ARTIFICIAL CHEMISTRY AND MOLECULAR DARWINIAN EVOLUTION *in silico*

Vladimír KVASNIČKA[1,*] and Jiří POSPÍCHAL[2]

Department of Mathematics, Slovak Technical University, 812 37 Bratislava, Slovakia; e-mail: [1] kvasnic@stuba.sk, [2] pospich@stuba.sk

Received July 10, 2002
Accepted October 10, 2002

Dedicated to Professor Petr Čársky, Professor Ivan Hubač and Professor Miroslav Urban on the occasion of their 60th birthdays.

A simplified model of Darwinian evolution at the molecular level is studied by applying the methods of artificial chemistry. A chemical reactor (chemostat) contains molecules that are represented by binary strings, the strings being capable of replication with a probability proportional to their fitness. Moreover, the process of replication is not fully precise, sporadic mutations may produce new offspring strings, which are slightly different from their parent templates. The dynamics of such an autoreplicating system is described by Eigen's differential equations. These equations have a unique asymptotically stable state, which corresponds to those strings that have the highest rate constants (fitness). Fitness of binary string is calculated as a graph-theory similarity between a folding (phenotype) of respective string and the so-called required folding. The presented method offers a detailed view of mechanisms of the molecular Darwinian evolution, in particular of the meaning and importance of neutral mutations.

Keywords: Artificial life; Artificial chemistry; Fitness landscape; Eigen theory of replicators; Molecular Darwinian evolution; Neutral mutations; Neutral evolution.

...the central problem of evolution... is that of a mechanism by which the species may continually find its way from lower to higher peaks.

(*Sewall Wright*[1])

Darwinian evolution belongs to standard subjects of interest of *Artificial Life*[2]. In particular, the main stimulus was observed at the end of the eighties, when evolutionary algorithms suddenly emerged as a new paradigm of computer science based on the metaphor of Darwinian evolution. This paradigm may be traced back to 1931 when Wright[1] has postulated an *adaptive landscape* (nowadays called the *fitness landscape* or *fitness surface*) and char-

acterized Darwinian evolution as an adaptive process (nowadays we say an optimization process), where the genotype of population is adapted in such a way that it reaches a local (even global) maximum on the fitness surface. Forty years later, this ingenious idea was used by Holland[3] as a metaphor for creation of *genetic algorithms*, which may be now interpreted as an abstraction of Darwinian evolution in the form of universal optimization algorithm[4] (see Dennett's seminal book[5]).

In this paper we present a very simple computational model of Darwinian evolution, which may reflect some of its most elementary aspects appearing on biomacromolecular level (*e.g.* see experiments of Spiegelman[6] from the end of the sixties). This "molecular" model of Darwinian evolution is capable of offering a detailed quantitative view of many of its basic concepts and notions, *e.g.* the role of neutral mutations may be studied as an auxiliary device to overcome local valleys of fitness surface in the course of the adaptation process. Artificial chemistry plays a very important role in the present study[7–14], which may be considered as a subfield of artificial life based on the metaphor of chemical reactor (in our forthcoming discussions called the chemostat). The chemostat is composed of "molecules" that are represented by abstract objects (*e.g.* by strings composed of symbols, trees, formulae constructed within algebra, *etc.*), which are transformed stochastically into other feasible objects by "chemical reactions". The probability of these transformations is strictly determined by the structure of incoming objects, resulting – outgoing objects are returned to the chemostat. Kinetics of the processes running in the chemostat is well described by Eigen's replicator differential equations[15,16], which were constructed on the basis of the well-known physico-chemical law of mass action. The main objects of artificial chemistry are (i) to study formal systems that are based on the chemical metaphor and that are potentially able to perform parallel computation, and (ii) to generate formal autocatalytic chemical systems (molecules are represented by structured objects) for purposes of "*in silico*" simulations of an emergence of living systems.

Recently, Newman and Engelhardt[17] have shown that many aspects of the molecular Darwinian evolution *in silico* may be studied by making use of a fitness surface based on a generalization of Kauffman KN [18,19] rugged function with tunable degree of neutrality. They demonstrated that almost all basic results obtained by Schuster's Vienna group[20–24], based on a realistic physico-chemical model of RNA molecules and their folding, may be simply and immediately obtained by this simple model of fitness surface (we have to note that the first two above mentioned papers[20,21] also used a binary string representation of replicators). Recently, we have used[14] this

simple model of fitness surface for a simulation of the molecular Darwinian evolution and the obtained results were formally closely related to or even almost identical with the theoretical results predicted by Eigen's replicator differential equations. We repeat our previous results with binary strings in place of "molecules" that fill an artificial-chemistry chemostat. Each binary string is evaluated by a planar secondary structure called folding. The secondary structure of binary strings specifies their fitness function, which is directly related to the probability of a replication process of strings. The folding of binary strings – genotypes – is considered as a phenotype that specifies the fitness of the genotype. This simple triad **genotype – phenotype – fitness** is fundamental for our computer simulation of the molecular Darwinian evolution, such that the evolution is interpreted through changes of string phenotypes.

THEORY

Eigen's Replicators

Manfred Eigen published in the early seventies a seminal paper entitled "*Self organization of matter and the evolution of biological macromolecules*"[15], where he postulated a hypothetical chemical system composed of the so-called replicators. This system mimics Darwinian evolution even on an abiotic level. Later, Eigen and Schuster[16] discussed the proposed model as a possible abiotic mechanism for increasing complexity on a border of abiotic and biotic systems.

Let us consider biomacromolecules (called the *replicators*) X_1, X_2, ..., X_n, which are capable of the following chemical reactions:

$$X_i \xrightarrow{k_i} X_i + X_i \qquad (i = 1, 2, ..., n) \tag{1a}$$

$$X_i \xrightarrow{\phi} \varnothing \qquad (i = 1, 2, ..., n) \tag{1b}$$

The reaction (*1a*) means that a molecule X_i is replicated onto itself with a rate constant k_i, whereas the reaction (*1b*) means that X_i becomes extinct with a rate parameter ϕ (this parameter is called the "*dilution flux*" and will be specified further). Applying the mass-action law of chemical kinetics, we get the following system of differential equations

$$\dot{x}_i = x_i(k_i - \phi) \quad (i = 1, 2, \ldots, n) \tag{2a}$$

The dilution flux ϕ is a free parameter and it will be determined in such a way, that the following condition is satisfied: a sum of time derivatives of concentrations $\dot{x}$'s is vanishing, $\sum \dot{x}_i = 0$, we get

$$\dot{x}_i = x_i\left(k_i - \sum_{j=1}^{n} k_j x_j \right) \quad (i = 1, 2, \ldots, n) \tag{2b}$$

where the condition $\sum x_i = 1$ is used without a loss of generality of our considerations. Its analytical solution is as follows

$$x_i(t) = \frac{x_i(0)\,\mathrm{e}^{k_i t}}{\sum_{j=1}^{n} x_j(0)\,\mathrm{e}^{k_j t}} \tag{3}$$

This solution has an asymptotic property, where only one type of molecules (with the maximal rate constant $k_{\max}$) is surviving while other ones become extinct

$$\lim_{t\to\infty} x_i(t) = \begin{cases} 1 & (\text{for } k_i = k_{\max} = \max\{k_1, \ldots, k_n\}) \\ 0 & (\text{otherwise}) \end{cases} \tag{4}$$

Loosely speaking, each type of molecules may be considered as a type of species with a fitness specified by the rate constant k. In a chemostat "survive" only those molecules – species that are best fitted, *i.e.* that have the highest rate constant $k_{\max}$, and all other molecules with smaller rate constants become extinct (Fig. 1). The condition of invariability of the sum of concentrations (*i.e.* $\sum x_i = 1$) introduces a "selection pressure" to replicated molecules, only those molecules will survive that are best fitted with the maximal rate constant.

The proposed model may be simply generalized[15] in such a way that mutations are introduced into process of replications. The system (*2*) is modified as follows

$$\dot{x}_i = x_i(k_{ii} - \phi) + \sum_{\substack{j=1 \\ (j\neq i)}}^{n} k_{ji} x_j \quad (i = 1, 2, \ldots, n) \tag{5}$$

where k_{ij} is a rate constant assigned to a modified reaction (*1a*)

$$X_i \xrightarrow{k_{ij}} X_i + X_j \qquad (i, j = 1, 2, \ldots, n) \tag{6}$$

There is postulated that a rate constant matrix $\boldsymbol{K} = (k_{ij})$ has dominant diagonal elements, *i.e.* nondiagonal elements are much smaller than diagonal ones. This requirement directly follows from an assumption that imperfect replications (*6*) are very rare, the product X_j is considered as a weak mutation of autoreplicated X_i, $X_j = O_{\mathrm{mut}}(X_i)$. The dilution flux ϕ from (*5*) is determined by the condition that the sum of time derivatives of concentrations is vanishing, $\sum \dot{x}_i = 0$. We get

$$\phi = \sum_{i,j=1}^{n} k_{ij} x_j \tag{7}$$

Analytical solution of (*5*) with dilution flux specified by (*7*) is[25]

$$x_i(t) = \frac{\sum_{j=1}^{n} q_{ij} x_i(0) \mathrm{e}^{\lambda_i t}}{\sum_{m,j=1}^{n} q_{mj} x_j(0) \mathrm{e}^{\lambda_j t}} \tag{8}$$

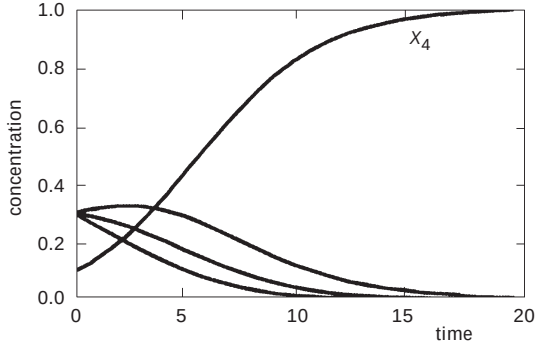

FIG. 1
A plot of relative concentrations of four component system with rate constants $k_1 = 1$, $k_2 = 2$, $k_3 = 3$, and $k_4 = 4$. We can see that only molecules X_4 survive in the end

where $\boldsymbol{Q} = (q_{ij})$ is a nonsingular matrix that diagonalizes the rate matrix $\boldsymbol{K}$, $\boldsymbol{Q}^{-1}\boldsymbol{K}\boldsymbol{Q} = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$.

Since we have postulated that nondiagonal elements of $\boldsymbol{K}$ are much smaller than its diagonal elements, its eigenvalues λ's are very close to diagonal elements, $\lambda_i \doteq k_{ii}$, and the transformation matrix $\boldsymbol{Q}$ is closely related to a unit matrix, $q_{ij} \doteq \delta_{ij}$ (a Kronecker's delta symbol). This means that introduction of weak mutations does not change dramatically general properties of the above simple replicator system without mutations. In particular, the final state of chemostat (for $t \to \infty$) will be composed almost entirely of molecules with the highest rate constant $k_{\max}$. These molecules are weakly accompanied by other replicators with rate constants k's slightly smaller than $k_{\max}$.

Replicators and Molecular Darwinian Evolution

Eigen's system of replicators with mutations (*i.e.* with imperfect replications) presented in the previous section can be simply used for description of the *molecular Darwinian evolution*. Let us consider a hypothetical reaction system composed of a sequence of n replicators X_1, X_2, X_3, ..., X_n. They are endowed with a property that X_i may produce by imperfect replication the juxtaposed replicators $X_{i\pm1}$ (Fig. 2, diagram A). If an initial concentration of X_1 is $x_1(0) = 1$, then in the course of evolution there exist concentration waves that are consecutively assigned X_2, X_3, ..., X_n (see Fig. 2, diagram B).

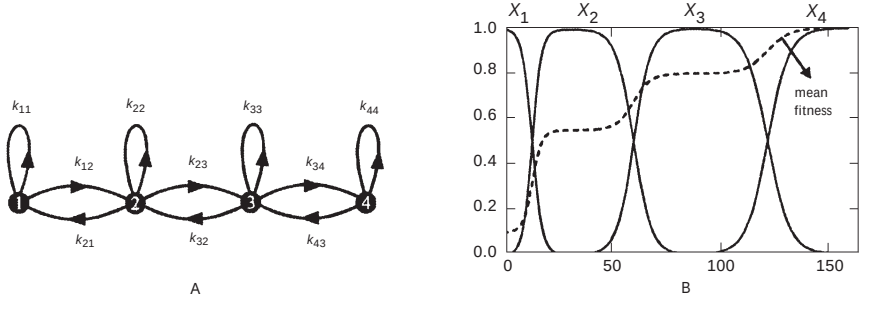

FIG. 2

Diagram A represents a four-replicator system, where a replicator X_i produces by imperfect replications juxtaposed replicators X_{i-1} and X_{i+1}. Edges of the diagram are evaluated by rate constants, their numerical values are specified by matrix $\boldsymbol{K}$ (see Eq. (*12*)), where diagonal elements are well separated and much greater than nondiagonal ones. Diagram B displays plots of replicator concentration profiles that form a sequence of concentration waves. This diagram also contains a plot of mean fitness specified by $\bar{k} = k_{11}x_1 + ... + k_{44}x_4$, which forms a typical nondecreasing "stair" function

This fact may be simply interpreted as a manifestation of the molecular Darwinian evolution, where fitness of single "species" are specified by diagonal rate constants k_{ii}. The evolution process was started by a population composed entirely of X_1. Since its replication is imperfect, it may occasionally produce (specified by the rate constant k_{12}) the next replicator X_2 with a greater fitness (rate constant k_{22}) than its predecessor X_1 ($k_{11} < k_{22}$). This means that in the course of the forthcoming evolution, "species" X_2 will survive (*i.e.* its concentration will climb almost to unity while the concentration of X_1 will be falling to a very small value). This process is repeated for the replicator X_2 considered now as an initial replicator (it plays the same role as the replicator X_1 in the previous stage). Since the replication of X_2 is imperfect, "species" X_3 is produced with a low probability. However, since X_3 has a greater fitness than its predecessor X_2 ($k_{33} > k_{22}$), this new "species" X_3 will survive. This process is finished when the last replicator X_n has appeared initially as a consequence of imperfect replication of X_{n-1} and then its concentration increases to unity by its autoreplication.

In order to formalize on a semiquantitative level the above considerations, let us assume that the replicator system in time t_0 is situated in such a transient state, where the concentration of X_i is almost unity, whereas concentrations of juxtaposed two replicators X_{i-1} and X_{i+1} are very small, *i.e.* $x_i(t_0) = 1 - 2\delta$, $x_{i-1}(t_0) = x_{i+1}(t_0) = \delta$, and $x_j(t_0) = 0$ for other remaining concentrations, where δ is a small positive number. This means that dilution flux $\phi(t)$ (*7*) is specified by

$$\begin{aligned}\phi(t_0) = {} & \delta(k_{i-1,i-2} + k_{i-1,i-1} + k_{i-1,i}) + (1 - 2\delta)(k_{ii} + k_{i,i-1} + k_{i,i+1}) + \\ & + \delta(k_{i+1,i+1} + k_{i+1,i} + k_{i+1,i+2})\end{aligned} \tag{9}$$

Then differential equation (*5*) is as follows:

$$\dot{x}_i = x_i(k_{ii} - \phi) + x_{i-1}k_{i-1,i} + x_{i+1}k_{i+1,i} \tag{10}$$

If we introduce here the above specifications of concentrations for the time t_0 and the assumption that $\dot{x}_i(t_0) = 0$, we get $0 = -(k_{i,i-1} + k_{i,i+1}) + \delta(k_{i-1,i} + k_{i+1,i})$, or

$$\delta = \frac{k_{i,i-1} + k_{i,i+1}}{k_{i-1,i} + k_{i+1,i}} \to 0 \quad \Rightarrow \quad k_{i,i-1} + k_{i,i+1} << k_{i-1,i} + k_{i+1,i} \tag{11}$$

This means that the assumption of good separability of concentration waves gives a strong condition for rate constants touching the replicator X_i in particular: we may say that for a particular replicator the sum of "outgoing" rate constants must be much smaller than the sum of "incoming" rate constants. In other words, loosely speaking, **the "probability" of creation of a particular replicator from its juxtaposed replicators by their imprecise replications must be much greater than the "probability" of destroying the respective replicator by its imprecise replication**.

The above simple considerations are numerically verified for a simple four-replicator system with rate constants specified by the following matrix of rate coefficients

$$\boldsymbol{K} = \begin{pmatrix} 0.1 & 10^{-3} & 0 & 0 \\ 10^{-7} & 0.55 & 10^{-7} & 0 \\ 0 & 10^{-11} & 0.8 & 10^{-11} \\ 0 & 0 & 10^{-7} & 1.0 \end{pmatrix} \qquad (12)$$

This matrix satisfies both the above postulated conditions: (i) Its diagonal matrix elements are much greater than its nondiagonal ones, and (ii) the nondiagonal rate constants satisfy inequalities of the type (*11*). This means that we may expect a Darwinian behavior of the replicator system. This expectation nicely coincides with our numerical results displayed in Fig. 2, diagram B, where concentration profiles of single replicators are shown, which form a sequence of concentration waves typical of Darwinian evolution. Summarizing the present section, we may say that **Eigen's phenomenological theory of replicators forms a proper theoretical framework for numerical studies of the molecular Darwinian evolutionary theory** (*i.e.* applied to biomolecules capable of replication process like RNA or DNA).

Artificial Chemistry and a Metaphor of Chemostat

Let us consider a *chemostat* (chemical reactor)[9,11] composed of formal objects called the molecules. It is postulated that the chemostat is not spatially structured (in chemistry it is said that the reactor is well stirred). Molecules are represented by formal structured objects (*e.g.* token strings, rooted trees, λ-expressions, *etc.*). An interaction between molecules is able to transform information, which is stored in the composition of the molecules. Therefore a chemical reaction (causing changes in the internal struc-

ture of reacting molecules) can be considered as an act of information processing. The capacity of the information processing depends on the complexity of molecules and chemical reactions between them.

General ideas of the chemostat approach will be explained by an example of chemostat as a binary function optimizer that resembles many features of the molecular Darwinian evolution emphasized in the previous section. Let us consider a binary function

$$f: \{0,1\}^n \rightarrow [0,1] \tag{13}$$

This function $f(g)$ maps binary strings $g = (g_1, g_2, ..., g_n) \in \{0,1\}^n$ of length n onto real numbers from the closed interval [0,1]. We look for an optimal solution

$$g_{\text{opt}} = \arg \max_{g \in \{0,1\}^n} f(g) \tag{14}$$

Since the cardinality of the set $\{0,1\}^n$ of solutions is equal to 2^n, a CPU time necessary for solution of the above optimization problem grows exponentially

$$t_{\text{CPU}} \approx 2^n \tag{15}$$

This means that the solution of the binary optimization problem (*14*) belongs to a class of hard numerical NP-complete problems. This is the main reason why the optimization problems (*14*) are solved by the so-called evolutionary algorithms[3,4], which represent very efficient numerical techniques of solving binary optimization problems. The purpose of this subsection is to demonstrate that a metaphor of replicator provides an efficient stochastic optimization algorithm.

Let a chemostat be composed of molecules that are realized by binary strings $g = (g_1, g_2, ..., g_n) \in \{0,1\}^n$. A monomolecular reaction is considered (*cf.* Eq. (*6*))

$$g \xrightarrow{f(g)} g + g' \tag{16}$$

where the formed molecule g' substitutes a randomly selected molecule from the chemostat. A term $f(g)$ assigned to the chemical reaction is interpreted as the probability (rate constant) of the occurrence of reaction (*16*). In evolutionary algorithms a selection pressure in population of solutions (chromosomes) is created by a reproduction process based on chromosome fitness. Chromosomes with a greater fitness have the greater chance to take part in a reproduction process (a measure of quality of chromosomes); on the other hand, chromosomes with a small fitness are rarely used in the reproduction process. This simple manifestation of Darwin's natural selection ensures a gradual evolution of the whole population. In the present approach the mentioned principle of fitness selection of molecules is preserved, but it is now combined with an additional selection pressure due to the constancy of number of molecules in the chemostat. A molecule incoming to the reaction is randomly selected from the chemostat. After evaluation of the quality of the selected molecule it is then stochastically decided whether the reaction is performed or not (see Algorithm 1) and, moreover, the resulting molecule substitutes another randomly selected molecule. Finally, we specify the product g' from the right-hand side of (*16*) as mutation[3] of an incoming molecule g

$$g' = O_{\mathrm{mut}}(g) \tag{17}$$

```
P: =randomly generated chemostat of molecules g;
epoch: = 0;
for epoch: =1 to epoch_max do
begin select randomly a molecule g;
      if random <f(g) then
      begin g': =O_mut(g);
            product g' substitutes in chemostat
            a randomly selected molecule;
      end
end;
```

ALGORITHM 1

A pseudo Pascal implementation of chemostat optimization. The algorithm is initialized by a chemostat composed of randomly generated molecules (binary strings of the fixed length). A copy of a randomly selected molecule g is transformed with a probability equal to its functional value $f(g)$ by a mutation operator onto a new molecule g'. The mutation operator causes an actual change only very rarely. This new mutated molecule substitutes a randomly selected molecule

where O_{mut} is a stochastic mutation operator that changes single bits with a probability P_{mut}. A pseudo Pascal code for the replicator algorithm is presented in Algorithm 1.

As an illustrative example we will study the chemostat approach specified for a simple unimodal function determined over binary strings of length 4. Let us postulate that a chemostat is formed by a multiset composed of binary strings of length 4

$$\{...,(1100),...\} \subset \{0,1\}^4 \tag{18}$$

Each binary vector α is evaluated by a rational number from the closed interval ⟨0,1⟩

$$\mathrm{real}(g) = \frac{1}{2^4 - 1}\mathrm{int}(g) \tag{19}$$

where int(g) is a nonnegative integer assigned to g. A rate constant k assigned to the binary string is specified as follows

$$k(g) = f(\mathrm{real}(g)) = \frac{1}{2}\left(1 - \sin(2\pi \cdot \mathrm{real}(g))\right) \tag{20}$$

with an optimal solution g_{opt} = (1011), where real(g_{opt}) = 11/15 and f(11/15) = 0.9973.

The chemostat is composed of 1000 randomly generated binary strings and the mutation operator O_{mut} is specified by a 1-bit probability P_{mut} = 10^{-5}. The obtained numerical results are displayed in Fig. 3. We can see that those binary strings are spontaneously emerging in the chemostat, which correspond to a suboptimal solution with a rational numerical value closely related to real(g_{opt}) = 0.25. Main results of this section may be summarized as follows: (i) **A metaphor of Eigen's replicators offers an effective stochastic optimization algorithm, where a proof of its convergence to a global solution immediately follows from the existence of a unique asymptotically stable solution with the greatest rate constant**, and (ii) **if the probability of mutations is a very small number, then the obtained results are very similar to those obtained by Eigen's replicator equations with mutations** (see Fig. 2).

Folding of Binary Strings

In an analogy with RNA molecules that are endowed with the so-called folding[26] (secondary structure), we will study a similar property specified also for binary strings (Fig. 4). The folding of a binary string may be specified by a list of matched pairs $i - j$ (for $i < j$) and unmatched singles k

$$\text{fold}(g) = \{i_1 - j_1, i_2 - j_2, \ldots, i_r - j_r; k_1, k_2, \ldots, k_q\} \tag{21a}$$

where the pairs are restricted by the following three conditions:

1. For any pair $i - j$ holds

$$j - i \geq 2 \tag{21b}$$

2. For any two pairs $i - j$ and $k - l$ (restricted by $i \leq k$) holds either

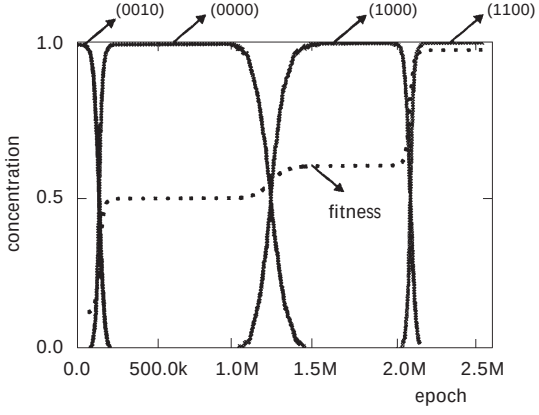

FIG. 3

Plot of frequencies of appearance of some dominant binary strings of length 4. The chemostat was initiated by 1000 randomly generated binary strings. At the beginning of the process, the chemostat was composed entirely of strings (0010). After 2.5×10^6 time steps, the most dominant final solution is g_{fin} = (1100), where real(g_{opt}) = 12/15 and f(12/15) = 0.9755. This final solution is juxtaposed in rational number evaluation to the optimal solution α_{opt} = (1011), where real(g_{opt}) = 11/15 and f(11/15) = 0.9973, but the Hamming distance d of binary strings (1100) and (1011) is great, d = 3 (in theory of GA this effect is called Hamming's cliff). This relatively great size of the Hamming distance is the main reason why the algorithm is unable to achieve the global solution (1011)

$$i = k \Leftrightarrow j = l \tag{21c}$$

$$k < j \Rightarrow i < k < l < j \tag{21d}$$

The first condition (*21b*) means that a minimal length of the so-called *hairpin* in the produced folding is two (see the right-hand part of secondary structure in Fig. 4). The second condition (*21c*) means that each string entry can take part in no more than one matched pair. Finally, the third condition (*21d*) implies that the so-called pseudoknots are forbidden (*i.e.* each folding may be represented by a planar graph; the appearance of pseudoknots in folding could cause its non-planarity). Moreover, the folding of g is defined such that it contains a maximal number of matched pairs; this condition reflects a physical meaning of folding as a most stable secondary structure. The last two conditions (*21c*), (*21d*) are very important for application of dynamic programming technique for construction of folding.

The folding fold(g) of binary string g may be alternatively expressed by a bracket formalism[26] (see Fig. 4). This means that each folding of the binary string may be expressed by a string composed of three symbols {(,),.}, formally fold(g) $\in \{(,),.\}^n$. Of course, actual forms of these strings are strongly restricted such that all three conditions (*21b*)–(*21d*) should be satisfied. The simplest way to ensure the correctness of bracket representation is to postulate that bracket strings are formulae of a simple context-free grammar

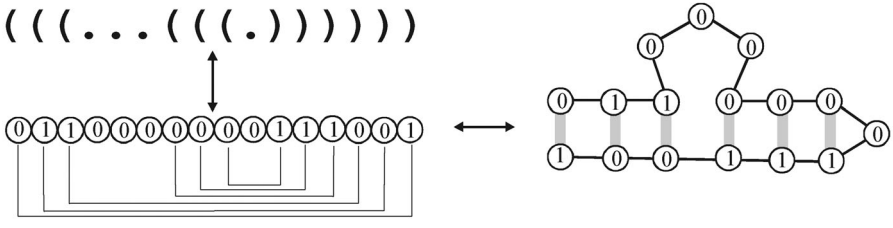

FIG. 4
A binary string (a linear graph with vertices assigned to binary entries that are connected by edges) may be folded into a two-dimensional structure (called the secondary structure) in such a way that complementary binary pairs are matched together. A bracket representation of folding is presented in the upper part of the drawing, where dot symbols correspond to noninteracting string entries and brackets "(" and ")" correspond to a pair of matched complementary entries

$\langle V_N, V_T, S, R\rangle$ where $V_N = \{S\}$ is a set of nonterminal symbols, $V_T = \{(,),.\}$ is a set of terminal symbols, S is a starting nonterminal symbol, and R is a rule set composed of the following three rules

$$S \rightarrow (S)|SS|. \qquad (22)$$

A feasible folding fold(g) which satisfies all three restrictions (*21b*)–(*21d*) is interpreted as a formula composed entirely of terminal symbols that belongs to the language L generated by the grammar (*22*), *i.e.* fold(g) $\in L$. An illustrative example of grammar production looks as follows

$$\begin{aligned} S &\rightarrow \underline{S}S \rightarrow (\underline{S})S \rightarrow ((\underline{S}))S \rightarrow ((\underline{SS}))S \rightarrow ((.\underline{S}))S \\ &\rightarrow ((..))\underline{S} \rightarrow ((..))S\underline{S} \rightarrow ((..))S\underline{S}S \rightarrow ((..))S(\underline{S})S \\ &\rightarrow ((..))S(\underline{SS})S \rightarrow ((..))S(.\underline{S})S \rightarrow ((..))\underline{S}(..)S \\ &\rightarrow ((..)).(..)\underline{S} \rightarrow ((..)).(..). \end{aligned}$$

We use a simple dynamic-programming technique[26] for the construction of folding of binary strings. In general, it can be used as an alternative to backtrack searching methods (these have an exponentially increasing CPU time; see, *e.g.*, the binary optimization problem (*13*), (*14*)), when a respective system may be decomposed into smaller subsystems such that the used evaluation of system is always additive with respect to its subparts (then an exponential time complexity e^n is reduced to a cubic complexity n^3). Let S_{ij} be a substring of bracket representation assigned to a folding between i and j (including) binary entries. These entries are initialized as follows

$$S_{ii} = '.' \quad \text{and} \quad S_{i,i+1} = '..' \qquad (23)$$

The forthcoming entries S_{ij} , for $j - i \geq 2$, are recurrently constructed by

$$S_{i,j} = \max_{i \leq k \leq j-1} S_{ik} \oplus S_{k+1,j} \qquad (24)$$

where $\oplus$ is the symbol of concatenation of "substrings" S_{ik} and $S_{k+1,j}$ which were already constructed in the previous stages of the algorithm (Algorithm 2). Moreover, if S_{ik} ($S_{k+1,j}$) contains in the leftmost (rightmost) position dot

symbol '.' and *i*th and *j*th binary entries are complementary, then the respective dot symbols are substituted by (and) symbols. Symbol "max" in (*24*) means that we select that index *k*, which produces maximal pairing in S_{ij}. The resulting folding in the form of bracket representation is stored at entry S_{1n}.

The main problem in actual implementation of this algorithm for construction of folding of binary strings consists in the fact that although it constructs a folding with maximal matchings, it offers as a result only one result of many possible. Therefore, the algorithm should be considered as a test-bed for development of phenotype of genotypes and also as an integral part of evaluation of genotypes by fitness.

The notion of folding allows us to introduce the triad of fundamental concepts from evolutionary biology, in particular *genotype*, *phenotype*, and *fitness*. The genotype *g* is represented by a binary string of length *n*, $g = (g_1, g_2, ..., g_n) \in \{0,1\}^n$, the phenotype $p(g)$ corresponds to a folding of *g*; formally, it may be expressed by the bracket notation, $p(g) = \text{fold}(g) \in \{(,),.\}^n$. Finally, the fitness, a numerical attribute of *g*, is specified by making use of the respective phenotype. For our forthcoming considerations the fitness will be specified as a similarity between a phenotype $p(g)$ and an ad-hoc *required phenotype* p_{req}

$$\text{fitness}(g) = s(p(g),p_{\text{req}}) \qquad (25)$$

This means that maximal fitness equal to *n* is achieved when a similarity between the respective phenotype $p(g)$ and the required phenotype p_{req} is maximal (*n*) (*i.e.*, $p(g) = p_{\text{req}}$). On the other hand, if $s(p(g),p_{\text{req}}) < n \Leftrightarrow p(g) \neq$

```
for i: =1 to n  do S[i,i]: = '.' ;
for i: =1 to n–1 do S[i,i+1]: = '. .' ;
for d: =2 to n–1 do
for i: =1 to n–d do
begin j: =d+i ;
      S[i,j] : =max{i≤k≤j–1, S[i,k]⊕S[k+1,j]} ;
end;
folding: =S[1,n] ;
```

ALGORITHM 2

A dynamic-programming scheme for the calculation of folding of a binary string of length *n*. The resulting folding is coded in the bracket representation and is stored at the string entry $S_{1,n}$

p_{req}, then the fitness is smaller than n; in a limit case, when the foldings are fully dissimilar (*i.e.* $s(p(g),p_{req}) = 0$), the fitness of g is vanishing, fitness(g) = 0. The notion of similarity will be specified in the forthcoming section. Table I contains illustrative results, where genotypes represented by all possible binary strings of length $n = 7$ are evaluated by phenotype foldings and fitness. We can see that many different binary strings are evaluated by the same folding, and similarly, many different foldings are evaluated by the same fitness. This simple observation immediately implies that mappings of genotypes onto phenotypes and phenotypes onto fitness are of the many-to-one type, *i.e.* there exist a huge redundancy in both mappings (Fig. 5). A single binary string could be theoretically evaluated by more foldings than one, but the mapping produced by an application of the Algorithm 2 is the first one, which was found with a particular maximal number of matched pairs.

An interrelationship between the elements of this triad "genotype-phenotype-fitness" is formally expressed as a sequence of two mappings (see Fig. 5)

$$G \xrightarrow{p} P \xrightarrow{\text{fitness}} [0,\infty) \qquad (26)$$

This means that the basic entity is the genotype: it is initially mapped onto the phenotype, and then the phenotype is mapped onto the fitness. An abbreviated form of this composite mapping is as follows

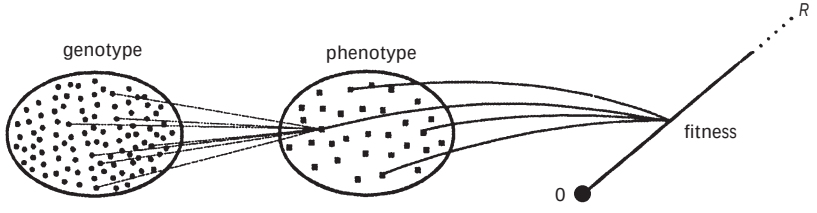

FIG. 5

Schematic outline of the composite mapping (*26*). Both respective mappings are of the many-to-one type, *i.e.* many gene strings are mapped onto one phenotype string, and similarly, many phenotype strings are mapped onto one value of fitness. This means that there exists a huge redundancy of the genotype coding, many different genes (strings) may be evaluated by one value of fitness. This property of the huge redundancy of genotype coding is of considerable importance for the existence of neutral stases in the Darwinian evolutionary theory

TABLE I
All possible genotypes represented by binary string (for n = 7), the respective phenotypes and fitness

No.	Genotype	Phenotype	Fitness[a]
1	(0000000) (1111111)	. . . . .	1
2	(0000001) (0000100) (1111011) (1111110)	. . . . (.)	1
3	(0000010) (1111101)	. . . (.) .	0
	(0000011) (0001001) (0001101) (0010010) (0010110)		
4	(0011000) (0011100) (0100011) (0100111) (0101001)	. . ((.))	0
	(0110010) (0110110) (0111000) (0111100) (1000011)		
	(1000111) (1001001) (1001101) (1010110) (1011000)		
	(1011100) (1100011) (1100111) (1101001) (1101101)		
	(1110010) (1110110) (1111100)		
5	(0000101) (0010001) (0101100) (0101110) (0110000)	. ((.) .)	3
	(0111010) (1000101) (1001111) (1010001) (1010011)		
	(1101110) (1111010)		
6	(0000110) (0011010) (1100101) (1111001)	. ((.)) .	5
7	(0000111) (0001111) (0010011) (0011011) (0100101)	(((.)))	7
	(0101101) (0110001) (0111001) (1000110) (1001110)		
	(1010010) (1011010) (1100100) (1101100) (1110000)		
	(1111000)		
8	(0001000) (1110111)	. . . (. .)	0
9	(0001010) (1110101)	((.) .) .	0
10	(0001011) (0001100) (0001110) (0011001) (0011110)	. (.) (.)	0
	(0100001) (0100100) (0100110) (0110011) (0110100)		
	(1001011) (1001100) (1011001) (1011011) (1011110)		
	(1100001) (1100110) (1110001) (1110011) (1110100)		
11	(0010000) (1101111)	. . (. . .)	1
12	(0010100) (0111011) (0111110) (1000001) (1000100)	(.) . (.)	1
	(1101011)		
13	(0010101) (0101010) (1010101) (1101010)	. ((. .))	1
14	(0010111) (0011101) (0100010) (0101000) (1010111)	. (. (.))	0
	(1011101) (1100010) (1101000)		
15	(0011111) (1100000)	((.) . .)	2
16	(0100000) (1011111)	. (. . . .)	1
17	(0101011) (1010100)	(. .) (.)	0
18	(0101111) (1010000)	(. (.) .)	5
19	(0110101) (0110111) (1001000) (1001010)	(.) (. .)	0
20	(0111101) (1000010)	(. . (.))	2
21	(0111111) (1000000)	(. . . . .)	3

[a] Fitness is calculated as a graph-theory similarity between the respective phenotype and the so-called required phenotype represented by p_{req} = (((.))), see Eqs (*32a*)–(*32d*)

$$G \xrightarrow{f=\text{fitness}\circ p} [0,\infty) \tag{27}$$

This new mapping immediately maps the genotype strings onto fitness without the necessity to consider explicitly an intermediate called the phenotype. In this connection one may ask why there is worthwhile to introduce the phenotype as a mediator between the genotype and the fitness. Of course, such a question is fully acceptable from the pure mathematical point of view, but it must be noted that the concept of phenotype is a very effective and fruitful heuristic for interpretation of the Darwinian evolutionary theory. In particular, a given form of phenotype is usually considered as an evolutionary goal, and therefore we may say that the Darwinian evolution is represented mainly as a sequence of phenotypes, which are progressively closer and closer to the evolutionary phenotype goal.

Wright[1] in 1931 introduced one of the most fundamental concepts of the Darwinian evolution called the *fitness surface* (Fig. 6). Moreover, by making use of this concept he characterized the Darwinian evolution as an optimization process, where the evolved population seeks a global maximum (or another solution closely related to this global one)

$$g_{\text{opt}} = \arg\max f(g) \tag{28}$$

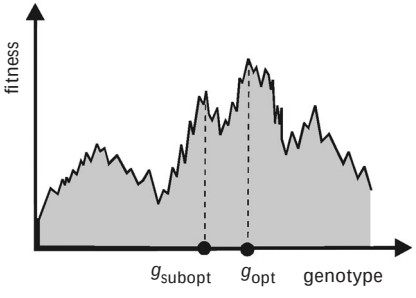

FIG. 6

Fitness landscape (or more precisely, fitness surface) was initially introduced into theory of the Darwinian evolution in 1931 by Wright[1], who characterized the Darwinian process as an optimization process over the fitness landscape specified by a composite mapping f (*27*). The resulting population genotype corresponds to a point – optimal genotype g_{opt} – with a maximal fitness. Loosely speaking, the Darwinian evolution should have tools to find a way from a suboptimal solution g_{subopt} to optimal solution g_{opt}

This complex optimization combinatorial problem will be solved in the forthcoming part of this paper by methods of artificial chemistry based on the metaphor of Eigen's replicators. It will be demonstrated that the "chemostat" formalism offers an effective tool for optimization of problems of the form (*28*), *i.e.* "replicator" methods of artificial chemistry are very well suitable to mimic the molecular Darwinian evolution.

To summarize the main results of this section, **we have observed that both mappings of genotype onto phenotype as well as phenotype onto fitness are of strongly stochastic nature with huge redundancy (both mappings are of the many-to-one type). General properties of these mappings may be specified only by making use of computer-simulation tools**, *i.e.* **instead of exact "deterministic" description of properties of mappings, we generate histograms of appearances of their different entities and then we deduce from them some general properties of statistical validity.**

A Set-Theory Formalism

A basic concept of the present formal theory is *genotype* g represented by a binary string of length n, $g = (g_1 g_2 ... g_n) \in G = \{0,1\}^n$. For pairs of genotypes $g = (g_1 g_2 ... g_n) \in G$, $g = (g'_1 g'_2 ... g'_n) \in G$, an L_1 (Hamming) *metrics* (*distance*) is determined

$$d(g, g') = \sum_{i=1}^{n} |g_i - g'_i| \qquad (29)$$

A *mutation* of the genotype g onto another genotype g', or formally $g' = O_{\mathrm{mut}}(g)$, is called neutral if both genotypes have the same fitness. In our forthcoming considerations this type of neutrality will be classified as a *weak neutrality*

$$g \xrightarrow{\text{weak}} g' = O_{\mathrm{mut}}(g) \Rightarrow \text{fitness}(g) = \text{fitness}(g') \qquad (30a)$$

On the other hand, if mutation of a genotype does not change a corresponding folding (phenotype), then we speak about a *strong neutrality*

$$g \xrightarrow{\text{strong}} g' = O_{\mathrm{mut}}(g) \Rightarrow \text{folding}(g) = \text{folding}(g') \qquad (30b)$$

Since the main driving force in the molecular Darwinian evolution is fitness of genotypes, the appearance of mutations that are weakly neutral, will be considered as one of hidden reasons for a continuation of evolution. Furthermore, if the appearance of neutral mutations is restricted to strongly neutral mutations, then, loosely speaking, their appearance is indeed a hidden stage of evolution, they could not be recorded by an external observer.

A *genotype graph* $\hat{G} = (V,E)$ has a vertex set V composed of 2^n vertices identified with binary strings of length n, $V = G$, and an edge set E is composed of edges that connect two vertices – binary strings with Hamming distance equal to one

$$E(\hat{G}) = \{e = (g,g'); d(g,g') = 1\} \tag{31}$$

The genotype graph $\hat{G}$ may be formally identified as a hypercube composed of 2^n vertices (Fig. 7).

Formally, each genotype $g = (g_1 g_2 \ldots g_n) \in G$ is evaluated by its folding, it will be called the *phenotype*, $p(g) = \text{fold}(g) \in \{(,),.\}^n$. As was already mentioned in the previous section, the folding (*i.e.* phenotype) is represented by a formula belonging to language L (see Eqs (*22a*, *22b*)). All possible phenotypes form a set P, we put $P \subset \{(,),.\}$.

For the set P we may define a similarity (metrics) as follows[27]: Let $p(g) = (p_1 p_2 \ldots p_n) \in L$ and $p(g') = (p'_1 p'_2 \ldots p'_n) \in L$ be two phenotypes, their set-theory specification is (see Eq. (*21a*))

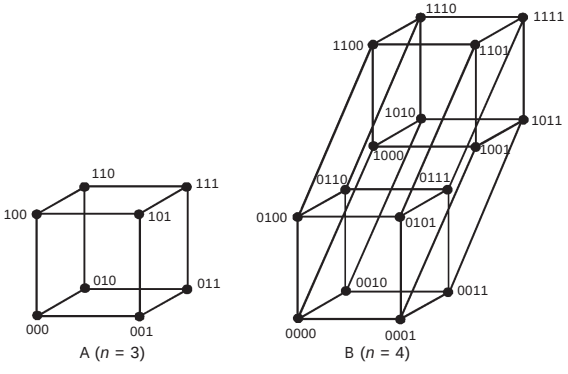

FIG. 7

Two hypercubes for $n = 3$ (A) and $n = 4$ (B) that visualize the notion of the genotype graph $\hat{G}$. Vertices correspond to binary strings, two vertices are connected by an edge if their Hamming distance is equal to one

$$p = p_{\text{pair}} \cup p_{\text{single}}, \quad p' = p'_{\text{pair}} \cup p'_{\text{single}} \tag{32a}$$

$$p_{\text{pair}} = \{[i_1, j_1], \ldots, [i_r, j_r]\}, \quad p_{\text{single}} = \{k_1, \ldots, k_q\} \tag{32b}$$

$$p'_{\text{pair}} = \{[i'_1, j'_1], \ldots, [i'_r, j'_r]\}, \quad p'_{\text{single}} = \{k'_1, \ldots, k'_q\} \tag{32c}$$

A concept of *similarity* is determined graph-theoretically[27] by maximal common subgraph of p and p' determined by intersections of subsets for pair matching and subsets for unmatched vertices, respectively (Fig. 8)

$$s(p, p') = \left(2|\, p_{\text{pair}} \cap p'_{\text{pair}}\,| + |\, p_{\text{single}} \cap p'_{\text{single}}\,|\right) \tag{32d}$$

Let $p_{\text{req}} = \left(p_1^{(\text{req})} p_2^{(\text{req})} \ldots p_n^{(\text{req})}\right) \in P$ be a required phenotype, then a *fitness* of the genotype g may be defined as follows (see Eq. (*25*))

$$\text{fitness}(g) = s(p(g), p_{\text{req}}) \tag{33}$$

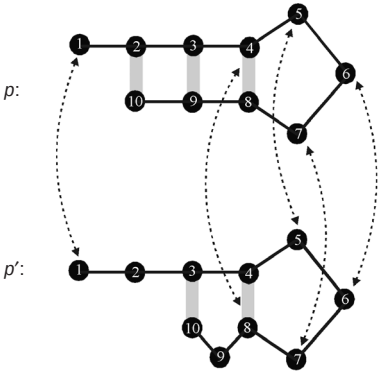

FIG. 8
An illustrative example of similarity calculation between two phenotypes p and p', $p_{\text{pair}} \cap p'_{\text{pair}} = \{[4,8]\}$, $p_{\text{single}} \cap p'_{\text{single}} = \{1,5,6,7\}$, $s(p,p') = 2 \times 1 + 4 = 6$

This means that each genotype is evaluated by a fitness represented by a nonnegative integer; a largest fitness n is achieved when the phenotype $p(g)$ is identical to the required phenotype p_{req}.

An evolutionary goal is to look for such an optimal genotype g_{opt} such that the fitness function (*33*) achieves maximal value

$$g_{\text{opt}} = \arg \max_{g \in G} \text{fitness}(g) \tag{34}$$

For larger values on n, this optimization problem represents an extremely complicated computational problem, which will be approached in the forthcoming part of this paper by a chemostat technique specified already above in the previous section.

The vertex set V may be decomposed into disjoint subsets of vertices that are evaluated by the same fitness

$$V = \bigcup_f V_f \tag{35}$$

where $V_f = \{g; \text{fitness}(g) = f\} \subset V$. Applying this decomposition to edge set, we get its disjoint decomposition

$$E = \bigcup_{f,f'(f \leq f')} E_{f,f'} \tag{36}$$

where $E_{f,f'}$ is an edge subset composed of edges with one vertex from V_f and the other vertex from $V_{f'}$, $E_{f,f'} = \{e = (g, g'); g \in V_f \wedge g' \in V_{f'}\}$. A neutral subgraph with specified fitness f is determined by $\hat{G}_f = (V_f, E_{f,f'})$. Formally, the genotype graph $\hat{G}$ may be expressed in a "reduced" form called the fitness graph $\hat{F}$, where a vertex set $V(\hat{F})$ is composed of all neutral subgraphs, $V(\hat{F}) = \{\hat{G}_f, \hat{G}_{f'}, ...\}$, and an edge $e = (\hat{G}_f, \hat{G}_{f'}) \in E(\hat{F})$ exists if both neutral subgraphs $\hat{G}_f$ and $\hat{G}_{f'}$ have at least one common edge in the genotype graph $\hat{G}$ (Figs 9 and 10)

$$e = \left(\hat{G}_f, \hat{G}_{f'}\right) \in E\left(\hat{F}\right) \Rightarrow \exists\left(g \in V_f\right) \exists\left(g' \in V_{f'}\right) : (g, g') \in E_{f,f'} \tag{37}$$

In a schematic condensed representation of genotype graphs as fitness graphs with vertices assigned to single subgraphs with given fitness (*e.g.* see Fig. 10), each edge may be evaluated by the probability of transition from one neutral subgraph to the other neutral subgraph. Single edges in the condensed genotype graph may be evaluated by probabilities $\mathrm{pr}(f \to f')$ of stochastic transition from one neutral graph $\hat{G}_f$ to another neutral graph $\hat{G}_{f'}$ (see Fig. 10) (another alternative way of definition of this probability is given in ref.[23])

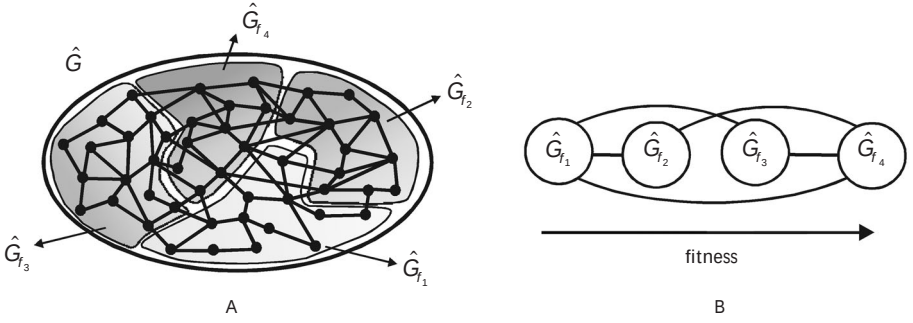

FIG. 9

A An illustrative representation of the genotype graph $\hat{G}$ composed of vertices (genotypes – binary strings) and edges that connect them (if the respective Hamming distance is equal to one). A vertex set $V(\hat{G})$ is decomposed into four disjoint vertex subsets that are evaluated by the same fitness. These subsets induce four neutral subgraphs $\hat{G}_f$. B Fitness graph $\hat{F}$ may be understood as a condensed form of the genotype graph $\hat{G}$, its vertices are neutral subgraphs that are connected by an edge if both its vertices – subgraphs are in $\hat{G}$ connected by an edge

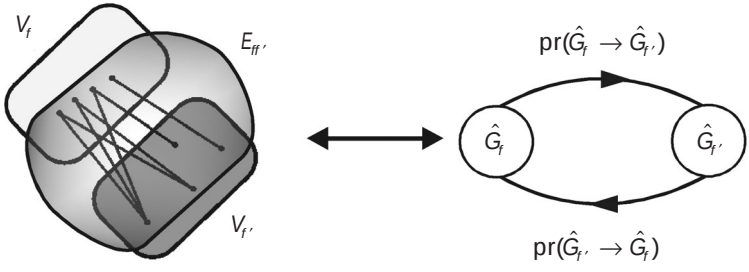

FIG. 10

Set-theory visualization of a transition probability from a neutral subgraph $\hat{G}_f$ to another neutral subgraph $\hat{G}_{f'}$. Since the probabilities are not symmetric, $\mathrm{pr}(\hat{G}_f \to \hat{G}_{f'}) \neq \mathrm{pr}(\hat{G}_{f'} \to \hat{G}_f)$, an edge in the condensed genotype graph should be substituted by two edges that are oriented in opposite ways, and each of them is evaluated by the respective probability

$$\mathrm{pr}(f \rightarrow f') = \frac{|E_{f,f'}|}{\omega_f} \tag{38a}$$

$$\omega_f = \sum_{f''} E_{f,f''} \tag{38b}$$

It is simple to prove that the above probabilities satisfy

$$0 \leq \mathrm{pr}(f \rightarrow f') \leq 1 \tag{39a}$$

$$\sum_{f'} \mathrm{pr}(f \rightarrow f') = 1 \tag{39b}$$

Especially, an expression $\mathrm{pr}(f \rightarrow f)$ specifies a probability of a neutral transition, when the fitness f remains unchanged.

An alternative decomposition of the genotype graph $\hat{G}$ into subgraphs with specified phenotypes may be introduced by decomposition of the vertex set V into disjoint subsets of vertices with the same folding – phenotype

$$V = \bigcup_p V_p \tag{40}$$

where $V_p = \{g; \mathrm{folding}(g) = p\} \subset V$. Applying this decomposition, we get an alternative decomposition of the edge set E

$$E = \bigcup_{p,p'(p \leq p')} E_{p,p'} \tag{41}$$

where $E_{p,p'}$ is an edge subset composed of edges with one vertex from V_p and other vertex from $V_{p'}$, $E_{p,p'} = \{e = (g,g'); g \in V_p \wedge g' \in V_{p'}\}$. A neutral subgraph with specified folding p is determined by $\hat{G}_p = (V_p, E_{p,p'})$. Formally, the genotype graph $\hat{G}$ may be expressed in a "reduced" form called the phenotype graph $\hat{P}$, where a vertex set $V(\hat{P})$ is composed of all neutral subgraphs, $V(\hat{P}) = \{\hat{G}_p, \hat{G}_{p'}, ...\}$, and an edge $e = (\hat{G}_p, \hat{G}_{p'}) \in E(\hat{P})$ exists if both neutral subgraphs $\hat{G}_p$ and $\hat{G}_{p'}$ have at least one common edge in the genotype graph $\hat{G}$ (Fig. 11)

$$e = \left(\hat{G}_p, \hat{G}_{p'}\right) \in E\left(\hat{P}\right) \Rightarrow \exists\left(g \in V_p\right) \exists\left(g' \in V_{p'}\right) : \left(g, g'\right) \in E_{p,p'} \qquad (42)$$

In a similar way as above we introduce probabilities of transitions from one phenotype neutral subgraph $\hat{G}_p$ to another neutral phenotype subgraph $\hat{G}_{p'}$

$$\mathrm{pr}(p \to p') = \frac{|E_{p,p'}|}{\omega_p} \qquad (43a)$$

$$\omega_p = \sum_{p''} E_{p,p''} \qquad (43b)$$

In the remaining part of this section we turn our attention to relationships between probabilities $\mathrm{pr}(f \to f')$ and $\mathrm{pr}(p \to p')$. After simple algebra, we get

$$\mathrm{pr}(f \to f') = \sum_{p \in f} \frac{\omega_p}{\omega_f} \sum_{p' \in f'} \mathrm{pr}(p \to p') \qquad (44a)$$

where, by making use of (*39b*) and (*43b*), we may prove that

$$\omega_f = \sum_{p \in f} \omega_p \qquad (44b)$$

From this property immediately follows that ratios ω_p/ω_f satisfy $0 \le \omega_p/\omega_f \le 1$ and $\sum_{p \in f} \omega_p / \omega_f = 1$. This means that (*44a*) can be understood as a convex

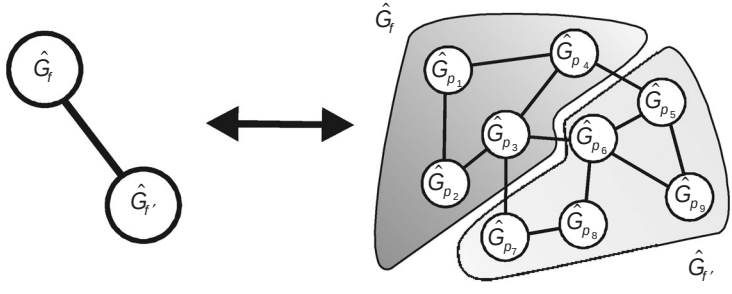

FIG. 11
Decomposition of two adjacent neutral subgraphs $\hat{G}_f$ and $\hat{G}_{f'}$ into phenotype subgraphs $\hat{G}_p$'s

combination of probabilities $\mathrm{pr}(p \to p')$. Consequently, a value of $\mathrm{pr}(f \to f')$ is bounded from above by a maximal value of probabilities $\mathrm{pr}(p \to p')$

$$\mathrm{pr}(f \to f') \leq \max_{p' \in f'} \{\mathrm{pr}(p \to p')\} \tag{45}$$

Let us consider two different genotypes g_{ini} and g_{fin}; we say that the genotype g_{fin} is *achievable* from the genotype g_{ini} if there exists in the genotype graph $\hat{G}$ a *trajectory* $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$ that connects the genotypes

$$T(g_{\mathrm{ini}},g_{\mathrm{fin}}) : g_{\mathrm{ini}} = g_1 \to g_2 \to \ldots \to g_{k-1} \to g_k = g_{\mathrm{fin}} \tag{46a}$$

such that fitness of all participated genotypes creates a nondecreasing sequence

$$\mathrm{fitness}(g_1) \leq \mathrm{fitness}(g_2) \leq \ldots \leq \mathrm{fitness}(g_{k-1}) \leq \mathrm{fitness}(g_k) \tag{46b}$$

Each trajectory $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$ can be evaluated by probability of its performance defined as follows

$$\mathrm{pr}(g_{\mathrm{ini}} \to g_{\mathrm{fin}}) = \prod_{i=1}^{k-1} \mathrm{pr}(\mathrm{fitness}(g_k) \to \mathrm{fitness}(g_{k+1})) \tag{47}$$

where the probability $\mathrm{pr}(f \to f')$ is determined by Eq. (*38a*). Of course, there may exist many different trajectories $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$; then the resulting probability $\mathrm{pr}(g_{\mathrm{ini}} \to g_{\mathrm{fin}})$ is equal to the maximal value of the probabilities assigned to these different trajectories.

A particular trajectory $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$ can be interpreted also as a sequence of neutral subgraphs $\hat{G}_f = (V_f, E_{f,f'})$

$$T(f_{\mathrm{ini}}, f_{\mathrm{fin}}) : \hat{G}_{f_{\mathrm{ini}}} \to \hat{G}_{f_2} \to \ldots \to \hat{G}_{f_{k-1}} \to \hat{G}_{f_{\mathrm{fin}}} \tag{48}$$

where $f_{\mathrm{ini}} = \mathrm{fitness}(g_{\mathrm{ini}})$ and $f_{\mathrm{fin}} = \mathrm{fitness}(g_{\mathrm{fin}})$. Genotype g_i $(1 \leq i \leq k)$ belongs to the neutral subgraph $\hat{G}_{f_i}$. We have to emphasize once again, that a se-

quence of used fitnesses is nondecreasing, $f_1 \le f_2 \le \ldots \le f_k$. If a pair of juxtaposed neutral subgraphs exists in Eq. (*48*) such that their fitness is equal, $f_i = f_{i+1}$, but respective phenotypes are different, then we say that the transition $g_i \to g_{i+1}$ is weakly neutral.

An alternative look at the trajectory $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$ may be provided by using the concept of neutral subgraph with specified phenotype $\hat{G}_p = (V_p, E_{p,p})$. We get

$$T(p_{\mathrm{ini}}, p_{\mathrm{fin}}) : \hat{G}_{p_{\mathrm{ini}}} \to \hat{G}_{p_2} \to \ldots \to \hat{G}_{p_{k-1}} \to \hat{G}_{p_{\mathrm{fin}}} \qquad (49)$$

where $p_{\mathrm{ini}} = \mathrm{folding}(g_{\mathrm{ini}})$ and $p_{\mathrm{fin}} = p_{\mathrm{reg}} = \mathrm{folding}(g_{\mathrm{fin}})$, *i.e.* the final genotype is unknown, we require only its phenotype p_{reg}. A transient genotype $\mathbf{g}_i$ ($1 \le i \le k$) belongs to a neutral subgraph $\hat{G}_{p_i}$, *i.e.* $p_i = \mathrm{folding}(g_i)$. A single transition $g_i \to g_{i+1}$ is called *strongly neutral* if both genotypes belong to the same neutral subgraph with specified phenotype, $\hat{G}_{p_i} = \hat{G}_{p_{i+1}}$.

The trajectory $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$ and its probability $\mathrm{pr}(g_{\mathrm{ini}} \to g_{\mathrm{fin}})$, defined by (*46a*) and (*47*), respectively, can be evaluated by the entity called the entropy as follows

$$S(g_{\mathrm{ini}} \to g_{\mathrm{fin}}) = -\ln \mathrm{pr}(g_{\mathrm{ini}} \to g_{\mathrm{fin}}) = -\sum_{i=1}^{k-1} \mathrm{pr}(f_k \to f_{k+1}) \ln[\mathrm{pr}(f_k \to f_{k+1})] \qquad (50)$$

where $f_k = \mathrm{fitness}(g_k)$. We expect that all probabilities $0 < \mathrm{pr}(f_k \to f_{k+1}) << 1$. This means that the most feasible trajectory is that one with maximal entropy. Unfortunately, the situation is not such simple as it seems at first sight, in particular, a loop of repeating genotypes with the same fitness can exist. In order to avoid loop problems in trajectories, we have to introduce the additional constraint that each contributing genotype g_k appears just once in the respective path $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$. The concept of trajectory entropy $S(g_{\mathrm{ini}} \to g_{\mathrm{fin}})$ can be used as a guiding principle for construction of trajectory $T(g_{\mathrm{ini}},g_{\mathrm{fin}})$. Formally, the Darwinian evolution may be considered as a metaphor how to effectively solve this complex combinatorial problem of construction of the required trajectory; within evolutionary algorithms, we have a huge class of methods that are capable of solving the problem.

RESULTS AND DISCUSSION

Chemostat Simulation of Molecular Darwinian Evolution

The chemostat approach outlined in the first part of this communication will be used as an algorithmic framework for simulation of the molecular Darwinian evolution. An initial composition of chemostat is made by identical strings of length 39

$$g_{\mathrm{ini}} = \mathbf{111111111111111000000000000000000000000000000} \quad (51a)$$

The required phenotype is specified by

$$p_{\mathrm{req}} = (((((((.)))))))(((((((.)))))))(((((((.))))))) \quad (51b)$$

Its secondary propeller-like structure is composed of three shafts (see secondary structure Q in Fig. 13). This phenotype will serve as a goal phenotype, all fitness calculations of binary strings being realized with respect to the required phenotype. This means that we expect a genotype with folding fully specified by the required phenotype p_{req} to emerge in the chemostat. The mutation operator O_{mut} (*17*) is specified by a probability of 1-bit mutation $P_{\mathrm{mut}} = 0.0001$. The used probability of 1-bit mutation is sufficiently small to ensure that the produced trajectory is composed of adjacent strings (with unit Hamming distance) in the respective genotype graph $\hat{G}$ (see Fig. 13). Further important parameter, the size of chemostat (number of strings), is equal to 500.

The composition of chemostat is characterized by two different types of entropies that are determined as follows: Let $w(x)$ be the probability that the chemostat string is evaluated by entity x (fitness or genotype). Then (*cf.* ref.[17])

$$S_x = -\sum_x w(x) \ln[w(x)] \quad (52)$$

These entropic parameters are very sensitive to composition changes of the chemostat. For a homogenous chemostat (composed of the same strings), both entropies are vanishing. A *fitness entropy* S_{fitness} may serve as a proper

indicator of intermediate transient states where a new genotype with a higher fitness substitutes an older genotype with lower fitness. This moment of transition can be detected from a fast temporary increase of the fitness entropy S_{fitness} from its usual zero value. The genotype entropy is capable of detecting such transient stages where fitness remains unchanged, but the genotype composition of chemostat is temporarily changed due to the existence of the neutral mutation. This means that the genotype entropy is capable of detecting very rare events when fitness remains invariant but there appear temporarily different neutral genotypes. Both entropies have nonnegative values; it is possible to assess their maximal values as follows. At most time stages the chemostat composition is homogeneous, then $S_x = 0$. A deviation from this homogeneity is usually realized as a two-component system (*e.g.* two different types of the genotype or phenotype, with respective concentrations α and $1 - \alpha$). Then the entropy is specified by $S_x = -\alpha \ln \alpha - (1 - \alpha) \ln (1 - \alpha)$, for $0 < \alpha < 1$. Its maximal value is achieved for $\alpha = 1/2$, $S_x(1/2) = \ln 2 \approx 0.7$. This usually means that great positive values of the respective entropy indicate that the chemostat is composed of two components with roughly same concentrations.

Numerical results of our computer simulation of the molecular Darwinian evolution are displayed in Figs 12 and 13. We can see that the produced plot of the chemostat mean fitness (diagram D in Fig. 12) is a nondecreasing step function, with a few relatively long neutral stases. Diagrams A and B correspond to a genotype and fitness entropy, respectively, where these plots unambiguously indicate that at the end of the first neutral stasis (about 300 000 epochs) a considerable deviation from genotype homogeneity exists, *i.e.* in this evolutionary stage the chemostat is likely composed of two different genotypes with the same fitness. This deviation from the chemostat homogeneity can be considered as a necessary preliminary stage before an evolutionary transition to a next stage with higher fitness. The same situations can be found also in the previous stages of evolution, in particular for phenotype transitions A→B and E→F. Diagram C represents concentration profiles of different genotypes with respective fitness; we see these profiles to nicely agree with the Eigen's theory prediction of their form (see Fig. 2, diagram B). In particular, they form a sequence of well separated profiles. The dotted vertical auxiliary lines mark evolutionary transitions well indicated by entropies. Figure 13 shows a sequence of phenotypes that appeared for single evolutionary stages (see Fig. 12). We got 17 different genotypes that are evaluated by different phenotypes (sec-

ondary folding structures). Their similarity to the required phenotype (*51b*) specifies directly their fitness.

We can see that neutral mutation plays an important role for the molecular Darwinian evolution to be able to escape local maxima on fitness hypersurface (Figs 14 and 15). In many cases, when a respective string g is not adjacent in the genotype graph $\hat{G}$ to another string g' with higher fitness (*i.e.* a transition $g \to g'$ causes an evolutionary jump), it is useful to stochastically look for another string $\tilde{g}$ in the neighborhood of g which has

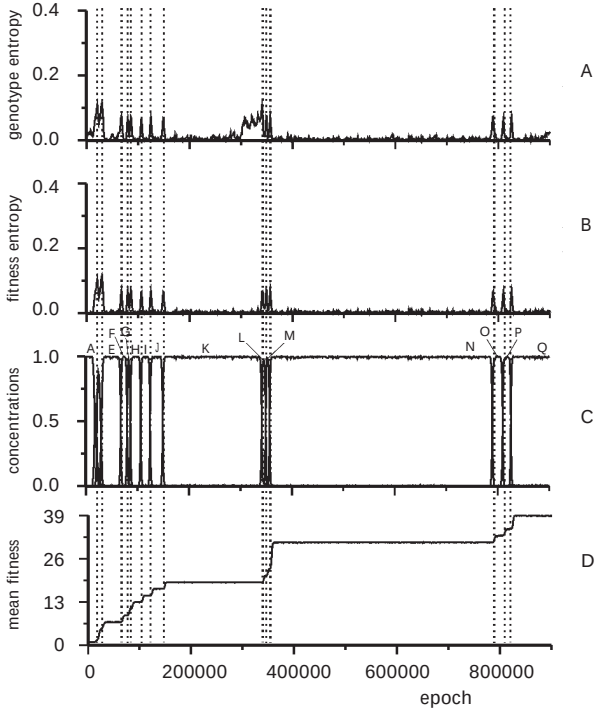

FIG. 12

Plots of A genotype entropy, B fitness entropy, C concentrations of strings with respective fitness, and D mean fitness. Mean fitness is a nondecreasing plot with two relatively long neutral stases. Its stairs are well indicated by fitness and genotype entropies. The plot of genotype entropy (A) at the end of the first neutral stasis indicates a substantial increase in genotype entropy whereas the fitness entropy is simultaneously almost vanishing. From this observation we may conclude that there exists huge appearance of "neutral mutants" that have the capability of producing by 1-bit mutations a new string with a higher fitness. Plot C displays relative concentrations of strings with given fitness. Each concentration profile is identified with labels of phenotypes in Fig. 13

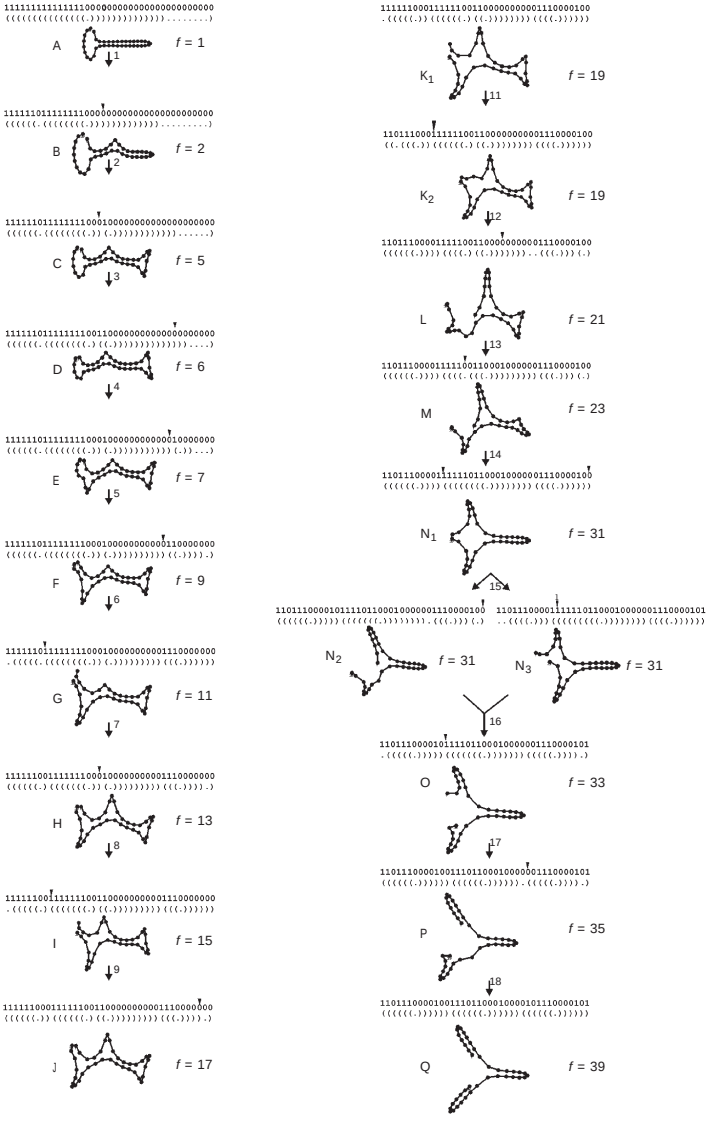

FIG. 13

A trajectory $T(g_{ini},p_{req})$, where the initial genotype is specified by (*48*) and the required phenotype is outlined by diagram Q. The length of trajectory is 18, in all transient states the used mutations are of the 1-bit type (between respective strings there is unit Hamming distance). There exist two cases of neutral mutations, in particular transitions 11 and 15. These transitions are different, the former corresponding to a sequence of neutral mutations, whereas the latter to a "parallel" appearance of two neutral mutations (*i.e.* the trajectory is split into two paths which are connected in the next step)

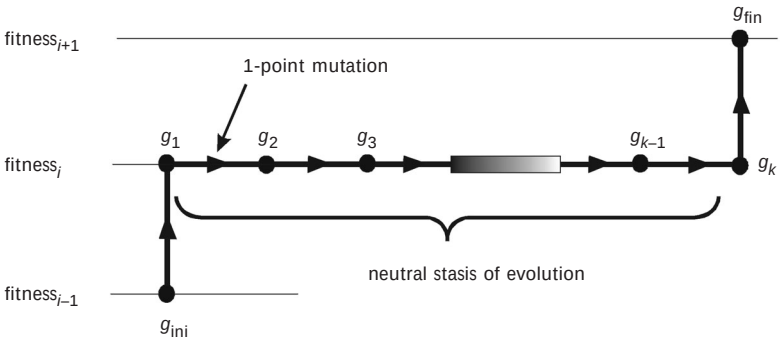

FIG. 14

Schematic outline of the neutral stasis between "jumps" from the neutral graph with *fitness*$_i$ into another neutral graph with *fitness*$_{i+1}$. At the neutral stage shown in details, hill-climbing forms a sequence of actual solutions within a cluster. The first solution in this sequence corresponds to a result from the previous "jump" and the last solution enables a "jump" to the forthcoming neutral graph *fitness*$_{i+1}$. Then the neutral stasis is repeated until a proper solution is found, which can be used for the next "jump" to further neutral graph

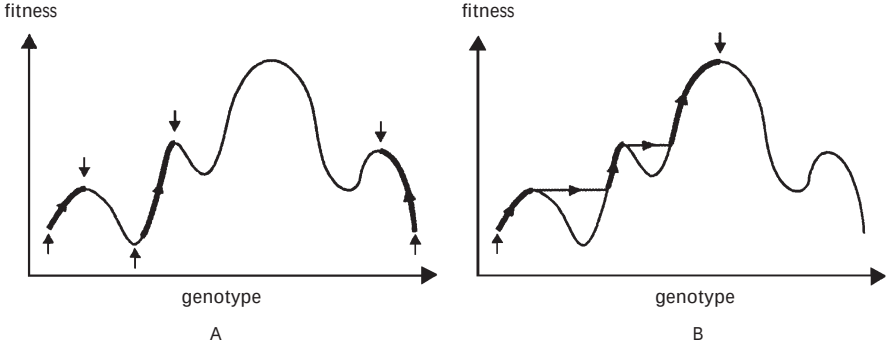

FIG. 15

A Schematic illustrative plot of the molecular Darwinian evolution when neutral mutations do not exist. An evolutionary adaptation ends at the nearest local maximum (in this case we may say that the Darwinian evolution is a local optimizer, population fitness is "scrambled" on fitness hypersurface to nearest maxima). B On the other hand, when neutral mutations are considered (fitness hypersurface is composed of many "plateaus" with constant value of fitness), a local character of Darwinian adaptation may be abridged by neutral "plateaus", which allow to overcome valleys with lower fitness. This means that the molecular Darwinian evolution has a good chance to achieve a global maximum on fitness hypersurface in the course of adaptation process. (After Schuster[23])

the same fitness, but for which an evolutionary jump $\tilde{g} \rightarrow g'$ with an increase of fitness, fitness(g') > fitness($\tilde{g}$), already exists.

Robustness of Phenotype

Recently, the problem of robustness of an evolutionarily emerged phenotype[26,28–32] is considered by theoreticians as a very important aspect of evolutionary theory. In these concepts it is postulated that the evolutionary goal is not only a best fitted phenotype (closely related to the required one) but it should simultaneously manifest a minimal vulnerability on mutations. That is, the resulting phenotype would not be very dependent on 1-bit mutations of the particular genotype. It is expected that most of the mutations are neutral or nearly neutral (with small impact on the shape of phenotype).

First of all, we have to specify this concept of robustness, which will play a fundamental role in our forthcoming considerations. Let us consider a genotype g; its phenotype and assigned fitness are denoted by $p(g)$ and fitness(g), respectively. *Robustness* of g is determined as follows

$$\text{robust}(g) = -\sum_{g' \in U(g)} \gamma^{s(p(g),p(g'))} \qquad (53)$$

where $0 < \gamma < 1$ is a robustness parameter (we used $\gamma = 0.85$) and $U(g)$ is a neighborhood of g composed of all its 1-bit mutations. We see that the robustness robust(g) is most sensitive to those contributions that are caused by mutated genotypes g', which phenotypes have a very small similarity to the phenotype of the original genotype g, *i.e.* for $s(p(g),p(g')) \approx 0$ closely related to zero. Summarizing, we can say that the evolutionary goal is specified by two conditions: (1) fitness(g) = max and (2) robust(g) = max. Of course, this two-criterion evolutionary objective function can be reduced to one-criterion in an obvious way such that the fitness is slightly increased by the robustness, *i.e.* fitness$'(g)$ = fitness(g) + $\omega \times$ robust(g).

In order to understand the role of robustness in the molecular Darwinian evolution we repeated our computer simulations such that we have included also the influence of the robustness on the probability of replication of genotypes in the chemostat. More explicitly, in our standard computational experiments the probability of replication was slightly decreased by the respective robustness. Our results are summarized in Fig. 16. We can see that inclusion of robustness in the probability of replication has a dramatic

impact on the genotype representation of the required phenotype. In particular, if the robustness is included and the fitness is slightly increased by it, then the final genotype (with folding equal to that required) is endowed with substantially higher robustness than that one with ignored robustness.

Recently, the concept of phenotype robustness has been extensively studied by Ancel and Fontana[28,29]. They used an interesting analogy between the robustness of RNA folding and the Baldwin effect[33–35]. They showed that the modularity of folding (the whole secondary structure is divided into smaller parts called *modules* such that 1-bit mutations cause only shape

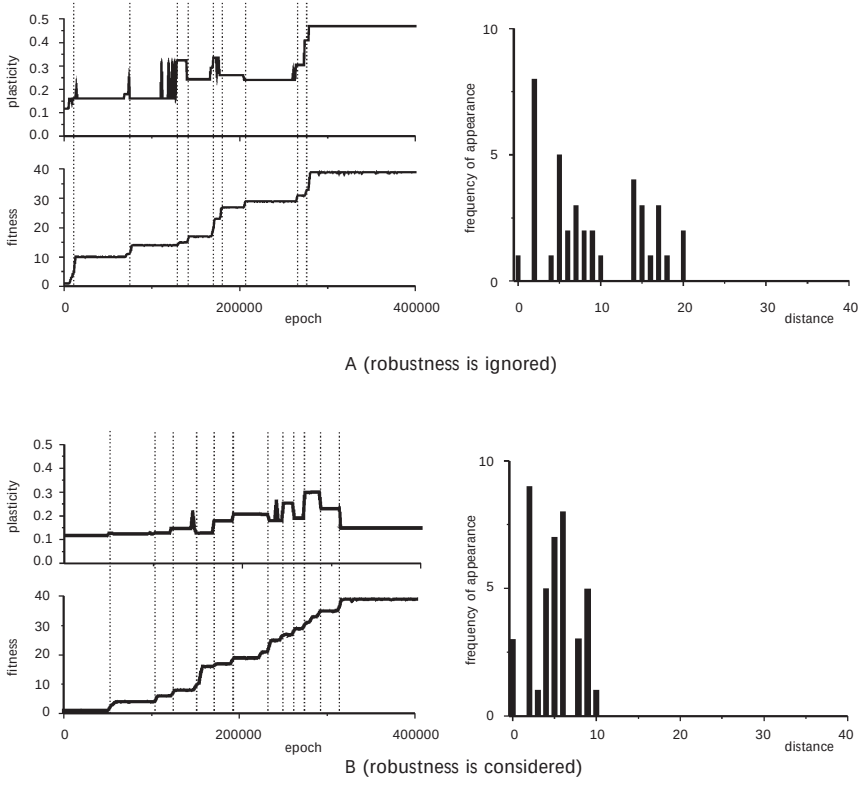

FIG. 16

Two different plots with robustness ignored (A) or robustness minimized (B) in the course of chemostat evolution. In both cases the maximal fitness was achieved, but in the case A the final robustness is substantially higher than in the case B. Moreover, both histograms nicely illustrate this conclusion, in particular the histogram B manifests the expected feature that the biggest part of distance changes is localized in small-value changes, whereas in the first case, distance changes were distributed through twice larger interval of distance-value changes

changes within one module) arises as a by-product when natural selection increases their robustness by stabilizing their shapes. These authors concluded[28,29] that while it is unclear how natural selection could generate modularity directly, there are many scenarios in which natural selection favors the increase of robustness. In their model, modularity arises as a necessary by-product of that reduction. It is the nature of modules to resist changes, hence the process that produces the modularity simultaneously leads populations into evolutionary dead ends. In other words, modularity may correspond to a steep local maximum evolutionary trap, from which there is a small chance to escape to a global maximum.

SUMMARY AND CONCLUSIONS

The main purpose of this paper was to study a simple model of molecular evolution based on binary strings and their folding into secondary structures. Binary strings are evaluated by fitness that reflects a graph-theory similarity between the respective folding and the required folding considered as an evolutionary goal. Binary strings – genotypes are able to replicate with the probability proportional to their fitness. The replication process is not entirely perfect, there exists a very small probability of mutations that produce binary strings slightly different from their parent templates. The dynamics of the replication is fully specified by Eigen's kinetic differential equations that form sound phenomenological basis of theory of molecular replications with mutations. Our results and observations can be summarized as follows:

1. Eigen's theory of replicators forms a sound phenomenological foundation of the molecular Darwinian evolution. Its differential-equation solutions that are specified by a proper selection of the required rate constants offer concentration plots very similar to those experimentally observed[6,36] (see Figs 2 and 3).

2. The Darwinian evolution is an interplay between Monodian[37] **chance and necessity**, between stochastic and deterministic processes. The Darwinian evolution is composed of parts that are fully deterministic, which are fully predictable (*e.g.* genotype–phenotype mapping), but integral parts of evolution involve also processes of a strong stochastic character that could not be well predicted, we can speak only about their basic statistic characteristics (*e.g.* mutations).

3. Wright's idea[1] of **fitness hypersurface** (adaptive or fitness landscape) is of a great heuristic importance and can be considered as one of the greatest achievements of the theory of Darwinian evolution. Then, after this

concept, the Darwinian evolution may be interpreted as a kind of an **evolutionary algorithm**[3,4] for solution of complicated (obviously NP-complete) combinatorial optimization problems.

4. The used model provides a simple mechanism for an explanation of **neutral mutations** and their importance for the Darwinian evolution (see Figs 14 and 15). The existence of neutral mutations on the fitness surface is of great importance for overcoming evolutionary traps of local maxima. The Darwinian evolution is divided into long-term neutral stasis, where mean fitness of population remains unchanged, but the composition of population strings slowly stochastically wanders towards strings with the possibility of 1-point mutational jumps to strings with greater fitness.

5. **Time orientation** of the Darwinian evolution is unambiguous. This is manifested, for example, by the existence of nondecreasing population mean fitness plot. Fitter strings have an evolutionary advantage with respect to other string, they are reproduced more frequently than strings with smaller fitness. In other words, we can say that the Darwinian evolution is a progressive change of mean population genotype such that the corresponding mean fitness is nondecreasing during the whole evolution.

6. Two different time scales[23] may be distinguished in the Darwinian evolution, in particular adaptive stages and neutral stages. An **adaptive stage** corresponds to a sudden change of mean fitness (see Fig. 13), where two different phenotypes coexist; old phenotype has a smaller fitness than a new phenotype. Since the probability of replication of strings is proportional to their fitness, strings corresponding to the new phenotype have a greater chance to be reproduced than the old ones. New strings with a greater fitness win in the course of several evolutionary steps; consequently, an adaptive stage, when a string species is substituted by another string species, seems to an external observer as an extremely short evolutionary stage. On the other hand, a **neutral stage** of the Darwinian evolution consists in a long-term stasis, where the appearance of neutral mutations stochastically prepares a sudden emergence of the next adaptive stage.

7. Evolutionary properties of the presented chemostat model are nicely specified by two types of entropies[17] that are introduced for specification of changes in the chemostat composition. The **fitness entropy** reflects composition changes of binary strings with respect to their fitness, *i.e.* the so-called neutral mutations do not affect this type of entropy. On the other hand, the **genotype entropy** is very sensitive to each composition change in the chemostat; consequently, it is affected by neutral mutations. Therefore, their simultaneous observation in the course of evolution allows us to

simply distinguish between nonneutral and neutral mutations on the evolutionary trajectories.

8. The concept of **robustness** of the phenotype assigned to binary strings is consistent with recent efforts to explain the so-called modularity of secondary structures of RNA molecules. We have demonstrated that the same required phenotype may be achieved such that robustness is maximized. The resulting phenotype is substantially insensitive to mutation of the respective binary string.

9. What are the **limits of the present model**? The used forms of genotype and its mapping onto phenotype are extremely simple, they do not reflect more complex organisms other than viruses, bacteriophages, and some prokaryotic bacteria. For complex organism, the model of genotype (and its mapping onto phenotype) must be much more complex, it should take into account such concepts as variable length, hierarchical structure, other mutations than 1-point ones, *etc*. Most important problem of the recent theory of Darwinian evolution is a lack of a theoretical model explaining the emergence of the so-called "**irreducible complexities**"[38]. The present model is unable to explain this problem even on an elementary level. Recent efforts in Artificial Life are concentrated on areas that might be of importance for elaboration of a more general theory of the Darwinian evolution than the one presented here. In particular, modular aspects of the genotype are recently very intensively studied[39–41] and the problem of symbiosis is modeled[42]. Both problems require a much more complex genotype than the linear strings of symbols with constant lengths which are studied here.

The work was supported by the Scientific Grant Agency of the Slovak Republic (grants 1/7336/20 and 1/8107/01).

REFERENCES

1. Wright S.: *Genetics* **1931**, *16*, 97.
2. Langton C.: *Artificial Life: An Overview*. MIT Press, Cambridge (MA) 1995.
3. Holland J. H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor 1975.
4. Fogel D. B.: *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. The IEEE Press, New York 1995.
5. Dennett D. C.: *Darwin's Dangerous Idea – Evolution and the Meaning of Life*. Penguin Press, London 1995.
6. Spiegelman S.: *Q. Rev. Biophys*. **1971**, *4*, 213.
7. Banzhaf W., Dittrich P., Eller B.: *Physica D (Amsterdam)* **1999**, *125*, 85.

8. Berry G., Boudol G.: *Theor. Comput. Sci.* **1992**, *96*, 217.
9. Dittrich P.: Tutorial held at *ECAL'99, September 13–17, 1999, Lausanne, Switzerland*.
10. Fontana W. in: *Artificial Life II* (C. G. Langton, Ed.), p. 159. Addison–Wesley, Reading (MA) 1991.
11. Gillespie D. T.: *J. Phys. Chem.* **1977**, *81*, 2340.
12. Kvasnička V., Pospíchal J. in: *Advances in Artificial Life* (J. Kelemen and P. Sosik, Eds), ECAL 2001, LNAI 2159, p. 37. Springer Verlag, Berlin 2001.
13. Kvasnička V., Pospíchal J.: *J. Mol. Struct. (THEOCHEM)* **2001**, *547*, 119.
14. Kvasnička V., Pospíchal J.: Artificial Chemistry, Replicators, and Molecular Darwinian Evolution *in Silico*. Presented as a plenary lecture on *2nd Euro-International Symposium on Computational Intelligence, June 16–19, 2002, Košice, Slovakia* (to be published in: *Quo Vadis Machine Intelligence*, World Scientific, Singapore 2003).
15. Eigen M.: *Naturwissenschaften* **1971**, *58*, 465.
16. a) Eigen M., Schuster P.: *Naturwissenschaften* **1977**, *64*, 541; b) Eigen M., Schuster P.: *Naturwissenschaften* **1977**, *65*, 7; c) Eigen M., Schuster P.: *Naturwissenschaften* **1977**, *65*, 341.
17. Newman M. E. J., Engelhardt R.: *Proc. R. Soc. London, Ser. B* **1998**, *265*, 1333.
18. Kauffman S. A.: *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York 1993.
19. Kauffman S. A., Hohnsen S.: *J. Theor. Biol.* **1991**, *149*, 467.
20. Fontana W., Schuster P.: *Biophys. Chem.* **1987**, *26*, 123.
21. Fontana W., Schnabl W., Schuster P.: *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *40*, 3301.
22. Fontana W., Schuster P.: *Science* **1998**, *80*, 1451.
23. Schuster P. in: *Evolutionary Dynamics – Exploring the Interplay of Selection, Accident, Neutrality, and Function* (J. P. Crutchfield and P. Schuster, Eds). Oxford University Press, Oxford (U.K.) 2002 (see many Vienna group's citations therein).
24. Schuster P., Fontana W.: *Physica D (Amsterdam)* **1999**, *133*, 427.
25. Jones B. L., Enns R. H., Rangnekar S. S.: *Bull. Math. Biol.* **1976**, *38*, 15.
26. Waterman M. S.: *Introduction to Computational Biology. Maps, Sequences and Genomes*. Chapman & Hall/CRC, Cambridge (U.K.) 1995.
27. Koča J., Kratochvíl M., Kvasnička V., Matyska L., Pospíchal J.: *Synthon Model of Organic Chemistry and Computer Synthesis Design*. Springer Verlag, Berlin 1989.
28. Ancel L. W., Fontana W.: *J. Exp. Zool. (Mol. Dev. Evol.)* **2000**, *288*, 242.
29. Ancel L. W., Fontana W. in: *Modularity: Understanding the Development and Evolution of Complex Natural Systems* (W. Callebout and D. Raskin-Gutman, Eds). MIT Press, Cambridge (MA), in press.
30. van Nimwegen E., Crutchfield J. P., Huyen M.: *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9716.
31. Wilke C. O.: *Bull. Math. Biol.* **2001**, *63*, 715.
32. Wilke C. O.: *Evolution* **2001**, *55*, 2412.
33. Baldwin J. M.: *Am. Nat.* **1896**, *30*, 441.
34. Belew R. K., Mitchel M. (Eds): *Adaptive Individuals in Evolving Populations: Models and Algorithms*. Addison–Wesley, Reading (MA) 1996.
35. Hinton G. E., Nowlan S. J.: *Complex Systems* **1987**, *1*, 495.
36. Lenski R. E., Travisano M.: *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6808.
37. Monod J.: *Chance and Necessity. An Essay on the Natural Philosophy of Modern Biology*. Vintage, New York 1971.

38. Behe M. J.: *Darwin's Black Box*. Simon & Schuster, New York 1996.
39. Callebout W., Raskin-Gutman D. (Eds): *Modularity: Understanding the Development and Evolution of Complex Natural Systems*. MIT Press, Cambridge (MA), in press.
40. Kvasnička V.: *Neural Network World* **2001**, *5*, 473.
41. Kvasnička V., Pospíchal J.: *Artif. Life* **2003**, *8*, 295.
42. Kvasnička V. in: *Quo Vadis Computational Intelligence?* (P. Sinčák and J. Vasčák, Eds), p. 293. Physica-Verlag, Heidelberg 2000.
43. Wagner A., Stadler P.: *J. Exp. Zool. (Mol. Dev. Evol.)* **1999**, *285*, 119.