Optimization, Selection, and Neutrality
What we can learn from Nature

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

EvoStar 2009
Tübingen, 15.– 17.04.2009
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
1. From Darwin to molecular biology

2. Selection in the test tube

3. Chemical kinetics of molecular evolution

4. Evolutionary biotechnology

5. The RNA model and neutrality

6. Simulation of molecular evolution
Color patterns on animal skins
Bates' mimicry

Müller's mimicry

Different forms of mimicry observed in nature
Bates' mimicry

Different forms of mimicry observed in nature
Three necessary conditions for Darwinian evolution are:

1. **Multiplication**, 

2. **Variation**, and 

3. **Selection**.

Biologists distinguish the **genotype** – the genetic information – and the **phenotype** – the organisms and all its properties. The **genotype** is unfolded in development and yields the **phenotype**.

**Variation** operates on the **genotype** – through mutation and recombination – whereas the **phenotype** is the target of **selection**.

One important property of the Darwinian mechanism is that variations in the form of mutation or recombination events occur **uncorrelated** to their **effects** on the **selection** of the **phenotype**.
Two variants with a mean progeny of ten or eleven descendants

\[ s = \frac{f_2 - f_1}{f_1} = 0.1 \]
Selection of advantageous mutants in populations of $N = 10\,000$ individuals

$N_1(0) = 9999, N_2(0) = 1; \quad s = 0.1, 0.02, 0.01$
Genotype, Genome

Collection of genes

Developmental program

Highly specific environmental conditions

Unfolding of the genotype

Phenotype

Evolution explains the origin of species and their interactions
Genotype, Genome

GCGGATTAGCTCAGTTGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

systems biology

‘the new biology is the chemistry of living matter’

Thomas Cech
RNA catalysis

Linus Pauling and Emile Zuckerkandl
molecular evolution

Gerhard Braunitzer
hemoglobin sequence

Max Perutz
DNA structure

Manfred Eigen

John Kendrew

DNA

RNA
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
Three necessary conditions for Darwinian evolution are:

1. **Multiplication**, 
2. **Variation**, and 
3. **Selection**.

**Darwinian evolution in the test tube**

All three conditions are fulfilled not only by cellular organisms but also by **nucleic acid molecules** - DNA or RNA - in suitable **cell-free experimental assays**.
James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004
Nobel prize 1962

1953 – 2003 fifty years double helix

The three-dimensional structure of a short double helical stack of B-DNA
DNA structure and DNA replication
Replication fork' in DNA replication

The mechanism of DNA replication is 'semi-conservative'.
Complementary replication is the simplest copying mechanism of RNA. Complementarity is determined by Watson-Crick base pairs:

\[ G \equiv C \quad \text{and} \quad A \equiv U \]
\[
\begin{align*}
\frac{dx_1}{dt} &= f_2 x_2 \quad \text{and} \quad \frac{dx_2}{dt} = f_1 x_1 \\
\eta(t) &= \eta(0) e^{-ft} \\
\zeta(t) &= \zeta(0) e^{ft}
\end{align*}
\]

\( x_1 = \sqrt{f_2} \xi_1, \quad x_2 = \sqrt{f_1} \xi_2, \quad \zeta = \xi_1 + \xi_2, \quad \eta = \xi_1 - \xi_2, \quad f = \sqrt{f_1 f_2} \)

Complementary replication as the simplest molecular mechanism of reproduction
Kinetics of RNA replication

C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
*Biochemistry* **22**:2544-2559, 1983
Evolution of RNA molecules based on Qβ phage


F. Öhlenschlager, M. Eigen, 30 years later – A new approach to Sol Spiegelman‘s and Leslie Orgel‘s in vitro evolutionary studies. Orig. Life Evol. Biosph. 27 (1997), 437-457
Application of the serial transfer technique to RNA evolution in the test tube

RNA sample

Stock solution: Qβ RNA-replicase, ATP, CTP, GTP and UTP, buffer
Decrease in mean fitness due to quasispecies formation

The increase in RNA production rate during a serial transfer experiment
Results from molecular evolution in laboratory experiments:

• Evolutionary optimization does not require cells and occurs in molecular systems too.

• *In vitro* evolution allows for production of molecules for predefined purposes and gave rise to a branch of biotechnology.
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
The Hypercycle

A Principle of Natural Self-Organization

Part A: Emergence of the Hypercycle

Máthias Eigen
Max-Planck-Institut für Biophysikalische Chemie, D-3400 Göttingen

Peter Schuster
Institut für Theoretische Chemie und Biologische Kinetik der Universität, A-1090 Wien

The paper deals with the question of whether the concept of the hypercycle, introduced by Eigen, is also applicable to higher organisms. It is argued that the hypercycle hypothesis can be extended to encompass higher organisms if the basic principles of self-replication and self-organization are suitably modified. The hypercycle hypothesis is then used to explain the emergence of complex biological systems from simple molecular precursors.

Introduction

1.1. Nature and Effect

The origin of life on Earth is a fundamental question in biology. The molecular evolution of biological macromolecules provides us with a key to understanding the origin of life. The hypothesis of molecular evolution suggests that life emerged from simple molecular precursors through processes of self-replication and self-organization.

1.2. The Hypercycle Hypothesis

The hypercycle hypothesis, introduced by Eigen, is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

1.3. Molecular Evolution

Molecular evolution is the process by which the genetic material of a population changes over time. It is driven by natural selection and is the basis for the evolution of new species.

1.4. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Hypercycle Hypothesis

2.1. Emergence of the Hypercycle

The hypercycle hypothesis is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

2.2. The Emergence of Life

The emergence of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

2.3. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

2.4. The Hypercycle Hypothesis

The hypercycle hypothesis is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

3. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Hypercycle Hypothesis

4. The Emergence of Life

The emergence of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

5. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

6. The Hypercycle Hypothesis

The hypercycle hypothesis is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

7. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Hypercycle Hypothesis

8. The Emergence of Life

The emergence of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

9. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Hypercycle Hypothesis

10. The Emergence of Life

The emergence of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Origin of Life

11. The Hypercycle Hypothesis

The hypercycle hypothesis is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

12. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Hypercycle Hypothesis

13. The Emergence of Life

The emergence of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

The Origin of Life

14. The Hypercycle Hypothesis

The hypercycle hypothesis is a model for the self-organization of complex biological systems from simple molecular precursors. It is based on the concept of a self-replicating system that is capable of self-organization and self-reproduction.

15. The Origin of Life

The origin of life is a fundamental question in biology. The most widely accepted theory is the RNA-world hypothesis, which suggests that RNA molecules were the first self-replicating systems on Earth.

Chemical kinetics of molecular evolution
A point mutation is caused by an incorrect incorporation of a nucleobase into the growing chain during replication.
Replication and mutation are parallel chemical reactions.
Chemical kinetics of replication and mutation as parallel reactions
**Mutation-selection equation:** \([I_i] = x_i \geq 0, \ f_i > 0, \ Q_{ij} \geq 0\)

\[
\frac{dx_i}{dt} = \sum_{j=1}^{n} f_j Q_{ji} x_j - x_i \phi, \quad i=1,2,\ldots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \bar{f}
\]

*Solutions* are obtained after integrating factor transformation by means of an eigenvalue problem

\[
x_i(t) = \frac{\sum_{k=0}^{n-1} \ell_{ik} \cdot c_k(0) \cdot \exp(\lambda_k t)}{\sum_{j=1}^{n} \sum_{k=0}^{n-1} \ell_{jk} \cdot c_k(0) \cdot \exp(\lambda_k t)}; \quad i=1,2,\ldots,n; \quad c_k(0) = \sum_{i=1}^{n} h_{ki} x_i(0)
\]

\[
W \div \{f_i Q_{ij}; \ i, j=1,2,\ldots,n\}; \quad L = \{\ell_{ij}; \ i, j=1,2,\ldots,n\}; \quad L^{-1} = H = \{h_{ij}; \ i, j=1,2,\ldots,n\}
\]

\[
L^{-1} \cdot W \cdot L = \Lambda = \{\lambda_k; \ k=0,1,\ldots,n-1\}
\]
Perron-Frobenius theorem applied to the value matrix $W$

$W$ is primitive: (i) $\lambda_0$ is real and strictly positive

(ii) $\lambda_0 > |\lambda_k|$ for all $k \neq 0$

(iii) $\lambda_0$ is associated with strictly positive eigenvectors

(iv) $\lambda_0$ is a simple root of the characteristic equation of $W$

(v-vi) etc.

$W$ is irreducible: (i), (iii), (iv), etc. as above

(ii) $\lambda_0 \geq |\lambda_k|$ for all $k \neq 0$
Formation of a quasispecies in sequence space

\[ p = 0 \]

Master sequence
Formation of a quasispecies in sequence space

\[ p = 0.25 \ p_{cr} \]
Formation of a quasispecies
in sequence space

\[ p = 0.50 \, p_{cr} \]
Formation of a quasispecies in sequence space

\[ p = 0.75 \, p_{cr} \]
$p \geq p_{cr}$

Mutant cloud

Uniform distribution in sequence space
Quasispecies Driving virus populations through threshold

The error threshold in replication
Antiviral strategy on the horizon

Error catastrophe laid conceptual origins in the middle of the 20th century, when the consequences of mutations on an enzyme involved in reovirus synthesis, as a theory of aging. In these times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be discussed through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variations in different organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to control the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the theory of cancerous viruses and cells. Population genetics rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The term of genetic polymorphism among individuals of the same biological species came to be a surprise when the first results on comparative electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA manipulation, and rapid nucleic acid sequencing techniques, molecular analyses of enzymes reinforced the conclusion of extreme inter-individual genetic variation within the same species. New, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly diverse. Most cellular processes, including such essential information-bearing and transcribing events as gene regulation, replication, transcription and translation, are increasingly perceived in a quantitatively accurate fashion, and in particular RNA viruses, are among the most active examples of exploitation of replication accuracy for survival.

Error catastrophe, or the loss of mass, genetic information through accidental genetic variation, was formulated in qualitative terms as a consequence of quasiparasites theory, which was first developed to explain self-regulation and adaptability of parasites to exploit early stages of life. However, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged in a possible natural strategy. This is the topic of the current special issue of Virus Research.

Preparation and editing of this issue provided an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I am conscious (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasiparasites dynamics in infected individuals to learn more about the maturation of viral diversity and its effect on the evolution of viruses, as well as the interactions of viral diversity with the host's immune response. (ii) Another on a deeper understanding of the mechanism of viral evolution, particularly in the context of the host's immune response. (iii) Finally, an analysis of the role of viruses in the evolution of quasiparasites dynamics and their interaction with the host's immune response. The latter can be summarized by saying that a quasiparasite virus, by virtue of its ability to interact with host cells, is able to modulate the host's immune response, leading to a complex interplay between viral and cellular factors. This interplay can result in the evolution of quasiparasites dynamics and their interaction with the host's immune response. Therefore, a better understanding of the role of viruses in the evolution of quasiparasites dynamics and their interaction with the host's immune response is essential for the development of effective antiviral strategies.
Molecular evolution of viruses
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
**Evolutionary design of RNA molecules**


An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules.
The SELEX-technique for evolutionary design of strongly binding molecules called aptamers
Formation of secondary structure of the tobramycin binding RNA aptamer with $K_D = 9$ nM

The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel,
Christian Jäckel, Peter Kast, and Donald Hilvert.
Protein design by directed evolution.
Application of molecular evolution to problems in biotechnology
Artificial evolution in biotechnology and pharmacology


Results from kinetic theory of molecular evolution and evolution experiments:

- Evolutionary optimization does not require cells and occurs as well in cell-free molecular systems.

- Replicating ensembles of molecules form stationary populations called quasispecies, which represent the genetic reservoir of asexually reproducing species.

- For stable inheritance of genetic information mutation rates must not exceed a precisely defined and computable error-threshold.

- The error-threshold can be exploited for the development of novel antiviral strategies.

- In vitro evolution allows for production of molecules for predefined purposes and gave rise to a branch of biotechnology.
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
5'-end GCGGAUUAGCUCAUGGAGACGCAGAACUGAUAGACUCUGGAGGUGUCUGUUGUUGAUCAGAUCACAGAUAUUCGACCA 3'-end

Definition of RNA structure
Criterion: Minimum free energy (mfe)

Rules: \(_(\_\_\_)\_ \in \{AU, CG, GC, GU, UA, UG\}\)

A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs
RNA sequence: GUAUCGAAAUAUCGUAGCCUAUGGGGAUGCUGGACCGGUCCCAUCCGUACUCCA

RNA folding:
Structural biology, spectroscopy of biomolecules, understanding molecular function

Inverse folding of RNA:
Biotechnology, design of biomolecules with predefined structures and functions

Inverse Folding Algorithm

Iterative determination of a sequence for the given secondary structure

RNA structure of minimal free energy:

Sequence, structure, and design
The inverse folding algorithm searches for sequences that form a given RNA structure.
many genotypes $\Rightarrow$ one phenotype
Prediction of RNA secondary structures: from theory to models and real molecules

Peter Schuster\textsuperscript{1,2}

\textsuperscript{1}Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria
\textsuperscript{2}The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

E-mail: pks@tbi.univie.ac.at
The surrounding of **GUCAAUCA** in sequence space
One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule of chain length $n=50$ in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule of chain length $n=50$ in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space
<table>
<thead>
<tr>
<th>Number of Structures:</th>
<th>Number</th>
<th>Mean Value</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>52.31</td>
<td>85.30</td>
<td>9.24</td>
</tr>
<tr>
<td>2</td>
<td>99875</td>
<td>16.949991</td>
<td>30.757651</td>
<td>5.545958</td>
</tr>
<tr>
<td>3</td>
<td>99875</td>
<td>11.647973</td>
<td>23.140715</td>
<td>4.810480</td>
</tr>
</tbody>
</table>

Total Hamming Distance: 150000
Nonzero Hamming Distance: 99875
Degree of Neutrality: 50125

1. (((((((((.(((((.(((......)))..)))).))).))).))).))............. 50125 0.334167
2. .......................................................... 2856 0.019040
3. …………………….................................................................. 2799 0.018660
4. ..................................................................................... 2417 0.016113
5. …………………………………….…………………………………………………. 2265 0.015100
6. .................................................................. 2233 0.014887
7. …………………………………….…………………………………………………. 1442 0.009613
8. …………………………………….…………………………………………………. 1081 0.007207
9. .................................................................. 1025 0.006833
10. …………………………………….…………………………………………………. 1003 0.006687
11. …………………………………….…………………………………………………. 963 0.006420
12. .................................................................. 860 0.005733
13. …………………………………….…………………………………………………. 800 0.005333
14. .................................................................. 548 0.003653
15. .................................................................. 362 0.002413
16. …………………………………….…………………………………………………. 337 0.002247
17. …………………………………….…………………………………………………. 241 0.001607
18. .................................................................. 231 0.001540
19. …………………………………….…………………………………………………. 225 0.001500
20. .................................................................. 202 0.001347

Shadow – Surrounding of an RNA structure in shape space:
**AUGC** alphabet, chain length n=50
ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNEAN, ETC., SOCIETIES;
AUTHOR OF "JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE
BOUND THE WORLD."

LONDON:

JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.
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, as perhaps we see in certain polymorphic species, or would ultimately become fixed, owing to the nature of the organism and the nature of the conditions.
Motoo Kimura's Populationsgenetik der neutralen Evolution.


The average time of replacement of a dominant genotype in a population is the reciprocal mutation rate, $1/v$, and therefore independent of population size.

Is the Kimura scenario correct for virus populations?

Fixation of mutants in neutral evolution (Motoo Kimura, 1955)
Fitness landscapes showing error thresholds
STATIONARY MUTANT DISTRIBUTIONS AND EVOLUTIONARY OPTIMIZATION

PETER SCHUSTER and JÖRG SWETINA
Institut für theoretische Chemie und Strahlenchemie der Universität Wien, Währingerstraße 17, A 1090 Wien, Austria

Molecular evolution is modelled by erroneous replication of binary sequences. We show how the selection of two species of equal or almost equal selective value is influenced by its nearest neighbours in sequence space. In the case of perfect neutrality and sufficiently small error rates we find that the Hamming distance between the species determines selection. As the error rate increases the fitness parameters of neighbouring species become more and more important. In the case of almost neutral sequences we observe a critical replication accuracy at which a drastic change in the "quasispecies", in the stationary mutant distribution occurs. Thus, in frequently mutating populations fitness turns out to be an ensemble property rather than an attribute of the individual.

In addition we investigate the time dependence of the mean excess production as a function of initial conditions. Although it is optimized under most conditions, cases can be found which are characterized by decrease or non-monotonous change in mean excess productions.

1. Introduction. Recent data from populations of RNA viruses provided direct evidence for vast sequence heterogeneity (Domingo et al., 1987). The origin of this diversity is not yet completely known. It may be caused (a) by the low replication accuracy of the polymerizing enzyme, commonly a virus specific, (b) RNA dependent RNA synthetase, or it may be the result of a high degree of selective neutrality of polynucleotide sequences. Eventually, both factors contribute to the heterogeneity observed. Indeed, mutations occur much more frequently than previously assumed in microbiology. They are by no means rare events and hence, neither the methods of conventional population genetics (Ewens, 1979) nor the neutral theory (Kimura, 1983) can be applied to these virus populations. Selectively neutral variants may be close with respect to Hamming distance and then the commonly made assumption that the mutation backflow from the mutants to the wild type is negligible does not apply.

A kinetic theory of polynucleotide evolution which was developed during the past 15 years (Eigen, 1971; 1985; Eigen and Schuster, 1979; Eigen et al., 1987; Schuster, 1986; Schuster and Sigmund, 1985) treats correct replication and mutation as parallel reactions within one and the same reaction network.
Pairs of genotypes in neutral replication networks

Neutral network

$\lambda = 0.01$, $s = 367$

$d_H = 1$

$\lim_{p \to 0} x_1(p) = x_2(p) = 0.5$

Neutral network

$\lambda = 0.01$, $s = 877$

$d_H = 2$

$\lim_{p \to 0} x_1(p) = a$

$\lim_{p \to 0} x_2(p) = 1 - a$

$d_H \hat{=} 3$

random fixation in the sense of Motoo Kimura
A fitness landscape including neutrality
Neutral network: Individual sequences

\( n = 10, \sigma = 1.1, d = 1.0 \)
Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
...... AUUAACGAA ......
...... ACAUGCGCA ......
...... GCAUAACGAA ......
...... ACAUGCUAA ......
...... ACAUGCGGG ......
...... ACAUGCGGAA ......
...... ACGUACGAA ......
...... ACAUAGGAA ......
...... ACAUACGAA ......
```

```
...... ACAUGGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
```

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 1$. 

```
...... ACAUGCGAA ......
Neutral network: Individual sequences

\[ n = 10, \sigma = 1.1, d = 1.0 \]
Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i, X_j) = 2$. 
1. From Darwin to molecular biology
2. Selection in the test tube
3. Chemical kinetics of molecular evolution
4. Evolutionary biotechnology
5. The RNA model and neutrality
6. Simulation of molecular evolution
Evolution in silico

Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Phenylalanyl-tRNA as target structure

Structure of randomly chosen initial sequence

Phenylalanyl-tRNA as target structure
Replication rate constant (Fitness):
\[ f_k = \gamma / [\alpha + \Delta d_s^{(k)}] \]
\[ \Delta d_s^{(k)} = d_H(S_k, S_\tau) \]

Selection pressure:
The population size, 

\[ N = \# \text{ RNA molecules,} \]

is determined by the flux:

\[ N(t) \approx \bar{N} \pm \sqrt{N} \]

Mutation rate:
\[ p = 0.001 / \text{Nucleotide} \times \text{Replication} \]

The flow reactor as a device for studying the evolution of molecules \textit{in vitro} and \textit{in silico}. 
In silico optimization in the flow reactor: Evolutionary Trajectory
28 neutral point mutations during a long quasi-stationary epoch

Transition inducing point mutations change the molecular structure

Neutral point mutations leave the molecular structure unchanged

Neutral genotype evolution during phenotypic stasis
Randomly chosen initial structure

Phenylalanyl-tRNA as target structure
Evolutionary trajectory

Spreading of the population on neutral networks

Drift of the population center in sequence space

[Graph showing Hamming Distance to Target and Hamming Distance over Replications]
Spreading and evolution of a population on a neutral network: $t = 150$
Spreading and evolution of a population on a neutral network: $t = 170$
Spreading and evolution of a population on a neutral network: $t = 200$
Spreading and evolution of a population on a neutral network: $t = 350$
Spreading and evolution of a population on a neutral network: $t = 500$
Spreading and evolution of a population on a neutral network: $t = 650$
Spreading and evolution of a population on a neutral network: $t = 820$
Spreading and evolution of a population on a neutral network: $t = 825$
Spreading and evolution of a population on a neutral network: $t = 830$
Spreading and evolution of a population on a neutral network: $t = 835$
Spreading and evolution of a population on a neutral network: $t = 840$
Spreading and evolution of a population on a neutral network: $t = 845$
Spreading and evolution of a population on a neutral network: $t = 850$
Spreading and evolution of a population on a neutral network: $t = 855$
A sketch of optimization on neutral networks
Is the degree of neutrality in GC space much lower than in AUGC space?

Number | Mean Value | Variance | Std.Dev.
--- | --- | --- | ---
Total Hamming Distance: | 150000 | 11.647973 | 23.140715 | 4.810480
Nonzero Hamming Distance: | 99875 | 16.949991 | 30.757651 | 5.545958
Degree of Neutrality: | 50125 | **0.334167** | 0.006961 | 0.083434
Number of Structures: | 1000 | 52.31 | 85.30 | 9.24

1 ((((((((((((((((((((((......)))))))))))))))).)))))))))))))))................. 50125 **0.334167**
2 ................. 2856 0.019040
3 ................. 2799 0.018660
4 ................. 2417 0.016113
5 ................. 2265 0.015100
6 ................. 2233 0.014887

Number | Mean Value | Variance | Std.Dev.
--- | --- | --- | ---
Total Hamming Distance: | 150000 | 11.647973 | 23.140715 | 4.810480
Nonzero Hamming Distance: | 99875 | 16.949991 | 30.757651 | 5.545958
Degree of Neutrality: | 50125 | **0.334167** | 0.006961 | 0.083434
Number of Structures: | 1000 | 52.31 | 85.30 | 9.24

1 ((((((((((((((((((((((......)))))))))))))))).)))))))))))))))................. 50125 **0.334167**
2 ................. 2856 0.019040
3 ................. 2799 0.018660
4 ................. 2417 0.016113
5 ................. 2265 0.015100
6 ................. 2233 0.014887

Shadow – Surrounding of an RNA structure in shape space – **AUGC and GC alphabet**
Neutrality in evolution

Charles Darwin: „... neutrality might exist ...“

Motoo Kimura: „... neutrality is unaviodable and represents the main reason for changes in genotypes and leads to molecular phylogeny ...“

Current view: „... neutrality is essential for successful optimization on rugged landscapes ...“

Proposed view: „... neutrality provides the genetic reservoir in the rare and frequent mutation scenario ...“
Neutrality in molecular structures and its role in evolution:

• Neutrality is an essential feature in biopolymer structures at the resolution that is relevant for function.

• Neutrality manifests itself in the search for minimum free energy structures.

• Diversity in function despite neutrality in structures results from differences in suboptimal conformations and folding kinetics.

• Neutrality is indispensable for optimization and adaptation on rugged landscapes.
Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)
Projects No. 09942, 10578, 11065, 13093
13887, and 14898

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

Austrian Genome Research Program – GEN-AU

Siemens AG, Austria

Universität Wien and the Santa Fe Institute
Coworkers

Walter Fontana, Harvard Medical School, MA

Christian Forst, Christian Reidys, Los Alamos National Laboratory, NM

Peter Stadler, Bärbel Stadler, Universität Leipzig, GE

Jord Nagel, Kees Pleij, Universiteit Leiden, NL

Christoph Flamm, Ivo L.Hofacker, Andreas Svrček-Seiler, Universität Wien, AT

Kurt Grünberger, Michael Kospach, Andreas Wernitznig, Stefanie Widder, Michael Wolfinger, Stefan Wuchty, Universität Wien, AT

Stefan Bernhart, Jan Cupal, Lukas Endler, Ulrike Langhammer, Rainer Machine, Ulrike Mückstein, Hakim Tafer, Universität Wien, AT

Ulrike Göbel, Walter Grüner, Stefan Kopp, Jaqueline Weber, Institut für Molekulare Biotechnologie, Jena, GE
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks