Self-Organization and Evolution

Peter Schuster
Institut für Theoretische Chemie und Molekulare Strukturbioologie der Universität Wien

Wissenschaftliche Gesellschaft: Dynamik – Komplexität – menschliche Systeme
AKH Wien, 12.06.2002
Equilibrium thermodynamics is based on two major statements:

1. The energy of the universe is a constant (first law).
2. The entropy of the universe never decreases (second law).

Carnot, Mayer, Joule, Helmholtz, Clausius, ……

Time fluctuations around equilibrium

Approach towards equilibrium

Spontaneous processes $\Delta S > 0$

Enlarged scale

Fluctuations around equilibrium

$S_{\text{max}}$

Entropy and fluctuations at equilibrium

$(d^2S)_{U,V,\text{equil}} < 0$

Entropy and time
Self-organization is spontaneous creation of order.

Entropy is equivalent to disorder. Hence there is no spontaneous creation of order at equilibrium.

Self-organization requires export of entropy to an environment which is almost always tantamount to an energy flux or transport of matter in an open system.

Entropy production and self-organization in open systems
Four examples of self-organization and spontaneous creation of order

• Hydrodynamic pattern formation in the atmosphere of Jupiter

• Fractal pattern in the solution manifold of mathematical equations

• Chaotic dynamics in model equations for atmospheric flow

• Pattern formation in chemical reactions

Examples of self-organization and pattern formation
Jupiter: Observation of the gigantic vortex

Computer simulation of the gigantic vortex on Jupiter

View from south pole

Particles turning counterclockwise

Particles turning clockwise

Jupiter: Computer simulation of the giant vortex

Mandelbrot set

$$z \rightarrow z^2 + c$$

with \( z = x + i y \)

The Mandelbrot set as an example of fractal patterns in mathematics

Mandelbrot set

\[ z \rightarrow z^2 + c \]

with \( z = x + iy \)

The Mandelbrot set as an example of fractal patterns in mathematics: Enlargement no.1

The Mandelbrot set as an example of fractal patterns in mathematics: Enlargement no.2

Mandelbrot set

\[ z \rightarrow z^2 + c \]

with \( z = x + i \, y \)

The Mandelbrot set as an example of fractal patterns in mathematics: Enlargement no.3

Lorenz attractor

\[ \frac{dx}{dt} = \sigma (y - x) \]
\[ \frac{dy}{dt} = \rho x - y - xz \]
\[ \frac{dz}{dt} = \beta z + xy \]

A trajectory of the Lorenz attractor in the chaotic regime
Entropy changes in different thermodynamic systems with chemical reactions
Reactions in the continuously stirred tank reactor (CSTR)
Reversible first order reaction in the flow reactor
Autocatalytic second order reaction in the flow reactor

\[ A + B \rightleftharpoons 2B \]
Autocatalytic second order and uncatalyzed reaction in the flow reactor
Autocatalytic third order reaction in the flow reactor
Autocatalytic third order and uncatalyzed reaction in the flow reactor
Autocatalytic third order reactions

Direct, $A + 2X \rightarrow 3X$, or hidden in the reaction mechanism (Belousow-Zhabotinskii reaction).

- Multiple steady states
- Oscillations in homogeneous solution
- Deterministic chaos
- Turing patterns

Spatiotemporal patterns (spirals)

Deterministic chaos in space and time

Pattern formation in autocatalytic third order reactions

Formation of target waves and spirals in the Belousov-Zhabotinskii reaction

Winding number:

\[
\text{number of left-handed spirals} - \text{number of right-handed spirals}
\]

Target waves and spirals in the Belousov-Zhabotinskii reaction

Autocatalytic second order reactions

Direct, \( A + I \rightarrow 2I \), or hidden in the reaction mechanism

Selection of molecular or organismic species competing for common sources

Chemical self-enhancement

Combustion and chemistry of flames

Selection of laser modes

Autocatalytic second order reaction as basis for selection processes.
The autocatalytic step is formally equivalent to replication or reproduction.
Replication in the flow reactor

Flow rate $r = \frac{1}{t_{R}}$

Selection in the flow reactor: Reversible replication reactions

$$\begin{align*}
A + I_1 & \xleftrightarrow{d_1} 2I_1 \\
A + I_2 & \xleftrightarrow{d_2} 2I_2 \\
A + I_3 & \xleftrightarrow{d_3} 2I_3 \\
A + I_4 & \xleftrightarrow{d_4} 2I_4 \\
A + I_5 & \xleftrightarrow{d_5} 2I_5
\end{align*}$$

$k_1 > k_2 > k_3 > k_4 > k_5$
Flow rate \( r = \frac{1}{t_R} \)

Selection in the flow reactor: Irreversible replication reactions

\[
\begin{align*}
\text{A + I}_1 & \xrightarrow{f_1} 2 \text{I}_1 \\
\text{A + I}_2 & \xrightarrow{f_2} 2 \text{I}_2 \\
\text{A + I}_3 & \xrightarrow{f_3} 2 \text{I}_3 \\
\text{A + I}_4 & \xrightarrow{f_4} 2 \text{I}_4 \\
\text{A + I}_5 & \xrightarrow{f_5} 2 \text{I}_5 \\
\end{align*}
\]

\( f_1 > f_2 > f_3 > f_4 > f_5 \)
dx_j / dt = f_j x_j - x_j \Phi = (f_j - \Phi) x_j

\Phi = \Sigma_i f_i x_i ; \quad \Sigma_i x_i = 1 ; \quad i,j = 1,2,\ldots,n

[A] = a = constant

f_m = \max \{f_j; j=1,2,\ldots,n\}

x_m(t) \to 1 \text{ for } t \to \infty

s = (f_{m+1} - f_m) / f_m

succession of temporarily fittest variants:

m \to m+1 \to \ldots

Selection of the “fittest” or fastest replicating species \( I_m \)
Selection of advantageous mutants in populations of N = 10 000 individuals
Thermodynamics of isolated systems: Entropy is a non-decreasing state function

Second law \( S \rightarrow S_{\text{max}} \)

Valid in the limit of infinite time, \( \lim_{t \to \infty} \).

Evolution of Populations: Mean fitness is a non-decreasing function

Ronald Fisher’s conjecture \( f = \frac{S_k x_k(t) f_k}{S_k x_k(t)} \)

Optimization heuristics in the sense that it only almost always true and the process not reach the optimum in finite times.

is need
Combinatorial diversity of heteropolymers illustrated by means of an RNA aptamer that binds to the antibiotic tobramycin

$4^{27} = 1.801 \times 10^{16}$ possible different sequences

Combinatorial diversity of sequences: $N = 4^7$
Complementary replication as the simplest copying mechanism of RNA.
Mutations represent the mechanism of variation in nucleic acids
\[
\frac{dx_j}{dt} = \sum_i f_i Q_{ji} x_i - x_j \Phi
\]

\[
\Phi = \sum_i f_i x_i \quad \Sigma_i x_i = 1 \quad \Sigma_i Q_{ij} = 1
\]

\[
Q_{ij} = (1-p)^{n-d(i,j)} p^{d(i,j)}
\]

p .......... Error rate per digit

d(i,j) .... Hamming distance between I_i and I_j

\[
[A] = a = \text{constant}
\]

Chemical kinetics of replication and mutation
The molecular quasispecies in sequence space
The molecular quasispecies and mutations producing new variants
Ronald Fisher’s conjecture of **optimization of mean fitness in populations** does not hold in general for **replication-mutation systems**: In general evolutionary dynamics the mean fitness of populations may also decrease monotonously or even go through a maximum or minimum. It does also not hold in general for **recombination of many alleles** and general multi-locus systems in population genetics.

**Optimization of fitness** is, nevertheless, fulfilled in most cases, and can be understood as a useful heuristic.
Optimization of RNA molecules in silico


Three-dimensional structure of phenylalanyl-transfer-RNA
Sequence:
GCGGAUUAGCUAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCUGUGTPCGAUCACAGAUAUCGCACCA

Secondary Structure:

Symbolic Notation:
5'-End ((((((...((........)))))(((........))))....(((........)))).... 3'-End

Definition and formation of the secondary structure of phenylalanyl-tRNA
Evolutionary dynamics including molecular phenotypes
Criterion of Minimum Free Energy

UUUAGCCAGCGCGAGUCGUGCGGACGGGGUUAUCUCUGUCGGGCUAGGGCGC
GUGAGCGCGGGGCACAGUUUCUCAAGGAUGUAAGUUUUUGCCGUUUUAUCUGG
UUAGCGAGAGAGGAGGCUUCUAGACCCAGCUCUCUGGGUCGUUGCUGAUGCG
CAUUGGUGCUAUAGAUUUAGGCCUGUAUUGCAGUAGCGAUCAGUGUCGG
GUAGGCCUCUUGACAUAAGAUUUUUCCAUGGUGGAGAUGGCCAUUGCAG

Sequence Space

Shape Space
The flowreactor as a device for studies of evolution *in vitro* and *in silico*
In silico optimization in the flow reactor: Trajectory
Endconformation of optimization
Average structure distance to target DdS

Evolutionary trajectory

Relay steps

Number of relay step

Reconstruction of the last step 43 Å  44
Reconstruction of last-but-one step 42 Å 43 (Å 44)
Reconstruction of step 41 Å 42 (Å 43 Å 44)
Reconstruction of step 40 Å 41 (Å 42 Å 43 Å 44)
Evolutionary process

Reconstruction of the relay series
In silico optimization in the flow reactor: Trajectory and relay steps
*In silico* optimization in the flow reactor: Uninterrupted presence
*In silico* optimization in the flow reactor: Main transitions
Main transition leading to clover leaf

Reconstruction of a main transitions $36 \leftrightarrow 37 \leftrightarrow 38$
Main transition leading to clover leaf

Evolutionary process

Final reconstruction 36 Å 44
In silico optimization in the flow reactor

Average structure distance to target $d_s$

Time (arbitrary units)

Relay steps
Main transitions
Uninterrupted presence
Evolutionary trajectory

In silico optimization in the flow reactor
Variation in genotype space during optimization of phenotypes
The number of main transitions or evolutionary innovations is constant.
Three important steps in the formation of the tRNA clover leaf from a randomly chosen initial structure corresponding to three main transitions.
„...Variations neither useful not 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)
Evolution in genotype space sketched as a non-descending walk in a fitness landscape.
Evolution of RNA molecules based on Qβ phage


The serial transfer technique applied to RNA evolution \textit{in vitro}
Reproduction of the original figure of the serial transfer experiment with Qβ RNA


---

**Fig. 9.** Serial transfer experiment. Each 0.25 ml standard reaction mixture contained 40 μg of Qβ replicase and ^32P-UTP. The first reaction (0 transfer) was initiated by the addition of 0.2 μg ts-1 (temperature-sensitive RNA) and incubated at 35 °C for 20 min, whereupon 0.02 ml was drawn for counting and 0.02 ml was used to prime the second reaction (first transfer), and so on. After the first 13 reactions, the incubation periods were reduced to 15 min (transfers 14-29). Transfers 30–38 were incubated for 10 min. Transfers 39–52 were incubated for 7 min, and transfers 53–74 were incubated for 5 min. The arrows above certain transfers (0, 8, 14, 29, 37, 53, and 73) indicate where 0.001–0.1 ml of product was removed and used to prime reactions for sedimentation analysis on sucrose. The inset examines both infectious and total RNA. The results show that biologically competent RNA censuses to appear after the 4th transfer (Mills *et al*. 1967).
Decrease in mean fitness due to quasispecies formation

The increase in RNA production rate during a serial transfer experiment
Bacterial Evolution


Epochal evolution of bacteria in serial transfer experiments under constant conditions

Variation of genotypes in a bacterial serial transfer experiment

Evolutionary design of RNA molecules


Selection cycle used in applied molecular evolution to design molecules with predefined properties.
The SELEX technique for the evolutionary design of aptamers
Formation of secondary structure of the tobramycin binding RNA aptamer

The three-dimensional structure of the tobramycin aptamer complex

A ribozyme switch

Two ribozymes of chain lengths n = 88 nucleotides: An artificial ligase (A) and a natural cleavage ribozyme of hepatitis-d-virus (B)
The sequence at the **intersection**:

An RNA molecules which is 88 nucleotides long and can form both structures
THEOREM 5. INTERSECTION-THEOREM. Let $s$ and $s'$ be arbitrary secondary structures and $C[s], C[s']$ their corresponding compatible sequences. Then,

$$C[s] \cap C[s'] \neq \emptyset.$$ 

Proof. Suppose that the alphabet admits only the complementary base pair $[XY]$ and we ask for a sequence $x$ compatible to both $s$ and $s'$. Then $(s, s') \cong D_n$ operates on the set of all positions $(x_1, \ldots, x_n)$. Since we have the operation of a dihedral group, the orbits are either cycles or chains and the cycles have even order. A constraint for the sequence compatible to both structures appears only in the cycles where the choice of bases is not independent. It remains to be shown that there is a valid choice of bases for each cycle, which is obvious since these have even order. Therefore, it suffices to choose an alternating sequence of the pairing partners $X$ and $Y$. Thus, there are at least two different choices for the first base in the orbit.

Remark. A generalization of the statement of theorem 5 to three different structures is false.

Random graph theory is used to model and analyse the relationships between sequences and secondary structures of RNA molecules, which are understood as mappings from sequence space into shape space. These maps are non-invertible since there are always many orders of magnitude more sequences than structures. Sequences folding into identical structures form neutral networks. A neutral network is embedded in the set of sequences that are compatible with the given structure. Networks are modeled as graphs and constructed by random choice of vertices from the space of compatible sequences. The theory characterizes neutral networks by the mean fraction of neutral neighbors ($\lambda$). The networks are connected and percolate sequence space if the fraction of neutral nearest neighbors exceeds a threshold value ($\lambda > \lambda^*$). Below threshold ($\lambda < \lambda^*$), the networks are partitioned into a largest “giant” component and several smaller components. Structures are classified as “common” or “rare” according to the sizes of their pre-images, i.e. according to the fractions of sequences folding into them. The neutral networks of any pair of two different common structures almost touch each other, and, as expressed by the conjecture of shape space covering sequences folding into almost all common structures, can be found in a small ball of an arbitrary location in sequence space. The results from random graph theory are compared to data obtained by folding large samples of RNA sequences. Differences are explained in terms of specific features of RNA molecular structures. © 1997 Society for Mathematical Biology
Two neutral walks through sequence space with conservation of structure and catalytic activity
Sequence of mutants from the intersection to both reference ribozymes
From sequences to shapes and back: a case study in RNA secondary structures

PETER SCHUSTER\textsuperscript{1,2,3}, WALTER FONTANA\textsuperscript{2}, PETER F. STADLER\textsuperscript{2,3} and IVO L. HOFACKER\textsuperscript{2}

\textsuperscript{1} Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, PF 100813, D-07708 Jena, Germany
\textsuperscript{2} Institut für Theoretische Chemie, Universität Wien, Austria
\textsuperscript{3} Santa Fe Institute, Santa Fe, U.S.A.

SUMMARY

RNA folding is viewed here as a map assigning secondary structures to sequences. At fixed chain length the number of sequences far exceeds the number of structures. Frequencies of structures are highly non-uniform and follow a generalized form of Zipf's law: we find relatively few common and many rare ones. By using an algorithm for inverse folding, we show that sequences sharing the same structure are distributed randomly over sequence space. All common structures can be accessed from an arbitrary sequence by a number of mutations much smaller than the chain length. The sequence space is percolated by extensive neutral networks connecting nearest neighbours folding into identical structures. Implications for evolutionary adaptation and for applied molecular evolution are evident: finding a particular structure by mutation and selection is much simpler than expected and, even if catalytic activity should turn out to be sparse in the space of RNA structures, it can hardly be missed by evolutionary processes.

Figure 4. Neutral paths. A neutral path is defined by a series of nearest neighbour sequences that fold into identical structures. Two classes of nearest neighbours are admitted: neighbours of Hamming distance 1, which are obtained by single base exchanges in unpaired stretches of the structure, and neighbours of Hamming distance 2, resulting from base pair exchanges in stacks. Two probability densities of Hamming distances are shown that were obtained by searching for neutral paths in sequence space: (i) an upper bound for the closest approach of trial and target sequences (open circles) obtained as endpoints of neutral paths approaching the target from a random trial sequence (185 targets and 100 trials for each were used); (ii) a lower bound for the closest approach of trial and target sequences (open diamonds) derived from secondary structure statistics (Fontana \textit{et al.} 1993a; see this paper, §4); and (iii) longest distances between the reference and the endpoints of monotonously diverging neutral paths (filled circles) (500 reference sequences were used).

Reference for postulation and \textit{in silico} verification of \textit{neutral networks}
Coworkers

Walter Fontana, Santa Fe Institute, NM

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

Peter Stadler, Universität Wien, AT
   Ivo L.Hofacker
   Christoph Flamm

Bärbel Stadler, Andreas Wernitznig, Universität Wien, AT
Michael Kospach, Ulrike Mückstein, Stefanie Widder, Stefan Wuchty
Jan Cupal, Kurt Grünberger, Andreas Svrček-Seiler

Ulrike Göbel, Institut für Molekulare Biotechnologie, Jena, GE
Walter Grüner, Stefan Kopp, Jaqueline Weber