Evolution and Design

Peter Schuster

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

Traunkirchner Gedankenexperimente

Traunkirchen, 13.09.2005
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks
Cardinal Christoph Schönborn:  *The New York Times*, July 07, 2005

... Any system of thought that denies or seeks to explain away the overwhelming evidence for design in biology is ideology, not science.

... Scientific theories that try to explain away the appearance of design as the result of “chance and necessity” are not scientific at all, but, as John Paul put it, an abdication of human intelligence.
1. History of evolutionary thinking
2. Probabilities in biology
3. Complex patterns from simple rules
4. Mechanisms of evolution
5. Origins of complexity - The eye
1. History of evolutionary thinking
2. Probabilities in biology
3. Complex patterns from simple rules
4. Mechanisms of evolution
5. Origins of complexity - The eye
Charles Darwin

Origin of evolutionary biology
1859

Gregor Mendel

Origin of genetics
1865
Charles Darwin

Origin of evolutionary biology 1859

First unification: Population genetics 1930

Ronald Fisher

Gregor Mendel

Origin of genetics 1865

‘Rediscovery’ 1900

Sewall Wright

JSB Haldane
Charles Darwin
Origin of evolutionary biology
1859
Gregor Mendel
Origin of genetics
1865
‘Rediscovery’ 1900
First unification: Population genetics 1930
Ernst Mayr
Synthetic or Neo-Darwinian theory
1940 - 1950
Theodosius Dobzhansky
Origin of evolutionary biology 1859

Origin of genetics 1865

‘Rediscovery’ 1900

Gregor Mendel

Friedrich Woehler

First unification: Population genetics 1930

Ernst Mayr

Synthetic or Neo-Darwinian theory 1940 - 1950

Theodosius Dobzhansky

Origin of biochemistry 1828
Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

‘Rediscovery’ 1900

Gregor Mendel
Friedrich Woehler

First unification: Population genetics 1930

Ernst Mayr

Synthetic or Neo-Darwinian theory
1940 - 1950

Theodosius Dobzhansky
Jacques Monod
James Watson and Francis Crick

Max Perutz

Biology of the 21st century

Friedrich Woehler

Origin of biochemistry
1828

Origin of molecular biology 1953

François Jacob
Max Perutz
John Kendrew
Charles Darwin
Origin of evolutionary biology 1859

Greger Mendel
Origin of genetics 1865

‘Rediscovery’ 1900

Friedrich Woehler
Origin of biochemistry 1828

First unification: Population genetics 1930

Ernst Mayr

Synthetic or Neo-Darwinian theory 1940 - 1950

Theodosius Dobzhansky

Jacques Monod

Max Perutz

James Watson and Francis Crick

Manfred Eigen

François Jacob

Sydney Brenner

John Kendrew

Biology of the 21st century

Biomathematics, bioinformatics, … , biophysics, biochemistry, … , molecular genetics, … , systems biology, biomedicine, macroscopic biology, evolutionary biology, sociobiology, anthropology, …
Ernst Mayr and others:

Can we explain the observations in biology without the assumption of a *causa finalis*?

The answer is *yes*, adaptation through variation and selection leads to the same result as rational design.

“Teleonomy replaces teleology”

Evolutionary biotechnology was able to prove this statement.
William of Ockham, ~1285 - 1349

Ockham's razor: "... plurality should not be assumed without necessity,"

or in modern English: "... keep it simple, unsophisticated, even stupid."

No concept, construct or variable should be used that is not required for the explanation of phenomena.

If we don't need a causa finalis, we have to dismiss it without replacement.
1. History of evolutionary thinking

2. Probabilities in biology

3. Complex patterns from simple rules

4. Mechanisms of evolution

5. Origins of complexity – The eye
Eugene Wigner’s argument applied to a bacterium:

All genomes have equal probability

5'-end GCGGATTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCTGTGTTCGAUCCACAGAATTG...GCACCA 3'-end

Alphabet size: 4

Chain length: \( \approx 1\ 000\ 000\ \text{nucleotides} \)

Number of possible genomes: \( 4^{1000000} \)

Probability to find a given bacterial genome:

\[
4^{-1000000} \approx 10^{-600000} = 0.000\ldots\ 001
\]

\( \left\llap{<} \ \ 600000 \ \ \right\rrap \)
Wigner’s paradox

The golf course landscape

Solution to Wigner’s paradox

The funnel landscape

Eugene Wigner’s argument revisited:

Every single point mutation leads to an improvement and is therefore selected

\[5'-\text{end GC}@\text{ATTAGCTAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCTGTGTTGCAUCCACAGAATC...GCAC}@3'-\text{end}\]

Alphabet size: 4

Chain length: \(\approx 1\,000\,000\) nucleotides

Length of longest path to the optimum: \(3 \times 1000000\)

Probability to find the optimal bacterial genome:

\[0.333.. \times 10^{-6} = 0.000000333..\]
Solution to Wigner’s paradox

An “all-roads-lead-to-Rome” landscape

The reconstructed folding landscape of a real biomolecule: “Lysozyme”

But (!) landscapes of evolution in nature and in the laboratory are unlike all the four examples shown here!
1. History of evolutionary thinking
2. Probabilities in biology
3. Complex patterns from simple rules
4. Mechanisms of evolution
5. Origins of complexity – The eye
John Horton Conway’s *Game-of-Life*

**Cell and neighborhood**
John Horton Conway's Game-of-Life

Cell and neighborhood

Populated cell: (1) each cell with one or no neighbors dies
(2) each cell with two or three neighbors survives
(3) each cell with four or more neighbors dies

Empty cell: (4) each empty cell with three neighbors becomes populated.
1. History of evolutionary thinking
2. Probabilities in biology
3. Complex patterns from simple rules
4. **Mechanisms of evolution**
5. Origins of complexity – The eye
Punctuated Equilibrium:
Evolution occurs through abrupt changes and not gradual.
Gradual change versus punctuated equilibrium in butterfly colors
Phyletic tree as pictured by the gradualists’ and the punctuated equilibrium approach
Falling meteorites:  
An example is the Chicxulub crater in Mexico dated 65 million years ago

<table>
<thead>
<tr>
<th></th>
<th>Generation time</th>
<th>Selection and adaptation 10 000 generations</th>
<th>Genetic drift in small populations 10⁶ generations</th>
<th>Genetic drift in large populations 10⁷ generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA molecules</td>
<td>10 sec</td>
<td>27.8 h = 1.16 d 6.94 d</td>
<td>115.7 d</td>
<td>3.17 a</td>
</tr>
<tr>
<td></td>
<td>1 min</td>
<td></td>
<td>1.90 a</td>
<td>19.01 a</td>
</tr>
<tr>
<td>Bacteria</td>
<td>20 min</td>
<td>138.9 d</td>
<td>38.03 a</td>
<td>380 a</td>
</tr>
<tr>
<td></td>
<td>10 h</td>
<td></td>
<td>1 140 a</td>
<td>11 408 a</td>
</tr>
<tr>
<td>Multicellular organisms</td>
<td>10 d</td>
<td>274 a</td>
<td>27 380 a</td>
<td>273 800 a</td>
</tr>
<tr>
<td></td>
<td>20 a</td>
<td>20 000 a</td>
<td>2 × 10⁷ a</td>
<td>2 × 10⁸ a</td>
</tr>
</tbody>
</table>

Time scales of evolutionary change
Bacterial Evolution


Serial transfer of Escherichia coli cultures in Petri dishes

1 day \( \Rightarrow \) 6.67 generations
1 month \( \Rightarrow \) 200 generations
1 year \( \Rightarrow \) 2400 generations
Epochal evolution of bacteria in serial transfer experiments under constant conditions


---

**Fig. 1.** Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (22). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these data (Table 1).

**Fig. 2.** Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (6, 7). The open symbols indicate the only two samples assigned to different steps by the cell size and fitness data.
Variation of genotypes in a bacterial serial transfer experiment

Innovation after 33 000 generations:

One out of 12 *Escherichia coli* colonies adapts to the environment and starts spontaneously to utilize citrate in the medium.
Evolution of RNA molecules based on Qβ phage


RNA sample

Stock solution: Qβ RNA-replicase, ATP, CTP, GTP and UTP, buffer

The serial transfer technique applied to RNA evolution *in vitro*
The increase in RNA production rate during a serial transfer experiment

Decrease in mean fitness due to quasispecies formation
Evolutionary design of RNA molecules


Y. Wang, R.R.Rando, Specific binding of aminoglycoside antibiotics to RNA. Chemistry & Biology 2 (1995), 281-290

An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules
The SELEX technique for the evolutionary preparation of aptamers
Formation of secondary structure of the tobramycin binding RNA aptamer with $K_D = 9 \text{ nM}$

The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel,
Computer simulation of RNA optimization

Walter Fontana and Peter Schuster,
Biophysical Chemistry 26:123-147, 1987

Walter Fontana, Wolfgang Schnabl, and

A computer model of evolutionary optimization

Walter Fontana and Peter Schuster
Institute für theoretische Chemie und Strahlenchemie der Universität Wien, Wiltenringstrasse 17, A-1090 Wien, Austria

Accepted 27 February 1987

Molecular evolution; Optimisation; Polyribonucleotide folding; Quasi-species; Selective value; Stochastic reaction kinetics

Molecular evolution is viewed as a typical combinatorial optimization problem. We analyse a chemical reaction model which consists of RNA replication occurring and point mutations together with hydrolytic degradation and the dilution flux of a flow reactor. The corresponding stochastic reaction network is implemented on a computer in order to investigate some basic features of evolutionary optimization dynamics. Characteristic features of real molecular systems are mimicked by folding binary sequences into underlying two-dimensional structures. Selective values are derived from these molecular 'phenotypes' by an evaluation procedure which assigns numerical values to different elements of the secondary structure. The fitness function obtained thereby contains nonthermal long-range interactions which are typical for real systems. The fitness landscape also reveals quite involved and bizarre local topologies which we consider also representative of polyribonucleotide replication in actually occurring systems. Optimization proceeds on an ensemble of sequences via mutations and natural selection. The strategy observed in the simulation experiments is fairly general and resembles closely a heuristic widely applied in operations research areas. Despite the relative smallness of the system – we study 2000 molecules of chain length $\sim 10$ in a typical simulation experiment – features typical for the evolution of real populations are observed as there are error thresholds for replication, evolutionary steps and quasistationary sequence distributions. The relative importance of selectively neutral or almost neutral variants is discussed quantitatively. Four characteristic ensemble properties, namely the distribution, ensemble correlation, mean Hamming distance and diversity of the population, are computed and checked for their sensitivity in recording major optimization events during the simulation.

I. Molecular evolution and optimization

Conventional population genetics treats mutation as an external stochastic source. Moreover, mutations are considered as very rare events. In the absence of genetic recombination populations of haploid organisms are expected to be usually heterogeneous. Experimental evidence on viral and bacterial populations is available now and it contradicts these expectations. Mutations appear much more frequently than was originally assumed.

Dedicated to Professor Maxted Eigen on the occasion of his 80th birthday.

Correspondence address: F. Schuster, Institut für theoretische Chemie und Strahlenchemie der Universität Wien, Wiltenringstrasse 17, A-1090 Wien, Austria.

0031-4622/87/$03.50 © 1987 Elsevier Science Publishers B.V. (Biomedical Division)
Fig. 1. The reaction network. Synthesis on template $I_k$ proceeds with the rate constant $d_k$ and leads with frequency $Q_{ik}$ to a new template $I_k$ preserving the old copy. Materials A needed for polymerization are assumed to be buffered. Degradation to waste products B occurs with rate $d_k$ and a controlled unspecific flux $\Phi(t)$ removes templates from the system.

$k = 1, 2, \ldots, n$

Fig. 2. The evolution reactor. This kind of flow reactor consists of a reaction vessel which allows for temperature and pressure control. Its walls are impermeable to polynucleotides. Energy rich material is poured from the environment into the reactor. The degradation products are removed steadily. Material transport is adjusted in such a way that the concentration of monomers is constant in the reactor. A dilution flux $\Phi$ is installed in order to remove excess of polynucleotides produced by replication. Thus, the sum of the numbers of individual particles $\sum_{i}X_i(t) = N(t)$ may be controlled by the flux $\Phi$. Under "constant organization" $\Phi$ is adjusted such that $N(t) = \Theta$ is essentially constant. By this we indicate that fluctuations with standard deviation $\sigma = \sqrt{N}$ occur regularly. The regulation of $\Phi$ requires internal control, which can be achieved by logistic coupling.
Evolution in silico


Institut für Theoretische Chemie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria, Sam Schuster, Institute for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.

Continuity in Evolution: On the Nature of Transitions
Walter Fontana and Peter Schuster

To distinguish continuous from discontinuous evolutionary change, a relation of nearness between genotypes is needed. Such a relation is based on the probability of one phenotype being accessible from another through changes in the genotype. This nearness relation is exemplified by calculating the shape neighborhood of a transfer RNA secondary structure and provides a characterization of discontinuous shape transformations in RNA. The replication of simulating and mutating RNA populations under selection shows that sudden adaptive progress coincides mostly, but not always, with discontinuous shape transformations. The nature of these transformations illuminates the key role of neutral genetic drift in their realization.

A much-debated issue in evolutionary biology concerns the extent to which the history of life has proceeded gradually or has been punctuated by discontinuous transitions at the origin of new species (1). Our goal is to make the notion of a discontinuous transition more precise and to understand how it arises in a model of evolutionary adaptation.

We focus on the narrow domain of RNA secondary structures, which is currently the simplest computationally tractable, yet realistic phenotype (2). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a class of gene products, which are the product of the DNA tertiary structure at atomic resolution. Yet, secondary structures are empirically well defined and their biophysical and biochemical importance from being a scaffold for the tertiary structure. For the sake of brevity, we shall refer to secondary structures as "shapes". RNA combines in a single molecule both genotype (replicable sequence) and phenotype (selectable shape), making it ideally suited for in vitro evolution experiments (3, 4).

To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (5, 6). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (4). Although in the experiment the evolutionary end product was unknown, we thought of its shape as being specified implicitly by the imposed selection criterion. Because our intent is to study evolutionary histories rather than end products, we defined a target shape in advance and assumed the replication rate of a sequence to be a function of the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

An instance representing in its qualitative features all the simulations performed is shown in Fig. 1A. Starting with identical sequences folding into a random shape, the simulation was stopped when the population became dominated by the target shape. The black curve traces the average distance to the target (inversely related to fitness) in the population against time. Aside from a short initial phase, the entire history is dominated by that step, that is, the period of no apparent adaptive progress, interrupted by sudden approaches toward the target shape (7). However, the dominant shape in the population not only changes at these marked events but undergoes several fitness-neutral transformations during the periods of no apparent progress. Although the black curve traces the average distance to the target (inversely related to fitness) in the population against time, it is entirely unclear when and on the basis of what, the series of successive phenotypes itself is called continuous or discontinuous.

A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally organized into a space because point mutations induce a canonical neighborhood. The neighborhood of a sequence consists of all its one-error mutants. The problem is how to organize the set of possible shapes into a space. The issue arises because, in contrast to sequences, there are
Replication rate constant:
\[ f_k = \gamma / [\alpha + \Delta d_{S}^{(k)}] \]
\[ \Delta d_{S}^{(k)} = d_H(S_k, S_\tau) \]

Selection constraint:

# RNA molecules is controlled by the flow
\[ N(t) \approx \bar{N} \pm \sqrt{N} \]

The flowreactor as a device for studies of evolution in vitro and 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
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.

Charles Darwin, Origin of species (1859)
Motoo Kimura’s Population genetics of neutral evolution.


Mount Fuji

Example of a smooth landscape on Earth
Examples of rugged landscapes on Earth
Evolutionary optimization in absence of neutral paths in sequence space
Evolutionary optimization including neutral paths in sequence space
Example of a landscape on Earth with ‘neutral’ ridges and plateaus
Conformational and mutational landscapes of biomolecules as well as fitness landscapes of evolutionary biology are **rugged**.

**Adaptive or non-descending walks** on rugged landscapes end commonly at one of the low lying local maxima.

Selective neutrality in the form of **neutral networks** plays an active role in evolutionary optimization and enables populations to reach high local maxima or even the global optimum.
1. History of evolutionary thinking
2. Probabilities in biology
3. Complex patterns from simple rules
4. Mechanisms of evolution
5. Origins of complexity – The eye
Molecular genetics shows that the development of all different forms of eyes have the same evolutionary origin, which can be traced back to a simple form of light-sensitivity found already in primitive bacteria.

Walter J. Gehring, The genetic control of eye development and its implications for the evolution of the various eye-types. Zoology 104 (2001), 171-183

Fig. 1. Different types of eyes. (A) Camera-type eye from the Lemur Propithecus verreauxi. (B) Compound eye of the praying Mantis. (C) Camera-type eye from the Cephalopod Sepia erostata. (D) Mirror eye from the clam Chlamys nobilis. (Courtesy of Dr. Kazuto Kato; photographs kindly provided by Masahiro Iijima, Susumu Yamaguchi and Isamu Soyama.)
Fig. 1. Schematic diagram of cephalopod eye development (Left) and vertebrate eye development (Right) as explained in more detail in refs. 7 and 8. Development proceeds from top to bottom. Even though the adult structures are fairly similar, excepting certain obvious features such as the placement of the photoreceptors and lentigencells, the development is very different. The cephalopod eye forms from an epidermal placode through a series of successive infoldings, while the vertebrate eye emerges from the neural plate and induces the overlying epidermis to form the lens.
Linear chain

Network

Processing of information in cascades and networks
Analysis of nodes and links in a step by step evolved network
The reaction network of cellular metabolism published by Boehringer-Ingelheim.
The citric acid or Krebs cycle (enlarged from previous slide).

Web-Page for further information:

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