Landscapes and Molecular Evolution

Peter Schuster

SFI WORKING PAPER: 1996-07-047

SANTA FE INSTITUTE

LANDSCAPES AND MOLECULAR EVOLUTION

Peter SCHUSTER*

Santa Fe Institute, Santa Fe, New Mexico 87501, USA,
Institut für Molekulare Biotechnologie, D-07708 Jena, Germany, and
Institut für Theoretische Chemie der Universität Wien, A-1090 Wien, Austria

Evolution of RNA molecules *in vitro* is visualized as a hill-climbing process on a fitness landscape that can be derived from molecular properties and functions. The optimization process is shaped by a high degree of redundance in sequence-to-structure mappings: there are many more sequences than structures and sequences folding into the same structure are (almost) randomly distributed in sequence space. Two consequences of this redundance are important for evolution: **shape space covering** by small connected regions in sequence space and the existence of extended **neutral networks**. Both results together explain how nature can fast and efficiently find solutions to complex optimization problems by trial and error while the number of possible genotypes exceeds all imagination. In the presence of neutral networks populations avoid being caught in evolutionary traps and eventually reach the global optimum through a composite dynamics of adaptive walks **and** random drift. Results derived from mathematical analysis are confronted with the results of computer simulation and available experimental data.

1. Evolution and landscapes

Sewall Wright [72] created the metaphor of hill-climbing on a fitness landscape for Darwinian evolution. Optimization then corresponds to an adaptive walk, i.e., to a process that allows to choose the direction for the next step at random from all directions along which fitness does not decrease. Stationary states of populations correspond to local optima of the fitness landscape. Evolution is seen as a series of transitions between optima with increasing fitness values. Wright's metaphor saw a recent revival when sufficiently simple models of fitness landscapes became available [1, 41]. These models are based on spin glass theory [63, 66] or closely related to it like Kauffman's Nk model [42]. Evolution of RNA molecules has been studied by more realistic models that deal explicitly with molecular structures obtained from folding RNA sequences [23, 24]. Fitness values serving as input parameters for evolutionary dynamics were derived through evaluation of the structures. The complexity of RNA fitness landscapes originates from conflicting consequences of structural changes that are reminiscent of "frustration" in the theory of spin glasses [2]. Fitness in the Darwinian sense means numerous progeny. Large fitness values can be achieved by high replication rate or

*Correspondence should be sent to: Institut für Molekulare Biotechnologie, PF 100813, D-07708 Jena, Germany. E-Mail: pks@imb-jena.de, Phone: +49 (3641) 65 6444, Fax: +49 (3641) 65 6446

longevity, i.e., small degradation rate. Structures that are optimal for replication in the RNA model are also easily degraded and hence there is no simple optimal solution with high replication and low degradation rate.

Biological evolution is a highly sophisticated dynamical phenomenon, and its complexity is often confusing. For the purpose of analysis and better understanding it is partitioned into the three simpler processes shown in figure 1 [56]: population dynamics, (population) support dynamics and genotype-phenotype mapping. The three processes are properly visualized in three abstract metric spaces:
(1) the *concentration space* of biochemical reaction kinetics,
(2) the *sequence space* of polynucleotide sequences, and
(3) the *shape space* of biopolymer structures.
Kinetics of chemical reactions or changes in populations are recorded in concentration space, the conventional space of chemical reaction kinetics. It was formalized and put into precise mathematical terms by Martin Feinberg [19]. Variables count numbers of particles, molecules, virions, cells or organisms, for example, and display their changes over time. A population of N individuals distributed over m variants is denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ with $\sum_{i=1}^{m} x_i = N$. Population dynamics is commonly described by differential equations,

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{dt}} \;=\; \dot{\mathbf{x}} \;=\; \mathbf{f}(\mathbf{x})\,; \quad \mathbf{f}(\mathbf{x}) = \Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\Big)\,, \tag{1}$$

by difference equations in case of synchronized generations or by a stochastic process in case of small particle numbers. Concentration space is restricted to the classes of genotypes actually present. When a mutant is produced, a new variable appears in concentration space; when a variant dies out, the corresponding variable disappears. The number of variables (m) matches the number of currently existing genotype classes. As a well known example we mention the selection mutation equation for asexually reproducing individuals introduced by Eigen [13]:

$$\frac{\mathrm{d}x_i}{\mathrm{dt}} \;=\; x_i\Big(k_i Q_{ii} - d_i - \Phi(\mathbf{x})\Big) \;+\; \sum_{j\neq i} k_j Q_{ji} x_j \;; \quad i = 1, 2, \ldots, m \tag{2}$$

Replication and degradation rate constants are denoted by k_i and d_i, respectively, replication accuracies and mutation frequencies are contained in the (bistochastic) matrix $\mathbf{Q} \doteq \{Q_{ij}\}$ ($\sum_{j=1}^{m} Q_{ij} = 1$, $\sum_{i=1}^{m} Q_{ij} = 1$), and $\Phi(\mathbf{x}) = \sum_{i=1}^{m}(k_i - d_i)x_i = \sum_{i=1}^{m} e_i x_i$ is the mean excess production of the population. Population variables are assumed to be normalized and thus the physically meaningful range of variables is confined to the concentration simplex: $S_m \doteq \big\{x_i \geq 0\,\forall\, i = 1, \ldots, m;\, \sum_{i=1}^{m} x_i = 1\big\}$.

The sequence space is a metric point space $(\mathcal{S}; h)$ containing all possible κ^n genotypes of given chain length n. By κ we denote the size of the polynucleotide alphabet (**AUGC**:

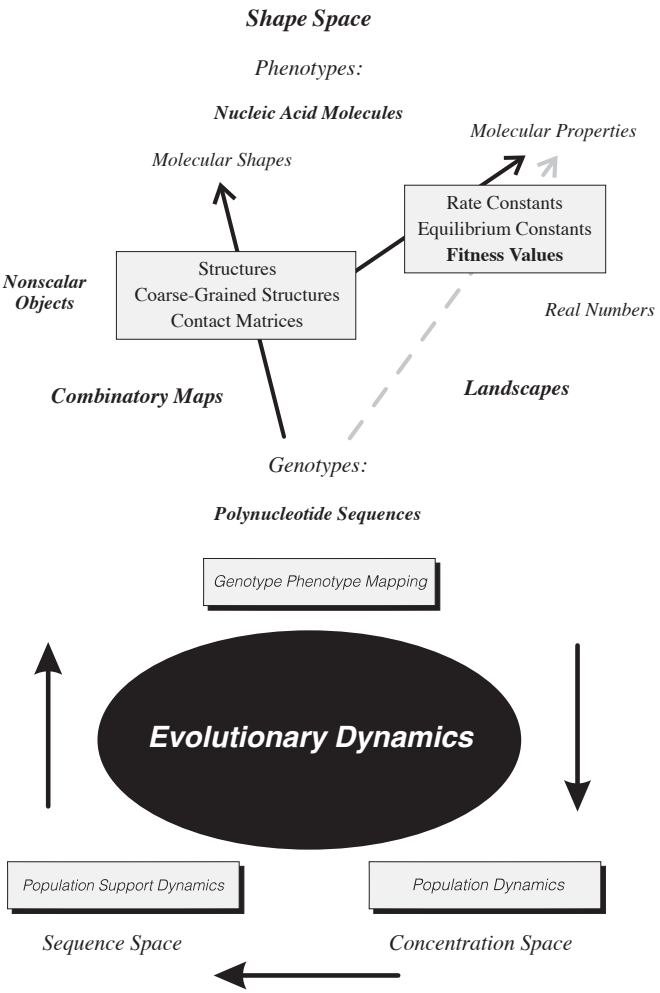


Figure 1: Evolutionary dynamics. *In vitro* evolution of biomolecules is partitioned into three simpler processes: (i) population dynamics, (ii) population support dynamics, and (iii) genotype-phenotype mapping. Population dynamics is tantamount to chemical reaction kinetics of replication, mutation and selection. Population support dynamics describes the migration of populations in the space of genotypes (see section 6). Genotype-phenotype mapping unfolds biological information stored in polynucleotide sequences. Two classes of mappings are distinguished: combinatory maps from one genotype space into another vector space or another space of non-scalar objects and landscapes that map genotype space into the real numbers. In molecular evolution landscapes provide rate constants, equilibrium constants and other more complex scalar properties of phenotypes, fitness values for example. These landscapes are commonly constructed in two steps: (i) a mapping of polynucleotide sequences into molecular structures and (ii) an evolution of structures to yield the (scalar) molecular properties.

$\kappa = 4$, and **GC**: $\kappa = 2$). Distances in sequence space are expressed by the minimum number of single nucleotide exchanges or point mutations converting one sequence into another. This distance, called Hamming distance h [27], induces a metric in sequence space. Connecting all pairs of binary sequences of chain length n with Hamming distance one by straight lines one obtains a hypercube of dimension n. Figure 2 shows how the sequence space in the four letter alphabet can be related to the corresponding binary sequence space. The notion of sequence space is useful for modeling mutation frequencies because genotypes are grouped according to error classes. The uniform error rate model [15, 16], for example, allows to express all mutation rates in terms of only one tunable parameter, the single digit accuracy q (or the error rate $p = 1 - q$), the chain length n and the Hamming distance between template and mutant h_{ij}:

$$Q_{ij} \;=\; q^n \left(\frac{1-q}{q}\right)^{h_{ij}} \;=\; q^n \left(\frac{p}{1-p}\right)^{h_{ij}} \quad . \tag{3}$$

Here the accuracy of replication ($Q_{ii} = q^n$) is independent of the nucleotide sequence.

The shape space contains all phenotypes that are formed by processing genotypes in a given context. In molecular evolution phenotypes are tantamount to spatial structures of biopolymers [65]. Meaningful notions of structure, however, depend on the context: successful descriptions of active sites of enzymes (or ribozymes) or specific binding sites for regulatory elements on DNA, for example, require high precision at atomic resolution as provided by x-ray crystallography whereas studies on phylogenetic conservation of structures can be done much better on the coarse-grained level of *ribbon* diagrams. Some levels of coarse-graining, secondary structures of RNA molecules, for example, are not only of physical relevance but also suitable for mathematical modeling. We assume that phenotypes are elements of some metric space $(\mathcal{Y};\eta)$. In other words, the existence of a distance, η, measuring the relatedness of phenotypes is postulated (although it is not explicitly required for most applications).

The assignment of fitness values to individual genotypes is done in two separate steps (several models, Kauffman's Nk model [42] for example, omit the consideration of a phenotype and assign fitness values directly to genotypes):

$$\textbf{\textit{genotype}} \;\Longrightarrow\; \textbf{\textit{phenotype}} \;\Longrightarrow\; \textbf{fitness}\,.$$

The first step, genotype-phenotype mapping, maps one vector space onto another non-scalar space

$$\Sigma : \quad \big(\mathcal{S};h\big) \;\Longrightarrow\; \big(\mathcal{Y};\eta\big)\;, \tag{4}$$

and has been characterized as a combinatory map [21, 22] in order to indicate that it is no proper landscape. Fitness values are functions of the evolutionarily relevant properties of phenotypes and a fitness landscape, accordingly, is a mapping from shape space into the real numbers (fig.1):

$$\Lambda : \quad \big(\mathcal{Y};\eta\big) \;\Longrightarrow\; \mathbb{R}_1\;. \tag{5}$$

The term "landscape" in this general sense will be used here for a mapping from sequence space into the real numbers irrespectively of its meaning for evolutionary dynamics [60].

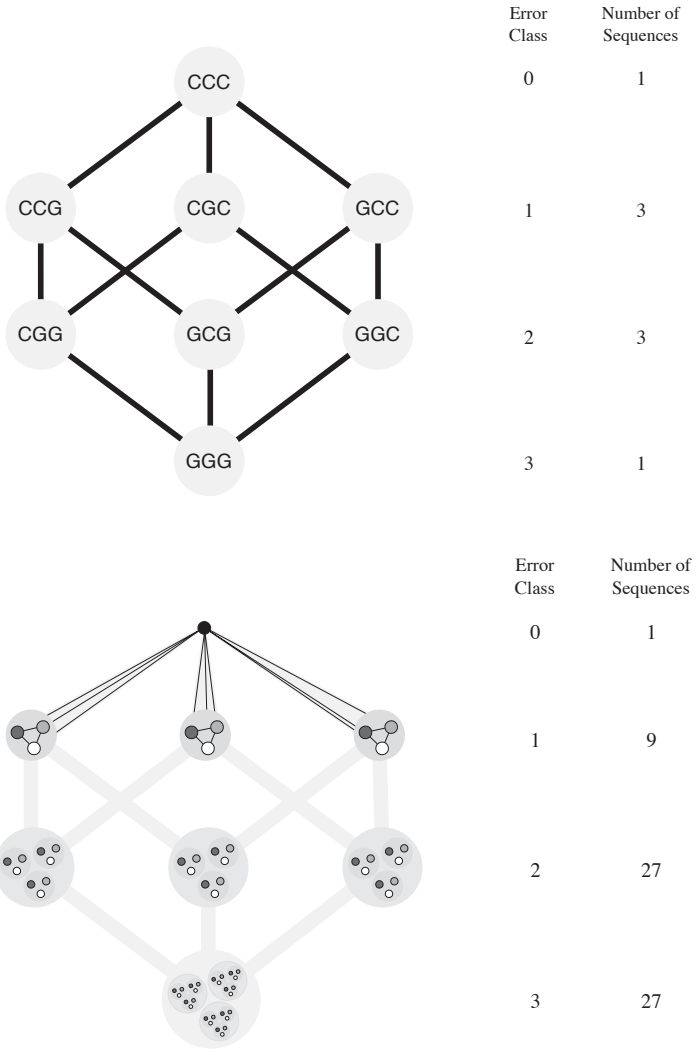


Figure 2: Sequence space. Sequences are represented by points in sequence space. Edges connect sequences of Hamming distance one, i.e., sequences that differ in a single position only. The index of the error class counts the Hamming distance from the reference sequence. The upper part shows the sequence space of **GC**-only sequences of chain length n=3. It has the form of a cube in ordinary three-dimensional space. In general, the sequence spaces of binary sequences are hypercubes of dimension n. The lower part contains the sequence space of natural (**AUGC**) sequences of chain length n=3. Note, that every base can be replaced by three other bases, for example, **A**$\rightarrow$**U**, **A**$\rightarrow$**G**, and **A**$\rightarrow$**C**. The different replacements are coded in different grey levels. Edges are only shown for the error classes 0 and 1.

2. Gradient dynamics

In case of optimization through Darwin's principle of variation and selection leading to "survival of the fittest" population dynamics consists simply of a monotonous approach towards a steady state consisting of a homogeneous population of the fittest genotype. This phenomenon was formulated in mathematical terms through Fisher's fundamental theorem of natural selection [20] which states that the mean fitness of a population, $\bar{\varphi}(t) = \sum_{i=1}^{m} \varphi_i x_i(t)$, is steadily increasing. The theorem is tantamount to saying that mean fitness is a Ljapounov function of the dynamical system (1). In the simplest cases the r.h.s. of the differential equation describing population dynamics represents a gradient

$$\dot{\mathbf{x}} \;=\; \mathrm{grad}\, V(\mathbf{x}) \;, \tag{6}$$

implying that all trajectories of the ODE are (locally) perpendicular to the constant level sets of $V(\mathbf{x})$. In other cases a Riemannian metric of sequence space has been introduced (instead of the conventional Euclidean metric) in order to transform (1) into a generalized gradient system [32, 55, 62, 64]. In particular, Shahshahani [62] defined such a gradient (*Grad*) based on the scalar product $|\mathbf{x}, \mathbf{y}|_{\mathbf{z}} = \sum_{i=1}^{m} x_i y_i \,/\, z_i$ whose application turns the selection equation (equ.2 with $\mathbf{Q} = \mathbb{I}$, the unit matrix) into a gradient: $\dot{\mathbf{x}} = \mathit{Grad}\,\Phi(\mathbf{x})$. It is worth noticing that the introduction of non-zero mutation rates restricts optimization of $\Phi(\mathbf{x})$ to parts of the concentration simplex. In other parts of concentration space $\Phi\big(\mathbf{x}(t)\big)$ may be monotonously decreasing or even a non-monotonous function of time [55, 61].

The gradient form of the selection equation is indeed the mathematical basis for considering evolving populations as walkers climbing uphill on a (rugged) fitness landscape in sequence space [40, 72] (see section 6). Many natural processes, however, interfere with pure selection and restrict the hill-climbing scenario to parts of the concentration simplex or make it entirely obsolete. Examples are mutation, recombination and frequency dependent reproduction. The latter case has been studied extensively by means of replicator equations (for a survey see e.g. Hofbauer & Sigmund [33]). Population dynamics may indeed be as complex as deterministic chaos. If one focusses on replicator dynamics [59] oscillations of concentrations [17], deterministic chaos in homogeneous solutions [53], and spiral waves in spatially resolved systems [6] have been reported.

3. Evolution of RNA molecules

Experimental studies on biological evolution are facing three tantalizing problems: (i) evolutionary phenomena like adaptation need tenthousands to millions of generations in order to become observable in populations and the times required for experiments much too long, (ii) combinatorial explosion leads to numbers of possible genotypes that exceed

all imaginations, and (iii) relations between genotypes and phenotypes are so complex that realistic modeling is impossible. Even in the case of fast replicating bacteria with generation times of approximately one hour tenthousand generations require more than a year and evolution experiments would last several years (Patient microbiologists, for example, have just arrived at some tenthousand generations of *Escherichia coli* under controlled conditions [46]). The numbers of possible sequences are prohibitive for systematic studies: viroid genoms are about threehundred nucleotides long and this implies already 4×10^{180} different RNA sequences of this chain length. The complexity of the relation between genotypes and phenotypes is even more discouraging: to predict the spatial structure of a single biopolymer molecule from its sequence is a very hard (and still unsolved) problem, and the simplest bacteria consist of several thousand different protein and nucleic acids molecules.

The first *in vitro* evolution experiments were performed in the sixties by Sol Spiegelman (for a review see [65]): replication rate constants of RNA from the bacteriophage Qβ were optimized in serial transfer experiments in suitable replication media. These studies demonstrated that evolutionary phenomena are not restricted to the existence of cellular life: molecules capable of reproduction and mutation fulfil the prerequisites for Darwin's principle and behave like asexually replicating individuals (as far as selection and adaptation to environmental conditions are concerned). About the same time Manfred Eigen [13] developed a theoretical frame for molecular evolution which had its roots in chemical reaction kinetics. *In vitro* evolution of RNA molecules circumvents the three problems indicated above: (i) generation times can be reduced to a few seconds under favorable conditions and evolutionary phenomena become observable in a few days, (ii) many genotypes form the same phenotype and are thus selectively neutral in the sense of Motoo Kimura [43], and (iii) genotype and phenotype are two properties of the same molecule, sequence and spatial structure, respectively. Evolution of RNA in the test-tube represents the simplest conceivable system that allows to study adaptation in the laboratory.

Since Spiegelman's pioneering works many different studies on evolutionary phenomena were performed with RNA molecules. RNA replication kinetics, for example, has been studied in detail. It is represented by a multistep polycondensation mechanism and the individual steps are now understood equally well as with other chemical reactions [5]. Serial transfer experiments in the presence of increasing concentration of the RNA degrading enzyme RNase A yielded "resistant" RNA molecules [67]. Catalytic RNA molecules of the group I intron family were trained to cleave DNA rather than RNA [4, 45], the SELEX technique [69] was applied to the selection of RNA molecules which bind to predefined targets with high specificity [36], and ribozymes with novel catalytic functions were derived from libraries of random RNA sequences [3, 7, 48]. Other approaches made use of *in vivo* selection to derive new variants of biopolymers [8, 49] or exploited the capacities of the immune system for evolutionary optimization in the design of catalytic antibodies [47, 54].

In fact, *in vitro* evolution experiments laid out the basis of a new discipline called evolutionary biotechnology (or applied molecular evolution) [14, 37, 38, 39]. In particular, RNA based evolutionary design of biomolecules has already become routine [18, 57]. In order to "breed" molecules for predefined purposes one cannot be satisfied with natural selection being tantamount to a search for the fastest replicating molecular species. In "artificial selection" the properties to be optimized must be disconnected from mere replication. The principle of evolutionary design of biomolecules is shown in figure 3. Molecular properties are optimized iteratively in selection cycles. The first cycle is initiated by a sample of random sequences or alternatively by a population derived from error prone replication of RNA (or DNA) sequences. Each selection cycle consists of three steps: (i) selection of suitable RNA molecules, (ii) amplification through replication, and (iii) diversification through mutation (with artificially increased error rates). The first step, selection of the genotypes that fulfil the predefined criteria best, requires biochemical and biophysical intuition or technological skill. Two strategies are commonly used: selection of best suited candidates from a large sample of different genotypes in homogeneous solutions or spatial separation of genotypes and massively parallel screening. Variants are tested and discarded in case they do not fulfil the predefined criteria. Selected genotypes are amplified, diversified by mutation with error rates adjusted to the current problem, and then the new population is subjected to selection. Optimally adapted genotypes are usually obtained after some twenty to fifty selection cycles. After isolation they can be processed by the conventional methods of molecular biology and genetic engineering.

4. RNA phenotypes and shape space covering

In *in vitro* evolution of RNA molecules genotypes and phenotypes are features of the same molecule, the sequence being the genotype and the spatial structure the phenotype. Mappings of genotypes into phenotypes are then reduced to the relation between sequences and structures of RNA molecules (fig.4). Current knowledge on three-dimensional structures of RNA molecules is rudimentary: only very few structures have been determined so far by crystallogaphy and nmr spectroscopy. Needless to say, spatial structures of RNA molecules are also very hard to predict. The so-called secondary structure of RNA is a coarse grained version of structure that lists only Watson-Crick and **GU** base pairs. It is conceptionally much simpler and allows to perform rigorous mathematical analysis and large scale computations. RNA secondary structure predictions are more reliable than those of full spatial structures. RNA secondary structures make it also possible to define formally consistent distance measures (η) in shape space [22, 34, 44, 51]. In addition, some statistical properties of RNA secondary structures were found to depend very little on choices of algorithms and parameter sets [68].

RNA secondary structures provide an excellent model system to study general relations between genotypes and phenotypes. The conventional approach of structural

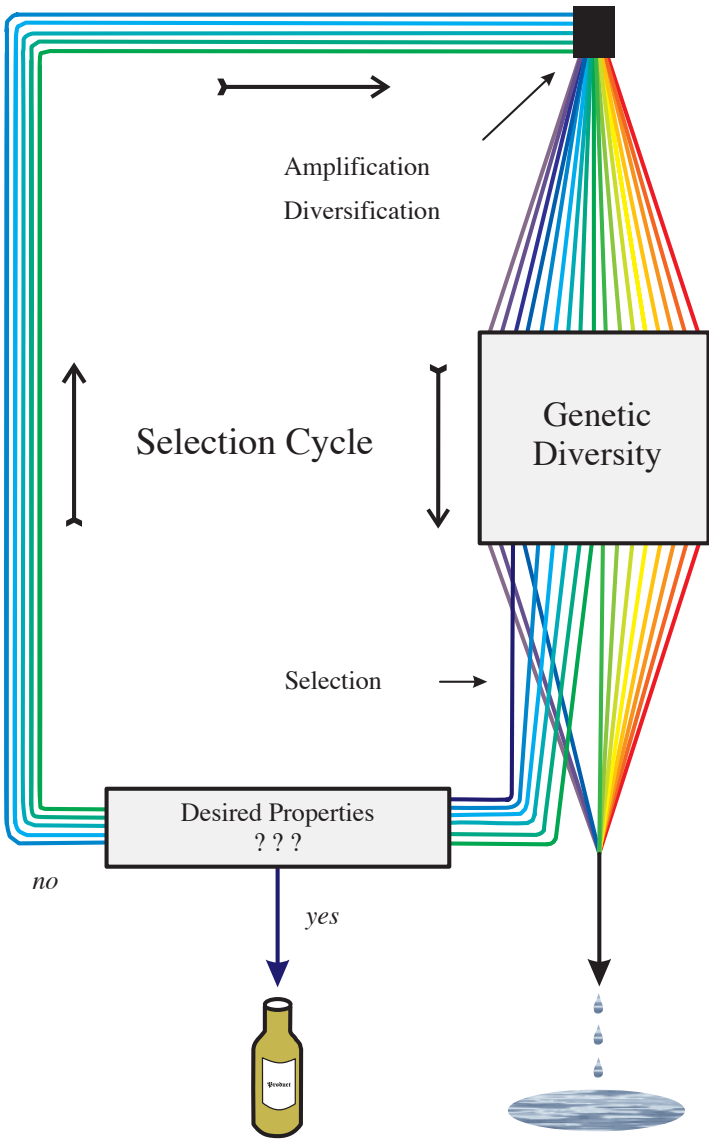


Figure 3: Evolutionary design of biopolymers. Properties and functions of biomolecules are optimized iteratively through selection cycles. Each cycle consists of three different phases: amplification, diversification by replication with high error rates or random synthesis, and selection. Currently successful selection techniques apply one of two strategies: (i) selection in (homogeneous) mixtures using binding to solid phase targets (SELEX) or reactive tags that allow to separate suitable molecules from the rest and (ii) spatial separation of individual molecular genotypes and large scale screening.

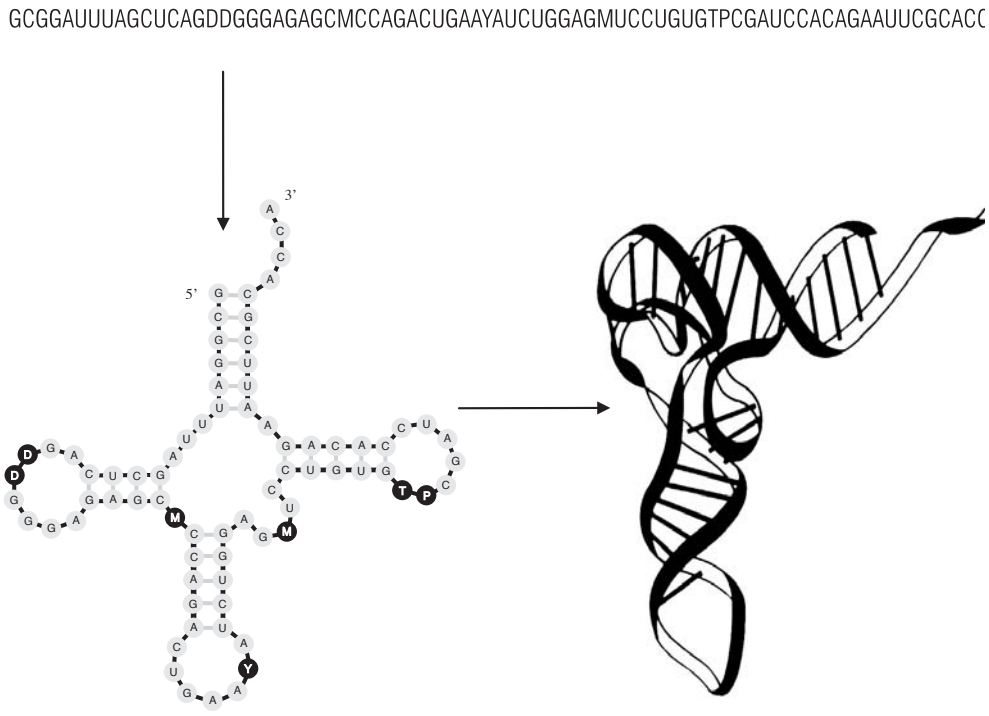


Figure 4: Folding of the natural t-RNA$^{\mathrm{phe}}$ sequence into its three-dimensional structure. The formation of the spatial structure is partitioned into two steps: at first the secondary structure, a coarse-grained version of structure that can be represented by a graph, is built by forming Watson-Crick and **GU** base pairs, and then the secondary structure is folded to yield the full three-dimensional structure.

biology which is essentially concerned with the folding of a single sequence into one structure under minimum free energy conditions has to be extended to a global concept that considers sequence structure relations as (non-invertible) mappings from sequence space into shape space [22, 52, 60]. Application of combinatorics allows to derive an asymptotic expression for the numbers of acceptable structures as a function of the chain length n [31, 58]:

$$S_n \;\approx\; 1.4848 \times n^{3/2}\,(1.8488)^n \;. \tag{7}$$

This expression is based on two assumptions: (i) the minimum stack length is two base pairs ($n_{stack} \geq 2$, i.e., isolated base pairs are excluded) and (ii) the minimal size of hairpin loops is three ($n_{loops} \geq 3$). The numbers of sequences are given by 4^n for natural RNA molecules and by 2^n for **GC**-only or **AU**-only sequences. In both cases there are

Table 1. Common secondary structures of **GC**-only sequences.

	#Sequences		#Struct.	**GC** *		
n	4^n	2^n	S_n	$S_{\mathbf{GC}}$	R_c	n_c
7	16,384	128	6	2	1	120
10	1.05×10^6	1,024	22	11	4	859
15	1.07×10^9	32 768	258	116	43	28 935
20	1.10×10^{12}	1.05×10^6	3 613	1 610	286	902 918
25	1.13×10^{15}	3.36×10^7	55 848	18 590	2 869	30 745 861
30	1.15×10^{18}	1.07×10^9	917 665	218 820	22 718	999 508 805

* The total number of minumum free energy secondary structures formed by **GC**-only sequences is denoted by $S_{\mathbf{GC}}$, R_c is the rank of the least frequent common structure and thus is tantamount to the number of common structures, and n_c is the number of sequences folding into common structures.

more RNA sequences than secondary structures and we are dealing with neutrality in the sense that many RNA sequences form the same (secondary) structure.

Not all acceptable secondary structures are found as minimum free energy structures. The numbers of stable structures were determined by exhaustive folding [25, 26] of all **GC**-only sequences with chain lengths up to $n = 30$ (table 2). The fraction of acceptable structures actually obtained through folding **GC** sequences is between 20% and 50%. It is decreasing with increasing chain length n. Secondary structures are properly grouped into two classes, common ones and rare ones. A straightforward definition of common structures was found to be very useful:

$$\mathcal{C} := \textit{common} \quad \text{iff} \quad \mathcal{N}_{\mathcal{C}} \geq \overline{\mathcal{N}_{\mathcal{S}}} = \frac{\kappa^n}{S_n} = \#(\text{Sequences}) \;/\; \#(\text{Structures}) \;, \qquad (8)$$

wherein $\mathcal{N}_{\mathcal{C}}$ is the number of sequences forming structure $\mathcal{C}$ and κ denotes the size of the alphabet ($\kappa = 2$ for **GC**-only or **AU**-only sequences and $\kappa = 4$ for natural RNA molecules). A structure $\mathcal{C}$ is common if it is formed by more sequences than the average structure. Considering the special example of **GC**-only sequences of chain length $n = 30$ we find that 10.4% of all structures are common. The most common ones (rank 1 and rank 2) are formed by some 1.5 million sequences; this amounts to about 0.13% of sequence space. In total, nearly a billion sequences making up 93.1% of sequence space fold into the common structures. It is worth looking also to the rare end of the structure distribution (table 2). More than 50% of all structures are formed by 100 sequences each or less, and 12 362 structures occur only once in the entire sequence space.

The results of exhaustive folding suggest two important general properties of the above given definition of common structures [25, 26]: (i) the common structures represent only a small fraction of all structures and this fraction decreases with increasing chain length,

Table 2. Distribution of rare structures with **GC**-only sequences of chain length $n = 30$.

Size of the neutral network (m)	Number of different structures $\mathcal{N}(m)$ *
1	12 362
$\leq$ 5	41 487
$\leq$ 10	60 202
$\leq$ 20	80 355
$\leq$ 50	106 129
$\leq$ 100	124 187

* Cumulative numbers $\mathcal{N}(m)$ are given that count the numbers of structures which are formed by m or less (neutral) sequences.

and (ii) the fraction of sequences folding into the common structures increases with chain length and approaches unity in the limit of long chains. Thus, for sufficiently long chains almost all RNA sequences fold into a small fraction of the secondary structures. The effective ratio of sequences to structures is larger than computed from equ.(8) since only common structures play a role in natural evolution and in evolutionary biotechnology.

Inverse folding determines the sequences that fold into a given structure. Application of inverse folding to RNA secondary structures [30] has shown that sequences folding into the same structure are (almost) randomly distributed in sequence space. It is straightforward then to compute a spherical environment (around any randomly chosen reference point in sequence space) that contains at least one sequence (on the average) for every common structure. The radius of such a sphere, called the covering radius r_{cov}, can be estimated from simple probability arguments [57]:

$$r_{cov} \quad = \quad \min\left\{h = 1, 2, \ldots, n \;\middle|\; B_h \geq \kappa^n \,/\, \overline{\mathcal{N}_{\mathcal{S}}} = S_n\right\} \;, \tag{9}$$

with B_h being the number of sequences contained in a ball of radius h. The covering radius is much smaller than the radius of sequence space. The covering sphere represents only a small connected subset of all sequences but contains, nevertheless, all common structures (fig.5) and forms an evolutionarily representative part of shape space.

Numerical values of covering radii are presented in table 3. In the case of natural sequences of chain length $n = 100$ a covering radius of $r_{cov} = 15$ implies that the number of sequences that have to be searched in order to find all common structures is about 4×10^{24}. Although 10^{24} is a very large number (and exceeds the capacities of all currently available polynucleotide libraries), it is negligibly small compared to the size of the entire sequence space that contains 1.6×10^{60} sequences. Exhaustive folding allows to test the estimates derived from simple statistics [26]. The agreement

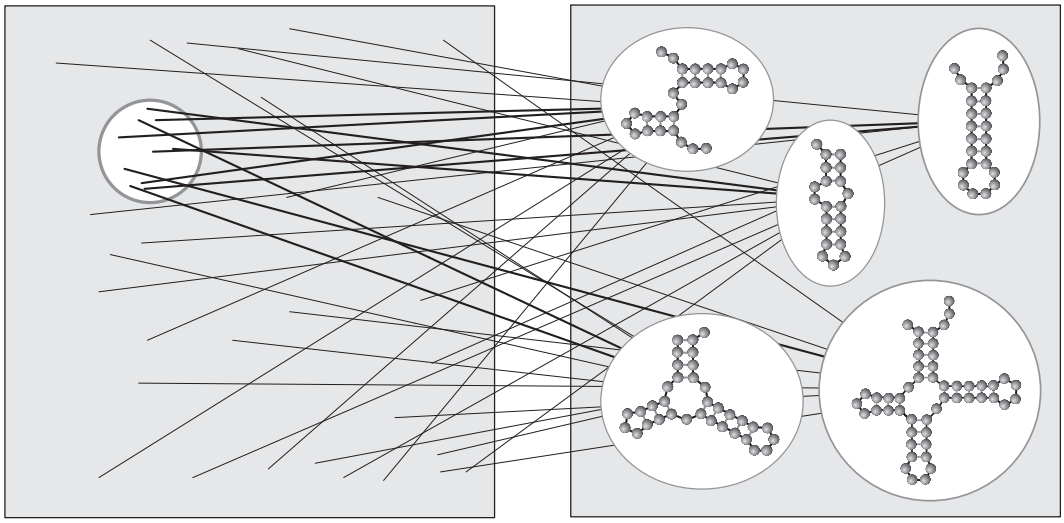


Figure 5: Shape space covering. Only a (relatively small) spherical environment around any arbitrarily chosen reference sequence has to be searched in order to find RNA sequences for every common secondary structure.

for **GC**-only sequences of short chain lengths is surprisingly good. The covering radius increases linearly with chain length with a factor around 1/4. The fraction of sequence space that is required to cover shape space thus decreases exponentially with increasing size of RNA molecules (table 3). We remark that, nevertheless, the absolute numbers of sequences contained in the covering sphere increase also (exponentially) with the chain length.

5. Neutral networks

Since every common structure is formed by a large number of sequences we ought to know also how sequences folding into the same structure are organized in sequence space. Sets of sequences forming the same structure have been called **neutral networks**. So far two approaches were applied to study neutral networks: a mathematical model based on random graph theory [52] and exhaustive folding [26]. The mathematical model assumes that sequences forming the same structure are distributed randomly in the space of compatible sequences. A sequence is compatible with a structure when it can, in principle, fold into the structure in question. It has complementary bases

Table 3. Shape space covering radius for common secondary structures.

n	Covering Radius r^*_{cov} Exhaustive Folding **GC**†	**AU**	Asymptotic Value $S_n(9)$ $\kappa = 2$	$\kappa = 4$	$B_{r_{cov}} \,/\, 4^{\kappa}$ ‡
20	3 (3.4)	2	4	2	3.29×10^{-9}
25	4 (4.7)	2	4	3	4.96×10^{-11}
30	6 (6.1)	3	7	4	7.96×10^{-13}
50			12	6	7.32×10^{-20}
100			26	15	4.52×10^{-37}

* The covering radius is estimated by means of a straightforward statistical estimate based on the assumption that sequences folding into the same structure are randomly distributed in sequence space.

† Exact values derived from exhaustive folding are given in parantheses.

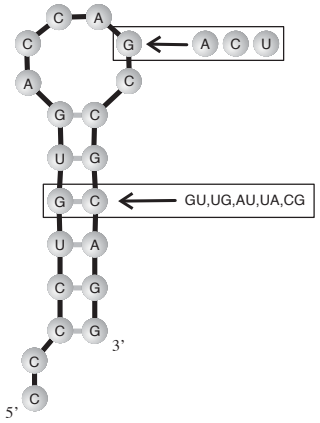
‡ Fraction of **AUGC** sequence space that has to be searched on the average in order to find a minimum at least one sequences for every common structure.

in all pairs of positions which form base pairs in the structure (fig.6). The neutral network of a structure is the subset of its compatible sequences that form the structure under the minimum free condition. Neutral networks are modelled by random graphs in sequence space. The analysis is simplified through partitioning of sequence space into a subspace of unpaired bases and a subspace of base pairs (fig.6). Neutral neighbors in both subspaces are chosen at random and connected to yield the edges of the random graph. The fraction of neighboring pairs that are assigned to be neutral is controlled by a parameter λ. In other words, λ measures the mean fraction of neutral neighbors in sequence space. The statistics of random graphs is studied as a function of λ. The connectivity of networks, for example, changes drastically threshold when λ passes a threshold value:

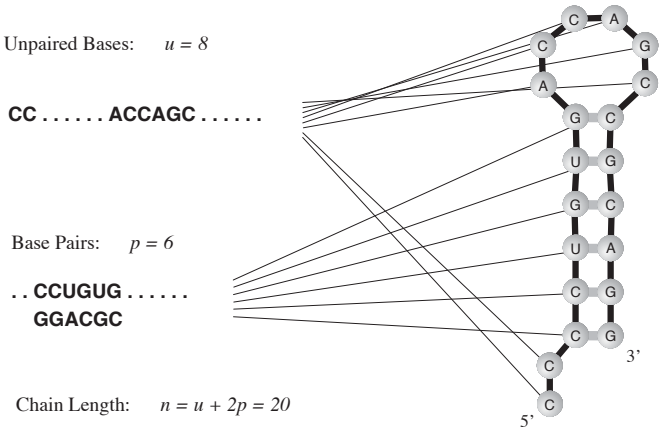
$$\lambda_{cr}(\kappa) \;=\; 1 \;-\; \sqrt[\kappa-1]{\frac{1}{\kappa}}\;. \tag{10}$$

The parameter κ in this equation represents the size of the alphabet. As shown in figure 6 we have $\kappa = 4$ (**A**,**U**,**G**,**C**) for bases in single stranded regions of RNA molecules and $\kappa = 6$(**AU**,**UA**,**UG**,**GU**,**GC**,**CG**) for base pairs. Depending on the particular structure considered the fraction of neutral neighbors is commonly different in the two subspaces of unpaired and paired bases and we are dealing with two different parameter values, λ_u and λ_p, respectively. Neutral networks consist of a single component that spans whole sequence space if $\lambda > \lambda_{cr}$ and below threshold, $\lambda < \lambda_{cr}$, the network is partitioned into a great number of components, in general, a giant component and many small ones (fig.7).

Exhaustive folding allows to check the predictions of random graph theory and reveals further details of neutral networks. The typical sequence of components for neutral



Admissible Exchanges in Compatible Sequences



Partitioning of Sequence Space

Figure 6: Compatible sequences and partitioning of sequence space. A sequence is called compatible with a structure if it contains two matching bases wherever there is a base pair in the structure (upper part). A compatible sequence need not form the structure in question under minimum free energy conditions. The structure, however, will always be found in the set of suboptimal foldings of the sequence. The lower part of the sketch shows partitioning of sequence space into a space of unpaired nucleotides and a space of base pairs as it is used in the application of random graph theory to neutral networks of RNA secondary structures.

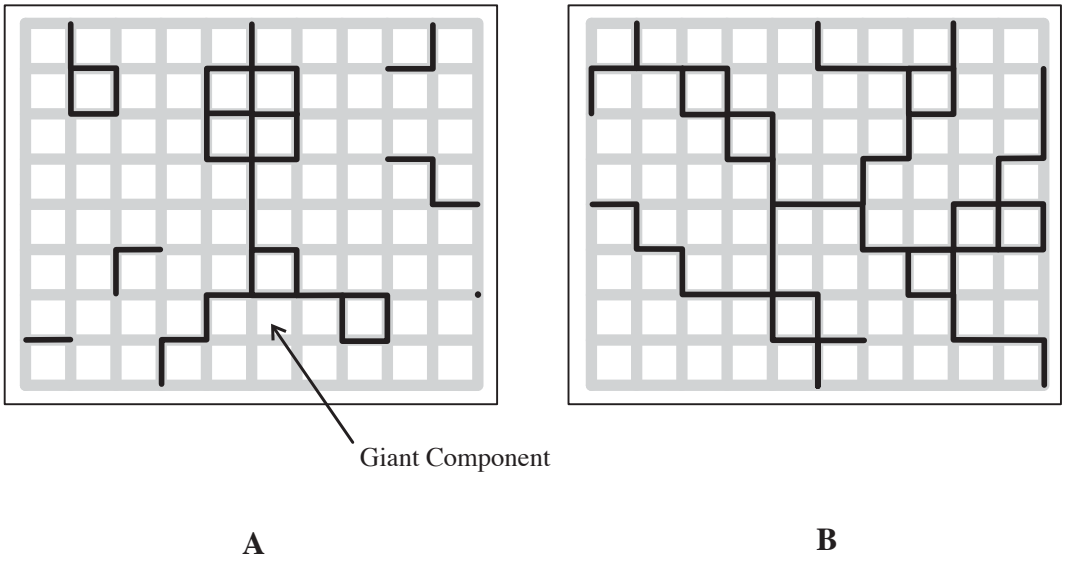


Figure 7: Connectivity of neutral networks. A neutral network consists of many components if the average fraction of neutral neighbors in sequence space (λ) is below a threshold value (λ_{cr}). Random graph theory predicts the existence of one giant component that is much larger than any other component (**A**). If λ exceeds the threshold value (**B**) the network is connected and spans the entire sequence space.

networks (either a connected network spanning whole sequence space or a very large component accompanied by several small ones) is indeed found with many common structures. There are, however, also numerous networks with significantly different sequences of components. We find networks with two as well as four equal sized large components, and three components with an approximate size ratio of 1:2:1. Differences between the predicitions of random graph theory and the results of exhaustive folding were readily explained in terms of special properties of RNA secondary structures [26].

The deviations from the ideal network (as predicted by random graph theory) can be identified as structural features that are not accounted for by some simple base pairing logics. All structures that cannot readily form additional base pairs when sequence requirements are fulfilled behave perfectly normal (class I in figure 8). There are, however, structures that can form additional base pairs (and will generally do so under the minimum free energy criterion) whenever the sequences carry complementary bases at the corresponding positions. Class II structures (fig.8), for example, are least likely to be formed when the overall base composition is 50% **G** and 50% **C**, because the probability for forming an additional base pair and folding into another structure is largest then. When there is an excess of **G** ($\{50+\delta\}\%$) it is much more likely that such a struc-

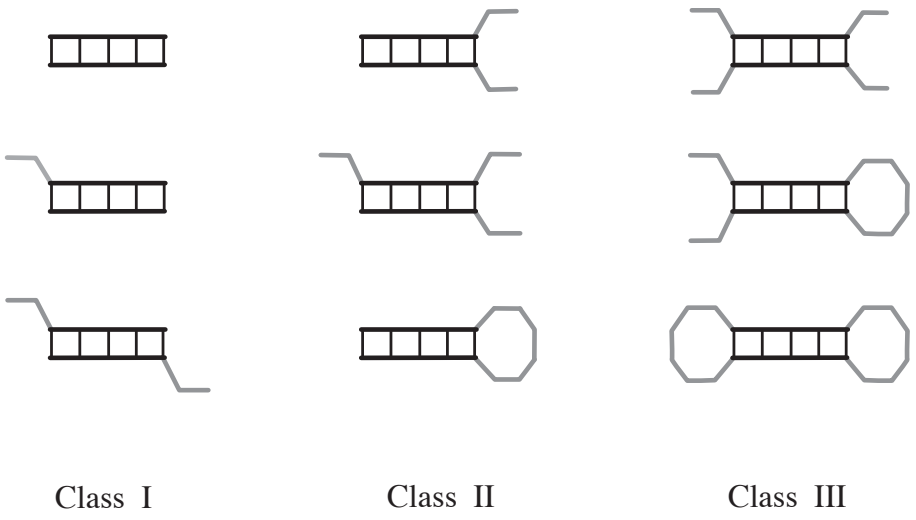


Figure 8: Three classes of RNA secondary structures forming different types of neutral networks. Structures of class I contain no mobile elements (free ends, large loops or joints) or have only mobile elements that cannot form additional base pairs. The mobile elements of structures of class II allow the extension of stacks by additional base pairs at one position. Stacks in class III structures can be extended in two positions. In principle, there are also structures that allow extensions of stacks in more than two ways but they play no role for short chain length (n<30).

ture will actually be formed. The same is true for an excess of **C** and this is precisely reflected by the neutral networks of class II structures with two (major) components: the maximum probabilities for forming class II structures are **G**:**C**=$(50+\delta)$:$(50-\delta)$ for one component and **G**:**C**=$(50+\delta)$:$(50-\delta)$ for the second one. By the same token structures of class III have two (independent) possibilities to form an additional base pair and thus they have the highest probability to be formed if the sequences have excess δ and ε. If no additional information is available we can assume $\varepsilon = \delta$. Independent superposition yields then four equal sized components with G:C compositions of (50+2δ):(50-2δ), 2×(50:50), and (50-2δ):(50+2δ) precisely as it is observed indeed with four component neutral networks. Three component networks are *de facto* four component networks in which the two central (50:50) components have merged to a single one. Neutral networks are thus described well by the random graph model: The assumption that sequences folding into the same structure are randomly distributed in the space of compatible sequences is justified unless special structural features lead to systematic biasses.

6. Optimization on combinatory landscapes

In the Darwinian view of support dynamics (fig.1) populations are thought to optimize mean fitness by adaptive walks through sequence space. The optimization process takes place on two time scales: fast establishment of a quasi-equilibrium between genotypes within the population and slow migration of populations driven by appearence and fixation of rare mutants of higher fitness. The former issue is modelled by the concept of the **molecular quasispecies** [13, 15, 16] which describes the stationary state of populations at a kinetic equilibrium between the fittest type called the master sequence and its frequent mutants (fig.9). Frequencies of mutants are determined by their fitness values as well as by their Hamming distance from the master. The quasispecies represents the genetic reservoir in asexually reproducing species. The concept was originally developed for infinite haploid populations but can be readily extended to finite population sizes [10, 50] and diploid populations [71]. Intuitively, it would pay in evolution to produce as many variant offspring as possible in order to adapt as fast as possible to environmental changes. Intuition, however, is misleading in this case: there is a well defined (genotypic) error threshold; at mutation rates above threshold inheritance breaks down since the number of correct copies of the fittest genotype is steadily decreasing until it is eventually lost. The slow process is based on occasional formation of rare mutants that have higher fitness than the previous master genotype. Population support dynamics was modelled by computer simulation on landscapes related to spin glass theory [42] or on those derived from folding RNA molecules into structures and evaluating fitness related properties [21, 23, 24].

An extension of adaptive evolution to migration of populations through sequence space in absence of fitness differences is straightforward. The genotypic error threshold becomes zero in this limiting case indicating that there are no stationary populations at constant fitness. Neutral evolution has been studied on model landscapes by analytical approaches [12] derived from the random-energy model [11] as well as by computer simulation [28, 29]. More recently the computer simulations were extended to neutral evolution on RNA folding landscapes [35]. In case of selective neutrality populations drift randomly in sequence space by a diffusion-like mechanism. Populations corresponding to large areas in sequence space are partitioned into smaller patches which have some average life time. These feature of population dynamics in neutral evolution was seen in analogy to the formation and evolution of biological species [29]. Population dynamics on neutral networks has been analysed by means of stochastic processes and computer simulation [70]. In analogy to the genotypic error threshold there exists also a phenotypic error threshold that sets a limit to the error rate which sustains a stationary master phenotype (despite always changing genotypes).

In order to visualize the course of adaptive walks on fitness landscapes derived from RNA folding we distinguish single walkers from migration of populations and "non-neutral" landscapes from those built upon extended neutral networks (fig.10):

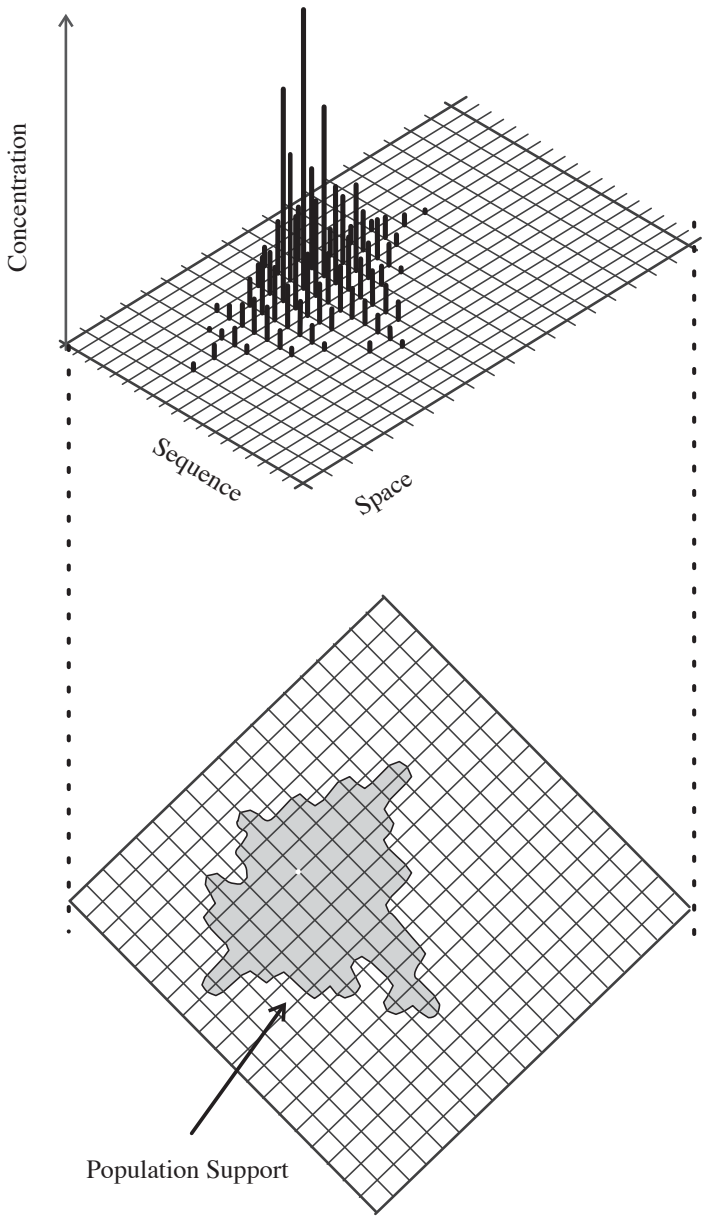


Figure 9: Molecular quasispecies in sequence space. The quasispecies is a stationary mutant distribution surrounding a (fittest and most frequent) master sequence. The frequencies of individual mutants are determined by their fitness values and by their Hamming distances from the master. A quasispecies occupies some region in sequence space called the population support. In the non-stationary case the (population) support migrates through sequence space.

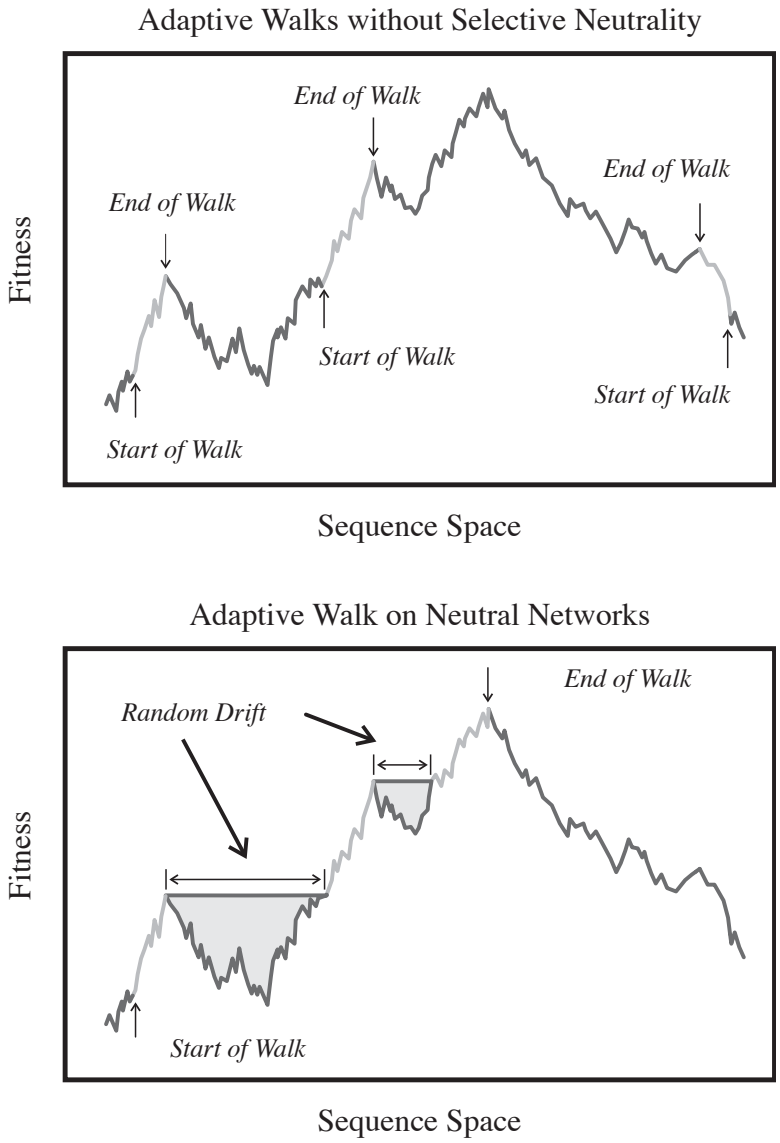


Figure 10: Optimization in sequence space through adaptive walks of populations. Adaptive walks allow to choose the next step arbitrarily from all directions of where fitness is (locally) non-decreasing. Populations can bridge over narrow valleys with widths of a few point mutations. In absence of selective neutrality (upper part) they are, however, unable to span larger Hamming distances and thus will approach only the next major fitness peak. Populations on rugged landscapes with extended neutral networks evolve by a combination of adaptive walks and random drift at constant fitness along the network (lower part). Eventually populations reach the global maximum of the fitness landscape.

(i) Single walkers in the non-neutral case can reach only nearby lying local optima since they are trapped in any local maximum of the fitness landscape [21]. Single walkers are unable to bridge any intermediate value of lower fitness and hence the walk ends whenever there is no one-error variant of higher fitness.
(ii) Populations thus have an "smoothening" effect on landscapes. Even in the non-neutral case sufficiently large populations will be able to escape from local optima provided the Hamming distance to the nearest point with a non-smaller fitness value can be spanned by mutation. In computer simulations of populations with about 3000 RNA molecules jumps of Hamming distances up to six were observed [24].
(iii) In the presence of extended neutral networks optimization follows a combined mechanism: adaptive walks leading to minor peaks are supplemented by random drift along networks that enable populations to migrate to areas in sequence space with higher fitness values (fig.10). Eventually, a local maximum of high fitness or the global fitness optimum is reached.

It is worth noticing how the greatest scholar of evolution, Charles Darwin himself, saw the role of neutral variants [9]: "... This preservation of favourable individual differences and variations, and the destruction of those which are injurious, I have called Natural Selection, or the Survival of the Fittest. Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, ..., or would ultimately become fixed, owing to the nature of the organism and the nature of the conditions." This clear recognition of selective neutrality in evolution by Darwin is remarkable. What he could not be aware of are the extent of neutrality detected in molecular evolution [43] and the positive role neutrality plays in supporting adaptive selection through random drift.

7. Concluding remarks

Molecular evolution experiments with RNA molecules provided essentially two important insights into the nature of evolutionary processes: (i) the Darwinian principle of (natural) selection is no priviledge of cellular life since it is valid also in evolution in the test-tube and (ii) adaptation to the environment and optimization of molecular properties can be observed in a few days or weeks during the course of a typical laboratory experiment.

Most of the studies discussed here were based on structures evaluated by means of the minimum free energy criterium of RNA folding. The validity of the statistical results, like shape space covering or the existence of extended neutral networks, is, however, not limited to minimum free energy folding since they belong to the (largely) algorithm independent properties of RNA secondary structures [68]. RNA secondary structures provide us with a toy world of vast neutrality that allows to study the powerful interplay of adaptive and neutral evolution.

Whether or not the results obtained with secondary structures can be transferred to three-dimensional structures of RNA molecules is an open question. There are strong indications that this will be so although the degree of neutrality is expected to be somewhat smaller than with the secondary structures. The answer to this question will be given only by carrying out suitable experiments on screening of RNA structures and properties in sequence space. Corresponding experiments dealing with binding of RNA molecules to predefined targets as the function to be optimized are under way in several groups. Preliminary data confirm the existence of neutral networks with respect to (coarse-grained) binding constants. A whole wealth of data on protein folding and resilience of protein structures against exchanges of amino acid residues seems to provide evidence for the validity of shape space covering and the existence of extended neutral networks for proteins too.

Neutral evolution, apparently, is not a dispensable addendum to evolutionary optimization as it has been often suggested. In contrary, neutral networks provide a powerful medium through which evolution can become really efficient. Adaptive walks of populations, usually ending in one of the nearby minor peaks of the fitness landscape, are supplemented by random drift on neutral networks. The phase of neutral diffusion ends when the population reaches an area of higher fitness values. Series of adaptive walks interrupted by (selectively neutral) random drift periods allow to approach the global minimum provided the neutral networks are sufficiently large.

8. Acknoledgements

The work reported here has been supported financially by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung* (Projects No.9942-PHY, No.10578-MAT, and No.11065-CHE), by the Commission of the European Community (CEC Contract Study PSS*0884) and by the Santa Fe Institute. Fruitful discussions with Drs. Walter Fontana, Christian Forst and Peter Stadler are gratefully acknowledged.

9. References

[1] C. Amitrano, L. Peliti, and M. Saber. A spin-glass model of evolution. In A. S. Perelson and S. A. Kauffman, editors, *Molecular Evolution on Rugged Landscapes*, volume IX of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 27–38. Addison-Wesley Publ. Co., Redwood City, CA, 1991.

[2] P. W. Anderson. The concept of frustration in spin glasses. *J. Less-Common Metals*, 62:291–294, 1978.

[3] D. P. Bartel and J. W. Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261:1411–1418, 1993.

[4] A. A. Beaudry and G. F. Joyce. Directed evolution of an RNA enzyme. *Science*, 257:635–641, 1992.

[5] C. K. Biebricher and M. Eigen. Kinetics of RNA replication by Qβ replicase. In E. Domingo, J. J. Holland, and P. Ahlquist, editors, *RNA Genetics. Vol.I: RNA Directed Virus Replication*, pages 1–21. CRC Press, Boca Raton, FL, 1988.

[6] M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution - hypercycles stable against parasites. *Physica D*, 48:17–28, 1991.

[7] K. B. Chapman and J. W. Szostak. *In vitro* selection of catalytic RNAs. *Curr. Opinion in Struct. Biol.*, 4:618–622, 1994.

[8] F. C. Christians and L. A. Loeb. Novel human DNA alkyltranferases obtained by random substitution and genetic selection in bacteria. *Proc. Natl. Acad. Sci. USA*, 93:6124–6128, 1996.

[9] C. R. Darwin. *The Origin of Species*, page 81. Number 811 in Everyman's Library. J. M. Dent & Sons, Aldine House, Bedford Street, London, UK, 1928.

[10] L. Demetrius, P. Schuster, and K. Sigmund. Polynucleotide evolution and branching processes. *Bull. Math. Biol.*, 47:239–262, 1985.

[11] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, 1981.

[12] B. Derrida and L. Peliti. Evolution in a flat fitness landscape. *Bull. Math. Biol.*, 53:355–382, 1991.

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

[14] M. Eigen and W. C. Gardiner. Evolutionary molecular engineering based on RNA replication. *Pure Appl. Chem.*, 56:967–978, 1984.

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

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

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

[18] A. D. Ellington. Aptamers achieve the desired recognition. *Current Biology*, 4:427–429, 1994.

[19] M. Feinberg. Mathematical aspects of mass action kinetics. In L. Lapidus and N. R. Amundson, editors, *Chemical Reactor Theory. A Review*, pages 1–78. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[20] R. A. Fisher. *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford, UK, 1930.

[21] W. Fontana, T. Griesmacher, W. Schnabl, P. Stadler, and P. Schuster. Statistics of landscapes based on free energies, replication and degradation rate constants of RNA secondary structures. *Mh. Chem.*, 122:795–819, 1991.

[22] W. Fontana, D. A. Konings, P. Stadler, and P. Schuster. Statistics of RNA secondary structures. *Biopolymers*, 33:1389–1404, 1993.

[23] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A*, 40:3301–3321, 1989.

[24] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophys. Chem.*, 26:123–147, 1987.

[25] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. Hofacker, P. Stadler, and P. Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Mh.Chem.*, 127:355–374, 1996.

[26] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. Hofacker, P. Stadler, and P. Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structure of neutral networks and shape space covering. *Mh.Chem.*, 127:375–389, 1996.

[27] R. W. Hamming. Error detecting and error correcting codes. *Bell Syst. Tech. J.*, 29:147–160, 1950.

[28] P. G. Higgs and B. Derrida. Stochastic models for species formation in evolving populations. *J. Physics A*, 24:L985–L991, 1991.

[29] P. G. Higgs and B. Derrida. Genetic distance and species formation in evolving populations. *J. Mol. Evol.*, 35:454–465, 1992.

[30] I. Hofacker, W. Fontana, P. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Mh. Chem.*, 125:167–188, 1994.

[31] I. Hofacker, P. Schuster, and P. Stadler. Combinatorics of RNA secondary structures. *Submitted to SIAM J. Disc. Math.*, 1996.

[32] J. Hofbauer. The selection mutation equation. *J. Math. Biol.*, 23:41–53, 1985.

[33] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems.* Cambridge University Press, Cambridge UK, 1988.

[34] P. Hogeweg and B. Hesper. Energy directed folding of RNA sequences. *Nucleic Acids Research*, 12:67–74, 1984.

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

[36] R. D. Jenison, S. C. Gill, A. Pardi, and B. Polisky. High-resolution molecular discrimination by RNA. *Science*, 263:1425–1429, 1994.

[37] G. F. Joyce. Directed molecular evolution. *Sci. Am.*, 267(6):48–55, 1992.

[38] S. A. Kauffman. Autocatalytic sets of proteins. *J. Theor. Biol.*, 119:1–24, 1986.

[39] S. A. Kauffman. Applied molecular evolution. *J. Theor. Biol.*, 157:1–7, 1992.

[40] S. A. Kauffman. *The Origins of Order. Self-Organization and Selection in Evolution.* Oxford University Press, Oxford, UK, 1993.

[41] S. A. Kauffman and S. Levine. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11–45, 1987.

[42] S. A. Kauffman and E. D. Weinberger. The n-k model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.*, 141:211–245, 1989.

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

[44] D. Konings and P. Hogeweg. Pattern analysis of RNA secondary structure. Similarity and consensus of minimal-energy folding. *J. Mol. Biol.*, 207:597–614, 1989.

[45] N. Lehman and G. F. Joyce. Evolution *in vitro*: Analysis of a lineage of ribozymes. *Current Biology*, 3:723–734, 1993.

[46] R. E. Lenski and M. Travisano. Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA*, 91:6808–6814, 1994.

[47] R. A. Lerner, S. J. Benkovic, and P. G. Schultz. At the crossroads of chemistry and immunology: Catalytic antibodies. *Science*, 252:659–667, 1991.

[48] J. R. Lorsch and J. W. Szostak. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature*, 371:31–36, 1994.

[49] K. M. Munir, D. C. French, and L. A. Loeb. Thymidine kinase mutants obtained by random selection. *Proc. Natl. Acad. Sci. USA*, 90:4012–1016, 1993.

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

[51] C. Reidys and P. F. Stadler. Bio-molecular shapes and algebraic structures. *Computers Chem.*, 20:85–94, 1996.

[52] C. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps - Neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 1995. Submitted. SFI-Preprint Series No. 95-07-058.

[53] W. Schnabl, P. F. Stadler, C. Forst, and P. Schuster. Full characterization of a strange attractor. Chaotic dynamics on low-dimensional replicator systems. *Physica D*, 48:65–90, 1991.

[54] P. G. Schultz. Catalytic antibodies. *Angew.Chem.*, 28:1283–1295, 1989.

[55] P. Schuster. Potential functions and molecular evolution. In M. Markus, S. C. Müller, and G. Nicolis, editors, *From Chemical to Biochemical Organization*, volume 39 of *Springer Series in Synergetics*, pages 149–165. Springer-Verlag, Berlin, 1988.

[56] P. Schuster. Artificial life and molecular evolutionary biology. In F. Morán, A. Moreno, J. J. Morelo, and P. Chacón, editors, *Advances in Artificial Life*, volume 929 of *Lecture Notes in Artificial Intelligence*, pages 3–19. Springer-Verlag, Berlin, 1995.

[57] P. Schuster. How to search for RNA structures. Theoretical concepts in evolutionary biotechnology. *Journal of Biotechnology*, 41:239–257, 1995.

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

[59] P. Schuster and K. Sigmund. Replicator dynamics. *J. Theor. Biol.*, 100:533–538, 1983.

[60] P. Schuster and P. Stadler. Landscapes: Complex optimization problems and biopolymer structures. *Computers Chem.*, 18:295–314, 1994.

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

[62] S. Shahshahani. A new mathematical framework for the study of linkage and selection. *Mem. Am. Math. Soc.*, 211, 1979.

[63] D. Sherrington and S. Kirkpatrick. Solvable model of a spin glass. *Phys. Rev. Letters*, 35:1792–1796, 1975.

[64] K. Sigmund. The maximum priciple for replicator equations. In W. Ebeling and M. Peschel, editors, *Lotka-Volterra Approach to Cooperation and Competition in Dynamical Systems*, pages 63–71. Akademie-Verlag, Berlin, 1985.

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

[66] D. L. Stein. Disordered systems: Mostly spin glasses. In D. L. Stein, editor, *Lectures in the Sciences of Complexity*, volume 1, pages 301–353. Addison-Wesley Publ. Co., Redwood City, CA, 1989.

[67] G. Strunk. *Automatized evolution experiments* in vitro *and natural selection under controlled conditions by means of the serial transfer technique.* PhD thesis, Universität Braunschweig, 1993.

[68] M. Tacker, P. Stadler, E. Bornberg-Bauer, I. Hofacker, and P. Schuster. Algorithm independent properties of RNA secondary structure predictions. Submitted to Eur.Biophys.J., 1996.

[69] C. Tuerk and L. Gold. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249:505–510, 1990.

[70] J. Weber, C. Reidys, C. Forst, and P. Schuster. Evolutionary dynamics on neutral networks. Preprint, IMB Jena, Germany, 1996.

[71] T. Wiehe, E. Baake, and P. Schuster. Error propagation in reproduction of diploid organisms. A case study in single peaked landscapes. *J. Theor. Biol.*, 177:1–15, 1995.

[72] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones, editor, *Int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.