CrossMark

ORIGINAL ARTICLE

Theoretical Analysis of a Self-Replicator With Reduced Template Inhibition Based on an Informational Leaving Group

Erwan Bigan[1,2] · Henri-Philippe Mattelaer[3] · Piet Herdewijn[3]

Abstract The first non-enzymatic self-replicating systems, as proposed by von Kiedrowski (Angew Chem Int Ed Engl 25(10):932–935, 1986) and Orgel (Nature 327(6120):346–347, 1987), gave rise to the analytical background still used today to describe artificial replicators. What separates a self-replicating from an autocatalytic system is the ability to pass on structural information (Orgel, Nature 358(6383):203–209, 1992). Utilising molecular information, nucleic acids were the first choice as prototypical examples. But early self-replicators showed parabolic over exponential growth due to the strongly bound template duplex after template-directed ligation of substrates. We propose a self-replicating scheme with a weakly bound template duplex, using an informational leaving group. Such a scheme is inspired by the role of tRNA as leaving group and information carrier during protein synthesis, and is based on our previous experience with nucleotide chemistry. We analyse theoretically this scheme and compare it to the classical minimal replicator model. We show that for an example hexanucleotide template mirroring that is used by von Kiedrowski (Bioorganic chemistry frontiers, 1993) for the analysis of the classical minimal replicator, the proposed scheme is expected to result in higher template self-replication rate. The proposed self-replicating scheme based on an informational leaving group is expected to outperform the classical minimal replicator because of a weaker template duplex bonding, resulting in reduced template inhibition.

Keywords Minimal replicator · Self-replication · Autocatalysis · Oligonucleotide · Template inhibition · Informational leaving group model

Erwan Bigan and Henri-Philippe Mattelaer contributed equally to this work. Piet Herdewijn is the contact author.

✉ Piet Herdewijn
chantal.biernaux@rega.kuleuven.be;
piet.herdewijn@rega.kuleuven.be

Erwan Bigan
bigan.erwan@orange.fr

Henri-Philippe Mattelaer
henriphilippe.mattelaer@rega.kuleuven.be

[1] Laboratoire d'Informatique (LIX), École Polytechnique, 91128 Palaiseau Cedex, France

[2] Laboratoire Matière et Systèmes Complexes, UMR7057 CNRS, Université Paris Diderot, 75205 Paris Cedex 13, France

[3] Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium

Introduction

The search for efficient non-enzymatic replication of nucleotides is of vital importance for origin-of-life research especially regarding the RNA world hypothesis (Gilbert 1986), based on RNA's ability to be both information carrier and catalytic agent (Kruger et al. 1982; Paul and Joyce 2002). Although theories vary, it is generally accepted that RNA ensured continuity of genetic information at some point during the prebiotic era. Starting from a series of non-enzymatic template-directed condensation reactions (Orgel 1992; Joyce 1987) and after a few simplifications, an autocatalytic oligonucleotide (von Kiedrowski 1986) was developed. Subsequently, more artificial replicators, either being nucleotides (Kruger et al. 1982; Paul and Joyce 2002; von Kiedrowski et al. 1991, 1989), peptides (Issac and Chmielewski 2002; Li and Chmielewski 2003; Lee et al. 1996) or small molecules (Vidonne and Philp 2009), appeared and began

🍃 Springer

Fig. 1 The minimal replicator (**a**) and the ILG model (**b**). Participating species are as follows: **a** precursors (**A**, **B**), template **C**, termolecular complex **ABC** and duplex C_2, **b** precursors (**A**, **B**), template **C**, termolecular complexes **ABC** and **D**, duplex C_2 and leaving group **L**. K_n denotes the dissociation constants of the respective complexes, k represents the (pseudo) irreversible step where a covalent bond is formed (and broken). **a** Reproduced from von Kiedrowski (1993), with permission of Springer
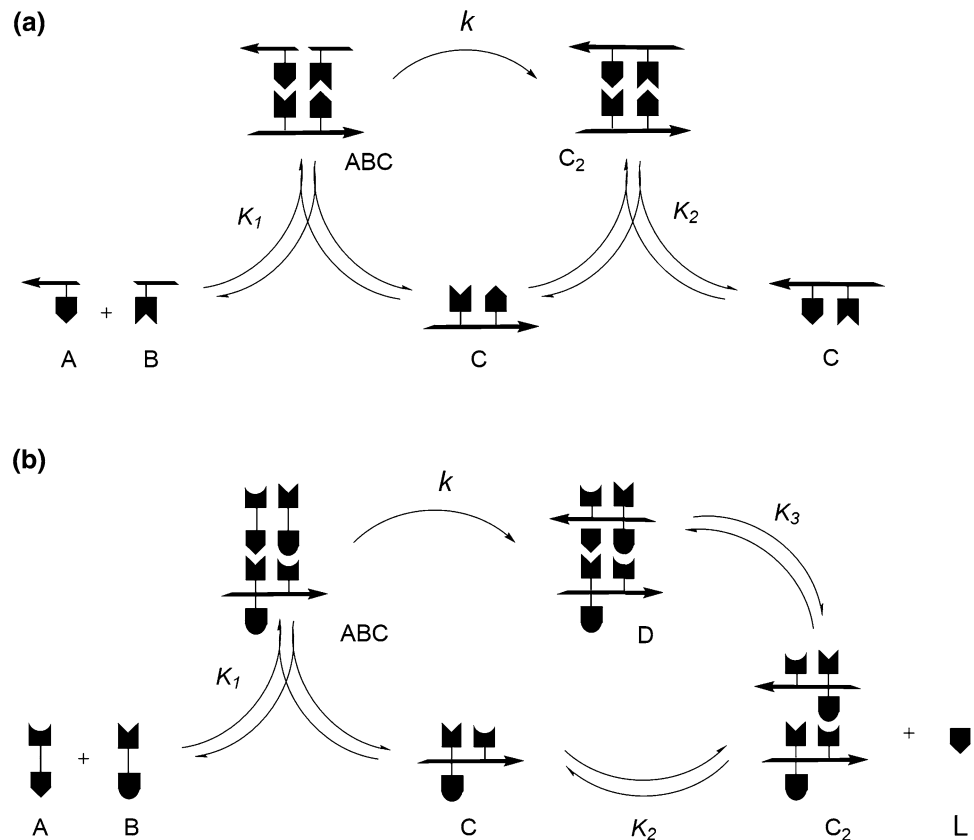
participating in more complex systems (Bissette and Fletcher 2013). But the analytical evaluation of all these replicators is still based on the same principles as discussed by von Kiedrowski in his minimal replicator theory (von Kiedrowski 1993).

The minimal replicator is based on the two, experimentally used, simplifications: a reaction cycle containing only one condensation step and using a self-complementary template. The latter is in contrast to natural nucleic acid replication which is cross-catalytic, where one strand catalyses the formation of the other and vice versa. The minimal replicator model consists of two precursors, **A** and **B**, and a template molecule **C**, complementary to **A** and **B** and facilitating the unidirectional ligation reaction between them by the reversible formation of a ternary complex **ABC**. The reversible dissociation of resulting duplex C_2 gives two **C** molecules as product, each capable of beginning a new replication cycle (Fig. 1a).

The analytical treatment provides tools to quantify and compare structurally different catalyst based on thermodynamic and kinetic data. One derived parameter is the autocatalytic reaction order p, describing the autocatalytic behaviour of the system. At $p = 1$, the reaction rate increases linearly with increasing amounts of template as expected with exponential growth. On the other hand, the

square root law of autocatalysis (von Kiedrowski 1986) ($p = 0.5$) describes the rate of autocatalysis as being proportional to the square root of initial template concentration. The latter case is a frequently witnessed phenomenon (von Kiedrowski 1986; Zielinski and Orgel 1987; von Kiedrowski et al. 1991, 1989; Lee et al. 1996) attributed to a strongly bound template duplex and is generally known as template inhibition. Much research has gone in to overcoming or avoiding such inhibition, by destabilising the **ABC** complex. Common strategies include distorting the template backbone (Li and Chmielewski 2003; Szostak 2012) and reducing template size (Paul and Joyce 2002; Issac and Chmielewski 2002).

Here we propose and analyse theoretically a chemical scheme to counter template inhibition by reintroducing an old concept. Inspired by the functionality of a few noncoding RNAs (Cech and Steitz 2014), we propose to promote the leaving group (LG) in self-replicating systems from plain activating agent to information carrier. This informational leaving group (ILG) approach is best illustrated by tRNA, acting as LG and information carrier during protein synthesis. Other examples from nature can be found among self-splicing, introns. Moreover, these oligonucleotide LGs express ribozyme activity besides recognition elements needed to excise the intron (Hausner

et al. 2014; Roitzsch et al. 2010). However, while tRNA acts as conduit for information transfer, introns serve a more selfish role consisting of self-recognition, cleavage and product release. Implementing an ILG strategy in autocatalysis was first suggested by Song and coworkers in a paper reporting the use of pyrophosphate-linked dinucleotides as substrates for DNA polymerases (Song et al. 2011). Furthermore, these compounds seem to be an important intermediate in non-enzymatic oligonucleotide synthesis (Sulston et al. 1968). Pyrophosphates not only spark the interest of prebiotic scientists, they are also known for playing a role in cellular processes such as energy metabolism (cofactor NAD^+) and DNA repair (poly(ADP-ribose)) (Schreiber et al. 2006). Thus, in light of these old problems and new interests, we propose to expand Song et al.'s LG strategy to non-enzymatic replication and analytically compare this proposed informational leaving group model (ILG model) with the minimal replicator model.

This paper is organised as follows. "Model" section describes the chemical reaction systems and mathematical models used to compare the proposed ILG scheme with the minimal replicator. "Results" section gives the results of this comparison and provides a discussion. And Section 4 gives a "conclusion". It should be noted that this paper contains equations and results for the minimal replicator model, which is not new material but is recalled here for the sake of clarity and consistency. In particular, Eqs. (20)–(24) are the same as Eqs. (29)–(33) in von Kiedrowski (1993). Likewise, Fig. 2a, top and bottom, mirrors Figs. 8a and 9 in von Kiedrowski (1993), respectively. And because the thermodynamic parameters used in Table 1a are very close to those used in von Kiedrowski (1993), the numerical results in Figs. 5a–c, 6a–c, 7a, 9a–c, 10a–c and 11a are very close to those of Figs. 10, 11, 12, 13, 14 and 15 in von Kiedrowski (1993), respectively.
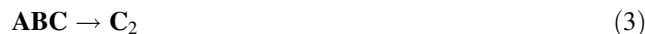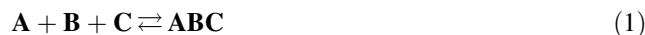
Model

Chemical Reaction System

Figure 1 illustrates both von Kiedrowski's original minimal replicator model (von Kiedrowski 1993) and the proposed ILG model. In both models, precursors (**A**, **B**) and template (**C**) reversibly form a termolecular complex **ABC**, and a subsequent series of kinetic and/or equilibrated steps lead to template self-replication. Similar to von Kiedrowski's original work, the approach followed here is that of a minimal model where all transient complexes between the template and the precursors are neglected. This results in the minimal replicator and ILG models given below.

Minimal Replicator Model

It consists of two fast quasi-instantaneously equilibrated bidirectional reactions and one unidirectional rate-limiting step:

$$\mathbf{A} + \mathbf{B} + \mathbf{C} \rightleftarrows \mathbf{ABC} \tag{1}$$

$$2\mathbf{C} \rightleftarrows \mathbf{C_2} \tag{2}$$

$$\mathbf{ABC} \rightarrow \mathbf{C_2} \tag{3}$$

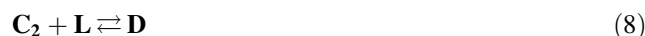with associated equilibrium constants for the equilibrated bidirectional reactions:

$$K_1 = \frac{[\mathbf{ABC}]}{[\mathbf{A}][\mathbf{B}][\mathbf{C}]} \tag{4}$$

$$K_2 = \frac{[\mathbf{C_2}]}{[\mathbf{C}]^2} \tag{5}$$

and with kinetic constant k for the unidirectional rate-limiting step.

ILG Model

It consists of three fast quasi-instantaneously equilibrated bidirectional reactions and one unidirectional rate-limiting step:

$$\mathbf{A} + \mathbf{B} + \mathbf{C} \rightleftarrows \mathbf{ABC} \tag{6}$$

$$2\mathbf{C} \rightleftarrows \mathbf{C_2} \tag{7}$$

$$\mathbf{C_2} + \mathbf{L} \rightleftarrows \mathbf{D} \tag{8}$$

$$\mathbf{ABC} \rightarrow \mathbf{D} \tag{9}$$

with associated equilibrium constants for the equilibrated bidirectional reactions:

$$K_1 = \frac{[\mathbf{ABC}]}{[\mathbf{A}][\mathbf{B}][\mathbf{C}]} \tag{10}$$

$$K_2 = \frac{[\mathbf{C_2}]}{[\mathbf{C}]^2} \tag{11}$$

$$K_3 = \frac{[\mathbf{D}]}{[\mathbf{C_2}][\mathbf{L}]} \tag{12}$$

and with kinetic constant k for the unidirectional rate-limiting step.

The reactions that were listed above for the ILG model are necessary for template self-replication. However, the formation of unwanted complexes is also possible:

1. Precursors **A** and **B** may bind to one another in two possible ways: looking at the schematic in Fig. 1b, either **B** binds on top of **A** with equilibrium constant K_2, or **A** binds on top of **B** with equilibrium constant K_3. This may lead to a chain polymerization giving sequences $(\mathbf{AB})_n$, where n is an integer. It can be easily

Fig. 2 Autocatalytic order p (*top*) and [**ABC**]$/c$ ratio (*bottom*) as a function of the decimal logarithm of K_1 and K_2 for the minimal replicator model (**a**), and K_1 and $K = K_2 = K_3$ for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For each plot of p, the vertical axis scale has been adjusted to the range of variation

(a) Minimal replicator

(b) ILG, l=0

(c) ILG, l=10 ×c
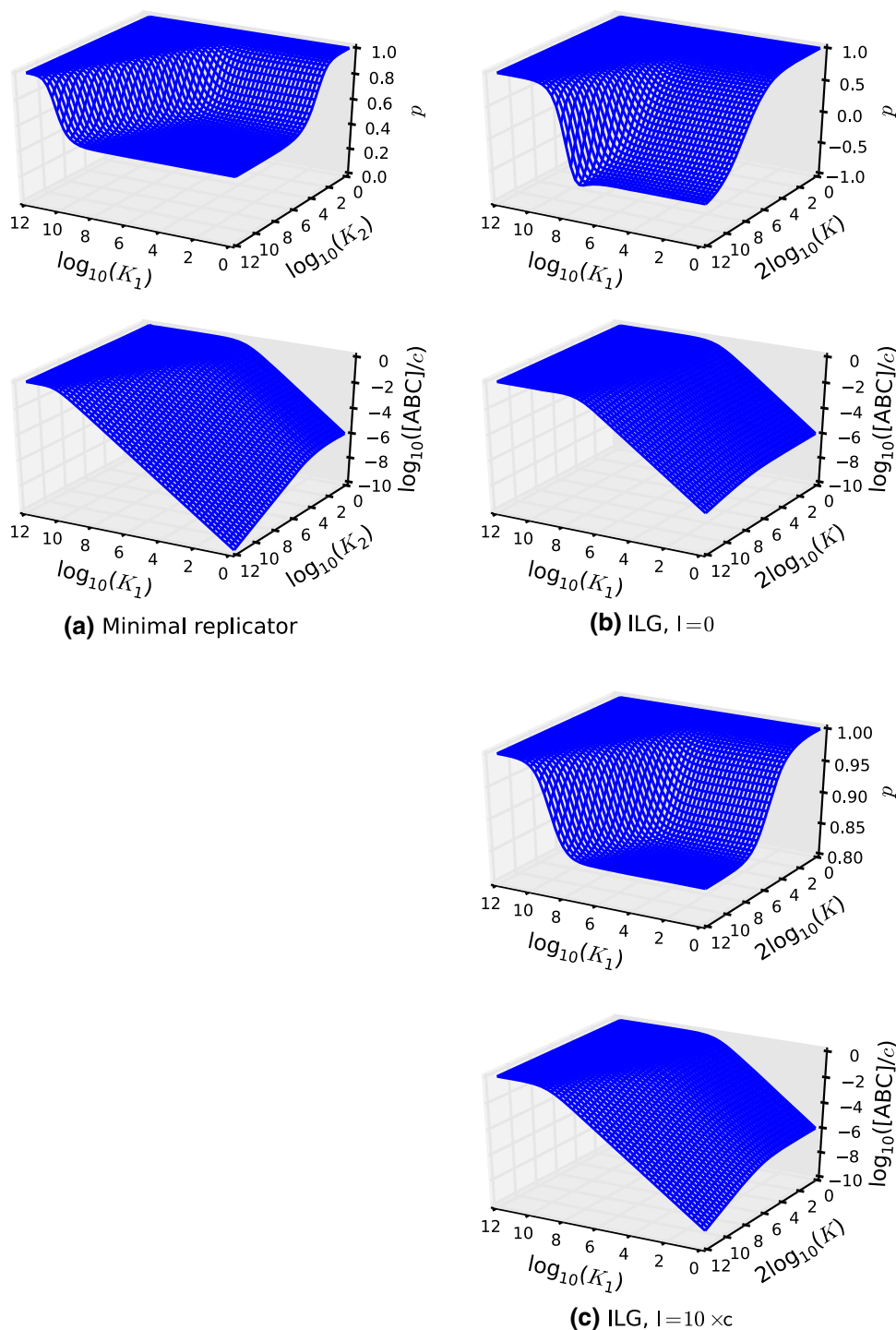
shown that if $K_2[\mathbf{B}]K_3[\mathbf{A}] \approx 1$, then **A** and **B** freeze into such chain polymers and are no longer available as free precursors. On the contrary, if $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$, then the amount of chain-polymerized $(\mathbf{AB})_n$ complexes is negligible. In the remainder of this text, we shall assume the latter.

2. Complexes **ABC**, **D** and **C₂** may also chain polymerize by binding to one another in any order,

with associated equilibrium constant K_2. Precursors **A** and **B** may also participate in such higher-order complexes. In the remainder of this text, we shall assume that the total amount of template c is much smaller than the total amount of precursor, a or b, so that in the above-assumed regime, $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$, the contribution of all such complexes can be neglected.

Table 1 Numerical parameters used for the comparison between the two models. Enthalpy changes (ΔH) are given in kcal mol^{-1} and entropy changes (ΔS) in cal mol^{-1} K^{-1}

	Best case	Worst case
(a) Minimal replicator		
ΔH_1	−56.45	$2 \times \Delta H_0$
ΔS_1	−160.72	$2 \times \Delta S_0$
ΔH_2	−56.45	
ΔS_2	−148.17	
ΔH_0	−22.23	
ΔS_0	−66.37	
(b) ILG		
ΔH_1	−58.36	$\Delta H_2 + \Delta H_3$
ΔS_1	−163.94	$\Delta S_2 + \Delta S_3$
ΔH_2	−22.23	
ΔS_2	−66.37	
ΔH_3	−24.06	
ΔS_3	−69.05	

3. The leaving group **L** may bind to the template **C** to give the complex **LC** with equilibrium constant K_3. Such a complex may block further self-replication of C and should therefore be taken into account. Precursor B may also bind to this complex to give the complex **LBC** but in the above-assumed regime, $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$, the contribution of this complex can be neglected.

Within the assumed regime of $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$ the only unwanted reaction that should also be taken into account is

$$\mathbf{L} + \mathbf{C} \rightleftarrows \mathbf{LC} \tag{13}$$

with associated equilibrium constants for the equilibrated bidirectional reactions:

$$K_3 = \frac{[\mathbf{LC}]}{[\mathbf{L}][\mathbf{C}]}. \tag{14}$$

Rapid Equilibria

The model aims at quantifying the rate of template replication at short times after mixing a solution containing only the template with a solution containing the precursors. It assumes that right after mixing, the chemical system reaches thermodynamic equilibrium (as the result of quasi-instantaneously equilibrated reactions) before the rate-limiting, unidirectional step starts to proceed.

Minimal Replicator Model

This starting equilibrium state can be determined knowing the initial template concentration called c (c is the total amount of template in whichever form—single-strand **C** or

duplex $\mathbf{C_2}$ divided by the total volume) as well as the initial precursor concentrations a and b (which are the total amount of precursors **A** and **B**, respectively). This is because the quantities a, b and c are conserved by the equilibrated reactions, which results in

$$a = [\mathbf{A}] + [\mathbf{ABC}] \tag{15}$$

$$b = [\mathbf{B}] + [\mathbf{ABC}] \tag{16}$$

$$c = [\mathbf{C}] + [\mathbf{ABC}] + 2[\mathbf{C_2}]. \tag{17}$$

Assuming that the initial amount of template is much smaller than the amount of precursors, $c \ll a$ or b, we have $[\mathbf{ABC}] < c \ll a$ or b so that

$$a \approx [\mathbf{A}] \tag{18}$$

$$b \approx [\mathbf{B}]. \tag{19}$$

Defining the dimensionless quantity q as

$$q = \frac{1}{K_1 ab} \tag{20}$$

and using it in Eqs. (4) and (5) gives $[\mathbf{C}]$ and $[\mathbf{C_2}]$ as a function of the single unknown $[\mathbf{ABC}]$:

$$[\mathbf{C}] = q[\mathbf{ABC}] \tag{21}$$

$$[\mathbf{C_2}] = K_2(q[\mathbf{ABC}])^2. \tag{22}$$

Feeding the above equations in Eq. 17 makes $[\mathbf{ABC}]$ solution of the following quadratic equation:

$$c = (1 + q)[\mathbf{ABC}] + 2K_2(q[\mathbf{ABC}])^2 \tag{23}$$

which has a single positive solution:

$$[\mathbf{ABC}] = \frac{\sqrt{8K_2cq^2 + (1+q)^2} - (1+q)}{4K_2q^2}. \tag{24}$$

ILG Model

Besides the conserved quantity a, b and c, the system of equilibrated reactions has an additional conserved quantity which is the amount of leaving group l (which might be initially present as the result of residual hydrolysis in the starting precursor solutions). These conserved quantities are given by

$$a = [\mathbf{A}] + [\mathbf{ABC}] \tag{25}$$

$$b = [\mathbf{B}] + [\mathbf{ABC}] \tag{26}$$

$$c = [\mathbf{C}] + [\mathbf{LC}] + [\mathbf{ABC}] + 2[\mathbf{C_2}] + 2[\mathbf{D}] \tag{27}$$

$$l = [\mathbf{D}] + [\mathbf{L}] + [\mathbf{LC}]. \tag{28}$$

As for the minimal replicator model, the initial amount of template is also assumed to be much smaller than the amount of precursors, $c \ll a$ or b, so that

$$a \approx [\mathbf{A}] \tag{29}$$

$$b \approx [\mathbf{B}]. \tag{30}$$

As for the minimal replicator model, the dimensionless quantity q is also defined as

$$q = \frac{1}{K_1 ab}. \tag{31}$$

Using it in Eqs. (10), (11) and (12) gives $[\mathbf{C}]$, $[\mathbf{C_2}]$ and $[\mathbf{D}]$ as a function of two unknowns $[\mathbf{ABC}]$ and $[\mathbf{L}]$:

$$[\mathbf{C}] = q[\mathbf{ABC}] \tag{32}$$

$$[\mathbf{LC}] = K_3 q[\mathbf{ABC}][\mathbf{L}] \tag{33}$$

$$[\mathbf{C_2}] = K_2 (q[\mathbf{ABC}])^2 \tag{34}$$

$$[\mathbf{D}] = K_2 K_3 (q[\mathbf{ABC}])^2 [\mathbf{L}]. \tag{35}$$

Feeding the above equations in Eqs. (27) and (28) makes $[\mathbf{ABC}]$ and $[\mathbf{L}]$ solutions of the set of coupled equations:

$$c = (1 + q + K_3 q[\mathbf{L}])[\mathbf{ABC}] \\ + 2K_2 (q[\mathbf{ABC}])^2 + 2K_2 K_3 (q[\mathbf{ABC}])^2 [\mathbf{L}] \tag{36}$$

$$l = (1 + K_3 q[\mathbf{ABC}] + K_2 K_3 (q[\mathbf{ABC}])^2)[\mathbf{L}] \tag{37}$$

which has a single positive solution, but without any analytical expression.

AB Complex Formation

For both the minimal replicator and ILG models, the above analysis neglects the possible formation of complexes by direct association between the precursors $\mathbf{A}$ and $\mathbf{B}$. If the formation of such an $\mathbf{AB}$ has an equilibrium constant K_0 and is no longer neglected, then the conserved quantities a and b in the above equations should be replaced by the unbound concentrations of $\mathbf{A}$ and $\mathbf{B}$, which are given by

$$a = \frac{K_0(c_A - c_B) - 1 + \sqrt{\Delta}}{2K_0} \tag{38}$$

$$b = c_B - c_A + a, \tag{39}$$

where

$$\Delta = K_0^2 (c_A - c_B)^2 + 2K_0(c_A + c_B) + 1 \tag{40}$$

and where c_A and c_B are the total known concentrations in bound and unbound form, respectively. If there are multiple different such complexes (as is the case for the ILG model, see Fig. 8 in "Detailed Comparison for a Hexanucleotide Template Reference Example" section below), then K_0 should be replaced in the above two equations by the sum of association constants for all such complexes.

Template Self-replication

Starting from the above equilibrium that is reached quasi-instantaneously after mixing the template and precursor solutions, the total template concentration c (in whichever form—unbound or bound) starts to increase as the result of the unidirectional rate-limiting step. The rate of increase in c is given by

$$\frac{dc}{dt} = k[\mathbf{ABC}]. \tag{41}$$

Another quantity of interest is the autocatalytic order p defined as the exponent such that

$$\frac{dc}{dt} = \alpha c^p \tag{42}$$

which by differentiation gives

$$p = \frac{d\log(\frac{dc}{dt})}{d\log(c)} = \frac{d(\log([\mathbf{ABC}]))}{d(\log(c))} = \frac{c}{[\mathbf{ABC}]}\frac{d[\mathbf{ABC}]}{dc}. \tag{43}$$

Minimal Replicator Model

Differentiating Eq. 23 gives

$$p = \frac{4K_2 cq^2}{8K_2 cq^2 + (1+q)^2 - (1+q)\sqrt{8K_2 cq^2 + (1+q)^2}}. \tag{44}$$

ILG Model

The two conserved quantities c and l are both varying as the result of the unidirectional rate-limiting step, and at the same rate.

$$\frac{dc}{dt} = \frac{dl}{dt} \tag{45}$$

which is given by Eq. 41. Differentiating Eqs. (36) and (37) gives

$$\frac{dc}{dt} = (1 + q + K_3 q[\mathbf{L}])\frac{d[\mathbf{ABC}]}{dt} + K_3 q[\mathbf{ABC}]\frac{d[\mathbf{L}]}{dt} \\ + 4K_2 q^2 [\mathbf{ABC}]\frac{d[\mathbf{ABC}]}{dt} \\ + 2K_2 K_3 q^2 \left(2[\mathbf{ABC}]\frac{d[\mathbf{ABC}]}{dt}[\mathbf{L}] + [\mathbf{ABC}]^2 \frac{d[\mathbf{L}]}{dt} \right) \tag{46}$$

$$\frac{dl}{dt} = (2K_2 K_3 q^2 [\mathbf{ABC}] + K_3 q)\frac{d[\mathbf{ABC}]}{dt}[\mathbf{L}] \\ + (1 + K_2 K_3 (q[\mathbf{ABC}])^2 + K_3 q[\mathbf{ABC}])\frac{d[\mathbf{L}]}{dt}. \tag{47}$$

Equating the above two expressions (because of Eq. 45) gives the following expression for $d[\mathbf{L}]/dt$:

$$\frac{d[\mathbf{L}]}{dt} = \frac{d[\mathbf{ABC}]}{dt}$$
$$\times \frac{4K_2q^2[\mathbf{ABC}] + (1+q) + 2K_2K_3q^2[\mathbf{ABC}][\mathbf{L}]}{1 - K_2K_3q^2[\mathbf{ABC}]^2}.$$

(48)

Feeding this back into Eqs. (46) or (47) and rearranging terms gives

$$\frac{d[\mathbf{ABC}]}{dt} = \frac{dc}{dt} \times \frac{1 - K_2K_3q^2[\mathbf{ABC}]^2}{Q}$$

(49)

from which the autocatalytic order p can be determined using Eq. 43:

$$p = \frac{c}{[\mathbf{ABC}]} \times \frac{\left(1 - K_2K_3q^2[\mathbf{ABC}]^2\right)}{Q},$$

(50)

where

$$\begin{aligned}Q = \big(&4K_2q^2[\mathbf{ABC}]\\ &+ (1+q)\big)\big(1 + K_2K_3q^2[\mathbf{ABC}]^2 + K_3q[\mathbf{ABC}]\big)\\ &+ 4K_2K_3q^2[\mathbf{ABC}][\mathbf{L}]\\ &+ K_3q[\mathbf{L}]\big(1 + K_2K_3q^2[\mathbf{ABC}]^2\big).\end{aligned}$$

(51)

Results

The following numerical values are used in this section for both the minimal replicator and ILG models: $a = b = 1 \times 10^{-3}$ M and $c = 1 \times 10^{-4}$ M as in von Kiedrowski (1993). The total amount of leaving group l is an additional conserved quantity in the case of the ILG model, for which two extreme values are considered: $l = 0$ and $l = 10 \times c = 1 \times 10^{-3}$ M. The former assumes that hydrolysis of precursors prior to mixing the precursor and template solutions is negligible. The latter corresponds to a situation where the purity of the initial precursor $\mathbf{A}$ solution is only 50 % due to significant hydrolysis prior to mixing the precursor and template solutions. Further details on the source of l will be elucidated in "Detailed Comparison for a Hexanucleotide Template Reference Example" section.

General Comparison

For a given k, the ratio $[\mathbf{ABC}]/c$ quantifies the relative rate of increase of the total template quantity c (see Eq. 41). Figure 2 shows the autocatalytic order (top) and the ratio $[\mathbf{ABC}]/c$ (bottom) as a function, of K_1 and K_2 for the minimal replicator model, and of K_1 and $K = K_2 = K_3$ for the ILG model for two different choices of total amount of leaving group $l = 0$ or $l = 10 \times c$. The rationale for the choice $K_2 = K_3$ for the ILG model is that the bonds involved in the corresponding equilibrated reactions are

comparable. This will be made clearer with the reference example described in the following section.

Comparing the bottom plots of Fig. 2b, c, the rate of template replication decreases on increasing the total amount of leaving group, l. This is because the concentration of template bound to a leaving group $[\mathbf{LC}]$ and that of template duplex bound to a leaving group, $[\mathbf{D}]$, both increase on increasing l, which in turn contributes to decrease $[\mathbf{ABC}]$ because of the conservation relation 27.

Comparing the top plots of Fig. 2b, c, the autocatalytic order p increases on increasing the total amount of leaving group, l. Although surprising at first sight, this may be understood as follows. The lower the initial l, the higher the initial rate of template replication. Any template replication results in an equal leaving group production rate, see Eq. 45. This produced leaving group binds to the duplex, thus reducing the rate of increase of template replication. And the lower the initial l, the relatively larger this inhibition mechanism. In the extreme case where $l \gg c$, then $[\mathbf{L}] \approx l$ and from Eq. 36 $[\mathbf{ABC}]$ is solution of

$$c = (1 + q(1 + K_3l))[\mathbf{ABC}] + 2K_2(q[\mathbf{ABC}])^2(1 + K_3l)$$

(52)

which has the same form as Eq. 23 for the minimal replicator model, replacing q by $q(1 + K_3l)$ (or equivalently, replacing K_1 by $K_1/(1 + K_3l)$) and K_2 by $K_2/(1 + K_3l)$. Increasing l is thus formally equivalent to decreasing both K_1 and K_2, which asymptotically makes the autocatalytic order close to unity as can be seen from Fig. 2a.

This is also the reason why, where $p \geq 0.5$ for the minimal replicator model, p may turn negative for the ILG model (which would correspond to a sublinear c trajectory). However, we shall see in the next section below that for a realistic numerical example mirroring that of von Kiedrowski (1993), the autocatalytic order p for the ILG model can actually exceed that for the minimal replicator model.

It should also be noted that the assumed regime of $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$ (see end of "Chemical Reaction System" section) imposes $K_2 < 1/b = 1 \times 10^3$ M and $K_3 < 1/a = 1 \times 10^3$ M. This restricts the applicability of Fig. 2b, c to the quadrant defined by $2\log_{10}(K) < 6$ and $\log_{10}(K_1) < 6$ (assuming for the latter that K_1 is comparable to K_2K_3, which is itself comparable to K^2).

Detailed Comparison for a Hexanucleotide Template Reference Example

Figure 3 shows the reference hexanucleotide example used for the detailed comparison between the minimal replicator and ILG models. Whereas a DNA hexamer was used in von Kiedrowski (1993), the proposed ILG model relies upon
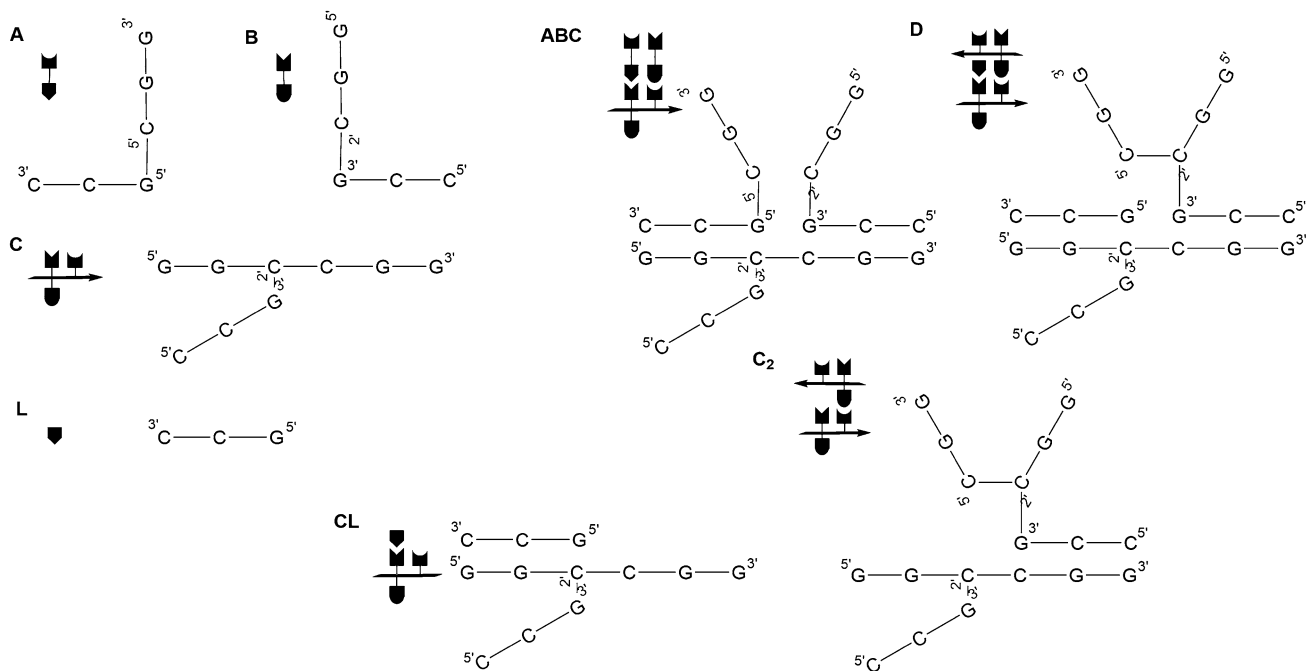
Fig. 3 ILG model. Example template and precursors used for the detailed comparison with the minimal replicator model. Other molecules and complexes that are formed in the course of template self-replication are also shown

RNA (because of the necessary 5′–5′ and 3′–2′ bonds) instead of DNA. For the sake of consistency, we chose to compare the proposed RNA-based ILG model with an RNA-based minimal replicator model using the same hexamer template as in von Kiedrowski (1993) (GCGCGC).

Such a palindromic sequence was taken from previous experimental work (von Kiedrowski 1986) where its self-complementary nature was needed in order to create a minimal replicator. In contrast, the example template chosen for the proposed ILG model should not be self-complementary, so as to avoid the possibility of a strongly bound duplex and the resulting template inhibition.

Besides, specific binding sites for each monomer are necessary in order to obtain the correct intermediate for ligation, exactly as in the example used for the minimal replicator model. The branched nonanucleotide template shown in Fig. 3 was chosen to exclude other template–precursor associations which would conflict with the regiochemistry of the ligation step. For example, choosing GCGCGC instead of GGCCGG as a template sequence (and CGC instead of GCC as a precursor trimer sequence) could result in the trimer CGC binding in two different locations.

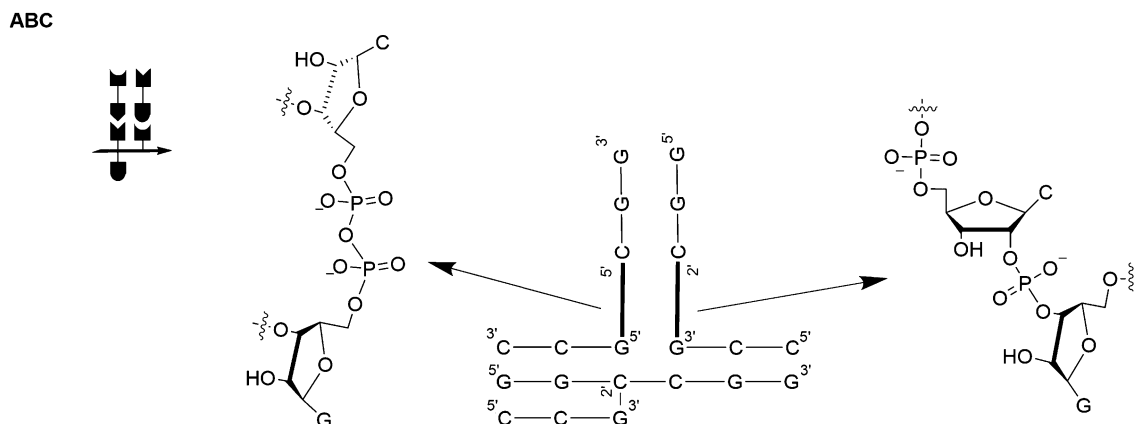Starting from the chosen template for the ILG model, precursors were chosen as follows:

Fig. 4 ILG model. Schematic of the 5′–5′ pyrophosphate GC and of the 2′–3′ phosphate CG bonds

1. Precursor **A** consists of two GCC and CGG trimers connected with a 5′–5′ pyrophosphate GC bond (see Fig. 4). Although pyrophosphates can be made in situ, the choice of a pre-activated monomer simplifies the model. The active bond makes it more prone to hydrolysis. It can be considered as the possible source of **L** and, thus, the previously discussed conserved quantity *l*.

2. Precursor **B** consists of two GGC and CCG trimers connected with a 2′–3′ phosphate CG bond. The resulting ligation of **B** with the non-LG part of **A** results in the branched template **C** (see Fig. 4). Hydrolysis of phosphate bonds will not be considered in both ILG and minimal replicator models.

To the best of our knowledge, the use of branched nucleotides in self-replication has not been explored yet, but such molecules have been used in ligation experiments, be it enzymatically (Mendel-Hartvig et al. 2004) or by chemical ligation (Carriero and Damha 2003). Moreover,
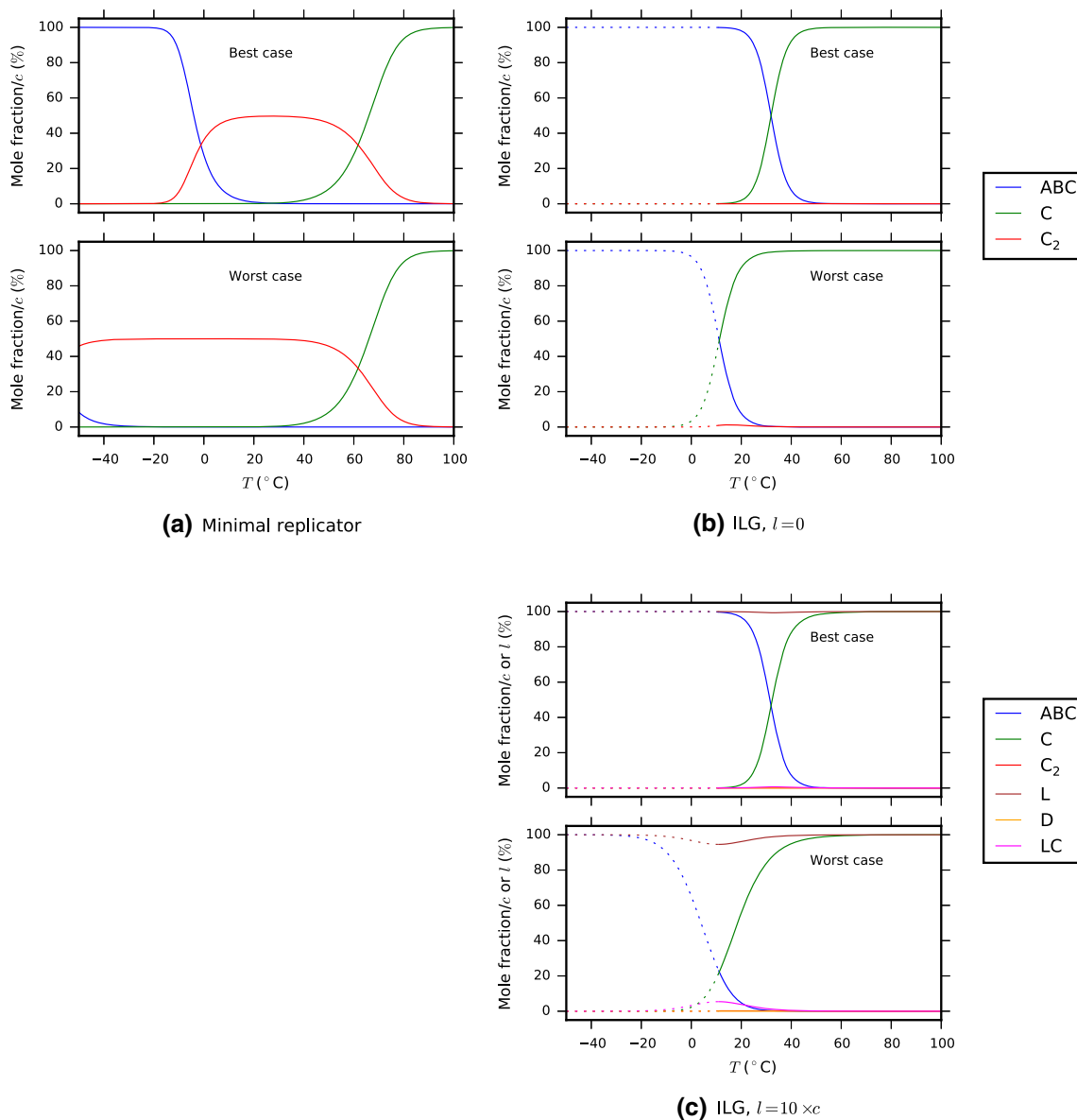

Fig. 5 Best case (*top*) and worst case (*bottom*) equilibrium distributions of the various species as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,°C$). The *vertical scale* is the concentration of each species divided by the corresponding conserved quantity: **ABC**/c, [**C**]/c and [**C**$_2$]/c for the minimal replicator model; **ABC**/c, [**C**]/c, [**C**$_2$]/c, [**L**]/l, [**D**]/l and [**LC**]/l for the ILG model

Without AB complex formation

(a) Minimal replicator

(b) ILG, $l = 0$

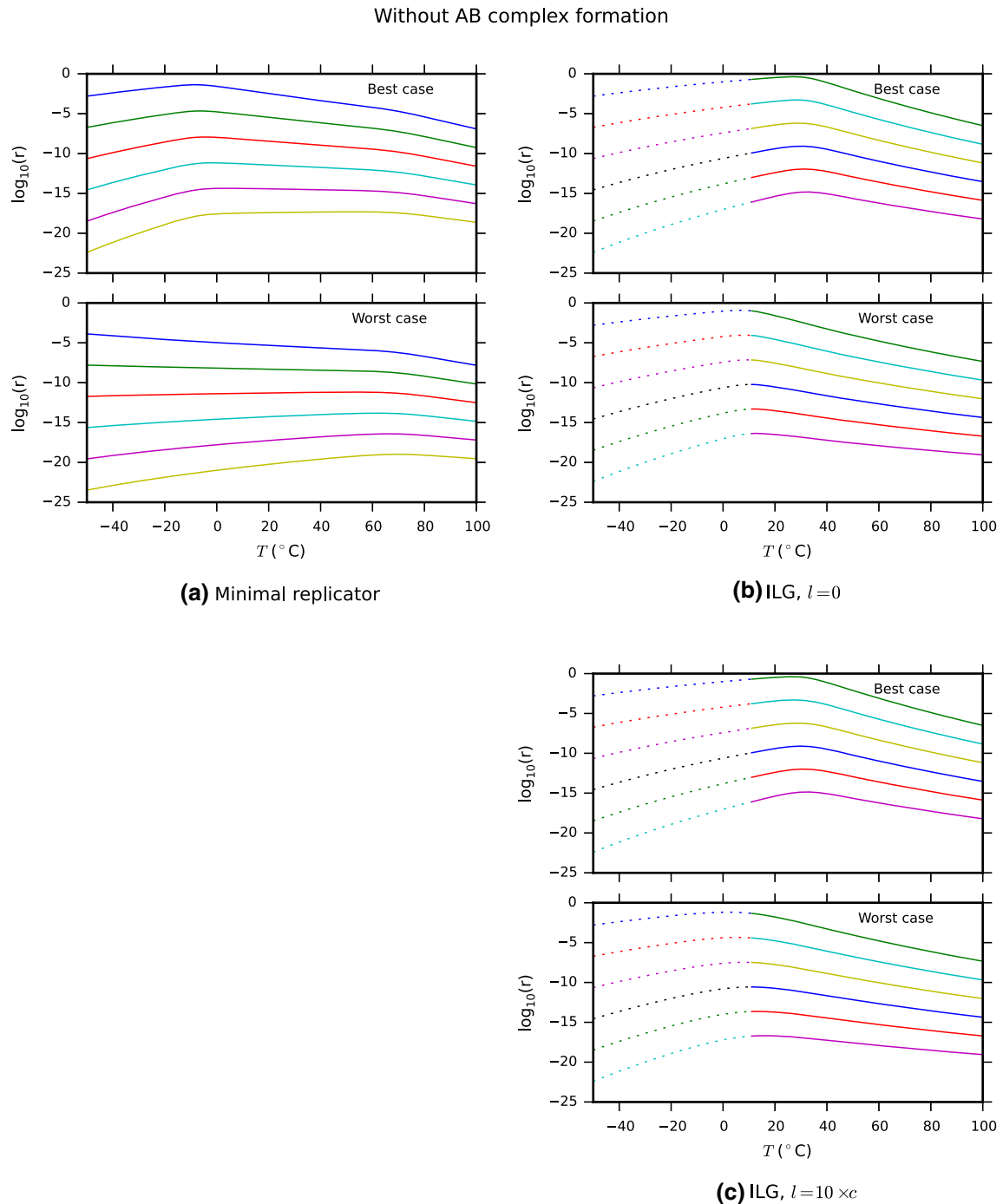**(c)** ILG, $l = 10 \times c$

Fig. 6 Best case (*top*) and worst case (*bottom*) decimal logarithm of the rate r of template formation as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,^\circ\text{C}$). The different curves correspond to different activation energies E_a ranging from 10 to 30 kcal mol^{-1} in 4 kcal mol^{-1} increments. The multiplicative coefficient in Eq. 53 was arbitrarily taken as $A = 1 \times 10^{11}$ s^{-1} as in von Kiedrowski (1993)

the reaction occurring in complex ABC, resulting in the branched nucleotide, reminds of the lariat structure in RNA splicing (Padgett et al. 1984; Peters and Toor 2015). The key difference between the previously discussed and

inspirational, self-splicing RNAs and our model lies in replacing the ribozyme activity by a purely chemical model using activated monomers. Although current synthetic methodology (Braich and Damha 1997) allows us to
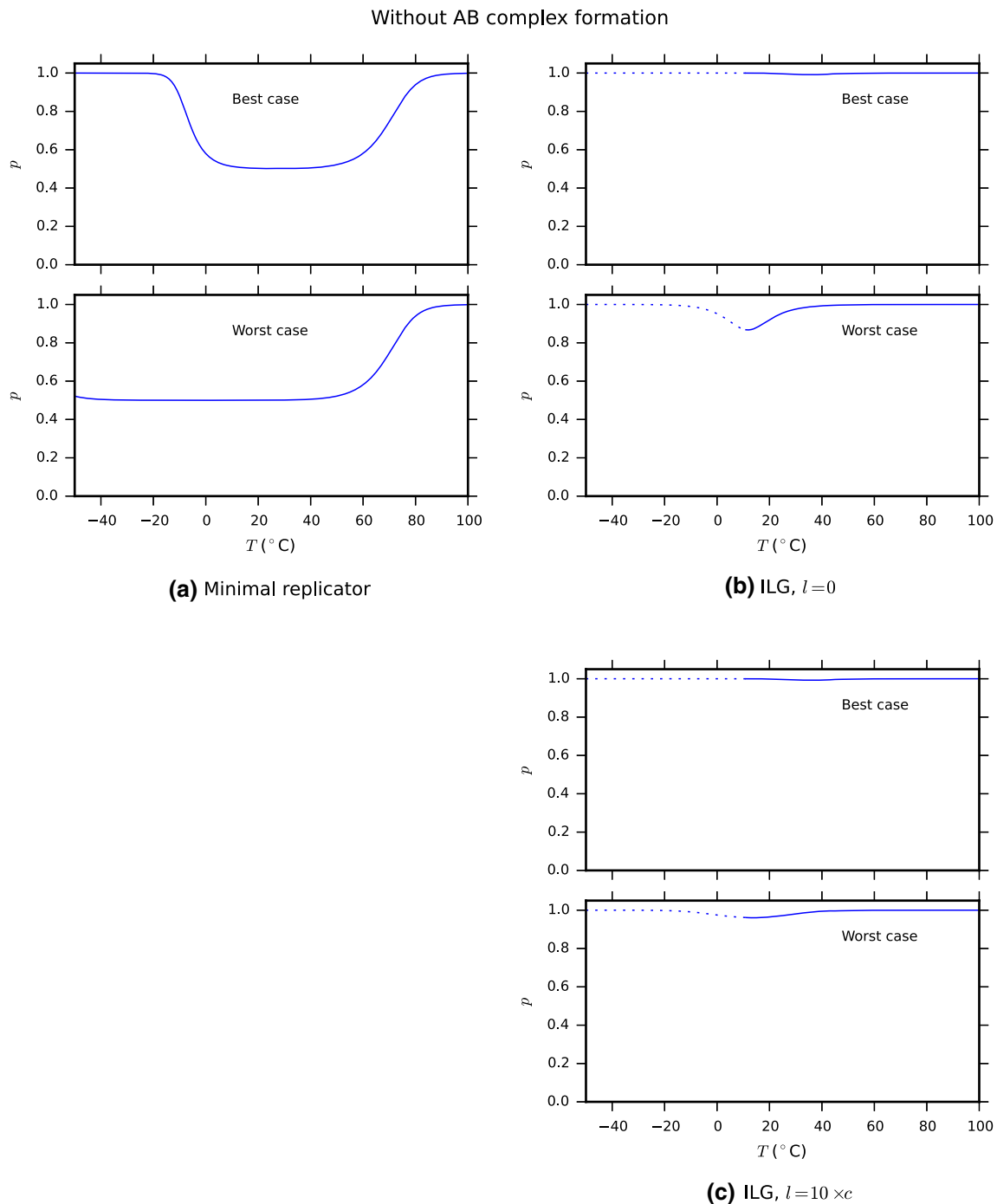
Without AB complex formation

(a) Minimal replicator

(b) ILG, $l=0$

(c) ILG, $l=10 \times c$

Fig. 7 Best case (*top*) and worst case (*bottom*) autocatalytic order p as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,^\circ$C)

translate our model into a hypothetical nucleotide-based replicator example, one important *caveat* should be kept in mind. Thermodynamic data are needed to compare the two models, but such experimental data are lacking for branched nucleotides. This is the reason why we used data and estimation methods of unbranched RNA as an approximation (Freier et al. 1985).

The numerical values used for comparing the two models are given in Table 1. The values for the RNA minimal replicator model are very close to those used in von Kiedrowski (1993) for DNA. $\Delta G_i = \Delta H_i - T\Delta S_i$, $i = 1$ or 2 for the minimal replicator model and $i = 1, 2,$ or 3 for the ILG model, is the standard Gibbs energy drop across the corresponding equilibrated reaction identified by
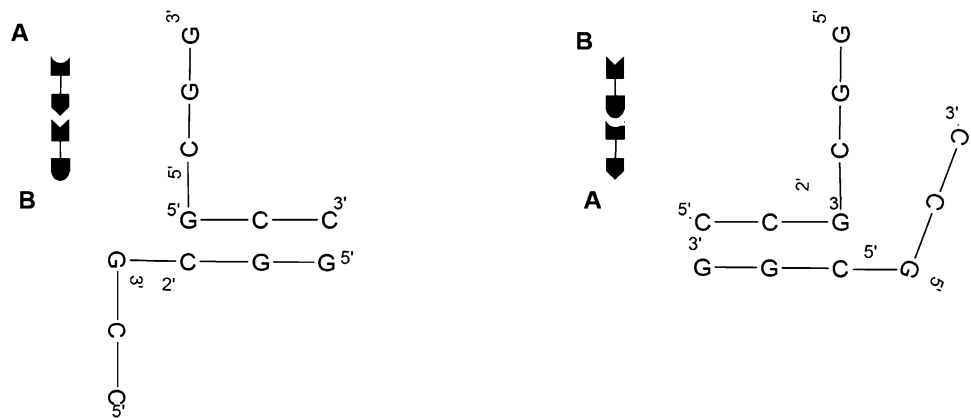
its index i, so that the corresponding equilibrium constant is given by $K_i = \exp(-\Delta G_i/\mathcal{R}T)$, where $\mathcal{R}$ is the ideal gas constant and T is the absolute temperature. As in von Kiedrowski (1993), a best case and a worst case were considered for the estimation of the precursors-to-template association parameters (ΔH_1 and ΔS_1). The best case assumes a cooperative binding between precursors upon addition to the template, whereas the worst case assumes non-cooperative binding.

It should be noted that the assumed regime defined by $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$ (see end of "Chemical Reaction System" section) restricts the applicability of the analysis to temperatures above 284 K (11 °C) for the ILG model.

Without Taking **AB** Complex Formation into Account

Figure 5 shows how the equilibrium distribution of the different species varies with temperature, for the minimal replicator model (left) and for the ILG model (right). For both minimal replicator and ILG models, a main difference between the best case and the worst case is the temperature below which the **ABC** complex can exist. Unlike for the minimal replicator model, for the ILG model the template duplex $\mathbf{C}_2$ is never present in any significant amount compared to the **ABC** complex or to the unbound template **C**. Increasing the amount of leaving group impacts only marginally this distribution.

As in von Kiedrowski (1993), both k and **C** contribute to the temperature dependence of the rate r of template formation given by Eq. 41. The rate constant k is expected to vary with temperature as

$$k = A \times \exp\left(-\frac{E_a}{\mathcal{R}T}\right), \tag{53}$$

where A is a constant. Figure 6 shows how the rate r of template formation varies with temperature, for the minimal replicator model (left) and for the ILG model (right), for various activation energies E_a.

Figure 7 shows how the autocatalytic order p varies with temperature, for the minimal replicator model (left) and for the ILG model (right).

von Kiedrowski 1993 distinguishes three regimes:

1. *Strong exponential* Most of the template resides in the **ABC** complex form, the autocatalytic order p is close to unity, and the rate of template formation r increases on increasing temperature.
2. *Weak exponential* Most of the template resides either in the **ABC** complex or in the unbound **C** forms, but not in the duplex $\mathbf{C}_2$ form, the autocatalytic order p is intermediate between 0.5 and 1, and the rate of template formation typically decreases on increasing temperature.
3. *Parabolic* Most of the template resides either in the **ABC** complex or in the duplex $\mathbf{C}_2$ forms, but not in the unbound **C** form, the autocatalytic order p is close to 0.5, and the rate of template formation either increases or decreases on increasing temperature.

Whereas the minimal replicator model behaves as strong exponential at low temperature, weak exponential at high temperature and parabolic at intermediate temperature, the ILG model behaves as strong exponential at low temperature and as weak exponential at high temperature. The autocatalytic order p stays close to unity throughout the investigated temperature range, and the rate of template formation r reaches its maximum at a temperature that is nearly independent of the activation energy. Besides, this behaviour is only marginally impacted by the initial amount of leaving group.

It thus appears that the ILG model is more favourable than the minimal replicator model because (i) its autocatalytic order remains close to unity over the entire temperature range, and (ii) the rate of template formation r is always higher for the ILG model than for the minimal replicator model. Besides, these properties hold for both best and worst cases (regarding the estimation of ΔG_1) as well as for a low or high initial amount of leaving group.
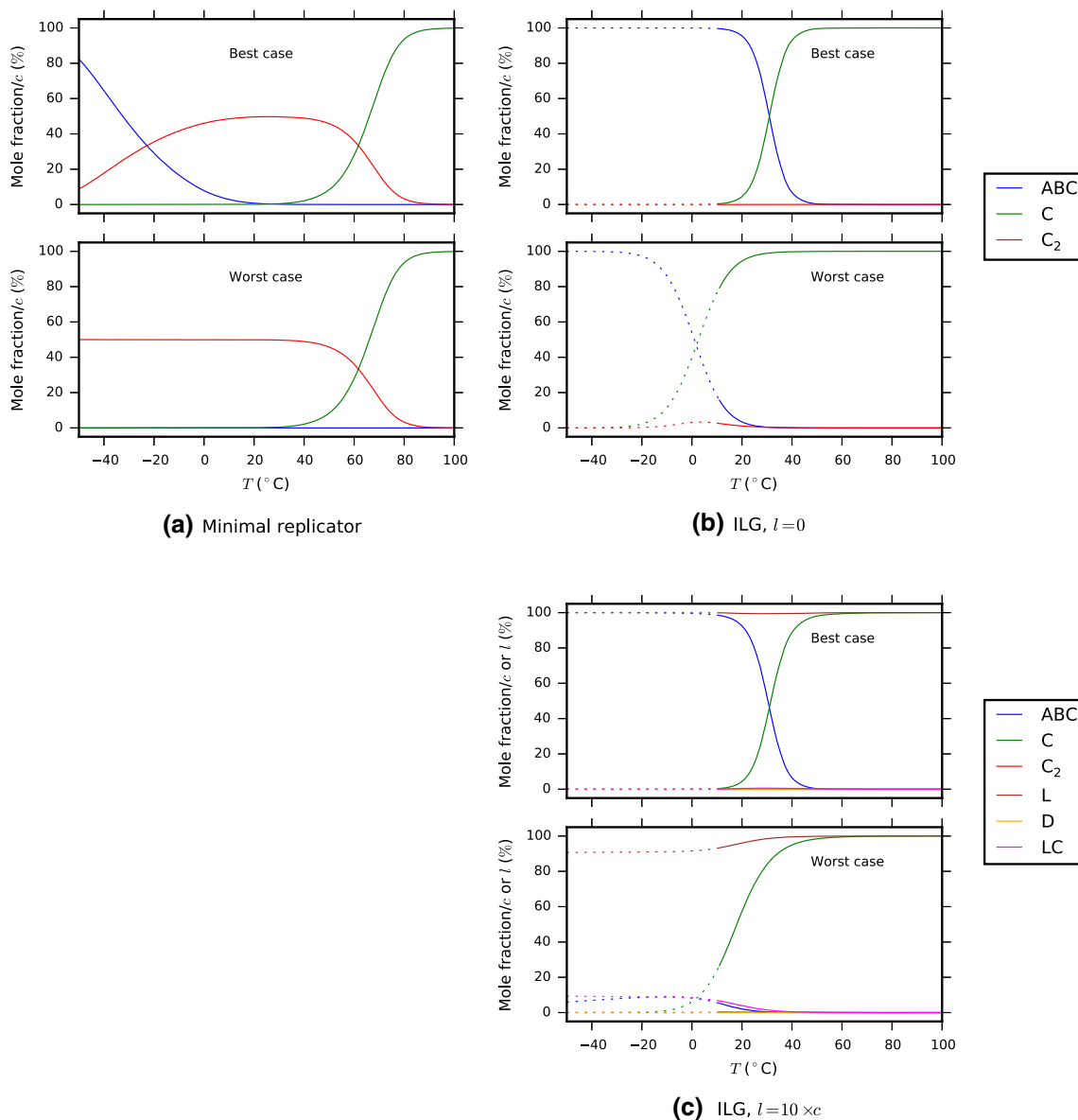
With AB complex formation

Fig. 9 Taking **AB** complex formation into account. Best case (*top*) and worst case (*bottom*) equilibrium distributions of the various species as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,^\circ$C). The *vertical scale* is the concentration of each species divided by the corresponding conserved quantity: $[\mathbf{ABC}]/c$, $[\mathbf{C}]/c$ and $[\mathbf{C_2}]/c$ for the minimal replicator model; $[\mathbf{ABC}]/c$, $[\mathbf{C}]/c$, $[\mathbf{C_2}]/c$, $[\mathbf{L}]/l$, $[\mathbf{D}]/l$ and $[\mathbf{LC}]/l$ for the ILG model

This advantage holds even when only considering temperatures above 284 K (11 °C) for the ILG model (solid lines in Figs. 5, 6 and 7, so as to ensure $K_2[\mathbf{B}]K_3[\mathbf{A}] \ll 1$, see end of "Chemical Reaction System" section). However, within this restricted temperature range, the maximum self-replication rate strongly depends on the estimated ΔG_1: it is significantly higher in the best case than in the worst case, because the maximum self-replication rate occurs very close to the lowest applicable temperature. Similarly, Fig. 5 shows that for the ILG model, strong exponential self-replication might only be observed close to this lowest temperature (freezing point of precursors into long chain-polymerized assemblies) and for the best case. For the worst case, the template resides mostly in the unbound **C** form in the applicable temperature range, which results in weak exponential self-replication.

This result is not a general property of the ILG model as seen in Fig. 2. Depending on the values of K_1 and $K = K_2 = K_3$, the autocatalytic order p may even turn

With AB complex formation

(a) Minimal replicator

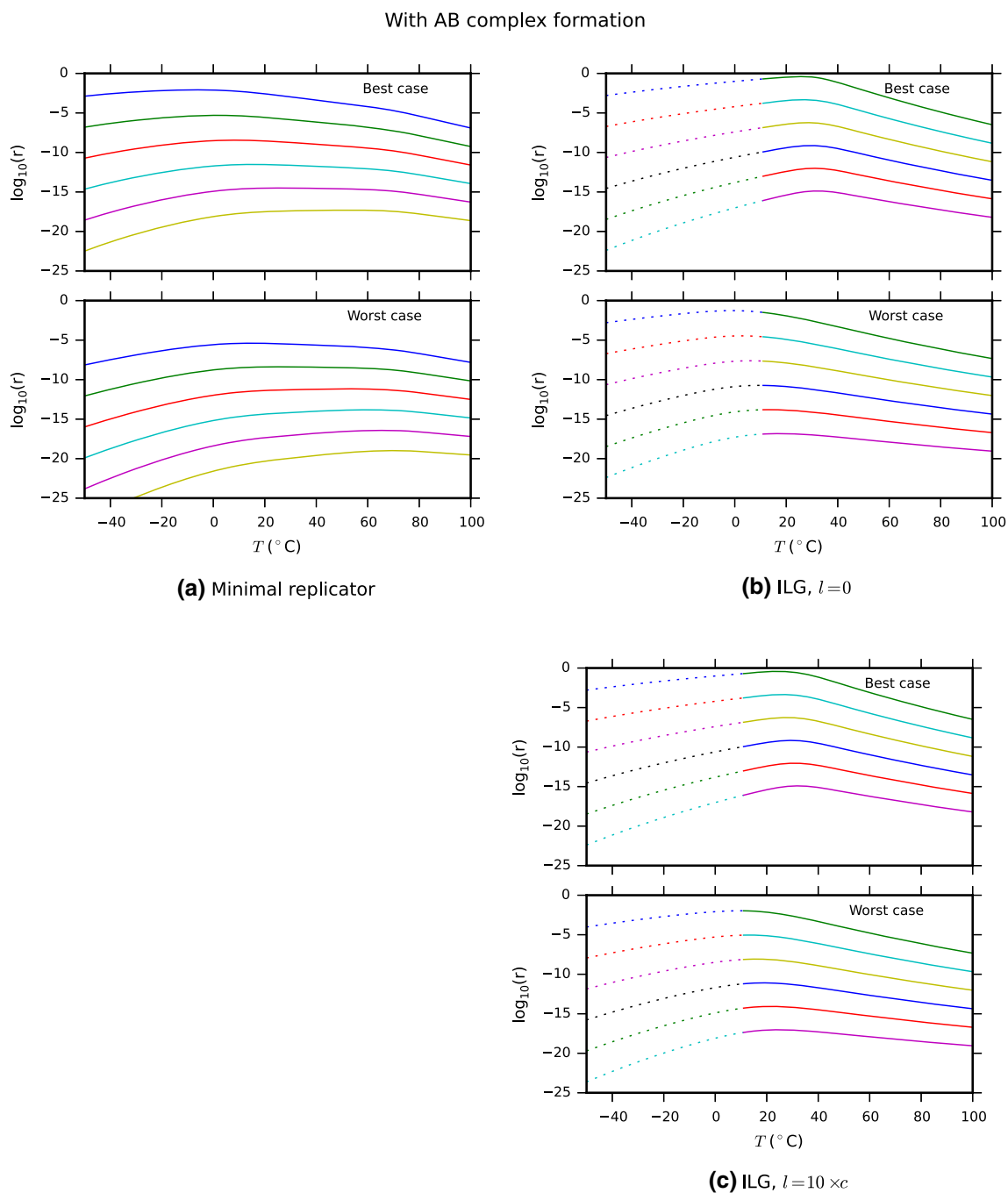**(b)** ILG, $l=0$

(c) ILG, $l=10 \times c$

Fig. 10 Taking **AB** complex formation into account. Best case (*top*) and worst case (*bottom*) decimal logarithm of the rate r of template formation as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l = 0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,°C$). The

different curves correspond to different activation energies E_a ranging from 10 to 30 kcal mol^{-1} in 4 kcal mol^{-1} increments. The multiplicative coefficient in Eq. 53 was arbitrarily taken as $A = 1 \times 10^{11}$ s^{-1} as in von Kiedrowski (1993)

negative, whereas it is always higher than 0.5 in the case of the minimal replicator model.

The reason why the ILG model is advantageous for the considered hexanucleotide example may be understood as follows.

From Eqs. (35) and (37), we have

$$[\mathbf{D}] = \frac{K_2 K_3 (q[\mathbf{ABC}])^2 l}{1 + K_3 q[\mathbf{ABC}] + K_2 K_3 (q[\mathbf{ABC}])^2} < \frac{K_2 K_3 (q[\mathbf{ABC}])^2 l}{1 + K_2 K_3 (q[\mathbf{ABC}])^2}.$$

$$(54)$$

From Eq. 36, we have $q[\mathbf{ABC}] < c$ which fed into the above equation gives:

With AB complex formation

(a) Minimal replicator
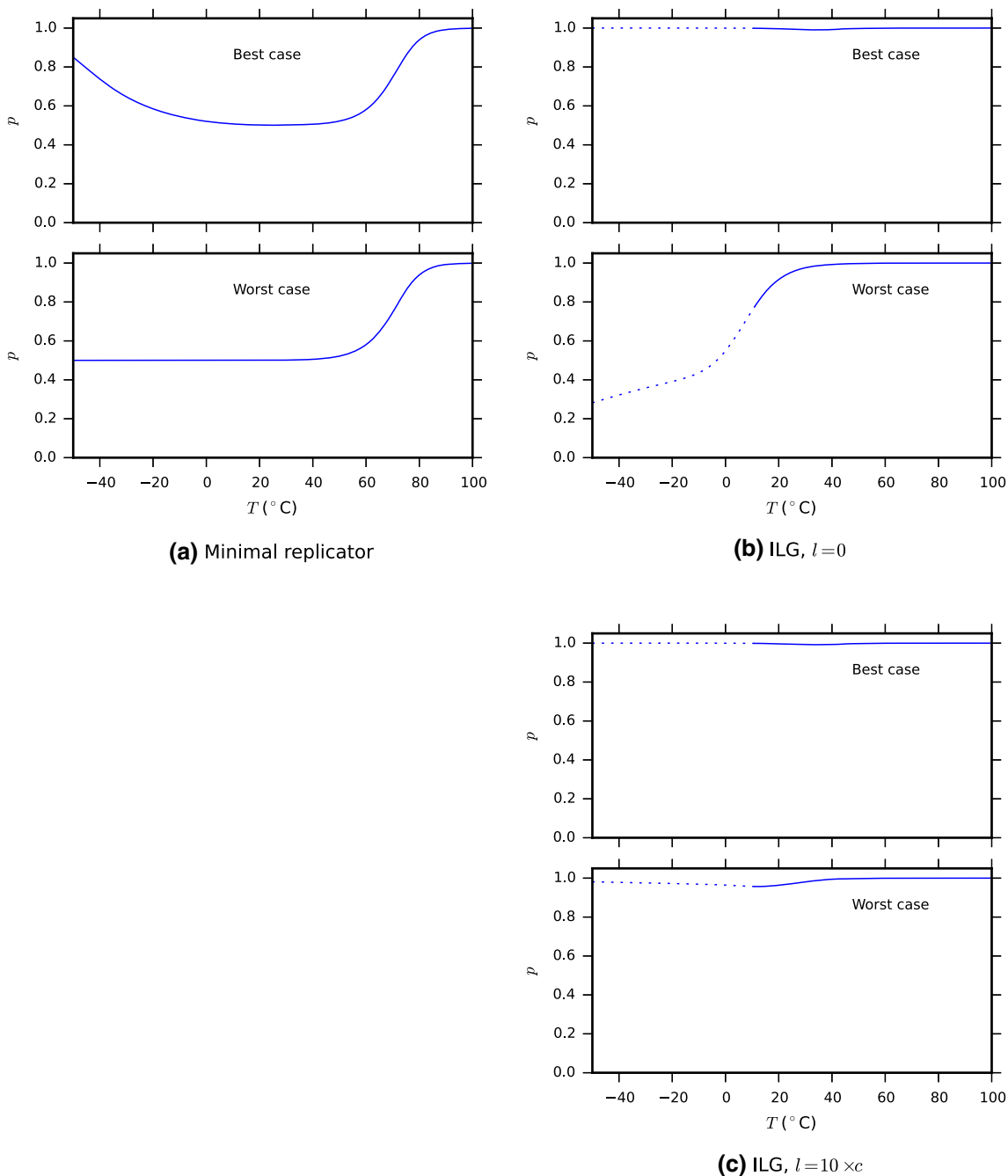
(b) ILG, $l=0$

(c) ILG, $l=10 \times c$

Fig. 11 Taking **AB** complex formation into account. Best case (*top*) and worst case (*bottom*) autocatalytic order p as a function of temperature, for the minimal replicator model (**a**) and for the ILG model with a total amount of leaving group of $l=0$ (**b**) and $l = 10 \times c$ (**c**). For the ILG scheme, *dotted lines* are used outside the applicable temperature range ($T < 11\,°C$)

$$[\mathbf{D}] < \frac{K_2 K_3 c^2 l}{1 + K_2 K_3 c^2} . \tag{55}$$

The contribution of $[\mathbf{D}]$ to Eq. 36 would thus be negligible if the following condition were verified:

$$\frac{K_2 K_3 c^2 l}{1 + K_2 K_3 c^2} \ll c. \tag{56}$$

This is indeed the case for the considered hexanucleotide example. At room temperature (T=298 K), we have $K_2 \approx$

$60\,M^{-1}$ and $K_3 \approx 360\,M^{-1}$. This gives $K_2K_3c^2 \approx 2 \times 10^{-4} \ll 1$ so that the above condition is equivalent to $K_2K_3cl \ll 1$. We have $K_2K_3cl \approx 2 \times 10^{-3}$ even in the extreme case where $l = 10 \times c$. With the contribution of $[\mathbf{D}]$ to Eq. 36 being thus negligible, the rate of template replication given by Eq. 41 approximately takes the same expression for the minimal replicator and ILG models. The key advantage of the ILG model arises from the significantly lower duplex stability (and associated lower K_2 equilibrium constant) because it only involves one $\mathbf{A} - \mathbf{B}$ bond instead of two for the minimal replicator model.

Taking **AB** Complex Formation into Account

For the minimal replicator model, there is only one such possible complex and the corresponding association constant K_0 is given in Table 1. In this case, a and b should be replaced by the expressions given by Eqs. (38) and (39).

For the ILG model, Fig. 8 reveals that there are two possible such complexes, with association constants K_2 and K_3 as given in Table 1. In this case, a and b should be replaced by the expressions given by Eqs. (38) and (39), replacing K_0 by $K_2 + K_3$.

Figures 9, 10 and 11 show how the equilibrium distribution of the different species, the rate r of template formation and the autocatalytic order p, respectively, vary with temperature, for the minimal replicator model (left) and for the ILG model (right).

As stressed out in von Kiedrowski (1993) for the minimal replicator model, the formation of complex **AB** makes it difficult to observe strong exponential growth which otherwise might be found at low temperatures. For the ILG model, **AB** complex formation does not result in any significant difference in the applicable temperature range.

Conclusion

We have proposed a self-replicating scheme based on an informational leaving group (ILG), inspired by the role of tRNA as leaving group and information carrier during protein synthesis. The potential advantage of such scheme is the weaker bonding of the template duplex and thus reduced template inhibition. We have carried out a theoretical analysis of this ILG scheme following the same approach as that of von Kiedrowski (1993) for the minimal replicator model, and have compared theoretical predictions for this scheme with those for the minimal replicator model. Although the autocatalytic order p may even turn theoretically negative for certain values of dissociation constants, when comparing this ILG scheme with the minimal replicator model for a hexanucleotide template

sequence mirroring that is used by von Kiedrowski in his original theoretical work (von Kiedrowski 1993), we have found that the ILG scheme was expected to outperform the minimal replicator model, with a higher replication rate and a higher autocatalytic order. Although this ILG scheme is expected to be sensitive to initial hydrolysis, this advantage is expected to hold even with significant hydrolysis in the initial mix. For the minimal replicator model, direct precursor-to-precursor complex formation should prevent strong exponential growth which otherwise might be found at low temperatures (von Kiedrowski 1993). For the proposed ILG model, the applicable temperature range is further restricted to temperatures above the freezing point of precursors into long chain-polymerized self-assemblies. Whether strong exponential growth might be observed in this applicable temperature range strongly depends on the estimated thermodynamic data for the template–precursors complex, for which experimental data are lacking. This stresses the need to confirm these theoretical predictions by future experimental work.

References

Bissette AJ, Fletcher SP (2013) Mechanisms of autocatalysis. Angew Chem Int Ed 52(49):12800–12826

Braich RS, Damha MJ (1997) Regiospecific solid-phase synthesis of branched oligonucleotides. Effect of vicinal 2′, 5′-(or 2′, 3′-) and 3′, 5′-phosphodiester linkages on the formation of hairpin DNA. Bioconj Chem 8(3):370–377

Carriero S, Damha MJ (2003) Template-mediated synthesis of lariat RNA and DNA. J Org Chem 68(22):8328–8338

Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157(1):77–94

Freier SM, Sinclair A, Neilson T, Turner DH (1985) Improved free energies for G.C base-pairs. J Mol Biol 185(3):645–647

Gilbert W (1986) Origin of life: the RNA world. Nature 319(6055):618

Hausner G, Hafez M, Edgell DR (2014) Bacterial group I introns: mobile RNA catalysts. Mob DNA 5:8

Issac R, Chmielewski J (2002) Approaching exponential growth with a self-replicating peptide. J Am Chem Soc 124(24):6808–6809

Joyce GF (1987) Nonenzymatic template-directed synthesis of informational macromolecules. Cold Spring Harb Symp Quant Biol 52:41–51

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell 31(1):147–157

Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382(6591):525–528

Li X, Chmielewski J (2003) Peptide self-replication enhanced by a proline kink. J Am Chem Soc 125(39):11820–11821

Mendel-Hartvig M, Kumar A, Landegren U (2004) Ligase-mediated construction of branched DNA strands: a novel DNA joining

activity catalyzed by T4 DNA ligase. Nucl Acids Res 32(1):e2–e2

Orgel LE (1992) Molecular replication. Nature 358(6383):203–209

Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA (1984) Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science 225(4665):898–903

Paul N, Joyce GF (2002) A self-replicating ligase ribozyme. Proc Natl Acad Sci 99(20):12733–12740

Peters JK, Toor N (2015) Group II intron lariat: structural insights into the spliceosome. RNA Biol 12(9):913–917

Roitzsch M, Fedorova O, Pyle AM (2010) The 2'-OH group at the group II intron terminus acts as a proton shuttle. Nat Chem Biol 6(3):218–224

Schreiber V, Dantzer F, Ame JC, De Murcia G (2006) Poly (ADP-ribose): novel functions for an old molecule. Nat Rev Mol Cell Biol 7:517–528

Song XP, Maiti M, Herdewijn P (2011) Enzymatic synthesis of DNA employing pyrophosphate-linked dinucleotide substrates. J Syst Chem 2(1):3

Sulston J, Lohrmann R, Orgel L, Miles HT (1968) Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. Proc Natl Acad Sci USA 59(3):726

Szostak JW (2012) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3(2):1–14

Vidonne A, Philp D (2009) Making molecules make themselves-the chemistry of artificial replicators. Eur J Org Chem 5:593–610

von Kiedrowski G (1986) A self-replicating hexadeoxynucleotide. Angew Chem Int Ed Engl 25(10):932–935

von Kiedrowski G (1993) Minimal replicator theory I: parabolic versus exponential growth. In: Bioorganic chemistry frontiers, Springer, New York, pp 113–146

von Kiedrowski G, Wlotzka B, Helbing J (1989) Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3'-5' pyrophosphate linkage. Angew Chem Int Ed Engl 28(9):1235–1237

von Kiedrowski G, Wlotzka B, Helbing J, Matzen M, Jordan S (1991) Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3'-5'-phosphoamidate linkage. Angew Chem Int Ed Engl 30(4):423–426

Zielinski WS, Orgel LE (1987) Autocatalytic synthesis of a tetranucleotide analogue. Nature 327(6120):346–347