Early evolution as an exercise in physics and chemistry

Peter Schuster

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

Shneior Lifson Memorial Lecture

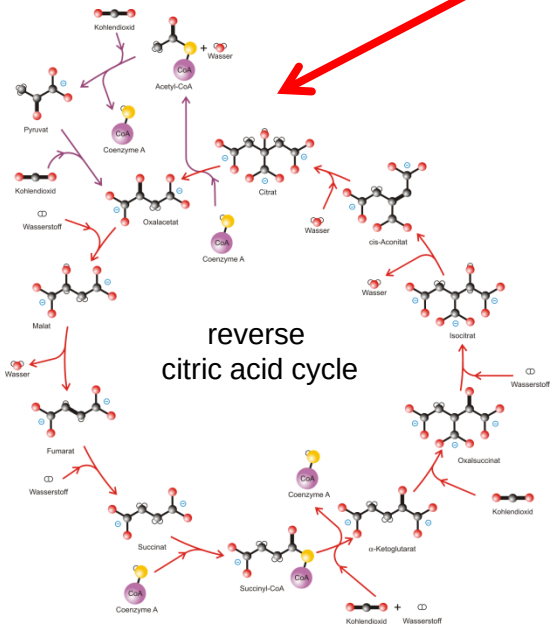
Weizmann Institute of Science, 04.03.2013

Web-Page for further information:

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

Prologue

prebiotic chemistry



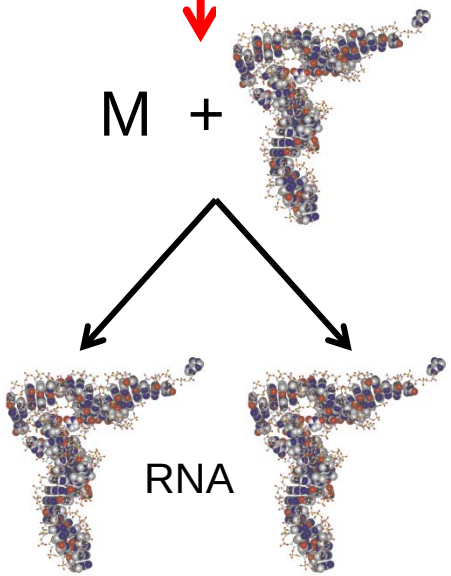
reverse
citric acid cycle

metabolism

RNA world

DNA + RNA + protein world

prebiotic chemistry

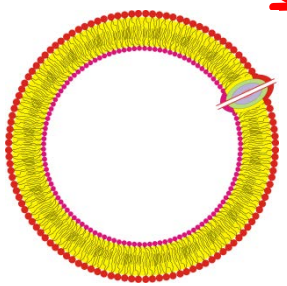


replication

RNA world

DNA + RNA + protein world

prebiotic chemistry



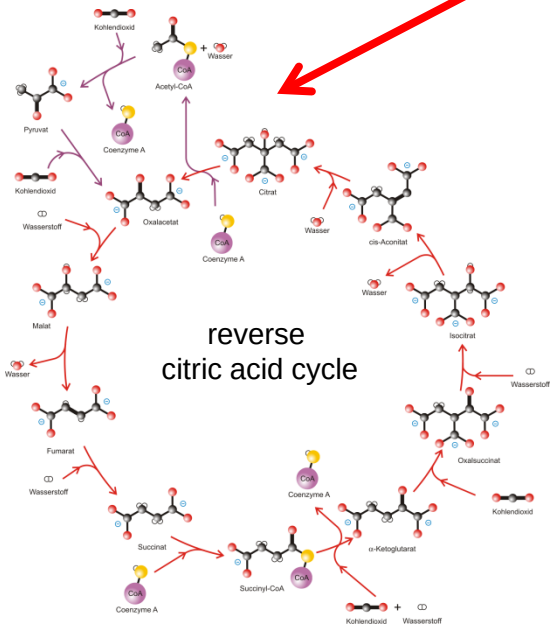
vesicles, composoms, …

multiphase systems

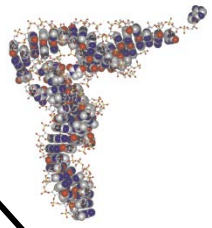
RNA world

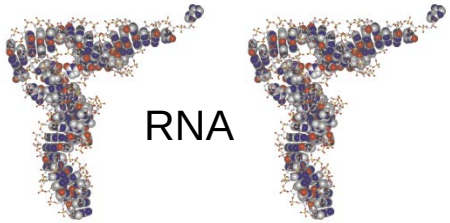
DNA + RNA + protein world

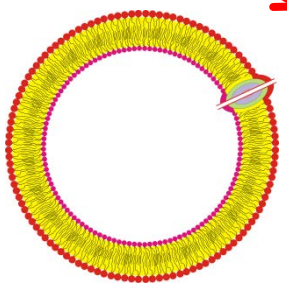
prebiotic chemistry



M +







metabolism

replication

vesicles, composoms, …

multiphasic systems

RNA world

DNA + RNA + protein world

Shneior Lifson. 1987. Chemical selection, diversity, teleonomy and the second law of thermodynamics. Reflections on Eigen‘s theory of self-organization of matter. *Biophyscial Chemistry* **26**:303-311.

Shneior and Hanna Lifson, Klosters 1995

Shneior Lifson. 1997. On the crucial stages in the origin of animate matter. *J.Mol.Evol.* **44**:1-8.

Shneior Lifson, Hanna Lifson. 1999. A model of prebiotic replication: Survival of the fittest versus extinction of the unfittest. *J.Theor.Biol.* **199**:425.

Shneior Lifson and the origin of life

J Mol Evol (1997) 44:1–8

JOURNAL OF MOLECULAR EVOLUTION

On the Crucial Stages in the Origin of Animate Matter

Shneior Lifson

Chemical Physics Department, Weizmann Institute of Science, Rehovot 76100, Israel

Received: 29 March 1996 / Accepted: 30 May 1996

Key words: Origin — Animate matter — Autocatalysis — Natural selection — Sequels — Complexity — Metabolism — Cellular organization — Genetic code

Shneior Lifson, 1914 - 2001

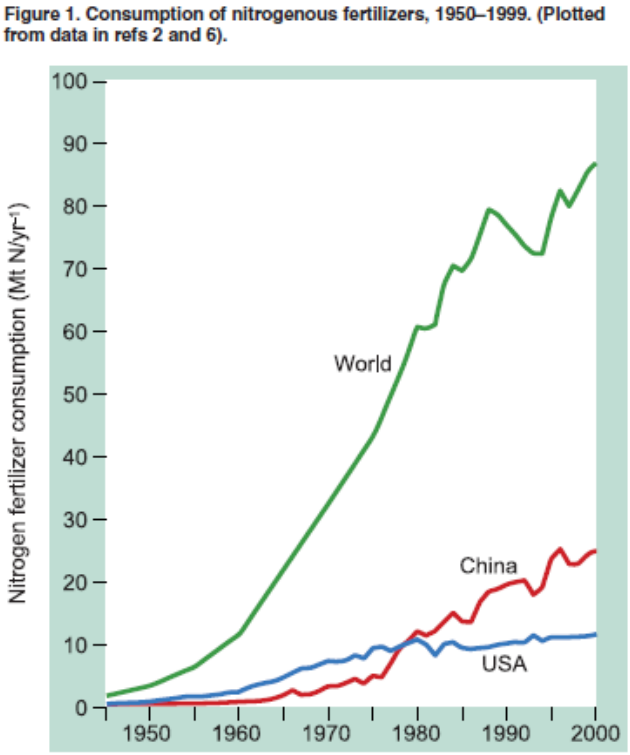
$$A + X \rightarrow 2X ; X \rightarrow D$$

Here, suffice it to recognize that *adaptation of autocatalysts to their changing environment by incorporating sequels into the autocatalytic process yields a great selective advantage.*

$$A + X \rightarrow 2X ; X \rightarrow D ; D \rightarrow A \quad \leftarrow \text{ recycling}$$

Shneior Lifson and the origin of life

Alexis Madrigal. 2008. How to make fertilizer appear out of thin air. 100 years Haber – Bosch process.



Vaclav Smil. 2002. *Ambio* **31**:126-131

Every fifth nitrogen atom in our body has seen a Haber-Bosch plant from inside at least once!

The importance of recycling in the modern world

1. History of molecular evolution and its applications
2. Why RNA is suitable for molecular evolution
3. Evolutionary dynamics of replication and mutation
4. Evolution and complexity

1. **History of molecular evolution and its applications**
2. Why RNA is suitable for molecular evolution
3. Evolutionary dynamics of replication and mutation
4. Evolution and complexity

Sol Spiegelman,
1914 - 1983

Evolution in the test tube:

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

G. F. Joyce

DOI: 10.1002/anie.200701369

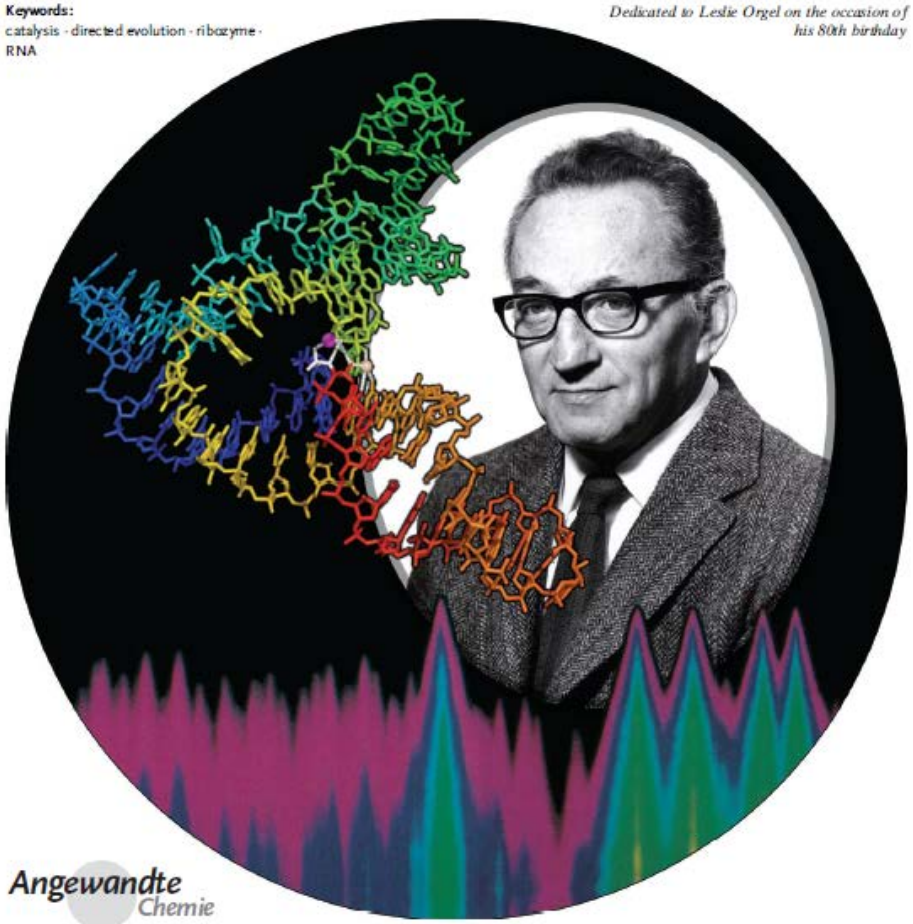
Molecular Evolution

Forty Years of In Vitro Evolution**

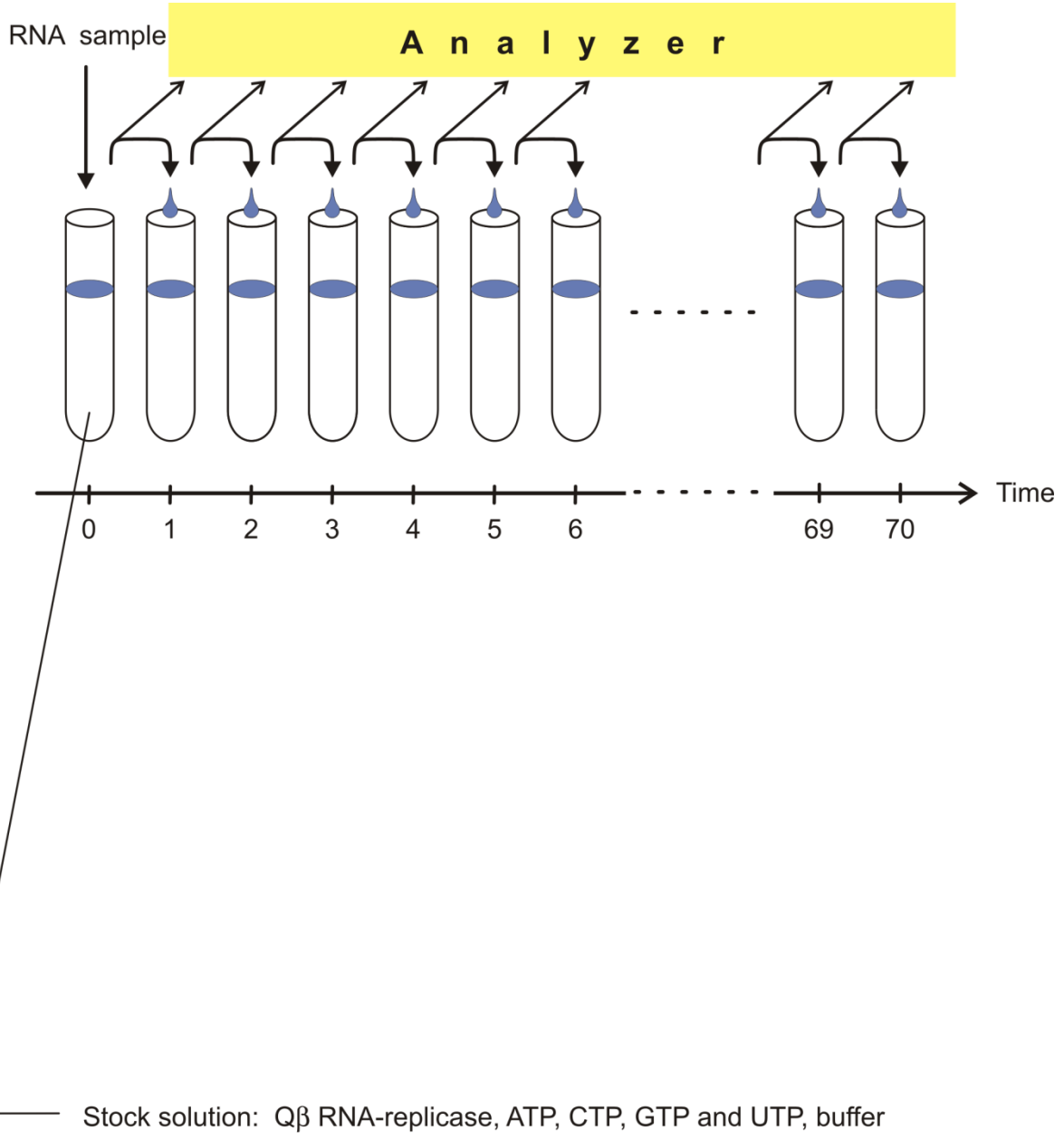
*Gerald F. Joyce**

Keywords:
catalysis · directed evolution · ribozyme · RNA

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



The serial transfer technique for *in vitro* evolution



Reproduction of the original figure of the serial transfer experiment with Qβ RNA

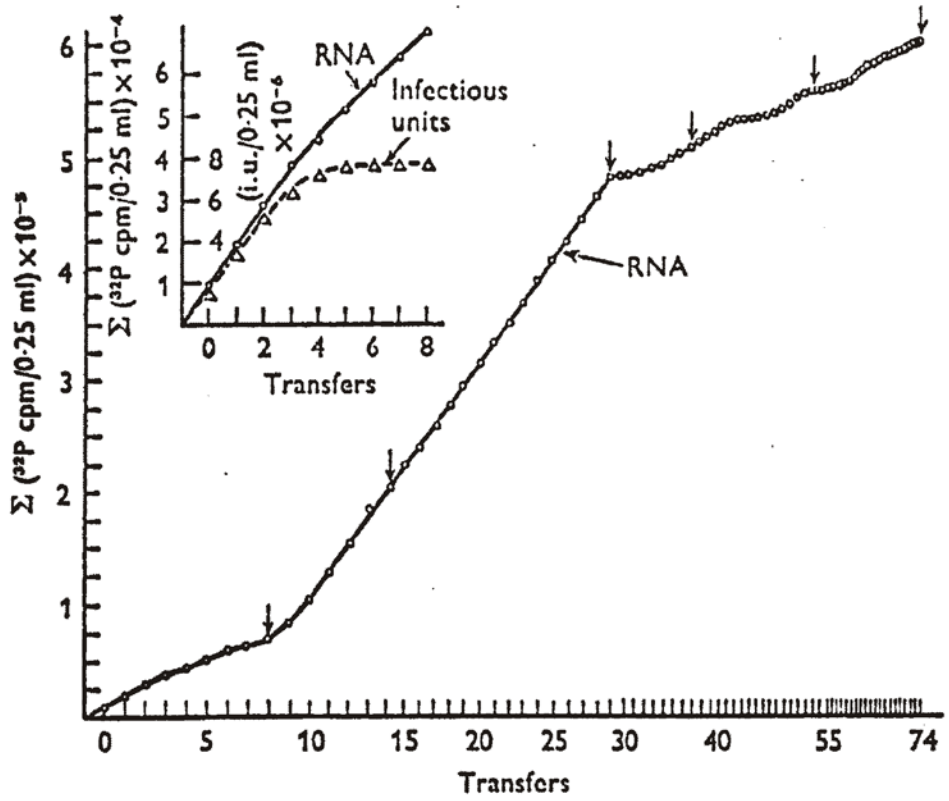
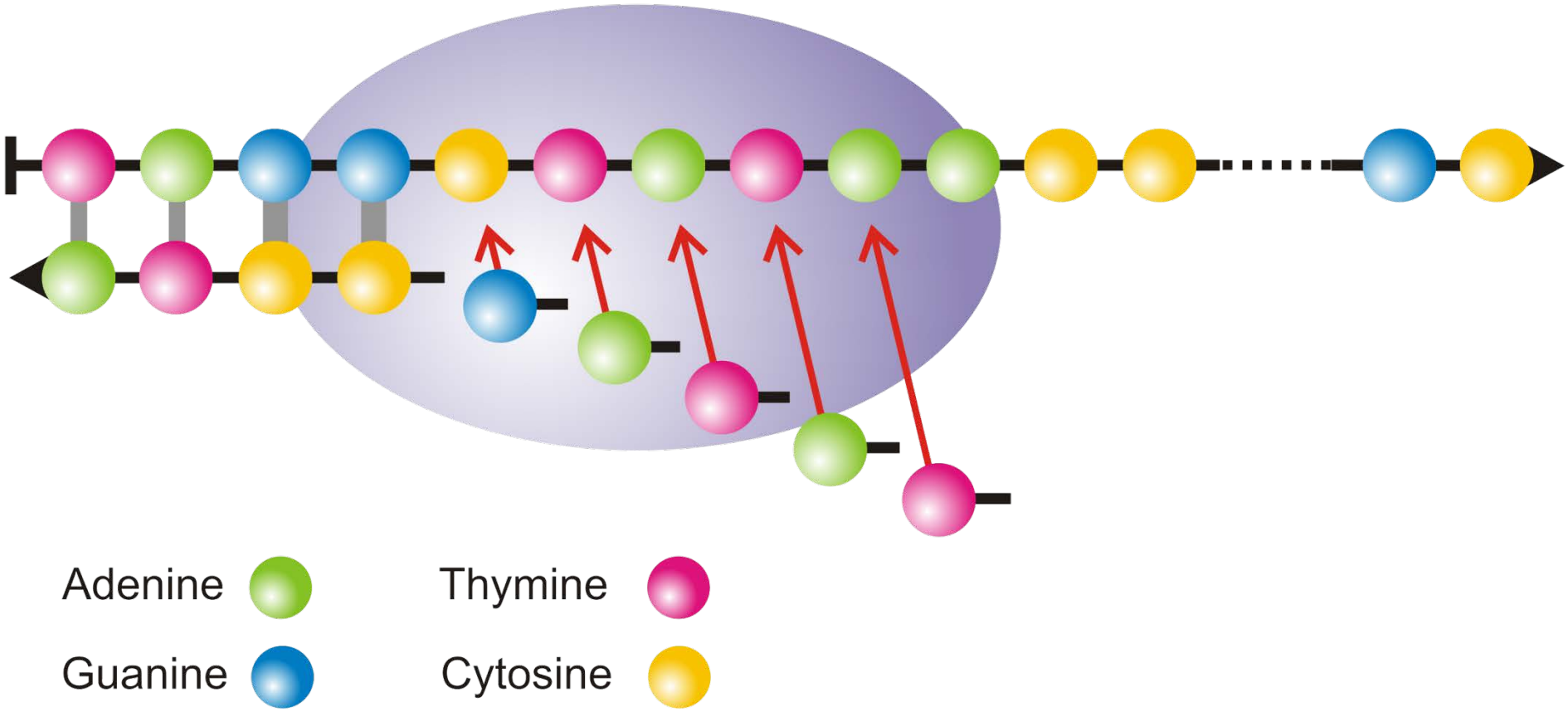
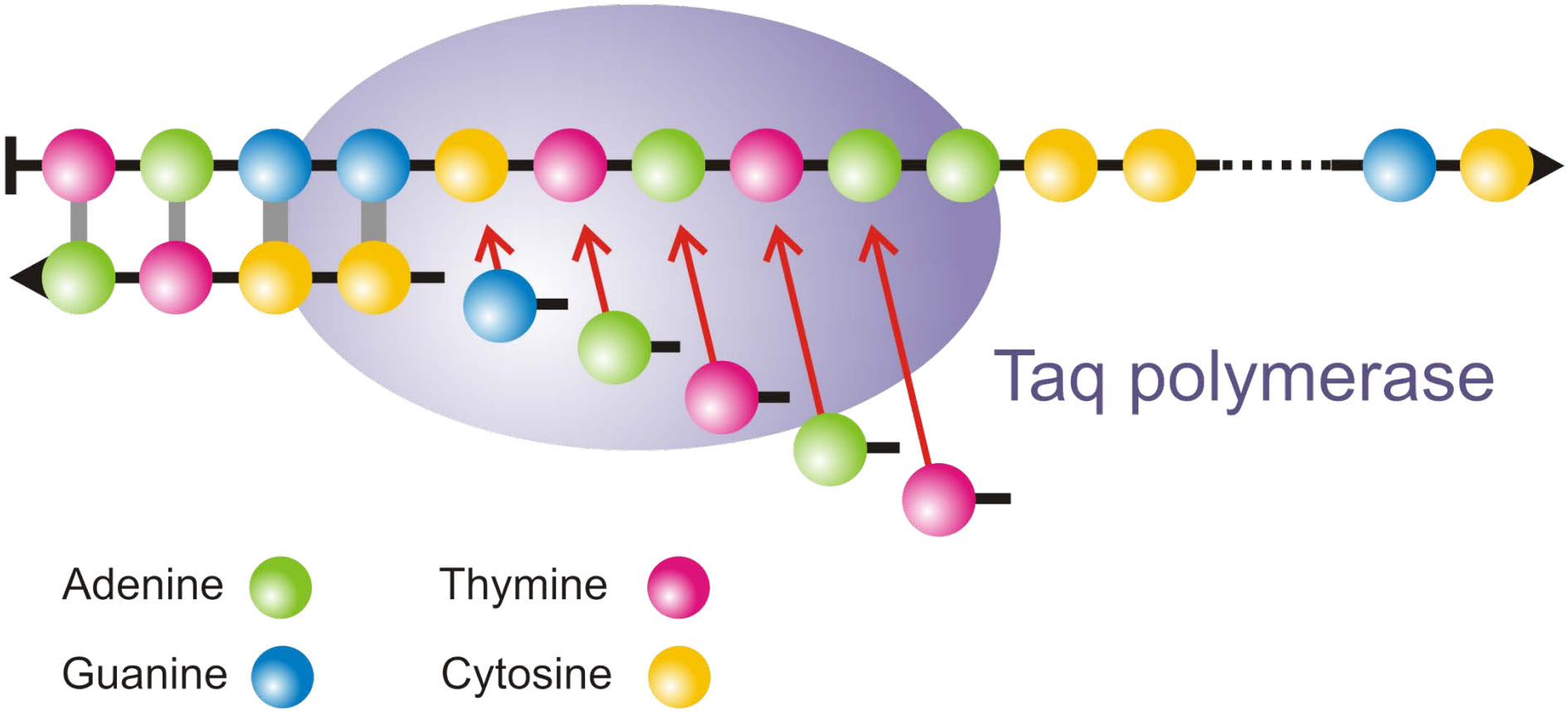


Fig. 9. Serial transfer experiment. Each 0·25 ml standard reaction mixture contained 40 μg of Qβ replicase and ^{32}P-UTP. The first reaction (0 transfer) was initiated by the addition of 0·2 μg ts-1 (temperature-sensitive RNA) and incubated at 35 °C for 20 min, whereupon 0·02 ml was drawn for counting 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).

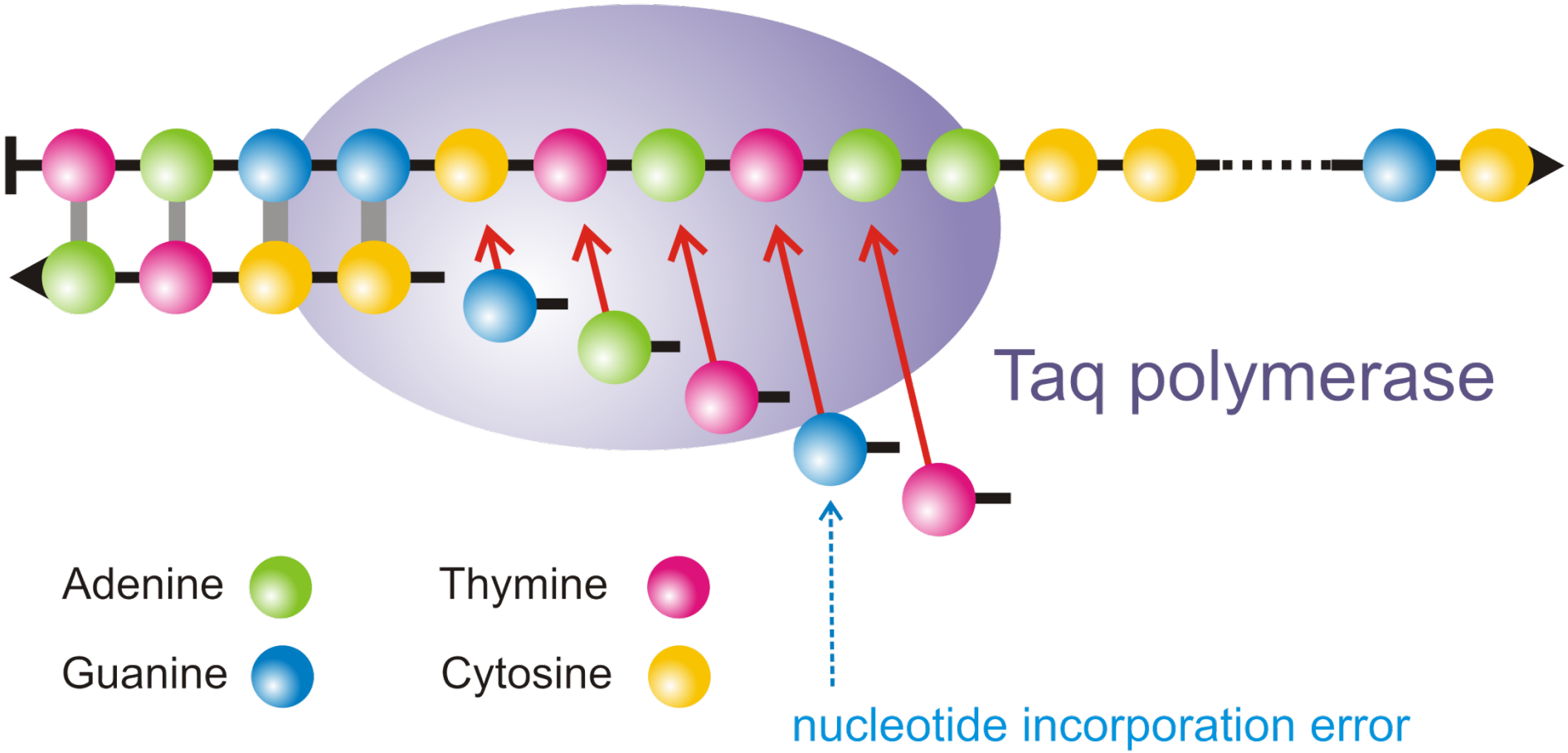
D.R.Mills, R,L,Peterson, S.Spiegelman, *An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule*. Proc.Natl.Acad.Sci.USA **58** (1967), 217-224



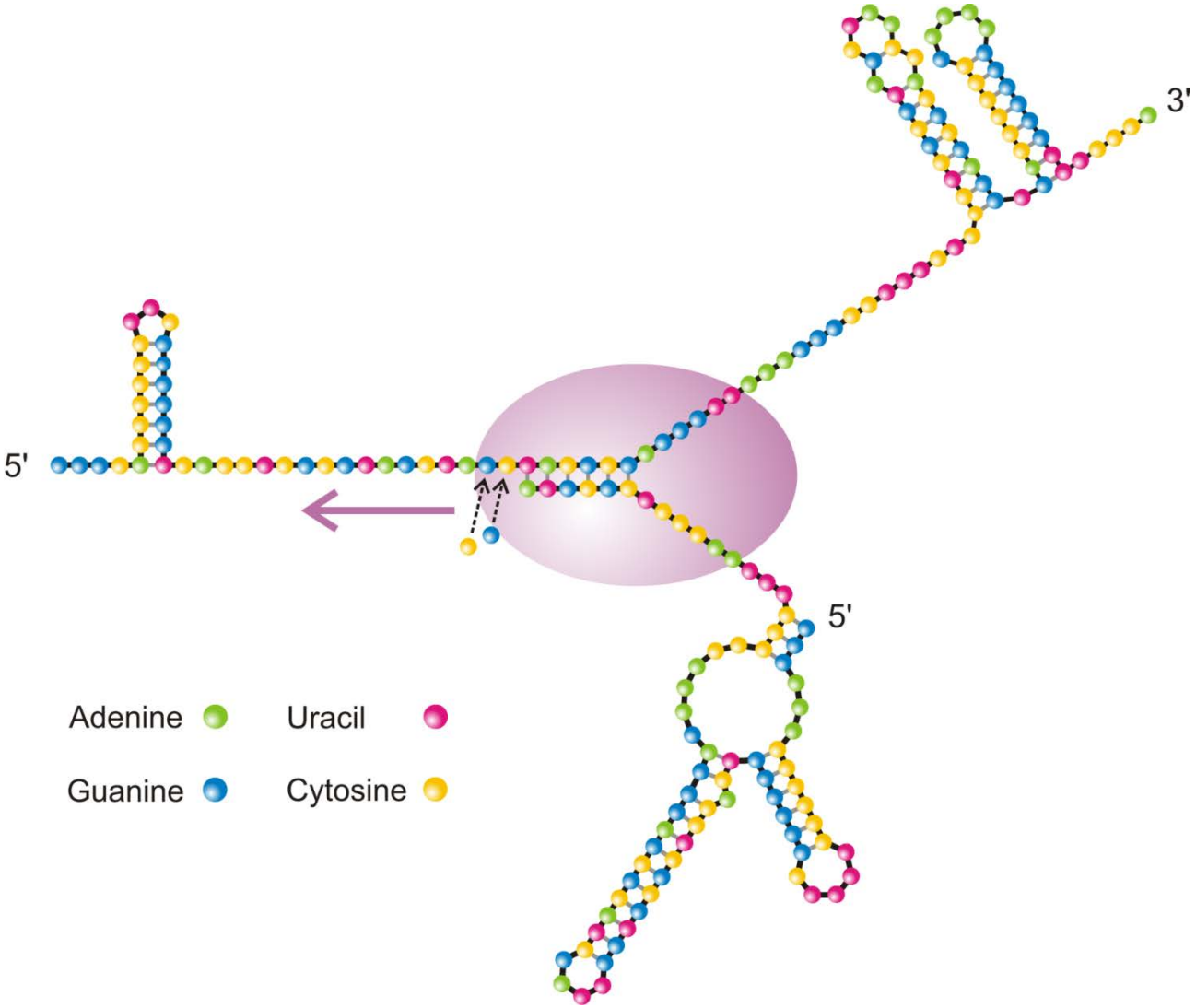
The logic of DNA (or RNA) replication and mutation

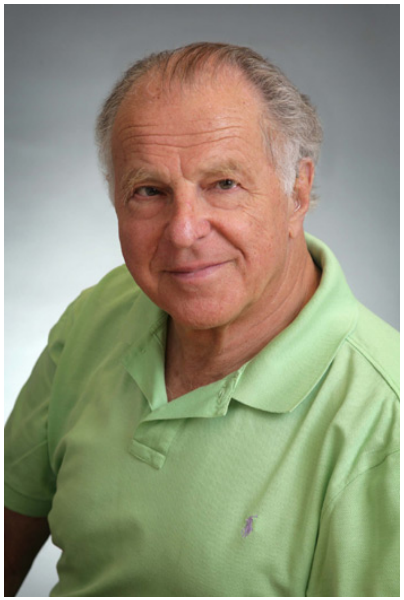


The logic of DNA (or RNA) replication and mutation



The logic of DNA (or RNA) replication and mutation



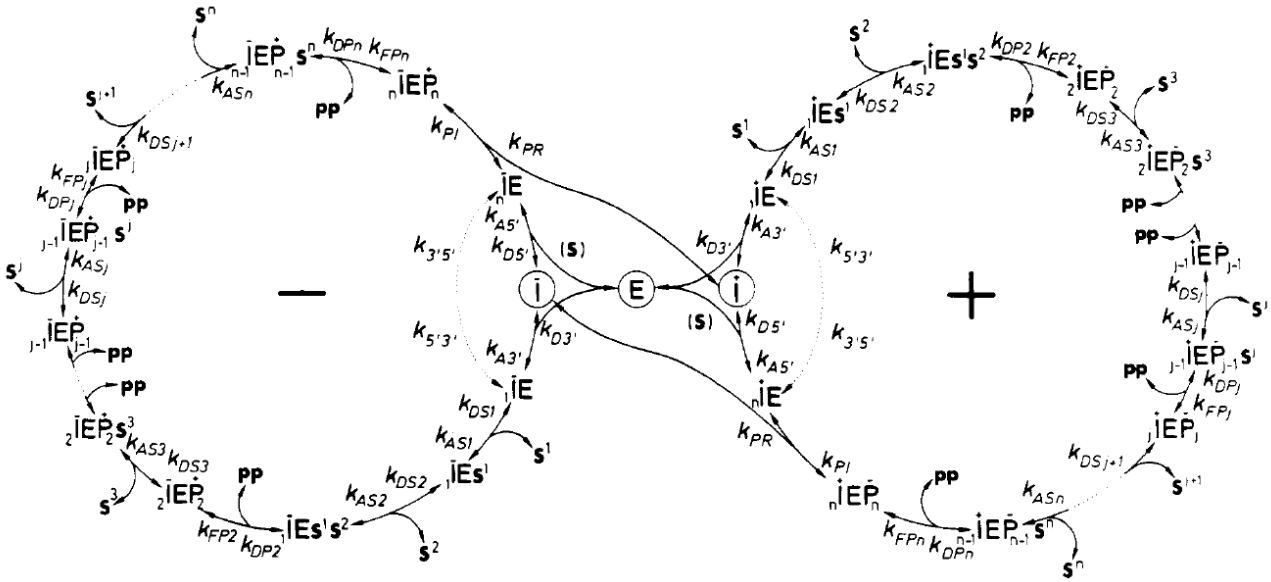


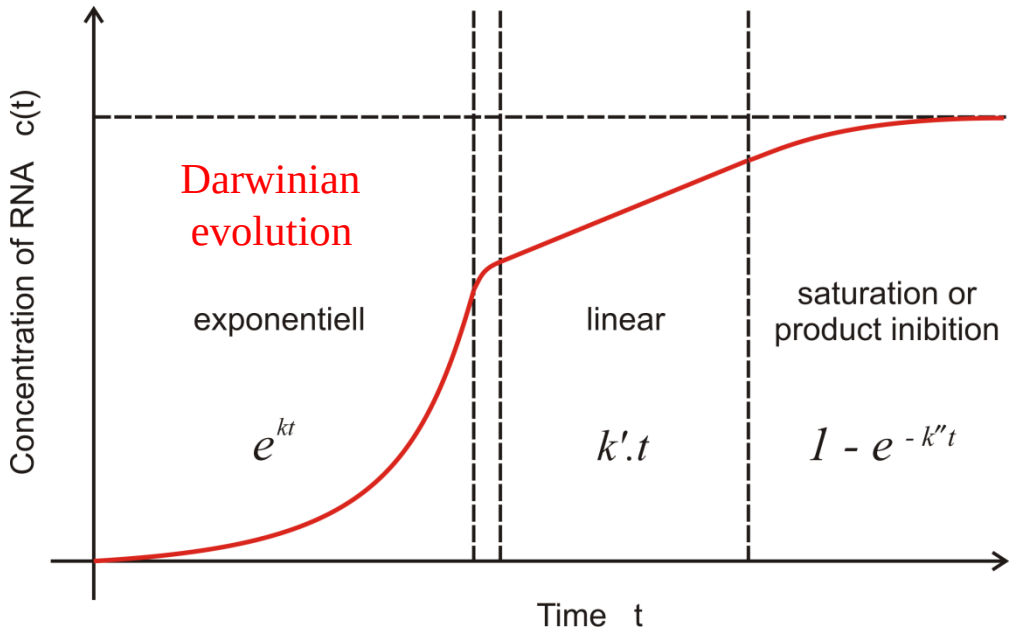
Charles Weissmann
1931-

RNA replication by Qβ-replicase

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

Christof K. Biebricher,
1941-2009





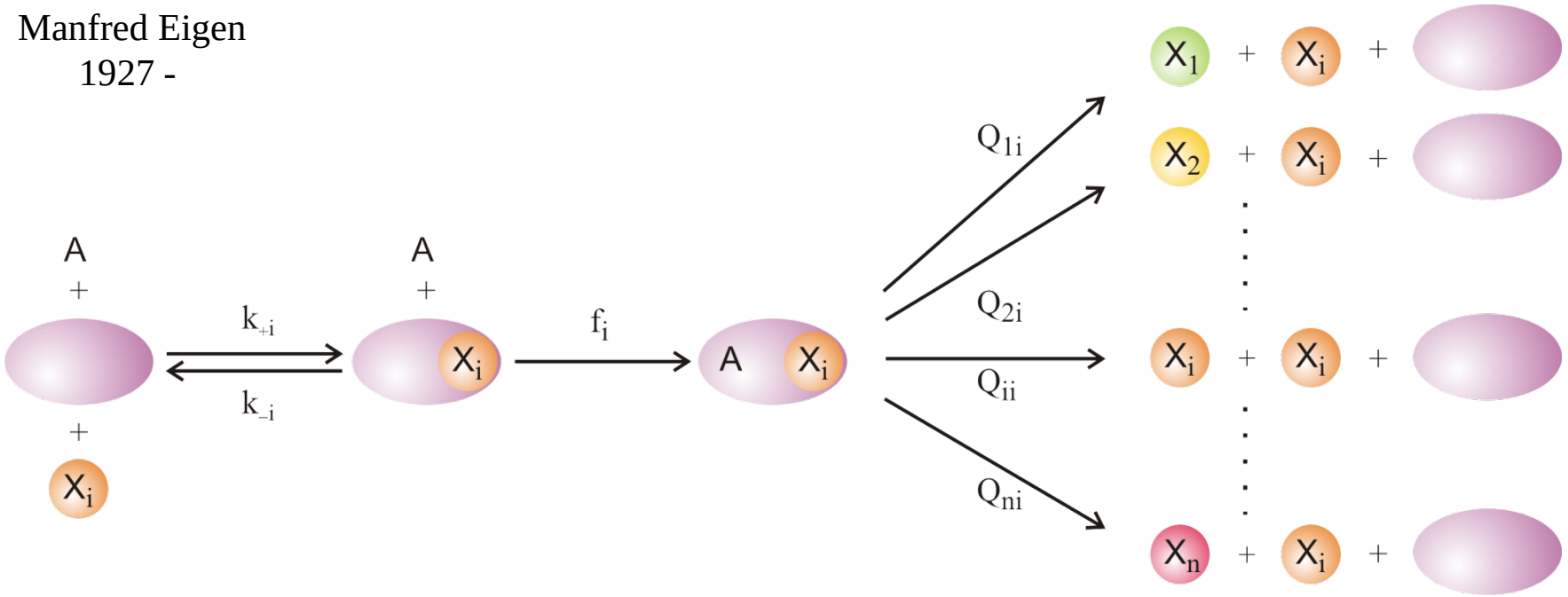
Kinetics of RNA replication

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

Manfred Eigen
1927 -

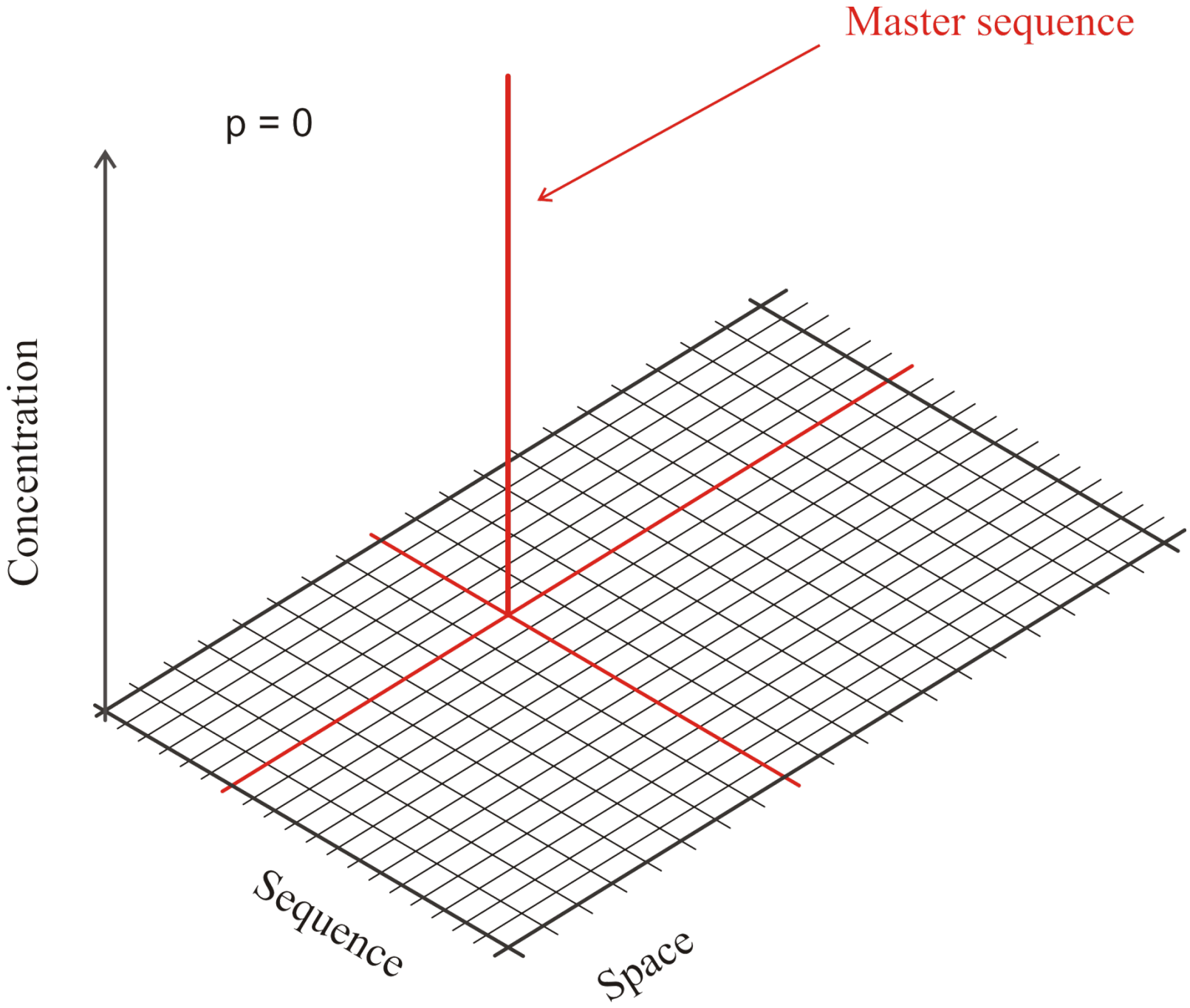
$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

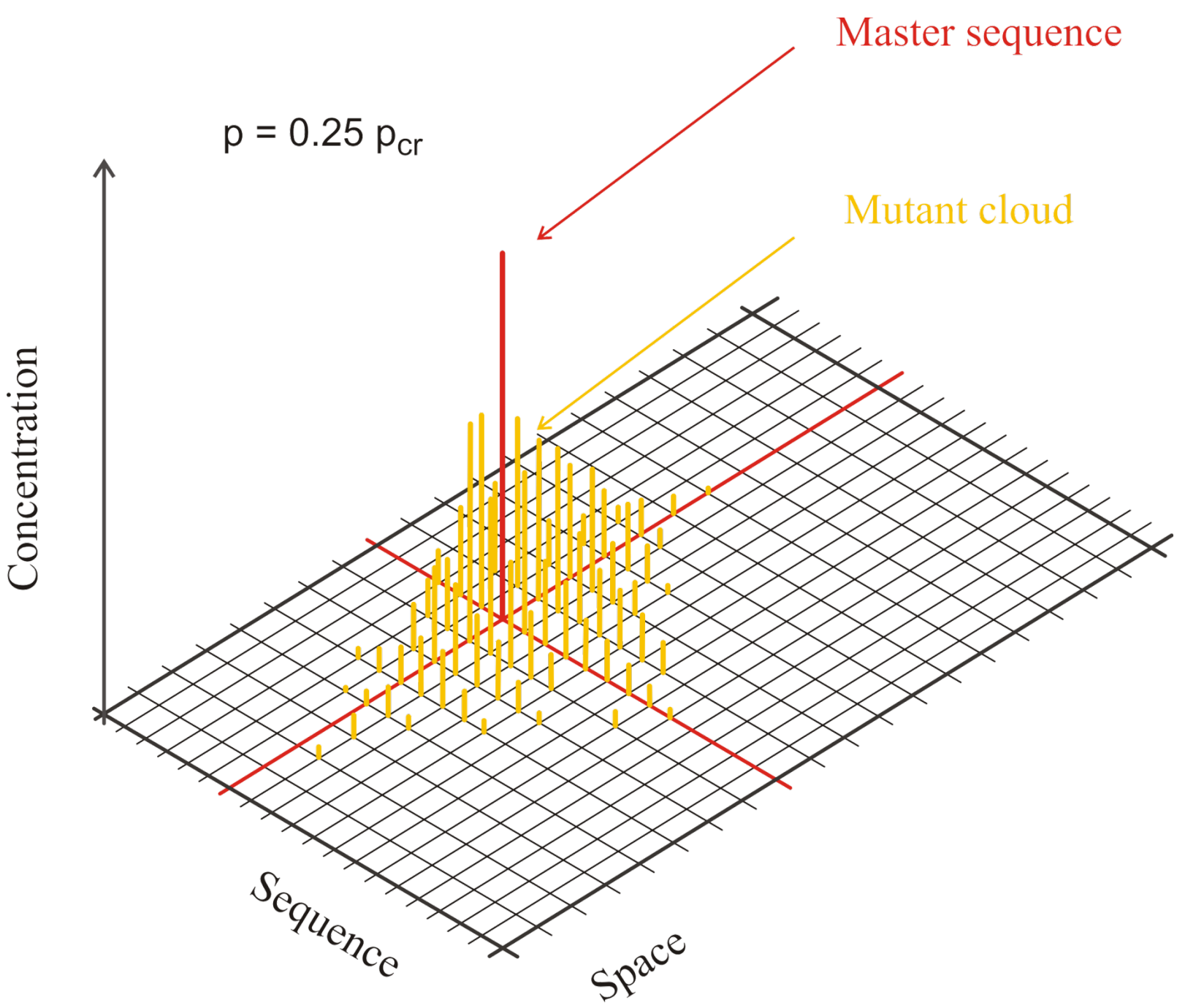
$$W_{ji} = Q_{ji} \cdot f_i\,, \quad \sum_{i=1}^{n} x_i = 1, \quad \Phi = \sum_{i=1}^{n} f_i\, x_i$$

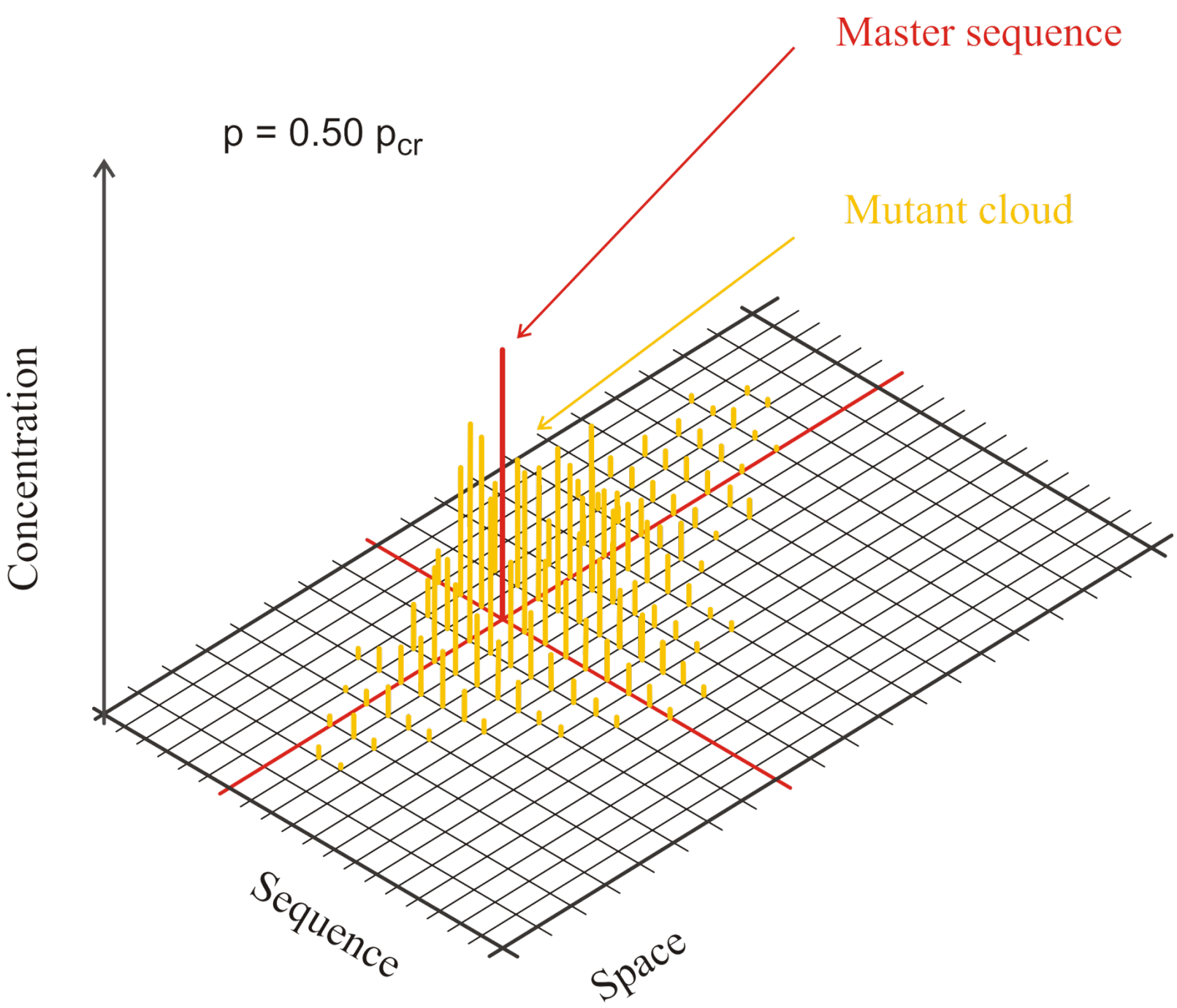


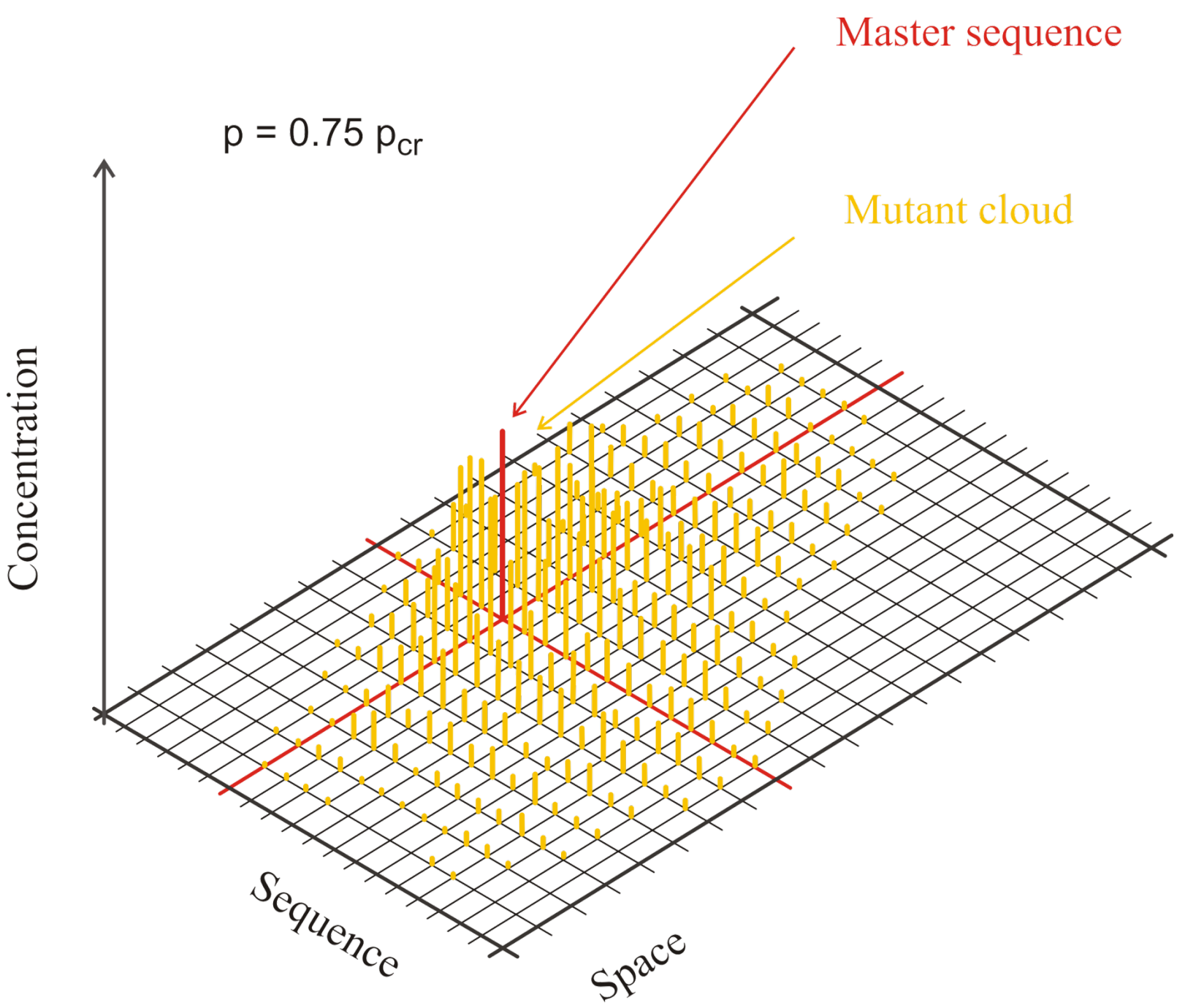
Mutation and (correct) replication as parallel chemical reactions

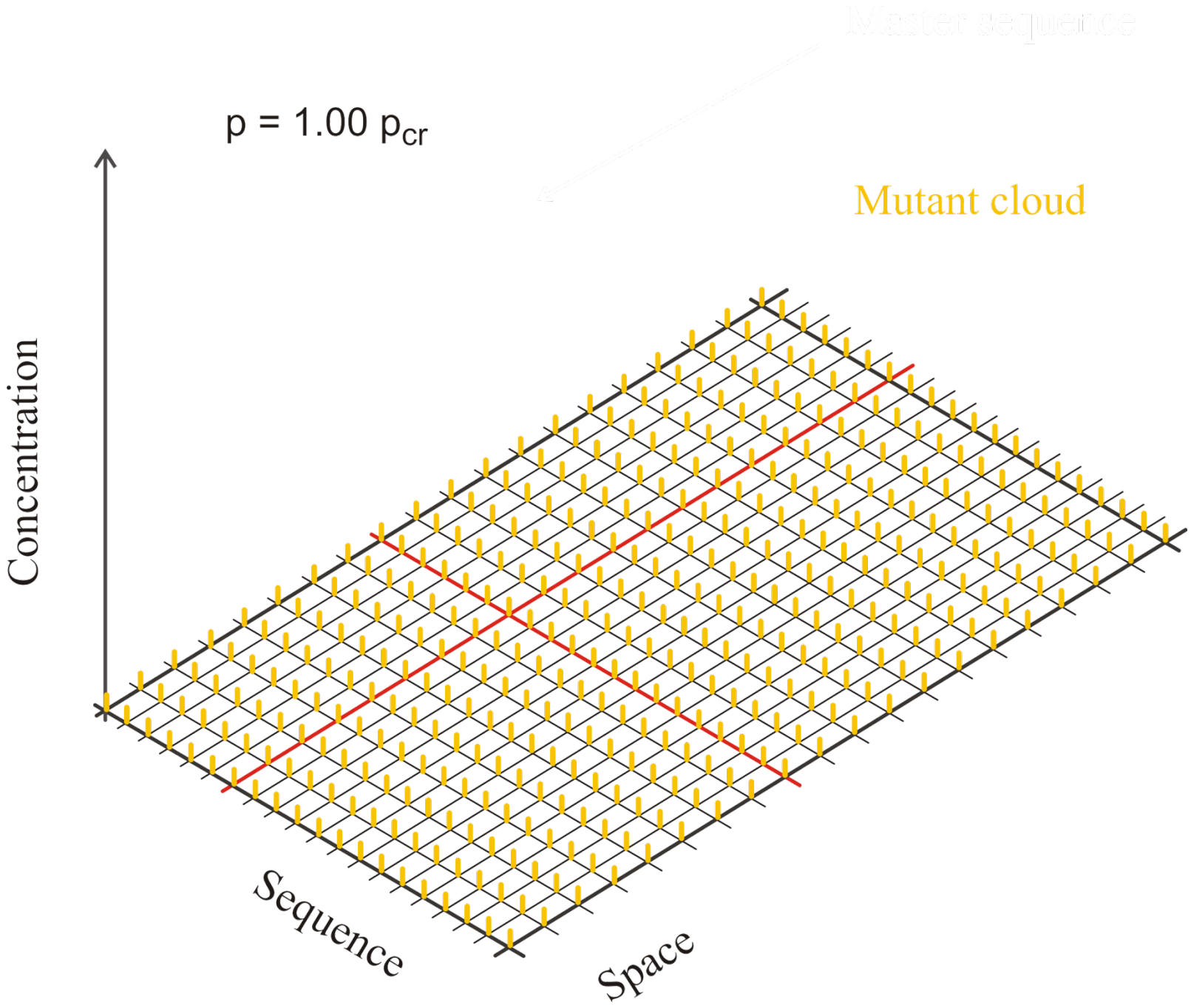
M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977. *Naturwissenschaften* 64:541, 65:7 und 65:341

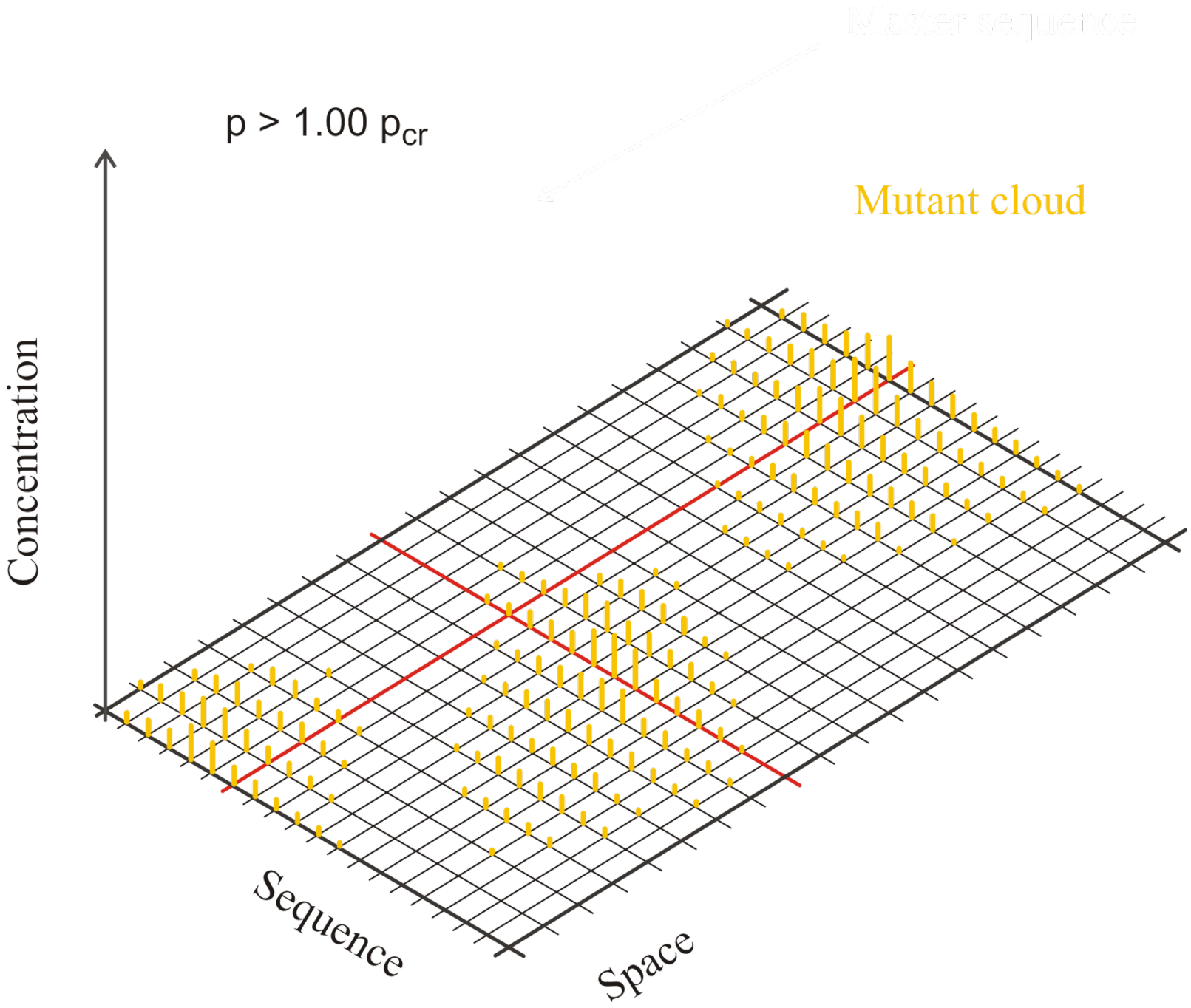

Master sequence
p = 0
Concentration
Sequence
Space

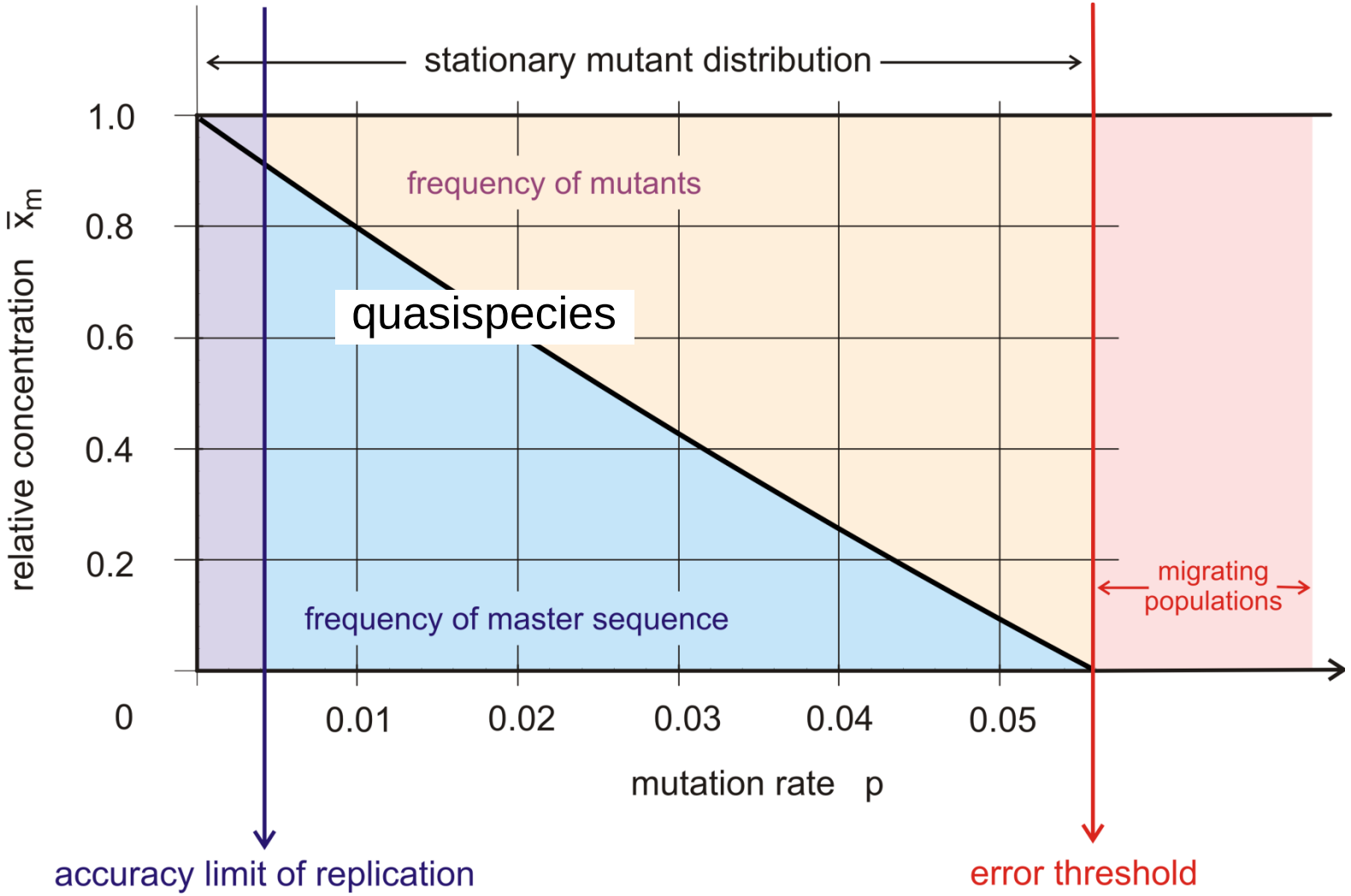

Master sequence
p = 0.25 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p = 0.50 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p = 0.75 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p = 1.00 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p > 1.00 p_cr
Mutant cloud
Concentration
Sequence
Space



The error threshold in replication and mutation

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

$$\sigma = \frac{(1-x_m)\, f_m}{\sum_{j \neq m} f_j\, x_j} \;\ldots\; \text{superiority of master sequence},$$

n … chain length , p … error rate ,

$Q = (1-p)^n$ … replication accuracy

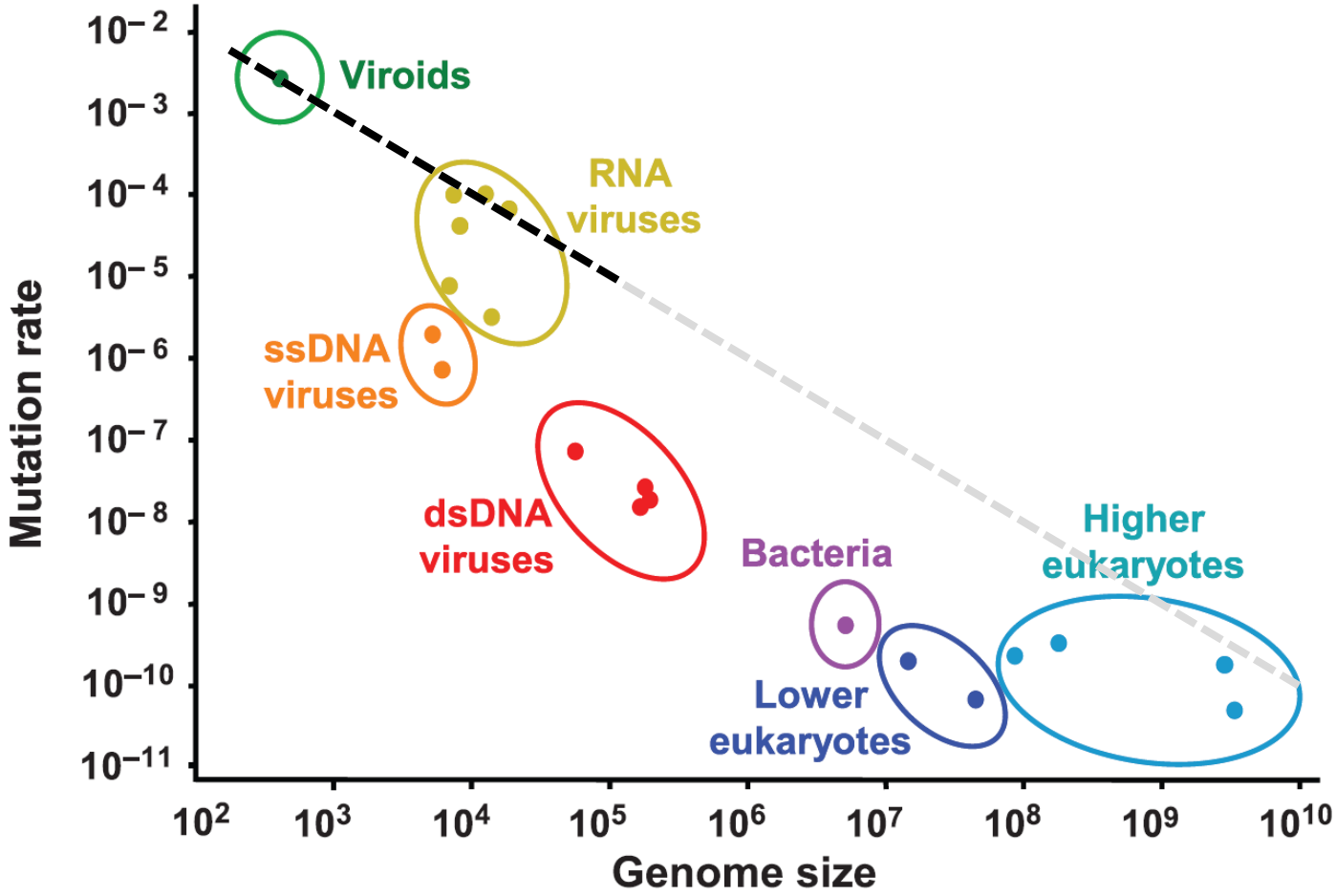
p … constant : $n_{\max} \approx \dfrac{\ln\sigma}{p}$ maximum chain length

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

Consequences of the error threshold

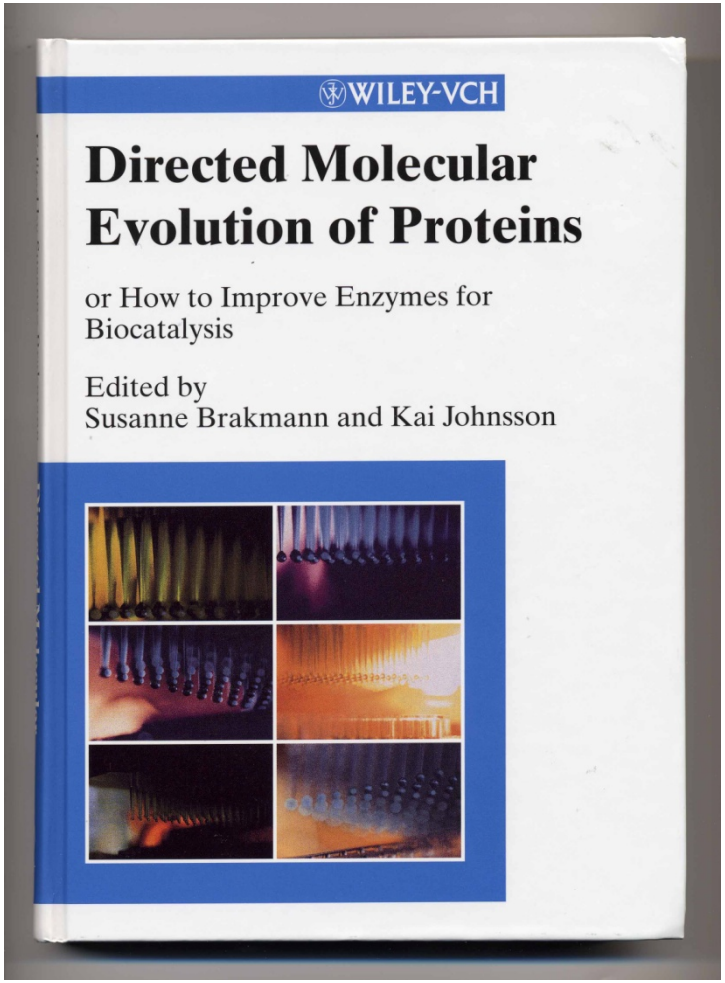
organism	mutation rate per genome	reproduction event
RNA virus	1	replication
retroviruses	0.1	replication
bacteria	0.003	replication
eukaryotes	0.003	cell division
eukaryotes	0.01 – 0.1	sexual reproduction

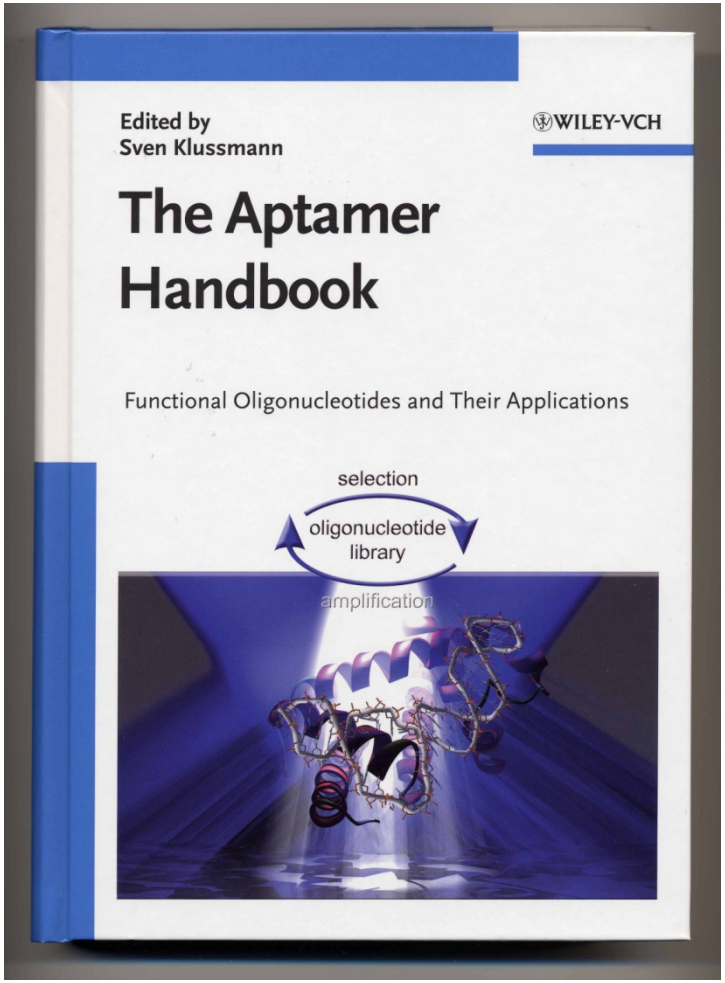
John W. Drake, Brian Charlesworth, Deborah Charlesworth and James F. Crow. 1998. Rates of spontaneous mutation. *Genetics* 148:1667-1686.



Selma Gago, Santiago F. Elena, Ricardo Flores, Rafael Sanjuán. 2009. Extremely high mutation rate of a hammerhead viroid. Science 323:1308.

Mutation rate and genome size





Application of molecular evolution to problems in biotechnology

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 107 (2005) 115–116
Virus Research
www.elsevier.com/locate/virusres

Preface

Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations on enzymes involved in protein synthesis, as a theory of aging. In those times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be eliminated through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variation in differentiated organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to unveil the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Population geneticists rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparison of electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA recombination, and rapid nucleic acid sequencing techniques, molecular analyses of genomes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Most cellular processes, including such essential information-bearing and transferring events as genome replication, transcription and translation, are increasingly perceived as inherently inaccurate. Viruses, and in particular RNA viruses, are among the most extreme examples of exploitation of replication inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in quantitative terms as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Recently, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged as a possible antiviral strategy. This is the topic of the current special issue of ***Virus Research***.

Few would nowadays doubt that one of the major obstacles for the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principles that have shaped biological evolution over eons, that is, repeated rounds of reproduction with genetic variation, competition and selection, often perturbed by random events such as statistical fluctuations in population size. However, with viruses the consequences of the operation of these very same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be also observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens begs for alternative antiviral strategies, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect different lines of evidence (both theoretical and experimental) on which antiviral designs based on genetic deterioration inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted to progeny. Closely related to the theoretical developments have been numerous experimental studies on quasispecies dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of existing as mutant spectra rather than defined genetic entities, remarkably expand their potential to overcome selective pressures intended to limit their replication. Indeed, the use of antiviral inhibitors in clinical practice and the design of vaccines for a number of major RNA virus-associated diseases, are currently presided by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replicases, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy received an important impulse by the observation that ribavirin (a licensed antiviral nucleoside analogue) may be exerting, in some systems, its antiviral activity through enhanced mutagenesis. This has encouraged investigations on new mutagenic base analogues, some of them used in anticancer chemotherapy. Some chapters summarize these important biochemical studies on cell entry pathways and metabolism of mutagenic agents, that may find new applications as antiviral agents.

This volume intends to be basically a progress report, an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I can envisage (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific mutagens and inhibitors in an effective way, finding synergistic combinations and avoiding antagonistic ones as well as severe clinical side effects. (ii) Another on a deeper understanding of the metabolism of mutagenic agents, in particular base and nucleoside analogues. This includes identification of the transporters that carry them into cells, an understanding of their metabolic processing, intracellular stability and alterations of nucleotide pools, among other issues. (iii) Still another line of needed research is the development of new mutagenic agents specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with anticancer research, but others should result from the designs of new molecules, based on the structures of viral polymerases. I really hope that the reader finds this issue not only to be an interesting and useful review of the current situation in the field, but also a stimulating exposure to the major problems to be faced.

The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of ***Virus Research***, and then taken enthusiastically by Luis Enjuanes, recently appointed as Editor of ***Virus Research***. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of ***Virus Research***, Brian Mahy, for their continued interest and support to the research on virus evolution over the years.

My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
Universidad Autónoma de Madrid
Centro de Biología Molecular "Severo Ochoa"
Consejo Superior de Investigaciones Científicas
Cantoblanco and Valdeolmos
Madrid, Spain
Tel.: +34 91 497 84858/9; fax: +34 91 497 4799
E-mail address: edomingo@cbm.uam.es

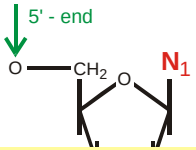
Available online 8 December 2004

doi:10.1016/j.virusres.2004.11.001

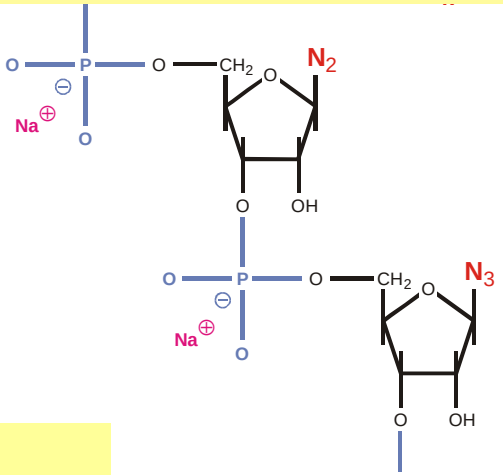
Esteban Domingo
1943 -

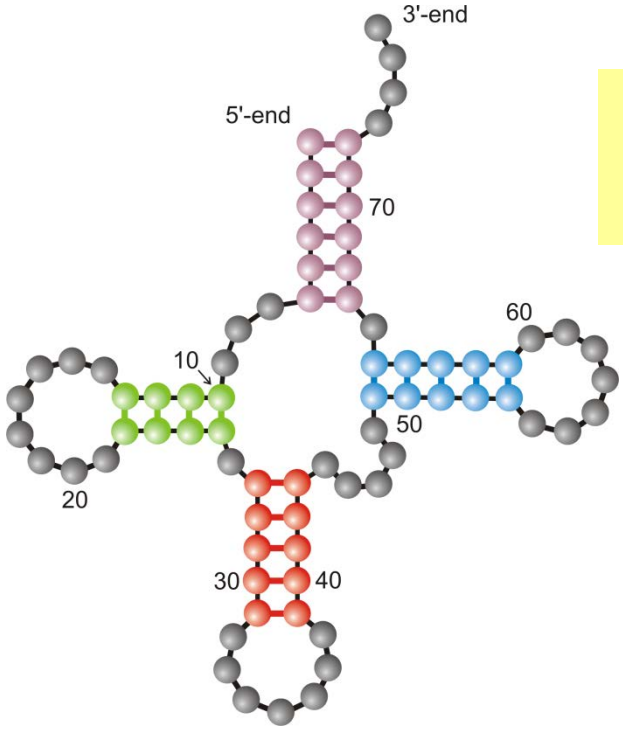
Application of quasispecies theory to the fight against viruses

1. History of molecular evolution and its applications
2. **Why RNA is suitable for molecular evolution**
3. Evolutionary dynamics of replication and mutation
4. Evolution and complexity



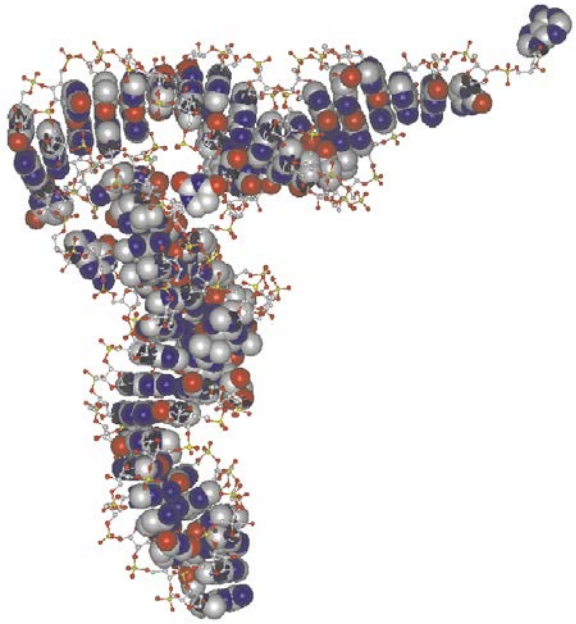
5'-end GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA 3'-end

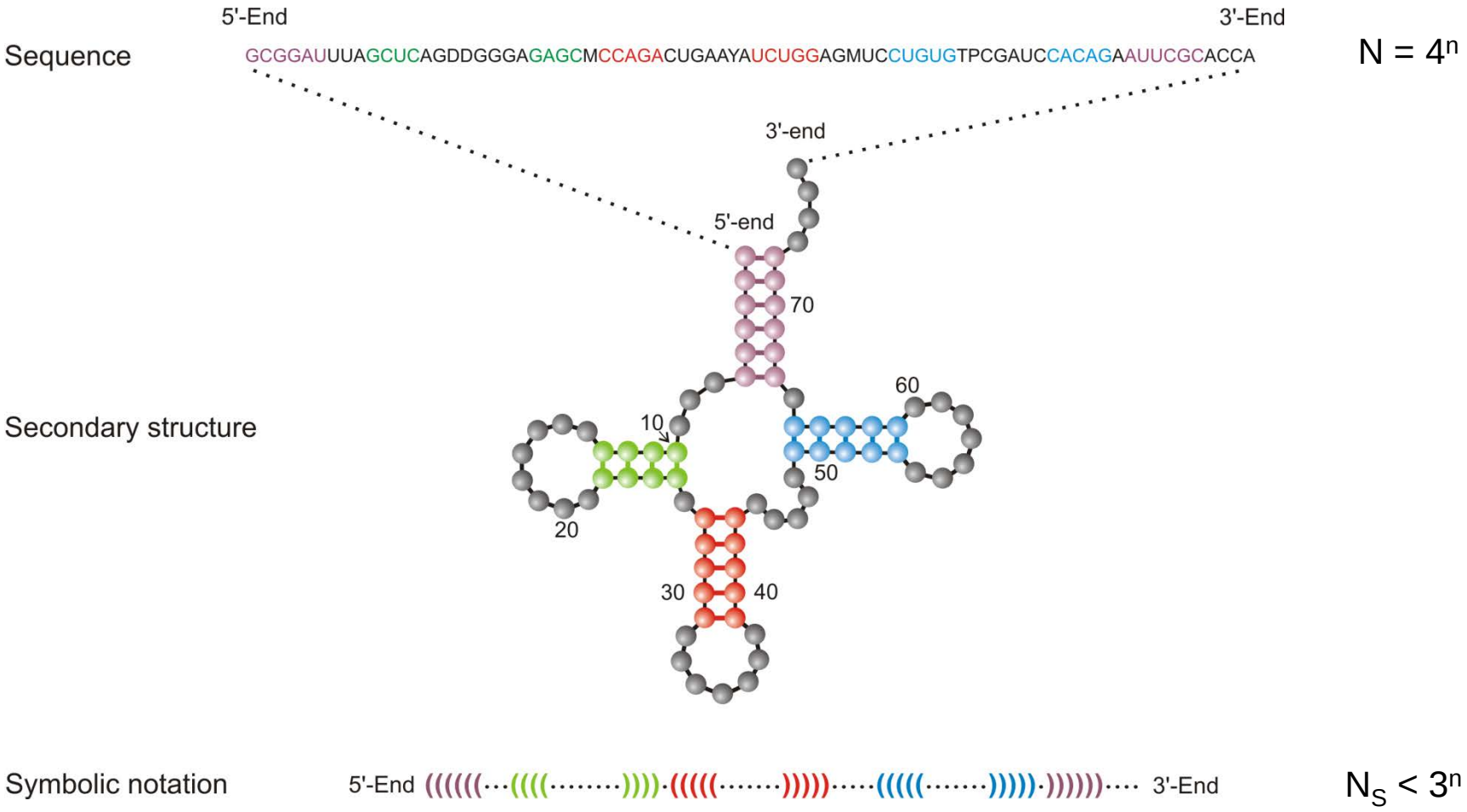




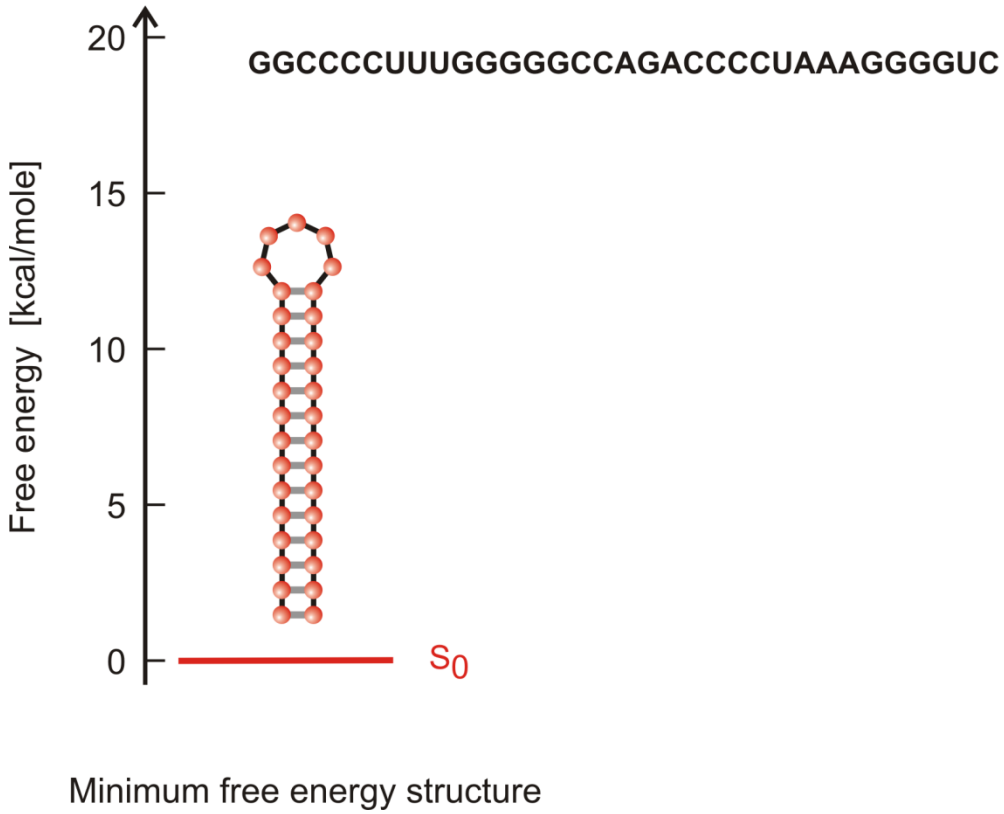
RNA structure

The molecular phenotype





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



The notion of structure

RNA sequence GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

linear programming

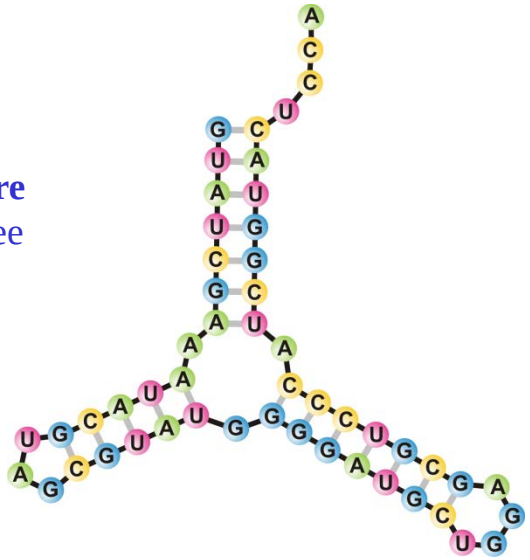
RNA folding:

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**

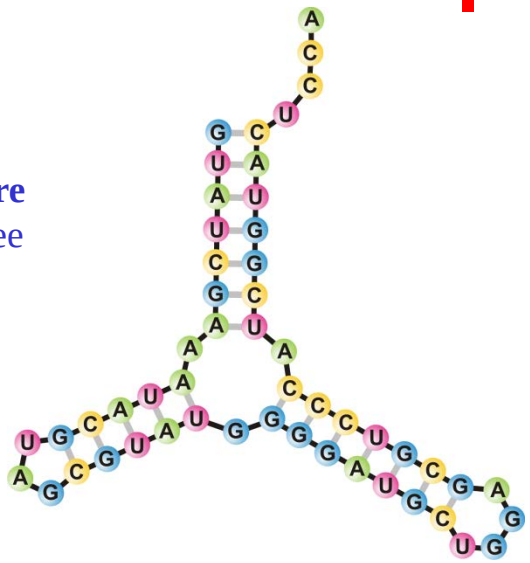
iterative determination of a sequence for the given secondary structure

inverse Folding Algorithm

inverse folding of RNA:

biotechnology, **design of biomolecules** with predefined structures and functions

RNA structure of minimal free energy

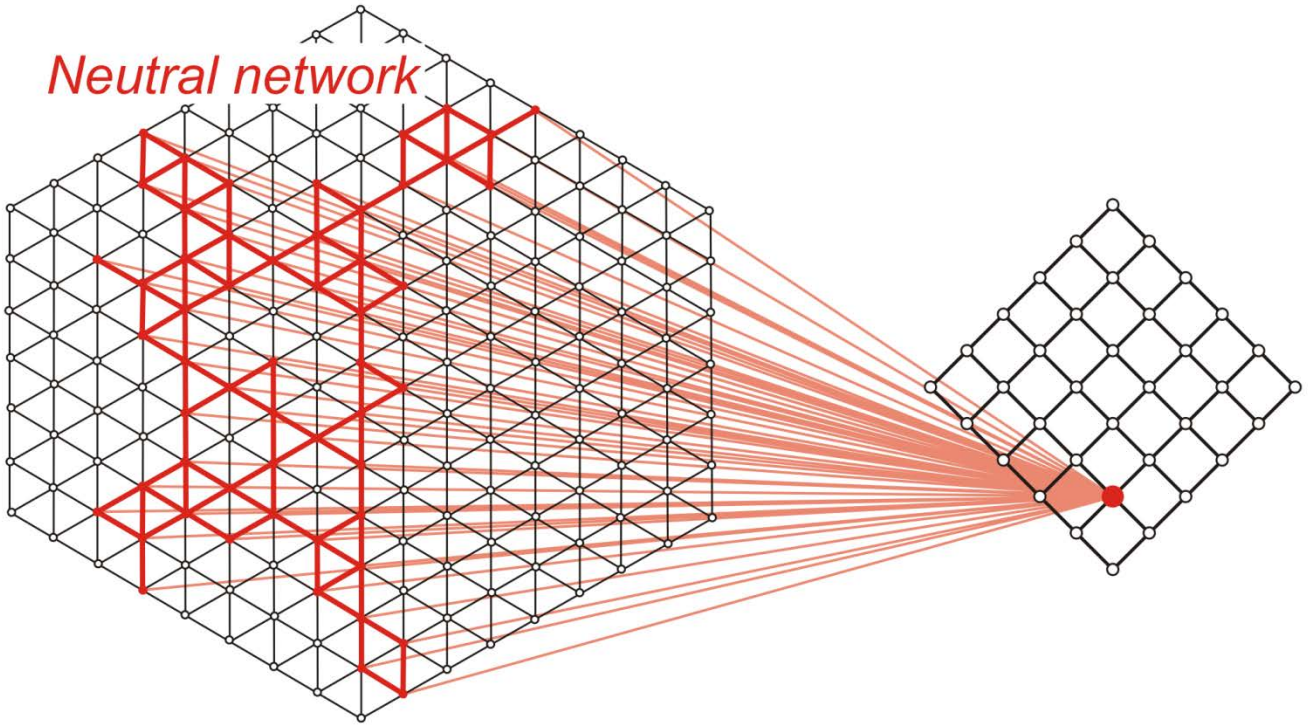


From RNA structure to sequence

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 Siederissen, Hakim Tafer, Christioh Flamm, Peter F. Stadler, and Ivo L. Hofacker.
ViennaRNA Package 2.0.
Algorithms Mol. Biol. **6**:26, 2011

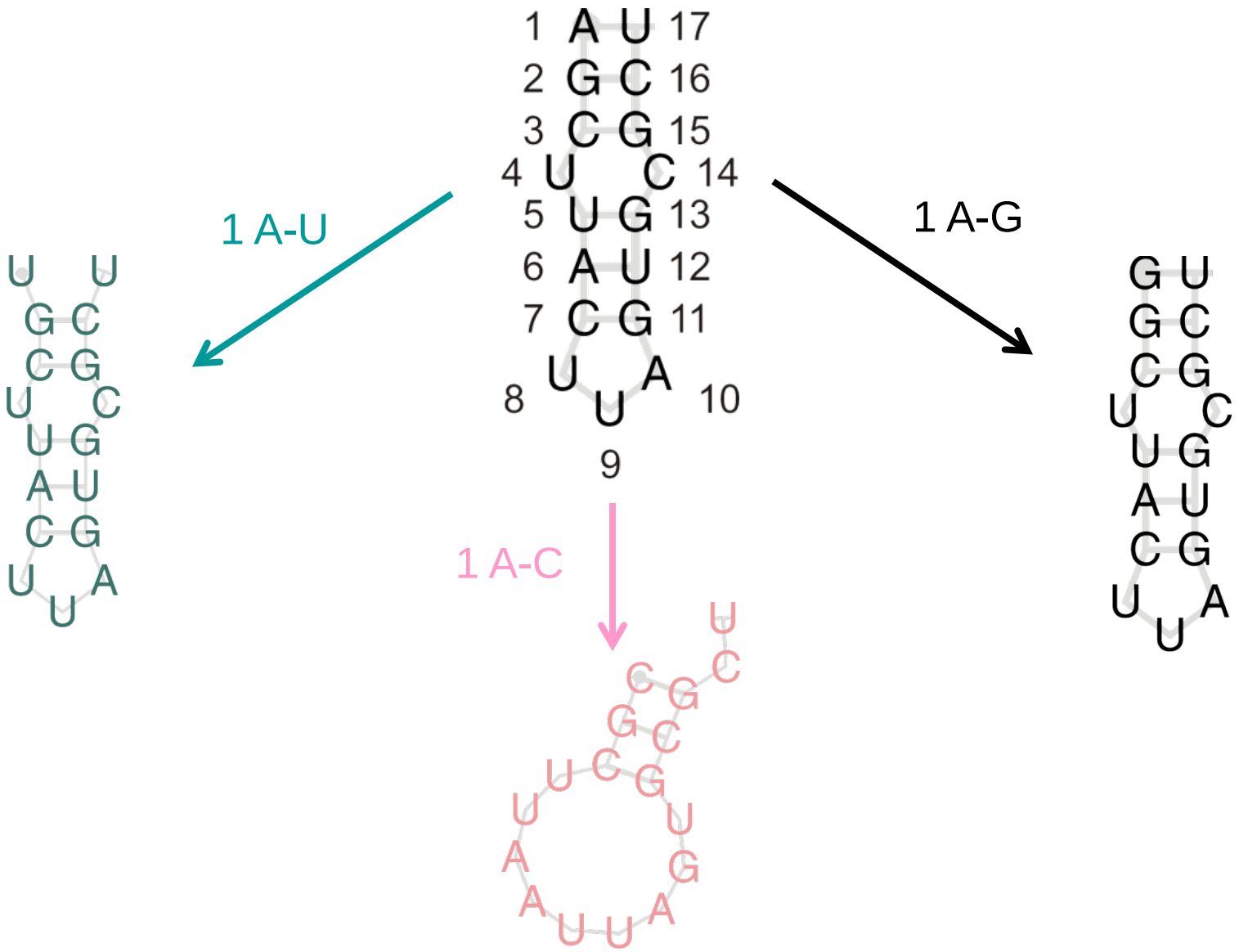


Sequence space

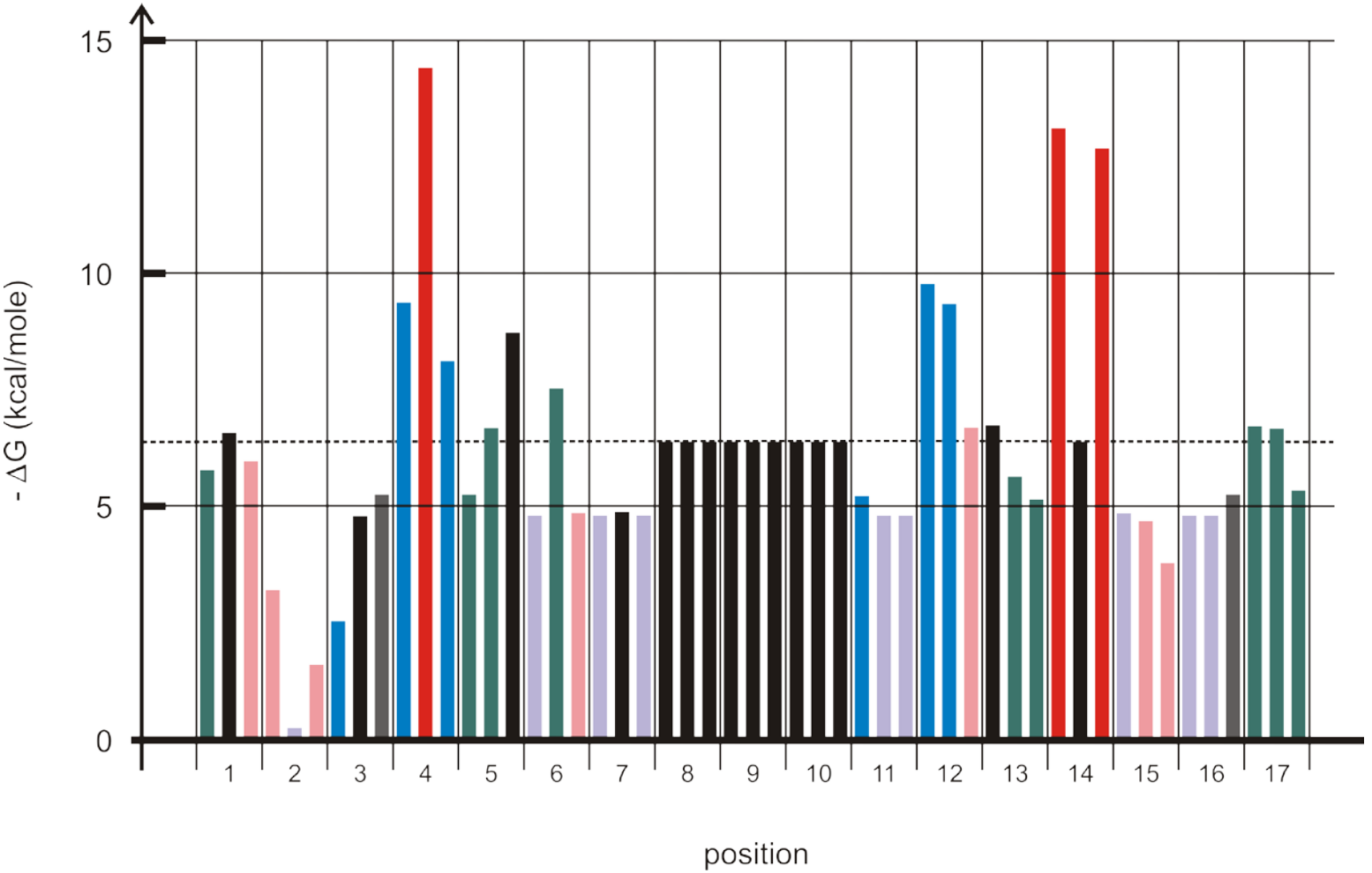
Structure space

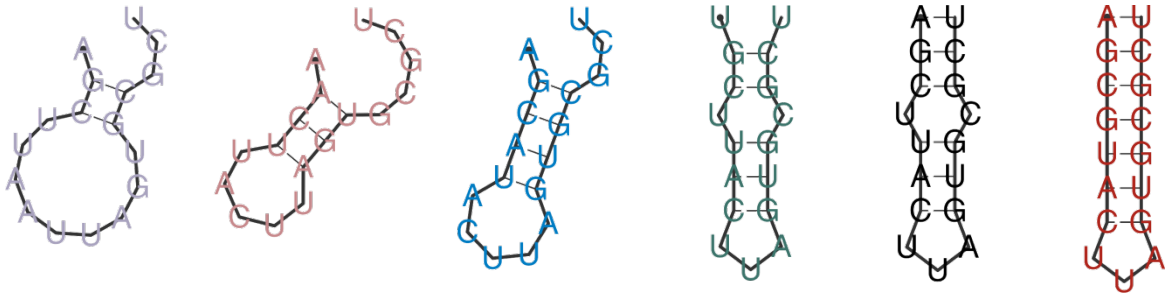
many genotypes ⇒ one phenotype

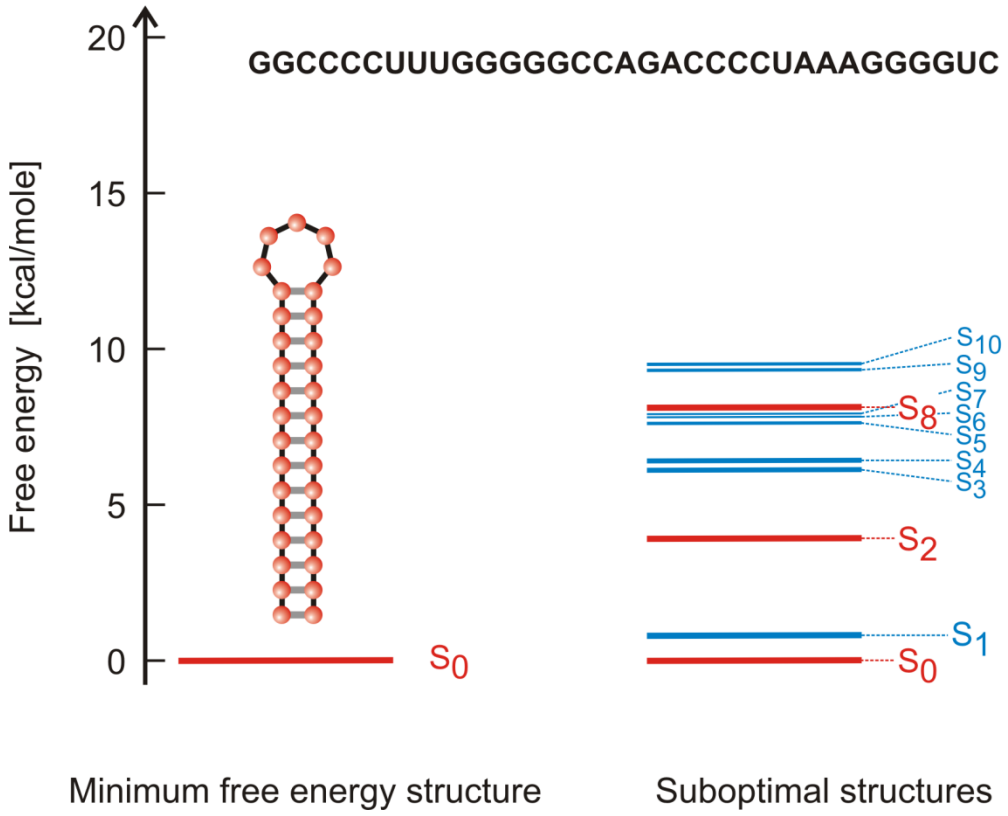
AGCUUAACUUAGUCGCU



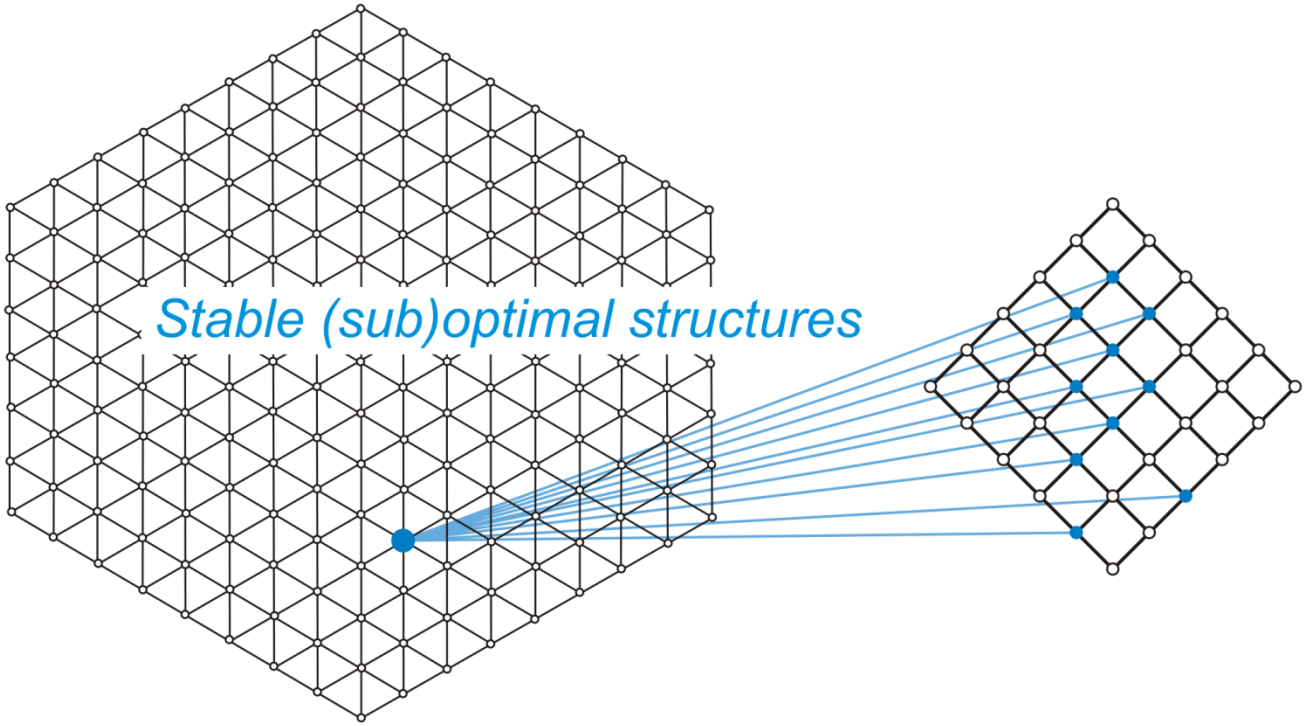

1 A U 17
2 G C 16
3 C G 15
4 U C 14
5 U G 13
6 A U 12
7 C G 11
U A
8 U 10
9

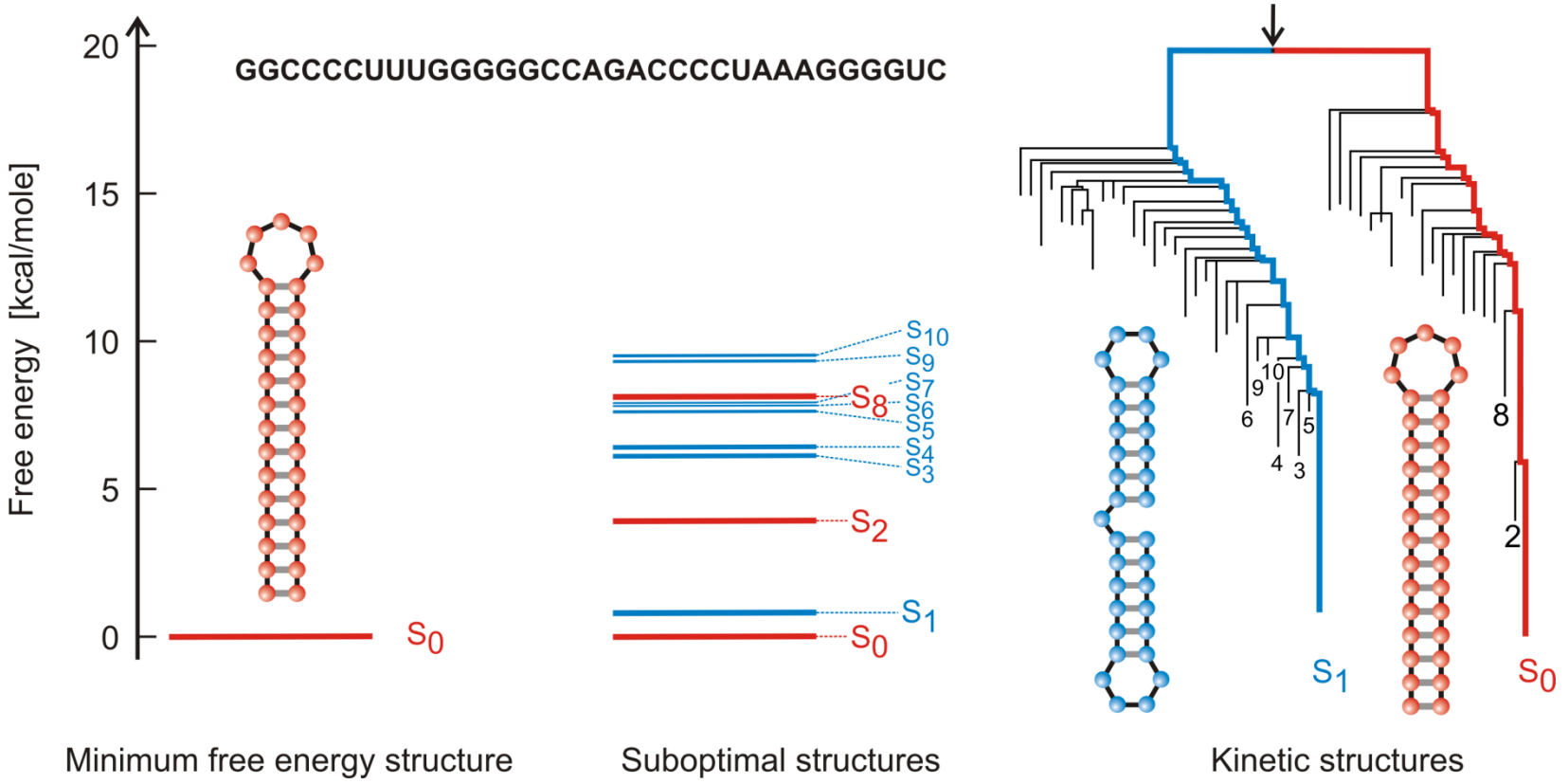

15
10
5
0
- ΔG (kcal/mole)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
position





Extension of the notion of structure


Stable (sub)optimal structures
Sequence space
Structure space



Interconversion of suboptimal structures

3568–3576 Nucleic Acids Research, 2006, Vol. 34, No. 12
doi:10.1093/nar/gkl445

Structural parameters affecting the kinetics of RNA hairpin formation

J. H. A. Nagel, C. Flamm[1], I. L. Hofacker[1], K. Franke[2], M. H. de Smit, P. Schuster[1] and C. W. A. Pleij*

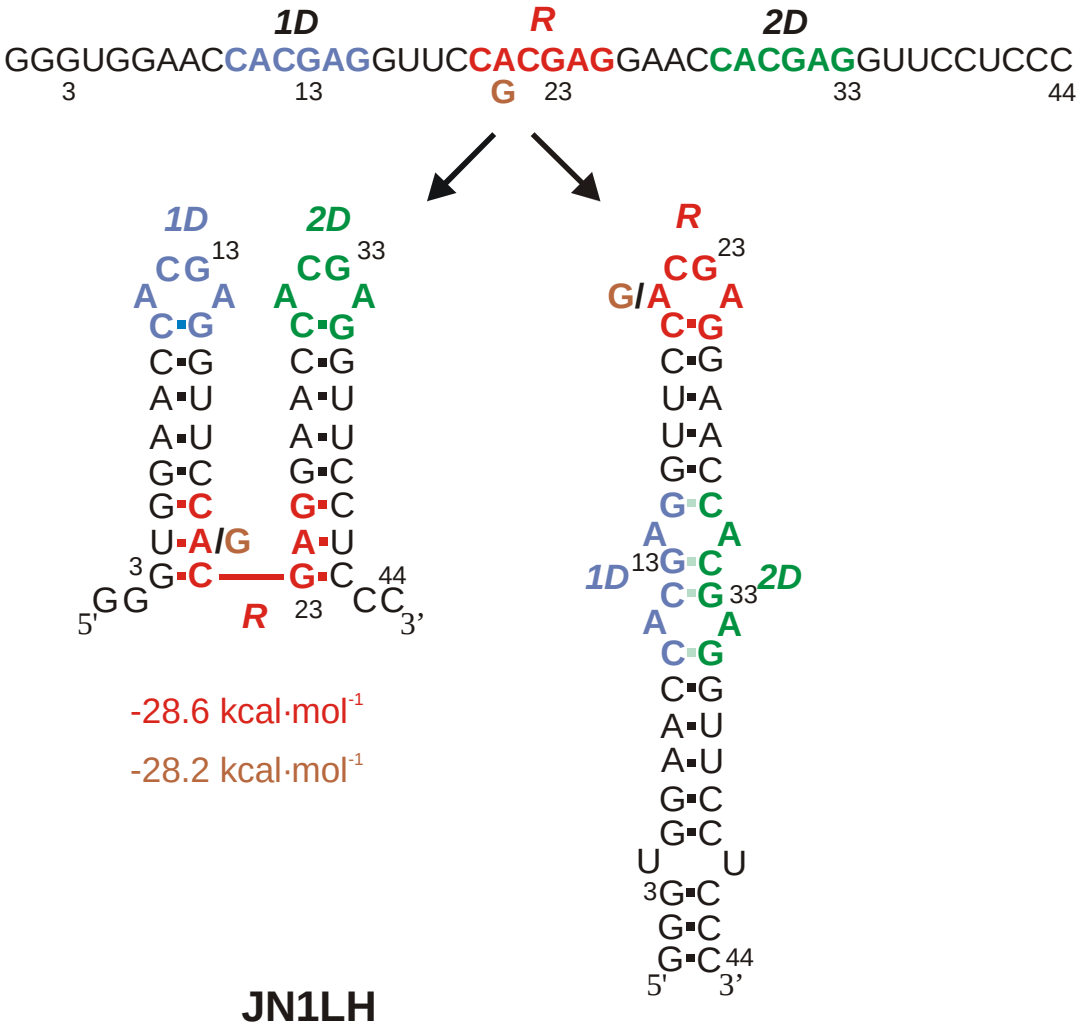
Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2300 RA Leiden, The Netherlands, [1]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, A-1090 Vienna, Austria and [2]IBA NAPS GmbH Rudolf-Wissell-Strasse 28 D-37079 Göttingen, Germany

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

ABSTRACT

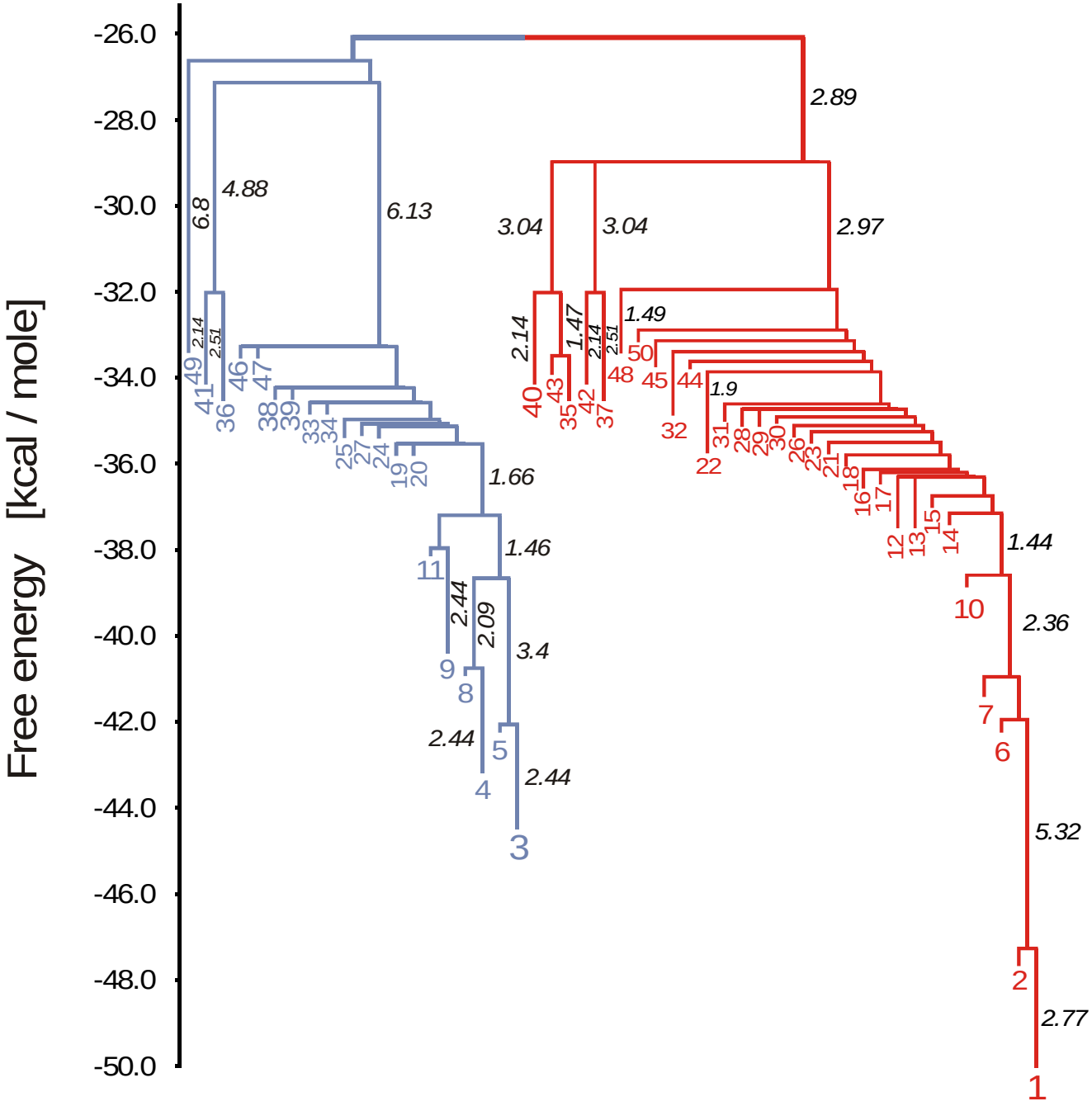
There is little experimental knowledge on the sequence dependent rate of hairpin formation in RNA. We have therefore designed RNA sequences that can fold into either of two mutually exclusive hairpins and have determined the ratio of folding of the two conformations, using structure probing. This folding 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

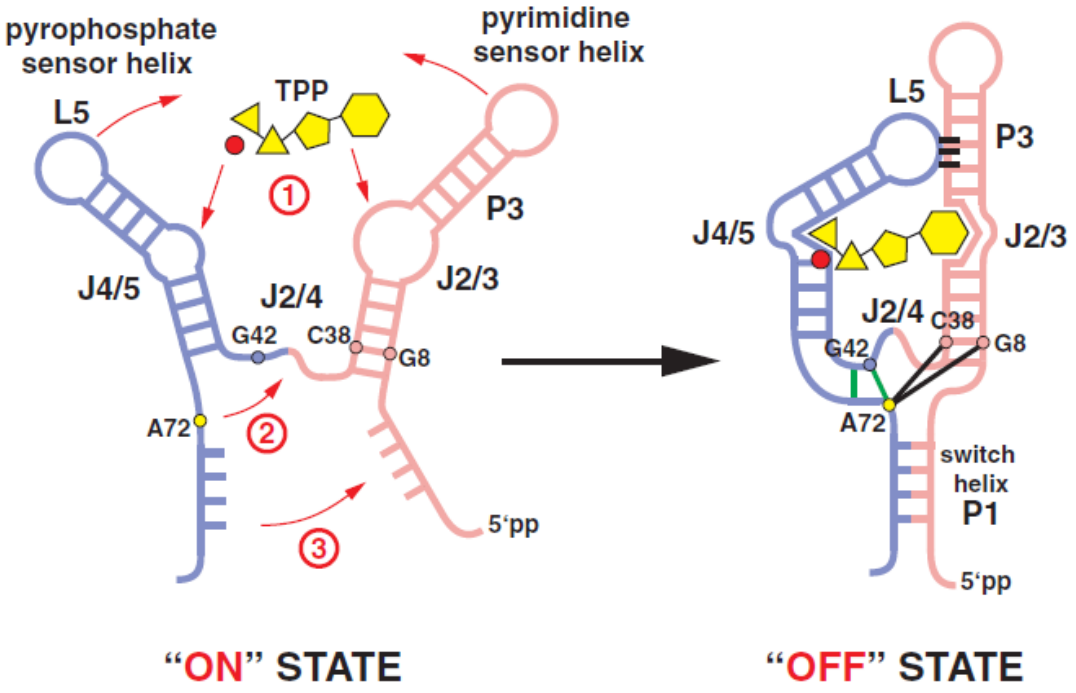


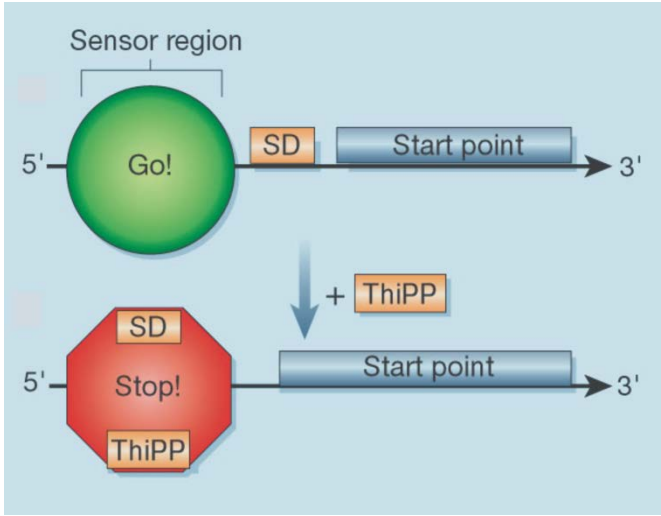
J.H.A. Nagel, C. Flamm, I.L. Hofacker, K. Franke, M.H. de Smit, P. Schuster, and C.W.A. Pleij.

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



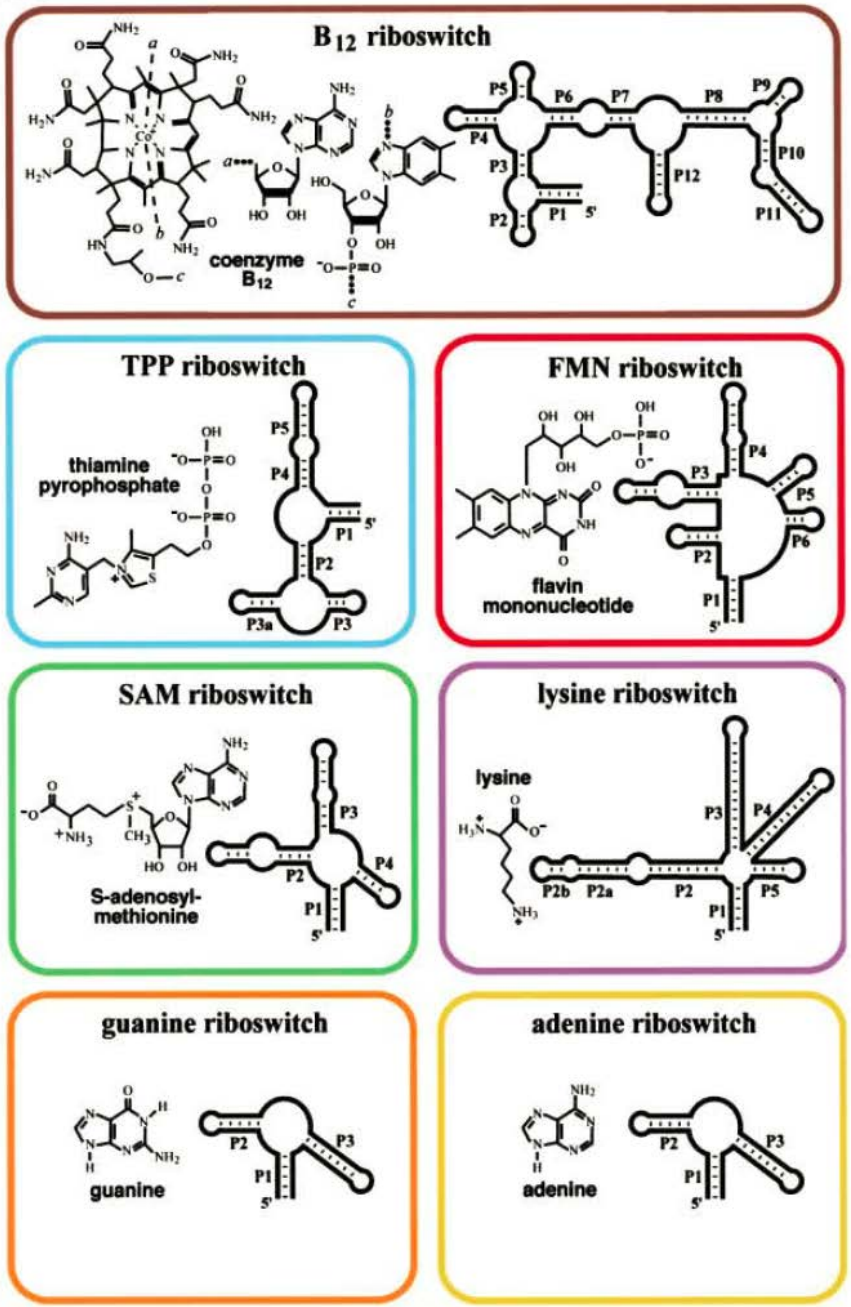
J1LH barrier tree

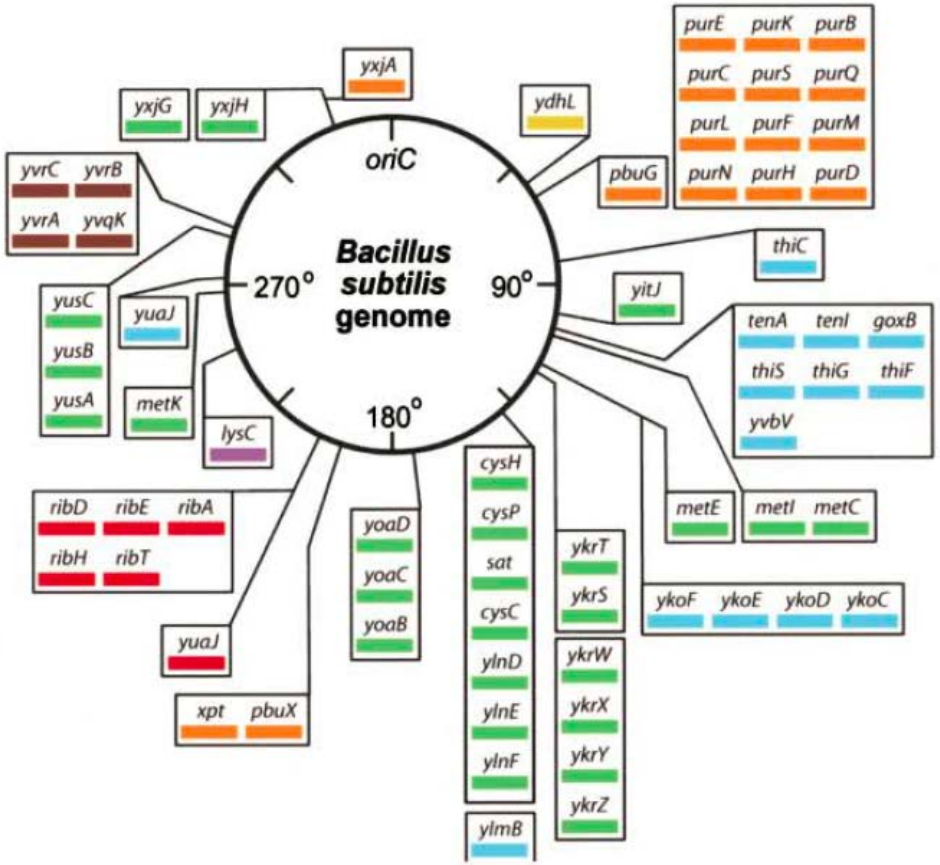




The thiamine-pyrophosphate riboswitch

S. Thore, M. Leibundgut, N. Ban.
Science **312**:1208-1211, 2006.





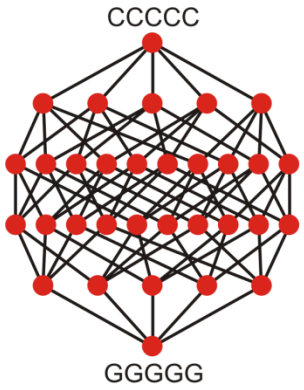
M. Mandal, B. Boese, J.E. Barrick, W.C. Winkler, R.R, Breaker. Cell 113:577-586 (2003)

1. History of molecular evolution and its applications
2. Why RNA is suitable for molecular evolution
3. **Evolutionary dynamics of replication and mutation**
4. Evolution and complexity

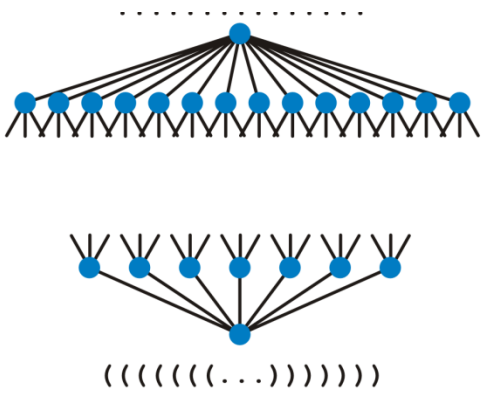
Sewall Wright, 1889 - 1988

Sewall Wright. 1932. *The roles of mutation, inbreeding, crossbreeding and selection in evolution*. In: D.F.Jones, ed. Int. Proceedings of the Sixth International Congress on Genetics. Vol.1, 356-366. Ithaca, NY.

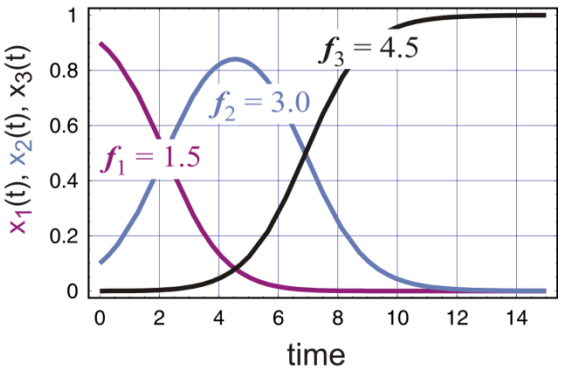
Sewall Wrights fitness landscape as metaphor for Darwinian evolution



sequence space



shape space



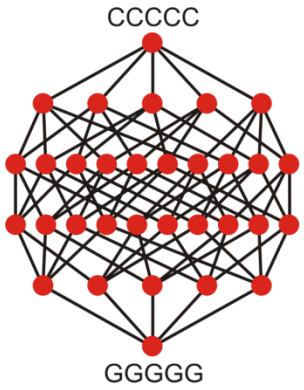
parameter space

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

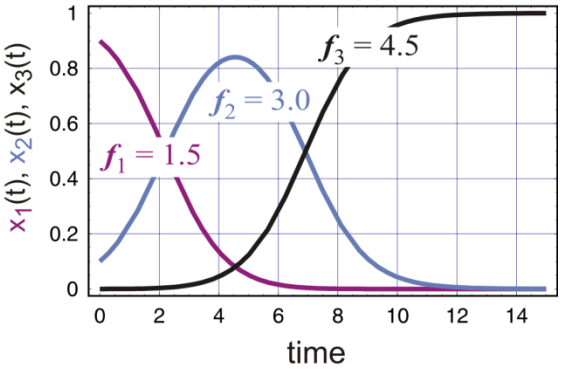
S $\longrightarrow$ Y = Φ(S) $\longrightarrow$ $f = \Psi(Y)$

sequence $\quad$ structure $\quad$ function

The paradigm of structural biology



sequence space



parameter space

S

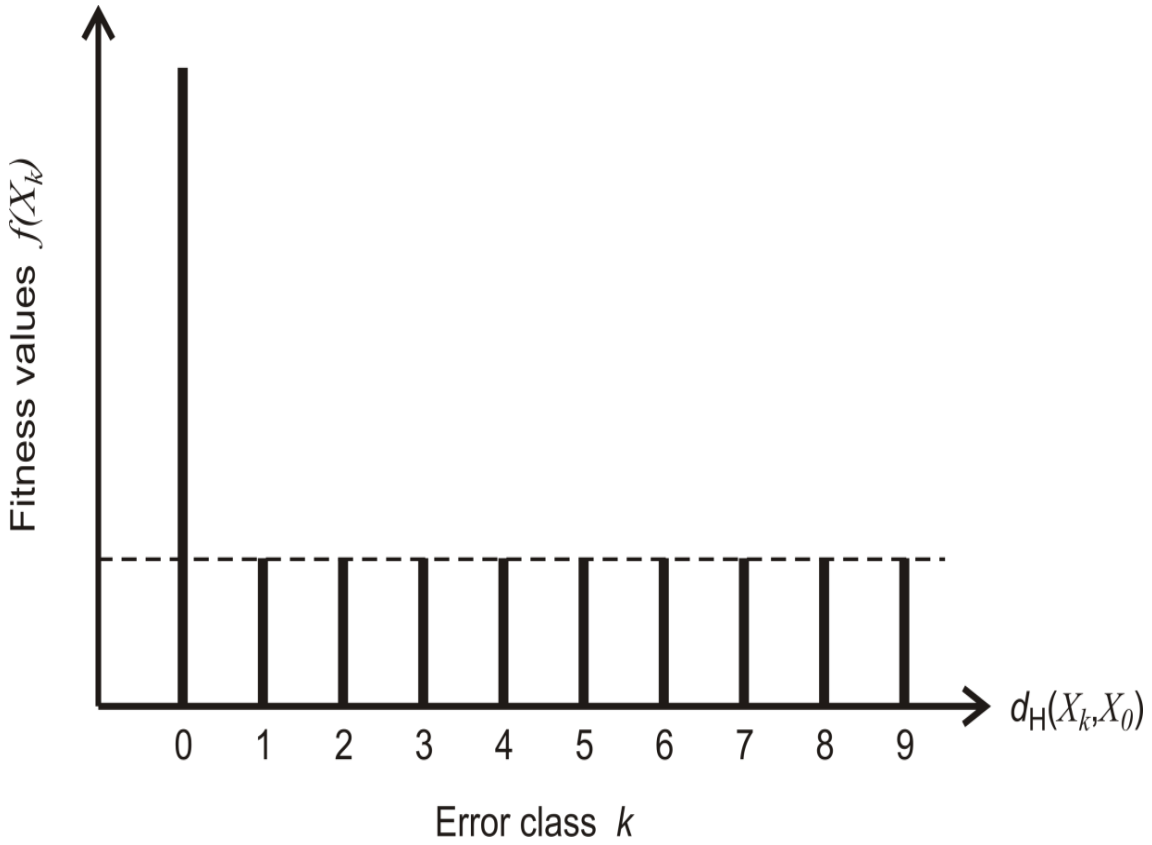
sequence

$f = \Psi(Y)$

function

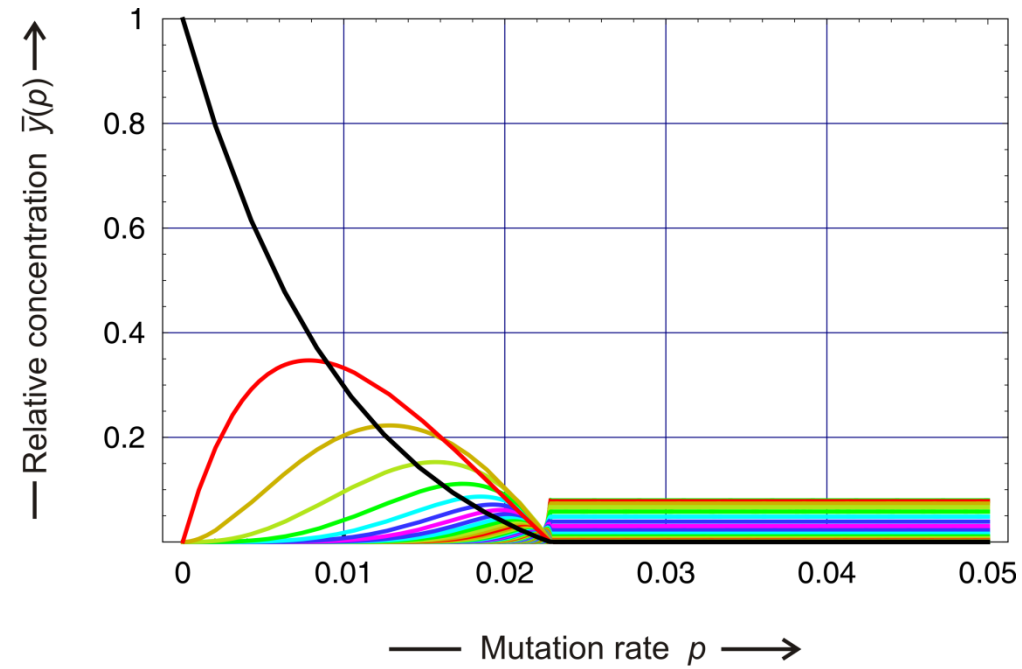
The simplified model

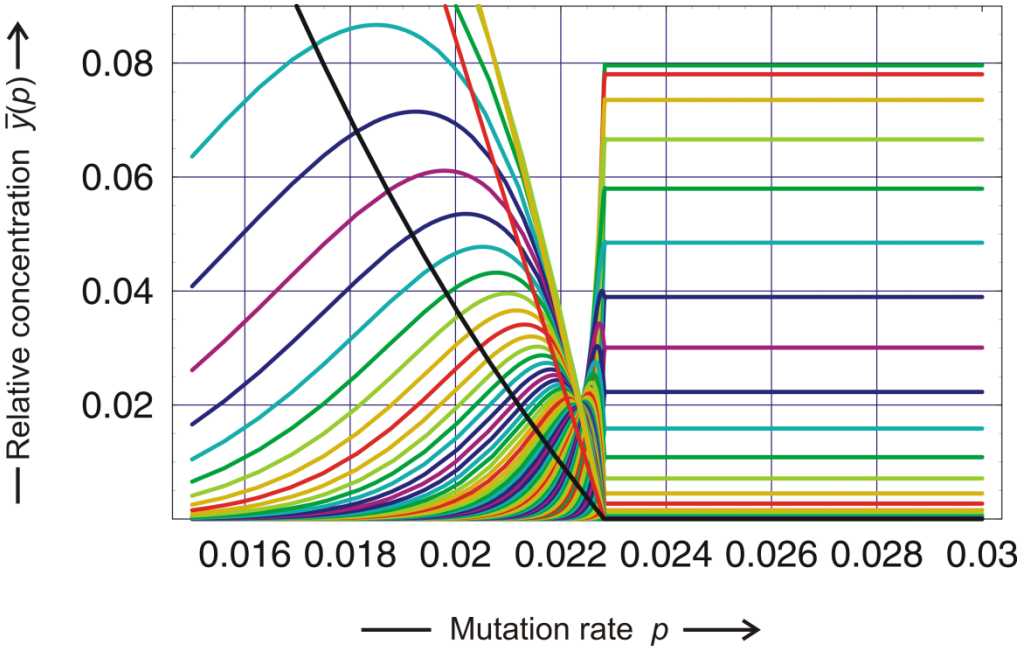
single peak landscape



A model fitness landscape that was accessible to computation in the nineteen eighties

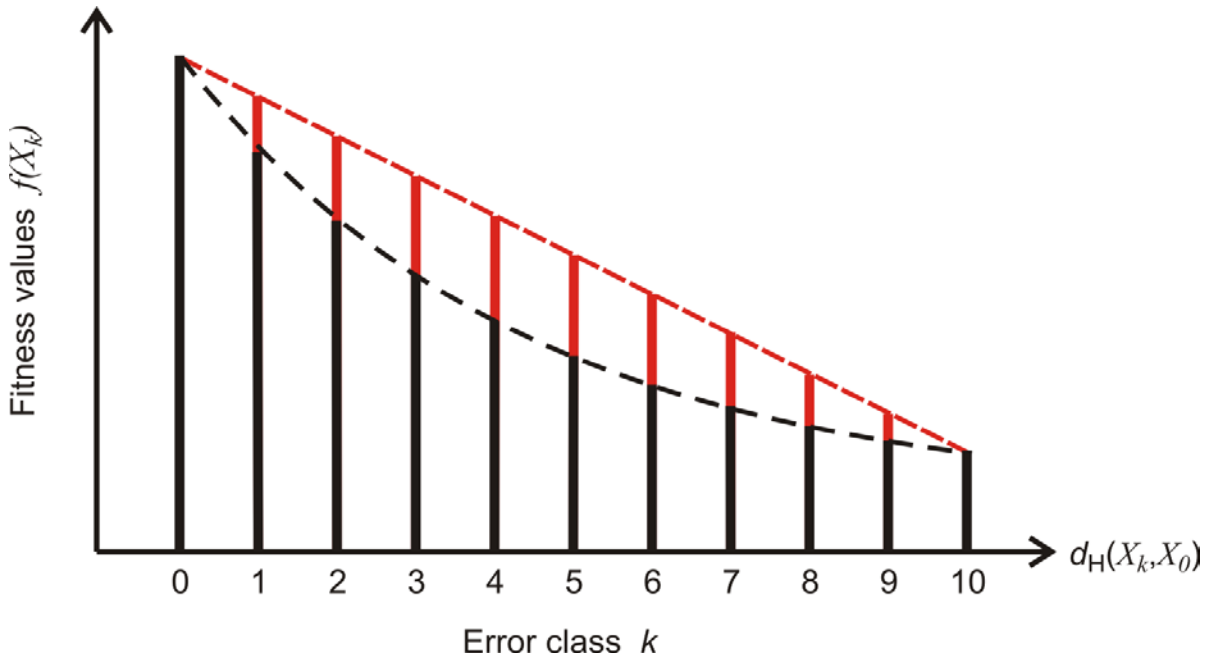
Error threshold on the single peak landscape



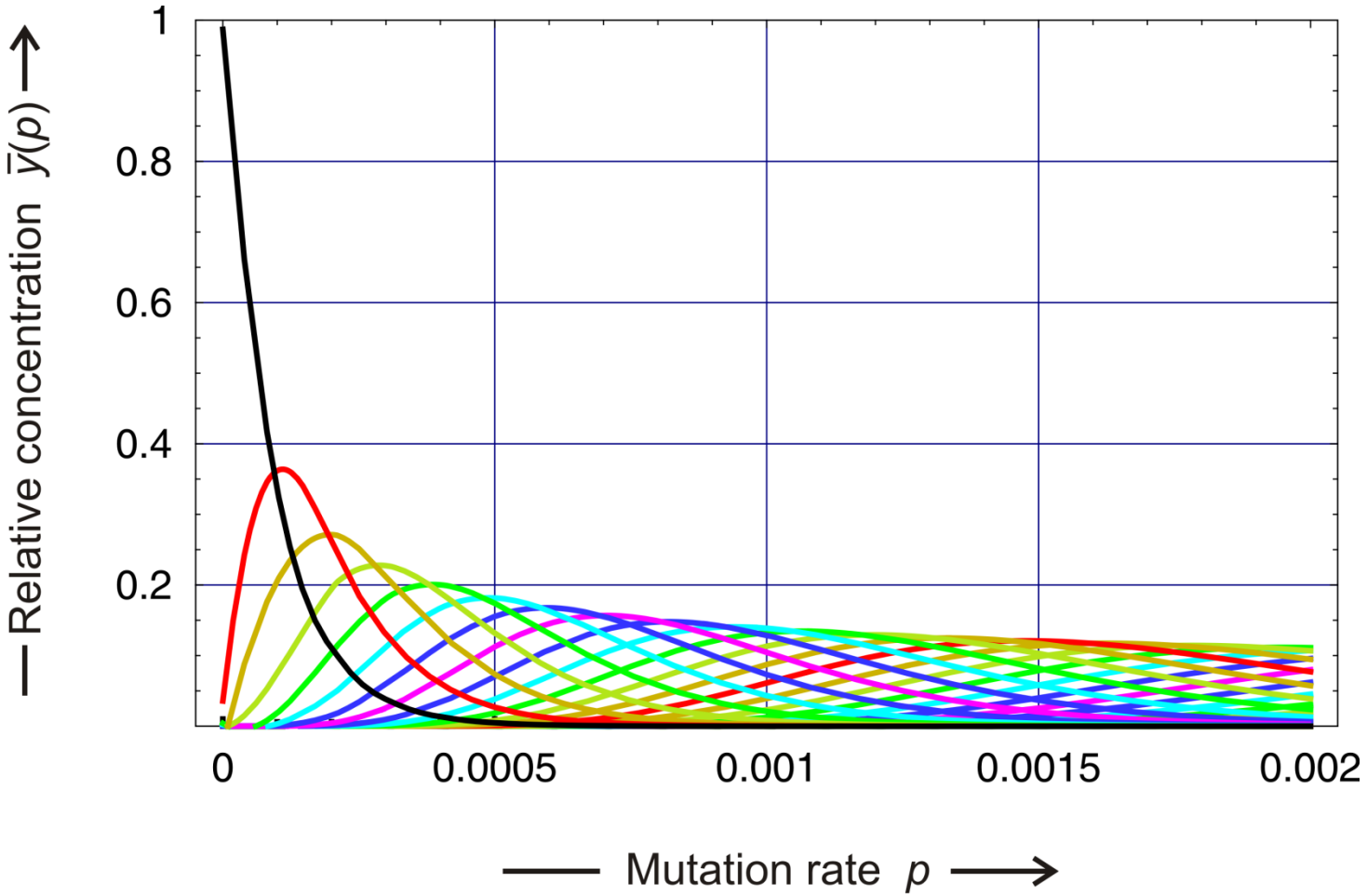


Thomas Wiehe. 1997. Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.* 69:127-136

linear and multiplicative



Model fitness landscapes that do not sustain error thresholds



The linear fitness landscape shows no error threshold

Realistic fitness landscapes

1. Ruggedness: nearby lying genotypes may develop into very different phenotypes

2. Neutrality: many different genotypes give rise to phenotypes with identical selection behavior

3. Combinatorial explosion: the number of possible genomes is prohibitive for systematic searches

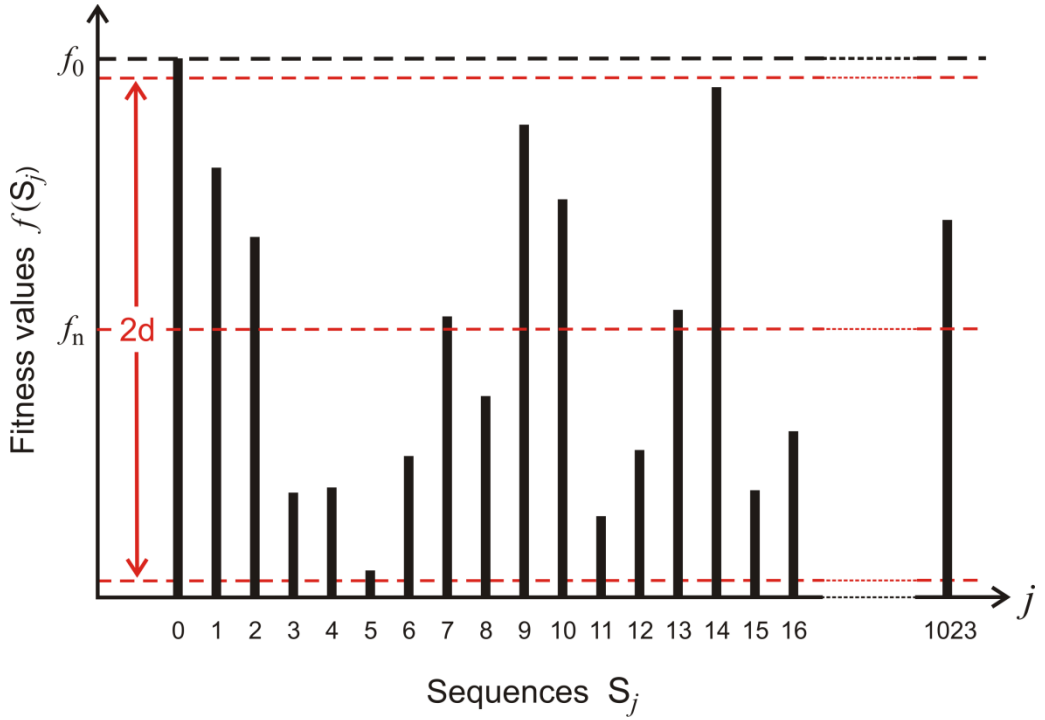
Facit: Any successful and applicable theory of molecular evolution must be able to predict evolutionary dynamics from a small or at least in practice measurable number of fitness values.

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

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

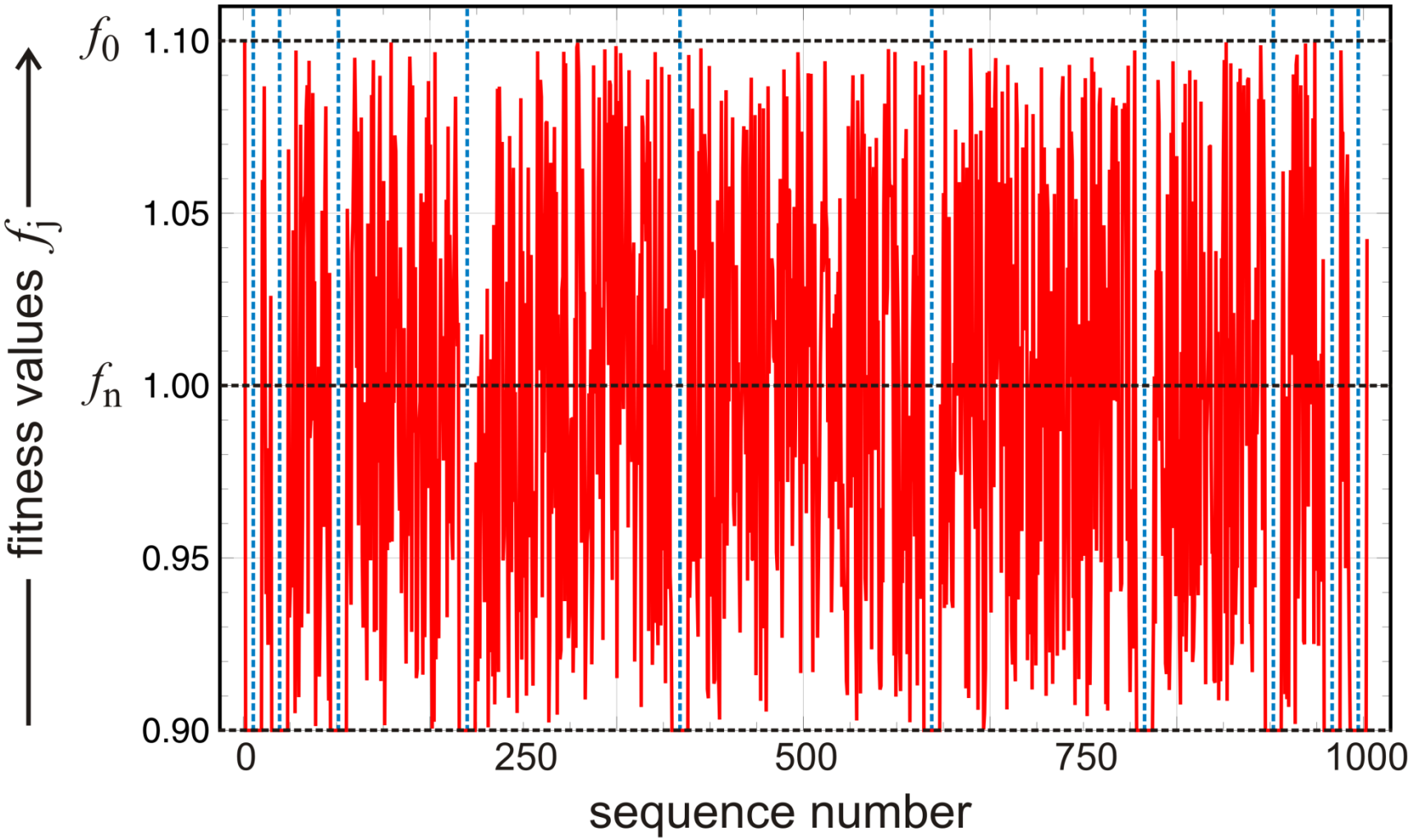
η … random number

s … seeds

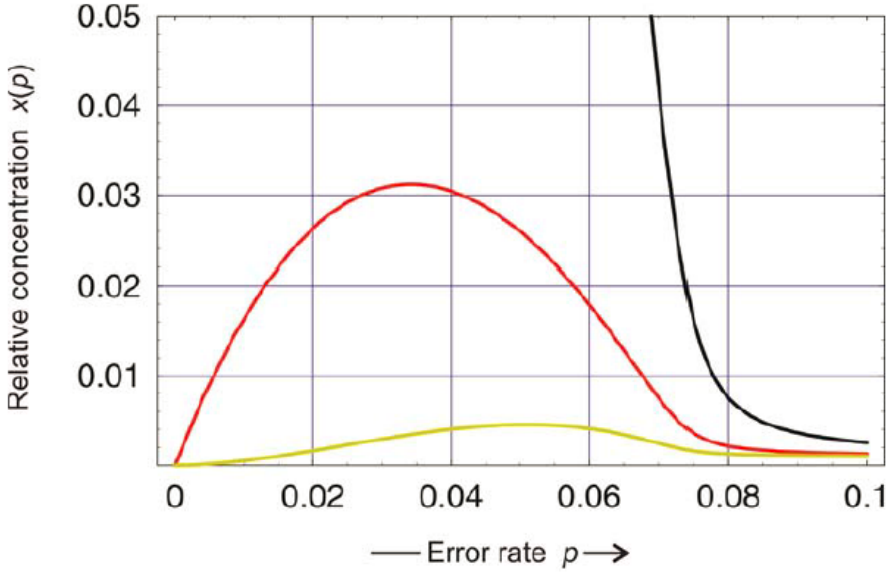


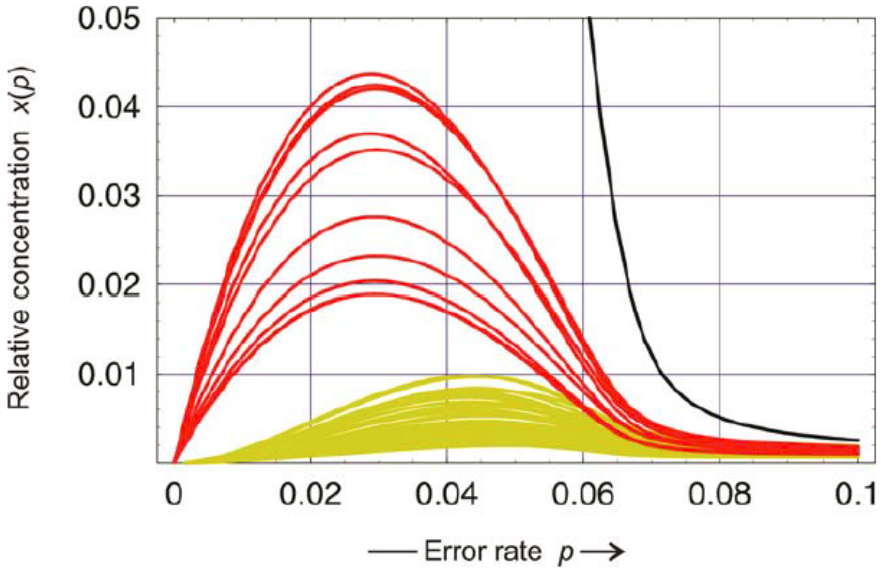
„realistic“ landscape

Rugged fitness landscapes over individual binary sequences with n = 10

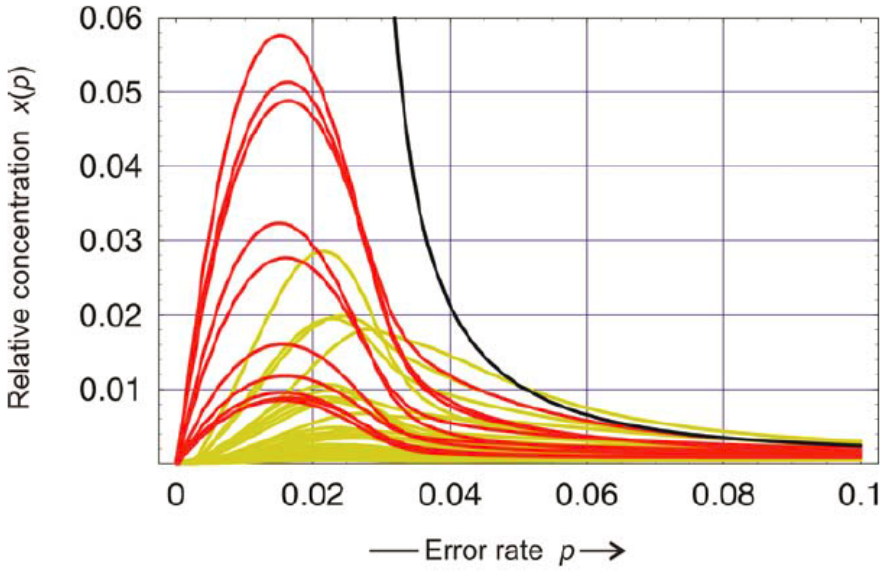


Random distribution of fitness values: d = 1.0 and s = 637



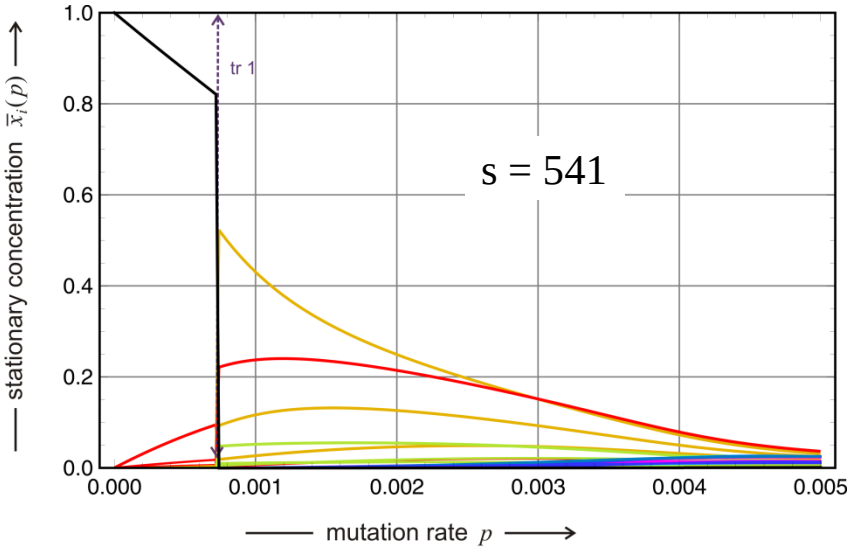


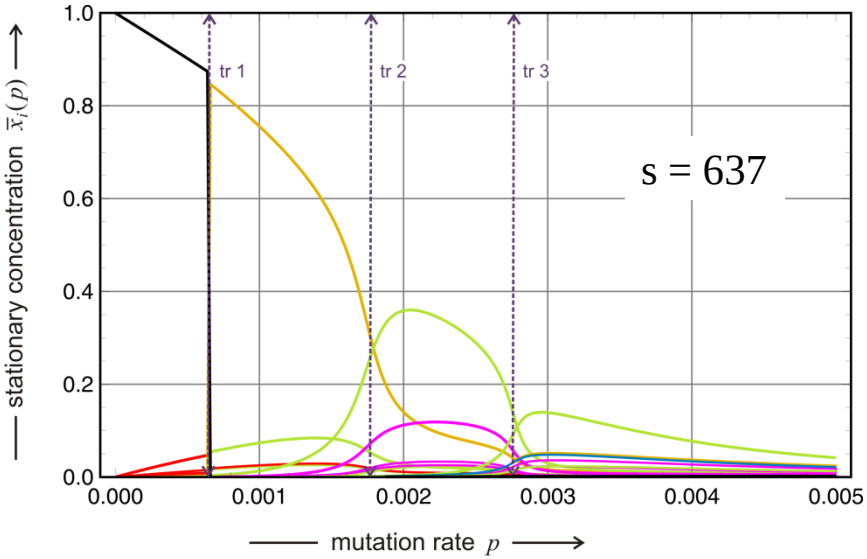
Quasispecies with increasing random scatter d



Error threshold: Individual sequences

n = 10, σ = 2, s = 491 and d = 0, 0.5, 0.9375



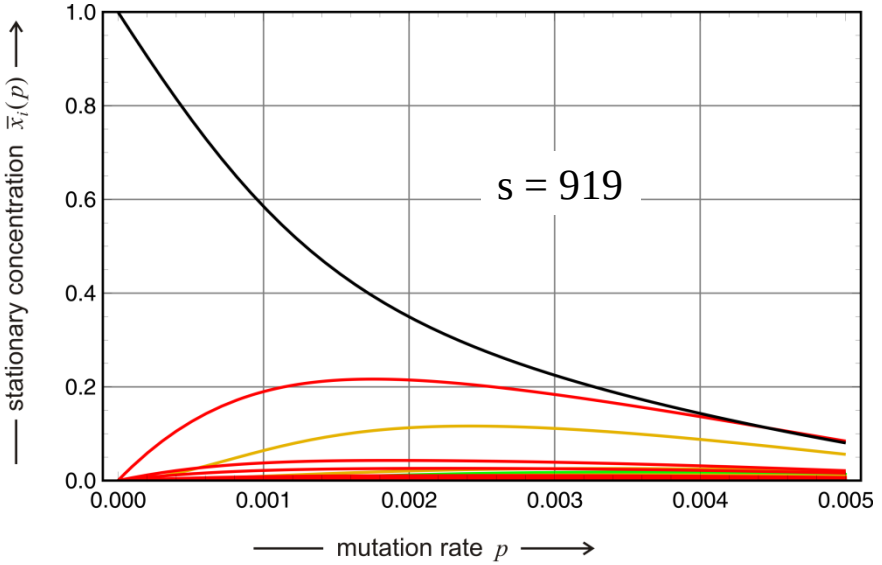


Three different choices of random scatter:

s = 541 , s = 637 , s = 919

Error threshold on ‚realistic' landscapes

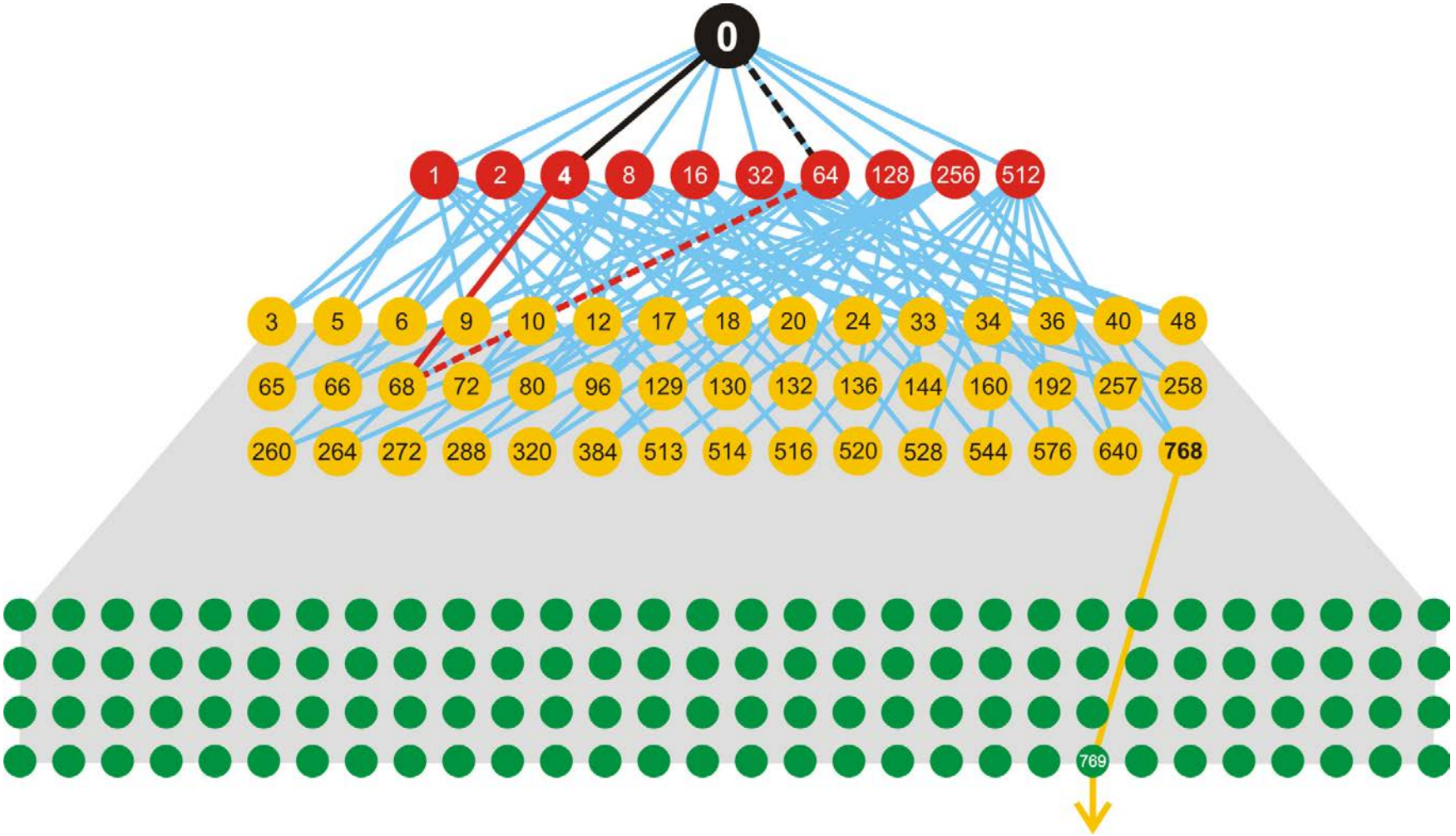
n = 10, $f_0 = 1.1$, $f_n = 1.0$, d = 1.0



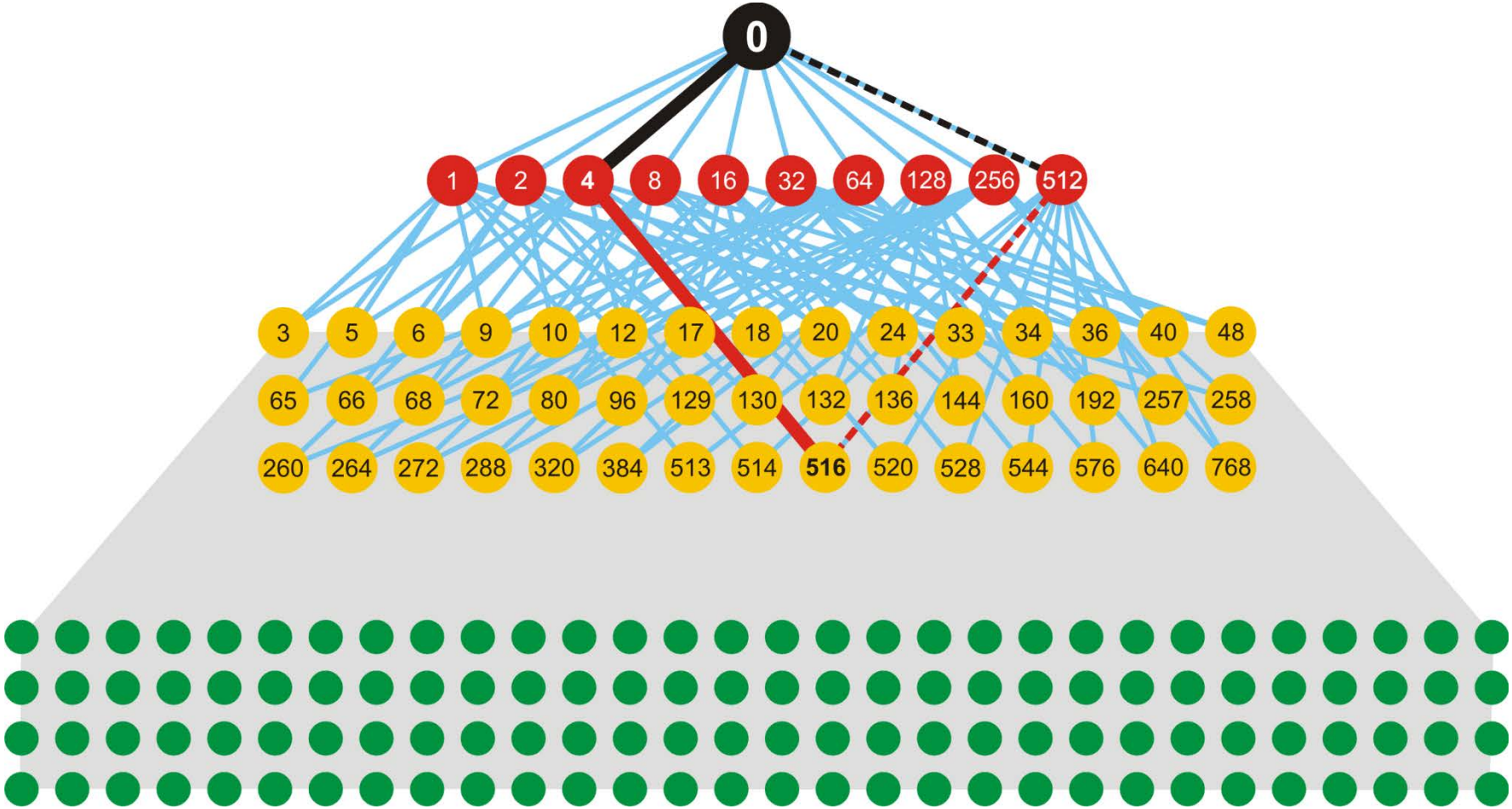
Two questions:

1. Why are quasispecies on some particular fitness landscapes so stable?

2. What happens if the dominant sequences are neutral with respect to selection?



Determination of the dominant mutation flow: $d = 1$, $s = 613$



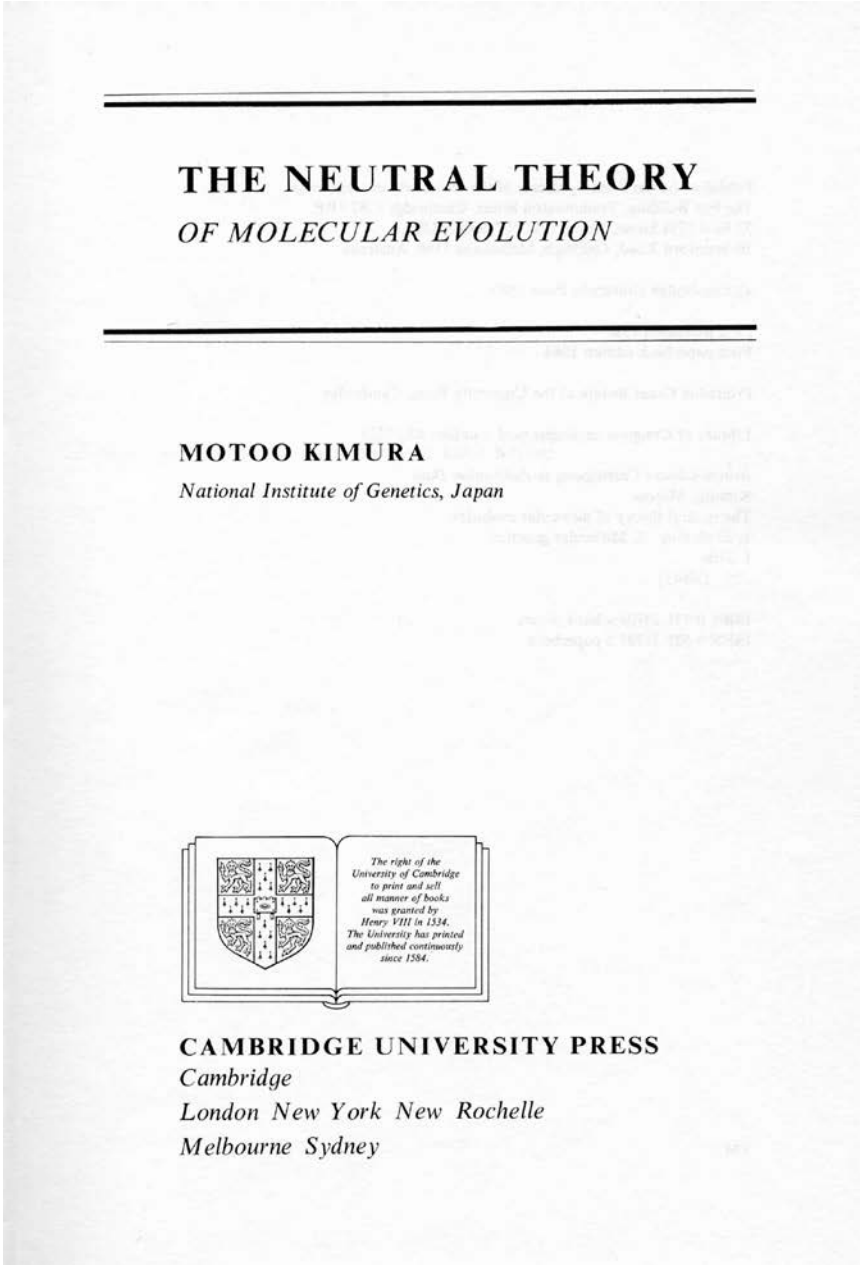
Determination of the dominant mutation flow: $d = 1$, $s = 919$

Motoo Kimura, 1924 - 1994

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

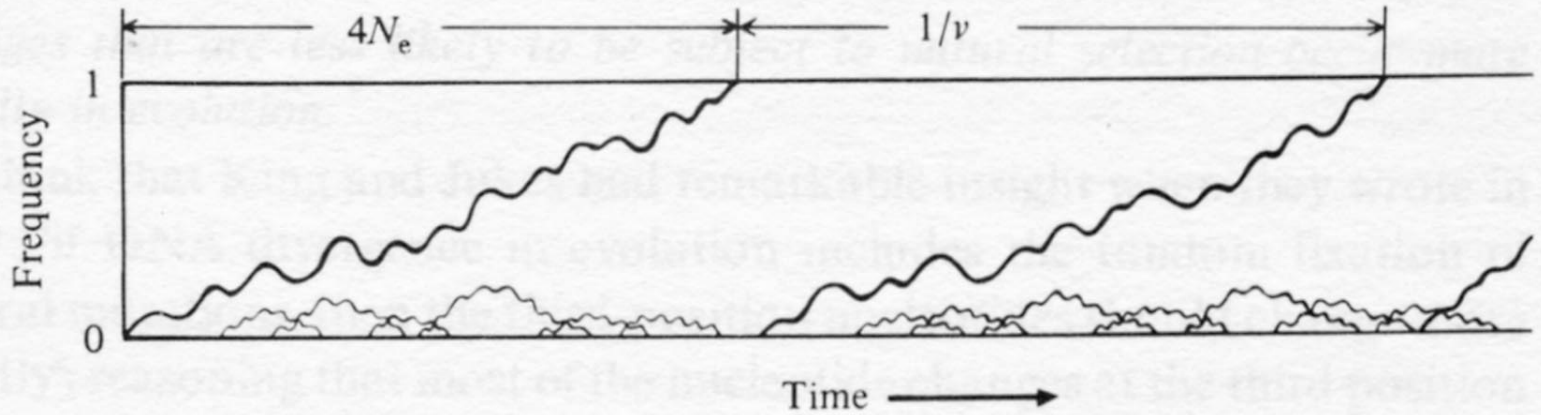
OF MOLECULAR EVOLUTION

MOTOO KIMURA

National Institute of Genetics, Japan

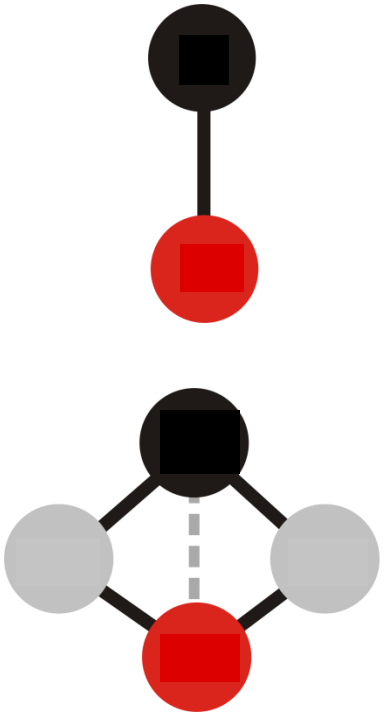
CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

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?



$\mathbf{d_H = 1}$

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

$\mathbf{d_H = 2}$

$$\lim_{p\to 0} x_1(p) = \alpha/(1+\alpha)$$

$$\lim_{p\to 0} x_2(p) = 1/(1+\alpha)$$

$\mathbf{d_H \geq 3}$

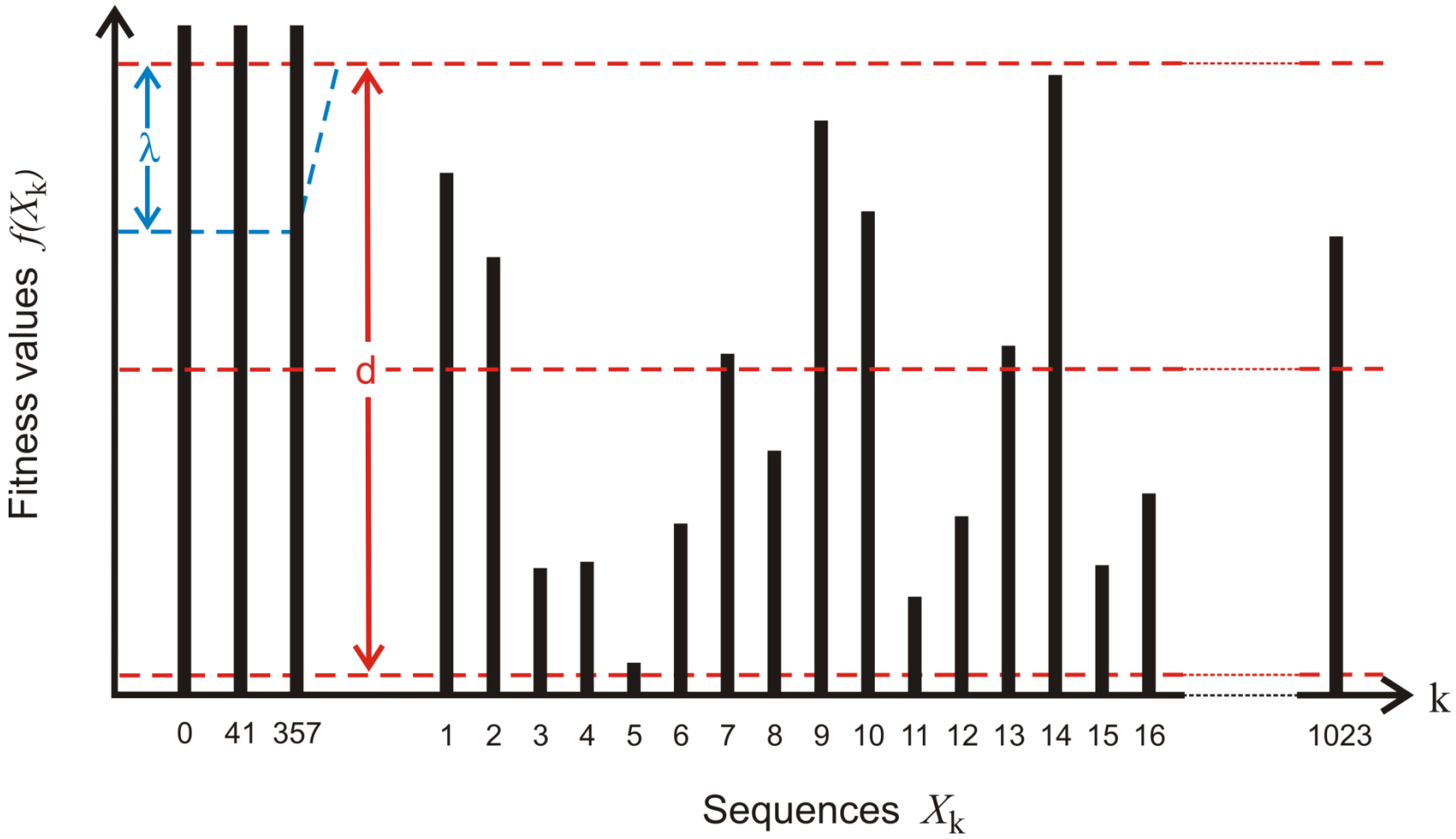
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

Pairs of neutral sequences in replication networks

Random fixation in the sense of Motoo Kimura

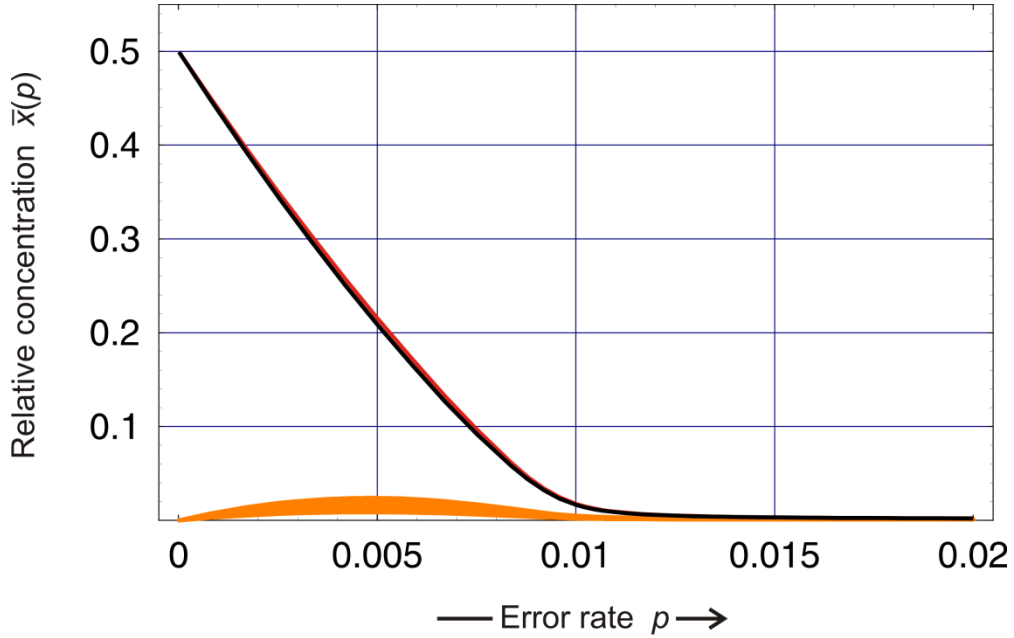
P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650



A fitness landscape including neutrality

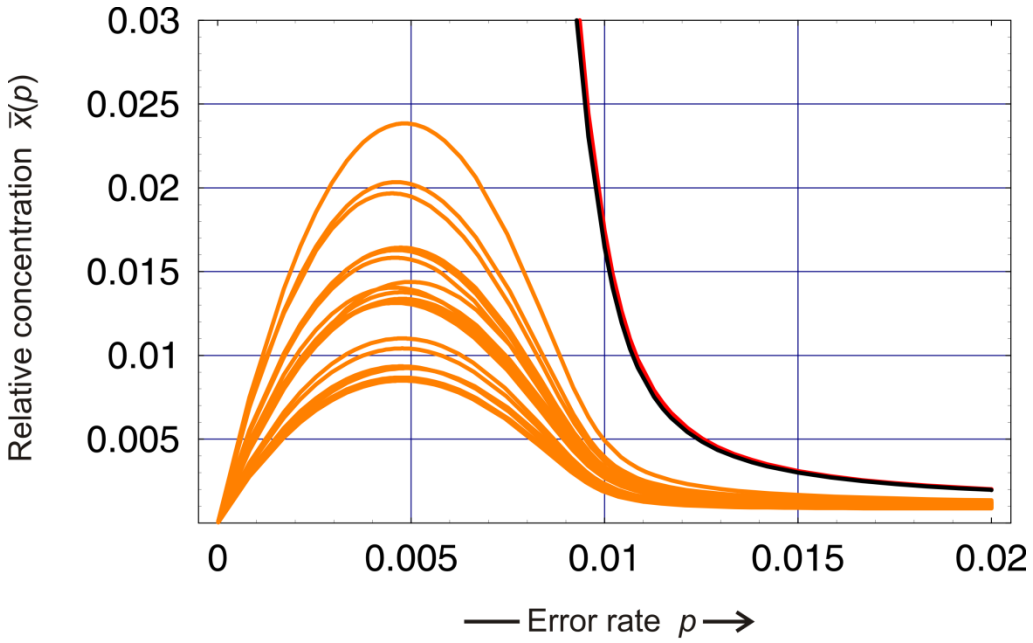
Neutral network

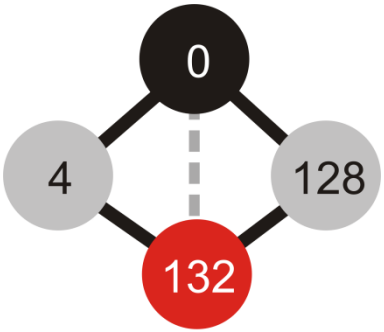
$\lambda = 0.01,\ s = 367$



Neutral network: Individual sequences

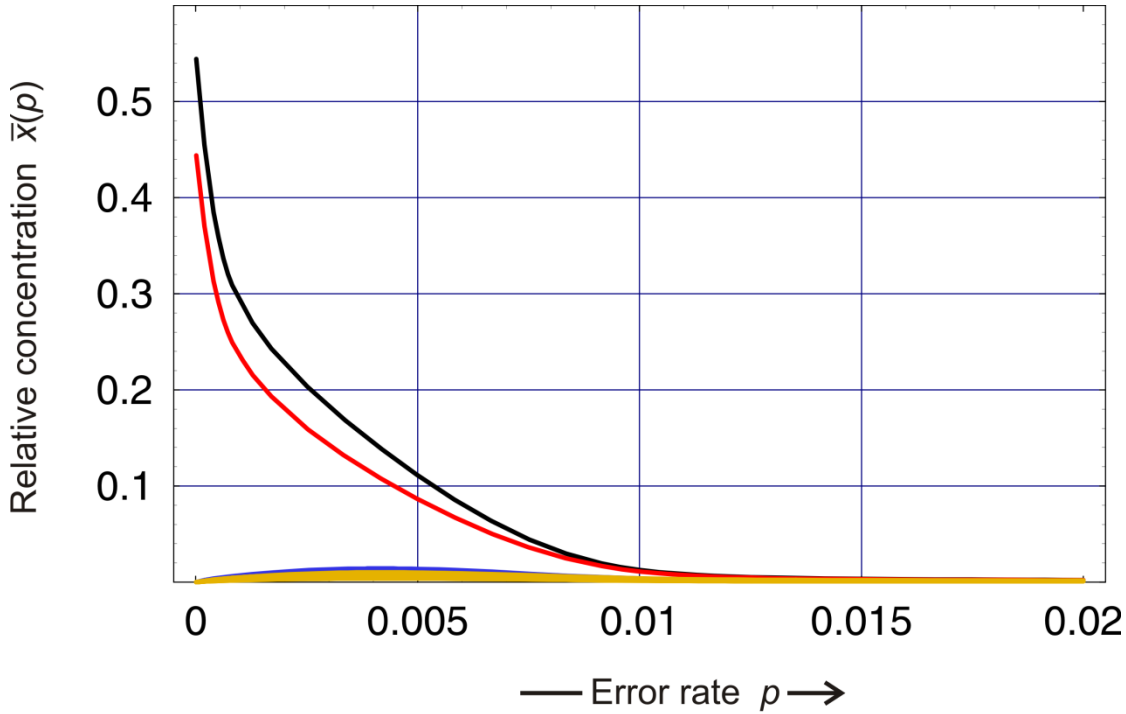
$n = 10,\ \sigma = 1.1,\ d = 1.0$





Neutral network

$\lambda = 0.01,\ s = 877$



Neutral network: Individual sequences

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

Consensus sequences of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i,X_j) = 1$ and 2.

1. History of molecular evolution and its applications
2. Why RNA is suitable for molecular evolution
3. Evolutionary dynamics of replication and mutation
4. **Evolution and complexity**

Fitness landscapes are becoming experimentally accessible!

Protein landscapes: Yuuki Hayashi, Takuyo Aita, Hitoshi Toyota, Yuzuru Husimi, Itaru Urabe, Tetsuya Yomo. 2006. Experimental rugged fitness landscape in protein seqeunce space. *PLoS One* 1:e96.

RNA landscapes: Sven Klussman, Ed. 2005. The aptamer handbook. Wiley-VCh, Weinheim (Bergstraße), DE.
Jason N. Pitt, Adrian Ferré-D'Amaré. 2010. Rapid construction of empirical RNA fitness landscapes. *Science* 330:376-379.

RNA viruses: Esteban Domingo, Colin R. Parrish, John J. Holland, Eds. 2007. Origin and evolution of viruses. Second edition. Elesvier, San Diego, CA.

Retroviruses: Roger D. Kouyos, Gabriel E. Leventhal, Trevor Hinkley, Mojgan Haddad, Jeannette M. Whitcomb, Christos J. Petropoulos, Sebastian Bonhoeffer. 2012. Exploring the complexity of the HIV-I fitness landscape. *PLoS Genetics* 8:e1002551

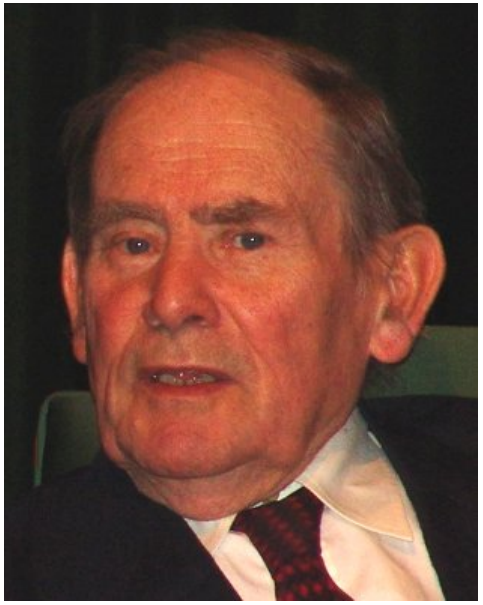
The new biology provides a hitherto unknown challenge for mathematicians, computer scientists, and theorical biologists for mainly two reasons

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

Coworkers

Universität Wien

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

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

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

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

Universität Wien

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

Thank you for your attention!

Web-Page for further information:

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