymers like RNA and proteins. In the case of these polymers the variation of catalytic function with sequence displays an orderliness, albeit scrambled, that satisfies the fundamental principle of molecular chemistry: similar structures tend to have similar properties – molecular structure-function mappings are far from random.

The original description of probabilistic autocatalytic sets of polymers (Kauffman, 1971, 1986) was susceptible to criticism along these lines: chemical reactions and catalytic functions were enumerated without any regard for the structural (and therefore the likely chemical) similarities that polymer sequences share. Considerations of the role of complementary sequence matching as a model of substrate recognition in catalytic processes have gone some way to addressing this criticism (Kauffman, 1993; Hordijk et al, 2011; Hordijk and Steel, 2012b), but there is another aspect of specific molecular recognition processes that is not very realistically represented when polymers (of all lengths up to a certain size) are randomly assigned as catalysts of ligation or cleavage reactions. With random assignments it is possible for a molecule comprising only a few atoms to be required to recognise the exact sequences of two much bigger molecules that it has the task of ligating, or creating through cleavage. In the absence of any other constraints, it is hard to imagine how a polymer sequence of length  $k$  could specifically recognise portions of other polymers comprised of many more than  $k$  monomers in total. In the most extreme case, how could a monomer or dimer be a ligation catalyst for the ligation of two particular polymers of length significantly greater than 2, but not others with the same end-sequences at the site of ligation or cleavage? This could only occur in a real chemical system if there were other factors, in addition to the direct sequence recognition capabilities of the catalyst, that made those two particular polymers and not others prone to such catalytic ligation or cleavage. Special chemical constraints of this sort cannot be accommodated in a random mapping from polymer sequence to catalytic function.

In this work, we take steps to address this problem. In the first place, we demand that any polymer that acts as a catalyst must contain a structure that is realistically capable of recognising the molecular features on which it acts. The minimum recognition structure for a catalyst is taken to be an oligomeric sequence complementary (in the two-letter alphabet) to the ligation/cleavage sequence that it acts on. We also consider cases in which the recognition structure may be required to contain more bits of information than the complementary sequence alone, while remaining contiguous with it. This addresses two aspects of chemical realism: (i) that the properties of a particular local structure, as long as it is intact, will not usually be unduly affected by remote structural features; and (ii) that an orderly variation in function correlates with an orderly variation in structure. The second demand we make is that the functional complexity of structures should be commensurate with their structural complexity. We achieve this in the simplest possible way, by making the restriction that only molecules of maximum length can act as catalysts and the maximum sum of features they can recognise is of the same size. This restriction is rather crude but it achieves the desired result without adding elaborate details of indeterminate effect on our elementary model. We consider the effect of these requirements individually and then together. We reach

the conclusion that autocatalytic systems that do not involve information storage and coded transfer do not have zero capacity for the maintenance of biochemical specificity, a conclusion at apparent variance with the Sequence Hypothesis of Crick (1958).

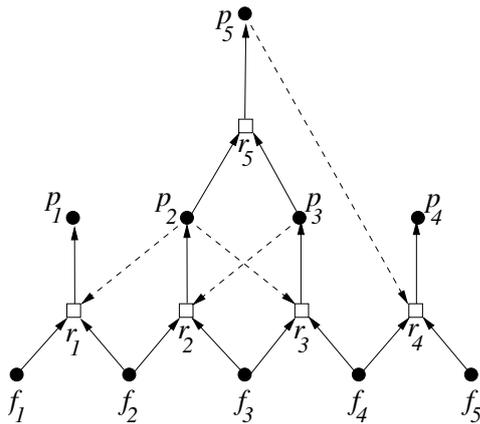
## 2 Chemical reaction systems and autocatalytic sets

We briefly review the relevant definitions and main results of autocatalytic set theory. First, a *chemical reaction system* (CRS) is defined as a tuple  $Q = \{X, \mathcal{R}, C\}$  consisting of a set of molecule types  $X$ , a set of chemical reactions  $\mathcal{R}$  and a catalysis set  $C$  that indicates which molecule types catalyse which reactions. We also include the notion of a food set  $F \subset X$ , which is a subset of molecule types that are assumed to be freely available from the environment. An *autocatalytic set* (or reflexively autocatalytic and food-generated (RAF) set) is now defined as a subset  $\mathcal{R}' \subseteq \mathcal{R}$  of reactions and associated molecule types which are:

1. *Reflexively autocatalytic* (RA): each reaction  $r \in \mathcal{R}'$  is catalysed by at least one molecule type involved in  $\mathcal{R}'$ , and
2. *Food-generated* (F): all reactants in  $\mathcal{R}'$  can be created from the food set  $F$  by using a series of reactions only from  $\mathcal{R}'$  itself.

A more formal definition of RAF sets is provided in Hordijk and Steel (2004); Hordijk et al (2011), including an efficient algorithm for finding RAF sets in general chemical reaction systems. It was shown that RAF sets are highly likely to exist in a simple model of chemical reactions systems known as the binary polymer model (Hordijk and Steel, 2004; Mossel and Steel, 2005), and that this result also holds when more realistic assumptions are included in the model (Hordijk et al, 2011; Hordijk and Steel, 2012b). An example of a simple CRS that contains RAF sets of size two and three is shown in Fig. 1. In this figure, the reaction network is represented as a bipartite graph (Bonchev and Mekenyan, 1994; Mincheva and Roussel, 2007). Black dots represent molecular species and white boxes represent reactions. Solid arrows indicate reactants and products going in and out of reactions, and dashed arrows represent catalysis.

The RAF sets that are found by our RAF algorithm are called *maximal* RAF sets (maxRAFs). However, it turns out that a maxRAF can often be decomposed into several smaller subsets which themselves are RAF sets (subRAFs) (Hordijk et al, 2012). If such a subRAF cannot be reduced any further without losing the RAF property, it is referred to as an *irreducible* RAF (irrRAF). The existence of multiple autocatalytic subsets can actually give rise to an evolutionary process (Vasas et al, 2012), and the emergence of larger and larger autocatalytic sets over time (Hordijk et al, 2012; Hordijk and Steel, 2012a). Recently, the formal RAF framework was also applied to an experimental chemical system of catalytic RNA molecules in which autocatalytic sets emerged spontaneously (Vaidya et al, 2012). The formal model is capable of reproducing the main experimental results and also provided additional insights and predictions about the system's behaviour (Hordijk and Steel, 2013).



**Fig. 1** A CRS that contains two RAF sets, the maxRAF  $\{r_1, r_2, r_3\}$  and the irrRAF  $\{r_2, r_3\}$ . Here  $F = \{f_1, \dots, f_5\}$ ,  $X = F \cup \{p_1, \dots, p_5\}$ ,  $\mathcal{R} = \{r_1, \dots, r_5\}$ , and catalysis is indicated by dashed arrows.

### 3 Models of chemical reaction systems

Here, we apply the RAF framework to two related models of chemical reaction systems, both of which are variants and extensions of the binary polymer model used previously. First, we briefly review the basic model, and then describe the two variants.

#### 3.1 The binary polymer model

The binary polymer model was originally introduced by Kauffman in the context of studying autocatalytic sets (Kauffman, 1986, 1993). Polymers are represented by strings of 0s and 1s, and the possible reactions are cleavage and ligation. Catalysis is assigned at random.

##### 3.1.1 The molecule set

The molecule set  $X$  consists of all bit strings up to (and including) a maximum length  $n$ :

$$X = \{0, 1\}^{\leq n}.$$

Therefore, there are  $|X| = 2^{n+1} - 2$  molecule types.

##### 3.1.2 The food set

The food set  $F$  consists of all bit strings up to (and including) a certain length  $t$ :

$$F = \{0, 1\}^{\leq t}.$$

Usually,  $t \ll n$  (e.g.,  $t = 2$  or  $t = 3$  is used).

### 3.1.3 The reaction set

The reaction set  $\mathcal{R}$  consists of all possible ligations (i.e., ways of “gluing” two bit strings together without violating the maximum length constraint) and cleavages (breaking a bit string into two parts). An example of a ligation reaction is  $000 + 1111 \rightarrow 0001111$ ; one for a cleavage reaction is  $010110101 \rightarrow 0101 + 10101$ .

There are  $|\mathcal{R}| = (n-2)2^{n+1} + 4$  possible ligation/cleavage reaction pairs, which can also be considered as one bi-directional reaction (although in terms of finding RAF sets, this does not make a difference).

### 3.1.4 The catalysis set

The catalysis set  $C$  is made up of combinations of molecules (bit strings) and reactions:

$$C = \{(x, r) | x \in X, r \in \mathcal{R}\},$$

In the model, these catalysis events are assigned independently and with equal probability  $p(n)$  across all possible  $(x, r)$  pairs (there are  $|X||R|$  such pairs, where the reactions  $r$  are considered to be bi-directional).

### 3.1.5 RAF sets

The binary polymer model was introduced to show that autocatalytic sets are highly likely to exist for a large enough diversity of molecule types, i.e., a large enough value of  $n$  (Kauffman, 1986, 1993). These arguments and results were refined later on, showing that RAF sets have a high probability of existence even for very moderate levels of catalysis – between one and two reactions catalysed per molecule, on average, for values of  $n$  at least up to 50 (Hordijk and Steel, 2004). Furthermore, despite the number of reactions growing exponentially with increasing  $n$ , a growth rate in the level of catalysis that is linear (with increasing  $n$ ) is sufficient (and also necessary) to maintain a high probability of RAF existence (Hordijk and Steel, 2004; Mossel and Steel, 2005).

## 3.2 An extended binary polymer model

Consider the following extended version of the binary polymer model in which the catalysis events  $\mathcal{E}(x, r, n)$  (i.e.,  $x$  catalyses  $r$ , for a maximum molecule length  $n$ ) are still independent across  $x$  and  $r$ , but where  $\Pr[\mathcal{E}(x, r, n)]$  can also depend on (some property of)  $x$  and  $r$ , instead of only on  $n$ . Allowing arbitrary dependence, however, is problematic. For example, suppose that one molecule  $m$  catalyses all reactions, or suppose that all reactions except the ones required in the last step to form  $m$  and all other molecules catalyse no reactions. Then the probability of an RAF can be arbitrarily close to 1 or 0, respectively. To obtain a balance between realism and tractability, we consider the following extended model in which  $0 \leq p(n) \leq 1$  and  $0 \leq m(x, r, n) \leq 1$  for all  $x, r, n$ :

$$\Pr[\mathcal{E}(x, r, n)] = p(n) \cdot m(x, r, n), \quad (1)$$

where  $m(x, r, n)$  is the probability that  $x$  and  $r$  conform to a given set of constraints (possibly involving  $n$ ), and  $p(n)$  is the probability that  $x$  catalyses  $r$  *given* that they conform to those constraints.

Here, we consider four versions of this extended model:

- **RAND**: The original (purely random) binary polymer model:

$$m(x, r, n) = 1$$

- **TMPL**: A template-based catalysis model, where  $x$  is considered a candidate catalyst for  $r$  only if, somewhere along its sequence, it matches the reaction template of  $r$  (or, equivalently, the complement of the reaction template). This reaction template could, for example, consist of the four bits (two on either side) around the cleavage/ligation site. We use the notation  $x \sim r$  to indicate such a template match between  $x$  and  $r$ . We thus have:

$$m(x, r, n) = \begin{cases} 1, & x \sim r; \\ 0, & \text{otherwise.} \end{cases}$$

- **MLEN**: Only molecules of maximum length  $n$  are considered as candidate catalysts:

$$m(x, r, n) = \begin{cases} 1, & \text{if } |x| = n; \\ 0, & \text{otherwise.} \end{cases}$$

- **BOTH**: A combination of the template-based and maximum-length constraints:

$$m(x, r, n) = \begin{cases} 1, & \text{if } x \sim r \text{ and } |x| = n; \\ 0, & \text{otherwise.} \end{cases}$$

Note that the **RAND** version of the model (i.e., the original model) was already described and investigated in detail in Kauffman (1986, 1993); Hordijk and Steel (2004); Mossel and Steel (2005) and the **TMPL** version (with a four-bit template) in Hordijk et al (2011); Hordijk and Steel (2012b). However, we have included these versions here for completeness and comparison (and as specific instances of the more general extended model), while the main interest is in the **MLEN** and **BOTH** versions of the model.

### 3.3 The Wills–Henderson Model

The Wills–Henderson (W-H) model, originally introduced in Wills and Henderson (2000), is another variant of the binary polymer model. It is defined as follows.

#### 3.3.1 The molecule set

The molecule set  $X = X(n)$  consists of all bit strings up to (and including) a maximum length  $n$ :

$$X = \{0, 1\}^{\leq n}$$

Therefore, there are  $|X(n)| = 2^{n+1} - 2$  molecule types.

### 3.3.2 The food set

The food set  $F$  consists of the two monomers (single bits), i.e.,  $t = 1$ :

$$F = \{0, 1\}.$$

### 3.3.3 The reaction set

The reaction set  $\mathcal{R} = \mathcal{R}(n)$  consists of additions (ligations) of a monomer to an already existing polymer (bit string) that has a length smaller than the maximum length  $n$ . Polymers are considered directional (left to right), and the monomer is added to its end. A distinction is made between adding a 0 to a 0, a 0 to a 1, a 1 to a 0, and a 1 to a 1. There are thus four ‘‘categories’’ of reactions, as follows:

1.  $\mathcal{R}_1 : b0 + 0 \rightarrow b00$ ,
2.  $\mathcal{R}_2 : b0 + 1 \rightarrow b01$ ,
3.  $\mathcal{R}_3 : b1 + 0 \rightarrow b10$ ,
4.  $\mathcal{R}_4 : b1 + 1 \rightarrow b11$ ,

where  $b$  is any bit string of length at most  $n - 2$  (including the empty string), i.e.,  $b \in \{0, 1\}^{\leq n-2}$ .

So, there are  $|\mathcal{R}(n)| = 2^{n+1} - 4$  reactions, and each category contains exactly one-quarter  $(2^{n-1} - 1)$  of these reactions. Reactions are again considered to be bi-directional (i.e., for each ligation reaction, there is the equivalent cleavage reaction).

### 3.3.4 The catalysis set

The catalysis set  $C$  is made up of combinations of molecules (bit strings) of maximum length  $n$  and reaction categories:

$$C = \{(x, \mathcal{R}_i) | x \in X, |x| = n, i = 1, 2, 3, 4\}.$$

Notice that if a maximum-length molecule  $x$  catalyses a reaction category  $\mathcal{R}_i$ , it catalyses *all* reactions in that category.

In the model, these catalysis events are assigned independently and with equal probability  $p(n)$  across all possible  $(x, \mathcal{R}_i)$  pairs ( $2^n \times 4 = 2^{n+2}$  such pairs).

### 3.3.5 RAF sets

RAF sets in the W-H model can contain reactions from any combination of reaction categories. For example, if the (maximum length) molecule  $0 \cdots 0$  catalyses the reaction category  $\mathcal{R}_1$ , then there exists an RAF set  $\mathcal{R}' = \{0 + 0 \rightarrow 00, 00 + 0 \rightarrow 000, \dots, 0 \cdots 0 + 0 \rightarrow 0 \cdots 00\}$ . In other words, all reactions involving only polymers of 0s are included in the RAF set (but not all reactions of  $\mathcal{R}_1$ ). Similarly, if the (maximum length) molecule  $0 \cdots 0$  catalyses the reaction category  $\mathcal{R}_4$  and (maximum length) molecule  $1 \cdots 1$  catalyses the reaction category  $\mathcal{R}_1$ , then there exists an RAF set  $\mathcal{R}' = \{0 + 0 \rightarrow 00, 00 + 0 \rightarrow 000, \dots, 0 \cdots 0 + 0 \rightarrow 0 \cdots 00, 1 + 1 \rightarrow 11, 11 + 1 \rightarrow 111, \dots, 1 \cdots 1 + 1 \rightarrow 1 \cdots 11\}$ . If all four reaction categories are catalysed by at least one molecule, then the entire reaction set  $\mathcal{R}$  becomes an RAF set. An RAF set that contains at least some (but not necessarily all) reactions from exactly  $j$  different reaction categories ( $j = 1, 2, 3, 4$ ) is hereafter referred to as a  $j$ -category RAF.

## 4 Results

### 4.1 The extended binary polymer model

#### 4.1.1 Theoretical results

We start with a theoretical result that generalises the original result of Mossel and Steel (2005) to the extended binary polymer model. First, we require some definitions and a slight modification of lemma 4.3(iii) of that paper.

Let

$$\lambda_r(n) = p(n) \cdot \sum_{x \in X(n)} m(x, r, n).$$

Then  $\lambda_r(n)$  is the expected number of molecules that catalyse ligation reaction  $r$  (for a given  $n$ ). Notice that, from (1) we have:

$$\lambda_r(n) = \sum_{x \in X(n)} \Pr[\mathcal{E}(x, r, n)], \quad (2)$$

and if we let  $\bar{\lambda}(n)$  be the average of these  $\lambda_r(n)$  values over all ligation reactions, then we have:

$$\bar{\lambda}(n) = \frac{1}{|\mathcal{R}_+(n)|} \sum_{r \in \mathcal{R}_+(n)} \lambda_r(n) = \frac{1}{|\mathcal{R}_+(n)|} \sum_{r \in \mathcal{R}_+(n)} \sum_{x \in X(n)} \Pr[\mathcal{E}(x, r, n)], \quad (3)$$

where  $\mathcal{R}_+(n)$  is the total set of ligation reactions.

Similarly, if we consider the dual quantities that were the focus of Mossel and Steel (2005), namely the expected number  $\mu_n(x)$  of ligation reactions that molecule  $x$  catalyses, and  $\bar{\mu}(n)$  the average value of these quantities, then we have:

$$\bar{\mu}(n) = \frac{1}{|X(n)|} \sum_{x \in X(n)} \mu_n(x) = \frac{1}{|X(n)|} \sum_{x \in X(n)} \sum_{r \in \mathcal{R}_+(n)} \Pr[\mathcal{E}(x, r, n)], \quad (4)$$

where  $X(n)$  is the total set of molecule types.

Comparing Eqns. (3) and (4), and noting that  $|\mathcal{R}_+(n)|/n|X(n)|$  converges exponentially quickly to 1 with increasing values of  $n$  (from Eqn. (2) and (3) in Mossel and Steel (2005), with  $\kappa = 2$ ), we obtain the following asymptotically exact link between these two averages:

$$\bar{\lambda}(n) \approx \bar{\mu}(n)/n.$$

We now state the modified lemma as follows.

**Lemma 1** *Under the extended binary polymer model, the probability that a ligation reaction  $r$  is catalysed by at least one molecule is:*

$$1 - \prod_{x \in X(n)} (1 - \Pr[\mathcal{E}(x, r, n)]) \geq 1 - \exp(-\lambda_r(n)).$$

*Proof:*  $1 - \Pr[\mathcal{E}(x, r, n)]$  is the probability that  $r$  is not catalysed by  $x$ , and so, by the independence assumption,  $\prod_{x \in X(n)} (1 - \Pr[\mathcal{E}(x, r, n)])$  is the probability that no molecule catalyses  $r$ . If we now apply the inequality:

$$\prod_i (1 - y_i) \leq \exp(-\sum_i y_i),$$

which holds when the  $y_i$  values are all non-negative, and invoke Eqn. (2), the probability that  $r$  fails to be catalysed by any molecule is, at most,

$$\exp(-\sum_{x \in X(n)} \Pr[\mathcal{E}(x, r, n)]) = \exp(-\lambda_r(n)).$$

The lemma now follows.

Finally, the generalised result can now be stated as the following theorem. Its proof follows a parallel argument to that provided for proposition 4.4 (ii) of Mossel and Steel (2005), based on Lemma 1.

**Theorem 1** *Given an instance  $\mathcal{Q}_n$  of the binary polymer model with food set  $F$ , suppose that for all reactions  $r \in \mathcal{R}$ ,  $\lambda_r(n) \geq \lambda$ . Then the probability that  $\mathcal{Q}_n$  contains an RAF involving all molecules is at least  $f(\lambda) = 1 - \frac{2(2^{-\lambda})^t}{1-2e^{-\lambda}}$ , which is independent of  $n$  and which converges to 1 exponentially fast as  $\lambda$  increases.*

This theorem, together with the fact that  $\bar{\mu}(n) \approx n\bar{\lambda}(n)$ , implies that in the extended binary polymer model, there is also a linear (in  $n$ ) upper bound on the growth rate in the level of catalysis ( $\bar{\mu}(n)$ ) required to get RAF sets with high probability. The argument from Mossel and Steel (2005) that provides a linear (in  $n$ ) lower bound on the growth rate in the level of catalysis required to get RAF sets also applies here, too. In other words, adding constraints on which molecules can catalyse which reactions (in the form of  $m(x, r, n)$ ), does not change this main result.

Next, we consider the question of whether the level of catalysis required to get RAF sets with high probability in the extended binary polymer model can be predicted from the (observed) required levels in the original (RAND) model. In Hordijk and Steel (2012b), we showed that this is possible for the TMPL version of the model by using an analytical approximation based on a mathematical technique called the *transfer matrix method*. This technique provides a way to calculate the number of bit strings of a certain length that contain a given substring analytically. From this, the probability can be derived that an arbitrary bit string (of length  $n$  at most, or exactly of length  $n$ ) matches the reaction template of an arbitrary reaction. These probabilities are then used as analytical approximations  $\hat{m}(x, r, n)$  of  $\bar{m}(x, r, n)$ , with which the required probability  $p(n)$  can be predicted (see Hordijk and Steel (2012b) for details of this analytical calculation).

Generalising this to the extended binary polymer model, we have:

$$\bar{\mu}(n) = |\mathcal{R}_+(n)| \cdot p(n) \cdot \bar{m}(x, r, n).$$

For the purely random model (RAND),  $\bar{m}(x, r, n) = 1$ , and the required value for  $p(n)$  to get, say, a probability  $P_n = 0.5$  to find RAF sets can be obtained from the simulation results (Hordijk and Steel, 2004). This provides a corresponding value for the average number of reactions catalysed per molecule,  $\bar{\mu}(n)$ . Now, for

the other model versions, we assume that a similar value for  $\bar{\mu}(n)$  is required to get a similar probability  $P_n$  of finding RAF sets. However, for these alternative model versions,  $\bar{m}(x, r, n) < 1$ , and so one would expect that the required value for  $p(n)$  needs to increase relative to that in the random model. Using the notation  $p(n)$  for the observed required probability in the random model and  $\hat{p}(n)$  for the predicted (or expected) required probability in the alternative model (TMPL, MLEN, or BOTH), we then get:

$$|\mathcal{R}_+(n)| \cdot p(n) = |\mathcal{R}_+(n)| \cdot \hat{p}(n) \cdot \hat{m}(x, r, n),$$

and thus:

$$\hat{p}(n) = \frac{p(n)}{\hat{m}(x, r, n)}.$$

Table 1 gives the analytically calculated values  $\hat{m}(x, r, n)$  (using the calculations described in Hordijk and Steel (2012b)) for several values of  $n$  for the model versions TMPL, MLEN, and BOTH.

$n$	TMPL	MLEN	BOTH
8	0.155	0.500	0.095
9	0.201	0.500	0.118
10	0.245	0.500	0.140
11	0.288	0.500	0.161
12	0.329	0.500	0.181
13	0.368	0.500	0.200

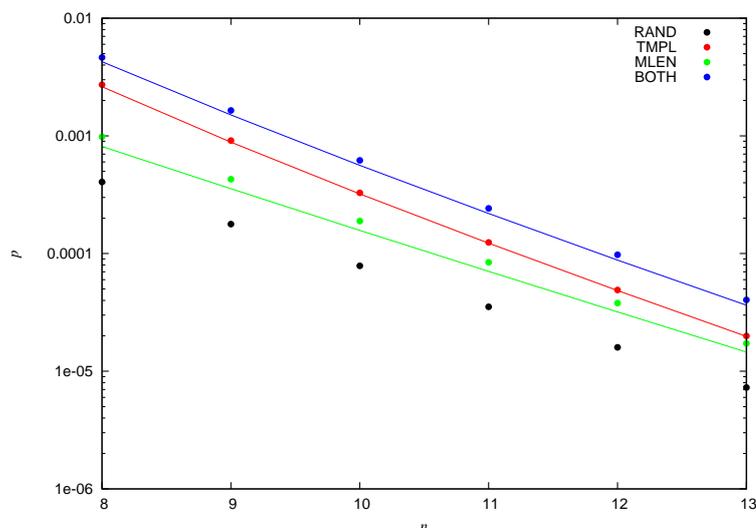
**Table 1** Analytical estimates  $\hat{m}(x, r, n)$  (rounded to three digits) for various values of  $n$  for the different model versions.

#### 4.1.2 Computational results

Figure 2 shows the corresponding predicted values  $\hat{p}(n)$  for these models (represented by the solid lines), based on the observed values  $p(n)$  for the random model (RAND; the black dots in Fig. 2). To see how accurate the predicted values  $\hat{p}(n)$  are, we performed computer simulations with the alternative models as well, using our RAF algorithm (Hordijk and Steel, 2004; Hordijk et al, 2011) to search for RAF sets in instances of these models. The observed  $p(n)$  values (to get  $P_n \approx 0.5$ ) are shown with dots in the same figure.

There are clear differences in the prediction accuracy between the different model versions. Figure 3 shows these differences in terms of the percentage of the predicted values. As the figure shows, the predictions for the TMPL (template-based) model are the most accurate, increasingly so for larger values of  $n$ . This confirms the observation already made in Hordijk and Steel (2012b) that larger molecules have a higher chance of matching a given (fixed-length) template, somewhere along their sequence. Therefore, for larger values of  $n$ , the template matching requirement becomes less and less of a constraint, and the predicted values for  $p(n)$  get more and more accurate.

The MLEN model is the least accurate, but also improves somewhat for larger values of  $n$ . It is, however, not surprising that for this model, the predictions

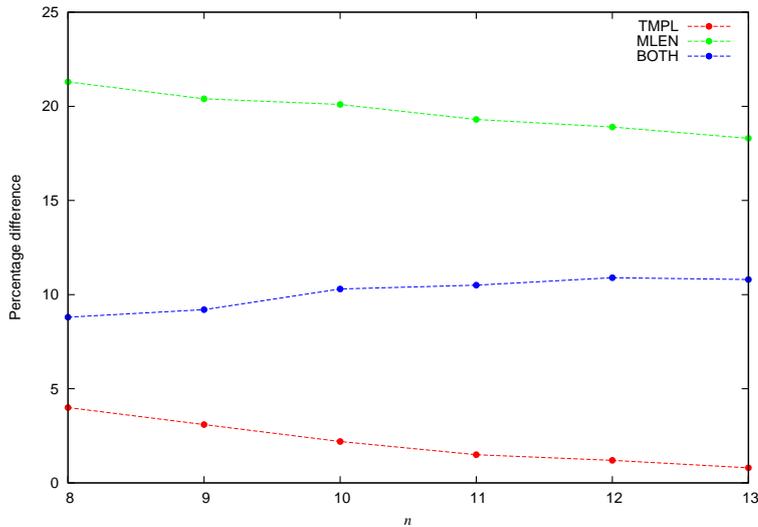


**Fig. 2** The analytically predicted (solid lines) and empirically observed (dots) values for the required probability  $p(n)$  (given the respective constraints  $m(x, r, n)$ ) to get  $P_n \approx 0.5$  (over 1000 instances) for the various model versions. Note that for the **RAND** model, there are only observed values (dots), on which the analytical predictions for the other model versions are based.

are less accurate. Using the analytical approximation  $\hat{m}(x, r, n)$  implies that this probability is independent and identical for each  $(x, r)$  pair. Obviously, this assumption is violated to a large extent in this model version, where only the largest molecules with a length of exactly  $n$  can be catalysts (which comprise exactly half of all molecules; hence  $\hat{m}(x, r, n) = 0.5$  for all  $n$ , as shown in Table 1).

Finally, the combination of template-based and maximum-length catalysis (the **BOTH** model) is somewhere in between in terms of accuracy. Interestingly, the accuracy actually *decreases* with larger  $n$ , but seems to level off eventually. This can be explained by the fact that, over all strings that match a given reaction template, the fraction of maximum-length strings is larger than 0.5 for smaller values of  $n$ , but converges to 0.5 with increasing  $n$ . For example, for  $n = 8$ , this fraction is 0.612 but for  $n = 13$ , it decreases to 0.534. The maximum-length requirement becomes, therefore, more of a constraint for larger values of  $n$ .

This last observation suggests an interesting measure for how much of a structural constraint a given requirement (such as template-based or maximum-length catalysis) imposes on the system in terms of its ability to form RAF sets. The (percentage) discrepancy between the analytically predicted value and the corresponding empirically observed value of  $p(n)$  can be taken as a measure of the severity of the imposed constraint. The more a given  $m(x, r, n)$  distribution deviates from being uniform over all  $(x, r)$  pairs, the larger the imposed structural constraint to form RAF sets will be, and, supposedly, the larger the discrepancy



**Fig. 3** The percentage difference between the theoretical and empirical values of  $p(n)$  in Fig. 2 for the various constrained model versions.

between the predicted and observed  $p(n)$  values. We return to this issue below by considering the “constructability” of RAF sets.

## 4.2 The Wills–Henderson (W-H) model

### 4.2.1 Theoretical results

We start again with some theoretical results, in particular on the probability of RAF sets existing in the W-H model. First, some definitions are required.

Given a subset  $\mathcal{R}'$  of  $\mathcal{R}$ , let  $J(\mathcal{R}') = \{j \in \{1, 2, 3, 4\} : \mathcal{R}' \cap \mathcal{R}_j \neq \emptyset\}$  denote the categories of reactions that are represented by at least one reaction in  $\mathcal{R}'$ . For a subset  $J$  of  $\{1, 2, 3, 4\}$ , let  $P_n^J$  be the probability that the W-H model (for polymers of length up to  $n$ ) has an RAF  $\mathcal{R}'$  with  $J(\mathcal{R}') = J$ .

Recall that a  $j$ -category RAF is defined as an RAF  $\mathcal{R}'$  that contains at least some (but not necessarily all) reactions from exactly  $j$  different reaction categories ( $j = 1, 2, 3, 4$ ). Thus, the probability that the W-H model (for polymers of length up to  $n$ ) has a  $j$ -category RAF is:

$$P_n^{(j)} = \sum_{J \subseteq \{1, 2, 3, 4\}: |J|=j} P_n^J,$$

The probability the model has an RAF is then  $\sum_j^4 P_n^{(j)}$ . We now derive theoretical approximations for these various probabilities.

– **1-category RAFs**

If  $J = \{2\}$  or  $J = \{3\}$ , only strings of length two can be created from the food set; therefore  $P_n^J = 0$  in both these cases.

If  $J = \{1\}$  or  $J = \{4\}$  then we generate exactly one sequence  $x$  of length  $n$  (either the all-0 string or the all-1 string). In this case, we have:

$$\begin{aligned} P_n^J &= \mathbb{P}(x \text{ catalyses } \mathcal{R}_j) \mathbb{P}(x \text{ doesn't catalyse any } \mathcal{R}_k \text{ for } k \neq j) \\ &= p(n)(1 - p(n))^3, \end{aligned}$$

and so

$$P_n^{(1)} \sim \binom{4}{1} p(n) = 4p(n).$$

- **2-category RAFs** If  $|J| = 2$ , say  $|J| = \{i, j\}$ , then in all cases, exactly two sequences,  $x$  and  $x'$ , of length  $n$  can be generated from the food set. Thus, in this case,  $P_n^J$  equals:

$$\mathbb{P}(x \text{ or } x' \text{ catalyses } \mathcal{R}_i) \mathbb{P}(x \text{ or } x' \text{ catalyses } \mathcal{R}_j) \mathbb{P}(x \text{ and } x' \text{ doesn't catalyse } \mathcal{R}_k \text{ or } \mathcal{R}_l),$$

where  $\{k, l\} = \{1, 2, 3, 4\} - \{i, j\}$ . Therefore,

$$P_n^J = (1 - (1 - p(n))^2)(1 - (1 - p(n))^2)(1 - p(n))^4 \sim 4p(n)^2.$$

Thus we have:

$$P_n^{(2)} \sim \binom{4}{2} \cdot 4p(n)^2 = 24p(n)^2.$$

- **3-category RAFs** The dominant and most interesting case is where  $J = \{1, 2, 3\}$  or  $J = \{2, 3, 4\}$ . Consider the first possibility (the other is similar). Here the number of molecules of maximal length  $n$  we can generate is precisely the number of sequences of length  $n$  in which two ‘1’s never appear consecutively (i.e.,  $\dots 11 \dots$  is forbidden). It is a classical result in enumerative combinatorics that this number is simply the Fibonacci number  $F_{n+1}$  where  $F_0 = F_1 = 1$  and  $F_n = F_{n-1} + F_{n-2}$  for all  $n > 1$ . The easiest way to see this is by virtue of an alternative description of the Fibonacci recursion:

$$F_{n+1} = 2F_n - F_{n-2}.$$

Thus we have:

$$P_n^J = \left[1 - (1 - p(n))^{F_{n+1}}\right]^3 (1 - p(n)) \sim (1 - e^{-c\mu})^3,$$

where  $p(n) = \mu / ((1 + \sqrt{5})/2)^n$  and  $c = \lim_{n \rightarrow \infty} F_{n+1} / \mu^n$  (notice that  $(1 + \sqrt{5})/2 = 1.618$ , the ‘golden ratio’).

The other 3-category RAFs are for  $J = \{1, 2, 4\}$  and  $\{1, 3, 4\}$ . For these the number of molecules of maximal length that are generated from  $F$  grows (only) linearly in  $n$ . So the 3-category RAFs are dominated in probability by the two interesting cases above.

- **4-Category RAFs** A 4-category RAF exists if and only if the entire set  $\mathcal{R} = \cup_{j=1}^4 \mathcal{R}_j$  of all reactions is an RAF. For the ‘only if’ part of this claim, note that if  $\mathcal{R}'$  is an RAF with  $J(\mathcal{R}') = \{1, 2, 3, 4\}$  then all of the reactions in  $\mathcal{R}$  are catalysed, and all the reactants of  $\mathcal{R}$  can be constructed from the food set  $F = \{0, 1\}$  using  $\mathcal{R}$ . Thus, for  $J = \{1, 2, 3, 4\}$ :

$$\begin{aligned} P_n^J &= \prod_{i=1}^4 \mathbb{P}(\mathcal{R}_i \text{ is catalysed by at least one polymer of length } n) \\ &= \prod_{i=1}^4 [1 - \mathbb{P}(\mathcal{R}_i \text{ is not catalysed by any polymer of length } n)] \\ &= (1 - (1 - p(n))^N)^4, \end{aligned}$$

for  $N = 2^n$ . Let us now write  $p(n) = \lambda/N$ . Then we have:

$$P^{(4)} = (1 - e^{-\lambda})^4 + o(1),$$

where  $o(1)$  refers to a term that converges exponentially quickly to 0 as  $n$  increases. Thus, for  $P_n^{(4)} = 0.5$ , we have  $\lambda = -\ln(1 - 2^{-0.25}) = 1.838\dots$ , and therefore:

$$p(n) \sim 1.838/2^n,$$

where (here and below)  $\sim$  denotes asymptotic equivalence as  $n$  grows.

Notice that when  $p(n) = 1.838/2^n$  (the 0.5 threshold for a 4-category RAF), we have the following:

- the probability of a 1-category RAF is  $\sim 7.35/2^n$ ;
- the probability of a 2-category RAF is  $\sim 81/4^n$  (a much smaller probability than for a 1-category RAF);
- the probability of a 3-category RAF of the most probable type (i.e.  $J = \{1, 2, 3\}$  or  $J = \{2, 3, 4\}$ ) is  $P_n^J \sim (1.838 \frac{F_{n+1}}{2^n})^3$  which converges to 0 exponentially quickly with  $n$  (but much slower than for a 1-category or 2-category RAF).

In summary, the  $P_n = 0.5$  threshold for an RAF in the W-H model converges asymptotically (and exponentially quickly with  $n$ ) to the  $P_n = 0.5$  threshold for 4-category RAFs. Any RAFs that are not 4-category RAFs are most likely to be 3-category RAFs. Of the remaining two, a 1-category RAF is much more probable than a 2-category RAF (but still much less than 3-category RAF). Thus the ordering is:

4-category  $\gg$  3-category  $\gg$  1-category  $\gg$  2-category.

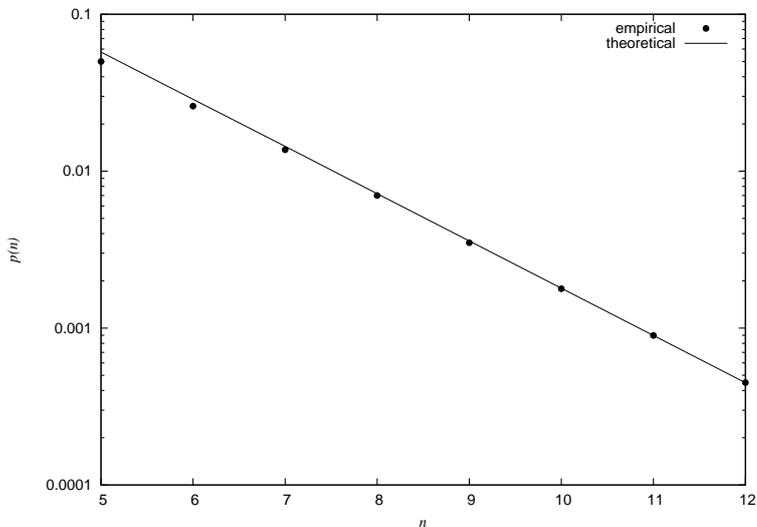
Finally, we consider the existence of irreducible RAF sets. In Steel et al (2013), we showed that, in general, finding the *smallest* irrRAF is a hard problem, and we introduced a randomised algorithm to find (arbitrary) irrRAFs and sample their sizes. However, in the specific case of the W-H model, it is actually possible to construct a polynomial-time algorithm to find the size of the smallest possible irrRAFs (they need not be unique) within a maxRAF  $\mathcal{R}'$ :

1. Take the set  $M$  of molecules of maximum length  $n$  that catalyse at least one reaction in  $\mathcal{R}'$ .
2. For each *minimal* subset  $S$  of  $M$  that includes exactly one catalyst for each catalysed reaction category (so  $S$  is a subset of four molecules at most from  $M$ ), do the following:
  - (a) For each  $x \in S$ , let  $R(x)$  be the sequence of  $n - 1$  reactions that generates  $x$  from  $F$  by adding monomers;
  - (b) Take  $R_S = \bigcup_{x \in S} R(x)$ .
3. The size of the smallest possible irrRAF's of  $\mathcal{R}'$ , is the size of the smallest set  $R_S$  generated in Step 2.

Note that this algorithm is polynomial in  $|M|$  (the number of catalysts in  $\mathcal{R}'$ ). Moreover, the algorithm implies that the size of the smallest irrRAF's of  $\mathcal{R}'$  must lie between  $n - 1$  and  $4(n - 1)$ .

#### 4.2.2 Computational results

We performed computer simulations with the W-H model for various values of  $n$  to check the accuracy of the theoretical predictions. Figure 4 shows the results, where the solid line represents the theoretical values and the dots the empirically observed values for  $p(n)$  to get a probability of around  $P_n = 0.5$  to find RAF sets (averaged over 1000 instances). As the plot shows, the theoretical predictions are very accurate, and increasingly so for larger values of  $n$ .



**Fig. 4** The theoretically predicted (solid line) and empirically observed (dots) values for the required probability  $p(n)$  to get  $P_n \approx 0.5$  (over 1000 instances) in the W-H model.

Furthermore, to check the prediction on the ordering of the four categories in terms of their likelihood, Table 2 shows the percentage of RAF sets for  $n = 8$  (and a value of  $p(n)$  that gives  $P_n = 0.5$ ) that are  $j$ -category RAFs for  $j = 1, 2, 3, 4$ . Indeed, 4-category RAFs dominate, the remainder consisting of 3-category, 1-category and 2-category RAFs (in that order), as predicted. However, for  $n > 10$ , basically all RAFs that are found are 4-category RAFs.

$j$	1	2	3	4
%	0.786	0.196	1.768	97.250

**Table 2** The percentage of instances where the found RAF set contains  $j = 1, 2, 3, 4$  reaction categories in the W-H model with  $n = 8$  and  $p(n) = 0.0070$ .

Even though the maximal RAF sets in the W-H model are predominantly 4-category RAFs, consisting of the entire reaction set  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3 \cup \mathcal{R}_4$ , they do contain many smaller RAF subsets. Figure 5 shows a histogram of the sizes (in number of reactions) of 300 irrRAF sets as found by the randomised algorithm (Steel et al, 2013) within one particular maximal RAF set for  $n = 10$ . The size of this 4-category maxRAF is  $|\mathcal{R}'| = |\mathcal{R}(10)| = 2^{11} - 4 = 2044$  reactions. However, as the figure shows, the sizes of the irrRAF sets found range from about 25 to 65 reactions (i.e., they are much smaller than the maxRAF). Indeed, the smallest irrRAF size found by the randomised algorithm is equal to the minimum irrRAF size calculated by the exact algorithm for the W-H model introduced above.

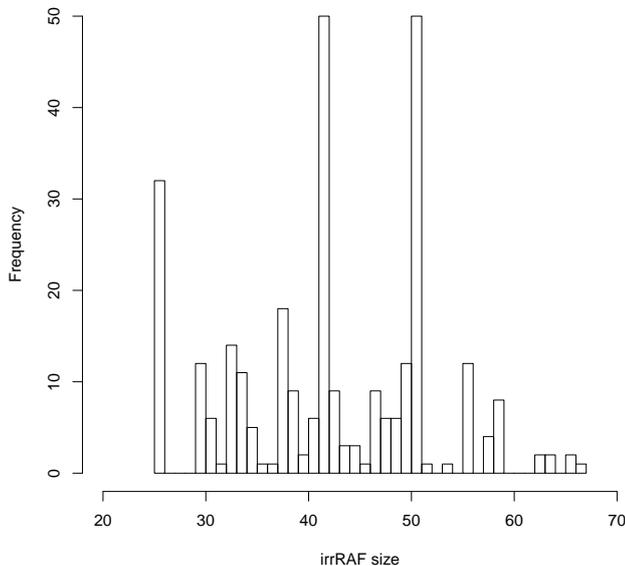
### 4.3 Constructability of RAFs

Note that the formation of an RAF starting from the food set (of polymers up to length  $t$ ) requires a minimum of  $\log_2(n) - t$  reactions to proceed uncatylsed before the first catalyst can be produced in the MLEN version of the binary polymer model. Similarly, there need to be at least  $n - 1$  such uncatylsed reactions in the W-H model. The definition of RAF sets allows for this to occur, since reactions can still proceed uncatylsed, albeit at a much slower rate. However, a chemical network in which catalysts are produced before they are needed is likely to have a significant advantage over the types where catalysis comes late (as in the MLEN model), because in the former case, an RAF would form more quickly, and before the reactants dissipate. In the extreme case, we have the notion of a ‘constructively autocatalytic F-generated set’ (CAF), studied in Mossel and Steel (2005), which can be built up in such a way that each reaction is catalysed by molecules already available. A more formal definition follows.

Given a chemical reaction system,  $\mathcal{Q} = (X, \mathcal{R}, C)$  with food set  $F$ , recall that  $\mathcal{R}'$  is a CAF if there is a linear ordering of  $\mathcal{R}'$ ,  $r_1, r_2, \dots$ , so that, for each  $i > 1$ :

- (P1) all reactants of  $r_i$  are contained in the closure of  $F$  relative to  $\{r_1, \dots, r_{i-1}\}$ ;
- (P2) at least one catalyst of  $r_i$  lies in the closure of  $F$  relative to  $\{r_1, \dots, r_{i-1}\}$ .

If  $\mathcal{R}'$  is an RAF but not a CAF, an interesting question then is whether or not there exists an ordering that satisfies (P1), and which requires at most  $k$  violations of (P2). When such an ordering exists, we say that the RAF is *constructible from  $F$  modulo  $k$  catalysations*.



**Fig. 5** Histogram of the sizes of 300 randomly generated irrRAF sets within one particular 4-category maxRAF in the W-H model with  $n = 10$ .

**Example:**

The maxRAF  $\{r_1, r_2, r_3\}$  shown in Fig. 1 is constructible from  $F$  modulo one catalysation. For example, the ordering  $r_2, r_1, r_3$  fails (P2) for just the first reaction ( $r_2$ ), and clearly satisfies (P1).

Now consider the following decision problem.

**$k$ -cat-RAF**

INSTANCE: A chemical reaction system and food set  $(Q, F)$ , an RAF  $\mathcal{R}'$  for  $(Q, F)$  and a positive integer  $k$ .

QUESTION: Is  $\mathcal{R}'$  constructible from  $F$  modulo  $k$  catalysations?

Determining whether a given RAF is constructible from  $F$  modulo  $k$  catalysations turns out to be an intractable problem, as the following result shows. The proof is provided in the Appendix.

**Theorem 2**  $k$ -cat-RAF is NP-hard.

The importance of this theorem is that it tells us that, rather than searching for a general exact algorithm for min- $k$ -RAF, we should consider special cases, or try to obtain upper and lower bounds for the solution that can be calculated efficiently. For example, an easily computable upper bound on the smallest value of  $k$  for which  $\mathcal{R}'$  is constructible from  $F$  modulo  $k$  catalysations is to construct a nested sequence

$$F = X_0 \subset X_1 \subset X_2 \cdots \subset X_m = F \cup \pi(\mathcal{R}')$$

of subsets of  $X$  in which  $X_{i+1}$  (for  $0 \leq i < m$ ) is the set of molecules in  $X_m$  that can be generated from reactants in  $X_i$  by applying a reaction from  $\mathcal{R}'$ . Let us say that a molecule  $x \in X_{i+1}$  is *premature* if  $x$  is not in  $X_i$  and if none of the reactions from  $\mathcal{R}'$  that generate  $x$  from reactants in  $X_i$  is catalysed by any molecule in  $X_i$ . Then an upper bound on the smallest value of  $k$  for which  $\mathcal{R}'$  is constructible from  $F$  modulo  $k$  catalysations is the sum of the number of premature molecules in the sequence  $X_1 \subset X_2 \cdots \subset X_m$ .

## 5 Concluding comments

In this paper, we have studied the consequences of constraining the formation of self-sustaining autocatalytic (RAF) sets of polymers so that the variation of the catalysts' properties with their structure conforms with the main features of chemistry that underpin the maintenance of functional specificity in molecular biological systems. The original binary polymer model, as first described by Kauffman (1971, 1986) and later refined by others, provides important insights into the probability of RAF formation. However, the simplicity of the binary polymer model comes at the price of biochemical realism. It gives a very short molecule, or one with no matching template or other generically defined features conferring recognition capability, the same probability of catalysing a given cleavage-ligation reaction as a long sequence. An exactly matching template or "keyhole" active site is the sort of structure most likely capable of precisely discriminating substrates. Thus, it is important to ask how results derived from the simple binary polymer model might be affected if the action of catalysis were more specifically dictated by the fit between the reaction and the potential catalyst.

Here, we have investigated two types of extensions of the binary polymer model.

The first is the extended binary polymer model, for which a catalyst is required to have either a matching template or to be of maximal length, or subject to both these requirements. The maximum length model (MLEN) represents an extreme case of models which are constrained by the plausible demand that longer molecules have higher probability of catalysing a reaction than shorter ones; we study this extreme case, as we expect to find the greatest difference from the original binary polymer model. In the extended binary polymer models, we demonstrated that the degree of catalysation required for the likely emergence of RAFs grows linearly with the length of the sequences, as has already been established for the original binary polymer model (Mossel and Steel, 2005).

We then asked whether or not we can predict the density of catalytic functionality in the polymer sequence space required for the emergence of RAFs in these extended models, by substituting in its place the simple binary polymer model with the density adjusted to match the degree of catalysation of the corresponding more complex model averaged over all reactions. Calculating these average densities of catalytic function is possible (by standard methods from combinatorics) and, for the template-matching model, the predicted degree of catalysation required for RAF formation can be estimated quite precisely (with a discrepancy of approximately 1% at  $n = 13$ ) by the surrogate binary polymer model. For the other two models, the discrepancy is higher (20% for the MLEN model and 10% for both combined).

The reason for this increased discrepancy may be explained, at least in part, by the increased heterogeneity of the distribution of catalysts in the polymer sequence space for these latter two models. For example, when only maximum-length molecules are catalysts (the MLEN model), we can exactly fit the expected degree of catalysation in this model using a simple binary polymer model (with the degree of catalysation chosen appropriately) but the distribution of catalysis in the MLEN model shows higher variance than that in the surrogate simple model. More precisely, suppose we select a molecule  $\chi$  uniformly at random and, conditional on  $\chi = x$ , consider the number  $N(x)$  of reactions that molecule  $x$  catalyses. Consider the variance of the compound random variable  $N(\chi)$ . Under the MLEN model, this variance  $\sigma_{\text{MLEN}}^2$ , is greater than the variance  $\sigma_{\text{RAND}}^2$  of a matching simple binary polymer (where the average degree of catalysation per molecule in both models is  $m$ ). This can be seen by comparing the following equations<sup>1</sup>:

$$\sigma_{\text{RAND}}^2 = m \text{ and } \sigma_{\text{MLEN}}^2 = m + m^2.$$

The second extension we investigated was the W-H model in which the reactions are more restrictive than the original binary polymer model. Instead of allowing molecules to combine freely by ligation operations, the reactions in the W-H model attach just one monomer at a time to a polymer; moreover, catalysis is possible only by maximal length molecules, with each such molecule having the same probability of catalysing any one of the four classes of ligation reactions (i.e.  $\dots x + y \rightarrow \dots xy$  for  $x, y = 0, 1$ ).

The W-H model has an advantage over the other models in that one can mathematically calculate the exact probability that an RAF set exists, and specify its  $j$ -category type. Also, one can compute the size of the smallest RAF exactly, which was recently shown to be an NP-hard problem for the simple binary polymer model (Steel et al, 2013). Moreover, the smallest RAF in the W-H model is always small (linear in  $n$ ) but for the random binary polymer model, it was recently proved that, at the level of catalysis where RAFs are starting to emerge, the smallest RAFs are almost certain to be of a size that is exponential in  $n$  (Steel et al, 2013).

These attractive features of the W-H model are tempered by the rather coarse way in which RAFs emerge – these typically include all of the reactions (or initially at least one of the four classes of reactions). Thus, the probability of achieving system specificity of the type common in biological systems, i.e., catalysis of a selection of reactions, rather than all them, is very small. The original consideration of the W-H model (Wills and Henderson, 2000) focused on the bias away from a random distribution of catalytic functions in the sequence space of maximum length polymers (the ratio  $f'/f$  in the nomenclature of that paper) needed for an  $i$ -category RAF to survive dynamic competition with  $k$ -category RAFs for  $k > i$ .

Although W-H systems with an unbiased (random) distribution of catalysations in the MLEN sequence space have a strong tendency toward the maximal RAF set, much smaller RAF subsets will generally exist. The specification of these small RAFs in terms of reaction sets is rather artificial, in that any molecules

<sup>1</sup> The first equation holds because in RAND, and for every molecule  $x$ ,  $N(x)$  has a Poisson distribution with mean – and therefore variance – equal to  $m$ ; the second equation is from the identity  $\text{Var}[N(\chi)] = E[\text{Var}[N|\chi]] + \text{Var}[E[N|\chi]]$ , together with the fact that half the molecules have maximal length.

catalysing a reaction specific to a small RAF will also catalyse all of the other reactions in the same category. However, the existence of small RAFs demonstrates the small number of exactly specified reactions, perhaps coincidental byproducts of some other catalytic process of the same broad genre, that are needed to seed the generation of the maximal RAF. It is a feature of both approaches (the extended binary polymer model and the W-H model) that a number of reactions must proceed uncatalysed, or as a result of other processes external to the system in which autocatalysis eventually occurs, until the catalysts that contribute to the RAF are formed. This is most obvious for the W-H model, where a minimum of  $n - 1$  such uncatalysed reactions are required.

For the extended binary polymer model in which only maximal length molecules are catalysts we need a minimum of  $\log_2(n) - t$  uncatalysed reactions to generate the molecules capable of maintaining an RAF. Such steps represent obstacles to the formation of an RAF; reactions that are uncatalysed can proceed, but only at a slow rate, and this may be too slow in the presence of dissipation or degrading side-reactions. Thus it would be helpful to be able to compute, for any RAF, the smallest number of reactions that need to proceed uncatalysed before the RAF can be established. Our final result was to show that this problem is NP-hard, so it is unlikely that a polynomial-time algorithm exists for it, and suggests that alternative strategies should instead be explored. As a first example, we described a simple upper bound on the the number of uncatalysed reactions required, which can be computed in polynomial time.

It should be possible to extend the results further; some extensions would likely be straightforward and lead to similar results (for example, polymers over a non-binary alphabet, where results are typically similar to the binary case (Mossel and Steel, 2005)), while other extensions would probably introduce new complications (for example, allowing molecules to inhibit reactions, or introducing degrading side-reactions). The dynamics of RAF sets, which have been studied for the simple binary polymer model (Hordijk and Steel, 2012a), would also be of interest in these extended models. These and other studies should help provide increasing biological relevance of RAF theory, with the ultimate aim of providing a better understanding of how the first primitive self-sustaining autocatalytic systems may have become established.

## Acknowledgments

We thank the *Allan Wilson Centre for Molecular Ecology and Evolution* and the *Alexander von Humboldt Foundation* for helping fund part of this research. We also thank an anonymous reviewer for several helpful suggestions.

## References

- Bonchev D, Mekenyan O (1994) Graph theoretical approaches to chemical reactivity. Kluwer
- Crick FHC (1958) On protein synthesis. *Symposia of the Society for Experimental Biology* 12:138–163
- Crick FHC (1970) Central dogma of molecular biology. *Nature* 227:561–563

- Eigen M (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523
- Eigen M, Schuster P (1979) *The Hypercycle*. Springer, Berlin
- Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman
- Hordijk W, Steel M (2004) Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology* 227(4):451–461
- Hordijk W, Steel M (2012a) Autocatalytic sets extended: Dynamics, inhibition, and a generalization. *Journal of Systems Chemistry* 3:5
- Hordijk W, Steel M (2012b) Predicting template-based catalysis rates in a simple catalytic reaction model. *Journal of Theoretical Biology* 295:132–138
- Hordijk W, Steel M (2013) A formal model of autocatalytic sets emerging in an RNA replicator system. *Journal of Systems Chemistry* 4:3
- Hordijk W, Kauffman SA, Steel M (2011) Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *International Journal of Molecular Sciences* 12(5):3085–3101
- Hordijk W, Steel M, Kauffman S (2012) The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica* 60(4):379–392
- Kauffman SA (1971) Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics* 1(1):71–96
- Kauffman SA (1986) Autocatalytic sets of proteins. *Journal of Theoretical Biology* 119:1–24
- Kauffman SA (1993) *The Origins of Order*. Oxford University Press
- Mincheva M, Roussel MR (2007) Graph-theoretic methods for the analysis of chemical and biochemical networks. I. Multistability and oscillations in ordinary differential equation models. *Journal of Mathematical Biology* 55:61–86
- Mossel E, Steel M (2005) Random biochemical networks: The probability of self-sustaining autocatalysis. *Journal of Theoretical Biology* 233(3):327–336
- Schrödinger E (1944) *What is Life?* Cambridge University Press, Cambridge
- Steel M (2000) The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters* 3:91–95
- Steel M, Hordijk W, Smith J (2013) Minimal autocatalytic networks. *Journal of Theoretical Biology* 332:96–107
- Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N (2012) Spontaneous network formation among cooperative RNA replicators. *Nature* 491:72–77
- Vasas V, Fernando C, Santos M, Kauffman S, Sathmáry E (2012) Evolution before genes. *Biology Direct* 7:1
- Watson JD, Crick FHC (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171:964–967
- Wills P, Henderson L (2000) Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: ligation systems. In: Bar-Yam Y (ed) *Unifying Themes in Complex Systems: Proceedings of the First International Conference on Complex Systems*, Perseus Books, pp 613–623

## Appendix: Proof of Theorem 2

*Proof:* We will reduce the graph theoretic problem VERTEX COVER to  $k$ -cat-RAF (a similar reduction was employed in Steel et al (2013) for a quite different problem). Recall that for a graph  $G = (V, E)$ , a *vertex cover* of  $G$  is a subset  $V'$  of  $V$  with the property that each edge of  $G$  is incident with at least one vertex in  $V'$ ; VERTEX COVER has as its instance a graph  $G = (V, E)$  and an integer  $K$  and we ask whether or not  $G$  has a vertex cover of size at most  $K$ . This is a well-known NP-complete problem (Garey and Johnson, 1979) (indeed, it is one of Karp's original 21 NP-complete problems). Given an instance  $(G = (V, E), K)$  of VERTEX COVER we show how to construct an instance  $(X_G, \mathcal{R}_G, C_G, F_G, k)$ , of  $k$ -cat-RAF for which the answers to the two decision problems are identical (here,  $\mathcal{R}_G$  is an RAF).

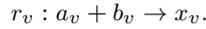
First we construct  $F_G$  and  $X_G$ . For each  $v \in V$  let  $a_v, b_v$  be two distinct elements of  $F_G$  and let  $x_v$  be an element of  $X_G - F_G$ . Order  $E$  as  $e^1, \dots, e^{|E|}$  and for each  $j = 1, \dots, |E|$  let  $d_j$  be a distinct element of  $F$  and  $y_j$  an element of  $X_G - F_G$ . Let  $d_0$  be another distinct element of  $F_G$ . Thus  $F_G$  consists of the  $2|V| + |E| + 1$  elements:

$$F_G := \{d_j : 0 \leq j \leq |E|\} \cup \{a_v, b_v : v \in V\},$$

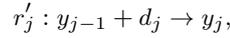
$X_G - F_G$  consists of  $|V| + |E|$  elements:

$$X_G - F_G := \{x_v : v \in V\} \cup \{y_j : 1 \leq j \leq |E|\}.$$

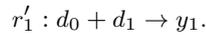
For each  $v \in V$ , define a reaction



For each  $1 < j \leq |E|$ , define the reaction:



and for  $j = 1$  let:



For any subset  $U$  of  $V$ , let  $\mathcal{R}_U = \{r_v : v \in U\}$ , let

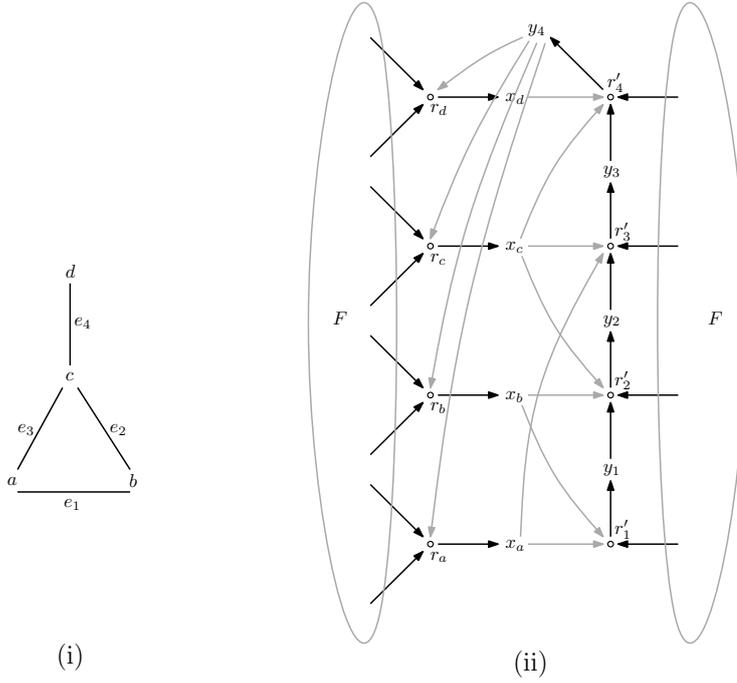
$$\mathcal{R}_V := \{r_v : v \in V\} \text{ and } \mathcal{R}_E := \{r'_j : 1 \leq j \leq |E|\},$$

and set  $\mathcal{R}_G = \mathcal{R}_V \cup \mathcal{R}_E$ . Thus we have specified  $X_G, F_G$  and  $\mathcal{R}_G$  and it remains to define the catalysis ( $C_G$ ) assignment, which is as follows:

- If  $e^j = (u^j, v^j)$  (where  $u^j, v^j \in V$ ) then  $r'_j$  is catalysed by both  $x_{u^j}$  and  $x_{v^j}$  (but by no other molecules).
- In addition, each reaction  $r_v : v \in V$  is catalysed by  $y_{|E|}$  and by no other molecule – we call the molecule  $y_{|E|}$  the *super-catalyst*.

An example of this construction is illustrated in Fig. 6. We have now fully specified the catalysation and thereby the pair  $(\mathcal{Q}_G, F_G)$  constructed from  $G$  ( $\mathcal{Q}_G = (X_G, \mathcal{R}_G, C_G)$ ).

**CLAIM:** A finite graph  $G$  has a vertex cover of size at most  $K$  if and only there is an ordering of  $\mathcal{R}_G$  that satisfies (P1) and involves at most  $K$  violations of (P2).



**Fig. 6** (i) A graph  $G$  and (ii) the associated CRS  $\mathcal{Q}_G$ , consisting of 8 reactions that form a RAF, and with the super-catalyst ( $y_4$ ) at the top.

To establish this claim, first suppose that  $V'$  is a vertex cover of  $G$  of size at most  $K$ . Then order  $\mathcal{R}_{V'}$  arbitrarily and place these as the first reactions in a linear ordering, followed by the reactions in  $\mathcal{R}_E$  in the order  $r'_1, \dots, r'_E$ , and finally the remaining reactions in  $\mathcal{R}_{V-V'}$  in arbitrary order as the final segment of the ordering. This ordering just requires  $|V'| = K$  violations of (P2) for the initial reactions (i.e.  $\mathcal{R}_{V'}$ ), and it also satisfies (P1), and so provides the required ordering of  $\mathcal{R}_G$ .

Conversely, suppose that there is an ordering of  $\mathcal{R}_G$ ,  $r_1, \dots, r_{|V|+|E|}$  that satisfies (P1) and involves at most  $K$  violations of (P2). Let  $J$  denote the set of  $j$  for which (P2) fails for  $r_j$ , and let

$$J_V = \{j \in J : r_j \in \mathcal{R}_V\} \text{ and } J_E = \{j \in J : r_j \in \mathcal{R}_E\}.$$

Each  $j \in J_E$  corresponds to some edge  $e$  of  $G$ , so we will let  $v(j)$  denote any vertex of  $G$  incident with  $e$ .

Now,  $\{r_j : j \in J_V\} \cup \{r_{v(j)} : j \in J_E\}$  is a subset of  $\mathcal{R}_V$  of size at most  $K$ , and so corresponds to  $\mathcal{R}_{V'}$  for a subset  $V'$  of  $V$  of size at most  $K$ . We show that  $V'$  is an edge cover of  $G$ , by showing that any given edge  $e$  contains at least one vertex from  $V'$ .

First, observe that the reaction  $r'_e \in \mathcal{R}_E$  is one of the reactions  $r_j$  in the above ordering of  $\mathcal{R}$ . We consider two cases: (i)  $j \in J_E$  and (ii)  $j \notin J_E$ . In Case (i),  $v(j) \in V'$  and so  $e$  contains this vertex from  $V'$ . In Case (ii),  $r_j$  is catalysed by a product of a reaction  $r_i$  that appears earlier in the ordering. This implies that

$r_i = r_v$  for some vertex  $v$  of  $V$ ; therefore, if  $v \in V'$  then  $e$  contains an element of  $V'$ . It remains to consider the case where  $v \notin V'$  (i.e.,  $i \notin J_V$ ). We will show that this case never arises by deriving a contradiction on the assumption that it does. If  $i \notin J_V$  then  $r_i$  is catalysed by a reaction  $r_k$  that appears earlier than  $i$  in the given ordering of  $\mathcal{R}$ . However, the only reaction that can catalyse  $r_i$  is  $r'_{[E]}$ , which must therefore appear as  $r_k$  for some  $k < i$  in the ordering (since we are assuming that  $i \notin J_V$ ). Summarising, we have:

$$k < i < j. \tag{5}$$

It is at this point that we invoke (P1). Notice that the reactants for  $r_k$  do not become available until all the other reactions in  $\mathcal{R}_E$  – including  $r_j$  – have occurred. By (P1), this requires that  $j < k$ . Combining this with Inequality (5) we obtain the required contradiction required to exclude the last case. This completes the proof.