The mathematics of Darwin‘s theory of evolution

1859 and 150 years later

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Österreich

und

The Santa Fe Institute, Santa Fe, New Mexico, USA

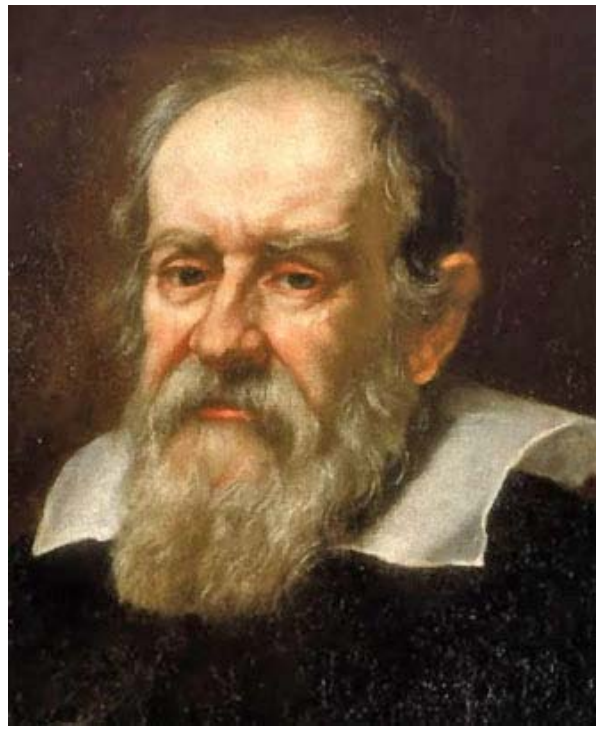
„The Mathematics of Darwin Legacy“

Centro Internacional de Mathemática, Lisbon, 23.-24.11.2009

Web-Page for further information:

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

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



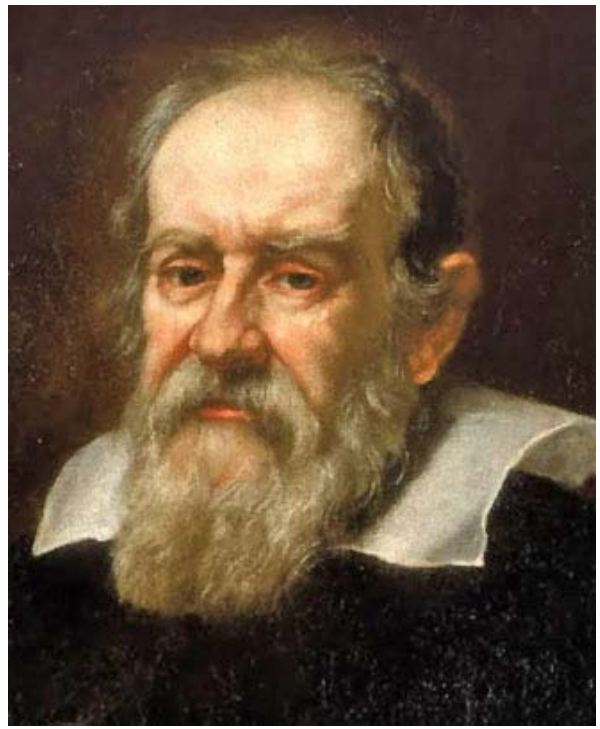
Galileo Galilei, 1564 - 1642

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

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

„Philosophy [science] is written in this grand book, the universe ... . It is written in the language of mathematics, and ist characters are triangles, circles and other geometric figures; …. „

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



Galileo Galilei, 1564 - 1642

ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

If Charles Darwin would have written the „Origin"
in mathematical language, how would he have done it?

What did we learn about evolution from *in vitro* experiments?

Quantitative systems biology – A challenge for biologists,
chemists, physicists, and mathematicians !

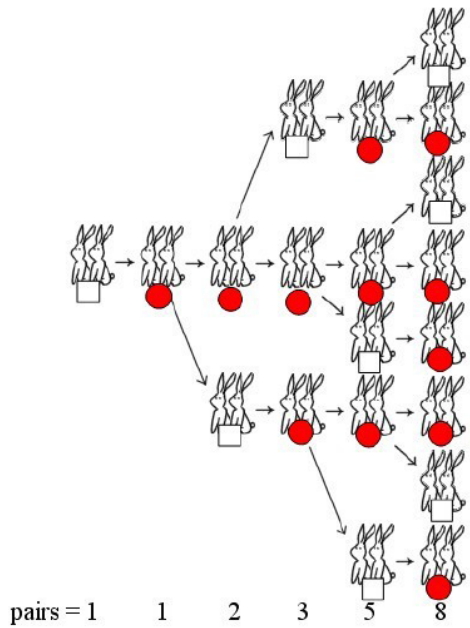
If Charles Darwin would have written the „Origin"
in mathematical language, how would he have done it?

What did we learn about evolution from *in vitro* experiments?

Quantitative systems biology – A challenge for biologists, chemists, physicists, and mathematicians !

$$F_{n+1} = F_n + F_{n-1};\ F_0 = 0,\ F_1 = 1$$

Thomas Robert Malthus
1766 – 1834

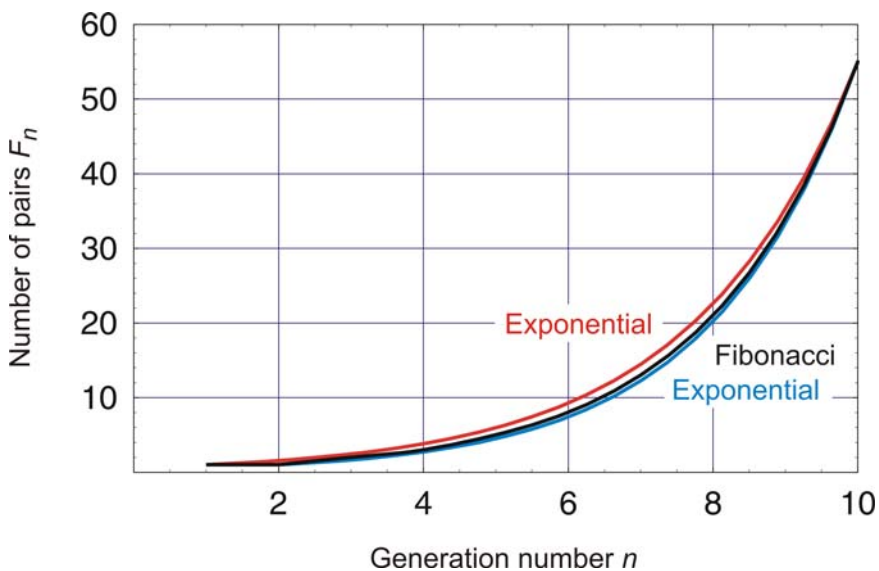


Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

1, 2 , 4 , 8 ,16 , 32 , 64, 128 , ...

geometric progression

exponential growth



The history of
exponential growth

INTRODUCTIO

IN ANALYSIN

INFINITORUM.

AUCTORE

LEONHARDO EULERO,

Professore Regio BEROLINENSI, *& Academiæ Imperialis Scientiarum* PETROPOLITANÆ *Socio.*

TOMUS PRIMUS.

LAUSANNÆ,

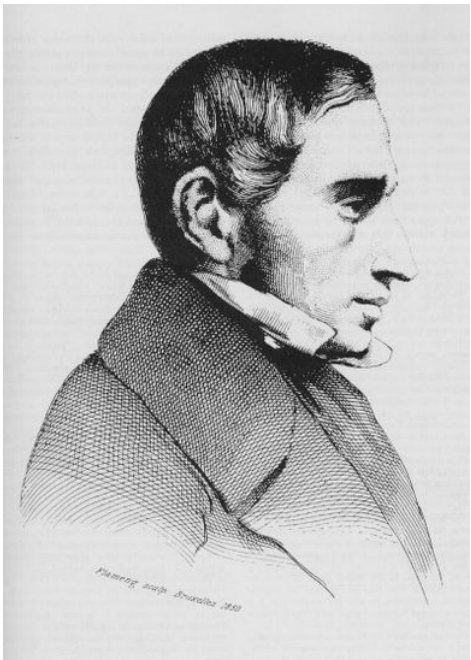
Apud MARCUM-MICHAELEM BOUSQUET & Socios.

MDCCXLVIII.

Leonhard Euler, 1717 - 1783

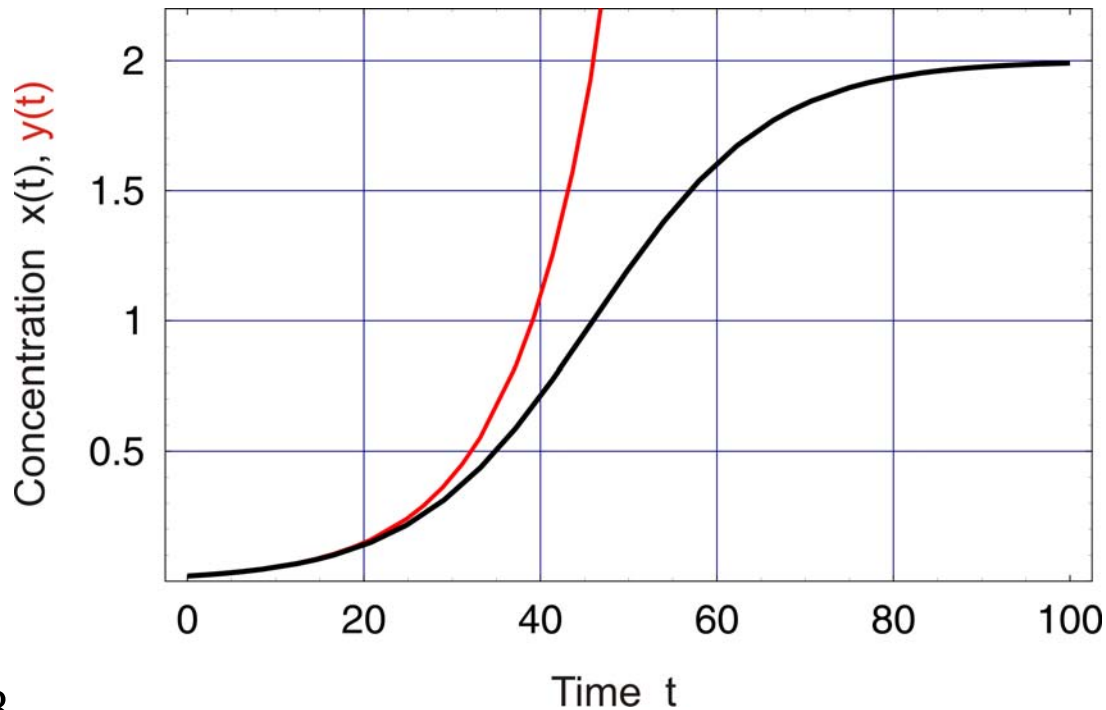
$$\exp(x) \equiv \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n$$

Exponential function and exponential growth



Pierre-François Verhulst,
1804-1849

$$\frac{dx}{dt} = r\,x\left(1-\frac{x}{C}\right), \quad x(t) = \frac{x(0)\,C}{x(0)+\left(C-x(0)\right)e^{-r\,t}}$$



The logistic equation, 1828

$$\frac{\mathrm{d}x}{\mathrm{dt}} = r\,x\left(1-\frac{x}{C}\right) \quad \Rightarrow \quad \frac{\mathrm{d}x}{\mathrm{dt}} = r\,x - \frac{x}{C}\,r\,x$$

$$r\,x \equiv \Phi(\mathrm{t}),\, C = 1: \quad \frac{\mathrm{d}x}{\mathrm{dt}} = x\left(r - \Phi\right)$$

$$\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n: \quad \left[\mathrm{X}_i\right] = x_i; \quad \sum_{i=1}^{n} x_i = C = 1$$

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} = x_j\left(f_j - \sum_{i=1}^{n} f_i\,x_i\right) = x_j\left(f_j - \Phi\right); \quad \Phi = \sum_{i=1}^{n} f_i\,x_i$$

Darwin

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

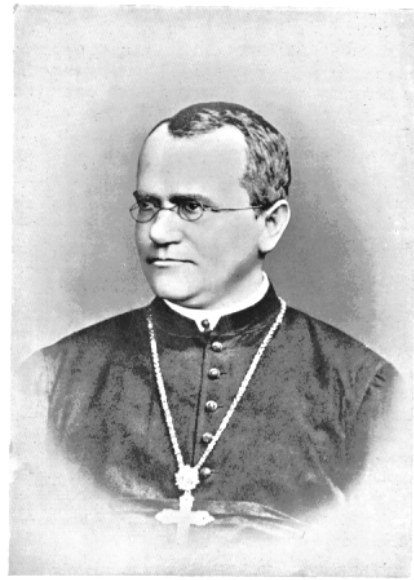
Generalization of the logistic equation to *n* variables yields selection

Three necessary conditions for Darwinian evolution are:

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

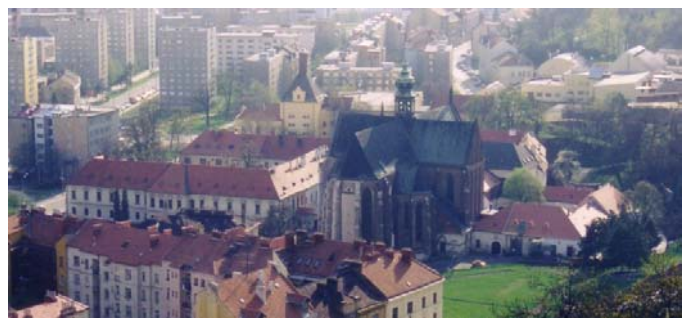
Darwin discovered the principle of **natural selection** from empirical observations in nature.

1. flower color is purple or white
2. flower position is axil or terminal
3. stem length is long or short
4. seed shape is round or wrinkled
5. seed color is yellow or green
6. pod shape is inflated or constricted
7. pod color is yellow or green



Gregor Mendel (1822-1884)

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



Gregor Mendel‘s experiments on plant genetics

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

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

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

Gregor Mendel concluded correctly from his experiments:

1. that the inheritance of each trait is determined by "units" or "factors" that are passed on to descendents unchanged (these units are now called genes)
2. that an individual inherits one such unit from each parent for each trait
3. that a trait may not show up in an individual but can still be passed on to the next generation.

Gregor Mendel‘s experiments on plant genetics

Ronald Fisher (1890-1962)

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

frequencies: $x_i = [A_i]$; genotypes: $A_i \cdot A_j$

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

Mendel

Darwin

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

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

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

Ronald Fisher‘s selection equation: The genetical theory of natural selection. Oxford, UK, Clarendon Press, 1930.

If Charles Darwin would have written the „Origin"
in mathematical language, how would he have done it?

What did we learn about evolution from *in vitro* experiments?

Quantitative systems biology – A challenge for biologists,
chemists, physicists, and mathematicians !

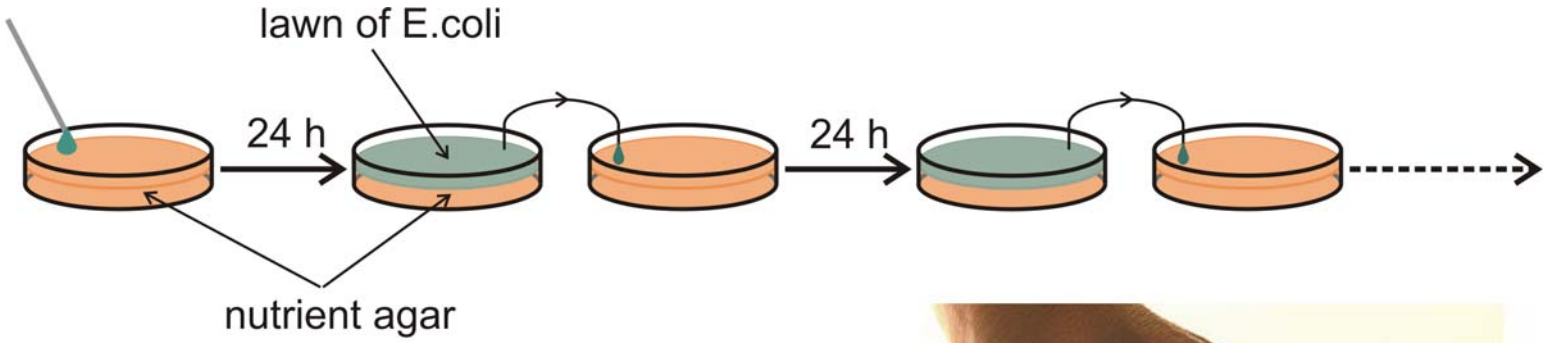
	Generation time (optimal)	Population size (maximal)	Mutation per replication event
Bacteria	20 min	10^{10}	1/400 – 1/300
Viruses	variable	10^{12}	≈ 1
RNA molecules	1 –10 sec	10^{15}	tunable

The world of *in vitro* evolution experiments

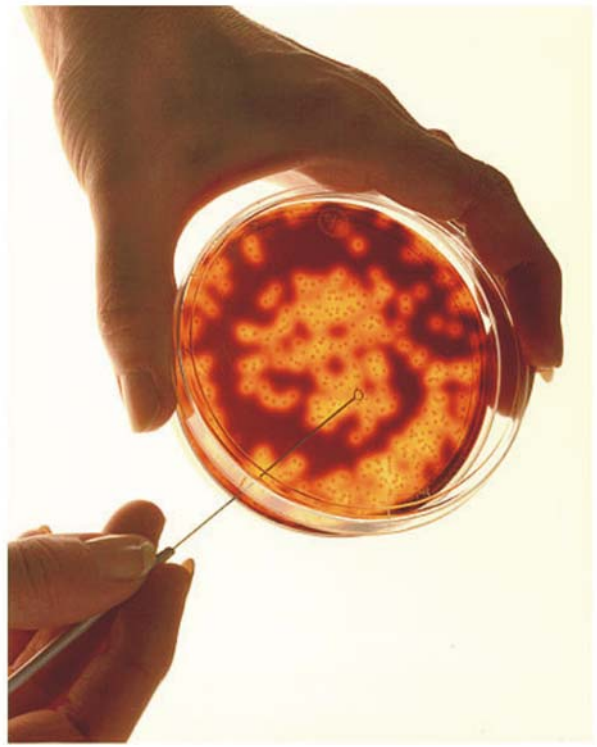
Richard Lenski, 1956 -

Bacterial evolution under controlled conditions: A twenty years experiment.

Richard Lenski, University of Michigan, East Lansing



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



Serial transfer of Escherichia coli cultures in Petri dishes

Bacterial evolution under controlled conditions: A twenty years experiment.

Richard Lenski, University of Michigan, East Lansing

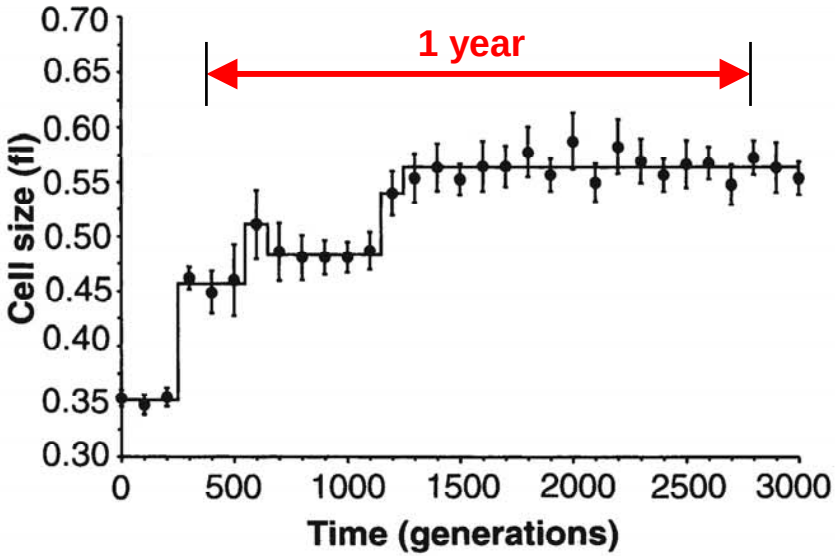


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

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

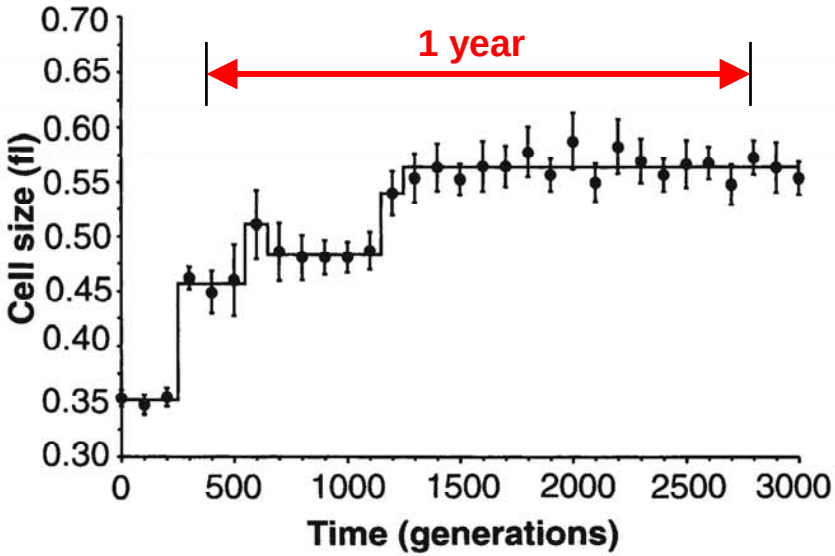


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

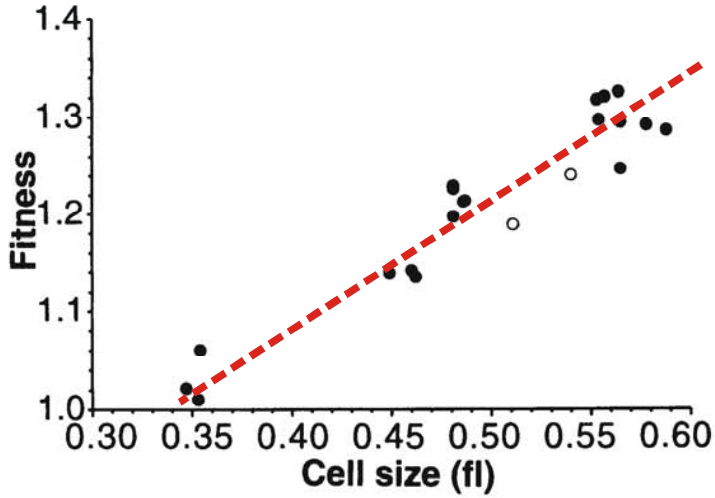
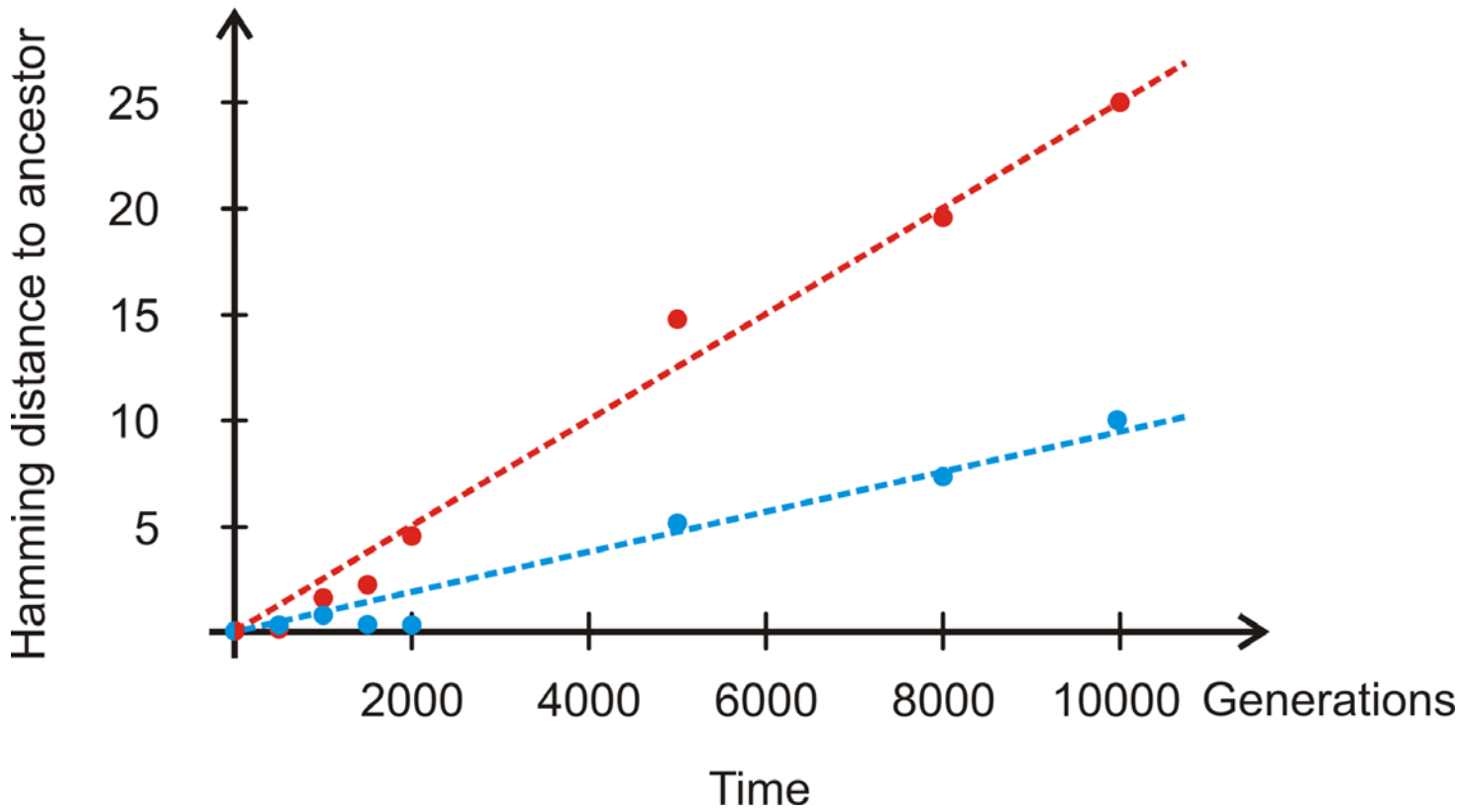


Fig. 2. Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (*6*, *7*). The open symbols indicate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

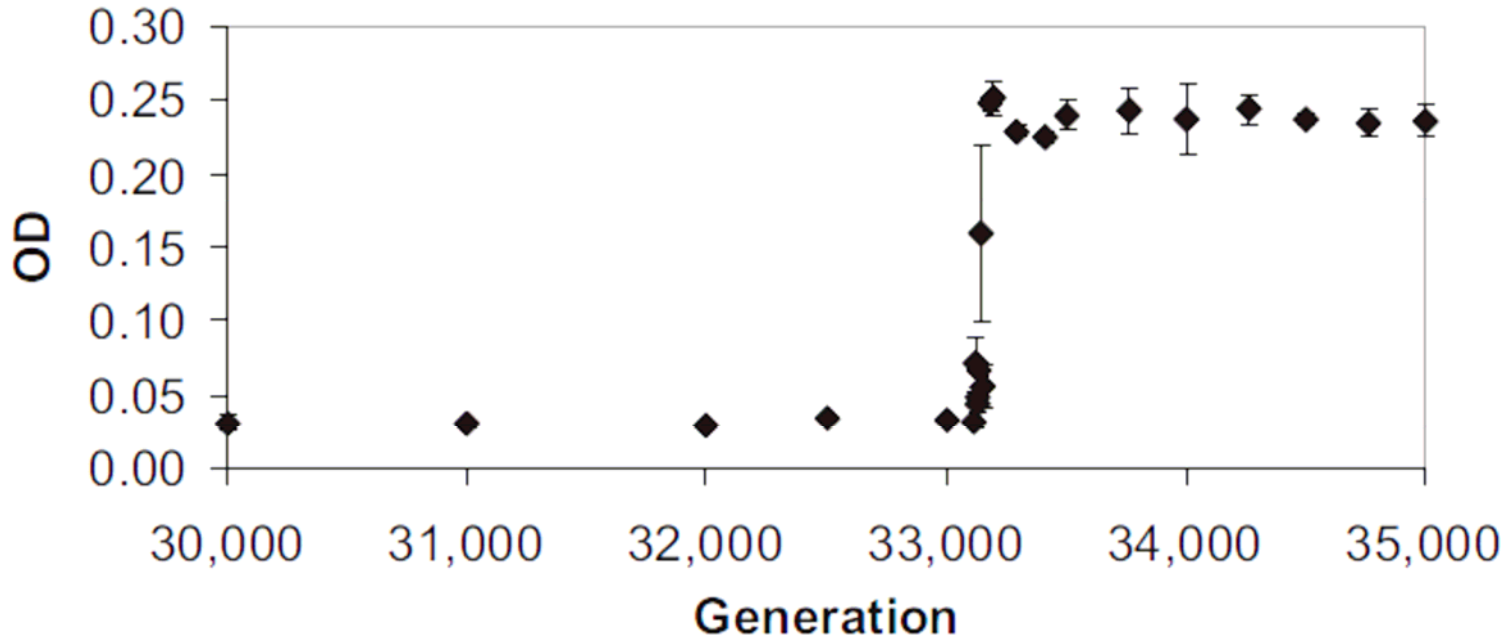


Fig. 1. Population expansion during evolution of the Cit$^+$ phenotype. Samples frozen at various times in the history of population Ara-3 were revived, and three DM25 cultures were established for each generation. Optical density (OD) at 420 nm was measured for each culture at 24 h. Error bars show the range of three values measured for each generation.

Innovation by mutation in long time evolution of Escherichia coli in constant environment

Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008. Proc.Natl.Acad.Sci.USA 105:7899-7906

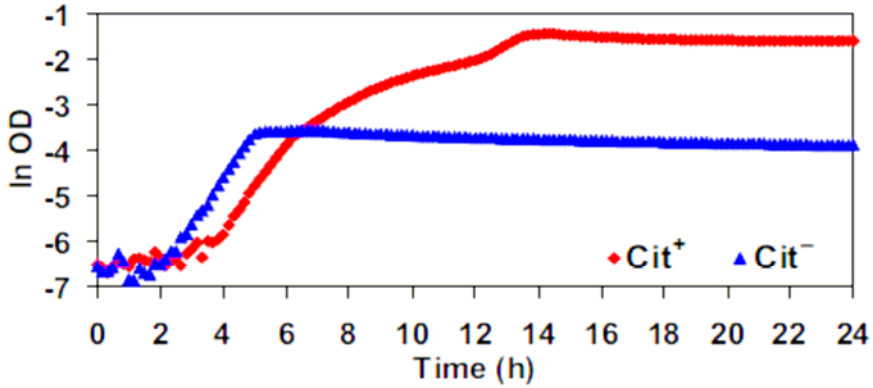


Fig. 2. Growth of Cit⁻ (blue triangles) and Cit⁺ (red diamonds) cells in DM25 medium. Each trajectory shows the average OD for eight replicate mixtures of three clones, all from generation 33,000 of population Ara-3.

Innovation by mutation in long time evolution of Escherichia coli in constant environment

Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008. Proc.Natl.Acad.Sci.USA 105:7899-7906

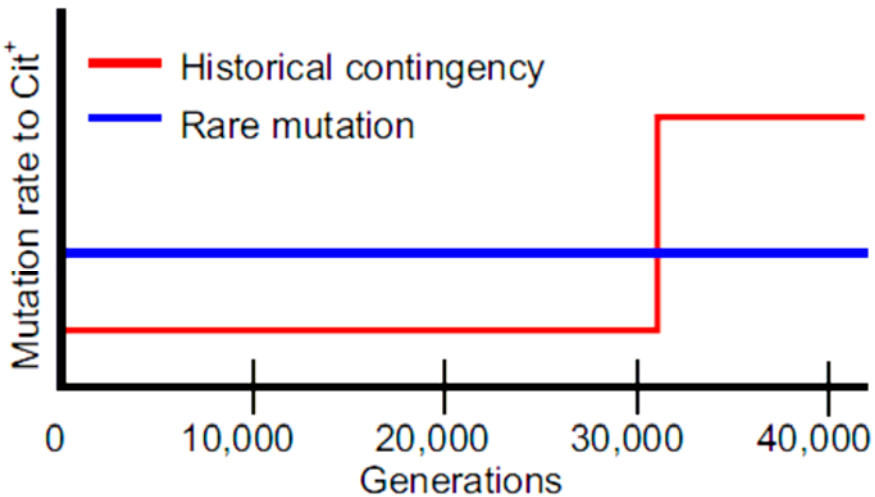


Fig. 3. Alternative hypotheses for the origin of the Cit⁺ function. According to the rare-mutation hypothesis, the probability of mutation from Cit⁻ to Cit⁺ was low but constant over time. Under the historical-contingency hypothesis, the probability of this transition increased when a mutation arose that produced a genetic background with a higher mutation rate to Cit⁺.

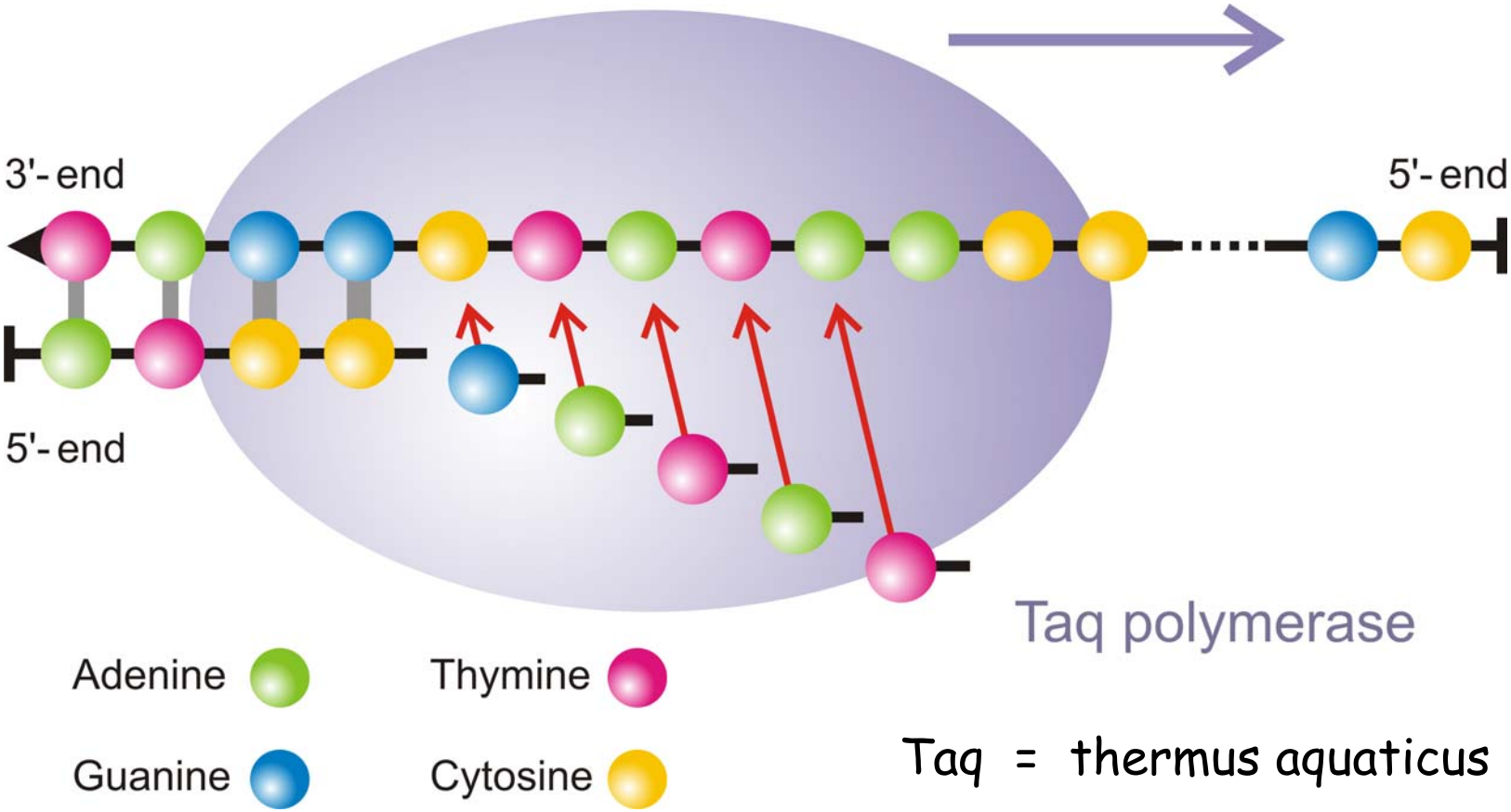
Charles Darwin, 1809-1882

Three necessary conditions for Darwinian evolution are:

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

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**:

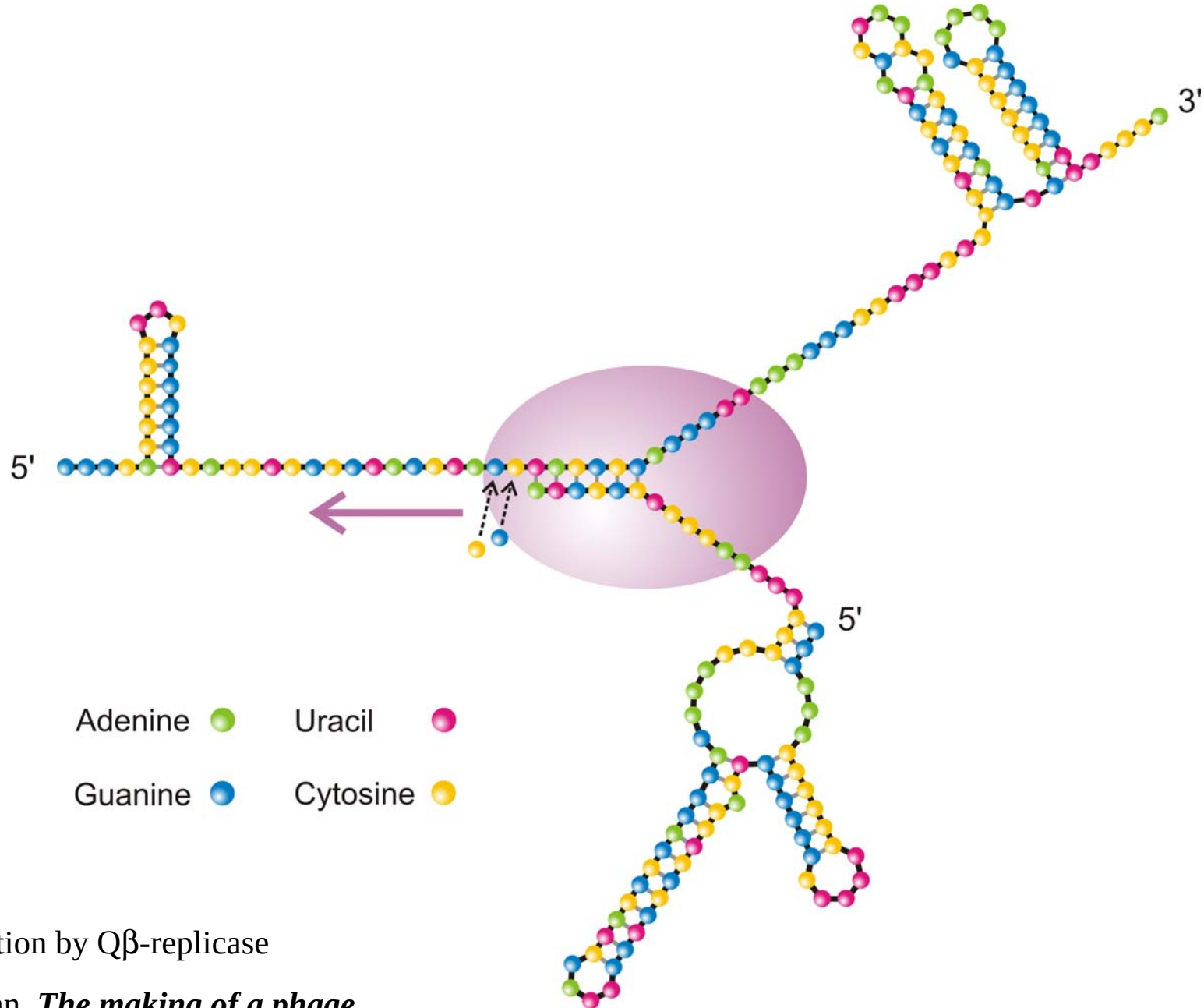
Darwinian evolution in the test tube



Taq = thermus aquaticus

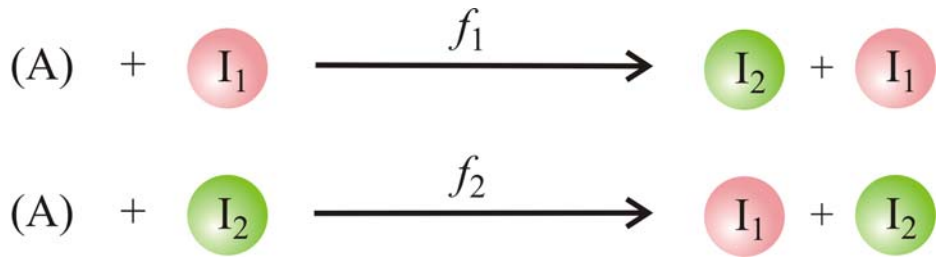
Accuracy of replication: $Q = q_1 \cdot q_2 \cdot q_3 \cdot \ldots \cdot q_n$

The logics of DNA replication



RNA replication by Qβ-replicase

C. Weissmann, ***The making of a phage***.
FEBS Letters **40** (1974), S10-S18



$$\frac{dx_1}{dt} = f_2\, x_2 \quad \text{and} \quad \frac{dx_2}{dt} = f_1\, x_1$$

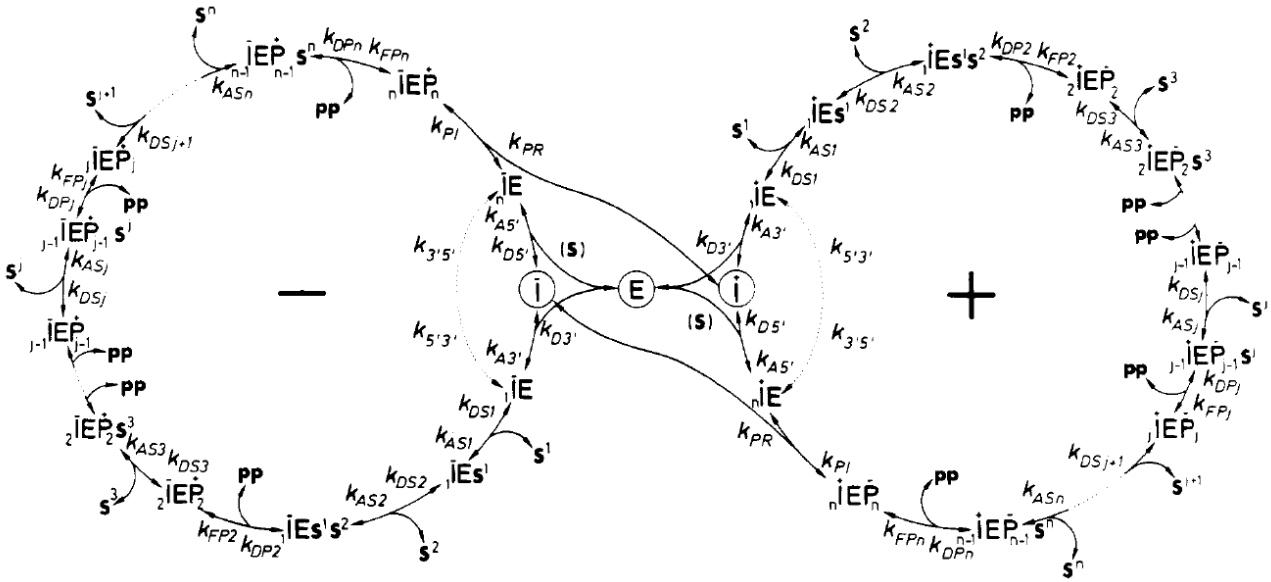
$$x_1 = \sqrt{f_2}\;\xi_1\,, \quad x_2 = \sqrt{f_1}\;\xi_2\,, \quad \zeta = \xi_1 + \xi_2\,, \quad \eta = \xi_1 - \xi_2\,, \quad f = \sqrt{f_1 f_2}$$

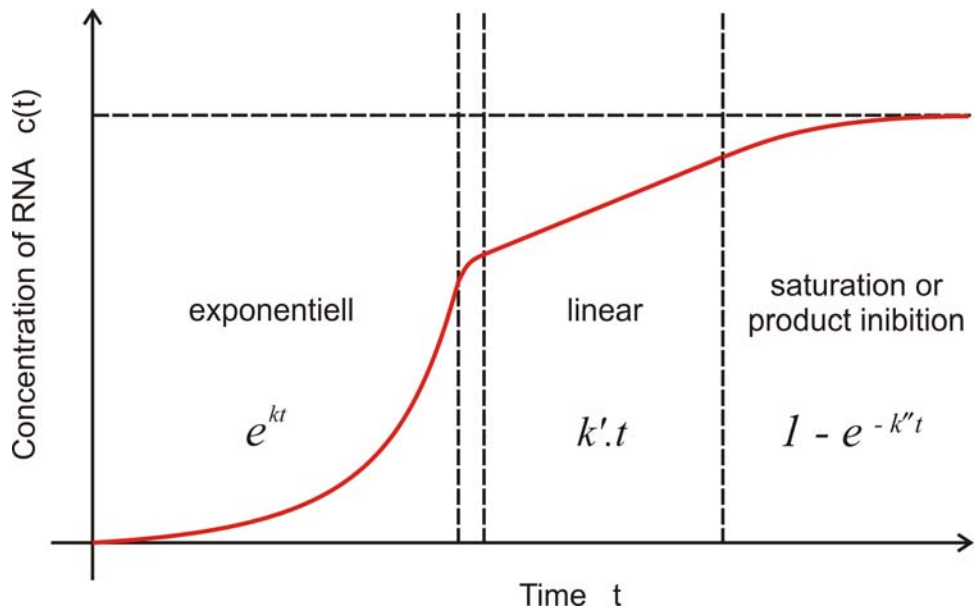
$$\eta(t) = \eta(0)\, e^{-ft}$$

$$\zeta(t) = \zeta(0)\, e^{ft}$$

Complementary replication as the simplest molecular mechanism of reproduction

Christof K. Biebricher,
1941-2009





Kinetics of RNA replication

C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
Biochemistry **22**:2544-2559, 1983

Evolution in the test tube:

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

DOI: 10.1002/anie.200701369

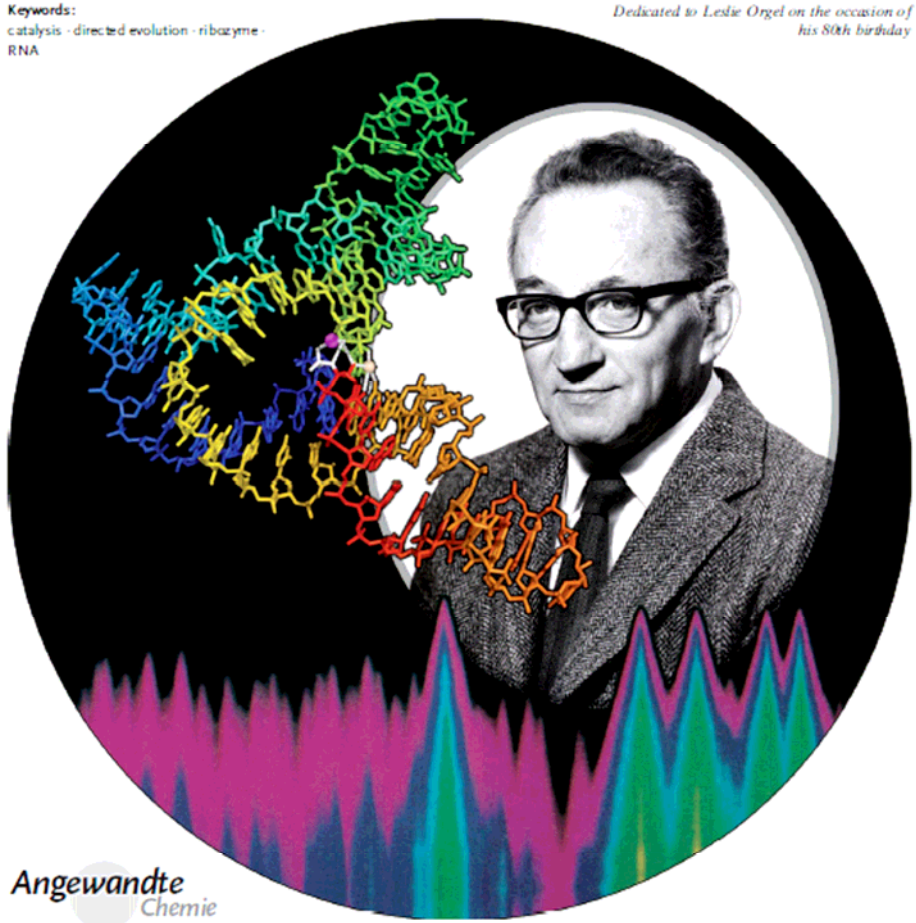
Molecular Evolution

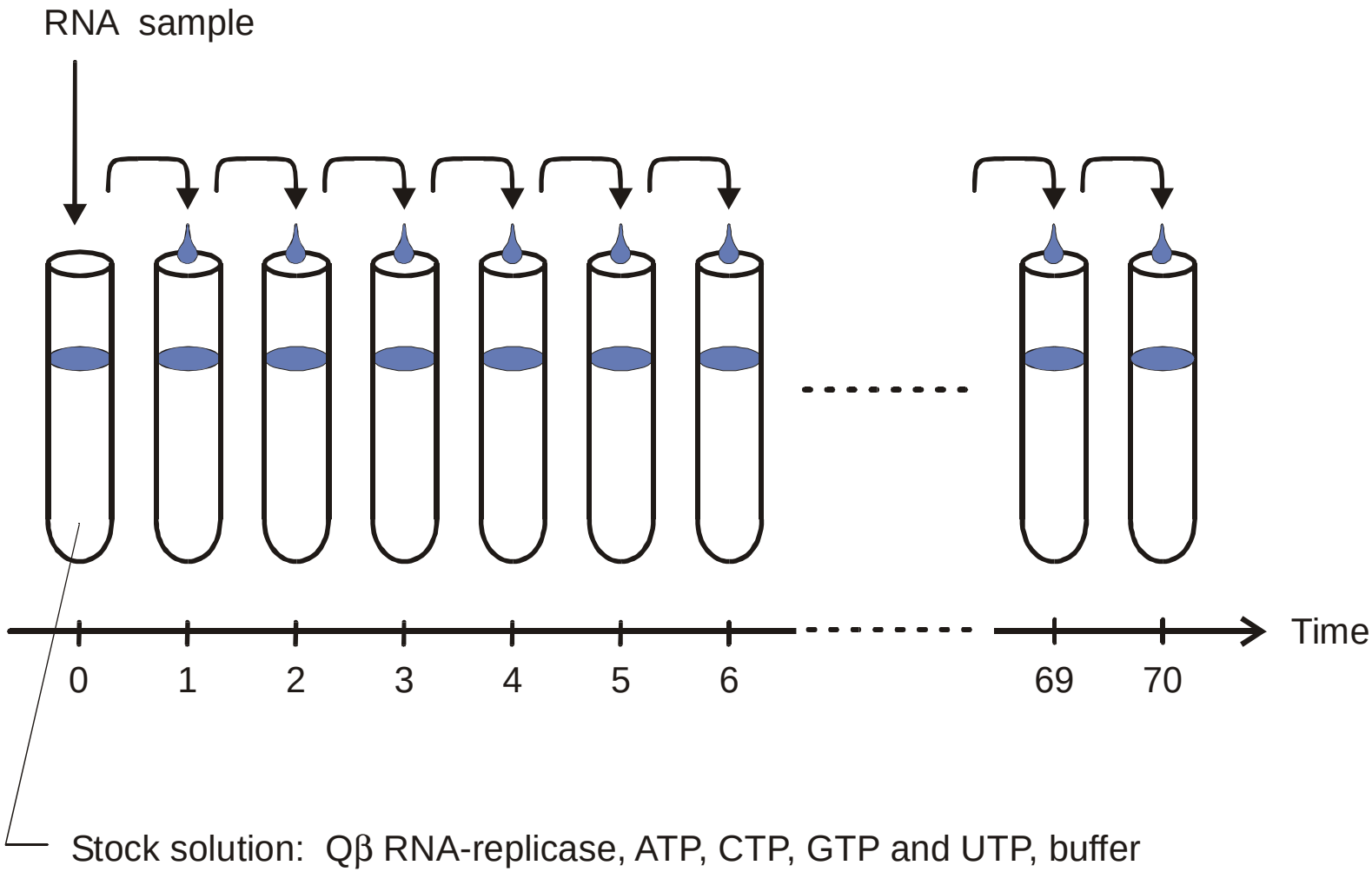
Forty Years of In Vitro Evolution**

*Gerald F. Joyce**

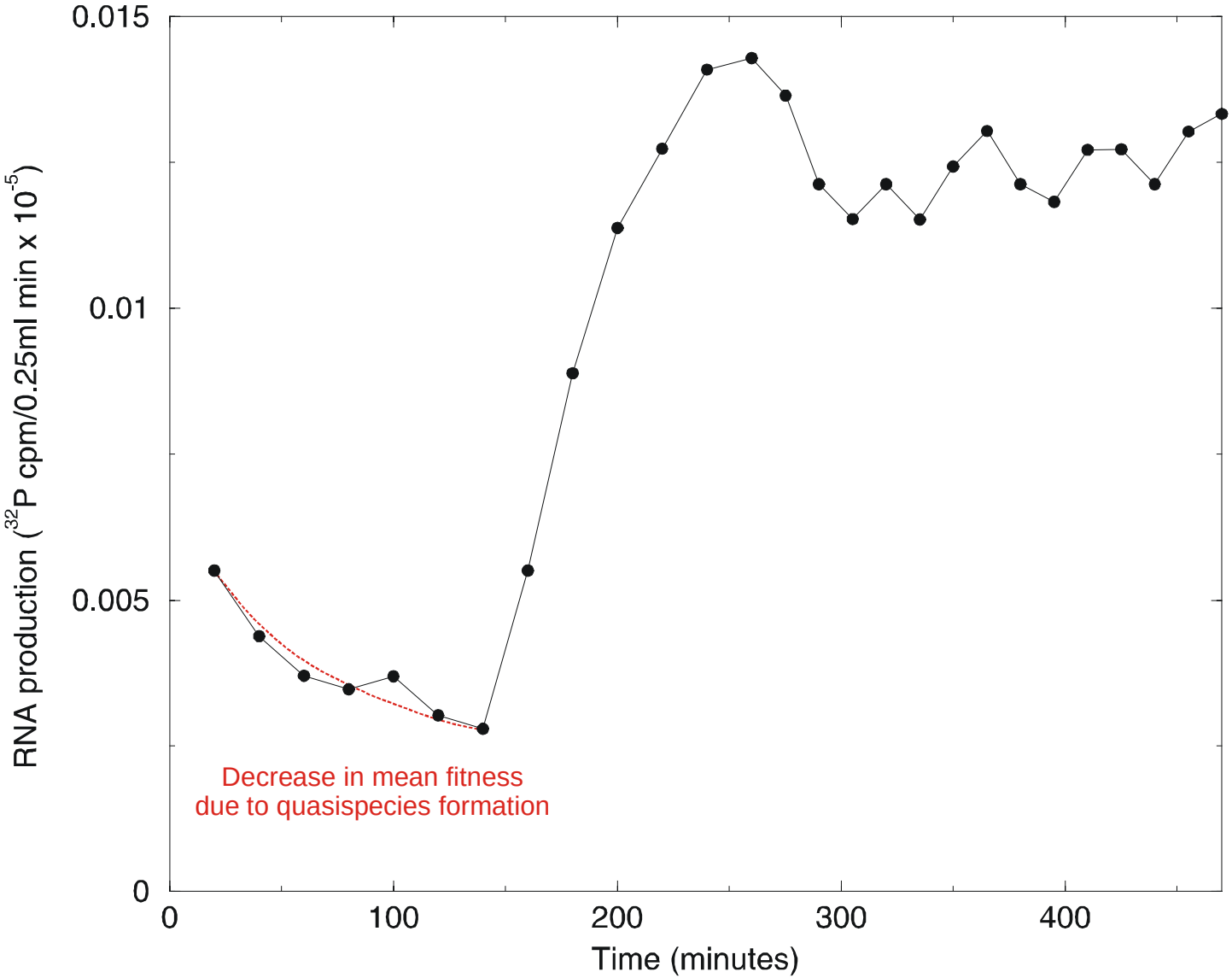
Keywords:
catalysis · directed evolution · ribozyme · RNA

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





Application of serial transfer technique to evolution of RNA in the test tube

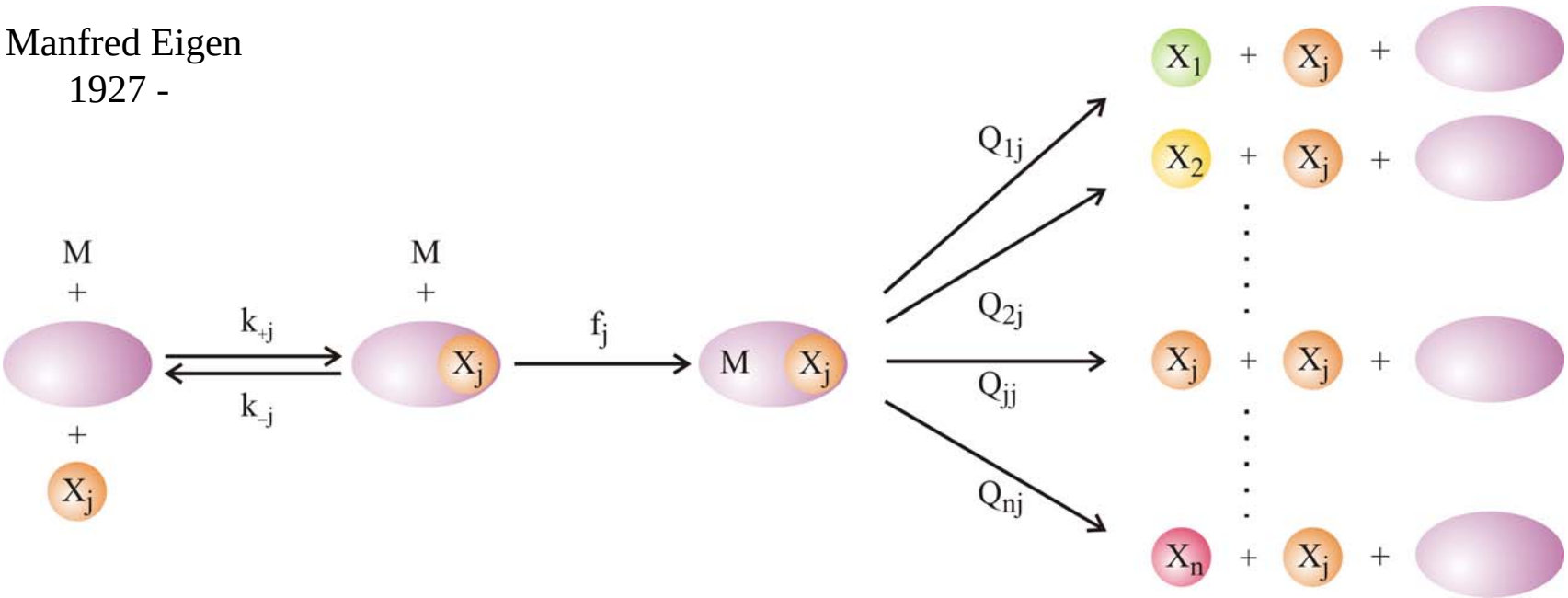


The increase in RNA production rate during a serial transfer experiment

Manfred Eigen
1927 -

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$\Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$



Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977. *Naturwissenschaften* 64:541, 65:7 und 65:341

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi = \sum_{i=1}^{n} Q_{ji}\, f_i\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$\Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$

Decomposition of matrix W

$$\mathrm{W} = \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1n} \\ w_{21} & w_{22} & \ldots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \ldots & w_{nn} \end{pmatrix} = \mathrm{Q} \cdot \mathrm{F} \quad \text{with}$$

$$\mathrm{Q} = \begin{pmatrix} Q_{11} & Q_{12} & \ldots & Q_{1n} \\ Q_{21} & Q_{22} & \ldots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \ldots & Q_{nn} \end{pmatrix} \quad \text{and} \quad \mathrm{F} = \begin{pmatrix} f_1 & 0 & \ldots & 0 \\ 0 & f_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & f_n \end{pmatrix}$$

Factorization of the value matrix W separates mutation and fitness effects.

Mutation-selection equation: $[\mathrm{I}_i] = x_i \geq 0,\ f_i \geq 0,\ Q_{ij} \geq 0$

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} Q_{ij}\, f_j\, x_j - x_i\, \phi, \quad i = 1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j\, x_j = \overline{f}$$

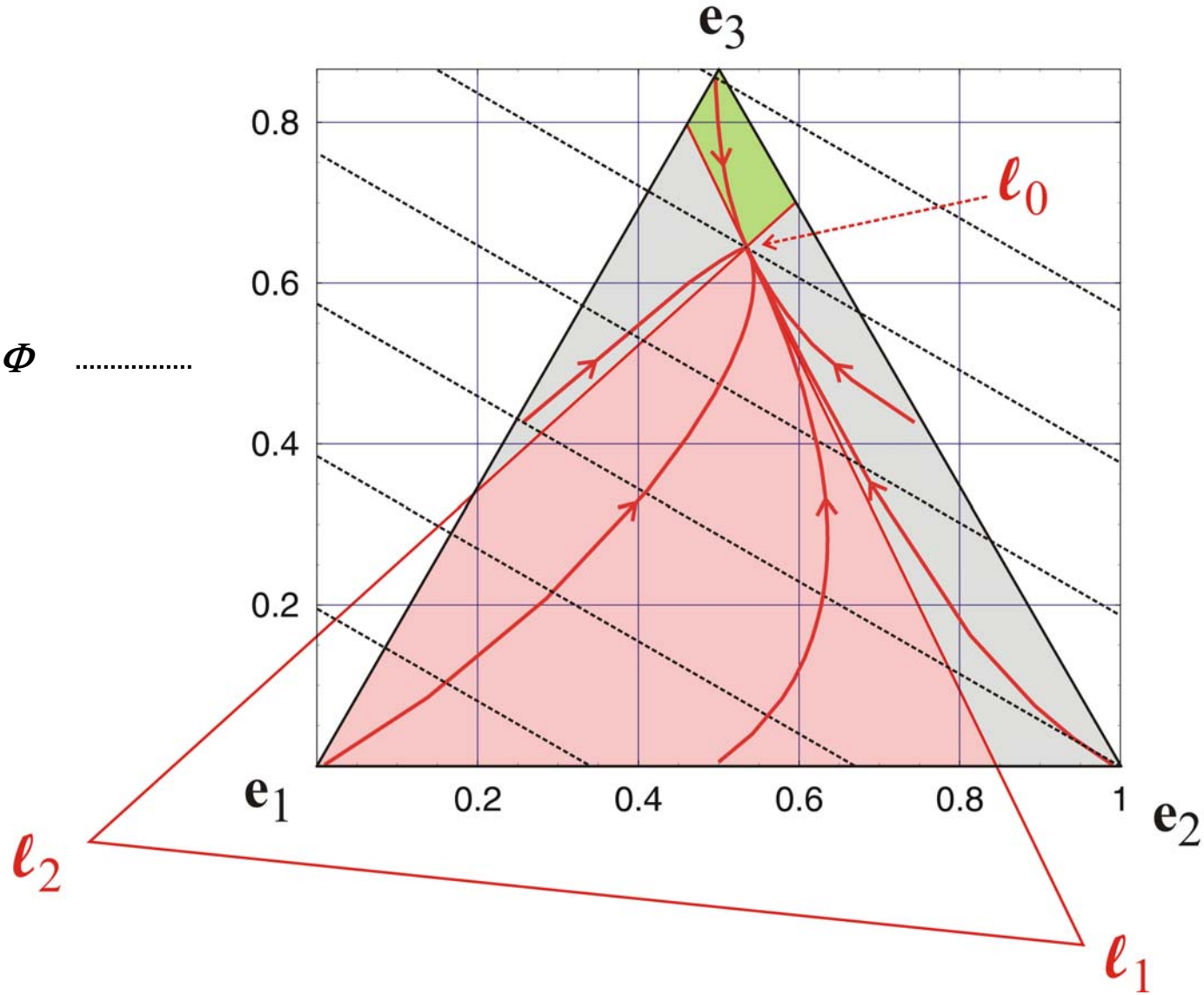
solutions are obtained after integrating factor transformation by means of an eigenvalue problem

$$x_i(t) = \frac{\sum_{k=0}^{n-1} \ell_{ik} \cdot c_k(0) \cdot \exp(\lambda_k t)}{\sum_{j=1}^{n} \sum_{k=0}^{n-1} \ell_{jk} \cdot c_k(0) \cdot \exp(\lambda_k t)}; \quad i = 1,2,\cdots,n; \quad c_k(0) = \sum_{i=1}^{n} h_{ki}\, x_i(0)$$

$$W \div \left\{ f_i Q_{ij};\ i, j = 1,2,\cdots,n \right\};\ L = \left\{ \ell_{ij};\ i, j = 1,2,\cdots,n \right\};\ L^{-1} = H = \left\{ h_{ij};\ i, j = 1,2,\cdots,n \right\}$$

$$L^{-1} \cdot W \cdot L \quad = \quad \Lambda \quad = \quad \left\{ \lambda_k; k = 0,1,\cdots,n-1 \right\}$$

constant level sets of Φ .....



Selection of quasispecies with $f_1 = 1.9$, $f_2 = 2.0$, $f_3 = 2.1$, and $p = 0.01$, parametric plot on S_3

Phenomenon	Optimization of fitness	Unique selection outcome
Selection	**yes**	**yes**
Recombination and selection Independent genes	**yes**	**no**
Recombination and selection Interacting genes	**no**	**no**
Mutation and selection	**no**	**yes**

The Darwinian mechanism of variation and selection is a very powerful **optimization heuristic**.

The Darwinian mechanism and optimization of fitness

Chain length and error threshold

$$Q \cdot \sigma \;=\; (1-p)^n \cdot \sigma \;\geq\; 1 \;\Rightarrow\; n \cdot \ln(1-p) \geq -\ln\sigma$$

p … constant : $n_{\max} \approx \dfrac{\ln\sigma}{p}$

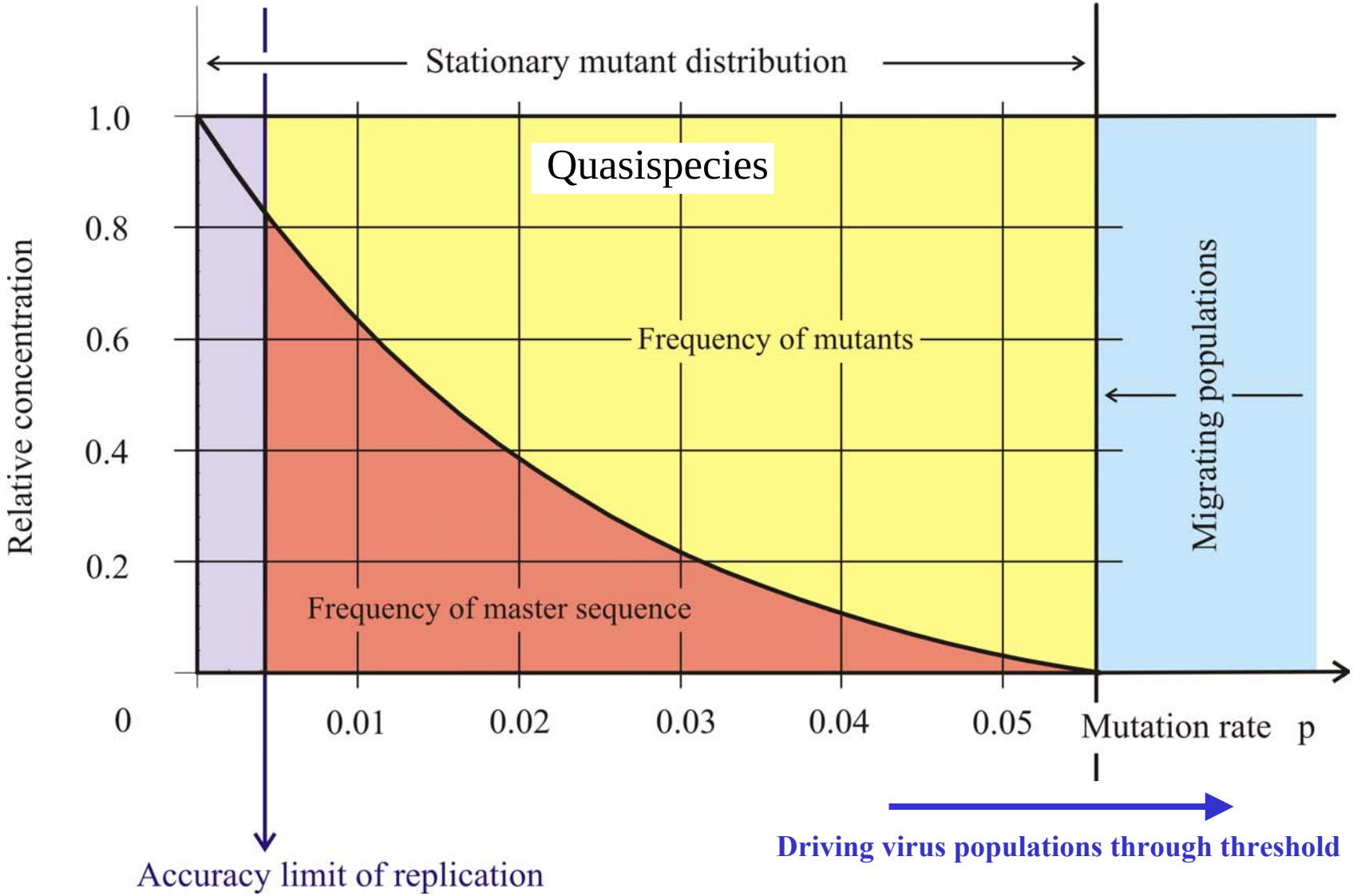
n … constant : $p_{\max} \approx \dfrac{\ln\sigma}{n}$

$Q = (1-p)^n$ … replication accuracy

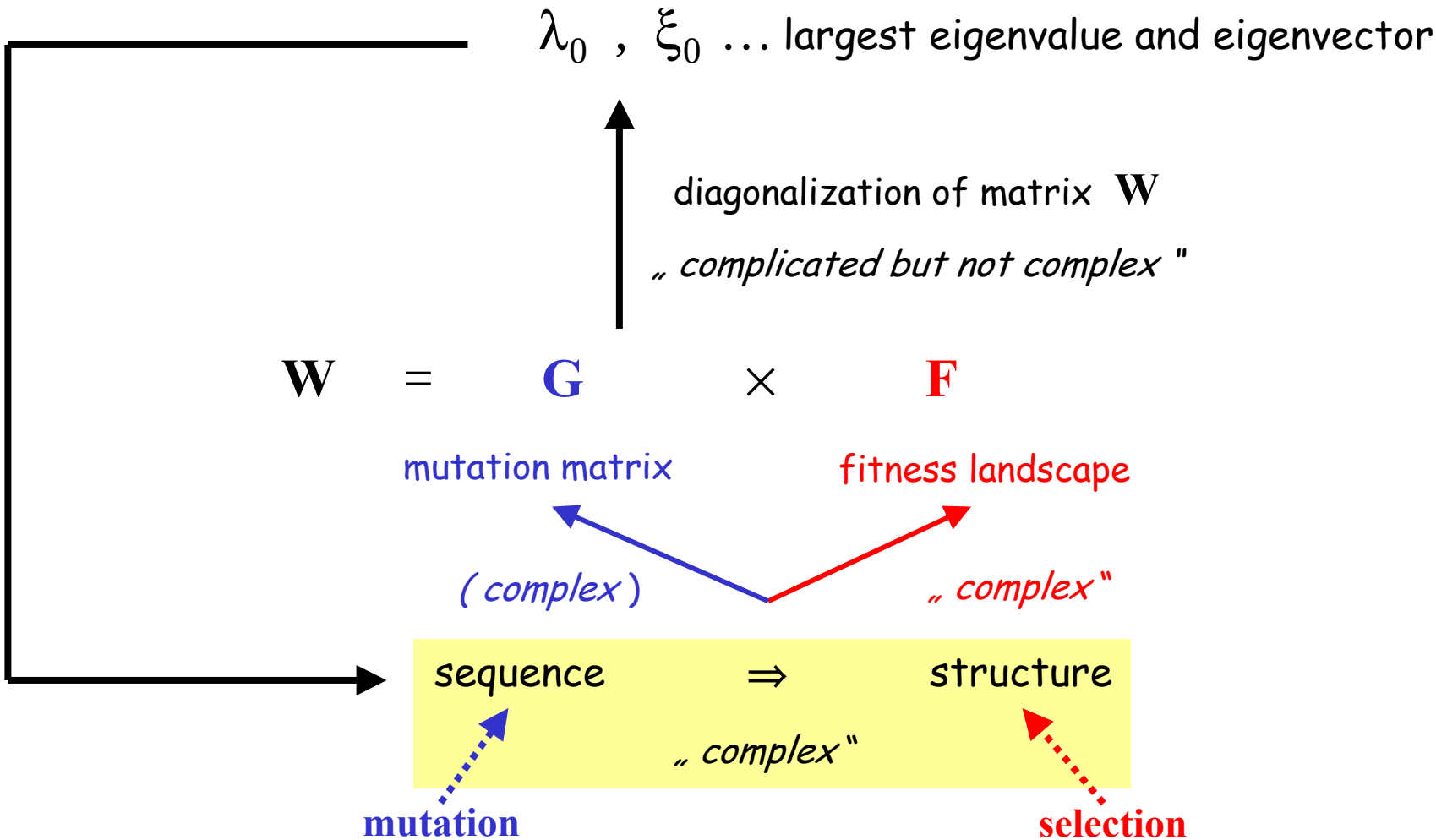
p … error rate

n … chain length

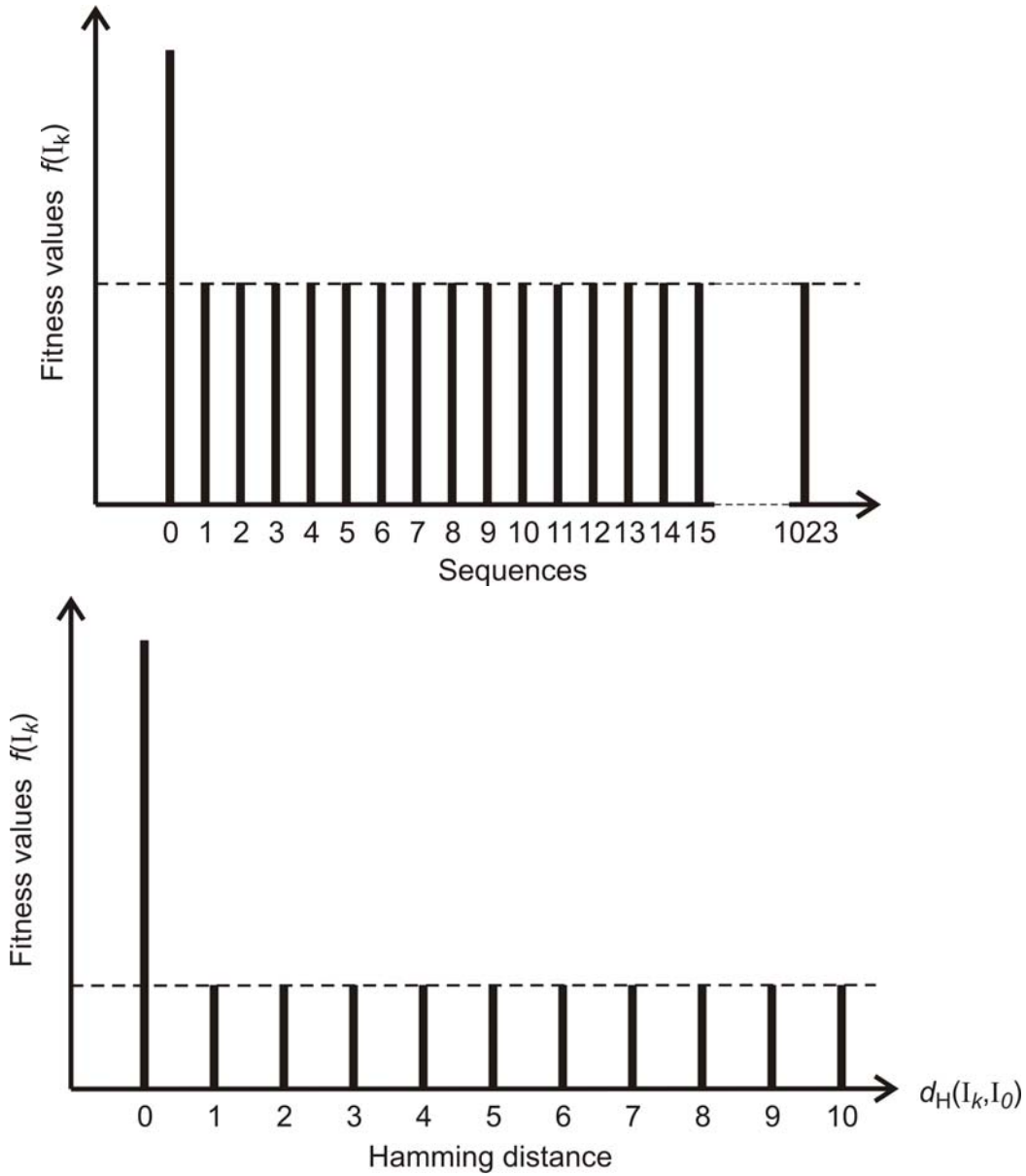
$\sigma = \dfrac{f_m}{\sum_{j \neq m} f_j}$ … superiority of master sequence



The error threshold in replication: No mutational backflow approximation



Complexity in molecular evolution



The single peak fitness landscapes corresponding to a mean field approximation

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

SELF-REPLICATION WITH ERRORS

A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *
Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria

Received 4th June 1982
Revised manuscript received 23rd August 1982
Accepted 30th August 1982

Key words: Polynucleotide replication; Quasi-species; Point mutation; Mutant class; Stochastic replication

A model for polynucleotide replication is presented and analyzed by means of perturbation theory. Two basic assumptions allow handling of sequences up to a chain length of $\nu = 80$ explicitly: point mutations are restricted to a two-digit model and individual sequences are subsumed into mutant classes. Perturbation theory is in excellent agreement with the exact results for long enough sequences ($\nu > 20$).

1. Introduction

Eigen [8] proposed a formal kinetic equation (eq. 1) which describes self-replication under the constraint of constant total population size:

$$\frac{dx_i}{dt} = \dot{x}_i = \sum_j w_{ij} x_j - \frac{x_i}{c}\phi; \; i = 1, \ldots, n \; ^{\dagger} \qquad (1)$$

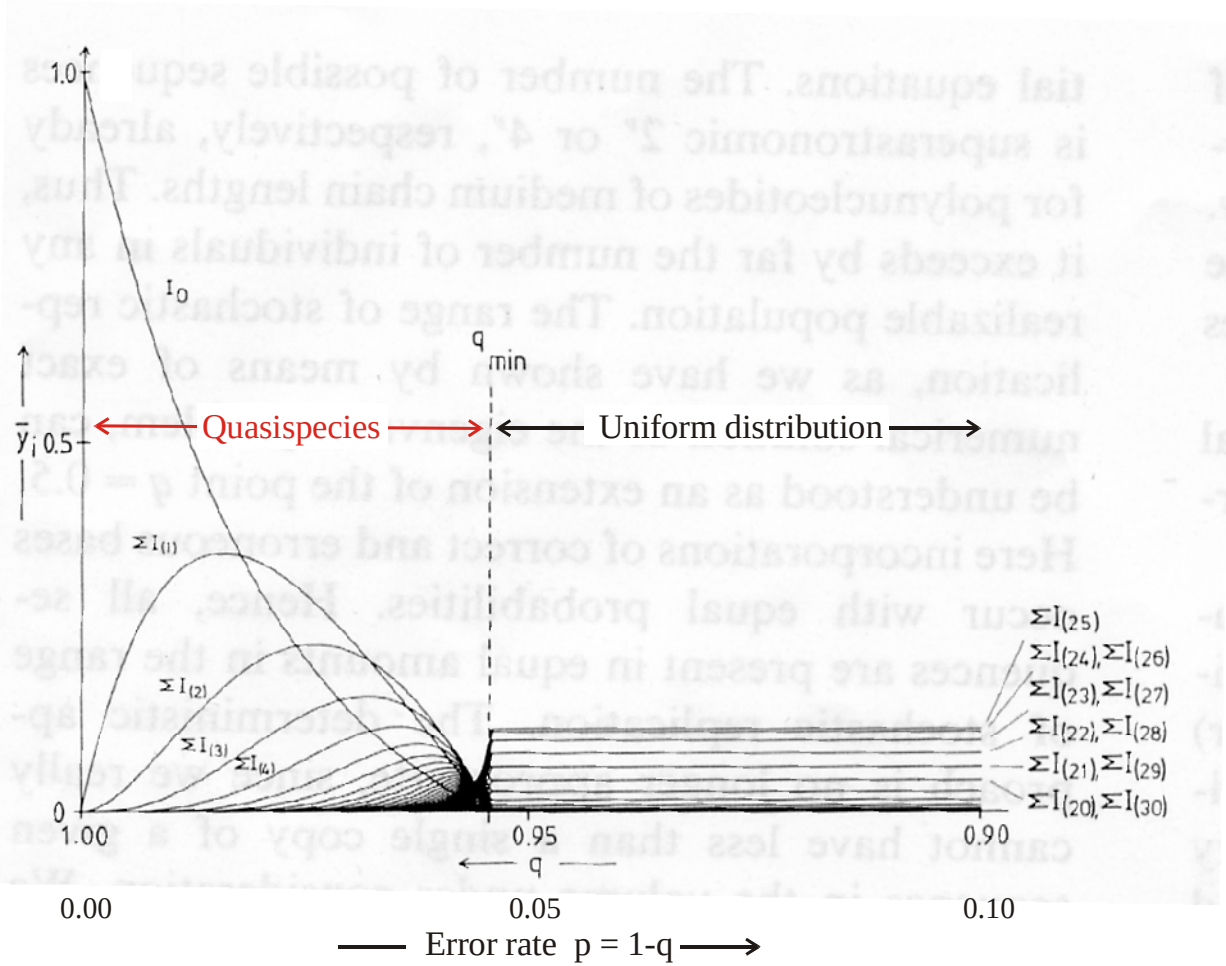
By x_i we denote the population number or concentration of the self-replicating element $\mathbf{I}_i$, i.e., $x_i = [\mathbf{I}_i]$. The total population size or total concentration $c = \Sigma_i x_i$ is kept constant by proper adjustment of the constraint ϕ: $\phi = \Sigma_i \Sigma_j w_{ij} x_j$. Characteristically, this constraint has been called 'constant organization'. The relative values of diagonal (w_{ii}) and off-diagonal (w_{ij}, $i \neq j$) rates, as we shall see in detail in section 2, are related to the accuracy of the replication process. The specific properties of eq. 1 are essentially based on the fact that it leads to exponential growth in the absence of constraints ($\phi = 0$) and competitors ($n = 1$).

The non-linear differential equation, eq. 1 – the non-linearity is introduced by the definition of ϕ at constant organization – shows a remarkable feature: it leads to selection of a defined ensemble of self-replicating elements above a certain accuracy threshold. This ensemble of a master and its most frequent mutants is a so-called 'quasi-species' [9]. Below this threshold, however, no selection takes place and the frequencies of the individual elements are determined exclusively by their statistical weights.

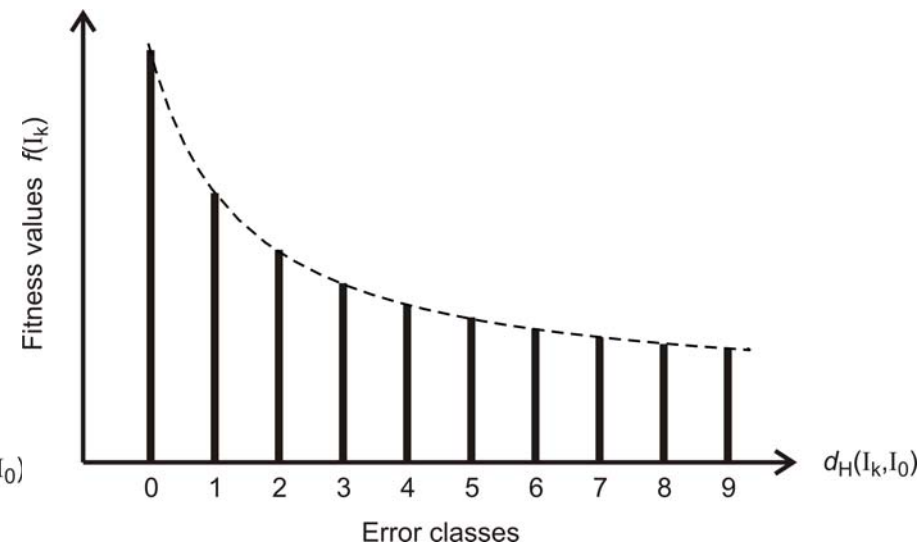
Rigorous mathematical analysis has been performed on eq. 1 [7,15,24,26]. In particular, it was shown that the non-linearity of eq. 1 can be removed by an appropriate transformation. The eigenvalue problem of the linear differential equation obtained thereby may be solved approximately by the conventional perturbation technique

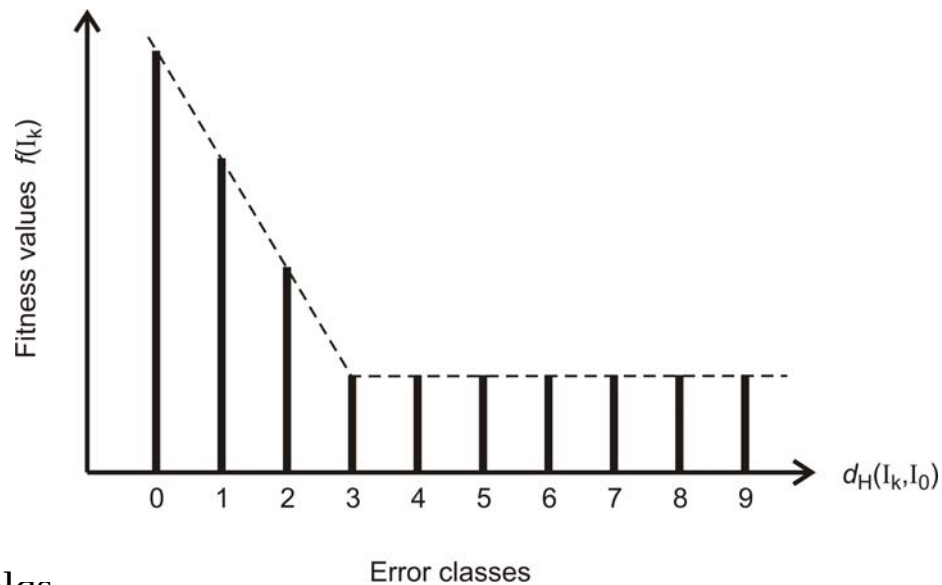
* Dedicated to the late Professor B.L. Jones who was among the first to do rigorous mathematical analysis on the problems described here.
** This paper is considered as part II of Model Studies on RNA replication. Part I is by Gassner and Schuster [14].
† All summations throughout this paper run from 1 to n unless specified differently: $\Sigma_i = \Sigma_{i=1}^{n}$ and $\Sigma_{i,i\neq j} = \Sigma_{i=1}^{j-1} + \Sigma_{i=j+1}^{n}$, respectively.

0301-4622/82/0000-0000/$02.75

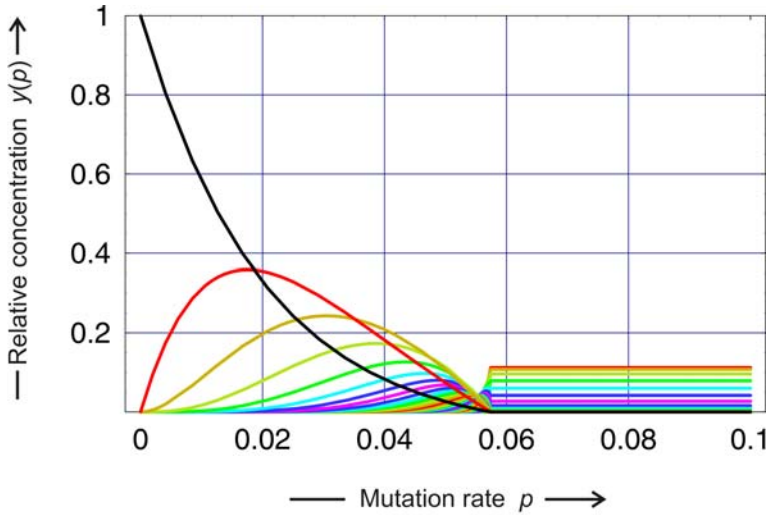


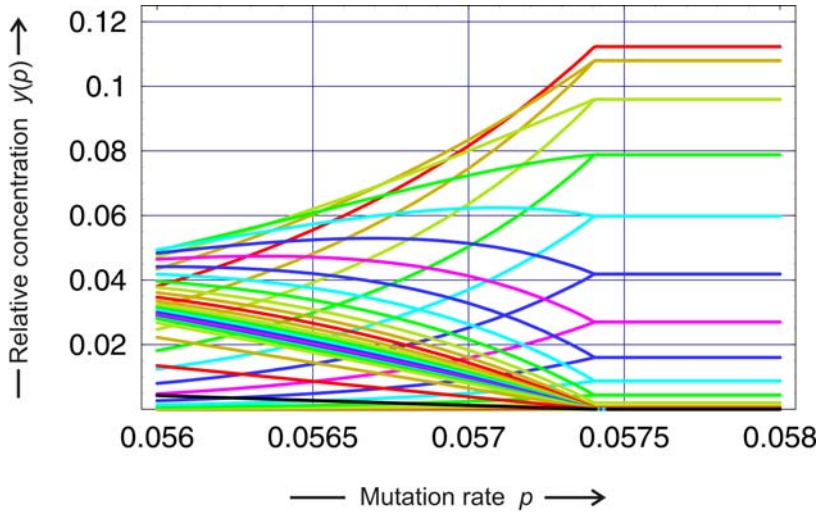
Stationary population or quasispecies as a function of the mutation or error rate *p*

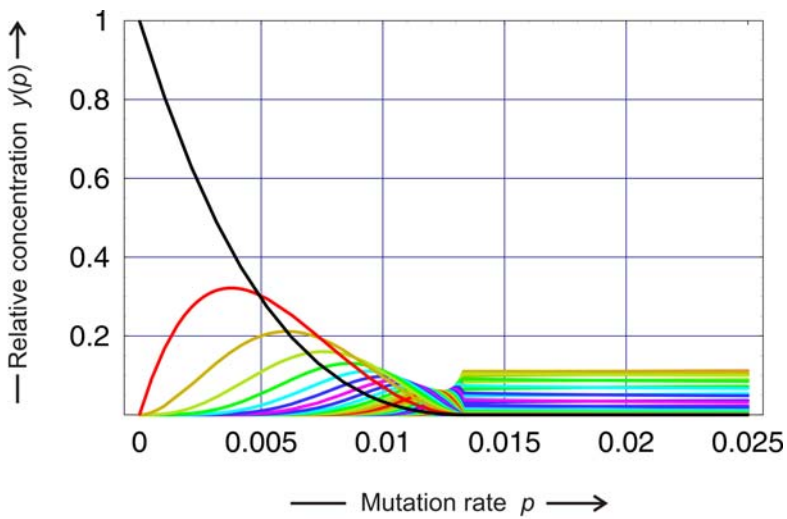


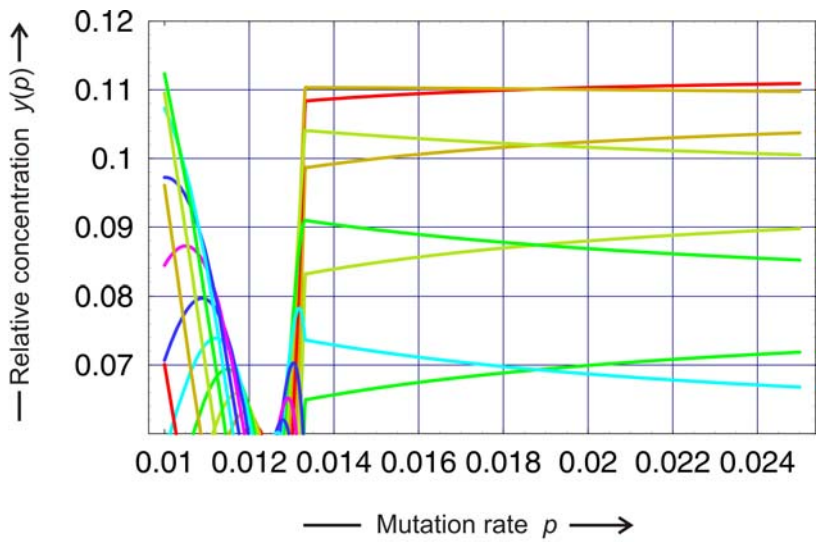


Fitness landscapes and the search for error thresholds

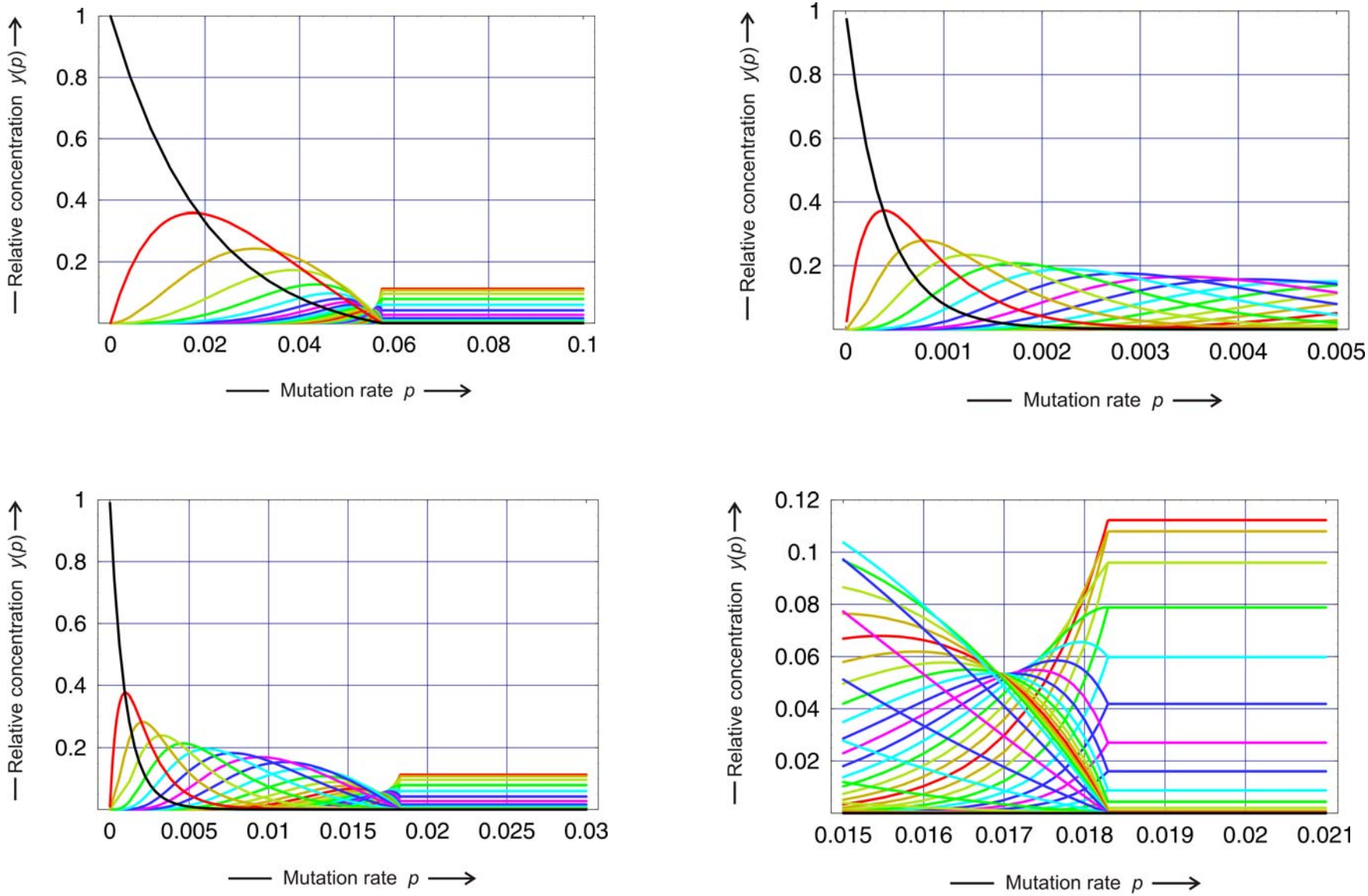




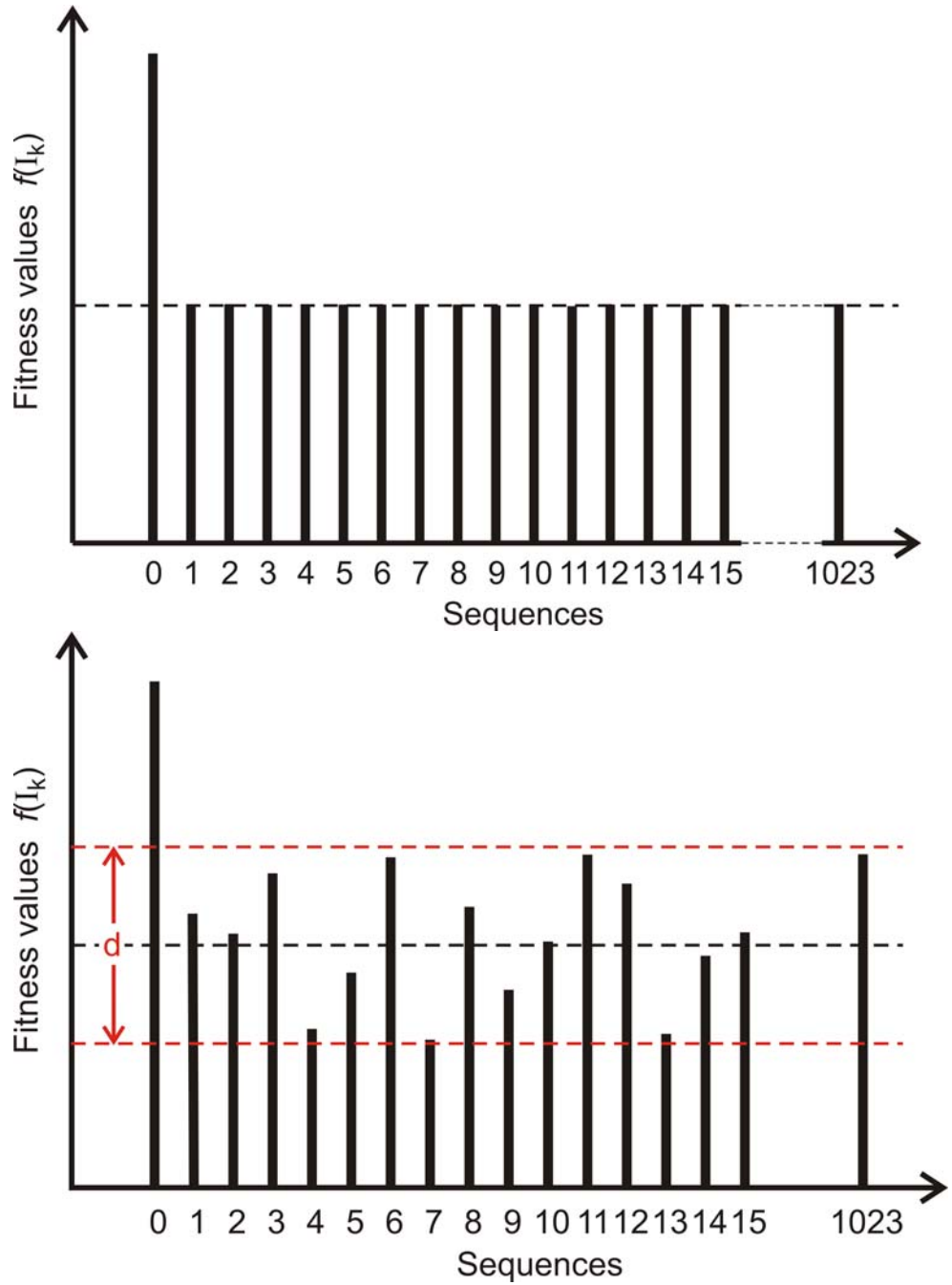




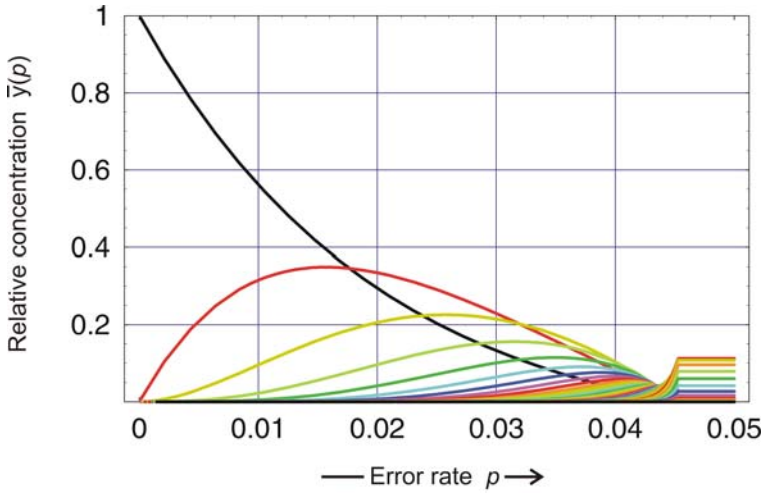
Error threshold on single-peak and hyperbolic landscapes

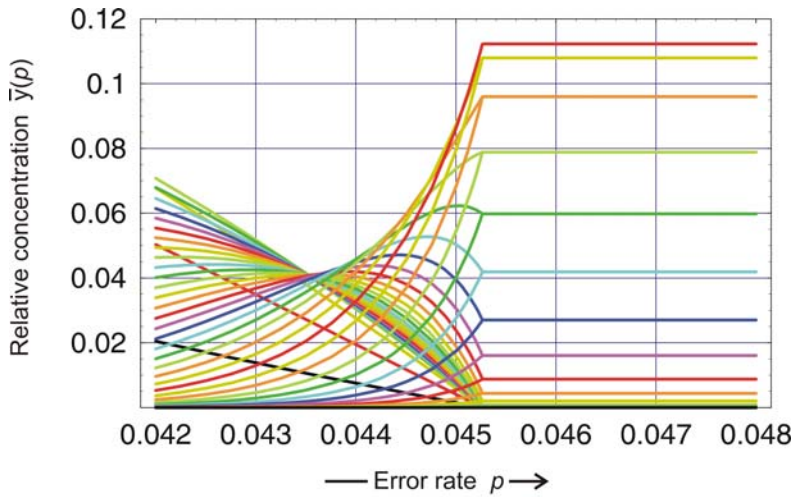


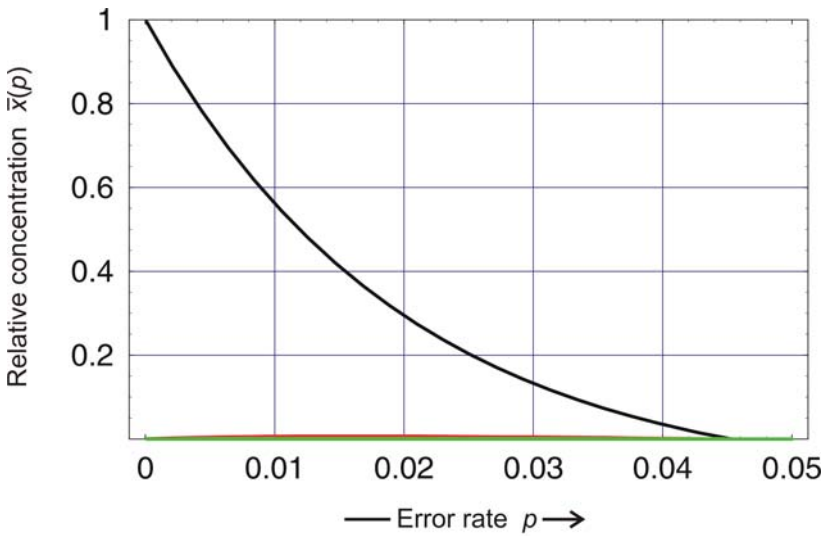
Error threshold on single-peak, linear, and step-linear landscapes

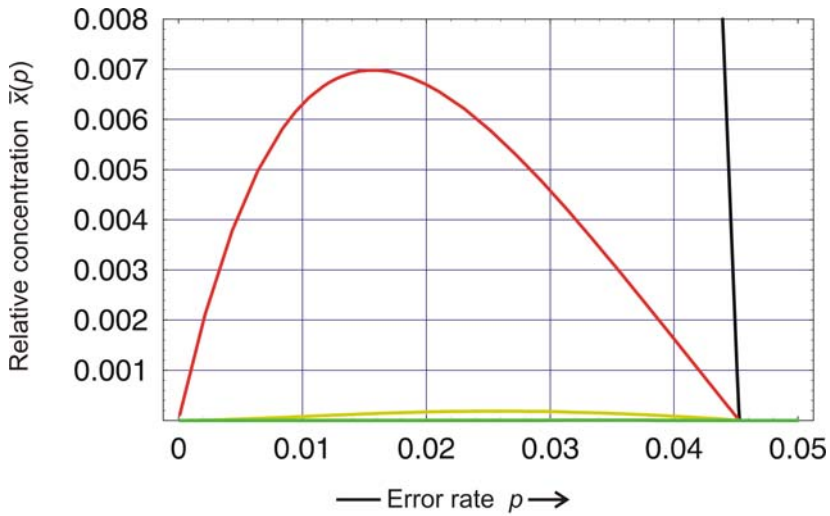


Fitness landscapes showing error thresholds

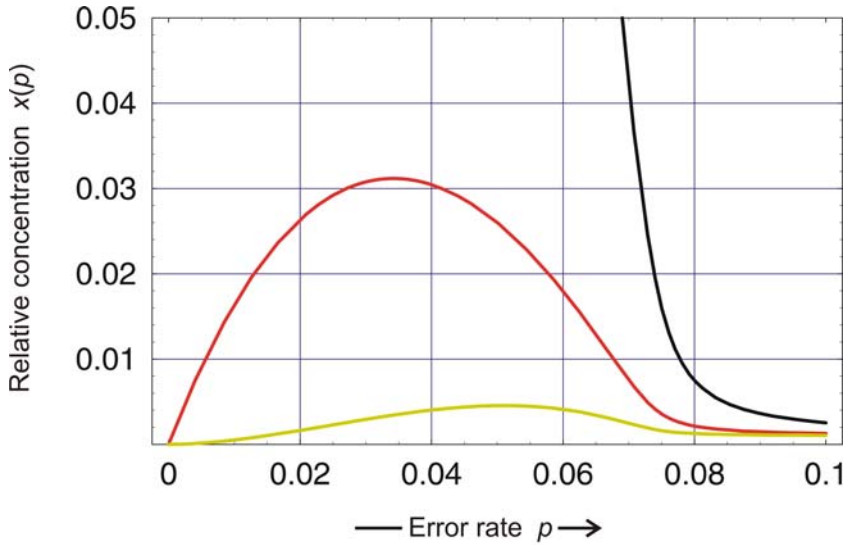


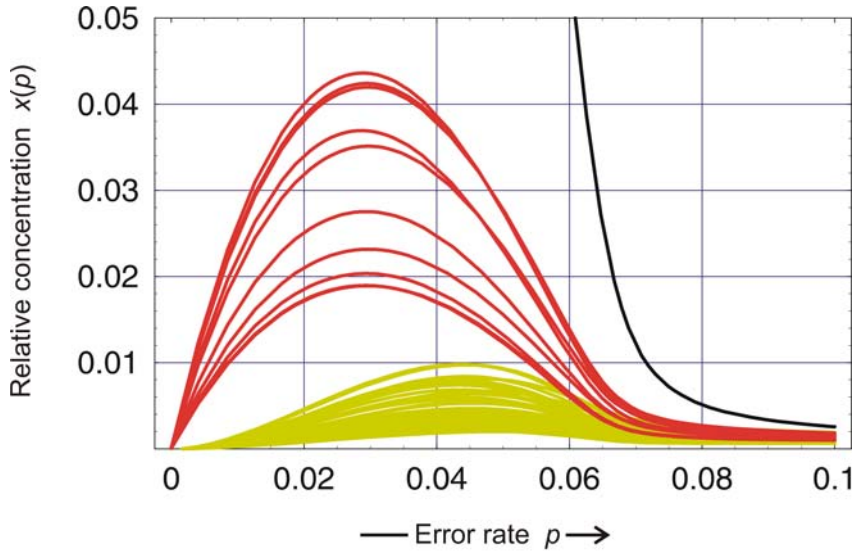


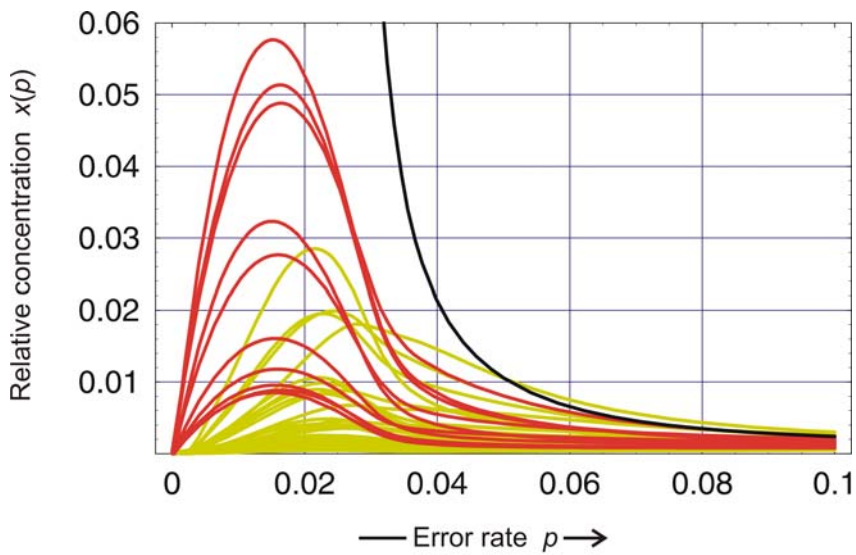




Error threshold on a single peak fitness landscape with n = 50 and σ = 10







Error threshold: Individual sequences

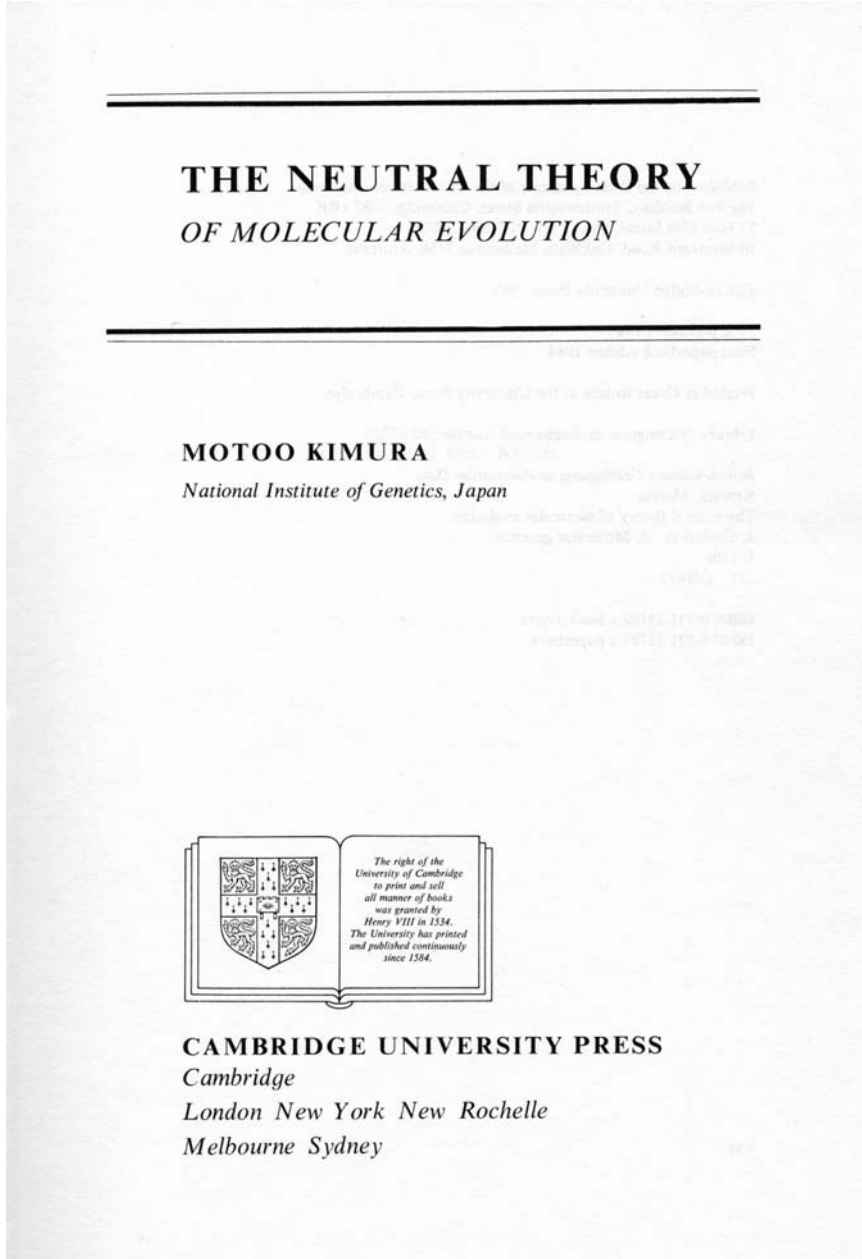
n = 10, σ = 2 and d = 0, 1.0, 1.85

MOTOO KIMURA

Motoo Kimuras population genetics of neutral evolution.

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

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



THE NEUTRAL THEORY

OF MOLECULAR EVOLUTION

MOTOO KIMURA

National Institute of Genetics, Japan

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

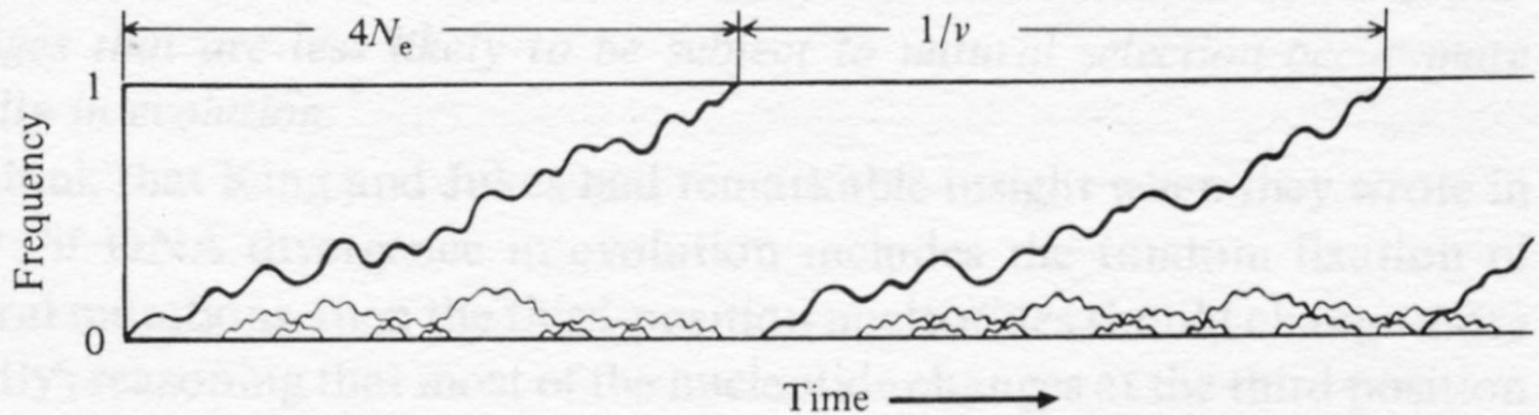
CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York New Rochelle

Melbourne Sydney

Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.

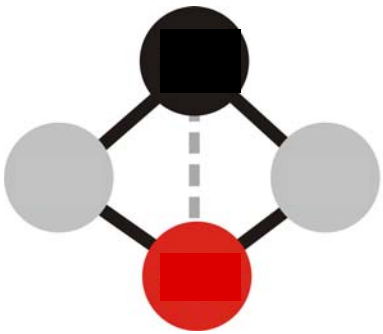


Motoo Kimura

Is the Kimura scenario correct for frequent mutations?

$\mathbf{d_H = 1}$

$$\lim_{p\to 0} x_1(p) = x_2(p) = 0.5$$



$\mathbf{d_H = 2}$

$$\lim_{p\to 0} x_1(p) = a$$

$$\lim_{p\to 0} x_2(p) = 1 - a$$

$\mathbf{d_H \geq 3}$

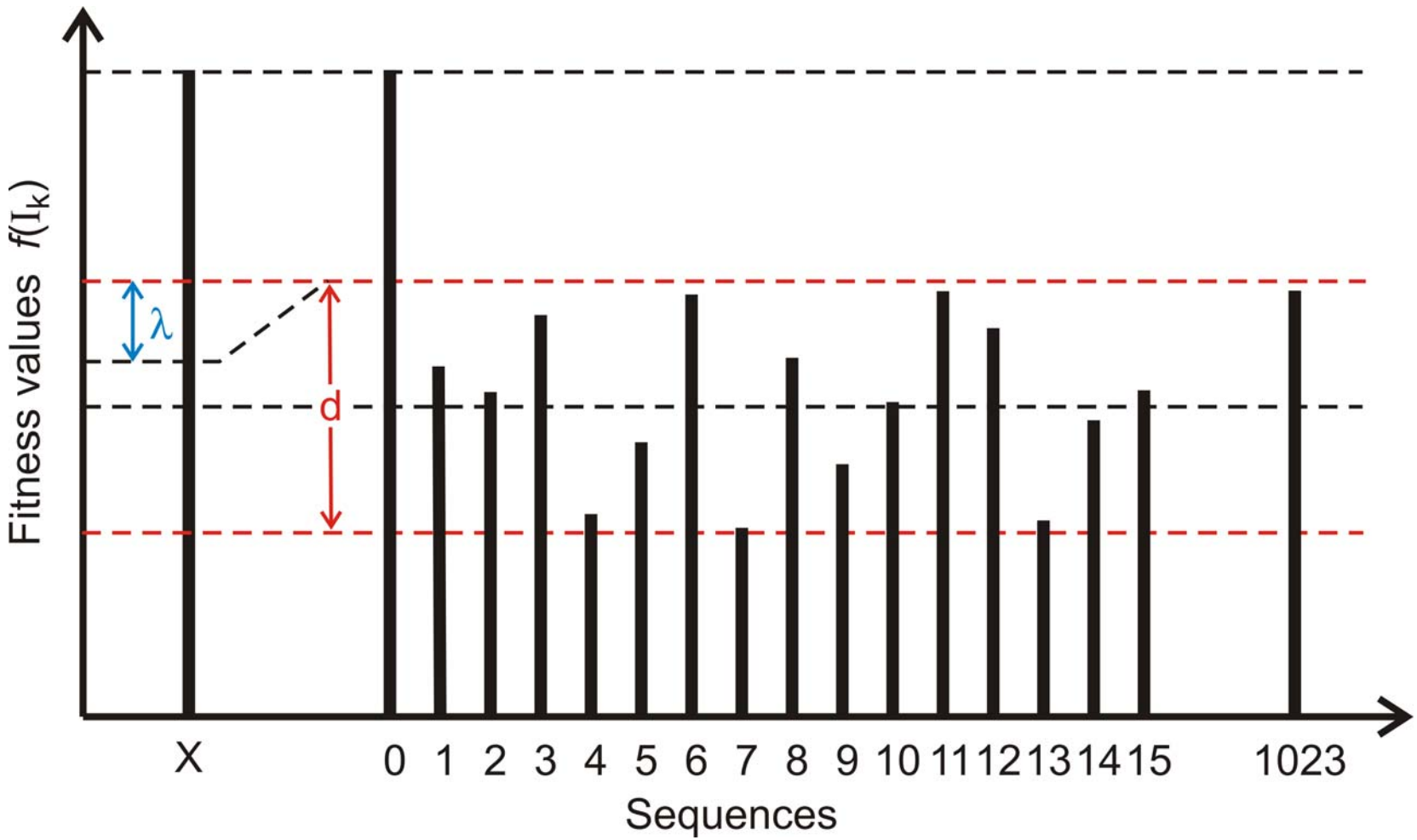
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

$$\lim_{p\to 0} x_1(p) = 0, \lim_{p\to 0} x_2(p) = 1$$

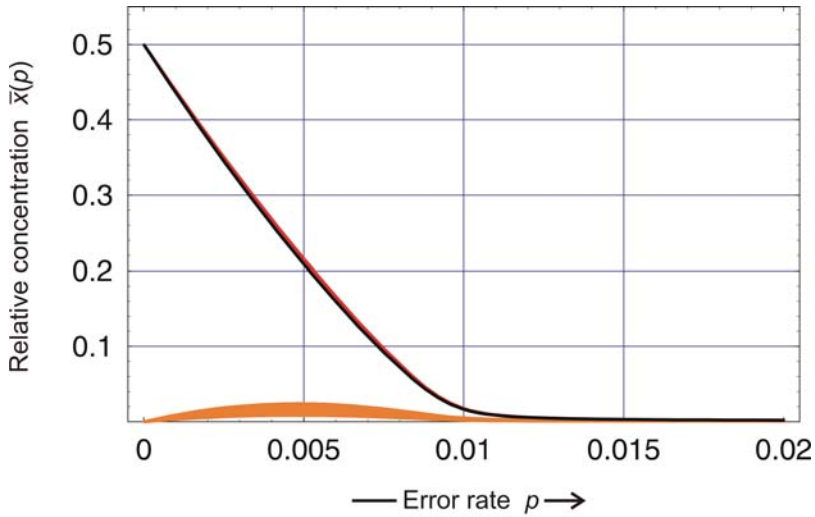
Pairs of neutral sequences in replication networks

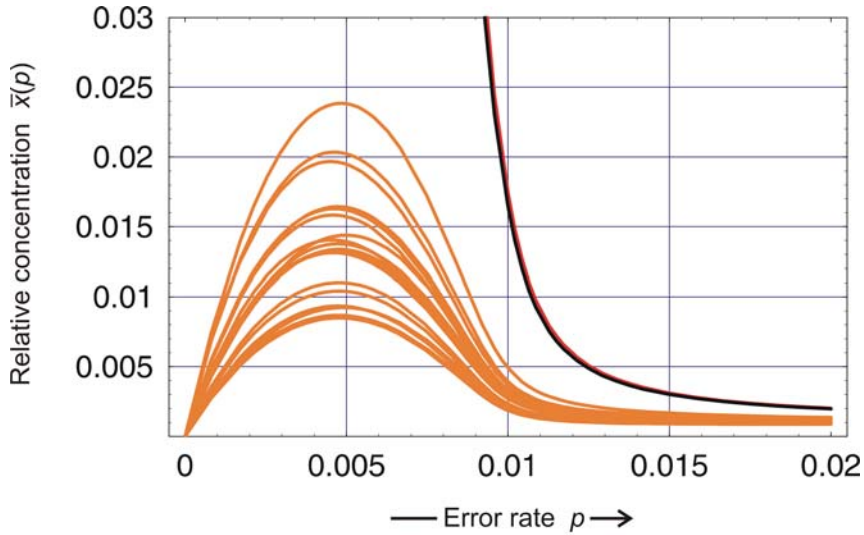
Random fixation in the sense of Motoo Kimura

P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650



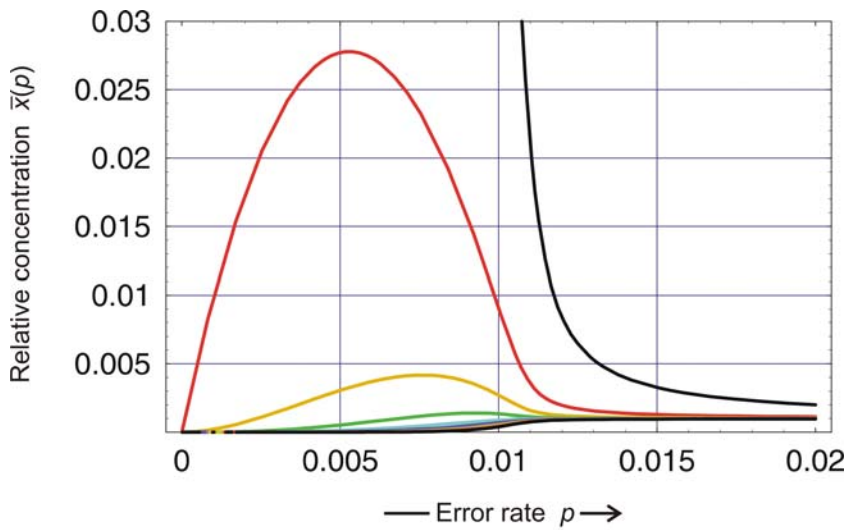
A fitness landscape including neutrality





Neutral network

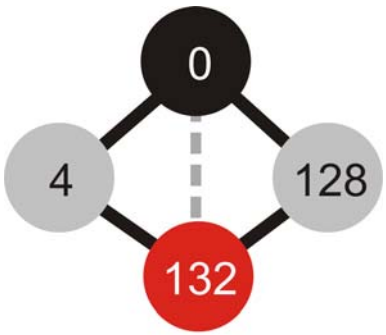
$\lambda = 0.01,\ s = 367$



Neutral network: Individual sequences

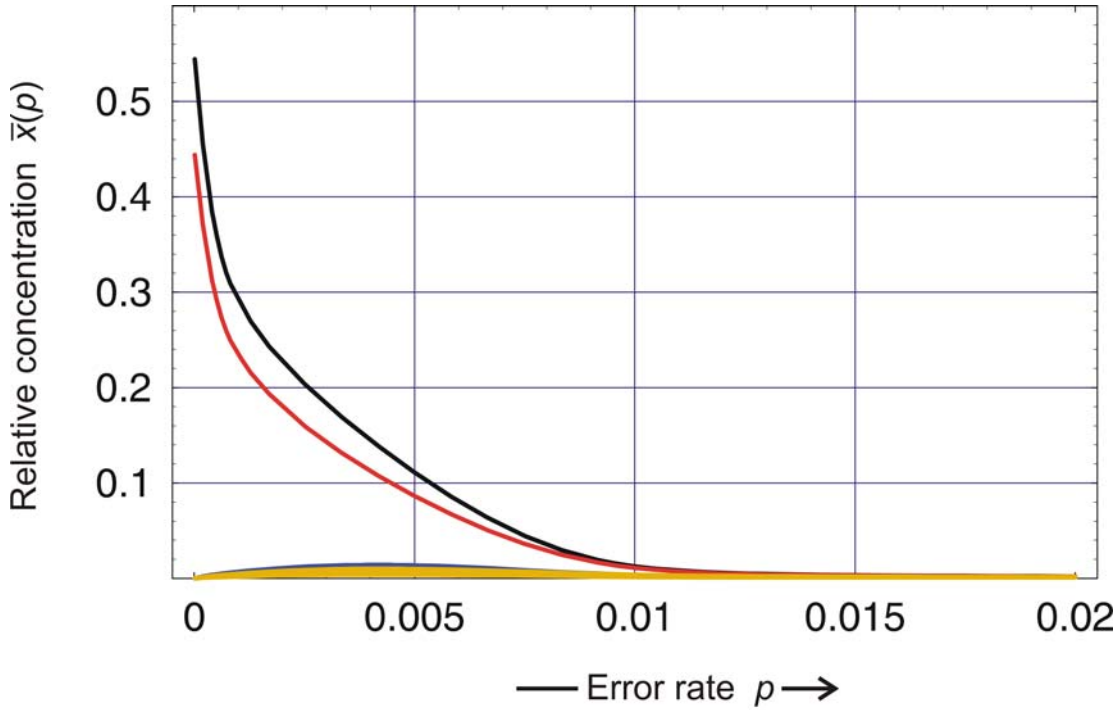
$n = 10,\ \sigma = 1.1,\ d = 1.0$

Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 1$.



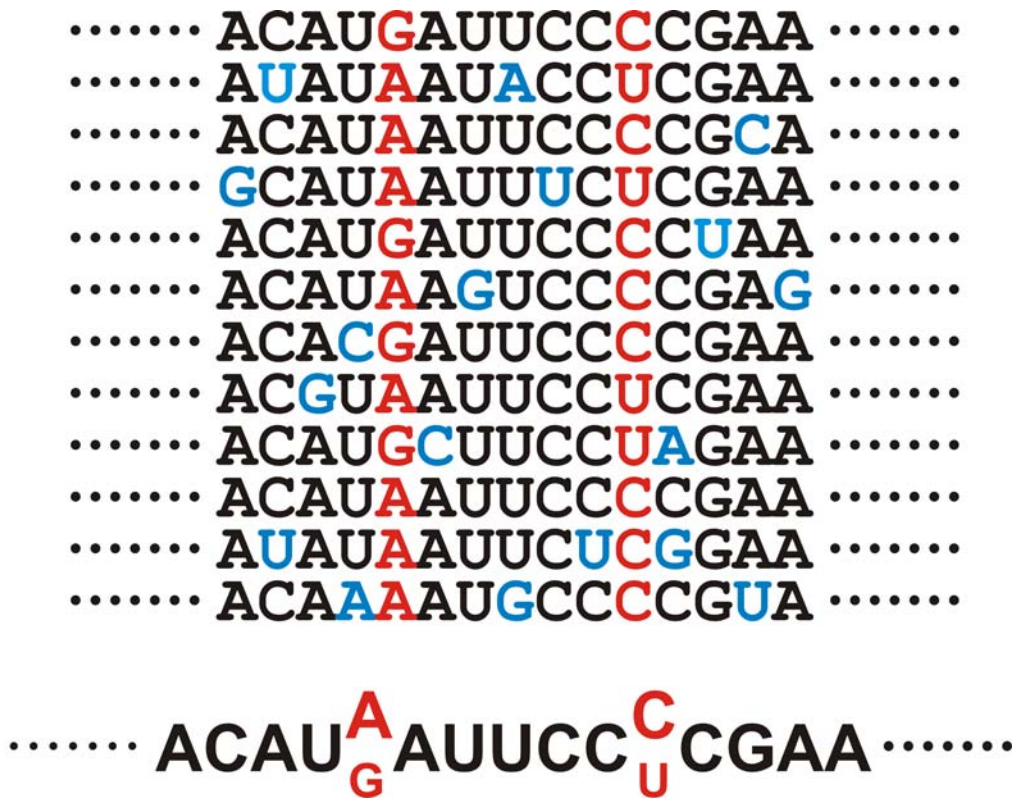
Neutral network

$\lambda = 0.01,\ s = 877$

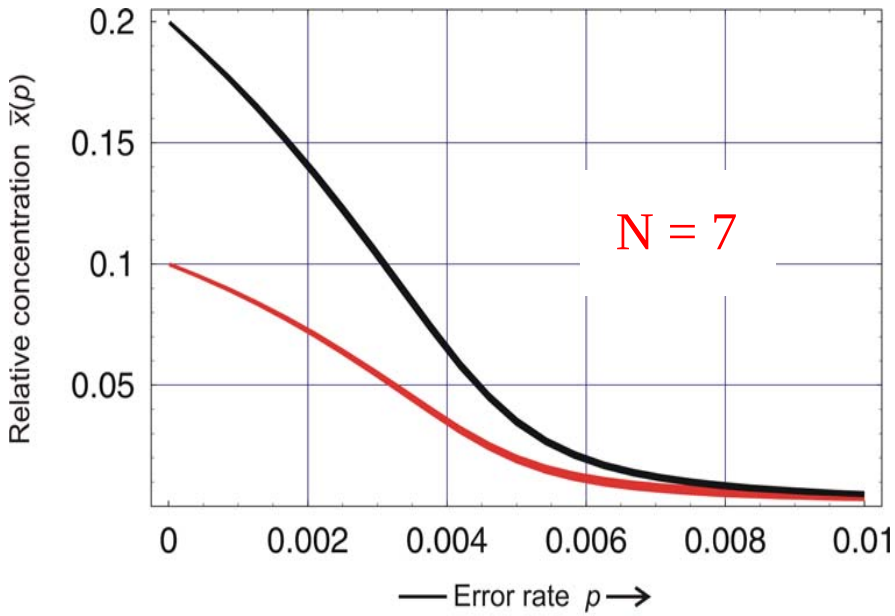


Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0

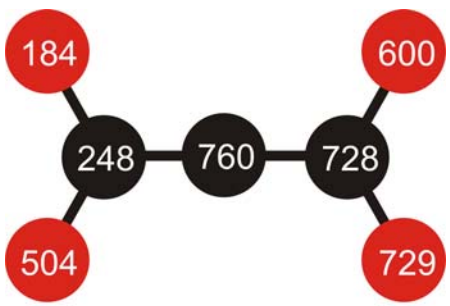


Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 2$.



Perturbation matrix W

$$W = \begin{pmatrix} f & 0 & \varepsilon & 0 & 0 & 0 & 0 \\ 0 & f & \varepsilon & 0 & 0 & 0 & 0 \\ \varepsilon & \varepsilon & f & \varepsilon & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & f & \varepsilon & 0 & 0 \\ 0 & 0 & 0 & \varepsilon & f & \varepsilon & \varepsilon \\ 0 & 0 & 0 & 0 & \varepsilon & f & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 0 & f \end{pmatrix}$$



Neutral network

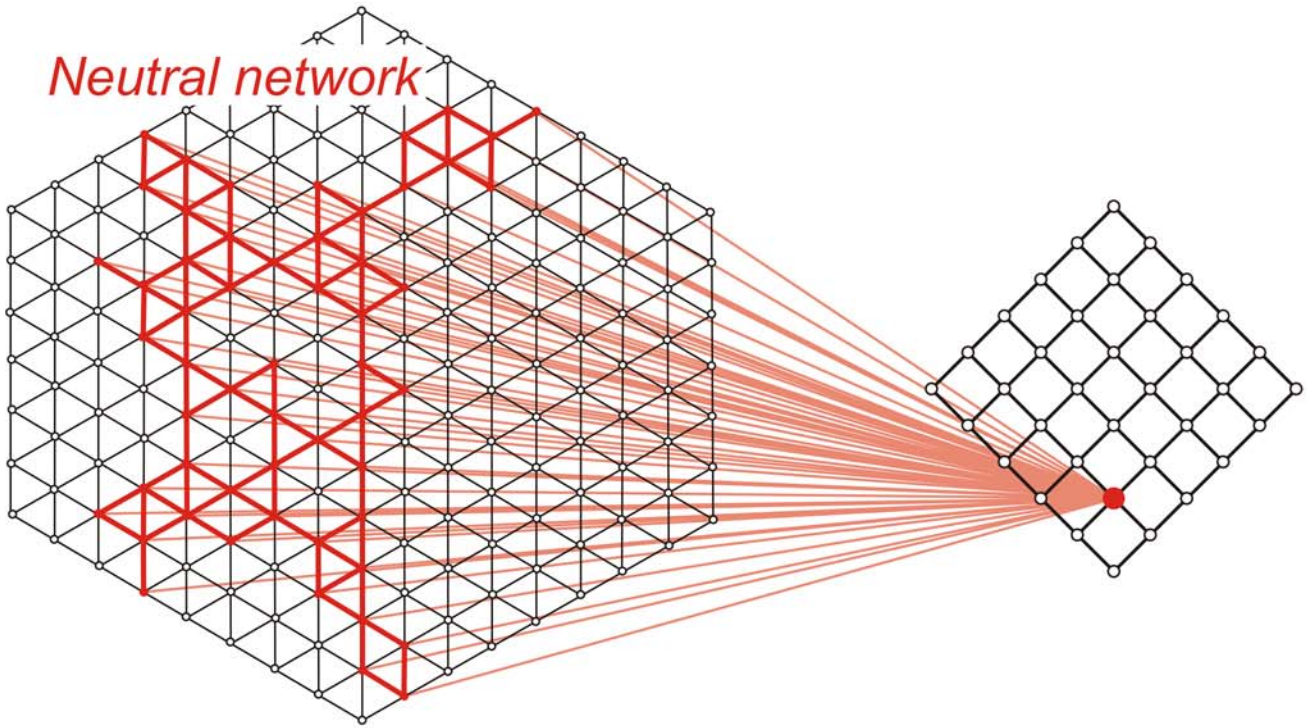
λ = 0.10, s = 229

Adjacency matrix

Largest eigenvector of W

$$\xi_0 = (0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1) .$$

Neutral networks with increasing λ: λ = 0.10, s = 229



many genotypes ⇒ one phenotype

Space of genotypes: $\mathcal{I} = \{I_1, I_2, I_3, I_4, \ldots, I_N\}$; Hamming metric

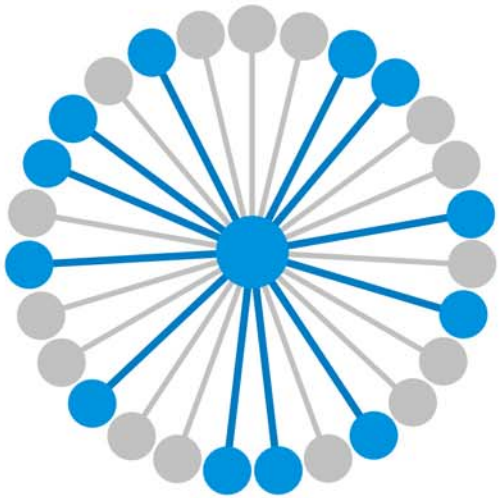
Space of phenotypes: $S = \{S_1, S_2, S_3, S_4, \ldots, S_M\}$; metric (not required)

$$N >> M$$

$$\psi(I_j) = S_k$$

$$\mathbf{G}_k = \psi^{-1}(S_k) \cup \{ I_j \mid \psi(I_j) = S_k \}$$

A mapping ψ and its inversion



$$G_k = \psi^{-1}(S_k) \doteq \{ I_j \mid \psi(I_j) = S_k \}$$

$$\bar{\lambda}_k = \frac{\sum_{j \in |G_k|} \lambda_j(k)}{|G_k|}$$

$\lambda_j = 12 / 27 = 0.444$

Alphabet size κ :

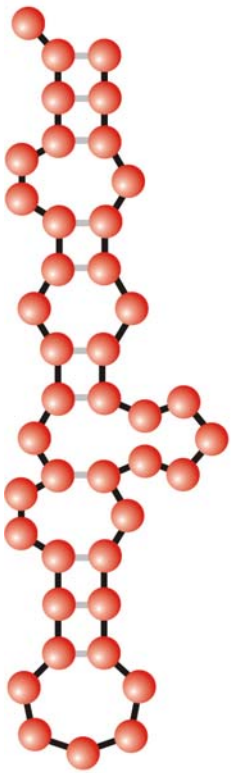
κ	λ_{cr}	
2	0.5	**AU,GC,DU**
3	0.423	**AUG , UGC**
4	0.370	**AUGC**

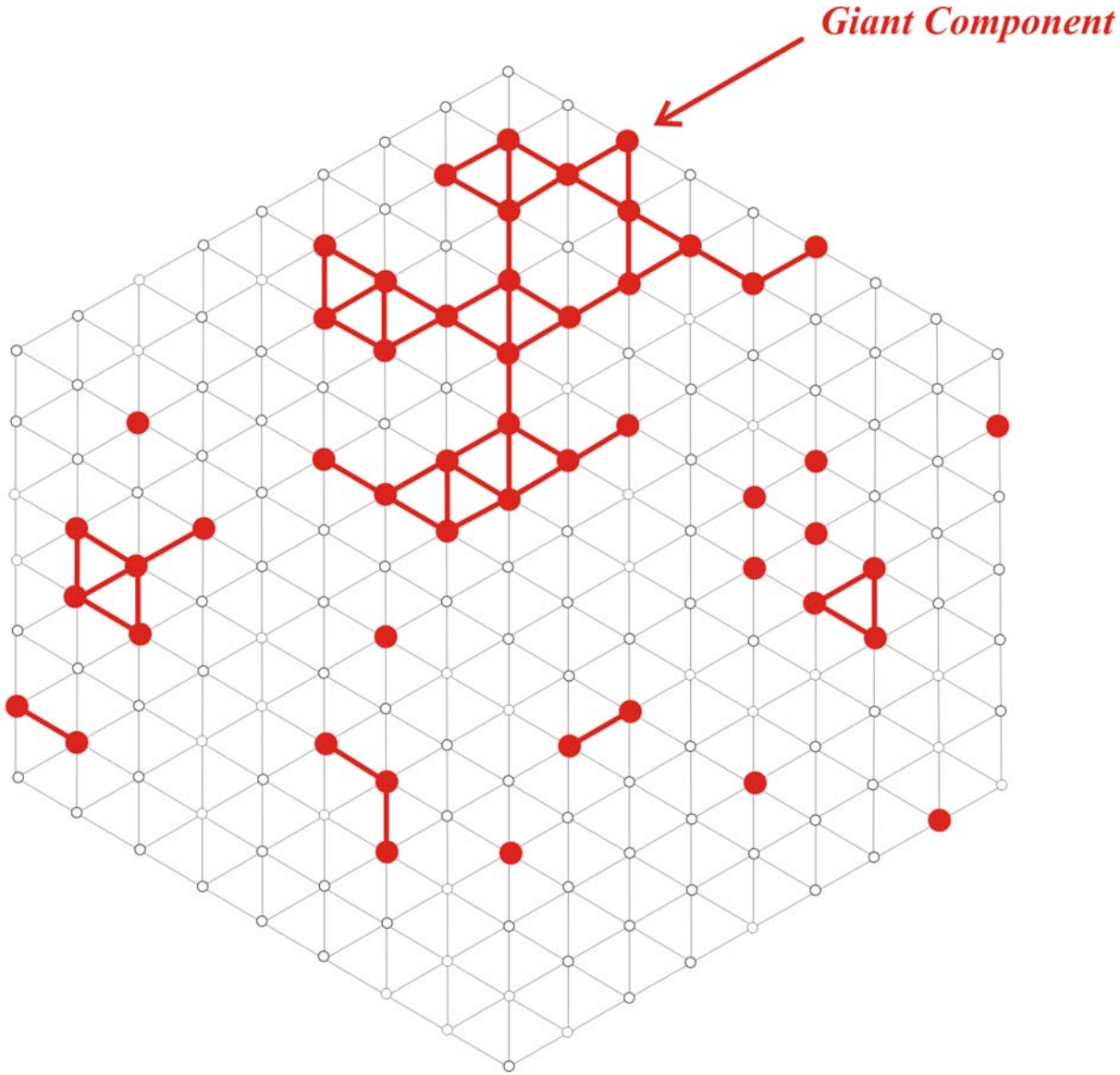
$\bar{\lambda}_k > \lambda_{cr}$ .... network G_k is connected

$\bar{\lambda}_k < \lambda_{cr}$ .... network G_k is **not** connected

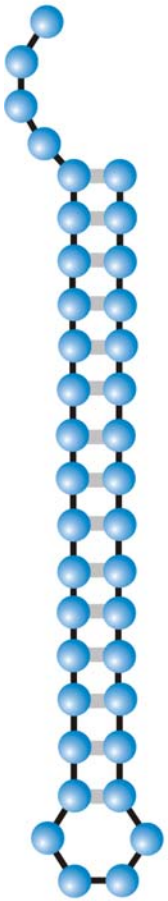
Connectivity threshold: $\lambda_{cr} = 1 - \kappa^{-1/(\kappa-1)}$

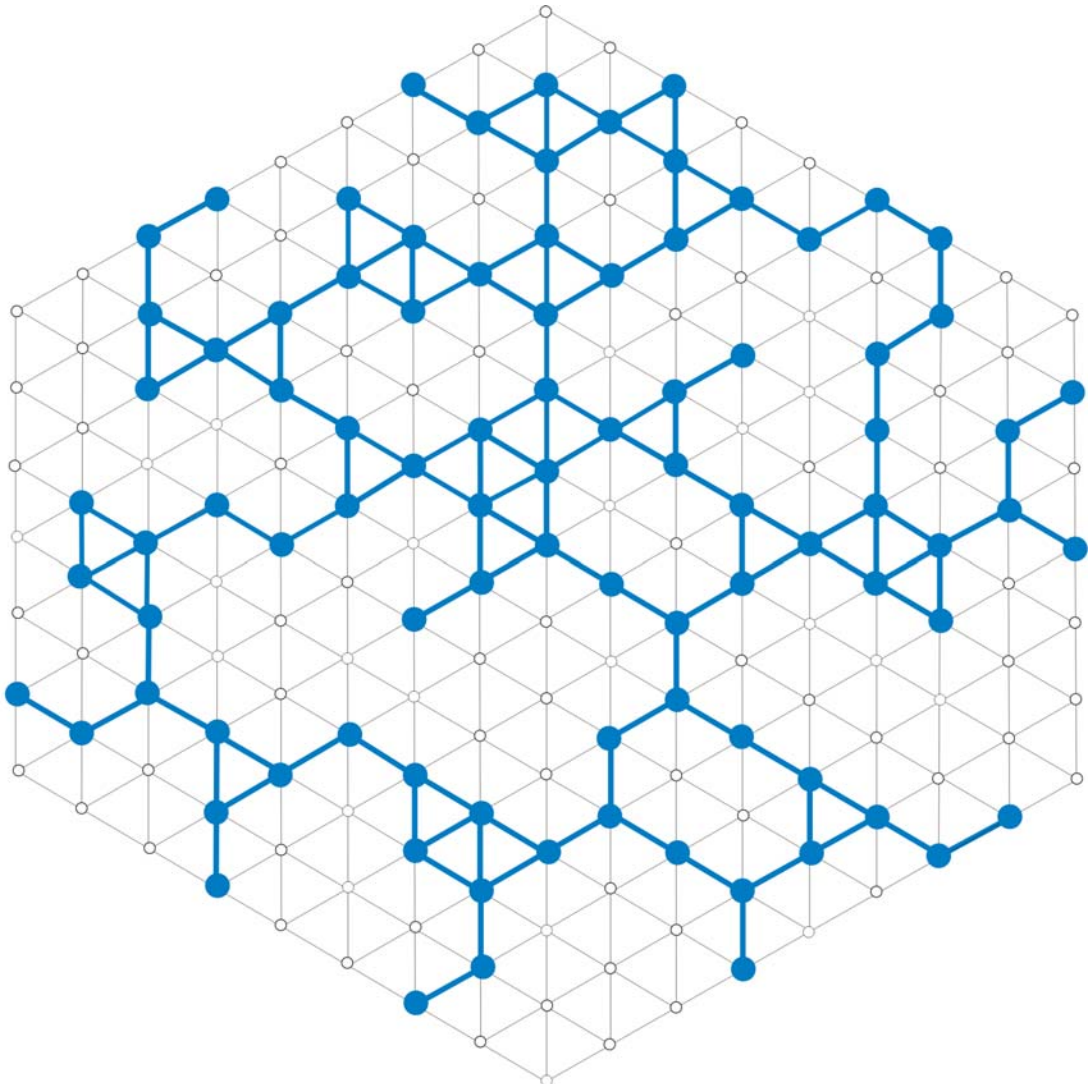
Degree of neutrality of neutral networks and the connectivity threshold





A multi-component neutral network formed by a rare structure: $\lambda < \lambda_{cr}$





A connected neutral network formed by a common structure: $\lambda > \lambda_{cr}$

INSTITUTE OF PHYSICS PUBLISHING REPORTS ON PROGRESS IN PHYSICS

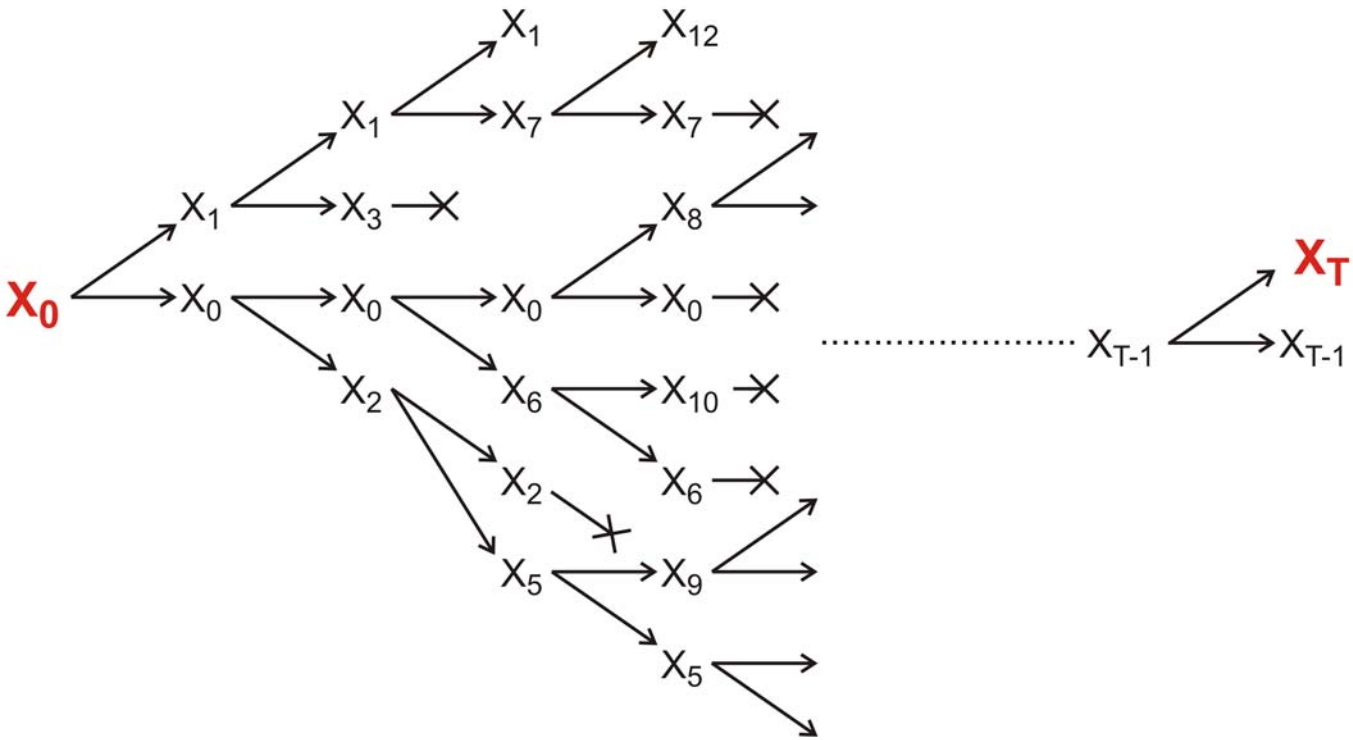
Rep. Prog. Phys. **69** (2006) 1419–1477 doi:10.1088/0034-4885/69/5/R04

Prediction of RNA secondary structures: from theory to models and real molecules

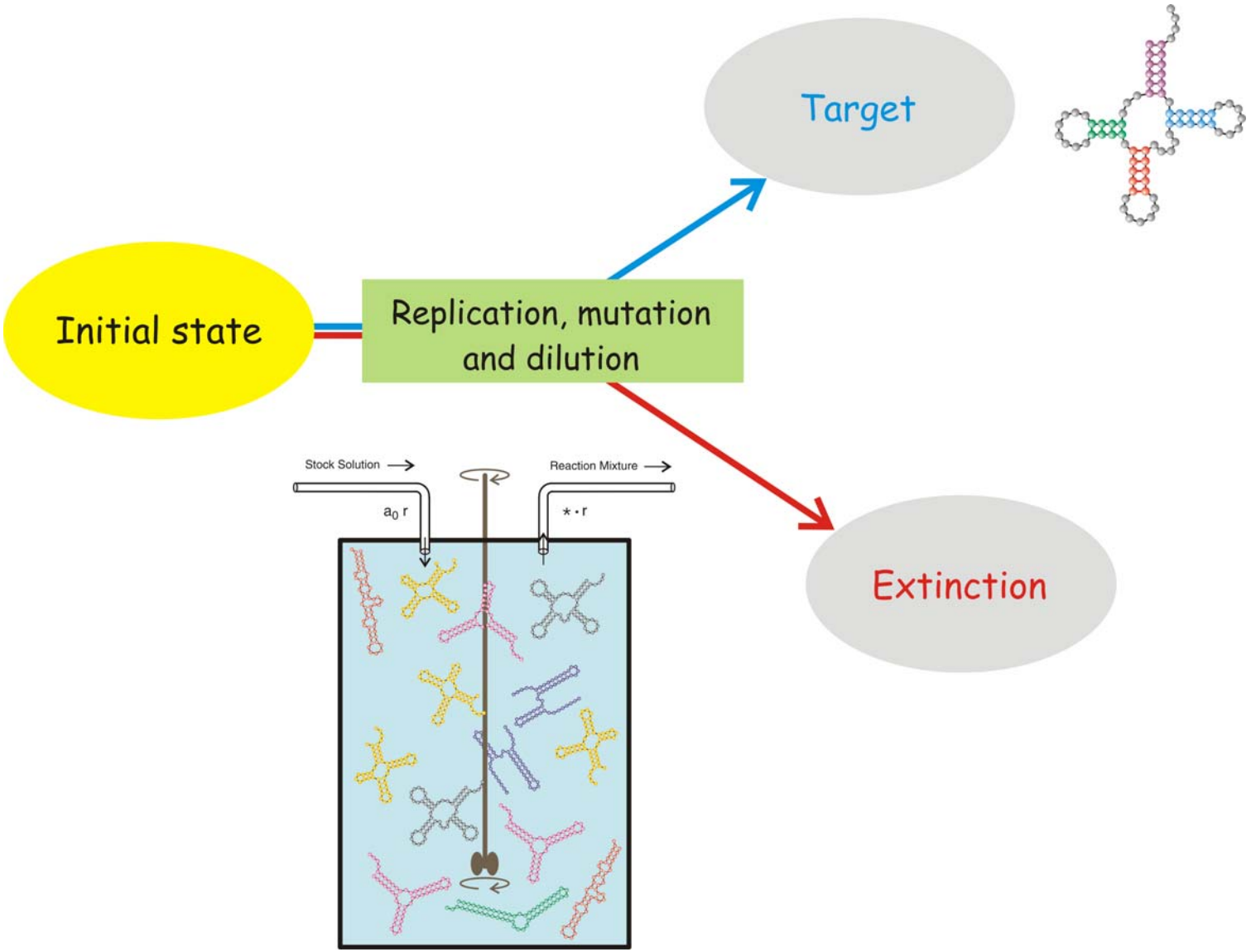
Peter Schuster[1,2]

[1]Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria
[2]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

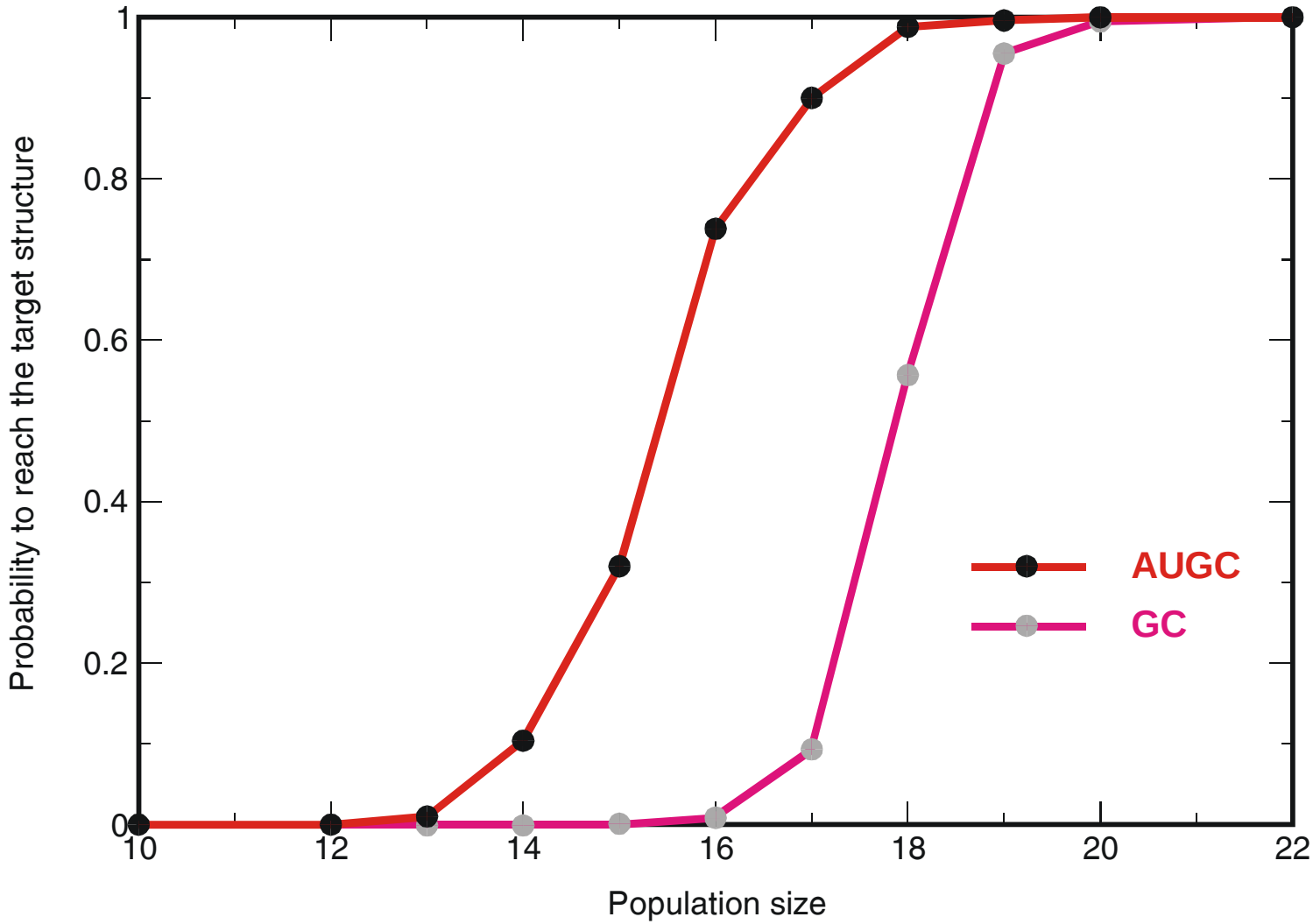
E-mail: pks@tbi.univie.ac.at



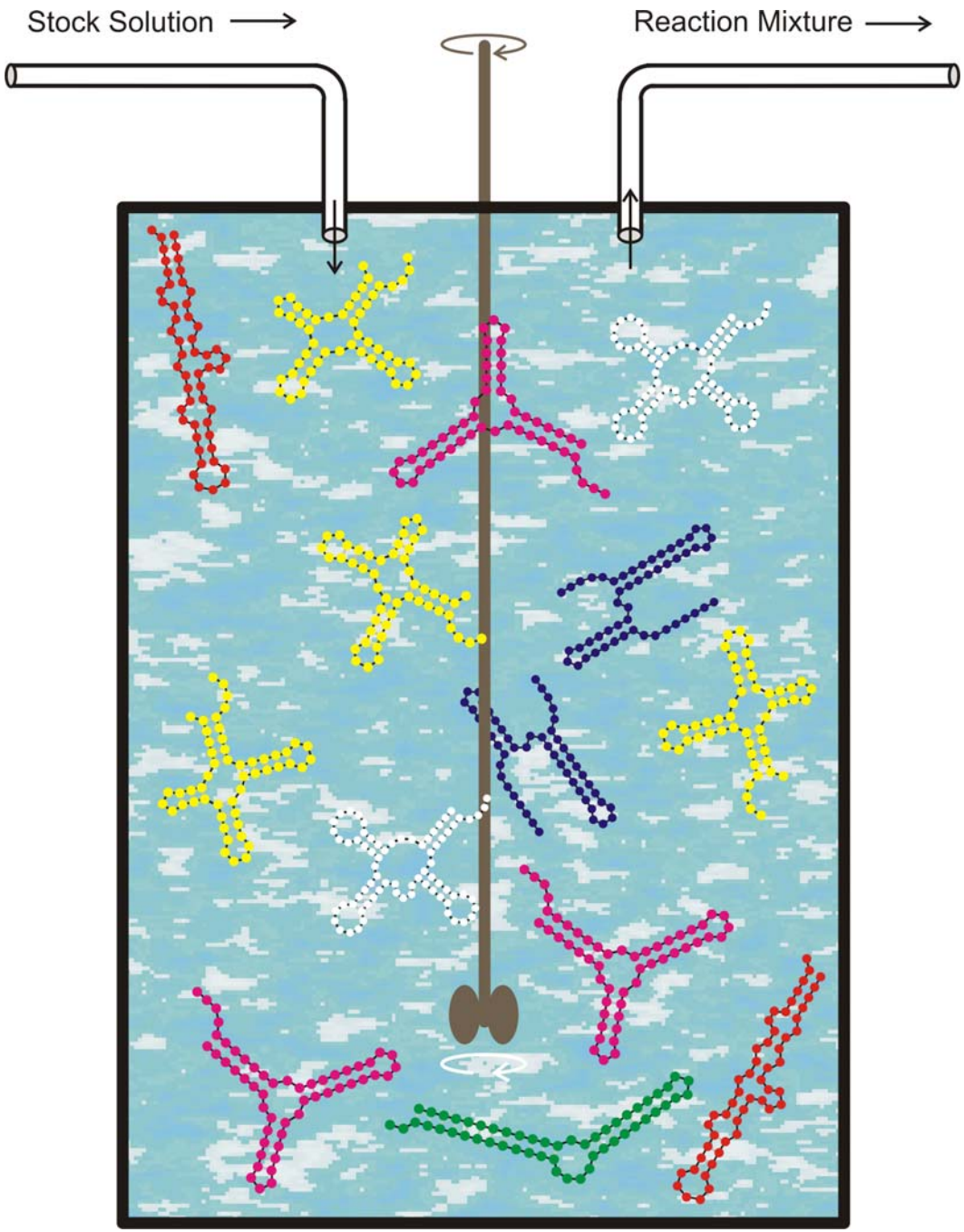
Evolution of RNA molecules as a Markow process



Replication in the flow reactor as a stochastic process with two absorbing barriers



Probability of a single trajectory to reach the target structure



Computer simulation using Gillespie's algorithm:

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***

Evolution *in silico*

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

random individuals. The primer pair used for genomic DNA amplification is 5′-TCTCCCTGGATTCTCATTTA-3′ (forward) and 5′-TCTTTGTCTTCTGTTCCACC-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′-GCATGACCTGCCGGCTAATGGG-3′; reverse, 5′-CTCACGGCTTCTGCATGGTGCTCGGCTGGC-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.

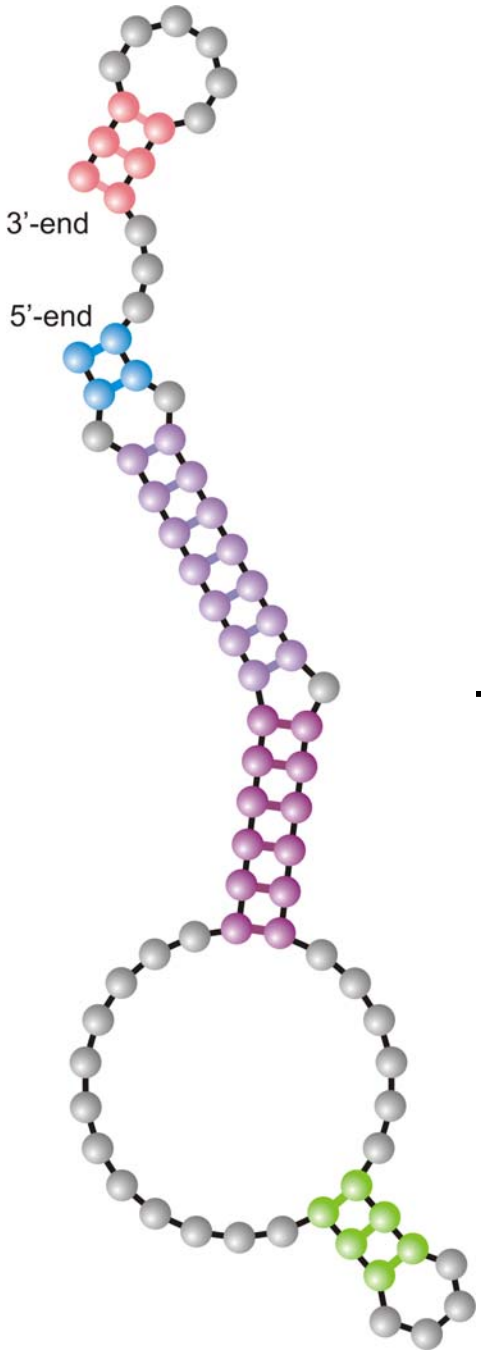
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level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

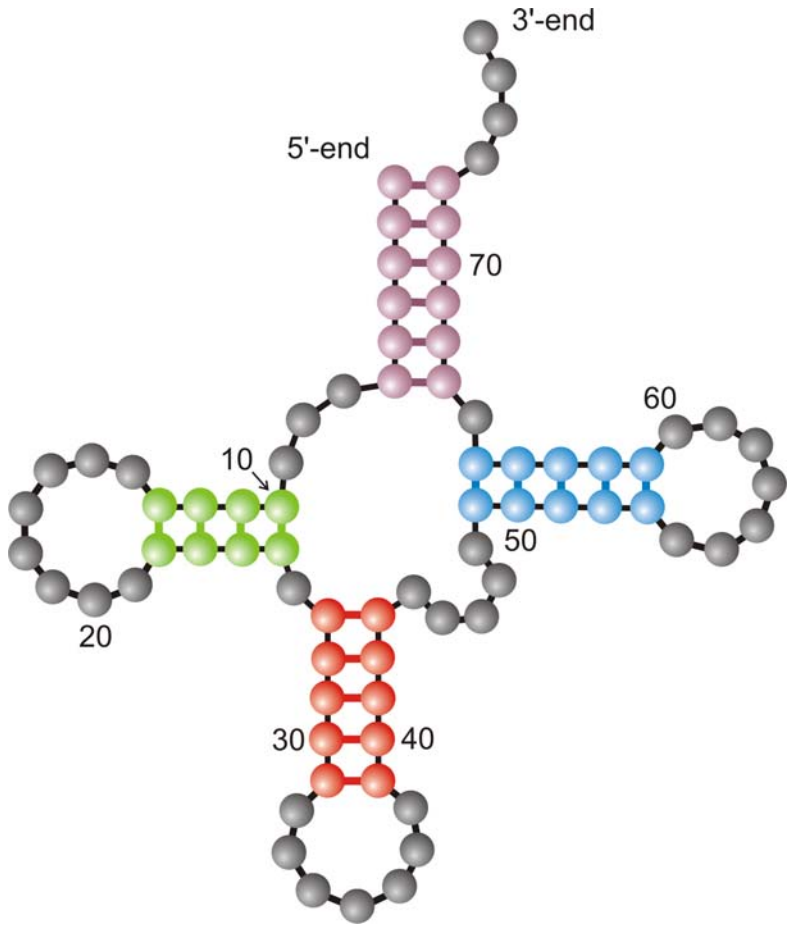
To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (*5*, *6*). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (*4*). Although 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

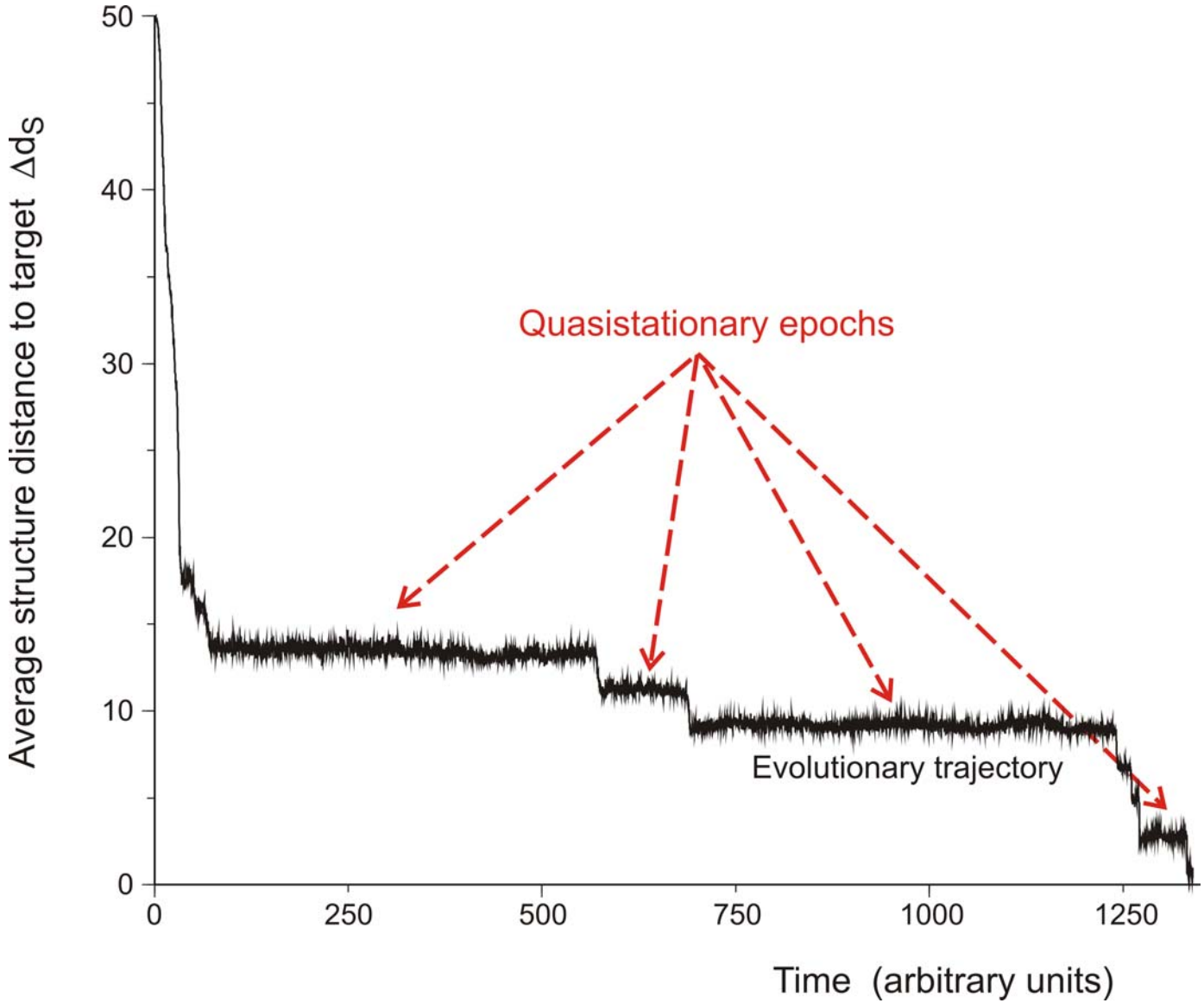
Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.





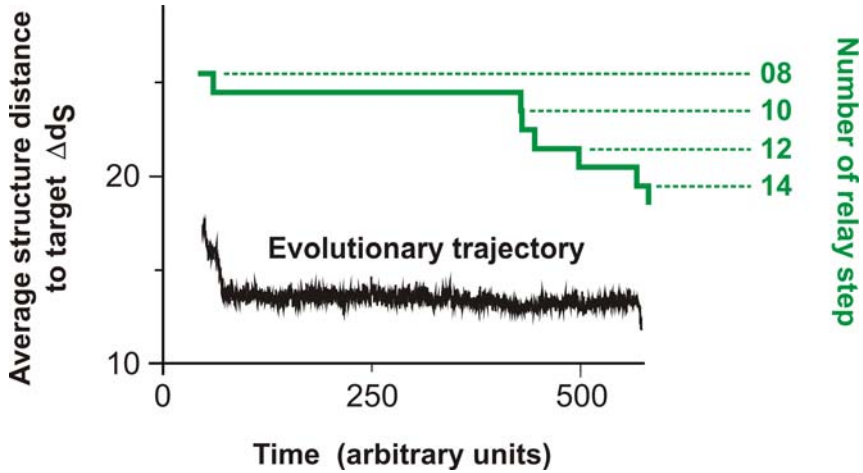
Structure of randomly chosen initial sequence

Phenylalanyl-tRNA as target structure



In silico optimization in the flow reactor: Evolutionary Trajectory

28 neutral point mutations during a long quasi-stationary epoch




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

Transition inducing point mutations change the molecular structure

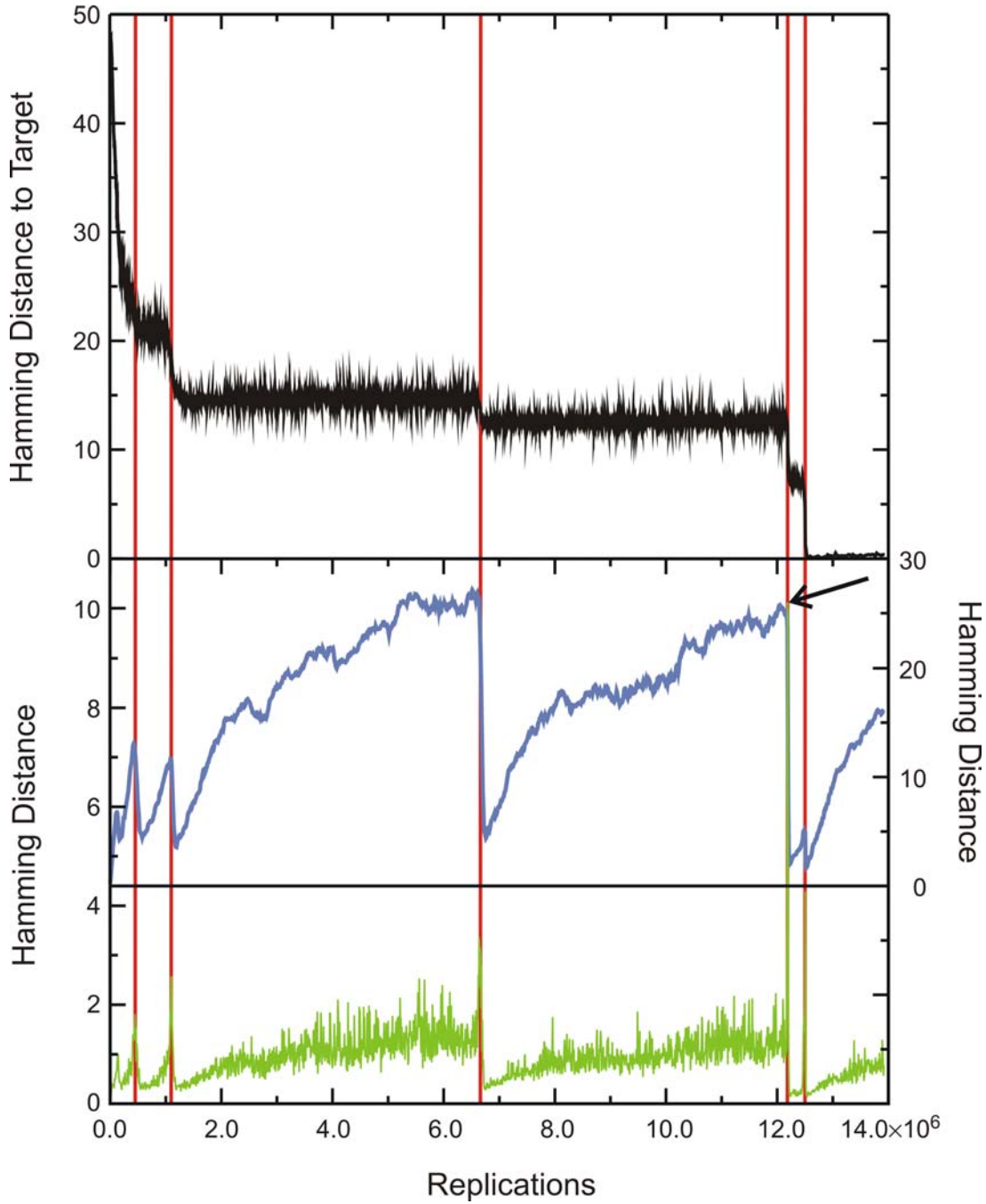
Neutral point mutations leave the molecular structure unchanged

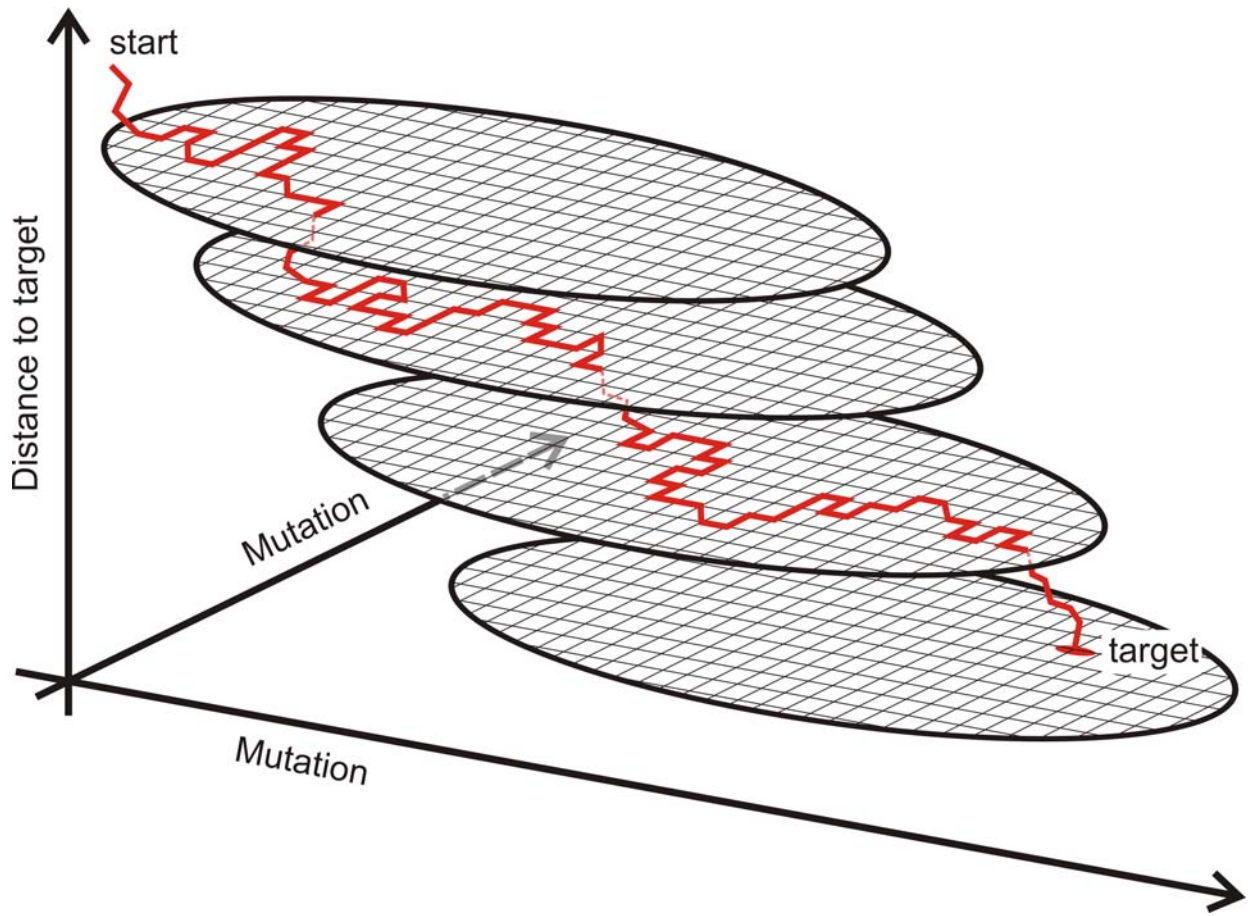
Neutral genotype evolution during phenotypic stasis

Evolutionary trajectory

Spreading of the population on neutral networks

Drift of the population center in sequence space





A sketch of optimization on neutral networks

If Charles Darwin would have written the „Origin"
in mathematical language, how would he have done it?

What did we learn about evolution from *in vitro* experiments?

Quantitative systems biology – A challenge for biologists, chemists, physicists, and mathematicians !

Systems biology or quantitative biology is the chemistry of whole cells and organisms

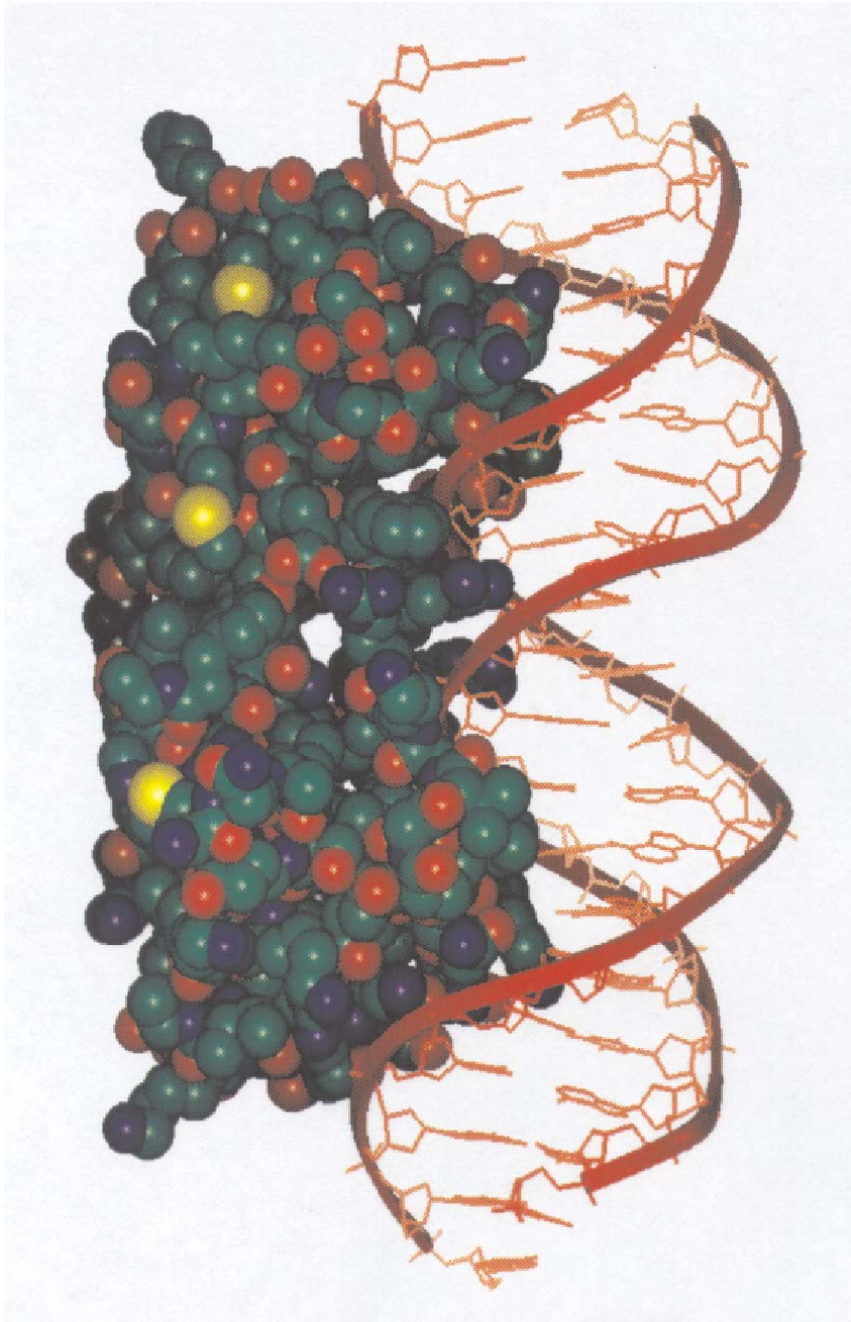
Challenges for theorists and mathematicians:

1. Very large numbers of variables and parameters in ODE modeling

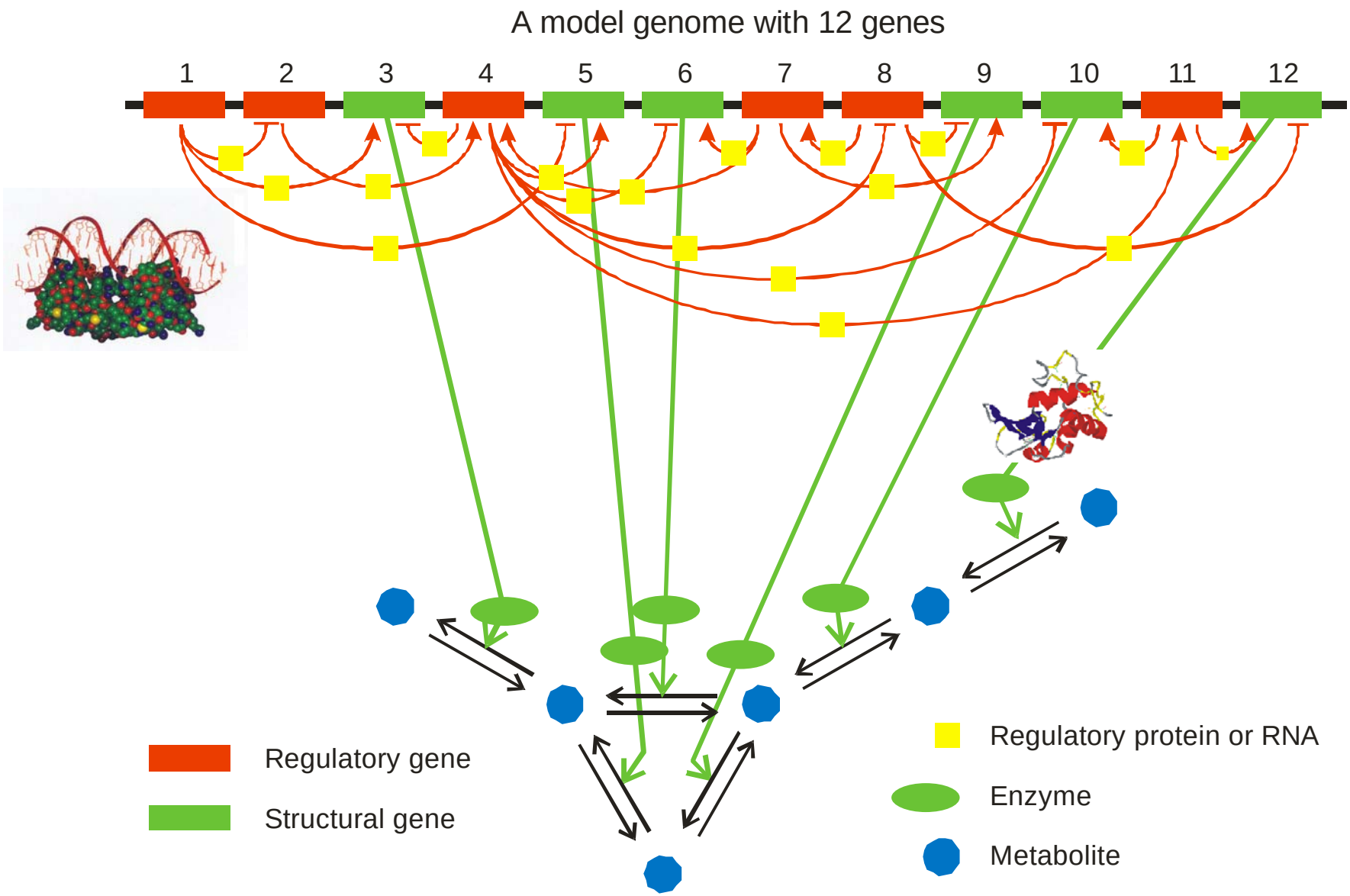
2. Stochastic effects because of very low particle numbers

3. Complex nonlinear reaction networks

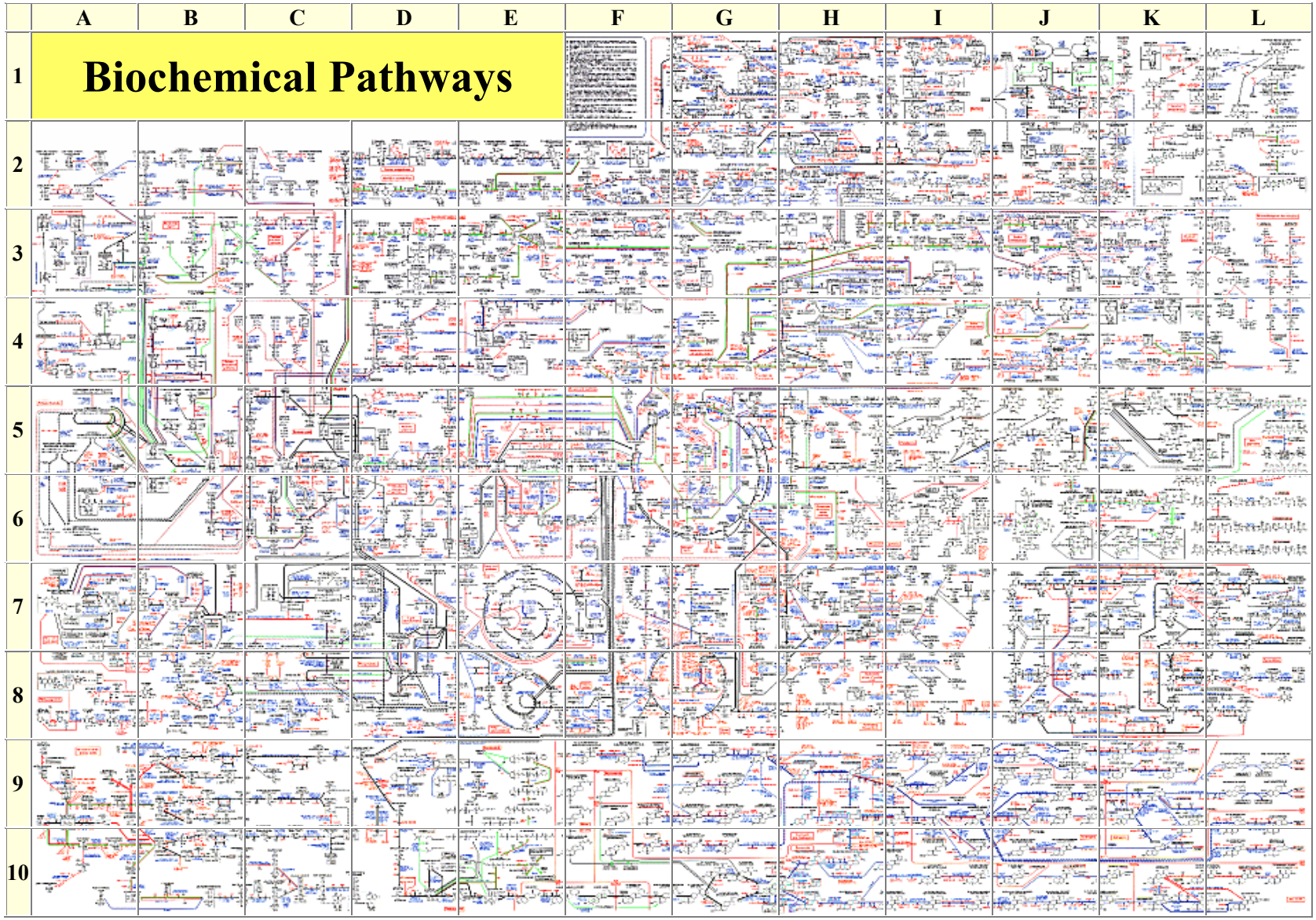
4. Complex spatial structures in specific aggregates and compartments



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



Sketch of a genetic and metabolic network



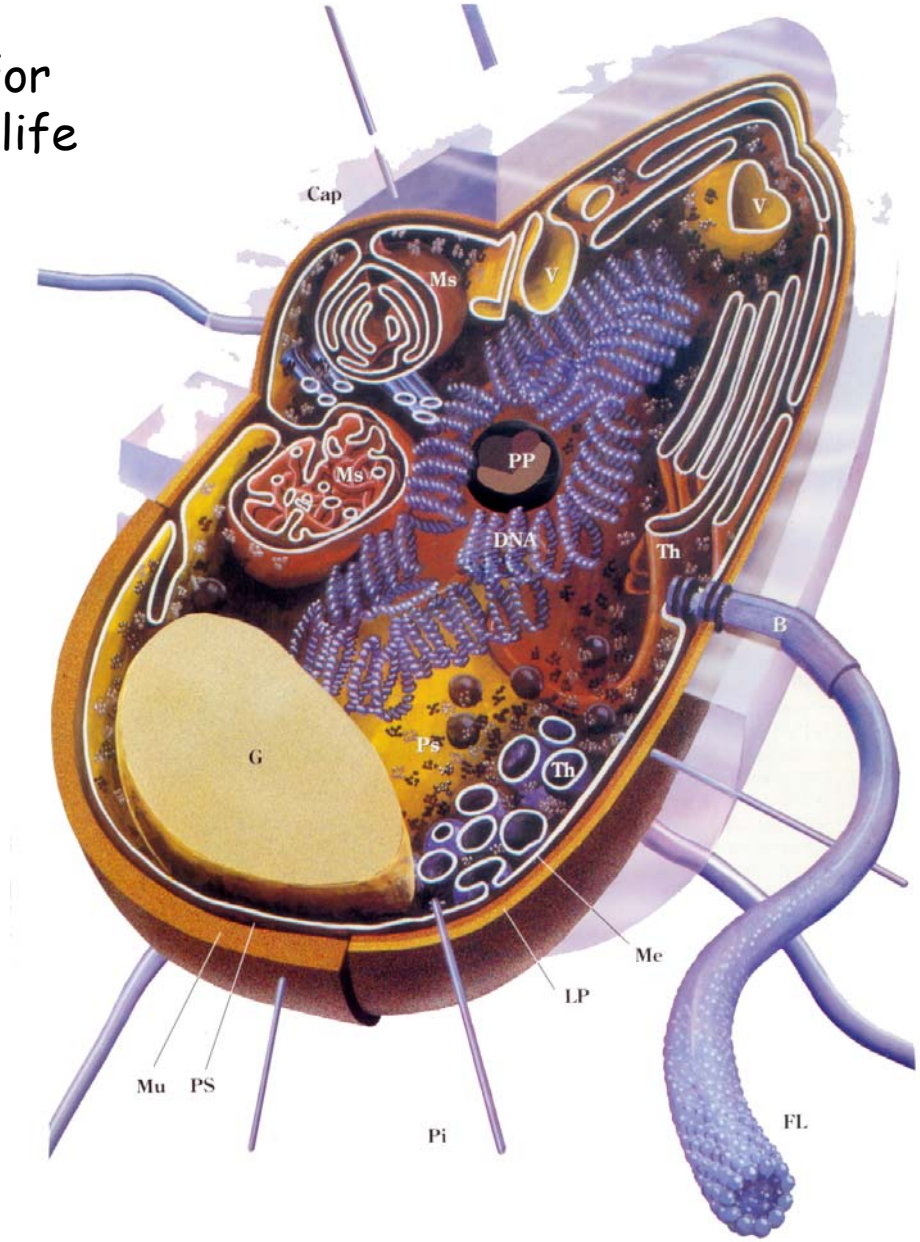
The reaction network of cellular metabolism published by Boehringer-Mannheim.

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

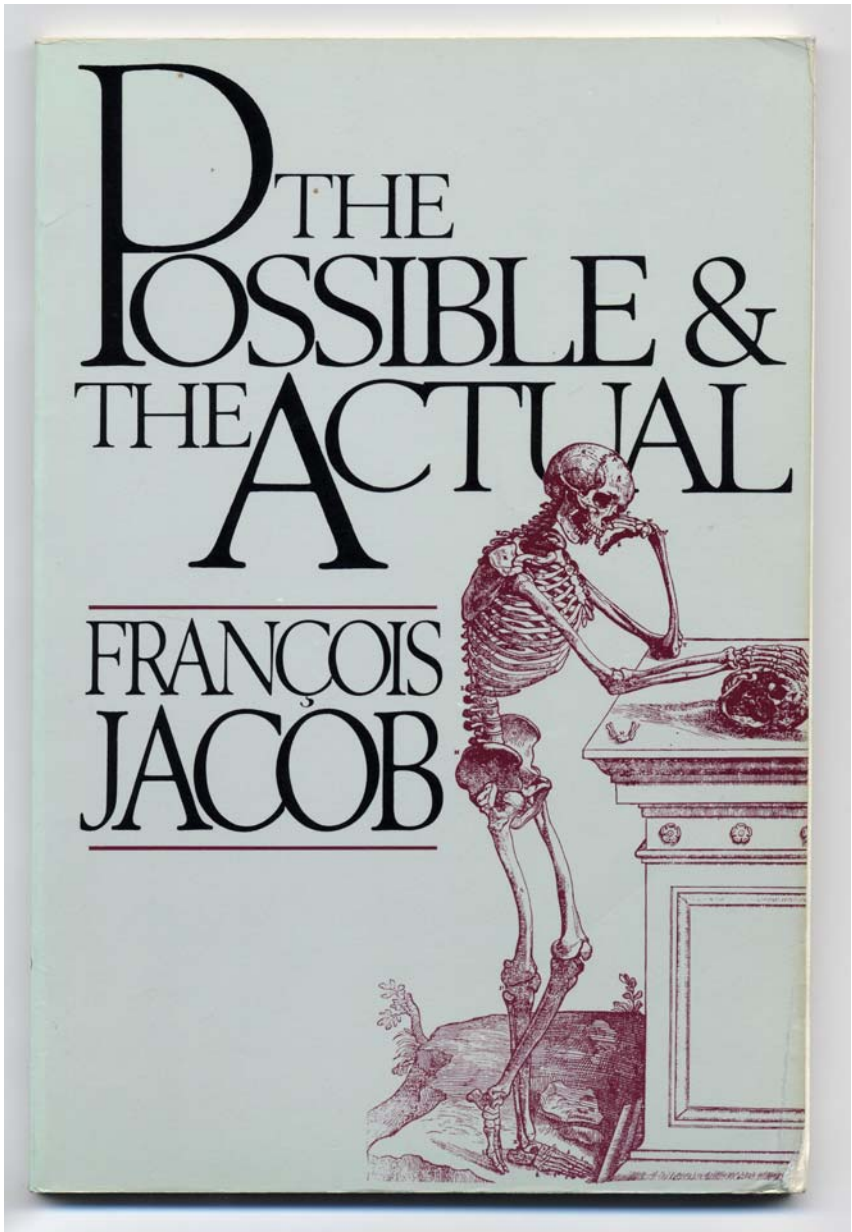
Escherichia coli genome:

4 million nucleotides

4460 genes



The structure of the bacterium *Escherichia coli*



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

François Jacob. *The Possible and the Actual*. Pantheon Books, New York, 1982, and

Evolutionary tinkering. *Science* **196** (1977), 1161-1166.

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.

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.

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

Coworkers

Universität Wien

Karl Sigmund, Universität Wien, AT

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

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

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

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

Walter Fontana, Harvard Medical School, **Martin Nowak**, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

Christian Forst, Los Alamos National Laboratory, NM

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

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

Kurt Grünberger, **Michael Kospach** , **Andreas Wernitznig**, **Stefanie Widder**,
Stefan Wuchty, **Jan Cupal**, **Stefan Bernhart**, **Lukas Endler**, **Ulrike Langhammer**,
Rainer Machne, **Ulrike Mückstein**, **Erich Bornberg-Bauer**,
Universität Wien, AT

Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)
Projects No. 09942, 10578, 11065, 13093
13887, and 14898

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

Austrian Genome Research Program – GEN-AU: Bioinformatics Network (BIN)

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Universität Wien

Thank you for your attention!

Web-Page for further information:

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