Some Problems in Molecular Evolution

An Overview

Peter Schuster

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

Institute Para Limes Workshop:
Genes, Infections, and Epidemics

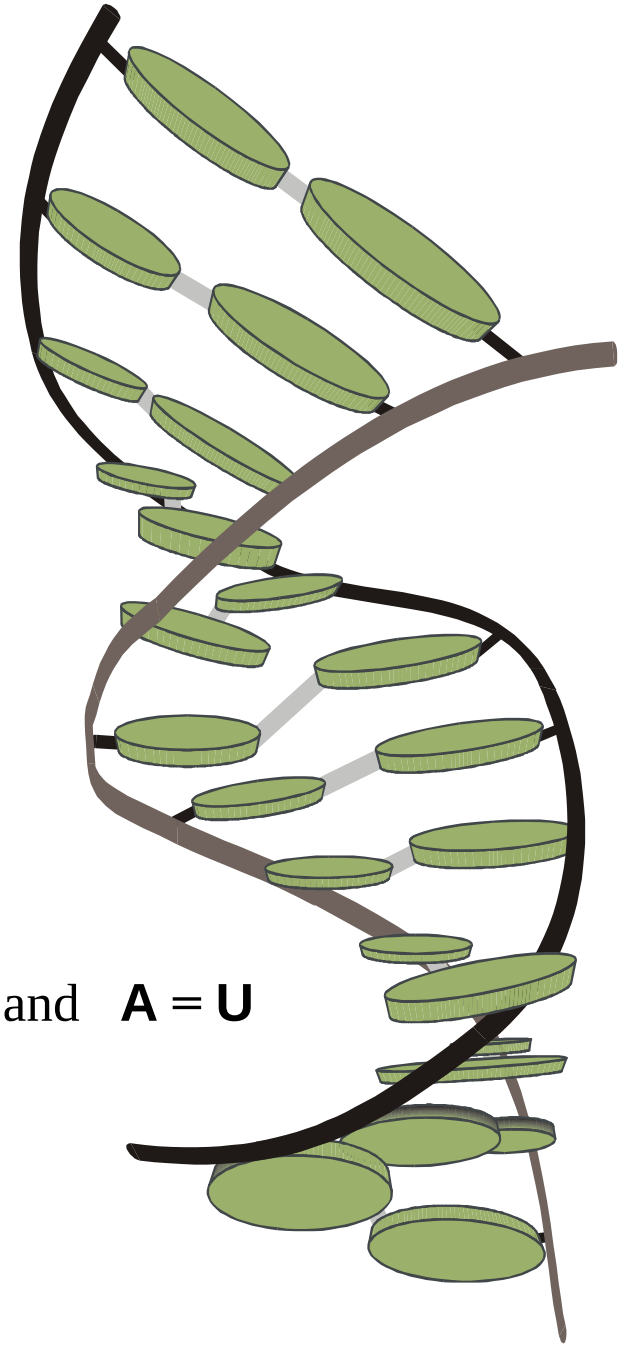
Wageningen, 13.– 15.06.2007

Web-Page for further information:

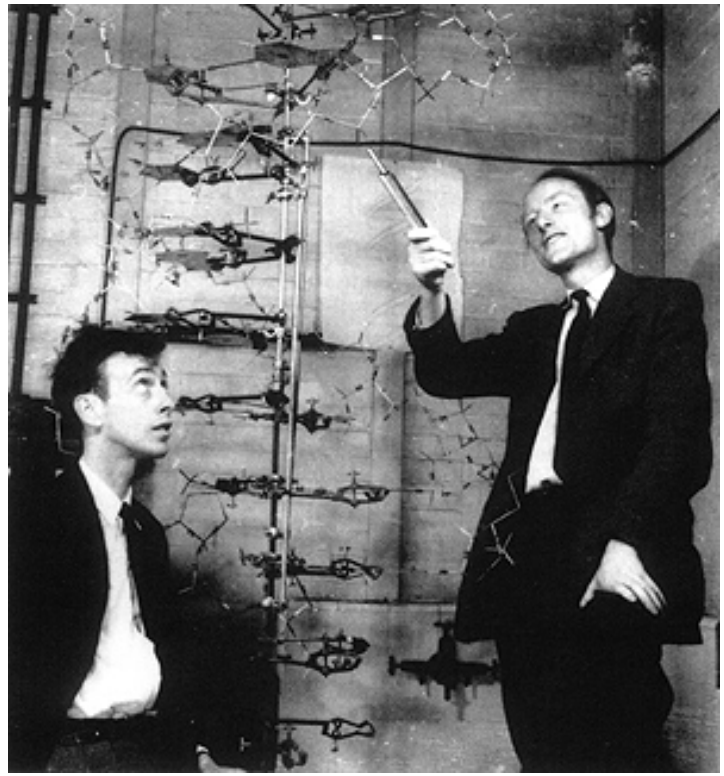
http://www.tbi.univie.ac.at/~pks

Three avenues to extinction

1. Dilution

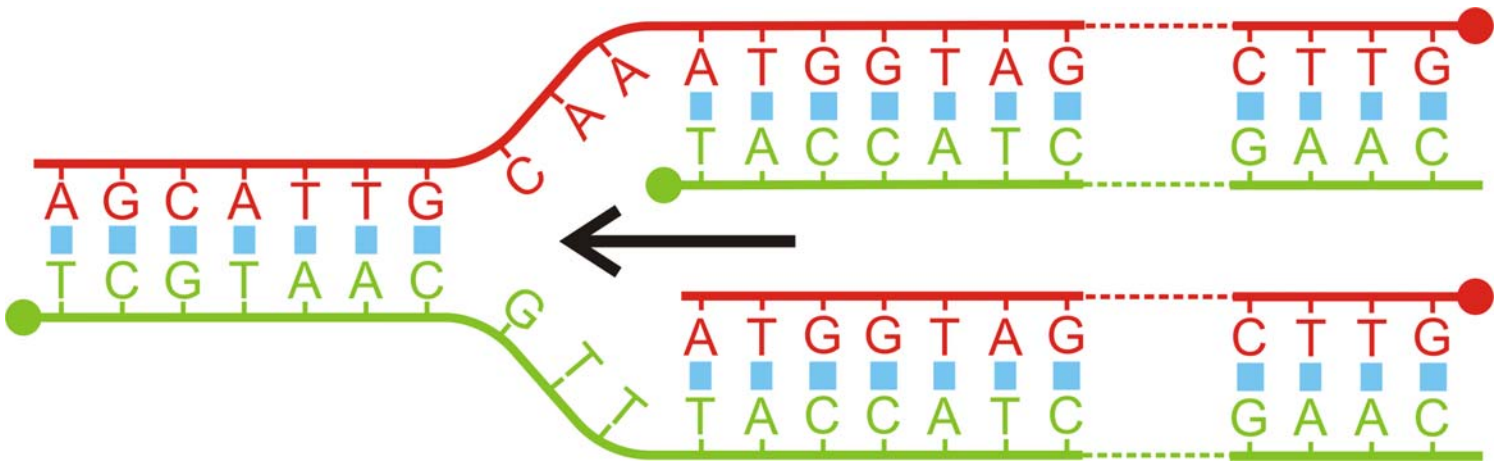


G ≡ C and **A = U**



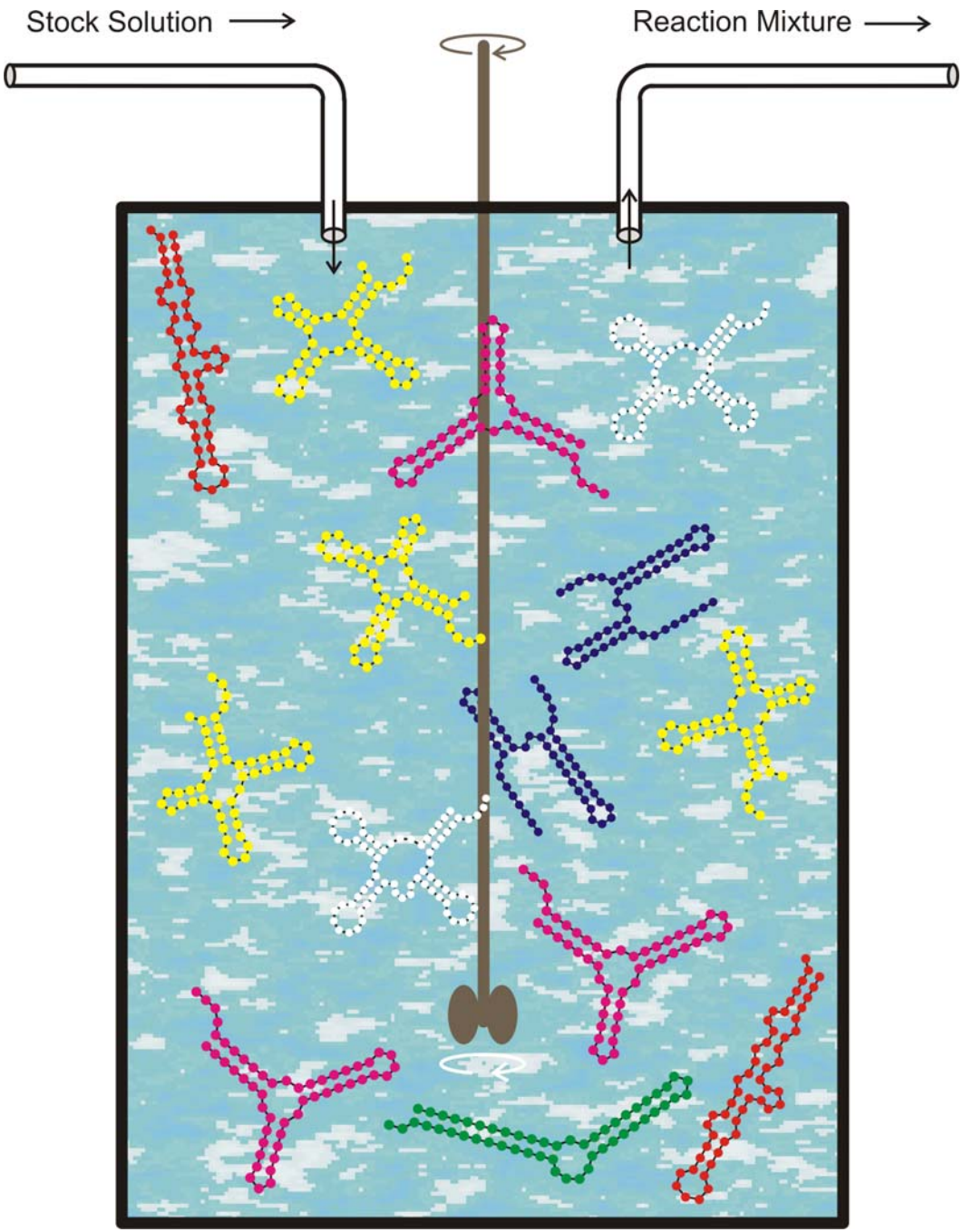
James D. Watson, 1928- , and Francis Crick, 1916-2004,
Nobel Prize 1962

The three-dimensional structure of a short double helical stack of B-DNA



‚Replication fork' in DNA replication

The mechanism of DNA replication is ‚semi-conservative'



Stock solution:

$[A] = a = a_0$

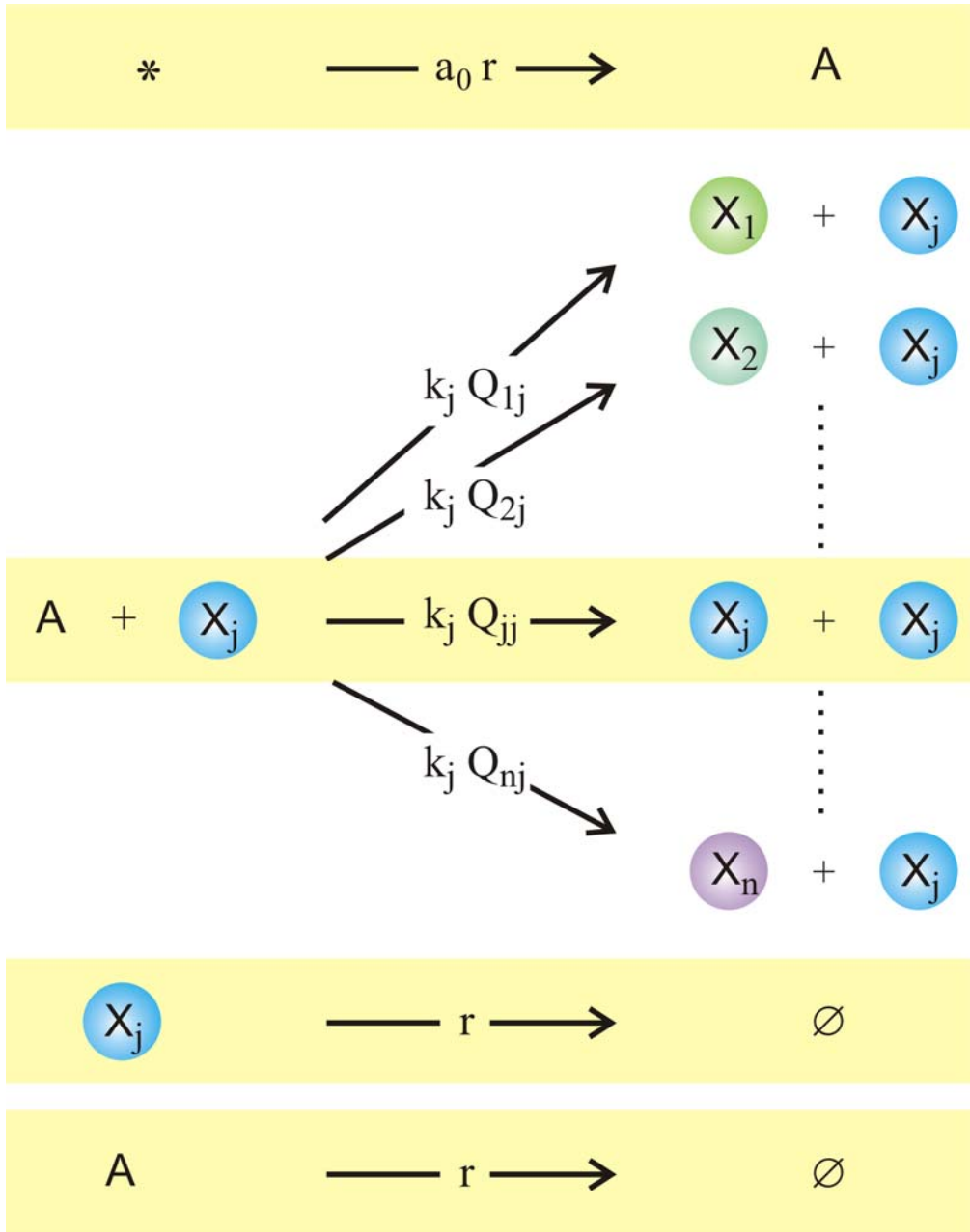
Flow rate:

$r = \tau_R^{-1}$

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***



$j = 1,2, \ldots ,n$

$$
\begin{aligned}
* &\xrightarrow{a_0 \cdot r} \mathbf{A} && : \text{influx} \\
\mathbf{A} + \mathbf{X}_i &\xrightarrow{k_i\, Q_{ii}} 2\,\mathbf{X}_i \;; \quad i = 1, \ldots, n && : \text{replication} \\
\mathbf{A} + \mathbf{X}_i &\xrightarrow{k_i\, Q_{ji}} \mathbf{X}_i + \mathbf{X}_j \quad i, j = 1, \ldots, n;\; i \neq j && : \text{mutation} \\
\mathbf{A} &\xrightarrow{r} \emptyset && : \text{outflux} \\
\mathbf{X}_j &\xrightarrow{r} \emptyset \;; \quad j = 1, \ldots, n && : \text{outflux}
\end{aligned}
$$

$$
\frac{da}{dt} \;=\; -\,a \sum_{i=1}^{n} \sum_{j=1}^{n} k_i Q_{ji}\, x_i \;+\; r\,(a_0 - a) \;=\; -\,a \sum_{i=1}^{n} k_i\, x_i \;+\; r\,(a_0 - a)
$$

$$
\frac{dx_j}{dt} \;=\; a \sum_{i=1}^{n} k_i Q_{ji}\, x_i \;-\; r\, x_j
$$

Origin of the replication-mutation equation from the flowreactor

Stationary solutions of the flow reactor:

$$\frac{da}{dt} = 0 = -\tilde{a}\left(\sum_{i=1}^{n} k_i\,\tilde{x}_i + r\right) + r\,\tilde{a}$$

$$\frac{dx_j}{dt} = 0 = \tilde{a}\sum_{i=1}^{n} k_i Q_{ji}\tilde{x}_i - r\,\tilde{x}_j\,;\; c = \sum_{i=1}^{n} x_i\,;\; \bar{k} = \frac{\sum_{i=1}^{n} k_i\,x_i}{c}$$

$$\frac{dc}{dt} = 0 = \tilde{c}\left(\bar{k}\,\tilde{a} - r\right)$$

Stationary solutions: 1. active state

$$r < \bar{k}\,a_0$$

$$\tilde{a} = \frac{r}{\bar{k}}$$

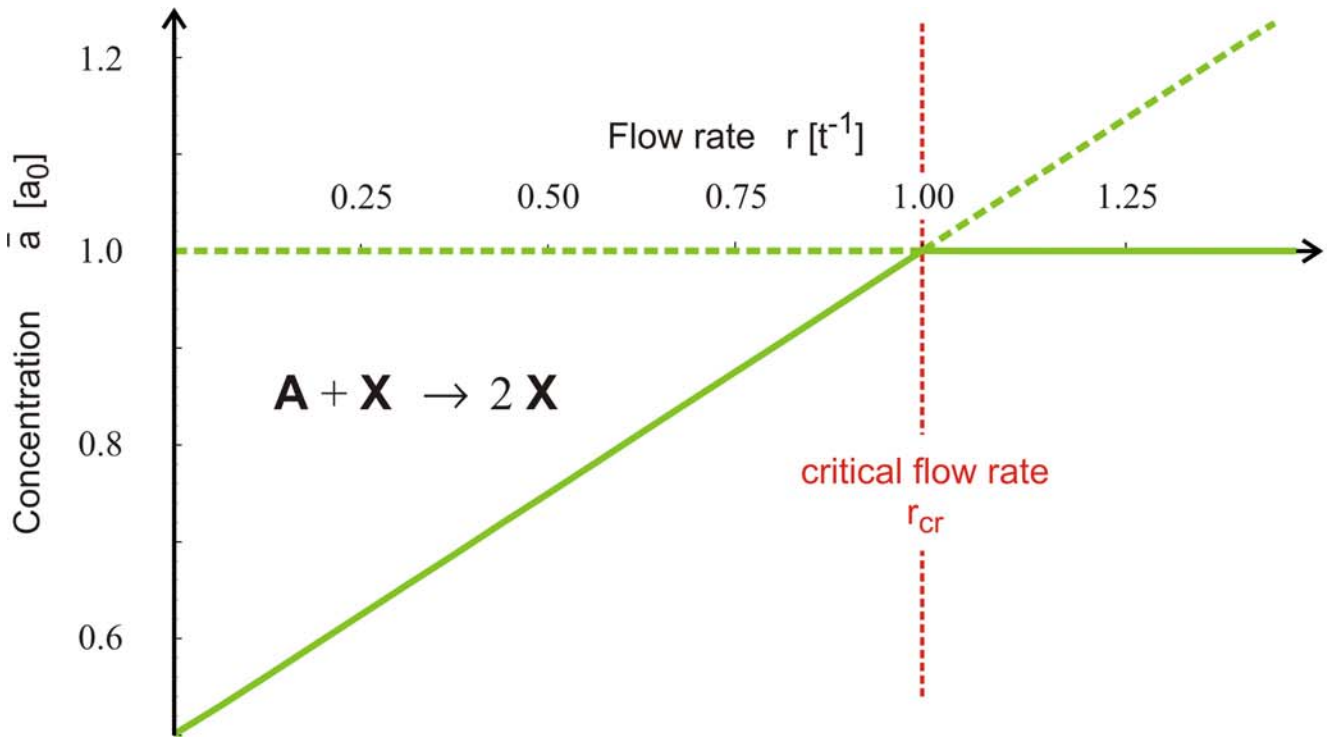
$$\tilde{c} = \frac{\bar{k}\,a_0 - r}{\bar{k}}$$

Stationary solutions: 2. extinction

$$r > \bar{k}\,a_0$$

$$\tilde{a} = a_0$$

$$\tilde{x}_j = 0\,;\; j = 1, 2, \ldots, n$$


Concentration $\bar{a}$ [a_0]
1.2
1.0
0.8
0.6
Flow rate r [t^{-1}]
0.25
0.50
0.75
1.00
1.25
A + X → 2 X
critical flow rate
r_{cr}

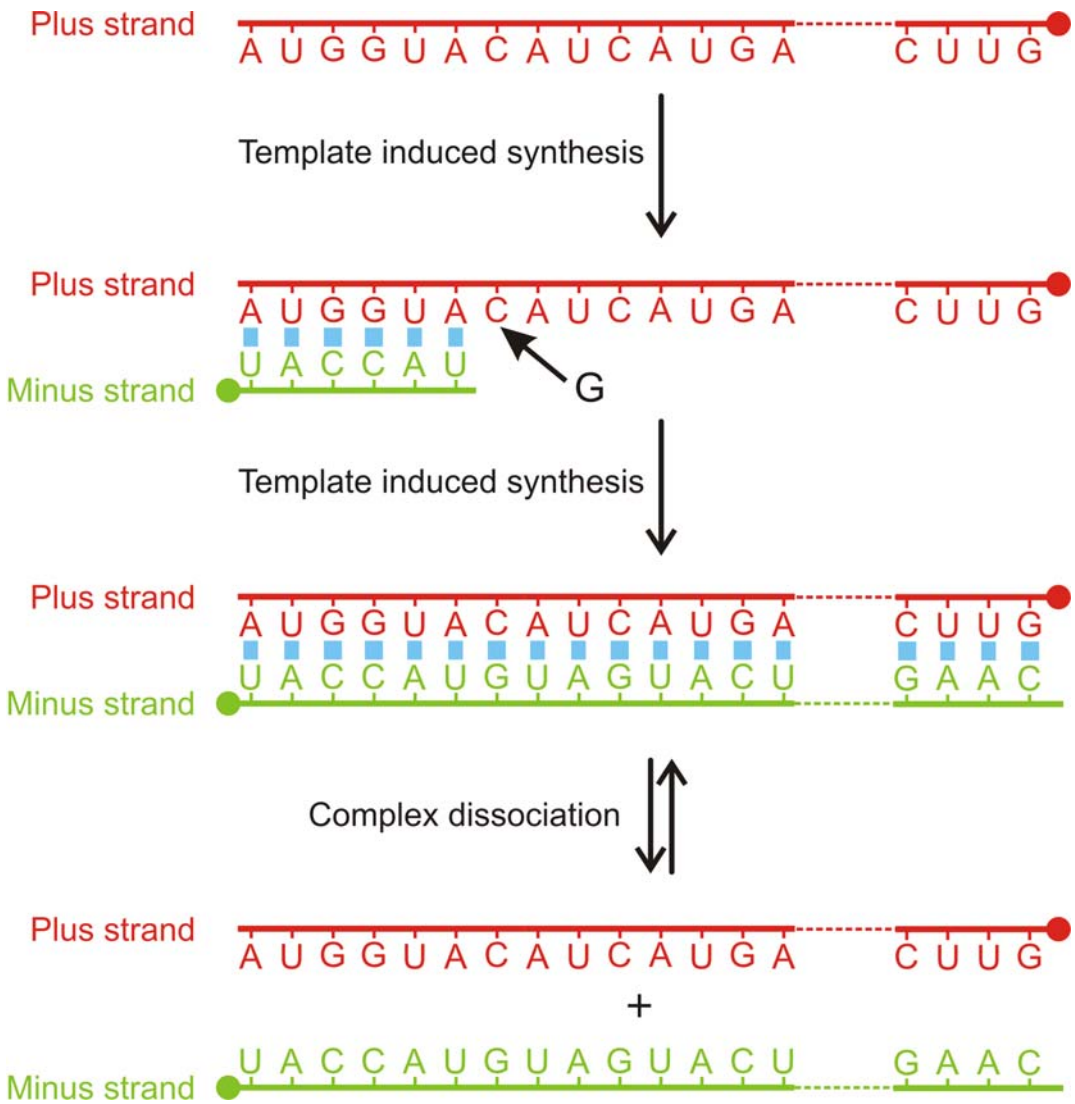
Find $r(t)$ such that $a(t) = \bar{a} = const.$

$$\frac{da}{dt} \;=\; 0 \;=\; -\,\bar{a}\sum_{i=1}^{n}\sum_{j=1}^{n} k_i Q_{ji}\, x_i \;+\; r(t)\,(a_0 - \bar{a})$$

$$r(t) \;=\; \frac{\bar{a}}{a_0 - \bar{a}}\sum_{i=1}^{n} k_i\, x_i\,;\;\; f_i \;=\; k_i\,\bar{a}$$

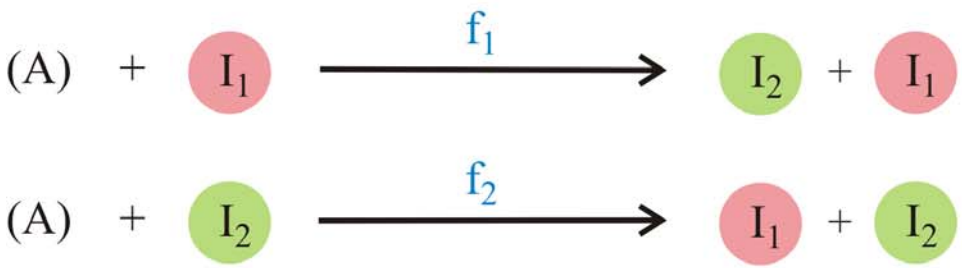
$$\frac{dx_j}{dt} \;=\; \sum_{i=1}^{n} f_i Q_{ji}\, x_i \;-\; x_j\,\frac{\sum_{i=1}^{n} f_i\, x_i}{\sum_{i=1}^{n} x_i} \;=\; \sum_{i=1}^{n} f_i Q_{ji}\, x_i \;-\; x_j\,\bar{f}$$

Origin of the replication-mutation equation from the flowreactor



Complementary replication is the simplest copying mechanism of RNA.
Complementarity is determined by Watson-Crick base pairs:

G≡C and **A=U**



$$dx_1 / dt = f_2 x_2 - x_1 \Phi$$
$$dx_2 / dt = f_1 x_1 - x_2 \Phi$$

$$\Phi = \Sigma_i f_i x_i ; \quad \Sigma_i x_i = 1 ; \quad i = 1,2$$

Complementary replication as the simplest molecular mechanism of reproduction

Equation for complementary replication: $[I_i] = x_i \geq 0\ ,\ f_i > 0\ ;\ i=1,2$

$$\frac{dx_1}{dt} = f_2 x_2 - x_1\phi\ ,\quad \frac{dx_2}{dt} = f_1 x_1 - x_2\phi\ ;\quad x_1 + x_2 = 1;\quad \phi = f_1 x_1 + f_2 x_2 = \overline{f}$$

Solutions are obtained by integrating factor transformation

$$x_{1,2}(t) = \frac{\sqrt{f_{2,1}}\left(\gamma_1(0)\cdot\exp(ft) + \gamma_2(0)\cdot\exp(-ft)\right)}{(\sqrt{f_1}+\sqrt{f_2})\,\gamma_1(0)\cdot\exp(ft) - (\sqrt{f_1}-\sqrt{f_2})\,\gamma_2(0)\cdot\exp(-ft)}$$

$$\gamma_1(0) = \sqrt{f_1}\,x_1(0) + \sqrt{f_2}\,x_2(0),\ \gamma_2(0) = \sqrt{f_1}\,x_1(0) - \sqrt{f_2}\,x_2(0);\quad f = \sqrt{f_1\,f_2}$$

$$x_1(t) \to \frac{\sqrt{f_2}}{\sqrt{f_1}+\sqrt{f_2}} \text{ and } x_2(t) \to \frac{\sqrt{f_1}}{\sqrt{f_1}+\sqrt{f_2}} \text{ as } \exp(-ft) \to 0$$

Evolution of RNA molecules based on Qβ phage

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

S.Spiegelman, ***An approach to the experimental analysis of precellular evolution***. Quart.Rev.Biophys. **4** (1971), 213-253

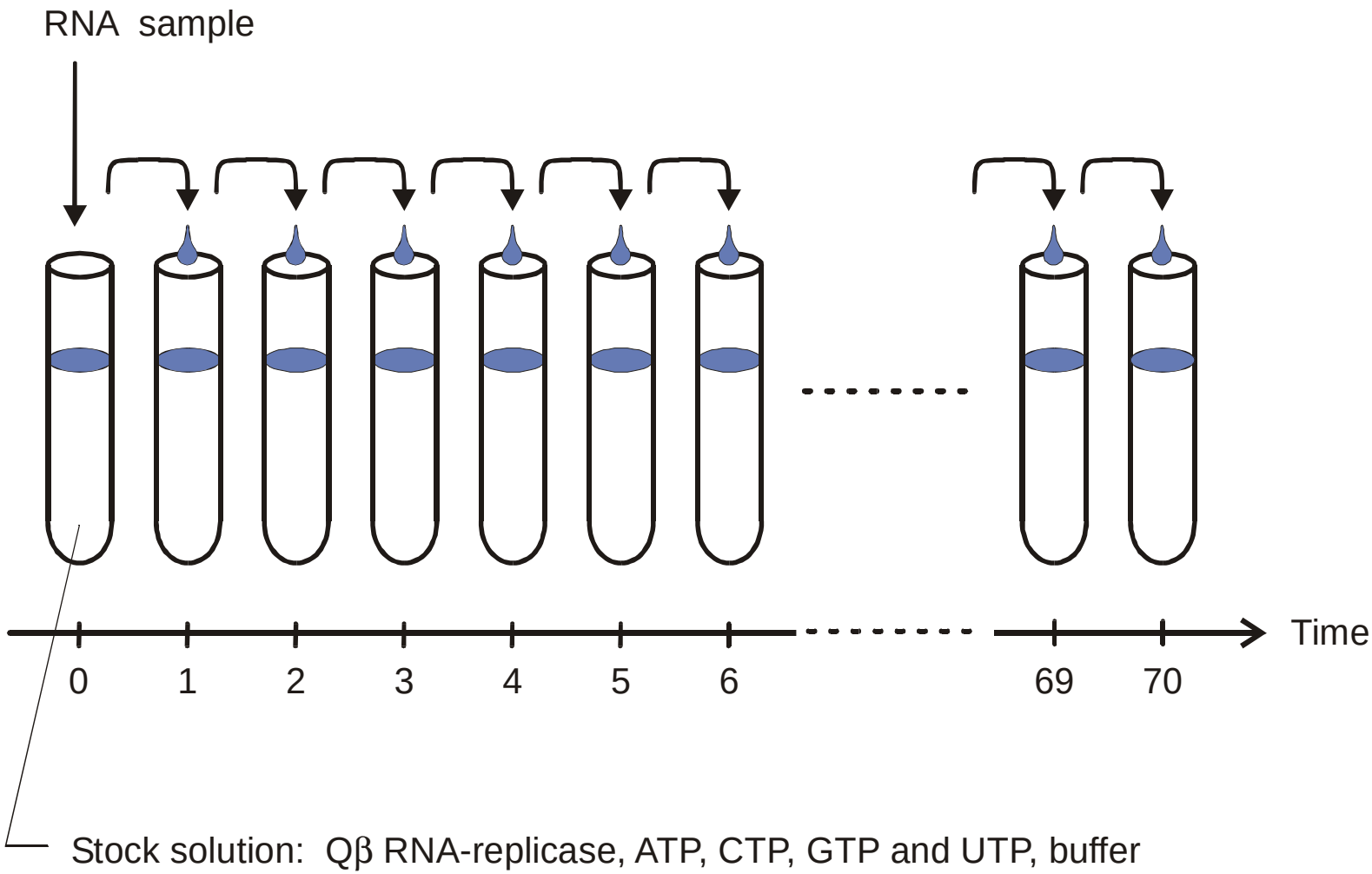
C.K.Biebricher, ***Darwinian selection of self-replicating RNA molecules***. Evolutionary Biology **16** (1983), 1-52

G.Bauer, H.Otten, J.S.McCaskill, ***Travelling waves of* in vitro *evolving RNA.*** *Proc.Natl.Acad.Sci.USA* **86** (1989), 7937-7941

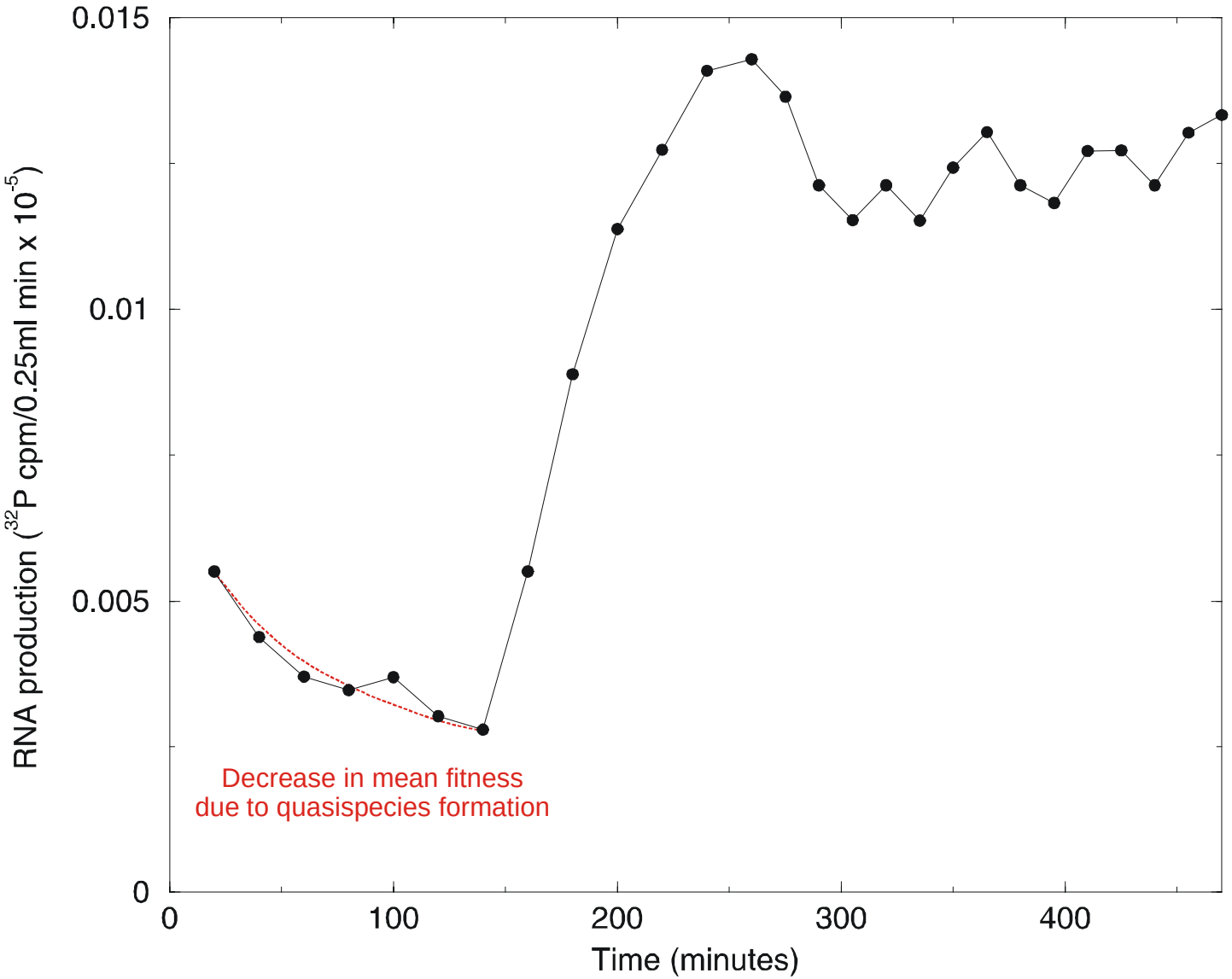
C.K.Biebricher, W.C.Gardiner, ***Molecular evolution of RNA* in vitro**. Biophysical Chemistry **66** (1997), 179-192

G.Strunk, T.Ederhof, ***Machines for automated evolution experiments* in vitro *based on the serial transfer concept***. Biophysical Chemistry **66** (1997), 193-202

F.Öhlenschlager, M.Eigen, ***30 years later – A new approach to Sol Spiegelman‘s and Leslie Orgel‘s* in vitro *evolutionary studies***. Orig.Life Evol.Biosph. **27** (1997), 437-457



The serial transfer technique applied to RNA evolution ***in vitro***



The increase in RNA production rate during a serial transfer experiment

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.

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

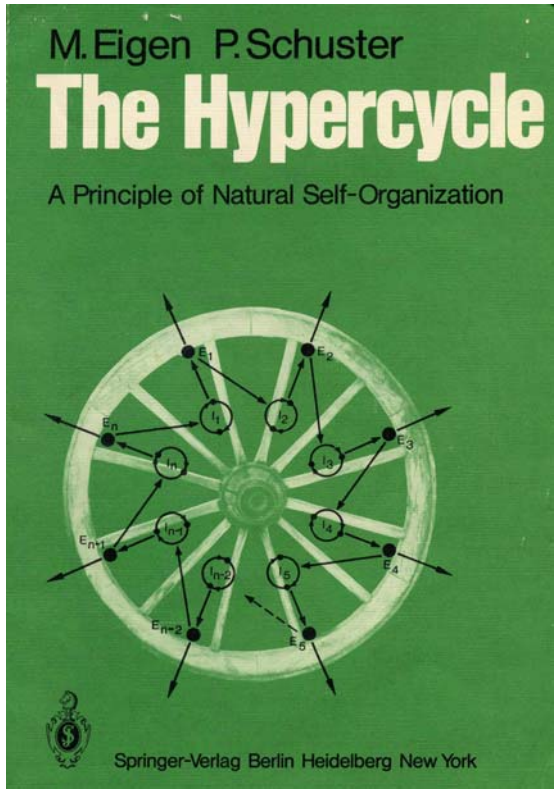
The mathematical analysis of dynamical systems using methods of differential topology, yields the result that there is only one type of mechanisms which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective 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 functions.
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 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.

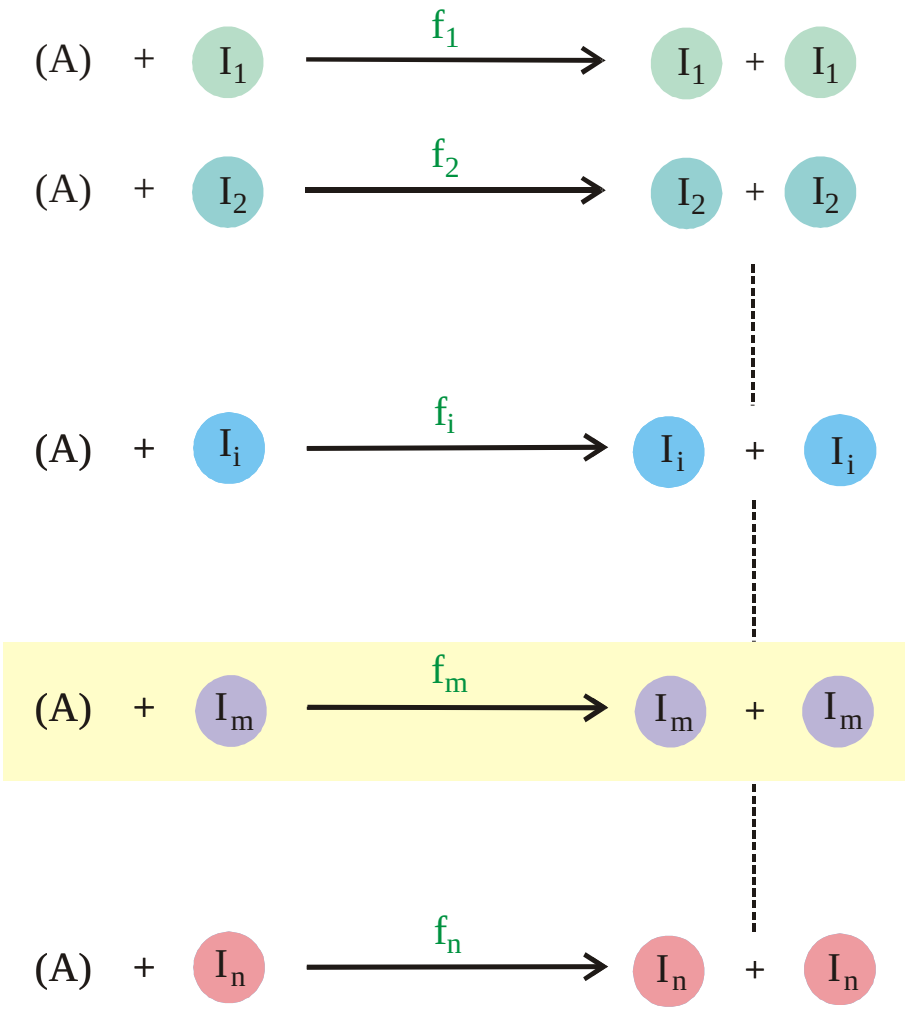
I. The Paradigm of Unity and Diversity in Evolution

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-



Chemical kinetics of molecular evolution

M. Eigen, P. Schuster, `The Hypercycle´, Springer-Verlag, Berlin 1979



$$dx_i / dt = f_i x_i - x_i \Phi = x_i (f_i - \Phi)$$

$$\Phi = \Sigma_j f_j x_j ; \quad \Sigma_j x_j = 1 ; \quad i,j = 1,2,...,n$$

$$[I_i] = x_i \geq 0 ; \quad i = 1,2,...,n ;$$

$$[A] = a = \text{constant}$$

$$f_m = \max \{f_j; j=1,2,...,n\}$$

$$x_m(t) \rightarrow 1 \quad \text{for} \quad t \rightarrow \infty$$

Reproduction of organisms or replication of molecules as the basis of selection

Selection equation: $[\mathrm{I}_i] = x_i \geq 0\ ,\ f_i > 0$

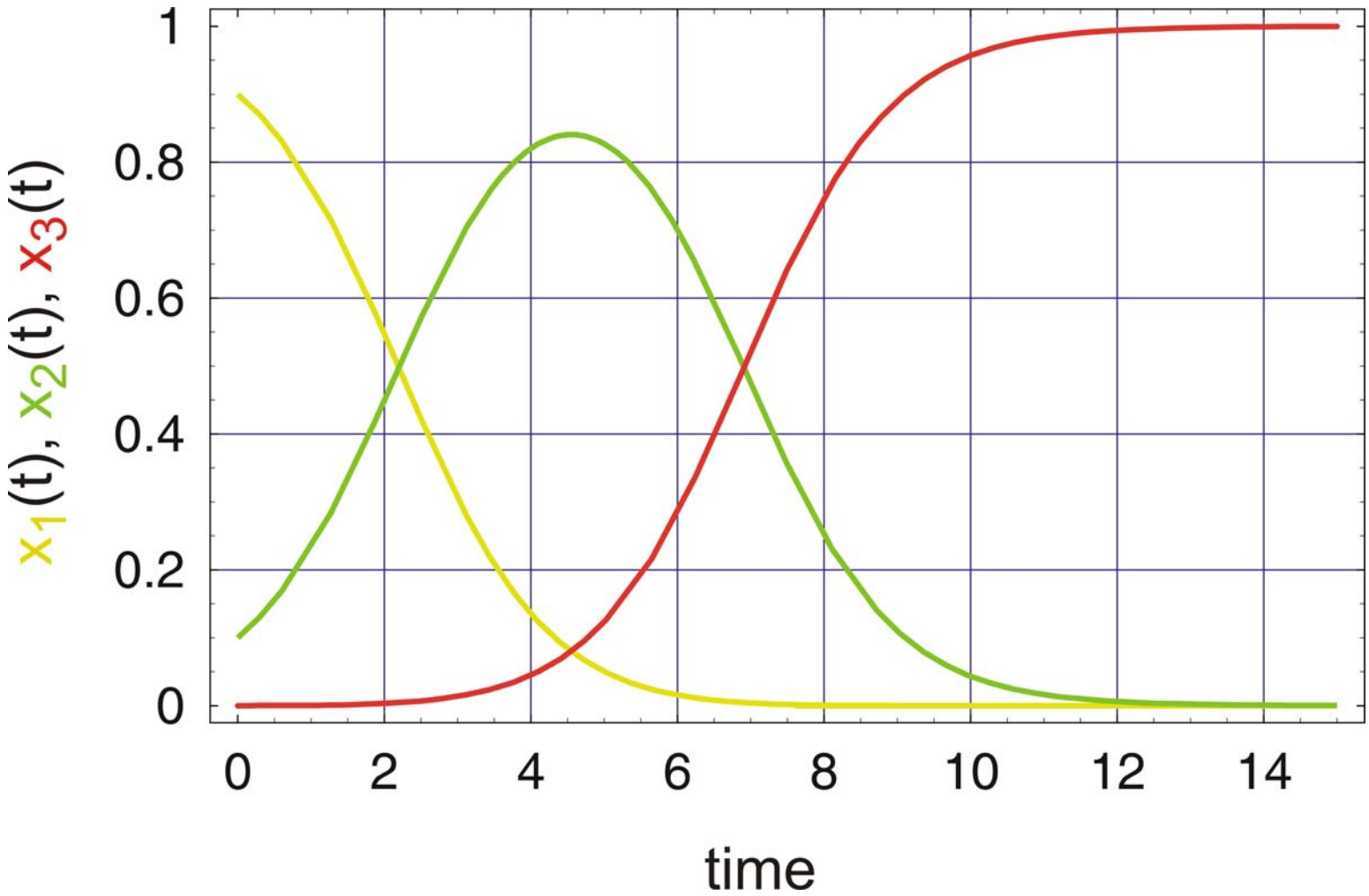
$$\frac{dx_i}{dt} = x_i\left(f_i - \phi\right), \quad i=1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \overline{f}$$

Mean fitness or dilution flux, $\phi(t)$, is a **non-decreasing function** of time,

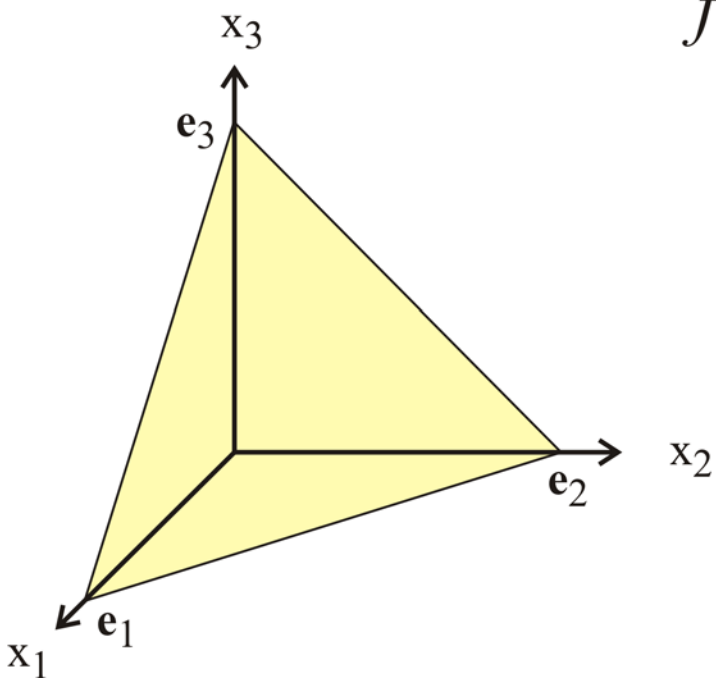
$$\frac{d\phi}{dt} = \sum_{i=1}^{n} f_i \frac{dx_i}{dt} = \overline{f^2} - \left(\overline{f}\right)^2 = \mathrm{var}\{f\} \geq 0$$

Solutions are obtained by integrating factor transformation

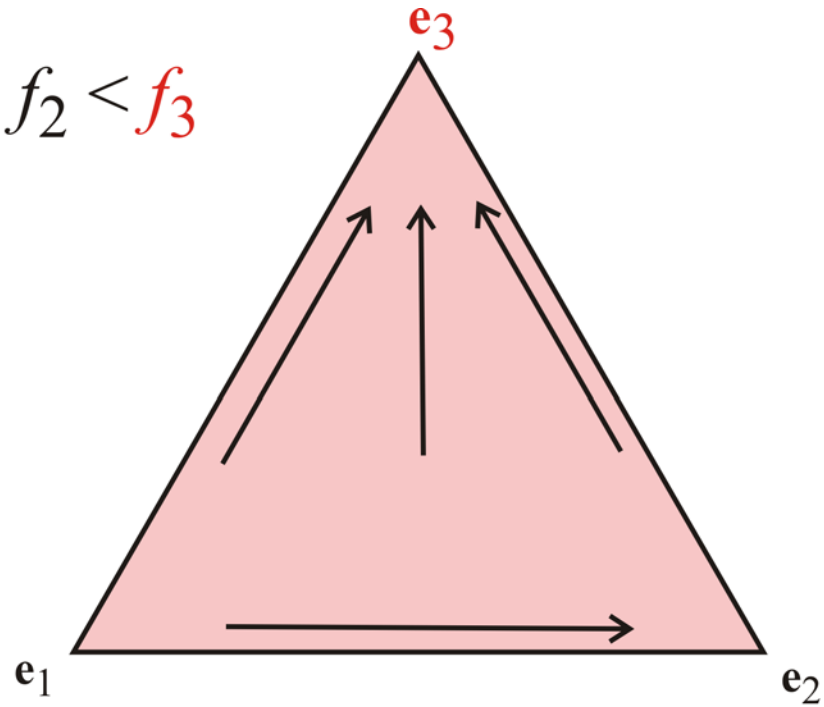
$$x_i(t) = \frac{x_i(0)\cdot \exp\left(f_i t\right)}{\sum_{j=1}^{n} x_j(0)\cdot \exp\left(f_j t\right)}; \quad i = 1,2,\cdots,n$$



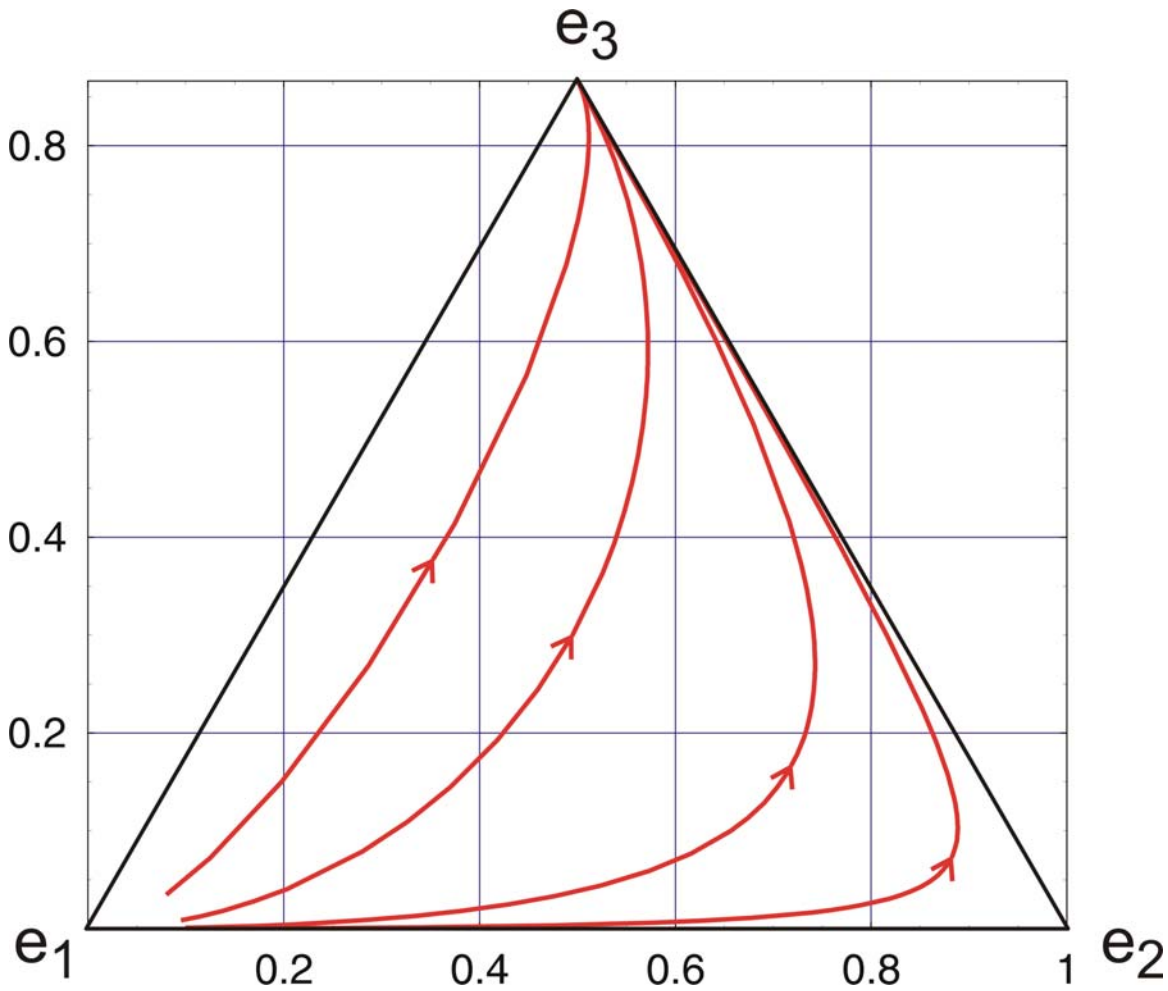
Selection between three species with $f_1 = 1$, $f_2 = 2$, and $f_3 = 3$



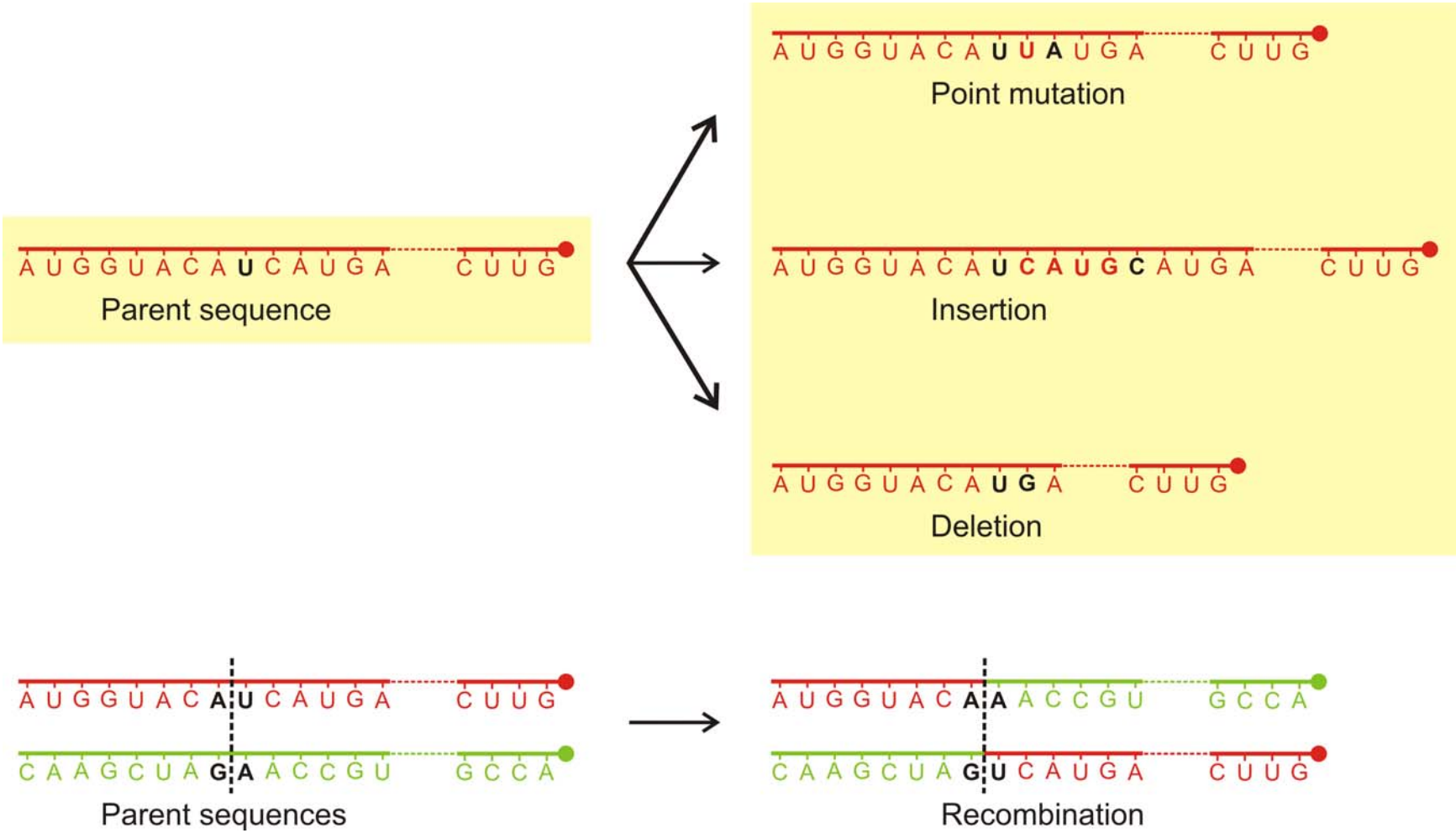
$$f_1 < f_2 < f_3$$



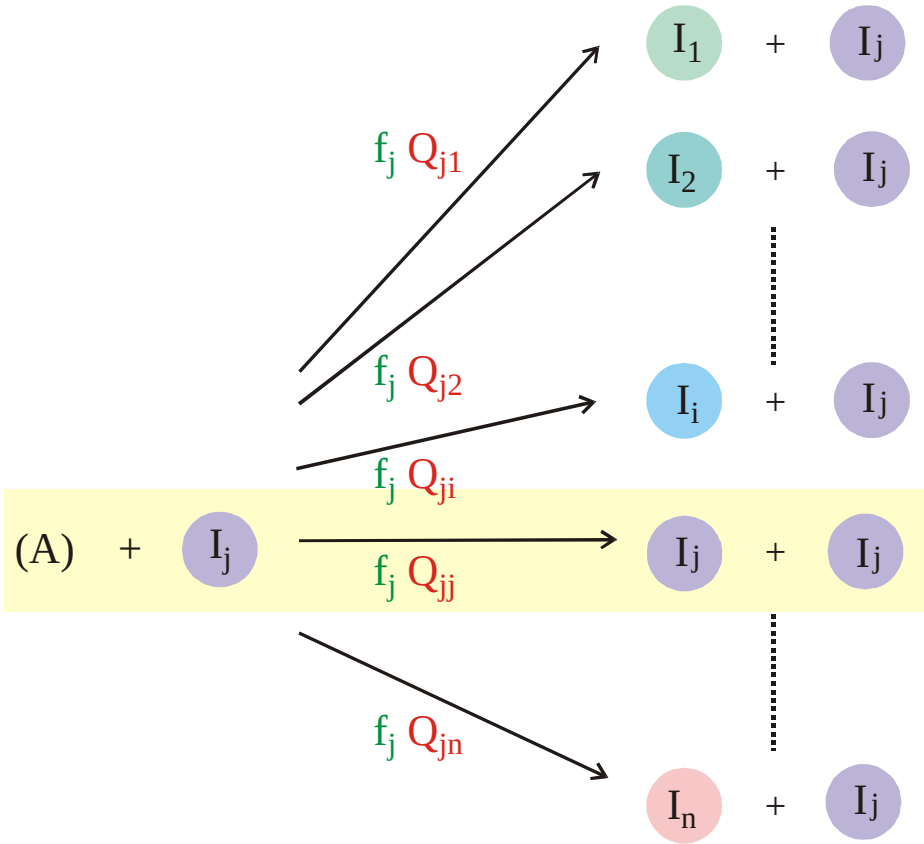
Selection on the concentration simplex: $S_3 = \left\{ x_i \geq 0,\ i = 1,2,3;\ \sum_{i=1}^{3} x_i = 1 \right\}$



Selection between three species with $f_1 = 1$, $f_2 = 2$, and $f_3 = 3$, parametric plot on S_3



Variation of genotypes through mutation and recombination



$$dx_i / dt = \Sigma_j f_j Q_{ji} x_j - x_i \Phi$$

$$\Phi = \Sigma_j f_j x_i ; \quad \Sigma_j x_j = 1 ; \quad \sum_i Q_{ij} = 1$$

$$[I_i] = x_i \geq 0 ; \quad i = 1,2,...,n ;$$

$$[A] = a = \text{constant}$$

$$Q_{ij} = (1-p)^{\ell - d(i,j)} \, p^{d(i,j)}$$

p ..... Error rate per digit

ℓ ..... Chain length of the polynucleotide

d(i,j) .... Hamming distance between I_i and I_j

Chemical kinetics of replication and mutation as parallel reactions

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

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

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

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

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

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

Matrix W and Frobenius theorem:

$$\mathrm{W} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix}$$

Primitive matrix W:

A nonnegative square matrix $\mathrm{W} = \{w_{ij}\}$ is said to be a primitive matrix if there exists k such that $\mathrm{W}^k \gg 0$, i.e., if there exists k such that for all i, j, the (i, j) entry of W^k is positive.

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

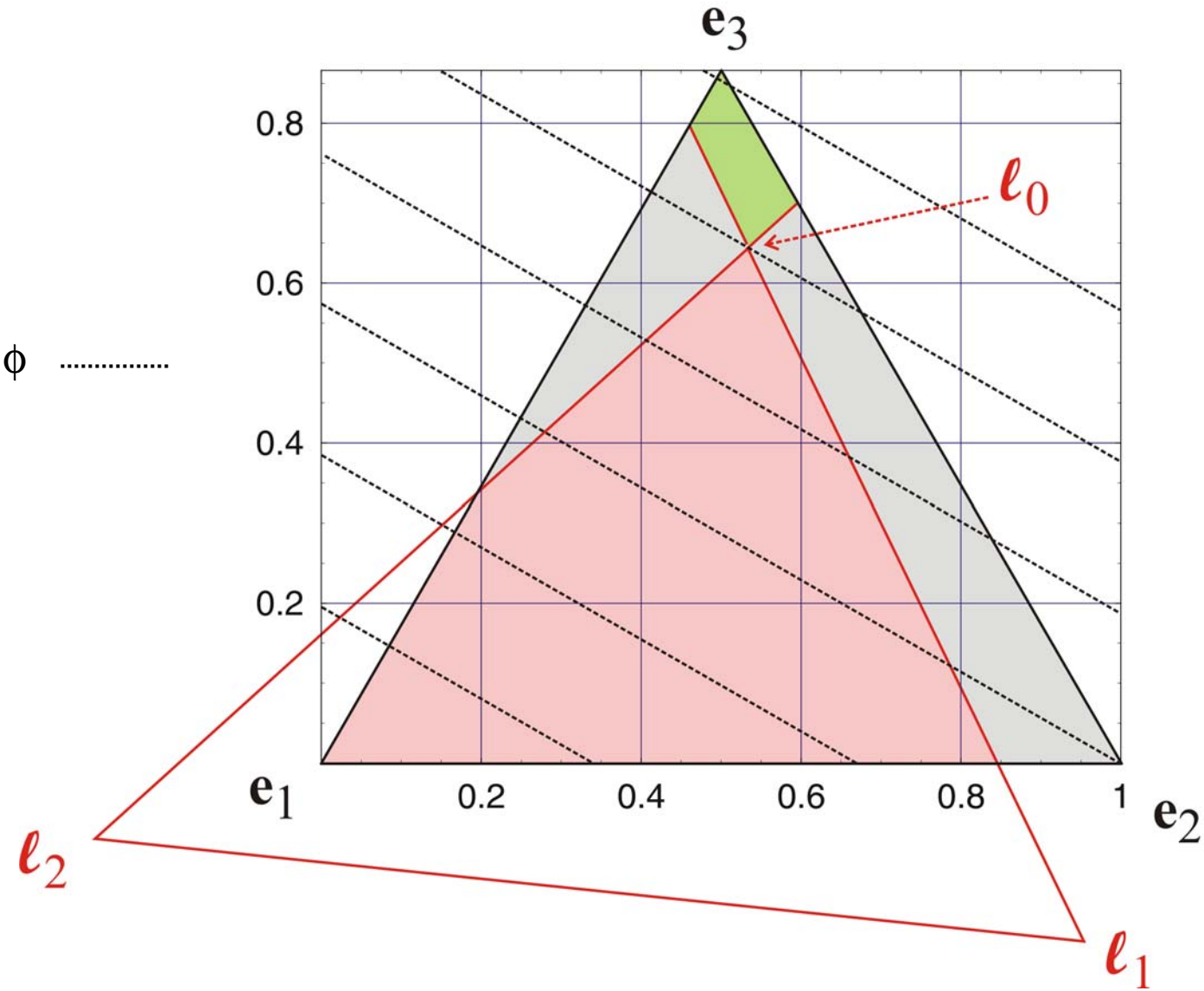
(iv) λ_0 is a simple root of the characteristic equation of W

(v-vi) etc.

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

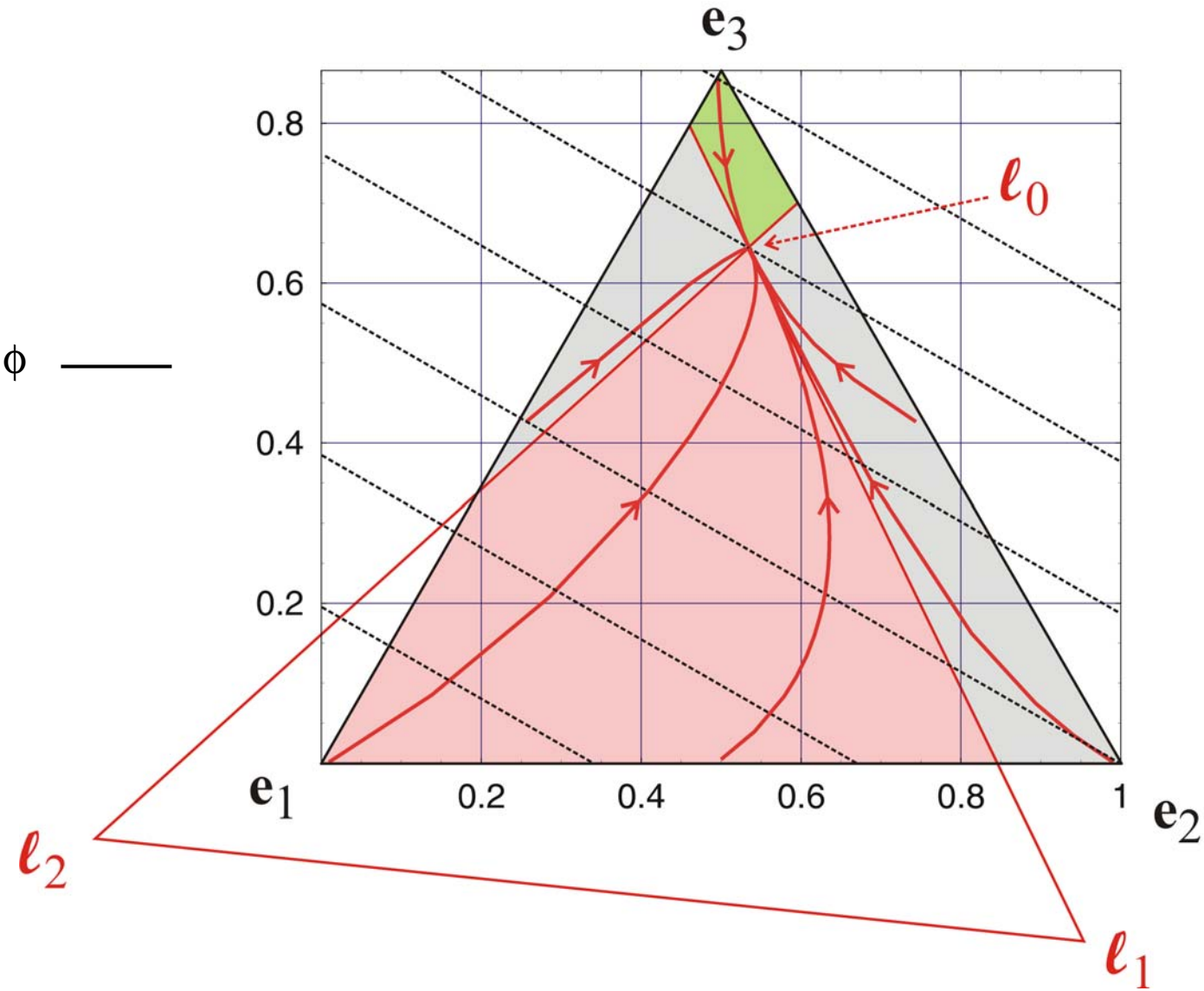
(ii) $\lambda_0 \geq |\lambda_k|$ for all $k \neq 0$

constant level sets of ϕ .....

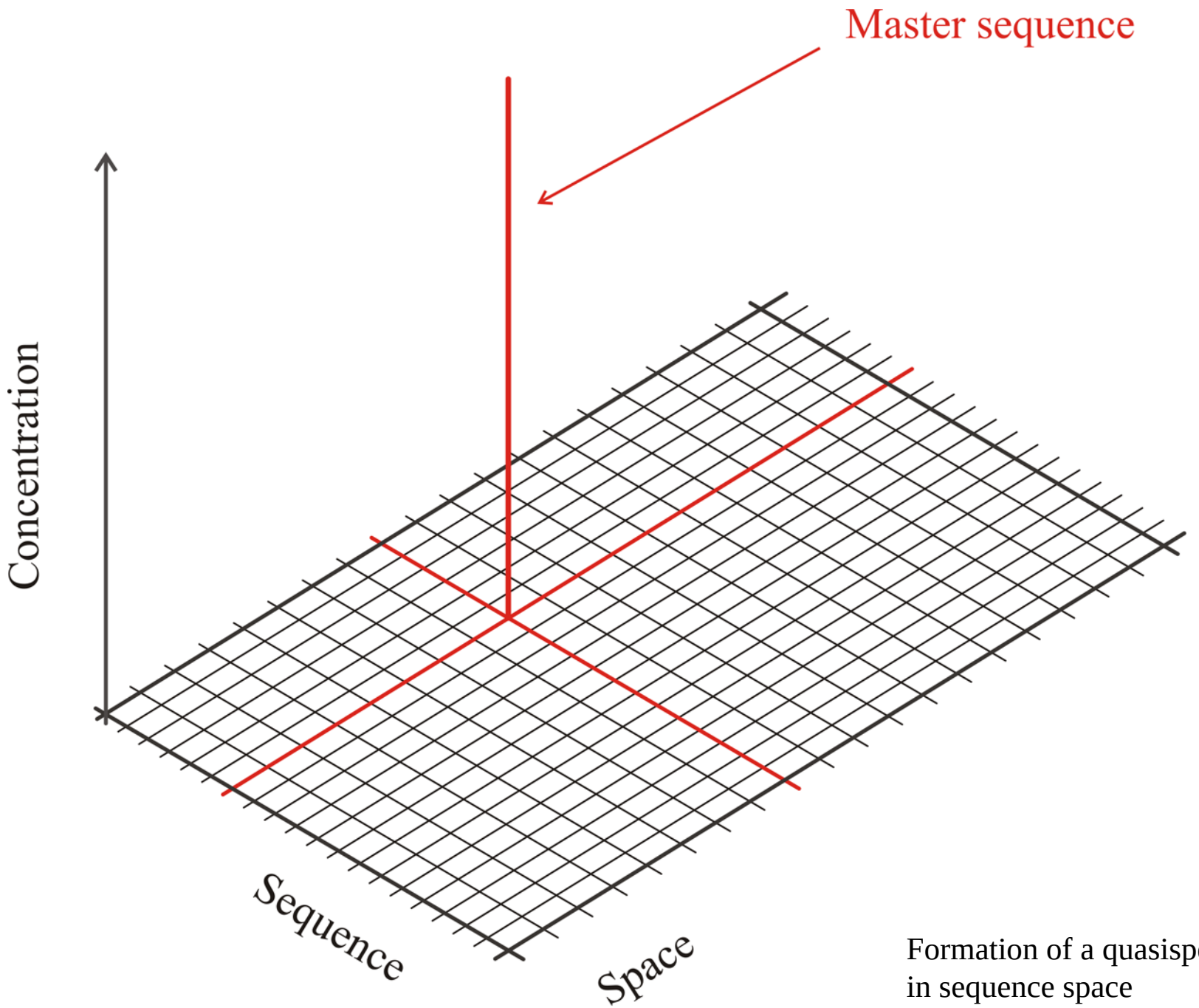


Selection of quasispecies with $f_1 = 1.9$, $f_2 = 2.0$, $f_3 = 2.1$, and $p = 0.01$

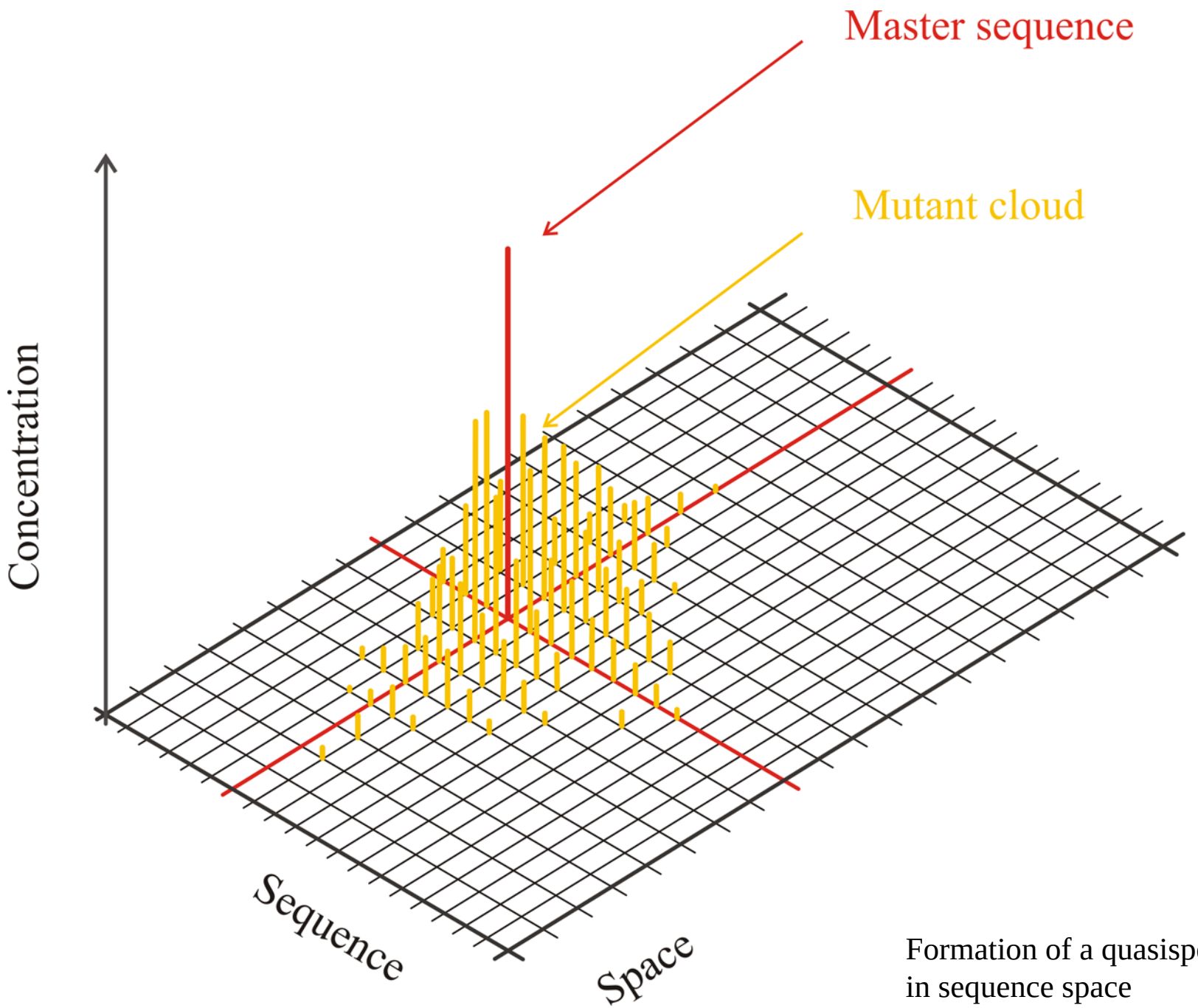
constant level sets of ϕ ——



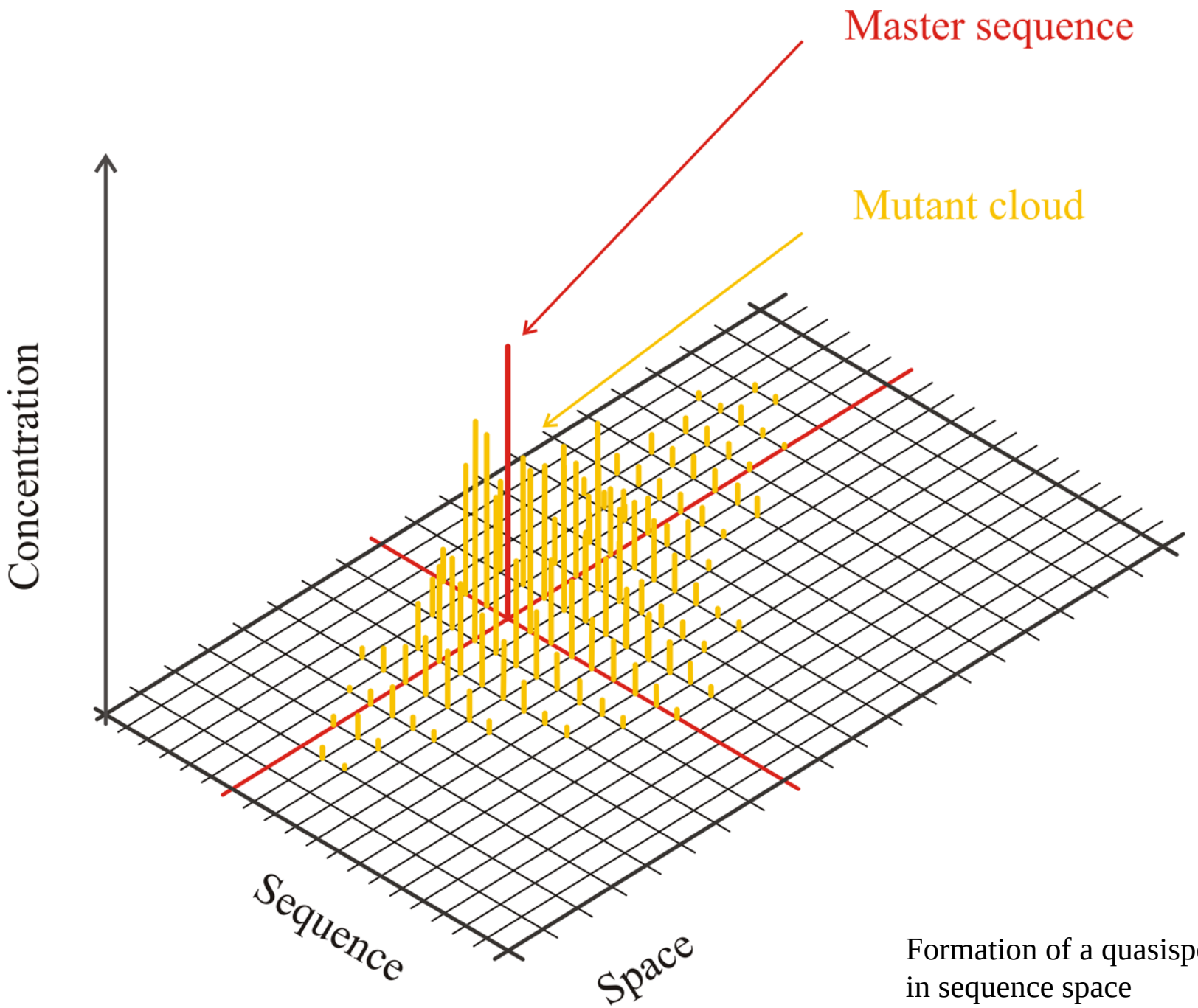
Selection of quasispecies with $f_1 = 1.9$, $f_2 = 2.0$, $f_3 = 2.1$, and $p = 0.01$, parametric plot on S_3



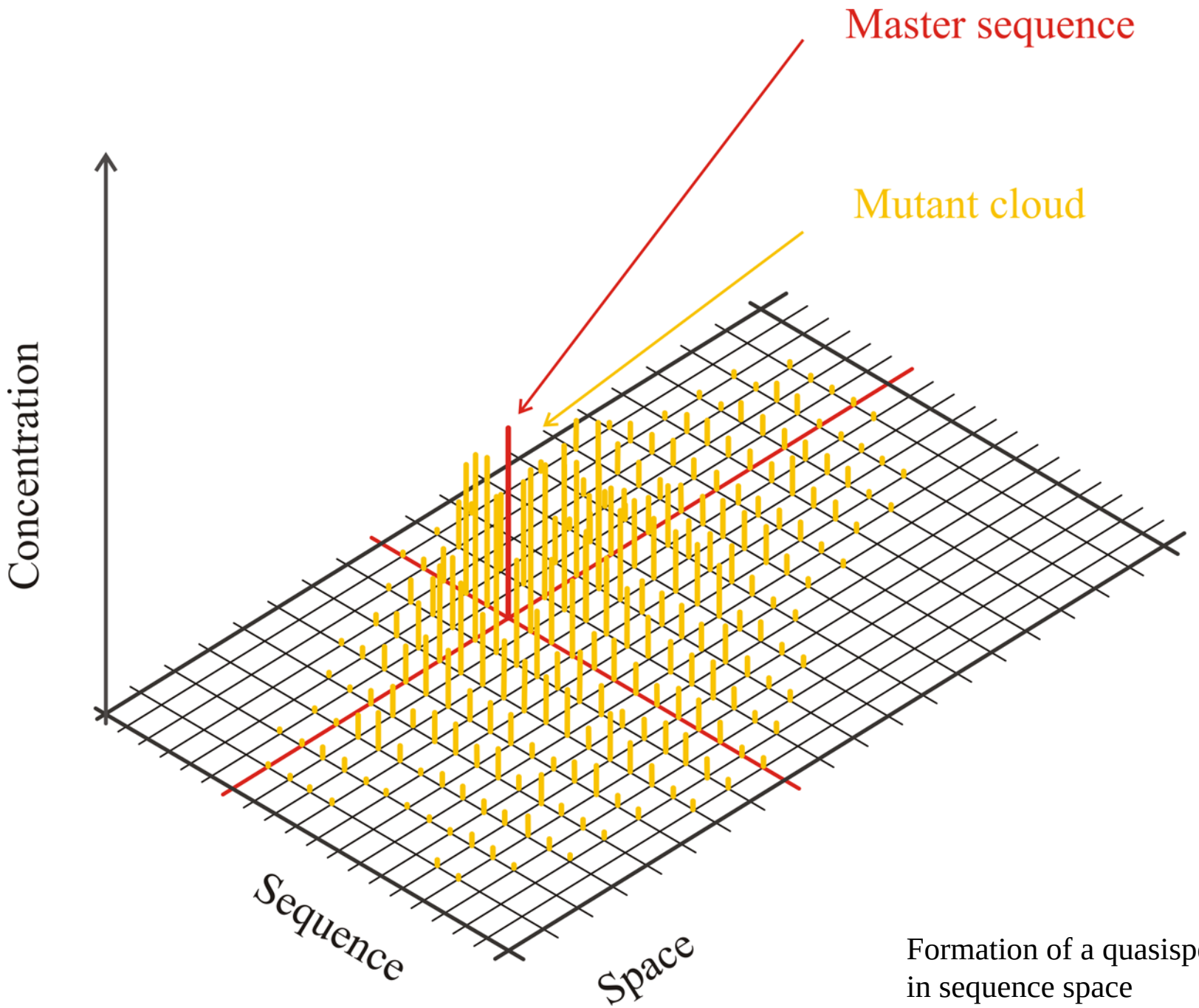
Formation of a quasispecies in sequence space



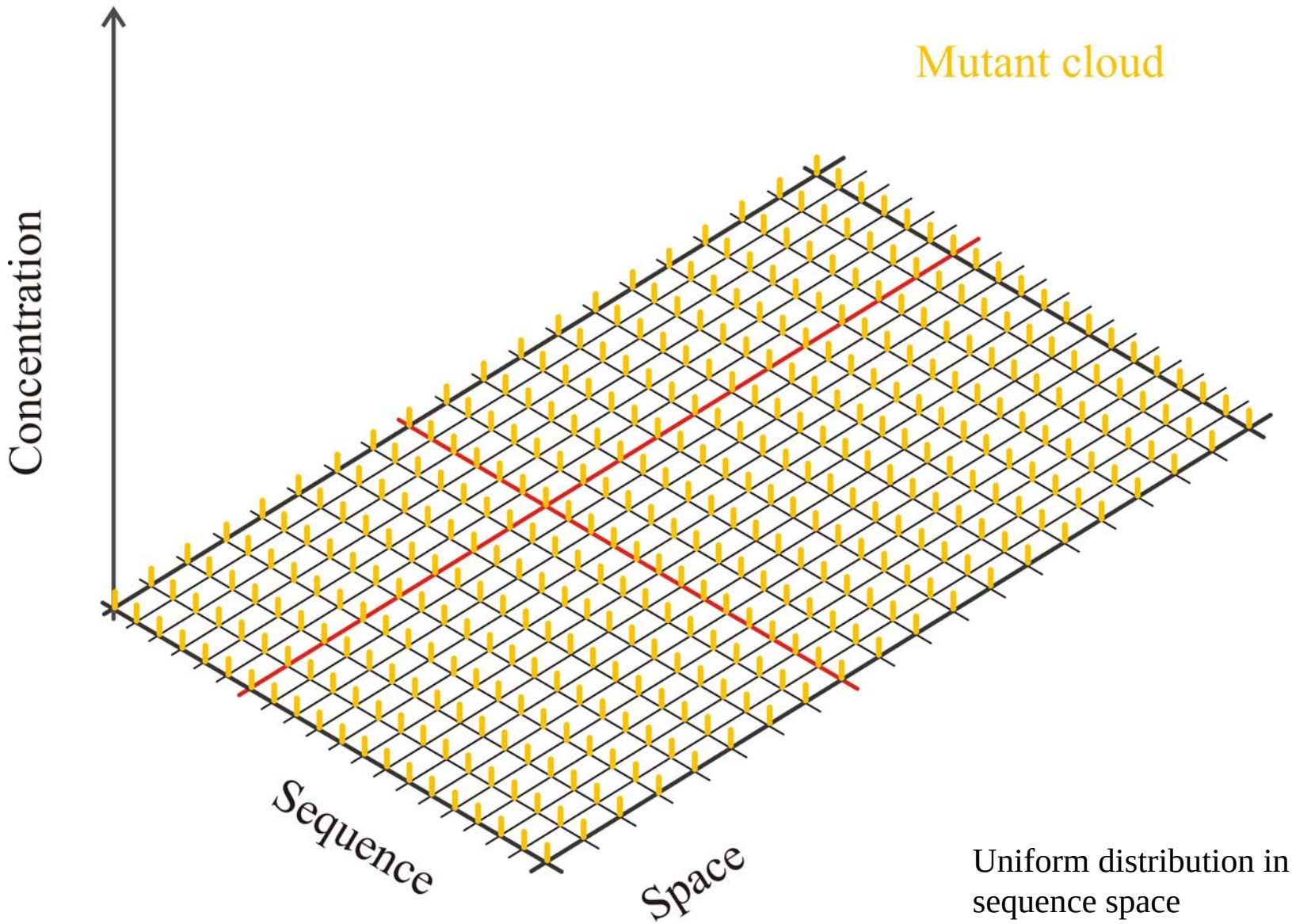
Formation of a quasispecies in sequence space



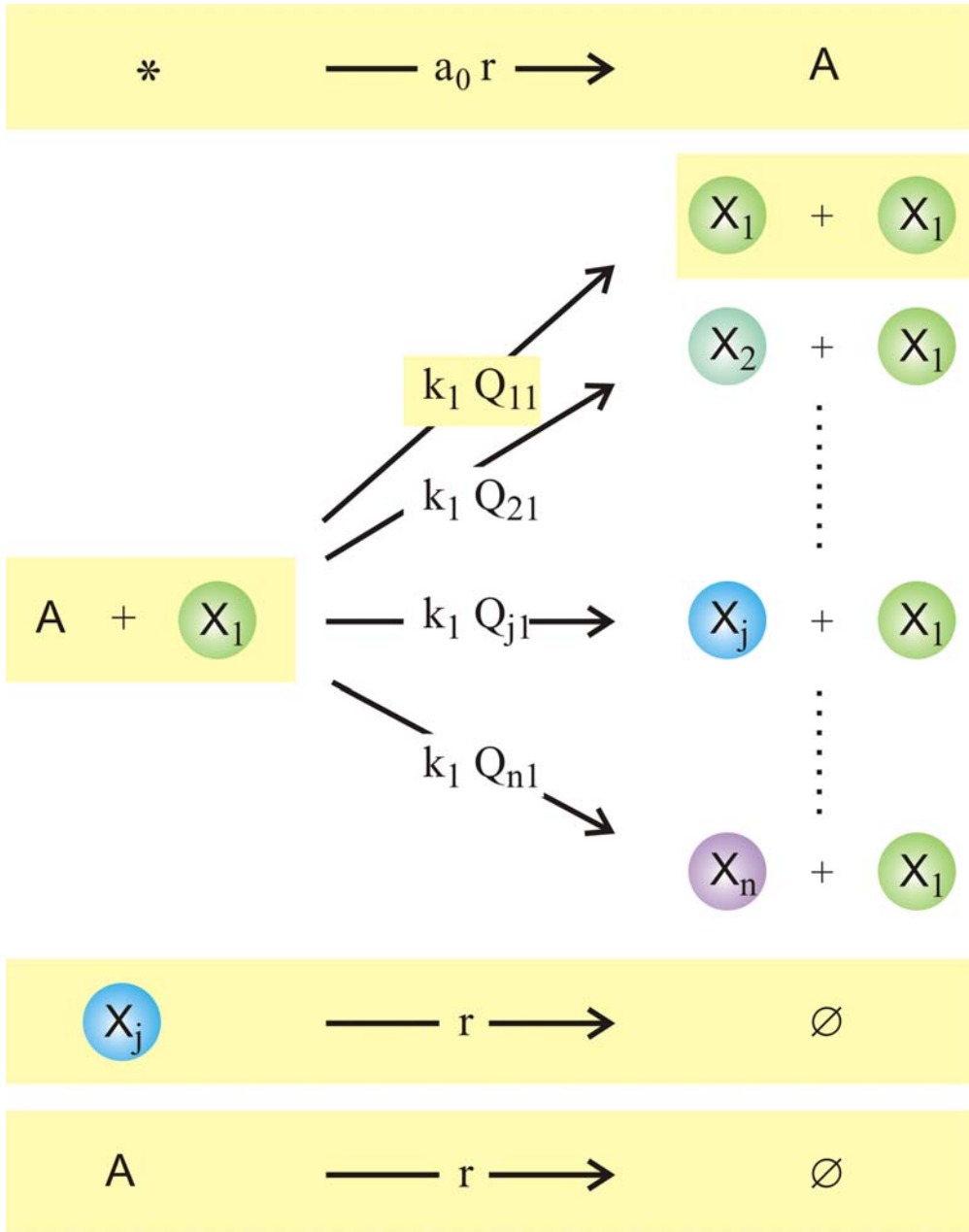
Formation of a quasispecies in sequence space



Formation of a quasispecies in sequence space



Uniform distribution in sequence space



$j = 1,2, \ldots ,n$

Lethal mutants and Frobenius theorem:

$$\mathrm{W} = \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ w_{21} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & 0 & \dots & 0 \end{pmatrix} = w_{11} \begin{pmatrix} 1 & 0 & \dots & 0 \\ \frac{w_{21}}{w_{11}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_{n1}}{w_{11}} & 0 & \dots & 0 \end{pmatrix}$$

$$\mathrm{W}^k = w_{11}^k \begin{pmatrix} 1 & 0 & \dots & 0 \\ \frac{w_{21}}{w_{11}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_{n1}}{w_{11}} & 0 & \dots & 0 \end{pmatrix}$$

$$
\begin{aligned}
* &\xrightarrow{a_0 \cdot r} \mathbf{A} && : \text{influx} \\
\mathbf{A} + \mathbf{X}_1 &\xrightarrow{k_1 Q_{11}} 2\,\mathbf{X}_1\ ; && : \text{replication} \\
\mathbf{A} + \mathbf{X}_1 &\xrightarrow{k_1 Q_{j1}} \mathbf{X}_1 + \mathbf{X}_j \quad j = 2, \ldots, n && : \text{mutation} \\
\mathbf{A} &\xrightarrow{r} \emptyset && : \text{outflux} \\
\mathbf{X}_j &\xrightarrow{r} \emptyset\ ; \quad j = 1, \ldots, n && : \text{outflux}
\end{aligned}
$$

$$
\begin{aligned}
\frac{da}{dt} &= -\,a \sum_{j=1}^{n} k_1 Q_{j1}\, x_1 \;+\; r\,(a_0 - a) \;=\; -\,a\,k_1\,x_1 \;+\; r\,(a_0 - a) \\
\frac{dx_j}{dt} &= a\,Q_{j1}\,x_1 \;-\; r\,x_j
\end{aligned}
$$

Stationary solutions: 1. active state

$$r < k_1 Q_{11} a_0$$

$$\tilde{a} = \frac{r}{k_1 Q_{11}}$$

$$\tilde{x}_1 = Q_{11}(a_0 - \tilde{a}) = Q_{11} a_0 - \frac{r}{k_1}$$

$$\tilde{x}_j = Q_{j1}(a_0 - \tilde{a}) = Q_{j1}\left(a_0 - \frac{r}{k_1 Q_{11}}\right); \quad j = 2, 3, \ldots, n$$

Stationary solutions: 2. extinction

$$r > k_1 Q_{11} a_0$$

$$\tilde{a} = a_0$$

$$\tilde{x}_j = 0; \; j = 1, 2, \ldots, n$$

Find $r(t)$ such that $a(t) = \bar{a} = const.$

$$\frac{da}{dt} \;=\; 0 \;=\; -\,\bar{a} \sum_{j=1}^{n} k_1 Q_{j1}\, x_1 \;+\; r(t)\,(a_0 - \bar{a})$$

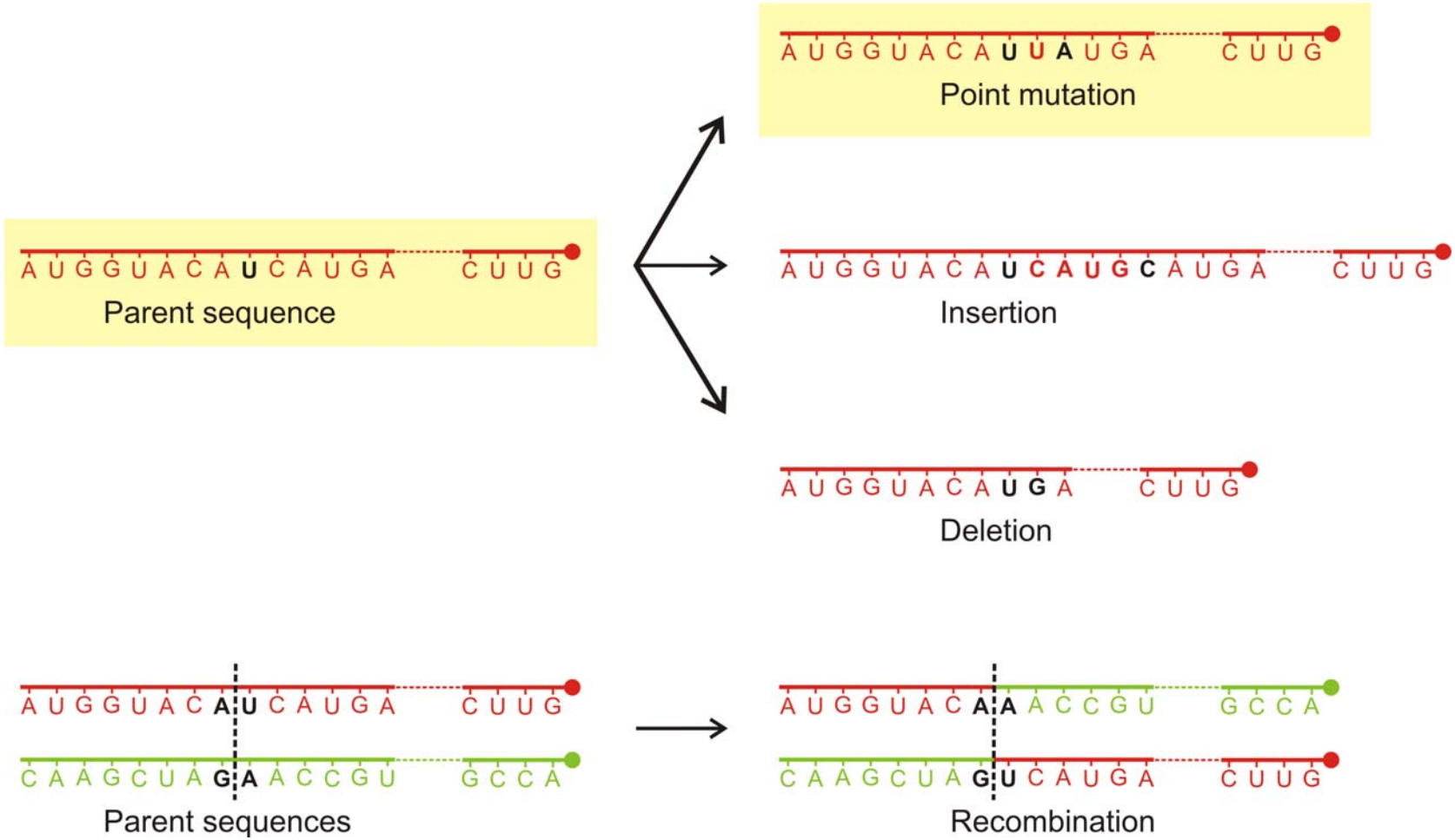
$$r(t) \;=\; \frac{\bar{a}}{a_0 - \bar{a}}\, k_1\, x_1\,;\;\; f_1 \;=\; k_1\,\bar{a}\,;\;\; \sum_{i=1}^{n} x_i \;=\; c \;=\; a_0 - \bar{a}$$

$$\frac{dx_j}{dt} \;=\; f_1 Q_{j1}\, x_1 \;-\; x_j\, \frac{f_1\, x_1}{\sum_{i=1}^{n} x_i} \;=\; f_1\, x_1 \left(Q_{j1} \;-\; \frac{x_j}{c} \right)$$

Stationary solutions:

$$\bar{x}_j \;=\; Q_{j1} \sum_{i=1}^{n} \bar{x}_i \;=\; Q_{ji}\, c$$

2. Crossing the error threshold

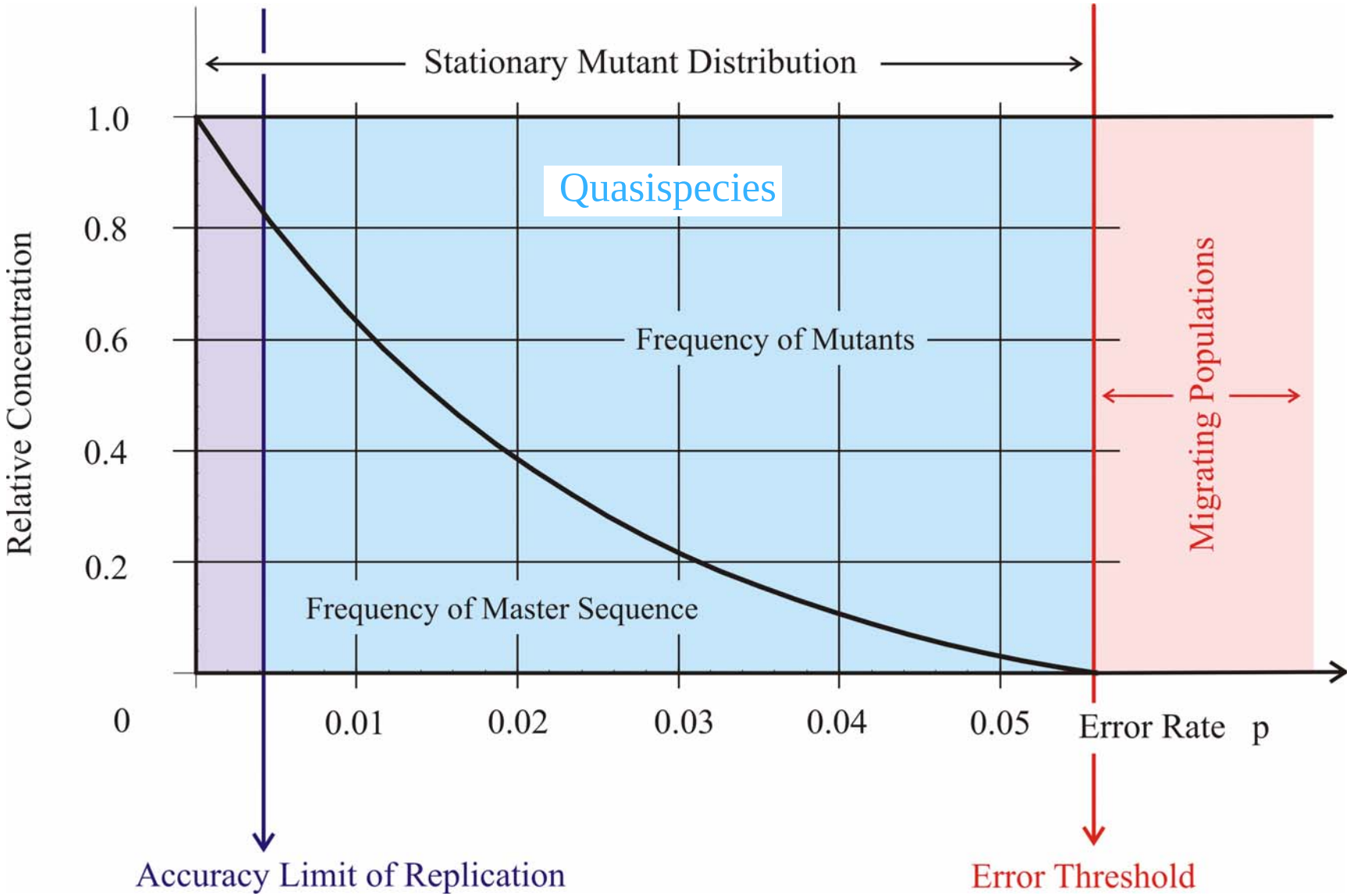


Variation of genotypes through mutation and recombination

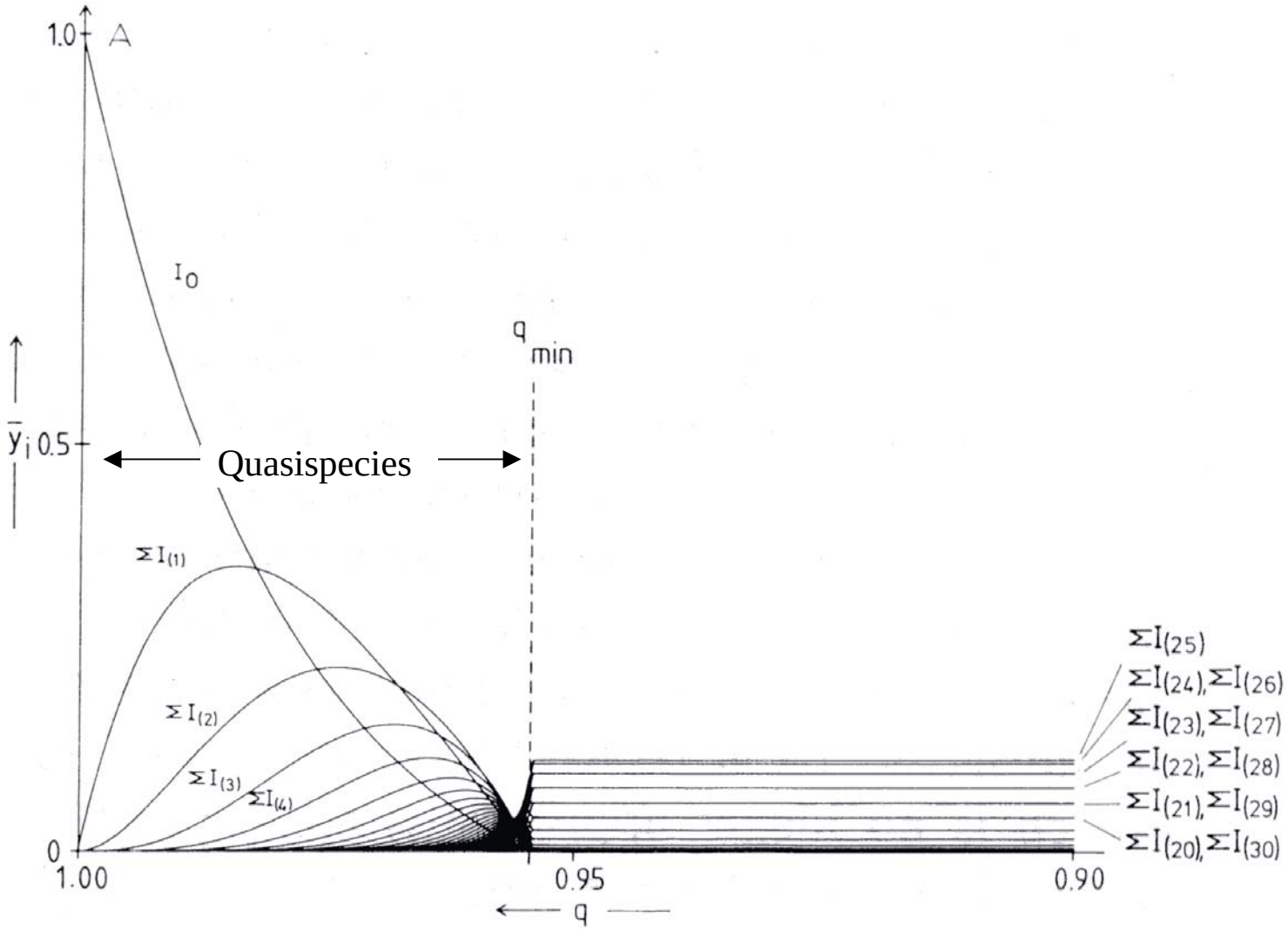
Uniform error rate model:

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

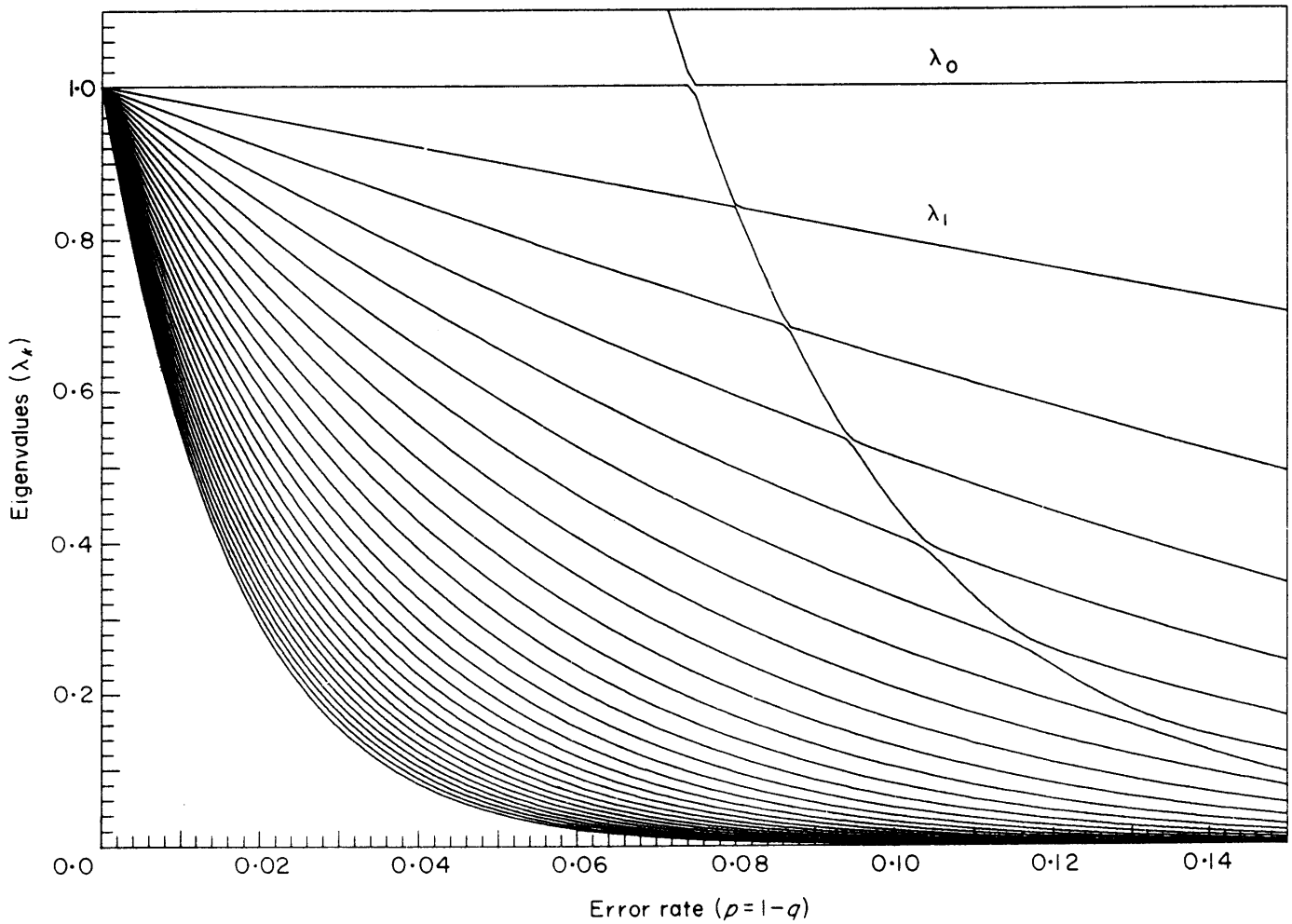
$d_H(\mathbf{X}_i, \mathbf{X}_j)$ ... Hamming distance



The error threshold in replication



Quasispecies as a function of the replication accuracy q


λ_0
λ_1
1·0
0·8
0·6
0·4
0·2
0·0
Eigenvalues (λ_k)
0·02
0·04
0·06
0·08
0·10
0·12
0·14
Error rate ($p = 1-q$)

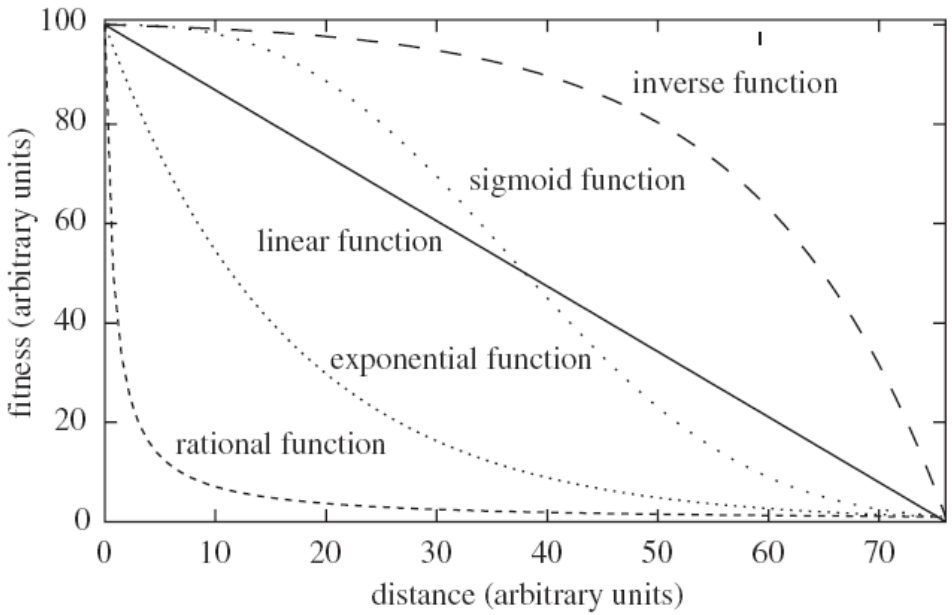
(1) linear $f^1_{scale}(d) = 100(1 - d/l)$,

(2) exponential $f^2_{scale}(d) = 100^{1-d/l}$,

(3) rational $f^3_{scale}(d) = \dfrac{1}{0.01 + d/l}$,

(4) sigmoid $f^4_{scale}(d) = 100^{1-(d/l)^\sigma}$,

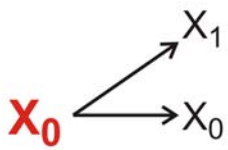
(5) inverse $f^5_{scale}(d) = 100 - 100^{d/l} + 1$.



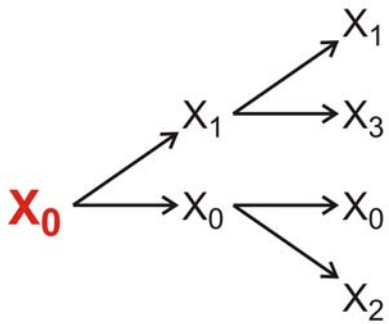
Anne Kupczok, Peter Dittrich, Determinats of simulated RNA evolution.
J.Theor.Biol. **238**:726-735, 2006

3. Stochasticity

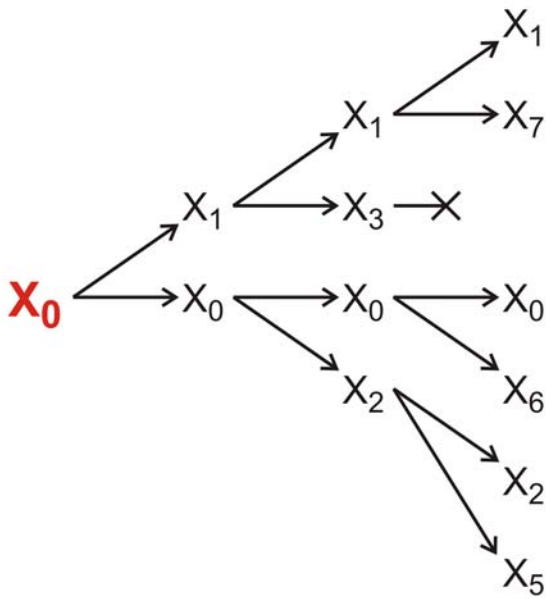
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



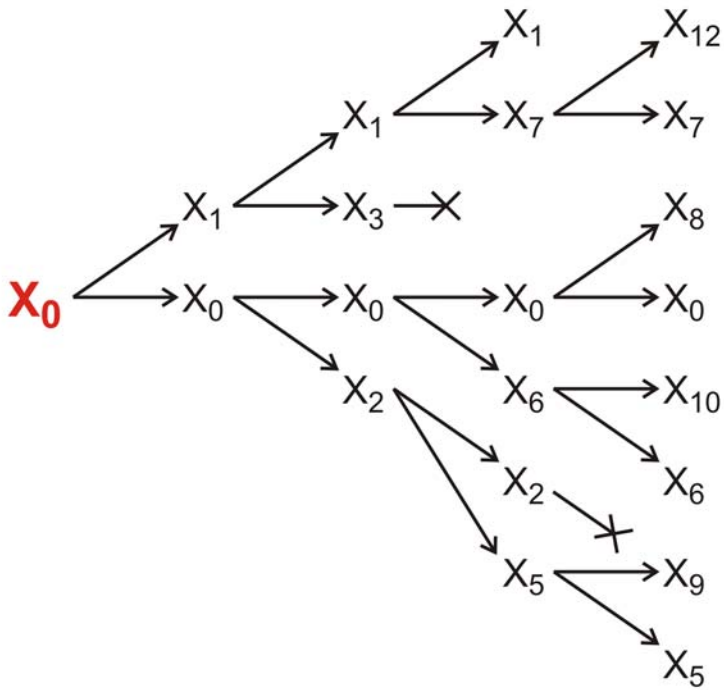
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



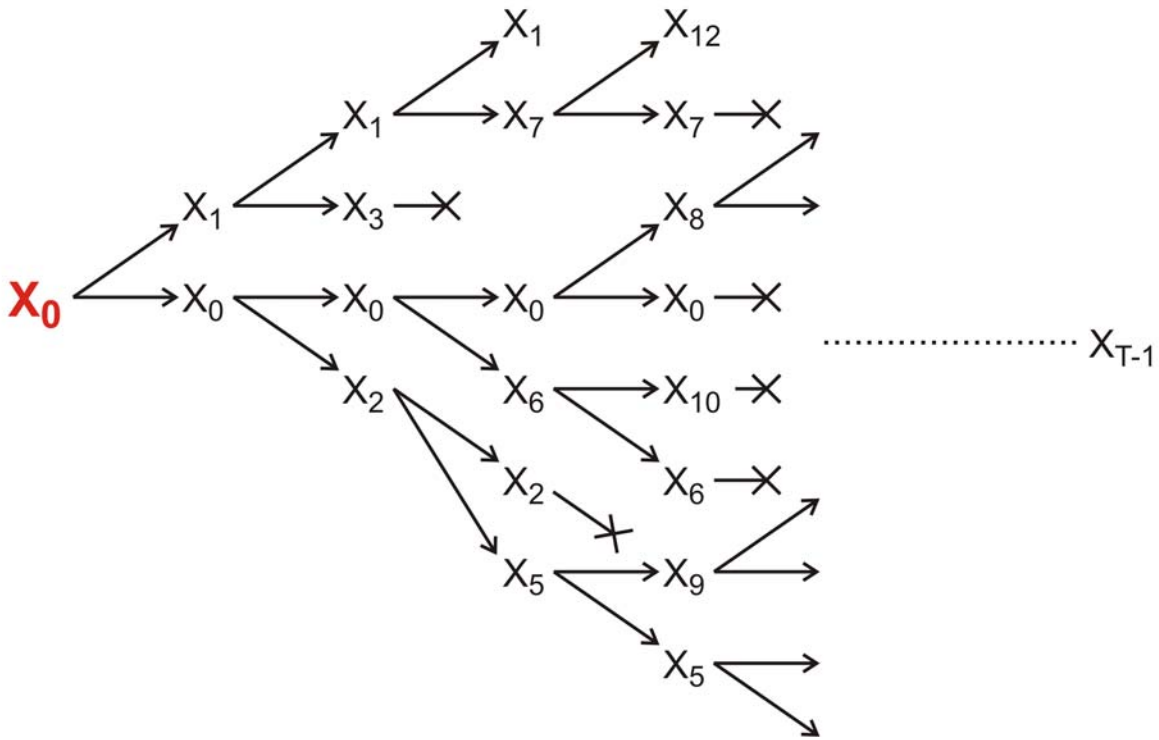
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



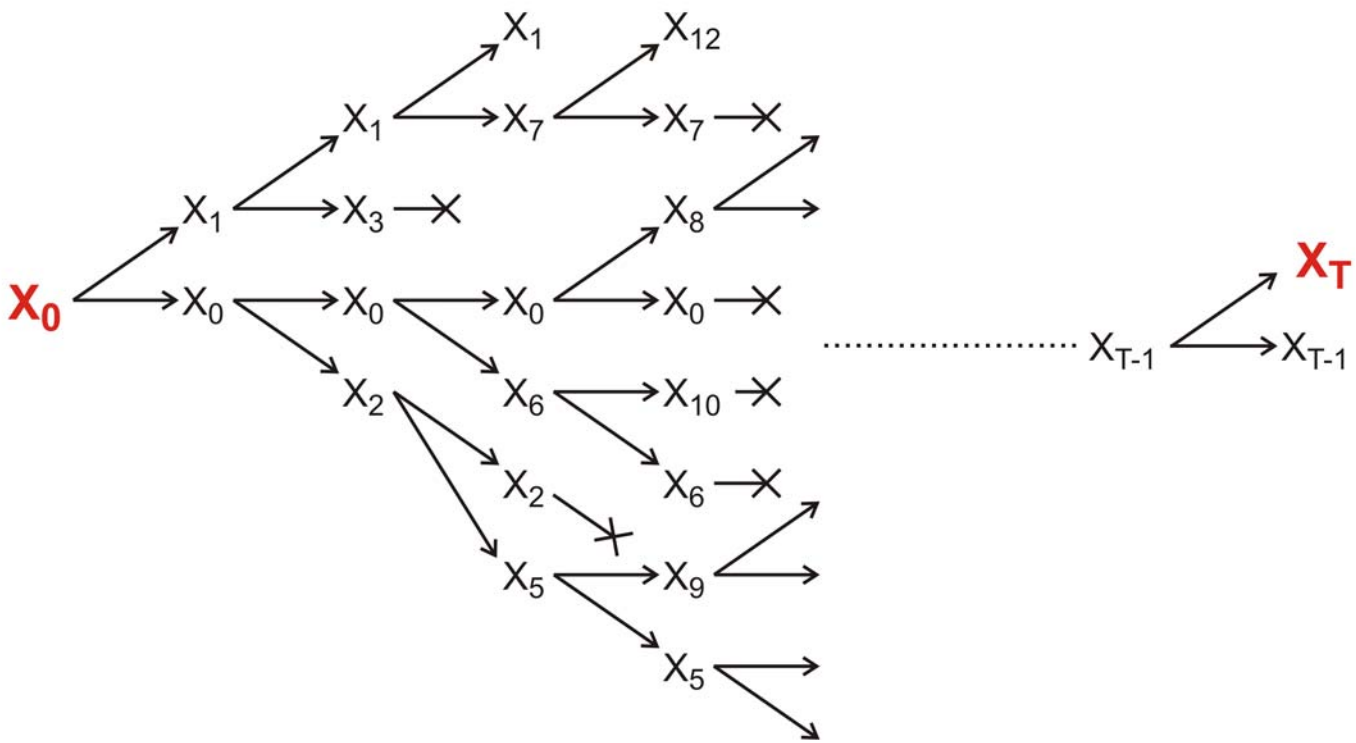
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



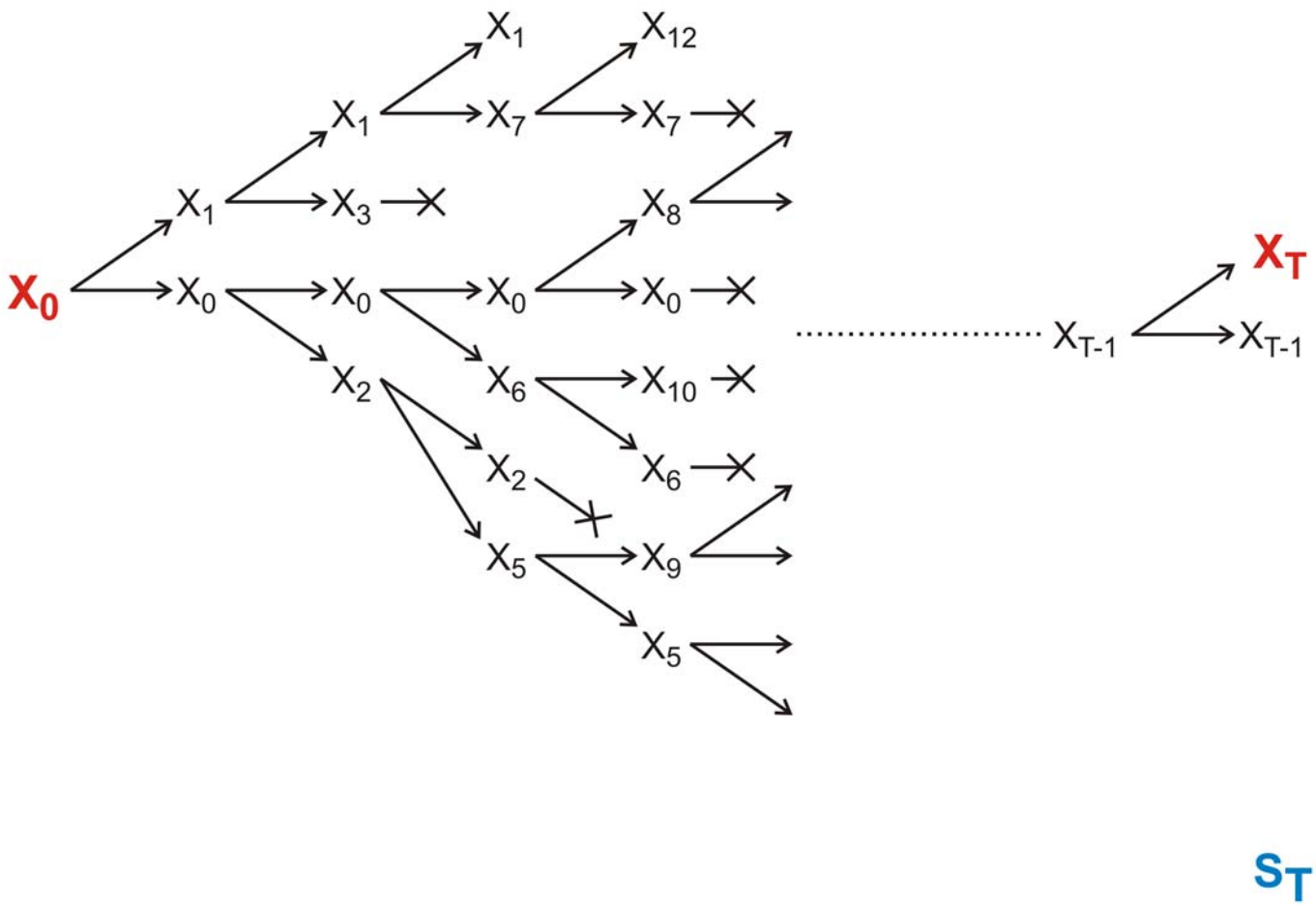
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



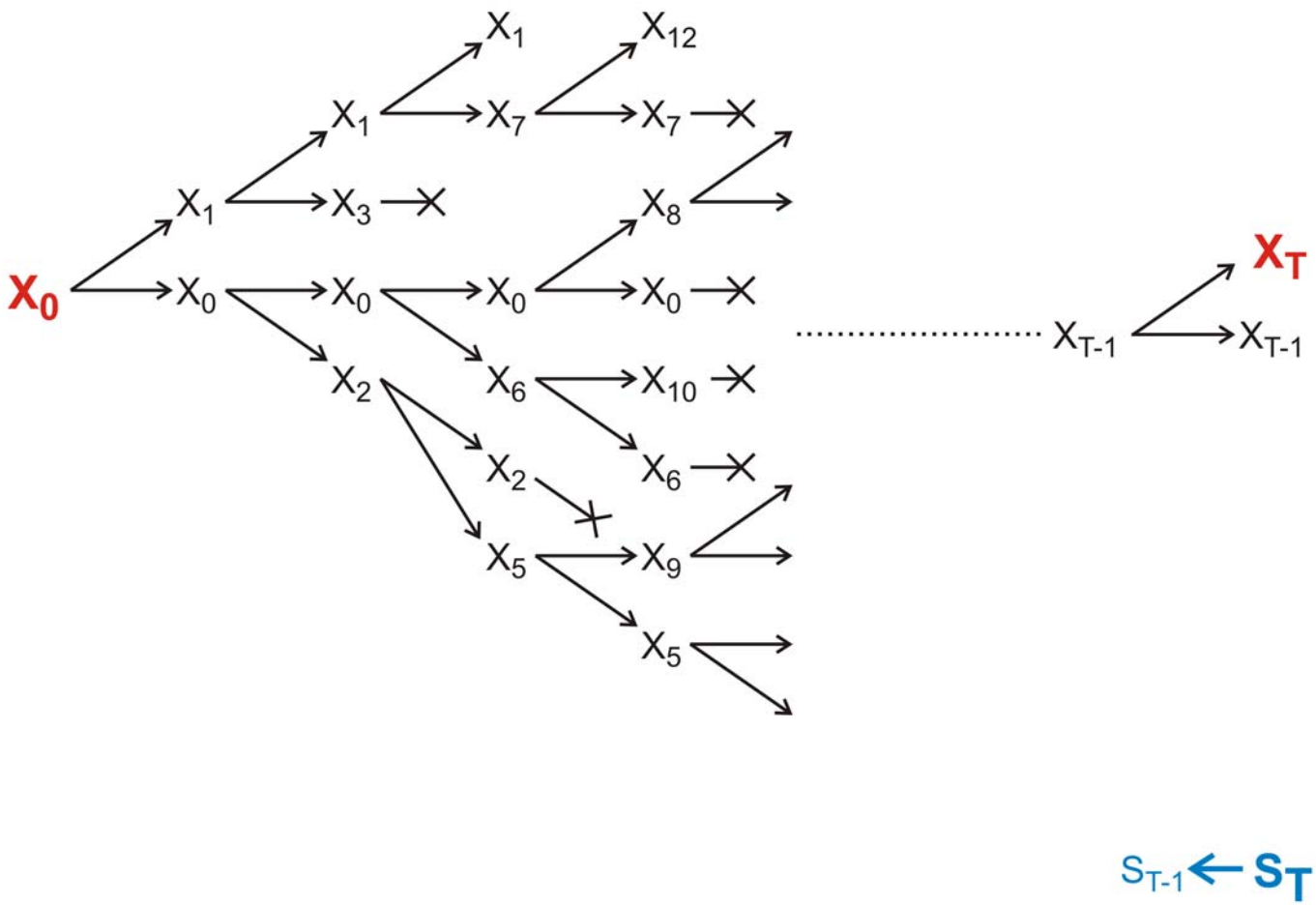
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



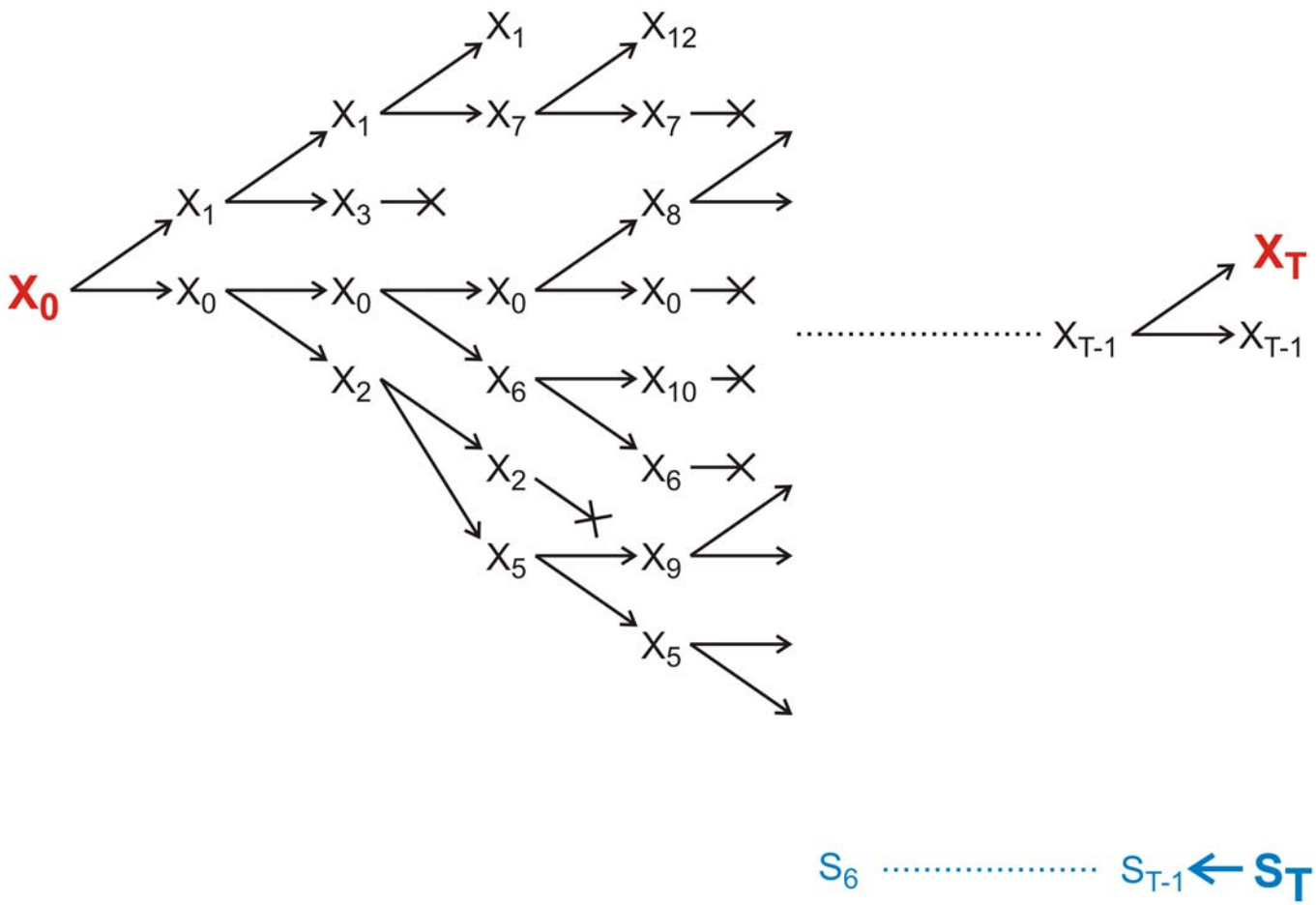
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



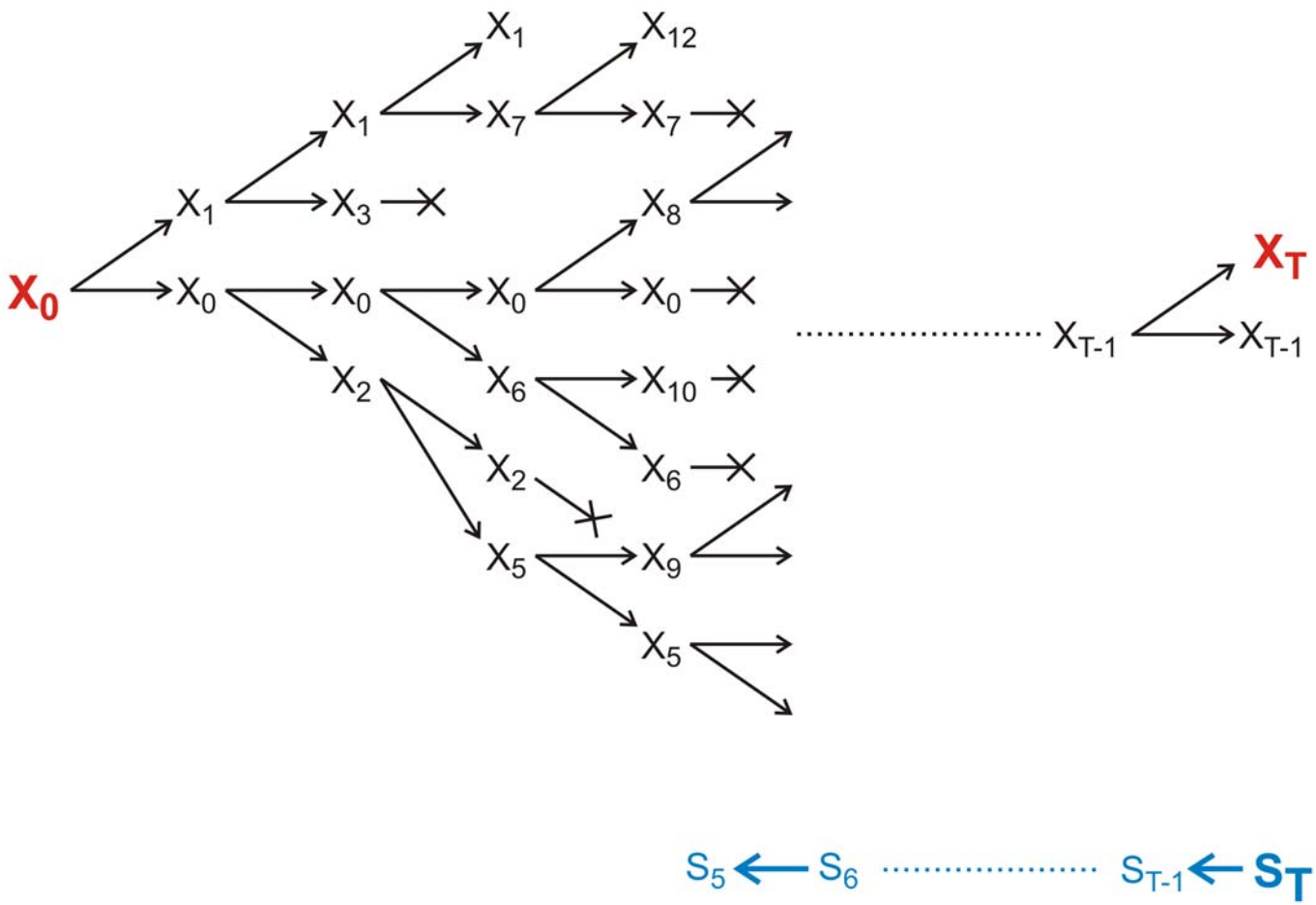
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



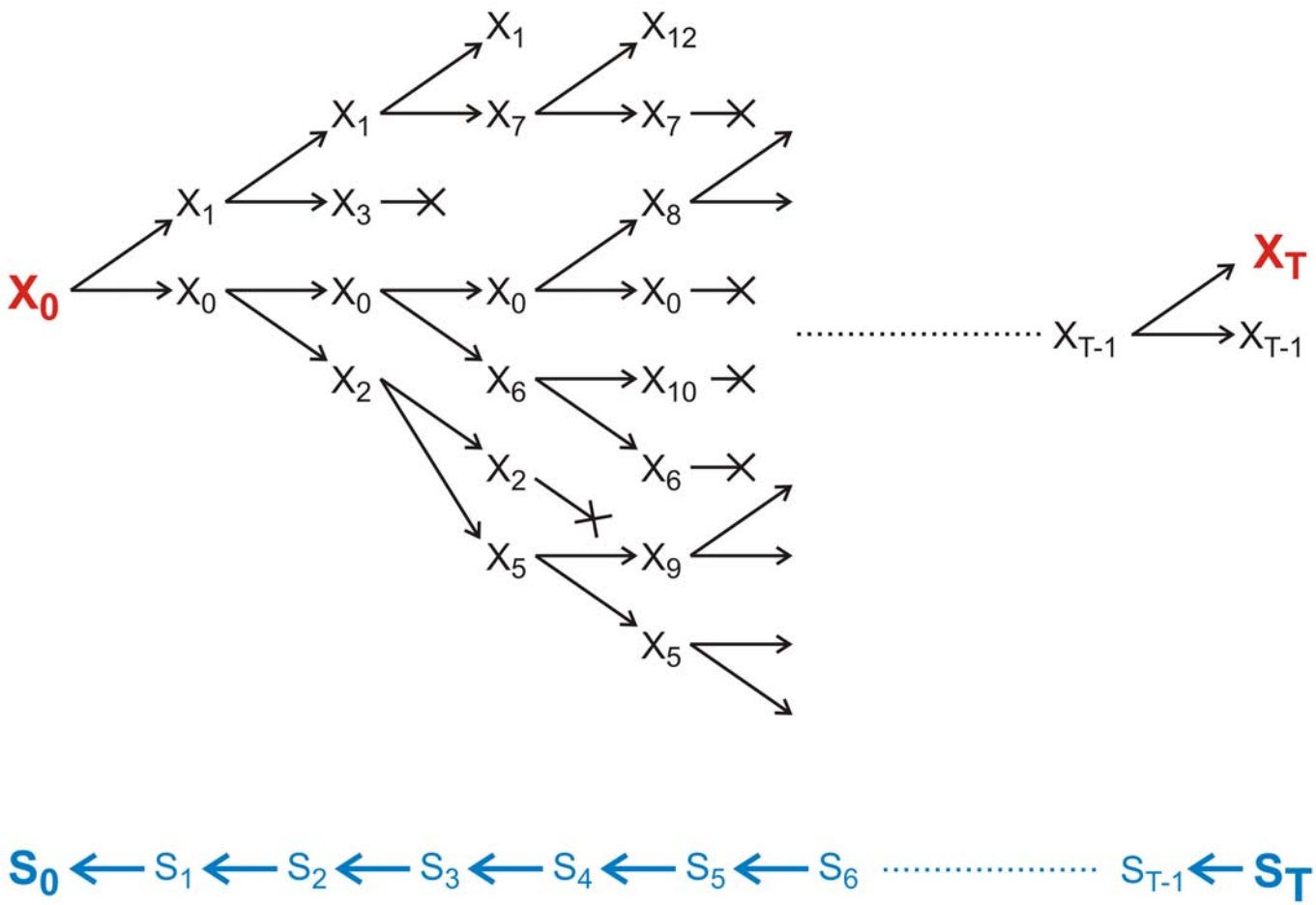
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



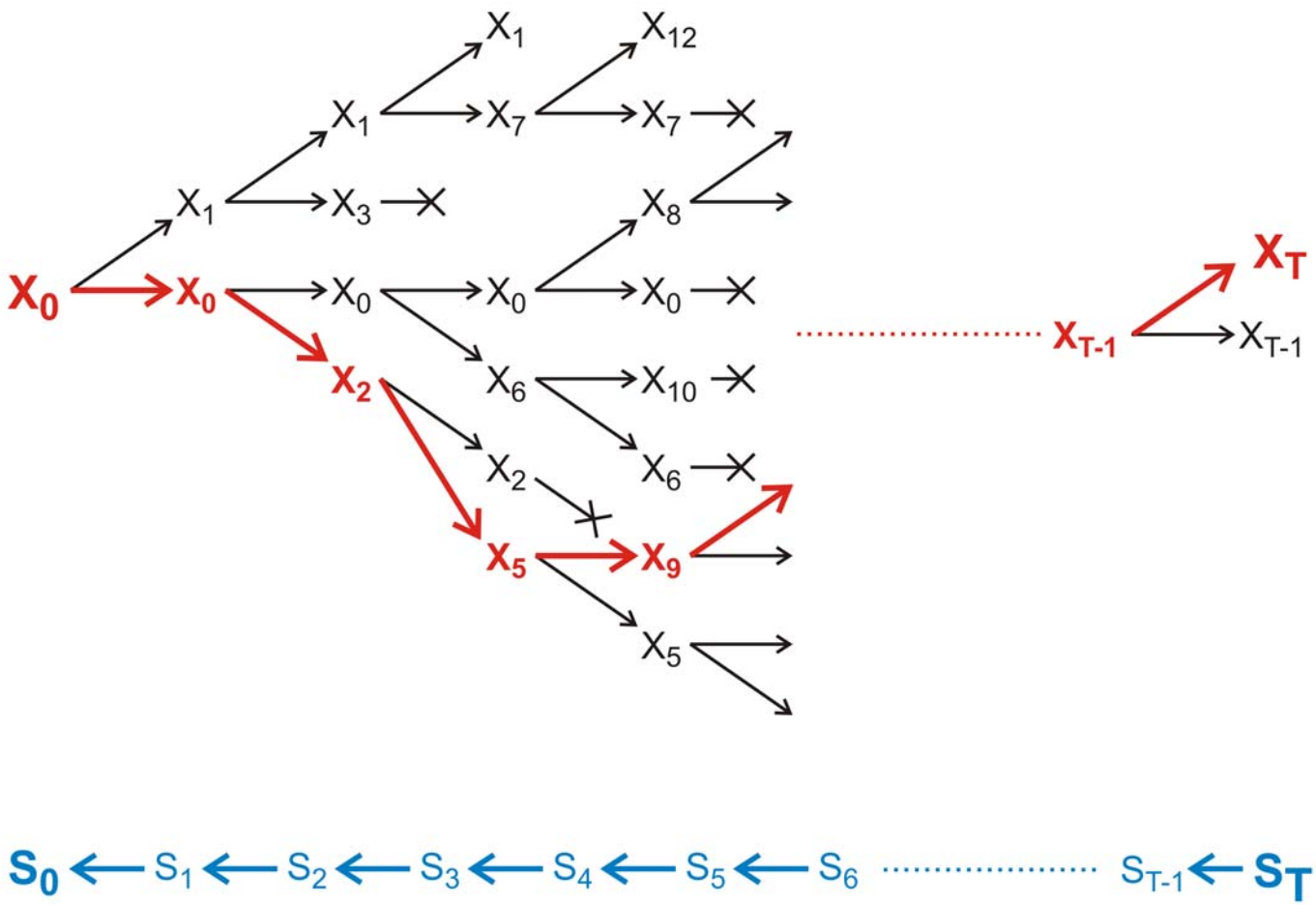
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series

Evolution *in silico*

W. Fontana, P. Schuster,
Science **280** (1998), 1451-1455

random individuals. The primer pair used for genomic DNA amplification is 5′-TCTCCCTGGATTCT-CATTTA-3′ (forward) and 5′-TCTTTGTCTTCTGT-TCCACC-3′ (reverse). Reactions were performed in 25 μl using 1 unit of Taq DNA polymerase with each primer at 0.4 μM; 200 μM each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM tris-HCl (pH 8.3), 50 mM KCl_2,1.5 mM $MgCl_2$] in a cycle condition of 94°C for 1 min and then 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by 72°C for 6 min. PCR products were purified (Qiagen), digested with Xmn I, and separated in a 2% agarose gel.

32. A nonsense mutation may affect mRNA stability and result in degradation of the transcript [L. Maquat, *Am. J. Hum. Genet.* **59**, 279 (1996)].
33. Data not shown; a dot blot with poly (A)$^+$ RNA from 50 human tissues (The Human RNA Master Blot, 7770-1, Clontech Laboratories) was hybridized with a probe from exons 29 to 47 of *MYO15* using the same condition as Northern blot analysis (*13*).
34. Smith–Magenis syndrome (SMS) is due to deletions of 17p11.2 of various sizes, the smallest of which includes *MYO15* and perhaps 20 other genes [(*6*); K-S Chen, L. Potocki, J. R. Lupski, *MRDD Res. Rev.* **2**, 122 (1996)]. *MYO15* expression is easily detected in the pituitary gland (data not shown). Haploinsufficiency for *MYO15* may explain a portion of the SMS phenotype such as short stature. Moreover, a few SMS patients have sensorineural hearing loss, possibly because of a point mutation in *MYO15* in trans to the SMS 17p11.2 deletion.
35. R. A. Fridell, data not shown.
36. K. B. Avraham *et al.*, *Nature Genet.* **11**, 369 (1995); X-Z. Liu *et al.*, *ibid.* **17**, 268 (1997); F. Gibson *et al.*, *Nature* **374**, 62 (1995); D. Weil *et al.*, *ibid.*, p. 60.
37. RNA was extracted from cochlea (membranous labyrinths) obtained from human fetuses at 18 to 22 weeks of development in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Only samples without evidence of degradation were pooled for poly (A)$^+$ selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-for-PCR kit (Clontech Laboratories). A portion of the first-strand cDNA (4%) was amplified by PCR with Advantage cDNA polymerase mix (Clontech Laboratories) using human *MYO15*-specific oligonucleotide primers (forward, 5′-GCATGACCTGCCGGCTAAT-GGG-3′; reverse, 5′-CTCACGGCTTCTGCATGGT-GCTCGGCTGGC-3′). Cycling conditions were 40 s at 94°C; 40 s at 66°C (3 cycles), 60°C (5 cycles), and 55°C (29 cycles); and 45 s at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 688-bp PCR product is expected from amplification of the human *MYO15* cDNA. Amplification of human genomic DNA with this primer pair would result in a 2903-bp fragment.

38. We are grateful to the people of Bengkala, Bali, and the two families from India. We thank J. R. Lupski and K.-S. Chen for providing the human chromosome 17 cosmid library. For technical and computational assistance, we thank N. Dietrich, M. Fergusson, A. Gupta, E. Sorbello, R. Torkzadeh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Sternberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and T. Barber, S. Sullivan, E. Green, D. Drayna, and J. Battey for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (Z01 DC 00035-01 and Z01 DC 00038-01 to T.B.F. and E.R.W. and R01 DC 03402 to C.C.M.), the National Institute of Child Health and Human Development (R01 HD30428 to S.A.C.) and a National Science Foundation Graduate Research Fellowship to F.J.P. This paper is dedicated to J. B. Snow Jr. on his retirement as the Director of the NIDCD.

9 March 1998; accepted 17 April 1998

Continuity in Evolution: On the Nature of Transitions

Walter Fontana and Peter Schuster

To distinguish continuous from discontinuous evolutionary change, a relation of nearness between phenotypes is needed. Such a relation is based on the probability of one phenotype being accessible from another through changes in the genotype. This nearness relation is exemplified by calculating the shape neighborhood of a transfer RNA secondary structure and provides a characterization of discontinuous shape transformations in RNA. The simulation of replicating and mutating RNA populations under selection shows that sudden adaptive progress coincides mostly, but not always, with discontinuous shape transformations. The nature of these transformations illuminates the key role of neutral genetic drift in their realization.

A much-debated issue in evolutionary biology concerns the extent to which the history of life has proceeded gradually or has been punctuated by discontinuous transitions at the level of phenotypes (*1*). Our goal is to make the notion of a discontinuous transition more precise and to understand how it arises in a model of evolutionary adaptation.

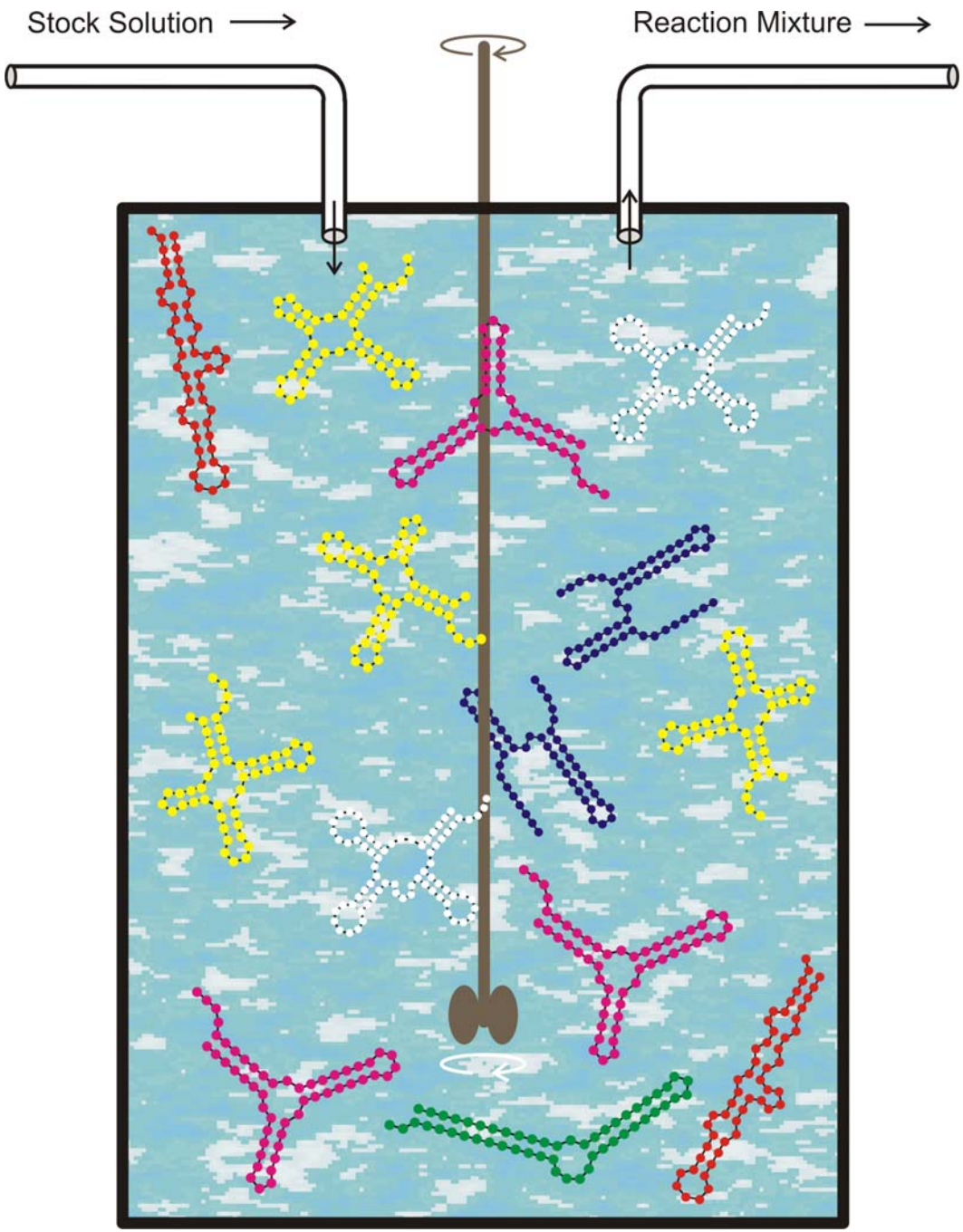
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level of analysis compared with the three-dimensional structure at atomic resolution. Yet, secondary structures are empirically well defined and obtain their biophysical and biochemical importance from being a scaffold for the tertiary structure. For the sake of brevity, we shall refer to secondary structures as "shapes." RNA combines in a single molecule both genotype (replicatable sequence) and phenotype (selectable shape), making it ideally suited for in vitro evolution experiments (*3*, *4*).

To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (*5*, *6*). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (*4*). Although in the experiment the evolutionary end product was unknown, we thought of its shape as being specified implicitly by the imposed selection criterion. Because our intent is to study evolutionary histories rather than end products, we defined a target shape in advance and assumed the replication rate of a sequence to be a function of the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

An instance representing in its qualitative features all the simulations we performed is shown in Fig. 1A. Starting with identical sequences folding into a random shape, the simulation was stopped when the population became dominated by the target, here a canonical tRNA shape. The black curve traces the average distance to the target (inversely related to fitness) in the population against time. Aside from a short initial phase, the entire history is dominated by steps, that is, flat periods of no apparent adaptive progress, interrupted by sudden approaches toward the target structure (*7*). However, the dominant shapes in the population not only change at these marked events but undergo several fitness-neutral transformations during the periods of no apparent progress. Although discontinuities in the fitness trace are evident, it is entirely unclear when and on the basis of what the series of successive phenotypes itself can be called continuous or discontinuous.

A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally organized into a space because point mutations induce a canonical neighborhood. The neighborhood of a sequence consists of all its one-error mutants. The problem is how to organize the set of possible shapes into a space. The issue arises because, in contrast to sequences, there are

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



Replication rate constant

(Fitness):

$$f_k = \gamma / [\alpha + \Delta d_S^{(k)}]$$

$$\Delta d_S^{(k)} = d_H(S_k, S_\tau)$$

Selection pressure:

The population size,

N = # RNA moleucles,

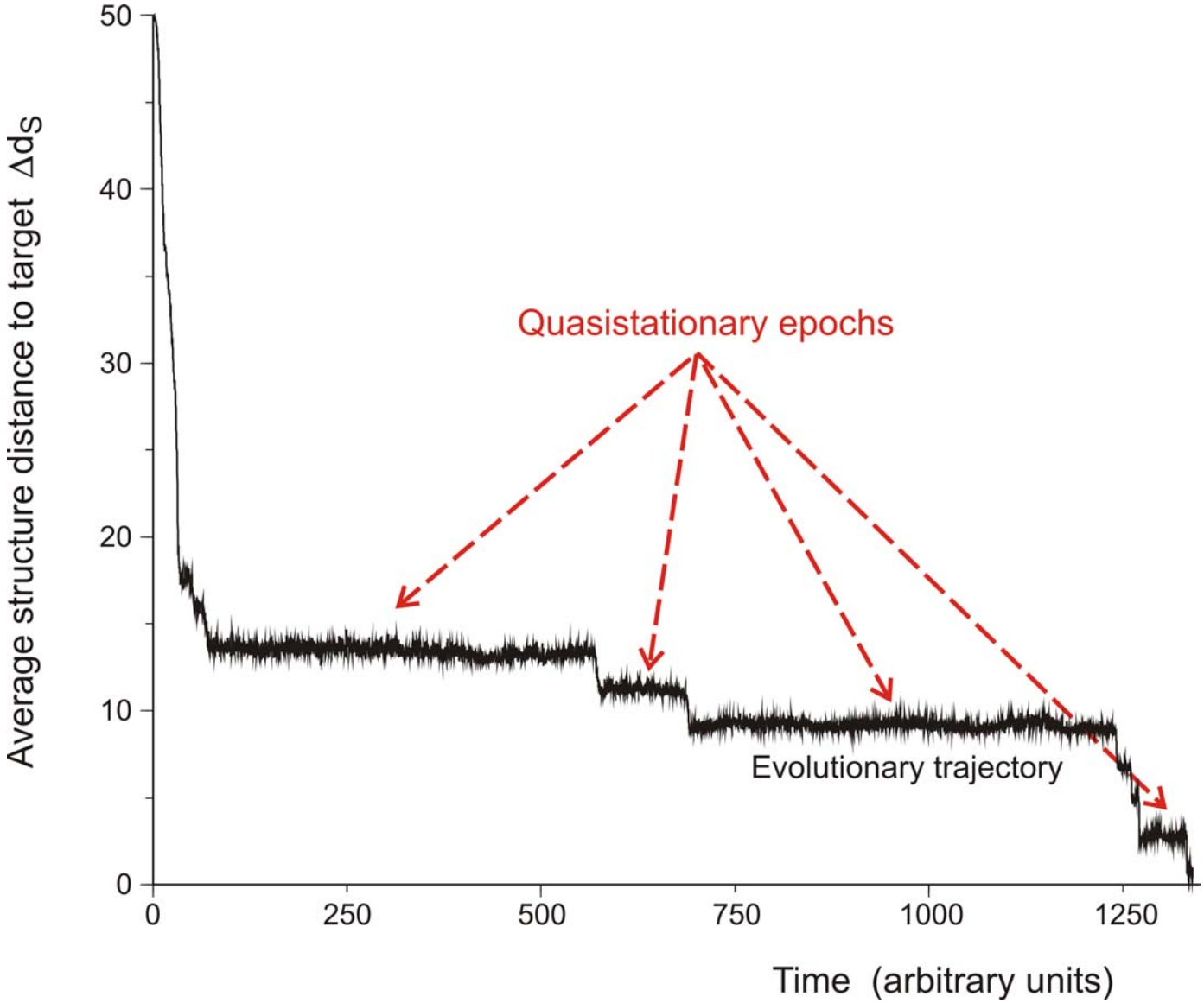
is determined by the flux:

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

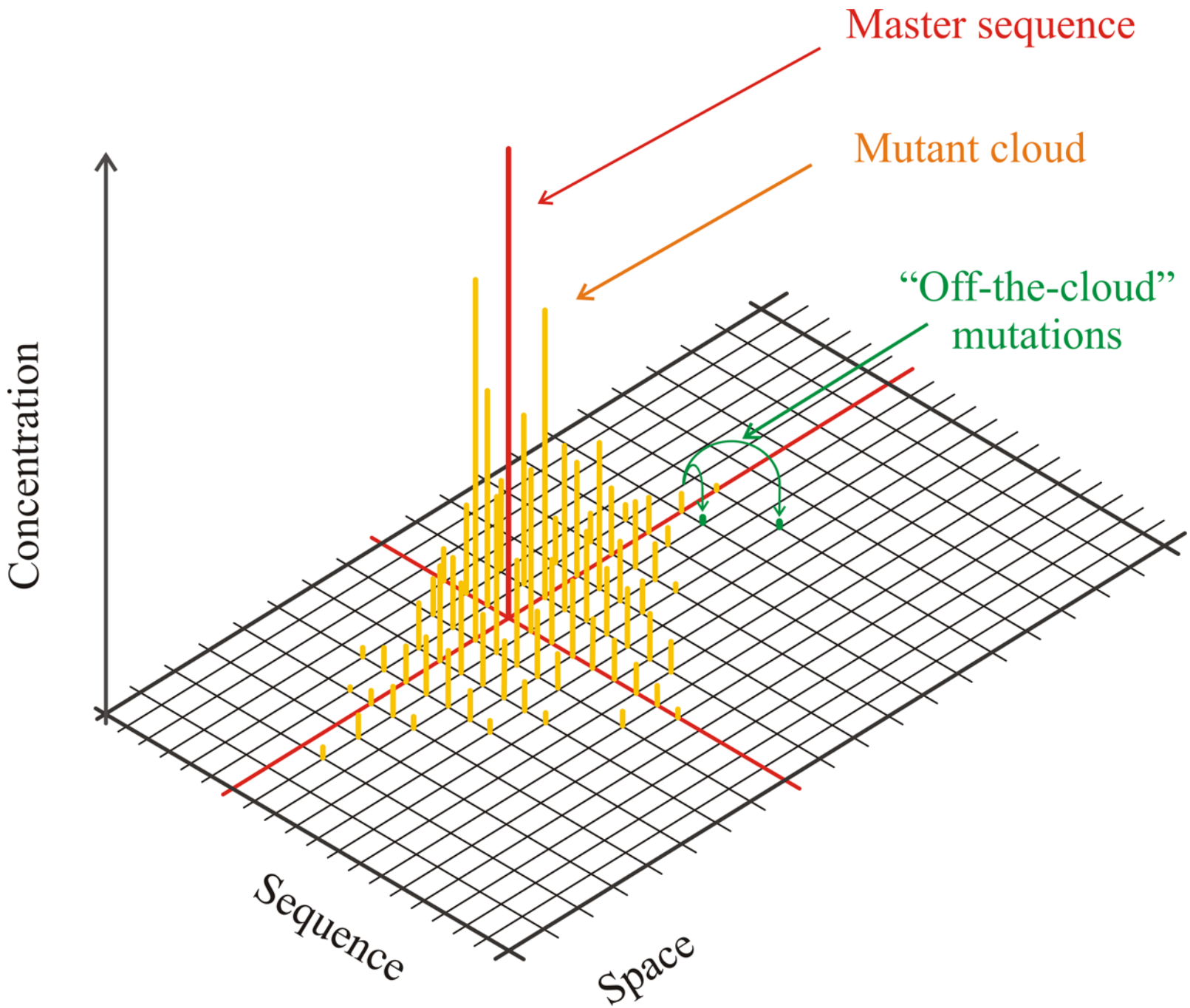
Mutation rate:

p = 0.001 / Nucleotide × Replication

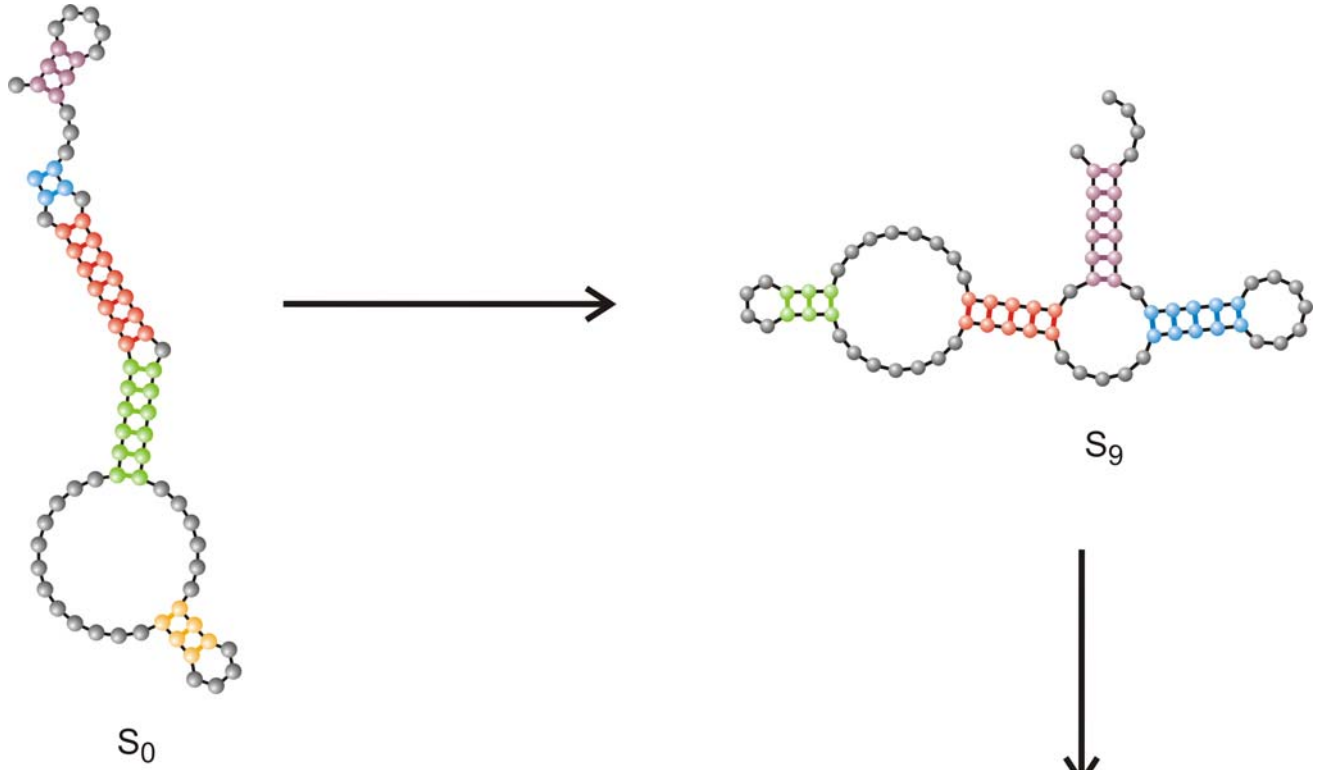
The flow reactor as a device for studying the evolution of molecules ***in vitro*** and ***in silico***.



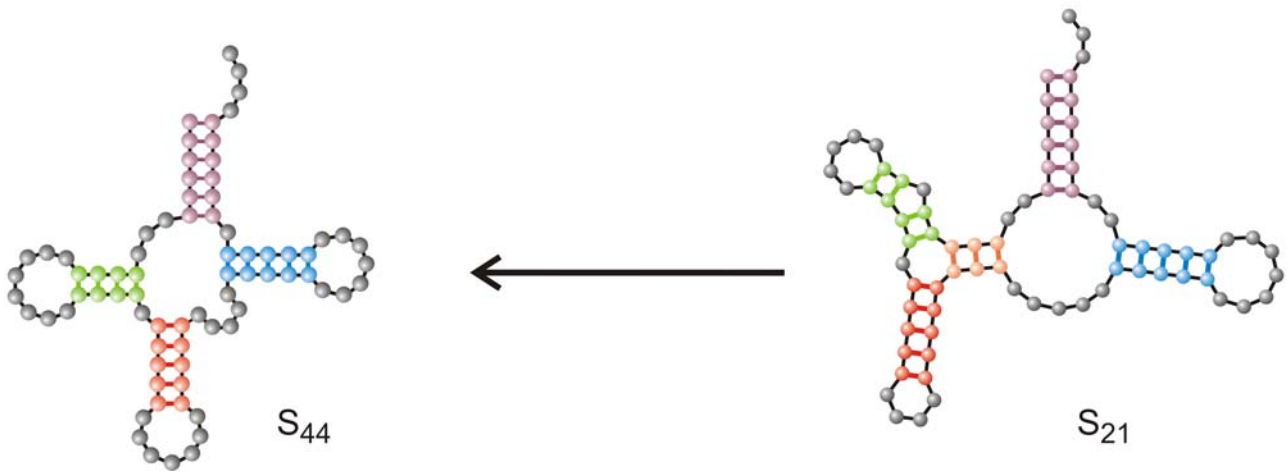
In silico optimization in the flow reactor: Evolutionary Trajectory


Master sequence
Mutant cloud
"Off-the-cloud" mutations
Concentration
Sequence
Space

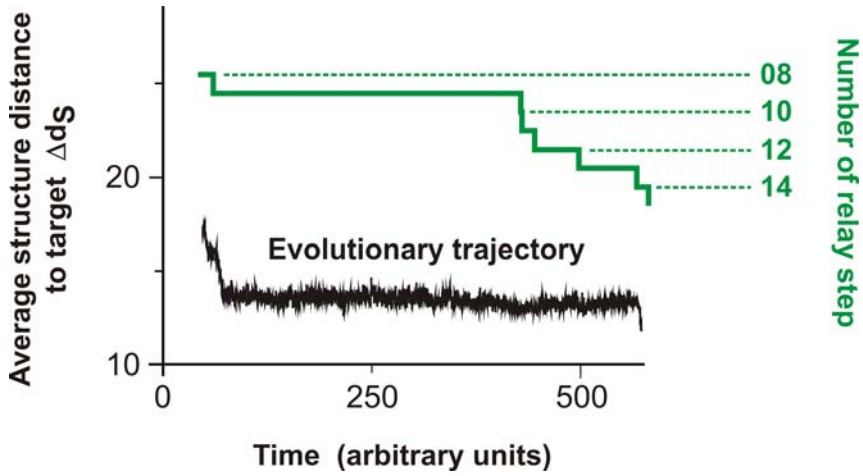
Randomly chosen
initial structure



Phenylalanyl-tRNA
as target structure



28 neutral point mutations during a long quasi-stationary epoch




```
entry   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8       .((((((((((((.........(((....)))).......)))))).....((((.......))))))))))....
exit    GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCAUACAGAA
entry   GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUACCAUACAGAA
9       .((((((.(((((.........(((....)))).......)))))).....((((.......))))).))))))....
exit    UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACACCGUCCCAAG
entry   UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACAGCGUCCCAAG
10      .(((((..(((((.........(((....)))).......)))))).....((((.......))))).))))))....
exit    UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUACGCCCCACACAGCGUCCCAAG
```

Transition inducing point mutations change the molecular structure

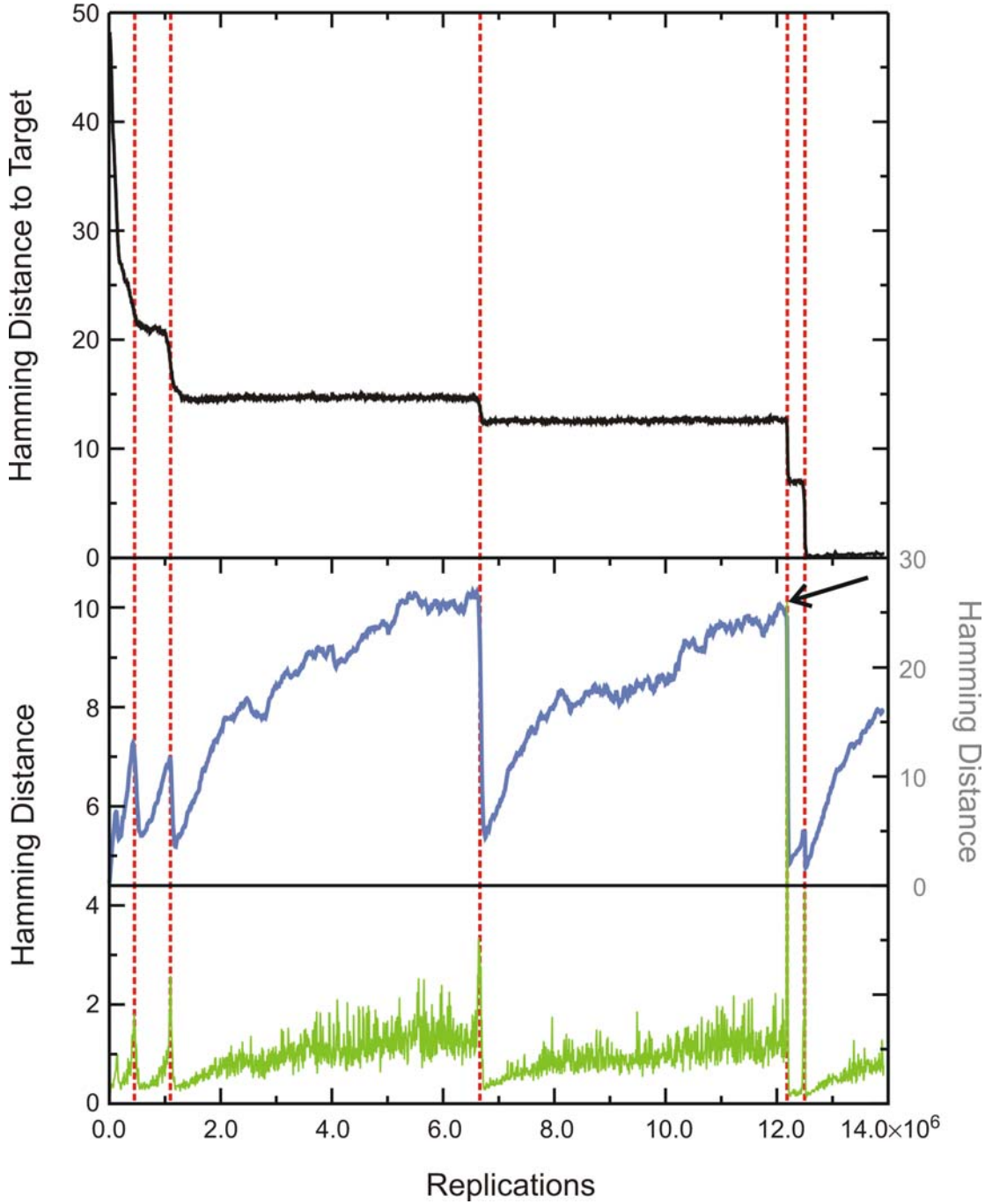
Neutral point mutations leave the molecular structure unchanged

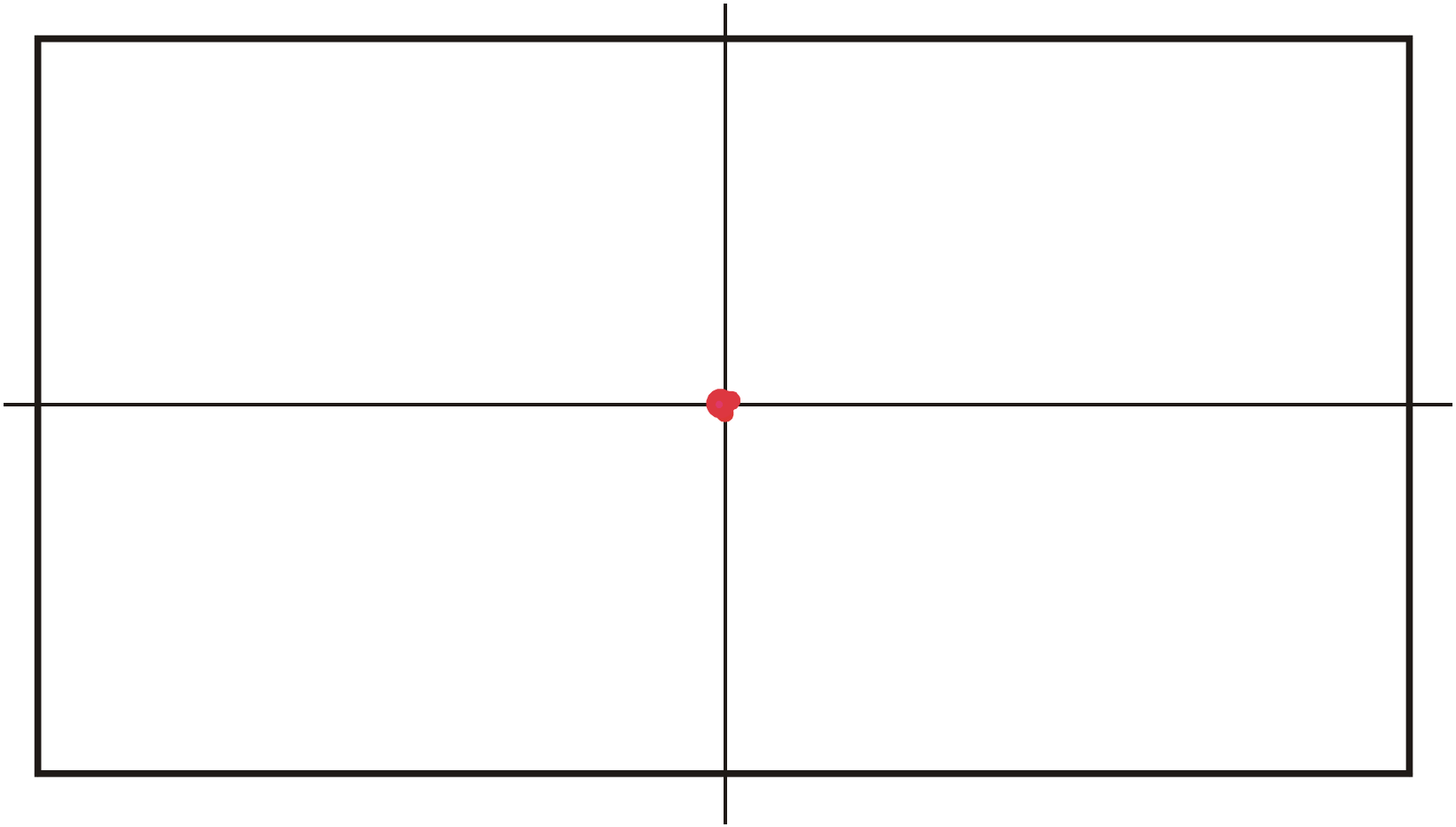
Neutral genotype evolution during phenotypic stasis

Evolutionary trajectory

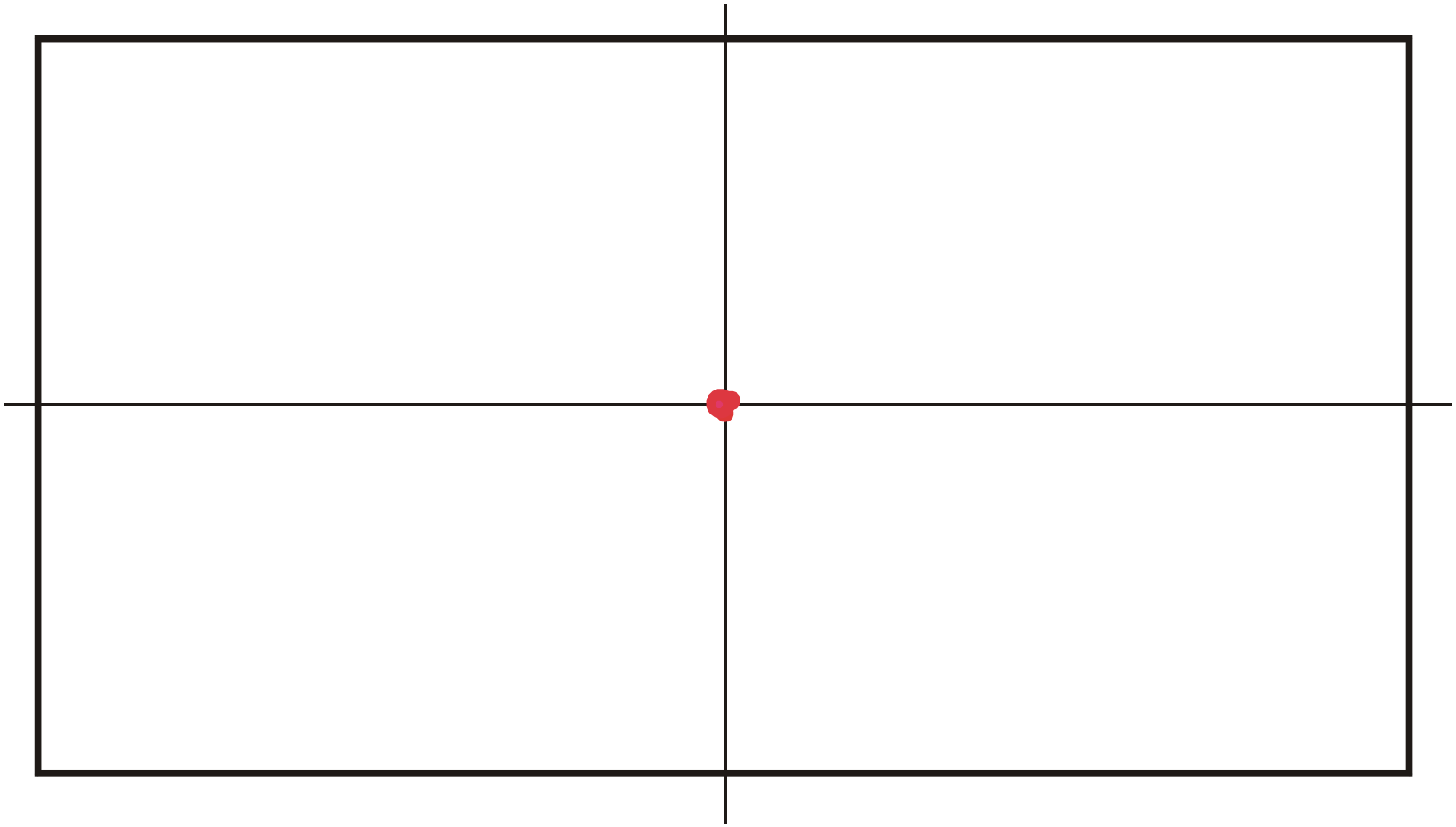
Spreading of the population on neutral networks

Drift of the population center in sequence space

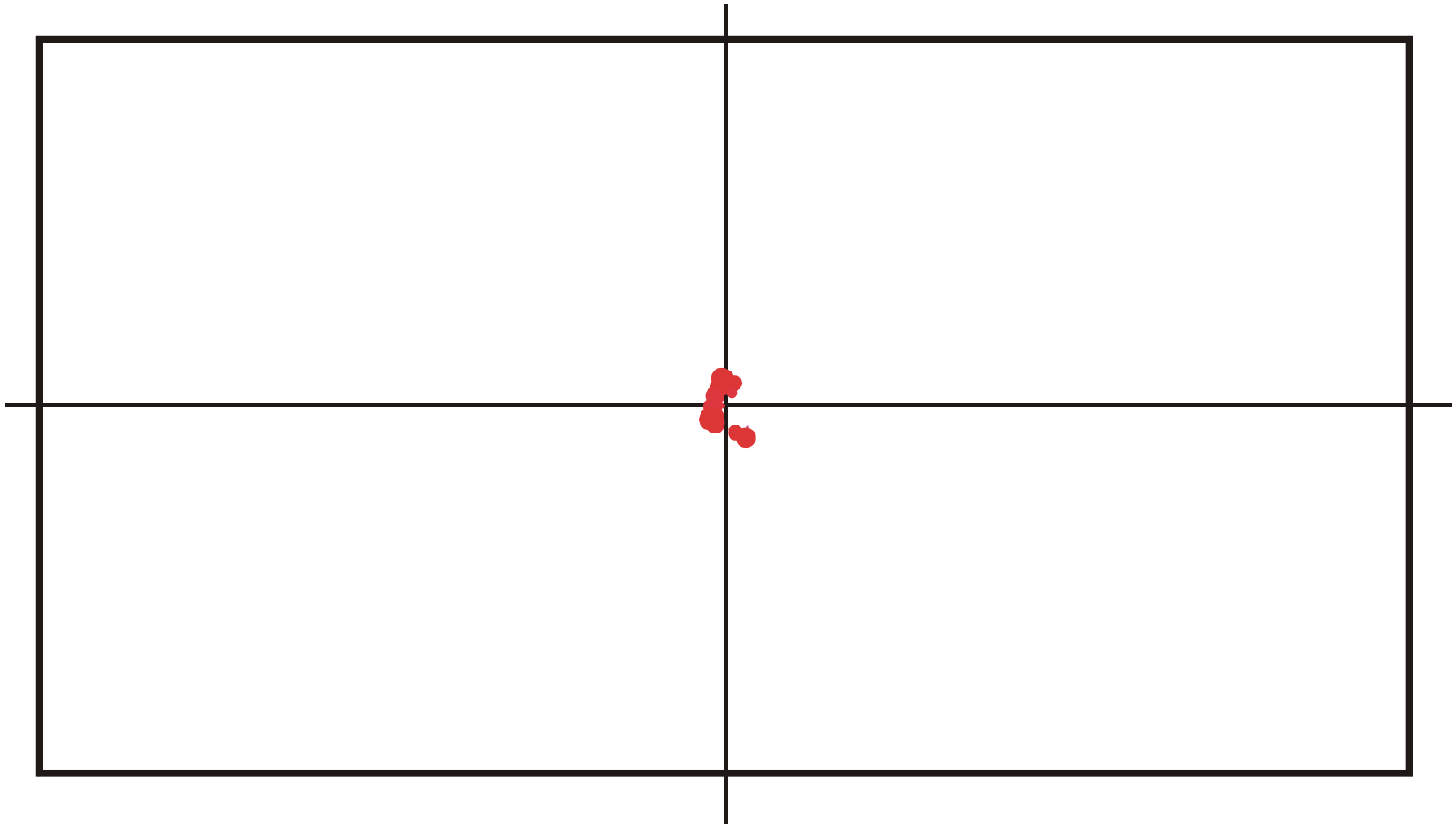




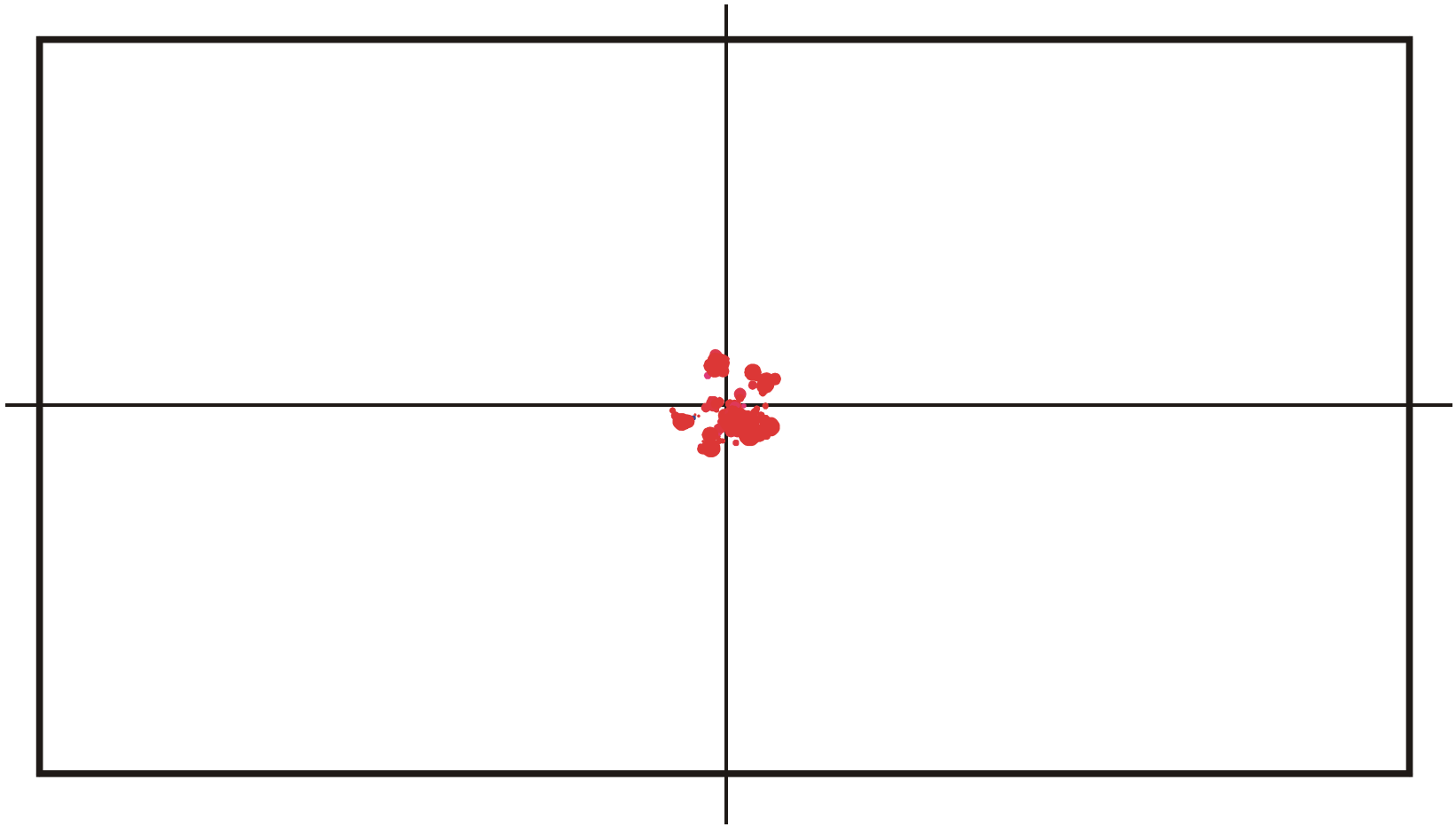
Spreading and evolution of a population on a neutral network: t = 150



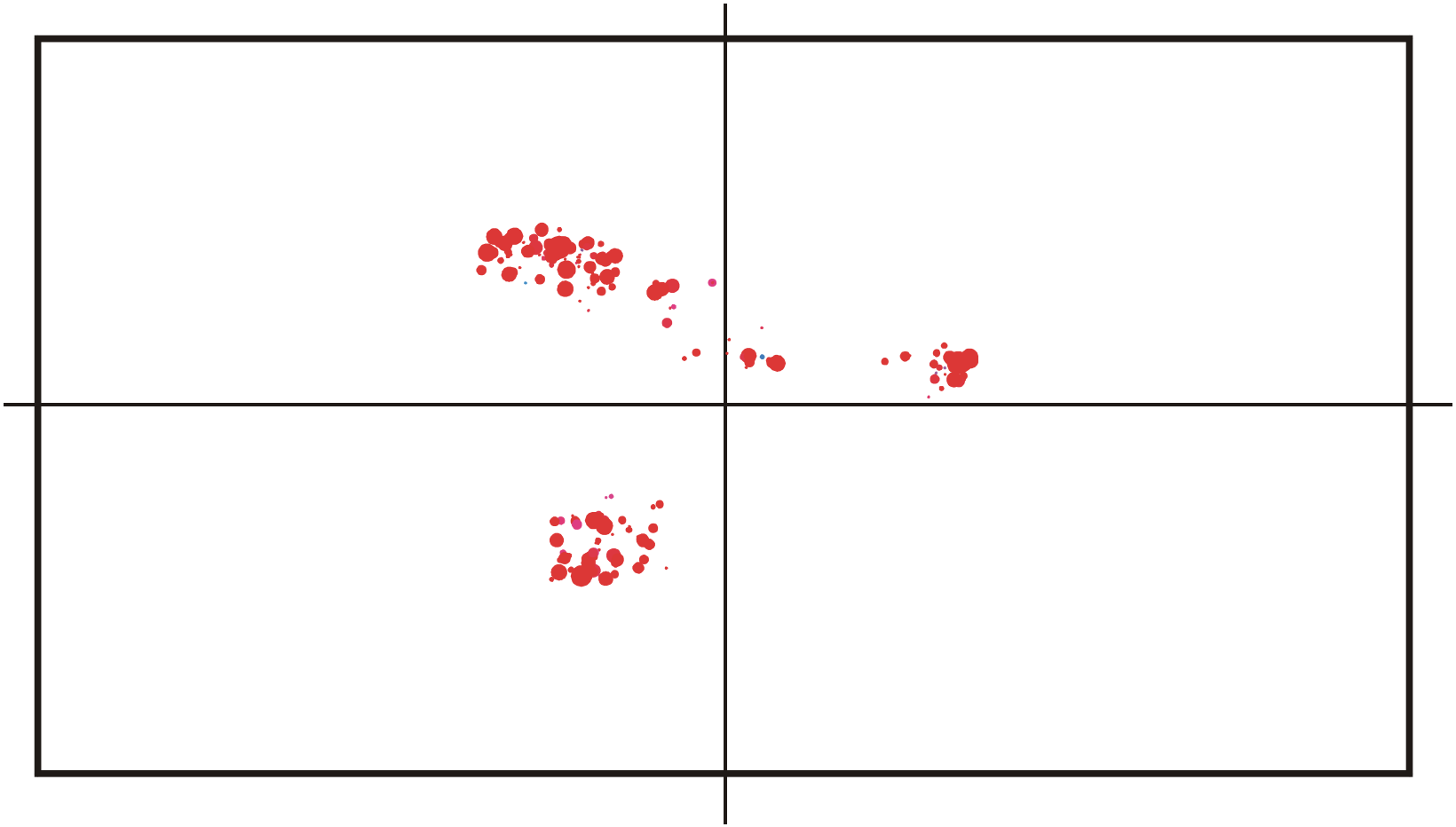
Spreading and evolution of a population on a neutral network: t = 150



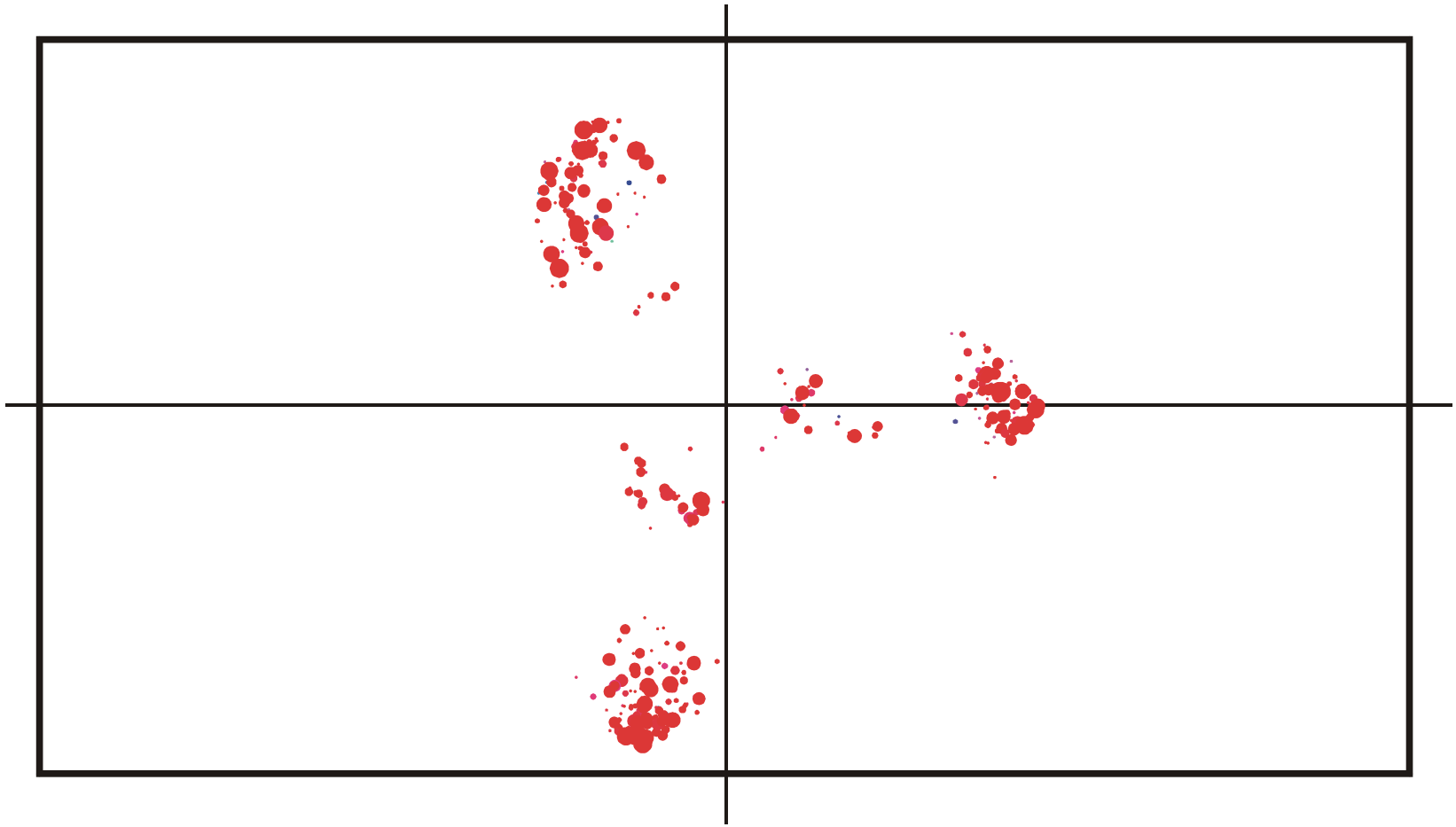
Spreading and evolution of a population on a neutral network : t = 170



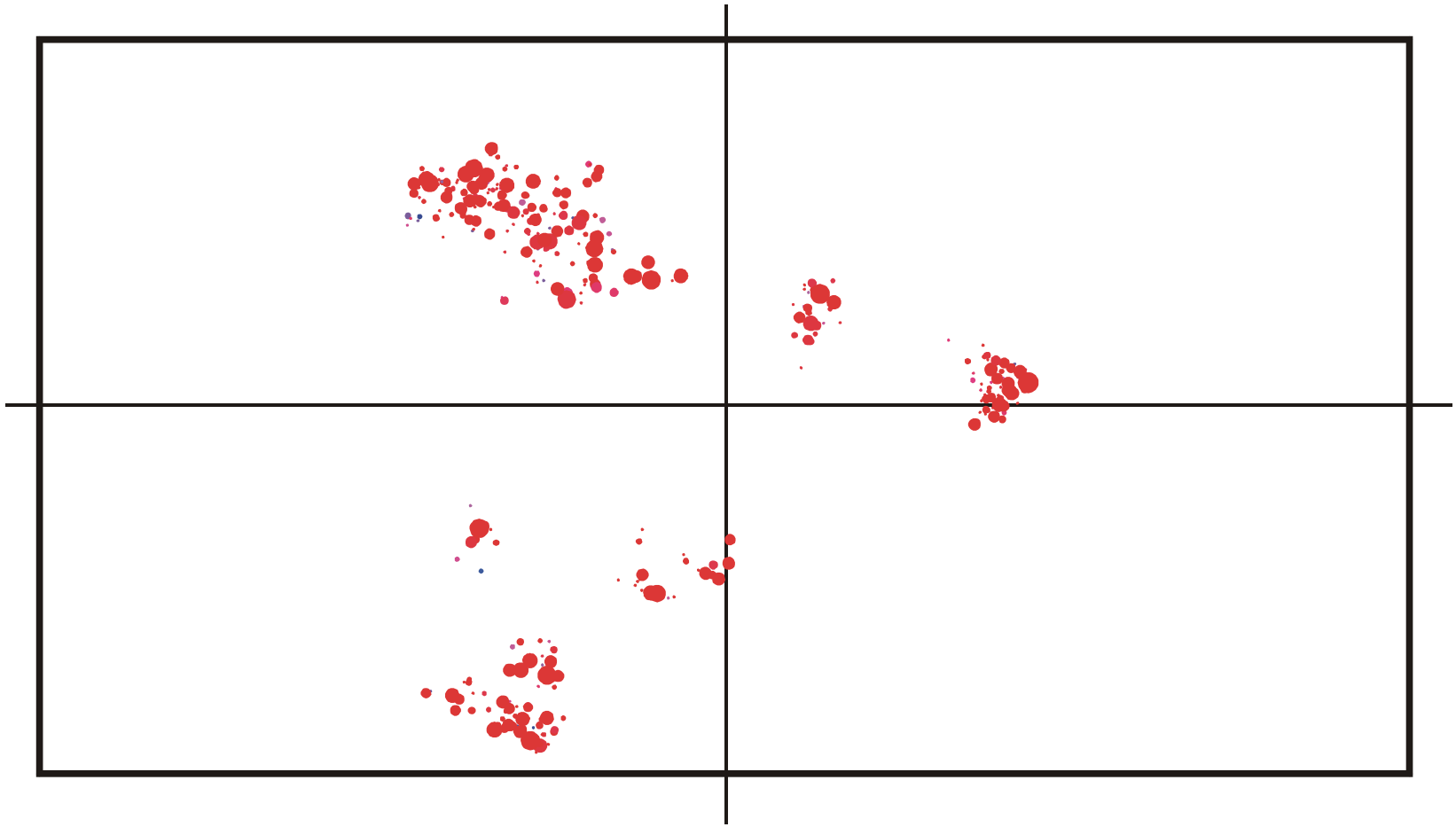
Spreading and evolution of a population on a neutral network : t = 200



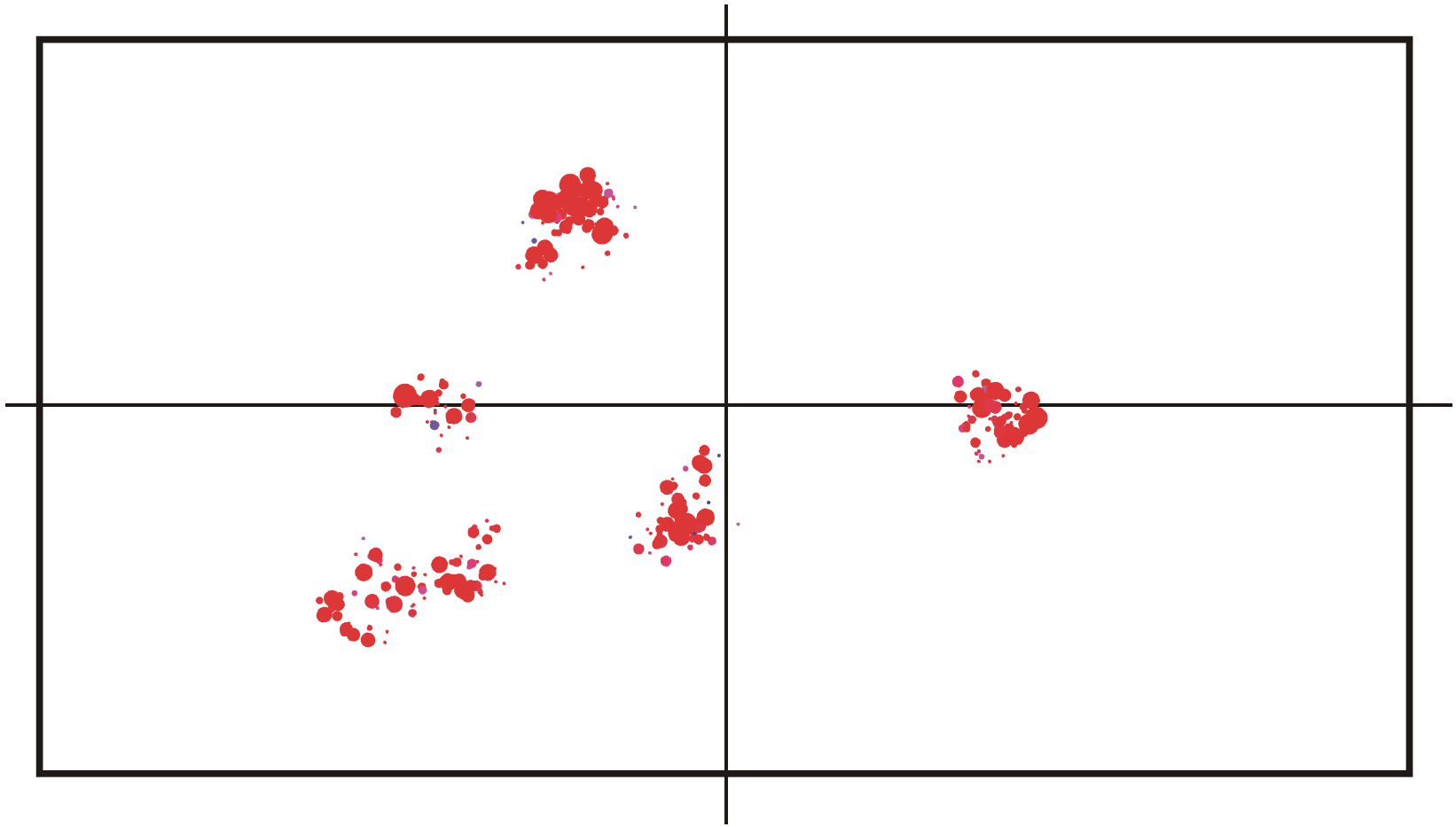
Spreading and evolution of a population on a neutral network : t = 350



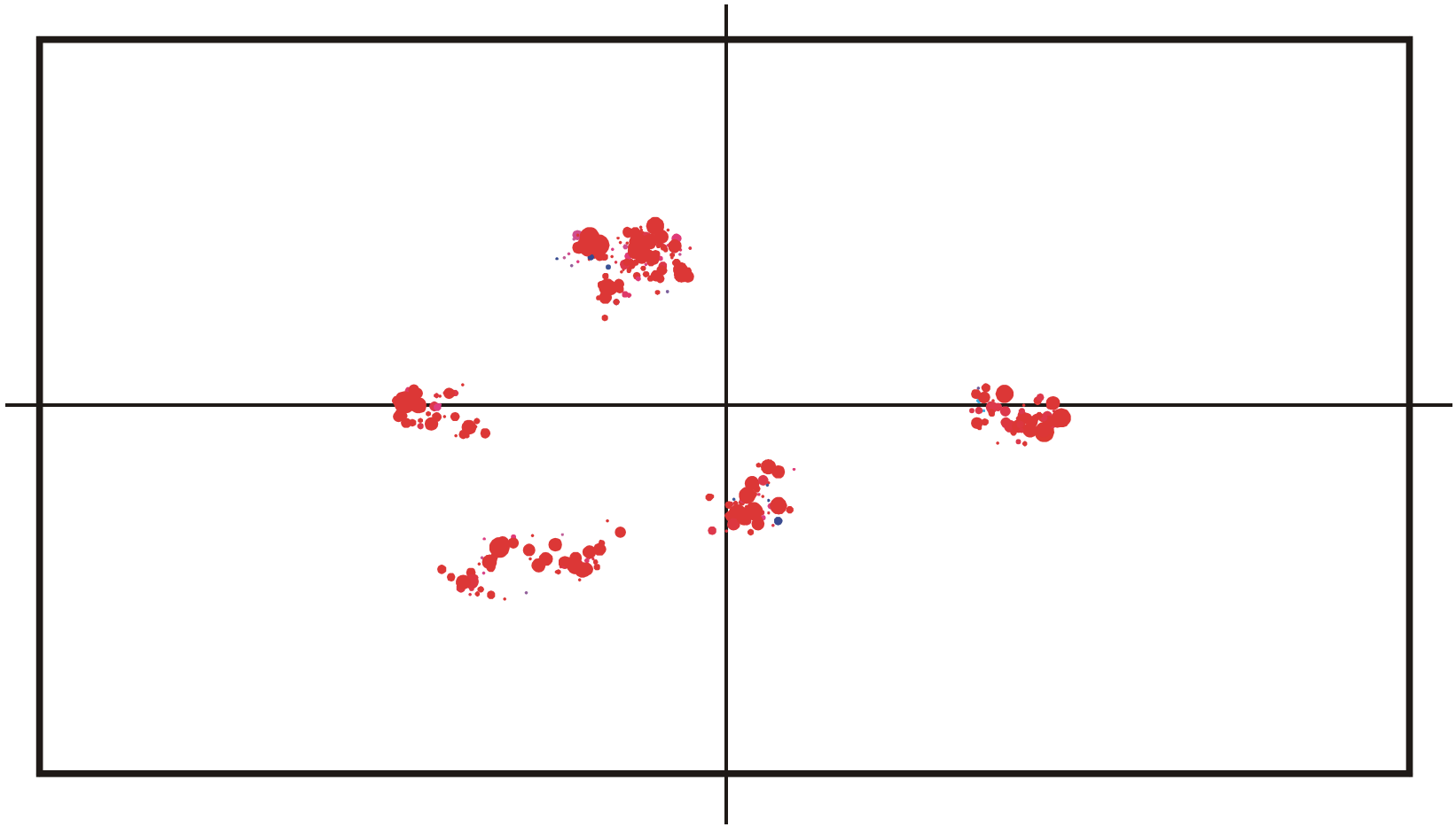
Spreading and evolution of a population on a neutral network : t = 500



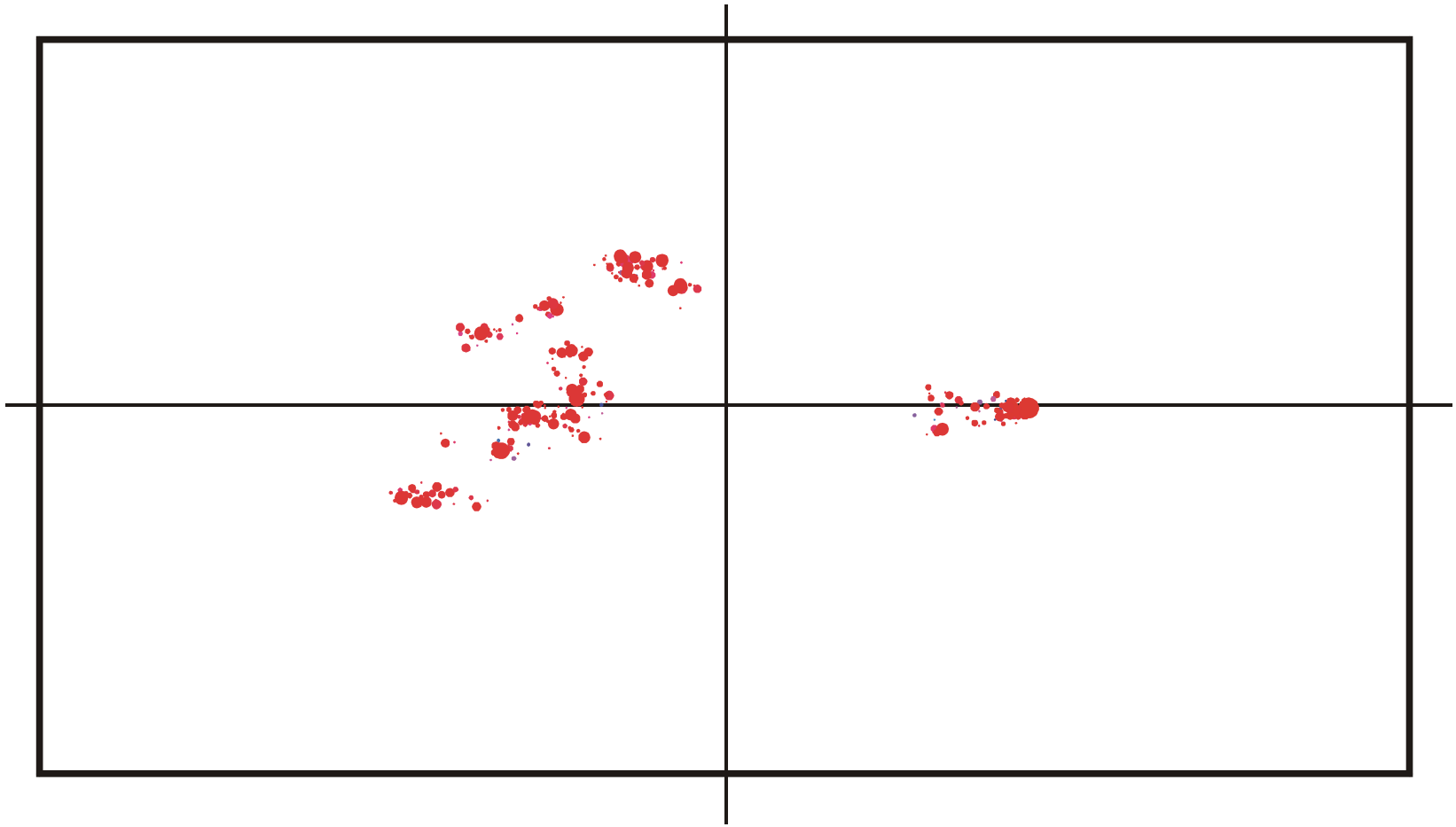
Spreading and evolution of a population on a neutral network : t = 650



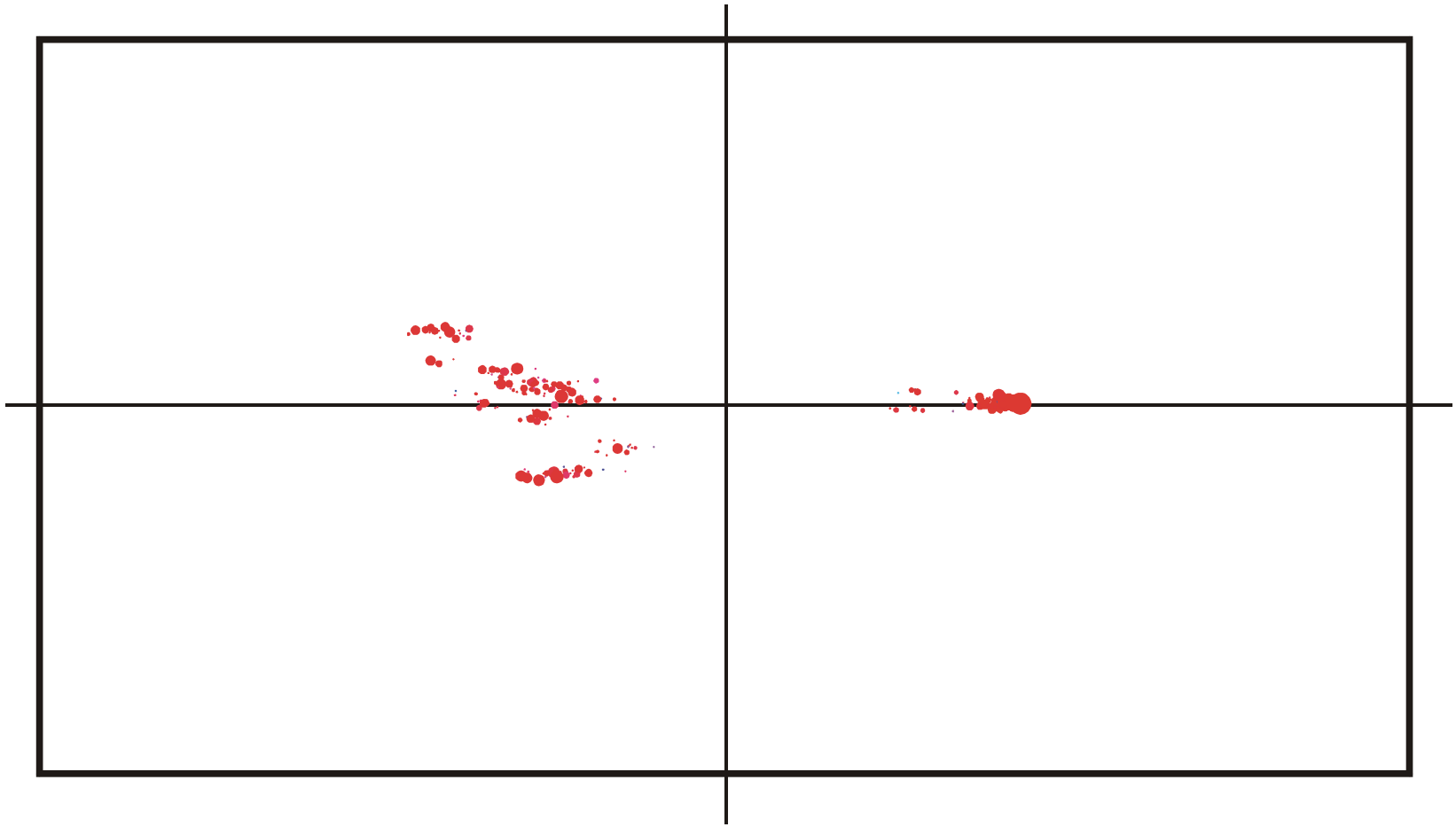
Spreading and evolution of a population on a neutral network : t = 820



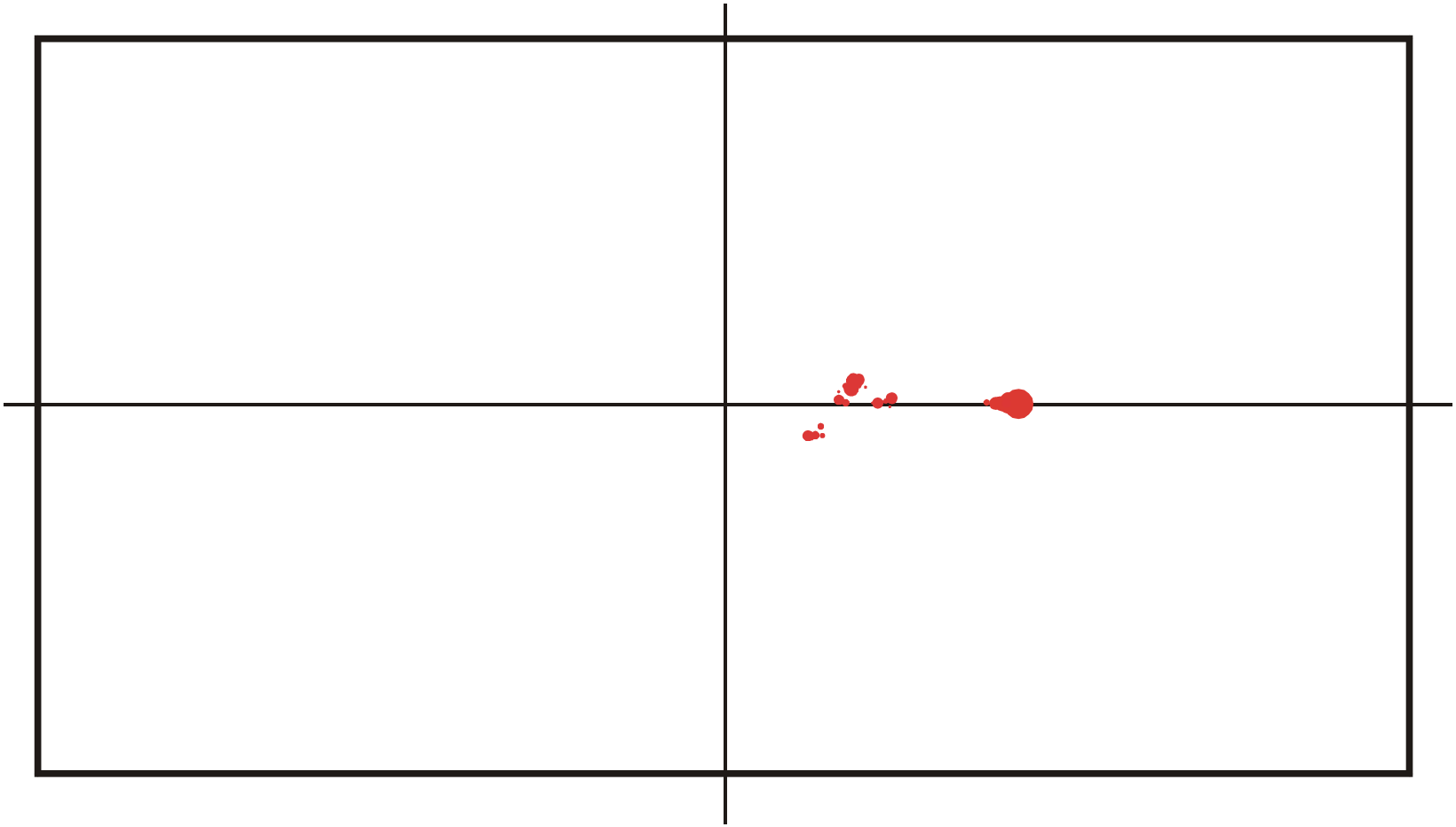
Spreading and evolution of a population on a neutral network : t = 825



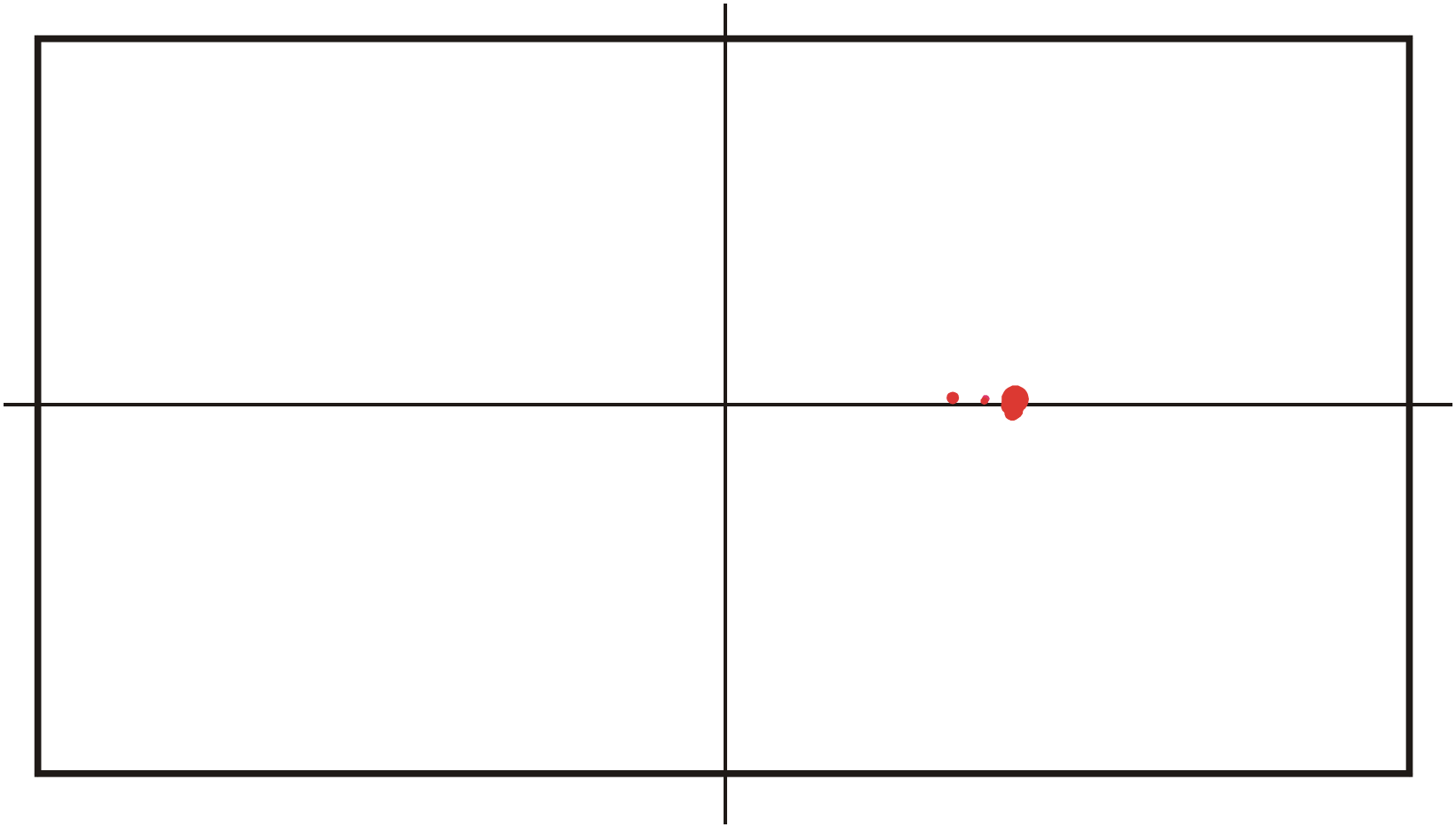
Spreading and evolution of a population on a neutral network : t = 830



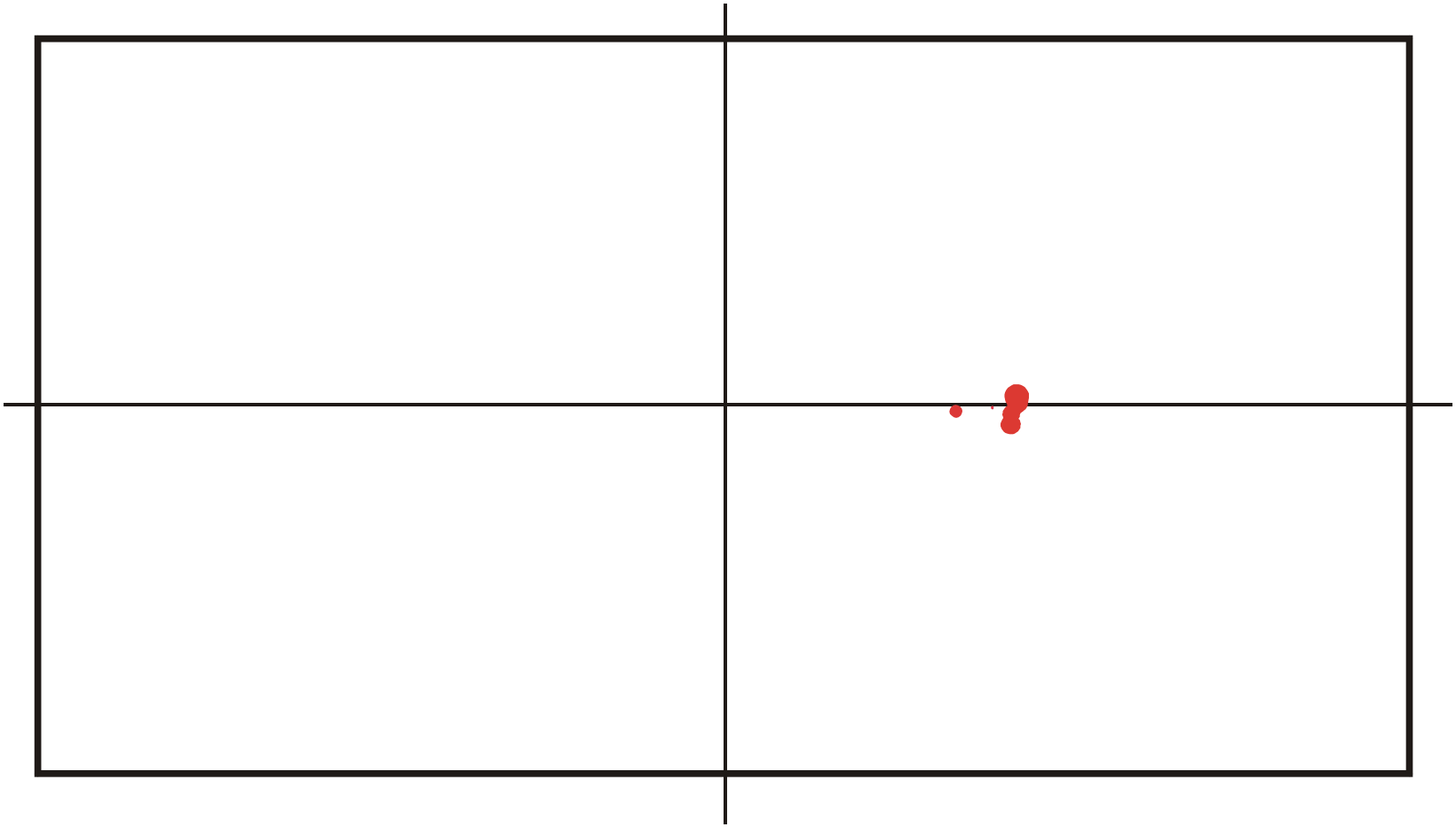
Spreading and evolution of a population on a neutral network : t = 835



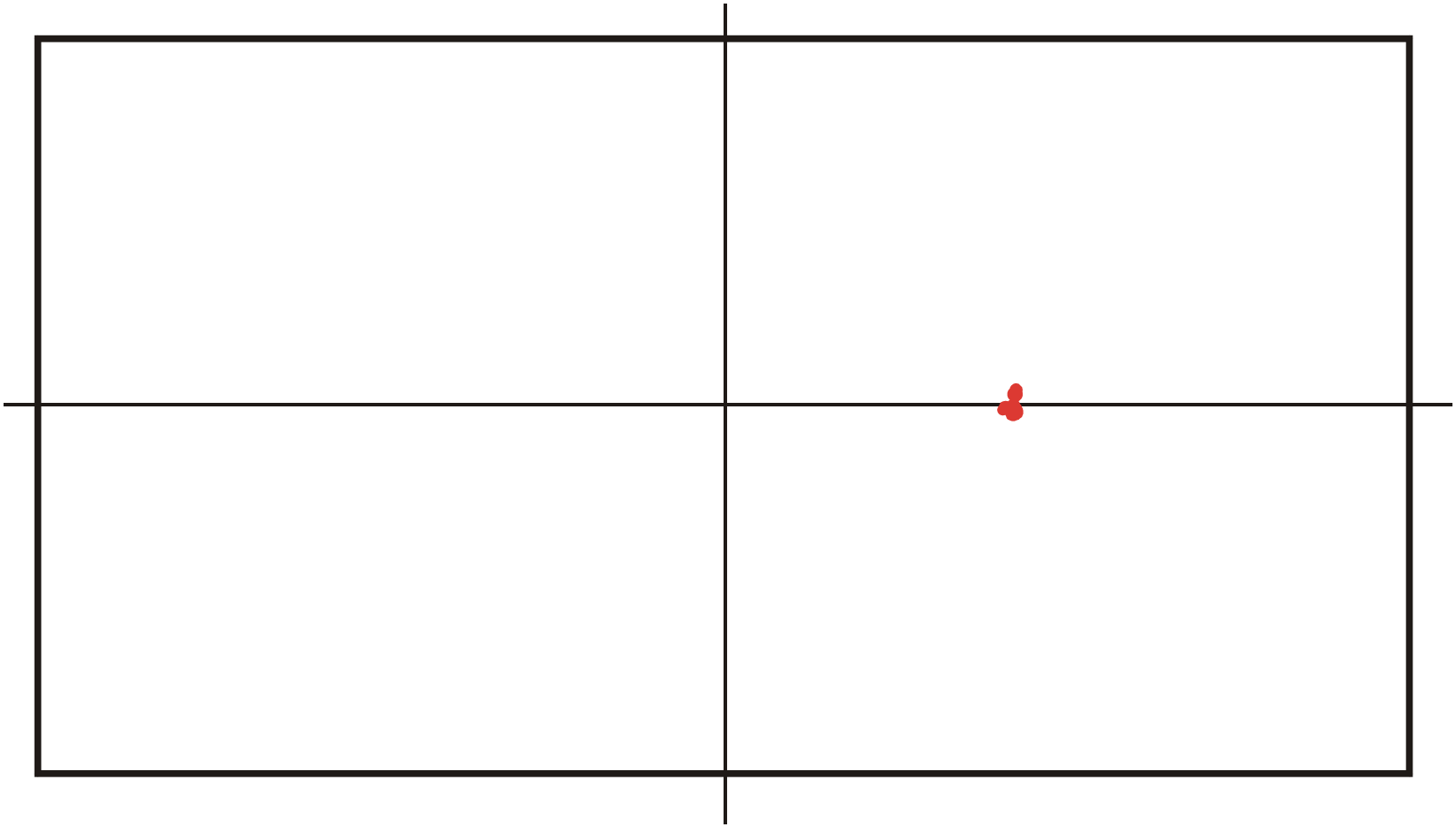
Spreading and evolution of a population on a neutral network : t = 840



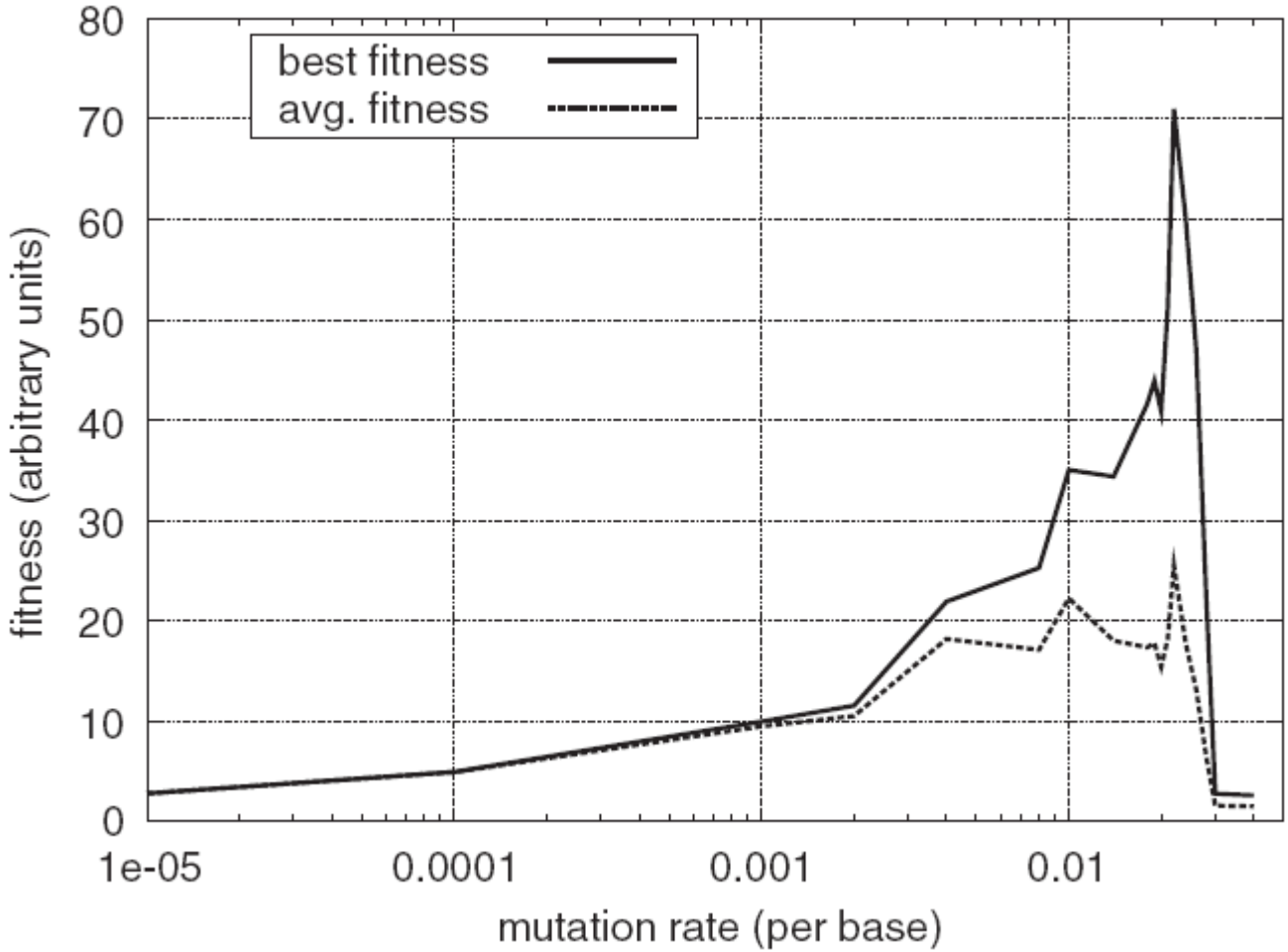
Spreading and evolution of a population on a neutral network : t = 845



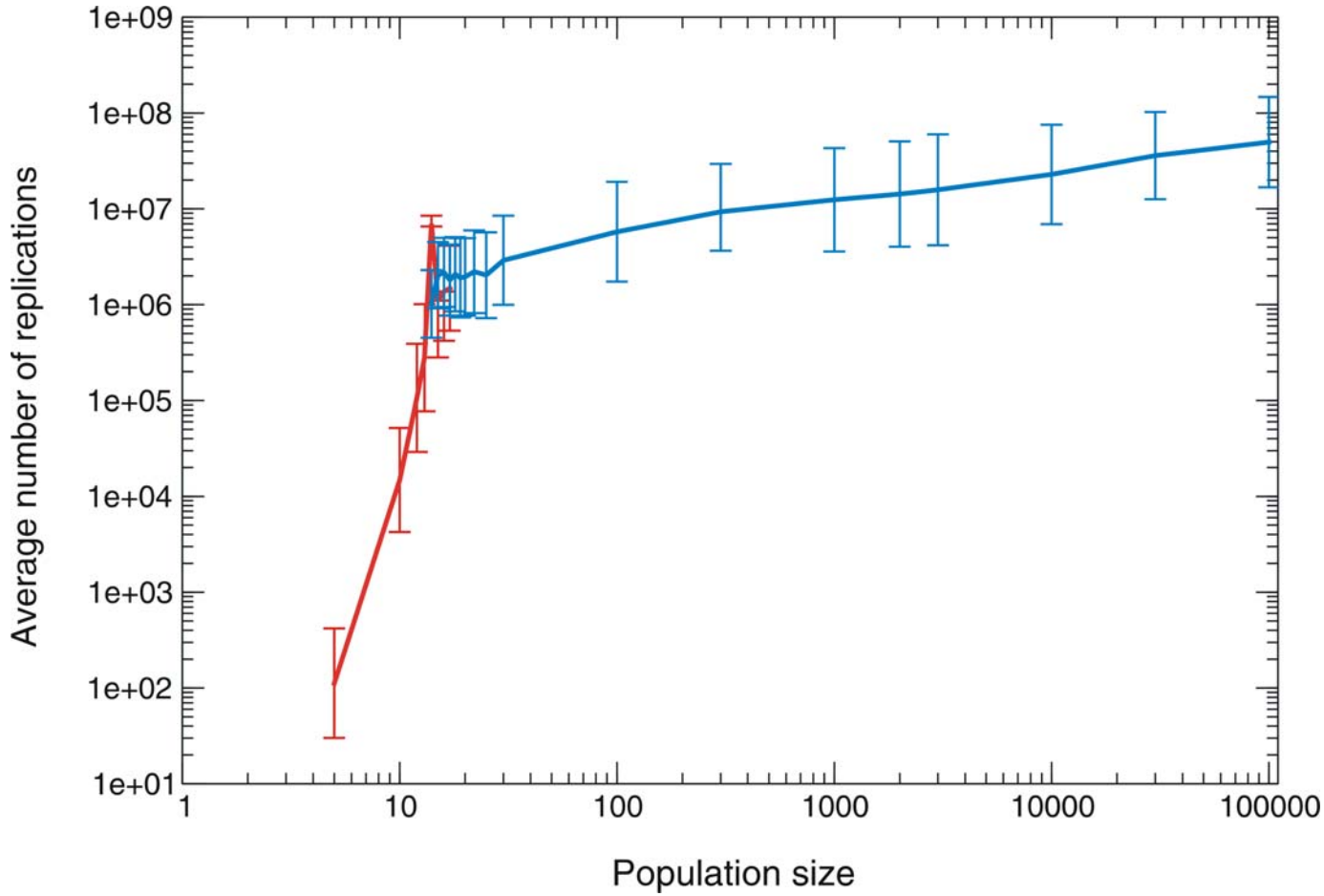
Spreading and evolution of a population on a neutral network : t = 850

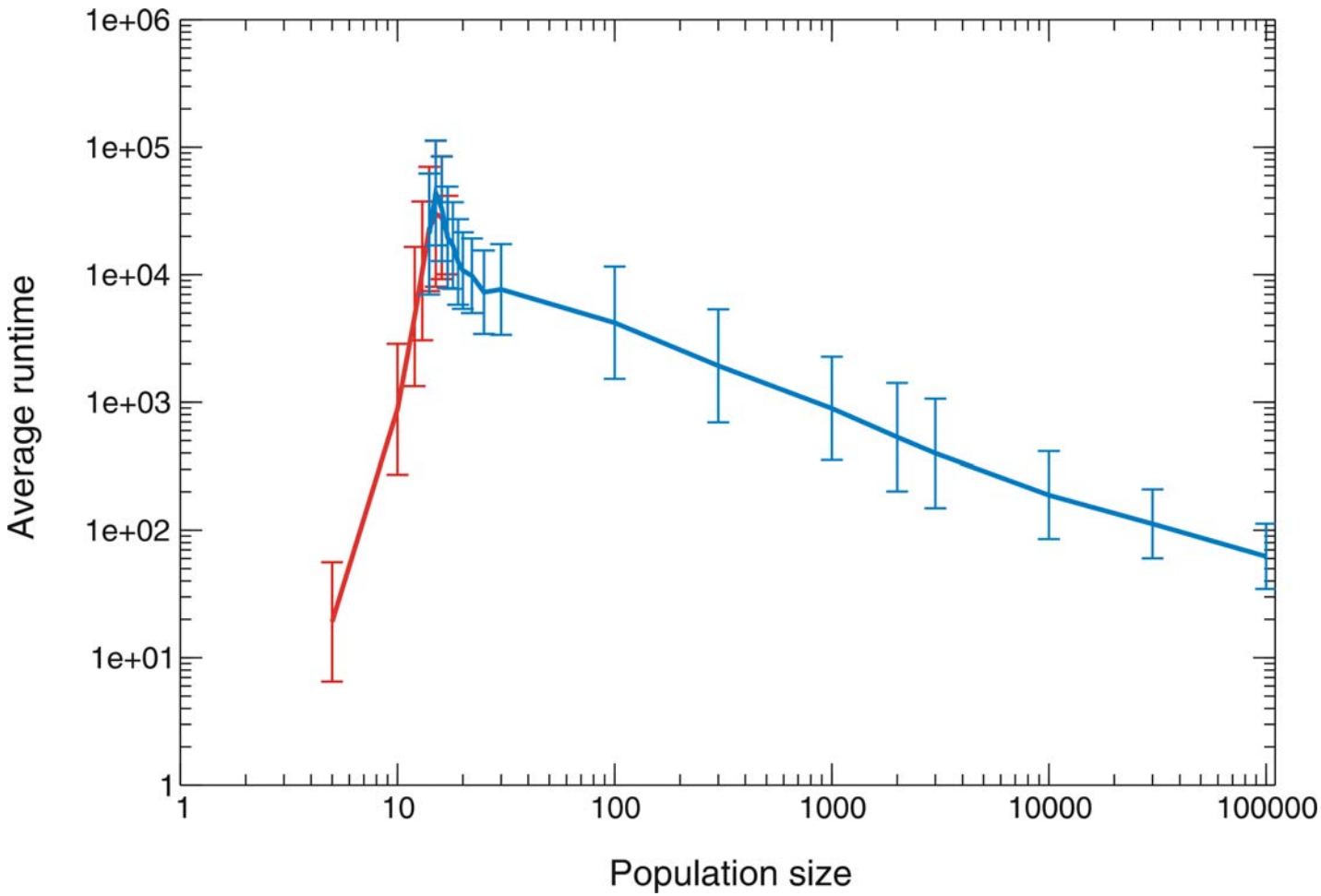


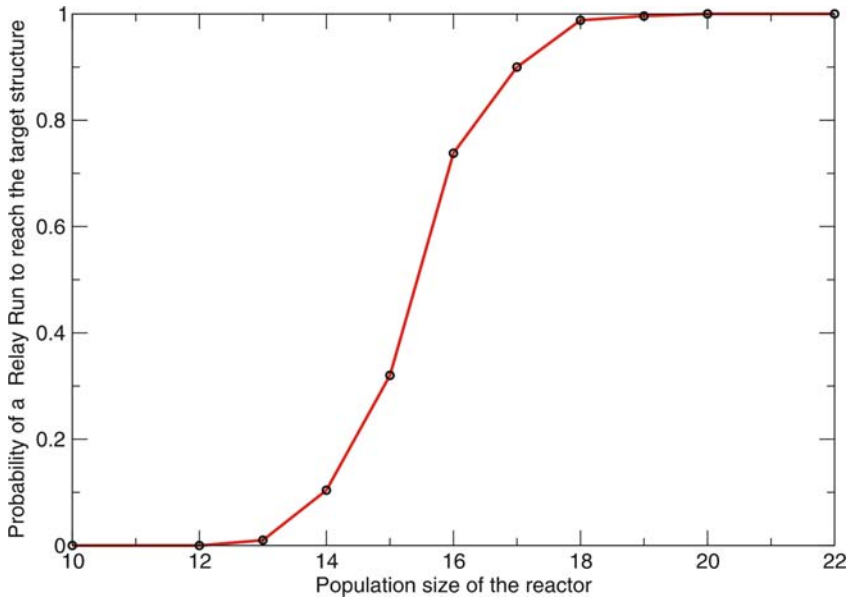
Spreading and evolution of a population on a neutral network : t = 855

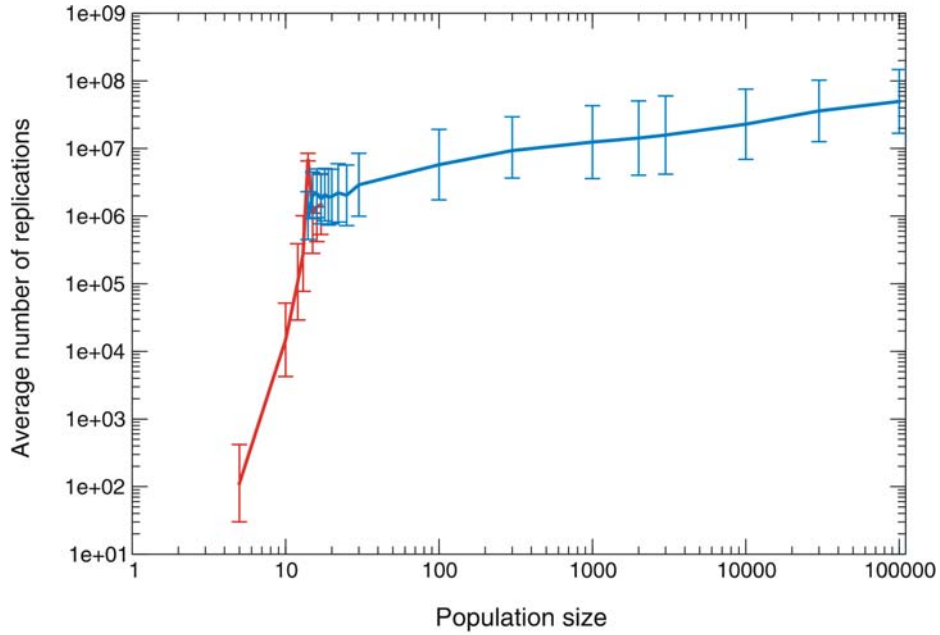


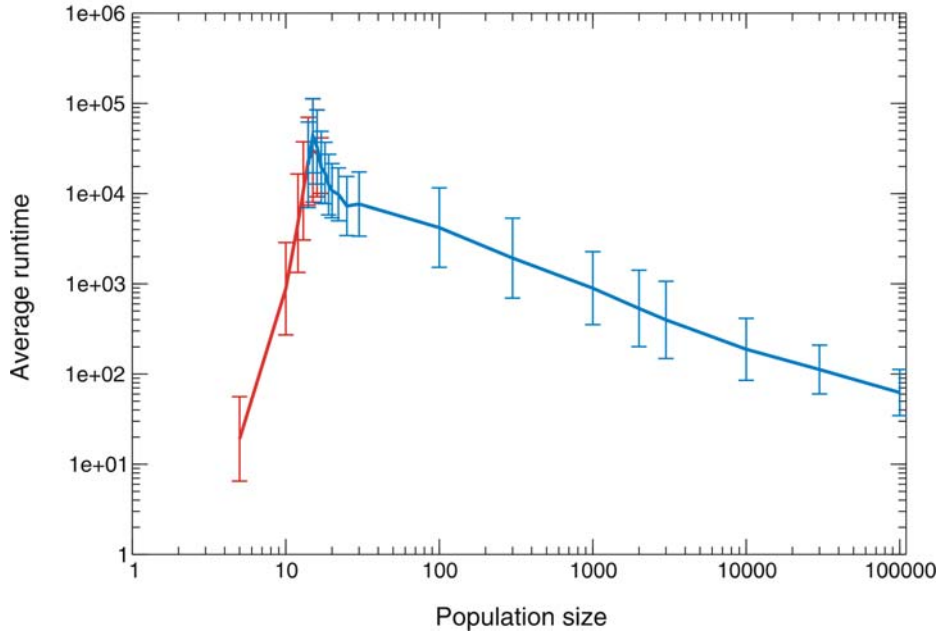
Anne Kupczok, Peter Dittrich, Determinats of simulated RNA evolution. J.Theor.Biol. **238**:726-735, 2006

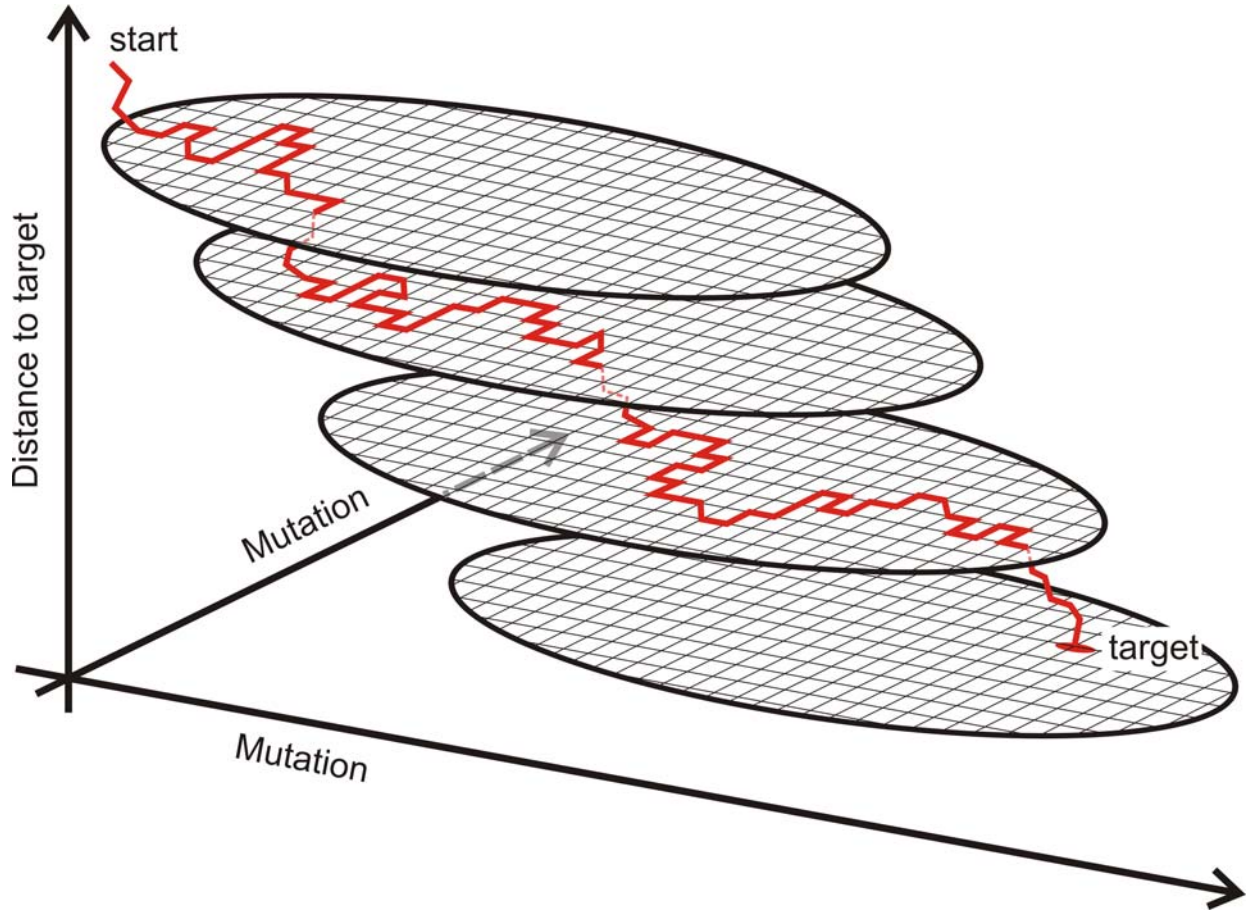

Average number of replications
1e+09
1e+08
1e+07
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1
10
100
1000
10000
100000
Population size


1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1
Average runtime
1
10
100
1000
10000
100000
Population size

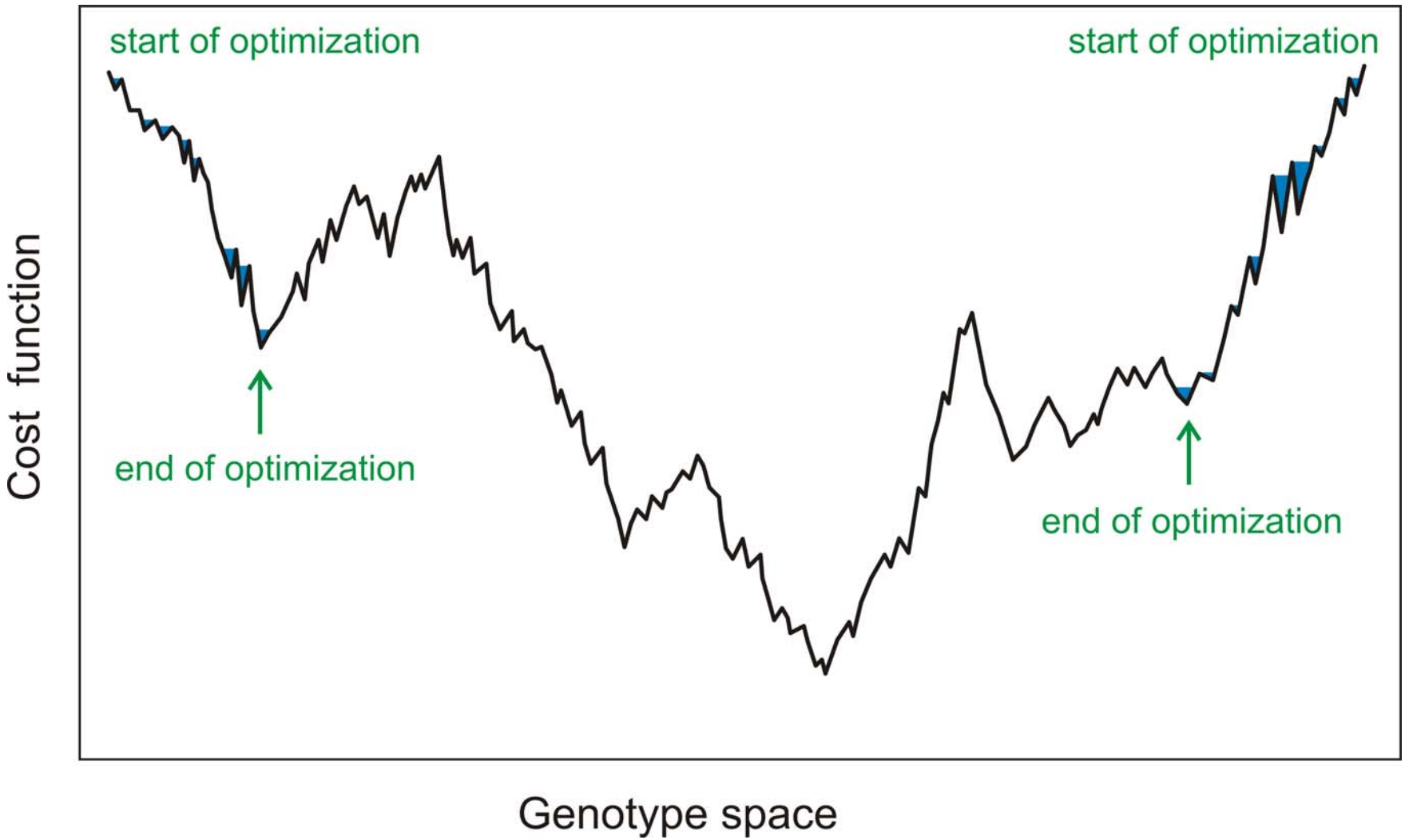

Probability of a Relay Run to reach the target structure
0
0.2
0.4
0.6
0.8
1
10
12
14
16
18
20
22
Population size of the reactor

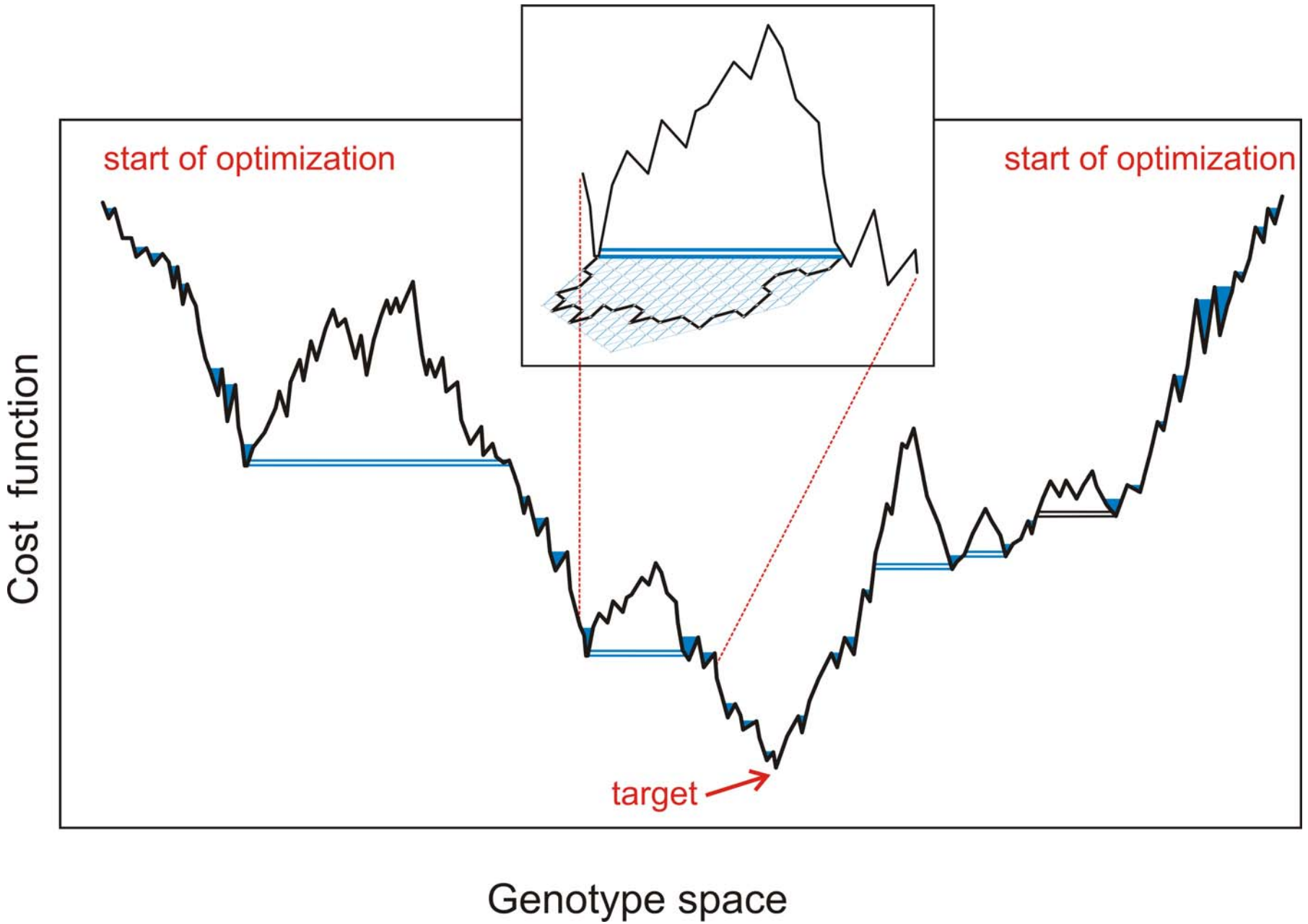

Average number of replications
1e+09
1e+08
1e+07
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1
10
100
1000
10000
100000
Population size


Average runtime
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1
10
100
1000
10000
100000
Population size



A sketch of optimization on neutral networks


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


start of optimization
start of optimization
target
Cost function
Genotype space

Landscapes and neutrality

RNA sequence GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA folding:

Structural biology, spectroscopy of biomolecules, understanding **molecular function**

Biophysical chemistry: Thermodynamics and kinetics

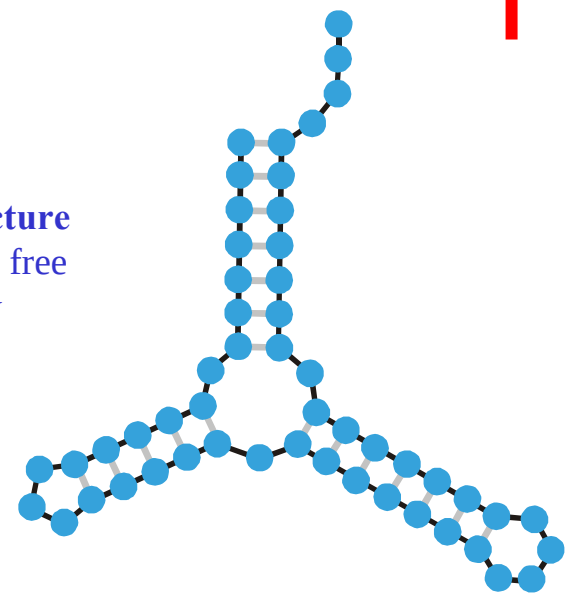
Empirical parameters

Inverse folding of RNA:

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

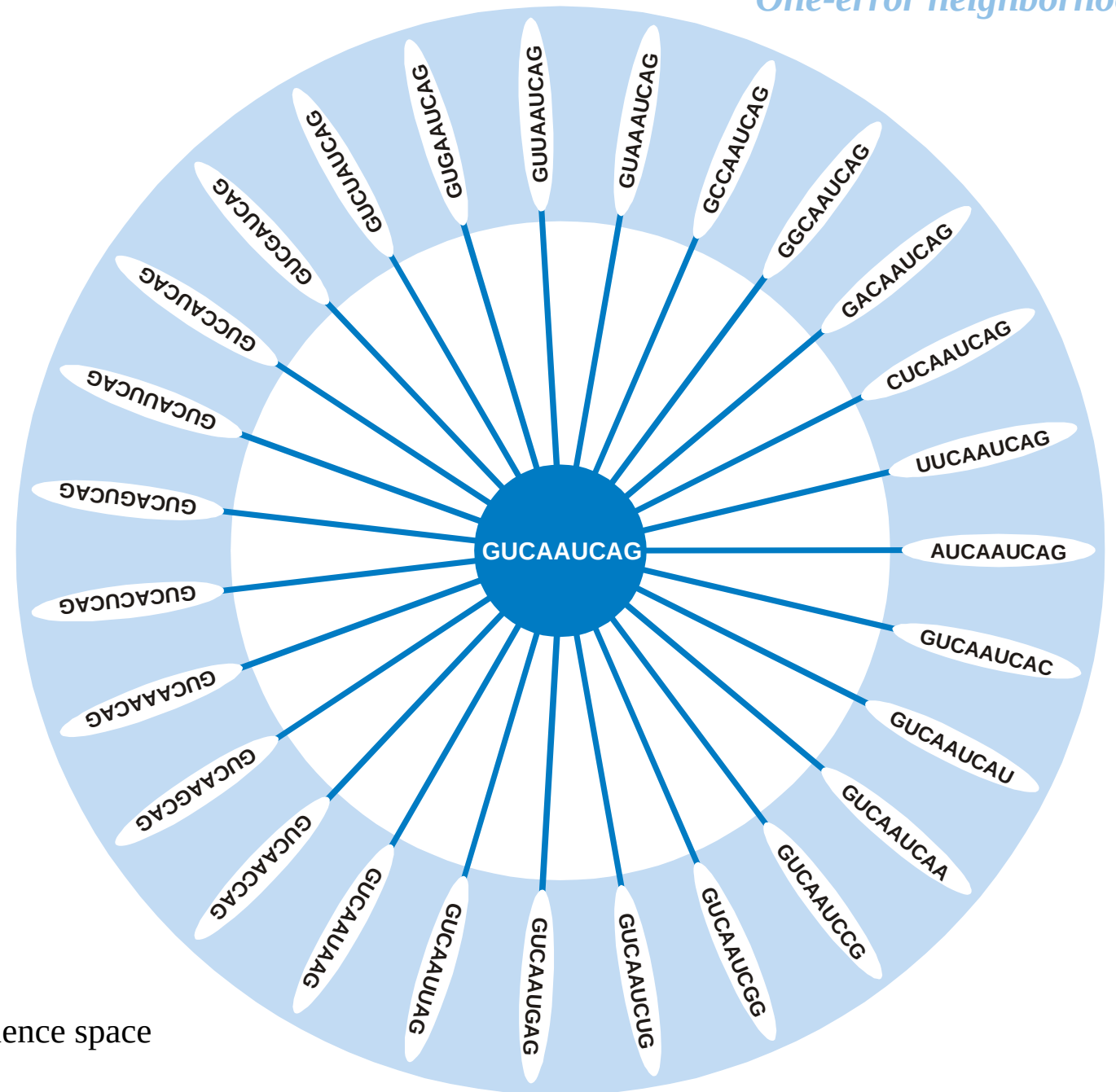
One-sequence-one structure paradigm

RNA structure of minimal free energy



Sequence, structure, and design

One-error neighborhood

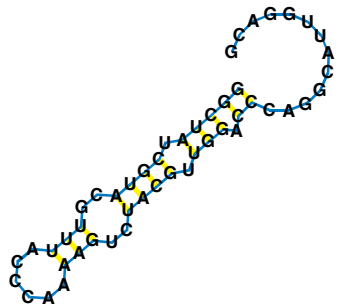


The surrounding of
GUCAAUCAG in sequence space

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

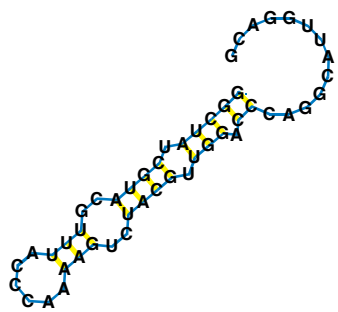
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



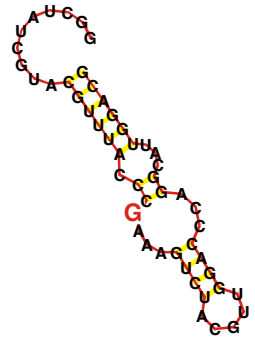
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

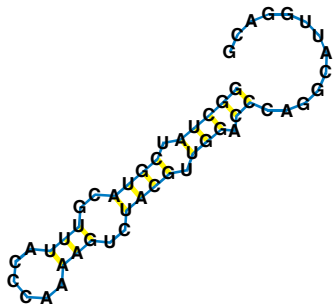


One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

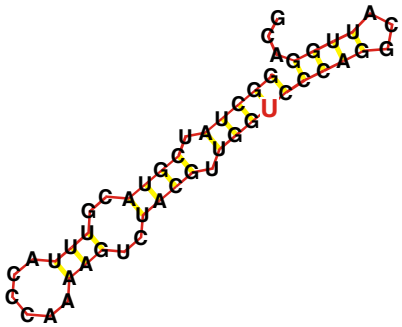


GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

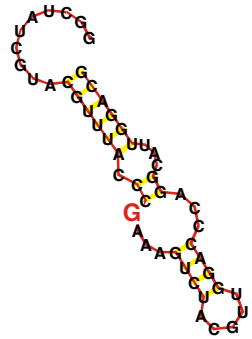
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

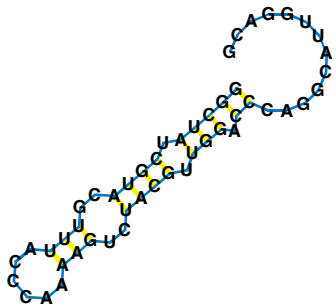


GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGUCCCAGGCAUUGGACG

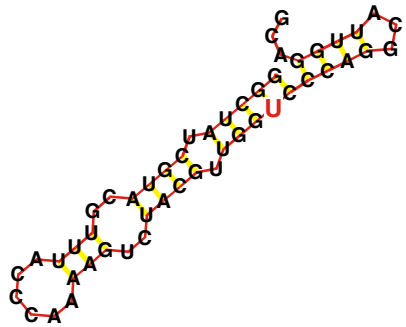


GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

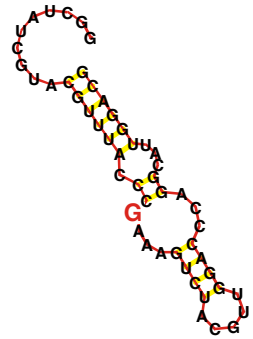
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



One error neighborhood – Surrounding of an RNA molecule in sequence and shape space



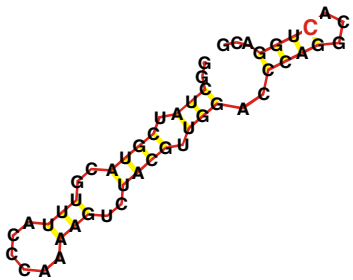
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGUCCCAGGCAUUGGACG

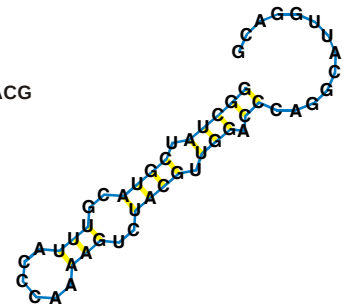


GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

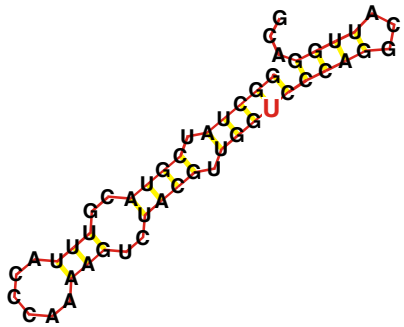
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCACUGGACG

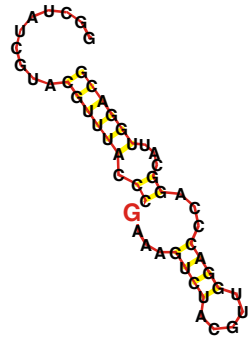




One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

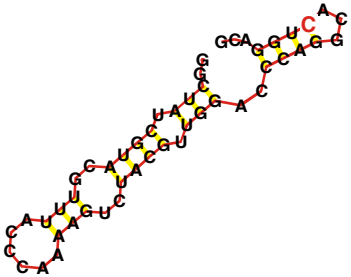


GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGUCCCAGGCAUUGGACG

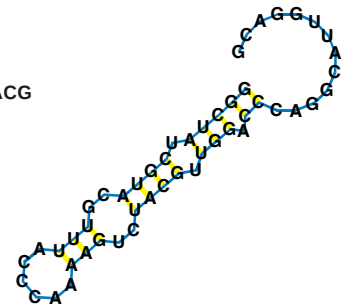


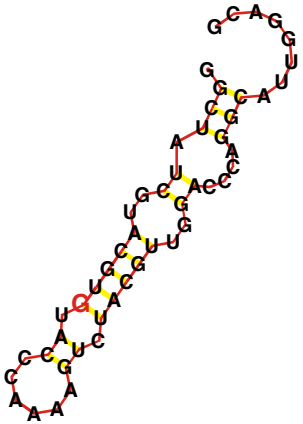
GGCUAUCGUACGUUUACCCGAAAGUCUACGUUGGACCCAGGCAUUGGACG

GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG



GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCACUGGACG

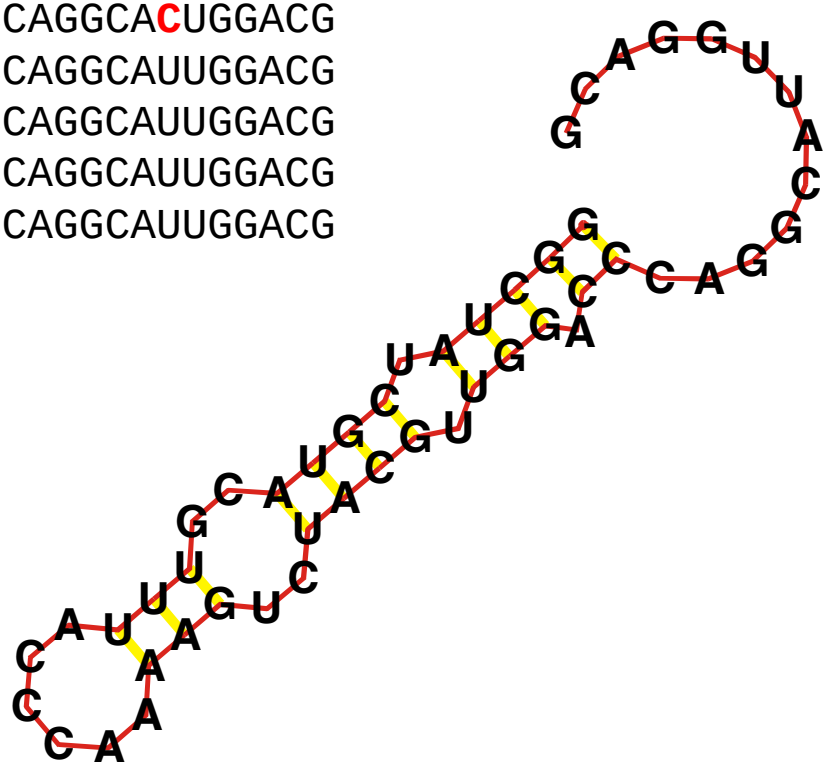




GGCUAUCGUACGUGUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG

One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

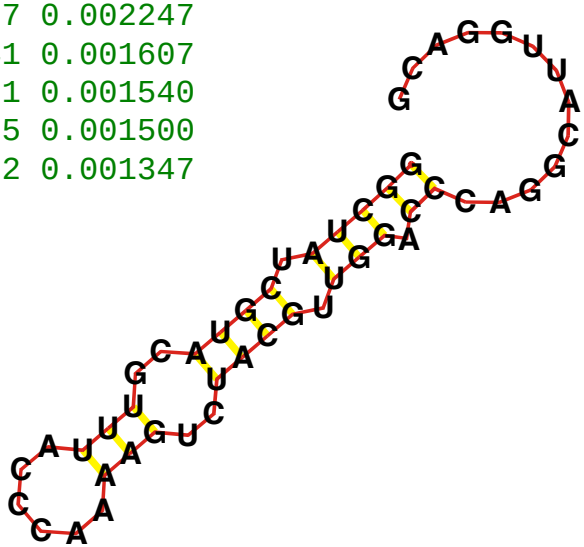
GGCUAUCGUA**U**GUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUU**A**GACG
GGCUAUCGUACGUUUAC**U**CAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACG**C**UUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGC**C**AUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG
GGCUAUCGUACGUUUACCCAAAAGUCUACGUUGGACCCAGGCAUUGGACG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



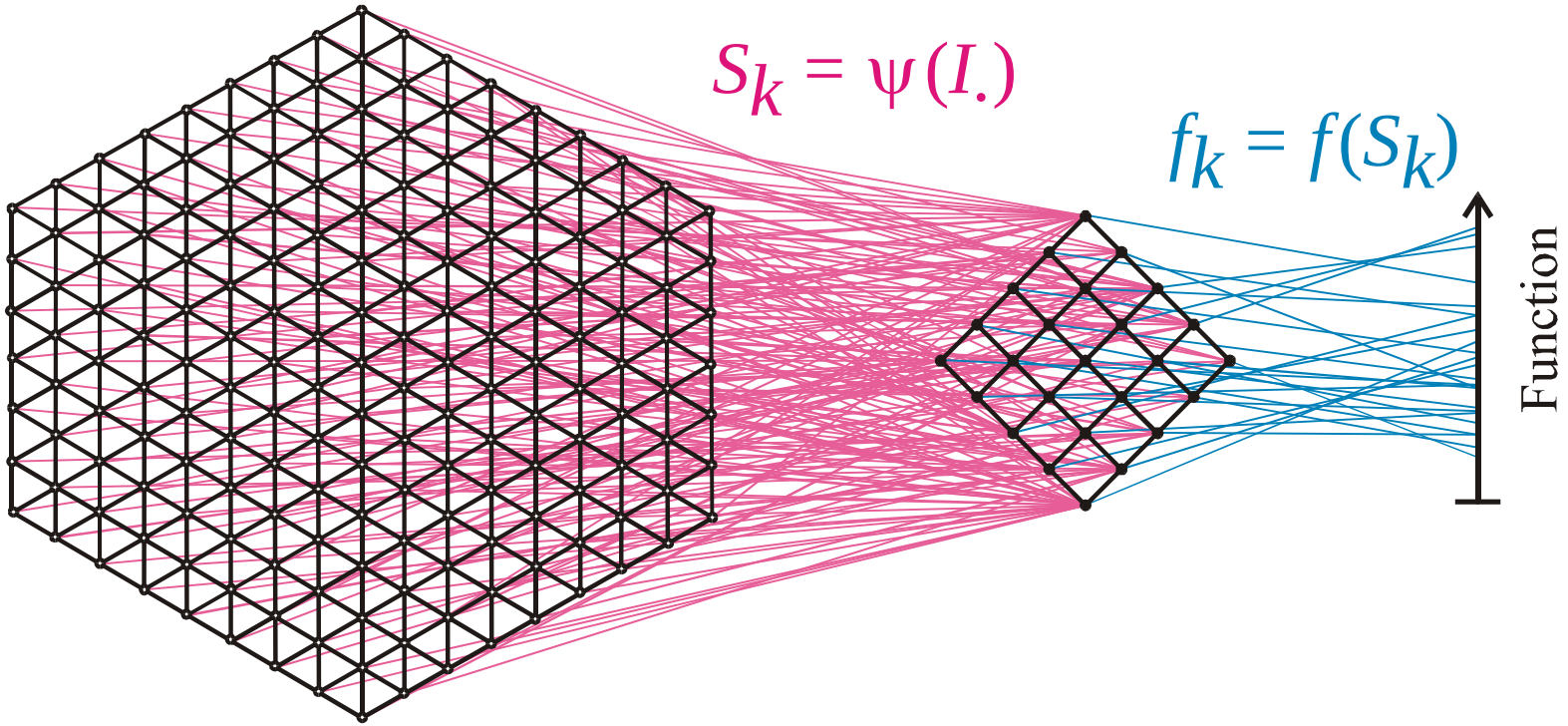
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

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

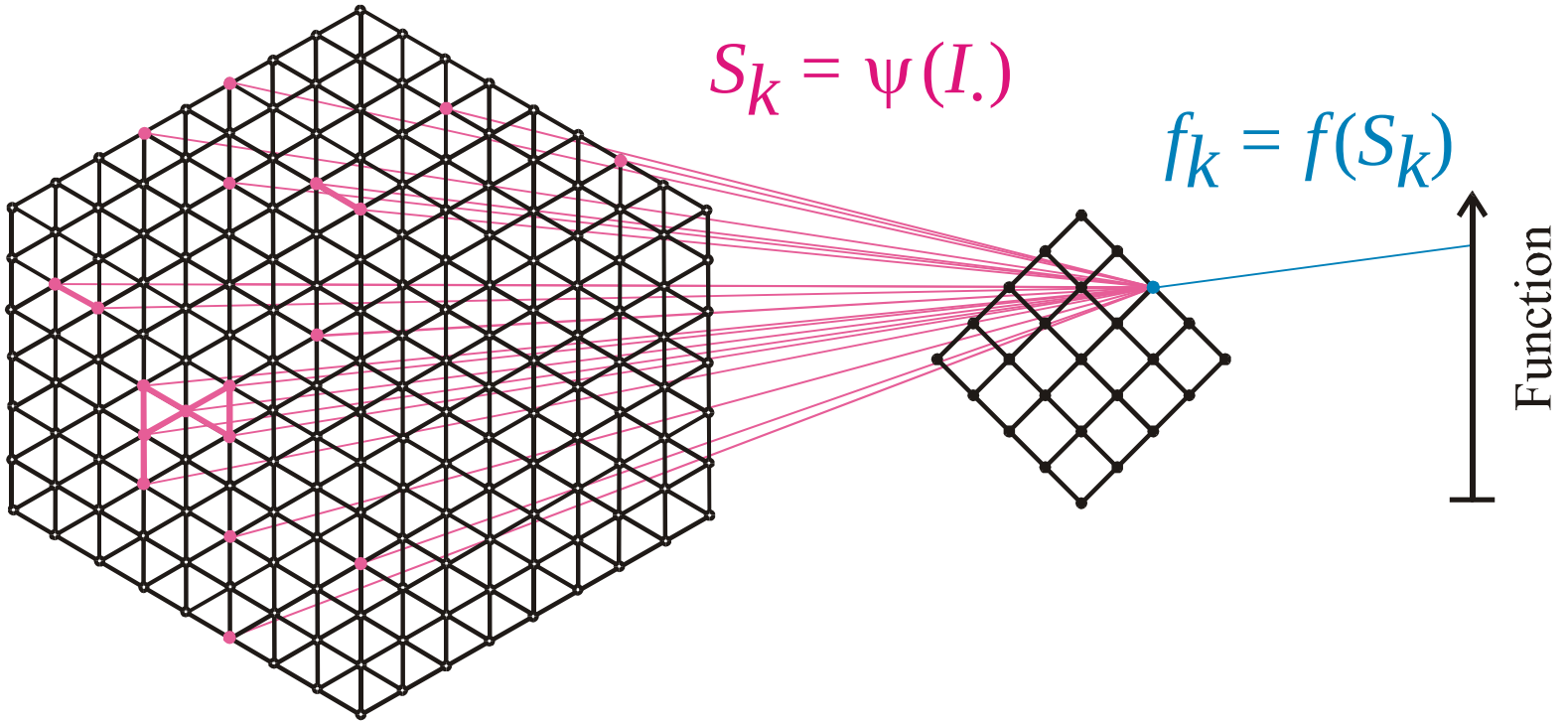
1	(((((.((((..(((.....))))..)))))..))).).....	**50125**	**0.334167**
2	..(((.((((..(((.....))))..)))))..))).).....	**2856**	**0.019040**
3	((((((((((..(((.....))))..)))))))).)).....	**2799**	**0.018660**
4	(((((.((((..((((....))))..)))))..))).)).....	**2417**	**0.016113**
5	(((((.((((.((((.....))))).)))))..))).)).....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

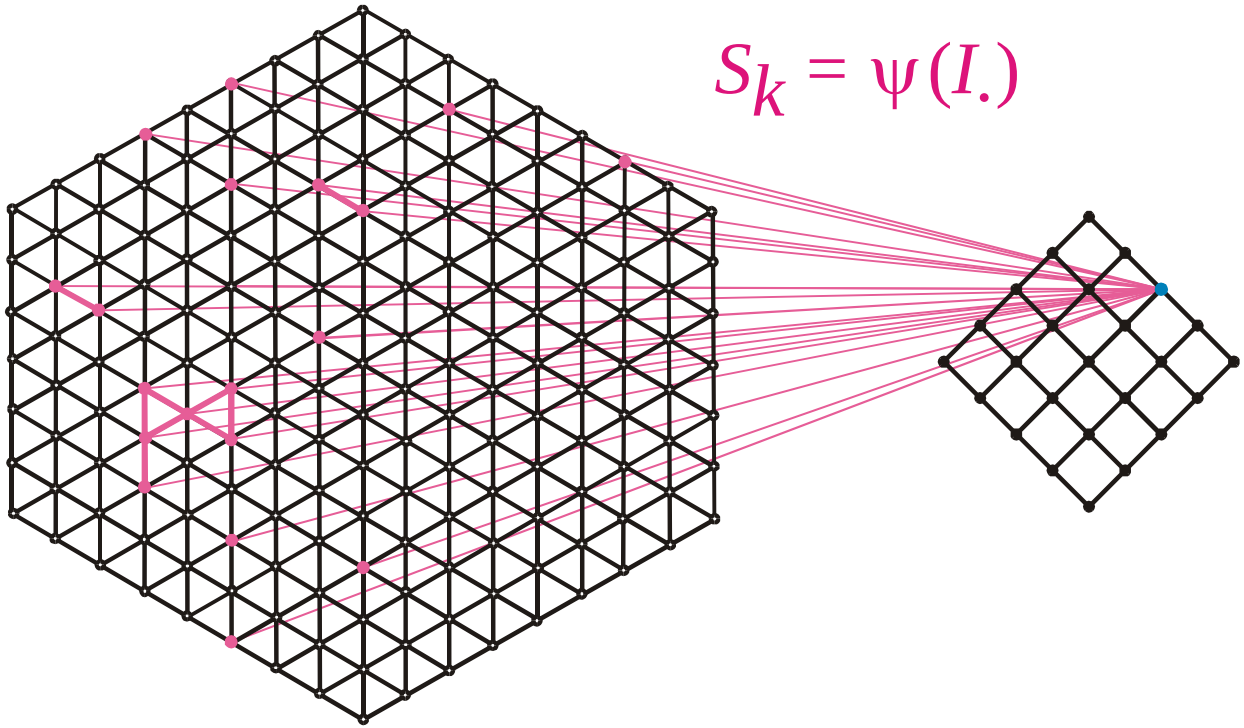


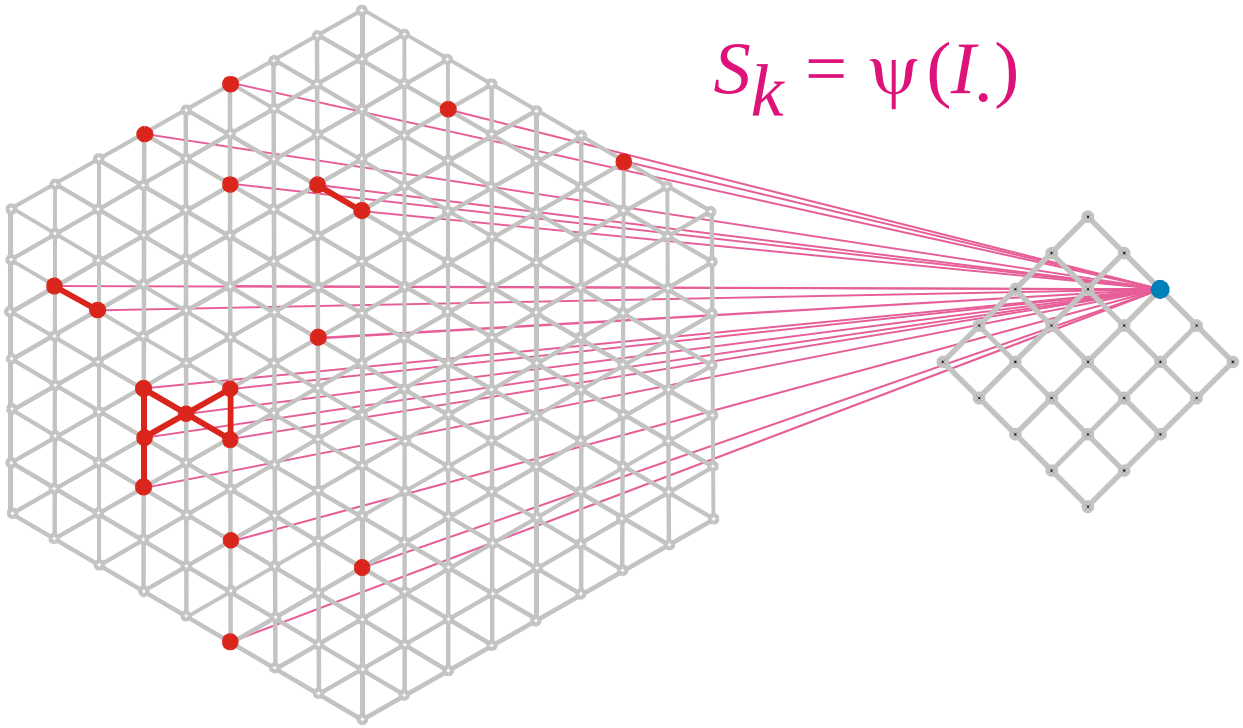
Shadow – Surrounding of an RNA structure in shape space – **AUGC** alphabet



Mapping from sequence space into structure space and into function


$S_k = \psi(I_.)$
$f_k = f(S_k)$
Function
Sequence space
Structure space
Real numbers


$S_k = \psi(I_.)$
Sequence space
Structure space



Sequence space

Structure space

The pre-image of the structure S_k in sequence space is the **neutral network G_k**

A ribozyme switch

E.A.Schultes, D.B.Bartel, Science **289** (2000), 448-452

minus the background levels observed in the HSP in the control (Sar1-GDP–containing) incubation that prevents COPII vesicle formation. In the microsome control, the level of p115-SNARE associations was less than 0.1%.
46. C. M. Carr, E. Grote, M. Munson, F. M. Hughson, P. J. Novick, *J. Cell Biol.* **146**, 333 (1999).
47. C. Ungermann, B. J. Nichols, H. R. Pelham, W. Wickner, *J. Cell Biol.* **140**, 61 (1998).
48. E. Grote and P. J. Novick, *Mol. Biol. Cell* **10**, 4149 (1999).
49. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
50. GST-SNARE proteins were expressed in bacteria and purified on glutathione-Sepharose beads using standard methods. Immobilized GST-SNARE protein (0.5 μM) was incubated with rat liver cytosol (20 mg) or purified recombinant p115 (0.5 μM) in 1 ml of NS buffer containing 1% BSA for 2 hours at 4°C with rotation. Beads were briefly spun (3000 rpm for 10 s) and sequentially washed three times with NS buffer and three times with NS buffer supplemented with 150 mM NaCl. Bound proteins were eluted three times in 50 μl of 50 mM tris-HCl (pH 8.5), 50 mM reduced glutathione, 150 mM NaCl, and 0.1% Triton X-100 for 15 min at 4°C with intermittent mixing, and elutes were pooled. Proteins were precipitated by MeOH/CH_3Cl and separated by SDS–polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting using p115 mAb 13F12.
51. V. Rybin *et al.*, *Nature* **383**, 266 (1996).
52. K. G. Hardwick and H. R. Pelham, *J. Cell Biol.* **119**, 513 (1992).
53. A. P. Newman, M. E. Groesch, S. Ferro-Novick, *EMBO J.* **11**, 3609 (1992).
54. A. Spang and R. Schekman, *J. Cell Biol.* **143**, 589 (1998).
55. M. F. Rexach, M. Latterich, R. W. Schekman, *J. Cell Biol.* **126**, 1133 (1994).
56. A. Mayer and W. Wickner, *J. Cell Biol.* **136**, 307 (1997).
57. M. D. Turner, H. Plutner, W. E. Balch, *J. Biol. Chem.* **272**, 13479 (1997).
58. A. Price, D. Seals, W. Wickner, C. Ungermann, *J. Cell Biol.* **148**, 1231 (2000).
59. X. Cao and C. Barlowe, *J. Cell Biol.* **149**, 55 (2000).
60. G. G. Tall, H. Hama, D. B. DeWald, B. F. Horazdovsky, *Mol. Biol. Cell* **10**, 1873 (1999).
61. C. G. Burd, M. Peterson, C. R. Cowles, S. D. Emr, *Mol. Biol. Cell* **8**, 1089 (1997).
62. M. R. Peterson, C. G. Burd, S. D. Emr, *Curr. Biol.* **9**, 159 (1999).
63. M. G. Waters, D. O. Clary, J. E. Rothman, *J. Cell Biol.* **118**, 1015 (1992).
64. D. M. Walter, K. S. Paul, M. G. Waters, *J. Biol. Chem.* **273**, 29565 (1998).
65. N. Hui *et al.*, *Mol. Biol. Cell* **8**, 1777 (1997).
66. T. E. Kreis, *EMBO J.* **5**, 931 (1986).
67. H. Plutner, H. W. Davidson, J. Saraste, W. E. Balch, *J. Cell Biol.* **119**, 1097 (1992).
68. D. S. Nelson *et al.*, *J. Cell Biol.* **143**, 319 (1998).
69. We thank G. Waters for p115 cDNA and p115 mAbs; G. Warren for p97 and p47 antibodies; R. Scheller for rbet1, membrin, and sec22 cDNAs; H. Plutner for excellent technical assistance; and P. Tan for help during the initial phase of this work. Supported by NIH grants GM 33301 and GM42336 and National Cancer Institute grant CA58689 (W.E.B.), a NIH National Research Service Award (B.D.M.), and a Wellcome Trust International Traveling Fellowship (B.B.A.).

20 March 2000; accepted 22 May 2000

One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds

Erik A. Schultes and David P. Bartel*

We describe a single RNA sequence that can assume either of two ribozyme folds and catalyze the two respective reactions. The two ribozyme folds share no evolutionary history and are completely different, with no base pairs (and probably no hydrogen bonds) in common. Minor variants of this sequence are highly active for one or the other reaction, and can be accessed from prototype ribozymes through a series of neutral mutations. Thus, in the course of evolution, new RNA folds could arise from preexisting folds, without the need to carry inactive intermediate sequences. This raises the possibility that biological RNAs having no structural or functional similarity might share a common ancestry. Furthermore, functional and structural divergence might, in some cases, precede rather than follow gene duplication.

Related protein or RNA sequences with the same folded conformation can often perform very different biochemical functions, indicating that new biochemical functions can arise from preexisting folds. But what evolutionary mechanisms give rise to sequences with new macromolecular folds? When considering the origin of new folds, it is useful to picture, among all sequence possibilities, the distribution of sequences with a particular fold and function. This distribution can range very far in sequence space (*1*). For example, only seven nucleotides are strictly conserved among the group I self-splicing introns, yet secondary (and presumably tertiary) structure within the core of the ribozyme is preserved (*2*). Because these disparate isolates have the same fold and function, it is thought that they descended from a common ancestor through a series of mutational variants that were each functional. Hence, sequence heterogeneity among divergent isolates implies the existence of paths through sequence space that have allowed neutral drift from the ancestral sequence to each isolate. The set of all possible neutral paths composes a "neutral network," connecting in sequence space those widely dispersed sequences sharing a particular fold and activity, such that any sequence on the network can potentially access very distant sequences by neutral mutations (*3–5*).

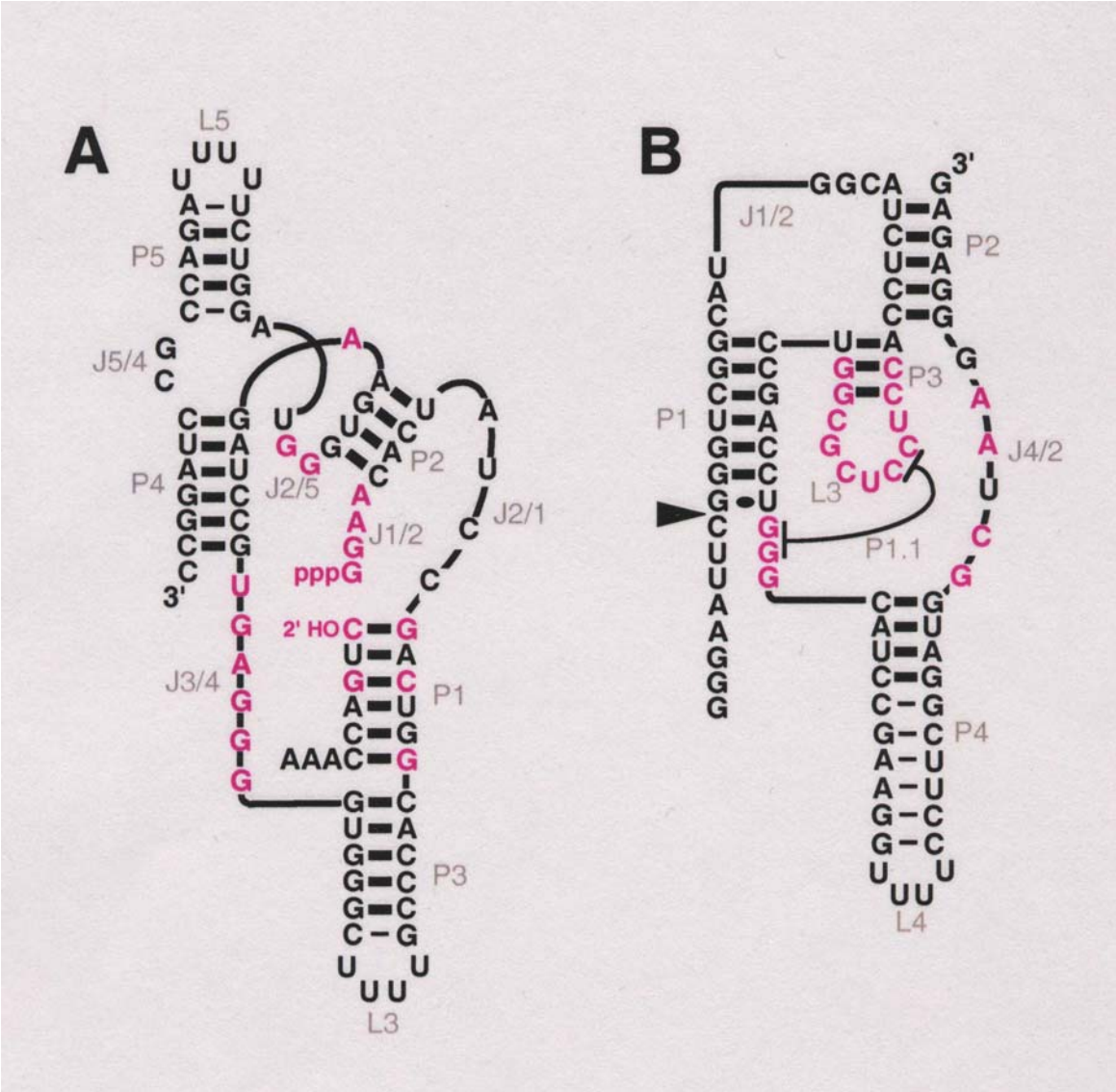
Theoretical analyses using algorithms for predicting RNA secondary structure have suggested that different neutral networks are interwoven and can approach each other very closely (*3*, *5–8*). Of particular interest is whether ribozyme neutral networks approach each other so closely that they intersect. If so, a single sequence would be capable of folding into two different conformations, would have two different catalytic activities, and could access by neutral drift every sequence on both networks. With intersecting networks, RNAs with novel structures and activities could arise from previously existing ribozymes, without the need to carry nonfunctional sequences as evolutionary intermediates. Here, we explore the proximity of neutral networks experimentally, at the level of RNA function. We describe a close apposition of the neutral networks for the hepatitis delta virus (HDV) self-cleaving ribozyme and the class III self-ligating ribozyme.

In choosing the two ribozymes for this investigation, an important criterion was that they share no evolutionary history that might confound the evolutionary interpretations of our results. Choosing at least one artificial ribozyme ensured independent evolutionary histories. The class III ligase is a synthetic ribozyme isolated previously from a pool of random RNA sequences (*9*). It joins an oligonucleotide substrate to its 5′ terminus. The prototype ligase sequence (Fig. 1A) is a shortened version of the most active class III variant isolated after 10 cycles of in vitro selection and evolution. This minimal construct retains the activity of the full-length isolate (*10*). The HDV ribozyme carries out the site-specific self-cleavage reactions needed during the life cycle of HDV, a satellite virus of hepatitis B with a circular, single-stranded RNA genome (*11*). The prototype HDV construct for our study (Fig. 1B) is a shortened version of the antigenomic HDV ribozyme (*12*), which undergoes self-cleavage at a rate similar to that reported for other antigenomic constructs (*13*, *14*).

The prototype class III and HDV ribozymes have no more than the 25% sequence identity expected by chance and no fortuitous structural similarities that might favor an intersection of their two neutral networks. Nevertheless, sequences can be designed that simultaneously satisfy the base-pairing requirements

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

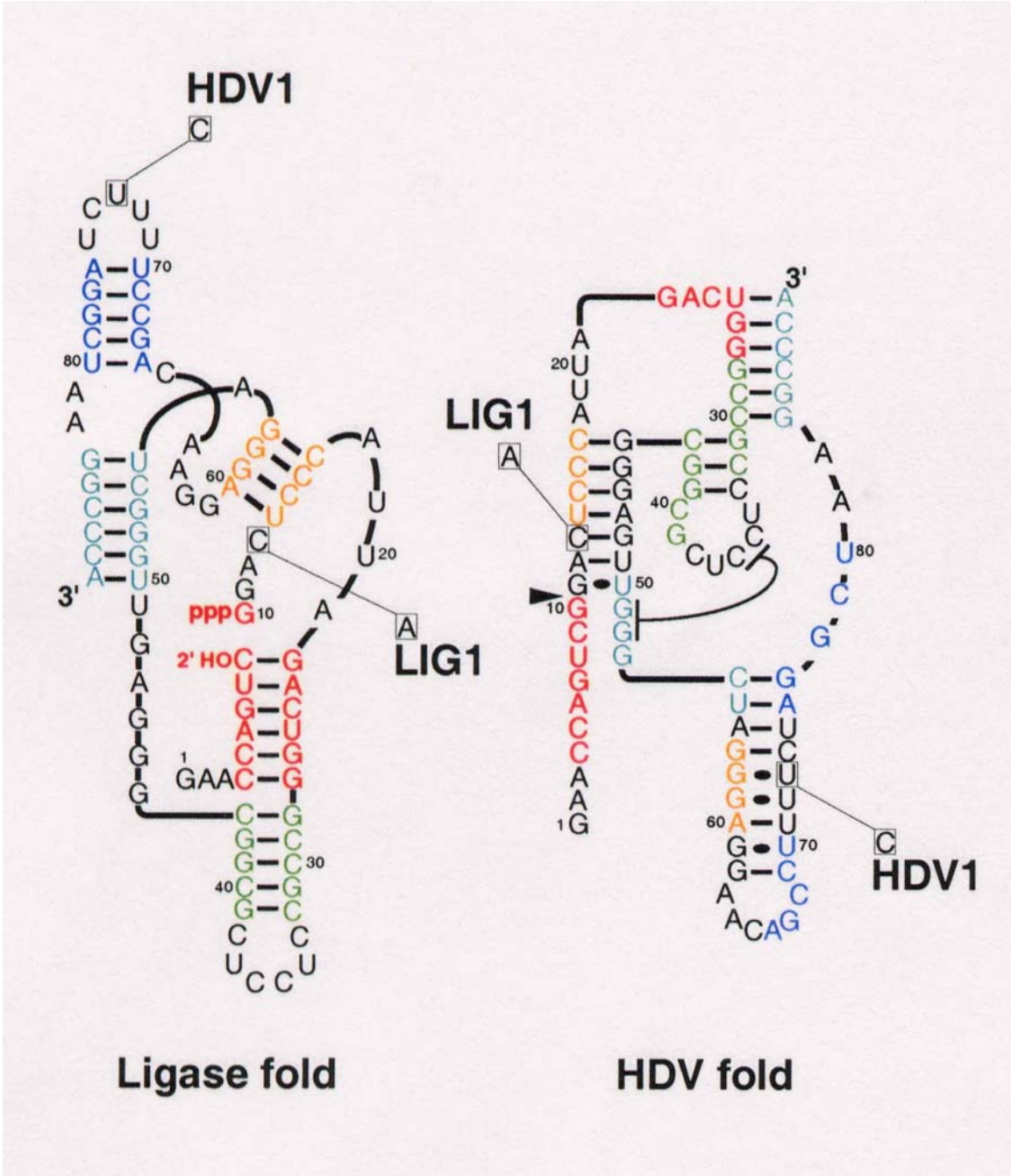
*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

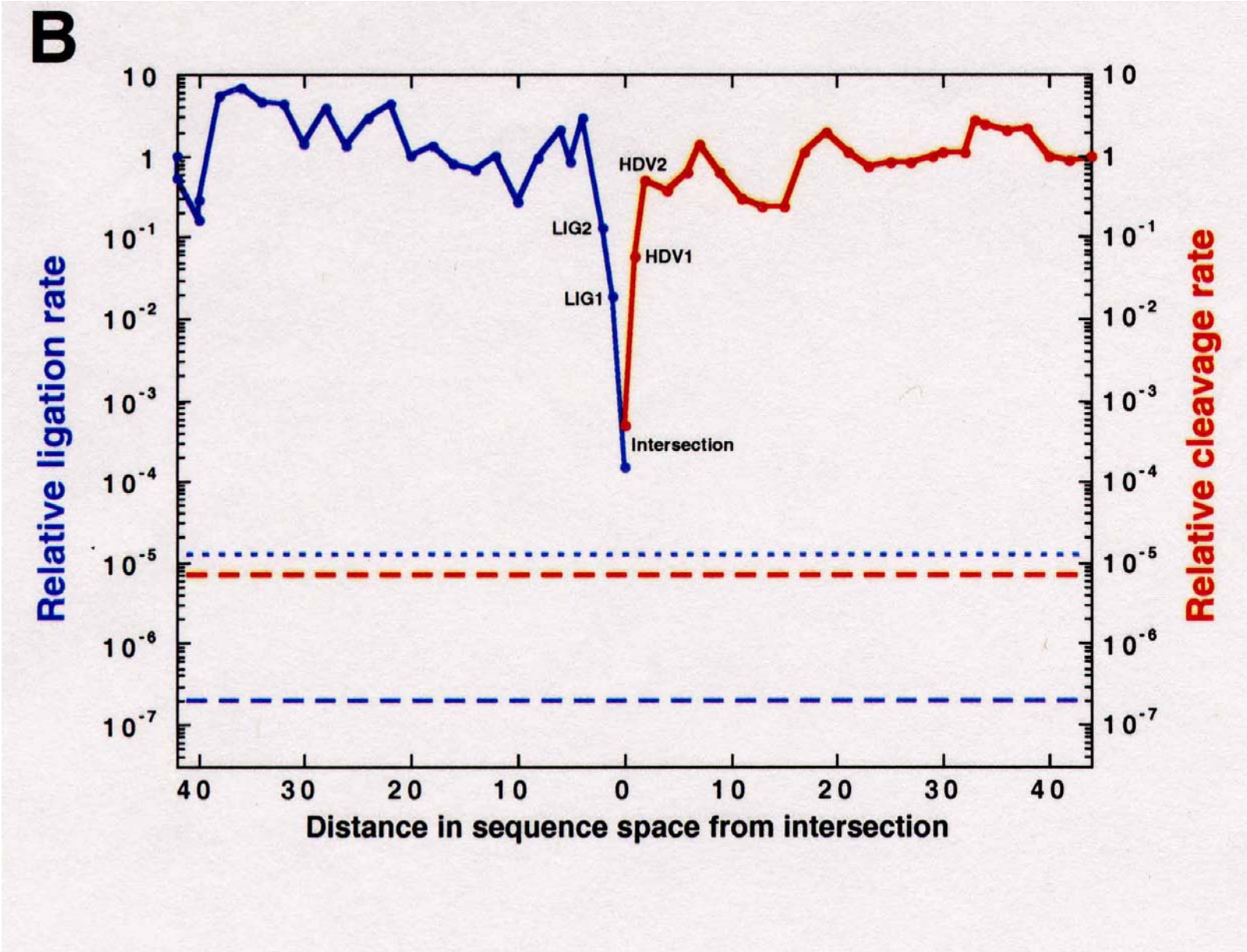


Two ribozymes of chain lengths n = 88 nucleotides: An artificial ligase (**A**) and a natural cleavage ribozyme of hepatitis-δ-virus (**B**)

The sequence at the ***intersection***:

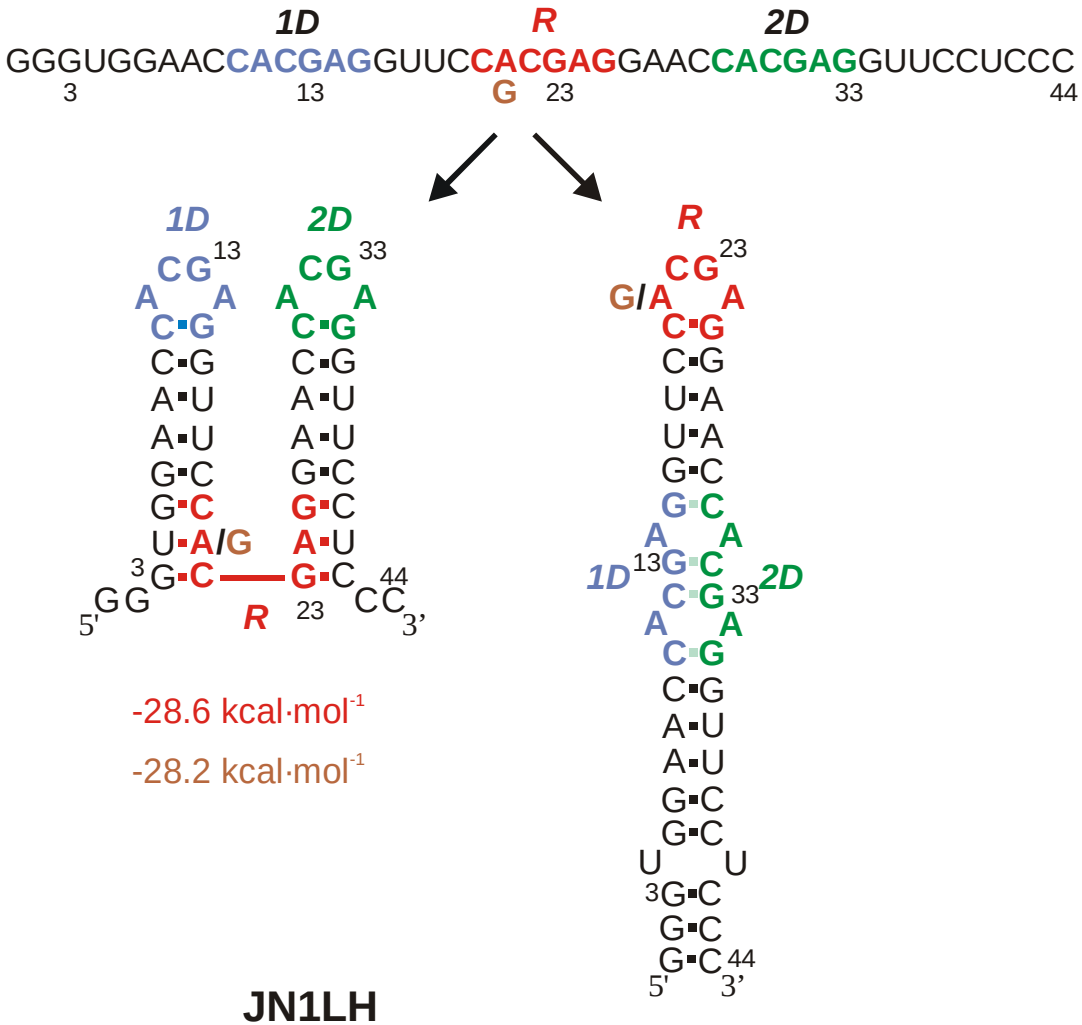
An RNA molecules which is 88 nucleotides long and can form both structures





Two neutral walks through sequence space with conservation of structure and catalytic activity

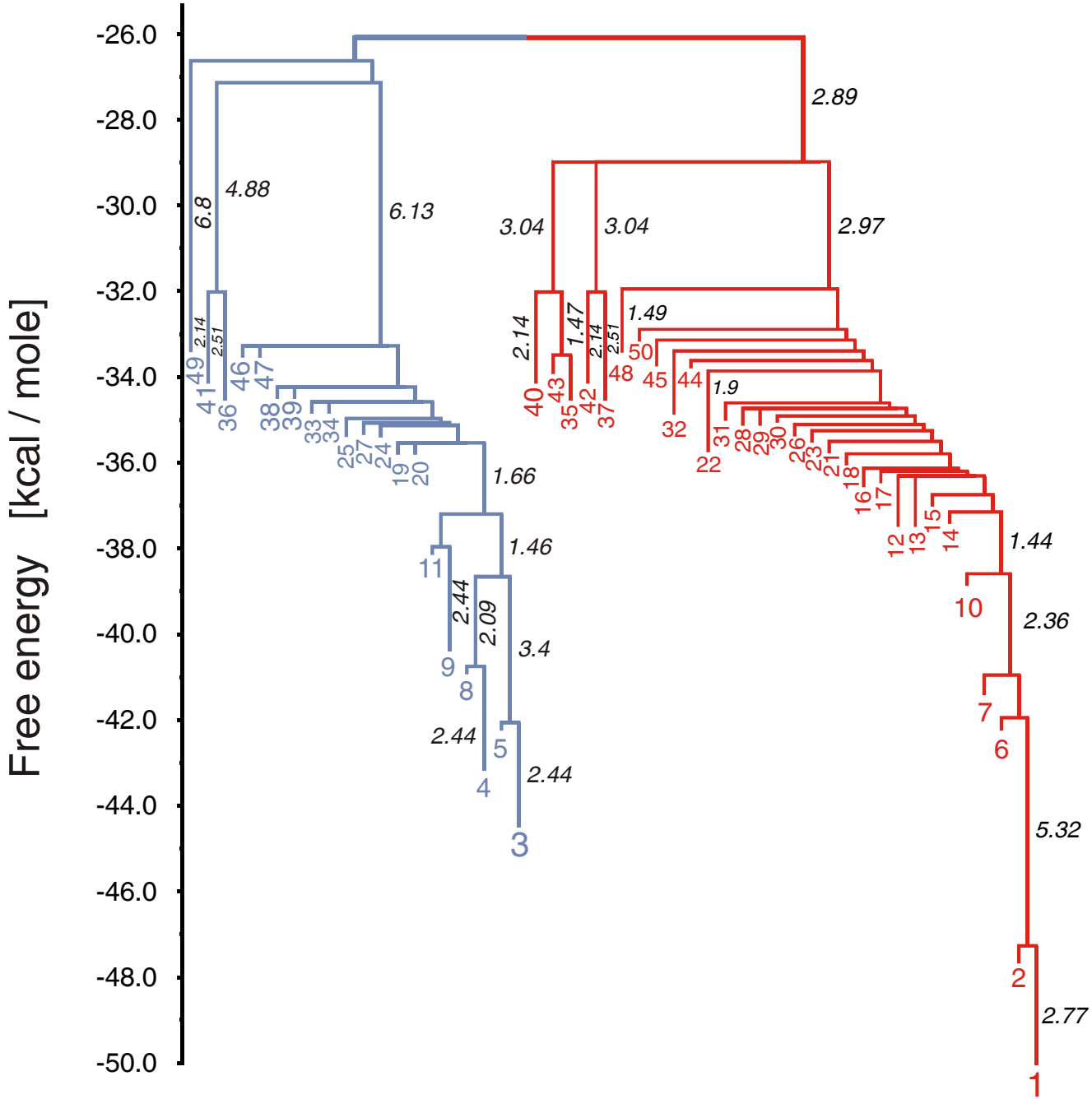
An experimental RNA switch



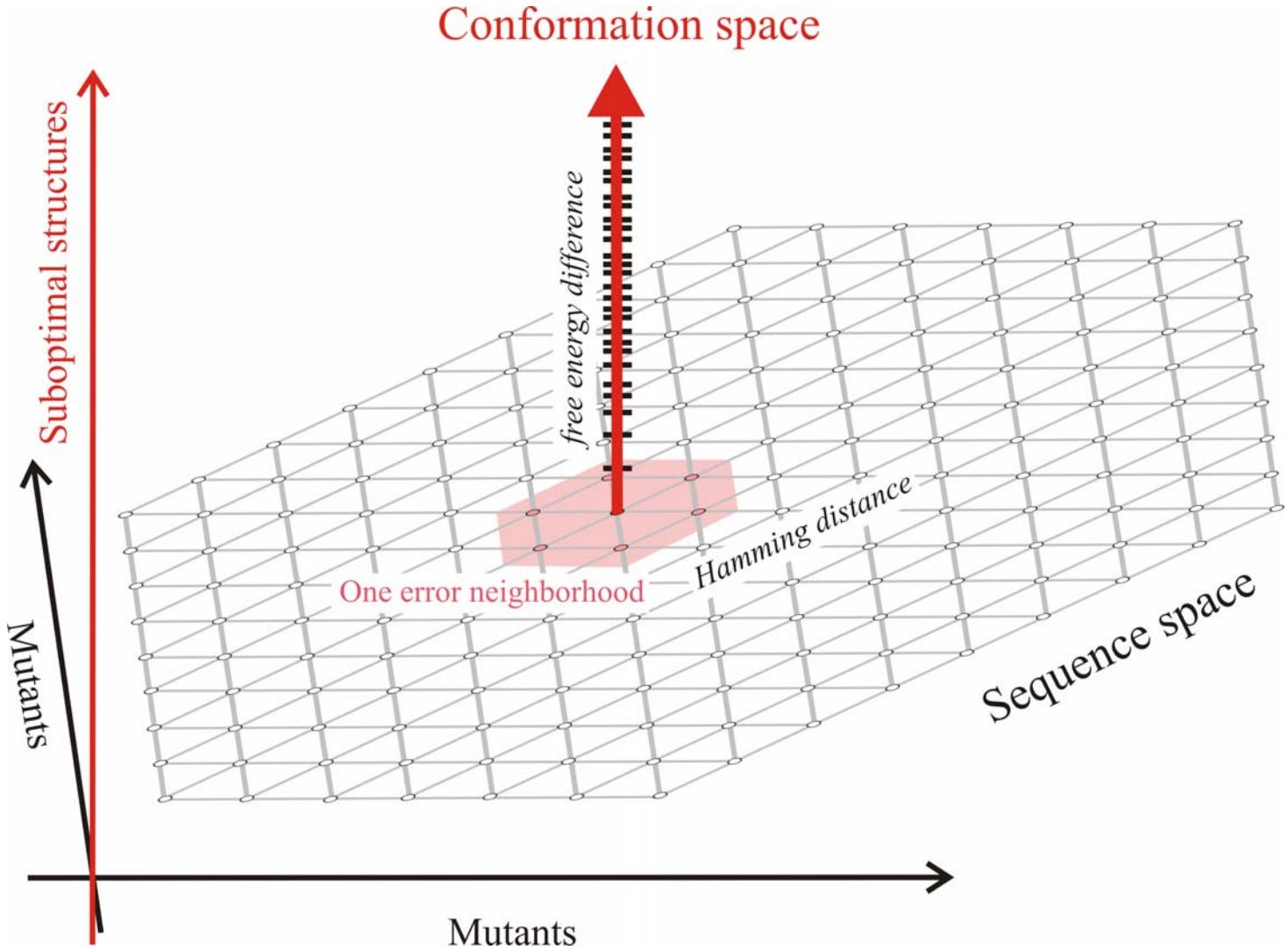
JN1LH

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

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



J1LH barrier tree



Construction of a combined landscape for folding and evolution

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

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

Christian Reidys, **Christian Forst**, Los Alamos National Laboratory, NM

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**, Universität Wien, AT

Jan Cupal, **Stefan Bernhart**, **Lukas Endler, Ulrike Langhammer**, **Rainer Machne, Ulrike Mückstein**, **Hakim Tafer, Thomas Taylor,** Universität Wien, AT

Web-Page for further information:

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