

How computation has changed research in chemistry and biology

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria

and

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



IWR - 25 Jahre-Jubiläum

Heidelberg, 21. – 22.02.2013

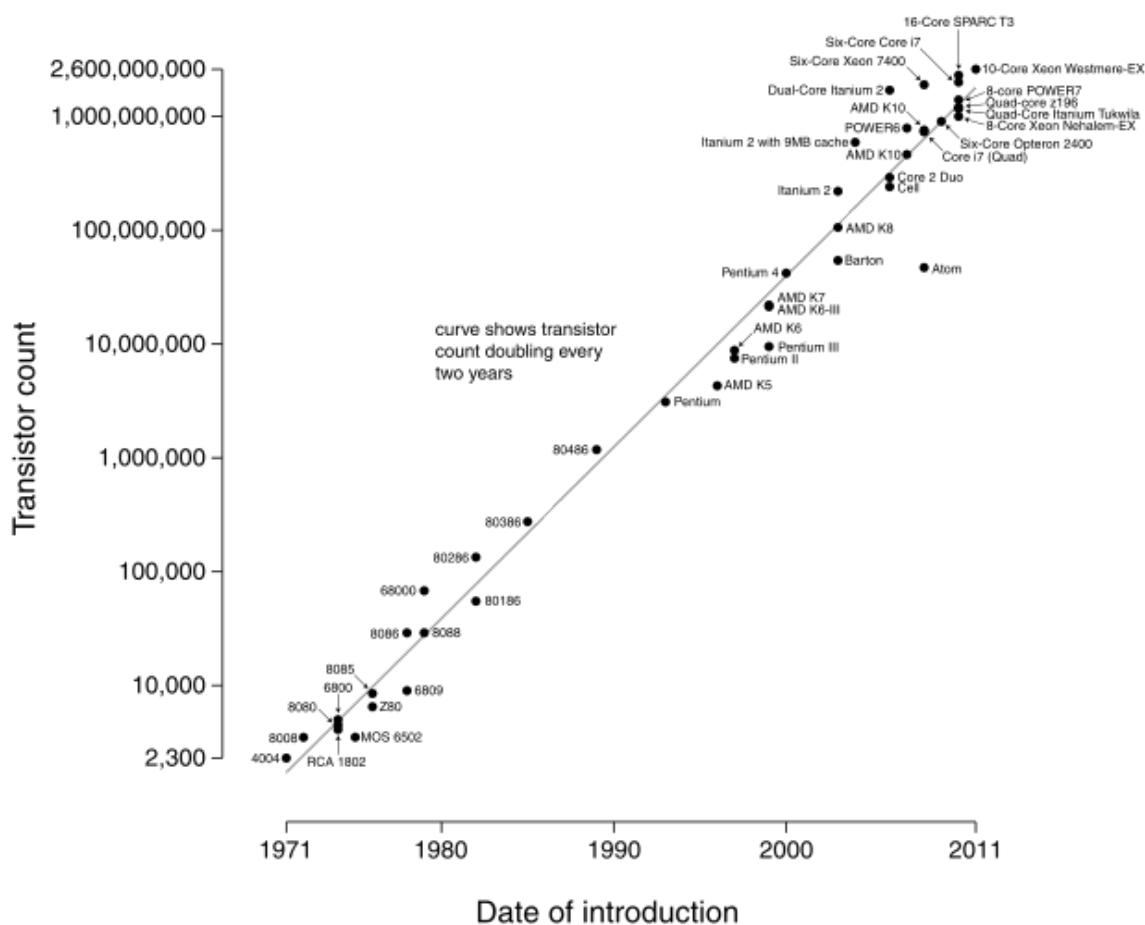
Web-Page for further information:

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

Some technological revolutions in 20th century science:

1. molecular spectroscopy,
2. micro-technology,
3. electronic computation,
4. molecular revolution in biology,
5. computational quantum chemistry, and
6. holistic chemistry of biological entities.

Microprocessor Transistor Counts 1971-2011 & Moore's Law



Gordon E. Moore, 1929 -

The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

Exponential increase in hardware power

Electronics 38 (8), 4-7, 1965

... Grötschel, an expert in optimization, observes that a benchmark production planning model solved using linear programming would have taken **82 years to solve in 1988**, using the computers and the linear programming algorithms of the day. Fifteen years later - **in 2003** - the same model could be solved in roughly **1 minute**, an improvement by a factor of roughly **43 million**.



Of this, a factor of roughly **1000** was due to increased **processor speed**, whereas a factor of roughly **43000** was due to improvements in **algorithms** !

Martin Grötschel, 1948 -

Grötschel also cites an algorithmic improvement of roughly **30000** for mixed integer programming between **1991** and **2008**.

PCIT Report to the President, 2010. **Progress in Algorithms Beats Moore's Law.**

J.P. Holdren, E. Lander, H. Varmus. Designing a digital future: Federally funded research and development in networking and information technology. President's council on science and technology, Washington, DC, p.71, 2010

Four selected examples

1. Parameter determination in chemical kinetics
2. Design of ribonucleic acid (RNA) structures
3. Kinetic folding of RNA molecules
4. Modeling evolution

Four selected examples

1. Parameter determination in chemical kinetics
2. Design of ribonucleic acid (RNA) structures
3. Kinetic folding of RNA molecules
4. Modeling evolution

L. Michaelis, M. Menten. Die Kinetik der Invertin-Wirkung. *Biochemische Zeitschrift* **49**, 333-369, 1913



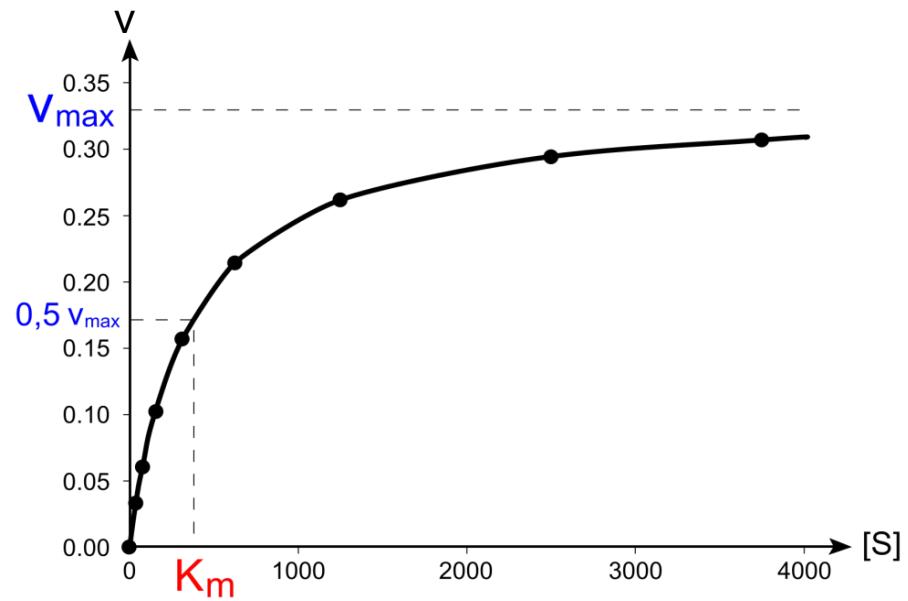
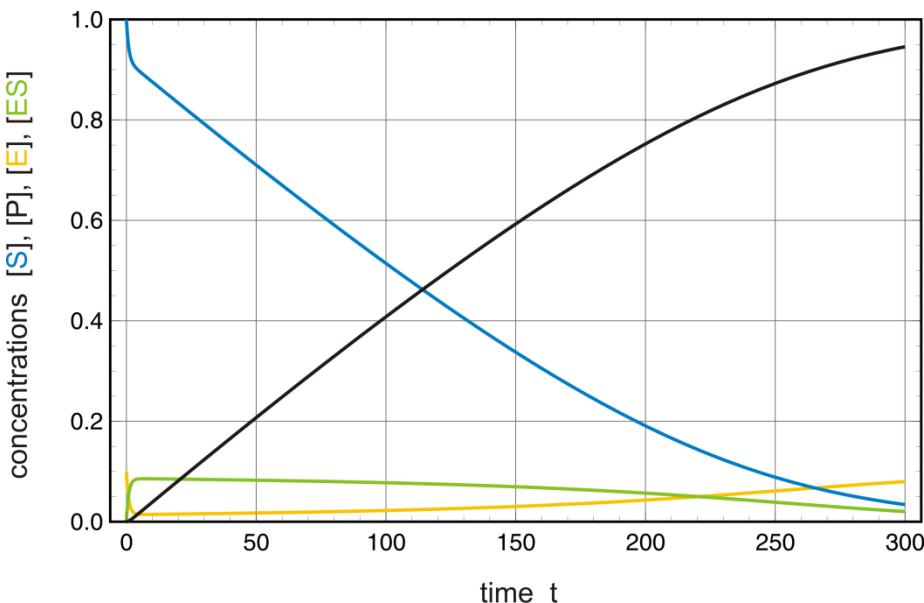
basic assumptions: $k_r < k_d$

$$[E]_0 \ll [S]_0$$

$$\frac{d[P]}{dt} = v([S]) = \frac{v_{\max} \cdot [S]}{K_M + [S]}$$

$$K_M = \frac{k_r + k_d}{k_f},$$

$$v([S]) = k_r \cdot [ES] \quad \text{and} \quad v_{\max} = k_r \cdot [E]_0$$



Michaelis-Menten mechanism of enzyme reactions

Linearization of a hyperbola:

$$v([S]) = \frac{v_{\max} \cdot [S]}{K_M + [S]}$$

Lineweaver-Burk:

$$1/v = f(1/[S])$$

Eadie-Hofstee:

$$v = f(1/[S])$$

Scatchard:

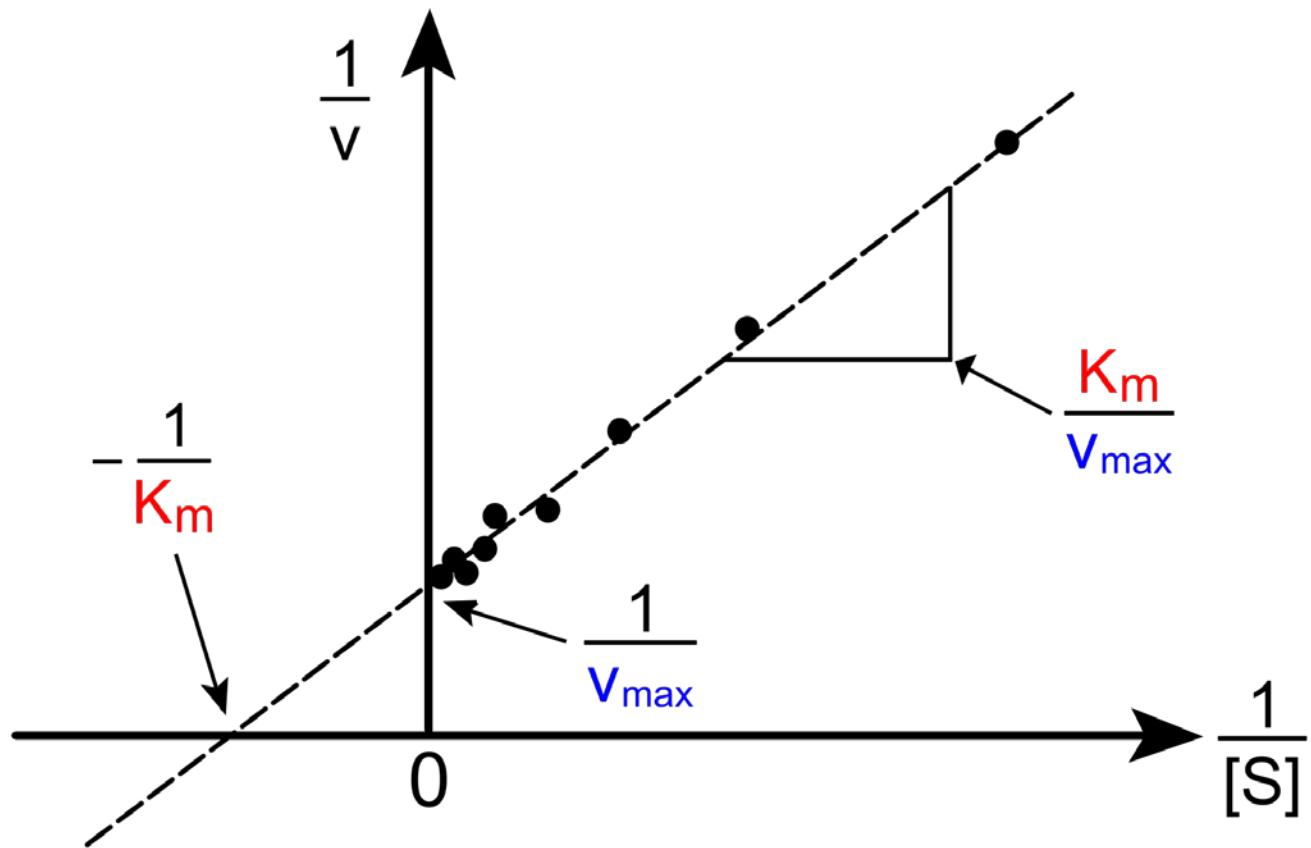
$$1/[S] = f(v)$$

Hanes:

$$[S] / v = f([S])$$

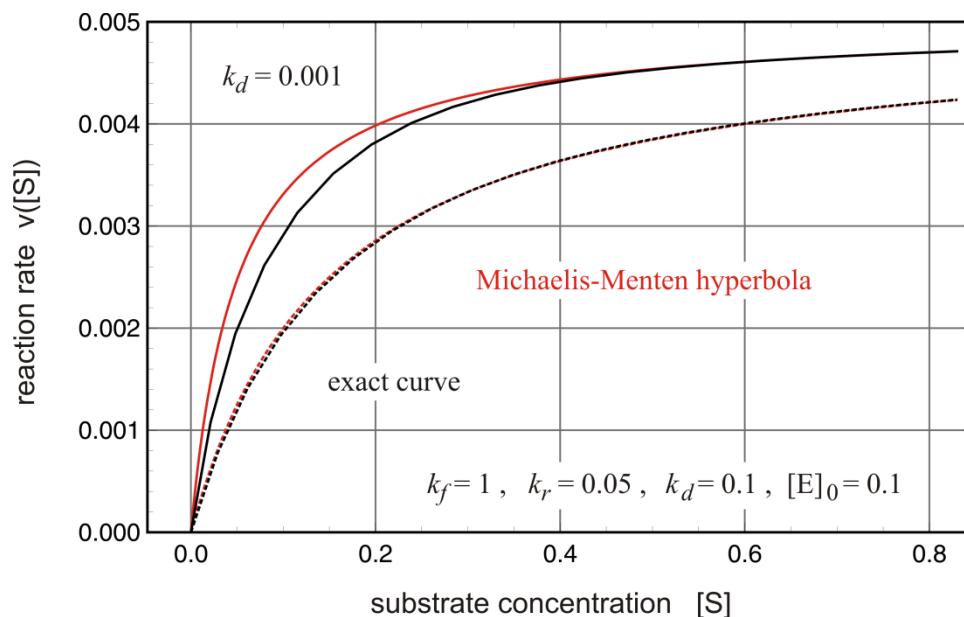
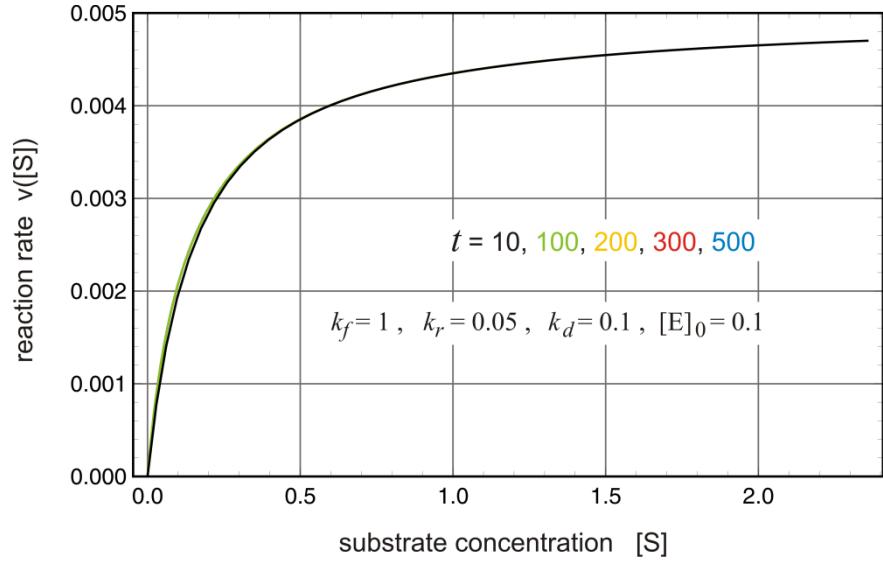
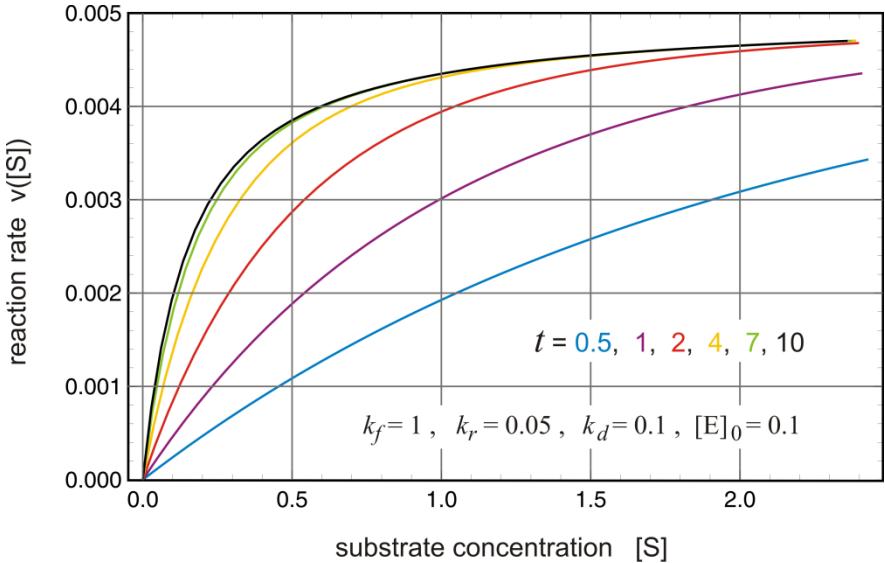
Hill:

$$\log \left(v/(v_{\max} - v) \right) = f(\log [S])$$



The Lineweaver-Burke plot of Michaelis-Menten kinetics

Source: Wikipedia, “Enzymkinetik”



Validity of the Michaelis-Menten approximation

Kinetic differential equations

$$\frac{dx}{dt} = f(x; k); x = (x_1, \dots, x_n); k = (k_1, \dots, k_m)$$

Reaction diffusion equations

$$\frac{\partial x}{\partial t} = D \nabla^2 x + f(x; k)$$

Parameter set

$$k_j(T, p, pH, I, \dots); j=1, 2, \dots, m$$

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

Initial conditions : $x(0)$

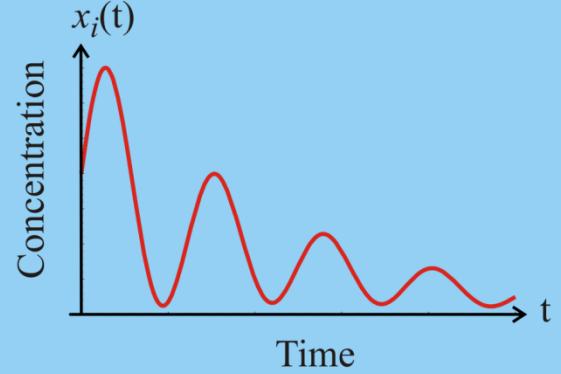
Boundary conditions :

boundary ... S , normal unit vector ... \hat{u}

Dirichlet : $x^S = g(r, t)$

Neumann : $\frac{\partial x}{\partial u} = \hat{u} \cdot \nabla x^S = g(r, t)$

Solution curves: $x(t)$



The forward problem of chemical reaction kinetics

Parameter identification and determination
is an ill-posed problem

Inverse problem solution techniques

Parameter set
 $k_j(I_G; T, p, pH, l, \dots); j=1, 2, \dots, m$

Kinetic differential equations

$$\frac{dx}{dt} = f(x; k); x=(x_1, \dots, x_n); k=(k_1, \dots, k_m)$$

Reaction diffusion equations

$$\frac{\partial x}{\partial t} = D \nabla^2 x + f(x; k)$$

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

Initial conditions : x(0)

Boundary conditions :

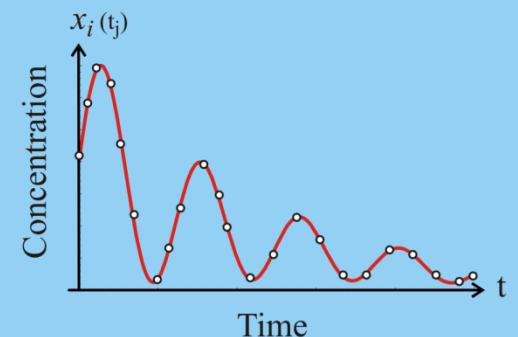
boundary ... S , normal unit vector ... \hat{u}

Dirichlet : $x^S = g(r, t)$

Neumann : $\frac{\partial x}{\partial u} = \hat{u} \cdot \nabla x^S = g(r, t)$

Data from measurements

$x(t_j); j = 1, 2, \dots, N$



The inverse problem of
chemical reaction kinetics

$$F(\vec{q}) = \vec{y}^\delta$$

$$\left\| \vec{y}^\delta - F(\vec{q}) \right\|_Y^2 \rightarrow \min_{\vec{q} \in Q} \quad \text{ill-conditioned problem}$$

$$\left\| \vec{y}^\delta - F(\vec{q}) \right\|_Y^2 + \alpha \mathcal{R}(\vec{q}, \vec{q}_0) \rightarrow \min_{\vec{q} \in Q} \quad \text{with} \quad \mathcal{R}(\vec{q}, \vec{q}_0) = \left\| \vec{q} - \vec{q}_0 \right\|_Q^2$$

regularization term \mathcal{R} - here Tikhonov regularization - with q_0 being an initial parameter guess and α the regularization parameter

Parameter identification and determination as an inverse problem

TOPICAL REVIEW

Inverse problems in systems biology

Heinz W Engl¹, Christoph Flamm², Philipp Kögler³, James Lu¹,
Stefan Müller¹ and Peter Schuster²

¹ Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, 4040 Linz, Austria

² Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Wien, Austria

³ Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

E-mail: heinz.engl@oeaw.ac.at, xtof@tbi.univie.ac.at, philipp.kuegler@jku.at,
james.lu@oeaw.ac.at, stefan.mueller@oeaw.ac.at and pks@tbi.univie.ac.at

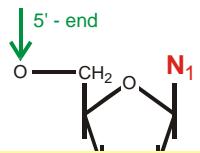
Received 7 July 2009, in final form 12 November 2009

Published 3 December 2009

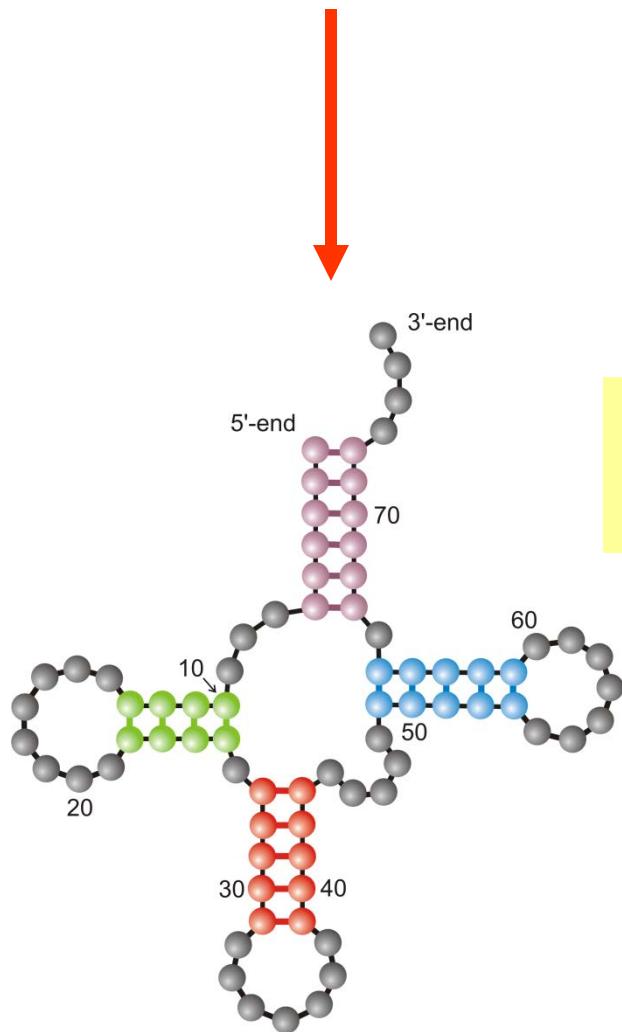
Online at stacks.iop.org/IP/25/123014

Four selected examples

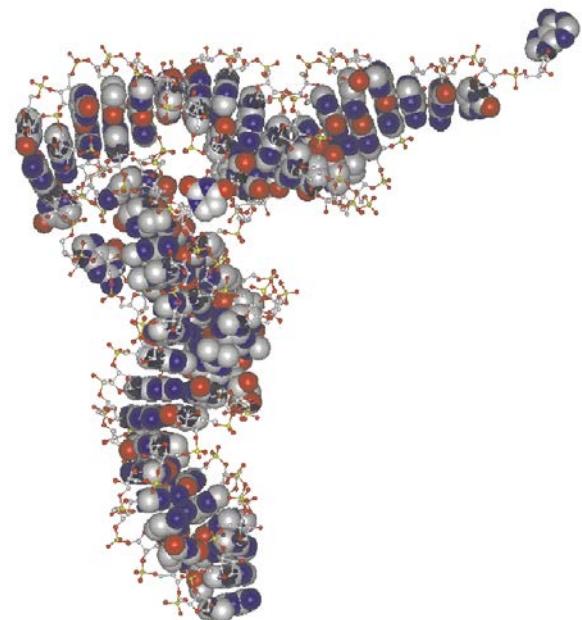
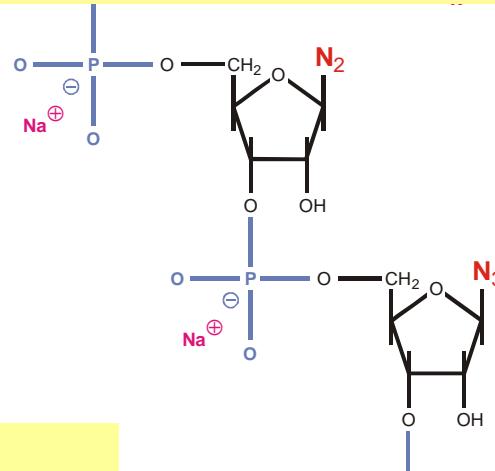
1. Parameter determination in chemical kinetics
2. Design of ribonucleic acid (RNA) structures
3. Kinetic folding of RNA molecules
4. Modeling evolution

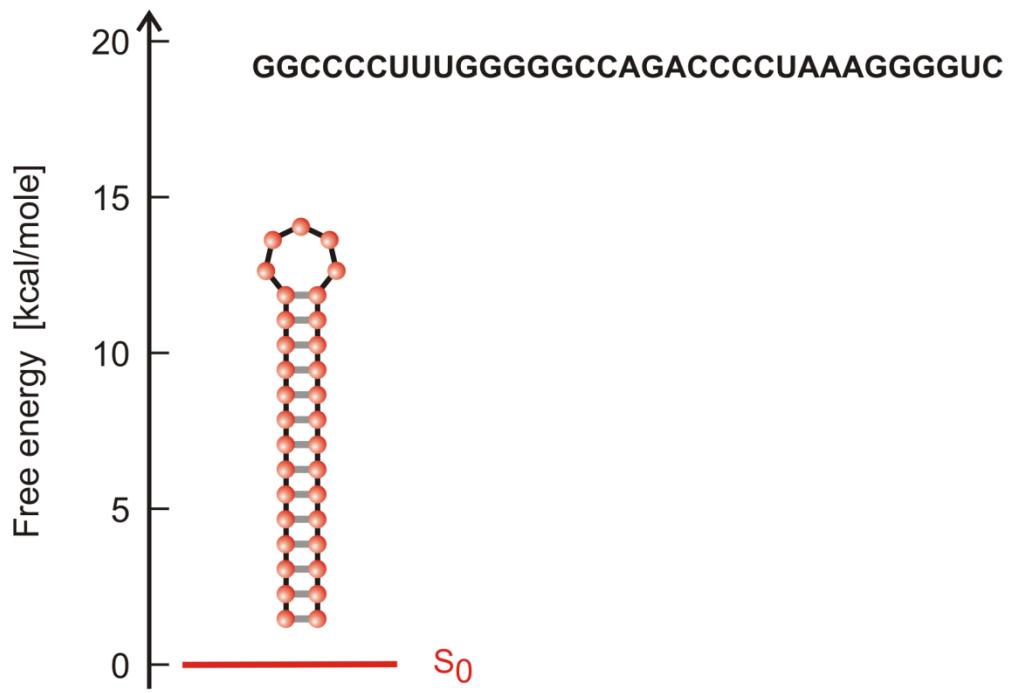


5'-end **GC****GGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA** 3'-end



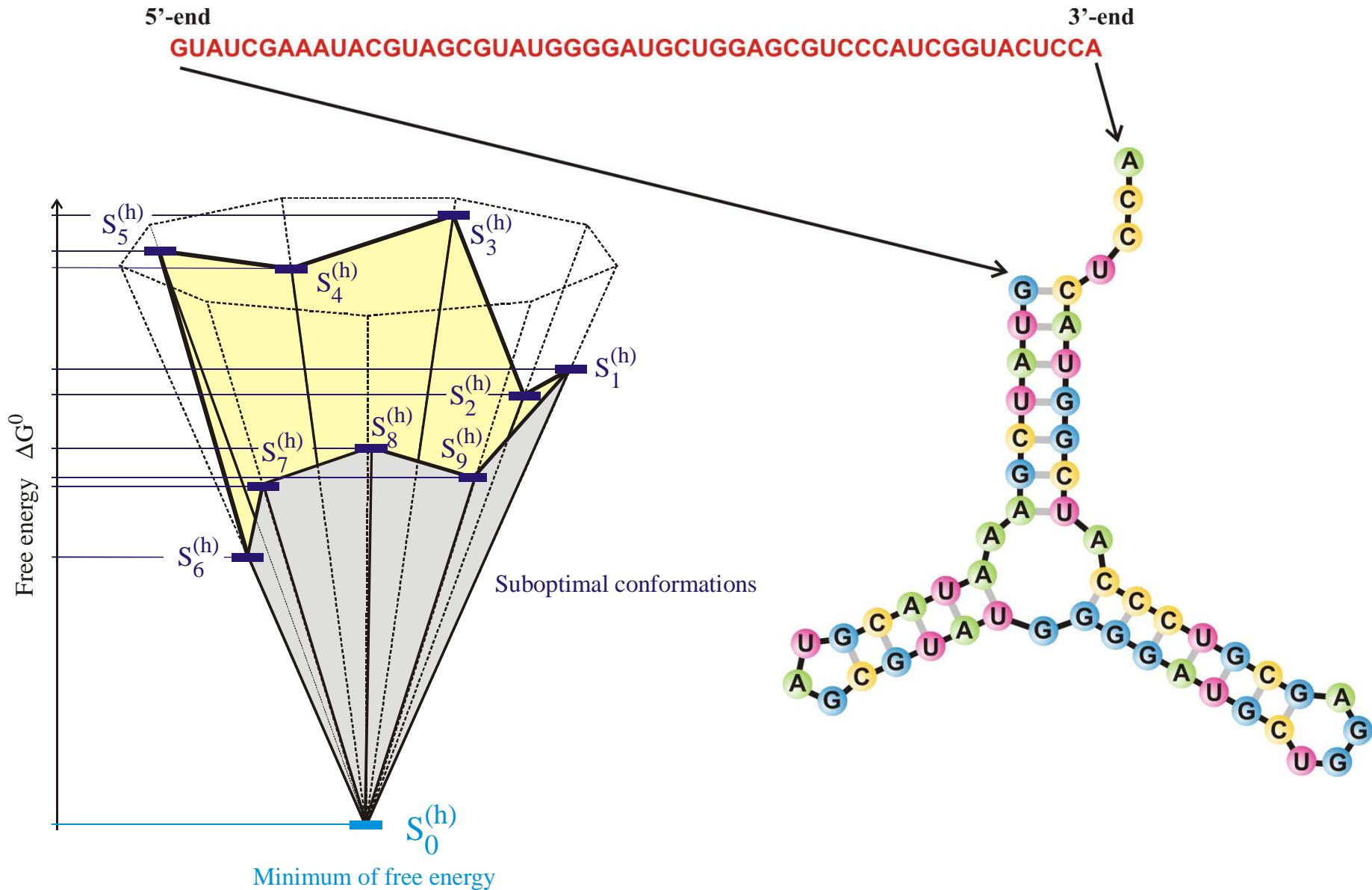
RNA structure
The molecular phenotype





Minimum free energy structure

The notion of structure



The minimum free energy structures on a discrete space of conformations

RNA sequence

GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

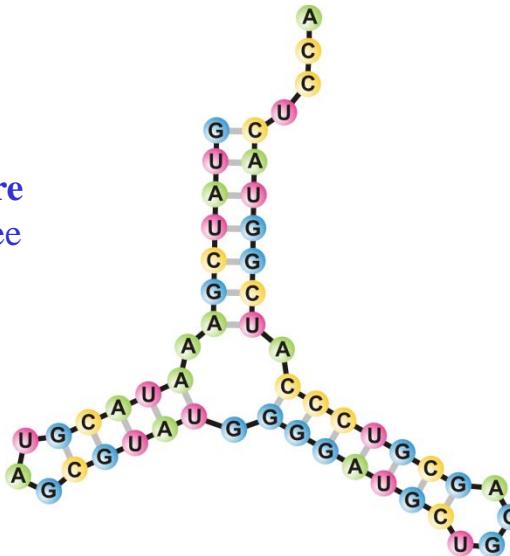
linear programming

RNA folding:
structural biology,
spectroscopy of
biomolecules,
understanding
molecular function

biophysical chemistry:
thermodynamics and
kinetics

empirical parameters

RNA structure
of minimal free
energy



From RNA sequence to structure

RNA sequence

GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

Linear programming

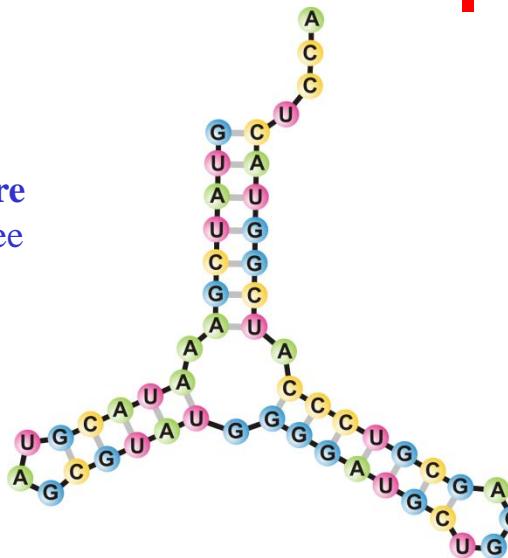
RNA folding:
Structural biology,
spectroscopy of
biomolecules,
understanding
molecular function

iterative determination
of a sequence for the
given secondary
structure

inverse Folding Algorithm

inverse folding of RNA:
biotechnology,
design of biomolecules
with predefined
structures and functions

RNA structure
of minimal free
energy



From RNA structure to sequence

ViennaRNA Package:

Ivo L. Hofacker, Walter Fontana, Peter F. Stadler, Sebastian Bonhoeffer,
Manfred Tacker, and Peter Schuster.

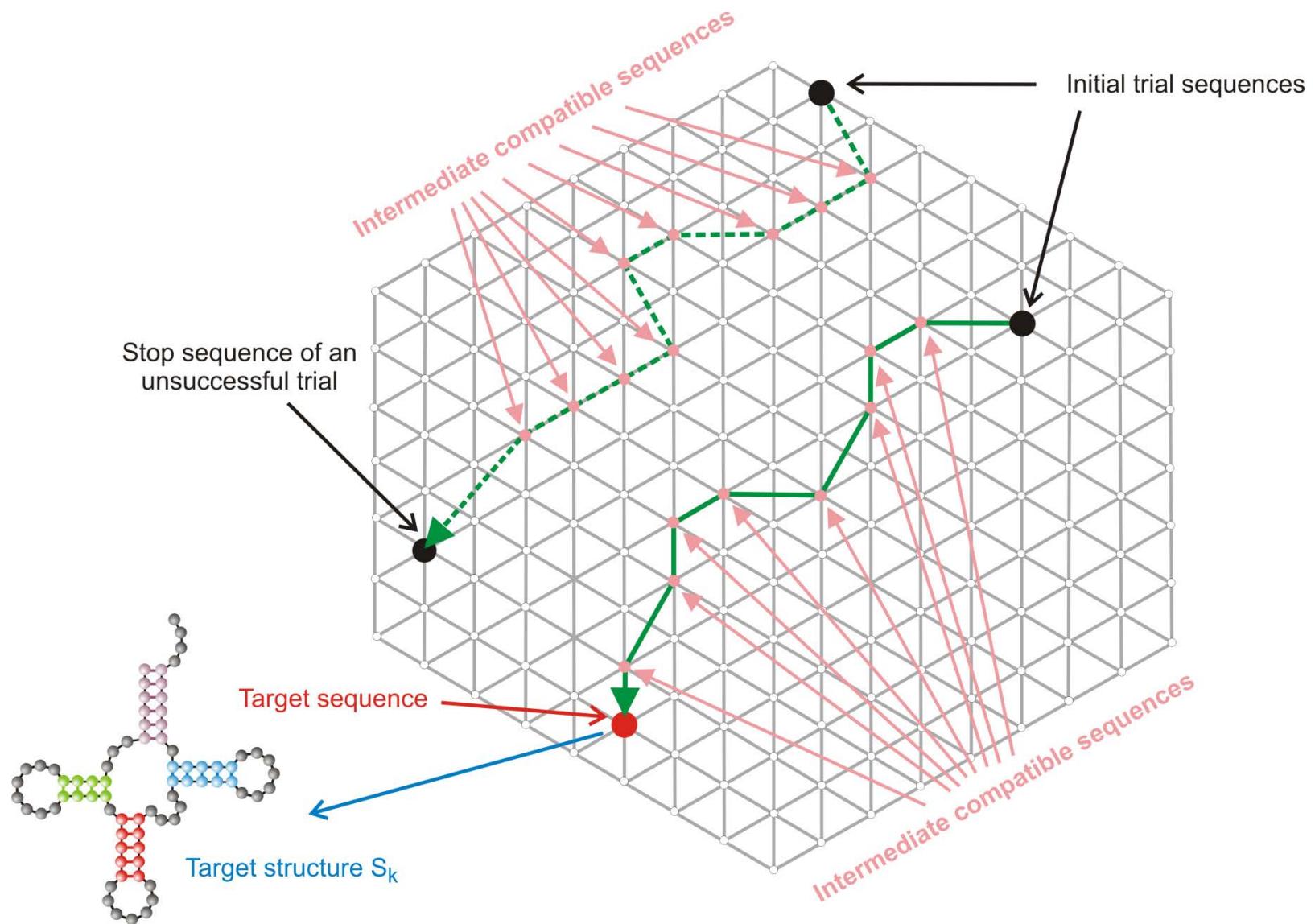
Fast folding and comparison of RNA secondary structures.

Mh.Chem. **125**:167-188, 1994

Ronny Lorenz, Stephan H. Bernhart, Christian Höner zu Siederssen,
Hakim Tafer, Christioh Flamm, Peter F. Stadler, and Ivo L. Hofacker.

ViennaRNA Package 2.0.

Algorithms Mol. Biol. **6**:26, 2011



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

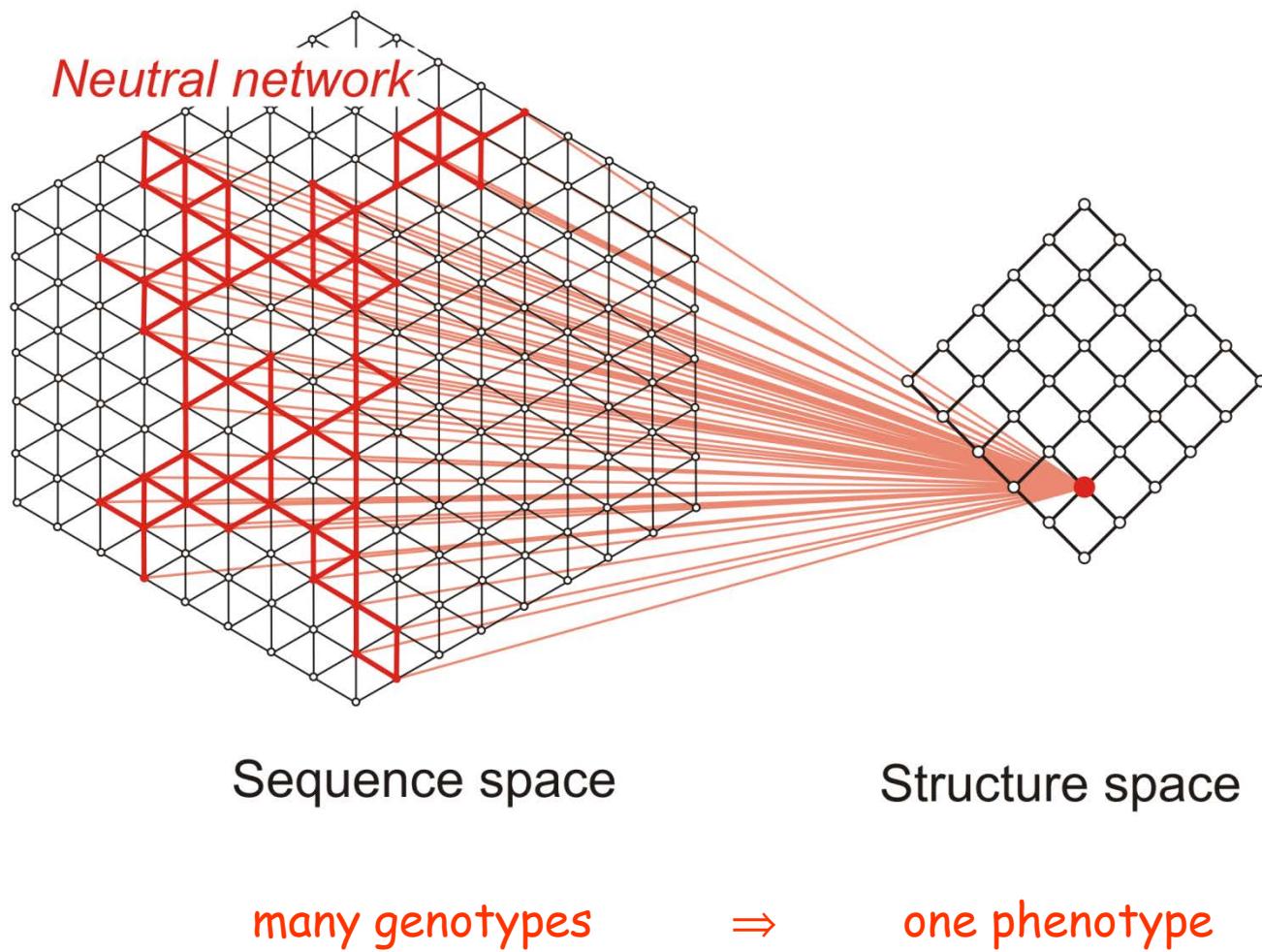
Space of phenotypes: $S = \{S_1, S_2, S_3, S_4, \dots, S_M\}$; metric (not required)

$$N \gg M$$

$$\psi(I_j) = S_k$$

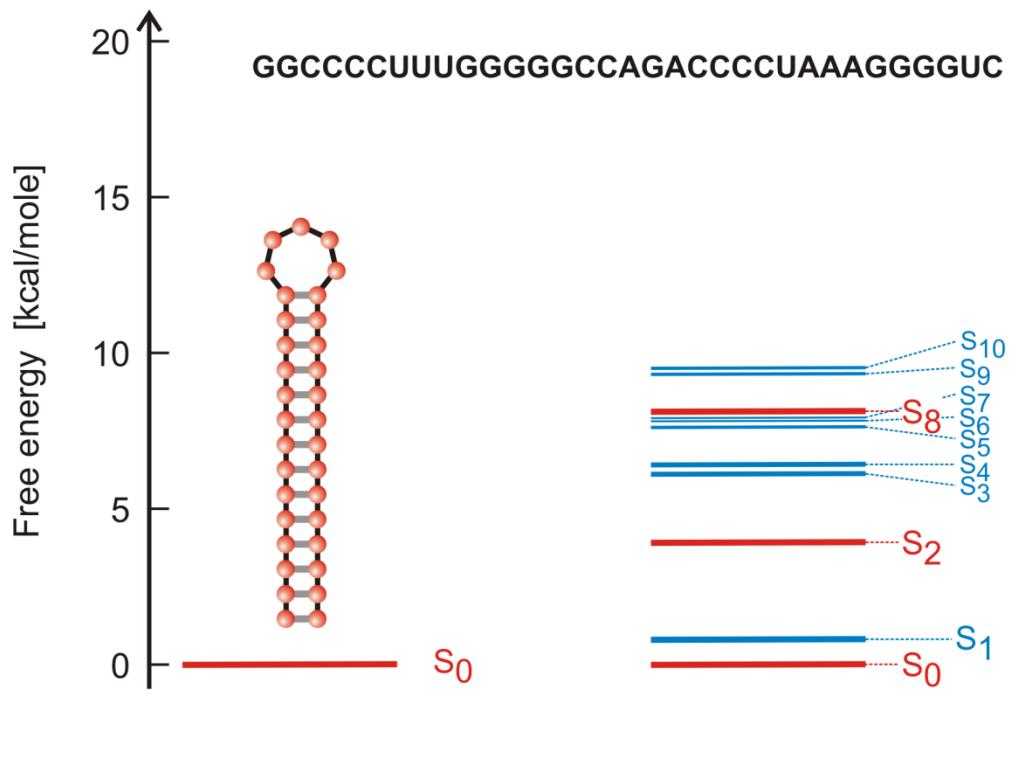
$$G_k = \psi^{-1}(S_k) = \{ I_j \mid \psi(I_j) = S_k \}$$

A mapping ψ and its inversion

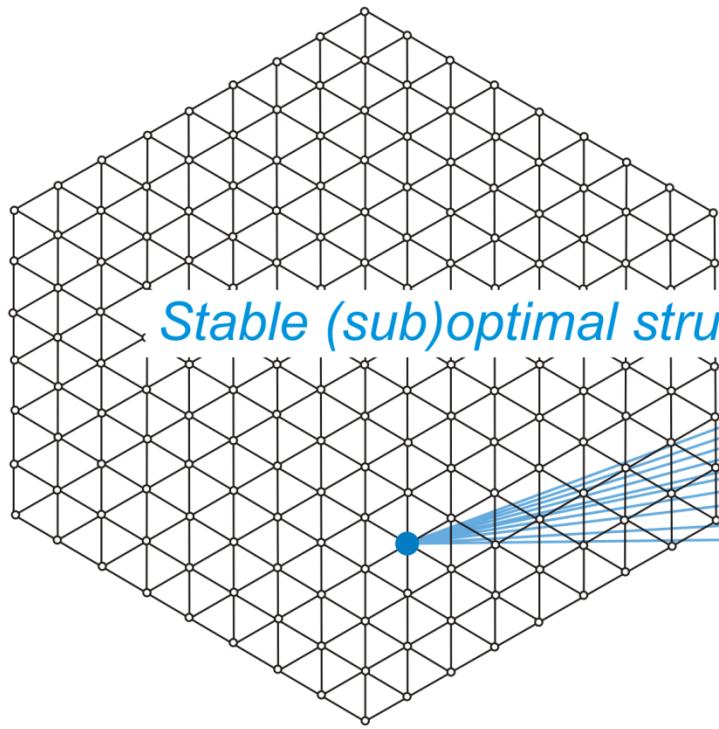


Four selected examples

1. Parameter determination in chemical kinetics
2. Design of ribonucleic acid (RNA) structures
- 3. Kinetic folding of RNA molecules**
4. Modeling evolution

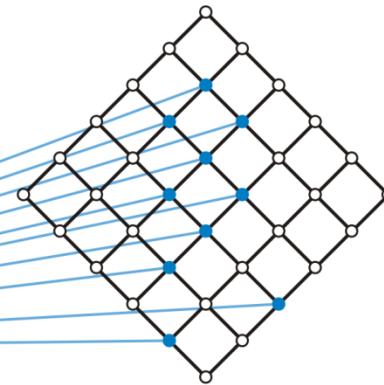


Extension of the notion of structure

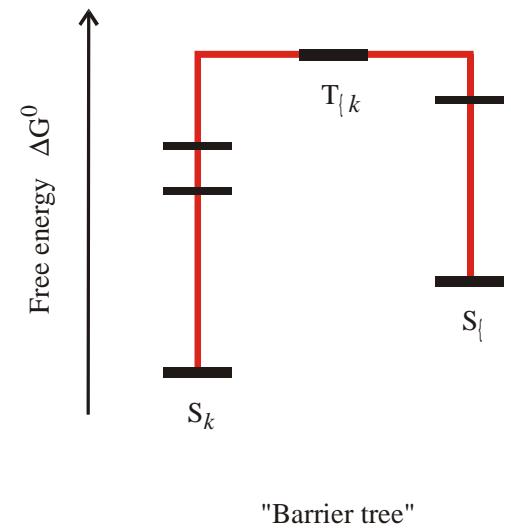
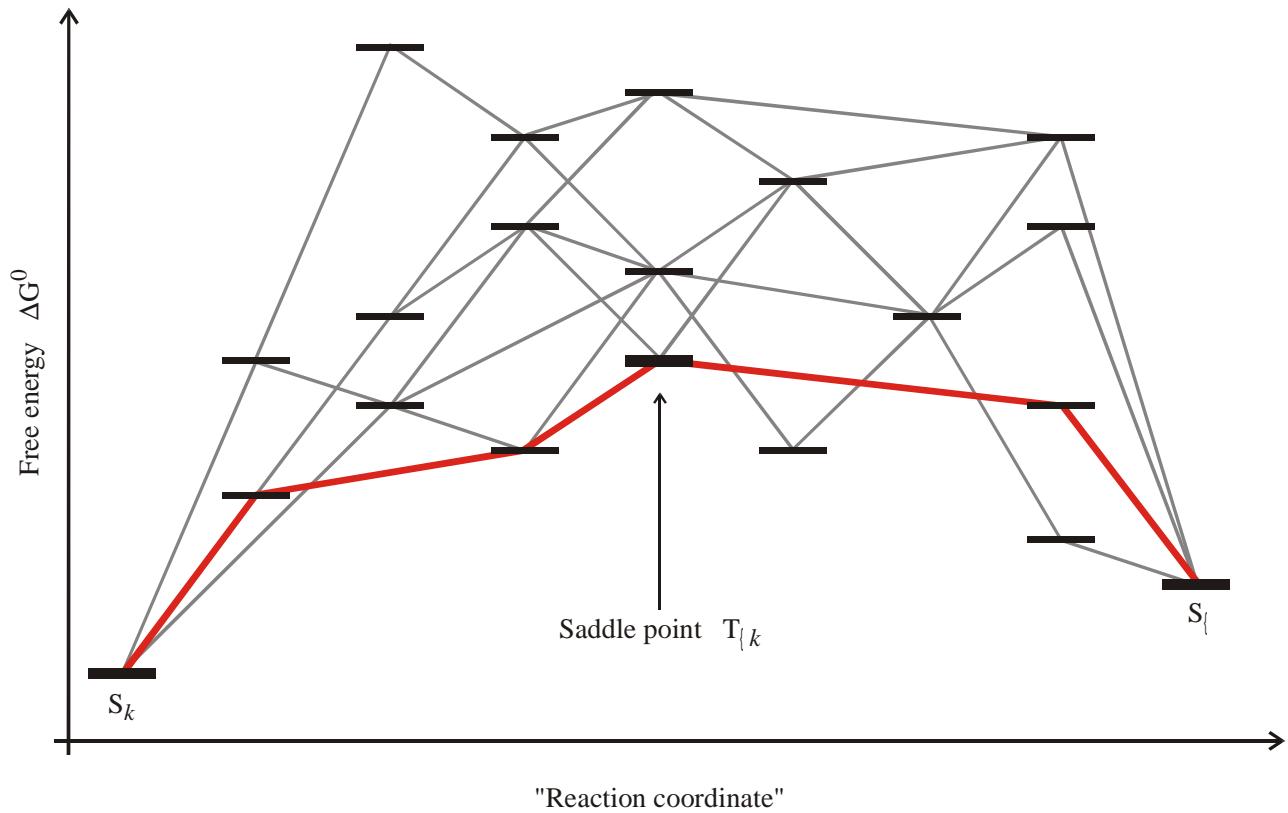


Stable (sub)optimal structures

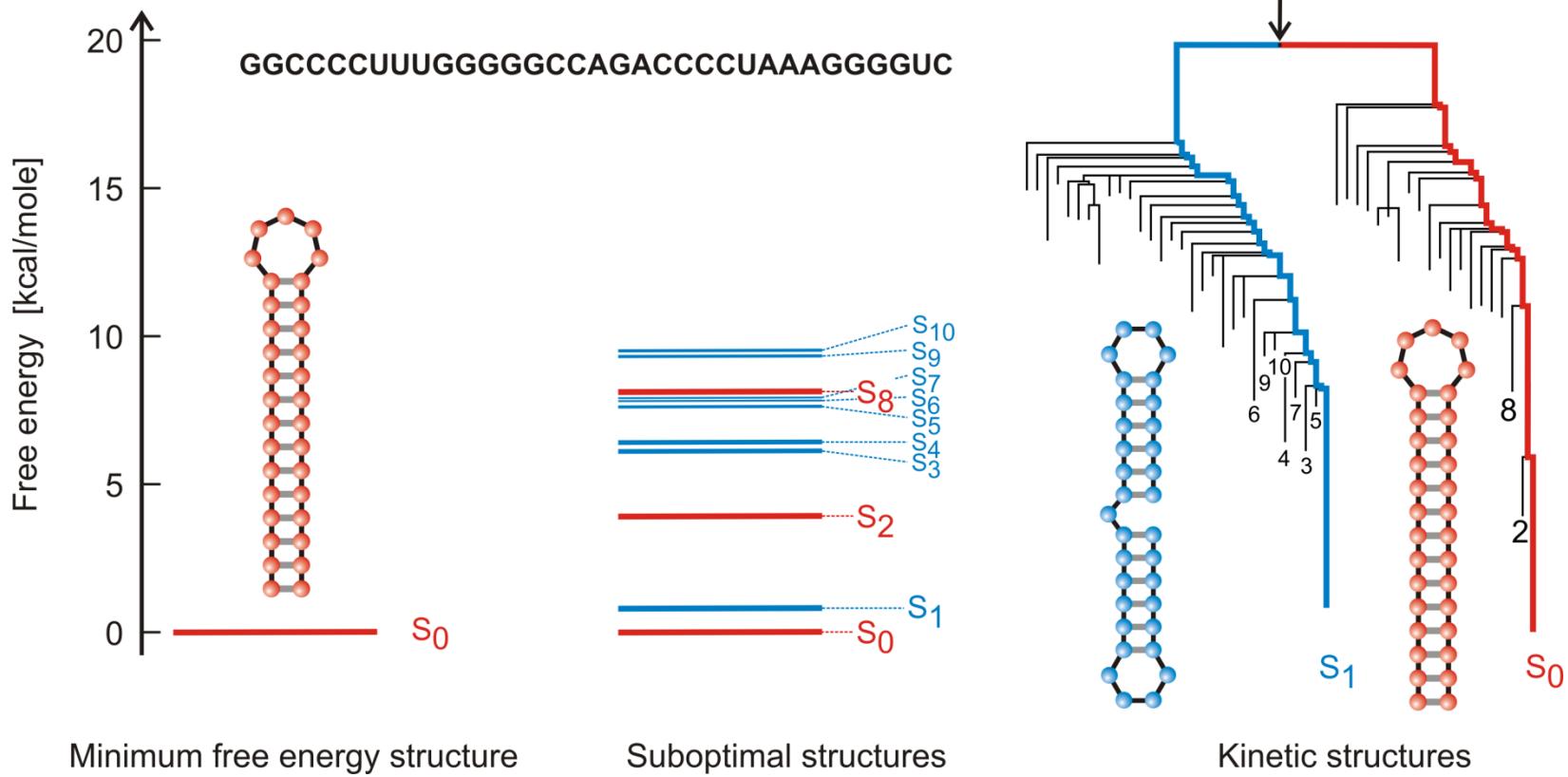
Sequence space



Structure space



Definition of a 'barrier tree'



Interconversion of suboptimal structures

Efficient computation of RNA folding dynamics

Michael T Wolfinger¹, W Andreas Svrcek-Seiler¹, Christoph Flamm¹,
Ivo L Hofacker¹ and Peter F Stadler²

¹ Institut für Theoretische Chemie und Molekulare Strukturbioologie, Universität Wien,
Währingerstraße 17, A-1090 Wien, Austria

² Bioinformatik, Institut für Informatik, Universität Leipzig, D-04103 Leipzig, Germany

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

Received 22 December 2003, in final form 8 March 2004

Published 14 April 2004

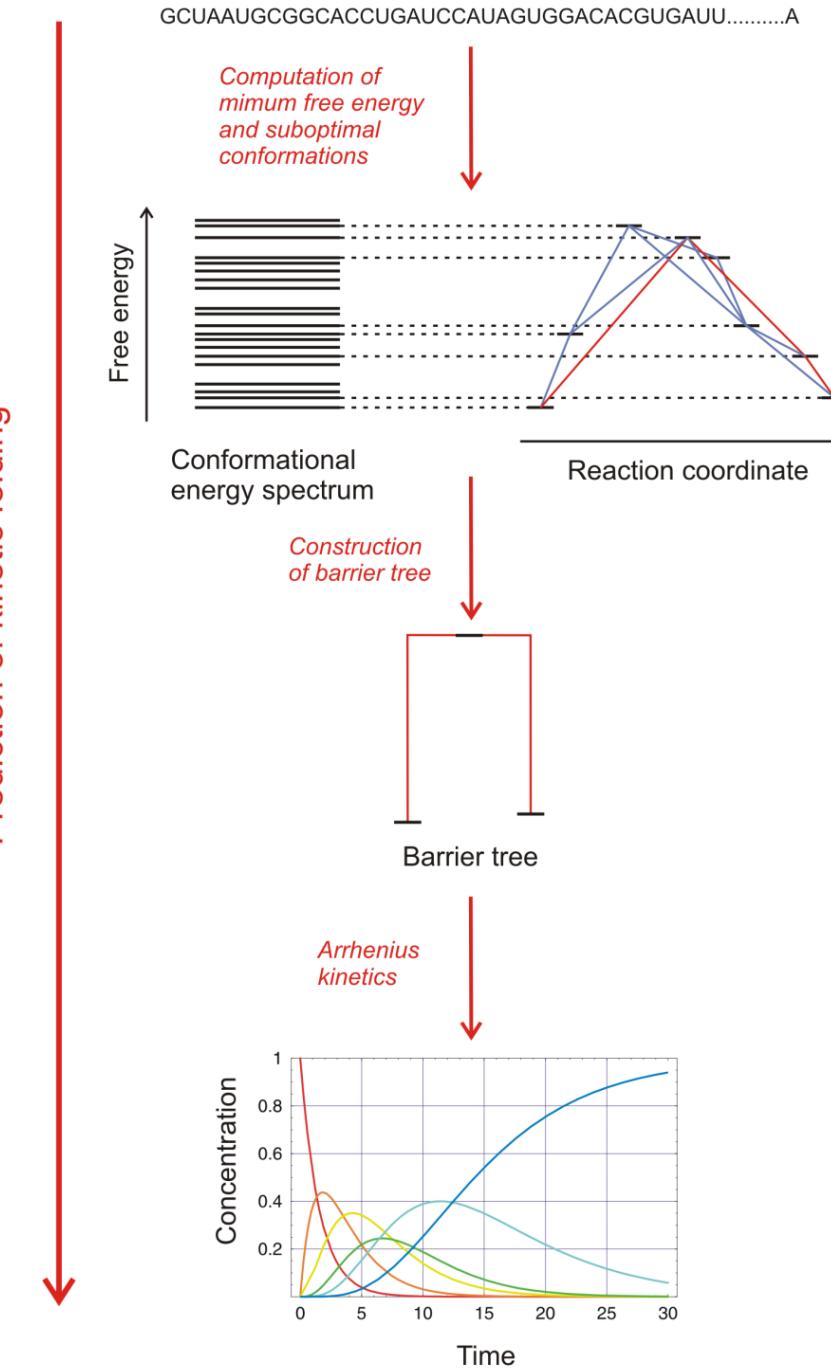
Online at stacks.iop.org/JPhysA/37/4731 (DOI: 10.1088/0305-4470/37/17/005)

Abstract

Barrier trees consisting of local minima and their connecting saddle points imply a natural coarse-graining for the description of the energy landscape of RNA secondary structures. Here we show that, based on this approach, it is possible to predict the folding behaviour of RNA molecules by numerical integration. Comparison with stochastic folding simulations shows reasonable agreement of the resulting folding dynamics and a drastic increase in computational efficiency that makes it possible to investigate the folding dynamics of RNA of at least tRNA size. Our approach is readily applicable to bistable RNA molecules and promises to facilitate studies on the dynamic behaviour of RNA switches.

PACS numbers: 87.14.Gg, 87.15.He, 87.15.Aa, 87.15.Cc

Computation of kinetic folding



Structural parameters affecting the kinetics of RNA hairpin formation

J. H. A. Nagel, C. Flamm¹, I. L. Hofacker¹, K. Franke², M. H. de Smit,
P. Schuster¹ and C. W. A. Pleij*

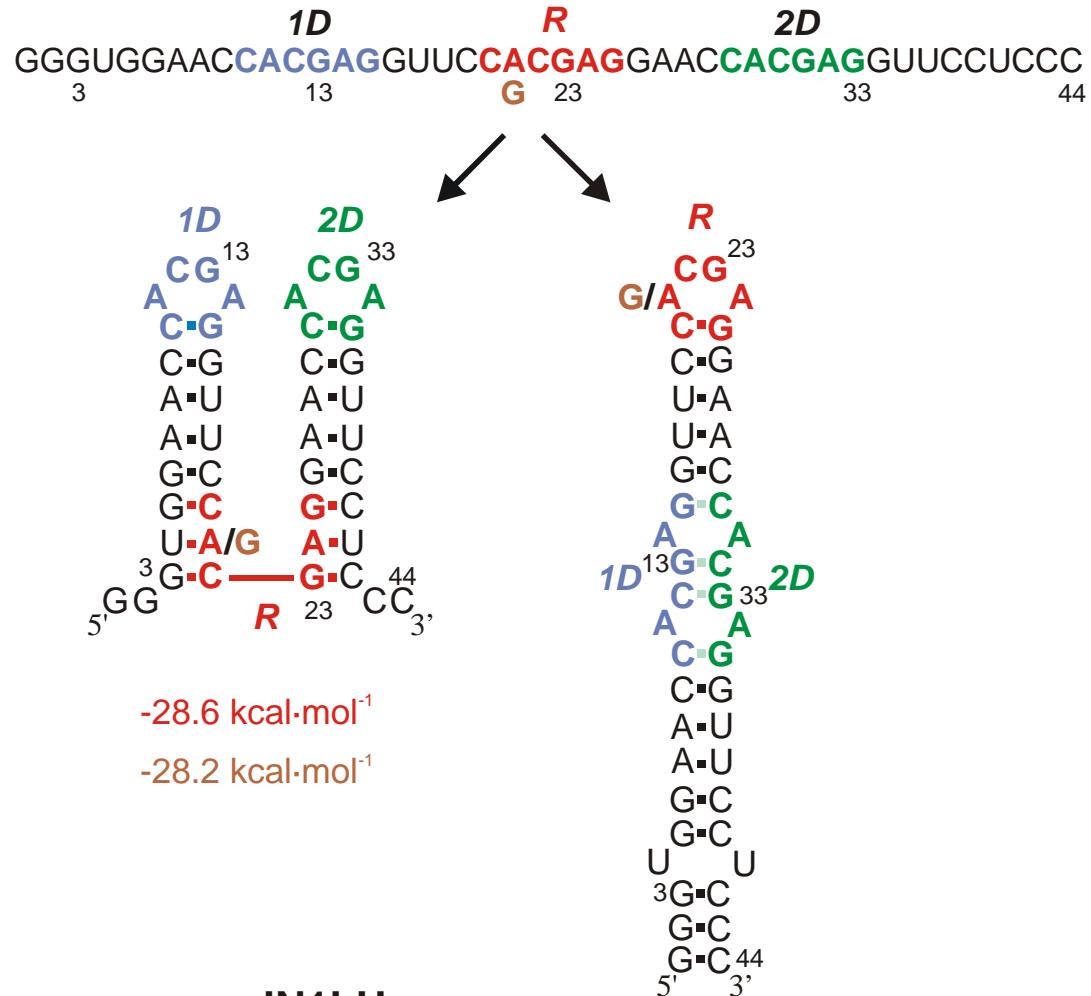
Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2300 RA Leiden, The Netherlands,

¹Institut für Theoretische Chemie und Molekulare Strukturbioologie, Universität Wien, A-1090 Vienna, Austria
and ²IBA NAPS GmbH Rudolf-Wissell-Strasse 28 D-37079 Göttingen, Germany

Received January 28, 2005; Revised and Accepted June 7, 2006

ABSTRACT

There is little experimental knowledge on the sequence dependent rate of hairpin formation in RNA. We have therefore designed RNA sequences that can fold into either of two mutually exclusive hairpins and have determined the ratio of folding of the two conformations, using structure probing. This folding ratio reflects their respective folding rates. Changing one of the two loop sequences from a purine- to a pyrimidine-rich loop did increase its folding rate, which corresponds well with similar observations in DNA hairpins. However, neither changing one of the loops from a regular non-GNRA tetra-loop into a stable GNRA tetra-loop, nor increasing the loop size from 4 to 6 nt did affect the folding rate. The folding kinetics of these RNAs have also been simulated with the program ‘Kinfold’. These simulations were in agreement with the experimental results if the additional stabilization energies for stable tetra-loops were not taken into account. Despite the high stability of the stable tetra-loops, they apparently do not affect folding kinetics of these RNA hairpins. These results show that it is possible to experimentally determine relative folding rates of hairpins and to use these data to improve the computer-assisted simulation of the folding kinetics of stem-loop structures.



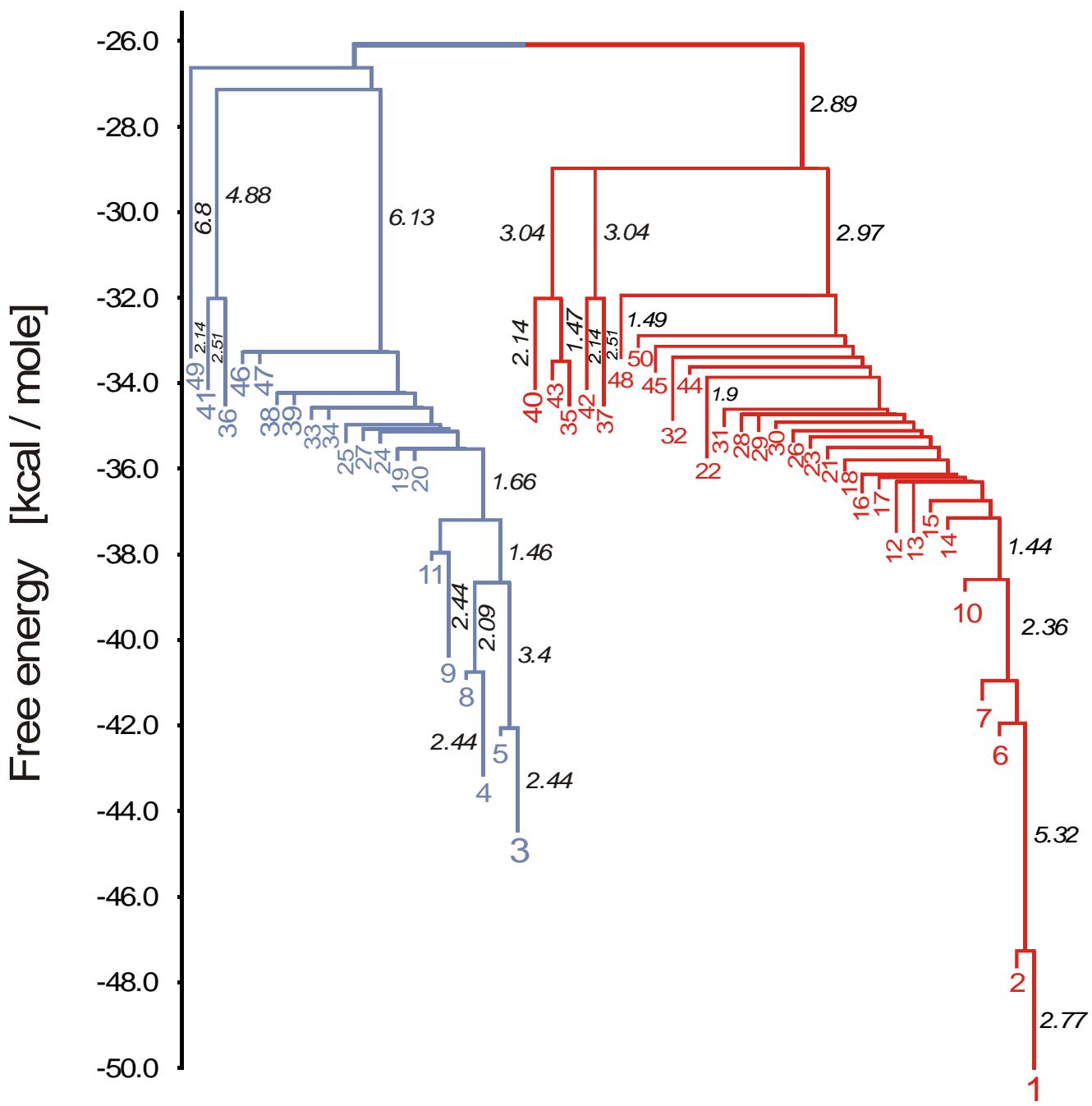
An experimental RNA switch

JN1LH

J.H.A. Nagel, C. Flamm, I.L. Hofacker, K. Franke,
 M.H. de Smit, P. Schuster, and C.W.A. Pleij.

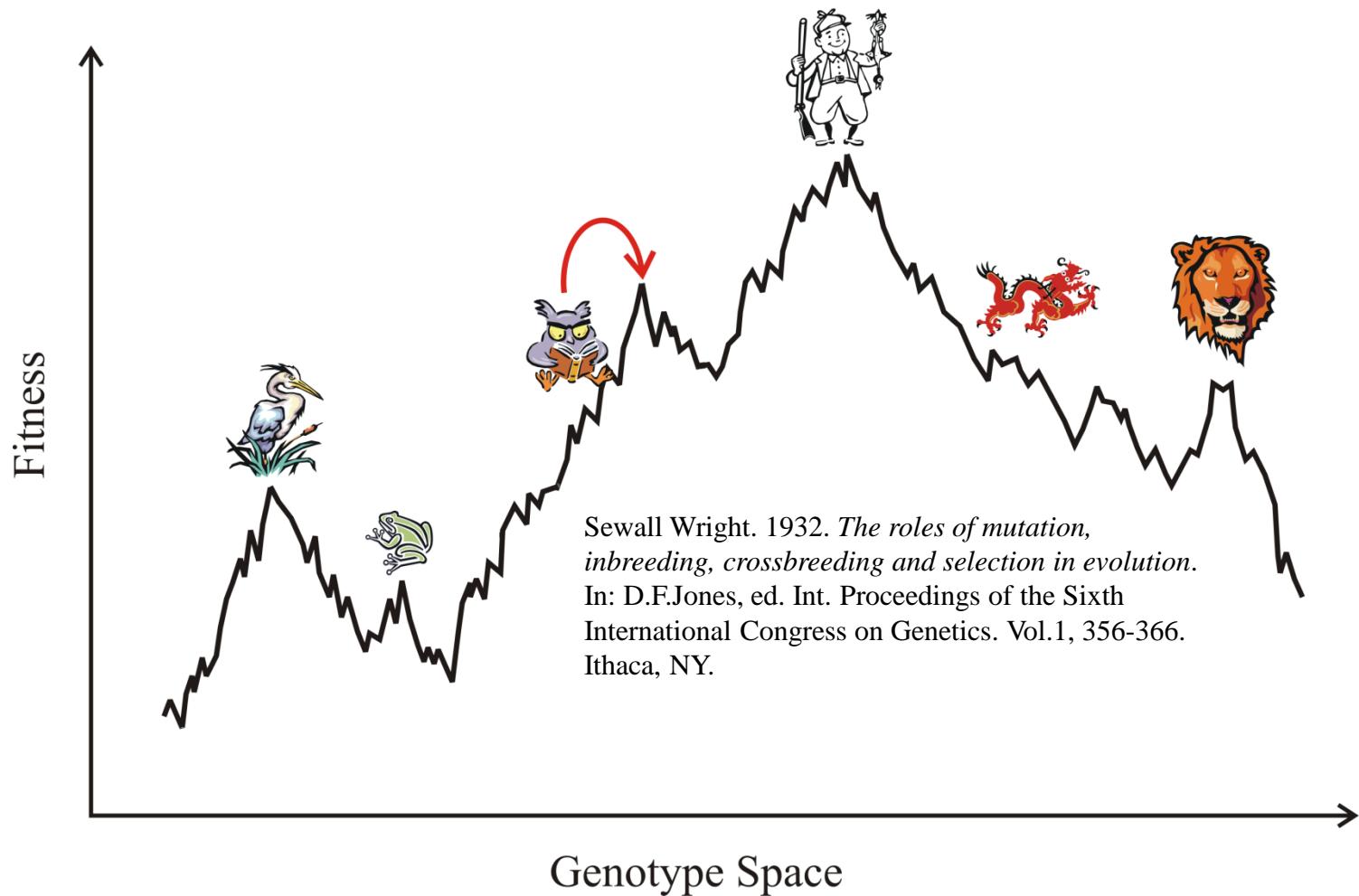
Structural parameters affecting the kinetic competition of RNA hairpin formation. *Nucleic Acids Res.* 34:3568-3576 (2006)

J1LH barrier tree



Four selected examples

1. Parameter determination in chemical kinetics
2. Design of ribonucleic acid (RNA) structures
3. Kinetic folding of RNA molecules
- 4. Modeling evolution**



Sewall Wright's fitness landscape as metaphor for Darwinian evolution



Sewall Wright, 1889 - 1988

$+$ wild type

a alternative allele
on locus A

:

:

:

$abcde$... alternative alleles
on all five loci

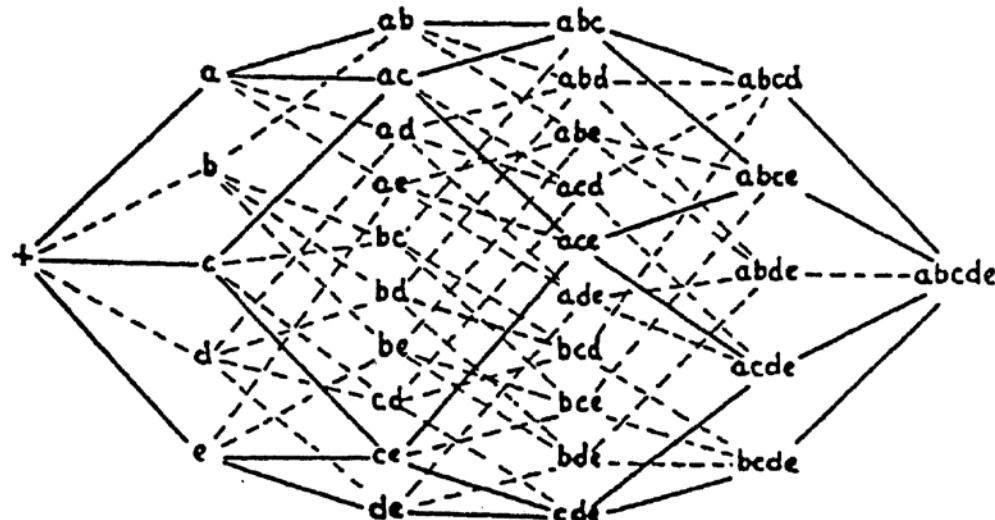
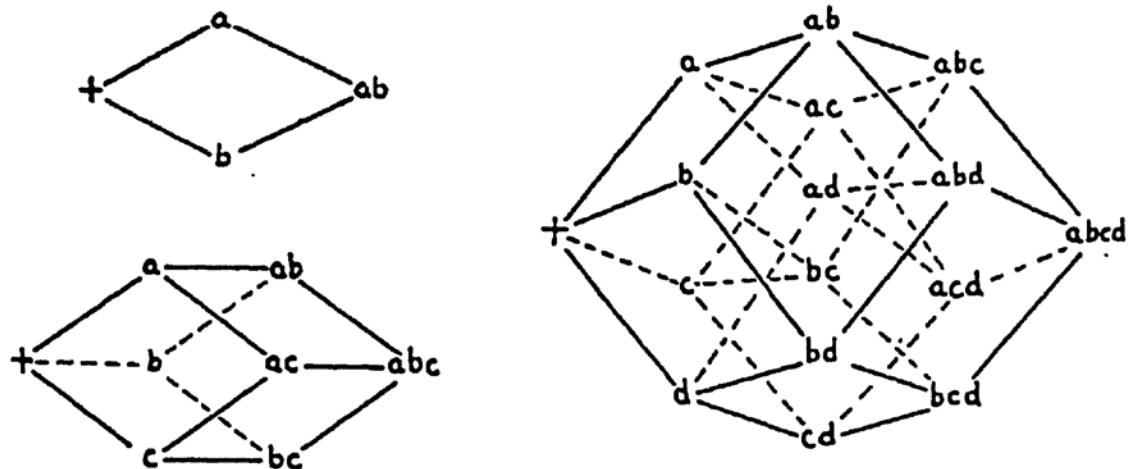


FIG. 1.—The combinations of from 2 to 5 paired allelomorphs.

The multiplicity of gene replacements with two alleles on each locus

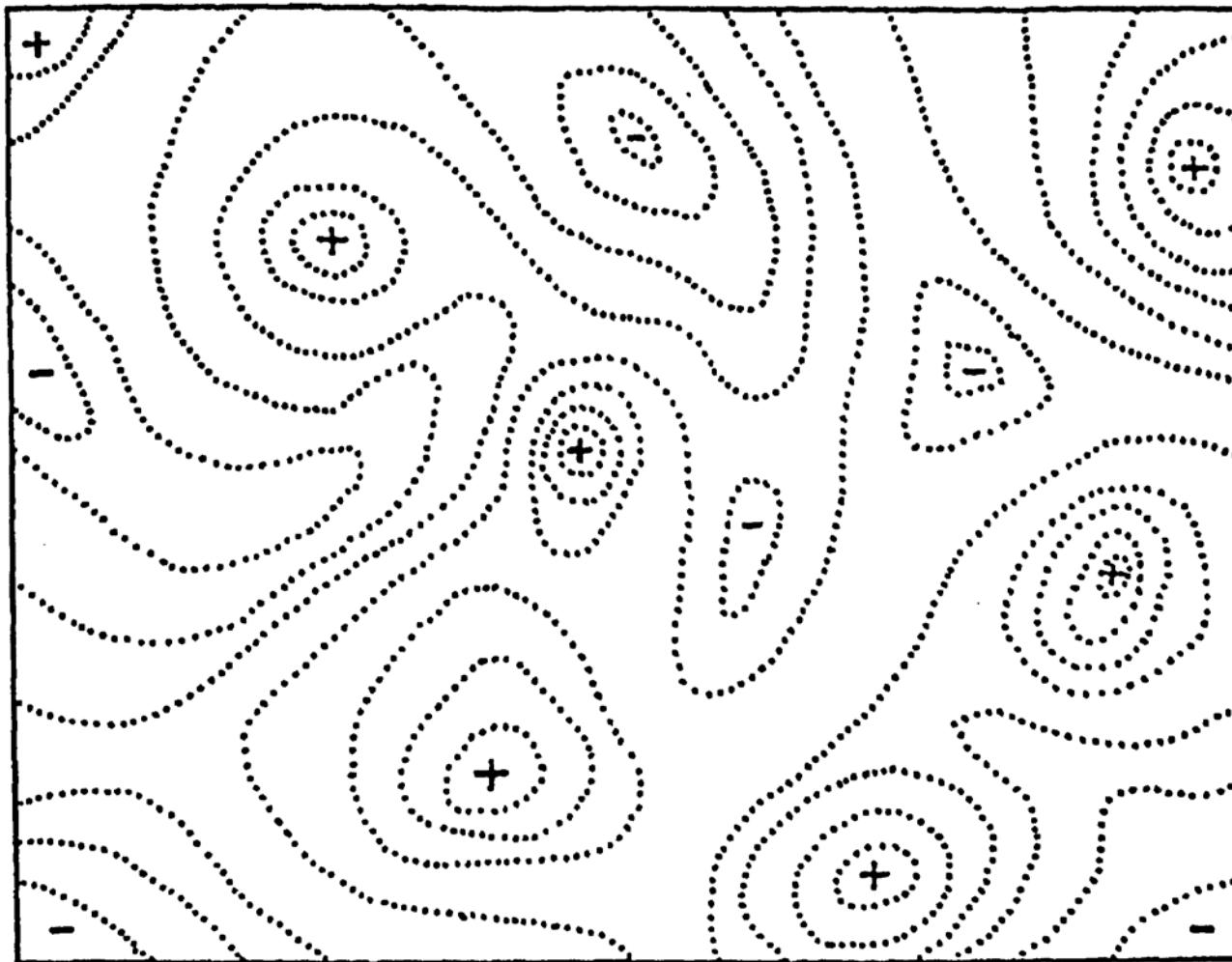
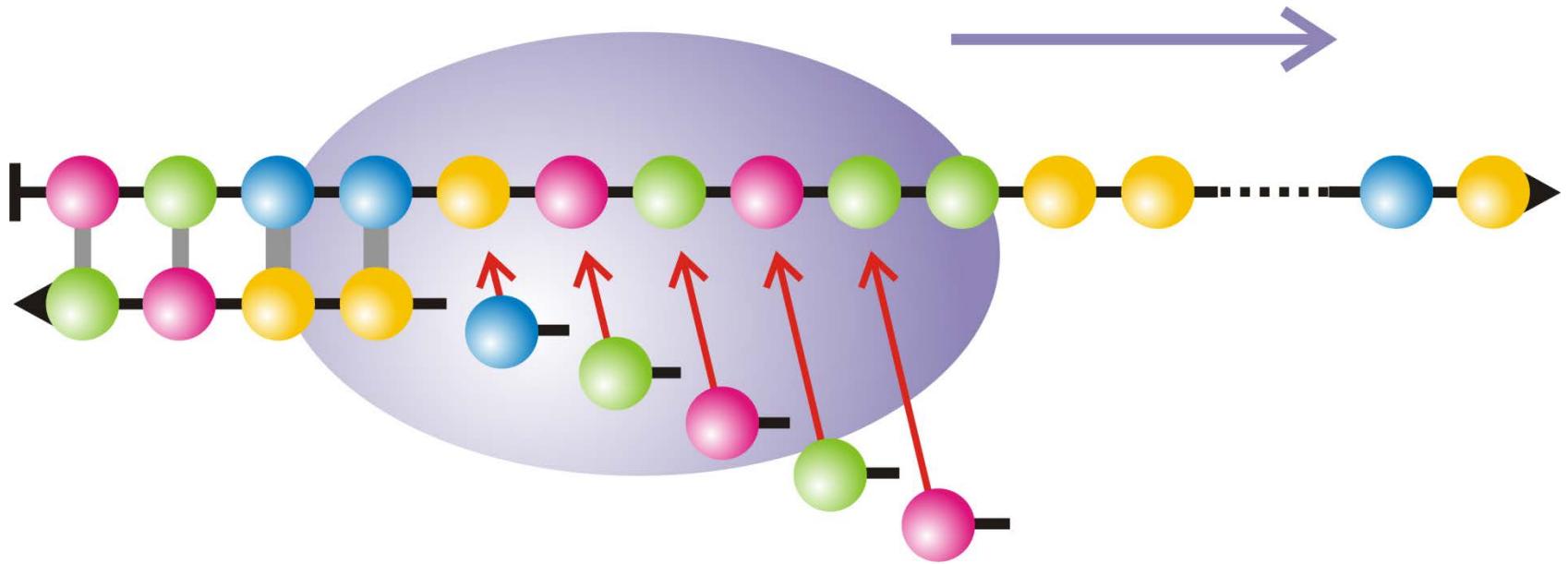


FIG. 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

Evolution is hill climbing of populations or subpopulations



Adenine



Thymine



Guanine



Cytosine



$$\text{Accuracy of replication: } Q = q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot \dots$$

The logics of DNA (or RNA) replication

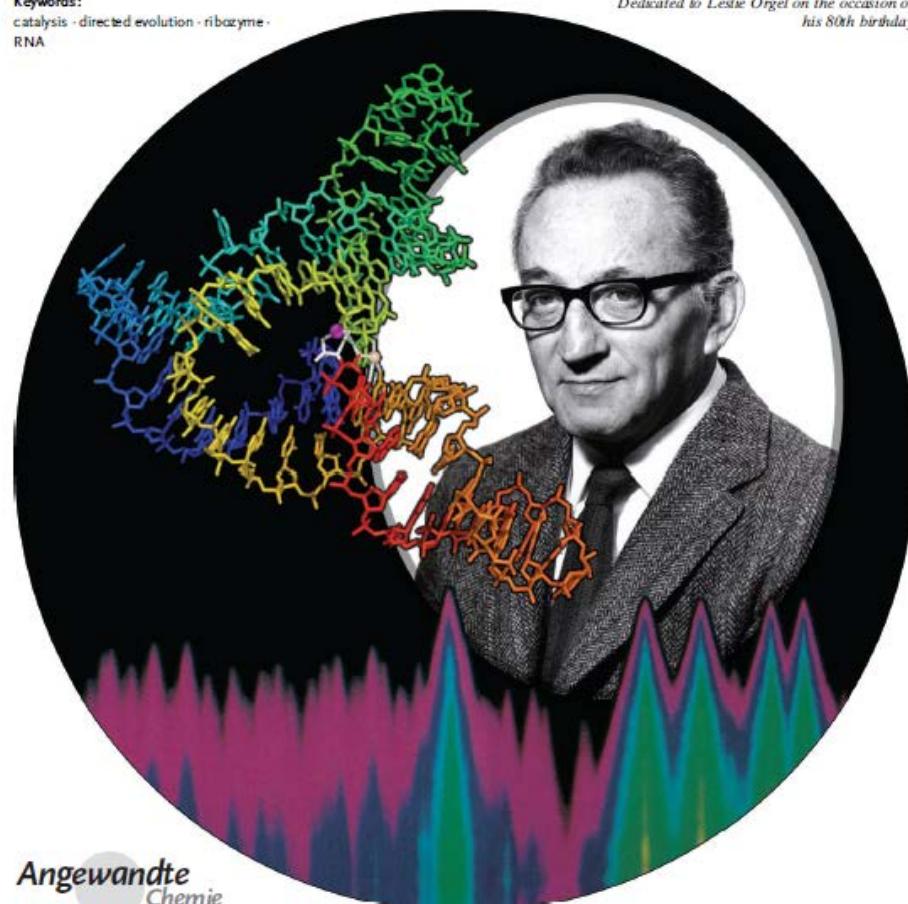
Forty Years of In Vitro Evolution**

Gerald F. Joyce*

Keywords:
catalysis · directed evolution · ribozyme · RNA

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

Sol Spiegelman,
1914 - 1983



Angewandte
Chemie

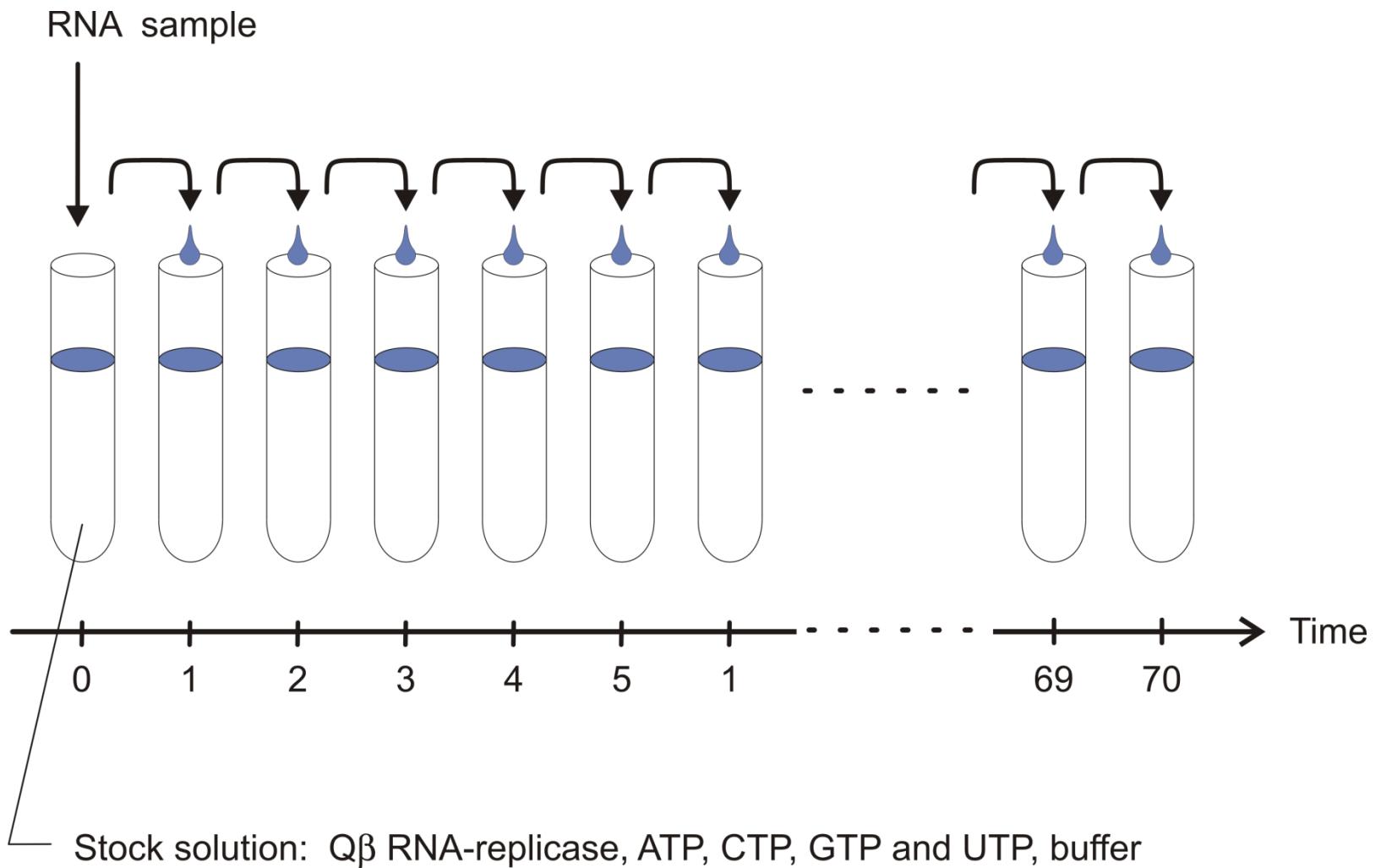
6420 [www.angewandte.org](http://www angewandte org)

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Angew. Chem. Int. Ed. 2007, 46, 6420–6436

Evolution in the test tube:

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



Reproduction of the original figure of the serial transfer experiment with Q β RNA

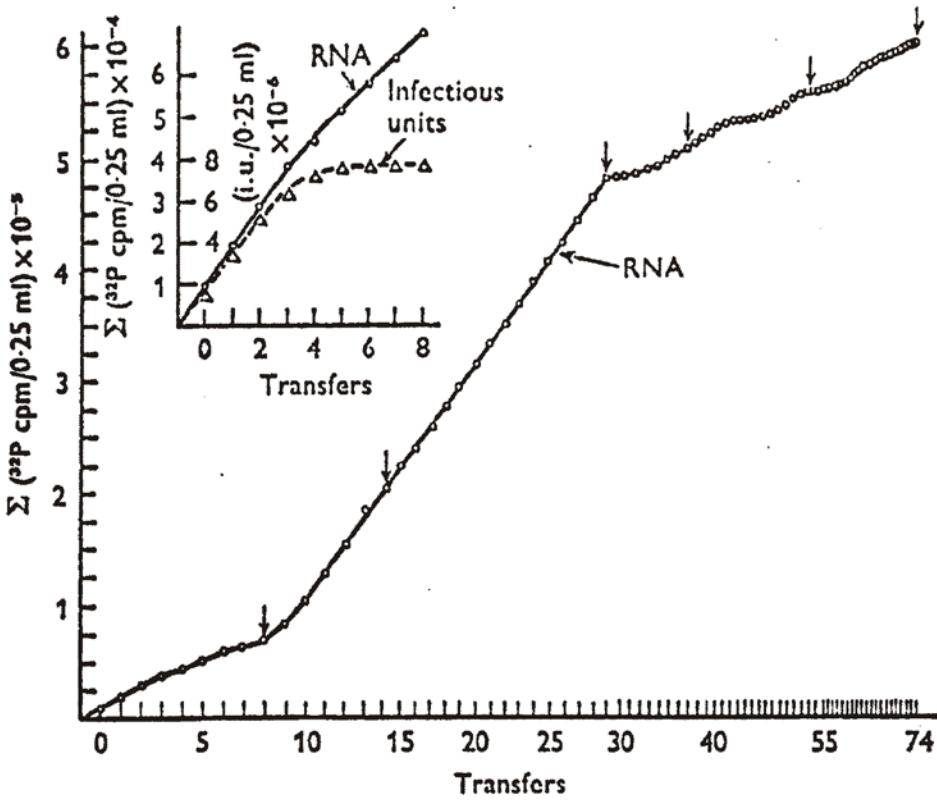
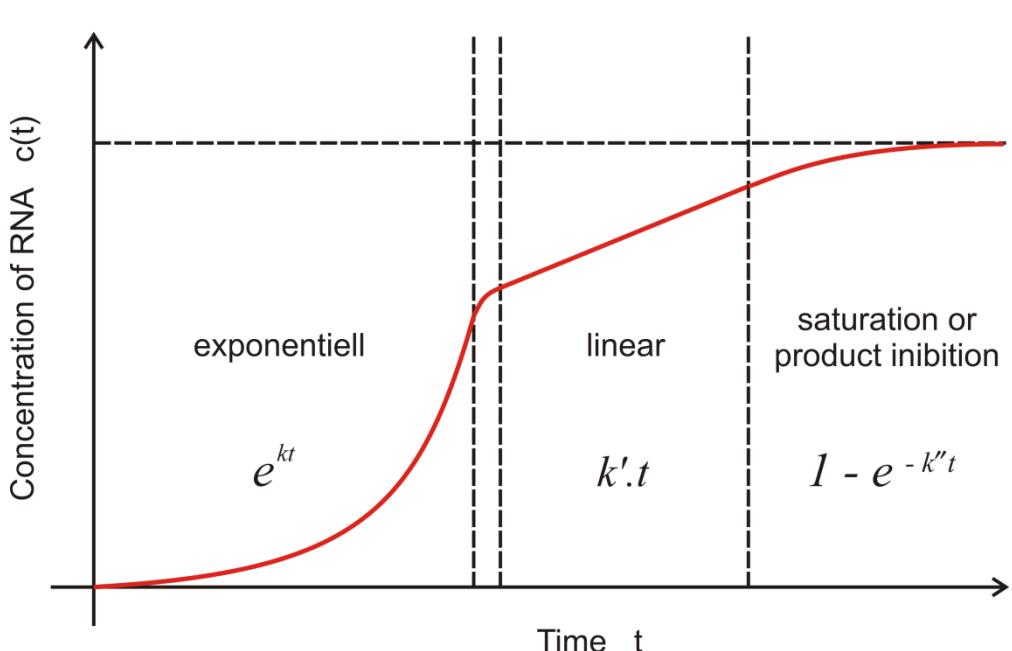
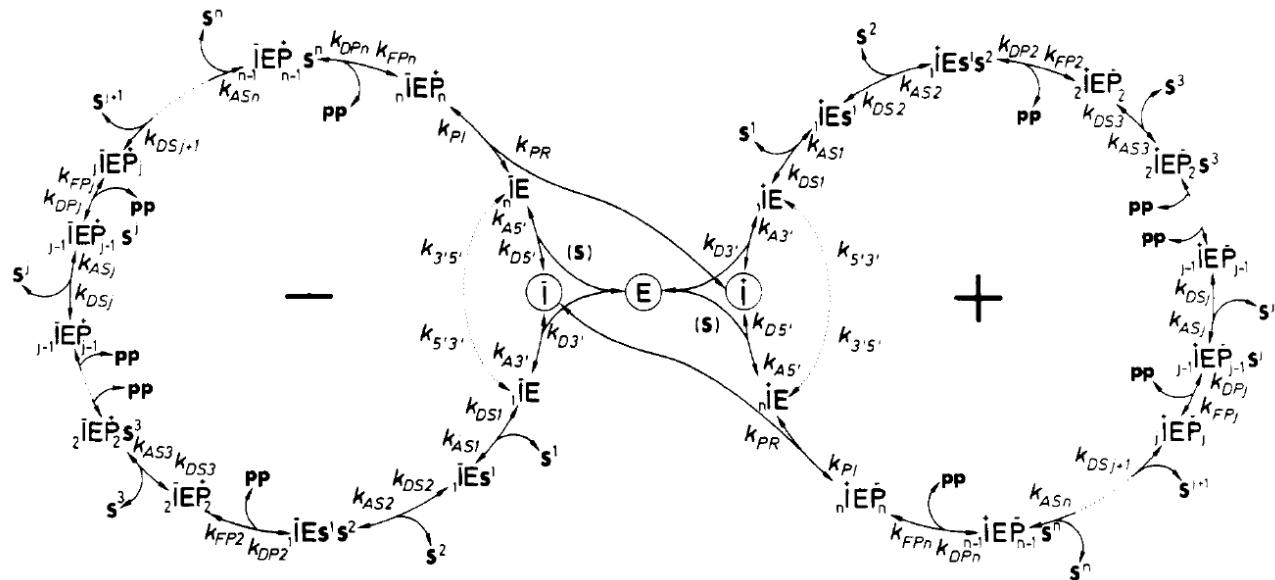


Fig. 9. Serial transfer experiment. Each 0.25 ml standard reaction mixture contained 40 μ g of Q β replicase and 32 P-UTP. The first reaction (0 transfer) was initiated by the addition of 0.2 μ g ts-1 (temperature-sensitive RNA) and incubated at 35 °C for 20 min, whereupon 0.02 ml was drawn for counting and 0.02 ml was used to prime the second reaction (first transfer), and so on. After the first 13 reactions, the incubation periods were reduced to 15 min (transfers 14–29). Transfers 30–38 were incubated for 10 min. Transfers 39–52 were incubated for 7 min, and transfers 53–74 were incubated for 5 min. The arrows above certain transfers (0, 8, 14, 29, 37, 53, and 73) indicate where 0.001–0.1 ml of product was removed and used to prime reactions for sedimentation analysis on sucrose. The inset examines both infectious and total RNA. The results show that biologically competent RNA ceases to appear after the 4th transfer (Mills *et al.* 1967).

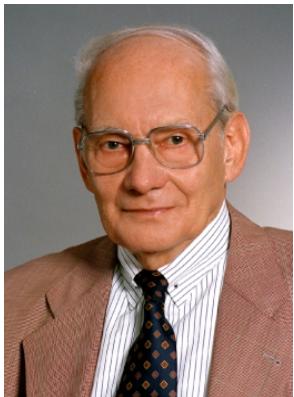
D.R.Mills, R.L.Peterson, S.Spiegelman,
An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc.Natl.Acad.Sci.USA
58 (1967), 217-224



Christof K. Biebricher,
1941-2009



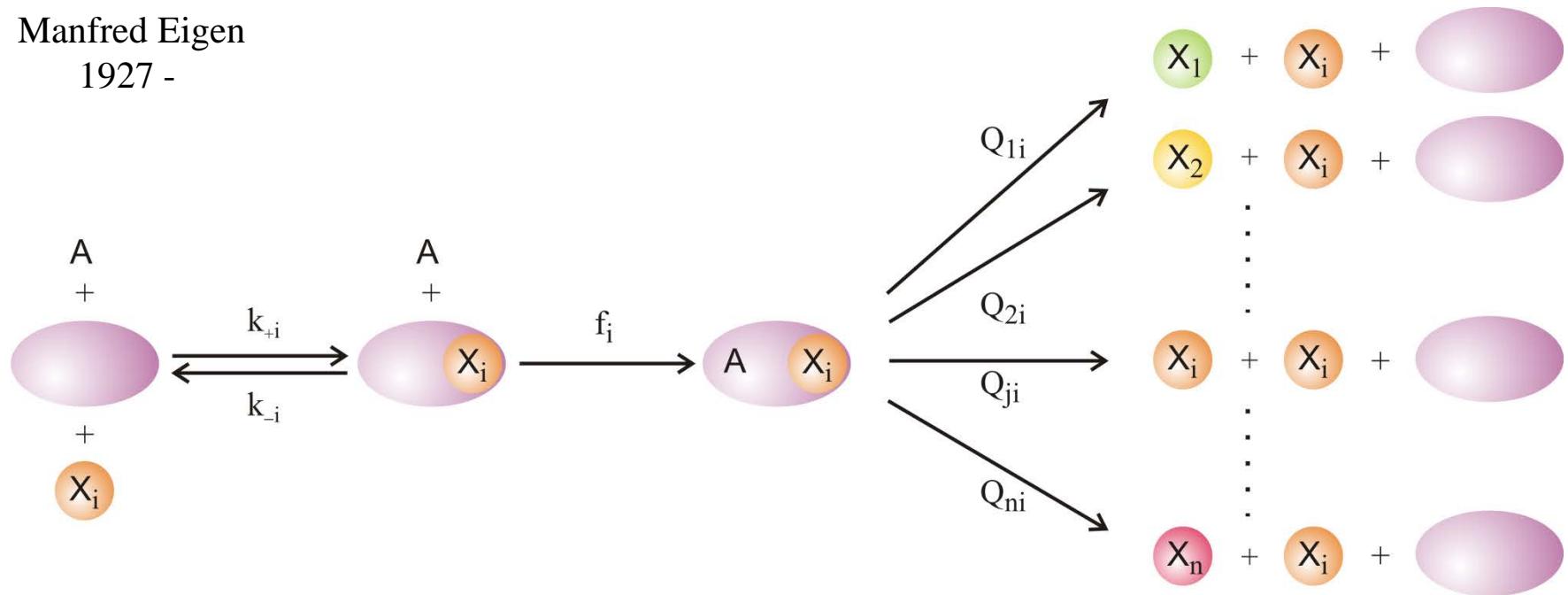
Kinetics of RNA replication
C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
Biochemistry 22:2544-2559, 1983



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

$$W_{ji} = Q_{ji} \cdot f_i, \quad \Phi = \sum_{i=1}^n f_i x_i / \sum_{i=1}^n x_i$$

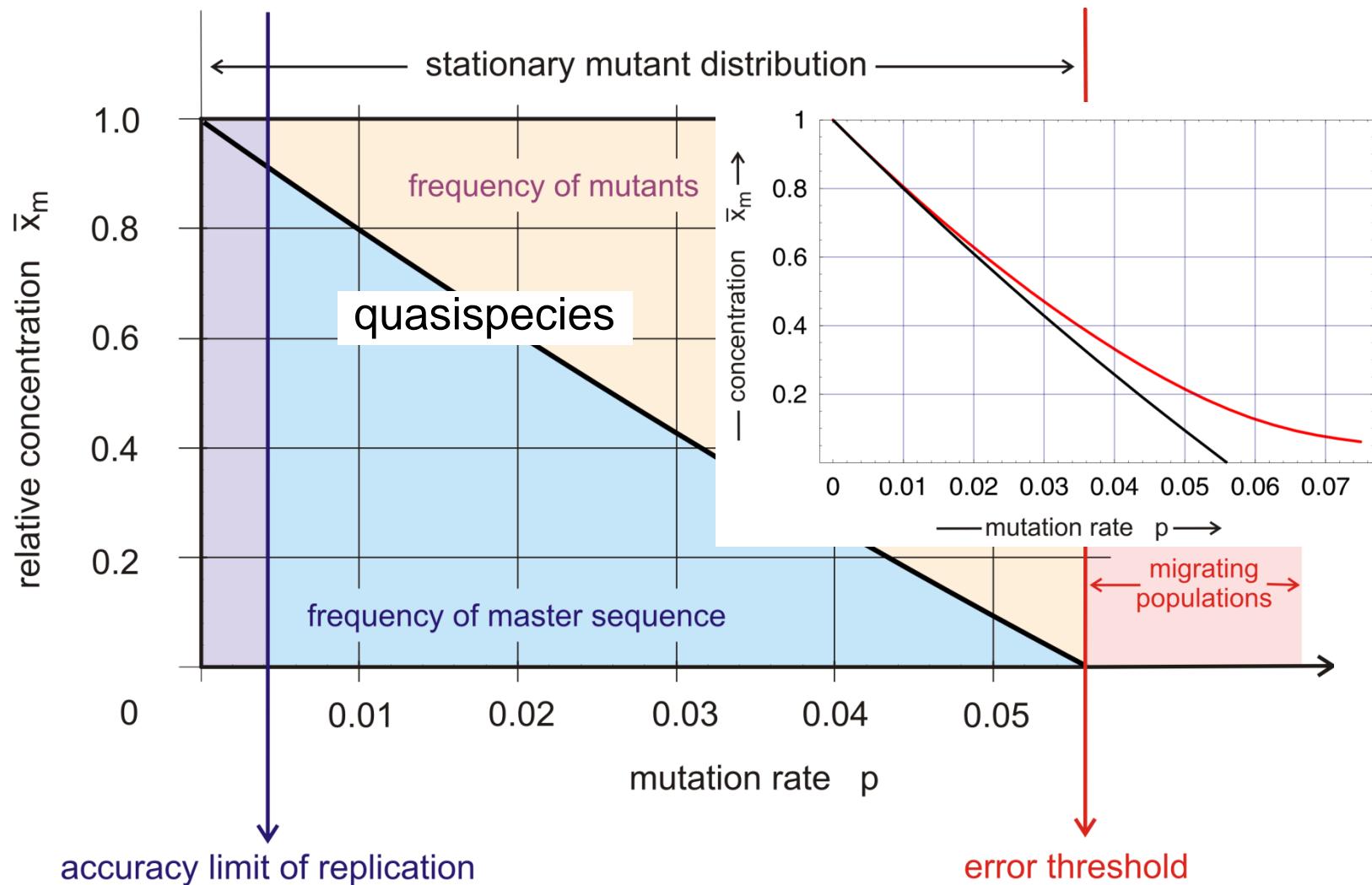
Manfred Eigen
1927 -



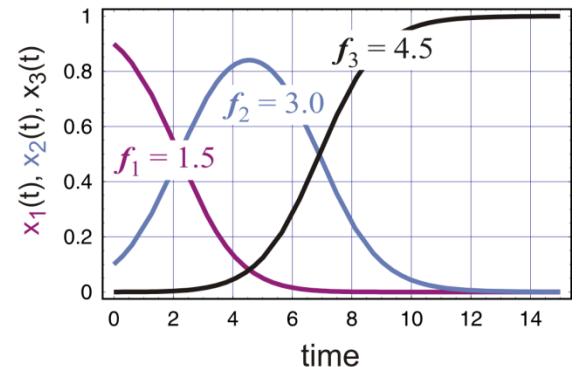
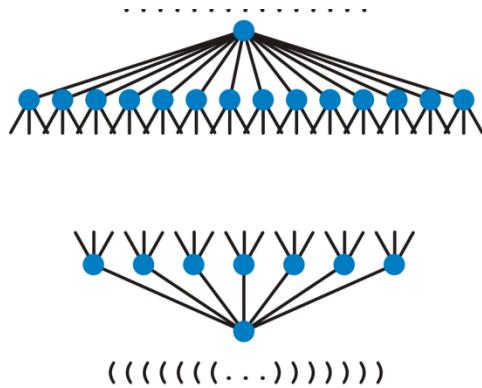
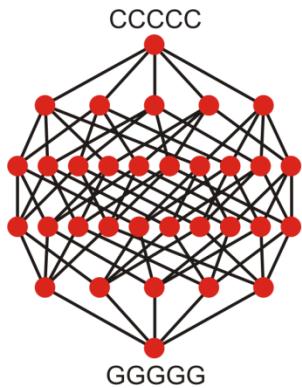
Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,

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



The error threshold in replication and mutation



$$\Phi: (\mathcal{Q}, d_H) \Rightarrow (\mathcal{Y}, d_Y)$$

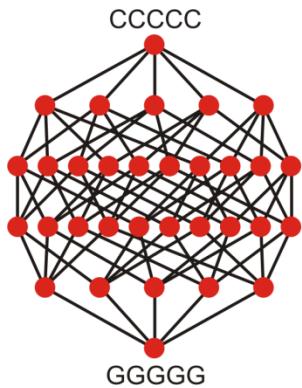
$$\Psi: (\mathcal{Y}, d_Y) \Rightarrow \mathbf{R}^1$$

S
sequence

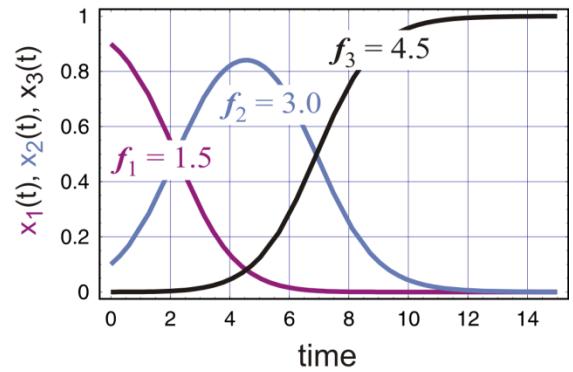
$\mathbf{Y} = \Phi(S)$
structure

$f = \Psi(Y)$
function

The paradigm of structural biology



sequence space



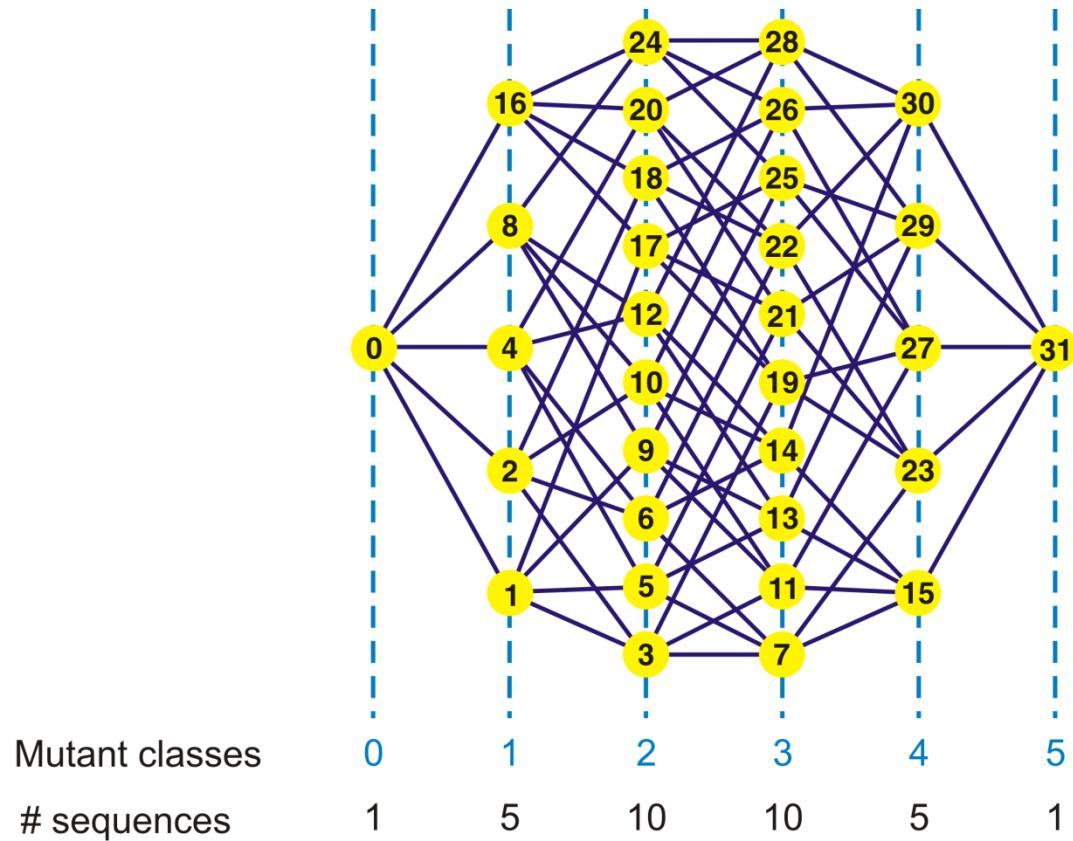
parameter space

S
sequence



$f = \Psi(Y)$
function

The simplified model



Binary sequences are encoded by their decimal equivalents:

C = 0 and **G** = 1, for example,

"0" ≡ 00000 = **CCCCC**,

"14" ≡ 01110 = **CGGGC**,

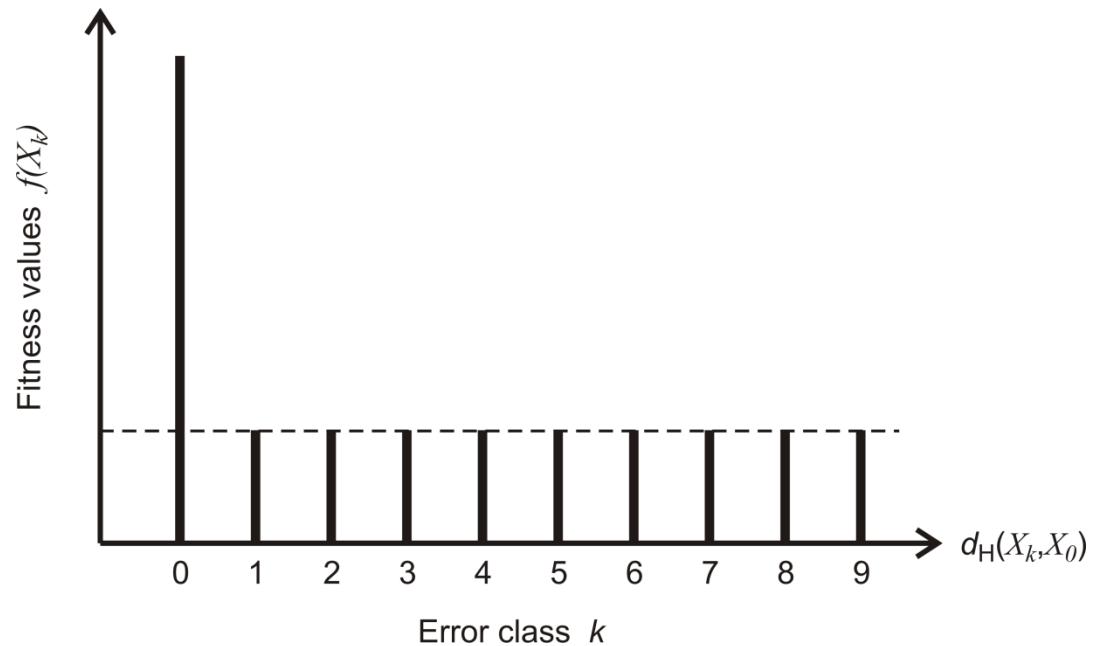
"29" ≡ 11101 = **GGGCG**, etc.

Concentrations of entire error classes: $[\Gamma_k] = y_k(p)$, $k = 0, 1, \dots, n$

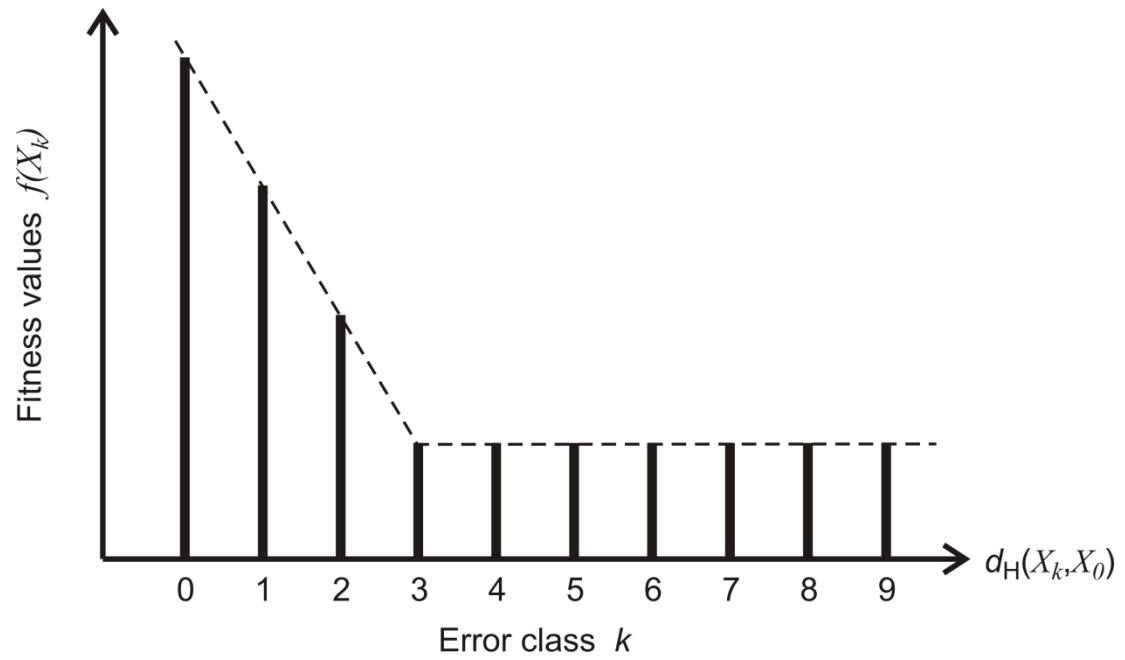
$$y_k(p) = \sum_{i=1, d_H(X_i, X_k)=k}^N x_i(p), \quad |\Gamma_k| = \binom{n}{k}$$

Model fitness landscapes I

single peak landscape



step linear landscape

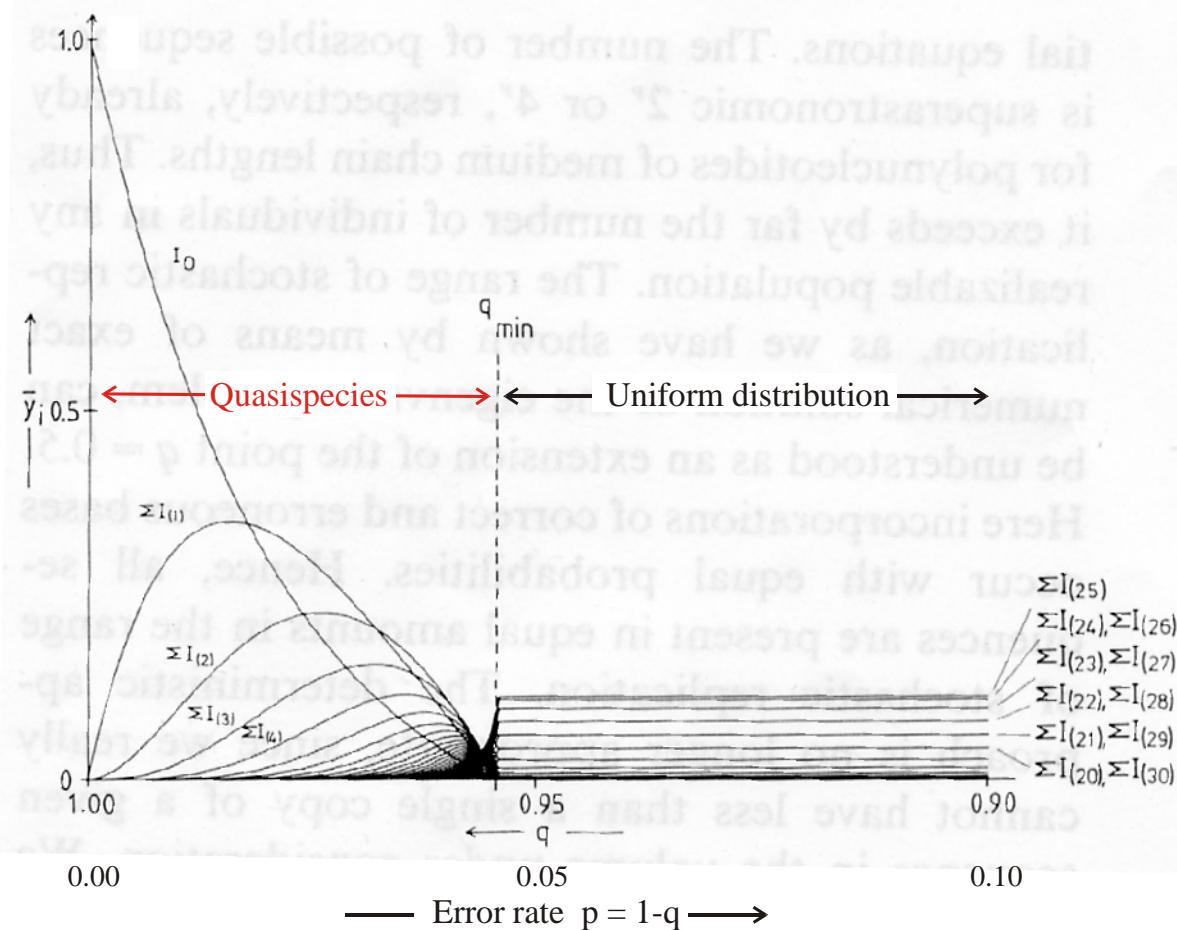


SELF-REPLICATION WITH ERRORS

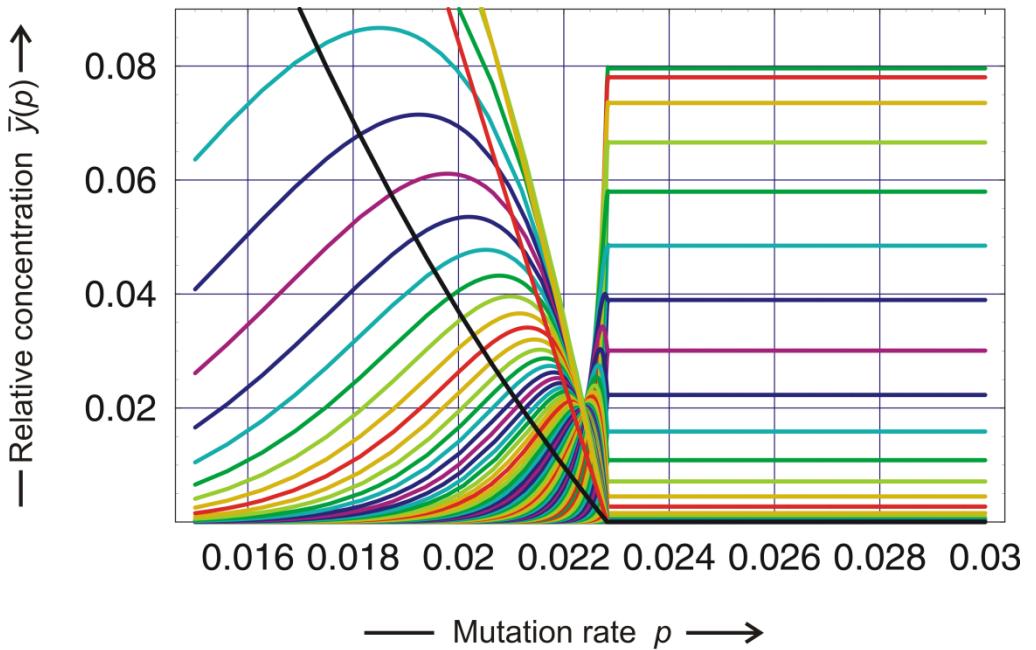
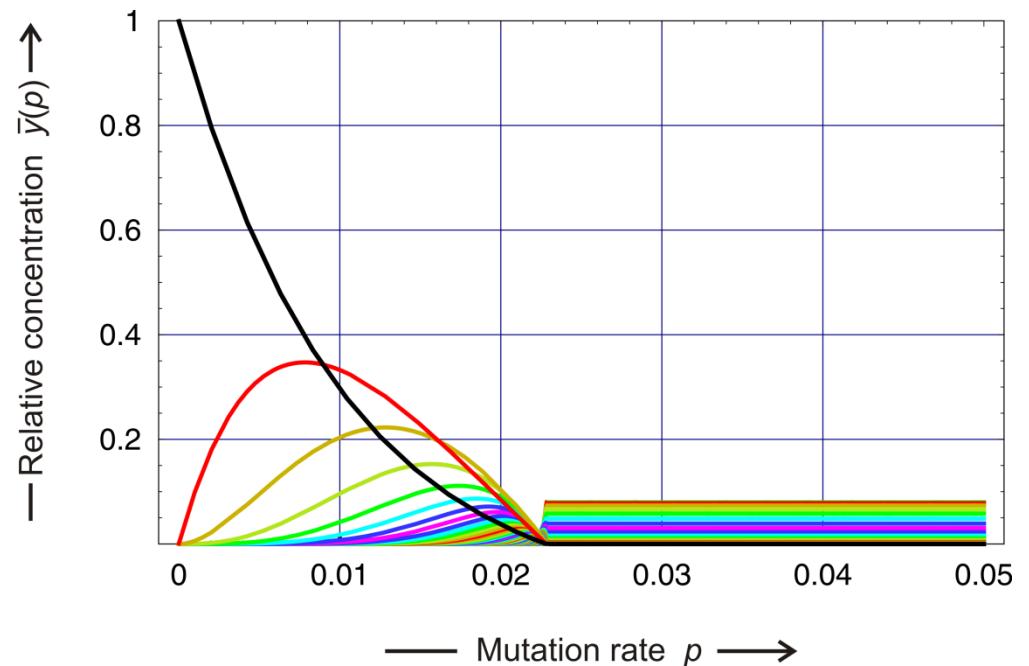
A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

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

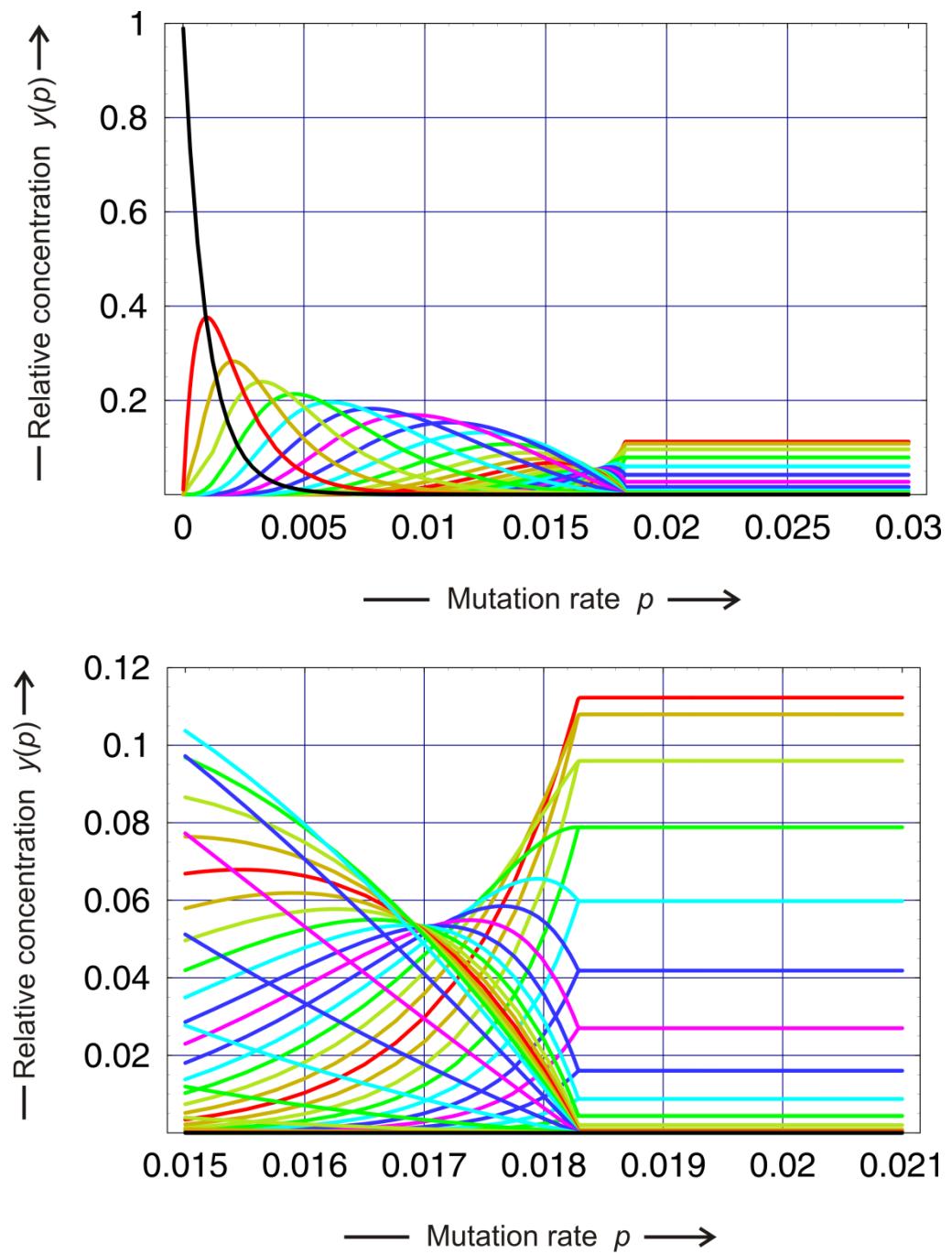


Stationary population or
quasispecies as a function
of the mutation or error
rate p

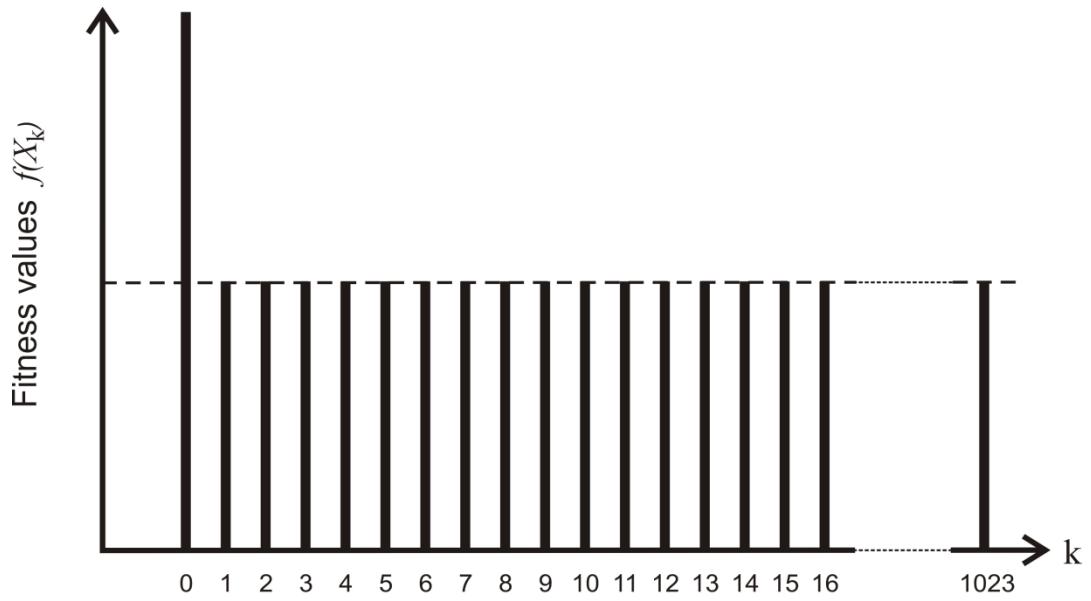


Error threshold on the
single peak landscape

Error threshold on the
step linear landscape



single peak landscape



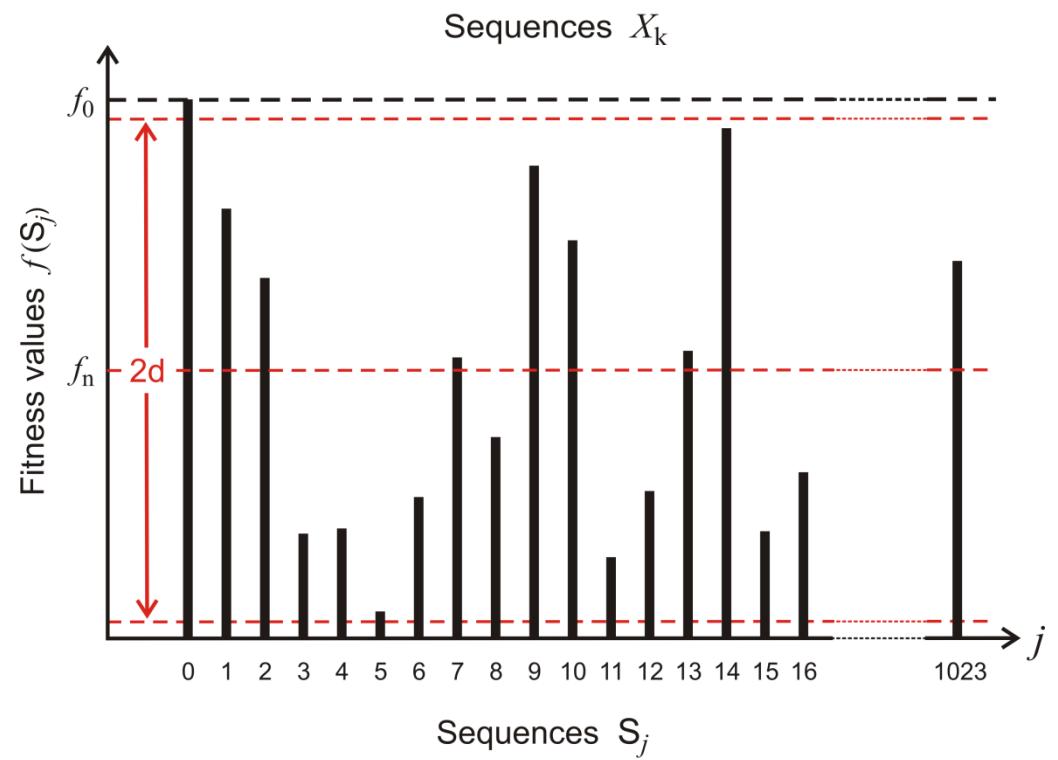
$$f(S_j) = f_n + 2d(f_0 - f_n) \left(\eta_j^{(s)} - 0.5 \right)$$

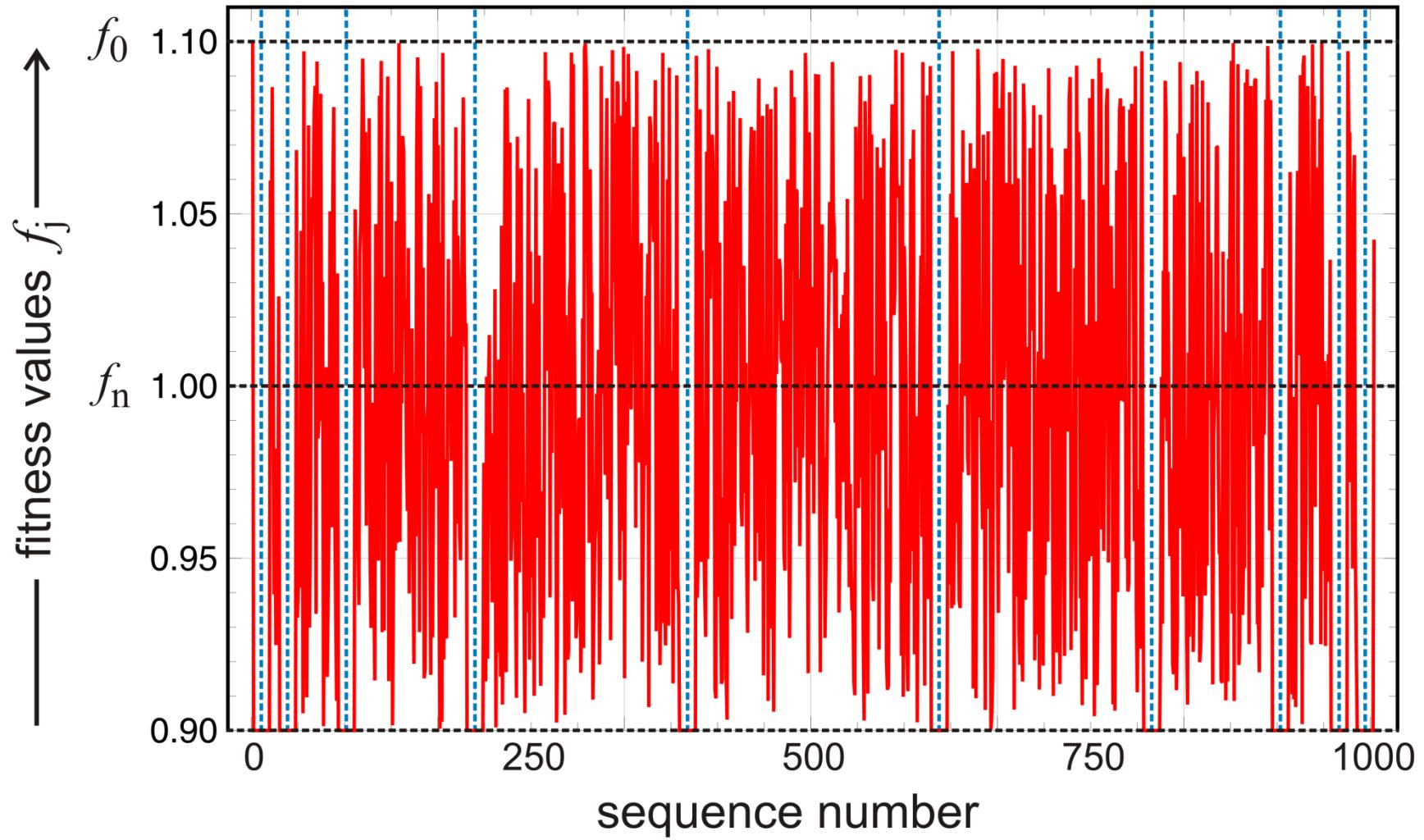
$j = 1, 2, \dots, N; j \neq m,$

$\eta \dots$ random number; $s \dots$ seeds

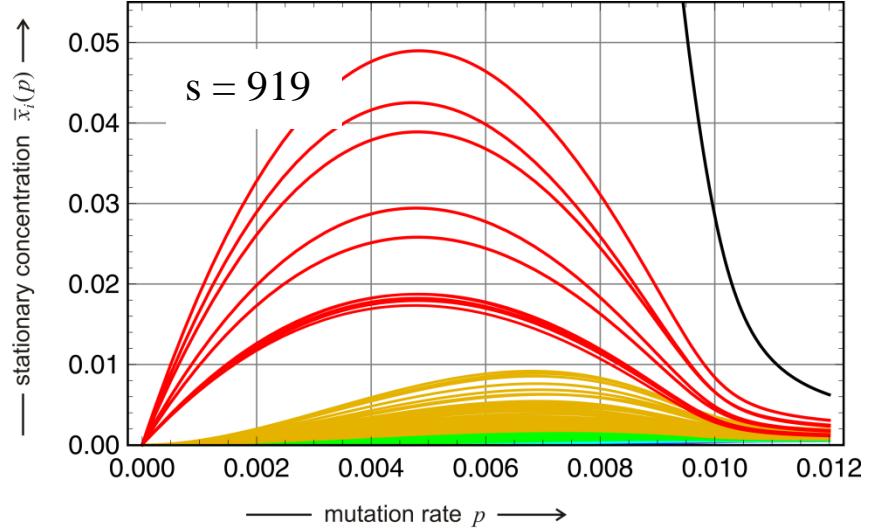
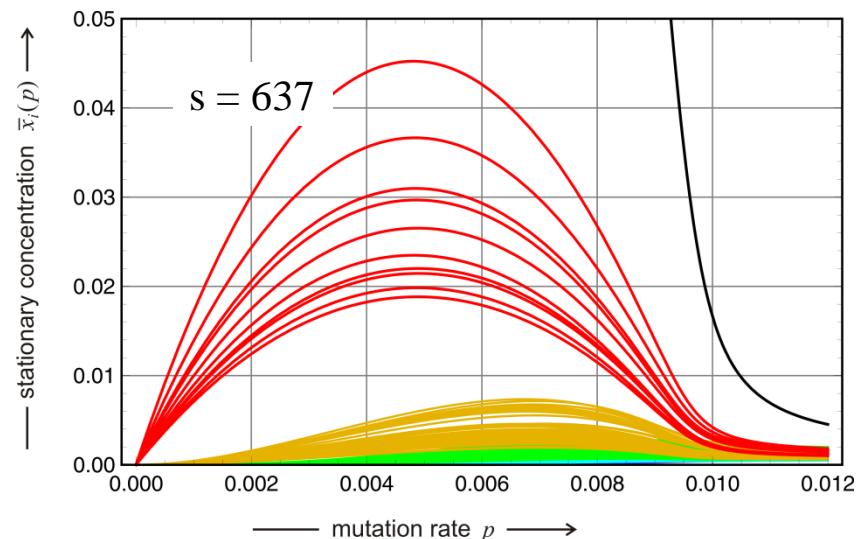
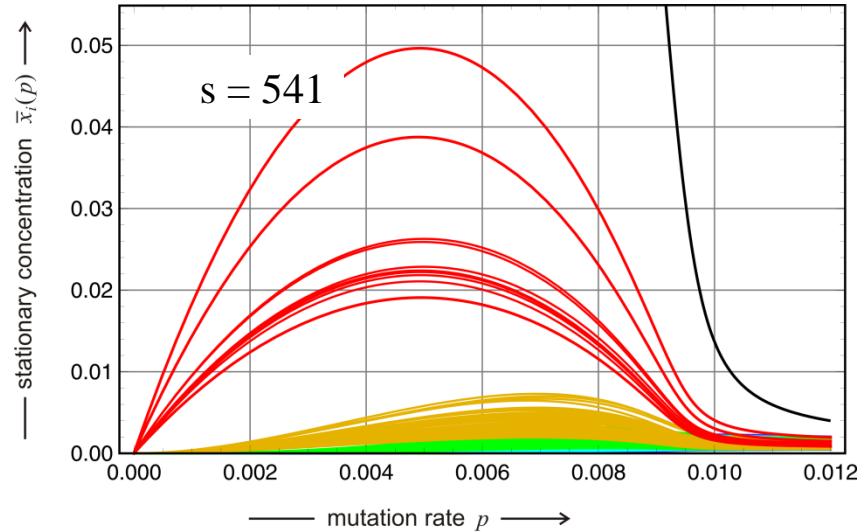
„realistic“ landscape

Rugged fitness landscapes
over individual binary sequences
with $n = 10$



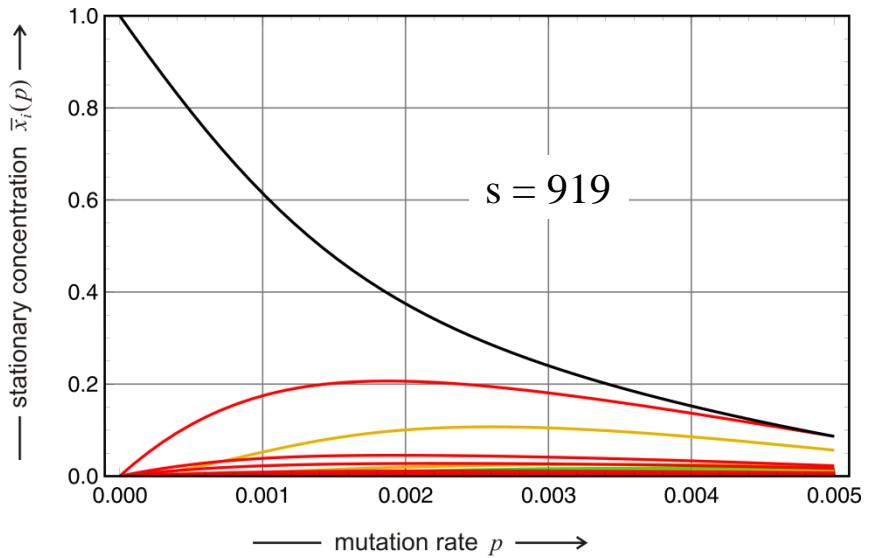
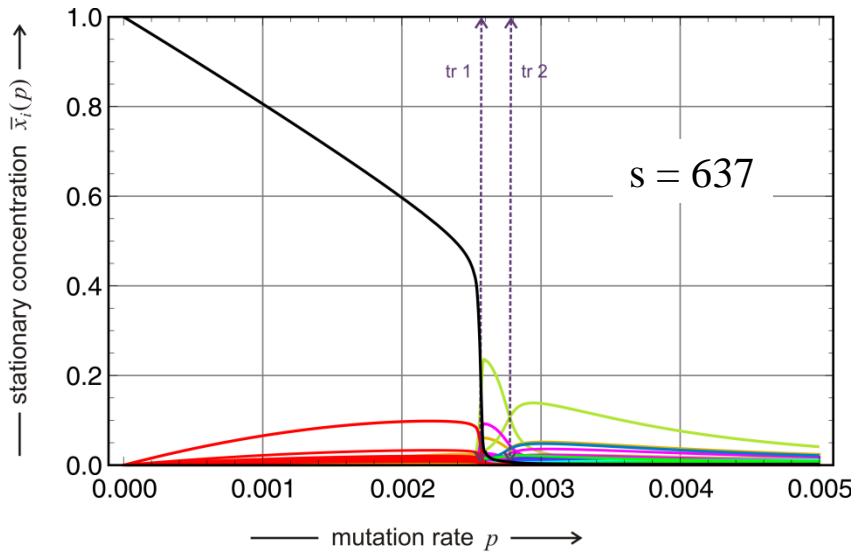
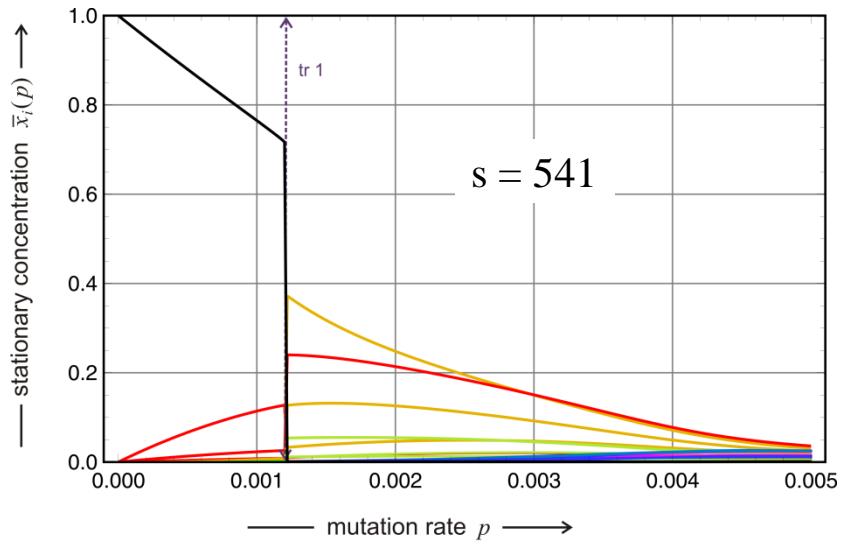


Random distribution of fitness values: $d = 1.0$ and $s = 637$



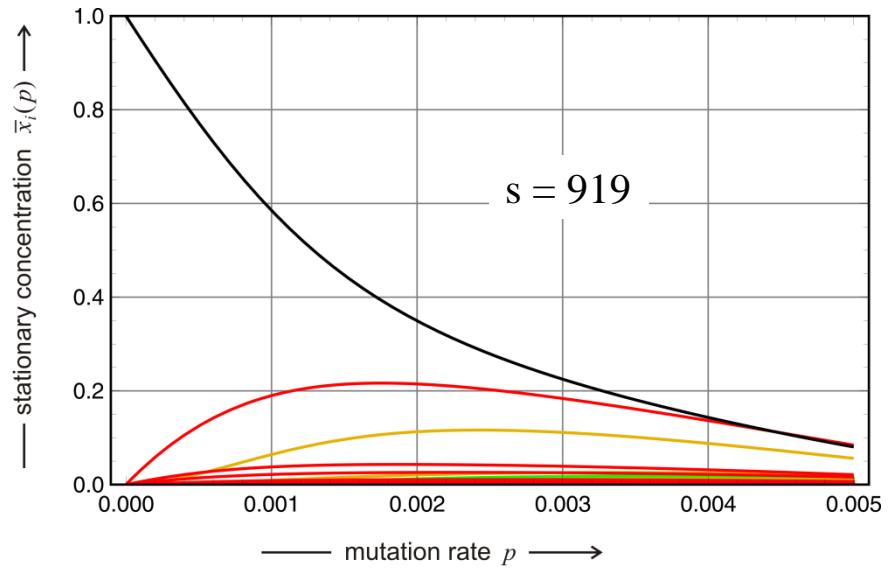
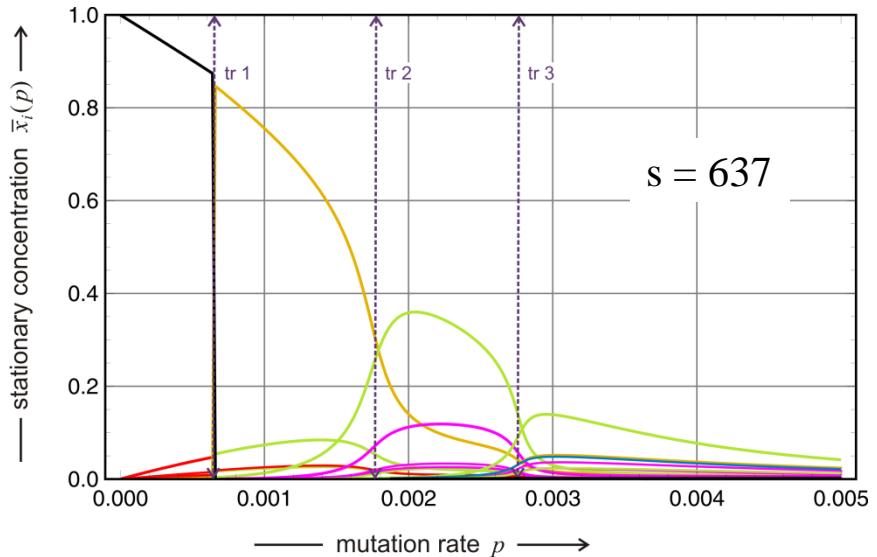
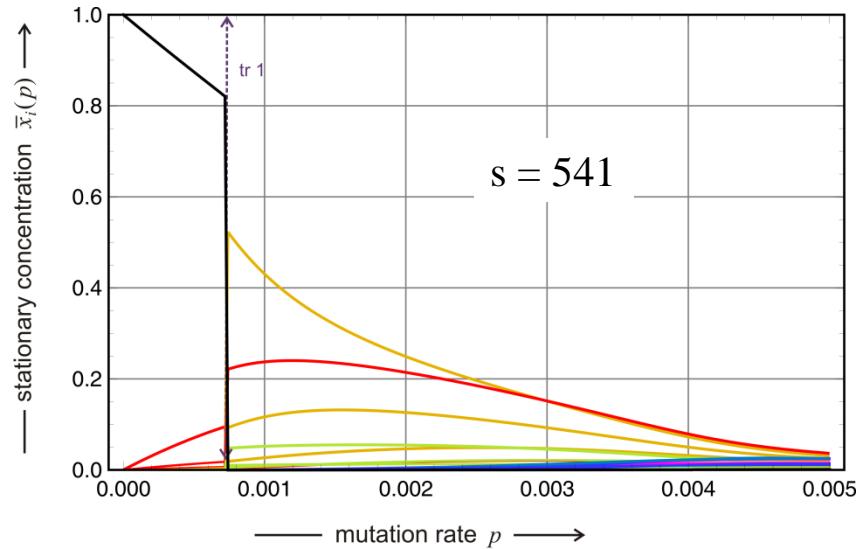
Error threshold on 'realistic' landscapes

$$n = 10, f_0 = 1.1, f_n = 1.0, d = 0.5$$



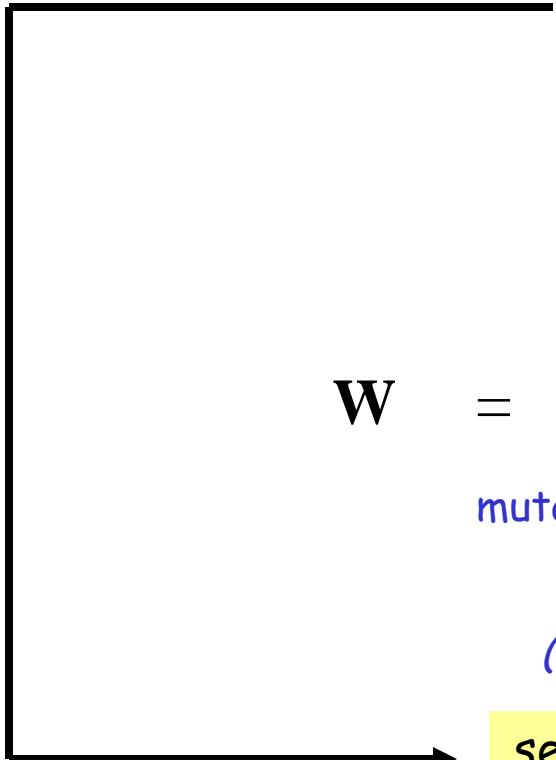
Error threshold on 'realistic' landscapes

$$n = 10, f_0 = 1.1, f_n = 1.0, d = 0.995$$



Error threshold on 'realistic' landscapes

$$n = 10, f_0 = 1.1, f_n = 1.0, d = 1.0$$



diagonalization of matrix \mathbf{W}

„ complicated but not complex ”

Complexity in molecular evolution

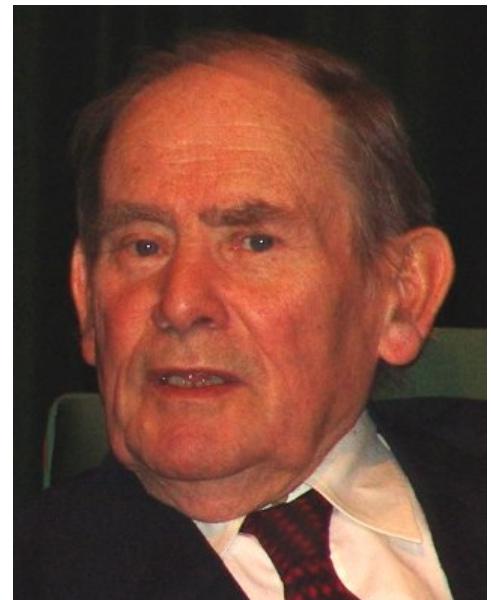
The new biology provides a hitherto unknown challenge for mathematicians, computer scientists, and theoretical biologists for mainly two reasons

enormous amount of data and

complexity of structure and dynamics:

… I was taught in the pregenomic era to be a hunter. I learnt how to identify the wild beasts and how to go out, hunt them down and kill them. We are now urged to be gatherers, to collect everything lying around and put it into storehouses.

Someday, it is assumed, someone will come and sort through the storehouses, discard all the junk, and keep the rare finds. The only difficulty is how to recognize them.

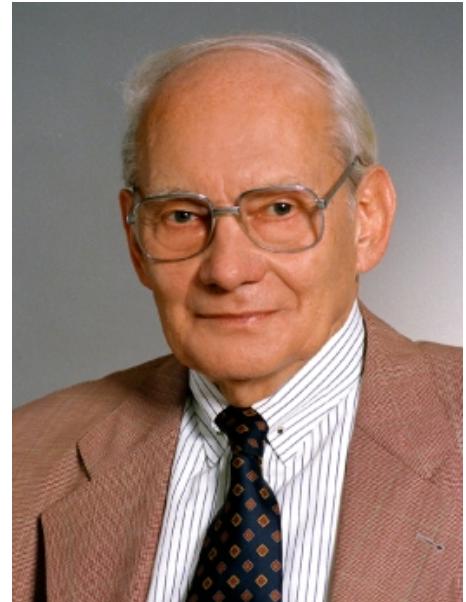


Sydney Brenner, 1927 -

Sydney Brenner. Hunters and gatherers. *The Scientist* **16**(4): 14, 2002

The „big data“ problem in bioinformatics

Theory - mathematics and computation
- cannot remove complexity, but it shows what kind of „regular“ behavior can be expected and what experiments have to be done to get a grasp on the irregularities.



Manfred Eigen, 1927 -
Preface to E. Domingo,
C.R. Parrish, J.J.Holland, eds.
Origin and Evolution of
Viruses. Academic Press 2008

Theory, mathematics and complexity

Coworkers

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

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

Heinz Engl, Philipp Kögler, James Lu, Stefan Müller, RICAM Linz, AT

Jord Nagel, Kees Pleij, Universiteit Leiden, NL

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, University of Southern Denmark, Odense, DK

Christian Forst, University of Texas, Southwestern Medical Center, TX

Thomas Wiehe, Ulrike Göbel, Walter Grüner, Stefan Kopp, Jacqueline 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



Universität Wien

Acknowledgement of support

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



Universität Wien

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

Happy 25th birthday IWR and
ad multos annos.

Thank you for your attention!

Web-Page for further information:

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

