Boltzmann and Evolution: Some Basic Questions of Biology seen with Atomistic Glasses

Peter Schuster* [a,b]

[a] *Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria,*

[b] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Abstract: Ludwig Boltzmann's visionary view of life being understandable on the molecular level as the interplay of complex molecules performing highly sophisticated chemical syntheses is discussed in the light of current molecular biology. Boltzmann's high esteem for Charles Darwin's theory of evolution is complemented by an account of the current view of selection as a general phenomenon not restricted to cellular life. In two more technical sections we present a brief overview of a molecular theory of evolution that provides a conceptual frame for understanding evolution and a tool for quantitative description of evolutionary phenomena.

Key words: Adaptation – Darwinian mechanism – neutral evolution – quasispecies – selection.

1 Boltzmann and biology

In his popular lecture on the second law of thermodynamics Ludwig Boltzmann made two highly remarkable statements about evolution and biology.[1] The first is dealing with the role of Charles Darwin and his theory of evolution [Boltzmann, 1979, p. 29]:

> ... Wenn Sie mich nach meiner innersten Überzeugung fragen ob man unser (das 19.) Jahrhundert einmal das eiserne Jahrhundert oder das Jahrhundert des Dampfes oder der Elektrizität nennen wird, so antworte ich ohne Bedenken, das Jahrhundert der mechanischen Naturauffassung, das Jahrhundert Darwins wird es heißen. ...
>
> ... If you ask me about my innermost conviction whether our century will be called the century of iron or the century of steam or electricity, I answer without hesitation: It will be called the century of the mechanical view of Nature, the century of Darwin. ...

*Corresponding author, E-mail: pks@tbi.univie.ac.at

[1] Ludwig Boltzmann, *Der zweite Hauptsatz der mechanischen Wärmetheorie.* Lecture presented at the 'Festive Session' of the Imperial Academy of Sciences in Vienna, May 29, 1886. The German text is taken from Boltzmann [1979] and the English translation from Broda [1983].

In order to interpret this sentence properly in our current terminology *mechanical* should be replaced *mechanistic*. Boltzmann went even further in his thoughts and made clear statements that indicate the beginning of en evolutionary theory of cognition. The second statement deals with energy, entropy and photosynthesis and is even more remarkable, because it anticipates the molecular view of present day biology [Boltzmann, 1979, p. 41]:

> ... Der allgemeine Daseinskampf der Lebewesen ist daher nicht ein Kampf um die Grundstoffe – die Grundstoffe aller Organismen sind in Luft, Wasser und Erdboden im Überflusse vorhanden – auch nicht um Energie, welche in Form von Wärme leider unverwandelbar in jedem Körper reichlich vorhanden ist, sondern ein Kampf um die Entropie, welche durch den Übergang der Energie von der heißen Sonne zur kalten Erde disponibel wird. Diesen Übergang möglichst auszunutzen, breiten die Pflanzen die unermeßliche Fläche ihrer Blätter aus und zwingen die Sonnenenergie in noch unerforschter Weise, ehe sie auf das Temperaturniveau der Erdoberflche herabsinkt, chemische Synthesen auszuführen, von denen man in unseren Laboratorien noch keine Ahnung hat. Die Produkte dieser chemischen Küche bilden das Kampfobjekt für die Tierwelt. ...

> ... The general struggle for existence of living beings is therefore not a fight for the elements – the elements of all organisms are available in abundance in air, water, and soil – nor for energy, which is plentiful in the form if heat, unfortunately untransformably, in every body. Rather it is a struggle for entropy that becomes available through the flow of energy from the hot Sun to the cold Earth. To make the fullest use of this energy, the plants spread out the immeasurable areas of their leaves and harness the Sun's energy by a process that is still unexplored, before it sinks down to the temperature level of the Earth, to drive chemical syntheses of which one has no inkling as yet in our laboratories. The products of this chemical kitchen are the object of the struggle ion the animal world. ...

'Entropy' in Boltzmann's lecture should presumably be replaced here by 'Negentropy' as it has been used by Erwin Schrödinger in his famous lectures on 'What is life' in Dublin [Schödinger, 1944]:

> ... What an organism feeds upon is negative entropy. Or to put it less paradoxically, the essential thing in metabolism is that the organism succeeds in freeing itself from all the entropy it cannot help producing while alive. ...

The recognition of negentropy, or more correctly free energy, as the object at stake in the struggle for existence and in the evolution of life was not a privilege of physicists. The famous biologist and mathematician D'Arcy Thompson gives a very clear account on the energetic basis of life in his book 'On Growth and Form' [Thompson, 1942] that was published first in 1917. Although Boltzmann's statement about the – for life on Earth for practical reasons – unlimited availability of mineral components is true under most circumstances, limitations of growth because of shortage in phosphorous have been reported and are well documented for more than seventy years [Riddell et al., 1934; Eaton, 1952; Wardle et al., 2004]. The essential components of biomass production as we know them nowadays are indeed sunlight and water. It is interesting that liquid water is addressed as a *conditio sine qua non* rather rarely by the scholars of physics.

Charles Darwin's theory of evolution [Darwin, 1859] can be casted in five statements [Kutschera, 2006, p. 34]:

(i) Evolution is a real historic process. Species are subject to change and have evolved over millions of years from precursor species.

(ii) The driving force for the evolution of species is the interplay of variation and selection in populations.

(iii) All living beings descend from a common ancestor that represents the root of the *tree of life*.

(iv) Phylogeny proceeds gradually and not stepwise.

(v) Precursor species are split into daughter species during phylogeny and lead thereby to branches in the tree of life.

Evidence for all five concepts has been found in Nature. There is an exception of statement (iv), because punctuation in the appearance of species seems to occur as well. Statements number (i), (iii), and (v) form the conventional view of biological evolution giving rise to the tree of life. A slight modification is necessary in the light of horizontal gene transfer.[2] Statement (ii) provides the mechanism of optimization in the Darwinian scenario: Multiplication of organisms leads to more progeny than the numbers of individuals that can be sustained by the ecosystem. As an highly relevant side effect it produces a distribution of variants by unperfect coping and recombination.[3] Variation and selection of the fittest in the sense of maximum number of progeny leads, inevitably, to an optimization of the number of fertile offspring. In section 4 we shall present and discuss mathematical models, in section 5 computer simulation of optimization through variation and selection.

Boltzmann's pleading for evolution as the key to understand biology and the world was precisely to the point. It was formulated almost ninety years later by Theodosius Dobzhansky, one of the scholars of the synthetic theory of evolution. He said the famous sentence [Dobzhansky et al., 1977]:[4]

'Nothing in biology makes sense except in the light of evolution'.

Another interesting detail concerns atomism or discreteness and biological evolution. Charles Darwin and the selectionists of the first half of the twentieth century thought that evolution proceeds in small steps (statement iv). They were influenced by the dominant view of geologists in the nineteenth century who believed in gradualism [Ruse, 1979; Mayr, 1982; Kutschera, 2006]. Large numbers of small changes shape the evolutionary process. Gregor Mendel's discoveries of the laws of inheritance introduced discreteness into evolutionary theory. As a matter of fact Mendel had introduced 'atoms

[2]Horizontal gene transfer is the exchange of genetic material between organisms living at the same time.

[3]Neither mutations nor recombination were known as sources of variation in Darwin's days. He himself believed in the hereditary acquisition of acquired capabilities as the source of variation. Thus, we would call him today a Lamarckian.

[4]The famous biologist and deep thinker of evolution Ernst Mayr gives an extensive account of the growth of biological thinking [Mayr, 1982].

of inheritance' – nowadays called genes – into biology. Would it not have been natural for Ludwig Boltzmann to join the party of the geneticists rather than the selectionists? Although Mendel's work stayed almost unnoticed during the second half of the nineteenth century, it was rediscovered around the turn of the century and Boltzmann could well have been aware of it.[5]

An even more important concept that influenced further development of evolutionary biology was August Weismann's hypothesis on the strict separation of the potentially immortal cells of the germline, which can be transferred to children, and the somatic cells, which die at the latest together with their carrier organism [Weismann, 1892; Wallace, 1889]. Together with Darwin's natural selection and Mendel's rules of inheritance the germline hypothesis – now fully confirmed by cellular and molecular biology – builds the basis of the so-called Neodarwinian theory of evolution. An interesting historical detail concerns Charles Darwin's thoughts about inheritance: All his life and even more outspoken in the later issues of the 'Origin of Species' Darwin believed in the transmission of acquired properties to the progeny and considered it as the factor providing inheritable variation. In this sense Darwin was a Lamarckian. In the Neodarwinian theory inheritable modifications are only possible through changes of germ cells. Biology was split into two camps during the first decades of the twentieth century since selectionists and geneticists, as the distinguished biologist Wolfgang Freiherr von Buddenbrock-Hettersdorf said in 1930, were heavily fighting:

> ... The controversy ... is as undecided now as it was 70 years ago ... neither party had been able to refute the arguments of their opponents and one must assume that the situation is not going to change soon. ...

Fortunately, Buddenbrock's prediction did not become true. Already in the nineteen thirties a formal mathematical unification of Darwinian selection and Mendelian genetics was performed by the scholars of population genetics, Ronald Fisher, J.B.S. Haldane and Sewall Wright. Later, before and during World War II, the ultimate unification occurred in the synthetic theory of evolution that has been further extended by molecular life sciences [Mayr and Provine, 1980; Reif et al., 2000; Kutschera and Niklas, 2004].

2 Holism and reductionism

The holism versus reductionism debate is an old theme in philosophy and science. Sometimes holism is even traced back to Aristotle's Metaphysics that contains the famous sentence: "The whole is more than the sum of its parts". The problem here is the usage of the term *sum*. When sum implies a simple arithmetic sum, the statement expresses nothing more and nothing less that the parts of the whole do interact. For every property $F(\Xi)$ of an ensemble $\Xi = \{X_1, X_2, \ldots, X_n\}$ we can write

$$F(\Xi) = \sum_{i\in\Xi} f^{(1)}(X_i) + \sum_{i\in\Xi}\sum_{j<i,j\in\Xi} f^{(2)}(X_iX_j) + \ldots + f^{(n)}(X_1\cdots X_n) \quad (1)$$

[5]Several biologists repeated and confirmed Gregor Mendel's breeding experiments. One of them, William Bateson [1902] coined the term *genetics* and published a monograph that has become a classic on the origin of genetics.

The terms in the cluster expansion $F(\cdot)$ corresponds to one body, two body and higher interaction contributions up to the n body term, for example

$$f^{(1)}(X_i) = F(X_i)\,,\; f^{(2)}(X_iX_j) = F(X_iX_j) - \Big(F(X_i) + F(X_j)\Big)\;\ldots\;.$$

In the naïve interpretation Aristotle's sentence then says that there non-vanishing two or more body terms and therefore additivity is violated.

Apart from trivial non-additivity Aristotle's sentence has three deeper interpretations that can be casted into questions:

(i) Are new properties emerging in the progression from a lower to a higher hierarchical level?

(ii) Can we describe the phenomena on a higher hierarchical level by means of the laws operating on the lower level or do we need new laws of Nature that become operational in the form of specific forces only on the higher level?

(iii) Are there limits in the predicability of complex systems that cannot be compensated by improved knowledge on the parts of the system?

The answer to question (i) will be almost always *yes*. We consider, for example, the world of atoms and the hierarchically higher world of molecules being composed of atoms. The chemical bond is part of the notions needed to describe the properties of molecules but does not exist in the world of atoms: The chemical bond is an emergent property.

Question (ii) asks whether the postulation of special laws and forces like the notorious vital force - *vis vitalis* - behind living organisms is indispensable. Additional forces causing essentially novel regularities on the higher level, which need new fundamental laws to describe them, are much harder to argue and to verify. In the life sciences this second version of holism has become very unpopular and the majority of scientists would currently agree that it is extremely unlikely to discover new fundamental forces in biology, psychology or sociology. In other words, there is a common belief that neither biology nor psychology nor sociology will lead to observations that contradict contemporary physics.[6]

Question (iii) addresses so-called scientific holism and finds its confirmation in the existence of principle reasons like quantum physical uncertainty or deterministic chaos and technical limitations, for example incomplete information on initial and boundary conditions.

Historically, the idea of reductionism has been introduced in the seventeenth century by René Descartes. He thought that the world was like a machine whose operations can be fully understood in terms of the mechanical parts like the operation of a mechanical clock can be explained if all its

[6] A related but more radically formulated view comparing bottom-up explanations with *deus ex machina* solutions is found under the heading 'Skyhooks or cranes?' in Daniel Dennett [1995, p. 73].

pieces and their relative positions are known. The most extreme form of reductionism is known as ontological reductionism and states that the ultimate explanation of everything has to be given in terms of most fundamental entities being elementary particles or strings. Daniel Dennett [1995] calls this concept also "greedy reductionism". A milder form of reductionism is called hierarchical reductionism [Dawkins, 1986, p. 13]. It is related to the idea of a unity of science and states that complex systems can be described by a hierarchy of levels in which each form of organization is described in terms of the objects of the next lower level. Within science hierarchical reductionism is expressed, for example, by the statements: Fundamental chemistry is based on physics, fundamental biology is based on chemistry, psychology is based on biology, sociology is based on psychology and, eventually, political science, anthropology and economics are based on sociology. A majority of scientists is accepting the first two reductions, chemistry $\Rightarrow$ physics and biology $\Rightarrow$ chemistry, but at present the other reductions are strongly opposed by many researchers. Examples are the controversial discussions of the interpretations of observations from sociobiology or evolutionary psychology [Caplan, 1978]. Methodological reductionism or the reductionists' program, on the other hand, is the method of handling problems in science if one aims at going beyond pure narrative descriptions. Physics is the discipline that has most experience with reduction but everywhere in science experimental exploration of regularities requires reduction in the sense of simplification and constant environments in particular in many variable systems. Even for understanding how and why the whole is more than the sum of its parts profound knowledge of the parts is indispensable. It seems useful to end the academic holism versus reductionism debate by referring to the famous biologist John Maynard Smith [1986] who had a rather pragmatic view on the subject. He compares macroscopic biologists as pursuing a holistic strategy by means of a top-down approach to describe the phenomena observed in biology with the reductionists' program of molecular biologists who perform a bottom-up approach to interpret biological phenomena by the methods of chemistry and physics. He rejects holistic arguments boiling down to the claim that, because we do not understand some phenomena at present, there must be some special vital force, which is responsible for them. He says [Maynard Smith, 1986, p. vii]:

> ... As it happens, I do not understand how modern sewing-machines work, but this does not lead me to suppose that the laws of topology have been broken: Indeed, I feel confident I could find out if someone would let me take one into pieces. ...

Although the holistic view is preferable to John Maynard Smith, he makes a visionary statement about molecular biology [Maynard Smith, 1986, p. vii]:

> ... Holists are, I think, in a weaker position, if only because in recent years progress has been so much faster from the bottom up than from the top down. Yet I do share their conviction that there are laws that can only be discovered by research on whole organisms, and on populations of organisms. Almost all my own work has been done at those levels. What should be the attitude of a biologist working on whole organisms to molecular biology? It is, I think, foolish to argue that we are discovering

> things that disprove molecular biology. It would be more sensible to say to molecular biologists that there are phenomena that they will one day have to interpret in their terms. ...

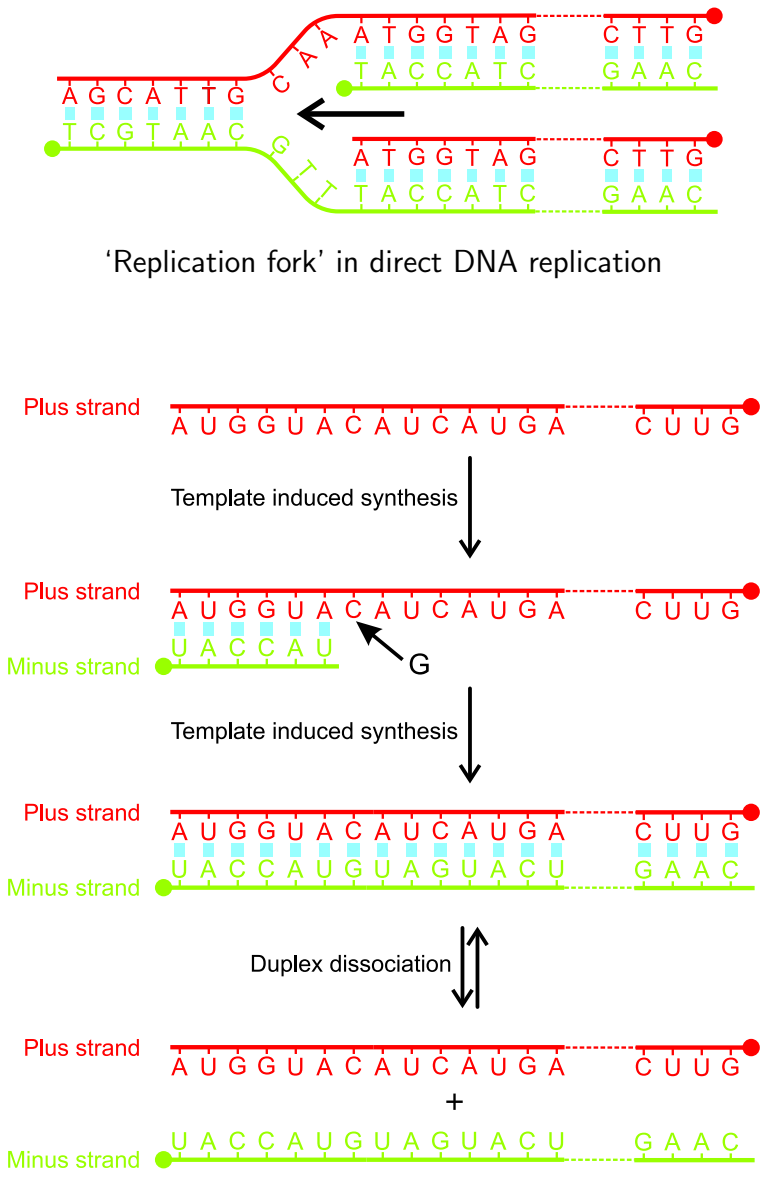
The expectation of Maynard Smith has become almost true nowadays in systems biology. What is said here about chemistry and biology, I believe,is likewise true for the relation between neurobiology and psychology or population biology and sociology.

Recalling Ludwig Boltzmann's statements on evolution and biology [Boltzmann, 1979, p. 29] we would assign him to the community of hierarchical reductionists. In other words he was convinced that one day the most complex biological phenomena would find there ultimate explanation in physics and chemistry of living matter. His high esteem of evolution makes clear his appreciation for phenomena that are unique in the realm of living beings, because self-organization and selection in the inanimate world were no theme in the nineteenth century. With our current knowledge the view is more subtle since we have plenty of examples for pattern formation, mode selection, and other spontaneous ordering processes in pure physics as well as fully developed theories to deal with them [Nicolis and Prigogine, 1977; Haken, 1977, 1983]. What distinguishes life sciences from the sciences of inanimate matter is biological or genetic information and how it originates through evolution [Eigen, 1971; Eigen and Schuster, 1977; Eigen, 1993].

3 Molecular biology and evolution

The beginning of biochemistry and the start of the unification of chemistry of minerals and biology is commonly dated 1828 when Friedrich Woehler succeeded to synthesize urea by heating the salt ammonium cyanate. For more than a century biochemists isolated, purified, and studied molecules obtained from living beings. The chemical compositions of biomolecules has been established. Biochemical reactions, in particular catalysis by means of enzymes, were investigated as rigorously as other chemical reactions and biochemical kinetics became an important field in its own right. A real breakthrough in understanding the molecules of life was the introduction of the concept of macromolecules by the German organic chemist Hermann Staudinger in the nineteen twenties. He was awarded the nobel prize in chemistry 1953. Staudinger characterized macromolecules as polymers consisting of a large number of small molecular units that are linked together by chemical bonds. He correctly recognized biological macromolecules, in particular proteins, nucleic acids and carbohydrates as heteropolymers built from several classes of monomeric units. The enormous variability of biopolymers is a result of combinatorics: A polymer of length n built from κ classes on monomers can exist in κ^n different sequences, each of which having the possibility to give rise to different molecular properties.

The second breakthrough in understanding the molecular mechanisms of life occurred after World War Two when the methods of structure determination through application of crystallography were extended to biomolecules. The two most important landmarks were: (i) The three-dimensional structural model for the DNA double helix by James Watson and Francis Crick



Complementary RNA replication

Figure 1: Basic mechanisms of nucleic acid replication. DNA replication (top) is a complex process involving a machinery of some twenty protein molecules. It is semi-conservative in the sense that every daughter molecule carries one strand of the parent DNA. RNA replication (bottom) commonly follows a complementary mechanism: A double-helical (plus-minus) duplex is synthesized from a (single) plus strand by making use of the complementarity of Watson-Crick base pairs. The critical step in replication is the separation of the duplex into two single strands because long double helical stretches are bound strongly. In RNA evolution experiments separation into single strand is performed by the replicase that prevents the formation of long double helical stretches through separating them into the two singe strands that form their own structures.

and (ii) the determination of the spatial structures of the proteins myoglobin and hemoglobin at atomic resolution by John Kendrew and Max Perutz, respectively. All four researchers were awarded nobel prizes in 1962, Watson

and Crick in medicine and Kendrew and Perutz in chemistry. A close look at the three-dimensional structures of biomolecules, in particular at the DNA double helix, gave immediate hints on their biological function. Watson and Crick state at the end of their seminal letter to Nature [Watson and Crick, 1953, see figure 1]:

> ... It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. ...

These structure determinations are seen as the beginning of molecular biology[7] and initiated the remarkable development of structural biology. Larger and larger structures were determined at atomic resolution and revealed Nature's tricks to perform the most sophisticated chemical reactions in highly specific and efficient ways. In very short time after the determination of DNA structure the whole genetic machinery including the genetic code relating DNA, RNA, and protein was discovered and early nineteen sixties molecular biology was fully established.

A third landmark in the development of molecular life sciences was the invention of new techniques for DNA sequencing by Walter Gilbert in USA and Frederick Sanger in England. Both received the nobel prize in chemistry 1980. Sequencing of whole genomes[8] of organisms became possible and molecular genetics got a new basis. The goal was now to investigate whole cells and whole multicellular organisms rather than individual biomolecules. Genomics, proteomics, metabolomics, and systems biology study all genes, all proteins, all metabolites, and all reactions of a cell or an organism together. The various *omics* aim at investigations of the chemistry of entire cells or organisms. The reductionistic bottom-up approach is currently reaching the situation John Maynard Smith was addressing in the quotation. Chemistry and physics have conquered biology but the physicists and chemists entering are becoming biologists, because they are now asking biological questions and they are analyzing the problems of biology with the new techniques they brought form their original disciplines. Notions and concepts genuine to biology find explanations in terms of chemistry and physics. Examples are Gregor Mendel's laws of inheritance and systematic deviations from the simple ratios, regulation of gene activities including pleiotropy and epistasis, epigenetic phenomena, and many others. Genetic information was found to be one of the the most relevant properties distinguishing living and inorganic matter, others are homeostasis, resilience and many more.

The source of variation in populations has always been some kind of mystery. Charles Darwin, as we have said, believed in a kind of Lamarckian mechanism. Molecular genetics was able to explain all kinds of variations as types of chemical reactions giving rise to deviations from correct copying of DNA or RNA. In figure 2 we summarize the most frequently occurring classes of mutations and recombination. Error prone replication and recombination

[7]The fascinating story of the sequence of discoveries in molecular biology is told in Judson [1979].

[8]The genome or the genotype of an organism is the complete genetic information that is stored in its DNA.

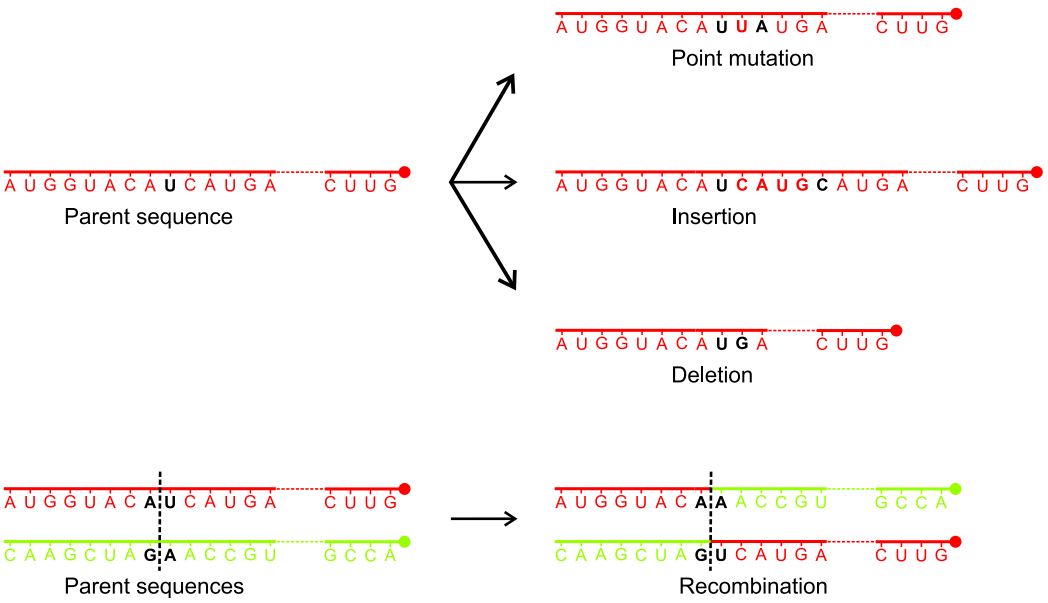


Figure 2: Basic mechanisms of sequence variation. The upper part of the figure sketches three classes of mutations: (i) A point mutations where a single digit is changed and the sequence length remains constant, (ii) an insertion where part of the sequence is replicated twice, and (iii) a deletion where part of the sequence is not replicated. In the lower part we show a case of symmetric recombination between two sequences of equal lengths leading to two new sequences both with the same number of nucleotides – for the sake of simplicity recombination is shown here for two single stranded molecules; in case of two double stranded molecules the molecular mechanism is more complex. Asymmetric recombination (not shown) leads to sequences of different chain lengths. In recombination the genetic information of two parent strands is reassembled in the two daughter molecules.

are the sources of diversity in populations. When selected, mutations introduce genetic novelty in Nature.

Evolution in the Darwinian sense is based on multiplication, variation, and selection. All three prerequisites can be fulfilled by molecules outside a cellular environment. Therefore there is no reason why the observation of evolution should be restricted to organisms. Indeed, Sol Spiegelman was able to show already in the nineteen sixties that RNA molecules evolve in test-tubes provided an environment that supports replication is provided and consumed materials are replenished [Mills et al., 1967; Spiegelman, 1971]. He took a series of test-tubes filled with stock solution, which contains all substances that are required for Qβ viral RNA replication including an enzyme, Qβ replicase. A sample of viral RNA is injected into the first test-tube, replication starts, material is consumed, and after some while a small sample of this test-tube is transferred into test-tube no.2. The procedure is repeated over and over again and after a sufficiently large number of serial transfers an RNA molecule is isolated that replicates much faster than the original one. Selection for the fastest replicating molecule has taken place and, as sequence analysis shows, the faster replicating molecules are much smaller than the original viral RNA. The explanation is straightforward: The genes

necessary for survival under natural conditions are not needed in the artificial laboratory environment and, therefore, they are eliminated through deletions yielding smaller and faster replicating molecules. Later on these experiments were repeated, carefully analyzed, and substantially extended [Biebricher, 1983; Biebricher and Eigen, 1988; Bauer et al., 1989; Biebricher and Gardiner, 1997; Strunk and Ederhof, 1997; Öhlenschlager, 1997].

4 Modelling evolution of molecules

The success of evolutionary optimization is based on the dichotomy of genotype and phenotype. The genotype of an individual is its DNA or RNA sequence, the phenotype is the totality of its properties and consequently all variation, mutation or recombination, involves the genotype, whereas selection being based on fitness values or numbers of (fertile) offspring is a property of the phenotype. Commonly, the phenotype is a highly complex object but in case of serial transfer or flow reactor optimization of RNA, the phenotype is the three-dimensional structure of the RNA molecules together with its properties in the laboratory experiments. In other words, variation operates on the genotype and selection weights fitness relative to mean fitness in the population. Here we review an attempt to see evolution with with the glasses of a physicist [Reidys et al., 1997; Fontana and Schuster, 1998a; Schuster, 2006].

Definitions of genotype and phenotype spaces allow for a formalization of evolutionary processes. The genotype or sequence space $\mathcal{Q}_n$ is a discrete space comprising all sequences of chain length n with the Hamming distance $d_H(\mathbf{X}_i, \mathbf{X}_j)$ representing a metric.[9] The points in phenotype or shape space $\mathcal{S}_n$ correspond to individual phenotypes and $d_S(S_i, S_j)$ is some metric in phenotype space. Fitness values, f_k, are the result of two consecutive mappings from sequence space into shape space and from shape space into non-negative real numbers:

$$\begin{aligned} \psi : \{\mathcal{Q}_n; d_H(X_i, X_j)\} &\stackrel{\text{fold}}{\Longrightarrow} \{\mathcal{S}_n; d_S(S_i, S_j)\} \quad \text{or} \quad S_k = \psi(X_k) \ , \\ \phi : \{\mathcal{S}_n; d_S(S_i, S_j)\} &\stackrel{\text{eval}}{\Longrightarrow} \mathbb{R}^1_+ \quad \text{or} \quad f_k = \phi(S_k) \ . \end{aligned} \tag{2}$$

Evolution takes place in genotype *and* phenotype space: A population migrates in sequence space as a consequence of mutation and recombination, the population in genotype space is mapped by means of ψ into an ensemble of phenotypes, which in turn have their individual fitness values described by ϕ. Based on this mapping ϕ selection operating on the population through reducing the numbers of genotypes of less fit phenotypes and eventually eliminating them. Hence evolution can be understood as a process in two spaces where the dynamics is coupled through the two mappings ψ and ϕ.

[9] The Hamming distance between to sequences, $d_H(\mathbf{X}_i, \mathbf{X}_j)$ counts the number of the positions in which the two sequences $\mathbf{X}_i$ and $\mathbf{X}_j$ differ. Restriction of sequence spaces to constant chain length n has the consequence that only point mutations and symmetric recombinations are considered. Extensions to shape spaces with variable chain lengths are not simple but possible.

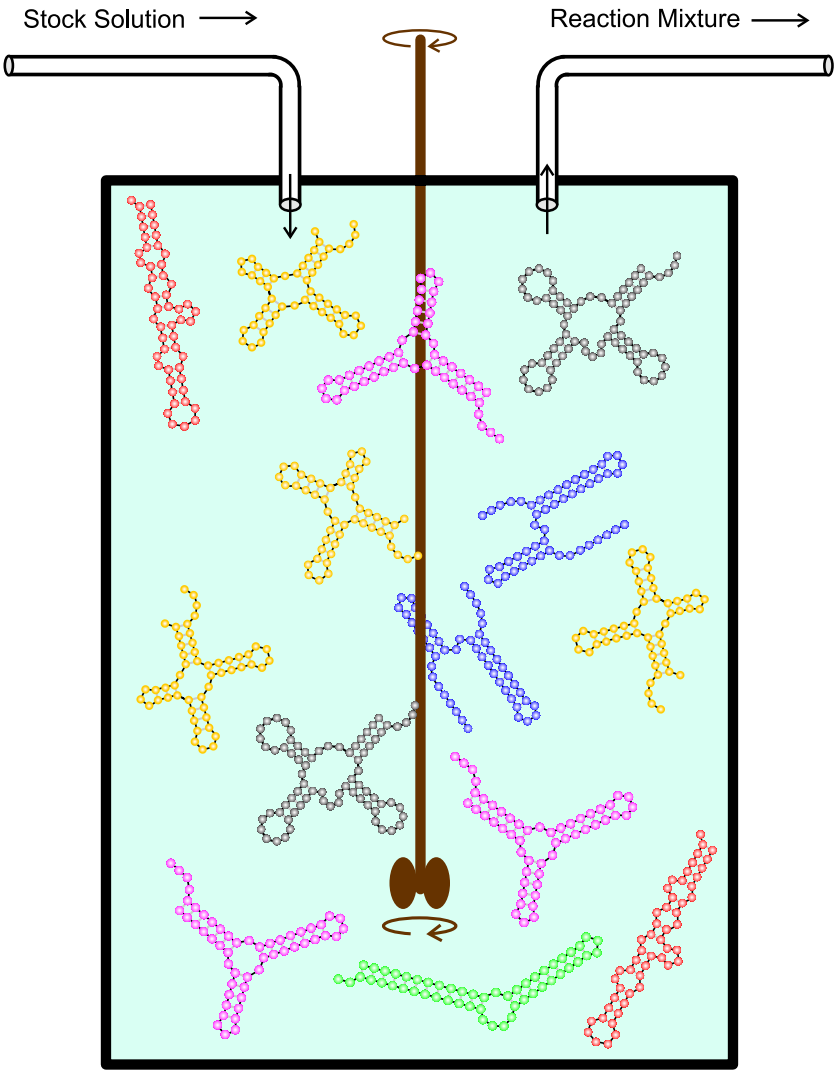


Figure 3: The flow reactor as a device for studying *in vitro* evolution of molecules. The reactor maintains off-equilibrium conditions by means of a influx of stock solution that supplies material (**A**) for replication. Molecules produced in excess are removes from the reaction mixture through an unspecific outflow. A population of N RNA molecules is subjected to replication and mutation. RNA structures are computed and evaluated by means of the function $f_k = \phi(S_k)$ for all mutant sequences. The clover-leaf shaped yeast tRNA^{phe} (grey shape in the reactor) was chosen as target structure. Inputs of an evolution experiment *in silico* are the parameters (i) population size N, (ii) chain length n of the RNA molecules, (iii) the mutation rate p, and (iv) the initial population.

Studying evolution of molecules in the laboratory reduces the enormous complexity of *in vivo* conditions and allows for straightforward modelling by means of chemical kinetics. Mathematical analysis is possible for ODE based models. Stochasticity can be investigated by means of computer simulations. Here we shall sketch the theory of molecular evolution only briefly and supplement the rigorous results by computer simulations providing hints on statistics in evolution (section 5).

The currently simplest model systems for studying evolution *in vitro* are Spiegelman's experiment and various other selection assays [Klussmann, 2006]. For the mathematical model we choose a flow reactor (figure 3) in which serial transfers are replaced by a continuous influx of stock solution

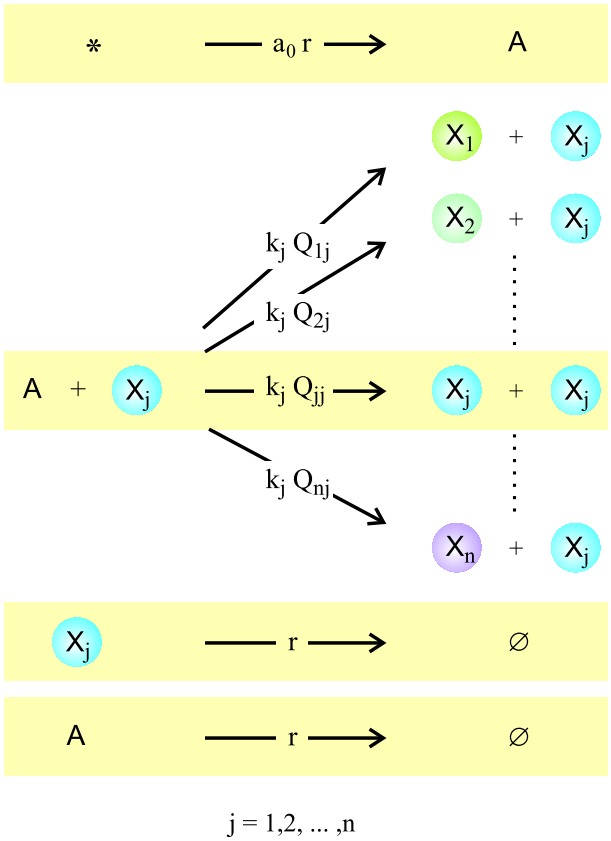


Figure 4: Replication and mutation in the flow reactor. Replication and mutation are considered as parallel chemical reactions. Stock solution containing **A** at concentration a_0 and all other material required for replication flows into the reactor with flow rate r. The volume compensating outflux reduces all concentrations with the same flow rate. The reaction rate parameter for a replication of template $\mathbf{X}_j$ is denoted by k_j. The dimensionless factors Q_{ij} represent the probabilities for the synthesis of molecule $\mathbf{X}_i$ as an error copy of template $\mathbf{X}_j$. Accordingly, Q_{jj} is the frequency of correct copying of $\mathbf{X}_j$. Since every copy has to be either correct or error prone, we have $\sum_{i=1}^{n} Q_{ij} = 1$ and Q is a (column) stochastic matrix. The pure replication case is underlaid in yellow.

containing **A** at a concentration a_0. The influx is compensated in volume by an outflux of the reaction mixture:

$$* \xrightarrow{a_0 \cdot r} \mathbf{A} \quad : \quad \text{influx} \tag{3}$$

$$\mathbf{A} \xrightarrow{r} \emptyset \quad : \quad \text{outflux} \tag{4}$$

$$\mathbf{X}_j \xrightarrow{r} \emptyset \; ; \quad j = 1, \ldots, n \quad : \quad \text{outflux} \tag{5}$$

The symbol **A** stands for the material, which is consumed in the synthesis of the RNA molecules, $\mathbf{X}_j$ $(j = 1, \ldots, n)$. Dilution effects all molecular species in the same way. The parameter r represents the flow rate or, in other words, $\tau = r^{-1}$ is the mean residence time of a volume element in the flow reactor. It is straightforward to show that complementary replication, $\mathbf{A} + \mathbf{X}_+ \rightarrow \mathbf{X}_+ + \mathbf{X}_-$ and $\mathbf{A} + \mathbf{X}_- \rightarrow \mathbf{X}_- + \mathbf{X}_+$ (figure 1) is characterized by two phases: (i) an initial phase that leads to a stationary value of the relative amounts of both strands and (ii) a quasi-stationary phase, during which the two strands forming the plus-minus ensemble grow together with a rate parameter that

is given by the geometric mean: $k = k_{\pm} = \sqrt{k_+ \cdot k_-}$ [Eigen, 1971], where k_+ and k_- are the rate parameters for the individual strands. Replication and mutation are considered as parallel chemical reactions (figure 4):

$$(\mathbf{A}) \; + \; \mathbf{X}_j \; \xrightarrow{k_j \cdot Q_{jj}} \; 2\,\mathbf{X}_j \,; \; j = 1, \ldots, n \qquad : \text{ replication and} \tag{6}$$

$$(\mathbf{A}) \; + \; \mathbf{X}_j \; \xrightarrow{k_j \cdot Q_{ij}} \; \mathbf{X}_j \; + \; \mathbf{X}_i \,; \; i, j = 1, \ldots, n \,; \; j \neq i \quad : \text{ mutation} \,. \tag{7}$$

Although full reaction kinetics of RNA replication follows a complicated multistep mechanism [Biebricher and Eigen, 1988], it is sufficient for our purposes to consider the overall processes (6) and (7). The material required for polymer synthesis is put in parentheses in order to indicate that commonly a mixture of compounds is needed and we dispense here from considering explicit stoichiometry. Then, the replication-mutation system is determined by a vector of replication rate parameters, $\mathbf{k} = (k_1, k_2, \ldots, k_n)$ and a matrix of mutation frequencies,

$$\mathrm{Q} \; = \; \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nn} \end{pmatrix} \; .$$

All components of $\mathbf{k}$ and all entries of Q are real and nonnegative.[10] The matrix Q is a (column) stochastic matrix since every replication has to be either correct or not: $\sum_{i=1}^{n} Q_{ij} = 1$. A uniform error rate model is useful for studying replication-mutation systems in general or for cases where more detailed information is missing. Three assumptions are made: (i) Only point mutations are considered, (ii) the frequency of mutation does not depend on the particular kind of base exchange, and (iii) the frequency of mutation does not depend on the position of the mutated nucleotide in the sequence. Then, the mutation frequency from $\mathbf{X}_j$ to $\mathbf{X}_i$ can be computed from the chain length n, the single nucleotide error rate p, and the Hamming distance of $\mathbf{X}_j$ and $\mathbf{X}_i$, $d_H(\mathbf{X}_i, \mathbf{X}_j)$:

$$Q_{ij} \; = \; (1-p)^{n - d_H(\mathbf{X}_i, \mathbf{X}_j)} \, p^{\, d_H(\mathbf{X}_i, \mathbf{X}_j)} \; . \tag{8}$$

Accordingly, the whole mutation matrix depends only on the relative position of the sequences in sequence space as expressed by the Hamming distance and on the single nucleotide error rate.

The dependence of the rate of RNA synthesis on the concentration of the required material $\mathbf{A}$, $[\mathbf{A}] = a$, is assumed to be some monotonically increasing function $F(a)$.[11] It is straightforward to write down the kinetic differential equation of replication and mutation for the mechanism of figure 4 in the

[10] This is a consequence from chemical reaction kinetics where all rate parameters are real and nonnegative.

[11] In case of simple stoichiometry $F(a) = a$.

flow reactor shown in figure 3 ($[\mathbf{X}_j] = x_j$ with $j = 1, \ldots, n$):

$$\begin{aligned} \frac{da}{dt} &= -\sum_{j=1}^{n} k_j\, F(a)\, x_j \;+\; r\,(a_0 - a) \quad \text{and} \\ \frac{dx_j}{dt} &= \sum_{i=1}^{n} k_i Q_{ij}\, F(a)\, x_i \;-\; r\, x_j\,;\; j = 1, \ldots, n\ . \end{aligned} \tag{9}$$

The flow is adjustable and we can use programmed flow $r(t)$ such that the concentration of $\mathbf{A}$ becomes constant, $a(t) \to \bar{a} = \text{const}$. Straightforward computation yields

$$\begin{aligned} r(t) &= \frac{1}{\bar{c}} \sum_{i=1}^{n} f_i x_i(t) \quad \text{with} \quad f_i = k_i F(\bar{a}) \text{ and } \bar{c} = a_0 - \bar{a} = \sum_{i=1}^{n} x_i\ , \\ \frac{dx_j}{dt} &= \sum_{i=1}^{n} f_i Q_{ij}\, x_i \;- x_j\, \bar{f};\; \bar{f} = \frac{\sum_{k=1}^{n} f_k x_k}{\bar{c}} \quad \text{and } j = 1, \ldots, n\ . \end{aligned} \tag{10}$$

A transformation of the time axis with a strictly positive function does not change the outcome of the selection process [Eigen and Schuster, 1977; Schuster and Sigmund, 1985] and hence replication-mutation in the flow reactor and in the idealized system (10) lead to the same stationary distribution of RNA molecules.

Equation (10) has been studied in great detail as a model for selection based on replication and mutation [Eigen, 1971; Eigen and Schuster, 1977, 1978a,b; Swetina and Schuster, 1982; Schuster and Swetina, 1988; Nowak and Schuster, 1989; Eigen et al., 1989]. The cases with different behavior of the longtime solutions can be classified with respect to the properties of the value matrix W, which is derived from the vector $\mathbf{f} = F(\bar{a}) \cdot \mathbf{k}$ and the mutation matrix Q according to equation (10): $\mathrm{W} \doteq \{f_i Q_{ij}; i, j = 1, \ldots, n\}$. We mention here the most important cases:

(i) **Selection of single variant**: The vector $\mathbf{f}$ has one single largest component ($f_m > f_i; i = 1, \ldots, n; m \neq i$), Q is diagonal and hence it is the unit matrix (all mutation rates are zero and we are dealing with replication alone). W is diagonal too and the variant $\mathbf{X}_m$ with the largest fitness value, $f_m = \max\{f_i; i = 1, \ldots, n; m \neq i\}$, called *master sequence*, is selected.

(ii) **Selection of quasispecies**: The matrix W is primitive implying that Perron-Frobenius theorem holds [Seneta, 1981, p.3,p.22],[12] and the error rate is below a threshold value $p < p_{\max}$ (see iv). The population converges to a stationary state with a mutant distribution, which is determined by the largest eigenvector of matrix W corresponding to the largest eigenvalue λ_0 with the following properties:

[12] A nonnegative square matrix $\mathrm{A} = (a_{ij}$ is said to be a *primitive* matrix if there exists k such that $A^k \gg 0$, i.e., if there exists k such that for all i, j, the (i, j) entry of A^k is positive. A sufficient condition for a matrix to be a primitive matrix is for the matrix to be an irreducible matrix with positive trace.

1. λ_0 is real and strictly positive,
2. $\lambda_0 > |\lambda_k|$ for all $k \neq 0$,
3. λ_0 is associated with strictly positive eigenvectors, and
4. λ_0 is a simple root of the characteristic equation of W.

Hence, the unique stationary population consists of a master sequence and a distribution of mutants in which all variants are present.

(iii) **Random drift through neutral evolution**: Drifting populations may be the result of neutrality in fitness values. Provided the degree of neutrality is sufficiently large the populations do not reach stationary distributions but drift randomly through sequence space in the sense of neutral evolution [Kimura, 1983; Huynen et al., 1996]. The process can be described successfully as diffusion of the population in sequence space.

(iv) **Random drift through error accumulation**: No stationary population is approached if the error rate exceeds a threshold value, $p > p_{\max} = 1 - \sigma_m^{-1/n}$. Here σ_m denotes the superiority of the master sequence:

$$\sigma_m = f_m/\overline{f_{-m}} \quad \text{with} \quad \overline{f_{-m}} = \frac{\sum_{i=1,i\neq m} f_i\, x_i}{\sum_{i=1,i\neq m} x_i} \; .$$

If the critical value p_{max} is exceeded maintenance of sequences in consecutive replications breaks down because of error accumulation.

Neutrality in the sense of (iii) has been postulated for the interpretation of mutation frequencies found in Nature [Kimura, 1968; King and Jukes, 1969]. More than thirty years later discussions between so-called neutralists and selectionists claiming that the majority of non-deleterious mutations is selectively neutral or adaptive, respectively, have not yet come to an end in molecular evolution [Li and Graur, 1991; Hartl and Clark, 1997; Page and Holmes, 1998; Nei and Kumar, 2000]. Data on optimization of RNA based enzymes, so-called *ribozymes*, however, provide clear evidence for vast selective neutrality of RNA structures and properties [Schultes and Bartel, 2000] as it has been predicted earlier from large scale folding computations [Schuster et al., 1994].

The random drift phenomenon described in (iv) is due to finite sizes of all real populations. The analysis of the stationary solution of the replication-mutation equation (10) in terms of the eigenvalues of the matrix W shows avoided crossing of two eigenvalues at the error threshold, $p = p_{\max}$ [Nowak and Schuster, 1989]. At higher error rates ($p > p_{\max}$) the largest eigenvector is very close to the uniform distribution, $\bar{x}_1 = \bar{x}_2 = \ldots = \bar{x}_m = \ldots = \bar{x}_n$. Maximum population sizes in evolution experiments with molecules are in the range of $N = 10^{16}$ molecules, natural populations of viruses or bacteria are much smaller and hardly exceed $N = 10^{10}$. The numbers of possible sequences (n), however, are much larger: For rather small chain length of 100 $n = 4^{100} \approx 10^{60}$ different polynucleotide sequences are possible. Consequently, $n \gg N$ and a uniform population can never exist. Instead, the population occupies only a tiny fraction of sequence space and mutation drives it to drift randomly.

The error threshold phenomenon (iv) turned out to be highly relevant for virology. In particular the maximum error rate sets a limit to the maximum length of sequences, $n_{\max} \approx \ln \sigma_m / p$ which is reflected by an empirically found relation between replication accuracy and genome length in RNA viruses. Depending on the distribution of the fitness values f_k in sequence space the transition from the stationary quasispecies to the random drift regime may be sharp or smooth. The former case is fulfilled on sufficiently steep and rugged landscapes that give rise to a phase transition in sequence space. The latter case is found with for smooth landscapes [Wiehe, 1997; Baake and Wagner, 2001].

The kinetic theory of molecular evolution in form of the quasispecies concept is based on two implicit assumptions: (i) infinite population size and (ii) sufficient time to reach the stationary distributions of variants. The first assumption is standard in chemical kinetics: Although we are never dealing with infinite populations fluctuations become small for concentrations in molar range.[13] This is often not the case in biology where we have self-enhancing systems like replication, which amplify fluctuations. The second point is no problem in laboratory systems but may be important for coevolution in ecosystem, in particular in host-parasite systems.

5 Computer simulation of molecular evolution

Computer simulations were performed in order to provide data for evolution on realistic biological fitness landscapes. Neither population genetics [Hartl and Clark, 1997] nor the kinetic theory of evolution deal with intrinsic genotype-phenotype relations that allow for comprehensive analysis. For evolution of molecules such landscapes are provided by the mapping (2) for which suitable approximations exist [Schuster, 2003]. The flow reactor (figure 3) is used also for the computer simulations of RNA evolution [Fontana and Schuster, 1998b]. Population sizes up to $N = 100\,000$ can be handled. The algorithm applied [Gillespie, 1976, 1977] computes individual trajectories, which simulate directly chemical reactions through encounters of molecules in homogeneous medium, gas phase or solution. Sampling of trajectories provides (statistical) approximations to the solutions of the corresponding Master equation. Accordingly, the time of computation is proportional to real time.

Basic results of the simulations are shown by means of a single trajectory showing the mean distance of the population from the target (figure 5). Evolutionary optimization of RNA structure does not occur gradually but shows a stepwise or punctuated progress, which is the result of processes on two time scales. Fast adaptive phases during which the distance to target decreases are interrupted by long quasi-stationary epochs or plateaus with no progress in the approach to the target structure. Explanation of this finding is straightforward: The mapping of RNA sequences into structures is characterized by a high degree of neutrality and therefore the population spreads on

[13] For all systems at equilibrium and for most chemical reactions the fluctuations are in the order of the square root of particle numbers: $\bar{N} \pm \sigma(N) \pm \sqrt{\bar{N}}$. For one mole particles $\bar{N} \approx 6 \times 10^{23}$ and $\sigma(N) \approx 8 \times 10^{11}$, which is too small to detect.

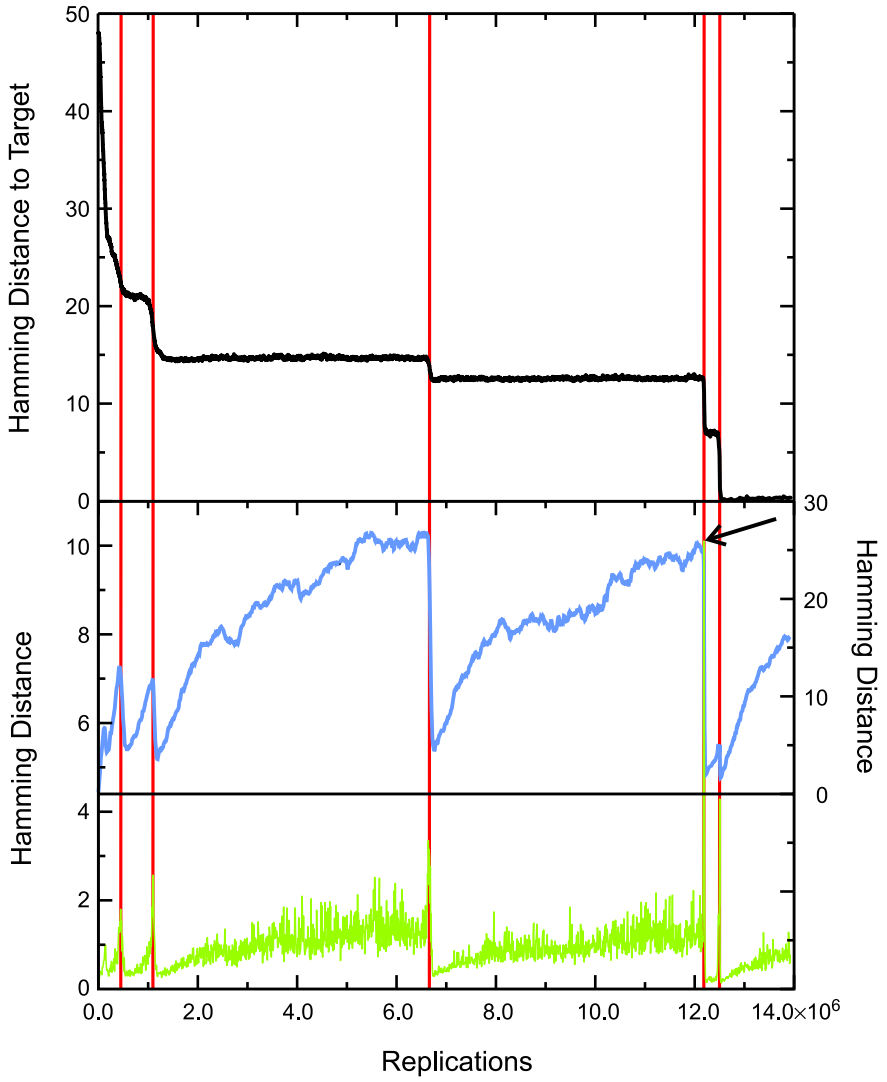


Figure 5: Evolutionary optimization of RNA structure. Shown is a single trajectory of a simulation of RNA optimization towards a tRNA^{phe} target with population size $n = 3000$ and mutation rate $p = 0.001$ per site and replication. The figure shows as functions of time: (i) the distance to target averaged over the whole population, $\overline{d_S(S_i, S_T)}(t)$ (black), (ii) the mean Hamming distance within the population, $\overline{d_P}(t)$ (blue, right ordinate), and (iii) the mean Hamming distance between the populations at time t and $t + \Delta t$, $\overline{d_C}(t, \Delta t)$ (green)with a time increment of $\Delta t = 8\,000$. The end of plateaus (vertical red lines) are characterized by a collapse in the width of the population and a peak in the migration velocity corresponding to a jump in sequence space. The arrow indicates a remarkably sharp peak of $d_C(t, 8\,000)$ around Hamming distance 10 at the end of the second long plateau ($t \approx 12.2 \times 10^6$ replications). In other words, every adaptive phase is accompanied by a drastic reduction in genetic diversity, $d_P(t)$. The diversity increases during quasi-stationary epochs. On the plateaus the center of the cloud migrates only at a speed of Hamming distance 0.125 per 1 000 replications.

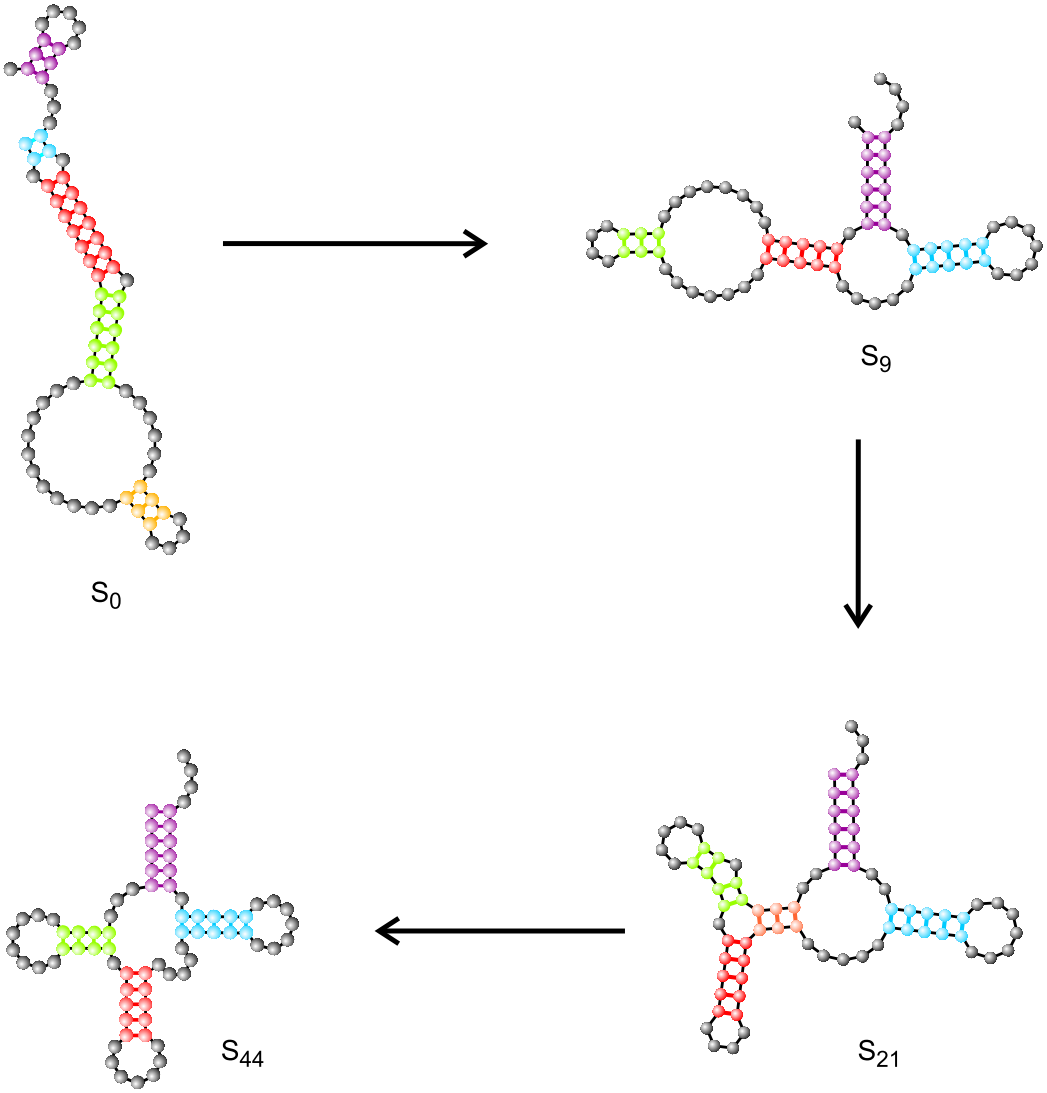


Figure 6: Intermediate structures of structure optimization in the flow reactor. We show the initial structure S_0, structure S_9, which is characterized by a particularly long plateau in the trajectory, structure S_{21} that is one step before the formation of the cloverleaf, and the target structure S_{44}.

a set of selectively neutral sequences within the time span of the horizontal plateaus. During this diffusion in sequence space many new genotypes are formed and the process continues until a mutant is found that allows for a continuation of the approach towards target. Adaptive phases are commonly initiated by a major rearrangement in structure that brings the population closer to the target structure. Examples of such major changes in structure are shown in figure 6. It is worth mentioning that a similar stepwise optimization process has been observed with evolution of bacterial populations in serial transfer experiments under constant conditions [Elena et al., 1996; Papadopoulos et al., 1999].

Sampling of trajectories and analysis revealed that (i) above a sharp threshold in population size around $N = 16$ almost all trajectories reach the target structure, (ii) the particular sequence that forms the target structure S_T is almost always different in different simulations, and (iii) at population

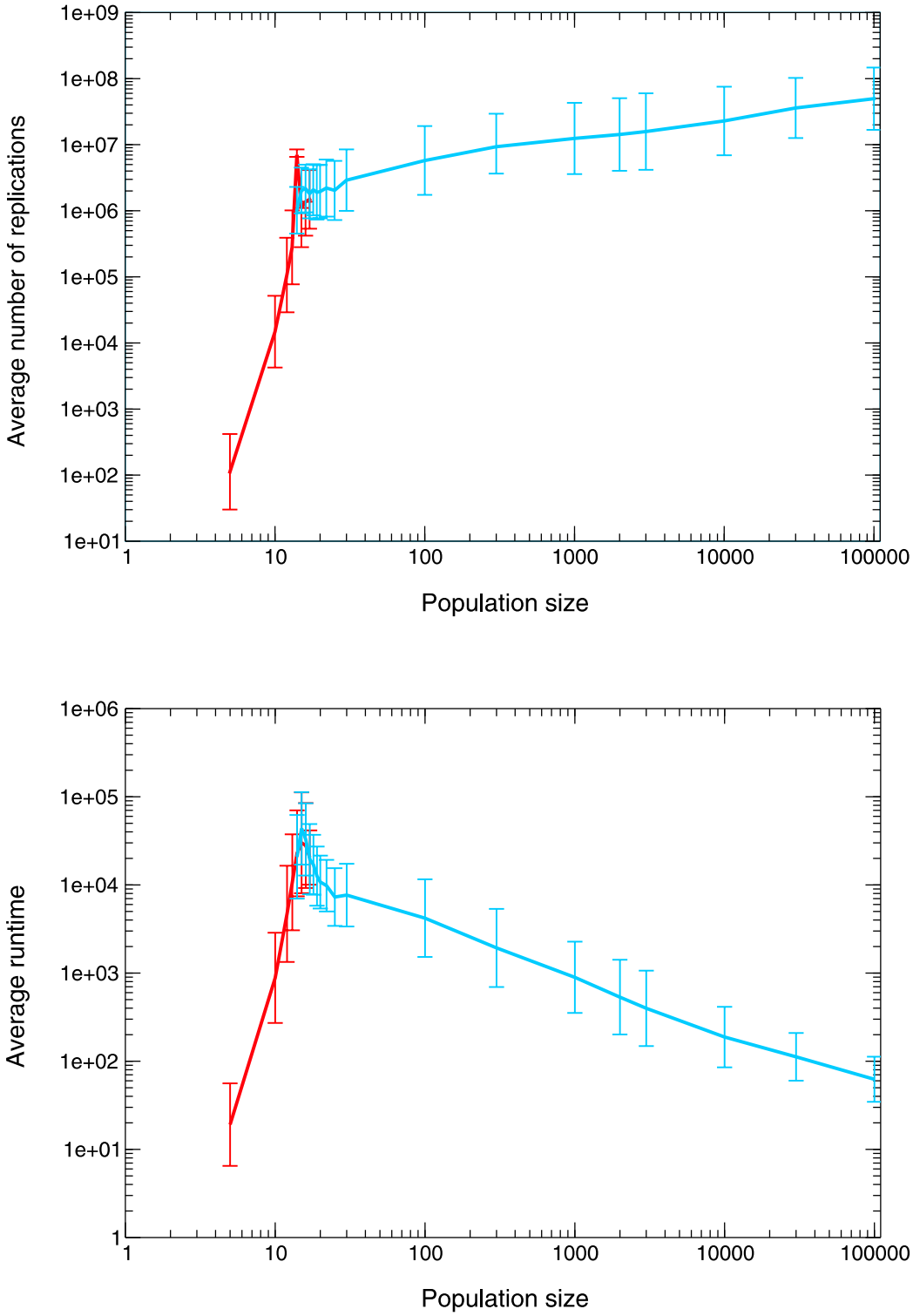


Figure 7: Scatter in length of trajectories in evolutionary optimization. The stochastic process underlying the simulation has only two absorbing states: (i) extinction, $\Upsilon_0 : \{X_i = 0 \text{ for all } i = 1, 2, \ldots, n\}$, and (ii) successful approach to target, $\Upsilon_1 : \{\exists \mathbf{X}_k \text{ such that } \psi(X_k) = S_T \text{ and } X_k \neq 0\}$. Hence, trajectories end either in Υ_0 or in Υ_1. The two plots show the average number of replications (top) or the average runtime (bottom) to reach either (red) or Υ_1 (blue). The transition from almost all trajectories going to Υ_0 to almost all trajectories going to Υ_1 is sharp an lies in the range between $N = 13$ and $N = 19$. The error bars refer to $\pm\sigma$ with σ being the standard deviations in log-normal distributions.

sizes above threshold the runtimes required to reach the target structure and the number of replications show vast scatter. Figure 7 presents mean values and error bars corresponding to $\bar{t}_T \pm \sigma(t_T)$ where $(t_T)_j$ is the time at which trajectory 'j' reaches the target and '$\sigma^2(t_T)$' represents the variance or the square of the standard deviation of the first passage times to target. In contrast to the common $\sqrt{N}$ law for standard deviations, here the scatter remains large with increasing population size. The mean number of replications increases slightly with population size whereas we observe a pronounced decrease in the time to reach target when the population becomes larger. The take-home lesson for carrying out evolution in the test-tube is therefore: Large population sizes should by applied if time is the limiting factor but population sizes as small as possible are recommended when the limitation is in the material resources.

A recent computer simulation [Kupczok and Dittrich, 2006] using the flow reactor, trajectory sampling and a similar RNA target as in Fontana and Schuster [1998b] proved the existence of the error threshold predicted by the kinetic theory of molecular evolution.

6 Concluding remarks

The great success of Ludwig Boltzmann was to introduce proper statistics that starts out from random molecular encounters and leads to the deterministic quantities of thermodynamics. As expressed by the $\sqrt{N}$-law, fluctuations that are highly important at low particle numbers become negligibly small at molar concentrations. Self-enhancement, in particular when it is caused by reproduction, gives rise to different laws for fluctuations and introduces indeterminism on macroscopic scales. A prominent example is the hypothetical approach to the uniform distribution where the number of possible molecular species exceeds the accessible population sizes by many orders of magnitude (section 4). The problems that were encountered in attempts to do proper statistics in evolution provide, at the same time, the basis for the enormous diversity and beauty of Nature. Biological information space built upon sequence diversity in inexhaustible.

Since its spectacular beginnings in 1953 molecular biology revealed piece after piece the magnificent chemistry and physics of life. Determinations of three-dimensional molecular architectures progressed to larger and larger units and provided deep insights in to the mysteriously successful chemistry performed by cells and organisms in the sense of Ludwig Boltzmann's statement cited in the beginning of this chapter [Boltzmann, 1979, p. 41]. The new developments initiated in the nineteen eighties pushed molecular biology and molecular genetics up to the level of entire cells or organisms: Genomics, proteomics, metabolomics, and systems biology are aiming at determining and analyzing all genes, all proteins and their interactions, all metabolic reactions and eventually the dynamics of complete regulatory and synthesizing networks.

In a way John Maynard Smith's vision [Maynard Smith, 1986, p. vii] has become reality. The reductionists' programm has reached a state where cellular and organismic problems can be handled on the molecular level. The new

discipline systems biology [Klipp et al., 2005; Palsson, 2006] aims at modelling complete genetic regulatory and metabolic networks of whole cells. Holistic questions like homeostasis, autopoiesis, cell cycle regulation, circadian rhythms and many others can now be addressed by the molecular approach. Explanations can be given by means of known processes at the lower hierarchical level that is biochemistry and chemical kinetics. The physicists and the chemists who have entered biology became biologists meanwhile studying biological problems with the techniques they brought with them from their original fields of research.

References

Baake, E. and Wagner, H.: 2001, Mutation-selection models solved exactly with metjods of statistical mechanics, *Genet. Res. Camb.* **78**, 93–117.

Bateson, W.: 1902, *Mendel's Priciples of Heridity. A Defense*, first edn, Cambridge University Press, London.

Bauer, G., McCaskill, J. and Otten, H.: 1989, Traveling waves of *in vitro* evolving RNA, *Proc. Natl. Acad. Sci.* **86**, 7937–7941.

Biebricher, C. K.: 1983, Darwinian selection of self-replicating RNA molecules, *in* M. K. Hecht, B. Wallace and G. T. Prance (eds), *Evolutionary Biology, Vol. 16*, Plenum Publishing Corporation, pp. 1–52.

Biebricher, C. K. and Eigen, M.: 1988, Kinetics of RNA replication by Qβ replicase, *in* E. Domingo, J. J. Holland and P. Ahlquist (eds), *RNA Genetics. Vol.I: RNA Directed Virus Replication*, CRC Press, Boca Raton, FL, pp. 1–21.

Biebricher, C. K. and Gardiner, W. C.: 1997, Molecular evolution of RNA *in vitro*, *Biophys. Chem.* **66**, 179–192.

Boltzmann, L.: 1979, *Populäre Schriften*, Friedr.Vieweg & Sohn, Braunschweig, DE. In German. Eingeleitet und ausgewählt von Engelbert Broda.

Broda, E.: 1983, *Ludwig Boltzmann. Man, Physicist, Philosopher*, Ox Bow Press, Woodbridge, CT.

Caplan, A. L. (ed.): 1978, *The Sociobiology Debate*, Harper & Row, New York.

Darwin, C.: 1859, *Of the Origin of Species by means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, Vol. 811 of *Everyman's Library (Printing 1967, J.M.Dent & Sons, London)*, Murray, London.

Dawkins, R.: 1986, *The Blind Watchmaker*, Longman Scientific & Technical, Harlow, Essex, UK.

Dennett, D. C.: 1995, *Darwin's Dangerous Idea. Evolution and the Meanings of Life*, Simon & Schuster, New York.

Dobzhansky, T., Ayala, F. J., Stebbins, G. L. and Valentine, J. W.: 1977, *Evolution*, W. H. Freeman & Co., San Francisco, CA.

Eaton, S. V.: 1952, Effects of phosphorous deficiency on growth and metabolism of black mustard, *Botanical Gazette* **113**, 301–309.

Eigen, M.: 1971, Selforganization of matter and the evolution of biological macromolecules, *Naturwissenschaften* **58**, 465–523.

Eigen, M.: 1993, The origin of genetic information: Viruses as models, *Gene* **135**, 37–47.

Eigen, M., McCaskill, J. and Schuster, P.: 1989, The molecular quasispecies, *Adv. Chem. Phys.* **75**, 149–263.

Eigen, M. and Schuster, P.: 1977, The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle., *Naturwissenschaften* **64**, 541–565.

Eigen, M. and Schuster, P.: 1978a, The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle., *Naturwissenschaften* **65**, 7–41.

Eigen, M. and Schuster, P.: 1978b, The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle, *Naturwissenschaften* **65**, 341–369.

Elena, S. F., Cooper, V. S. and Lenski, R. E.: 1996, Punctuated evolution caused by selection of rare beneficial mutants, *Science* **272**, 1802–1804.

Fontana, W. and Schuster, P.: 1998a, Shaping space. The possible and the attainable in RNA genotype-phenotype mapping, *J.Theor.Biol.* **194**, 491–515.

Fontana, W. and Schuster, P.: 1998b, Continuity in evolution. On the nature of transitions, *Science* **280**, 1451–1455.

Gillespie, D. T.: 1976, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comp. Phys.* **22**, 403–434.

Gillespie, D. T.: 1977, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* **81**, 2340–2361.

Haken, H.: 1977, *Synergetics. An Introduction. Nonequilibrium Phase Transitions and Self-Orgaization in Physics, Chemistry and Biology*, Springer-Verlag, Berlin.

Haken, H.: 1983, *Advanced Synergetics. Istability Hierarchies of Self-Organizing Systems and Devices*, Springer-Verlag, Berlin.

Hartl, D. L. and Clark, A. G.: 1997, *Principles of Population Genetics*, third edn, Sinauer Associates, Sunderland, MA.

Huynen, M. A., Stadler, P. F. and Fontana, W.: 1996, Smoothness within ruggedness: The role of neutrality in adaptation, *Proc. Natl. Acad. Sci. USA* **93**, 397–401.

Judson, H. F.: 1979, *The Eighth Day of Creation. Makers of the Revolution in Biology*, Jonathan Cape, London.

Kimura, M.: 1968, Evolutionary rate at the molecular level, *Nature* **217**, 624–626.

Kimura, M.: 1983, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, UK.

King, J. L. and Jukes, T. H.: 1969, Non-Darwinian evolution, *Science* **164**, 788–798.

Klipp, E., Herwig, R., Kowald, A., Wierling, C. and Lehrach, H.: 2005, *System Biology in Practice. Concepts, Implementation, and Application*, Wiley-VCh, Weinheim, DE.

Klussmann, S. (ed.): 2006, *The Aptamer HAndbook. Functional Oligonucleotides and Their Applications*, Wiley-VCh Verlag, Weinheim, DE.

Kupczok, A. and Dittrich, P.: 2006, Determinants of simulated RNA evolution, *J. Theor. Biol.* **238**, 726–735.

Kutschera, U.: 2006, *Evolutionsbiologie*, 2nd edn, Verlag Eugen Ulmer, Stuttgart, DE. In German.

Kutschera, U. and Niklas, K. J.: 2004, The modern theory of biological evolution. An expanded synthesis, *Naturwissenschaften* **91**, 225–276.

Li, W.-H. and Graur, D.: 1991, *Fundamentals of Molecular Evolution*, Sinauer Associates, Sunderland, MA.

Maynard Smith, J.: 1986, *The Problems of Biology*, Oxford University Press, Oxford, UK.

Mayr, E.: 1982, *The Growth of Biological Thought. Diversity, Evolution, and Inheritance*, Harvard Univesity Press, Cambridge, MA.

Mayr, E. and Provine, W. (eds): 1980, *The Evolutionary Synthesis: Perspectives on the Unification of Biology*, Harvard University Press, Cambridge, MA.

Mills, D. R., Peteron, R. L. and Spiegelman, S.: 1967, An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule, *Proc. Natl. Acad. Sci. USA* **58**, 217–224.

Nei, M. and Kumar, S.: 2000, *Molecular Evolution and Phylogenetics*, Oxfor University Press, New York.

Nicolis, G. and Prigogine, I.: 1977, *Self-Organization in Nonequilibrium Systems. From Dissipative Structures to Order through Fluctuations*, John Wiley & Sons, New York.

Nowak, M. and Schuster, P.: 1989, Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet., *J. Theor. Biol.* **137**, 375–395.

Öhlenschlager, F.: 1997, 30 years later – A new approach to Sol Spiegelman's and Leslie Orgel's *in vitro* evolutionary studies, *Orig. Life Evol. Biosph.* **27**, 437–457.

Page, R. D. M. and Holmes, E. C.: 1998, *Molecular Evolution. A Phylogenetic Approach*, Blackwell Science, Oxford, UK.

Palsson, B. Ø.: 2006, *Systems Biology. Properties of Reconstructed Networks*, Cambridge University Press, New York.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E. and Blot, M.: 1999, Genomic evolution during a 10 000-generation experiment with bacteria, *Proc. Natl. Acad. Sci. USA* **96**, 3807–3812.

Reidys, C., Stadler, P. F. and Schuster, P.: 1997, Generic properties of combinatory maps. Neutral networks of RNA secondary structure., *Bull.Math.Biol.* **59**, 339–397.

Reif, W.-E., Junker, T. and Hoßfeld, U.: 2000, The synthetic theory of evolution: General problems and the German contribution to the synthesis, *Theory Biosci.* **119**, 41–91.

Riddell, W. H., Hughes, J. S. and Fitch, J. B.: 1934, The relation of phorphorous deficiency to the utilization of feed in dairy cattle, *Technical Bulletin 36*, Kansas State College of Agriculture and Applied Science, Manhattan, KS.

Ruse, M.: 1979, *The Darwinian Revolution*, The University of Chicago Press, Chicago, IL.

Schödinger, E.: 1944, *What is Life? The Physical Aspect of the Living Cell*, Cambridge University Press, Cambridge, UK.

Schultes, E. and Bartel, D.: 2000, One sequence, two ribozymes: Implications for the emergence of new ribozyme folds, *Science* **289**, 448–452.

Schuster, P.: 2003, Molecular insight into the evolution of phenotypes, *in* J. P. Crutchfield and P. Schuster (eds), *Evolutionary Dynamics – Exploring the Interplay of Accident, Selection, Neutrality, and Function*, Oxford University Press, New York, pp. 163–215.

Schuster, P.: 2006, Prediction of RNA secondary structures: From theory to models and real molecules, *Reports on Progress in Physics* **69**, 1419–1477.

Schuster, P., Fontana, W., Stadler, P. F. and Hofacker, I. L.: 1994, From sequences to shapes and back: A case study in RNA secondary structures, *Proc. Roy. Soc. London B* **255**, 279–284.

Schuster, P. and Sigmund, K.: 1985, Dynamics of evolutionary optimization., *Ber. Bunsenges. Phys. Chem.* **89**, 668–682.

Schuster, P. and Swetina, J.: 1988, Stationary mutant distribution and evolutionary optimization, *Bull. Math. Biol.* **50**, 635–660.

Seneta, E.: 1981, *Non-Negative Matrices and Markov Chains*, second edn, Springer-Verlag, New York.

Spiegelman, S.: 1971, An approach to the experimental analysis of precellular evolution, *Quart. Rev. Biophys.* **4**, 213–253.

Strunk, G. and Ederhof, T.: 1997, Machines for automated evolution experiments in vitro based on the serial tranfer concept, *Biophys. Chem.* **66**, 193–202.

Swetina, J. and Schuster, P.: 1982, Self-replication with errors - A model for polynucleotide replication, *Biophys. Chem.* **16**, 329–345.

Thompson, D. W.: 1942, *On Growth and Form*, Cambridge University Press, Cambridge, UK. The first edition was published 1917 in Cambridge, UK.

Wallace, A. R.: 1889, *Darwinism. An Exposition of the Theory of Natural Selection with Some of its Applications*, Macmillan & Co, London.

Wardle, D. A., Walker, L. R. and Bardgett, R. D.: 2004, Ecosystem properties and forest decline in contrasting long-term chronosequences, *Sciecne* **305**, 509–513.

Watson, J. D. and Crick, F. H. C.: 1953, Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid, *Nature* **171**, 737–738.

Weismann, A.: 1892, *Das Keimplasma. Eine Theorie der Vererbung*, G. Fischer-Verlag, Jena, DE. In German.

Wiehe, T.: 1997, Model dependency of error threshold: The role of fitness functions and contrasts between finite and infinite size models, *Genet. Res. Camb.* **69**, 127–136.