Günter von Kiedrowski

Institute for Organic Chemistry, Georg-August-University, Tammannstraße 2, 37077 Göttingen, Germany

The paper describes an analytical treatment of artificial self-replicating systems. All artificial selfreplicating systems known today are minimal replicators in the sense that their kinetic behavior can be rationalized by a common, minimal reaction model which is outlined in the introduction. In the second section, empirical rate equations are introduced which have proven useful for the evaluation of experimental concentration-time profiles. The third section begins with an discussion of reaction models which have been described earlier to explain the autocatalytic synthesis of self-replicating template molecules. It is followed by an analytical treatment of the minimal reaction model: A + B $+ C \rightleftharpoons ABC \rightarrow C_2 \rightleftharpoons 2C$, where C is a self-complementary template molecule, A and B its precursor molecules, ABC a termolecular complex, and C_2 a template duplex. It is assumed that the irreversible formation of C_2 from ABC is the rate limiting step and that the total template concentration is small as compared to its precursors. The analytical expressions derived allow us to estimate the rate and autocatalytic reaction order for synthetic self-replicating systems from the elementary rate and equilibria constants involved. Three limit growth laws for minimal self-replicating systems-termed as parabolic, weak exponential, and strong exponential-can be distinguished. The following section deals with the influence of temperature. Strong exponential growth is to be expected for low temperatures, whereas weak exponential growth should occur at high temperatures. Parabolic growth is expected for average temperatures. Depending on the activation energy of the irreversible step as well as on the enthalpies of the formation of ABC and C_2 , the maximum of the autocatalytic rate occurs either at the temperature of the transition from strong exponential to parabolic growth, or, at the temperature of transition from parabolic to weak exponential growth, or, at an average temperature. The analytical results from the treatment of the above minimal reaction model are then compared to results from more realistic models. In particular, it is shown that the formation of a complex AB from A and B makes it difficult to observe strong exponential growth which otherwise might be found at low temperatures.

1	Introduction	i
2	On the Kinetic Behavior of Minimal Replicators	1
	2.1 Empirical Rate Equations and the Square Root Law	
	of Autocatalysis	1
	2.2 Nomenclature of Autocatalytic Growth Functions	
	2.3 Differences Between Parabolic and Exponential Growth 123	l
3	How to Explain the Kinetic Behavior?	
	3.1 Minimal Versus Full Modeling	
	3.2 Rapid Equilibria	
	3.3 Calculation of the Autocatalytic Reaction Order p 130	

Bioorganic Chemistry Frontiers, Vol. 3 © Springer-Verlag Berlin Heidelberg 1993

	3.4 Thermodynamic Requirements for Parabolic and																
	Exponential Growth																131
	3.5 Classification of Minir	nal	Re	epli	ica	tor	s .										133
4	The Effect of the Tempera	ture	e 01	n tl	ne	Equ	ili	bria	1 , A	lut	oca	ata	lyti	c			
	Rate and Reaction Order																135
5	Summary and Conclusions					•								·			143
6	References																145

1 Introduction

Recent years have seen considerable progress in the development of artificial self-replicating systems (review and overview articles: [1–8]). From a general point of view, self-replicating systems can be defined as autocatalytic reaction systems capable of passing on structural information [2, 9]. This requires a certain kind of autocatalyst, namely one which can act as a molecular template [10] in the autocatalytic event. Autocatalytic template molecules preorganize their precursor constituents in a defined spatial arrangement which allows them to transfer the structural information from the molecular template to the molecular copy [2, 4]. Thus, the term information is understood in a most elementary sense: to inform means to give a thing a form. Structural information can be stored as constitution, configuration, or even long-living conformation. Although these terms refer to the network of covalent bonds between atoms they may be extended to the network of covalent bonds between residues which replace the atoms as superatoms. The constitutional information of a linear network is given by the sequence of the residues. All informational [11] selfreplicating systems known so far are based on sequential templates which have a linear topology. Future developments may explore other topologies, for example, branched, cyclic, or two-dimensional arrays of residues. In any case, information transfer necessitates a scheme for molecular recognition allowing the constituents to be ordered on the template. Molecular recognition refers to weak, noncovalent interactions based on charge, shape, and solvent effects [12–14]. Binding is achieved by the attraction of opposite charges arranged on the surfaces of the shapes. In dyadic schemes for molecular recognition the interacting partner P^+ and P^- are thus necessarily complementary (P^+ binds P^{-}) although other schemes are conceivable, in which non-complementary partners may interact. For example, triadic schemes may utilize the arrangements $P^+ - C^- - P^+$, and $P^- - C^+ - P^-$ where C^+ and C^- are charged auxiliaries such as metal ions or anions.

The present design of artificial self-replicating systems is clearly inspired by nucleic acids as natural prototypes [1, 4, 7]. Logically, the search for artificial

Fig. 1. Minimal reaction model of the autocatalytic reaction cycle in synthetic self-replicating systems. K_1 and K_2 define the overall formation constants of the complexes **ABC** and **C**₂, respectively



replicators originated in the field of nucleic acid chemistry [5, 15, 16]. It was guided by the view of an RNA-world [9, 17–21] which became popular among molecular biologists [22–24] and bioorganic chemists [1, 25–27] after the seminal discovery of catalytic RNA [28, 29]. According to this view, life existed



a

Fig. 2a. Self-replicating hexadeoxynucleotides of von Kiedrowski et al. [32, 34, 36]. The autocatalytic templates shown are formed from two trideoxynucleotides: a trimer 3'-phosphate whose 3'phosphate group is activated in situ by the water soluble carbodiimide EDC, and another trimer which bears a nucleophilic group at its 5'-terminus. (b) Computer-generated model of the termolecular complex ABC of System I (standard B-DNA geometry). The attacking 5'-hydroxy group of the first trimer (indicated by a *small arrow*) is in close spatial proximity to the activated 3'-phosphate group of the second trimer



Fig. 2. (Continued)

on our planet before nature invented the genetic code and a translation machinery to synthesize proteins. Central to the view of an RNA-world is the hypothesis that nucleic acids are able to replicate in the absence of a proteinogenic replicase. This central hypothesis became the subject of experimental tests in many laboratories. Molecular biologists began to search for ribozymes acting as a primitive replicase [30, 31], and bioorganic chemists began to search for nonenzymatic template-directed condensation reactions leading to an autocatalytic oligonucleotide [1, 4-6]. The latter approach proved successful after reducing the dimension of the problem to a managable scale [32]. Following the design of the first real working example [32], all synthetic replicators so far published are based on two simplifications [33-39]. First, the number of necessary condensation steps to yield a template molecule has been cut down to a minimum of one. Second, and contrary to the natural prototype of nucleic acid replication, the template is self-complementary; it is structurally undistinguishable from its copy. In nucleic acid replication, the original strand acts as the template for the complementary copy. The copy strand in turn facilitates the synthesis of the original strand. Consequently, the synthesis proceeds in the sense of a cross-catalytic reaction here whereas the artificial templates are autocatalysts in the direct sense.

A minimal model which explains the autocatalytic synthesis in these systems is shown in Fig. 1 [32]: The self-complementary template C catalyzes its own formation from two complementary constituents A and B. A total of three steps is required to generate the autocatalytic cycle. In the first step, which is



Fig. 3a. Self-replicating tetraribonucleotide analogue of Zielinski and Orgel [33]. As above (Fig. 2), the bond formation is mediated by the water-soluble carbodiimide **EDC**. (b) Synthetic replicators of Rebek et al. [35, 37]. The templates shown are hybride molecules in the sense that they are composed of a nucleic acid constituent (a 5'-aminoadenosine derivative) and an artificial receptor for adenine (a derivative of Kemp's triacid)



Fig. 4. Termolecular complex of the fully synthetic replicator of Terfort and von Kiedrowski [39]. Molecular recognition is achieved by means of amidinium-carboxylate salt bridges which substitute complementary nucleobase pairing. The formation of the Schiff base proceeds reversibly. Note that System I is autocatalytic, whereas System II is catalytic. Very recently, Goodwin and Lynn described a similar example of a reversible template directed reaction based on a Schiff base formation [55]. Another related example was reported by Persico and Wuest [56].

reversible, the template C binds its constituents A and B to yield a termolecular complex ABC. In this complex, the reactive ends of the precursors are held in close spatial proximity which makes it easier to form a covalent bond between them. In the second step, the termolecular complex ABC is irreversibly transformed into the duplex C_2 . In the third step, which again is reversible, the dissociation of the duplex C_2 leads to two template molecules C. Both of them can begin a new round of synthesis. The above scheme was originally applied to achieve non-enzymatic replication of a self-complementary hexadeoxynucleotide in 1986 (Figure 2a,b, System I) [32], later to oligonucleotide analogues (Figure 2a, System III and IV; Figure 3a, b) [33–37], and recently to template molecules in which the nucleobases were replaced by other functionalities for molecular recognition (Figure 4, System I) [39].

2 On the Kinetic Behavior of Minimal Replicators

2.1 Empirical Rate Equations and the Square Root Law of Autocatalysis

Typical for an autocatalytic reaction is a sigmoid (S-shaped) curve for the timecourse of the autocatalyst formation. Initial experiments on the formation of a self-complementary hexadeoxynucleotide from two trideoxynucleotides (Figure

G. von Kiedrowski

(2)

2a, System I) however did not reveal any sigmoid curve for the production of the hexameric template at all [32]. Nevertheless, an autocatalytic contribution to the formation of template molecules could be confirmed in an indirect way. In experiments, in which the template was added initially, an increase in the rate of template synthesis was observed [32]. Initial rates of template synthesis, r_0 , were determined from concentration-time profiles at early reaction times [32]. Data from a series of experiments – in which the initial concentrations of the template, c_0 , was varied while all other concentrations were kept constant – allowed an empirical relationship to be established [32]:

$$\mathbf{r}_0 = \alpha \sqrt{\mathbf{c}_0 + \beta} \tag{1}$$

Equation (1) means that the template **C** is formed from **A** and **B** via two pathways: an autocatalytic pathway which is characterized by the empirical constant α , and a non-autocatalytic, viz. template-independent bimolecular pathway which is measured with the empirical constant β . The latter accounts for the initial rate of template synthesis in the absence of template. According to Eq. (1) the rate of autocatalytic template synthesis follows the square-root of the initial template concentration (and not the initial template concentration itself). In other words, the autocatalytic reaction order is not 1, but 1/2 ($\sqrt{x} = x^{1/2}$). This peculiar rate law was named the square-root law of autocatalysis [32].

More recently, another empirical rate equation was used to evaluate concentration-time profiles of self-replicating systems [36]:

$$\frac{dc}{dt} = (a_0 - c)(b_0 - c)[k_a(c + c_0)^P + k_b]e^{-k_c t}$$

Here, dc/dt describes the rate of template synthesis, whereas c contains the concentration of template molecules that have been synthesized at the reaction time t. a₀, b₀, c₀ are the initial concentrations of A, B, and C, respectively. Note that the concentrations given in Eq. (2) do not reflect the equilibrium concentrations of the corresponding species. The latter cannot be measured in these experiments [40]. Instead, the expressions $a_0 - c$, $b_0 - c$, and $c + c_0$ refer to measurable total concentrations. For example, $c + c_0$ refers to the sum of the concentrations of C, ABC, and C_2 , whereby the concentration of the latter is taken twice, Implicit in Eq. (2) is the mass-balance between A, B, and C: Each molecule of A and B which is consumed in the autocatalytic or non-autocatalytic pathway adds to the pool of existing template-molecules C. The implication of mass-balance restricts the application of Eq. (2) to those systems, in which A and B react solely to the template C. In other words, side reactions which lead to any other consumption of A and B are forbidden if Eq. (2) is chosen for the kinetic evaluation. This restriction however does not apply to Eq. (1). The apparent rate constants k_a and k_b refer to the autocatalytic and the nonautocatalytic formation pathway, respectively. The term apparent means that these constants are complex quantities which comprise elementary rate and equilibrium constants as well as environmental factors (pH, ion strength) and

the concentrations of other species (i square-root law, in which the reaction constants k_a and k_b are connected w follows:

$$\alpha = k_a a_0 b_0$$
$$\beta = k_b a_0 b_0$$

The exponential factor and its appar first-order (or pseudo first-order) de although it does not have any effect or allows to correct for hydrolysis effec based on condensation reactions in w sing agent whose hydrolysis is catalyz When studying the self-replication of organic phosphate which quickly re to form a carbodiimide-adduct, A* [3 antly hydrolyzed to regenerate A. So low degrees of activation (meaning the degrees of conversion (from A^* to C), approximated as a pseudo-first order sidered to be a pre-activated precurso been freshly synthesized prior to the account for its hydrolysis. To preserve the quantification method must not non-hydrolyzed form of A. For efficient hydrolysis is negligible on the time so factor can be omitted.

In summary, if one deals with a sel observable product, it is recommended without) the exponential factor. As "empirical" gross model function such derived are more reliable than those the complete mechanism. This is due constants, i.e. K1 in Fig. 1, cannot b have to be chosen more or less arbit values of other rate constants to be de under which mechanistic circumstan equations based on it are valid. One the experimentalist an idea what the e for the case of a square-root law comp for the reaction scheme in Fig. 1. The classification scheme, which allows us of the internal equilibria involved.

the concentrations of other species (i.e. condensing agents). For the case of a square-root law, in which the reaction order p is set to 1/2, the apparent rate constants k_a and k_b are connected with the empirical constants α and β as follows:

$$\alpha = k_a a_0 b_0 \tag{3}$$

$$\beta = k_b a_0 b_0 \tag{4}$$

The exponential factor and its apparent constant k_e in Eq. (2) accounts for a first-order (or pseudo first-order) decay of a species, which affects the rate although it does not have any effect on the mass balance between A, B, and C. It allows to correct for hydrolysis effects when studying self-replicating systems based on condensation reactions in water [36]. The decay of an excess condensing agent whose hydrolysis is catalyzed by A can be taken into consideration. When studying the self-replication of oligonucleotides, for example, A is an organic phosphate which quickly reacts with a water-soluble carbodiimide to form a carbodiimide-adduct, A* [32-34, 36]. The latter in turn is predominantly hydrolyzed to regenerate A. Sometimes, A* reacts with B to give C. For low degrees of activation (meaning that A is always in excess of A*) and low degrees of conversion (from A* to C), the decay of the condensing agent can be approximated as a pseudo-first order process [36]. Alternatively, if A is considered to be a pre-activated precursor (i.e. an active azolide or ester which has been freshly synthesized prior to the experiment), the exponential factor may account for its hydrolysis. To preserve the mass-balance in this particular case, the quantification method must not distinguish between the hydrolyzed and non-hydrolyzed form of A. For efficient self-replicating systems in which the hydrolysis is negligible on the time scale of the measurements, the exponential factor can be omitted.

In summary, if one deals with a self-replicating system in which C is the only observable product, it is recommended to use Eq. (2) with or without (if possible without) the exponential factor. As one will see later, the advantage of an "empirical" gross model function such as (2) is that the apparent rate constants derived are more reliable than those which can be derived from simulations of the complete mechanism. This is due to the fact that some rate and equilibrium constants, i.e. K₁ in Fig. 1, cannot be measured directly. Instead, their values have to be chosen more or less arbitrarily which in turn affects the numerical values of other rate constants to be determined. One should be aware, however, under which mechanistic circumstances a square-root law and empirical rate equations based on it are valid. One objective of this paper is therefore to give the experimentalist an idea what the experimental time courses should look like for the case of a square-root law compared to other rate laws which are possible for the reaction scheme in Fig. 1. The other objective is to provide a simple classification scheme, which allows us to predict the rate law from an inspection of the internal equilibria involved.

2.2 Nomenclature of Autocatalytic Growth Functions

The autocatalytic reaction order p in rate Eq. (2) determines the type of the autocatalytic growth curve. As it will be shown later, there are basically only two autocatalytic reaction orders which can be expected for self-replicating systems which work as depicted in Fig. 1. The first case is p = 1, the second p = 1/2. Note that these are extreme cases; transient reaction orders, 1/2 , may be found in real systems as well. Let us now consider a pure autocatalytic synthesis of template**C**at an very early stage of the reaction where the consumption of**A**and**B**is negligible. We may rewrite Eq. (2) in the form:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \alpha c^{\mathrm{p}} \tag{5}$$

Integration of Eq. (5) in the time interval between 0 and t and the corresponding concentration interval between c_0 and c gives:

$$\int_{c_0}^{c} \frac{dc}{c^p} = \int_{0}^{t} \alpha dt$$
(6)

Note that the elementary integrals which must be chosen for the left side of Eq. (6) are case-dependant. For the case p = 1, dc/c is integrated to give ln c, whereas in the case p = 1/2, the differential equation $\frac{dy}{dx} = x^n$ yielding the elementary integral $\frac{1}{n+1}x^{n+1}$ is the prototype. It follows from Eq. (6): p = 1: lnc - lnc₀ = α t (7)

$$p = 1/2; \quad \sqrt{c} - \sqrt{c_0} = \frac{\alpha}{2}t \tag{8}$$

From Eqs. (7) and (8), the concentration c can be expressed as a function of the time t:

$$\mathbf{p} = 1; \quad \mathbf{c} = \mathbf{c}_0 \mathbf{e}^{\alpha t} \tag{9}$$

$$p = 1/2: \quad c = \left(\sqrt{c_0} + \frac{\alpha}{2}t\right)^2 \tag{10}$$

It becomes evident that Eq. (9) describes an exponential growth function. - Equation (10) yields, for infinitesimal small concentrations c_0 , the growth function:

$$c = \frac{\alpha^2}{4} t^2, \qquad (11)$$

according to which the concentration increases with the square of the time. Since the graph of function (11) shows a parabola (parabola equation: $y = mx^2$),

autocatalytic growth based on the square-root law has been recently termed parabolic growth [1, 36]. Again, the growth functions (9)–(11) approximate the autocatalytic growth curves only at a very early stage of the reaction since the consumption of the precursors is not taken into consideration. "Real" autocatalytic growth curves are typically (but not necessarily) sigmoid which is the consequence of the decrease of the precursors.

2.3 Differences Between Parabolic and Exponential Growth

Parabolic growth is always a "weaker" type of autocatalytic growth when compared to exponential growth. This is illustrated in Fig. 5 for the case of pure autocatalytic syntheses. The dashed curves account for the ideal case of unlimited growth; they represent the graphs of the functions (9) and (10). The solid curves refer to the case of autocatalytic growth which is limited by the consumption of the precursors. Both solid curves were generated using the integrated form of Eq. (2) in which k_b and k_c were set to zero. Both curves start at the same initial concentration c_0 with the same initial slope, thus, simulating equal initial rates of template synthesis. To account for the latter, the apparent constants k_a were set as follows: k_a (parabolic) = $\sqrt{c_0}k_a$ (exponential). As it is evident from Figure 5, the change of the autocatalytic reaction order p produces a dramatic effect on the S-shape of the curves. The maximum of the autocatalytic rate occurs at the point of inflection at which the second derivative of the concentration as the function of the time becomes zero:

$$\frac{\mathrm{d}^2 \mathbf{c}}{\mathrm{d}t^2} = 0 \tag{12}$$

Generally, the autocatalyst concentration c_i at the point of inflection is calculated from the first derivative of the corresponding rate equation [41]. c_i is then used in the integrated rate equation to calculate the time of inflection t_i . When applying condition (12) to Eq. (2), however, no explicit analytical equation for c_i can be deduced. We have therefore employed a numerical approximation technique to calculate the inflection points shown in the following figures. Figure 5 illustrates, that the inflection concentration c_i for the case of parabolic growth (b) is typically smaller than in the case of exponential growth (a).

In real experimental systems however, the autocatalytic synthesis of templates C is never completely "pure". Figure 6a shows in the exponential and Figure 6b in the parabolic case how the growth curves respond to an increase of k_b which accounts for the nonautocatalytic formation of template molecules. The ratio:

$$\varepsilon = \frac{k_a}{k_b},\tag{13}$$

C₀



Fig. 6a, b. Dependancy of the growth curves from the autocatalytic efficiency ε for (a) p = 1, and (b) p = 1/2. All curves were generated by integration of Eq. (2) using the conditions: $a_0 = b_0 = 10^{-3}$ M, $c_0 = 0$, $k_c = 0$, $k_a = 10$, $k_b = k_a/\epsilon$. Dots indicate points of inflection. Note that in the case of parabolic growth (b), the growth curves for $\epsilon = 10^4 - 10^6$ overlap within the thickness of the curves

measures the autocatalytic efficiency, it decreases with an increase of k_b . Again, Eq. (2) was used to generate the set of curves. Both, in Fig. 6a and in Fig. 6b, the S-shape becomes visible only above a critical threshold value of ϵ . This explains why, in some experimental systems, a sigmoid production of the template could not be detected although an autocatalytic pathway was established [32–35]. In order to be able to observe a sigmoid curve in an experimental system, ϵ should be above 10^2 for p = 1/2 or above 10^4 for p = 1 if one accounts for the uncertainities of experimental profiles. There is a striking difference between the corresponding growth curves when ϵ increases towards large values. In Fig. 6b, the growth curves become more and more similar finally approaching the curve for a pure autocatalytic synthesis whereas in Fig. 6a, the growth curves are steadily shifted towards longer induction periods. In other words, for the case p = 1/2, the inflection time t_i becomes finite ϵ . The experimental consequence of this





fact is that parabolic replicators behave almost "purely" at values of ε which are experimentally accessible. The latter in turn makes it difficult to differentiate between an efficient and a very efficient replicator. The opposite applies to exponential replicators. Here, attempts to increase the autocatalytic synthesis at the cost of the non-autocatalytic counterpart will be surely repaid the outcome of the experimental curves.

From an organic chemist's point of view, the easiest way to detect a parabolic or exponential replicator is to follow the formation of template molecules in the presence of increasing initial concentrations c_0 at an early stage of the reaction (Fig. 7a, b). Typically, four experiments are performed in which c_0 is varied as follows: $c_0(1) = 0$, $c_0(2) = y$, $c_0(3) = 2y$, $c_0(4) = 4y$, where y is typically between 2–10% of the initial concentration of the precursors [32–34, 36, 39]. Note that from the second experiment, the initial concentrations c_0 are doubled when proceeding to the next experiment. In the exponential case (Fig. 7a), the concentration differences c = c(i) - c(1), measured at any early time, approximately reflect the ratio 1:2:4 whereas in the parabolic case (Fig. 7b), the corresponding ratio is given by $1:\sqrt{2}:\sqrt{4}$.

3 How to Explain the Kinetic Behavior?

3.1 Minimal Versus Full Modeling

In the previous section, it was demonstrated that a simple empirical rate equation, Eq. (2), can be used to generate a whole variety of autocatalytic concentration-time profiles. The question arises how the autocatalytic reaction p depends on the concentrations, rate- and equilibrium constants involved. In the following sections, it will be shown how the reaction order p can be calculated for systems in which the transition from the termolecular complex **ABC** to the template duplex C_2 is the rate determining step of the cycle. The assumption of rapid equilibria holds for all experimental systems so far described. According to Fig. 1, the autocatalytic synthesis is modeled by a total of three steps:

$$\mathbf{A} + \mathbf{B} + \mathbf{C} \rightleftharpoons \mathbf{ABC} \tag{14}$$

$$2C \equiv C_2 \tag{15}$$

$$ABC \to C_2 \tag{16}$$

We refer to Eqs. (14)–(16) as a minimal model for the reactions taking place in the autocatalytic channel [32]. In contrast, a full model for the autocatalytic

synthesis accounts for all possible transient complexes. Here, the net equation, Eq. (14), becomes resolved into:

$$\mathbf{A} + \mathbf{C} \rightleftharpoons \mathbf{A}\mathbf{C} \tag{17}$$

$$\mathbf{B} + \mathbf{C} \rightleftharpoons \mathbf{B}\mathbf{C} \tag{18}$$

$$AC + B \rightleftharpoons ABC \tag{19}$$

$$BC + A \rightleftharpoons ABC \tag{20}$$

Nowick et al. modeled Rebeks self-replicating system of 1990 [42]. The reaction model consisted of Eqs. (15–20) which account for the autocatalytic reaction cycle whereas the synthesis in the nonautocatalytic reaction channel was described by:

$$\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{A}\mathbf{B} \tag{21}$$

$$\mathbf{A} + \mathbf{B} \to \mathbf{C} \tag{22}$$

$$AB \rightarrow C$$
 (23)

Nowick et al. assumed rapid equilibria for the reversible reactions expressed by Eqs. (15), and (17)–(21) [42]. Theoretical concentration-time profiles were computed by numerical integration of the set of differential equations derived from Eqs. (16), (22) and (23) [42]. Prior to an integration step, the equilibrium distribution of all species was calculated iteratively [42]. Although some insights into the dependancy of the growth curves from the initial concentrations and the rate constants of the irreversible steps were presented, no general explanation was given how the growth curves and the autocatalytic reaction order depend on the internal equilibria [42]. It was nevertheless stated that the experimental kinetic data fit Eq. (2) quite well if a square-root law of autocatalysis (p = 1/2) is assumed [42].

Similar modeling studies were performed in 1985 for the full reaction model of a self-replicating hexadeoxynucleotide [43]. In contrast to the method employed by Nowick et al., the computation of the rapid equilibria was not separated from the numerical integration. Instead, the set of differential equations for both, reversible and irreversible steps was integrated using a solver for stiff differential equations as described by Gear [44]. These simulations allowed us to evaluate the conditions that lead to a square-root law. Later simulations confirmed that, for the autocatalytic reaction channel, the minimal model as described by Eqs. (14)–(16) is indeed sufficient to derive a reasonable value for the autocatalytic reaction order p in many cases. This was expected from thermodynamic considerations. As a consequence of the latter, the differential equations describing the reaction model could be further compressed into a single "empirical" rate equation, Eq. (2).

The advantage of such dynamic simulations is that they can be performed with any conceivable full reaction model and that the numerical concentration values from which p can be computed are rather precise. However, dynamic

simulations do not reveal any analytical expression—which lets one understand the internal dependencies—by themselves. Insight derived from simulations is always based on numbers and not on the relationships which lead to those numbers. The latter may be found from an inspection of the computed data. The opposite ("classical") approach is to derive the theoretical relationships first and then to compute some numbers which may illustrate the relationships. This approach, however, necessitates that the reaction model is minimal – at least for the case of self-replicating systems and at least if interpretable explicit equations are to be derived.

In the following, the fundamental insights into the dynamics of a self-replicating system come from an analytical treatment of the minimal reaction model as described by Eqs. (14)–(16). The results of this analytical treatment are later compared with those which can be computed numerically for more realistic reaction models.

3.2 Rapid Equilibria

The calculation of the autocatalytic reaction order p makes it necessary to express the equilibrium concentration of the termolecular complex, [ABC], as a function of the total concentrations and the equilibrium constants. Let [A], [B], [C], [ABC], and [C₂] be the equilibrium concentrations of the corresponding species. The total concentrations of molecules **A**, **B**, and **C** are approximated by:

$$a = [A] + [ABC] \cong [A]$$
⁽²⁴⁾

$$\mathbf{b} = [\mathbf{B}] + [\mathbf{ABC}] \cong [\mathbf{B}] \tag{25}$$

$$c = [C] + [ABC] + 2[C_2]$$
(26)

In Eqs. (24) and (25) it is assumed that the relative contribution of [ABC] to the total concentrations a and b is negligible. This is always the case if the total template concentration c is small compared to the precursor concentrations a and b. The assumption is valid for the initial stage of autocatalytic growth as well as for experiments in which the initial rate of template synthesis is monitored in the presence of small concentrations of added template. Note that the concentrations a and b do approximate the total concentrations of the precursors **A** and **B** only for the case that **A** and **B** do not interact. If the latter interaction is taken into account (cf. Eqs. (34)–(35)), a and b refer to unbound precursor concentrations.

The overall formation constants of the complexes ABC and C_2 are:

$$K_{1} = \frac{[ABC]}{[A][B][C]}$$

$$K_{2} = \frac{[C_{2}]}{[C]^{2}}$$
(27)
(28)

With

$$q = \frac{1}{K_1 ab}$$
(29)

equations (27) and (28) allow us to calculate [C] and [C₂] from [ABC]:

$$[C] = \frac{[ABC]}{K_1[A][B]} \cong [ABC]q \tag{30}$$

$$[C_2] = K_2[C]^2 \cong K_2[ABC]^2 q^2$$
(31)

[C] and $[C_2]$ in Eq. (26) are substituted by expressions (30) and (31), respectively. This leads to the quadratic equation (32) whose physical meaningful solution is expression (33):

$$[ABC]^{2} + \frac{1+q}{2K_{2}q^{2}}[ABC] - \frac{c}{2K_{2}q^{2}} = 0$$
(32)

$$[ABC] = \frac{\sqrt{8K_2cq^2 + (1+q)^2 - q - 1}}{4K_2q^2}$$
(33)

Equation (33) implies that no further equilibria except those expressed with K_1 and K_2 exist. Additional equilibria must be considered explicitly. It is very likely, for example, that A and B form a complex AB since A and B are complementary. Logically, the same recognition principle which allows the formation of the termolecular complex ABC will allow an interaction leading to a precursor complex AB. If this complex is formed, the quantities a and b no longer account for the approximate total concentrations, but for the unbound concentrations of A and B which are now unknown. The latter concentrations, however, can be estimated from the known total concentrations c_A and c_B assuming that AB has the formation constant K_0 :

$$a = \frac{K_0(c_A - c_B) - 1 + \sqrt{K_0^2(c_A - c_B)^2 + 2K_0(c_A + c_B) + 1}}{2K_0}$$
(34a)

$$\mathbf{b} = \mathbf{c}_{\mathbf{B}} - \mathbf{c}_{\mathbf{A}} + \mathbf{a} \tag{34b}$$

From Eqs. (34a) and (35a), the following expression which replaces expression (29) can be derived:

$$q = \frac{2K_0^2}{K_1 \left[1 + K_0(c_A + c_B) - \sqrt{1 + 2K_0(c_A + c_B) + K_0^2(c_A - c_B)^2}\right]}$$
(35)

Another possibility is the formation of hairpins [45]. If, for example, the template C forms a hairpin C' having the formation constant K_3 , the variable q in Eq. (33) must simply be replaced by the expression $q \cdot (1 + K_3)$ in order to derive a proper value for [ABC]. In any case, Eq. (35) holds for situations, in which [ABC] is small compared to c_A and c_B . If this is no longer valid, or, if more

coupled equilibria must be considered, it is recommended that [ABC] be computed by numerical approximation techniques instead. It should be mentioned that Eq. (33) does not necessarily reflect the real equilibrium concentration of the termolecular complex since intermediate complexes such as **AC** and **BC** are not considered. The calculation scheme implies a cooperative formation of the complex **ABC** which means that the concentrations of the intermediate complexes **AC** and **BC** can be neglected. This particular simplification however does not change the autocatalytic reaction order to a significant extent, as long as [ABC] is small compared to c_A and c_B .

3.3 Calculation of the Autocatalytic Reaction Order p

Each autocatalytically synthesized template molecule which adds to the concentration pool c is formed via the termolecular complex **ABC**. The rate of autocatalytic template synthesis is therefore given by:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = k[\mathrm{ABC}],\tag{36}$$

where k is the elementary rate constant of the irreversible step. At early reaction times, the autocatalytic reaction order p template is implicitly defined in the context of Eq. (5). An explicit equation for the reaction order p is given after "logarithmization" of Eqs. (5) and (36):

$$p = \frac{dln\left(\frac{dc}{dt}\right)}{dln(c)} = \frac{dln[ABC]}{dln(c)}$$
(37)

According to Eq. (37) the reaction order p does not depend on the rate constant k of the irreversible step. This is always the case if the latter is rate-determining. In order to find an analytical expression for p, one has to remember that $d\ln(c) = \frac{1}{c}dc$. Differentiation of Eq. (33) according to (37) needs concatenation, $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx}$, substituting $\ln[ABC]$ as y, $\ln(c)$ as x, the numerator of Eq. (33) as u, and the expression under the square-root of the numerator as v. It follows:

$$p = \frac{4K_2cq^2}{8K_2cq^2 + (1+q)^2 - (1+q)\sqrt{8K_2cq^2 + (1+q)^2}}$$
(38)

The practical value of Eq. (38) is to predict p from K_1 and K_2 for a given set of concentrations a, b, c. This is especially useful if the initial rate of template formation is measured as a function of the template concentration added initially. From these measurements, an experimental reaction order can be

determined which may be compared with the predicted value for p. A significant discrepancy between those numbers may indicate that either the kinetic data or the thermodynamic data are erroneous. Again, one has to be aware that the reaction order defined with Eq. (38) is valid for the condition that the total template concentration is small compared to the total concentration of the precursors. When applied to Eq. (2) the latter condition reads: $(a_0 - c)$, $(b_0 - c) \ge (c + c_0)$. This restricts the comparison to an early stage of the reaction and to those experiments in which the initial template concentrations, c_0 , are small compared to the initial precursor concentrations a_0 and b_0 . As we will understand later, the constraint turns out to be less demanding for the case of parabolic replicators.

3.4 Thermodynamic Requirements for Parabolic and Exponential Growth

Equation (38) describes the reaction order p as a function of five variables:

$$p = f(a, b, c, K_1, K_2)$$
(39)

For a fixed set of concentrations, a, b, c, the reaction order p is a function of just two variables: K_1 and K_2 . Self-replication experiments with C/G-rich hexadeoxynucleotides are typically carried out at millimolar concentrations of the trimers **A** and **B**. Growth of template **C** is usually followed at templateconcentrations between $c = 1 - 100 \mu mol/l$. Figure 8a shows for a typical set of concentrations ($a = b = 10^{-3}$ M (unbound concentrations, cf. Eqs. 34a, b), $c_1 = 5 \times 10^{-5}$ M, $c_2 = 1 \times 10^{-4}$ M) how the reaction order p depends on the complex formation constants K_1 , and K_2 .

It becomes evident that a square root law leading to parabolic growth (p = 1/2) is to be expected for systems in which the duplex C_2 is sufficiently stable $(K_2 > 10^6 \text{ M}^{-1})$ whereas the termolecular complex ABC must be sufficiently less stable than C_2 . Figure 8b shows the corresponding contour plot. Towards high values of K_2 and K_1 the contour lines run parallel following the equation:

$$\log K_1 = \frac{\log K_2}{2} + s,$$
 (40)

where s is the intercept of these lines at the log K₁-axis. To keep p below 0.55, for example, s must be below 3.4 (cf. line a). To keep p above 0.95, s must be above 4.7 (cf. line b). Multiplying equation (40) by the factor -2.303 RT gives:

$$\Delta G_1 = \frac{\Delta G_2}{2} - 2.303 \text{sRT},\tag{41}$$

where ΔG_1 , and ΔG_2 are the Gibbs free enthalpy changes upon formation of **ABC** and **C**₂, respectively. Let us now assume that the difference in the free enthalpy between these complexes is solely based on entropy. This means that



Fig. 8a. Surface and (b) contour plot of the autocatalytic reaction order p as a function of the decadic logarithm of K₁ and K₂. The values of p were calculated with Eq. (38) for $a = b = 10^{-3}$ M, $c = 10^{-4}$ M. The lines a and b in Fig. 8b represent Eq. (40) for s = 3.4, and s = 4.8, respectively. Line c represents Eq. (42) divided by -2.303 RT

the enthalpy of base-stacking across the reaction site (nick) in the termolecular complex is comparable to the duplex. In other words, it is assumed that **ABC** and **C**₂ adopt similar conformations. Note that the formation of **C**₂ requires one nucleation (helix initiation) step whereas **ABC** requires two nucleations. The entropy which accounts for the nick in **ABC** is thus given by the entropy change ΔS_{nuc} of the additional nucleation step:

$$\Delta G_1 = \Delta G_2 - T\Delta S_{nuc} \tag{42}$$

A typical value for ΔS_{nuc} of C/G-rich oligonucleotides is -12.6 e.u. [46]. The line c in Fig. 8b—representing Eq. (42) divided by -2.303RT—is based on this

value. The intersections of this line with the lines a and b can be calculated from Eqs. (41) and (42):

$$\Delta G_2 = 2T(\Delta S_{nuc} - 2.303 \text{sR}) \tag{43}$$

Equation (43) can now be used to estimate the free enthalpy of duplex formation that is needed to achieve a certain kind of autocatalytic growth at the given set of concentrations. The required free enthalpy change for the formation of ABC is then calculated from Eq. (42). Autocatalytic growth of the order p > 0.95(almost exponential), for example, should be observable at T = 298 K, if $-\Delta G_1$ is at least 19 kcal/mol ($-\Delta G_2 = 20.3$ kcal/mol). Smaller complex stabilities of ABC are required for parabolic growth (p < 0.55): At T = 298 K, $-\Delta G_1$ should be no larger than 15.5 kcal/mol ($-\Delta G_2 = 16.8$ kcal/mol). The required free enthalpies may be compared with those derived from the thermodynamic data of duplex-singlestrand transitions. The reversible formation of the hexadeoxynucleotide duplex (dGCGCGC)₂, for example, was shown to proceed with $\Delta H^{0} = -56.2 \text{ kcal/mol}$, and $\Delta S^{0} = -154 \text{ e.u.}$ at pH 7 and 1 M NaCl [47]. This corresponds to a change in free enthalpy of $-\Delta G^{0}_{298} = 10.3$ kcal/mol. According to Eq. (42), $-\Delta G_1$ at T = 298 K (formation of the termolecular complex ABC ($= dGCG \cdot dCGC \cdot dGCGCGC$)) is assumed to be 6.7 kcal/mol (log $K_1 = 4.9$) which is far below the maximum value allowing parabolic growth and even further below the minimum required for exponential growth at the given set of concentrations. Thus, C/G-rich hexadeoxynucleotides should exhibit parabolic growth at the given temperature and the given set of concentrations which indeed was found experimentally [36]. A more detailed analysis of the kinetic behavior of self-replicating hexadeoxynucleotides is the subject of Sect. 4.

3.5 Classification of Minimal Replicators

From an inspection of Eq. (33), three limit cases can be deduced, which allow us to classify the autocatalytic growth of synthetic replicators as parabolic, weak exponential and strong exponential. The conditions leading to the latter cases are listed together with the resulting rate laws in table 1. These limit cases become also evident from a triple-logarithmic plot of the autocatalytic rate, dc/dt, as a function of K_1 and K_2 at a given set of concentrations. Figure 9 illustrates this for the set of concentrations which were used to generate Fig. 8a, b. Thus, Figs 9 and 8a may be directly compared. The rates were calculated from Eqs. (33) and (36) using the rate constant k = 1. The overall surface shown in Fig. 9 can be divided into three sections indicated as se (= strong exponential), we (= weak exponential), and p (= parabolic). Each of the three sections approaches an asymptotic plane which is described by the logarithmic form of the corresponding limit rate law given in Table 1. The intersections between these planes define borderlines which refer to the conditions listed in

Table 1. Since each section is separated from each other by two borderlines, two conditions are required to separate each limit case from another.

In the case of strong exponential growth, the rate law neither depends on k_1 nor on K_2 . Each template molecule exists in the form of its termolecular complex **ABC** and an increase of the stability of the latter, for example, cannot increase its population. The case of strong exponential growth may thus be compared with the case of maximal velocity observed in enzyme kinetics when the enzyme becomes saturated with the substrate. The limit rate law of weak exponential growth does not depend on K_2 but does depend on K_1 . Template molecules may exist as termolecular complexes **ABC** or as unbound molecules **C** but not as duplexes **C**₂ here. An increase of K_1 shifts the population of termolecular complexes **ABC** towards higher densities. The limit rate law for

Table 1. Classification of autocatalysis for synthetic replicators at internal equilibrium. For systems in which the complex **AB** is involved, the expression K_1 ab has to be calculated with Eq. (35); K_1 ab $= q^{-1}$

1. Condition	2. Condition	Limit rate law	
$\sqrt{2 K_2 c} \gg K_1 ab$	$\sqrt{2 \text{ K}_2 \text{c}} \gg 1$	$\frac{dc}{dt} = \frac{kK_1ab\sqrt{c}}{\sqrt{2}K_2}$	
$\sqrt{2 \text{ K}_2 \text{c}} \ll 1$	$K_1 ab \ll 1$	$\frac{\mathrm{d}c}{\mathrm{d}t} = kK_1 \mathrm{abc}$	
$K_1ab \gg \sqrt{2K_2c}$	$K_1ab \gg 1$	$\frac{dc}{dt} = kc$	
	1. Condition $\sqrt{2 K_{2}c} \gg K_{1}ab$ $\sqrt{2 K_{2}c} \ll 1$ $K_{1}ab \gg \sqrt{2K_{2}c}$	1. Condition 2. Condition $\sqrt{2} K_2 c \gg K_1 ab$ $\sqrt{2} K_2 c \gg 1$ $\sqrt{2} K_2 c \ll 1$ $K_1 ab \ll 1$ $K_1 ab \gg \sqrt{2} K_2 c$ $K_1 ab \gg 1$	1. Condition2. ConditionLimit rate law $\sqrt{2} K_2 c \gg K_1 ab$ $\sqrt{2} K_2 c \gg 1$ $\frac{dc}{dt} = \frac{kK_1 ab\sqrt{c}}{\sqrt{2} K_2}$ $\sqrt{2} K_2 c \ll 1$ $K_1 ab \ll 1$ $\frac{dc}{dt} = kK_1 abc$ $K_1 ab \gg \sqrt{2K_2 c}$ $K_1 ab \gg 1$ $\frac{dc}{dt} = kc$



Fig. 9. Triple-logarithmic plot of the autocatalytic rate dc/dt as a function of K_1 and K_2 . The rates were calculated with Eqs. (36) and (33) for $a = b = 10^{-3}$ M, $c = 10^{-4}$ M and k = 1 day⁻¹. The grey lines above the surface indicate the intersections between asymptotic planes which account for the limit rate laws in Table 1. Abbreviations: se = strong exponential, we = weak exponential, p = parabolic

parabolic growth is a square-root law. Here, the rate increases with the stability of the termolecular complex, expressed by K_1 , whereas it decreases with the stability of the template duplex, since K_2 is in the denominator. Template molecules may exist either as termolecular complexes ABC or as duplexes C_2 but not (at least not to a significant amount) as unbound molecules C.

The conditions given in Table 1 allow to predict the type of autocatalysis from the quantities a, b, c, K_1 , and K_2 . This can be done for systems, in which K1 and K2 are either known from thermodynamic measurements, or estimated by comparison with similar equilibria. Rebek's first replicator, for example, was reported to have an dimerisation constant of $K_2 = 630 \text{ M}^{-1}$ (cf. Fig. 3b, system [] [35]. All other complexes—their formation being described by equations (17)-(21)-were assumed to have an formation constant of 60 M⁻¹ [35]. From the given numbers it follows that $K_1 = 60^2 = 3600 \text{ M}^{-2}$. The initial total concentrations of A and B were in the order of 0.01 M[35]. Using the reciprocal expression (35), $K_1ab = 0.18$. At a total template concentration of c = 0.001 M, which again is close to the experimental concentration, $\sqrt{2K_2c}$ (= 1.12) is close to 1. Table 1 reveals that under the given set of conditions, Rebek's replicator is expected to show a transient type of autocatalysis which is between the parabolic and the weak exponential case. Indeed, from Eq. (38), the reaction order p is predicted to be 0.74 at c = 1 mM. Higher values of p are expected at lower template concentrations: p = 0.93 at c = 0.1 mM, p = 0.99 at c = 0.01 mM. It should be mentioned that although the square-root law is putatively not applicable here, Eq. (2) with p = 1/2 gave a significantly better fit of the experimental curves as compared to evaluations based on higher values of p [36]. This discrepancy is not due to the limitations of the classification method given. Rebek's second replicator ($K_1 = 3600 \text{ M}^{-2}$, $K_2 = 80000 \text{ M}^{-1}$, $c_A = c_B$ = 0.050 M, cf. Figure 3b, system II) [37], however, can be classified as parabolic with certainty: $K_1ab = 0.03$; $\sqrt{2K_2c} = 13$ at c = 1 mM, $\sqrt{2K_2c} = 4$ at c = 10 mM. As to be expected, equation (2) based on the square-root law gave an excellent fit to the experimental data [37].

4 The Effect of the Temperature on the Equilibria, Autocatalytic Rate and Reaction Order

Experimental studies on self-replicating hexadeoxynucleotides revealed a remarkable temperature dependance of the autocatalytic synthesis [48]. In any case, the rate was found to exhibit a characteristic maximum at an optimal temperature T_{opt} . An understanding of this phenomenon makes it necessary to consider the influence of the temperature on the rate and equilibrium constants of a minimal selfreplicating system. This is done in the following for the minimal reaction model as defined with the reaction Eqs. (14)–(16). The main goal here is to explain T_{opt} in terms of the temperature-independant thermodynamic and kinetic quantities involved. The results from an analytical treatment of the coupled equilibria (14) and (15) are graphically illustrated and compared to those derived from a numerical treatment of the coupled equilibria (15), (17)–(20). In a later section, the reaction model is extended to account for the reaction equation (21).

Consider Eq. (36). Both, the rate constant k, and the equilibrium concentration [ABC] depend on the temperature. The rate constant k is expected to increase with the temperature according to the law of Arrhenius:

$$k = A \exp\left[-\frac{E_a}{RT}\right]$$
(44)

In the latter equation, A describes the frequency factor *viz*. the rate at infinite temperature and E_a the activation energy. The equilibrium concentration [ABC] is expected to decrease with the temperature due to the weakening of the noncovalent interactions which stabilize the complex. [ABC] can be expressed as a function of the temperature, if the equilibrium constants K_1 (cf. Eq. (29)) and K_2 in Eq. (33) are substituted using the thermodynamic relationships:

$$K_{1} = \exp\left[-\frac{\Delta H_{1}}{RT} + \frac{\Delta S_{1}}{R}\right]$$
(45)

$$K_2 = \exp\left[-\frac{\Delta H_2}{RT} + \frac{\Delta S_2}{R}\right]$$
(46)

 ΔH_1 , ΔH_2 are the enthalpy changes and ΔS_1 , ΔS_2 the entropy changes upon formation of the complexes **ABC** and **C**₂, respectively. For oligoribo- and oligodeoxyribonucleotides a wealth of thermodynamic information is available [46, 49, 50]. From multivariate statistical analyses of sequence-dependant data, tables of enthalpic and entropic increments were compiled which allow to estimate the required quantities for any given sequence [46, 50]. In many cases the estimates are astonishingly good, although for a general use of these tables care must be taken for the conditions (pH, ionic strength) on which the increments are based.

As an example, the thermodynamic data for the following complexes were estimated from sequence increments [46]:

d(C-C-G-C-G-G)₂:
$$\Delta H_2 = -56$$
 kcal/mol, $\Delta S_2 = -149$ e.u.

d(C–C–G)·d(C–G–G): $\Delta H_0 = -22 \text{ kcal/mol}, \Delta S_0 = -66 \text{ e.u.}$

Based on these data, two estimates for the termolecular complex **ABC** $(= d(C-C-G-C-G-G) \cdot d(C-C-G) \cdot d(C-G-G))$ can be made. In the best case, which refers to a cooperative binding of the trimers, the thermodynamic data of **ABC** are given by:

$$\Delta H_1 = \Delta H_2 = -56 \text{ kcal/mol}, \ \Delta S_1 = \Delta S_2 + \Delta S_{nuc} = -162 \text{ e.u.}$$



Fig. 10a, b. Partition of the species ABC, C₂, and C as a function of the temperature T for the case of (a) cooperative, and (b) non-cooperative formation of ABC. The solid curves were generated using the Eqs. (33), (31), (30) for ABC, C₂, C, respectively. The equilibrium constants in the latter equations were calculated from the enthalpic and entropic data given in the text; the total concentrations were $a = b = 10^{-3}$ M, $c = 10^{-4}$ M; quantity q was calculated using Eq. (29). The *dashed curves* (mostly hidden by the solid counterparts) were generated by an iterative calculation of the coupled equilibria (17)–(20) using the following data for the transient complexes AC and BC: $\Delta H_{AC} = \Delta H_{BC} = \Delta H_0$, $\Delta S_{AC} = \Delta S_D$. Moreover, the formation of ABC from AC and BC was calculated as follows: (a) $\Delta H_{ABC,AC} = \Delta S_2 - \Delta S_{BC} + \Delta S_{nuc}$; (b) $\Delta H_{ABC,AC} = \Delta H_0$, $\Delta H_{ABC,BC} = \Delta H_0$, $\Delta S_{ABC,AC} = \Delta S_0$, $\Delta S_{ABC,AC} =$

The rationale behind these assumptions is the one which leads to Eq. (42). A non-cooperative binding of the trimers must be considered as the worst case:

$$\Delta H_1 = 2\Delta H_0 = -44 \text{ kcal/mol}, \ \Delta S_1 = 2\Delta S_0 = -132 \text{ e.u.}$$

Figure 10a shows for the best case data and Fig. 10b for the worst case data how the equilibrium distribution of the species ABC, C, and C₂ depend on the temperature. The solid curves in both figures were generated using Eq. (33) for ABC, Eq. (30) for C, and Eq. (31) for C₂. The dashed curves result from an iterative computation of the equilibrium distribution when the transient complexes as described by the reaction Eqs. (17)–(20) are considered explicitly. It becomes evident that the models are barely distinguishable with respect to their results, both, in the cooperative and in the non-cooperative case. In any case, the termolecular complex ABC exists at low temperatures, the template duplex C₂ at average temperatures, and the unbound template C at high temperatures. The difference between the cooperative and the noncooperative case is basically determined by the temperature below which the termolecular complex ABC can exist.

Following Eqs. (36), (33), and (44)–(46), the autocatalytic rate can be expressed as a function of ten variables:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = f(T, a, b, c, A, E_a, \Delta H_1, \Delta S_1, \Delta H_2, \Delta S_2)$$
(47)



Fig. 11a, b. Logarithm of the autocatalytic rate r of template formation as a function of the temperature T for the case of (a) cooperative, and (b) noncooperative formation of ABC. The solid curves were generated for the given activaenergies E .; tion A $= 10^{11} \text{ s}^{-1}$. The dashed curves were generated for the . parabolic condition for growth as given in Table 2: (a) $E_a = 28 \text{ kcal/mol};$ (b) E. = 16 kcal/mol. *Dots* indicate rate maxima. Further data and conditions as given in the legend of Fig. 10

An interesting property of Eq. (47) is revealed when plotting the logarithm of the autocatalytic rate r as the function of the temperature T for different activation energies E_a (Fig. 11a,b). The curves in Fig. 11a,b were calculated based on the "standard" set of concentrations and on the set of thermodynamic data estimated above for the cooperative (Fig. 11a) and the non-cooperative case (Fig. 11b). The activation energies vary between 10 kcal/mol (top curves) and 30 kcal/mol (bottom curves), whereas the frequency factor A was kept constant. Since the latter quantity only influences the intercept of the curves at the log r-axis but does not affect the course of the slopes, it could be chosen arbitrarily. Fig. 12a,b show for the above data a plot of the autocatalytic reaction order p as a function of the temperature. Figures 12a,b were generated by Eq. (38) providing the equilibrium constants from Eqs. (45) and (46).

The overall temperature response of a minimal self-replicating system can be separated into three stages. At low temperatures the system behaves as a strong exponential replicator (cf. Table 1). At this stage the rate increases with an increase of the temperature. At high temperatures the system behaves as a weak exponential replicator. Here the rate decreases with an increase of the temper-



Fig. 12a, b. Autocatalytic reaction order p as a function of the temperature T for the case of (a) cooperative, and (b) non-cooperative formation of ABC. The curves were generated using Eq. (38). Further data and conditions as given in the legend of Fig. 10

ature. At average temperatures the system behaves as a parabolic replicator. Here the rate may either increase or decrease with the temperature.

The rate maximum is found for the condition:

$$\frac{d\left[\frac{dc}{dt}\right]}{dT} = 0 \tag{48}$$

The optimal temperature, T_{opt} , is the temperature at which the rate maximum occurs. Figure 11a reveals that for low activation energies E_a the rate maximum (points) is found close to the transition from strong exponential to parabolic growth whereas for high activation energies, the rate maximum occurs close to the transition from parabolic to weak exponential growth. There is a certain activation energy at which the rate maximum is found at the stage of parabolic growth (dashed curve). The activation energy required for parabolic growth in the case of noncooperative binding (Fig. 11b) is lower than in the case of cooperative binding (Fig. 11a). The condition which enables parabolic growth can be calculated from the corresponding limit rate law given in Table 1. We may rewrite the square-root law in logarithmic form substituting k, K₁, and K₂ using Eqs. (44)–(46):

$$n\left[\frac{dc}{dt}\right] = \frac{\Delta H_2 - 2(E_a + \Delta H_1)}{2RT} + \frac{2\Delta S_1 - \Delta S_2}{R} + \ln\left[A\frac{ab}{\sqrt{2c}}\right]$$
(49)

If the rate is maximal, the logarithm of the rate is maximal as well:

$$\frac{\mathrm{dln}\left\lfloor\frac{\mathrm{dc}}{\mathrm{dt}}\right\rfloor}{\mathrm{dT}} = \frac{2(\mathrm{E}_{\mathrm{a}} + \Delta\mathrm{H}_{\mathrm{1}}) - \Delta\mathrm{H}_{\mathrm{2}}}{2\mathrm{RT}^{2}} = 0$$
(50)

It follows from Eq. (50):

$$E_a = \frac{\Delta H_2 - 2\Delta H_1}{2} \tag{51}$$

Equation (51) describes the condition under which parabolic growth is optimal with respect to the temperature. In most cases condition (51) will not be satisfied. The rate maximum is then expected to be found either close to the temperature T_1 at which the growth changes from strong exponential to parabolic, or close to the temperature T_2 , at which the transition from parabolic to weak exponential growth takes place. The conditions for these transitions are those which separate the limit rate laws given in Table 1. For the temperature T_1 the condition reads:

$$\sqrt{2K_2c} = K_1 ab \tag{52}$$

whereas the condition for the temperature T_2 is given by:

$$\sqrt{2K_2c} = 1 \tag{53}$$

From Eqs. (52) and (53), the transition temperatures T_1 and T_2 can be calculated if the equilibria constants are substituted by Eqs. (45) and (46). The results are summarized in Table 2. The numerical values for the above data sets are: T_2 = 64 °C, $T_1 = -9$ °C (best case); $T_2 = 64$ °C, $T_1 = -64$ °C (worst case). Note that the transition temperatures T_1 and T_2 given in Table 2 correspond to the

Table 2. Estimation of the optimal temperatures T_{opt} for systems in which the formation of AB is negligible

Rate maximum at	Condition	Approximate T _{opt}
Strong expon.	$E_a = 0$	$T_{opt} < T_1$
Strong expon. to parabolic	$0 < E_a < \frac{\Delta H_2 - 2\Delta H_1}{2}$	$T_{1} = \frac{\Delta H_{2} - 2\Delta H_{1}}{\Delta S_{2} - 2\Delta S_{1} - R \ln \left[\frac{a^{2}b^{2}}{2c}\right]}$
Parabolic	$E_a = \frac{\Delta H_2 - 2\Delta H_1}{2}$	$T_1 < T_{opt} < T_2$
Parabolic to weak expon.	$-\Delta H_1 > E_a > \frac{\Delta H_2 - 2\Delta H_1}{2}$	$T_2 = \frac{\Delta H_2}{\Delta S_2 + R \ln(2c)}$
Weak expon.	$E_a \geq \ - \ \Delta H_1$	$T_{opt} > T_2$

melting temperatures of the complexes ABC and C_2 , respectively. The melting temperature of an oligonucleotide duplex is usually defined as the temperature at which half of the total oligonucleotide concentration exists in the form of the duplex [46, 48]. According to Table 2, a minimal replicator is expected to work at its rate maximum if the temperature is adjusted close to the melting temperature of either ABC or C_2 . The condition given in Table 2 allows us to decide whether T_1 or T_2 is the appropriate temperature.

The situation becomes more complicated if the formation of the complex AB is additionally taken into account. Again, it is likely that such a process occurs in most experimental systems since A and B are complementary. The analytical treatment begins with Eq. (35), in which K_0 is expressed by:

$$K_{0} = \exp\left[-\frac{\Delta H_{0}}{RT} + \frac{\Delta S_{0}}{R}\right]$$
(54)

 ΔH_0 and ΔS_0 describe the enthalpy- and entropy changes upon formation of AB. In order to illustrate the role of this equilibrium, we will make use of the enthalpic and entropic data given above.

Figures 13a,b-15a,b may be directly compared with Figs. 10a,b-12a,b. The partition of the complexes ABC, C and C_2 (Figs. 13a,b, solid curves) was calculated from the respective Eqs (33), (30), (31) taking the quantity q from Eq. (35). Again, the dashed curves—which account for the more complete set of reaction equations, Eqs. (15), (17)-(21)—do not exhibit a substantial deviation from the solid curves. This again means that the analytical equations are sufficient for an approximate description of the temperature response of an "equilibrated" self-replicating system. A comparison of Figs. 10a and 13a shows,



Fig. 13a, b. Partition of the species ABC, C_2 , and C as a function of the temperature T if the formation of AB is not neglected: (a) cooperative, and (b) non-cooperative formation of ABC. The calculation was performed as described in the legend of Fig. 10; Eq. (35) was used to calculate the quantity q from $c_A = c_B = 10^{-3}$ M. Further data and conditions as given in the legend of Fig. 10



Fig. 14a, b. Logarithm of the autocatalytic rate r of template formation as a function of the temperature T if the formation of AB is not neglected: (a) cooperative, and (b) non-cooperative formation of ABC. Activation energies as in Fig. 11. Further data and conditions as given in the legends of Figs. 10 and 13

that the additional equilibrium softens the transition from ABC to C_2 but does not influence the transition from C_2 to C. As a consequence of the former, the temperature range which allows for parabolic growth is increased towards lower temperatures (Fig. 15a). In the case of a non-cooperative formation of ABC, the stage of strong exponential growth becomes practically inaccessable, if the complex AB comes into play (Fig. 15b).

Interestingly to note, that the rate maximum for low values of E_a is no longer close to the transition from strong exponential to parabolic growth. Instead, if the formation of **AB** contributes to the internal equilibria, the rate maximum for low activation energies is always observed at the stage of parabolic growth, both, in the case of a cooperative (Fig. 14a) and a non-cooperative formation of **ABC** (fig. 14b). However, since the transition from C_2 to **C** is not affected by the additional equilibrium, high values of E_a allow us to observe the rate maximum close to the respective transition temperature T_2 . As a consequence, the experimental search for strong exponential growth will probably suffer from the equilibrium under consideration. The only opportunity for exponential growth



Fig. 15a, b. Autocatalytic reaction order p as a function of the temperature T if the formation of AB is not neglected: (a) cooperative, and (b) non-cooperative formation of ABC. Further data and conditions as given in the legends of Figs. 10 and 13

is then conceived for the stage of weak exponential growth which requires a temperature above T_2 . Whether or not this type of growth is observable experimentally, will depend on the temperature response of other reactions (i. e. non-autocatalytic template formation, precursor activation, hydrolysis reactions) which are not considered here.

5 Summary and Conclusions

The above analytical treatment gives an elementary understanding how the autocatalytic growth of a minimal self-replicating system depends on the thermodynamic properties of the complexes involved. If the irreversible step is rate determining, the autocatalytic reaction order p is solely determined by the coupled equilibria. The values of p which one can expect for a minimal replicator are always between 1/2 and 1. For p = 1/2, the autocatalytic growth has been termed parabolic, whereas for p = 1, growth is known as exponential. Based on a consideration of limit cases, two types of exponential growth can be distinguished, which have been introduced as strong and weak exponential growth. Strong exponential growth can only occur at low temperatures where the termolecular complex ABC is the predominant species. Weak exponential growth is predicted for high temperatures where the equilibrium concentration of all complexes is very small. Parabolic growth is expected for average temperatures where the template duplex C, is the predominant species. The

theory explains that the replication rate is maximal at a certain temperature. The optimal temperature depends on the activation energy of the irreversible step. Above a critical activation energy the rate optimum is found at a temperature which is close to the melting temperature of the template duplex C_2 . For activation energies which are substantially smaller than this threshold value, the rate optimum occurs close to the melting temperature of the termolecular complex ABC. Moreover, in the realistic case of a formation of a complex AB, the rate optimum occurs at the stage of parabolic growth.

The theory should prove useful for those chemists who aim to design selfreplicating systems. If thermodynamic data for the involved complexes are available, or, if reasonable estimates can be made, Eq. (33) in combination with Eq. (35) allows us to calculate the equilibrium concentration of the termolecular complex from the total concentrations of A, B and C. The equilibrium concentration of ABC is proportional to the rate of the autocatalytic synthesis. The autocatalytic rate law for the given set of concentrations is characterized by the reaction order p, which can be calculated from Eq. (38). Moreover, the conditions given in Table 1 allow us to predict, whether the growth at the given concentrations is parabolic, weak exponential or strong exponential. The corresponding equations are simple enough to be programmed on a benchtop calculator. The application of the analytical equations is however limited to those systems whose autocatalytic synthesis can be based on the reaction model Eqs. (14)-(16), and (21). Nevertheless, the reaction model can be extended to account for the formation of hairpins. If, for example, template C undergoes an equilibrium with its hairpin C' having the formation constant K_3 [45], the variable q from Eq. (29), or Eq. (35) must simply be multiplied by the factor $1 + K_{3}$.

The important limitation of the theory comes from the assumption that the reaction (14) is rate limiting. If the complexes **ABC** and C_2 are very stable, this assumption may not be valid. For oligonucleotide based systems the rate constant k can be estimated to be in the order between $1-10^{-3}$ s⁻¹, depending on the type of the condensation reaction. The association rate constants for the formation of oligonucleotide complexes were reported to be rather independant from the temperature and from the length of the oligonucleotides involved [51]. A typical value is $k_{ass} = 10^7 \text{ M}^{-1} \text{ s}^{-1}$ [47]. In order to keep the system at its internal equilibrium, the dissociation rate constant for C_2 (and also for ABC) must exceed the above values for k. This restricts the application of the theory to a range of K_2 between 10^6 – 10^9 M⁻¹ here. Clearly, the rates shown in Figs 11 and 14 do overestimate the real rates at low temperatures. Towards low temperatures the assumption of internal equilibrium is no longer valid, since the dissociation of C_2 becomes more and more rate limiting. As a consequence of the latter, the rate maximum is observed at the stage of parabolic growth even for low activation energies. Based on this consideration, it has already been stated that parabolic growth is an optimal type of growth for self-replicating oligonucleotides [1]. A more detailed analysis of self-replicating systems which are no longer equilibrated is the subject of the second part of this theory.

Finally, the above theory complements theoretical results of E. Szathmary [52–54]. It was shown, that self-replicating molecules which compete for their precursors under the constraint of constant organization (viz. in a flow reactor in which the total concentration of the competing templates is kept constant) cannot undergo Darwinian selection, if the growth is sub-exponential. Darwinian selection means that the most efficient replicator outgrows its less efficient competitors and finally supersedes them in the reaction chamber ("survival of the fittest"). Selection necessitates exponential growth; competition between parabolic replicators invariably results in coexistence. Here, the efficient replicator is no longer able to push away its less efficient competitors from the common resources. As a consequence, all replicators survive although at different levels of concentration ("survival of everybody").

From Szathmary's results it follows that exponential growth is *a conditio sine non* for any attempt to realize a Darwinian selection-process in an artificial self-replicating system. Our theory provides clues, how the autocatalytic growth is controlled by the dynamics of the reactions involved. Future experimental approaches may integrate these clues in the search for an exponential synthetic replicator.

Acknowledgments: This work was supported by Deutsche Forschungsgemeinschaft and Fonds der Chemischen Industrie. Comments on the manuscript of G. Boche, Marburg, L.E. Orgel, La Jolla, and F.P. Schmidtchen, München, are gratefully acknowledged.

6 References

- I von Kiedrowski G, Helbing J, Wlotzka B, Jordan S, Matzen M, Achilles T, Sievers D, Terfort A, Kahrs BC (1992) Nachr Chem Tech Lab 40:578
- von Kiedrowski G (1990) In: 40 Jahre Fonds der Chemischen Industrie 1950–1990. Verband der Chemischen Industrie, Frankfurt, p 197
- 3. von Kiedrowski G (1992) Chemie heute 1992/1993:66
- 4 Orgel LE (1992) Nature 358:203
- 5 Joyce GF (1987) In: Cold Spring Harbor Symp Quant Biol, Cold Spring Harbor Press, New York, vol 7, p 41
- 6 Hoffmann S (1992) Angew Chem 104:1032; Angew Chem Int Ed Engl 31:1013
- Z Famulok M, Nowick JS, Rebek J (1992) Acta Chem Scand 46:315
- Rebek J (1992) Chem Ind 1992:171
- 9. Eigen M, Schuster P (1979) The hypercycle, Springer, Berlin Heidelberg New York
- Hoffmann S (1978) Molekulare Matrizen, Akademie-Verlag, Berlin, vol I: Evolution, vol II: Proteine, vol III: Nucleinsäuren, vol. IV: Membrane
- **11.** The term informational self-replication was recently introduced [4] in order to distinguish autocatalytic template syntheses from other forms of autocatalytic reactions which were reported as self-replicating systems: Bachmann PA, Luisi PL, Lang J (1992) Nature 357:57
- Lehn JM (1988) Angew Chem 100:91; Angew Chem Int Ed Engl 27:89
- 13. Whitesides GM (1990) Angew Chem 102:1247; Angew Chem Int Ed Engl 29:1209
- 14. Lindsey JS (1991) New J Chem 15:153
- 15. Schramm G, Grötsch H, Pollmann W (1962) Angew Chem 74:53; Angew Chem Int Ed Engl 1:1
- Df. Orgel LE, Lohrmann R (1974) Acc. Chem. Res. 7:368

17. Woese CR (1967) The origins of the genetic code. Harper and Row, New York

- 18. Crick F (1968) J Mol Biol 38:367
- 19. Orgel LE (1968) J Mol Biol 38:381
- 20. Kuhn H (1981) Angew. Chem. 93:495
- 21. Joyce GF (1989) Nature 338:217
- 22. Pace NR, Marsh TL (1985) Origins of Life 16:97
- Szostak J (1988) In: Benner S (ed) Redesigning the molecules of life. Springer, Berlin Heidelberg New York, p 87
- Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM (1987) Molecular biology of the gene, Benjamin Cummings, Menlo Park, p 1097
- Benner SA, Ellington AD (1990) In: Dugas H (ed) Bioorganic chemistry frontiers, Springer, Berlin, Heidelberg New York, Vol 1, p 1
- 26. Eschenmoser A (1991) Nachr Chem Tech Lab 39:795
- Kellogg RM, Visser CM (1988) In: Benner S (ed) Redesigning the molecules of life, Springer, Berlin Heidelberg New York, p 1
- 28. Cech TR (1990) Angew Chem 102:745; Angew Chem Int Ed Engl 29:716
- 29. Altman S (1990) Angew Chem 102:735; Angew Chem Int Ed Engl 29:707
- 30. Doudna JA, Szostak JW (1989) Nature 339:519
- 31. Doudna JA, Conture J, Szostak JW (1991) Science 251:1605
- 32. von Kiedrowski G (1986) Angew Chem 93:932; Angew Chem Int Ed Engl 25:932
- 33. Zielinski WS, Orgel LE (1987) Nature 327:364
- von Kiedrowski G, Wlotzka B, Helbing J (1989) Angew Chem 101:1249; Angew Chem Int Ed Engl 28:1235
- 35. Tjivikua T, Ballister P, Rebek J (1990) J Am Chem Soc 112:1249
- von Kiedrowski G, Wlotzka B, Helbing J, Matzen M, Jordan S (1991) Angew Chem 103:456, 1066; Angew Chem Int Ed Engl 30:423, 892
- 37. Rotello V, Hong JI, Rebek J (1991) J Am Chem Soc 113:9422
- 38. Hong JI, Feng Q, Rotello V, Rebek J (1992) Science 255:848
- 39. Terfort A, von Kiedrowski G (1992) Angew Chem 104:626; Angew Chem Int Ed Engl 31:654
- 39a. Park TK, Feng Q, Rebek J (1992) J Am Chem Soc 104:4529
- 40. Data analysis has been carried using a computer program written for this purpose. A typical input data set consists of the experimental concentration-time profiles of **A**, **B**, and **C** from four experiments differing in the initial concentrations of **C**. Initial guesses for the apparent rate constants k_a , k_b , and if necessary k_c have to be provided as numerical parameters. The program performs numerical integrations of equation (2) to calculate the theoretical concentrations for the data set. A nonlinear least-square approximation procedure based on the Newton-Raphson algorithm gives the optimized parameters, whose standard errors and covariances are determined consequently.
- 41. Póta G, Bazsa G (1988) J Chem Soc, Faraday Trans I 84 (1):215
- 42. Nowick JS, Feng Q, Tjivikua T, Ballester P, Rebek J (1991) J Am Chem Soc 113:8831
- 43. von Kiedrowski G, unpublished results
- 44. Gear CW (1971) Numerical initial value problems in ordinary differential equations, Prentice Hall, New Jersey
- 45. Wu T, Orgel LE (1992) J Am Chem Soc 114:317, 7963
- 46. Freier SM, Sinclair A, Neilson T, Turner DH (1985) J Mol Biol 185:645
- 47. Freier SM, Albergo DD, Turner DH (1983) Biopolymers 22:1107
- 48. Wlotzka B (1992) Dissertation, University of Göttingen
- Breslauer KH (1988) In: Hinz HJ (ed) Thermodynamic data for biochemistry and biotechnology, Springer, Berlin Heidelberg New York
- 50. Breslauer KH, Frank R, Blöcker H, Marky LA (1986) Proc Natl Acad Sci USA 83:3746
- 51. Pörschke D, Uhlenbeck OC, Martin FH (1973) Biopolymers 12:1313
- 52. Szathmáry E, Gladkih I (1989) J theor Biol 138:55
- 53. Szathmáry E (1989) Oxf Surv Evol Biol 6:169
- 54. Szathmáry E (1991) Trends Ecol Evol 6, 366
- 55. Goodwin JT, Lynn DG (1992) J Am Chem Soc 114:9197
- 56. Persico F, Wuest JD (1993) J Org Chem 58:95