

Self-Organization and Evolution

Peter Schuster

Institut für Theoretische Chemie und Molekulare Strukturbiologie
der Universität Wien

Wissenschaftliche Gesellschaft: Dynamik – Komplexität –
menschliche Systeme

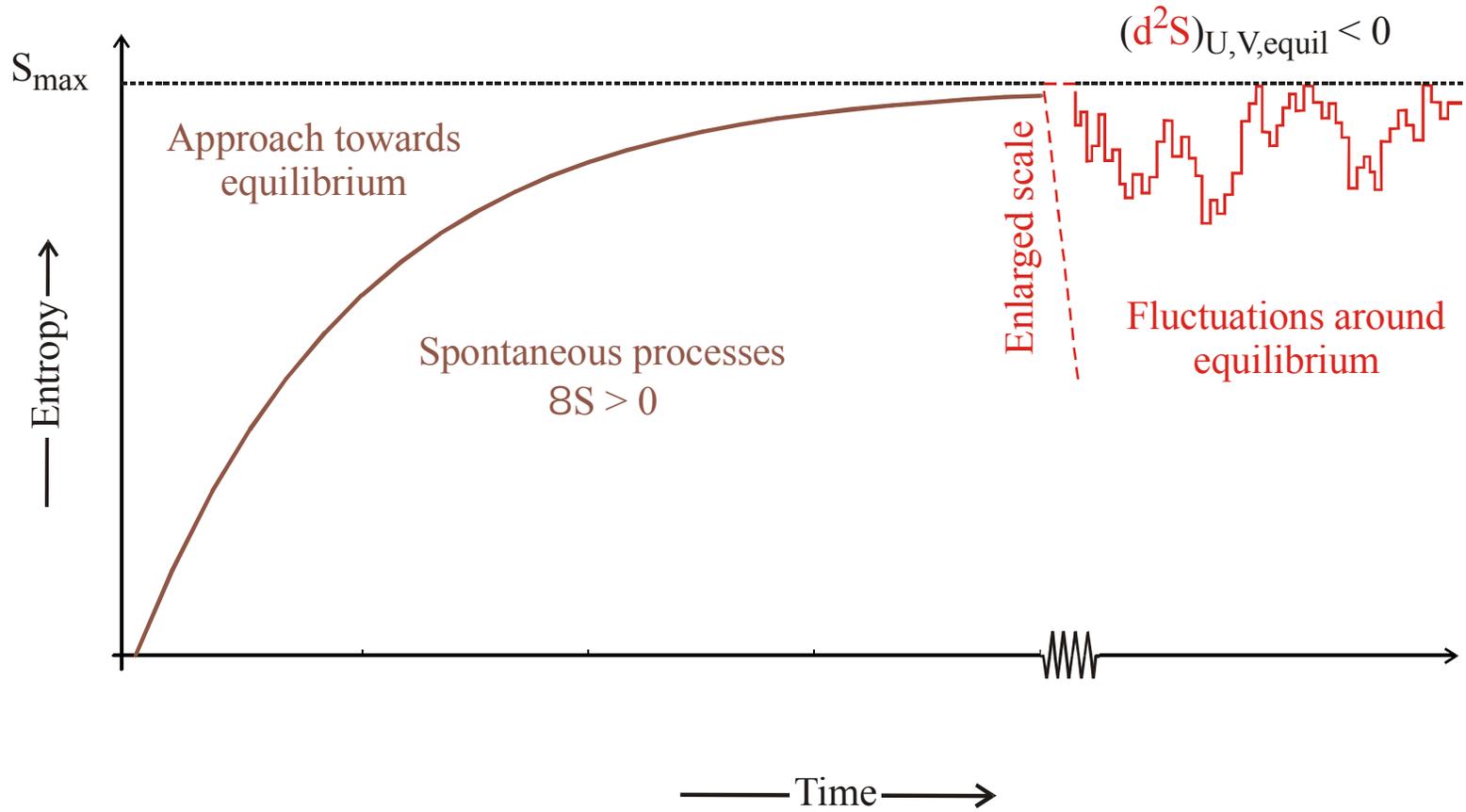
AKH Wien, 12.06.2002

Equilibrium thermodynamics is based on two major statements:

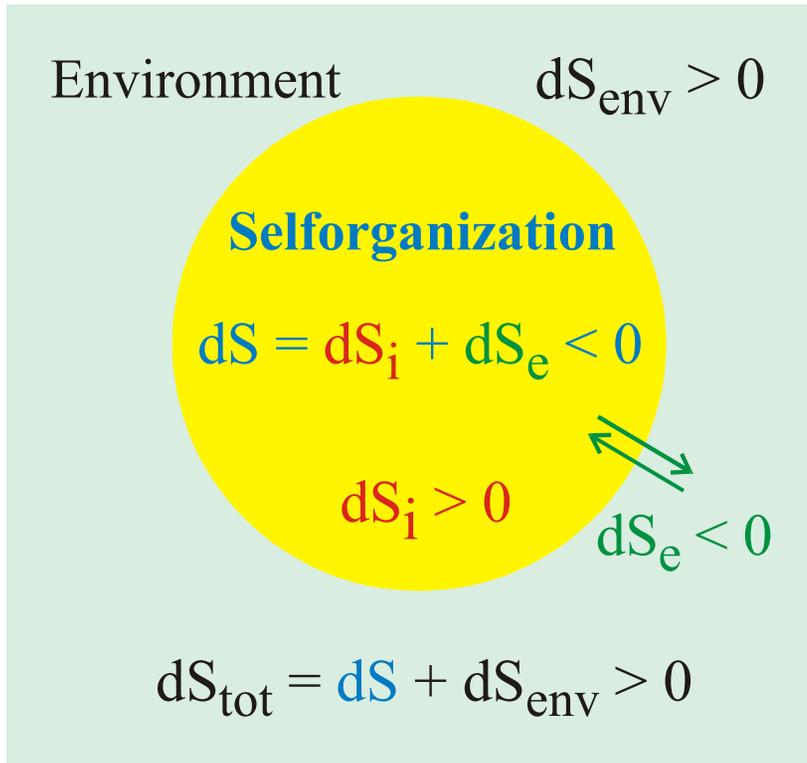
1. The energy of the universe is a constant (first law).
2. The entropy of the universe never decreases (second law).

Carnot, Mayer, Joule, Helmholtz, Clausius,

D.Jou, J.Casas-Vázquez, G.Lebon, *Extended Irreversible Thermodynamics*, 1996



Entropy and fluctuations at equilibrium



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

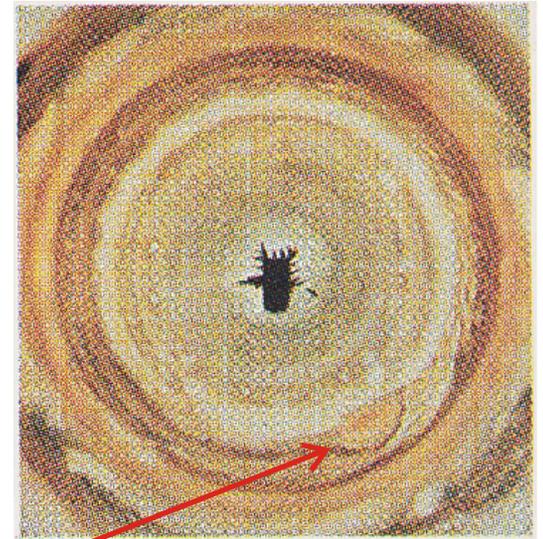
Four examples of self-organization and spontaneous creation of order

- 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



South pole



View from south pole

Red spot



Jupiter: Observation of the gigantic vortex

Picture taken from James Gleick, *Chaos*. Penguin Books, New York, 1988

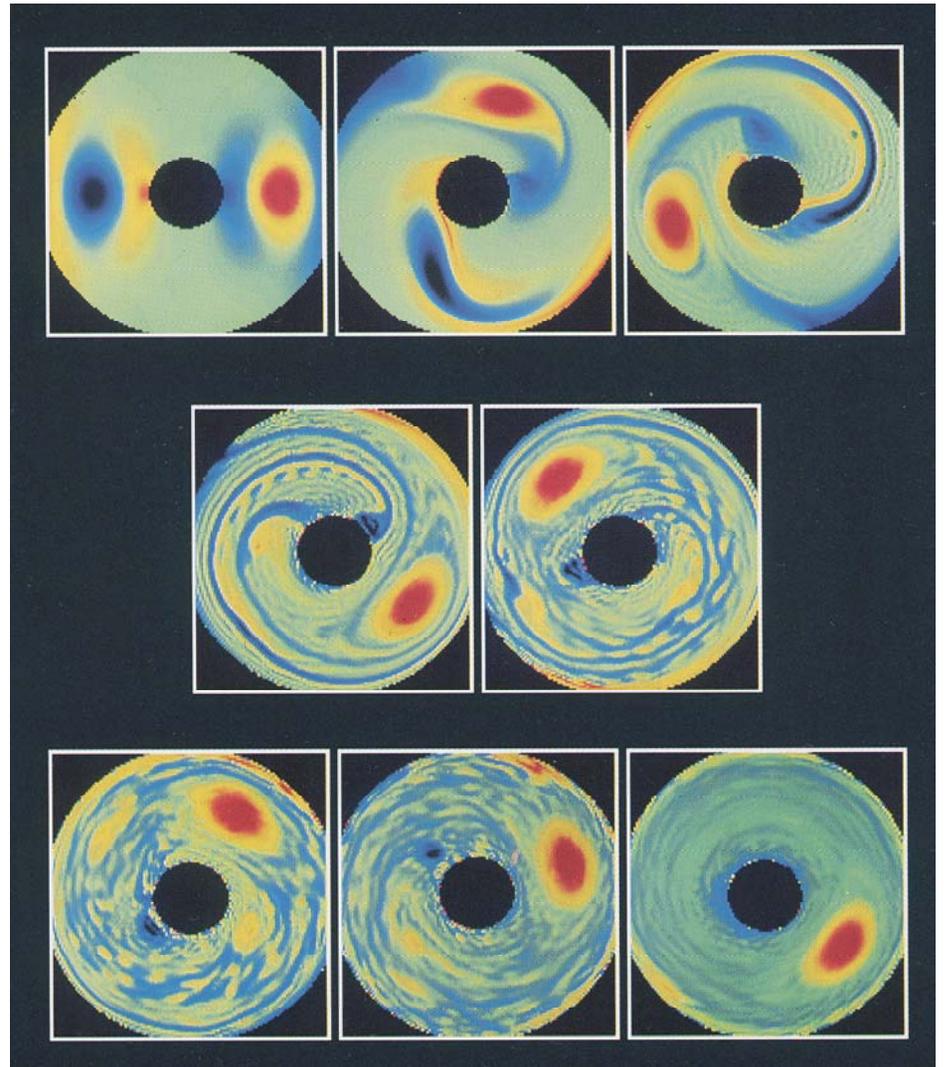
Computer simulation of the gigantic vortex on Jupiter

View from south pole

Particles turning
counterclockwise



Particles turning
clockwise



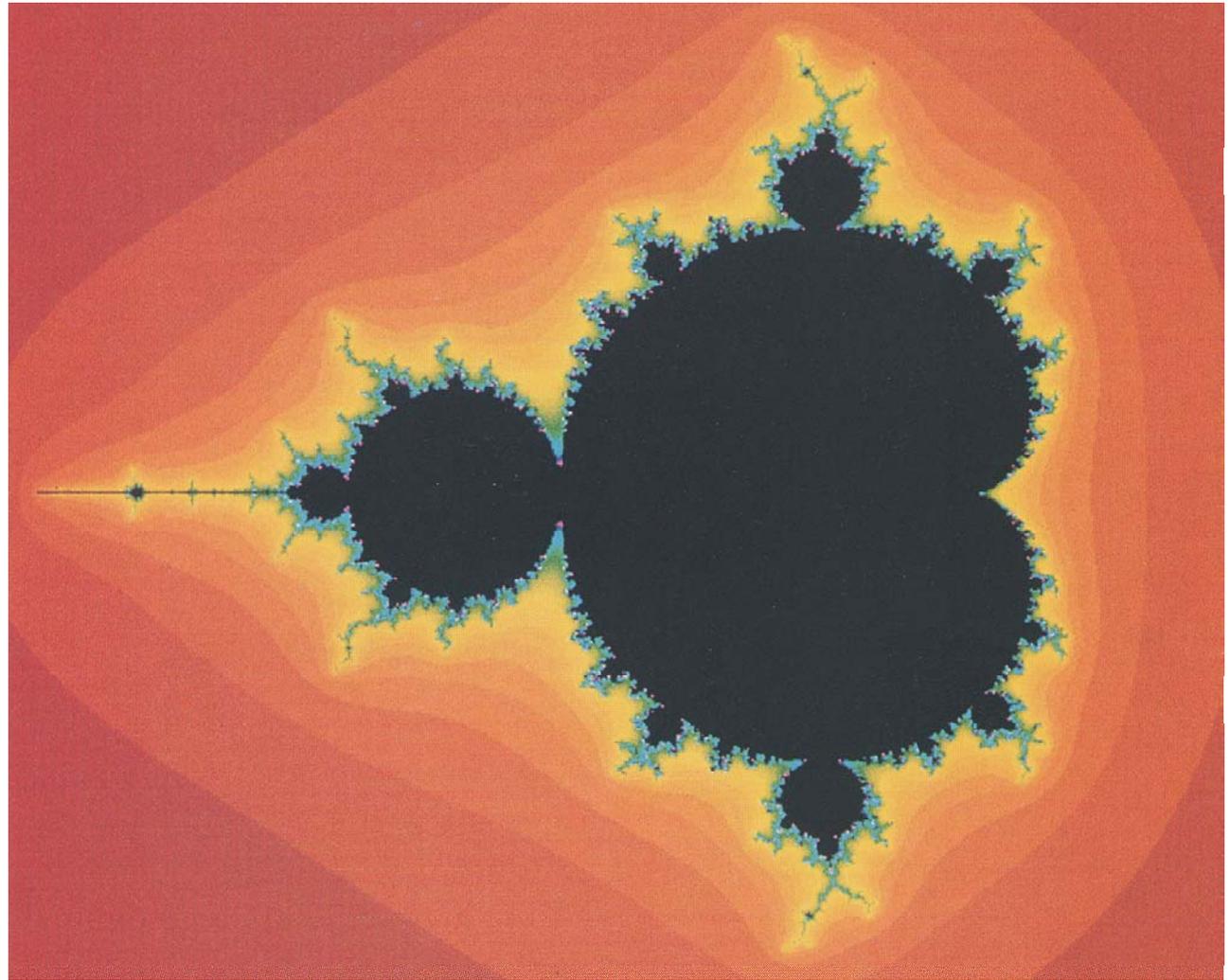
Jupiter: Computer simulation of the giant vortex

Philip Marcus, 1980. Picture taken from James Gleick, *Chaos*. Penguin Books, New York, 1988

Mandelbrot set

$$z \checkmark z^2 + c$$

with $z = x + iy$



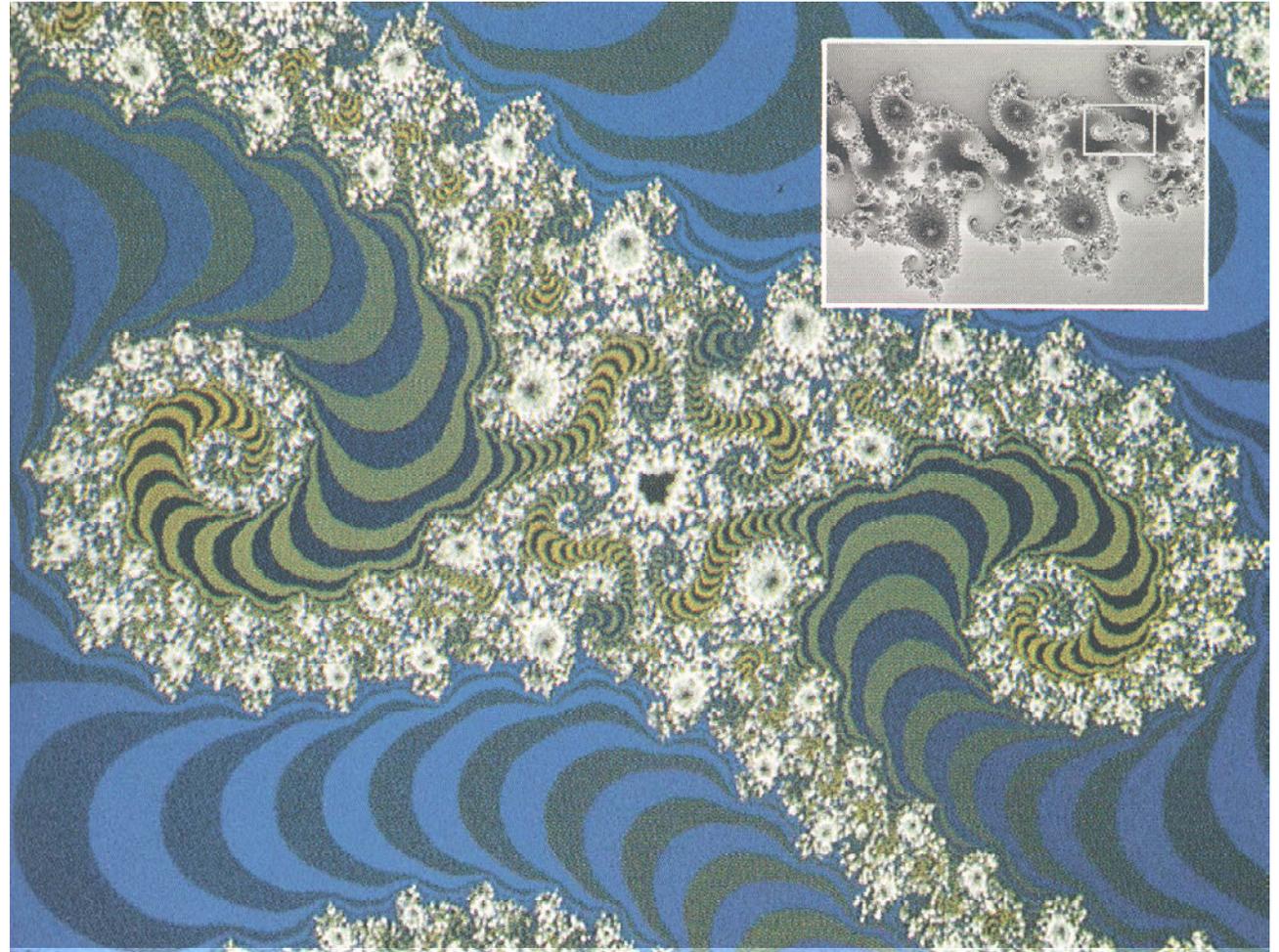
The Mandelbrot set as an example of fractal patterns in mathematics

Picture taken from James Gleick, *Chaos*. Penguin Books, New York, 1988

Mandelbrot set

$$z \dot{\sim} z^2 + c$$

with $z = x + iy$



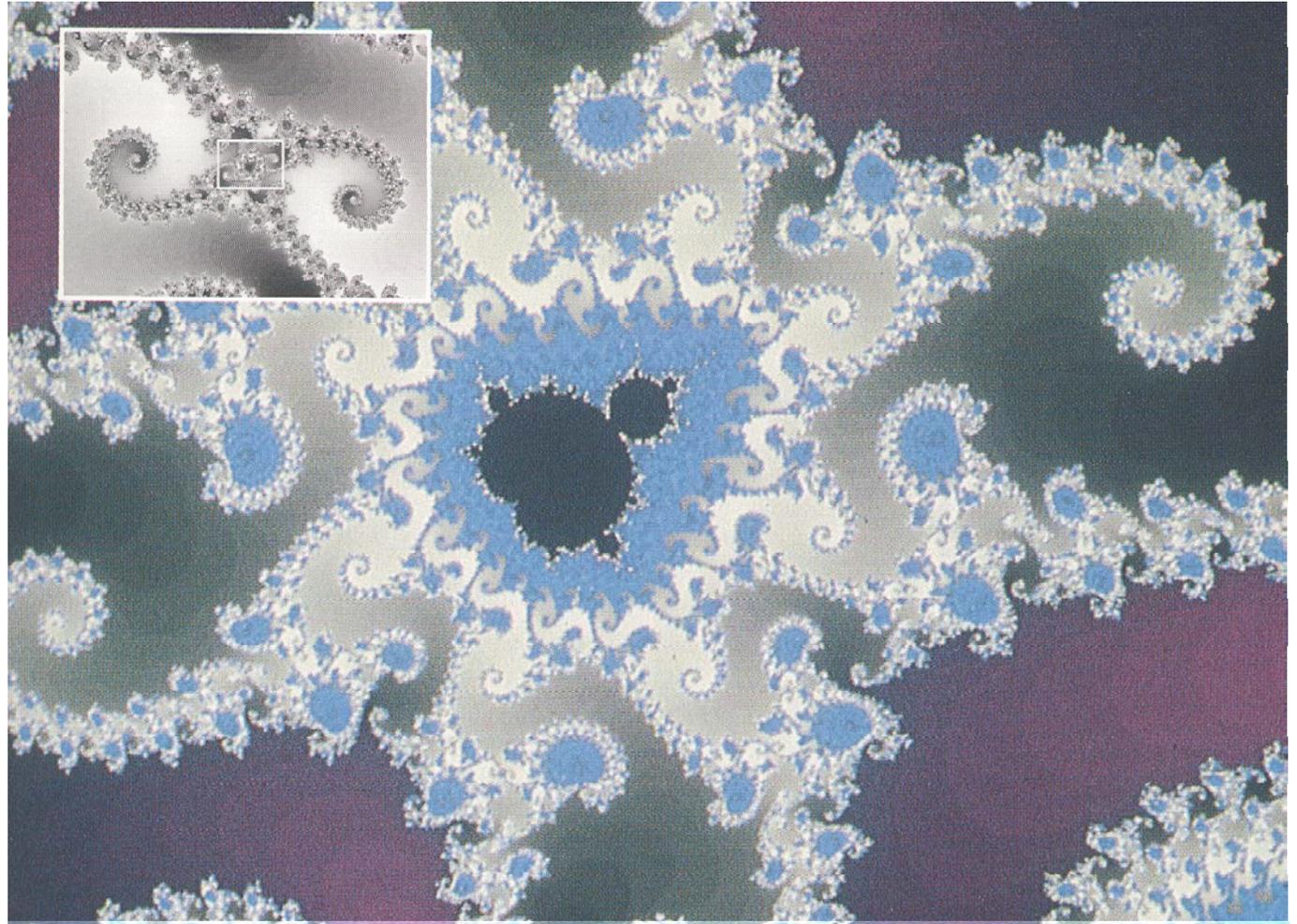
The Mandelbrot set as an example of fractal patterns in mathematics: Enlargement no.1

Picture taken from James Gleick, *Chaos*. Penguin Books, New York, 1988

Mandelbrot set

$$z \dot{=} z^2 + c$$

with $z = x + iy$



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

Picture taken from James Gleick, *Chaos*. Penguin Books, New York, 1988

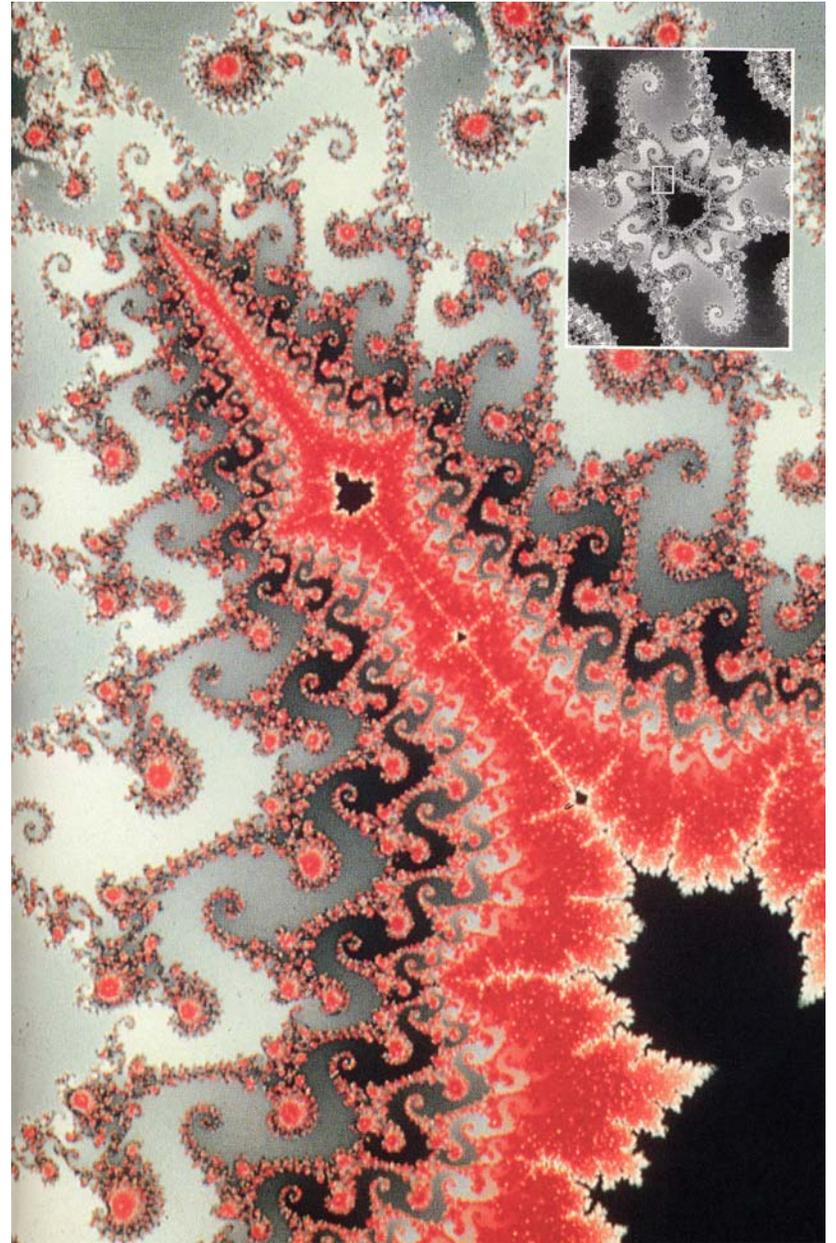
Mandelbrot set

$$z \checkmark z^2 + c$$

with $z = x + i y$

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

Picture taken from James Gleick, *Chaos*. Penguin
Books, New York, 1988

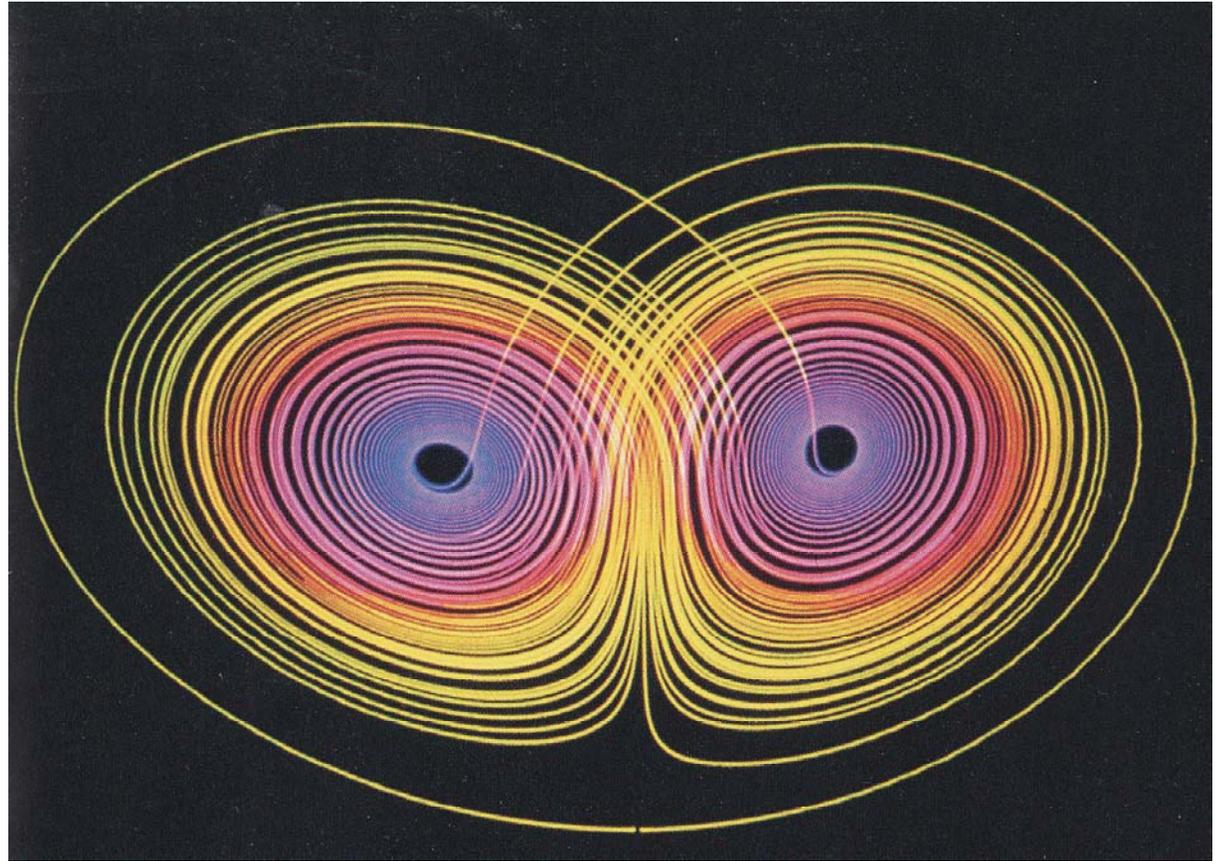


Lorenz attractor

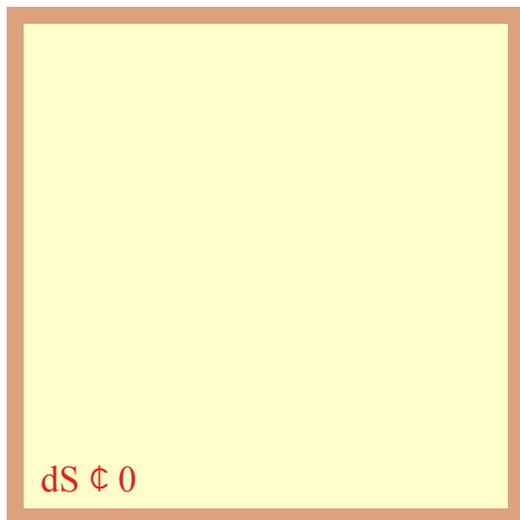
$$dx/dt = \sigma (y - x)$$

$$dy/dt = \rho x - y - xz$$

$$dz/dt = \beta z + xy$$



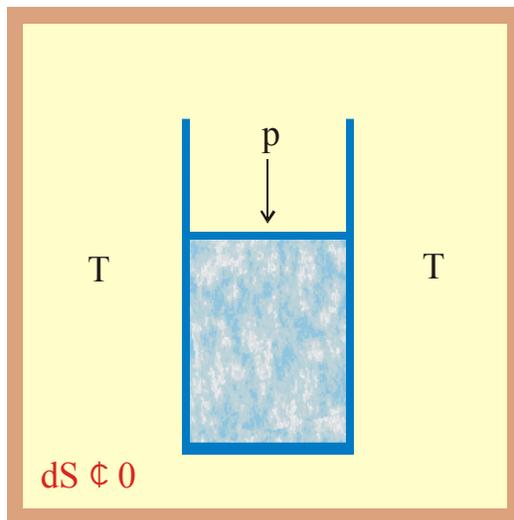
A trajectory of the Lorenz attractor in the chaotic regime



Isolated system

$$U = \text{const.}, V = \text{const.},$$

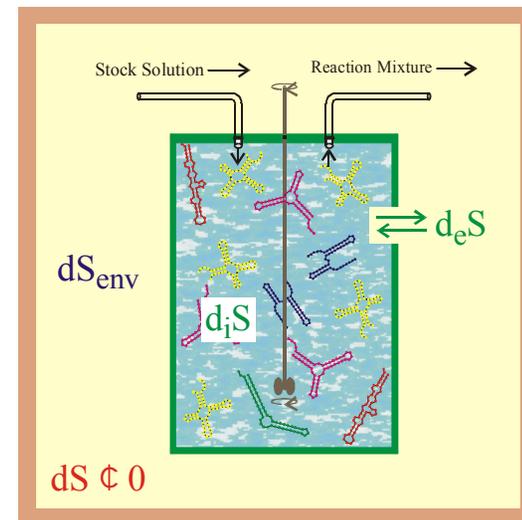
$$dS \leq 0$$



Closed system

$$T = \text{const.}, p = \text{const.},$$

$$dG = dU - pdV - TdS < 0$$



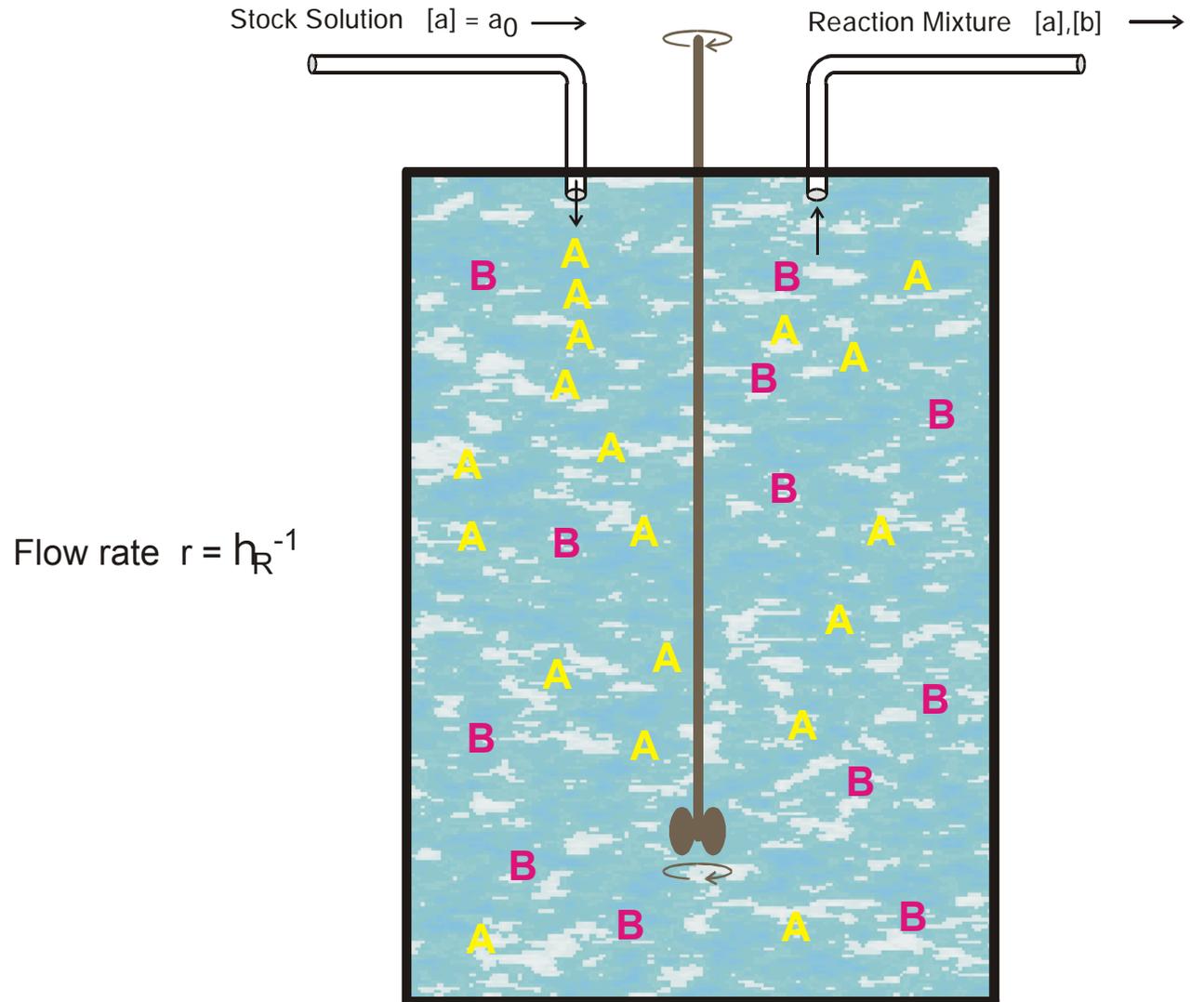
Open system

$$dS = dS_{\text{env}} + dS \leq 0$$

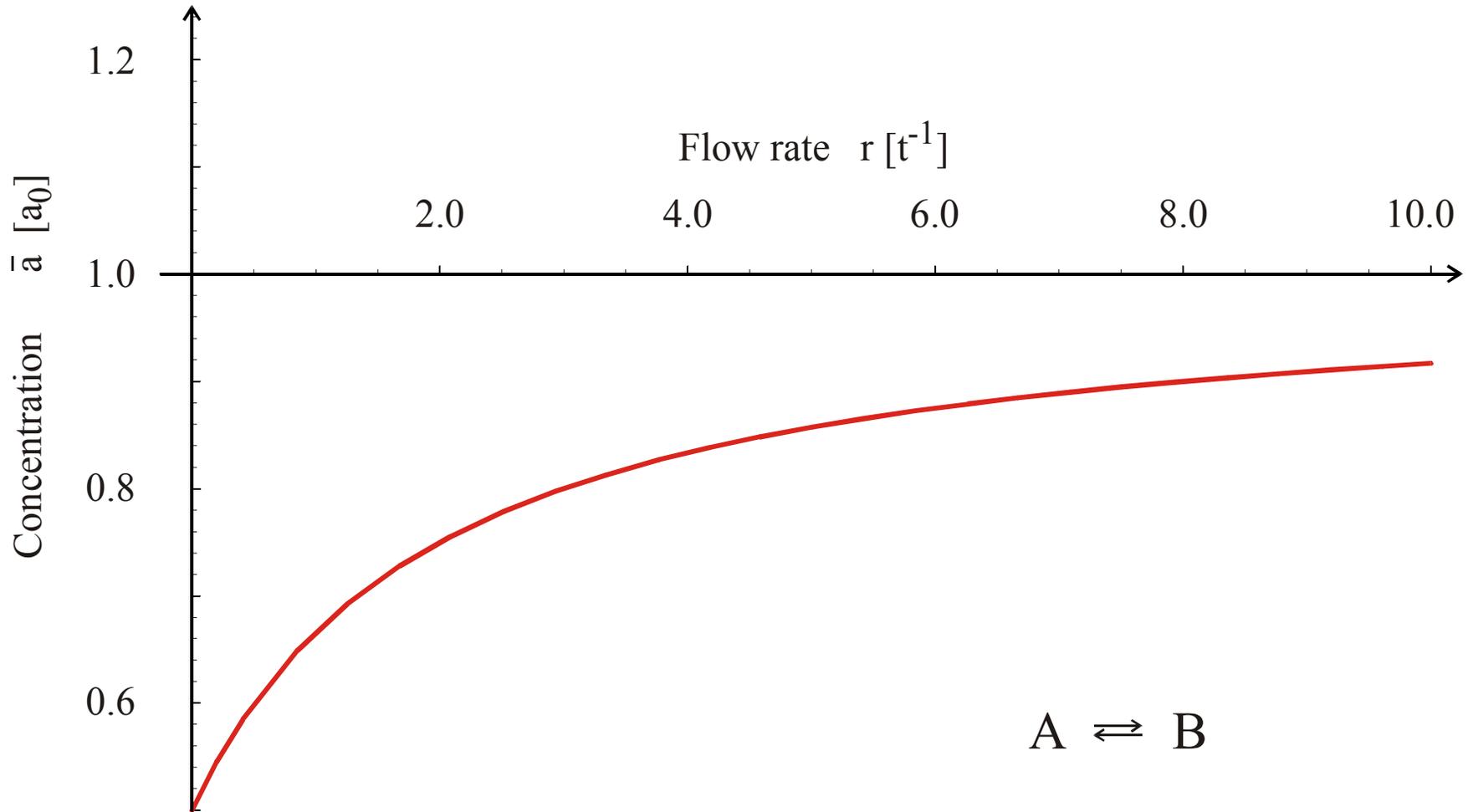
$$dS = d_i S + d_e S$$

$$d_i S \leq 0$$

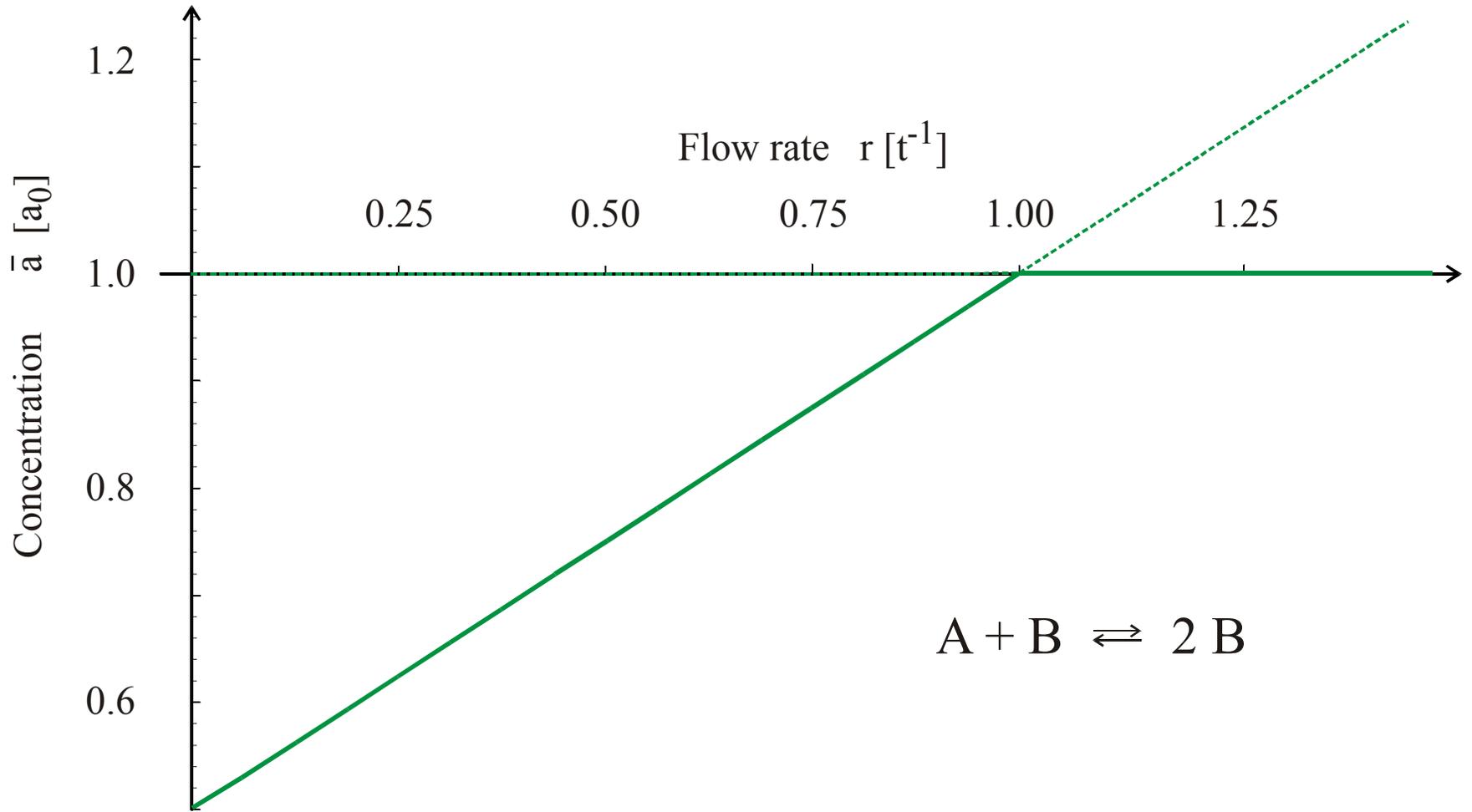
Entropy changes in different thermodynamic systems with chemical reactions



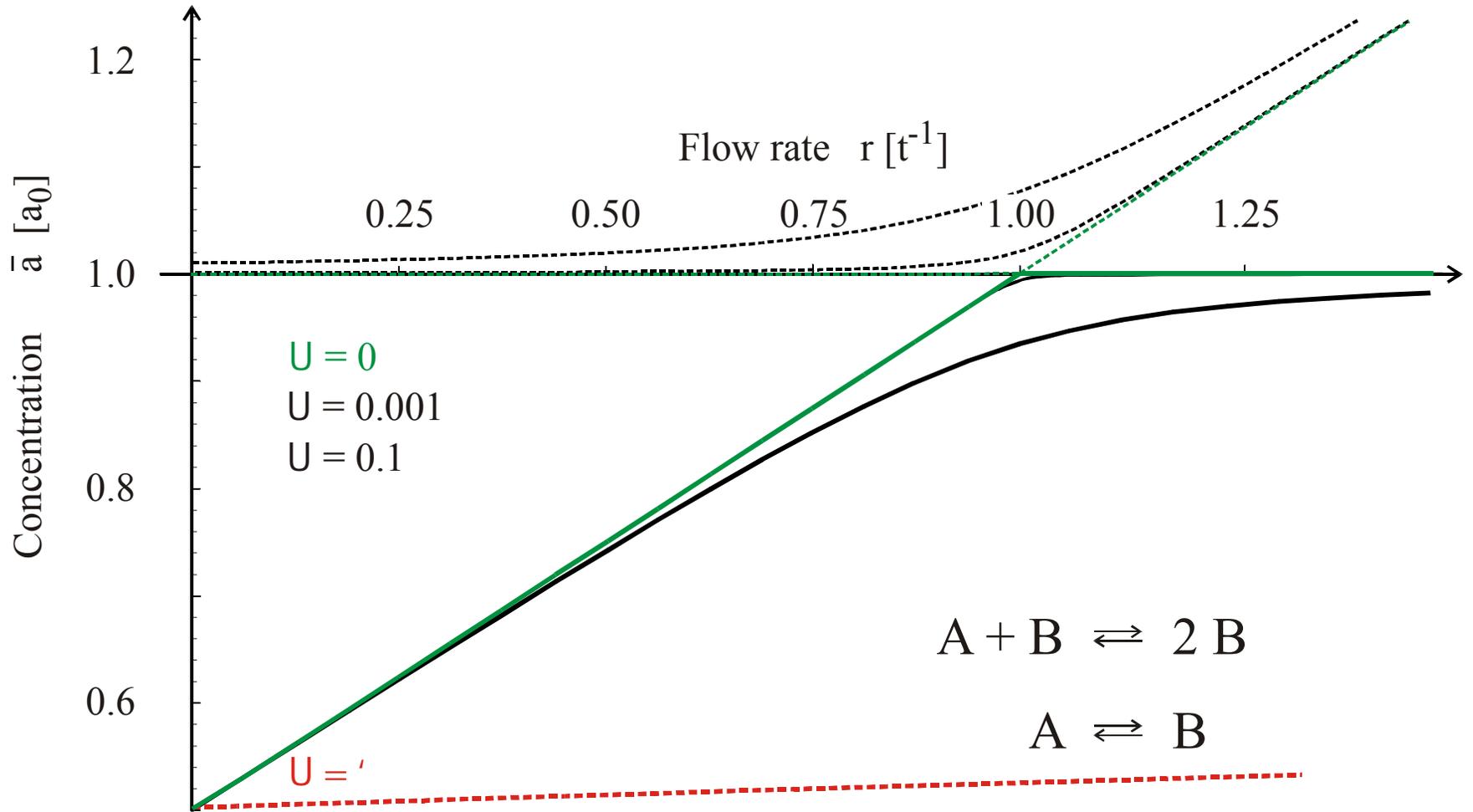
Reactions in the continuously stirred tank reactor (CSTR)



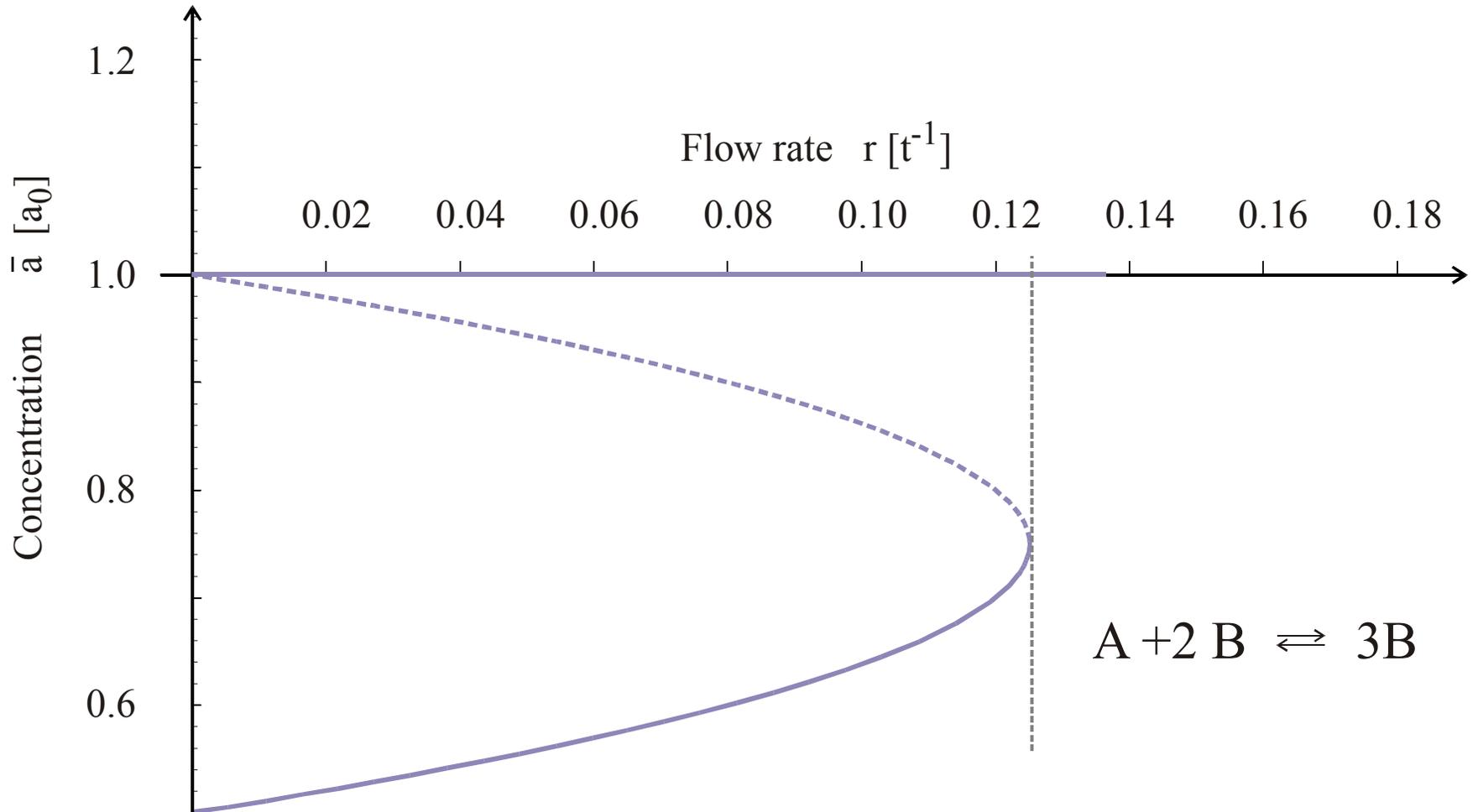
Reversible first order reaction in the flow reactor



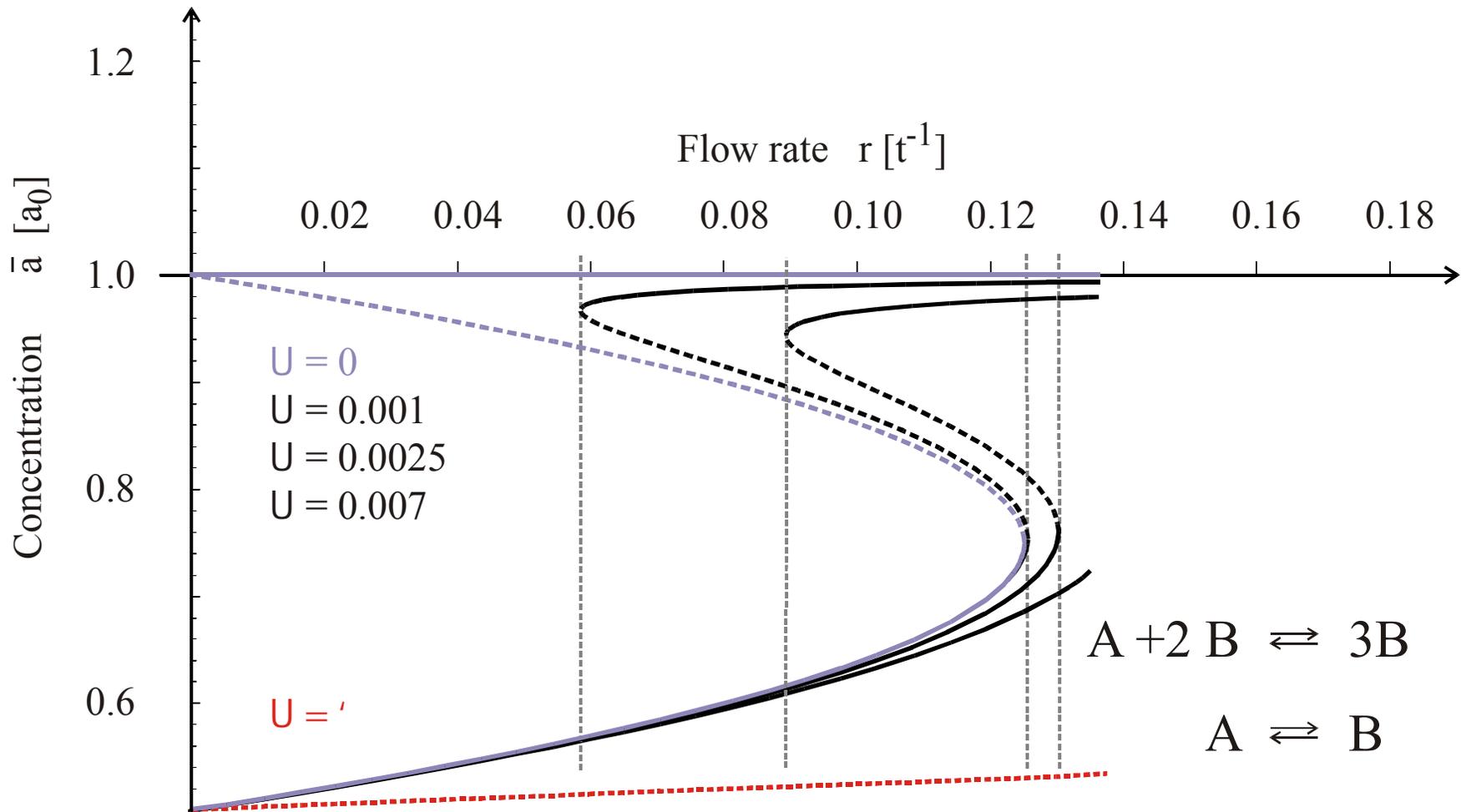
Autocatalytic second order reaction in the flow reactor



Autocatalytic second order and uncatalyzed reaction in the flow reactor



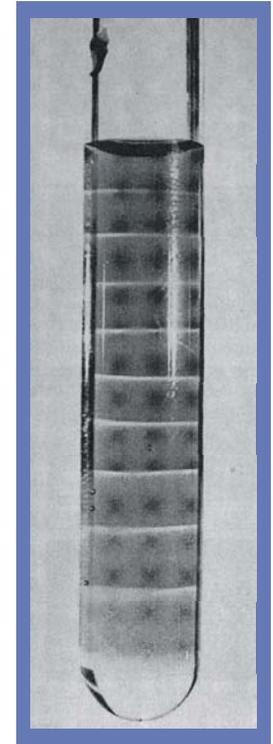
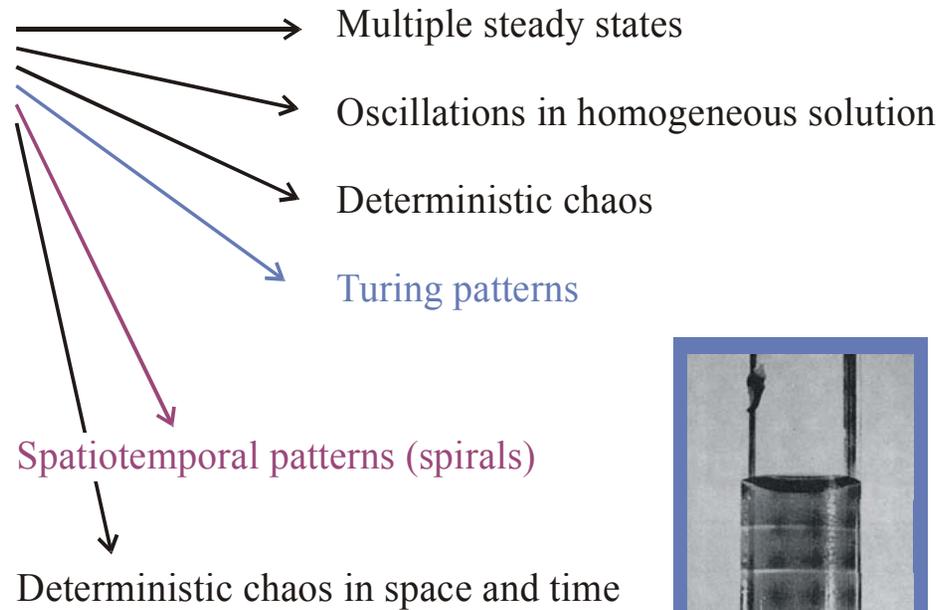
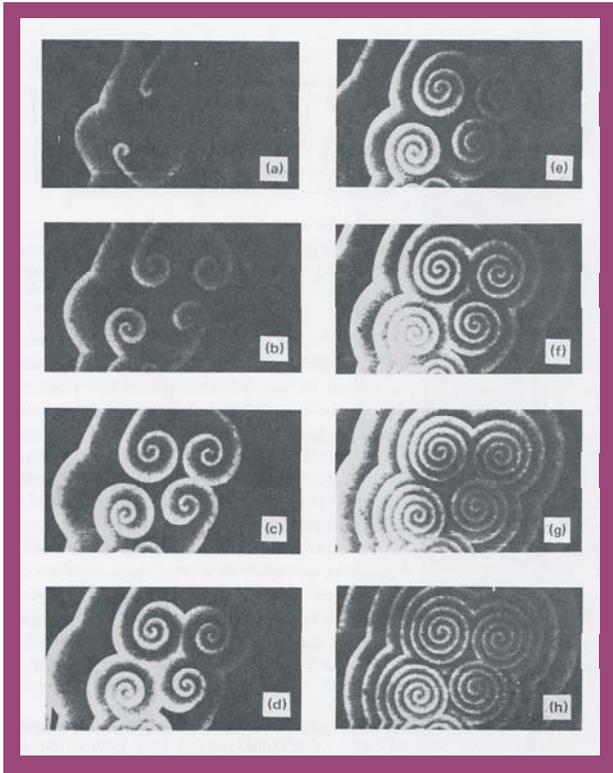
Autocatalytic third order reaction in the flow reactor



Autocatalytic third order and uncatalyzed reaction in the flow reactor

Autocatalytic third order reactions

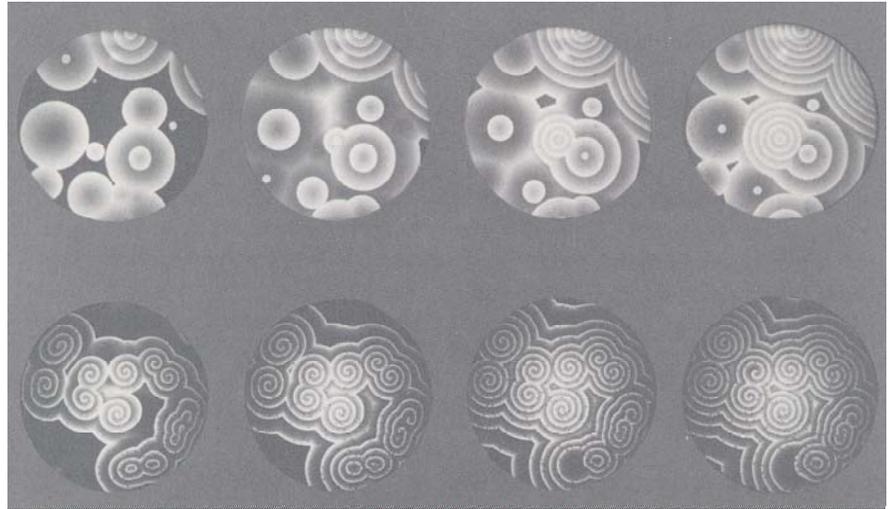
Direct, $A + 2X \rightarrow 3X$, or hidden in the reaction mechanism (Belousov-Zhabotinskii reaction).



Pattern formation in autocatalytic third order reactions

G.Nicolis, I.Prigogine. *Self-Organization in Nonequilibrium Systems. From Dissipative Structures to Order through Fluctuations*. John Wiley, New York 1977

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



Winding number:

number of left-handed spirals
minus
number of right-handed spirals

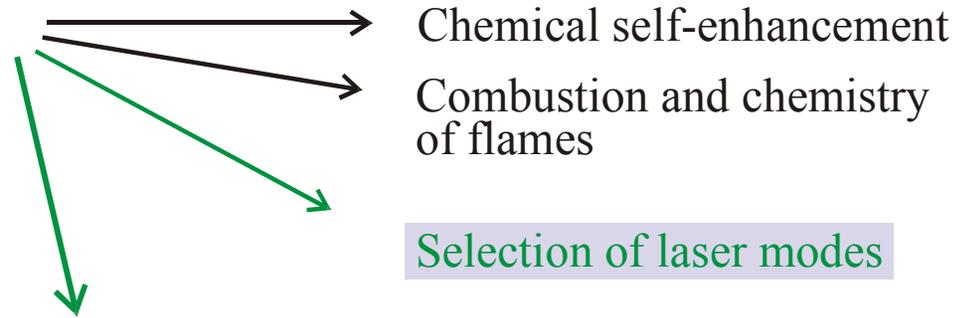


Target waves and spirals in the Belousov-Zhabotinskii reaction

Pictures taken from Arthur T. Winfree, *The geometry of biological time*. Springer-Verlag, New York, 1980.

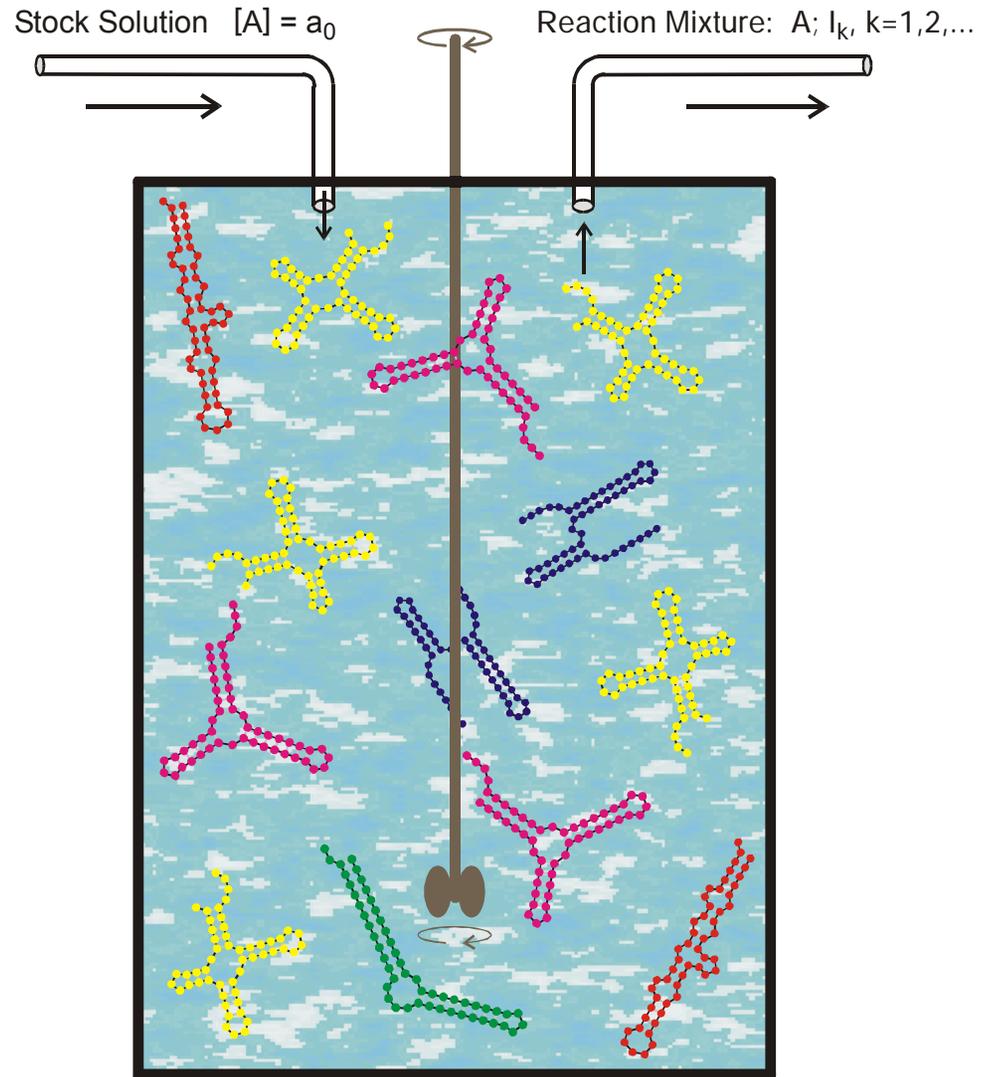
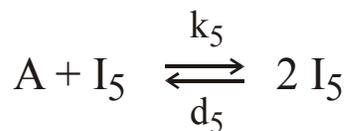
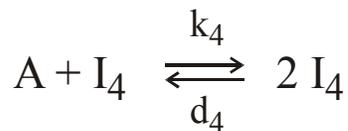
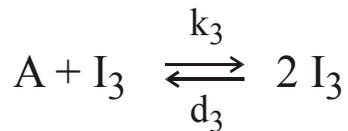
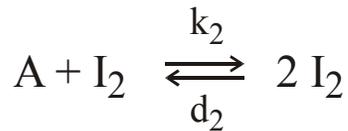
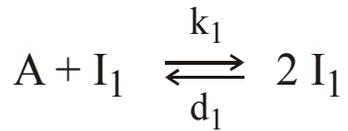
Autocatalytic second order reactions

Direct, $A + I \xrightarrow{k} 2I$, or hidden in the reaction mechanism



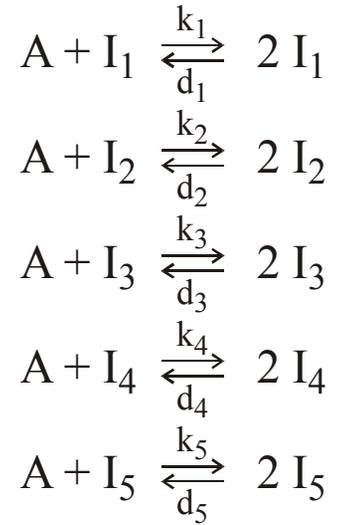
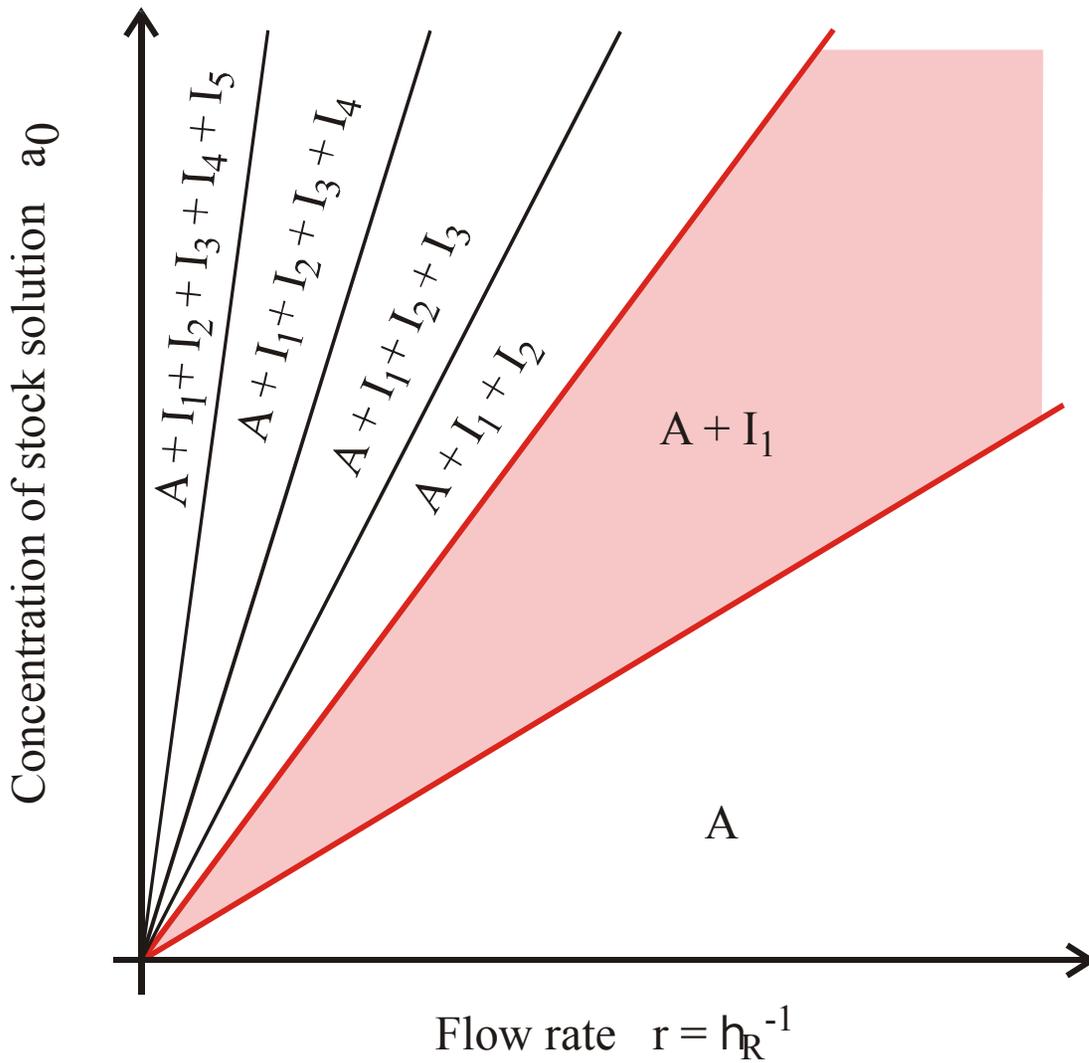
Autocatalytic second order reaction as basis for selection processes.

The autocatalytic step is formally equivalent to replication or reproduction.



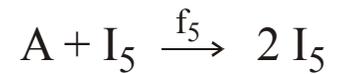
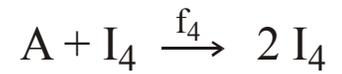
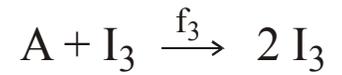
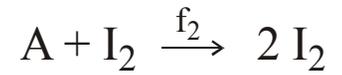
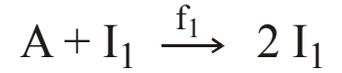
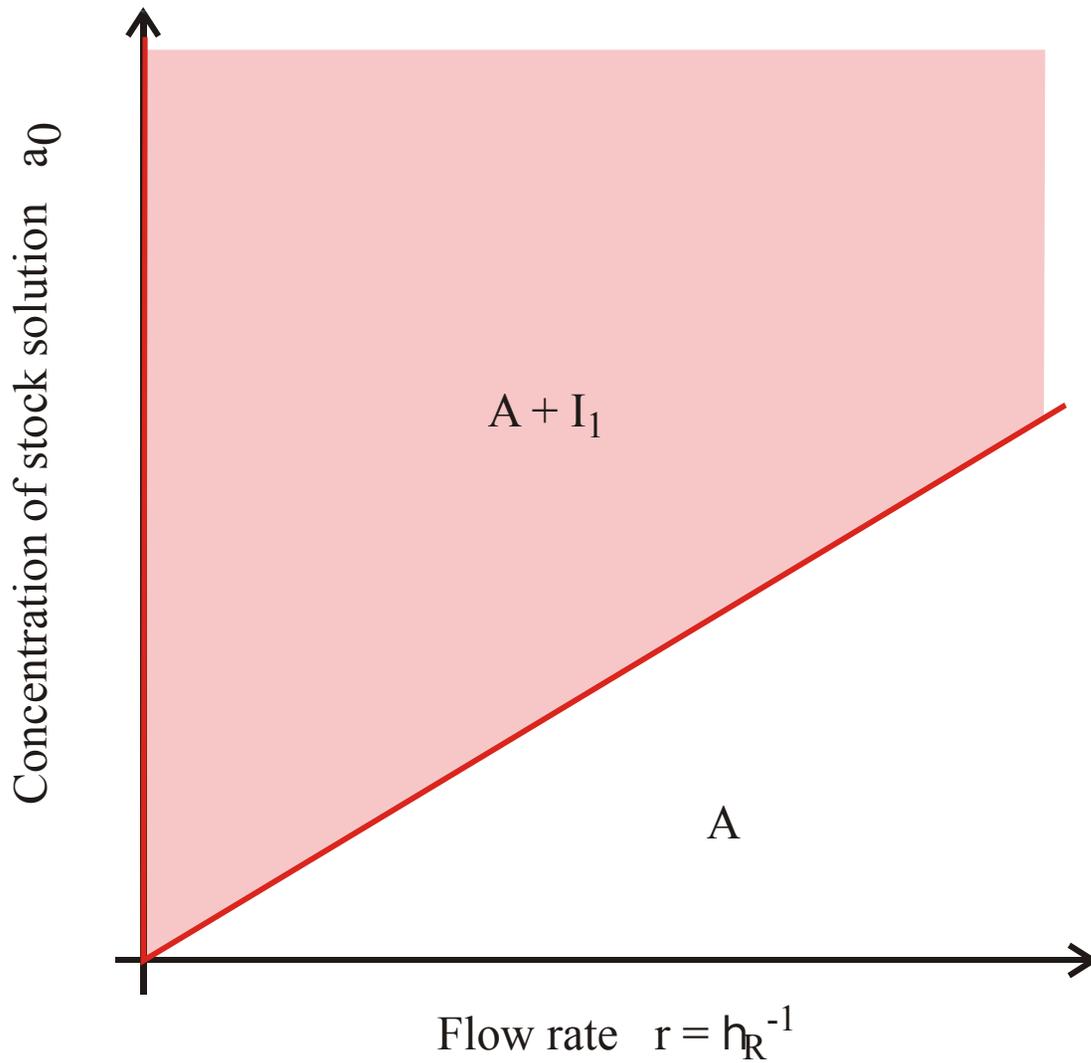
Replication in the flow reactor

P.Schuster & K.Sigmund, Dynamics of evolutionary optimization, *Ber.Bunsenges.Phys.Chem.* **89**: 668-682 (1985)



$$k_1 > k_2 > k_3 > k_4 > k_5$$

Selection in the flow reactor: Reversible replication reactions



$$f_1 > f_2 > f_3 > f_4 > f_5$$

Selection in the flow reactor: Irreversible replication reactions

$$dx_j / dt = f_j x_j - x_j \Phi = (f_j - \Phi) x_j$$

$$\Phi = \sum_i f_i x_i ; \quad \sum_i x_i = 1 ; \quad i, j = 1, 2, \dots, n$$

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

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

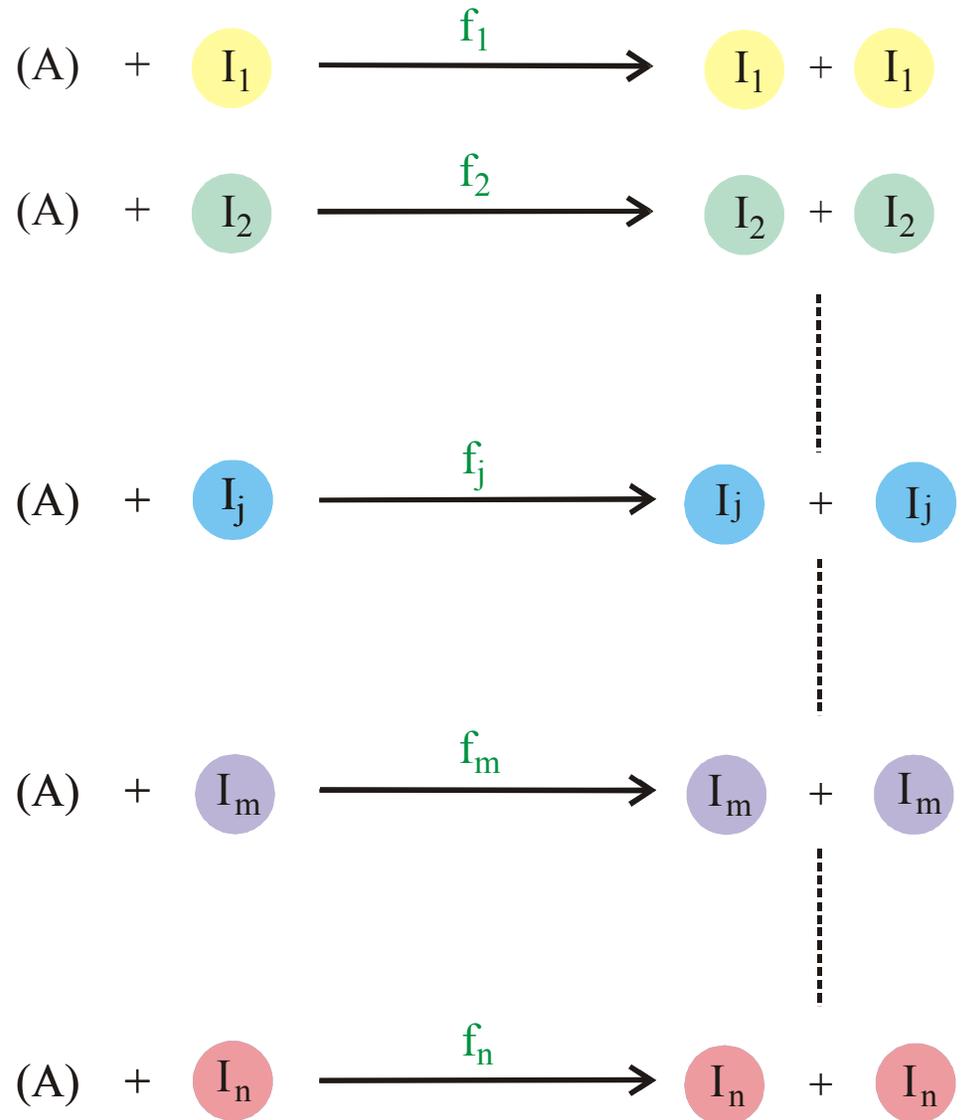
$$x_m(t) \xrightarrow{t \rightarrow \infty} 1$$

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

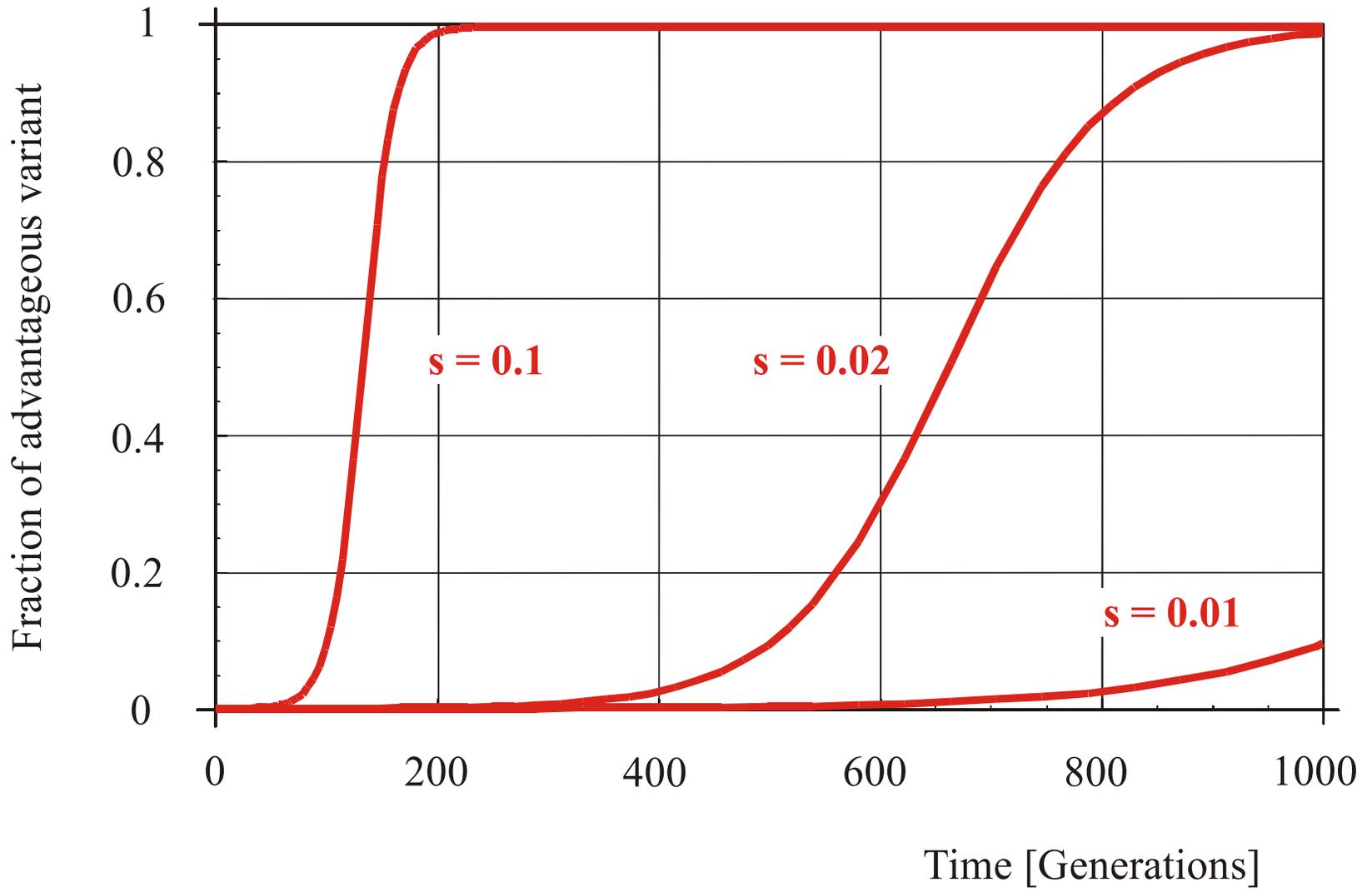
succession of temporarily

fittest variants:

$$m \checkmark m+1 \checkmark \dots$$



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:
state function

Entropy is a non-decreasing

Second law

$$S \rightarrow S_{\max}$$

Valid in the limit of infinite time, $\lim_{t \rightarrow \infty} S = S_{\max}$.

Evolution of Populations:

Mean fitness is a non-decreasing function

Ronald Fisher's conjecture

$$\bar{f} = \frac{\sum_k x_k(t) f_k}{\sum_k x_k(t)} \rightarrow f_{\max}$$

is
need

Optimization heuristics in the sense that it
only almost always true and the process

not reach the optimum in finite times.

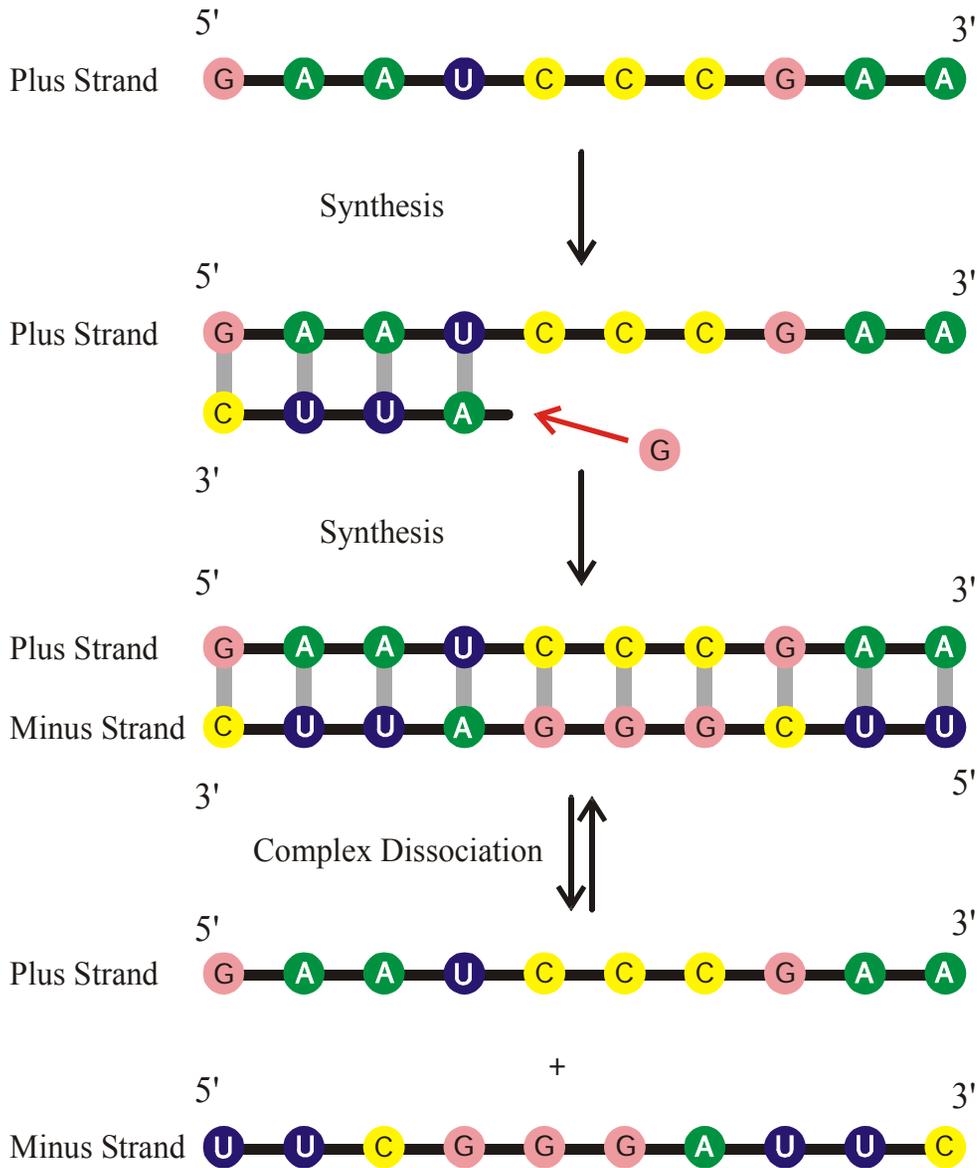


$4^{27} = 1.801 \times 10^{16}$ possible different sequences

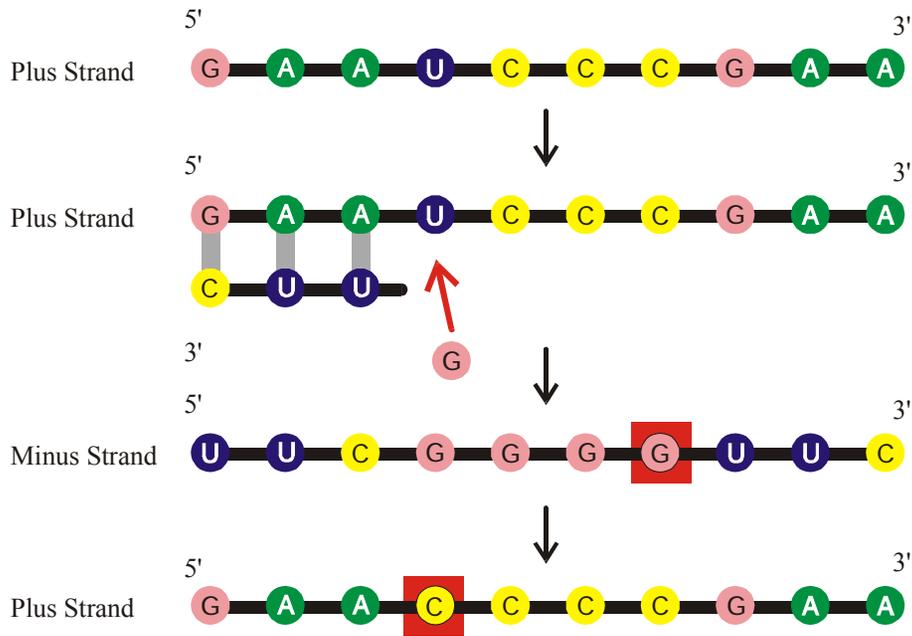
Combinatorial diversity of sequences: $N = 4^0$

- A** = adenylate
- U** = uridylate
- C** = cytidylate
- G** = guanylate

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



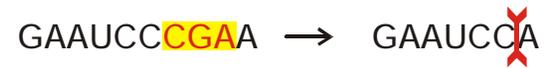
Complementary replication as the simplest copying mechanism of RNA



Point Mutation



Insertion



Deletion

Mutations represent the mechanism of variation in nucleic acids

$$dx_j / dt = \sum_i f_i Q_{ji} x_i - x_j \Phi$$

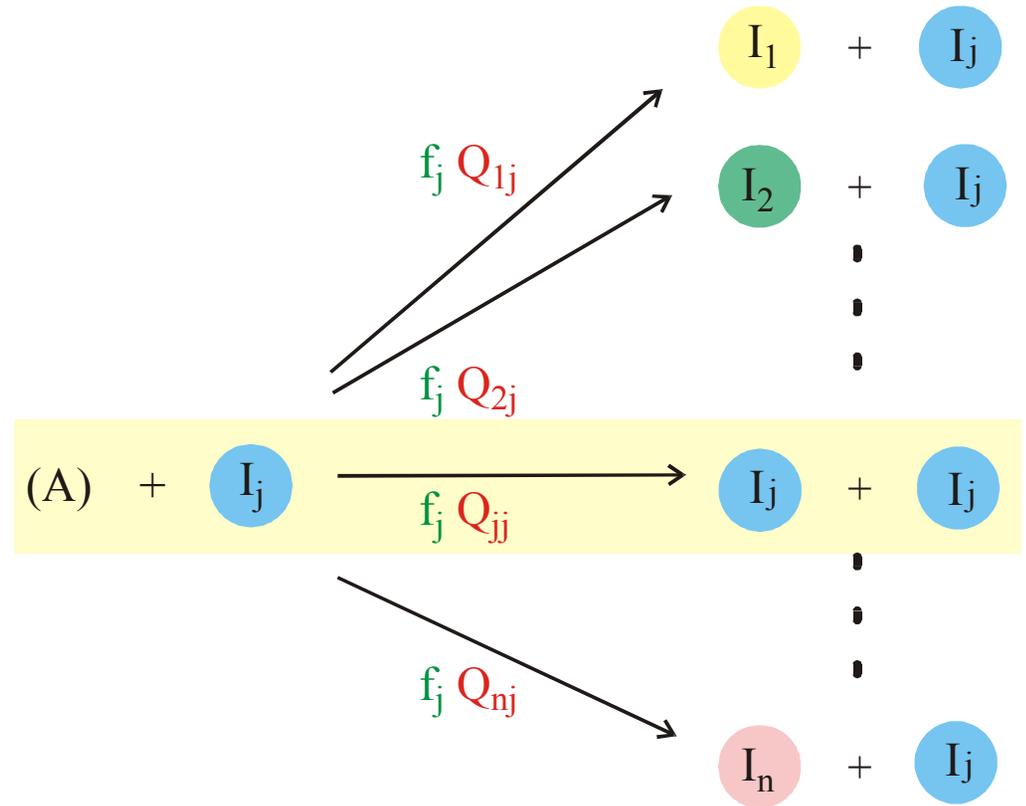
$$\Phi = \sum_i f_i x_i ; \quad \sum_i x_i = 1 ; \quad \sum_i Q_{ij} = 1$$

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

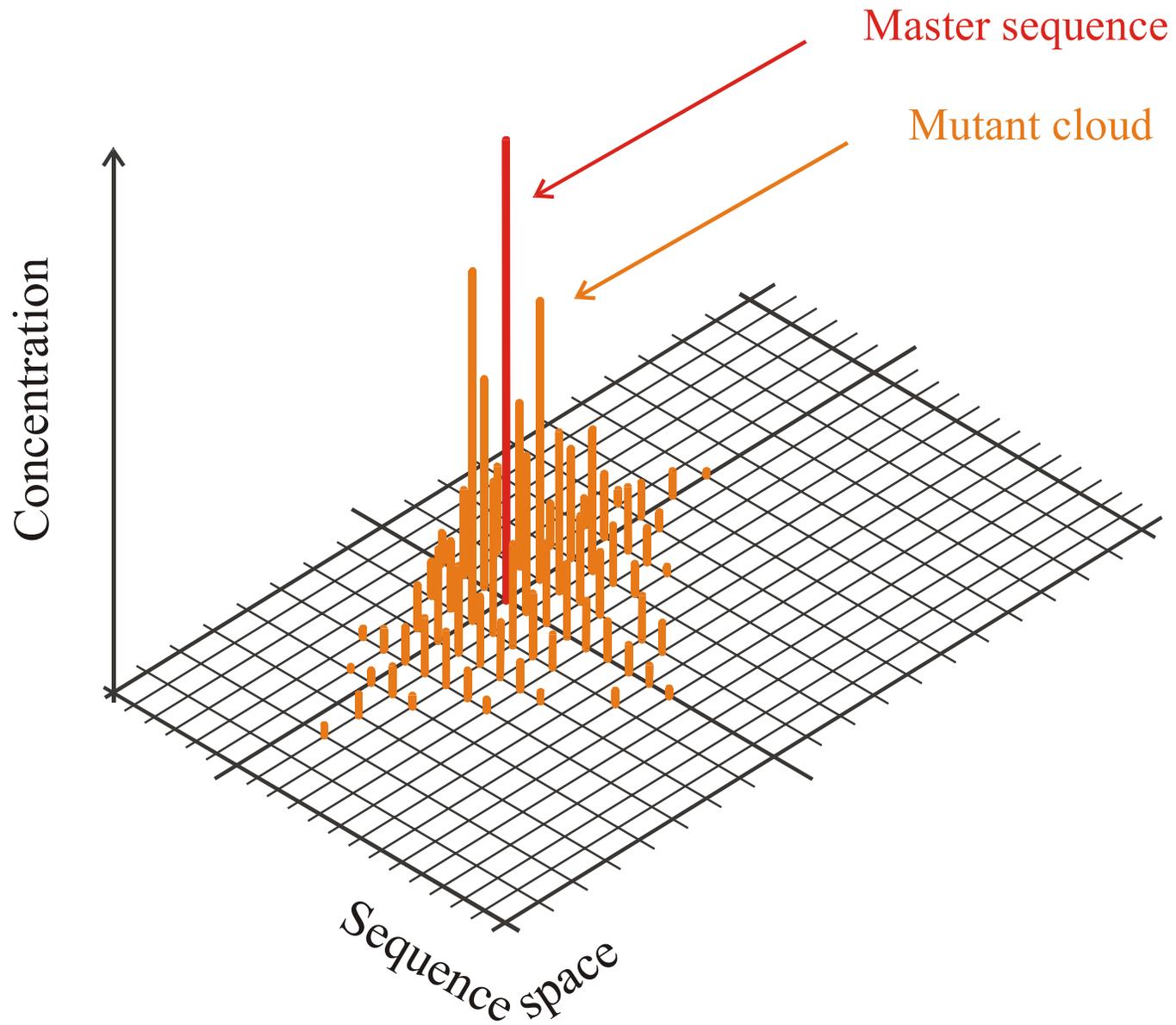
p Error rate per digit

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

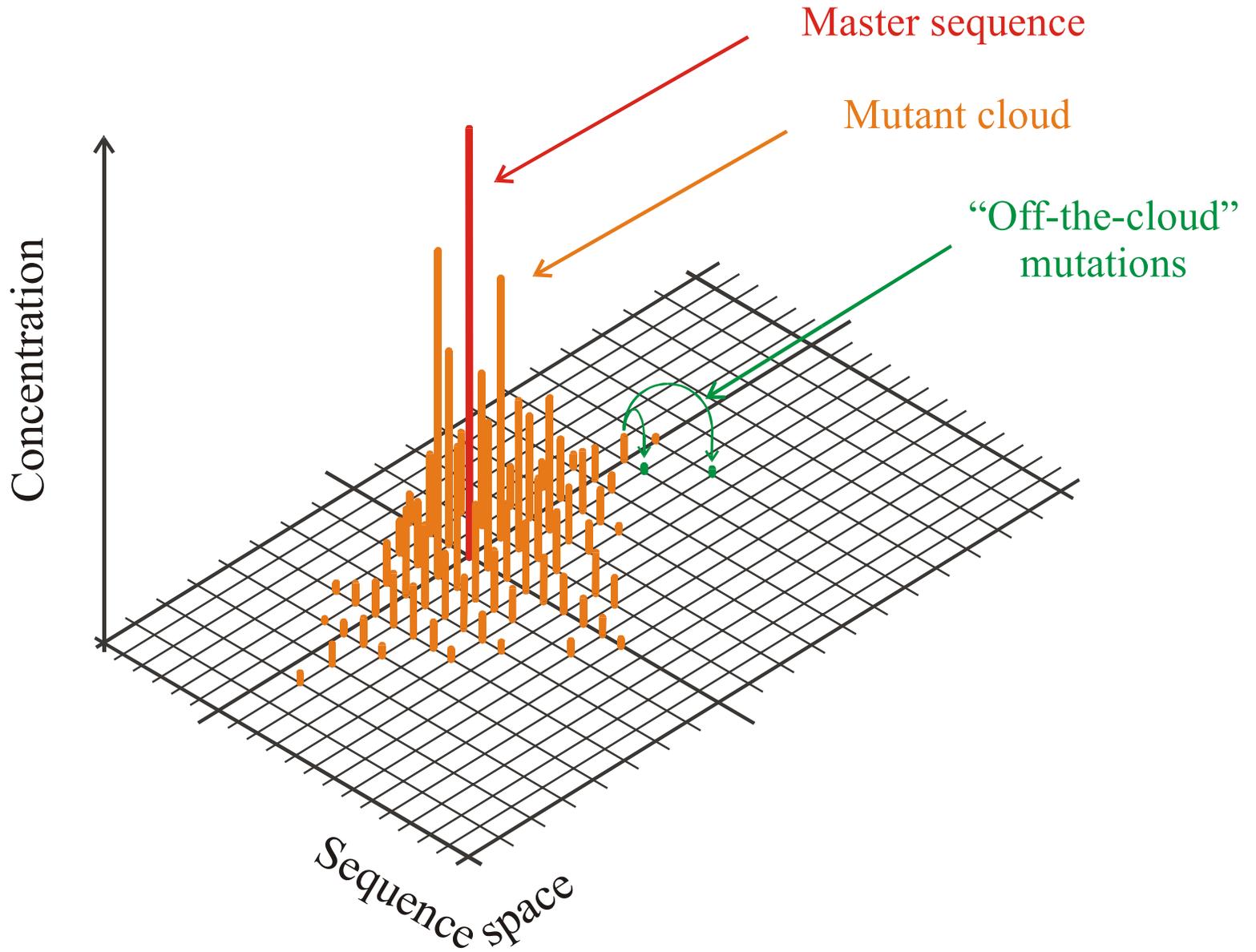
[A] = a = constant



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*

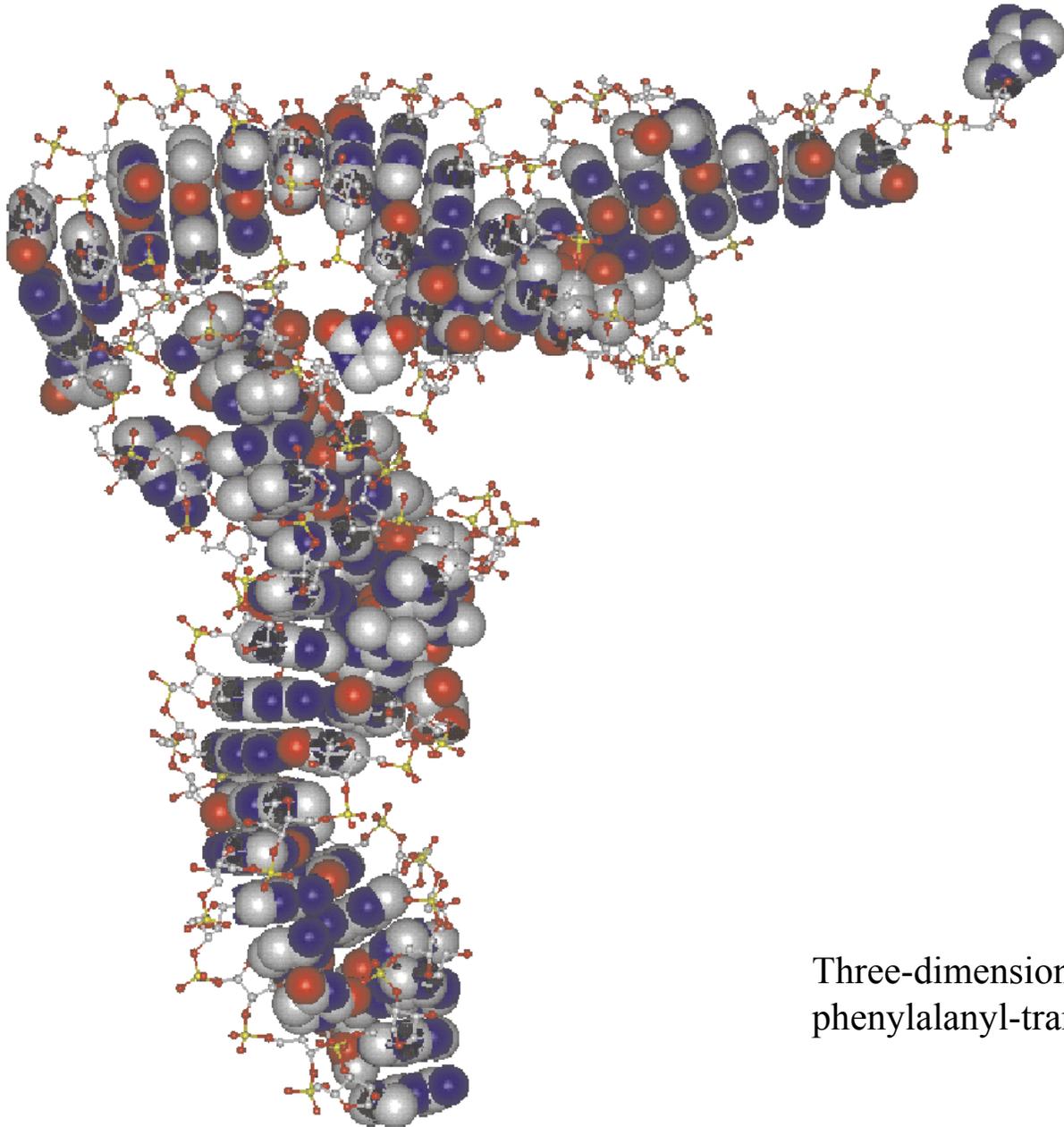
W.Fontana, P.Schuster, *A computer model of evolutionary optimization*. Biophysical Chemistry **26** (1987), 123-147

W.Fontana, W.Schnabl, P.Schuster, *Physical aspects of evolutionary optimization and adaptation*. Phys.Rev.A **40** (1989), 3301-3321

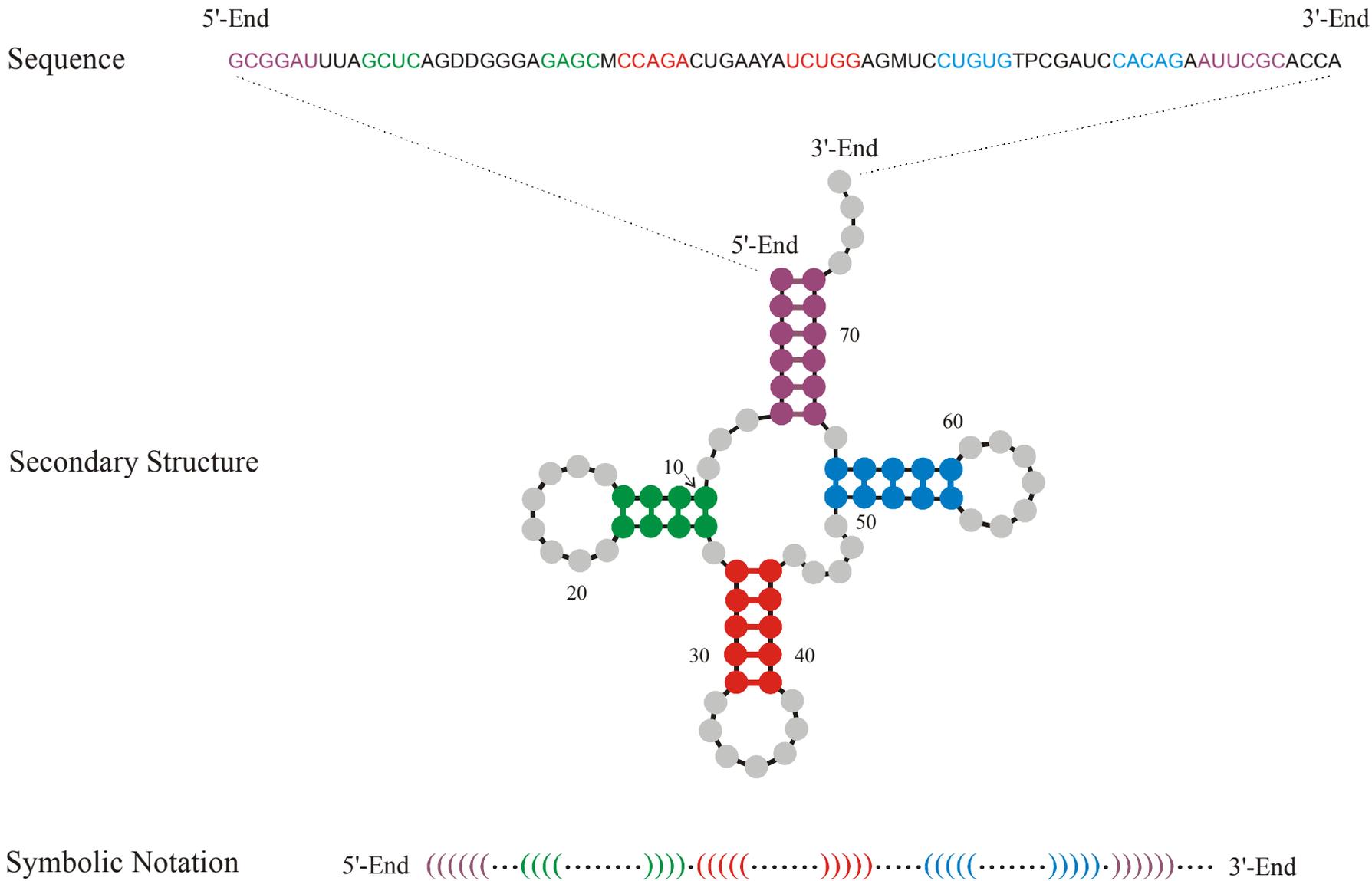
M.A.Huynen, W.Fontana, P.F.Stadler, *Smoothness within ruggedness. The role of neutrality in adaptation*. Proc.Natl.Acad.Sci.USA **93** (1996), 397-401

W.Fontana, P.Schuster, *Continuity in evolution. On the nature of transitions*. Science **280** (1998), 1451-1455

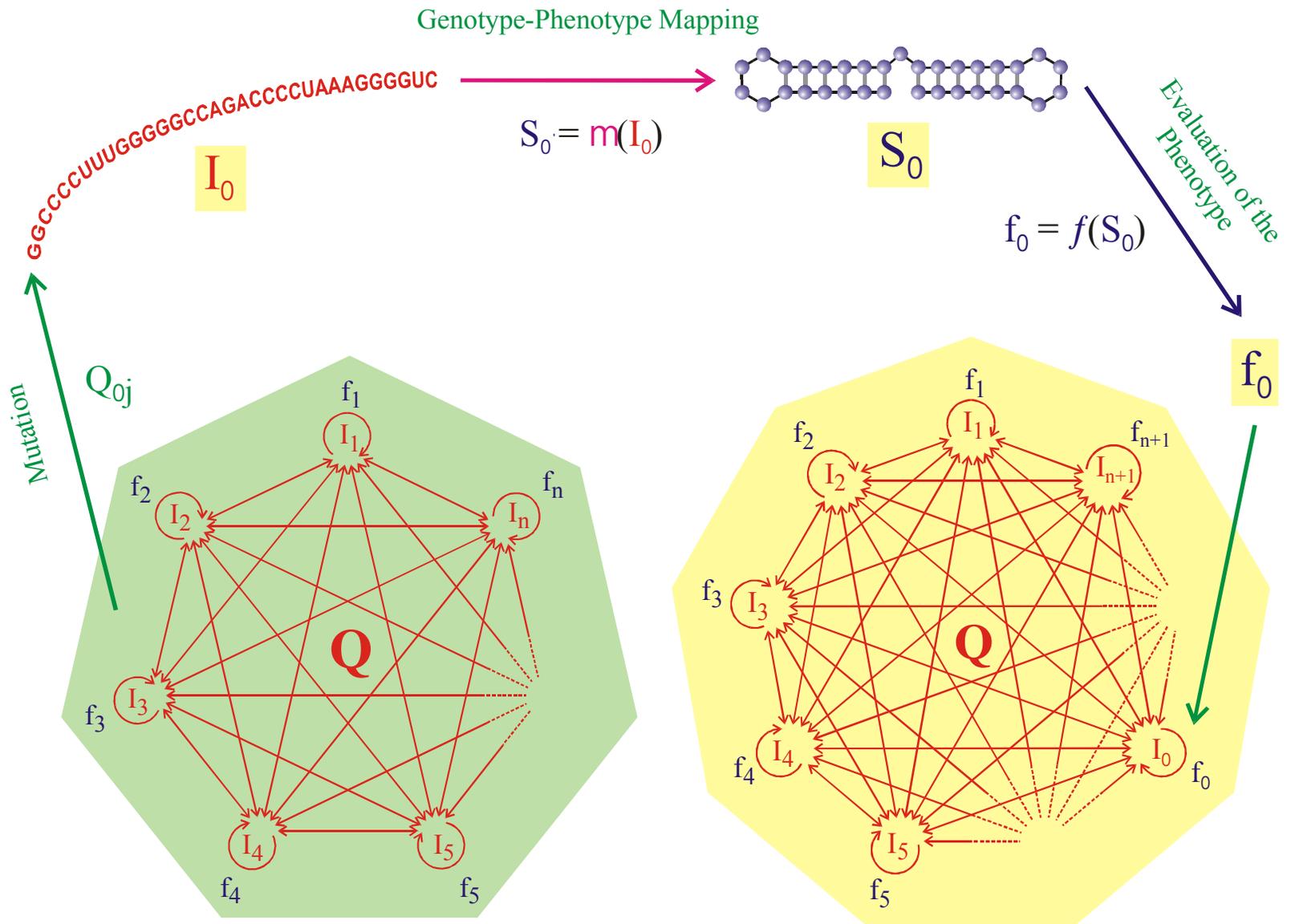
W.Fontana, P.Schuster, *Shaping space. The possible and the attainable in RNA genotype-phenotype mapping*. J.Theor.Biol. **194** (1998), 491-515



Three-dimensional structure of phenylalanyl-transfer-RNA



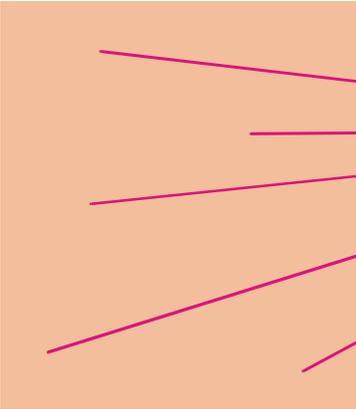
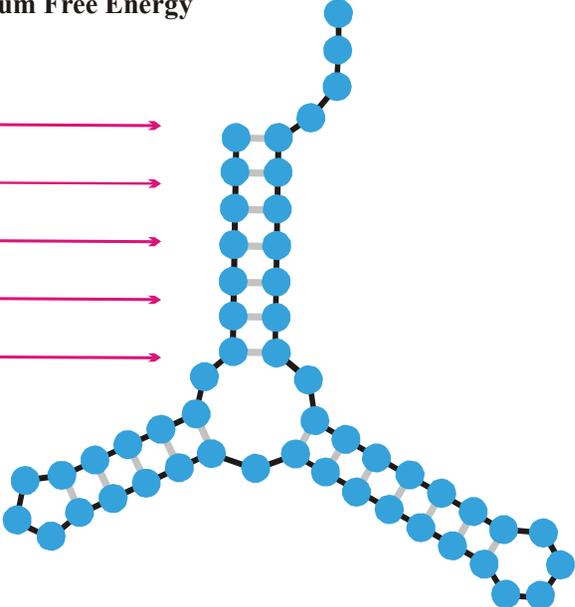
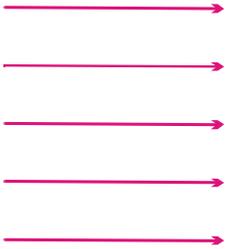
Definition and formation of the secondary structure of phenylalanyl-tRNA



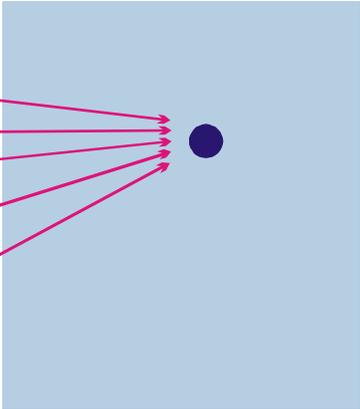
Evolutionary dynamics
including molecular phenotypes

**Criterion of
Minimum Free Energy**

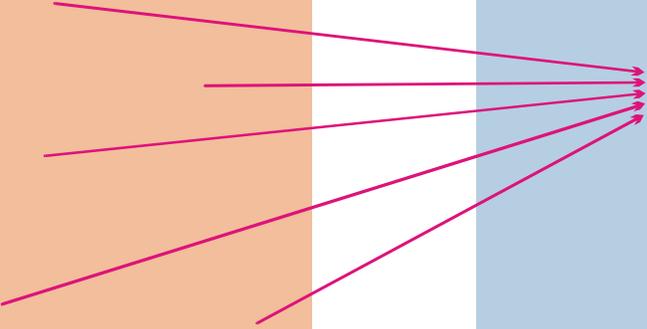
UUUAGCCAGCGCGAGUCGUGCGGACGGGGUUAUCUCUGUCGGGCUAGGGCGC
GUGAGCGCGGGGCACAGUUUCUCAAGGAUGUAAGUUUUUGCCGUUUUUCUGG
UUAGCGAGAGAGGAGGCUUCUAGACCCAGCUCUCUGGGUCGUUGCUGAUGCG
CAUUGGUGCUAAUGAUUUAGGGCUGUAUUCUGUAUAGCGAUCAGUGUCCG
GUAGGCCUCUUGACAUAAGAUUUUUCCAAUGGUGGGAGAUGGCCAUUGCAG

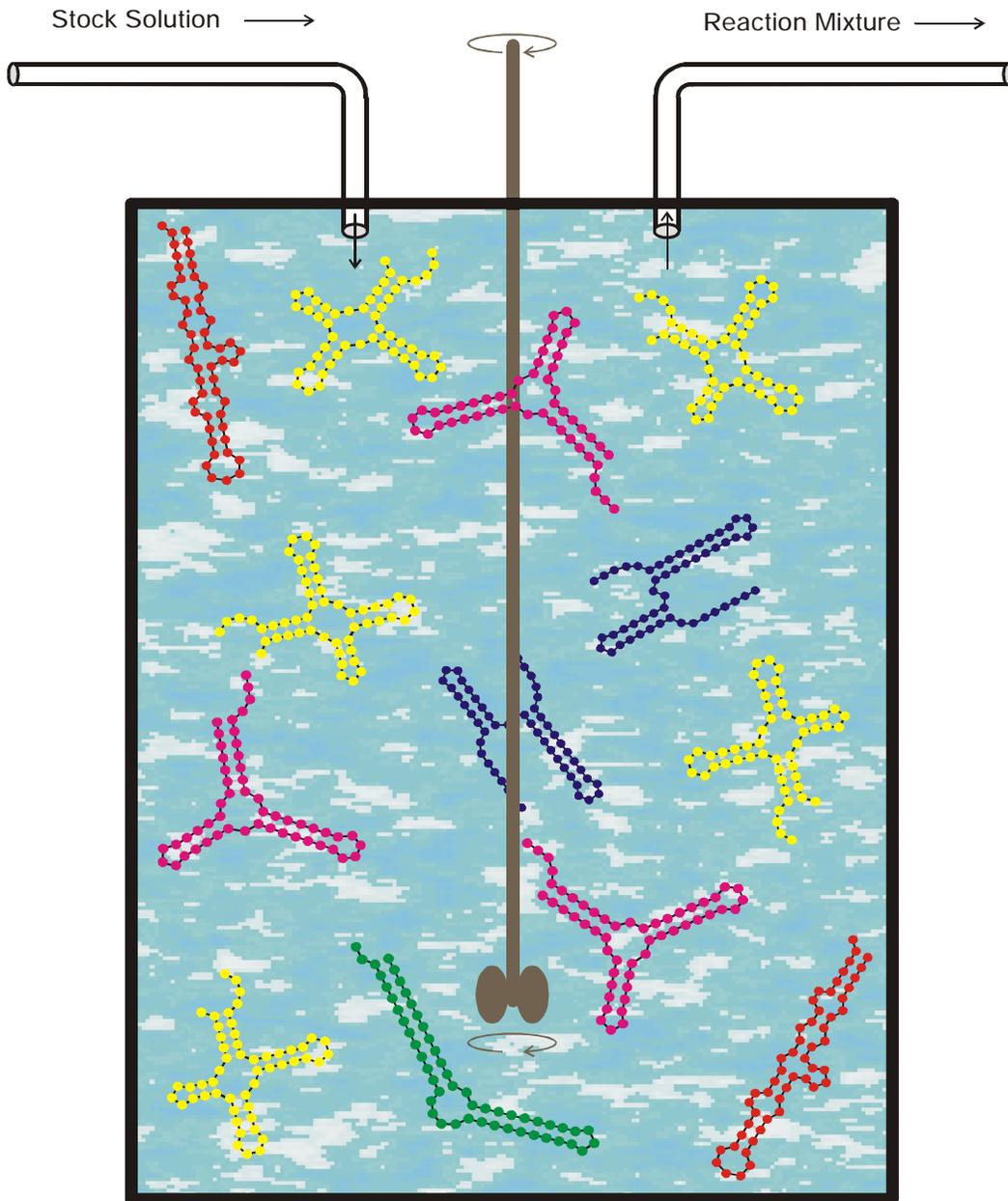


Sequence Space

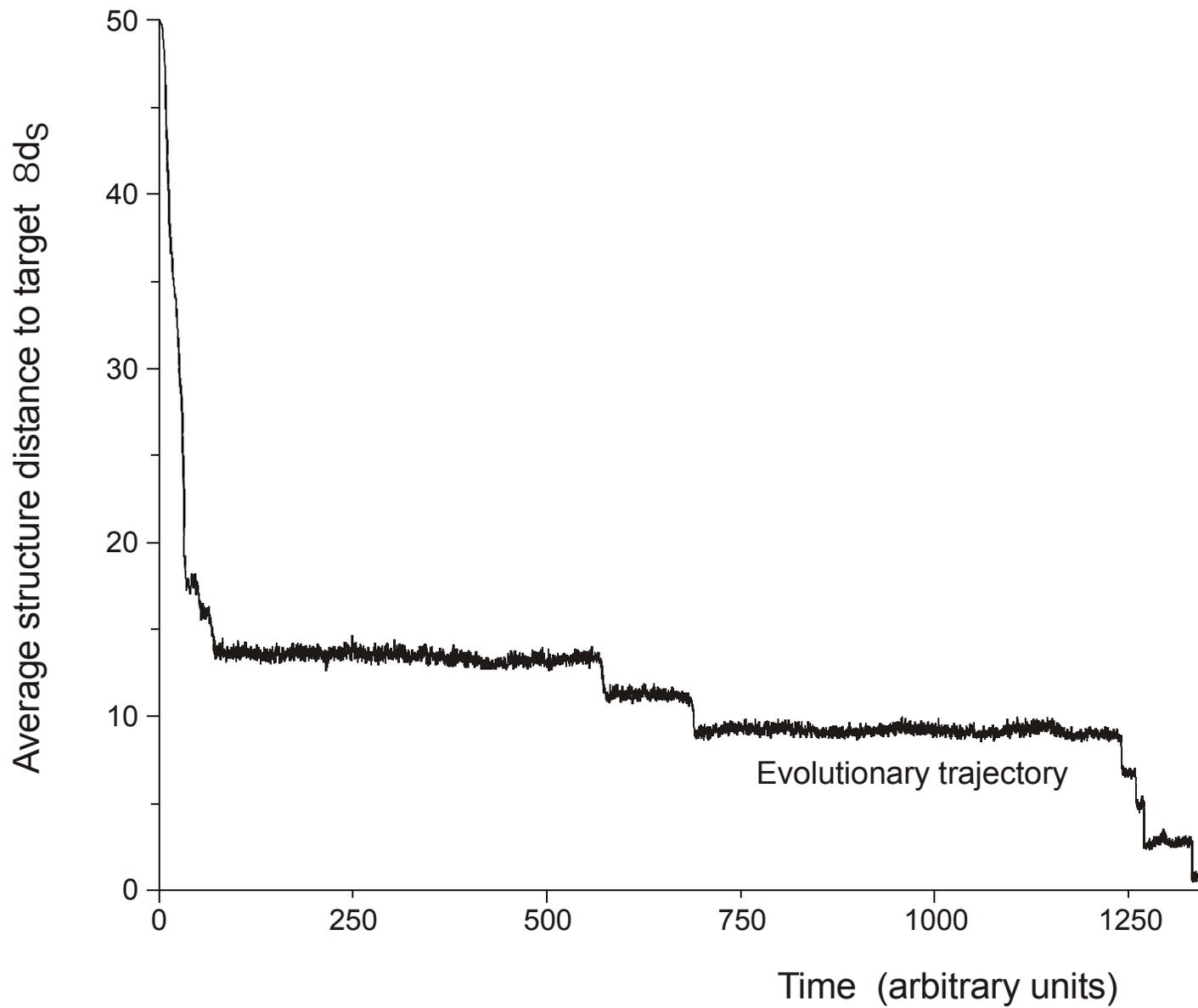


Shape Space

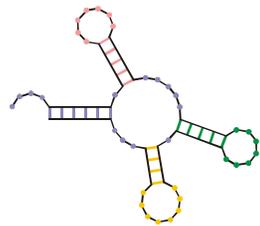
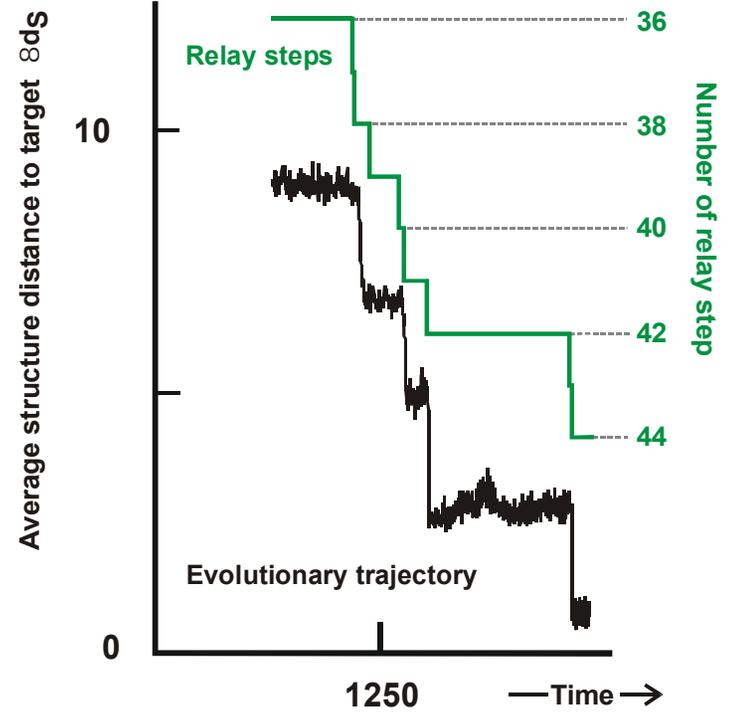




The flowreactor as a device for studies of evolution *in vitro* and *in silico*

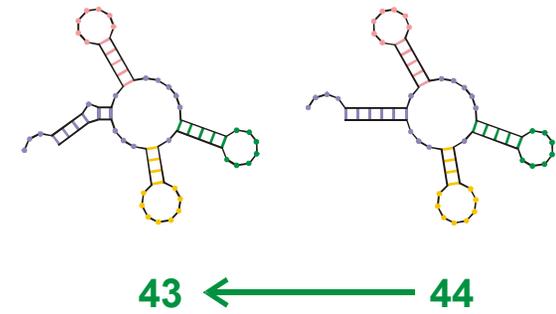
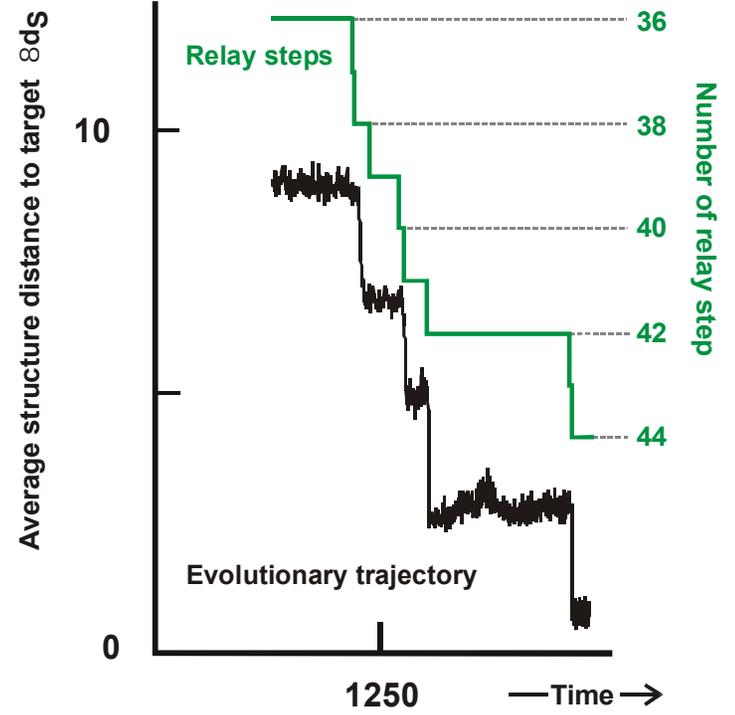


In silico optimization in the flow reactor: Trajectory

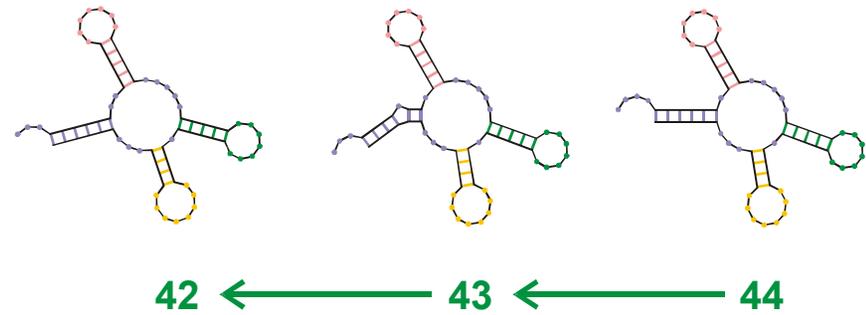
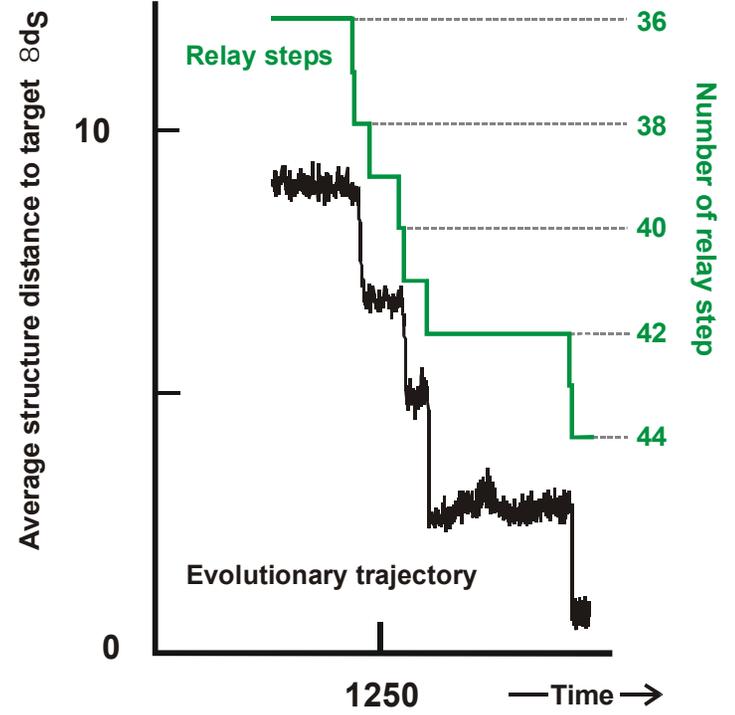


44

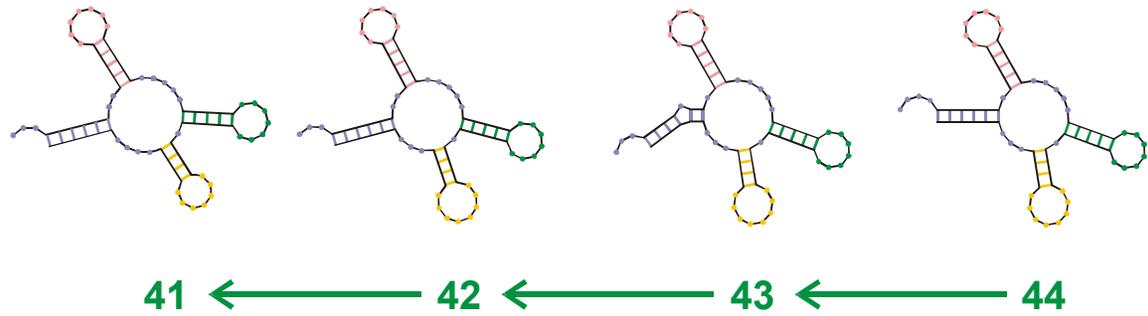
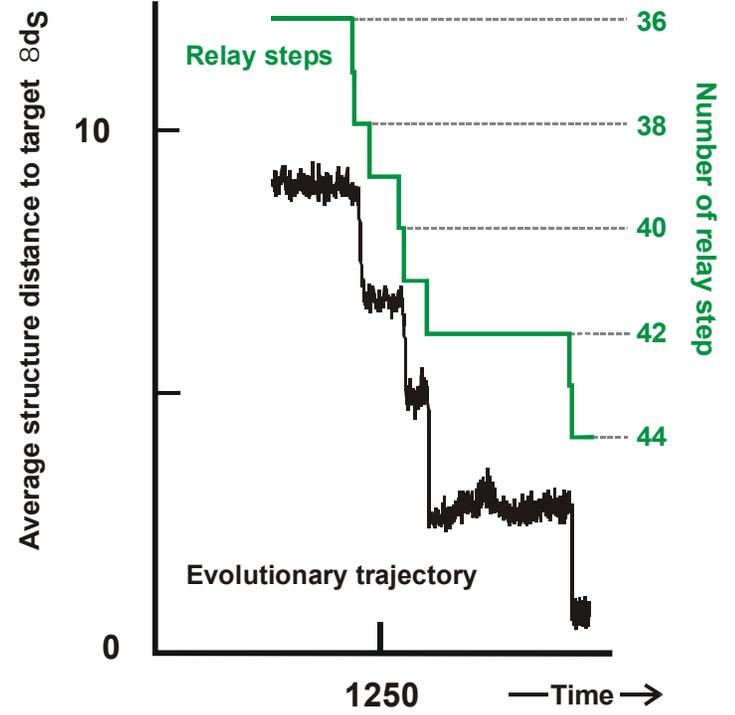
Endconformation of optimization



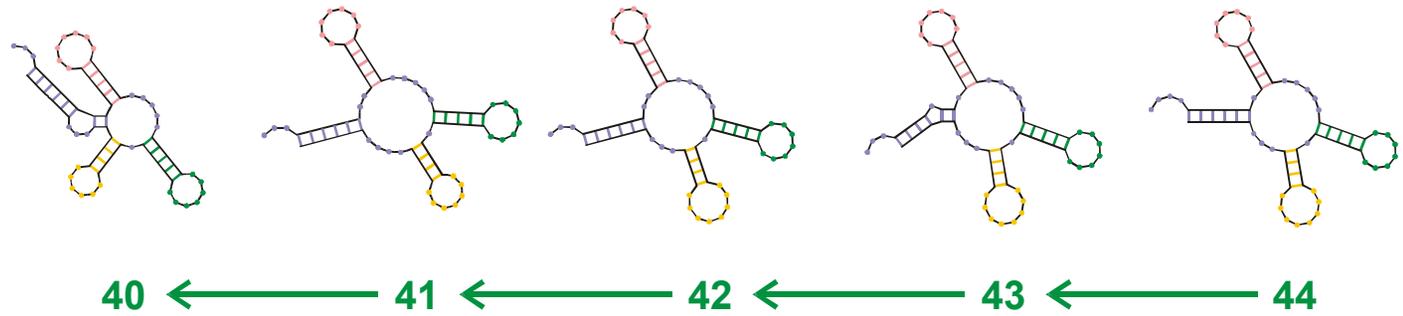
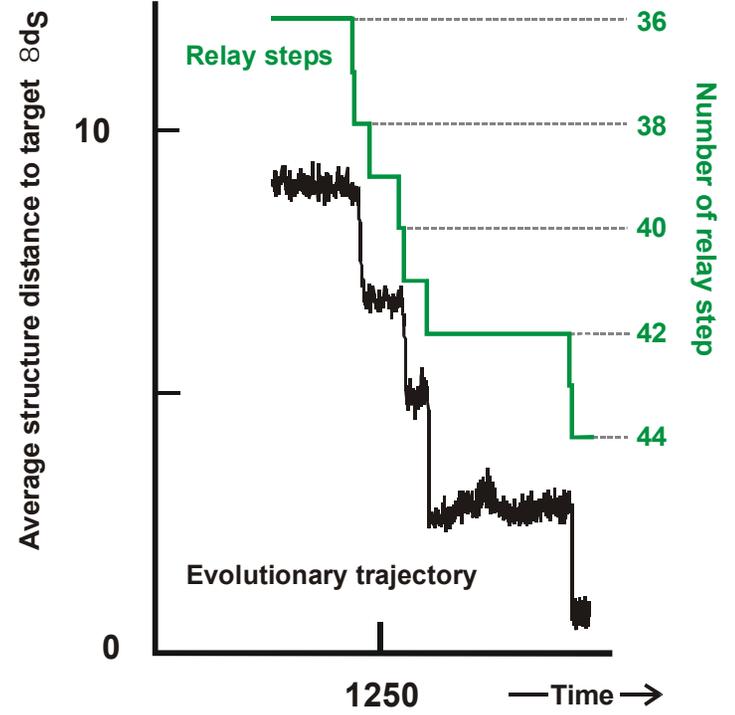
Reconstruction of the last step 43 \leftarrow 44



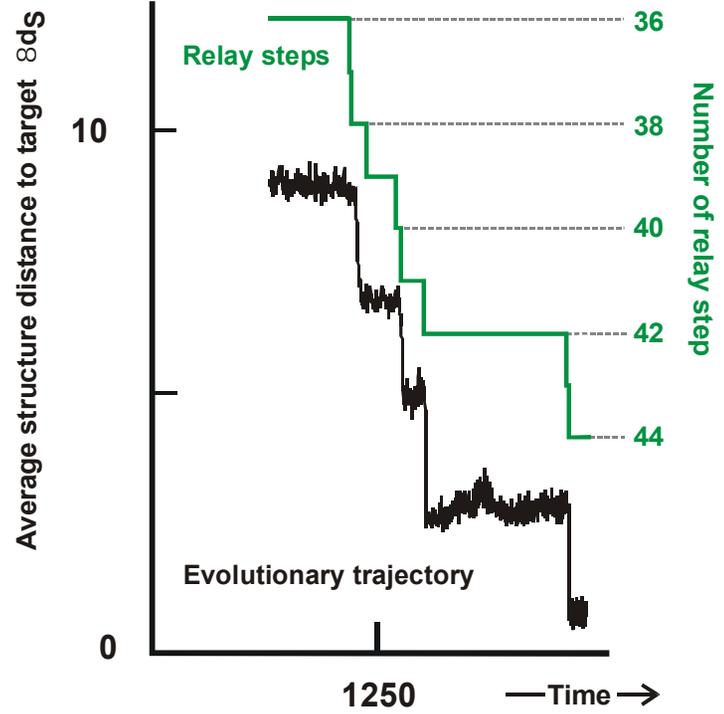
Reconstruction of last-but-one step 42 \checkmark 43 (\checkmark 44)



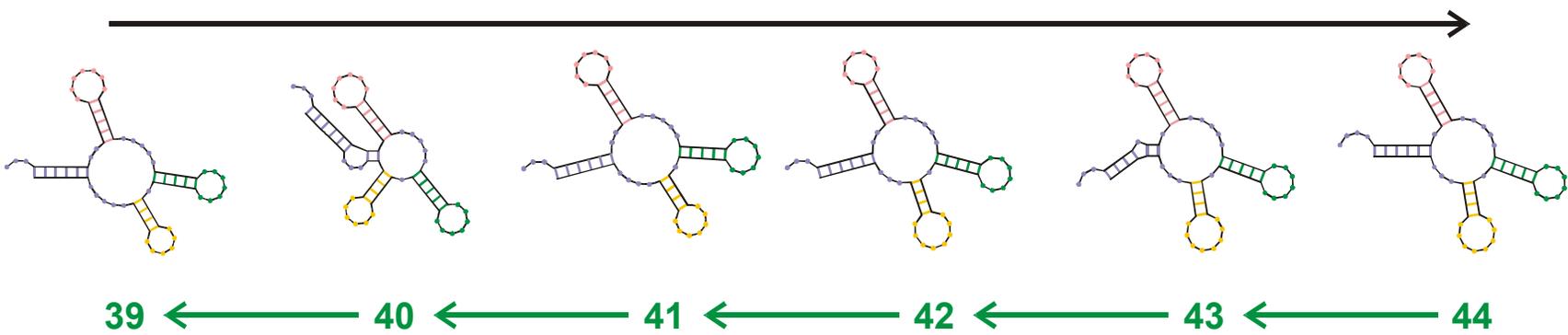
Reconstruction of step 41 š 42 (š 43 š 44)



Reconstruction of step 40 š 41 (š 42 š 43 š 44)

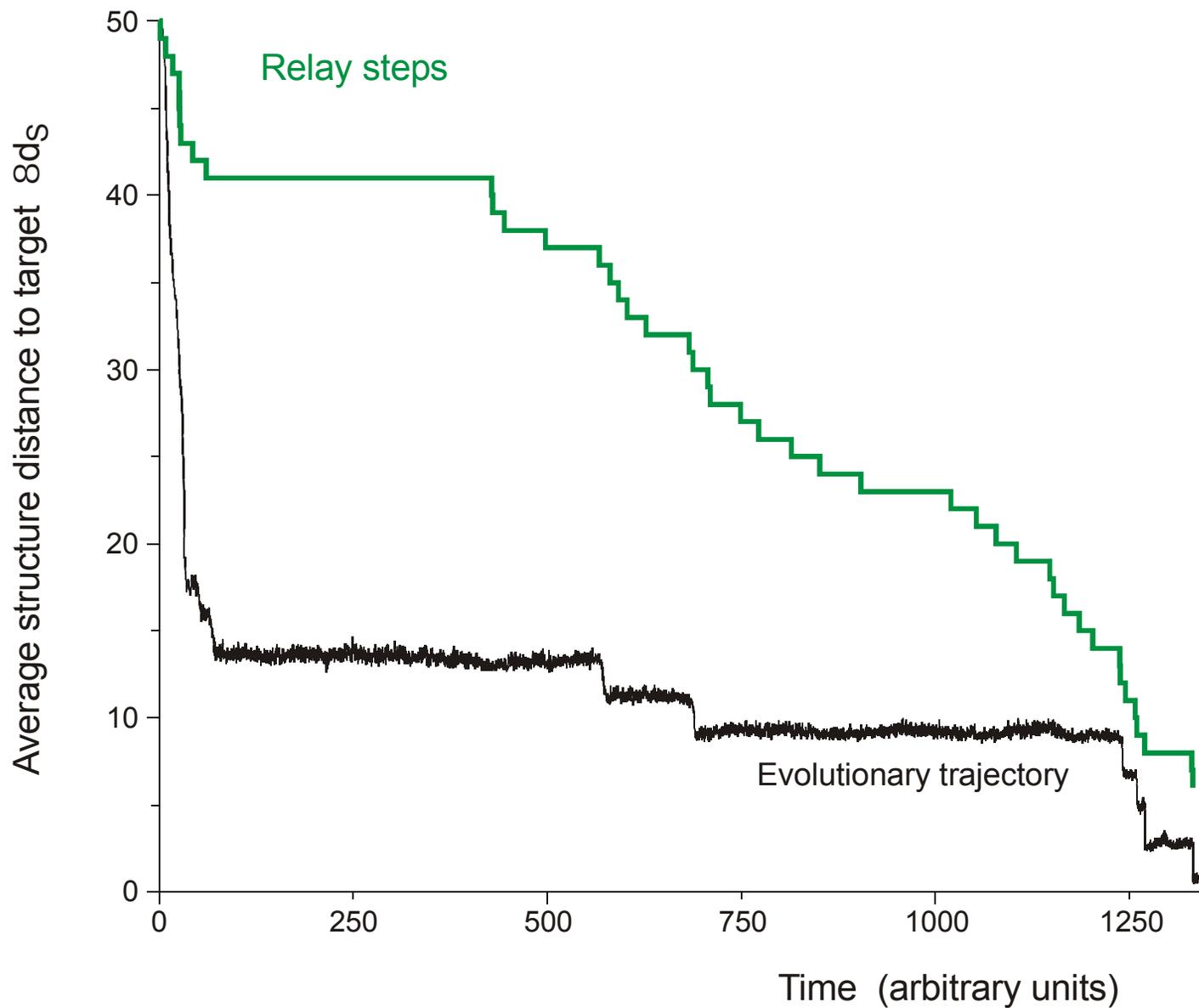


Evolutionary process

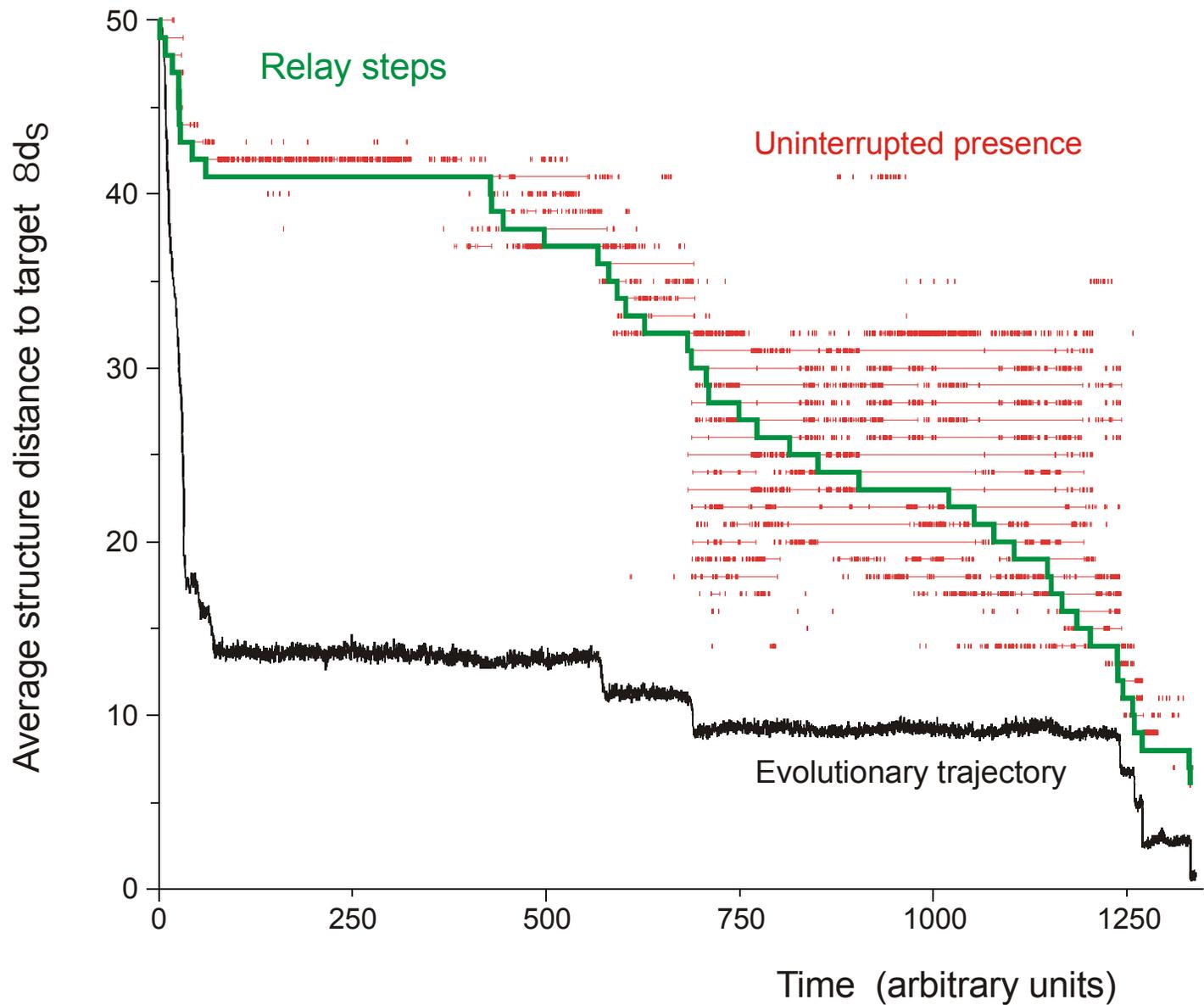


Reconstruction

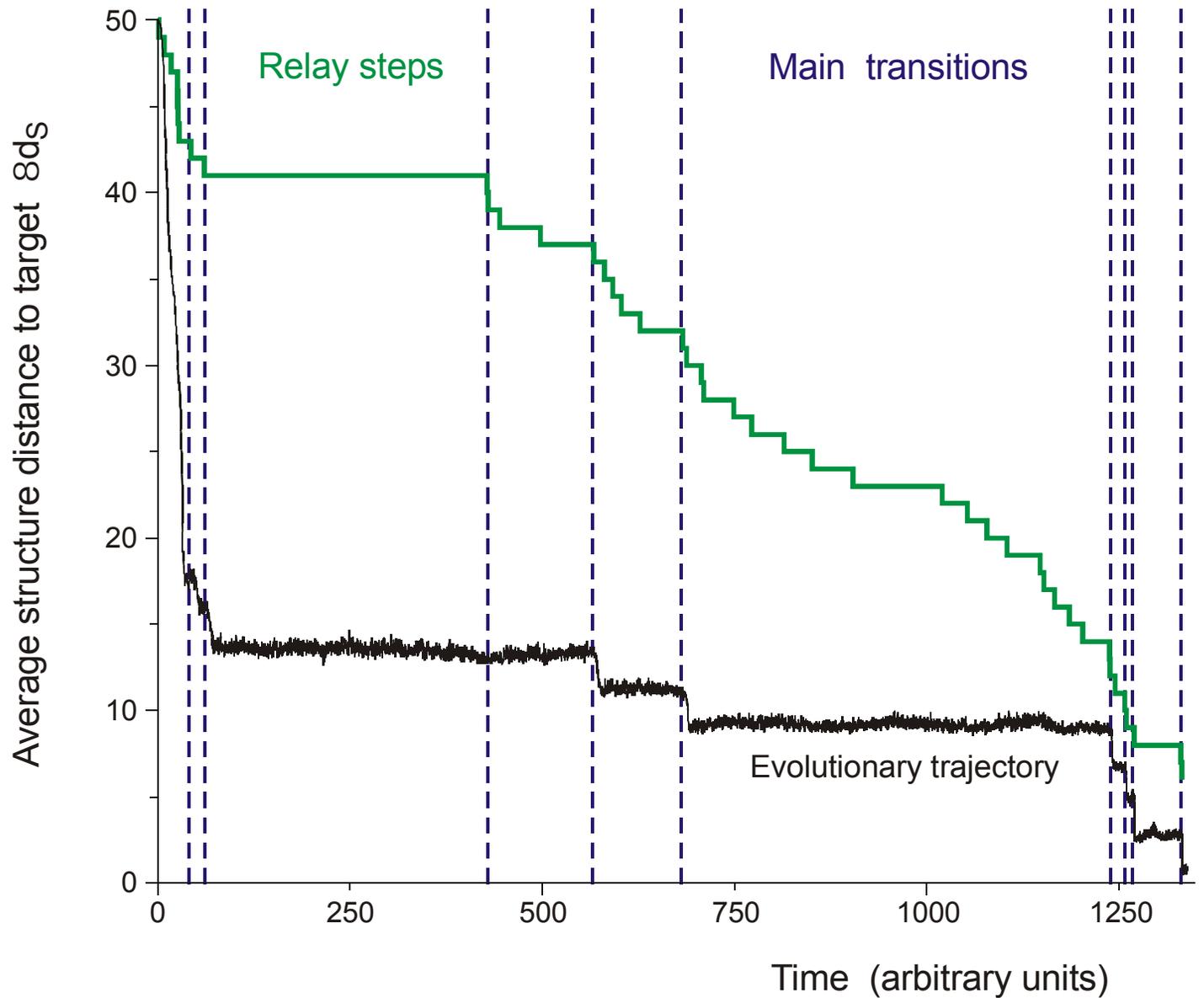
Reconstruction of the relay series



In silico optimization in the flow reactor: Trajectory and relay steps

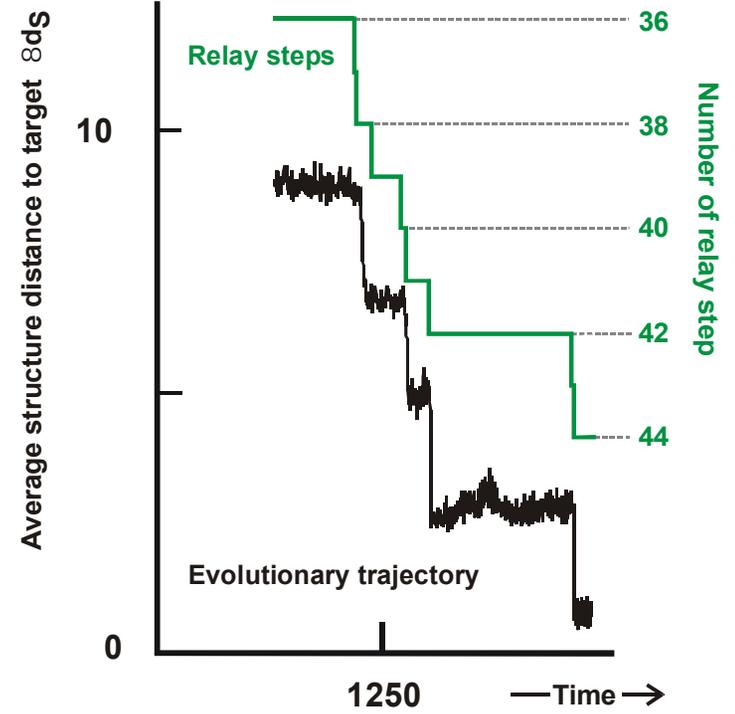
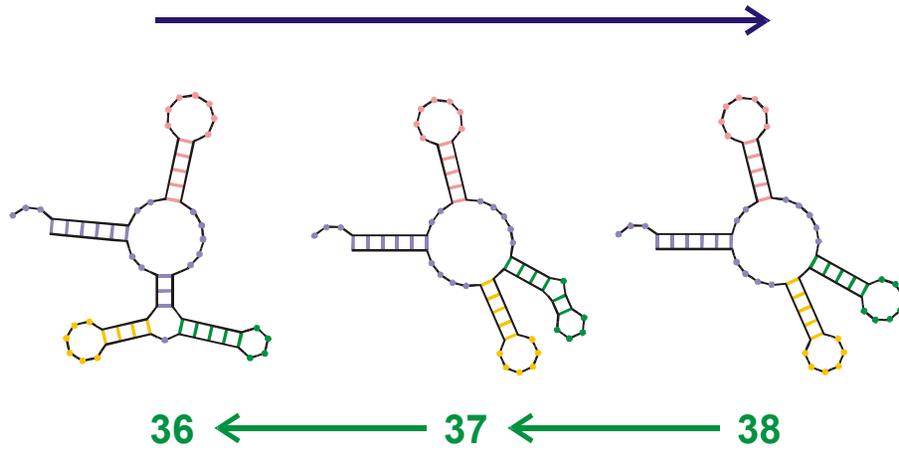


In silico optimization in the flow reactor: Uninterrupted presence



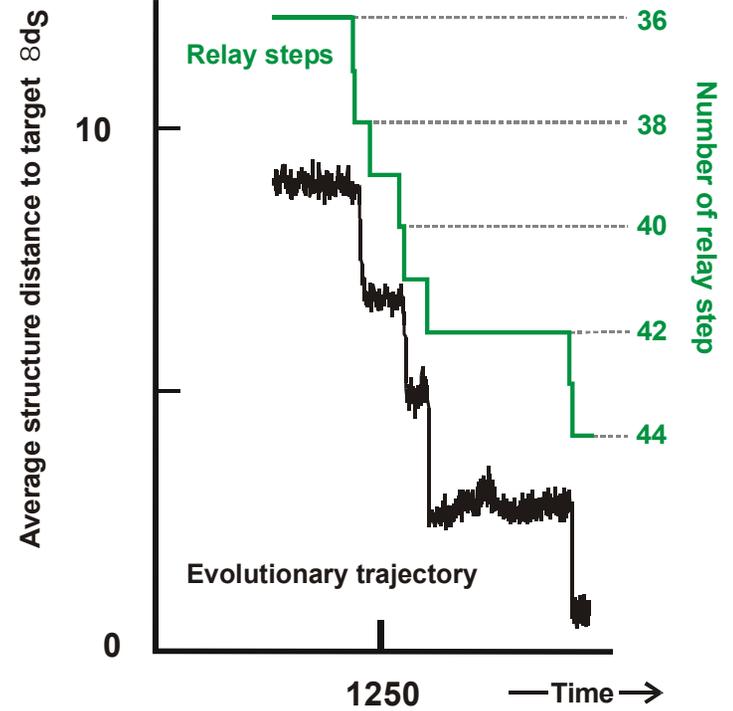
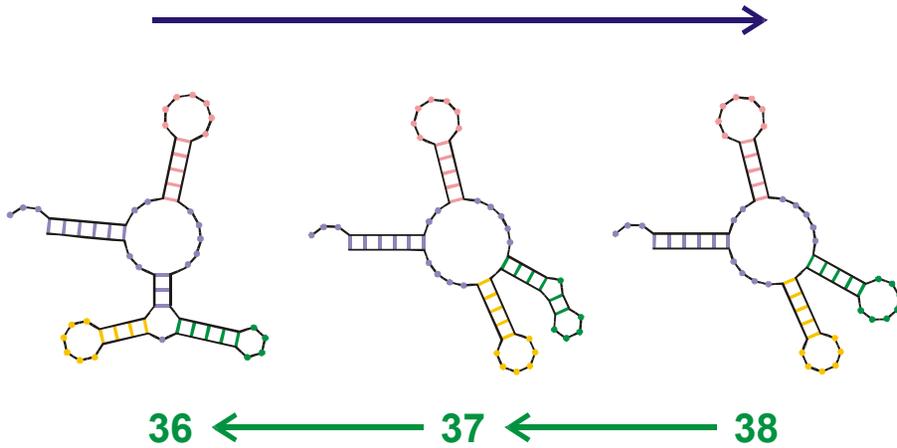
In silico optimization in the flow reactor: Main transitions

Main transition leading to clover leaf

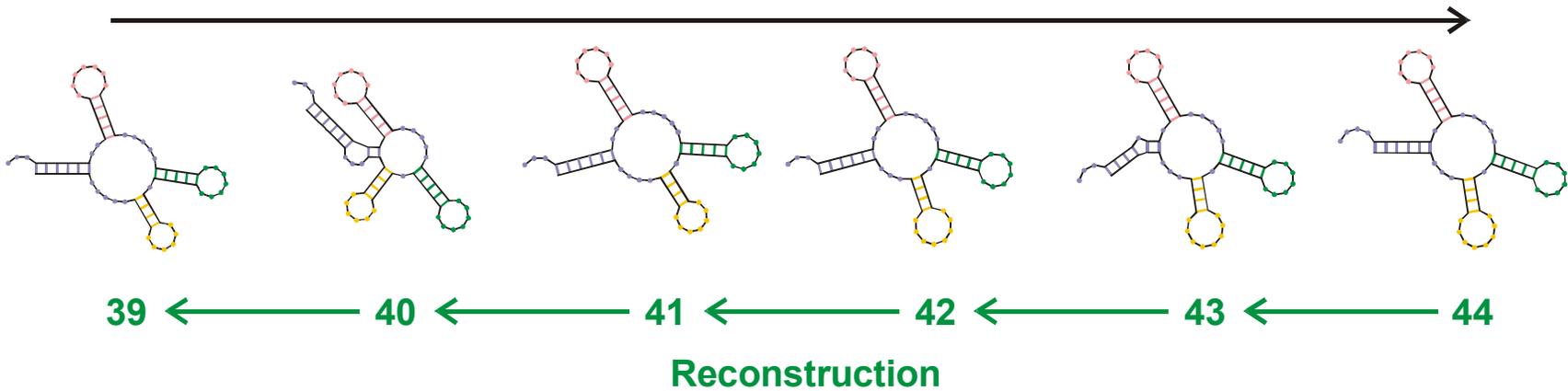


Reconstruction of a main transitions 36 \rightarrow 37 (\rightarrow 38)

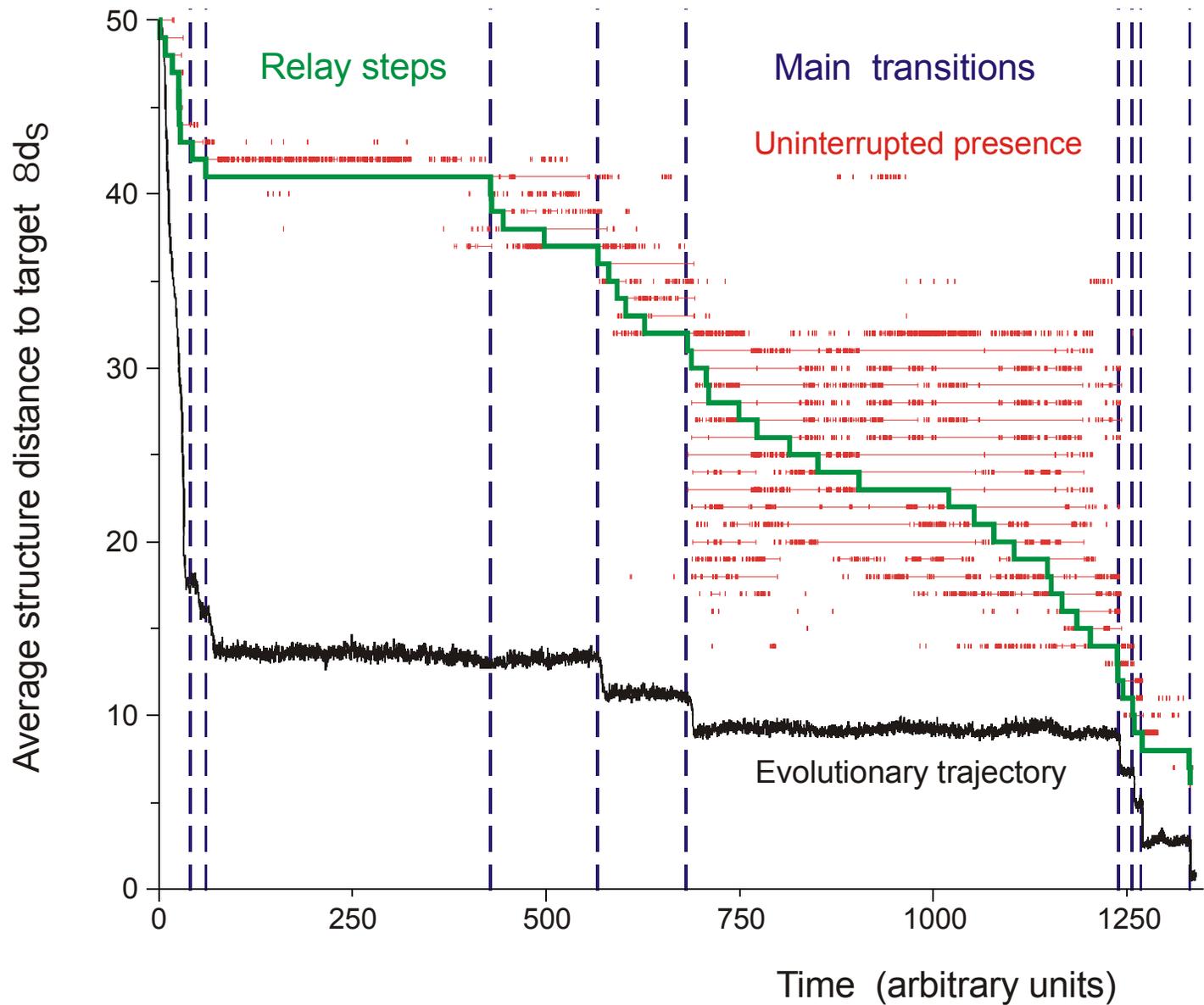
Main transition leading to clover leaf



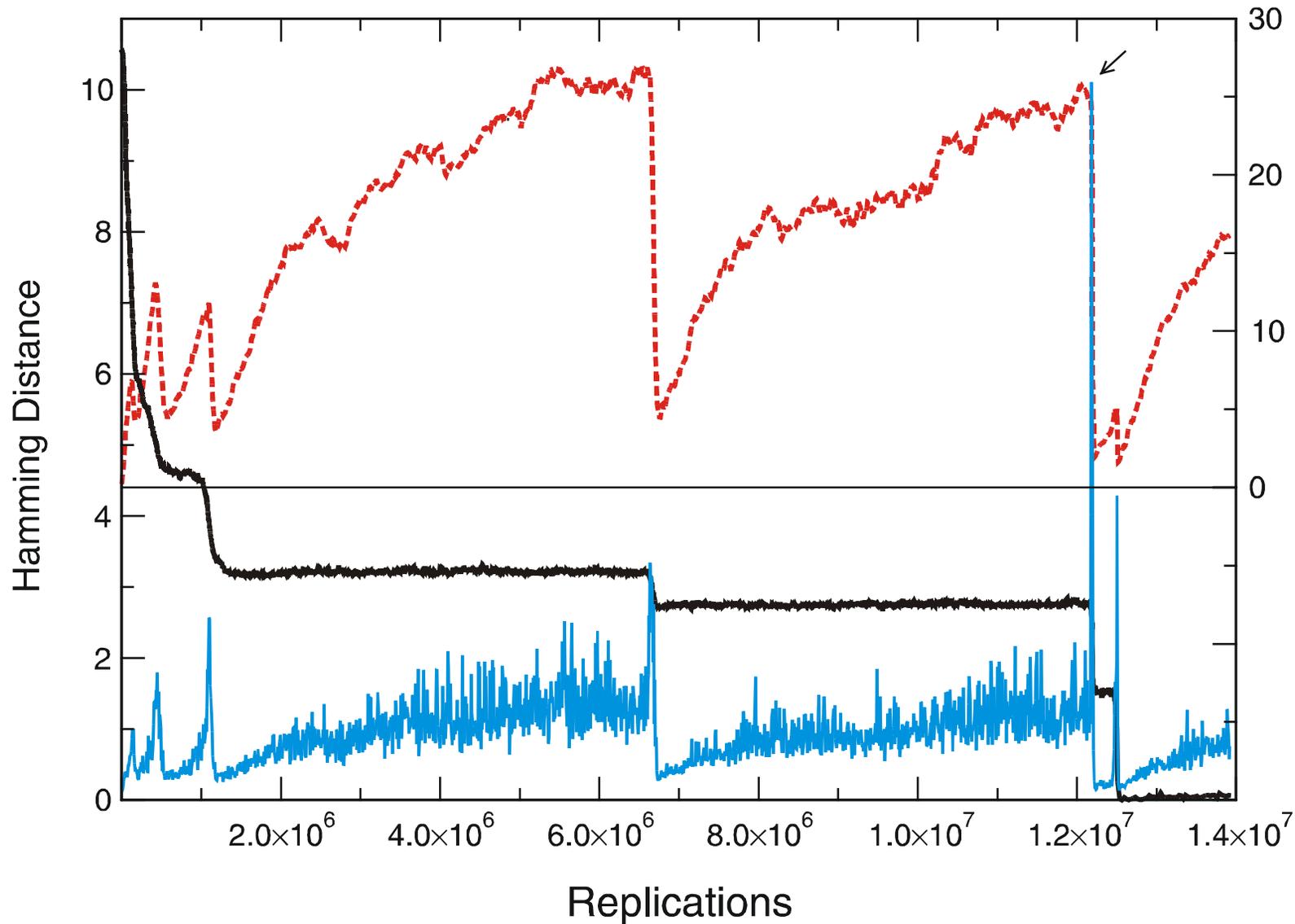
Evolutionary process



Final reconstruction 36 $\ddot{\sim}$ 44



In silico optimization in the flow reactor

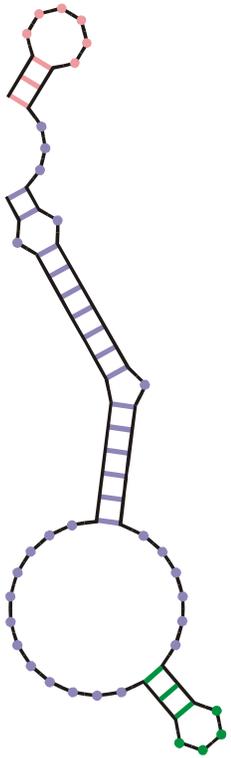


Variation in genotype space during optimization of phenotypes

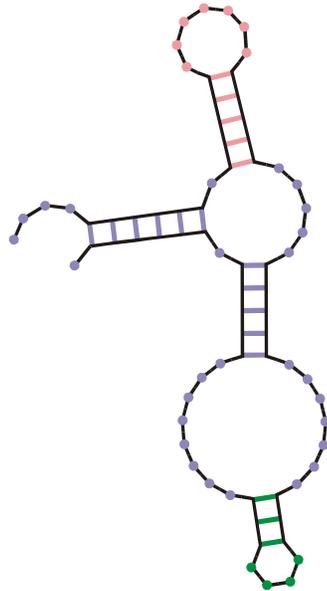
Statistics of evolutionary trajectories

Population size N	Number of replications < n _{rep} >	Number of transitions < n _{tr} >	Number of main transitions < n _{dtr} >
1 000	$(5.5 \pm [6.9, 3.1]) \times 10^7$	$92.7 \pm [80.3, 43.0]$	$8.8 \pm [2.4, 1.9]$
2 000	$(6.0 \pm [11.1, 3.9]) \times 10^7$	$55.7 \pm [30.7, 19.8]$	$8.9 \pm [2.8, 2.1]$
3 000	$(6.6 \pm [21.0, 5.0]) \times 10^7$	$44.2 \pm [25.9, 16.3]$	$8.1 \pm [2.3, 1.8]$
10 000	$(1.2 \pm [1.3, 0.6]) \times 10^8$	$35.9 \pm [10.3, 8.0]$	$10.3 \pm [2.6, 2.1]$
20 000	$(1.5 \pm [1.4, 0.7]) \times 10^8$	$28.8 \pm [5.8, 4.8]$	$9.0 \pm [2.8, 2.2]$
30 000	$(2.2 \pm [3.1, 1.3]) \times 10^8$	$29.8 \pm [7.3, 5.9]$	$8.7 \pm [2.4, 1.9]$
100 000	$(3 \pm [2, 1]) \times 10^8$	$24 \pm [6, 5]$	9 ± 2

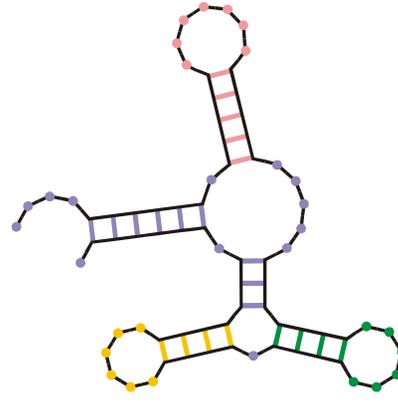
The number of **main transitions** or **evolutionary innovations** is constant.



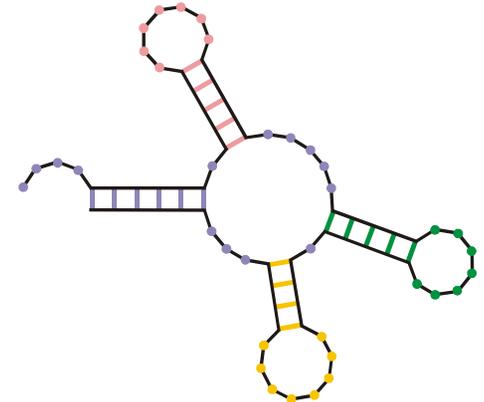
00



09



31

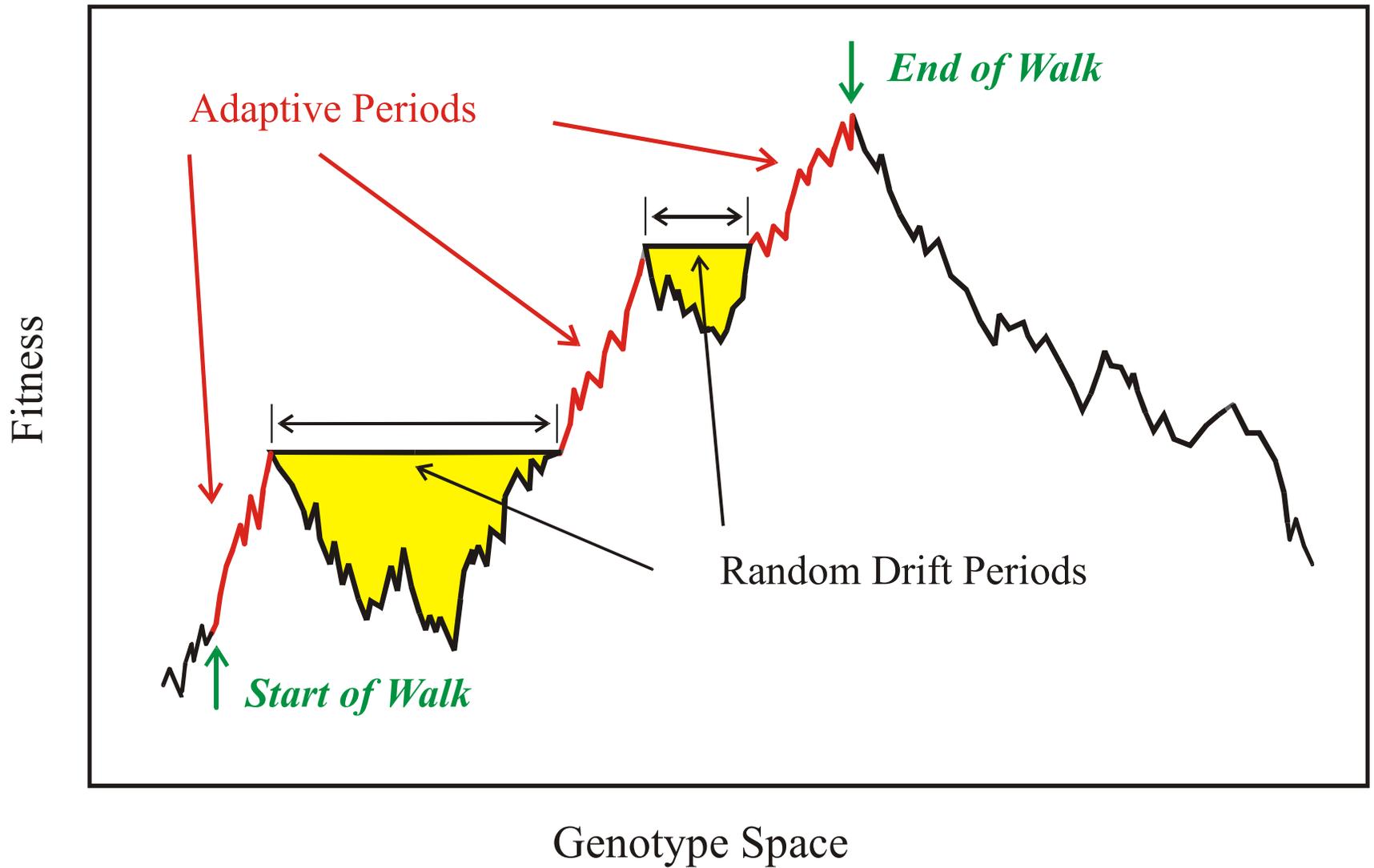


44

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 Q β phage

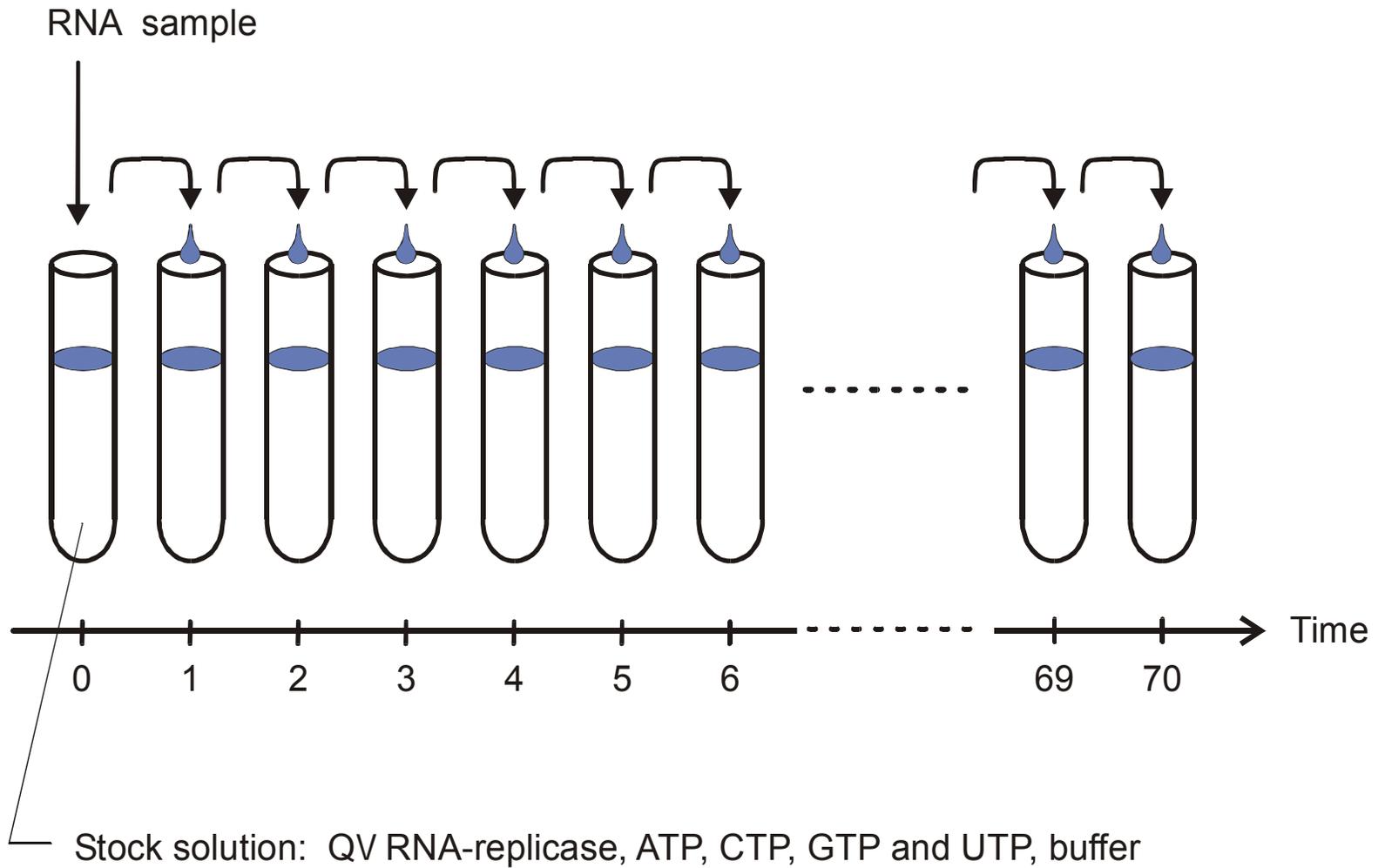
D.R.Mills, R.L.Peterson, S.Spiegelman, *An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule*. Proc.Natl.Acad.Sci.USA **58** (1967), 217-224

S.Spiegelman, *An approach to the experimental analysis of precellular evolution*. Quart.Rev.Biophys. **4** (1971), 213-253

C.K.Biebricher, *Darwinian selection of self-replicating RNA molecules*. Evolutionary Biology **16** (1983), 1-52

C.K.Biebricher, W.C. Gardiner, *Molecular evolution of RNA in vitro*. Biophysical Chemistry **66** (1997), 179-192

G.Strunk, T. Ederhof, *Machines for automated evolution experiments in vitro based on the serial transfer concept*. Biophysical Chemistry **66** (1997), 193-202



The serial transfer technique applied to RNA evolution *in vitro*

Reproduction of the original figure of the serial transfer experiment with Q β RNA

D.R.Mills, R.L.Peterson, S.Spiegelman,
*An extracellular Darwinian experiment
 with a self-duplicating nucleic acid
 molecule.* Proc.Natl.Acad.Sci.USA
58 (1967), 217-224

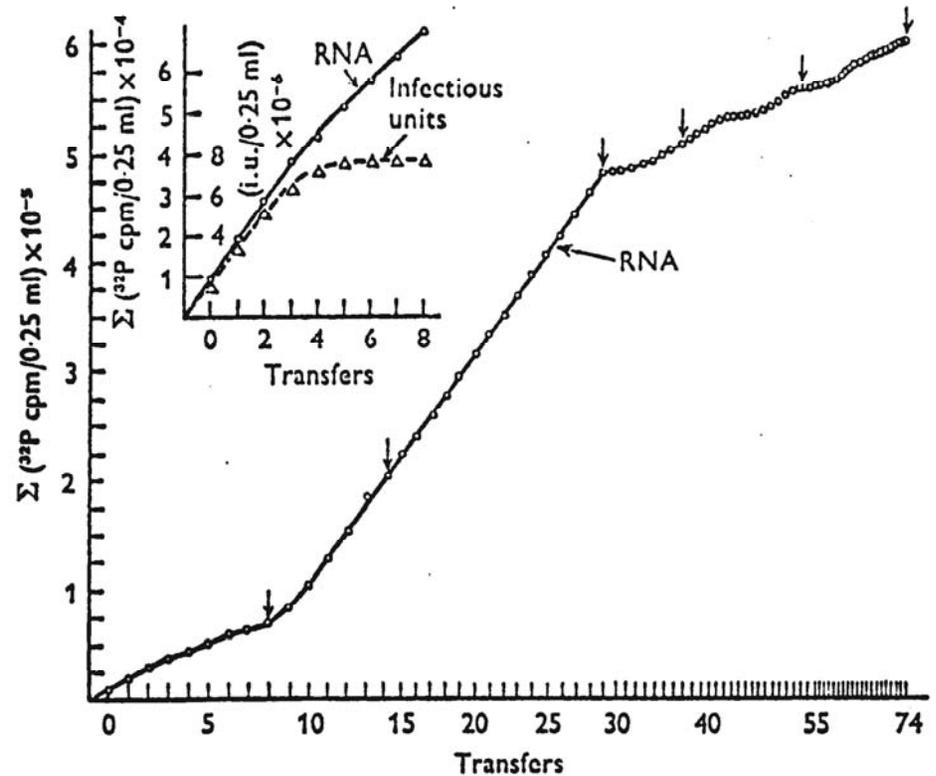
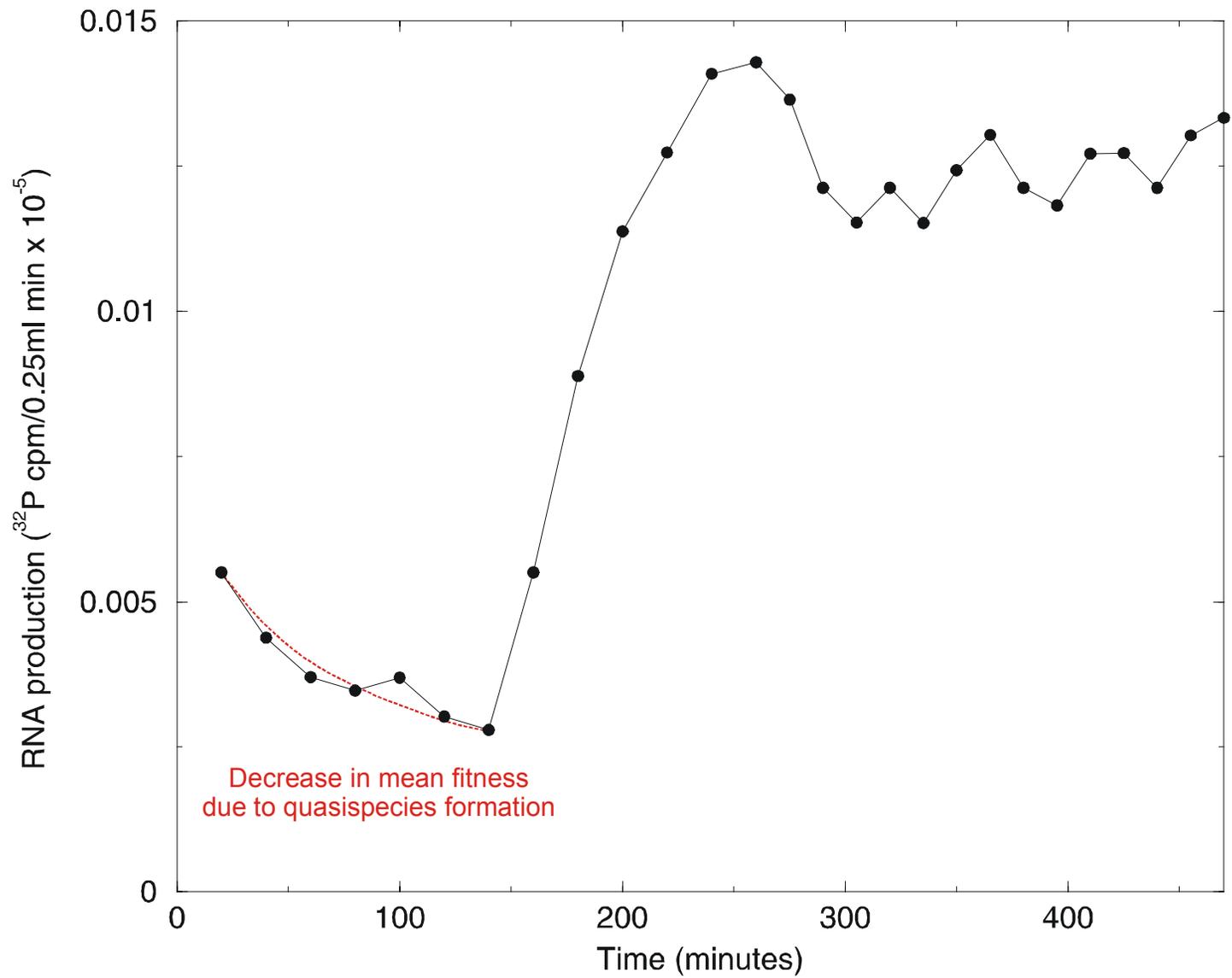


Fig. 9. Serial transfer experiment. Each 0.25 ml standard reaction mixture contained 40 μg of Q β replicase and ^{32}P -UTP. The first reaction (0 transfer) was initiated by the addition of 0.2 μg ts-1 (temperature-sensitive RNA) and incubated at 35 $^{\circ}\text{C}$ for 20 min, whereupon 0.02 ml was 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 ceases to appear after the 4th transfer (Mills *et al.* 1967).



The increase in RNA production rate during a serial transfer experiment

Bacterial Evolution

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants*. Science **272** (1996), 1802-1804

D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, M. Blot. *Genomic evolution during a 10,000-generation experiment with bacteria*. Proc.Natl.Acad.Sci.USA **96** (1999), 3807-3812

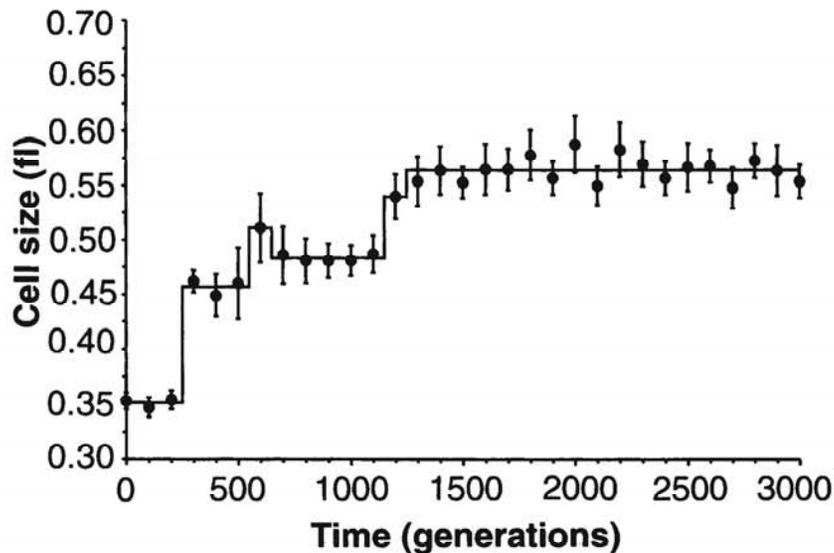


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (22). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

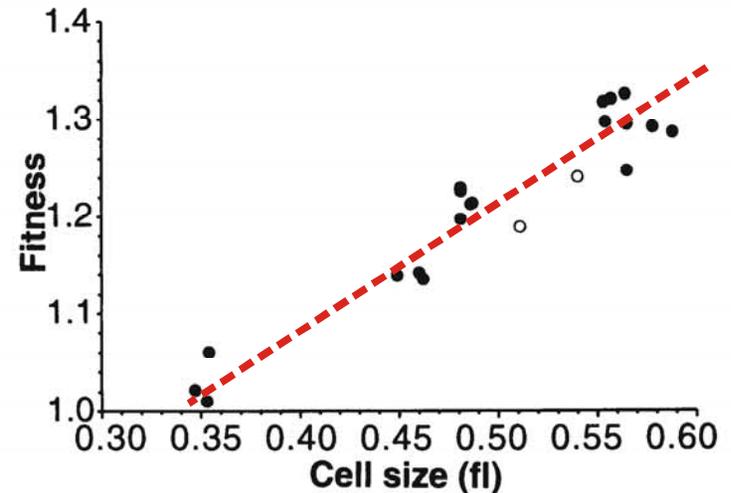
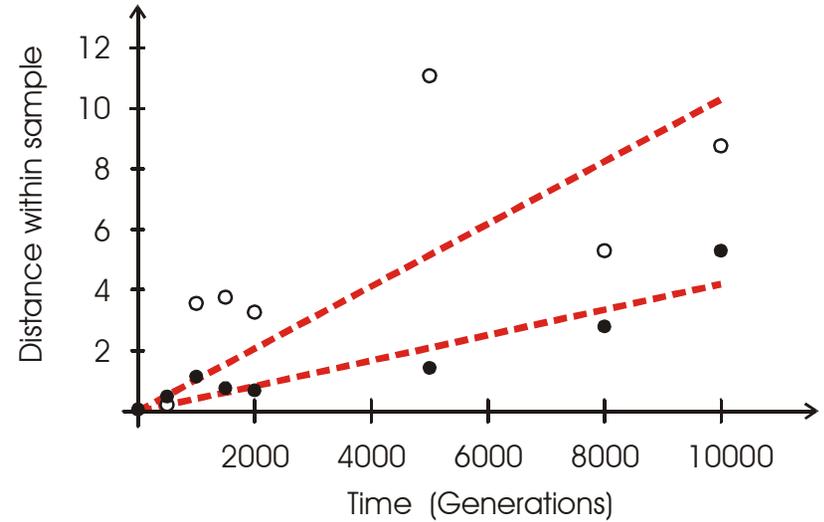
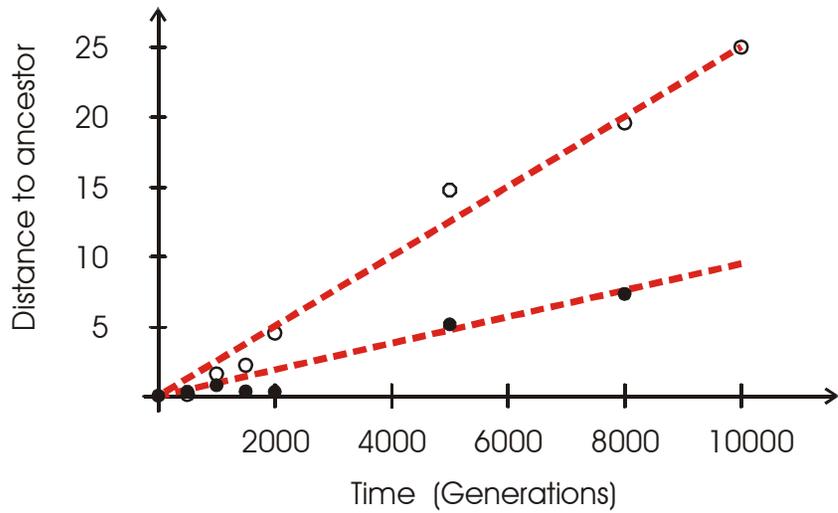


Fig. 2. Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (6, 7). The open symbols indicate the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

Epochal evolution of bacteria in serial transfer experiments under constant conditions

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants.* Science **272** (1996), 1802-1804



Variation of genotypes in a bacterial serial transfer experiment

D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, M. Blot. *Genomic evolution during a 10,000-generation experiment with bacteria*. Proc.Natl.Acad.Sci.USA **96** (1999), 3807-3812

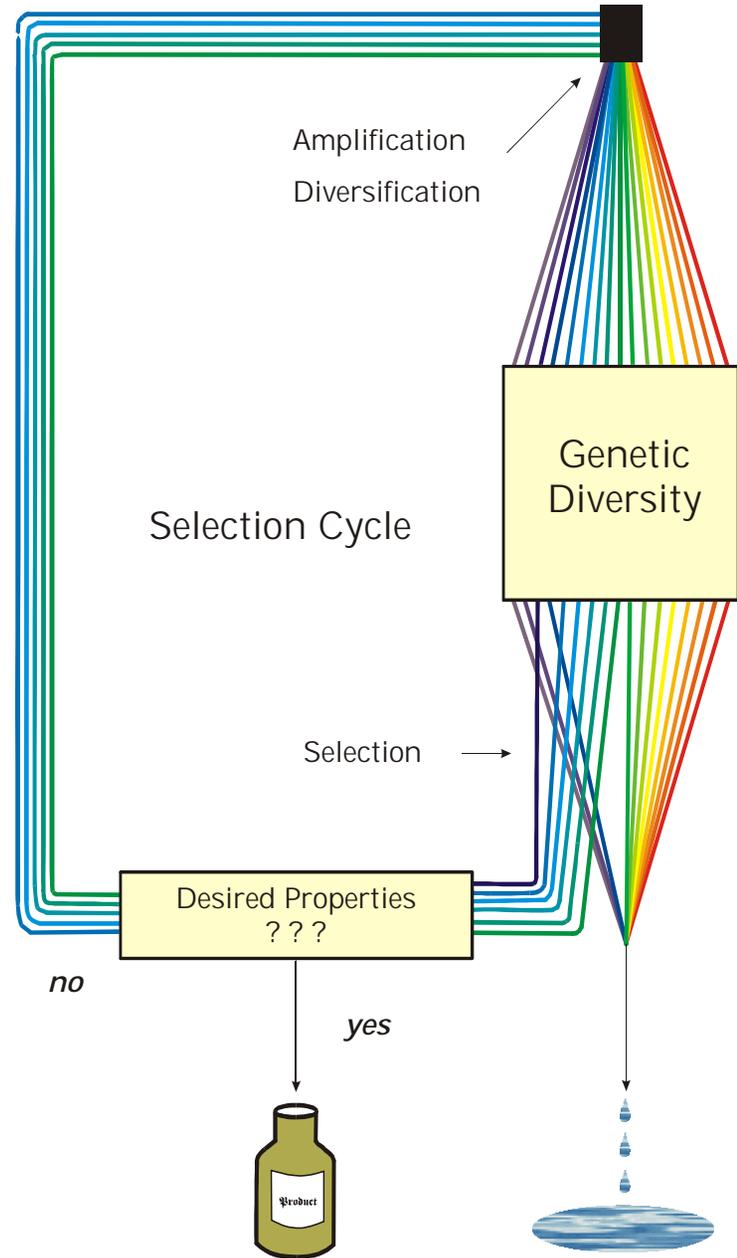
Evolutionary design of RNA molecules

D.B.Bartel, J.W.Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature **346** (1990), 818-822

C.Tuerk, L.Gold, *SELEX - Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science **249** (1990), 505-510

D.P.Bartel, J.W.Szostak, *Isolation of new ribozymes from a large pool of random sequences*. Science **261** (1993), 1411-1418

R.D.Jenison, S.C.Gill, A.Pardi, B.Poliski, *High-resolution molecular discrimination by RNA*. Science **263** (1994), 1425-1429

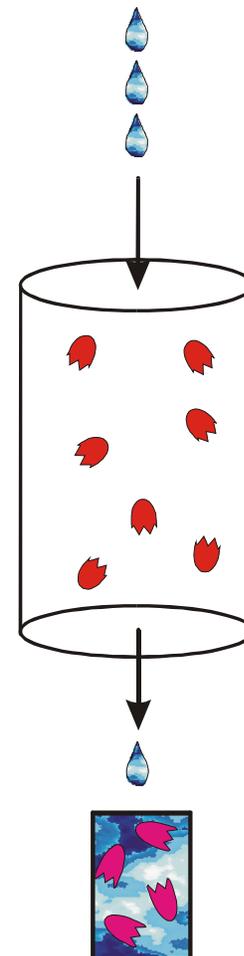
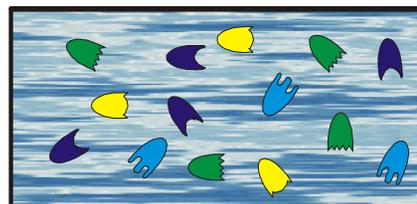
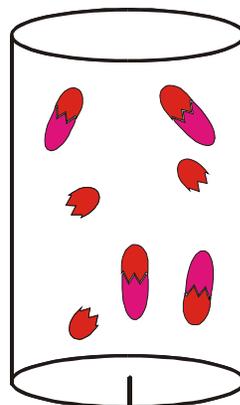
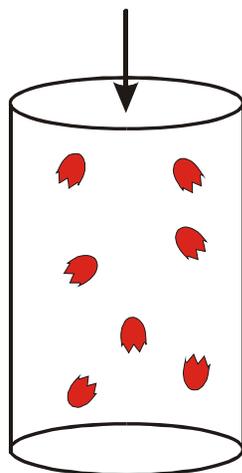
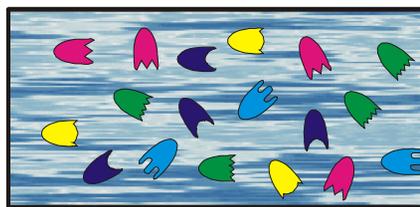


Selection cycle used in applied molecular evolution to design molecules with predefined properties

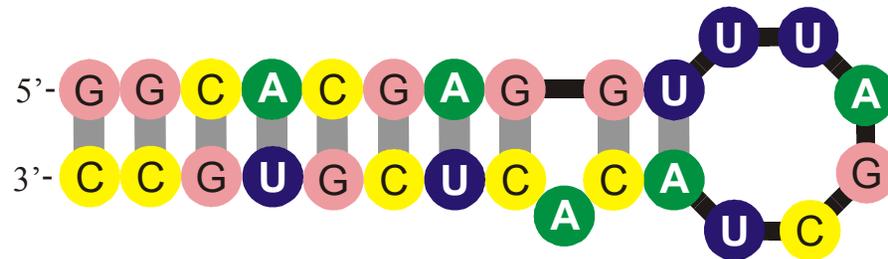
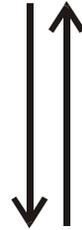
Retention of binders

Elution of binders

Chromatographic column

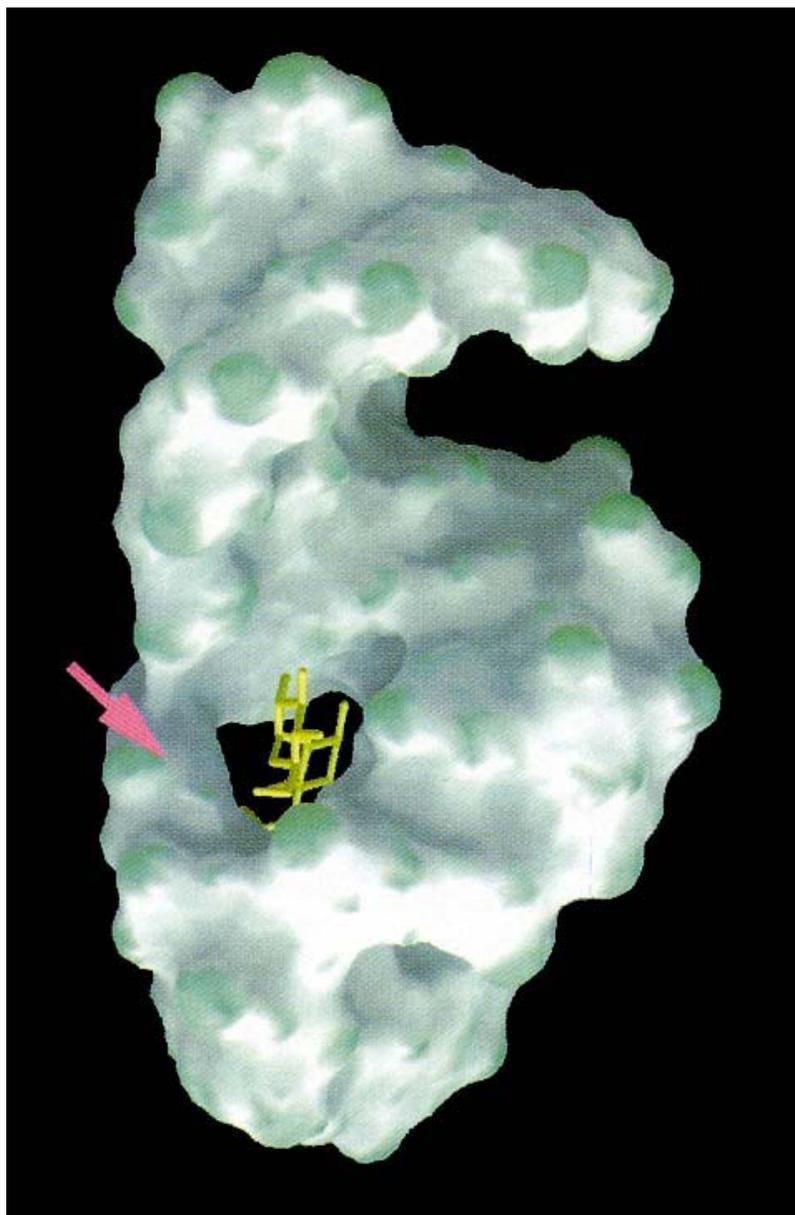


The SELEX technique for the evolutionary design of *aptamers*



Formation of secondary structure of the tobramycin binding RNA aptamer

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, *Chemistry & Biology* 4:35-50 (1997)

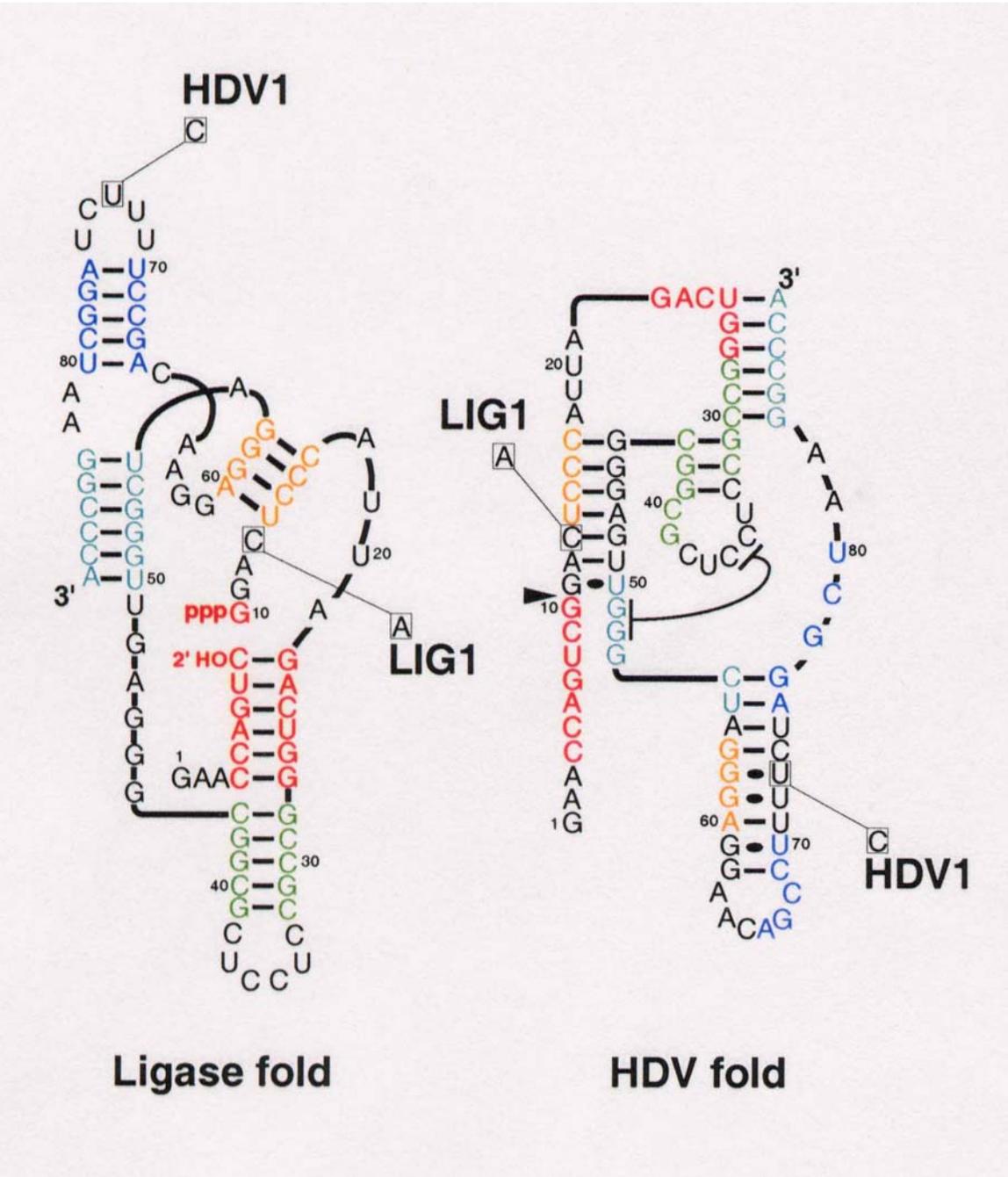


The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel,
Chemistry & Biology 4:35-50 (1997)

A ribozyme switch

E.A.Schultes, D.B.Bartel, *One sequence, two ribozymes: Implication for the emergence of new ribozyme folds*. Science **289** (2000), 448-452



The sequence at the *intersection*:

An RNA molecules which is 88 nucleotides long and can form both structures



S0092-8240(96)00089-4

GENERIC PROPERTIES OF COMBINATORY MAPS: NEUTRAL NETWORKS OF RNA SECONDARY STRUCTURES¹

■ CHRISTIAN REIDYS*, †, PETER F. STADLER*, ‡
 and PETER SCHUSTER*, ‡, §, ¶²

*Santa Fe Institute,
 Santa Fe, NM 87501, U.S.A.

†Los Alamos National Laboratory,
 Los Alamos, NM 87545, U.S.A.

‡Institut für Theoretische Chemie der Universität Wien,
 A-1090 Wien, Austria

§Institut für Molekulare Biotechnologie,
 D-07708 Jena, Germany

(E.mail: pks@tbi.univie.ac.at)

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 (λ). 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

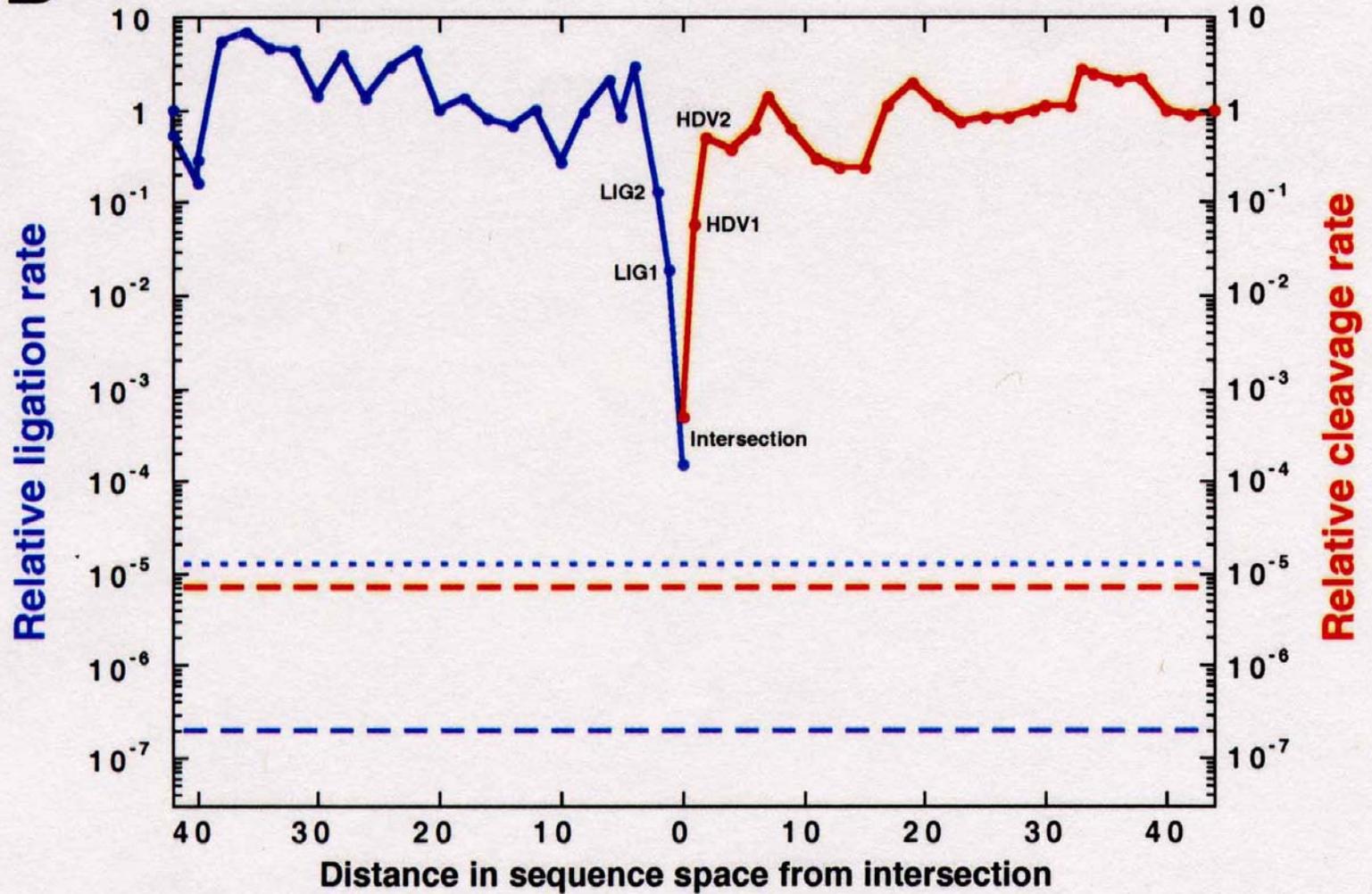
THEOREM 5. INTERSECTION-THEOREM. *Let s and s' be arbitrary secondary structures and $C[s], C[s']$ their corresponding compatible sequences. Then,*

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

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 $f(s, s') \cong D_m$ operates on the set of all positions $\{x_1, \dots, x_n\}$. 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.

Reference for the definition of the intersection and the proof of the *intersection theorem*

B

Two neutral walks through sequence space with conservation of structure and catalytic activity

From sequences to shapes and back: a case study in RNA secondary structures

PETER SCHUSTER^{1,2,3}, WALTER FONTANA³, PETER F. STADLER^{2,3}
AND IVO L. HOFACKER²

¹ Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, PF 100813, D-07708 Jena, Germany

² Institut für Theoretische Chemie, Universität Wien, Austria

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.

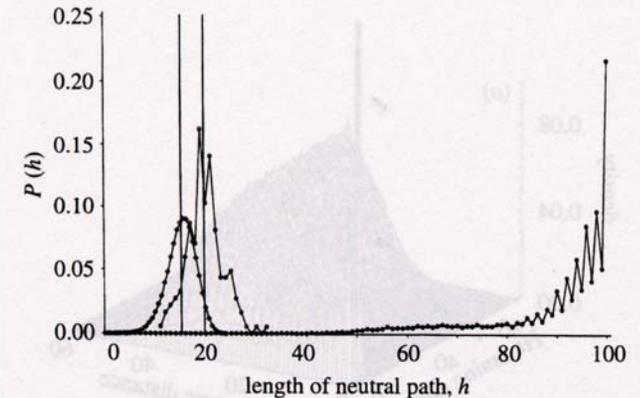


Figure 4. Neutral paths. A neutral path is defined by a series of nearest neighbour sequences that fold into identical structures. Two classes of nearest neighbours are admitted: neighbours of Hamming distance 1, which are obtained by single base exchanges in unpaired stretches of the structure, and 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 *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).

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