



# Molecular evolution between chemistry and biology

## The interplay of competition, cooperation, and mutation

Peter Schuster<sup>1</sup>

Received: 13 August 2017 / Revised: 18 December 2017 / Accepted: 27 January 2018 / Published online: 2 March 2018  
© The Author(s) 2018. This article is an open access publication

### Abstract

Biological evolution is reduced to three fundamental processes in the spirit of a minimal model: (i) Competition caused by differential fitness, (ii) cooperation of competitors in the sense of symbiosis, and (iii) variation introduced by mutation understood as error-prone reproduction. The three combinations of two fundamental processes each, ( $\mathcal{A}$ ) competition and mutation, ( $\mathcal{B}$ ) cooperation and competition, and ( $\mathcal{C}$ ) cooperation and mutation, are analyzed. Changes in population dynamics that are induced by bifurcations and threshold phenomena are discussed.

**Keywords** Competition · Cooperation · Darwinian optimization · Error threshold · Mutation · Neutral evolution · Selection

### Introduction

Biology is centered around evolutionary thinking or understanding biology implies understanding evolution as Theodosius Dobzhansky pointed out clearly in his famous book on evolution: “Nothing in biology makes sense except in the light of evolution” (Dobzhansky et al. 1977). Pure Darwinian evolution is a simple process but its embedding in nature renders it complex: Natural selection would follow uncomplicated laws in a simple environment. In the light of current molecular biology, there is need for a simple but comprehensive mathematical model of evolution to be able to account for modern genetics. The various epigenetic mechanisms have to be part of any comprehensive model of evolution and to keep such a model amenable to analysis and handling, molecular details must be reduced to a coarse-grained level. This article deals with a flexible model of evolution under defined environmental conditions. We present a concise and comprehensive review of work that was published elsewhere (Schuster 2016a, d, 2017a) together with a few new results.

The model focusses on three basic processes: (i) fitness-driven competition through differential reproductive success, (ii) reproduction-relevant cooperation between competitors, and (iii) reproduction-induced variation. The model is conceived with a modular structure and allows for the implementation of different, more or less complicated mechanisms for the processes (i), (ii), and (iii). For example, variation may be implemented by mutation, by recombination or by both. Here, we shall apply the simplest conceivable mechanisms: reproduction as single enzyme mediated replication (Biebricher 1983), cooperation of competitors as hypercycle dynamics (Eigen and Schuster 1978a, b), and variation as single point mutations based on the uniform error rate assumption (Swetina 1982).

Usage of the notion evolution is often ambiguous and a precise definition is desirable. Here evolution is understood as a process based on reproduction of a genotype being a DNA or an RNA sequence that carries encoded information on the formation of a phenotype, which is evaluated with respect to success in reproduction. The evolutionary process is built upon two foundations: (i) the dynamics at the population level and (ii) the environment dependent encoding of phenotypes in genotype. At the molecular level, the latter boils down to sequence–structure–function relations (Schuster 2016). The simplest systems that are capable of evolution in the sense of the given definition are nothing but special autocatalytic reactions involving polynucleotides under suitable conditions.

---

Special Issue: Chemical Kinetics, Biological Mechanisms and Molecular Evolution.

---

✉ Peter Schuster  
pks@tbi.univie.ac.at

<sup>1</sup> Institut für Theoretische Chemie, Universität Wien,  
Währingerstraße 17, 1090 Wien, Austria

In the next "Preliminaries", some prerequisites are presented for the model, which is introduced in "A minimal mathematical model for the evolution of molecules". The model comes in a deterministic version based on kinetic differential equations or it is formulated as a stochastic process modeled by means of chemical master equations (Schuster 2016). Although almost no analytic solutions are available for master equations derived from nonlinear reaction kinetics<sup>1</sup> in two or more variables, the stochastic version of the model can be studied by efficient simulation methods for the calculations of trajectories (Gillespie 1977, 2007). In three sections, solution curves for the three two-dimensional subspaces, (A) reproduction and mutation, (B) reproduction & cooperation, and (C) cooperation & mutation, are presented, analyzed and interpreted. Then follows a brief discussion of results obtained with the full three-dimensional model, reproduction & cooperation & mutation and in the final section we discuss the simple model in the context of the complex processes in nature.

## Preliminaries

Darwinian evolution is often—and incorrectly—seen as a synonym for optimization mainly because Ronald Fisher's fundamental theorem of natural selection (Fisher 1930, 1958; Price 1972; Ewens and Lessard 2015) focusses on non-decreasing mean fitness  $\phi(t)$  in evolving populations. In the simplest form derived from idealized models, the time derivative of the mean fitness is the variance of the fitness values and therefore a non-negative quantity. In reality, mean fitness is the result of many factors and as Fisher was certainly aware, only a few cases—like single locus genetics—fulfil the theorem in pure form (for elaborate discussions see the more recent literature Plutynski 2006; Okasha 2008).

Evolution is intimately related to the environment in which it takes place and accordingly environment and environmental changes are major factors shaping evolutionary processes. Here, we are primarily interested in the internal dynamics and hence a well-defined and controllable environment is required. For this goal, we introduce a flow reactor in "Idealized environments", which represents a simple device that is not only useful for performing evolution experiments but also provides at the same time a suitable setup for theoretical modeling. More complex environments can

be implemented as long as they can be cast in kinetic equations. In a separate "Basic processes", the processes that will be used to model evolution are introduced. The following section describes the deterministic and stochastic methods, which are applied to find solutions of the kinetic equations. Finally, we review some fundamental features of autocatalysis since this is the chemical counterpart of reproduction.

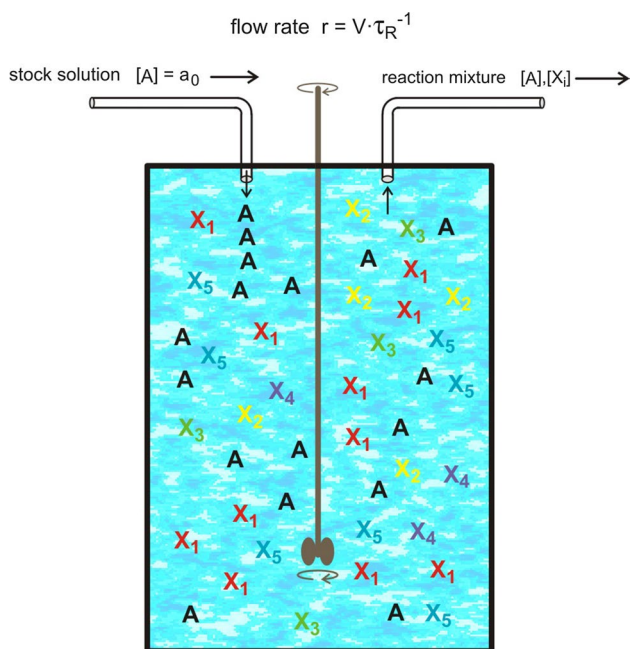
## Idealized environments

Environments that allow for investigations of observations as functions of one or few parameters with everything else being constant require elaborate design in the form of sophisticated experiments since devices controlling environmental conditions may be quite involved. In theoretical approaches, often the silent assumption is made that there exists a hypothetical machinery, which takes care of fixing parameter values as needed for the mathematical analysis.

Environmental influences on phenotypes are commonly large, manifold, and easy to observe. In this study, however, we are interested in the intrinsic driving forces of evolution, which result from reproduction, symbiotic cooperation and variation, and therefore impacts on evolution caused by changes in the environment are intended to be kept to a minimum. To reduce the influence of the environment as much as possible we shall assume a control device in the form of a simple flow reactor (Fig. 1; see also Schmidt 2005). More elaborate reactors, which keep, for example, the numbers of bacterial cells constant have been designed and implemented (Novick and Szillard 1950a, b; Bryson 1952). The flow reactors called chemostat, *cellstat* or turbidostat and other experimental devices for monitoring and controlling evolution in the laboratory may serve as examples (Husimi et al. 1982; Dykhuizen and Hartl 1983; Husimi 1989; Koltermann and Kettling 1997; Strunk and Ederhof 1997).

Implementation of a physical device rather than application of idealized assumptions like constant population size is required for the stochastic description of evolution. An illustrative example where the deterministic approach yields a stable solution whereas the corresponding stochastic system is unstable is provided by the linear birth-and-death process (Goel and Richter-Dyn 1974), which is described by the kinetic equations  $A + X \xrightarrow{f} 2X$  and  $X \xrightarrow{d} \emptyset$ . For equal birth and death rate parameters,  $f = d$ , the population number of the deterministic system stays constant whereas in the stochastic model the fluctuations are unregulated. With increasing amplitude of fluctuations, the populations will hit the death state, which is an absorbing boundary and where the system therefore remains caught forever. In other words, the system is unstable despite a (marginally) stable deterministic solution. The direct incorporation of constant population size into Fisher's selection (1930) or Eigen's equation (1971) leads to an instability of the same kind since

<sup>1</sup> The term *linear* reaction kinetics is used in this contribution for reactions and flow terms of zero and first order, which give rise to constants or linear functions in the kinetic ODEs. First-order kinetics gives rise to exponential functions, which become linear in  $\ln[X]/t$ -plots.



**Fig. 1** The continuous-flow stirred-tank reactor (CSTR). The figure sketches a device for controlling the environmental condition of evolution experiments. The material needed for reproduction is subsumed by **A**, it flows into the reactor with a (volumetric) flow rate  $r$  [ $V/t$ ]≡[volume / time] in form of a solution with concentration  $[A] = a_0$  or number density  $[A] = A_0$ . In the reactor molecules  $X_i$  ( $i = 1, \dots, n$ ) are reproduced and **A** is consumed. The volume  $V$  of the reactor is constant and hence reaction mixture compensating the volume increase through influx of stock solution has to flow out of the reactor. The mean residence time of a volume element in the CSTR is  $\tau_R = V/r$  [t]. Inflow and outflow of materials are handled as virtual chemical reactions or pseudoreactions:  $* \xrightarrow{a_0 \cdot r} A$  for inflow and,  $A \xrightarrow{r} \emptyset$  and  $X_i \xrightarrow{r} \emptyset$  ( $i = 1, \dots, n$ ) for outflow

fluctuations are not self-regulating (Jones and Leung 1981). Implementation of the system in the flow reactor provides stability due to balance control of inflow and outflow modeled by the pseudoreactions  $* \xrightarrow{a_0 \cdot r} A$  and  $A \xrightarrow{r} \emptyset$  as well as  $X_i \xrightarrow{r} \emptyset$  ( $i = 1, \dots, n$ ) (Fig. 1).

**Basic processes**

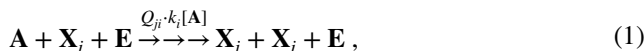
The minimal system for modeling evolution of molecules is based on the three classes of processes: (i) competitive reproduction, (ii) symbiotic cooperation, and (iii) reproduction based variation. For the minimal model we shall choose the simplest possible chemical reactions. Reproduction will be modeled by simple replication,  $A + X_i \xrightarrow{f_i} 2X_i$  with  $f_i$  being the fitness of species  $X_i$  and competition being introduced through living on the same resource **A**. Symbiotic cooperation is introduced as catalyzed replication,  $A + X_i + X_j \xrightarrow{h_{ij}} 2X_i + X_j$  ( $i, j = 1, \dots, n$ ), where  $X_i$  represents the template and  $X_j$  is the catalyst. In its most general form—every molecule  $X_j$  has the potential to act as catalyst

in the replication of every molecular species  $X_i$ —the number of catalytic terms is  $n^2$  and unrealistically large since specific catalysis is a rare property. The simplest example of stable cooperative catalytic networks with fewer catalytic reactions known so far is the catalytic hypercycle (Eigen 1971; Eigen and Schuster 1977, 1978, 1978): The catalyzed reactions  $2X_i + X_{i+1} \xleftarrow{h_i} A + X_i + X_{i+1}$  ( $i = 1, \dots, n; i \text{ mod } n$ )

$$\dots \leftarrow X_n \leftarrow X_1 \leftarrow X_2 \leftarrow \dots \leftarrow X_{n-1} \leftarrow X_n \leftarrow \dots$$

form a closed cycle with  $n$  members.

Genetic variation occurs at the level of a DNA or RNA genotype in forms of mutation and recombination. The simplest form of variation is the point mutation that consists of the exchange of a single nucleotide in the sequence and caused by the incorporation of a wrong nucleotide during the replication process. Correct and error-prone replication are considered as parallel reaction channels within one and the same replication mechanism (Eigen 1971). In terms of a simple over-all replication kinetics of the multistep process,



with **E** being a replicase enzyme, the reaction rate is obtained as the product of two parameters:  $Q_{ji} \cdot f_i$  where  $f_i = k_i [A]$  is the fitness of the template that depends on the availability of resources here the concentration  $[A]$ . The dimensionless factors  $Q_{ji}$  with  $i, j = 1, \dots, n$  are understood as the elements of a mutation matrix  $Q = \{Q_{ij}\}$  and provide the probability to obtain  $X_j$  as a copy of the template  $X_i$  that can be either correct ( $X_j \equiv X_i$ ) or error-prone ( $X_j, j \neq i$ ). Conservation of probabilities requires:  $\sum_{j=1}^n Q_{ji} = 1$  since each copy has to be either correct,  $Q_{ii}$ , or incorrect,  $\sum_{j=1, j \neq i}^n Q_{ji} = 1 - Q_{ii}$ . For the sake of simplicity, binary sequences will be considered here and we distinguish only between correct and incorrect base pairs.

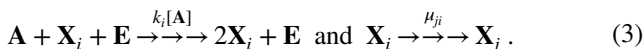
A useful simplifying approximation is made by the uniform error rate model (Swetina 1982): The error per nucleotide and replication event,  $p$ , is assumed to be independent of the position of the nucleotide along the sequence and the nature of the nucleotide to be complemented. Then all elements of the mutation matrix can be expressed by a simple formula,

$$Q_{ji} = p^{d_{ij}} (1 - p)^{l - d_{ij}} = p^l \epsilon^{d_{ij}} \text{ with } \epsilon = \frac{p}{1 - p}, \tag{2}$$

with only three parameters: (i) the sequence length of the RNA molecules  $l$ , (ii) the mutation rate  $p$ , and (iii) the Hamming distance (Hamming 1950, 1986) between the two sequences interrelated by the mutation process,  $d_{ij} = d_H(X_i, X_j)$ . Without changing important results for the purposes pursued here, the analysis of the model is substantially simplified by the assumption of constant chain lengths  $l$ , which is also consistent with the restriction to

point mutations since point mutations do not change chain lengths by definition.

As an alternative to Eq. (1), mutation can be seen as a consequence of DNA change—damage and imperfect repair—during the whole life time of an organism, which is the idea in the Crow–Kimura mutation model (2009), pp 264–266]:



Interestingly, both models (1) and (3) although different with respect to the underlying physics give rise to identical mathematical problems (see e.g. Baake and Gabriel 1999 and Schuster 2016, pp 76–78).

### Deterministic and stochastic approaches

Reaction mechanisms are commonly analyzed by deterministic and stochastic approaches. The former translate the chemical reaction equations into kinetic differential equations, which can be directly solved by mathematics, studied by means of qualitative analysis or investigated by numerical integration. The theorems of existence and uniqueness of solutions of differential equations are applicable and a single integration provides the complete information for a given input set. Competitive selection with nonzero mutation leads to one unique asymptotically stable stationary state (Thompson 1974; Jones et al. 1976), whereas the long-time dynamics of cooperative systems is much richer and multiple stationary states, oscillations or deterministic chaos may be observed (Schuster 1978; Schnabl et al. 1991).

Stochastic analysis in general is based on searching for a stochastic process that fits the model to be studied as closely as possible. Chemical reaction kinetics prefers master equations although the repertoire of analytical solutions is very limited. It is not difficult to write down a multivariate master equation but the derivation of analytical solutions is successful only in exceptional cases, for example for networks of monomolecular reactions (Jahnke et al. 2007; Deuffhard et al. 2008). If no analytical solutions are available information on the stochastic system can be obtained by trajectory sampling. The theoretical background for trajectory harvesting has been laid down by Andrey Kolmogorov (1931), Willy Feller (1940), and Joe Doob (1942, 1945). With electronic computers now being generally available elaborate simulations of stochastic processes became possible. The more recent conception, analysis, and implementation of a simple but highly efficient algorithm by Daniel Gillespie (1976, 1977, 2007) provides a very useful tool for investigations of stochastic effects in chemical kinetics. Distributions of trajectories are characterized by expectation values and higher moments, commonly only by variances or standard deviations.

Equilibrium fluctuations in conventional chemical reaction kinetics follow an approximate  $\sqrt{N}$ -law and hence play almost no role in systems with particle numbers that are typical for chemical systems. In biology a different scenario is encountered. For example, every new variant originating from mutation has to start out from a single copy. The deterministic approach commonly uses continuous variables, which can only be an approximation to reality, since numbers of molecules or biological entities are integers by definition. Continuous concentrations can adopt arbitrarily small values whereas stochastic variables cannot pass low values beyond unity, because then molecular species go extinct, and deterministic solutions become unrealistic and differ strongly from the stochastic results. Large population size alone is not sufficient, each variable has to be sufficiently large at every instant to guarantee similarity between stochastic and deterministic solutions. Two more phenomena are relevant in the stochastic approach at low particle numbers and may lead to large standard deviations in the variables: (i) nucleation steps in reactions involving two or more molecules and (ii) multiple (quasi)stationary states<sup>2</sup> of the stochastic system. The stochastic system may approach any stationary state and the distribution of the possible outcomes is determined by probabilities. Then the calculations of expectation values and higher distribution moments build averages over different final states and may give rise to enormous standard deviations.

For the purpose of illustration, we consider the equilibration of the flow reactor as an example of a stochastic system that can be fully analyzed by analytic calculations (Schuster 2016, pp 436–441). The two pseudoreactions,



are converted into the easy to solve kinetic differential equation

$$\frac{da}{dt} = a_0 \cdot r - a \cdot r = (a_0 - a)r \text{ with } a(t) = n_0 + (n_0 - a_0)e^{-rt} \quad (4c)$$

where  $n_0 = a(0)$ . The master equation for the probability  $P_n(t) = P(A(t) = n)$  with number density  $A(t)$  being the discrete pendant of concentration  $a(t)$ ,

$$\frac{\partial P_n}{\partial t} = r(a_0 P_{n-1} + (n+1)P_{n+1} - (a_0 + n)P_n), \quad n \in \mathbb{N}, \quad (4d)$$

<sup>2</sup> A stochastic quasistationary state is a state towards which the system converges stochastically in the long-time limit and around which it fluctuates. It is not an absorbing boundary, and if true asymptotically stable stationary states exist the system converges to one of them in the limit  $t \rightarrow \infty$  although the mean time of convergence may be extremely long.

has the analytical solution

$$P_n(t) = \sum_{k=0}^{\min\{n_0, n\}} \binom{n_0}{k} a_0^{n-k} \frac{e^{-krt}(1 - e^{-rt})^{n_0+n-2k}}{(n-k)!} e^{a_0(1-e^{-rt})} \tag{4e}$$

with  $n, n_0, a_0 \in \mathbb{N}$  for the sharp initial condition  $P_n(0) = \delta_{n,n_0}$ . In the limit  $t \rightarrow \infty$  the distribution (4e) converges to a Poisson distribution

$$\lim_{t \rightarrow \infty} P_n(t) = \bar{P}_n = \frac{a_0^n}{n!} \exp(-a_0). \tag{4f}$$

First and second moment yield expectation values and standard deviations

$$\begin{aligned} E(A(t)) &= a_0 + (n_0 - a_0)e^{-rt} \text{ and} \\ \sigma(A(t)) &= \sqrt{(a_0 + n_0 e^{-rt})(1 - e^{-rt})}. \end{aligned} \tag{4g}$$

The standard deviation starts out from  $\sigma(t=0) = 0$  and approaches the equilibrium value  $\bar{\sigma} = \sqrt{a_0}$  either from below for  $a_0 > n_0$  or from above for  $a_0 < n_0$  after having passed a maximum at

$$t_{\max} = \frac{1}{r} \ln \frac{2n_0}{n_0 - a_0}. \tag{4h}$$

In general the maximum if it exists is very flat (see Fig. 3) as can be easily checked from the analytical expressions through inspection:

$$E(A(t_{\max})) = a_0 + \frac{(n_0 - a_0)^2}{2n_0} \text{ and } \sigma(A(t_{\max})) = \frac{n_0 - a_0}{2\sqrt{n_0}}.$$

The flow reactor is one of the few rather exceptional cases where the stochastic approach can be done completely by mathematical analysis.

### Autocatalysis in the batch reactor

A batch reactor is an elaborate device that allows for performing chemical reactions under controlled conditions without inflow and outflow (Schmidt 2005). Here, the term batch reactor is used to indicate that reactions are carried out in a well-mixed closed system under constant temperature and pressure, and in the long run approach a thermodynamic equilibrium state.

In conventional chemistry autocatalysis is a rather rare phenomenon but in biology it represents the most important process since multiplication is just a special form of autocatalysis that in simplest form can be expressed by the reversible reaction



The forward reaction of (5) leads to molecular self-enhancement: The resource **A** is consumed in the synthesis of the replicator **X** and the process is catalyzed by the presence of further molecules **X**. Depending on the molecularity of the autocatalytic process as expressed by the value of  $m$  autocatalysis comes in different forms that, in essence, fall into two classes: (i) simple or first-order autocatalysis with  $m = 1$  and (ii) second- and higher-order autocatalysis with  $m \geq 2$  [(Phillipson and Schuster 2009, pp 18–27)]. For consistency, we add here also the uncatalyzed reaction and call it *zeroth order autocatalytic* ( $m = 0$ ).<sup>3</sup> All processes in closed systems converge to an equilibrium state and show mass conservation that implies a conservation relation  $A(0) + X(0) = A(t) + X(t) = N$ . The reversible autocatalytic reactions converge to the equilibrium state:

$$S_1: \bar{a} = N/(1 + K), \bar{x} = N \cdot K/(1 + K).$$

Although the expressions for the equilibria are the same for the all reactions independently of the stoichiometric coefficient  $m$ ,  $K = g_{\leftarrow}/g_{\rightarrow}$ , there are subtle differences in the probability distributions, which become important at small concentrations: The particle numbers are discrete quantities, change by integers only, and the stoichiometric factors are  $X(t)(X(t) - 1) \equiv (X(t))_2$  or  $X(t)(X(t) - 1)(X(t) - 2) \equiv (X(t))_3$  rather than  $X(t)^2$  and  $X(t)^3$ , respectively.<sup>4</sup> The states with  $X(t) = 1$  for  $m = 1$ , or the states  $X(t) = 1$  and  $X(t) = 2$  for  $m = 2$  cannot react because two or three molecules **X** are needed for the conversion of **X** into **A**. The inaccessibility of the state with  $X(t) = 0$  or the states with  $X(t) = 0$  and  $X(t) = 1$  in the first- or second-order autocatalytic reactions require different normalizations for the stationary probabilities of the three systems:

$$\bar{P}_n^{(0)} = \binom{N}{n} \frac{K^n}{(1 + K)^N}; n \in [0, N], \tag{6a}$$

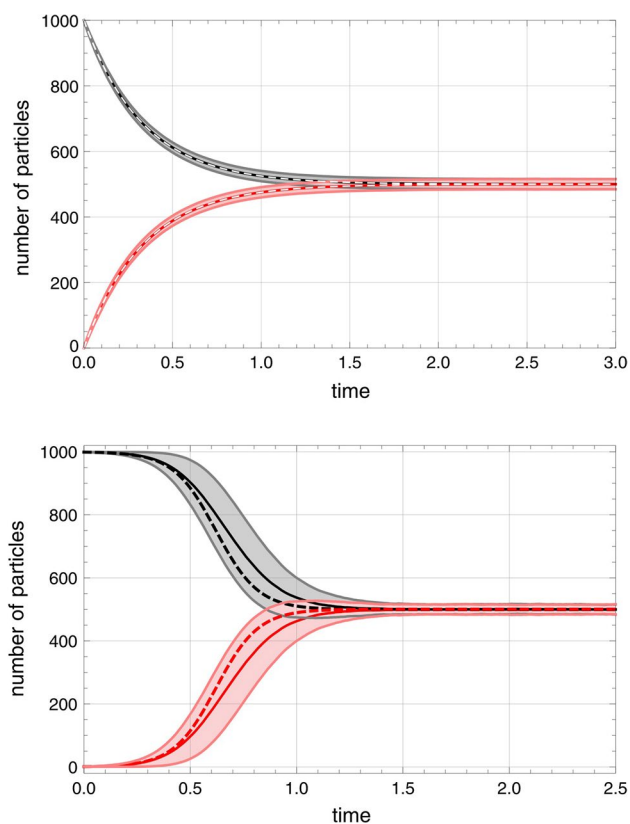
$$\bar{P}_n^{(1)} = \binom{N}{n} \frac{K^n}{(1 + K)^N - K^N}; n \in [0, N - 1], \tag{6b}$$

$$\bar{P}_n^{(2)} = \binom{N}{n} \frac{K^n}{(1 + K)^N - K^N - N K^{N-1} - 1}; n \in [1, N - 2], \tag{6c}$$

where as before the stochastic variable is  $n = A$ . Mass conservation provides  $X = N - n$ . As expected the truncation of  $\bar{P}_n^{(0)}$  is important for small values of  $N$  only. For first-order autocatalysis with  $N = 10$ ,  $k_{\leftarrow} = k_{\rightarrow} = 1.0$  we obtain

<sup>3</sup> We shall use different letters for the rate parameters: For  $m = 0$  we use  $g \equiv h$ , for  $m = 1$   $g \equiv k$ , and for  $m = 2$   $g \equiv l$ . It is worth recalling the different dimensions:  $h$  [ $t^{-1}$ ],  $k$  [ $M^{-1}t^{-1}$ ], and  $l$  [ $M^{-2}t^{-1}$ ] that are the same in both directions.

<sup>4</sup> Here the expression  $(x)_n = x!/(x - n)!$  denotes the falling Pochhammer symbol.



**Fig. 2** Comparison of fluctuations in the reversible zero- and first-order autocatalytic reaction  $A + mX \rightleftharpoons (m + 1)X$  with  $m = 0, 1$  in the batch reactor. The two plots show the expectation values  $E(A(t))$  (black) and  $E(X(t))$  (red) together with the one standard deviation band  $E(t) \pm \sigma(t)$  (gray and pink) obtained by sampling of 10,000 trajectories that were calculated by Gillespie's simulation method for the uncatalyzed ( $m = 0$ ; top plot) and the first-order autocatalytic reaction ( $m = 1$ ; bottom plot). Both reactions approach almost identical thermodynamic equilibrium states. Choice of parameters and initial conditions:  $N = 1000$ ;  $h = h^- = 1.5[r^{-1}]$ ,  $A(0)=1000$ ,  $X(0)=0$  (top plot);  $k_+ = k_- = 0.01[M^{-1} r^{-1}]$ ,  $A(0)=999$ ,  $X(0)=1$  (bottom). Solutions of the corresponding kinetic differential equations are shown as broken lines.

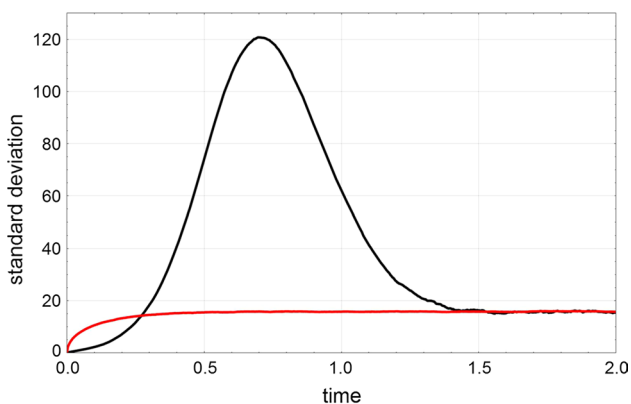
$E(\bar{A}) = 4.9951$ , for second-order autocatalysis  $E(\bar{A}) = 4.9605$  compared to  $E(\bar{A}) = 5$  for the uncatalyzed process.

In principle, there are three major sources of randomness in autocatalytic reactions: (i) thermal fluctuations, (ii) delayed onset of autocatalytic reactions and (iii) multiple stationary states. (ad i) In the transients towards equilibrium the thermal fluctuation bands are essentially the same in autocatalytic and conventional reactions as can be seen best from the comparison of standard deviations at equilibrium (Fig. 2). The differences between conventional and autocatalytic equilibrium densities can be recognized numerically for very small particle numbers only (). (ad ii) The reaction rate for an autocatalytic reaction of order  $m$  is  $v(t) = k A(t) (X(t))_m - l (X(t))_{m+1}$  (with  $m \geq 1$ ). At

small  $t$  the factor  $A(t)$  is large and the factor(s) containing  $X(t)$  are small and only the first term in  $v(t)$  matters in the early phase of the reaction. Thus  $X$  is produced and  $X(t)$  increases, which in turn leads to an increase  $v(t)$  yielding the typical scenario of self-enhancement. Self-enhancement in chemical reactions is tantamount to an increase of the reaction rate with concentration in the early phase and together with the late saturation phase gives rise to "s"-shaped or *sigmoid* curves whereas the uncatalyzed reaction ( $m = 0$ ) follows a simple exponential decay (Fig. 2). Higher values of  $m$  lead to steeper curves, which approach a step function with increasing  $m$ . The maximum standard deviation in the approach towards equilibrium  $\sigma_{\max} = \max\{\sigma(t)\}$  is a measure of the random scatter in the delay in the onset of the autocatalytic reaction (Table 1). (ad iii) Multiple final states give rise to an additional stochastic component often called anomalous fluctuations (de Pasquale et al. 1980) since "Autocatalysis in theow reactor").

In Fig. 2 transients for the two processes  $A + mX \rightleftharpoons (m + 1)X$  with  $m = 0$  and 1 are compared by means of expectation values and fluctuation bands. As initial conditions we apply an empty reactor,  $A(0) = 0$ , and the smallest possible values for  $X$ :  $X(0) = 0$  and  $X(0) = 1$  for the uncatalyzed and the autocatalytic reaction, respectively. A total population size of  $N = 1000$  was chosen so that the one-standard-deviation fluctuation band of order  $\sqrt{N}$  appears small and the deterministic solutions coincide with the expectation values in the reference process  $A \rightleftharpoons X$  ( $m = 0$ ). The transient of the autocatalytic process ( $m = 1$ ) is different: It shows substantial broadening of the fluctuation band (Table 1) because of delayed onset of the reaction before it narrows down to the equilibrium value. In the intermediate range the expectation values differ remarkably from the deterministic solution curves. As expected an increase in  $X_0$  leads to a decrease in the width of the fluctuation band. Second order autocatalysis ( $m = 2$ ) differs from first-order autocatalysis mainly by the size of the characteristic effects: Both autocatalytic reactions show broadening of the fluctuation bands but the band width in the second-order case is about four times as wide. In particular, the scatter in the waiting times until the first reaction events occurs is much larger because we are dealing with two small factors,  $(X - 1) \cdot (X - 2)$ , in the expression for  $v(0)$ .

The standard deviation in the course of the reactions,  $\sigma(t)$ , is shown in Fig. 3. Because of sharp initial conditions the fluctuation band starts out from zero— $\sigma(0) = 0$ , increases, becomes broad in the intermediate range and settles down at the equilibrium value (6). Substantial deviations between the deterministic solution and the stochastic expectation value,  $a(t)$  and the  $E(A(t))$  or  $x(t)$  and  $E(X(t))$ , respectively, are observed in the intermediate range. Accordingly, the standard deviation goes through a pronounced maximum that is



**Fig. 3** Comparison of the standard deviation in the reversible first-order autocatalytic reaction  $A + X \rightleftharpoons 2X$  and the reversible uncatalyzed reaction  $A \rightleftharpoons X$ . The reactions are recorded for the closed system fulfilling  $A(t) + X(t) = A_0 + X_0 = \text{const} = N$  where  $A(0) = A_0$  and  $X(0) = X_0$  for both reactions. The figure presents the results of statistical evaluation of 10 000 trajectories obtained by computer simulations with Gillespie’s method (Gillespie 2007) for the autocatalytic reaction  $\sigma_X^{(1)}(t) = \sigma^{(1)}(t)$  (black) and for the uncatalyzed reaction  $\sigma_X^{(0)}(t) = \sigma^{(0)}(t)$  (red). Choice of parameters:  $k_+ = k_- = 0.01[M^{-1} t^{-1}]$  and  $h_+ = h_- = 1.5[t^{-1}]$ ; equilibrium parameters  $K = k_+/k_- = h_+/h_- = 1$ ; initial conditions:  $N = 1000$ ,  $X_0 = 1$ ,  $A_0 = 999$  and  $N = 1000$ ,  $X_0 = 0$ ,  $A_0 = 1000$  yielding the numerical equilibrium values  $\bar{A} = \bar{X} = 500$ . The equilibrium value of the standard deviation is practically the same for both reactions:  $\lim_{t \rightarrow \infty} \sigma_X^{(1)}(t) \approx \lim_{t \rightarrow \infty} \sigma_X^{(0)}(t) \approx 15.8114$

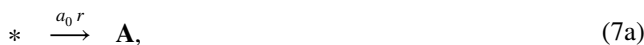
qualitatively different from the shallow maximum observed with the standard deviation of conventional chemical processes (see Eq. (4h)).

The irreversible processes,  $A + mX \rightarrow (m + 1)X$  obtained from Eq. (5) by putting  $g_- = 0$ , are illustrative because they are closer to biology where replication or reproduction occur always under the conditions of irreversibility. The shape of the solution curves compared to

those shown in Fig. 2 shows similarities except a twice as wide range for the values of the stochastic variables,  $A(0) + X(0) \geq A(t) \geq 0$  and  $1 \leq X(t) \leq A(0) + X(0)$  and expectation values and standard deviations approach zero at sufficiently long times. Again, we observe a sigmoid shape of the solution curves, for small initial values ( $X(0) < 5$ ) the standard deviation  $\sigma(t)$  becomes large in the intermediate range (Fig. 3), and the deterministic curve deviates substantially from the stochastic expectation values.

### Autocatalysis in the flow reactor

Implementation of autocatalytic reactions in the flow reactor provides additional insights into the different forms of randomness. In particular we are interested in multiple stationary states as a source of stochasticity (item iii). The reaction equations for first order autocatalysis are:



with the kinetic differential equations

$$\frac{da}{dt} = -kax + (a_0 - a)r \text{ and } \frac{dx}{dt} = x(ka - r). \tag{7e}$$

The simple first-order autocatalytic process in the flow reactor sustains two long time states: (i) The *state of extinction*  $S_0$  with  $\lim_{t \rightarrow \infty} a(t) = \bar{a} = a_0$  and  $\lim_{t \rightarrow \infty} x(t) = \bar{x} = 0$ , and (ii) the quasistationary state  $S_1$  with  $\lim_{t \rightarrow \infty} a(t) = \bar{a} = r/k$  and

**Table 1** Fluctuations in the two autocatalytic reactions  $A + X \rightleftharpoons 2X$  and  $A + 2X \rightleftharpoons 3X$ <sup>a</sup>

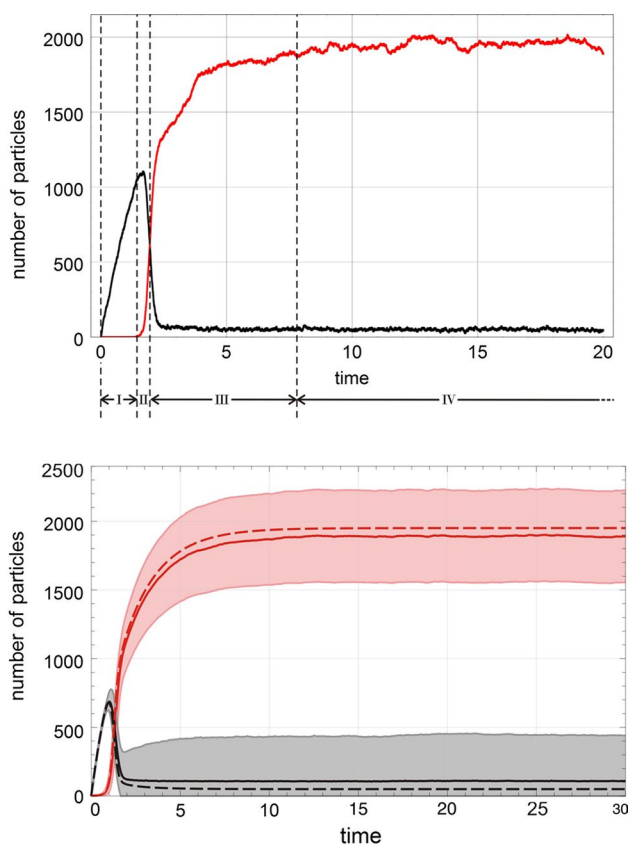
Autocatalysis order	Reaction	Initial conditions				Limit $t \rightarrow \infty$
		$X_0 = 1$	$X_0 = 2$	$X_0 = 5$	$X_0 = 10$	
Zero	$A \rightleftharpoons X$	–	–	–	–	15.8 <sup>b</sup>
First	$A + X \rightleftharpoons 2X$	123.2	89.8	59.6	44.4	15.8 <sup>b</sup>
Second	$A + 2X \rightleftharpoons 3X$	–	245.1	226.7	189.4	15.8 <sup>b</sup>

The table presents maximal standard deviations  $\sigma_{\text{max}}$  computed from 1 000 trajectories of the two autocatalytic reactions for different initial conditions  $X_0 = X(0)$ . For the autocatalytic reactions the standard deviation  $\sigma(t)$  passes through the maximum  $\sigma_{\text{max}}$  listed here whereas for the uncatalyzed process it increases monotonously from  $\sigma(0) = 0$  to the equilibrium value (see Fig. 3)<sup>a</sup>

The equilibrium fluctuations calculated from equations (7) are practically the same for all three reactions. Choice of parameters,  $N = 1000$ ,  $A \rightleftharpoons X$ :  $h_+ = h_- = 1.5[t^{-1}]$ ;  $A + X \rightleftharpoons 2X$ :  $k_+ = k_- = 0.01[M^{-1} t^{-1}]$ ;  $A + 2X \rightleftharpoons 3X$ :  $l_+ = l_- = 0.00001[M^{-2} t^{-1}]$

<sup>a</sup>Depending on rate parameters and initial conditions the trajectories may pass a flat maximum before they decrease to the equilibrium value (see Eq. (4h) and Schuster 2016, pp. 445–449)

<sup>b</sup>The accurate value obtained from the stationary master equation is  $\sigma = 15.8114$



**Fig. 4** The first-order autocatalytic reaction.  $\mathbf{A} + \mathbf{X} \rightleftharpoons 2\mathbf{X}$  in the flow reactor. The two plots show a single trajectory (top plot) and statistics consisting of expectation value within the one-standard deviation band,  $E(A(t)) \pm \sigma(A(t))$  and  $E(X(t)) \pm \sigma(X(t))$  taken from sample of 100 trajectories (bottom plot). The four phases of the approach to the long-time solution are indicated (top plot; see text). Choice of parameters and initial conditions:  $N = 2000$ ;  $k = 0.01[\text{M}^{-1} \text{t}^{-1}]$ ,  $r = 0.5 [\text{V} \text{t}^{-1}]$ ,  $A(0) = 0$ ,  $X(0) = 1$  (top) and  $X(0) = 3$  (bottom). Pseudorandom number generator: *ExtendedCA, Mathematica*, seeds  $s = 491$ . The sample size of the bottom plot was 100 trajectories, 3 led to  $S_0$  (extinction) and 97 reached the pseudo-stationary state  $S_1$ . Solutions of the corresponding kinetic differential equations are shown as dashed lines

$\lim_{t \rightarrow \infty} x(t) = \bar{x} = a_0 - r/k$ . Stability of  $S_1$  requires that the condition  $r < a_0 k$  is fulfilled.

Starting from an empty reactor containing no  $\mathbf{A}$  and the autocatalyst only in seeding quantities,  $A(0) = 0$  and  $X(0) = 1, 2, 3, \dots$ , the trajectory shown in the upper part of Fig. 4 allows for the distinction of several phases: (I) the flow reactor is filled with the resource  $\mathbf{A}$  in phase I, (II) in the random phase II the decision is made to which state the trajectory will converge, (III) the trajectory approaches the long-time state, and (IV) the trajectory fluctuates around the state (see Figs. 4 and 7). First-order autocatalysis sustains the two long-time solutions  $S_0 = (\bar{a}^{(0)}, 0)$  and  $S_1 = (\bar{a}^{(1)}, \bar{x}^{(1)})$ , stochastic trajectories approach either of the two states, and parameters and initial conditions

determine the probabilities to end up here or there. For sufficiently large population sizes, the long-time expectation values of the stochastic variables can be well approximated by linear combination of the deterministic values:

$$E(\bar{A}) = \xi_0 \bar{a}^{(0)} + \xi_1 \bar{a}^{(1)} \text{ and } E(\bar{X}) = \xi_0 \bar{x}^{(0)} + \xi_1 \bar{x}^{(1)}$$

with  $\xi_0 = N_0/(N_0 + N_1)$  and  $\xi_1 = N_1/(N_0 + N_1)$  where  $N_0$  and  $N_1$  are the counted numbers of trajectories ending up in  $S_0$  or  $S_1$  from a sufficiently large sample. Although only 3/100 trajectories go to extinction in the example shown in the lower plot of Fig. 4 the influence on the expectation values  $E(A(t))$  and  $E(X(t))$  and the standard deviations  $\sigma(A(t))$  and  $\sigma(X(t))$  is remarkable. This random component of processes has been intensively studied in the nineteen eighties by Paolo Tombesi, Francesco de Pasquale, Piero Tartaglia and the notion anomalous fluctuation caused by a chemical instability was coined for this kind of stochasticity (de Pasquale et al. 1980, 1982, 1982). It is advantageous to collect trajectories separately for the different final states, because then the anomalous fluctuations disappear. For example, the standard deviation of  $\mathbf{A}$  at  $t = 30$  is reduced from  $\sigma(A(30)) = 335.7$  to  $\sigma(A(30)) = 7.12$  if one changes from a sample of hundred trajectories with three extinction events (Fig. 4, Pseudorandom number generator: *ExtendedCA, Mathematica*, seeds  $s = 491$ ) to one without extinction ( $s = 919$ ).

The major difference between the two classes of autocatalytic reactions lies in the repertoire of possible dynamic behaviors. First-order autocatalysis gives rise to exponential growth in unconstrained systems and to selection and optimization of mean fitness in multispecies cases with finite resources (see "Competition, mutation and quasispecies"). Accordingly first-order autocatalysis leads to a Darwinian scenario of selection of the fittest. In contrast to first-order autocatalytic reaction networks, the dynamics in second-order systems is very rich and includes multiple stationary states, oscillations and deterministic chaos. The second-order autocatalytic elementary step,  $\mathbf{A} + 2\mathbf{X} \rightleftharpoons 3\mathbf{X}$ , represents a kind of generally used prototype for theoretical models, for example the Brusselator (Nicolis and Prigogine 1977). It provides a simple enough reaction step for studies by means of rigorous mathematics. Qualitative analysis of Brusselator dynamics is straightforward and numerical integration causes no problem provided the integration software can handle stiff differential equations. In reality, however, single-step autocatalytic reactions are extremely rare, instead we are commonly dealing with multistep-reaction networks (Noyes et al. 1972) (see also the review by Francesc Sagués and Irving Epstein (2003).

In biology, in particular in the theory of evolution, the process  $\mathbf{A} + 2\mathbf{X} \rightarrow 3\mathbf{X}$  plays a special role since in the simple form of hypercycles it is the basis for suppression of

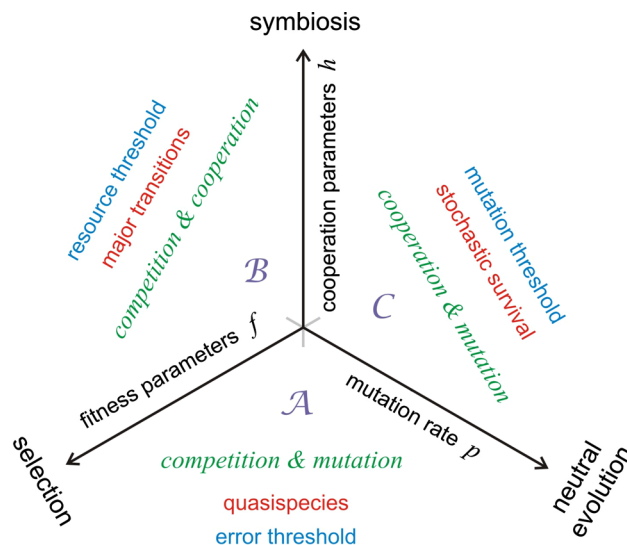
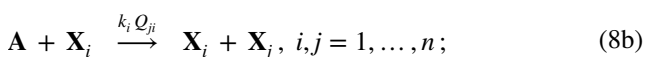


competitive selection without destroying template-induced reproduction. It provides one fundamental mechanism for major transitions (Maynard Smith and Szathmary 1995; Schuster 1996) and will be discussed extensively in "Competition and cooperation". The enormous flexibility of second-order autocatalysis follows, for example, from Fisher’s selection equation and the proof for the optimization of mean fitness in sexual reproduction under idealized condition. In a caricature model we may explain how the above mentioned reaction step could be related to sexual reproduction: 2 X on the l.h.s. are (at least stoichiometrically) related to the parents and 3 X on the r.h.s. model parents and one offspring. Apart from being illustrative toys autocatalytic processes set the stage for modeling evolution in the sense that we shall reencounter all special phenomena of autocatalysis in the more elaborate model for evolution to be presented and discussed in the next "A minimal mathematical model for the evolution of molecules".

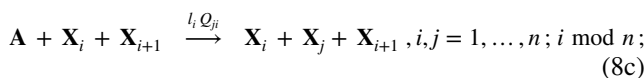
### A minimal mathematical model for the evolution of molecules

The minimal model is dealing with the time dependence of the distribution of genotypes in populations  $\Pi(t)$  and hence it is rooted in chemical kinetics of (i) competitive reproduction, (ii) symbiotic cooperation, and (iii) genetic variation. The quantification of these three properties yields three parameters or sets of parameters, which can be plotted on the three axes of a Cartesian coordinate system (see Fig. 5 and the previous publications (Schuster 2016a, b, d, 2017b)). We consider the simplest case here, where the three quantities are a fitness parameter  $f$ , a cooperation parameter  $h$ , and an mutation rate parameter  $p$ . For an implementation of the model in the flow reactor we need two additional external parameters measuring the accessible resources expressed, for example, as number density or concentration of a (hypothetical) compound A,  $A_0$  or  $a_0$ , respectively, and the mean residence time  $\tau_R = V/r$  of the reaction mixture in the reactor where  $V$  is the volume of the reactor and  $r$  is the (volumetric) flow rate. The parameter  $\tau_R$  defines the time resolution of the reactor since slow reactions, which do not progress appreciably during the time interval  $\tau_R$  cannot be studied.

In the next step, the model is implemented by means of a suitable and simple reaction mechanism. Based on "Idealized environments" and "Basic processes" we consider the following set of  $2n^2 + n + 2$  reactions in the flow reactor:



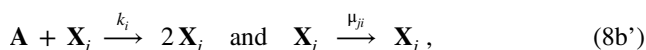
**Fig. 5** A minimal model for modeling evolution. Evolution is considered as an interplay of three processes: (i) competition through reproduction, (ii) cooperation through symbiosis, and (iii) mutation through error-prone replication. In parameter space, the intensity parameters of all three processes, (i) fitness parameters  $f$  corresponding to reaction rates for competition, (ii) reaction rates  $h$  for catalyzed reproduction, and (iii) an error rate parameter  $p$  for mutation are plotted on the axes of a Cartesian coordinate system. On the three two-dimensional faces of the coordinate system we are dealing with the three fundamental evolutionary processes: (A) competitive reproduction and mutation are the basis of Darwinian optimization through natural selection and give rise to the formation of quasispecies and eventually to the occurrence of error thresholds (Eigen 1971; Eigen and Schuster 1977; B) the interplay of competition and cooperation allows for the description of major transitions, which are seen as the consequences of crossing resource thresholds (Maynard Smith and Szathmary 1995; Szathmary and Maynard Smith 1995; Schuster 1996, 2016); (C) the combination of cooperation and mutation enables reintroduction of extinguished species provided the error rate is sufficiently large such that a mutation threshold for stochastic survival can be recognized (Schuster 2016)



The process (8a) supplies the material required for reproduction. A solution with A at concentration  $a_0$  flows into a continuously stirred tank reactor (CSTR) with a flow rate parameter  $r$  (Schmidt 2005, p. 87ff). The reactor operates at constant volume and this implies that the volume of solution flowing into the reactor per time unit  $[t]$  is compensated exactly by an outflow, which is described by the Eqs. (8d) and (8e) and concerns all  $(n + 1)$  molecular species, A and  $X_i, i = 1, \dots, n$ . Inflow and outflow are often characterized

as *pseudoreactions* because they are no chemical reactions in the strict sense, which are converting reactants into products. The two classes of reactions producing progeny, template induced replication (8b) and catalyzed template induced replication (8c), represent the core of the evolution model. In agreement with the conditions in biology, both reproduction steps are implemented irreversibly in the direction of polynucleotide synthesis. A basic assumption for both reproduction steps is that correct reproduction and mutation are parallel chemical reaction channels ("Basic processes"). In other words, there is no mutation under conditions that do not sustain reproduction.

As an alternative to the Eigen model (1) mutation can be seen, for example, as the result of DNA damage and imperfect damage repair during the whole life span of an organism, which is the idea underlying the Crow–Kimura mutation model (Crow and Kimura 2009, pp 264–266). Then reproduction and mutation are completely independent processes,



and in the kinetic differential equations they appear as additive terms. Interestingly, the Eigen and the Crow–Kimura model although being fundamentally different with respect to the assumptions about the nature of the mutation process give rise to the same mathematical problems [see, e.g., (Schuster 2016d, pp.76–78)]. The mutation matrix  $Q$  corresponds formally to the mutation matrix  $\mu$  but there are non-negligible differences  $Q$  covers correct and error-prone replication but the process  $X_i \rightarrow 2X_i$  is handled separately in the Crow–Kimura model and hence all diagonal terms are zero  $\mu_{ii} = 0 \forall i = 1, \dots, n$ .

The equations that will be applied in the analysis of the dynamics of the model (8) implement three processes along the coordinate axes: (i) Darwinian selection of the fittest on the competition axis, (ii) hypercycle dynamics on the cooperation axis, and (iii) neutral evolution on the mutation axis. The kinetic differential equations of the model mechanism (8) are of the form:

$$\frac{da}{dt} = -a \sum_{i=1}^n (k_i + l_i x_{i+1}) x_i + r(a_0 - a), \quad i \text{ mod } n \text{ and} \quad (9a)$$

$$\frac{dx_i}{dt} = a \left( \sum_{j=1}^n Q_{ij} (k_j + l_j x_{j+1}) x_j \right) - x_i r, \quad (9b)$$

$$i = 1, 2, \dots, n; j \text{ mod } n.$$

In (9a) we made use of the conservation relation  $\sum_{i=1}^n Q_{ij} = 1$ . No analytical solutions are available for (9) in general but numerical integration is straightforward as

long as  $n$  is not too large. In absence of cooperation terms,  $l_i = 0 \forall i = 1, \dots, n$ , Eq. (9) can be transformed into an eigenvalue problem of a symmetric matrix, which is readily diagonalized provided  $n$  is not very large ( $n < 10^6$ ; "Competition, mutation and quasispecies").<sup>5</sup> In mutation-free systems,  $p = 0$  ("Competition and cooperation"), qualitative analysis and determination of stationary states are straightforward, and the dynamics of the complete system can be derived by extrapolation from the error-free results to finite mutations rates. The cooperation system with mutation, (9) with  $k_i = 0 \forall i = 1, \dots, n$  is used here to study the relevance of mutation in symbiotic systems. It also serves as an example for the study of unconventional consequences of replication with frequent errors in the strong mutation scenario ("Cooperation and mutation").

### Processes on individual coordinate axes

The processes along the individual coordinate axes are considered in order to verify the initial statement on the three basic processes. For this goal it is easiest to set certain parameters zero: (I) no natural selection implies  $k_1 = k_2 = \dots = k_n = k (= 0)$ , (II) no cooperative coupling requires  $l_1 = l_2 = \dots = l_n = l = 0$ , and (III) no mutation leads to  $Q = \mathbb{1}$  where  $\mathbb{1}$  is the unit matrix.

A process taking place on the selection axis is given by (II) and (III) being true leads to the ODE

$$\frac{da}{dt} = -a \sum_{i=1}^n k_i x_i + r(a_0 - a) \quad \text{and} \quad (10a)$$

$$\frac{dx_i}{dt} = (k_i a - r) x_i; \quad i = 1, 2, \dots, n. \quad (10b)$$

Competition between the reproducing elements  $X_i$  leads to *survival of the fittest subspecies*  $X_m$ , which is defined by  $k_m a = f_m = \max\{f_i; i = 1, \dots, n\}$ . For constant  $[A] = a_0 = \text{const}$ , the mean fitness of the population,  $\phi(t) = \sum_{i=1}^n f_i x_i(t) / c(t)$  with  $c(t) = \sum_{i=1}^n x_i(t)$ , is a non-decreasing function of time:  $d\phi/dt = \text{var}\{f\} \geq 0$ . This result is the formal analogue of Fisher's fundamental theorem for asexual reproduction.

<sup>5</sup> We remark that the deterministic kinetic Eqs. for (8b) and (8c) were extensively studied under the simplifying assumption of constant population size  $\sum_{i=1}^n x_i(t) = c_0 = \text{const}$  (Eigen 1971; Eigen and Schuster 1978; Eigen and McCaskill 1989). The deterministic solution curves formulated in relative concentrations  $\xi_i(t) = x_i(t) / \sum_{i=1}^n x_i(t)$  are identical for the CSTR and for constant population size (Schuster and Sigmund 1985). As we mentioned already in "Idealized environments" the stochastic system is unstable for unregulated constant population size (Jones and Leung 1981) and a proper description has to account explicitly for the physical setup applied, here the flow reactor.

On the cooperation axis, the conditions (I) and (III) are fulfilled and we obtain the equations of hypercycle dynamics

$$\frac{da}{dt} = -a \sum_{i=1}^n l_i x_i x_{i+1} + r(a_0 - a) \quad \text{and} \quad (11a)$$

$$\frac{dx_i}{dt} = (l_i x_{i+1} a - r)x_i, \quad i = 1, 2, \dots, n, \quad i \text{ mod } n, \quad (11b)$$

which were studied in great detail in a number of previous papers (Eigen and Schuster 1978; Schuster 1978; Schuster et al. 1979, 1980; Hofbauer et al. 1991).

The third case—with (I) and (II) being true—yields a degenerate deterministic solution: All distributions of subspecies with fixed population size  $\sum_{i=1}^n x_i = c$  yield the same solutions and therefore constitute an  $(n - 1)$ -dimensional invariant manifold fulfilling

$$\frac{da}{dt} = -akc + r(a_0 - a) \quad \text{and} \quad (12a)$$

$$\frac{dc}{dt} = akc - cr = (ak - r)c \quad \text{and} \quad (12b)$$

$$\frac{d\gamma}{dt} = (a_0 - \gamma)r \quad \text{with } \gamma = a + c$$

In the absence of selection and cooperation neutral evolution in the sense of Motoo Kimura (1955, 1983) is observed on the mutation axis. For an adequate description of the process a stochastic treatment is required. Random selection of a single arbitrarily chosen subspecies is observed in the no-mutation limit,  $\lim p \rightarrow 0$ . For non-vanishing mutation rates once selected subspecies may be replaced by other subspecies and the mean time of replacement of one randomly selected subspecies is  $\bar{T}_{\text{rep}} = \mu^{-1}$  where  $\mu$  is the mutation rate per generation (Kimura 1955, 1983). Translated into our model, we find for single point mutations  $\mu \approx p$ .

### Master equation and simulation

Reaction (8) can be cast into chemical master equations. The particle numbers of the molecular species,  $[A] = A(t)$  and  $[X_i] = X_i(t)$  with  $i = 1, \dots, n$ , are integers and in the absence of flows they fulfil the conservation relations,  $A(t) + \sum_{i=1}^n X_i(t) = C$ . The variables of the master equation are the probabilities

$$P_m(t) = \text{Prob}(A(t) = m) \quad \text{and} \quad (13)$$

$$P_{s_i}(t) = \text{Prob}(X_i(t) = s_i); \quad i = 1, \dots, n,$$

and the indices are subsumed in an index vector,  $\mathbf{m} = (m, s_1, \dots, s_n)$ , which characterizes the state  $S_{\mathbf{m}}$  of the system. The chemical master equation is based on the

assumption that chemical processes occur one at a time, all jumps involve single steps and the particle numbers change by  $\pm 1$  unless the elementary steps involve more than a single molecule per class. The jumps  $S_{\mathbf{m}'} \rightarrow S_{\mathbf{m}}$  or  $S_{\mathbf{m}} \rightarrow S_{\mathbf{m}'}$  are denoted by the shorthand notation

$$\mathbf{m}' = (m \pm 1, s_1, \dots, s_n) \equiv (\mathbf{m}; m \pm 1) \quad \text{or}$$

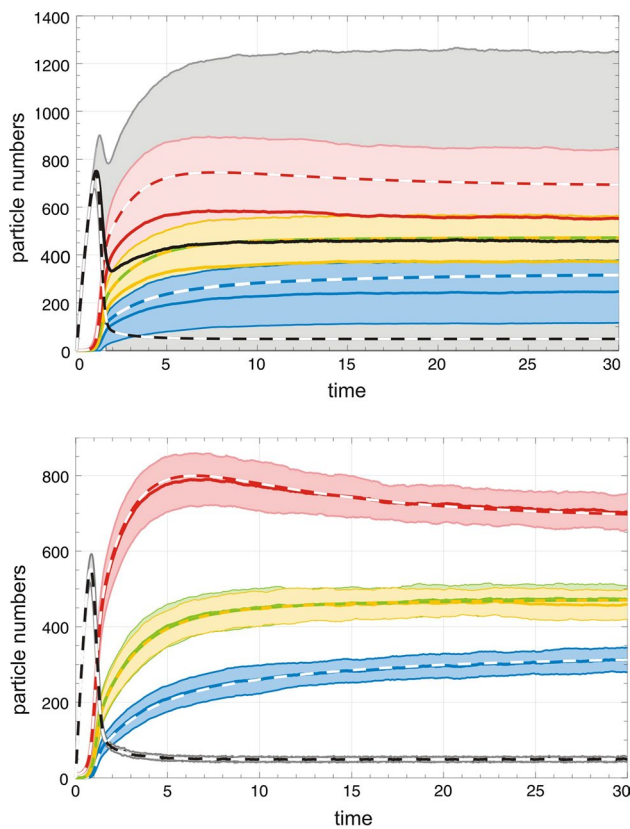
$$\mathbf{m}' = (m, s_1, \dots, s_k \pm 1, \dots, s_n) \equiv (\mathbf{m}; s_k \pm 1).$$

Then the master equation of mechanism (8) takes on the form

$$\begin{aligned} \frac{dP_{\mathbf{m}}}{dt} = & a_0 r (P_{(\mathbf{m}; m-1)} - P_{\mathbf{m}}) + r ((m+1)P_{(\mathbf{m}; m+1)} - mP_{\mathbf{m}}) \\ & + r \sum_{i=1}^n ((s_i+1)P_{(\mathbf{m}; s_i+1)} - s_i P_{\mathbf{m}}) \\ & + \sum_{i=1}^n Q_{ii}(k_i + l_i s_{i+1}) ((m+1)(s_i-1)P_{(\mathbf{m}; m+1, s_i-1)} - m s_i P_{\mathbf{m}}) \\ & + \sum_{i=1}^n \sum_{j=1, j \neq i}^n Q_{ij}(k_j + l_j s_{j+1}) s_j ((m+1)P_{(\mathbf{m}; m+1, s_i-1)} - mP_{\mathbf{m}}). \end{aligned} \quad (14)$$

Each reaction step involving  $S_{\mathbf{m}}$  changes the probability to be in state  $S_{\mathbf{m}}, P_{\mathbf{m}}$ , in two ways: It increases the probability through reactions or pseudoreactions  $S_{\mathbf{m}'} \rightarrow S_{\mathbf{m}}$  and decreases the probability through the reaction steps  $S_{\mathbf{m}} \rightarrow S_{\mathbf{m}'}$ , where  $\mathbf{m}'$  is summed over all states from which  $S_{\mathbf{m}}$  can be reached or vice versa. The two terms in the first line, for example, describe the two pseudoreactions modeling inflow and outflow of the material **A**, and further each reaction is represented by two steps. It is also worth noticing that stoichiometry requires two slightly different replication terms depending on whether the copy is correct or incorrect.

Master equations are easily written down and stationary solutions can be derived by generally applicable techniques as it was shown for the flow equilibrium in "Deterministic and stochastic approaches" but explicit time-dependent solutions are very hard to obtain and known only in exceptionally simple cases [Schuster 2016e, pp 347–568]. Here, we shall use the simulation technique of sampling trajectories introduced already in "Autocatalysis in the batch reactor". Expectation values and second moments of variables can be computed through sampling of trajectories—an example is shown in Fig. 6—but often this approach exceeds the available computational facilities and it is necessary to interpret single trajectories. As examples we consider single trajectories in Fig. 4 and in Fig. 7. The process for convenience starting from an empty flow reactor is split into four phases: (i) establishment of the flow equilibrium of **A**, (ii) random decision on the (quasi)stationary state towards which the trajectory

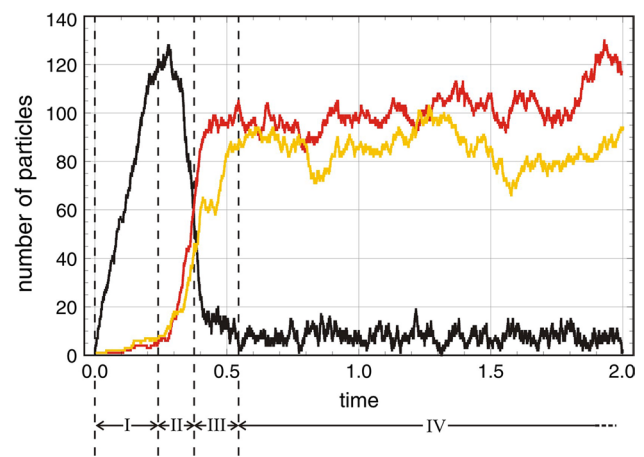


**Fig. 6** Quasispecies formation. The two plots show the formation of quasispecies from an initially empty reactor,  $A(0) = 0$ , with replicators in seeding amounts. The system in the upper plot was initiated to be a single copy of the master sequence,  $X_1(0) = 1$  and  $X_2(0) = X_3(0) = X_4(0) = 0$ . At the beginning the system might be extinguished by a single dilution event  $X_1 \rightarrow \emptyset$  and the high probability of extinction gives rise to enormously broad and overlapping one-standard-deviation bands. The initial values in the lower plot were chosen such that the probability of extinction is zero for all practical purposes:  $X_1(0) = 10$  and  $X_2(0) = X_3(0) = X_4(0) = 0$ . Choice of parameters:  $N = 2000$ ;  $k_1 = 0.011 [M^{-1} t^{-1}]$ ,  $k_2 = k_3 = 0.010 [M^{-1} t^{-1}]$ ,  $k_4 = 0.009 [M^{-1} t^{-1}]$ ,  $r = 0.5 [V t^{-1}]$ . Color code: **A** black, **X<sub>1</sub>** red, **X<sub>2</sub>** yellow, **X<sub>3</sub>** green, and **X<sub>4</sub>** blue. Expectation values are shown as full lines, deterministic solutions as broken lines. Since  $x_2(t) = x_3(t)$  is fulfilled for the parameter values used here the curve is shown as a yellow–green broken line

converges, (iii) the approach towards this (quasi)stationary state and (iv) fluctuations around the (quasi)stationary state. The separation into phases is made possible by the choice of suitable initial conditions.

### Competition, mutation and quasispecies

The bottom face of the three-dimensional parameter space— $\mathcal{A}$  in Fig. 5—is dealing with selection provided by the combination of competition and mutation. Natural selection at zero mutation rate leads to a homogeneous population of the



**Fig. 7** Sequence of phases in the approach towards a quasistationary state for  $n = 2$ . A stochastic trajectory simulating competition and cooperation (“Competition and cooperation”) of two species in the flow reactor is shown in the plot above. The corresponding master equation is derived from (14) by putting  $Q = 1$ . The stochastic process is assumed to start with an empty reactor except seeds for the two autocatalysts  $X_1$  and  $X_2$ . It can be partitioned into four phases: (I) fast raise in the concentration of **A**, (II) a random phase where the decision is made into which final state— $S_0$ ,  $S_1^{(1)}$ ,  $S_1^{(2)}$  or  $S_2$ —the trajectory progresses, (III) the approach towards the final state—here  $S_2$ —and (IV) fluctuations around the values of the (quasi) stationary state. Comparison with the simple autocatalytic process in Fig. 4 reveals great similarity. Choice of parameter values:  $k_1 = 0.099$ ,  $k_2 = 0.101 [M^{-1} t^{-1}]$ ,  $l_1 = 0.0050$ ,  $l_2 = 0.0045 [M^{-2} t^{-1}]$ ,  $a_0 = 200$ ,  $r = 4.0 [V t^{-1}]$ , pseudorandom number generator: *ExtendedCA*, *Mathematica*, seeds  $s = 631$ . Initial conditions:  $A(0) = 0$ ,  $X_1(0) = X_2(0) = 1$ . Color code: **A**(t) black, **X<sub>1</sub>**(t) red, and **X<sub>2</sub>**(t) yellow. The figure is adapted from Schuster (2017a)

fittest subspecies and at non-zero mutation rates this scenario yields selection of a fittest ensemble of subspecies, which has been characterized as quasispecies (Eigen and Schuster 1977; Domingo and Schuster 2016). Precisely, the quasispecies is the stable stationary long-time distribution of a population of subspecies undergoing replication and mutation: A population that consists of several genotypes present in time-dependent concentrations,

$$\Pi(t) = x_1(t)\mathbf{X}_1 \oplus x_2(t)\mathbf{X}_2 \oplus \dots \oplus x_n(t)\mathbf{X}_n \quad (15)$$

has the stationary solution,  $\lim_{t \rightarrow \infty} \Pi(t) = \bar{\mathbf{Y}} = \bar{x}_1\mathbf{X}_1 \oplus \bar{x}_2\mathbf{X}_2 \oplus \dots \oplus \bar{x}_n\mathbf{X}_n$ , which is called the *quasispecies*  $\bar{\mathbf{Y}}$ .

*Deterministic quasispecies* The deterministic or *continuous quasispecies* represents the unique deterministic long-time solution of the replication mutation problem, which is described in the flow reactor by the ODE (Schuster and Sigmund 1985):

$$\frac{da}{dt} = -a \sum_{i=1}^n k_i x_i + r(a_0 - a) \quad (16a)$$

$$\frac{dx_i}{dt} = a \sum_{j=1}^n Q_{ij} k_j x_j - r x_i, \quad i = 1, \dots, n. \tag{16b}$$

All early works on quasispecies dynamics were performed with the constraint of constant population size:  $\sum_{i=1}^n x_i(t) = c = \text{const}$  with  $a(t) = a_0$  and  $f_i = k_i a_0$  that gives rise to the differential equation

$$\frac{dx_i}{dt} = \sum_{j=1}^n Q_{ij} f_j x_j - \phi(t) x_i, \tag{16'}$$

with  $\phi = \frac{\sum_{j=1}^n f_j x_j}{\sum_{j=1}^n x_j}; i = 1, \dots, n,$

which fulfils  $dc/dt = \sum_{i=1}^n dx_i/dt = 0$ . Equations (16) and (16') have identical solutions in normalized variables  $\xi_i(t) = x_i / \sum_{i=1}^n x_i(t)$  but the stability properties of the corresponding stochastic systems are different. [For more details see (Thompson 1974; Jones et al. 1976; Ebeling and Mahnke 1979; Jones and Leung 1981; Schuster and Sigmund 1985)].

The quasispecies as a function of the mutation rate parameter  $\bar{Y}(p)$  begins at  $p = 0$  as a homogeneous population containing exclusively the fittest subspecies  $\mathbf{X}_m$  and becomes a distribution of subspecies at nonzero  $p$ . This distribution consists of a most frequent sequence called master sequence, which is surrounded by a mutant cloud. Commonly, the most frequent or master sequence is also the fittest one, but this is not necessarily so: In case of strong stationary mutation flow from the mutant cloud to the master sequence, a less fit master may outgrow a fitter sequence with less efficient mutational backflow. At further increase in  $p$  the distribution broadens, and eventually ends up as the uniform distribution  $\mathcal{U}: \{\hat{x}_1 = \hat{x}_2 = \dots = \hat{x}_n\}$  at  $\hat{p} = 1 - (\kappa - 1)\hat{p}/\kappa = 1/\kappa$  where  $n = \kappa^l$  and  $\hat{x}_i = 1/\kappa^l$  ( $i = 1, \dots, \kappa^l$ ) for sequences of chain length  $l$  over an alphabet with  $\kappa$  letters.<sup>6</sup> At the mutation rate  $p = \hat{p}$ , correct and wrong digits are incorporated with the same frequency. The uniform distribution is the dynamical answer to the absence of any preference for nucleotide assignment: All subspecies in the stationary distribution occur with the same frequency. The quasispecies in the intermediate range is determined by the fitness landscape—the distribution of fitness values  $f_i$  in sequence space, by the move set of allowed mutations as well as the mutation rate  $p$ . Typically, sharp transitions occur at some critical mutation rates  $p = p_{tr}$ : The distribution changes smoothly from  $p = 0$  to  $p = p_{tr}$ , where the distribution turns abruptly into another distribution. At the transition with the largest  $p$ -value  $p = 1/\kappa > p_{cr} > p_{tr}$ , the quasispecies becomes

an approximate uniform distribution. An explanation is straightforward: Above this critical transition, mutations occur too often to sustain sufficiently accurate reproduction of the template sequence and the result is *random replication*: In the long run, every sequence is obtained with the same probability. The transition has been characterized as *error threshold* (Eigen 1971; Eigen and Schuster 1977) since evolutionary dynamics does not sustain a structured long-time population at higher error rates,  $p > p_{cr}$ . On typical fitness landscapes, the error threshold sharpens with increasing chain length  $l$  and becomes a first-order phase transition in the limit  $l \rightarrow \infty$  (Tarazona 1992; Park et al. 2010; Huang et al. 2017).<sup>7</sup>

*Discrete quasispecies* In the continuous description, the quasispecies contains all species at finite positive concentrations no matter how small the concentrations might be. Dealing with less than one molecule per reactor volume, however, is unrealistic. Numbers of molecules are non-negative integers,  $X_i \in \mathbb{N}$ , subspecies distributions are truncated in reality and we call them stochastic or *discrete quasispecies*:

$$\tilde{\mathbf{Y}} = \bar{X}_1 \mathbf{X}_1 \oplus \bar{X}_2 \mathbf{X}_2 \oplus \dots \oplus \bar{X}_n \mathbf{X}_n \quad \text{with} \quad \bar{X}_i = \begin{cases} \lfloor \bar{x}_i \rfloor & \text{if } \bar{x}_i \geq 1 \\ 0 & \text{if } \bar{x}_i < 1 \end{cases} \tag{17}$$

The discreteness of the stochastic variables leads to a modification of the scenario for the mutation dependence of quasispecies  $\bar{Y}(p)$ . In the mutation-free system,  $p = 0$ , *survival of the fittest* is observed and the quasispecies consists of a single sequence:  $\tilde{\mathbf{Y}} = \bar{X}_m \mathbf{X}_m = N \mathbf{X}_m$  with  $m$  indicating maximal fitness,  $f_m = \max\{f_i; i = 1, \dots, n\}$ , and  $N$  being the population size. At sufficiently small values of the mutation rate parameter  $p$ , no subspecies except the master exceeds the threshold  $x_i \geq 1$  and the discrete quasispecies  $\tilde{\mathbf{Y}}$  still consists of a single fittest genotype  $\mathbf{X}_m$ . The scenario in the small mutation rate regime—populations are almost always homogeneous except short non-stationary periods during which advantageous mutations take over the population—is tantamount to the strong selection–weak mutation scenario discussed in evolutionary biology (Gillespie 1983; Joyce et al. 2008).<sup>8</sup> If the  $p$ -value is so large that for one or

<sup>7</sup> We mention that some model landscapes like the additive landscape or the multiplicative landscape sustain smooth rather than sharp transitions (Wiehe 1997). For a recent update and a review of the state of the art in the relation between fitness landscape and selection dynamics see (Schuster 2016d).

<sup>8</sup> Biologists (Dean and Thronton 2007; Sniegowski and Gerrish 2010) and computer scientists commonly distinguish strong and weak mutation. The weak mutation scenario assumes that adaptive mutations are sufficiently rare and do not interfere with the selection process but initiate replacements of currently fittest genotypes by still fitter variants. The strong mutation scenario is characterized by sufficiently large values of  $p$  that give rise to the quasispecies dynamics described here [for details see (Domingo and Schuster 2016)] and to mutation induced cooperative dynamics ("[Cooperation and mutation](#)").

<sup>6</sup> The value  $\kappa = 2$  is used here and applies for binary sequences. For the natural AUGC-nucleotide alphabet we have  $\kappa = 4$ .

more subspecies  $\bar{x}_i \geq 1$  is fulfilled, selection of a family of genotypes is observed:  $\tilde{\mathbf{Y}}$  consist of a master genotype  $\mathbf{X}_m$ , together with a stationary distribution of sufficiently frequent mutants  $\mathbf{X}_i$  ( $i \neq m$ ). The quasispecies becomes broader with increasing mutation parameter  $p$  until a threshold value  $p_{cr}$  is reached above which error propagation does not sustain a stationary state and the population  $\Pi$  drifts randomly through sequence space (Huynen et al. 1996; Higgs and Derida 1991). Interestingly, the critical mutation rate  $p_{cr}$  can be derived from the continuous quasispecies theory.

As calculations of the time dependencies of the first two moments of the probability distribution of subspecies in the population,  $P^{(n)}(t)$ , reveal (Jones and Leung 1981) (16') is only marginally stable:

$$\frac{d\langle N \rangle}{dt} = \sum_{i=1}^n f_i \langle X_i \rangle - \langle \phi(\mathbf{X}(t)) N \rangle = 0, \quad (18a)$$

$$\begin{aligned} \frac{d\langle N^2 \rangle}{dt} &= 2 \sum_{i=1}^n f_i \langle X_i N \rangle - 2 \langle \phi(\mathbf{X}(t)) N^2 \rangle + \\ &+ 2 \sum_{i=1}^n f_i \langle X_i \rangle - \frac{d\langle N \rangle}{dt} = 2 \sum_{i=1}^n f_i \langle X_i \rangle, \text{ and} \end{aligned} \quad (18b)$$

$$\text{var}(N) = \langle N^2 \rangle - \langle N \rangle^2 = 2 \int_0^t \sum_{i=1}^n f_i \langle X_i(\tau) \rangle d\tau \quad (18c)$$

The time derivative of the first moment vanishes as expected since the condition of a constant population size is fulfilled by the differential equation (16'). The variance, however, increases with time since the integrand in (18b) is always positive. After sufficiently long time, the fluctuation band becomes so large that the expectation value is irrelevant for the description of the system. The instability in Eigen's quasispecies equation (Eigen 1971) is well known from similar problems: (i) the neutral birth-and-death process with equal birth and death parameters,  $\lambda = \mu$ , and (ii) the Wiener process. In both cases a constant expectation value is jeopardized by a variance that grows linearly with time. Stability against fluctuation is easily introduced into (16'): one needs only to give up the condition of strictly constant population size and to replace the denominator in  $\phi(t)$  by a constant  $\Theta$ ,

$$\phi(t) = \sum_{j=1}^n f_j x_j / \Theta,$$

where  $\Theta$  is the population size towards which the population converges after sufficiently long time. The approach chosen here, the implementation of a flow reactor as a physical device, yields a stable system as well.

Fluctuations at small particle numbers have different origins ("Autocatalysis in the batch reactor"): (i) conventional thermal fluctuations, (ii) enhanced fluctuations related to autocatalytic self-enhancement, and (iii) anomalous fluctuations in the stochastic variables arising from two or more quasistationary states. The standard deviations  $\sigma(t)$  fulfil the  $\sqrt{N}$ -law for the resource  $\mathbf{A}$  but are larger for the autocatalysts  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ , and  $\mathbf{X}_4$ . The stationary states of the stochastic system are extinction  $S_0$  and quasispecies  $S_1$ . Fig. 6 shows a typical example: The anomalous fluctuations in the upper plot are in full analogy to first-order autocatalysis (Fig. 4). Since the initial condition  $X_1(0) = 1$  was chosen, one outflow step may extinguish the population and the probability of dying out is indeed as large as 20%. The fluctuations bands are extremely broad, and large differences between the deterministic solution and the corresponding expectation values are observed. In the lower plot initial conditions  $X_1(0) = 10$  were chosen, which are sufficient to reduce the contributions of anomalous fluctuations practically to zero. Then, for a population size of  $N = 2000$  the concentration  $a(t)$  coincides with the expectation value  $E(A(t))$  for all practical purposes and the fluctuations fall into a typical  $\pm\sqrt{N}$ -band. The autocatalysts,  $X_1, \dots, X_4$ , show broader than usual fluctuation bands because of self-enhancement as we saw in the intermediate range of the first-order autocatalytic reaction (Fig. 2). For long-times the standard deviation  $\sigma(t)$  stays large in the quasispecies because the flow reactor is an open system and does not approach an equilibrium state (see also Fig. 4).

It is worth recalling what means stochasticity for quasispecies: (i) continuous concentrations are replaced by discrete particle numbers, (ii) fluctuations replace single line trajectories by bands within which trajectories follow a probability distribution, (iii) subspecies can be diluted out of the flow reactor and if this happens for all subspecies the population goes extinct giving rise to anomalous fluctuations, and (iv) error thresholds introduce random reproduction that is closely related to Motoo Kimura's random drift. An increase in the error rate up to the error threshold leads to broadening of the mutant spectrum surrounding the master sequence. Above the thresholds, the populations migrate by random drift through sequence space.

## Competition and cooperation

The kinetic equations for replication describing the template induced, uncatalyzed and catalyzed processes are obtained from Eq. (1) by neglect of mutation. Then the kinetic differential equations for competition and cooperation of competitors result from (9) by setting  $Q = \emptyset$ :

$$\frac{da}{dt} = -a \sum_{i=1}^n (k_i + l_i x_{i+1}) x_i + r(a_0 - a) \text{ and} \quad (19a)$$

**Table 2** Asymptotically stable stationary states of Eq. (19) with  $n = 3$  Schuster (2016)

Symbol	Stationary values				Stability range
	$\bar{a}$	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	
$S_0$	$a_0$	0	0	0	$0 \leq a_0 \leq \gamma_3$
$S_1^{(3)}$	$\gamma_3$	0	0	$a_0 - \gamma_3$	$\gamma_3 \leq a_0 \leq \gamma_3 + \gamma_{32}$
$S_2^{(1)}$	$\gamma_3$	0	$a_0 - \gamma_3 - \gamma_{32}$	$\gamma_{32}$	$\gamma_3 + \gamma_{32} \leq a_0 \leq \gamma_3 + \gamma_{32} + \gamma_{31}$
$S_3$	$\alpha$	$\alpha_3$	$\alpha_1$	$\alpha_2$	$\gamma_3 + \gamma_{32} + \gamma_{31} \leq a_0$

The four stationary states are ordered with respect to increasing  $a_0$ -values of their asymptotically stable regime. The relations  $k_1 < k_2 < k_3$  and  $l_1 > l_2 > l_3$  between the rate parameters were assumed. The abbreviations  $\gamma_3 = r/k_3$ ,  $\gamma_{31} = (k_3 - k_1)/l_1$ , and  $\gamma_{32} = (k_3 - k_2)/l_2$  are used for the combination of parameters. For the cooperative state  $S_3$  the stationary concentration of **A** is obtained as one root of a quadratic equation (20) with two combinations of the rate parameters,  $\psi = \sum_{i=1}^3 k_i/l_i$  and  $\phi = \sum_{i=1}^3 1/l_i$ , and  $\alpha_i = (r - k_i\alpha)/(l_i\alpha)$  for  $i = 1, 2, 3$  and  $\alpha = \bar{a}_1$  from (20). The existence of the non-trivial stationary state requires a sufficiently small flow rate:  $r \leq (a_0 + \psi)^2/4\phi$

$$\frac{dx_i}{dt} = \left( a(k_i + l_i x_{i+1}) - r \right) x_i; \quad i = 1 \dots, n; \quad i \text{ mod } n. \tag{19b}$$

Both equations contain terms of different molecularity in **X** and this has the consequence that the dynamic behavior depends strongly on the total concentration  $c = \sum_{i=1}^n x_i$ , which in turn is determined by the amount of available resources,  $a_0$ : At sufficiently low concentration, the first-order terms,  $k_i x_i$ , dominate whereas the second-order terms,  $l_i x_{i+1} x_i$  become important at high concentrations. No direct analytical solutions are available but exhaustive qualitative analysis is possible and the concentrations of the stationary solutions,  $\bar{a}, \bar{x}_i$  ( $i = 1, \dots, n$ ), can be computed analytically Schuster (2016a, d) (Table 2).

The basis of the calculations of stationary states is the factorability of the r.h.s. of (19b). The relation  $\bar{x}_i \cdot ((k_i + l_i \bar{x}_{i+1}) \bar{a} - r) = 0$  with  $i \text{ mod } n$  is compatible with stationary states that fulfil either

$$\bar{x}_i^{(i)} = 0 \text{ or } \bar{x}_i^{(ii)} = \frac{1}{l_{i-1}} \left( \frac{r}{\bar{a}} - k_{i-1} \right), \quad \forall i = 1, \dots, n,$$

and the conservation condition  $\bar{a} + \sum_{i=1}^n \bar{x}_i = a_0$ . In total there are  $2^n$  possible states out of which a single one is asymptotically stable for a given set of parameters. In Table 2 the results for  $n = 3$  are shown. The number of subspecies present at the stationary state,  $n_s$ , is suitable for characterizing the state:  $n_s = 0$  is the state of *extinction*,  $n_s = 1$  is the state of *selection*,  $n_s = 2$  is the state of *exclusion*, etc., and finally  $n_s = 3 = n$  occurs at the state of *cooperation* where all three subspecies are present. The numbers of long-time subspecies  $n_s$  depend on the resource input  $a_0$ . The relative size ordering of the parameters determines the identity of the selected, excluded, etc., subspecies. For the calculations shown in Table 2 the orderings  $k_1 < k_2 < k_3$  and  $l_1 > l_2 > l_3$  were applied and hence **X**<sub>3</sub> is selected and **X**<sub>1</sub> is excluded. Considering steady states as functions of the available resources, the state of *extinction*  $S_0$  comes first at small

$a_0$  and is stable for  $a_0 < r/k_3$ , the state of *selection* of **X**<sub>3</sub>,  $S_1^{(3)}$ , is stable in the range  $r/k_3 < a_0 < r/k_3 + (k_3 - k_2)/l_2$ , followed by *exclusion* of **X**<sub>1</sub>,  $S_2^{(1)}$ , in the range  $k_3 + (k_3 - k_2)/l_2 < a_0 < k_3 + (k_3 - k_2)/l_2 + (k_3 - k_1)/l_1$ . Above this value for  $a_0$  *cooperation* of all three subspecies is observed provided the flow rate is not too large:  $r < r_{cr} = (a_0 + \psi)^2/4\phi$  with  $\psi = \sum_{i=1}^3 k_i/l_i$  and  $\phi = \sum_{i=1}^3 1/l_i$ . The free concentration of **A** is obtained as solution of a quadratic equation<sup>9</sup>

$$\bar{a}_{1,2} = \frac{1}{2} \left( a_0 + \psi \mp \sqrt{(a_0 + \psi)^2 - 4r\phi} \right), \tag{20}$$

where the minus sign corresponds to the cooperative state  $S_3$ . The second solution belongs to an unstable state  $S'_3$ , which separates the basins of attraction of the states of cooperation and extinction. Above the critical flow rate,  $r > r_{cr}$  the states  $S_3$  and  $S'_3$  do not exist. For  $n > 3$ , the situation becomes more complex since solutions may oscillate. Many systems with  $n = 4$  have oscillations with very weak damping factors,  $n \geq 5$  commonly leads to undamped oscillations. The properties of these systems have been discussed extensively in previous publications to which we refer here (Schuster 2016e; Schuster and Sigmund 1985; Schuster 1978).

Stochasticity has common effects on competition–cooperation systems like thermal fluctuations and fluctuation through autocatalytic enhancement. In addition, there are many more quasistationary states than the asymptotically stable states of the deterministic system. For example, states in which less efficient subspecies are selected show up as quasistationary states as well. In case of the smallest possible system with  $n = 2$  and  $k_2 > k_1$ , we have four states: (i) the absorbing boundary as state of extinction  $S_0$ , (ii) the state of natural selection  $S_1^{(2)}$  where the fittest variant **X**<sub>2</sub> is

<sup>9</sup> It is worth mentioning that Eq. (20) and the solutions for  $\bar{a}_{1,2}$  are valid for arbitrary  $n$  provided the summations in  $\psi$  and  $\phi$  are adjusted accordingly.

**Table 3** Probabilities to reach quasistationary states in the cooperative regime with  $n = 2$  and different initial conditions

Initial values		Counted numbers of states in final outcomes			
$X_1(0)$	$X_2(0)$	$N_{S_0}$	$N_{S_1^{(1)}}$	$N_{S_1^{(2)}}$	$N_{S_2}$
1	1	$385.1 \pm 23.6$	$1481.0 \pm 36.8$	$1719.6 \pm 37.8$	$6414.3 \pm 53.8$
2	1	$77.4 \pm 9.1$	$1822.6 \pm 41.6$	$367.6 \pm 17.0$	$7733.3 \pm 38.3$
1	2	$71.6 \pm 8.5$	$280.6 \pm 20.0$	$2075.8 \pm 28.9$	$7572.0 \pm 39.2$
3	1	$15.0 \pm 2.9$	$1900.4 \pm 30.9$	$74.69 \pm 10.0$	$8009.0 \pm 35.3$
1	3	$14.0 \pm 3.7$	$53.1 \pm 4.8$	$2180.5 \pm 48.4$	$7752.3 \pm 53.8$
2	2	$14.9 \pm 2.6$	$303.7 \pm 16.0$	$354.5 \pm 23.8$	$9326.8 \pm 44.9$
3	3	0	$70.2 \pm 10.0$	$106.2 \pm 10.9$	$9823.4 \pm 15.7$
4	4	0	$12.1 \pm 2.6$	$28.0 \pm 5.0$	$9959.9 \pm 6.4$
5	5	0	$2.5 \pm 1.1$	$6.3 \pm 2.6$	$9991.2 \pm 3.0$

The table provides probabilities of occurrence for all four possible long-term states: extinction  $S_0$ , selection of  $\mathbf{X}_1$  in  $S_1^{(1)}$ , selection of  $\mathbf{X}_2$  in  $S_1^{(2)}$ , or cooperation  $S_2$ . The counted numbers of events are sample means and unbiased standard deviations calculated from ten packages, each of them containing 10 000 trajectories computed with identical parameters and initial conditions, and differing only in the sequence of random events determined by the seeds of the pseudorandom number generator (*Extended CA, Mathematica*). Choice of parameters:  $k_1 = 0.09[\text{M}^{-1}\text{t}^{-1}]$ ,  $k_2 = 0.11[\text{M}^{-1}\text{t}^{-1}]$ ,  $l_1 = 0.0011[\text{M}^{-2}\text{t}^{-1}]$ ,  $l_2 = 0.0009[\text{M}^{-2}\text{t}^{-1}]$ ,  $a_0 = 200$ ,  $r = 0.5[\text{V t}^{-1}]$ . Initial value  $A(0) = 0$ . Probabilities are obtained by multiplication by  $10^{-4}$ .

**Table 4** Long-time behavior in the cooperation-mutation system with  $n = 3$  and different initial conditions

Mutation $p$	Initial values			Expectation values at $t = 30$				Counts $P(S_3)$
	$X_1(0)$	$X_2(0)$	$X_3(0)$	$E(A)$	$E(X_1)$	$E(X_2)$	$E(X_3)$	
0	1	1	1	279.3	44.56	36.06	39.58	0.301
0	2	2	2	81.53	117.7	94.9	105.4	0.814
0	4	4	4	2.023	146.1	119.6	132.2	0.996
0	10	10	10	0.376	147.0	120.0	132.6	1.0
0	Deterministic			0.377	147.0	120.3	132.3	1
0.05	1	1	1	177.2	81.66	68.16	72.64	0.432
0.05	2	2	2	28.57	136.1	113.8	121.5	0.937
0.05	4	4	4	0.383	147.0	122.2	130.4	1.0
0.05	Deterministic			0.377	146.7	122.3	130.5	1
0.1	1	1	1	161.6	86.95	74.39	76.96	0.599
0.1	2	2	2	28.57	136.1	113.8	121.5	0.957
0.1	4	4	4	0.383	147.0	122.2	130.4	1.0
0.1	Deterministic			0.377	146.7	122.3	130.5	1

The table provides expectation values at the time of the end of the simulation ( $t_{\text{end}} = 30$ ) for different mutation rate parameters  $p$  and different initial conditions. Choice of parameters:  $l_1 = 0.011$ ,  $l_2 = 0.010$ ,  $l_3 = 0.009[\text{M}^{-2}\text{t}^{-1}]$ ,  $a_0 = 400$ ,  $r = 0.5[\text{V t}^{-1}]$ . Initial value  $A(0) = 0$ .

selected, (iii) the state of selection of the less fit subspecies  $\mathbf{X}_1$ , and eventually (iv) the state of cooperation  $S_2$ . The relative stabilities of the individual states are reflected by the probabilities to reach these states by randomly chosen trajectories (Table 3). Parameters for the calculations shown in the table were chosen such that the corresponding deterministic system is situated in the cooperative domain in parameter space, and indeed the approach towards the cooperative state  $S_2$  has always the largest probability. Again, initial particle numbers around  $X_1(0) + X_2(0) = 4$  are sufficient for strong dominance of the state corresponding to the deterministic

solution. In case of  $n = 3$  we present expectation values of the four stochastic variables,  $A(t)$  and  $X_i(t)$  ( $i = 1, 2, 3$ ) at some predefined time of the simulation end,  $t_{\text{end}}$  for different initial conditions and mutation rate parameters  $p$ . The values for  $p = 0.0$  refer to the pure competition–cooperation case discussed here (Table 4). As expected extinction plays a major role at small initial values,  $X_i(0) = 1, 2, 3$  ( $i = 1, 2, 3$ ), but  $X_i(0) = 4$  is already sufficient for coming close to the expectation values obtained for large initial values,  $X_i(0) = 10$  is enough for reaching the deterministic values for practical purposes.

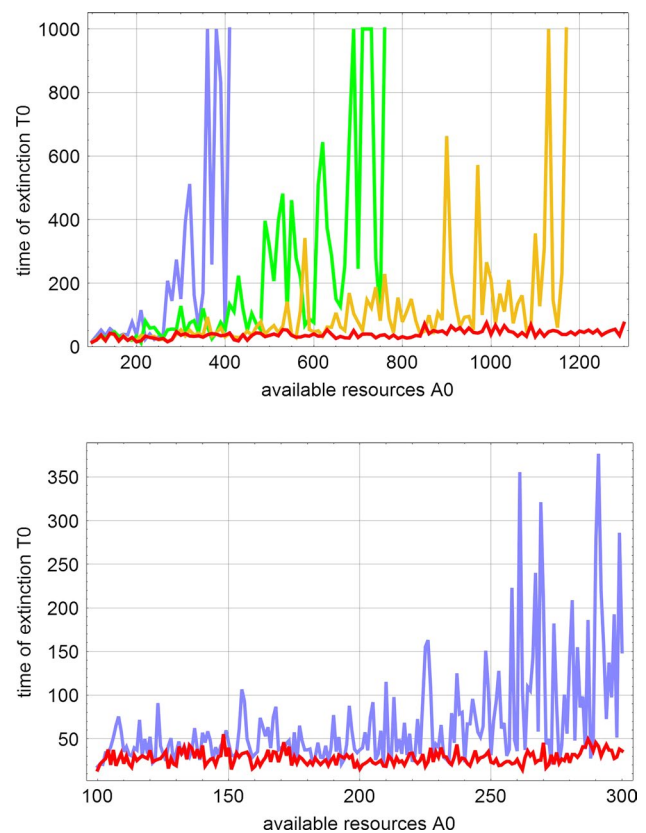


In systems with  $n \geq 4$  deterministic and stochastic solution curves oscillate. The solutions of the ODE's are different for  $n = 4$  where weakly damped oscillations occur and for  $n \geq 5$  showing undamped *relaxation oscillations* (Phillipson et al. 1984). In the stochastic approach, the systems die out after population numbers of individual subspecies went beyond  $X(t) = 1$ , and for sufficiently large population sizes, four-membered systems may survive for very long time whereas systems with five or more members go extinct. Cases with  $n = 5$  are well suited for studying the transition from selection to cooperative dynamics through increase of the parameter ratio  $h/f = l/k$  (Fig. 9). At dominant competition or small  $h$ -values the system approaches selection of the fittest as long-time solution. The upcoming role of cooperation in a series of systems with increasing parameters  $l$  can be nicely illustrated by a series of plots of trajectories from selection to somewhat chaotic looking intermediate scenarios and further to oscillatory hypercycle dynamics (see Fig. 9 where a nonzero mutation rate parameter  $p$  was chosen and where accordingly quasispecies formation instead of selection is observed).

### Cooperation and mutation

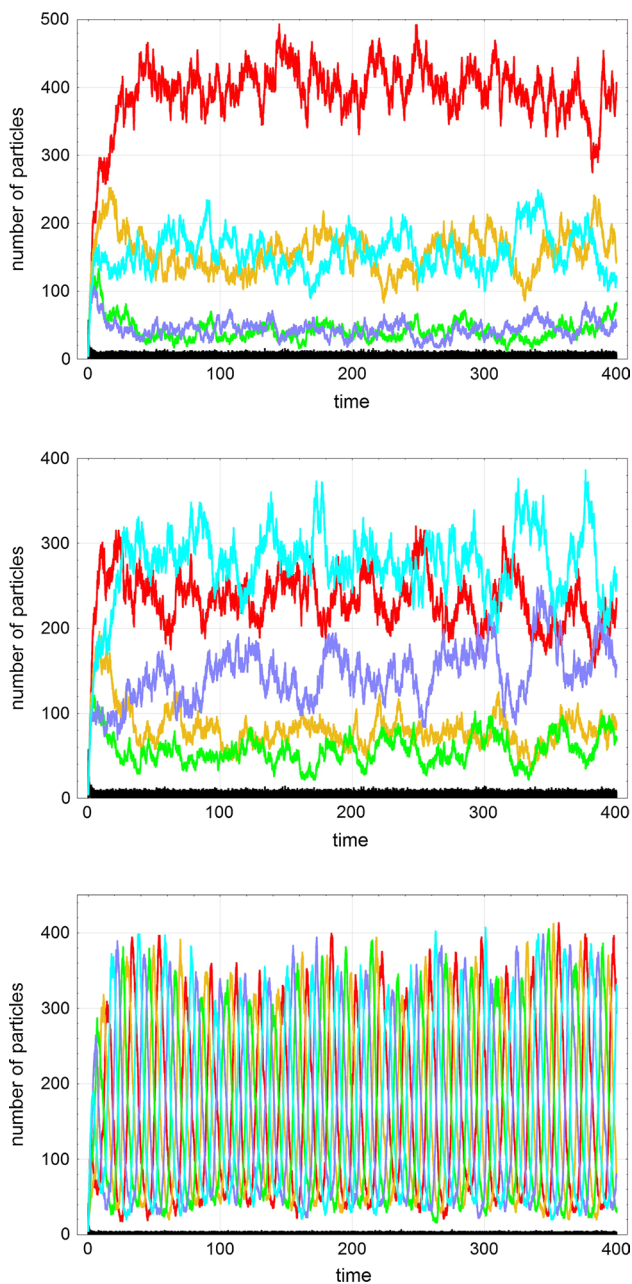
The combination of cooperation and mutation reveals a less common role of mutation in addition to the creation of diversity through variation. In principle, mutation can reintroduce extinguished subspecies into the population. Here, we shall focus on this aspect and, in particular, study the influence of the mutation rate parameter  $p$  on extinction times. To study the role of mutation in low-dimensional cooperative systems ( $n = 2, 3$ ) expectation values of the stochastic variables  $A(t)$ ,  $X_1(t)$  and  $X_2(t)$ , or  $X_2(t)$  and  $X_3(t)$  were calculated at predefined times ( $t_{\text{end}}$ ) and compared with the probabilities of trajectories to end up in the cooperative state,  $S_2$  or  $S_3$ , respectively. For the initial conditions  $X_1(0) = X_2(0) (= X_3(0)) > 4$  the results are practically indistinguishable from the deterministic values. An increase in the mutation rate parameter  $p$  shows the expected influence: Extinguished subspecies can be reintroduced and this increases the probability of reaching the cooperative state,  $S_2$  or  $S_3$ . The case  $n = 3$  is shown as an example in Table 4.

The oscillating systems are more difficult to investigate. Here, we consider the time of extinction of the entire population as a function of the mutation rate and the available resources,  $T_0(p, A_0)$ . The results are shown in Fig 8: The extinction times  $T_0$  show very strong scatter and their appearance is dependent on the resolution of the calculations. By resolution, we mean here the number of molecules  $\mathbf{A}$  in  $A_0$  between two neighboring points. The highest resolution is achieved when the calculations are performed with every (integer) number of  $A_0$  molecules, e.g.,  $\Delta A_0 = 1$  yields 100, 101, 102, ... Computations at somewhat lower resolutions



**Fig. 8** Times to extinction as a function of available resources in the five membered cooperative system ( $n = 5$ ). Extinction times  $T_0$  of the populations  $\Pi$  are shown for different amounts of available resources measured as inflow concentrations  $a_0$  or  $A_0$  when expressed in numbers of molecules per unit volume. The upper diagram presents the data at a resolution of ten molecules ( $\Delta A_0 = 10; 100, 110, 120, \dots$ ) for four different values of the mutation rate parameter:  $p = 0.0000$  (red),  $p = 0.0005$  (yellow),  $p = 0.0010$  (green), and  $p = 0.0020$  (blue).  $T_0$ -values above 1000 are truncated at this value. The lower diagram shows the two plots ( $p = 0.0000$  (red) and  $p = 0.0020$  (blue)) at the highest possible resolution ( $\Delta A_0 = 1$ ). Choice of other parameters:  $r = 0.5[V^{-1}t^{-1}]$ ,  $l_1 = l_2 = l_3 = l_4 = l_5 = 0.01[M^{-2}t^{-1}]$ . Pseudorandom number generator: *Extended CA, Mathematica*, seed:  $s = 491$ . Initial conditions:  $A(0) = 0$ ,  $X_1(0) = X_2(0) = X_3(0) = X_4(0) = X_5(0) = 5$

are less time consuming and provide in essence the same results. The plots shown in Fig. 8 show enormous scatter but, nevertheless, allow for drawing two conclusions: (i) In the mutation-free case the extinction time  $T_0$  is independent of the amount of available resources up to a value  $A_0 \approx 1300$ , and (ii) for non-zero mutation rates a kind of noisy or stochastic threshold phenomenon is observed. Considering the noisy function  $T_0(p, A_0)$  and taking  $A_0$  at the first value of  $T_0(p, A_0) \geq 1000$  we find for the parameter values applied in Fig. 8:  $A_0^{(T_0 \geq 1000)} = A_0^{(\text{cr})} = 1130, 690, 360$  for the mutation rates  $p = 0.0005, 0.0010, 0.0020$ , respectively. As expected the threshold moves to lower  $A_0^{(\text{cr})}$ -values with increasing mutation rate. The behavior of the extinction times  $T_0(A_0)$  is



**Fig. 9** The transitions from competition to cooperation in the system with  $n = 5$ . The three plots show single trajectories for three different scenarios in the flow reactor: (i, topmost plot) the quaspecies scenario with a dominant master sequence ( $\mathbf{X}_1$ ), (ii, middle plot) an intermediate scenario with irregular dynamics and two dominating species ( $\mathbf{X}_1$  and  $\mathbf{X}_5$ ), and (iii, bottom plot) the stochastic hypercycle scenario with irregular, undamped oscillations. Choice of parameter:  $k_1 = 0.150$ ,  $k_2 = k_5 = 0.125$ ,  $k_3 = k_4 = 0.100$  [ $M^{-1} t^{-1}$ ],  $l_1 = l_2 = l_3 = l_4 = l_5 = l$ ,  $l = 0.0$  (topmost plot),  $l = 0.002$  (middle plot),  $l = 0.01$  [ $M^{-2}t^{-1}$ ] (bottom plot),  $a_0 = 800$ ,  $r = 0.5$  [ $V t^{-1}$ ],  $p = 0.075$ , pseudorandom number generator: *ExtendedCA*, *Mathematica*, seeds  $s = 491$ . Initial conditions:  $A(0) = 0$ ,  $X_1(0) = X_2(0) = X_3(0) = X_4(0) = X_5(0) = 5$ . Color code:  $A(t)$  black,  $X_1(t)$  red,  $X_2(t)$  yellow,  $X_3(t)$  green,  $X_4(t)$  blue, and  $X_5(t)$  cyan

similar for  $n = 4$  but the critical concentrations  $A_0^{(cr)}$  for the different  $p$ -values lie much closer together and the analysis is more difficult. Considering survival at constant resources  $A_0$  reveals a *mutation threshold* above which the population survives to long times.

Hypercycle extinction is an example that reflects well the expected increase in lifetime with increasing mutation rate. One general remark nevertheless is important: This mechanism of reintroduction of extinguished subspecies requires that template and mutant are close relatives and that the Hamming distance between them is not too large. What we need in reality, however, is not a perfect revertant being genetically identical to the lost original, we need only a subspecies that can replace the original with respect to its phenotypic function. Suppression of deleterious mutations (Gorini and Beckwith 1966; Hartman and Roth 1973; Prelich 1999) as well as the relation between protein sequence, structure and functional efficiency (Albery and Knowles 1976, 1977) have been extensively studied in the last decades of the twentieth century.

### The complete model

Completion of the model brings together the three faces of the coordinate system in Fig. 5 and is concerned with an analysis of the dynamics in the interior. An appropriate strategy for analyzing the interior consists in choosing certain type of behavior on one of the three faces and increasing the third parameter from zero to the value of interest. Raising the third parameter will change the dynamic behavior either gradually or in threshold-like manner or stepwise through a cascade of bifurcations. Illustrative prototype examples are seen through rising the mutation rate in competitive or cooperative reproduction, or with the introduction of cooperation into Darwinian systems.

The characteristic dependence of the population dynamics on  $n$ , the number of subspecies, prevails also in the full model. For small numbers ( $n = 2, 3$ ) and  $p = 0$  the transition from the competitive to the cooperative system has been discussed in "Competition and cooperation". An increase of the cooperation parameter  $h = lA(t)$  leads in steps from selection of the fittest to a cooperative state with all subspecies present. Oscillating systems ( $n \geq 4$ ) are more spectacular since the hypercycles are unstable at  $p = 0$  in the stochastic approach and raising the cooperation parameter  $h$  leads from selection to extinction. In the intermediate parameter range where the deterministic system shows stepwise increase in the number of coexisting subspecies ( $1, 2, \dots, n$  or, expressed phenomenologically, selection, exclusion, ..., cooperation) the stochastic approach yields highly irregular dynamics with different numbers of non-extinguished subspecies whereby the number of species present increases

with increasing values of the cooperation parameter  $h$ . The scenario in parameter space is completed by considering the series of states at different mutation rate parameters  $p$ . The case  $n = 5$ , which for large  $h$ -values leads to undamped oscillations with a stochastic contribution to the amplitude, was chosen to facilitate the illustration (Fig. 9): At zero mutation rate  $p = 0$  the series with increasing  $h$  is selection  $\rightarrow$  irregular dynamics with two species  $\rightarrow$  irregular dynamics with three species  $\rightarrow \dots \rightarrow$  extinction. At intermediate  $p$ -values, we find quasispecies  $\rightarrow$  irregular dynamics  $\rightarrow$  oscillations with highly irregular spacings. At high mutation rates, we are dealing with quasispecies  $\rightarrow$  irregular dynamics with increasing numbers of dominant subspecies  $\rightarrow$  stochastic hypercycle dynamics (Fig. 9).

## Concluding remarks

The model presented here has been conceived with modular structure in the sense that different mechanisms can be applied for each of the three basic components. Here, it has been presented in its simplest form. Each of the three modules, competition, cooperation and variation, can be made arbitrarily complex. Variation, for example, can be extended to include more elaborate mutation mechanisms and recombination as well as environmental influences. Even in case of viruses the reproduction mechanism is commonly much more elaborate and comprises a whole molecular machinery instead of a single enzyme. Virus reproduction may include also the defense system of the host, epigenetic phenomena could be taken into account through simultaneous consideration of several generations, and for higher organisms the real challenge in reproduction is to deal with the enormous complexity of development in a form that is simple enough for modeling. Cooperation at the molecular level could also involve reproductive autocatalytic networks whereas social phenomena in reproductive groups or societies represent the currently highest step in the open ended complexity increase of biological evolution. Cooperation has been frequently modeled by game theory Maynard Smith (1982); Hofbauer and Sigmund (1998). There is no limitation to make the model more complex, the problem evidently is to include the desired phenomena but to keep the model sufficiently simple for mathematical analysis or simulation.

In the simple form in which the model was introduced here, it has been tested experimentally by *in vitro* evolution experiments [For an overview of early works on this subject see (Spiegelman 1971; Biebricher 1983); as a recent review we mention (Joyce 2007)]. The kinetic equations describing replication and mutation were introduced 1971 by Manfred Eigen in his scholarly written paper on self-organization of biological macromolecules (Eigen 1971). Eigen's mutation-selection equation describes the evolution of the distribution

of asexually reproducing genotypes in a population of constant size  $N$ . Correct replication and mutation are seen as parallel chemical reactions leading to a uniquely defined stationary population called *quasispecies* (Eigen and Schuster 1977). RNA replication catalyzed by single virus specific enzymes from RNA bacteriophages provides a bridge from chemistry to biology: The mechanism of the replication process is well understood in all molecular details (Biebricher 1983; Biebricher et al. 1984, 1985) and an appropriate replication assay serves for *in vitro* evolution studies (Mills et al. 1967; Biebricher 1983). The mutation-selection scenario was found to provide an appropriate molecular basis for understanding also virus evolution [For a recent survey see the contributed volume (Domingo and Schuster 2016)]. More complex systems, for example, bacteria and populations of cancer cells, were found to be describable by quasispecies theory as well (Bertels et al. 2017; Covacci and Rappuoli 1998; Napoletani et al. 2013; Brumer et al. 2006).

In the *strong mutation scenario*<sup>8</sup> Darwin's view of evolution has to be modified. Not a single fittest genotype is selected but a uniquely defined distribution of genotypes, which is represented by the largest eigenvector of a value matrix that represents the long time or stationary solution of Eigen's mutation-selection equation. The mean fitness of populations is not always optimized since situations can be constructed in which the fitness is decreasing in the approach towards the stationary state. A trivial but illustrative example of decreasing fitness during evolution considers a homogeneous population consisting exclusively of fittest genotypes as initial condition: Mutations introduce mutants into the population and since they have lower fitness by definition the mean fitness is doomed to decrease. Such situations, however, are rather rare and Darwinian optimization still remains a very powerful heuristic that applies to almost all scenarios. For error rate parameters exceeding a critical value  $p_{cr}$ , the largest eigenvector approaches the uniform distribution over the entire sequence space, which is the exact solution for the value  $p = \hat{p}$  leading to incorporations of correct and incorrect nucleotides with equal probabilities—for binary sequences this happens at  $\hat{p} = 1 - \hat{p} = \frac{1}{2}$ . In realistic populations, the uniform distribution is incompatible with a discrete quasispecies (17). Instead populations are observed that migrate randomly through sequence space Higgs and Derrida (1991); Huynen et al. (1996).

In the second half of the twentieth century, most of the molecular insights into reproduction and inheritance came from viruses and bacteria and a high percentage of molecular biologists thought that the basic regulation mechanisms of gene activities are understood. Eukaryotic cells, however, are not “giant bacteria”. Although the genetic code is the same, the gene expression and inheritance system of higher organisms are different from the prokaryotic one and much more complex. A true wealth of information on eukaryotic

cells has been discovered in the past fifty years, but gene expression in animals, plants, and fungi is still a subject of current cutting-edge research. Most of the recently revealed gene expression regulation mechanisms are subsumed under the notion of *epigenetics* for which an operational definition has been proposed at the Cold Spring Harbor Meeting in the year 2008 (Berger et al. 2009):

*An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence.*

The diversity of epigenetic effects on gene regulation is enormous. It ranges from specific methylation of DNA, in particular cytosine methylation in position 5 of CpG elements (Zemach et al. 2010), histone methylation and acetylation (Lawrence et al. 2016) to post-transcriptional RNA-methylation of adenine in position 6 (Barbosa Dogini et al. 2014; Yue et al. 2015) and small interfering RNAs (He and Hanon 2004). Epigenetics provides an extremely diverse, complex and flexible richness of regulatory actions on genes, which so far was not yet cast into a comprehensive theory and precisely this is one of the greatest challenges for the future of evolutionary biology.

There is neither a convincing theoretical model nor experimental evidence that Darwinian evolution leads to an obligatory increase in complexity. The combination of competitive selection and cooperation, however, may lead from one level of complexity to the next higher one by integration of competitors through cooperation (Szathmáry and Maynard Smith 1995; Schuster 1996). The evolution model presented here proposes a mechanism for this integration of competitors and identifies the abundance of resources as one driving force towards higher complexity. This simple model distinguishes four steps (Schuster 1996): (i) Initially the systems consists of independent replicators competing for a single resource, (ii) the capability of cooperative interaction allows to form an autocatalytic network, which couples the replicators and suppresses competition but makes the network vulnerable to exploitation by parasites, which consume resources without contributing a share to the common properties, (iii) the members of the autocatalytic network are separated from the environment by means of a suitable boundary that prevents the system from exploitation and allows for the formation of a new unit at a hierarchically higher level, and (iv) the individual units at the higher level diversify by variation, compete for common resources, Darwinian selection sets in and takes place now at the higher level. The previously autonomous units at the lower level lost their autonomy at least in part when they were integrated into the higher unit of selection. Although modeling major transitions as shown in "[Competition and cooperation](#)" is not difficult, suitable experimental molecular models are very hard to conceive, because second- and higher-order

autocatalytic systems consist almost always of complex reaction networks rather than single-step reactions [as examples for attempts to construct simple systems of this kind see (McCaskill 1997; Wlotzka and McCaskill 1997)]. John Maynard Smith and Eörs Szathmáry collected a true wealth of evidence for the historic occurrence of such major evolutionary transitions (Maynard Smith and Szathmáry 1995) in the evolution of life.

It is illustrative to think about transitions in terms of thresholds: Up to a certain value of the transition determining parameter the typical feature is hardly detectable and does not become evident before the parameter exceeds the transition value. Accordingly, such thresholds correspond to sharp transitions. Nevertheless, it appears useful to be less fussy and to accept the notion of threshold also for smooth transitions. On the three faces of the coordinate system (Fig. 5) we observe an error threshold between selection and random replication, a cooperation threshold between selection and symbiosis, and a mutation threshold that separates the regime of independent replication of all subspecies from mutual support through frequent mutation.

Understanding evolution implies knowledge on the relation between genotypes being DNA or RNA sequences and phenotypes giving rise to all fitness relevant parameters. The metaphor of an abstract fitness landscape has been originally introduced by Sewall Wright for the purpose of illustration (Wright 1922). In formal mathematical terms such a relation can be modeled as a mapping from a genotype or sequence space into fitness values. In molecular structural biology such a mapping is split into two parts: (i) a mapping from sequences into structures or genotypes into phenotypes, and (ii) a second mapping assigning fitness values to structures or phenotypes (Schuster 2016). Computer models of RNA evolution with sequence–secondary structure–fitness mappings have been studied in the past (Fontana and Schuster 1987; Fontana et al. 1989; Fontana and Schuster 1998a, b) and these studies provided the basis for a definition of evolutionary nearness of phenotypes in the presence of neutrality (Stadler et al. 2001). With more and more data on sequences and fitness values of mutants becoming currently available fitness landscapes may also be determined directly by experiment (Kouyos et al. 2012) and it is not risky to predict that genotype–phenotype relations will become an integral part of evolution models in the near future. Then evolution can be described in a self-contained manner where the genotype–phenotype mapping is a genuine part of the evolving system.

**Acknowledgements** Open access funding provided by University of Vienna. This contribution presents the continuation of a twenty five years lasting, wonderful cooperation with Manfred Eigen, which has been initiated in 1968 when the author was Post-Doc in the Eigen laboratory at the Max Planck-Institute for Physical Chemistry in Göttingen, Germany. The author wants to thank for the unique opportunity of

participating in the development of Manfred Eigen's visionary molecular view of life. Many years of cooperation and continuous discussions with Professors Karl Sigmund, Josef Hofbauer, Walter Fontana, Peter F. Stadler, Ivo L. Hofacker, Christoph Flamm and others are gratefully acknowledged.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Albery WJ, Knowles JR (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* 15:5631–5640
- Albery WJ, Knowles JR (1977) Efficiency and evolution of enzyme catalysis. *Angew Chem Internat Ed Engl* 16:285–293
- Baake E, Gabriel W (1999) Biological evolution through mutation, selection, and drift: an introductory review. In: Stauffer D (ed) *Annual review of computational physics VII*. World Scientific, Singapore, pp 203–264
- Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. *Genes Dev* 23:781–783
- Bertels F, Gokhale CS, Traulsen A (2017) Discovering complete quasispecies in bacterial genomes. *Genetics* 206:2149–2157
- Biebricher CK (1983) Darwinian selection of self-replicating RNA molecules. In: Hecht MK, Wallace B, Prance GT (eds) *Evolutionary biology*, vol 16. Plenum Publishing Corporation, New York, pp 1–52
- Biebricher CK, Eigen M, Gardiner WC Jr (1983) Kinetics of RNA replication. *Biochemistry* 22:2544–2559
- Biebricher CK, Eigen M, William C, Gardiner J (1984) Kinetics of RNA replication: plus-minus asymmetry and double-strand formation. *Biochemistry* 23:3186–3194
- Biebricher CK, Eigen M, William C, Gardiner J (1985) Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry* 24:6550–6560
- Brumer Y, Michor F, Shakhnovich EI (2006) Genetic instability and the quasispecies model. *J Theor Biol* 241:216–222
- Bryson V, Szybalski W (1952) Microbial selection. *Science* 116:45–46
- Covacci A, Rappuoli R (1998) *Helicobacter pylori*: molecular evolution of a bacterial quasi-species. *Curr Opt Microbiol* 1:96–102
- Crow JF, Kimura M (1970) *An introduction to population genetics theory*. Sinauer Associates, Sunderland
- Dogini BD, Pascoal VD, Avansini SH, Vieira AS, Campos TC, Lopes-Cendes I (2014) The new world of RNAs. *Genet Mol Biol* 37(1 suppl):285–293
- de Pasquale F, Tartaglia P, Tombesi P (1980) Stochastic approach to chemical instabilities. Anomalous fluctuations transient behavior. *Lettere al Nuovo Cimento* 28:141–145
- de Pasquale F, Tartaglia P, Tombesi P (1982) New expansion technique for the decay of an unstable state. *Phys Rev A* 25:466–471
- de Pasquale F, Tartaglia P, Tombesi P (1982) Transient-behavior near an instability point. Vectorial stochastic representation of the Malthus. *Nuovo Cimento B* 69:228–238
- Dean AM, Thronton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* 8:675–688
- Deuffhard P, Huisinga W, Jahnke T, Wulkow M (2008) Adaptive discrete Galerkin methods applied to the chemical master equation. *SIAM J Sci Comput* 30:2990–3011
- Dobzhansky T, Ayala FJ, Stebbins GL, Valentine JW (1977) *Evolution*. W. H. Freeman & Co., San Francisco
- Domingo E, Schuster P (eds) (2016) *Quasispecies: from theory to experimental systems, current topics in microbiology and immunology*, vol 392. Springer, Berlin
- Domingo E, Schuster P: What is a quasispecies? Historical origins and current scope. In: E. Domingo, P. Schuster (eds.) *Quasispecies: From Theory to Experimental Systems, Current Topics in Microbiology and Immunology*, vol. 392, chap. 1, pp. 1–22. Springer-Verlag, Berlin (2016)
- Doob JL (1942) Topics in the theory of Markoff chains. *Trans Am Math Soc* 52:37–64
- Doob JL (1945) Markoff chains—Denumerable case. *Trans Am Math Soc* 58:455–473
- Dykhuisen DE, Hartl DL (1983) Selection in chemostats. *Microbiol Rev* 46:150–168
- Ebeling W, Mahnke R (1979) Kinetics of molecular replication and selection. *Zagadnienia Biofizyki Współczesnej* 4:119–128
- Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523
- Eigen M, McCaskill J, Schuster P (1989) The molecular quasispecies. *Adv Chem Phys* 75:149–263
- Eigen M, Schuster P (1977) The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64:541–565
- Eigen M, Schuster P (1978) The hypercycle. A principle of natural self-organization. Part B: the abstract hypercycle. *Naturwissenschaften* 65:7–41
- Eigen M, Schuster P (1978) The hypercycle. A principle of natural self-organization. Part C: the realistic hypercycle. *Naturwissenschaften* 65:341–369
- Ewens WJ, Lessard S (2015) On the interpretation and relevance of the fundamental theorem of natural selection. *Theor Popul Biol* 104:59–67
- Feller W (1940) On the integro-differential equations of purely discontinuous Markoff processes. *Trans Am Math Soc* 48:488–515
- Fisher RA (1930) *The genetical theory of natural selection*. Oxford University Press, Oxford
- Fisher RA (1958) *The genetical theory of natural selection*, 2nd edn. Dover Publications Inc, New York
- Fontana W, Schnabl W, Schuster P (1989) Physical aspects of evolutionary optimization and adaptation. *Phys Rev A* 40:3301–3321
- Fontana W, Schuster P (1987) A computer model of evolutionary optimization. *Biophys Chem* 26:123–147
- Fontana W, Schuster P (1998a) Continuity in evolution. On the nature of transitions. *Science* 280:1451–1455
- Fontana W, Schuster P (1998b) Shaping space. The possible and the attainable in RNA genotype-phenotype mapping. *J Theor Biol* 194:491–515
- Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys* 22:403–434
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361
- Gillespie DT (2007) Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* 58:35–55
- Gillespie JH (1983) Some properties of finite populations experiencing strong selection and weak mutation. *Am Nat* 121:691–708
- Goel NS, Richter-Dyn N (1974) *Stochastic models in biology*. Academic Press, New York
- Gorini L, Beckwith JR (1966) Suppression. *Annu Rev Microbiol* 20:401–422
- Hamming RW (1950) Error detecting and error correcting codes. *Bell Syst Tech J* 29:147–160

- Hamming RW (1986) Coding and information theory, 2nd edn. Prentice-Hall, Englewood Cliffs
- Hartman PE, Roth JR (1973) Mechanisms of suppression. *Adv Genet* 17:1–105
- He L, Hanon GJ (2004) Micro RNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5:522–531
- Higgs PG, Derrida B (1991) Stochastic models for species formation in evolving populations. *J Phys A Math Gen* 24:L985–L991
- Hofbauer J, Mallet-Paret J, Smith HL (1991) Stable periodic solutions for the hypercycle system. *J Dyn Diff Equ* 3:423–436
- Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge
- Huang CI, Tu MF, Lin HH, Chen CC (2017) Variation approach to error threshold in generic fitness landscape. *Chin J Phys* 55:606–618
- Husimi Y (1989) Selection and evolution of a bacteriophages in cellstat. *Adv Biophys* 25:1–43
- Husimi Y, Nishigaki K, Kinoshita Y, Tanaka T (1982) Cellstat—a continuous culture system of a bacteriophage for the study of the mutation rate and the selection process at the DNA level. *Rev Sci Instr* 53:517–522
- Huynen MA, Stadler PF, Fontana W (1996) Smoothness within ruggedness. The role of neutrality in adaptation. *Proc Natl Acad Sci USA* 93:397–401
- Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol* 54:1–26
- Jones BL, Enns RH, Rangnekar SS (1976) On the theory of selection of coupled macromolecular systems. *Bull Math Biol* 38:15–28
- Jones BL, Leung HK (1981) Stochastic analysis of a non-linear model for selection of biological macromolecules. *Bull Math Biol* 43:665–680
- Joyce GF (2007) Forty years of in vitro evolution. *Angew Chem Internat Ed* 46:6420–6436
- Joyce P, Rokyta DR, Beisel CL, Orr HA (2008) A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. *Genetics* 180:1627–1643
- Kimura M (1955) Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci USA* 41:144–150
- Kimura M (1955) Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb Symp Quant Biol* 20:33–53
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- Kolmogorov AN (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math Ann* 104:415–458 In German
- Koltermann A, Kettling U (1997) Principles and methods of evolutionary biotechnology. *Biophys Chem* 66:159–177
- Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ, Bonhoeffer S (2012) Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genet* 8:e1002551
- Lawrence M, Daujat S, Schneider R (2016) Lateral thinking: how histone modifications regulate gene expression. *Trends Genet* 32:42–56
- Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge
- Maynard Smith J, Szathmáry E (1995) The major transitions in evolution. W. H. Freeman-Spektrum, Oxford
- McCaskill JS (1997) Spatially resolved in vitro molecular ecology. *Biophys Chem* 66:145–158
- Mills DR, Peterson RL, Spiegelman S (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci USA* 58:217–224
- Napoletani D, Signore M (2013) Cancer quasispecies and stem-like adaptive aneuploidy. *F1000Research* 2:e268
- Nicolis G, Prigogine I (1977) Self-organization in nonequilibrium systems. Wiley, New York
- Novick A, Szillard L (1950) Description of the chemostat. *Science* 112:715–716
- Novick A, Szillard L (1950) Experiments with the chemostat on spontaneous mutations of bacteria. *Proc Natl Acad Sci USA* 36:708–719
- Noyes RM, Field RJ, Körös E (1972) Oscillations in chemical systems. I. Detailed mechanism in a system showing temporal oscillations. *J Am Chem Soc* 94:1394–1395
- Okasha S (2008) Fisher’s fundamental theorem of natural selection—a philosophical analysis. *Br J Philos Sci* 59:319–351
- Park JM, Muñoz E, Deem MW (2010) Quasispecies theory for finite populations. *Phys Rev E* 81:e011902
- Phillipson PE, Schuster P (2009) Modeling by Nonlinear Differential Equations. Dissipative and Conservative Processes, World Scientific Series on Nonlinear Science A, 69th edn. World Scientific, Singapore
- Phillipson PE, Schuster P, Kemler F (1984) Dynamical machinery of a biochemical clock. *Bull Math Biol* 46:339–355
- Plutynski A (2006) What was Fisher’ fundamental theorem of natural selection and what was it for? *Stud Hist Phil Biol Biomed Sci* 37:50–82
- Prelich G (1999) Suppression mechanisms. themes from variations. *Trends Genet* 15:261–266
- Price GR (1972) Fisher’s ‘fundamental theorem’ made clear. *Ann Hum Genet* 36:129–140
- Sagués F, Epstein IR (2003) Nonlinear chemical dynamics. *J Chem Soc Dalton Trans* 2003:1201–1217
- Schmidt LD (2005) The engineering of chemical reactions, 2nd edn. Oxford University Press, New York
- Schnabl W, Stadler PF, Forst C, Schuster P (1991) Full characterization of a strang attractor. Chaotic dynamics on low-dimensional replicator systems. *Phys D* 48:65–90
- Schuster P (1996) How does complexity arise in evolution. Nature’s recipe for mastering scarcity, abundance, and unpredictability. *Complexity* 2/1(1):22–30
- Schuster P (2016) Increase in complexity and information through molecular evolution. *Entropy* 18:e397
- Schuster P (2016) Major transitions in evolution and in technology. What they have in common and where they differ. *Complexity* 21/4(4):7–13
- Schuster P (2016) Quasispecies on fitness landscapes. In: E. Domingo, P. Schuster (eds) Quasispecies: from theory to experimental systems, *Current Topics in Microbiology and Immunology*, vol. 392, chap. 4. Springer-Verlag, Berlin, pp 61–120.
- Schuster P (2016) Some mechanistic requirements for major transitions. *Phil Trans Roy Soc Lond B* 371:e20150439
- Schuster P (2016) Stochasticity in processes. Fundamentals and applications in chemistry and biology. Springer series in synergetics. Springer, Berlin
- Schuster P (2017) A mathematical model of evolution. *MATCH Communications in Mathematical and in Computer Chemistry* 78, submitted
- Schuster P (2017) Molecular evolution between chemistry and biology. The interplay of competition, cooperation, and mutation. *European Biophysics Journal* 46, submitted
- Schuster P, Sigmund K (1985) Dynamics of evolutionary optimization. *Ber Bunsenges Phys Chem* 89:668–682
- Schuster P, Sigmund K, Wolff R (1978) Dynamical systems under constant organization I. Topological analysis of a family of nonlinear differential equations - A model for catalytic hypercycles. *Bull Math Biol* 40:734–769

- Schuster P, Sigmund K, Wolff R (1979) Dynamical systems under constant organization III. Cooperative and competitive behavior of hypercycles. *J Diff Equ* 32:357–368
- Schuster P, Sigmund K, Wolff R (1980) Dynamical systems under constant organization II. Homogeneous growth functions of degree  $p = 2$ . *SIAM J Appl Math* 38:282–304
- Sniegowski PD, Gerrish PJ (2010) Beneficial mutations and the dynamics of adaptation in asexual populations. *Phil Trans R Soc B* 356:1255–1263
- Spiegelman S (1971) An approach to the experimental analysis of pre-cellular evolution. *Quart Rev Biophys* 4:213–253
- Stadler BRM, Stadler PF, Wagner GP, Fontana W (2001) The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J Theor Biol* 213:241–274
- Strunk G, Ederhof T (1997) Machines for automated evolution experiments in vitro based on the serial-transfer concept. *Biophys Chem* 66:193–202
- Swetina J, Schuster P (1982) Self-replication with errors - A model for polynucleotide replication. *Biophys Chem* 16:329–345
- Szathmáry E, Maynard Smith J (1995) The major evolutionary transitions. *Nature* 374:227–232
- Tarazona P (1992) Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin glasses. *Phys Rev A* 45:6038–6050
- Thompson CJ, McBride JL (1974) On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math Biosci* 21:127–142
- Wiehe T (1997) Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet Res Camb* 69:127–136
- Wlotzka B, McCaskill JS (1997) A molecular predator and its prey: coupled isothermal amplification of nucleic acids. *Chem Biol* 4:25–33
- Wright S (1922) Coefficients of inbreeding and relationship. *Am Nat* 56:330–338
- Yue Y, Liu J, He C (2015) RNA N<sup>6</sup>-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev* 29:1343–1355
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919