Complexity in Evolutionary Processes

Peter Schuster

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

7th Vienna Central European Seminar on
Particle Physics and Quantum Field Theory

Vienna, 26.– 28.11.2010

Web-Page for further information:

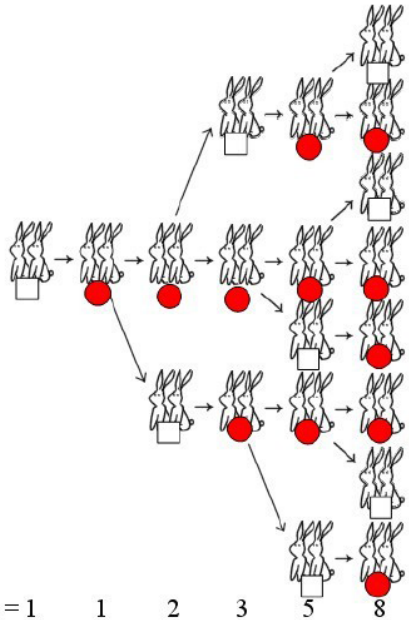
http://www.tbi.univie.ac.at/~pks

1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution

1. **Exponential growth and selection**
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution

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

Thomas Robert Malthus
1766 – 1834

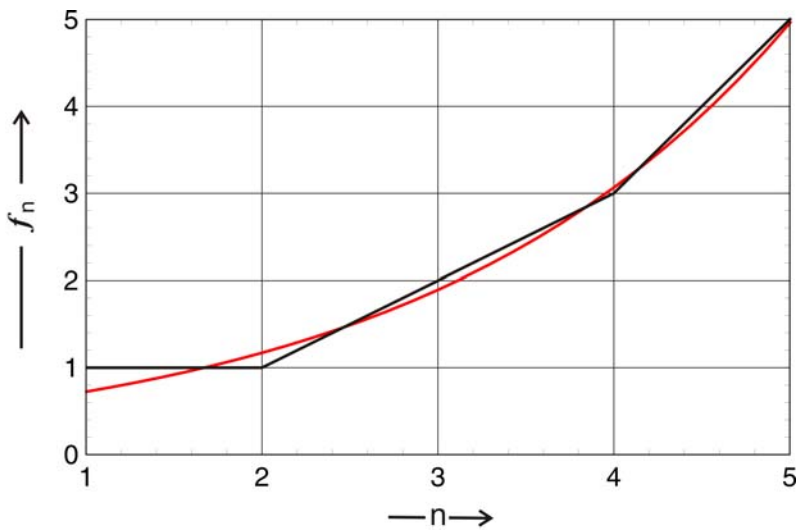


Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

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

geometric progression

exponential growth



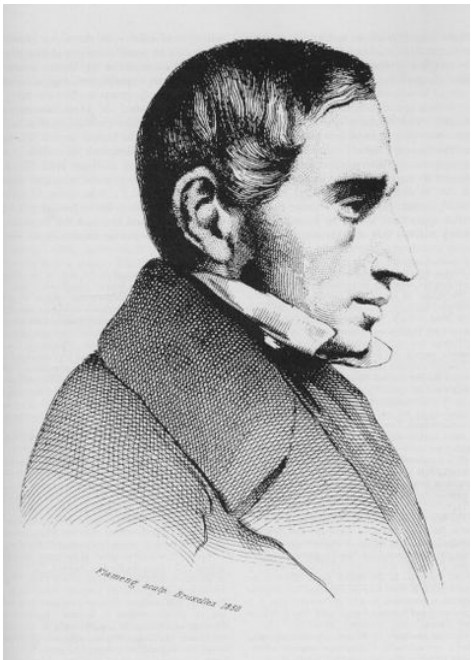
$$f_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

The history of
exponential growth

Three necessary conditions for Darwinian evolution are:

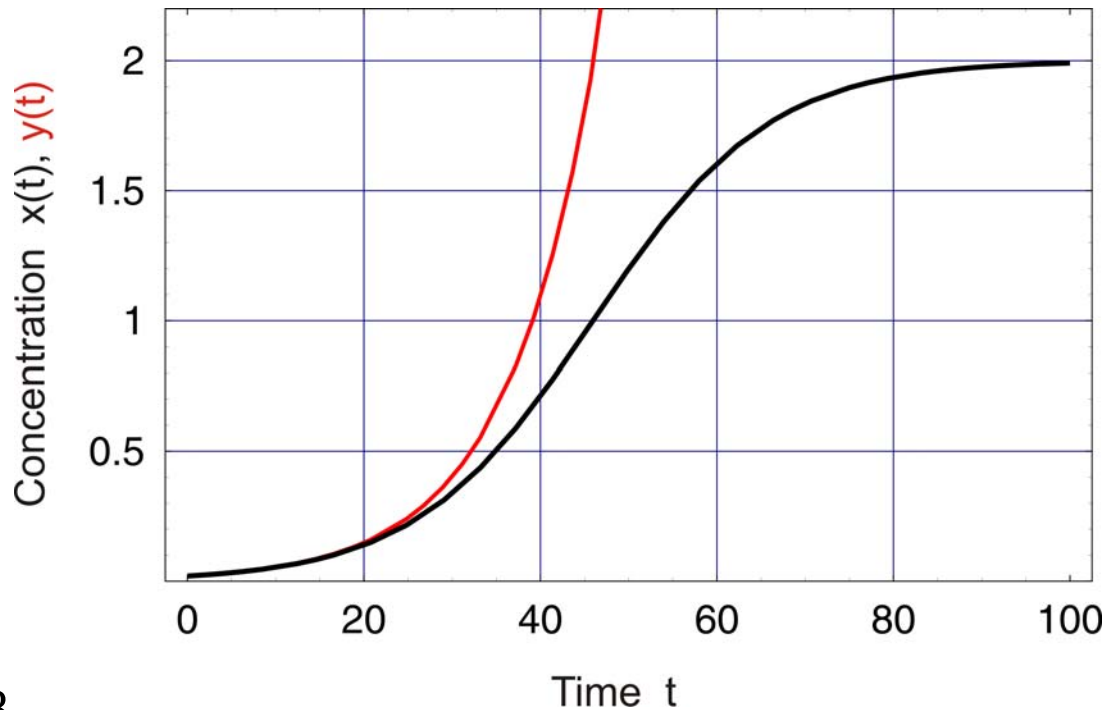
1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

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

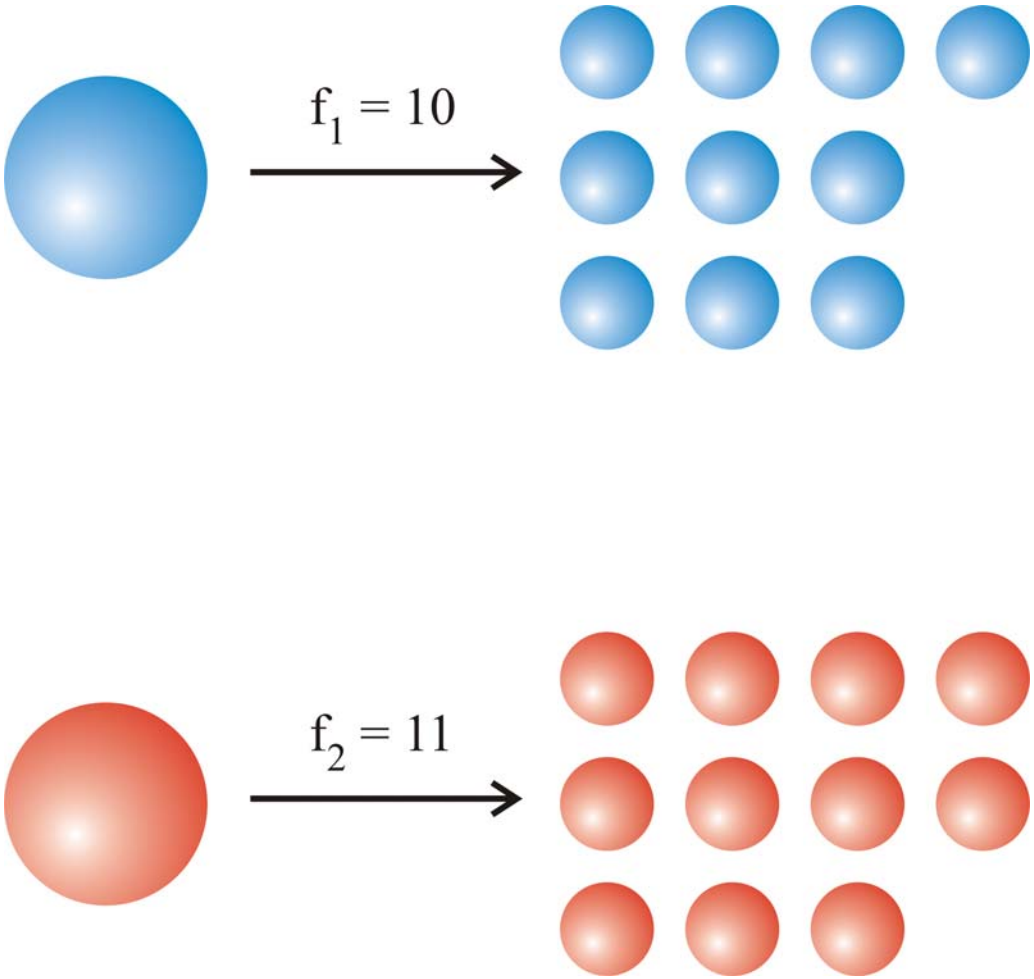


Pierre-François Verhulst,
1804-1849

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

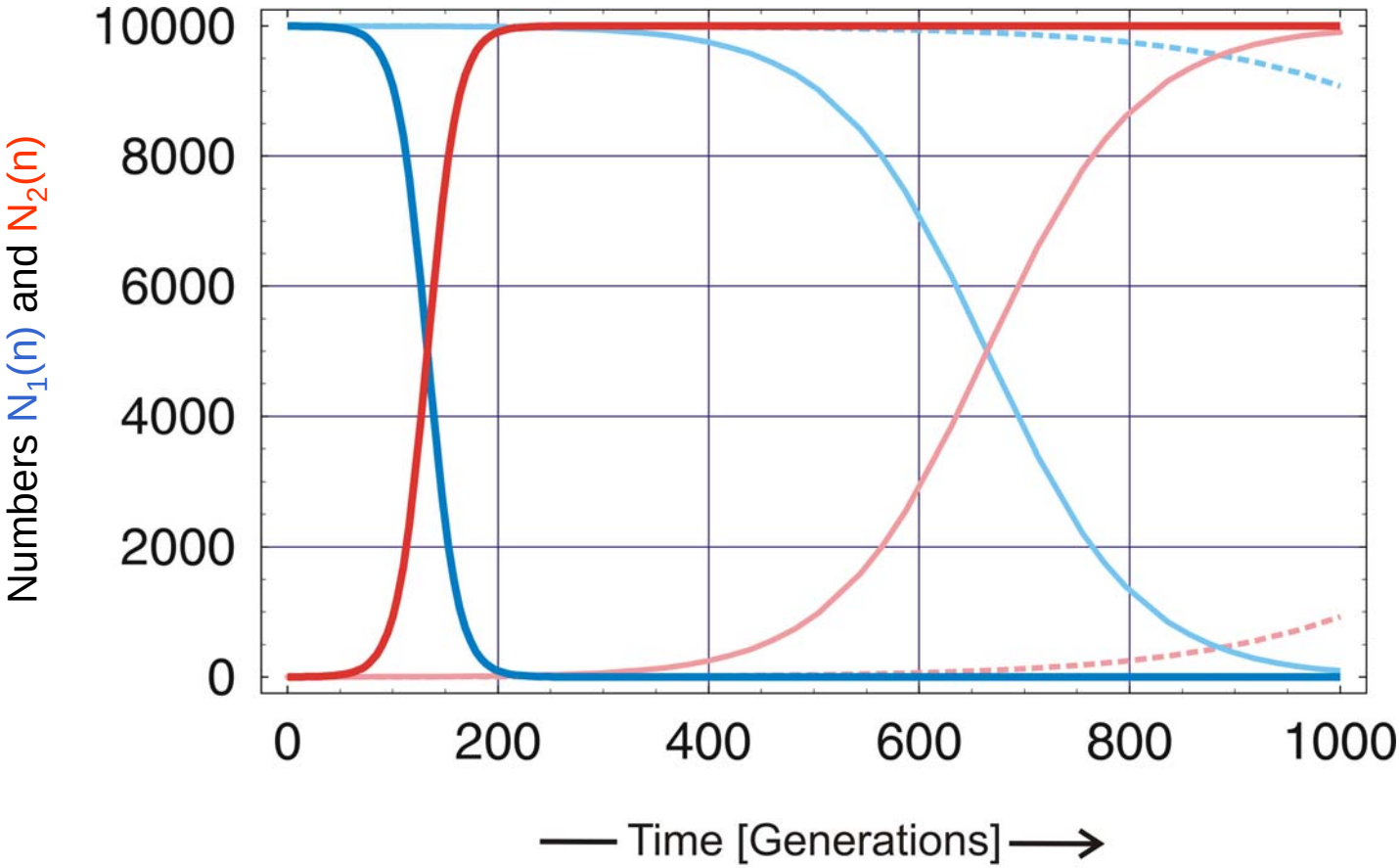


The logistic equation, 1828



$$s = \frac{f_2 - f_1}{f_1} = 0.1$$

Two variants with a mean progeny of ten or eleven descendants



$N_1(0) = 9999$, $N_2(0) = 1$; s = 0.1 , 0.02 , 0.01

Selection of advantageous mutants in populations of N = 10 000 individuals

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

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

$$X_1, X_2, \ldots, X_n: \quad [X_i] = x_i; \quad \sum_{i=1}^{n} x_i = C = 1$$

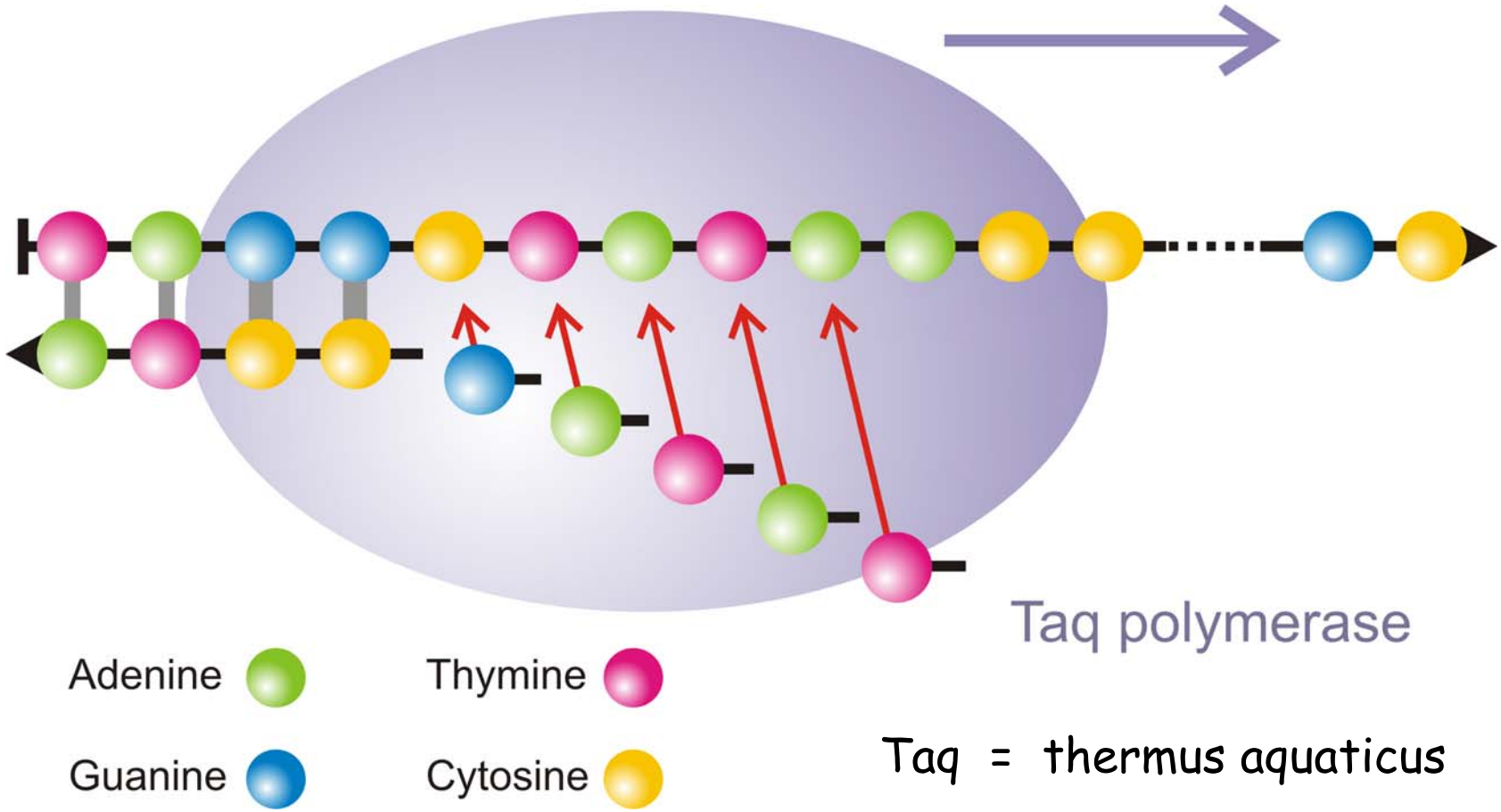
$$\frac{dx_j}{dt} = x_j\left(f_j - \sum_{i=1}^{n} f_i\,x_i\right) = x_j\left(f_j - \Phi\right); \quad \Phi = \sum_{i=1}^{n} f_i\,x_i$$

Darwin

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

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

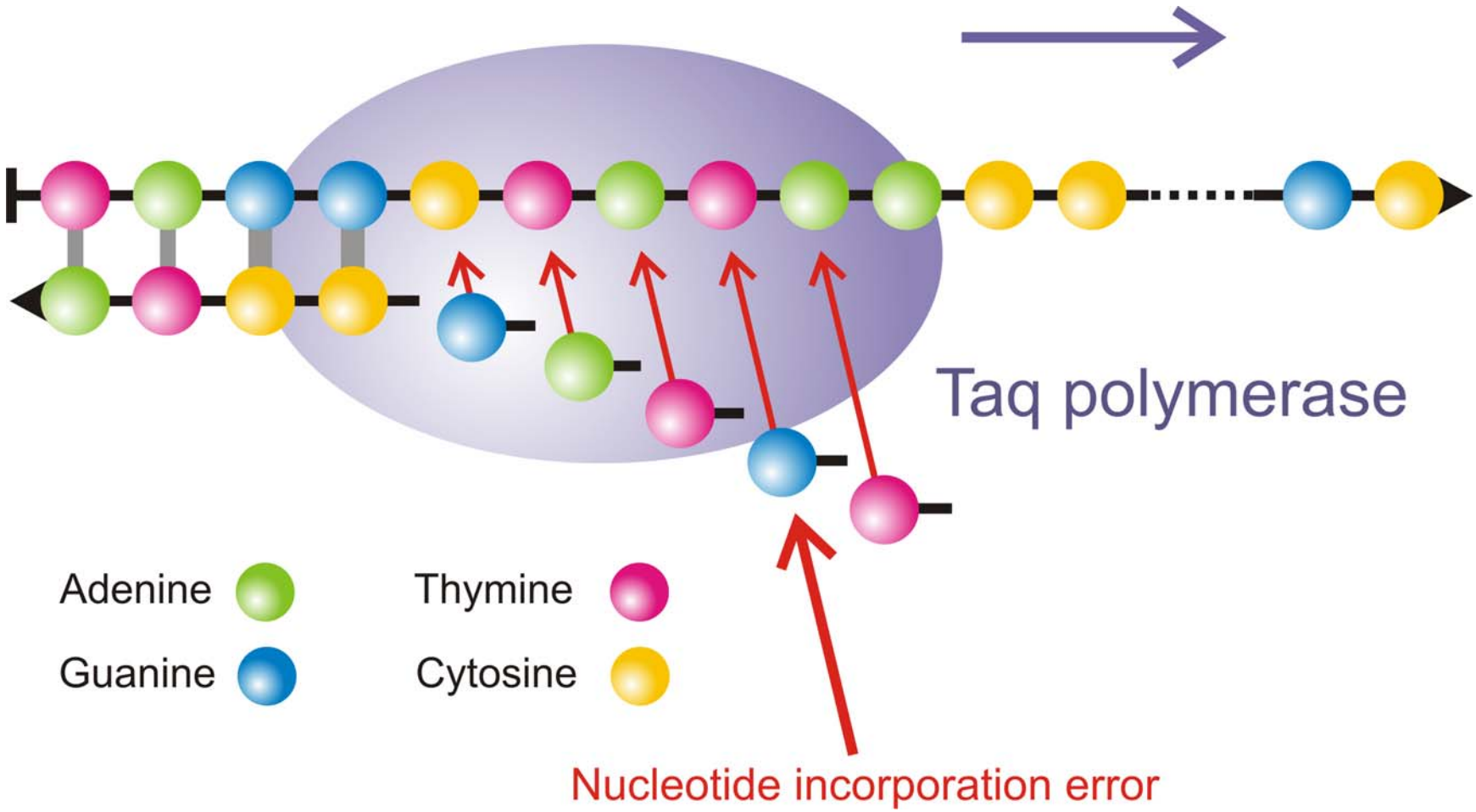
1. Exponential growth and selection
2. **Evolution as replication and mutation**
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution



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

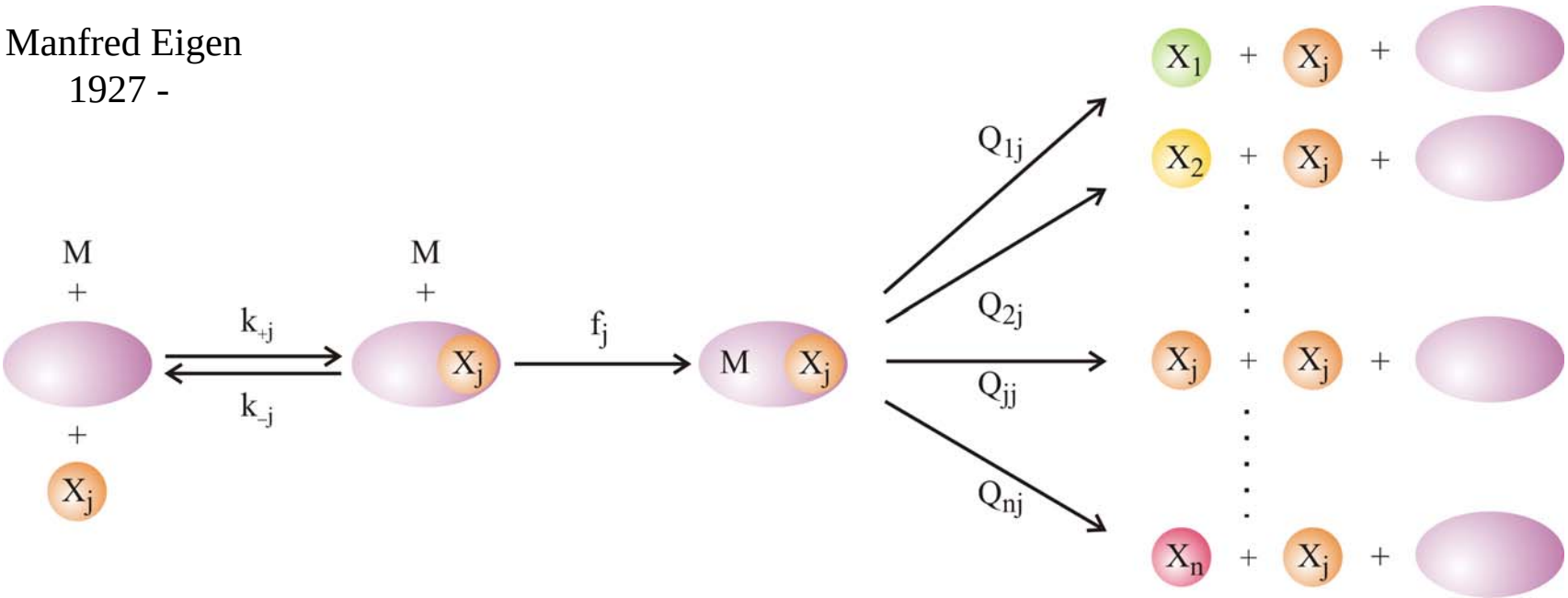


Point mutation

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.

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} Q_{ji} f_i x_i - x_j \phi(t)\,;\; j = 1, 2, \ldots, n\,;\; \sum_{i=1}^{n} x_i = 1\,;\; \phi = \sum_{i=1}^{n} f_i x_i = \bar{f}$$

$$x_j = [\mathrm{X}_j] \geq 0\,;\; f_j \geq 0\,;\; Q_{ji} \geq 0 \;\; \forall \;\; i, j = 1, 2, \ldots, n$$

$$z_j(t) = x_j(t) \exp\left(\int_0^t \phi(\tau) d\tau\right) \quad \text{with} \quad \exp\left(\int_0^t \phi(\tau) d\tau\right) = \left(\sum_{i=1}^{n} z_i(t)\right)^{-1}$$

integrating factor transformation

$$\mathrm{W} = \{W_{ij} = Q_{ij} F_j\}\,;\; \mathrm{L} = \{\ell_{ij}\}\,;\; \mathrm{L}^{-1} = \mathrm{H} = \{h_{ij}\}$$

$$\mathrm{L}^{-1} \cdot \mathrm{W} \cdot \mathrm{L} = \Lambda = \{\Lambda_{ik} = \lambda_k \delta_{ik}\}$$

eigenvalue problem

$$x_j(t) = \frac{\sum_{k=1}^{n} \ell_{jk}\, \zeta_k(0)\, \exp(\lambda_k\, t)}{\sum_{i=1}^{n} \sum_{k=1}^{n} \ell_{ik}\, \zeta_k(0)\, \exp(\lambda_k\, t)}\,;\; j = 1, 2, \ldots, n\,;\; \zeta_k(0) = \sum_{i=1}^{n} h_{ki} x_i(0)$$

Solution of the mutation-selection equation

1. Exponential growth and selection
2. Evolution as replication and mutation
3. **A phase transition in evolution**
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution

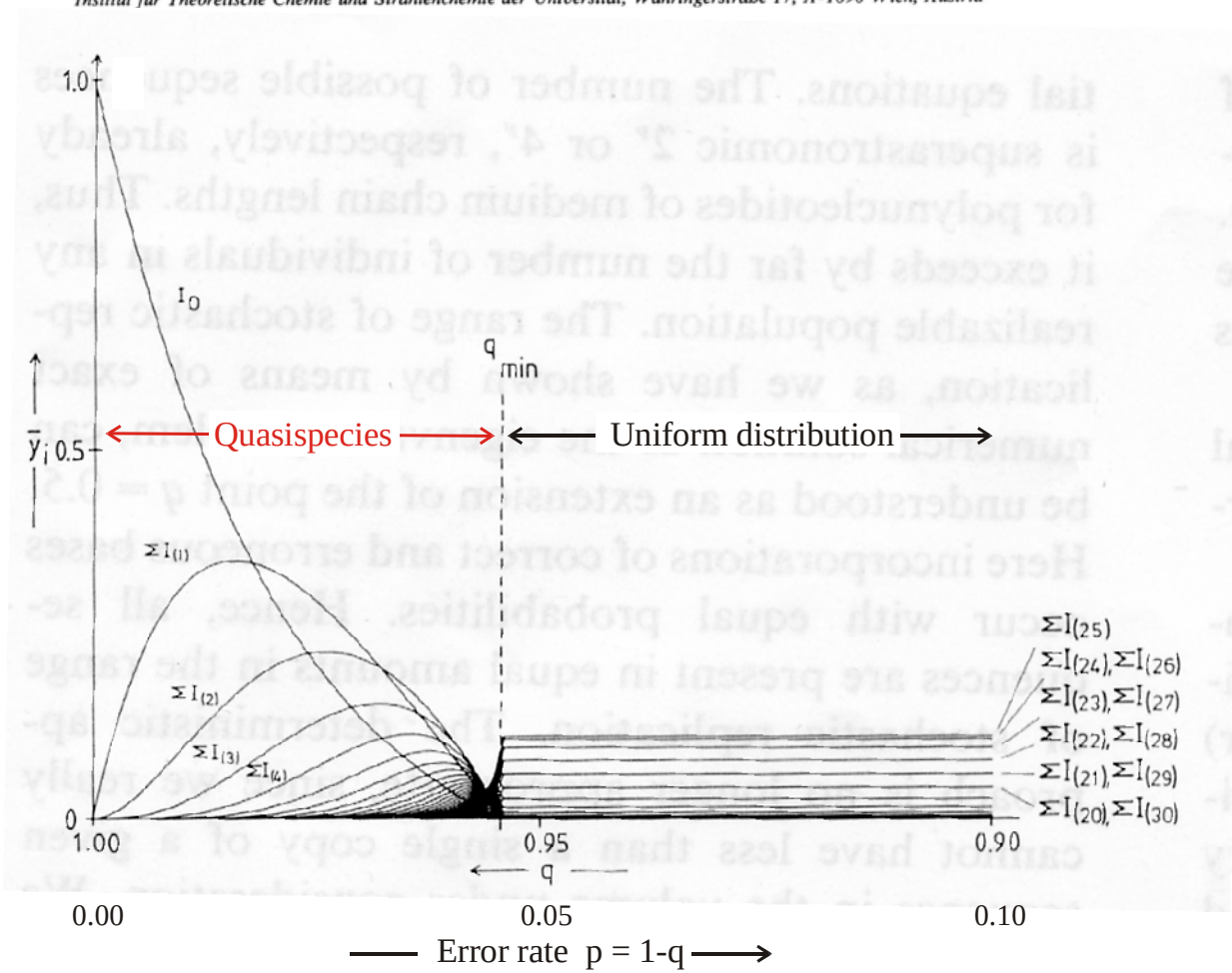
Biophysical Chemistry 16 (1982) 329–345
Elsevier Biomedical Press

SELF-REPLICATION WITH ERRORS

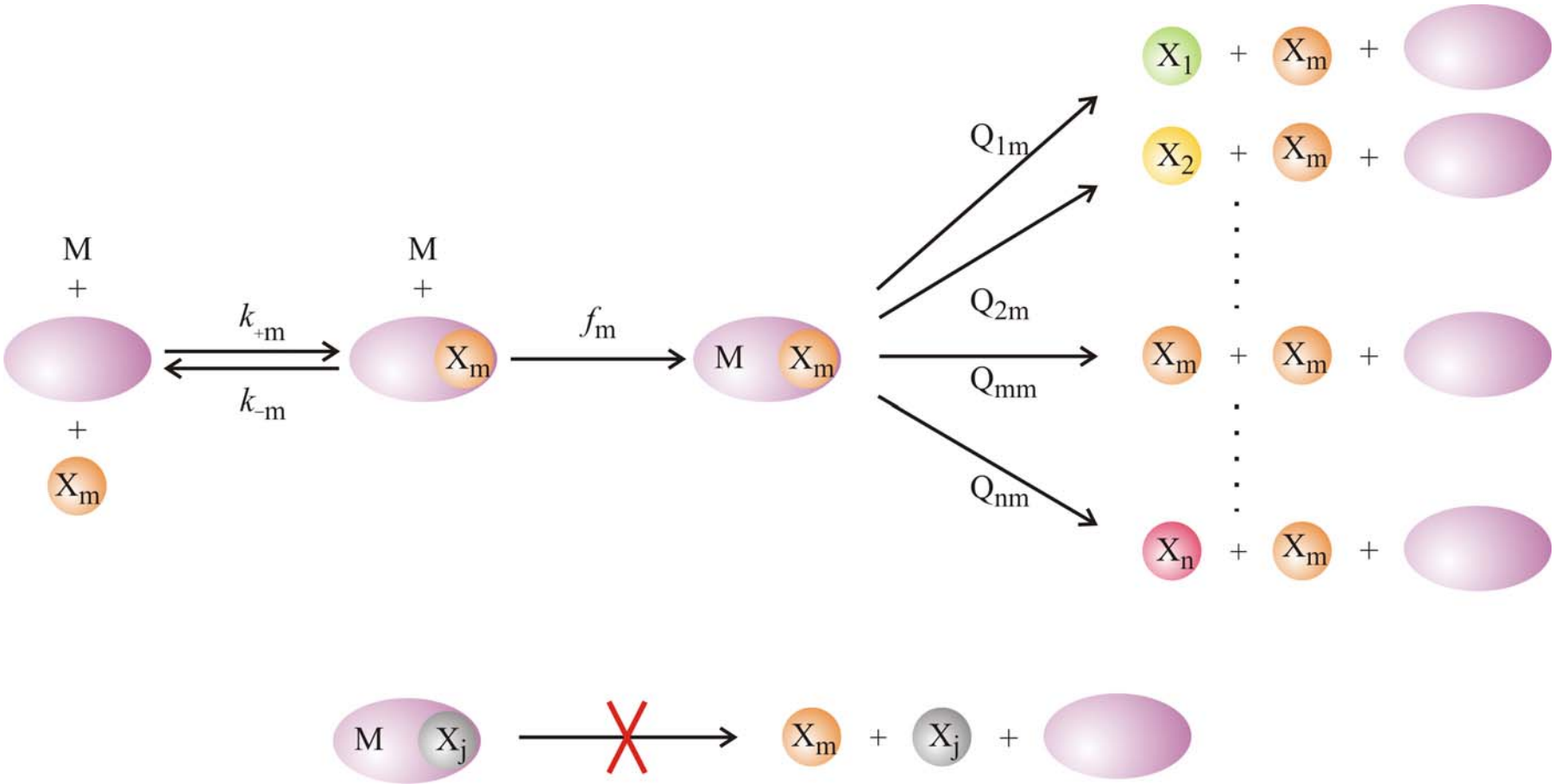
A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

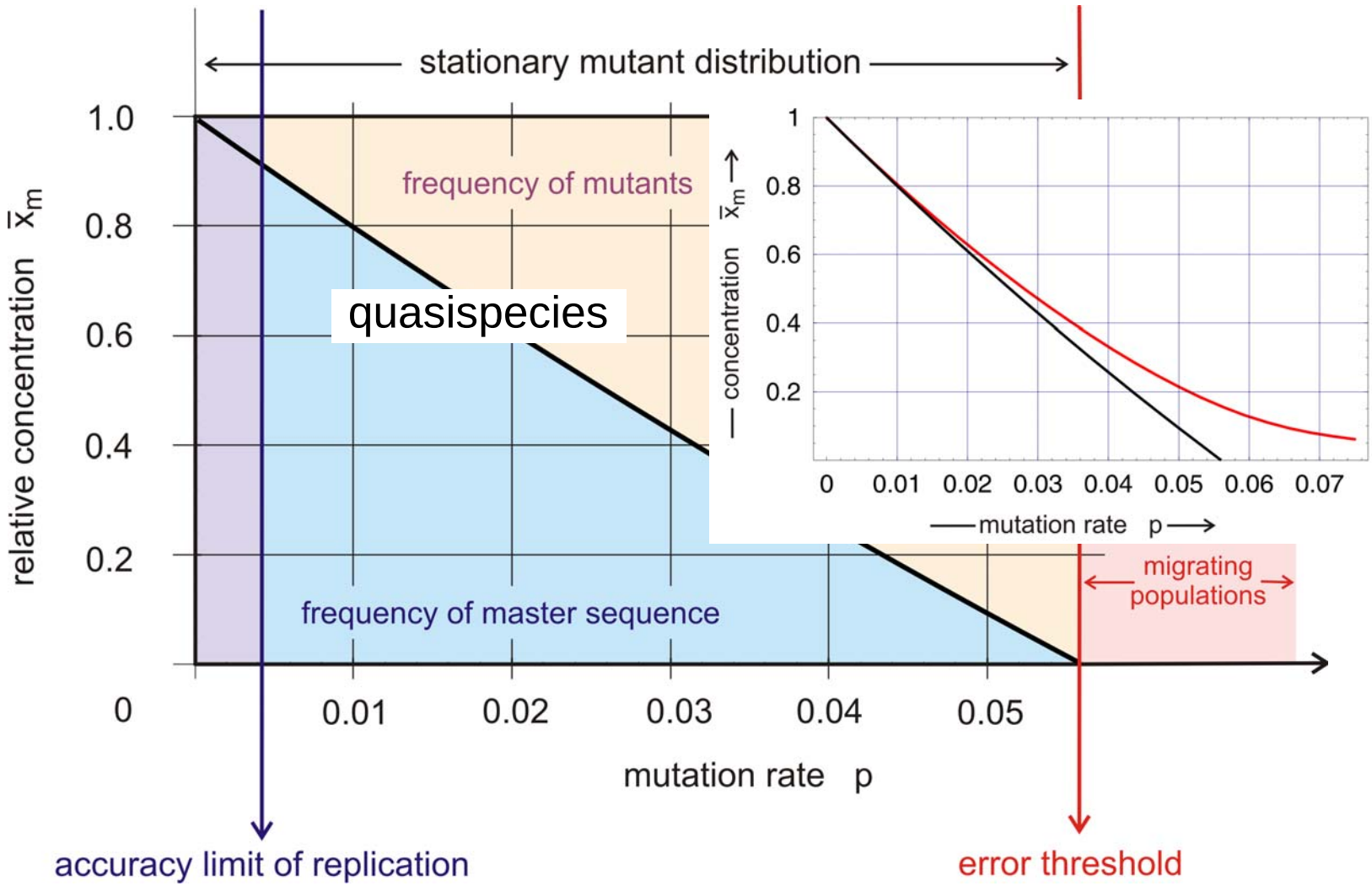
Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria



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



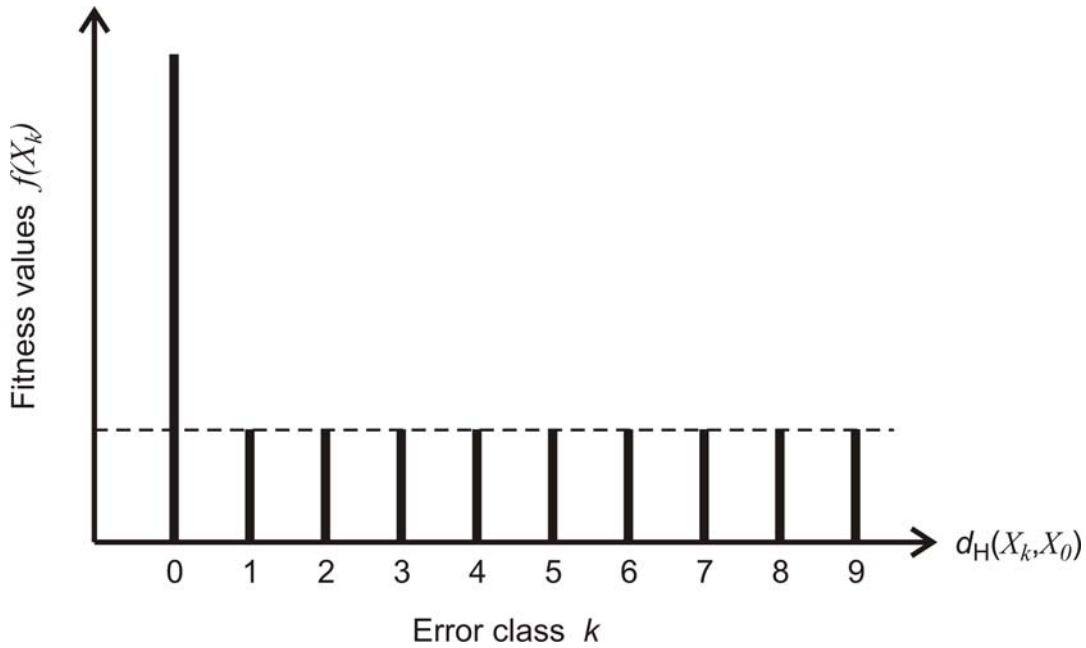
The no-mutational backflow or zeroth order approximation



The error threshold in replication and mutation

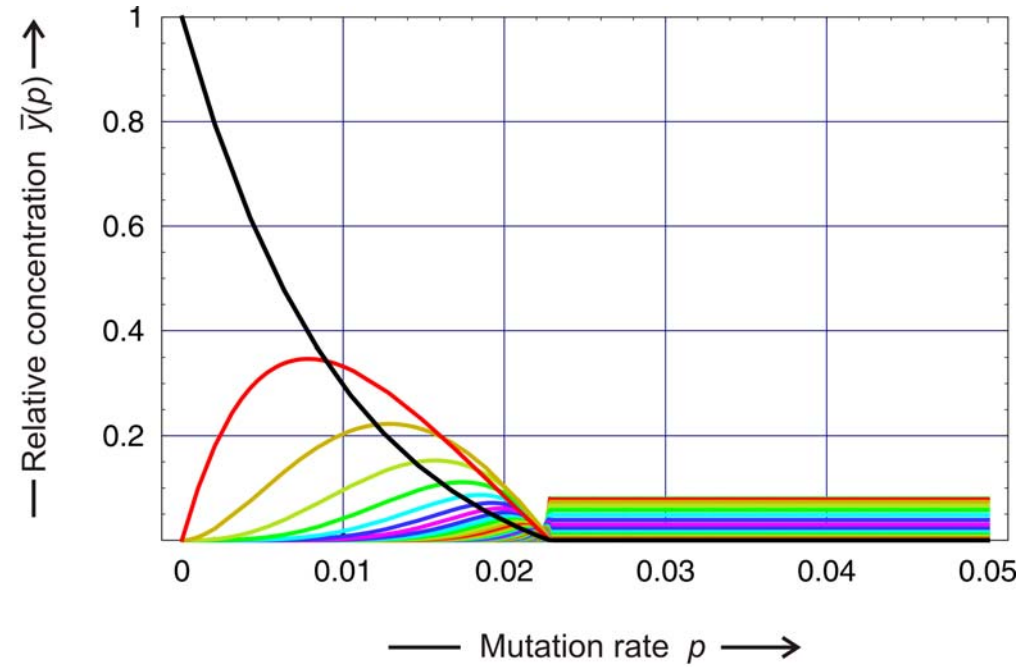
1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. **Fitness landscapes as source of complexity**
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution

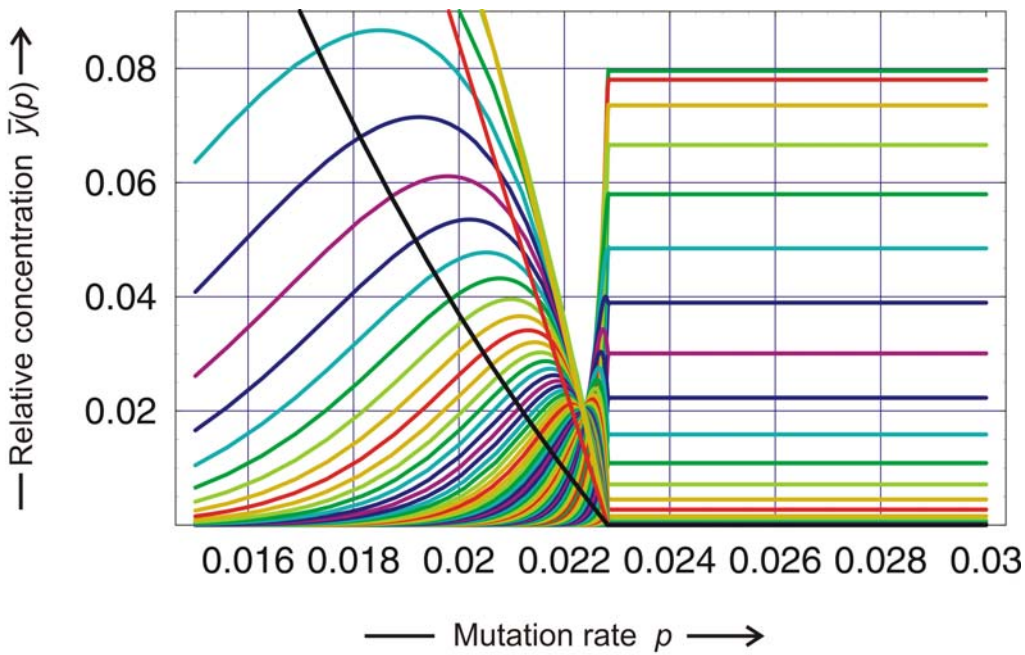
single peak landscape



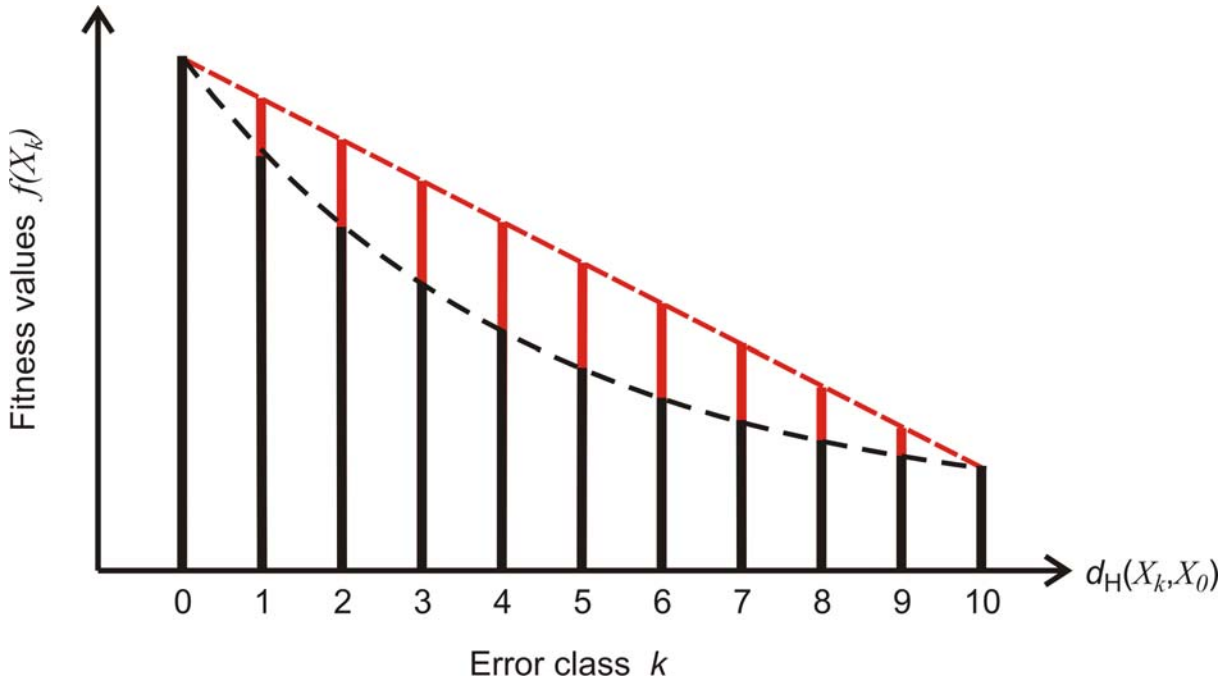
„Rugged" fitness landscapes

Error threshold on the single peak landscape

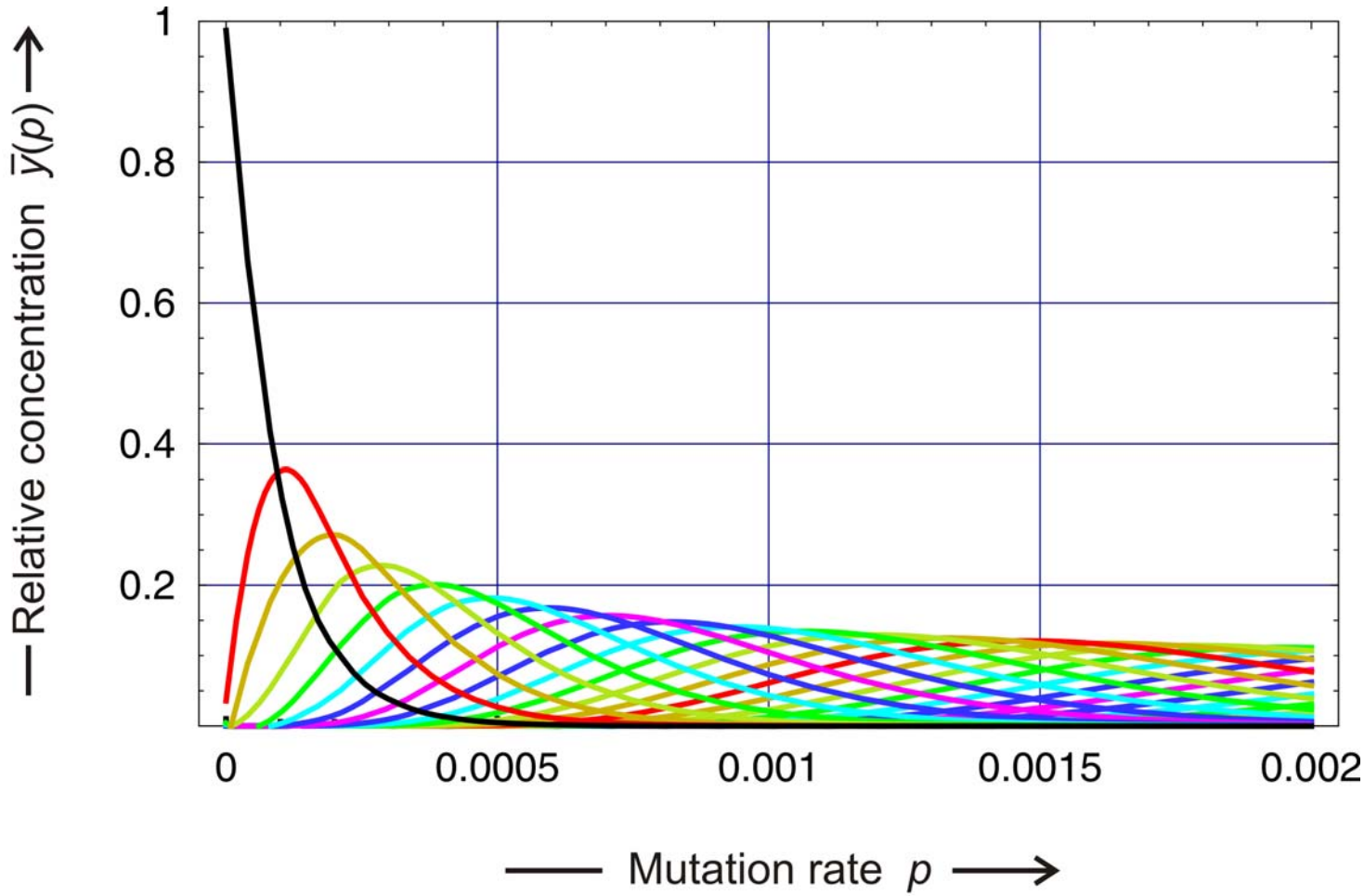




linear and multiplicative landscape



Smooth fitness landscapes



The linear fitness landscape shows no error threshold

Make things as simple as possible,
but not simpler !

Albert Einstein

Albert Einstein‘s razor, precise refence is unknown.

Vol. 131, No. 1 The American Naturalist January 1988

NOTES AND COMMENTS

SURFACES OF SELECTIVE VALUE REVISITED

Provine, in his generally favorable discussion of my shifting-balance theory of evolution, severely criticized the concept of "surfaces of selective value" (1986, p. 307). I think that he was looking for something more mathematical than was intended. Professor E. M. East, as organizer of the program of the Sixth International Congress of Genetics (held in 1932 in Ithaca, New York), had asked me to present a brief, nonmathematical account of the views on evolution that I had presented in a long (63-page) paper in 1931. I agreed to do this.

Most early geneticists thought of the phenotype as if it were a mosaic of unit characters, each determined by a single locus, with effects as conspicuous as those that they used in their experiments. They thought of alleles as having constant relative selective values. The consequences of this assumption were worked out most exhaustively by Haldane in a series of papers beginning in 1924 and summarized in 1932. In addition, he worked out less fully some of the consequences of various other assumptions, also summarized in this book.

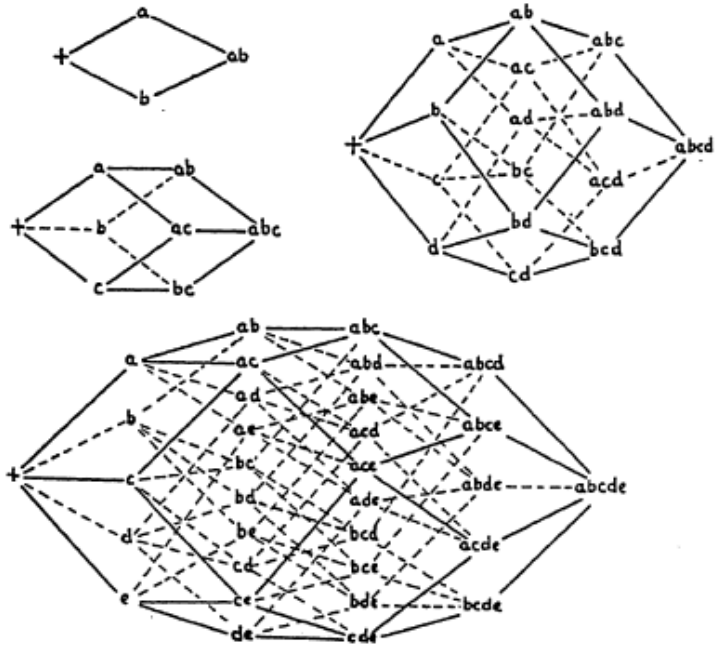


FIG. 1.—The combinations of from 2 to 5 paired allelomorphs.

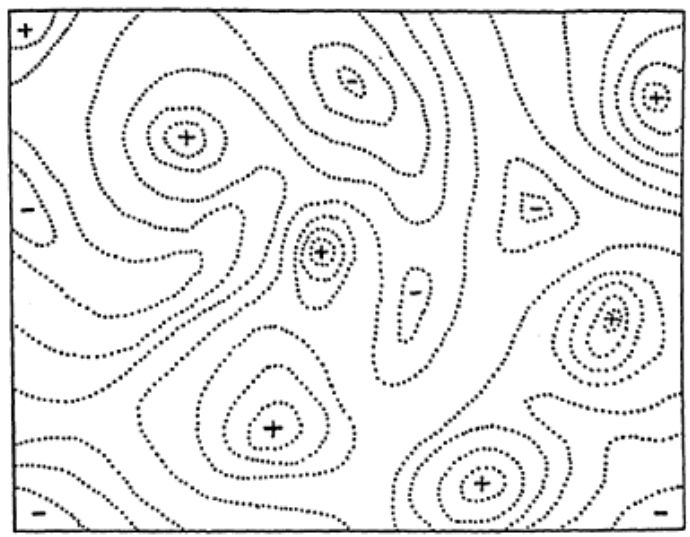
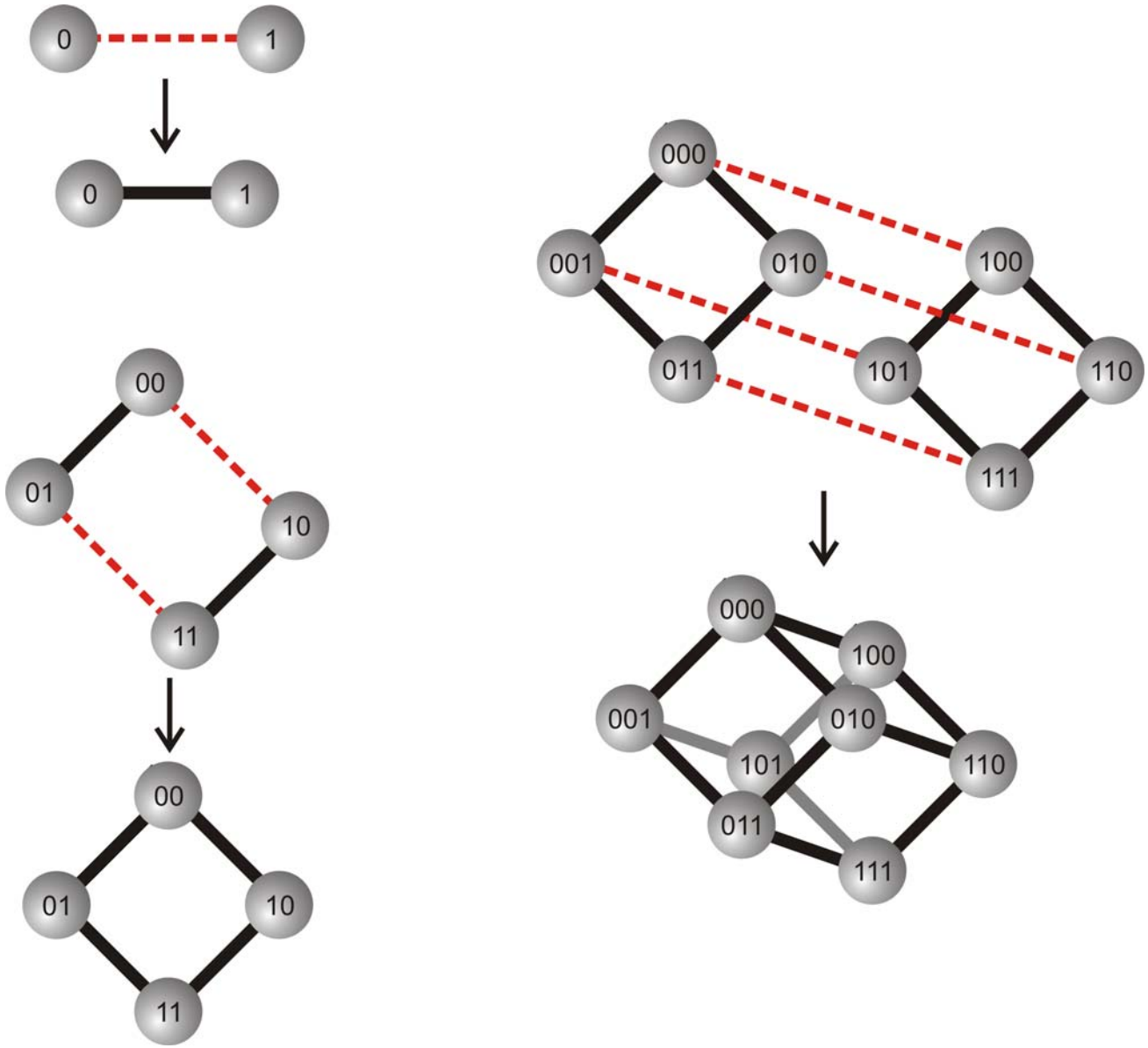


FIG. 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

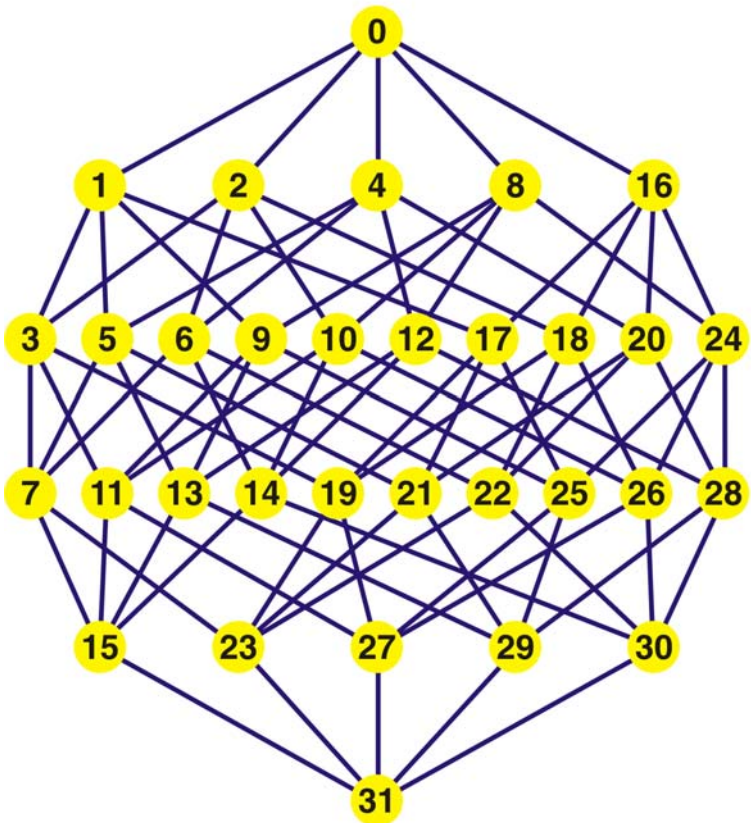
Sewall Wright. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.

-- --. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In: D.F.Jones, ed. *Proceedings of the Sixth International Congress on Genetics, Vol.I.* Brooklyn Botanical Garden. Ithaca, NY, pp. 356-366.

-- --. 1988. Surfaces of selective value revisited. *The American Naturalist* 131:115-131.



Build-up principle of binary sequence spaces



Mutant class

0

1

2

3

4

5

Binary sequences can be encoded by their decimal equivalents:

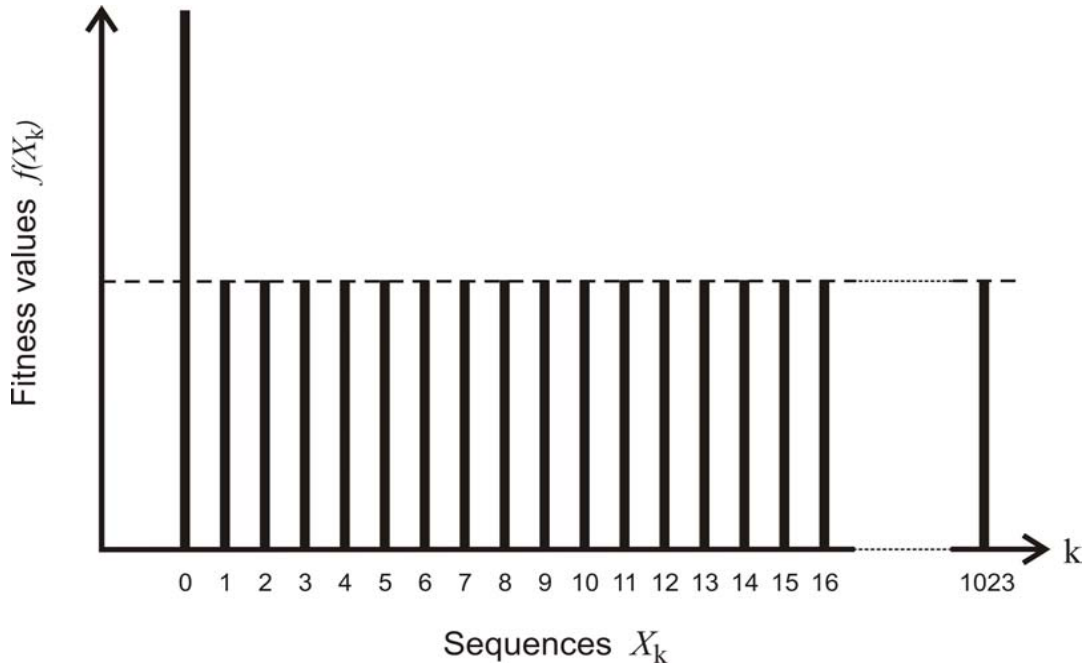
C = 0 and G = 1, for example,

"0" ≡ 00000 = CCCCC,

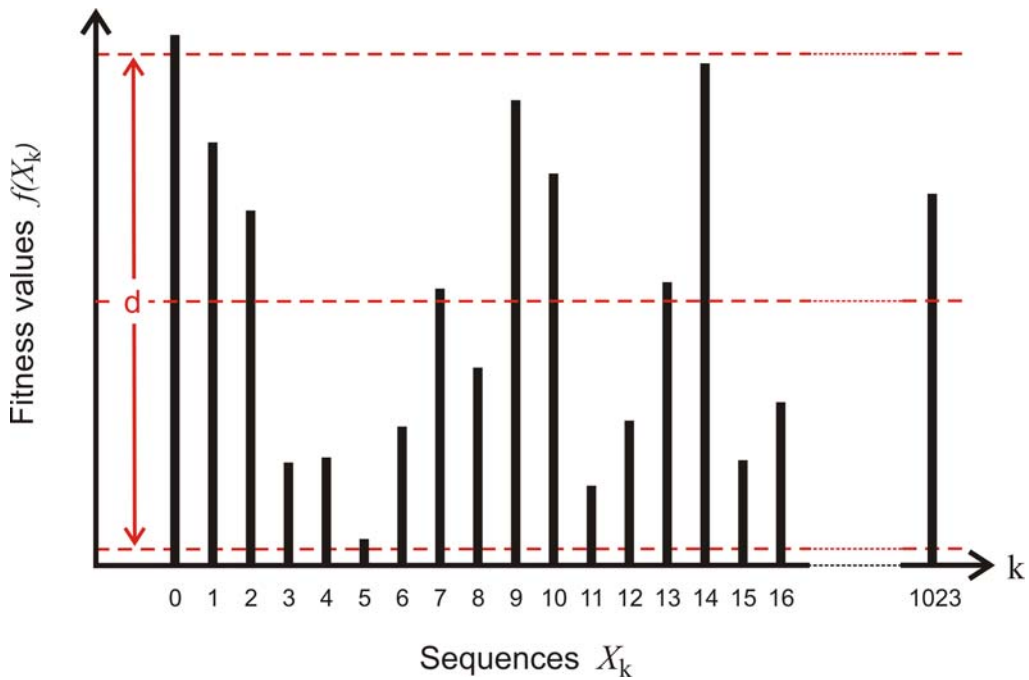
"14" ≡ 01110 = CGGGC,

"29" ≡ 11101 = GGGCG, etc.

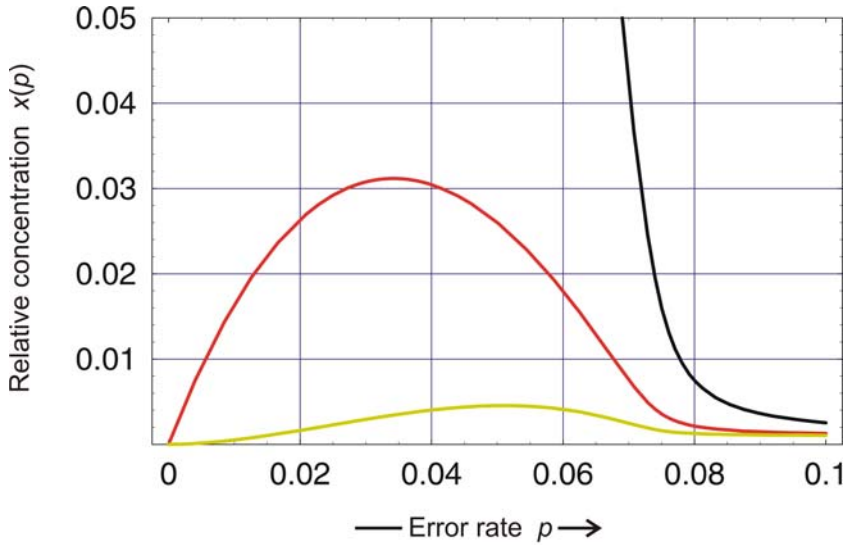
single peak landscape

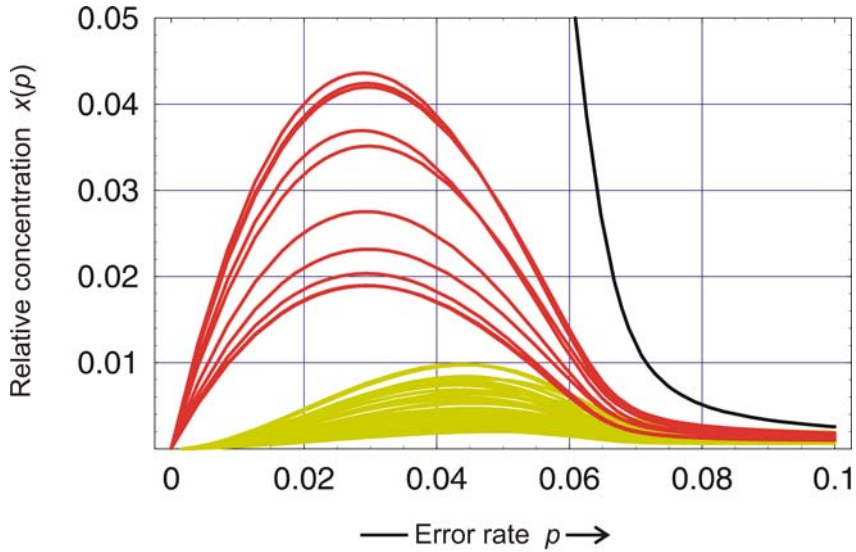


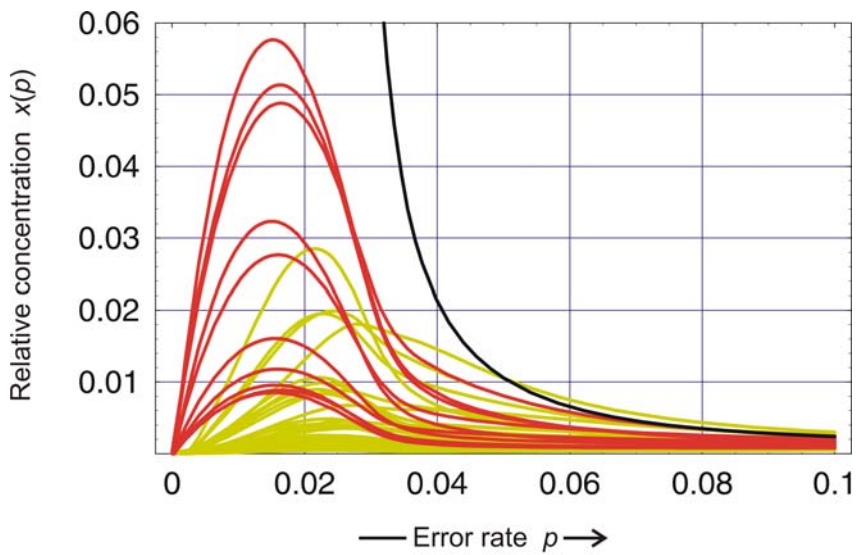
„realistic" landscape



Rugged fitness landscapes over individual binary sequences with n = 10



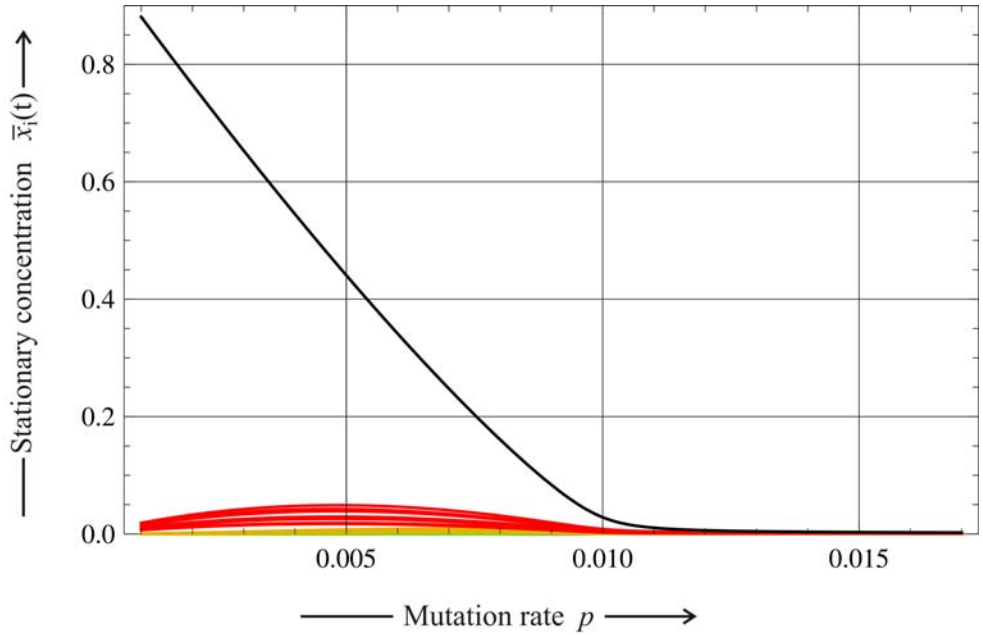




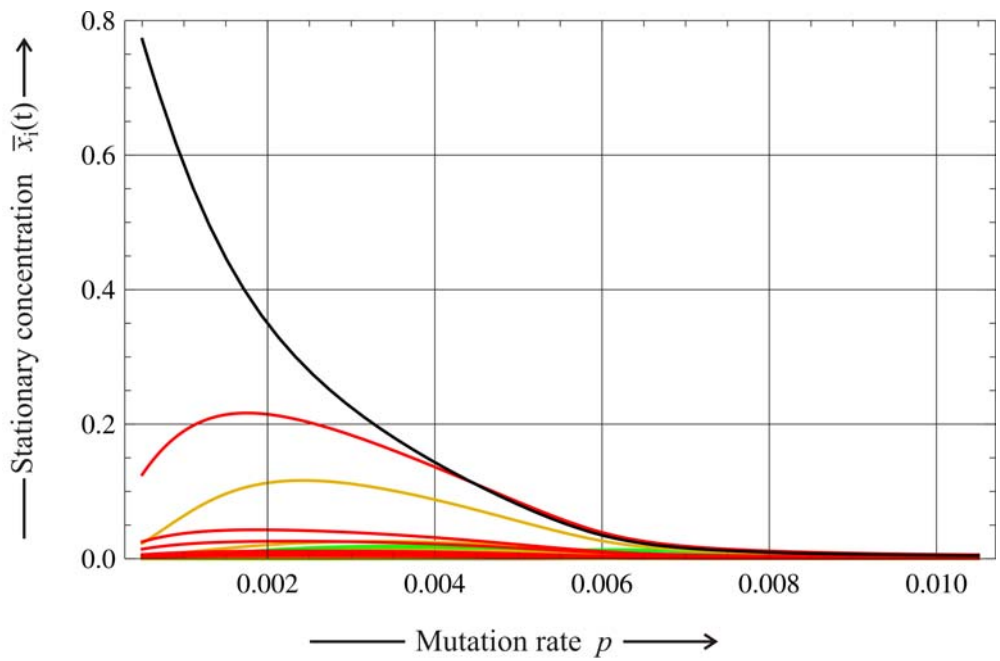
Error threshold: Individual sequences

n = 10, σ = 2, s = 491 and d = 0, 1.0, 1.875

d = 0.100



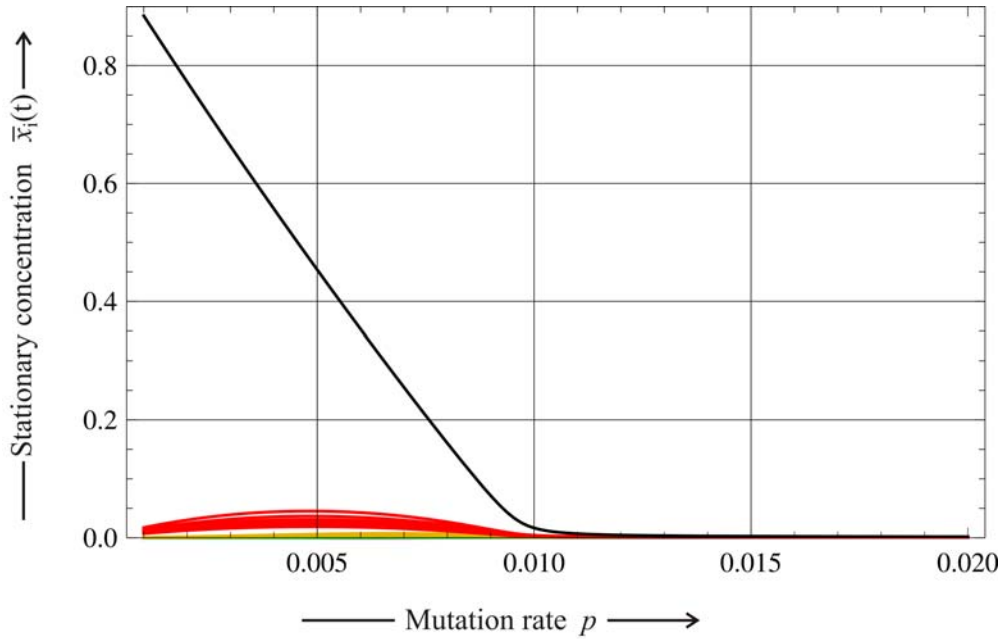
d = 0.200



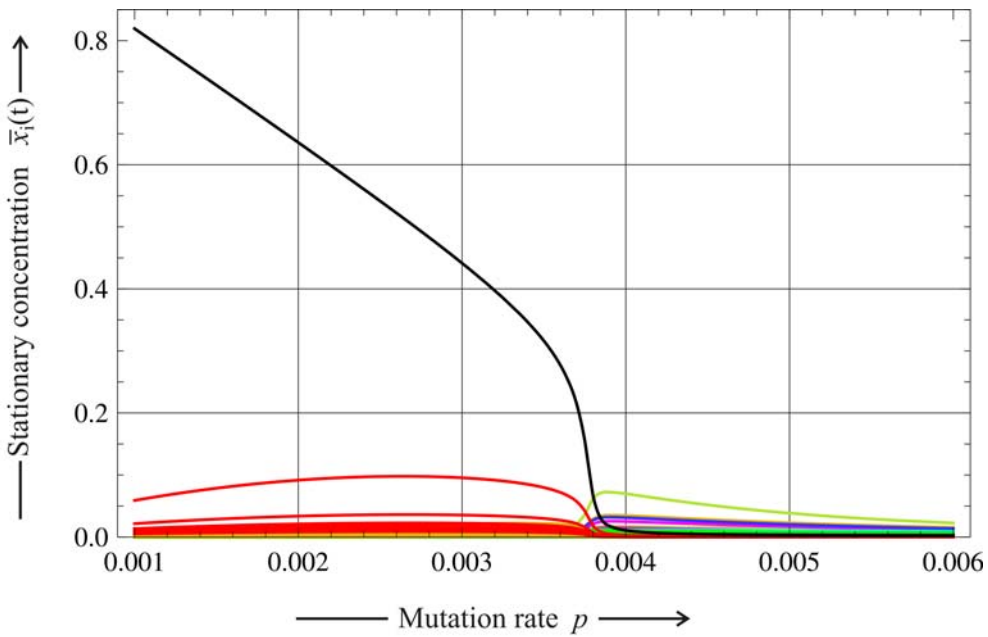
Case I: Strong Quasispecies

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 919

d = 0.100



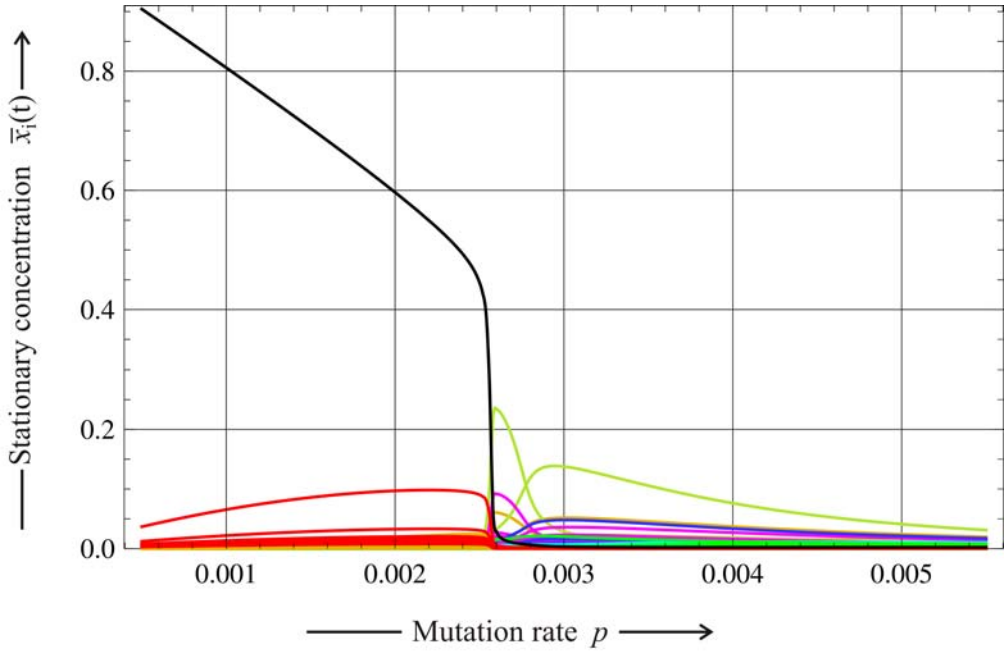
d = 0.195



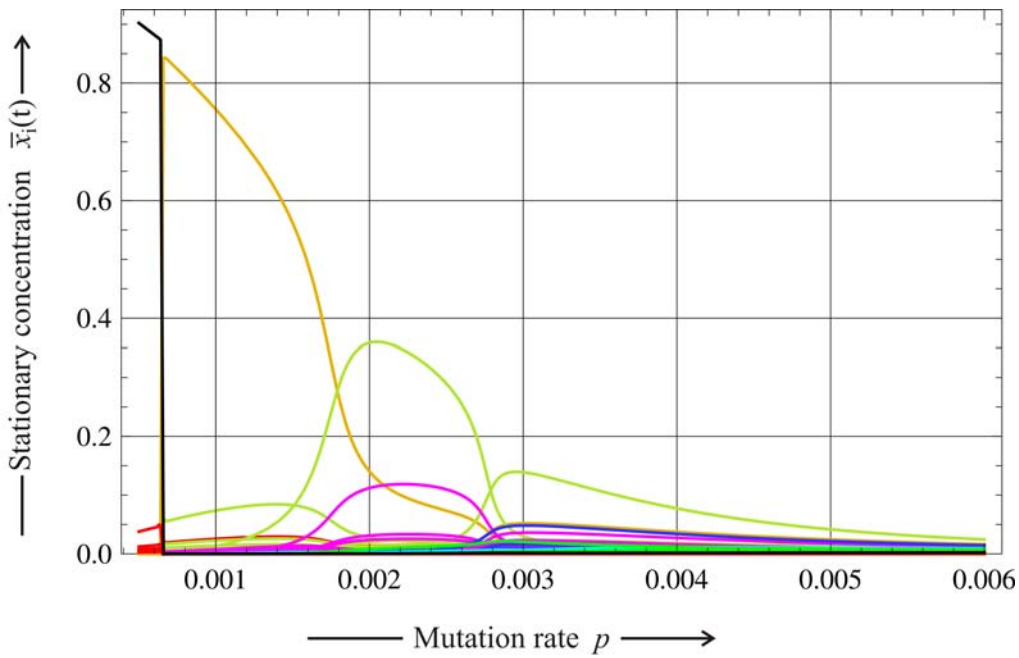
Case III: Multiple transitions

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 637

d = 0.199

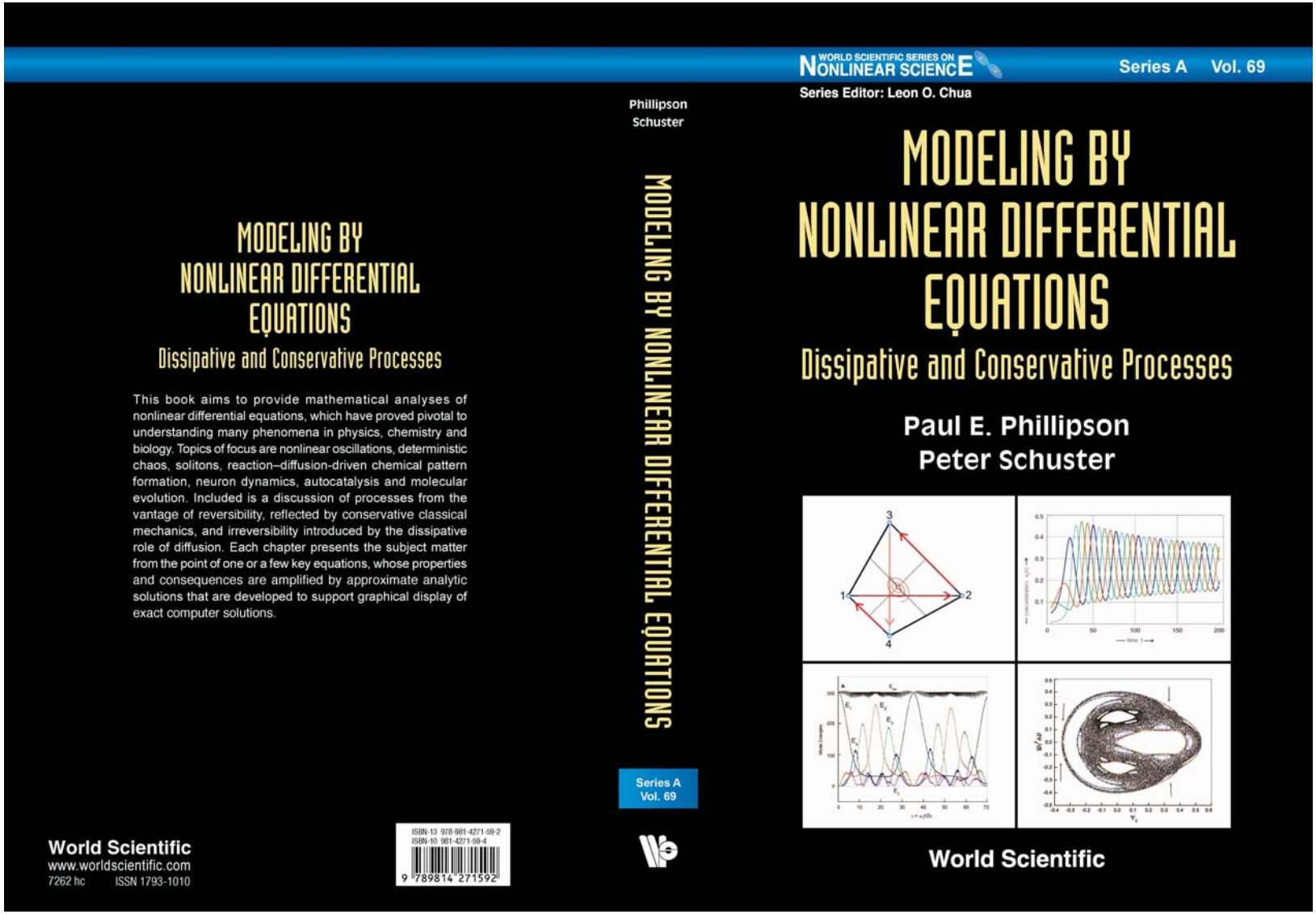


d = 0.200

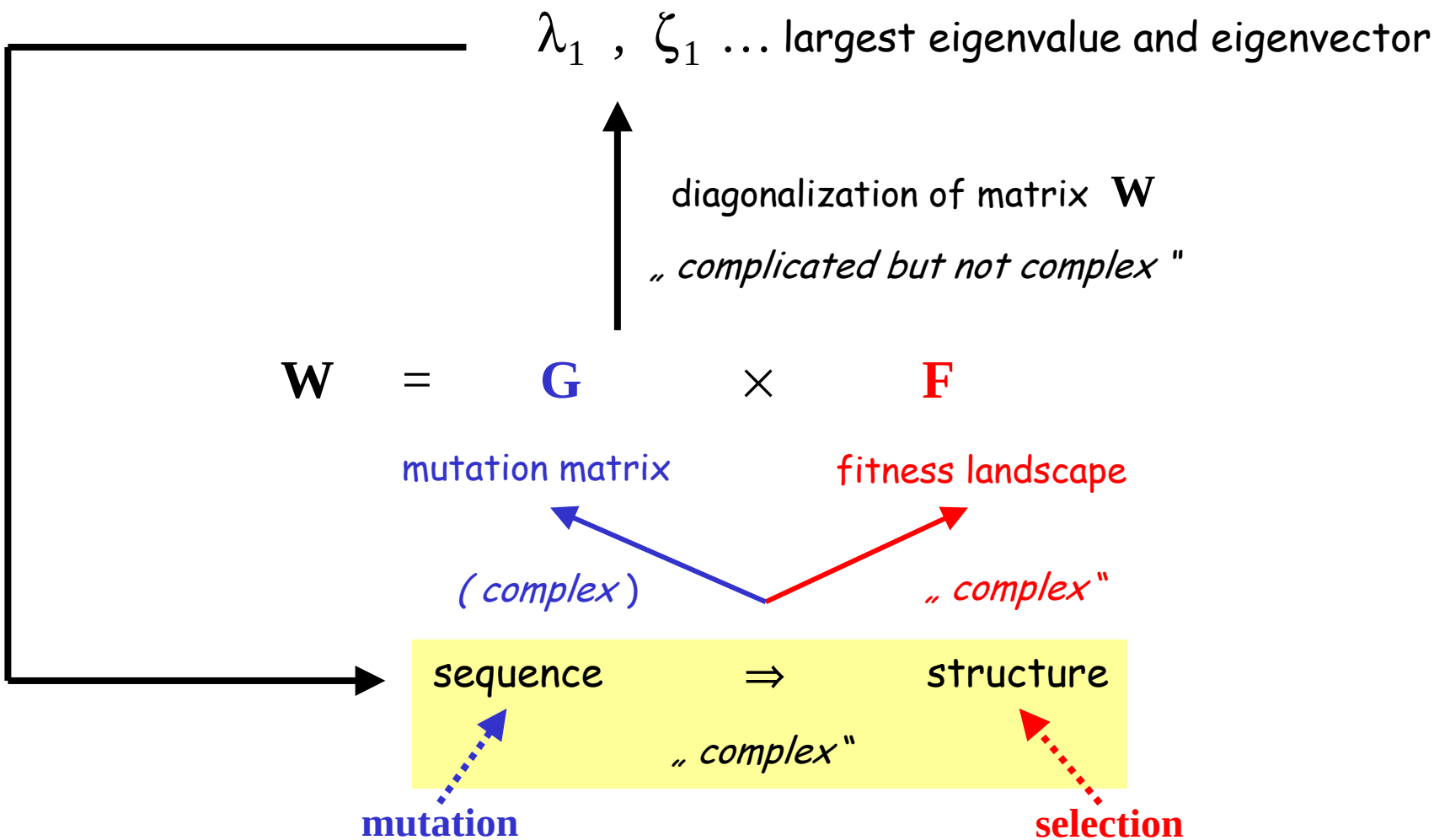


Case III: Multiple transitions

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 637

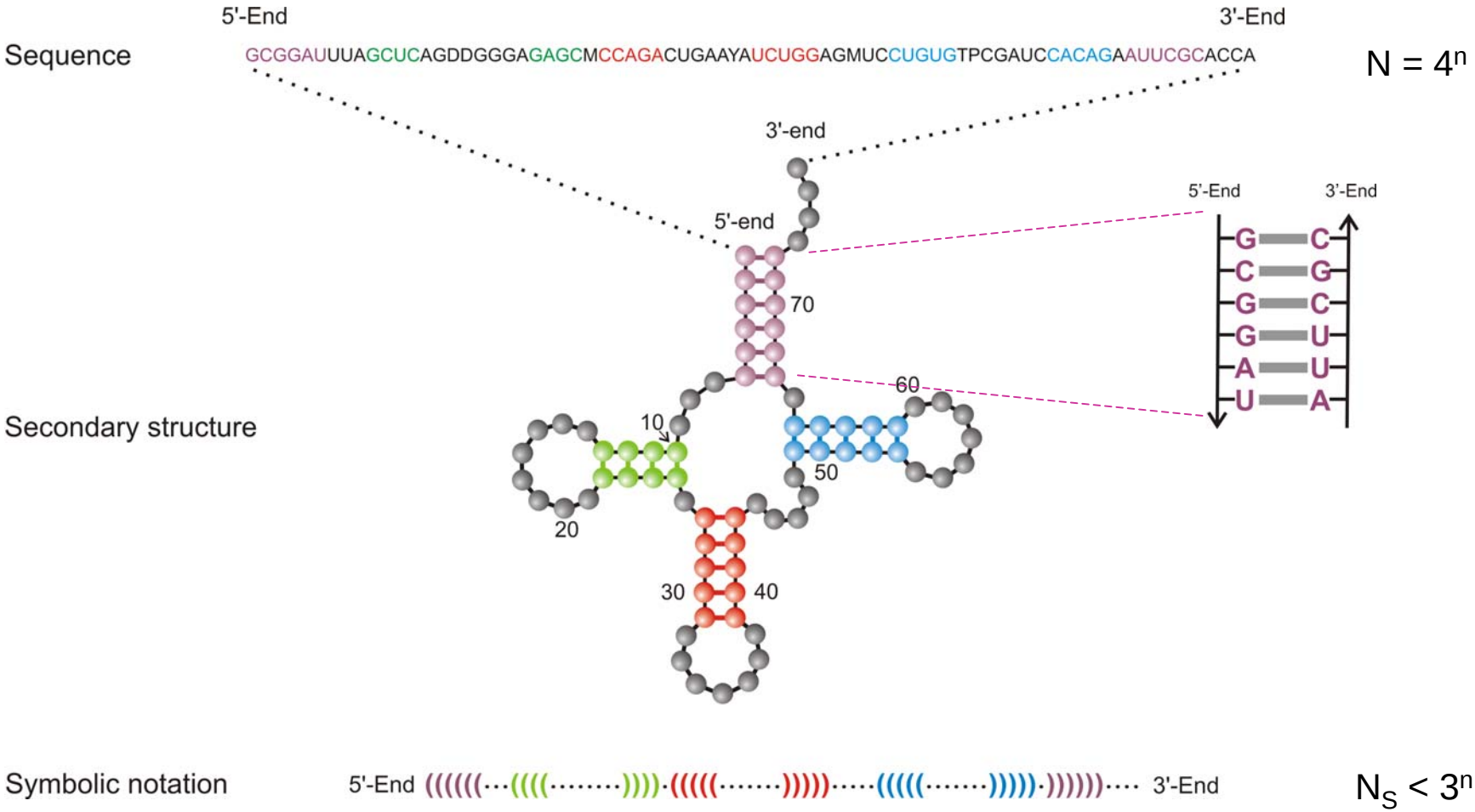


Paul E. Phillipson, Peter Schuster. (2009) Modeling by nonlinear differential equations. Dissipative and conservative processes. World Scientific, Singapore, pp.9-60.



Complexity in molecular evolution

1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. **Molecular landscapes from biopolymers**
6. The role of stochasticity
7. Neutrality and selection
8. Computer simulation of evolution



Criterion: Minimum free energy (mfe)

Rules: _ (_) _ ∈ {**AU**,**CG**,**GC**,**GU**,**UA**,**UG**}

A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs

INSTITUTE OF PHYSICS PUBLISHING REPORTS ON PROGRESS IN PHYSICS

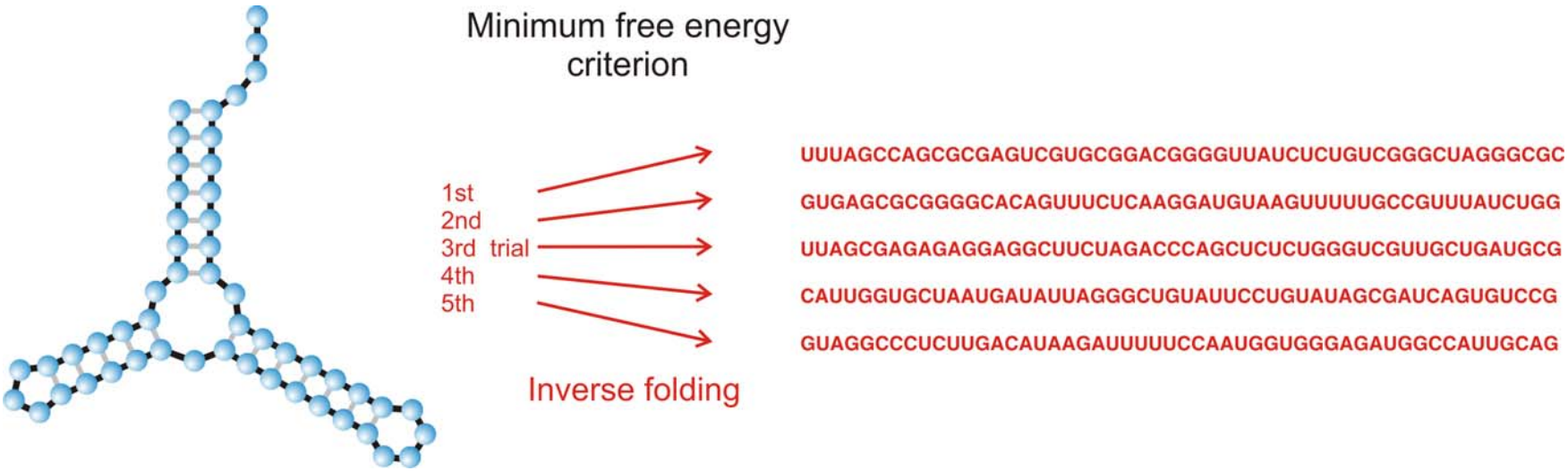
Rep. Prog. Phys. **69** (2006) 1419–1477 doi:10.1088/0034-4885/69/5/R04

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



The **inverse folding algorithm** searches for sequences that form a given RNA secondary structure under the minimum free energy criterion.

What is neutrality ?

Selective neutrality =
= several genotypes having the same fitness.

Structural neutrality =
= several genotypes forming molecules with the same structure.

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

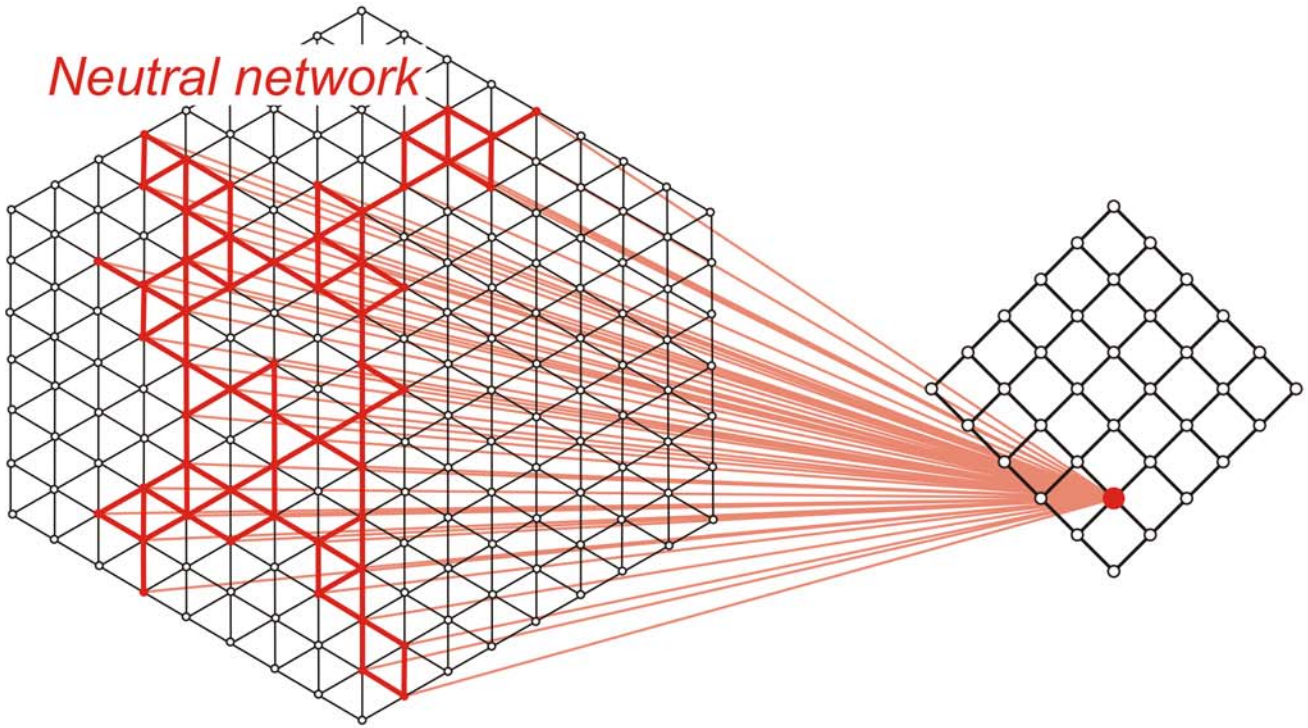
Space of phenotypes: $S = \{S_1, S_2, S_3, S_4, \ldots, S_M\}$; metric (not required)

$$N >> M$$

$$\psi(I_j) = S_k$$

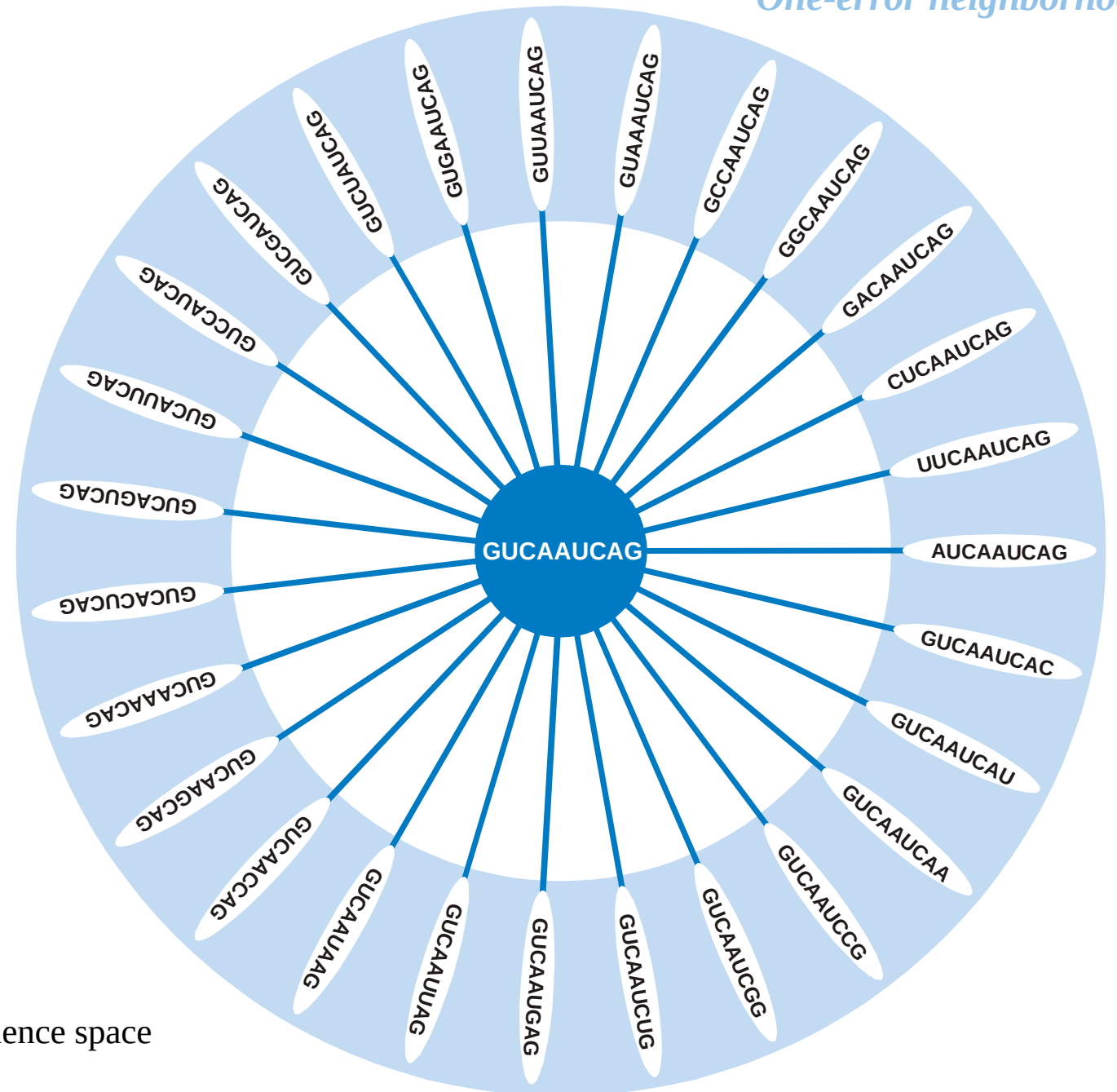
$$\mathbf{G}_k = \psi^{-1}(S_k) \cup \{ I_j \mid \psi(I_j) = S_k \}$$

A mapping ψ and its inversion

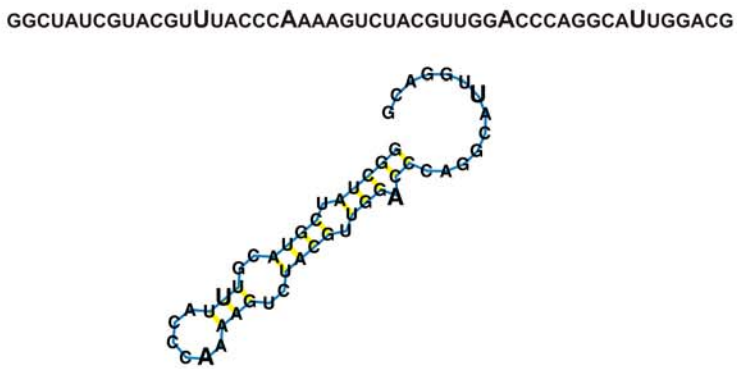


many genotypes ⇒ one phenotype

One-error neighborhood



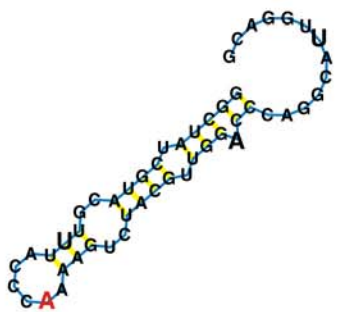
The surrounding of
GUCAAUCAG in sequence space



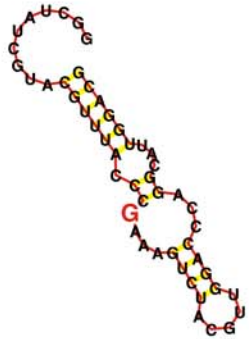
One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space

GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

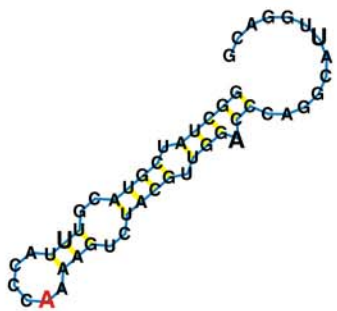


One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space

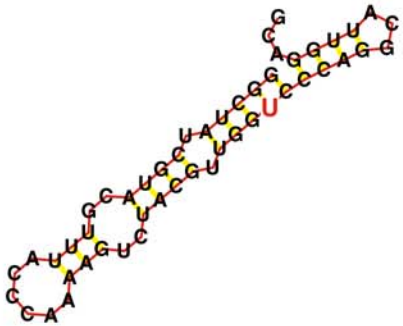


GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

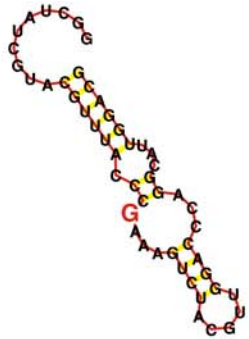
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space

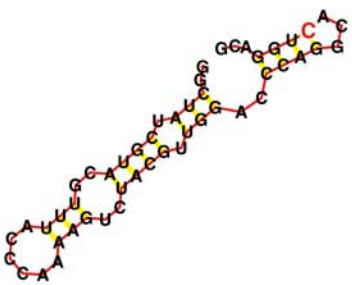


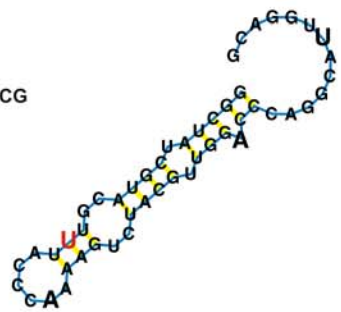
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGUCCCAGGCAUUGGACG



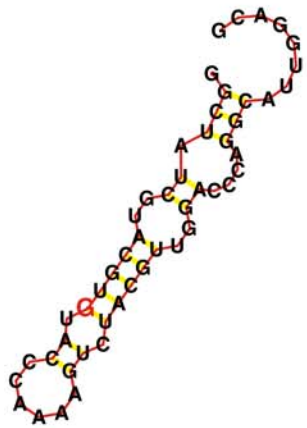
GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCACUGGACG





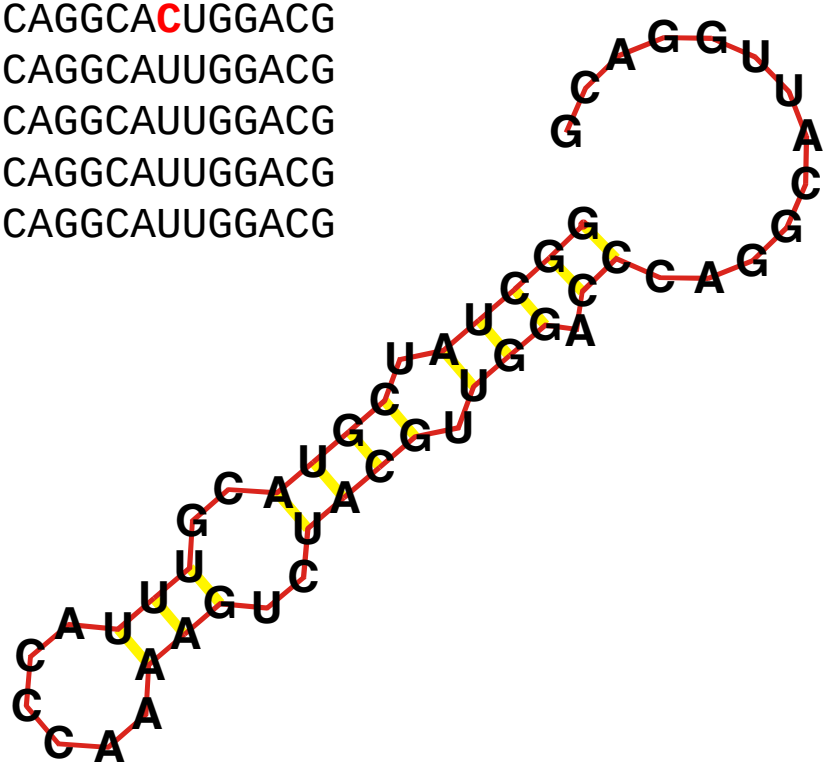
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space

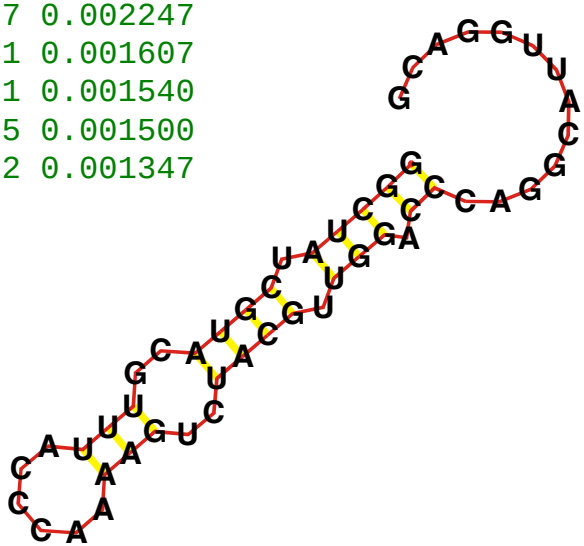
GGCUAUCGUA**U**GUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUU**A**GACG
GGCUAUCGUACGUUUAC**U**CAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACG**C**UUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGC**C**AUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGU**G**UACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUA**A**CGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCC**U**GGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCA**C**UGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGG**U**CCCAGGCAUUGGACG
GGCUA**G**CGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCC**G**AAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAG**C**CUACGUUGGACCCAGGCAUUGGACG



One error neighborhood – Surrounding of an RNA molecule of chain length n=50 in sequence and shape space

	Number	Mean Value	Variance	Std.Dev.
Total Hamming Distance:	150000	11.647973	23.140715	4.810480
Nonzero Hamming Distance:	99875	16.949991	30.757651	5.545958
Degree of Neutrality:	50125	**0.334167**	0.006961	**0.083434**
Number of Structures:	**1000**	**52.31**	85.30	**9.24**

1	(((((.((((..(((......))).)))))..))))..)).).....	**50125**	**0.334167**
2	..(((.((((..(((......))).)))))..))))..)).).....	**2856**	**0.019040**
3	((((((((((..(((......))).)))))))))..)).).....	**2799**	**0.018660**
4	(((((.((((..((((....)))).)))))..))))..)).).....	**2417**	**0.016113**
5	(((((.((((.((((......)))).)))))..))))..)).).....	**2265**	**0.015100**
6	(((((.(((((.(((......))).)))))..))))..)).).....	**2233**	**0.014887**
7	(((((..(((..(((......))).)))..)))..)))..)).).....	1442	0.009613
8	(((((.((((..((........)).)))))..))))..)).).....	1081	0.007207
9	((((..((((..(((......))).)))))..))))..)).).....	1025	0.006833
10	(((((.((((..(((......))).)))))..))))..))))).....	1003	0.006687
11	.((((.((((..(((......))).)))))..))))..)))).....	963	0.006420
12	(((((.(((...(((......))).)))...)))..)))..)).).....	860	0.005733
13	(((((.((((..(((......))).)))))..))))..)).))).....	800	0.005333
14	(((((.((((...(((......))).)))...))))..)))..)).).....	548	0.003653
15	(((((.((((..............))))..)))..)).).....	362	0.002413
16	((.((.((((..(((......))).)))))..))))..))..)).).....	337	0.002247
17	(.(((.((((..(((......))).)))))..))))..)))..).....	241	0.001607
18	(((((.(((((((((......)))))))))..)))..)).).....	231	0.001540
19	((((..((((..(((......))).)))))...))))..)).).....	225	0.001500
20	((....((((..(((......))).)))))....))..)).).....	202	0.001347



Shadow – Surrounding of an RNA structure in shape space:
AUGC alphabet, chain length n=50

1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. **The role of stochasticity**
7. Neutrality and selection
8. Computer simulation of evolution

Stochastic phenomena in evolutionary processes

1. Finite population size effects	ODEs (in population genetics) describe expectation values in infinite populations.
2. Low numbers of individual species	Every mutant starts from a single copy.
3. Selective neutrality	Populations drift randomly in the space of neutral variants.

$$P_k^{(j)}(t) \;=\; Prob\{\mathcal{X}_j \,=\, k\}\,,\;\; k = 0, 1, \ldots, N\,;\;\; j = 1, \ldots, n$$

probabilistic notion of particle numbers $\mathcal{X}_j$

$$\begin{aligned}\frac{dP_k^{(j)}}{dt} \;=\; & \left(\sum_{i=1}^{n} Q_{ji} f_i \sum_{s=1}^{N} s\, P_s^{(i)}\right) P_{k-1}^{(j)} \;-\; \phi(t)\, P_k^{(j)} \;- \\ & - \left(\sum_{i=1}^{n} Q_{ji} f_i \sum_{s=1}^{N} s\, P_s^{(i)}\right) P_k^{(j)} \;+\; \phi(t)\, P_{k+1}^{(j)}\end{aligned}$$

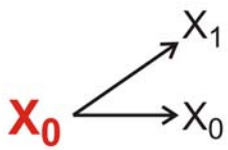
master equation

$$\begin{aligned}\frac{dP_k^{(j)}}{dt} \;=\; & \left(\sum_{i=1}^{n} Q_{ji} f_i \sum_{s=1}^{N} s\, P_s^{(i)}\right) P_{k-1}^{(j)} \;-\; r\, k\, P_k^{(j)} \;- \\ & - \left(\sum_{i=1}^{n} Q_{ji} f_i \sum_{s=1}^{N} s\, P_s^{(i)}\right) P_k^{(j)} \;+\; r\,(k+1)\, P_{k+1}^{(j)}\end{aligned}$$

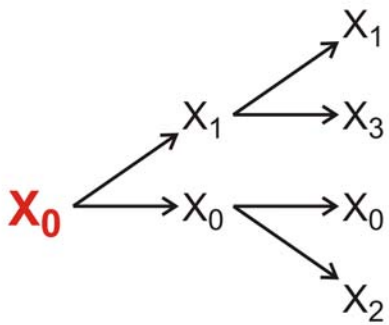
flow reactor

x_0

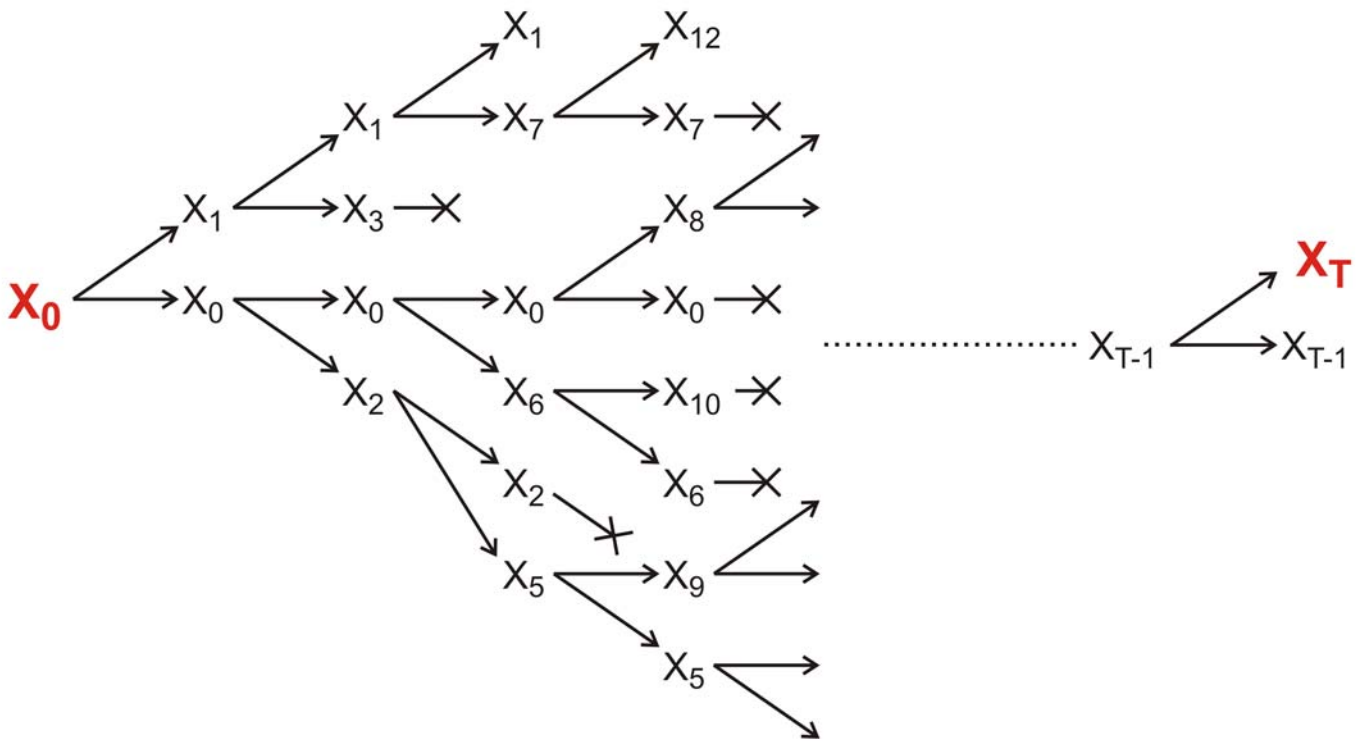
Evolution of RNA molecules as a Markow process



Evolution of RNA molecules as a Markow process



Evolution of RNA molecules as a Markow process



Evolution of RNA molecules as a Markow process

Bulletin of Mathematical Biology Vol. 47, No. 2, pp. 239–262, 1985.
Printed in Great Britain

0092-8240/85$3.00 + 0.00
Pergamon Press Ltd.

POLYNUCLEOTIDE EVOLUTION AND BRANCHING PROCESSES*

■ LLOYD DEMETRIUS
Max-Planck-Institut für Biophysikalische Chemie,
Göttingen, F.R.G.

■ PETER SCHUSTER
Institut für Theoretische Chemie und Strahlenchemie,
Universität Wien,
A-1090 Wien, Austria

■ KARL SIGMUND
Institut für Mathematik,
Universität Wien,
A-1090 Wien, Austria, and
IIASA, A-2361 Laxenburg, Austria

The theory of multitype branching processes is applied to the kinetics of polynucleotide replication. The results obtained are compared with the solutions of the deterministic differential equations of conventional chemical kinetics.

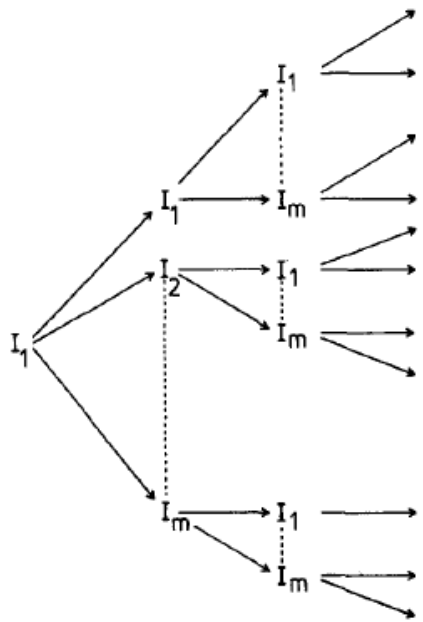
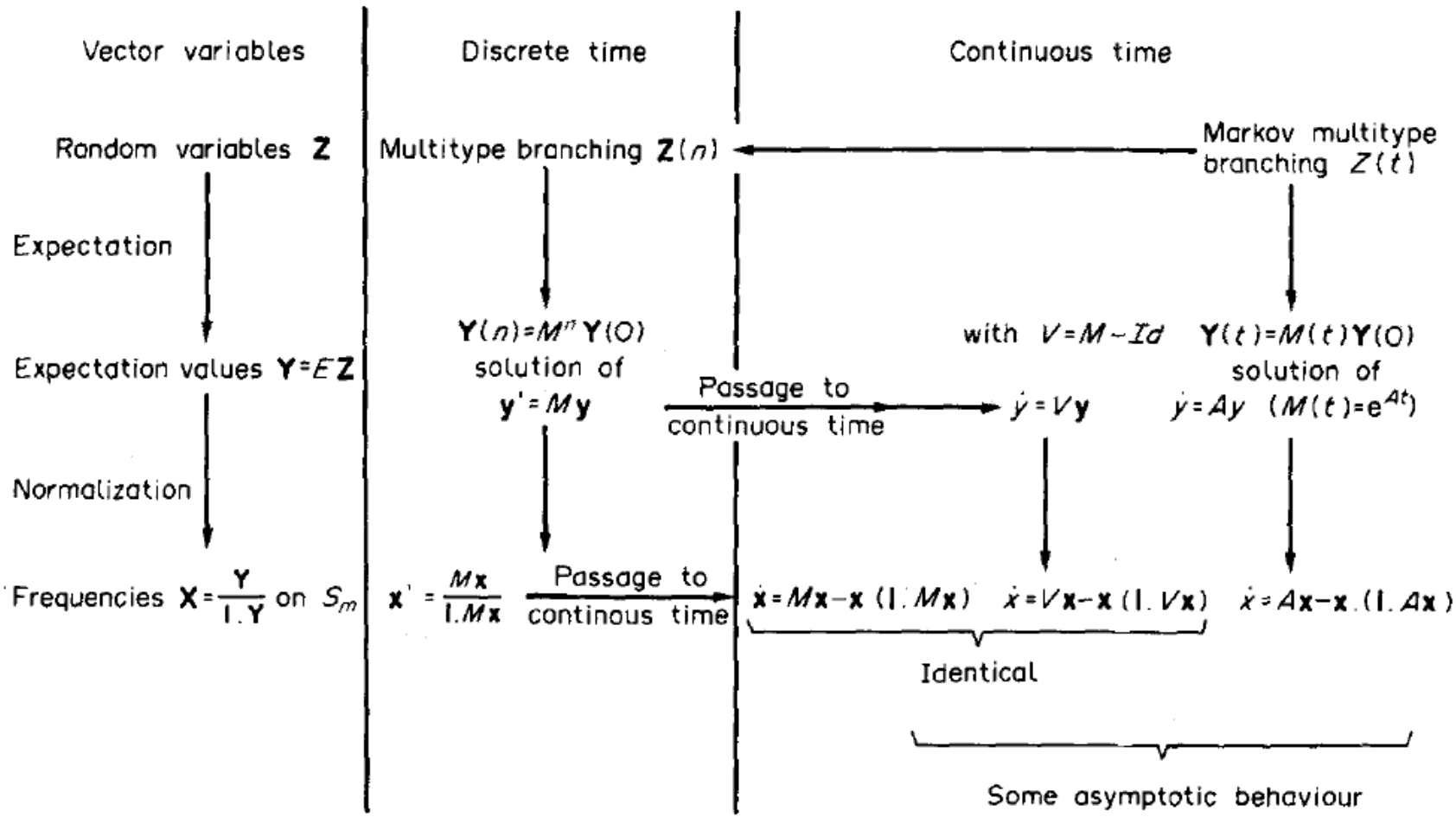
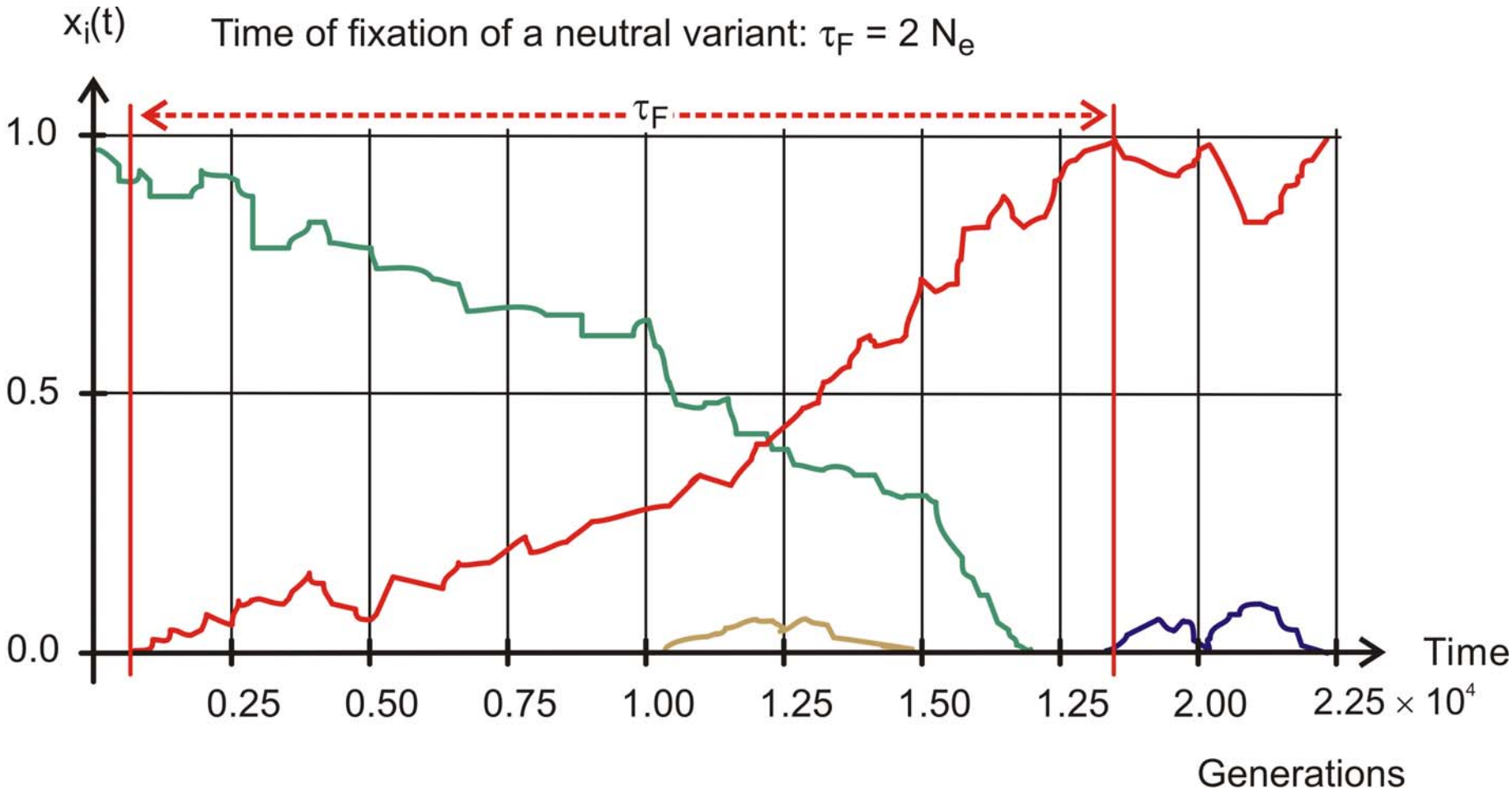


Figure 1. Replication as a multitype branching process.

RNA replication and mutation as a multitype branching process


Vector variables
Discrete time
Continuous time
Random variables Z
Multitype branching Z(n)
Markov multitype branching Z(t)
Expectation
Expectation values Y=EZ
Y(n)=M^n Y(0) solution of y'=My
with V=M−Id
Y(t)=M(t)Y(0) solution of ẏ=Ay (M(t)=e^{At})
Passage to continuous time
ẏ=Vy
Normalization
Frequencies X=Y/l.Y on S_m
x'=Mx/l.Mx
Passage to continous time
ẋ=Mx−x(l.Mx)
ẋ=Vx−x(l.Vx)
ẋ=Ax−x.(l.Ax)
Identical
Some asymptotic behaviour

1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. **Neutrality and selection**
8. Computer simulation of evolution



Population size $N_e = 10000$, s = 0

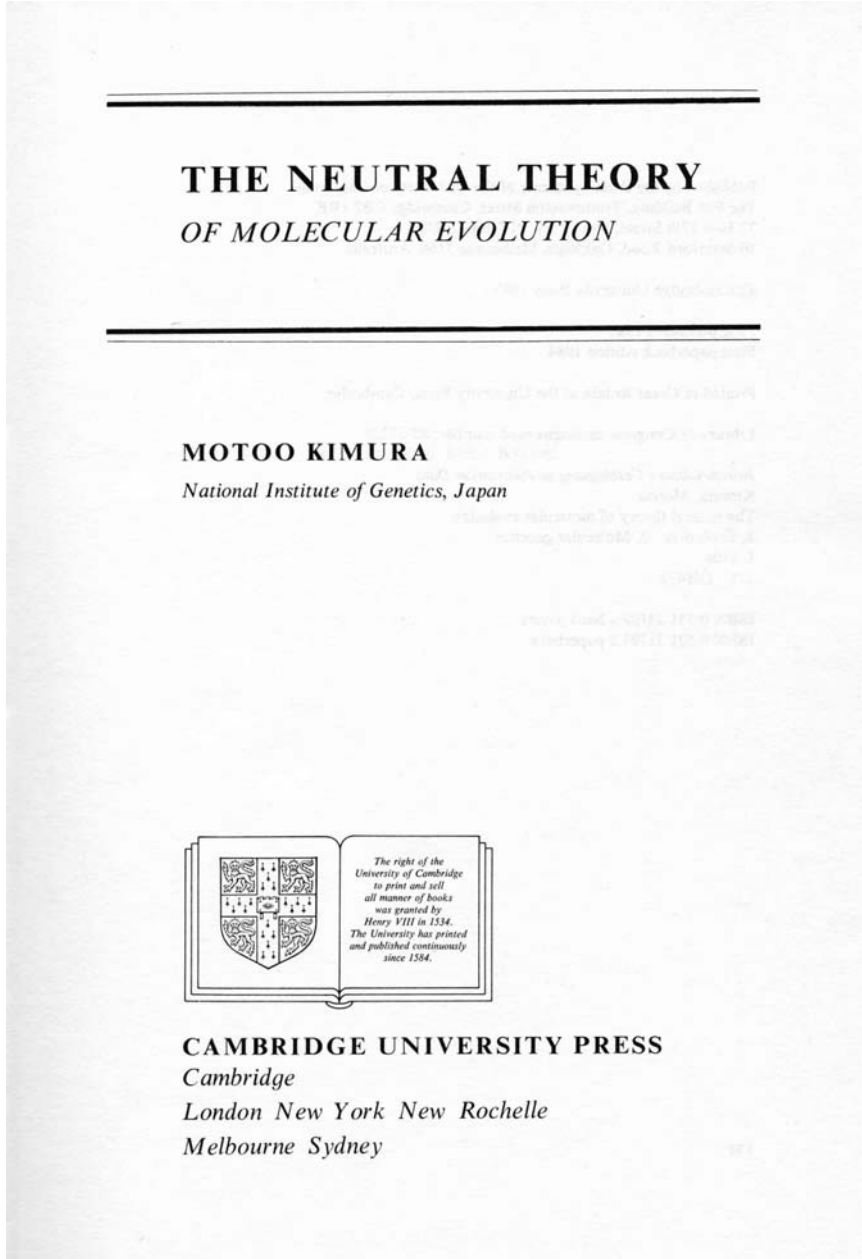
Stochastic population genetics of neutral, asexually reproducing species

MOTOO KIMURA

Motoo Kimura's 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.

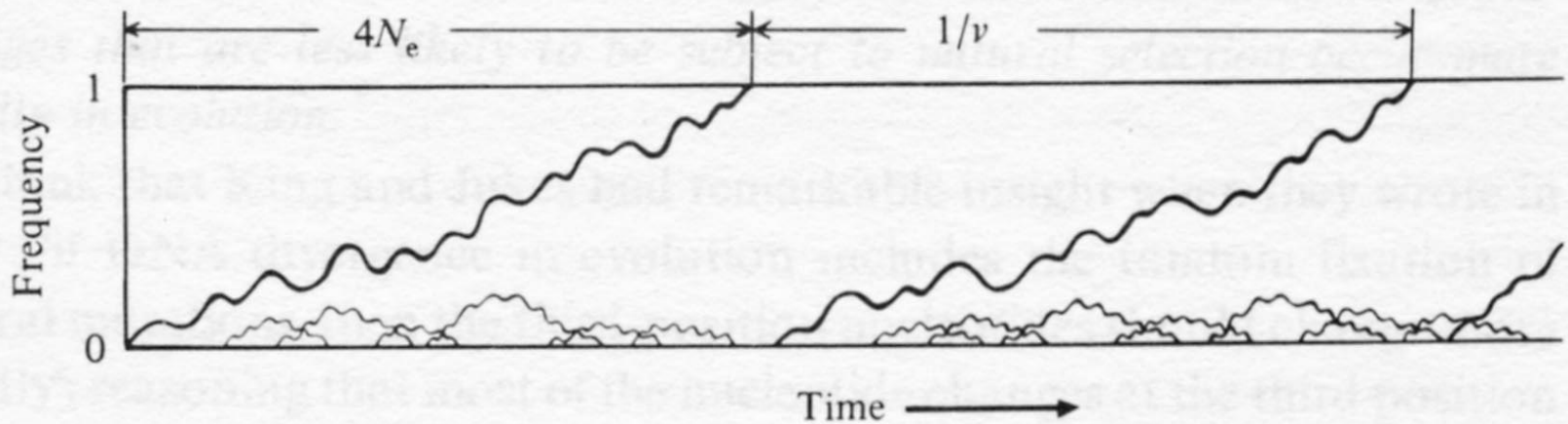
CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York New Rochelle

Melbourne Sydney

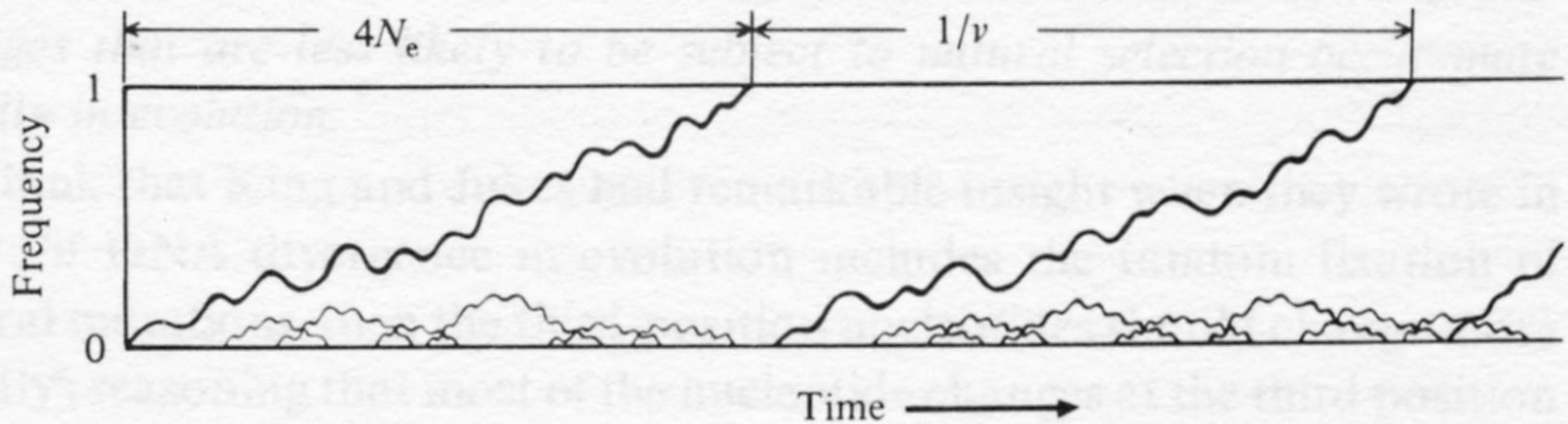
Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



The average time of replacement of a dominant genotype in a population is the reciprocal mutation rate, 1/ν, and therefore independent of population size.

Fixation of mutants in neutral evolution (Motoo Kimura, 1955)

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.

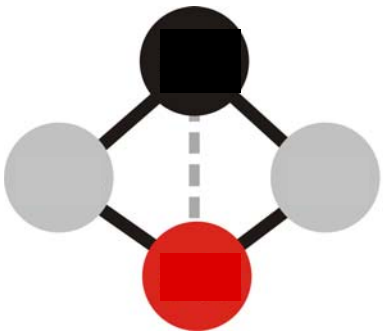


Is the Kimura scenario correct for frequent mutations?

Fixation of mutants in neutral evolution (Motoo Kimura, 1955)

$\mathbf{d_H = 1}$

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



$\mathbf{d_H = 2}$

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

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

$\mathbf{d_H \geq 3}$

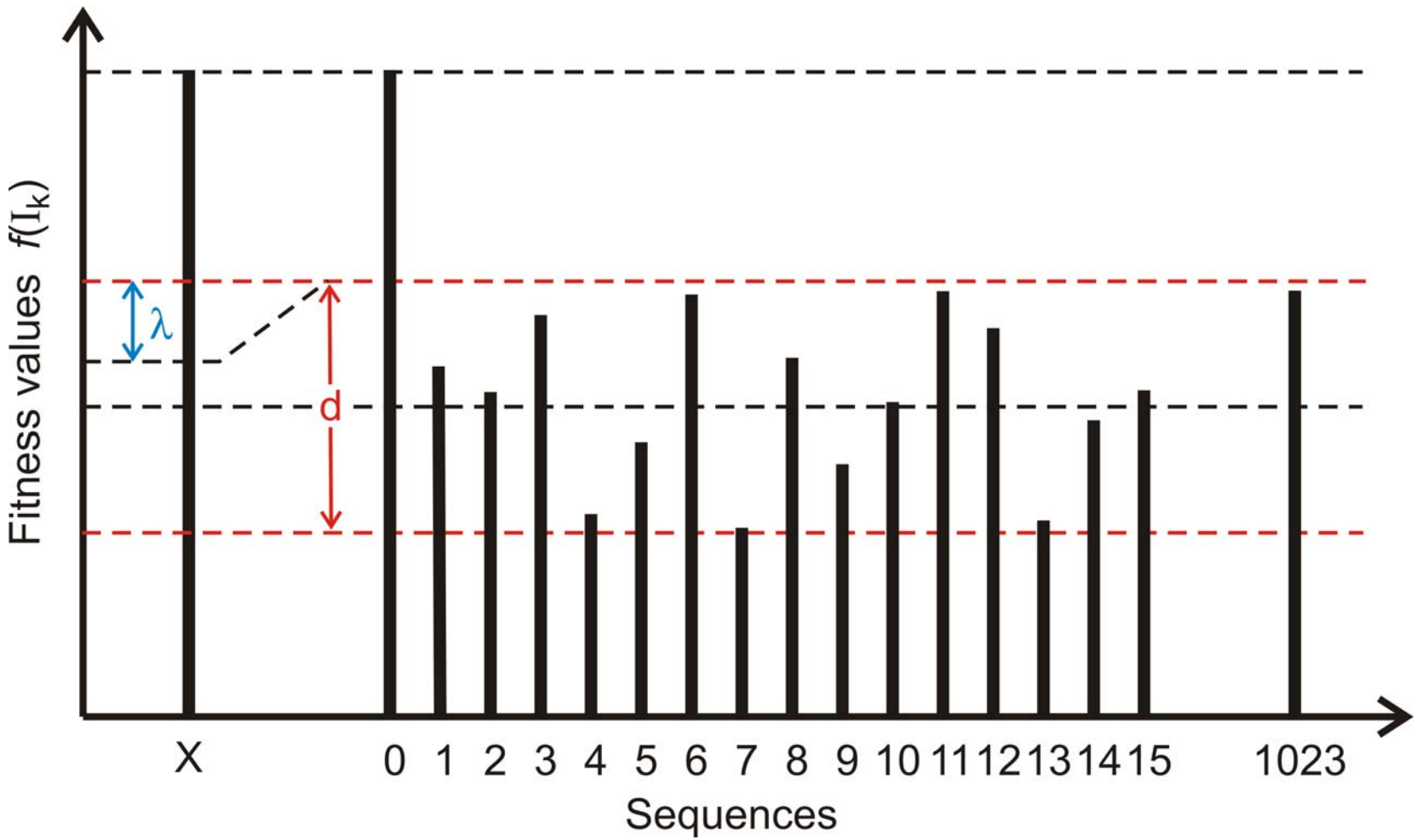
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

Pairs of neutral sequences in replication networks

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

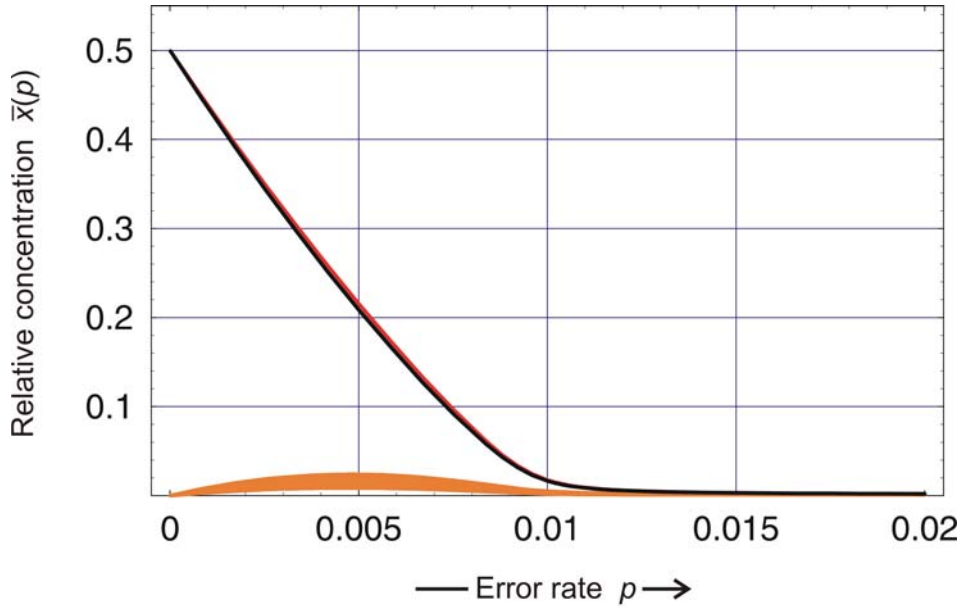
Random fixation in the sense of Motoo Kimura

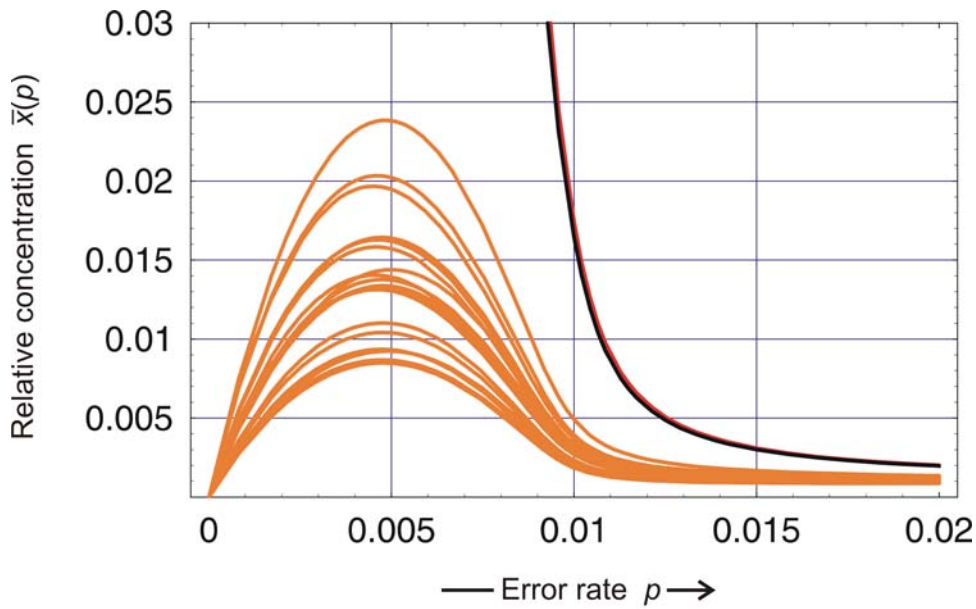


A fitness landscape including neutrality

Neutral network

$\lambda = 0.01,\ s = 367$

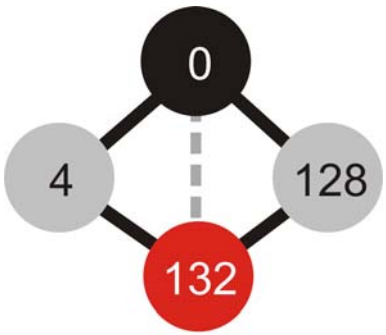




Neutral network: Individual sequences

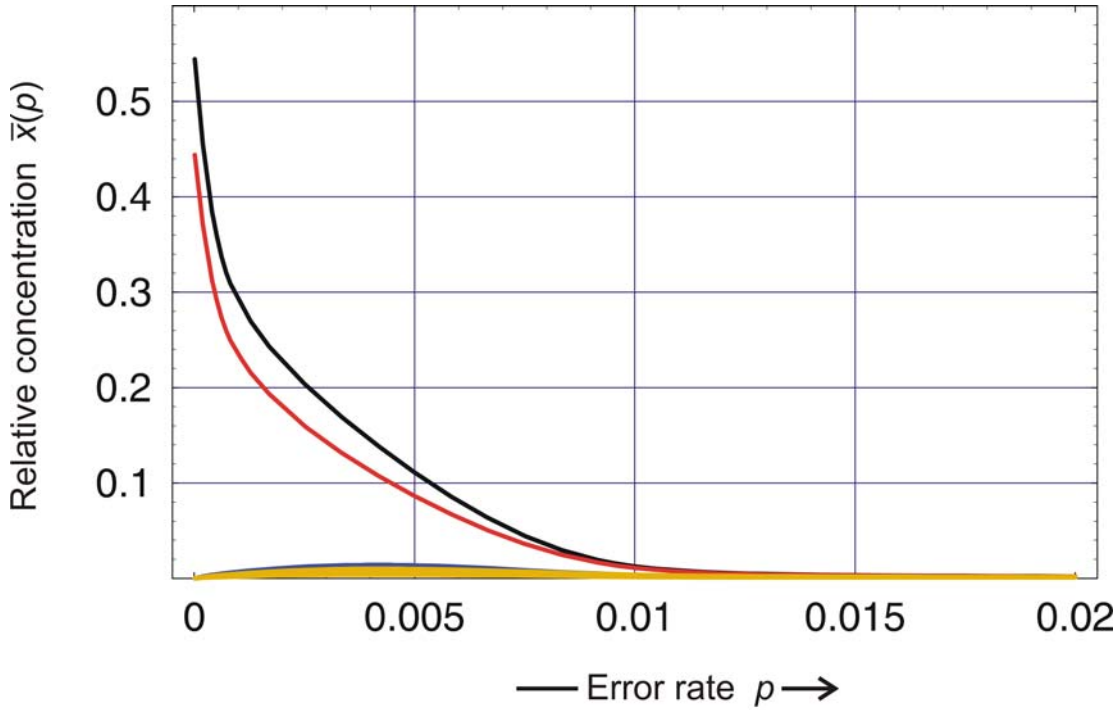
$n = 10,\ \sigma = 1.1,\ d = 1.0$

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



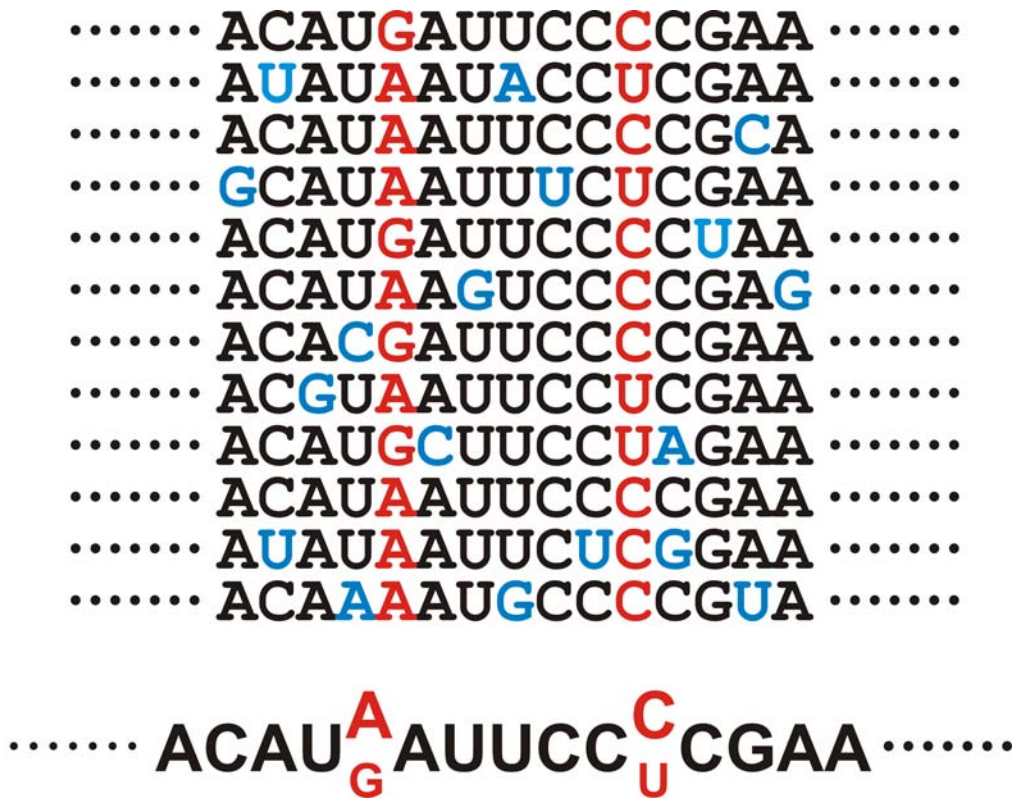
Neutral network

$\lambda = 0.01,\ s = 877$

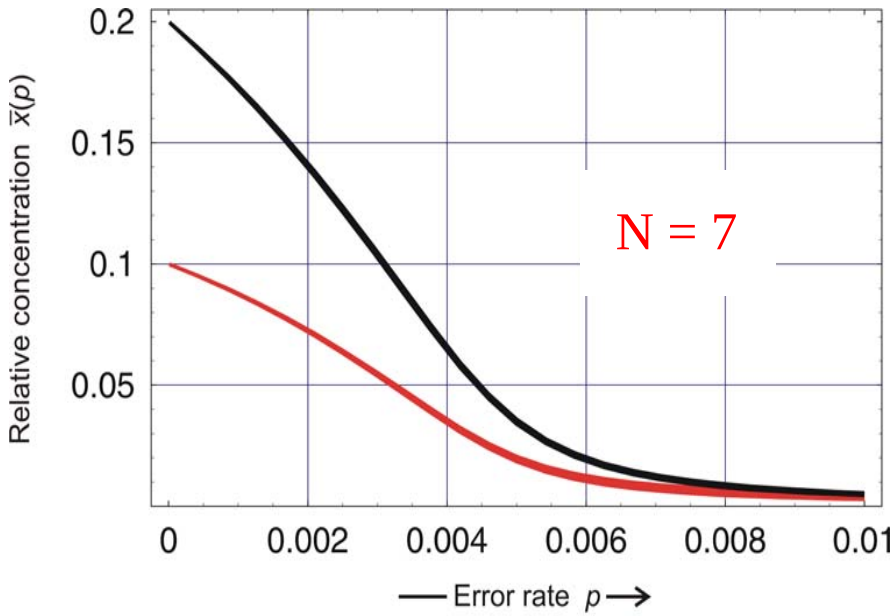


Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0

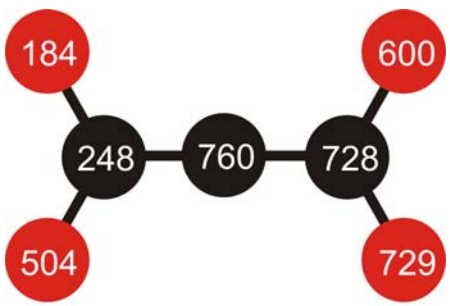


Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 2$.



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

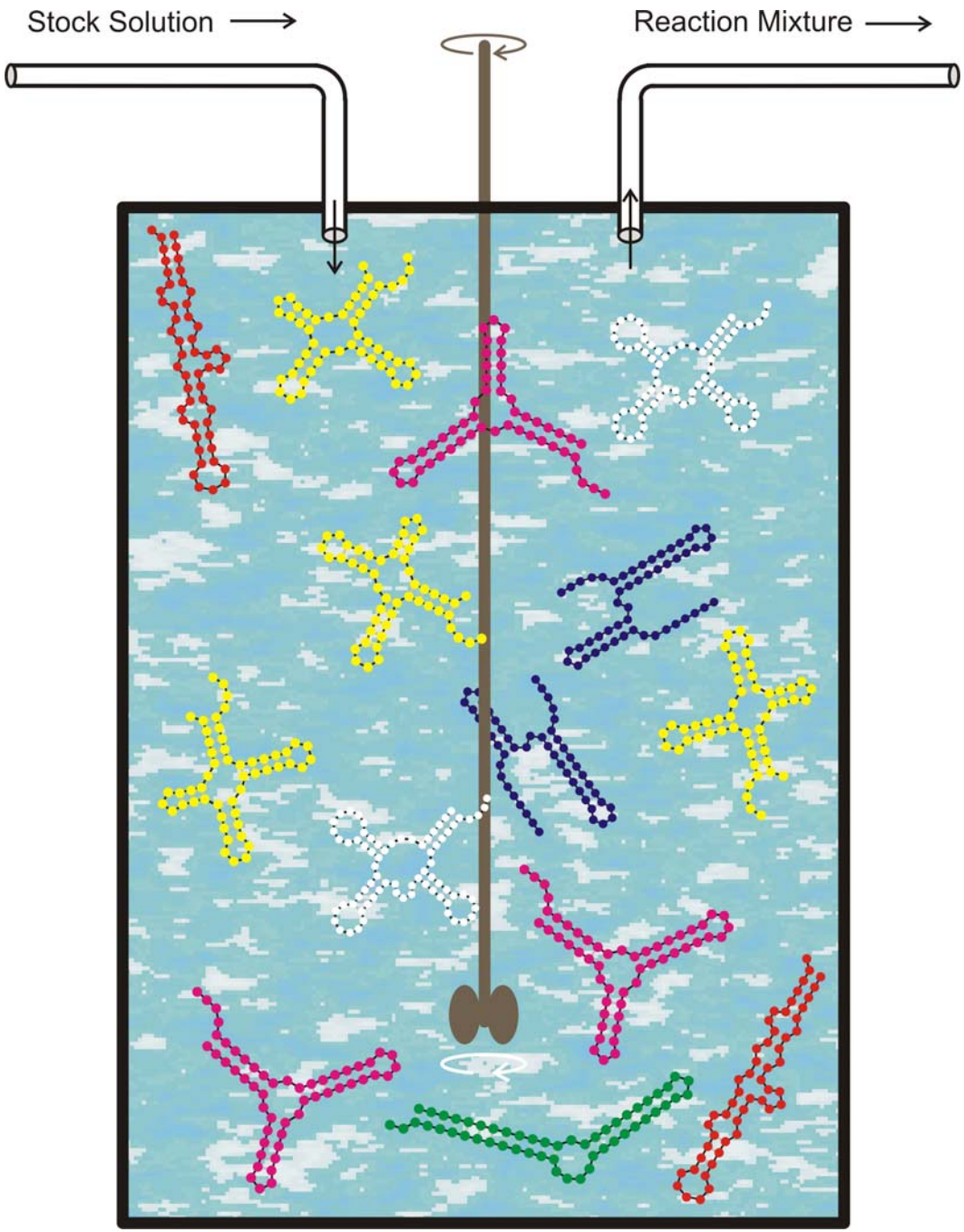
Adjacency matrix

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

1. Exponential growth and selection
2. Evolution as replication and mutation
3. A phase transition in evolution
4. Fitness landscapes as source of complexity
5. Molecular landscapes from biopolymers
6. The role of stochasticity
7. Neutrality and selection
8. **Computer simulation of evolution**



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.

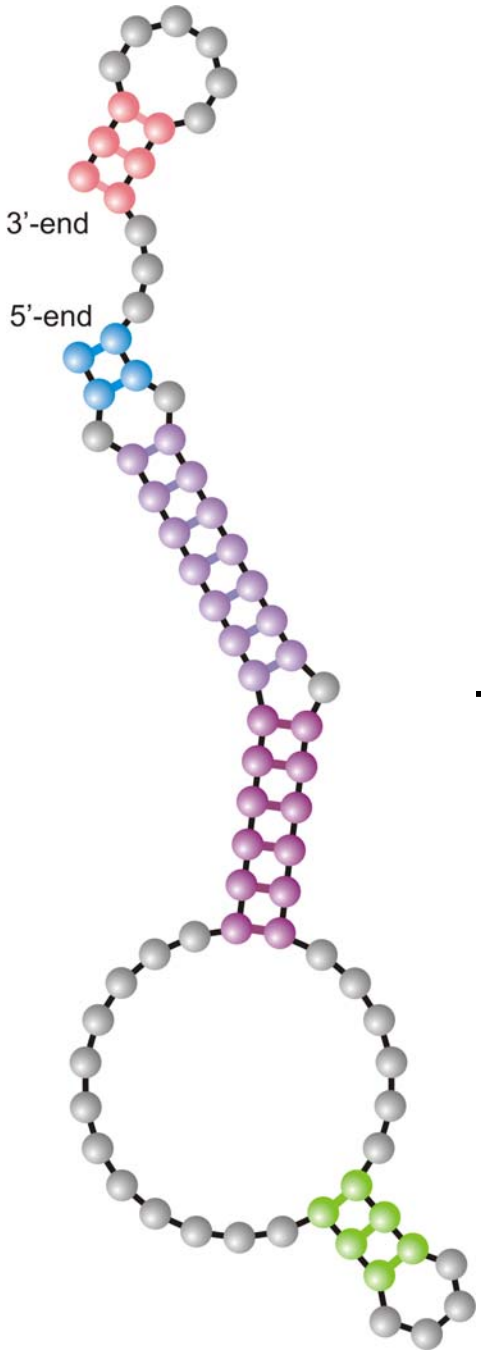
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level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

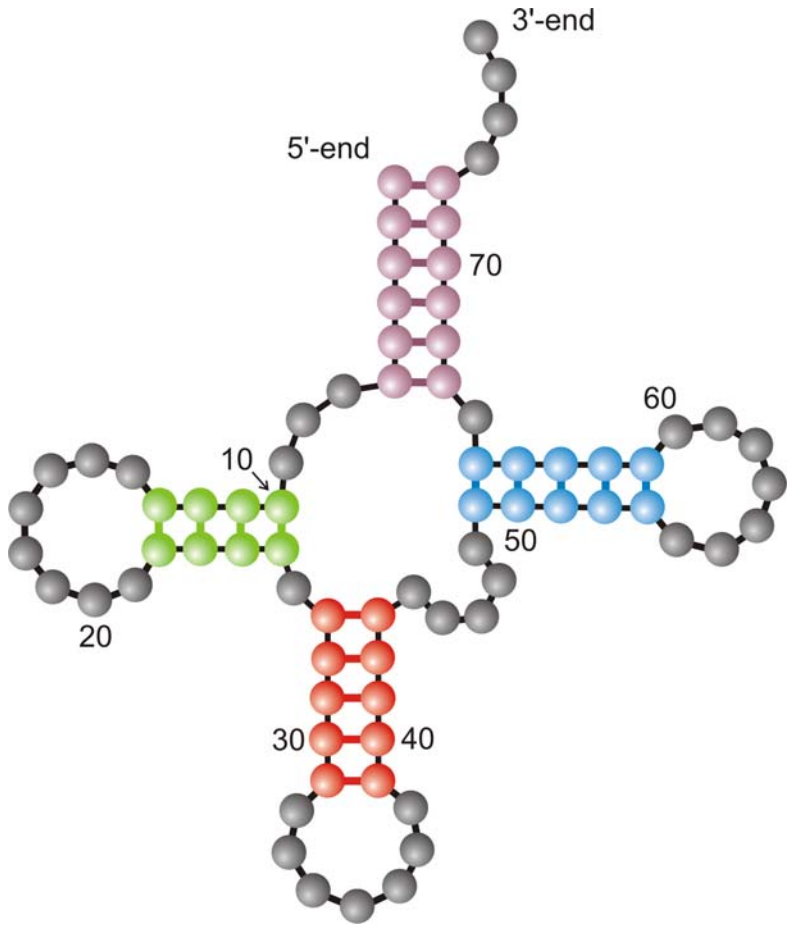
To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (*5*, *6*). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (*4*). Although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

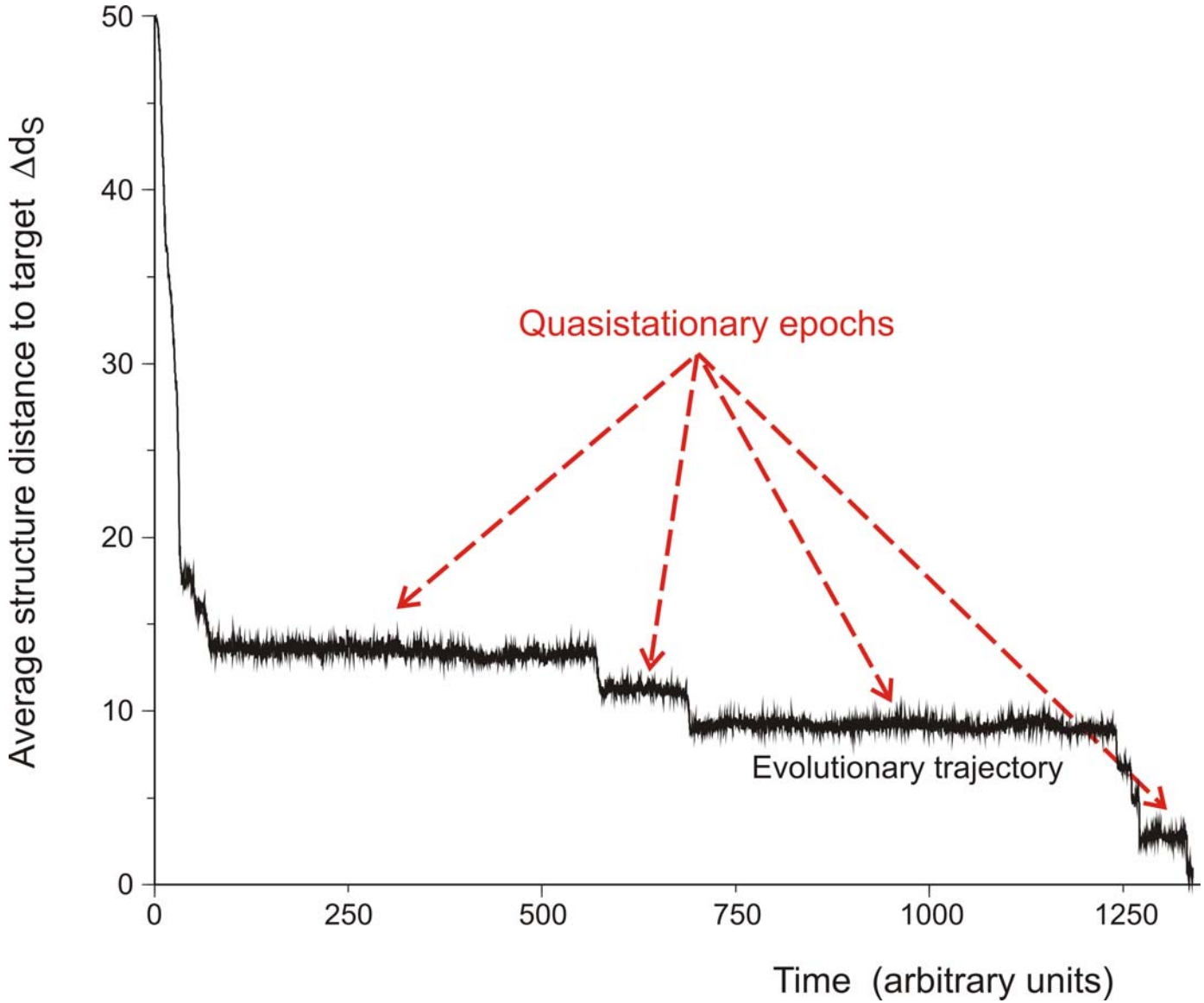
Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.





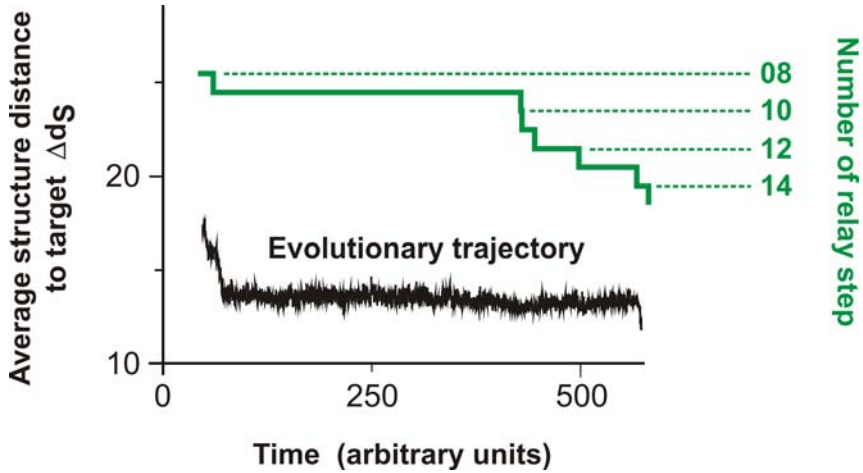
Structure of randomly chosen initial sequence

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

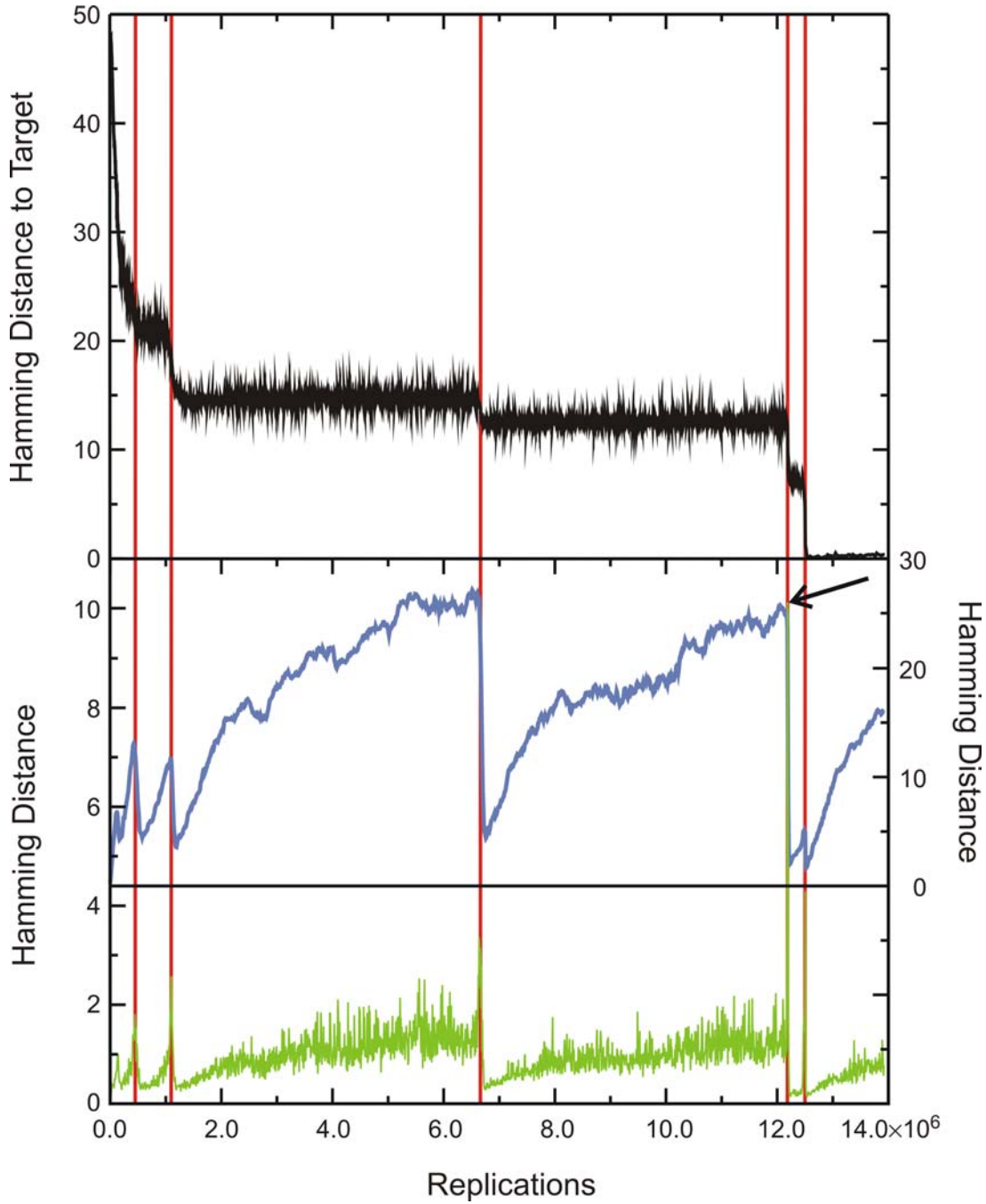
Neutral point mutations leave the molecular structure unchanged

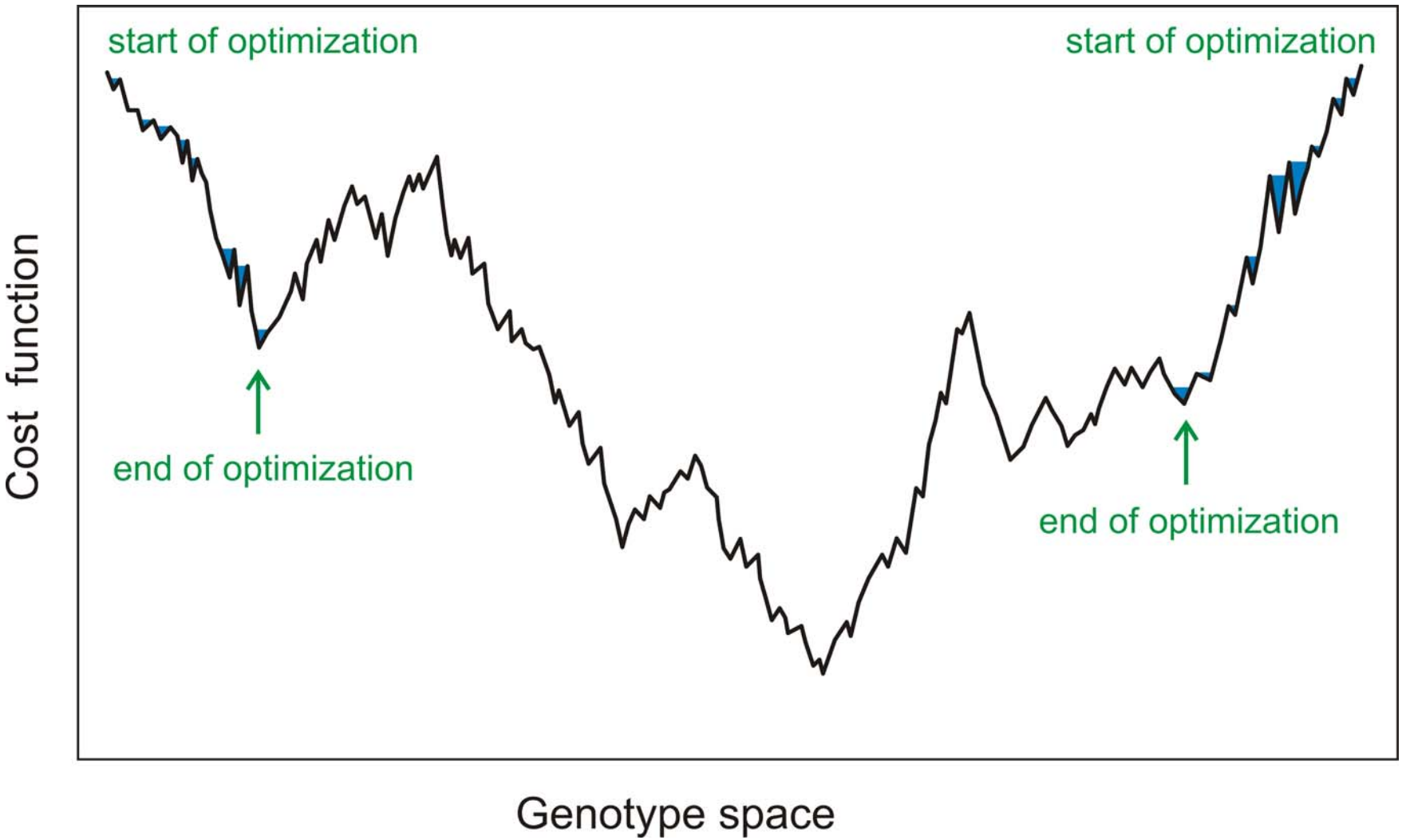
Neutral genotype evolution during phenotypic stasis

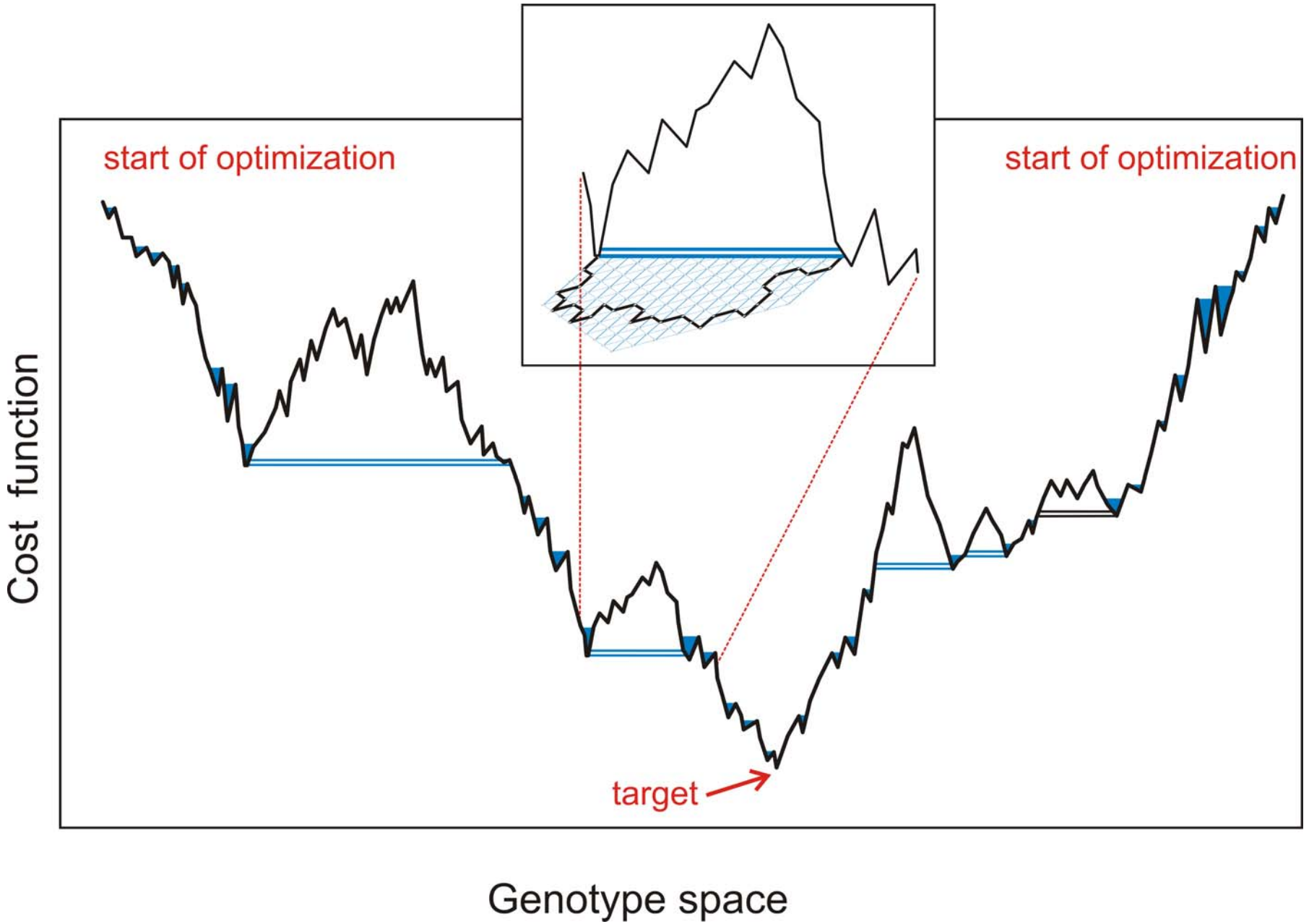
Evolutionary trajectory

Spreading of the population on neutral networks

Drift of the population center in sequence space




start of optimization
start of optimization
end of optimization
end of optimization
Cost function
Genotype space


start of optimization
start of optimization
target
Cost function
Genotype space

Coworkers

Universität Wien

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

Walter Fontana, Harvard Medical School, MA

Ivo L.Hofacker, **Christoph Flamm**, Universität Wien, AT

Martin Nowak, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

Christian Forst, Los Alamos National Laboratory, NM

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

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

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