Mit dem Computer auf Entdeckungsreisen in der Evolution

Peter Schuster

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

Leopoldina, Jahresversammlung 2009

Halle (Saale), 02.– 04.10.2009

Web-Page für weitere Informationen

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

Mathematischen Modellierung hat Vor- und Nachteile:

1. Ein durch **korrekte Beweisführung** erhaltenes Resultat ist gültig und bedarf keiner weiteren Absicherung (etwa durch Wiederholung der Messung oder durch Verfeinerung der Methode).

2. Die Ergebnisse sind nur im Rahmen der Korrektheit der Modellannahmen und Interpretationen gültig (schwarz=korrekt, rot=falsch, braun=kann korrekt sein).

Analyse $\Rightarrow$ Vorhersage $\Rightarrow$ Interpretation $\Rightarrow$ Erklärung

Mathematischen Modellierung hat Vor- und Nachteile:

1. Ein durch **korrekte Beweisführung** erhaltenes Resultat ist gültig und bedarf keiner weiteren Absicherung (etwa durch Wiederholung der Messung oder durch Verfeinerung der Methode).

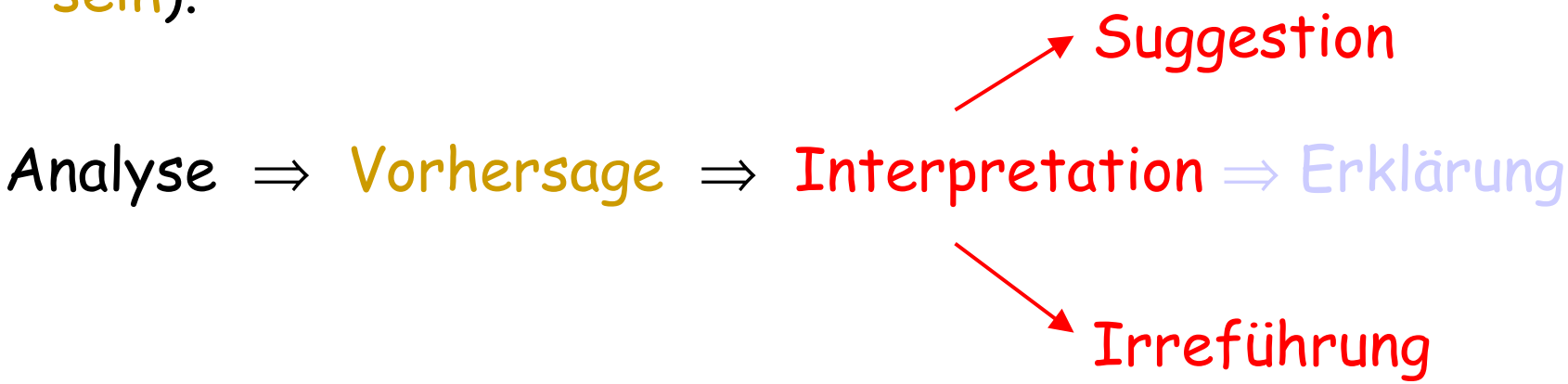
2. Die Ergebnisse sind nur im Rahmen der Korrektheit der Modellannahmen und Interpretationen gültig (schwarz=korrekt, rot=falsch, braun=kann korrekt sein).

Analyse $\Rightarrow$ Vorhersage $\Rightarrow$ Interpretation $\Rightarrow$ Erklärung

Mathematischen Modellierung hat Vor- und Nachteile:

1. Ein durch **korrekte Beweisführung** erhaltenes Resultat ist gültig und bedarf keiner weiteren Absicherung (etwa durch Wiederholung der Messung oder durch Verfeinerung der Methode).

2. Die Ergebnisse sind nur im Rahmen der Korrektheit der Modellannahmen und Interpretationen gültig (schwarz=korrekt, rot=falsch, braun=kann korrekt sein).



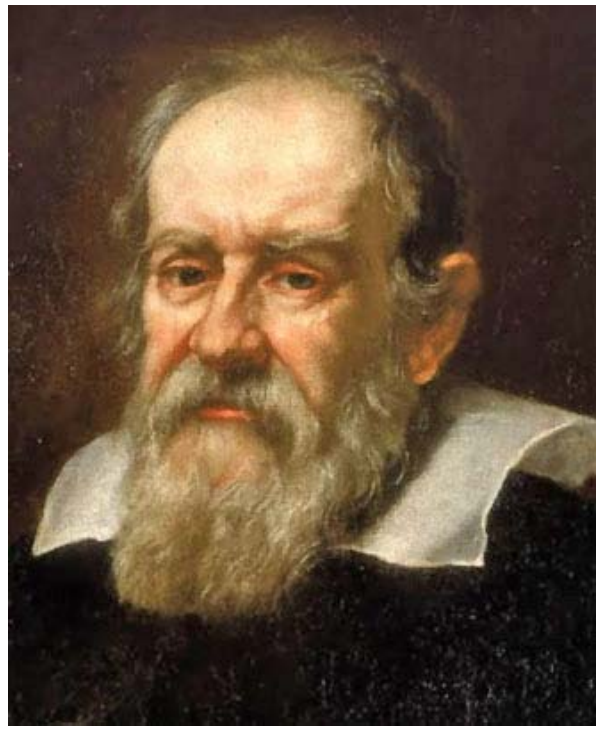
Computersimulation ist die Fortsetzung der

mathematischen Modellierung mit anderen Mitteln.

1. Mathematik und Physik
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'

1. **Mathematik und Physik**
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'

"La Filosophia è scritta in questo grandissimo libro, que continuamente ci stà aperto innanzi à gli occhi (io dico l'universo) ma non si può intendere se prima non s'impara à intender la lingua, e conoscer i caratteri, nei quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, e altre figure geometriche ...",

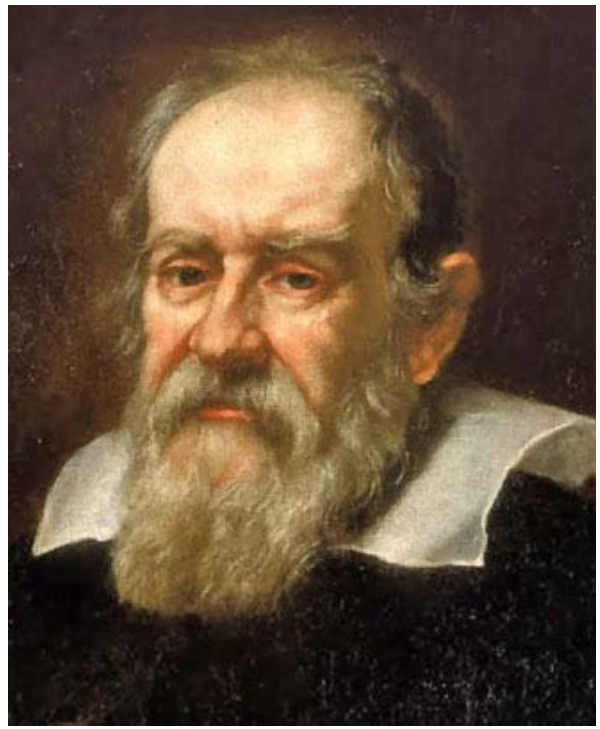


Galileo Galilei, 1564 - 1642

Galileo Galilei. 1632. ***Il Saggiatore***.
Edition Nationale, Bd.6, Florenz 1896, p.232.

"**Die Wissenschaft** [Filosophia] **ist in dem großartigen Buch geschrieben**, das ständig vor unseren Augen geöffnet ist (ich nenne es das Universum), aber man kann es nicht verstehen, wenn man nicht vorher seine Sprache erlernt, seine Zeichen, in denen es geschrieben ist. **Es ist in der Sprache der Mathematik geschrieben, und die Zeichen sind Dreiecke, Kreise und anderen geometrischen Figuren** …. „

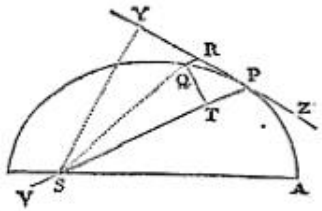
Galileo Galilei. 1632. ***Il Saggiatore***.
Edition Nationale, Bd.6, Florenz 1896, p.232.



Galileo Galilei, 1564 - 1642

De Motu Corporum.

Corol. 4. Iisdem positis, est vis centripeta ut velocitas bis directe, & chorda illa inverse. Nam velocitas est reciproce ut perpendiculum ST per corol. 1. prop. 1.

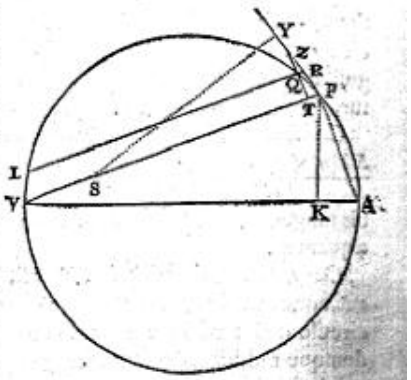


Corol. 5. Hinc si detur figura quævis curvilinea APQ, & in ea detur etiam punctum S, ad quod vis centripeta perpetuo dirigitur, inveniri potest lex vis centripetæ, qua corpus quodvis P a cursu rectilineo perpetuo retractum in figuræ illius perimetro detinebitur, eamque revolvendo describet. Nimirum computandum est vel solidum $\frac{SPq \times QTq}{QR}$ vel solidum $STq \times PV$ huic vi reciproce proportionale. Ejus rei dabimus exempla in problematis sequentibus.

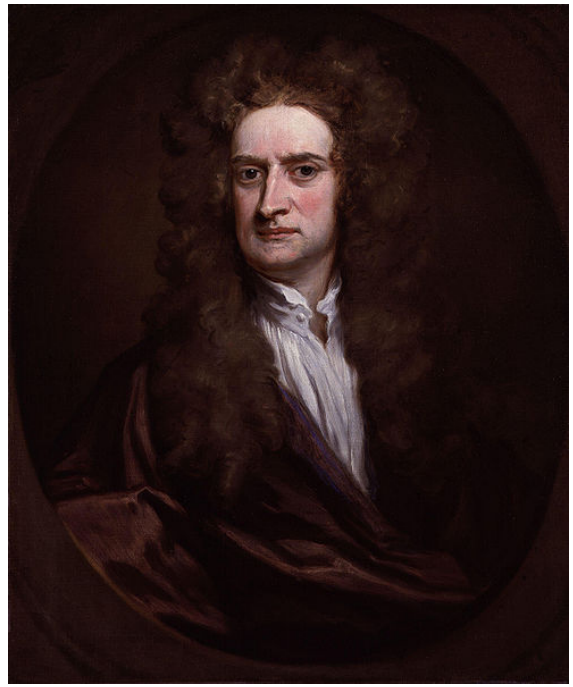
PROPOSITIO VII. PROBLEMA II.

Gyretur corpus in circumferentia circuli, requiritur lex vis centripetæ tendentis ad punctum quodcunque datum.

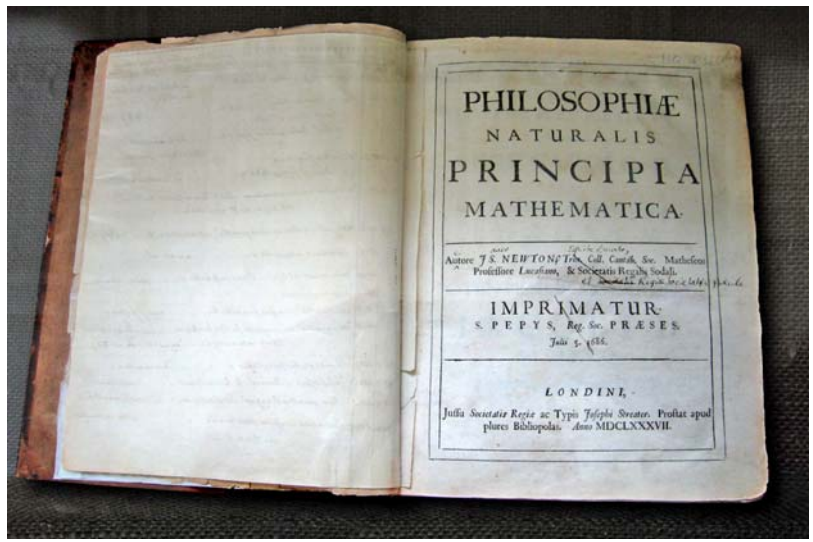
Esto circuli circumferentia $VQPA$; punctum datum, ad quod vis ceu ad centrum sum tendit, S; corpus in circumferentia latum P; locus proximus, in quem movebitur Q; & circuli tangens ad locum priorem PRZ. Per punctum S ducatur chorda PV; & acta circuli diametro VA, jungatur AP; & ad SP demittatur perpendiculum QT, quod productum occurrat tangenti PR in Z; ac denique per punctum Q agatur LR, quæ ipsi SP parallella sit, & occurrat tum circulo in L, tum tangenti PZ in R. Et ob similia triangula ZQR, ZTP, VPA; erit RP *quad.* hoc est QRL ad QT *quad.*



Isaac Newton. 1686. ***Principia mathematica***.



Isaac Newton, 1643 - 1727



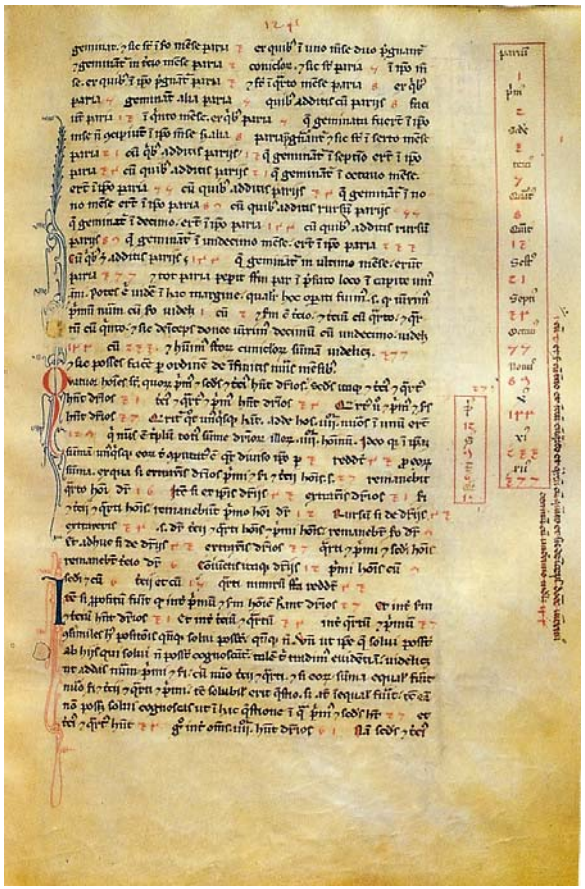
1. Mathematik und Physik
2. **Mathematik in der Biologie**
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'

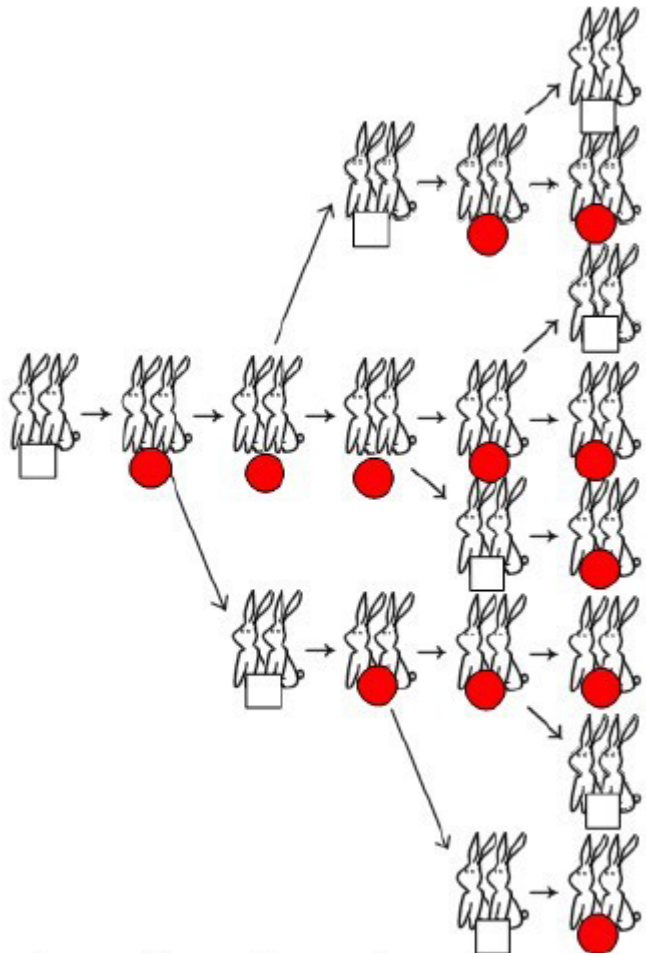
$$F_n := \begin{cases} 0 & \text{if } n = 0\,; \\ 1 & \text{if } n = 1\,; \\ F_{n-1} + F_{n-2} & \text{if } n > 1\,. \end{cases}$$

$$F_n = 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots, \text{ for}$$
$$n = 0, 1, 2, 3, 4, 5, 6, \ 7, \ 8, \ 9, 10, \ldots\ .$$

Leonardo da Pisa
„Fibonacci“ – Filius Bonacci
~1180 – ~1240

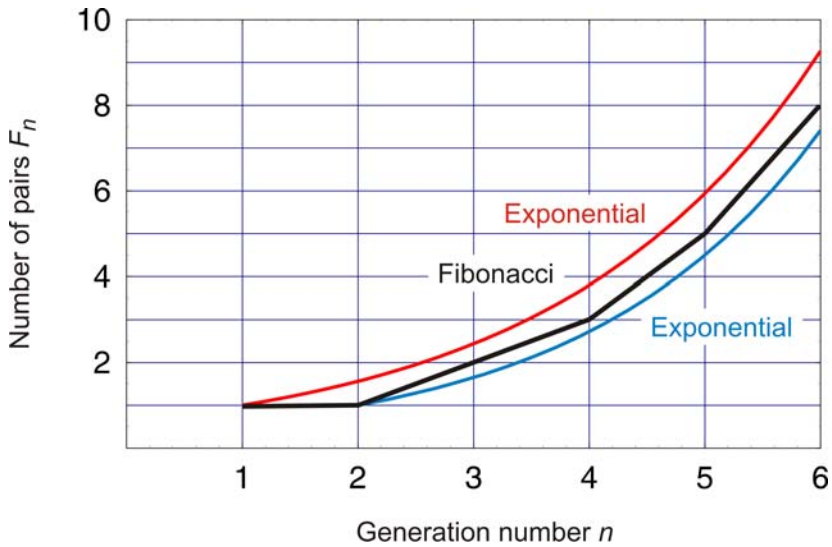
Die Fibonacci Reihe

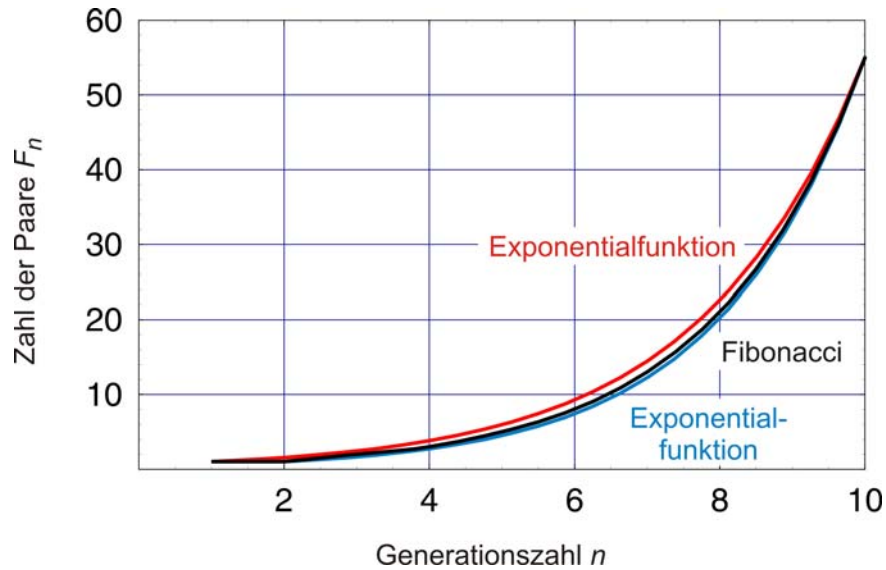




Die Fibonacci Reihe

Bodo Werner, Universität Hamburg, 2006





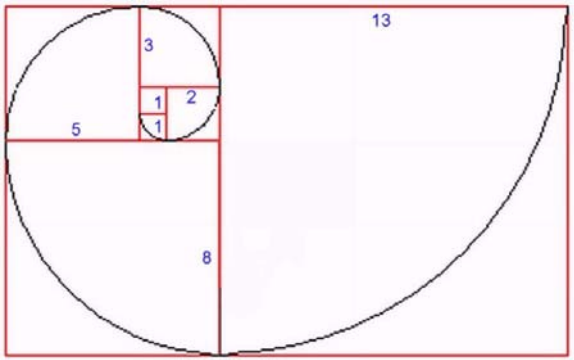
$$f_{\text{upper}}(n) = \exp\left(0.445259 \cdot (n-1)\right)$$

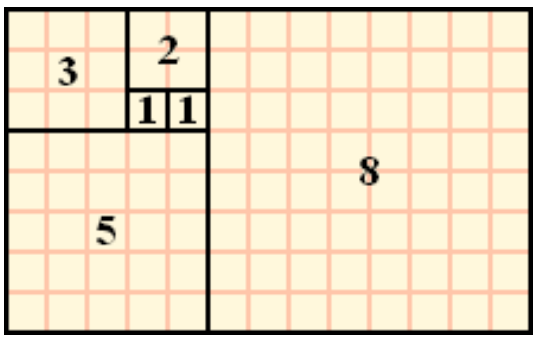
$$f_{\text{lower}}(n) = \exp\left(0.500917 \cdot (n-2)\right)$$

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \frac{1}{2}\left(1+\sqrt{5}\right) = 1.61803\ \ldots$$

Johannes Kepler (1571-1630)

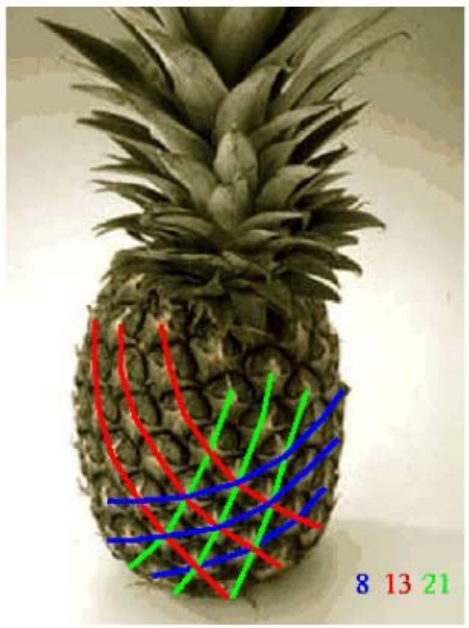
Die Fibonacci Reihe

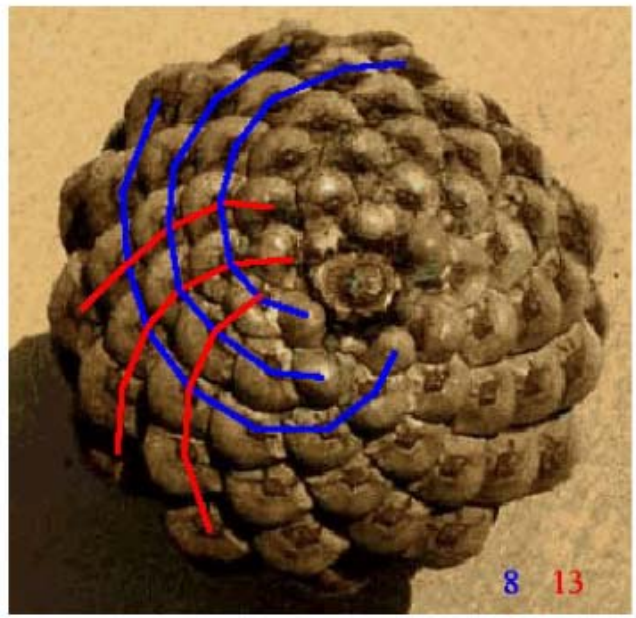




Raum erfüllende Quadrate

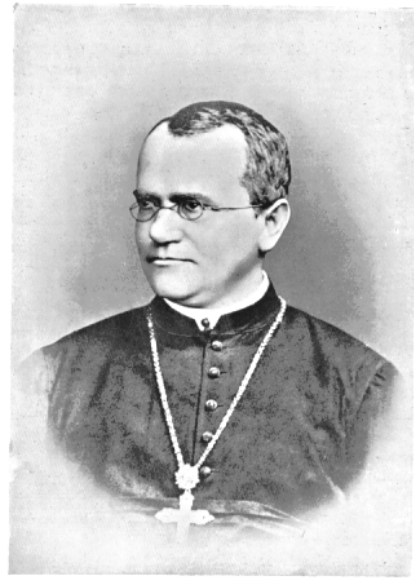
Die Fibonacci Spiralen





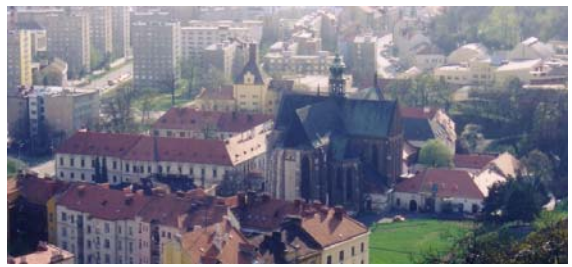
Gregor Mendels Merkmale an Erbsen:

1. Blütenfarbe purpurrot oder weiß,
2. Blüten am Stamm oder endständig,
3. Stamm kurz oder lang,
4. Samen rund oder runzelig,
5. Samenfarbe gelb oder grün,
6. Schoten voll oder eingeschnürt,
7. Schotenfarbe gelb oder grün.



Gregor Mendel (1882-1884)

```
1st experiment  ⇒  60 fertilizations on 15 plants
2nd experiment  ⇒  58 fertilizations on 10 plants
3rd experiment  ⇒  35 fertilizations on 10 plants
4th experiment  ⇒  40 fertilizations on 10 plants
5th experiment  ⇒  23 fertilizations on  5 plants
6th experiment  ⇒  34 fertilizations on 10 plants
7th experiment  ⇒  37 fertilizations on 10 plants
```



Gregor Mendels Experimente zur Genetik der Pflanzen

Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn* **4**: 3–47, 1866.
Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des naturforschenden Vereines in Brünn* **8**: 26–31, 1870.

	Experiment 1		**Experiment 2**	
	Form of Seed		Color of Albumen	
Plants	Round	Angular	Yellow	Green
1	45	12	25	11
2	27	8	32	7
3	24	7	14	5
4	19	10	70	27
5	32	11	24	13
6	26	6	20	6
7	88	24	32	13
8	22	10	44	9
9	28	6	50	14
10	25	7	44	18

- Expt. 1: Form of seed. From 253 hybrids 7324 seeds were obtained in the second trial year. Among them were 5474 round or roundish ones and 1850 angular wrinkled ones. Therefrom the ratio **2.96:1** is deduced.
- Expt. 2: Color of albumen.. 258 plants yielded 8023 seeds, 6022 yellow, and 2001 green; their ratio, therefore, is as **3.01:1**.

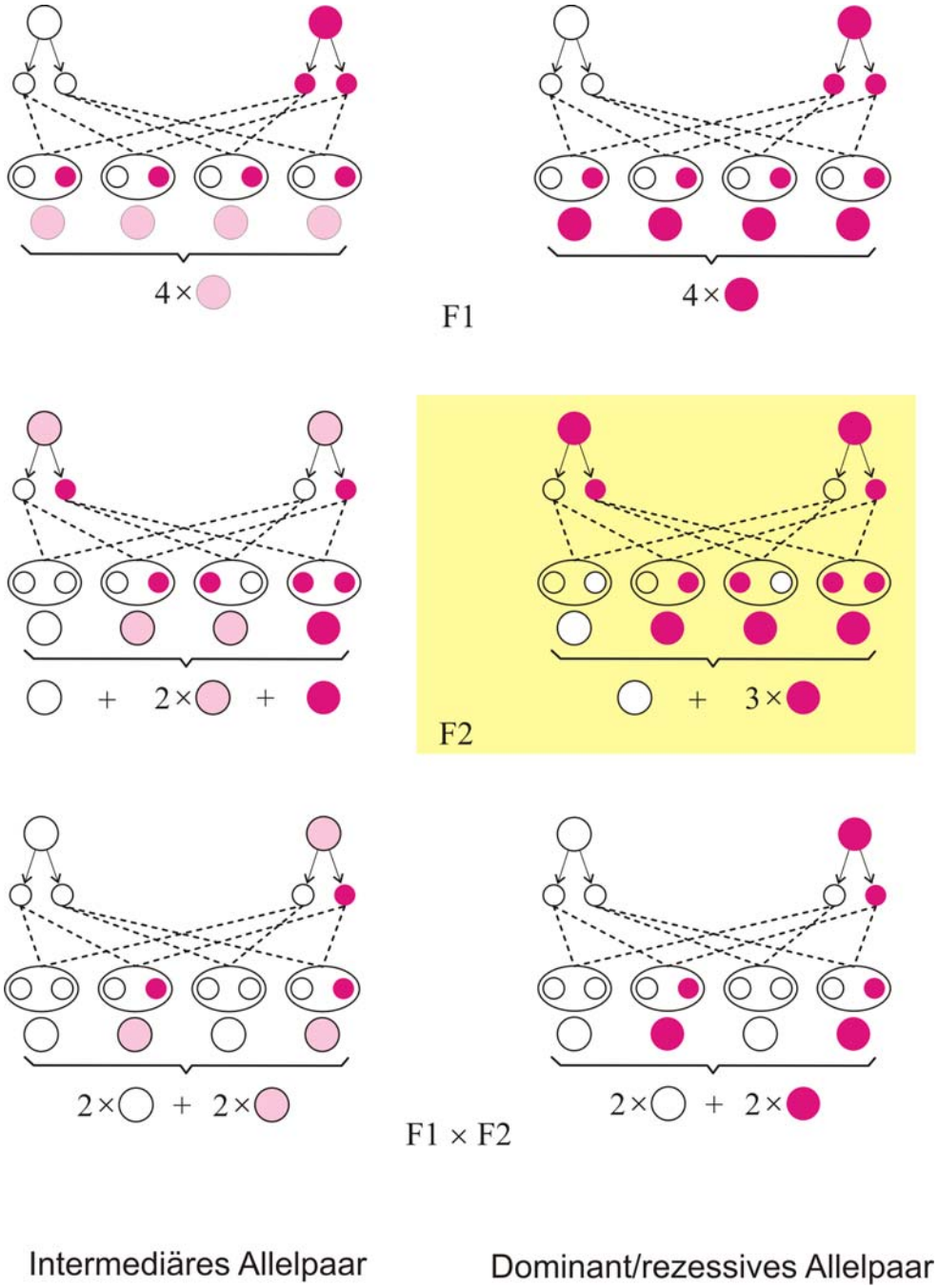
Gregor Mendels Experimente zur Genetik der Pflanzen

Gregor Mendel zog aus seinen Experimenten drei richtige Schlüsse:

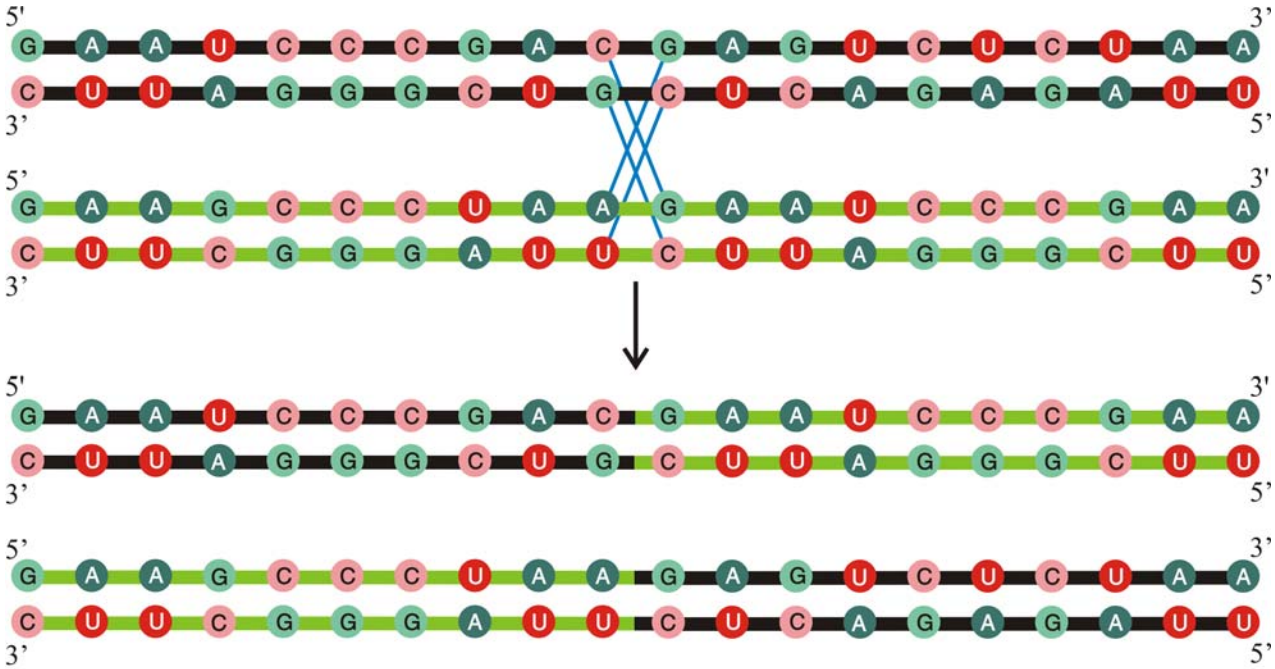
1. Die Vererbung jedes Merkmals wird durch „Elemente“ oder „Faktoren“ bestimmt, welche unverändert an die Nachkommen weitergegeben werden (Diese Faktoren nennen wir heute Gene).

2. Ein Nachkomme erbt je ein solches Element von jedem Elternteil für jedes Merkmal.

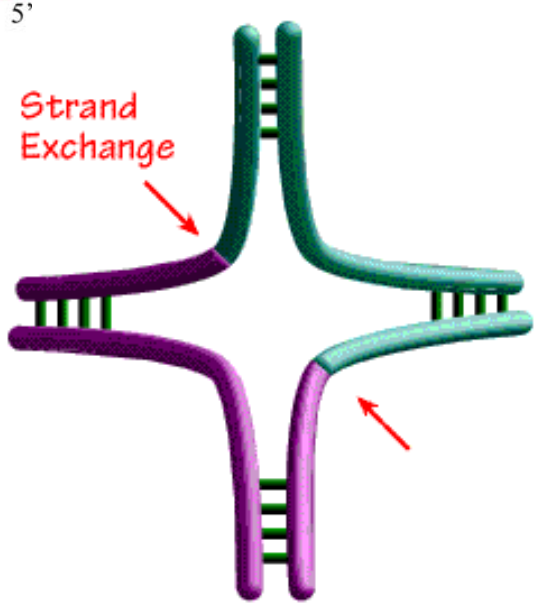
3. Ein Merkmal kann bei einem Individuum unsichtbar sein, aber dessen ungeachtet an die nächste Generation weitergegeben werden (Wir bezeichnen das heute als ein rezessives Merkmal).



Gregor Mendels Experimente
zur Genetik der Pflanzen



genetic recombination



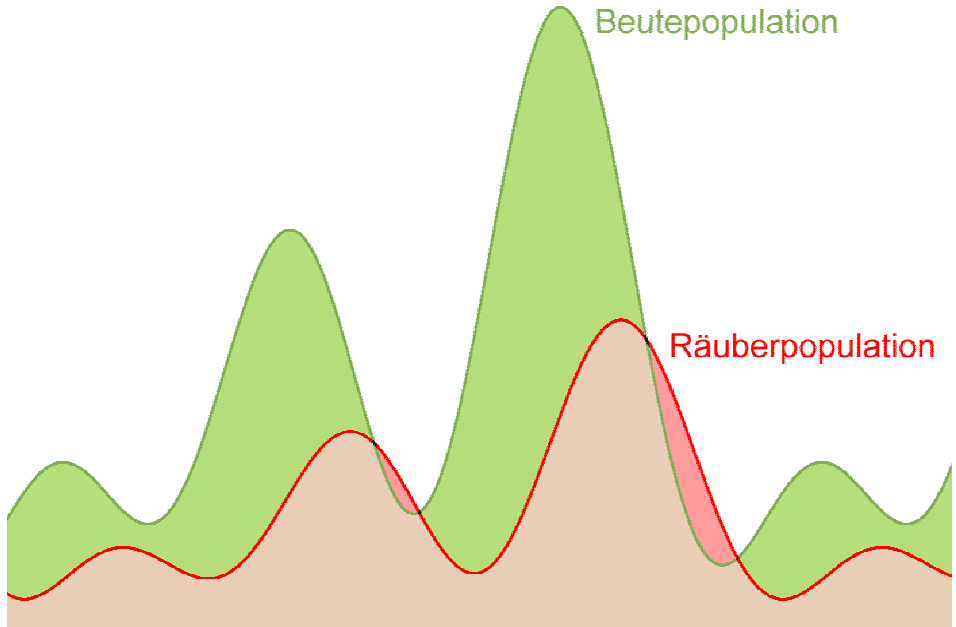
Molekulare Erklärung von Mendels Exprimenten – Rekombination

Alfred J. Lotka, 1880 - 1949

Vito Volterra, 1860 - 1940

$$\frac{dn_1}{dt} = \varepsilon_1 n_1 - \gamma_1 n_1 n_2 \qquad \text{Hasen}$$

$$\frac{dn_2}{dt} = -\varepsilon_2 n_2 + \gamma_2 n_1 n_2 \qquad \text{Füchse}$$



Räuber-Beute Beziehungen nach Lotka und Volterra

Ronald Fisher (1890-1962)

Allele: $A_1, A_2, \ldots, A_n$

Häufigkeiten: $x_i = [A_i]$; Genotypen: $A_i \cdot A_j$

Fitnesswerte: $a_{ij} = f(A_i \cdot A_j),\ a_{ij} = a_{ji}$

Mendel

Darwin

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} a_{ji}\, x_i\, x_j - \Phi\, x_j = x_j \left(\sum_{i=1}^{n} a_{ji}\, x_i - \Phi \right), \quad j = 1, 2, \ldots, n$$

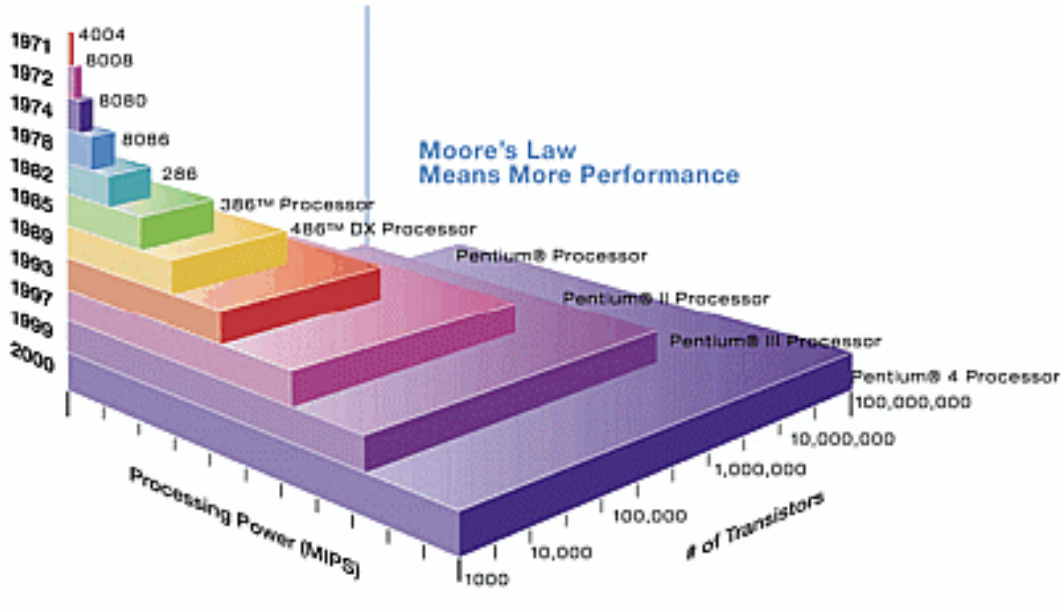
$$\text{mit} \quad \Phi(t) = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji}\, x_i\, x_j \quad \text{und} \quad \sum_{j=1}^{n} x_j = 1$$

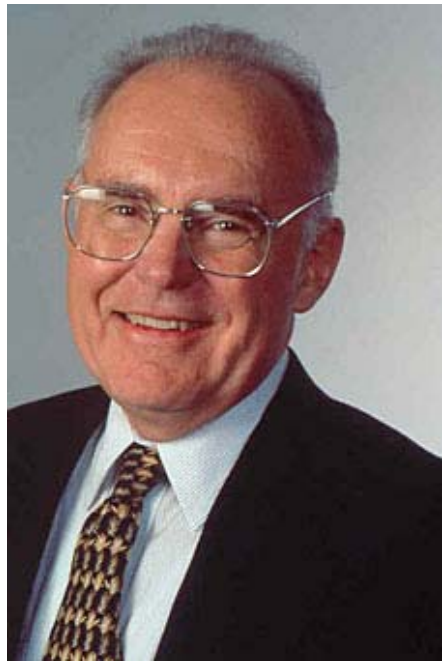
$$\frac{d\Phi}{dt} = 2\left(< \bar{a}^2 > - < \bar{a} >^2 \right) = 2 \text{ var } \{\bar{a}\} \geq 0$$

Ronald Fishers Selektionsgleichung: The genetical theory of natural selection. Oxford, UK, Clarendon Press, 1930.

1. Mathematik und Physik
2. Mathematik in der Biologie
3. **Das Zeitalter des Computers**
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'

CPU Transistor Counts 1971-2008 & Moore's Law



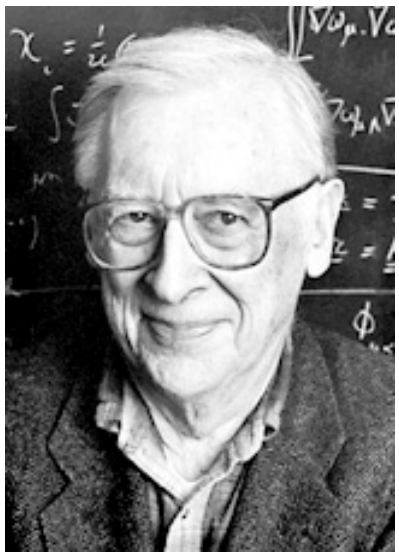


Gordon Moore, 1929 -

Steigerung des Leistungsfähigkeit der Computer

Computational Chemistry: 1960 – heute

1. Näherungsverfahren zur Berechnung der Elektronenstrukturen von Molekülen
2. *Ab initio*-Berechnung von Elektronenstrukturen (Hartree-Fock und CI)
3. Dichtefunktionaltheorie
4. Berechnung der molekularen Strukturen und anderer Moleküleigenschaften



John A. Pople, 1925 - 2004

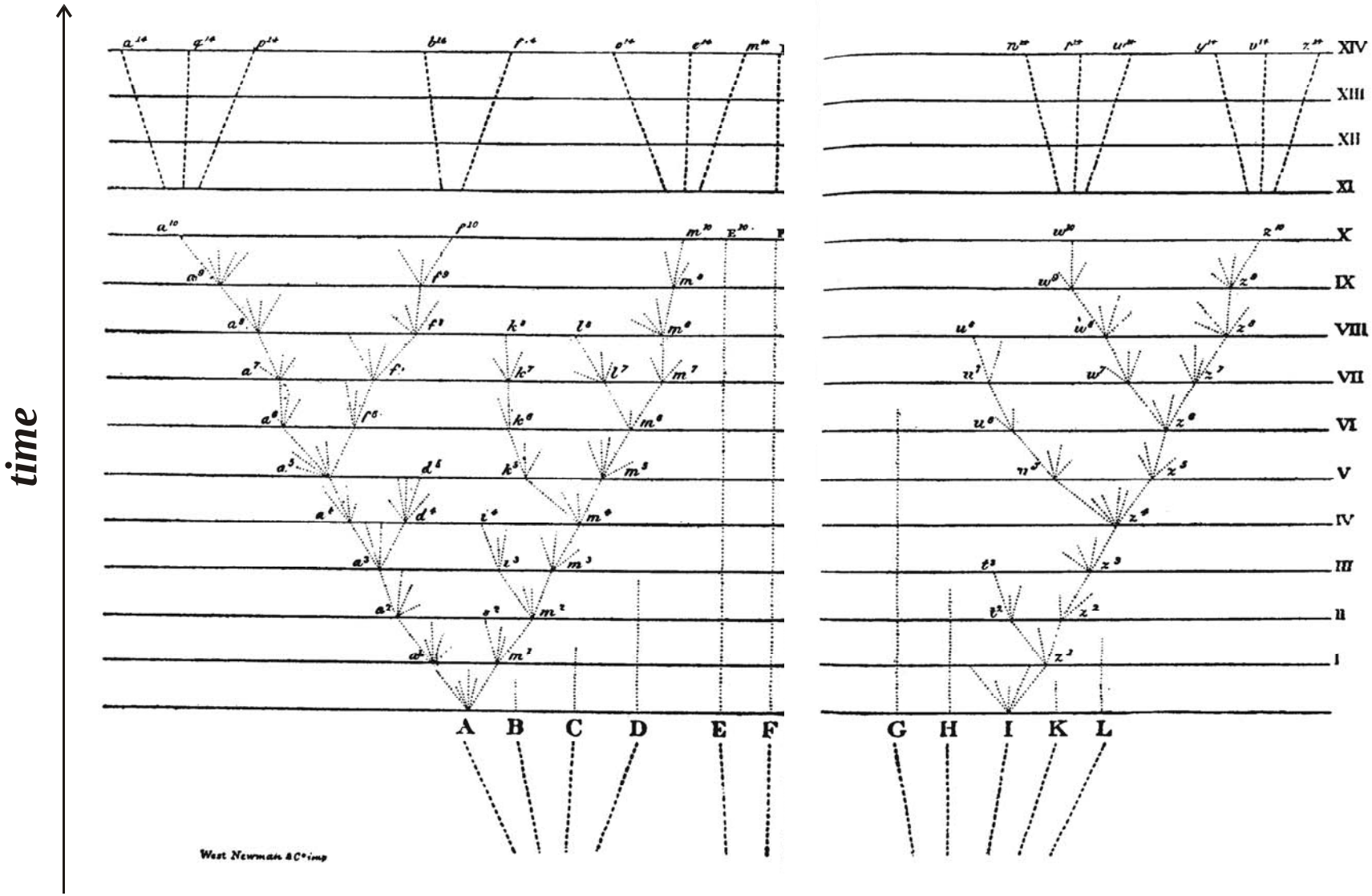
Nobelpreis für Chemie
1998

Walter Kohn, 1923 -

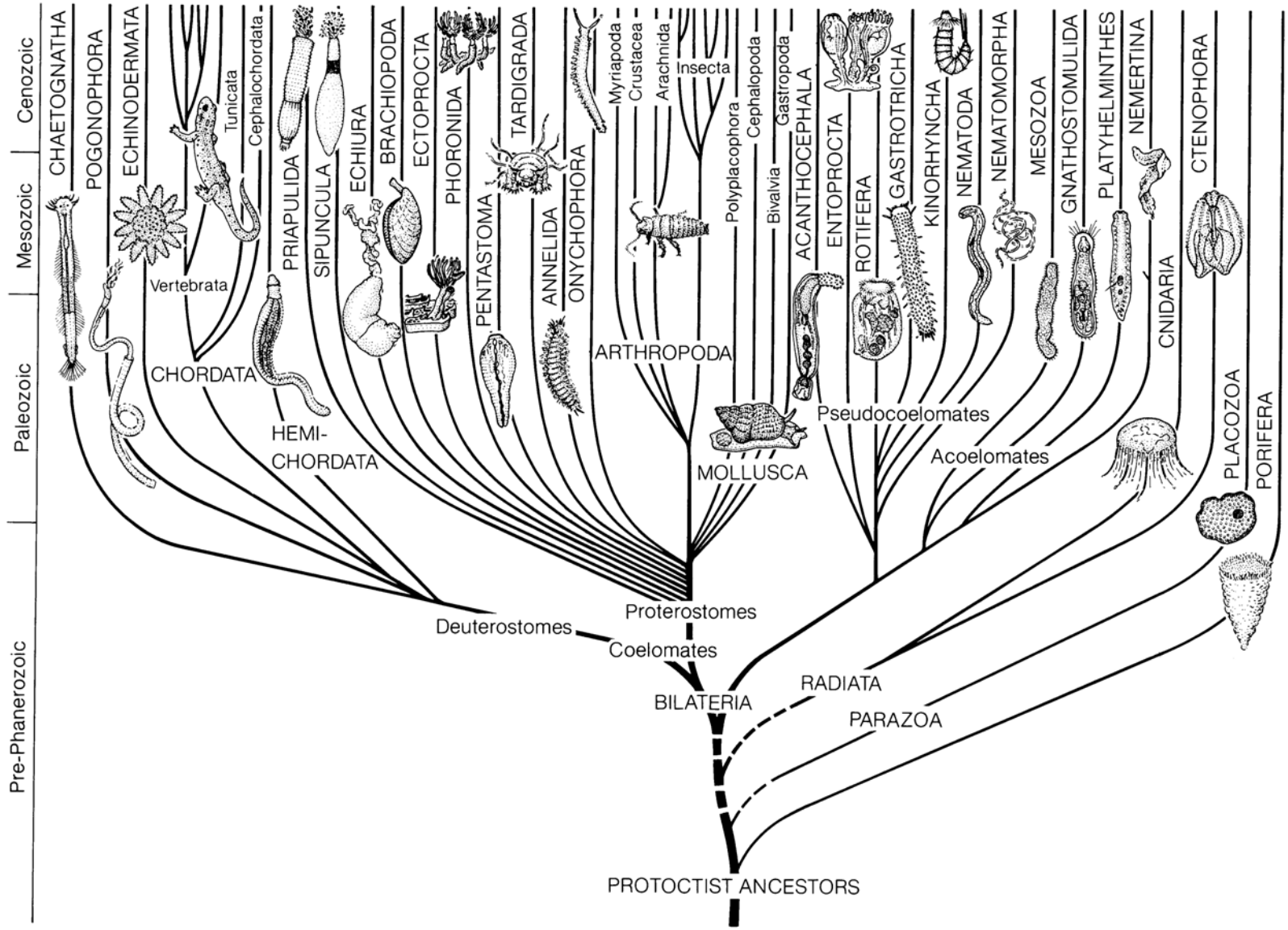
Computational Structural Biology: 1970 – heute

1. Vorhersage der Sekundärstrukturen von Proteinen
2. Vorhersage der Sekundärstrukturen von einsträngigen RNA-Molekülen
3. Berechnung von Protein 3D-Strukturen
4. Modellierung von RNA 3D-Strukturen
5. Homologie-Modellierung von Proteinstrukturen
6. Molekulardynamik von Proteinen und Nukleinsäuren
7. Ab initio Proteinstrukturrechnungen
8. Quantenchemische Rechnungen an Modellverbindungen

1. Mathematik und Physik
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. **Bioinformatik und Systembiologie**
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'



Charles Darwin, *The Origin of Species*, 6th edition.
Everyman‘s Library, Vol.811, Dent London, pp.121-122.



Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms*. *An Illustrated Guide to the Phyla of Life on Earth*. W.H. Freeman, San Francisco, 1982.

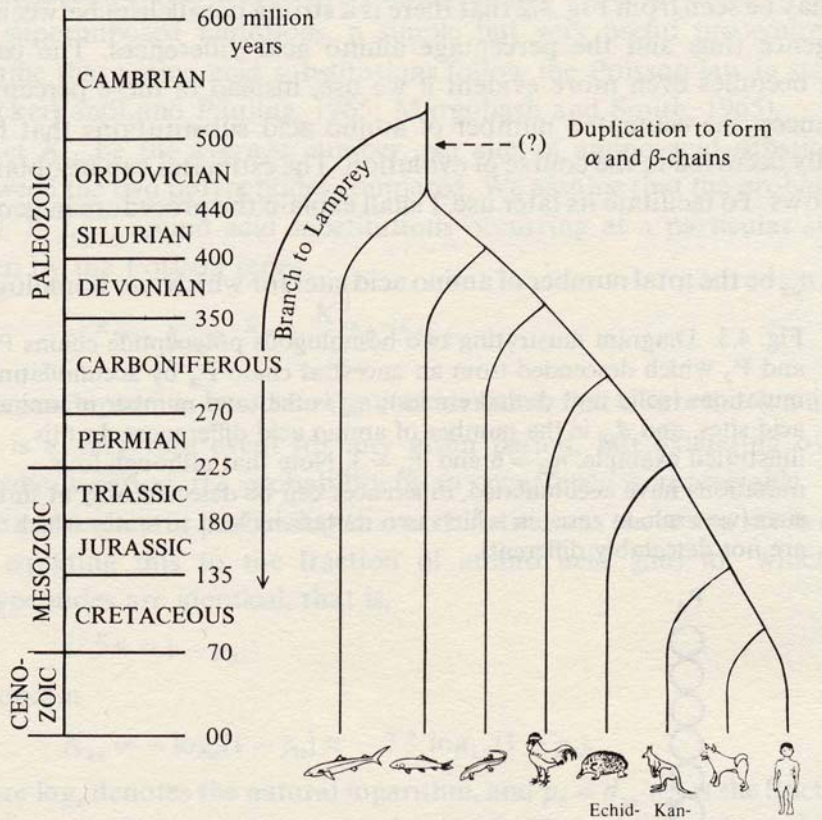
Motoo Kimura, 1924 - 1994

The molecular clock of evolution

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

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

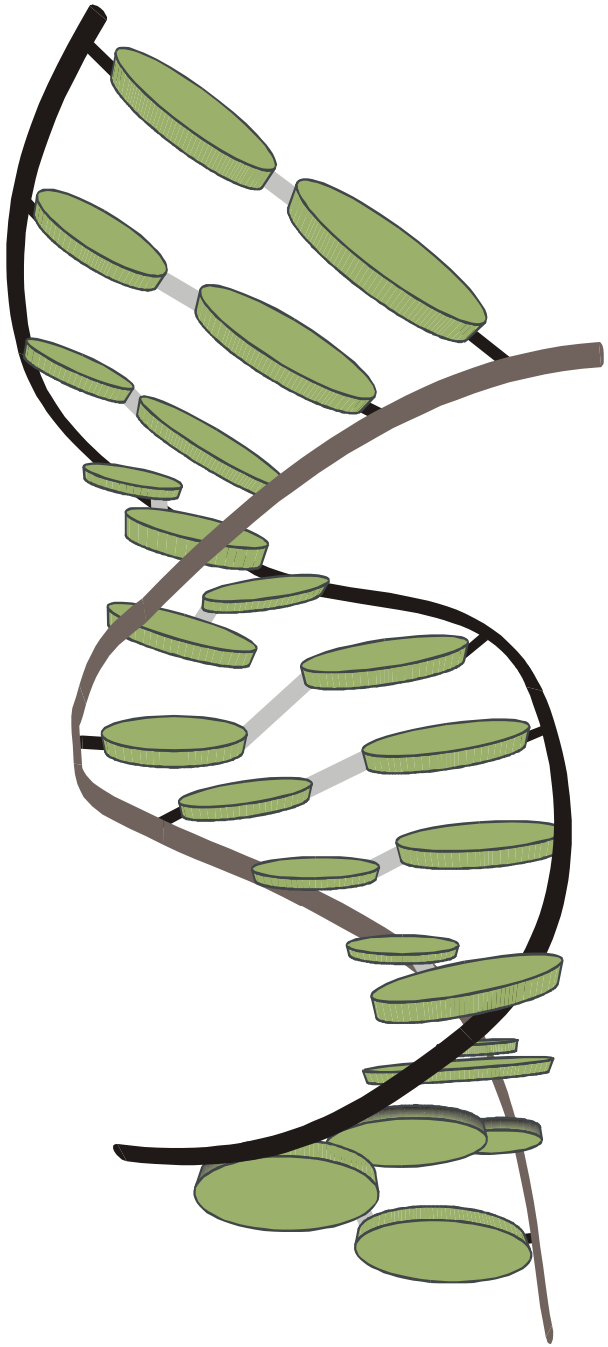
Fig. 4.2. Percentage amino acid differences when the α hemoglobin chains are compared among eight vertebrates together with their phylogenetic relationship and the times of divergence.

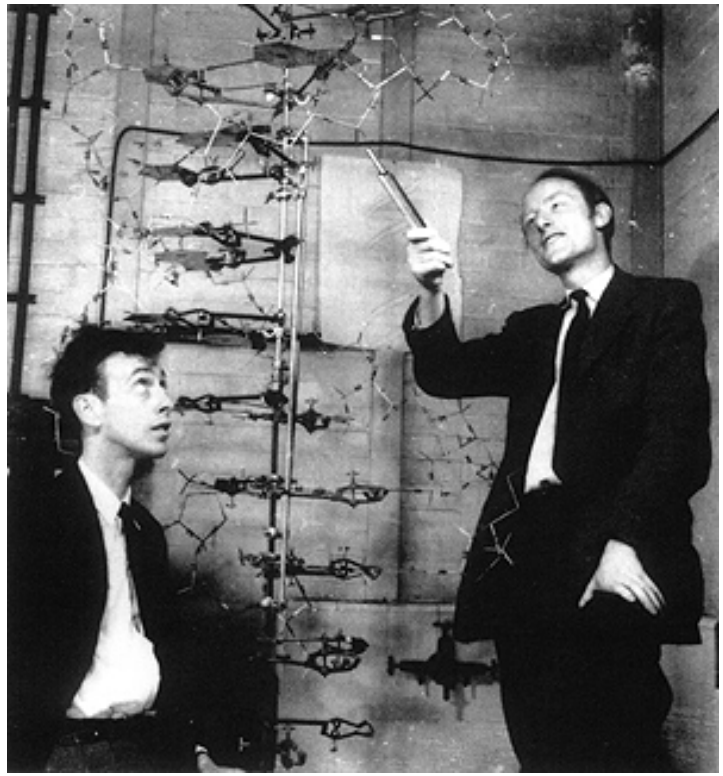


	Shark	Carp	Newt	Chicken	Echidna	Kangaroo	Dog	Human
Shark		59.4	61.4	59.7	60.4	55.4	56.8	53.2
Carp			53.2	51.4	53.6	50.7	47.9	48.6
Newt				44.7	50.4	47.5	46.1	44.0
Chicken					34.0	29.1	31.2	24.8
Echidna						34.8	29.8	26.2
Kangaroo							23.4	19.1
Dog								16.3
Human								

α-chain

Percent Amino Acid Differences



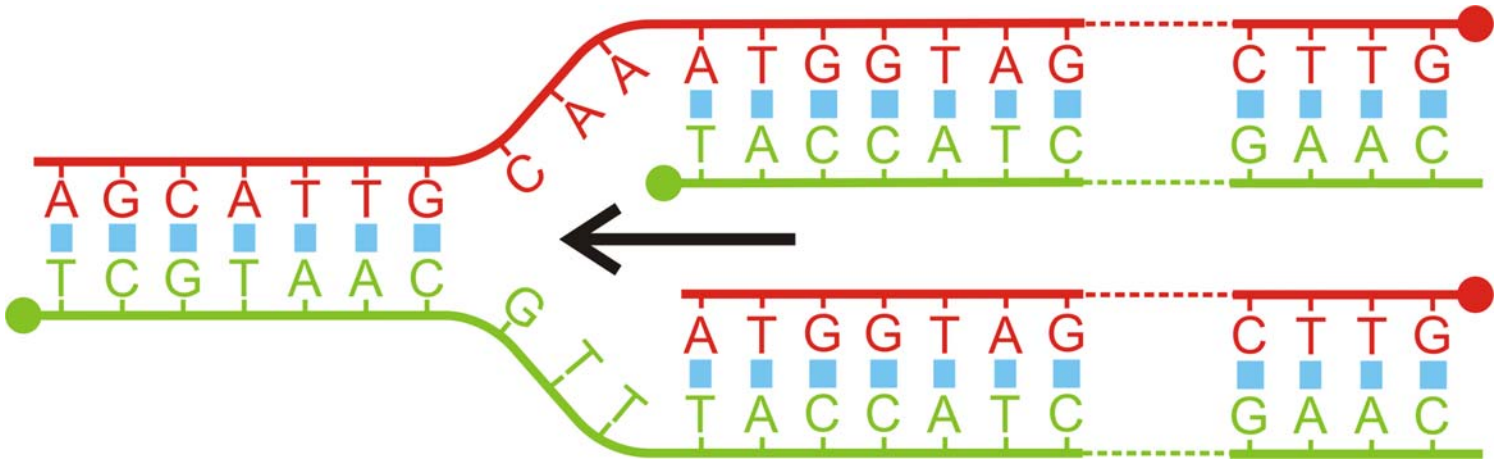


James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004

Nobel prize 1962

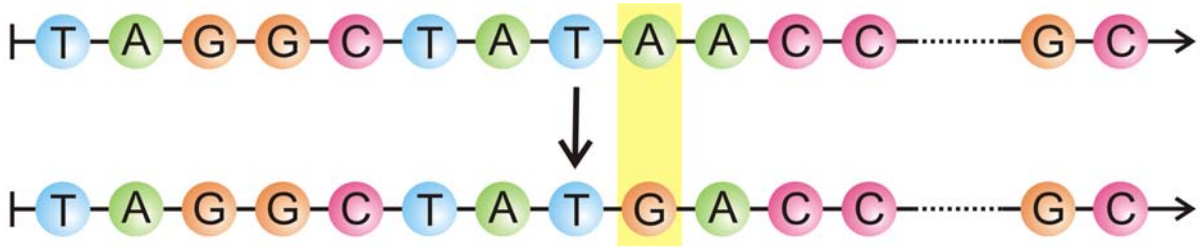
1953 – 2003 fifty years double helix

The three-dimensional structure of a short double helical stack of B-DNA

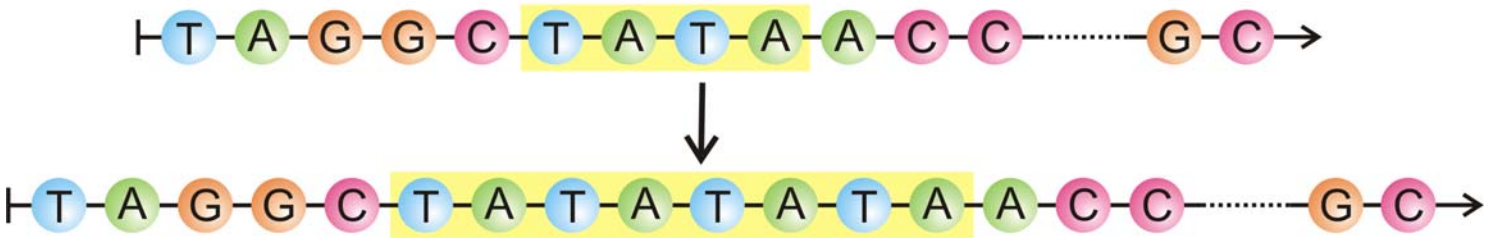


‚Replikationsgabel' bei der DNA-Verdoppelung

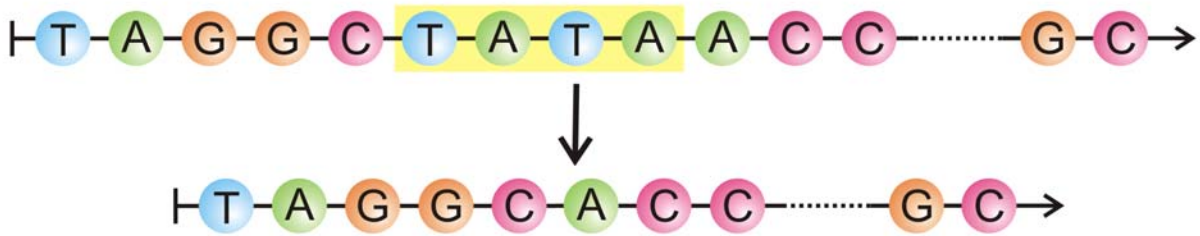
Der Mechanismus der DNA-Replication ist ‚semi-konservativ '



Point mutation

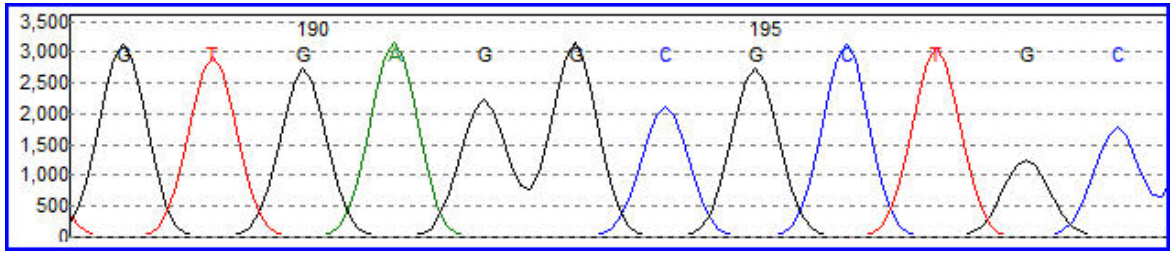


Insertion



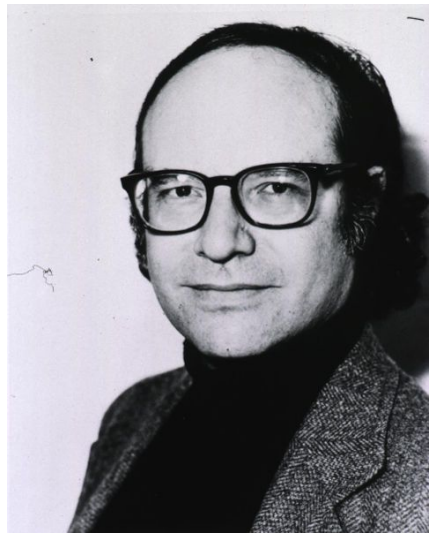
Deletion

Rekonstruktion phylogenetischer Bäume durch den Vergleich von Sequenzdaten



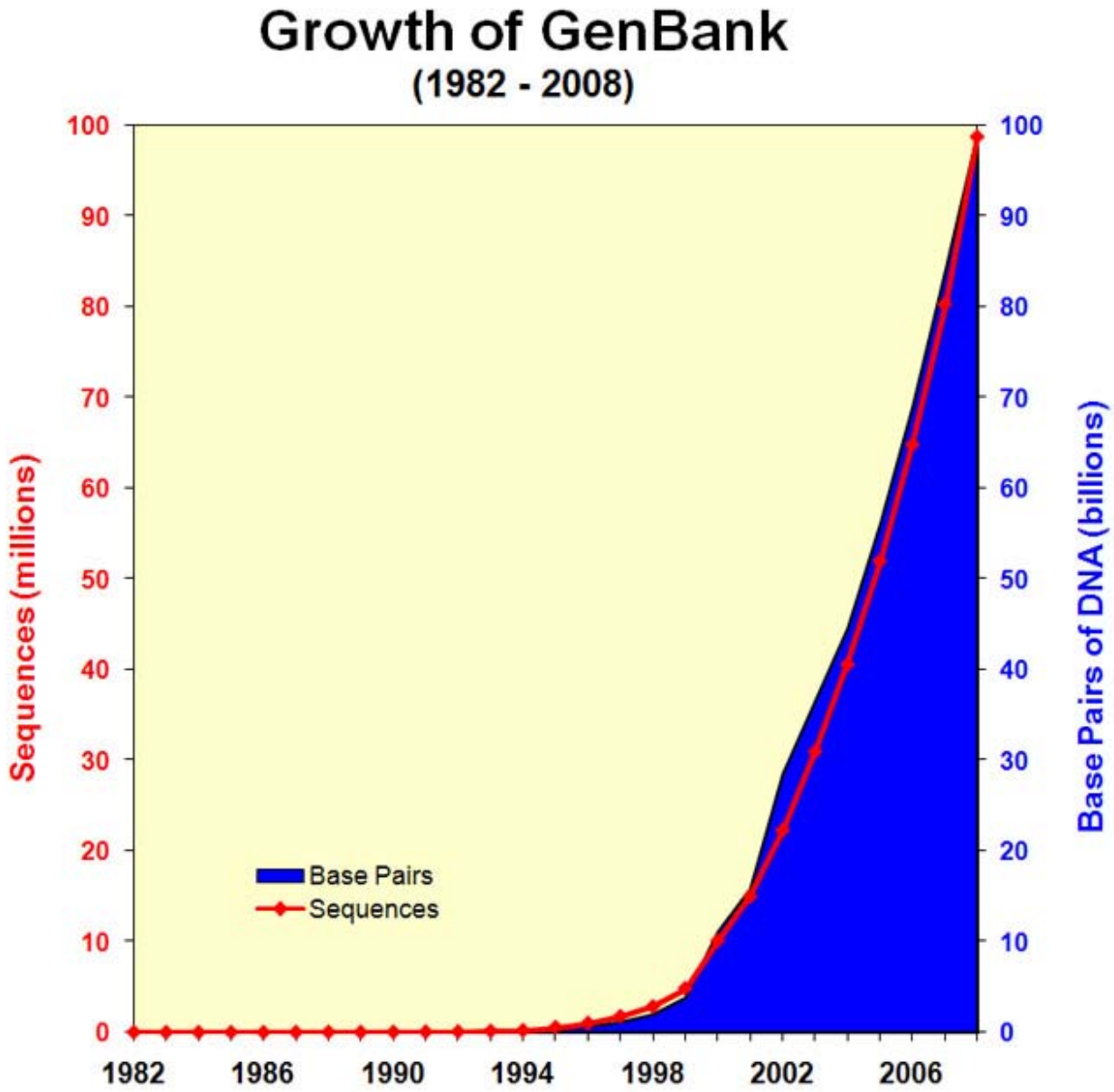
Frederick Sanger, 1918 -

Nobelpreis für Chemie
1980



Walter Gilbert, 1932 -

Neue DNA Sequenzierungstechniken: Sanger, 1977 (enzymatische Synthese) und Maxam & Gilbert, 1977 (chemischer Abbau)



Die atemberaubende Zunahme der Leistung der DNA-Sequenzierung nach der Einführung der neuen Techniken

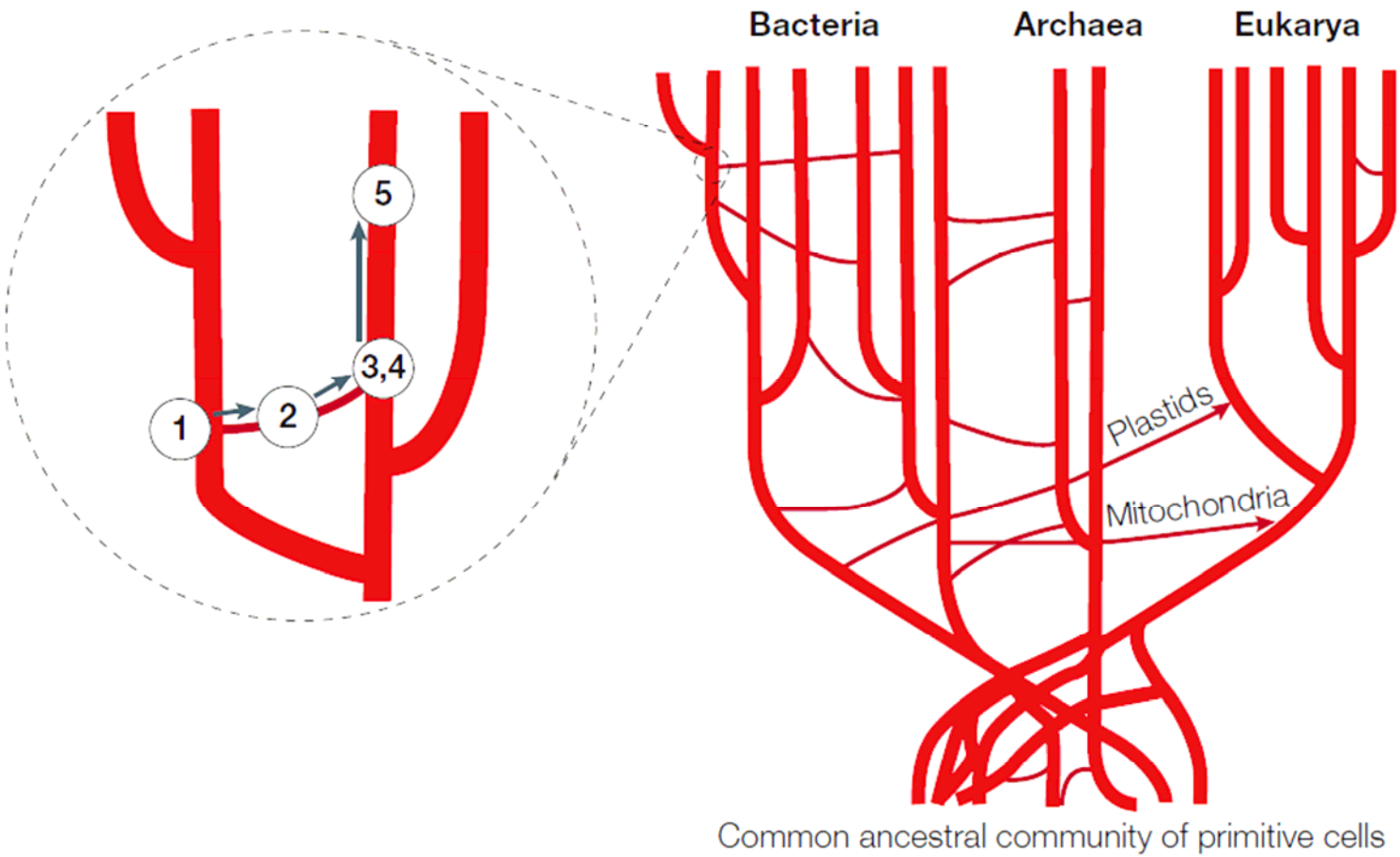


Figure 1 | **The 5 steps of horizontal gene flow.** Horizontal gene transfer and how it has impacted the evolution of life is presented through a web connecting bifurcating branches that complicate, yet do not erase, the tree of life. The inset illustrates the continuum of 5 steps that leads to the stable inheritance of a transferred gene in a new host.

Horizontal gene transfer

B.F. Smets, T. Barkay. 2005. Horizontal gene transfer: Perspectives at a crossroads of scientific disciplines. Nature Reviews Microbiology 3:675-678.

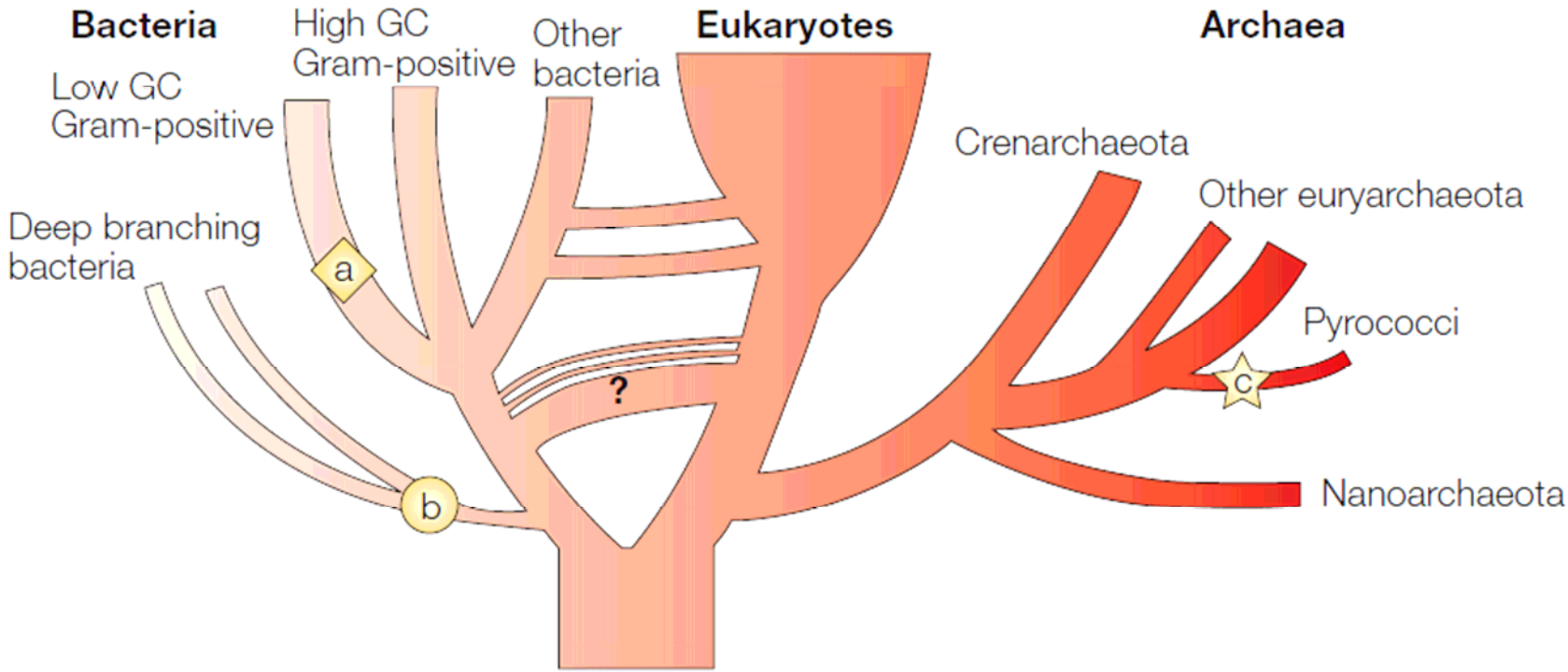
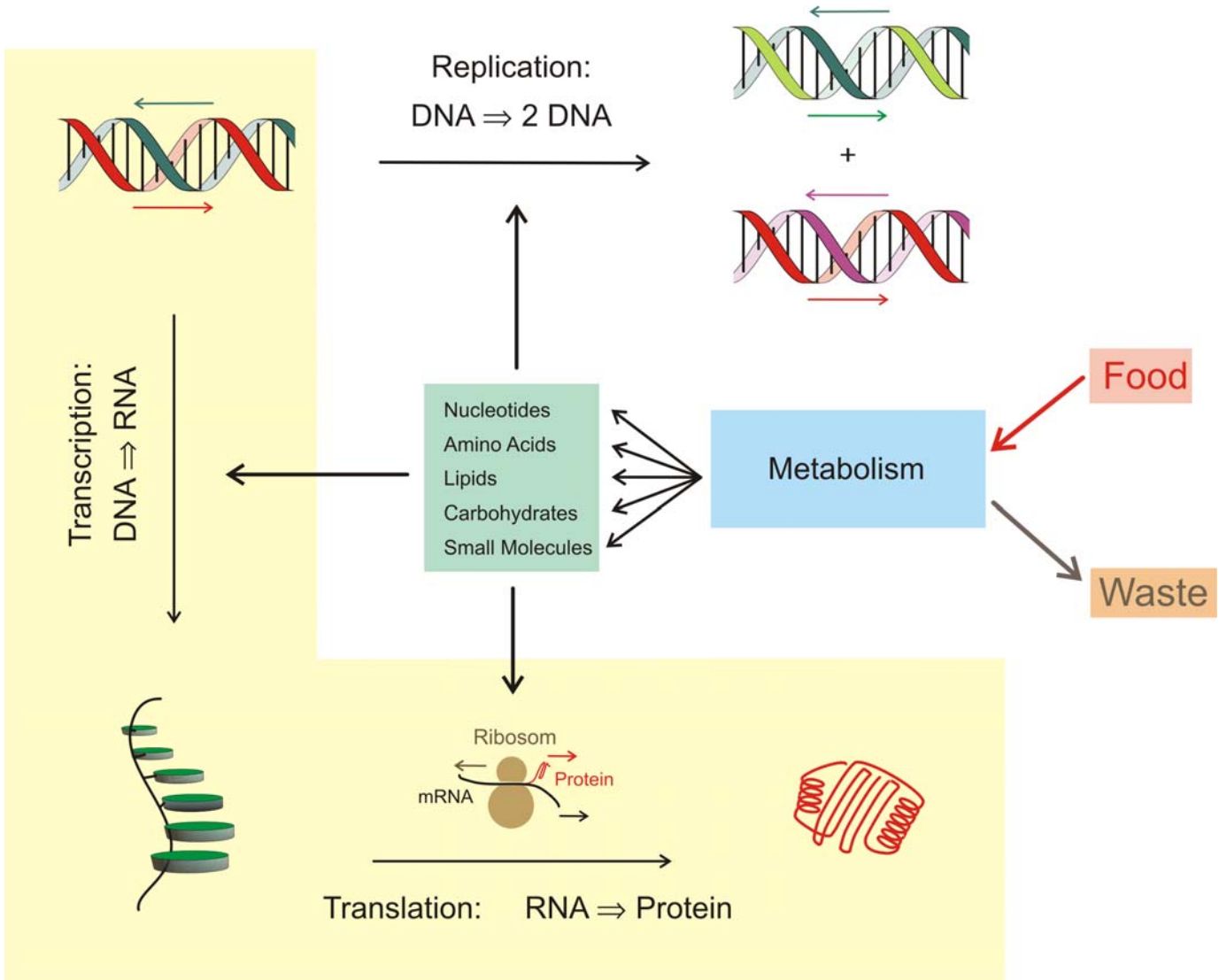


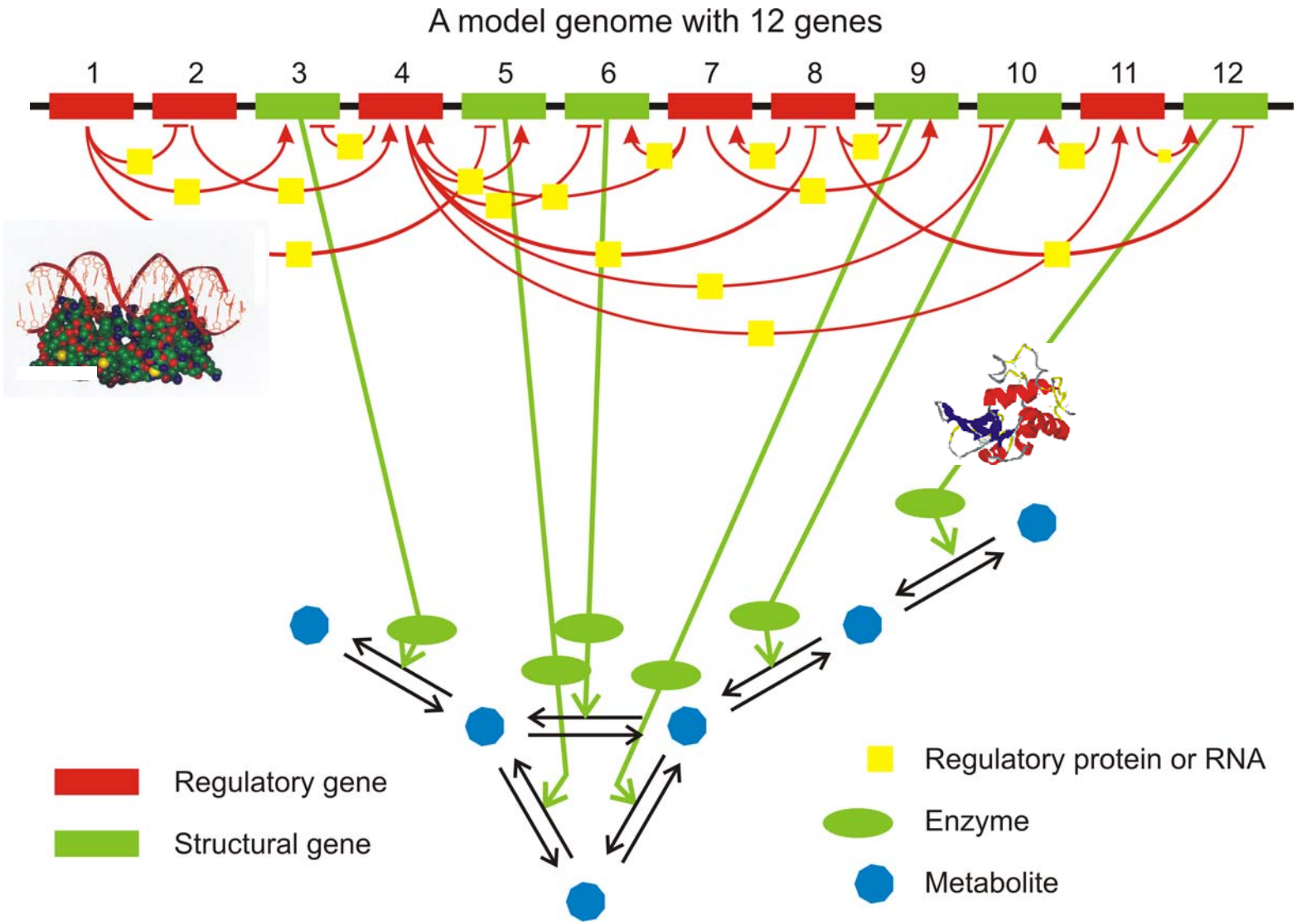
Figure 2 | **The tree of life.** A sketch of the tree of life as it is frequently derived from genome data (for example, REF. 26), with the three possible positions of *Thermotoga maritima* marked according to (**a**) 'concordant' genes (placed with the Gram-positives), (**b**) 16S RNA (and other conserved genes) and whole-genome analyses (placed as an early diverging lineage) and (**c**) phylogenetically discordant genes (placed with the Pyrococci among the Archaea). For further discussion see REF. 28 and text.

Horizontal gene transfer

J.P. Gogarten, J.P. Townsend. 2005. Horizontal gene transfer, genome innovation, and evolution. Nature Reviews Microbiology 3:679-687.



Skizze des zellulären Stoffwechsels



Skizze eines genetischen und metabolischen Netzwerks

Vollständige mathematische Analyse von kleinen genetischen Netzwerken

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 246 (2007) 395–419
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Dynamic patterns of gene regulation I: Simple two-gene systems

Stefanie Widder[a], Josef Schicho[b], Peter Schuster[a,c,*]

[a]*Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
[b]*RICAM—Johann Radon Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*
[c]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 24 February 2006; received in revised form 7 January 2007; accepted 8 January 2007
Available online 16 January 2007

Abstract

Regulation of gene activities is studied by means of computer assisted mathematical analysis of ordinary differential equations (ODEs) derived from binding equilibria and chemical reaction kinetics. Here, we present results on cross-regulation of two genes through activator and/or repressor binding. Arbitrary (differentiable) binding function can be used but systematic investigations are presented for gene–regulator complexes with integer valued Hill coefficients up to $n = 4$. The dynamics of gene regulation is derived from bifurcation patterns of the underlying systems of kinetic ODEs. In particular, we present analytical expressions for the parameter values at which one-dimensional (transcritical, saddle-node or pitchfork) and/or two-dimensional (Hopf) bifurcations occur. A classification of regulatory states is introduced, which makes use of the sign of a 'regulatory determinant' D (being the determinant of the block in the Jacobian matrix that contains the derivatives of the regulator binding functions): (i) systems with $D<0$, observed, for example, if both proteins are activators or repressors, to give rise to one-dimensional bifurcations only and lead to bistability for $n\geqslant 2$ and (ii) systems with $D>0$, found for combinations of activation and repression, sustain a Hopf bifurcation and undamped oscillations for $n>2$. The influence of basal transcription activity on the bifurcation patterns is described. Binding of multiple subunits can lead to richer dynamics than pure activation or repression states if intermediates between the unbound state and the fully saturated DNA initiate transcription. Then, the regulatory determinant D can adopt both signs, plus and minus.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Basal transcription; Bifurcation analysis; Cooperative binding; Gene regulation; Hill coefficient; Hopf bifurcation

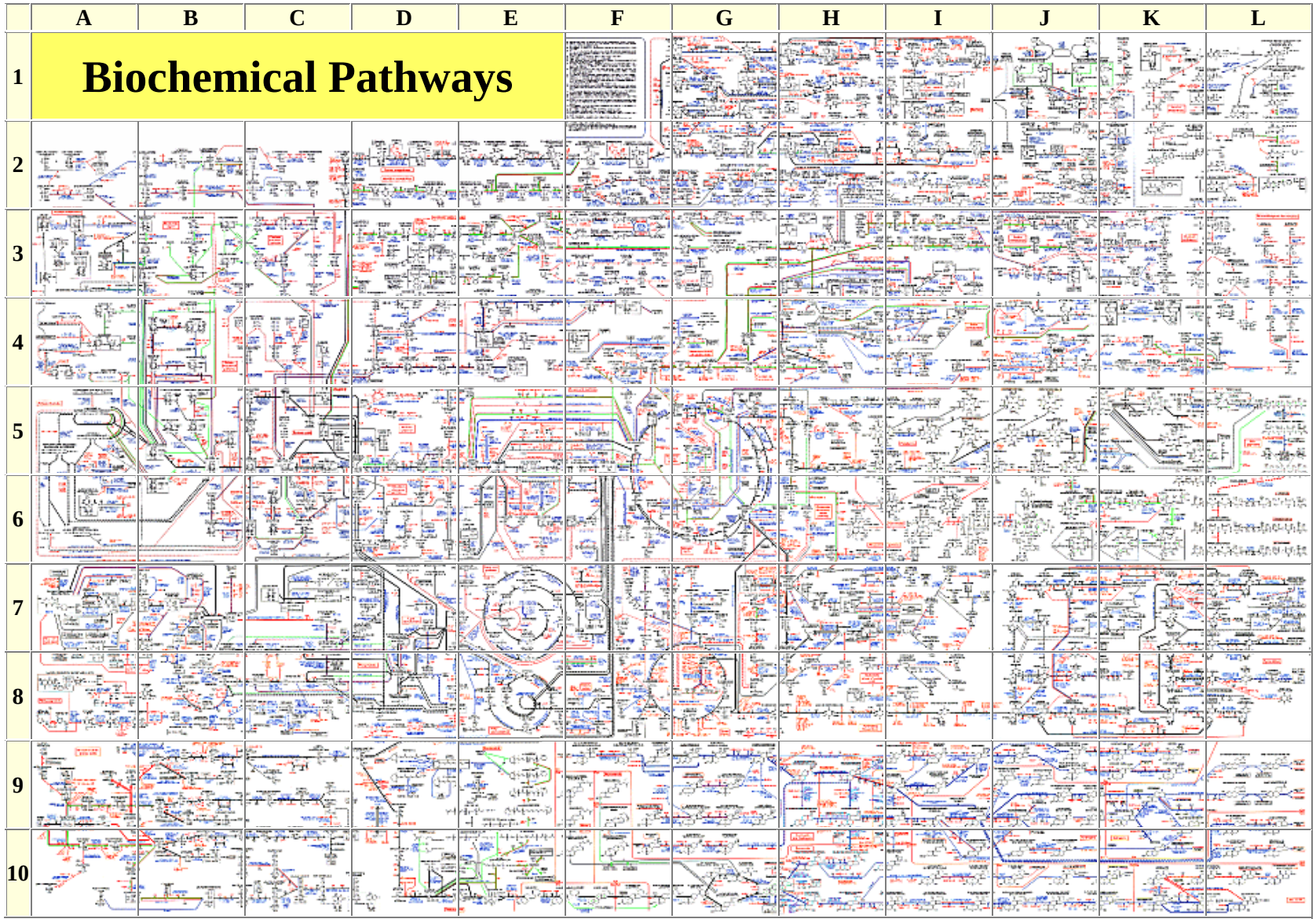
1. Introduction

Theoretical work on gene regulation goes back to the 1960s (Monod et al., 1963) soon after the first repressor protein had been discovered (Jacob and Monod, 1961). A little later the first paper on oscillatory states in gene regulation was published (Goodwin, 1965). The interest in gene regulation and its mathematical analysis never ceased (Tiwari et al., 1974; Tyson and Othmer, 1978; Smith, 1987) and saw a great variety of different attempts to design models of genetic regulatory networks that can be used in systems biology for computer simulation of *gen*(etic and met)*abolic* networks.[1] Most models in the literature aim at a minimalist dynamic description which, nevertheless, tries to account for the basic regulatory functions of large networks in the cell in order to provide a better understanding of cellular dynamics. A classic in general regulatory dynamics is the monograph by Thomas and D'Ari (1990). The currently used mathematical methods comprise application of Boolean logic (Thomas and Kaufman, 2001b; Savageau, 2001; Albert and Othmer, 2003), stochastic processes (Hume, 2000) and deterministic dynamic models, examples are Cherry and Adler (2000), Bindschadler and Sneyd (2001) and Kobayashi et al. (2003) and the recent elegant analysis of bistability (Craciun et al.,

*Corresponding author. Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.
Tel.: +43 1 4277 52743; fax: +43 1 4277 52793.
E-mail address: pks@tbi.univie.ac.at (P. Schuster).

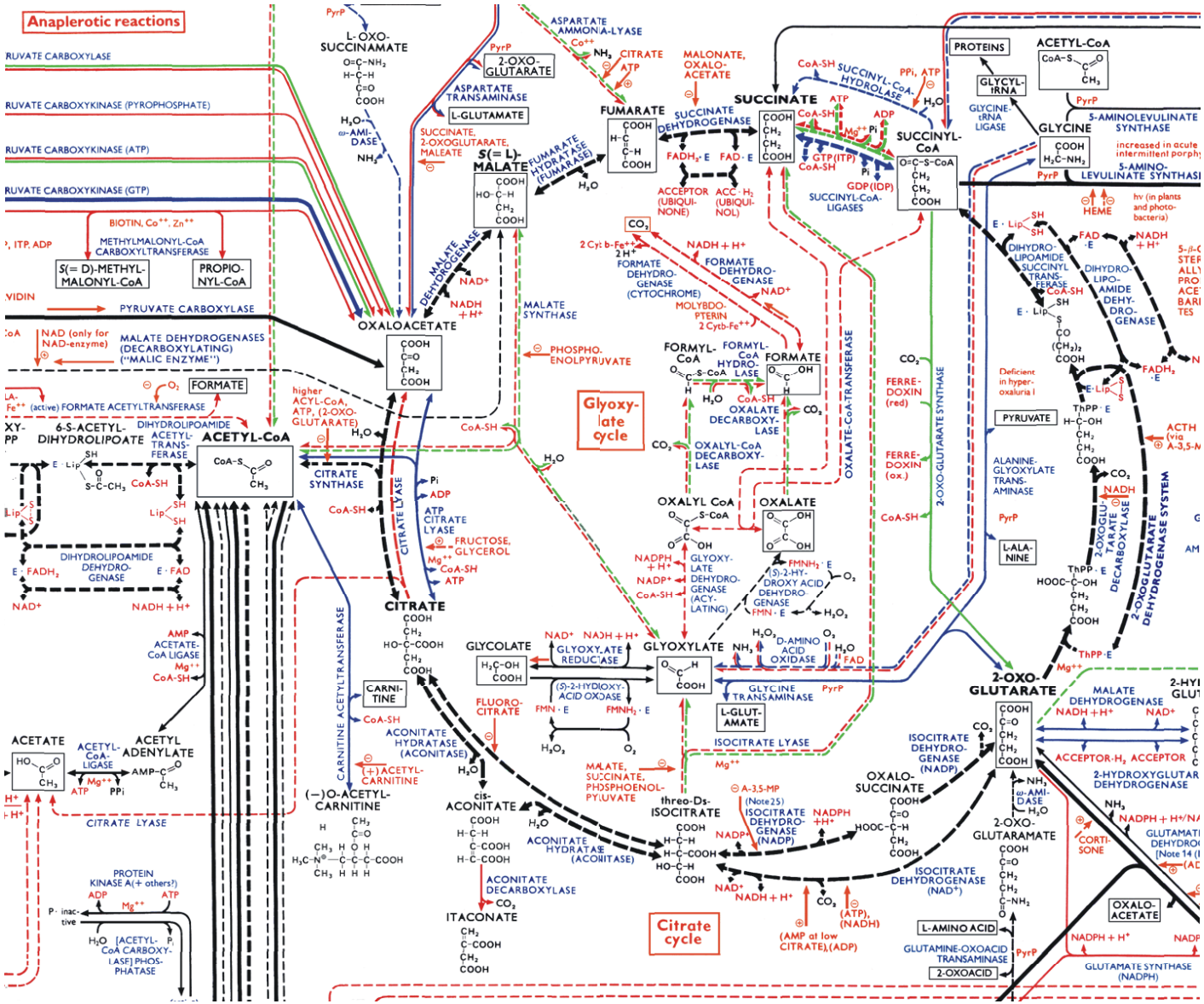
0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.01.004

[1]Discussion and analysis of combined genetic and metabolic networks has become so frequent and intense that we suggest to use a separate term, *genabolic networks*, for this class of complex dynamical systems.



Das Reaktionsnetzwerk des zellulären Metabolismus (Boehringer-Mannheim).

Der Zitronensäure oder Krebs Zyklus (vergrößert vom vorigen Bild).

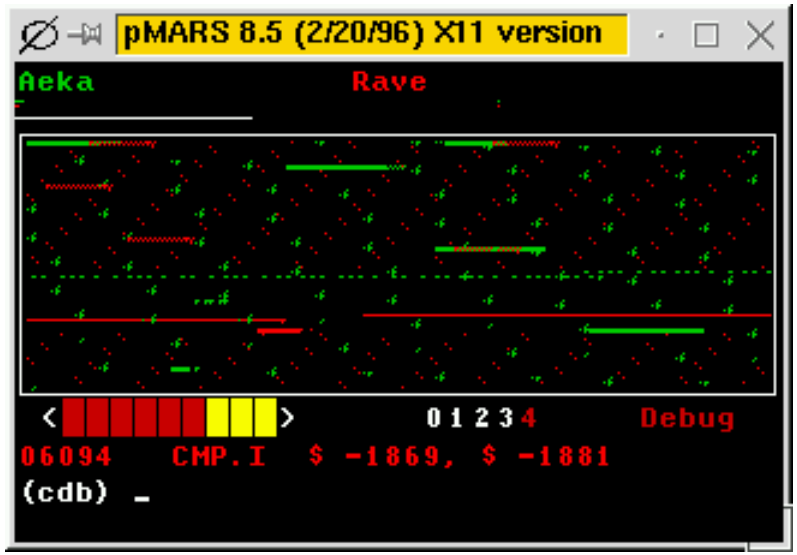


1. Mathematik und Physik
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. **Evolutionsforschung am Computer**
6. Evolution im ‚Flussreaktor'
7. Komplexität ‚ohne Ende'

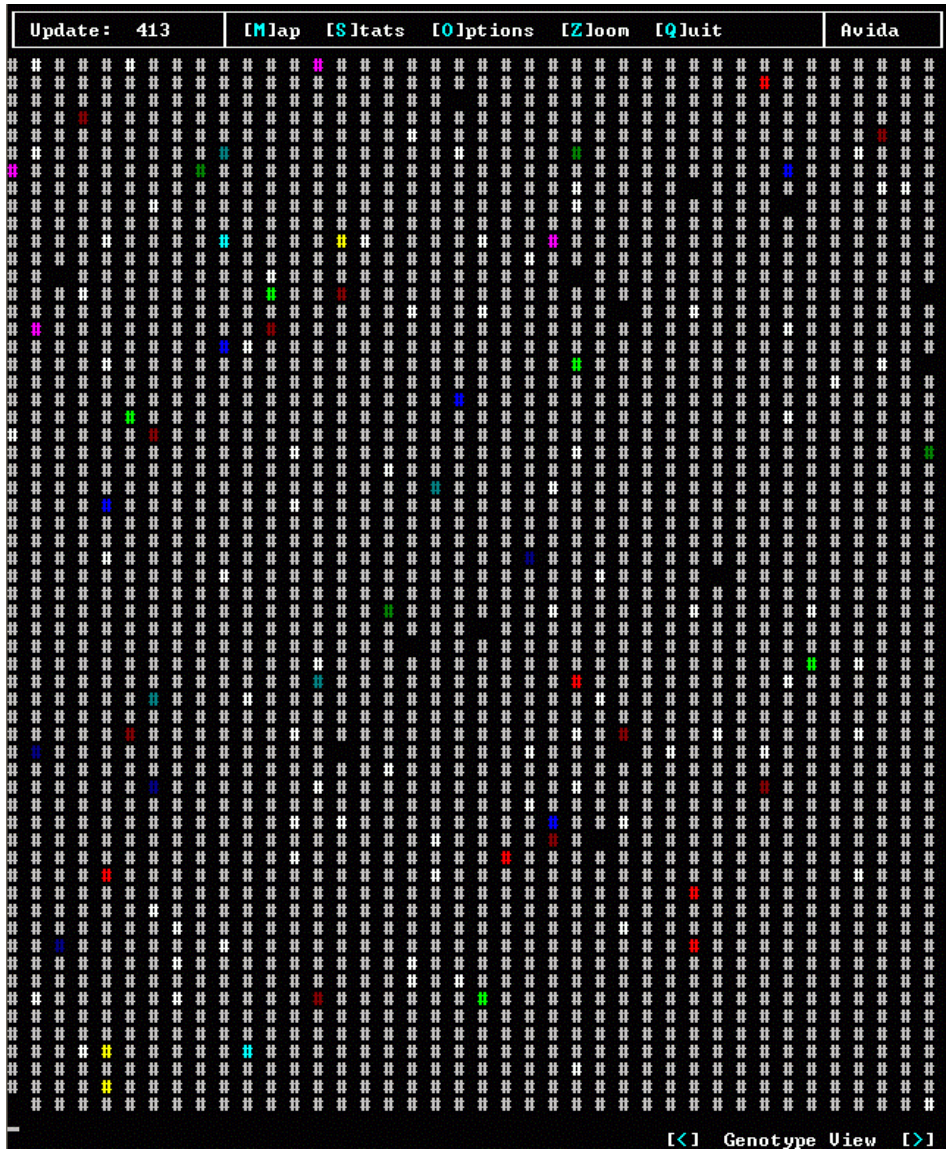
Simulation von Evolutionsprozessen

Flowreactor	chemische Kinetik, Gillespie algorithm	1987	Fontana, W., Schuster, P.
Tierra	CPU-core simulation, shared memory space, comes from Core-Wars	1991	Ray, T.
Avida	CPU-core simulation, modified Tierra protected memory space	1993	Ofria, C., Adami, C., Wilke, C.O.
Evolve	high-level language, comes from Game-of-Life (Conway) and Core-Wars	1996	Stauffer, K.
Framesticks	3D-world simulation	1996	Komosinski, M., Ulatowski, s.
Darwinbots	simulation on plane, no grid, stems from C-robots	2003	Comis, C.
breve	multi-purpose simulator, stems from Game-of-Life (Conway)	2006	Klein, J.

und kooperative Spiele (Maynard-Smith, Sigmund, Nowak) und vieles andere mehr.

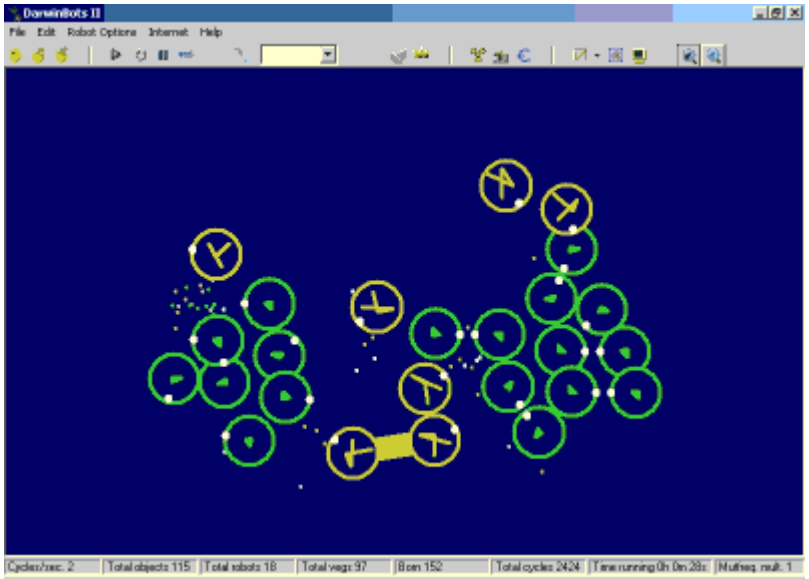


Core-Wars oder Tierra

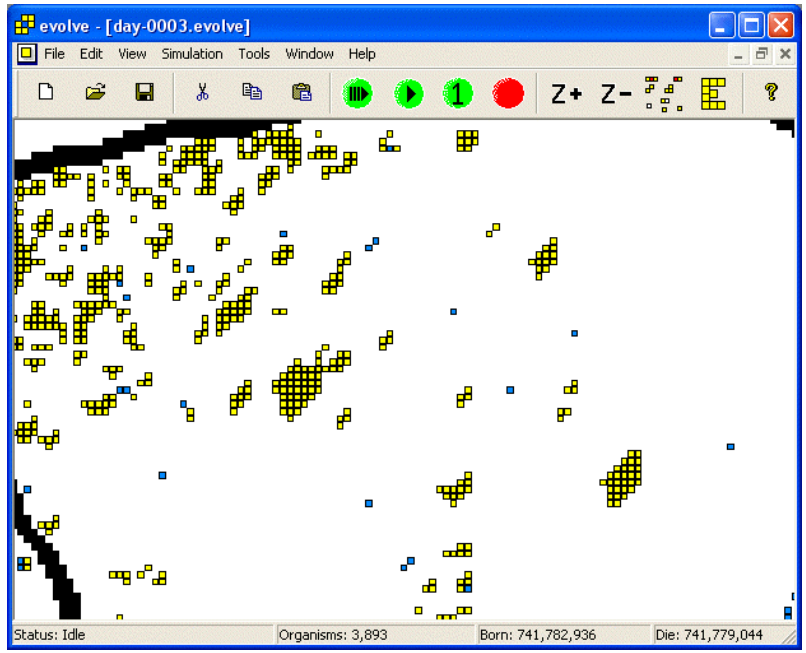


Avida

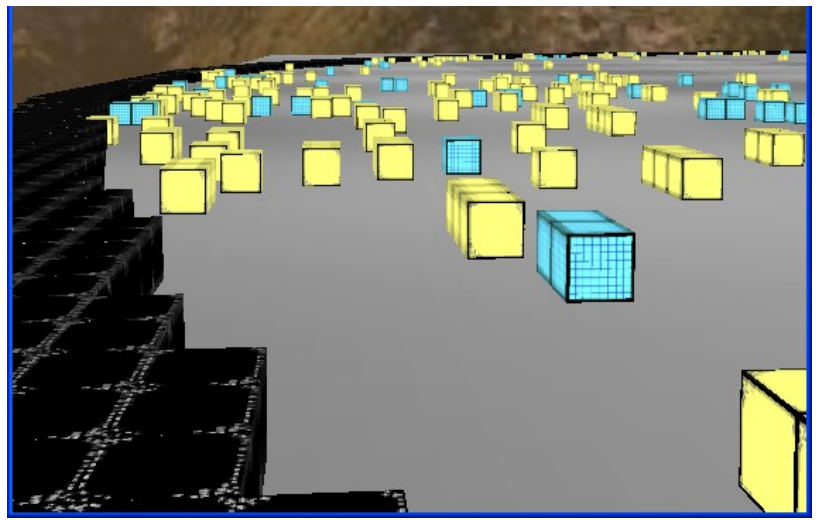
Momentaufnahmen von Simulationen der Evolutionsvorgänge im CPU-Core



Darwinbots

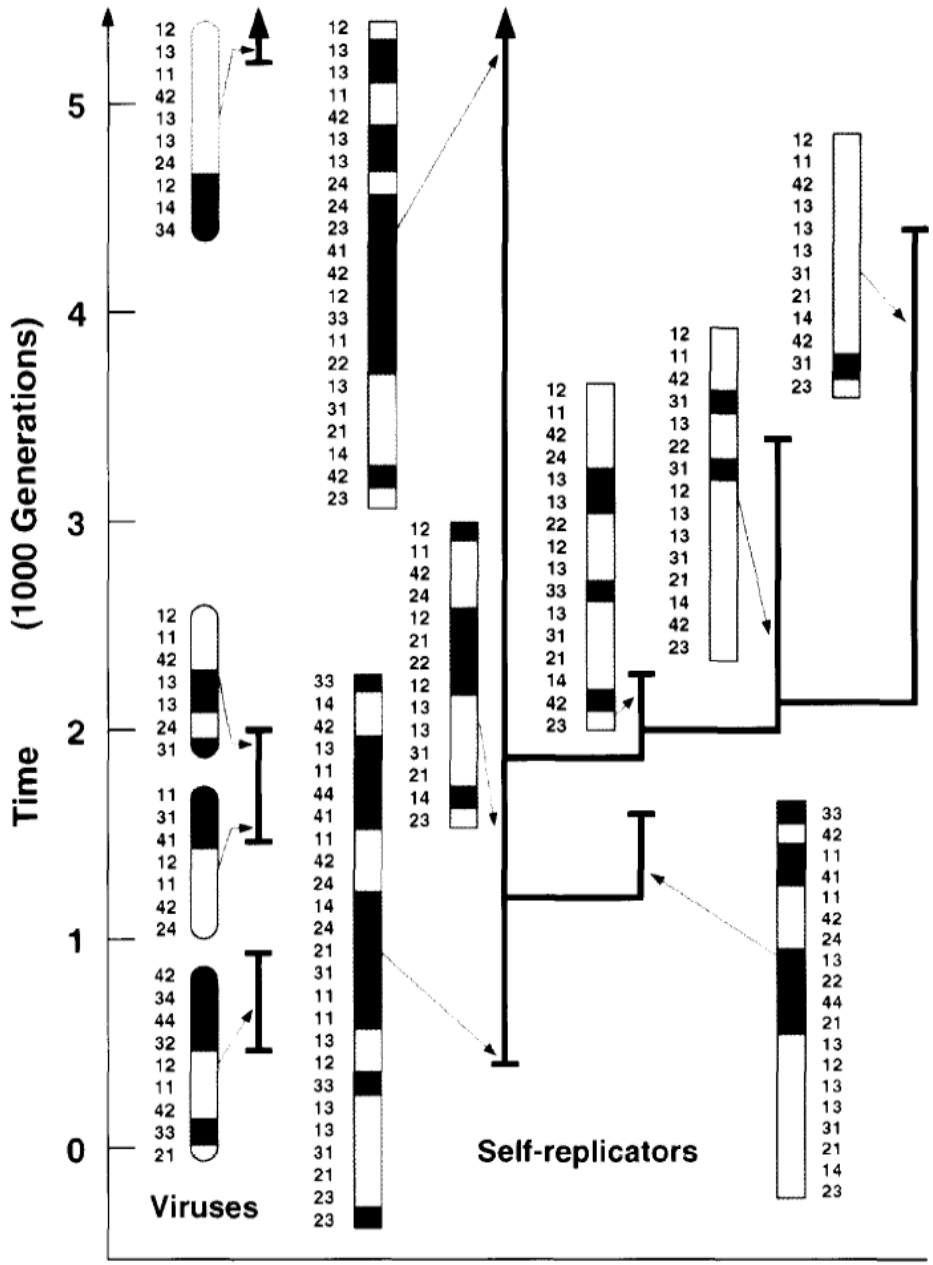


Evolve

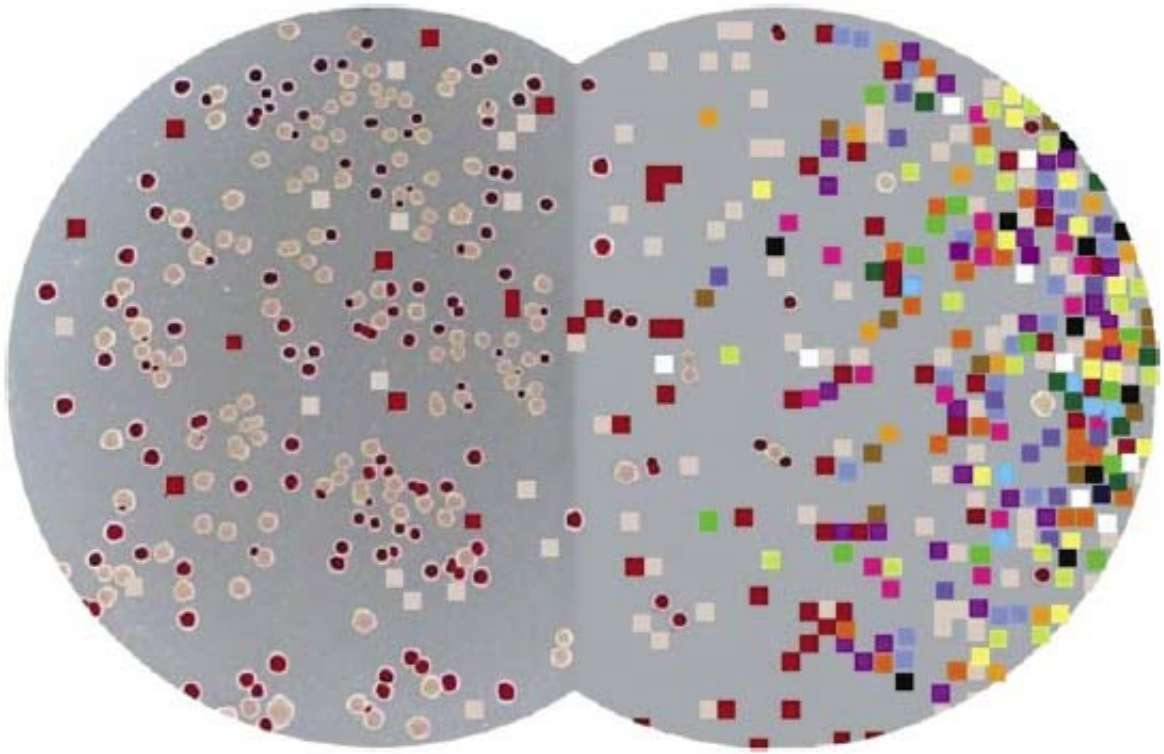


Momentaufnahmen von Simulationen der Evolutionsvorgänge

A.N. Pargellis. 1996. The spontaneous generation of digital „life“.
Physica D 91:86-96.



Phylogenie digitaler Organismen



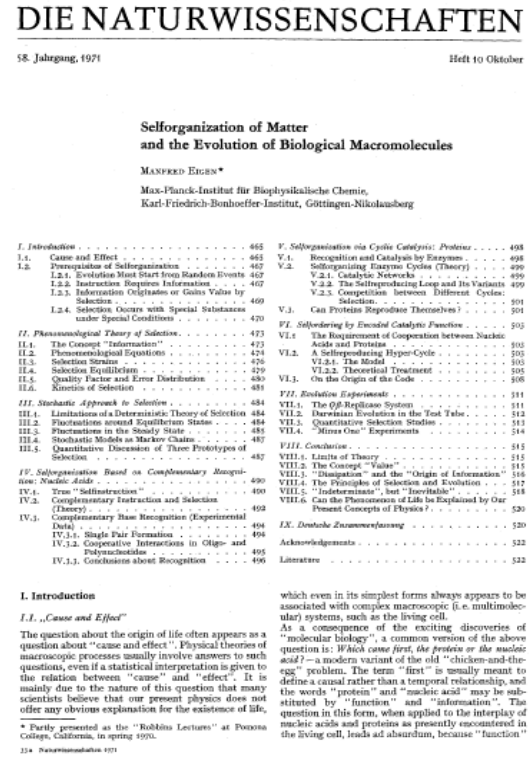
DOI: 10.1371/journal.pbio.0000018.g001

Figure 1. Hybrid Graphic of Petri Dishes with Bacteria Blending into Digital Organisms
Lenski spends as much research time with bacteria (left) as he does with digital organisms (right), balancing the strengths and limitations of the two systems in an effort to understand and explain the principles of evolutionary theory. (Hybrid graphic courtesy of Dusan Misevic, Michigan State University.)

Computersimulation einer Bakterienkultur mit „Digital Organisms“ (Avida).

B. O‘Neill. 2003. *Digital Evolution*. PLoS Biology 1:11-14, e18.

1. Mathematik und Physik
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. **Evolution im ‚Flussreaktor'**
7. Komplexität ‚ohne Ende'



DIE NATURWISSENSCHAFTEN

58. Jahrgang, 1971 — Heft 10 Oktober

Selforganization of Matter and the Evolution of Biological Macromolecules

MANFRED EIGEN*

Max-Planck-Institut für Biophysikalische Chemie,
Karl-Friedrich-Bonhoeffer-Institut, Göttingen-Nikolausberg

1. Introduction

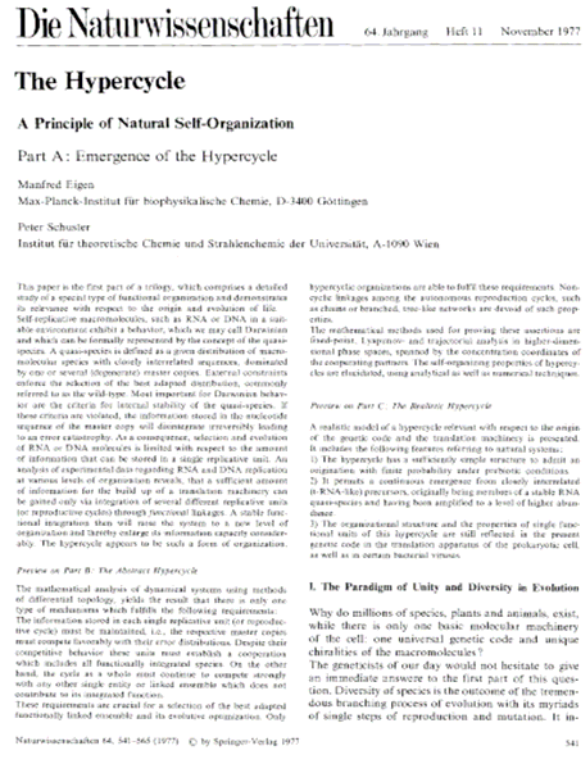
I.1. „Cause and Effect"

The question about the origin of life often appears as a question about "cause and effect". Physical theories of macroscopic processes usually involve answers to such questions, even if a statistical interpretation is given to the relation between "cause" and "effect". It is mainly due to the nature of this question that many scientists believe that our present physics does not offer any obvious explanation for the existence of life, which even in its simplest forms always appears to be associated with complex macroscopic (i.e. multimolecular) systems, such as the living cell.

As a consequence of the exciting discoveries of "molecular biology", a common version of the above question is: *Which came first, the protein or the nucleic acid?*—a modern variant of the old "chicken-and-the-egg" problem. The term "first" is usually meant to define a causal rather than a temporal relationship, and the words "protein" and "nucleic acid" may be substituted by "function" and "information". The question in this form, when applied to the interplay of nucleic acids and proteins as presently encountered in the living cell, leads ad absurdum, because "function"

* Partly presented as the "Robbins Lectures" at Pomona College, California, in spring 1970.

1971



Die Naturwissenschaften 64. Jahrgang Heft 11 November 1977

The Hypercycle

A Principle of Natural Self-Organization

Part A: Emergence of the Hypercycle

Manfred Eigen
Max-Planck-Institut für biophysikalische Chemie, D-3400 Göttingen

Peter Schuster
Institut für theoretische Chemie und Strahlenchemie der Universität, A-1090 Wien

This paper is the first part of a trilogy, which comprises a detailed study of a special type of functional organization and demonstrates its relevance with respect to the origin and evolution of life. Self-replicative macromolecules, such as RNA or DNA in a suitable environment exhibit a behavior, which we may call Darwinian and which can be formally represented by the concept of the quasi-species. A quasi-species is defined as a given distribution of macromolecular species with closely interrelated sequences, dominated by one or several (degenerate) master copies. External constraints enforce the selection of the best adapted distribution, commonly referred to as the wild-type. Most important for Darwinian behavior are the criteria for internal stability of the quasi-species. If these criteria are violated, the information stored in the nucleotide sequence of the master copy will deteriorate irreversibly leading to an error catastrophy. As a consequence, selection and evolution of RNA or DNA molecules is limited with respect to the amount of information that can be stored in a single replicative unit. An analysis of experimental data regarding RNA and DNA replication at various levels of organization reveals, that a sufficient amount of information for the build up of a translation machinery can be gained only via integration of several different replicative units (or reproductive cycles) through functional linkages. A stable functional integration then will raise the system to a new level of organization and thereby enlarge its information capacity considerably. The hypercycle appears to be such a form of organization.

Preview on Part B: The Abstract Hypercycle

The mathematical analysis of dynamical systems using methods of differential topology, yields the result that there is only one type of mechanisms which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective master copies must compete favorably with their error distributions. Despite their competitive behavior these units must establish a cooperation which includes all functionally integrated species. On the other hand, the cycle as a whole must continue to compete strongly with any other single entity or linked ensemble which does not contribute to its integrated function.

These requirements are crucial for a selection of the best adapted functionally linked ensemble and its evolutive optimization. Only hypercyclic organizations are able to fulfil these requirements. Non-cyclic linkages among the autonomous reproduction cycles, such as chains or branched, tree-like networks are devoid of such properties.

The mathematical methods used for proving these assertions are fixed-point, Lyapunov- and trajectorial analysis in higher-dimensional phase spaces, spanned by the concentration coordinates of the cooperating partners. The self-organizing properties of hypercycles are elucidated, using analytical as well as numerical techniques.

Preview on Part C: The Realistic Hypercycle

A realistic model of a hypercycle relevant with respect to the origin of the genetic code and the translation machinery is presented. It includes the following features referring to natural systems:

1) The hypercycle has a sufficiently simple structure to admit an origination with finite probability under prebiotic conditions.

2) It permits a continuous emergence from closely interrelated (t-RNA-like) precursors, originally being members of a stable RNA quasi-species and having been amplified to a level of higher abundance.

3) The organizational structure and the properties of single functional units of this hypercycle are still reflected in the present genetic code in the translation apparatus of the prokaryotic cell, as well as in certain bacterial viruses.

1. The Paradigm of Unity and Diversity in Evolution

Why do millions of species, plants and animals, exist, while there is only one basic molecular machinery of the cell: one universal genetic code and unique chiralities of the macromolecules?

The geneticists of our day would not hesitate to give an immediate answere to the first part of this question. Diversity of species is the outcome of the tremendous branching process of evolution with its myriads of single steps of reproduction and mutation. It in-

1977

Reprinted from The Journal of Physical Chemistry, 1988, 92, 6881.
Copyright © 1988 by the American Chemical Society and reprinted by permission of the copyright owner.

Molecular Quasi-Species†

Manfred Eigen,* John McCaskill,
Max Planck Institut für biophysikalische Chemie, Am Fassberg, D 3400 Göttingen-Nikolausberg, BRD
and Peter Schuster*
Institut für theoretische Chemie und Strahlenchemie, der Universität Wien, Währinger Strasse 17, A-1090 Wien, Austria (Received: June 9, 1988)

The molecular quasi-species model describes the physicochemical organization of monomers into an ensemble of heteropolymers with combinatorial complexity by ongoing template polymerization. Polynucleotides belong to the simplest class of such molecules. The quasi-species itself represents the stationary distribution of macromolecular sequences maintained by chemical reactions effecting error-prone replication and by transport processes. It is obtained deterministically, by mass-action kinetics, as the dominant eigenvalue of a *value* matrix, **W**, which is derived directly from chemical rate coefficients, but it also exhibits stochastic features, being composed to a significant fraction of unique individual macromolecular sequences. The quasi-species model demonstrates how macromolecular information originates through specific nonequilibrium autocatalytic reactions and thus forms a bridge between reaction kinetics and molecular evolution. Selection and evolutionary optimization appear as new features in physical chemistry. Concentration bias in the production of mutants is a new concept in population genetics, relevant to frequently mutating populations, which is shown to greatly enhance the optimization properties. The present theory relates to asexually replicating ensembles, but this restriction is not essential. A sharp transition is exhibited between a drifting population of essentially random macromolecular sequences and a localized population of close relatives. This transition at a threshold error rate was found to depend on sequence lengths, distributions of selective values, and population sizes. It has been determined generically for complex landscapes and for special cases, and, it was shown to persist generically in the presence of nearly neutral mutants. Replication dynamics has much in common with the equilibrium statistics of complex spin systems: the error threshold is equivalent to a magnetic order-disorder transition. A rational function of the replication accuracy plays the role of temperature. Experimental data obtained from test-tube evolution of polynucleotides and from studies of natural virus populations support the quasi-species model. The error threshold seems to set a limit to the genome lengths of several classes of RNA viruses. In addition, the results are relevant even in eucaryotes where they contribute to the exon–intron debate.

1. Molecular Selection

Our knowledge of physical and chemical systems is, in a final analysis, based on models derived from repeatable experiments. While none of the classic and rather besieged list of properties rounded up to support the intuition of a distinction between the living and nonliving—metabolism, self-reproduction, irritability, and adaptability, for example—intrinsically limit the application of the scientific method, a determining role by unique or *individual* entities comes into conflict with the requirement of repeatability. Combinatorial variety, such as that in heteropolymers based on even very small numbers of different bases, even just two, readily provides numbers of different entities so enormous that neither consecutive nor parallel physical realization is possible. The physical chemistry of finite systems of such macromolecules must deal with both known regularities and the advent of unique copolymeric sequences. Normally this would present no difficulty in a statistical mechanical analysis of typical behavior, where rare events play no significant role, but with autocatalytic polymerization processes even unique single molecules may be amplified to determine the fate of the entire system. Potentially creative, self-organizing around unique events, the dynamics of this simplest living chemical system is invested with regularities that both allow and limit efficient adaptation. The quasi-species model is a study of these regularities.

The fundamental regularity in living organisms that has invited explanation is adaptation. Why are organisms so well fitted to their environments? At a more chemical level, why are enzymes optimal catalysts? Darwin's theory of natural selection has provided biologists with a framework for the answer to this question. The present model is constructed along Darwinian lines but in terms of specific macromolecules, chemical reactions, and physical processes that make the notion of survival of the fittest precise. Not only does the model give an understanding of the physical limitations of adaptation, but also it provides new insight into the role of chance in the process. For an understanding of the structure of this minimal chemical model it is first necessary to recall the conceptual basis of Darwin's theory.

Darwin recognized that new inheritable adaptive properties were not induced by the environment but arose independently in the production of offspring. Lasting adaptive changes in a population could only come about by natural selection of the heritable traits or *genotype* based on the full characteristics or *phenotype* relevant for producing offspring. A process of chance, i.e., uncorrelated with the developed phenotype, controls changes in the genotype from one generation to the next and generates the diversity necessary for selection. Three factors have probably prevented chemists from gaining a clear insight into these phenomena in the past, despite the discovery of the polymeric nature of the genotype (DNA): the complexity of a minimum replication phenotype, the problem of dealing with a huge number of variants, and the nonequilibrium nature of these ongoing processes.

The formulation of a tractable chemical model based on Darwin's principle may be understood in several steps:

1. The major constituents of the system have to be inherently self-reproductive. Only two classes of molecules are presently

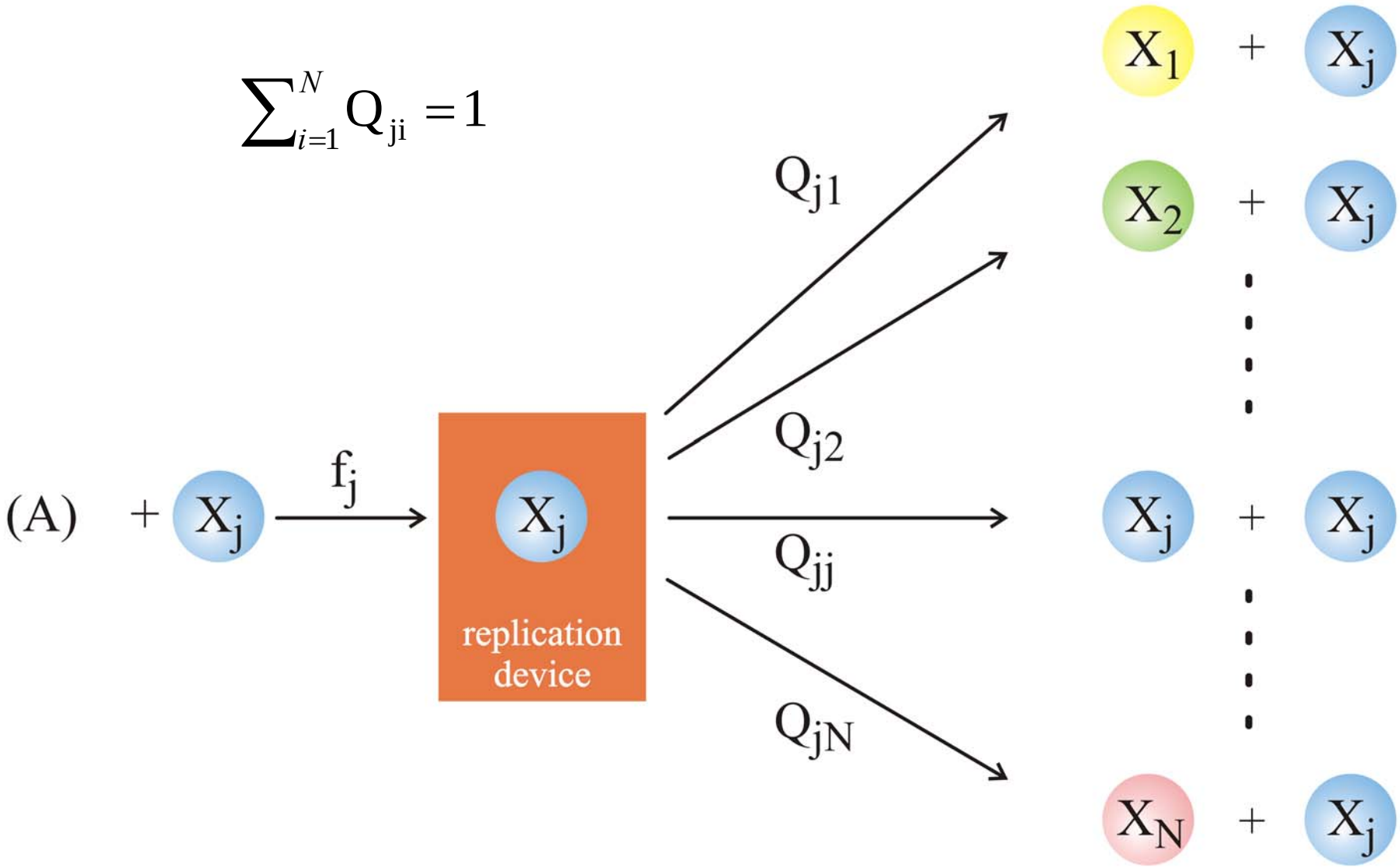
† This is an abridged account of the quasi-species theory that has been submitted in comprehensive form to *Advances in Chemical Physics*.[1]

(1) Eigen, M.; McCaskill, J. S.; Schuster, P. *Adv. Chem. Phys.*, in press.

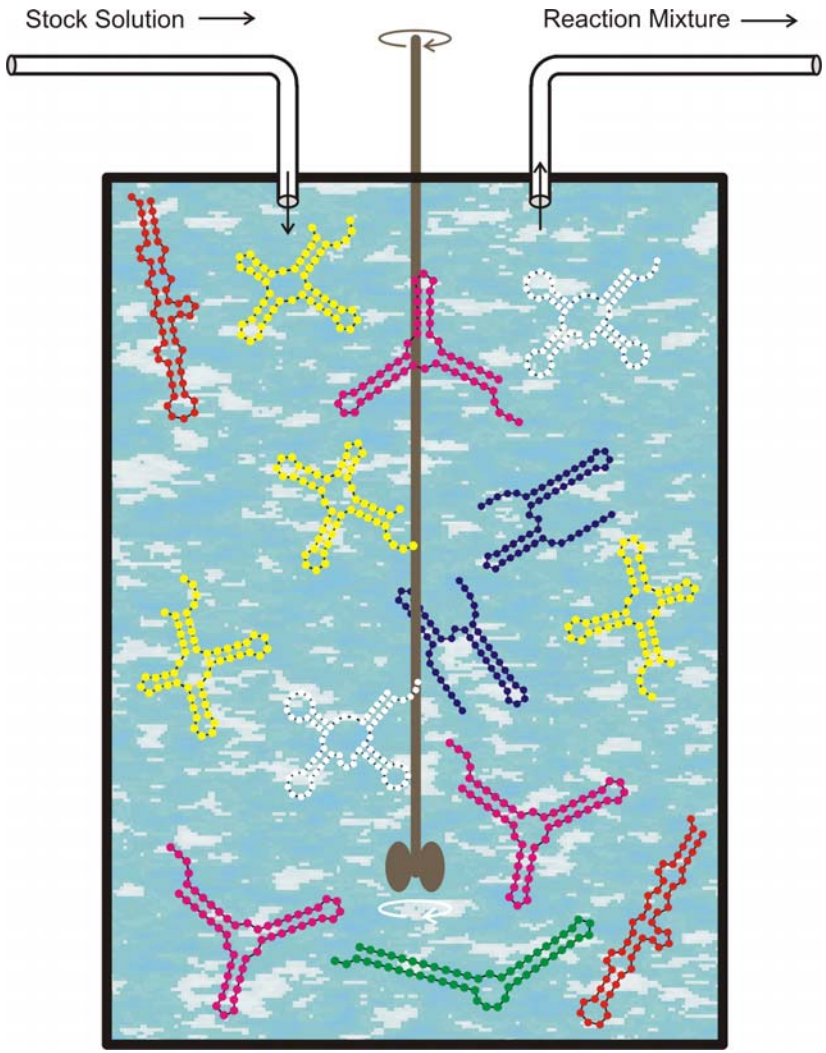
0022-3654/88/2092-6881$01.50/0 © 1988 American Chemical Society

1988

Chemische Kinetik der molekularen Evolution



Chemische Kinetik von Replikation und Mutation als Parallelreaktionen



Chemische Kinetik von Replikation und Mutation als Differentialgleichungen

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} = \sum_{i=1}^{N} Q_{ji}\, f_i\, x_i - \Phi\, x_j\,, \;\; j = 1,2,\ldots,N;\; \sum_{i=1}^{N} x_i = 1$$

$$\Phi(\mathrm{t}) = \sum_{i=1}^{N} f_i\; x_i$$

$$\lim_{t\to\infty} \frac{\mathrm{d}x_i}{\mathrm{dt}} = 0;\;\; \lim_{t\to\infty} x_i(\mathrm{t}) = \bar{x}_i$$

$$\left\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\right\} \ldots \text{Quasispezies}$$

Chemische Kinetik von Replikation und Mutation als Differentialgleichungen

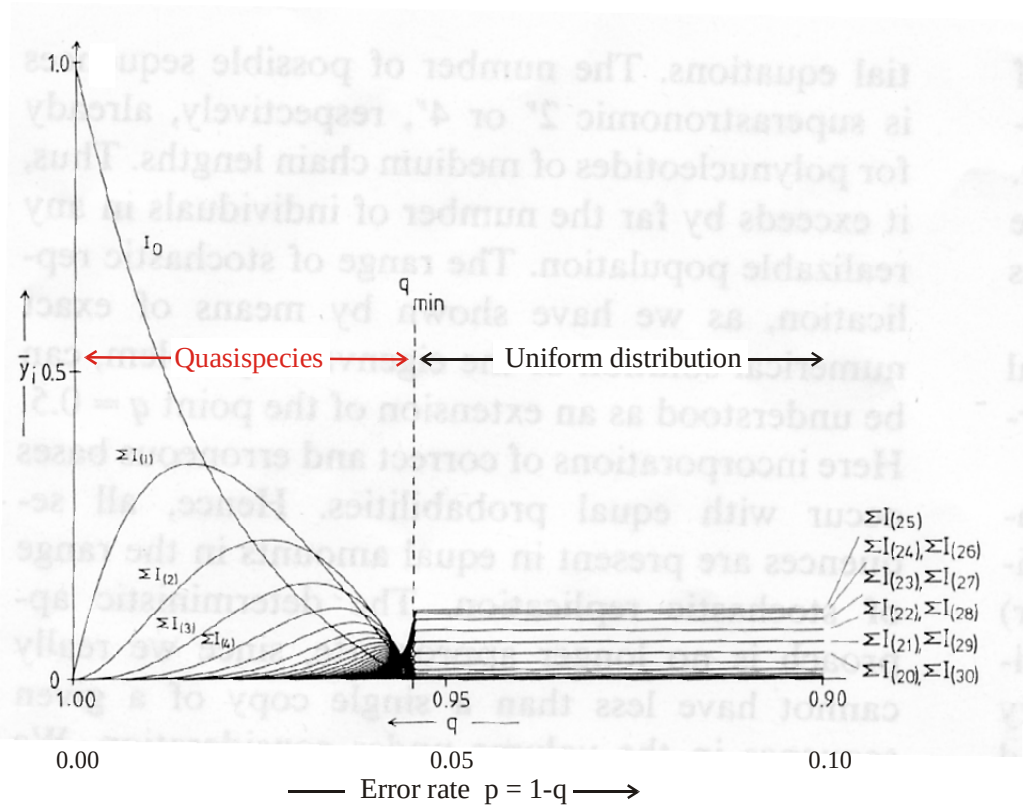
Biophysical Chemistry 16 (1982) 329–345
Elsevier Biomedical Press

SELF-REPLICATION WITH ERRORS

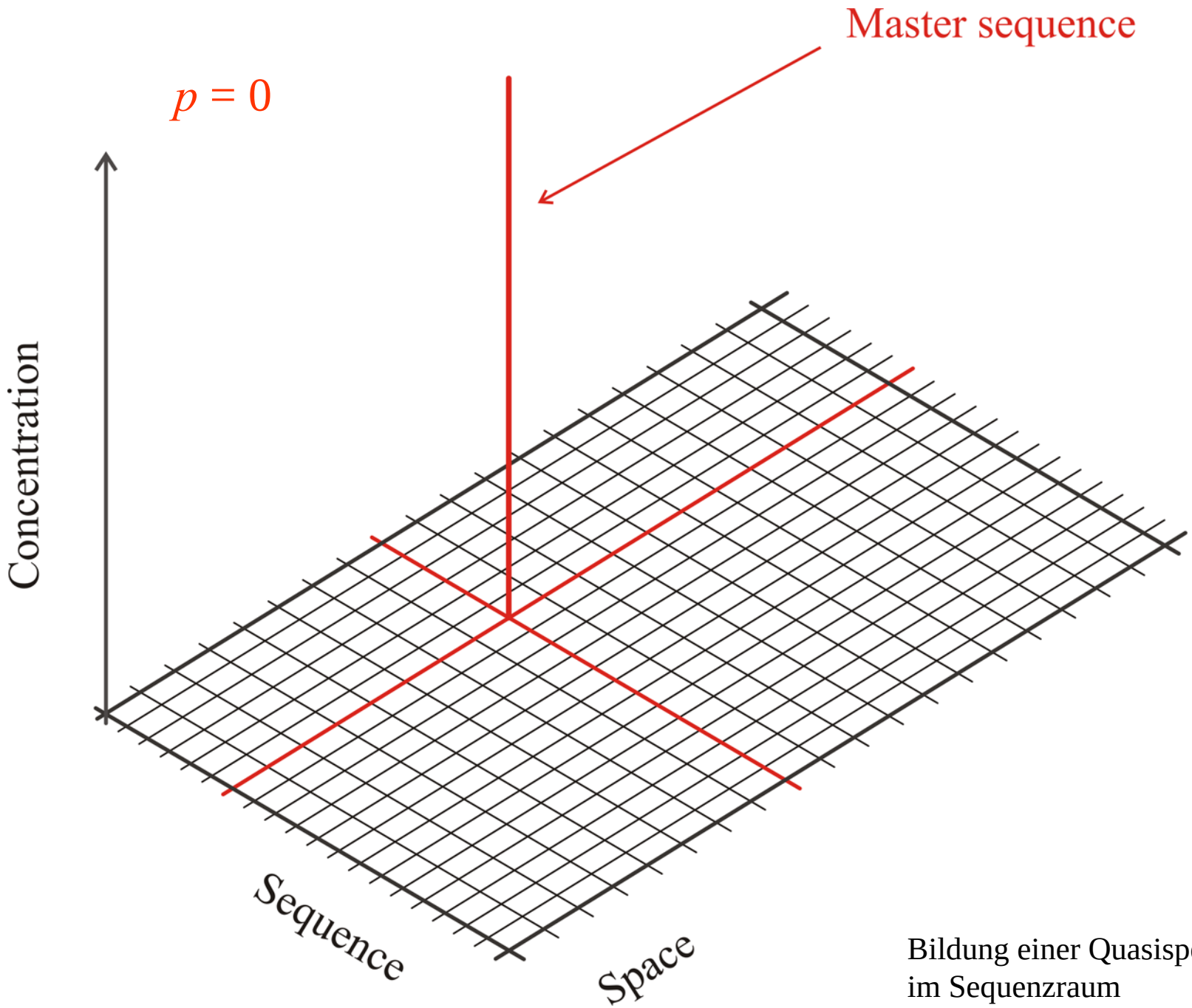
A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

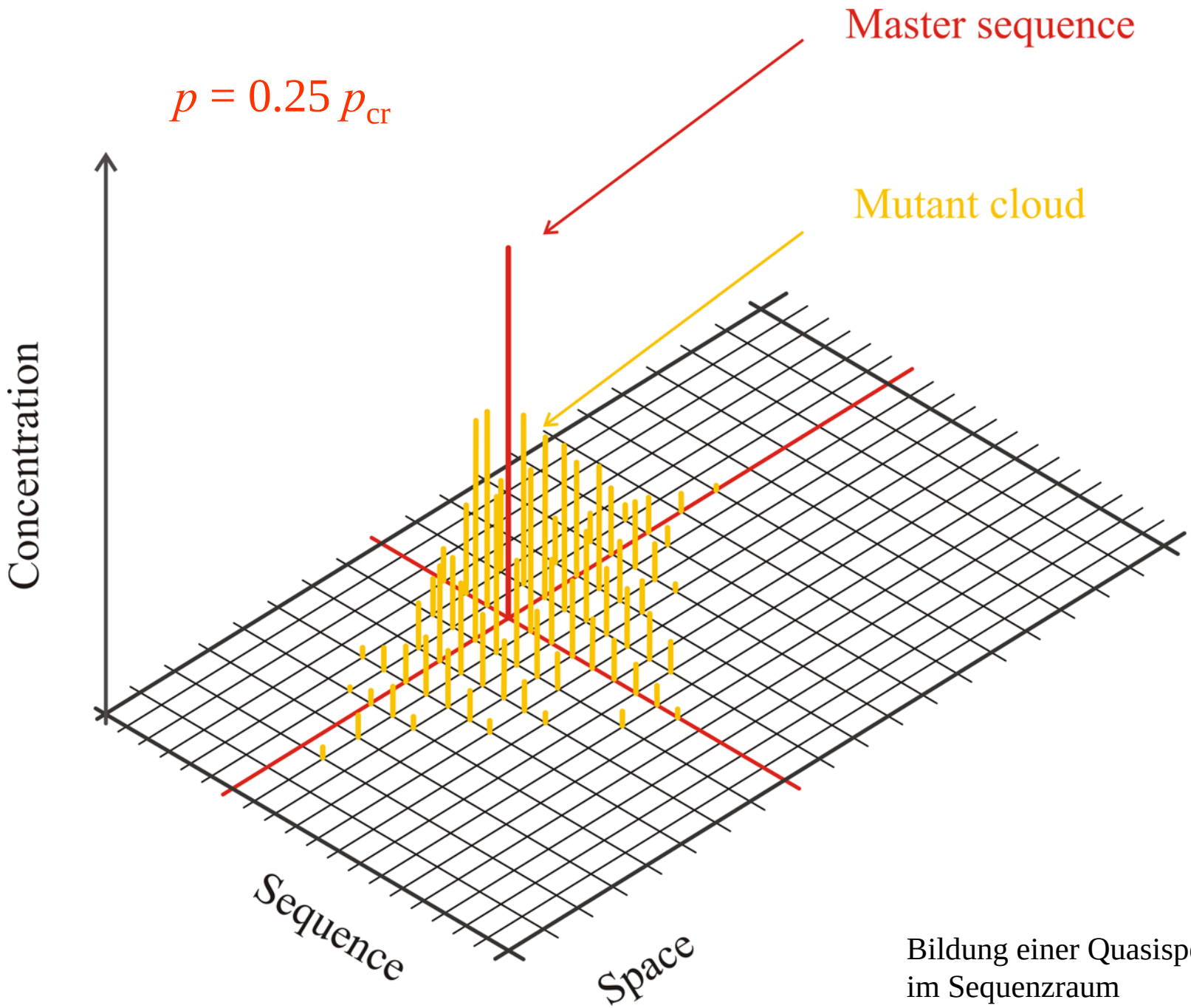
Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria



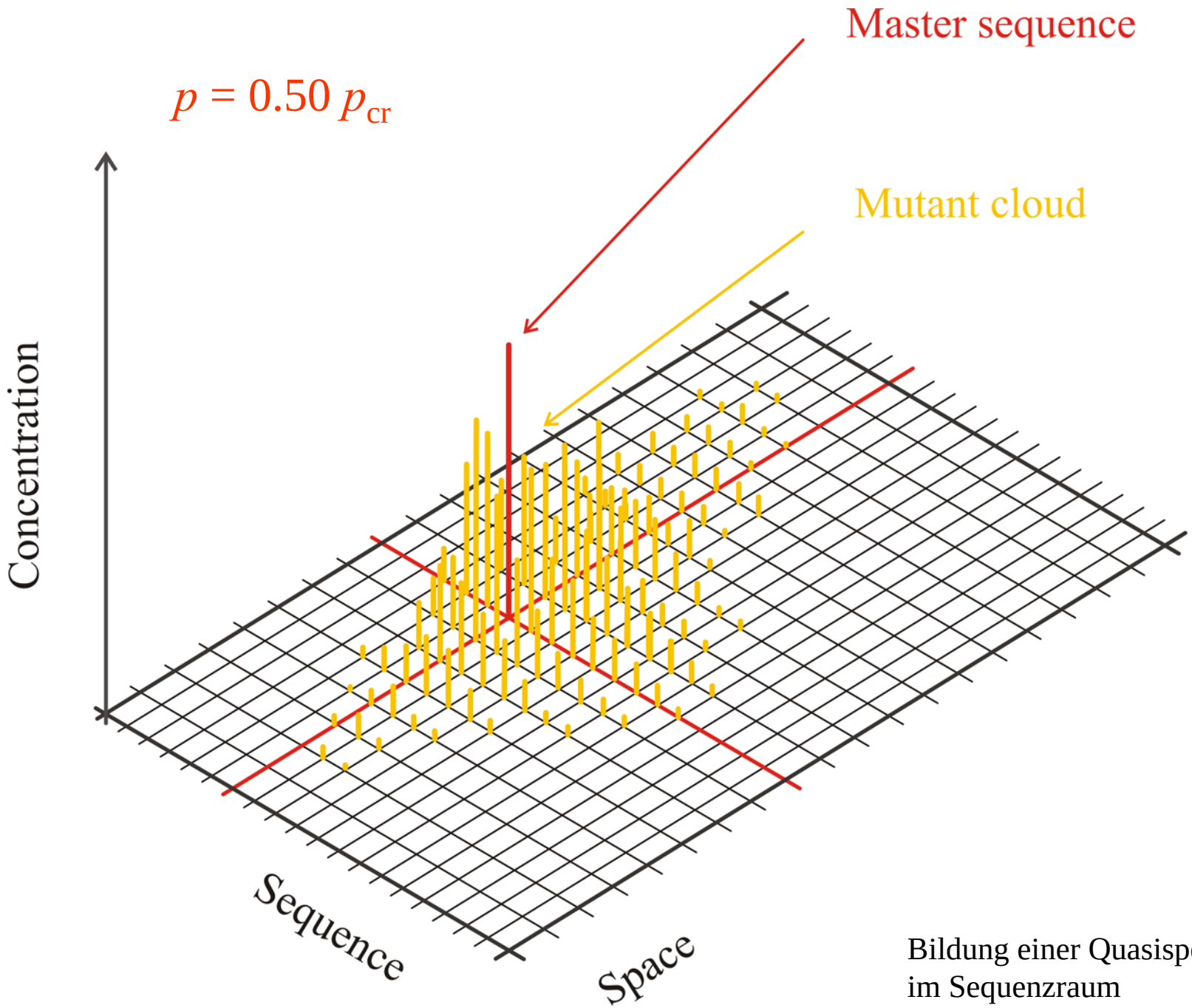
Die stationäre Population oder Quasispezies als Funktion der Mutationsrate *p*



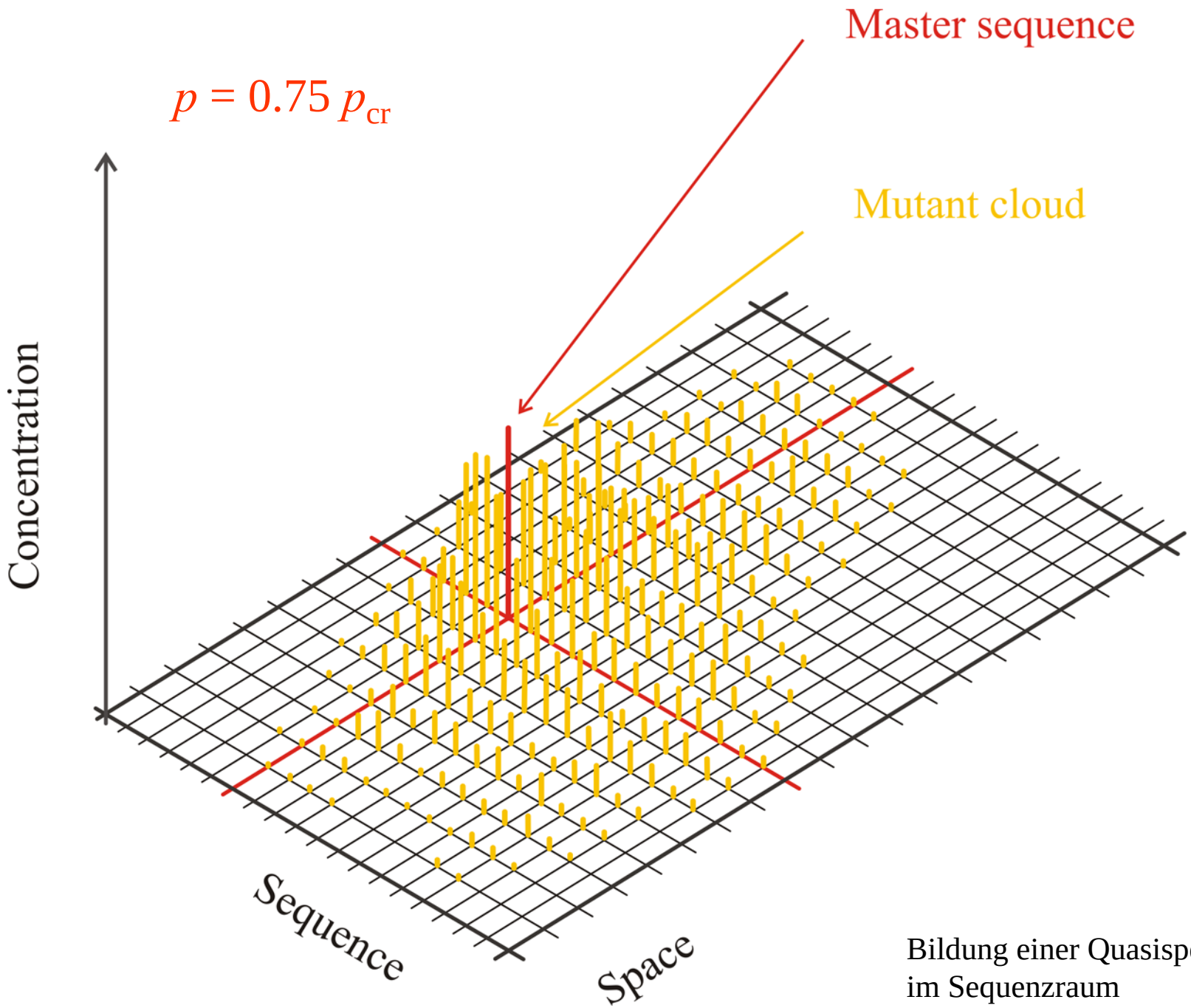
Bildung einer Quasispezies
im Sequenzraum



Bildung einer Quasispezies
im Sequenzraum

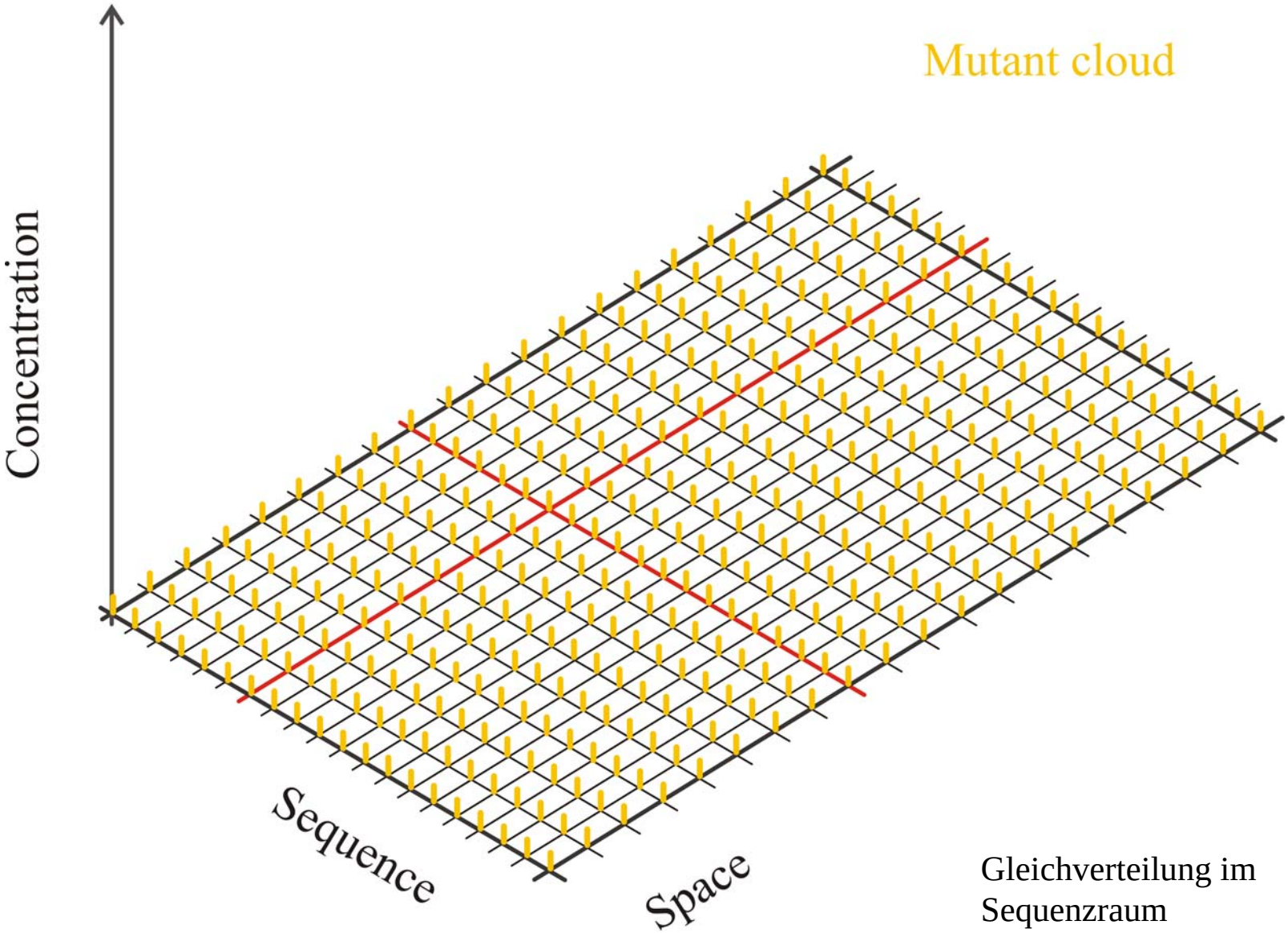


Bildung einer Quasispezies
im Sequenzraum

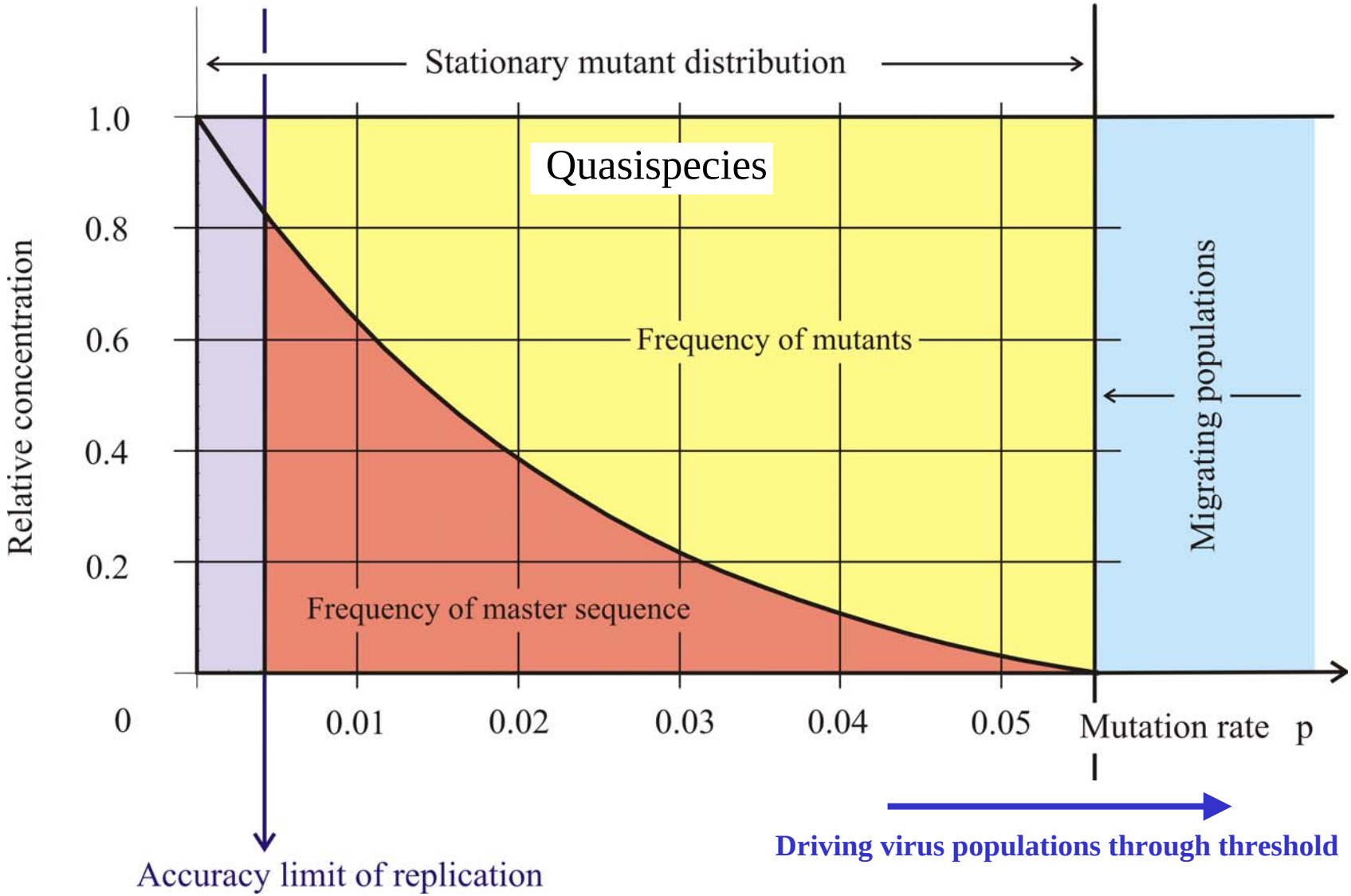


Bildung einer Quasispezies
im Sequenzraum

$p \geq p_{cr}$



Gleichverteilung im Sequenzraum



Fehlerschwellen oder Mutationsgrenzen bei der Replikation

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 107 (2005) 115–116
Virus Research
www.elsevier.com/locate/virusres

Preface

Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations on enzymes involved in protein synthesis, as a theory of aging. In those times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be eliminated through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variation in differentiated organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to unveil the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Population geneticists rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparison of electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA recombination, and rapid nucleic acid sequencing techniques, molecular analyses of genomes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Most cellular processes, including such essential information-bearing and transferring events as genome replication, transcription and translation, are increasingly perceived as inherently inaccurate. Viruses, and in particular RNA viruses, are among the most extreme examples of exploitation of replication inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in quantitative terms as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Recently, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged as a possible antiviral strategy. This is the topic of the current special issue of ***Virus Research***.

Few would nowadays doubt that one of the major obstacles for the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principles that have shaped biological evolution over eons, that is, repeated rounds of reproduction with genetic variation, competition and selection, often perturbed by random events such as statistical fluctuations in population size. However, with viruses the consequences of the operation of these very same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be also observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens begs for alternative antiviral strategies, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect different lines of evidence (both theoretical and experimental) on which antiviral designs based on genetic deterioration inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted to progeny. Closely related to the theoretical developments have been numerous experimental studies on quasispecies dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of existing as mutant spectra rather than defined genetic entities, remarkably expand their potential to overcome selective pressures intended to limit their replication. Indeed, the use of antiviral inhibitors in clinical practice and the design of vaccines for a number of major RNA virus-associated diseases, are currently presided by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replicases, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy received an important impulse by the observation that ribavirin (a licensed antiviral nucleoside analogue) may be exerting, in some systems, its antiviral activity through enhanced mutagenesis. This has encouraged investigations on new mutagenic base analogues, some of them used in anticancer chemotherapy. Some chapters summarize these important biochemical studies on cell entry pathways and metabolism of mutagenic agents, that may find new applications as antiviral agents.

This volume intends to be basically a progress report, an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I can envisage (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific mutagens and inhibitors in an effective way, finding synergistic combinations and avoiding antagonistic ones as well as severe clinical side effects. (ii) Another on a deeper understanding of the metabolism of mutagenic agents, in particular base and nucleoside analogues. This includes identification of the transporters that carry them into cells, an understanding of their metabolic processing, intracellular stability and alterations of nucleotide pools, among other issues. (iii) Still another line of needed research is the development of new mutagenic agents specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with anticancer research, but others should result from the designs of new molecules, based on the structures of viral polymerases. I really hope that the reader finds this issue not only to be an interesting and useful review of the current situation in the field, but also a stimulating exposure to the major problems to be faced.

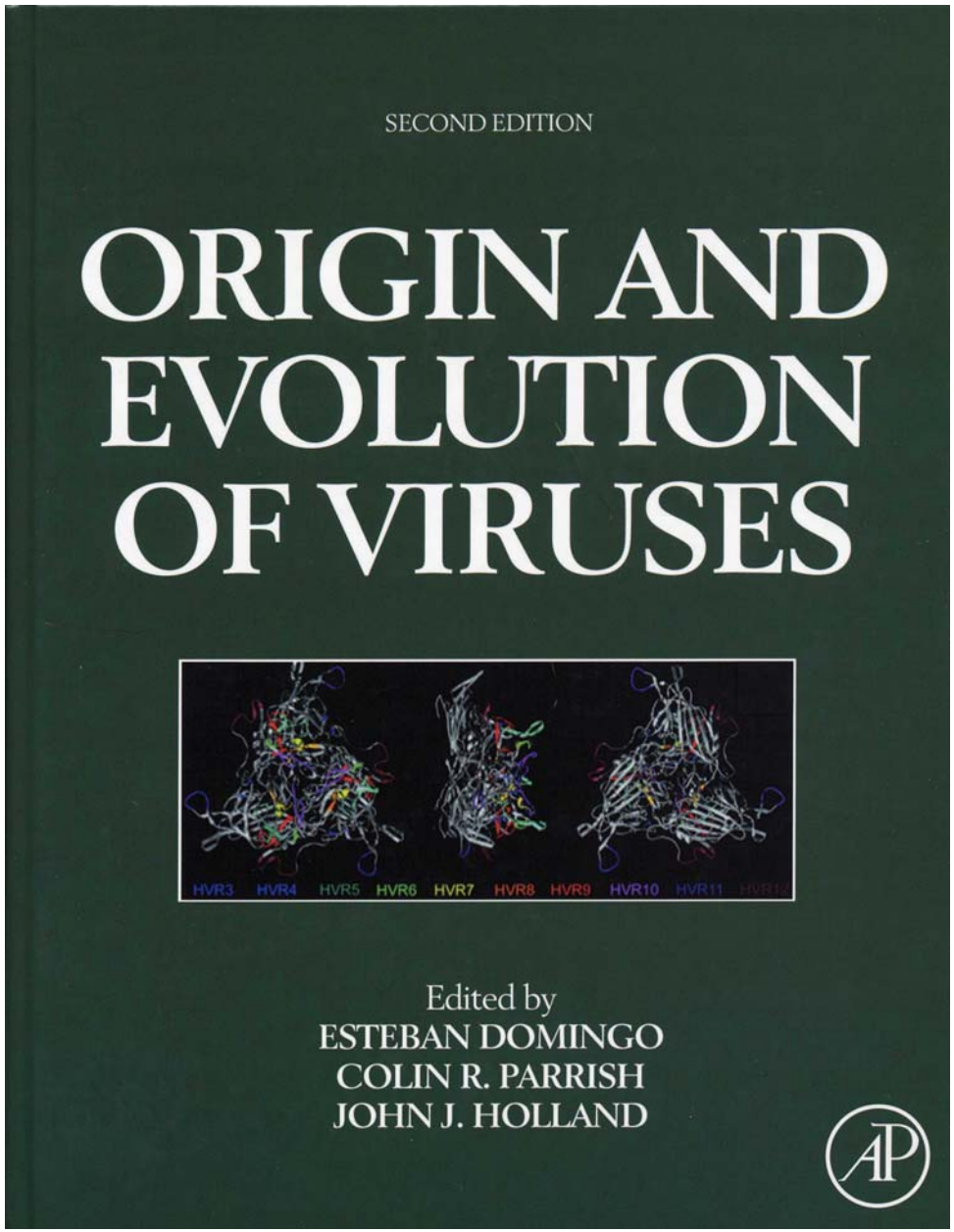
The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of ***Virus Research***, and then taken enthusiastically by Luis Enjuanes, recently appointed as Editor of ***Virus Research***. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of ***Virus Research***, Brian Mahy, for their continued interest and support to the research on virus evolution over the years.

My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
Universidad Autónoma de Madrid
Centro de Biología Molecular "Severo Ochoa"
Consejo Superior de Investigaciones Científicas
Cantoblanco and Valdeolmos
Madrid, Spain
Tel.: +34 91 497 84858/9; fax: +34 91 497 4799
E-mail address: edomingo@cbm.uam.es

Available online 8 December 2004

0168-1702/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2004.11.001



Molekulare Evolution von Viren

Evolution im Reagenzglas

G.F. Joyce, *Angew.Chem.Int.Ed.* **46** (2007), 6420-6436

DOI: 10.1002/anie.200701369

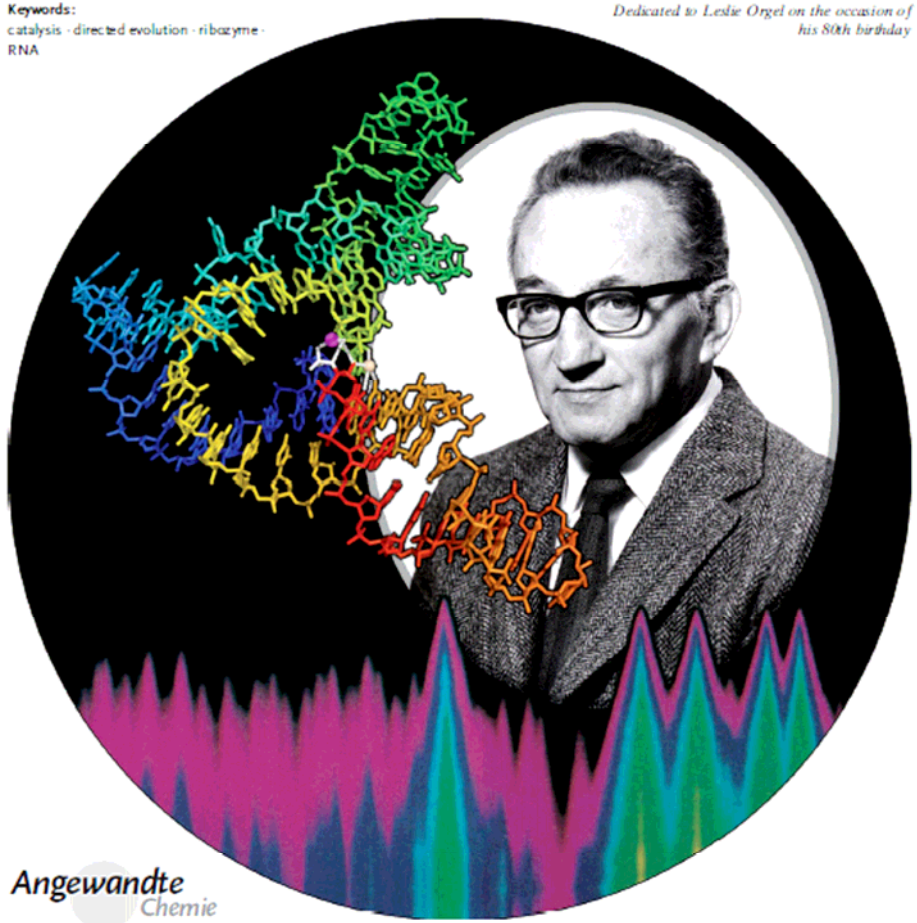
Molecular Evolution

Forty Years of In Vitro Evolution**

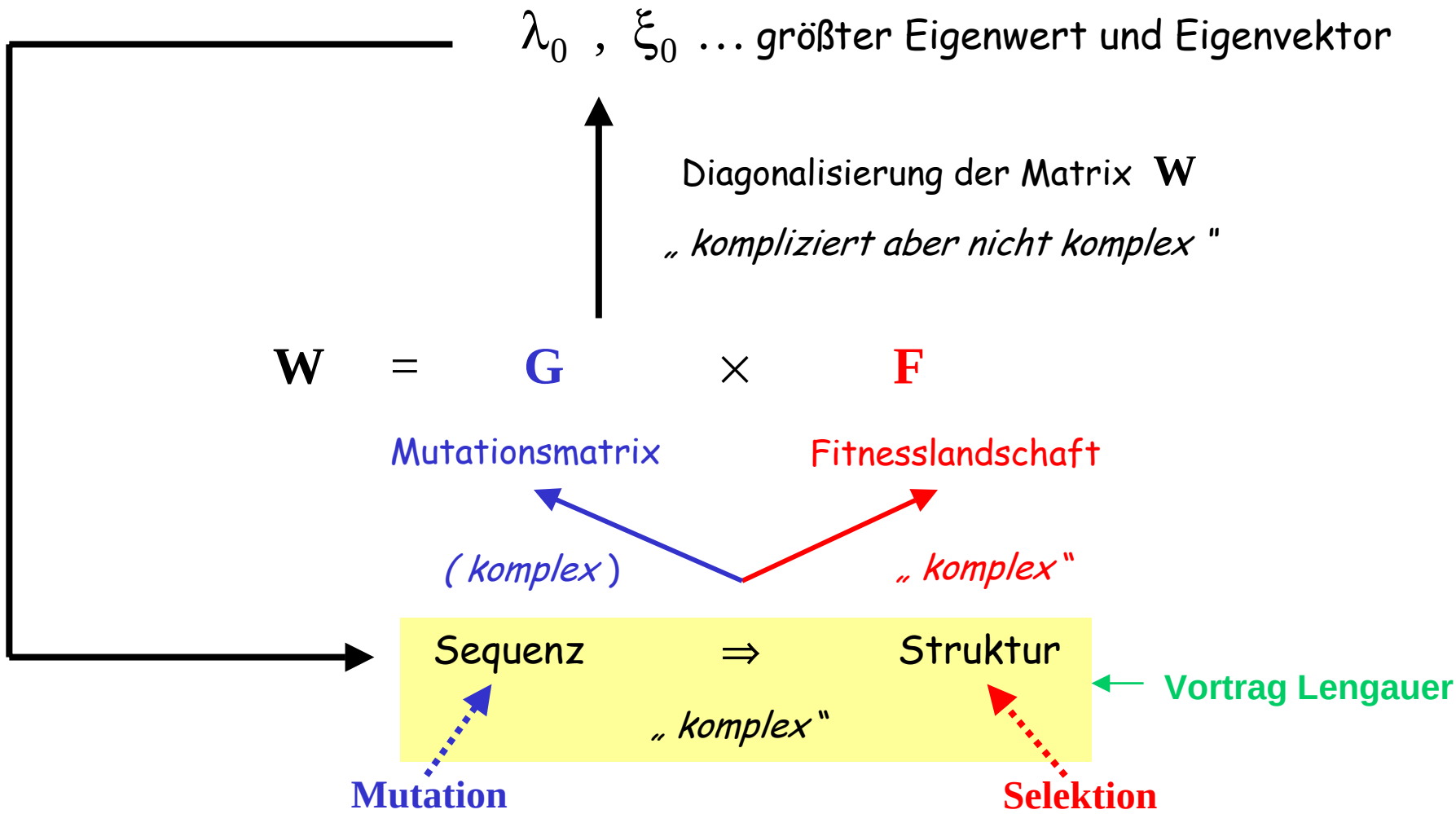
*Gerald F. Joyce**

Keywords:
catalysis · directed evolution · ribozyme · RNA

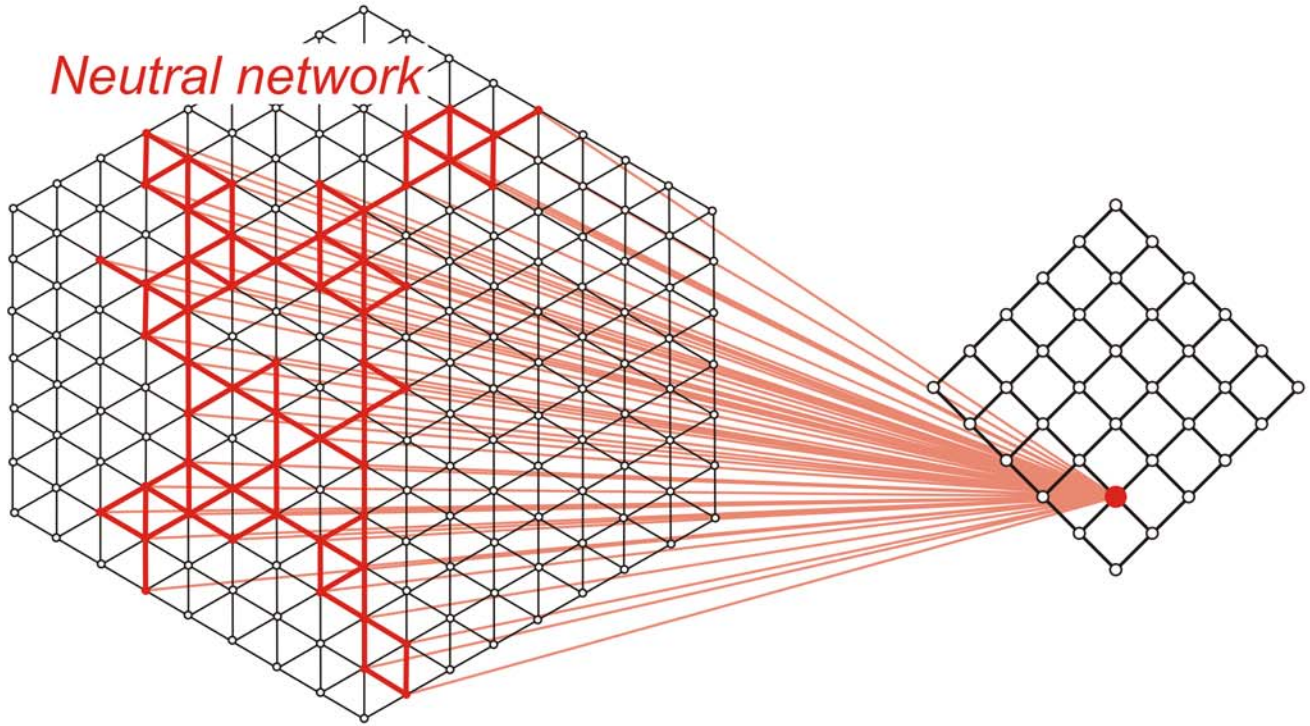
Dedicated to Leslie Orgel on the occasion of his 80th birthday



© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim



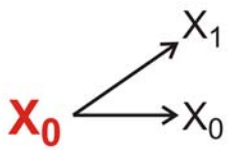
Komplexität der molekularen Evolution in der Modellierung mit Differentialgleichungen



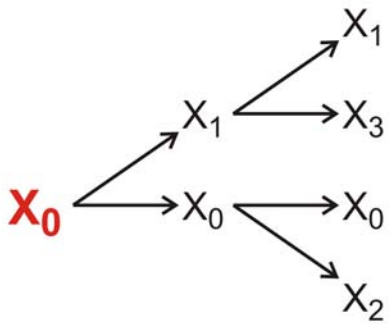
many genotypes ⇒ one phenotype

Die Beziehung zwischen Sequenzen und Strukturen als ein Abbildung vom Raum der Sequenzen in den Raum der Strukturen

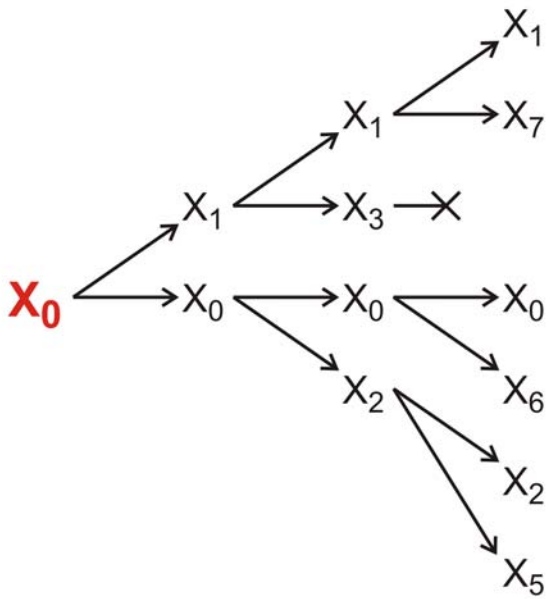
Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



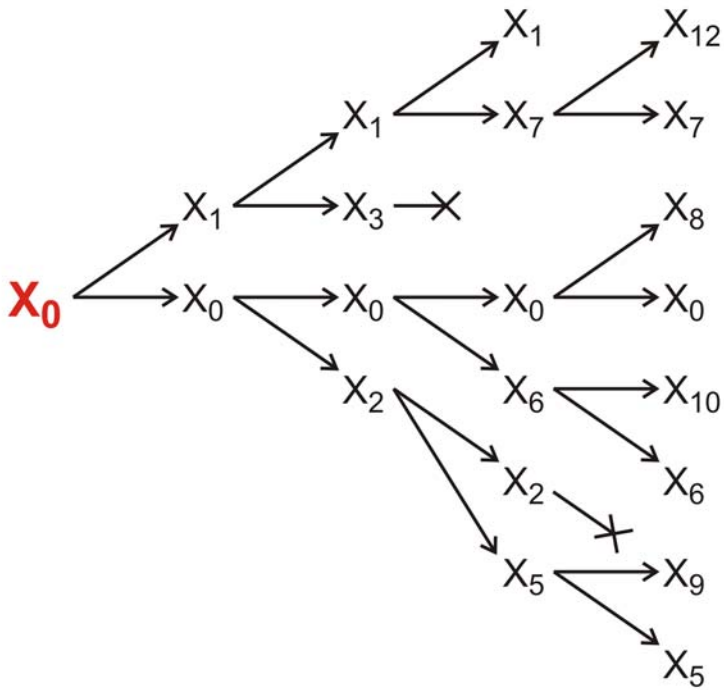
Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



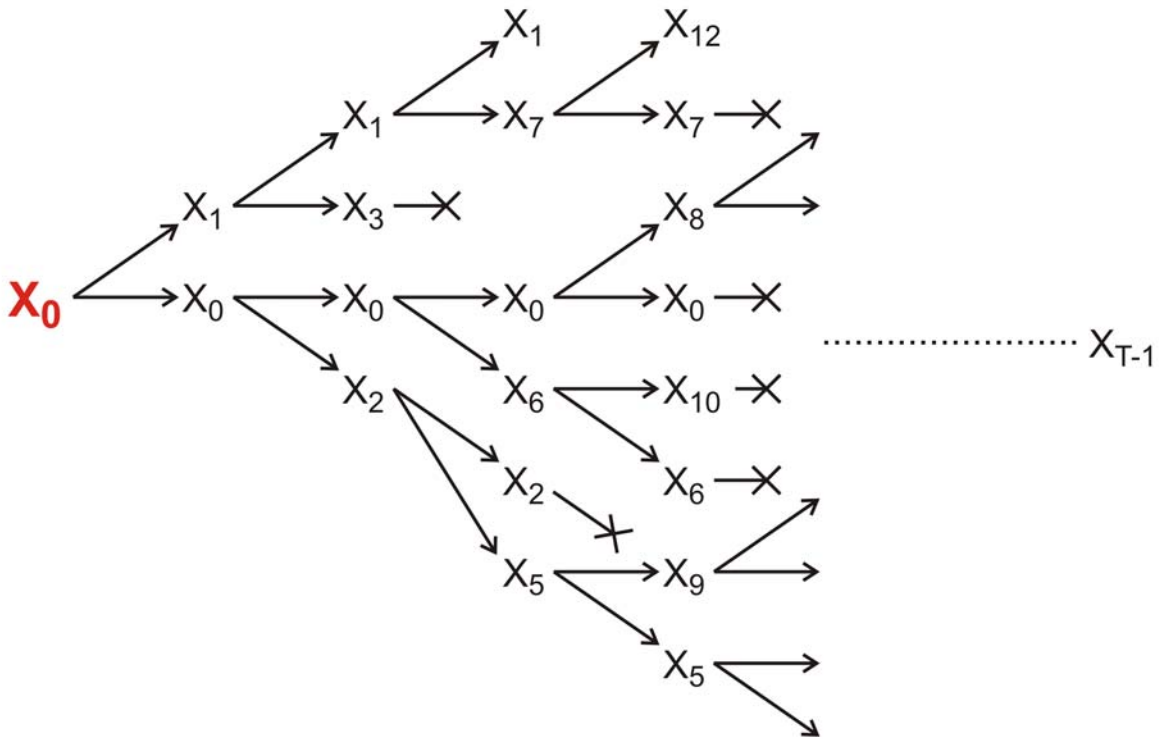
Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



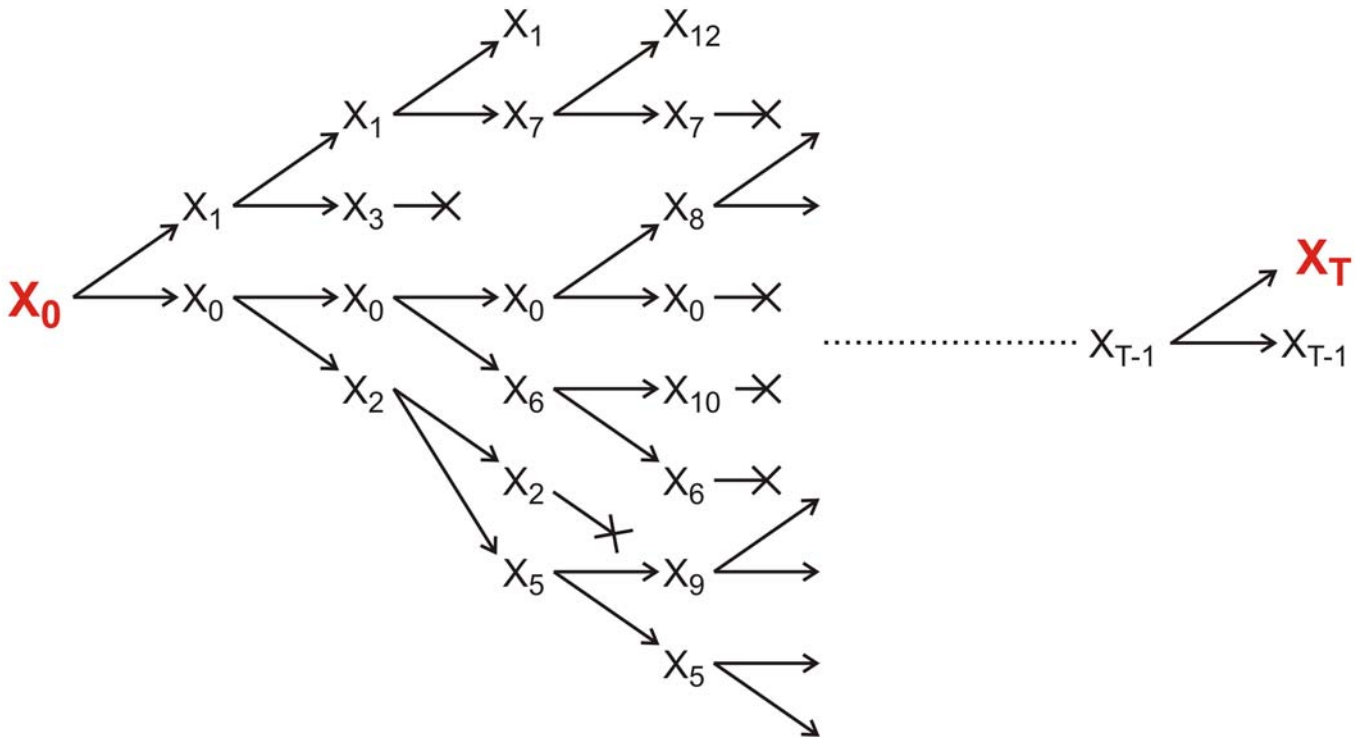
Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen



Evolution von RNA-Molekülen als Markovprozess: Modellierung mit Mastergleichungen

Biophysical Chemistry 26 (1987) 123–147
Elsevier

BPC 01133

A computer model of evolutionary optimization

Walter Fontana and Peter Schuster

Institut für theoretische Chemie und Strahlenchemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

Accepted 27 February 1987

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

PHYSICAL REVIEW A VOLUME 40, NUMBER 6 SEPTEMBER 15, 1989

Physical aspects of evolutionary optimization and adaptation

Walter Fontana, Wolfgang Schnabl, and Peter Schuster*

Institut für Theoretische Chemie der Universität Wien, Währingerstrasse 17, A 1090 Wien, Austria

(Received 2 February 1989; revised manuscript received 5 May 1989)

Computersimulation der RNA-Evolution im Flussreaktor

Walter Fontana and Peter Schuster, Biophysical Chemistry 26:123-147, 1987 und Walter Fontana, Wolfgang Schnabl, and Peter Schuster, Phys.Rev.A 40:3301-3321, 1989

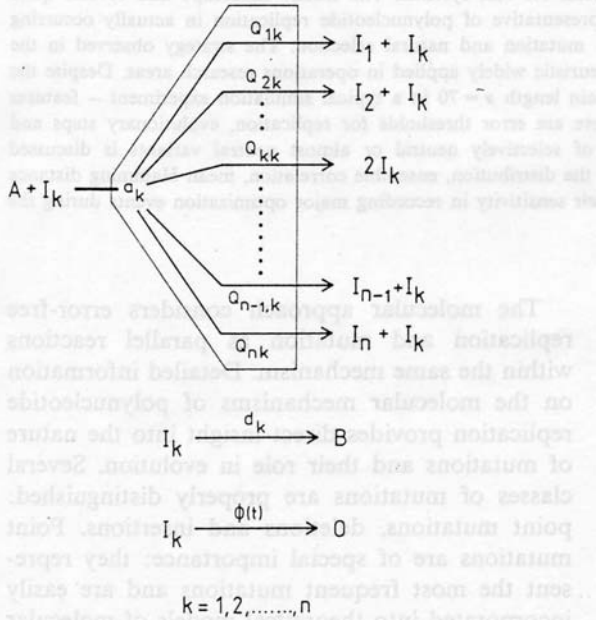


Fig. 1. The reaction network. Synthesis on template I_k proceeds with the rate constant a_k and leads with frequency Q_{ik} to a new template I_i preserving the old copy. Materials A needed for polymerization are assumed to be buffered. Degradation to waste products B occurs with rate d_k and a controlled unspecific flux $\Phi(t)$ removes templates from the system.

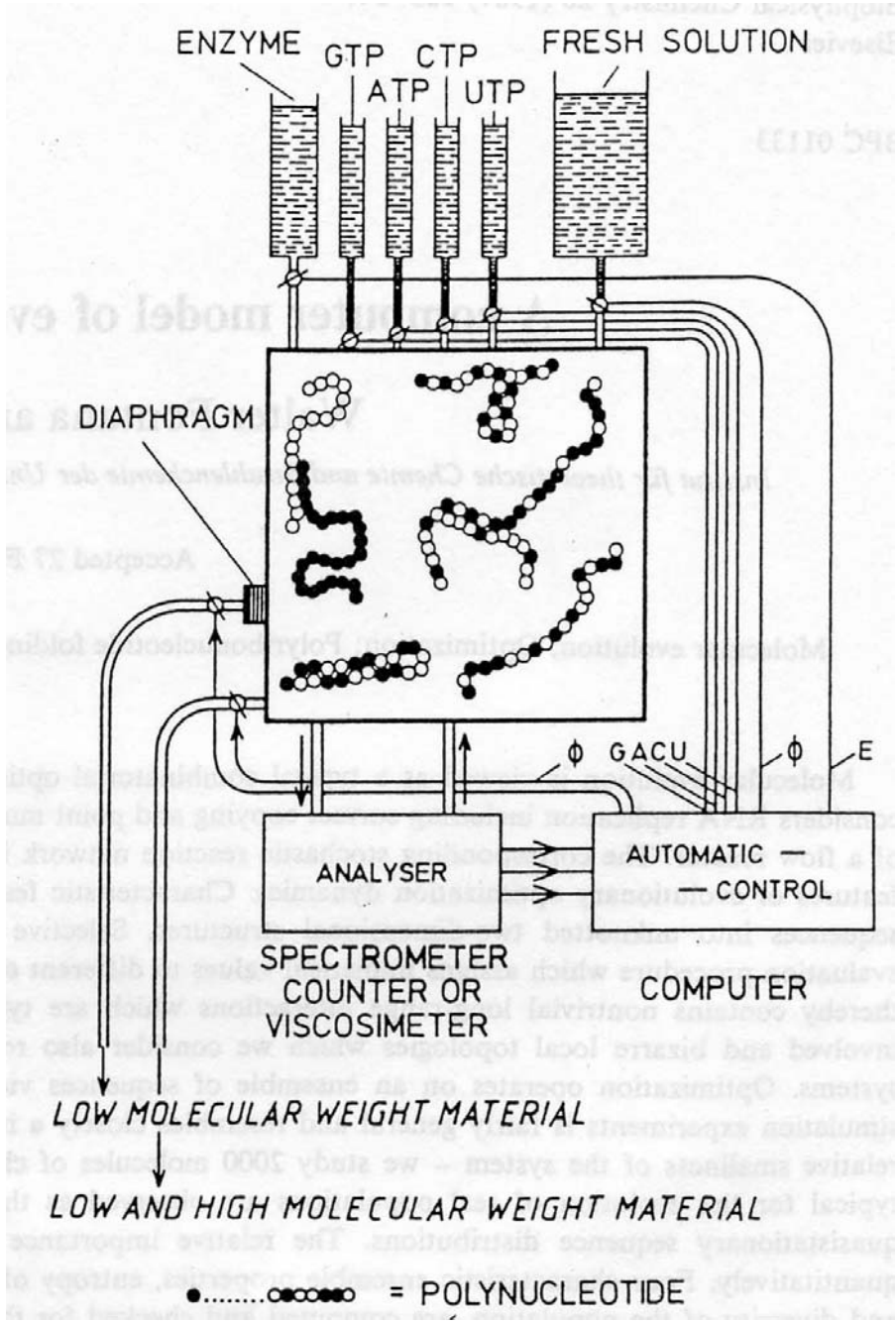
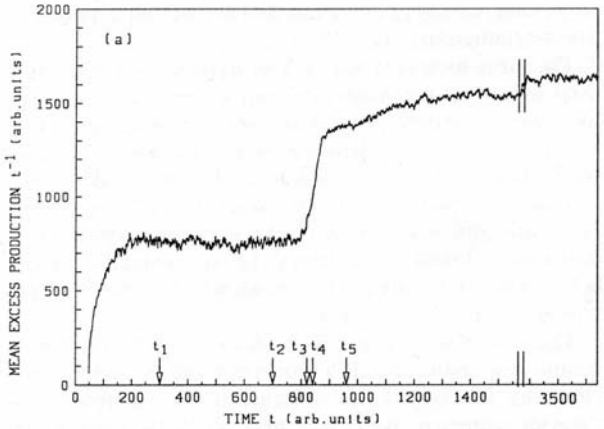
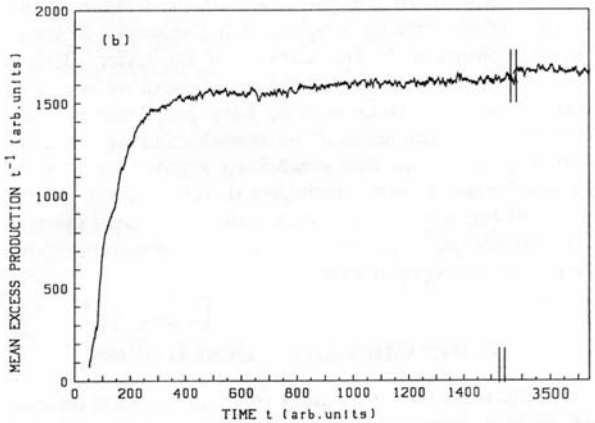
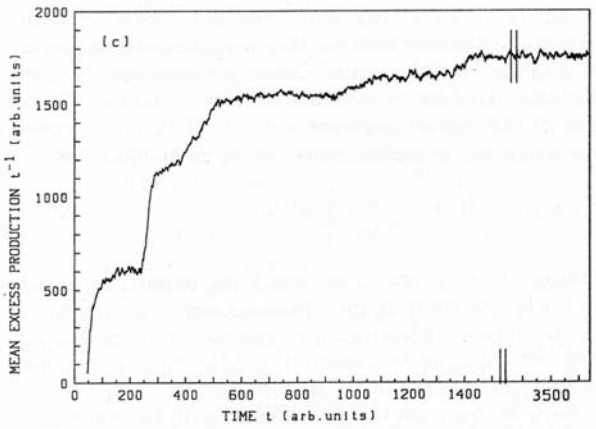


Fig. 2. The evolution reactor. This kind of flow reactor consists of a reaction vessel which allows for temperature and pressure control. Its walls are impermeable to polynucleotides. Energy-rich material is poured from the environment into the reactor. The degradation products are removed steadily. Material transport is adjusted in such a way that the concentration of monomers is constant in the reactor. A dilution flux Φ is installed in order to remove excess of polynucleotides produced by replication. Thus, the sum of the numbers of individual particles $\sum_i X_i(t) = N(t)$ may be controlled by the flux Φ. Under 'constant organization' Φ is adjusted such that $N(t) = \Theta$ is essentially constant. By this we indicate that fluctuations with standard deviation $\sigma = \sqrt{N}$ occur regularly. The regulation of Φ requires internal control, which can be achieved by logistic coupling.

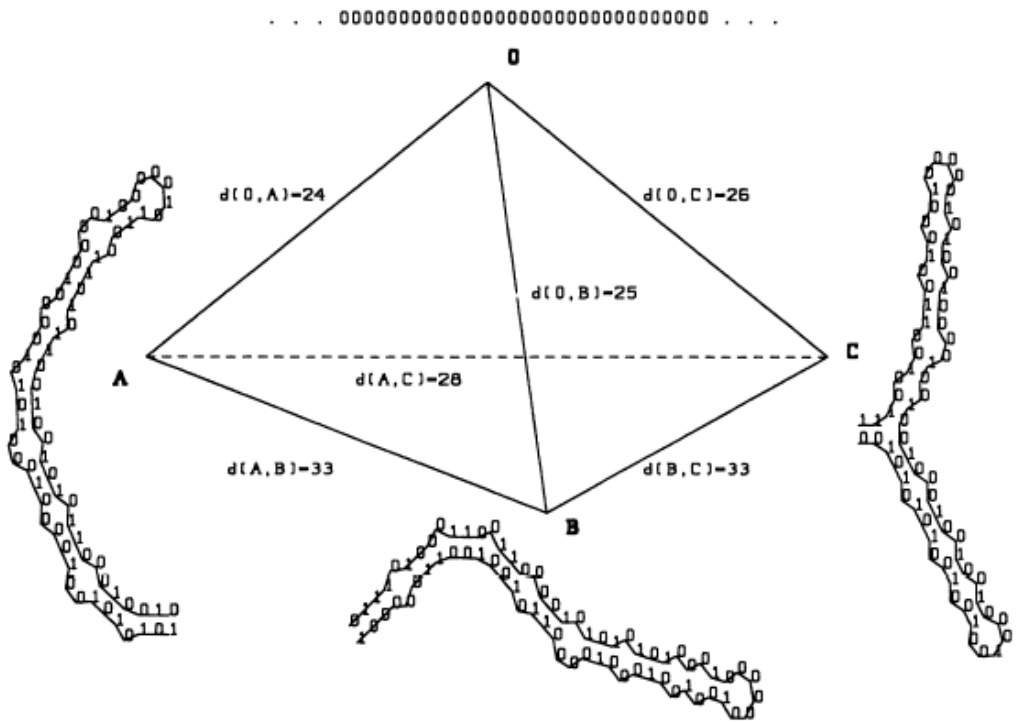






Optimizierung der Differenz von Replikations- und Abbaurate:

$$(f_i - d_i) x_i$$



Walter Fontana, Wolfgang Schnabl, and Peter Schuster, Phys.Rev.A 40:3301-3321, 1989

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.

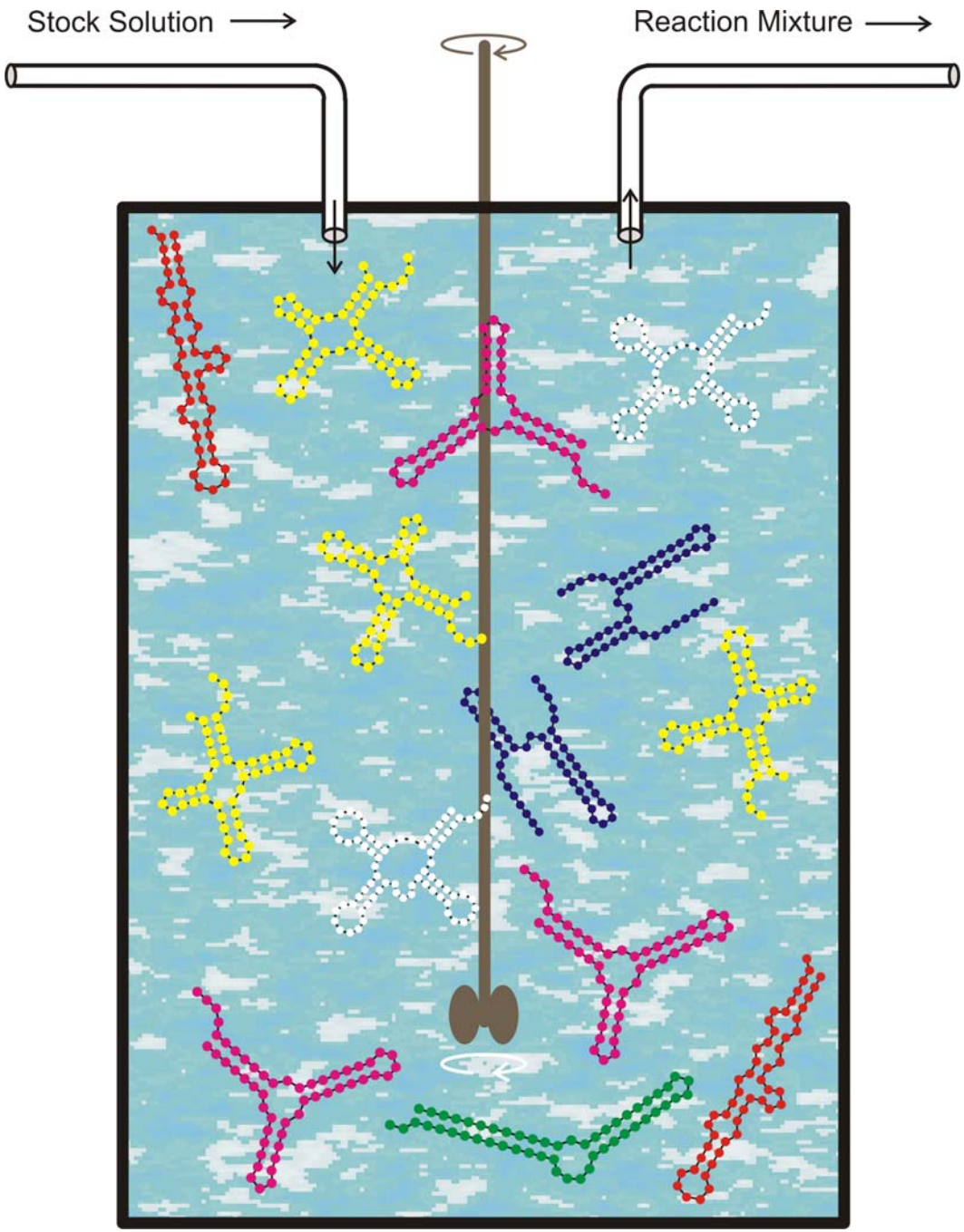
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level 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.



Replikationsparameter:

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

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

Selektionsbedingung:

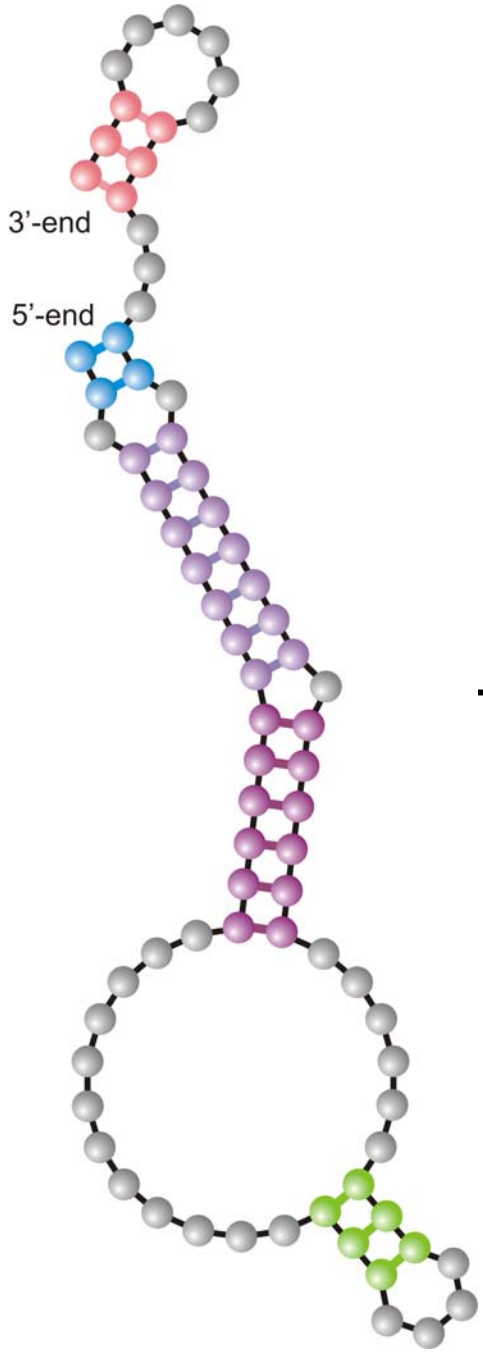
Die Populationsgröße, N = # RNA-Moleküle, wird durch den Fluss kontrolliert:

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

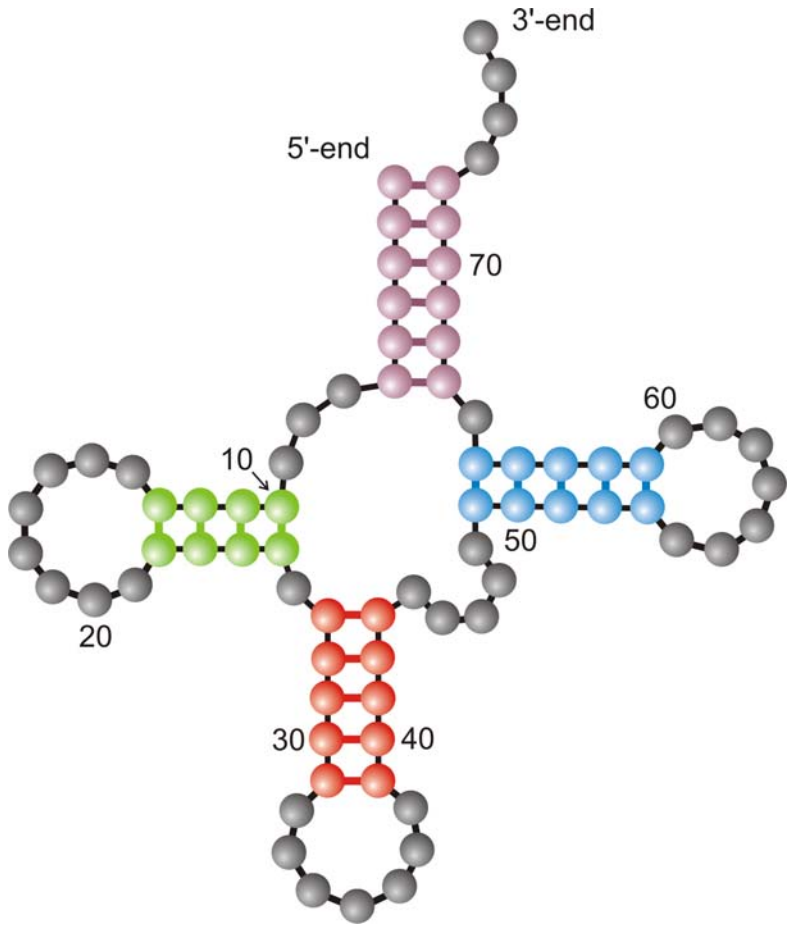
Mutationsrate:

p = 0.001 / Nukleotid × Replikation

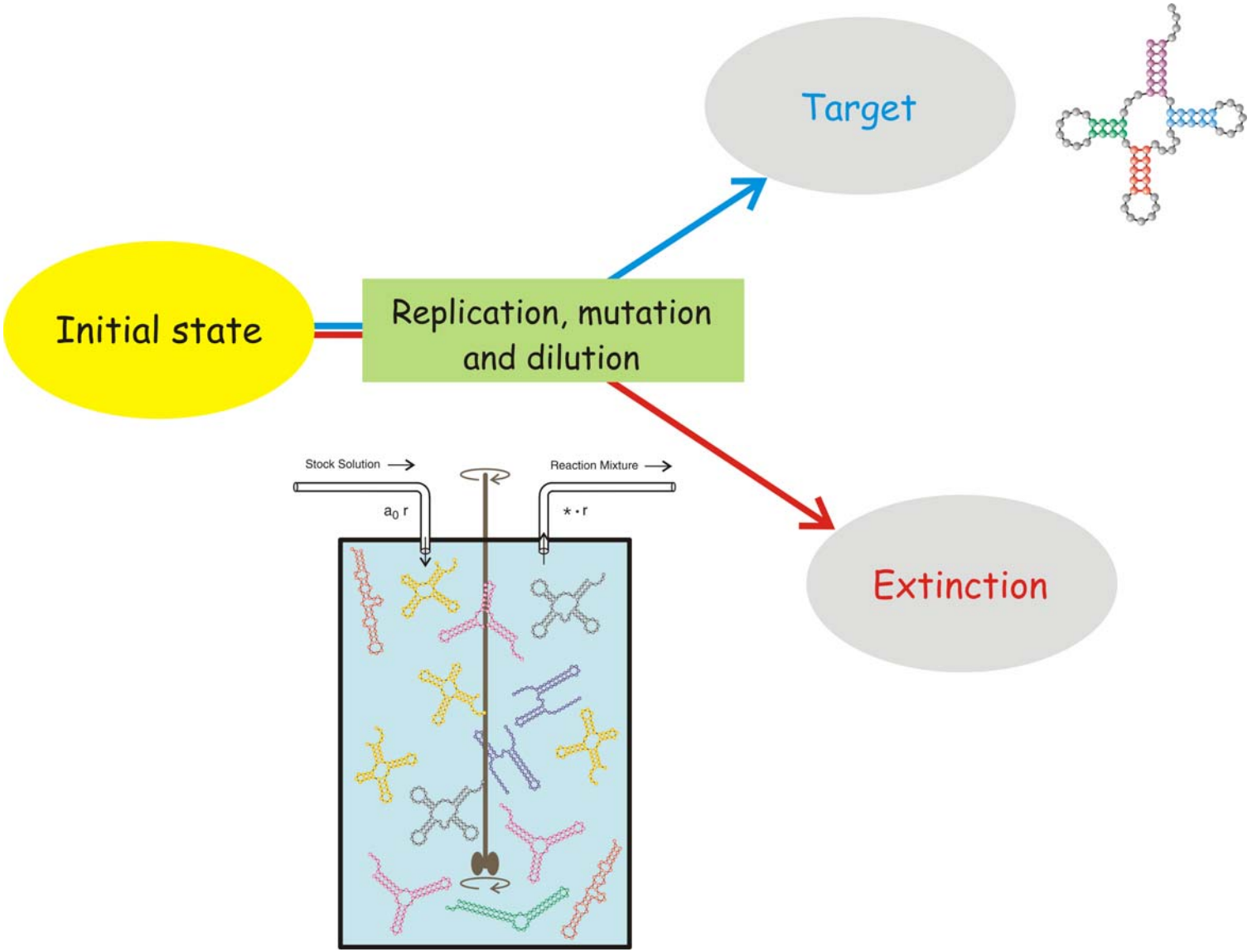
Der Flussreaktor zum Studium von Evolution ***in vitro*** and ***in silico***



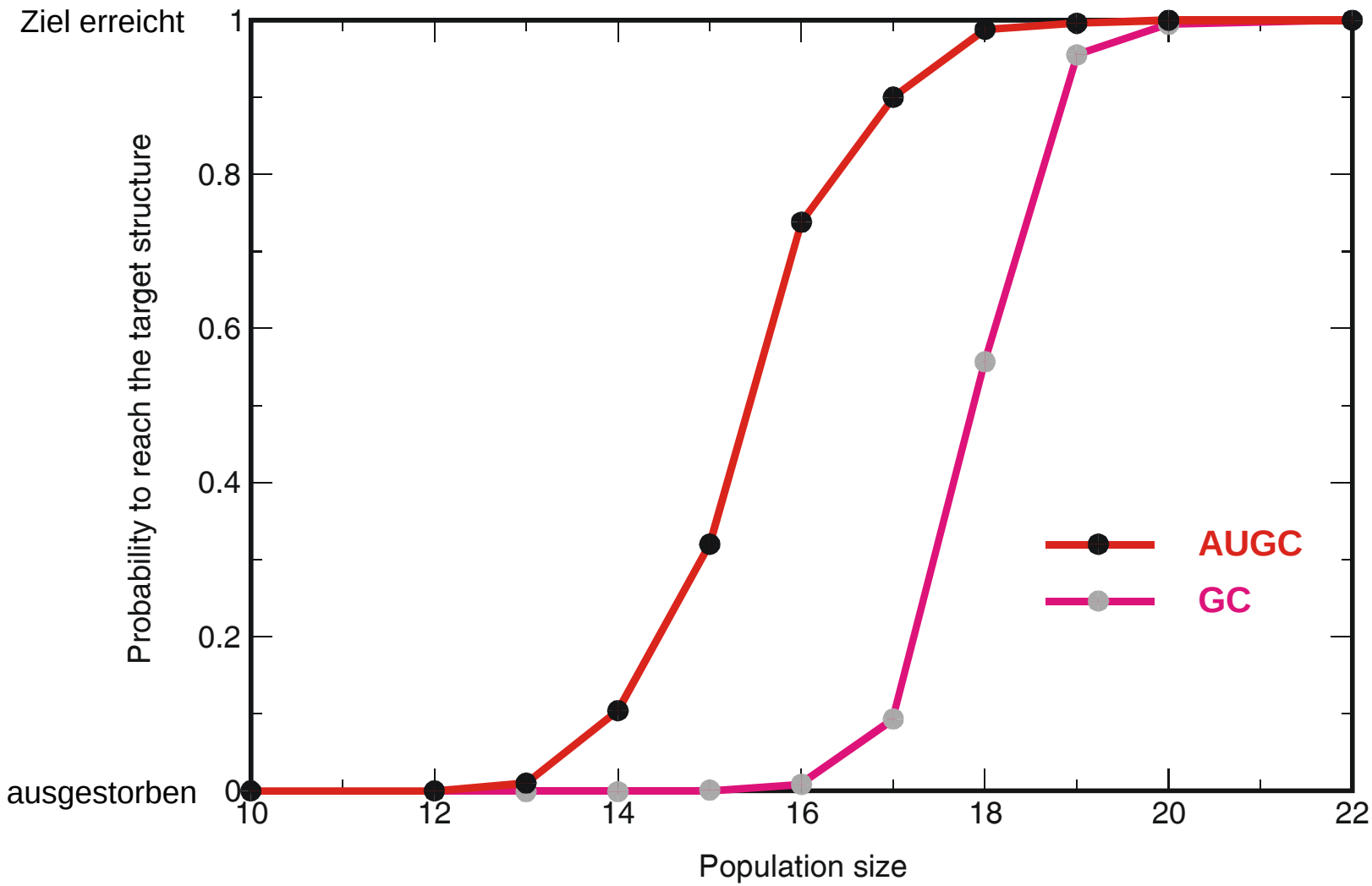
Struktur einer
zufällig gewählten
Anfangssequenz



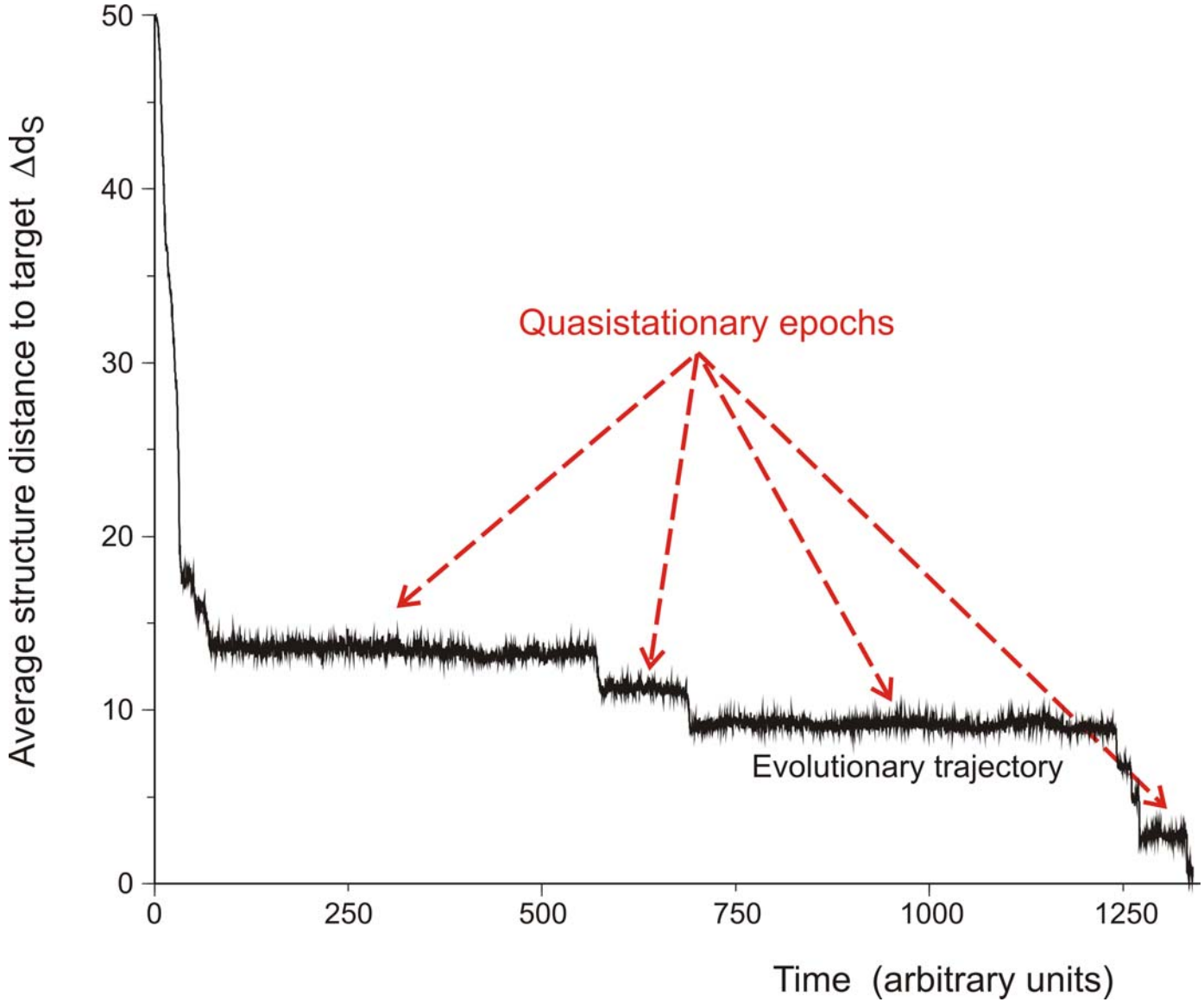
Phenylalanyl-tRNA
als Zielstruktur



Die stochastische Replikation im Flussreaktor hat eine positive Aussterbewahrscheinlichkeit.

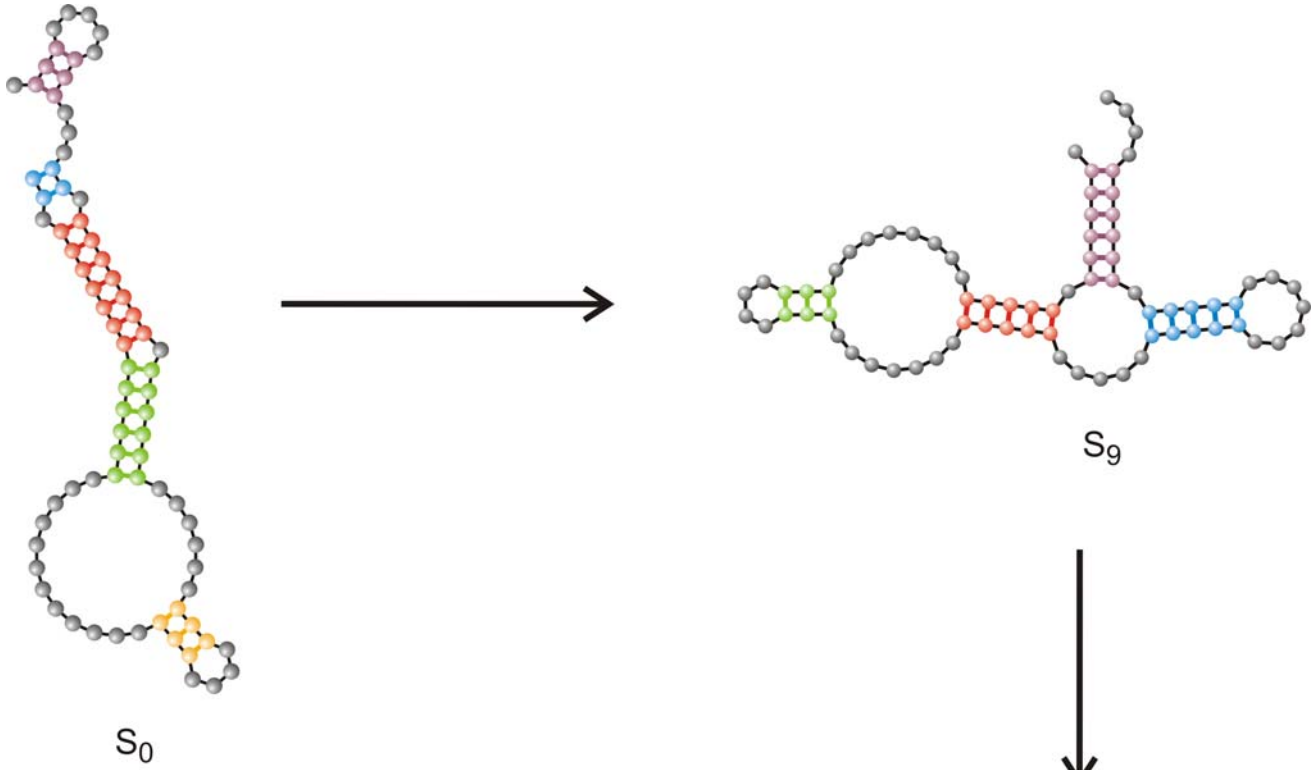


Die Wahrscheinlichkeit, dass eine einzelne Trajektorie das Ziel erreicht



In silico Strukturoptimierung im Flussreaktor: eine Trajektorie des Evolutionsprozesses

Zufällig gewählte
Anfangsstruktur



Phenylalanyl-tRNA
als Zielstruktur

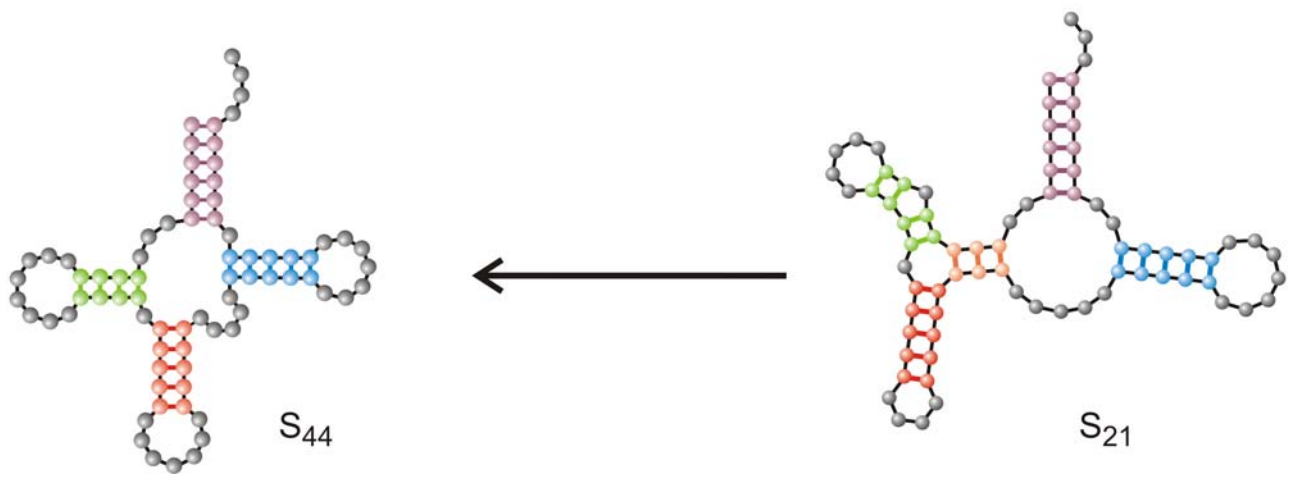


Table 8. Statistics of the optimization trajectories. The table shows the results of sampled evolutionary trajectories leading from a random initial structure, S_I, to the structure of $tRNA^{phe}$, S_T, as the target[a]. Simulations were performed with an algorithm introduced by Gillespie [55–57]. The time unit is here undefined. A mutation rate of $p = 0.001$ per site and replication were used. The mean and standard deviation were calculated under the assumption of a log-normal distribution that fits well the data of the simulations.

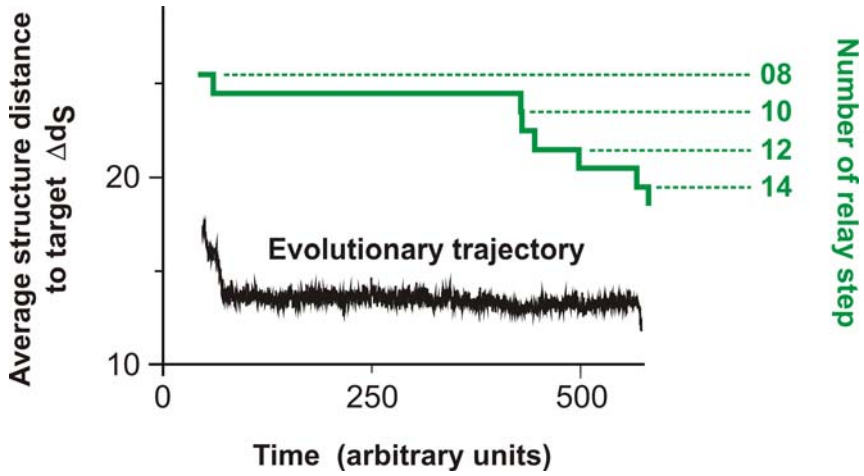
Alphabet	Population size, N	Number of runs, n_R	Real time from start to target		Number of replications $[10^7]$	
			Mean value	σ	Mean value	σ
AUGC	1 000	120	900	+1380 −542	1.2	+3.1 −0.9
	2 000	120	530	+880 −330	1.4	+3.6 −1.0
	3 000	1199	400	+670 −250	1.6	+4.4 −1.2
	10 000	120	190	+230 −100	2.3	+5.3 −1.6
	30 000	63	110	+97 −52	3.6	+6.7 −2.3
	100 000	18	62	+50 −28	–	–
GC	1 000	46	5160	+15700 −3890	–	–
	3 000	278	1910	+5180 −1460	7.4	+35.8 −6.1
	10 000	40	560	+1620 −420	–	–

[a] The structures S_I and S_T were used in the optimization:

S_I: ((.(((((((((((((.....((((....)))).....))))))).)))))))..))...(((.....)))
S_T: ((((((...((((........)))).((((.......)))))).....((((.......)))))..))))))....

Statistik der RNA-Strukturoptimierung: P. Schuster, Rep.Prog.Phys. 69:1419-1477, 2006

28 neutrale Punktmutationen
während einer langen quasi-stationären Epoche

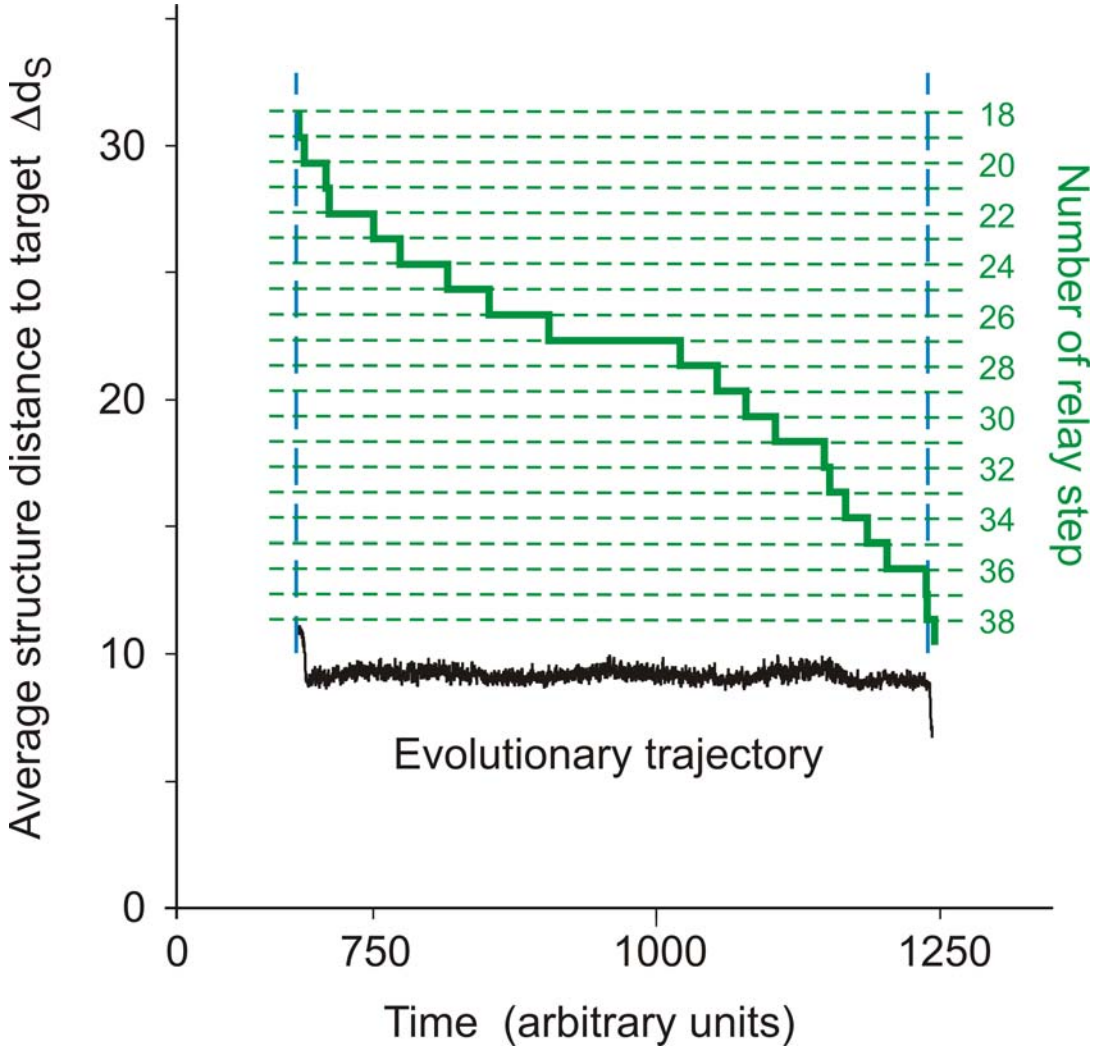



```
entry   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8       .((((((((((((.........(((....)))........)))))).....((((.......))))))))))....
exit    GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCAUACAGAA
entry   GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUACCAUACAGAA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

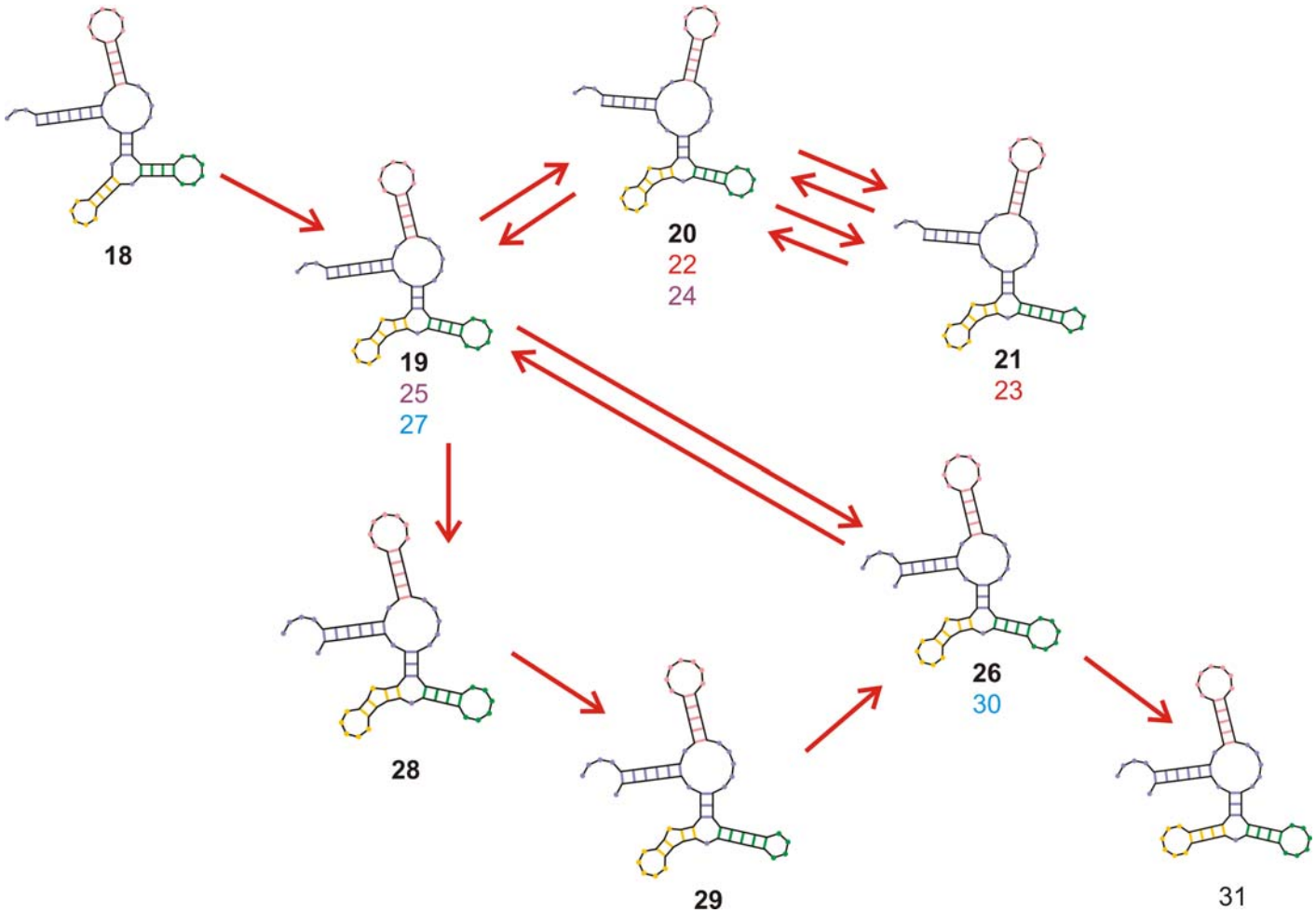
Übergänge induzierende Punktmutationen
ändern die molekulare Struktur

Neutrale Punktmutationen lassen die molekulare Struktur unverändert

Neutrale Evolution von Sequenzen bei konstanter Struktur



Eine quasistationäre Epoche mit wechselnden Strukturen

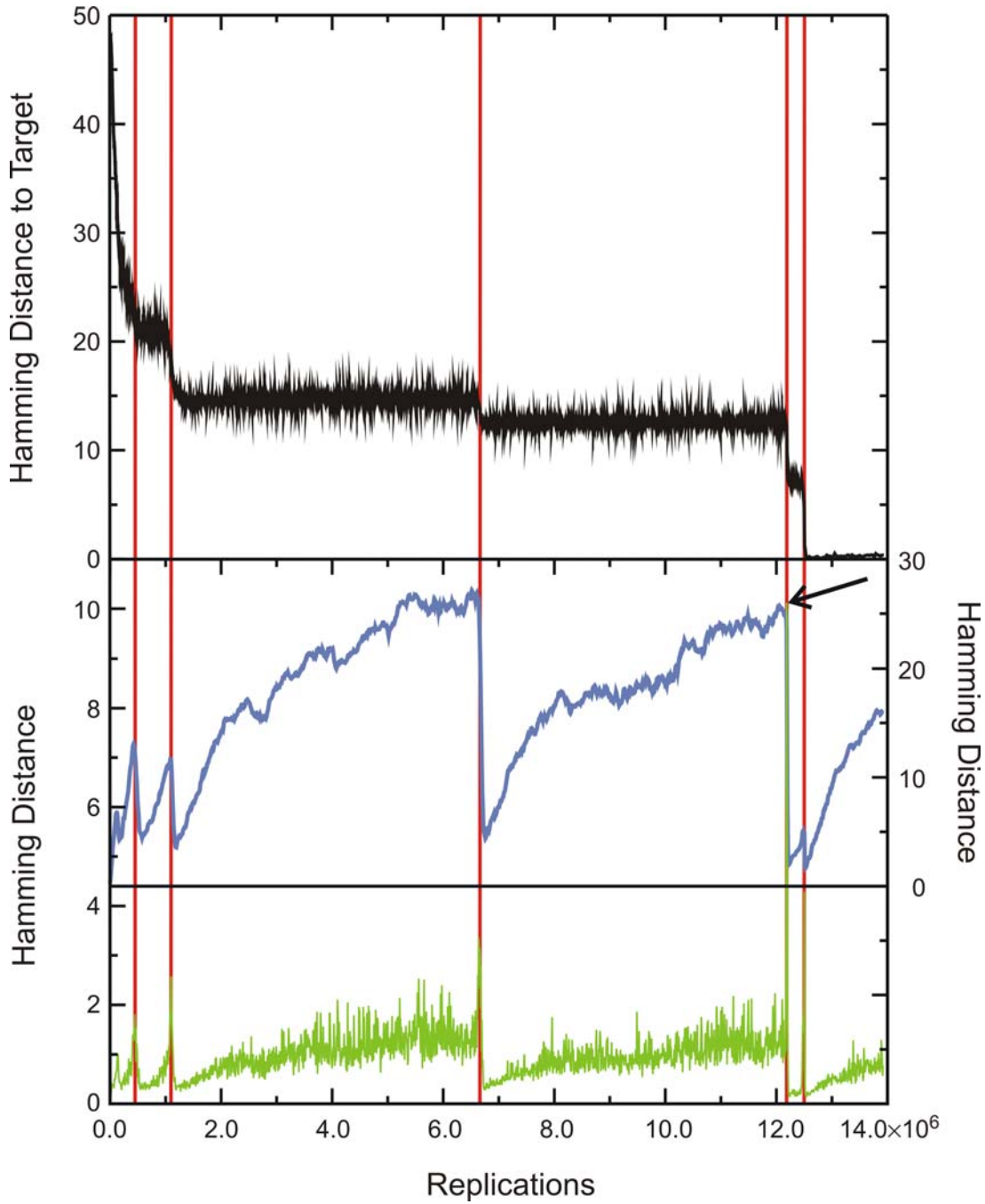


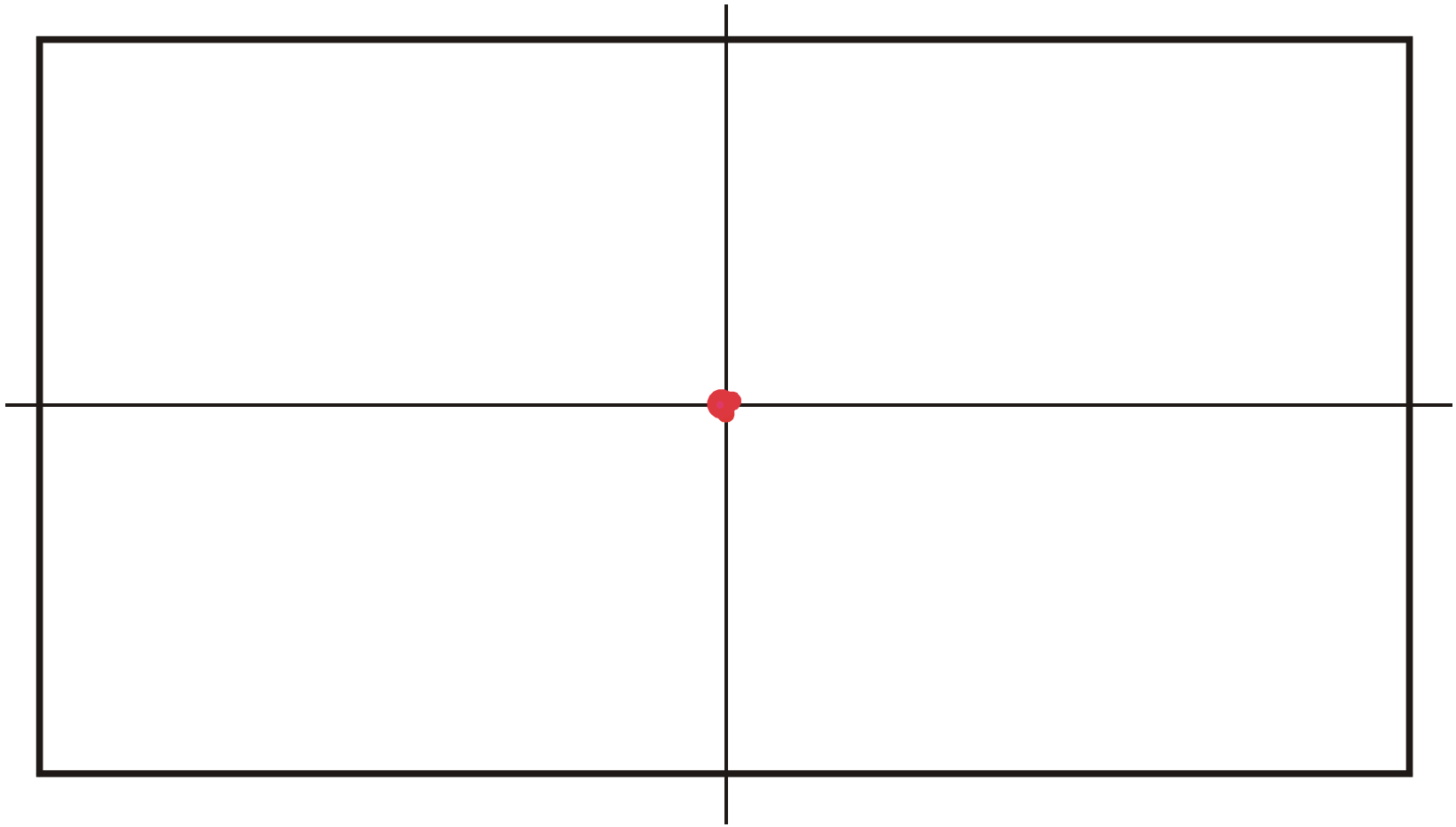
Ein ‚Irrflug‘ im Raum gleichwertiger Strukturen

Trajektorie eines
Evolutionsprozesses

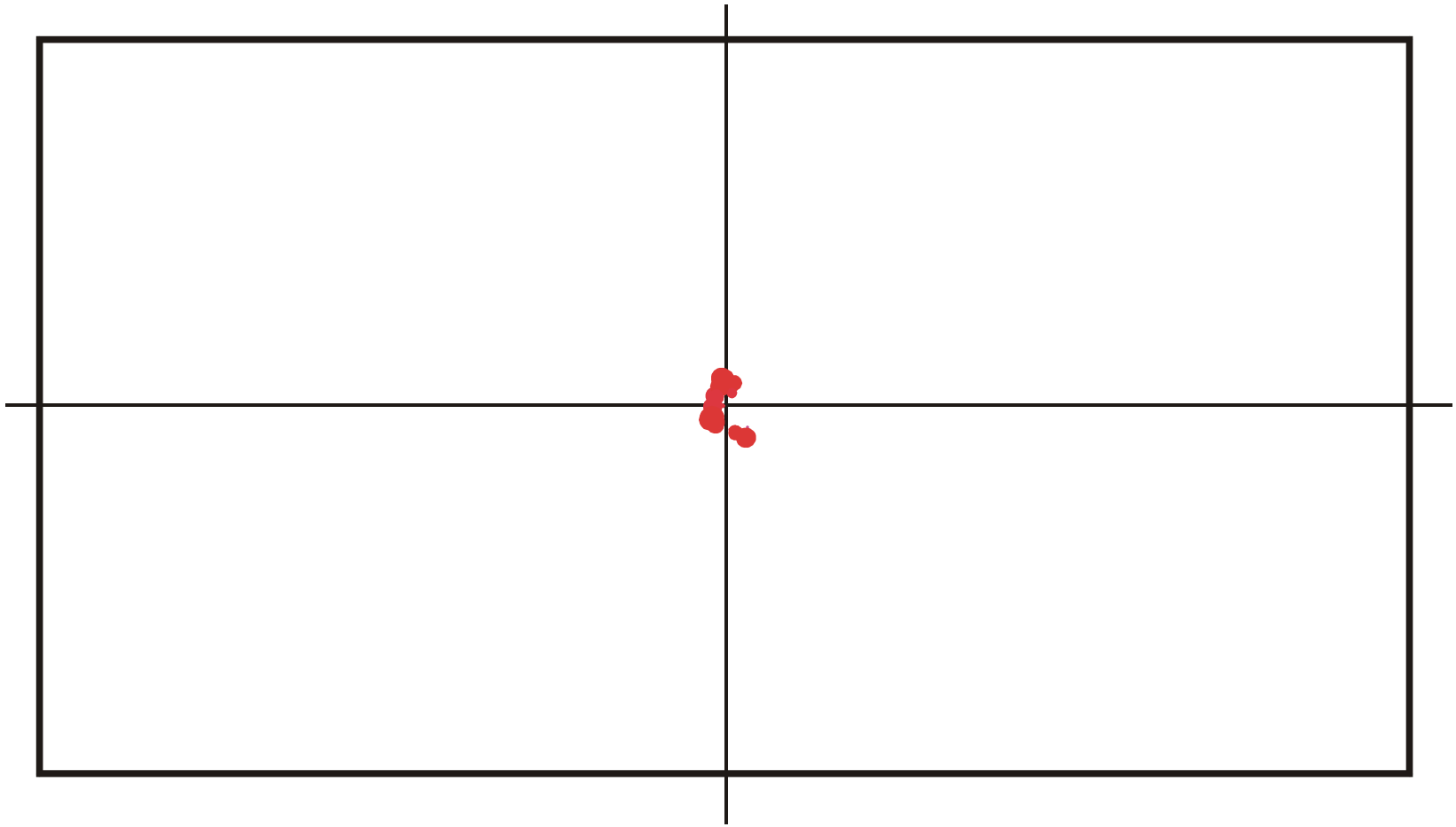
Ausbreitung der Population
auf neutralen Netzwerken

Drift des Populations-
schwerpunktes im
Sequenzraum

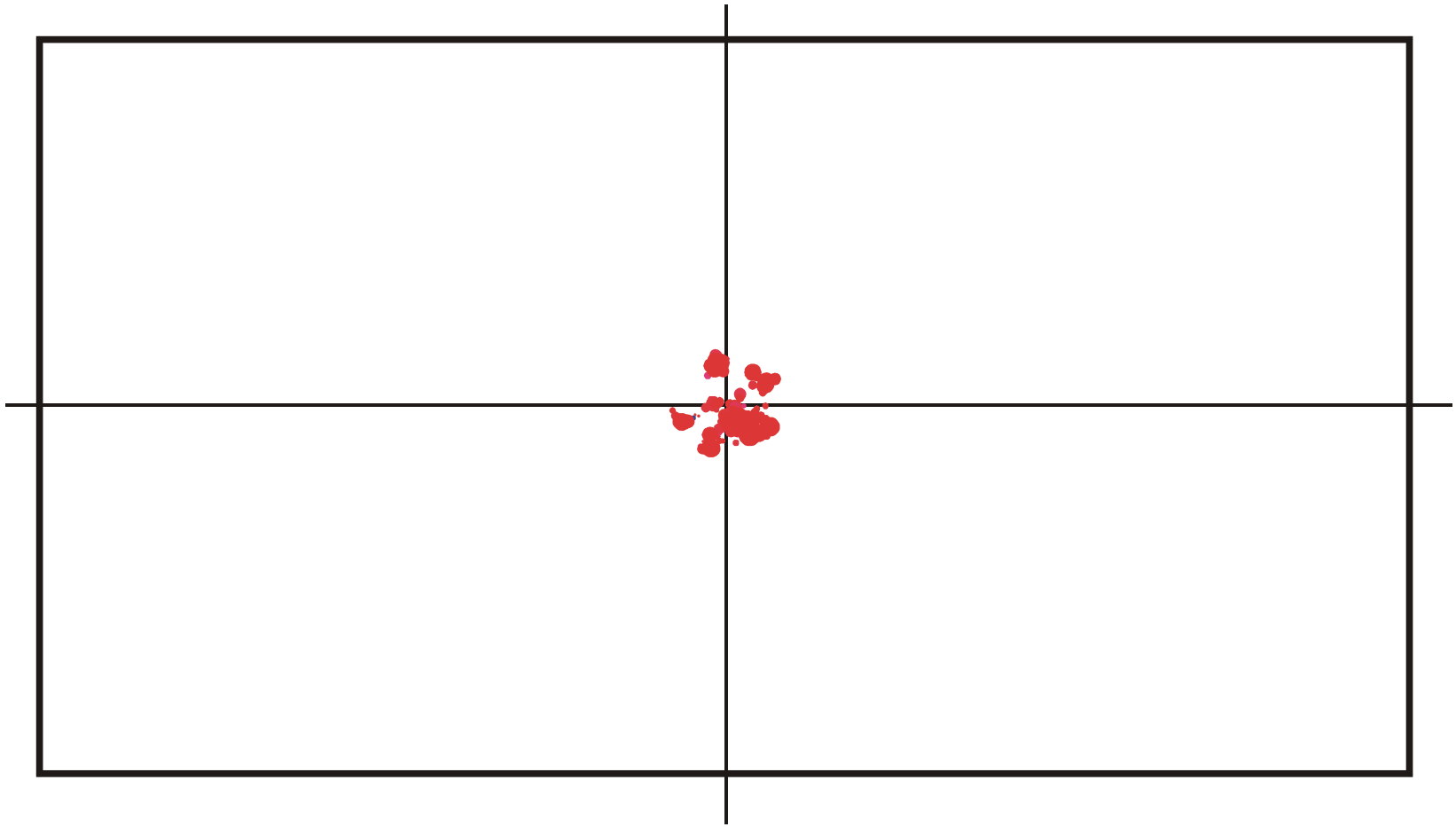




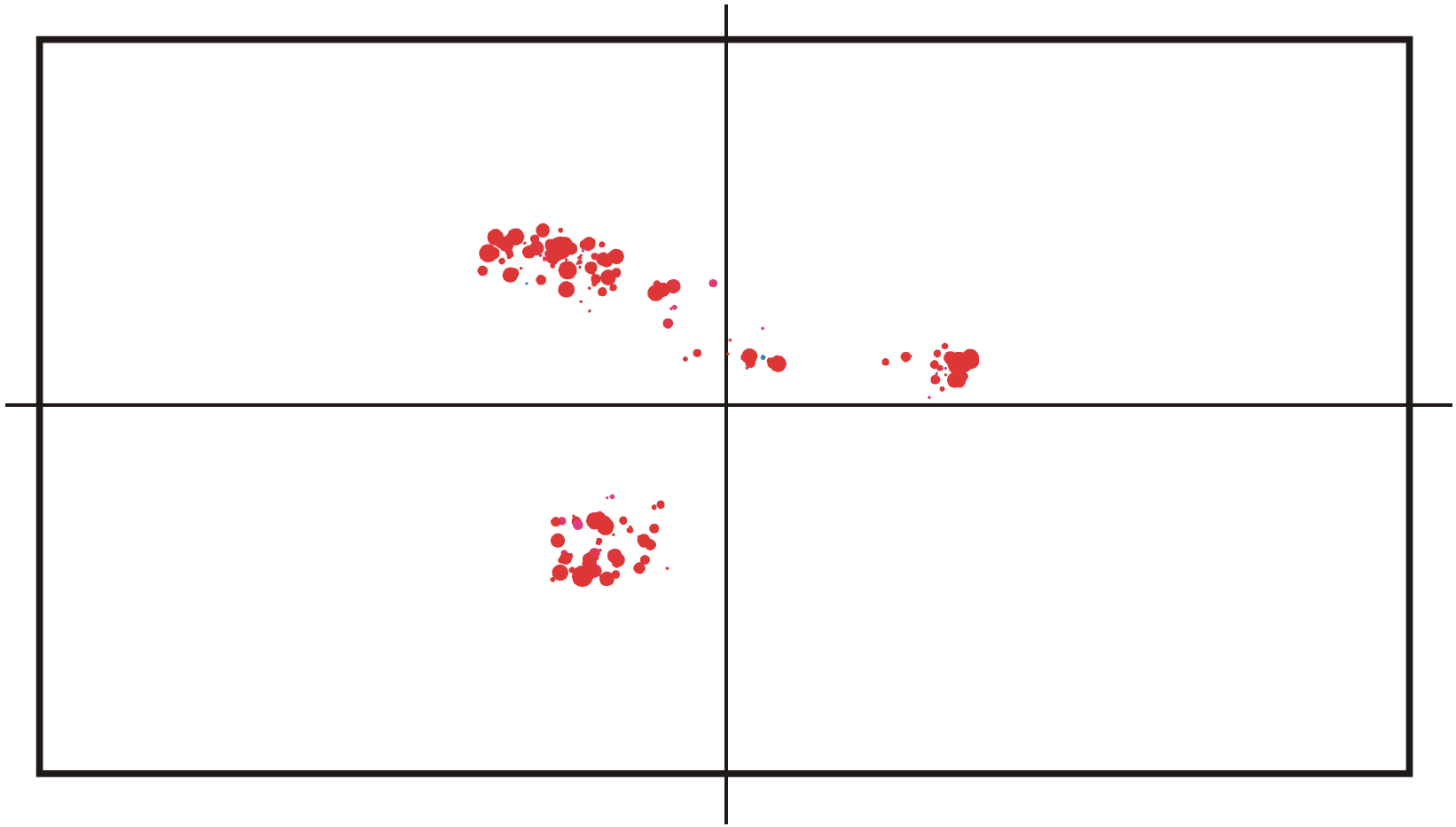
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 150



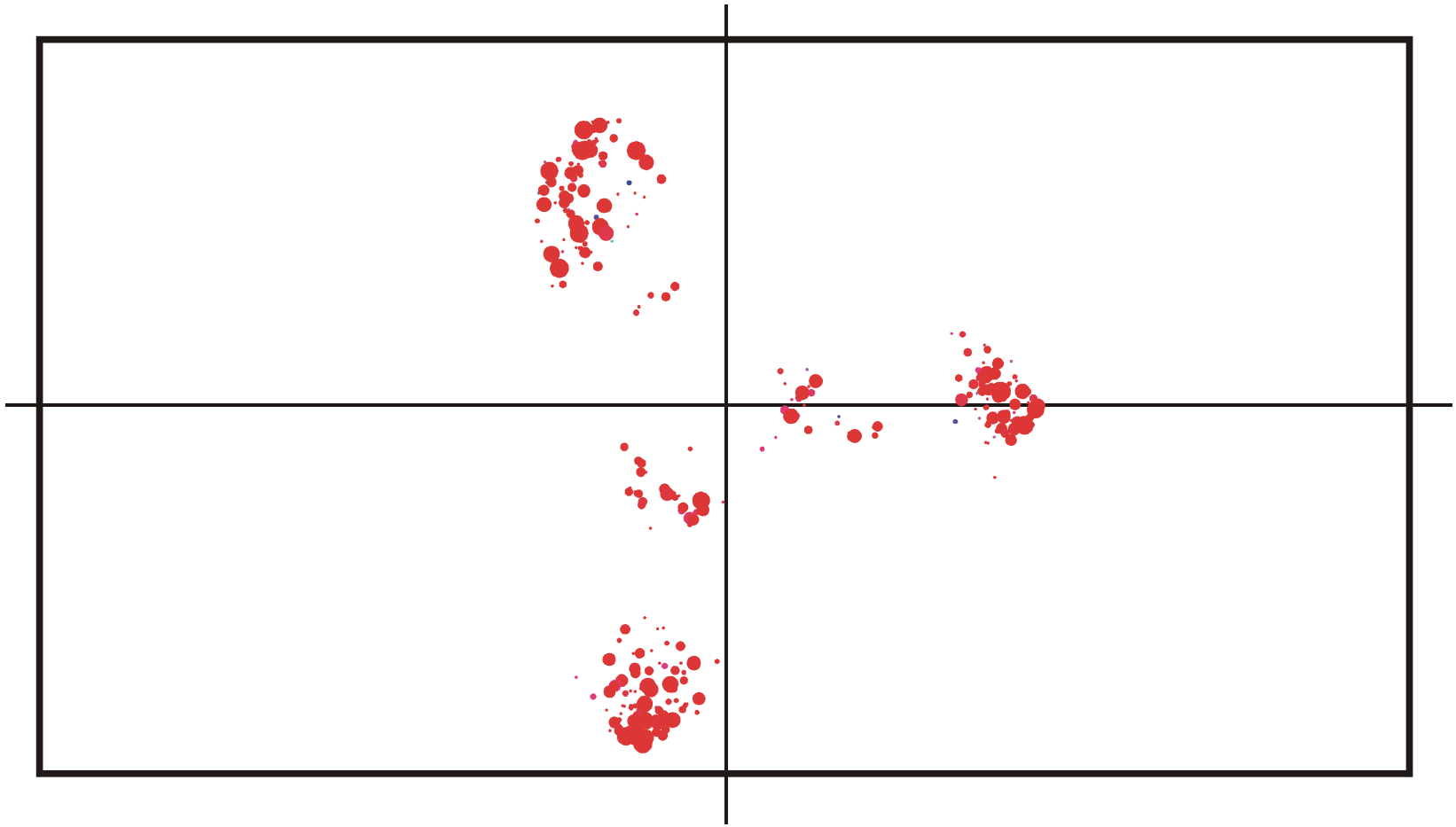
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 170



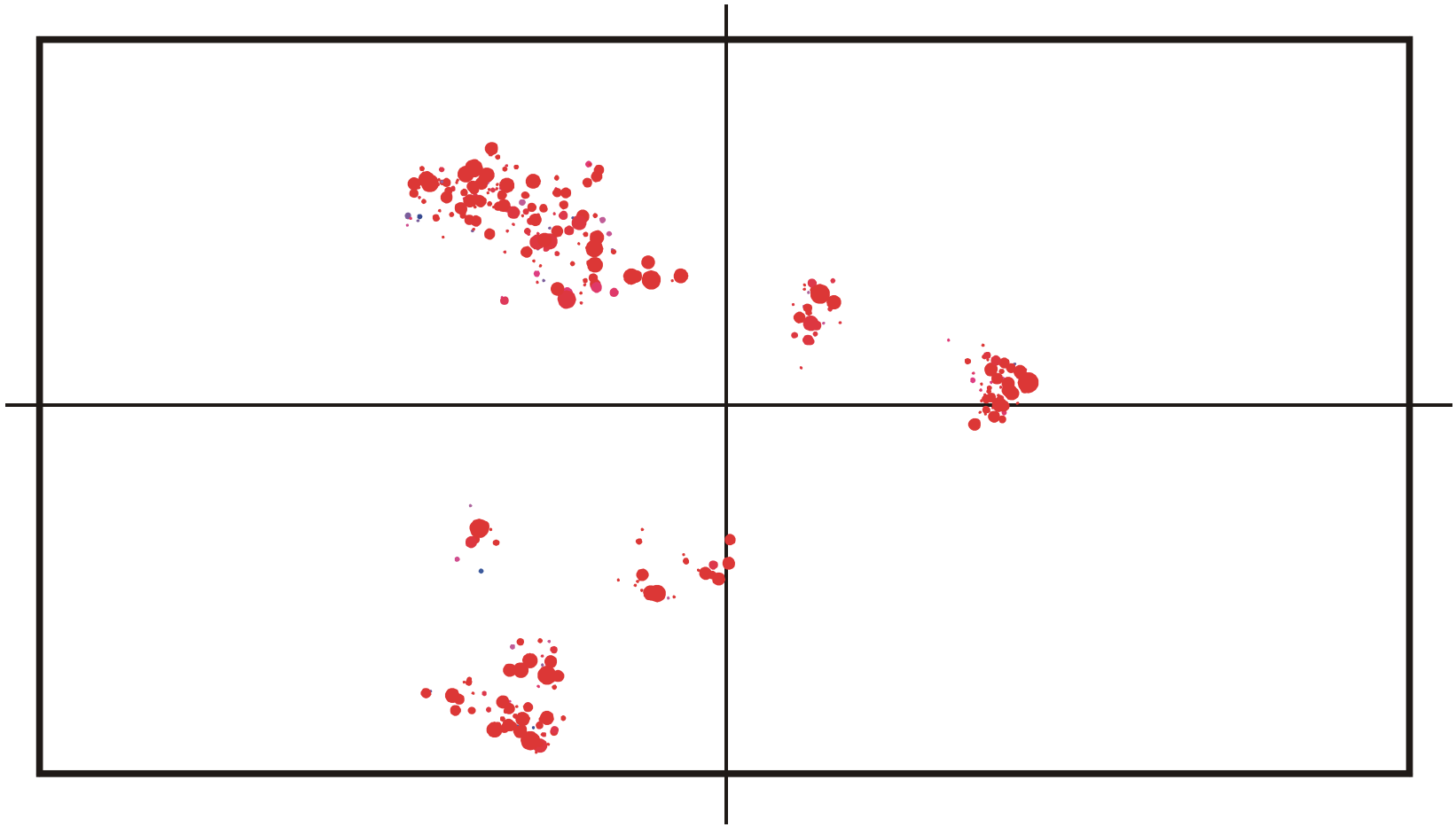
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 200



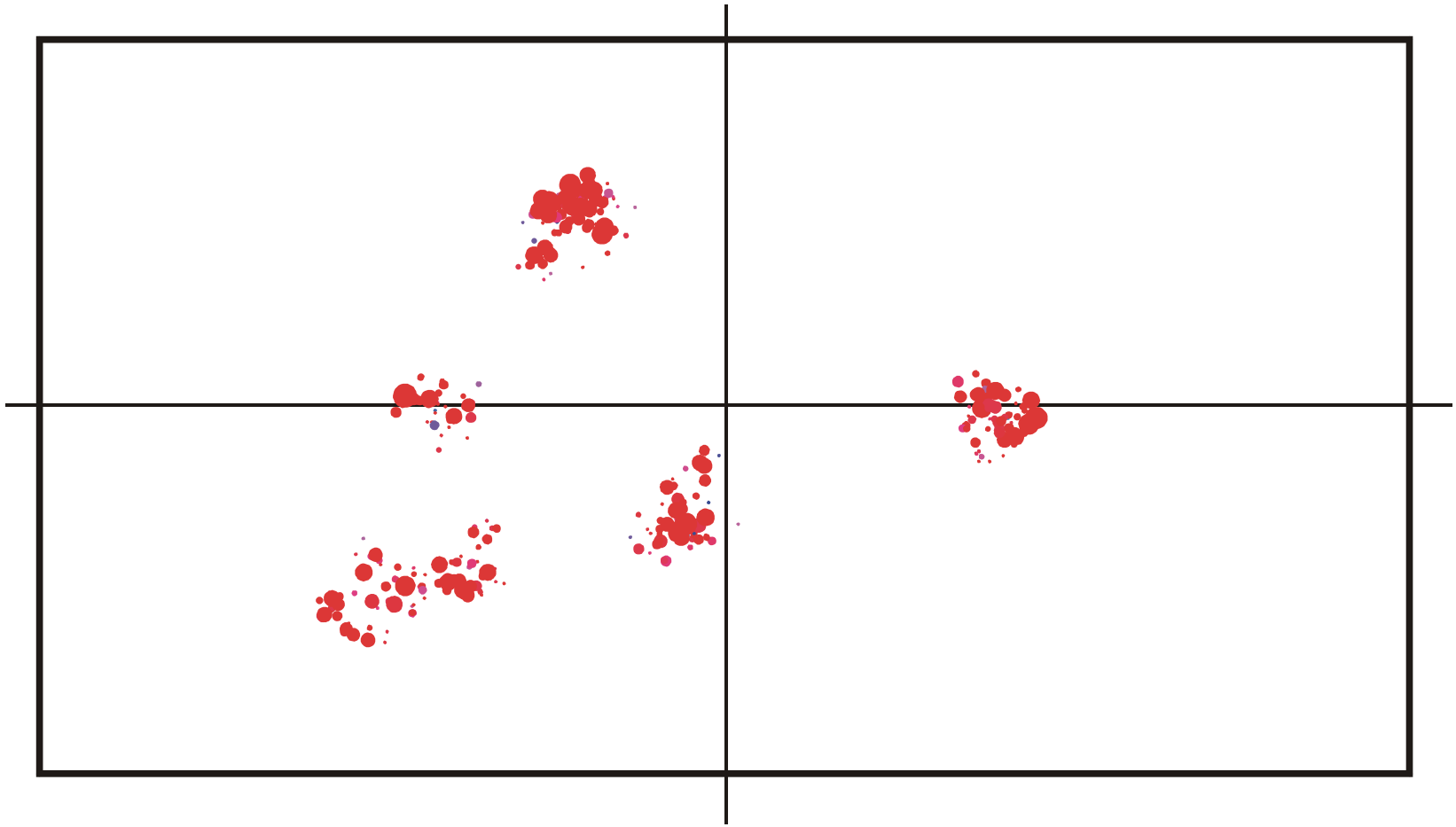
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 350



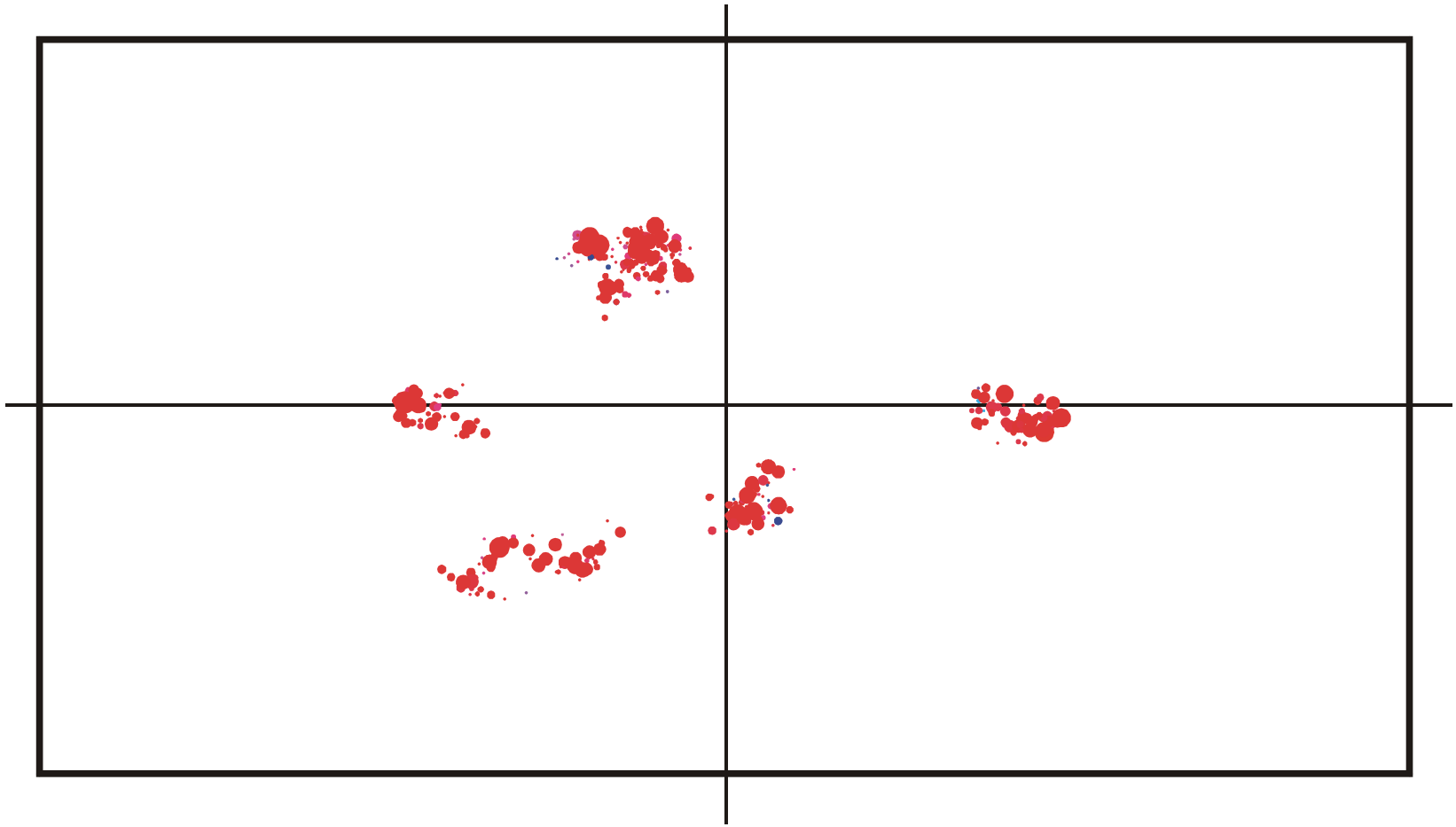
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 500



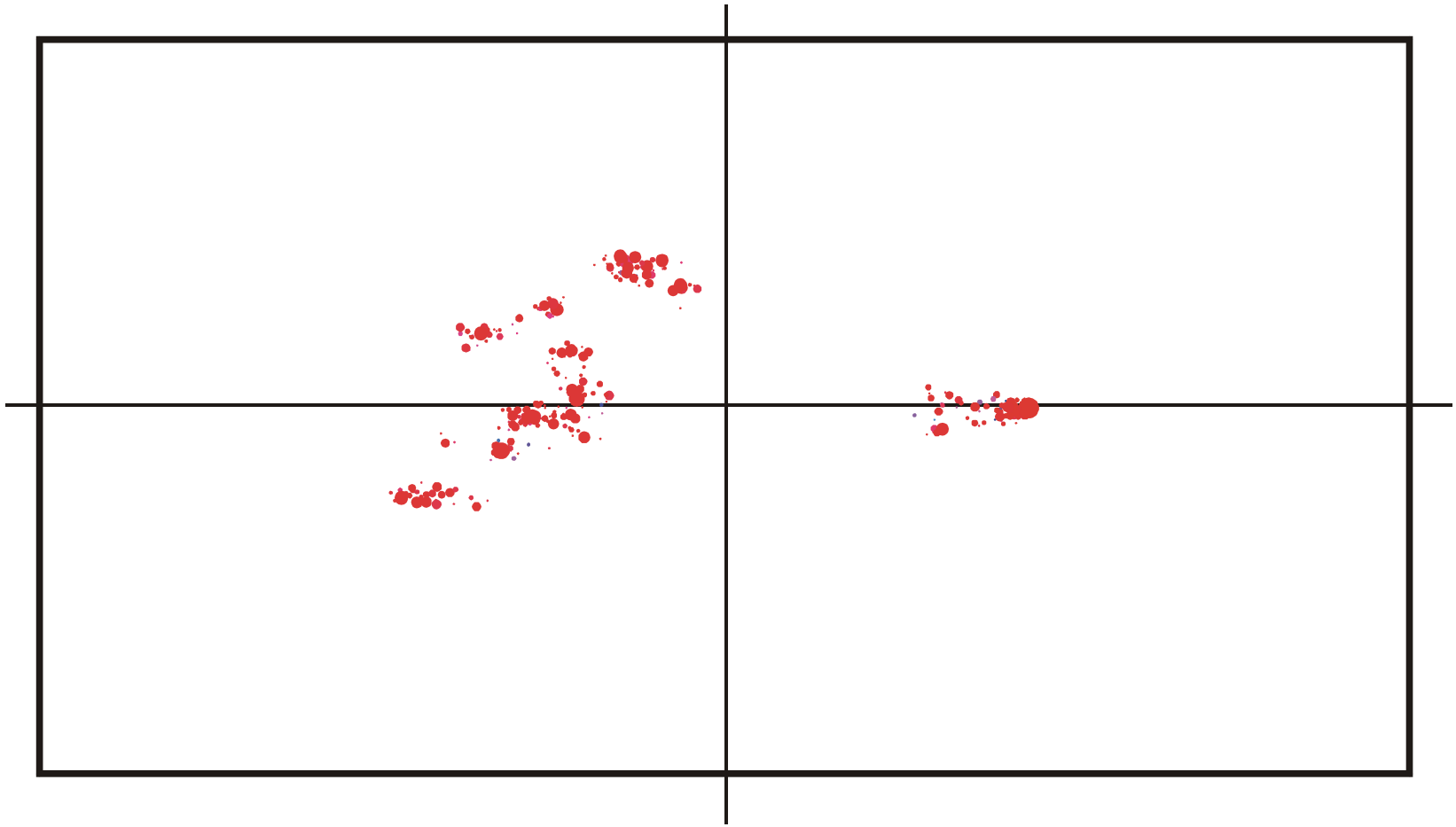
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 650



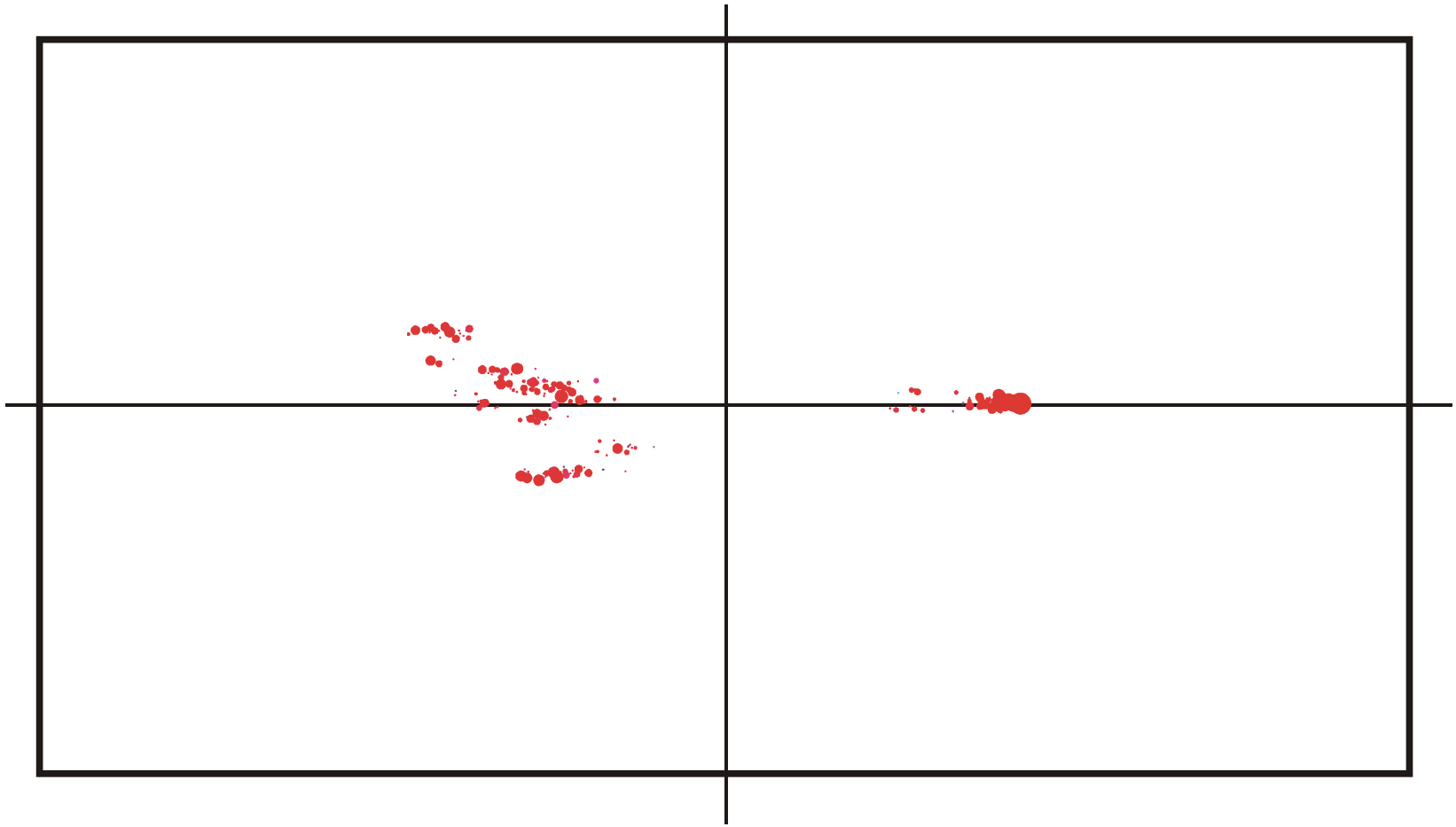
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 820



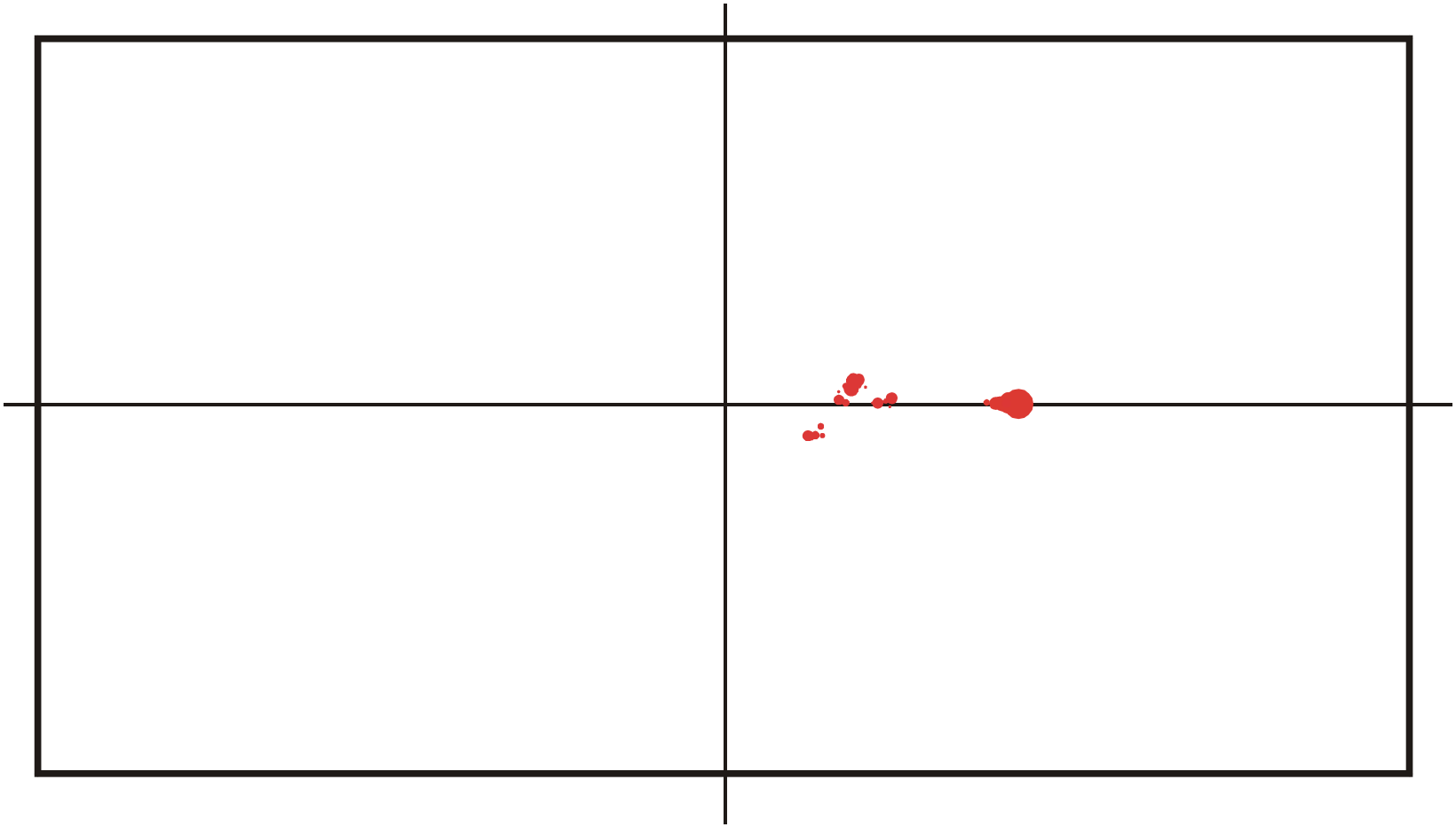
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 825



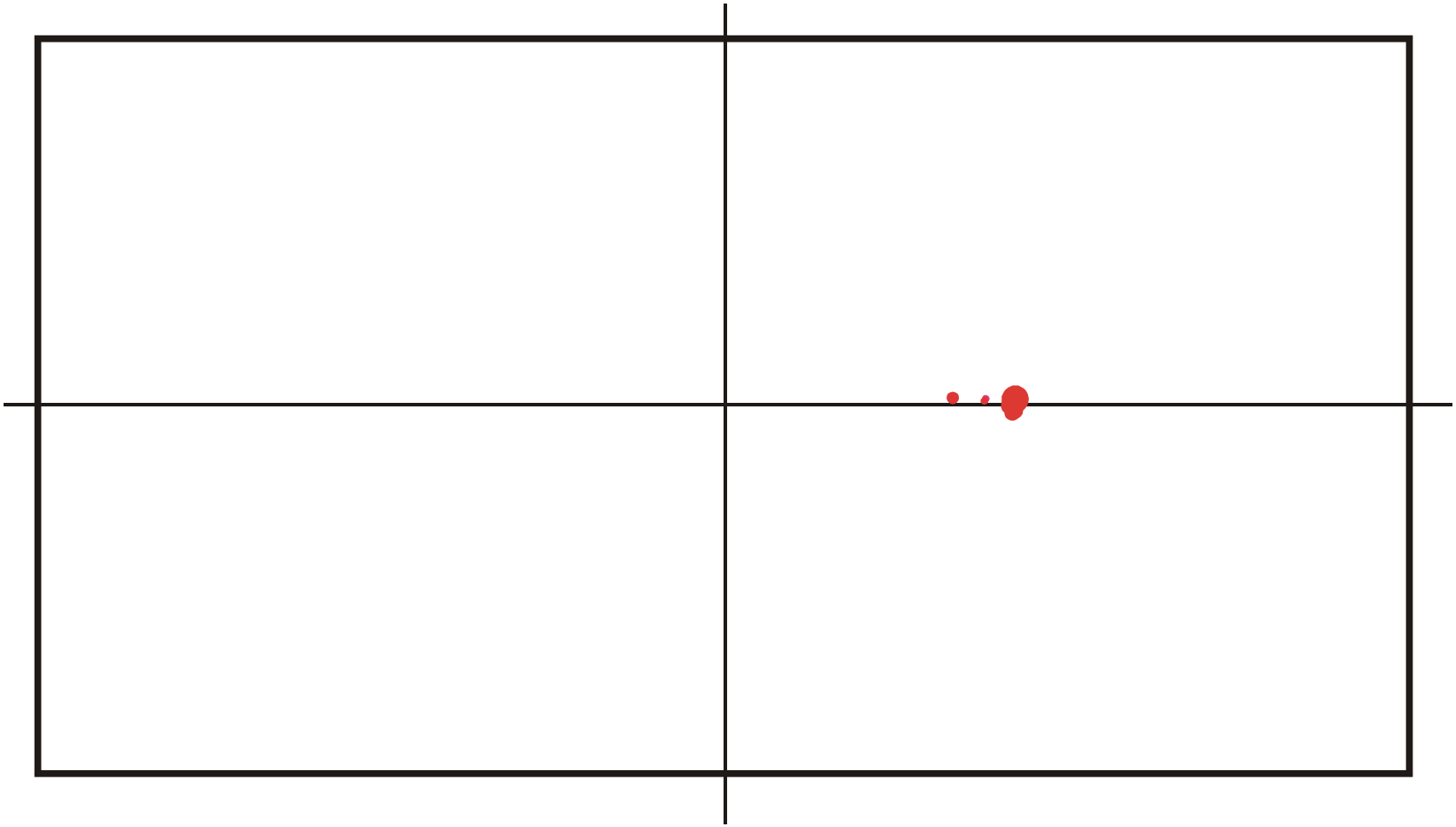
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 830



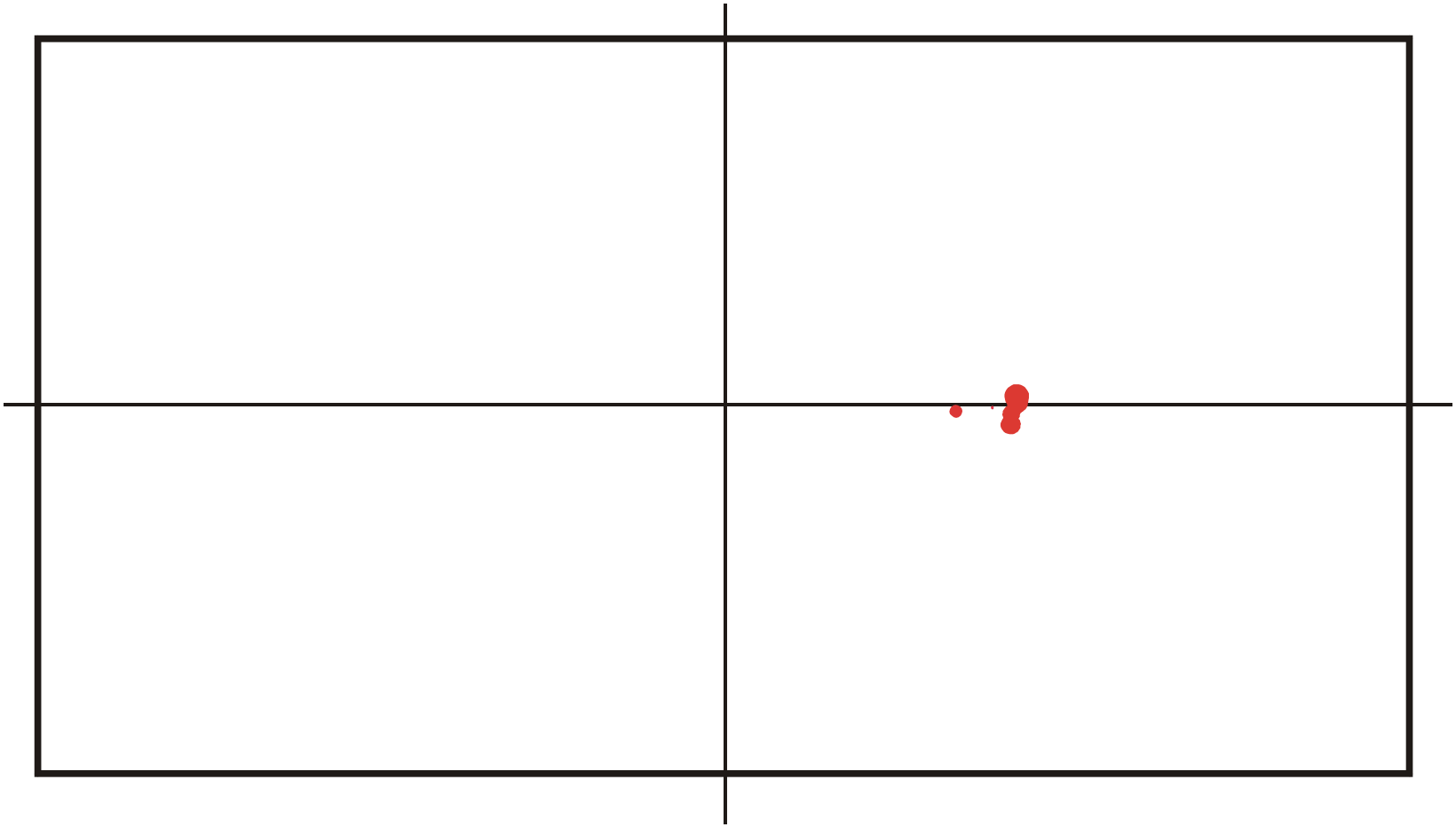
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 835



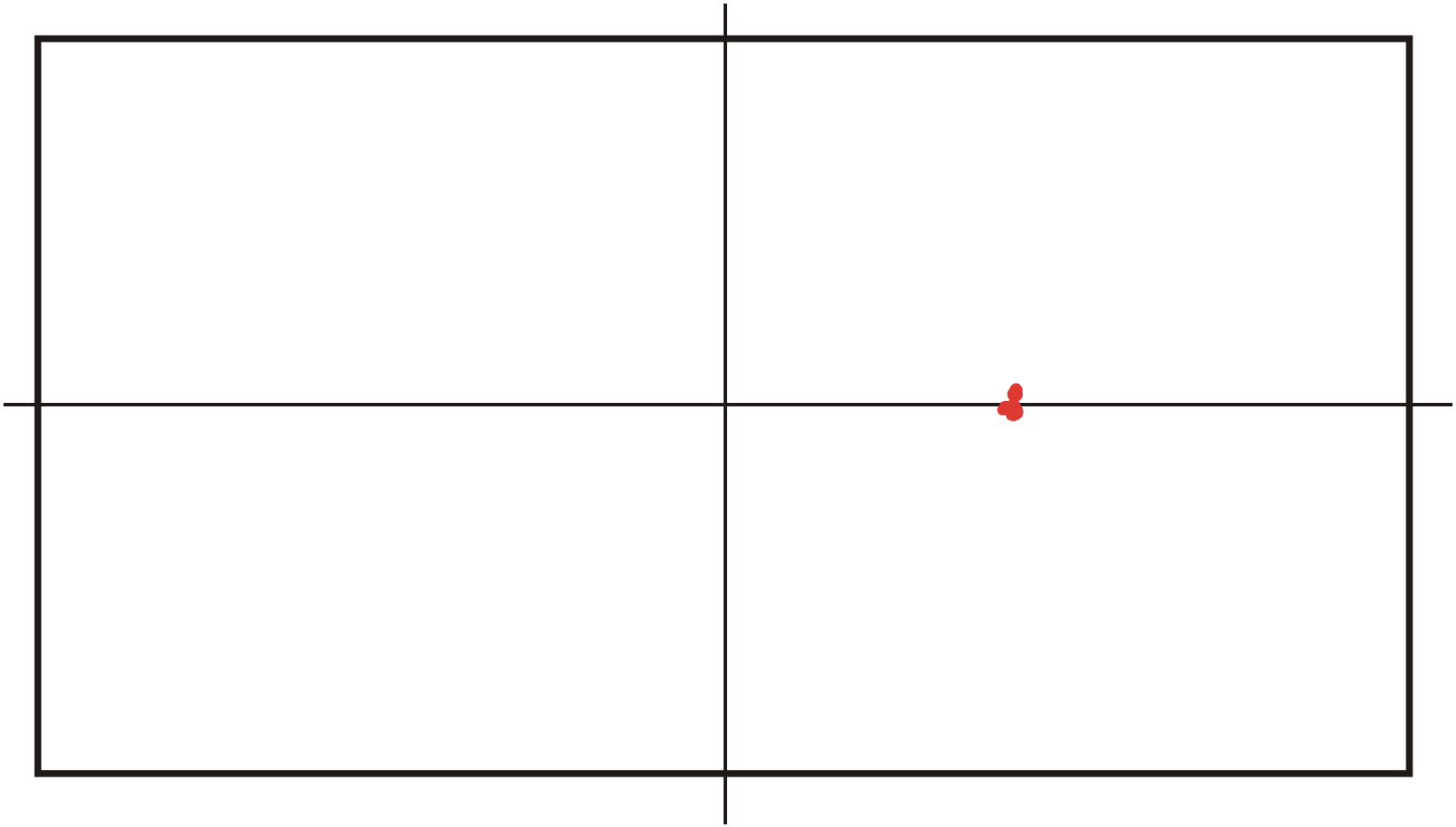
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 840



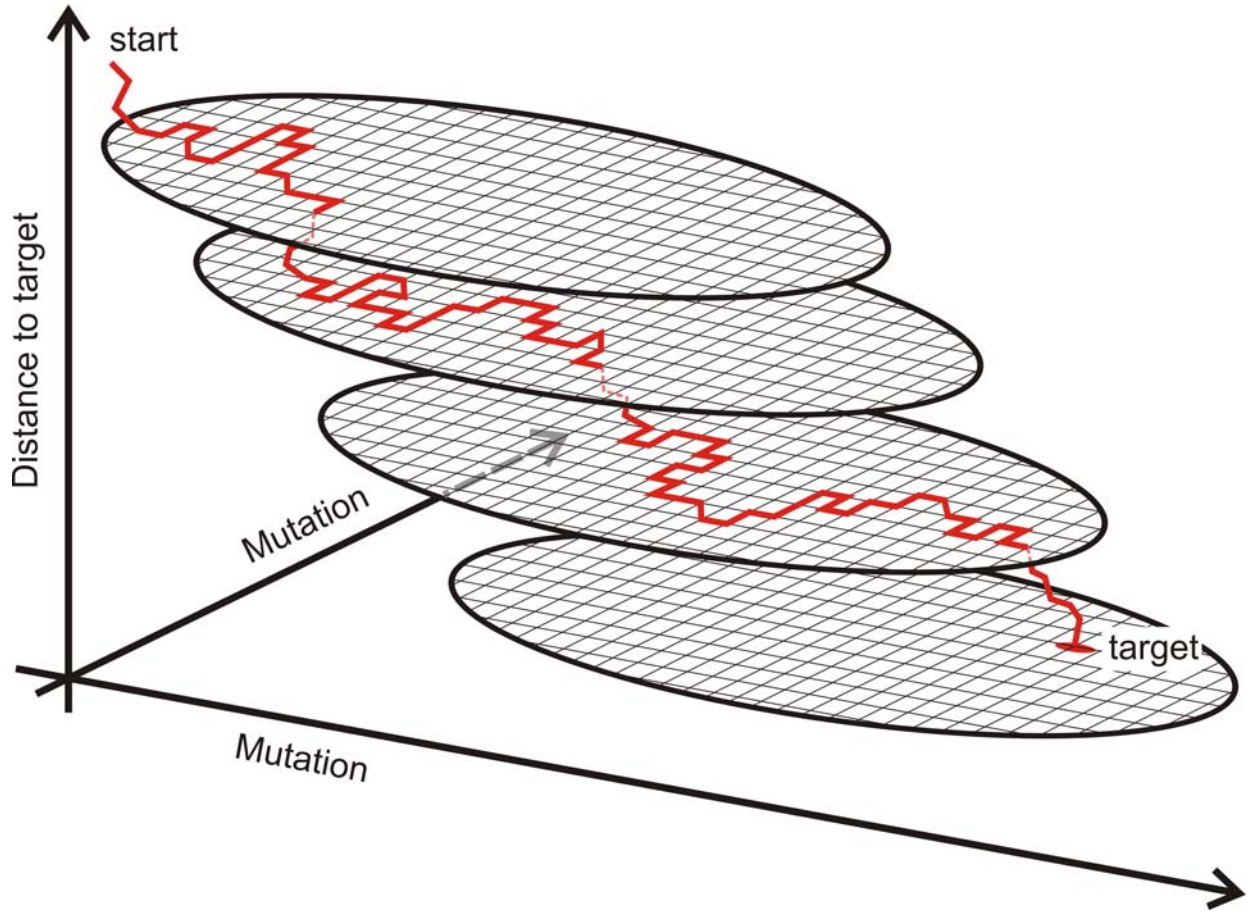
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 845



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

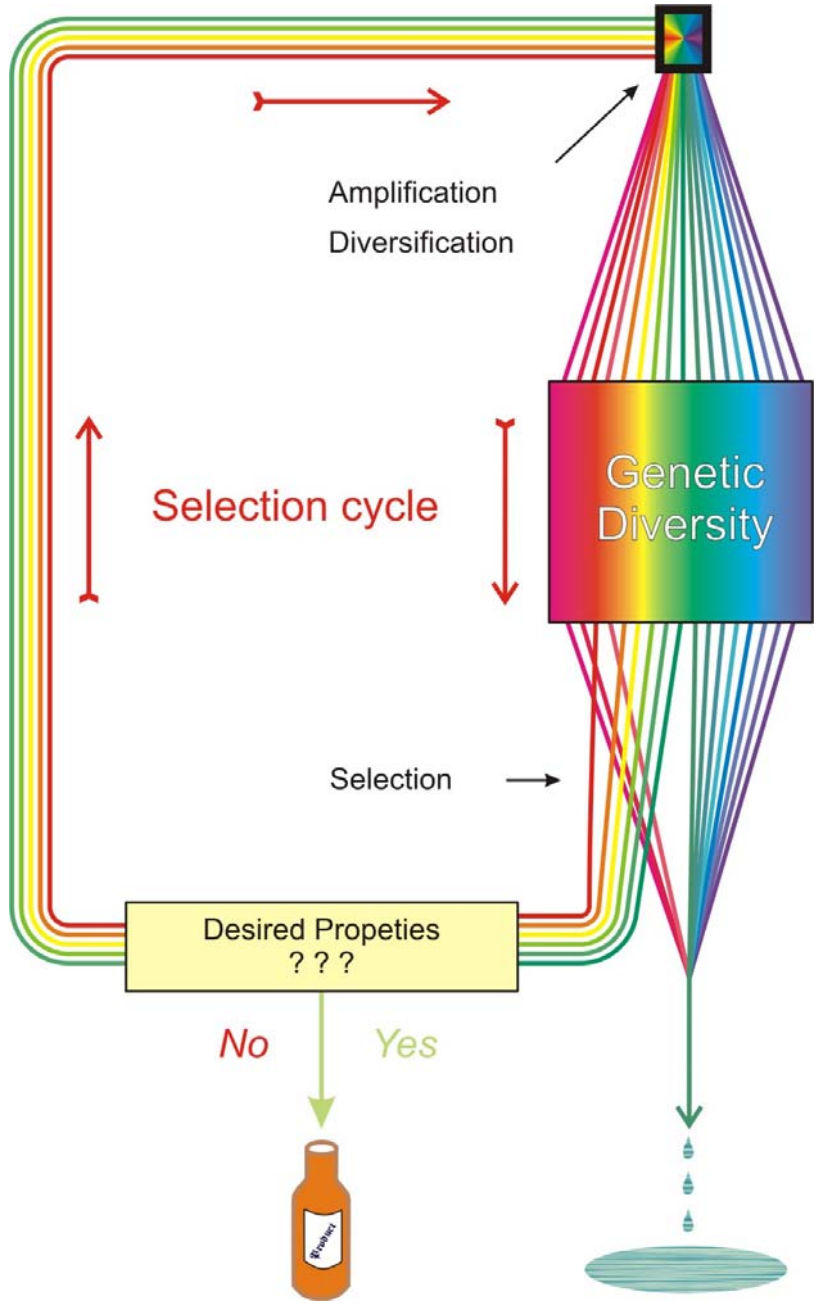


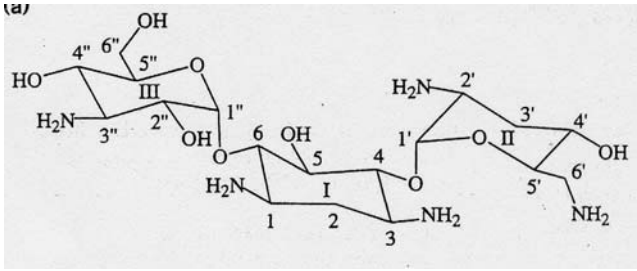
Ausbreitung und Evolution einer Population auf einem neutralen Netzwerk: t = 855



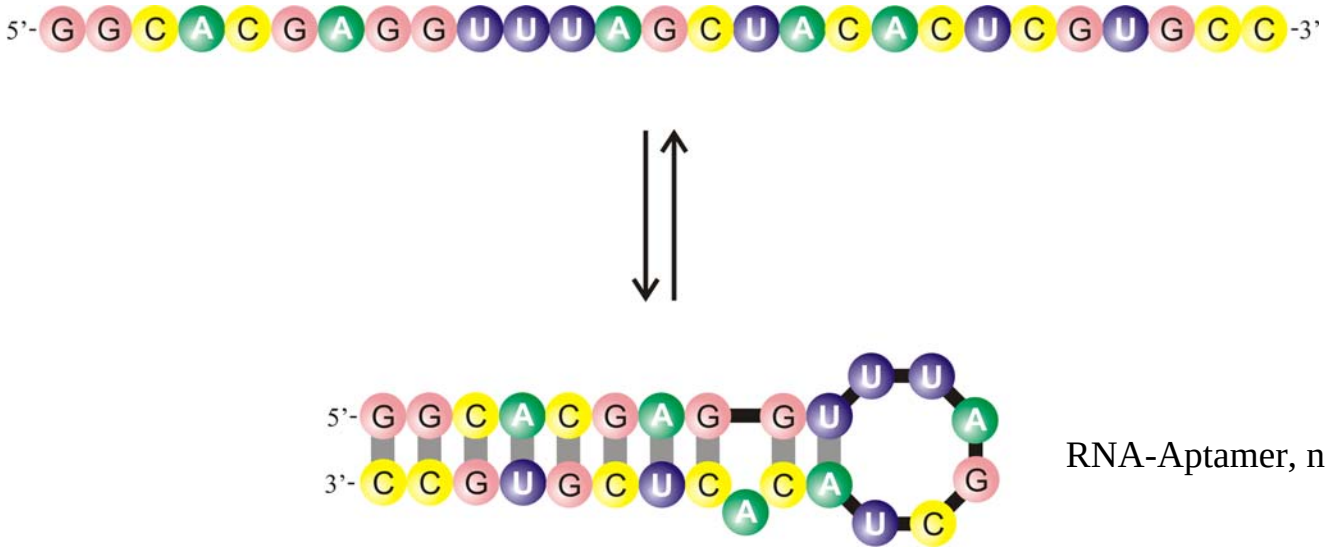
Skizze der Optimierung auf neutralen Netzwerken

Ein Beispiel künstlicher Selektion mit RNA-Molekülen oder das ‘Züchten’ von Biomolekülen





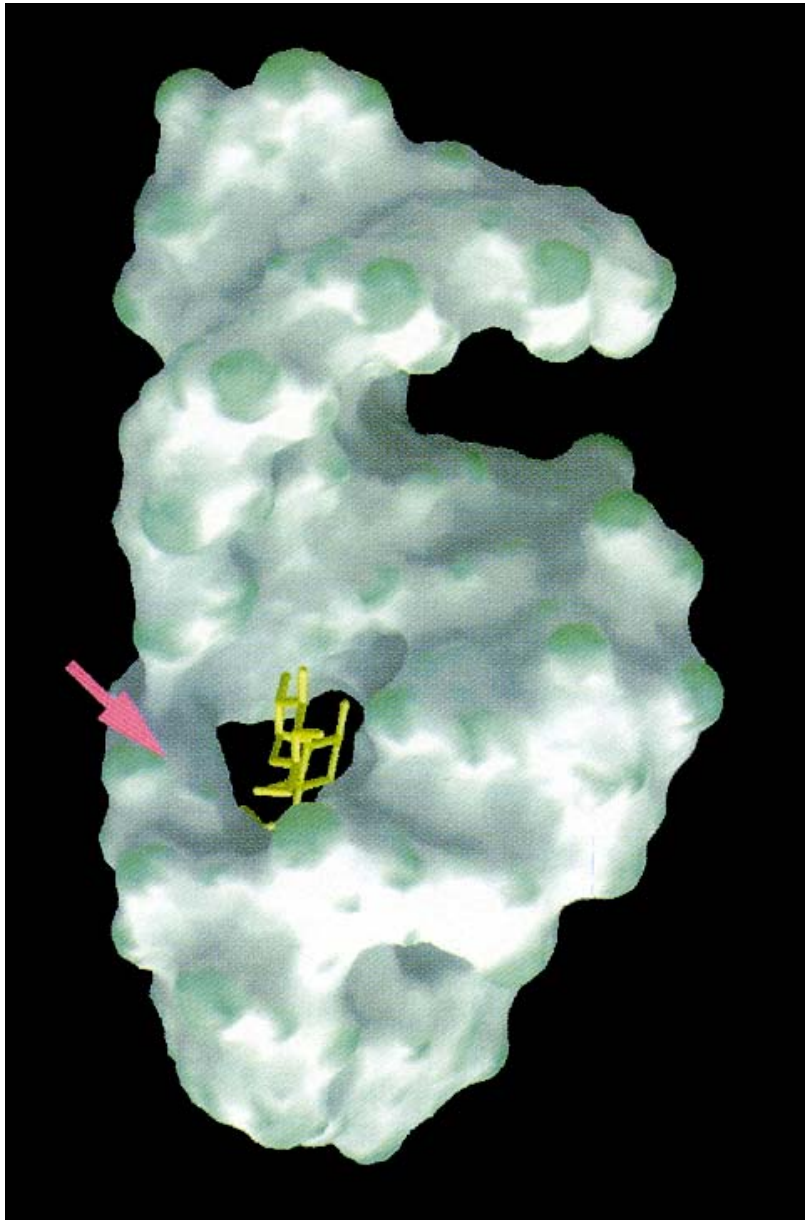
Tobramycin



RNA-Aptamer, n = 27

Ausbildung der Sekundärstruktur des an Tobramycin bindenden RNA-Aptameren mit einer Dissoziationskonstanten von **K_D = 9 nM**

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.*** Chemistry & Biology **4**:35-50 (1997)



Die räumliche Struktur des Tobramycin-Aptamer-Komplexes

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

Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer

ZHEN HUANG[1] and JACK W. SZOSTAK[2]

[1]Department of Chemistry, Brooklyn College, Ph.D. Programs of Chemistry and Biochemistry, The Graduate School of CUNY, Brooklyn, New York 11210, USA

[2]Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

ABSTRACT

Small changes in target specificity can sometimes be achieved, without changing aptamer structure, through mutation of a few bases. Larger changes in target geometry or chemistry may require more radical changes in an aptamer. In the latter case, it is unknown whether structural and functional solutions can still be found in the region of sequence space close to the original aptamer. To investigate these questions, we designed an in vitro selection experiment aimed at evolving specificity of an ATP aptamer. The ATP aptamer makes contacts with both the nucleobase and the sugar. We used an affinity matrix in which GTP was immobilized through the sugar, thus requiring extensive changes in or loss of sugar contact, as well as changes in recognition of the nucleobase. After just five rounds of selection, the pool was dominated by new aptamers falling into three major classes, each with secondary structures distinct from that of the ATP aptamer. The average sequence identity between the original aptamer and new aptamers is 76%. Most of the mutations appear to play roles either in disrupting the original secondary structure or in forming the new secondary structure or the new recognition loops. Our results show that there are novel structures that recognize a significantly different ligand in the region of sequence space close to the ATP aptamer. These examples of the emergence of novel functions and structures from an RNA molecule with a defined specificity and fold provide a new perspective on the evolutionary flexibility and adaptability of RNA.

Keywords: Aptamer; specificity; fold; selection; RNA evolution

RNA **9**:1456-1463, 2003

Evidenz für neutrale Netzwerke und ‚Shape space covering‘

Evidenz für neutrale Netzwerke und Überschneidung von Aptamerfunktionen

J Mol Evol (2003) 57:299–308
DOI: 10.1007/s00239-003-2481-y

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2003

Evolutionary Landscapes for the Acquisition of New Ligand Recognition by RNA Aptamers

Daniel M. Held, S. Travis Greathouse, Amit Agrawal, Donald H. Burke

Department of Chemistry, Indiana University, Bloomington, IN 47405-7102, USA

Received: 15 November 2002 / Accepted: 8 April 2003

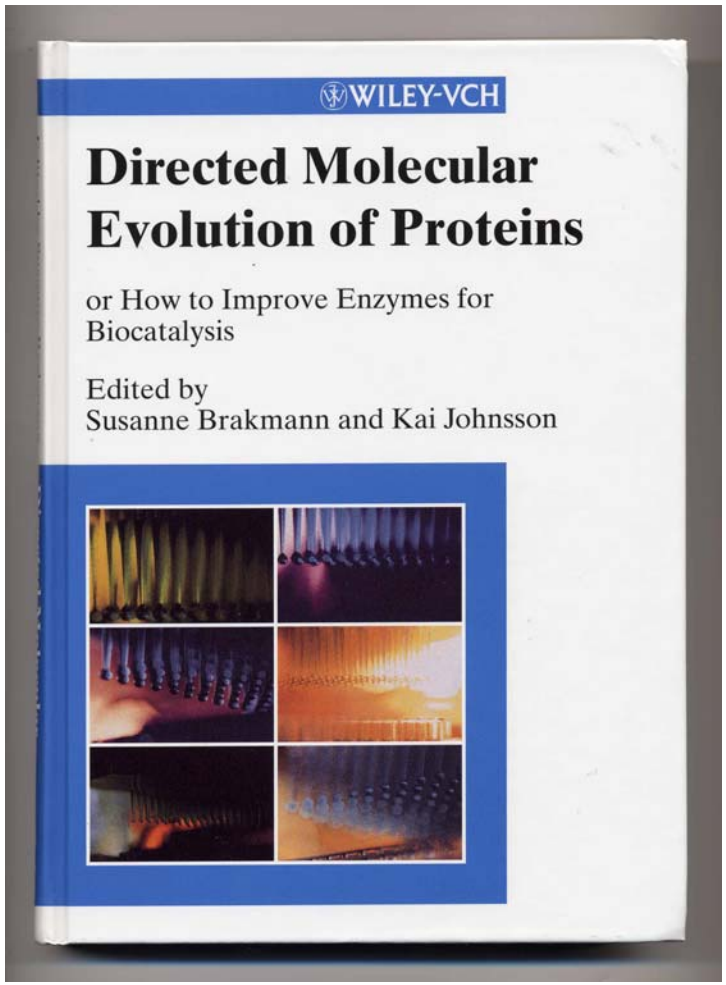
Abstract. The evolution of ligand specificity underlies many important problems in biology, from the appearance of drug resistant pathogens to the re-engineering of substrate specificity in enzymes. In studying biomolecules, however, the contributions of macromolecular sequence to binding specificity can be obscured by other selection pressures critical to bioactivity. Evolution of ligand specificity *in vitro*—unconstrained by confounding biological factors—is addressed here using variants of three flavin-binding RNA aptamers. Mutagenized pools based on the three aptamers were combined and allowed to compete during *in vitro* selection for GMP-binding activity. The sequences of the resulting selection isolates were diverse, even though most were derived from the same flavin-binding parent. Individual GMP aptamers differed from the parental flavin aptamers by 7 to 26 mutations (20 to 57% overall change). Acquisition of GMP recognition coincided with the loss of FAD (flavin-adenine dinucleotide) recognition in all isolates, despite the absence of a counter-selection to remove FAD-binding RNAs. To examine more precisely the proximity of these two activities within a defined sequence space, the complete set of all intermediate sequences between an FAD-binding aptamer and a GMP-binding aptamer were synthesized and assayed for activity. For this set of sequences, we observe a portion of a neutral network for FAD-binding function separated from GMP-binding function by a distance of three mutations. Furthermore, enzymatic probing of these aptamers revealed gross structural remodeling of the RNA coincident with the switch in ligand recognition. The capacity for neutral drift along an FAD-binding network in such close approach to RNAs with GMP-binding activity illustrates the degree of phenotypic buffering available to a set of closely related RNA sequences—defined as the set's functional tolerance for point mutations—and supports neutral evolutionary theory by demonstrating the facility with which a new phenotype becomes accessible as that buffering threshold is crossed.

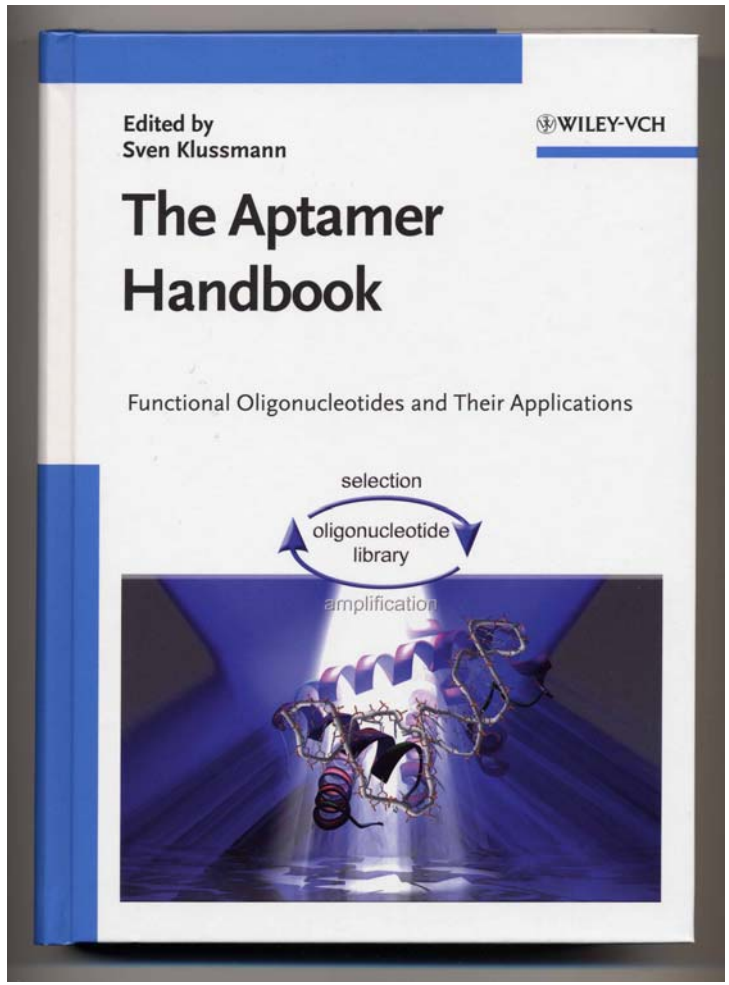
Key words: Aptamers — RNA structure — Phenotypic buffering — Fitness landscapes — Neutral evolutionary theory — Flavin — GMP

Correspondence to: Donald H. Burke; *email:* dhburke@indiana.edu

Introduction

RNA aptamers targeting small molecules serve as useful model systems for the study of the evolution and biophysics of macromolecular binding interactions. Because of their small sizes, the structures of several such complexes have been determined to atomic resolution by NMR spectrometry or X-ray crystallography (reviewed by Herman and Patel 2000). Moreover, aptamers can be subjected to mutational and evolutionary pressures for which survival is based entirely on ligand binding, without the complicating effects of simultaneous selection pressures for bioactivity, thus allowing the relative contributions of each activity to be evaluated separately.





Anwendung der molekularen Evolution auf Probleme in der Biotechnologie

1. Mathematik und Physik
2. Mathematik in der Biologie
3. Das Zeitalter des Computers
4. Bioinformatik und Systembiologie
5. Evolutionsforschung am Computer
6. Evolution im ‚Flussreaktor'
7. **Komplexität ‚ohne Ende'**

Komplexität in der molekularen Welt

(i) Suboptimale Konformationen und metastabile Zustände

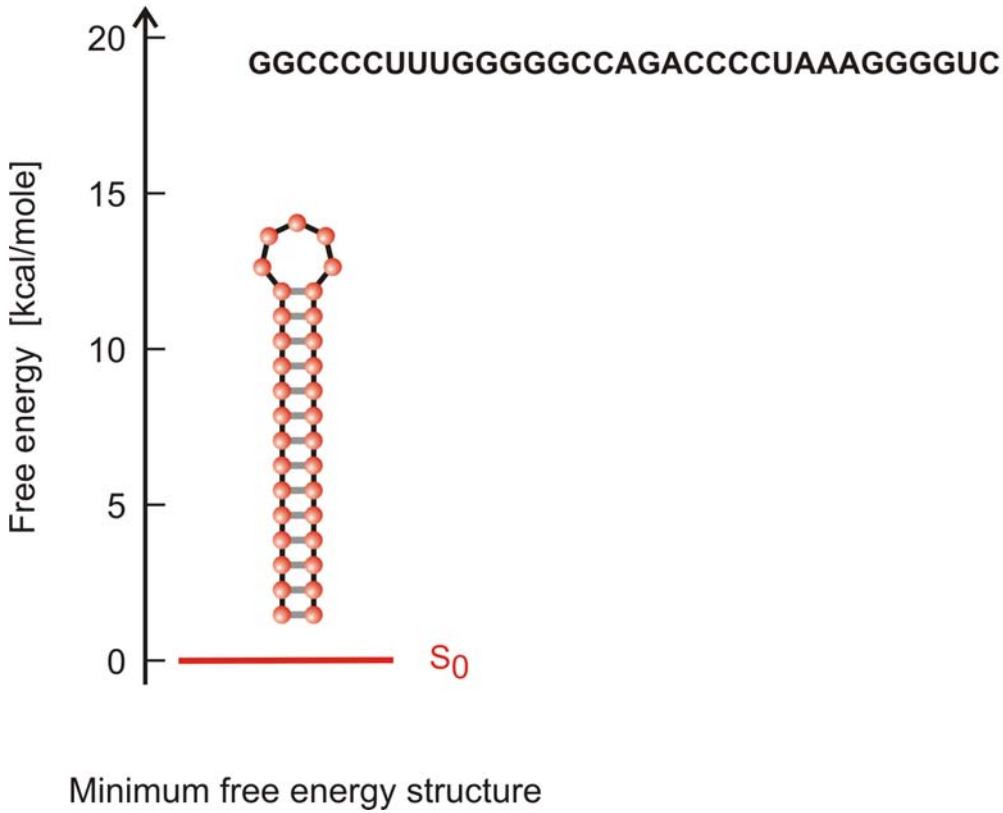
(ii) Genomverdopplungen und Genverlust

(iii) Alternative Prozessierung der transkribierten RNA

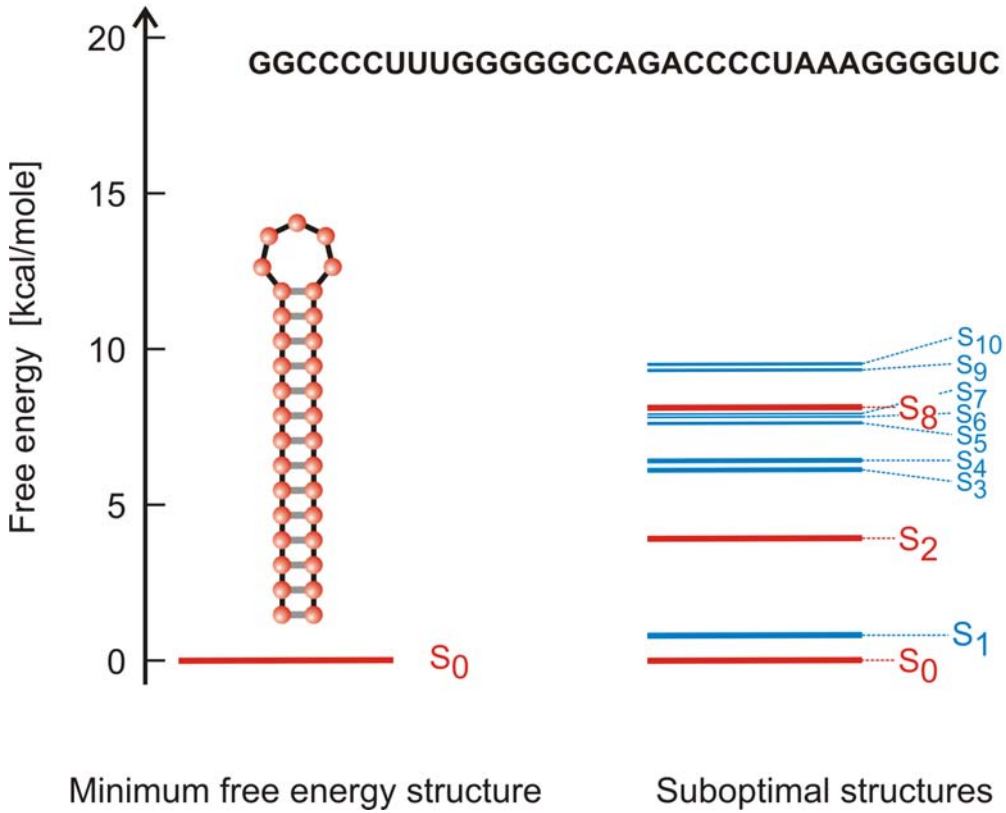
(iv) Räumliche Strukturen in der Zelle

(v) Regulation durch RNA-Moleküle

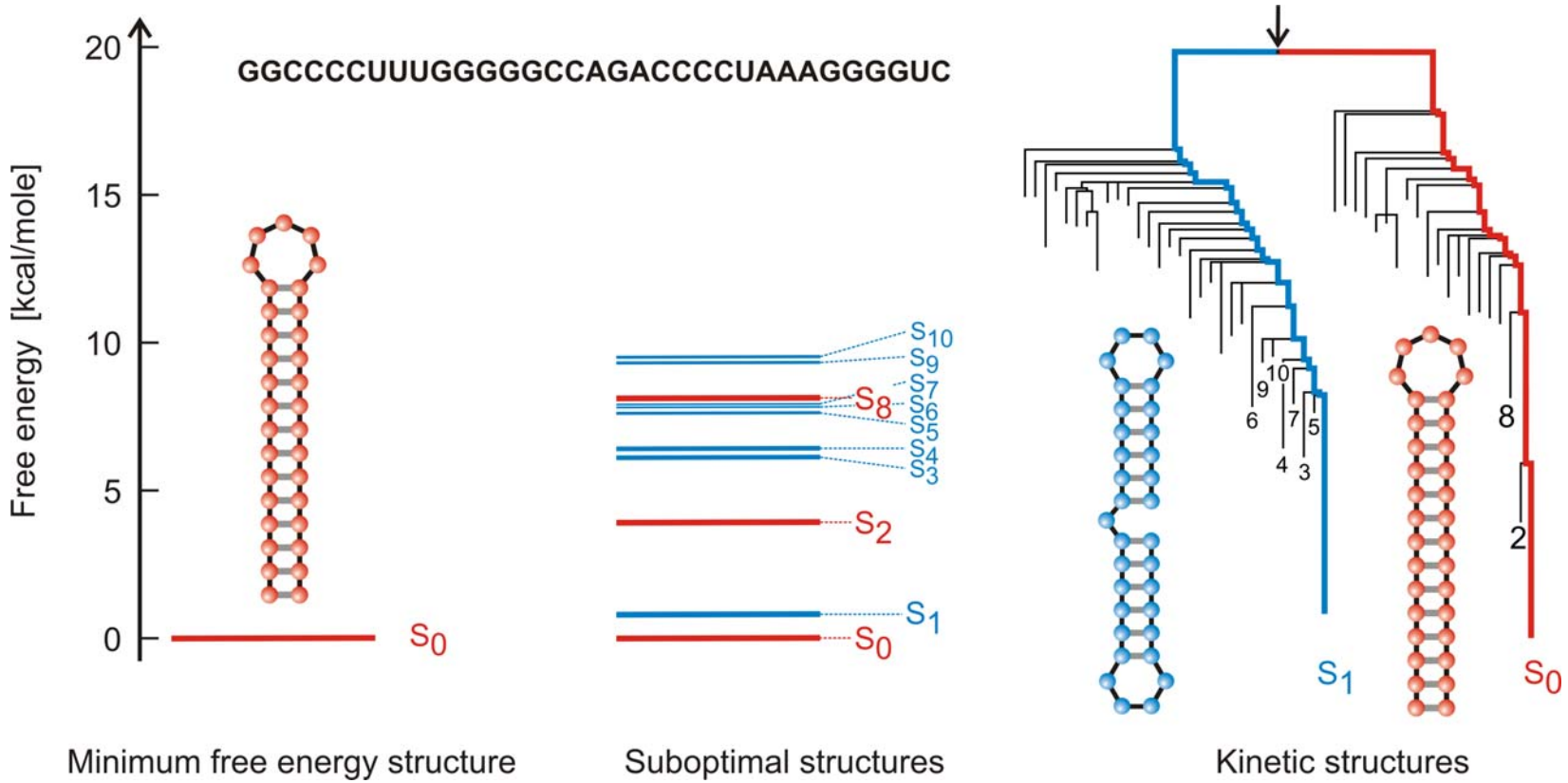
..... und vieles andere mehr.



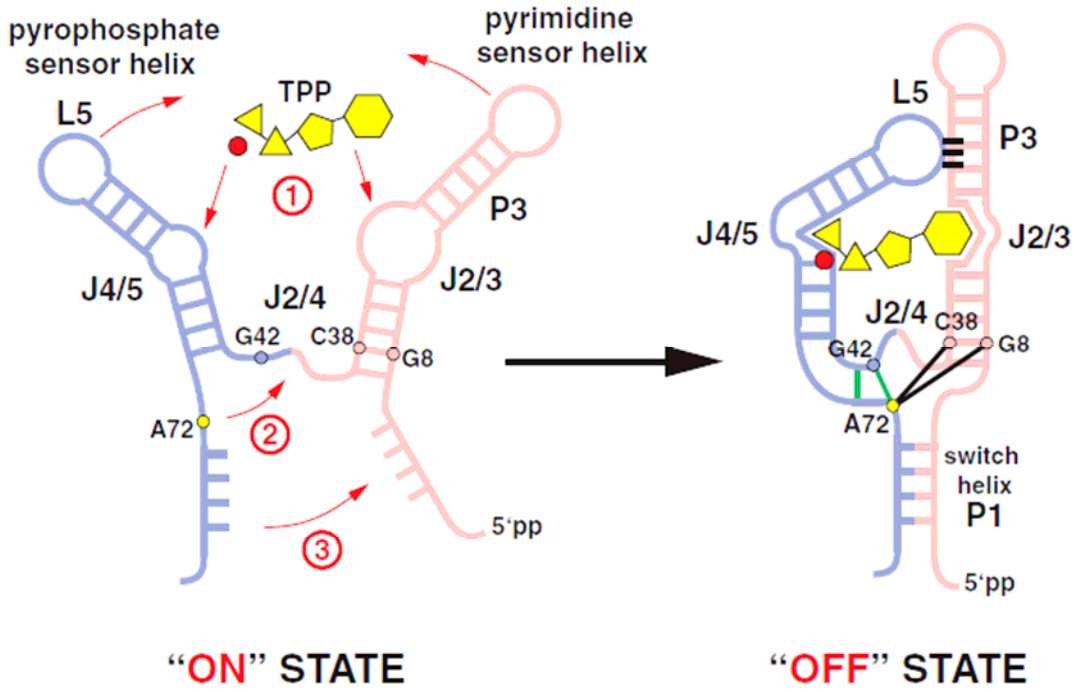
Erweiterung des molekularen Strukturbegriffes

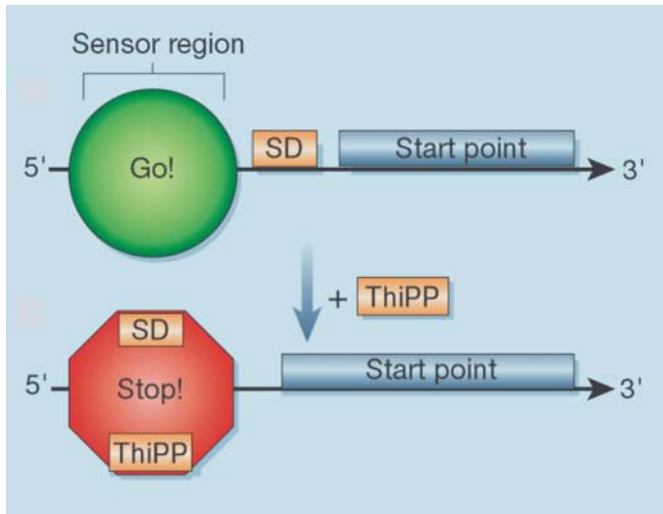


Erweiterung des molekularen Strukturbegriffes



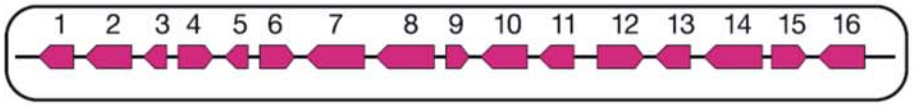
Erweiterung des molekularen Strukturbegriffes





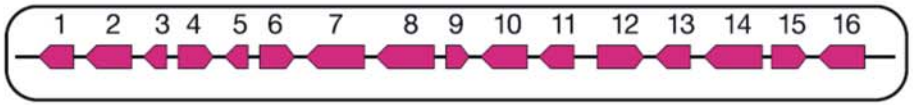
Der Thiamin-Pyrophosphat RNA-Schalter

S. Thore, M. Leibundgut, N. Ban.
Science **312**:1208-1211, 2006.



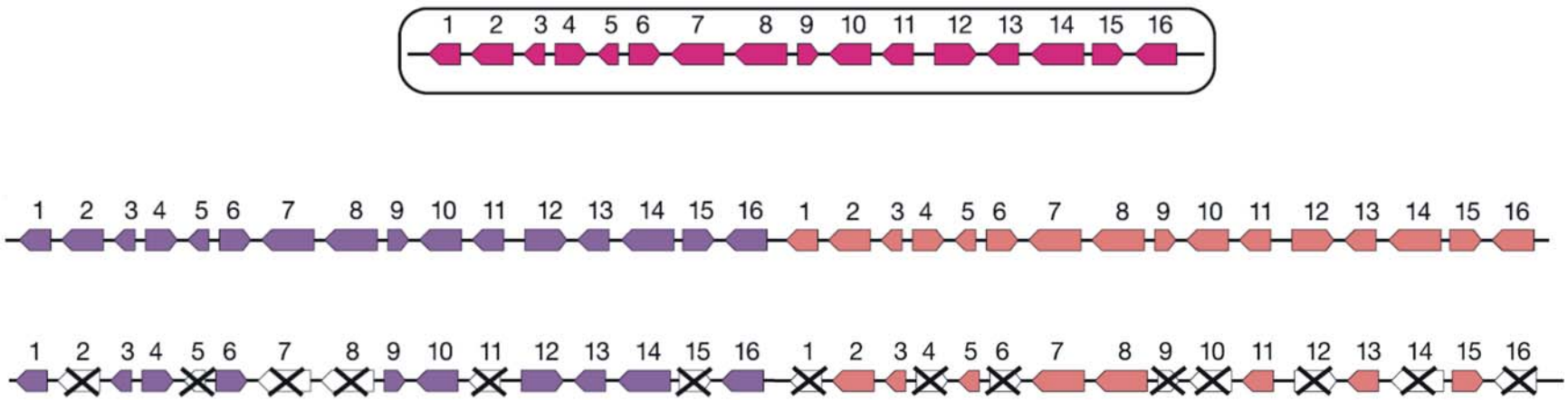
Ein Model für eine Genomverdopplung in Hefe vor 100 Millionen Jahren

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



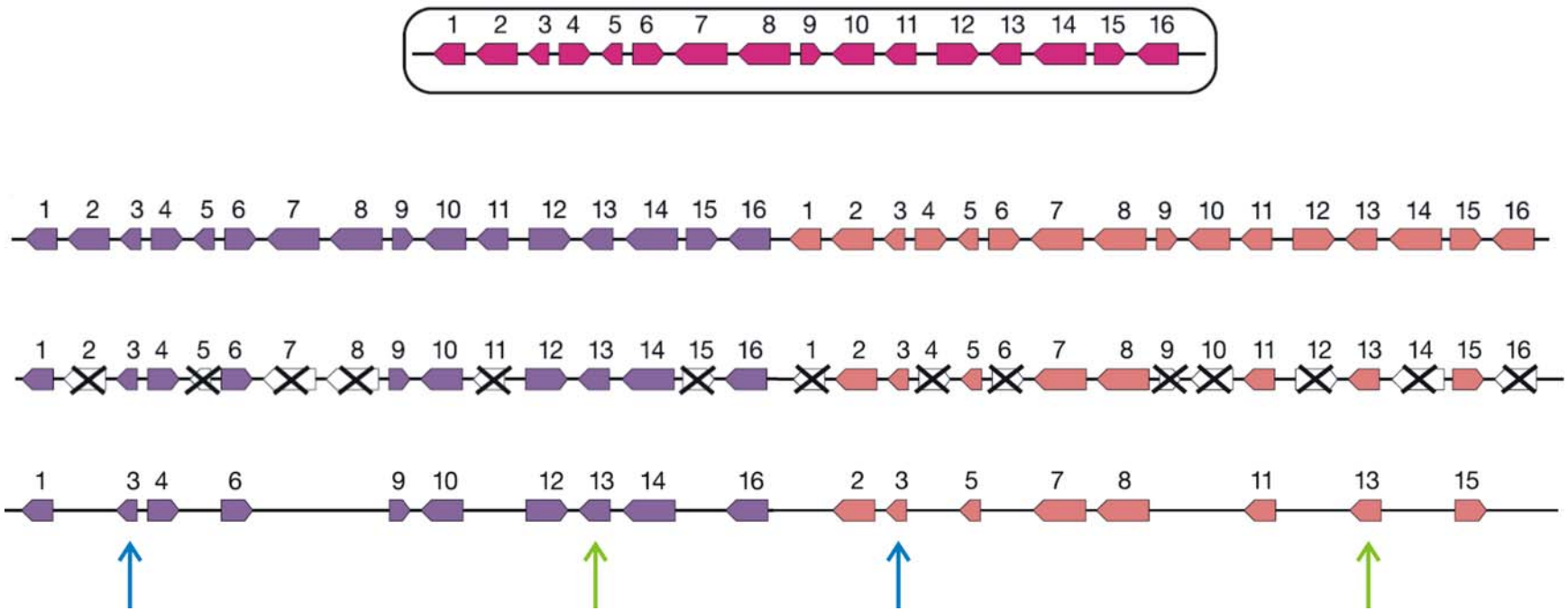
Ein Model für eine Genomverdopplung in Hefe vor 100 Millionen Jahren

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



Ein Model für eine Genomverdopplung in Hefe vor 100 Millionen Jahren

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



Ein Model für eine Genomverdopplung in Hefe vor 100 Millionen Jahren

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004

Die Schwierigkeit eine Definition für das ‚Gen' zu geben.

Helen Pearson,
Nature **441**: 399-401, 2006

WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past[1]. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals[2]. If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist Laurence Hurst at the University of Bath, UK.

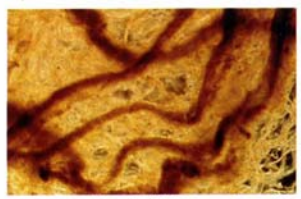
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

"We've come to the realization that the genome is full of overlapping transcripts."
— Phillip Kapranov

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the previously unimagined scope of RNA.



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

P. PLAILLY/SPL

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from[3].

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team[4], and one by geneticist Rotem Sorek[5], now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

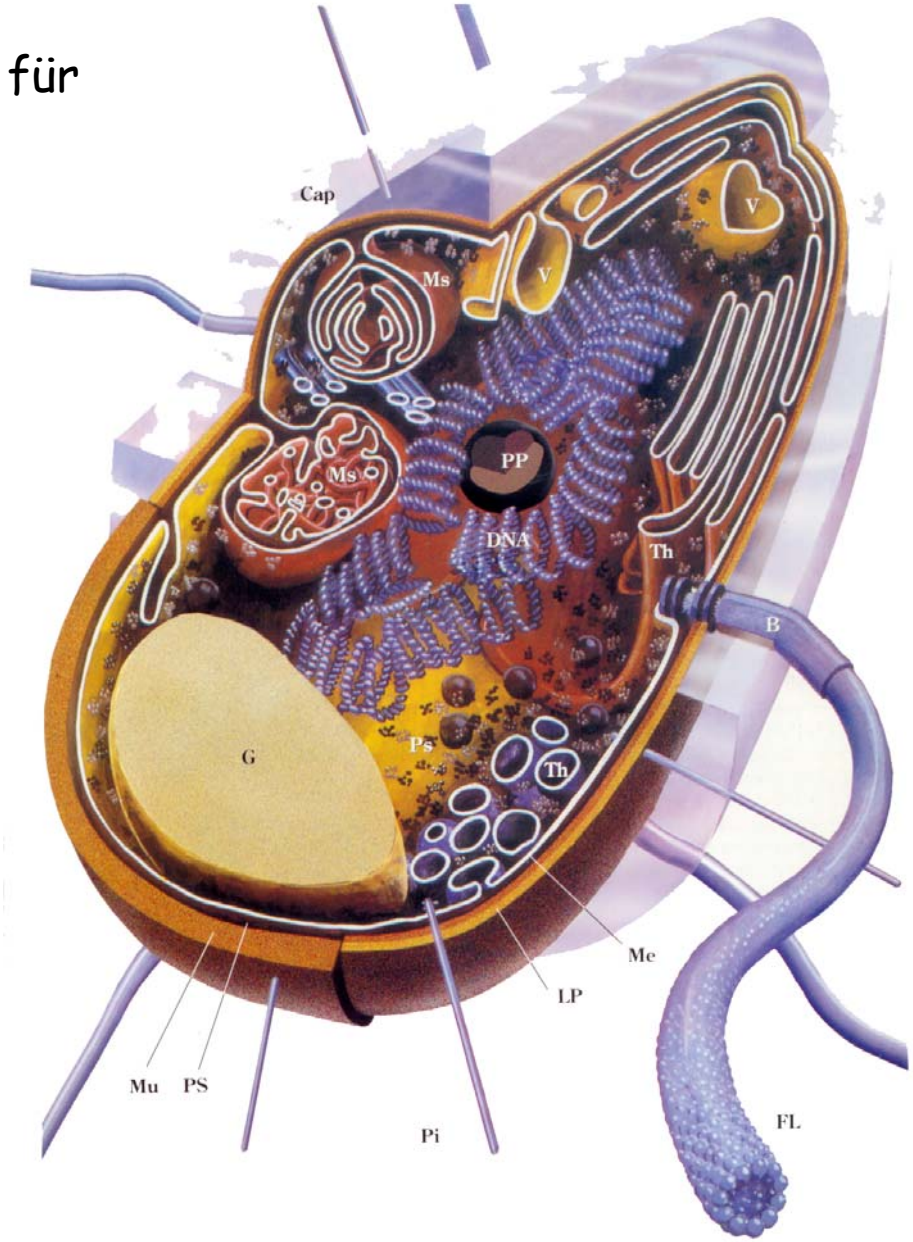
Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

Die Bakterienzelle als ein Beispiel für die einfachste Form autonomen Lebens.

Escherichia coli genome:

4 Millionen Nukleotide

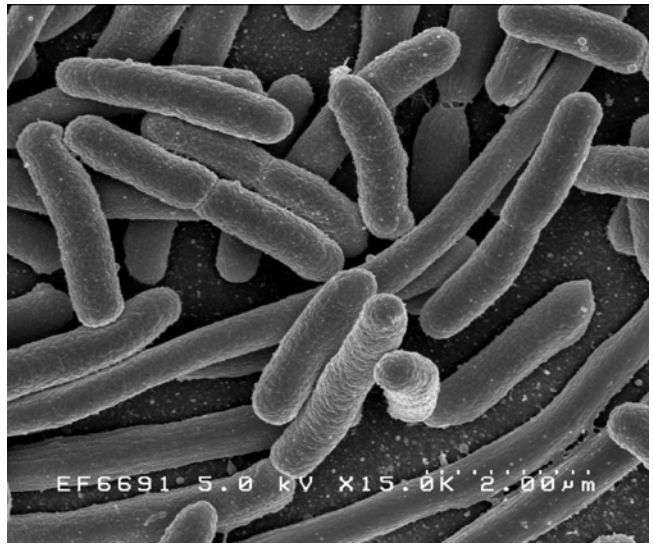
4460 Gene



Die Struktur des Bakteriums *Escherichia coli*

E. coli:	Genomlänge	4×10^6 nucleotides
	Zahl der Zelltypen	1
	Zahl der Gene	4 460

Vier Bücher zu je 300 Seiten



Man:	Genomlänge	3×10^9 nucleotides
	Zahl der Zelltypen	200
	Zahl der Gene	≈ 30 000

Eine Bibliothek mit 3000 Bänden zu je 300 Seiten

Zunahme der Komplexität in der Evolution

ENCODE steht für:

ENCyclopedia Of DNA Elements.

ENCODE Projekt Konsortium. Identifizierung und Analyse der funktionellen Elemente in 1% des menschlichen Genoms im Rahmen des ENCODE Pilotprojektes.

Nature **447**:799-816, 2007

Dank an die Mitarbeiter

Universität Wien

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

Paul E. Phillipson, University of Colorado at Boulder, CO

Heinz Engl, Philipp Kügler, **James Lu**, **Stefan Müller**, RICAM Linz, AT

Jord Nagel, **Kees Pleij**, Universiteit Leiden, NL

Walter Fontana, Harvard Medical School, MA

Christian Reidys, Nankai University, Tianjin, China

Christian Forst, Los Alamos National Laboratory, NM

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

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

Wolfgang Schnabl, Kurt Grünberger, Michael Kospach, **Andreas Wernitznig**, **Stefanie Widder, Stefan Wuchty**, Universität Wien, AT

Jan Cupal, **Stefan Bernhart**, **Lukas Endler, Ulrike Langhammer**, **Rainer Machne, Ulrike Mückstein**, **Hakim Tafer, Thomas Taylor,** Universität Wien, AT

Dank an die Geldgeber

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: Bioinformatics Network (BIN)

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Universität Wien

Danke für die Aufmerksamkeit !

Web-Page für weitere Informationen

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