Error thresholds on „realistic“ landscapes

An old story in a new setting

Peter Schuster

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

Seminar Lecture

Bioinformatik, Universität Leipzig, 11.05.2010

Web-Page for further information:

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

Prologue

The work on a molecular theory of evolution started 42 years ago .....

DIE NATURWISSENSCHAFTEN

58. Jahrgang, 1971 — Heft 10 Oktober

Selforganization of Matter and the Evolution of Biological Macromolecules

MANFRED EIGEN*

MAX-PLANCK-INSTITUT für Biophysikalische Chemie, Karl-Friedrich-Bonhoeffer-Institut, Göttingen-Nikolausberg

I. Introduction

I.1. „Cause and Effect"

The question about the origin of life often appears as a question about "cause and effect". Physical theories of macroscopic processes usually involve answers to such questions, even if a statistical interpretation is given to the relation between "cause" and "effect". It is mainly due to the nature of this question that many scientists believe that our present physics does not offer any obvious explanation for the existence of life, which even in its simplest forms always appears to be associated with complex macroscopic (i.e. multimolecular) systems, such as the living cell.

As a consequence of the exciting discoveries of "molecular biology", a common version of the above question is: *Which came first, the protein or the nucleic acid?* – a modern variant of the old "chicken-and-the-egg" problem. The term "first" is usually meant to define a causal rather than a temporal relationship, and the words "protein" and "nucleic acid" may be substituted by "function" and "information". The question in this form, when applied to the interplay of nucleic acids and proteins as presently encountered in the living cell, leads ad absurdum, because "function"

* Partly presented as the "Robbins Lectures" at Pomona College, California, in spring 1970.

1971

Die Naturwissenschaften 64. Jahrgang Heft 11 November 1977

The Hypercycle

A Principle of Natural Self-Organization

Part A: Emergence of the Hypercycle

Manfred Eigen
Max-Planck-Institut für biophysikalische Chemie, D-3400 Göttingen

Peter Schuster
Institut für theoretische Chemie und Strahlenchemie der Universität, A-1090 Wien

This paper is the first part of a trilogy, which comprises a detailed study of a special type of functional organization and demonstrates its relevance with respect to the origin and evolution of life. Self-replicative macromolecules, such as RNA or DNA in a suitable environment exhibit a behaviour, which we may call Darwinian and which can be formally represented by the concept of the quasi-species. A quasi-species is defined as a given distribution of macromolecular species with closely interrelated sequences, dominated by one or several (degenerate) master copies. External constraints enforce the selection of the best adapted distribution, commonly referred to as the wild-type. Most important for Darwinian behavior are the criteria for internal stability of the quasi-species. If these criteria are violated, the information stored in the nucleotide sequence of the master copy will deteriorate irreversibly leading to an error catastrophy. As a consequence, selection and evolution of RNA or DNA molecules is limited with respect to the amount of information that can be stored in a single replicative unit. An analysis of experimental data regarding RNA and DNA replication at various levels of organization reveals, that a sufficient amount of information for the build up of a translation machinery can be gained only via integration of several different replicative units (or reproductive cycles) through functional linkages. A stable functional integration then will raise the system to a new level of organization and thereby enlarge its information capacity considerably. The hypercycle appears to be such a form of organization.

Preview on Part B: The Abstract Hypercycle

The mathematical analysis of dynamical systems using methods of differential topology, yields the result that there is only one type of mechanisms which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective master copies must compete favorably with their error distributions. Despite their competitive behavior these units must establish a cooperation which includes all functionally integrated species. On the other hand, the cycle as a whole must continue to compete strongly with any other single entity or linked ensemble which does not contribute to its integrated function.
These requirements are crucial for a selection of the best adapted functionally linked ensemble and its evolutive optimization. Only hypercyclic organizations are able to fulfil these requirements. Non-cyclic linkages among the autonomous reproduction cycles, such as chains or branched, tree-like networks are devoid of such properties.
The mathematical methods used for proving these assertions are fixed-point, Lyapunov- and trajectorial analysis in higher-dimensional phase spaces, spanned by the concentration coordinates of the cooperating partners. The self-organizing properties of hypercycles are elucidated, using analytical as well as numerical techniques.

Preview on Part C: The Realistic Hypercycle

A realistic model of a hypercycle relevant with respect to the origin of the genetic code and the translation machinery is presented. It includes the following features referring to natural systems:
1) The hypercycle has a sufficiently simple structure to admit an origination with finite probability under prebiotic conditions.
2) It permits a continuous emergence from closely intercalated (t-RNA-like) precursors, originally being members of a stable RNA quasi-species and having been amplified to a level of higher abundance.
3) The organizational structure and the properties of single functional units of this hypercycle are still reflected in the present genetic code in the translation apparatus of the prokaryotic cell, as well as in certain bacterial viruses.

I. The Paradigm of Unity and Diversity in Evolution

Why do millions of species, plants and animals, exist, while there is only one basic molecular machinery of the cell: one universal genetic code and unique chiralities of the macromolecules?
The geneticists of our day would not hesitate to give an immediate answere to the first part of this question. Diversity of species is the outcome of the tremendous branching process of evolution with its myriads of single steps of reproduction and mutation. It in-

1977

Molecular Quasi-Species†

Manfred Eigen,* John McCaskill,
Max Planck Institut für biophysikalische Chemie, Am Fassberg, D 3400 Göttingen-Nikolausberg, BRD
and Peter Schuster*
Institut für theoretische Chemie und Strahlenchemie, der Universität Wien, Währinger Strasse 17, A-1090 Wien, Austria (Received: June 9, 1988)

The molecular quasi-species model describes the physicochemical organization of monomers into an ensemble of heteropolymers with combinatorial complexity by ongoing template polymerization. Polynucleotides belong to the simplest class of such molecules. The quasi-species itself represents the stationary distribution of macromolecular sequences maintained by chemical reactions effecting error-prone replication and by transport processes. It is obtained deterministically, by mass-action kinetics, as the dominant eigenvalue of a *value* matrix, **W**, which is derived directly from chemical rate coefficients, but it also exhibits stochastic features, being composed to a significant fraction of unique individual macromolecular sequences. The quasi-species model demonstrates how macromolecular information originates through specific nonequilibrium autocatalytic reactions and thus forms a bridge between reaction kinetics and molecular evolution. Selection and evolutionary optimization appear as new features in physical chemistry. Concentration bias in the production of mutants is a new concept in population genetics, relevant to frequently mutating populations, which is shown to greatly enhance the optimization properties. The present theory relates to asexually replicating ensembles, but this restriction is not essential. A sharp transition is exhibited between a drifting population of essentially random macromolecular sequences and a localized population of close relatives. This transition at a threshold error rate was found to depend on sequence lengths, distributions of selective values, and population sizes. It has been determined generically for complex landscapes and for special cases, and, it was shown to persist generically in the presence of nearly neutral mutants. Replication dynamics has much in common with the equilibrium statistics of complex spin systems: the error threshold is equivalent to a magnetic order–disorder transition. A rational function of the replication accuracy plays the role of temperature. Experimental data obtained from test-tube evolution of polynucleotides and from studies of natural virus populations support the quasi-species model. The error threshold seems to set a limit to the genome lengths of several classes of RNA viruses. In addition, the results are relevant even in eucaryotes where they contribute to the exon–intron debate.

1. Molecular Selection

Our knowledge of physical and chemical systems is, in a final analysis, based on models derived from repeatable experiments. While none of the classic and rather besieged list of properties rounded up to support the intuition of a distinction between the living and nonliving—metabolism, self-reproduction, irritability, and adaptability, for example—intrinsically limit the application of the scientific method, a determining role by unique or *individual* entities comes into conflict with the requirement of repeatability. Combinatorial variety, such as that in heteropolymers based on even very small numbers of different bases, even just two, readily provides numbers of different entities so enormous that neither consecutive nor parallel physical realization is possible. The physical chemistry of finite systems of such macromolecules must deal with both known regularities and the advent of unique copolymeric sequences. Normally this would present no difficulty in a statistical mechanical analysis of typical behavior, where rare events play no significant role, but with autocatalytic polymerization processes even unique single molecules may be amplified to determine the fate of the entire system. Potentially creative, self-organizing around unique events, the dynamics of this simplest living chemical system is invested with regularities that both allow and limit efficient adaptation. The quasi-species model is a study of these regularities.

The fundamental regularity in living organisms that has invited explanation is adaptation. Why are organisms so well fitted to their environments? At a more chemical level, why are enzymes optimal catalysts? Darwin's theory of natural selection has provided biologists with a framework for the answer to this question. The present model is constructed along Darwinian lines but in terms of specific macromolecules, chemical reactions, and physical processes that make the notion of survival of the fittest precise. Not only does the model give an understanding of the physical limitations of adaptation, but also it provides new insight into the role of chance in the process. For an understanding of the structure of this minimal chemical model it is first necessary to recall the conceptual basis of Darwin's theory.

Darwin recognized that new inheritable adaptive properties were not induced by the environment but arose independently in the production of offspring. Lasting adaptive changes in a population could only come about by natural selection of the heritable traits or *genotype* based on the full characteristics or *phenotype* relevant for producing offspring. A process of chance, i.e., uncorrelated with the developed phenotype, controls changes in the genotype from one generation to the next and generates the diversity necessary for selection. Three factors have probably prevented chemists from gaining a clear insight into these phenomena in the past, despite the discovery of the polymeric nature of the genotype (DNA): the complexity of a minimum replication phenotype, the problem of dealing with a huge number of variants, and the nonequilibrium nature of these ongoing processes.

The formulation of a tractable chemical model based on Darwin's principle may be understood in several steps:

1. The major constituents of the system have to be inherently self-reproductive. Only two classes of molecules are presently

† This is an abridged account of the quasi-species theory that has been submitted in comprehensive form to *Advances in Chemical Physics*.[1]

(1) Eigen, M.; McCaskill, J. S.; Schuster, P. *Adv. Chem. Phys.*, in press.

0022-3654/88/2092-6881$01.50/0

1988

Chemical kinetics of molecular evolution

Error Thresholds for Quasispecies on Dynamic Fitness Landscapes

Martin Nilsson
Institute of Theoretical Physics, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

Nigel Snoad
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
and The Australian National University, ACT 0200, Australia†
(Received 29 March 1999)

In this paper we investigate error thresholds on dynamic fitness landscapes. We show that there exists both a lower and an upper threshold, representing limits to the copying fidelity of simple replicators. The lower bound can be expressed as a correction term to the error threshold present on a static landscape. The upper error threshold is a new limit that only exists on dynamic fitness landscapes. We also show that for long genomes and/or highly dynamic fitness landscapes there exists a lower bound on the selection pressure required for the effective selection of genomes with superior fitness independent of mutation rates, i.e. there are distinct nontrivial limits to evolutionary parameters in dynamic environments.

PACS numbers: 87.23.Kg, 87.10.+e, 87.15.Aa

Maternal Effects in Molecular Evolution

Claus O. Wilke*
Digital Life Laboratory, Mail Code 136-93, Pasadena, California 91125
(Received 27 June 2001; published 31 January 2002)

We introduce a model of molecular evolution in which the fitness of an individual depends both on its own and on the parent's genotype. The model can be solved by means of a nonlinear mapping onto the standard quasispecies model. The dependency on the parental genotypes cancels from the mean fitness, but not from the individual sequence concentrations. For finite populations, the position of the error threshold is very sensitive to the influence from parent genotypes. In addition to biological applications, our model is important for understanding the dynamics of self-replicating computer programs.

DOI: 10.1103/PhysRevLett.88.078101 PACS numbers: 87.23.Kg

Quasispecies theory for multiple-peak fitness landscapes

David B. Saakian,[1,2] E. Muñoz,[3] Chin-Kun Hu,[1] and M. W. Deem[3]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
[3]*Department of Physics and Astronomy, Rice University, Houston, Texas 77005-1892, USA*
(Received 15 September 2005; revised manuscript received 13 December 2005; published 11 April 2006)

We use a path integral representation to solve the Eigen and Crow-Kimura molecular evolution models for the case of multiple fitness peaks with arbitrary fitness and degradation functions. In the general case, we find that the solution to these molecular evolution models can be written as the optimum of a fitness function, with constraints enforced by Lagrange multipliers and with a term accounting for the entropy of the spreading population in sequence space. The results for the Eigen model are applied to consider virus or cancer proliferation under the control of drugs or the immune system.

DOI: 10.1103/PhysRevE.73.041913 PACS number(s): 87.23.Kg, 02.50.−r, 87.10.+e, 87.15.Aa

Phase Diagrams of Quasispecies Theory with Recombination and Horizontal Gene Transfer

J.-M. Park[1,2] and M. W. Deem[1]
[1]*Department of Physics & Astronomy and Department of Bioengineering, Rice University, Houston, Texas 77005-1892, USA*
[2]*Department of Physics, The Catholic University of Korea, Bucheon, 420-743, Korea*
(Received 9 October 2006; published 29 January 2007)

We consider how transfer of genetic information between individuals influences the phase diagram and mean fitness of both the Eigen and the parallel, or Crow-Kimura, models of evolution. In the absence of genetic transfer, these physical models of evolution consider the replication and point mutation of the genomes of independent individuals in a large population. A phase transition occurs, such that below a critical mutation rate an identifiable quasispecies forms. We show how transfer of genetic information changes the phase diagram and mean fitness and introduces metastability in quasispecies theory, via an analytic field theoretic mapping.

DOI: 10.1103/PhysRevLett.98.058101 PACS numbers: 87.23.Kg, 87.15.Aa

PHYSICAL REVIEW E **75**, 061109 (2007)

Emergence of order in selection-mutation dynamics

Christoph Marx, Harald A. Posch,* and Walter Thirring†
Faculty of Physics, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria
(Received 7 March 2007; published 8 June 2007)

We characterize the time evolution of a d-dimensional probability distribution by the value of its final entropy. If it is near the maximally possible value we call the evolution mixing, if it is near zero we say it is purifying. The evolution is determined by the simplest nonlinear equation and contains a $d \times d$ matrix as input. Since we are not interested in a particular evolution but in the general features of evolutions of this type, we take the matrix elements as uniformly distributed random numbers between zero and some specified upper bound. Computer simulations show how the final entropies are distributed over this field of random numbers. The result is that the distribution crowds at the maximum entropy, if the upper bound is unity. If we restrict the dynamical matrices to certain regions in matrix space, to diagonal or triangular matrices, for instance, then the entropy distribution is maximal near zero, and the dynamics typically becomes purifying.

DOI: 10.1103/PhysRevE.75.061109 PACS number(s): 05.20.−y, 87.23.Kg, 05.45.Pq, 87.10.+e

PHYSICAL REVIEW E **76**, 041133 (2007)

Emergence of order in quantum extensions of the classical quasispecies evolution

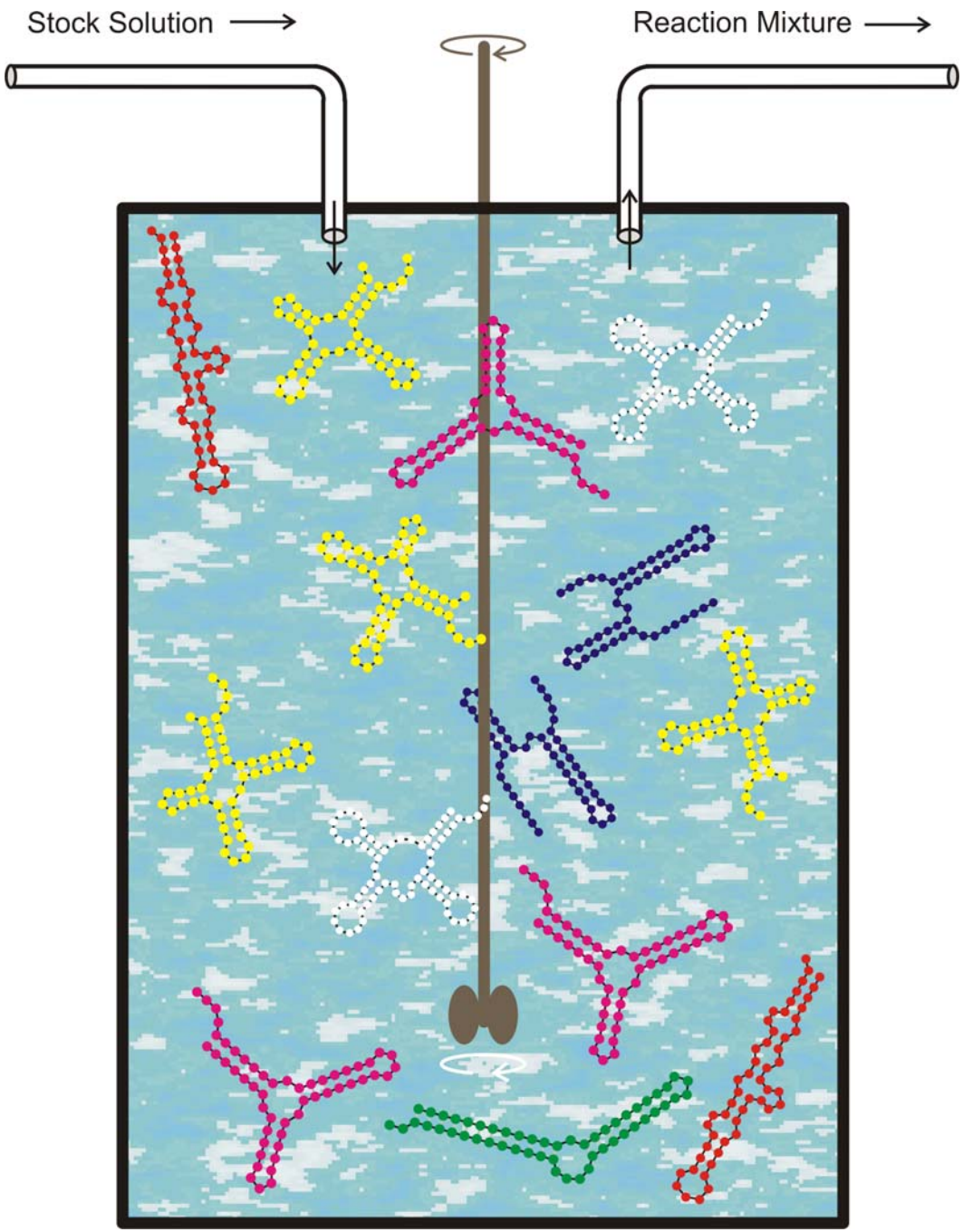
Heide Narnhofer,* Harald A. Posch,† and Walter Thirring‡
Faculty of Physics, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria
(Received 12 June 2007; published 24 October 2007)

We study evolution equations which model selection and mutation within the framework of quantum mechanics. The main question is to what extent order is achieved for an ensemble of typical systems. As an indicator for mixing or purification, a quadratic entropy is used which assumes values between zero for pure states and $(d-1)/d$ for fully mixed states. Here, d is the dimension. Whereas the classical counterpart, the quasispecies dynamics, has previously been found to be predominantly mixing, the quantum quasispecies (QS) evolution surprisingly is found to be strictly purifying for all dimensions. This is also typically true for an alternative formulation (AQS) of this quantum mechanical flow. We compare this also to analogous results for the Lindblad evolution. Although the latter may be viewed as a simple linear superposition of the purifying QS and AQS evolutions, it is found to be predominantly mixing. The reason for this behavior may be explained by the fact that the two subprocesses by themselves converge to different pure states, such that the combined process is mixing. These results also apply to high-dimensional systems.

DOI: 10.1103/PhysRevE.76.041133 PACS number(s): 05.30.−d, 87.23.Kg, 04.20.Ha, 87.10.+e

1. Chemical kinetics of replication and mutation
2. Complexity of fitness landscapes
3. Quasispecies on realistic landscapes
4. Neutrality and replication

1. **Chemical kinetics of replication and mutation**
2. Complexity of fitness landscapes
3. Quasispecies on realistic landscapes
4. Neutrality and replication



Enzyme immobilized

Stock solution:

$[A] = a = a_0$

Flow rate:

$r = \tau_R^{-1}$

The population size N , the number of polynucleotide molecules, is controlled by the flow r

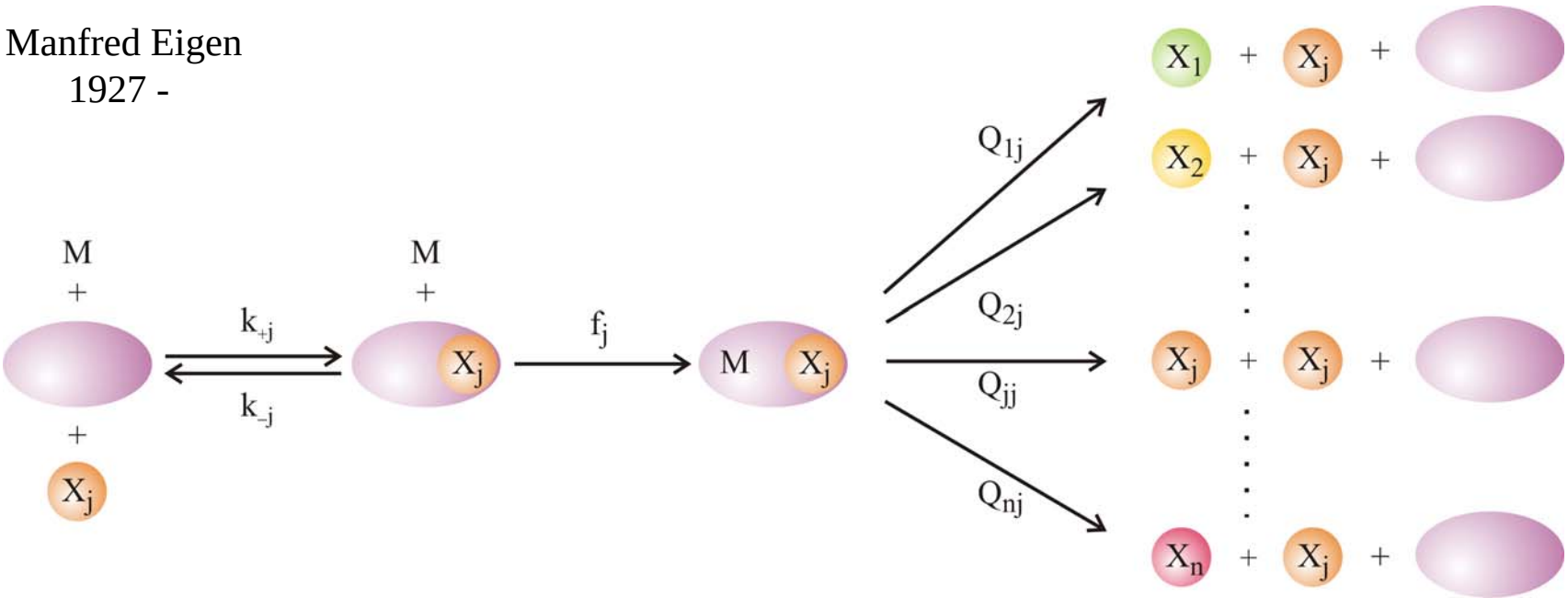
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

The flowreactor is a device for studying evolution ***in vitro*** and ***in silico***

Manfred Eigen
1927 -

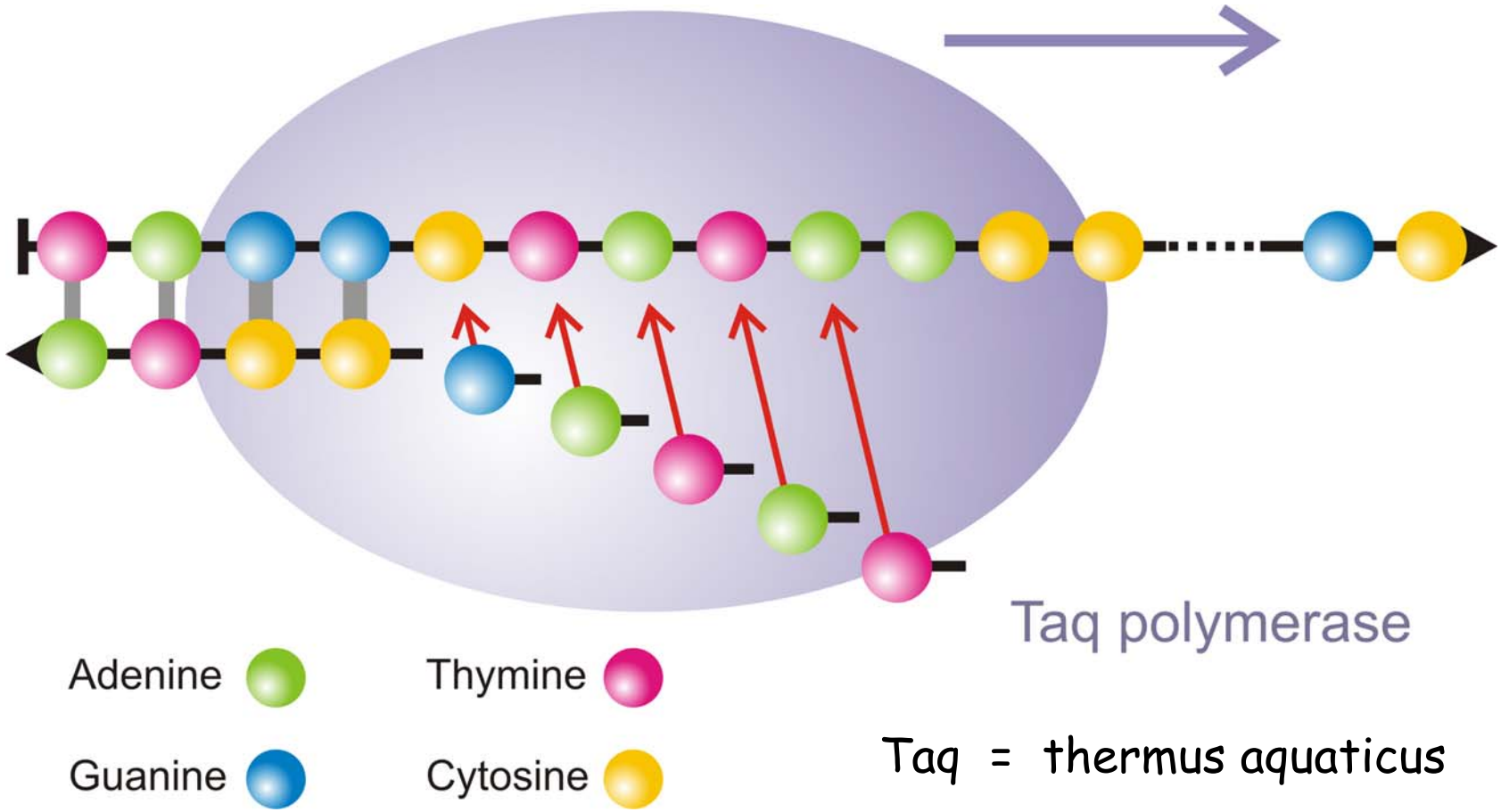
$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

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



Mutation and (correct) replication as parallel chemical reactions

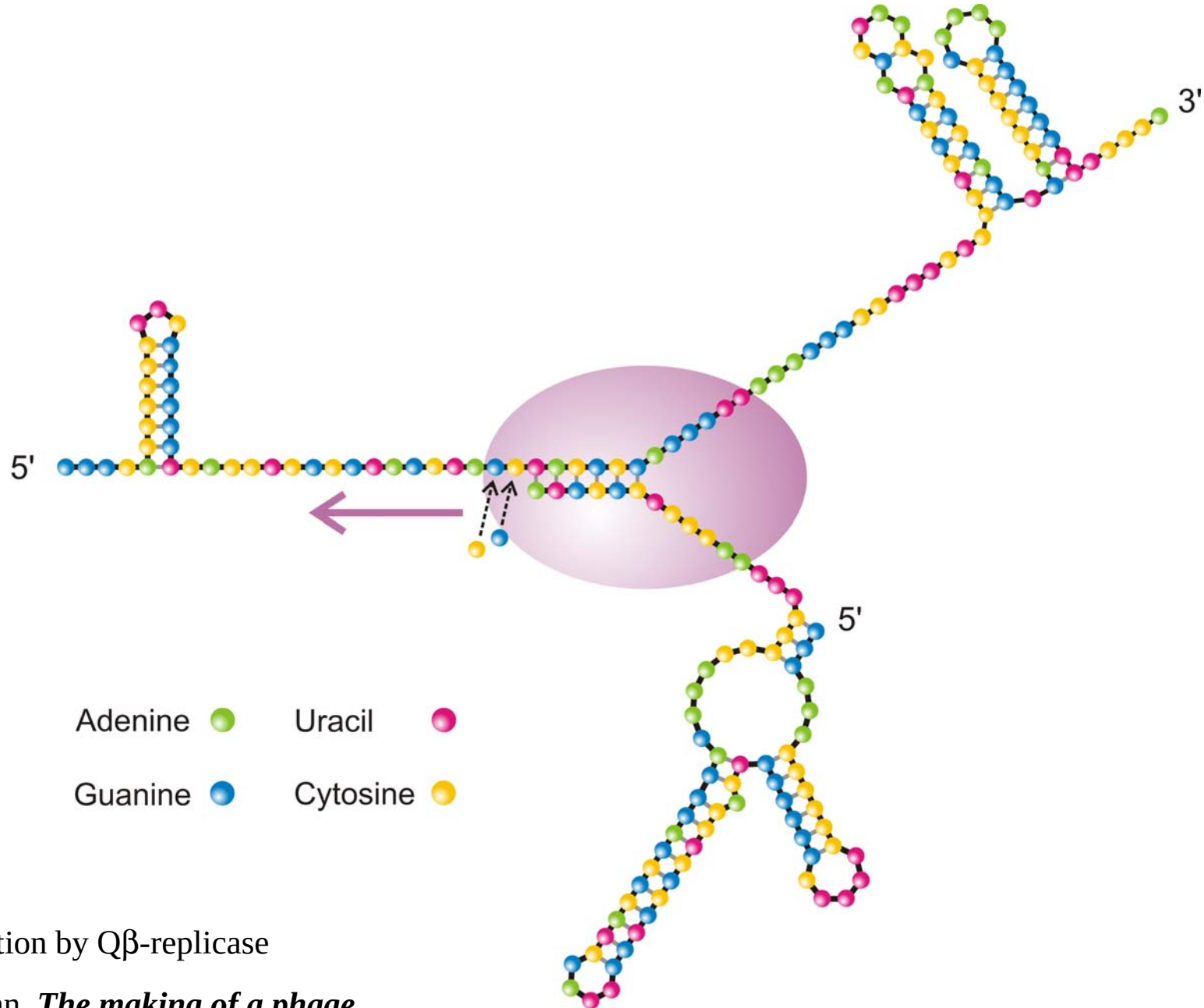
M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977. *Naturwissenschaften* 64:541, 65:7 und 65:341



Taq = thermus aquaticus

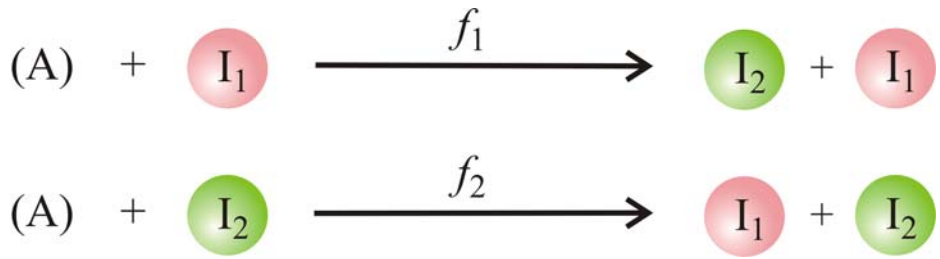
Accuracy of replication: $Q = q_1 \cdot q_2 \cdot q_3 \cdot \ldots \cdot q_n$

The logics of DNA replication



RNA replication by Qβ-replicase

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



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

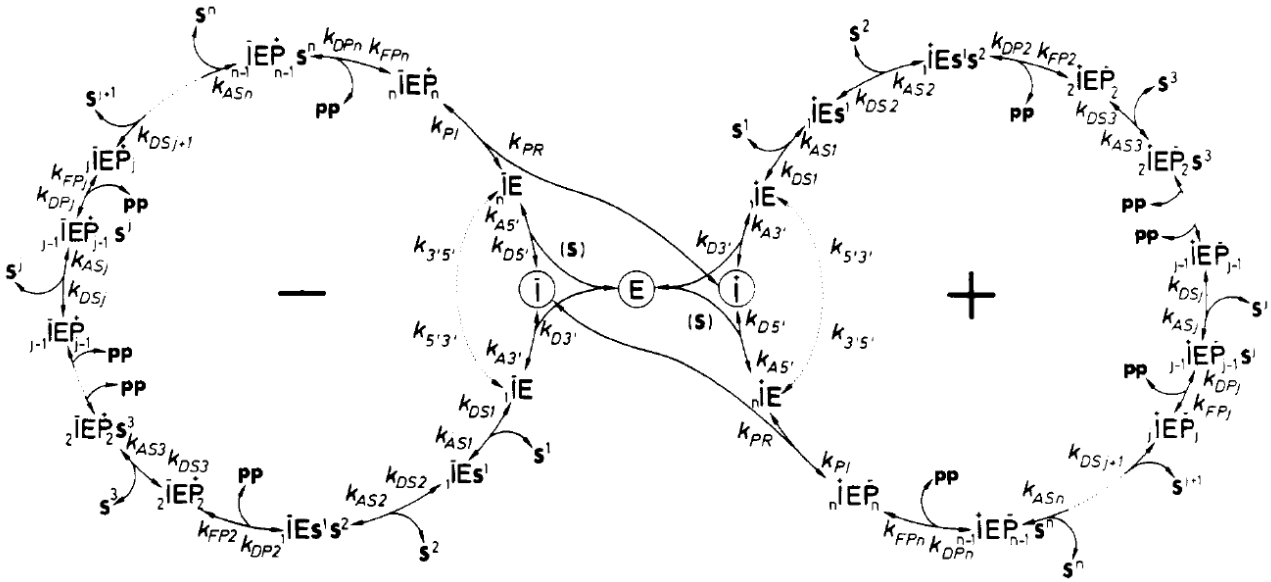
$$x_1 = \sqrt{f_2}\;\xi_1\,, \quad x_2 = \sqrt{f_1}\;\xi_2\,, \quad \zeta = \xi_1 + \xi_2\,, \quad \eta = \xi_1 - \xi_2\,, \quad f = \sqrt{f_1 f_2}$$

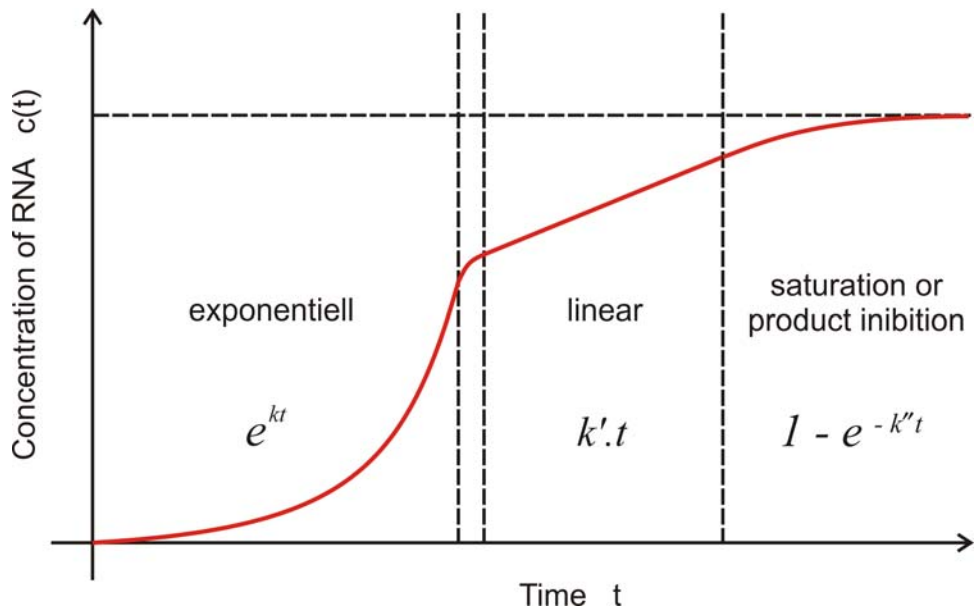
$$\eta(t) = \eta(0)\, e^{-ft}$$

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

Complementary replication as the simplest molecular mechanism of reproduction

Christof K. Biebricher,
1941-2009





Kinetics of RNA replication

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

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

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

Decomposition of matrix W

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

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

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

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

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

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

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

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

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

Perron-Frobenius theorem applied to the value matrix W

W is primitive: (i) λ_0 is real and strictly positive

(ii) $\lambda_0 > |\lambda_k|$ for all $k \neq 0$

(iii) λ_0 is associated with strictly positive eigenvectors

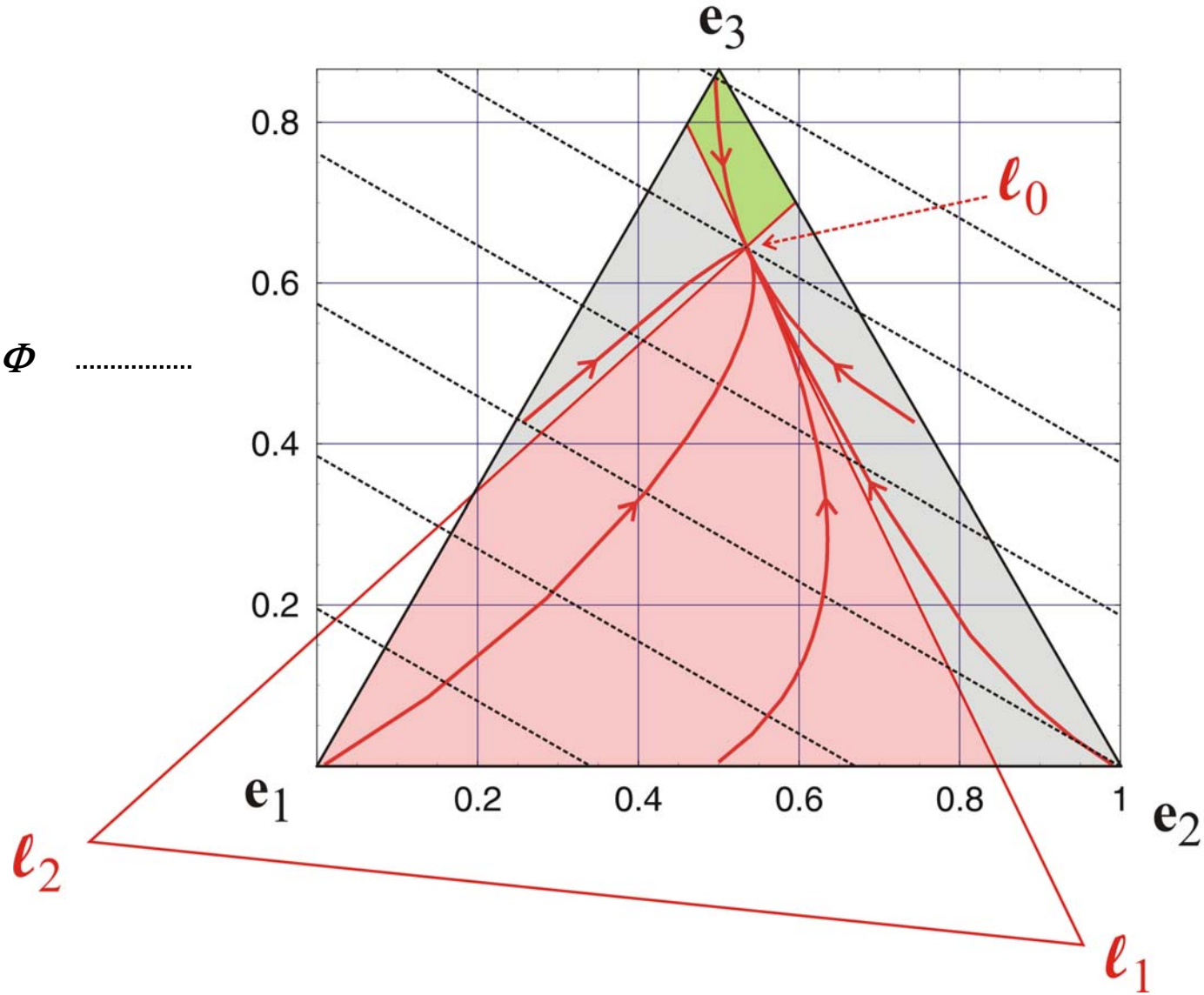
(iv) λ_0 is a simple root of the characteristic equation of W

(v-vi) etc.

W is irreducible: (i), (iii), (iv), etc. as above

(ii) $\lambda_0 \geq |\lambda_k|$ for all $k \neq 0$

constant level sets of Φ .....



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

Uniform error rate model:

$$Q_{ij} \;=\; p^{d_H(\mathbf{X}_i,\mathbf{X}_j)}\,(1-p)^{\left(n-d_H(\mathbf{X}_i,\mathbf{X}_j)\right)}$$

$d_H(\mathbf{X}_i, \mathbf{X}_j)$ ... Hamming distance

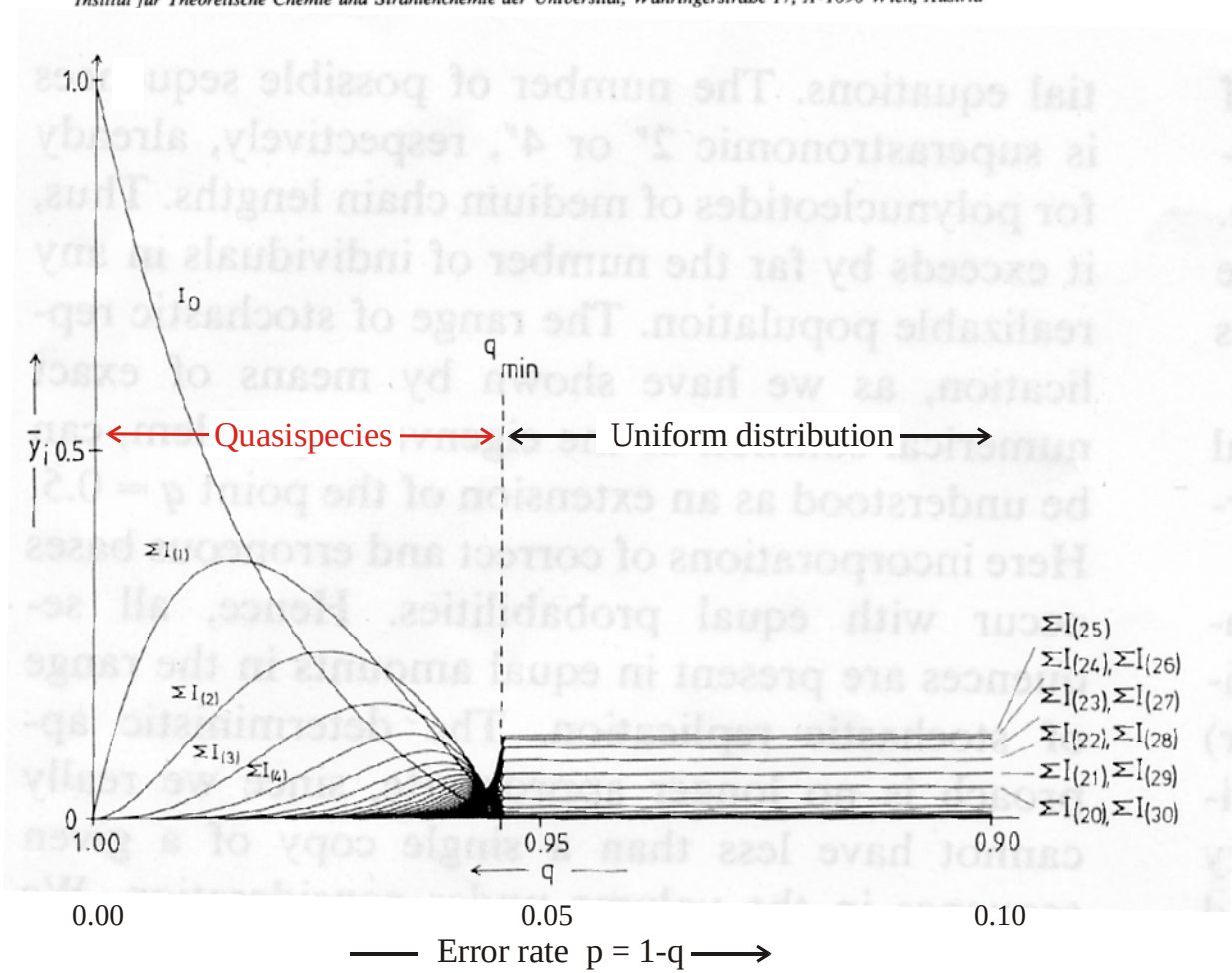
Biophysical Chemistry 16 (1982) 329–345
Elsevier Biomedical Press

SELF-REPLICATION WITH ERRORS

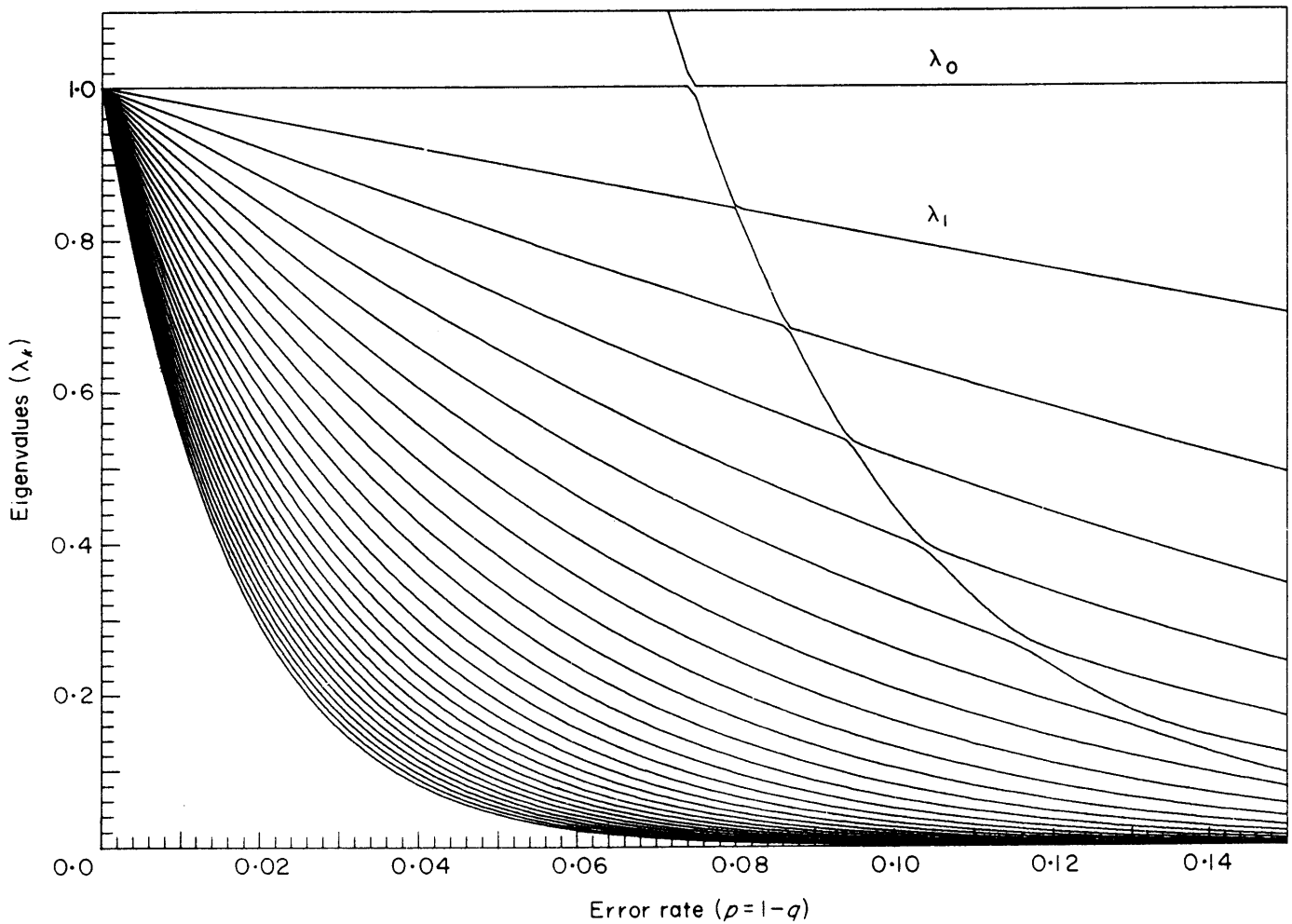
A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

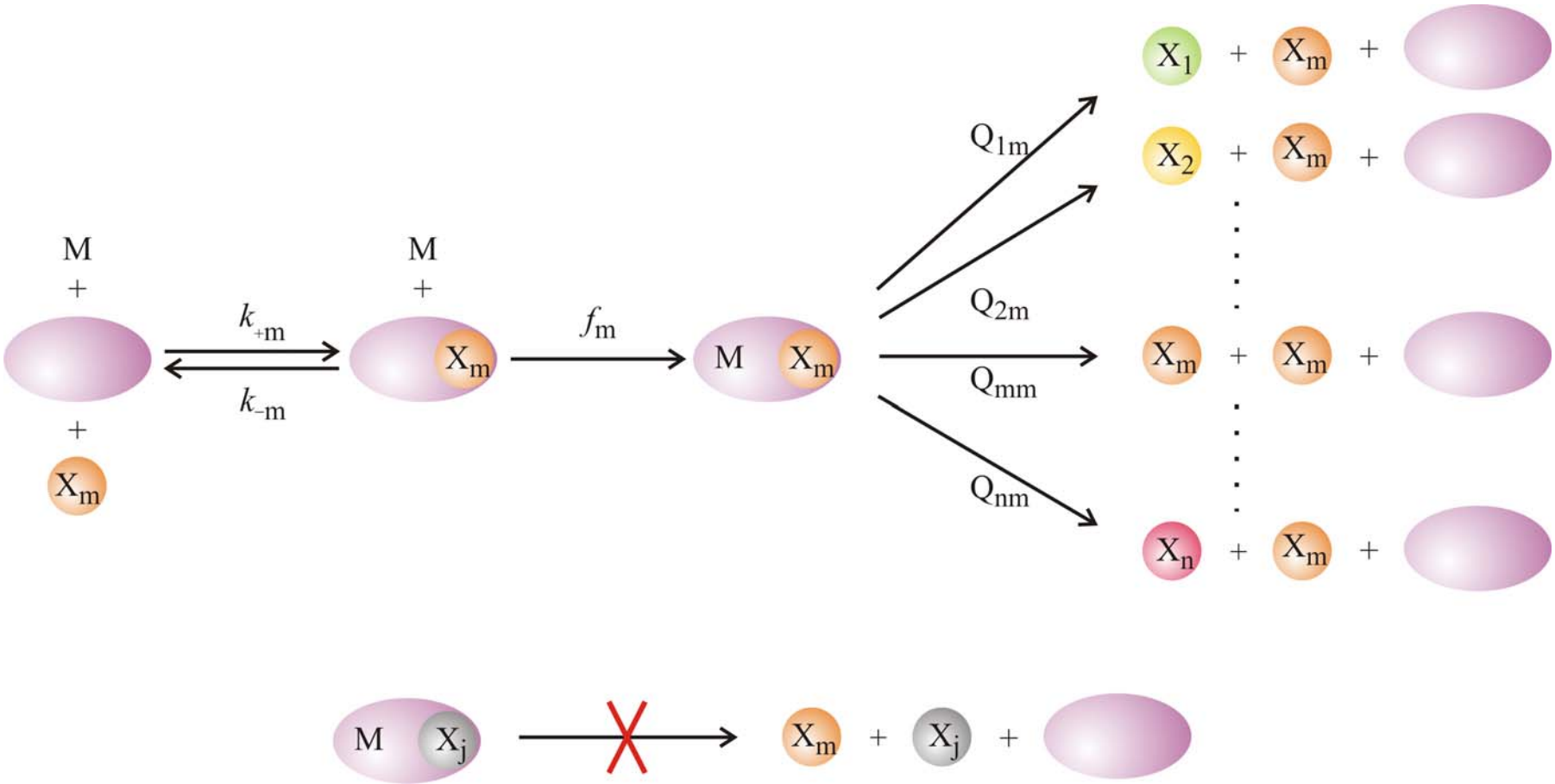
Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria



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



Eigenvalues of the matrix W as a function of the error rate p



The no-mutational backflow or zeroth order approximation

$$\frac{\mathrm{d}x_m^{(0)}}{\mathrm{dt}} = x_m^{(0)}\left(Q_{mm} f_m - \phi(t)\right) = 0 \quad \text{and} \quad \phi(t) = Q_{mm} f_m$$

The ‚no-mutational-backflow‘ or zeroth order approximation

$$\frac{\mathrm{d}x_m^{(0)}}{\mathrm{dt}} = x_m^{(0)}\left(Q_{mm} f_m - \phi(t)\right) = 0 \quad \text{and} \quad \phi(t) = Q_{mm} f_m$$

$$\bar{x}_m^{(0)} = \frac{Q_{mm} - \sigma_m^{-1}}{1 - \sigma_m^{-1}} = \frac{1}{\sigma_m - 1}\left(\sigma_m (1-p)^n - 1\right)$$

The ‚no-mutational-backflow‘ or zeroth order approximation

$$\frac{\mathrm{d}x_m^{(0)}}{\mathrm{dt}} = x_m^{(0)}\left(Q_{mm} f_m - \phi(t)\right) = 0 \quad \text{and} \quad \phi(t) = Q_{mm} f_m$$

$$\bar{x}_m^{(0)} = \frac{Q_{mm} - \sigma_m^{-1}}{1 - \sigma_m^{-1}} = \frac{1}{\sigma_m - 1}\left(\sigma_m (1-p)^n - 1\right)$$

$$\bar{x}_m^{(0)} = 0 \quad \Rightarrow \quad (1-p)^n = \sigma_m^{-1} \quad \text{and} \quad p_{\mathrm{cr}} \approx 1 - (\sigma_m)^{-1/n}$$

$$\sigma_m = \frac{f_m}{\bar{f}_{-m}} \quad \text{and} \quad \bar{f}_{-m} = \frac{1}{(1-x_m)} \sum_{i=1, i\neq m}^{N} x_i f_i$$

The ‚no-mutational-backflow‘ or zeroth order approximation

Chain length and error threshold

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

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

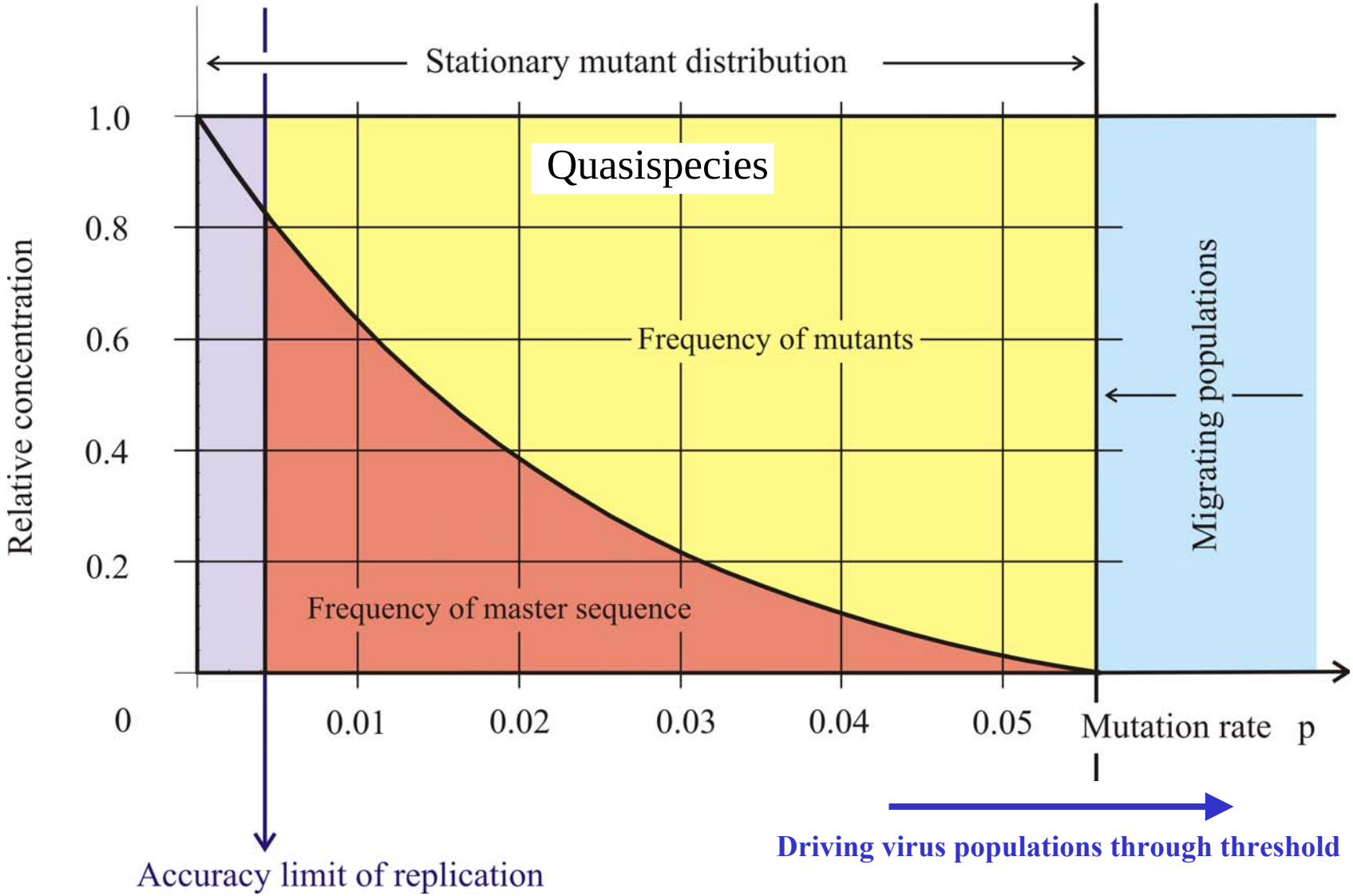
n … constant: $p_{\max} \approx \dfrac{\ln\sigma_m}{n}$

$Q = (1-p)^n$ … replication accuracy

p … error rate

n … chain length

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



The error threshold in replication: No mutational backflow approximation

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.

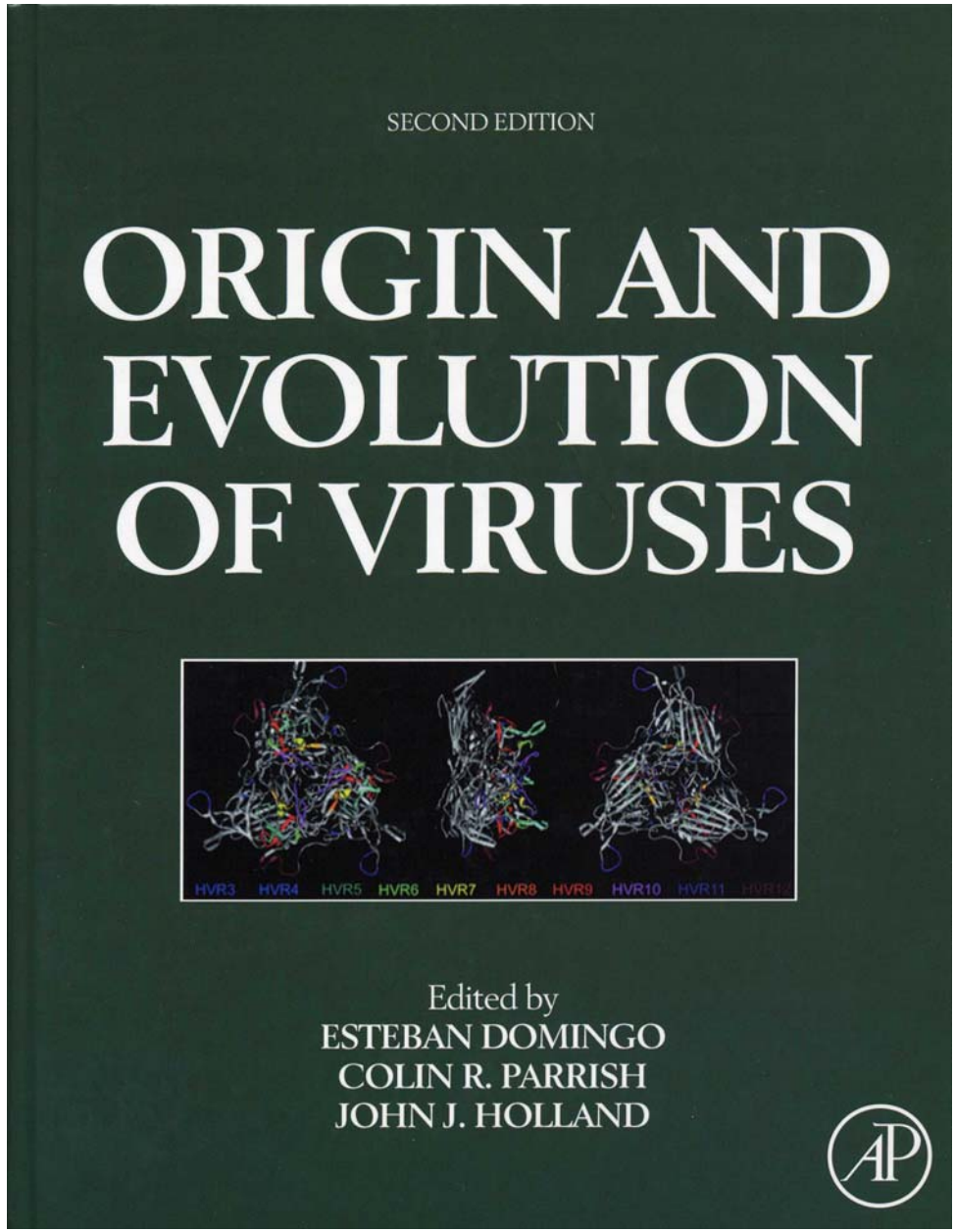
My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with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

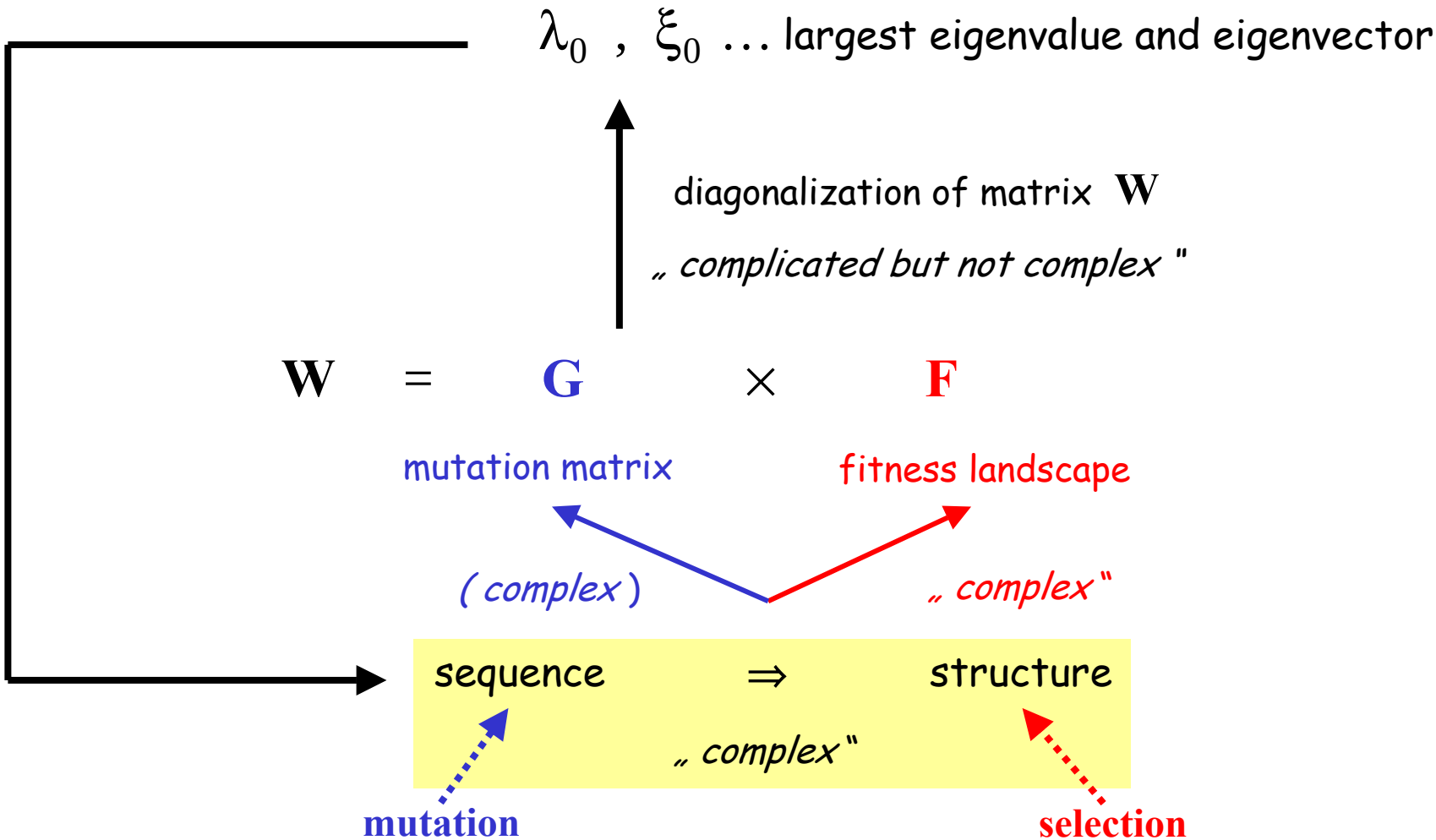
Available online 8 December 2004

doi:10.1016/j.virusres.2004.11.001

Molecular evolution of viruses



1. Chemical kinetics of replication and mutation
2. **Complexity of fitness landscapes**
3. Quasispecies on realistic landscapes
4. Neutrality and replication



Complexity in molecular evolution

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.

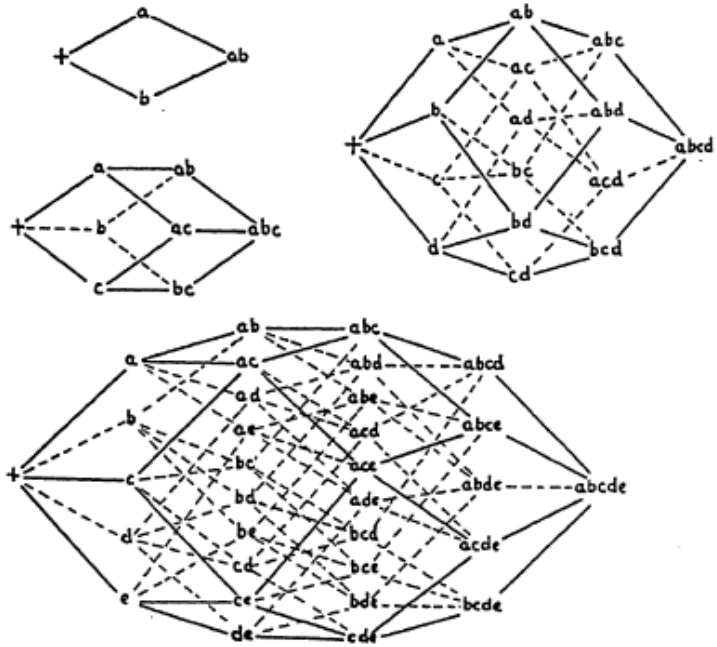


FIG. 1.—The combinations of from 2 to 5 paired allelomorphs.

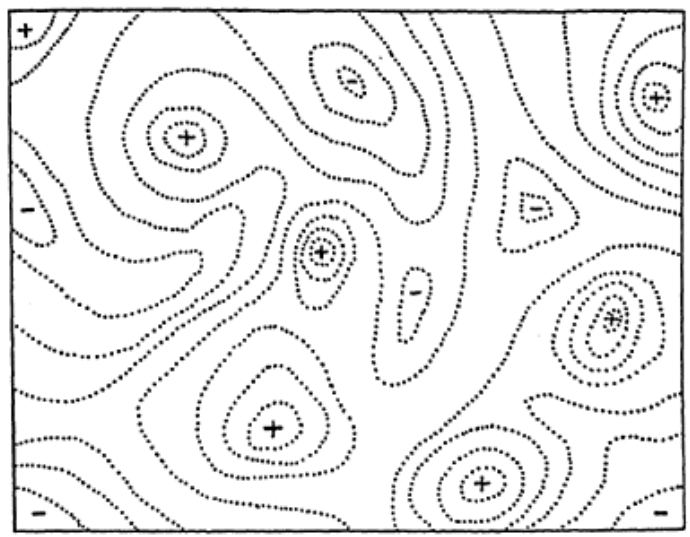
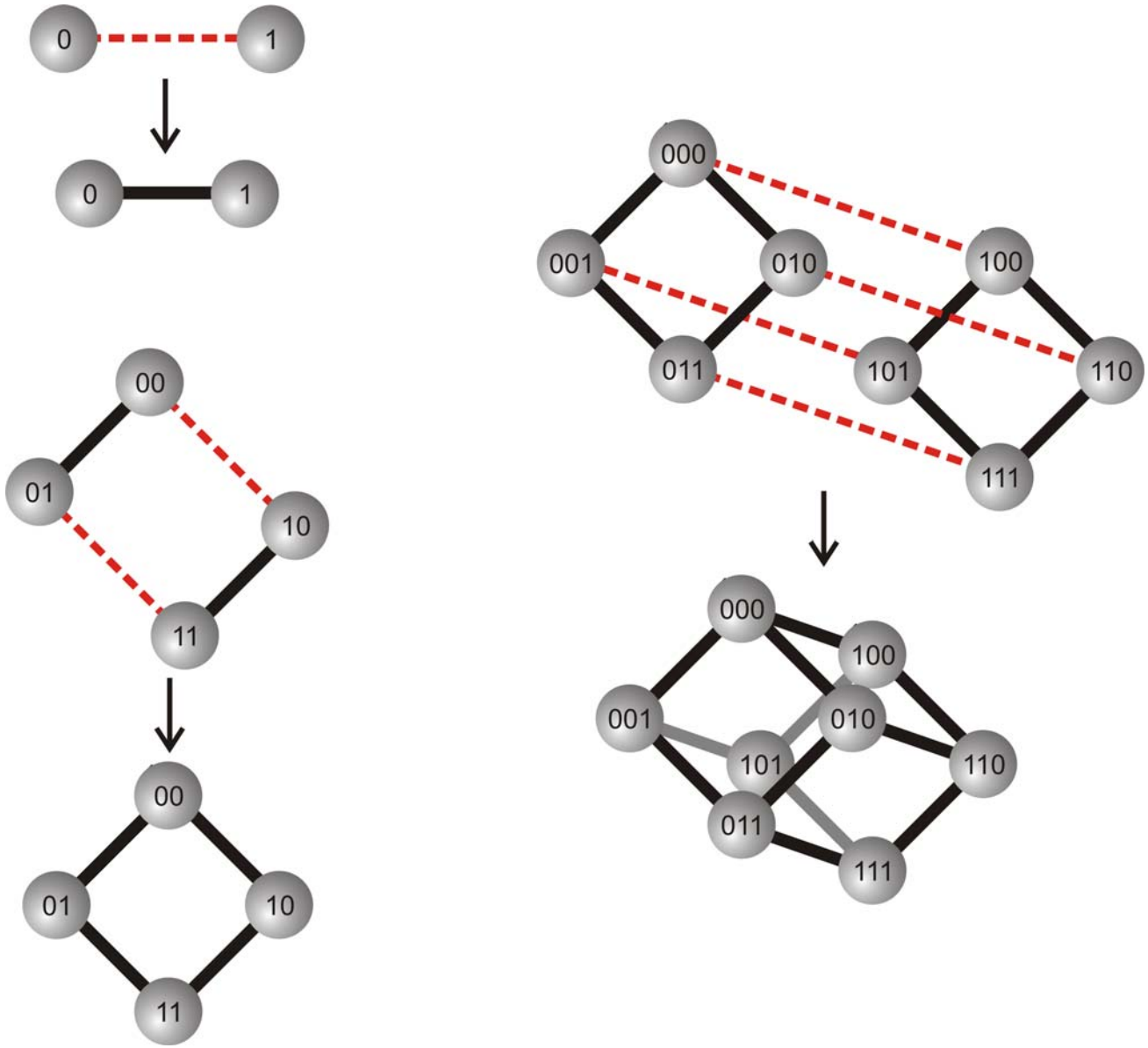


FIG. 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

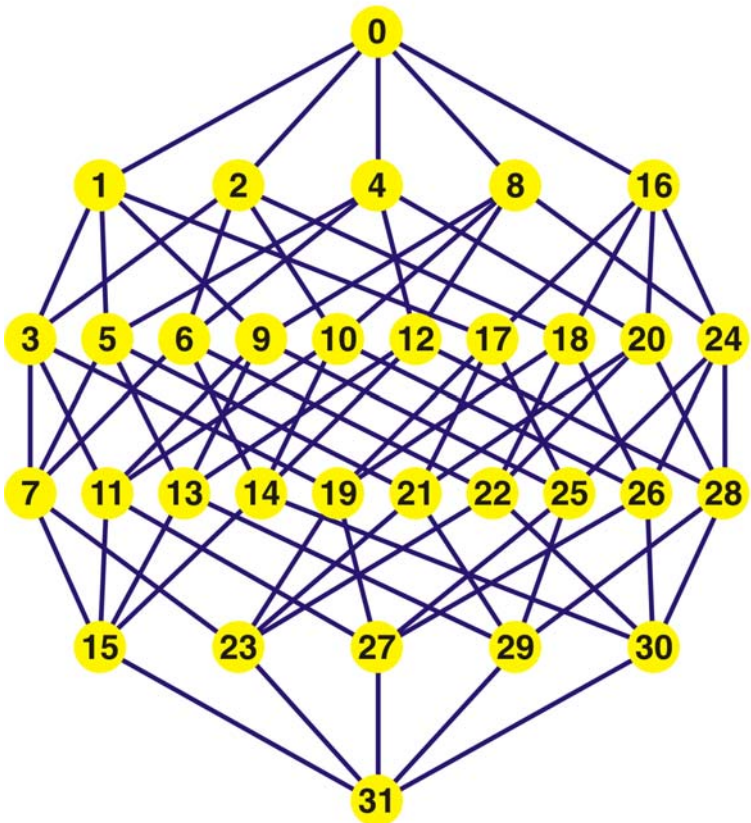
Sewall Wright. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.

-- --. 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

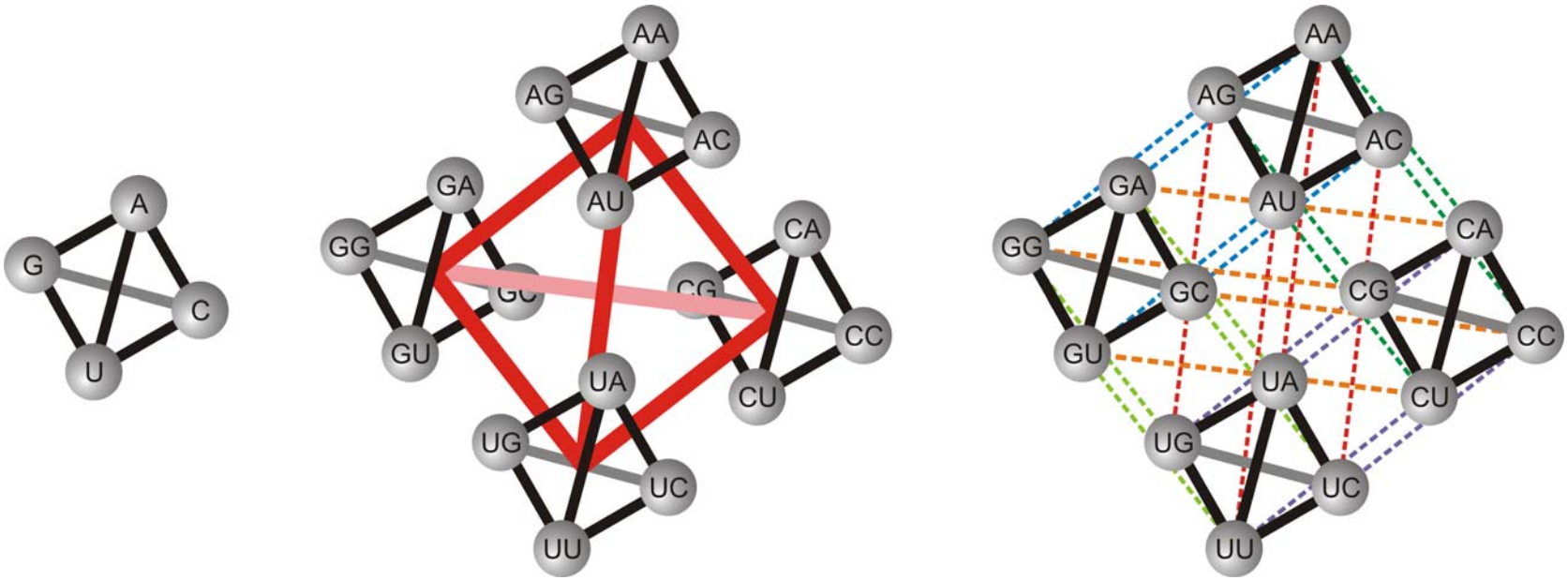
Binary sequences can be encoded by their decimal equivalents:

C = 0 and G = 1, for example,

"0" ≡ 00000 = CCCCC,

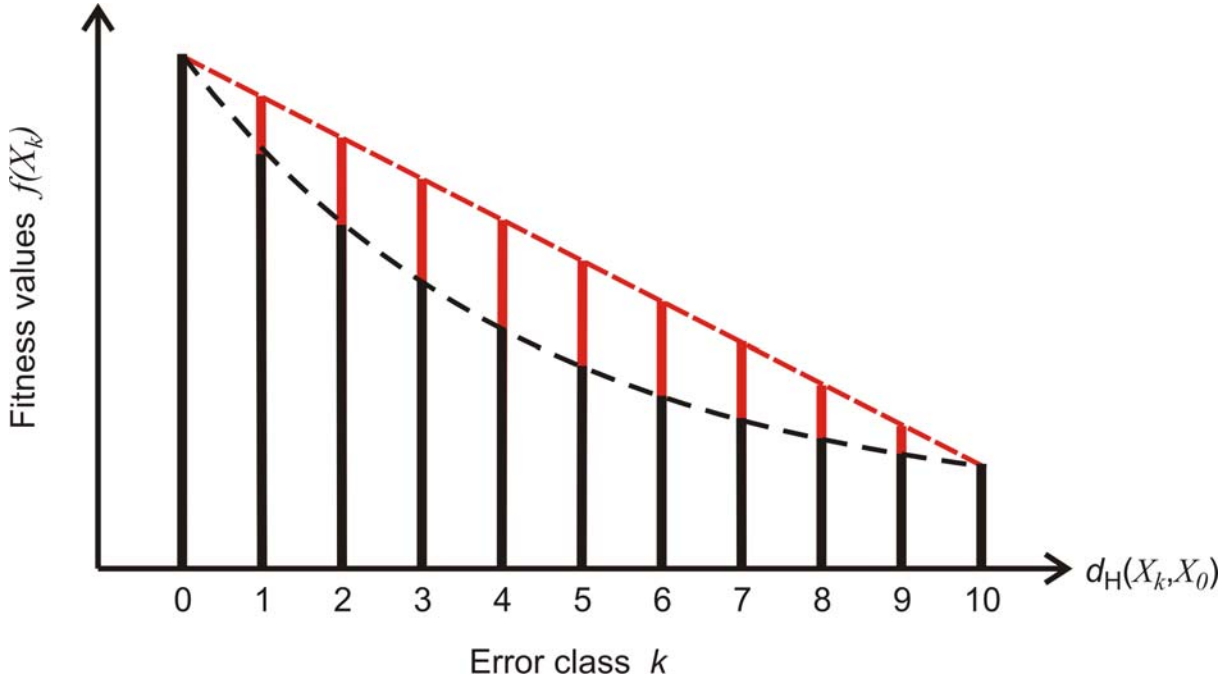
"14" ≡ 01110 = CGGGC,

"29" ≡ 11101 = GGGCG, etc.

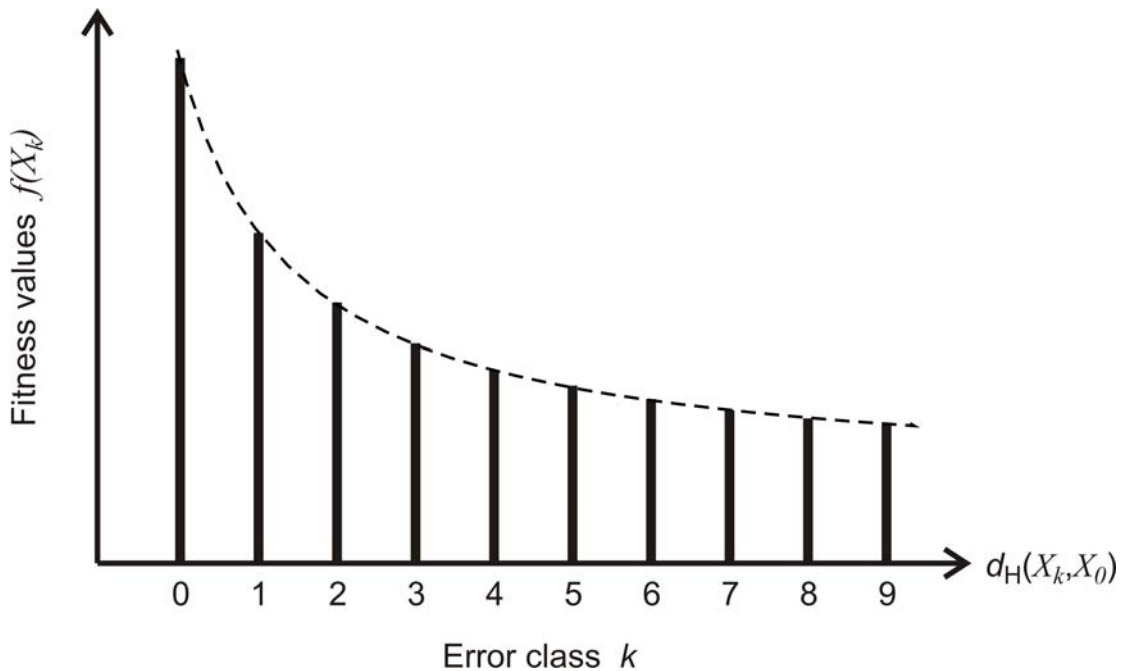


Build-up principle of four letter (AUGC) sequence spaces

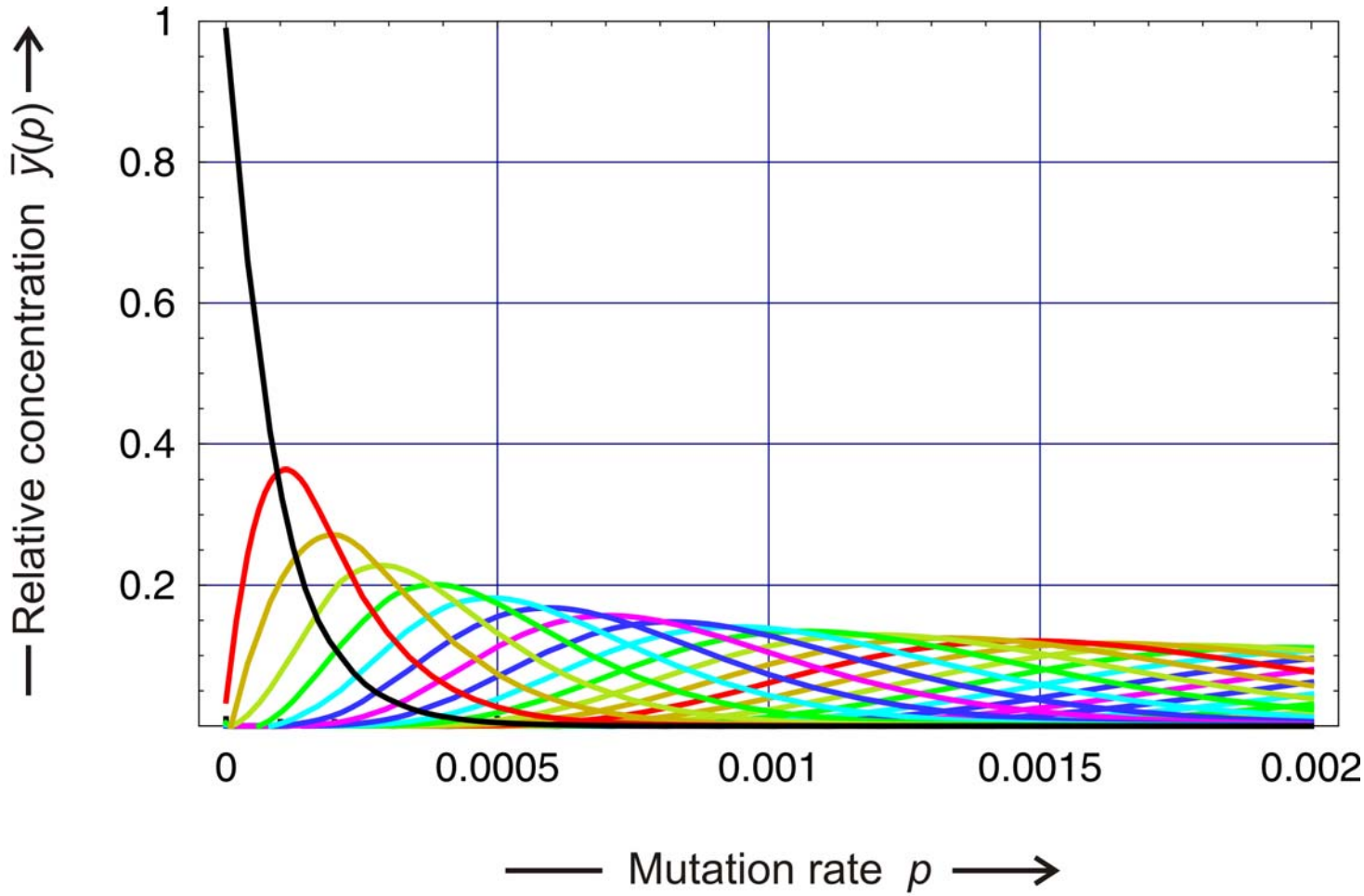
linear and
multiplicative



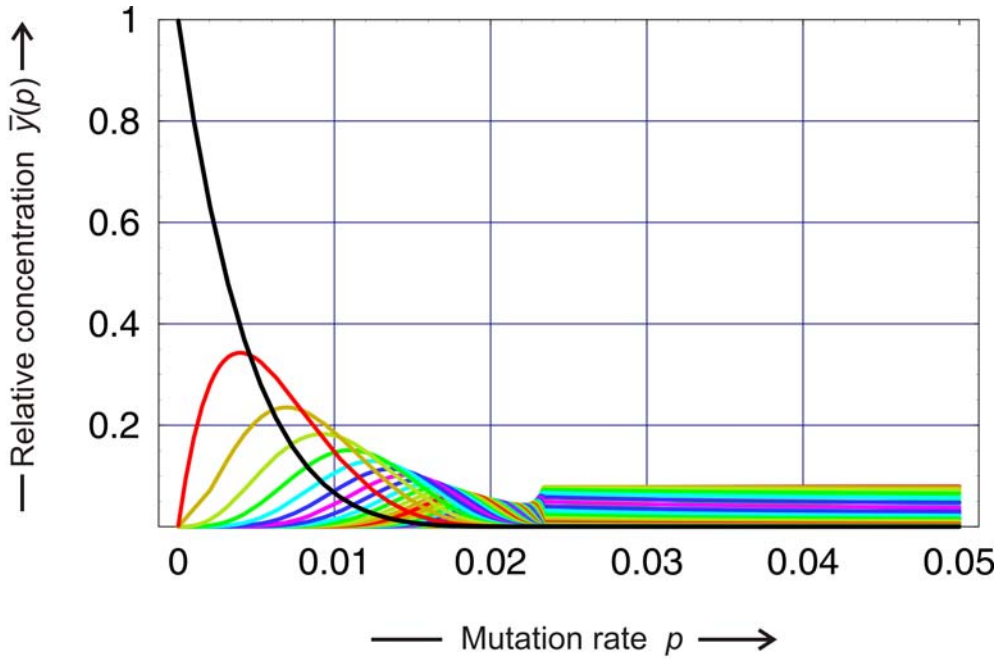
hyperbolic

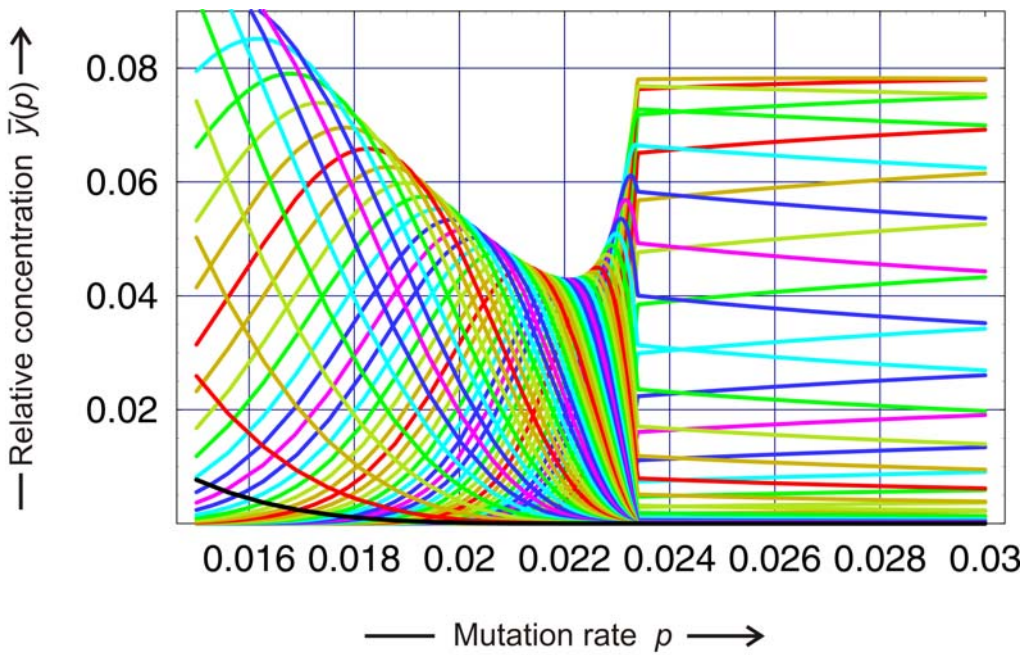


Smooth fitness landscapes



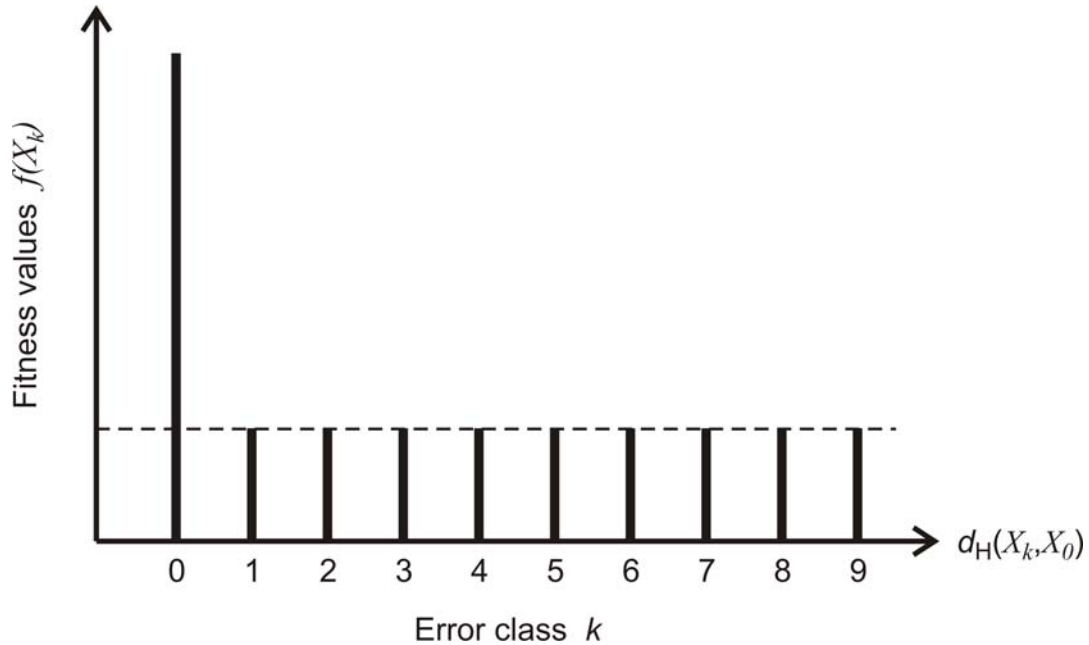
The linear fitness landscape shows no error threshold



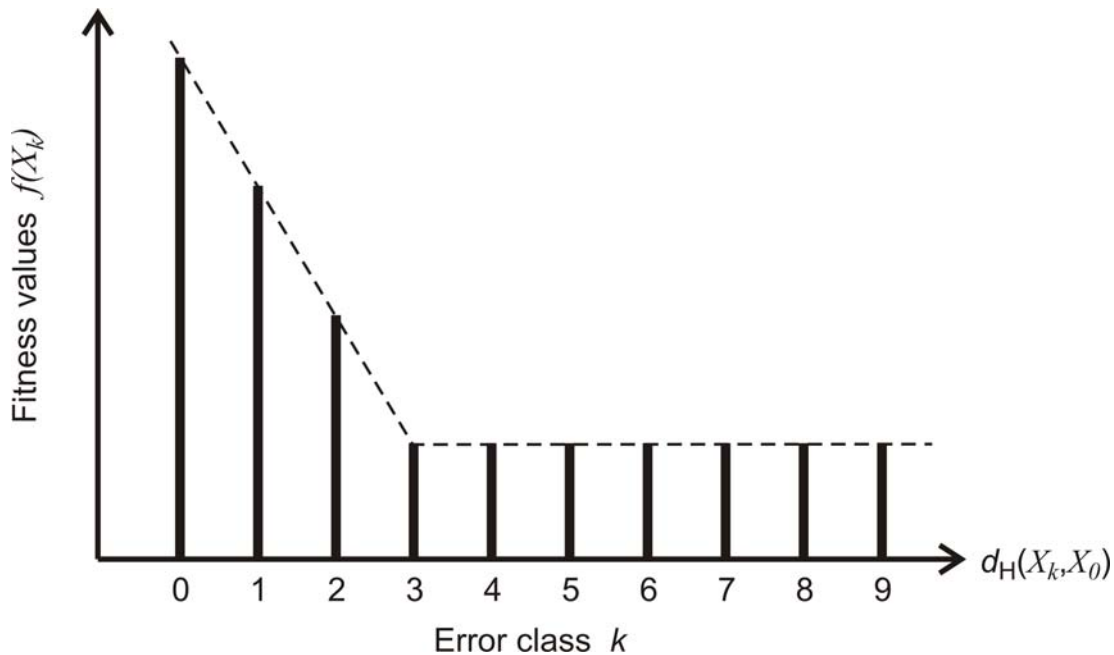


Error threshold on the hyperbolic landscape

single peak landscape

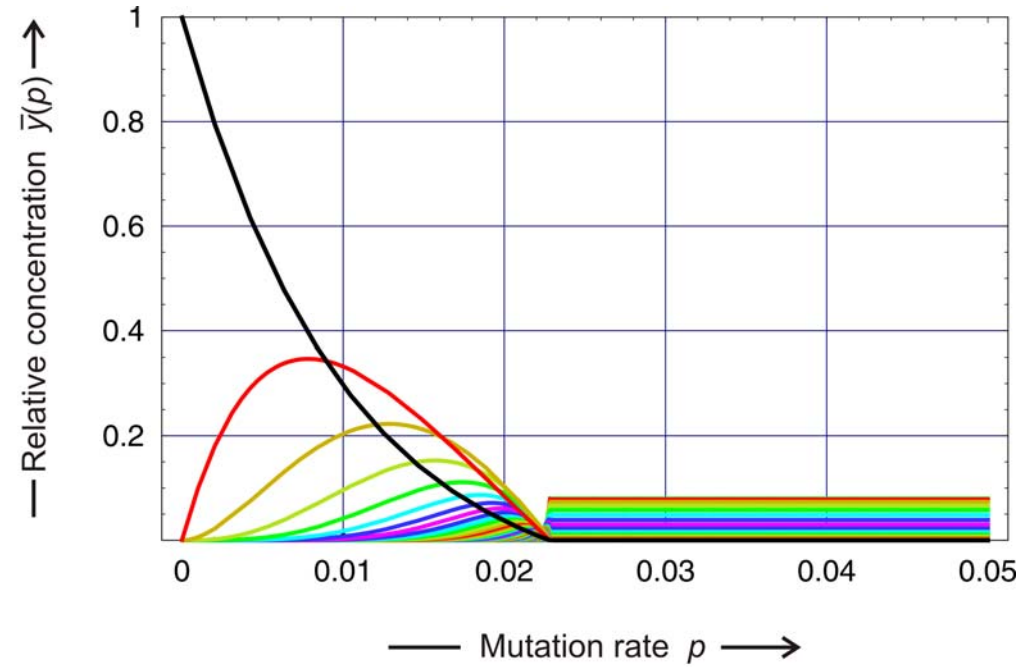


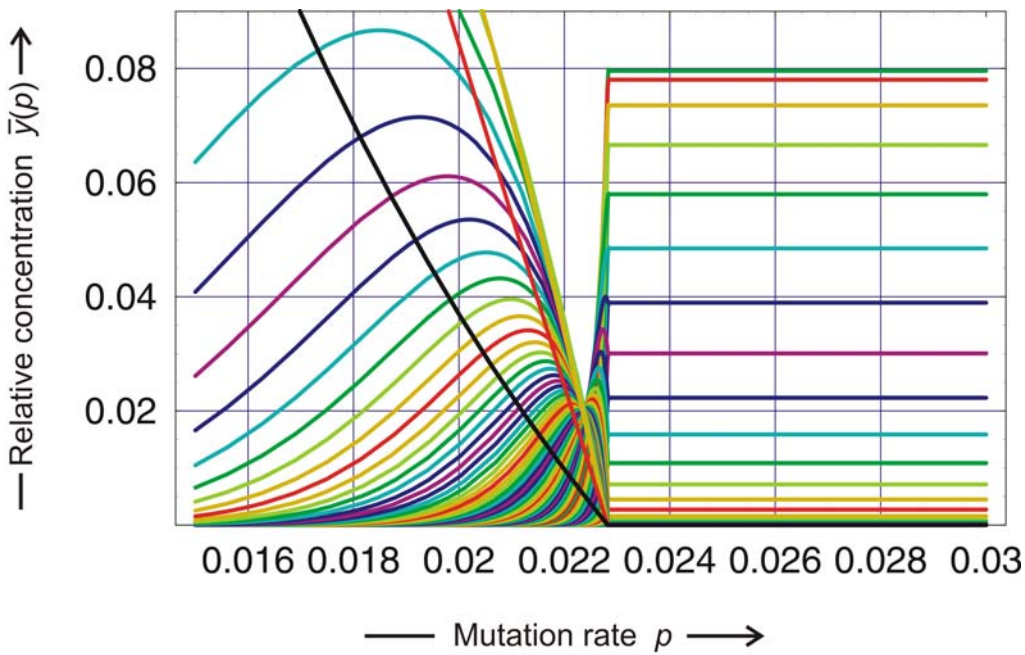
step linear landscape



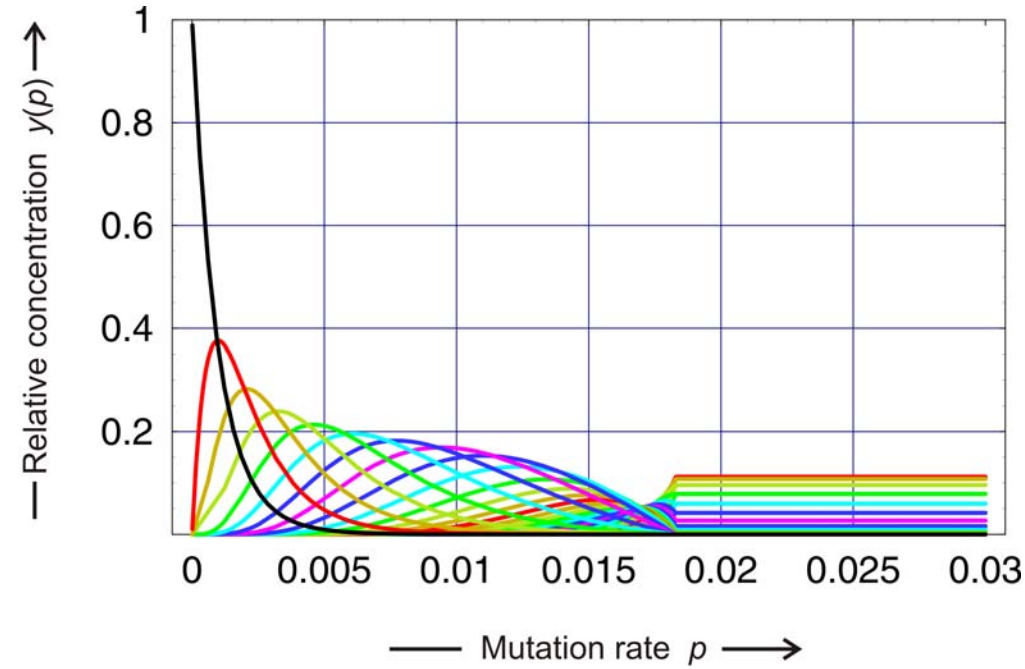
Rugged fitness landscapes

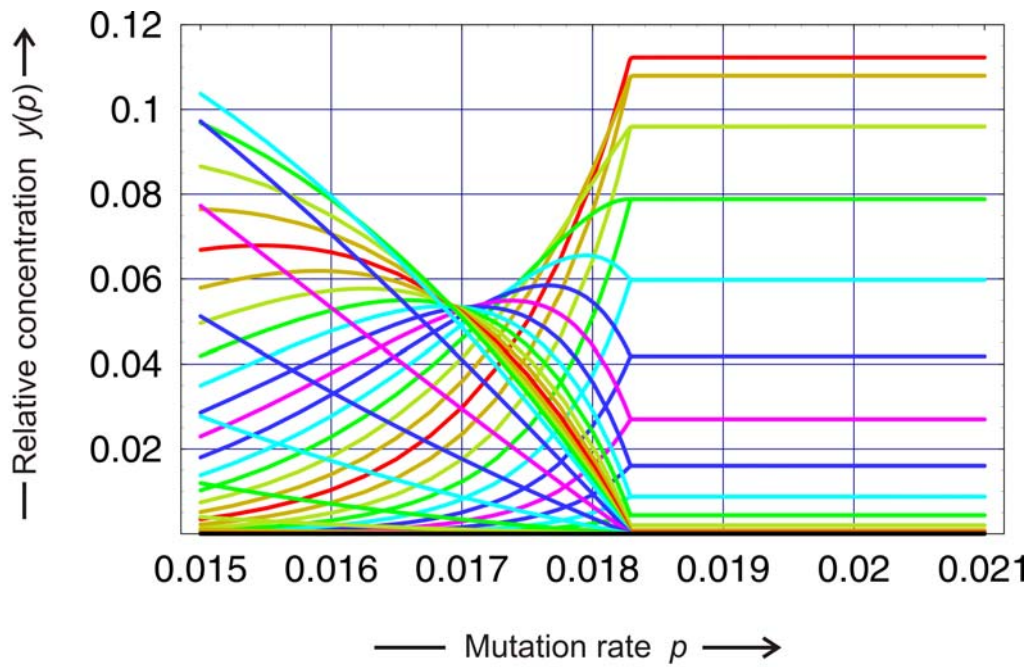
Error threshold on the single peak landscape





Error threshold on the step linear landscape





The error threshold can be separated into three phenomena:

1. Decrease in the concentration of the master sequence to very small values.

2. Sharp change in the stationary concentration of the quasispecies distribuiton.

3. Transition to the uniform distribution at small mutation rates.

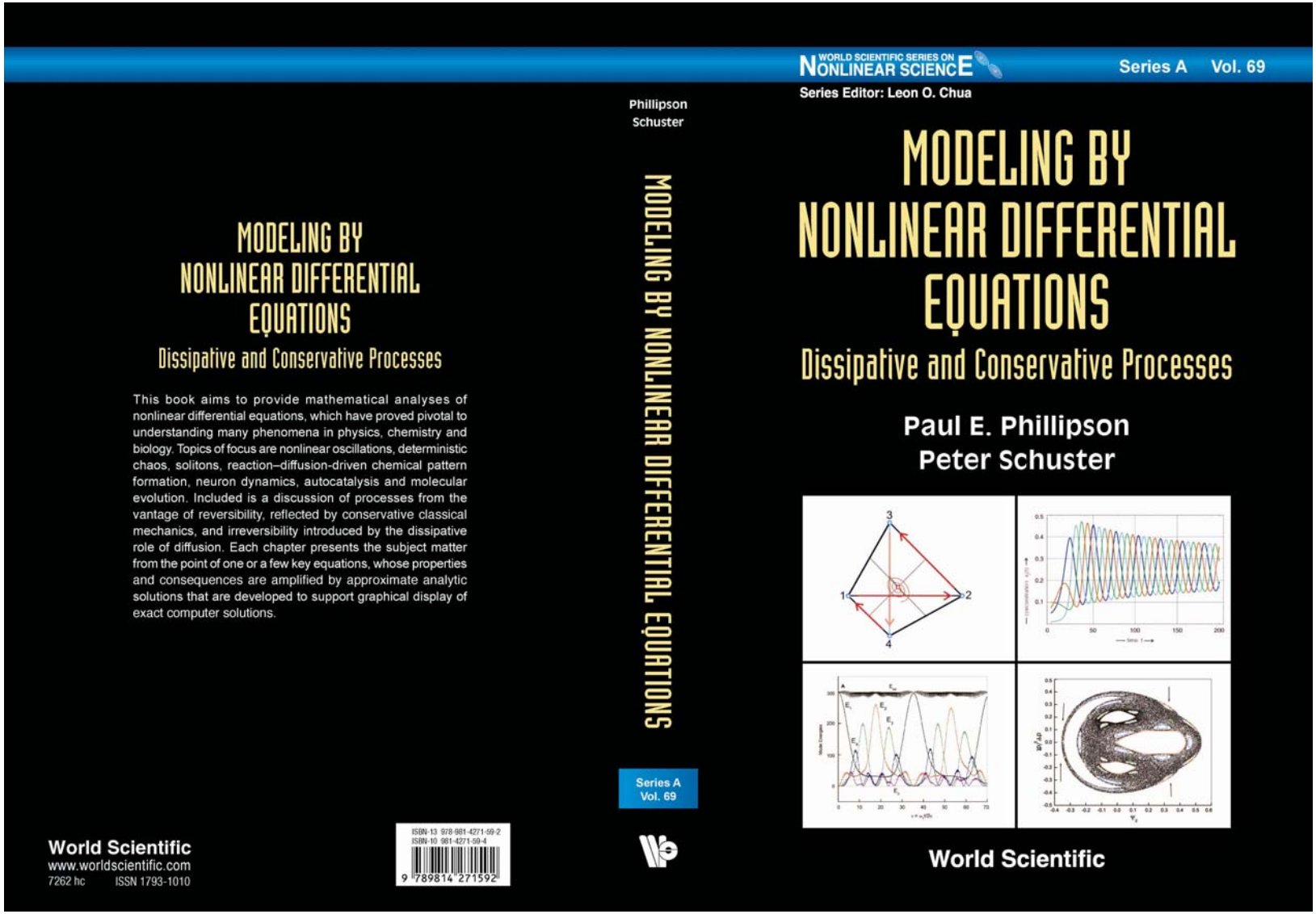
The error threshold can be separated into three phenomena:

1. Decrease in the concentration of the master sequence to very small values.

2. Sharp change in the stationary concentration of the quasispecies distribuiton.

3. Transition to the uniform distribution at small mutation rates.

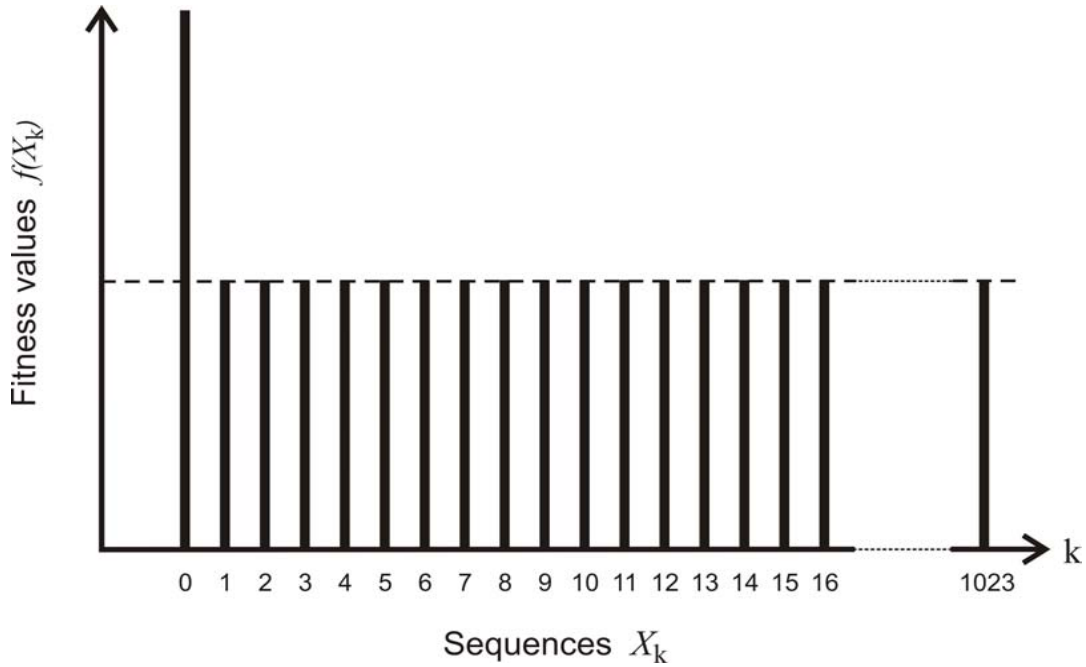
All three phenomena coincide for the quasispecies on the single peak fitness lanscape.



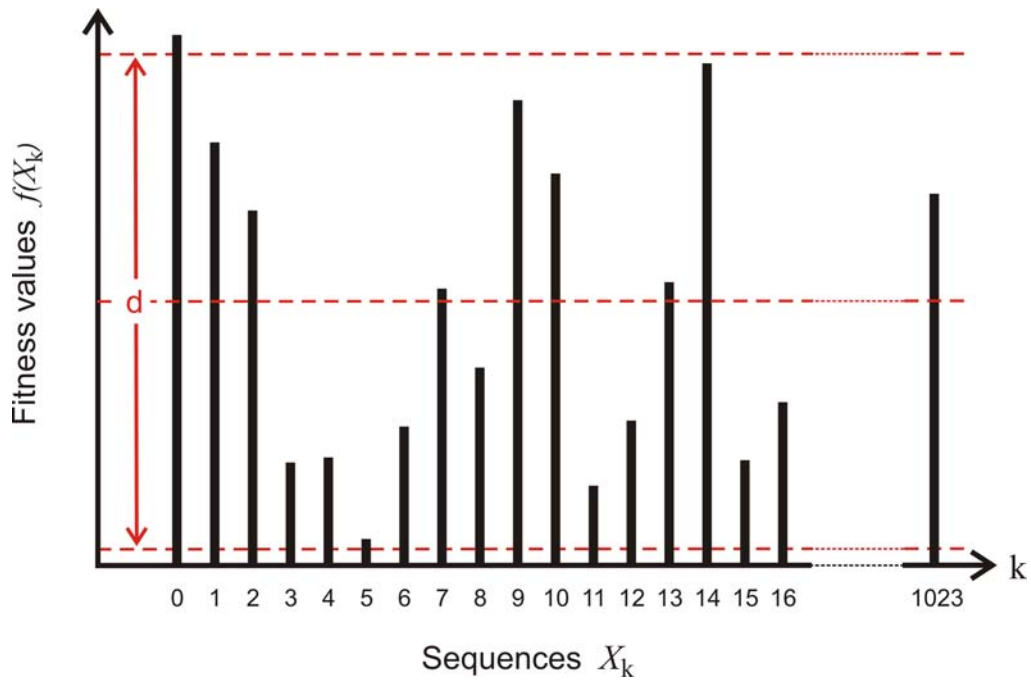
Paul E. Phillipson, Peter Schuster. (2009) Modeling by nonlinear differential equations. Dissipative and conservative processes. World Scientific, Singapore, pp.9-60.

1. Chemical kinetics of replication and mutation
2. Complexity of fitness landscapes
3. **Quasispecies on realistic landscapes**
4. Neutrality and replication

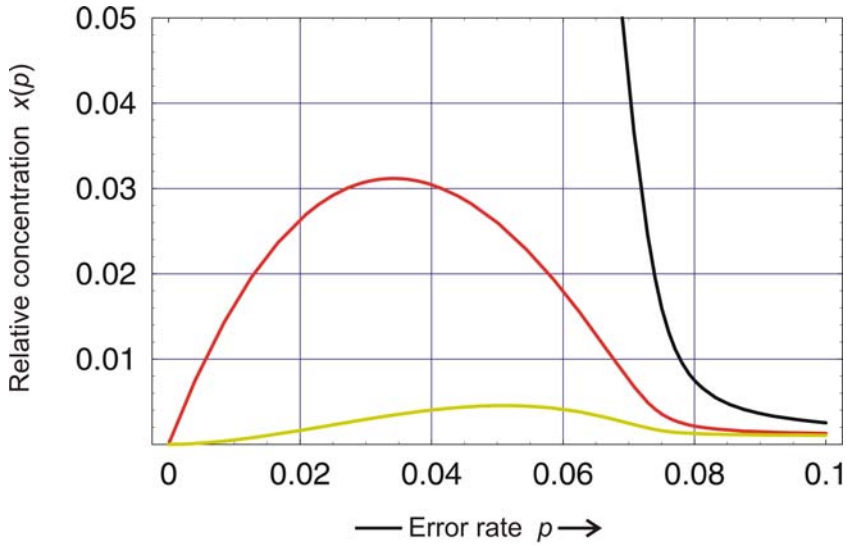
single peak landscape

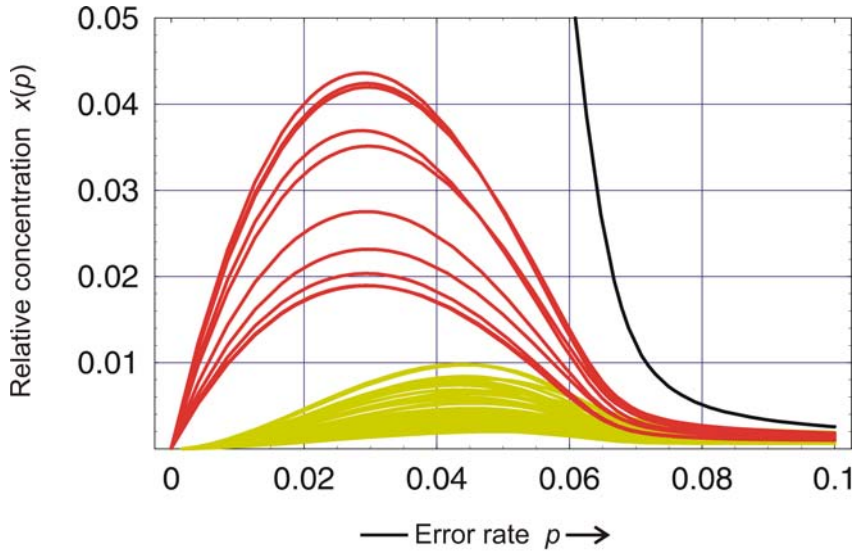


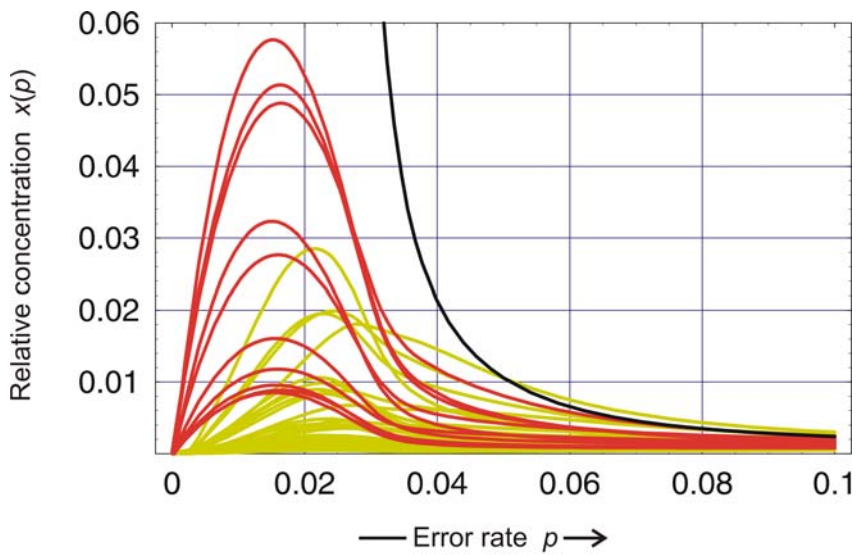
„realistic“ landscape



Rugged fitness landscapes
over individual binary sequences
with n = 10



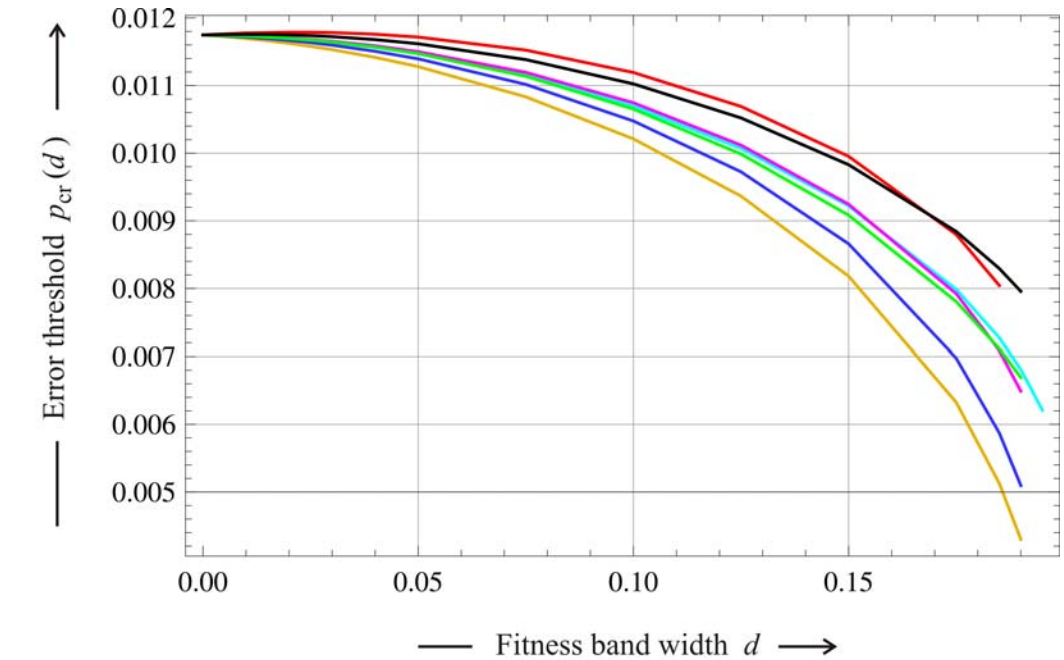


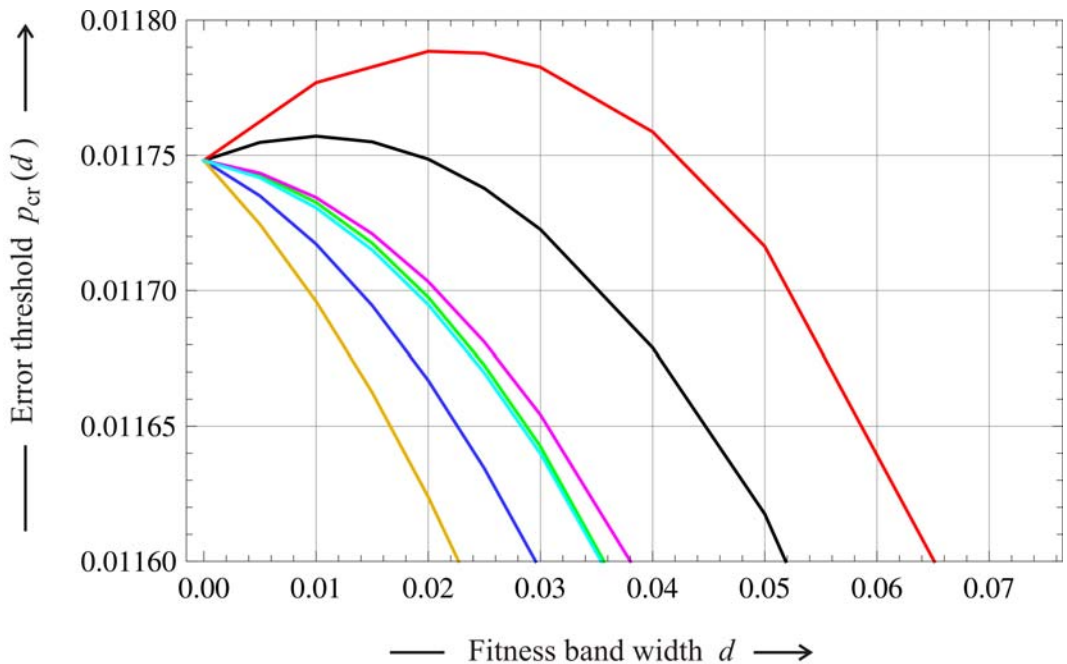


Error threshold: Individual sequences

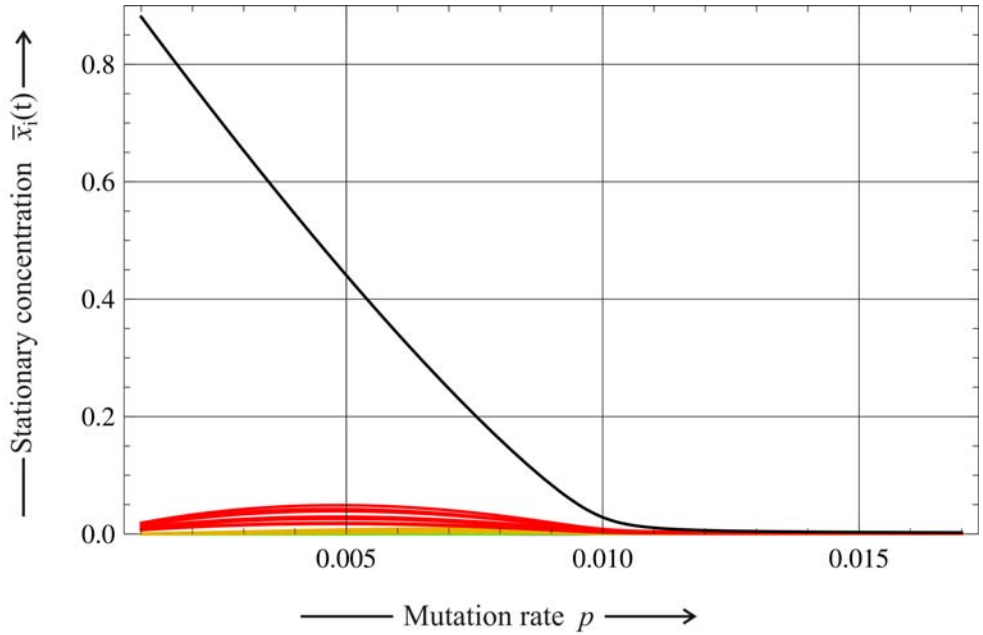
n = 10, σ = 2, s = 491 and d = 0, 1.0, 1.875

Shift of the error threshold with increasing ruggedness of the fitness landscape

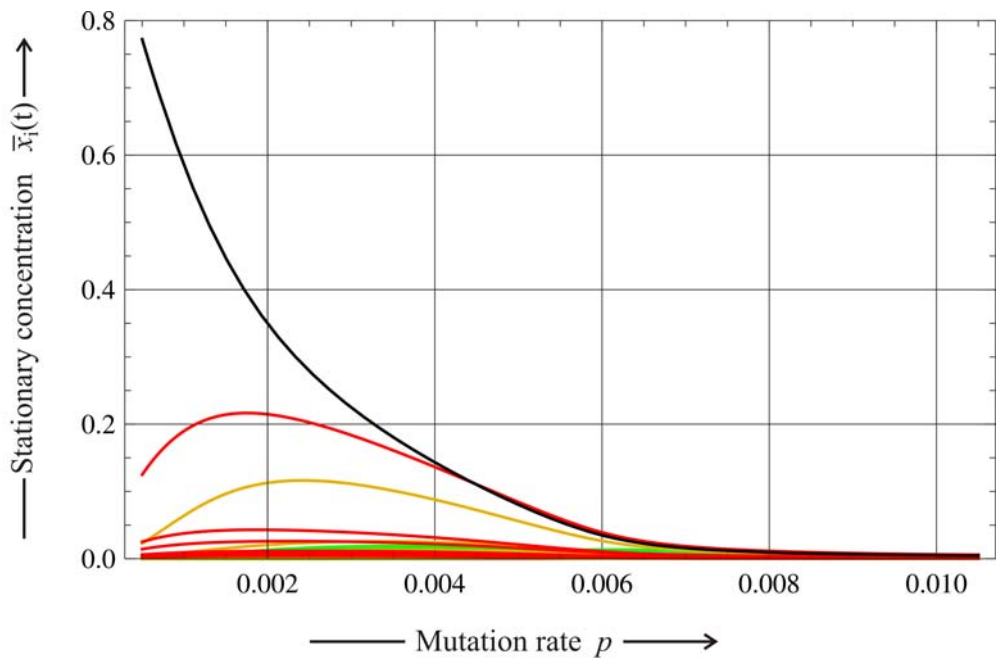




d = 0.100



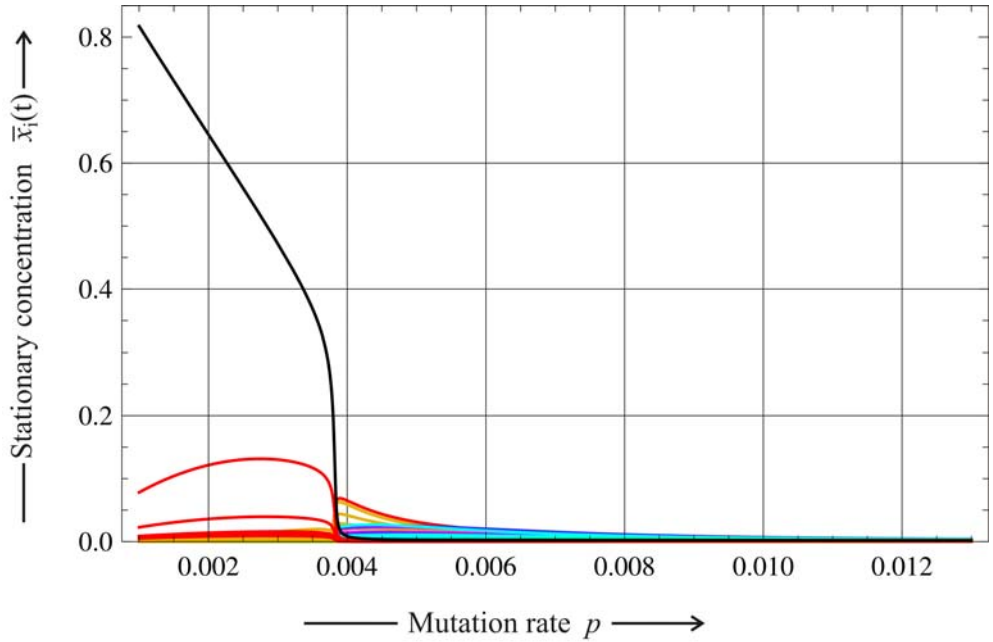
d = 0.200



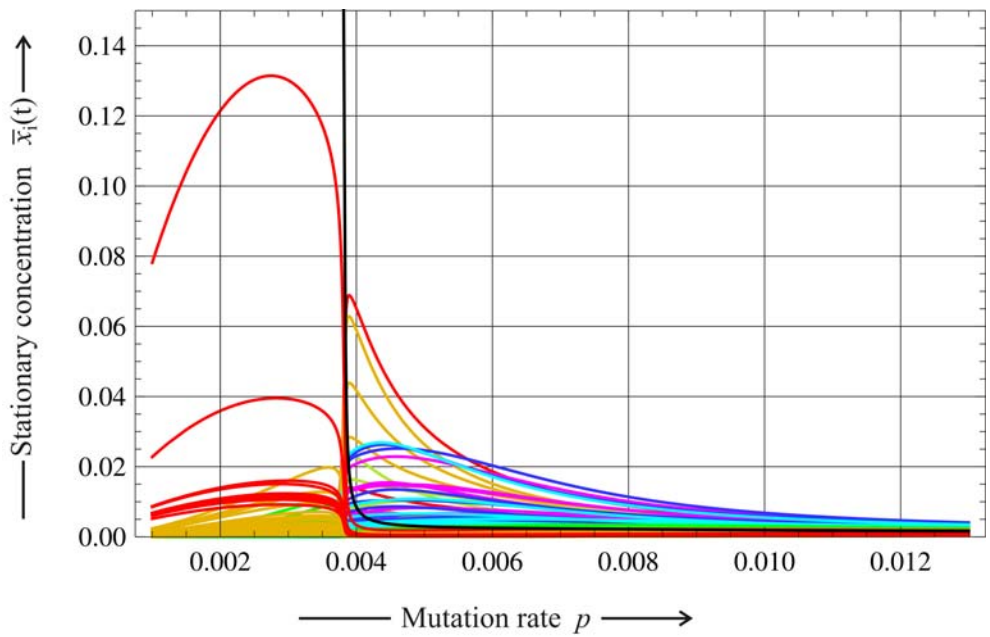
Case I: Strong Quasispecies

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

d = 0.190



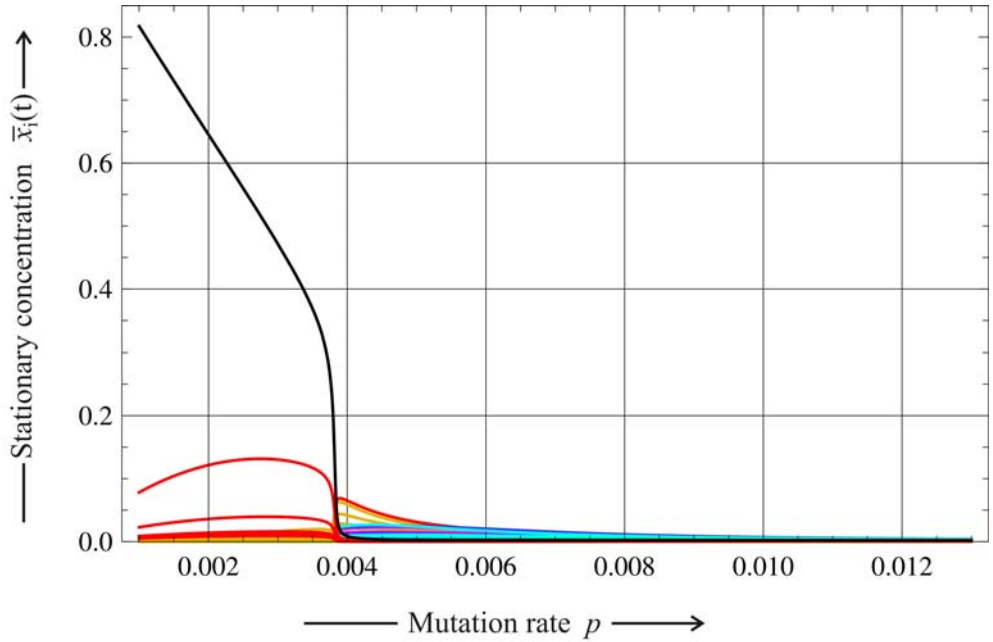
d = 0.190



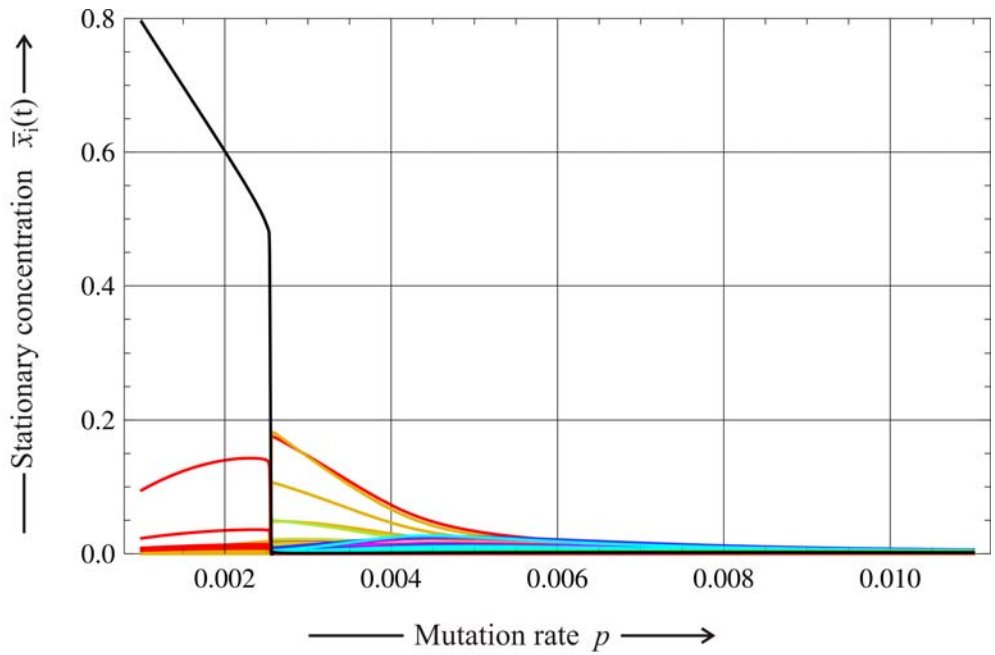
Case II: Dominant single transition

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 541

d = 0.190



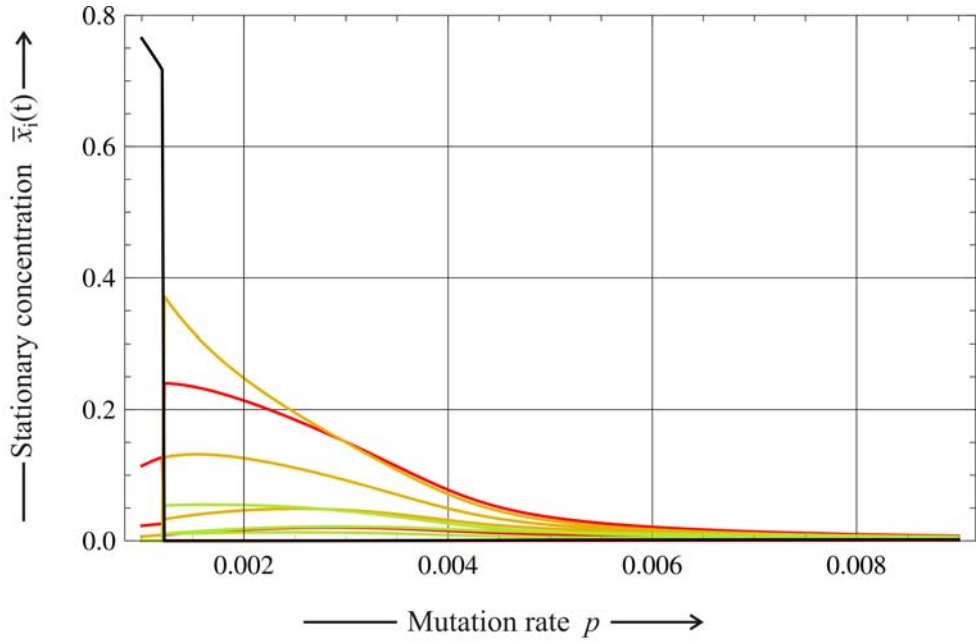
d = 0.195



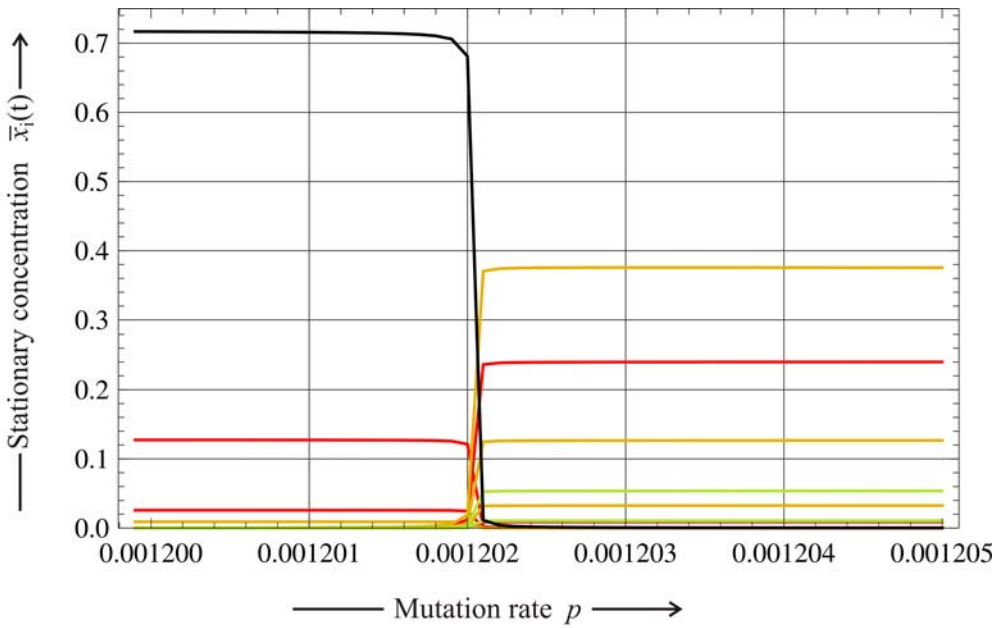
Case II: Dominant single transition

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 541

d = 0.199



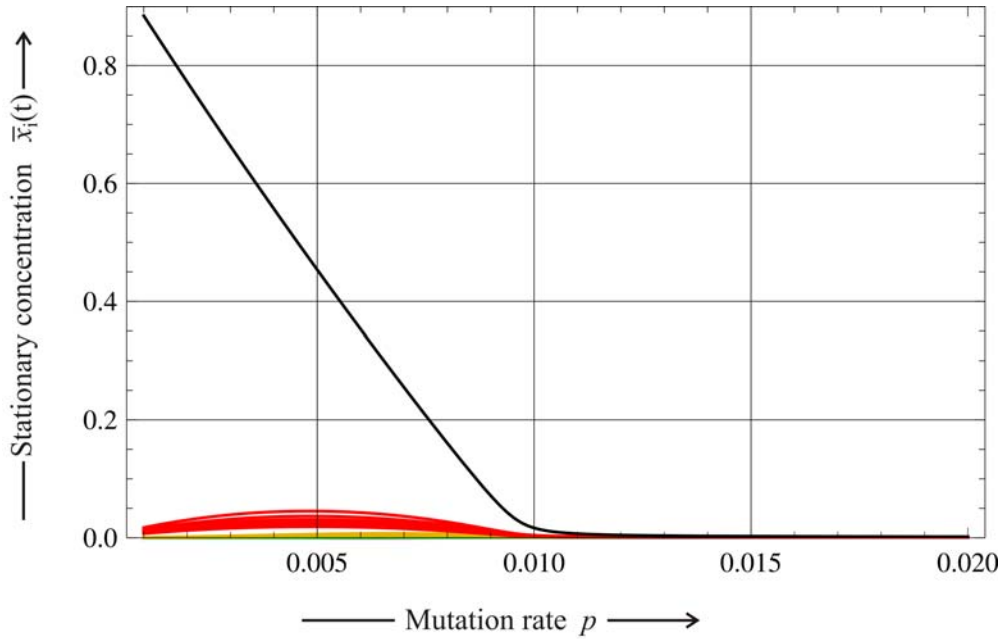
d = 0.199



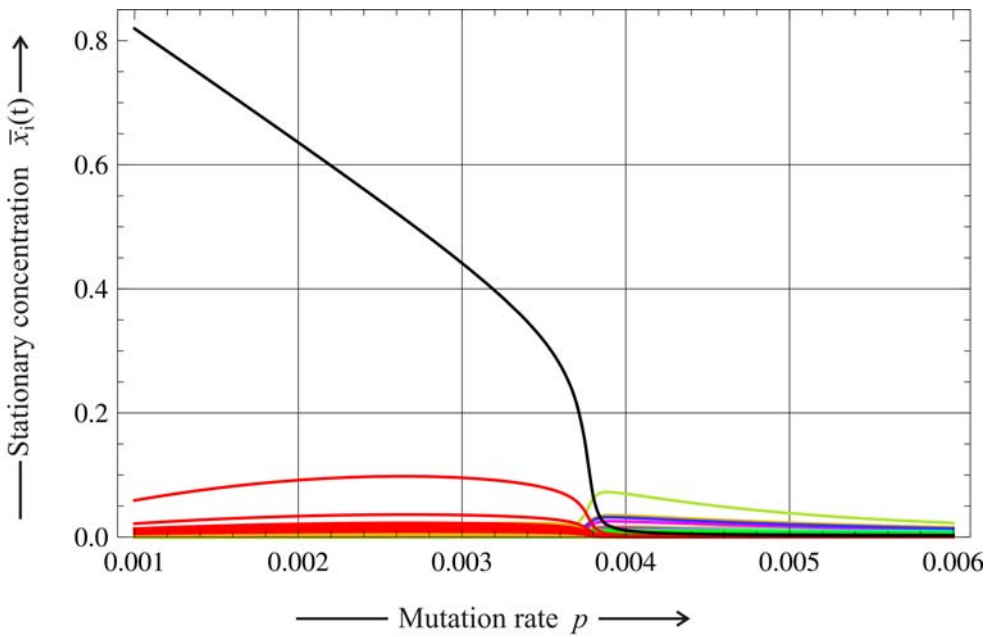
Case II: Dominant single transition

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 541

d = 0.100



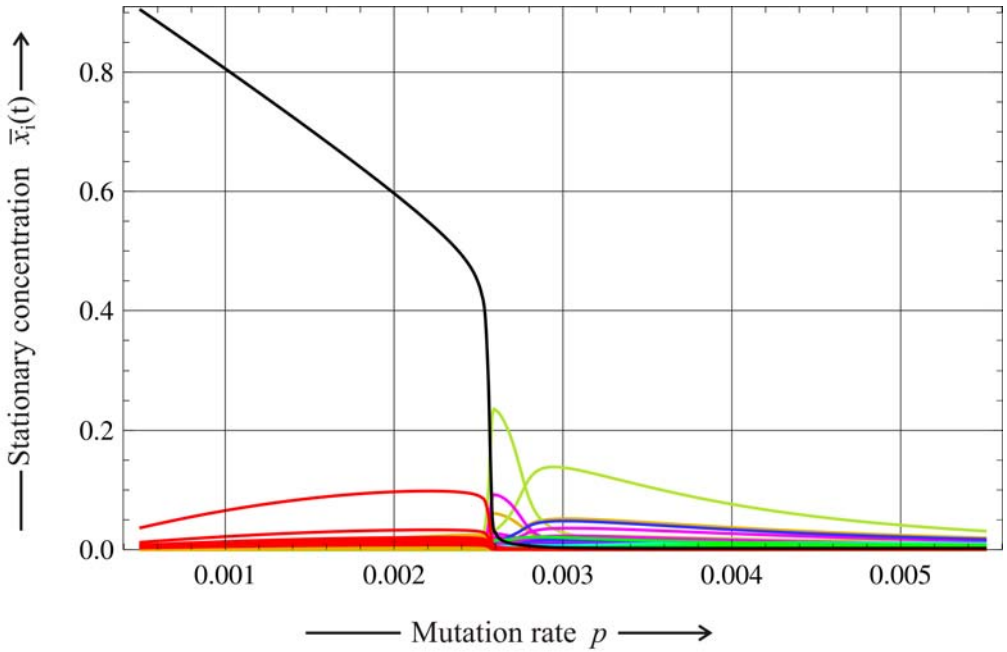
d = 0.195



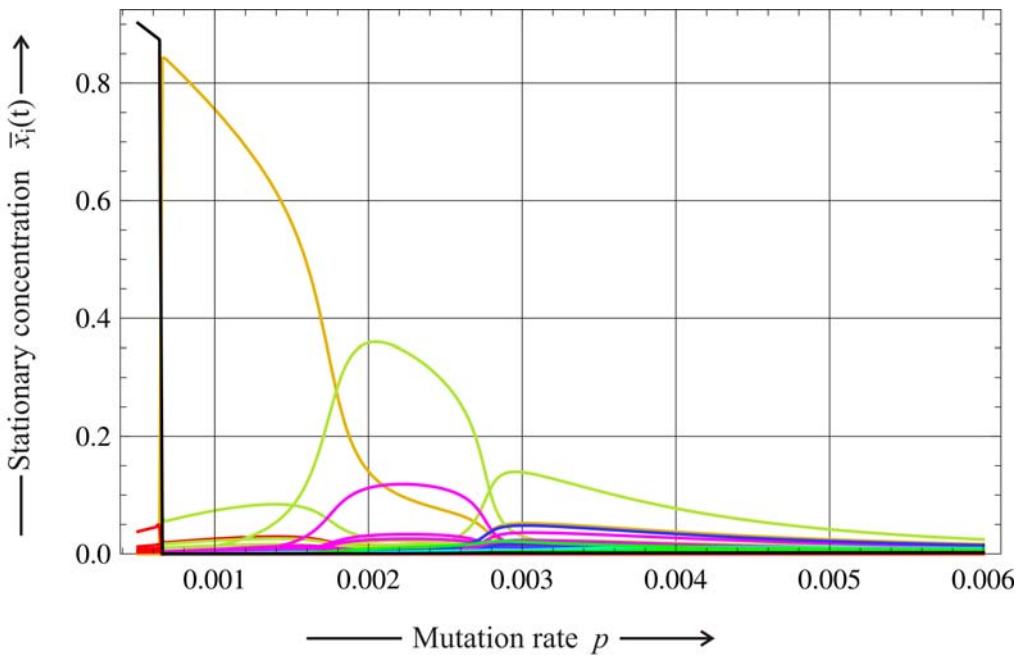
Case III: Multiple transitions

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

Fitness values of sequences and positions in the quasispecies

seed	class	no.	sequence	no.	fitness	neighborhood fitness
023	0	1	0	**1**	1.10000	1.01094 (0.90566 – 1.09036)
	1	1	512	**513**	1.09786	1.02411 (0.94456 – 1.10000)
	2	1	65	66	1.08449	0.99216 (0.90790 – 1.09215)
	2	2	576	577	1.08032	0.98794 (0.93090 – 1.09786)
	2	3	514	**515**	1.07903	1.02050 (0.92012 – 1.09786)
	6	1	627	628	1.09989	0.98010 (0.91012 – 1.07521)
	6	2	910	**911**	1.09850	1.04192 (0.92016 – 1.09642)
067	0	1	0	**1**	1.10000	0.99229 (0.92005 – 1.05296)
	1	1	2	**3**	1.05926	1.00795 (0.90287 – 1.10000)
	2	1	272	**273**	1.09740	0.99120 (0.91986 – 1.06725)
	5	1	109	**110**	1.09845	1.01442 (0.90510 – 1.09297)
401	0	1	0	**1**	1.10000	1.01094 (0.90566 – 1.09036)
	1	1	64	**65**	1.09812	1.00715 (0.90192 – 1.10000)
	2	1	80	**81**	1.09872	1.00132 (0.90458 – 1.09812)
541	0	1	0	**1**	1.10000	1.00134 (0.90880 – 1.07645)
	1	1	32	**33**	1.09322	1.01555 (0.96794 – 1.10000)
	2	1	384	385	1.09826	0.98012 (0.92088 – 1.06300)
	2	2	65	**66**	1.09803	1.03192 (0.96301 – 1.09402)
	8	1	1017	**1018**	1.09967	1.01796 (0.94269 – 1.09776)
	9	1	1021	**1022**	1.09776	1.02221 (0.91501 – 1.09967)

1. Chemical kinetics of replication and mutation
2. Complexity of fitness landscapes
3. Quasispecies on realistic landscapes
4. **Neutrality and replication**

MOTOO KIMURA

Motoo Kimuras population genetics of neutral evolution.

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

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

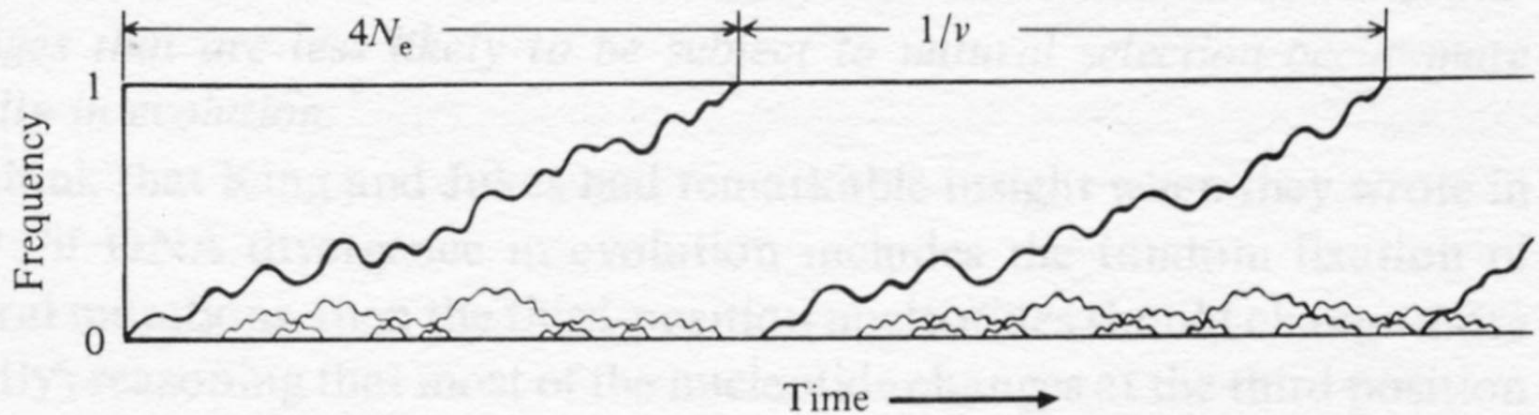
THE NEUTRAL THEORY
OF MOLECULAR EVOLUTION

MOTOO KIMURA
National Institute of Genetics, Japan

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

CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

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

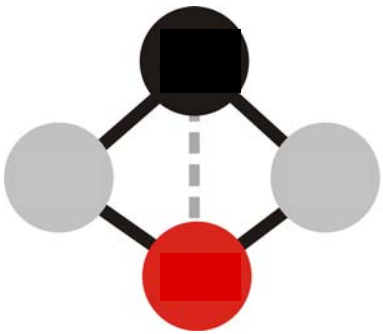


Motoo Kimura

Is the Kimura scenario correct for frequent mutations?

$\mathbf{d_H = 1}$

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



$\mathbf{d_H = 2}$

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

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

$\mathbf{d_H \geq 3}$

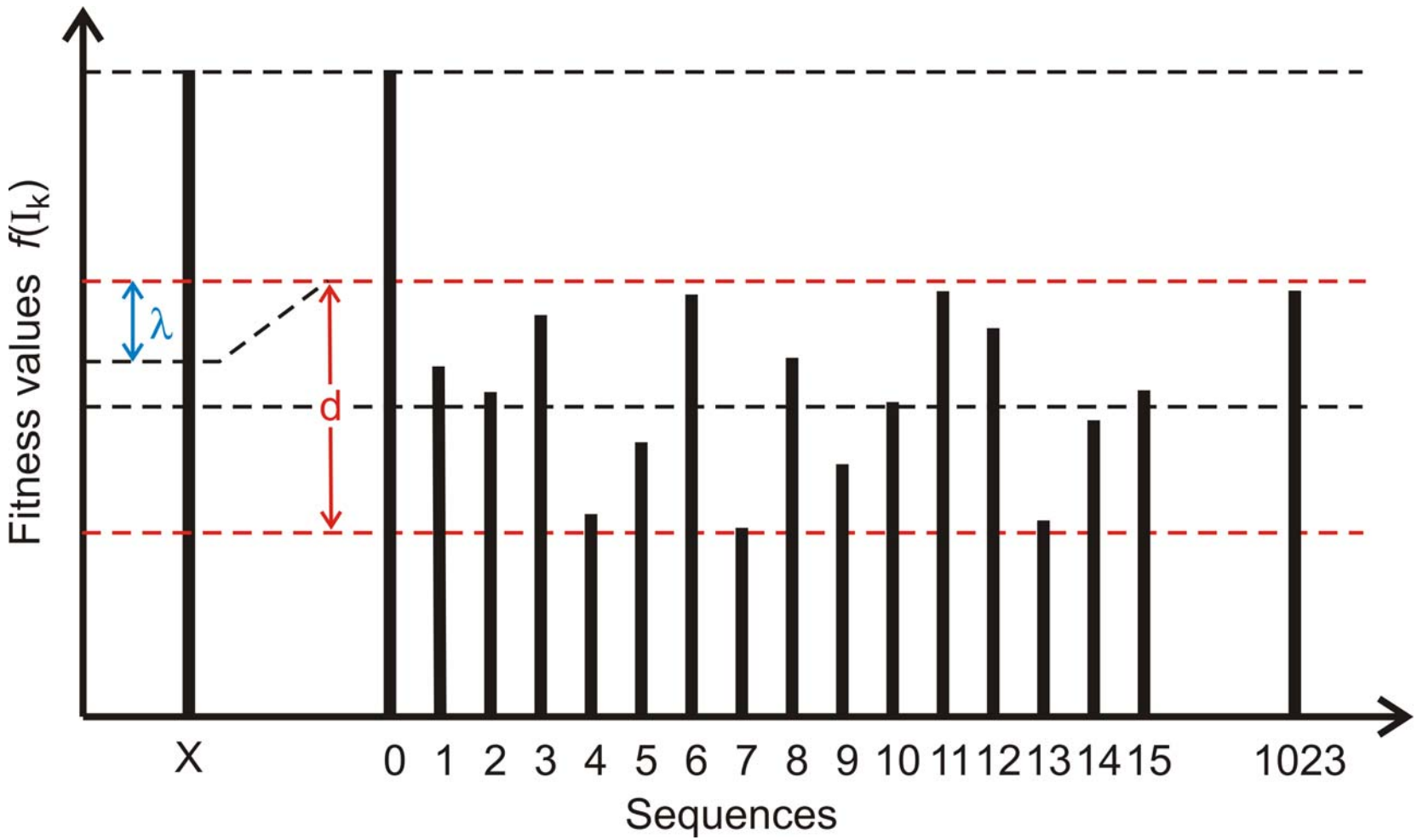
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

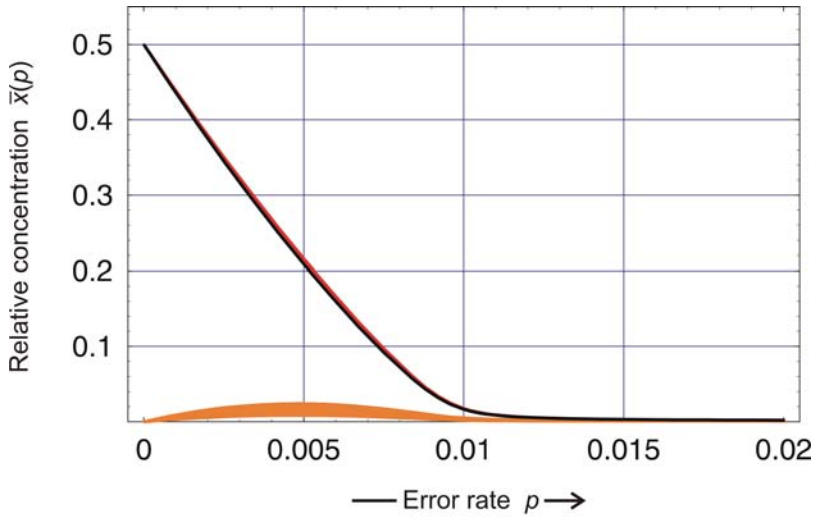
Pairs of neutral sequences in replication networks

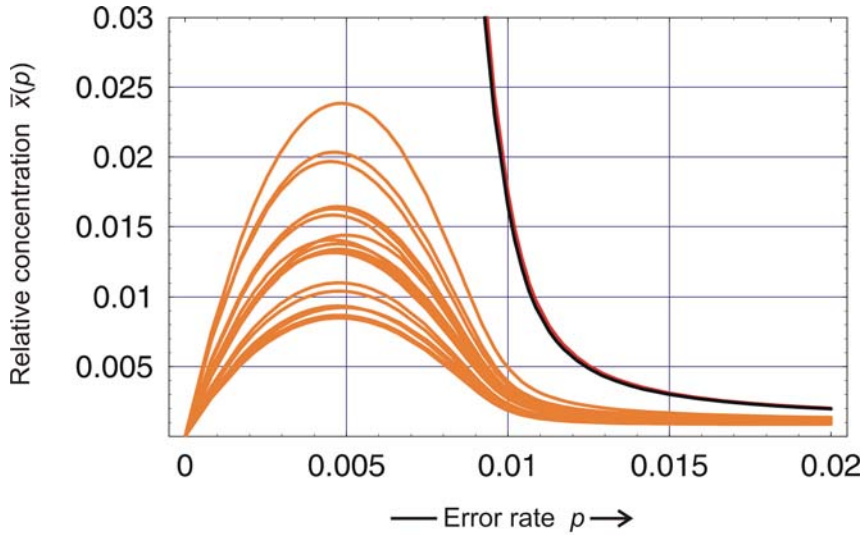
Random fixation in the sense of Motoo Kimura

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



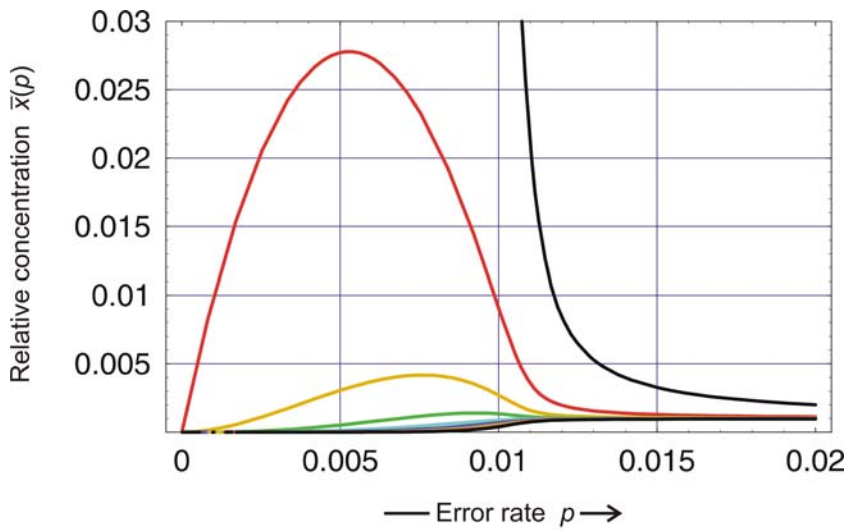
A fitness landscape including neutrality





Neutral network

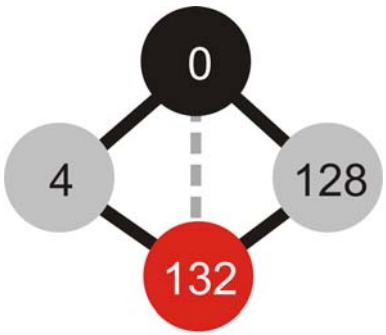
$\lambda = 0.01,\ s = 367$



Neutral network: Individual sequences

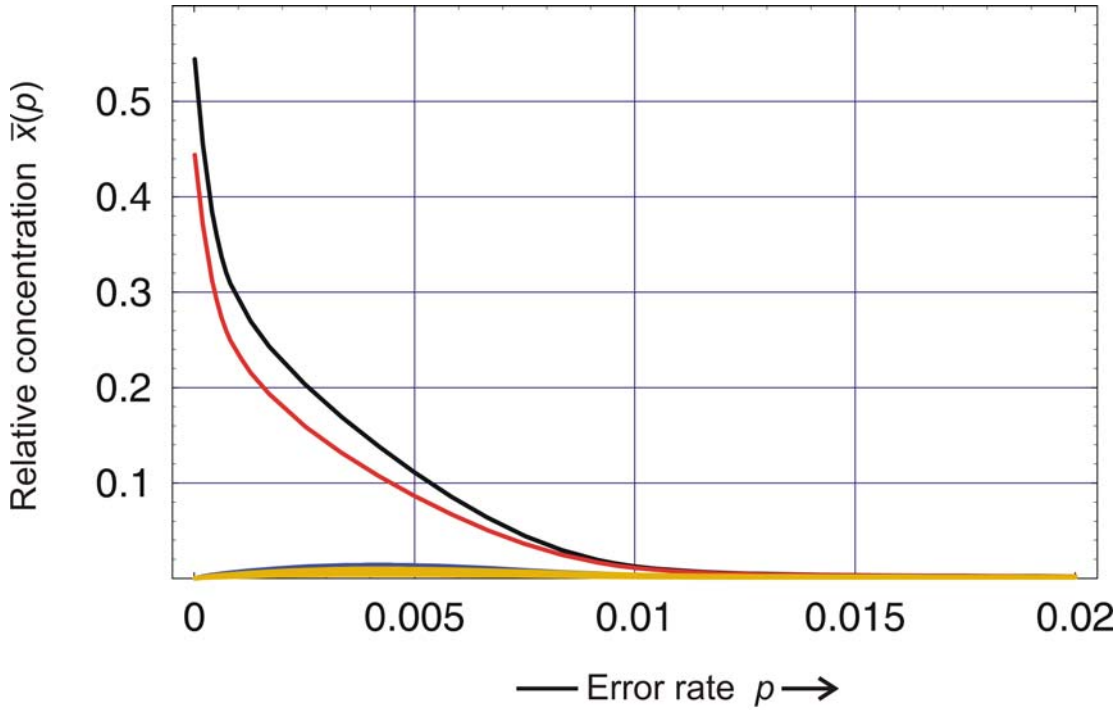
$n = 10,\ \sigma = 1.1,\ d = 1.0$

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



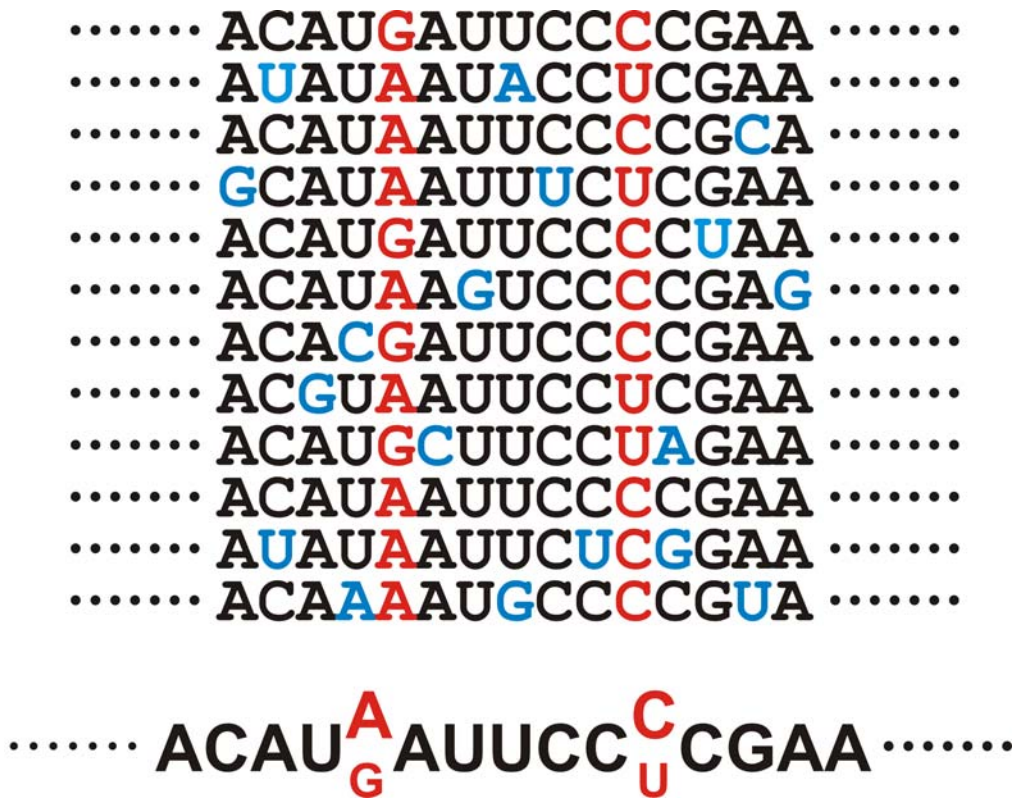
Neutral network

$\lambda = 0.01,\ s = 877$

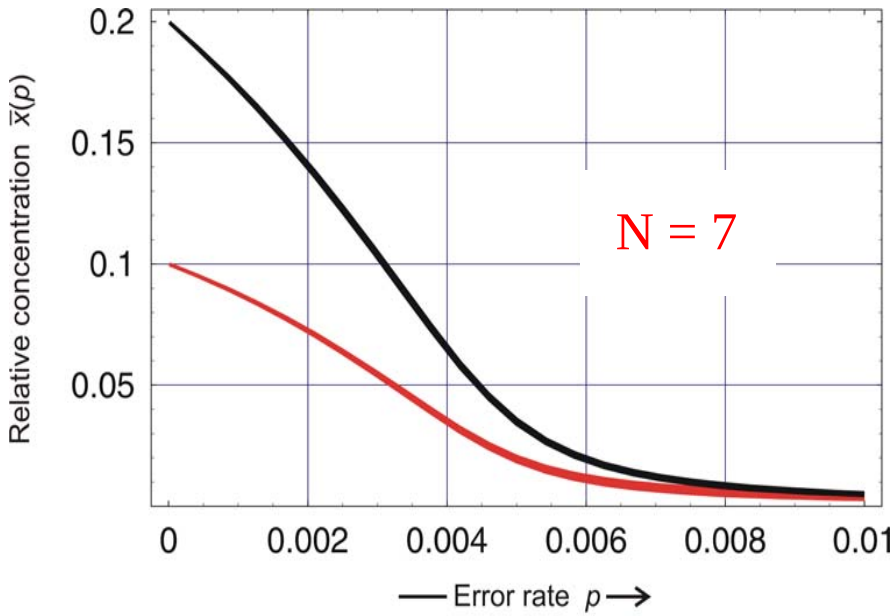


Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0

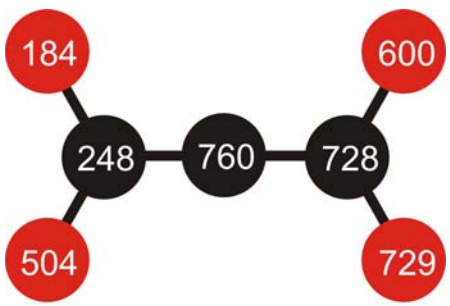


Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 2$.



Perturbation matrix W

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



Neutral network

λ = 0.10, s = 229

Adjacency matrix

Largest eigenvector of W

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

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

Coworkers

Universität Wien

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

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

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

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

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

Christian Forst, Los Alamos National Laboratory, NM

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

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

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

Acknowledgement of support

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

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

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

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

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

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Universität Wien

Thank you for your attention!

Web-Page for further information:

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