Origin of catalytic cycles

Peter Schuster

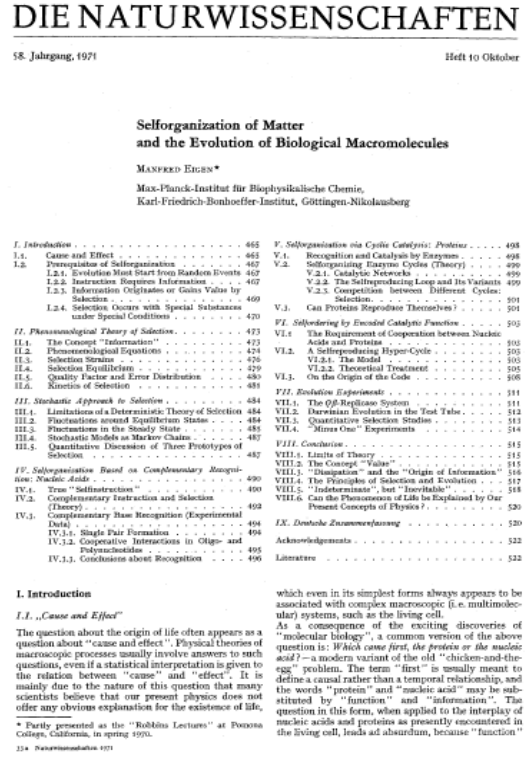
Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Open Questions on the Origin of Life

San Sebastian, 20.– 23.05.2009

Web-Page for further information:

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



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 465
I.1. Cause and Effect 465
I.2. Prerequisites of Selforganization 467
I.2.1. Evolution Must Start from Random Events 467
I.2.2. Instruction Requires Information 467
I.2.3. Information Originates or Gains Value by Selection 469
I.2.4. Selection Occurs with Special Substances under Special Conditions 470
II. Phenomenological Theory of Selection 473
II.1. The Concept "Information" 473
II.2. Phenomenological Equations 474
II.3. Selection Strains 476
II.4. Selection Equilibrium 479
II.5. Quality Factor and Error Distribution 480
II.6. Kinetics of Selection 481
III. Stochastic Approach to Selection 484
III.1. Limitations of a Deterministic Theory of Selection 484
III.2. Fluctuations around Equilibrium States 484
III.3. Fluctuations in the Steady State 485
III.4. Stochastic Models as Markov Chains 487
III.5. Quantitative Discussion of Three Prototypes of Selection 487
IV. Selforganization Based on Complementary Recognition: Nucleic Acids 490
IV.1. Term "Selfinstruction" 490
IV.2. Complementary Instruction and Selection (Theory) 492
IV.3. Complementary Base Recognition (Experimental Data) 494
IV.3.1. Single Pair Formation 494
IV.3.2. Cooperative Interactions in Oligo- and Polynucleotides 495
IV.3.3. Conclusions about Recognition 496
V. Selforganization via Cyclic Catalysis: Proteins 498
V.1. Recognition and Catalysis by Enzymes 498
V.2. Selforganizing Enzyme Cycles (Theory) 499
V.2.1. Catalytic Networks 499
V.2.2. The Selfreproducing Loop and Its Variants 499
V.2.3. Competition between Different Cycles: Selection. 501
V.3. Can Proteins Reproduce Themselves? 501
VI. Selfordering by Encoded Catalytic Function 503
VI.1. The Requirement of Cooperation between Nucleic Acids and Proteins 503
VI.2. A Selfreproducing Hyper-Cycle 503
VI.2.1. The Model 503
VI.2.2. Theoretical Treatment 505
VI.3. On the Origin of the Code 508
VII. Evolution Experiments 511
VII.1. The Qβ-Replicase System 511
VII.2. Darwinian Evolution in the Test Tube 512
VII.3. Quantitative Selection Studies 513
VII.4. "Mirror One" Experiments 514
VIII. Conclusion 515
VIII.1. Limits of Theory 515
VIII.2. The Concept "Value" 515
VIII.3. "Dissipation" and the "Origin of Information" 516
VIII.4. The Principles of Selection and Evolution . . . 517
VIII.5. "Indeterminate", but "Inevitable" 518
VIII.6. Can the Phenomenon of Life be Explained by Our Present Concepts of Physics? 520
IX. Deutsche Zusammenfassung 520
Acknowledgements 522
Literature 522

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

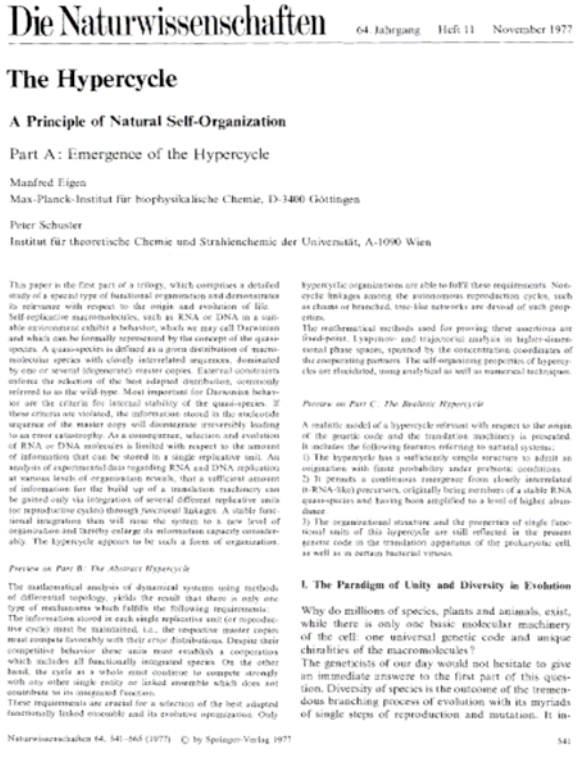
1. Introduction

1.1. „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"

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 behavior, 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-cycle 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 interrelated (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

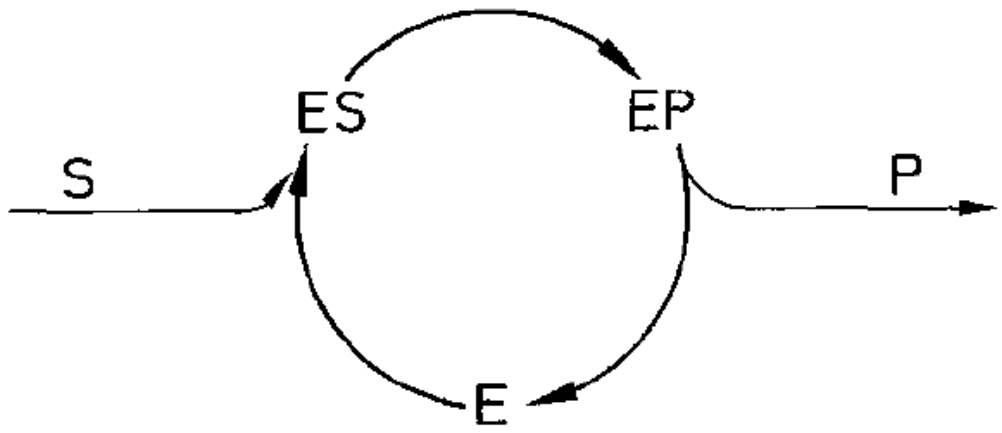
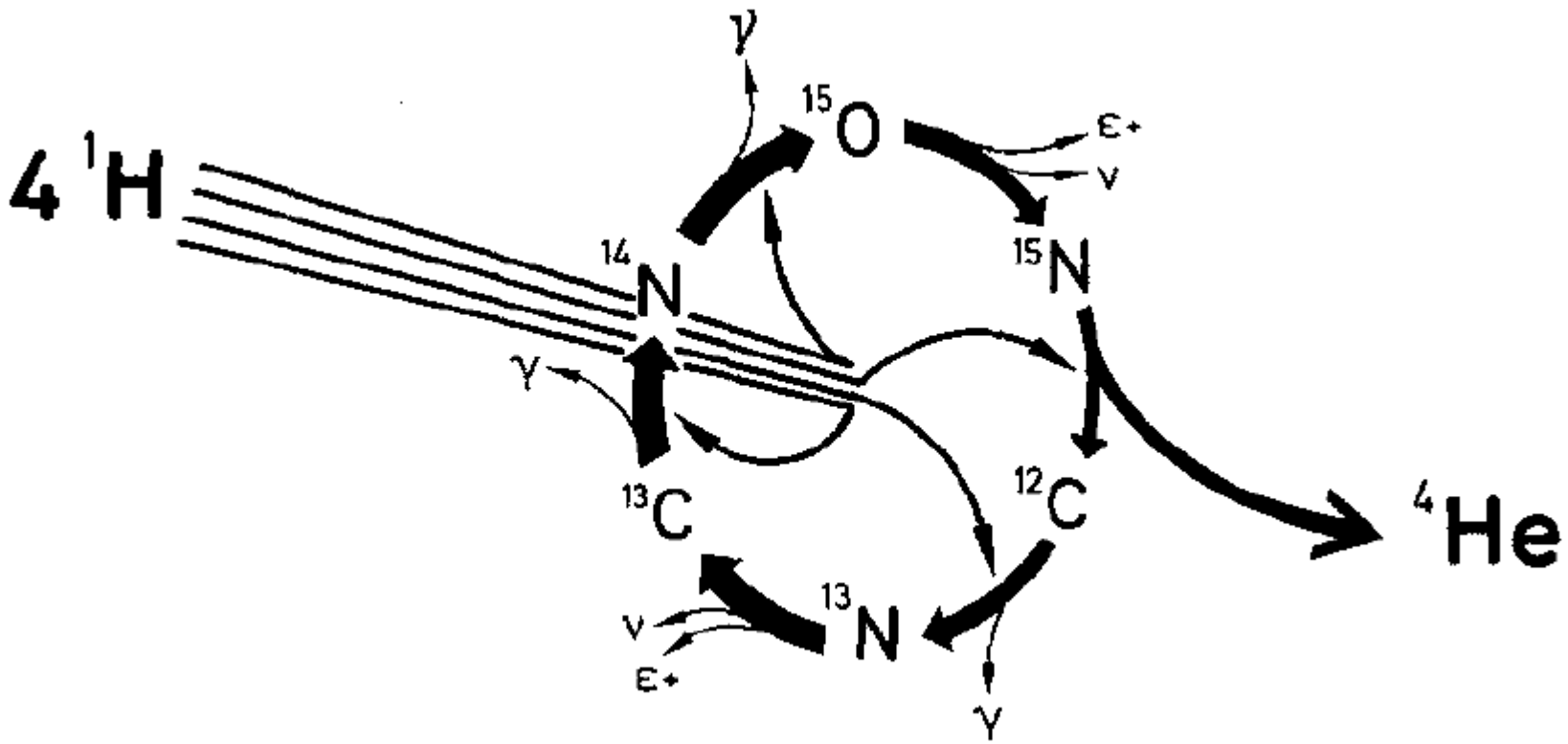
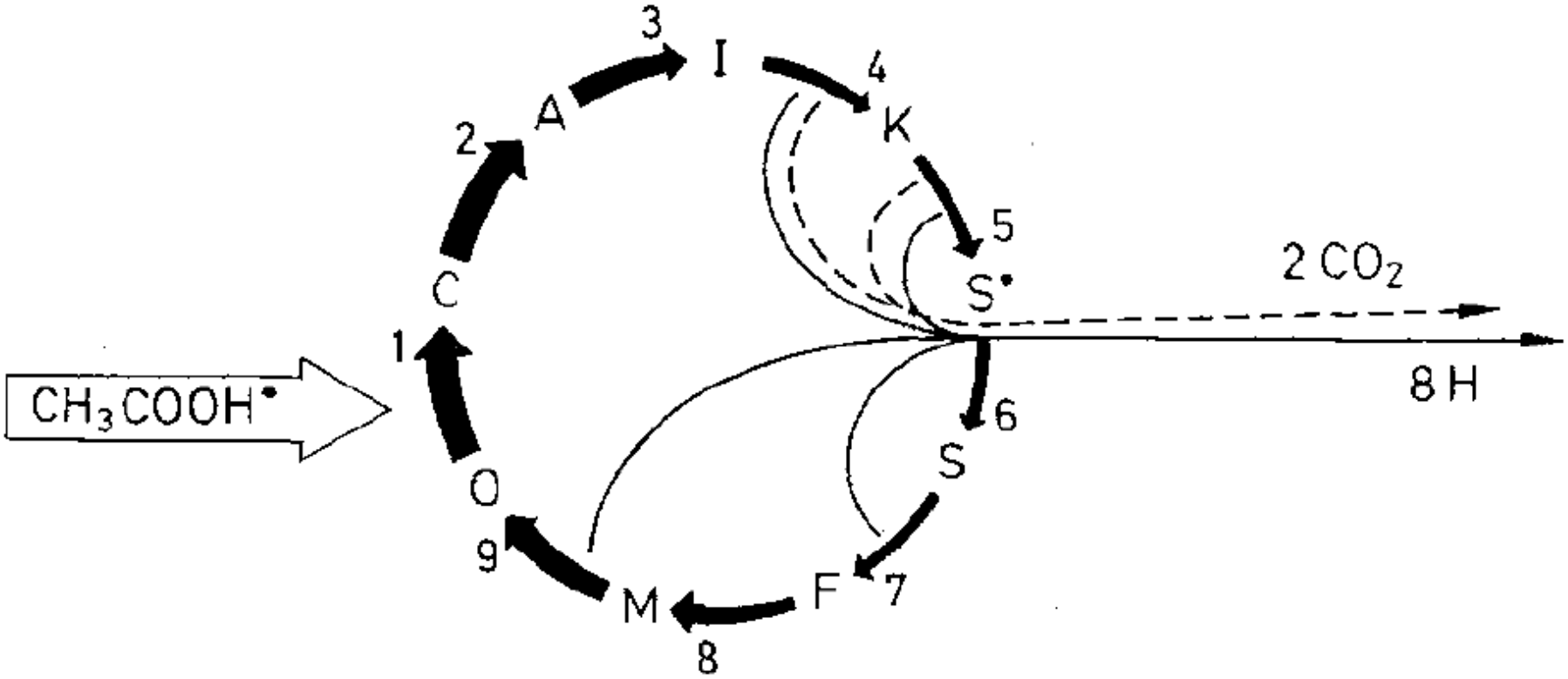


Fig. 1. *The common catalytic mechanism of an enzyme* according to Michaelis and Menten involves (at least) three intermediates: the free enzyme (E), the enzyme-substrate (ES) and the enzyme-product complex (EP). The scheme demonstrates the equivalence of catalytic action of the enzyme and cyclic restoration of the intermediates in the turnover of the substrate (S) to the product (P). Yet, it provides only a formal representation of the true mechanism which may involve a stepwise activation of the substrate as well as induced conformation changes of the enzyme.

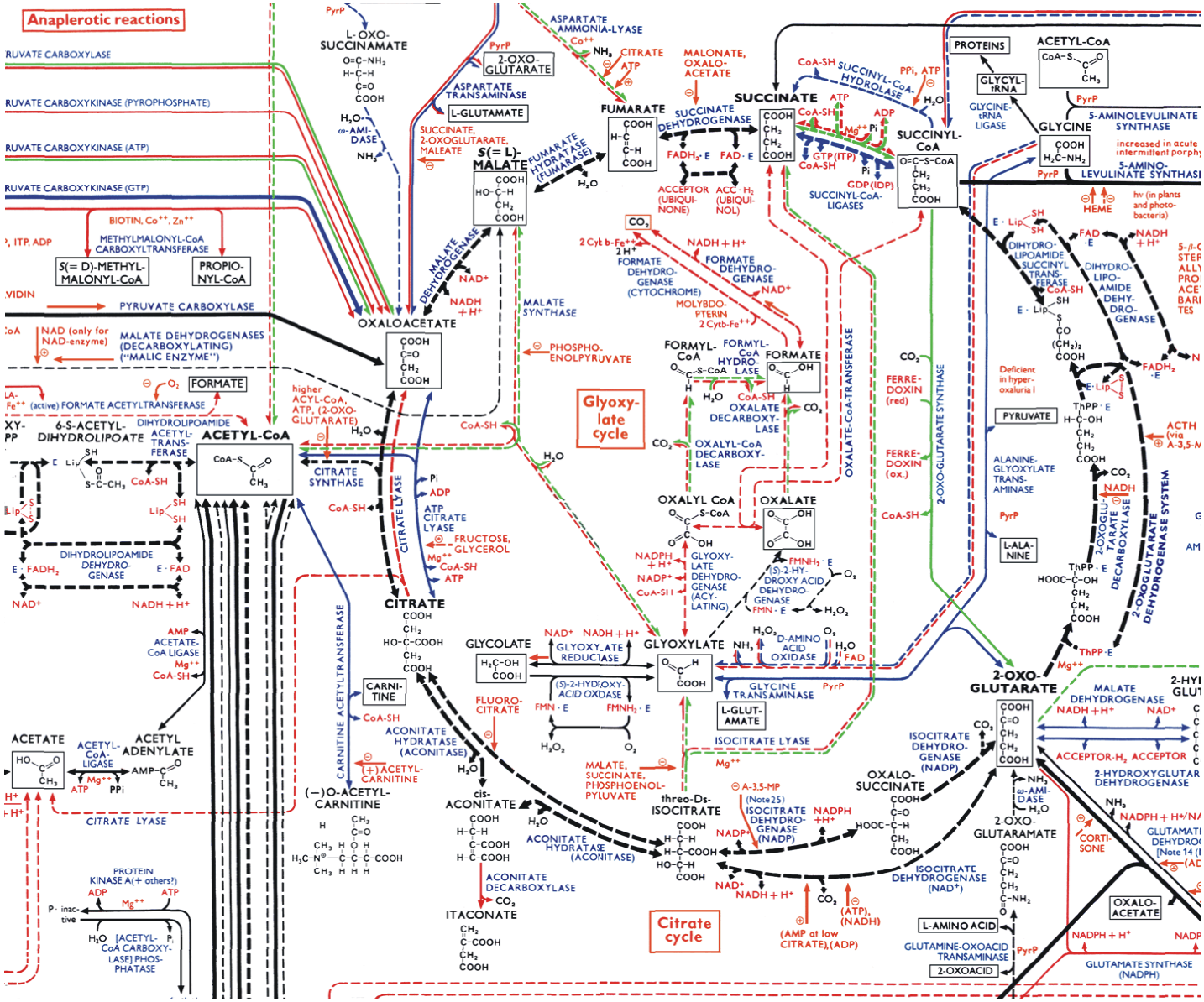


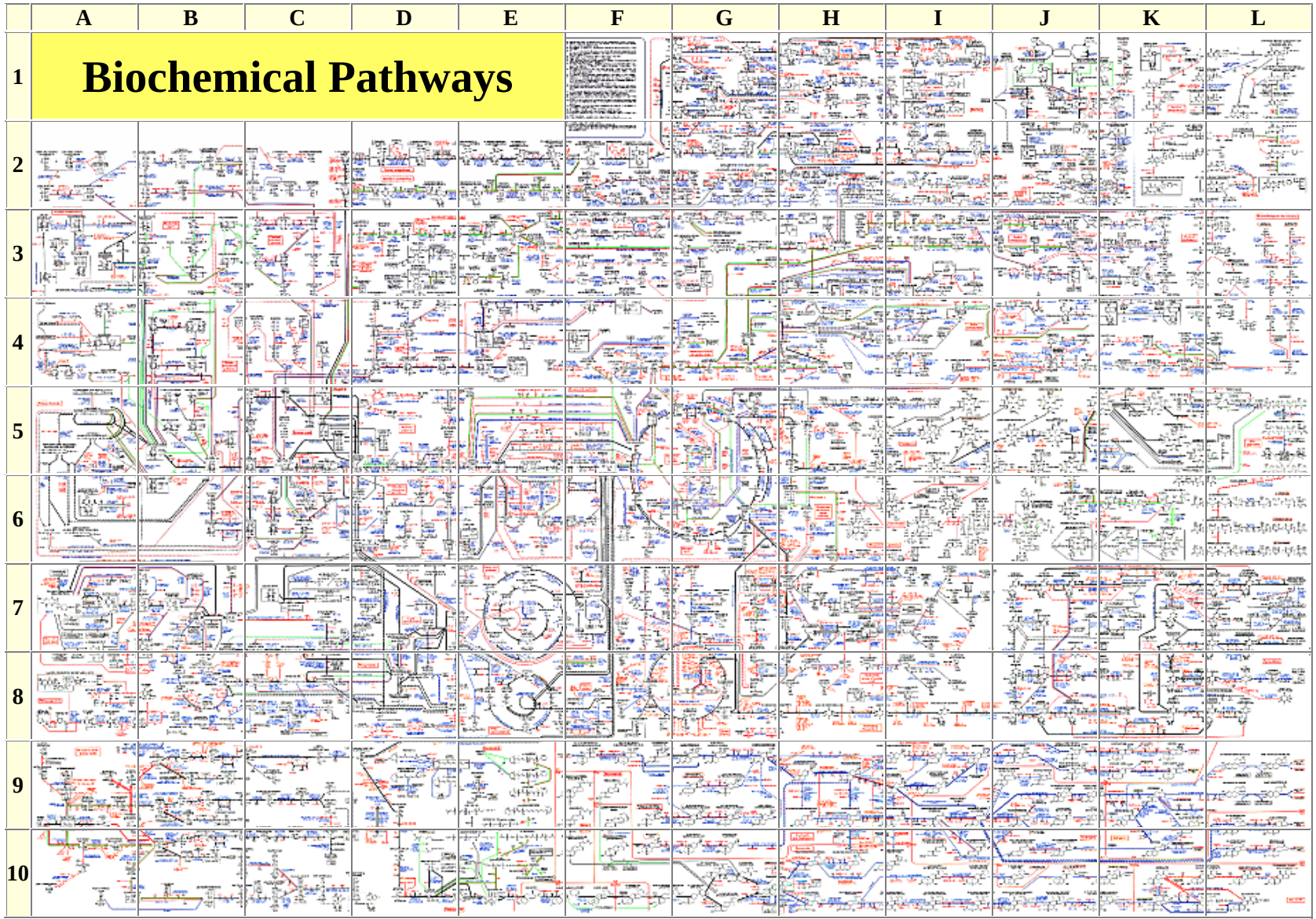
The Bethe - vonWeizsäcker catalytic cycle ist responsible – in part – for the energy production in massive stars.



The tricarboxylic acid or citric acid cycle is fuelling the metabolic reactions of the cell.

The citric acid or Krebs cycle (enlarged from previous slide).





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

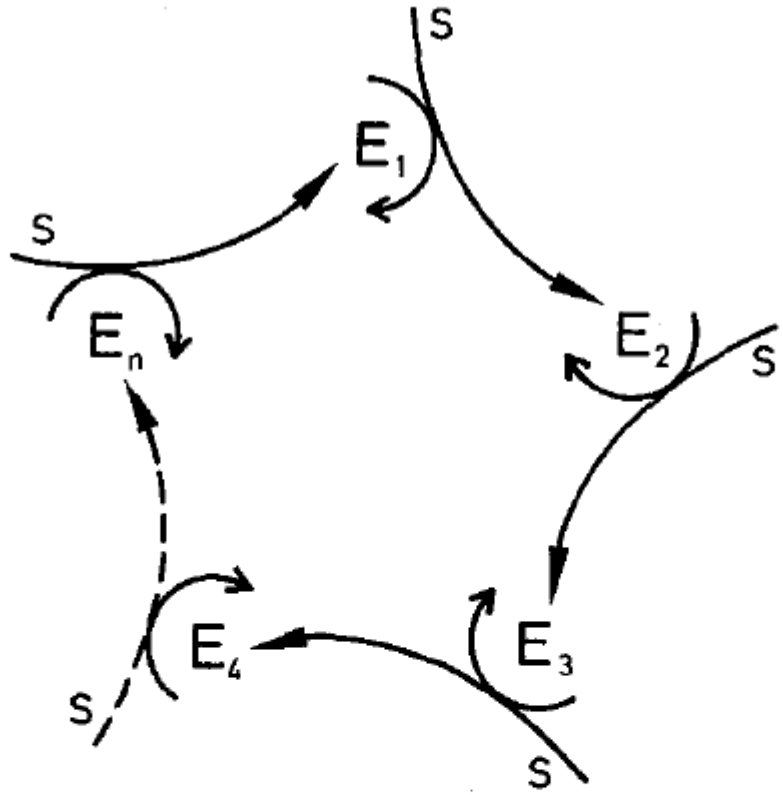
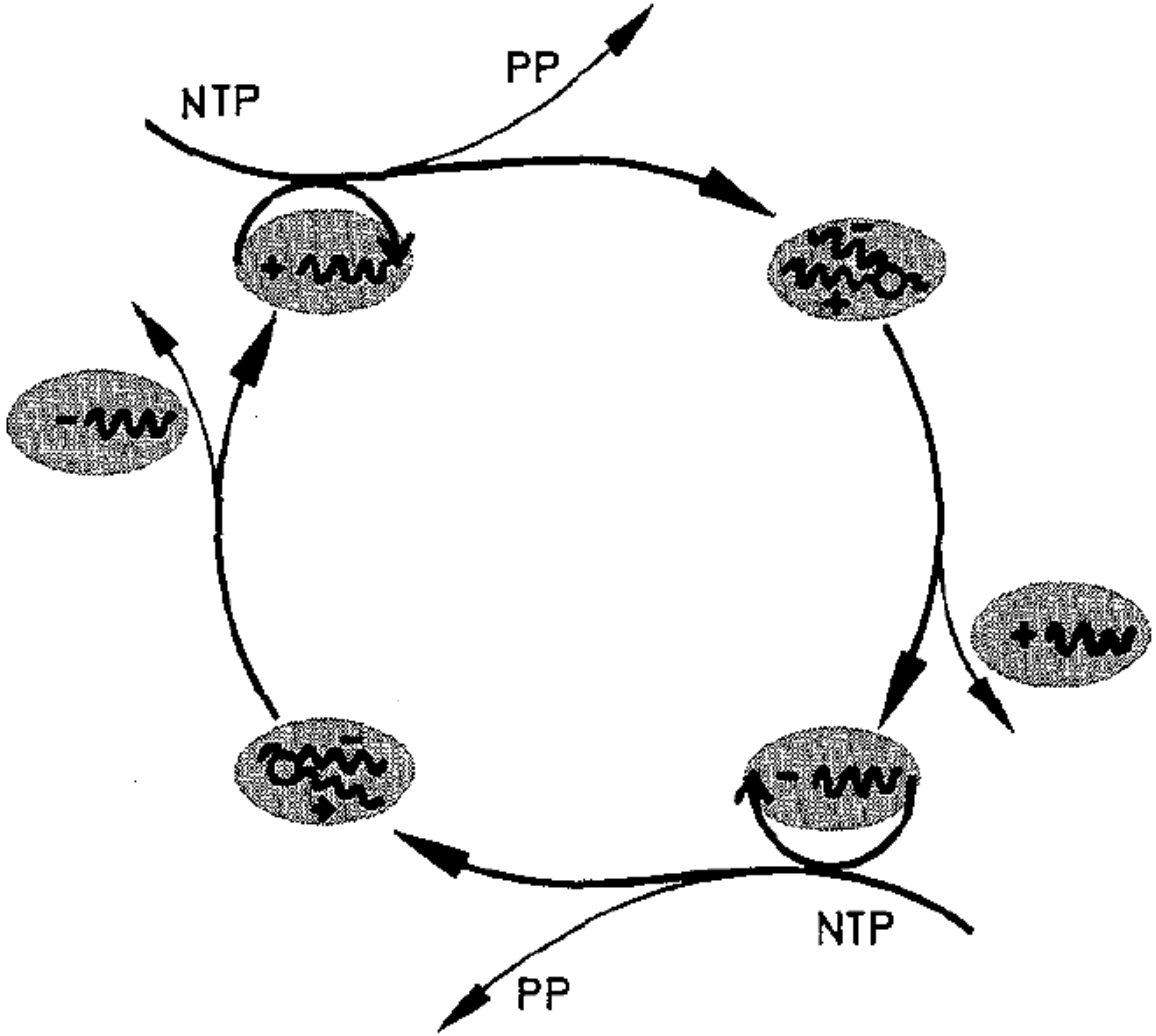


Fig. 4. *The catalytic cycle* represents a higher level of organization in the hierarchy of catalytic schemes. The constituents of the cycle $E_1 \rightarrow E_n$ are themselves catalysts which are formed from some energy-rich substrates (S), whereby each intermediate E_i is a catalyst for the formation of E_{i+1}. The catalytic cycle seen as an entity is equivalent to an autocatalyst, which instructs its own reproduction. To be a catalytic cycle it is sufficient, that only one of the intermediates formed is a catalyst for one of the subsequent reaction steps.



Complementary (±) replication of RNA as an example of an autocatalytic cycle.

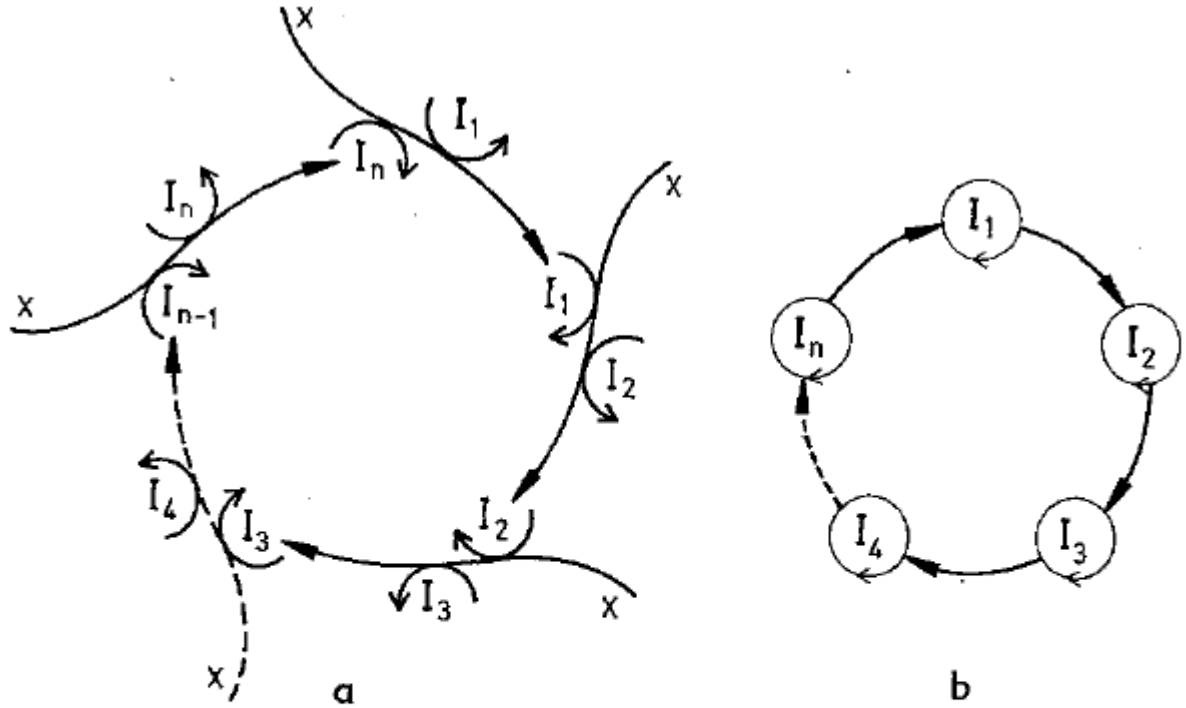
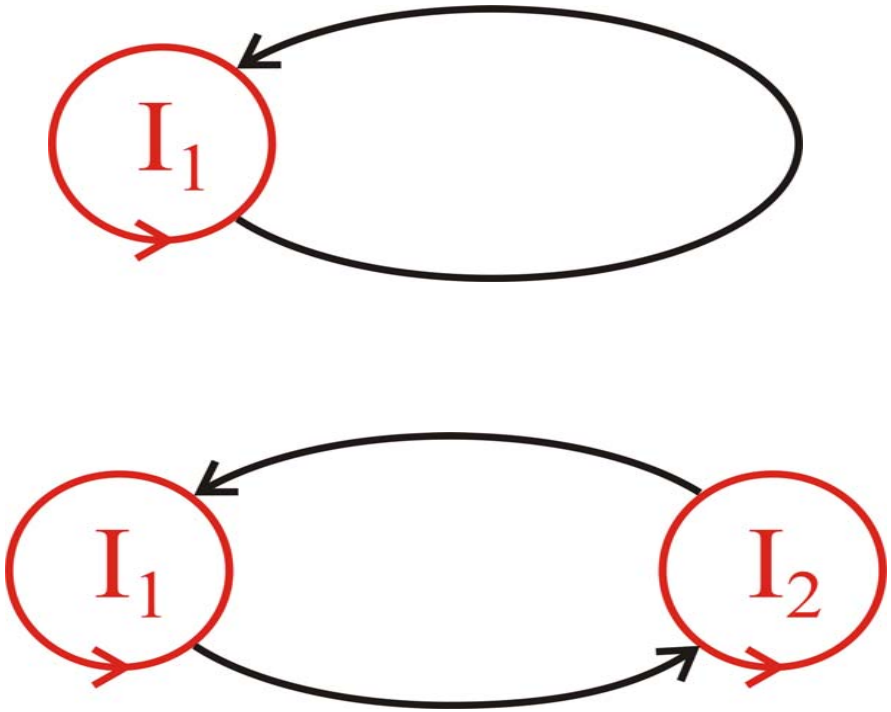
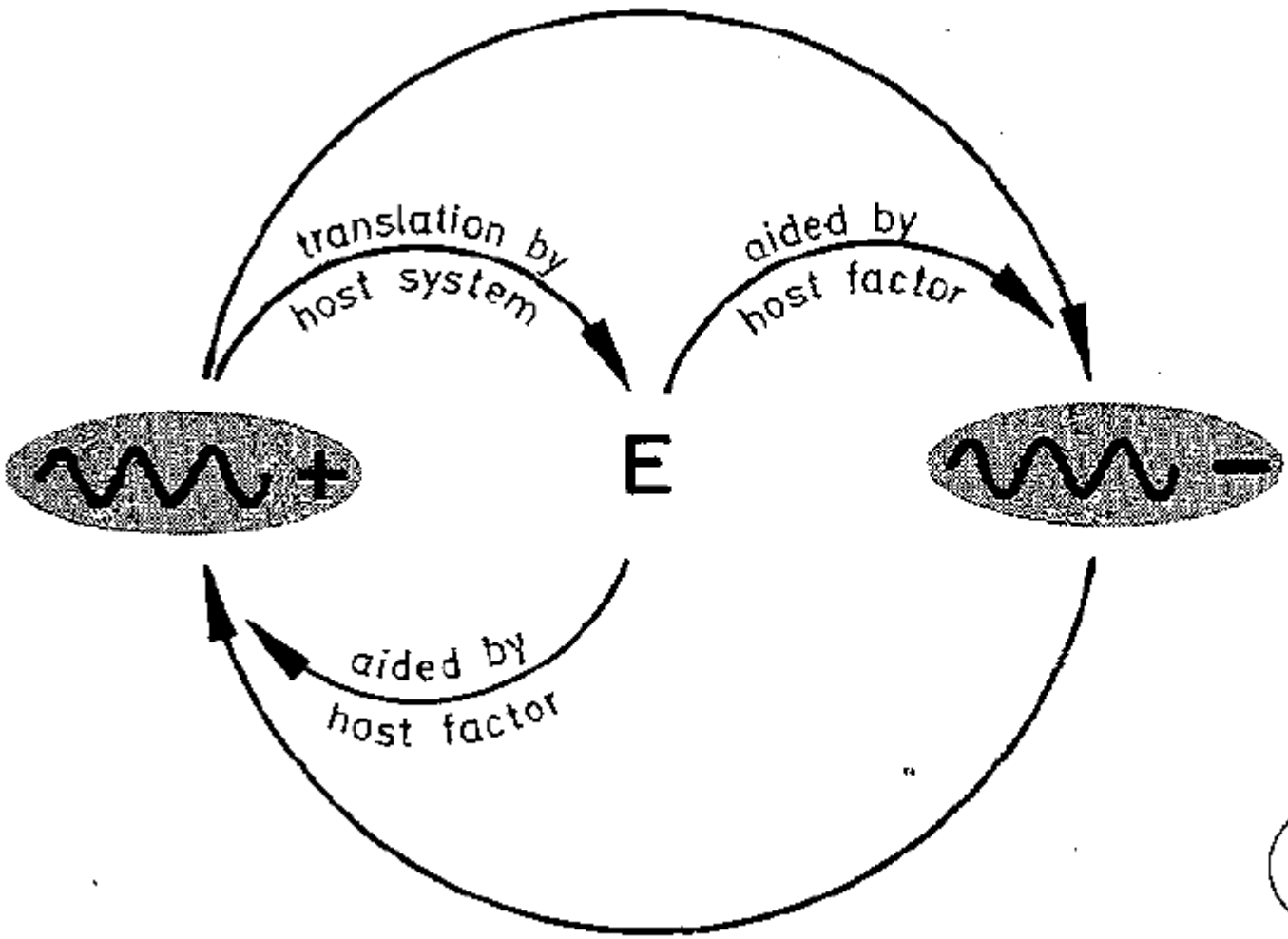


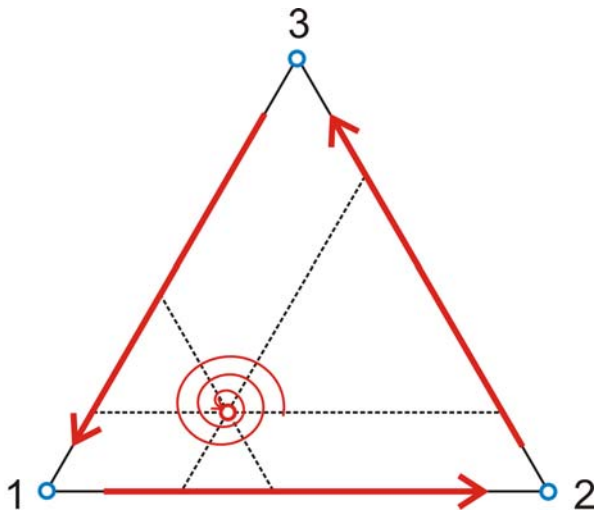
Fig. 7. *A catalytic hypercycle* consists of self-instructive units I_i with two-fold catalytic functions. As autocatalysts or—more generally—as catalytic cycles the intermediates I_i are able to instruct their own reproduction and, in addition, provide catalytic support for the reproduction of the subsequent intermediate (using the energy-rich building material X). The simplified graph (b) indicates the cyclic hierarchy

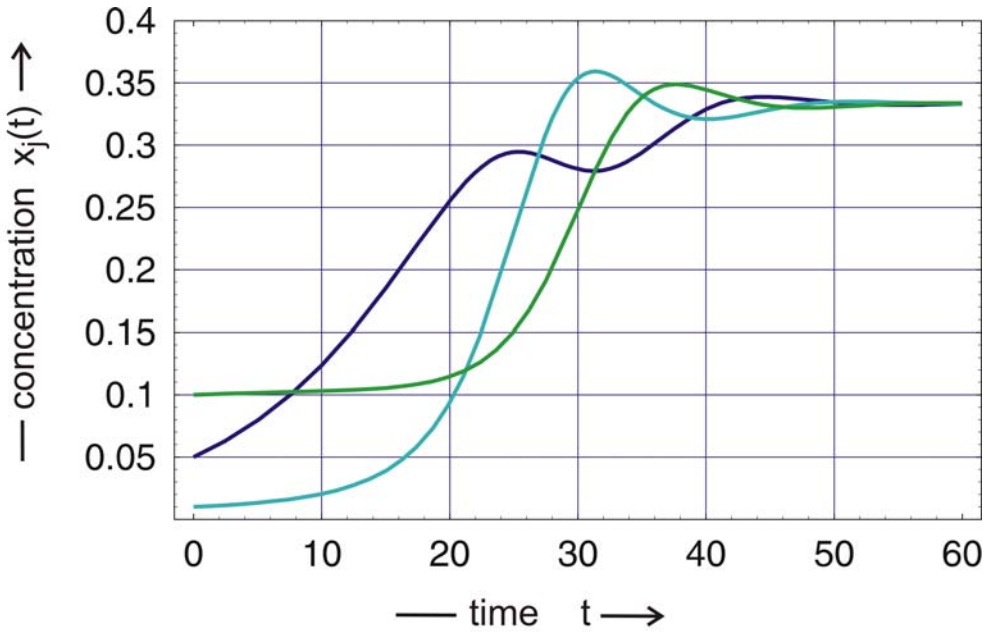


Hypercycles with one and two members are common in nature.

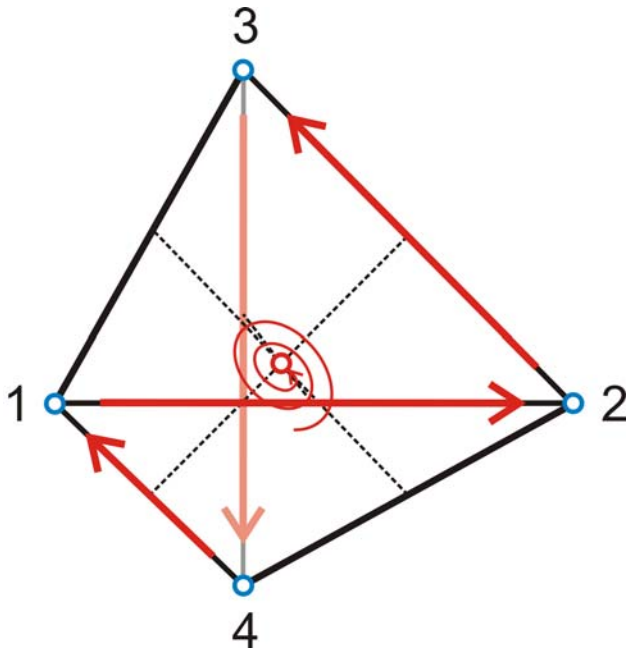

translation by
host system
aided by
host factor
+
E
−
aided by
host factor

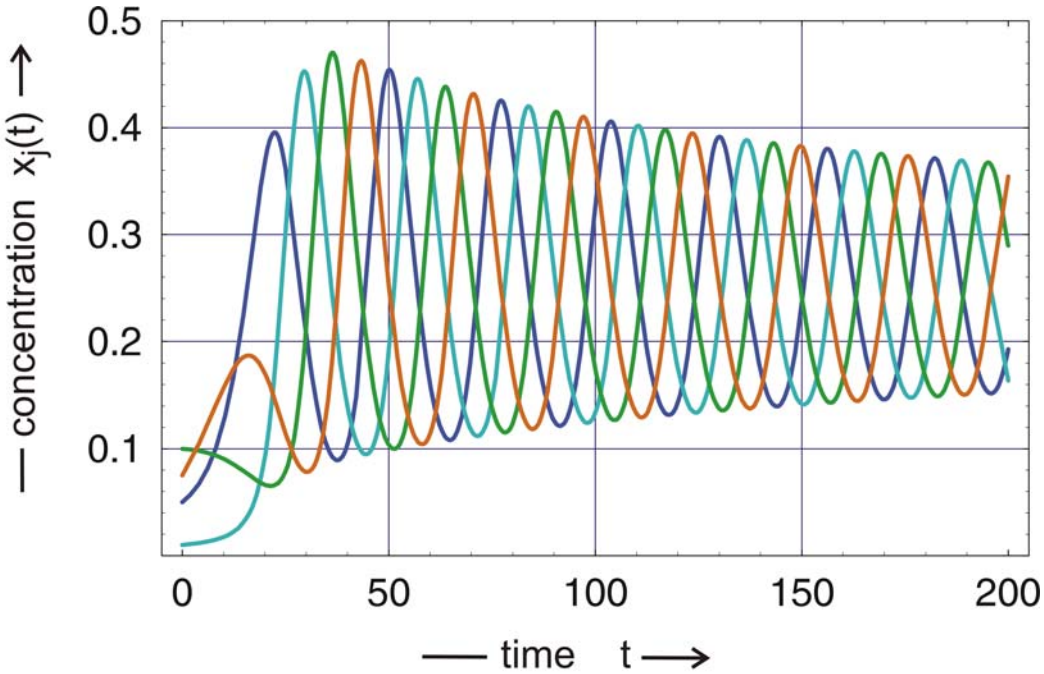

I
E



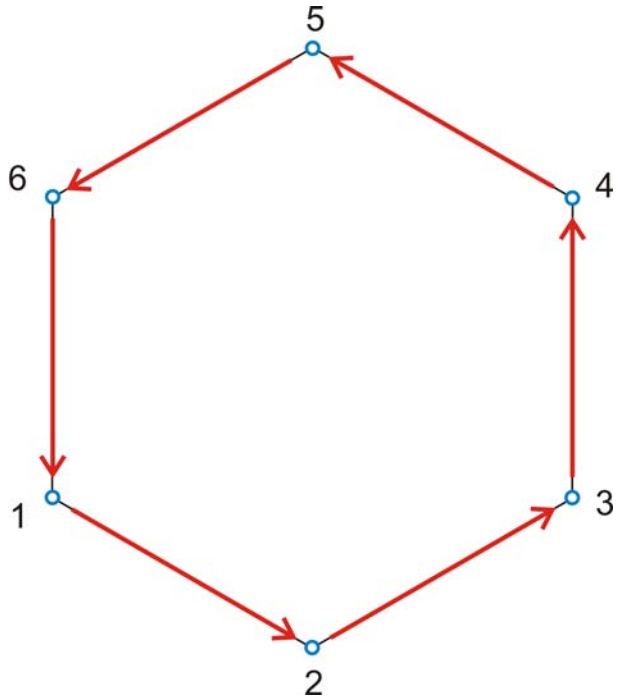


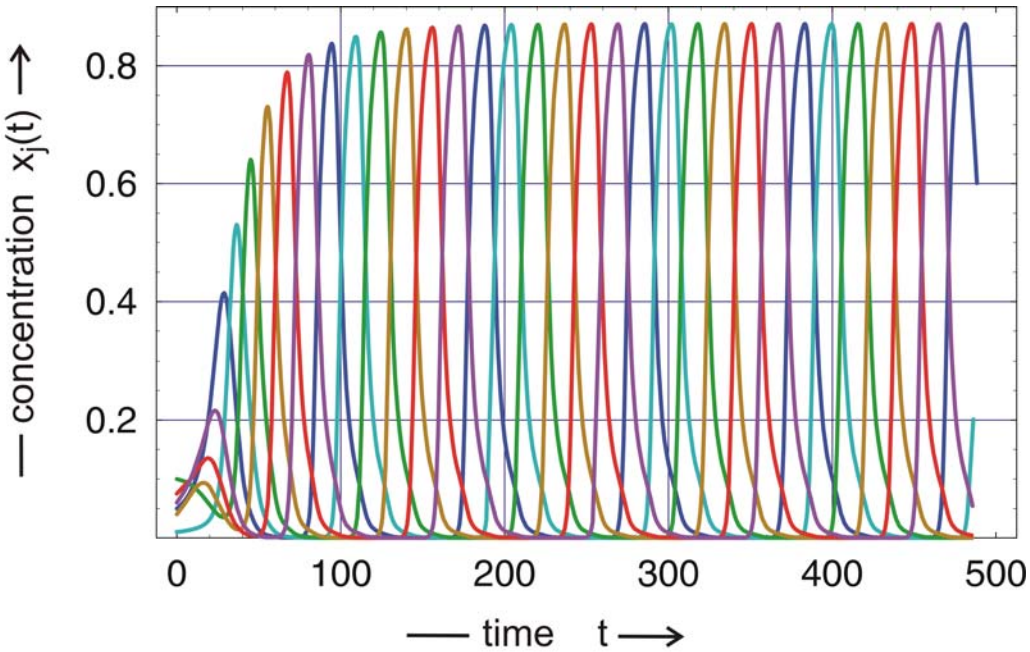
Hypercycle dynamics for n=3



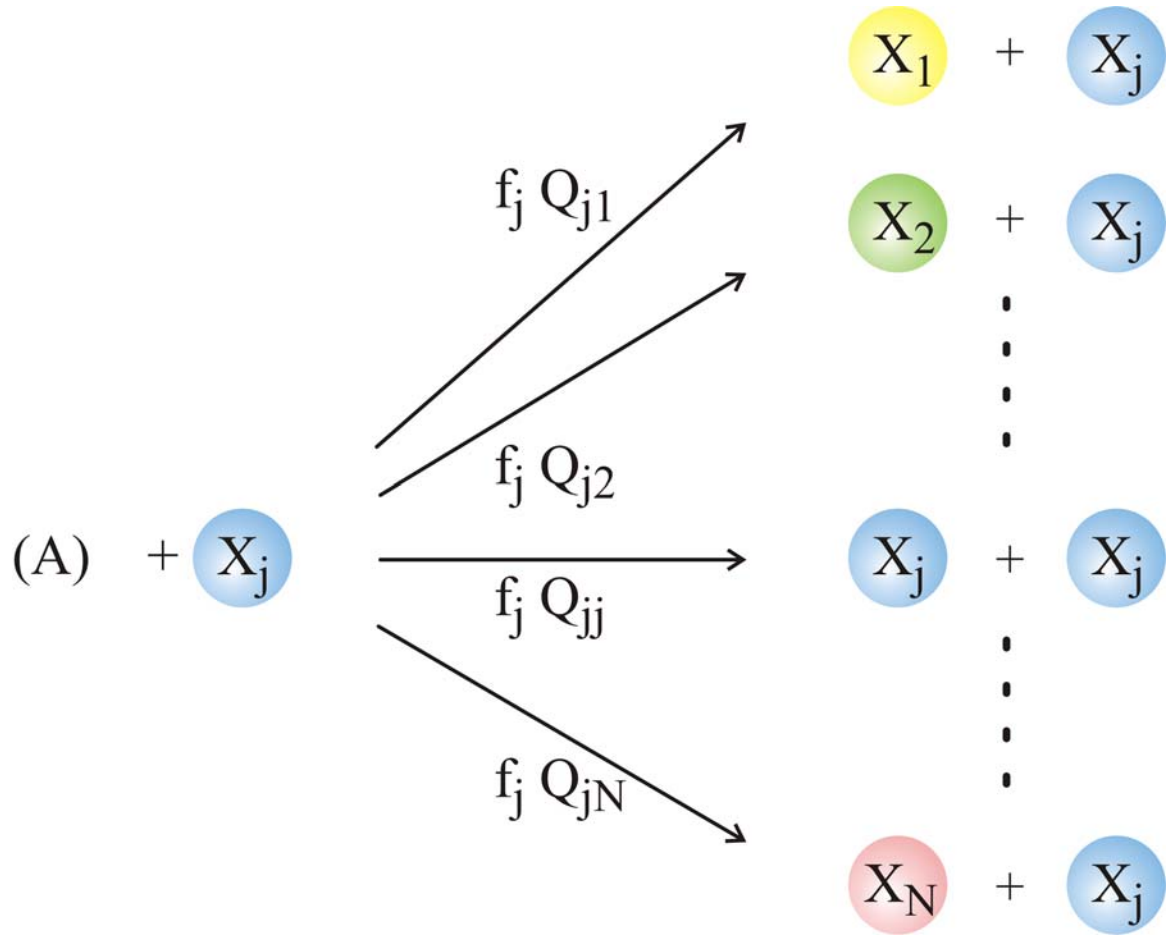


Hypercycle dynamics for n=4



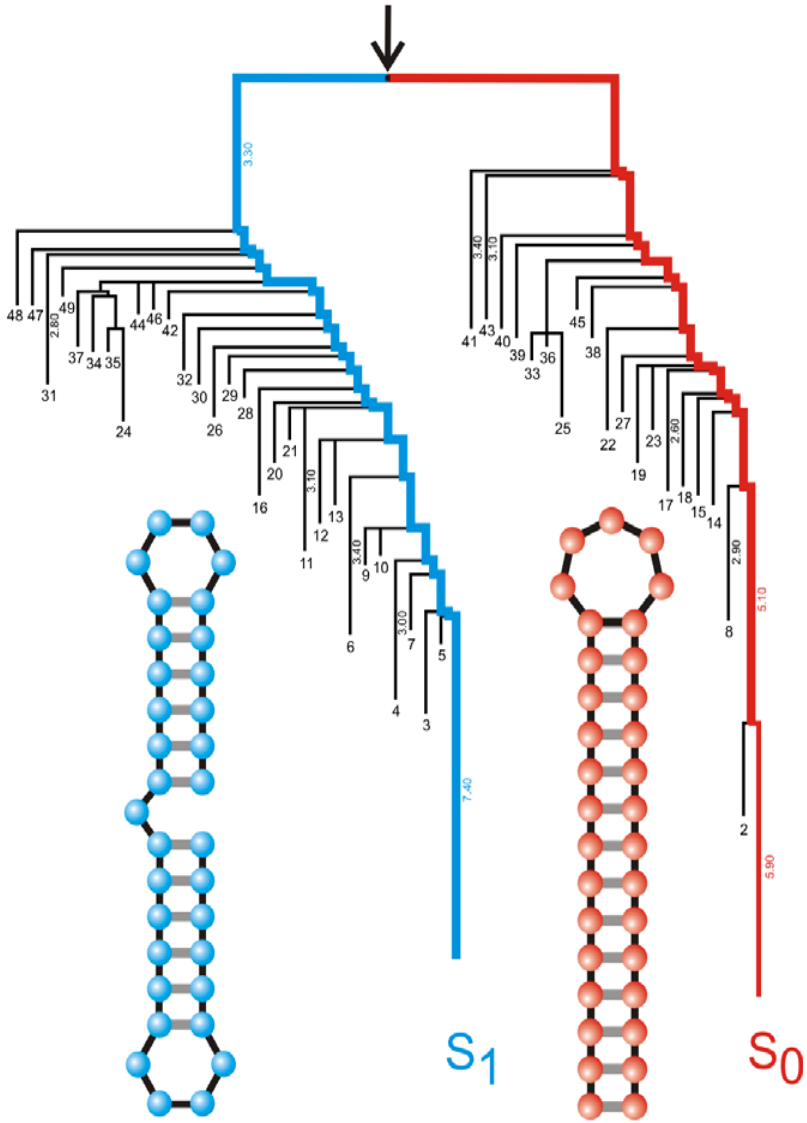


Hypercycle dynamics for n=6



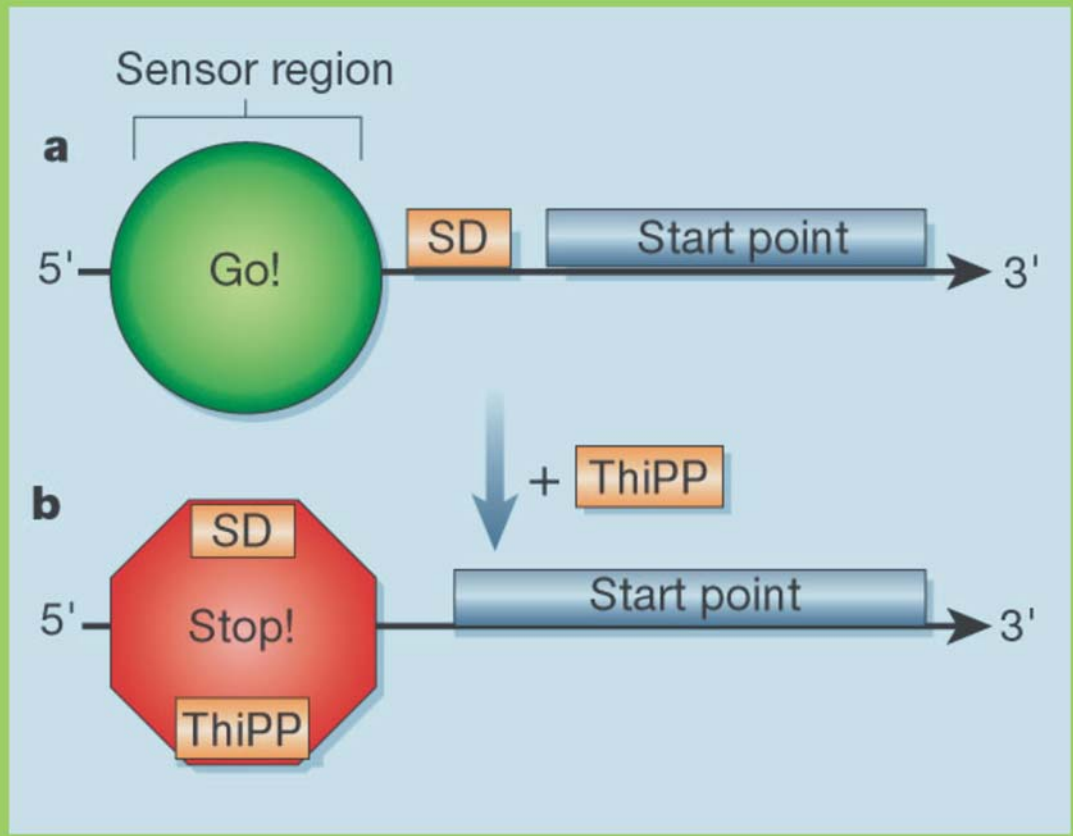
Chemical kinetics of replication and mutation as parallel reactions

Metastable structures
Conformational switches

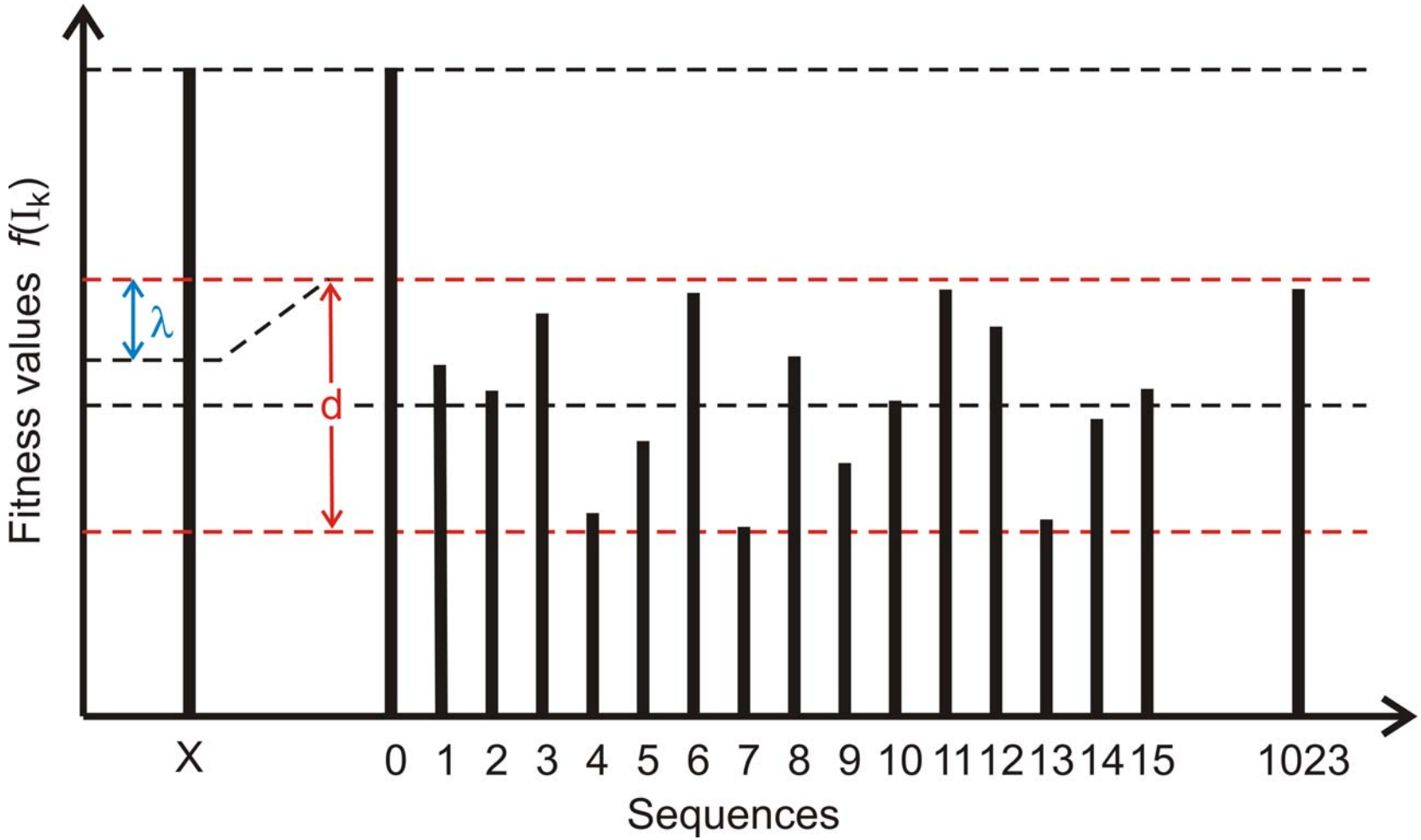


Kinetic structures

Allosteric control of transcribed RNA

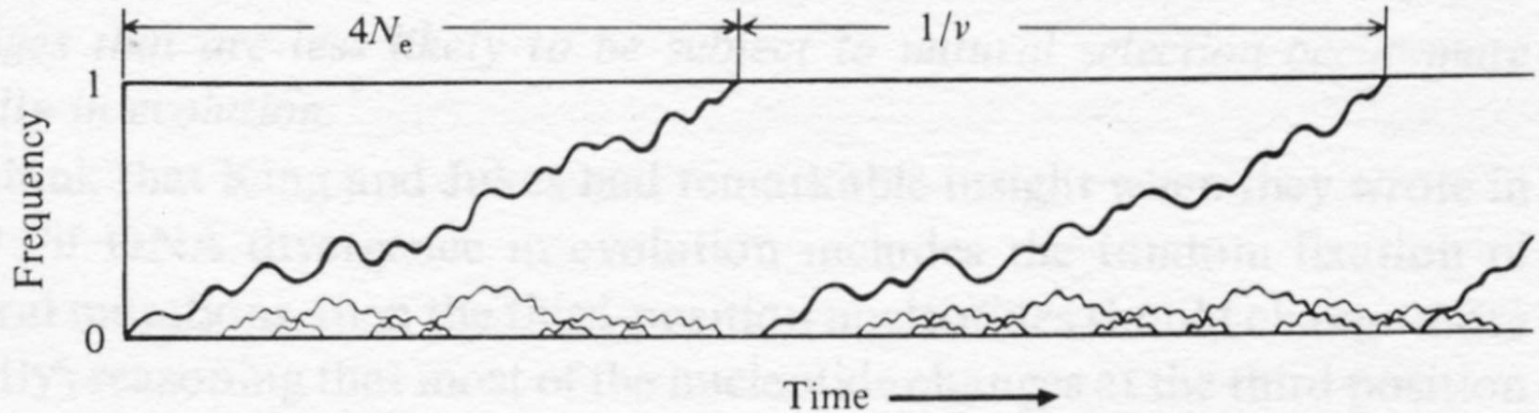


Riboswitches controlled by **metabolites**



A fitness landscape including neutrality

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?

Bulletin of Mathematical Biology Vol. 50, No. 6, pp. 635–660, 1988.
Printed in Great Britain.

0092-8240/88$3.00 + 0.00
Pergamon Press plc
Society for Mathematical Biology

STATIONARY MUTANT DISTRIBUTIONS AND EVOLUTIONARY OPTIMIZATION

■ PETER SCHUSTER and JÖRG SWETINA
Institut für theoretische Chemie
und Strahlenchemie der Universität Wien,
Währingerstraße 17,
A 1090 Wien,
Austria

Molecular evolution is modelled by erroneous replication of binary sequences. We show how the selection of two species of equal or almost equal selective value is influenced by its nearest neighbours in sequence space. In the case of perfect neutrality and sufficiently small error rates we find that the Hamming distance between the species determines selection. As the error rate increases the fitness parameters of neighbouring species become more and more important. In the case of almost neutral sequences we observe a critical replication accuracy at which a drastic change in the "quasispecies", in the stationary mutant distribution occurs. Thus, in frequently mutating populations fitness turns out to be an ensemble property rather than an attribute of the individual.

In addition we investigate the time dependence of the mean excess production as a function of initial conditions. Although it is optimized under most conditions, cases can be found which are characterized by decrease or non-monotonous change in mean excess productions.

1. Introduction. Recent data from populations of RNA viruses provided direct evidence for vast sequence heterogeneity (Domingo *et al.*, 1987). The origin of this diversity is not yet completely known. It may be caused by the low replication accuracy of the polymerizing enzyme, commonly a virus specific, RNA dependent RNA synthetase, or it may be the result of a high degree of selective neutrality of polynucleotide sequences. Eventually, both factors contribute to the heterogeneity observed. Indeed, mutations occur much more frequently than previously assumed in microbiology. They are by no means rare events and hence, neither the methods of conventional population genetics (Ewens, 1979) nor the neutral theory (Kimura, 1983) can be applied to these virus populations. Selectively neutral variants may be close with respect to Hamming distance and then the commonly made assumption that the mutation backflow from the mutants to the wilde type is negligible does not apply.

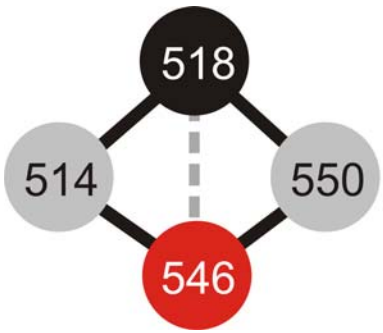
A kinetic theory of polynucleotide evolution which was developed during the past 15 years (Eigen, 1971; 1985; Eigen and Schuster, 1979; Eigen *et al.*, 1987; Schuster, 1986); Schuster and Sigmund, 1985) treats correct replication and mutation as parallel reactions within one and the same reaction network

Neutral network

$\lambda = 0.01,\ s = 367$

$\mathbf{d_H = 1}$

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



Neutral network

$\lambda = 0.01,\ s = 877$

$\mathbf{d_H = 2}$

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

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

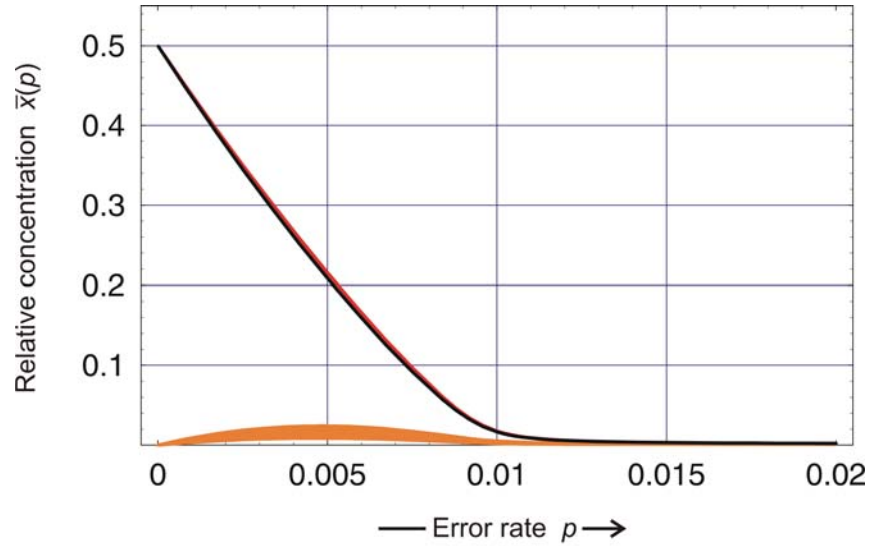
$\mathbf{d_H \geq 3}$

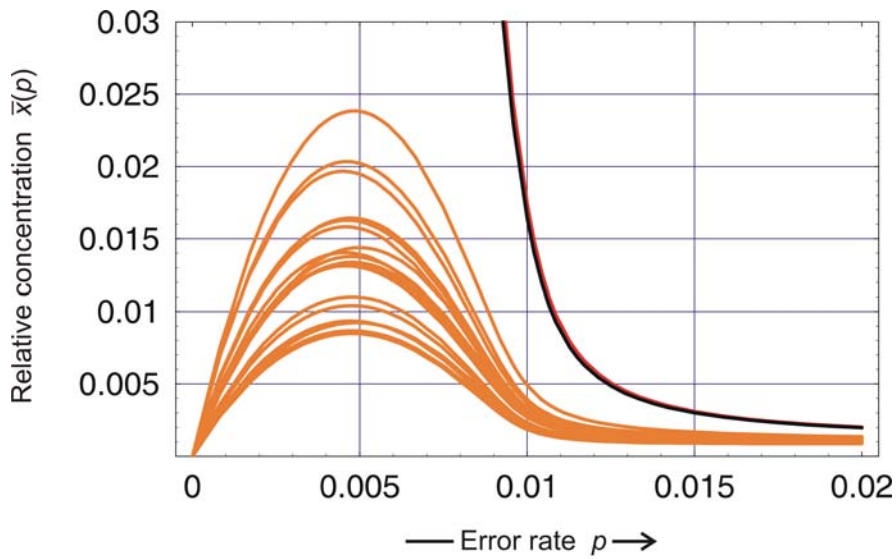
random fixation in the sense of Motoo Kimura

Pairs of genotypes in neutral replication networks

Neutral network

$\lambda = 0.01,\ s = 367$

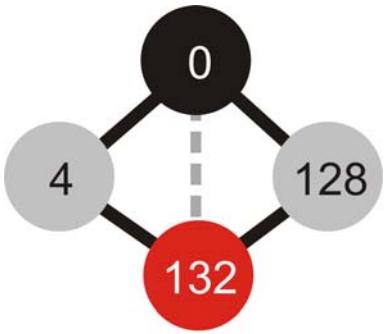




Neutral network: Individual sequences

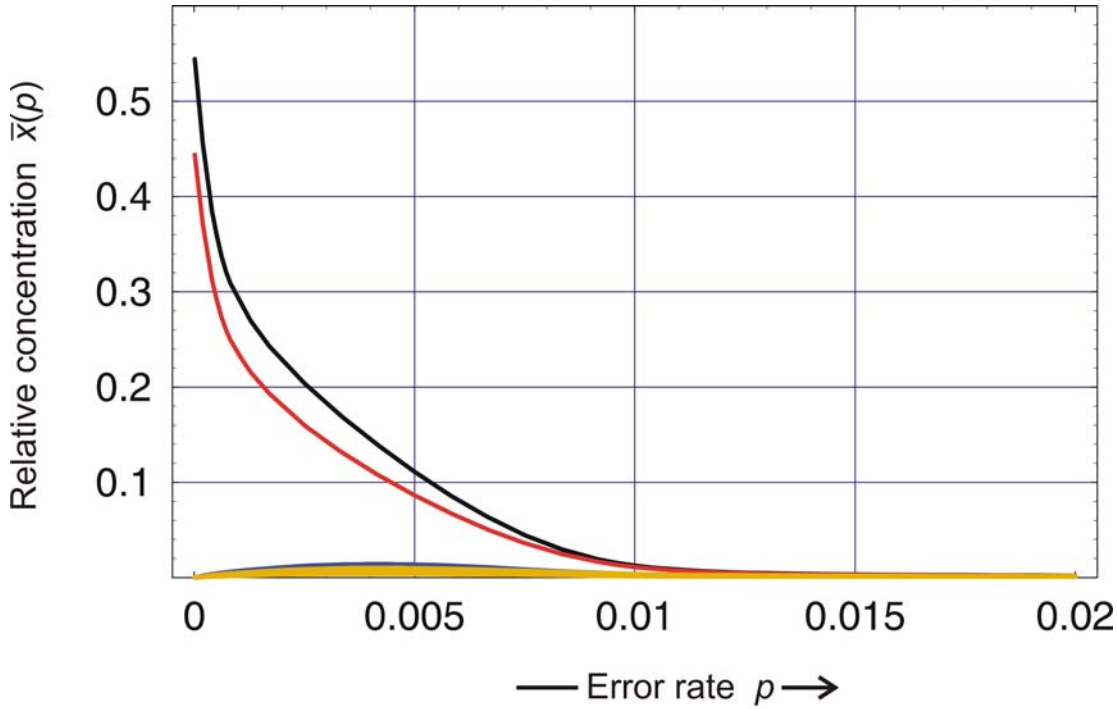
$n = 10,\ \sigma = 1.1,\ d = 1.0$

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



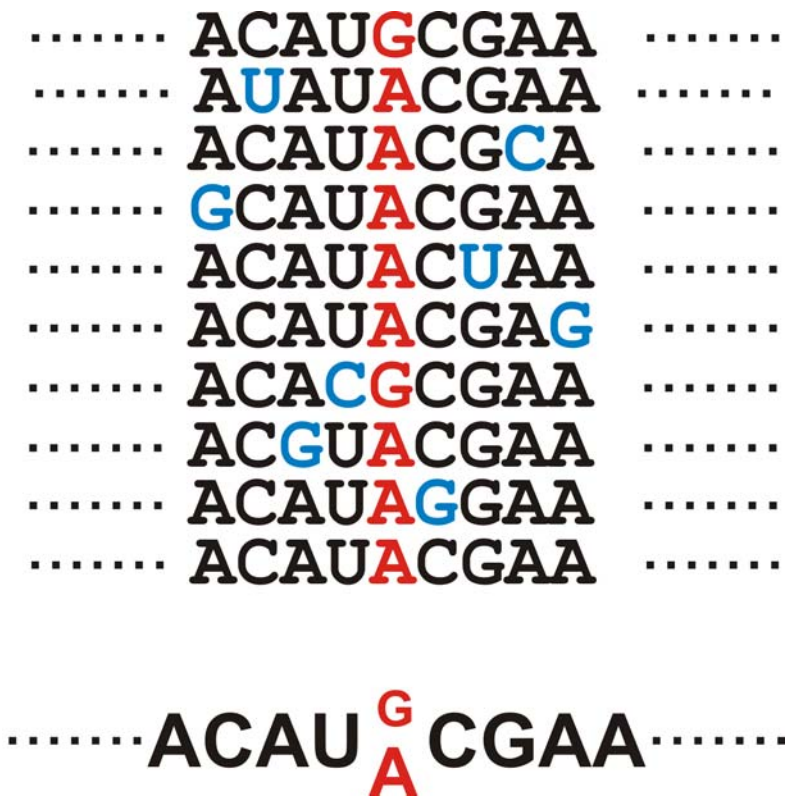
Neutral network

$\lambda = 0.01,\ s = 877$

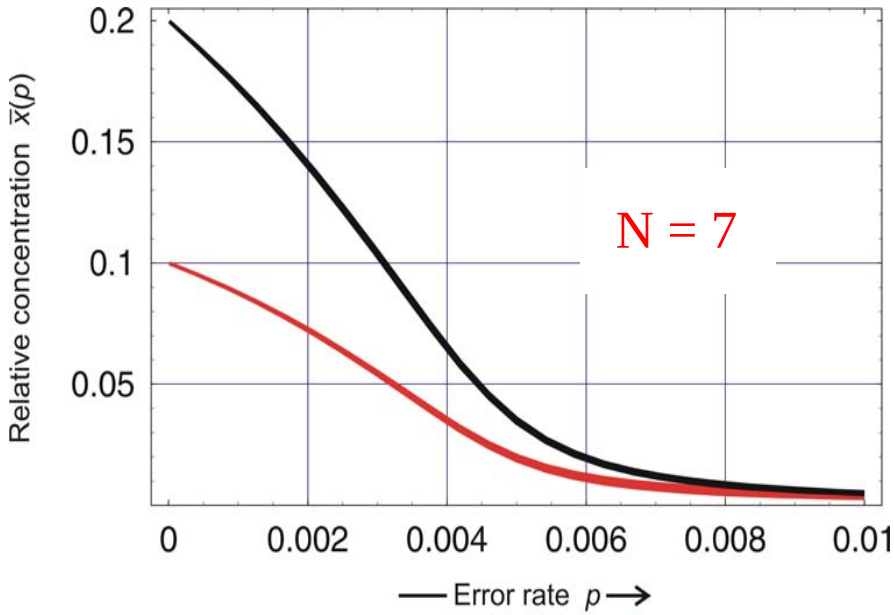


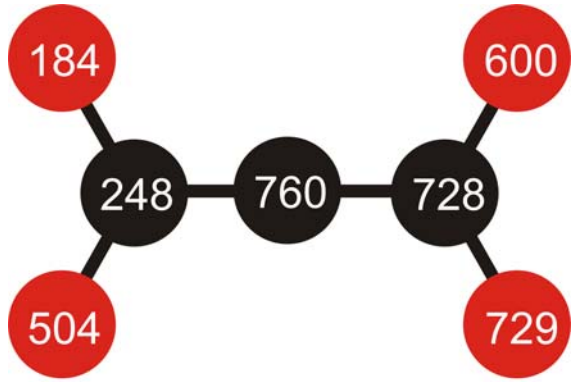
Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0



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

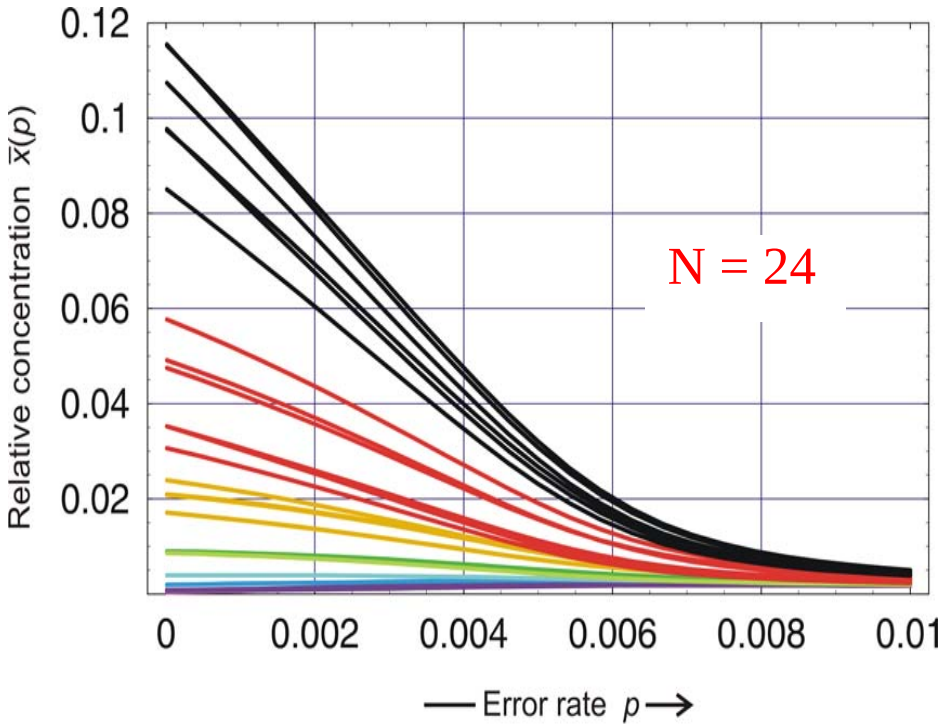


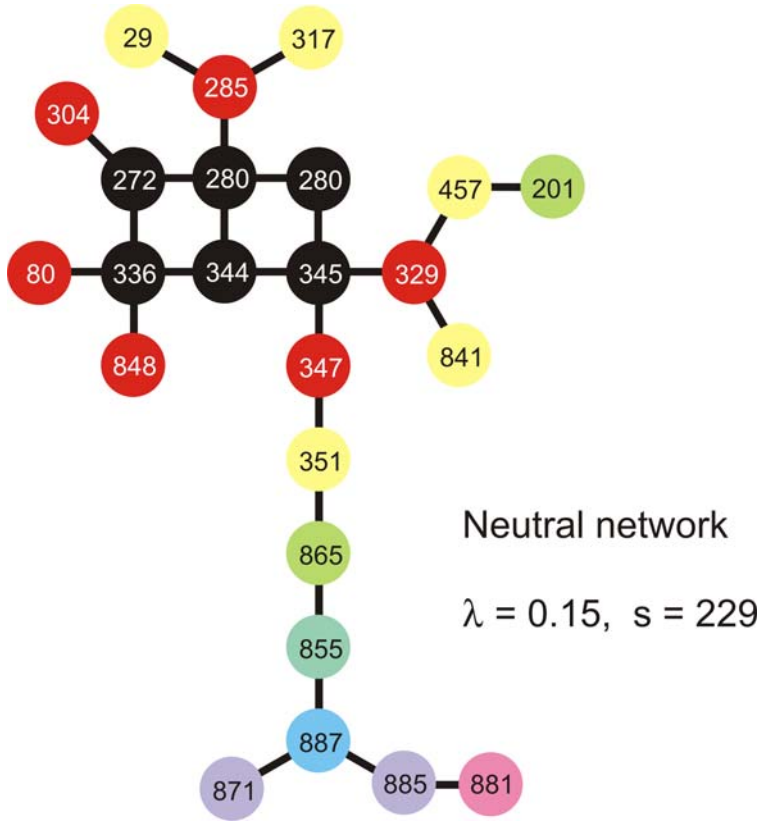


Neutral network

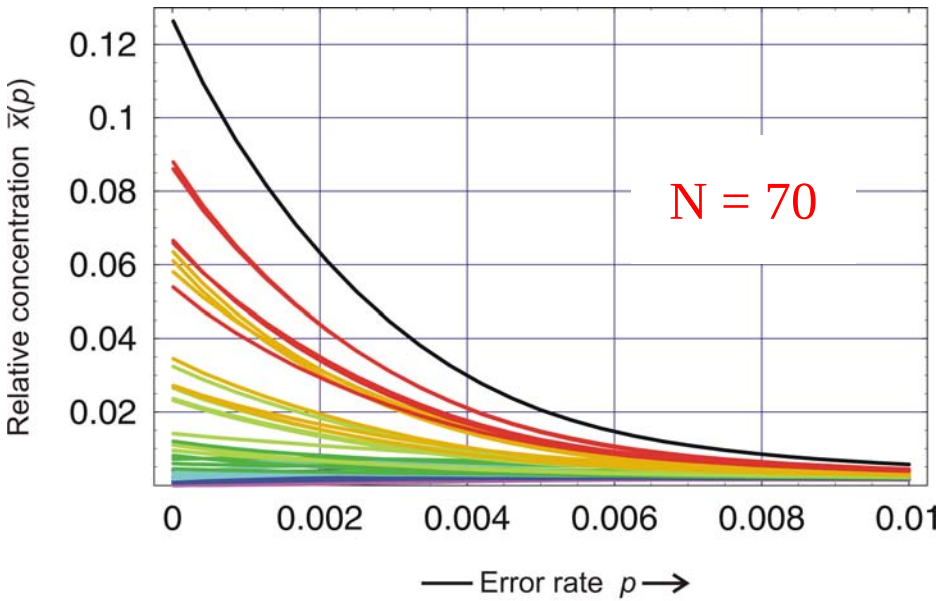
$\lambda = 0.10$, s = 229

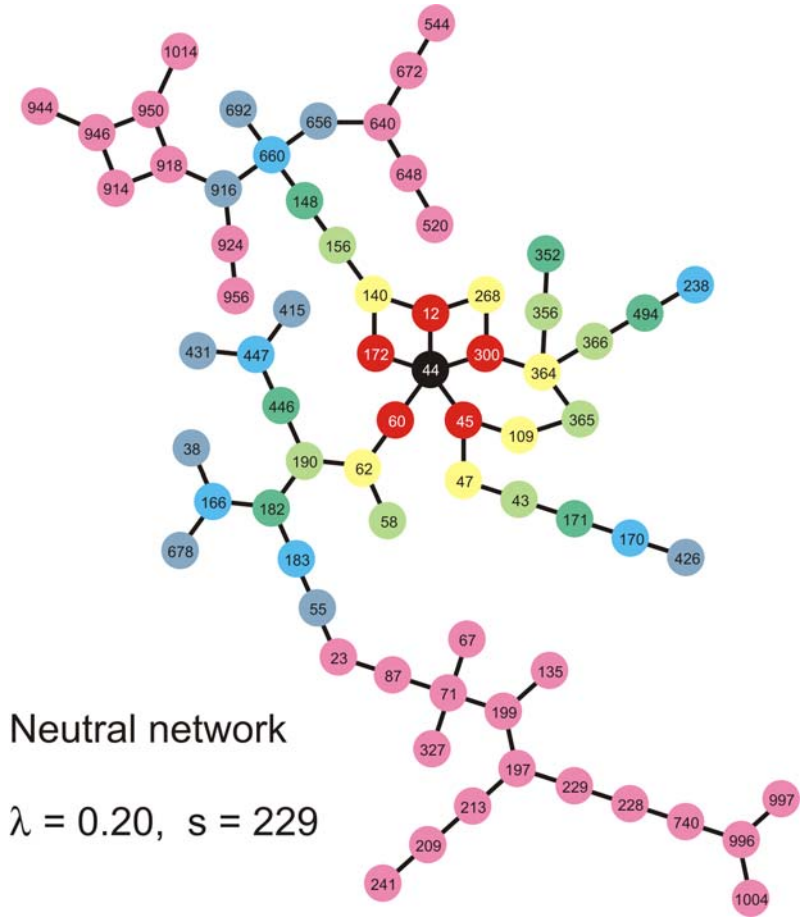
Neutral networks with increasing λ: $\lambda = 0.10$, s = 229





Neutral networks with increasing λ: λ = 0.15, s = 229





Neutral networks with increasing λ: λ = 0.20, s = 229

Web-Page for further information:

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