Leben – Ein Produkt von Evolution oder Design?

Peter Schuster

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

Erich-Regener-Vortragsreihe am
Max-Planck-Institut für Sonnenforschung

Katlenburg-Lindau, 09.06.2008

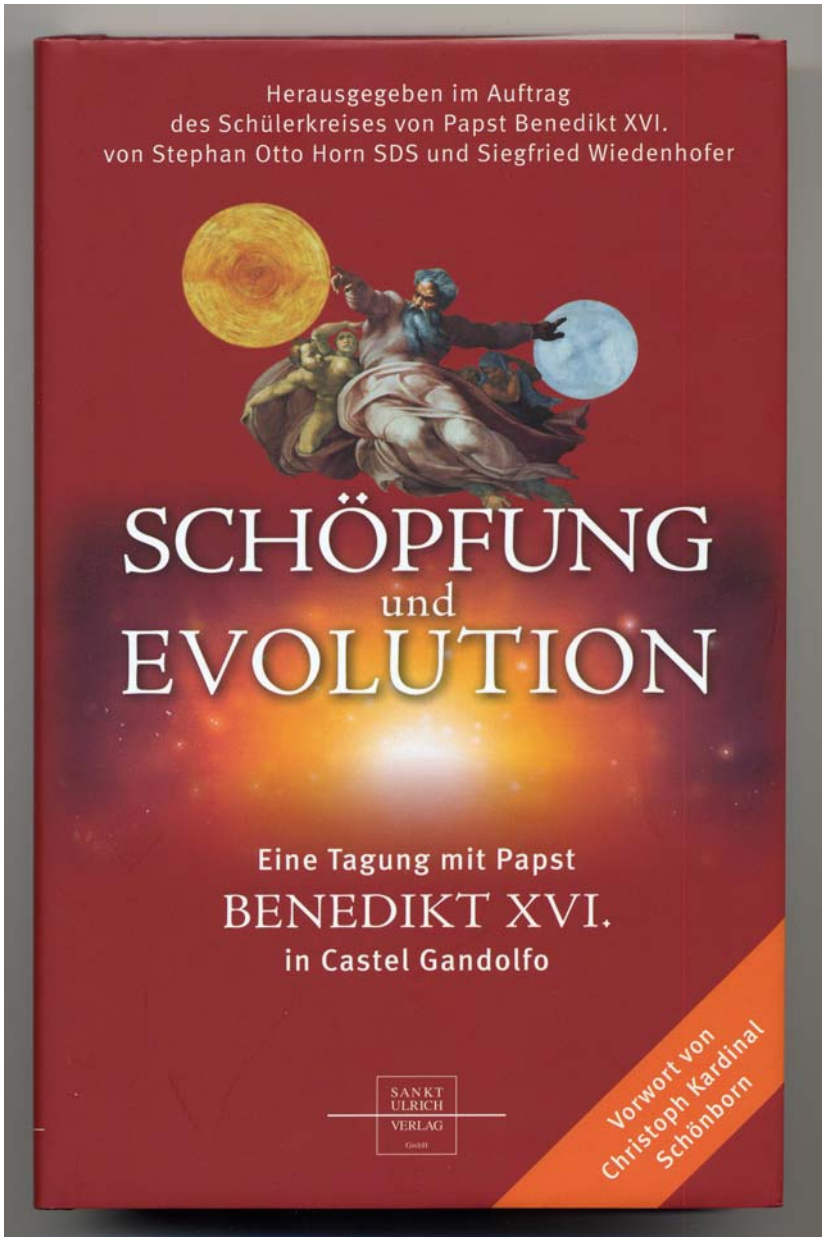
Web-Page for further information:

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

Kardinal Christoph Schönborn, ***Finding Design in Nature***, Gastkommentar in *The New York Times*, July 5, 2005

„ ... Evolution in the sense of common ancestry might be true, but evolution in the Neo-Darwinian sense – an unguided, unplanned process of random variation and natural selection – is not. Any system of thought that denies or seeks to explain away the overwhelming evidence for design in biology is ideology, not science.

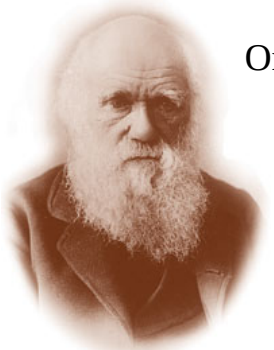
... Scientific theories that try to explain away the appearance of design as the result of ‚chance and necessity' are not scientific at all, but ... an abdication of human intelligence."



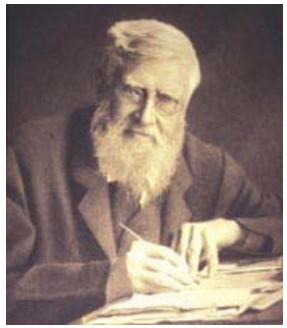
INHALT

1. Evolution und der „Baum des Lebens“
2. Wahrscheinlichkeiten und Zufall
3. Vermehrung, Mutation und Selektion
4. Evolution von Molekülen und Optimierung
5. Evolutionäres “Basteln” und Komplexität
6. Schlußbemerkungen

1. **Evolution und der „Baum des Lebens“**
2. Wahrscheinlichkeiten und Zufall
3. Vermehrung, Mutation und Selektion
4. Evolution von Molekülen und Optimierung
5. Evolutionäres “Basteln” und Komplexität
6. Schlußbemerkungen

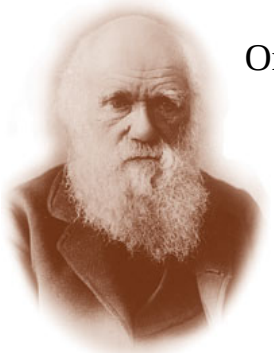


Origin of evolutionary biology
1859



Charles Darwin

Alfred Russel Wallace

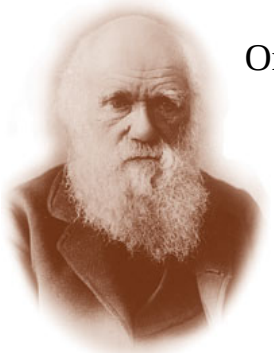


Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

Gregor Mendel



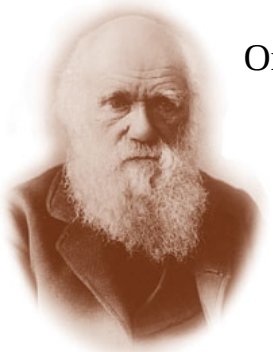
Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

‘Rediscovery’ 1900

Gregor Mendel



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

Gregor Mendel

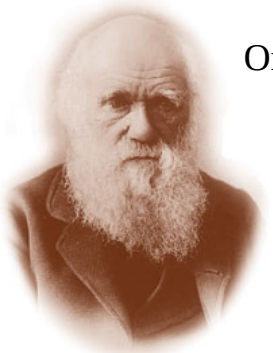
‘Rediscovery’ 1900

First unification: Population genetics 1930

Ronald Fisher

JSB Haldane

Sewall Wright



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

Gregor Mendel

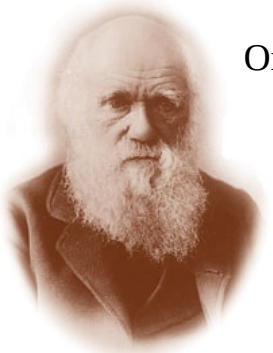
‘Rediscovery’ 1900

First unification: Population genetics 1930

Ernst Mayr

Theodosius
Dobzhansky

Synthetic theory
1940 - 1950



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

‘Rediscovery’ 1900

Gregor Mendel

Friedrich Woehler

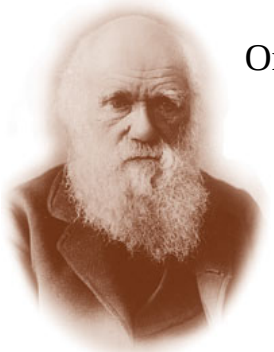
Origin of biochemistry
1828

First unification: Population genetics 1930

Ernst Mayr

Theodosius Dobzhansky

Synthetic theory
1940 - 1950



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

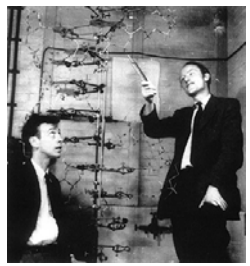
‘Rediscovery’ 1900

Gregor Mendel

Friedrich Woehler

Origin of biochemistry
1828

Origin of molecular biology 1953



James Watson and Francis Crick

First unification: Population genetics 1930

Ernst Mayr

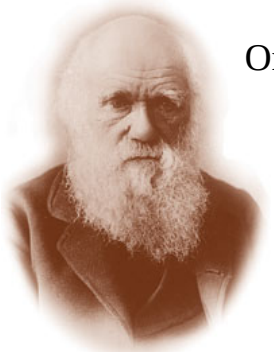
Theodosius Dobzhansky

Synthetic theory
1940 - 1950

Max Perutz

John Kendrew

Biology of the 21st century



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

‘Rediscovery’ 1900

Gregor Mendel

Friedrich Woehler

Origin of biochemistry
1828

First unification: Population genetics 1930

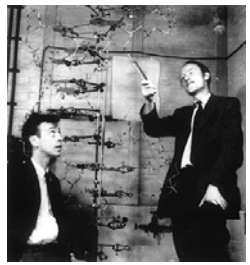
Ernst Mayr

Theodosius Dobzhansky

Synthetic theory
1940 - 1950

Origin of molecular biology 1953

Jacques Monod



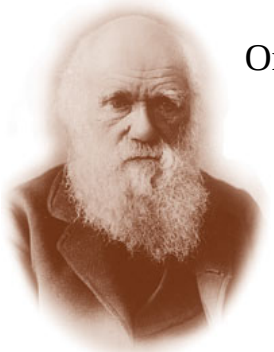
James Watson and Francis Crick

François Jacob

Max Perutz

John Kendrew

Biology of the 21st century



Charles Darwin

Origin of evolutionary biology
1859

Origin of genetics
1865

‘Rediscovery’ 1900

Gregor Mendel

Friedrich Woehler

Origin of biochemistry
1828

First unification: Population genetics 1930

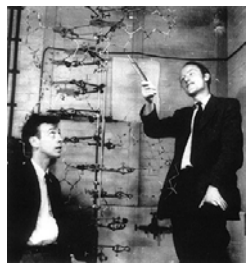
Ernst Mayr

Theodosius Dobzhansky

Synthetic theory
1940 - 1950

Origin of molecular biology 1953

Jacques Monod



James Watson and Francis Crick

François Jacob

Max Perutz

Manfred Eigen

Sydney Brenner

John Kendrew

Biology of the 21st century

Biomathematics, bioinformatics, … , biophysics, biochemistry, … , molecular genetics, … , systems biology, biomedicine, macroscopic biology, evolutionary biology, sociobiology, anthropology, …

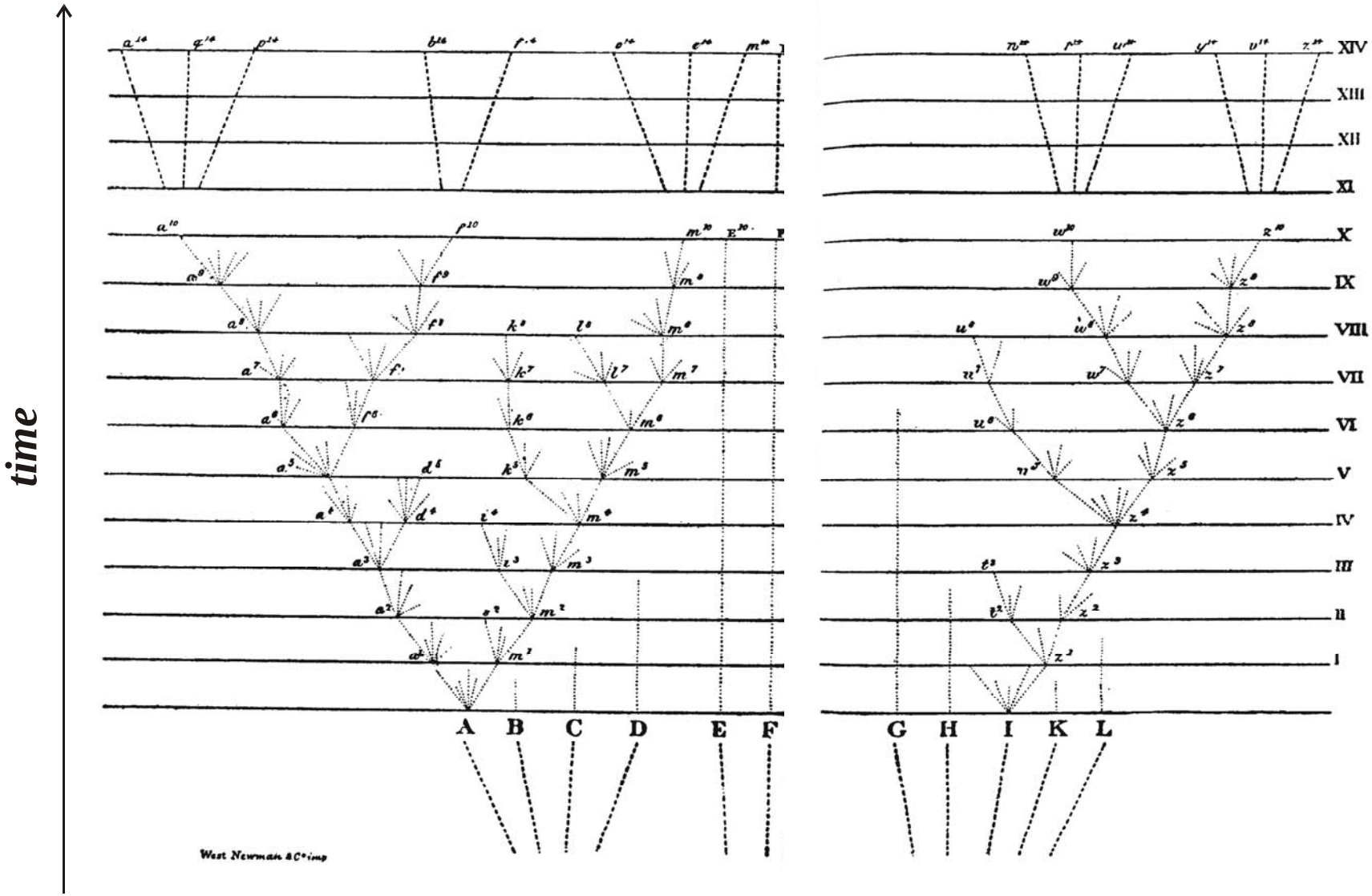
Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

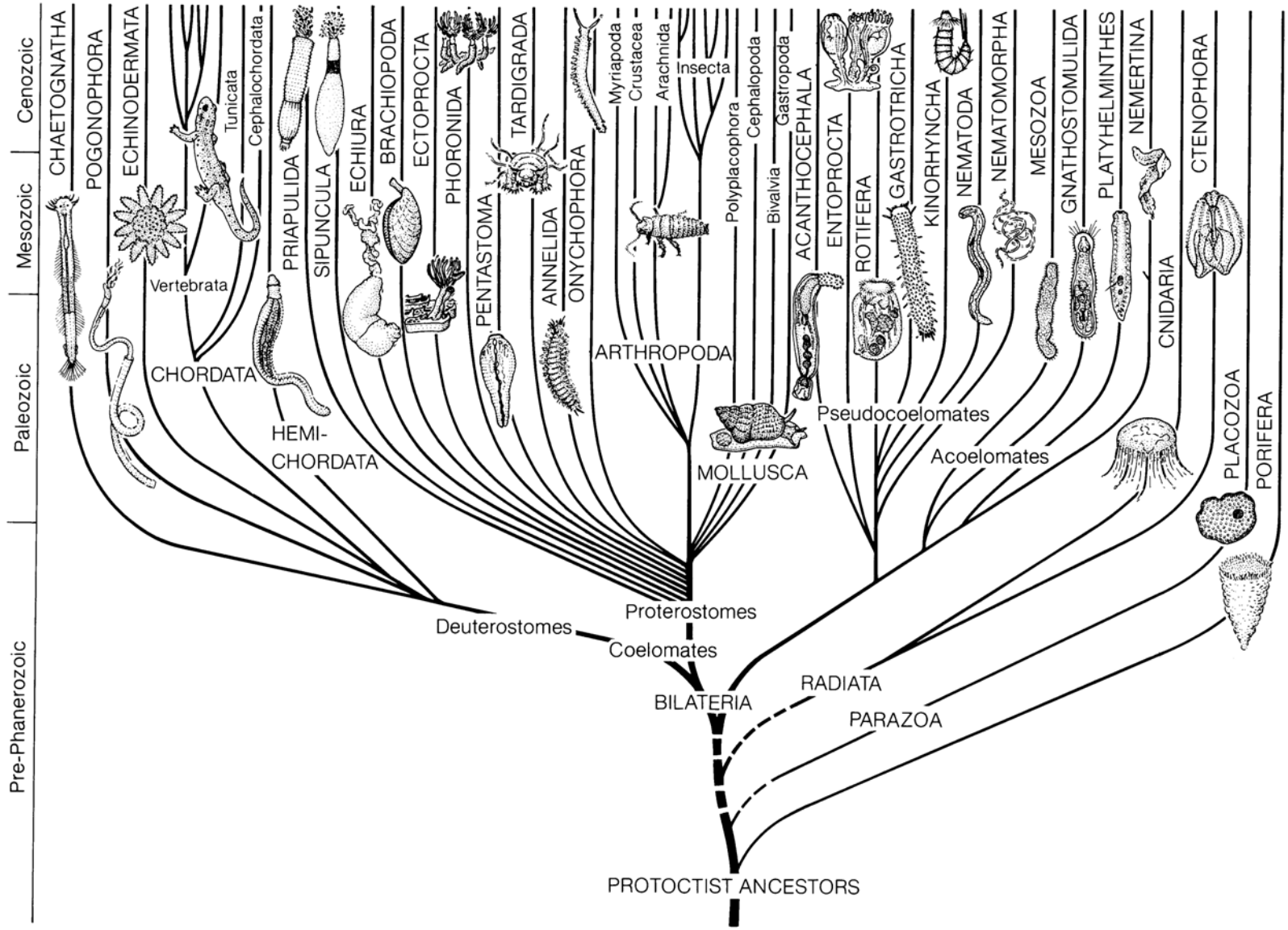
Biologists distinguish the **genotype** – the genetic information – and the **phenotype** – the organisms and all its properties. The **genotype** is unfolded in development and yields the **phenotype**.

Variation operates on the **genotype** – through mutation and recombination – whereas the **phenotype** is the target of **selection**.

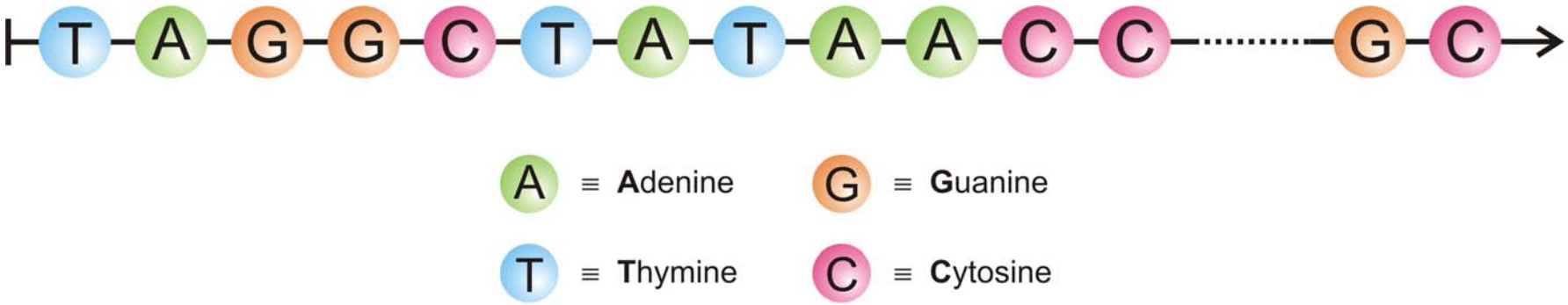
One important property of the Darwinian mechanism is that **variations** in the form of mutation or recombination events occur **uncorrelated** to their **effects** on the **selection** of the **phenotype**.



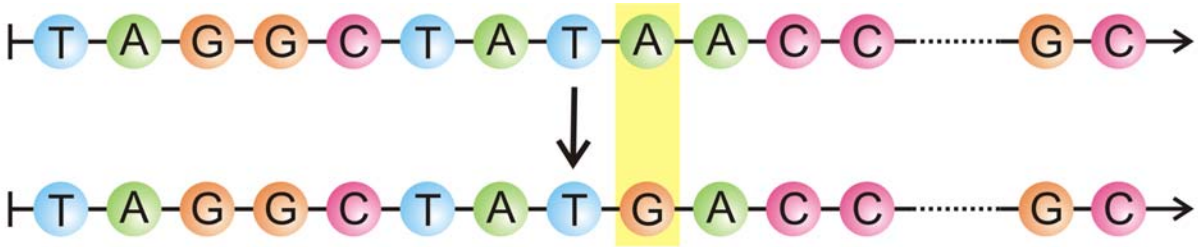
Charles Darwin, *The Origin of Species*, 6th edition.
Everyman‘s Library, Vol.811, Dent London, pp.121-122.



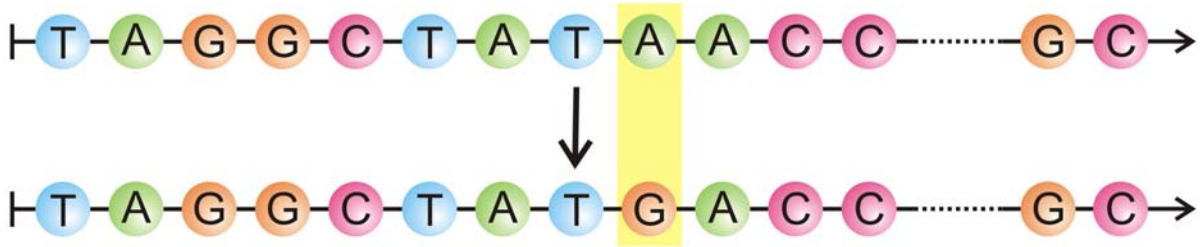
Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth*. W.H. Freeman, San Francisco, 1982.



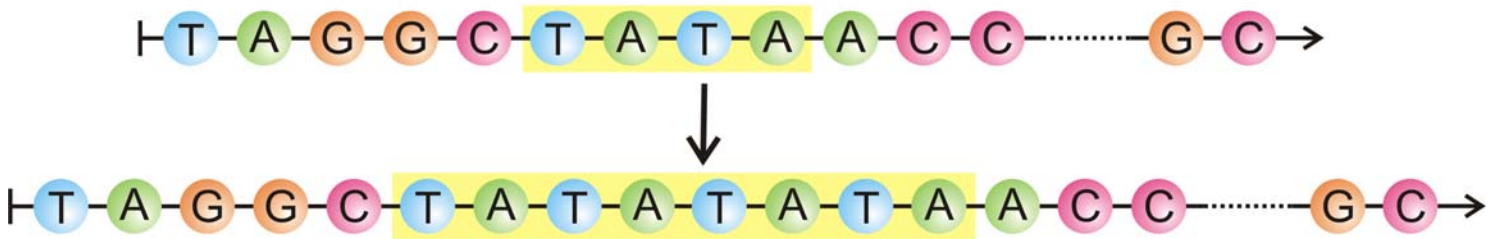
Deoxyribonucleic acid - DNA



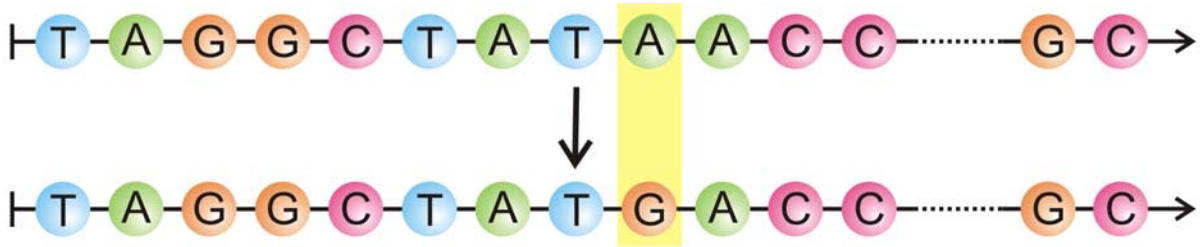
Punktmutation



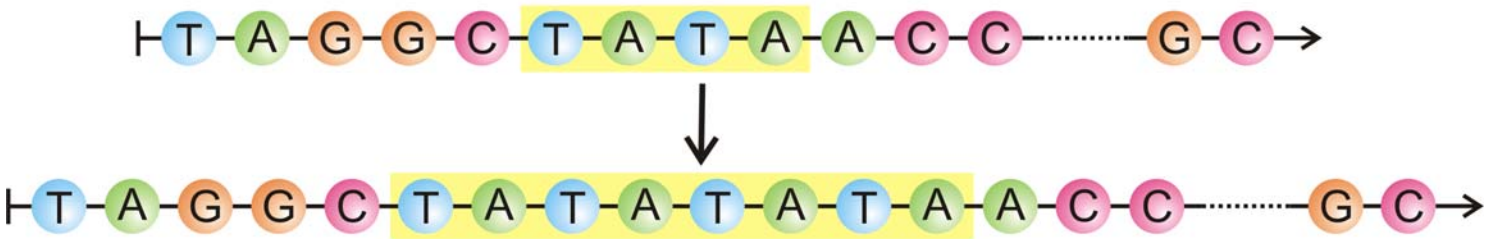
Punktmutation



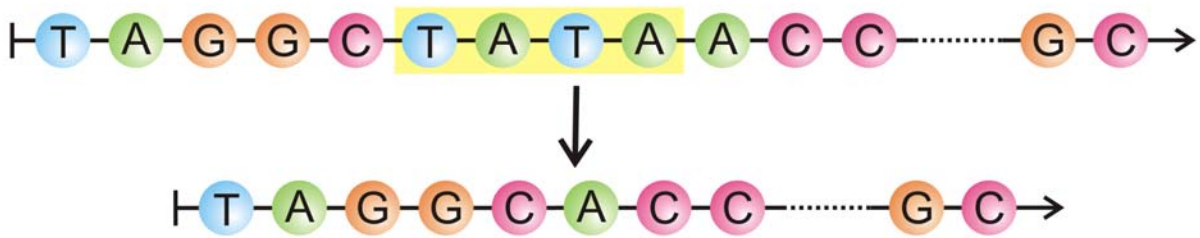
Insertion



Punktmutation



Insertion



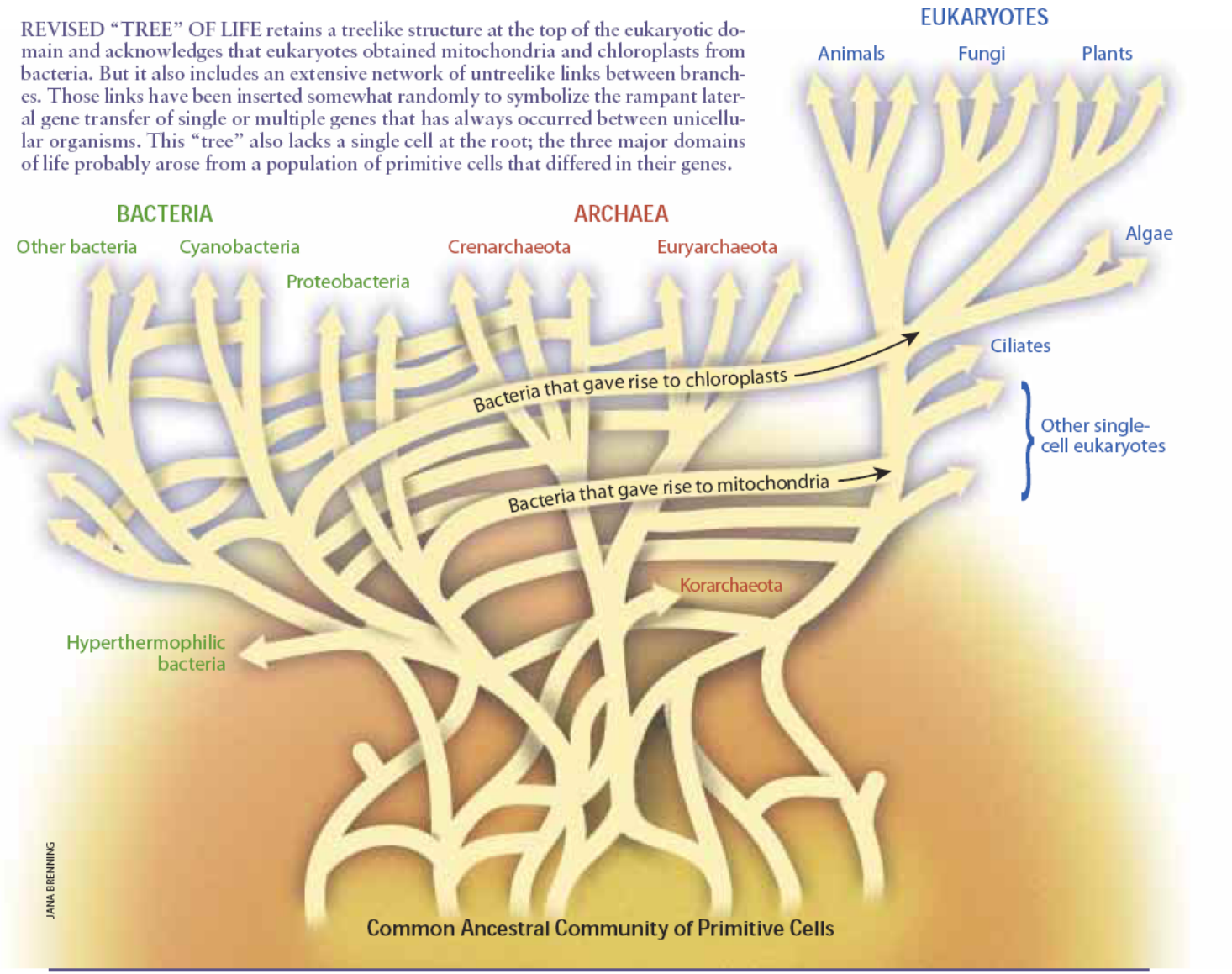
Deletion

Reconstruction of phylogenies through comparison of molecular sequence data

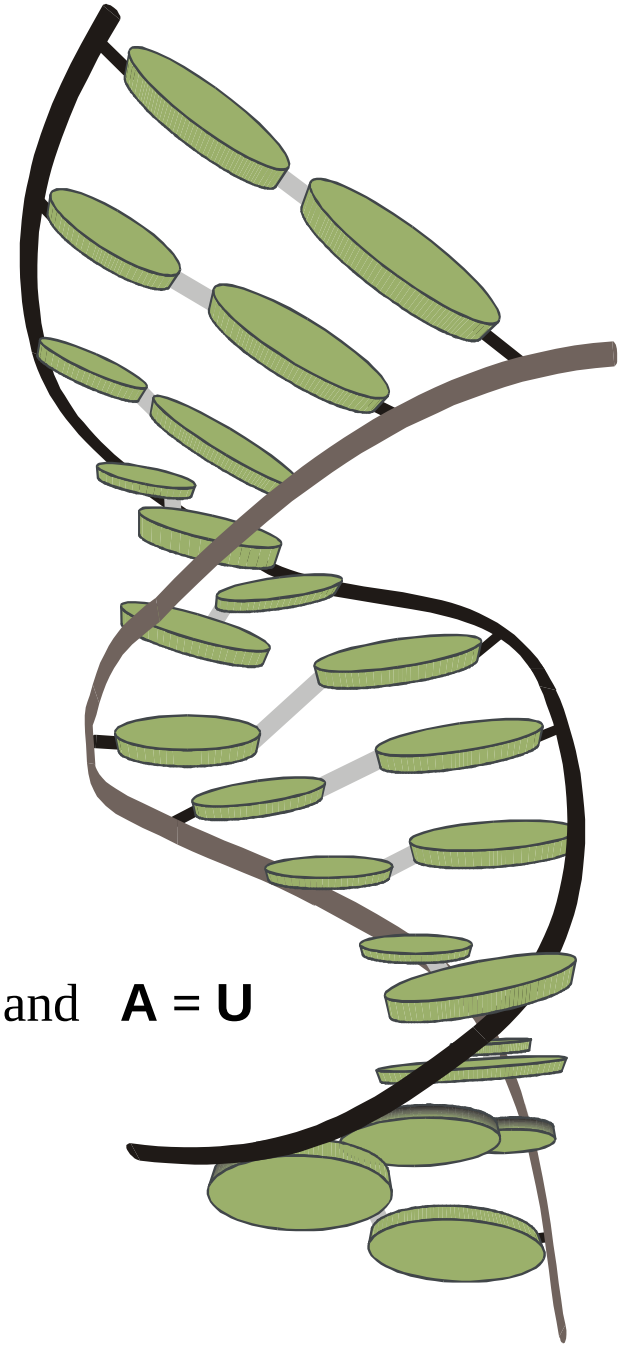
Molekulare Evolutionsforschung durch DNA-Sequenzierung

Aus dem Vergleich der heutigen DNA-Sequenzen kann die geschichtliche Abfolge der Mutationen rekonstruiert werden und diese ergibt phylogenetische Bäume, die jenen aus der vergleichenden Morphologie, welche durch Betrachtung von Formen und Gestalten der Organismen gewonnen wurden, weitest gehend entsprechen.

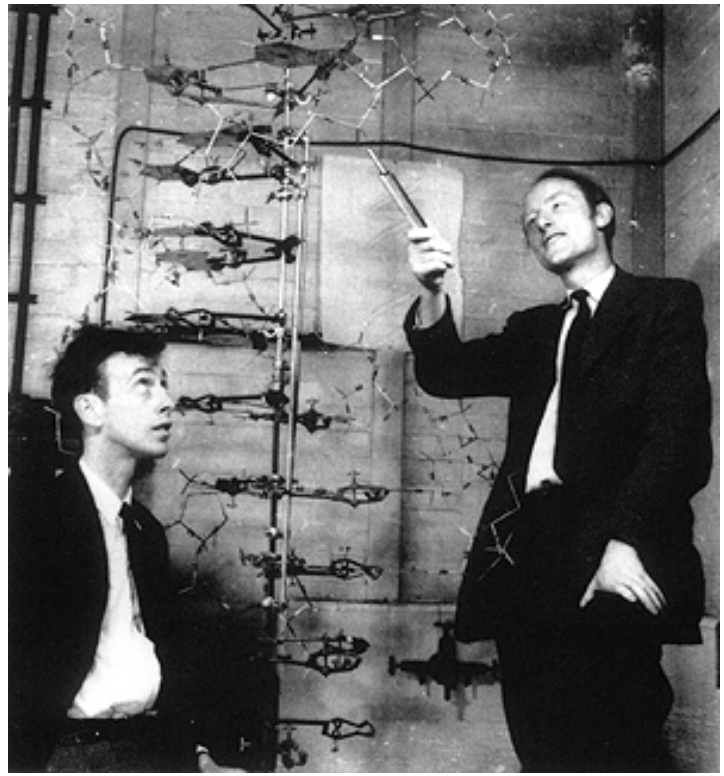
Eine in der Vergangenheit postulierte **molekulare Uhr der Evolution** verlangt, dass die Mutationshäufigkeiten auf den verschiedenen Ästen der phylogenetischen Bäume gleich groß sein soll. Die **molekulare Uhr** ist bei Wirbeltieren ganz gut erfüllt, trifft aber für die wirbellosen Tiere nicht zu. Je näher man an die Gegenwart herankommt, umso schneller tickt die molekulare Uhr.



W. F. Doolittle. Phylogenetic classification and the universal tree. *Science* **284**:2124-2128, 1999.

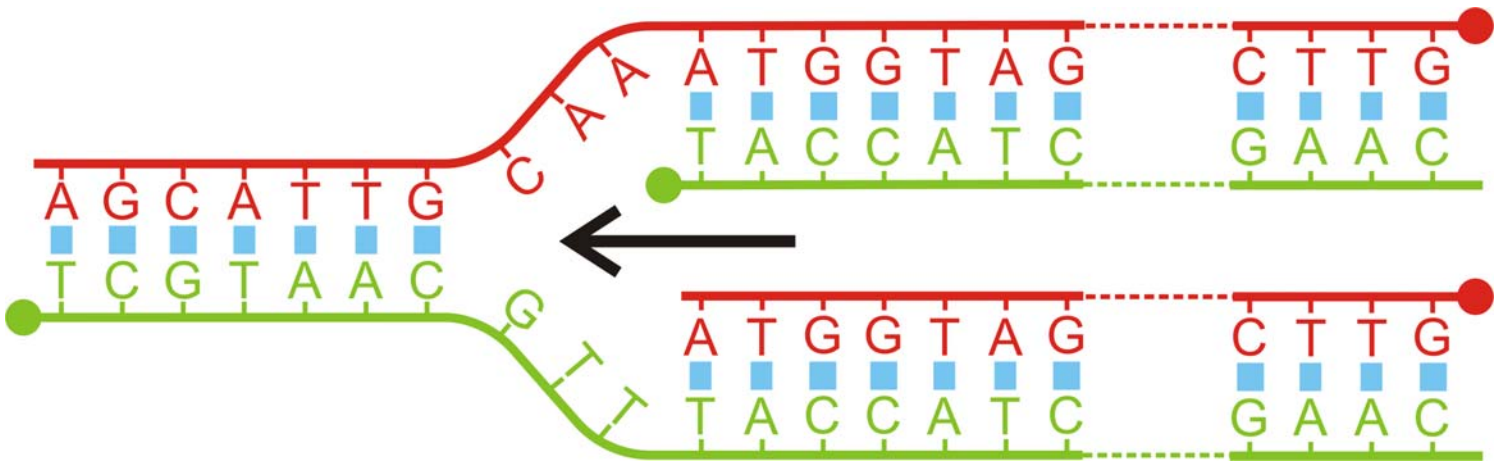


G ≡ C and **A = U**



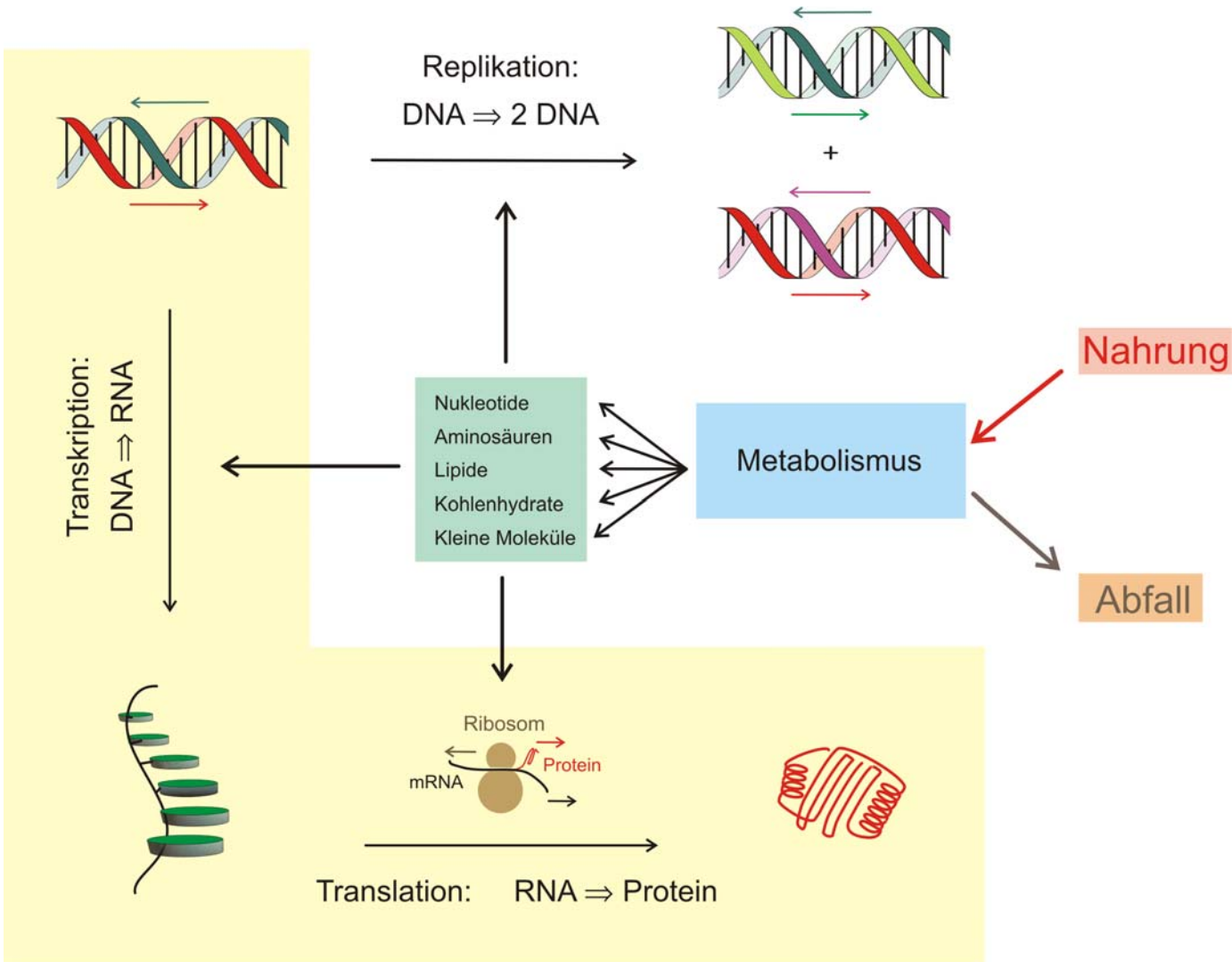
James D. Watson, 1928- , and Francis Crick, 1916-2004,
Nobel Prize 1962

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

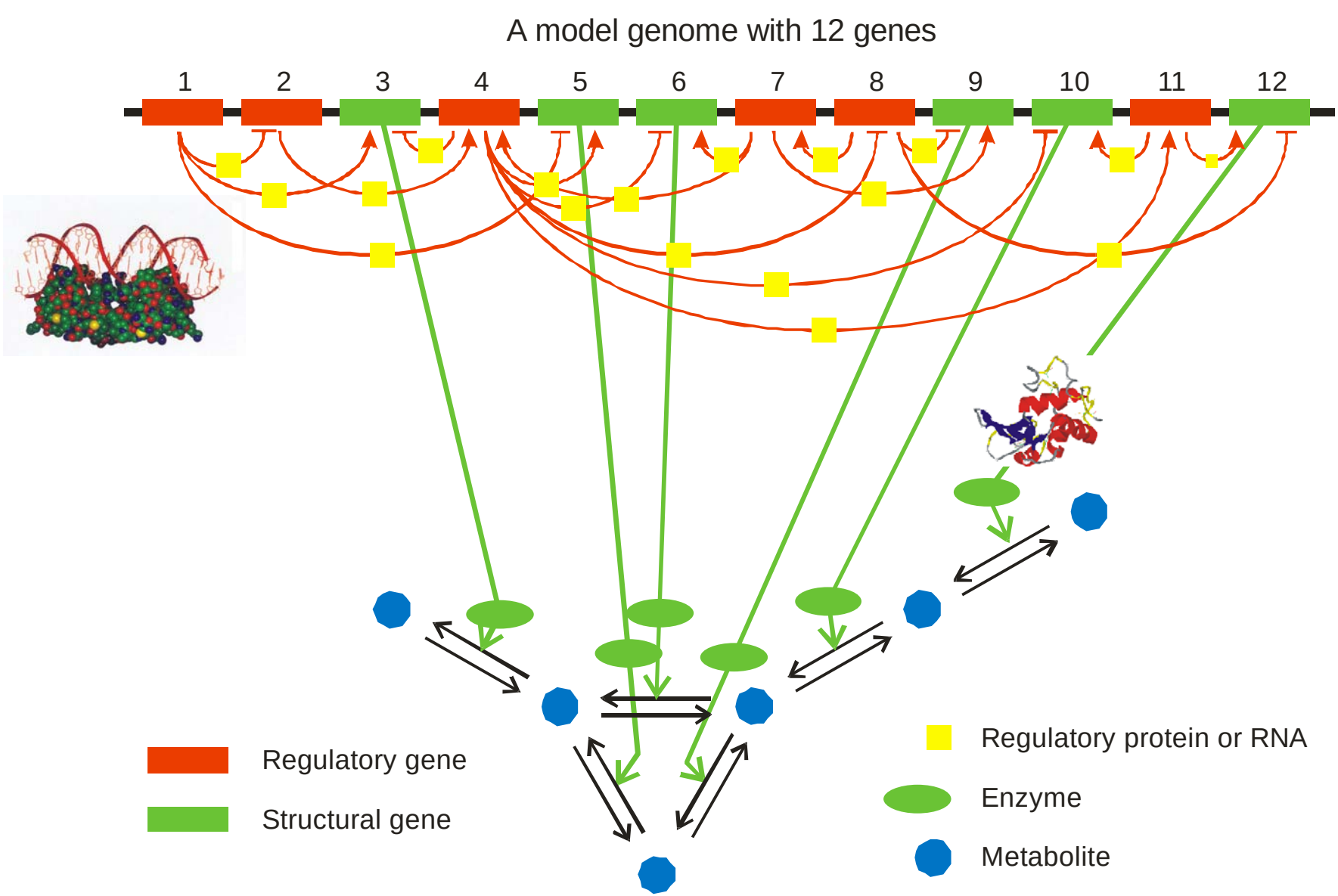


‚Replication fork' in DNA replication

The mechanism of DNA replication is ‚semi-conservative'



Skizze des zellulären Stoffwechsels

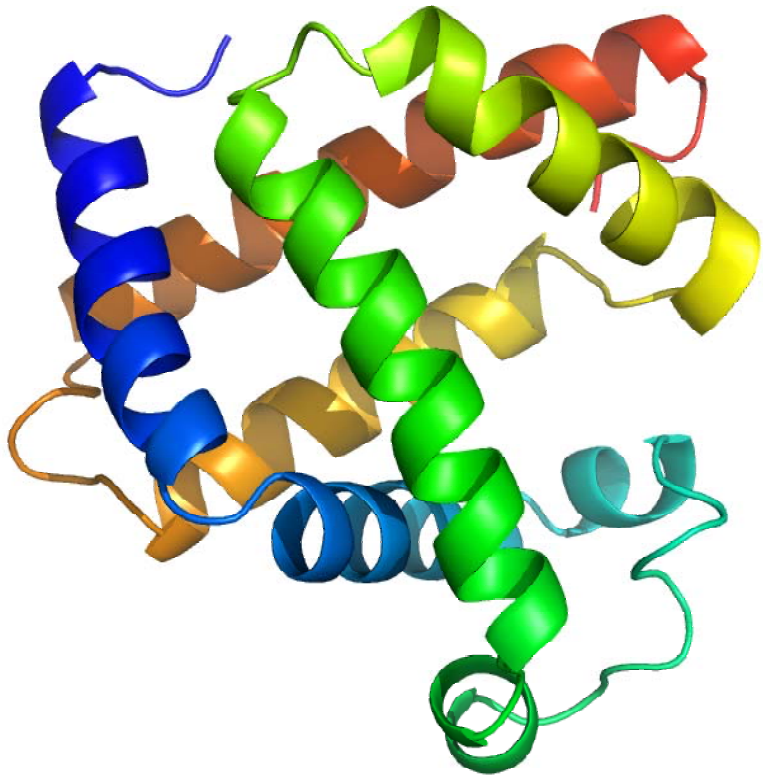


Sketch of a genetic and metabolic network

1. Evolution und der „Baum des Lebens“
2. **Wahrscheinlichkeiten und Zufall**
3. Vermehrung, Mutation und Selektion
4. Evolution von Molekülen und Optimierung
5. Evolutionäres “Basteln” und Komplexität
6. Schlußbemerkungen

Kette aus 153 Aminosäurerestenmit der Sequenz:

GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLK
SEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIP
VKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELG
FQG



Das Myglobinmolekül

Eugene Wigner's or Fred Hoyle's argument applied to myoglobin:

All sequences have equal probability and all except one have no survival value or are lethal

GLSDGEWQLVLNVWG.....FQG

Alphabet size: 20

Chain length: 153 amino acids

Number of possible sequences: $20^{153} = \mathbf{0.11 \times 10^{200}}$

Probability to find the myoglobin sequence:

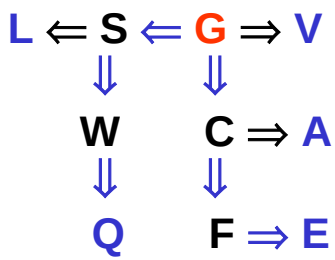
$20^{-153} = \mathbf{9 \times 10^{-200}} = 0.000\ldots\ldots009$

200

Eugene Wigner's and Fred Hoyle's arguments revisited:

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

GLSDGEWQLVLNVWG.....FQG



Alphabet size: 20

Chain length: 153 amino acids

Length of longest path to myoglobin sequence: $19 \times 153 = 2907$

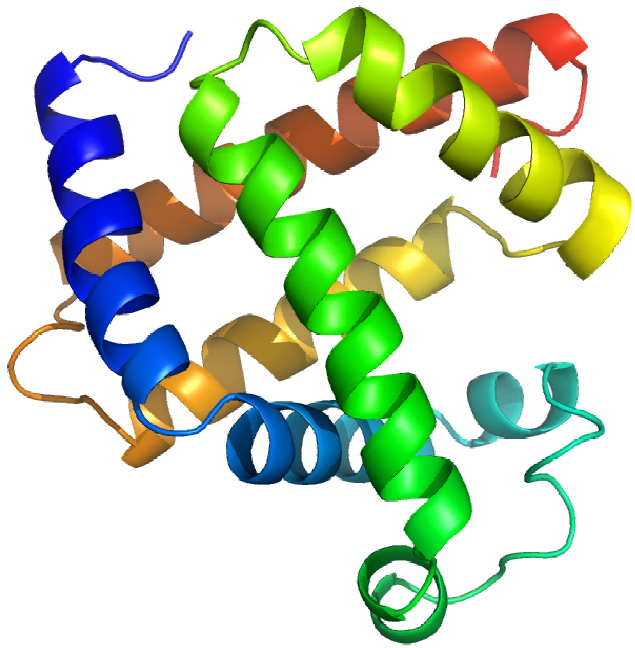
Probability to find the myoglobin sequence: **0.00034**

Das Faltungsproblem des Myoglobinmoleküls:

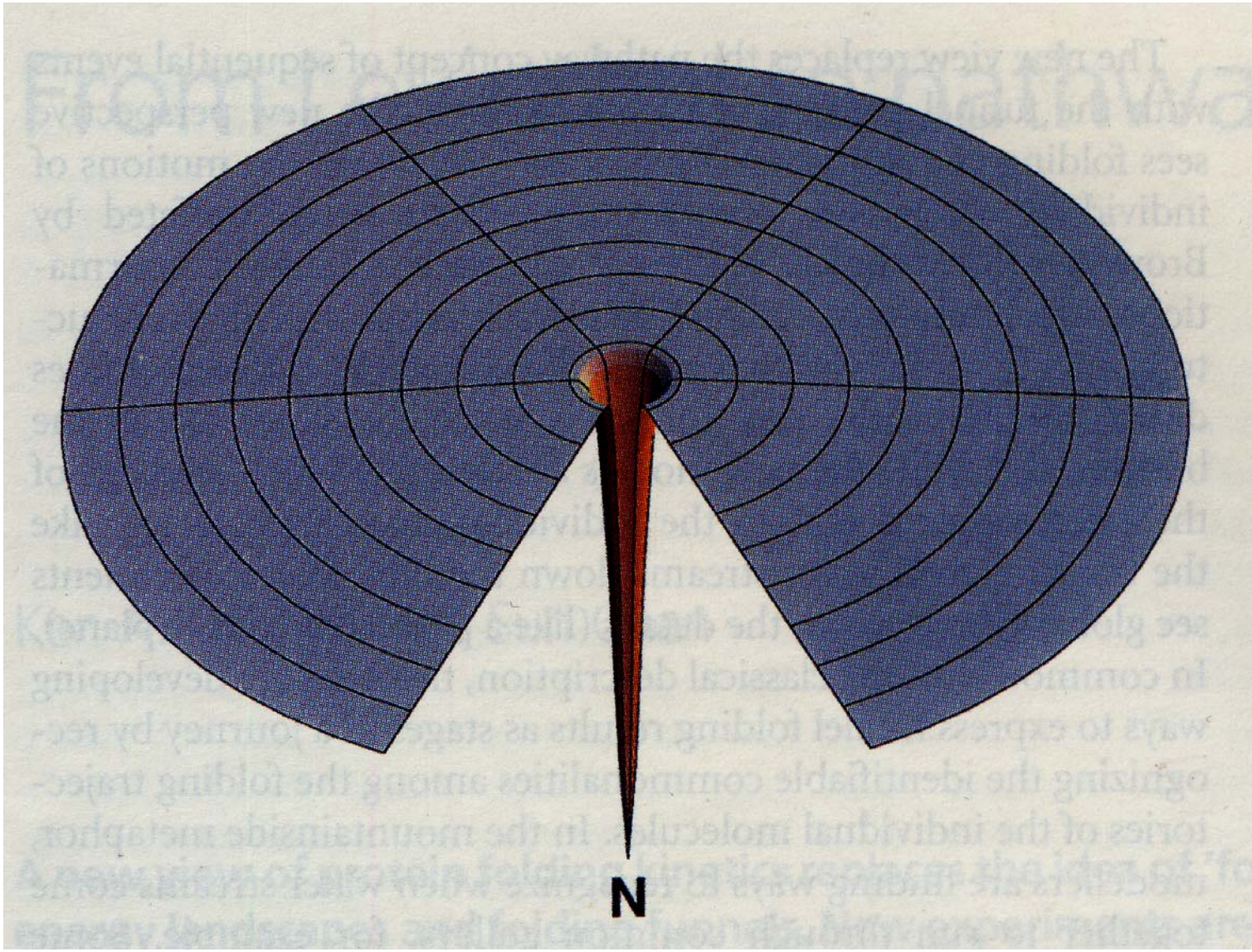
Eine Kette aus 153 Aminosäureresten, von welchen jeder im Mittel 15 verschiedene Konformationen einnimmt, kann in

$$15^{153} = 0.9 \times 10^{180} \text{ Zuständen}$$

vorkommen. Einer davon muss bei der Faltung in die stabile Struktur gefunden werden.



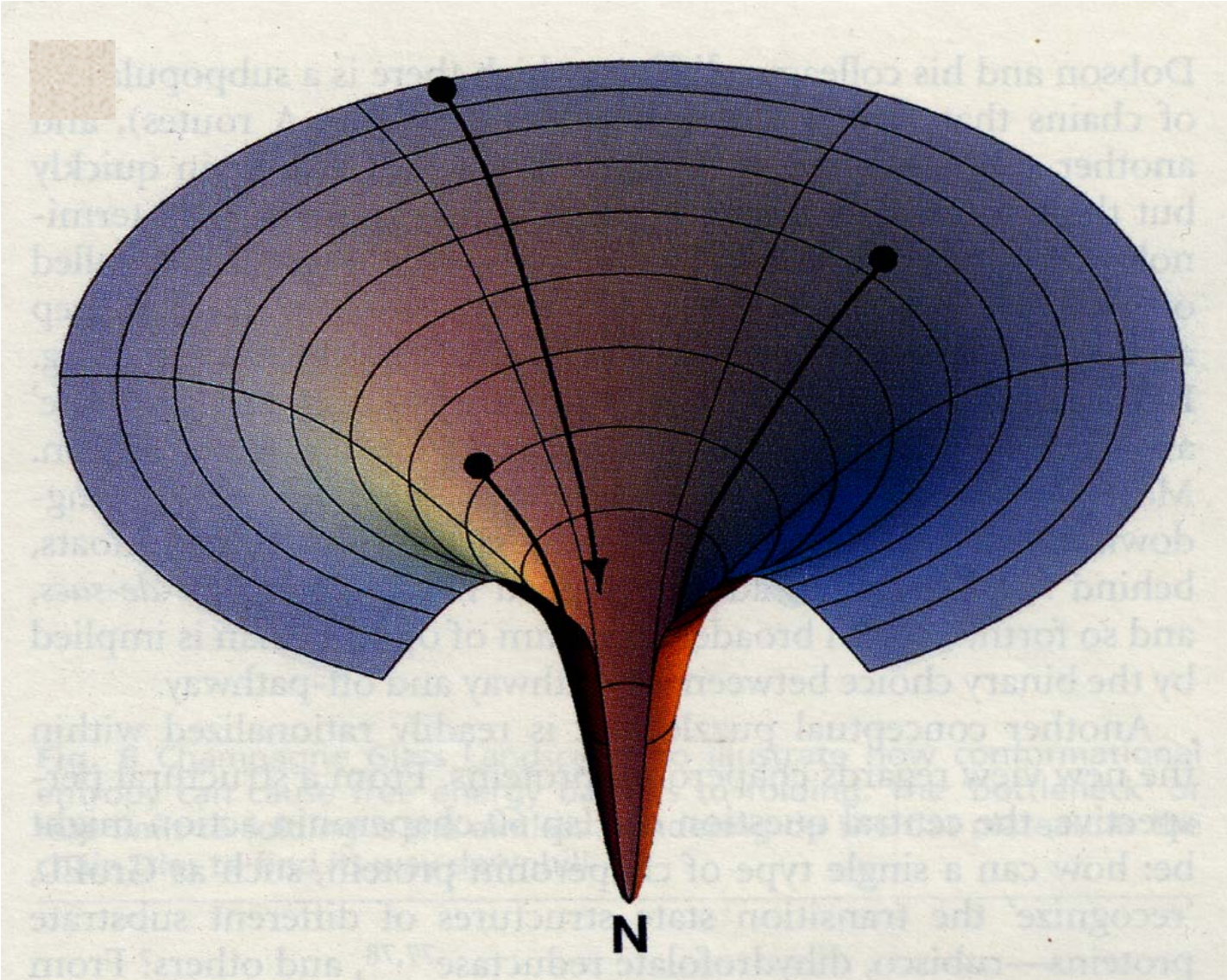
Das Levinthal-Paradoxon der Proteinfaltung



The gulf course landscape

Solution to Levinthal's paradox

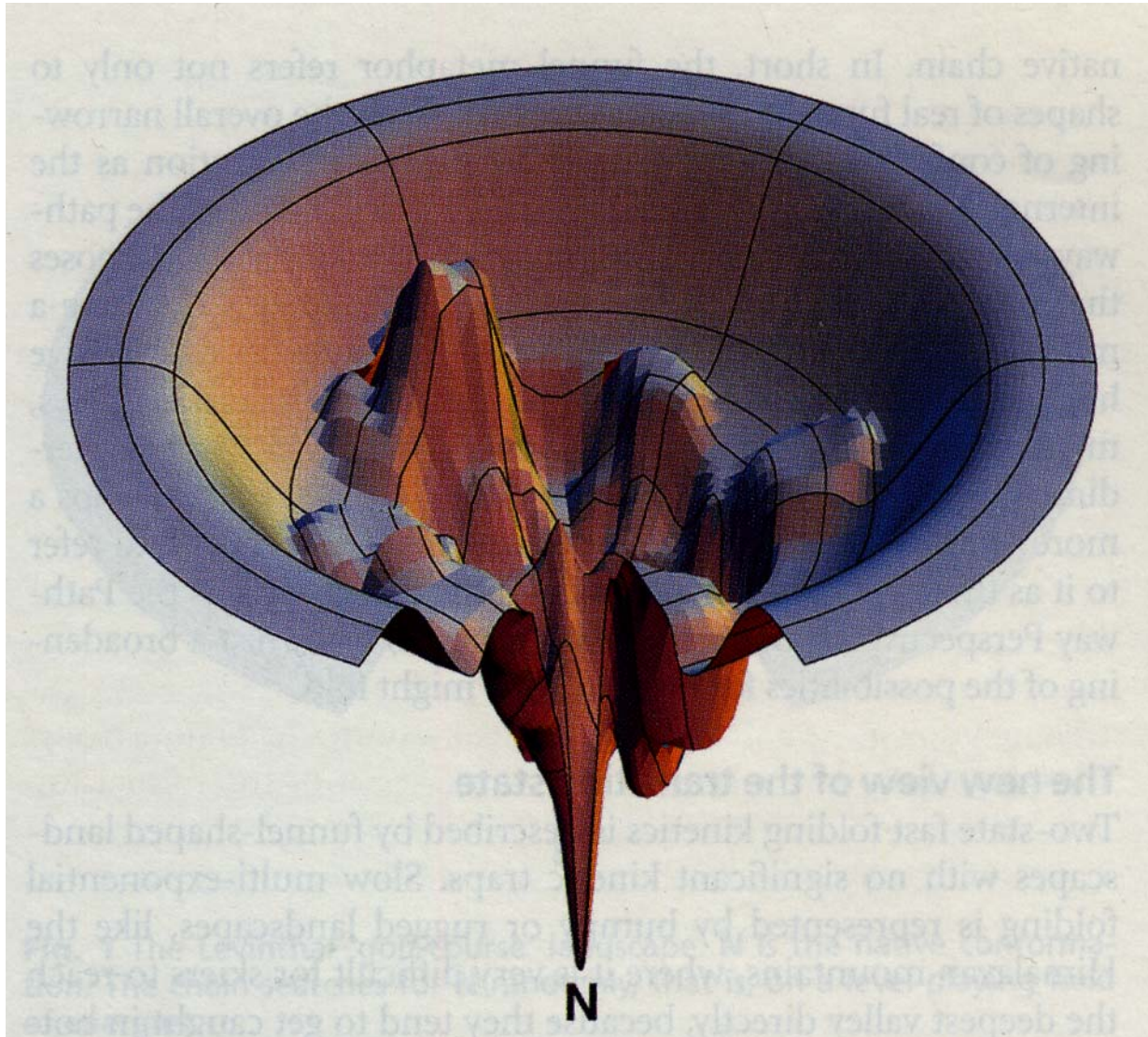
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19



The funnel landscape

Solution to Levinthal’s paradox

Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19

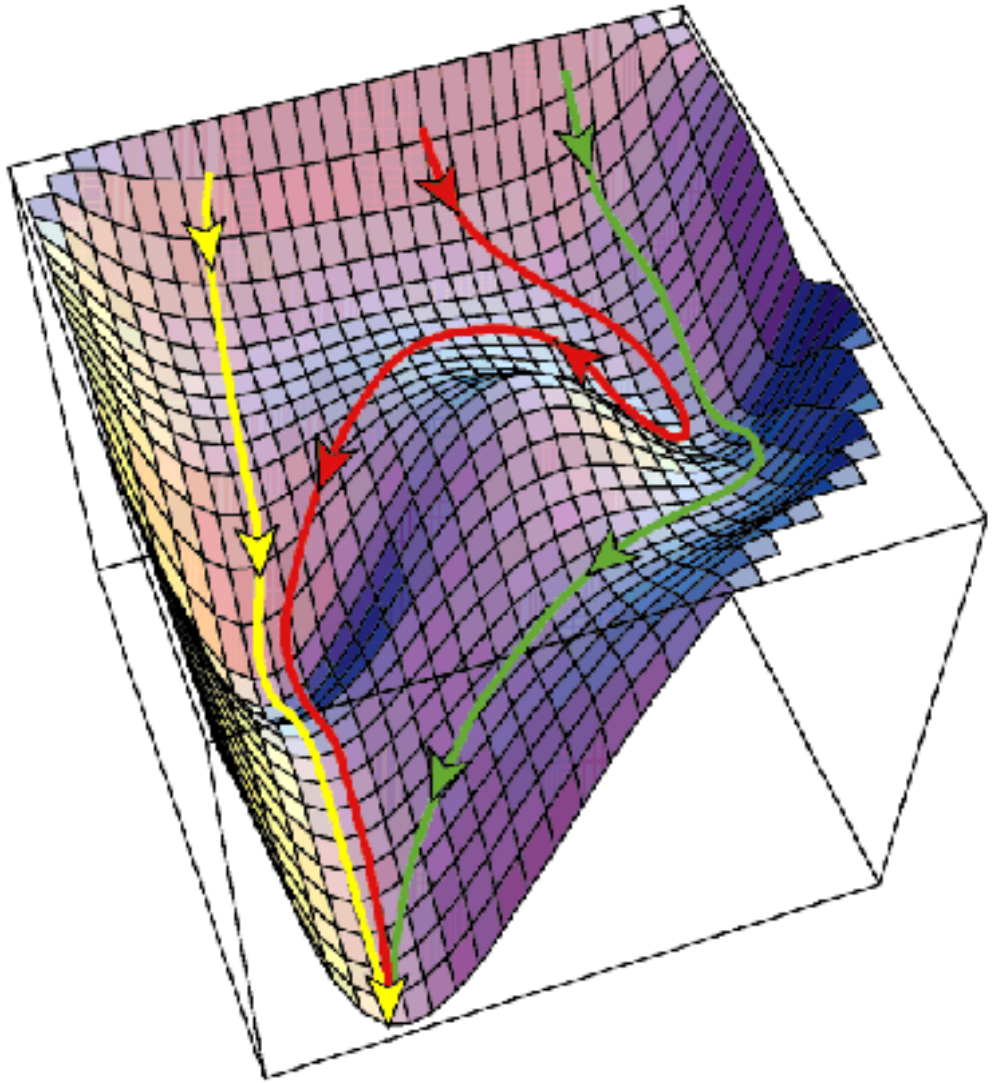


The structured funnel landscape

Solution to Levinthal’s paradox

Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19

An “all-roads-lead-to-Rome” landscape



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

Picture: C.M. Dobson, A. Šali, and M. Karplus, Angew.Chem.Internat.Ed. 37: 868-893, 1988

Das Brettspiel „Mensch ärgere dich nicht“ als ein Beispiel für das Zusammenwirken einer deterministischen (Regeln) und einer zufälligen (Würfel) Komponente:

Sicher ist, dass einer der vier Spieler dadurch gewinnen wird, dass er seine vier Figuren auf die vier vorgesehenen Plätze bringt.

Zufällig ist, welcher der vier Spieler das sein wird.

Die Dauer des Spieles zeigt eine für stochastische Prozesse typische Wahrscheinlichkeitsverteilung.

1. Evolution und der „Baum des Lebens“
2. Wahrscheinlichkeiten und Zufall
3. **Vermehrung, Mutation und Selektion**
4. Evolution von Molekülen und Optimierung
5. Evolutionäres “Basteln” und Komplexität
6. Schlußbemerkungen

Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

Darwinian evolution in the test tube

All three conditions are fulfilled not only by cellular organisms but also by **nucleic acid molecules** – **DNA** or **RNA** – **in** suitable **cell-free experimental assays**.

Evolution of RNA molecules based on Qβ phage

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

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

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

G.Bauer, H.Otten, J.S.McCaskill, ***Travelling waves of* in vitro *evolving RNA.*** *Proc.Natl.Acad.Sci.USA* **86** (1989), 7937-7941

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

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

F.Öhlenschlager, M.Eigen, ***30 years later – A new approach to Sol Spiegelman‘s and Leslie Orgel‘s* in vitro *evolutionary studies***. Orig.Life Evol.Biosph. **27** (1997), 437-457

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

I. 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.

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 disintegrate 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

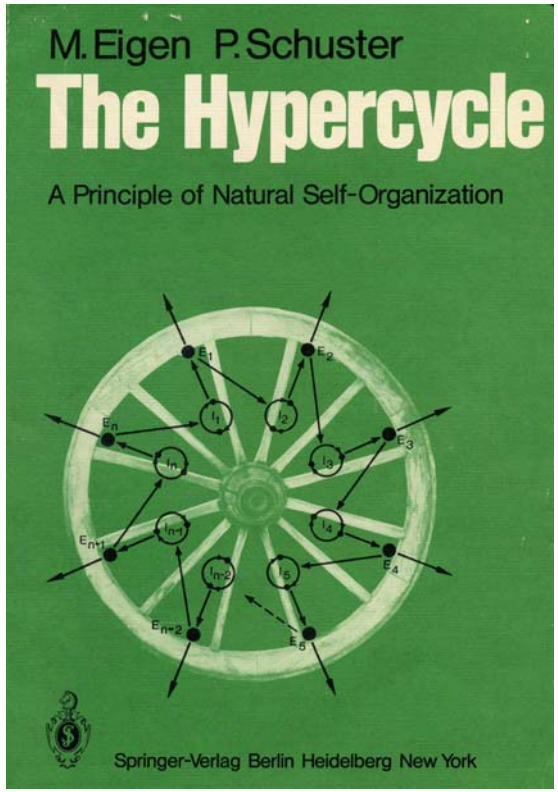
The mathematical analysis of dynamical systems using methods of differential topology, yields the result that there is only one type of mechanisms which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective 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 functions.
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.

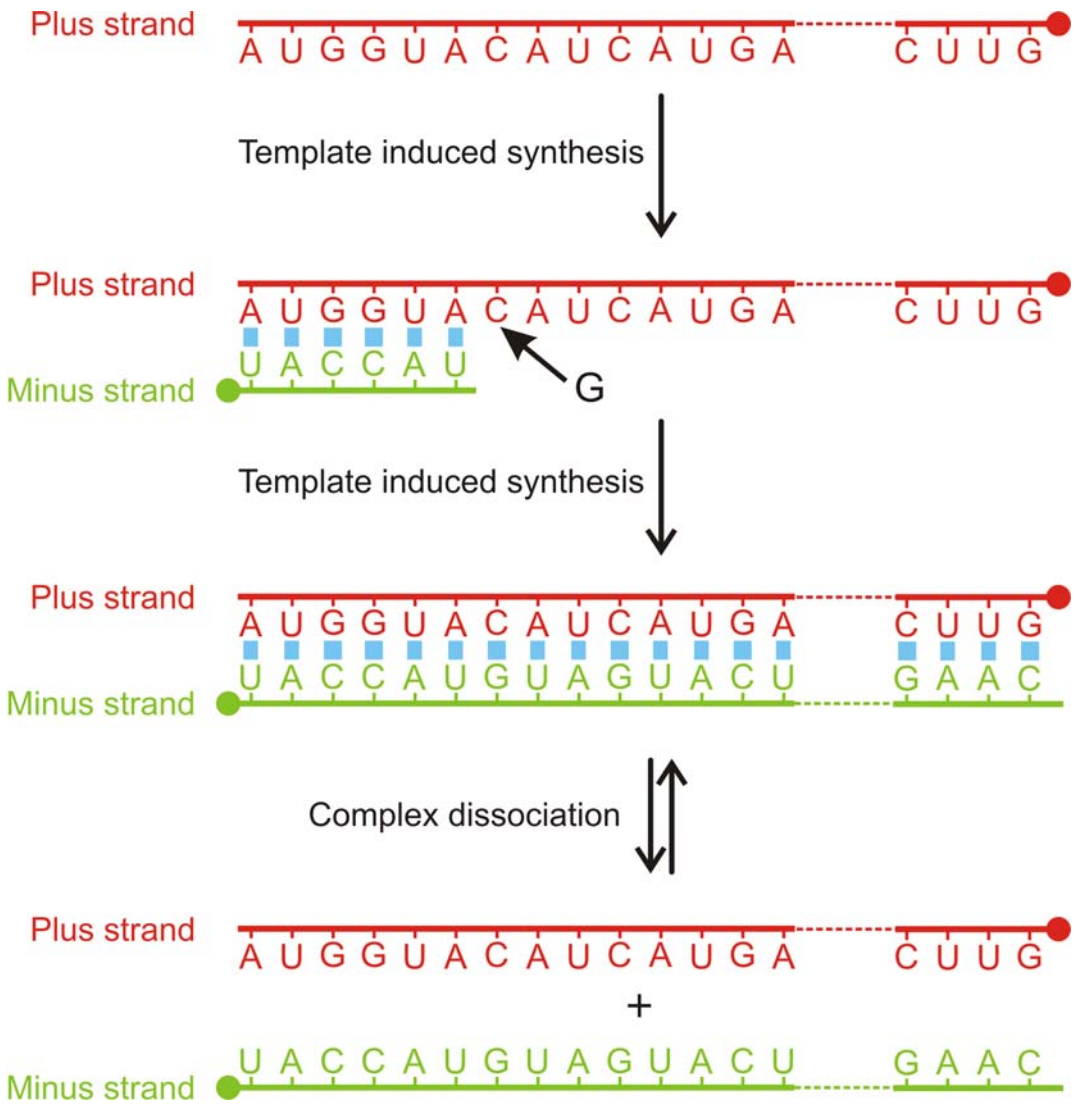
I. The Paradigm of Unity and Diversity in Evolution

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 answer 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-



Chemical kinetics of molecular evolution

M. Eigen, *Naturwissenschaften* **58**:465-523, 1971
M. Eigen, P. Schuster, *Naturwissenschaften* **64**:541-565, 1977
M. Eigen, P. Schuster, `The Hypercycle´, Springer-Verlag, Berlin 1979

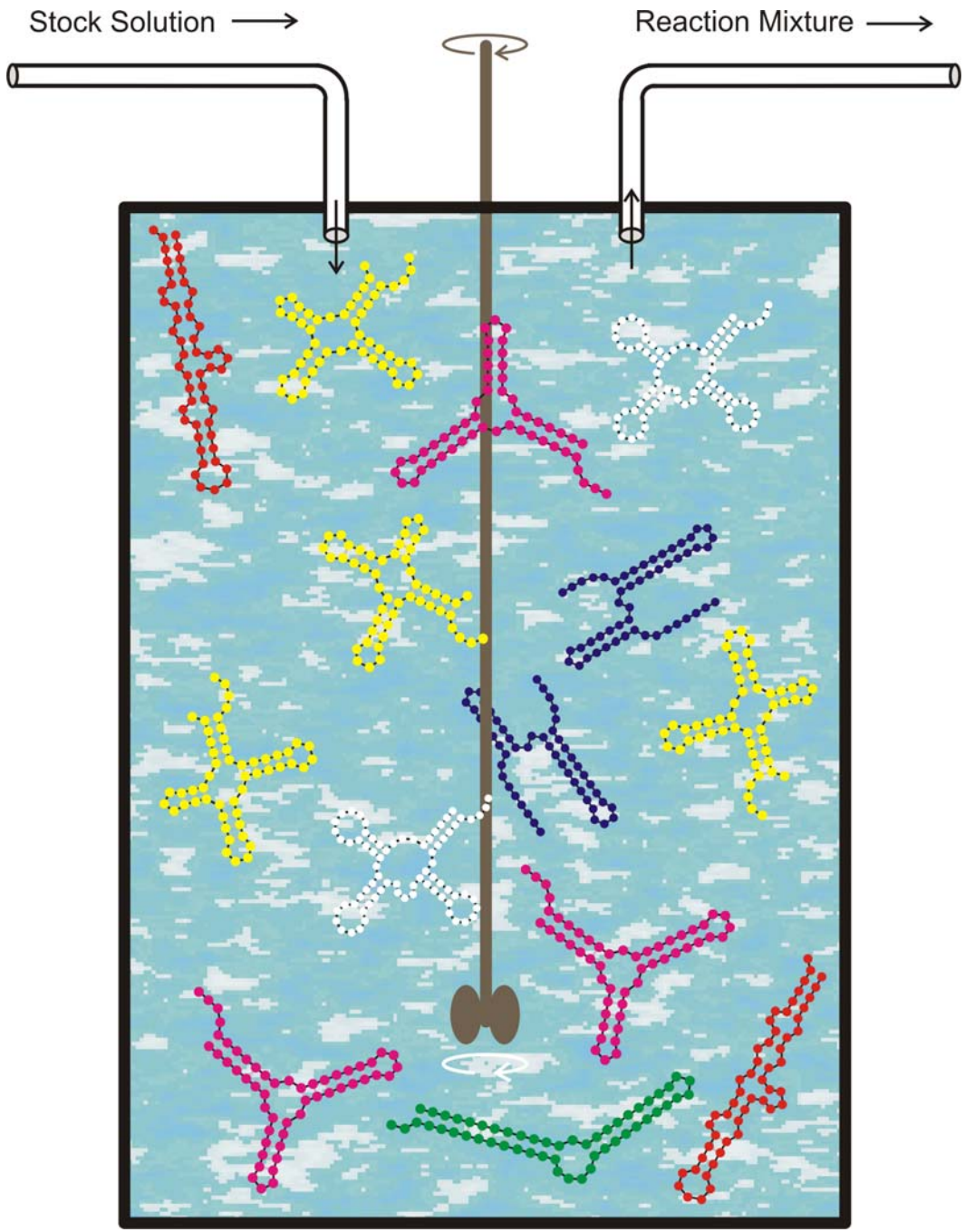


Complementary replication is the simplest copying mechanism of RNA.
Complementarity is determined by Watson-Crick base pairs:

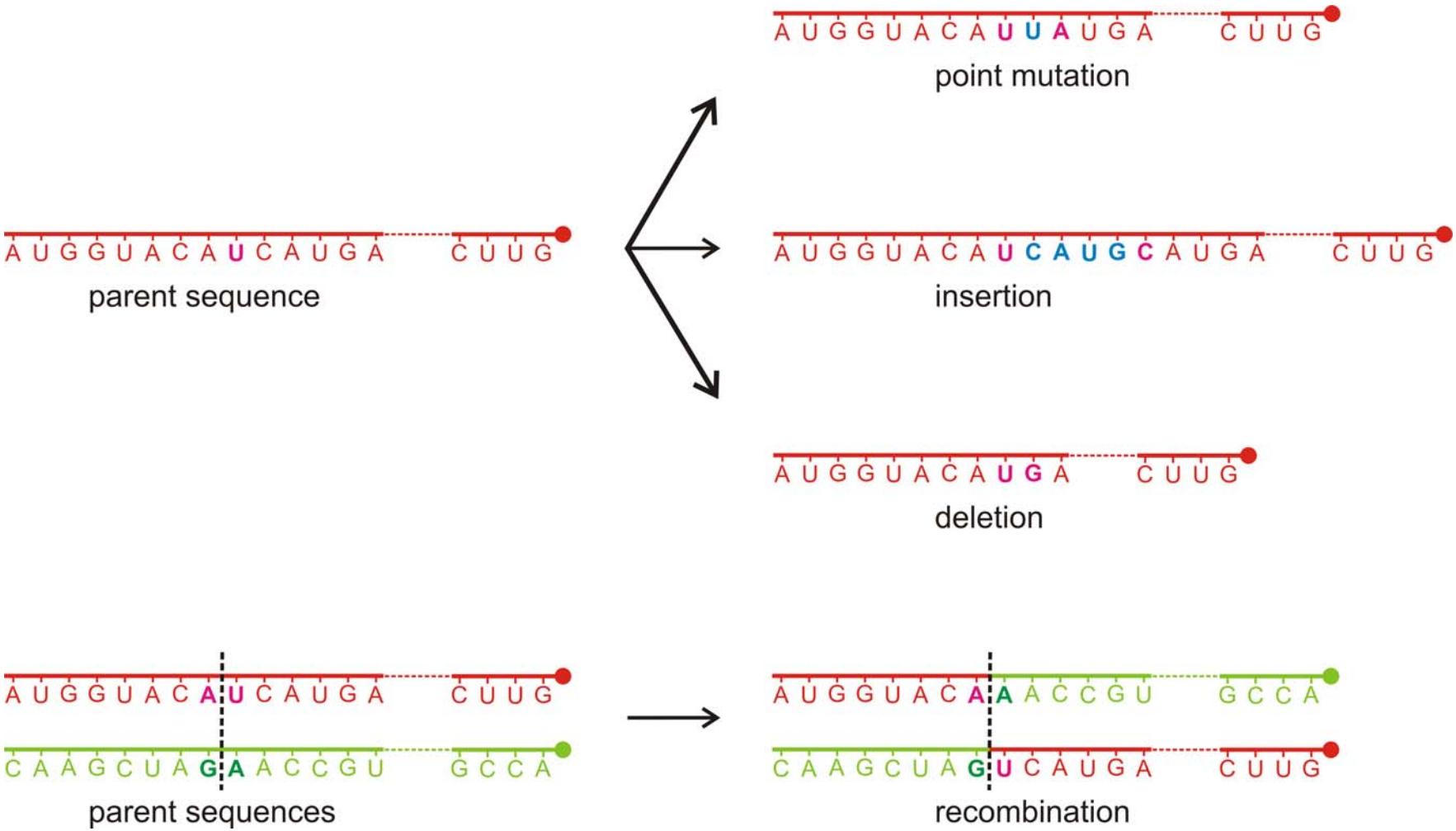
G≡C and **A=U**

Stock solution:

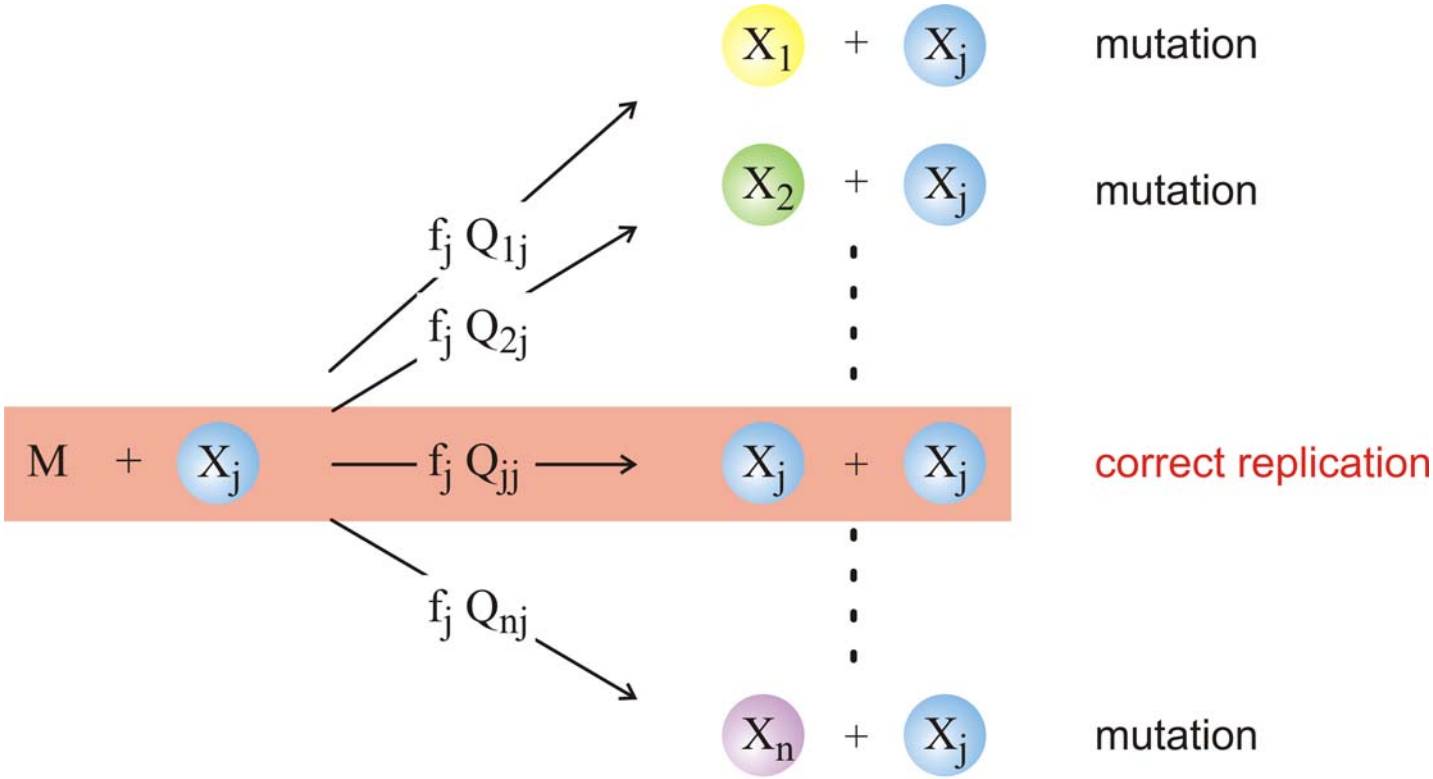
activated monomers, **ATP, CTP, GTP, UTP (TTP);**
a replicase, an enzyme that performs complemantary replication;
buffer solution



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



Variation of genotypes through mutation and recombination

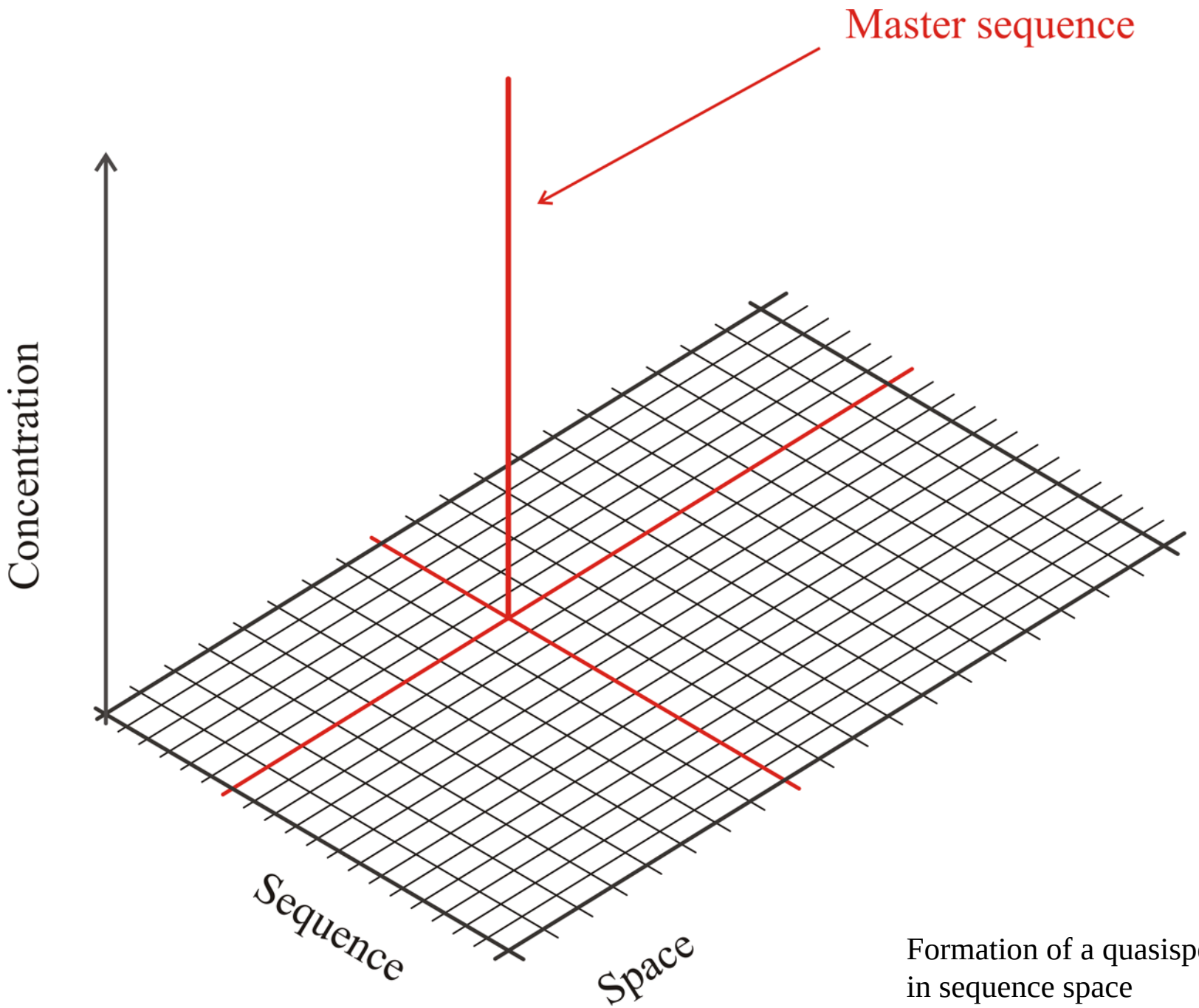


f_j ... replication rate function or fitness function

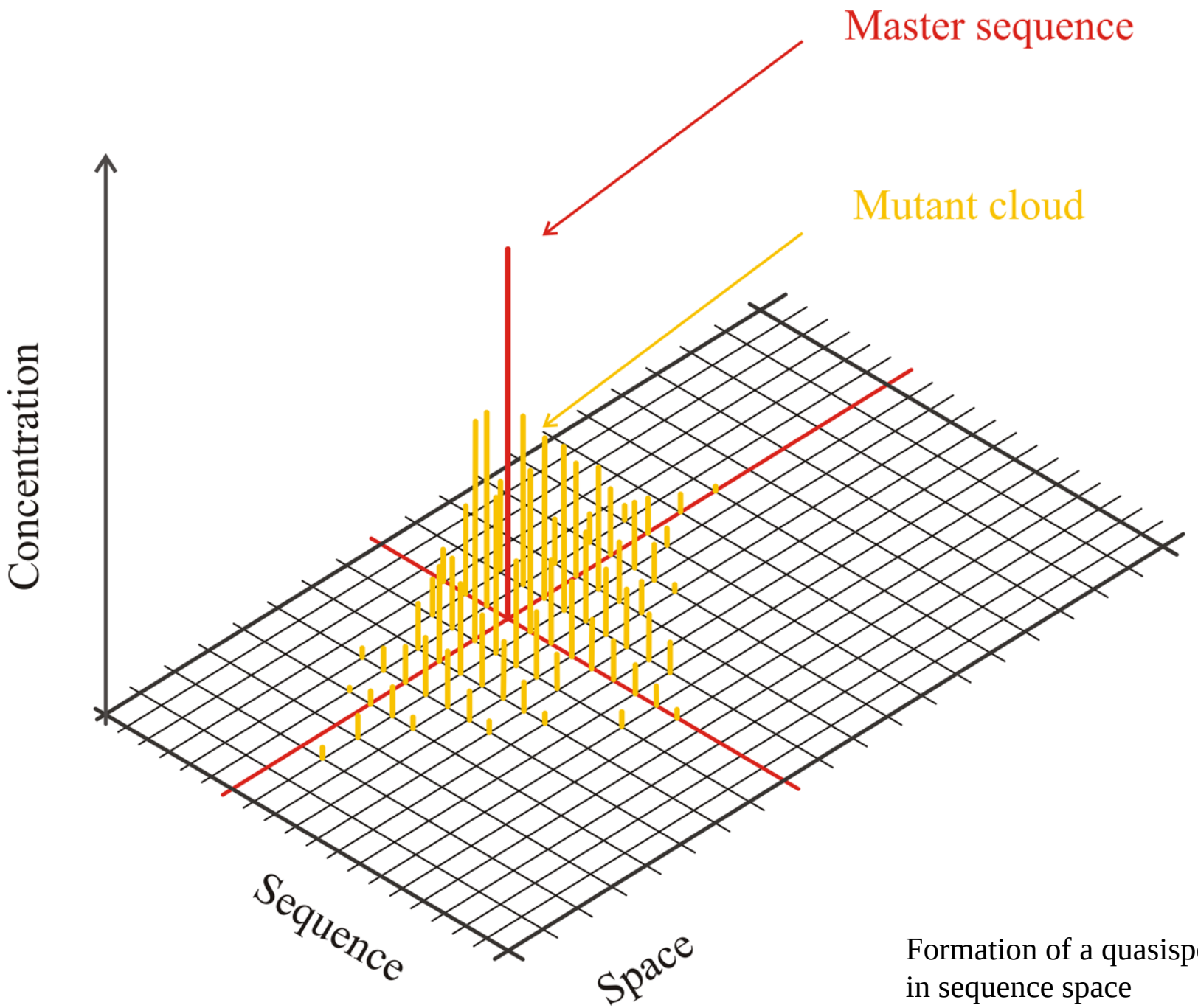
Q_{ij} ... mutation frequency: $X_j \rightarrow X_i$

Chemical kinetics of molecular evolution

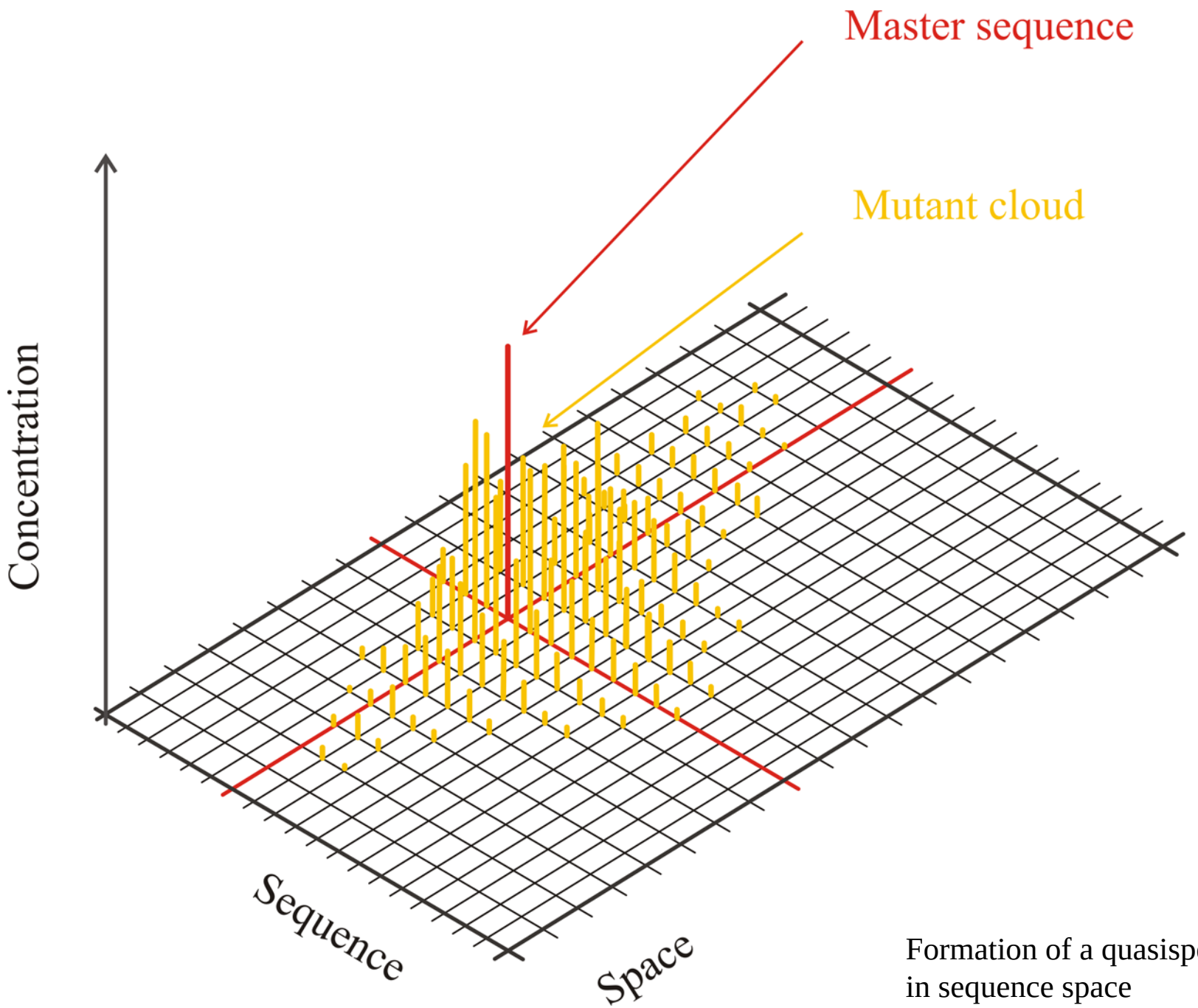
M. Eigen, P. Schuster, `The Hypercycle´, Springer-Verlag, Berlin 1979



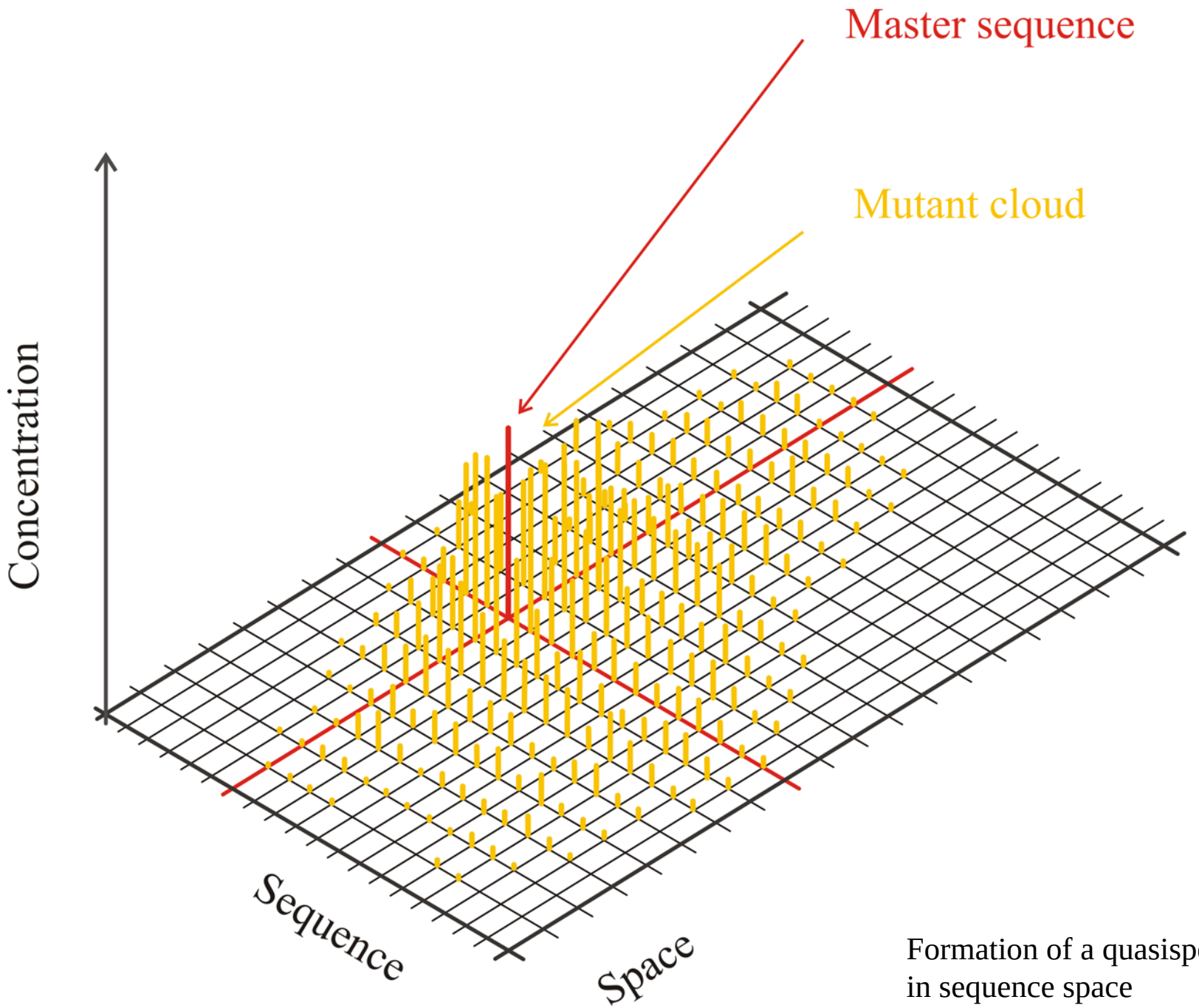
Formation of a quasispecies in sequence space



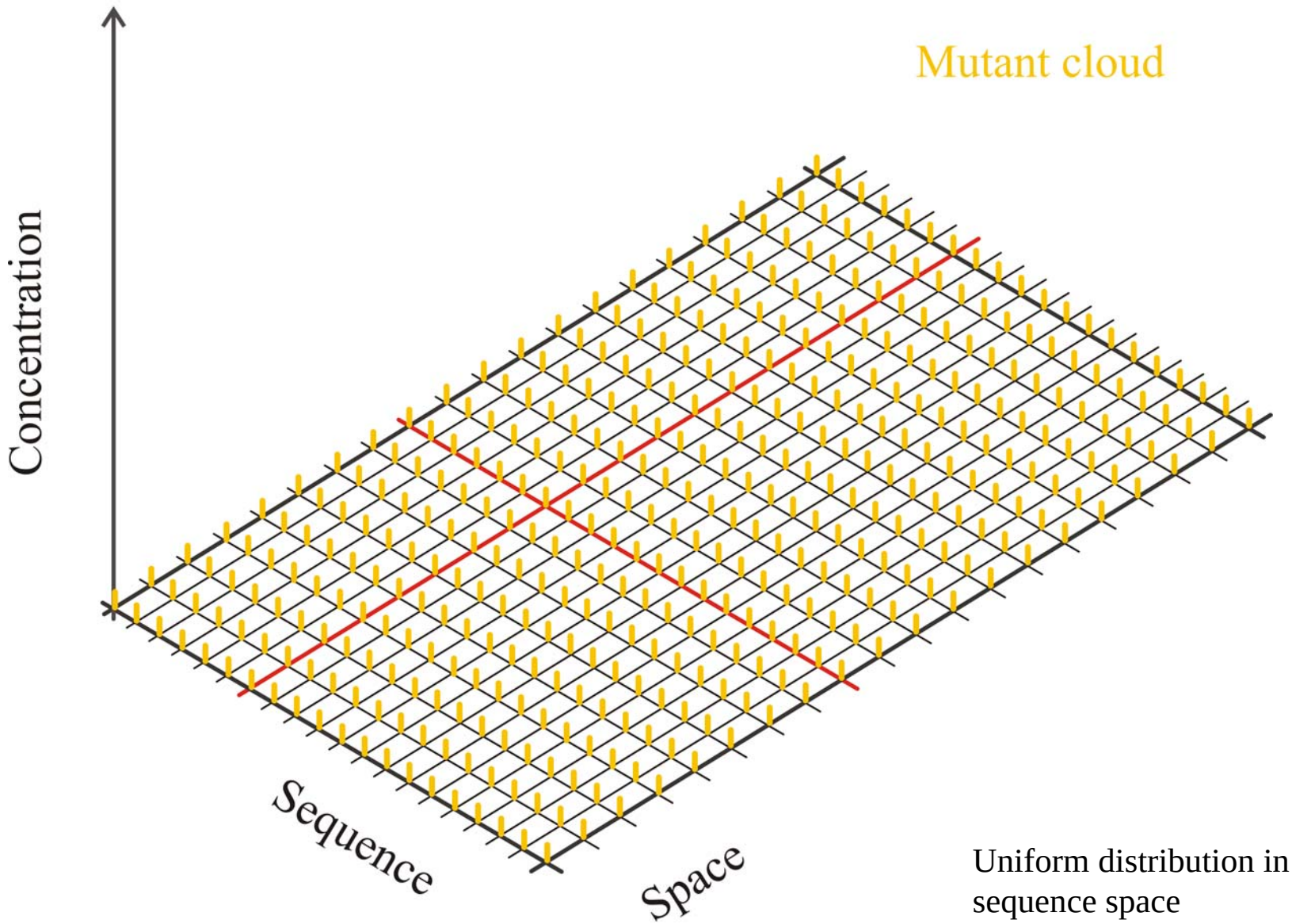
Formation of a quasispecies in sequence space



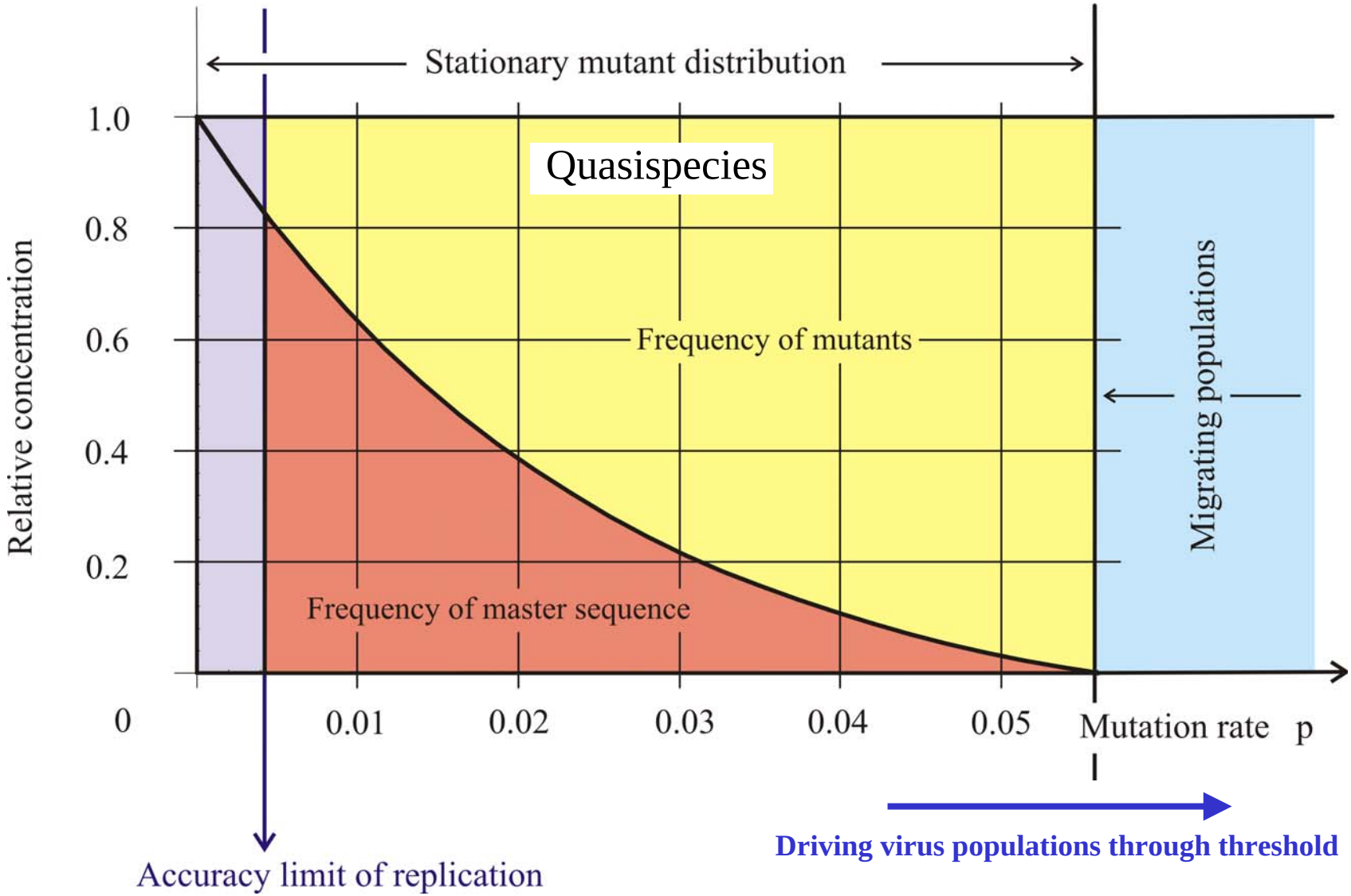
Formation of a quasispecies in sequence space



Formation of a quasispecies in sequence space

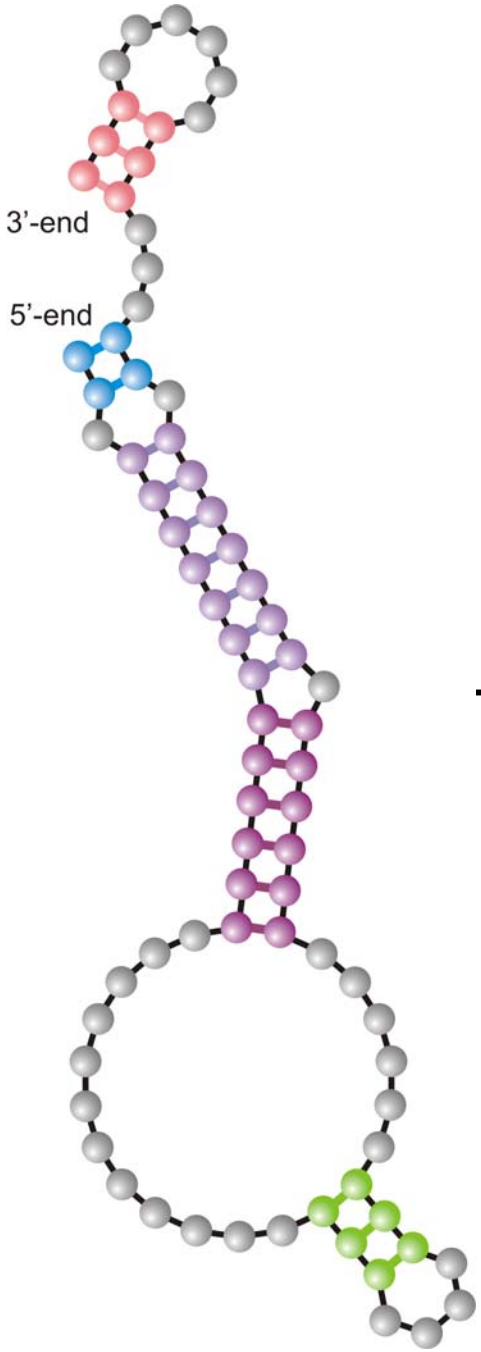


Uniform distribution in sequence space

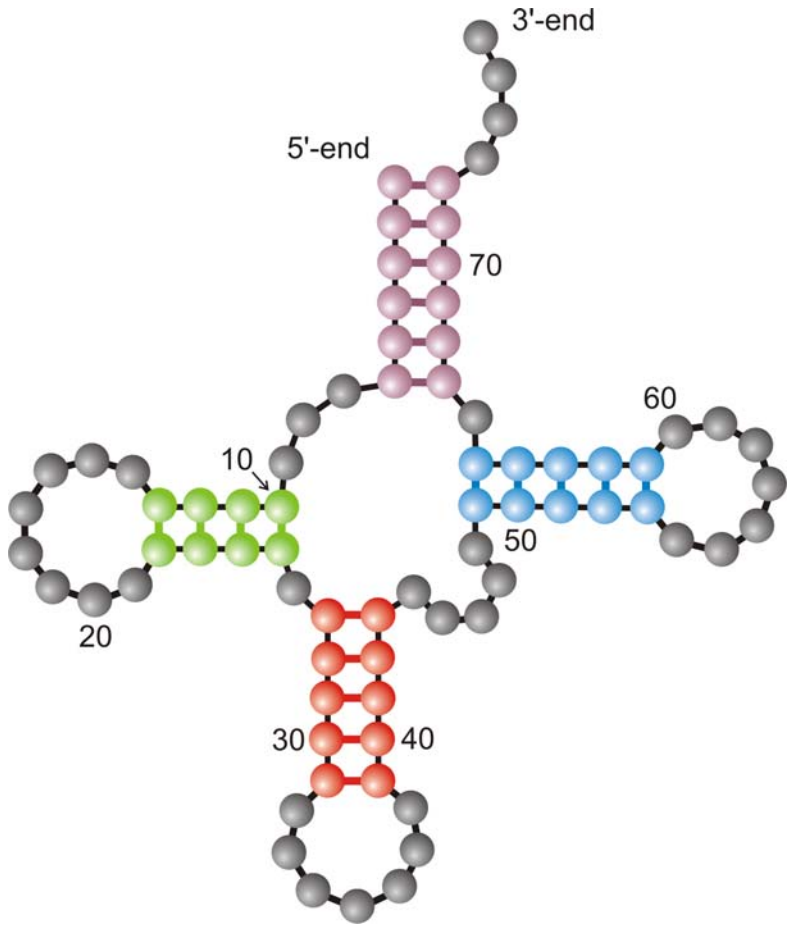


The error threshold in replication

1. Evolution und der „Baum des Lebens“
2. Wahrscheinlichkeiten und Zufall
3. Vermehrung, Mutation und Selektion
4. **Evolution von Molekülen und Optimierung**
5. Evolutionäres “Basteln” und Komplexität
6. Schlußbemerkungen



Structure of andomly chosen initial sequence



Phenylalanyl-tRNA as target structure

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.

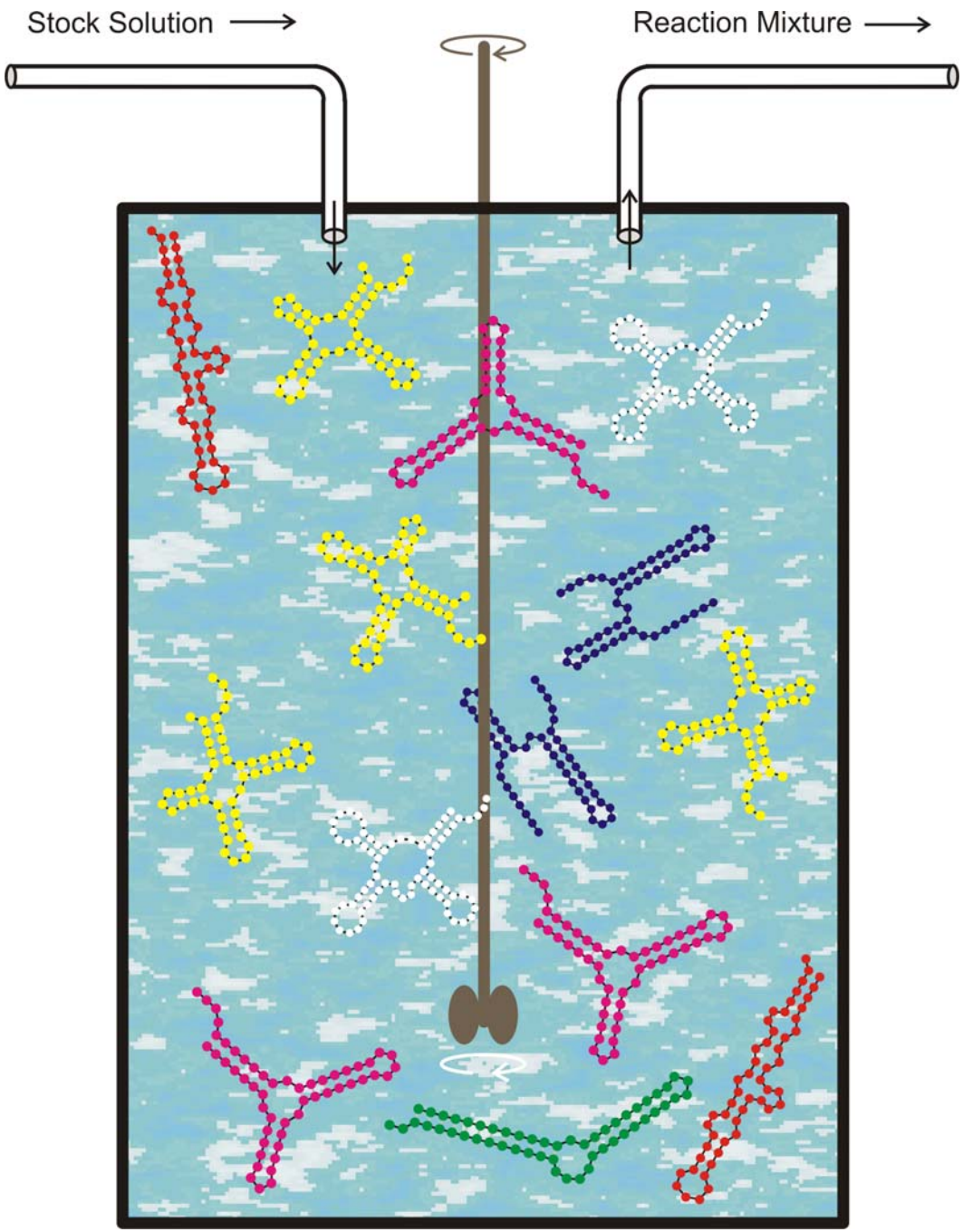
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level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



Replication rate constant:

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

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

Selection constraint:

Population size, N = # RNA molecules, is controlled by the flow

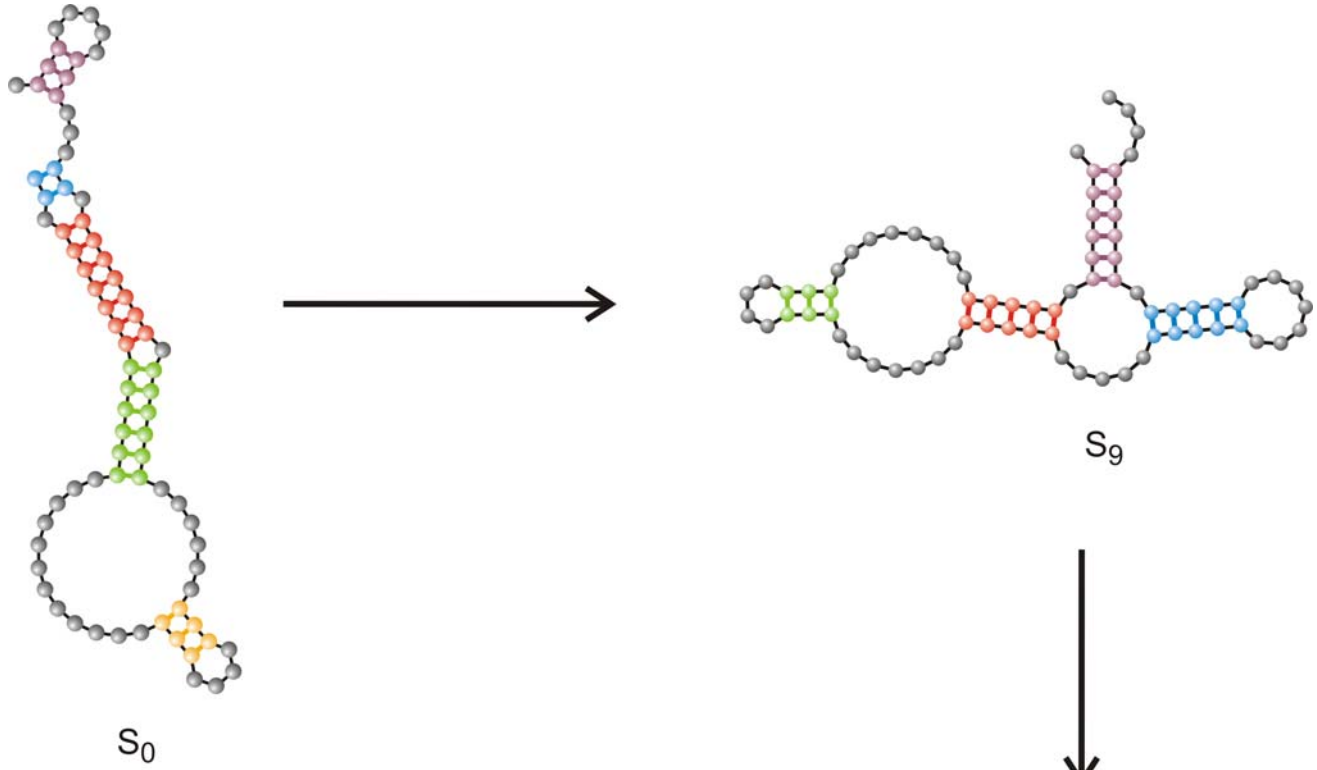
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

Mutation rate:

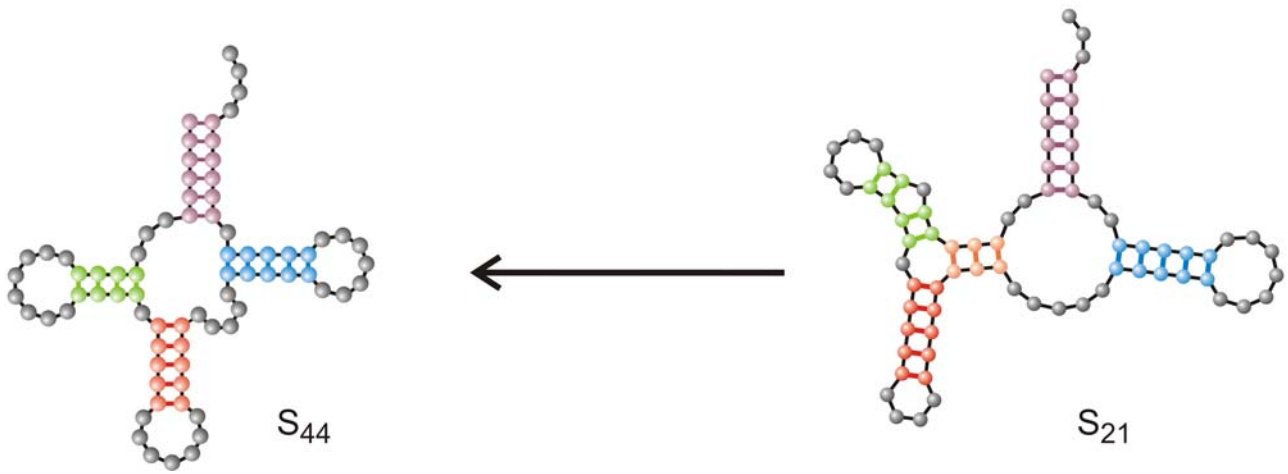
p = 0.001 / site × replication

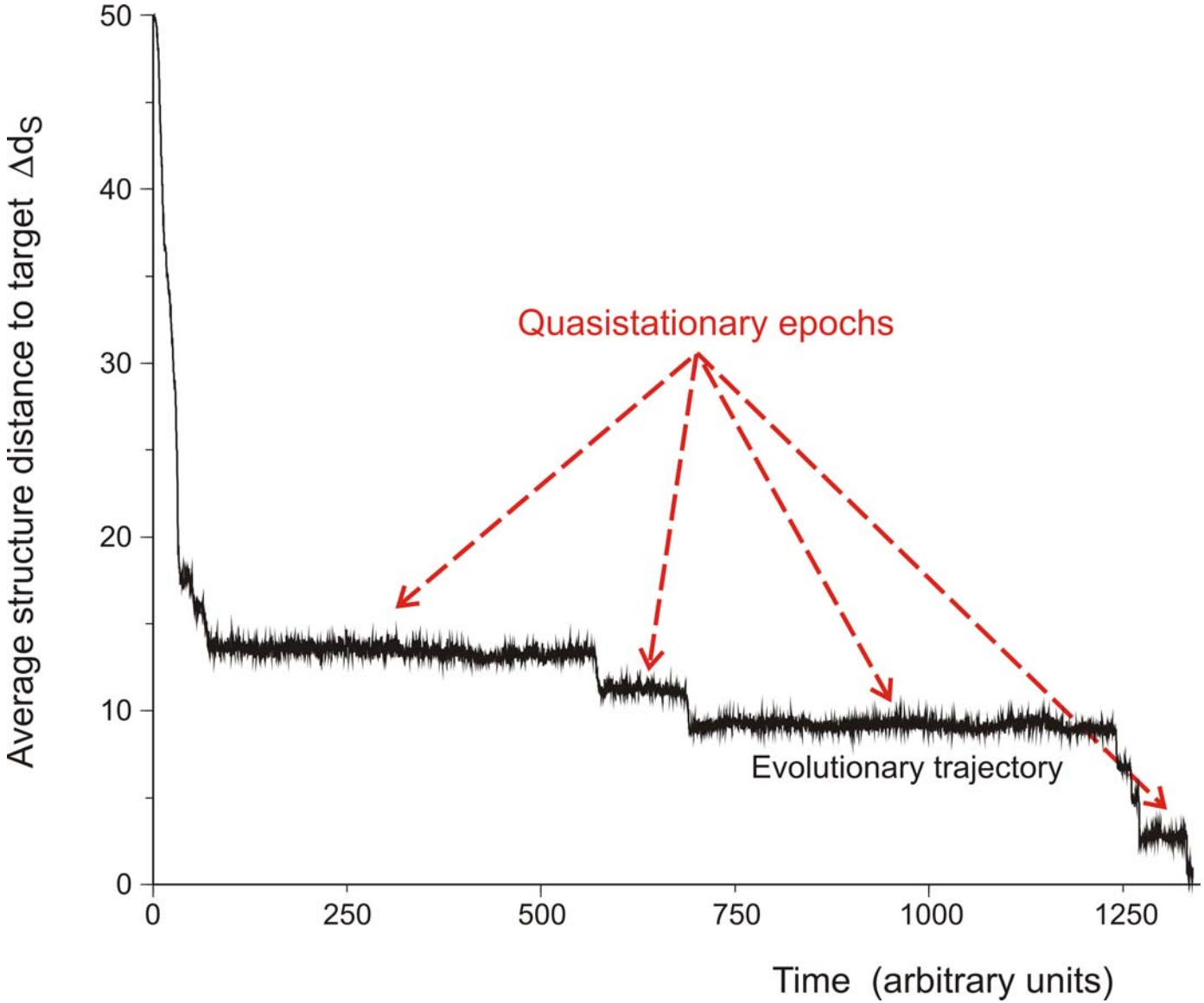
The flowreactor as a device for studies of evolution ***in vitro*** and ***in silico***

Randomly chosen
initial structure



Phenylalanyl-tRNA
as target structure





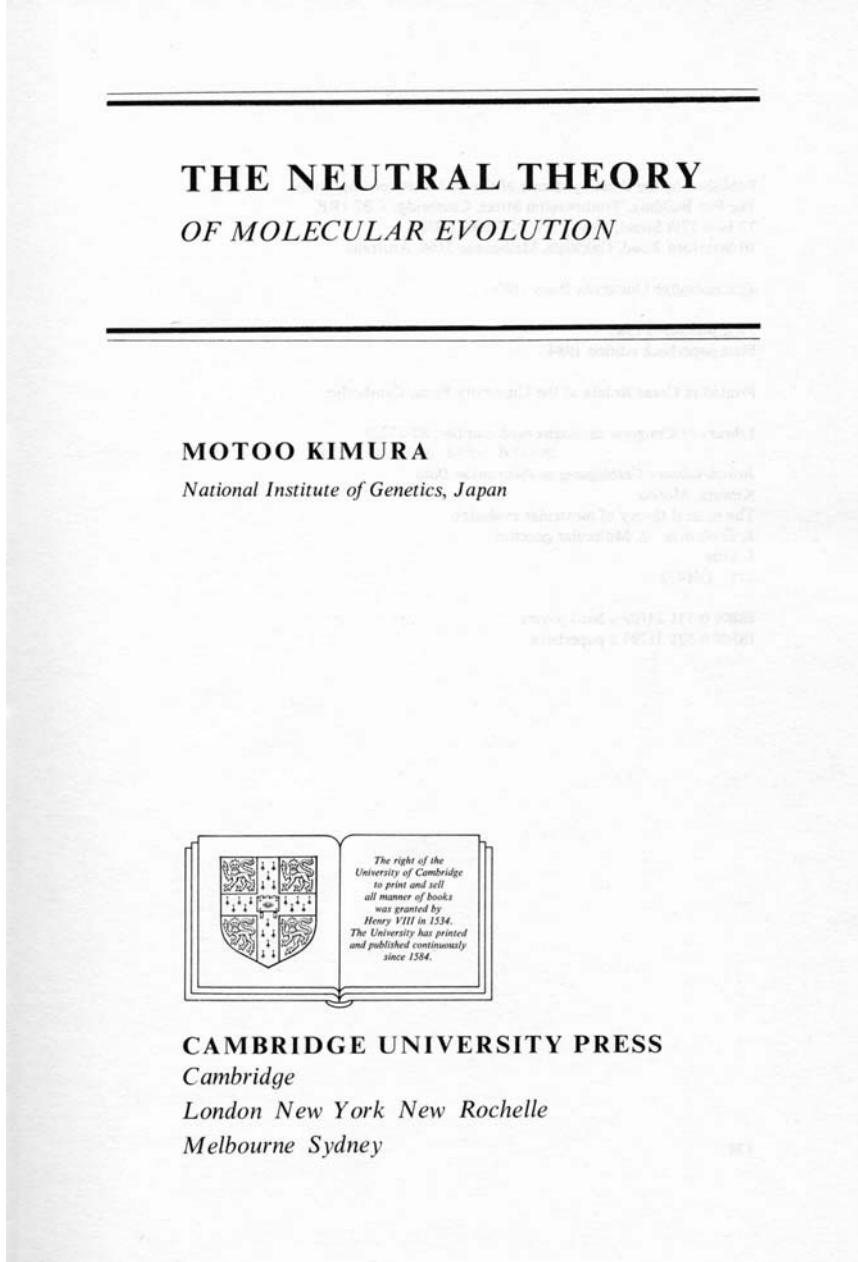
In silico optimization in the flow reactor: Evolutionary Trajectory

MOTOO KIMURA

Motoo Kimuras Populationsgenetik der neutralen Evolution.

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

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



THE NEUTRAL THEORY

OF MOLECULAR EVOLUTION

MOTOO KIMURA

National Institute of Genetics, Japan

CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

Evolutionary design of RNA molecules

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

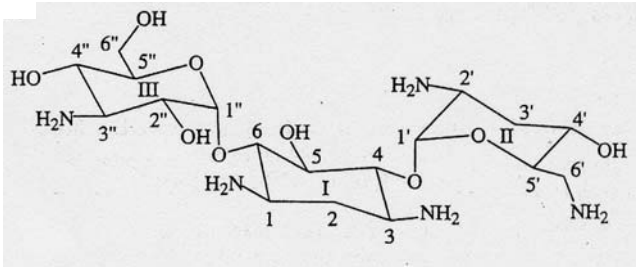
C.Tuerk, L.Gold, **SELEX - *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage* T4 *DNA polymerase***. Science **249** (1990), 505-510

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

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

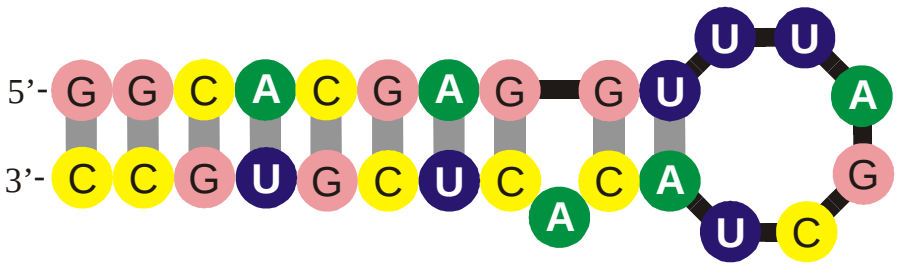
Y. Wang, R.R.Rando, ***Specific binding of aminoglycoside antibiotics to RNA***. Chemistry & Biology **2** (1995), 281-290

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



tobramycin

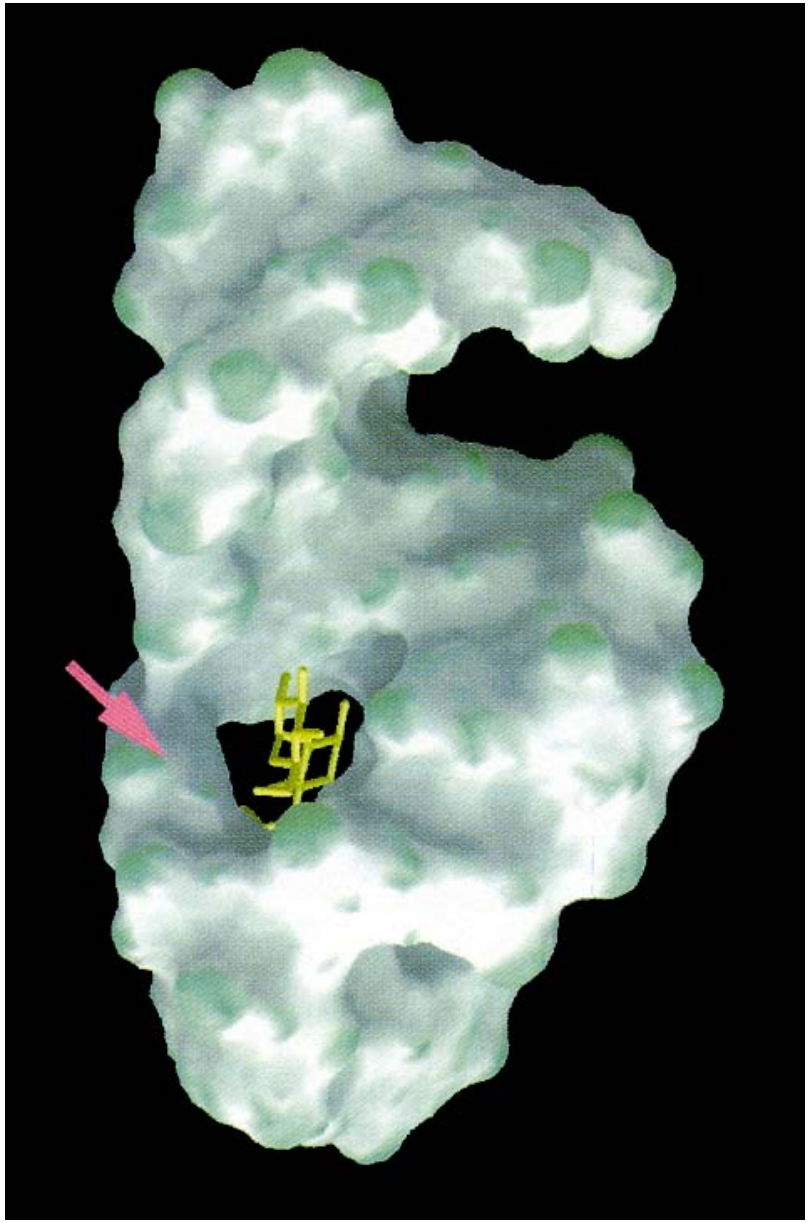
5’- G G C A C G A G G U U U A G C U A C A C U C G U G C C -3’



RNA aptamer

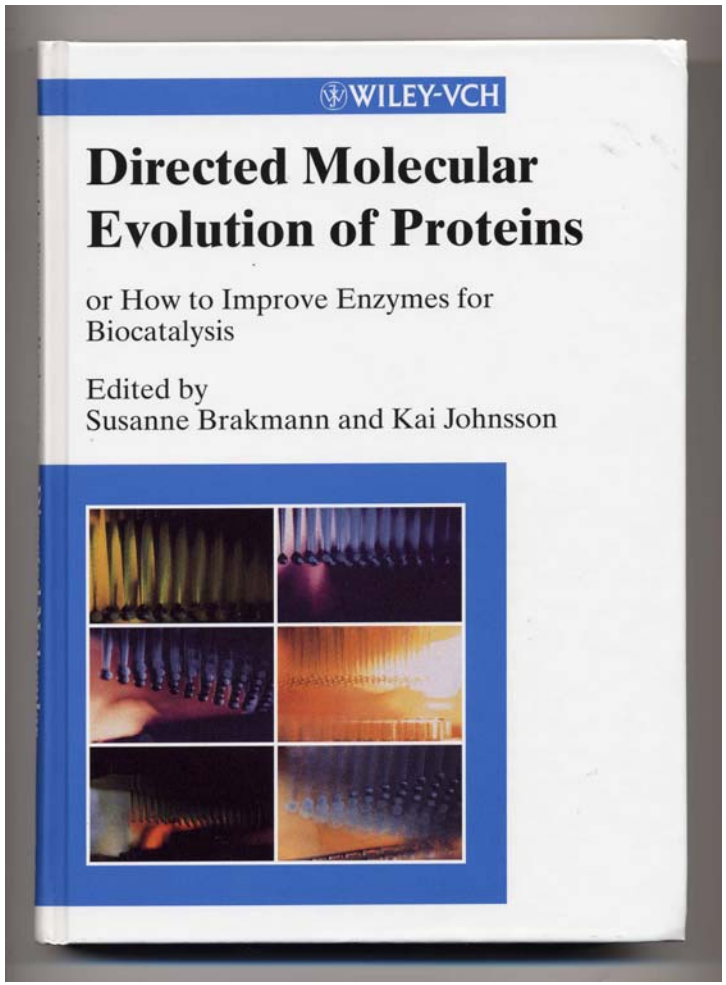
Formation of secondary structure of the tobramycin binding RNA aptamer with **K_D = 9 nM**

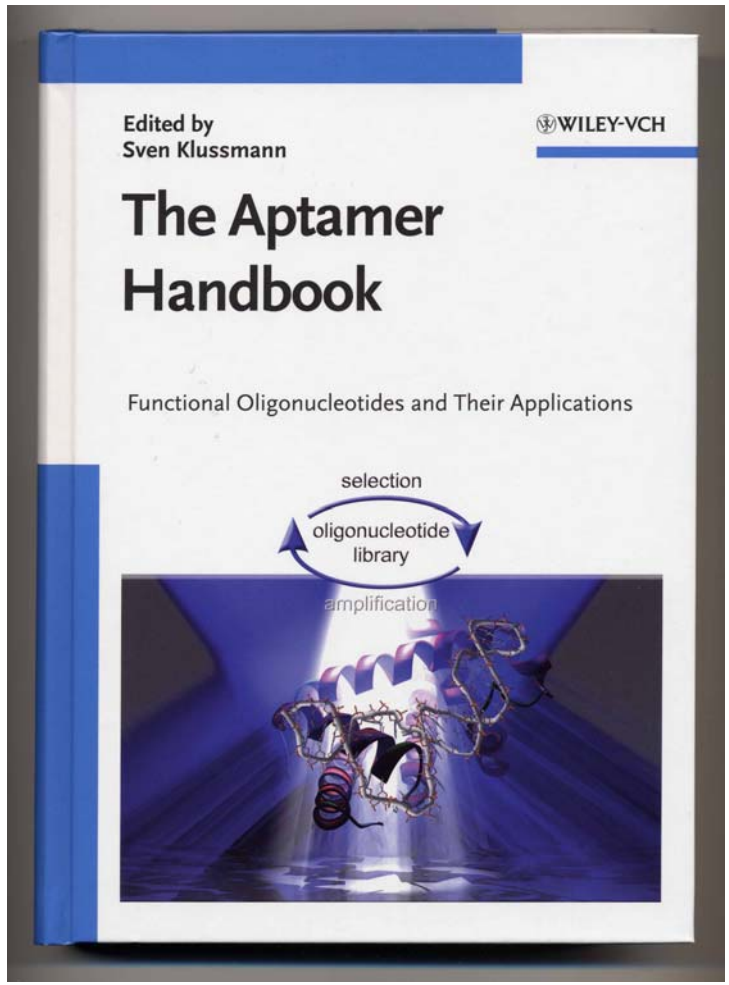
L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.*** Chemistry & Biology **4**:35-50 (1997)



The three-dimensional structure of the tobramycin aptamer complex

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





Application of molecular evolution to problems in biotechnology

1. Evolution und der „Baum des Lebens“
2. Wahrscheinlichkeiten und Zufall
3. Vermehrung, Mutation und Selektion
4. Evolution von Molekülen und Optimierung
5. **Evolutionäres “Basteln” und Komplexität**
6. Schlußbemerkungen

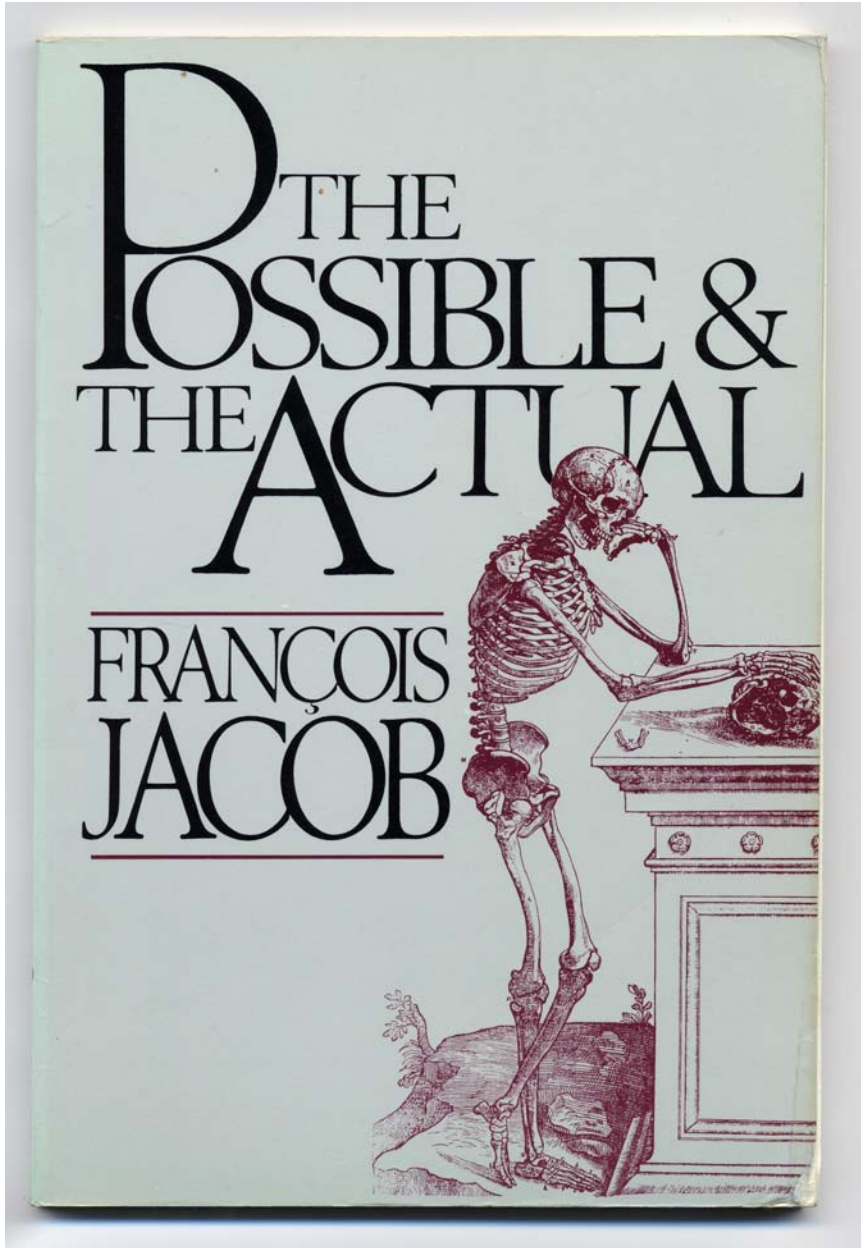
EVOLUTIONARY TINKERING

Blood . . . is the best possible thing to have coursing through one's veins.

—*Woody Allen*, Getting Even

Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

Francois Jacob, Pantheon Books, New York 1982



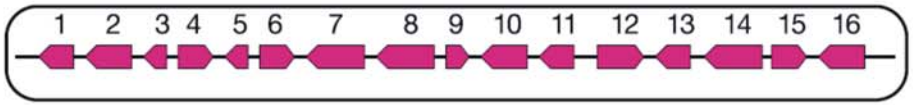
The evolution of 'bricolage'

DENIS DUBOULE (denis.duboule@zoo.unige.ch)

ADAM S. WILKINS (edoffice@bioessays.demon.co.uk)

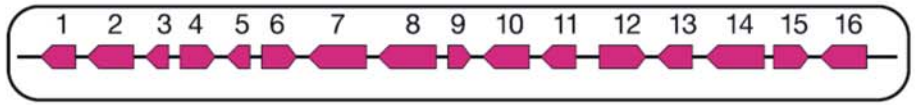
The past ten years of developmental genetics have revealed that most of our genes are shared by other species throughout the animal kingdom. Consequently, animal diversity might largely rely on the differential use of the same components, either at the individual level through divergent functional recruitment, or at a more integrated level, through their participation in various genetic networks. Here, we argue that this inevitably leads to an increase in the interdependency between functions that, in turn, influences the degree to which novel variations can be tolerated. In this 'transitionist' scheme, evolution is neither inherently gradualist nor punctuated but, instead, progresses from one extreme to the other, together with the increased complexity of organisms.

D. Duboule, A.S. Wilkins. 1998. The evolution of ‚bricolage'. Trends in Genetics **14**:54-59.



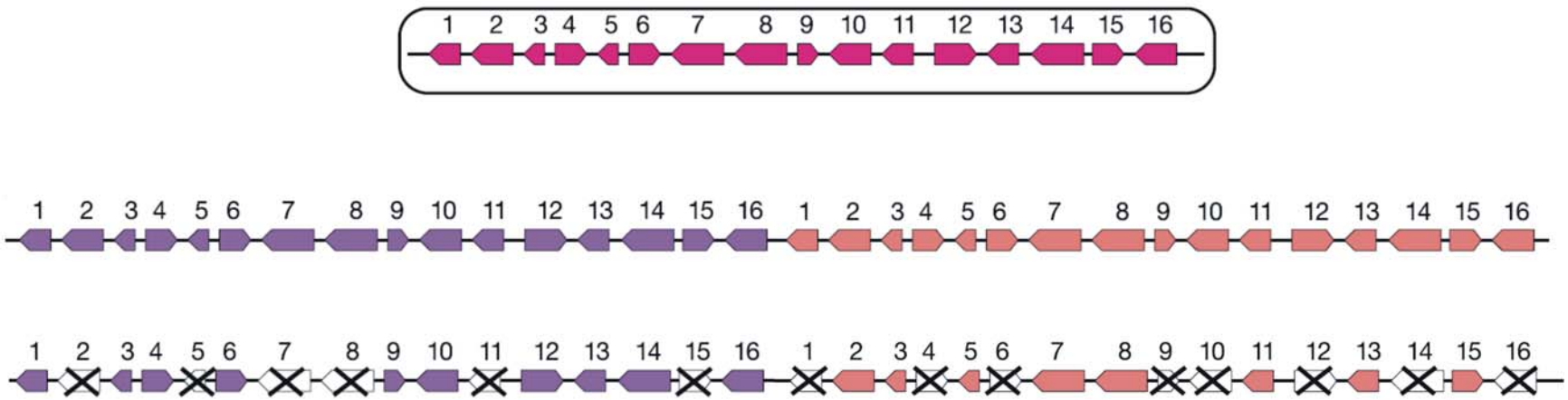
A model for the genome duplication in yeast 100 million years ago

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



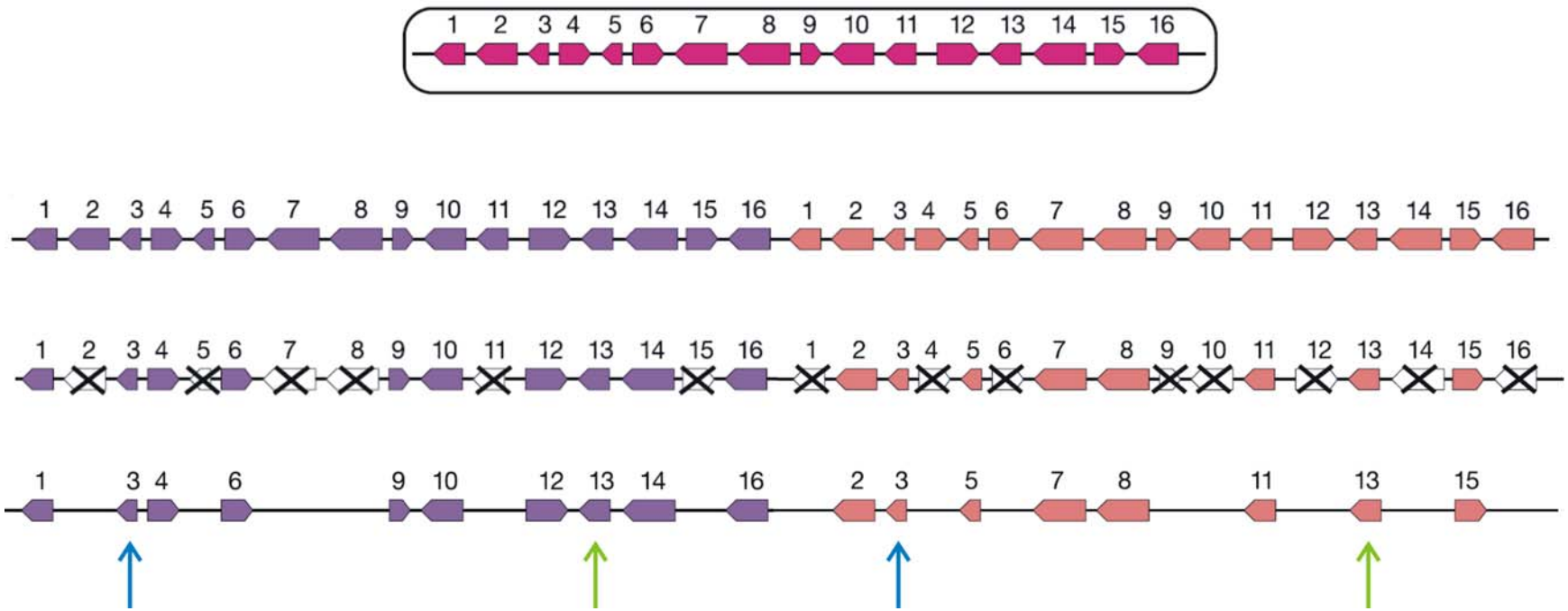
A model for the genome duplication in yeast 100 million years ago

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



A model for the genome duplication in yeast 100 million years ago

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



A model for the genome duplication in yeast 100 million years ago

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

Walter Gehring, Biozentrum, Universität Basel

Die Molekulargenetik zeigt, dass die Entwicklung aller verschieden geformten Augen denselben evolutionären Ursprung hat, welcher bis zu einer einfachen lichtempfindlichen Vorstufe eines Organs zurückverfolgt werden kann, das bereits in primitiven Bakterienstämmen gefunden wird.

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

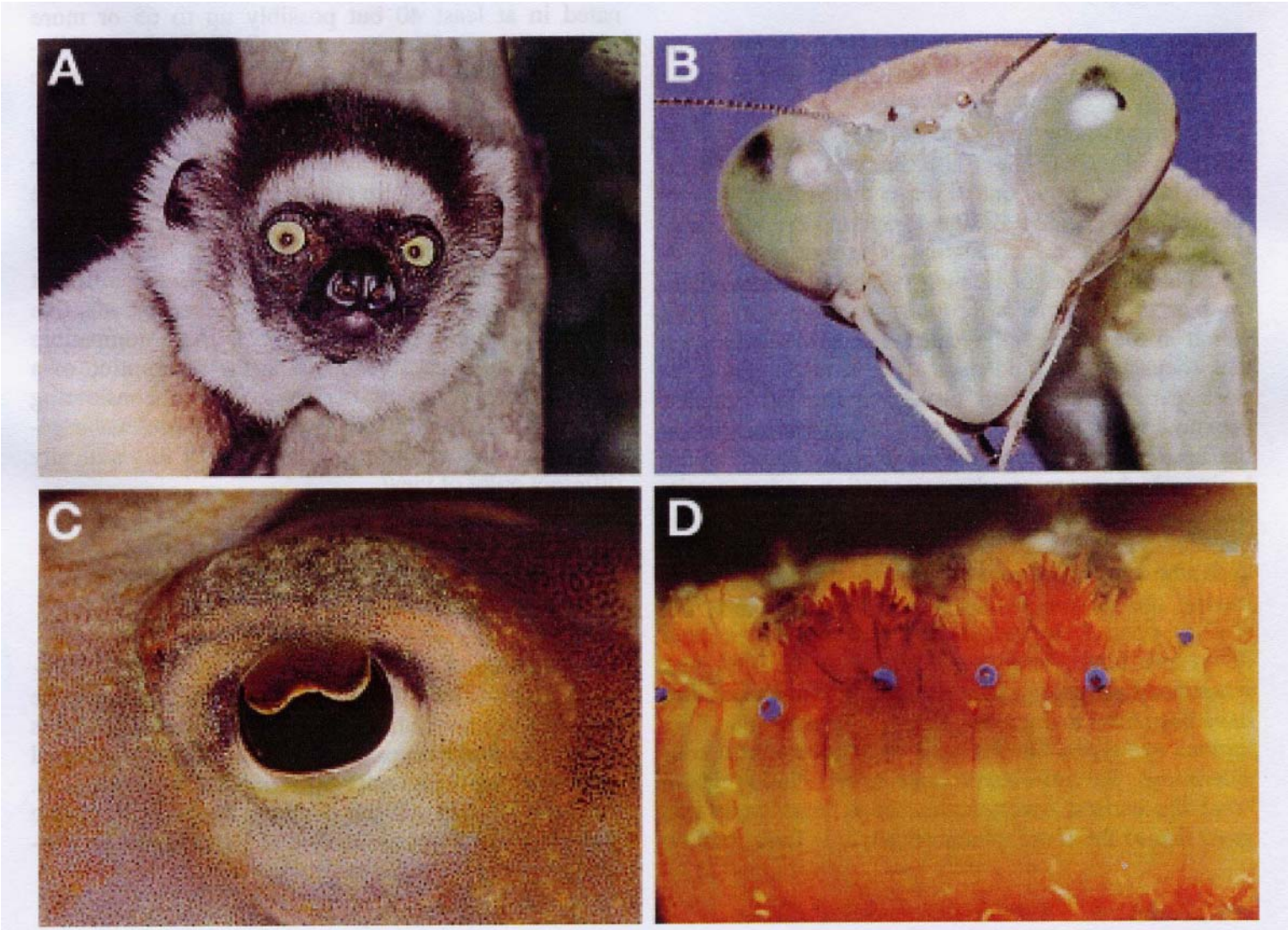


Fig. 1. Different types of eyes. **(A)** Camera-type eye from the Lemur *Propithecus verrauxi.* **(B)** Compound eye of the praying *Mantis.* **(C)** Camera-type eye from the Cephalopod *Sepia erostrata.* **(D)** Mirror eye from the clam *Chlamys nobilis.* (Courtesy of Dr. Kazuto Kato; photographs kindly provided by Masahiro Iijima, Susumu Yamaguchi and Isamu Soyama).

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

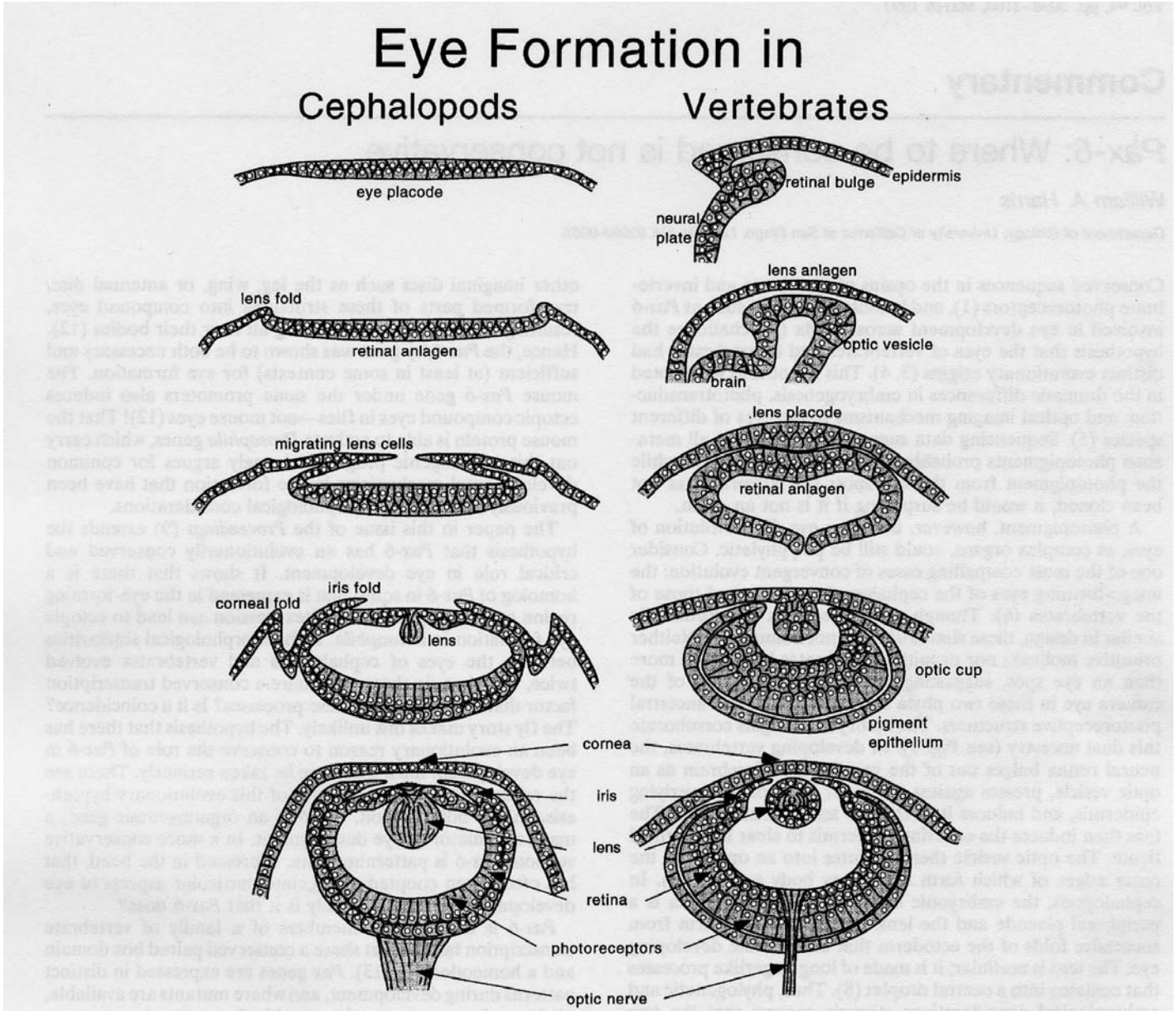


FIG. 1. Schematic diagram of cephalopod eye development (*Left*) and vertebrate eye development (*Right*) as explained in more detail in refs. 7 and 8. Development proceeds from top to bottom. Even though the adult structures are fairly similar, excepting certain obvious features such as the placement of the photoreceptors and lentigenic cells, the development is very different. The cephalopod eye forms from an epidermal placode through a series of successive infoldings, while the vertebrate eye emerges from the neural plate and induces the overlying epidermis to form the lens.

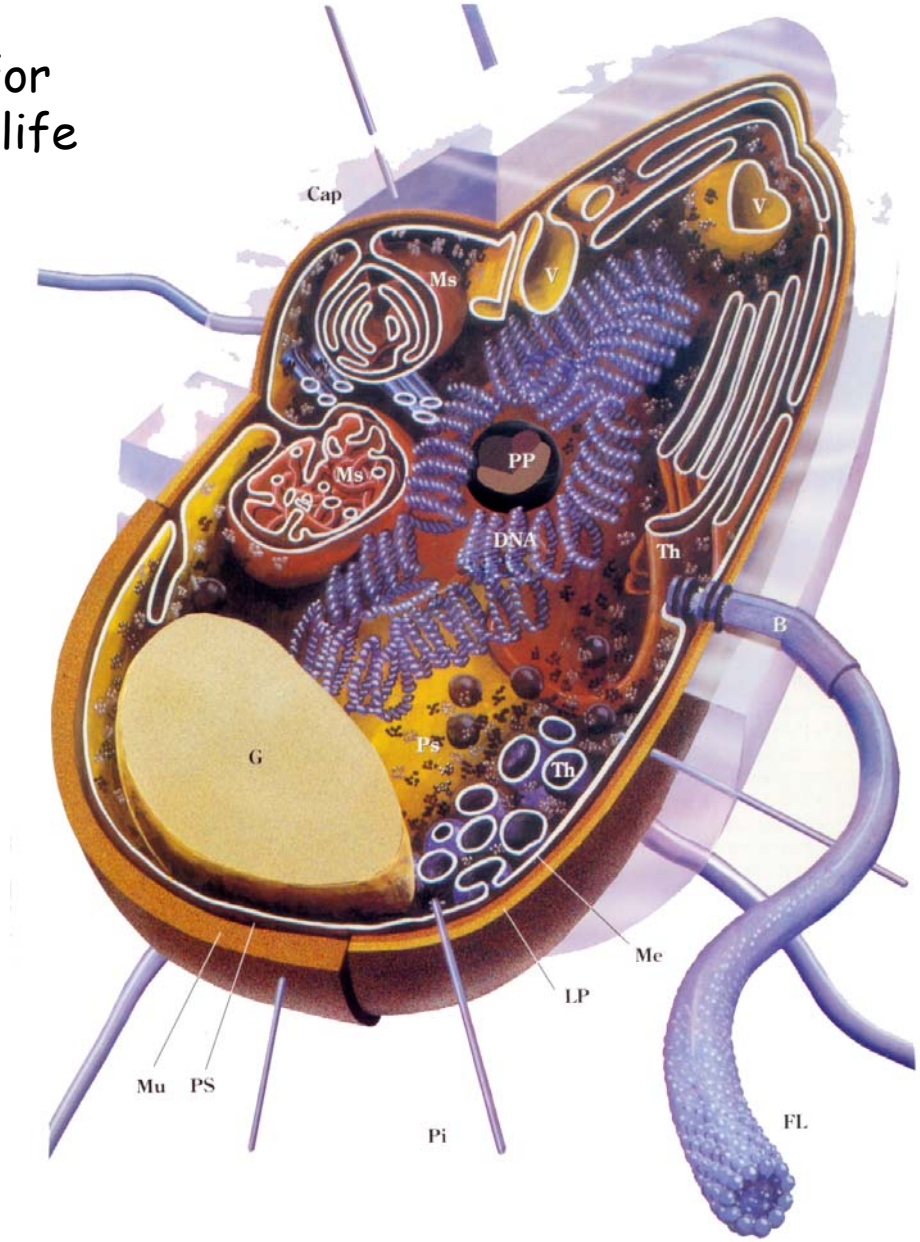
William A. Harris,
Proc.Natl.Acad.Sci.USA
94:2098-2100, 1997

The bacterial cell as an example for the simplest form of autonomous life

Escherichia coli genome:

4 million nucleotides

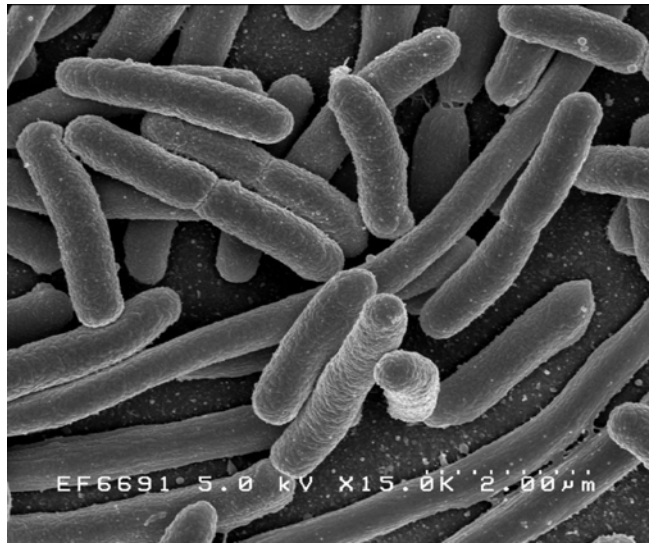
4460 genes



The structure of the bacterium *Escherichia coli*

E. coli:	Genome length	4×10^6 nucleotides
	Number of cell types	1
	Number of genes	4 460

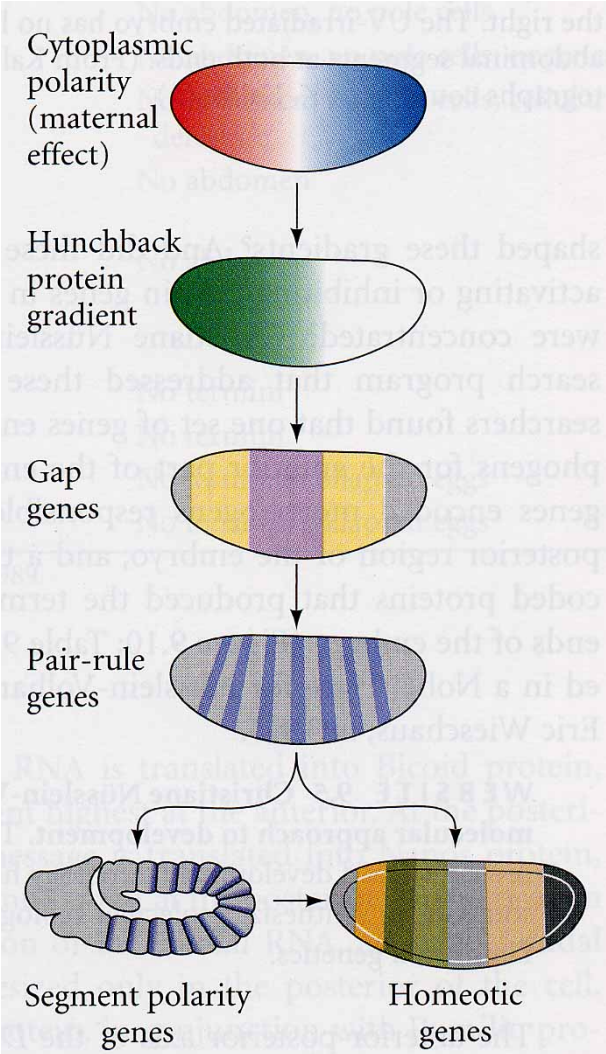
Four books, 300 pages each

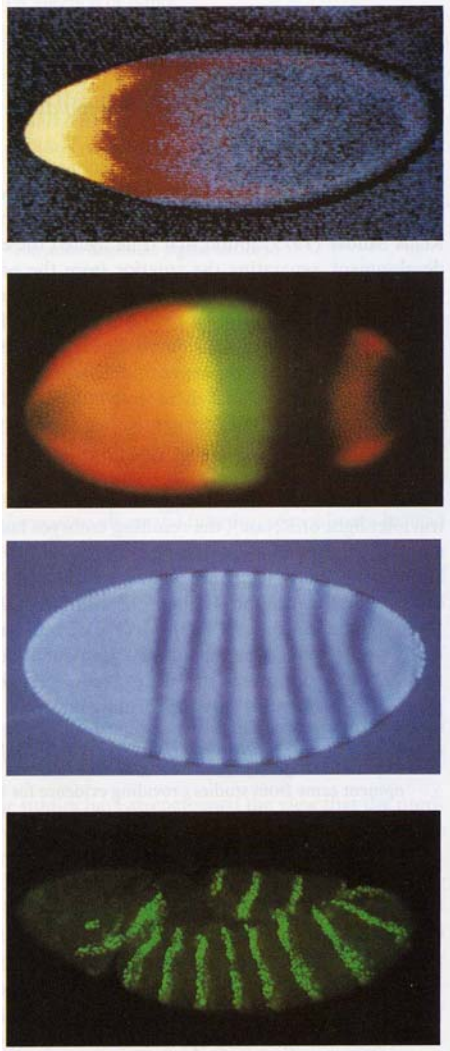


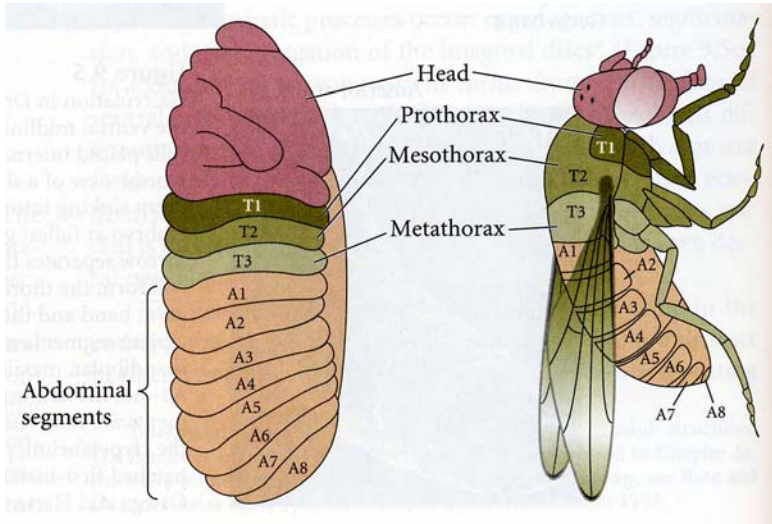
Man:	Genome length	3×10^9 nucleotides
	Number of cell types	200
	Number of genes	≈ 30 000

A library of 3000 volumes,
300 pages each

Complexity in biology







Cascades, A ⇒ B ⇒ C ⇒ ... , and networks of genetic control

Turing pattern resulting from reaction-diffusion equation ?

Intercelluar communication creating positional information

Development of the fruit fly *drosophila melanogaster*: Genetics, experiment, and imago

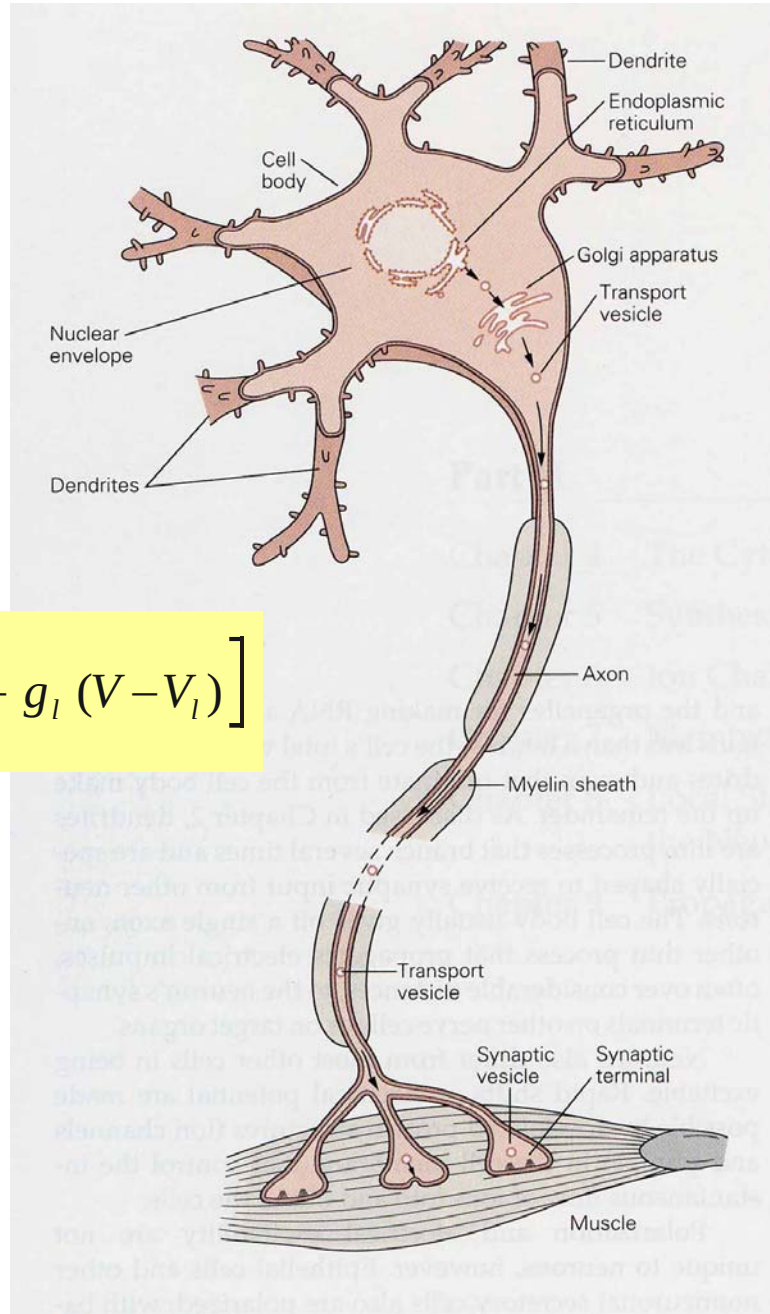
$$\frac{dV}{dt} = \frac{1}{C_M}\left[I - g_{Na}\, m^3\, h\,(V - V_{Na}) - g_K\, n^4\,(V - V_K) - g_l\,(V - V_l)\right]$$

$$\frac{dm}{dt} = \alpha_m\,(1-m) - \beta_m\, m$$

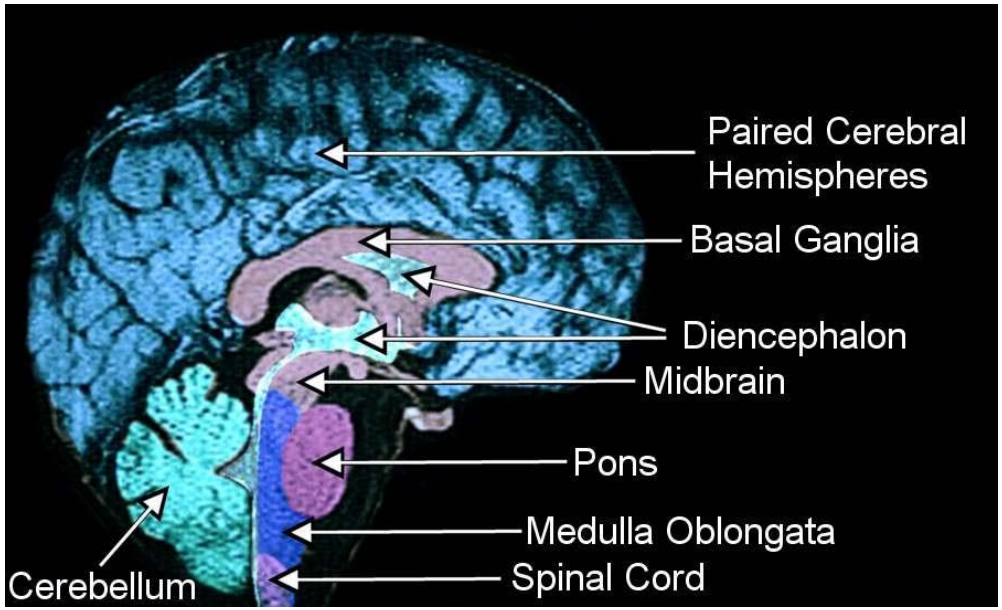
$$\frac{dh}{dt} = \alpha_h\,(1-h) - \beta_h\, h$$

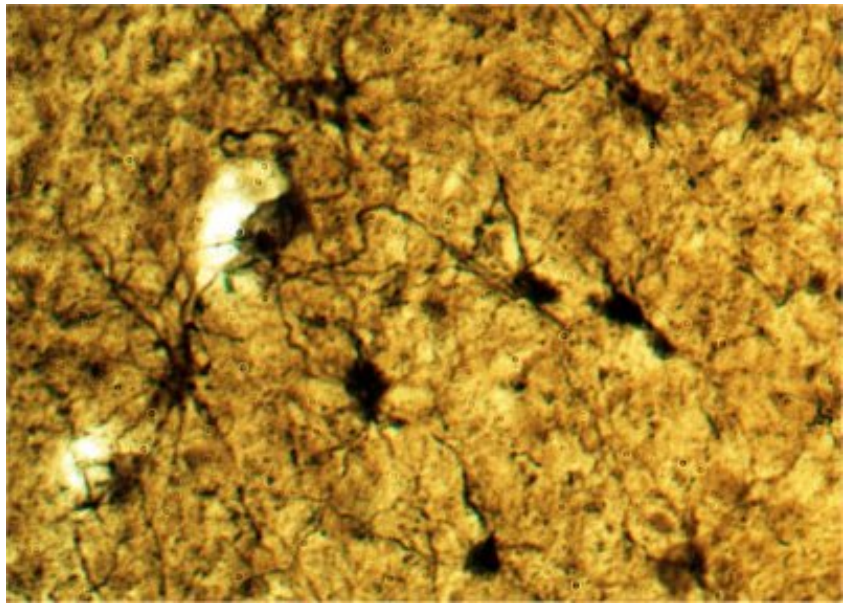
$$\frac{dn}{dt} = \alpha_n\,(1-n) - \beta_n\, n$$

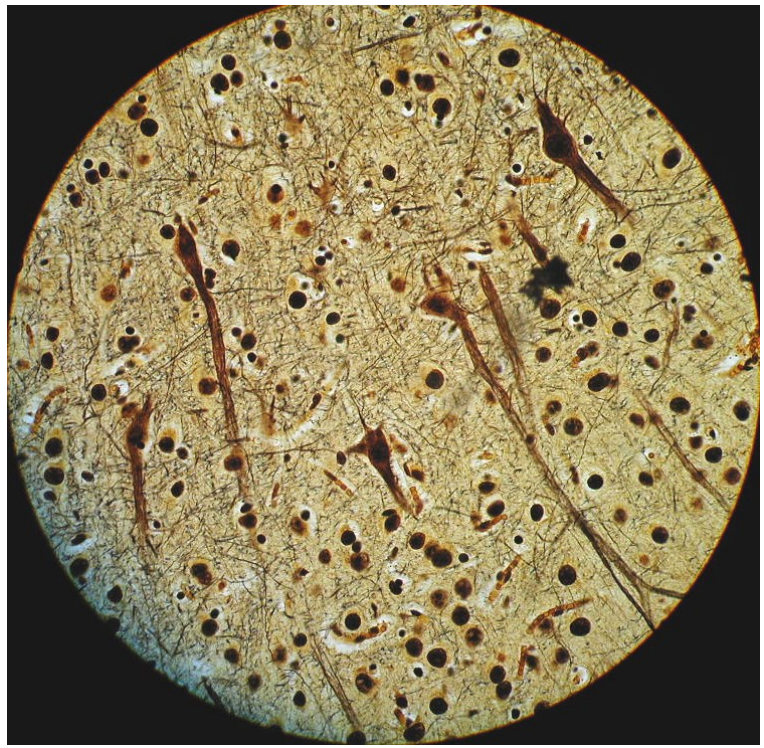
Hogdkin-Huxley OD equations



A single neuron signaling to a muscle fiber

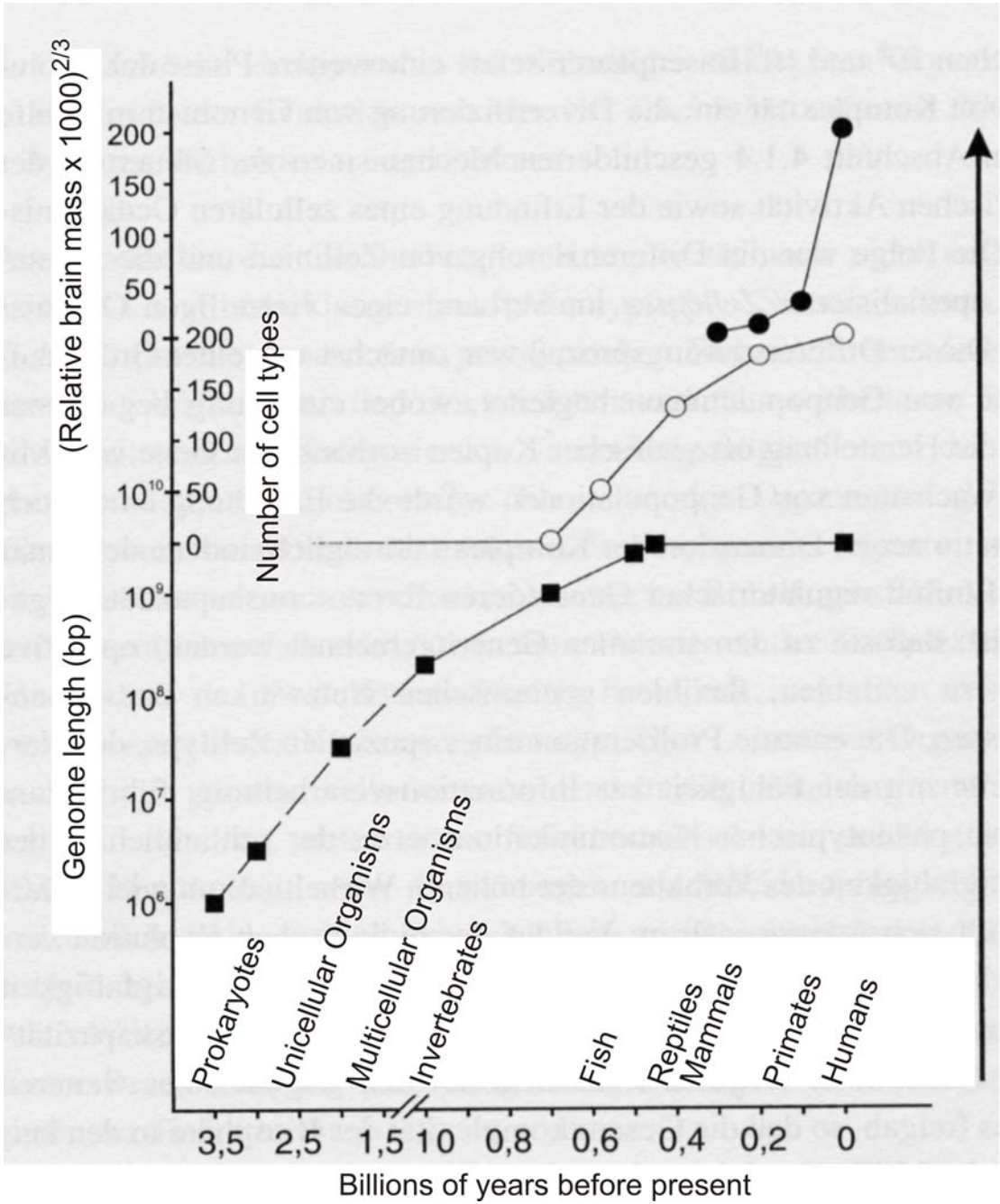






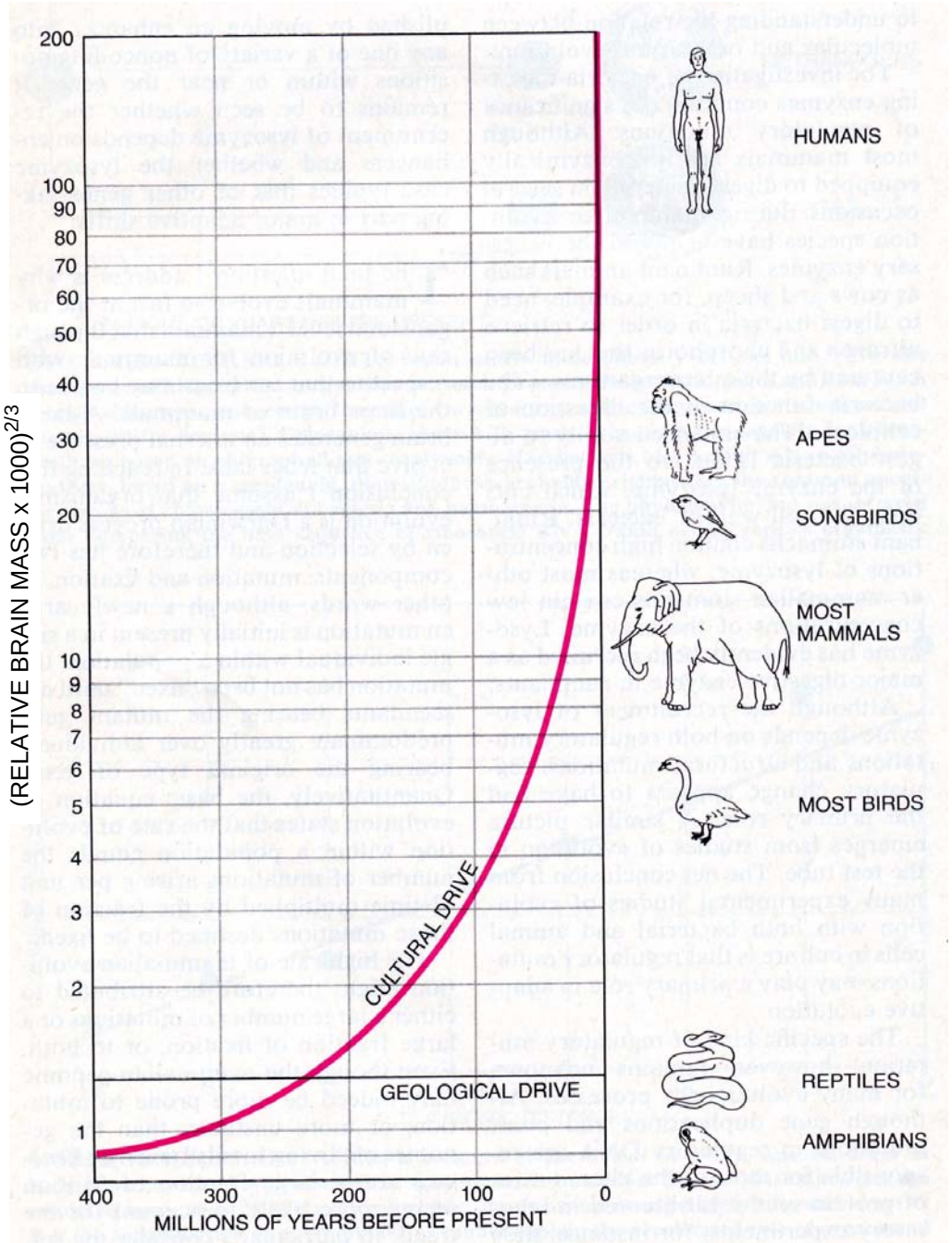
The human brain

10^{11} neurons connected by ≈ 10^{13} to 10^{14} synapses



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

BRITISH TIT



Alan C. Wilson.1985. The molecular basis of evolution.
Scientific American **253**(4):148-157.

The difficulty to define the notion of „gene".

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.

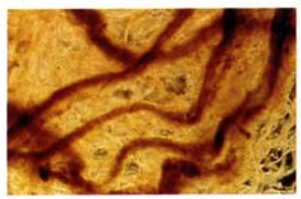
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia 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

ENCODE stands for
ENCyclopedia Of DNA Elements.

ENCODE Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
Nature **447**:799-816, 2007

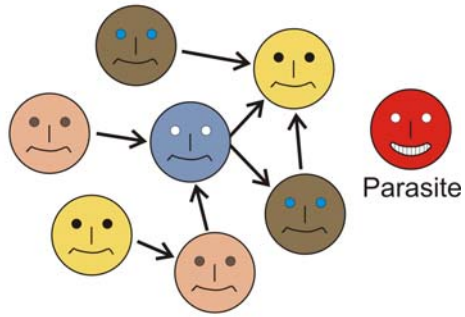
1. Evolution und der „Baum des Lebens“
2. Wahrscheinlichkeiten und Zufall
3. Vermehrung, Mutation und Selektion
4. Evolution von Molekülen und Optimierung
5. Evolutionäres “Basteln” und Komplexität
6. **Schlußbemerkungen**

Die großen Evolutionsschritte (nach John Maynard Smith und Eörs Szathmáry)

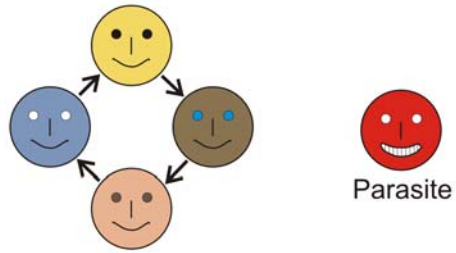
		Membranen, organisierte Teilung	
Replizierende Moleküle		⇒	Moleküle in Kompartments
		Molekülverkettung, gemeinsame Replikation	
Unabhängige Replikatoren		⇒	Chromosomen
		genetischer Code, Ribosom	
RNA als Gen und Enzyme		⇒	DNA und Protein
		Zusammenschluß durch Endosymbiose	
Prokaryoten		⇒	Eukaryoten
		Ursprung der sexuellen Vermehrung	
Asexuell vermehrende Klone		⇒	Sexuell vermehrende Populationen
		Zelldifferenzierung und Entwicklung	
Protisten		⇒	Pflanzen, Pilze und Tiere
		Entstehung nicht-reproduktiver Kasten	
Einzeln lebende Individuen		⇒	Tierkolonien
		Sprache, Schrift, Kultur, …	
Primatengesellschaften		⇒	menschliche Gesellschaften

Stufe I:
Unabhängige Replikatoren
in Konkurrenz

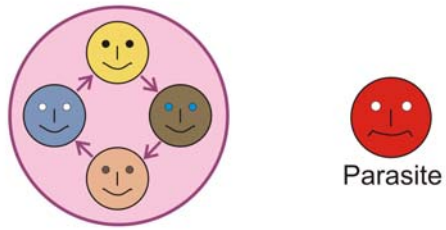
Stufe II:
Katalyse und Konkurrenz
bei der Replikation



Stufe III:
Funktionell verknüpfte
Replikatoren



Stufe IV:
Neue Einheit der
Selektion



Stufe V:
Unabhängige Einheiten
in Konkurrenz

Ein Mechanismus zur Überwindung
hierarchischer Stufen in der Evolution
(nach Manfred Eigen und Peter Schuster)

Darwin hatte in folgenden Punkten **nicht recht**:

• Der Darwinsche Vererbungsmechanismus war falsch. Mendel hatte die korrekte Lösung.

• Mutation und Rekombination können keine, kleine und große Auswirkungen haben und es besteht kein Grund, dass die biologische Evolution quasikontinuierlich oder anders ausgedrückt nur in verschwindend kleinen Schritten erfolgt.

• Im Verlaufe der biologischen Evolution gab es auch katastrophenartige Ereignisse terrestrischen und extraterrestrischen Ursprungs.

• Die Komplexität der höheren Lebewesen ist so groß, dass ihre Eigenschaften nicht voll optimiert sein können.

Darwins Theorie wurde in folgenden Punkten **voll bestätigt**:

•Das Auftreten von Varianten bei der Reproduktion wurde durch die Aufklärung der molekularen Mechanismen von Rekombination und Mutation auf eine solide wissenschaftliche Basis gestellt.

• Das Darwinsche Prinzip der Optimierung durch Variation und Selektion in endlichen Populationen gilt nicht nur in der Biologie sondern auch in der unbelebten Welt.

• Die natürliche Entstehung der Arten und die daraus resultierenden phylogenetischen Stammbäume wurde durch die Vergleiche der genetischen Informationsträger heute lebender Organismen voll bestätigt.

• Das Referat beschränkte sich auf die heutigen naturwissenschaftlichen Erkenntnisse.

• Die Vorstellung der biologischen Evolution ist eine empirisch begründete, naturwissenschaftliche Theorie.

• Die Evolutionstheorie ist in einigen wesentlichen Aussagen experimentell prüfbar und überprüft und baut auf Tatsachen aus mehreren Teildisziplinen auf.

• Die Evolutionstheorie ist daher vom selben Rang wie physikalische Theorien, etwa die Newtonsche Mechanik, die Relativitätstheorie oder die Quantentheorie.

• Wie die meisten naturwissenschaftlichen Theorien kann die biologische Evolutionstheorie nicht alle beobachteten Einzelheiten erklären insbesondere, da die Biologie zur Zeit in einer faszinierenden und raschen Entwicklung steht.

• Die Molekularbiologie führt die biologischen Befunde auf Gesetzmäßigkeiten aus Physik und Chemie zurück, ohne dadurch die Eigenständigkeit der Biologie in Frage zu stellen.

Web-Page for further information:

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