Die Evolution molekularer Phänotypen

Von Nukleinsäuren zu metabolischen Netzwerken

Peter Schuster

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

26. Berhard Rensch - Vorlesung

Münster, 14.06.2005

Web-Page for further information:

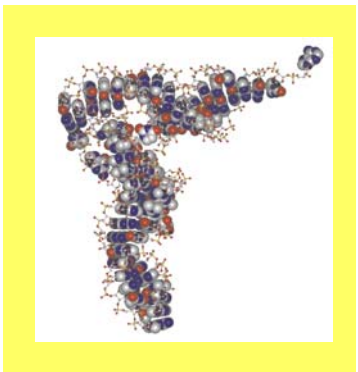
http://www.tbi.univie.ac.at/~pks

Genotyp, Genom

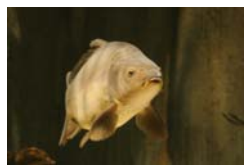
GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

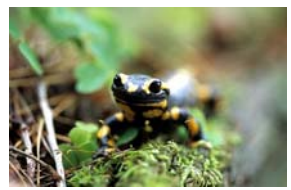
Biologische Strukturbildung

Hochspezifische Umweltbedingungen



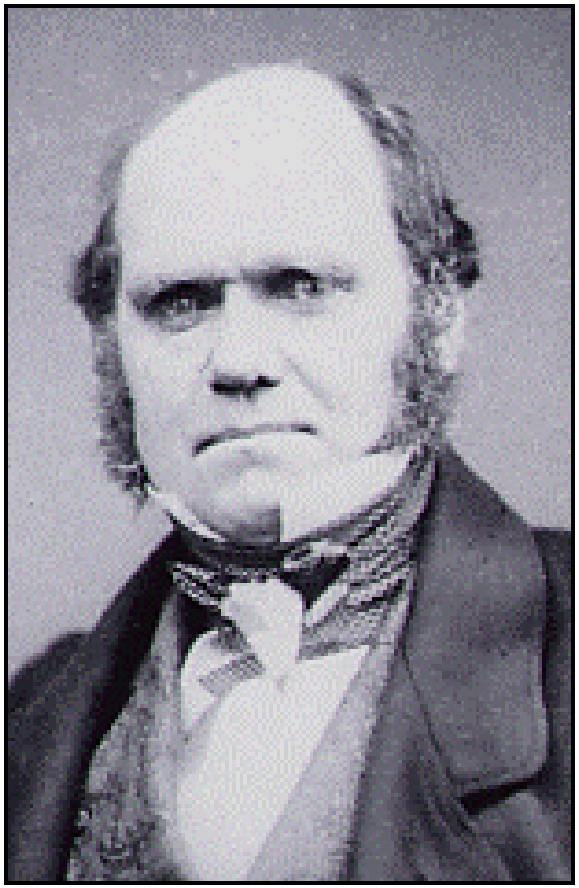
Phänotyp



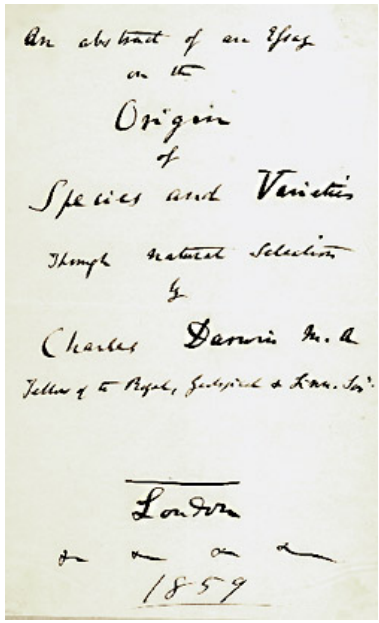


1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit
2. Ribonukleinsäuremoleküle als Phänotypen
3. Evolution im Reagenzglas und im Computer
4. Genetische und metabolische Netzwerke
5. Ein minimaler Metabolismus als Phänotyp
6. Evolution und Komplexität

1. **Der Darwinsche Mechanismus und seine universelle Anwendbarkeit**
2. Ribonukleinsäuremoleküle als Phänotypen
3. Evolution im Reagenzglas und im Computer
4. Genetische und metabolische Netzwerke
5. Ein minimaler Metabolismus als Phänotyp
6. Evolution und Komplexität



Charles Robert Darwin, 1809-1882



Earlier abstract of the
‚Origin of Species‘

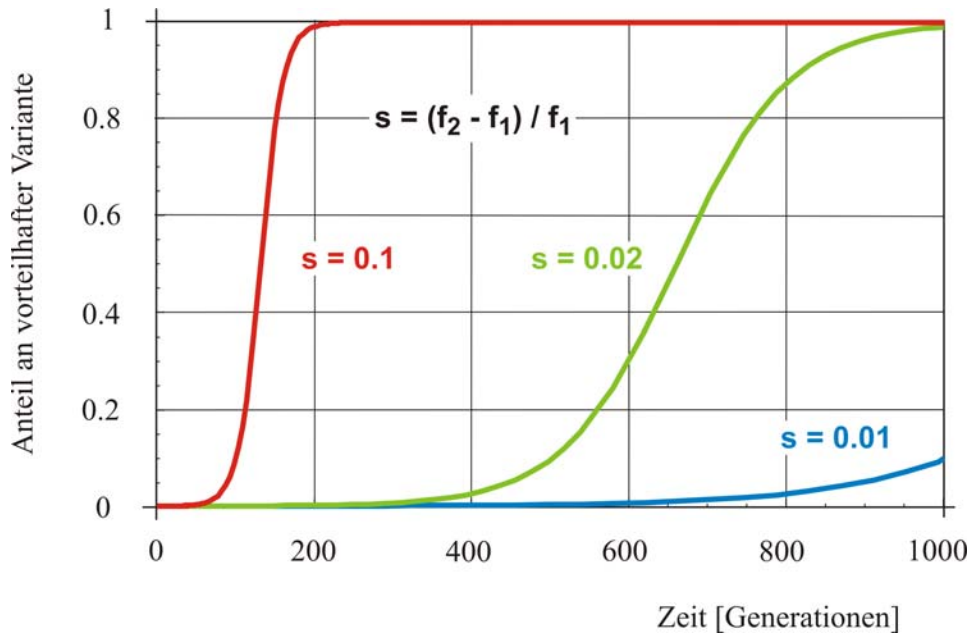
Alfred Russell Wallace, 1823-1913

Die beiden Konkurrenten bei der Formulierung des Selektionsprinzips

Darwinscher Evolutionsmechanismus

Vermehrung und Vererbung

Variation des Genotyps ✚ Selektion des Phänotyps



Deterministische Selektion einer Variante mit höherer Fitness

Ausgangssituation: Ein Exemplar I_2 in einer Population von zehntausend I_1

Vier Missverständnisse bei der Interpretation des Darwinschen Prinzips

1. Der Darwinsche Mechanismus macht statistische Aussagen über Populationen und Varianten aber nicht über Individuen.

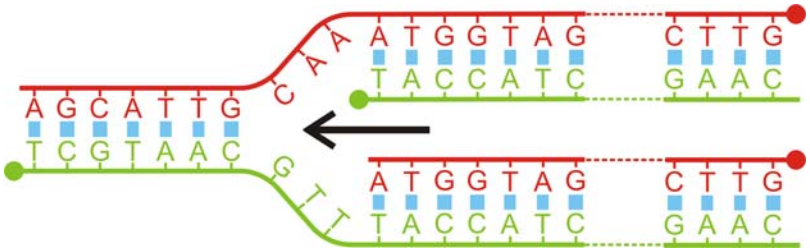
2. Als Kriterium für das Nicht-Aussterben einer Variante dient nur die Fitness, welche die Zahl der Nachkommen relativ zur mittleren Nachkommenzahl der Gesamtpopulation misst.

3. Die Natur befindet sich in einem dynamischen Prozess und Teilsysteme können weit von einem durch Variation und Selektion erreichbaren Optimum entfernt sein.

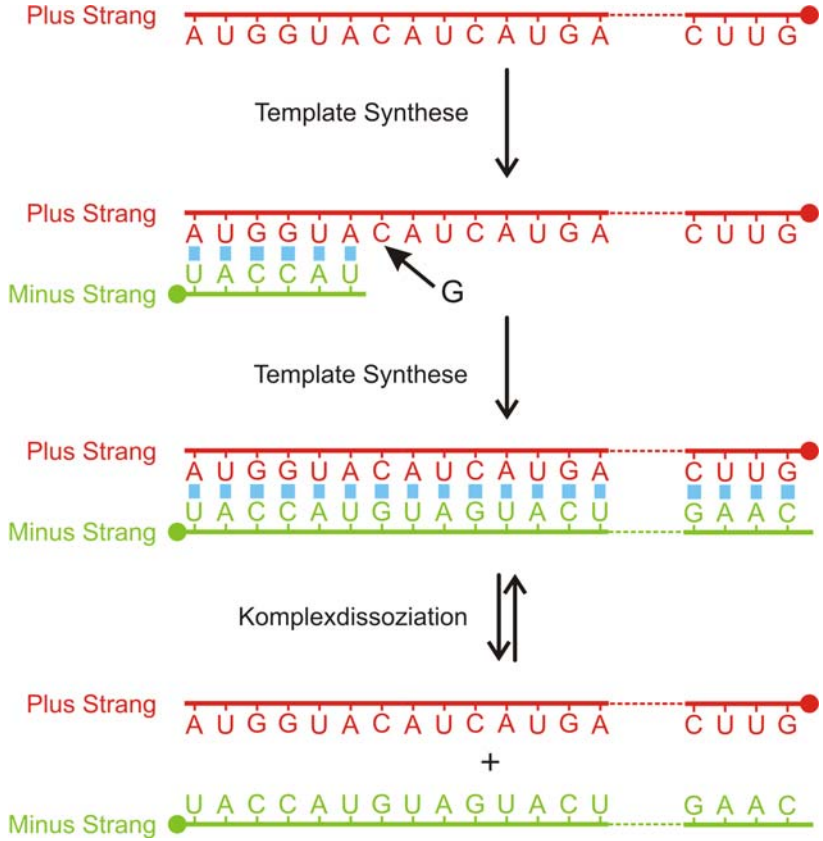
4. Der Darwinsche Mechanismus ist nicht das einzige Prinzip der biologischen Evolution. In den als „***Major Transitions***" bezeichneten Phasen der biologischen Evolution wurde die Konkurrenz von Varianten durch Kooperation außer Kraft gesetzt.

Erfolgsrezept des Darwinschen Mechanismus

Es zählt nur die Zahl der Nachkommen und deshalb wirken Vererbung, Variation und Selektion in Populationen von ganz einfachen und höchst komplexen Individuen in gleicher Weise.

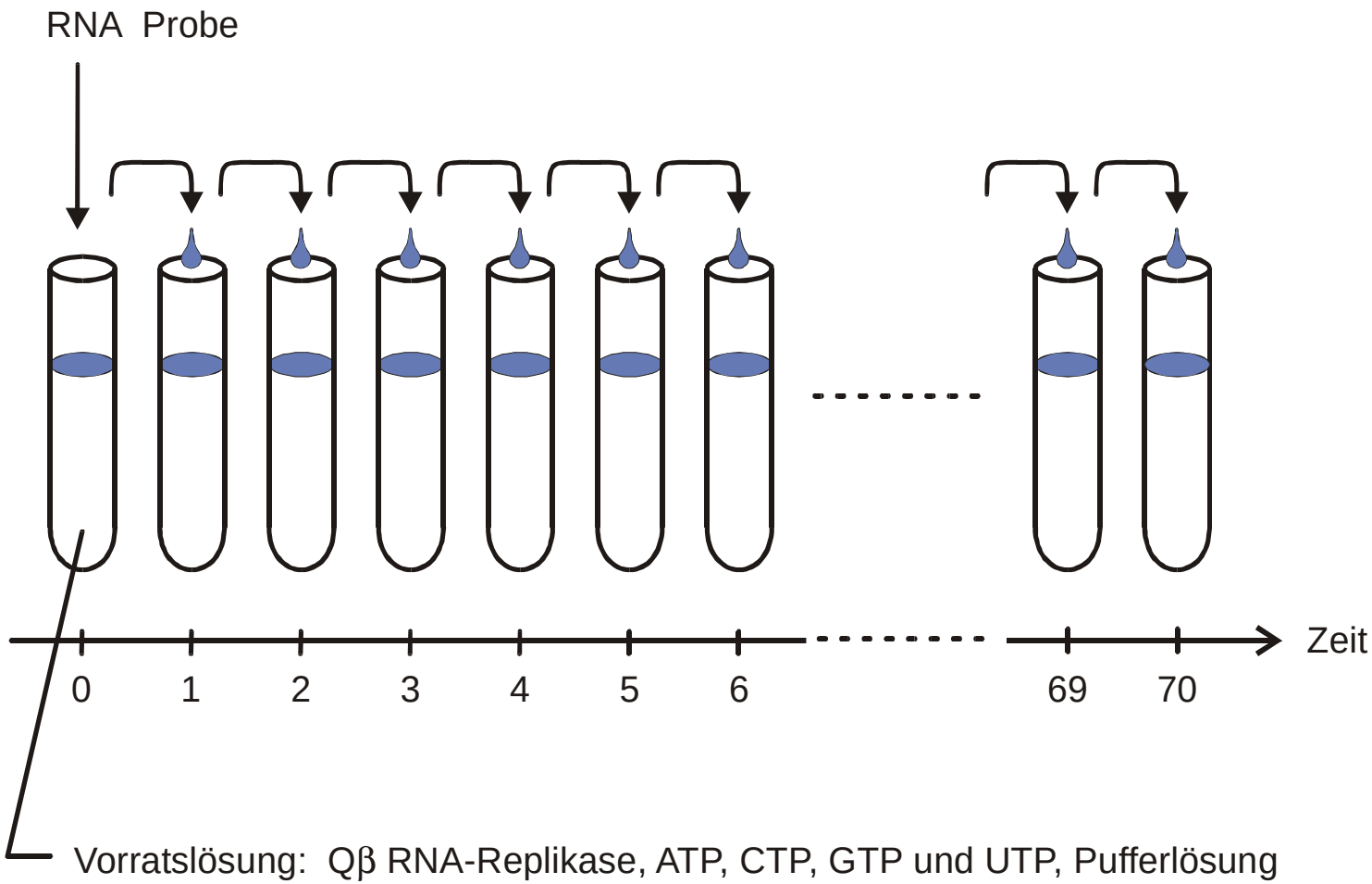


Replikationsgabel (DNA)



Komplementäre Synthese (RNA)

Zwei Mechanismen zur Vermehrung von Nukleinsäuremolekülen



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

Reproduktion der Abbildung des ersten Serial-Transfer-Experiments mit Qβ RNA

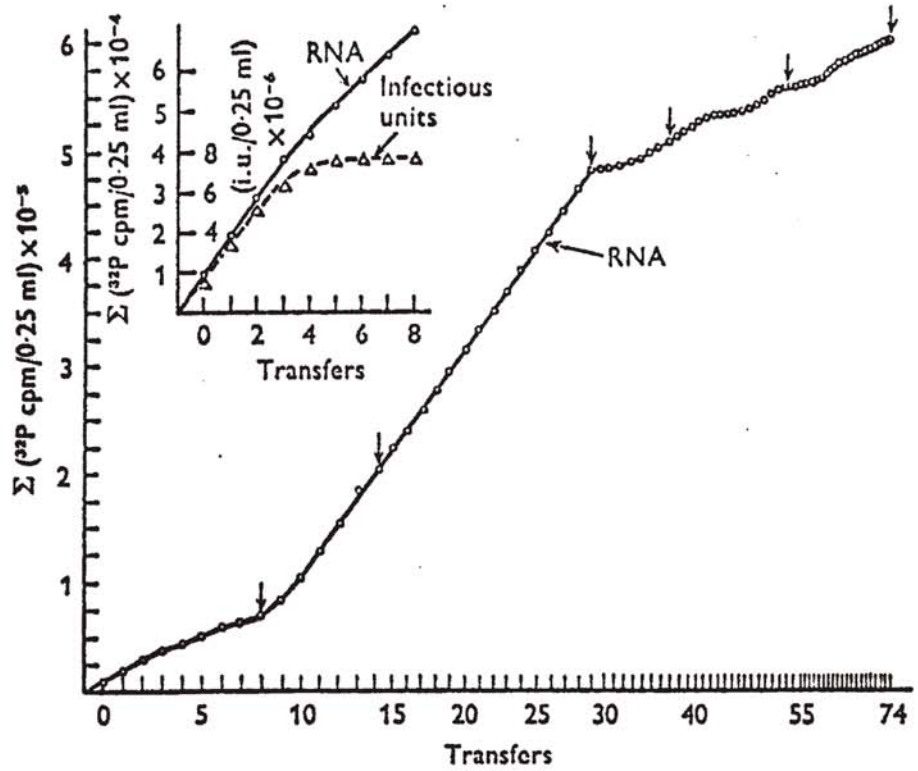
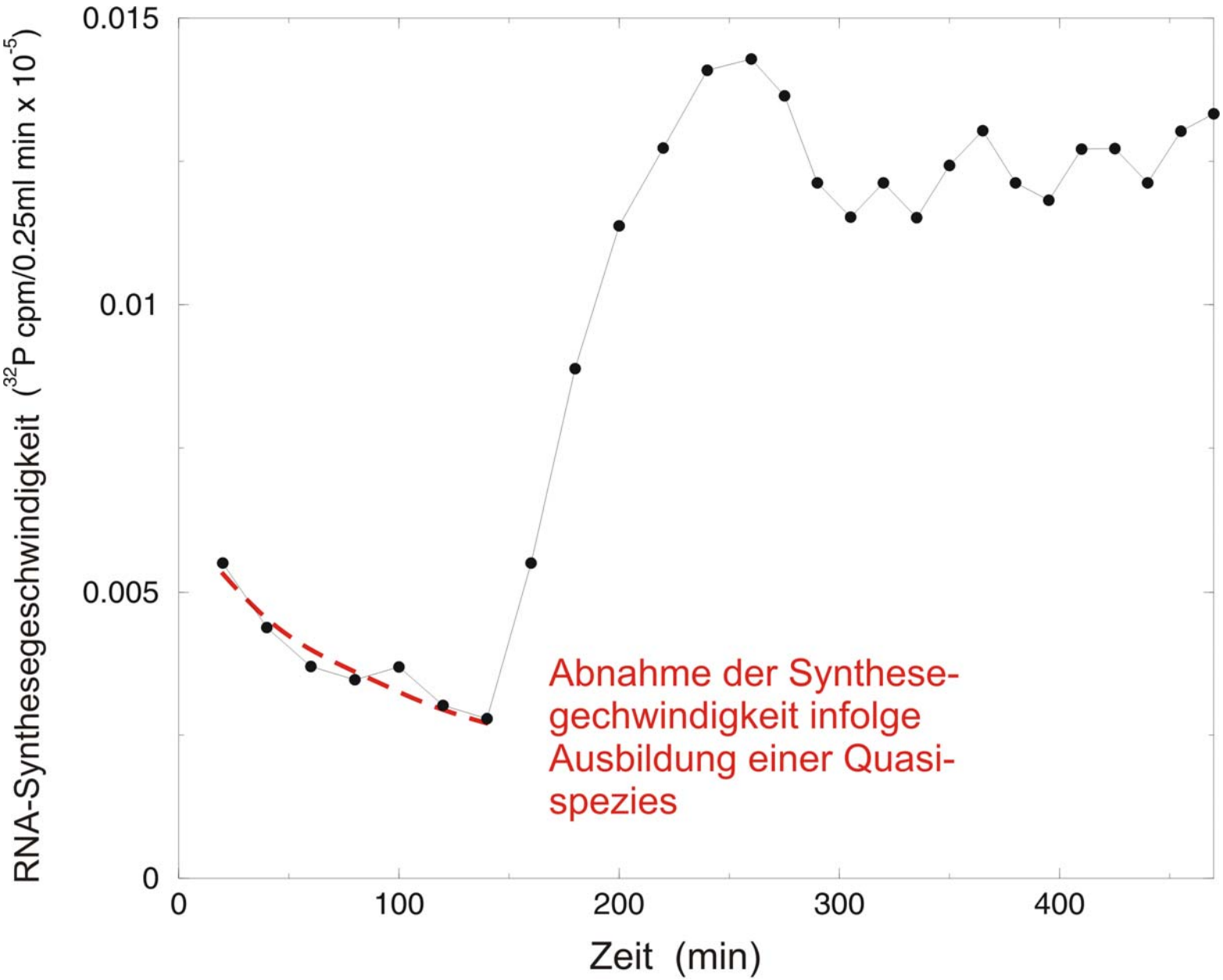


Fig. 9. Serial transfer experiment. Each 0·25 ml standard reaction mixture contained 40 μg of Qβ replicase and ^{32}P-UTP. The first reaction (0 transfer) was initiated by the addition of 0·2 μg ts-1 (temperature-sensitive RNA) and incubated at 35 °C for 20 min, whereupon 0·02 ml was drawn for counting 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).

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


RNA-Synthesegeschwindigkeit (^{32}P cpm/0.25ml min x 10^{-5})
0.015
0.01
0.005
0
0
100
200
300
400
Zeit (min)
Abnahme der Synthese-
gechwindigkeit infolge
Ausbildung einer Quasi-
spezies

1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit

2. **Ribonukleinsäuremoleküle als Phänotypen**

3. Evolution im Reagenzglas und im Computer

4. Genetische und metabolische Netzwerke

5. Ein minimaler Metabolismus als Phänotyp

6. Evolution und Komplexität

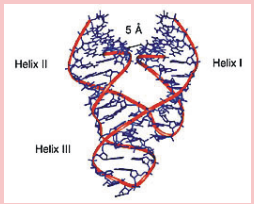
Sequenz

Molekulare Struktur

Biologische Funktion

RNA

RNA *as catalyst*



Ribozyme

RNA *as transmitter of genetic information*

DNA

transcription

...AGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUC...

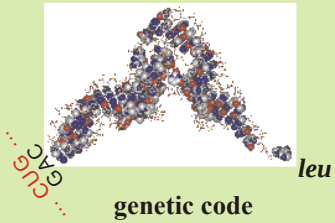
messenger-**RNA**

translation

protein

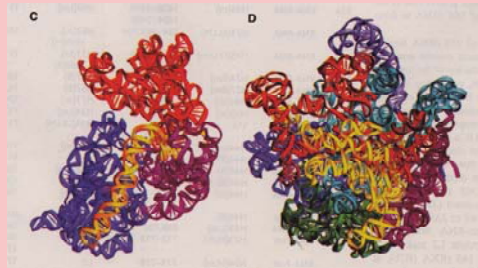
RNA as **working copy** of genetic information

RNA *as adapter molecule*



genetic code

RNA *is the catalytic subunit in supramolecular complexes*



The RNA *world as a precursor of the current* DNA + protein *biology*

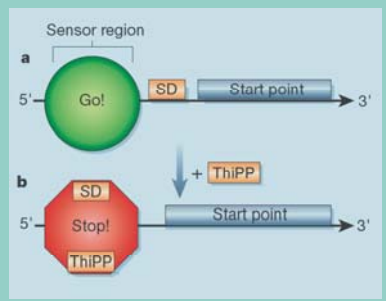
RNA *as carrier of genetic information*

RNA viruses and retroviruses

RNA evolution *in vitro*

Evolutionary biotechnology
RNA aptamers, artificial ribozymes, allosteric ribozymes

Allosteric control of transcribed RNA



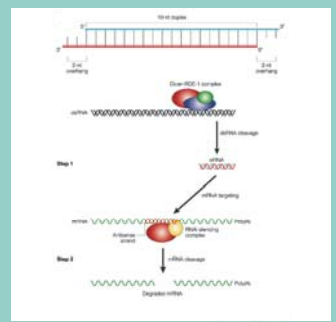
Riboswitches controlling transcription and translation through **metabolites**

RNA *is modified by epigenetic control*

RNA editing

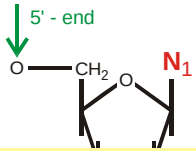
Alternative splicing of messenger **RNA**

RNA *as regulator of gene expression*

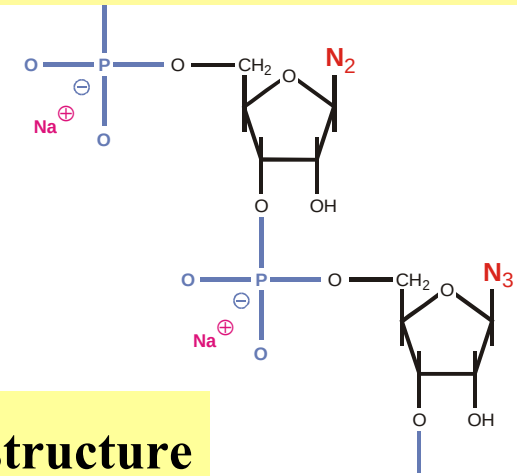


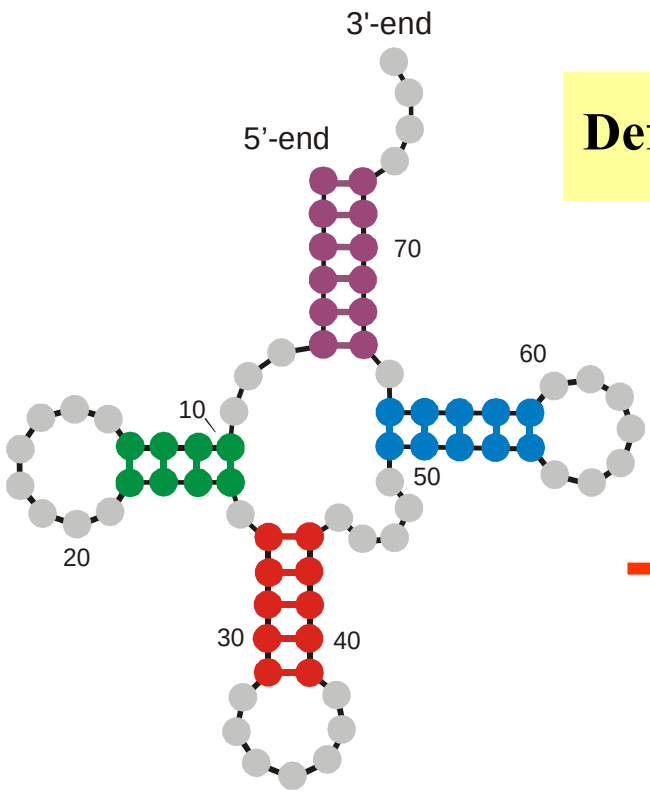
Gene silencing by small interfering RNAs

Functions of RNA molecules

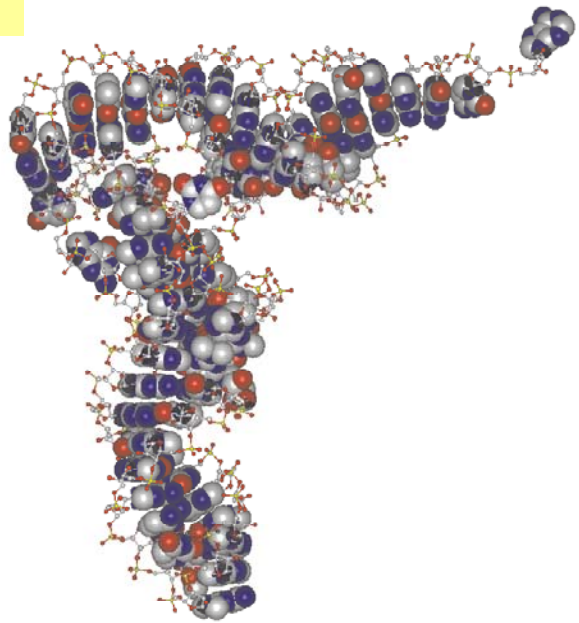


5'-end GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA 3'-end





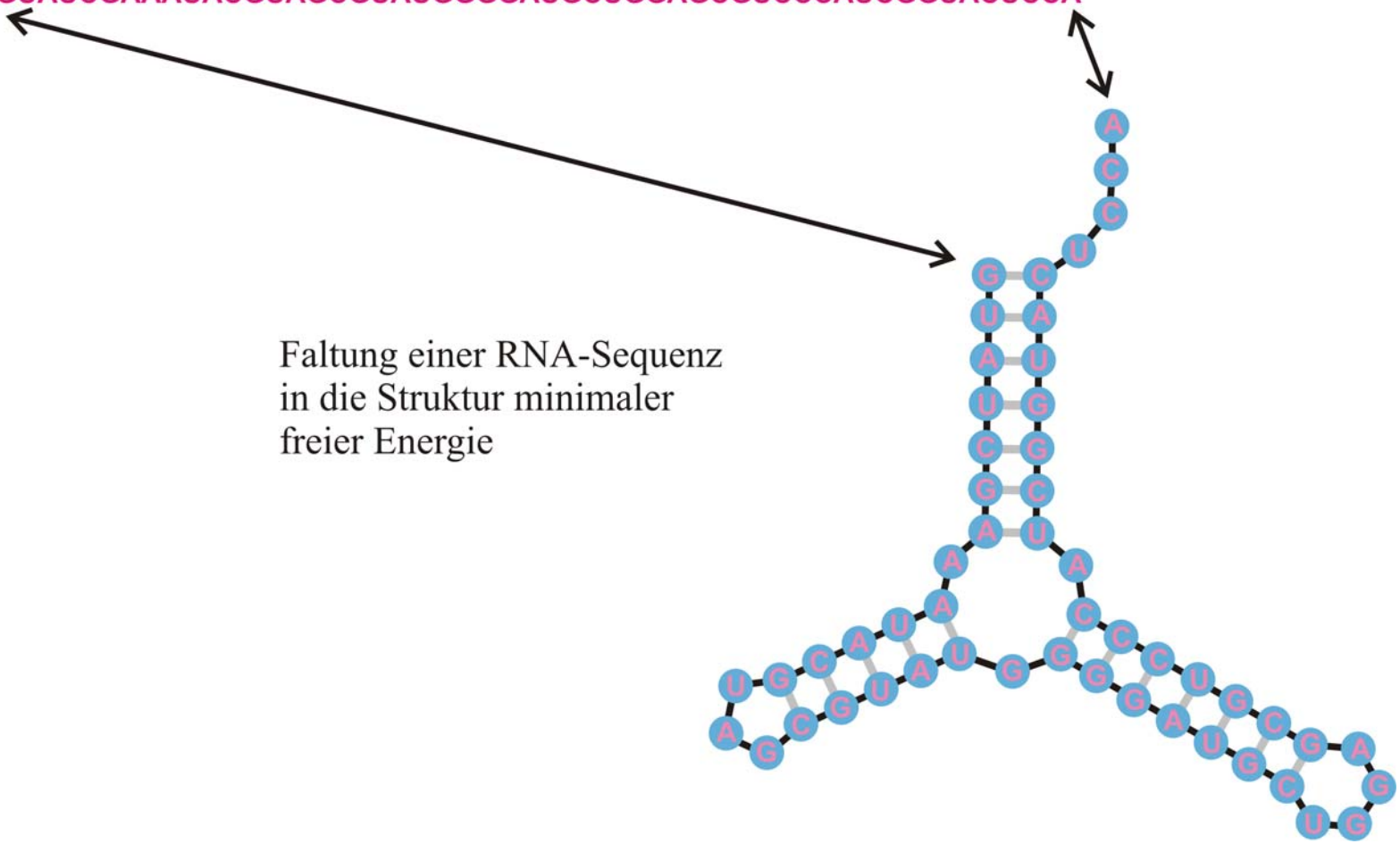
Definition of RNA structure



1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit
2. Ribonukleinsäuremoleküle als Phänotypen
3. **Evolution im Reagenzglas und im Computer**
4. Genetische und metabolische Netzwerke
5. Ein minimaler Metabolismus als Phänotyp
6. Evolution und Komplexität

5’-Ende 3’-Ende

GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGAGCGUCCCAUCGGUACUCCA



Faltung einer RNA-Sequenz
in die Struktur minimaler
freier Energie

The ***Vienna RNA-Package***:

A library of routines for folding, ***inverse folding***, sequence and structure alignment, ***kinetic folding***, ***cofolding***, …

Monatshefte für Chemie 125, 167–188 (1994)

Monatshefte für Chemie
Chemical Monthly

Printed in Austria

Fast Folding and Comparison of RNA Secondary Structures

I. L. Hofacker[1,*], W. Fontana[3], P. F. Stadler[1,3], L. S. Bonhoeffer[4], M. Tacker[1] and P. Schuster[1,2,3]

[1] Institut für Theoretische Chemie, Universität Wien, A-1090 Wien, Austria
[2] Institut für Molekulare Biotechnologie, D-07745 Jena, Federal Republic of Germany
[3] Santa Fe Institute, Santa Fe, NM 87501, U.S.A.
[4] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

Summary. Computer codes for computation and comparison of RNA secondary structures, the Vienna RNA package, are presented, that are based on dynamic programming algorithms and aim at predictions of structures with minimum free energies as well as at computations of the equilibrium partition functions and base pairing probabilities.

An efficient heuristic for the inverse folding problem of RNA is introduced. In addition we present compact and efficient programs for the comparison of RNA secondary structures based on tree editing and alignment.

All computer codes are written in ANSI C. They include implementations of modified algorithms on parallel computers with distributed memory. Performance analysis carried out on an Intel Hypercube shows that parallel computing becomes gradually more and more efficient the longer the sequences are.

Keywords. Inverse folding; parallel computing; public domain software; RNA folding; RNA secondary structures; tree editing.

Schnelle Faltung und Vergleich von Sekundärstrukturen von RNA

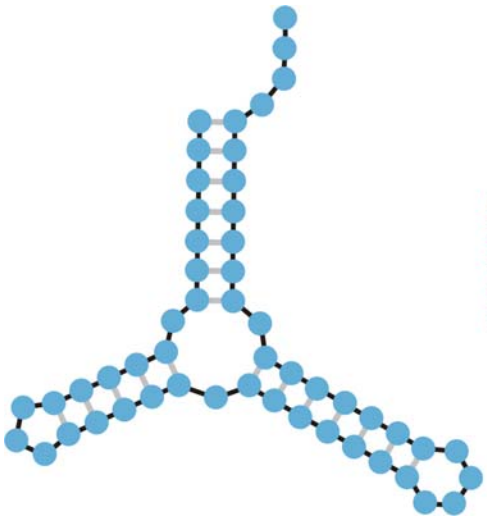
Zusammenfassung. Die im Vienna RNA package enthaltenen Computer Programme für die Berechnung und den Vergleich von RNA Sekundärstrukturen werden präsentiert. Ihren Kern bilden Algorithmen zur Vorhersage von Strukturen minimaler Energie sowie zur Berechnung von Zustandssumme und Basenpaarungswahrscheinlichkeiten mittels dynamischer Programmierung.

Ein effizienter heuristischer Algorithmus für das inverse Faltungsproblem wird vorgestellt. Darüberhinaus präsentieren wir kompakte und effiziente Programme zum Vergleich von RNA Sekundärstrukturen durch Baum-Editierung und Alignierung.

Alle Programme sind in ANSI C geschrieben, darunter auch eine Implementation des Faltungsalgorithmus für Parallelrechner mit verteiltem Speicher. Wie Tests auf einem Intel Hypercube zeigen, wird das Parallelrechnen umso effizienter je länger die Sequenzen sind.

1. Introduction

Recent interest in RNA structures and functions was caused by their catalytic capacities [1, 2] as well as by the success of selection methods in producing RNA



Minimum free energy criterion

1st
2nd
3rd trial
4th
5th

Inverse folding

UUUAGCCAGCGCGAGUCGUGCGGACGGGGUUAUCUCUGUCGGGCUAGGGCGC

GUGAGCGCGGGGCACAGUUUCUCAAGGAUGUAAGUUUUUGCCGUUUAUCUGG

UUAGCGAGAGAGGAGGCUUCUAGACCCAGCUCUCUGGGUCGUUGCUGAUGCG

CAUUGGUGCUAAUGAUAUUAGGGCUGUAUUCCUGUAUAGCGAUCAGUGUCCG

GUAGGCCCUCUUGACAUAAGAUUUUUCCAAUGGUGGGAGAUGGCCAUUGCAG

Evolution *in silico*

W. Fontana, P. Schuster,
Science **280** (1998), 1451-1455

random individuals. The primer pair used for genomic DNA amplification is 5′-TCTCCCTGGATTCT-CATTTA-3′ (forward) and 5′-TCTTTGTCTTCTGT-TCCACC-3′ (reverse). Reactions were performed in 25 μl using 1 unit of Taq DNA polymerase with each primer at 0.4 μM; 200 μM each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM tris-HCl (pH 8.3), 50 mM KCl_2,1.5 mM $MgCl_2$] in a cycle condition of 94°C for 1 min and then 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by 72°C for 6 min. PCR products were purified (Qiagen), digested with Xmn I, and separated in a 2% agarose gel.

32. A nonsense mutation may affect mRNA stability and result in degradation of the transcript [L. Maquat, *Am. J. Hum. Genet.* **59**, 279 (1996)].
33. Data not shown; a dot blot with poly (A)$^+$ RNA from 50 human tissues (The Human RNA Master Blot, 7770-1, Clontech Laboratories) was hybridized with a probe from exons 29 to 47 of *MYO15* using the same condition as Northern blot analysis (*13*).
34. Smith–Magenis syndrome (SMS) is due to deletions of 17p11.2 of various sizes, the smallest of which includes *MYO15* and perhaps 20 other genes [(*6*); K-S Chen, L. Potocki, J. R. Lupski, *MRDD Res. Rev.* **2**, 122 (1996)]. *MYO15* expression is easily detected in the pituitary gland (data not shown). Haploinsufficiency for *MYO15* may explain a portion of the SMS phenotype such as short stature. Moreover, a few SMS patients have sensorineural hearing loss, possibly because of a point mutation in *MYO15* in trans to the SMS 17p11.2 deletion.
35. R. A. Fridell, data not shown.
36. K. B. Avraham *et al.*, *Nature Genet.* **11**, 369 (1995); X-Z. Liu *et al.*, *ibid.* **17**, 268 (1997); F. Gibson *et al.*, *Nature* **374**, 62 (1995); D. Weil *et al.*, *ibid.*, p. 60.
37. RNA was extracted from cochlea (membranous labyrinths) obtained from human fetuses at 18 to 22 weeks of development in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Only samples without evidence of degradation were pooled for poly (A)$^+$ selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-for-PCR kit (Clontech Laboratories). A portion of the first-strand cDNA (4%) was amplified by PCR with Advantage cDNA polymerase mix (Clontech Laboratories) using human *MYO15*-specific oligonucleotide primers (forward, 5′-GCATGACCTGCCGGCTAAT-GGG-3′; reverse, 5′-CTCACGGCTTCTGCATGGT-GCTCGGCTGGC-3′). Cycling conditions were 40 s at 94°C; 40 s at 66°C (3 cycles), 60°C (5 cycles), and 55°C (29 cycles); and 45 s at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 688-bp PCR product is expected from amplification of the human *MYO15* cDNA. Amplification of human genomic DNA with this primer pair would result in a 2903-bp fragment.

38. We are grateful to the people of Bengkala, Bali, and the two families from India. We thank J. R. Lupski and K.-S. Chen for providing the human chromosome 17 cosmid library. For technical and computational assistance, we thank N. Dietrich, M. Fergusson, A. Gupta, E. Sorbello, R. Torkzadeh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Sternberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and T. Barber, S. Sullivan, E. Green, D. Drayna, and J. Battey for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (Z01 DC 00035-01 and Z01 DC 00038-01 to T.B.F. and E.R.W. and R01 DC 03402 to C.C.M.), the National Institute of Child Health and Human Development (R01 HD30428 to S.A.C.) and a National Science Foundation Graduate Research Fellowship to F.J.P. This paper is dedicated to J. B. Snow Jr. on his retirement as the Director of the NIDCD.

9 March 1998; accepted 17 April 1998

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 phenotypes 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 simulation of replicating 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 level of phenotypes (*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.

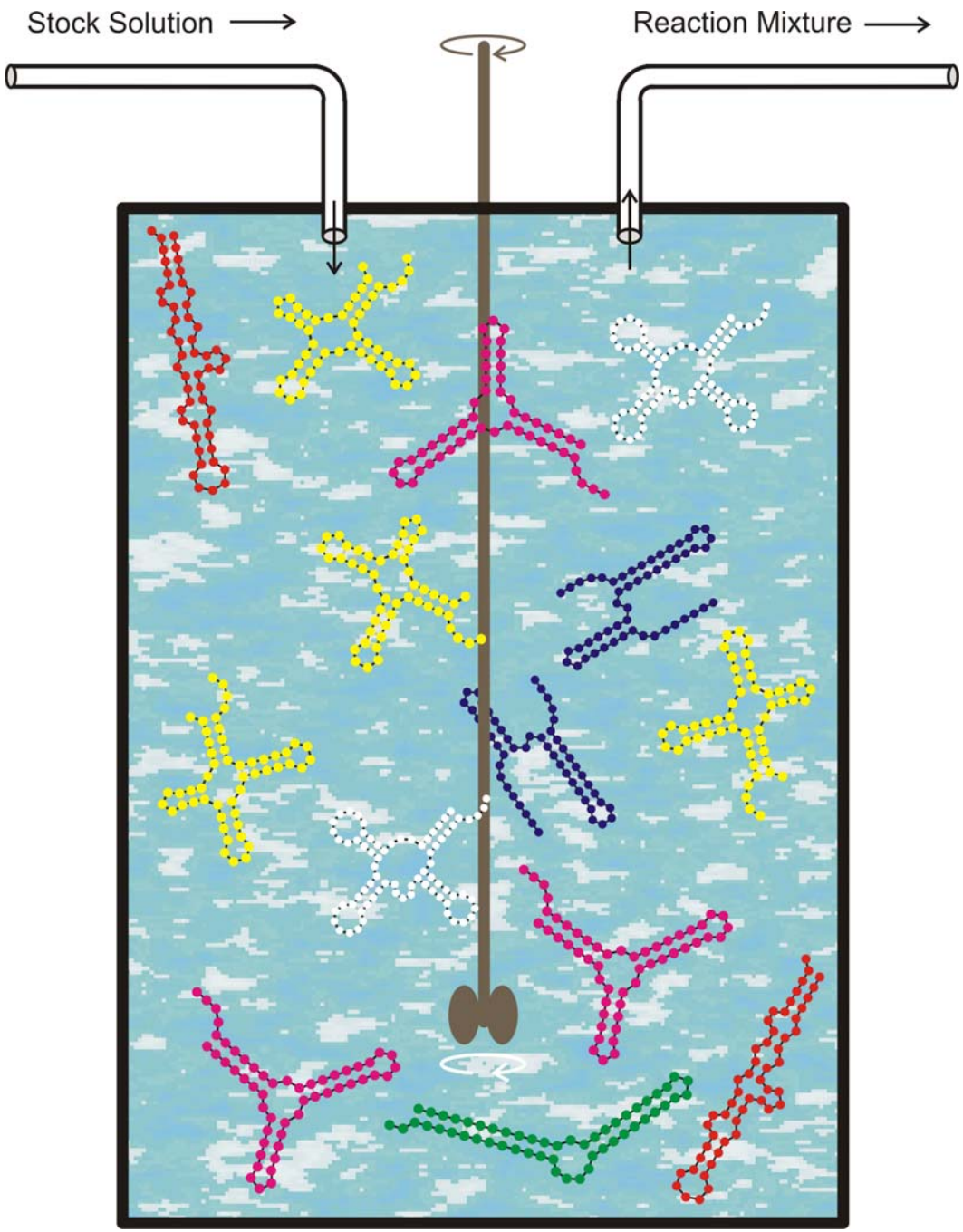
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level of analysis compared with the three-dimensional structure at atomic resolution. Yet, secondary structures are empirically well defined and obtain 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 (replicatable 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 we 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, here a canonical tRNA 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 steps, that is, flat periods of no apparent adaptive progress, interrupted by sudden approaches toward the target structure (*7*). However, the dominant shapes in the population not only change at these marked events but undergo several fitness-neutral transformations during the periods of no apparent progress. Although discontinuities in the fitness trace are evident, it is entirely unclear when and on the basis of what the series of successive phenotypes itself can be 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

Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.



Replikationsgeschwindigkeit

(Fitness):

$$f_k = \gamma / [\alpha + \Delta d_S^{(k)}]$$

$$\Delta d_S^{(k)} = d_H(S_k, S_\tau)$$

Selektionsdruck:

Die Populationsgröße,

N = # RNA Moleküle,

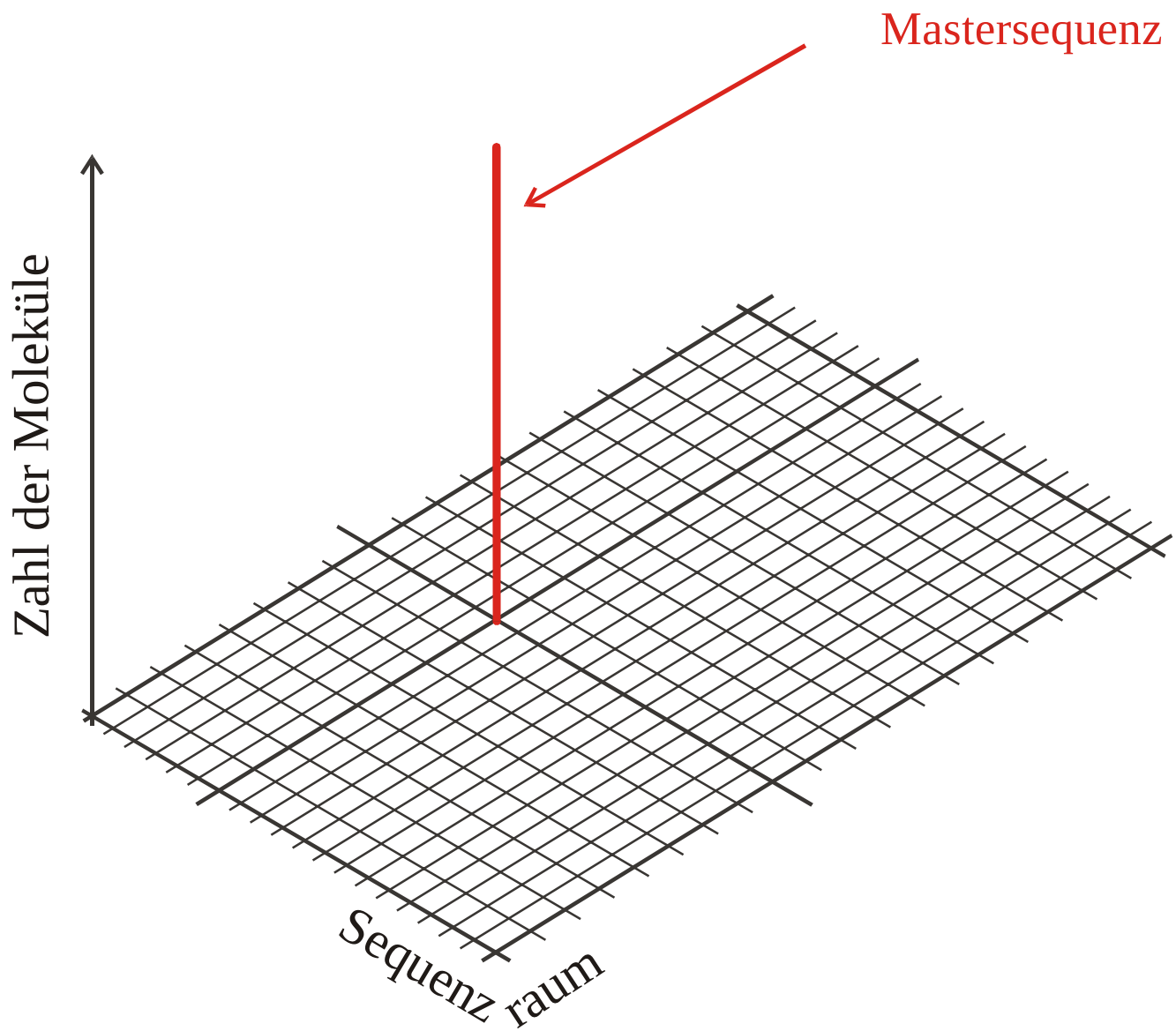
wird durch den Fluss bestimmt:

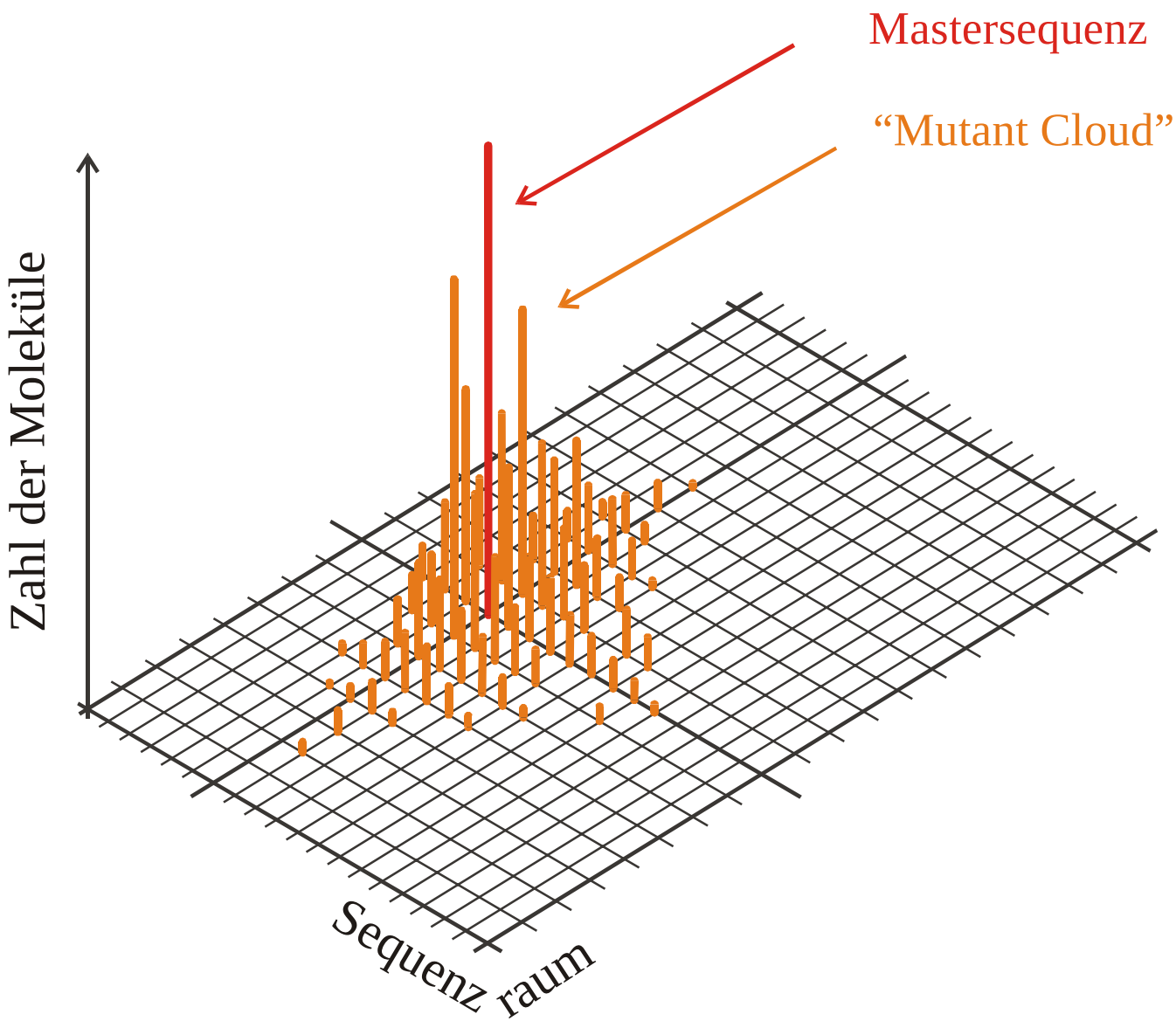
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

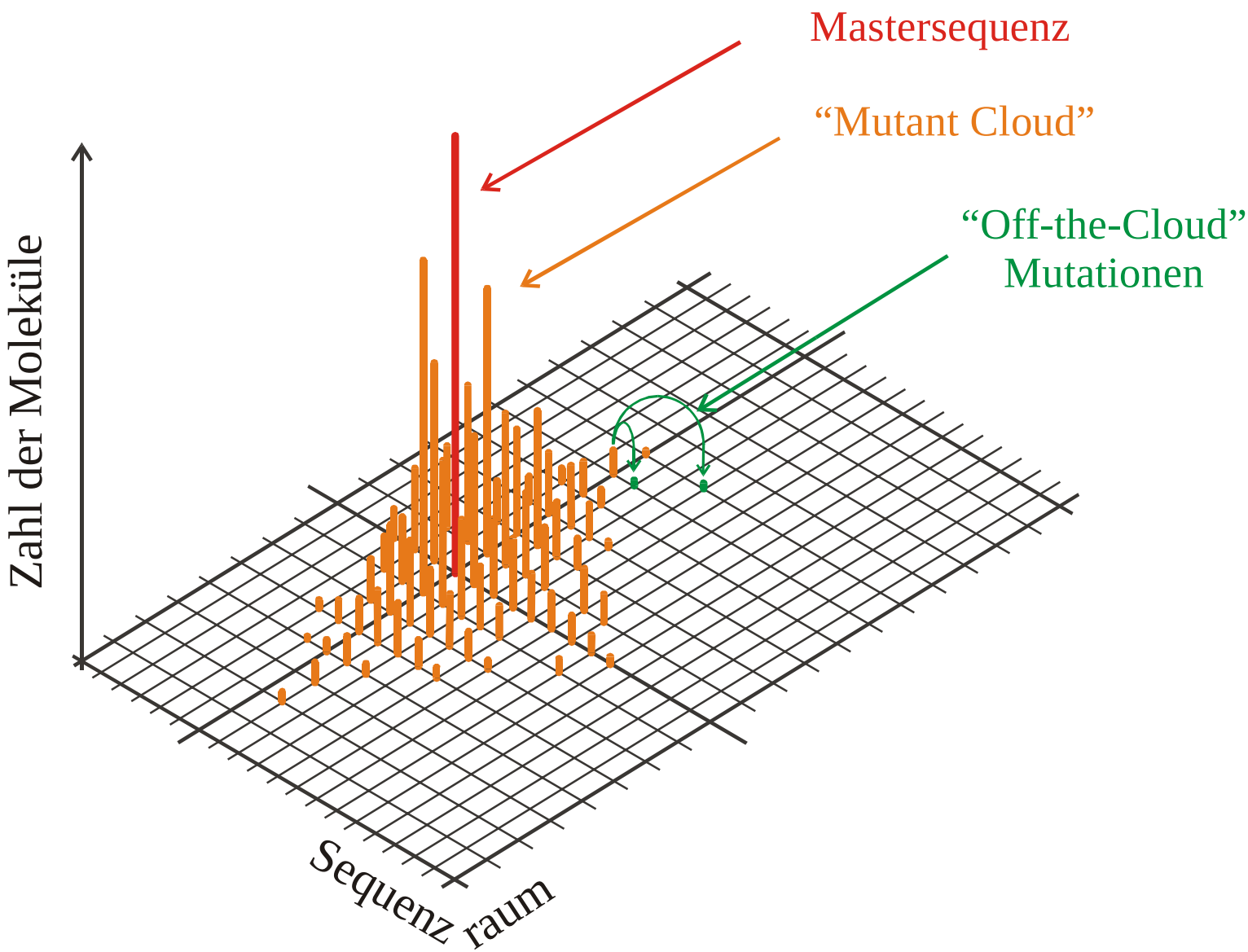
Mutationsrate:

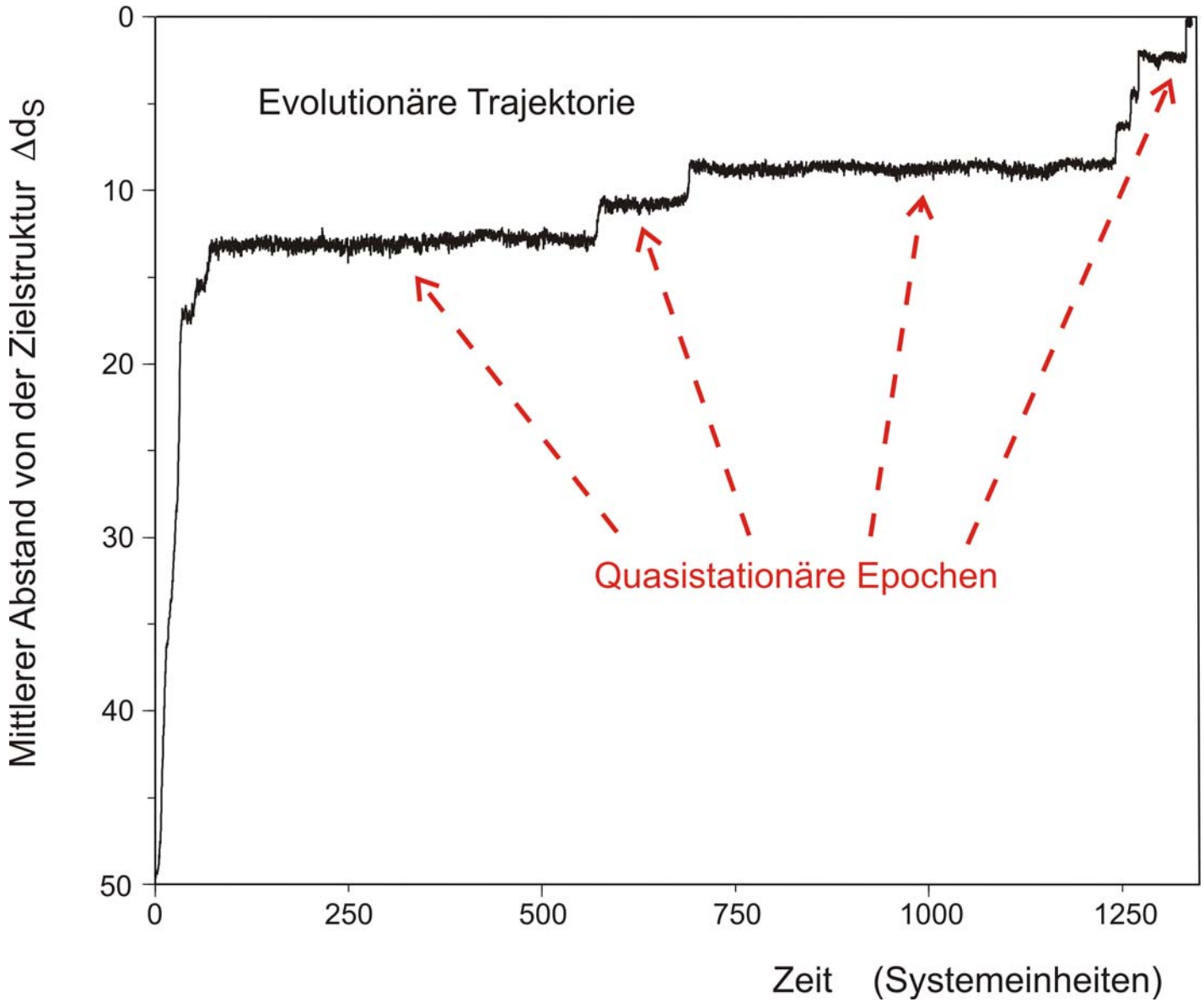
p = 0.001 / Nukleotid × Replikation

Der Flussreaktor als eine Apparatur zum Studium der Evolution von Molekülen ***in vitro*** und ***in silico***.


Mastersequenz
Zahl der Moleküle
Sequenz raum

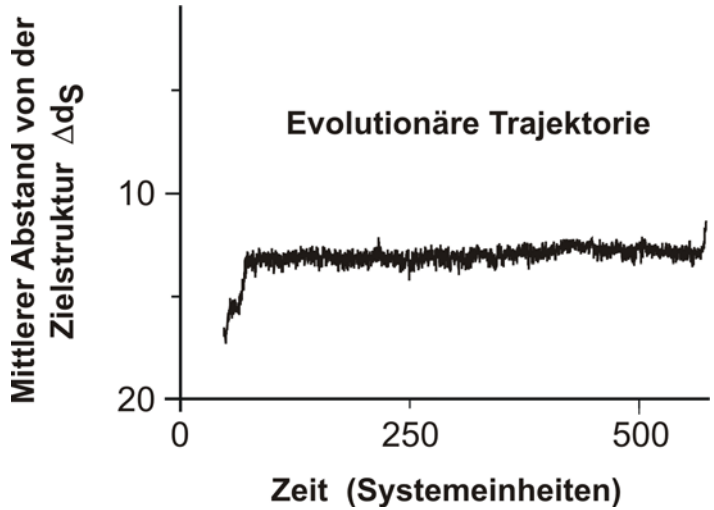

Mastersequenz
“Mutant Cloud”
Zahl der Moleküle
Sequenz raum


Mastersequenz
“Mutant Cloud”
“Off-the-Cloud”
Mutationen
Zahl der Moleküle
Sequenz raum



In silico Optimierung im Flußreaktor: Evolutionäre Trajektorie

28 neutrale Punktmutationen
während ein langen quasistationären
Epoche der Optimierung




```
entry   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8       .((((((((((((.........(((....)))........)))))).....((((.........))))))))))....
exit    GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCAUACAGAA
entry   GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUACCAUACAGAA
9       .((((((.((((.........(((....)))........)))))).....((((.........))))).))))))....
exit    UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACACCGUCCCAAG
entry   UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACAGCGUCCCAAG
10      .(((((..((((.........(((....)))........)))))).....((((.........))))).)))))....
exit    UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACAGCGUCCCAAG
```

Übergänge induzierende Punktmutationen verändern die Molekülstruktur

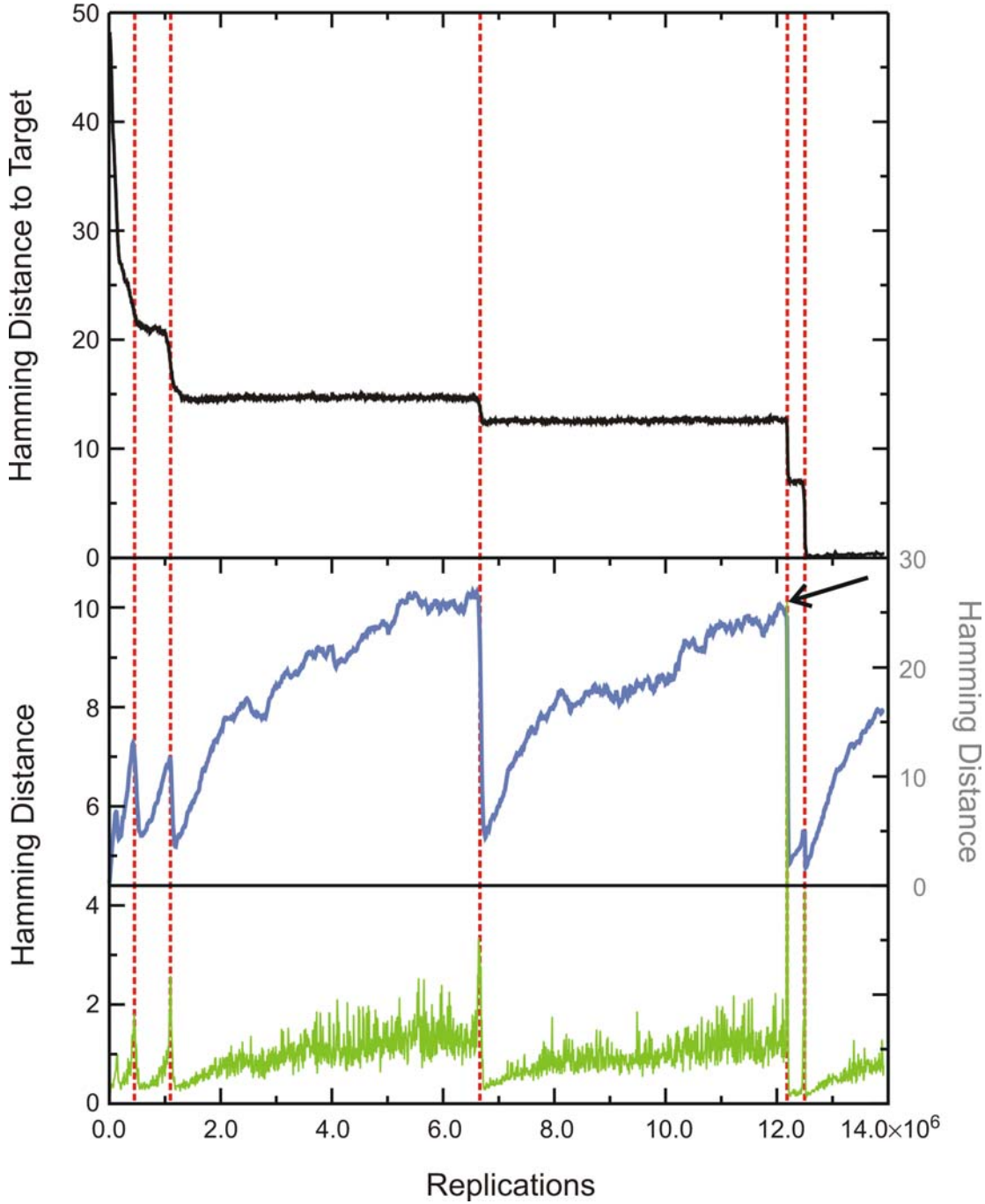
Neutrale Punktmutation lassen die Molekülstruktur unverändert

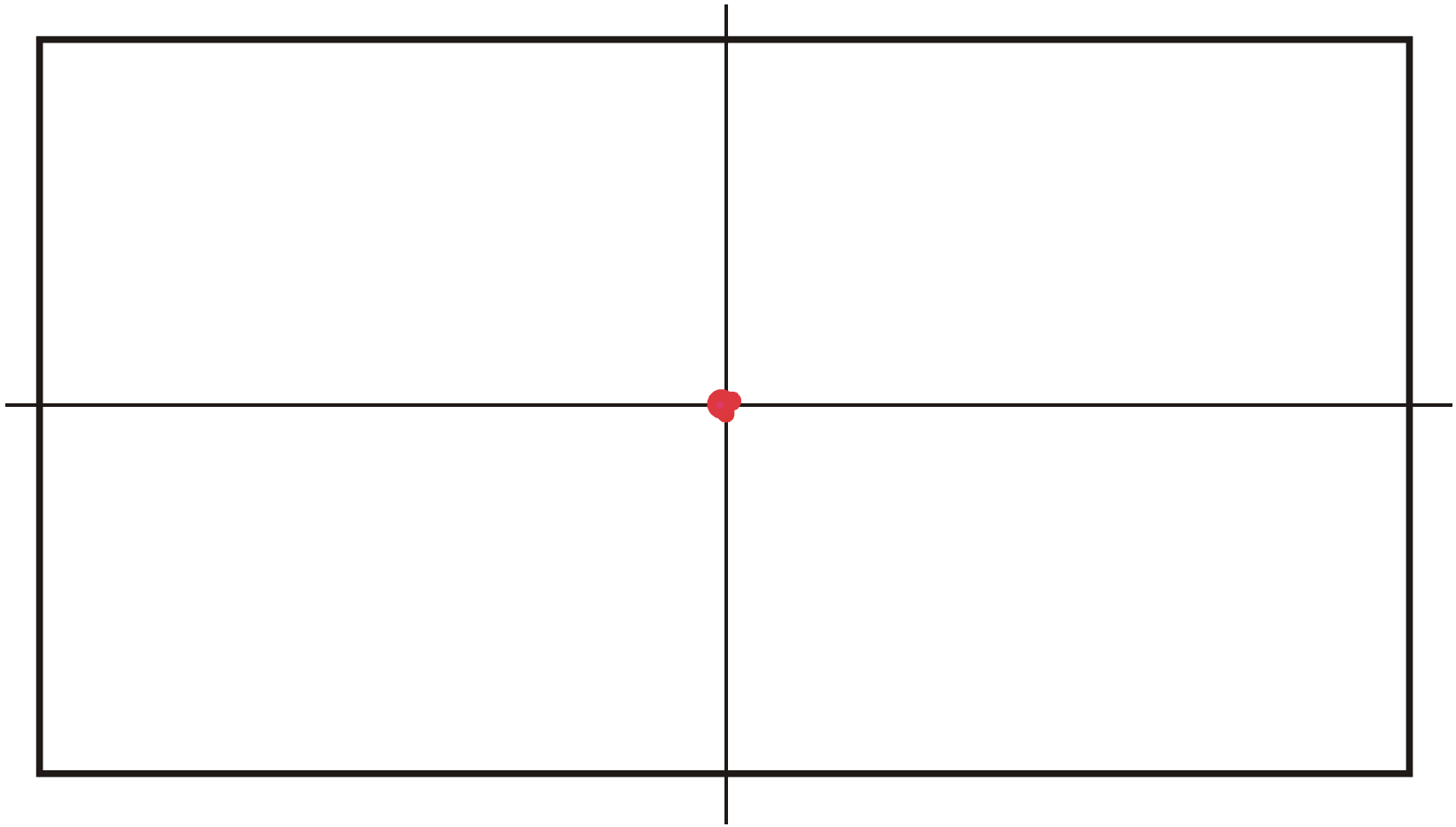
Evolution auf neutralen Netzwerken

Evolutionäre Trajectorie

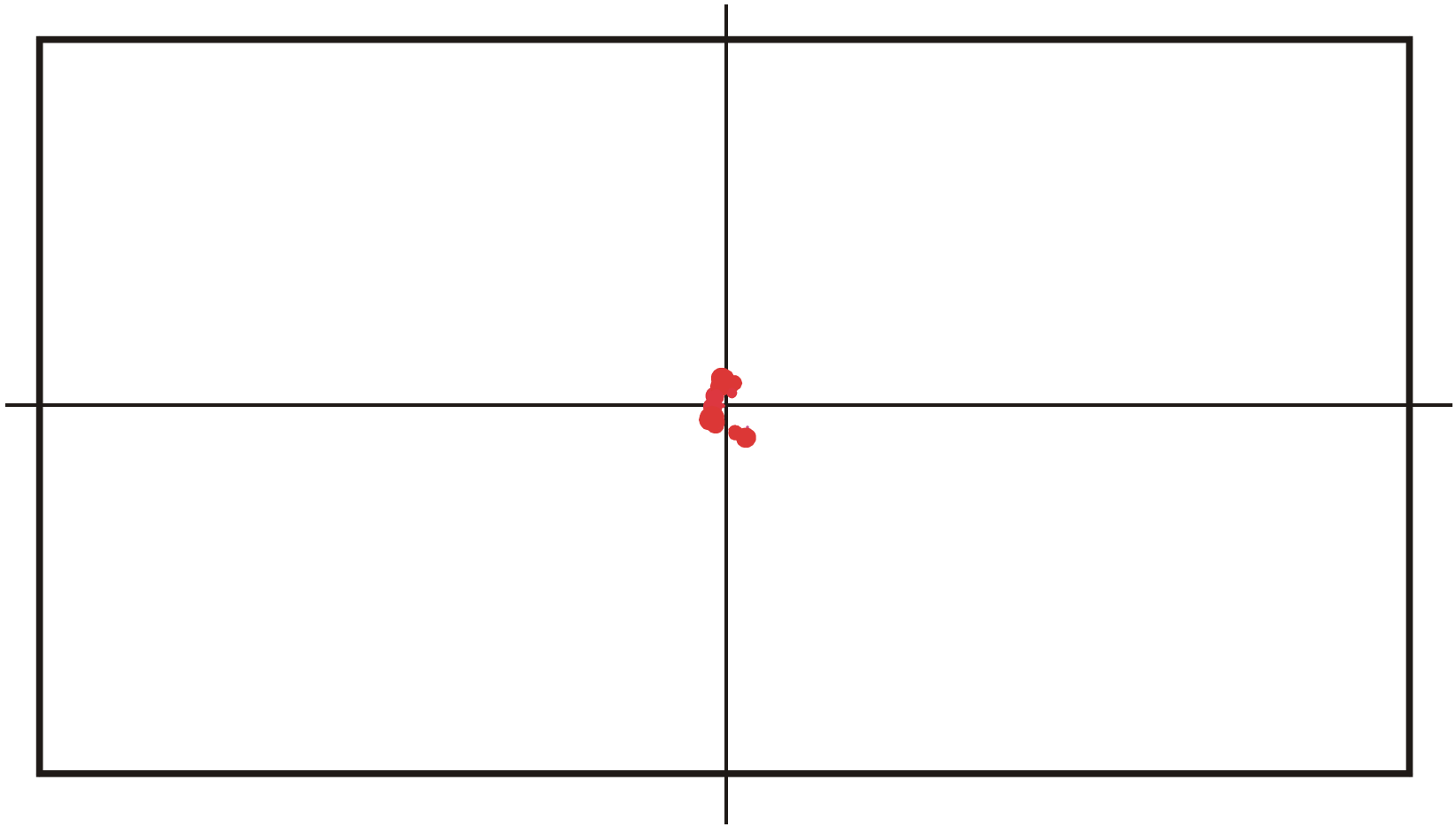
Ausbreitung der Population durch Diffusion auf einem neutralen Netzwerk

Drift des Mittelpunktes der Population im Sequenzraum

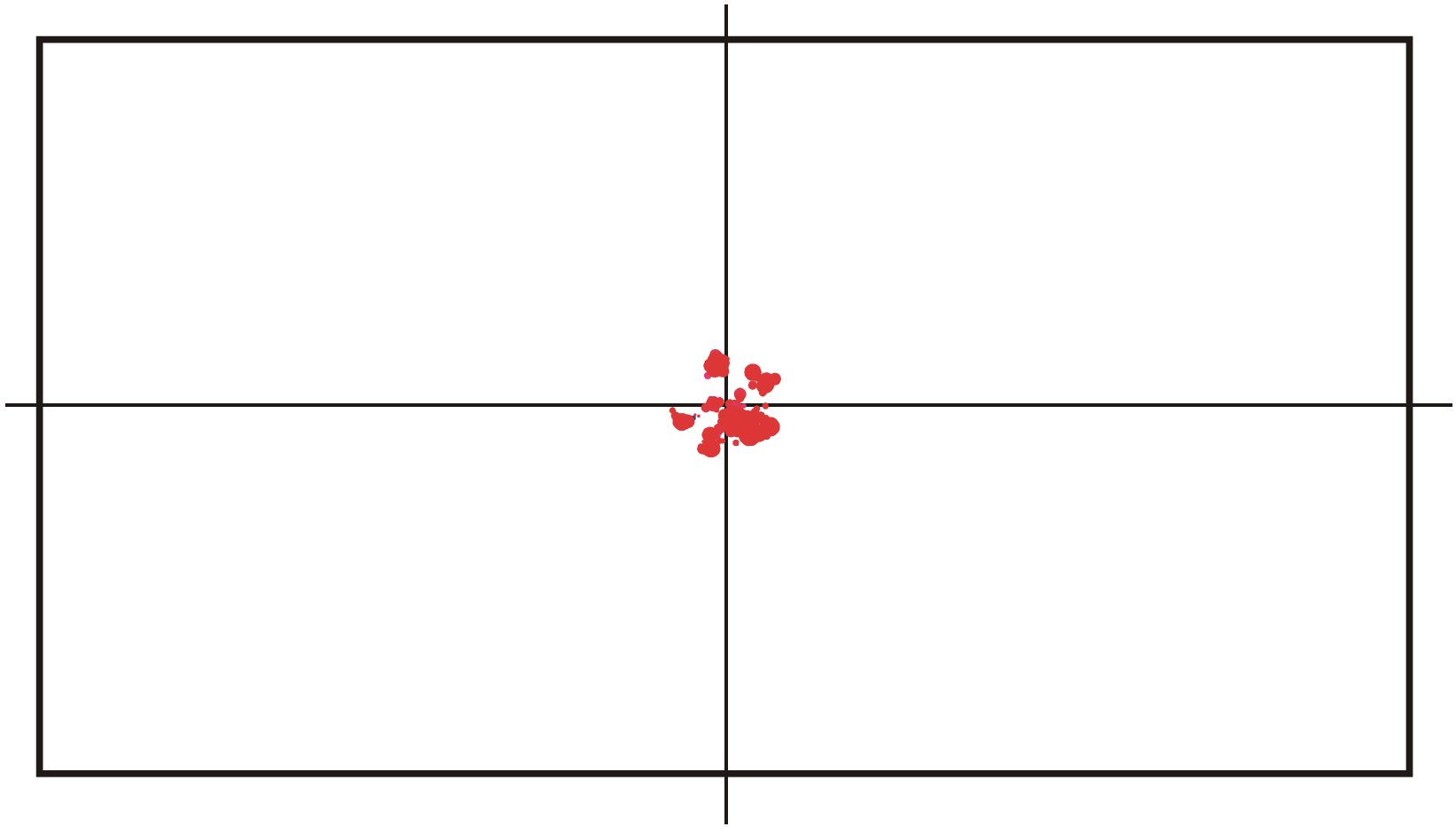




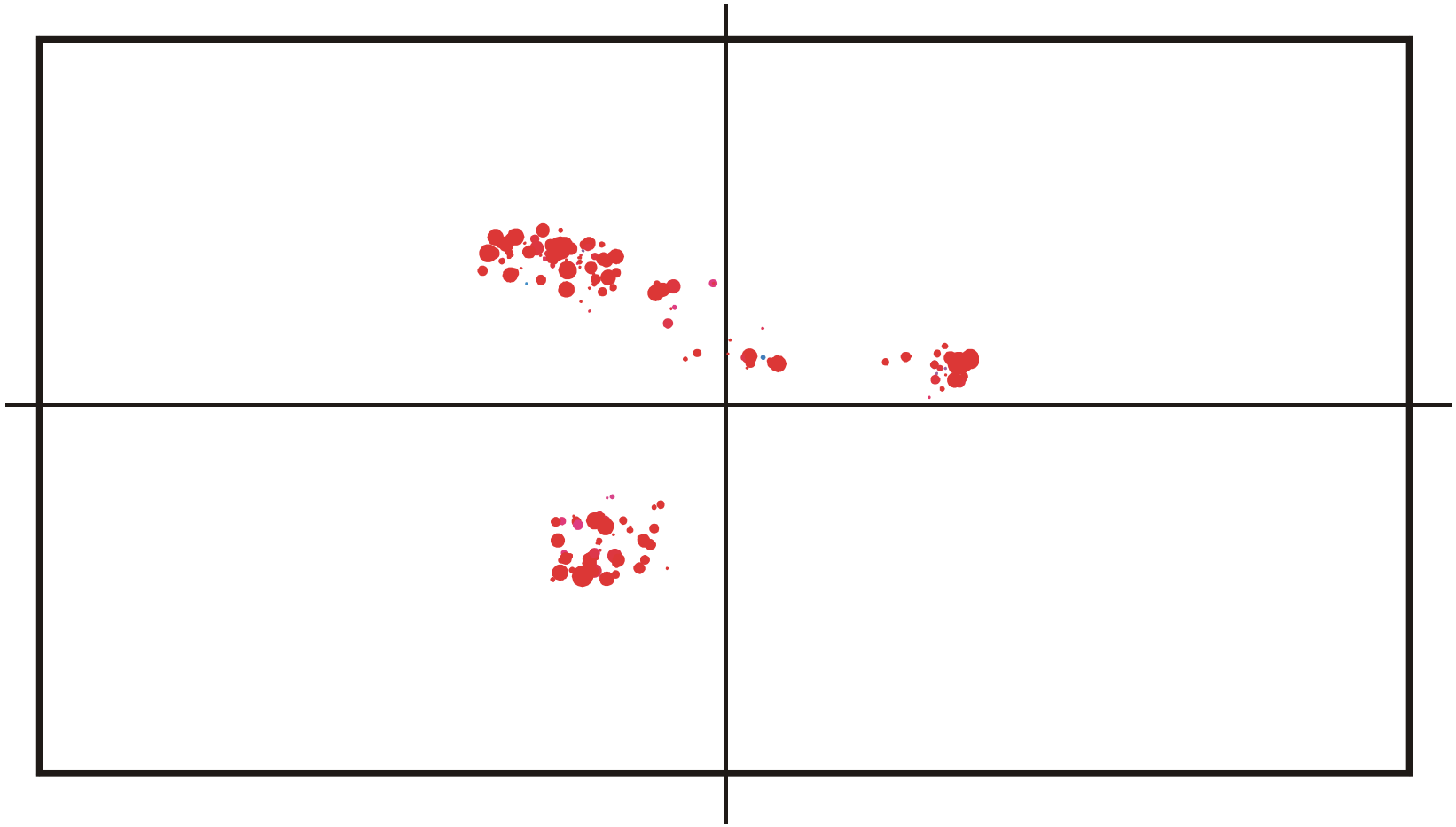
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 150



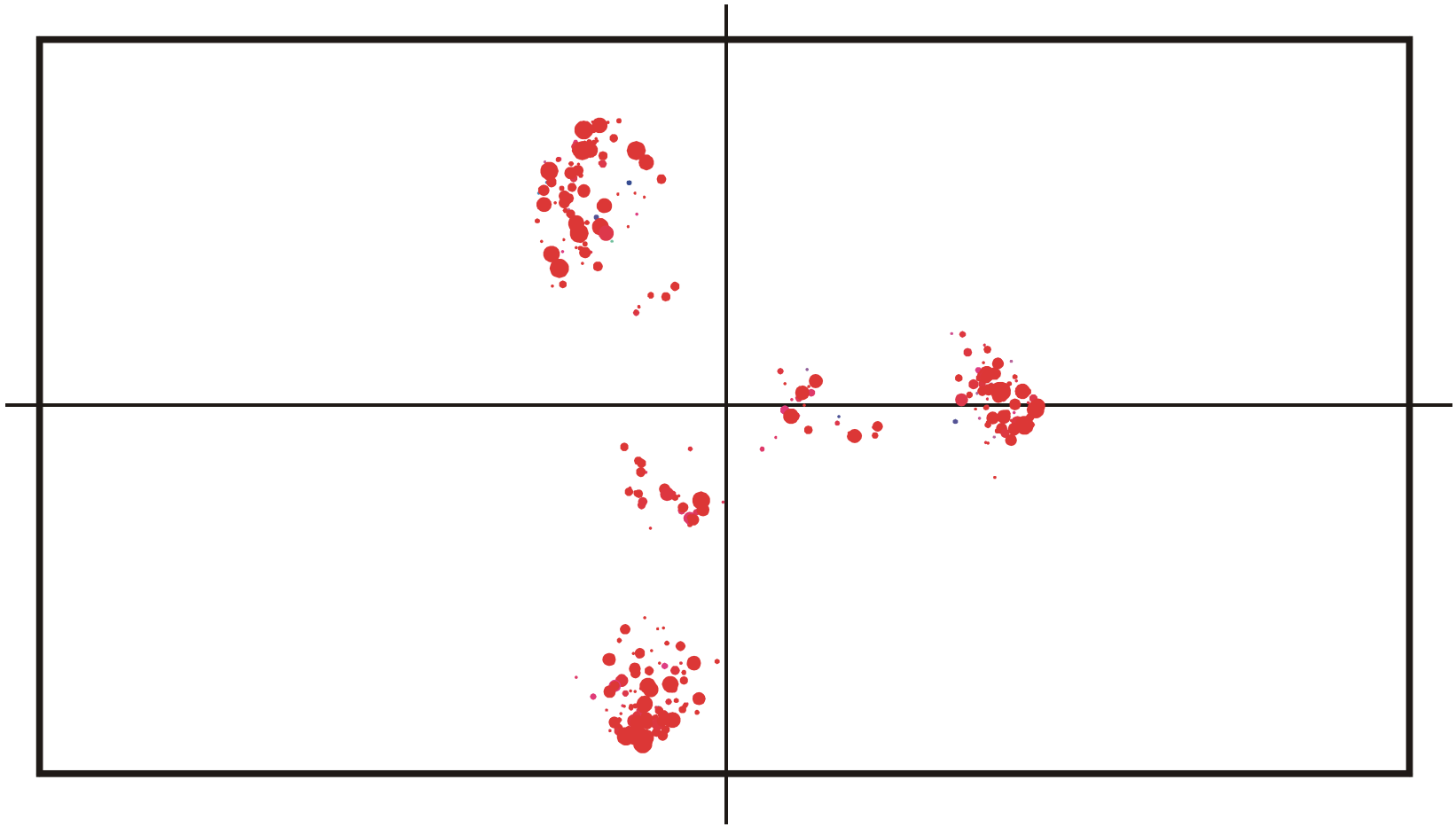
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 170



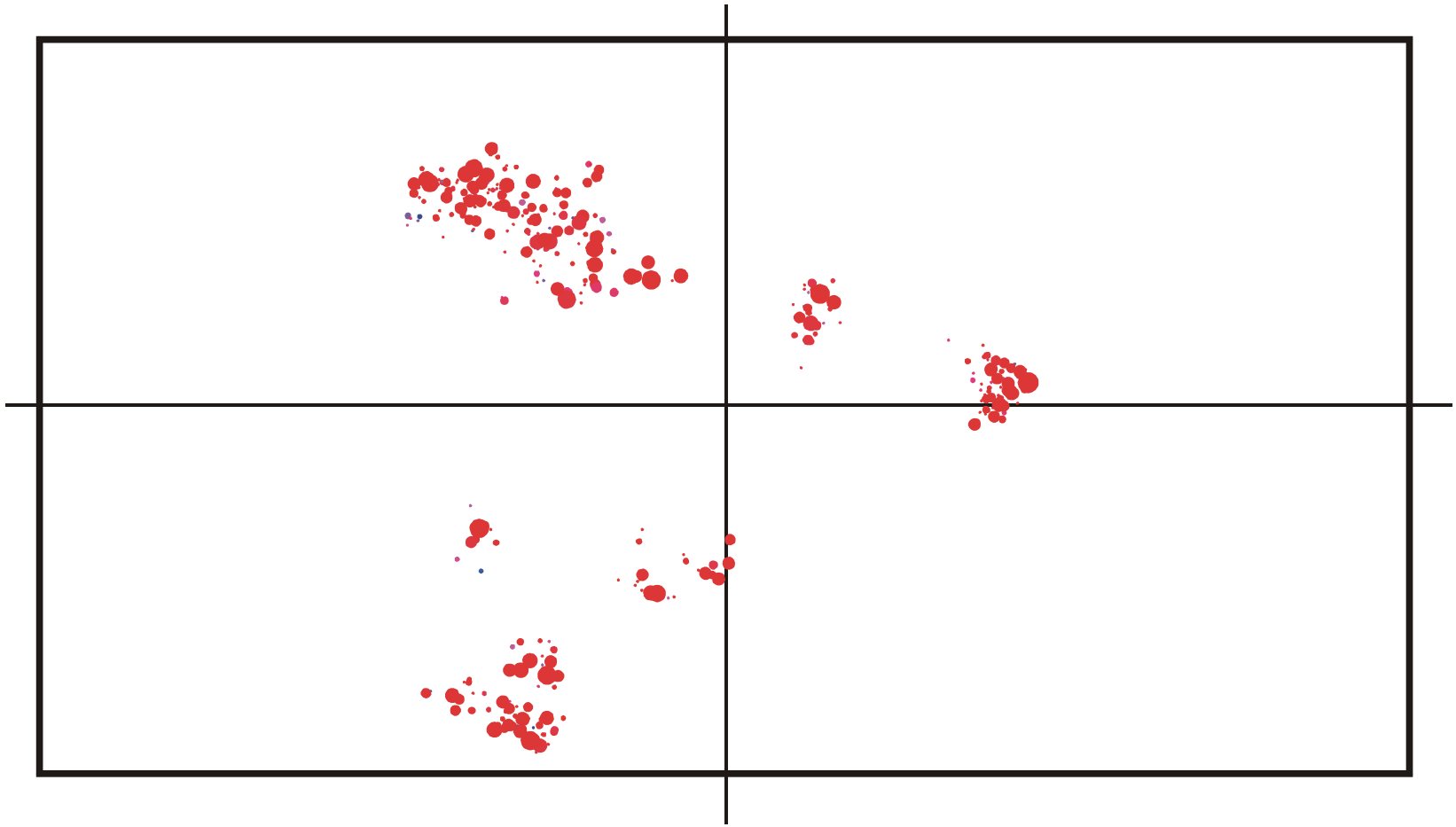
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 200



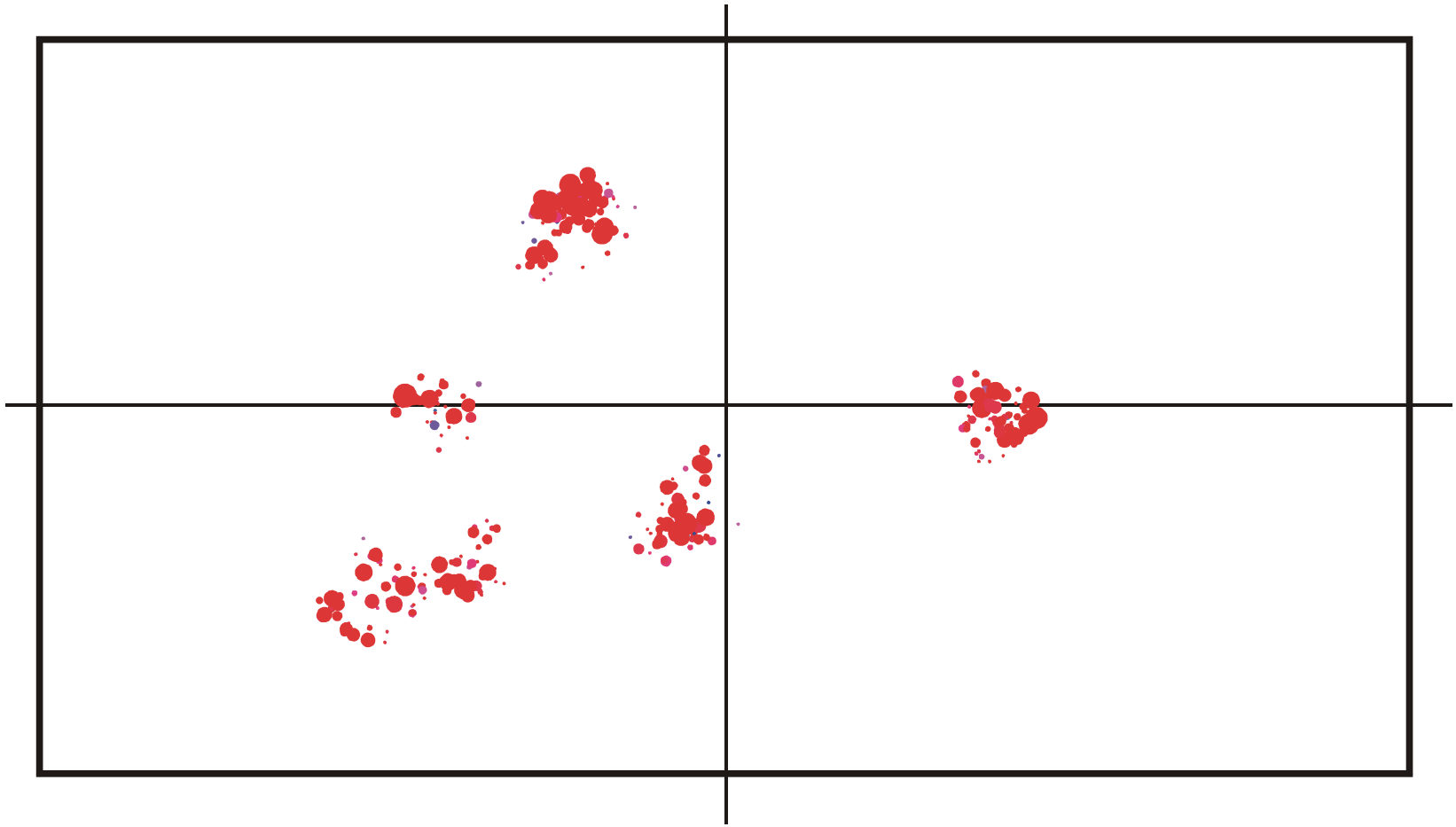
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 350



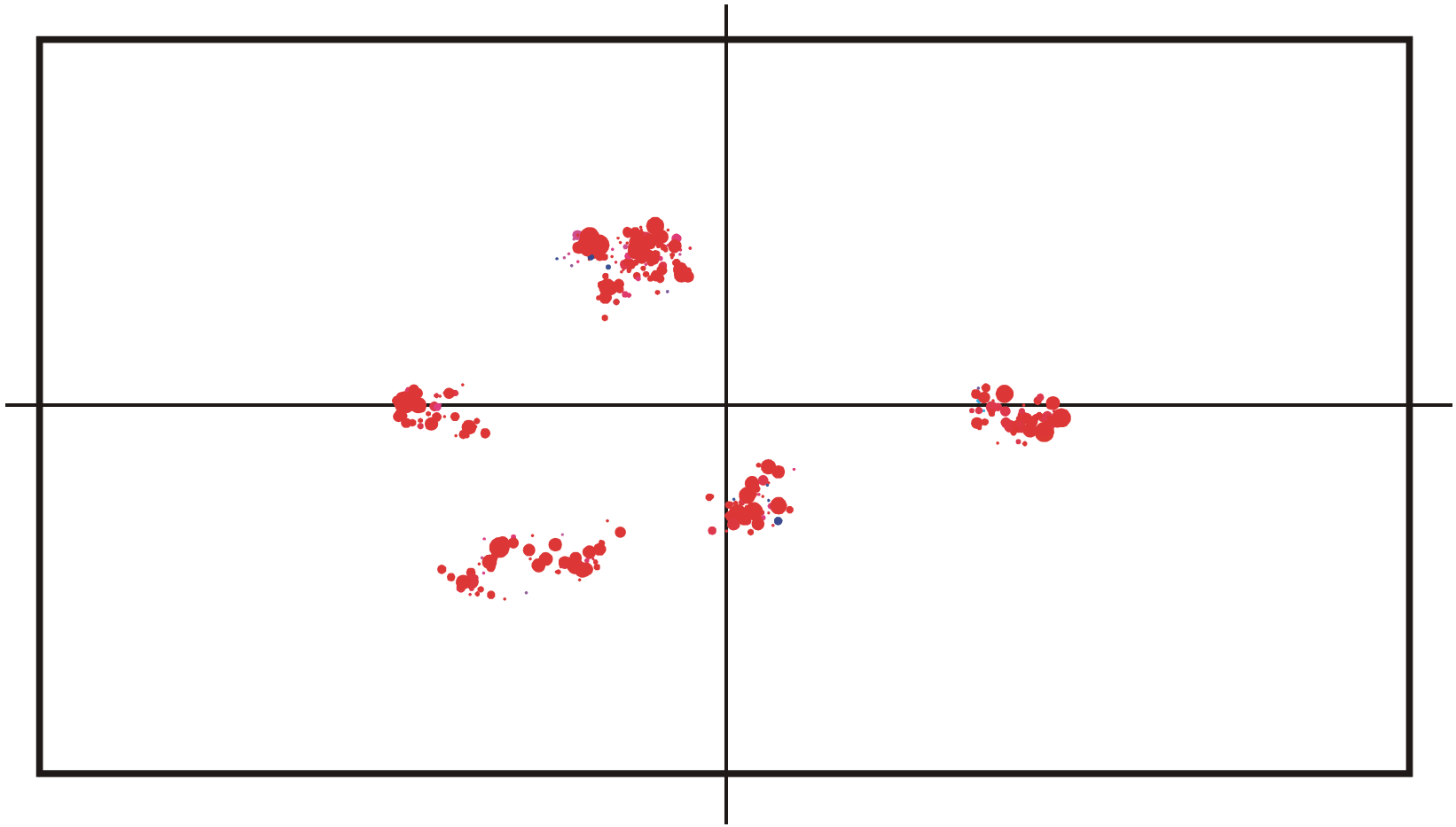
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 500



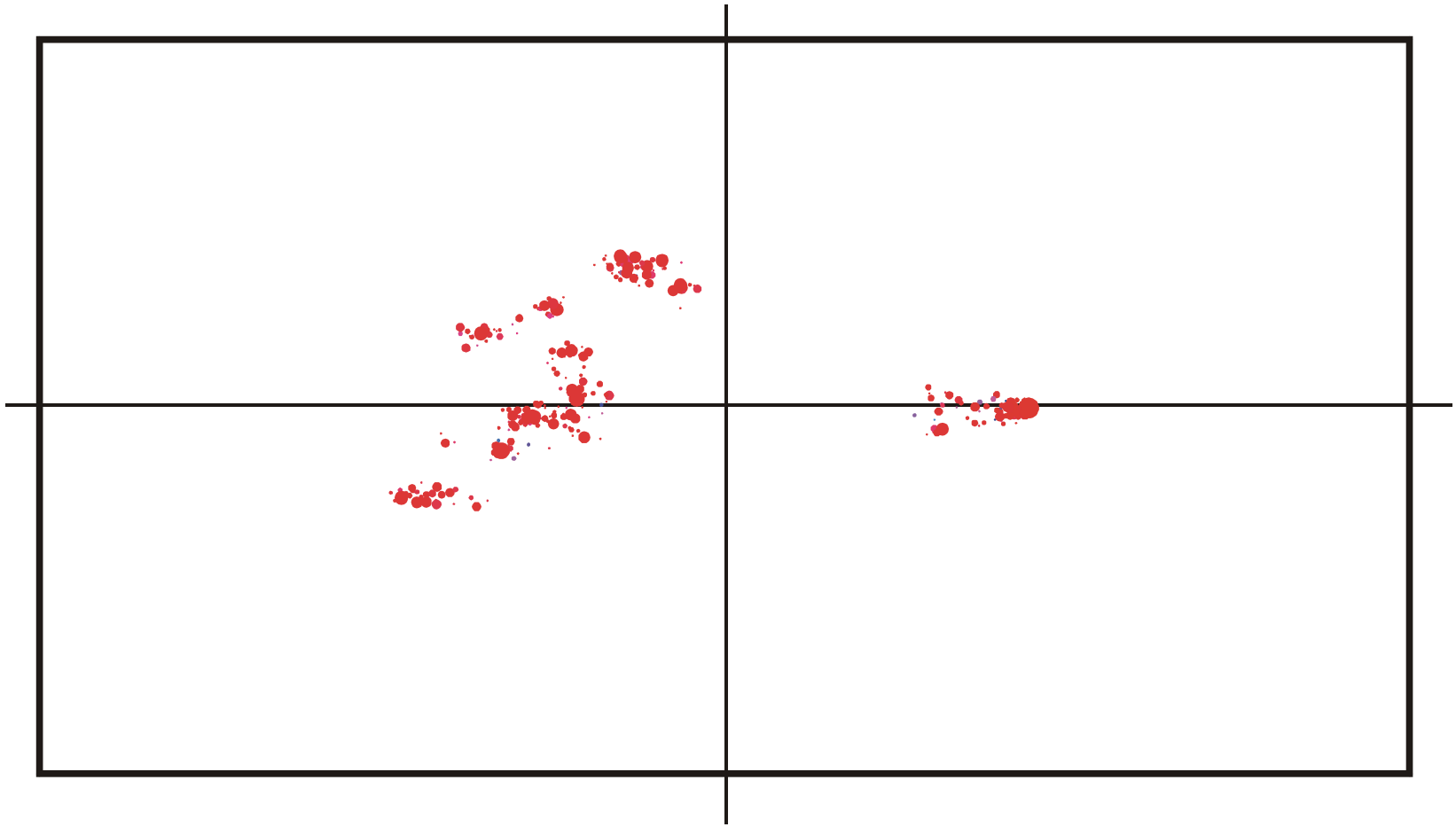
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 650



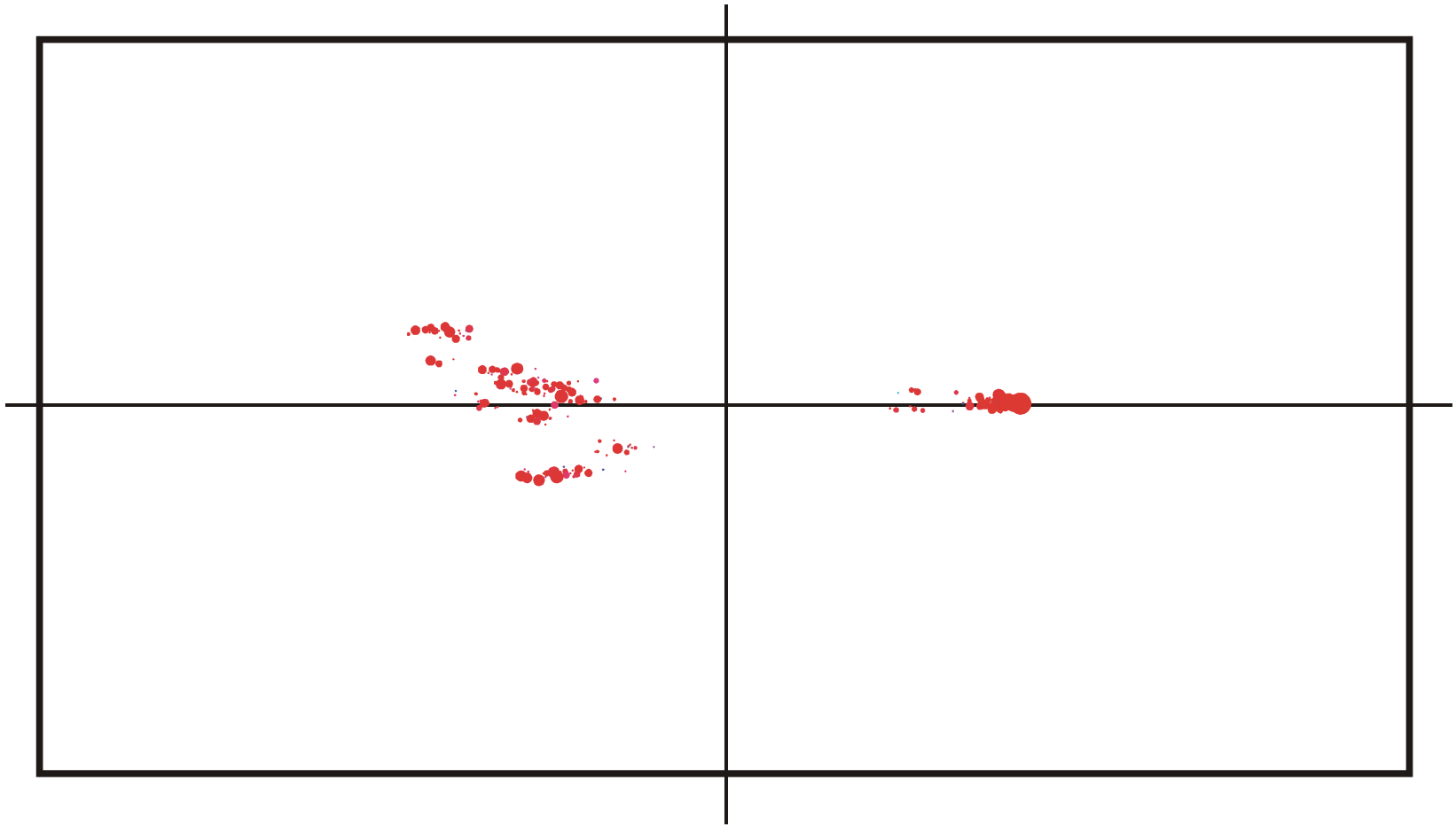
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 820



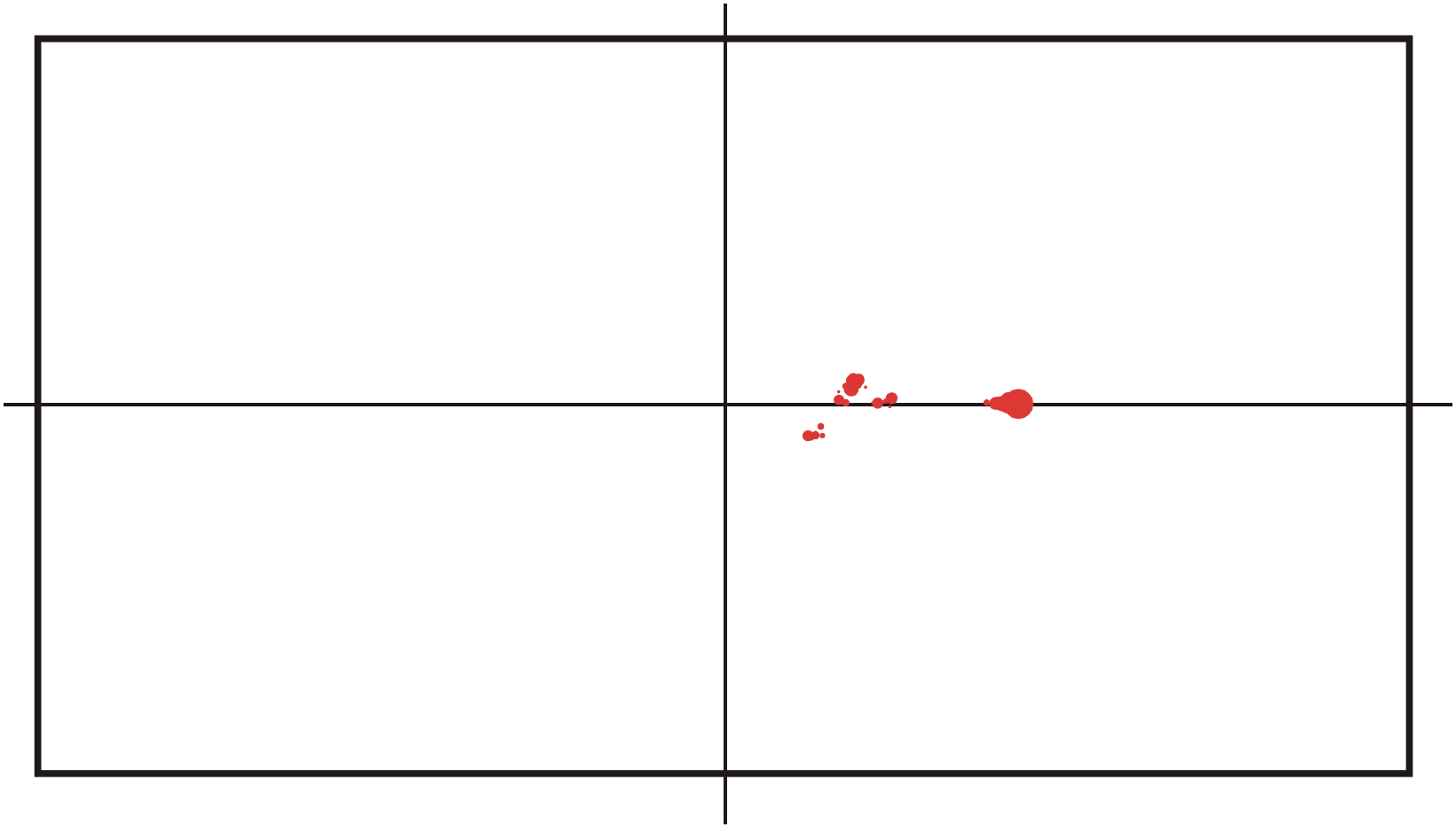
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 825



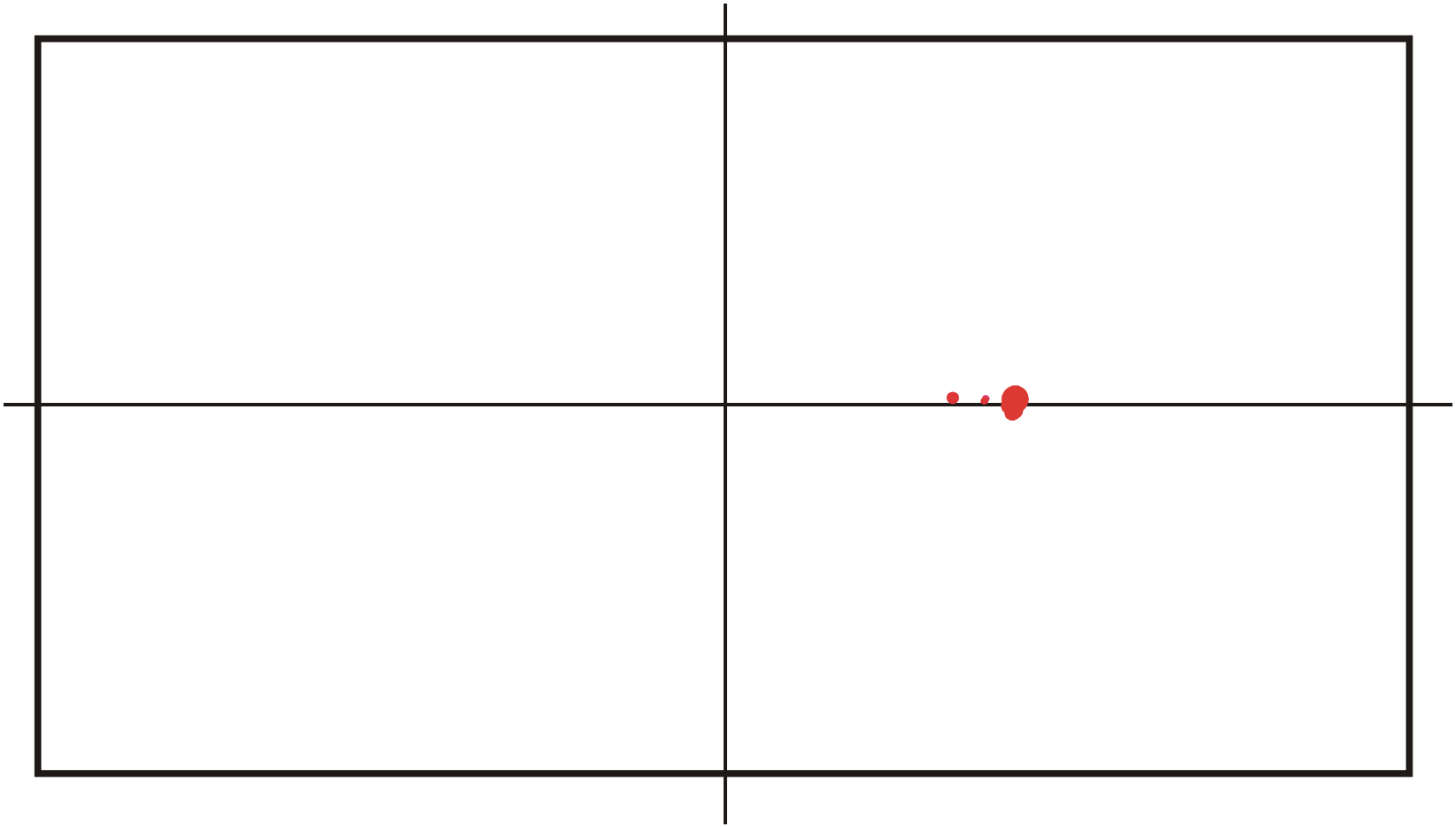
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 830



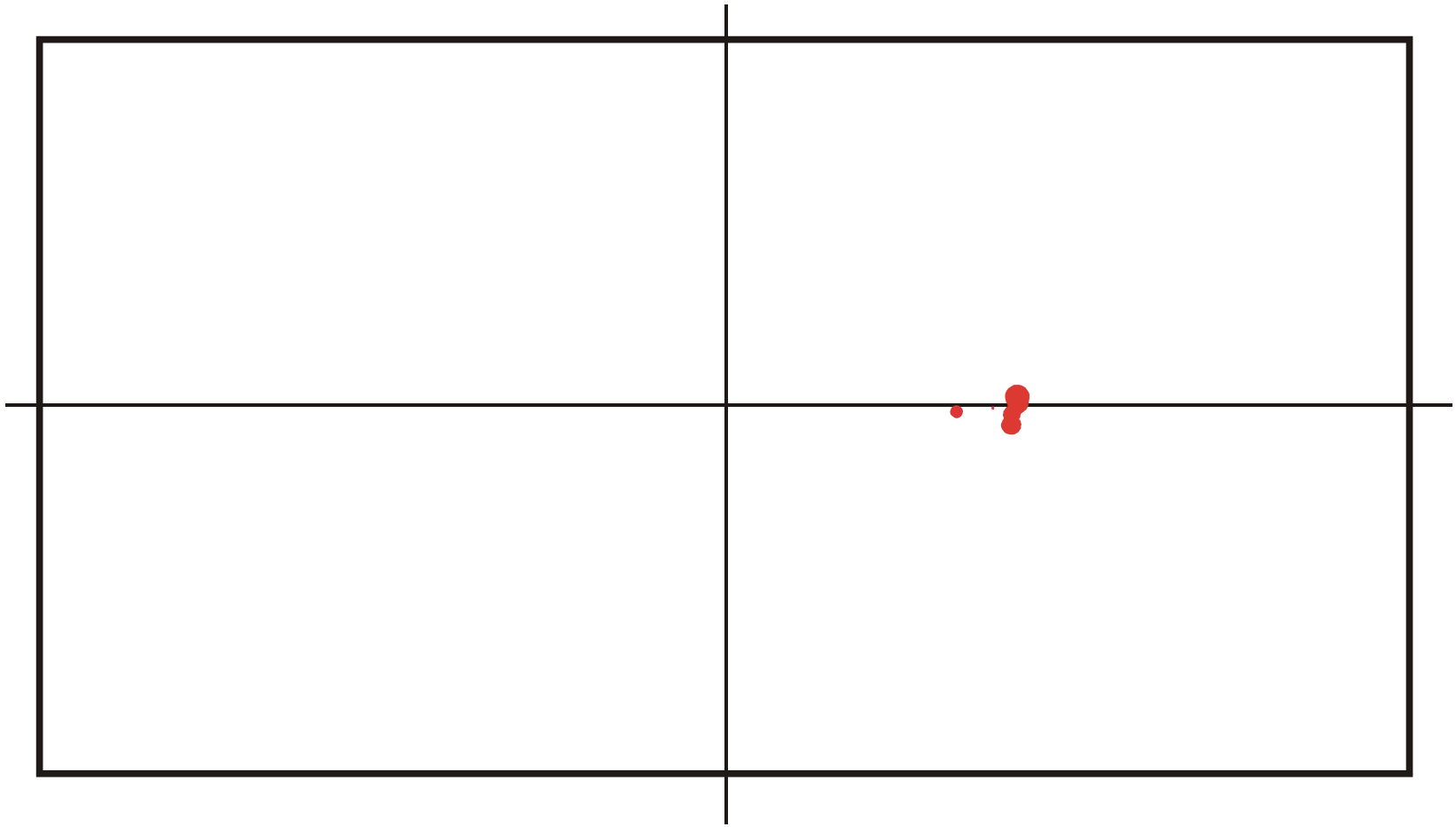
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 835



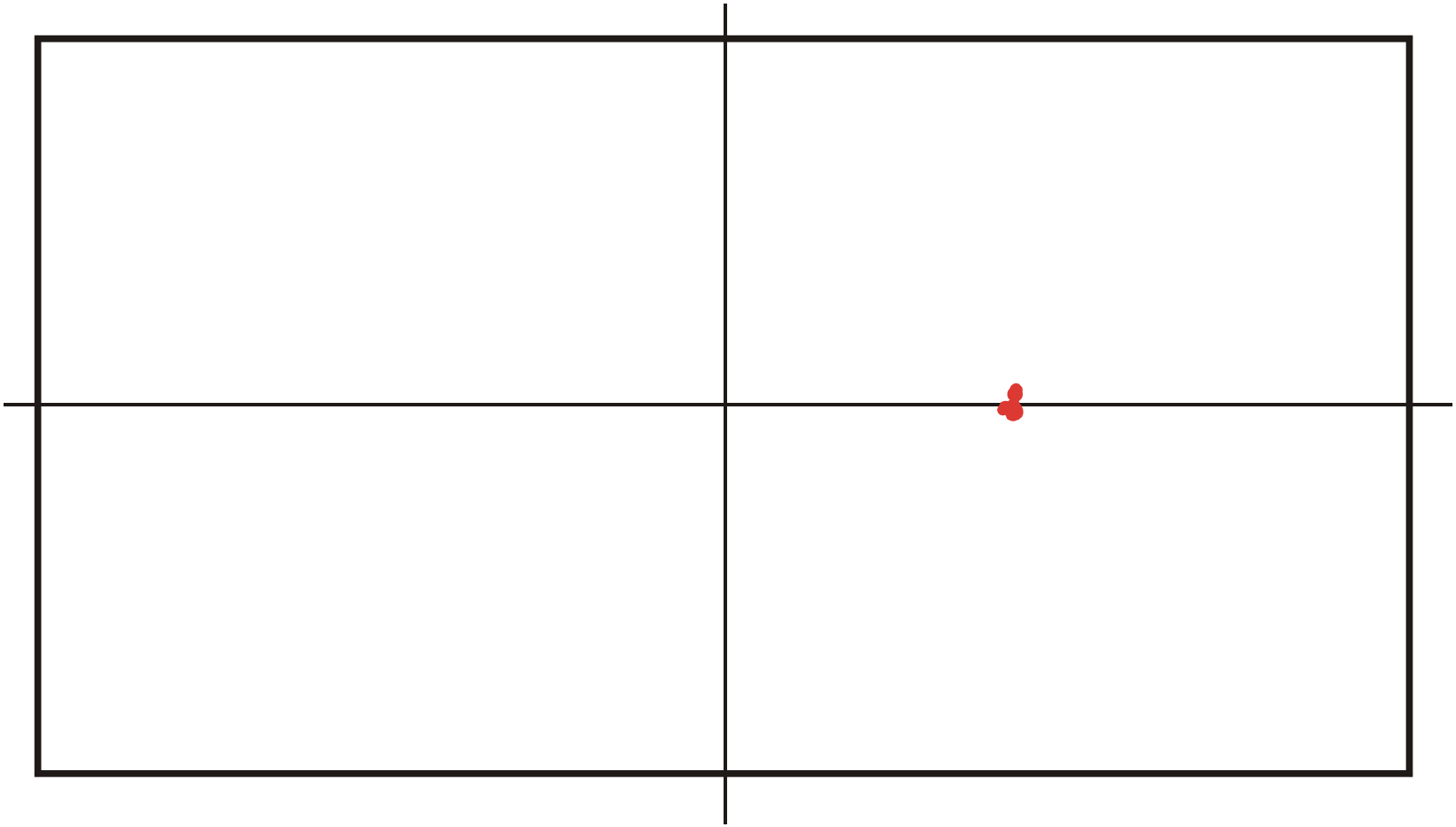
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 840



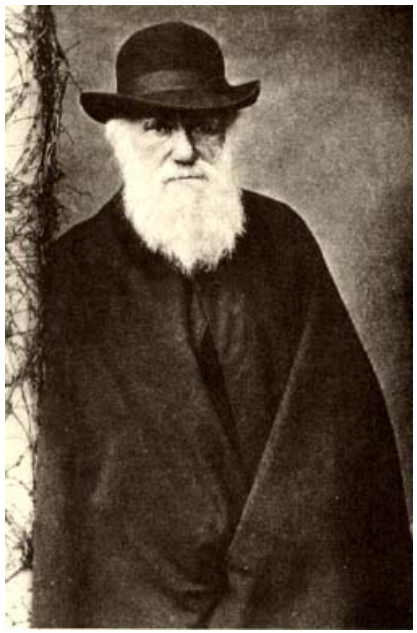
Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 845



Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 850



Ausbreitung und Entwicklung einer Population auf einem neutralen Netzwerk: t = 855



„…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)

MOTOO KIMURA

Motoo Kimura's Populationsgenetik der Neutral Evolution.

Evolutionary rate at the molecular level. *Nature* **217**: 624-626, 1955.

The Neutral Theory of Molecular Evolution. Cambridge University Press. Cambridge, UK, 1983.

THE NEUTRAL THEORY
OF MOLECULAR EVOLUTION

MOTOO KIMURA
National Institute of Genetics, Japan

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

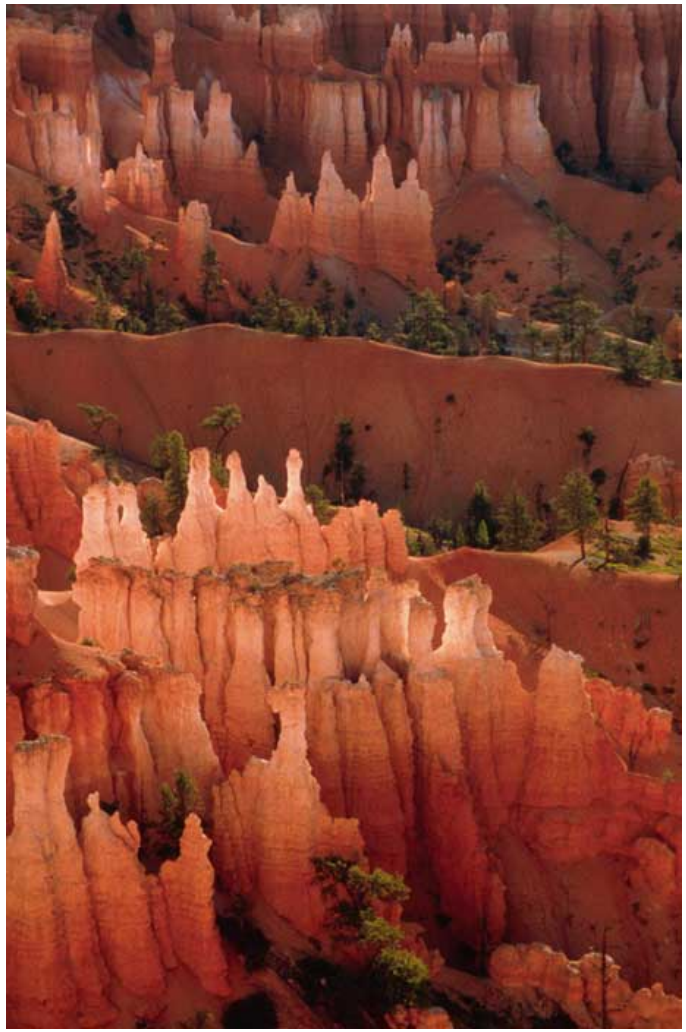
CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

Mount Fuji

Die Umgebungen von RNA-Molekülen im Sequenzraum sehen **nicht** so aus wie **glatte Landschaften**. Sie weisen einen hohen Grad von **Rauheit** oder Zerklüftetheit auf. ...

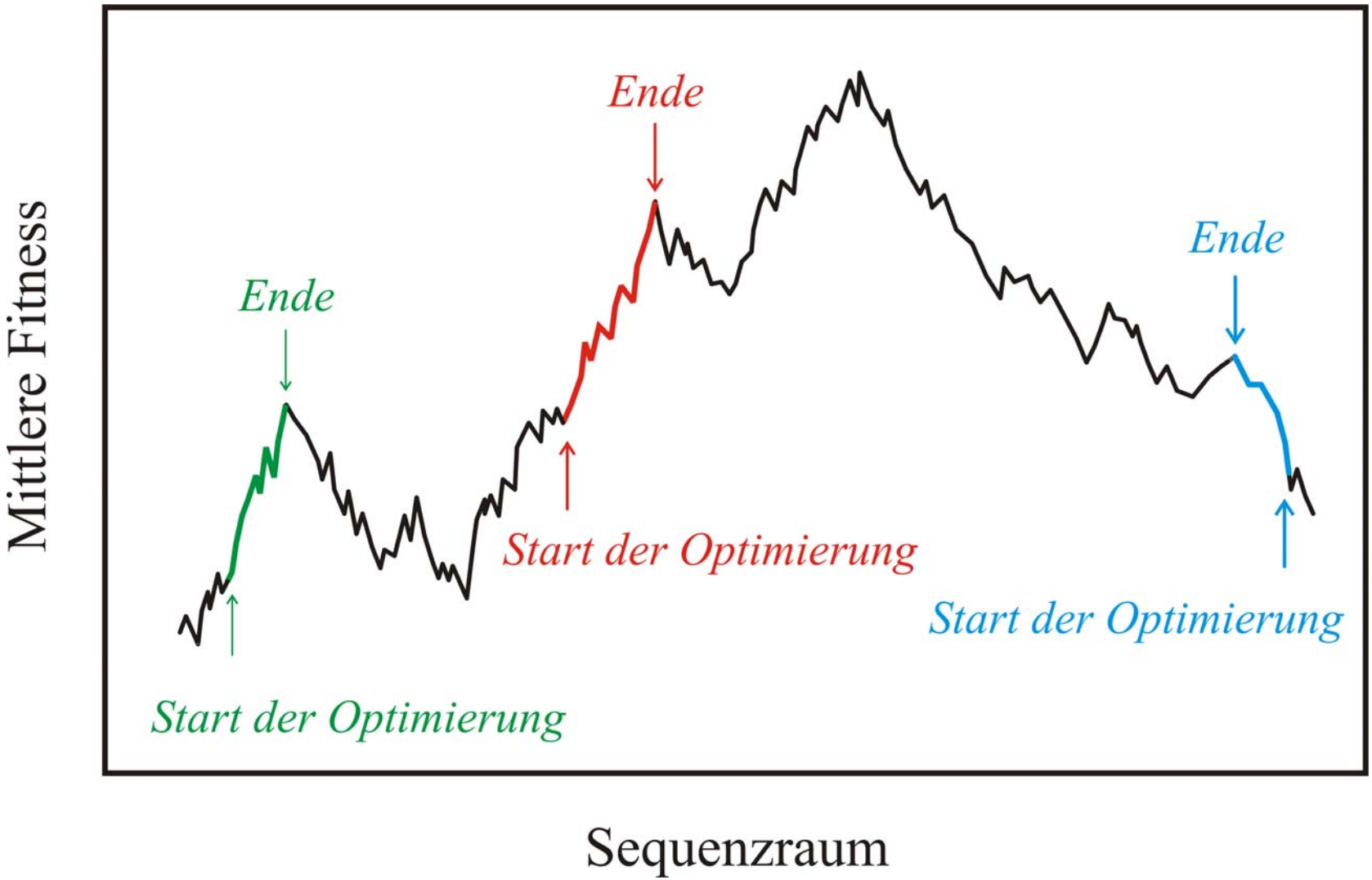
Beispiel einer “glatten” Landschaft auf der Erde

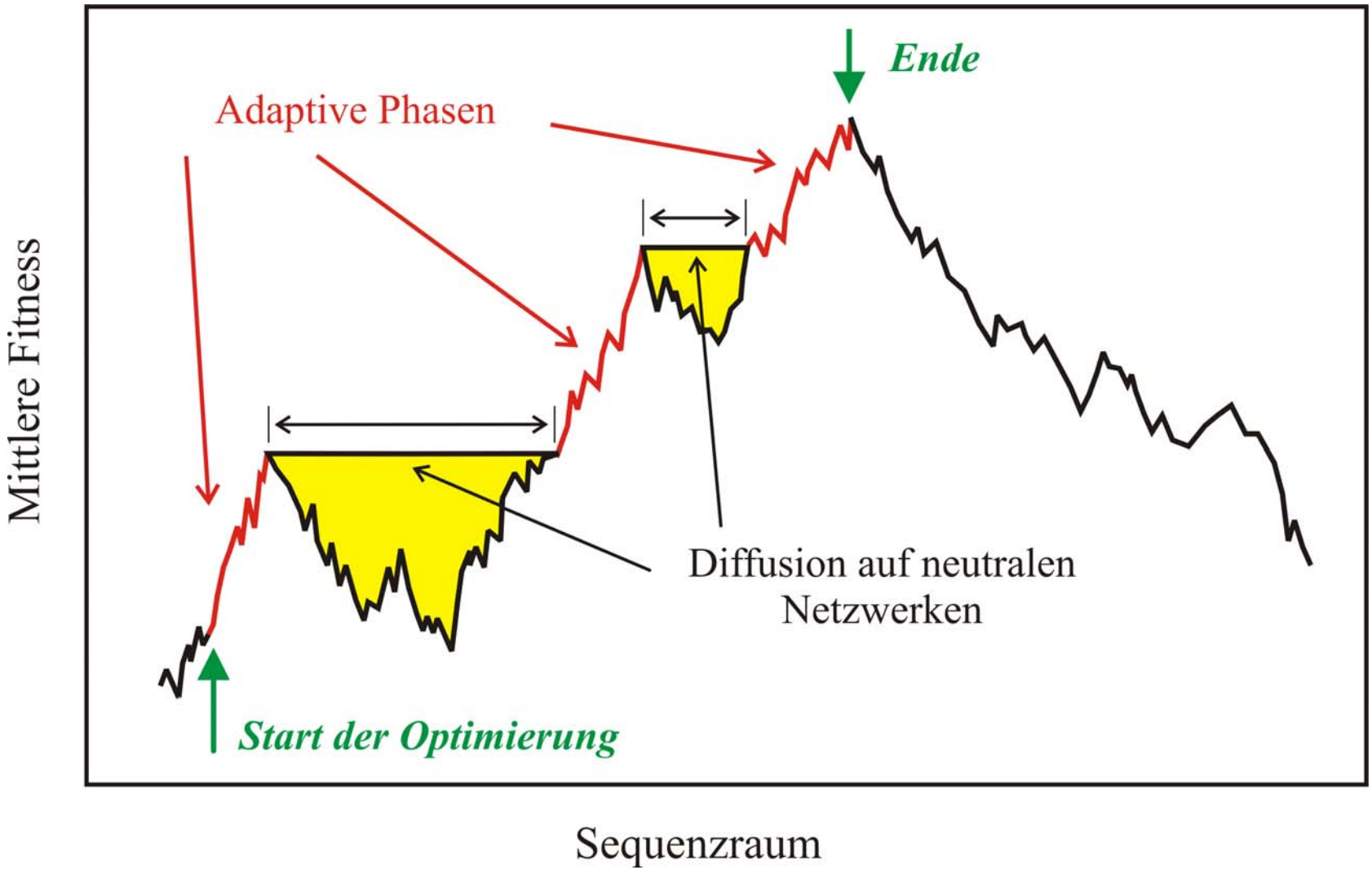
Dolomiten



Bryce Canyon

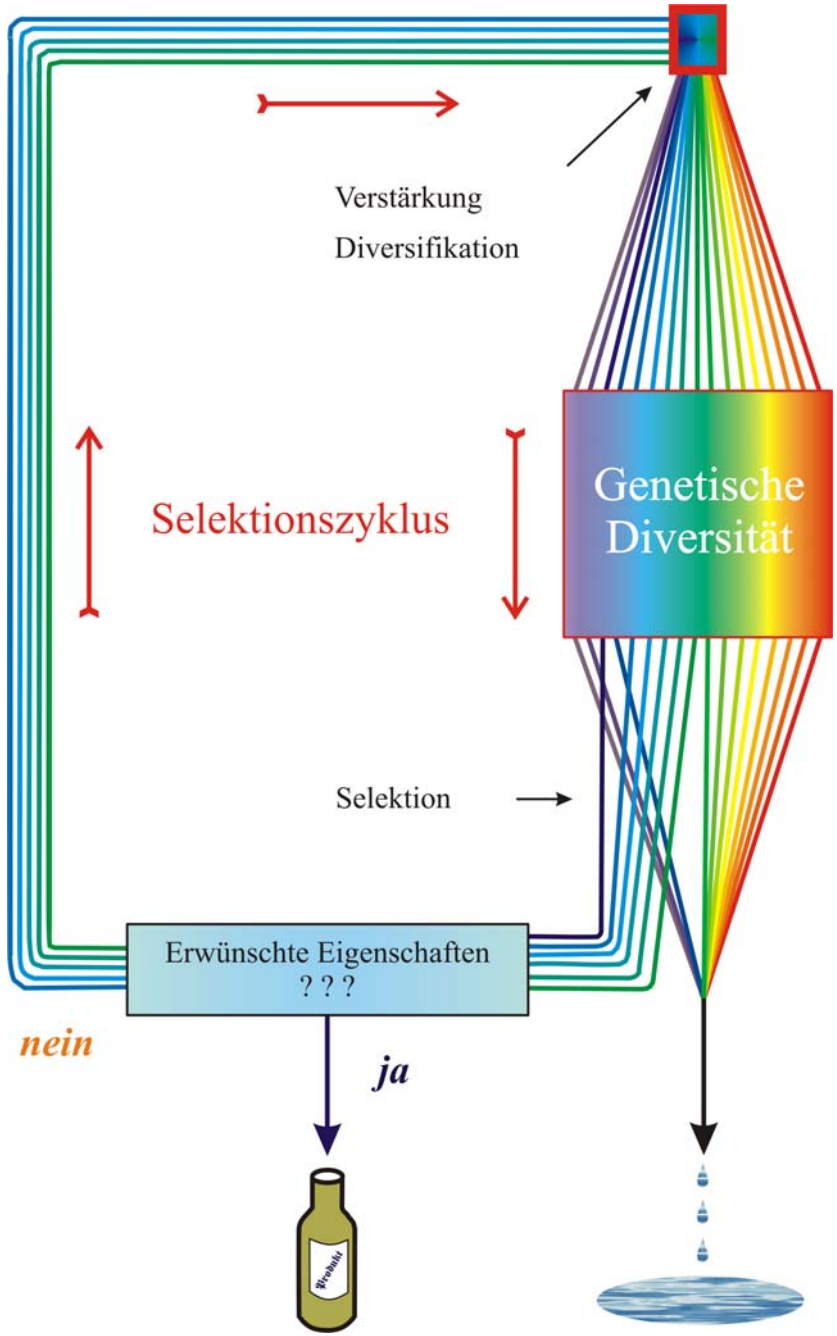
Beispiele rauher Landschaften auf der Erde


Mittlere Fitness
Ende
Ende
Ende
Start der Optimierung
Start der Optimierung
Start der Optimierung
Sequenzraum


Ende
Adaptive Phasen
Diffusion auf neutralen
Netzwerken
Start der Optimierung
Mittlere Fitness
Sequenzraum

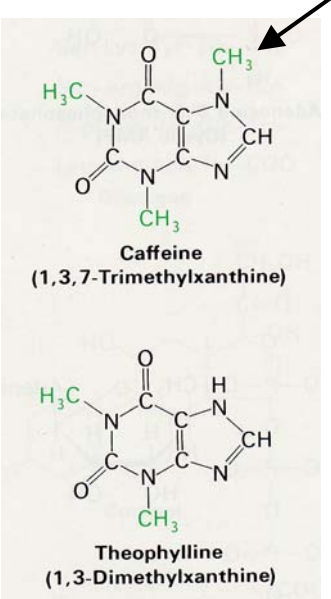
Grand Canyon

Beipiel einer irdischen Landschaft mit neutralen Graten und Plateaus



Schematischer Selektionszyklus
bei der evolutionären Optimierung
in der Biotechnologie

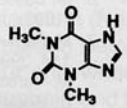
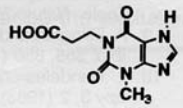
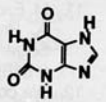
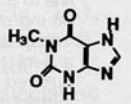
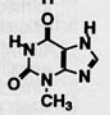
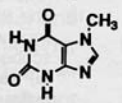
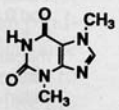
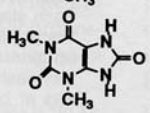
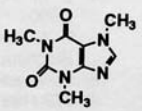
Zusätzliche Methylgruppe



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

Dissoziationskonstanten und Spezifität von Theophyllin, Coffein, und verwandten Harnsäurederivaten für die Bindung an ein diskriminierendes RNA-Molekül: Aptamer TCT8-4

Table 1. Competition binding analysis with TCT8-4 RNA. The chemical structures are shown for a series of derivatives used in competitive binding experiments with TCT8-4 RNA (Fig. 2) (*20*). The right column represents the affinity of the competitor relative to theophylline, $K_d(c)/K_d(t)$, where $K_d(c)$ is the individual competitor dissociation constant and $K_d(t)$ is the competitive dissociation constant of theophylline. Certain data (denoted by >) are minimum values that were limited by the solubility of the competitor. Each experiment was carried out in duplicate. The average error is shown.

Compound	Structure	$K_d(c)$ (μM)	$K_d(c)/K_d(t)$
Theophylline		0.32 ± 0.13	1
CP-theophylline		0.93 ± 0.20	2.9
Xanthine		8.5 ± 0.40	27
1-Methylxanthine		9.0 ± 0.30	28
3-Methylxanthine		2.0 ± 0.7	6.3
7-Methylxanthine		> 500	>1500
3,7-Dimethylxanthine		> 500	> 1500
1,3-Dimethyluric acid		> 1000	>3100
Hypoxanthine		49 ± 10	153
Caffeine		3500 ± 1500	10,900

Schematische Zeichnung der Aptamer Bindungsstelle für das Theophyllin Molekül

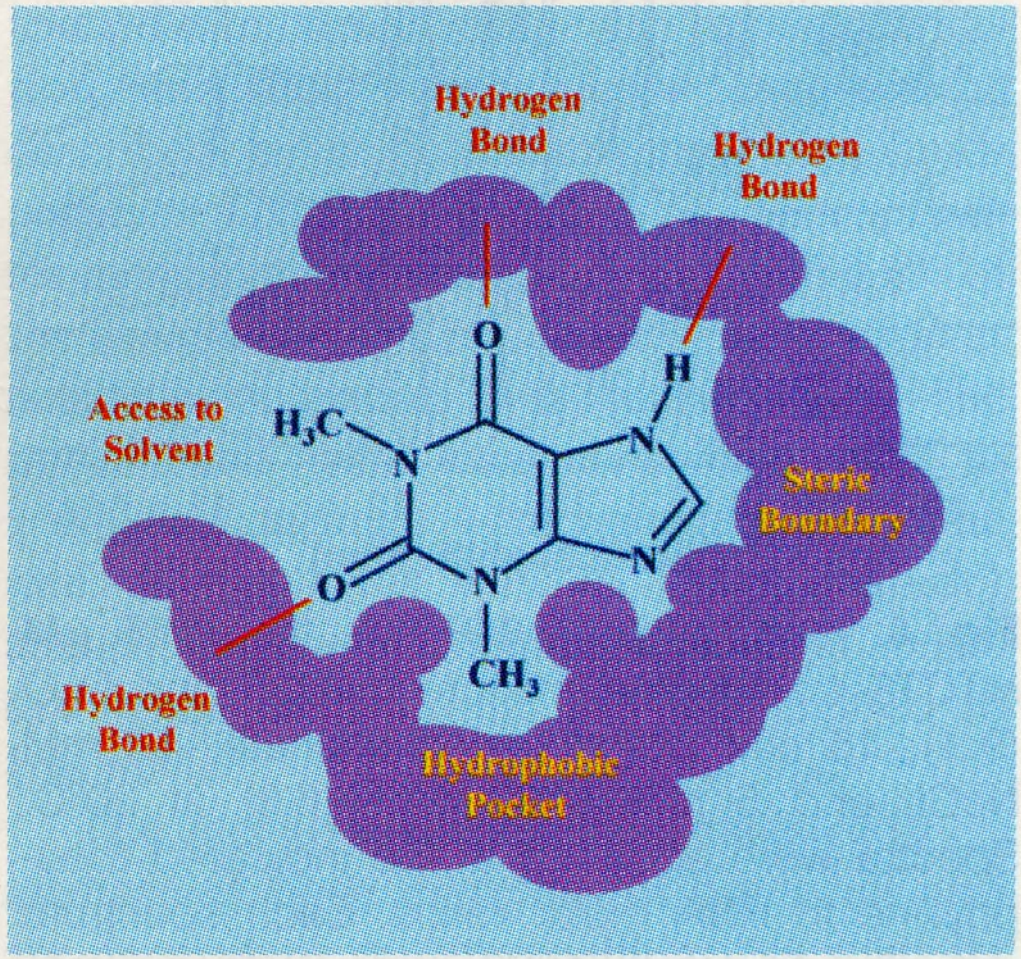


Fig. 3. Schematic representation of the RNA (purple) binding site for theophylline (blue).

Bulletin of Mathematical Biology, Vol. 59, No. 2, pp. 339–397, 1997
Elsevier Science Inc.

0092-8240/97 $17.00 + 0.00

S0092-8240(96)00089-4

GENERIC PROPERTIES OF COMBINATORY MAPS: NEUTRAL NETWORKS OF RNA SECONDARY STRUCTURES[1]

■ 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*

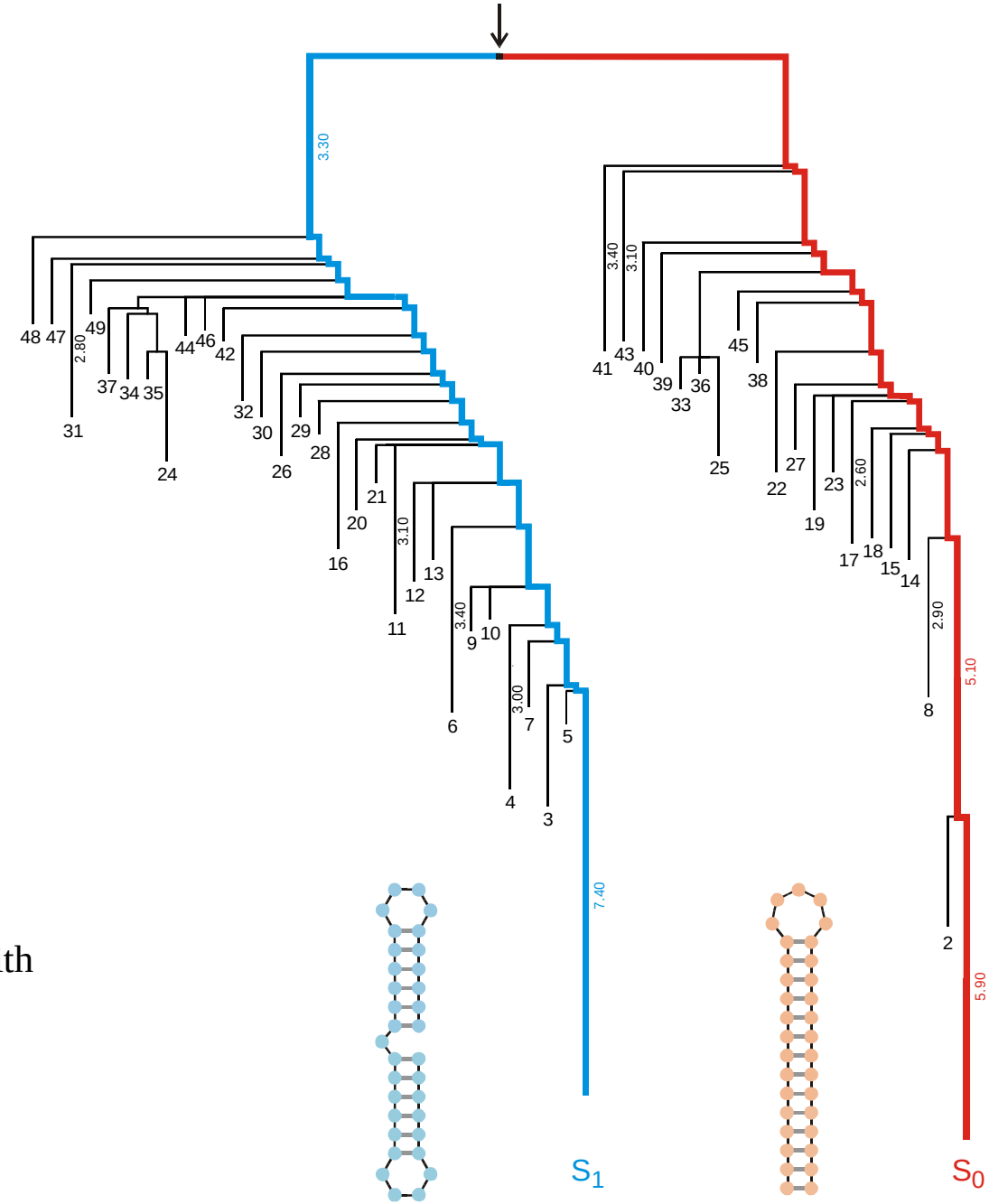
THEOREM 5. INTERSECTION-THEOREM. *Let* s *and* s′ *be arbitrary secondary structures and* **C**[s], **C**[s′] *their corresponding compatible sequences. Then,*

$$\mathbf{C}[s] \cap \mathbf{C}[s'] \neq \varnothing.$$

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 $\gamma(s, s') \cong D_m$ 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.

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



Barrier tree of a sequence with two conformations

Ein katalytischer RNA-Schalter

E.A.Schultes, D.B.Bartel, Science **289** (2000), 448-452

minus the background levels observed in the HSP in the control (Sar1-GDP–containing) incubation that prevents COPII vesicle formation. In the microsome control, the level of p115-SNARE associations was less than 0.1%.
46. C. M. Carr, E. Grote, M. Munson, F. M. Hughson, P. J. Novick, *J. Cell Biol.* **146**, 333 (1999).
47. C. Ungermann, B. J. Nichols, H. R. Pelham, W. Wickner, *J. Cell Biol.* **140**, 61 (1998).
48. E. Grote and P. J. Novick, *Mol. Biol. Cell* **10**, 4149 (1999).
49. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
50. GST-SNARE proteins were expressed in bacteria and purified on glutathione-Sepharose beads using standard methods. Immobilized GST-SNARE protein (0.5 μM) was incubated with rat liver cytosol (20 mg) or purified recombinant p115 (0.5 μM) in 1 ml of NS buffer containing 1% BSA for 2 hours at 4°C with rotation. Beads were briefly spun (3000 rpm for 10 s) and sequentially washed three times with NS buffer and three times with NS buffer supplemented with 150 mM NaCl. Bound proteins were eluted three times in 50 μl of 50 mM tris-HCl (pH 8.5), 50 mM reduced glutathione, 150 mM NaCl, and 0.1% Triton X-100 for 15 min at 4°C with intermittent mixing, and elutes were pooled. Proteins were precipitated by MeOH/CH_3Cl and separated by SDS–polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting using p115 mAb 13F12.
51. V. Rybin *et al.*, *Nature* **383**, 266 (1996).
52. K. G. Hardwick and H. R. Pelham, *J. Cell Biol.* **119**, 513 (1992).
53. A. P. Newman, M. E. Groesch, S. Ferro-Novick, *EMBO J.* **11**, 3609 (1992).
54. A. Spang and R. Schekman, *J. Cell Biol.* **143**, 589 (1998).
55. M. F. Rexach, M. Latterich, R. W. Schekman, *J. Cell Biol.* **126**, 1133 (1994).
56. A. Mayer and W. Wickner, *J. Cell Biol.* **136**, 307 (1997).
57. M. D. Turner, H. Plutner, W. E. Balch, *J. Biol. Chem.* **272**, 13479 (1997).
58. A. Price, D. Seals, W. Wickner, C. Ungermann, *J. Cell Biol.* **148**, 1231 (2000).
59. X. Cao and C. Barlowe, *J. Cell Biol.* **149**, 55 (2000).
60. G. G. Tall, H. Hama, D. B. DeWald, B. F. Horazdovsky, *Mol. Biol. Cell* **10**, 1873 (1999).
61. C. G. Burd, M. Peterson, C. R. Cowles, S. D. Emr, *Mol. Biol. Cell* **8**, 1089 (1997).
62. M. R. Peterson, C. G. Burd, S. D. Emr, *Curr. Biol.* **9**, 159 (1999).
63. M. G. Waters, D. O. Clary, J. E. Rothman, *J. Cell Biol.* **118**, 1015 (1992).
64. D. M. Walter, K. S. Paul, M. G. Waters, *J. Biol. Chem.* **273**, 29565 (1998).
65. N. Hui *et al.*, *Mol. Biol. Cell* **8**, 1777 (1997).
66. T. E. Kreis, *EMBO J.* **5**, 931 (1986).
67. H. Plutner, H. W. Davidson, J. Saraste, W. E. Balch, *J. Cell Biol.* **119**, 1097 (1992).
68. D. S. Nelson *et al.*, *J. Cell Biol.* **143**, 319 (1998).
69. We thank G. Waters for p115 cDNA and p115 mAbs; G. Warren for p97 and p47 antibodies; R. Scheller for rbet1, membrin, and sec22 cDNAs; H. Plutner for excellent technical assistance; and P. Tan for help during the initial phase of this work. Supported by NIH grants GM 33301 and GM42336 and National Cancer Institute grant CA58689 (W.E.B.), a NIH National Research Service Award (B.D.M.), and a Wellcome Trust International Traveling Fellowship (B.B.A.).

20 March 2000; accepted 22 May 2000

One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds

Erik A. Schultes and David P. Bartel*

We describe a single RNA sequence that can assume either of two ribozyme folds and catalyze the two respective reactions. The two ribozyme folds share no evolutionary history and are completely different, with no base pairs (and probably no hydrogen bonds) in common. Minor variants of this sequence are highly active for one or the other reaction, and can be accessed from prototype ribozymes through a series of neutral mutations. Thus, in the course of evolution, new RNA folds could arise from preexisting folds, without the need to carry inactive intermediate sequences. This raises the possibility that biological RNAs having no structural or functional similarity might share a common ancestry. Furthermore, functional and structural divergence might, in some cases, precede rather than follow gene duplication.

Related protein or RNA sequences with the same folded conformation can often perform very different biochemical functions, indicating that new biochemical functions can arise from preexisting folds. But what evolutionary mechanisms give rise to sequences with new macromolecular folds? When considering the origin of new folds, it is useful to picture, among all sequence possibilities, the distribution of sequences with a particular fold and function. This distribution can range very far in sequence space (*1*). For example, only seven nucleotides are strictly conserved among the group I self-splicing introns, yet secondary (and presumably tertiary) structure within the core of the ribozyme is preserved (*2*). Because these disparate isolates have the same fold and function, it is thought that they descended from a common ancestor through a series of mutational variants that were each functional. Hence, sequence heterogeneity among divergent isolates implies the existence of paths through sequence space that have allowed neutral drift from the ancestral sequence to each isolate. The set of all possible neutral paths composes a "neutral network," connecting in sequence space those widely dispersed sequences sharing a particular fold and activity, such that any sequence on the network can potentially access very distant sequences by neutral mutations (*3–5*).

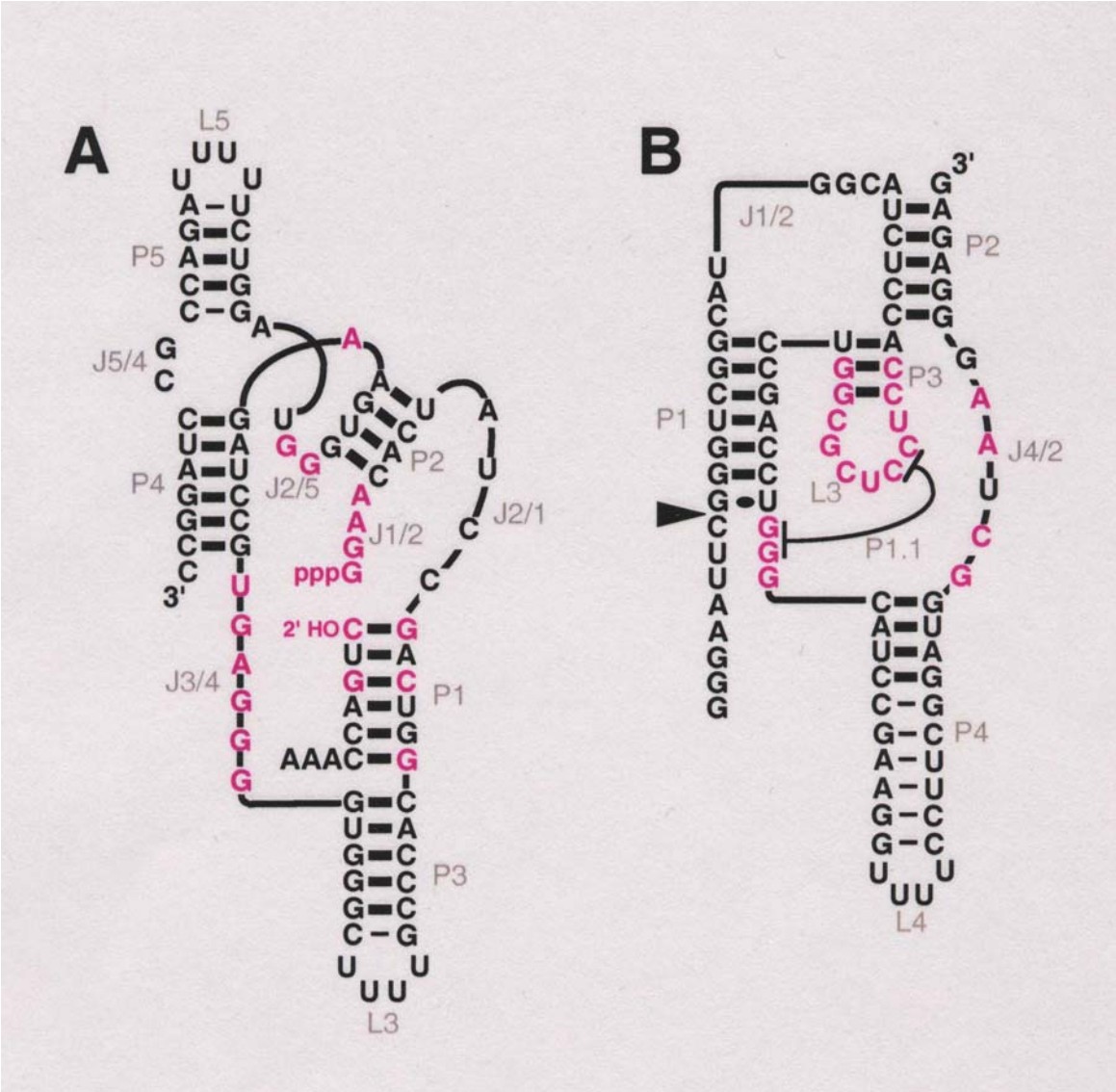
Theoretical analyses using algorithms for predicting RNA secondary structure have suggested that different neutral networks are interwoven and can approach each other very closely (*3*, *5–8*). Of particular interest is whether ribozyme neutral networks approach each other so closely that they intersect. If so, a single sequence would be capable of folding into two different conformations, would have two different catalytic activities, and could access by neutral drift every sequence on both networks. With intersecting networks, RNAs with novel structures and activities could arise from previously existing ribozymes, without the need to carry nonfunctional sequences as evolutionary intermediates. Here, we explore the proximity of neutral networks experimentally, at the level of RNA function. We describe a close apposition of the neutral networks for the hepatitis delta virus (HDV) self-cleaving ribozyme and the class III self-ligating ribozyme.

In choosing the two ribozymes for this investigation, an important criterion was that they share no evolutionary history that might confound the evolutionary interpretations of our results. Choosing at least one artificial ribozyme ensured independent evolutionary histories. The class III ligase is a synthetic ribozyme isolated previously from a pool of random RNA sequences (*9*). It joins an oligonucleotide substrate to its 5′ terminus. The prototype ligase sequence (Fig. 1A) is a shortened version of the most active class III variant isolated after 10 cycles of in vitro selection and evolution. This minimal construct retains the activity of the full-length isolate (*10*). The HDV ribozyme carries out the site-specific self-cleavage reactions needed during the life cycle of HDV, a satellite virus of hepatitis B with a circular, single-stranded RNA genome (*11*). The prototype HDV construct for our study (Fig. 1B) is a shortened version of the antigenomic HDV ribozyme (*12*), which undergoes self-cleavage at a rate similar to that reported for other antigenomic constructs (*13*, *14*).

The prototype class III and HDV ribozymes have no more than the 25% sequence identity expected by chance and no fortuitous structural similarities that might favor an intersection of their two neutral networks. Nevertheless, sequences can be designed that simultaneously satisfy the base-pairing requirements

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

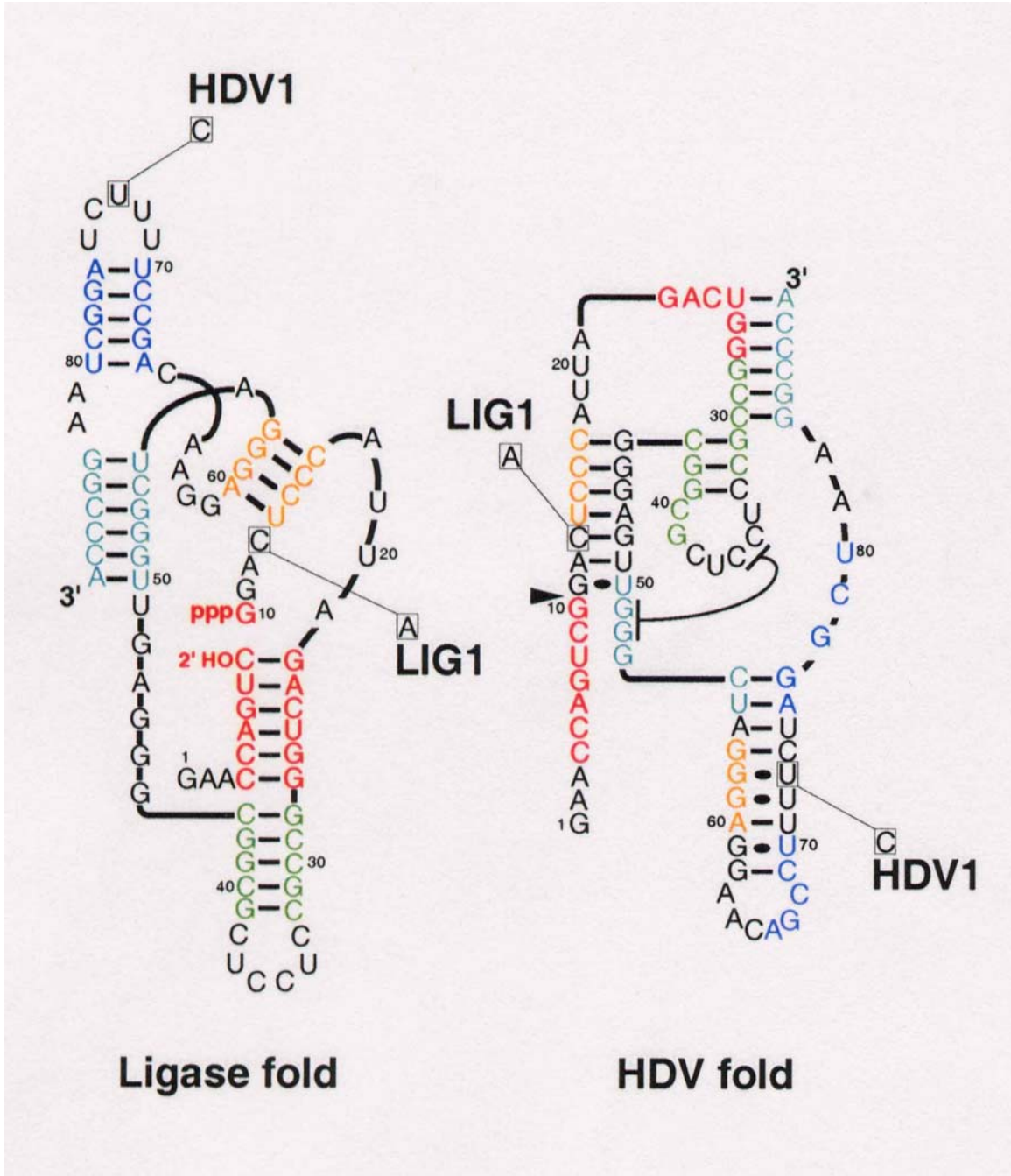
*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

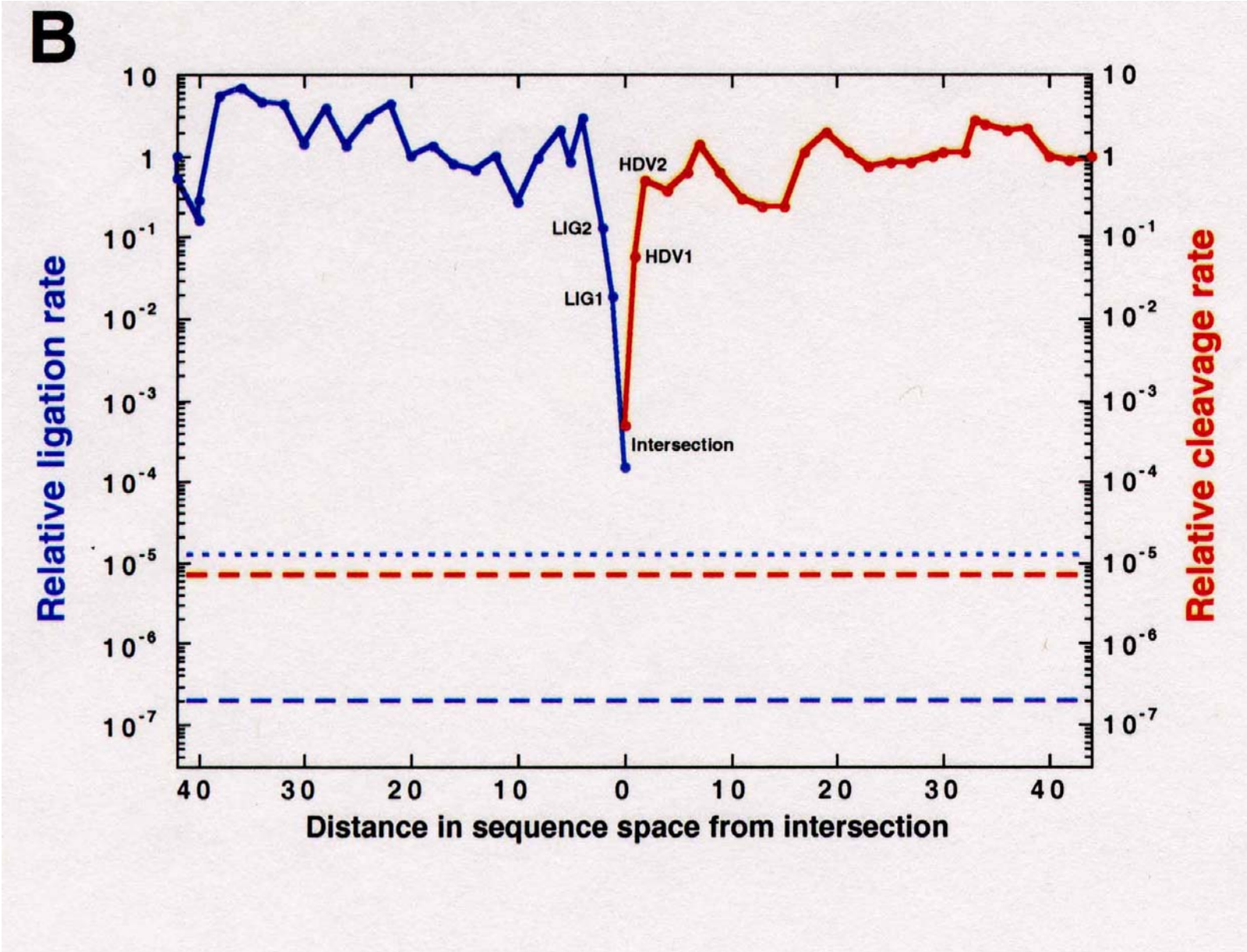


Zwei Ribozyme der Kettenlänge n = 88 Nukleotide: Eine künstliche Ligase (**A**) und ein natürliches Spaltungsribozym aus Hepatitis-δ-Virus (**B**)

Die Sequenz an der **Schnittstelle**:

Ein RNA-Molekül mit 88 Nukleotiden, welches beide Strukturen ausbilden kann



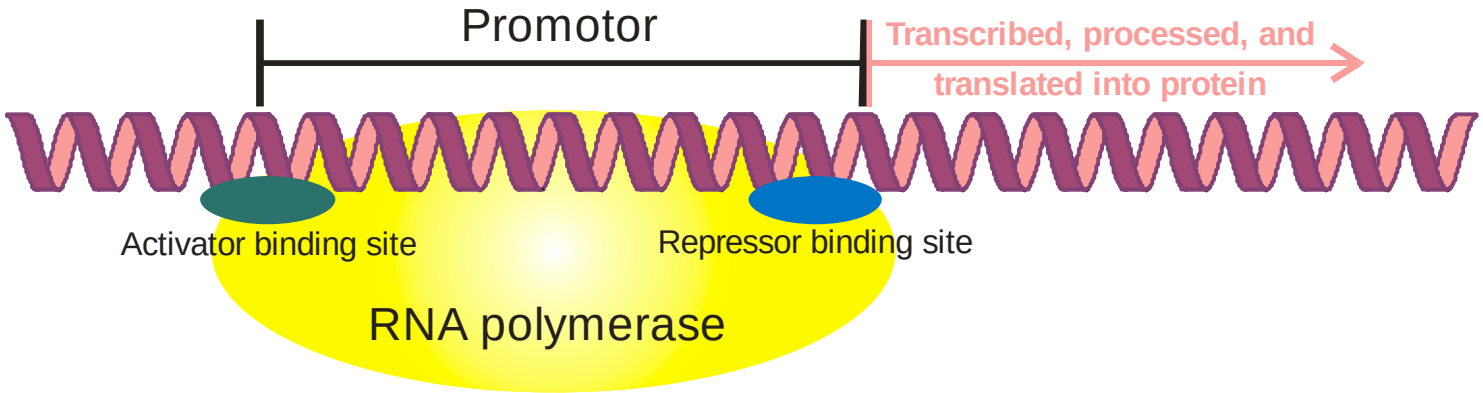


Zwei neutrale Pfade durch den Sequenzraum unter Erhalt von Struktur und katalytischer Aktivität

1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit
2. Ribonukleinsäuremoleküle als Phänotypen
3. Evolution im Reagenzglas und im Computer
4. **Genetische und metabolische Netzwerke**
5. Ein minimaler Metabolismus als Phänotyp
6. Evolution und Komplexität

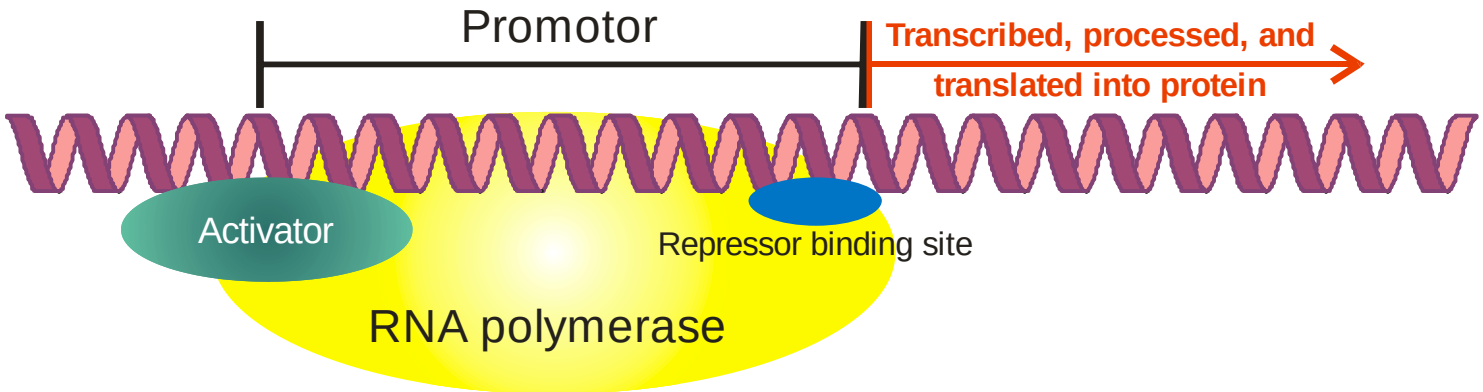
State I:

basal state



State II:

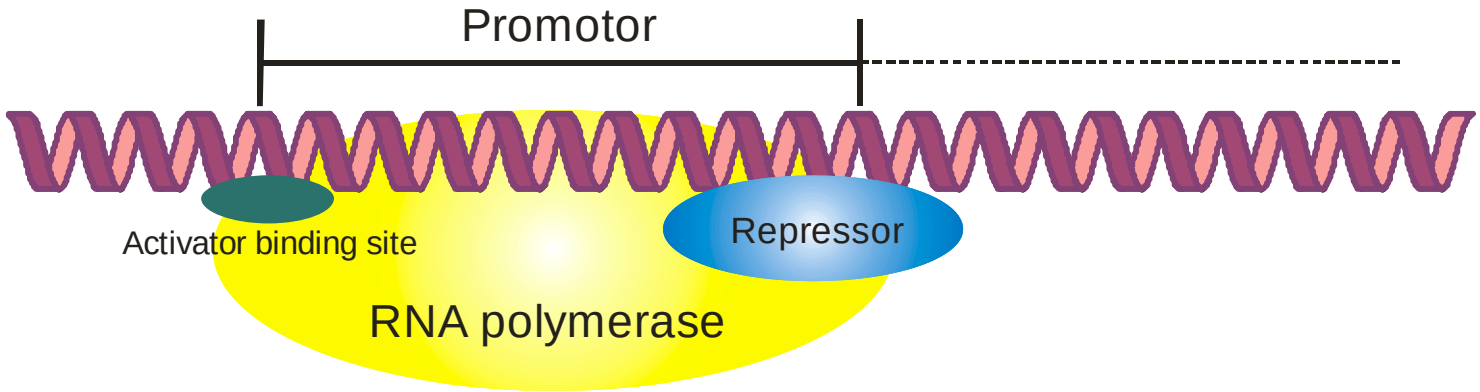
active state



Aktive Zustände einer einfachen Genregulation

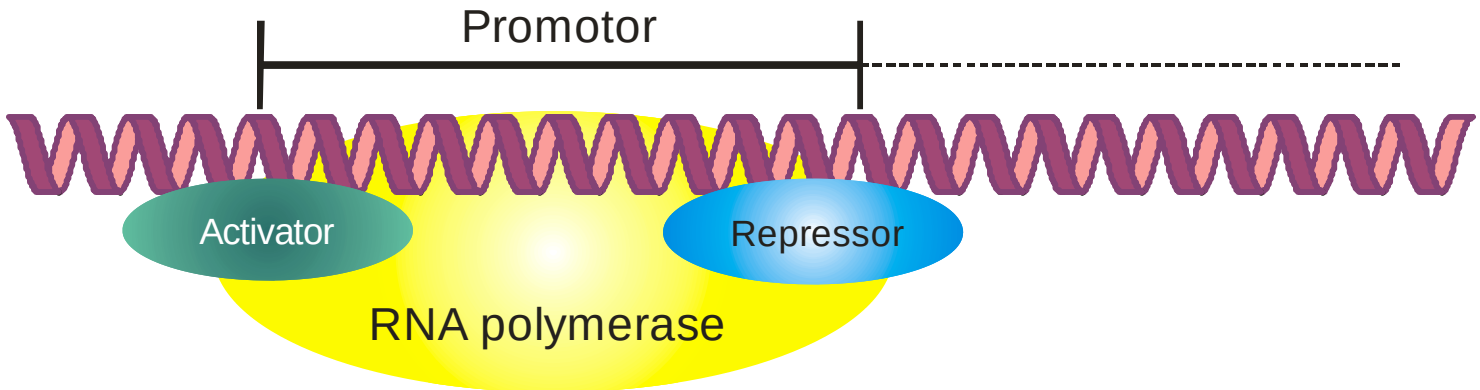
State III:

inactive state



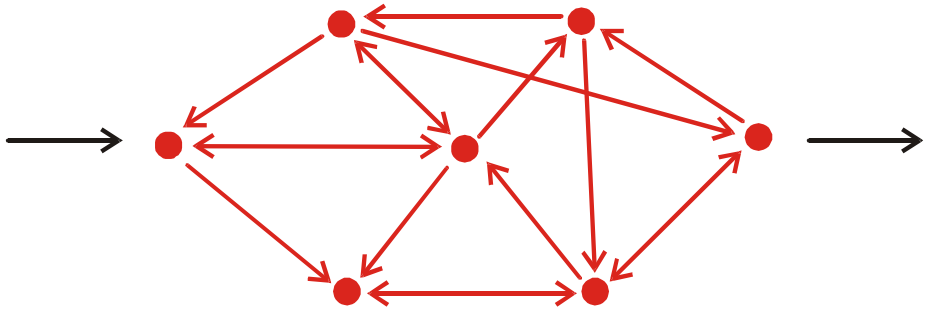
State III:

inactive state

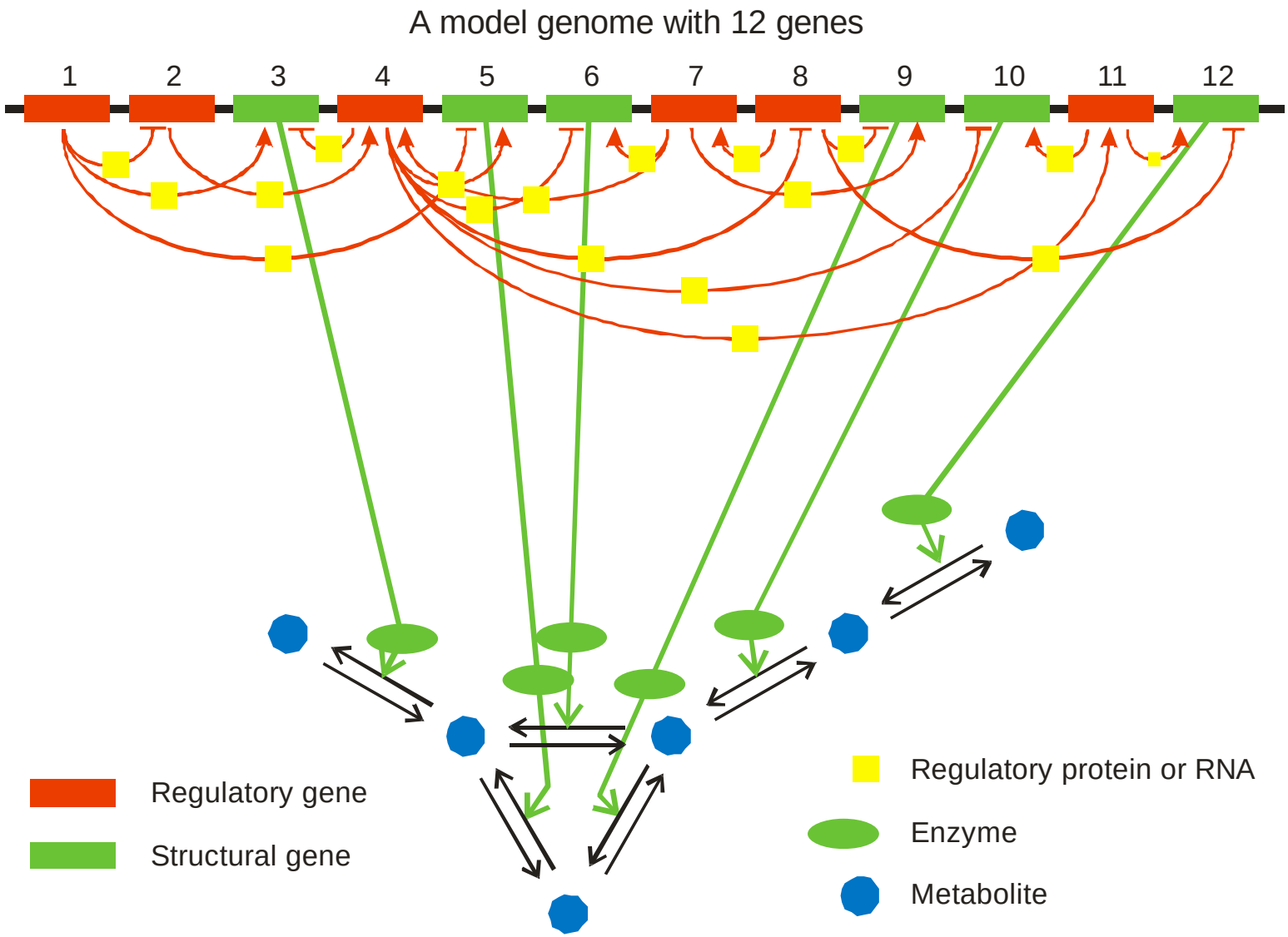


Abgeschalteter Zustande einer einfachen Genregulation

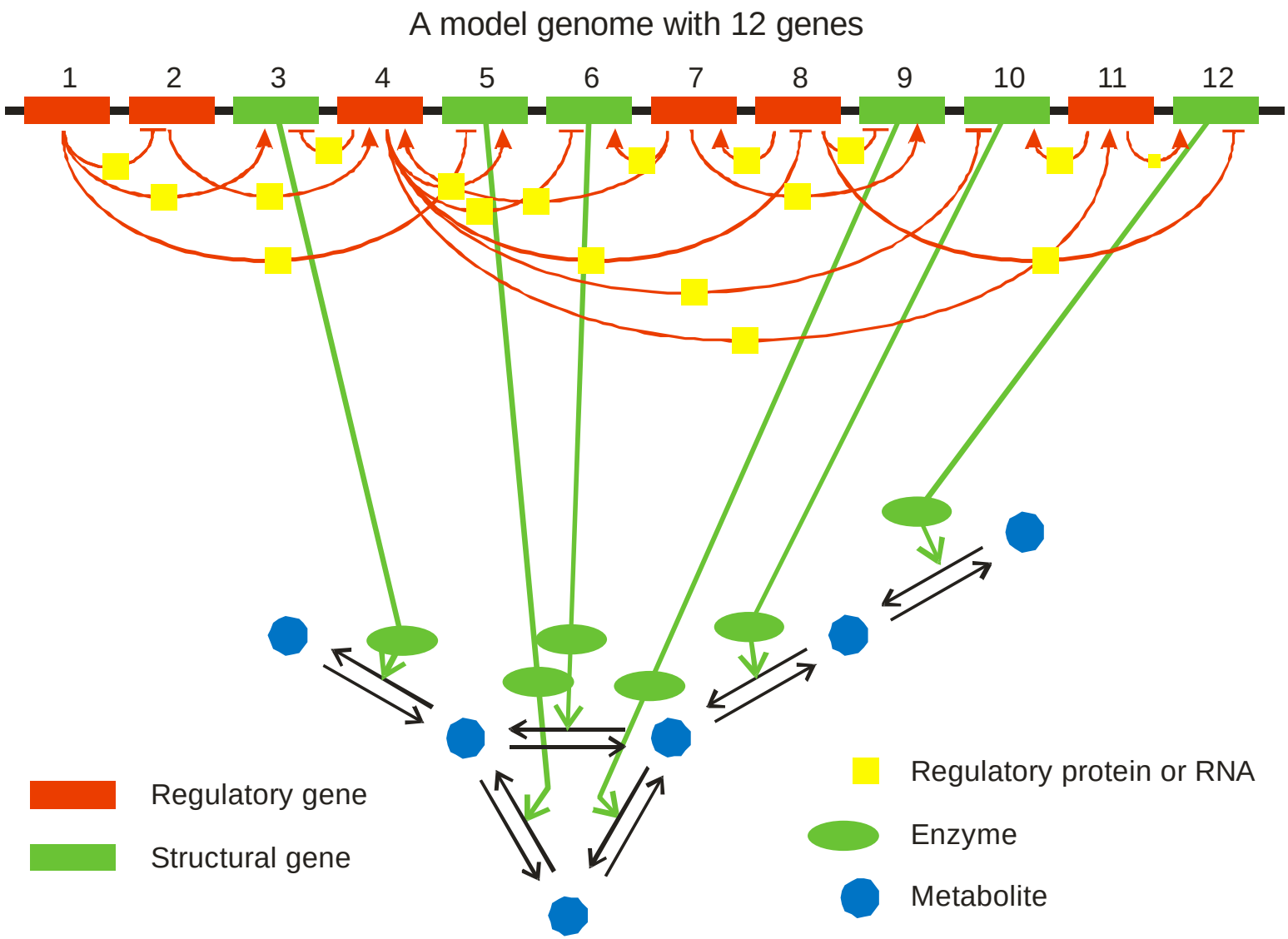
Linear chain



Network

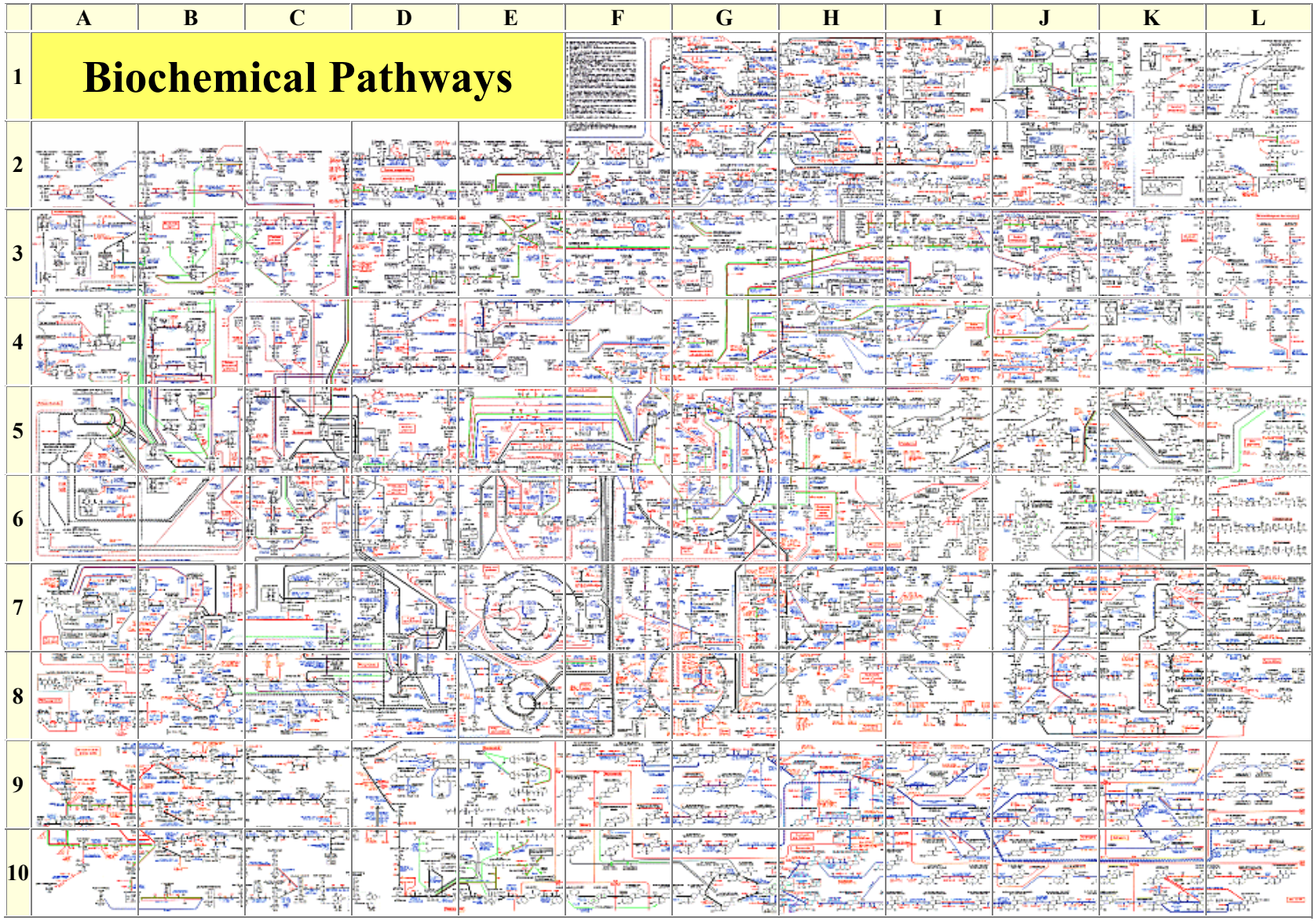


Skizze eines genetischen und metabolischen Netzwerks



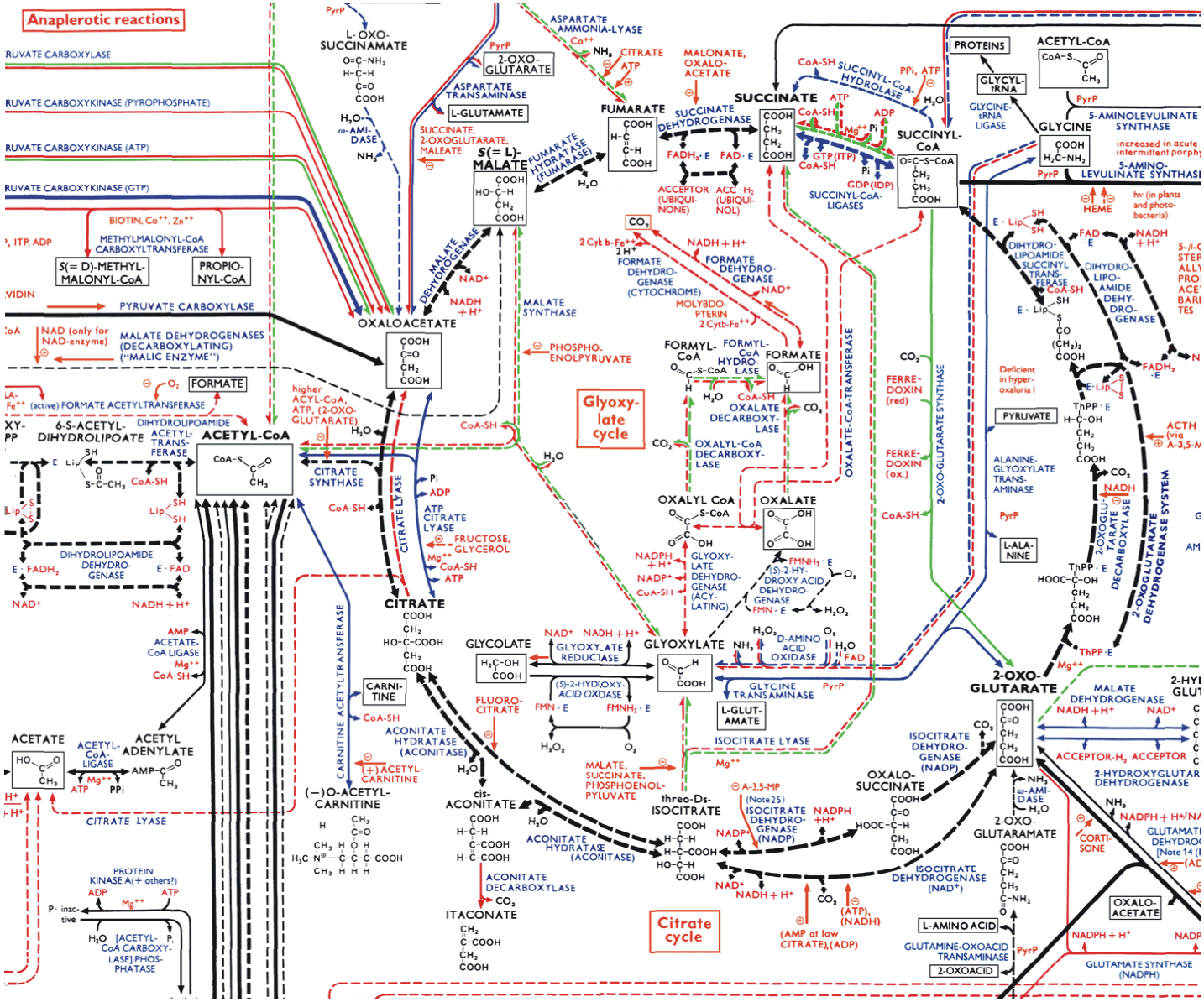
Skizze eines genetischen und metabolischen Netzwerks

„gen(etic and met)**abolic“**

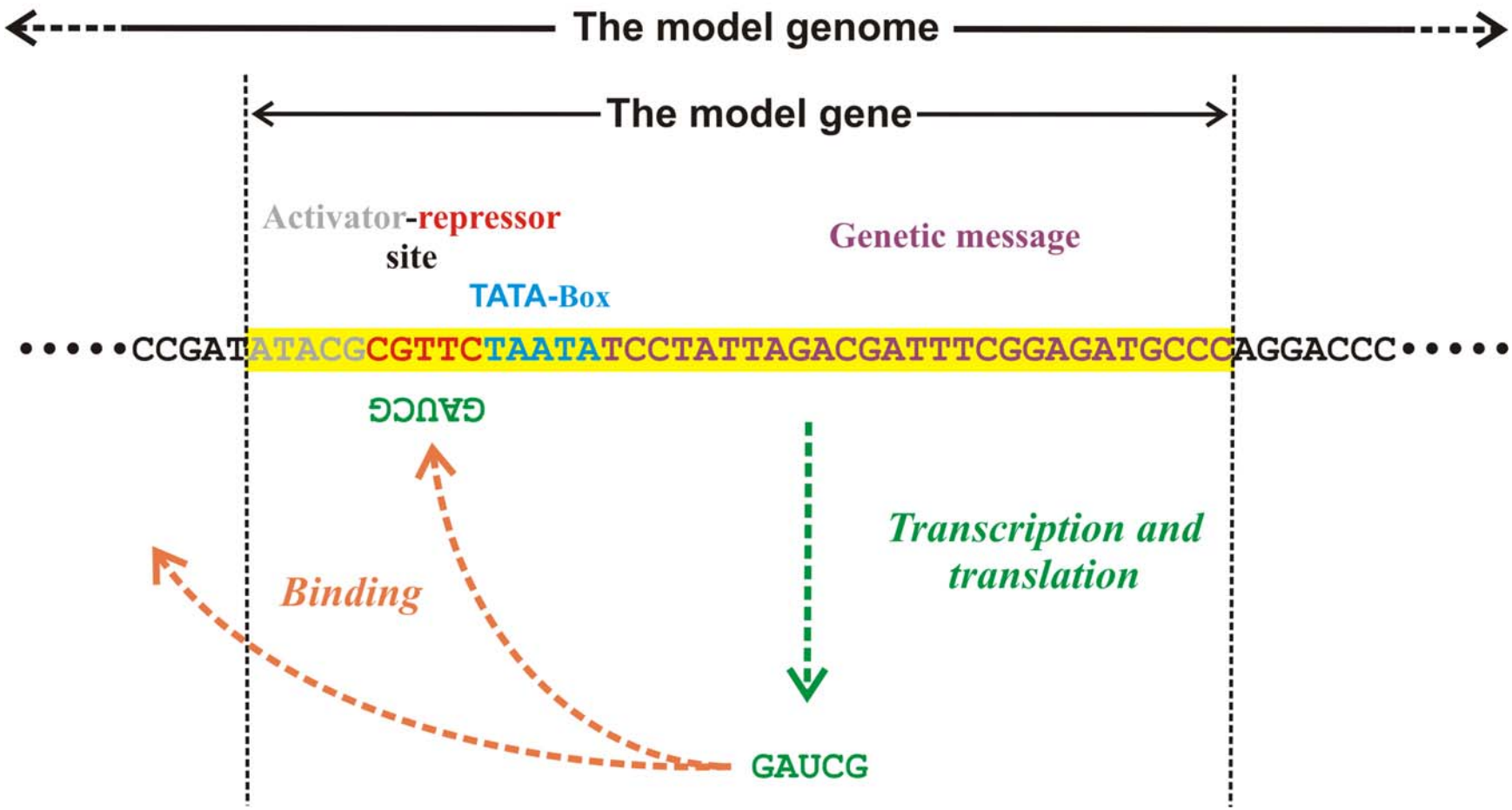


The reaction network of cellular metabolism published by Boehringer-Ingelheim.

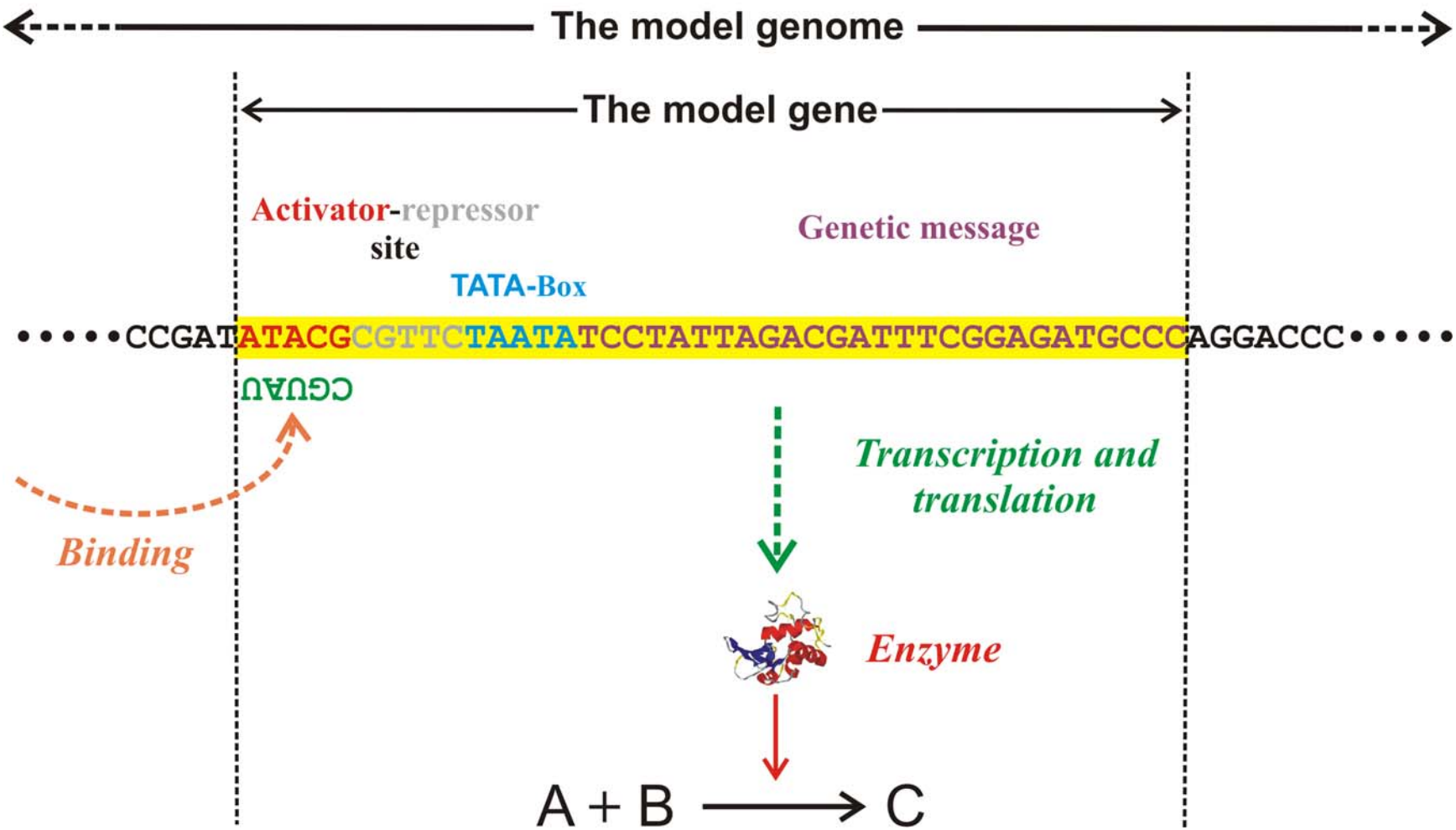
The citric acid or Krebs cycle (enlarged from previous slide).



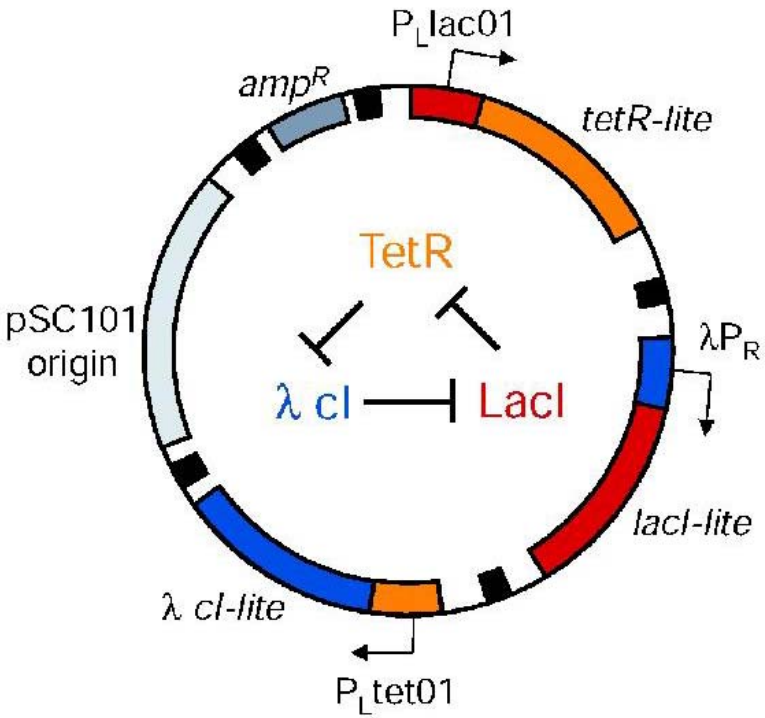
1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit
2. Ribonukleinsäuremoleküle als Phänotypen
3. Evolution im Reagenzglas und im Computer
4. Genetische und metabolische Netzwerke
5. **Ein minimaler Metabolismus als Phänotyp**
6. Evolution und Komplexität

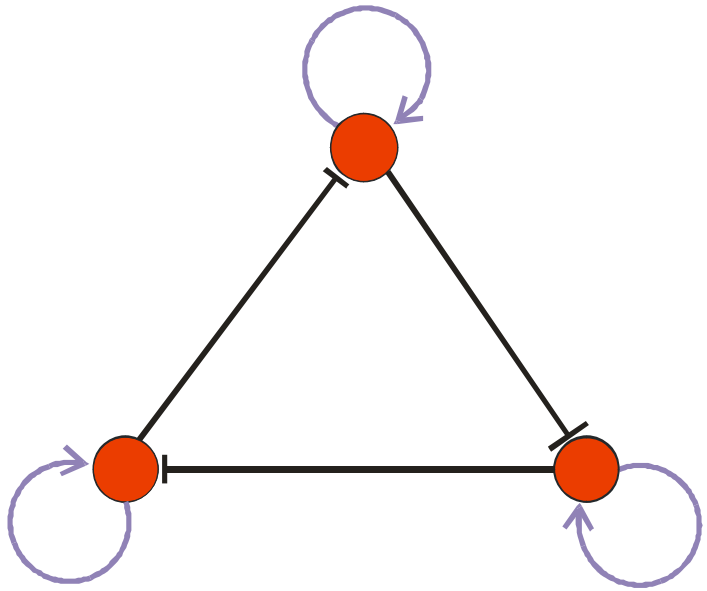


Modell eines regulatorischen Gens in ***MiniCellSim***



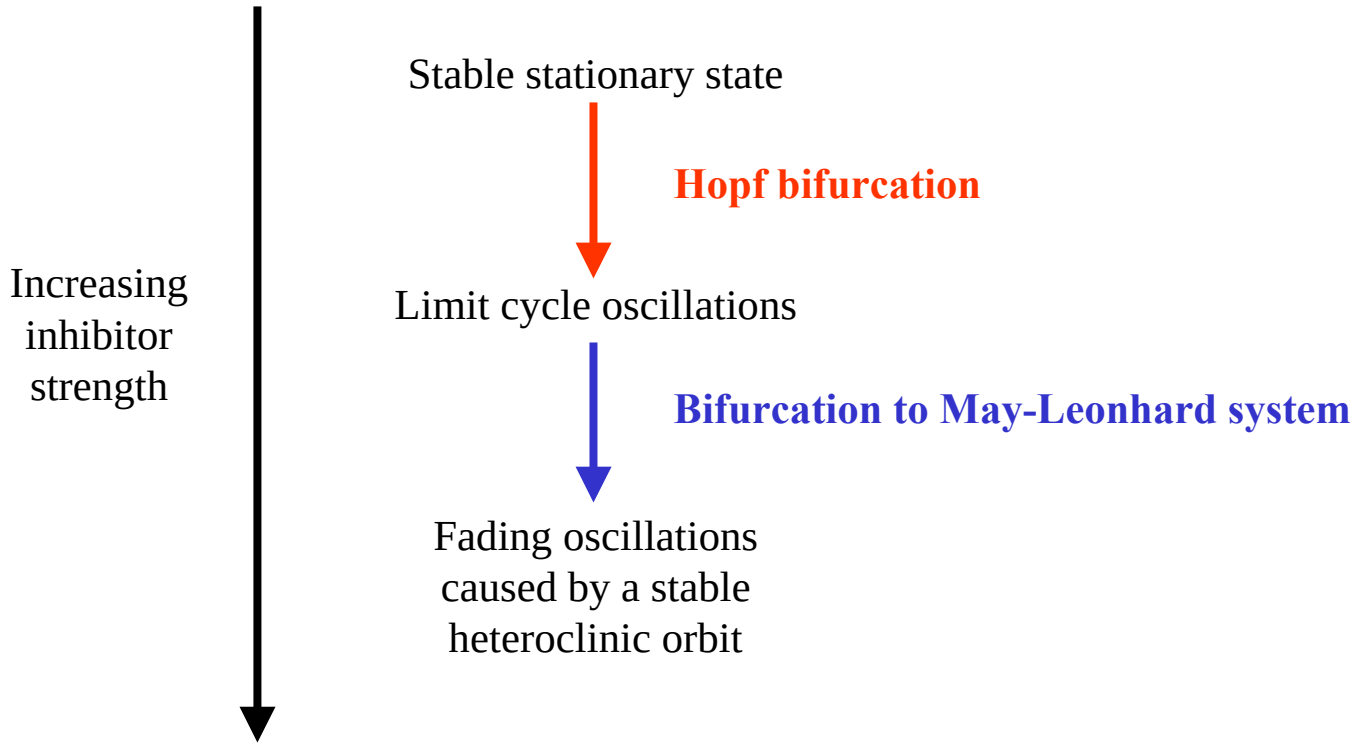
Modell eines strukturellen Gens in ***MiniCellSim***

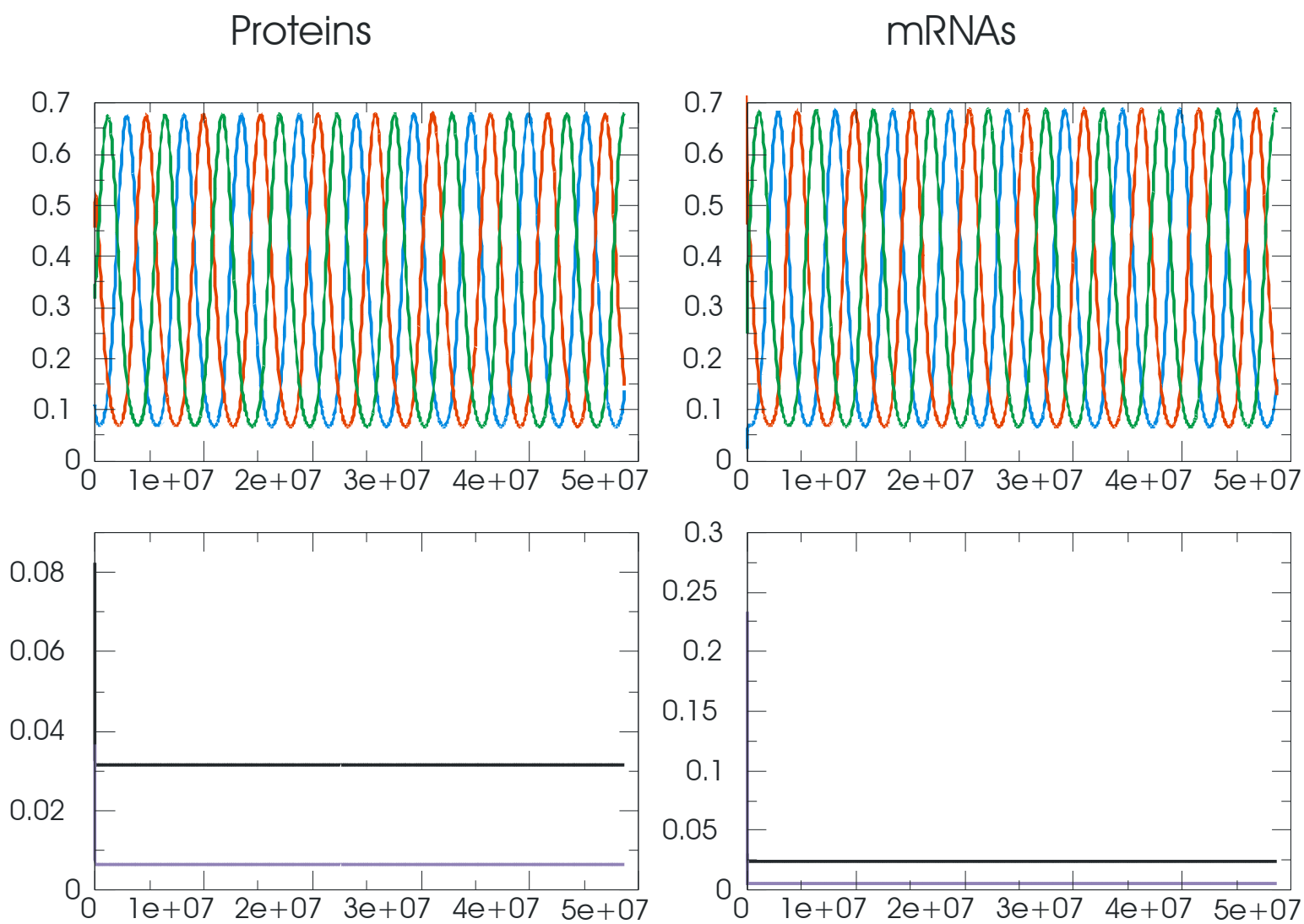




Ein experimentelles Beispiel simuliert durch ***MiniCellSim***

The repressilator: M.B. Ellowitz, S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**:335-338, 2002


Increasing inhibitor strength
Stable stationary state
Hopf bifurcation
Limit cycle oscillations
Bifurcation to May-Leonhard system
Fading oscillations caused by a stable heteroclinic orbit



Oszillationen im simulierten **Respressilatorsystem**

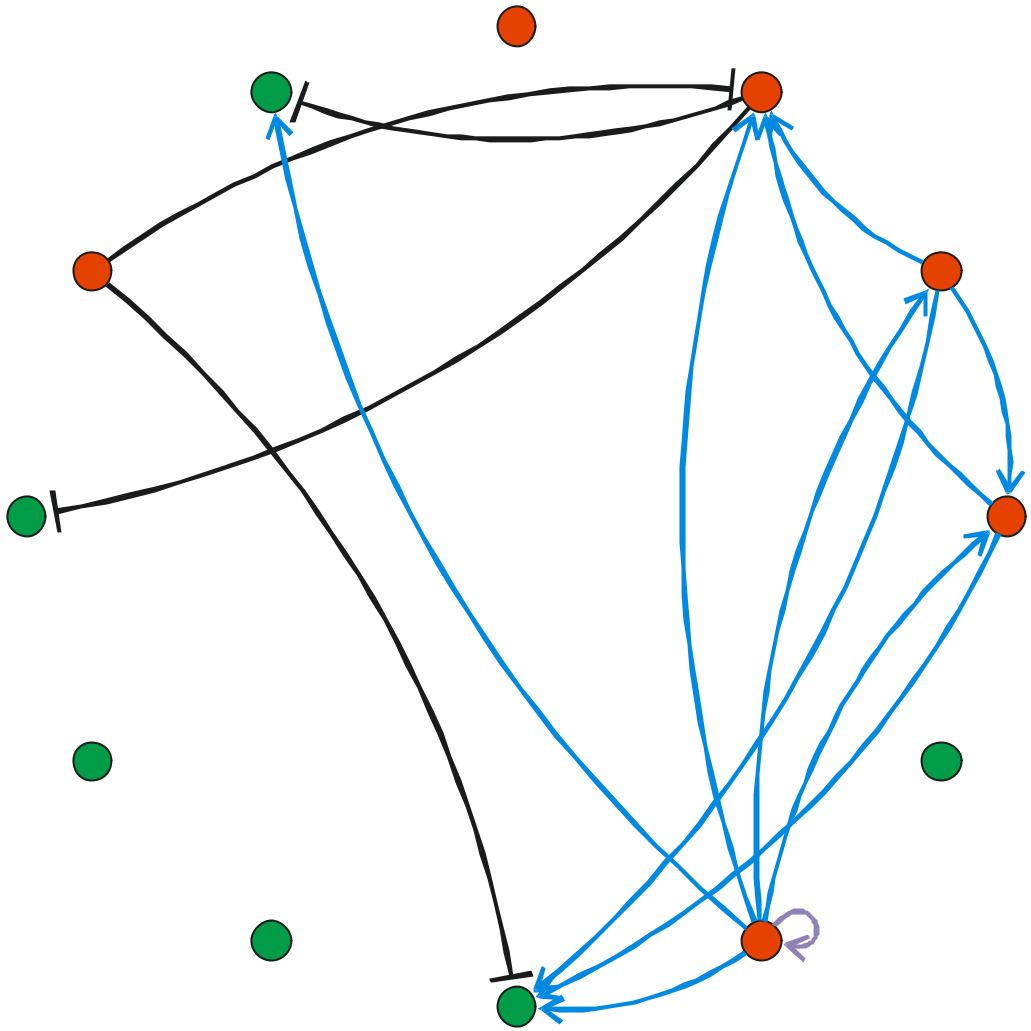
Evolutionary time: 0000

Number of genes: **12**

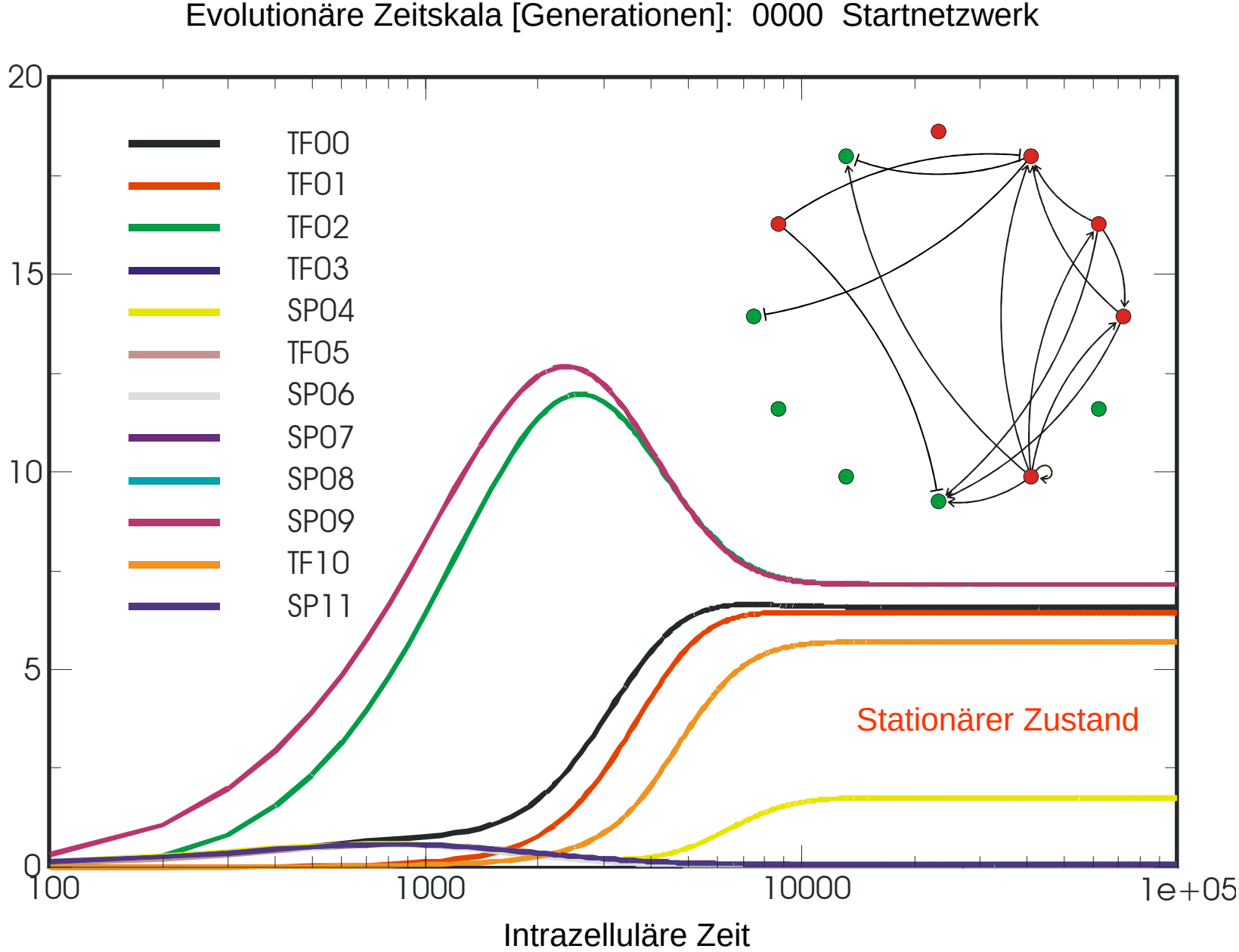
06 structural + 06 regulatory

Number of interactions: **15**

04 inhibitory + 10 activating +
+ 1 self-activating



Ein Reaktionsnetzwerk gebildet von einer Zufallssequenz aus n=200 Nukleotiden


Evolutionäre Zeitskala [Generationen]: 0000 Startnetzwerk
20
15
10
5
0
100
1000
10000
1e+05
Intrazelluläre Zeit
TF00
TF01
TF02
TF03
SP04
TF05
SP06
SP07
SP08
SP09
TF10
SP11
Stationärer Zustand

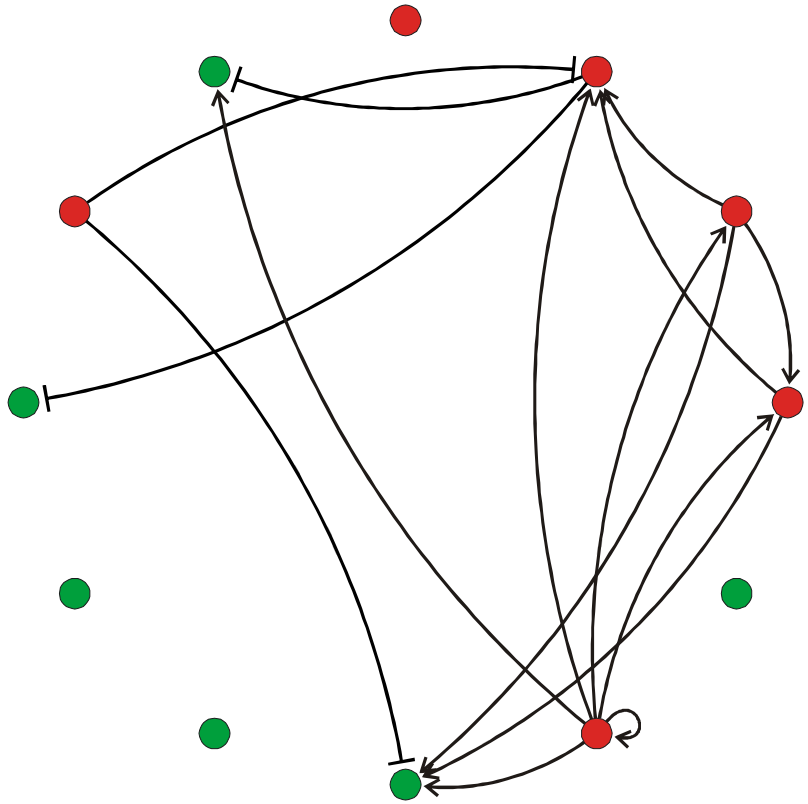
Evolution of a genetic and metabolic network:

Anfangsgenome: Zufallssequenz der Länge **n = 200**, **AUGC** alphabet

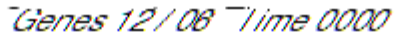
Genlänge: **n = 25**

Simulation with a mutation rate: **p = 0.01**

Evolutionäre Zeiteinheit >> Intrazelluläre Zeiteinheit



Zahl der Gene: gesamt / Strukturgene Regulatorgene



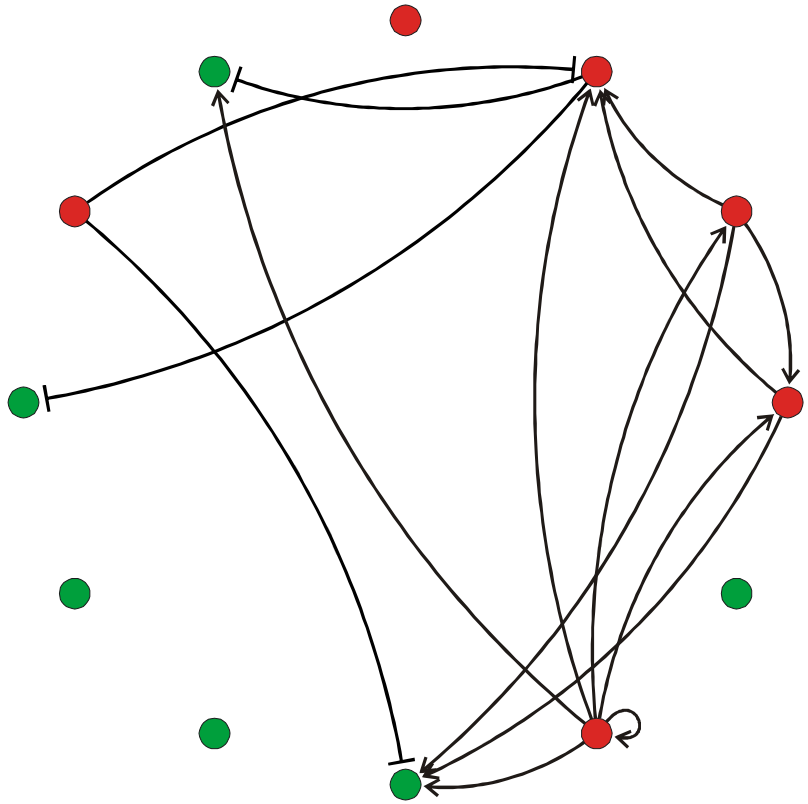
Evolution of a genetic and metabolic network:

Anfangsgenome: Zufallssequenz der Länge **n = 200**, **AUGC** alphabet

Genlänge: **n = 25**

Simulation with a mutation rate: **p = 0.01**

Evolutionäre Zeiteinheit >> Intrazelluläre Zeiteinheit



Beobachtete Ereignisse:

(i) Verlust eines Gens durch Zerstörung des Initiierungssignals "**TA**" (Analogon der üblichen "**TATA** Box")

(ii) Erzeugung eines Gens

(iii) Veränderung der Kanten durch mutations-bedingte Änderung der Affinitäten der Translationsprodukte zu den Bindungsstellen

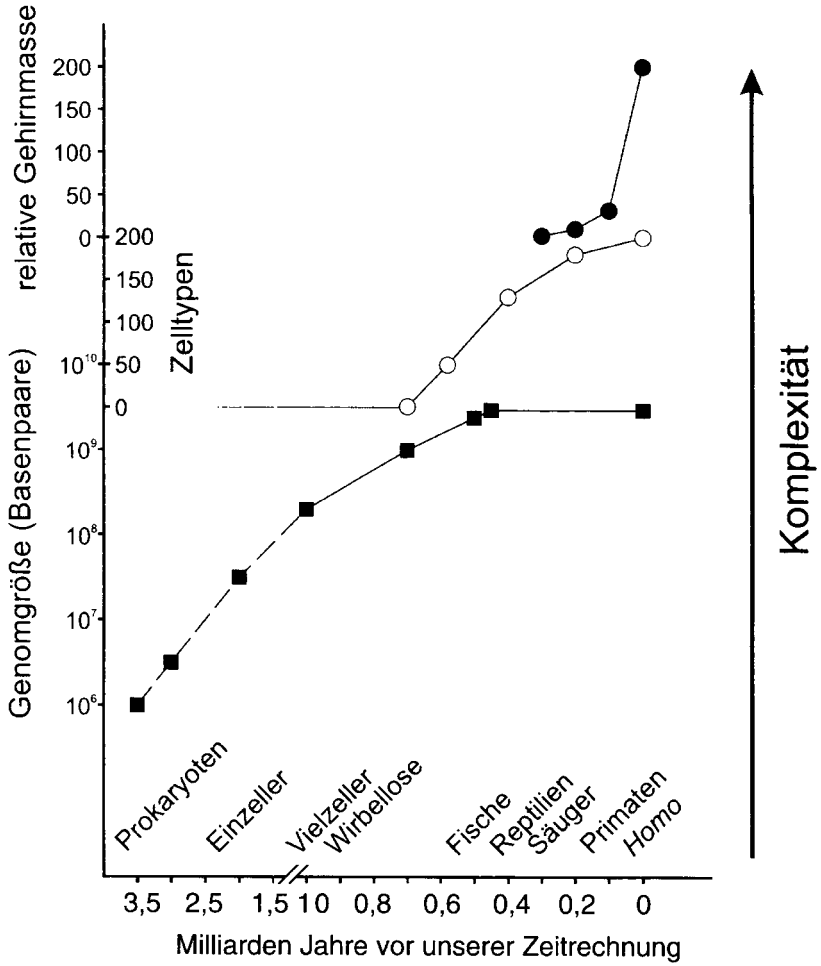
(iv) Klassenwechsel von Genen (tf ⇔ sp)

Statistik über eintausend Generationen

Gesamtzahl der Gene:	11.67 ± 2.69
Regulatorgene:	5.97 ± 2.22
Strukturgene:	5.70 ± 2.17

1. Der Darwinsche Mechanismus und seine universelle Anwendbarkeit
2. Ribonukleinsäuremoleküle als Phänotypen
3. Evolution im Reagenzglas und im Computer
4. Genetische und metabolische Netzwerke
5. Ein minimaler Metabolismus als Phänotyp
6. **Evolution und Komplexität**

E. coli:	Length of the Genome	4×10^6 Nucleotides
	Number of Cell Types	1
	Number of Genes	4 000
Man:	Length of the Genome	3×10^9 Nucleotides
	Number of Cell Types	200
	Number of Genes	40 000 - 60 000



4.10 Die Zunahme der Komplexität ist ein wesentlicher Aspekt der biologischen Evolution, wobei höhere Komplexität sowohl durch Vergrößerung der Zahl von miteinander in Wechselwirkung stehenden Elementen als auch durch Differenzierung der Funktionen dieser Elemente entstehen kann. In dieser Abbildung wird zwischen drei Phasen oder Strategien der Evolution von Komplexität unterschieden. *Untere Kurve*: Zunahme der Genomgröße; logarithmische Auftragung der Zahl der Basenpaare im Genom von Zellen seit Beginn der biologischen Evolution (Daten aus Abbildung 2.3). *Mittlere Kurve*: Zunahme der Zahl der Zelltypen in der Evolution der Metazoa (Daten aus Abbildung 4.8). *Obere Kurve*: Zunahme des relativen Gehirngewichts (bezogen auf die Körperoberfläche) bei Säugetieren (Daten aus Wilson 1985). Für die Abszisse wurden zwei Skaleneinteilungen verwendet, eine für den Zeitraum $>10^9$ Jahre, eine andere für den Zeitraum $<10^9$ Jahre vor der Gegenwart. Oberhalb der Abszisse sind die Namen einiger wichtiger taxonomischer Einheiten angeführt, deren Evolution in etwa beim jeweiligen Wortbeginn einsetzt.

Wolfgang Wieser. Die Erfindung der Individualität oder die zwei Gesichter der Evolution. Spektrum Akademischer Verlag, Heidelberg 1998.

A.C.Wilson. The Molecular Basis of Evolution. Scientific American, Oct.1985, 164-173.

Die großen Evolutionssprünge

1. der Übergang von einzelnen unabhängigen RNA-Genen zu einem Gesamtgenom,

2. die Entwicklung des genetischen Codes als Voraussetzung für den Übergang von einer RNA-Welt zu einer DNA(+RNA)+Protein-Welt,

3. die Ausbildung der Zelle mit Metabolismus und Kompartmentstruktur,

4. die Ausbildung der komplexen eukaryotischen Zelle durch einen Endosymbiose genannten Zusammenschluß von prokaryotischen Vorläufern,

5. die Ausbildung von Symbiosen zwischen Arten,

6. der Übergang vom Einzeller zum Vielzellerorganismus,

7. der Übergang von solitären Individuen zu Tiergesellschaften,

8. der Übergang von Tiergesellschaften zu primitiven menschlichen Gesellschaftsformen, und

9. die Entwicklung der heutigen menschlichen Gesellschaften mit Sprachen und Schriften.

Acknowledgement of support

Universität Wien

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, Austrian Academy of Sciences, and the
Santa Fe Institute

Coworkers

Universität Wien

Walter Fontana, Harvard Medical School, MA

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

Peter Stadler, **Bärbel Stadler**, Universität Leipzig, GE

Sebastian Bonhoeffer, ETH Zürich, **Erich Borberg-Bauer**, Universität Münster,
Martin Nowak, Harvard University, **Thomas Wiehe**, Universität Köln

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 Machne, 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