





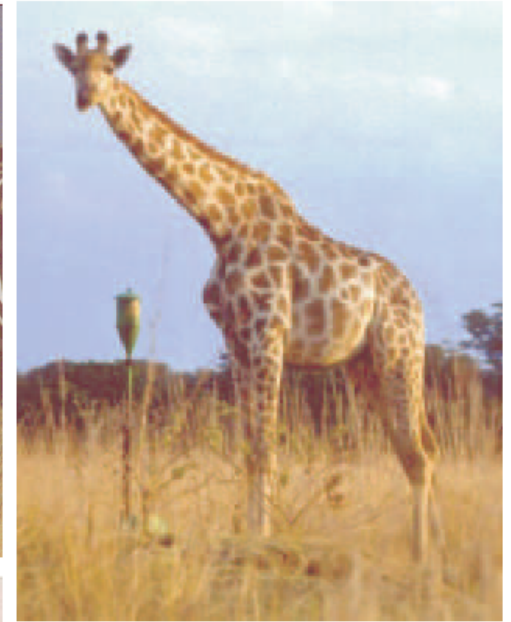
Web-Page for further information:

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

In der belebten Natur treffen wir überall zu Bewunderung und Erstaunen Anlass gebende Muster und Strukturen an. Seien es die vielfältigen Formen und Farben von Blüten, die als Tarnung oder als Signale wirkenden Felle, Federkleider, Panzer oder Häute von Tieren, überall finden wir auffällige Muster und zum Teil ornamentartige Verzierungen.

Bemerkenswerte Formen des Nachmachens der Erscheinungsformen anderer Arten – Mimikry genannt – wurden entdeckt, welche dem Schutz vor Fressfeinden dienen, die –getäuscht durch die äußere Erscheinung – vom Verzehr der vermeintlich als giftig angesehenen Beute Abstand nehmen.

Die aufgefundene Vielfalt und die augenscheinliche Leichtigkeit, mit der eine Erscheinungsform in eine andere umgewandelt werden kann, legt die Vermutung nahe, dass der Musterbildung ein weit verbreiteter und verhältnismäßig einfach zu steuernder Mechanismus zugrunde liegen muss. Es erhebt sich die Frage, ob Musterbildung auch außerhalb der Biologie beobachtet werden kann und, welchen Gesetzmäßigkeiten sie gehorcht.



Farbmuster auf Tierfellen, -flügeln und -panzern

Musterbildung in der nicht organischen Natur ist uns allen bekannt. Am spektakulärsten sind die verschiedenen Wolkenformationen, die insbesondere vor hereinbrechenden Schlechtwetterfronten auffallend regelmäßige Strukturierungen zeigen (Bild links oben).

In Gewittern kommt es zu bedeutenden Ladungstrennungen und elektrischen Potentialunterschieden oder Spannungen zwischen den Wolken oder zwischen Wolken und der Erdoberfläche (Bild rechts oben). Blitzentladungen erfolgen, wenn bestimmte Grenzspannungen überschritten werden. Der genaue Verlauf der Entladungen ist nicht vorhersagbar, da die lokalen elektrischen Leitfähigkeiten kleinster Luftvolumina eine Rolle spielen können. Für die Gefahr eines Blitzschlages auf der Erdoberfläche lassen sich auch nur Wahrscheinlichkeiten angeben. Diese Wahrscheinlichkeiten sind beispielweise größer bei höheren Lagen, bei Spitzen und bei elektrischen Leitern – auf dieser zuerst von Benjamin Franklin empirisch beobachteten Tatsache beruht seine Erfindung des Blitzableiters im Jahre 1752.

Bei genügend hohen Temperaturunterschieden zwischen Boden und Wolkenobergrenze und genügend feuchter Luft kann es zur Ausbildung von Tornados über Land (Windhosen; Bild links unten) oder über Wasser (Wasserhosen) kommen. Obwohl es eine recht gut entwickelte Theorie für die Entstehung von Tornados gibt, ist es unmöglich vorherzusagen, wann und wo ein Tornado auftreten wird.

Ein Hurrikan (Bild rechts unten) ist ein Wirbelsturm von gigantischen Ausmaßen, der sich über warmen Meeren ausbildet. Auch in diesem Fall gilt, dass ein genügend hoher Temperaturunterschied zwischen der Meeresoberfläche und der Wolkenobergrenze und eine hinreichend starke Rotationskraft (Corioliskraft) notwendig ist, um einen Wirbel entstehen zu lassen: Im allgemeinen werden mindestens  $26,5\text{ }^{\circ}\text{C}$  Oberflächentemperatur des Meeres und mehr als  $|\pm 5|^{\circ}$  geographischer Breite als Bedingungen genannt. Ebenso wie bei Blitz und Tornado sind dies notwendige aber keine hinreichenden Bedingungen – Die tatsächliche Ausbildung eines Hurrikans hängt auch von den unmöglich ermittelbaren, lokalen Gegebenheiten im Entstehungsgebiet ab.



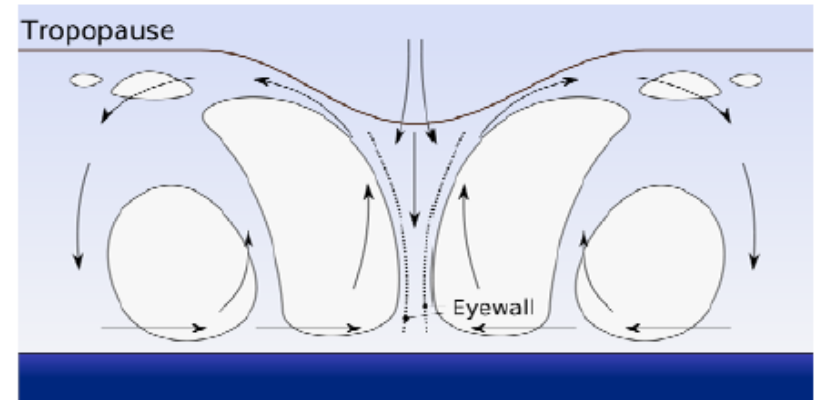
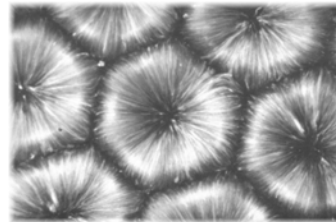
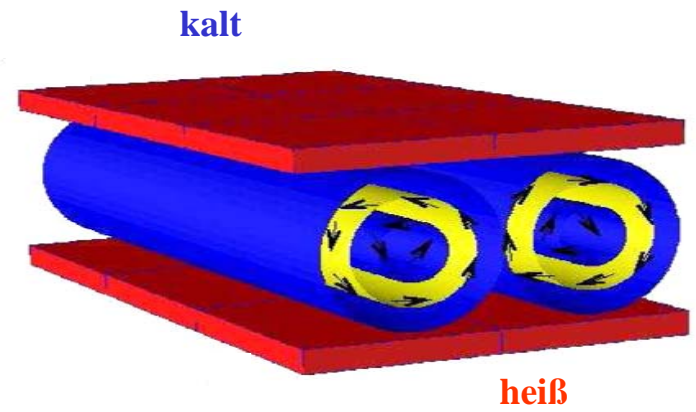
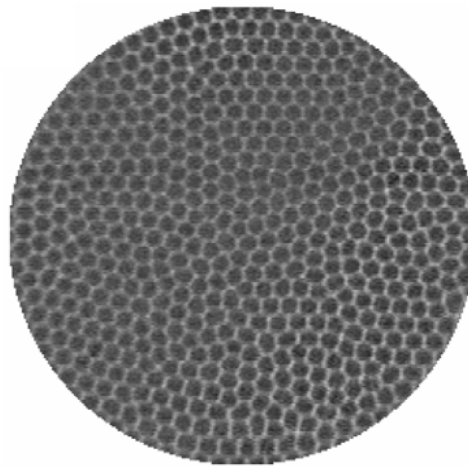
Wetterphänomene – Gewitterfront, Blitz, Tornado and Hurrikan

Die Musterbildung als Folge eines Temperaturunterschiedes – unter heiß und oben kalt – kann auch in viel kleinerem Maßstab beobachtet werden. Raleigh-Benard Konvektion entsteht, wenn man eine ruhende Flüssigkeit von unten erhitzt. Die heiße Flüssigkeit am Boden ist leichter als die kalte Flüssigkeit an der oberen Flüssigkeitsoberfläche. Bei geringen Temperaturunterschieden reicht die thermische Bewegung zur Kompensation des Temperaturunterschieds aus. Wird dieser durch stärkeres Erhitzen größer, dann kommt es zur thermischen Konvektion: Volumenelemente der Flüssigkeit bewegen sich, um einen rascheren Temperatur- und Dichteausgleich zu bewirken, als dies mit Wärmebewegung allein möglich wäre. Heiße Flüssigkeit steigt vom Boden nach oben und kalte Flüssigkeit sinkt von der oberen Oberfläche nach unten. Die Konvektion führt zur Ausbildung von zylindrischen Rollen (Bild oben rechts), an deren Berührungspunkten die Flüssigkeit abwechselnd auf- und absteigt.

Experimentell lässt sich die Musterbildung einfach beobachten (Bilder oben links). Die Rollen bilden an der oberen Oberfläche sichtbare Muster aus und tendieren zu Spiralenbildung, die sich mit beiden Drehsinnen einrollen. Bei weiterer Erhitzung entstehen aus den Rollen Wabenmuster, die bei geringen Störungen die Form regelmäßigen Sechsecken annehmen.

Im unteren Teil des Bildes ist eine gut entwickelte Theorie der Hurrikanausbildung skizziert (Bild unten rechts). Die Ausbildung von Zonen mit auf- und absteigenden Luftmassen entspricht ganz der Raleigh-Benard Konvektion. Durch die Wirkung der Corioliskraft rotiert ein Hurrikan auf der Nordhalbkugel gegen den Uhrzeigersinn und auf der Südhalbkugel im Uhrzeigersinn. Obwohl die Theorie der Hurrikanbildung zufriedenstellende Erklärungen für das Phänomen als solches geben kann, ist es nicht möglich vorherzusagen, wann und wo ein Hurrikan gebildet wird. Lokale nicht erfassbare Details spielen bei der Nukleation eines Wirbelsturmes eine wesentliche Rolle.



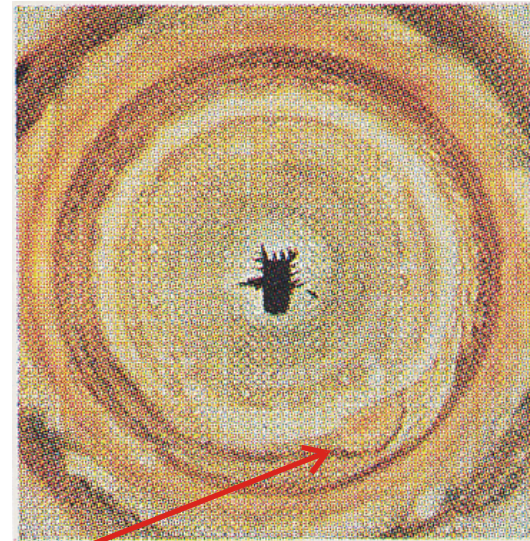


Raleigh-Bénard Konvektion und Ausbildung von Hurrikanen

Wirbelstürme ganz im Sinne der irdischen Hurrikane, Zyklone oder Taifune können gigantische Ausmaße annehmen, wie der „Rote Fleck“ am Jupiter zeigt. Derartig mächtige Wirbel am Jupiter weisen eine für irdische Verhältnisse unwahrscheinlich lange Lebensdauer auf, da dort die Energie der irdischen Hurrikane absorbierende Festland fehlt oder zumindest weit weniger Einfluss hat. Die Entstehung des Jupiterwirbels wurde am Computer simuliert und auch in einem einige Meter großen Tank experimentell nachgewiesen.

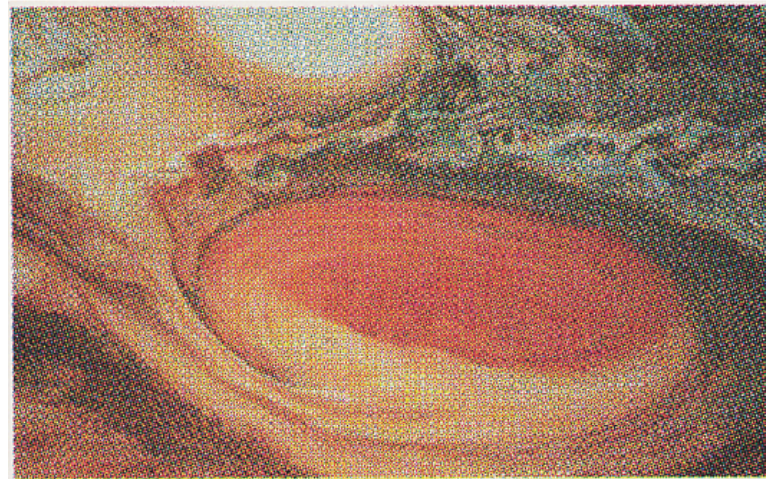


South pole



View from south pole

Red spot



Roter Fleck des Jupiters: Beobachtung eines gigantischen Wirbels

Bilder sind entnommen von James Gleick, *Chaos*. Penguin Books, New York, 1988

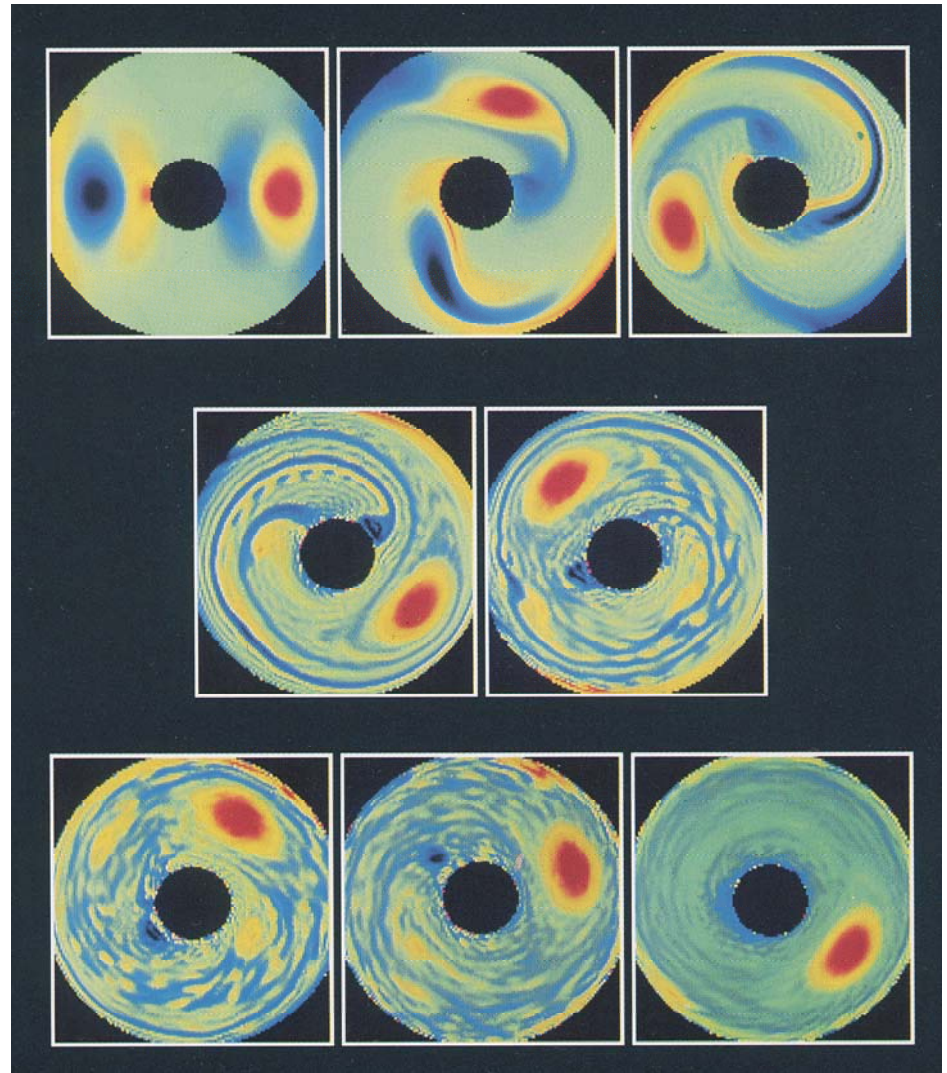
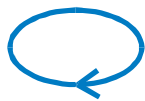
# Computer simulation of the gigantic vortex on Jupiter

View from south pole

Particles turning  
counterclockwise



Particles turning  
clockwise



Jupiter: Computersimulation des gigantischen Wirbels

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

Musterbildung bei chemischen Reaktionen wurde um 1950 von dem bekannten englischen Computerwissenschaftler Alan M. Turing vorhergesagt und als ein mögliches Prinzip der biologischen Musterbildung postuliert. Dem Turingschen Modell liegt eine einfache Modellierung (nicht-linearer) chemischer Reaktionen in Raum und Zeit zugrunde: Das Turing-Modell benutzt Reaktions-Diffusions-Gleichungen um die Produktion chemischer Verbindungen und ihre Ausbreitung im Raum zu beschreiben.

Für Zweiphasen-Systeme war das Auftreten von Turing-Muster artigen Bandstrukturen schon lange bekannt: Bei der Auskristallisation aus übersättigten Lösungen treten sogenannte Liesegang-Ringe auf, die einer den Turingschen Systemen ähnlichen Nichtlinearität ihre Entstehung verdanken.

Zehn Jahre nach Turings Vorhersage wurde die Ausbildung von Raum-Zeit-Mustern in Form von wandernden Spiralen und konzentrischen Mustern („Target-Waves“) bei der Belousov-Zhabotinskii-Reaktion nachgewiesen. Erst vierzig Jahre nach der Turingschen Vorhersage gelang es stationäre (stehende) Turing-Muster bei chemischen Reaktionen zu erzeugen.

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u, v)$$

$$\frac{\partial v}{\partial t} = D_v \nabla^2 v + g(u, v)$$

$$u = u(x, y, z, t) \quad \text{and} \quad v = v(x, y, z, t)$$

Veränderung in der lokalen Konzentration =

= Diffusion + Chemische Reaktion

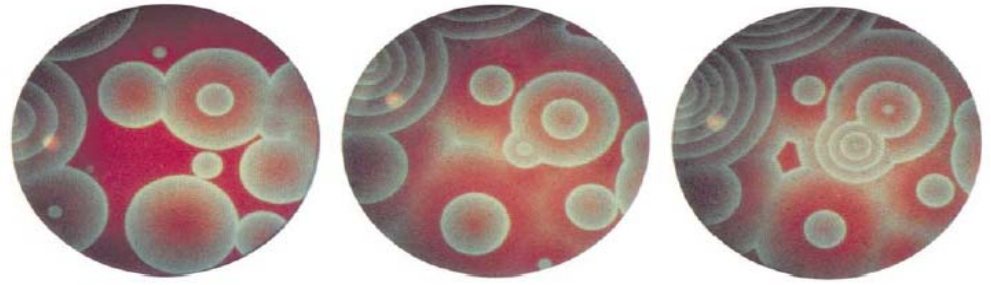


Alan M. Turing, 1912-1954

A.M. Turing. 1952. The chemical basis of morphogenesis.  
*Phil.Trans.Roy.Soc.London B* **237**:37-72.



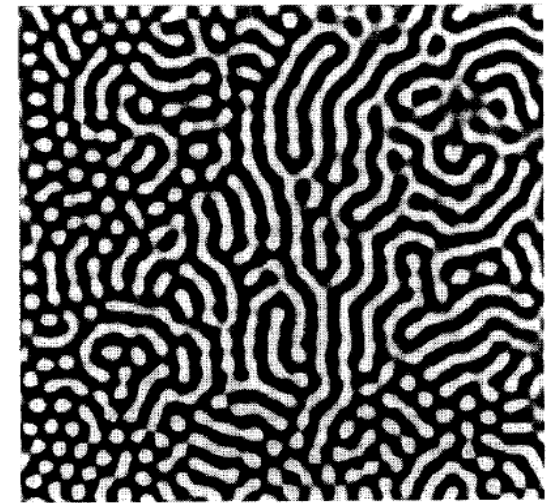
Liesegang Ringe 1895



Belousov-Zhabotinskii Reaktion 1959

Musterbildung durch chemische Selbstorganisation:

Liesegang Ringe durch Fällung aus übersättigten Lösungen,  
Raum-Zeit-Muster in der Belousov-Zhabotinskii Reaktion,  
und stationäre Turing Muster.



Turing Muster:  
Boissonade, De Kepper 1990

Blitze, Tornados, Hurrikans, Reaktions-Diffusionsmuster können nicht bis ins letzte Detail vorausgesagt beziehungsweise reproduziert werden, da lokale, mikroskopische Prozesse für die Nukleation oder Keimbildung verantwortlich sind, von welchen ausgehend der in seinen globalen Entstehungsbedingungen gut verstandene Prozess seinen Ausgang nimmt. In anderen Worten wir können notwendige Bedingungen angeben aber diese sind nicht hinreichend, um das genaue „Wann-und-Wo“ des Eintretens von Musterbildung vorherzusagen.

In der Biologie stoßen wir auf Muster, die hinsichtlich Form und Gestalt ganz ähnlich beschaffen sind wie die anorganischen Muster, auch gelten ähnliche Randbedingungen für ihre Ausbildung. Alan Turing lag mit seiner Vision der Musterbildung als Folge einer nicht-linearen chemischen Reaktion nicht ganz falsch. Es besteht aber ein grundlegender Unterschied hinsichtlich des Ergebnisses: Bei denselben Anfangsbedingungen (Genetik, Epigenetik, ...) ist das Endresultat fast identisch. Als Beispiel zeigen wir das Gesicht von drei verwilderten Katzen, die einer Inzuchtpopulation auf einer Hochfläche in den Nockbergen entstammen: Die Katzenmutter, das Junge und der mutmaßliche Katzenvater tragen ein bis ins Detail gehend gleiches Muster auf der Stirn.





Mutter



mutmaßlicher  
Vater

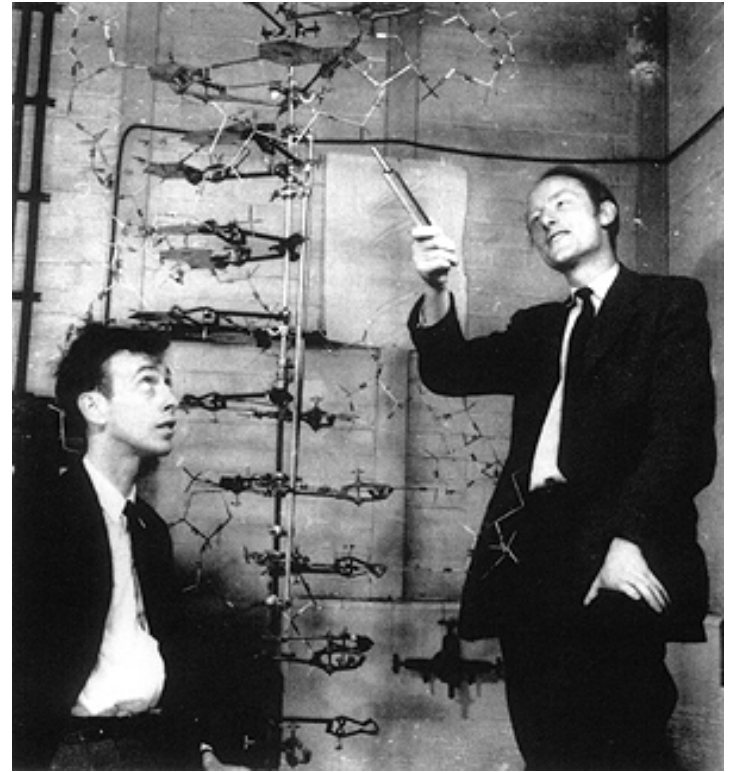
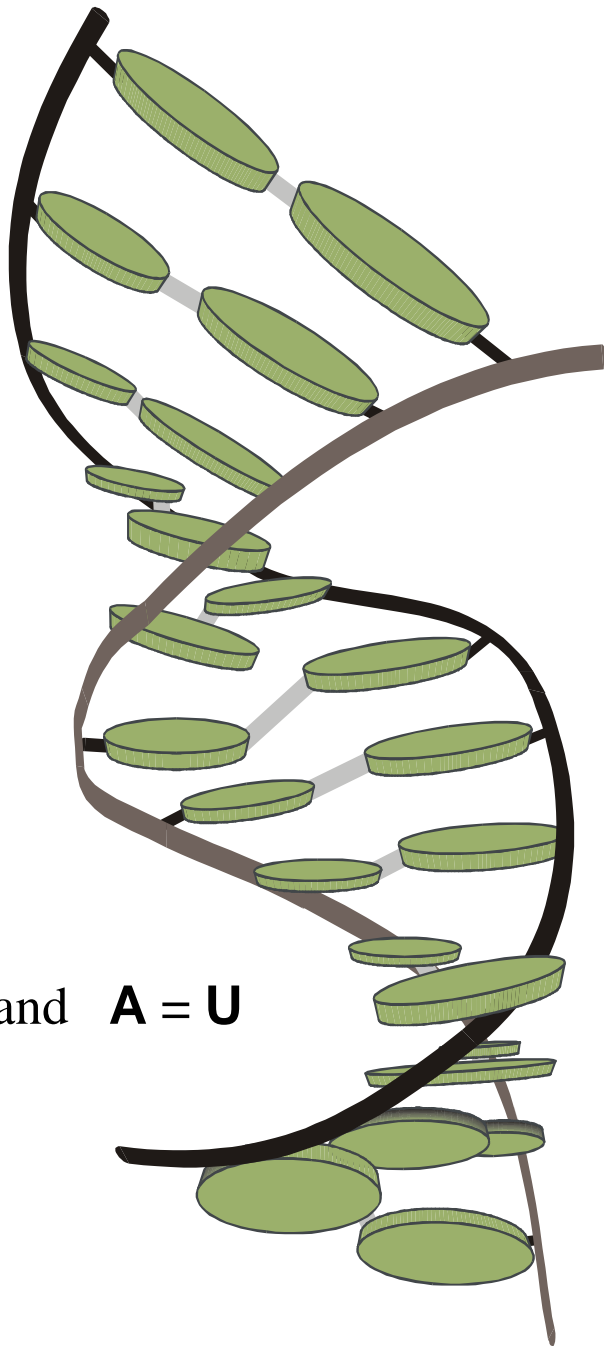


Tochter

Fellmuster in einer  
Inzuchtlinie von  
verwilderten Katzen

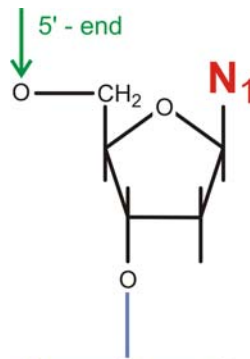
Eltern und Tochter

Die Erklärung des Unterschieds zwischen physikalischen und chemischen Strukturen einerseits und biologischen Mustern andererseits ist einfach: Beim Auftreten notwendiger Bedingungen kommt es in physikalisch-chemischen Systemen zur Musterbildung ausgehend von Keimen, deren Bildung von den lokalen Gegebenheiten eines zumeist stark fluktuierenden Hintergrundes abhängt. In der Biologie ist derselbe Sachverhalt überlagert von einem genetischen Programm, das – zusammen mit einigen epigenetischen Faktoren – die Musterentwicklung steuert. Der Einfluss lokaler Fluktuationen wird dabei fast gänzlich ausgeschaltet. Das entwicklungsbiologische Programm ist auf der Desoxyribonukleinsäure (DNA) einer jeden Zelle des Organismus gespeichert. Die Struktur des DNA-Moleküls ist auch die Basis für den Vermehrungsprozess.

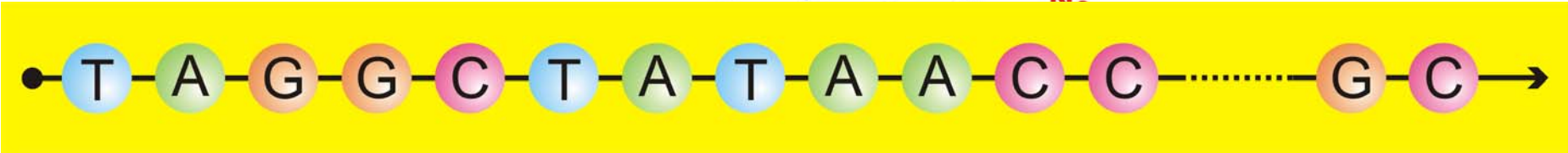


James D. Watson, 1928- , and Francis Crick, 1916-2004,  
Nobel Prize 1962

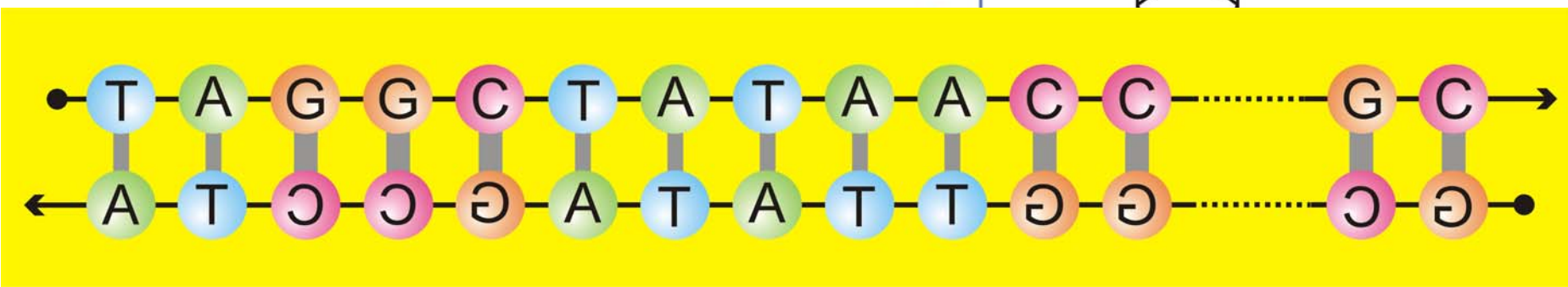
The three-dimensional structure of a  
short double helical stack of B-DNA



- $N_k =$
- A ≡ Adenine
  - T ≡ Thymine
  - G ≡ Guanine
  - C ≡ Cytosine

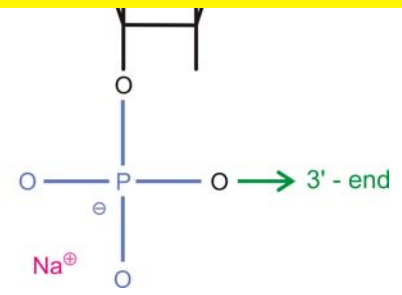


Duplication of genetic information



Deoxyribonucleic acid – DNA

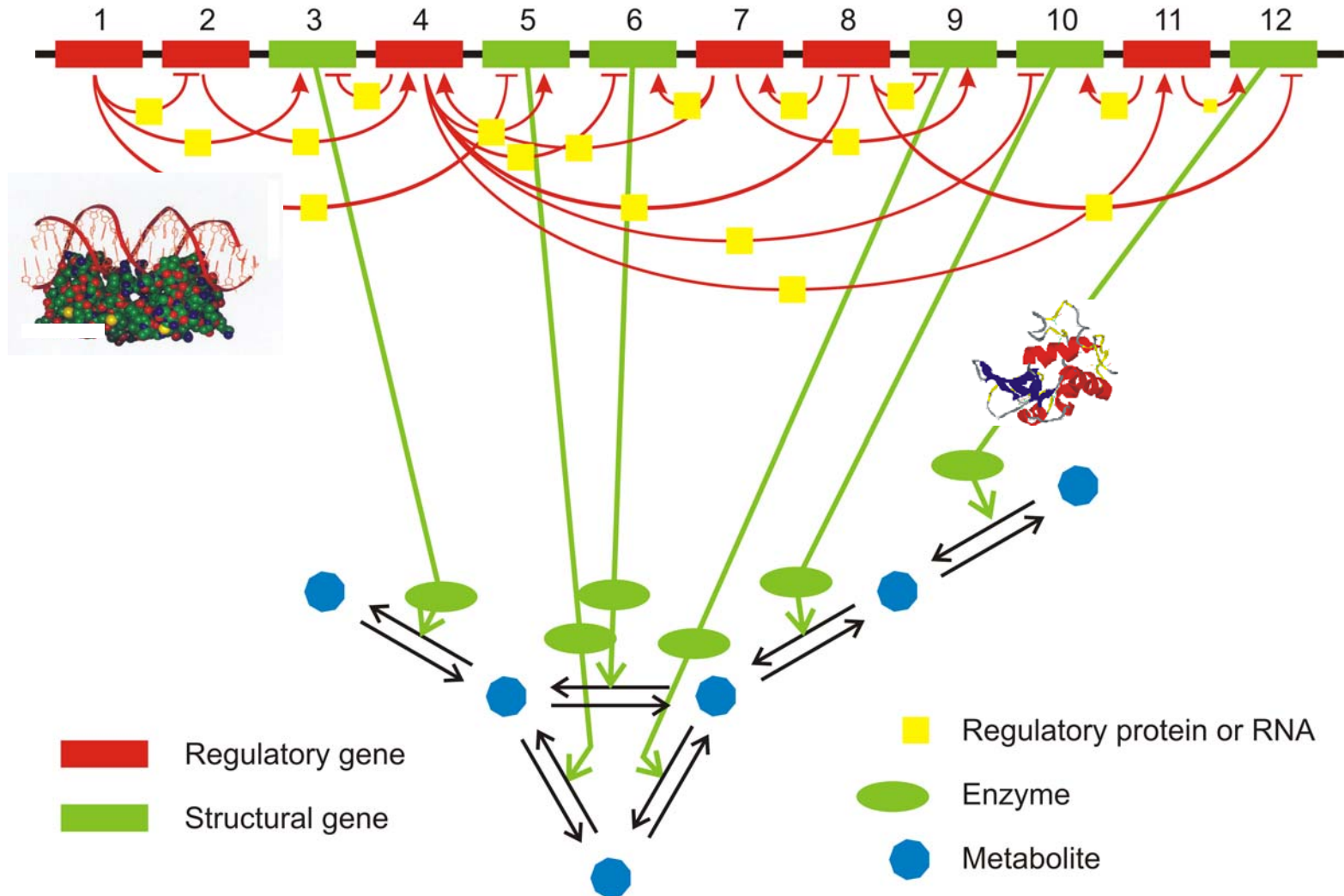
The carrier of digitally encoded information



Die DNA einer jeden Zelle wird teilweise transkribiert (abgeschrieben) und in Proteine oder Ribonukleinsäuren (RNA) translatiert (übersetzt). Die Protein- und RNA-Moleküle übernehmen verschiedene Aufgaben in den Zellen; ihre Strukturen und Eigenschaften sind letzten Endes in den DNA-Stücken oder Genen kodiert, aus welchen sie durch Übersetzung gebildet werden. Zur Steuerung des Zellstoffwechsels und der organismischen Entwicklung ist es notwendig die Aktivität der einzelnen Gene zu steuern. Dies geschieht bei Bakterien durch Regulatorgene, welche die Aktivitäten der Strukturgene kontrollieren. Bei höheren Organismen wird die Regulation sehr stark durch andere Faktoren beispielsweise kleine RNA-Moleküle mitbestimmt.

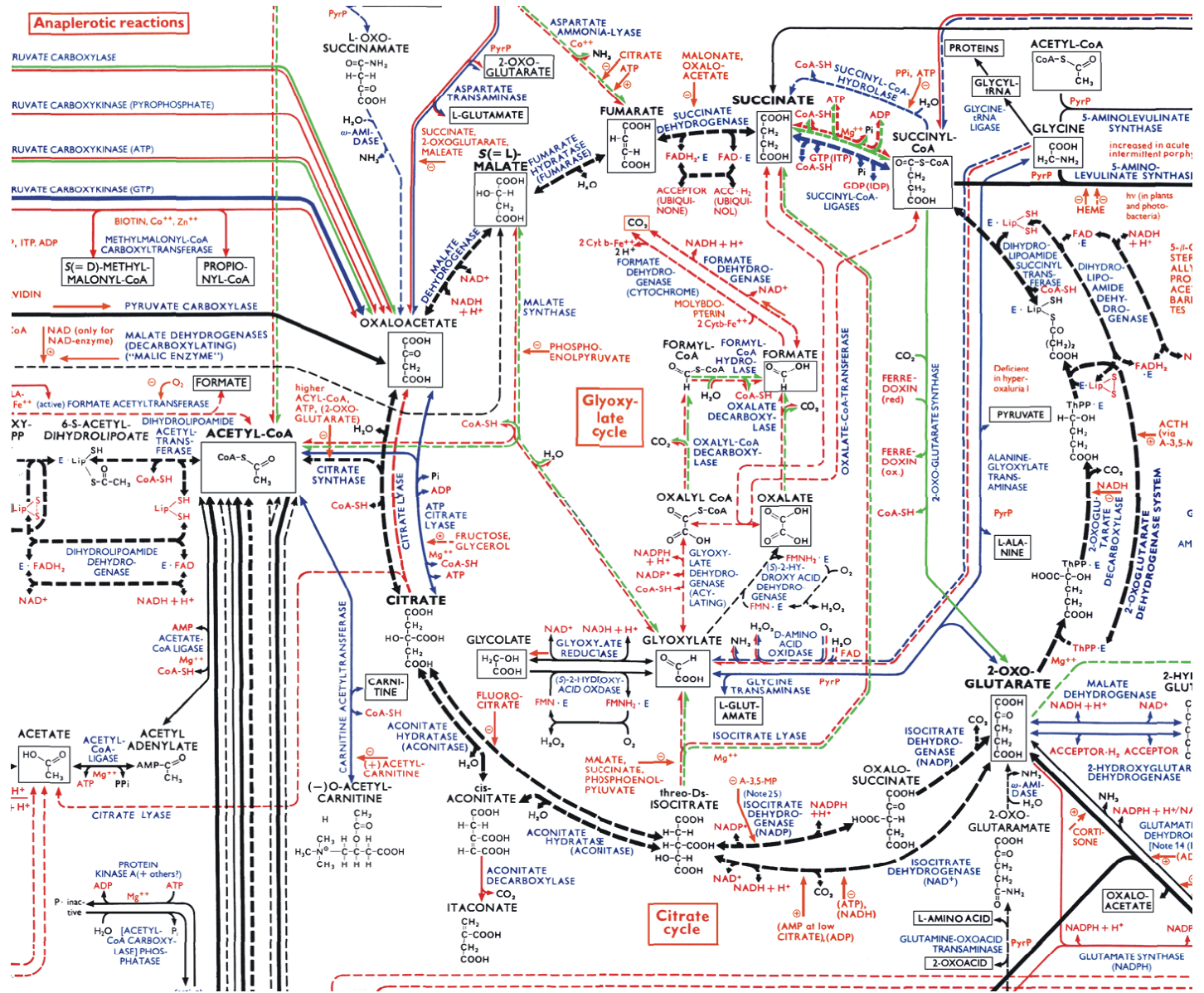
Der Metabolismus der Zellen muss alle notwendigen nicht eigens importierten Bausteine der Moleküle in der Zelle synthetisieren. Das entsprechende metabolische Netzwerk, welchen beim Menschen unter der Kontrolle von etwa 25 000 Regulator- und Strukturgenen steht, ist ziemlich kompliziert und unübersichtlich.

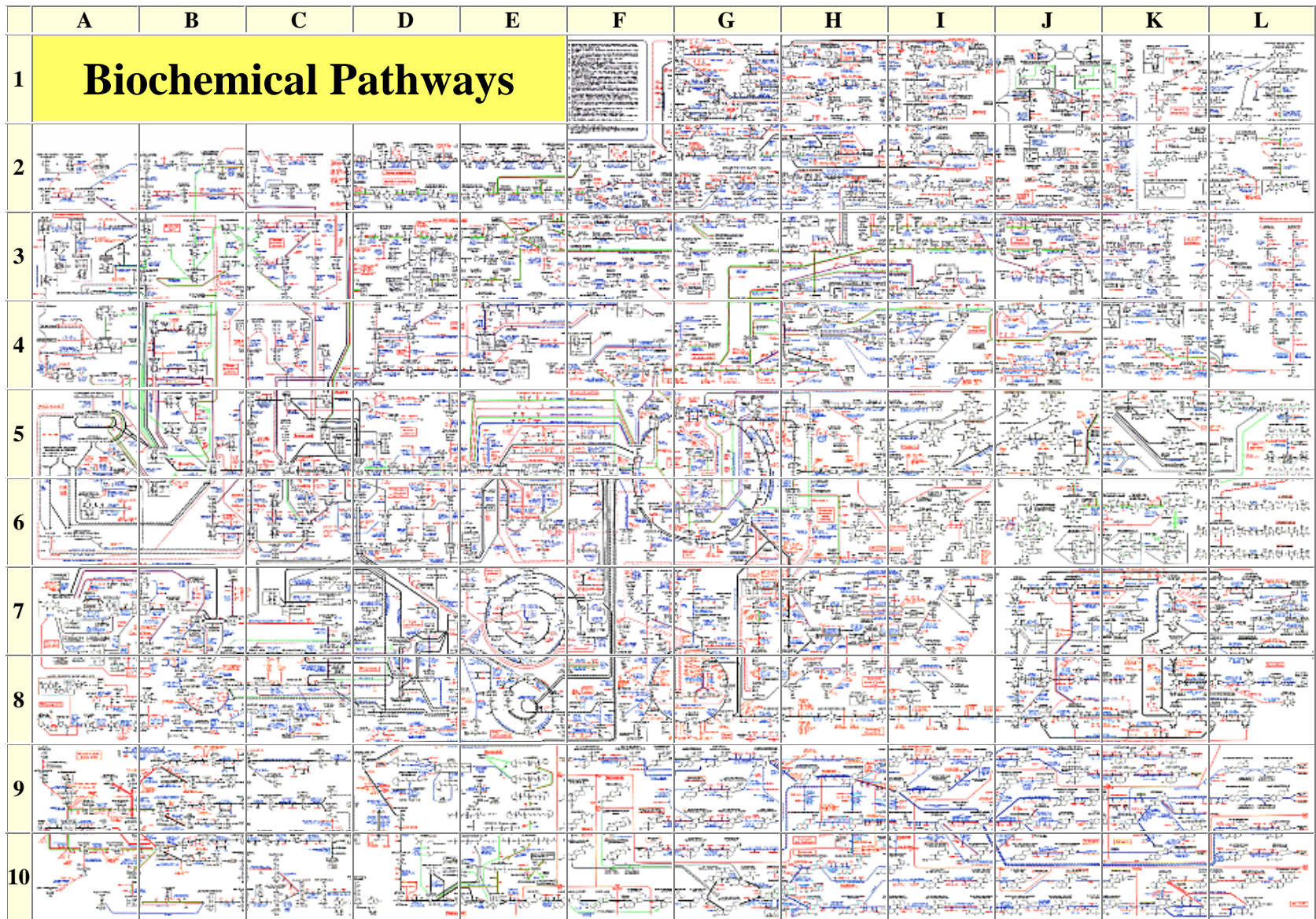
# A model genome with 12 genes



A sketch of a genetic and metabolic network

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





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



Das Darwinsche Prinzip der biologischen Evolution basiert auf den Fähigkeiten zu Vermehrung, Variation und Selektion und macht keine Voraussetzungen hinsichtlich der Einheiten, welche sich vermehren und variieren. Es sollte daher möglich sein, ohne zellulären Strukturen nur mit geeigneten Molekülen in Lösung evolutionäre Prozesse zu beobachten.

Seit den 1960iger Jahren werden laufend Evolutionsexperimente mit RNA-Molekülen in zellfreien Assays durchgeführt und diese Arbeiten haben schließlich zur Entstehung eines eigenen Zweiges der Biotechnologie, genannt Evolutionäre Biotechnologie oder „Applied Molecular Evolution“, geführt mit dem heute Moleküle nach Maß für bestimmte Zwecke hergestellt werden können.



Drei notwendige Bedingungen für Darwinsche Evolution:

1. Vermehrung,
2. Variation, and
3. Selektion.

**Variation** in Form von Rekombination und/oder Mutation verändert die **Genotypen** wogegen **Selektion** nur auf den **Phänotypen** operiert.

Im Darwinschen Szenario treten **Variationen** in Form von Rekombinations- und/oder Mutationsereignissen **unkorreliert** mit ihren **Effekt auf den Selektionsprozess** auf und erscheinen daher **zufällig**.

Alle drei Bedingungen werden nicht nur von zellulären Organismen erfüllt sondern auch von **Molekülen** in geeigneten **zellfreien Assays**.

random individuals. The primer pair used for genomic DNA amplification is 5'-TCTCGCTGGATTCCTCATTTA-3' (forward) and 5'-TCTTTGTCTTCTGTGTGGAC-3' (reverse). Reactions were performed in 25  $\mu$ l using 1 unit of Taq DNA polymerase with each primer at 0.4  $\mu$ M, 200  $\mu$ M each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>] in a cycle condition of 94°C for 1 min and then 95 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 90 s followed by 72°C for 5 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)<sup>+</sup> 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 conditions 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 [S; K.-S. Chen, L. Polakoff, J. R. Lupski, *Mol. Cell. Res.* 164, 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. Fridal, data not shown.

36. K. B. Avraham et al., *Nature Genet.* 11, 390 (1995); X.Z. Liu et al., *Hum. Mol. Genet.* 17, 268 (1997); F. Gibson et al., *Nature* 374, 62 (1996); D. Wall et al., *Hum. Mol. Genet.* 17, 268 (1997).

37. RNA was extracted from cochlea (membranes labyrinthis) 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)<sup>+</sup> selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-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'-GGATGACCTGGCGTAAATGGG-3'; reverse, 5'-CTCAGGGCTTGTGATGGTGGCTGGCTGGG-3'). Cycling conditions were 40 s at 94°C, 40 s at 55°C (5 cycles), 50°C (5 cycles), and 55°C (23 cycles), and 45 s at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 588-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 2003-bp fragment.

38. We are grateful to the people of Bengalia, 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. Ferguson, A. Gupta, E. Sorbello, R. Torkazad, C. Varnier, M. Walker, G. Bourlart, and S. Backstrom-Stenberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinman, L. N. Arhys, and S. Winita for assistance in Bali, and T. Barber, S. Sullivan, E. Green, D. Dayna, and J. Batty for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (ZD1 DC 00395-01 and ZD1 DC 00398-01 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.F. 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 empir-

ically 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 (replicable 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

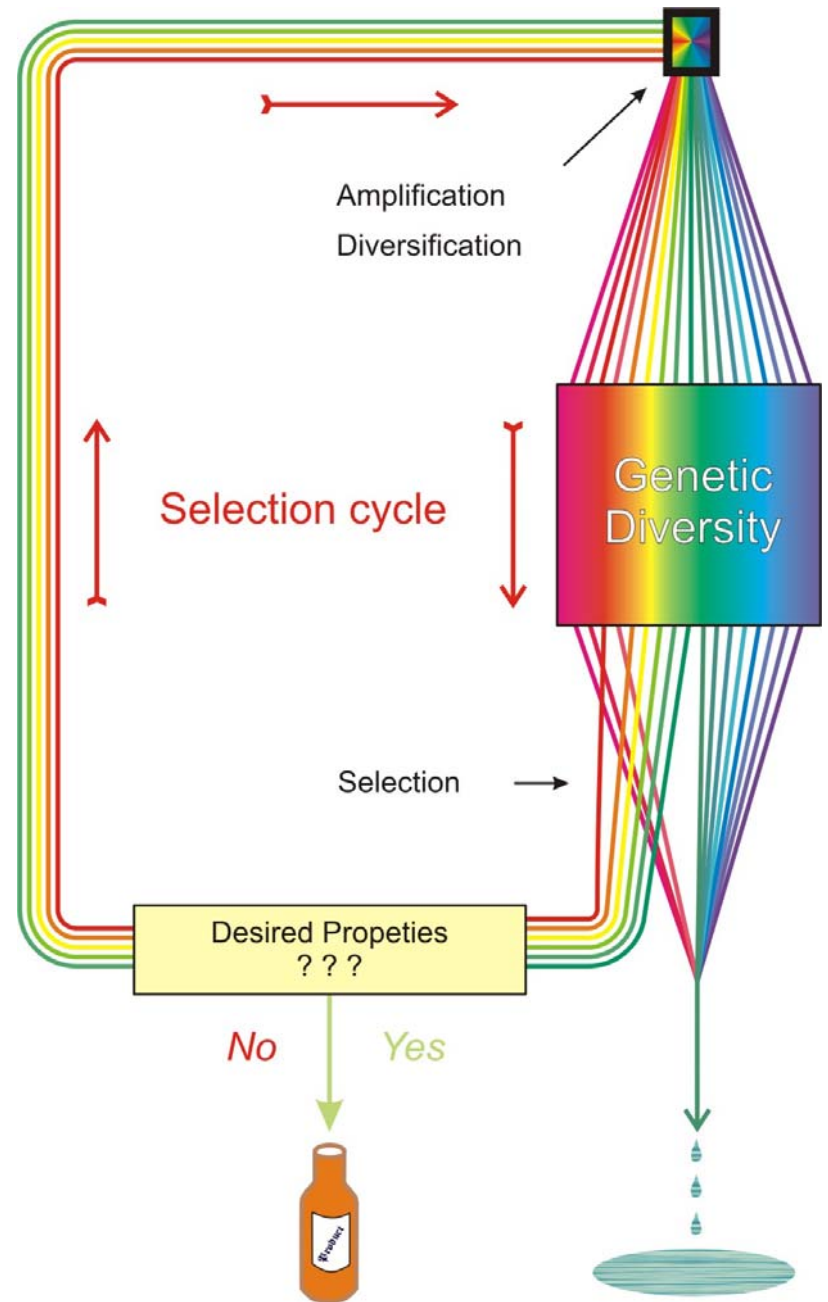
the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

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.

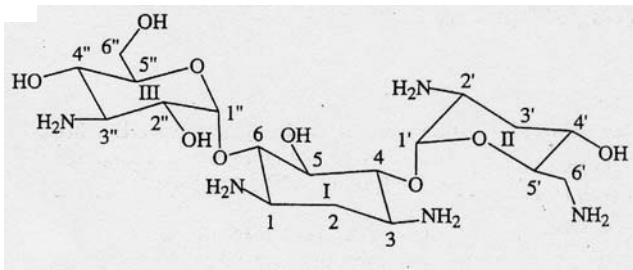
A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally 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

## Stochastic simulation of evolution of RNA molecules

Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1080 Wien, Austria, Santa Fe Institute, 1390 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.



An example of 'artificial selection' with RNA molecules or 'breeding' of biomolecules



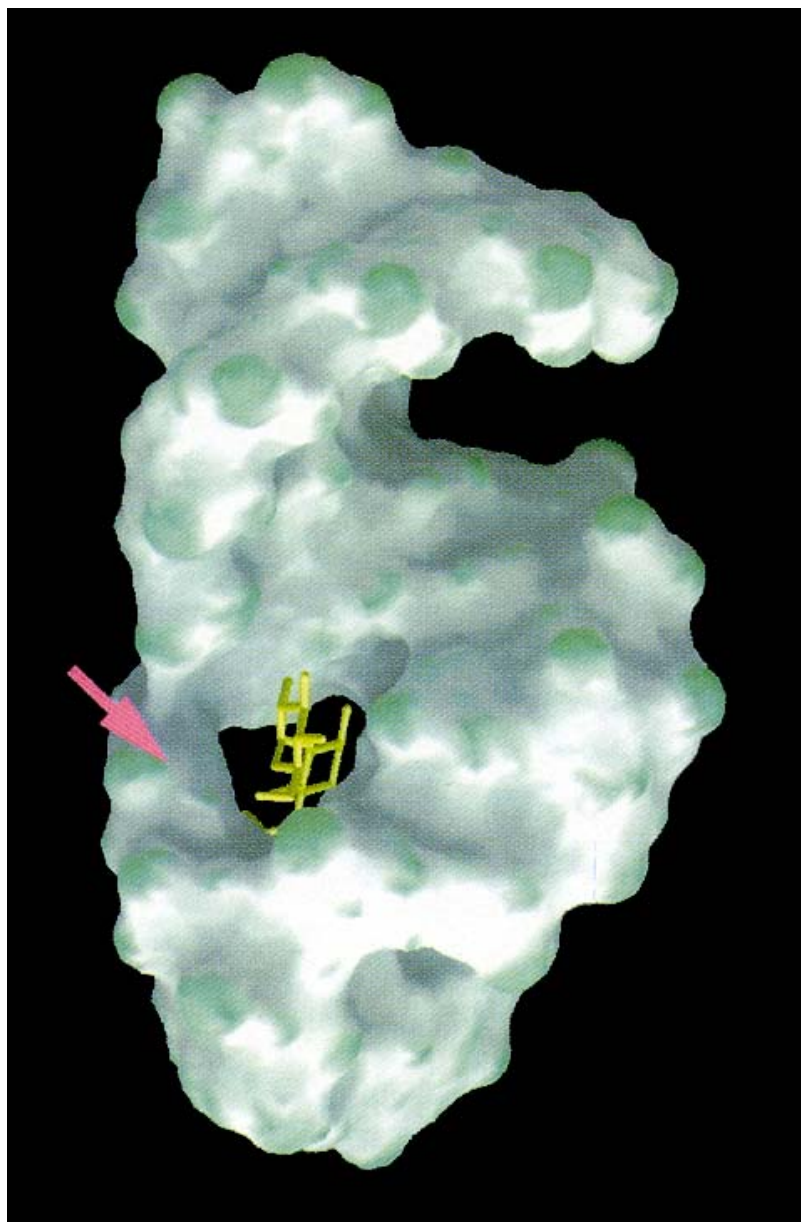
tobramycin



RNA aptamer, n = 27

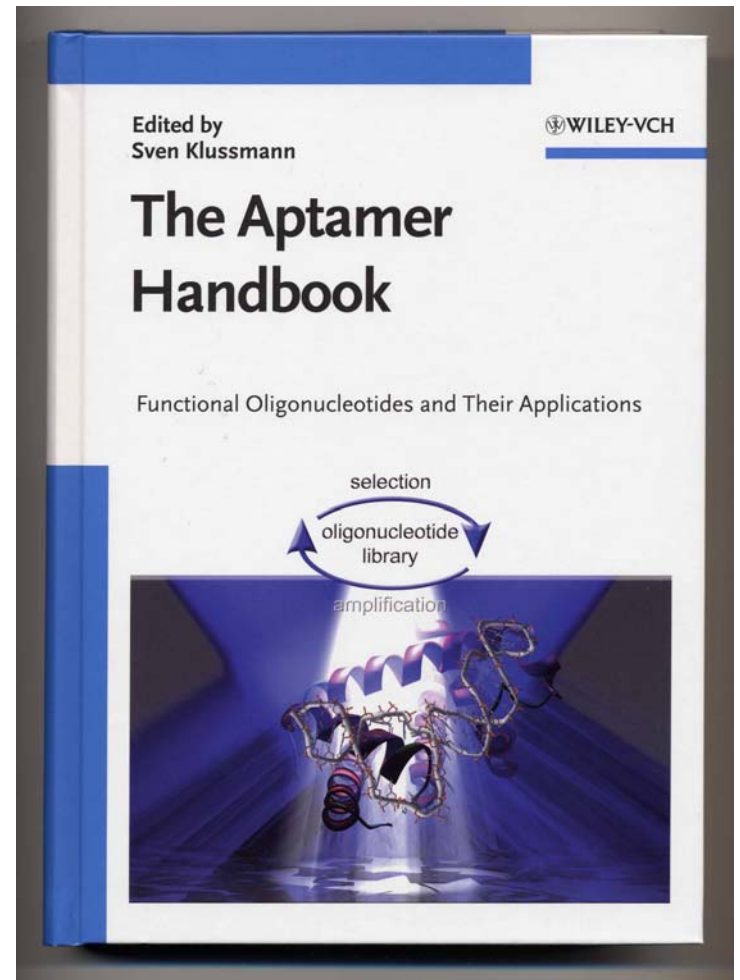
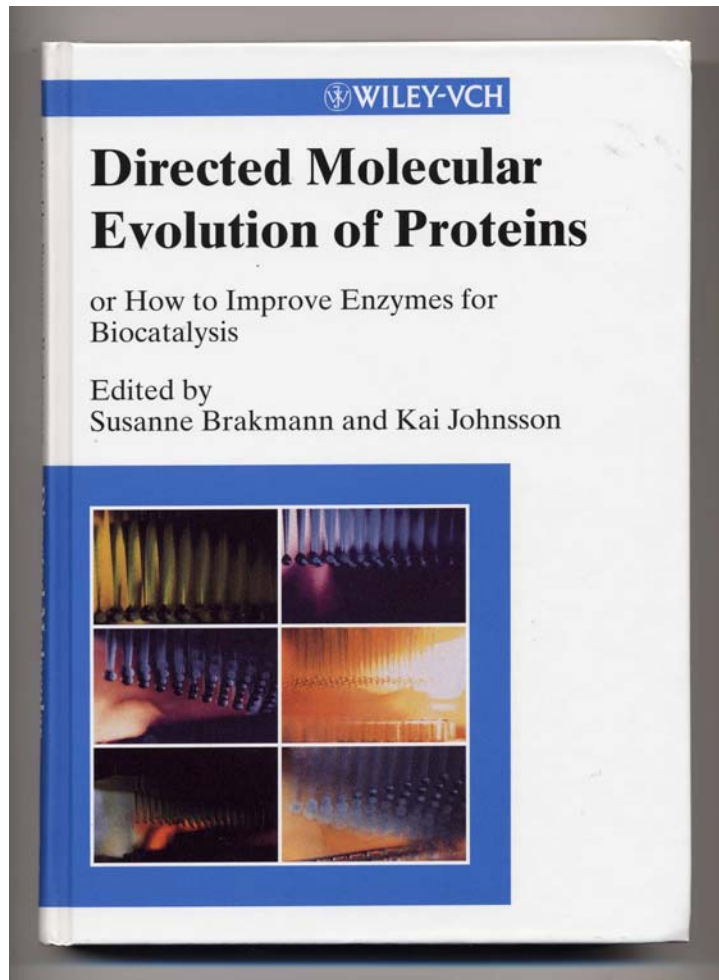
Formation of secondary structure of the tobramycin binding RNA aptamer with  $K_D = 9 \text{ nM}$

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

## Artificial evolution in biotechnology and pharmacology

G.F. Joyce. 2004. Directed evolution of nucleic acid enzymes. *Annu.Rev.Biochem.* **73**:791-836.

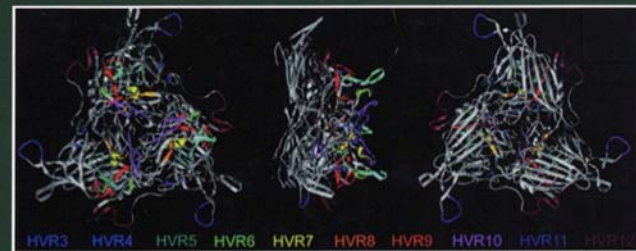
C. Jäckel, P. Kast, and D. Hilvert. 2008. Protein design by directed evolution. *Annu.Rev.Biophys.* **37**:153-173.

S.J. Wrenn and P.B. Harbury. 2007. Chemical evolution as a tool for molecular discovery. *Annu.Rev.Biochem.* **76**:331-349.



SECOND EDITION

# ORIGIN AND EVOLUTION OF VIRUSES



Edited by  
ESTEBAN DOMINGO  
COLIN R. PARRISH  
JOHN J. HOLLAND



Molecular evolution of viruses

Die gegenwärtige molekulare Evolutionsbiologie befasst sich mit einigen neuen Phänomenen, deren Bedeutungen in der Zeit der Ausarbeitung der Synthetischen Theorie (1940-1950) noch nicht abschätzbar waren und welche in den aktuellen Problemen der Molekulargenetik eine vorrangige Rolle spielen:

- (i) Neutralität: Ein Großteil, wenn nicht der größte Teil aller Mutationen sind für die Selektion nicht unterscheidbar – sie werden deshalb als neutral bezeichnet – und bilden das evolutionäre Variantenreservoir für geänderte Umweltbedingungen.
- (ii) Bastelprinzip: Die Evolution kennt keine zukünftigen Ziele und bastelt aus den vorhandenen Strukturen und Funktionen neue Organismen, die mit geänderten Bedingungen zurande kommen. Basteln führt auf der molekularen Ebene zu immer komplizierter werden Regulationsnetzwerken.
- (iii) Somatische Vielfalt: Dasselbe Stück DNA kann in den verschiedenen Geweben oder Organen eines Organismus in verschiedene Proteine übersetzt werden, wodurch der konventionelle Begriff des Gens in Frage zu stellen ist.
- (iv) Transkribierte aber nicht translatierte DNA: In der frühen Molekulargenetik wurde klar, dass nur ein ziemlich kleiner Teil der DNA von höheren Organismen in Protein übersetzt wird. Man sprach dann von „Junk-DNA“, für die keine Funktion erkennbar war. Beim Menschen lag der Anteil an „Junk-DNA“ sogar zwischen 90 – 95%. Mittlerweile wurde klar, dass der ganz überwiegende Anteil (bis zu 90%) dieser DNA transkribiert wird. Nach der Funktion der vielen, auf diese Weise gebildeten RNA-Moleküle wird gegenwärtig intensiv geforscht.

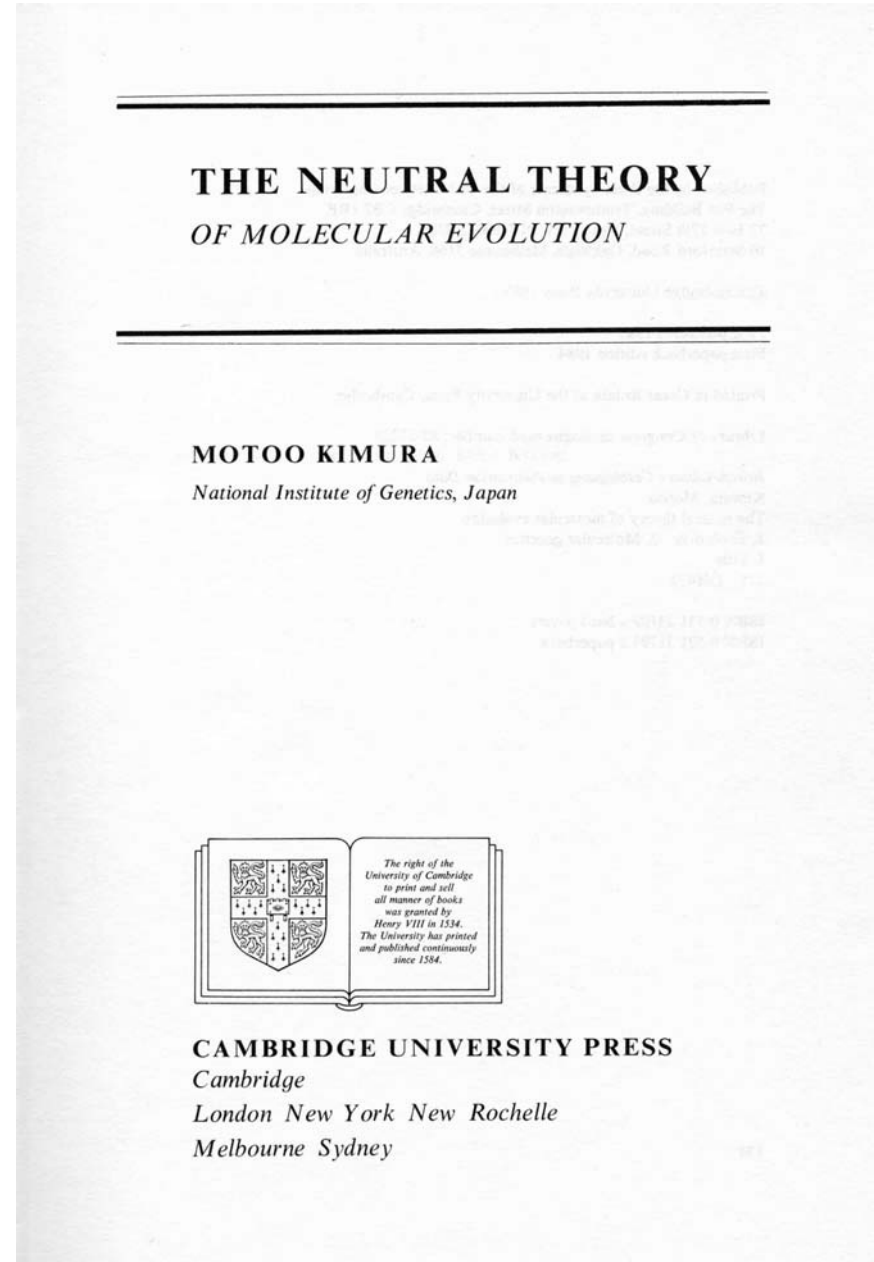
Eine erschöpfende Beantwortung aller dieser Fragen benötigt eine vollständige chemische Analyse aller Vorgänge in den Zellen, welche durch die vollständige Sequenzierung von Genomen eingeleitet wurde. Im Rahmen der Systembiologie wird dieses Programm zur Zeit auf die Erfassung der Funktionen aller Moleküle ausgedehnt.



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.



---

---

EVOLUTIONARY TINKERING

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

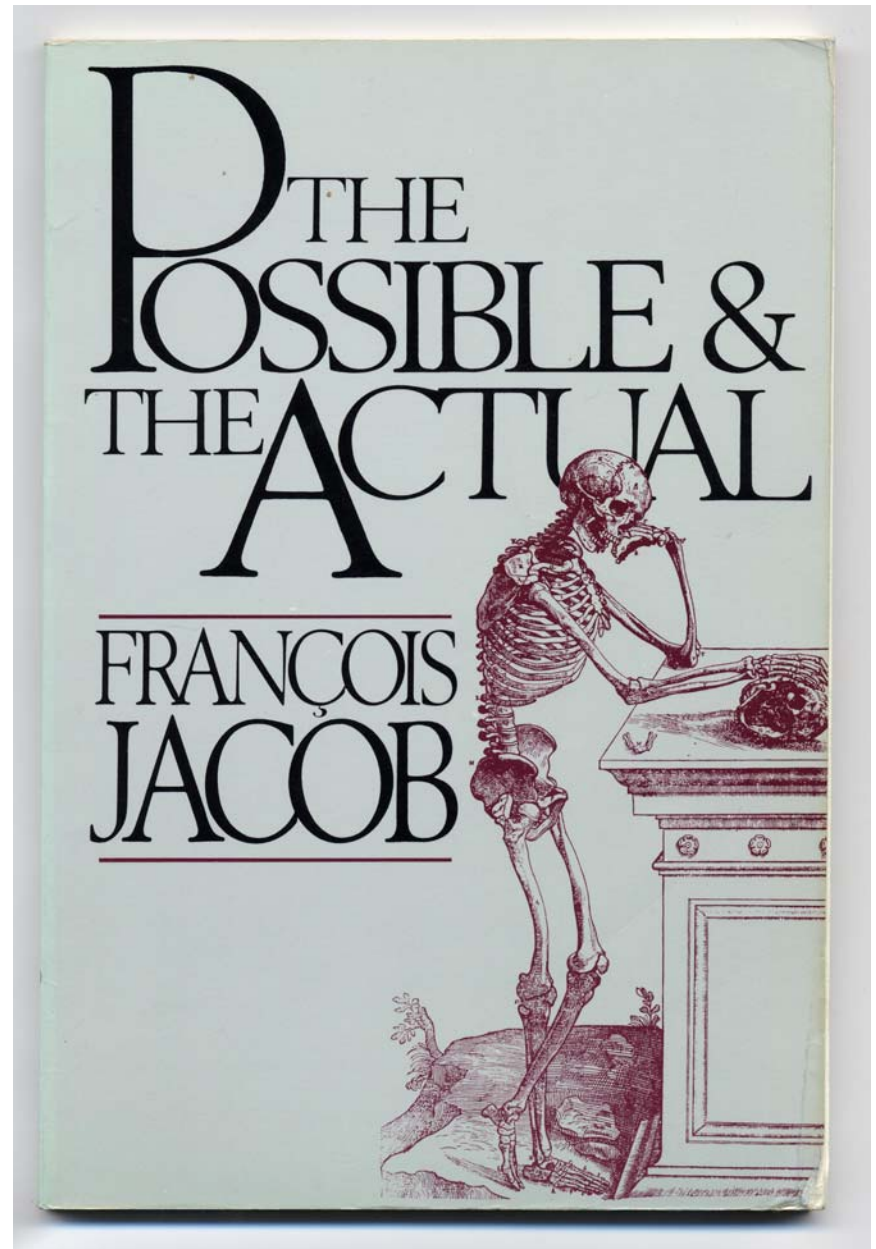
—Woody Allen, *Getting Even*

---

---

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

Francois Jacob, Pantheon Books,  
New York 1982



# 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<sup>1</sup>. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals<sup>2</sup>. 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."

## 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 pre-

viously unimagined scope of RNA.

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<sup>3</sup>.

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<sup>4</sup>, and one by geneticist Rotem Sorek<sup>5</sup>, 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

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

— Phillip Kapranov



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

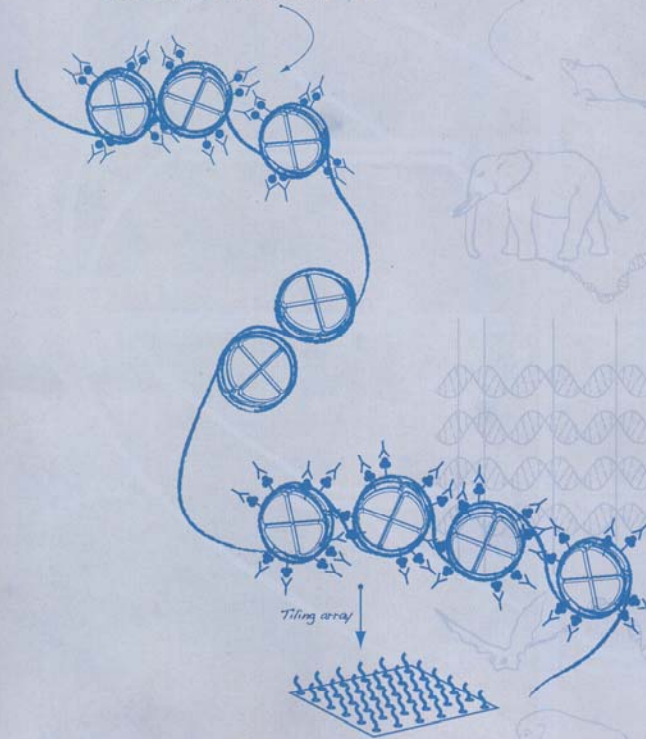
The difficulty to define the notion of „gene“.

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

# nature

*Hi-stone-modification chromatin IP*

*Comparative syntenic alignment*



**MARS'S  
ANCIENT OCEAN**  
Polar wander  
solves an enigma

**THE DEPTHS OF  
DISGUST**  
Understanding the  
ugliest emotion

**MENTORING**  
How to be top

**NATUREJOBS**  
Contract  
research

## DECODING THE BLUEPRINT

The ENCODE pilot maps  
human genome function



ENCODE stands for  
**ENC**yclopedia **O**f **DNA** **E**lements.

**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

Web-Page for further information:

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