Some Mathematical Challenges from Molecular Biology

Part I

Peter Schuster

Institut für Theoretische Chemie und Molekulare Strukturbiologie der Universität Wien

Mathematisches Kolloquium

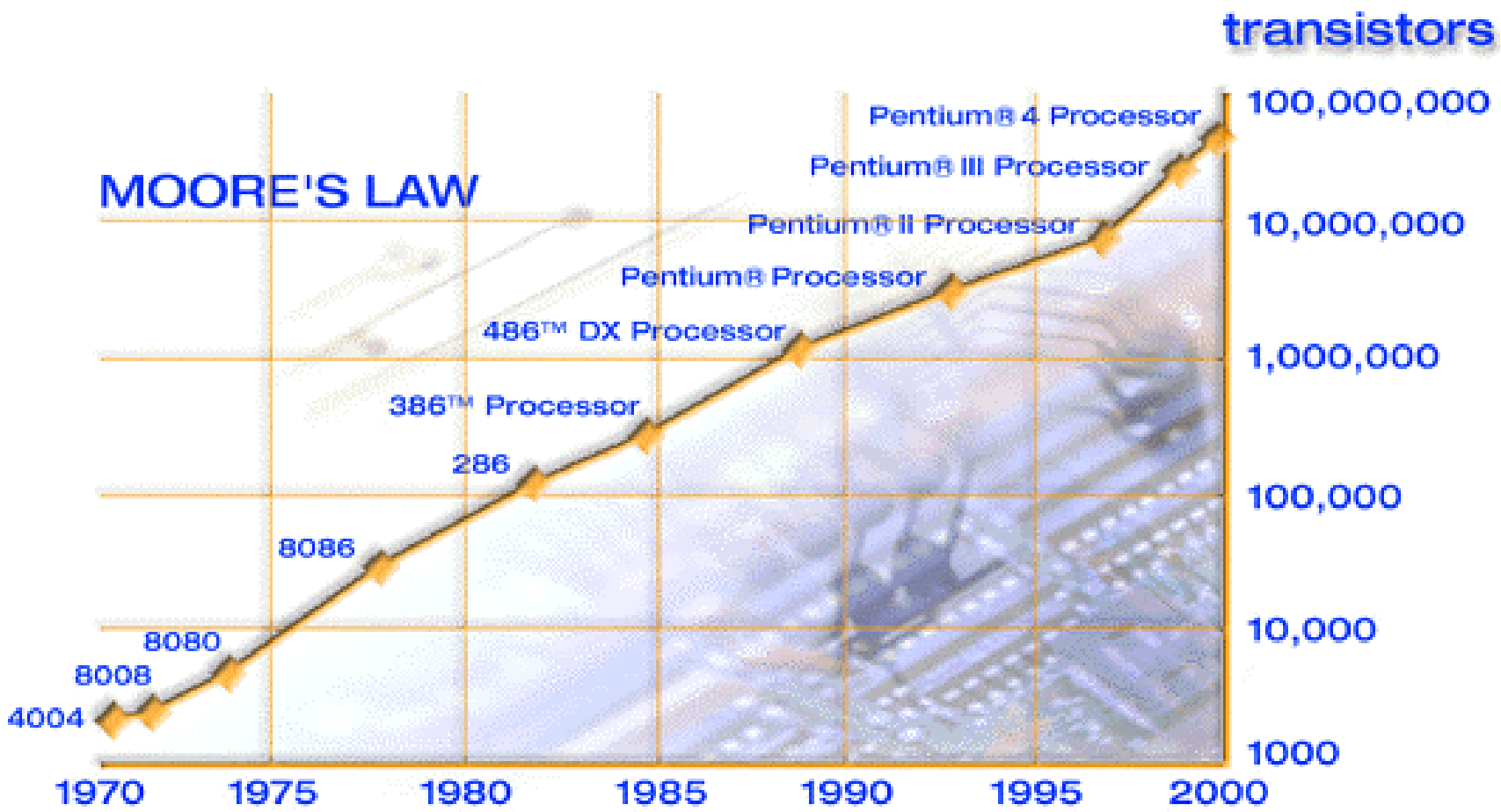
Zürich, 11.11.2003

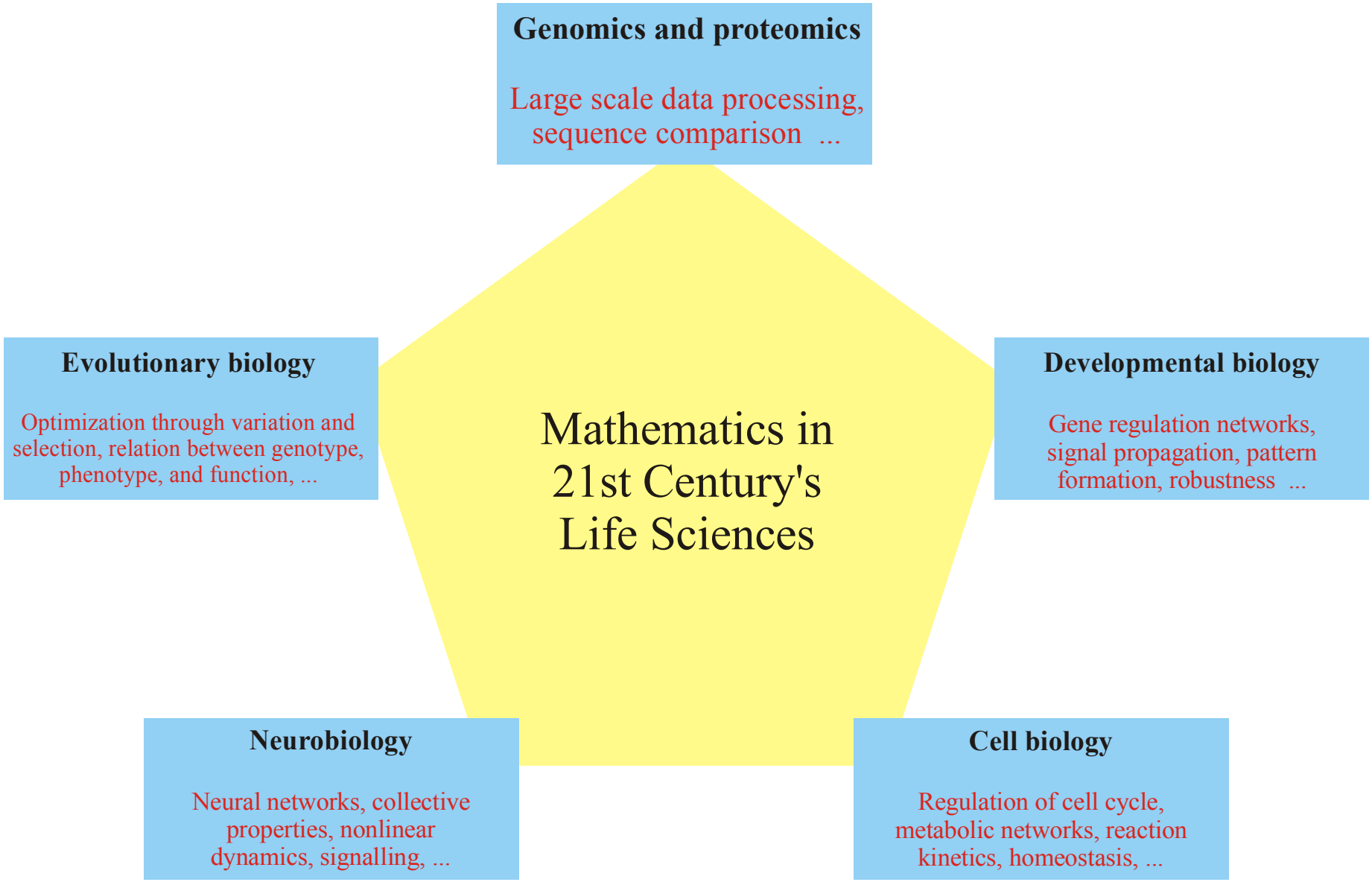
Web-Page for further information:

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

1. **Prolog – Mathematics and the life sciences in the 21st century**
2. **Replication kinetics of RNA molecules and evolution**
3. **RNA evolution *in silico***
4. **Sequence-structure maps, neutral networks, and intersections**
5. **Reference to experimental data**
6. **Summary**

1. **Prolog – Mathematics and the life sciences in the 21st century**
2. **Replication kinetics of RNA molecules and evolution**
3. **RNA evolution *in silico***
4. **Sequence-structure maps, neutral networks, and intersections**
5. **Reference to experimental data**
6. **Summary**

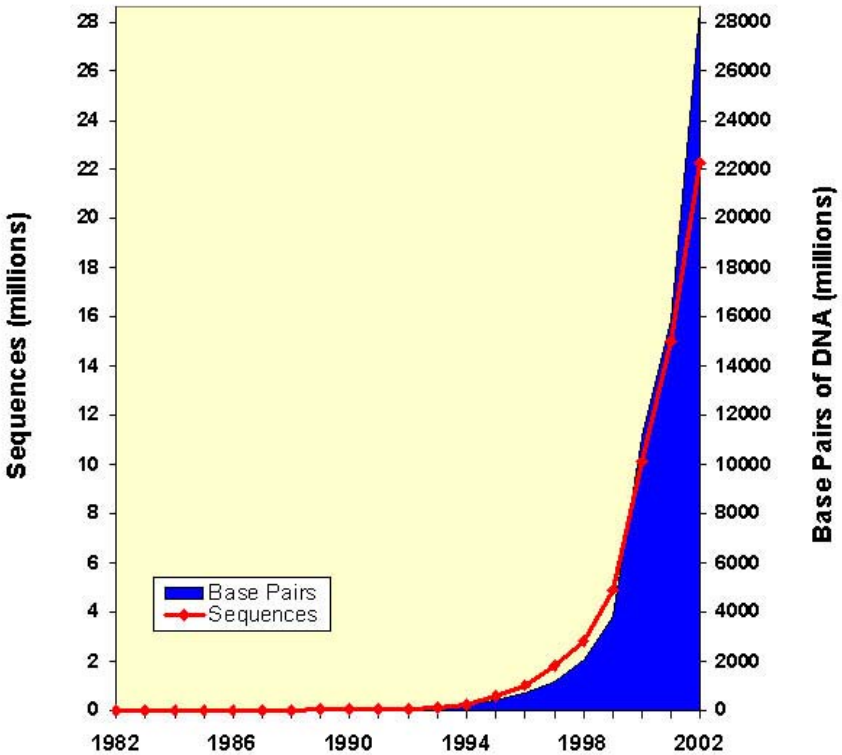

transistors
MOORE'S LAW
Pentium® 4 Processor
Pentium® III Processor
Pentium® II Processor
Pentium® Processor
486™ DX Processor
386™ Processor
286
8086
8080
8008
4004
100,000,000
10,000,000
1,000,000
100,000
10,000
1000
1970
1975
1980
1985
1990
1995
2000


Genomics and proteomics
Large scale data processing, sequence comparison ...
Evolutionary biology
Optimization through variation and selection, relation between genotype, phenotype, and function, ...
Mathematics in 21st Century's Life Sciences
Developmental biology
Gene regulation networks, signal propagation, pattern formation, robustness ...
Neurobiology
Neural networks, collective properties, nonlinear dynamics, signalling, ...
Cell biology
Regulation of cell cycle, metabolic networks, reaction kinetics, homeostasis, ...

Genomics and proteomics

Large scale data processing, sequence comparison ...

E. coli:	Length of the Genome	4×10^6 Nucleotides
	Number of Cell Types	1
	Number of Genes	4 000
Man:	Length of the Genome	3×10^9 Nucleotides
	Number of Cell Types	200
	Number of Genes	30 000 - 100 000



Source: NCBI

Fully sequenced genomes

- *Organisms* 751 projects

153 complete (16 A, 118 B, 19 E)

(*Eukarya* examples: mosquito (pest, malaria), sea squirt, mouse, yeast, homo sapiens, arabidopsis, fly, worm, …)

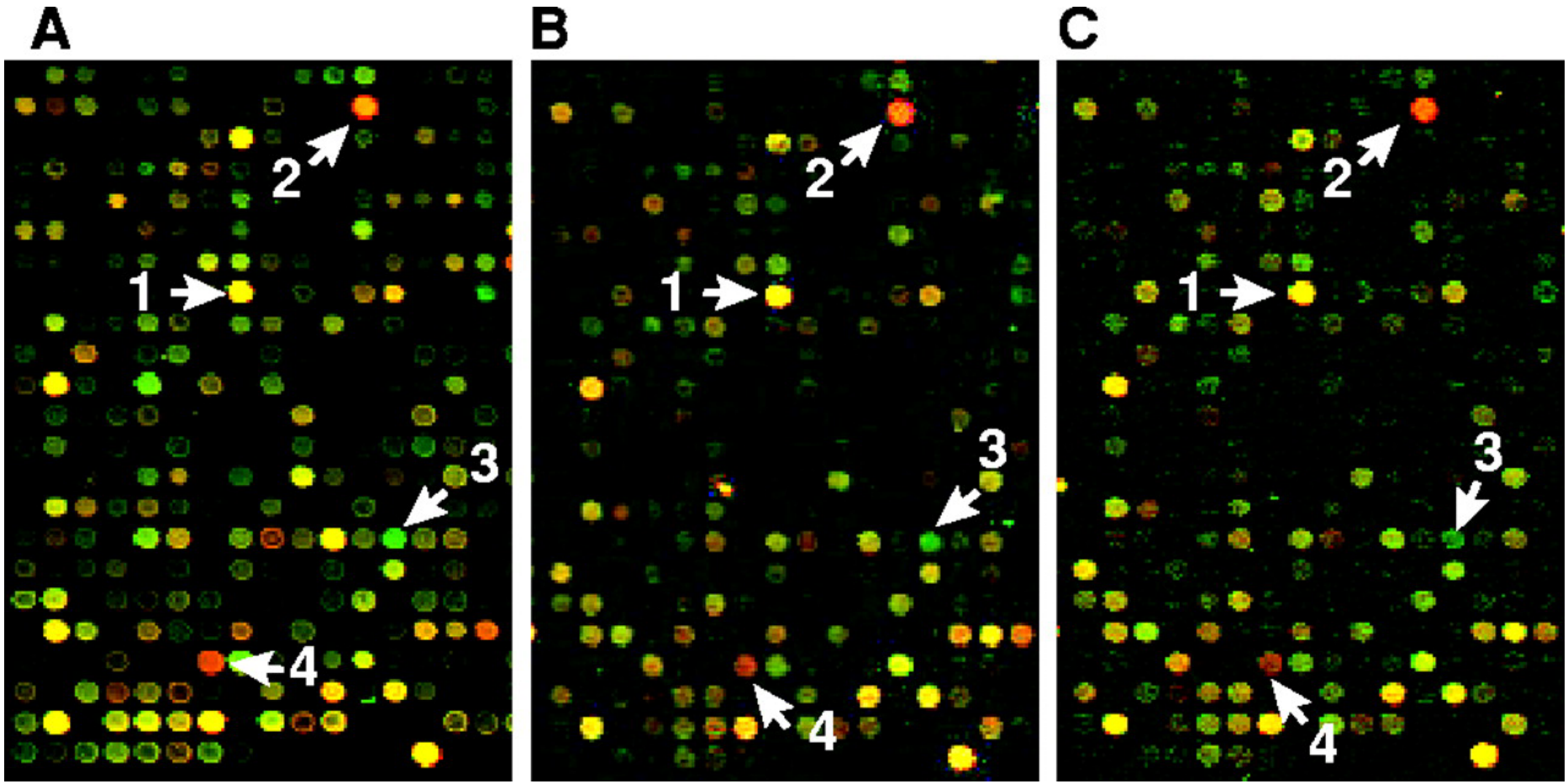
598 ongoing (23 A, 332 B, 243 E)

(*Eukarya* examples: chimpanzee, turkey, chicken, ape, corn, potato, rice, banana, tomato, cotton, coffee, soybean, pig, rat, cat, sheep, horse, kangaroo, dog, cow, bee, salmon, fugu, frog, …)

- *Other structures with genetic information*

68 phages
1328 viruses
35 viroids
472 organelles (423 mitochondria, 32 plastids, 14 plasmids, 3 nucleomorphs)

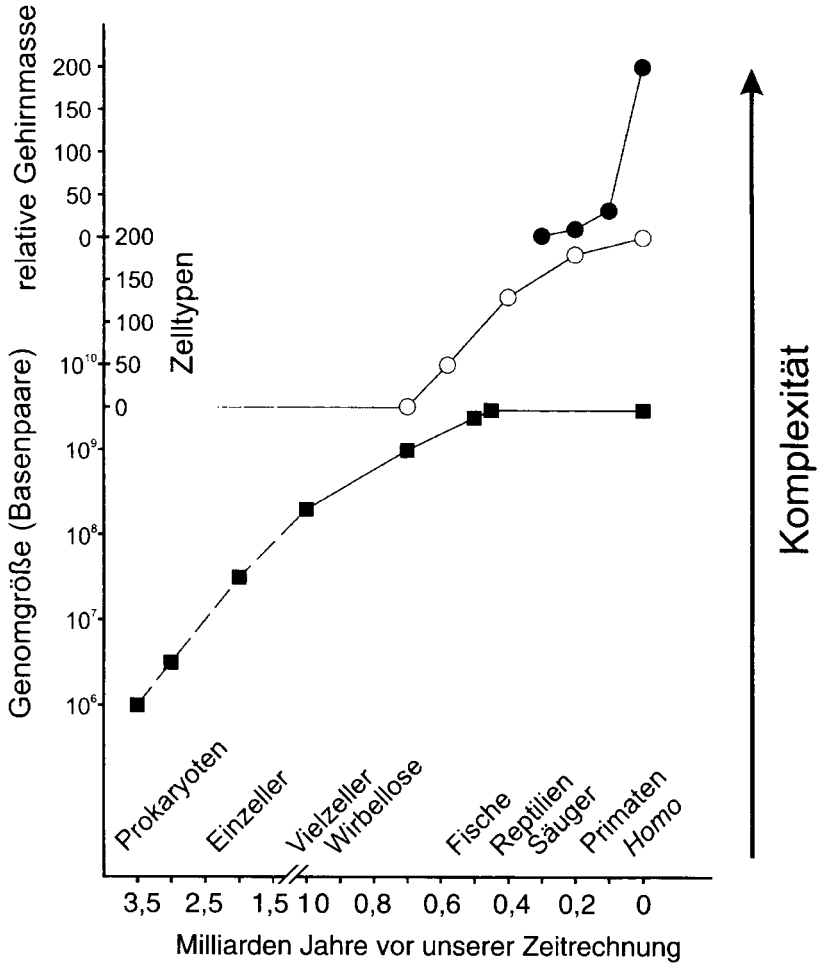
Source: Integrated Genomics, Inc.
August 12th, 2003



The same section of the microarray is shown in three independent hybridizations. Marked spots refer to: (1) protein disulfide isomerase related protein P5, (2) IL-8 precursor, (3) EST AA057170, and (4) vascular endothelial growth factor

Gene expression DNA microarray representing 8613 human genes used to study transcription in the response of human fibroblasts to serum

V.R.Iyer *et al*., *Science* **283**: 83-87, 1999



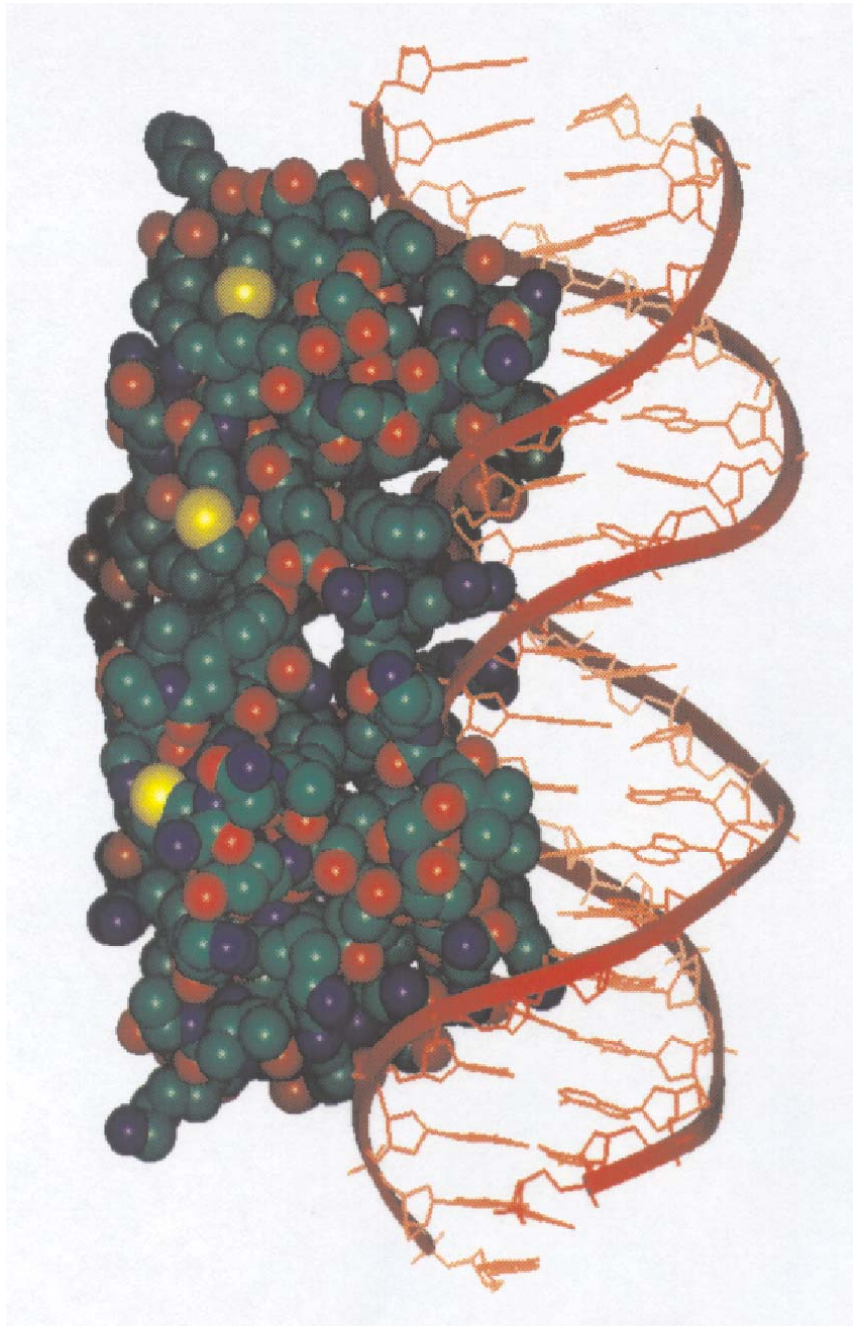
4.10 Die Zunahme der Komplexität ist ein wesentlicher Aspekt der biologischen Evolution, wobei höhere Komplexität sowohl durch Vergrößerung der Zahl von miteinander in Wechselwirkung stehenden Elementen als auch durch Differenzierung der Funktionen dieser Elemente entstehen kann. In dieser Abbildung wird zwischen drei Phasen oder Strategien der Evolution von Komplexität unterschieden. *Untere Kurve*: Zunahme der Genomgröße; logarithmische Auftragung der Zahl der Basenpaare im Genom von Zellen seit Beginn der biologischen Evolution (Daten aus Abbildung 2.3). *Mittlere Kurve*: Zunahme der Zahl der Zelltypen in der Evolution der Metazoa (Daten aus Abbildung 4.8). *Obere Kurve*: Zunahme des relativen Gehirngewichts (bezogen auf die Körperoberfläche) bei Säugetieren (Daten aus Wilson 1985). Für die Abszisse wurden zwei Skaleneinteilungen verwendet, eine für den Zeitraum $>10^9$ Jahre, eine andere für den Zeitraum $<10^9$ Jahre vor der Gegenwart. Oberhalb der Abszisse sind die Namen einiger wichtiger taxonomischer Einheiten angeführt, deren Evolution in etwa beim jeweiligen Wortbeginn einsetzt.

Wolfgang Wieser. Die Erfindung der Individualität oder die zwei Gesichter der Evolution. Spektrum Akademischer Verlag, Heidelberg 1998.

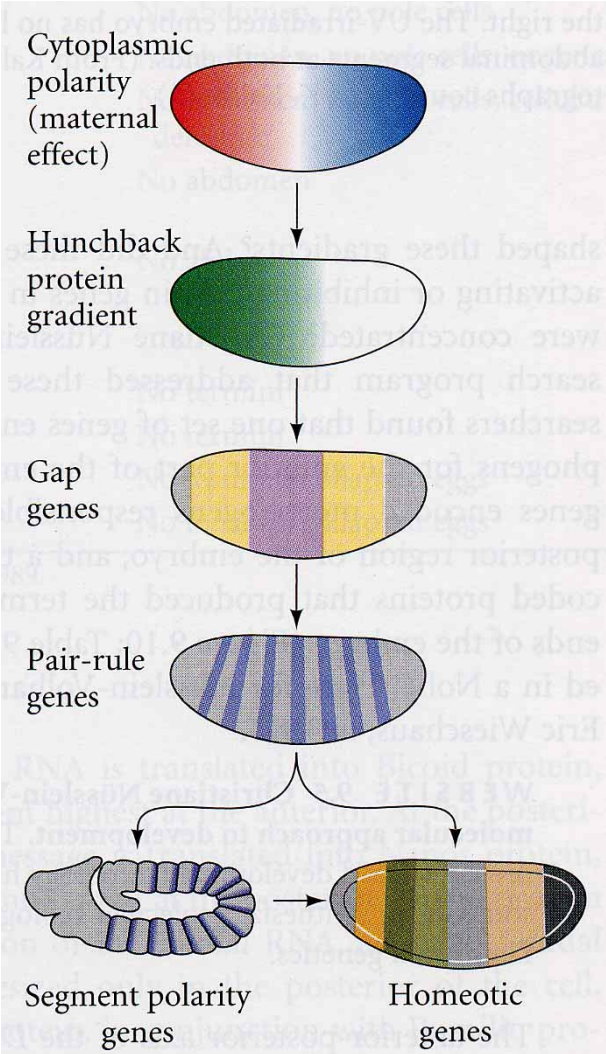
A.C.Wilson. The Molecular Basis of Evolution. Scientific American, Oct.1985, 164-173.

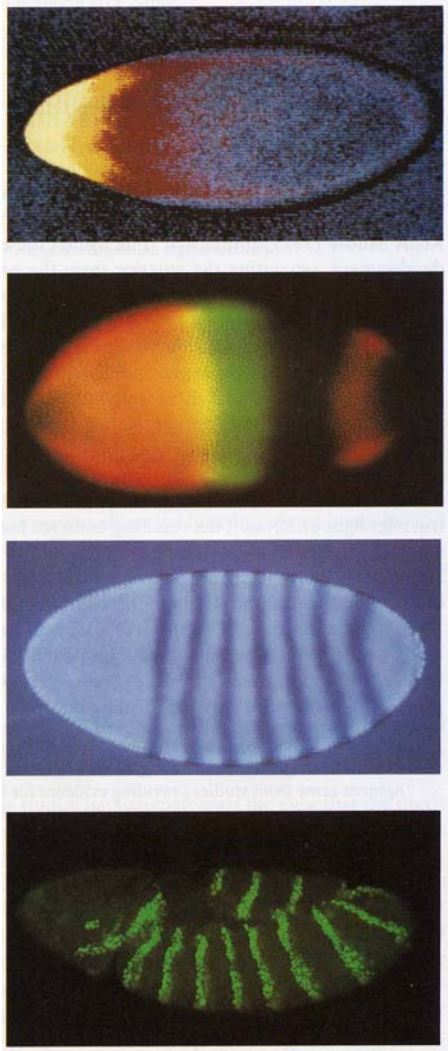
Developmental biology

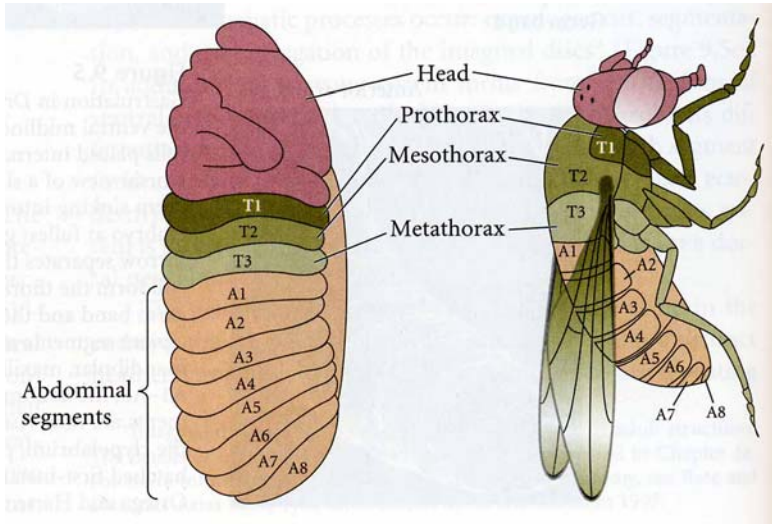
Gene regulation networks, signal propagation, pattern formation, robustness ...



Three-dimensional structure of the complex between the regulatory protein **cro-repressor** and the binding site on λ-phage **B-DNA**





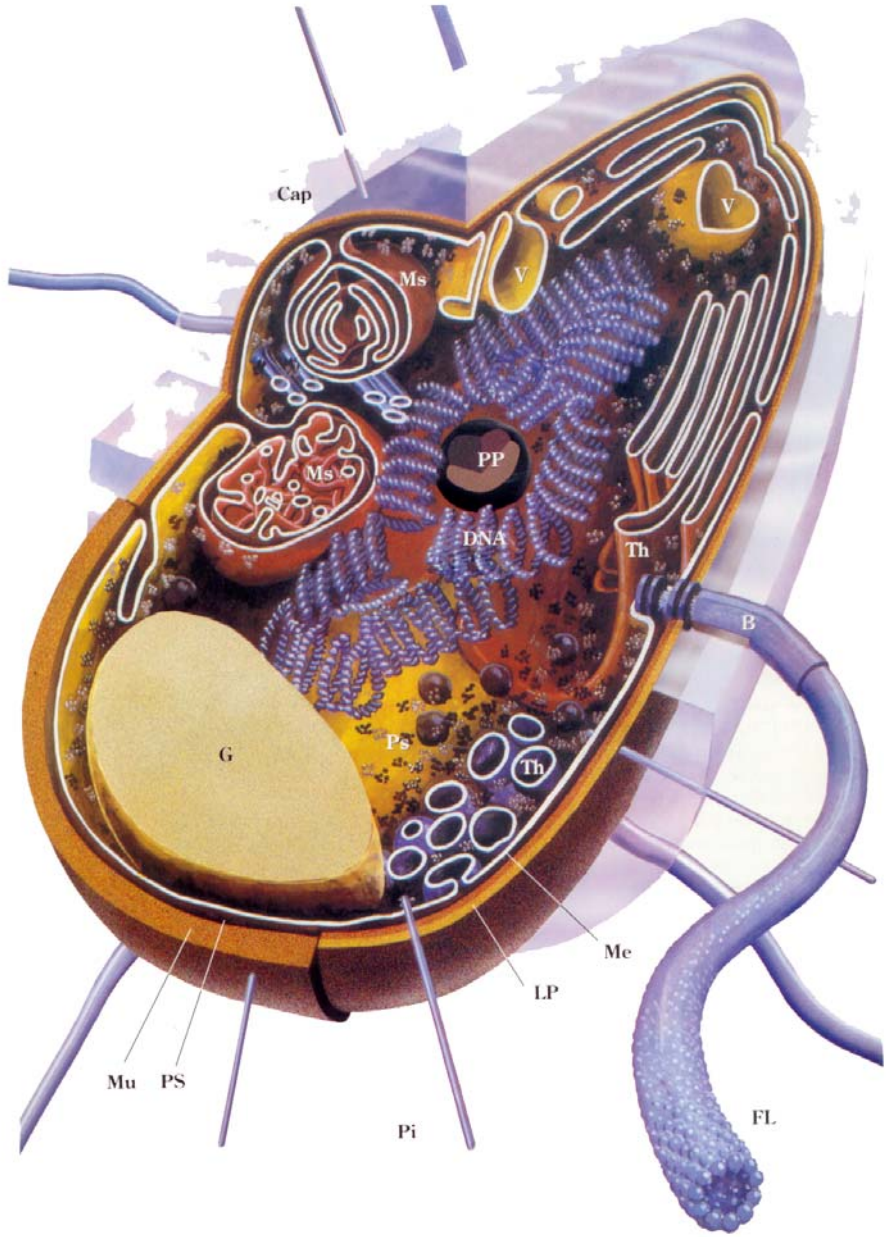


Development of the fruit fly *drosophila melanogaster*: Genetics, experiment, and imago

Cell biology

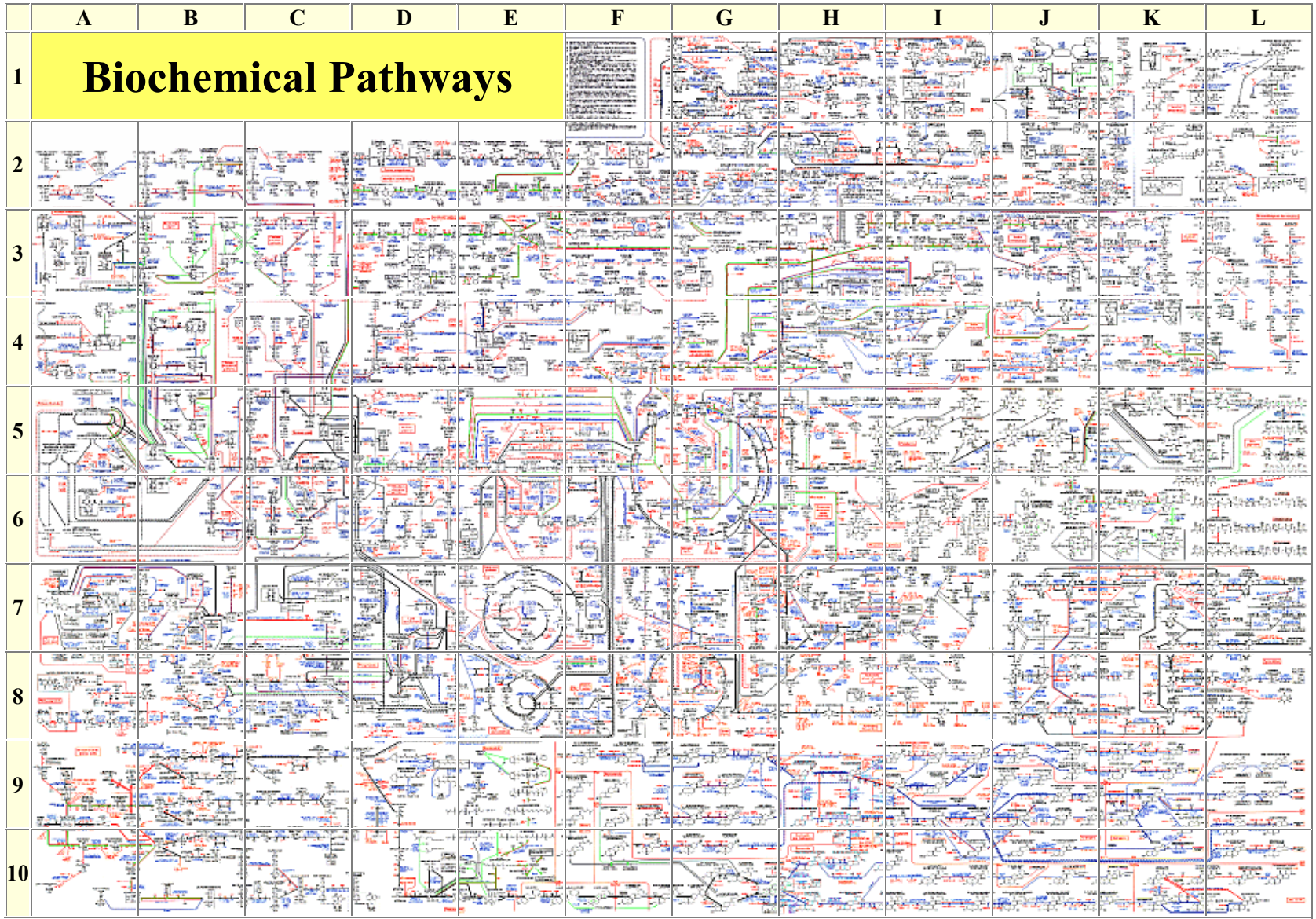
Regulation of cell cycle, metabolic networks, reaction kinetics, homeostasis, ...

The bacterial cell as an example for the simplest form of autonomous life



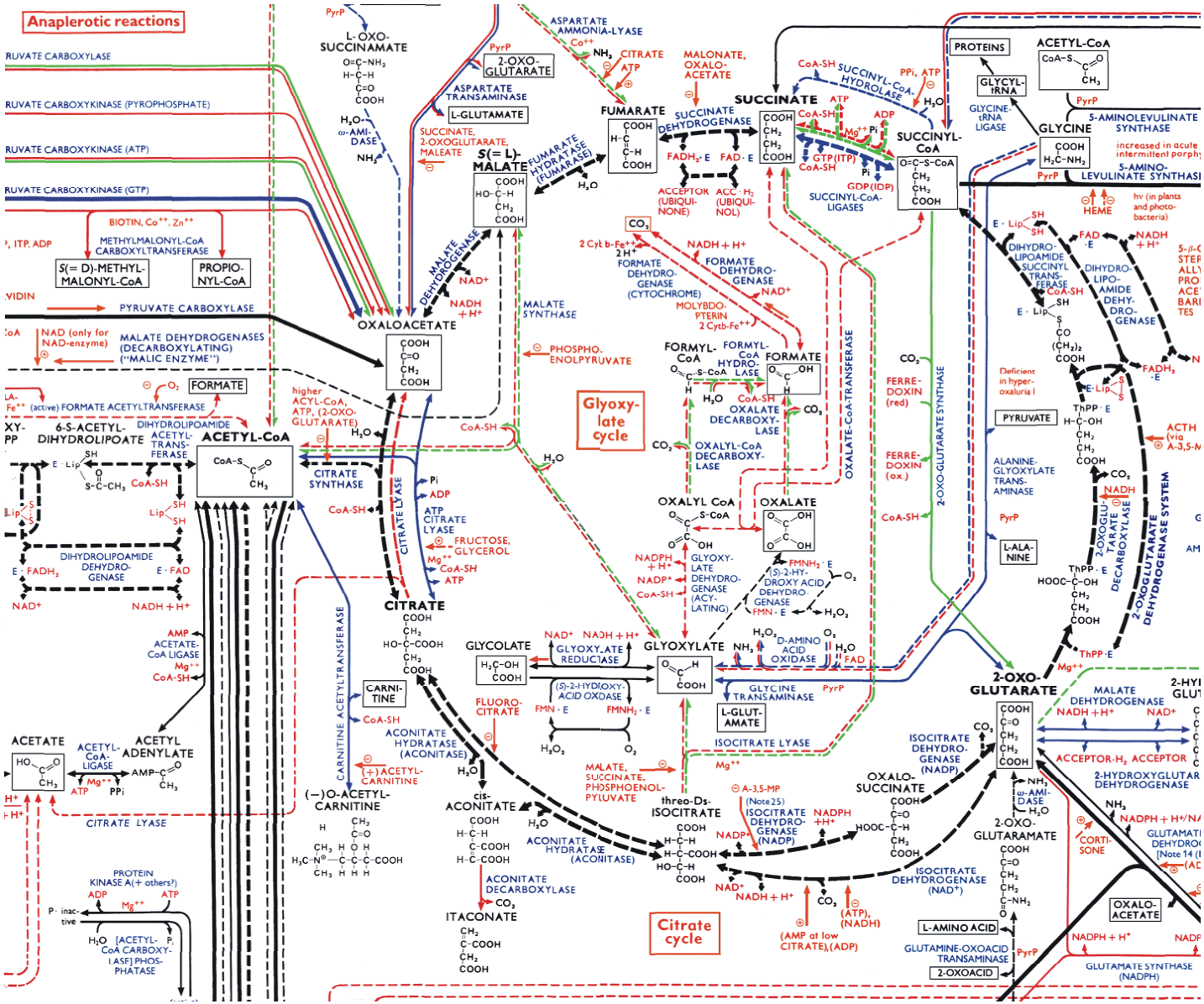
The human body:

10^{14} cells, 10^{13} eukaryotic cells and
Î 9Ç10^{13} bacterial (prokaryotic) cells,
and Î 200 eukaryotic cell types



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

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



Kinetic differential equations

$$\frac{d\,x_i}{d\,t} = f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

Reaction diffusion equations

$$\frac{\partial x_i}{\partial t} = D_i \nabla^2 x_i + f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

Parameter set

$$k_j(T, p, pH, I, \ldots; x_1, x_2, \ldots, x_n);\ j = 1, 2, \ldots, m$$

General conditions: T , p , pH , I , ...

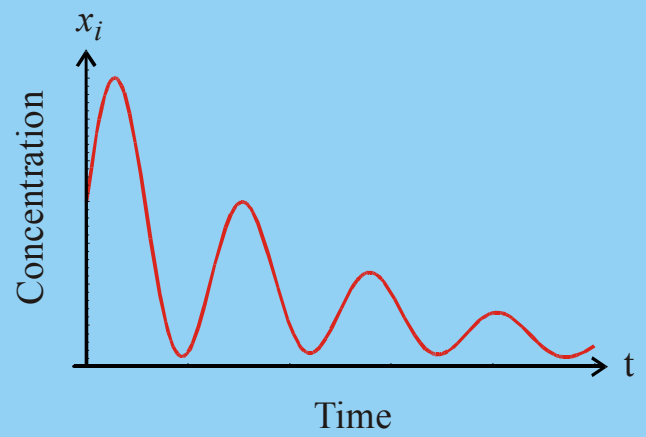
Initial conditions: $x_i(0);\ i = 1, 2, \ldots, n$

Boundary conditions: boundary ... $\vec{s}$
normal unit vector ... $\vec{u}$

Dirichlet , $x_i^{\vec{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

Neumann , $\frac{\partial x_i}{\partial u} = \hat{u} \cdot \nabla x_i^{\vec{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

Solution curves: $x_i(t);\ i = 1, 2, \ldots, n$



The forward-problem of chemical reaction kinetics

Kinetic differential equations

$$\frac{dx_i}{dt} = f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

Reaction diffusion equations

$$\frac{\partial x_i}{\partial t} = D_i \nabla^2 x_i + f(x_1, x_2, \ldots, x_n; k_1, k_2, \ldots, k_m);\ i = 1, 2, \ldots, n$$

General conditions: T, p, pH, I, ...

Initial conditions: $x_i(0);\ i = 1, 2, \ldots, n$

Boundary conditions: boundary ... $\hat{s}$
normal unit vector ... $\hat{u}$

Dirichlet, $x_i^{\hat{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

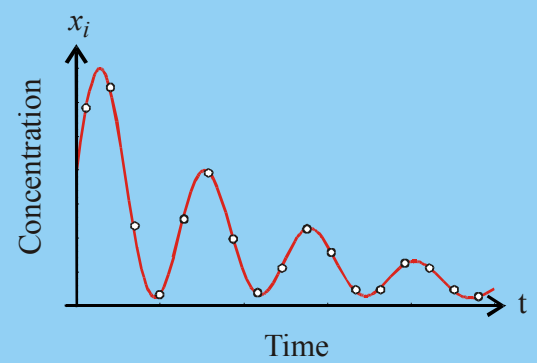
Neumann, $\frac{\partial x_i}{\partial u} = \hat{u} \cdot \nabla x_i^{\hat{s}} = f(\vec{r}, t);\ i = 1, 2, \ldots, n$

Parameter set

$$k_j(T, p, pH, I, \ldots; x_1, x_2, \ldots, x_n);\ j = 1, 2, \ldots, m$$

Data from measurements

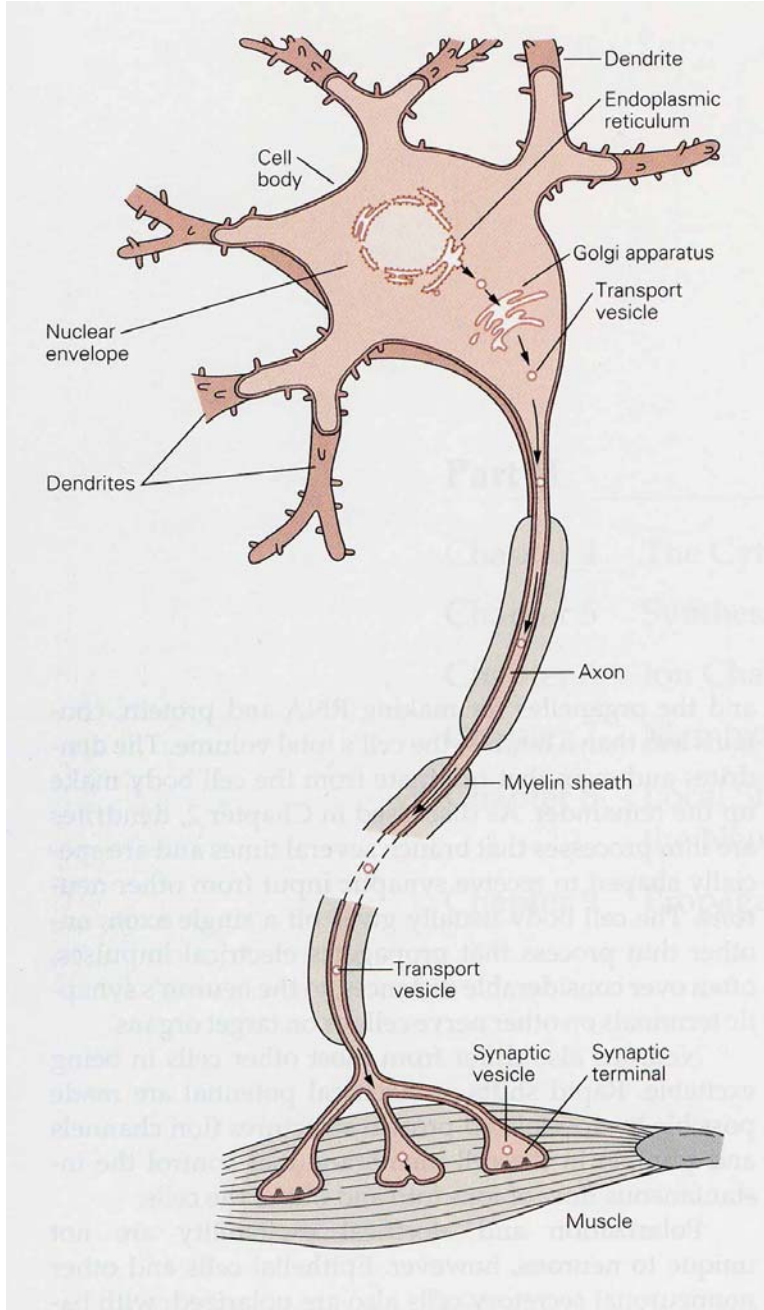
$$x_i(t_k);\ i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, N$$



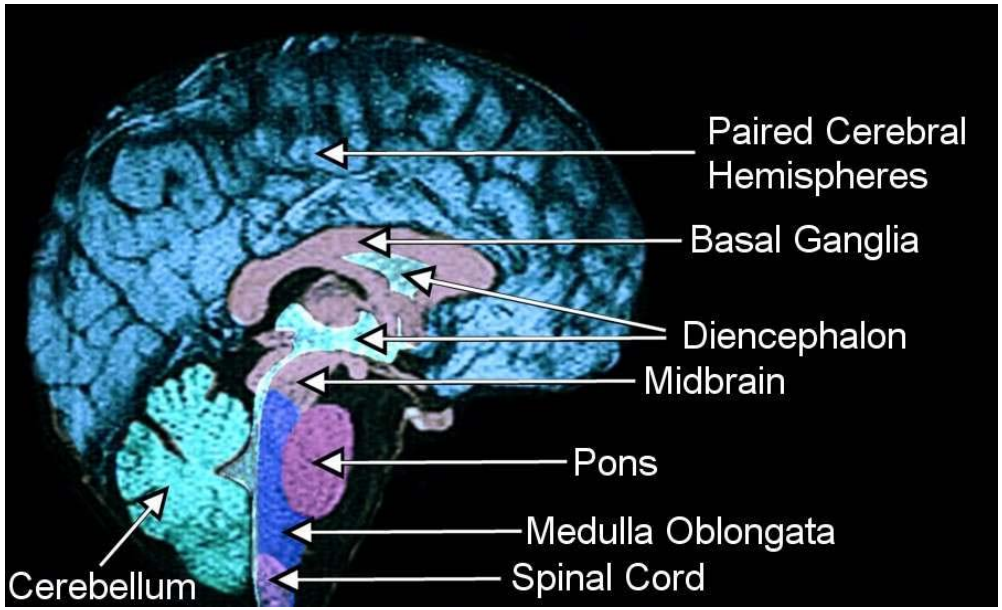
The inverse-problem of chemical reaction kinetics

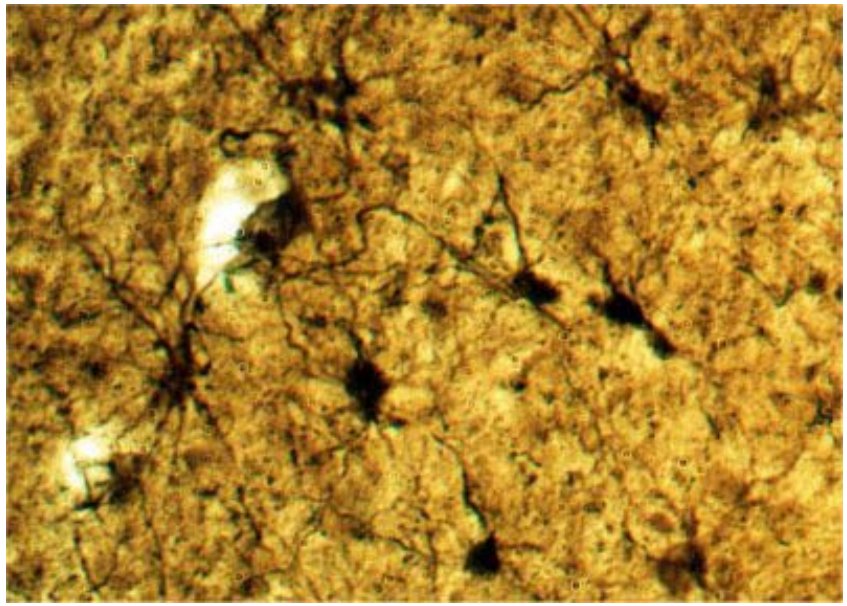
Neurobiology

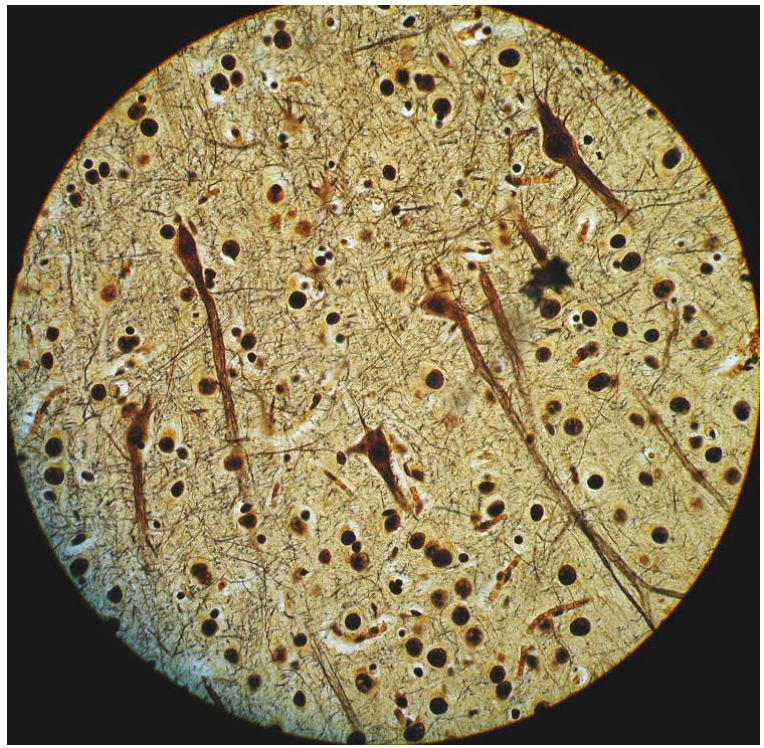
Neural networks, collective properties, nonlinear dynamics, signalling, ...



A single neuron signaling to a muscle fiber







The human brain

10^{11} neurons connected by Î 10^{13} to 10^{14} synapses

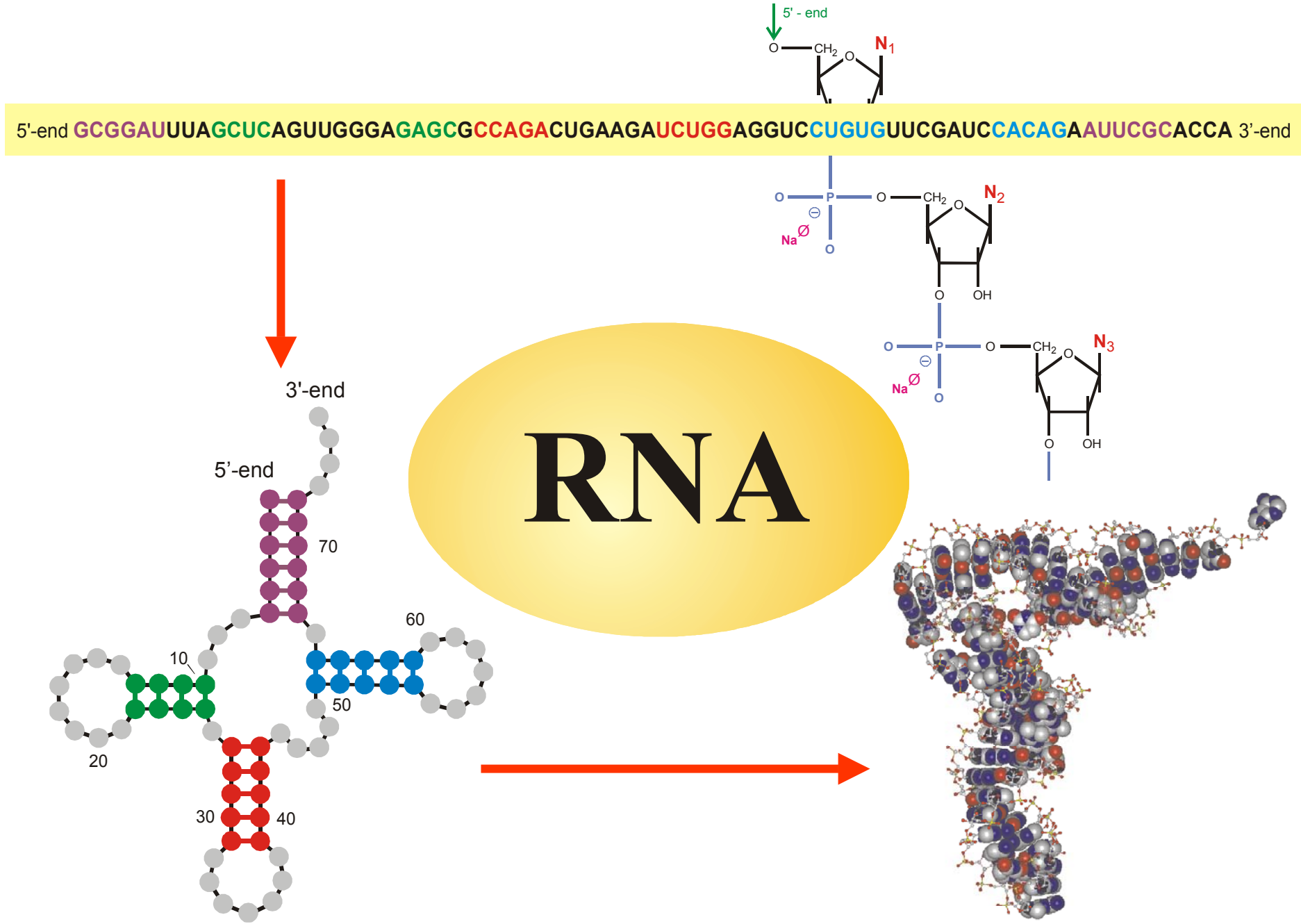
Evolutionary biology

Optimization through variation and selection, relation between genotype, phenotype, and function, ...

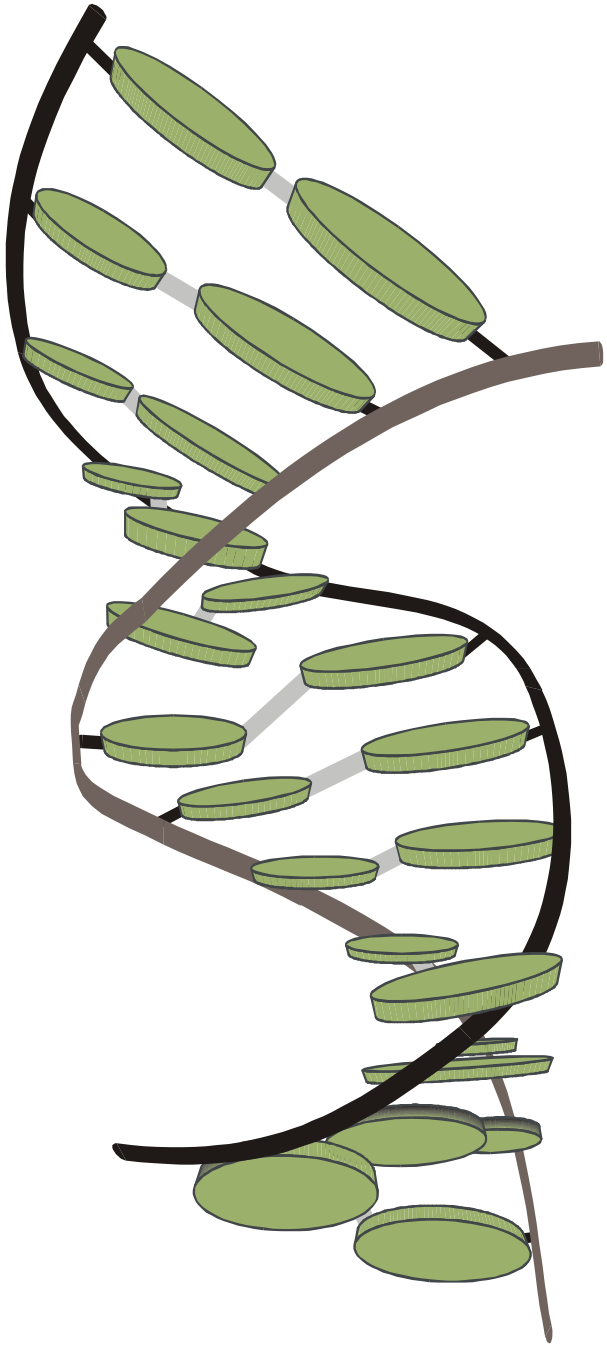
	Generation time	10 000 generations	10^6 generations	10^7 generations
RNA molecules	10 sec 1 min	27.8 h = 1.16 d 6.94 d	115.7 d 1.90 a	3.17 a 19.01 a
Bacteria	20 min 10 h	138.9 d 11.40 a	38.03 a 1 140 a	380 a 11 408 a
Higher multicelluar organisms	10 d 20 a	274 a 20 000 a	27 380 a 2×10^7 a	273 800 a 2×10^8 a

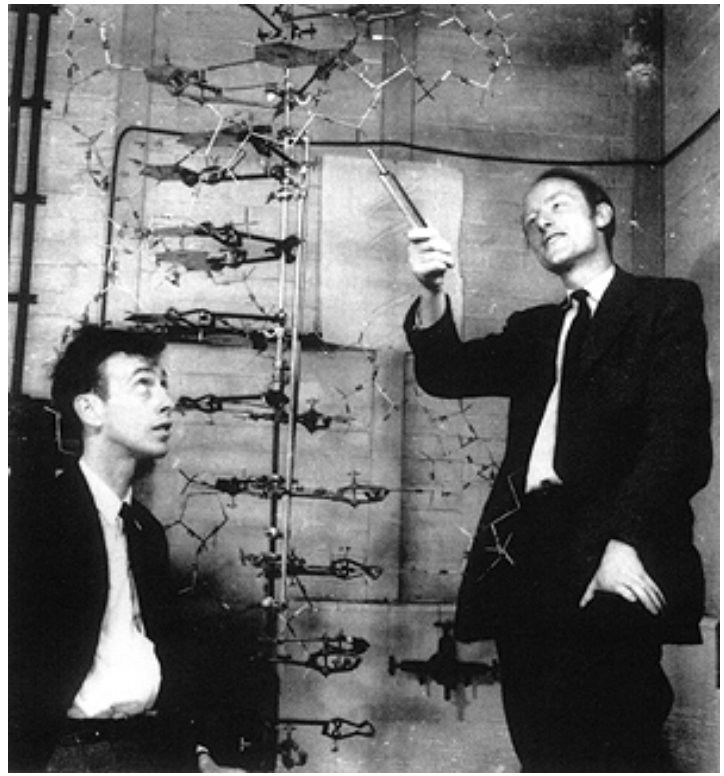
Time scales of evolutionary change

1. Prolog – Mathematics and the life sciences in the 21st century
2. **Replication kinetics of RNA molecules and evolution**
3. RNA evolution *in silico*
4. Sequence-structure maps, neutral networks, and intersections
5. Reference to experimental data
6. Summary



Definition of RNA structure





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

1953 – 2003 fifty years double helix

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

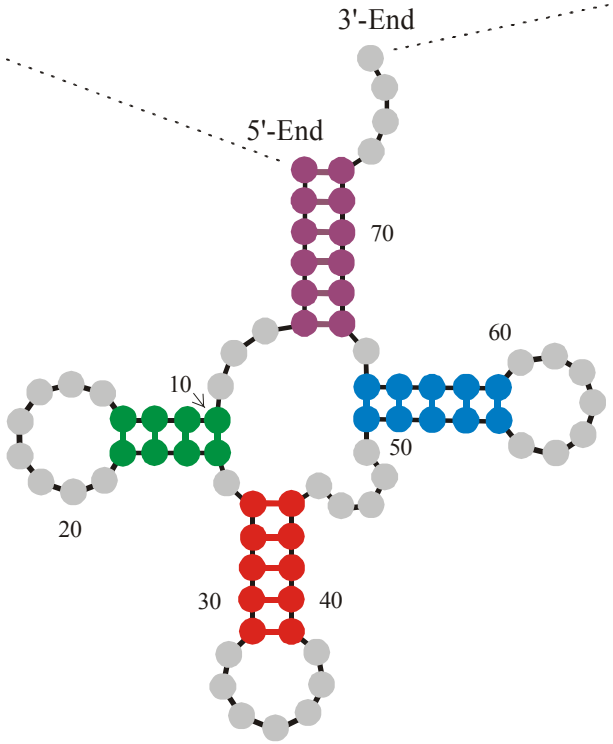
Sequence

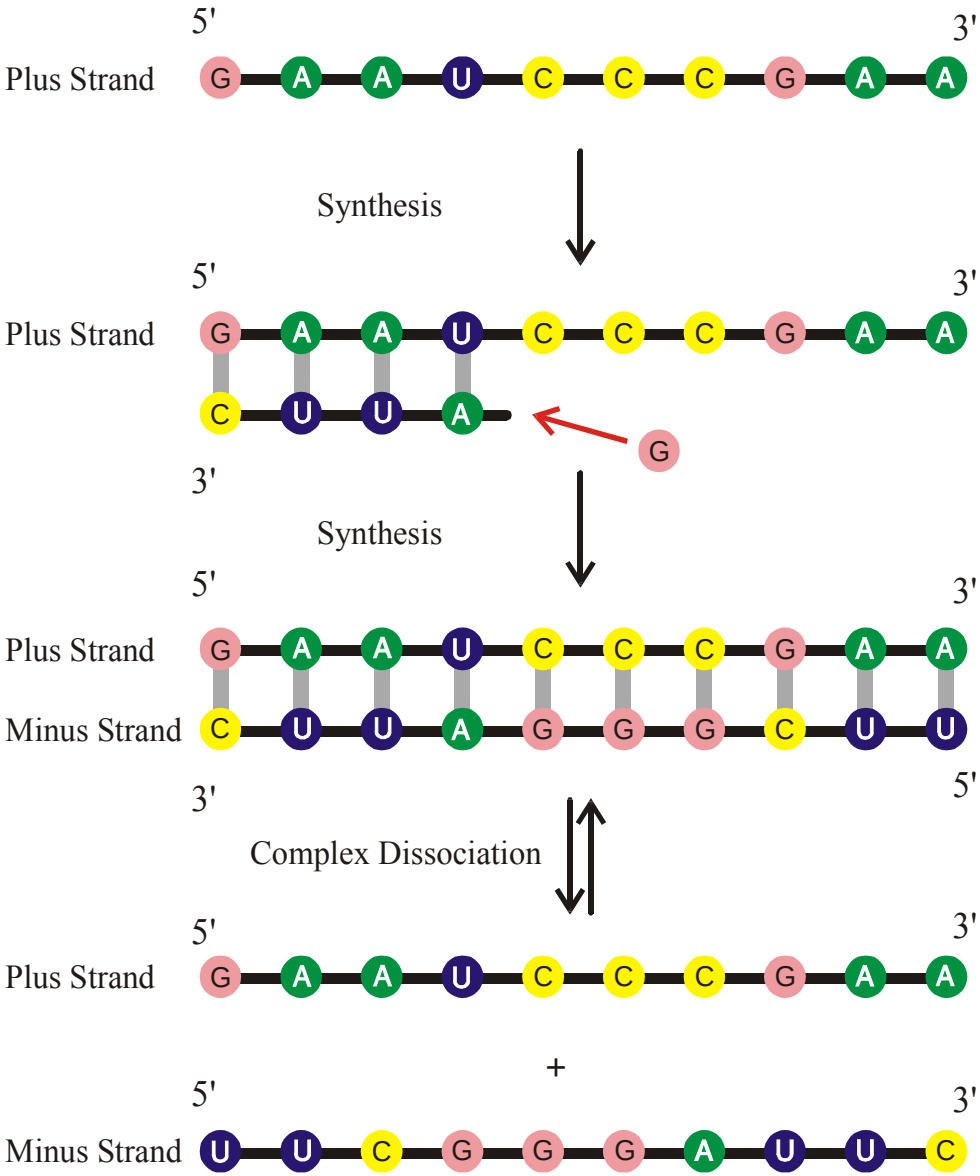
5'-End

3'-End

GCGGAUUUAGCUCAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCCUGUGTPCGAUCCACAGAAUUCGCACCA

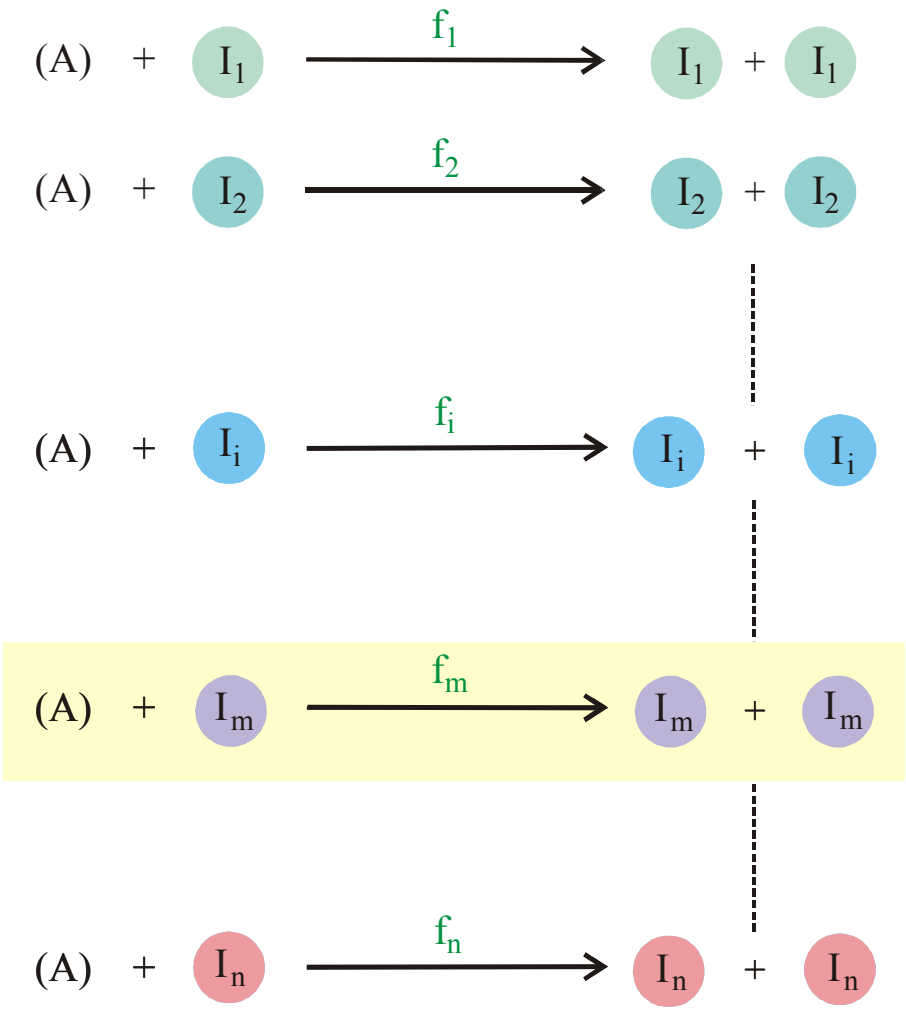
Secondary structure





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

G Í C and A=U



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

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

$[I_i] = x_i$ Æ 0 ; i =1,2,...,n ;

[A] = a = constant

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

$x_m(t)$ Á 1 for t Á ¸

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

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

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

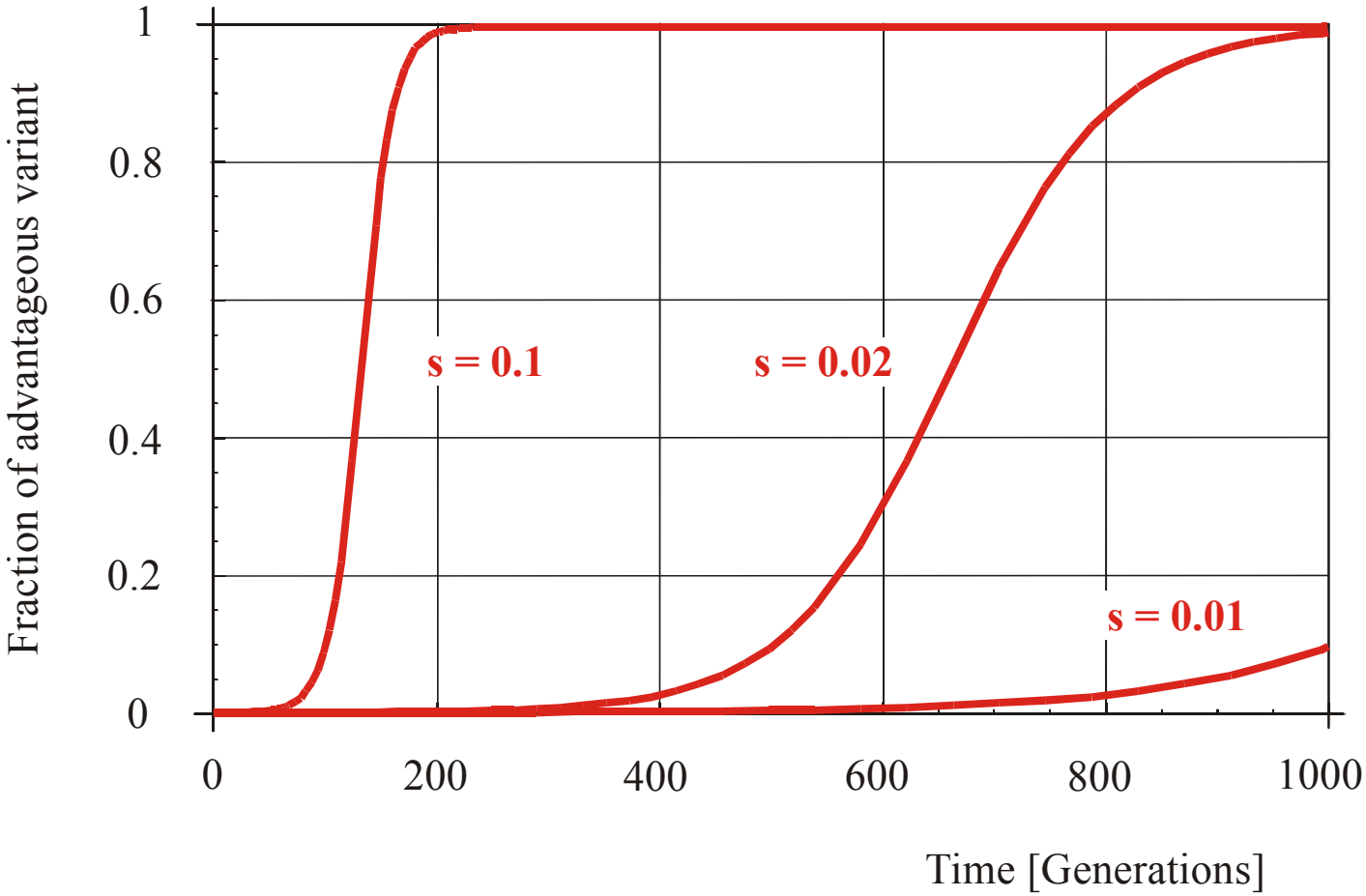
Mean fitness or dilution flux, ϕ(t), is a **non-decreasing function** of time,

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

Solutions are obtained by integrating factor transformation

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

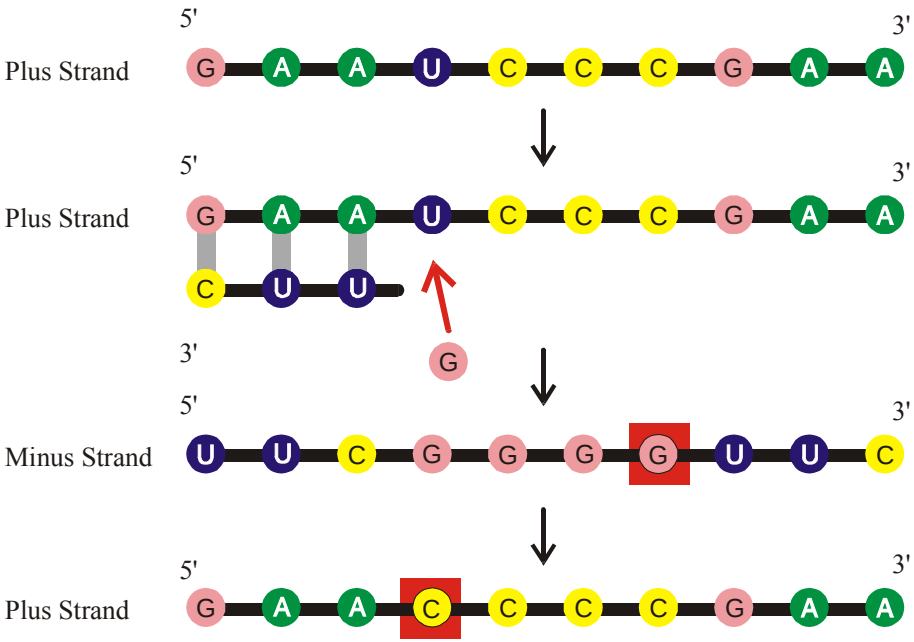
$$\mathbf{s} = (f_2 - f_1) / f_1;\ f_2 > f_1\ ;\ x_1(0) = 1 - 1/N\ ;\ x_2(0) = 1/N$$



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 → GAAUCCGA

Deletion

Mutations in nucleic acids represent the mechanism of **variation** of **genotypes**.

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

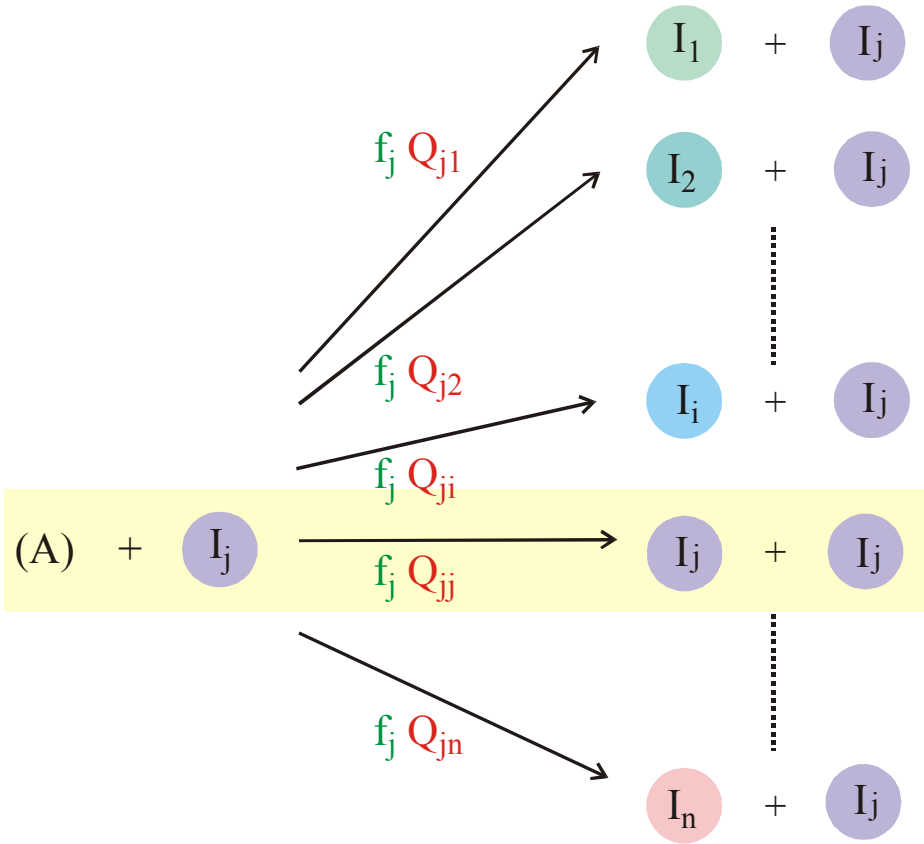
M.Eigen, P.Schuster, ***The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle***. Naturwissenschaften **65** (1978), 341-369

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



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

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

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

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

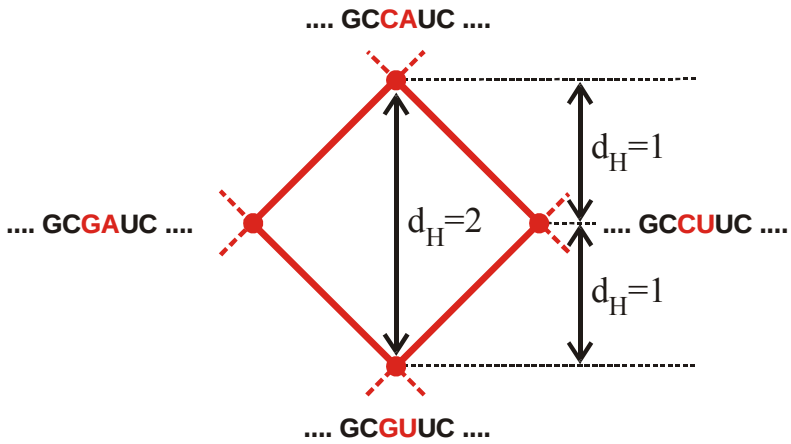
$$Q_{ij} = (1-p)^{\ell - d(i,j)} p^{d(i,j)}$$

p ..... Error rate per digit

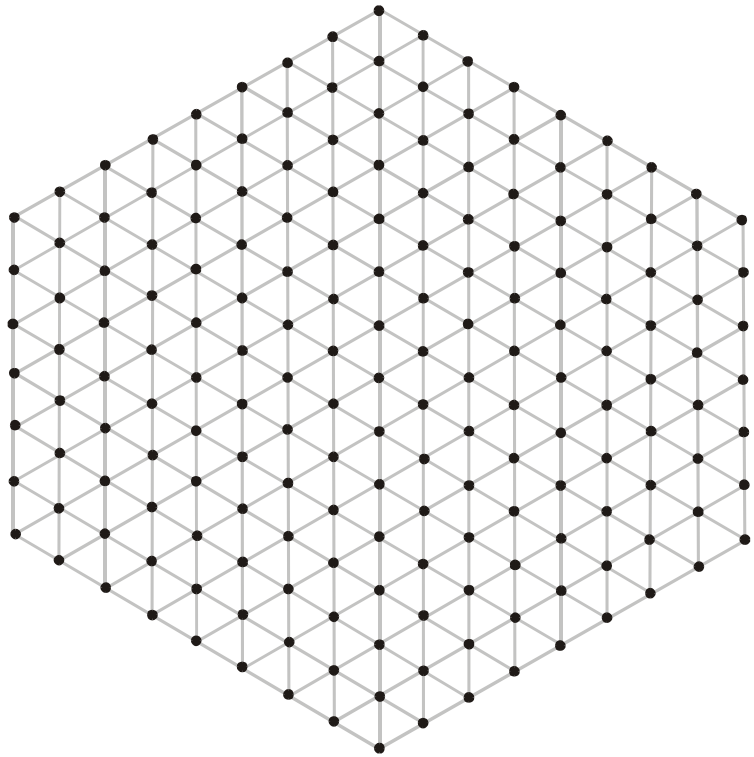
ℓ ..... Chain length of the polynucleotide

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

Chemical kinetics of replication and mutation as parallel reactions

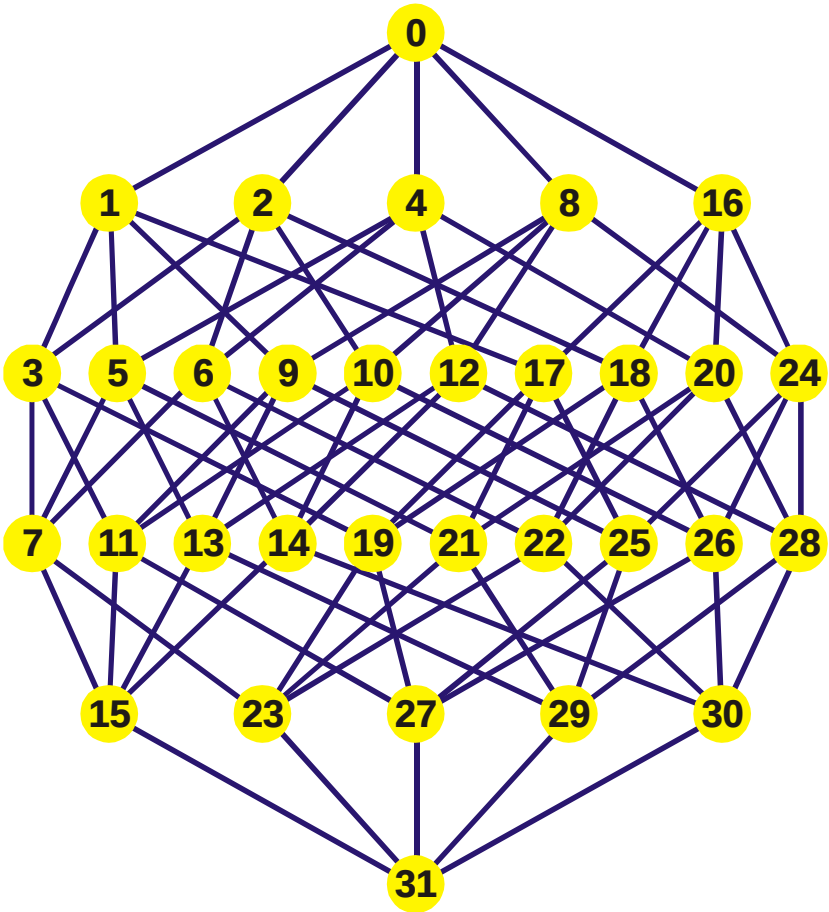


City-block distance in sequence space



2D Sketch of sequence space

Single point mutations as moves in sequence space



Mutant class

0

1

2

3

4

5

Binary sequences are encoded by their decimal equivalents:

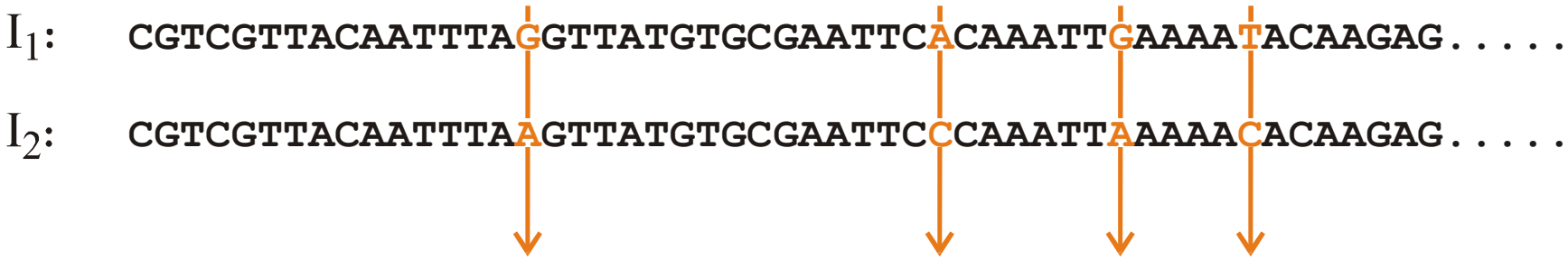
C = 0 and G = 1, for example,

"0" ≡ 00000 = CCCCC,

"14" ≡ 01110 = CGGGC,

"29" ≡ 11101 = GGGCG, etc.

Sequence space of binary sequences of chain lenght n=5



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

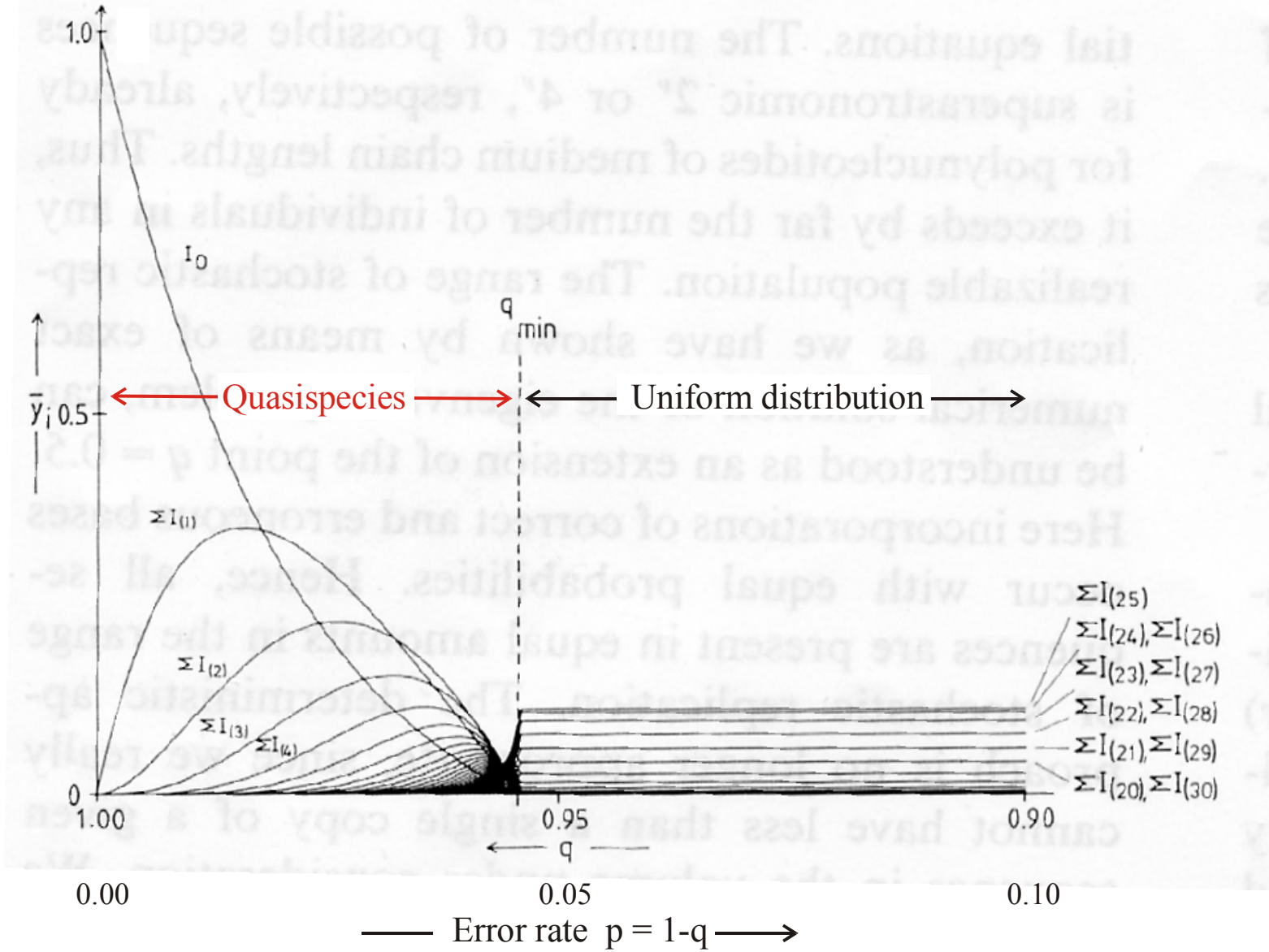
$$\frac{dx_i}{dt} = \sum_{j=1}^{n} f_j Q_{ji}\, x_j - x_i\, \phi, \quad i = 1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \overline{f}$$

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

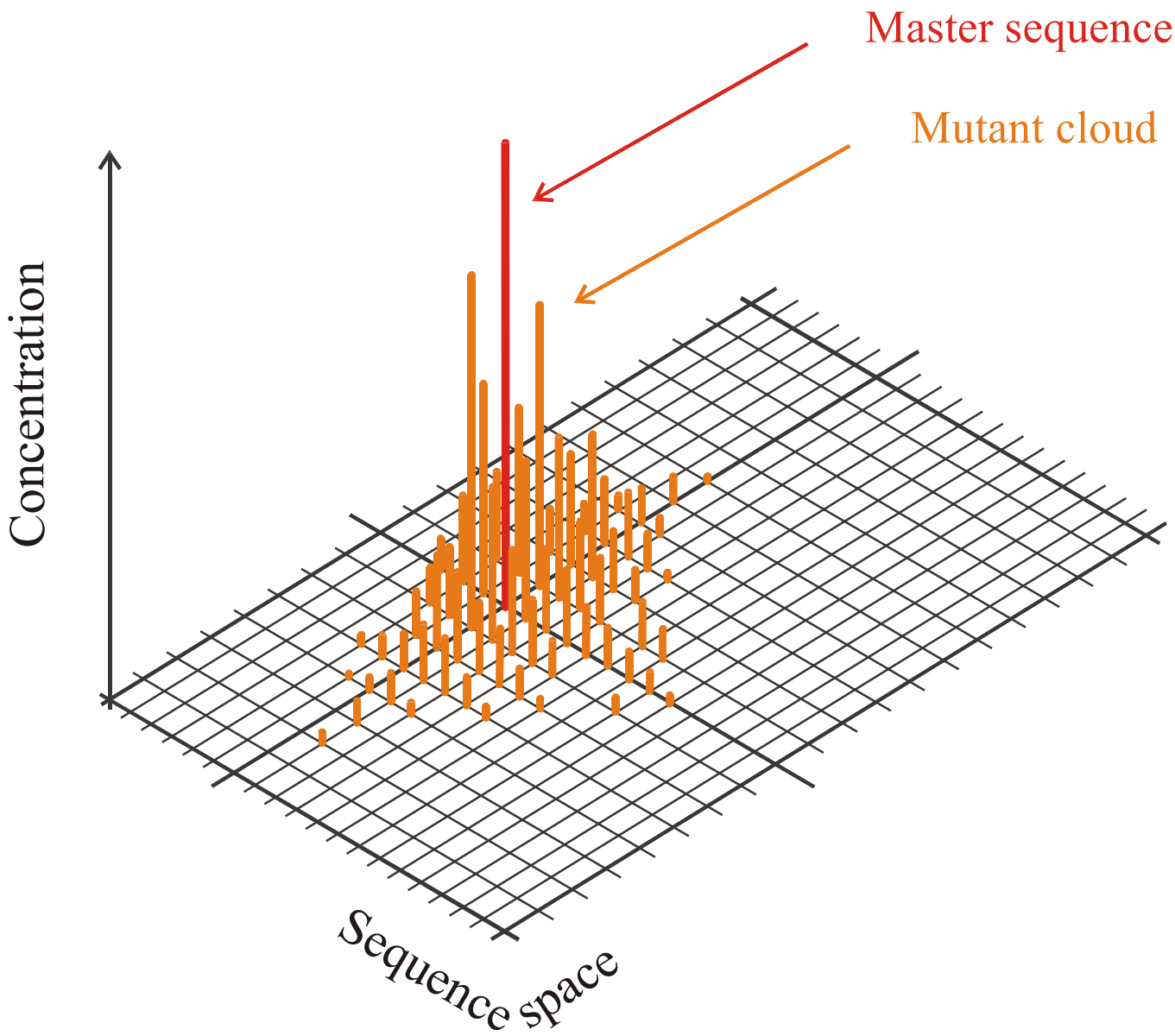
$$x_i(t) = \frac{\sum_{k=0}^{n-1} \ell_{ik} \cdot c_k(0) \cdot \exp(\lambda_k t)}{\sum_{j=1}^{n} \sum_{k=0}^{n-1} \ell_{jk} \cdot c_k(0) \cdot \exp(\lambda_k t)}; \quad i = 1,2,\cdots,n; \quad c_k(0) = \sum_{i=1}^{n} h_{ki}\, x_i(0)$$

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

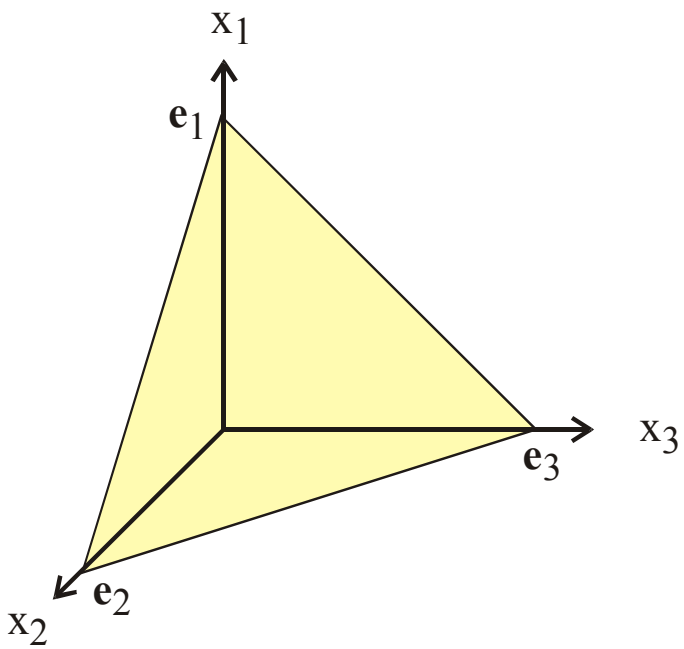
$$L^{-1} \cdot W \cdot L \quad = \quad \Lambda \quad = \quad \left\{ \lambda_k; k = 0, 1, \cdots, n-1 \right\}$$

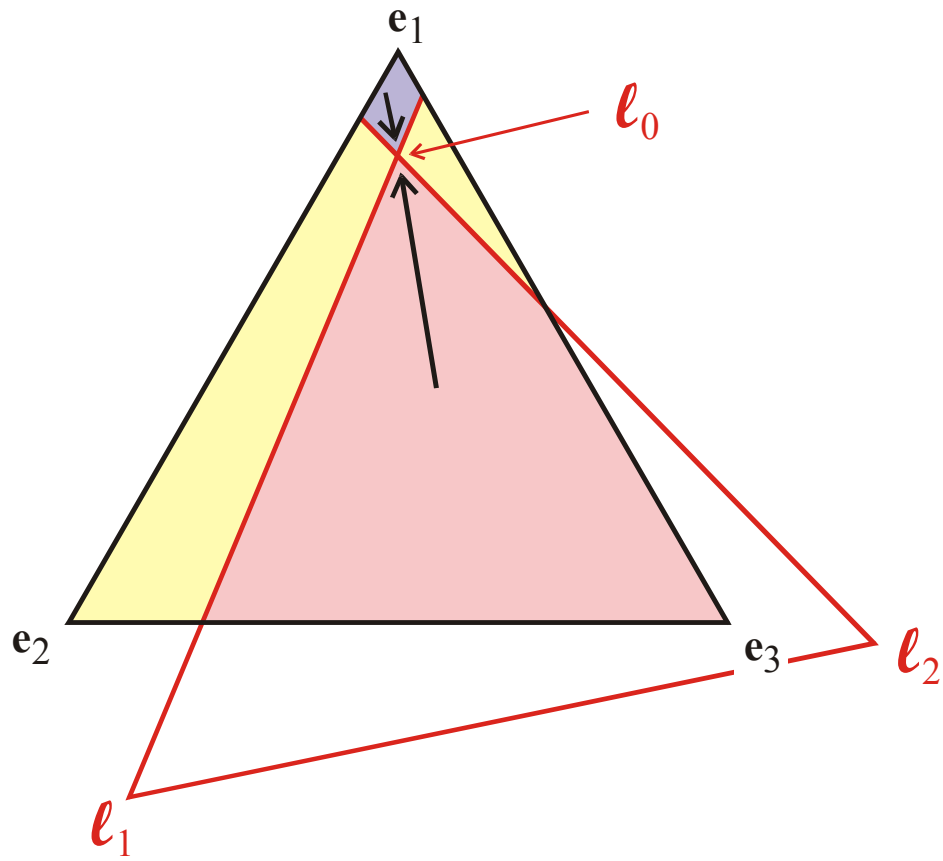


Quasispecies as a function of the replication accuracy q



The molecular quasispecies in sequence space





The quasispecies on the concentration simplex $\mathbb{S}_3 = \left\{ x_i \geq 0,\ i = 1,2,3;\ \sum_{i=1}^{3} x_i = 1 \right\}$

In the case of non-zero mutation rates ($p>0$ or $q<1$) the **Darwinian principle** of optimization of mean fitness can be understood only as an **optimization heuristic**. It is valid only on part of the concentration simplex. There are other well defined areas where the mean fitness decreases monotonously or where it may show non-monotonous behavior. The volume of the part of the simplex where mean fitness is non-decreasing in the conventional sense decreases with inreasing mutation rate p.

1. Prolog – Mathematics and the life sciences in the 21st century
2. Replication kinetics of RNA molecules and evolution
3. **RNA evolution *in silico***
4. Sequence-structure maps, neutral networks, and intersections
5. Reference to experimental data
6. Summary

In evolution **variation** occurs on **genotypes** but **selection** operates on the **phenotype**.

Mappings from genotypes into phenotypes are highly complex objects. The only computationally accessible case is in the evolution of RNA molecules.

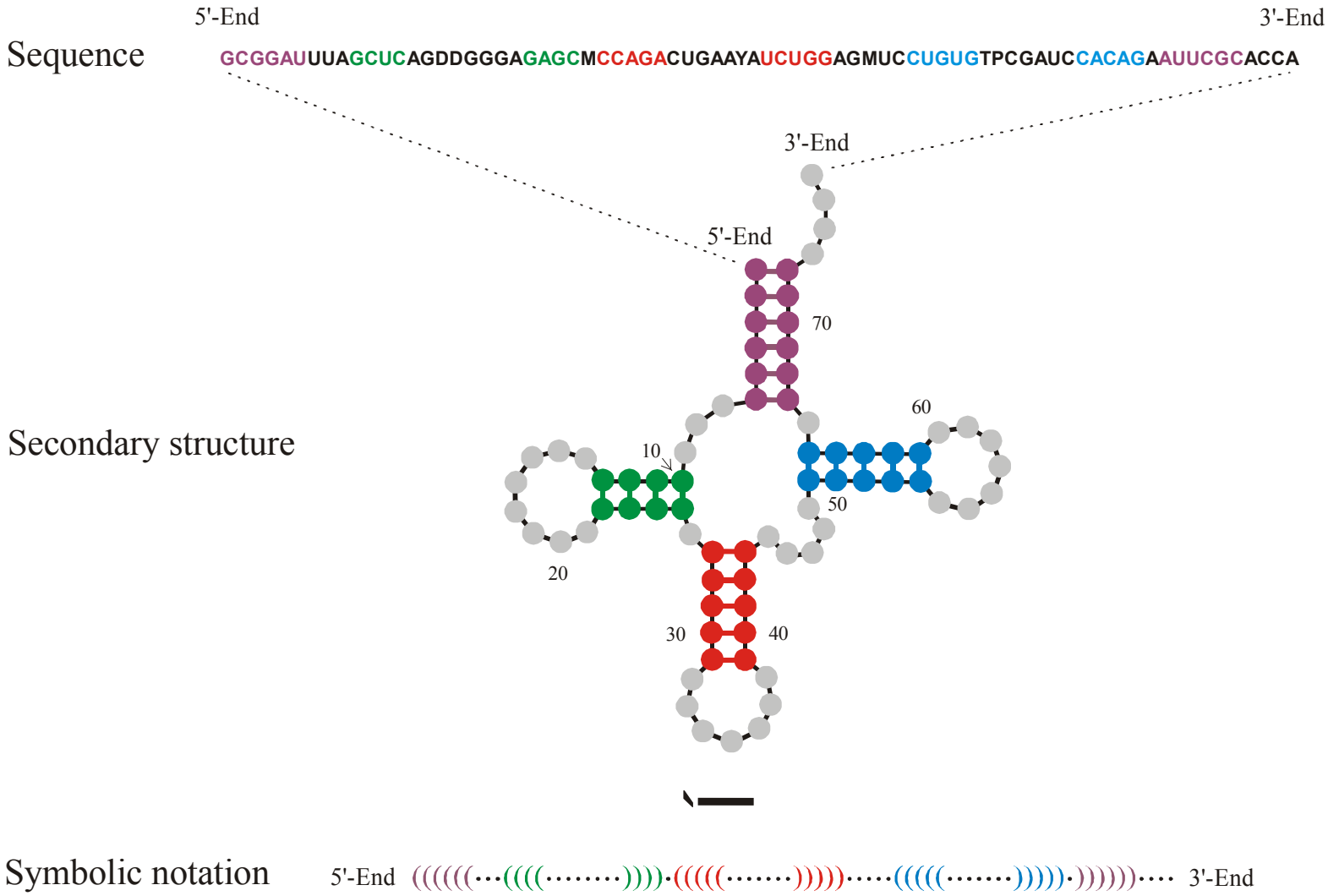
The mapping from RNA sequences into secondary structures and function,

sequence ñ **structure** ñ **function**,

is used as a model for the complex relations between genotypes and phenotypes. Fertile progeny measured in terms of **fitness** in population biology is determined quantitatively by **replication rate constants** of RNA molecules.

Population biology	Molecular genetics	Evolution of RNA molecules
Genotype	**Genome**	**RNA sequence**
Phenotype	**Organism**	**RNA structure and function**
Fitness	**Reproductive success**	**Replication rate constant**

The RNA model



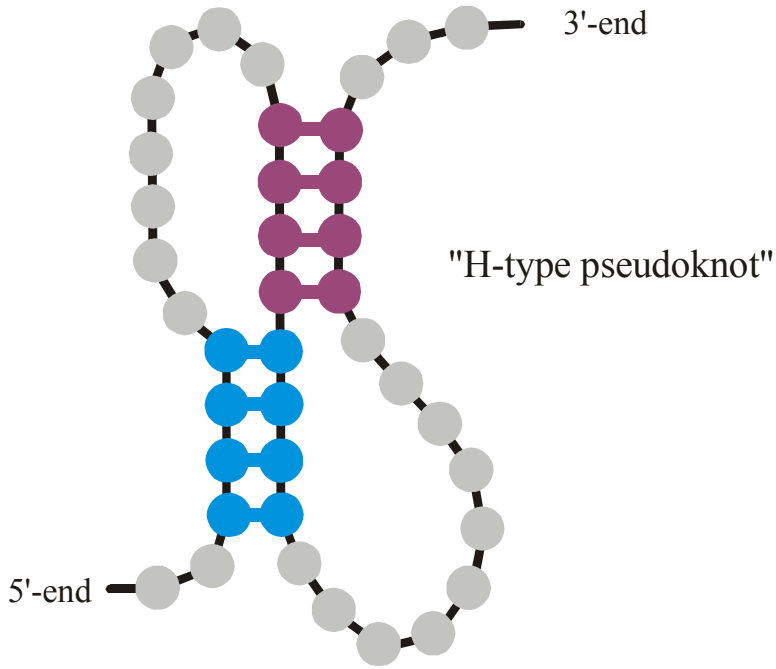
A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs

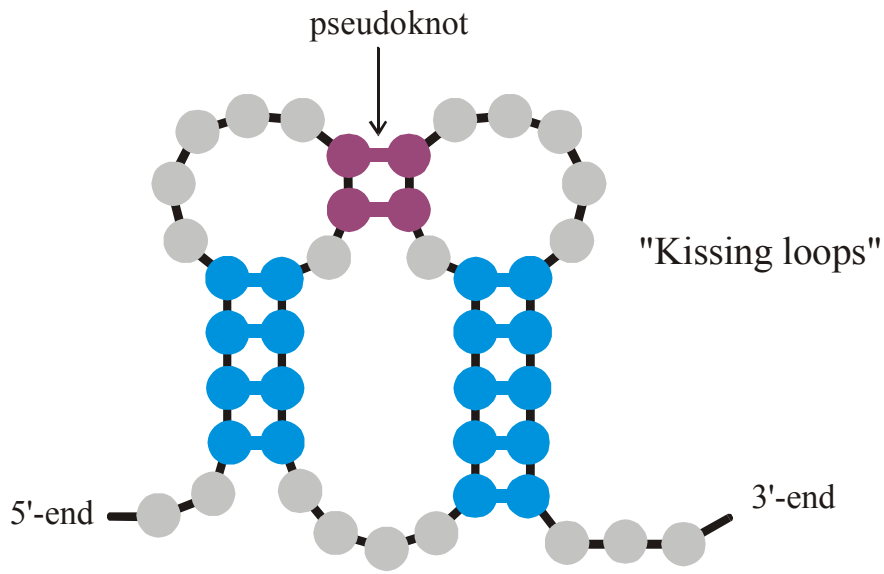
Definition and **physical relevance** of RNA secondary structures

RNA secondary structures are listings of Watson-Crick and GU wobble base pairs, which are free of knots and pseudokots.

D.Thirumalai, N.Lee, S.A.Woodson, and D.K.Klimov. *Annu.Rev.Phys.Chem*. **52**:751-762 (2001):

„**Secondary structures are folding intermediates in the formation of full three-dimensional structures**.“





..((((.....[[[[.)))).....]]]]...

..((((.....[[.))))....((((.].....))))...

Two classes of pseudoknots in RNA structures

RNA sequence: GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA folding:

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

Algorithm: Dynamic programming

Biophysical chemistry: thermodynamics and kinetics

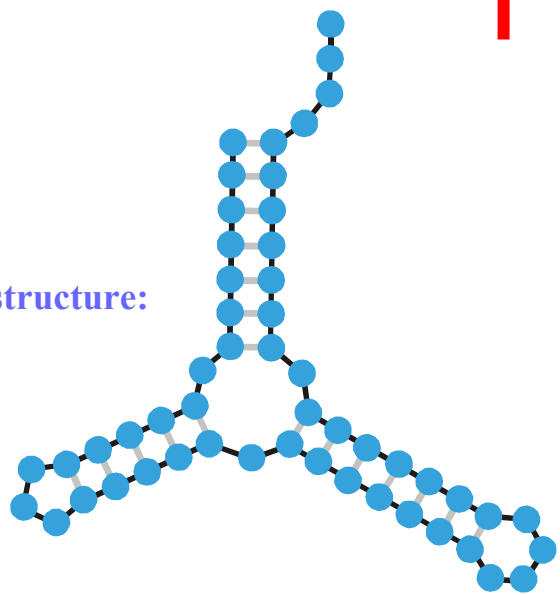
Empirical parameters

Inverse folding of RNA:

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

Algorithm: Trial-and-error search heuristic, dynamic programming

RNA structure:



Sequence and structure of RNA

How to compute RNA secondary structures

Efficient algorithms based on **dynamic programming** are available for computation of minimum free energy and **many** suboptimal secondary structures for given sequences.

M.Zuker and P.Stiegler. ***Nucleic Acids Res***. **9**:133-148 (1981)

M.Zuker, ***Science*** **244**: 48-52 (1989)

Equilibrium partition function and base pairing probabilities in Boltzmann ensembles of suboptimal structures.

J.S.McCaskill. ***Biopolymers*** **29**:1105-1190 (1990)

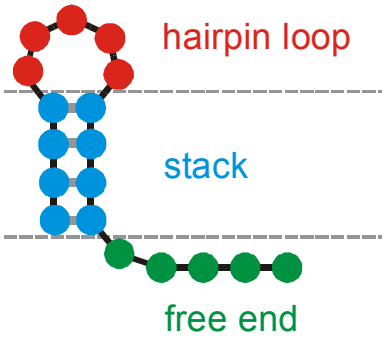
The **Vienna RNA Package** provides in addition: **inverse folding** (computing sequences for given secondary structures), computation of melting profiles from partition functions, **all** suboptimal structures within a given energy interval, barrier tress of suboptimal structures, **kinetic folding** of RNA sequences, RNA-hybridization and RNA/DNA-hybridization through **cofolding** of sequences, alignment, etc..

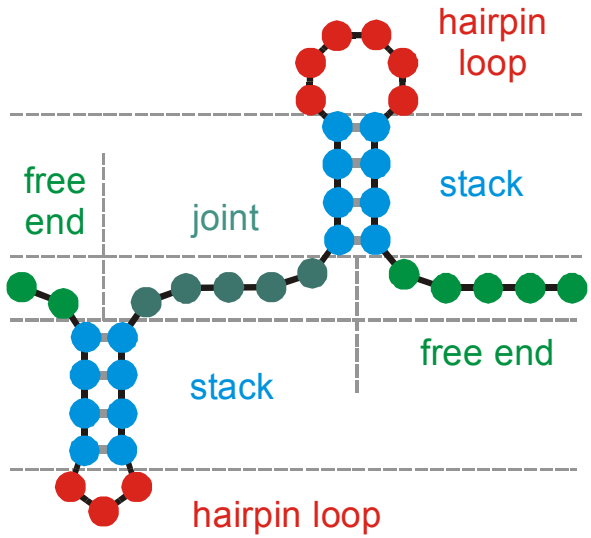
I.L.Hofacker, W. Fontana, P.F.Stadler, L.S.Bonhoeffer, M.Tacker, and P. Schuster. ***Mh.Chem.*** **125**:167-188 (1994)

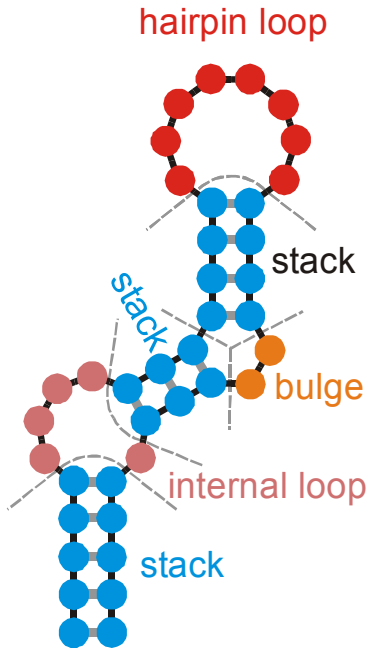
S.Wuchty, W.Fontana, I.L.Hofacker, and P.Schuster. ***Biopolymers*** **49**:145-165 (1999)

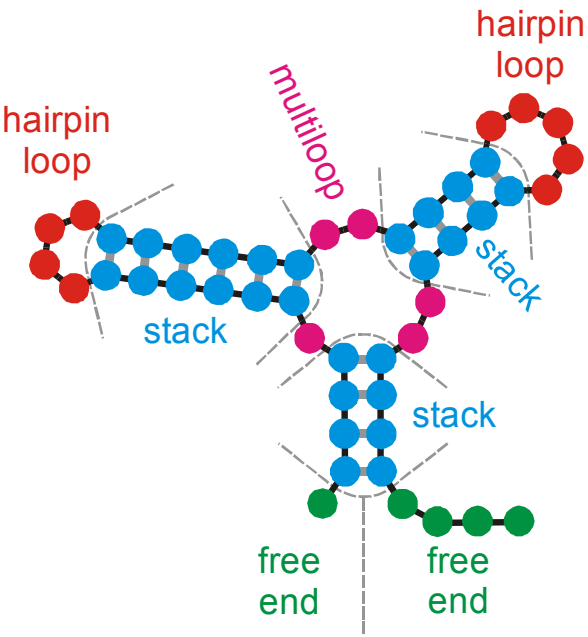
C.Flamm, W.Fontana, I.L.Hofacker, and P.Schuster. ***RNA*** **6**:325-338 (1999)

Vienna RNA Package: http://www.tbi.univie.ac.at





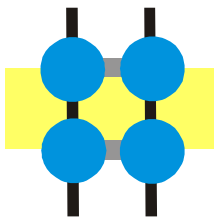




Elements of RNA secondary structures
as used in free energy calculations

5’-end 3’-end

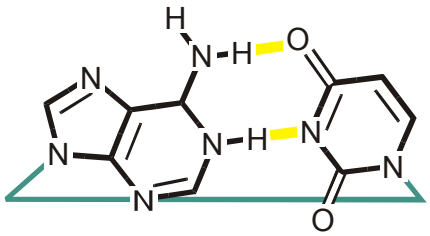
GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGAGCGUCCCAUCGGUACUCCA



free energy of stacking < 0

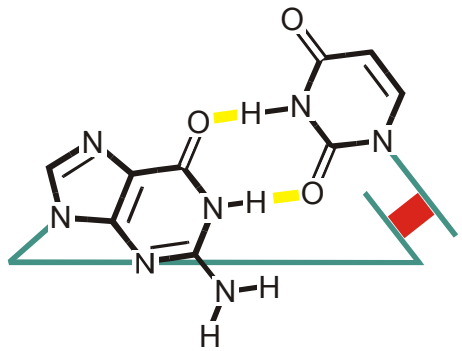
$$\Delta G_0^{300} = \underbrace{\sum g_{ij,kl}}_{\text{stacks of base pairs}} + \underbrace{\sum h(n_l)}_{\text{hairpin loops}} + \underbrace{\sum b(n_b)}_{\text{bulges}} + \underbrace{\sum i(n_i)}_{\text{internal loops}} + \cdots$$

Folding of RNA sequences into secondary structures of minimal free energy, DG_0^{300}

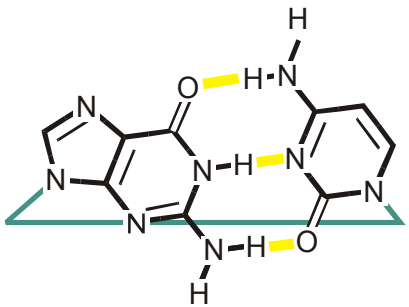


A=U

(U=A)

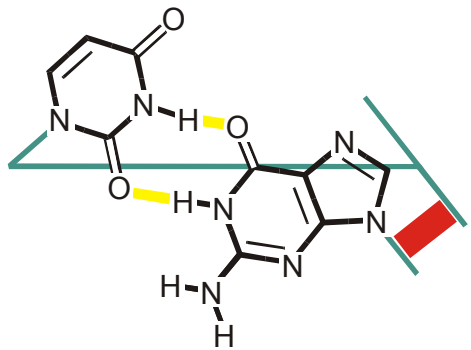


G=U



G Í C

(C Í G)



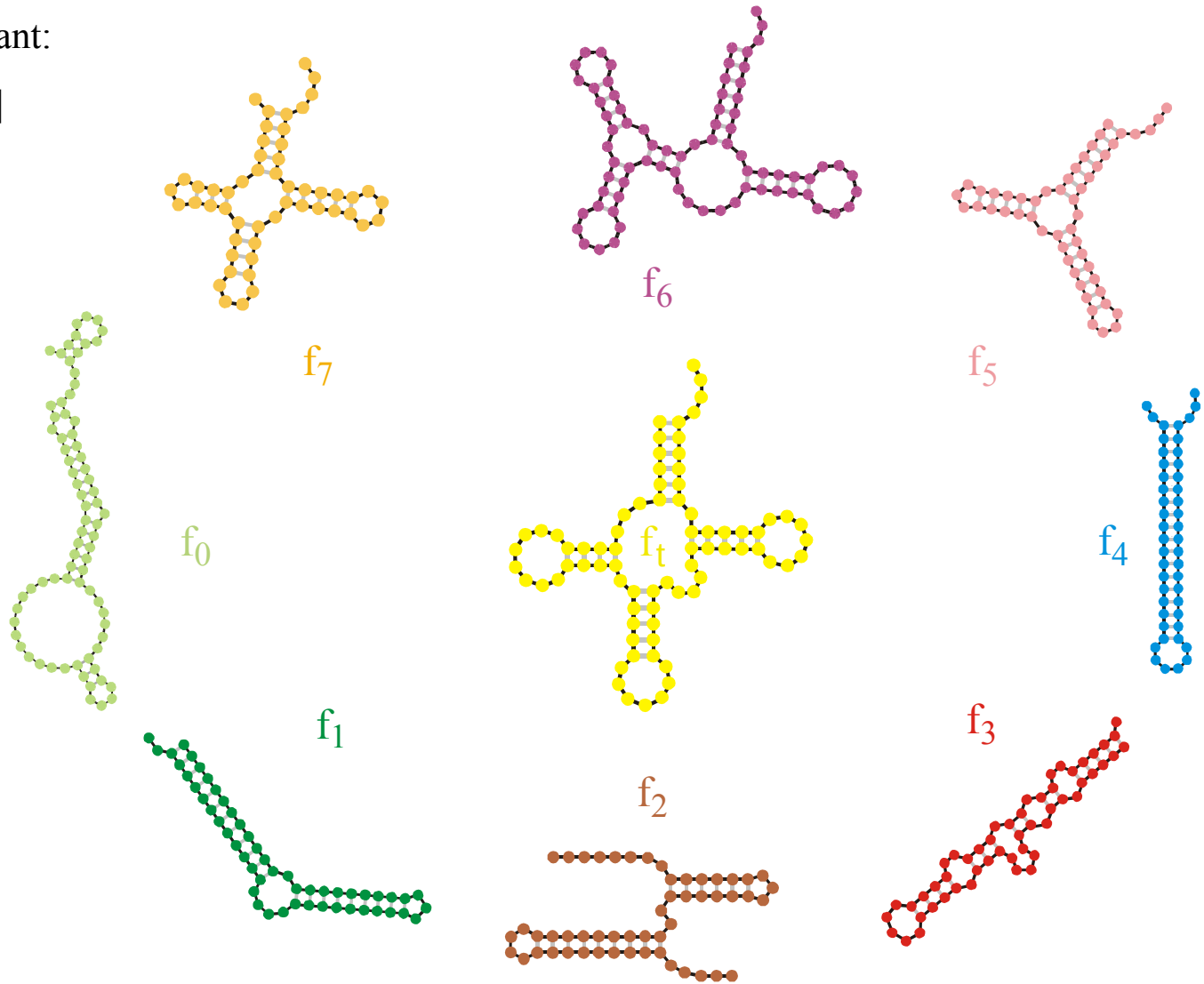
U=G

Three base pairing alphabets built from natural nucleotides **A**, **U**, **G**, and **C**

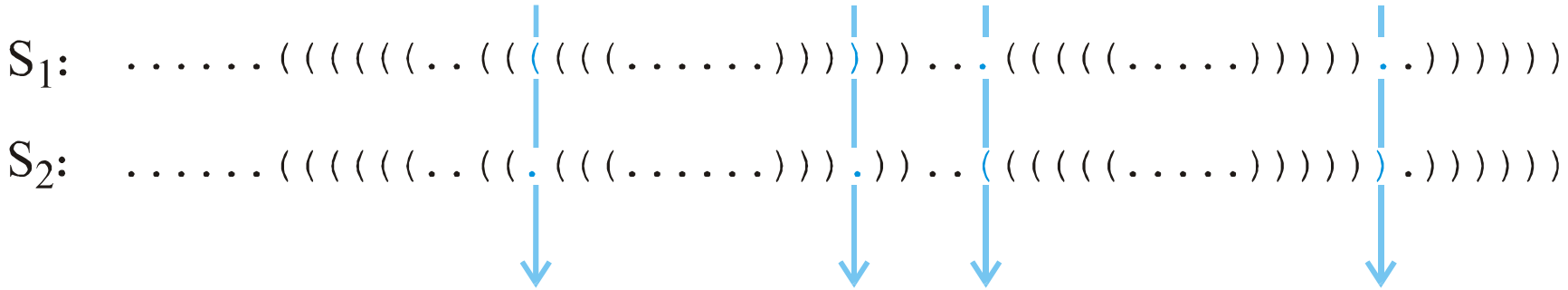
Replication rate constant:

$$f_k = g / [a + Dd_S^{(k)}]$$

$$Dd_S^{(k)} = d_H(S_k, S_t)$$



Evaluation of RNA secondary structures yields replication rate constants



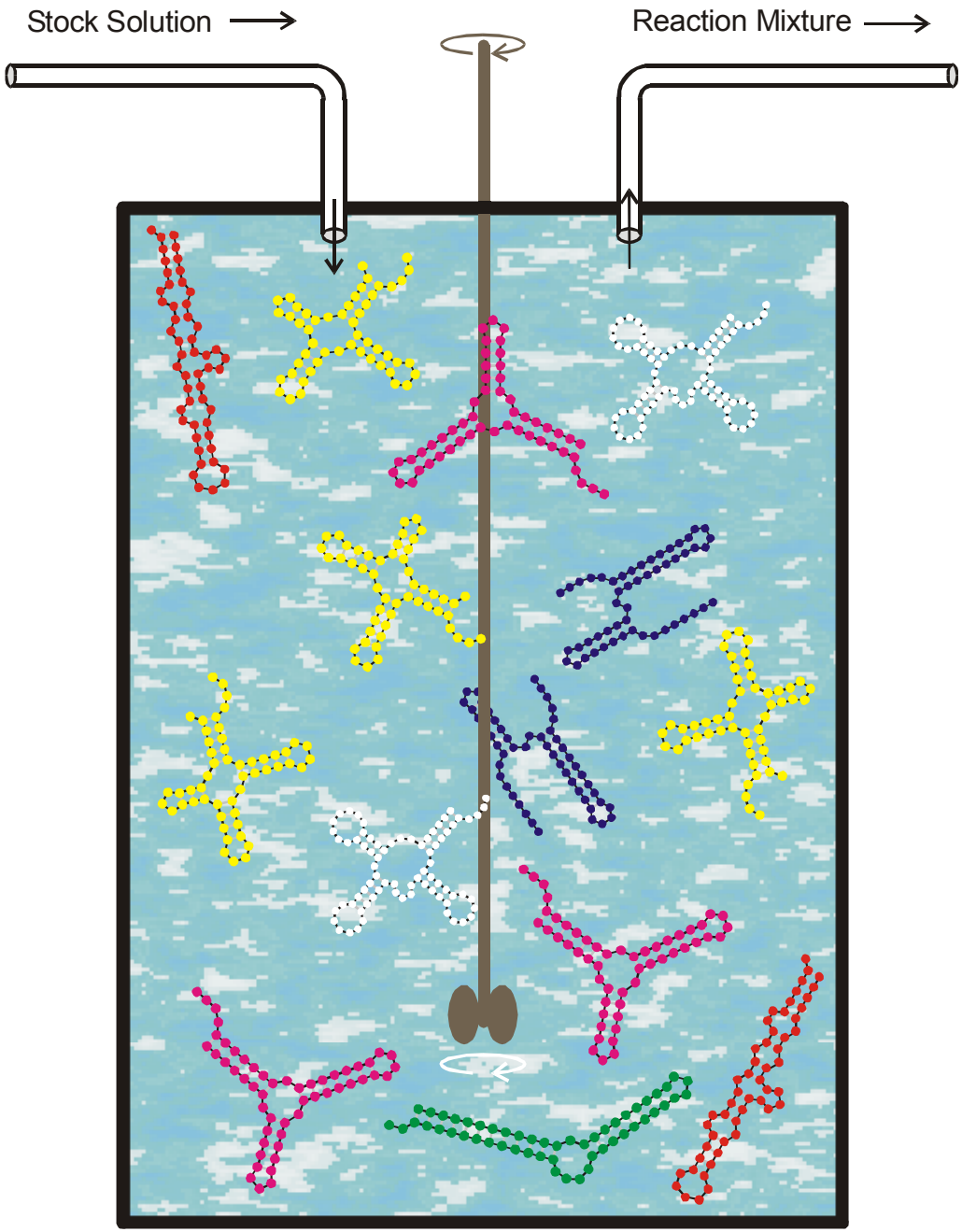
Hamming distance $d_H(S_1,S_2) = 4$

(i) $d_H(S_1,S_1) = 0$

(ii) $d_H(S_1,S_2) = d_H(S_2,S_1)$

(iii) $d_H(S_1,S_3) \P d_H(S_1,S_2) + d_H(S_2,S_3)$

The Hamming distance between structures in parentheses notation forms a metric in structure space



Replication rate constant:

$$f_k = g / [a + Dd_S^{(k)}]$$

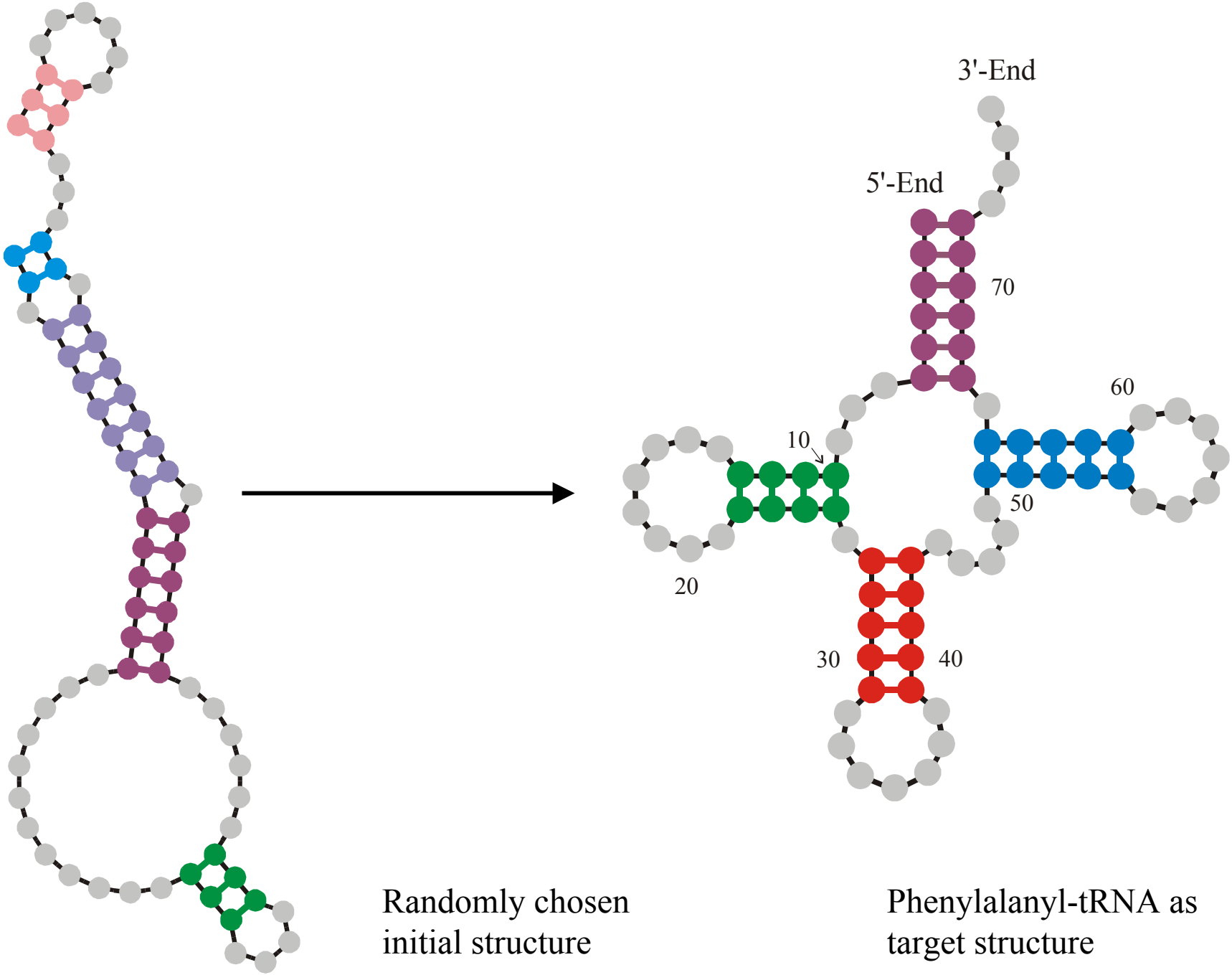
$$Dd_S^{(k)} = d_H(S_k, S_t)$$

Selection constraint:

RNA molecules is controlled by the flow

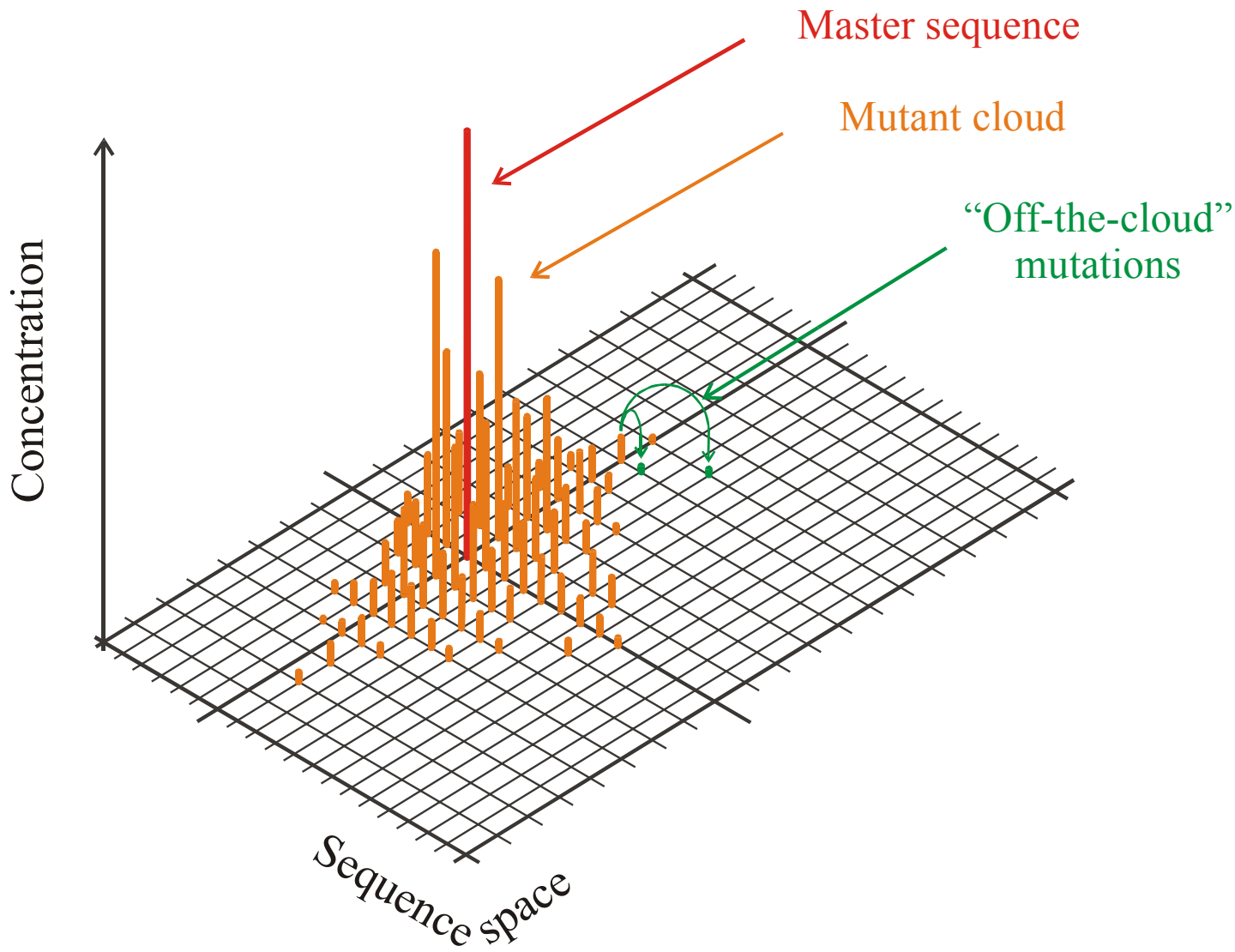
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

The flowreactor as a device for studies of evolution ***in vitro*** and ***in silico***

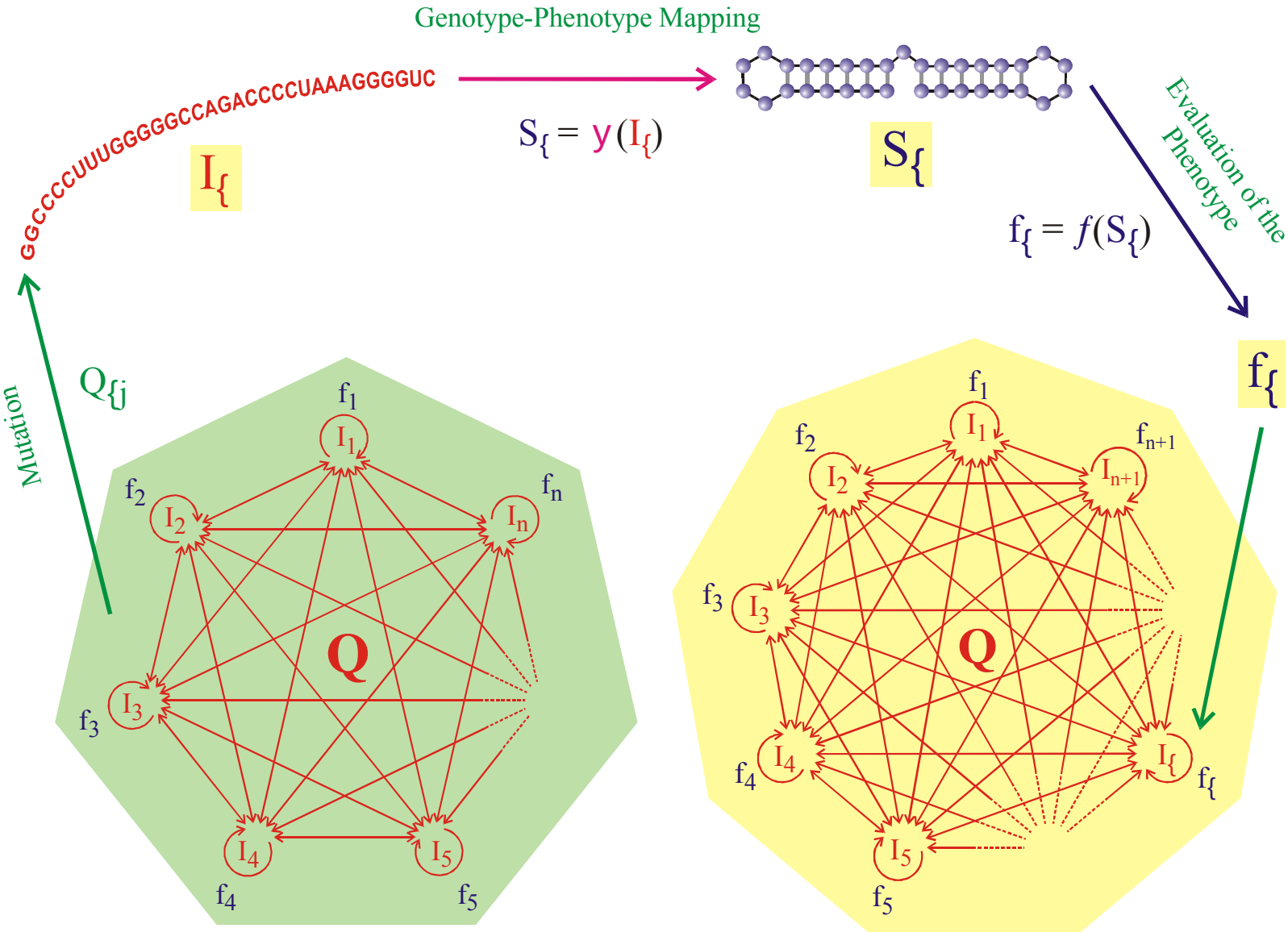


Randomly chosen
initial structure

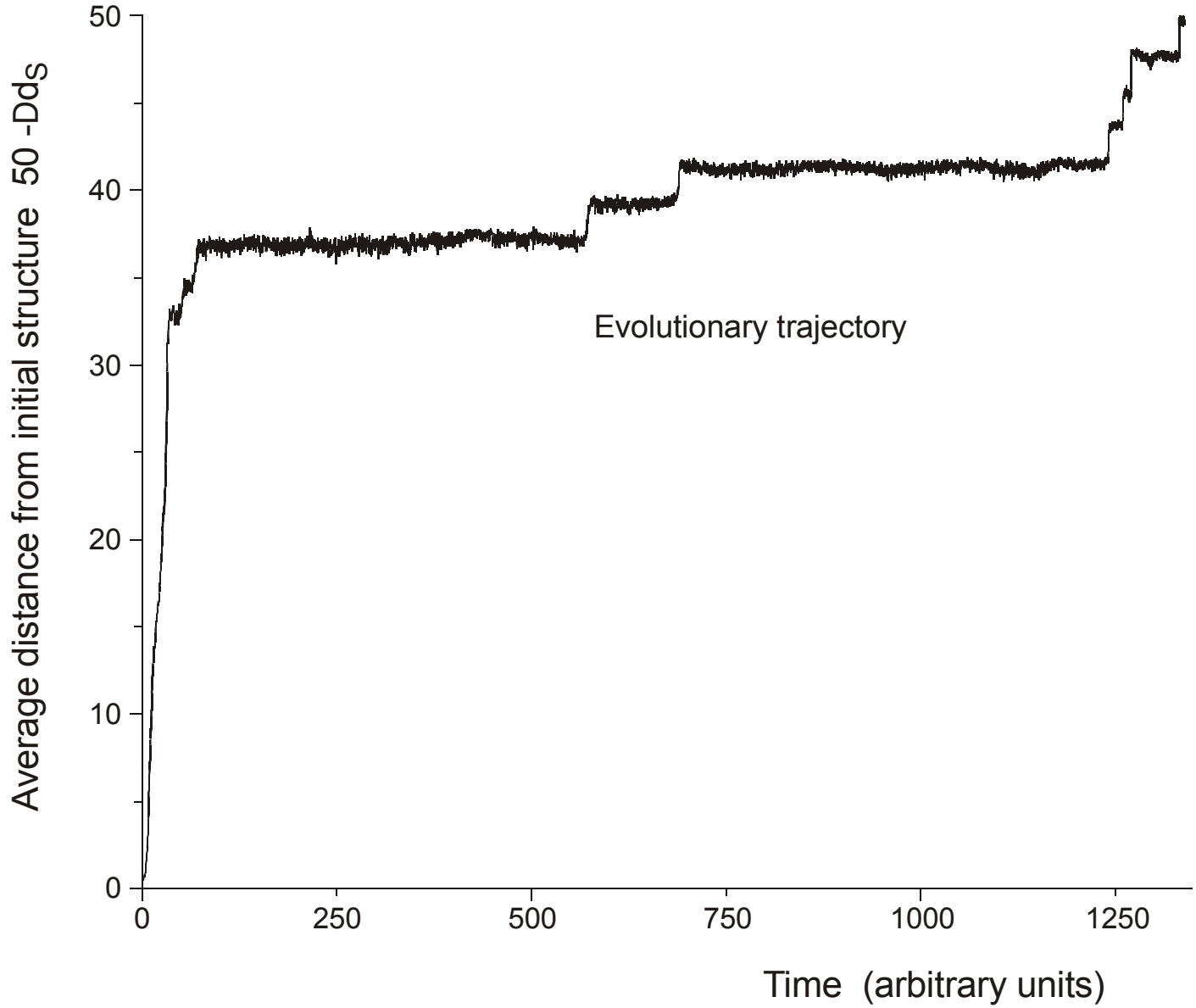
Phenylalanyl-tRNA as
target structure



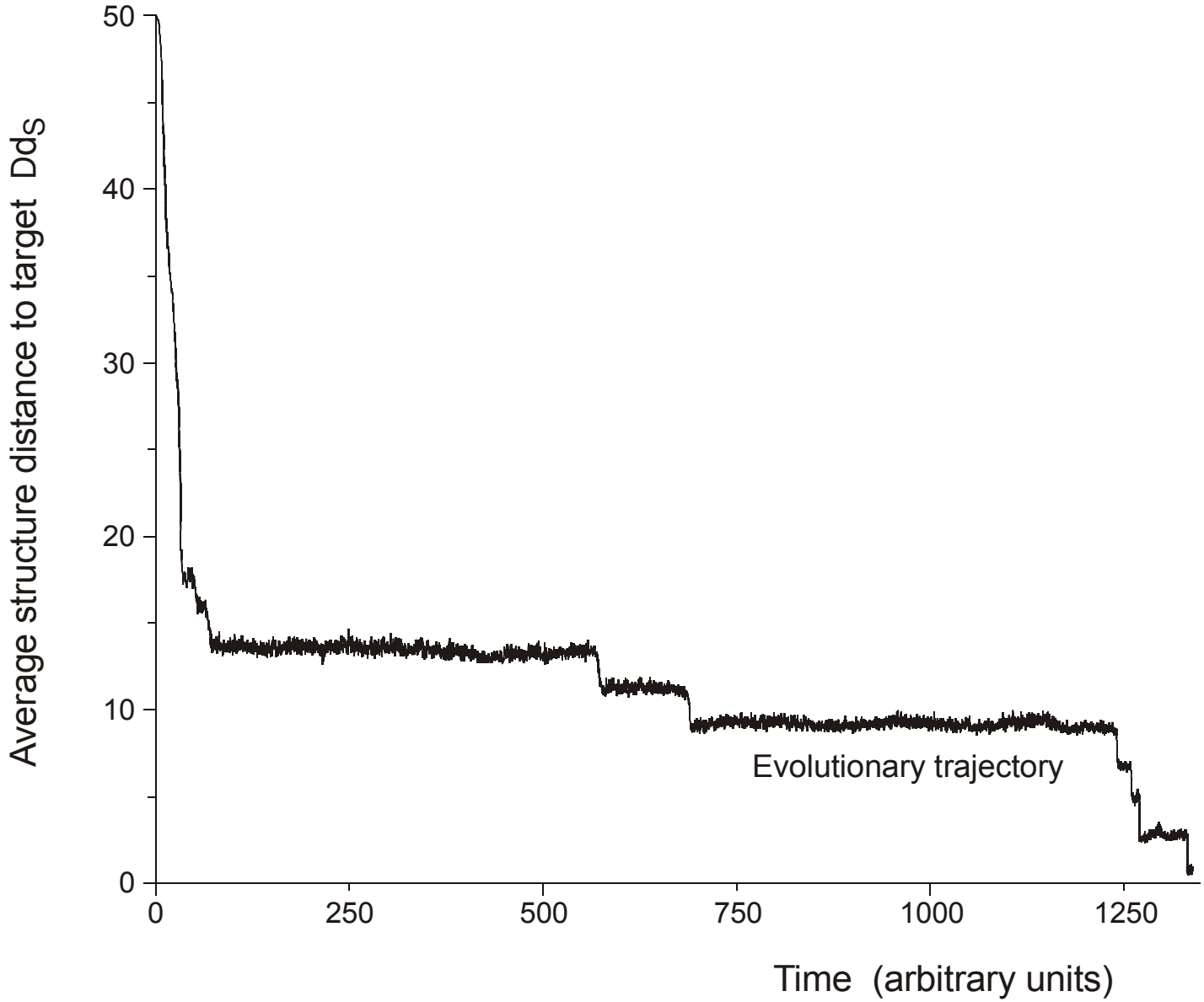
The molecular quasispecies in sequence space



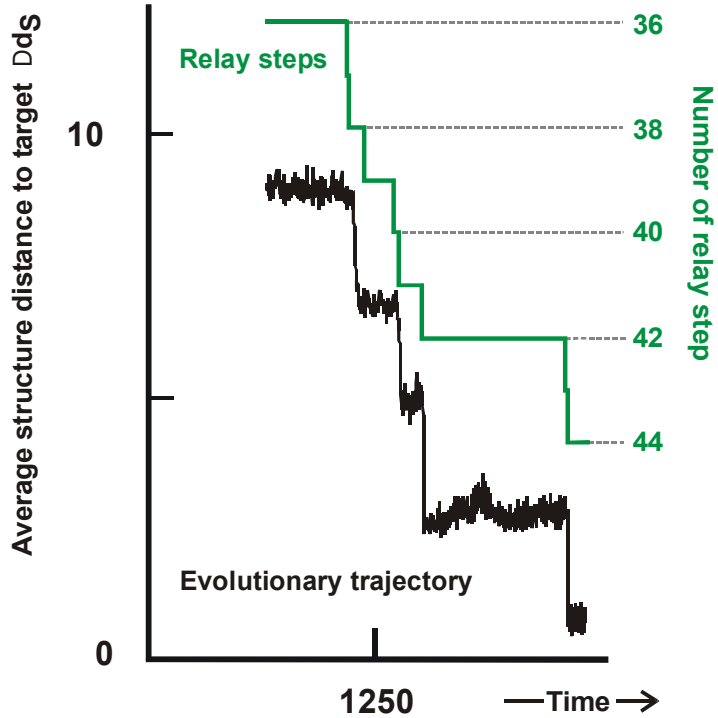
Evolutionary dynamics
including molecular phenotypes

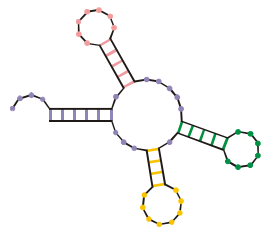


In silico optimization in the flow reactor: Trajectory (**biologists‘ view**)



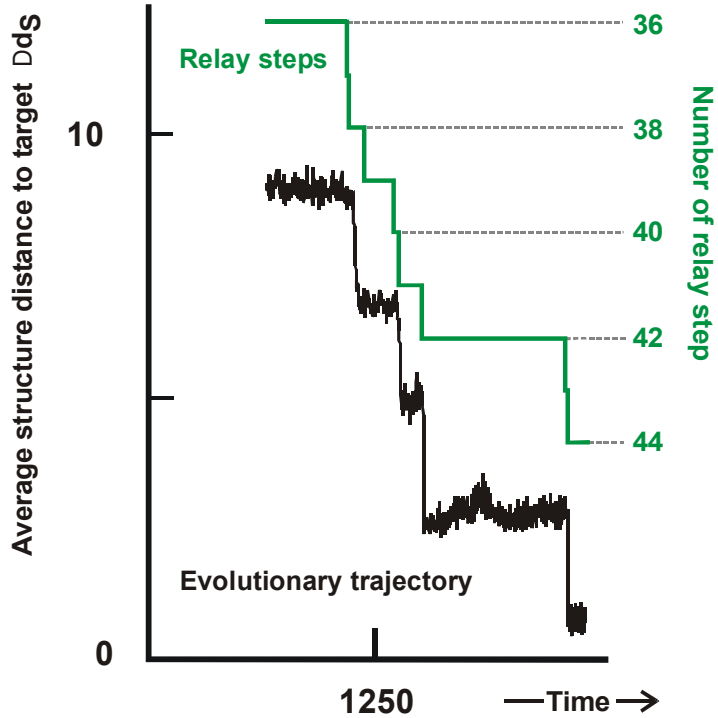
In silico optimization in the flow reactor: Trajectory (**physicists‘ view**)

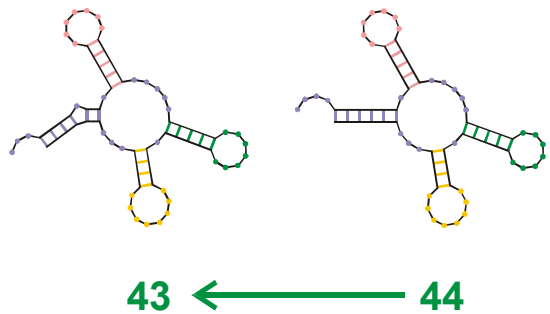




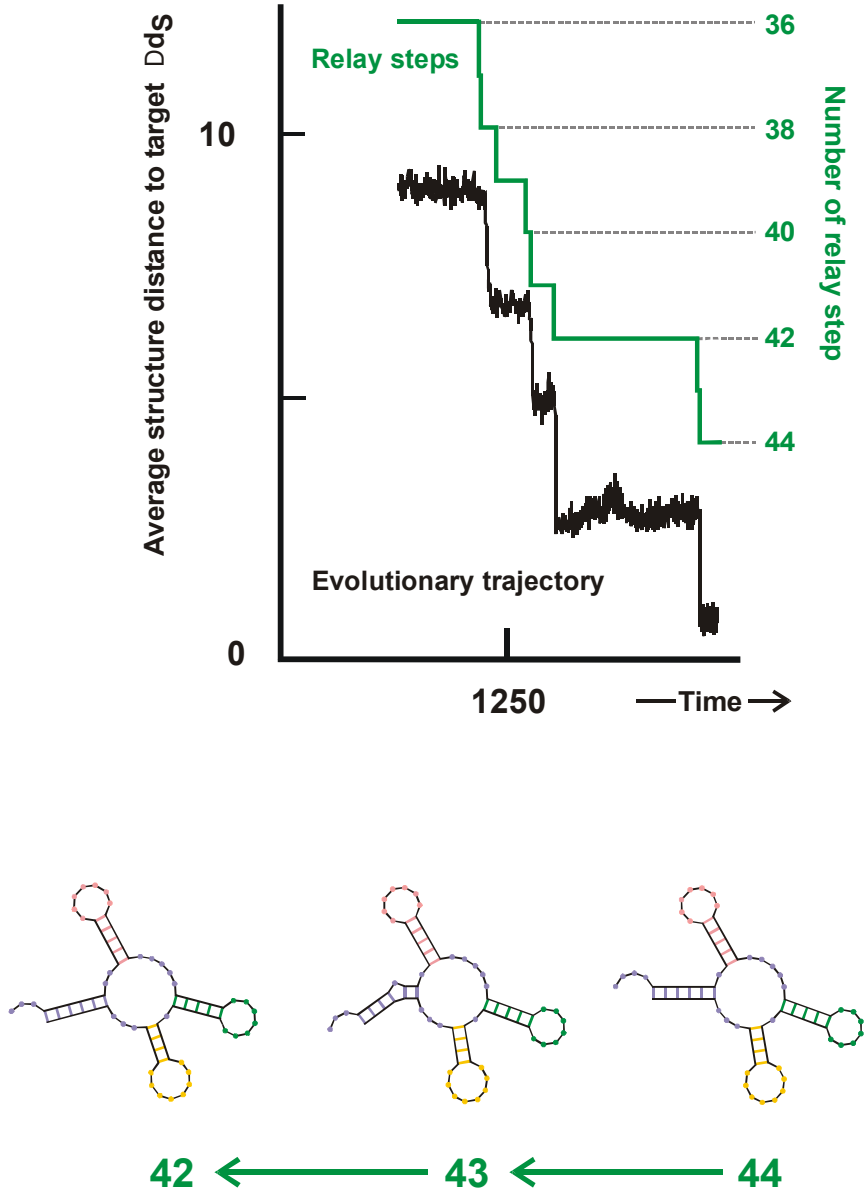
44

Endconformation of optimization

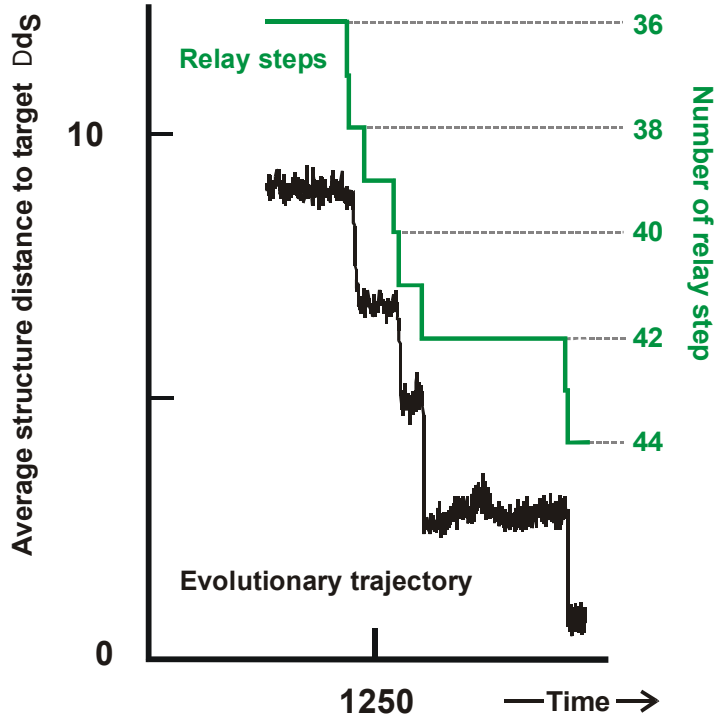


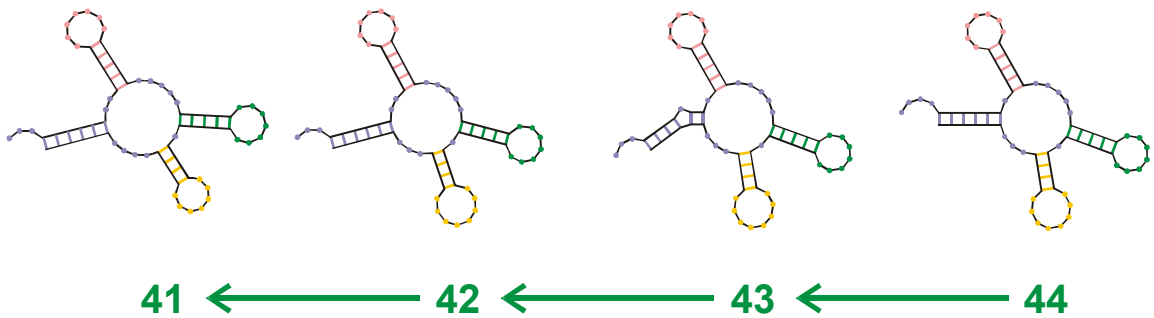


Reconstruction of the last step 43 Á 44

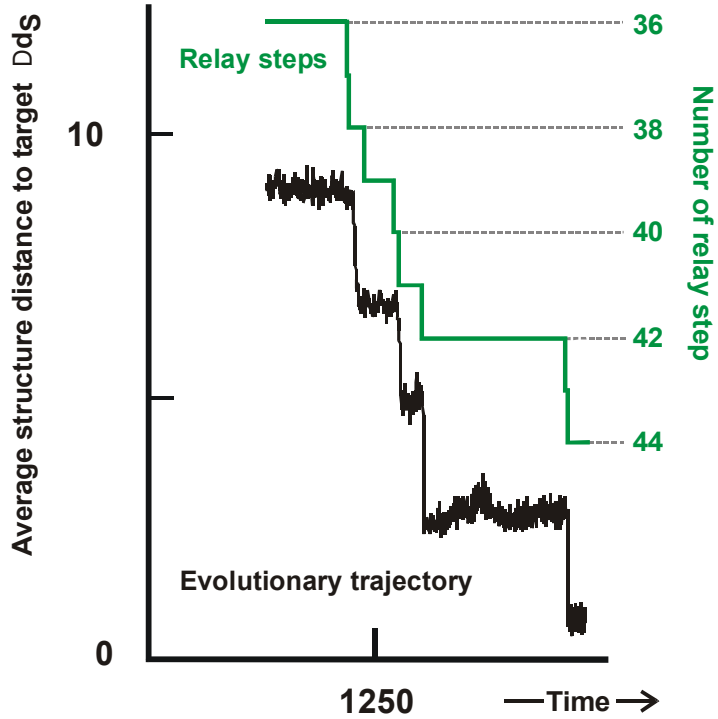


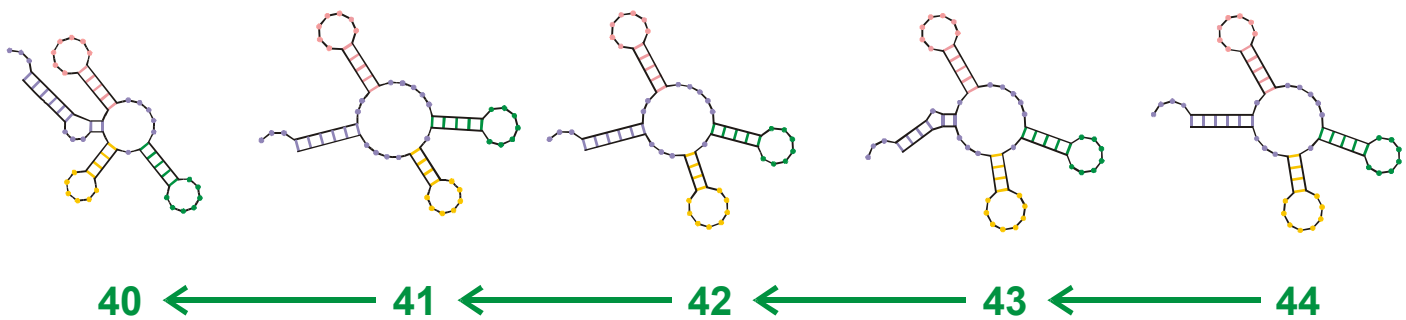
Reconstruction of last-but-one step 42 Á 43 (Á 44)



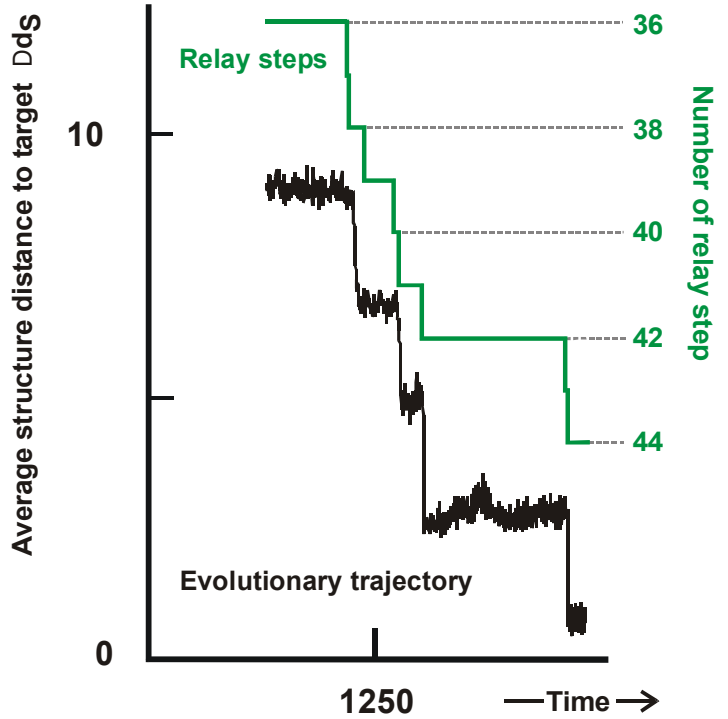


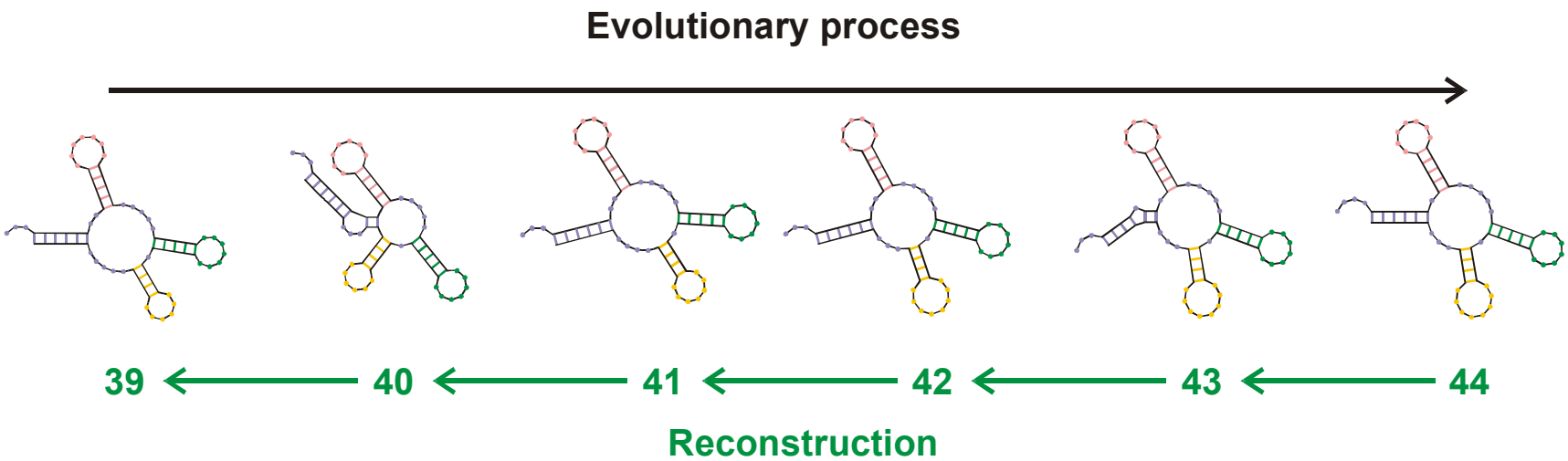
Reconstruction of step 41 Á 42 (Á 43 Á 44)



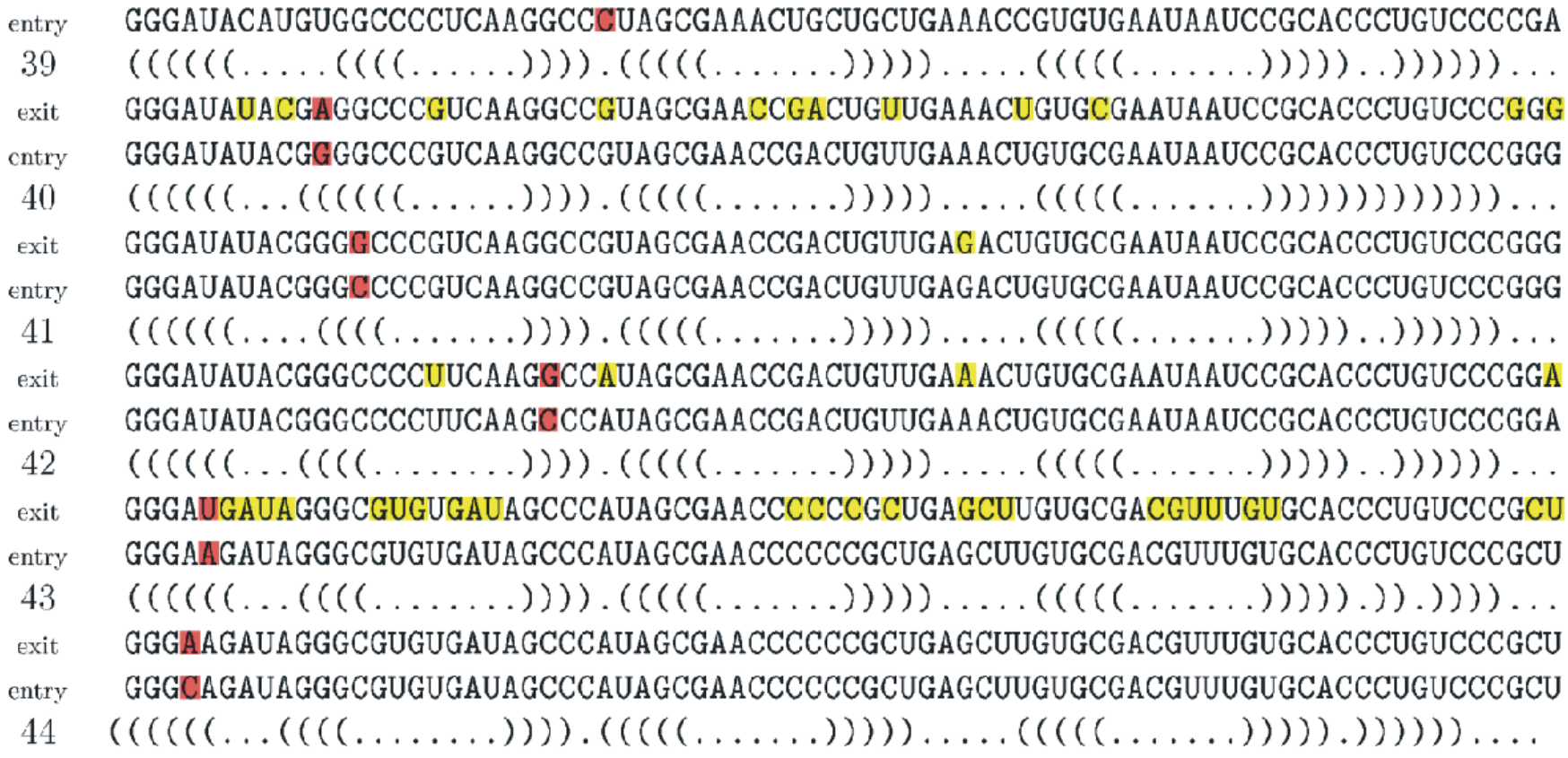


Reconstruction of step 40 Á 41 (Á 42 Á 43 Á 44)



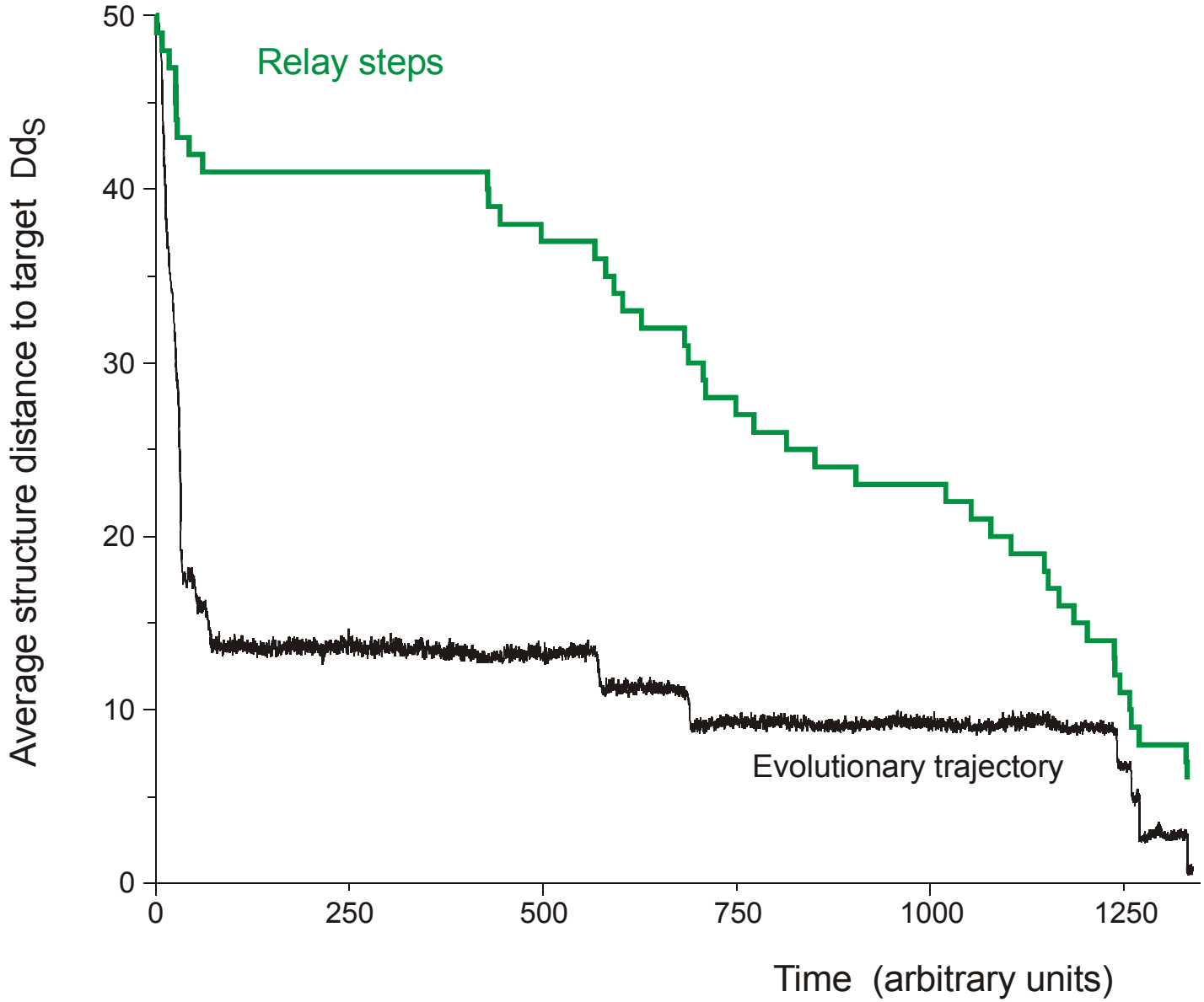


Reconstruction of the relay series



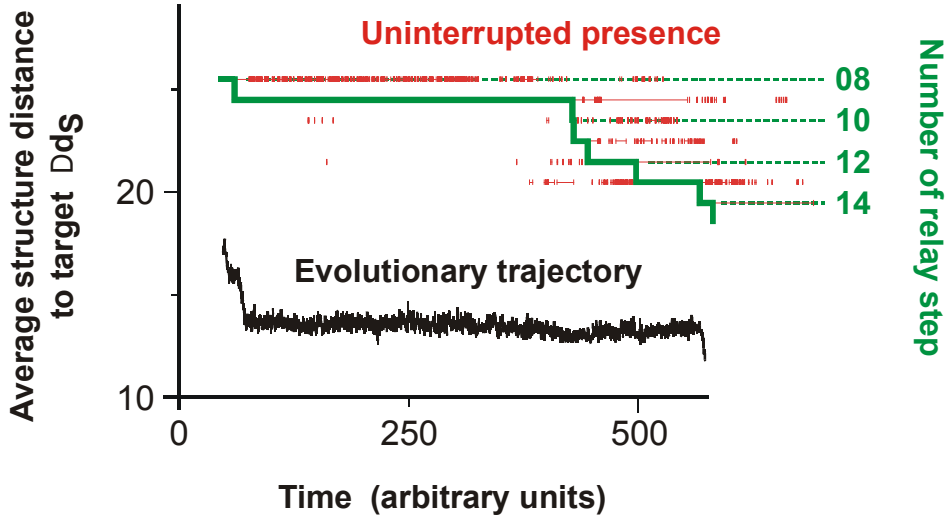
Transition inducing point mutations **Neutral point mutations**

Change in RNA sequences during the final five relay steps 39 Á 44



In silico optimization in the flow reactor: Trajectory and relay steps

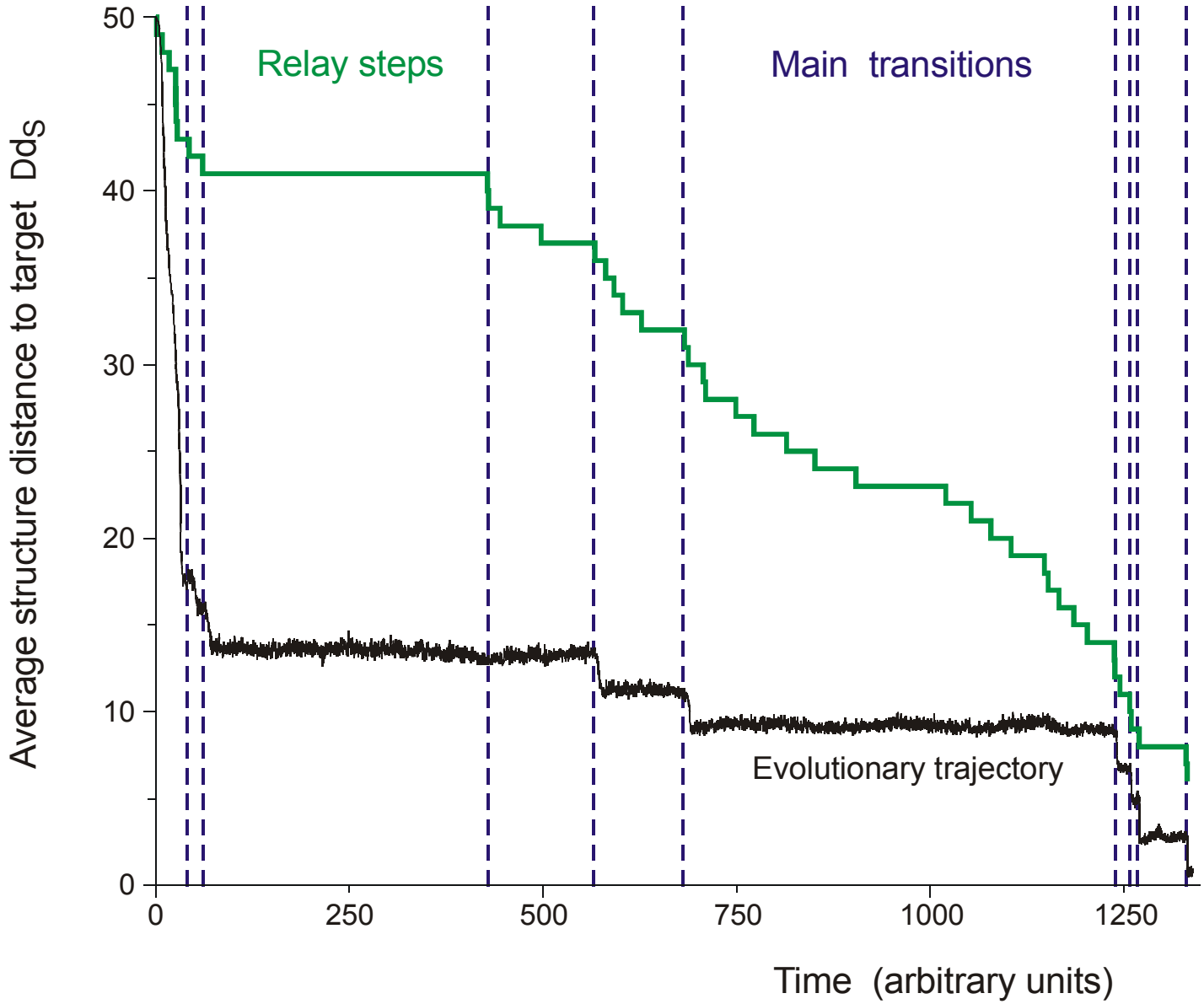
28 neutral point mutations during a long quasi-stationary epoch



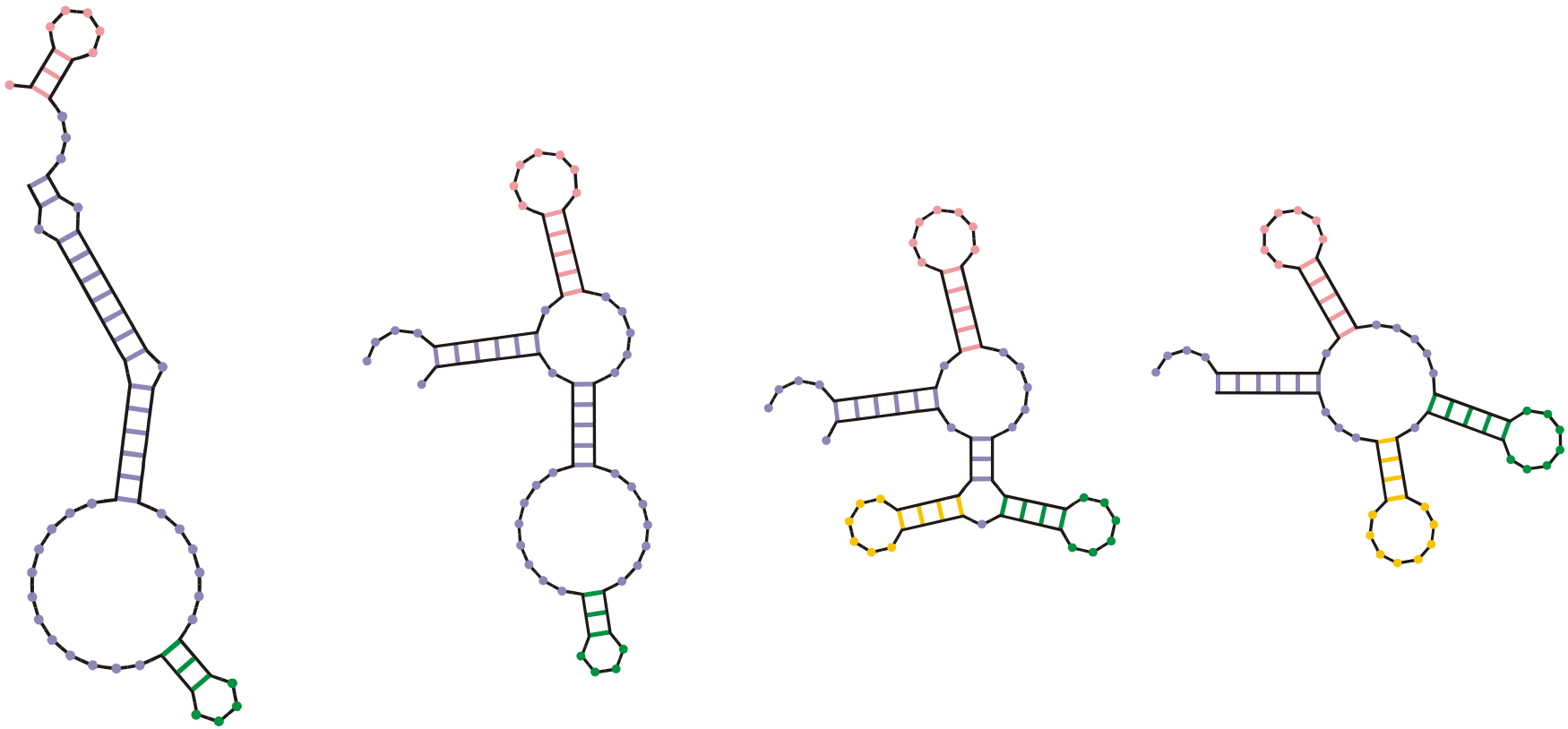

```
entry   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8       .((((((((((((.........(((....)))........)))))).....((((.........))))))))))....
exit    GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCAUACAGAA
entry   GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUACCAUACAGAA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

Transition inducing point mutations — **Neutral point mutations**

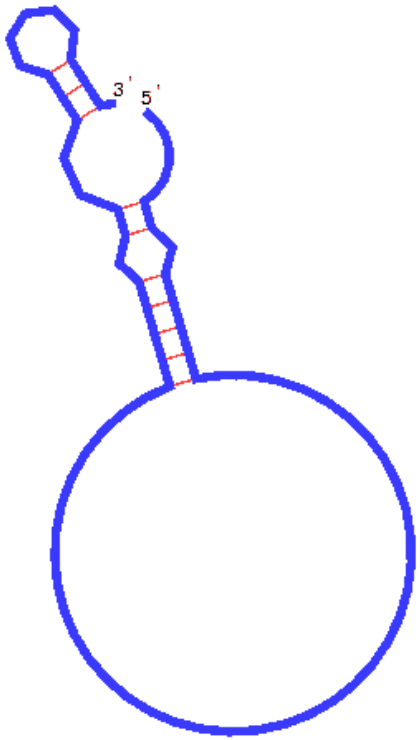
Neutral genotype evolution during phenotypic stasis



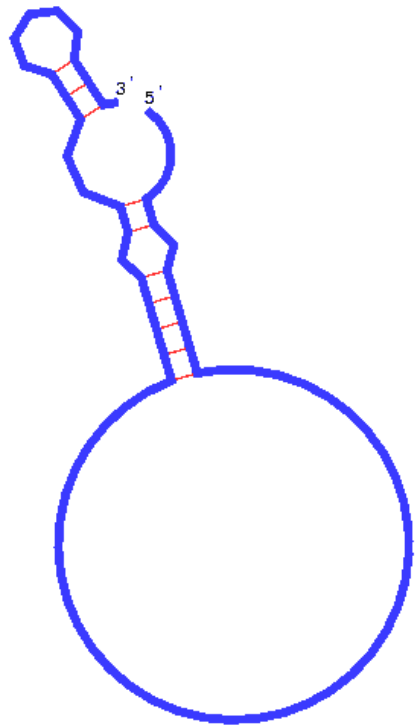
In silico optimization in the flow reactor: Main transitions



Three important steps in the formation of the tRNA clover leaf from a randomly chosen initial structure corresponding to three **main transitions**.

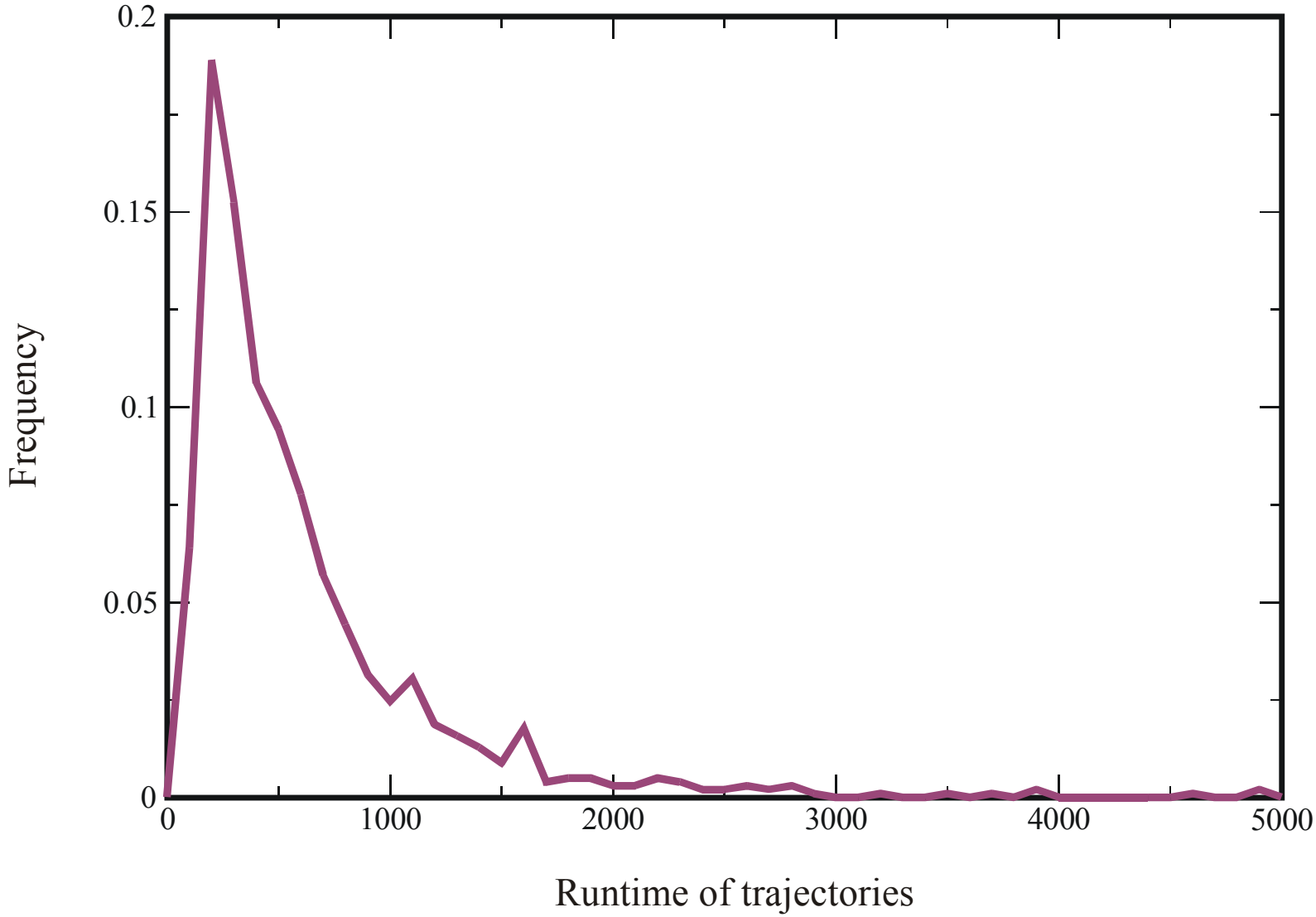


AUGC

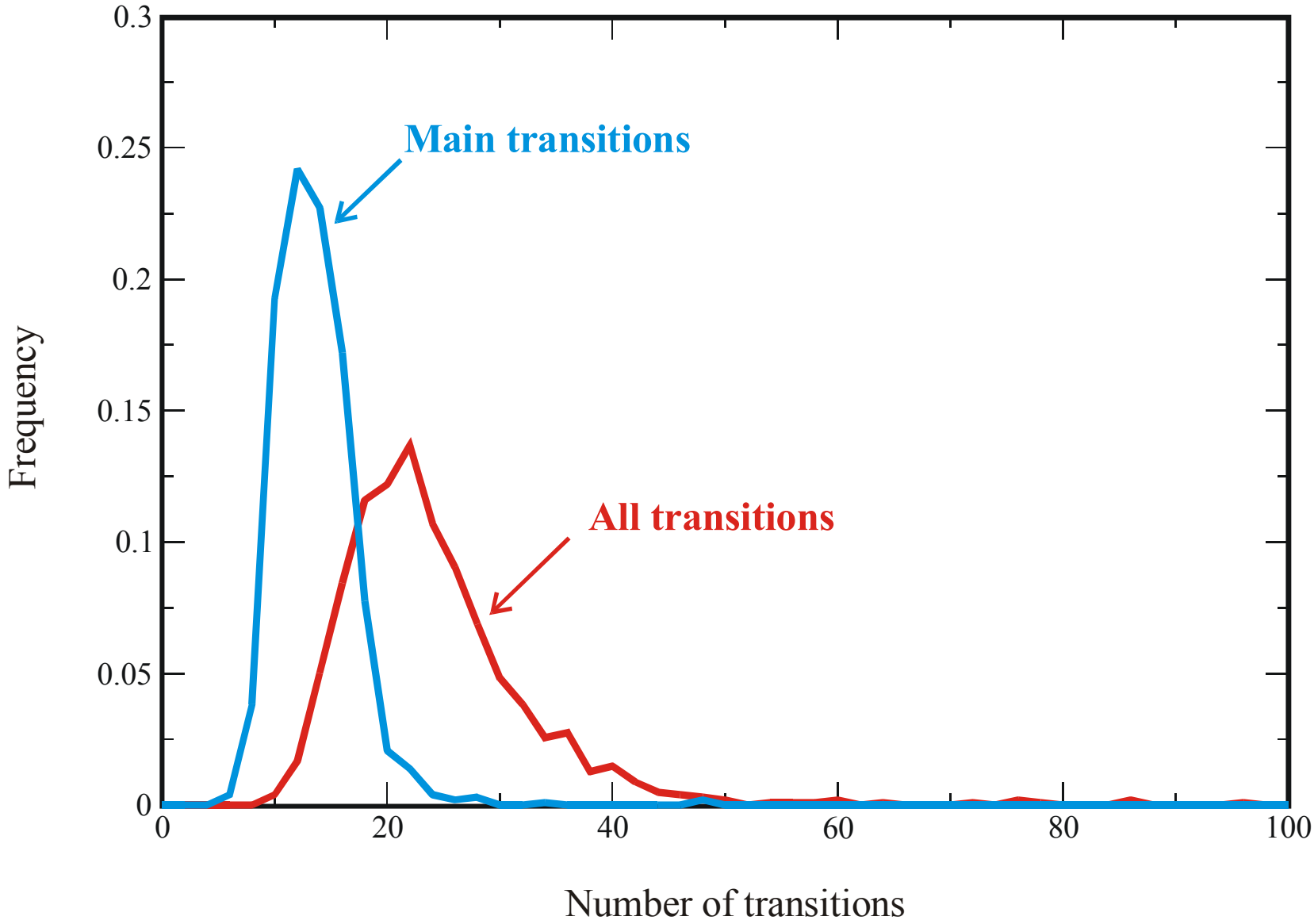


GC

Movies of optimization trajectories over the **AUGC** and the **GC** alphabet



Statistics of the lengths of trajectories from initial structure to target (**AUGC**-sequences)



Statistics of the numbers of transitions from initial structure to target (**AUGC**-sequences)

Alphabet	Runtime	Transitions	Main transitions	No. of runs
AUGC	385.6	22.5	12.6	1017
GUC	448.9	30.5	16.5	611
GC	2188.3	40.0	20.6	107

Statistics of trajectories and relay series (mean values of log-normal distributions)