Some Mathematical Challenges from Life Sciences

Part I

Peter Schuster, Universität Wien
Peter F.Stadler, Universität Leipzig
Günter Wagner, Yale University, New Haven, CT
Angela Stevens, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig
and
Ivo L. Hofacker, Universität Wien

Oberwolfach, GE, 16.-21.11.2003

Web-Page for further information:

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

1. **Mathematics and the life sciences in the 21st century**

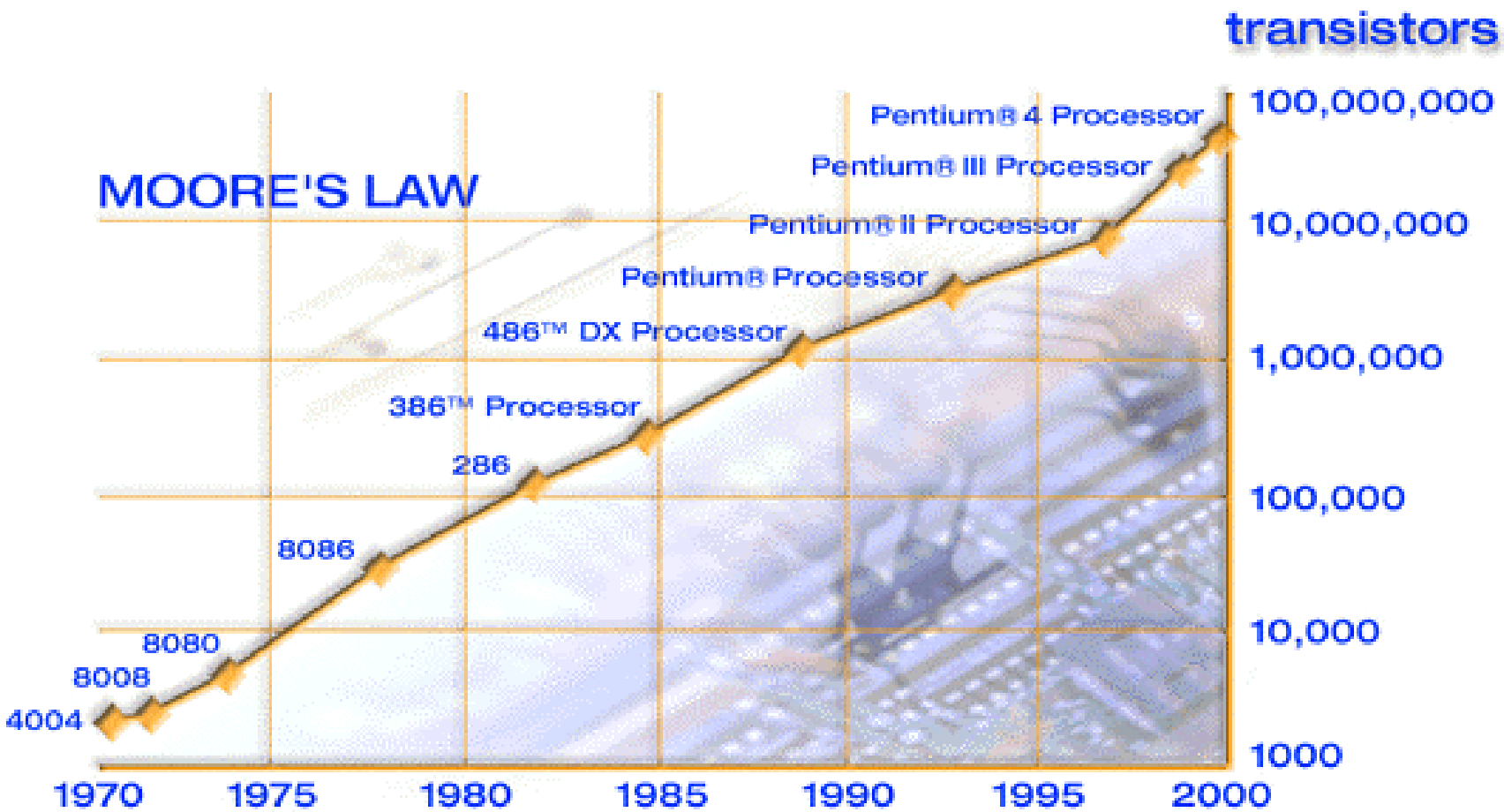
2. **Selection dynamics**

3. **RNA evolution *in silico* and optimization of structure and properties**

1. **Mathematics and the life sciences in the 21st century**

2. Selection dynamics

3. RNA evolution *in silico* and optimization of structure and properties


transistors
MOORE'S LAW
Pentium® 4 Processor
Pentium® III Processor
Pentium® II Processor
Pentium® Processor
486™ DX Processor
386™ Processor
286
8086
8080
8008
4004
100,000,000
10,000,000
1,000,000
100,000
10,000
1000
1970
1975
1980
1985
1990
1995
2000

Zur gleichen Zeit schreien viele nach einer neuen Biologie. Man liest, sie wollen „Integrative Biologie“ machen, oder „Systembiologie“. Kaum einer nennt es beim richtigen Namen: Theoretische Biologie. Weil diese einen schlechten Klang hat. Ich jedoch denke, ich kann die Sünden der Vergangenheit vergeben und nehme das Wort: Wir brauchen eine Theorie, die das alles einschließt. Stellen Sie sich doch nur mal vor, wir müssen am Ende all dieses Zeug nicht nur unter Fachleuten besprechen, sondern müssen es an Universitäten lehren, in der Schule, und es der Öffentlichkeit erklären. Wie sollen wir das machen ohne umfassende Theorie? Das, denke ich, ist die Herausforderung, der wir uns stellen müssen.

At the same time people are crying for a new biology. They say, they want to make “Integrative Biology” or “Systems Biology”. Hardly anyone calls it by its proper name: Theoretical Biology. Because it has a bad reputation. I think, however, I can remit the sins of the past and declare: We need a theory, which comprises all that (*Molecular, Structural, Cellular, Developmental, …… , and Evolutionary Biology*). Imagine, eventually, we not only need to discuss all this stuff with our expert colleagues, but we have to teach it at universities, at schools, and to the public. How could we manage without a comprehensive theory? This is the challenge we have to meet.

Sydney Brenner im Gespräch: „*Eine einsame Stimme aus der Prägenomik Ära*“. Laborjournal 2002, Heft 4:28 – 33.

Time to free tomorrow's biologists from pre-med tyranny?

10 September 2002 17:35 EST

by *Lois Wingerson*

Quality training for the biologists of the future depends on liberating life-science programs from the pre-med template and especially from the criteria of the Medical College Admissions Test (MCAT), according to a report from the US National Academy of Sciences (NAS), released today.

Asking colleges to rethink their entire undergraduate life-science curricula, the NAS committee also called for a greater focus on chemistry, physics, and math, more interdisciplinary subject materials, and mathematical curricula that go beyond calculus and statistics to embrace other quantitative skills relevant to life science not only today but tomorrow.

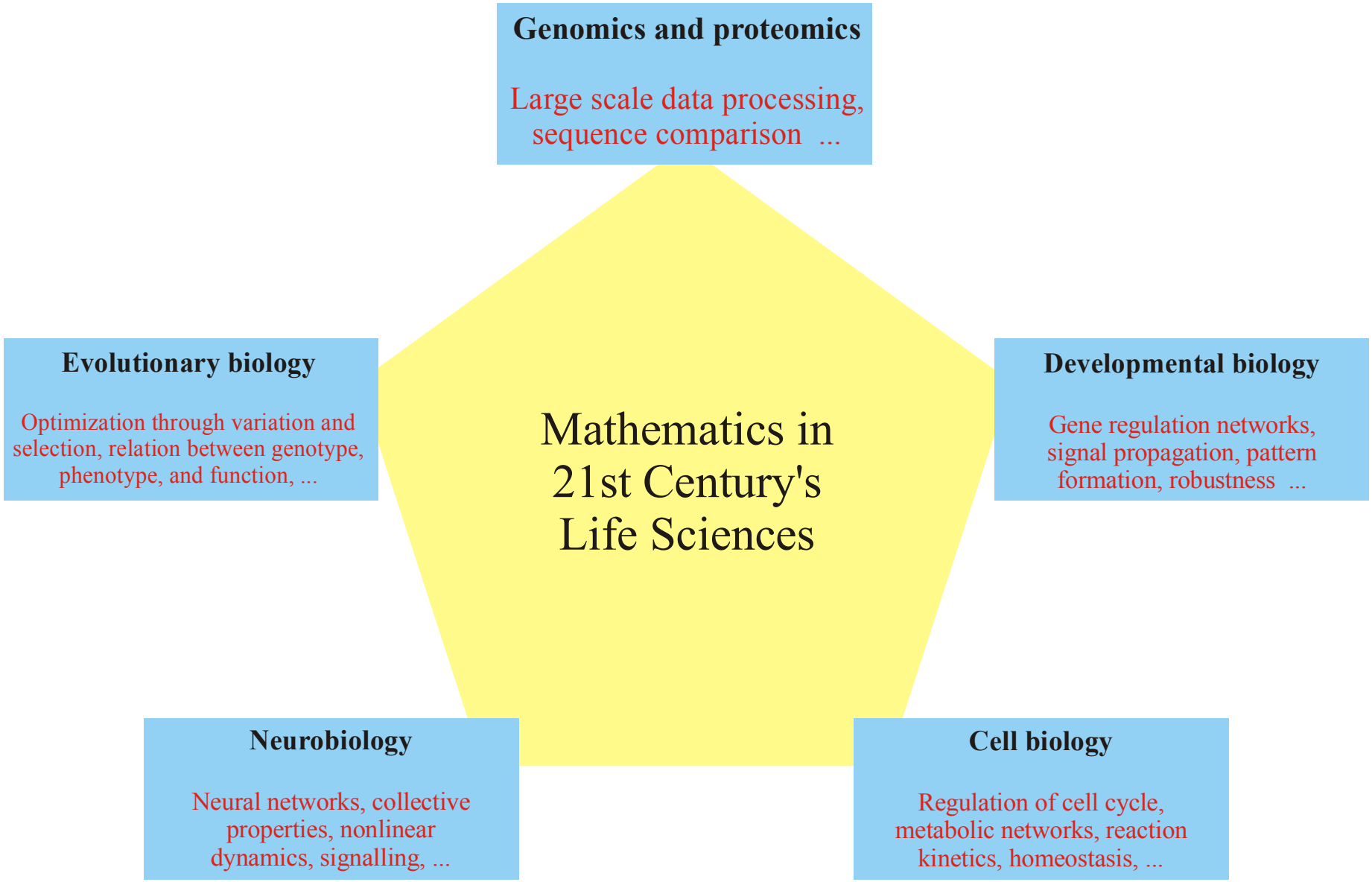
"Most biology students of today are being prepared for the biology of the past, not the future," said Stanford University neurology professor Lubert Stryer, chairman of the committee that wrote the report. Experiments such as imaging molecular motors, unimaginable 20 years ago, are now being carried out by graduate students, he noted, yet many Bio 101 students learn little more than "factoids."

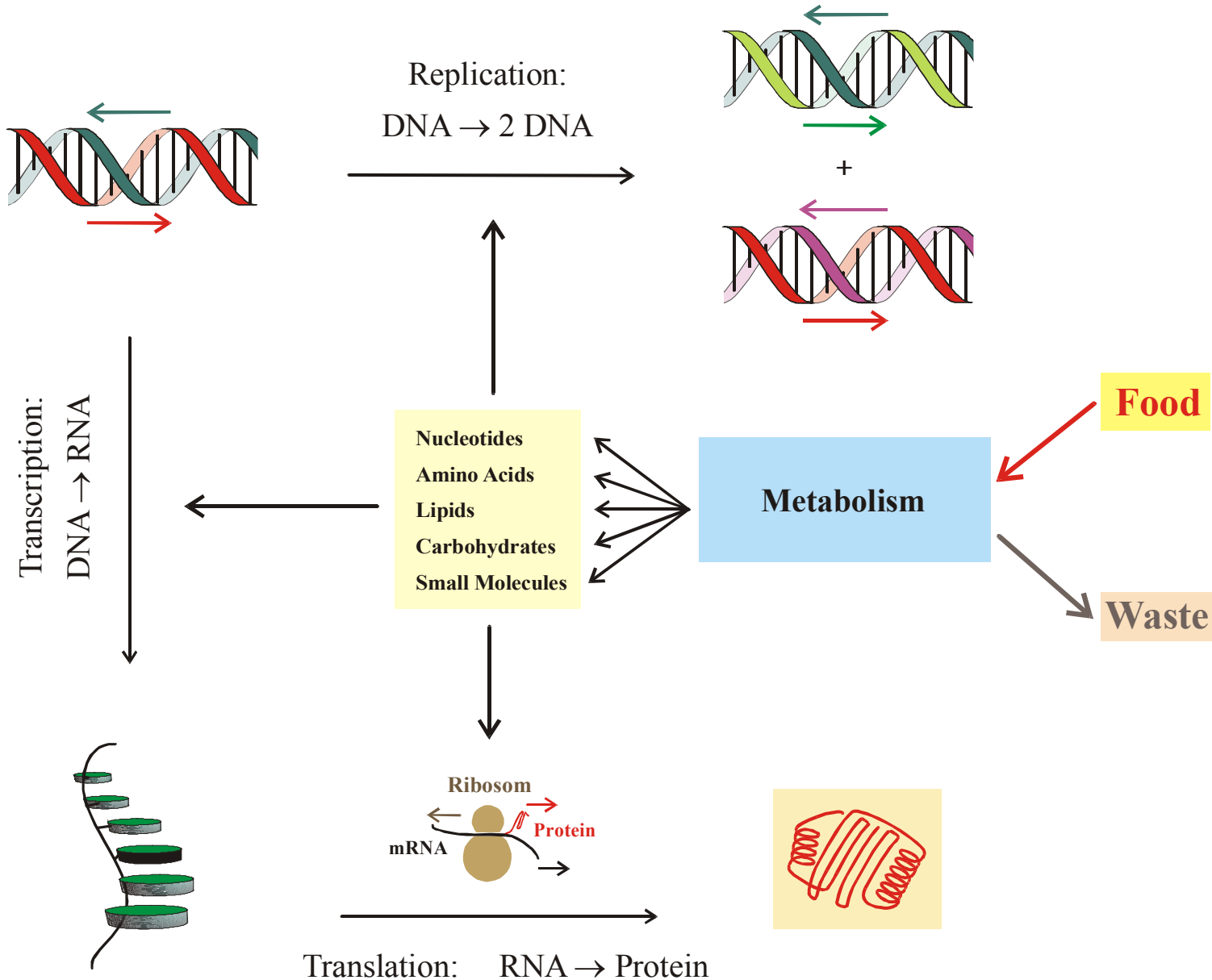
See also:

Comparison of problem-and lecture-based pharmacology teaching
[Opinion]
Martin C. Michel, Angela Bischoff and Karl H. Jakobs
Trends in Pharmacological Sciences, 2002, 23:4:168-170

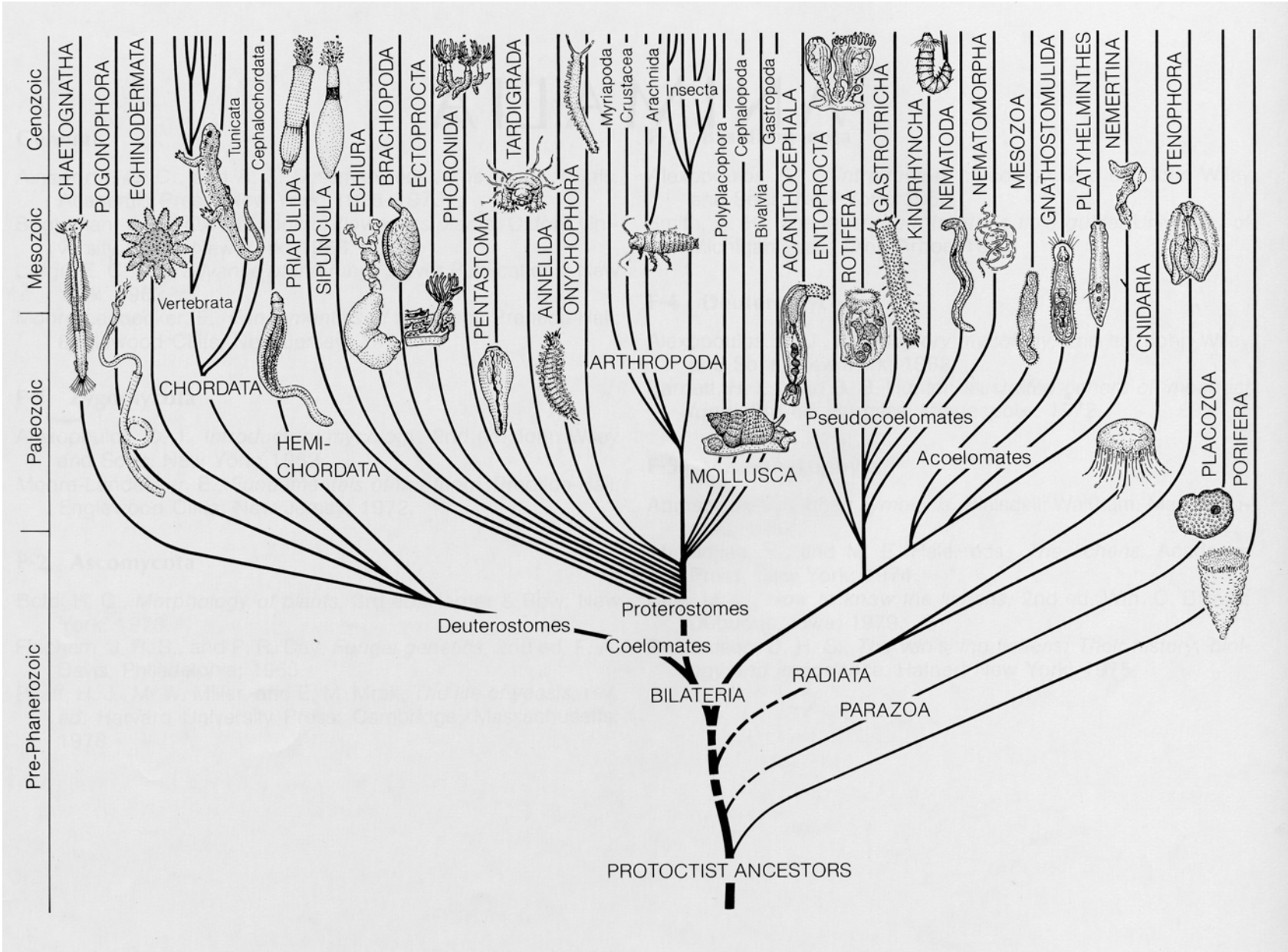
Teaching the scientific thrill
[In brief]
Stephanie Bono de
Trends in Biochemical Sciences, 2001, 26:11:647

Biochemistry and molecular biology teaching over the past 50 years
E.J. Wood
Nat Rev Mol Cell Biol, 2001 Mar 2:217-21


Genomics and proteomics
Large scale data processing, sequence comparison ...
Evolutionary biology
Optimization through variation and selection, relation between genotype, phenotype, and function, ...
Mathematics in 21st Century's Life Sciences
Developmental biology
Gene regulation networks, signal propagation, pattern formation, robustness ...
Neurobiology
Neural networks, collective properties, nonlinear dynamics, signalling, ...
Cell biology
Regulation of cell cycle, metabolic networks, reaction kinetics, homeostasis, ...



A sketch of cellular DNA metabolism



Five kingdoms.

L. Margulis, K.V. Schwartz, W.H.Freeman & Co., 1982

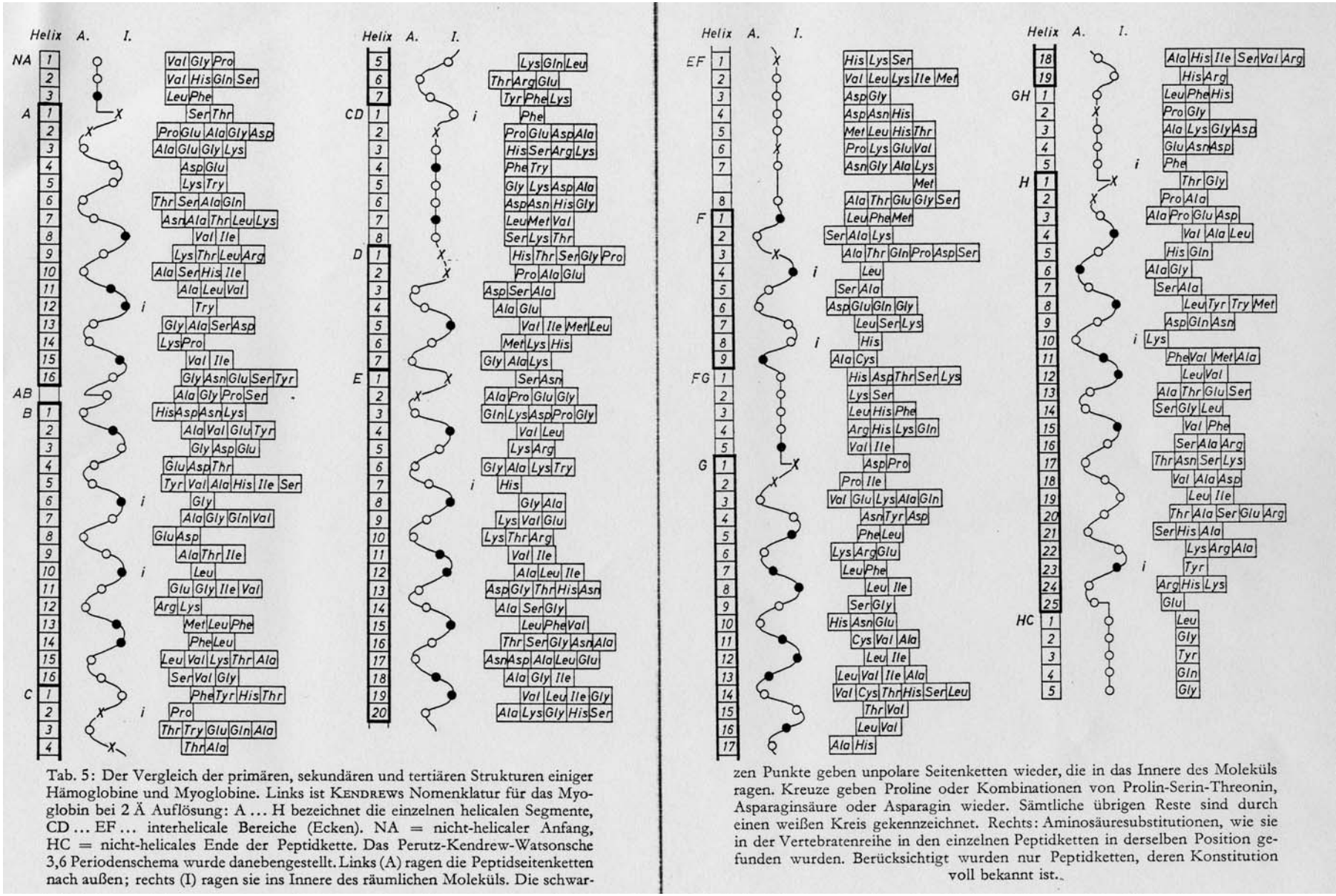
Five kingdoms.
L. Margulis, K.V. Schwartz,
W.H.Freeman & Co., 1982

Genomics and proteomics

Large scale data processing, sequence comparison ...

E. coli:	Length of the Genome	4×10^6 Nucleotides
	Number of Cell Types	1
	Number of Genes	4 000
Man:	Length of the Genome	3×10^9 Nucleotides
	Number of Cell Types	200
	Number of Genes	30 000 - 100 000

Gerhard Braunitzer, 1929 - 1989



Tab. 5: Der Vergleich der primären, sekundären und tertiären Strukturen einiger Hämoglobine und Myoglobine. Links ist KENDREWS Nomenklatur für das Myoglobin bei 2 Å Auflösung: A ... H bezeichnet die einzelnen helicalen Segmente, CD ... EF ... interhelicale Bereiche (Ecken). NA = nicht-helicaler Anfang, HC = nicht-helicales Ende der Peptidkette. Das Perutz-Kendrew-Watsonsche 3,6 Periodenschema wurde danebengestellt. Links (A) ragen die Peptidseitenketten nach außen; rechts (I) ragen sie ins Innere des räumlichen Moleküls. Die schwarzen Punkte geben unpolare Seitenketten wieder, die in das Innere des Moleküls ragen. Kreuze geben Proline oder Kombinationen von Prolin-Serin-Threonin, Asparaginsäure oder Asparagin wieder. Sämtliche übrigen Reste sind durch einen weißen Kreis gekennzeichnet. Rechts: Aminosäuresubstitutionen, wie sie in der Vertebratenreihe in den einzelnen Peptidketten in derselben Position gefunden wurden. Berücksichtigt wurden nur Peptidketten, deren Konstitution voll bekannt ist.

Sequence and structure of a-helices in hemoglobin

80

SONDERDRUCK

aus

Jahrbuch 1967 der Max-Planck-Gesellschaft
zur Förderung der Wissenschaften e.V.

*

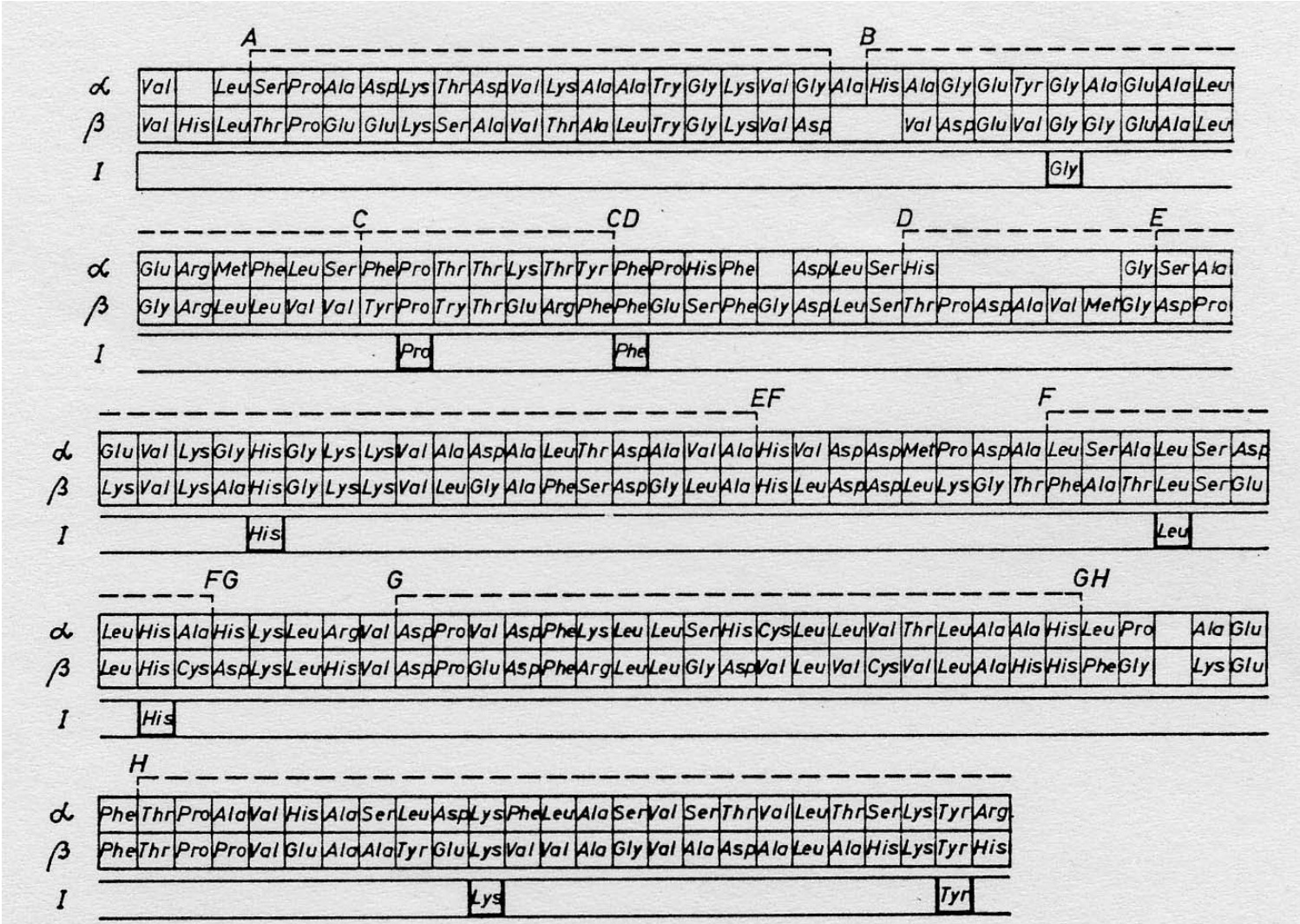
Molekularbiologie und Evolution

Von

Prof. Dr. GERHARD BRAUNITZER

Max-Planck-Institut für Biochemie, München

Molecular evolution through comparison of sequences from different organisms



Tab. 4: Die invarianten Reste (I) der Hämoglobine der Vertebraten.

Hemoglobin sequences in different vertebrates

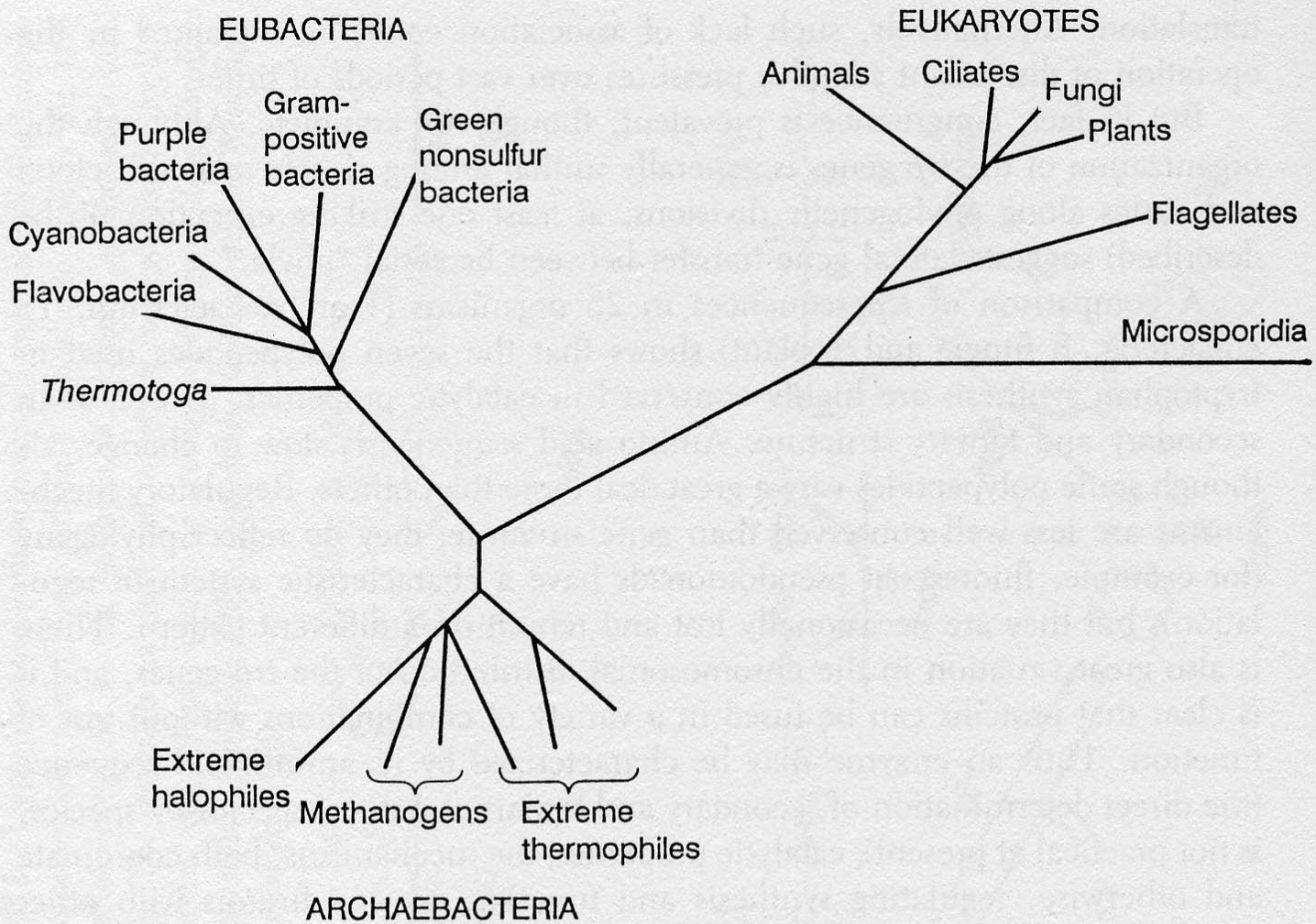
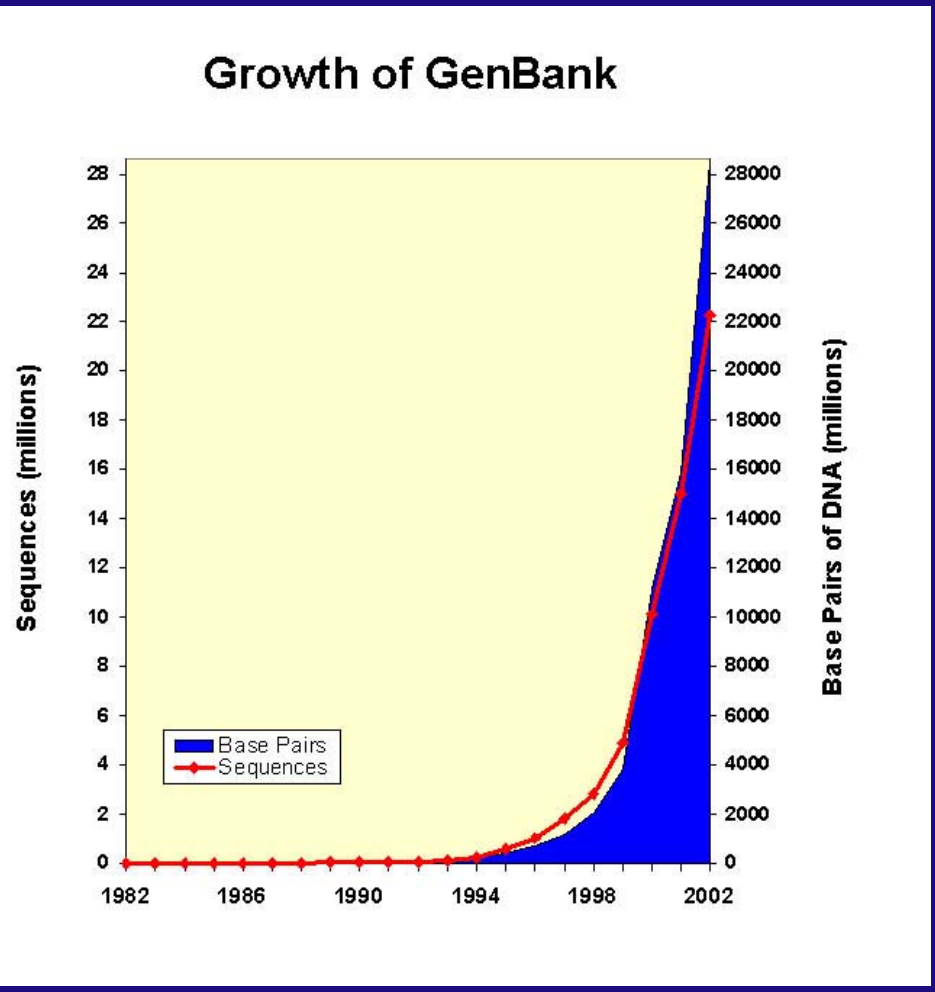


FIGURE 2. The two bacterial phylogenies, taken from the universal phylogenetic tree determined from rRNA sequence comparisons (Woese, 1987).

Evolution at the molecular level.
R.K. Selander, A.G. Clark, T.S. Whittam, eds. Sinauer Associates, 1991.



Source: NCBI

Fully sequenced genomes

- *Organisms* 751 projects

153 complete (16 A, 118 B, 19 E)

(*Eukarya* examples: mosquito (pest, malaria), sea squirt, mouse, yeast, homo sapiens, arabidopsis, fly, worm, …)

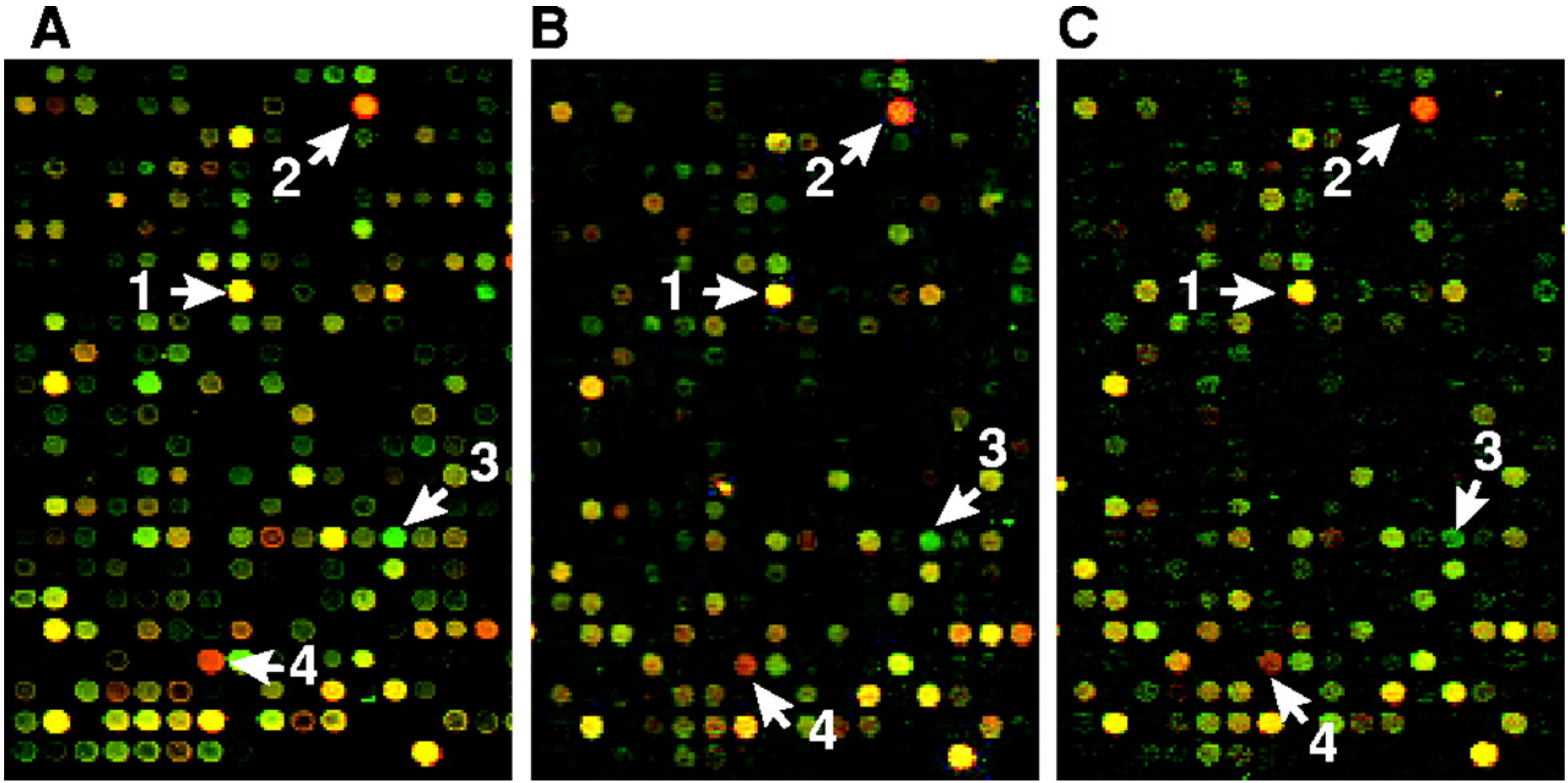
598 ongoing (23 A, 332 B, 243 E)

(*Eukarya* examples: chimpanzee, turkey, chicken, ape, corn, potato, rice, banana, tomato, cotton, coffee, soybean, pig, rat, cat, sheep, horse, kangaroo, dog, cow, bee, salmon, fugu, frog, …)

- *Other structures with genetic information*

68 phages
1328 viruses
35 viroids
472 organelles (423 mitochondria, 32 plastids, 14 plasmids, 3 nucleomorphs)

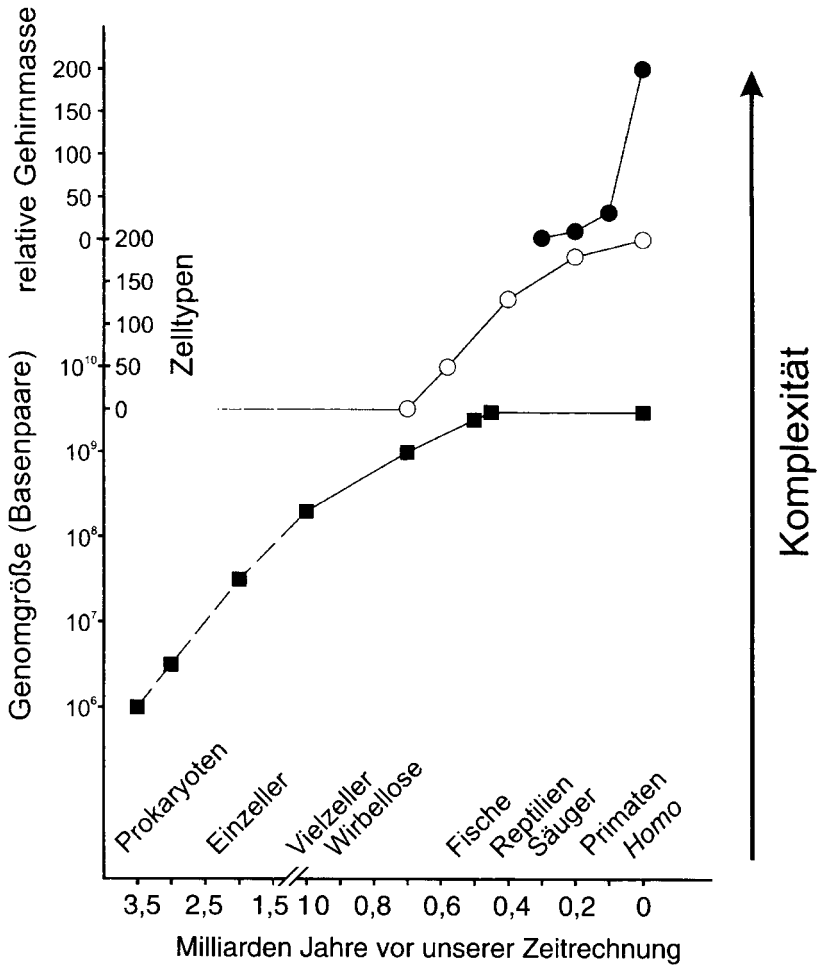
Source: Integrated Genomics, Inc.
August 12th, 2003



The same section of the microarray is shown in three independent hybridizations. Marked spots refer to: (1) protein disulfide isomerase related protein P5, (2) IL-8 precursor, (3) EST AA057170, and (4) vascular endothelial growth factor

Gene expression DNA microarray representing 8613 human genes used to study transcription in the response of human fibroblasts to serum

V.R.Iyer *et al*., *Science* **283**: 83-87, 1999

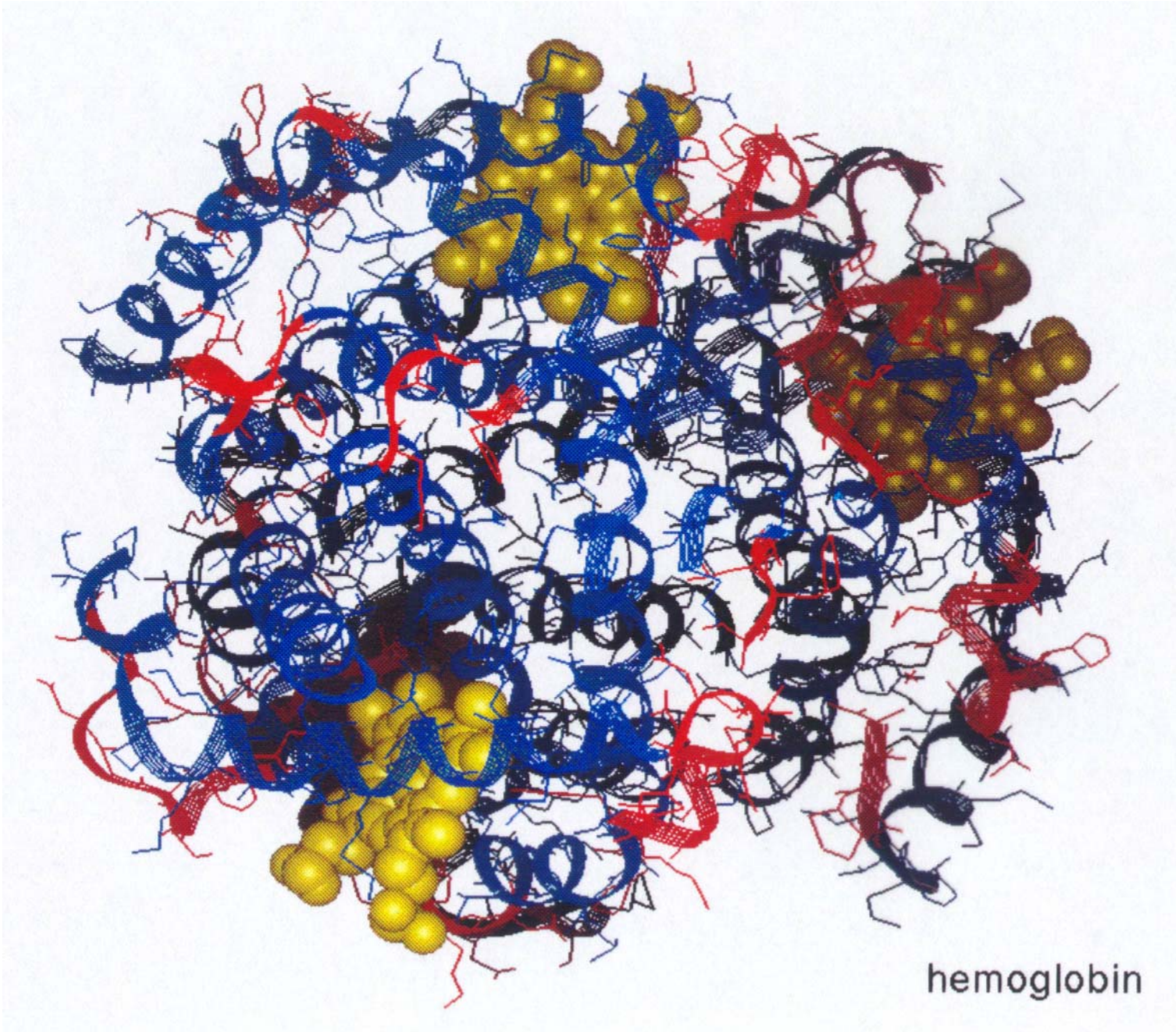


4.10 Die Zunahme der Komplexität ist ein wesentlicher Aspekt der biologischen Evolution, wobei höhere Komplexität sowohl durch Vergrößerung der Zahl von miteinander in Wechselwirkung stehenden Elementen als auch durch Differenzierung der Funktionen dieser Elemente entstehen kann. In dieser Abbildung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

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

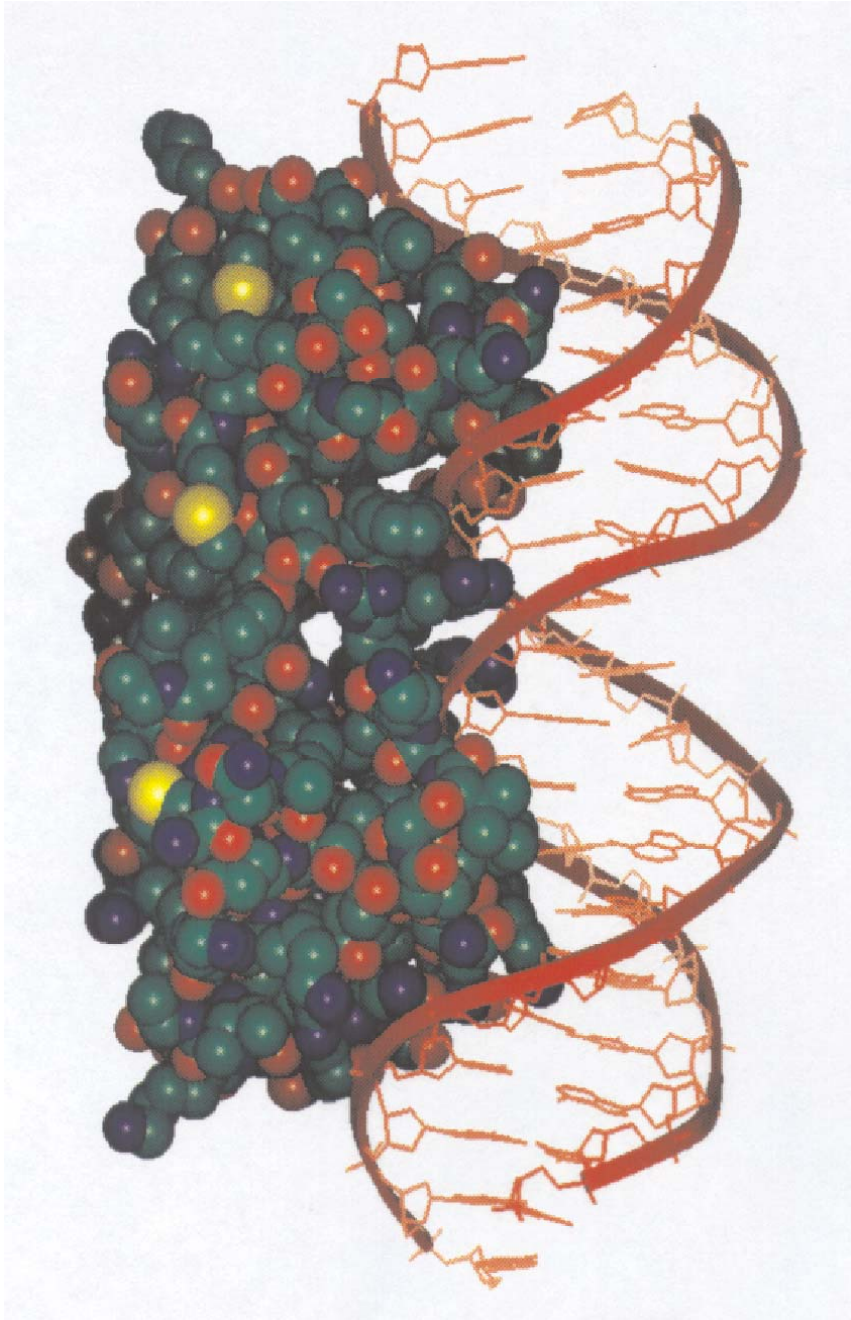
A.C.Wilson. The Molecular Basis of Evolution. Scientific American, Oct.1985, 164-173.

Max Perutz 1994 at the opening of the
Max Perutz-Library, Vienna BioCenter


hemoglobin

Developmental biology

Gene regulation networks, signal propagation, pattern formation, robustness ...



Three-dimensional structure of the complex between the regulatory protein **cro-repressor** and the binding site on λ-phage **B-DNA**

Autocatalytic chemical reactions

Direct, **A + 2 X Á 3 X** , or hidden in the reaction mechanism (Belousow-Zhabotinskii reaction).

Multiple steady states

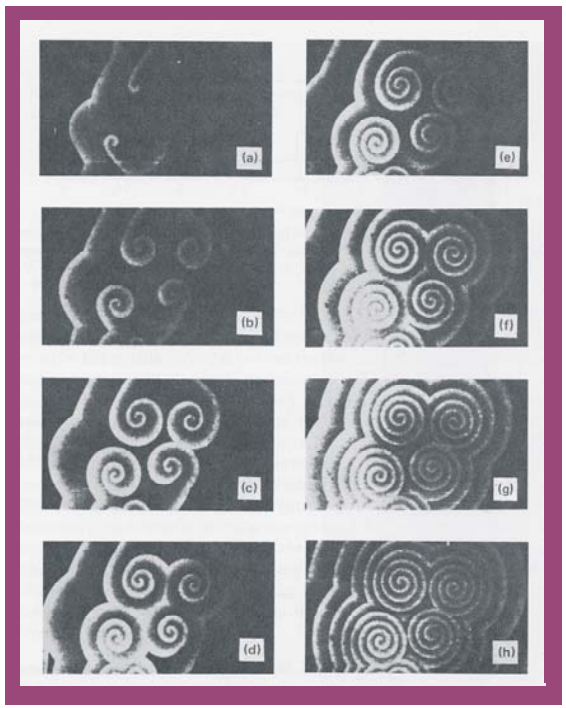
Oscillations in homogeneous solution

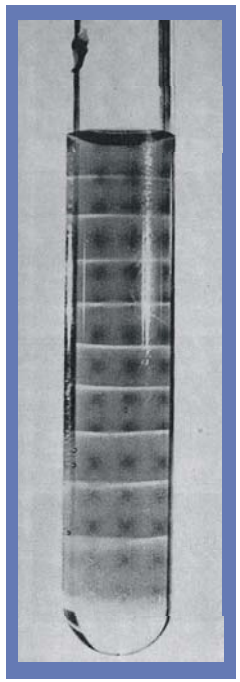
Deterministic chaos

Turing patterns

Spatiotemporal patterns (spirals)

Deterministic chaos in space and time

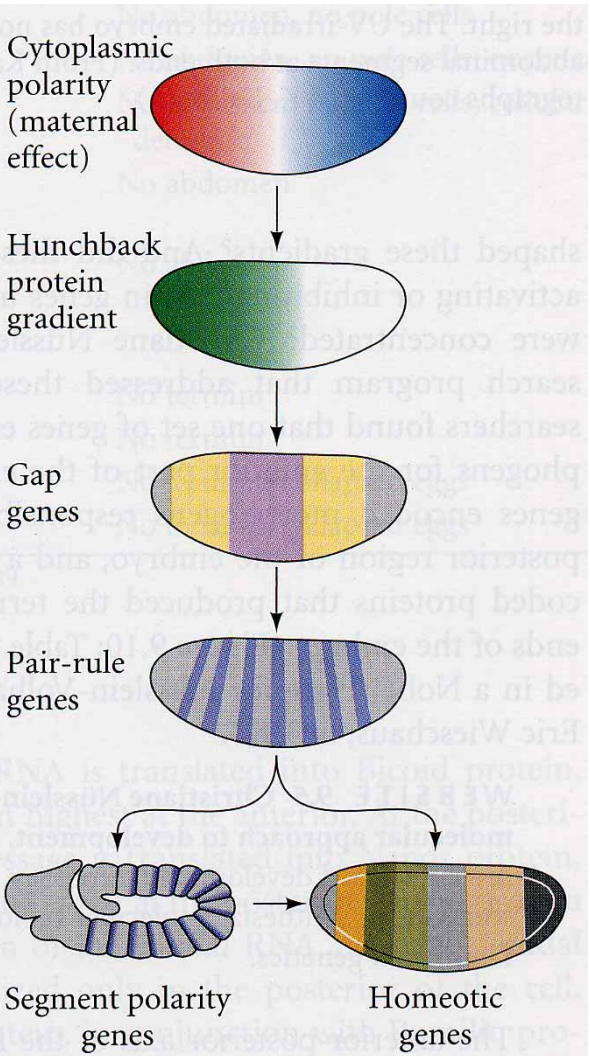


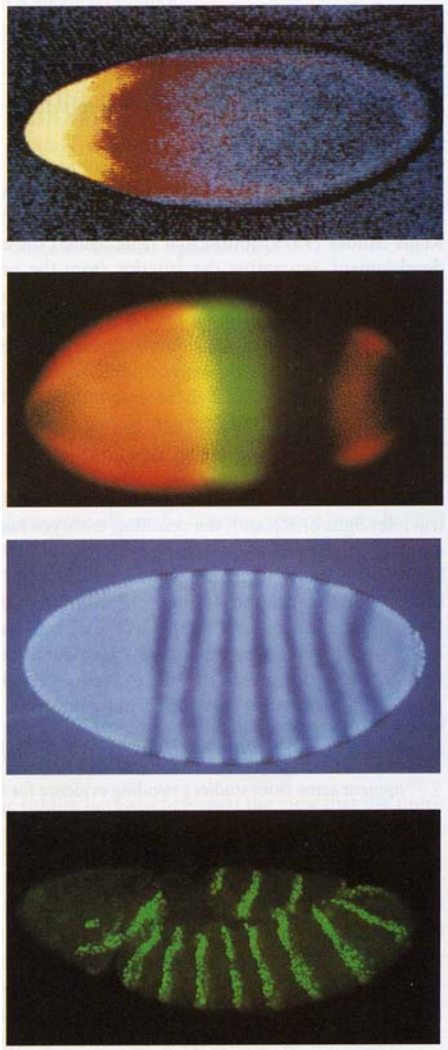


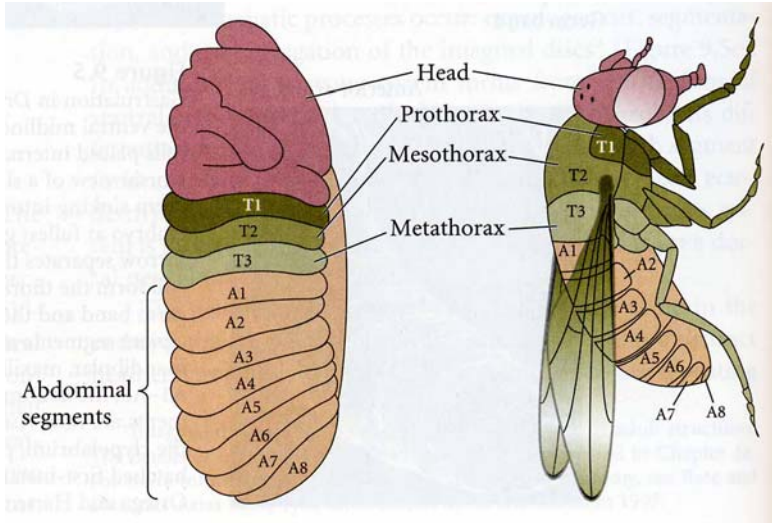
$$x_i(\vec{r},t)$$

$$\frac{\partial x_i}{\partial t} = D_i \nabla^2 x_i + F_i(\vec{r}, x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1,2,\ldots,n$$

Pattern formation in reaction-diffusion systems





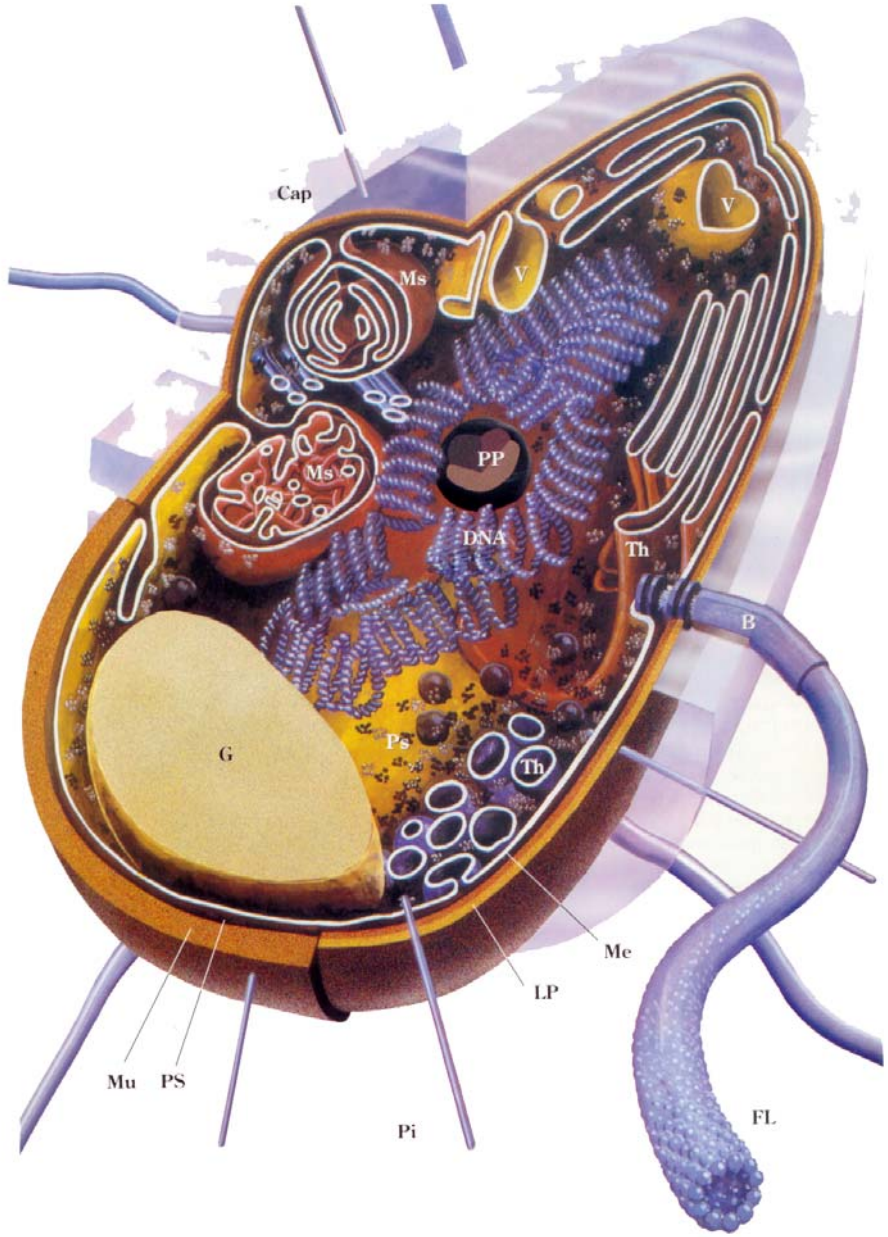


Development of the fruit fly *drosophila melanogaster*: Genetics, experiment, and imago

Cell biology

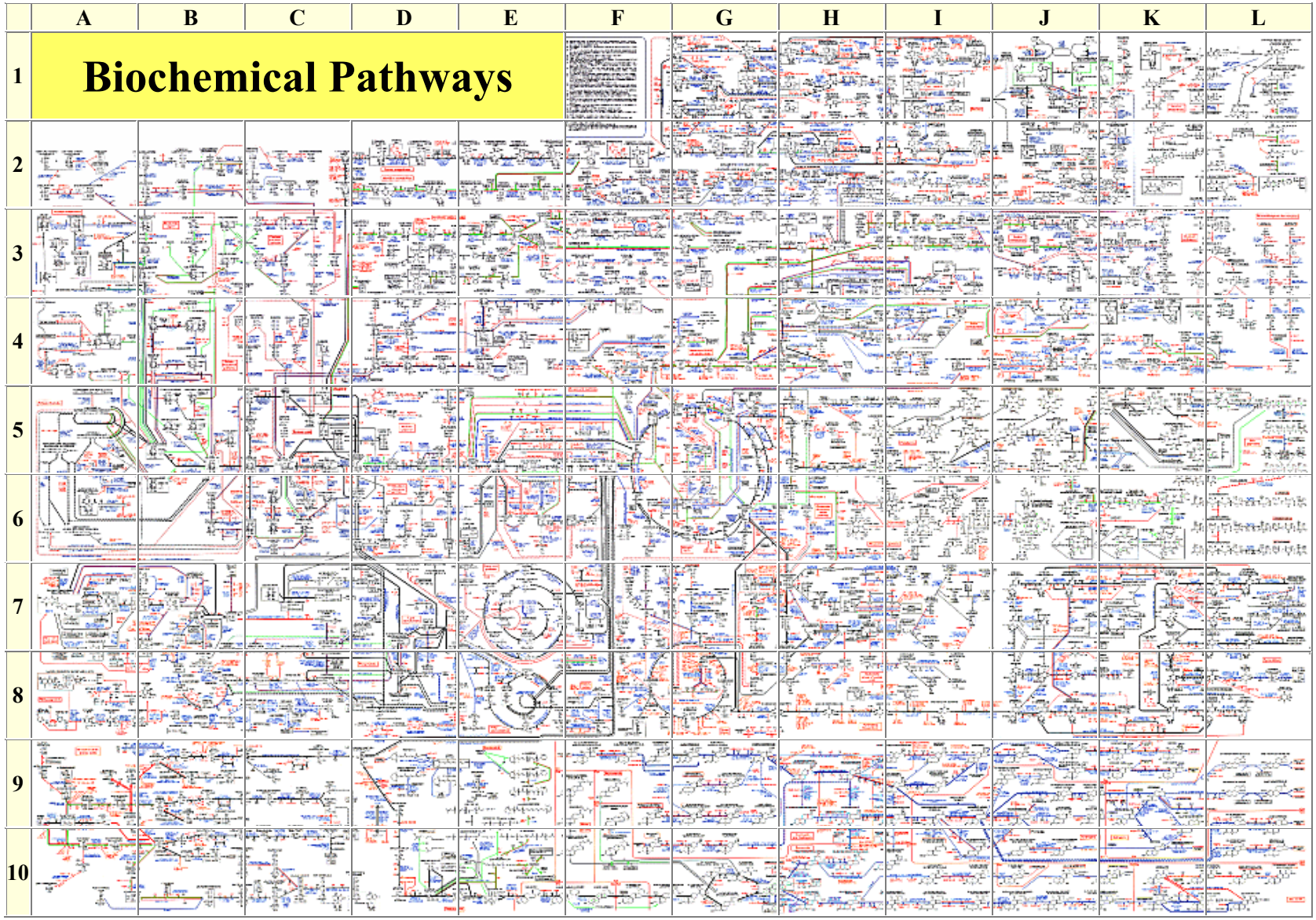
Regulation of cell cycle, metabolic networks, reaction kinetics, homeostasis, ...

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



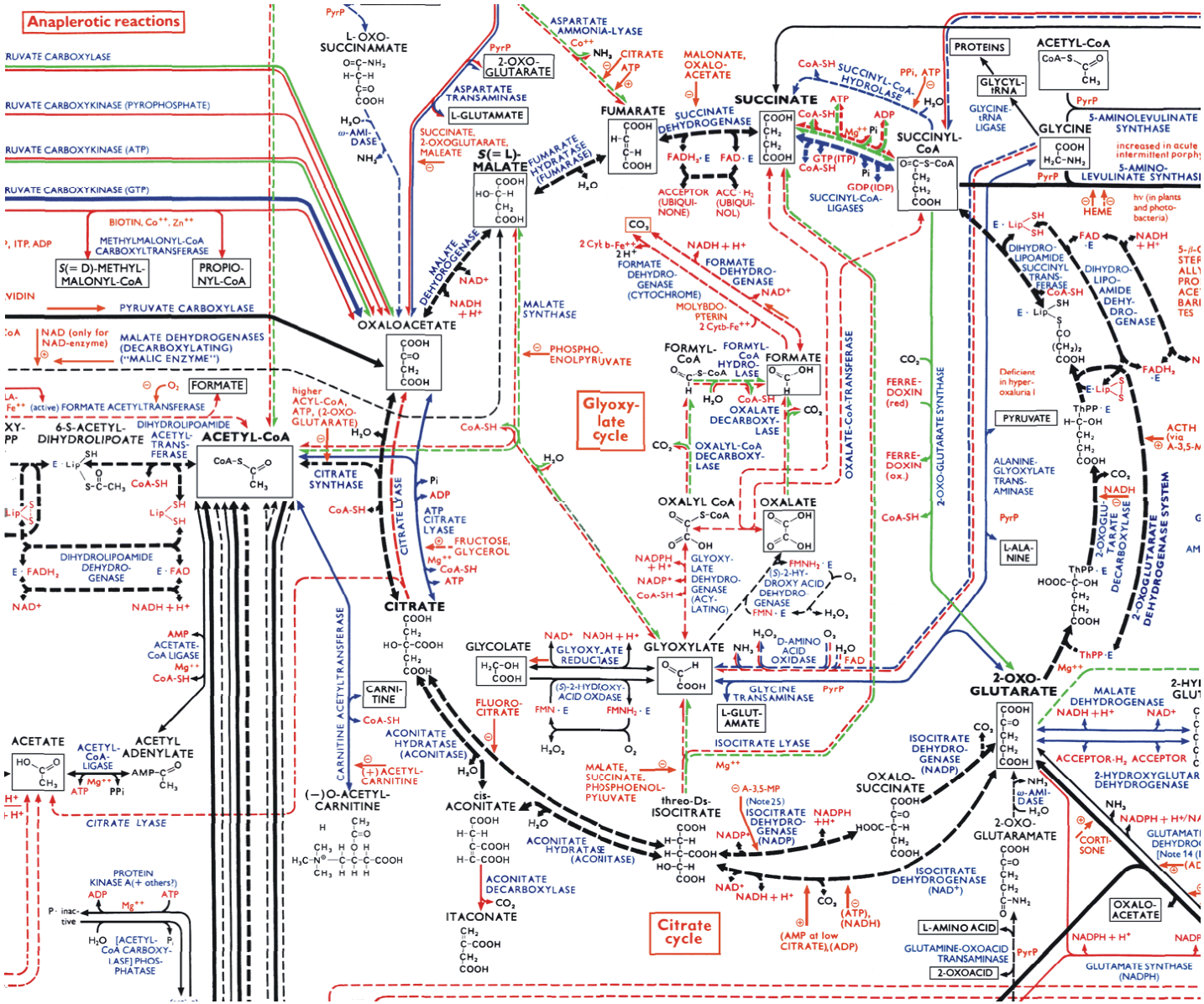
The human body:

10^{14} cells = 10^{13} eukaryotic cells + Î 9Ç10^{13} bacterial (prokaryotic) cells, and Î 200 eukaryotic cell types



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

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



Kinetic differential equations

$$\frac{d\,x_i}{d\,t} = f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

Reaction diffusion equations

$$\frac{\partial x_i}{\partial t} = D_i \nabla^2 x_i + f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

Parameter set

$$k_j(T, p, pH, I, \ldots; x_1, x_2, \ldots, x_n);\ j = 1, 2, \ldots, m$$

General conditions: T , p , pH , I , ...

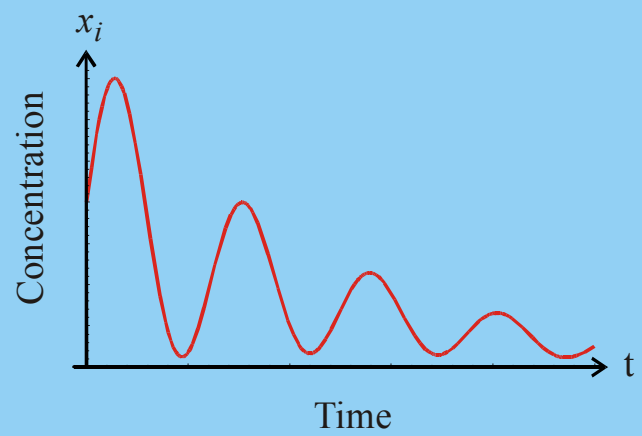
Initial conditions: $x_i(0);\ i = 1, 2, \ldots, n$

Boundary conditions: boundary ... $\hat{S}$
normal unit vector ... $\hat{u}$

Dirichlet , $x_i^{\vec{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

Neumann , $\frac{\partial x_i}{\partial u} = \hat{u} \cdot \nabla x_i^{\vec{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

Solution curves: $x_i(t);\ i = 1, 2, \ldots, n$



The forward-problem of chemical reaction kinetics

Kinetic differential equations

$$\frac{d x_i}{d t} = f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m); \; i = 1, 2, \ldots, n$$

Reaction diffusion equations

$$\frac{\partial x_i}{\partial t} = D_i \nabla^2 x_i + f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m); \; i = 1, 2, \ldots, n$$

General conditions: T , p , pH , I , ...

Initial conditions: $x_i(0); \; i = 1, 2, \ldots, n$

Boundary conditions: boundary ... $\hat{S}$
normal unit vector ... $\hat{u}$

Dirichlet , $x_i^{\hat{S}} = f(\vec{r}, t); \; i = 1, 2, \ldots, n$

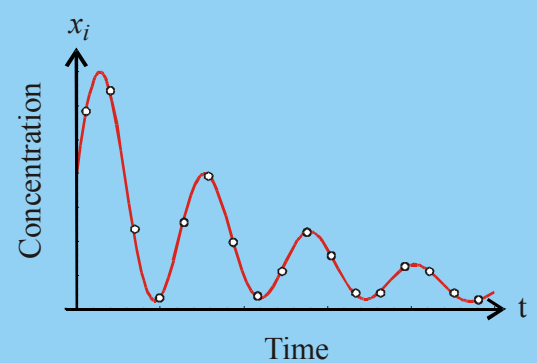
Neumann , $\frac{\partial x_i}{\partial u} = \hat{u} \cdot \nabla x_i^{\hat{S}} = f(\vec{r}, t); \; i = 1, 2, \ldots, n$

Parameter set

$$k_j(T, p, pH, I, \ldots; x_1, x_2, \ldots, x_n); \; j = 1, 2, \ldots, m$$

Data from measurements

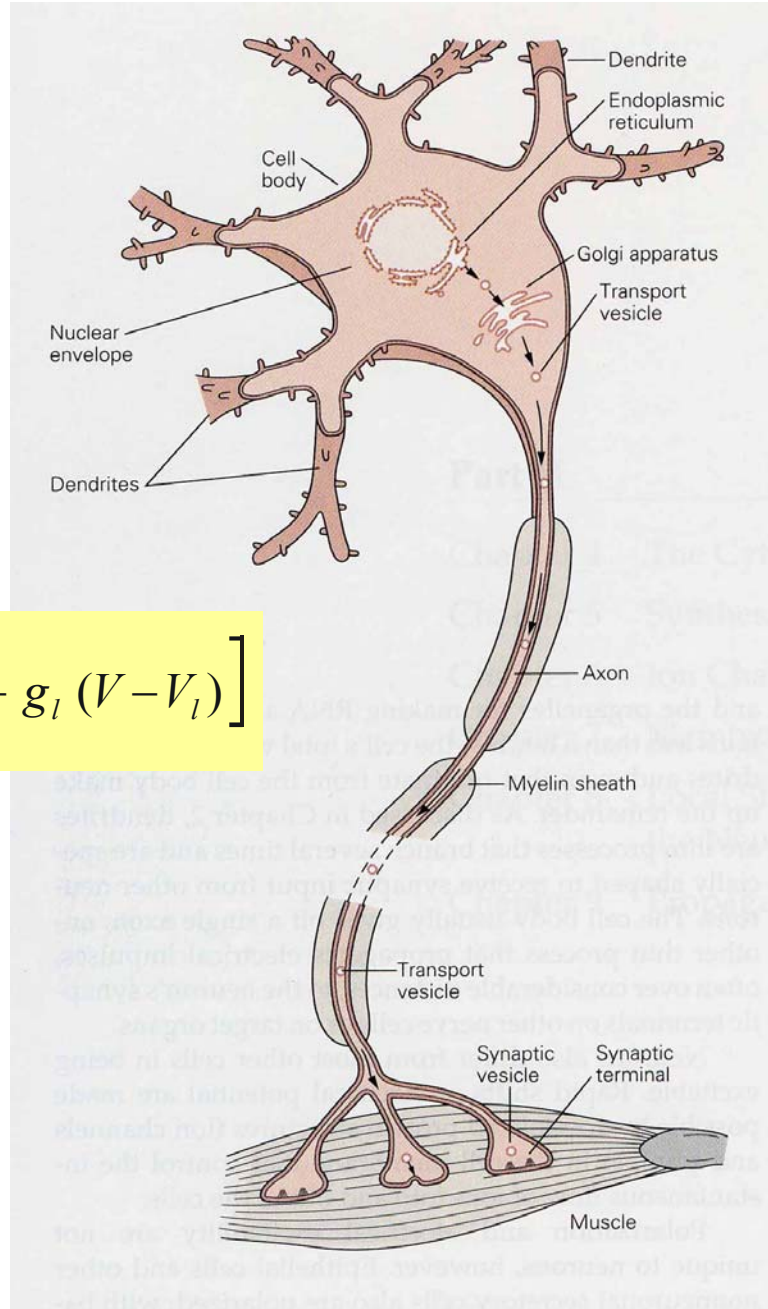
$$x_i(t_k); \; i = 1, 2, \ldots, n; \; k = 1, 2, \ldots, N$$



The inverse-problem of chemical reaction kinetics

Neurobiology

Neural networks, collective properties, nonlinear dynamics, signalling, ...



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

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

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

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

Hogdkin-Huxley OD equations

A single neuron signaling to a muscle fiber

$$\frac{1}{R}\frac{\partial^2 V}{\partial x^2} = C\frac{\partial V}{\partial t} + \left[g_{Na}\, m^3\, h (V - V_{Na}) + g_K\, n^4\, (V - V_K) + g_l\, (V - V_l) \right] 2\pi\, r\, L$$

$$\frac{\partial m}{\partial t} = \alpha_m (1 - m) - \beta_m\, m$$

$$\frac{\partial h}{\partial t} = \alpha_h (1 - h) - \beta_h\, h$$

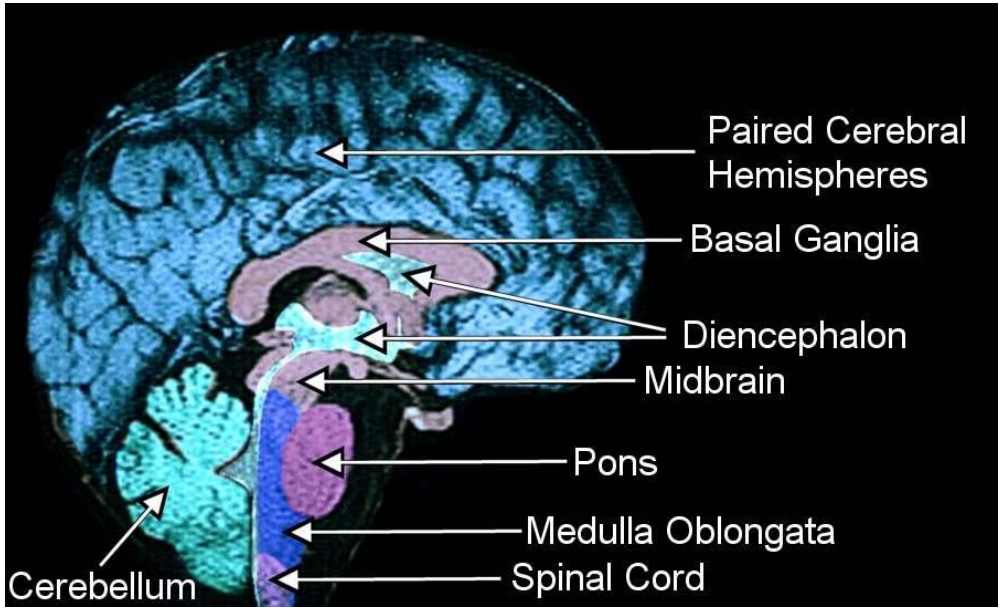
$$\frac{\partial n}{\partial t} = \alpha_n (1 - n) - \beta_n\, n$$

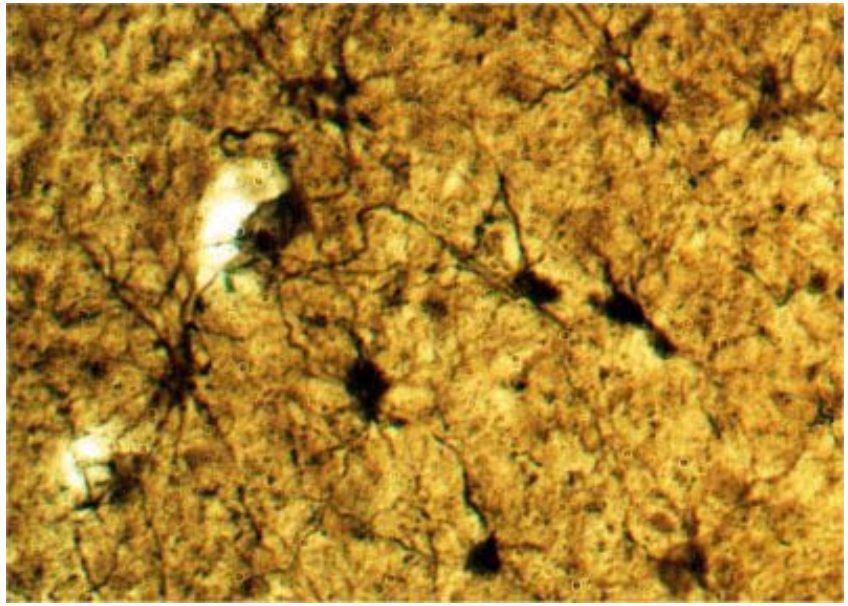
Hodgkin-Huxley PDEquations

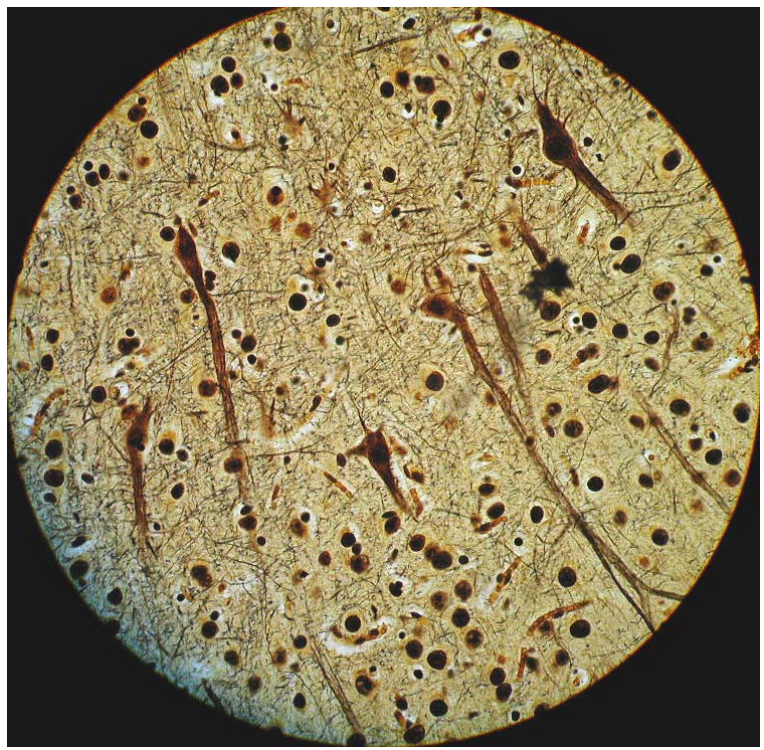
Travelling pulse solution: $V(x,t) = W(\mathrm{x})$ with

$$\mathrm{x} = x - \mathrm{q}\, t$$

Hodgkin-Huxley equations describing pulse propagation along nerve fibers







The human brain

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

Evolutionary biology

Optimization through variation and selection, relation between genotype, phenotype, and function, ...

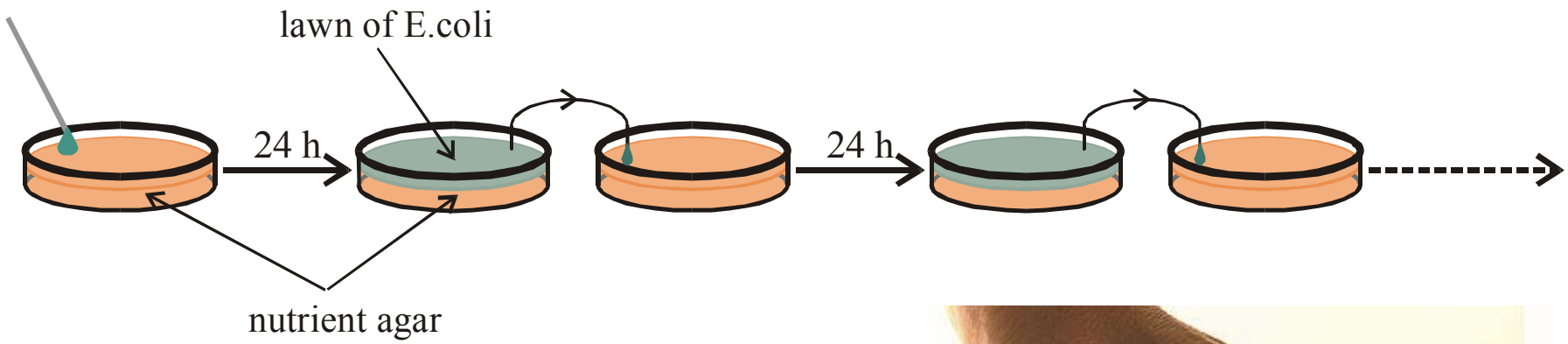
	Generation time	10 000 generations	10^6 generations	10^7 generations
RNA molecules	10 sec 1 min	27.8 h = 1.16 d 6.94 d	115.7 d 1.90 a	3.17 a 19.01 a
Bacteria	20 min 10 h	138.9 d 11.40 a	38.03 a 1 140 a	380 a 11 408 a
Higher multicelluar organisms	10 d 20 a	274 a 20 000 a	27 380 a 2×10^7 a	273 800 a 2×10^8 a

Time scales of evolutionary change

Bacterial Evolution

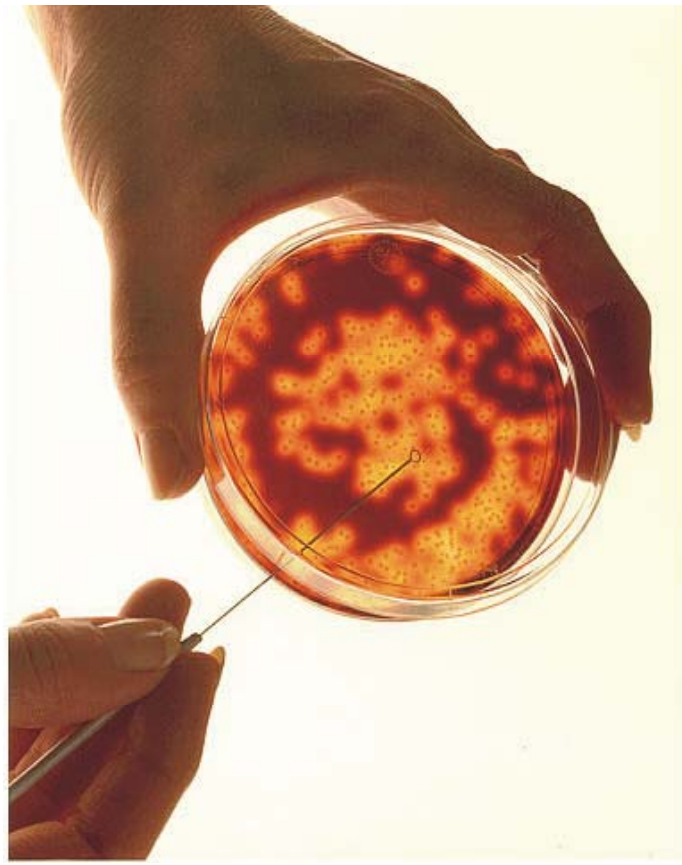
S. F. Elena, V. S. Cooper, R. E. Lenski. ***Punctuated evolution caused by selection of rare beneficial mutants***. Science **272** (1996), 1802-1804

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



Serial transfer of Escherichia coli cultures in Petri dishes

1 day Î 6.67 generations
1 month Î 200 generations
1 year Î 2400 generations



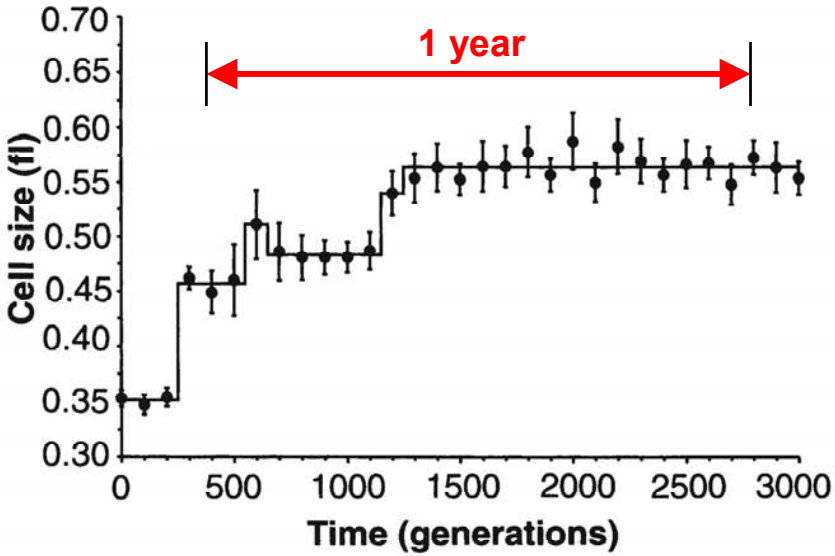


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

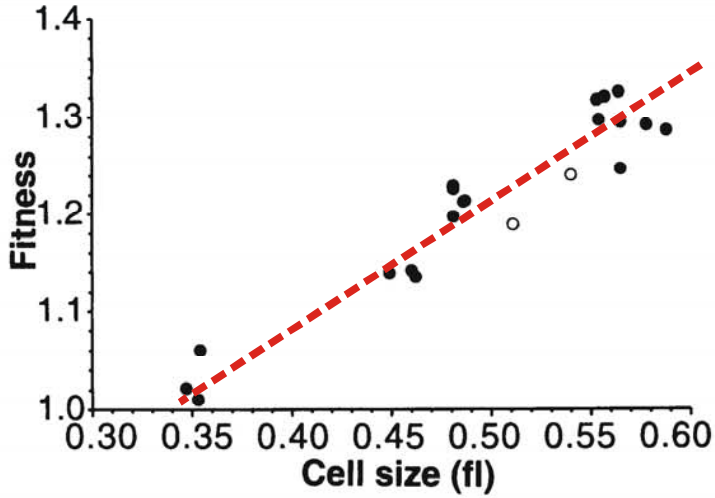
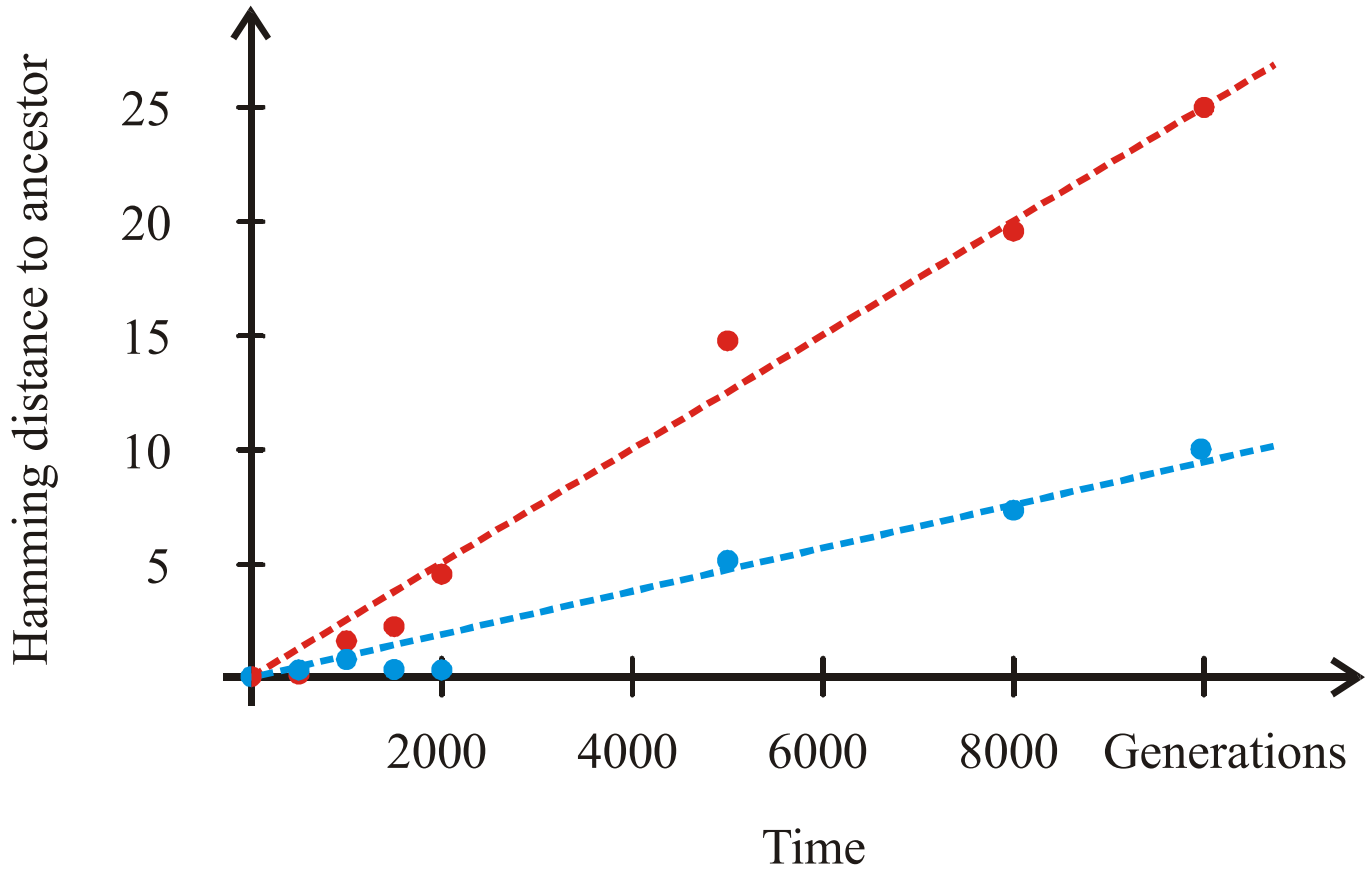


Fig. 2. Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (*6*, *7*). The open symbols indicate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

Epochal evolution of bacteria in serial transfer experiments under constant conditions

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



Variation of genotypes in a bacterial serial transfer experiment

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

In evolution **variation** occurs on **genotypes** but **selection** operates on the **phenotype**.

Mappings from genotypes into phenotypes are highly complex objects. The only computationally accessible case is in the evolution of RNA molecules.

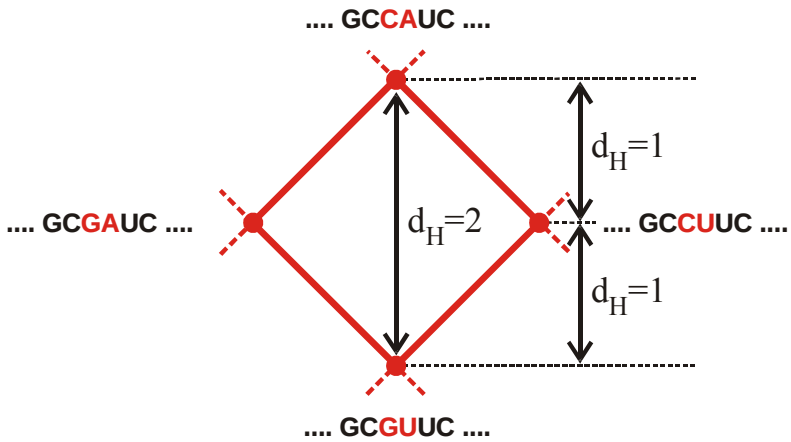
The mapping from RNA sequences into secondary structures and function,

sequence ñ **structure** ñ **function**,

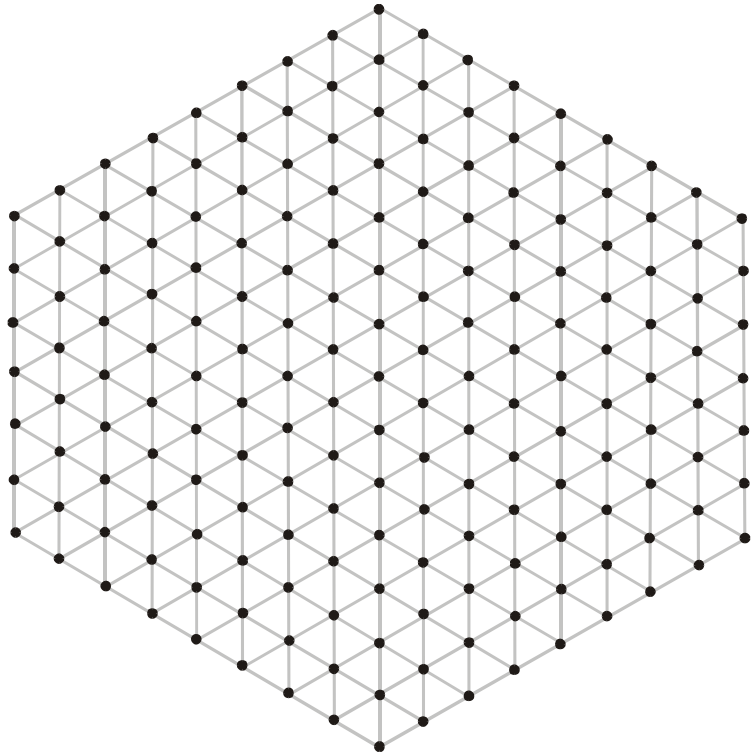
is used as a model for the complex relations between genotypes and phenotypes. Fertile progeny measured in terms of **fitness** in population biology is determined quantitatively by **replication rate constants** of RNA molecules.

Population biology	Molecular genetics	Evolution of RNA molecules
Genotype	**Genome**	**RNA sequence**
Phenotype	**Organism**	**RNA structure and function**
Fitness	**Reproductive success**	**Replication rate constant**

The RNA model

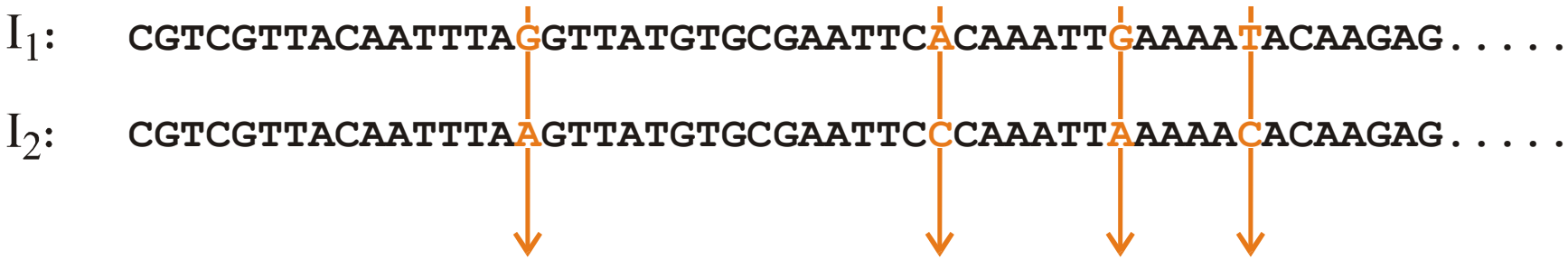


City-block distance in sequence space



2D Sketch of sequence space

Single point mutations as moves in sequence space



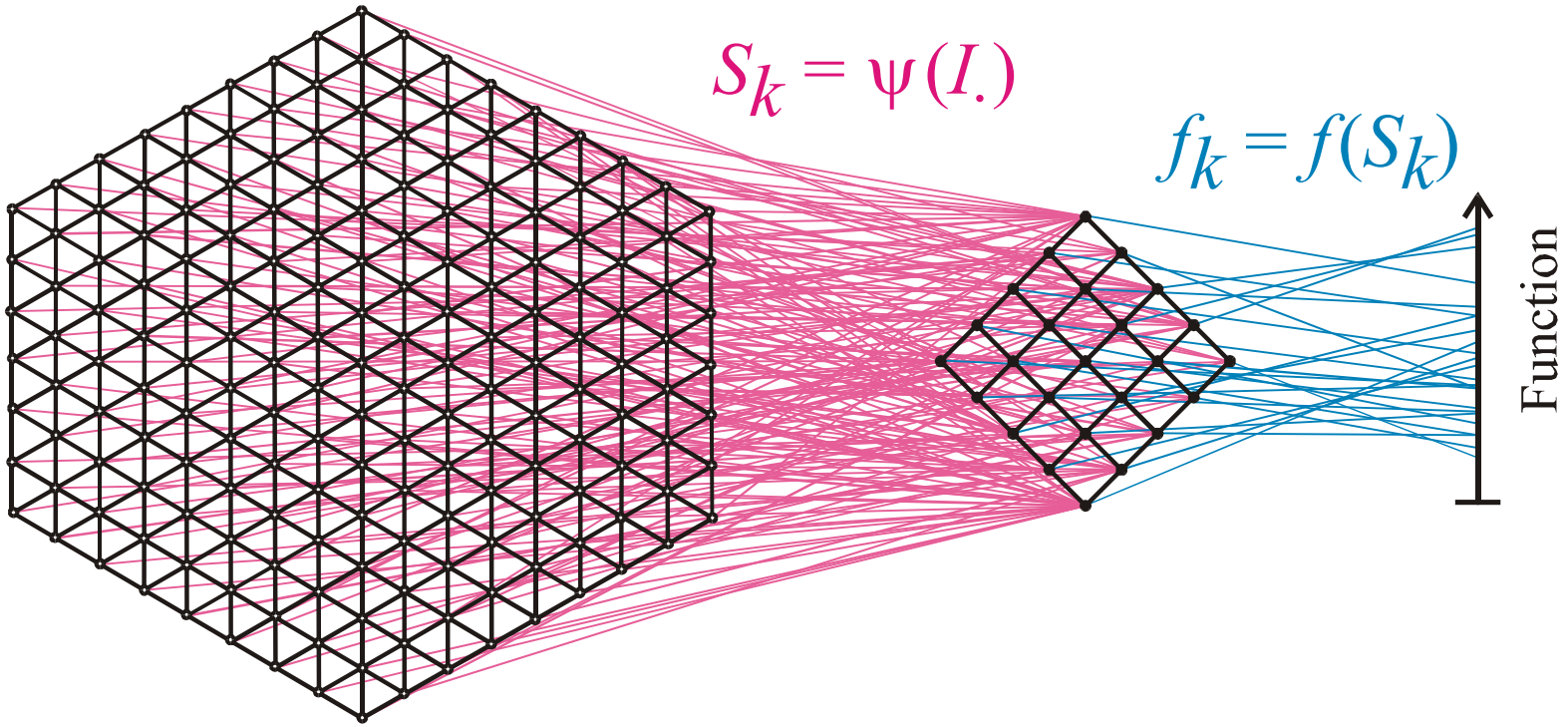
Hamming distance $d_H(I_1,I_2) = 4$

(i) $d_H(I_1,I_1) = 0$

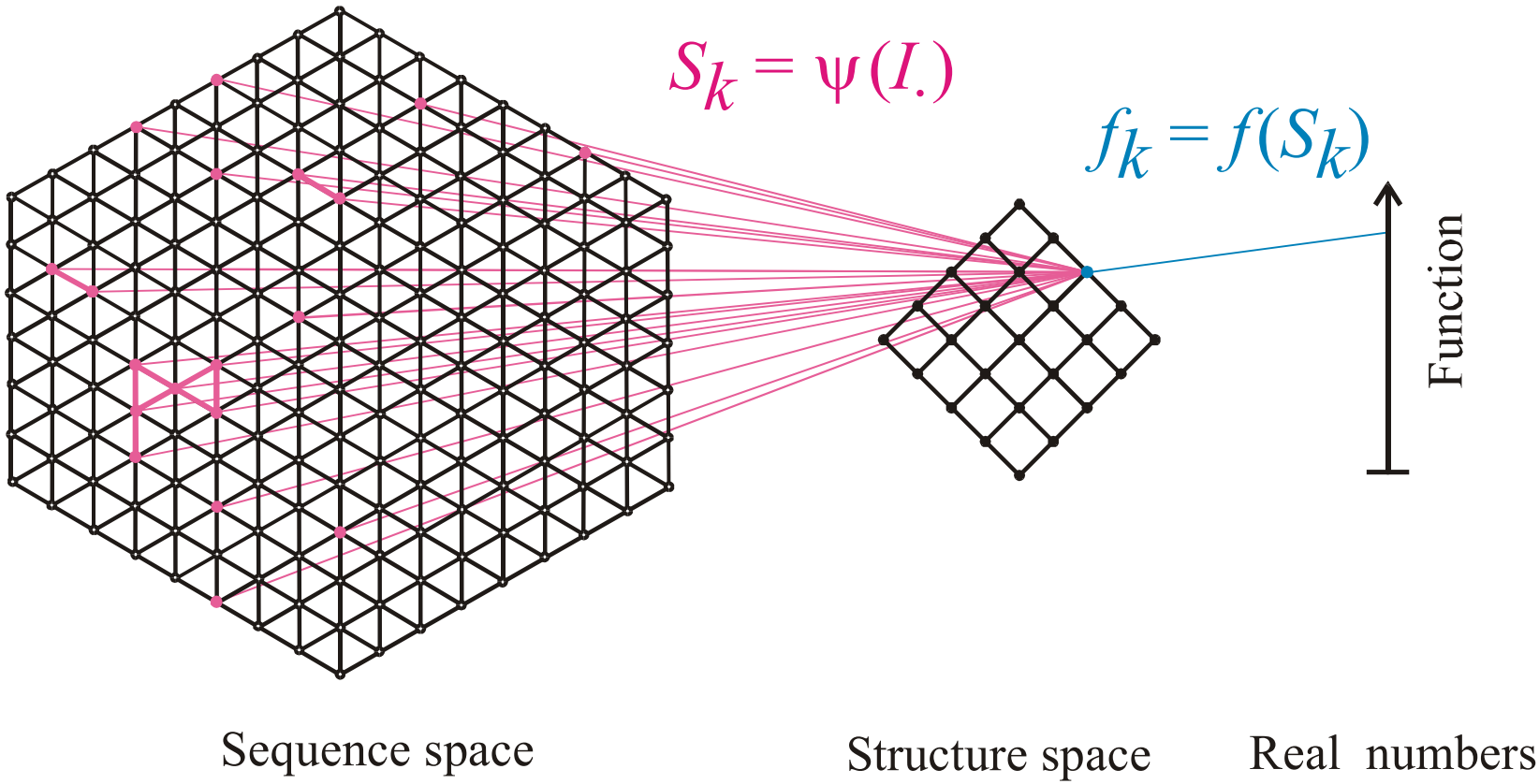
(ii) $d_H(I_1,I_2) = d_H(I_2,I_1)$

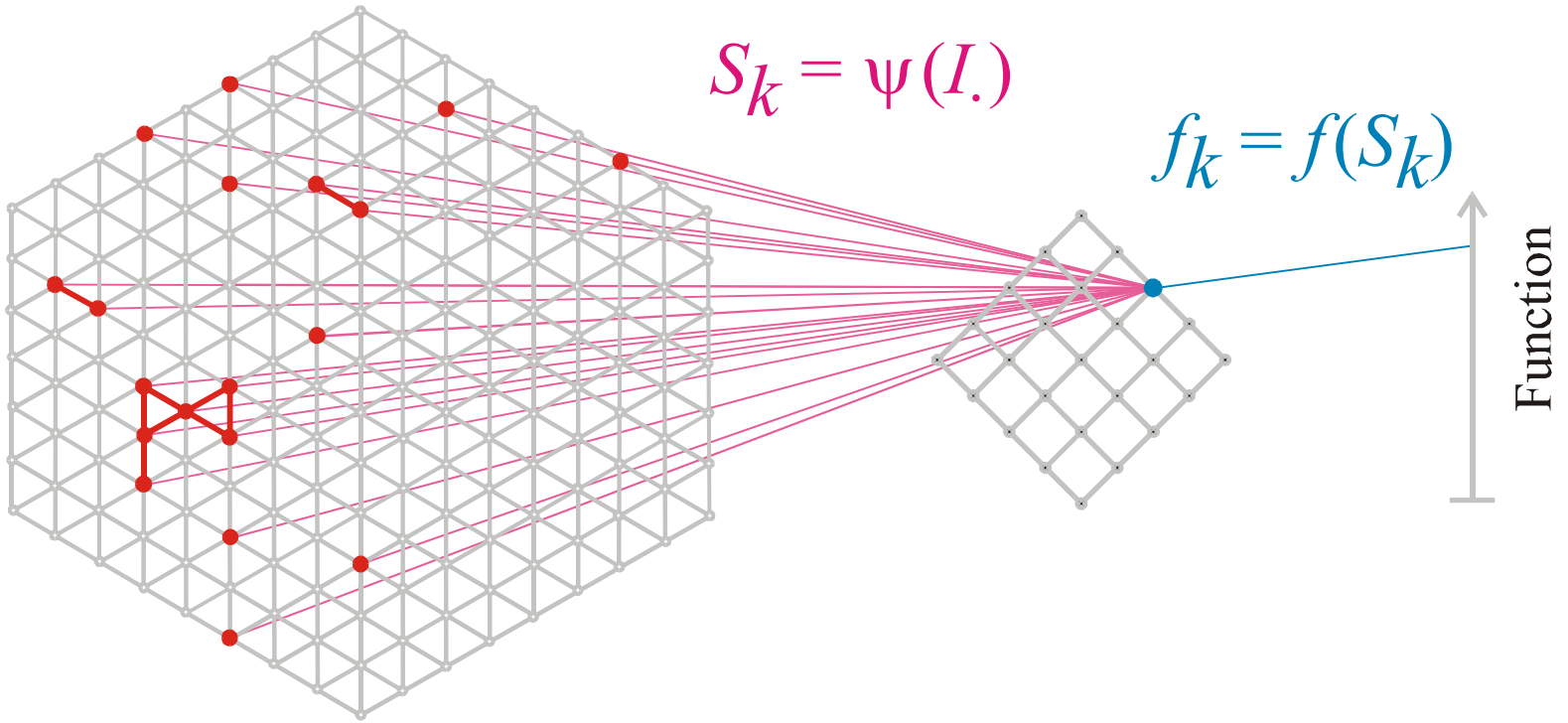
(iii) $d_H(I_1,I_3) \P d_H(I_1,I_2) + d_H(I_2,I_3)$

The Hamming distance between sequences induces a metric in sequence space



Mapping from sequence space into structure space and into function


$S_k = \psi(I_.)$
$f_k = f(S_k)$
Function
Sequence space
Structure space
Real numbers



The pre-image of the structure S_k in sequence space is the **neutral network G_k**

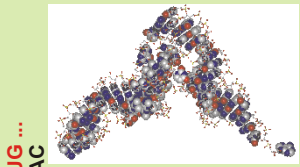
RNA

RNA *as transmitter of genetic information*

DNA
↓ transcription
...AGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUC...
messenger-RNA
↓ translation
protein

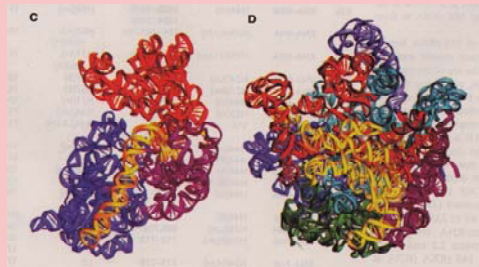
RNA as **working copy** of genetic information

RNA *as adapter molecule*

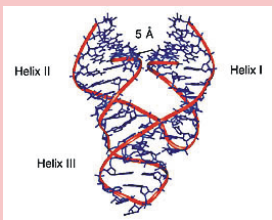


genetic code

RNA *is the catalytic subunit in supramolecular complexes*



RNA *as catalyst*



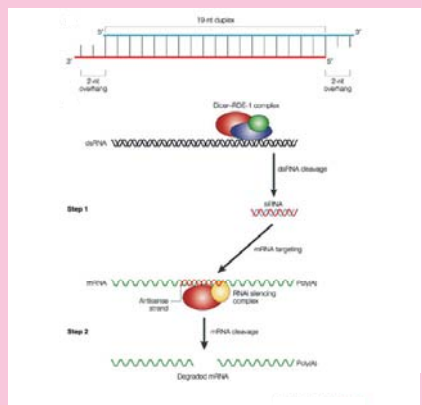
ribozyme

RNA *is modified by epigenetic control*

RNA editing

Alternative splicing of messenger **RNA**

RNA *as regulator of gene expression*



gene silencing by small interfering RNAs

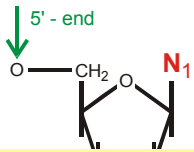
RNA *as carrier of genetic information*

RNA viruses and retroviruses

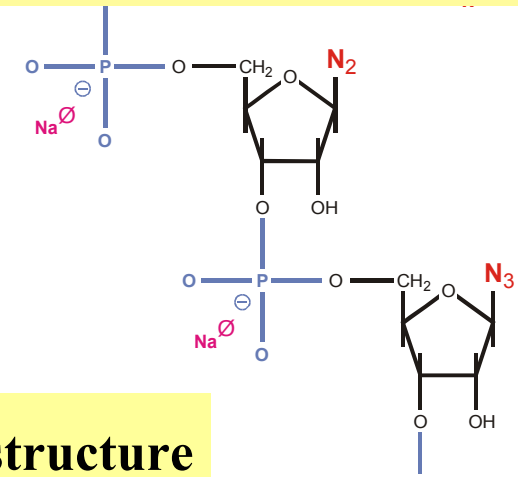
RNA as information carrier in evolution *in vitro* and evolutionary biotechnology

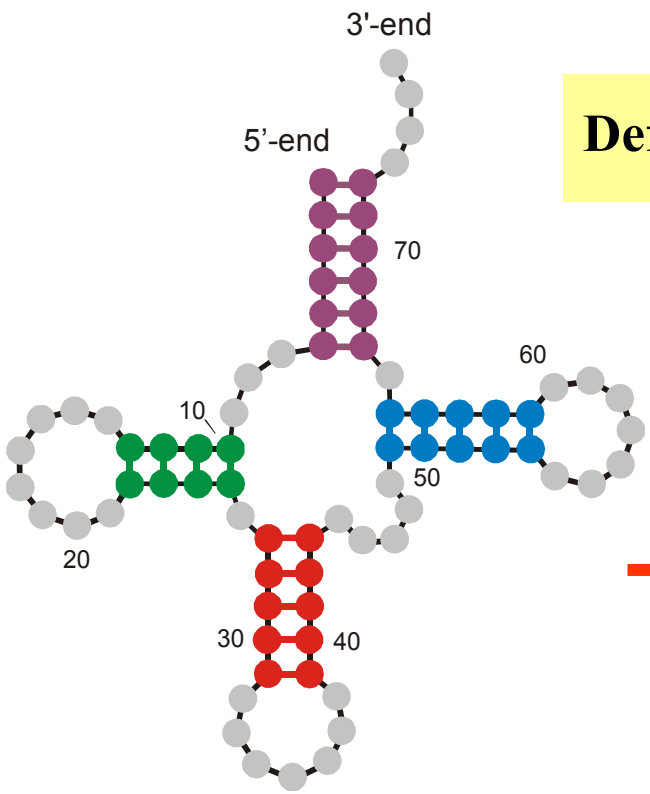
The RNA *world as a precursor of the current* DNA + protein *biology*

Functions of RNA molecules

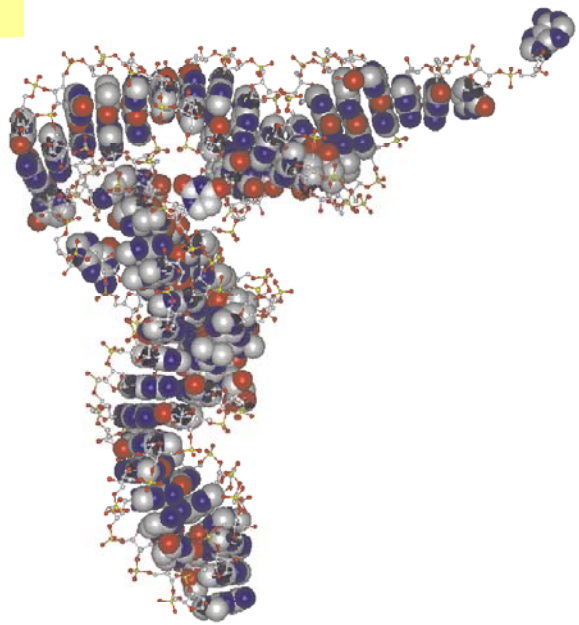


5'-end GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA 3'-end





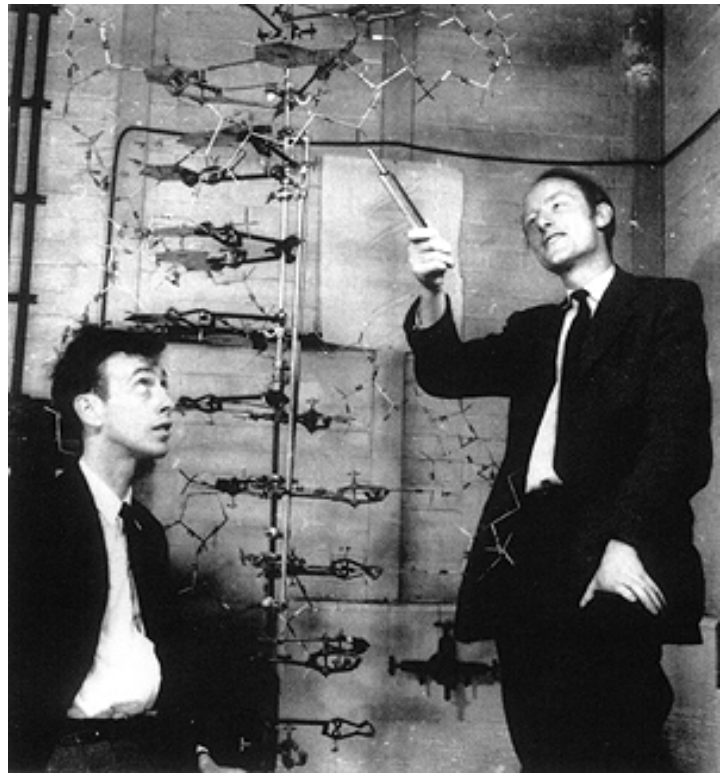
Definition of RNA structure



Stacking of free nucleobases or other planar heterocyclic compounds (N6,N9-dimethyl-adenine)

The stacking interaction as driving force of structure formation in nucleic acids

Stacking of nucleic acid single strands (poly-A)



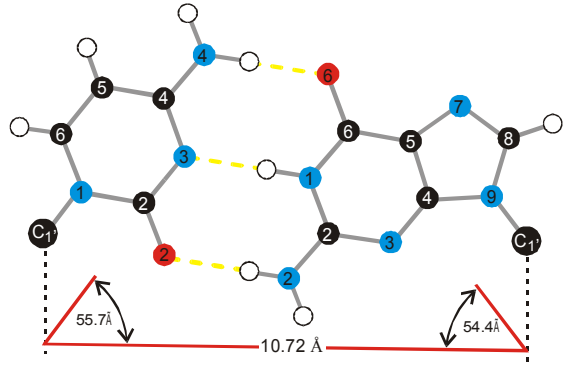
James D. Watson and Francis H.C. Crick

Nobel prize 1962

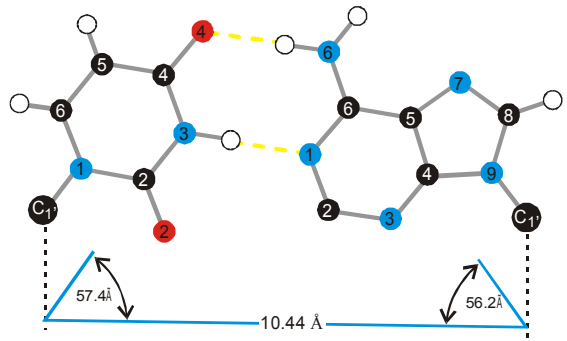
1953 – 2003 fifty years double helix

Stacking of base pairs in nucleic acid double helices (B-DNA)

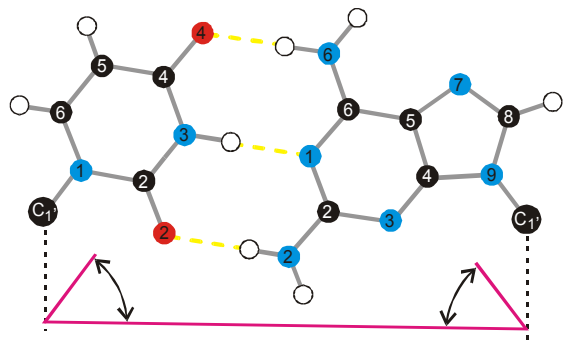
C Í G



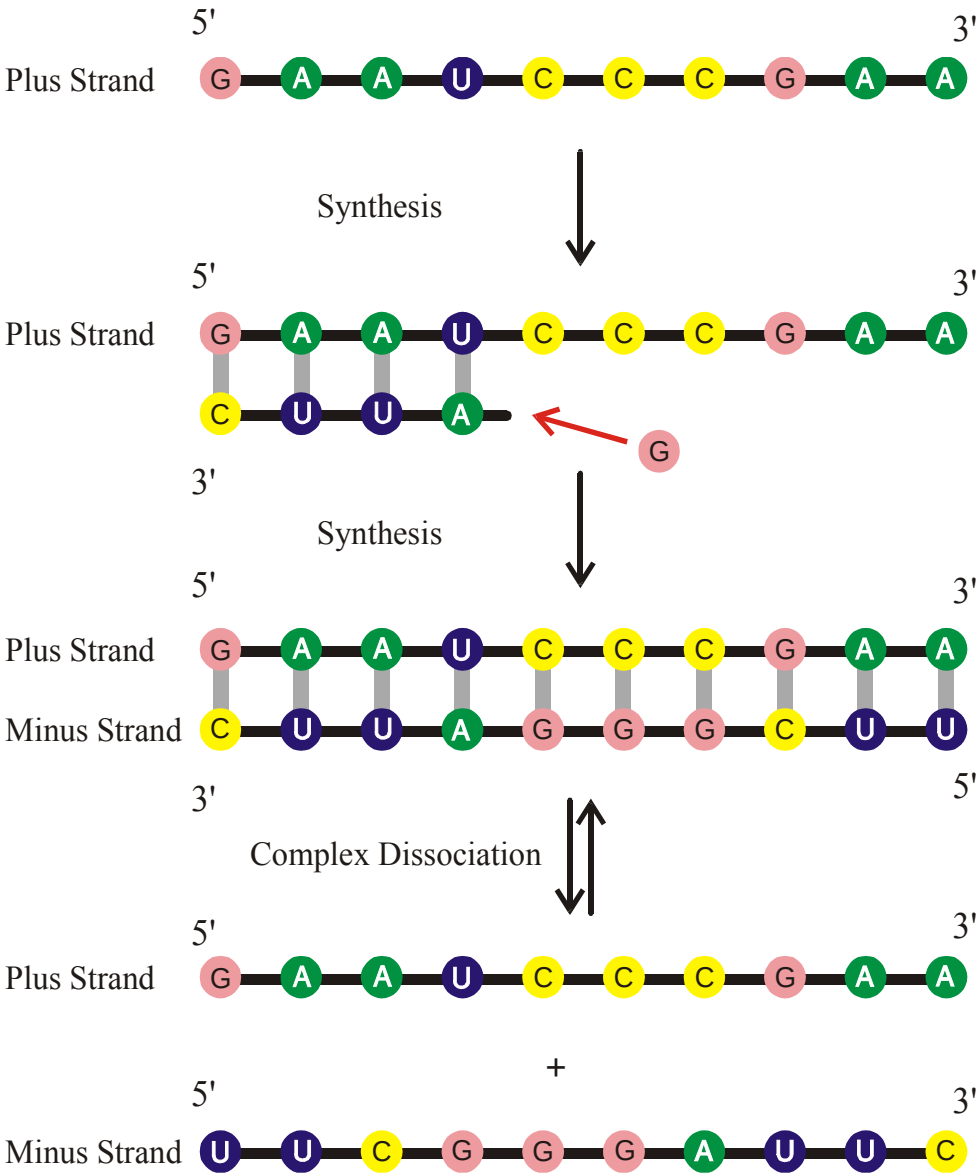
U = A



U Í D



Three Watson-Crick type base pairs



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

G Í C and A=U