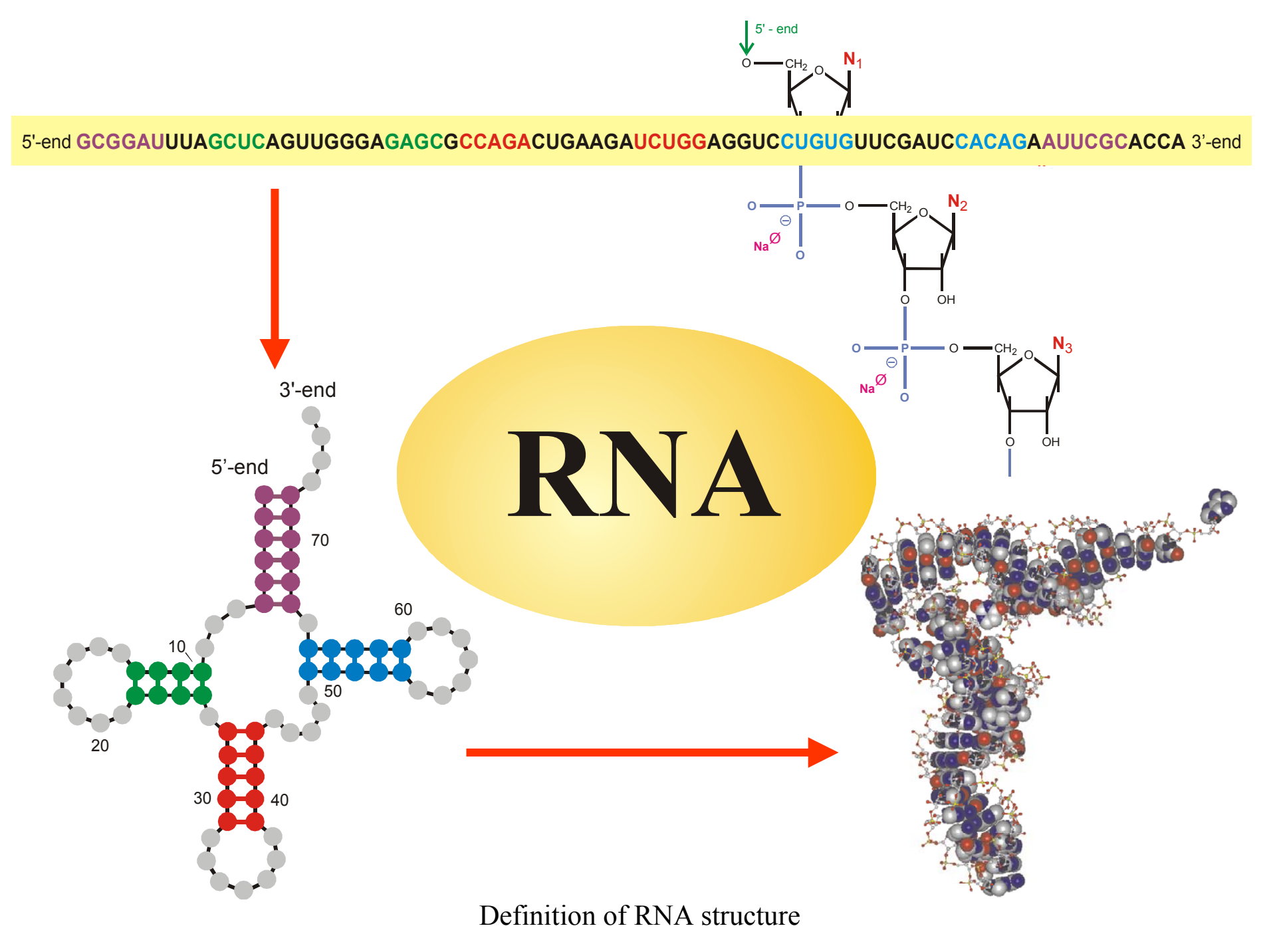
1. Autocatalytic chemical reactions in the flow reactor

2. Replication, mutation, selection and Shannon information

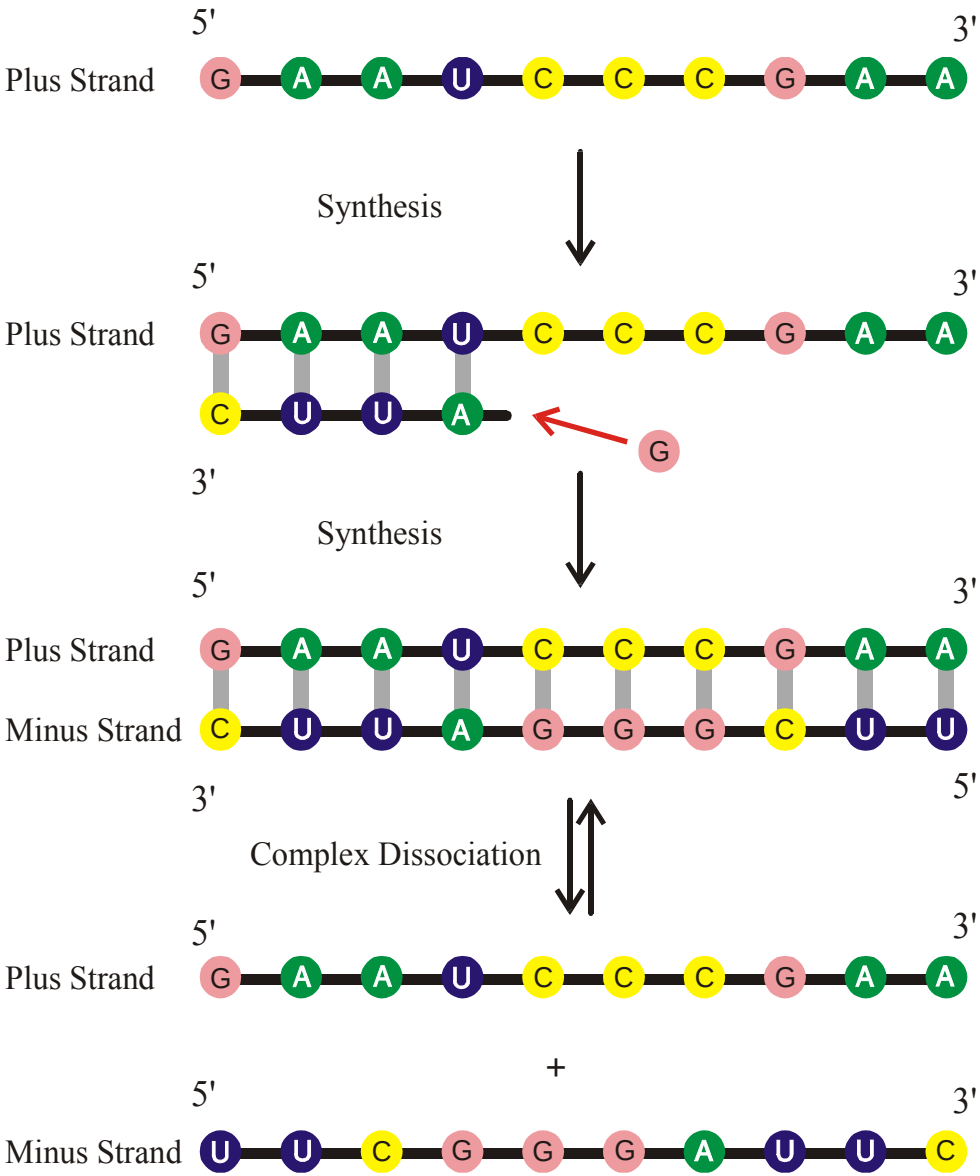
3. Evolution *in silico* and optimization of RNA structures

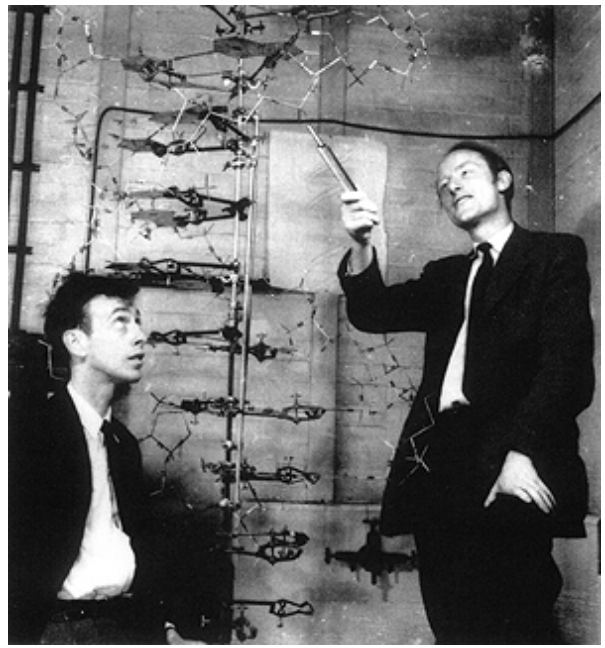
4. Random walks and ‚ensemble learning‘

5. Sequence-structure maps, neutral networks, and intersections



Definition of RNA structure





James Watson and Francis Crick, 1953

50 years double-helix: 1953-2003

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

G Í C and A=U

$$dx_j / dt = \Sigma_i k_i x_i - x_j \Phi$$

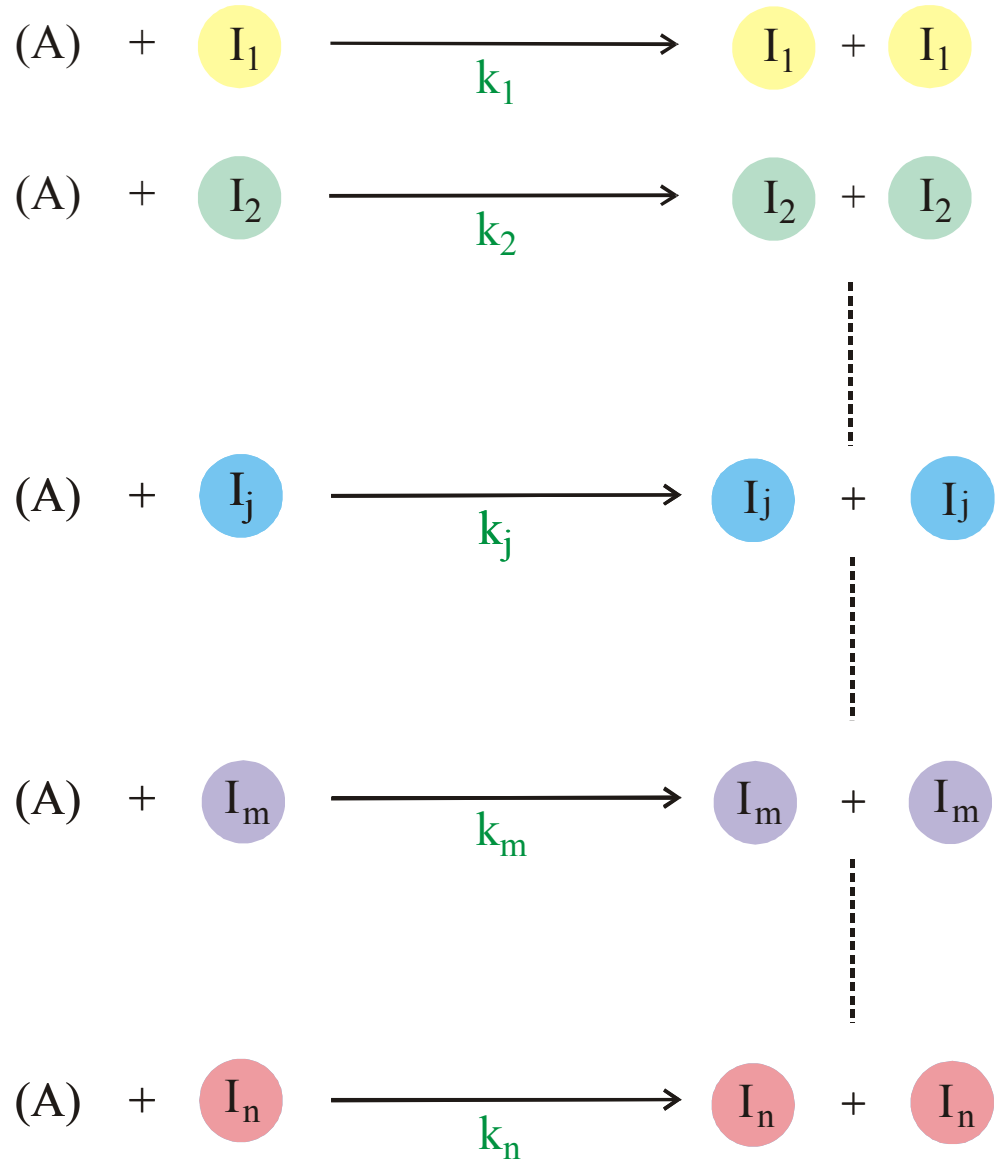
$$\Phi = \Sigma_i k_i x_i ; \quad \Sigma_i x_i = 1$$

[A] = a = constant

$k_m = \max \{k_j; j=1,2,...,n\}$

$x_m(t)$ Á 1 for t Á ¸

$s = (k_{m+1}-k_m)/k_m$



Selection of the „fittest“ or fastest replicating species

Selection equation: $[I_i] = x_i \neq 0\ ,\ f_i > 0$

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

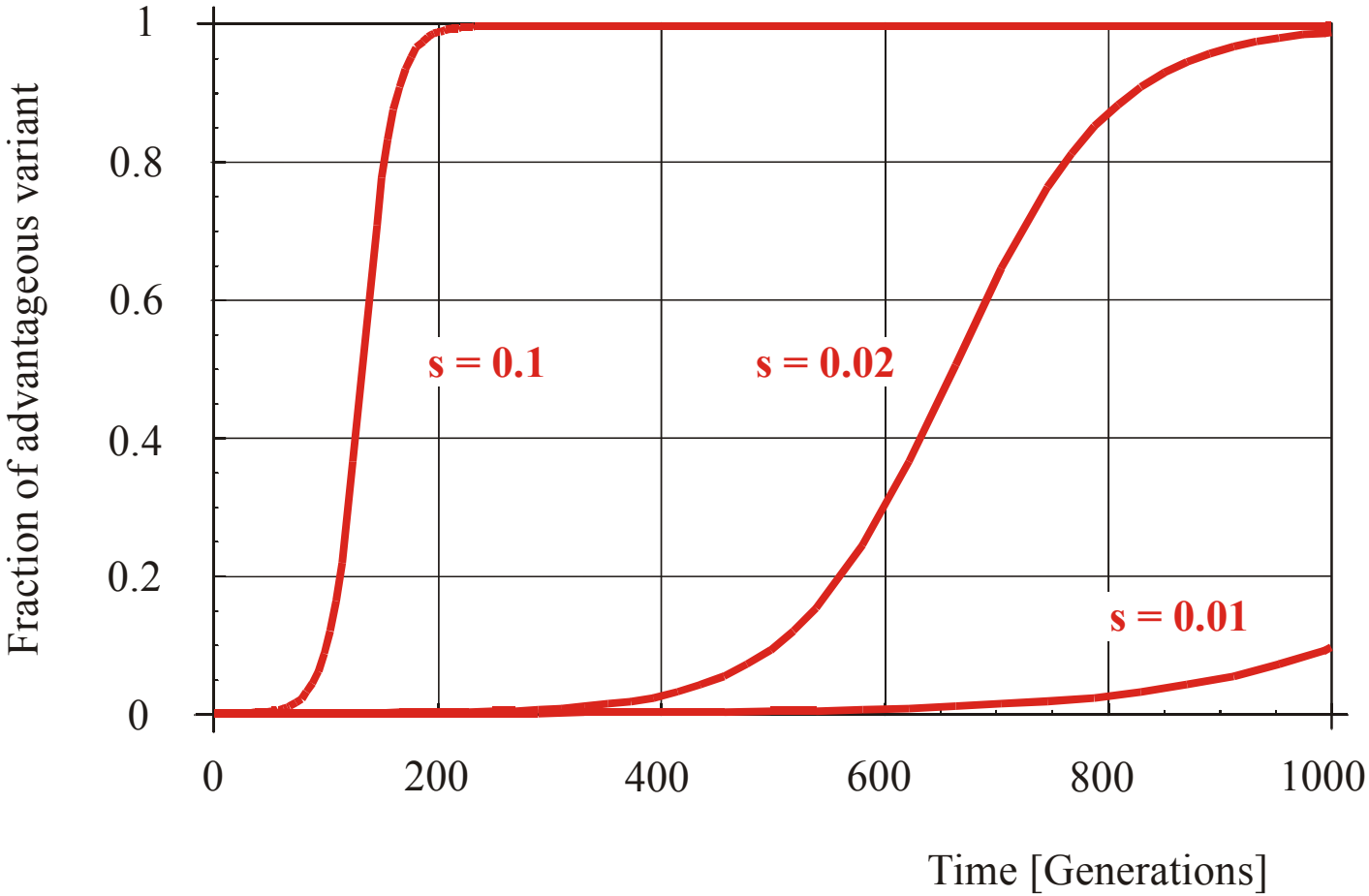
Mean fitness or dilution flux, ϕ(t), is a **non-decreasing function** of time,

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

Solutions are obtained by integrating factor transformation

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

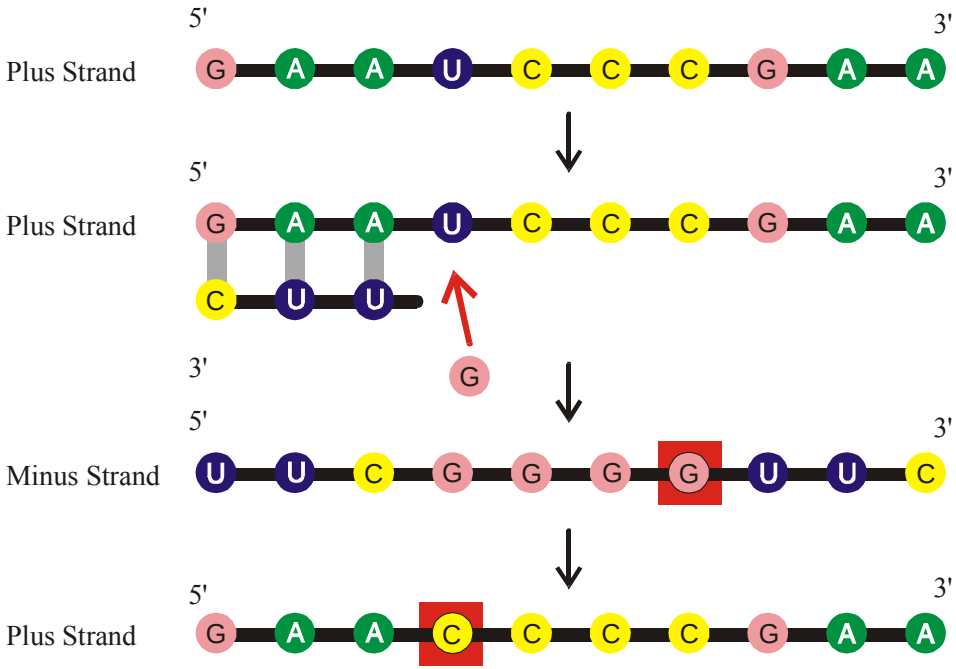
$\mathbf{s} = (f_2 - f_1) / f_1;\ f_2 > f_1\ ;\ x_1(0) = 1 - 1/N\ ;\ x_2(0) = 1/N$



Selection of advantageous mutants in populations of N = 10 000 individuals

Changes in RNA sequences originate from replication errors called **mutations**.

Mutations occur uncorrelated to their consequences in the selection process and are, therefore, commonly characterized as **random elements** of evolution.



Point Mutation

GAAUCCCGAA → GAAUCCCGUCCCGAA

Insertion

GAAUCCCGAA → GAAUCCA

Deletion

The origins of changes in RNA sequences are **replication errors** called **mutations**.

Theory of molecular evolution

M.Eigen, ***Self-organization of matter and the evolution of biological macromolecules***. Naturwissenschaften **58** (1971), 465-526

C.J. Thompson, J.L. McBride, ***On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules***. Math. Biosci. **21** (1974), 127-142

B.L. Jones, R.H. Enns, S.S. Rangnekar, ***On the theory of selection of coupled macromolecular systems.*** Bull.Math.Biol. **38** (1976), 15-28

M.Eigen, P.Schuster, ***The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle***. Naturwissenschaften **58** (1977), 465-526

M.Eigen, P.Schuster, ***The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle***. Naturwissenschaften **65** (1978), 7-41

M.Eigen, P.Schuster, ***The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle***. Naturwissenschaften **65** (1978), 341-369

J. Swetina, P. Schuster, ***Self-replication with errors - A model for polynucleotide replication.*** Biophys.Chem. **16** (1982), 329-345

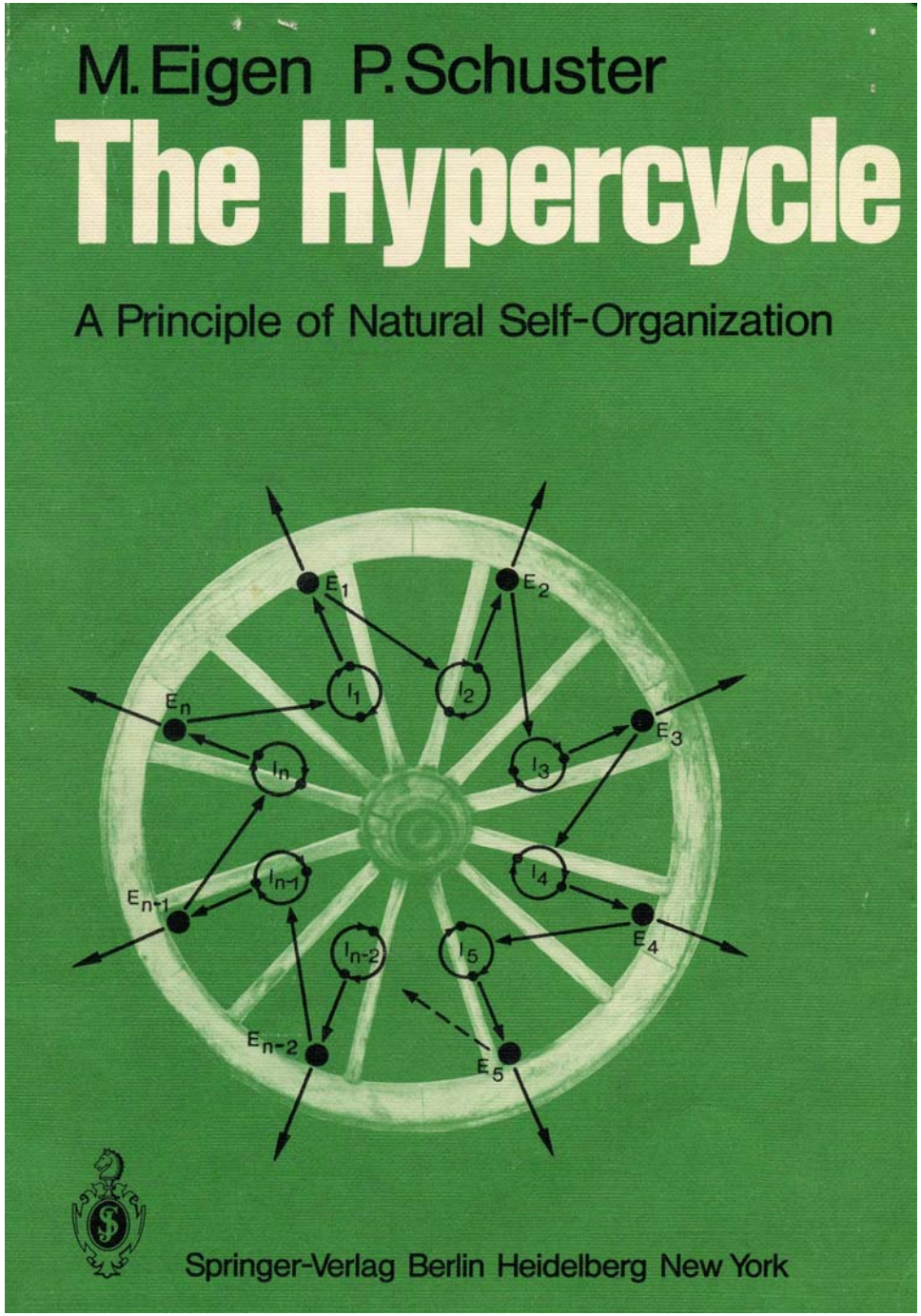
J.S. McCaskill, ***A localization threshold for macromolecular quasispecies from continuously distributed replication rates***. J.Chem.Phys. **80** (1984), 5194-5202

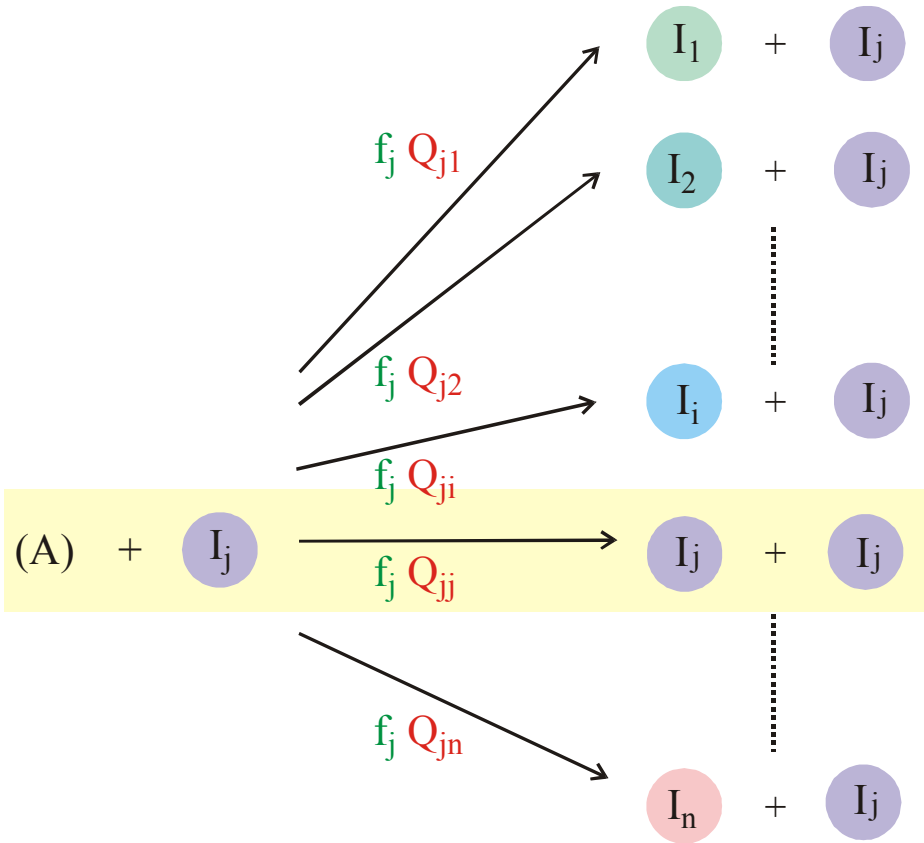
M.Eigen, J.McCaskill, P.Schuster, ***The molecular quasispecies***. Adv.Chem.Phys. **75** (1989), 149-263

C. Reidys, C.Forst, P.Schuster, ***Replication and mutation on neutral networks***. Bull.Math.Biol. **63** (2001), 57-94

Chemical kinetics of molecular evolution

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





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

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

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

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

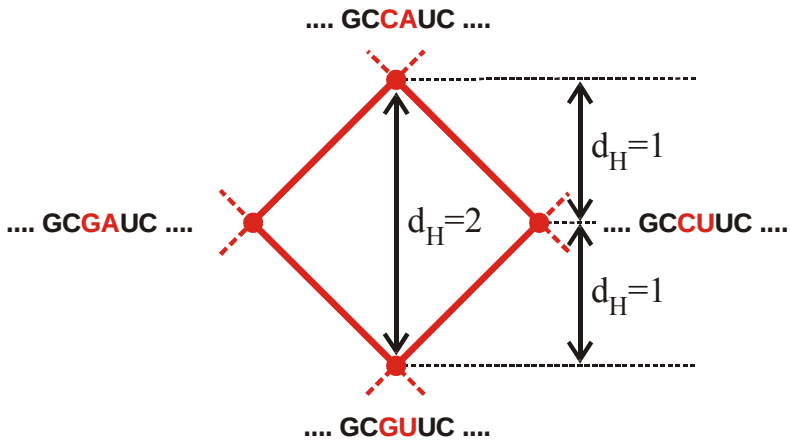
$$Q_{ij} = (1-p)^{\ell - d(i,j)} p^{d(i,j)}$$

p ..... Error rate per digit

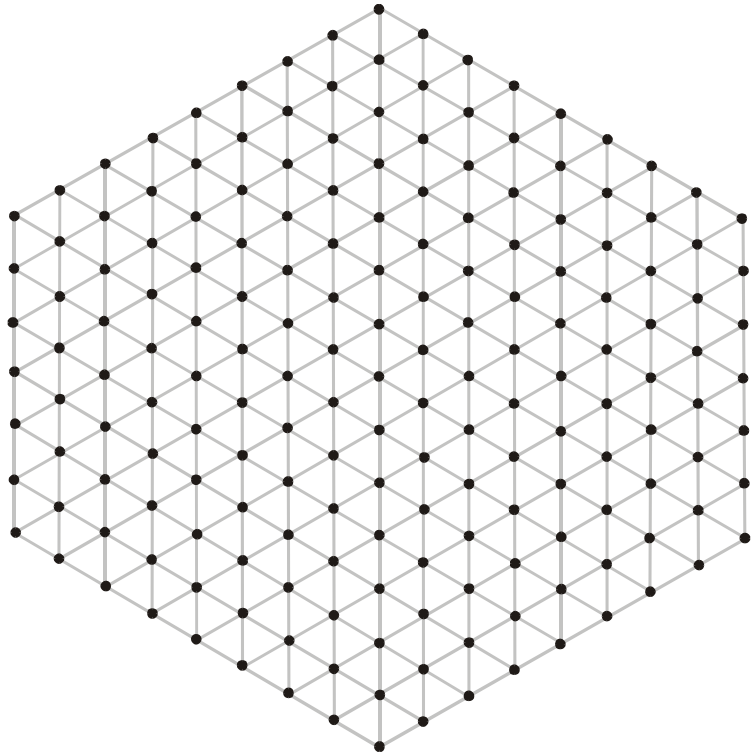
ℓ ..... Chain length of the polynucleotide

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

Chemical kinetics of replication and mutation as parallel reactions

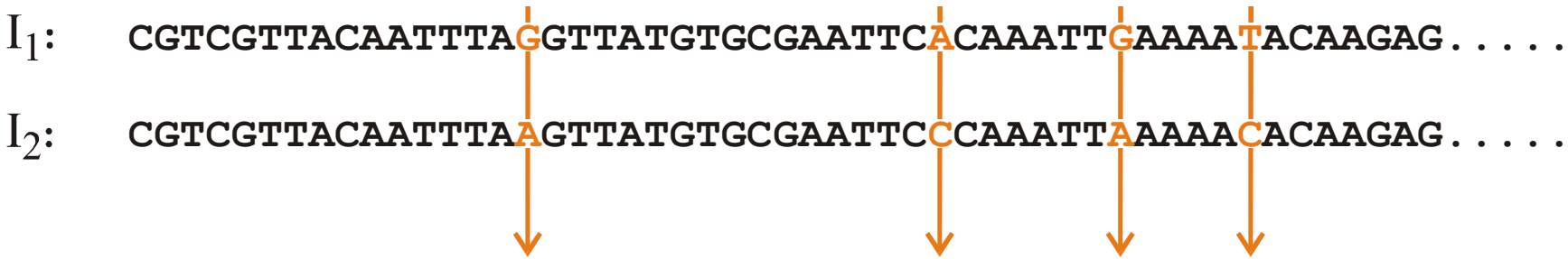


City-block distance in sequence space



2D Sketch of sequence space

Single point mutations as moves in sequence space



Hamming distance $d_H(I_1,I_2) = 4$

(i) $d_H(I_1,I_1) = 0$

(ii) $d_H(I_1,I_2) = d_H(I_2,I_1)$

(iii) $d_H(I_1,I_3) \P d_H(I_1,I_2) + d_H(I_2,I_3)$

The Hamming distance between sequences induces a metric in sequence space

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

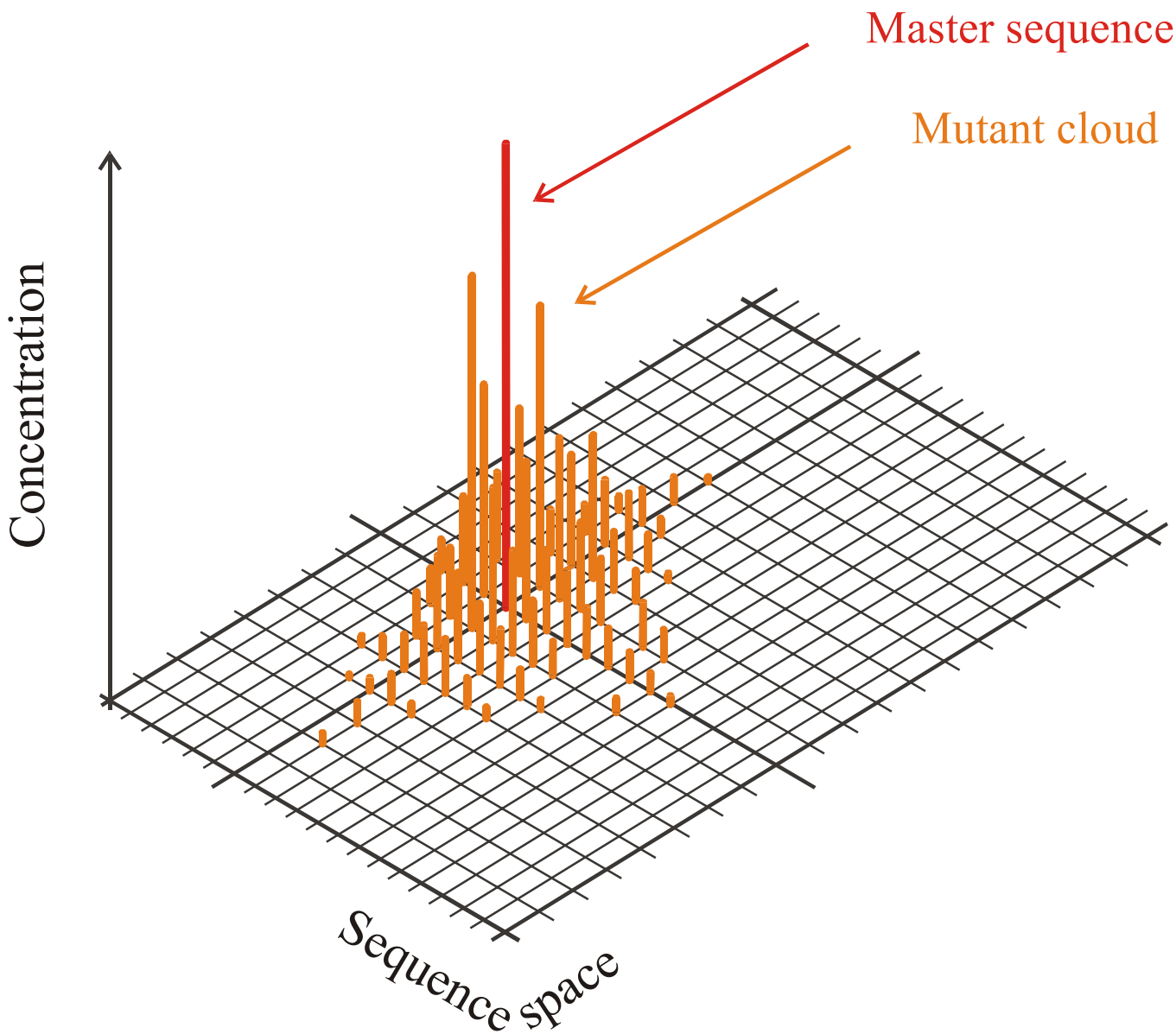
$$\frac{dx_i}{dt} = \sum_{j=1}^{n} f_j Q_{ji}\, x_j - x_i\, \phi, \quad i = 1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \overline{f}$$

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

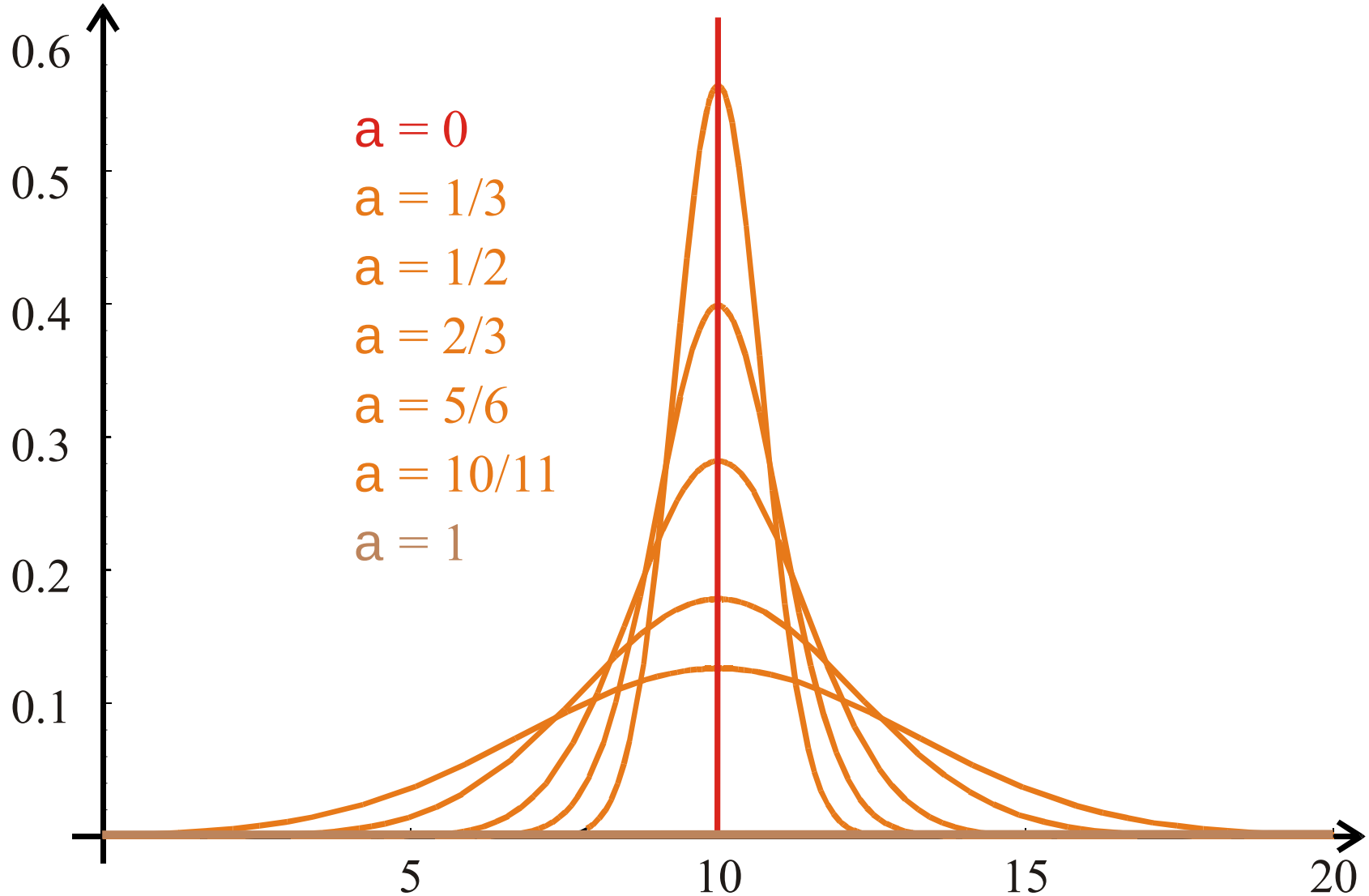
$$x_i(t) = \frac{\sum_{k=0}^{n-1} \ell_{ik} \cdot c_k(0) \cdot \exp(\lambda_k t)}{\sum_{j=1}^{n} \sum_{k=0}^{n-1} \ell_{jk} \cdot c_k(0) \cdot \exp(\lambda_k t)}; \quad i = 1,2,\cdots,n; \quad c_k(0) = \sum_{i=1}^{n} h_{ki}\, x_i(0)$$

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

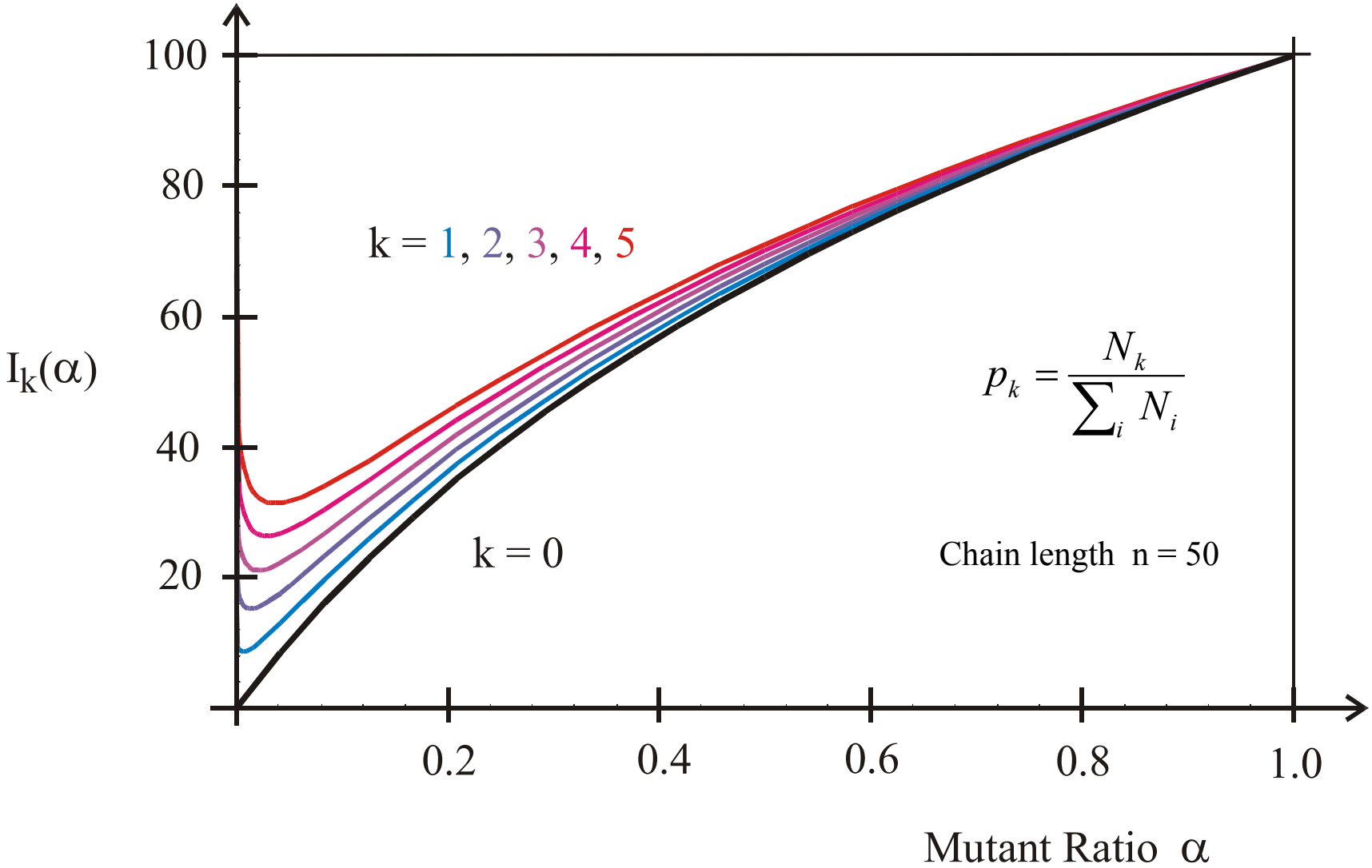
$$L^{-1} \cdot W \cdot L \quad = \quad \Lambda \quad = \quad \left\{ \lambda_k; k = 0, 1, \cdots, n-1 \right\}$$



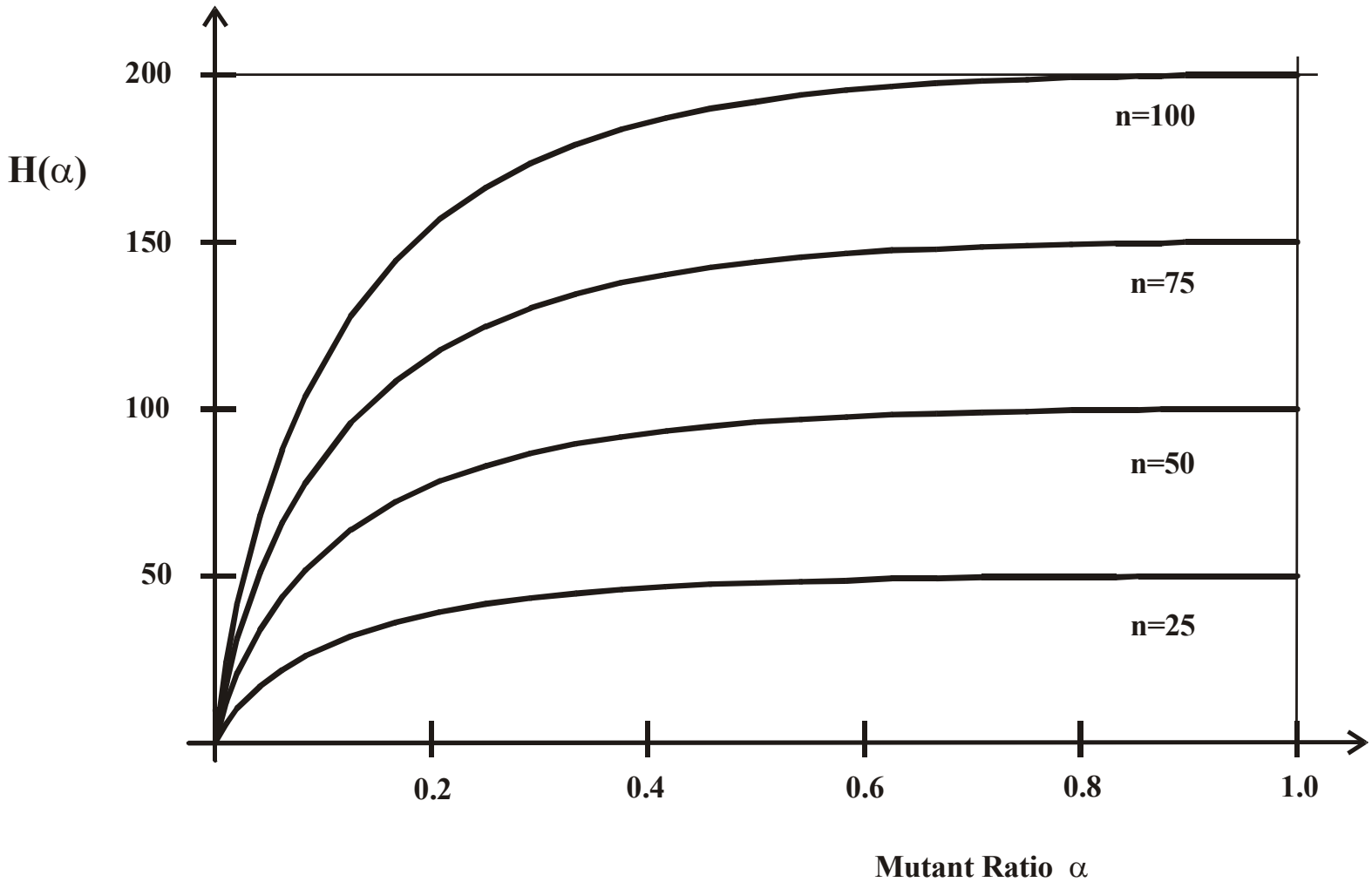
The molecular quasispecies in sequence space



Width of the population in sequence space



Shannon information content of individual sequences: $I_k = -1.4428 \ln p_k$ [bit]



Shannon‘s information entropy of the population: $H = \sum_k p_k I_k = -1.4428 \sum_k p_k \ln p_k$

Shannon information does not provide insight into the process creating information nor does it refer to the context that makes it useful.

Information on the environment is created in the **population** by means of the **selection process** through autocatalytic self-enhancement or replication of advantageous variants.

The **population** is visualized as a distribution of RNA molecules. In the evolutionary optimization process the population carries a **temporary memory** on its recent history in terms of **previously selected variants** that are still present.