Mathematische Probleme aus den „Life-Sciences“

Peter Schuster

Institut für Theoretische Chemie und Molekulare Strukturbiologie der Universität Wien

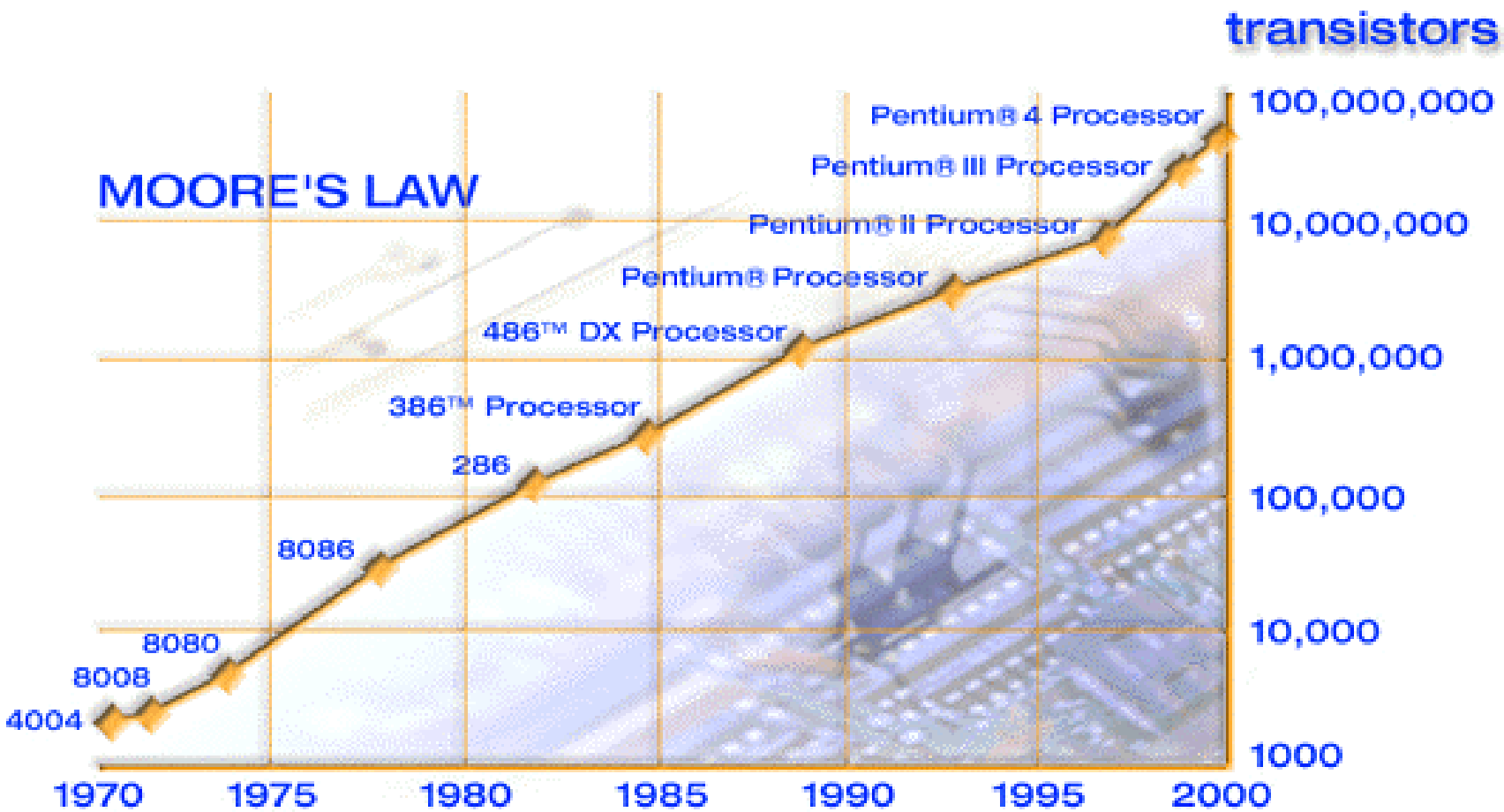
Vortragsreihe „Mathematik im Betrieb“

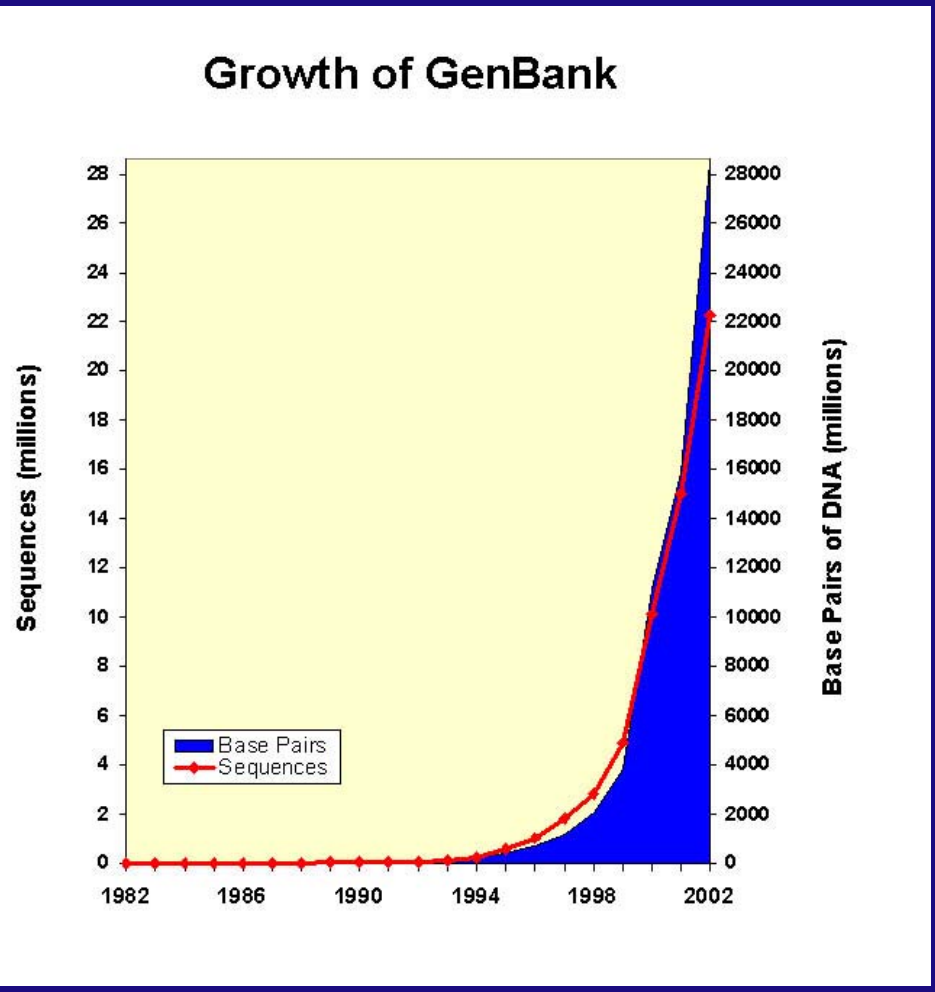
Dornbirn, 27.05.2004

Web-Page for further information:

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

1. Komplexität in der Biologie
2. Evolutionäre Optimierung und Lernen im Ensemble
3. Strukturbildung von Biomolekülen als kombinatorisches Problem
4. Modellbildung in der Neurobiologie


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



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)

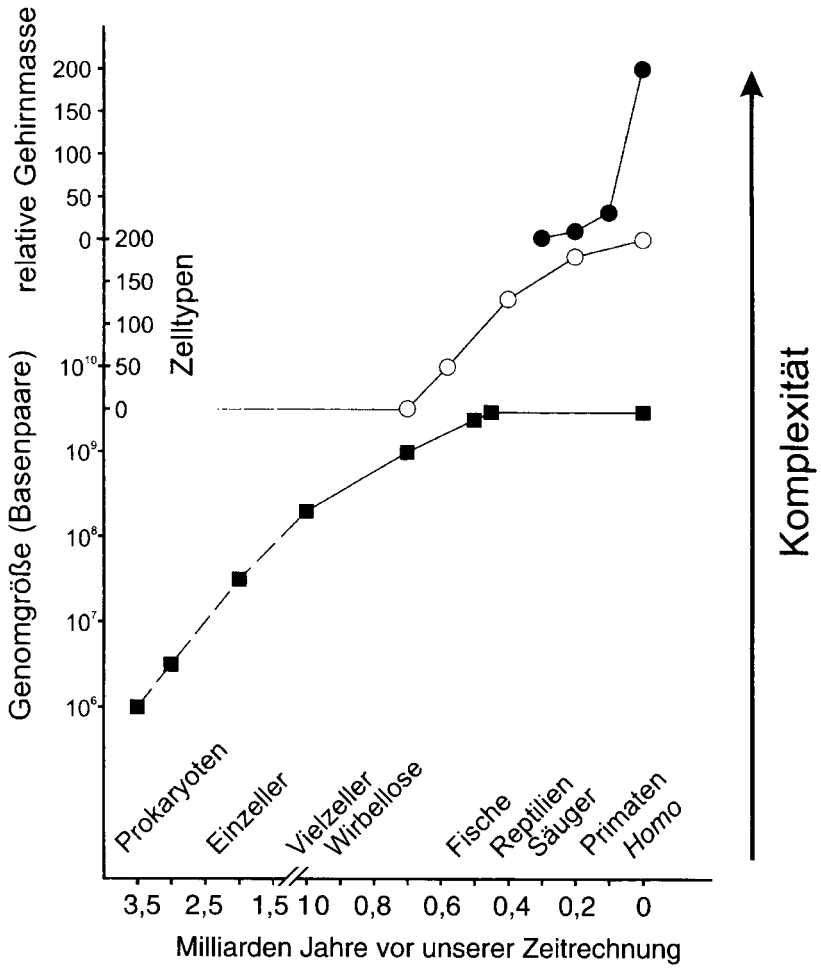
(*Eukarya* examples: chimpanzee, turkey, chicken, ape, corn, potato, rice, banana, tomato, cotton, coffee, soybean, pig, rat, cat, sheep, horse, kangaroo, dog, cow, bee, salmon, fugu, frog, …)

- *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

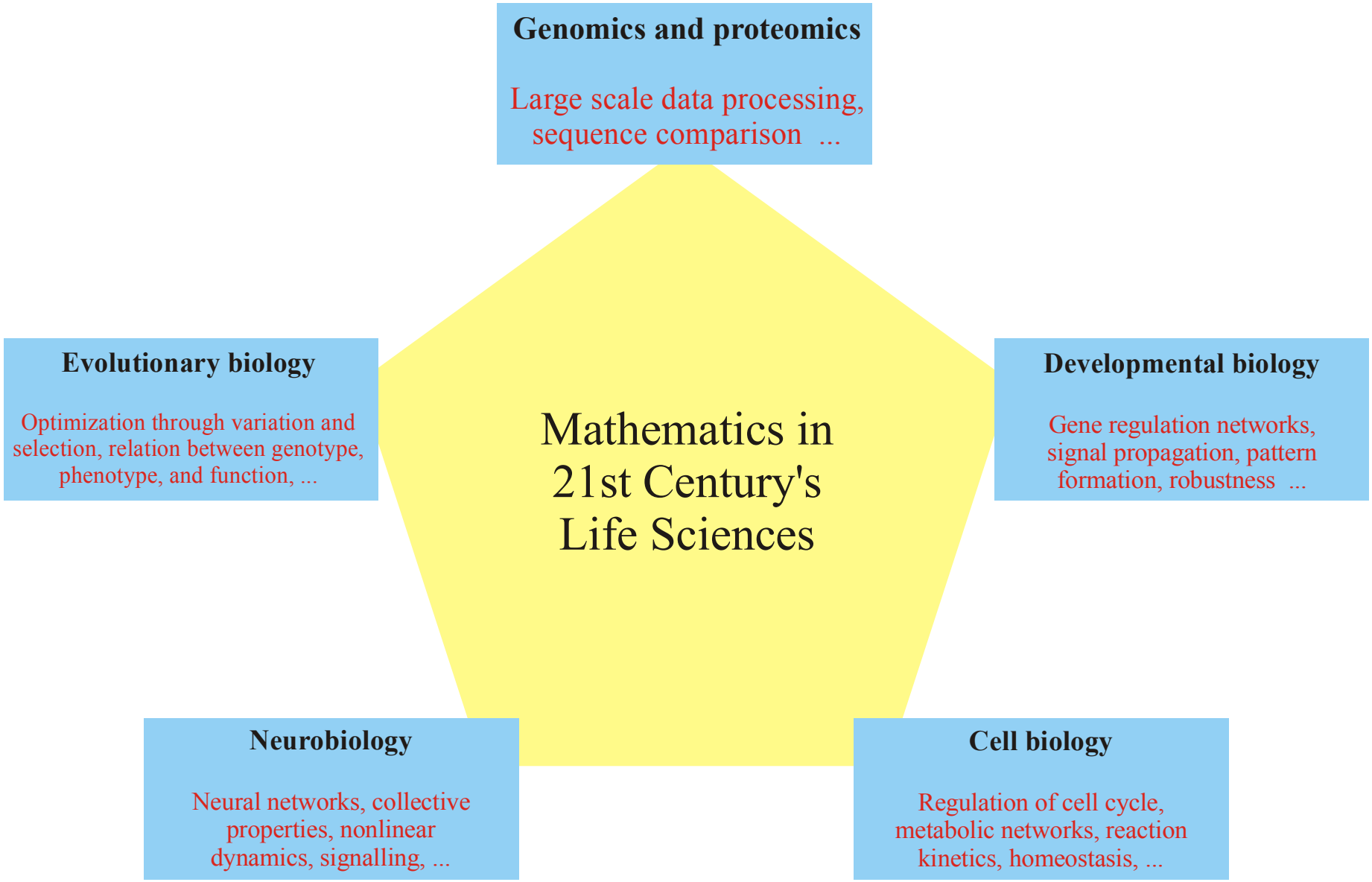
1. **Komplexität in der Biologie**
2. Evolutionäre Optimierung und Lernen im Ensemble
3. Strukturbildung von Biomolekülen als kombinatorisches Problem
4. Modellbildung in der Neurobiologie



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.

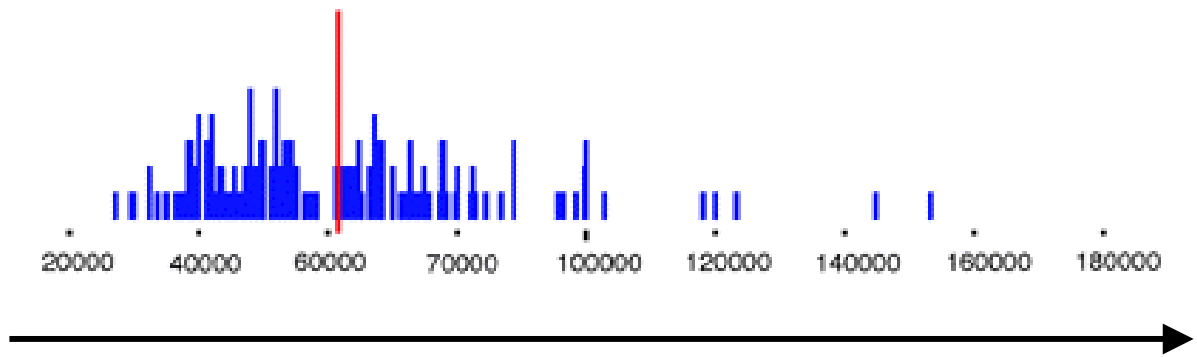
A.C.Wilson. The Molecular Basis of Evolution. Scientific American, Oct.1985, 164-173.


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

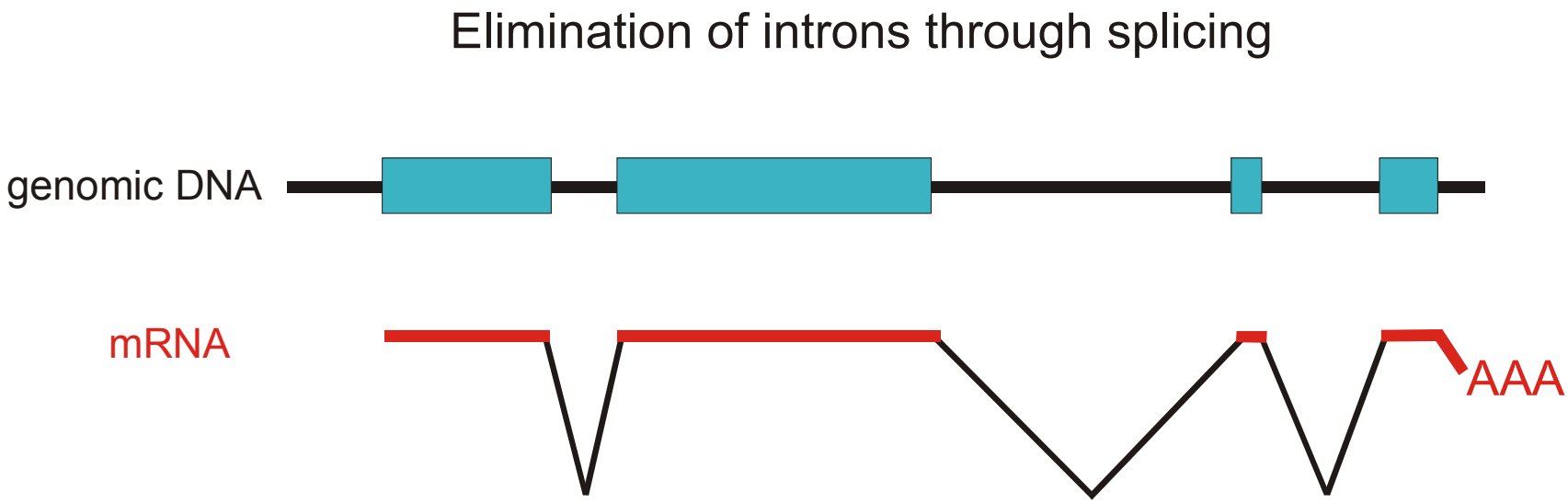
Large scale data processing, sequence comparison ...

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	40 000 - 60 000

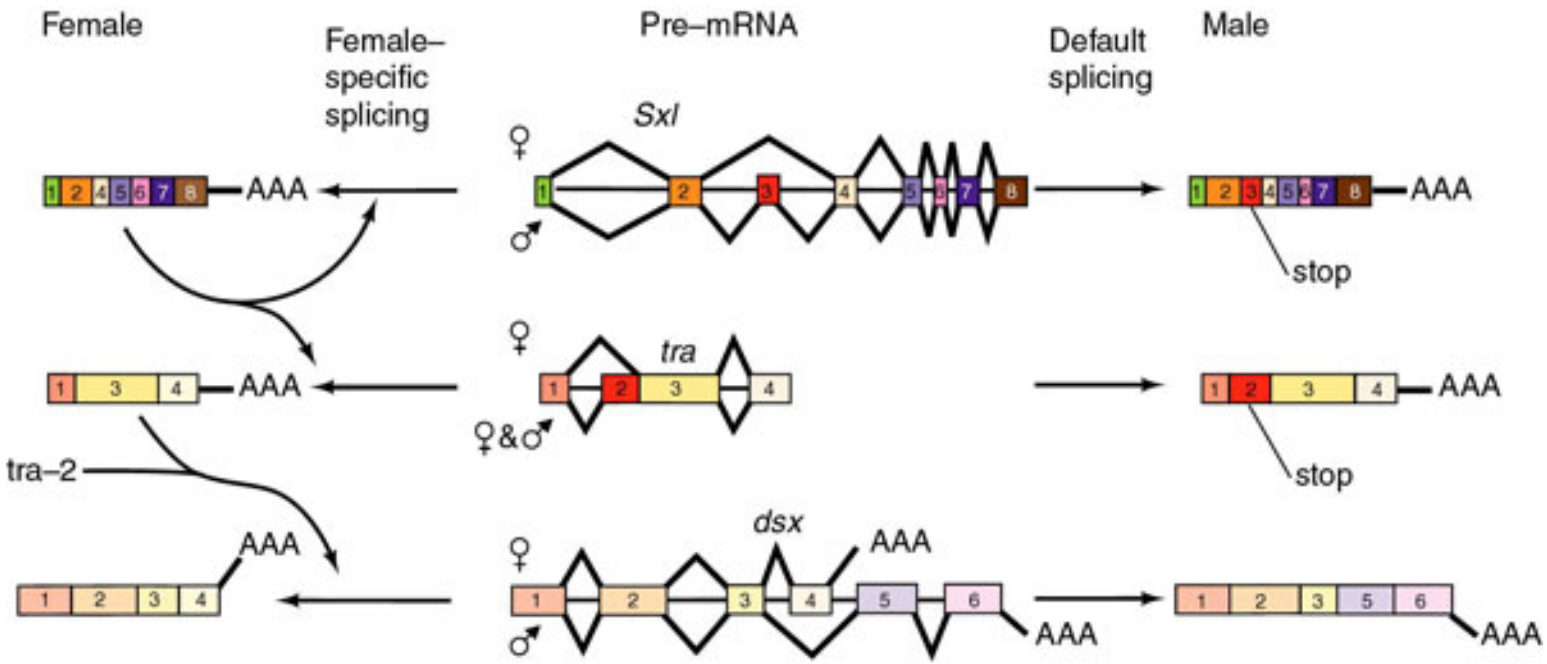


Number of genes in the human genome

The number of genes in the human genome is still only a very rough estimate

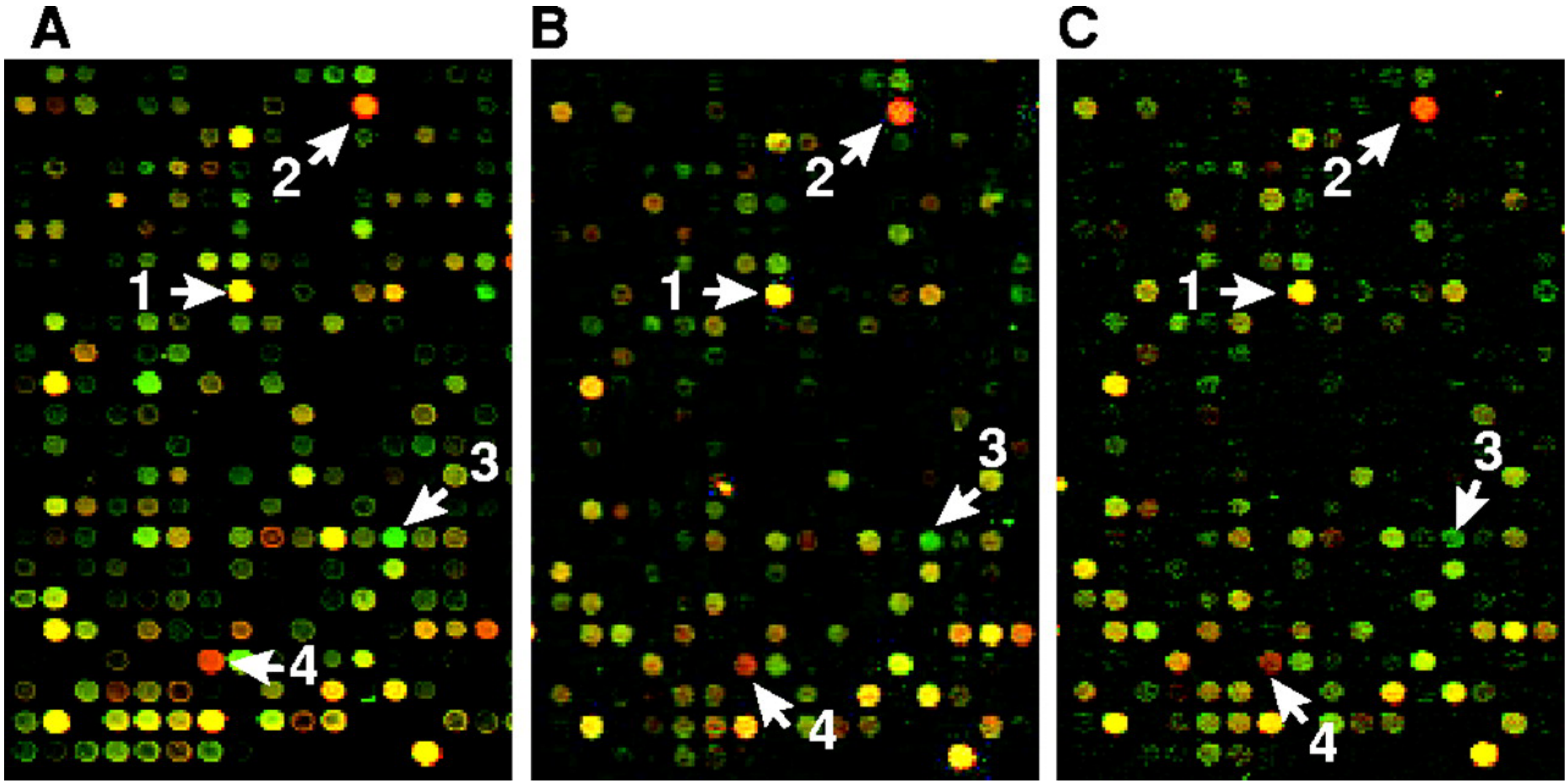


The gene is a stretch of DNA which after transcription and processing gives rise to a mRNA



Sex determination in ***Drosophila*** through alternative splicing

The process of protein synthesis and its regulation is now understood but the notion of the gene as a stretch of DNA has become obscure. The gene is essentially associated with the sequence of unmodified amino acids in a protein, and it is determined by the nucleotide sequence as well as the dynamics of the the process eventually leading to the m-RNA that is translated.



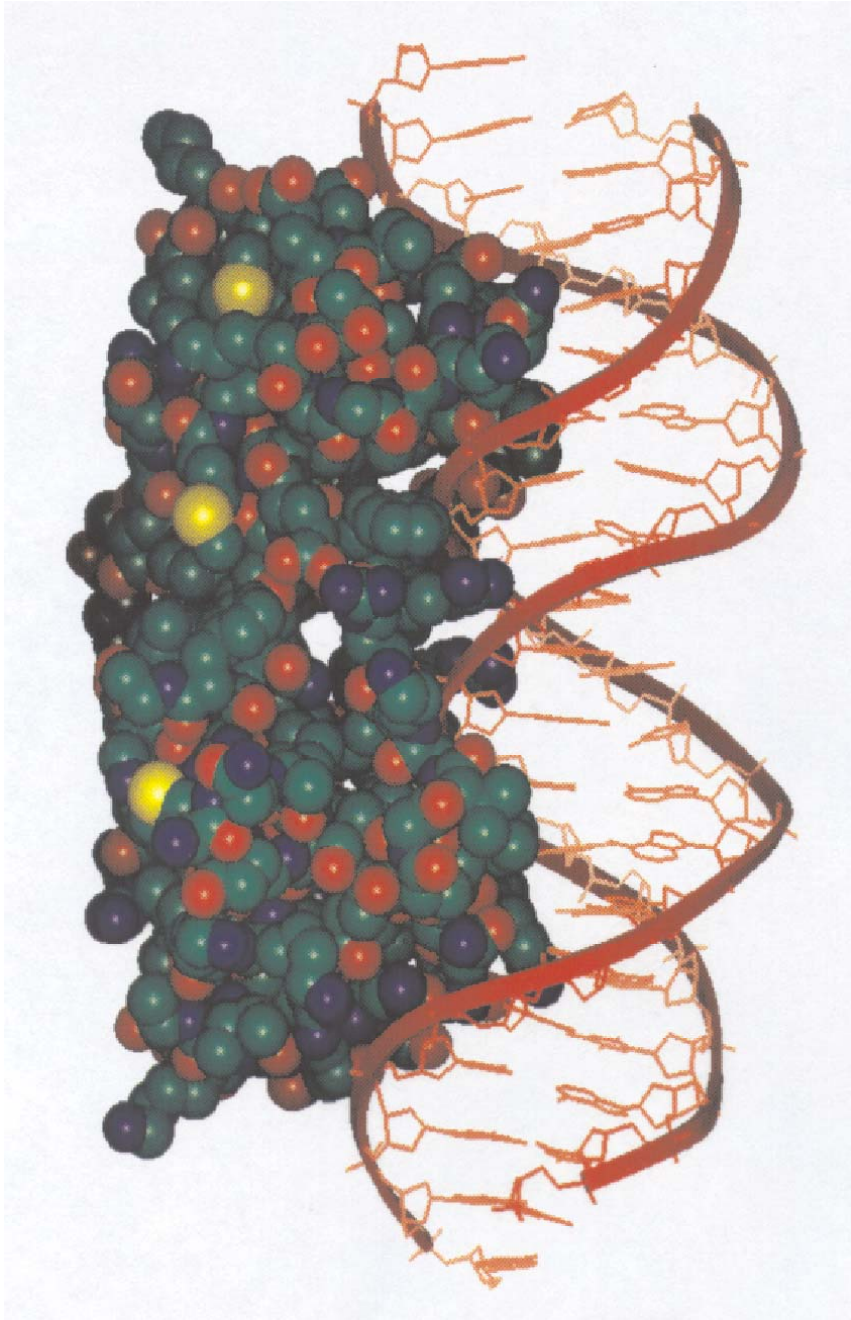
The same section of the microarray is shown in three independent hybridizations. Marked spots refer to: (1) protein disulfide isomerase related protein P5, (2) IL-8 precursor, (3) EST AA057170, and (4) vascular endothelial growth factor

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

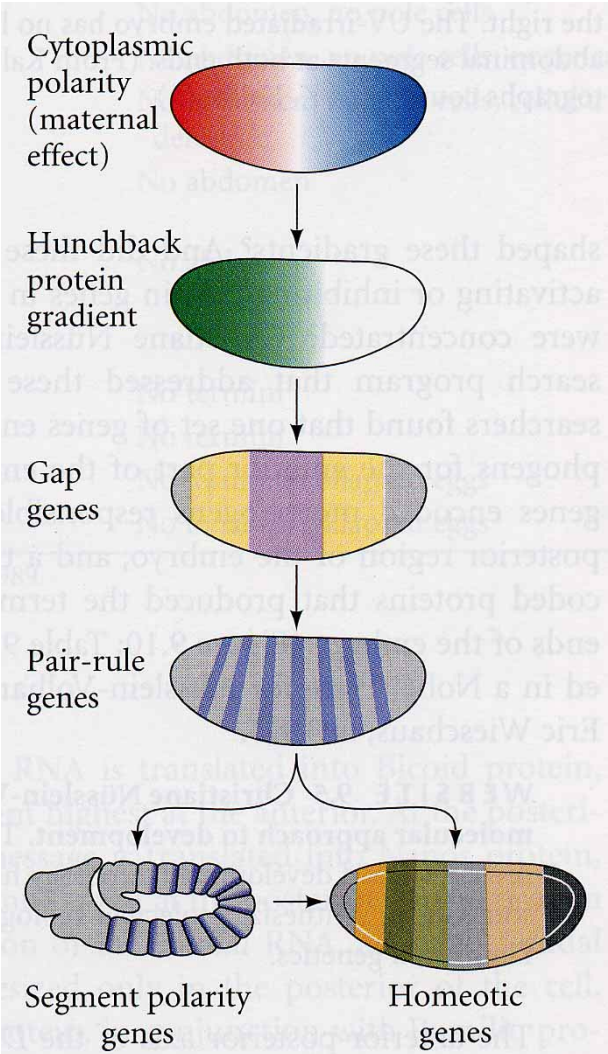
V.R.Iyer *et al*., *Science* **283**: 83-87, 1999

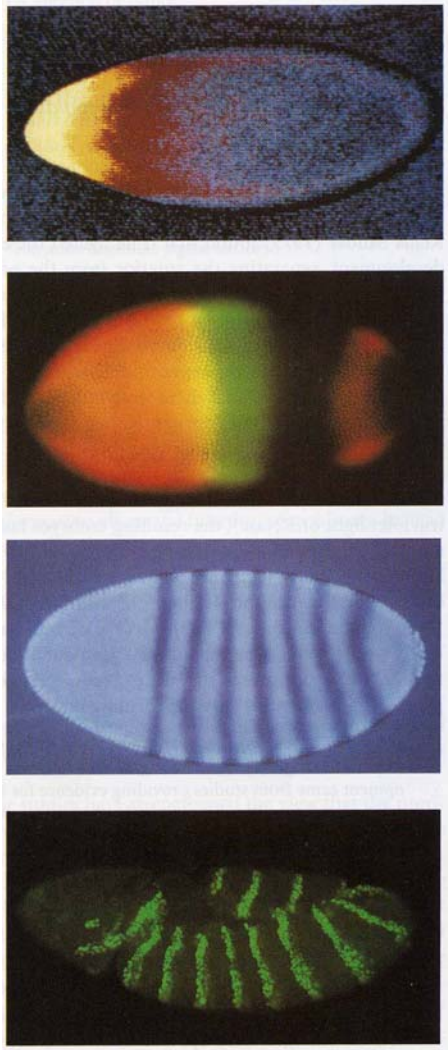
Developmental biology

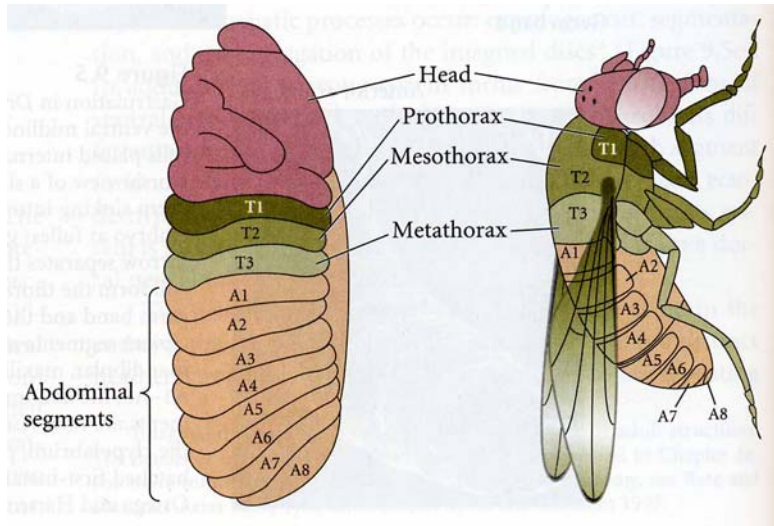
Gene regulation networks, signal propagation, pattern formation, robustness ...



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







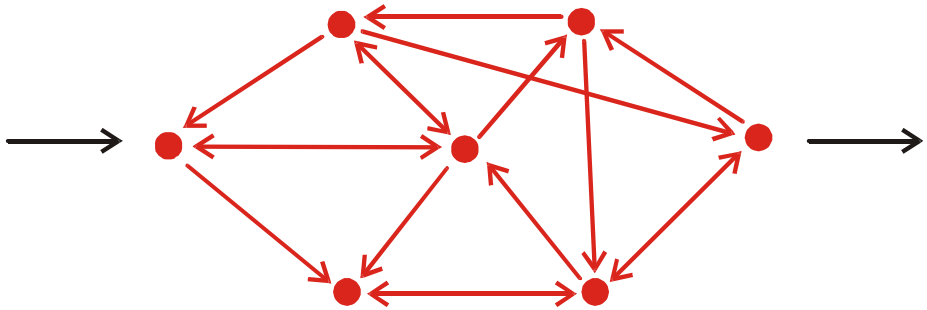
Cascades, A ñ B ñ C ñ ... , and networks of genetic control

Turing pattern resulting from reaction-diffusion equation ?

Intercelluar communication creating positional information

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

Linear chain



Network

Processing of information in cascades and networks

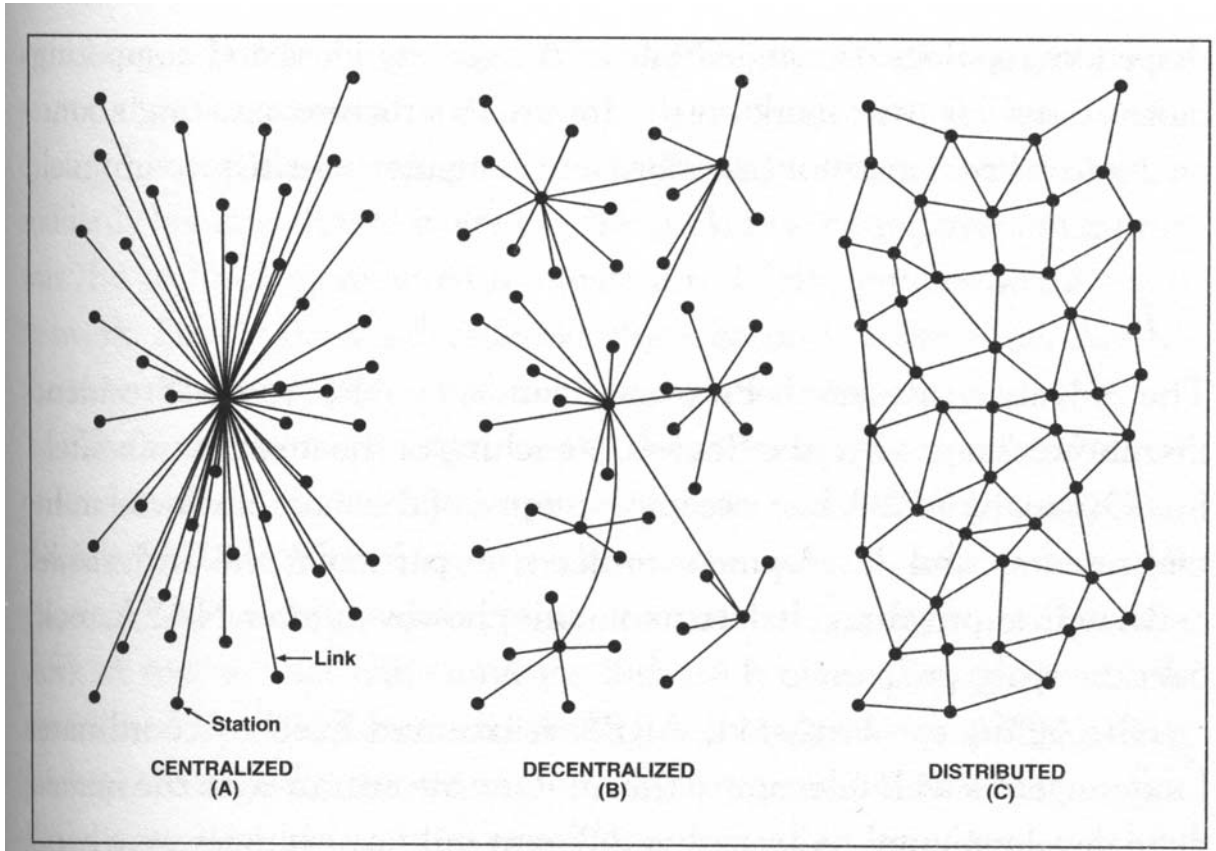


Figure 11.1 Paul Baran's Networks. *In 1964, Paul Baran began thinking about the optimal structure of the Internet. He suggested that there were three possible architectures for such a network—centralized, decentralized, and distributed—and warned that both the centralized and decentralized structures that dominated communications systems of the time were too vulnerable to attack. Instead, he proposed that the Internet should be designed to have a distributed, mesh-like architecture. (Reproduced with permission of Paul Baran.)*

Albert-László Barabási, Linked – The New Science of Networks. Perseus Publ., Cambridge, MA, 2002

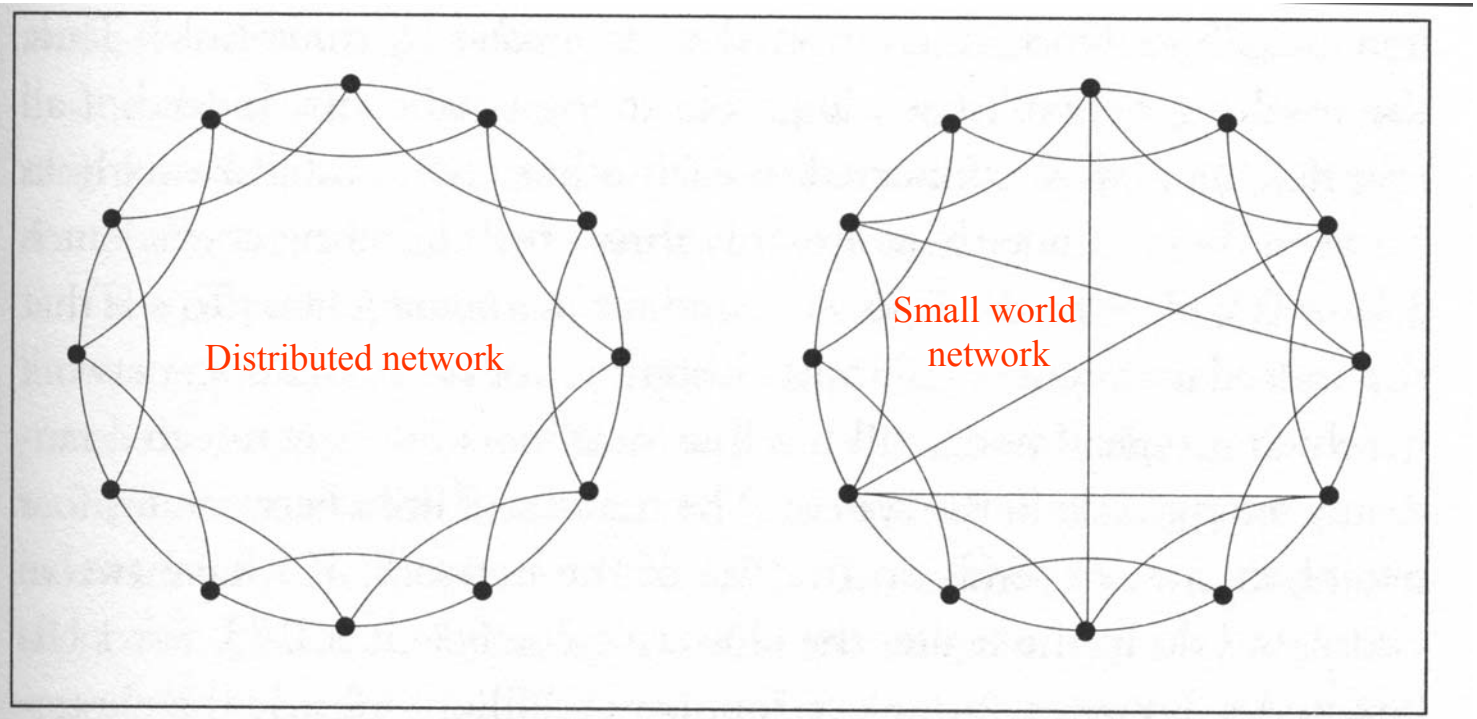


Figure 4.2 A Small and Clustered World. *To model networks with a high degree of clustering, Duncan Watts and Steven Strogatz started from a circle of nodes, where each node is connected to its immediate and next-nearest neighbors (left). To make this world a small one, a few extra links were added, connecting randomly selected nodes (right). These long-range links offer the crucial shortcuts between distant nodes, drastically shortening the average separation between all nodes.*

Albert-László Barabási, Linked – The New Science of Networks. Perseus Publ., Cambridge, MA, 2002

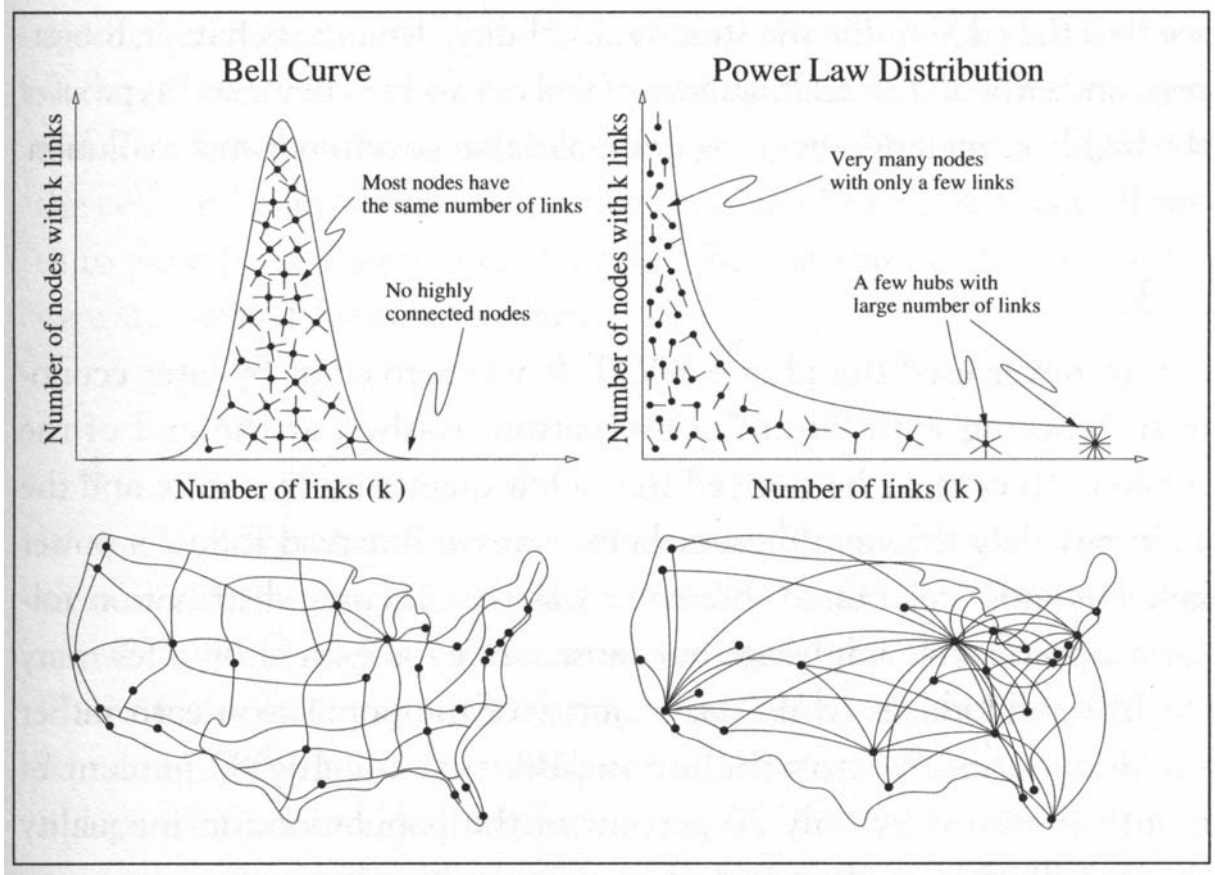
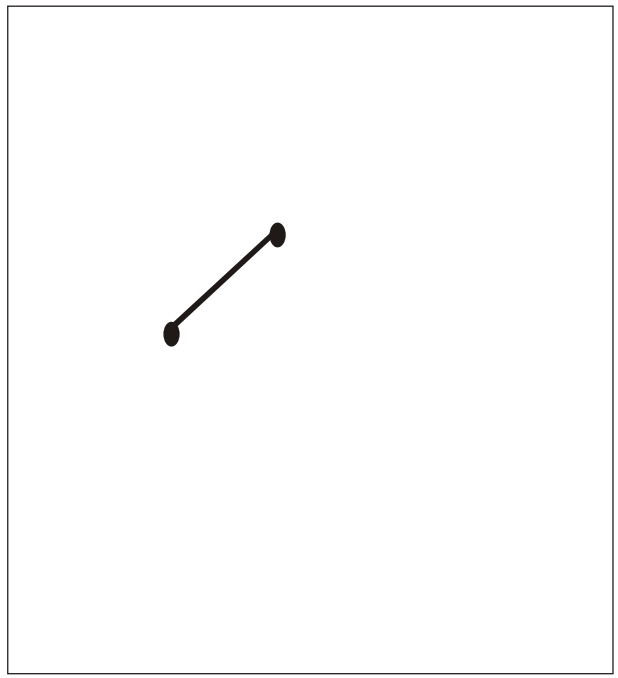
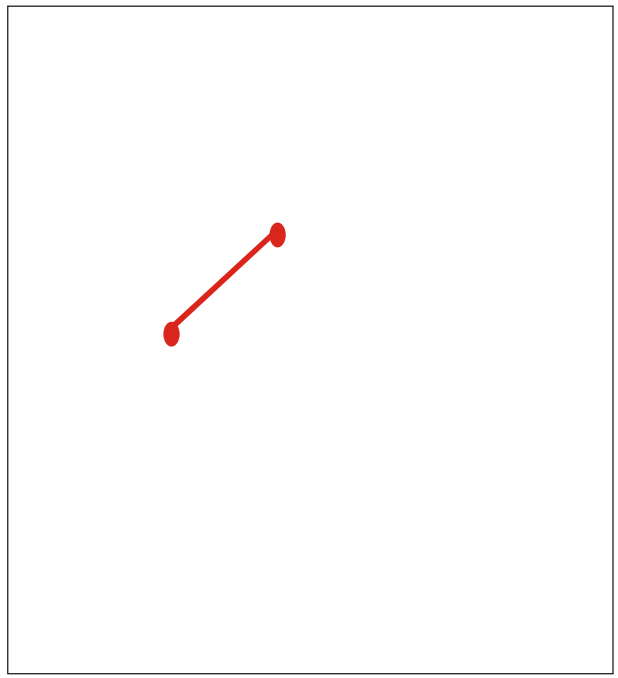


Figure 6.1 Random and Scale-Free Networks. *The degree distribution of a random network follows a bell curve, telling us that most nodes have the same number of links, and nodes with a very large number of links don't exist (top left). Thus a random network is similar to a national highway network, in which the nodes are the cities, and the links are the major highways connecting them. Indeed, most cities are served by roughly the same number of highways (bottom left). In contrast, the power law degree distribution of a scale-free network predicts that most nodes have only a few links, held together by a few highly connected hubs (top right). Visually this is very similar to the air traffic system, in which a large number of small airports are connected to each other via a few major hubs (bottom right).*

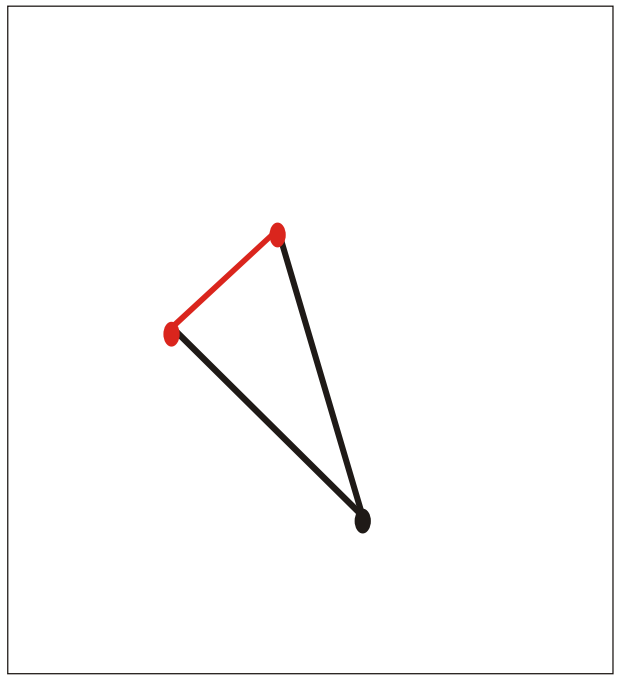
Albert-László Barabási,
Linked – The New Science of Networks
Perseus Publ., Cambridge, MA, 2002



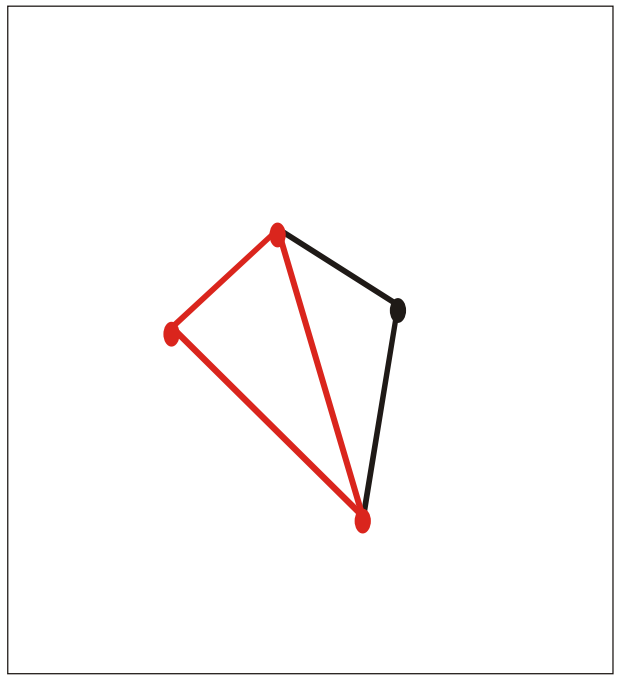
Formation of a scale-free network through evolutionary point by point expansion: Step 000



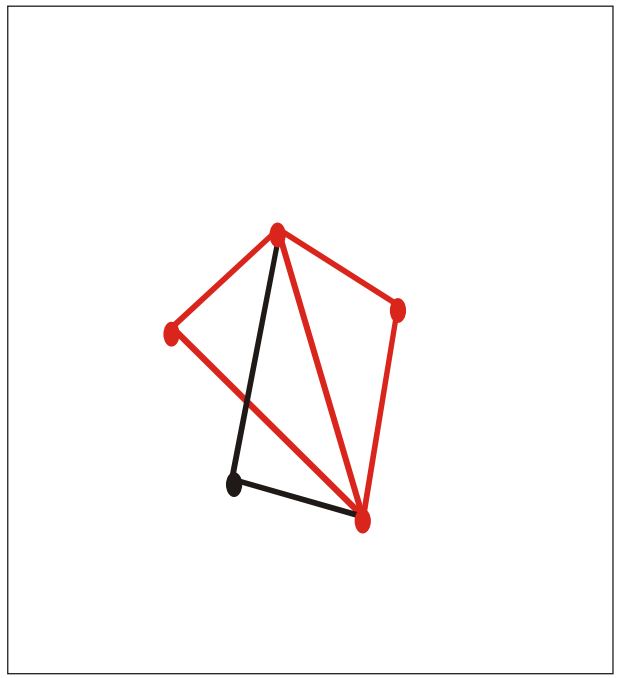
Formation of a scale-free network through evolutionary point by point expansion: Step 001



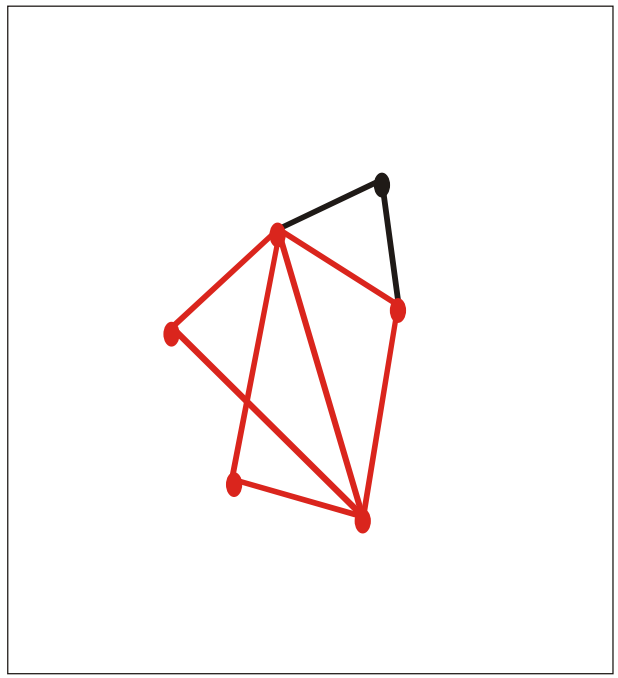
Formation of a scale-free network through evolutionary point by point expansion: Step 002



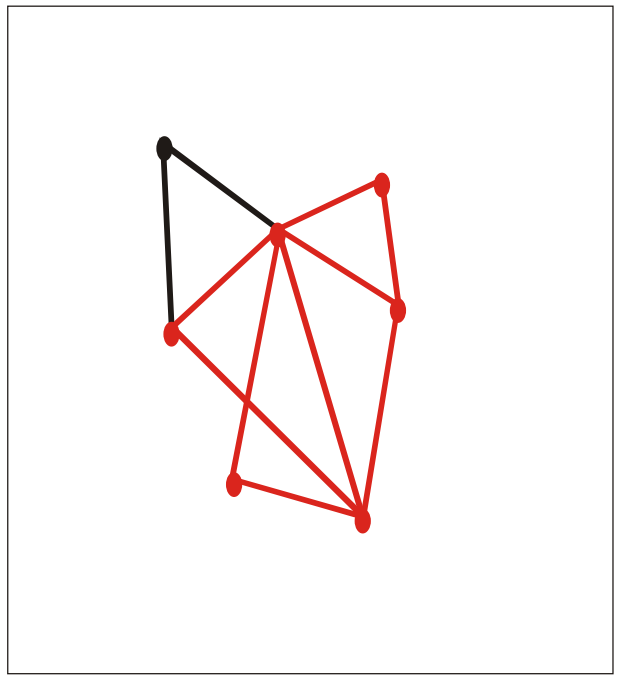
Formation of a scale-free network through evolutionary point by point expansion: Step 003



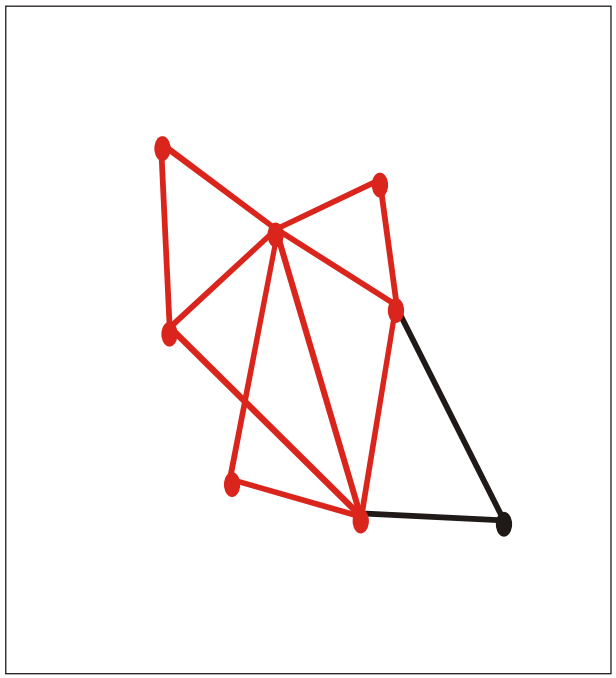
Formation of a scale-free network through evolutionary point by point expansion: Step 004



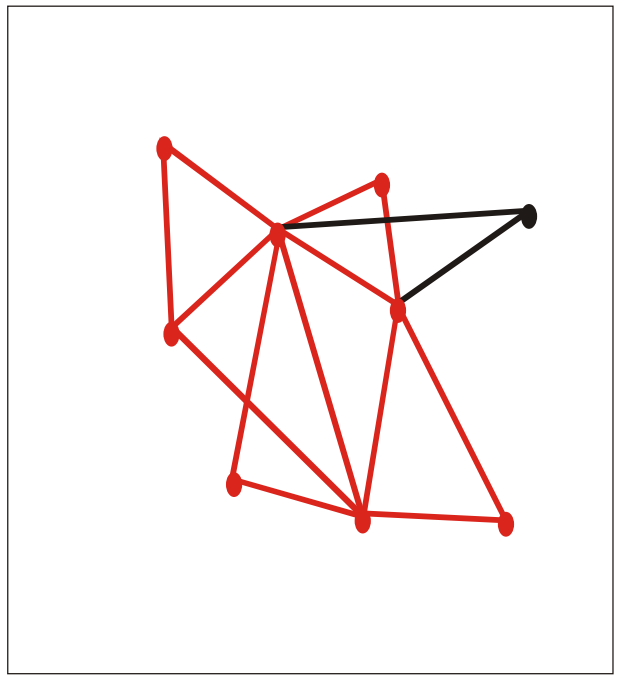
Formation of a scale-free network through evolutionary point by point expansion: Step 005



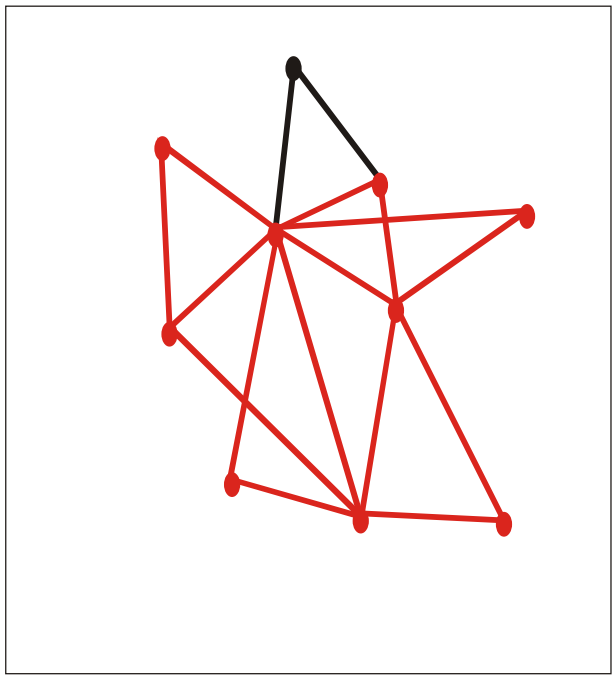
Formation of a scale-free network through evolutionary point by point expansion: Step 006



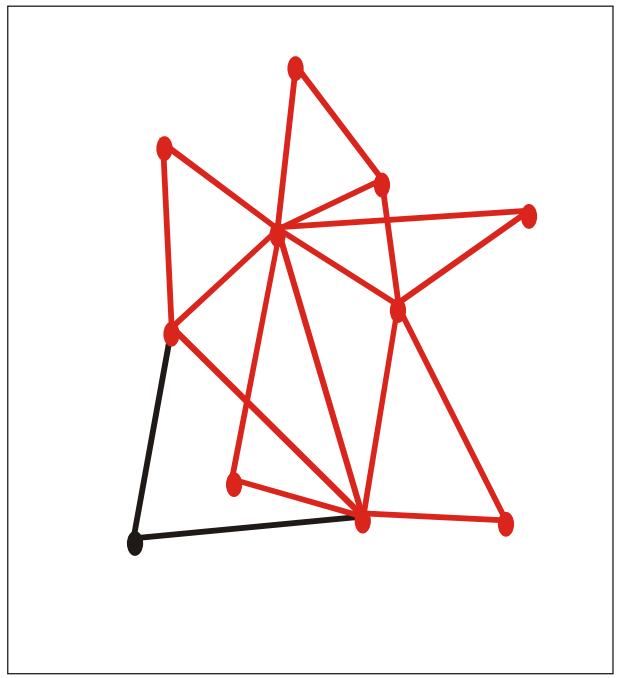
Formation of a scale-free network through evolutionary point by point expansion: Step 007



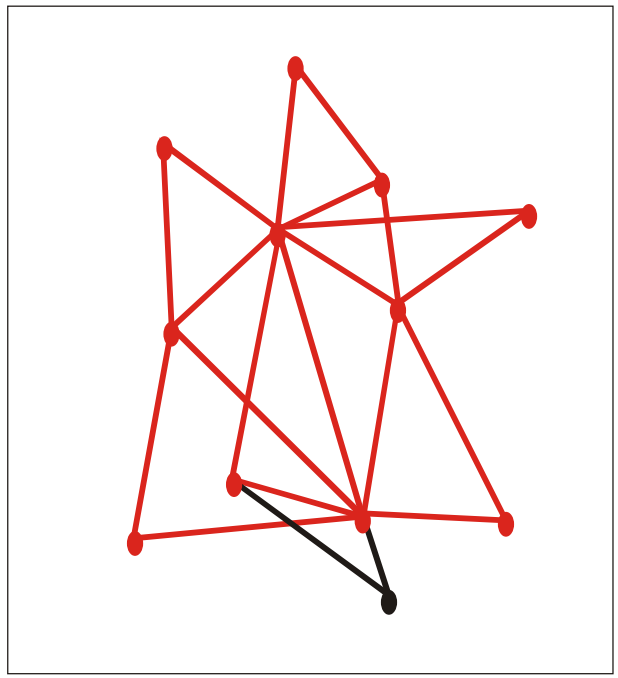
Formation of a scale-free network through evolutionary point by point expansion: Step 008



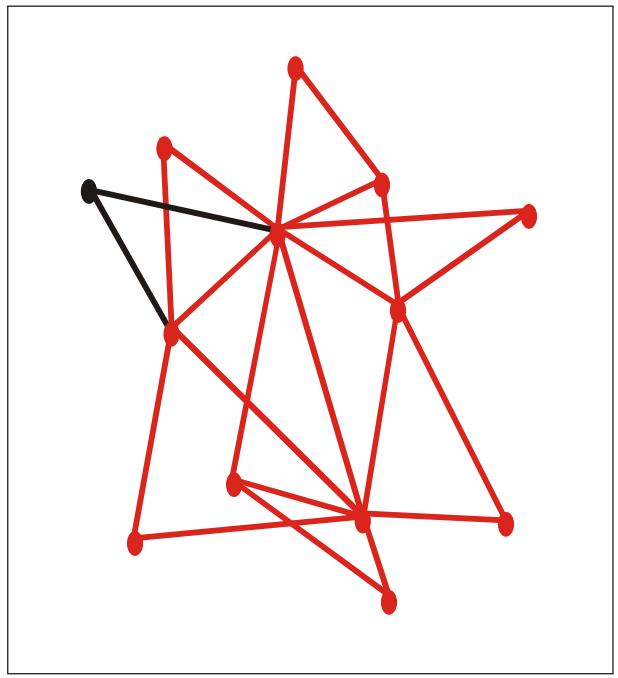
Formation of a scale-free network through evolutionary point by point expansion: Step 009



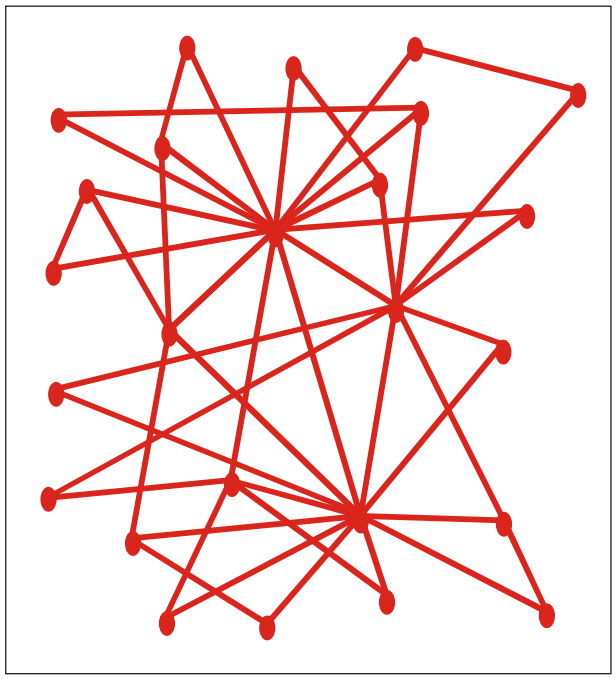
Formation of a scale-free network through evolutionary point by point expansion: Step 010



Formation of a scale-free network through evolutionary point by point expansion: Step 011

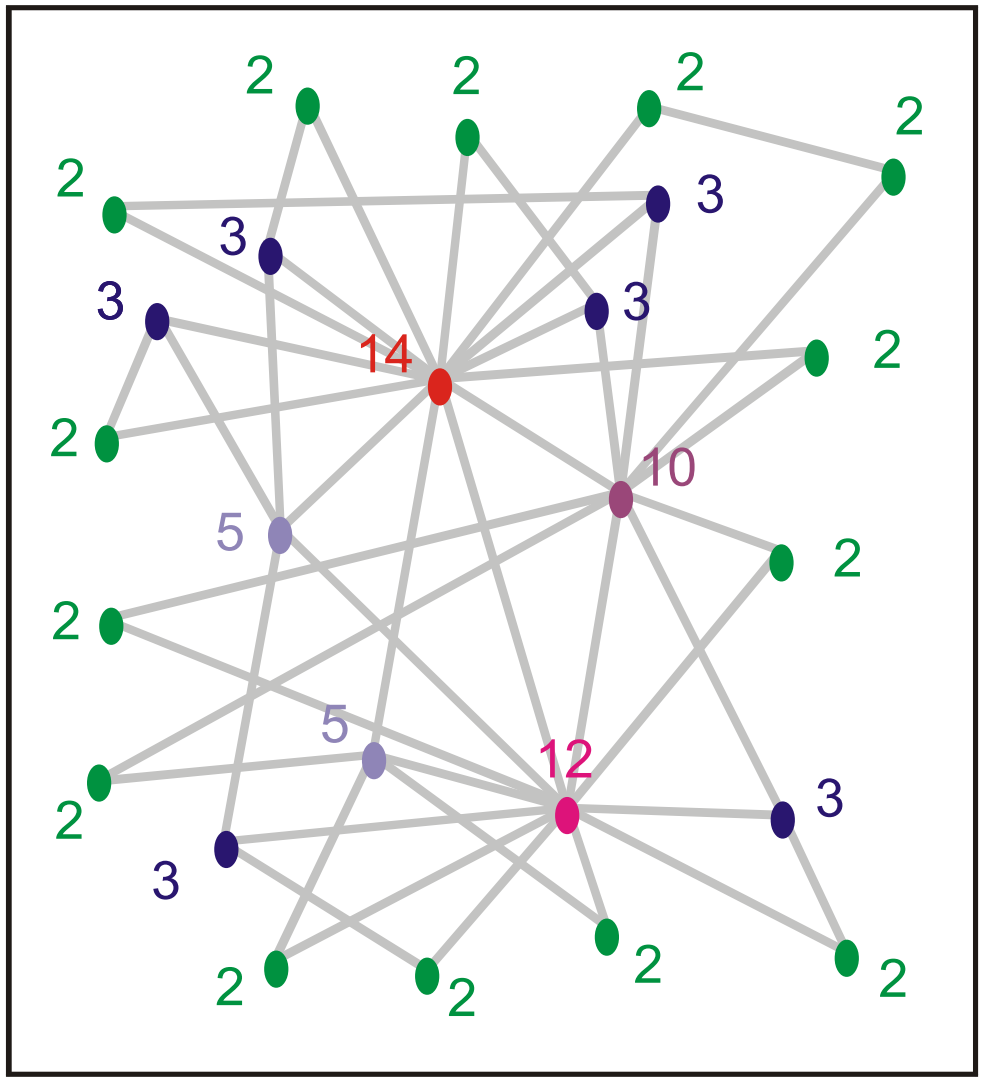


Formation of a scale-free network through evolutionary point by point expansion: Step 012



Formation of a scale-free network through evolutionary point by point expansion: Step 024

links	# nodes
2	14
3	6
5	2
10	1
12	1
14	1

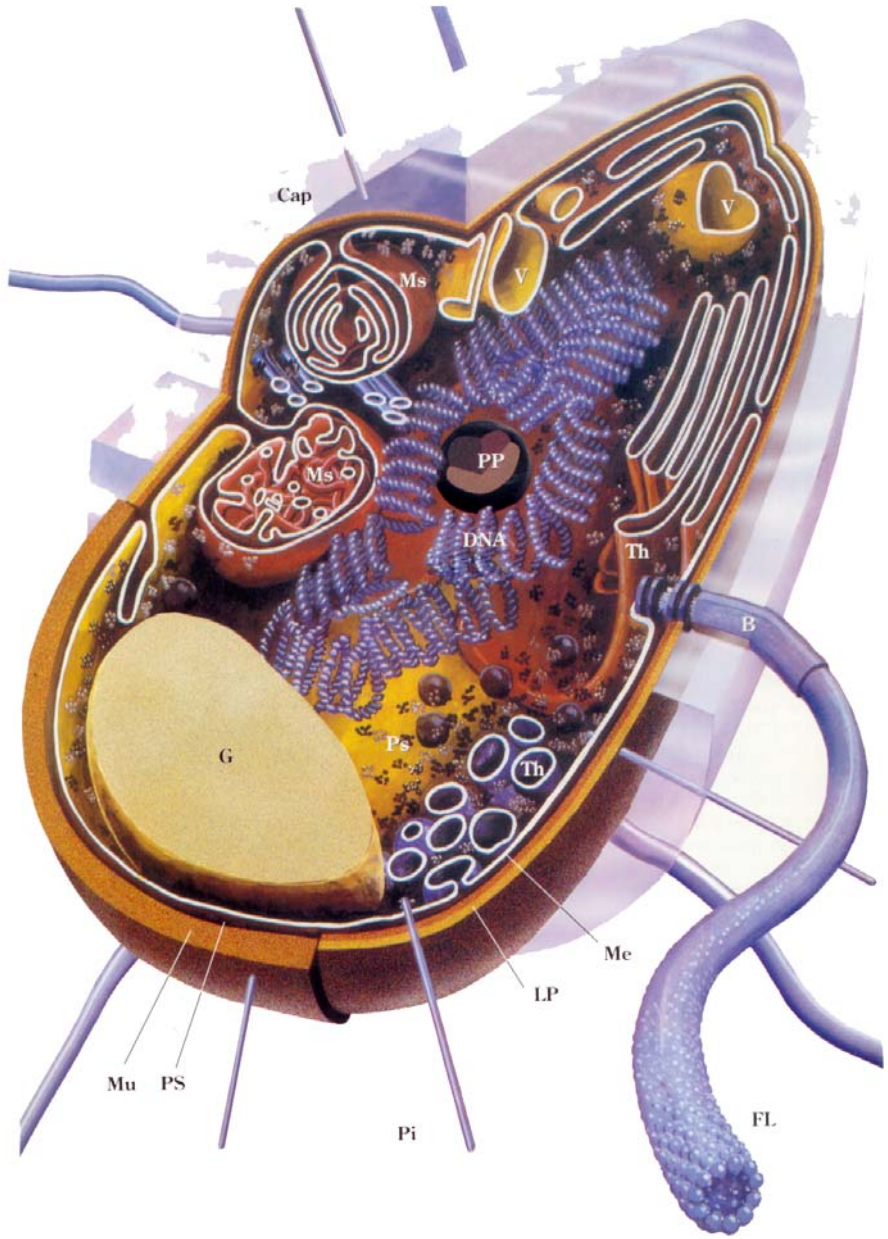


Analysis of nodes and links in a step by step evolved network

Cell biology

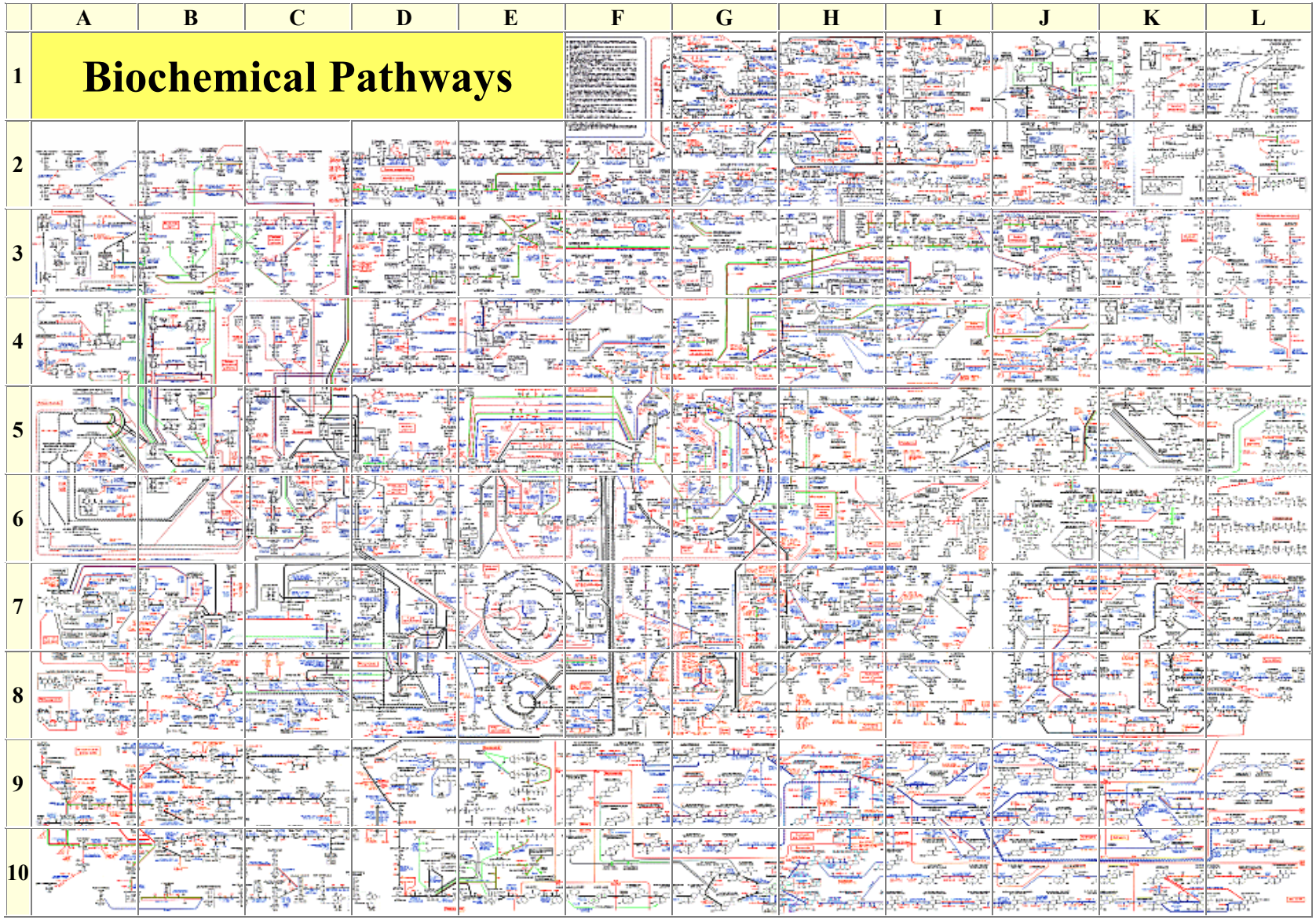
Regulation of cell cycle, metabolic networks, reaction kinetics, homeostasis, ...

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



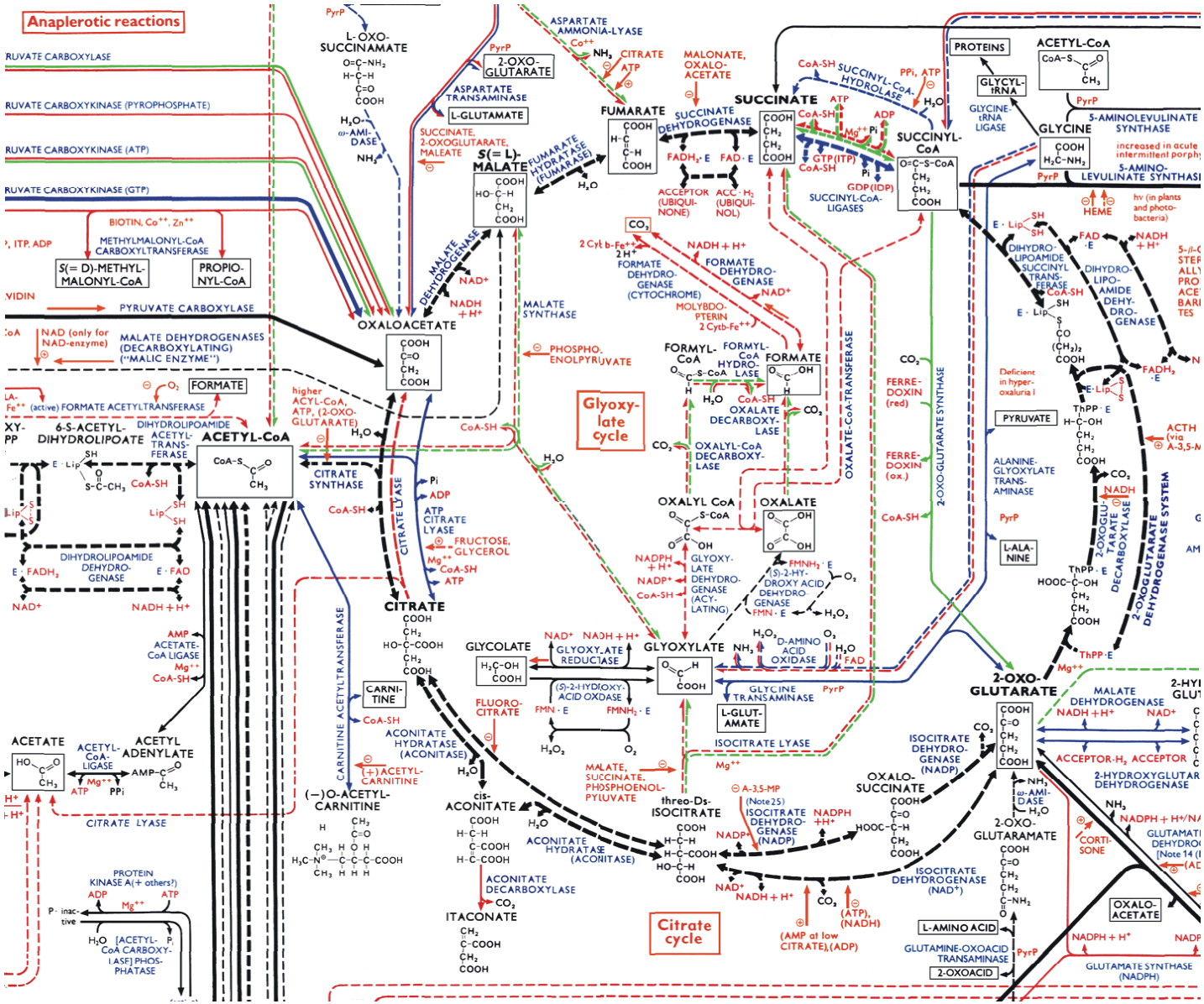
The human body:

10^{14} cells = 10^{13} eukaryotic cells +
Î 9 Ç 10^{13} bacterial (prokaryotic) cells;
Î 200 eukaryotic cell types



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

The citrate, tricarboxylic 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 , ...

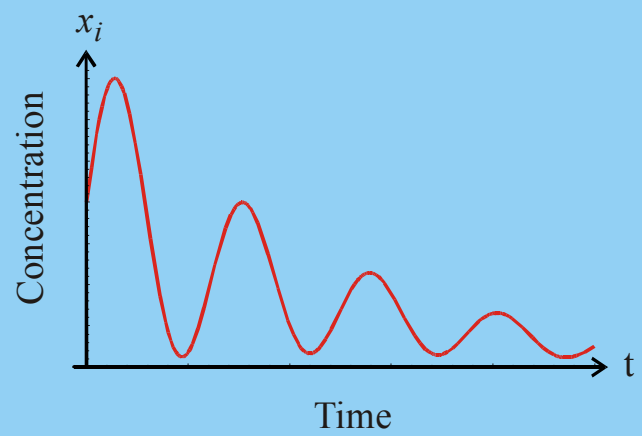
Initial conditions: $x_i(0);\ i = 1, 2, \ldots, n$

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

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

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

Reaction diffusion equations

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

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

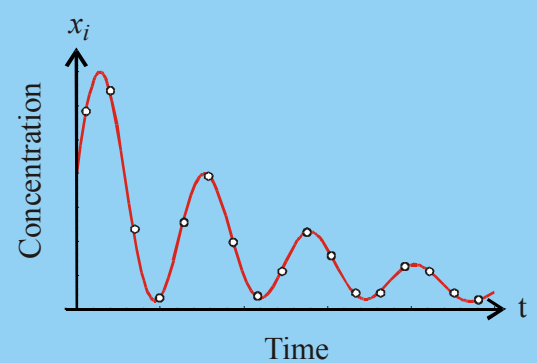
Neumann , $\frac{\partial x_i}{\partial u} = \hat{u} \cdot \nabla x_i^{\hat{S}} = f(\vec{r}, t); \; i = 1, 2, \ldots, n$

Parameter set

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

Data from measurements

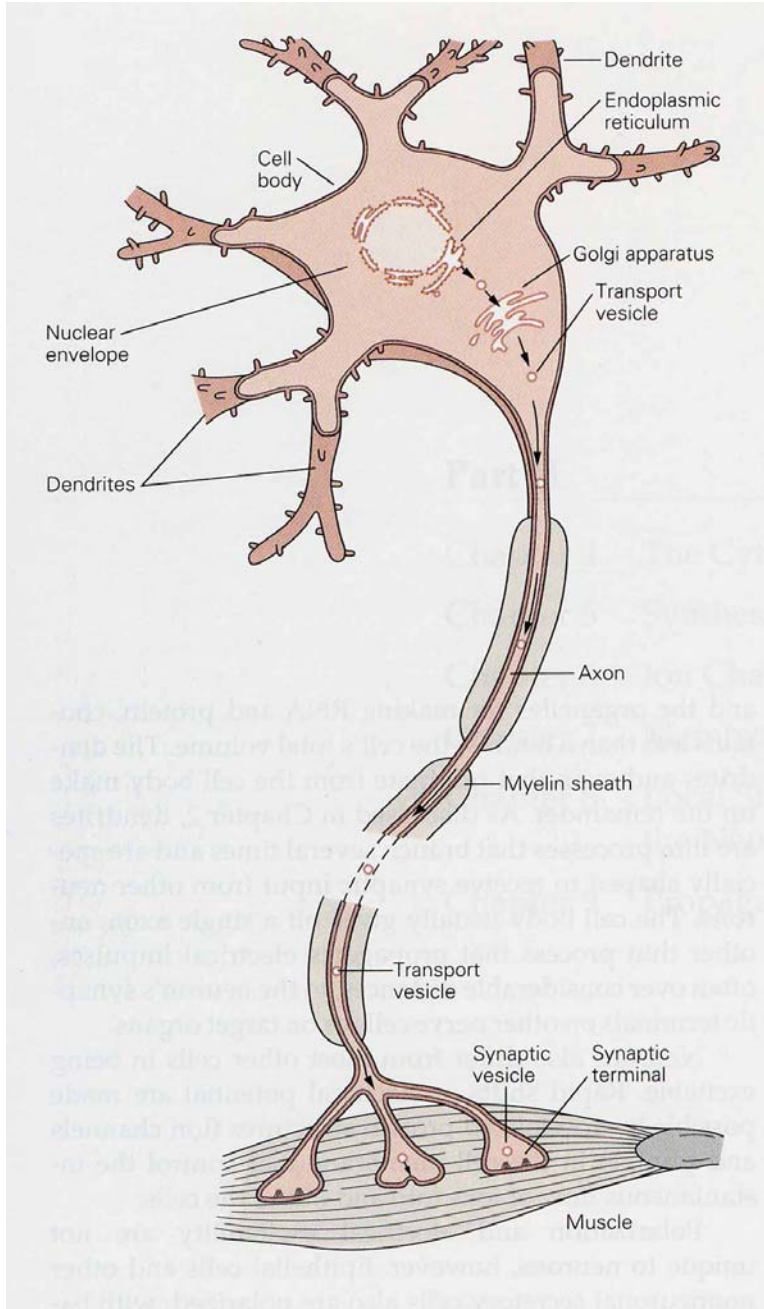
$$x_i(t_k); \; i = 1, 2, \ldots, n; \; k = 1, 2, \ldots, N$$



The inverse-problem of chemical reaction kinetics

Neurobiology

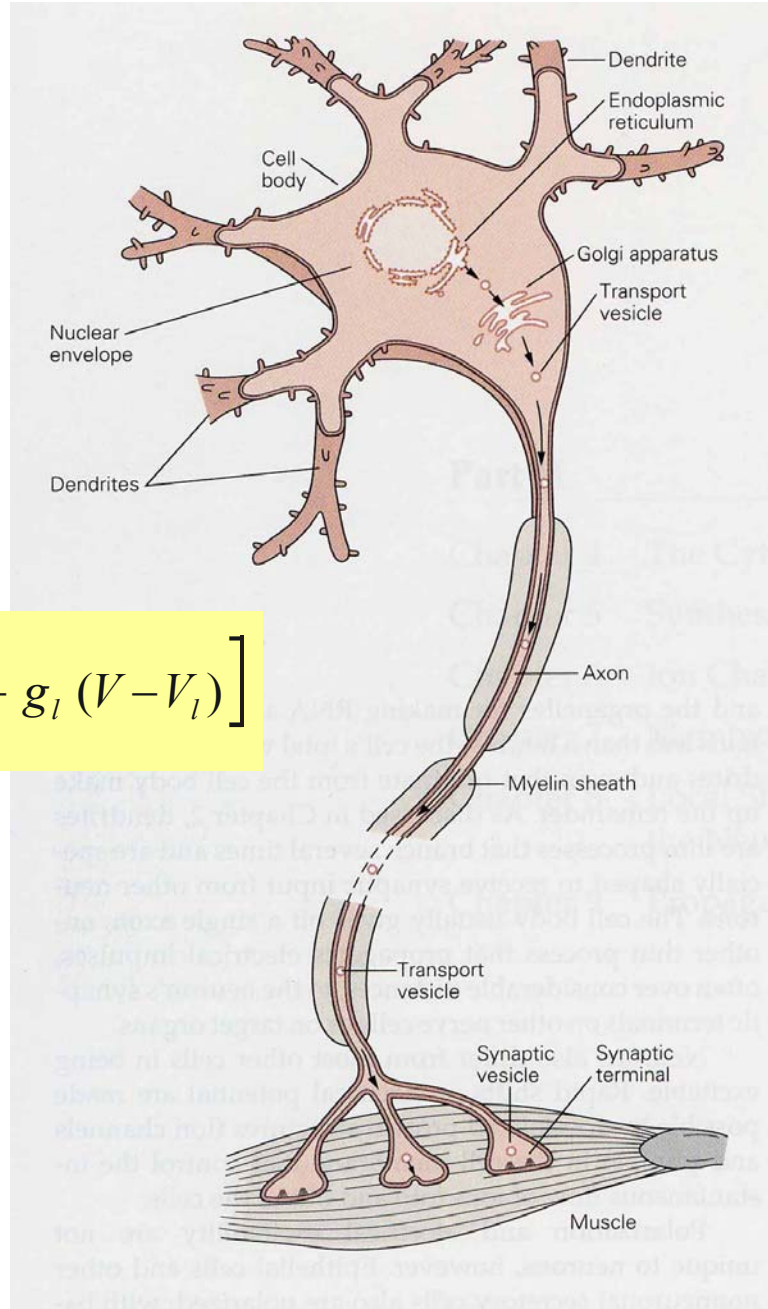
Neural networks, collective properties, nonlinear dynamics, signalling, ...



A single neuron signaling to a muscle fiber

Neurobiology

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



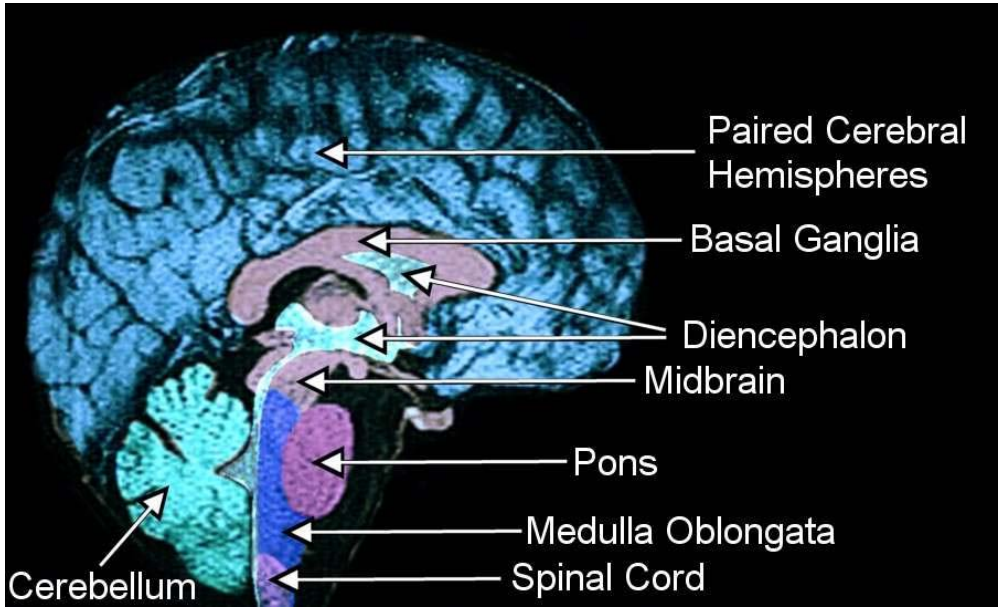
$$\frac{dV}{dt} = \frac{1}{C_M}\left[I - g_{Na}\, m^3\, h\,(V - V_{Na}) - g_K\, n^4\,(V - V_K) - g_l\,(V - V_l)\right]$$

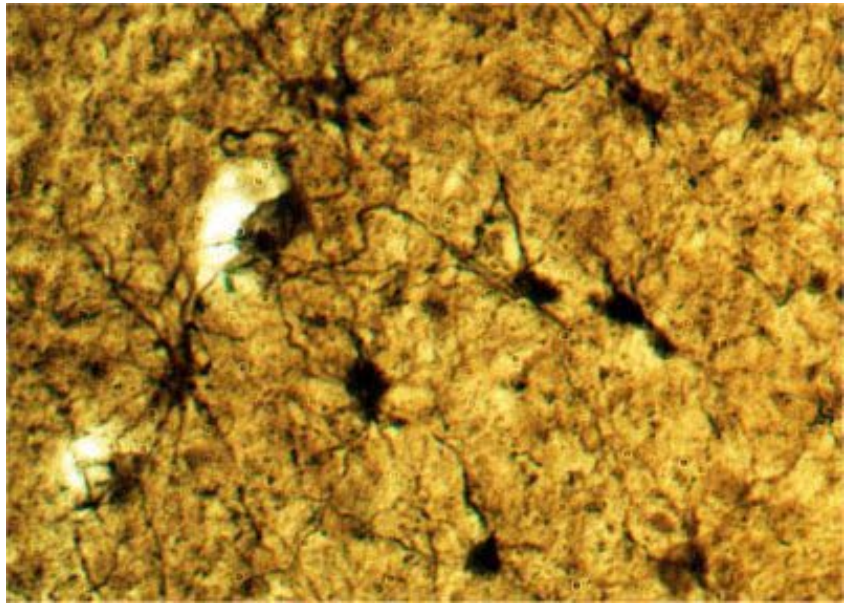
$$\frac{dm}{dt} = \alpha_m\,(1-m) - \beta_m\, m$$

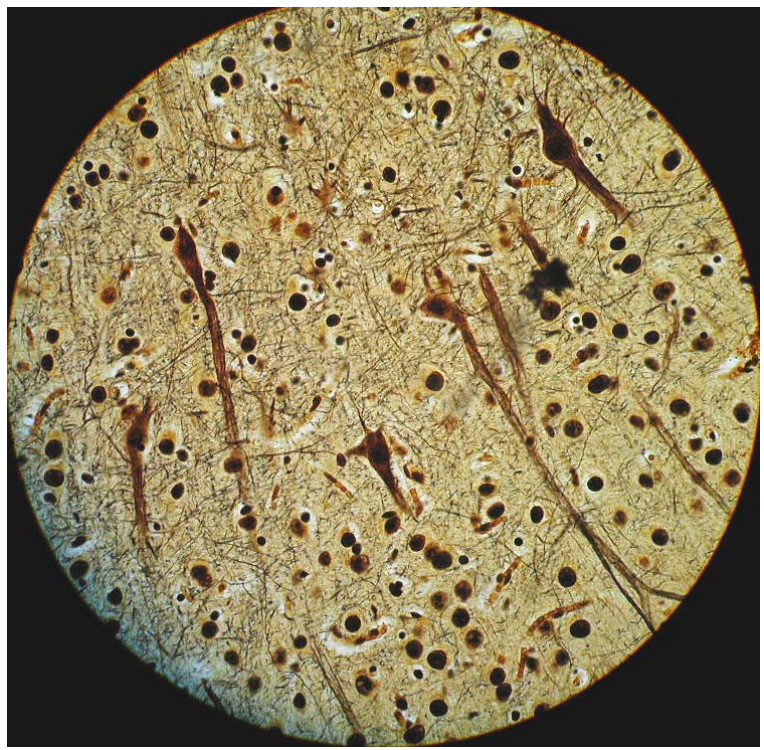
$$\frac{dh}{dt} = \alpha_h\,(1-h) - \beta_h\, h$$

$$\frac{dn}{dt} = \alpha_n\,(1-n) - \beta_n\, n$$

Hogdkin-Huxley OD equations







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

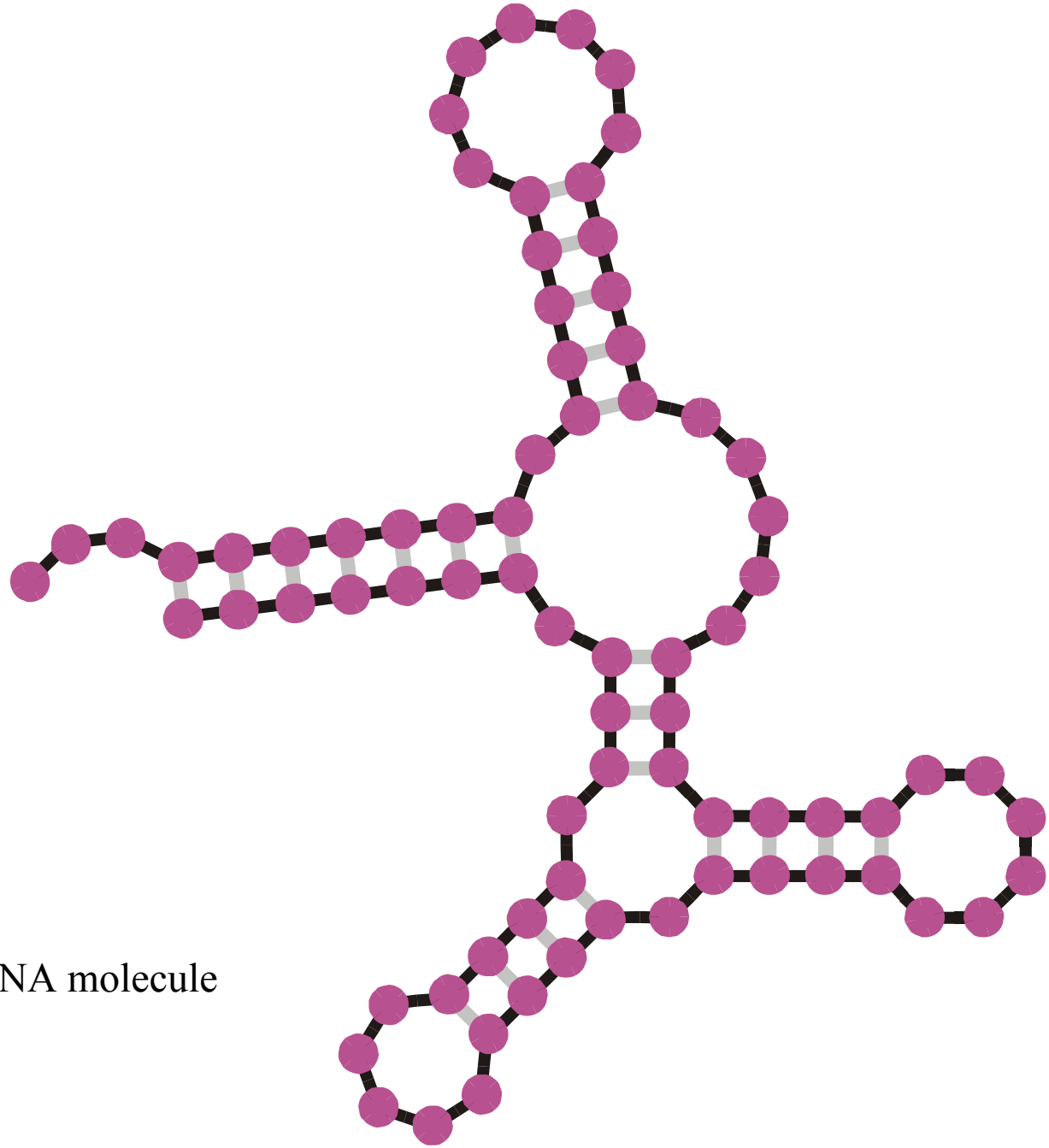
Time scales of evolutionary change

1. Komplexität in der Biologie

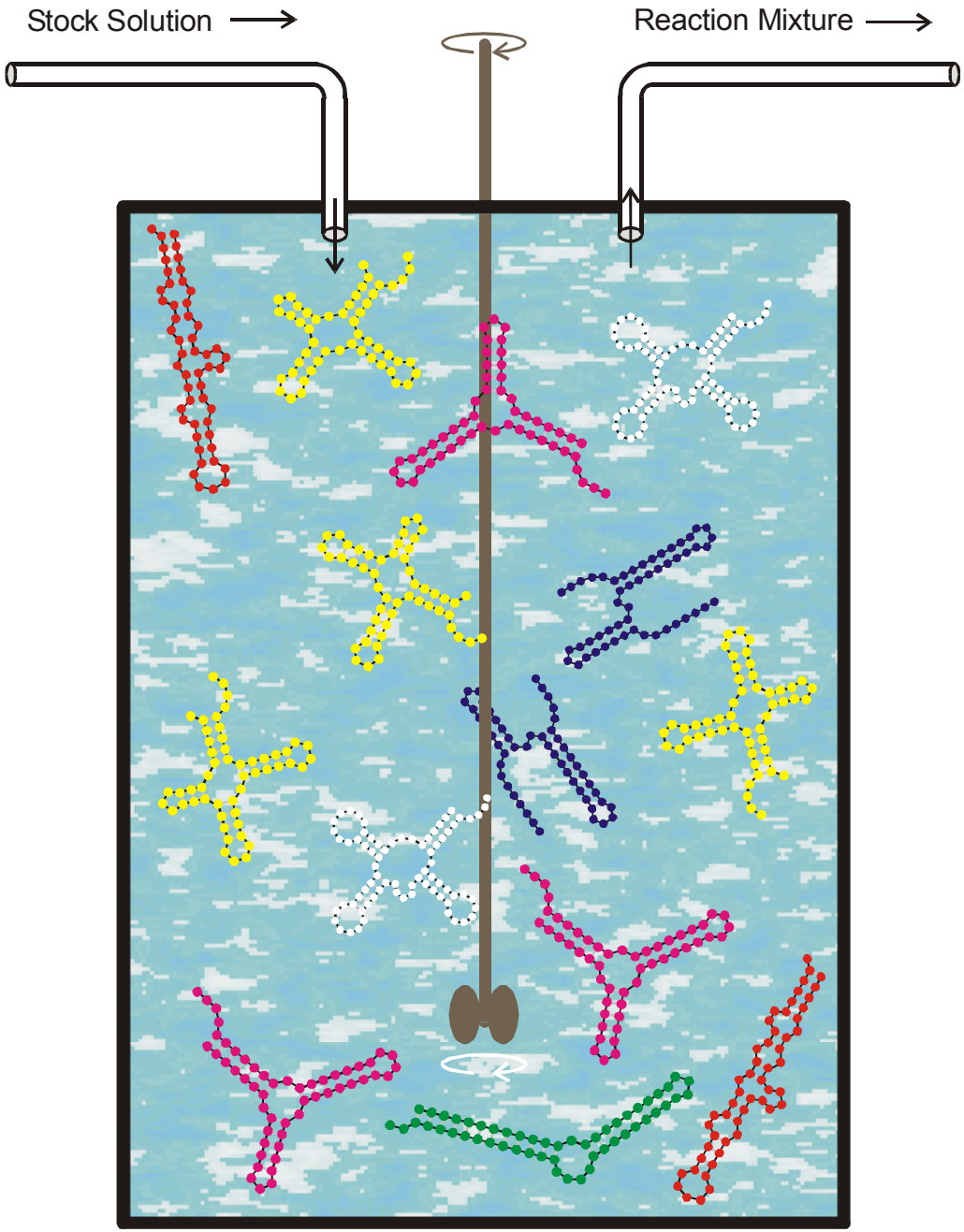
2. Evolutionäre Optimierung und Lernen im Ensemble

3. Strukturbildung von Biomolekülen als kombinatorisches Problem

4. Modellbildung in der Neurobiologie



Element of class 1: The RNA molecule



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

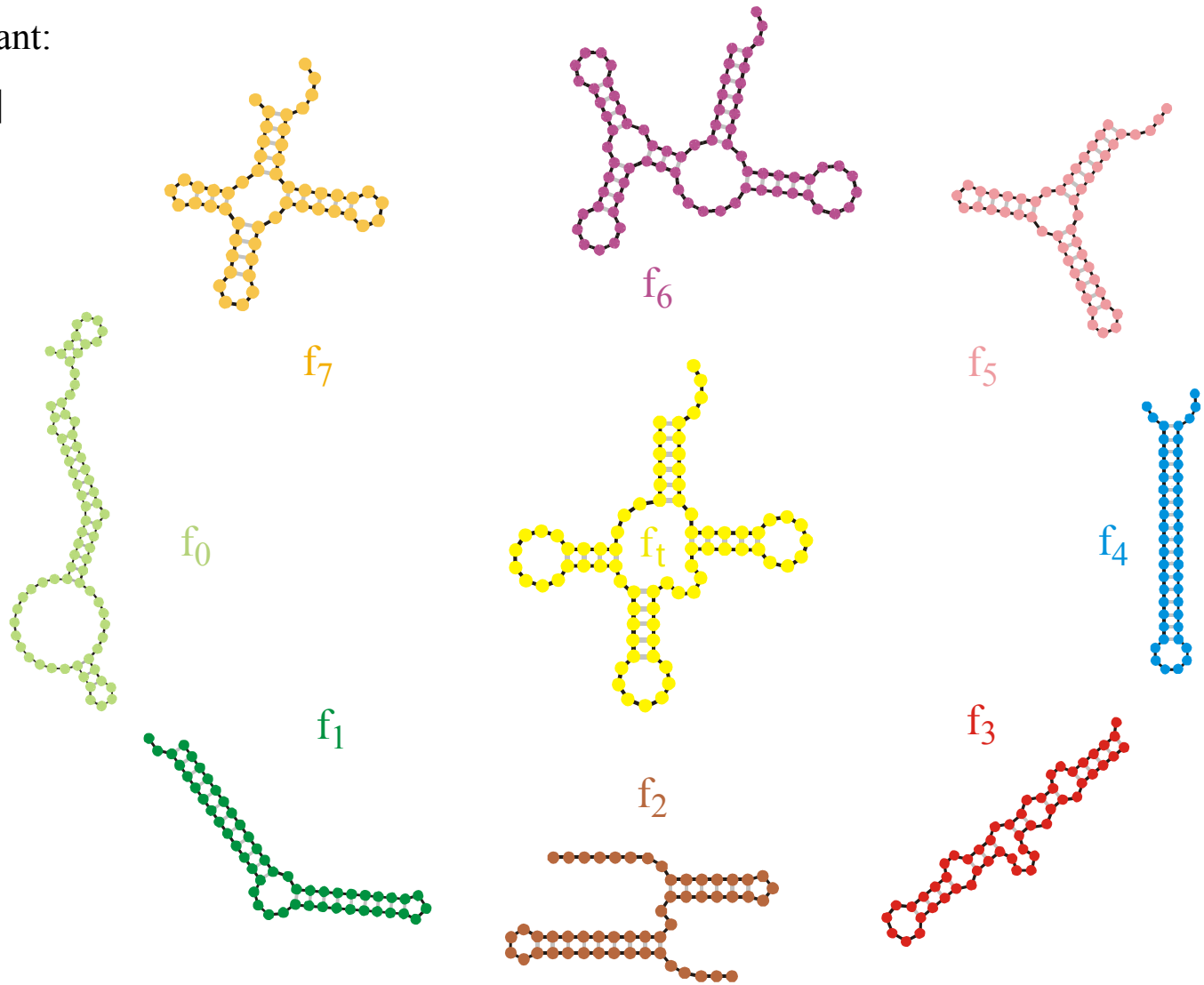
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

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

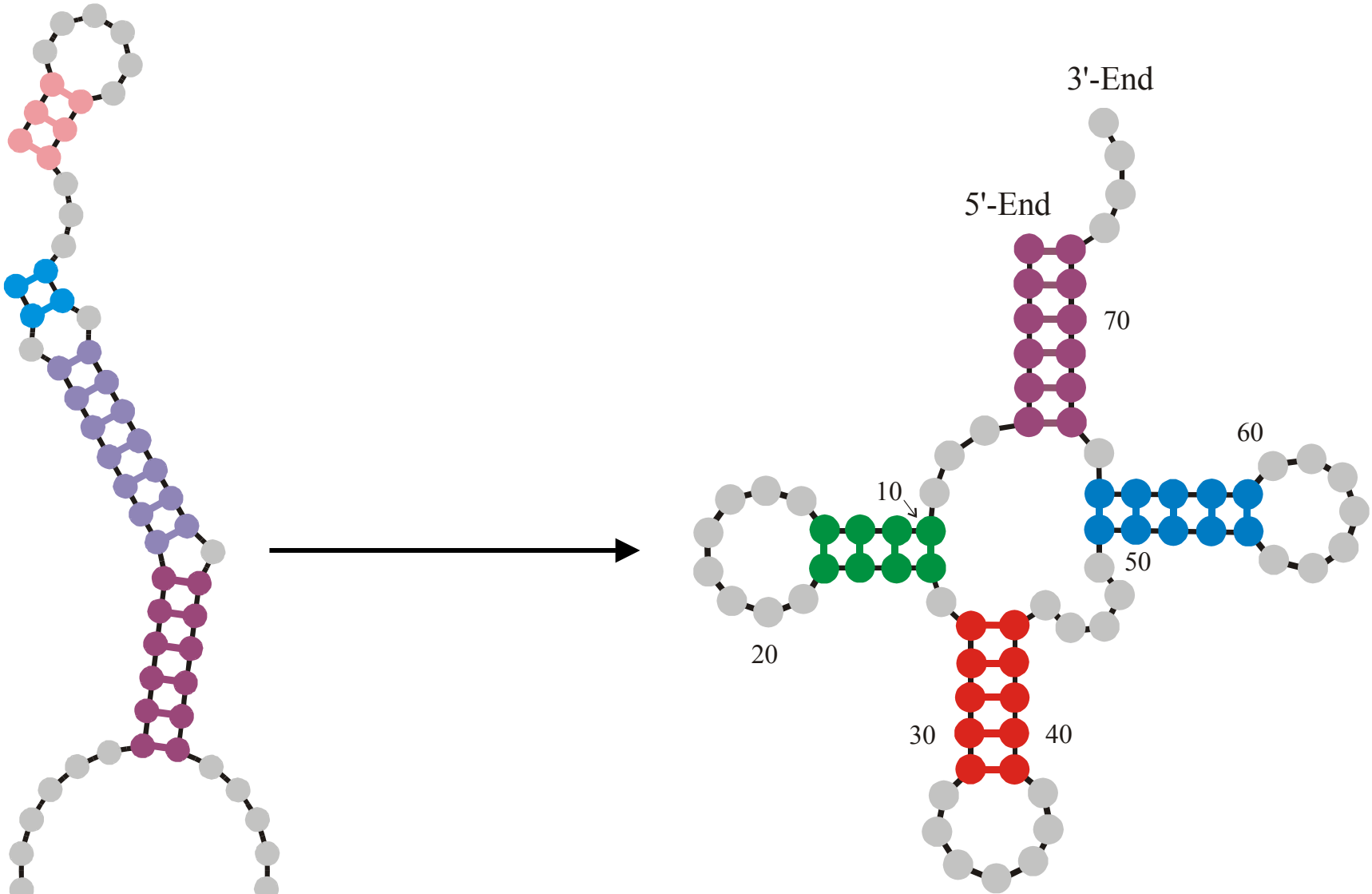
Replication rate constant:

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

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

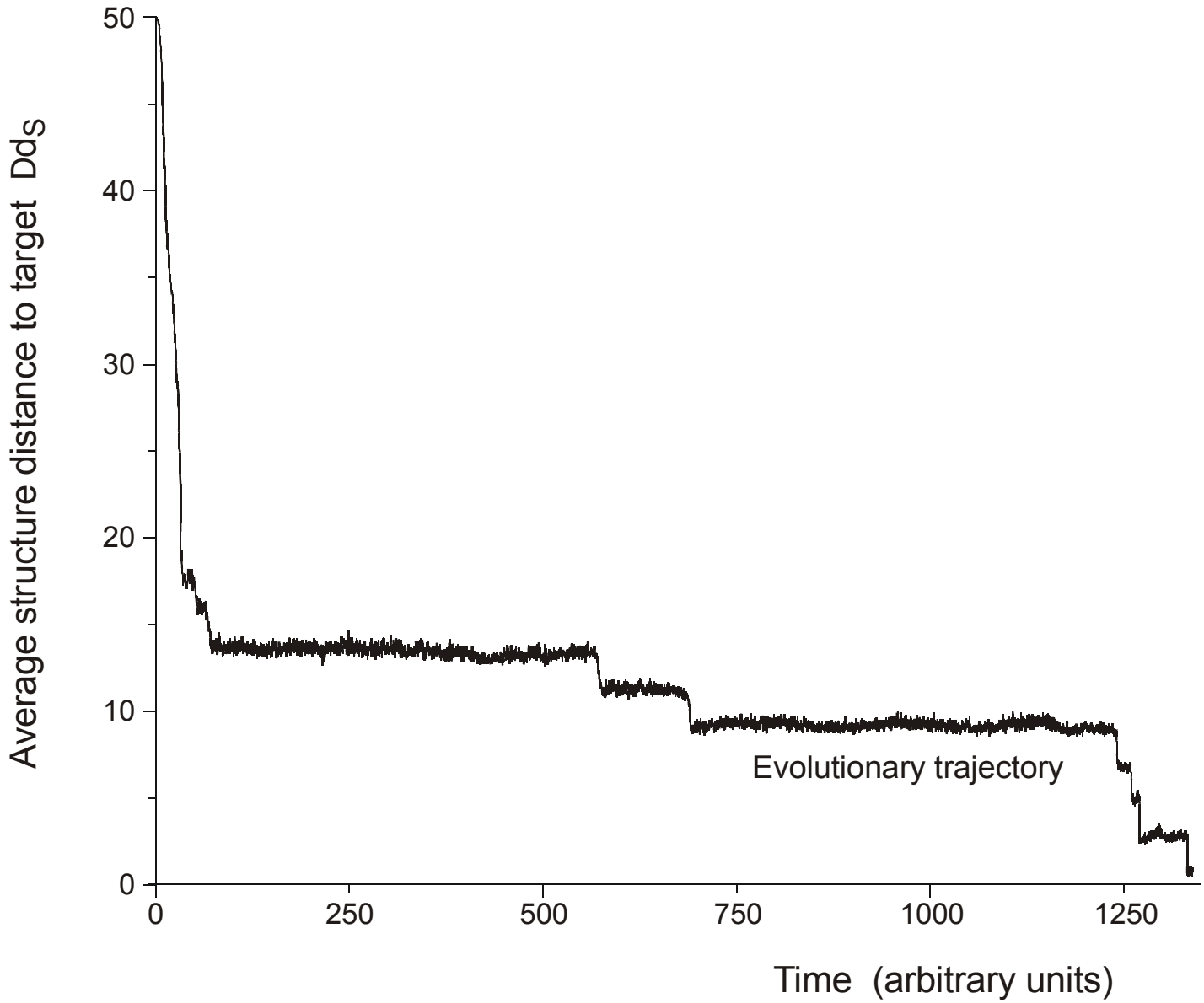


Evaluation of RNA secondary structures yields replication rate constants

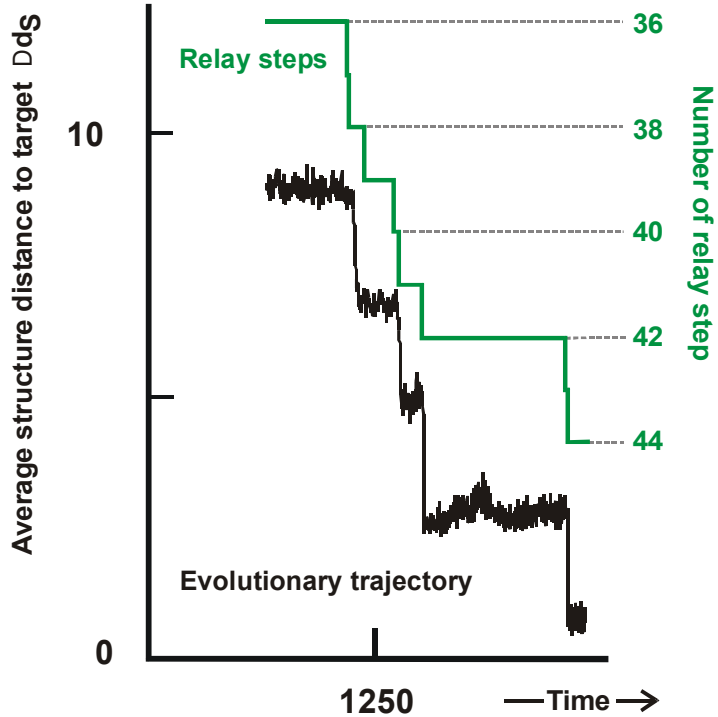


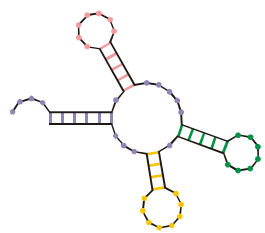
Randomly chosen initial structure

Phenylalanyl-tRNA as target structure



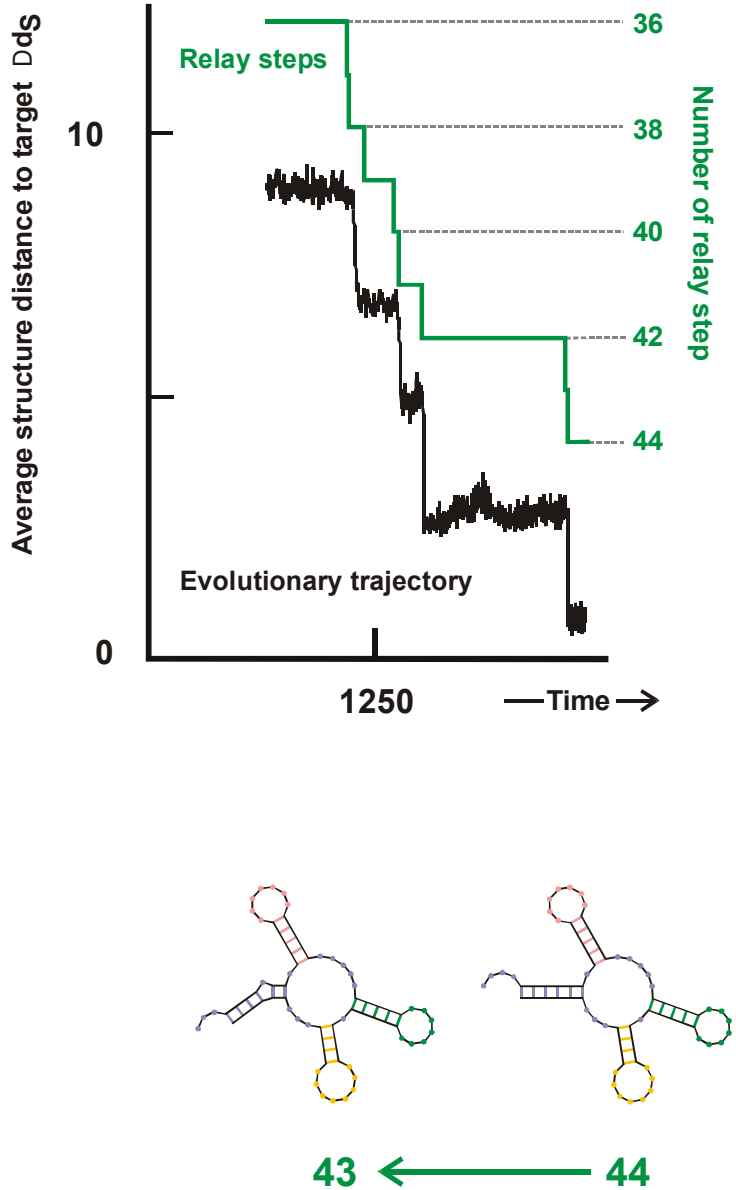
In silico optimization in the flow reactor: Trajectory (**physicists‘ view**)



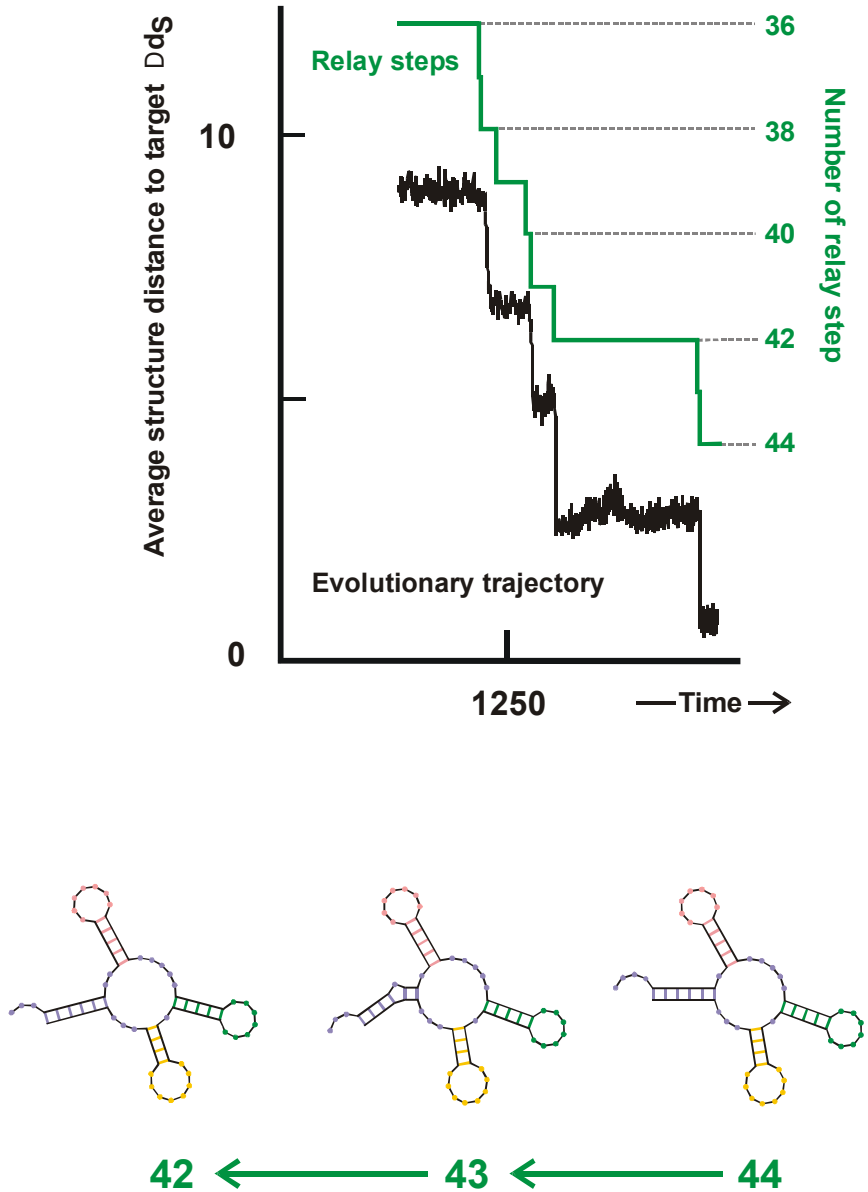


44

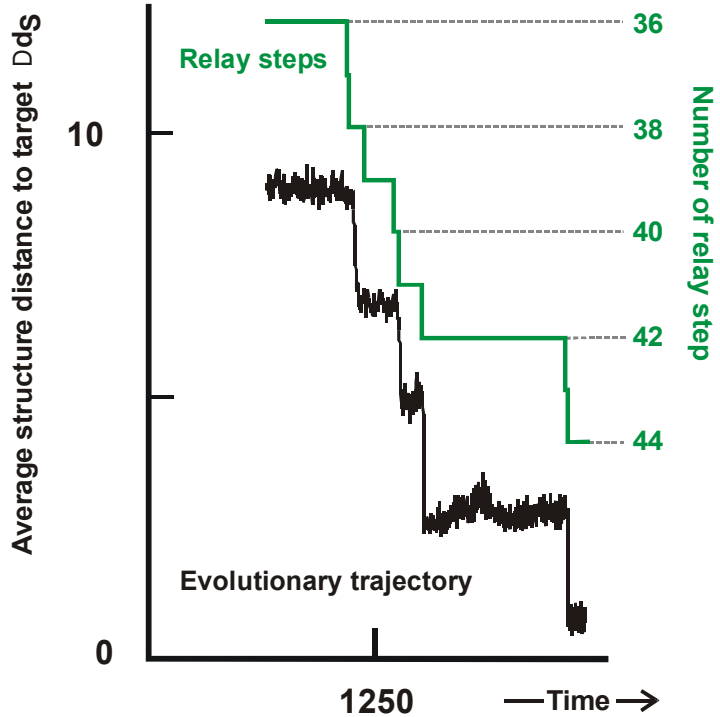
Endconformation of optimization

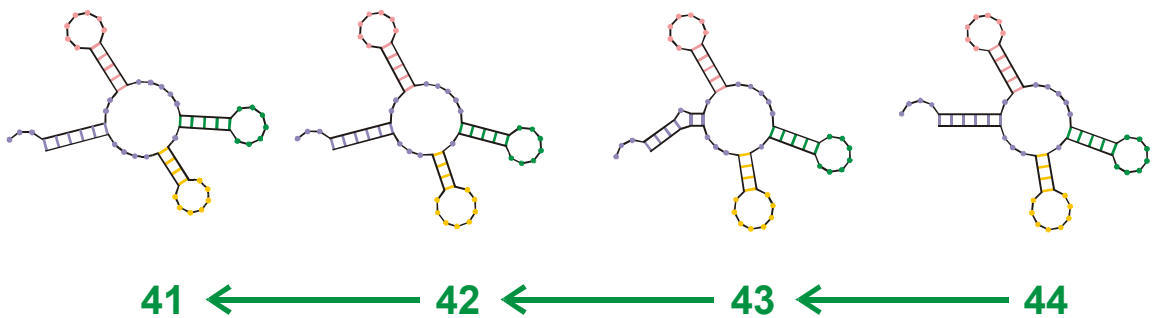


Reconstruction of the last step 43 Á 44

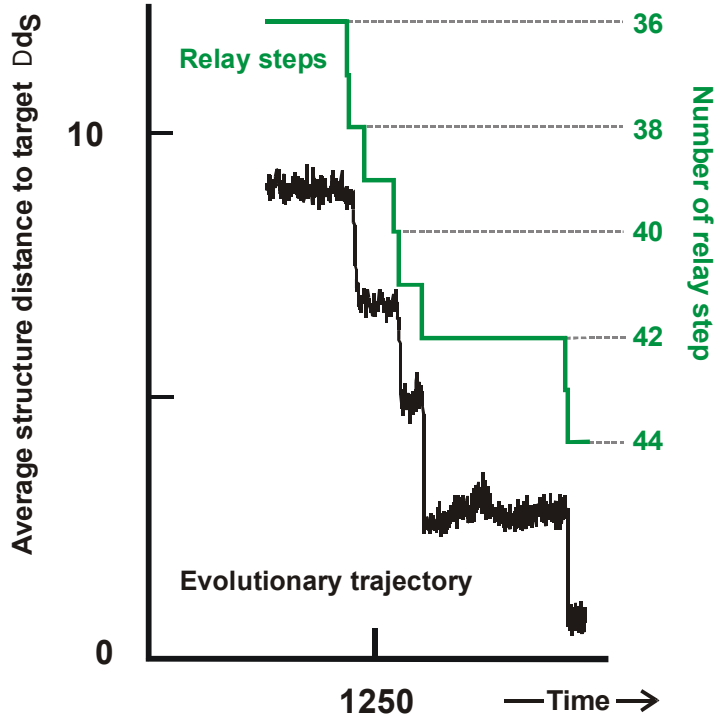


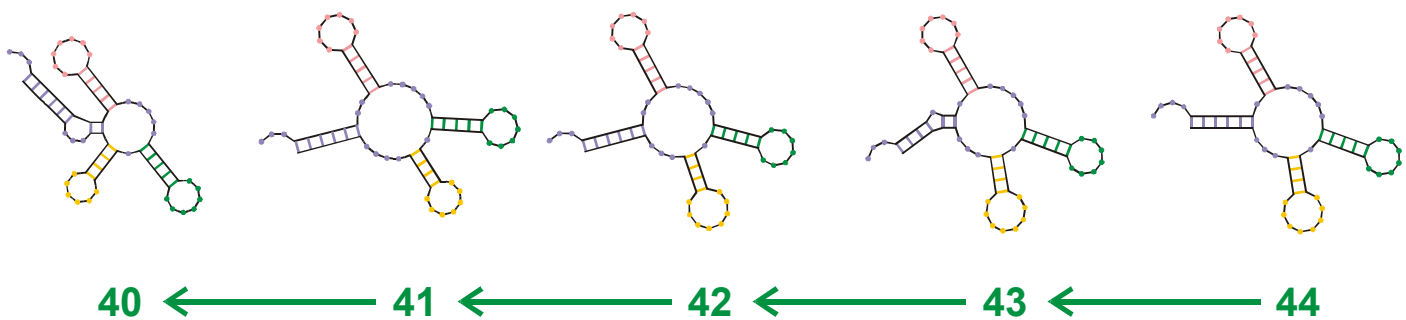
Reconstruction of last-but-one step 42 Á 43 (Á 44)



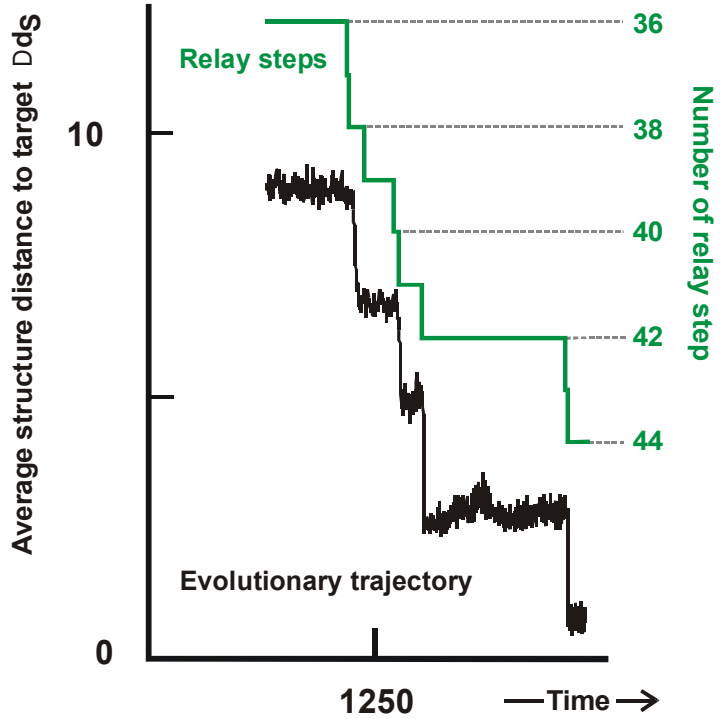


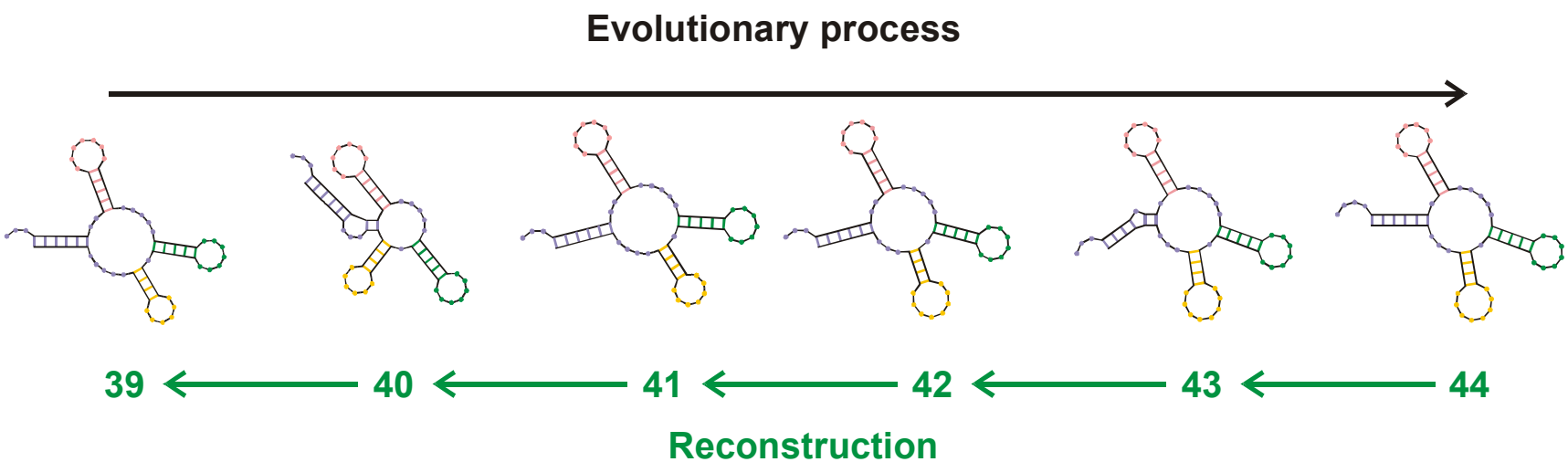
Reconstruction of step 41 Á 42 (Á 43 Á 44)



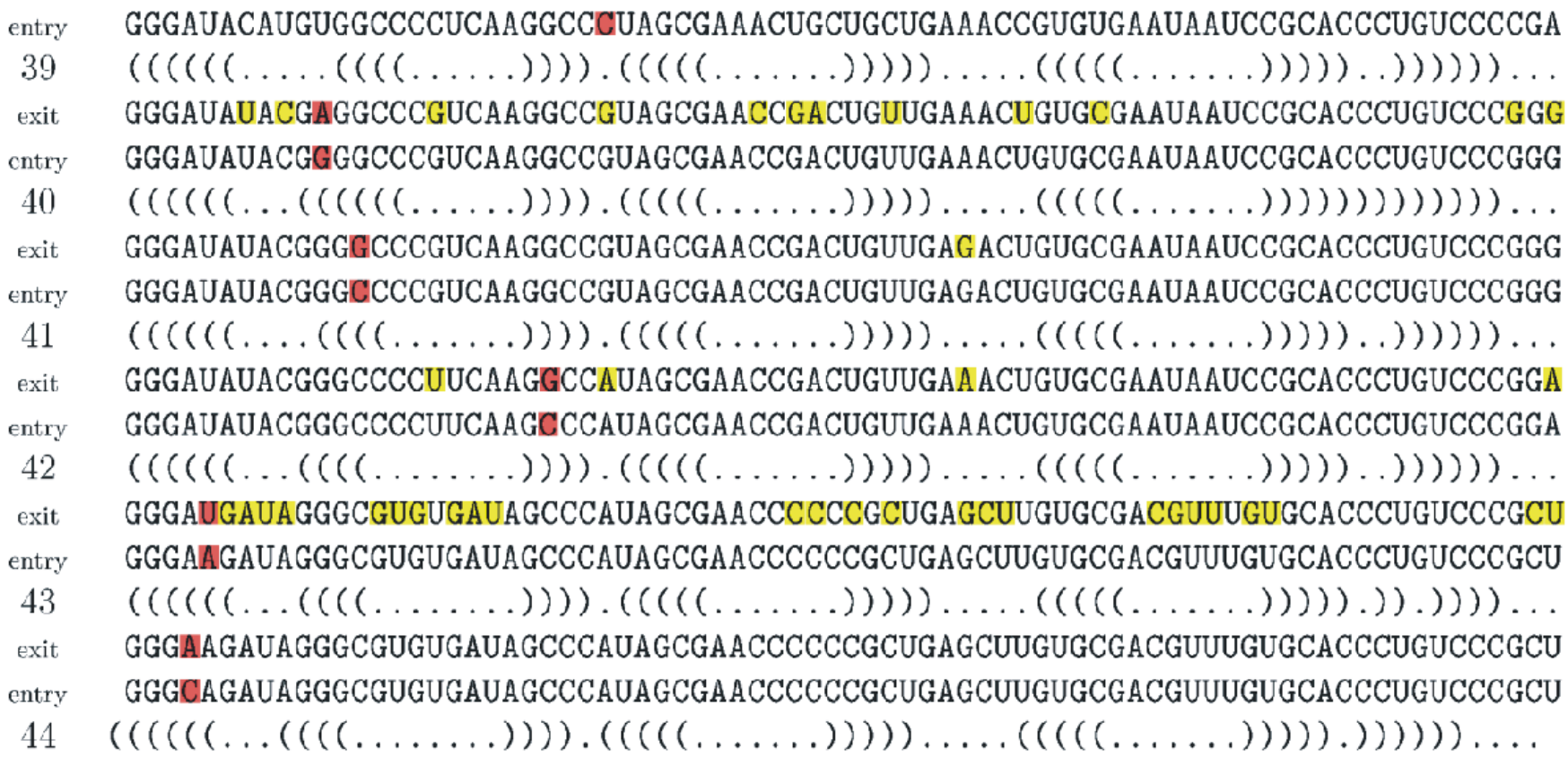


Reconstruction of step 40 Á 41 (Á 42 Á 43 Á 44)



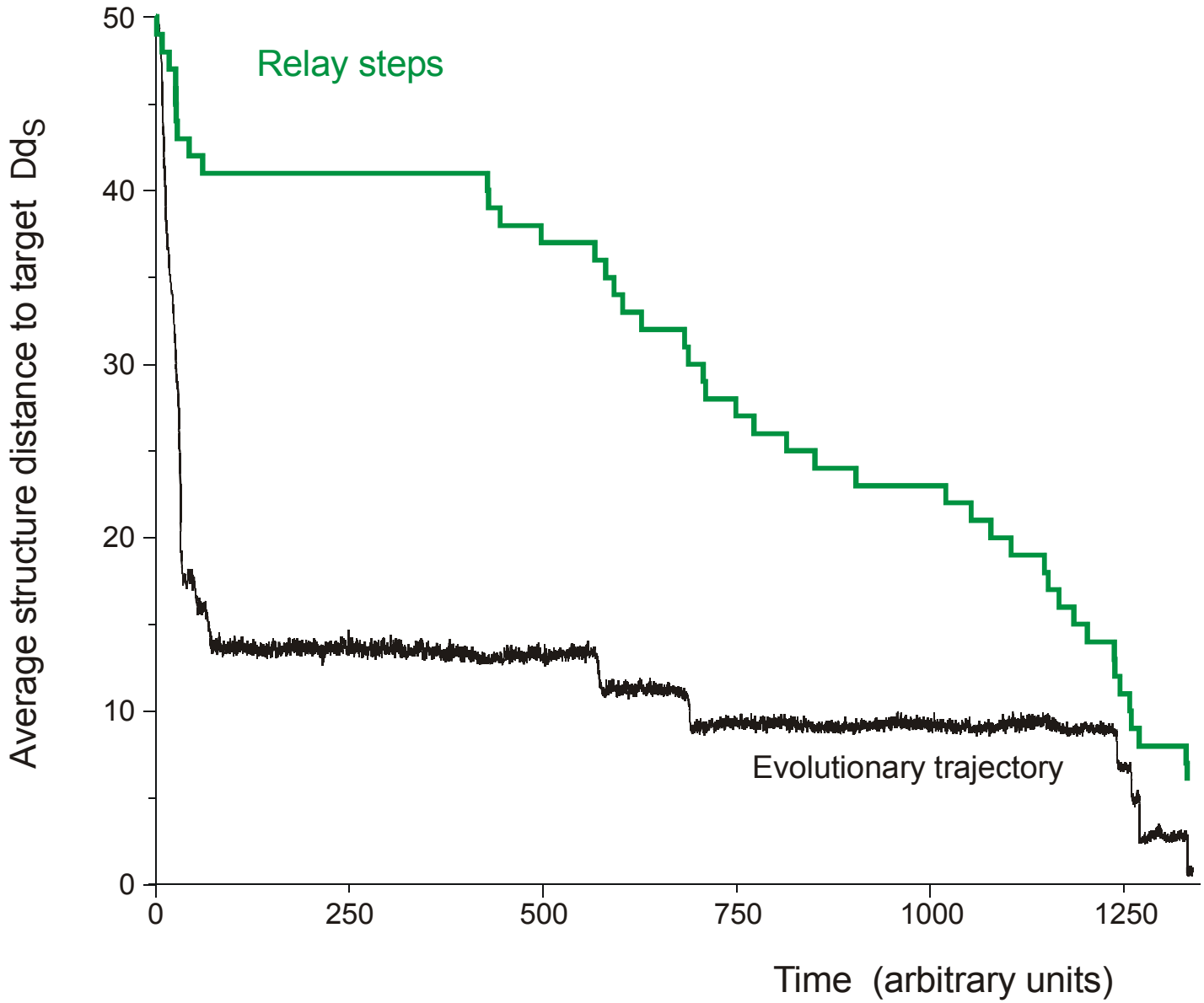


Reconstruction of the relay series



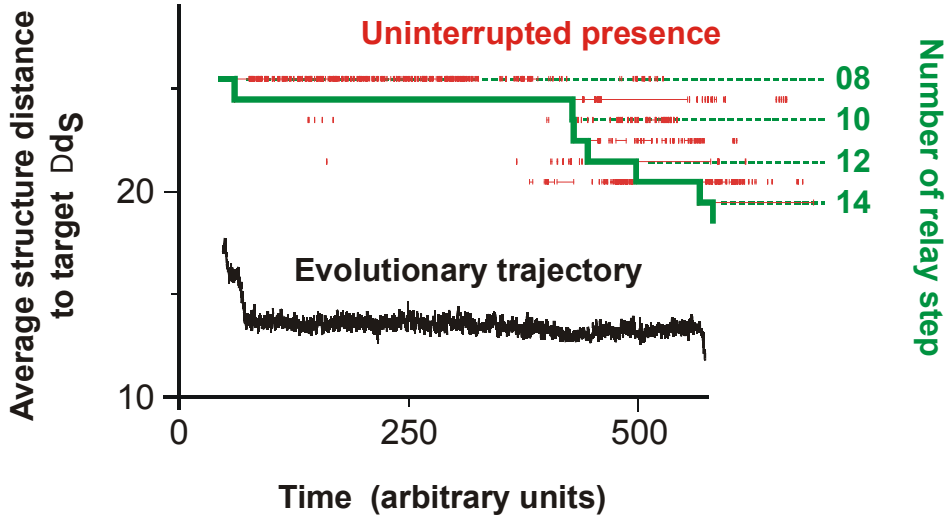
Transition inducing point mutations **Neutral point mutations**

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



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

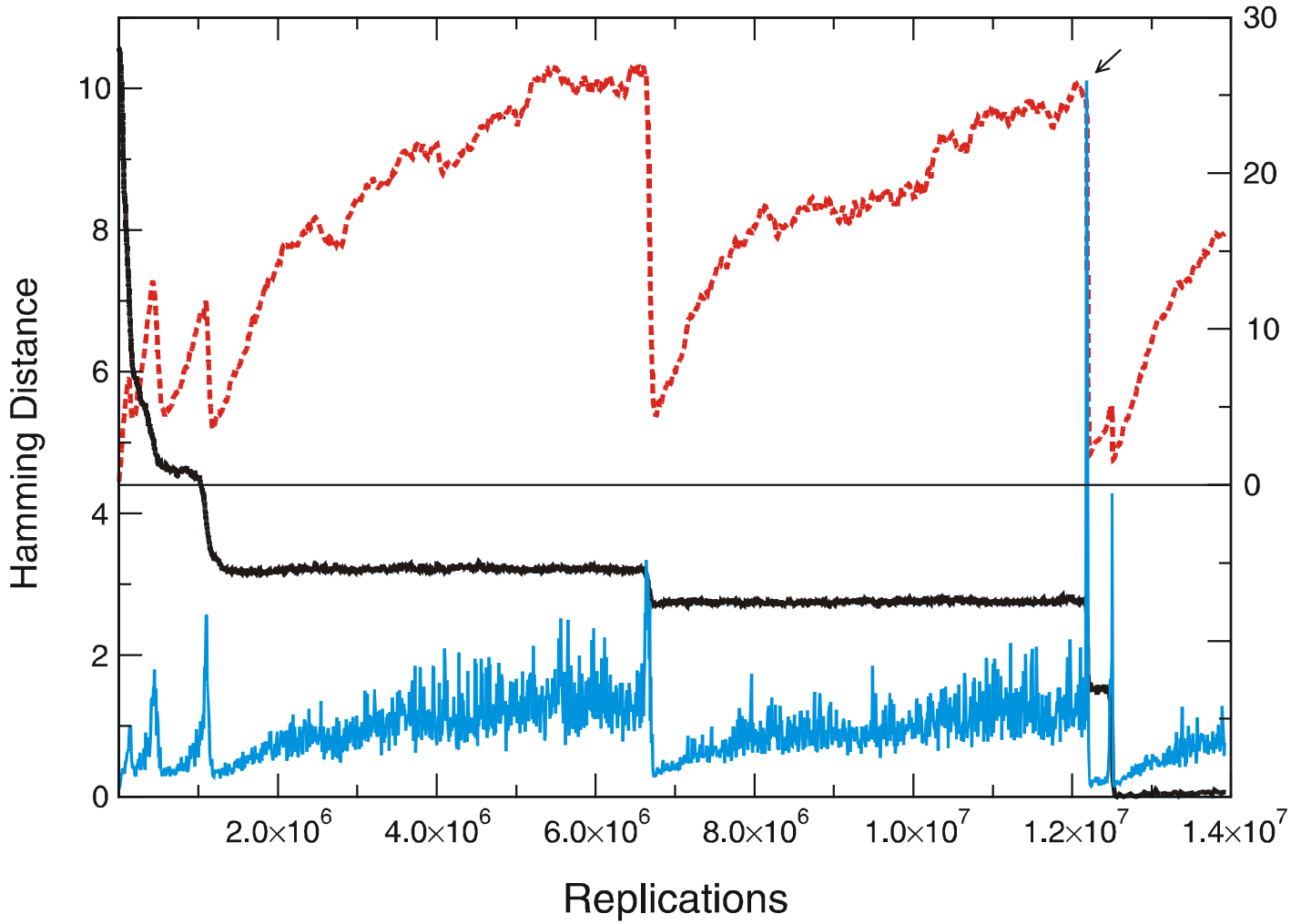
28 neutral point mutations during a long quasi-stationary epoch




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

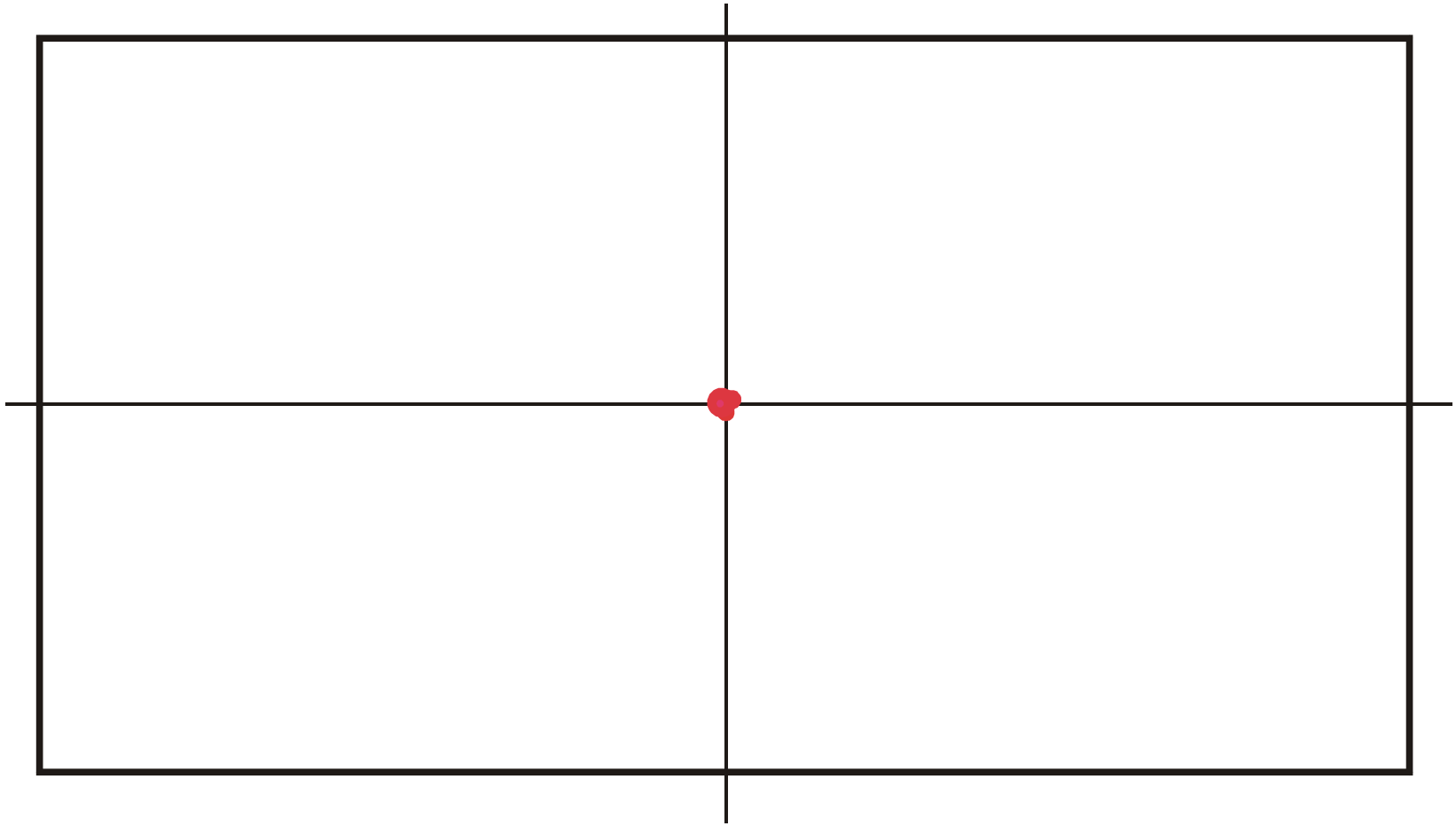
Transition inducing point mutations **Neutral point mutations**

Neutral genotype evolution during phenotypic stasis

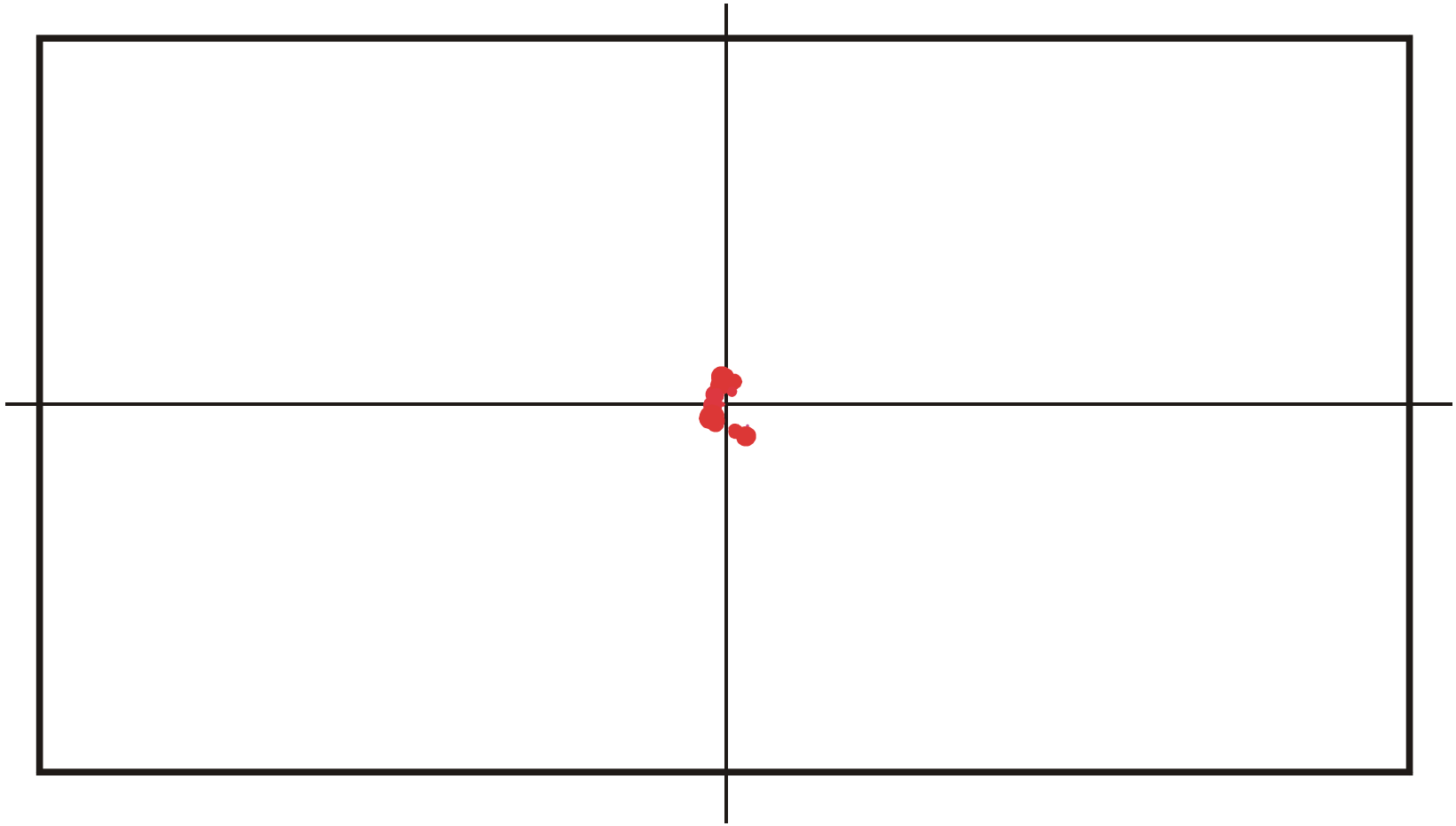


Variation in genotype space during optimization of phenotypes

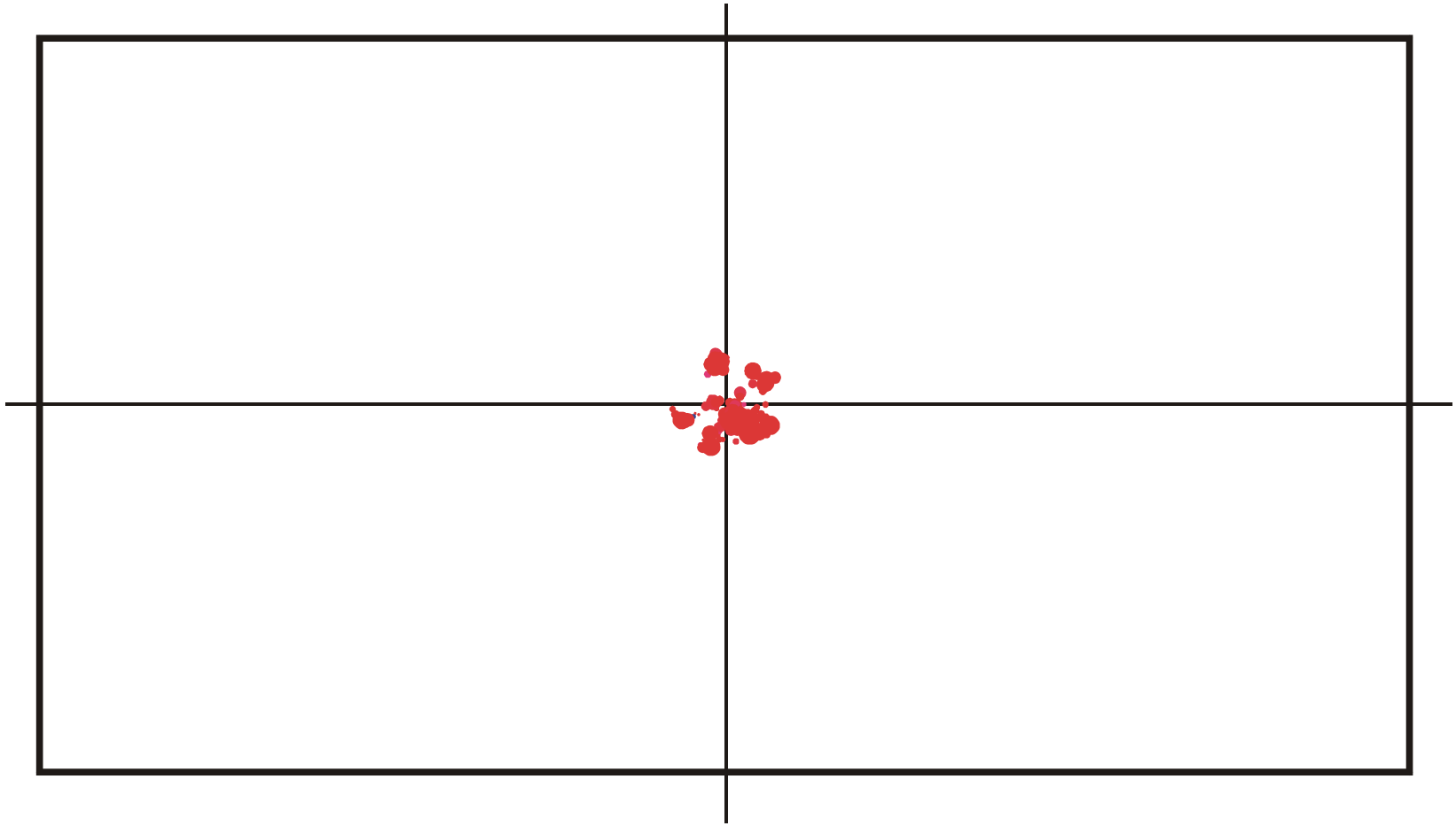
Mean Hamming distance within the population and **drift velocity of the population center** in sequence space.



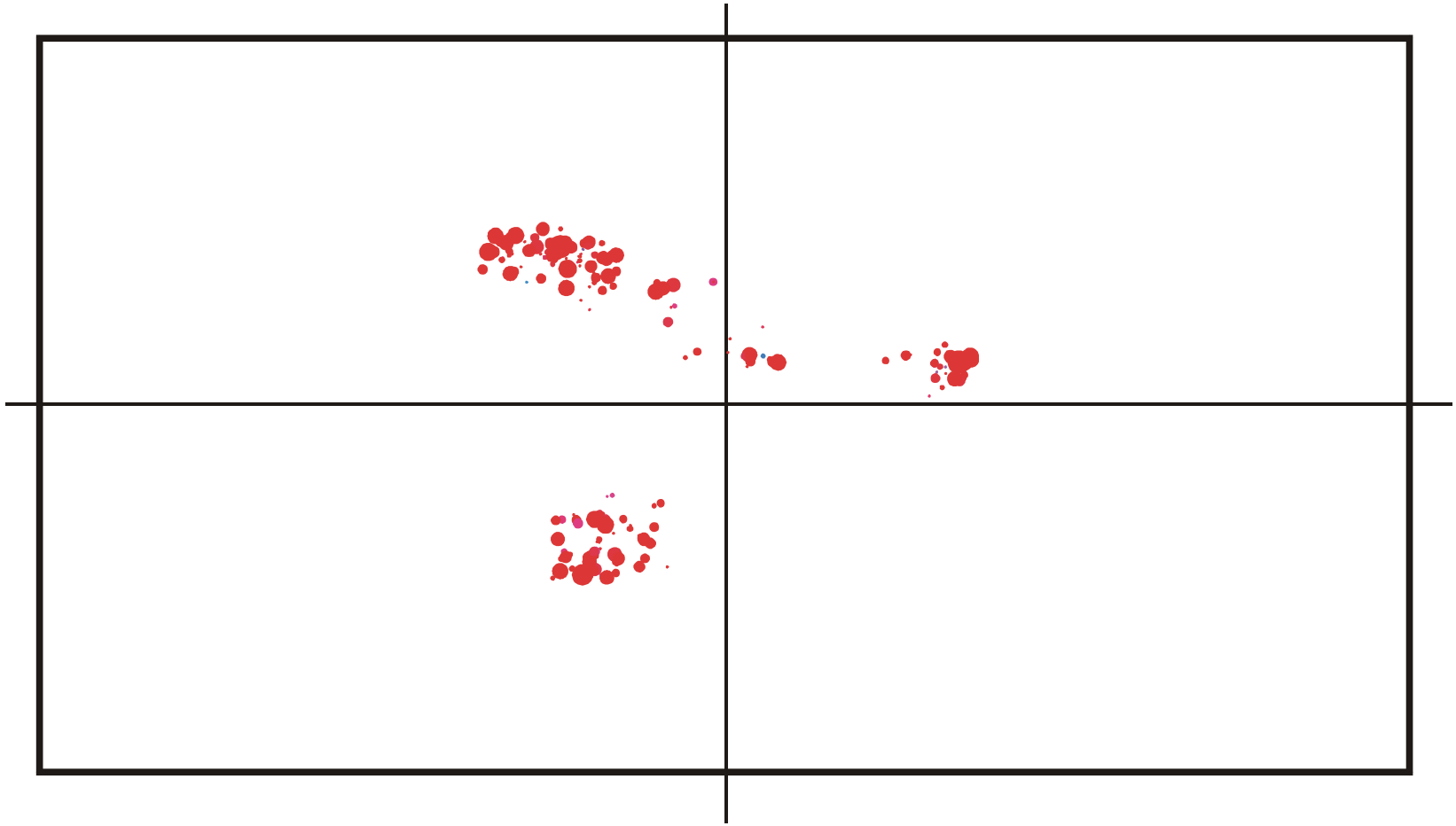
Spread of population in sequence space during a quasistationary epoch: t = 150



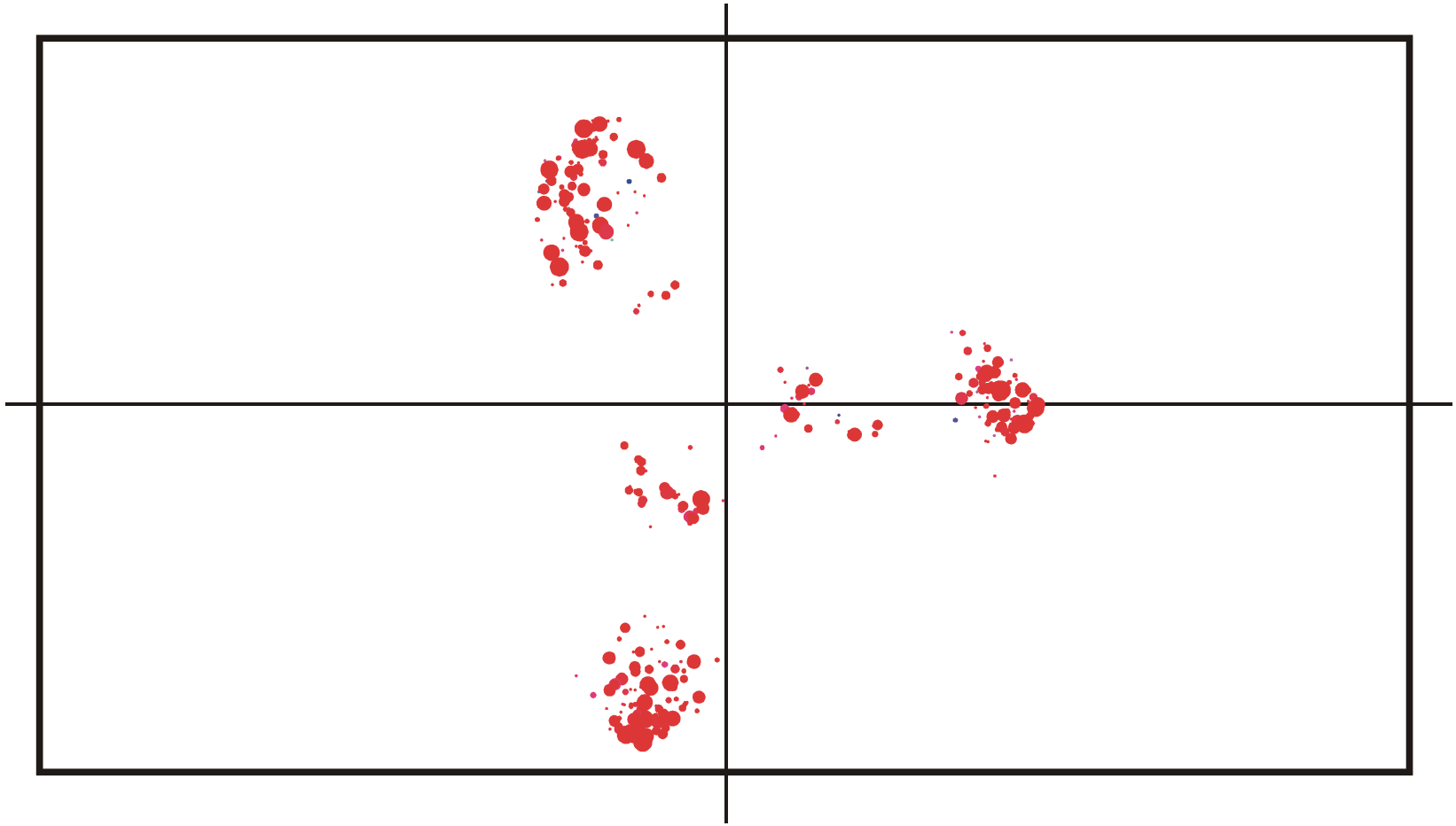
Spread of population in sequence space during a quasistationary epoch: t = 170



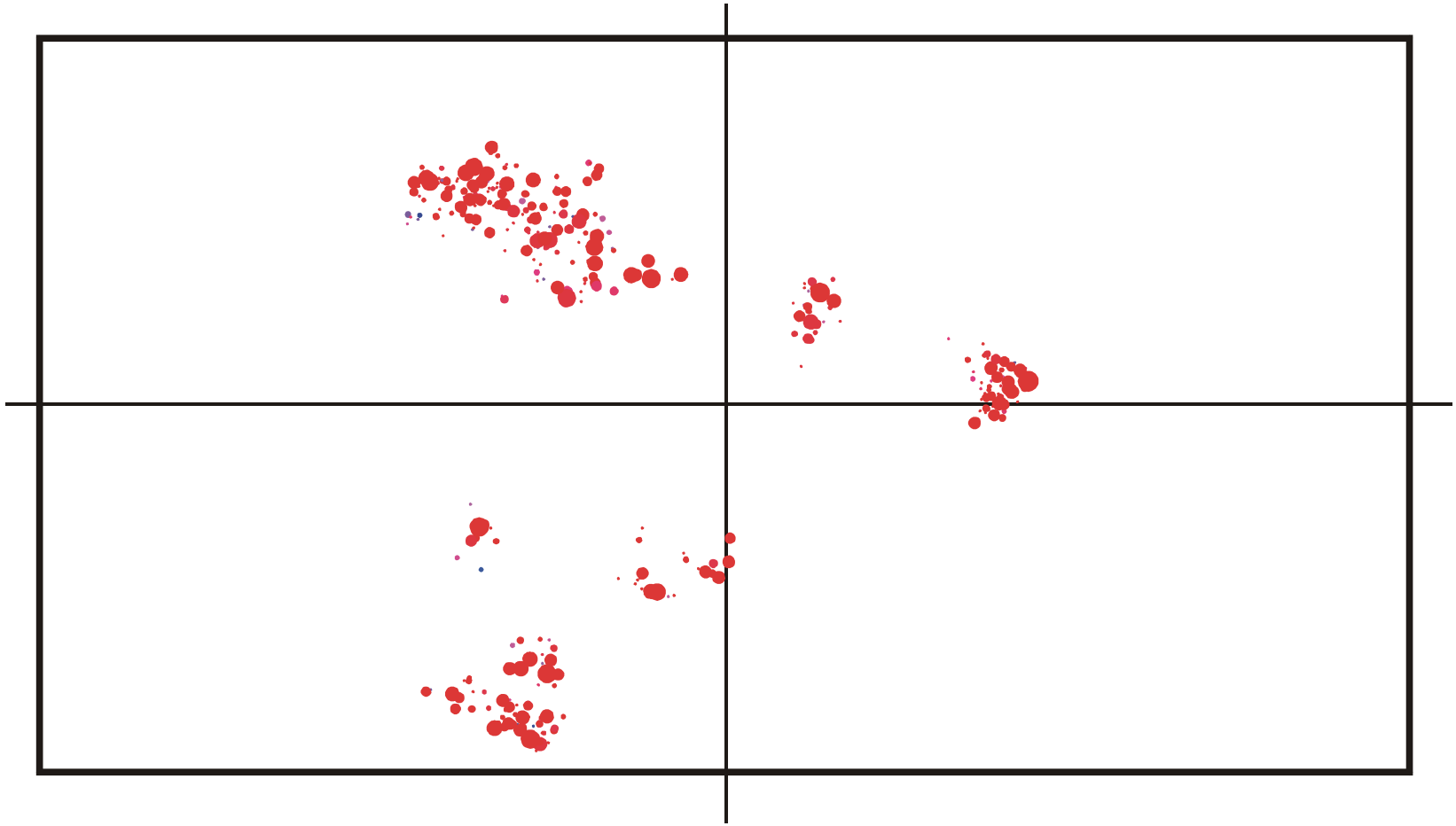
Spread of population in sequence space during a quasistationary epoch: t = 200



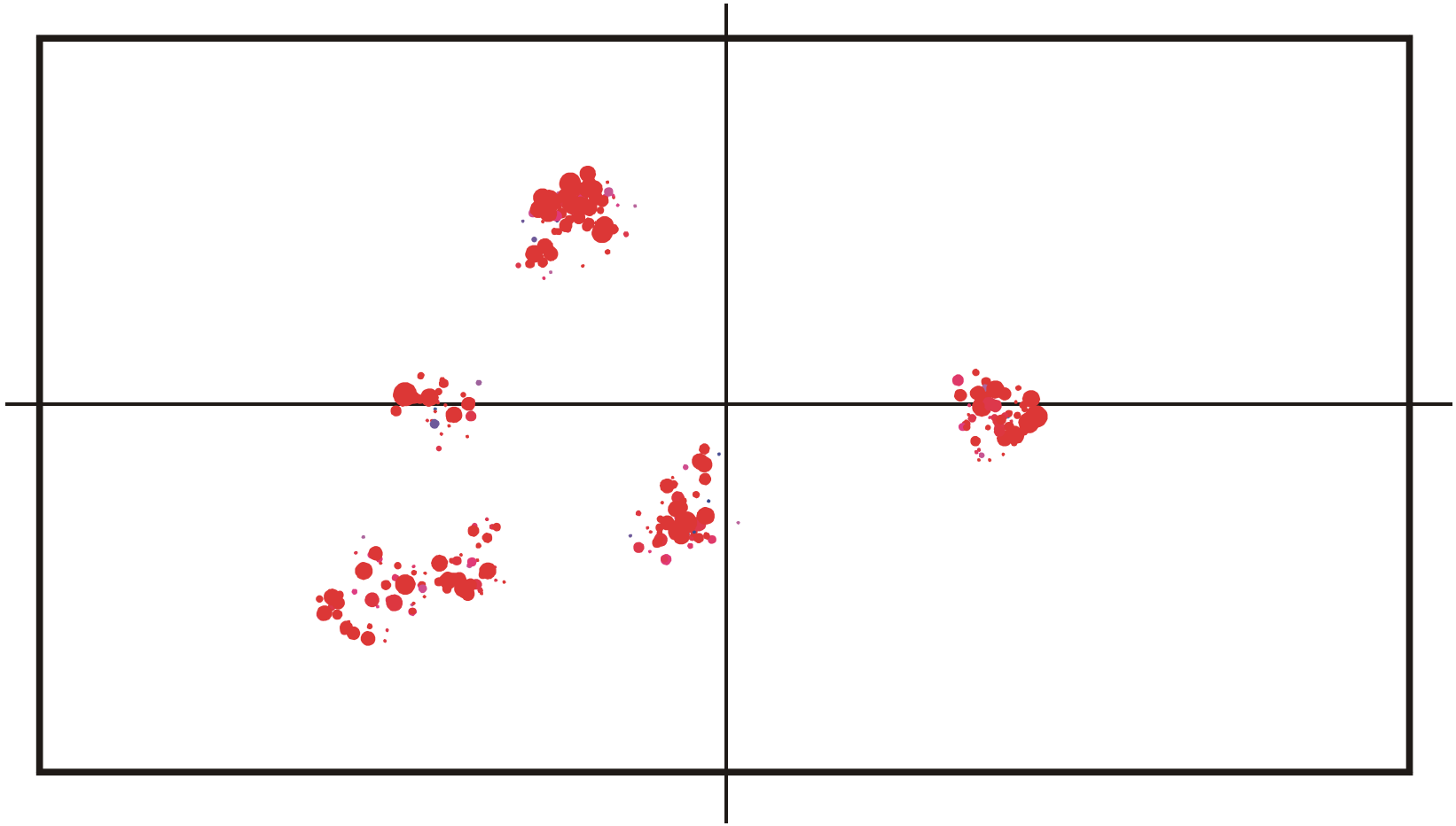
Spread of population in sequence space during a quasistationary epoch: t = 350



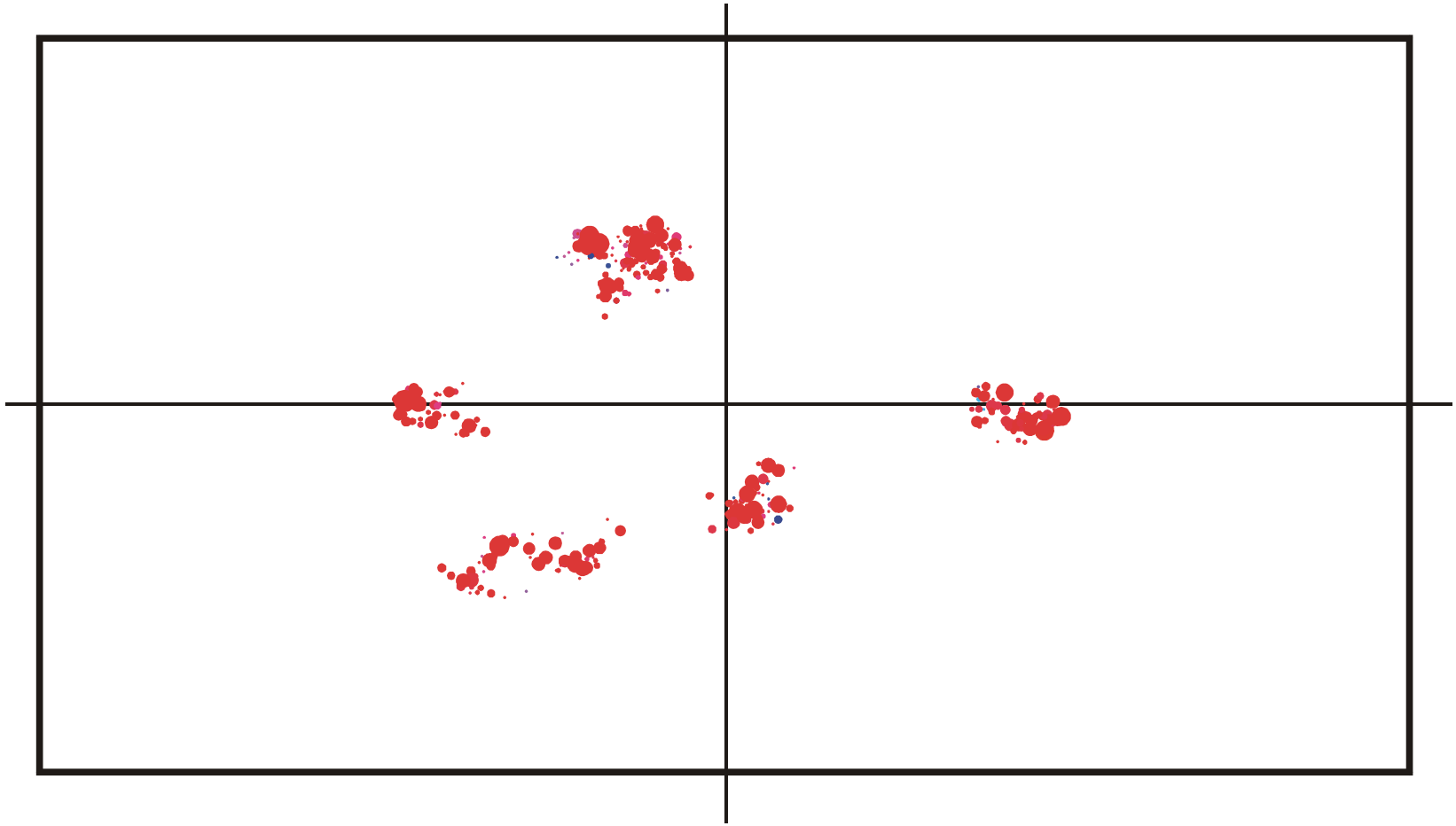
Spread of population in sequence space during a quasistationary epoch: t = 500



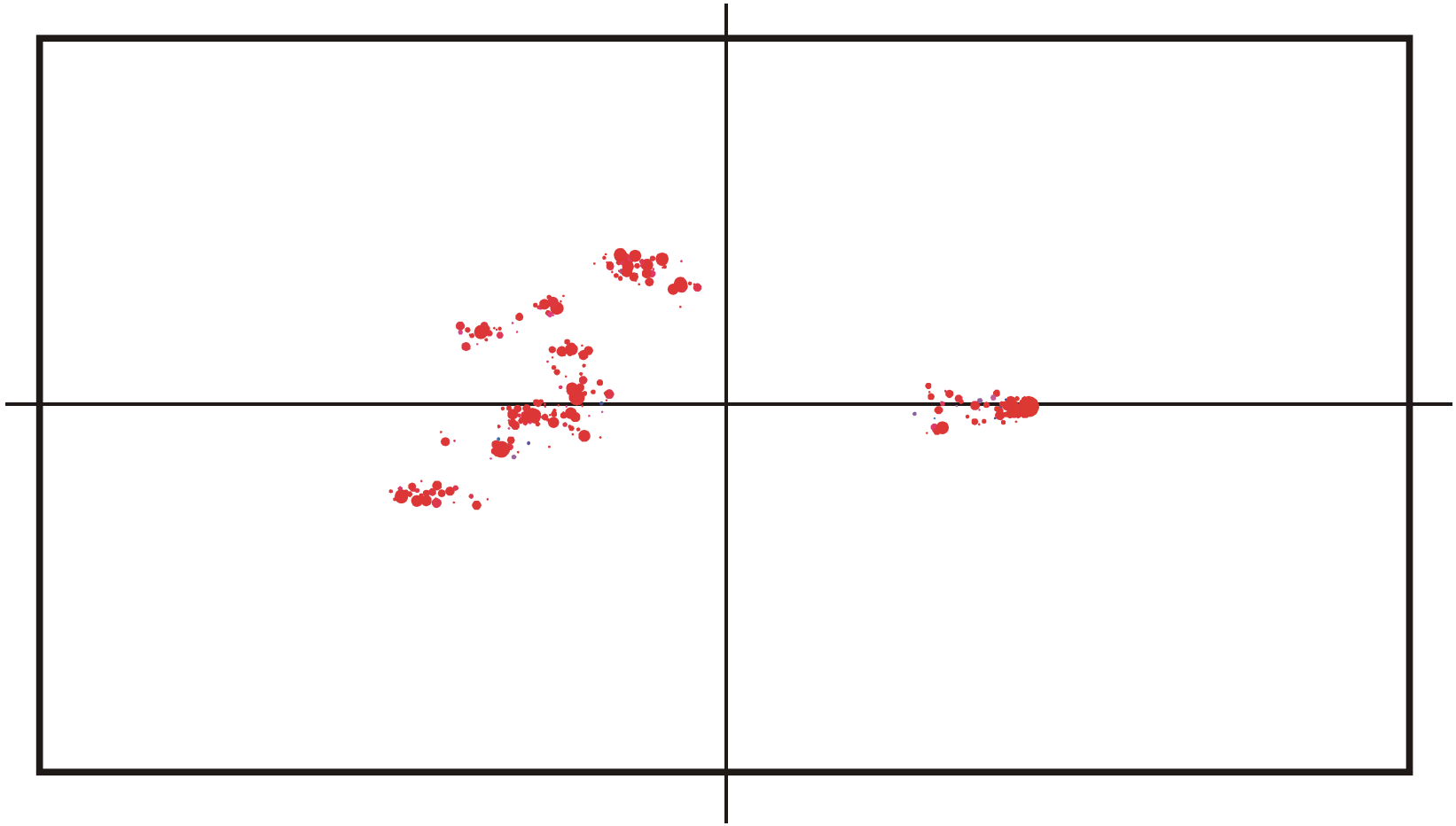
Spread of population in sequence space during a quasistationary epoch: t = 650



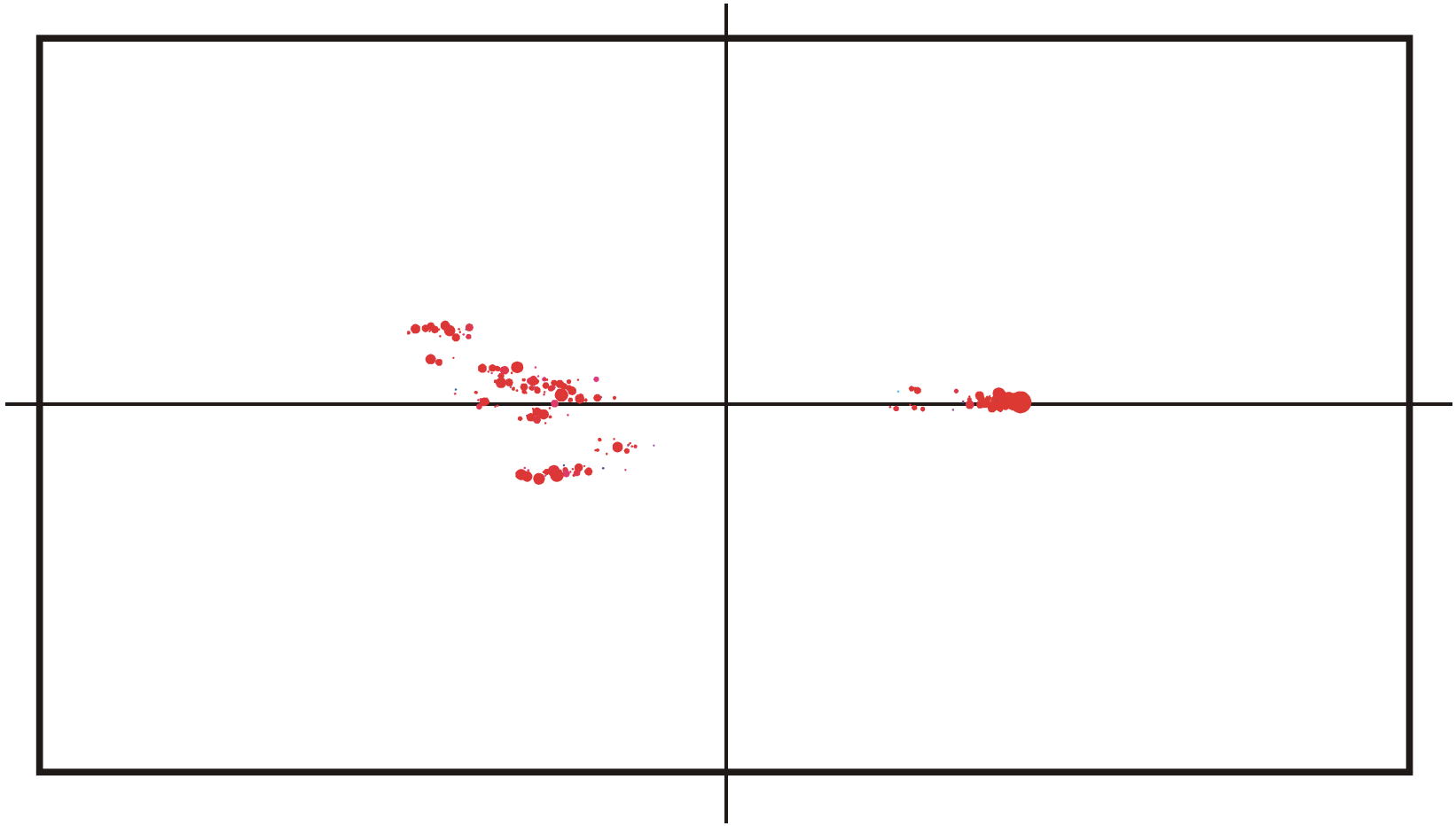
Spread of population in sequence space during a quasistationary epoch: t = 820



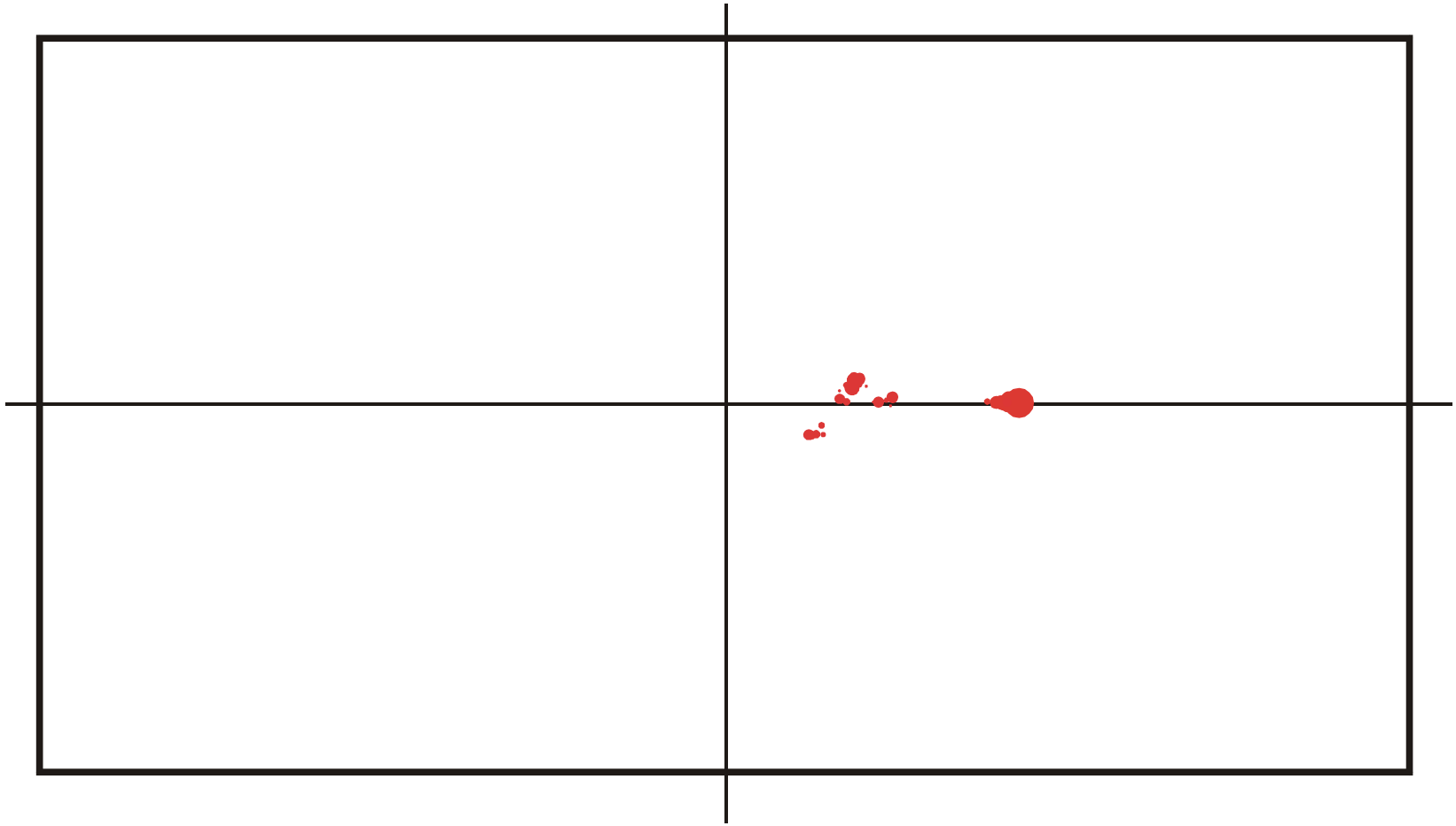
Spread of population in sequence space during a quasistationary epoch: t = 825



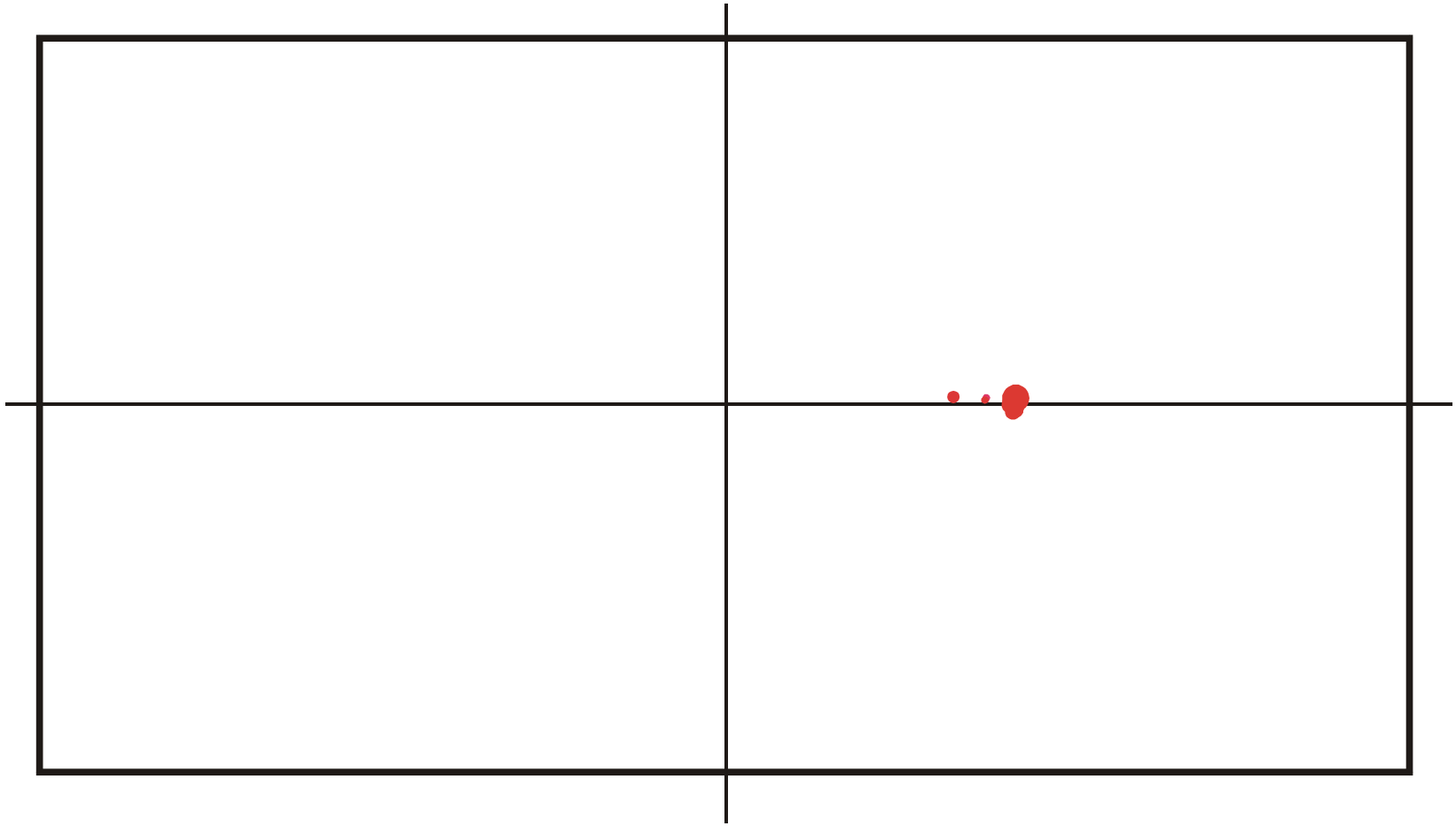
Spread of population in sequence space during a quasistationary epoch: t = 830



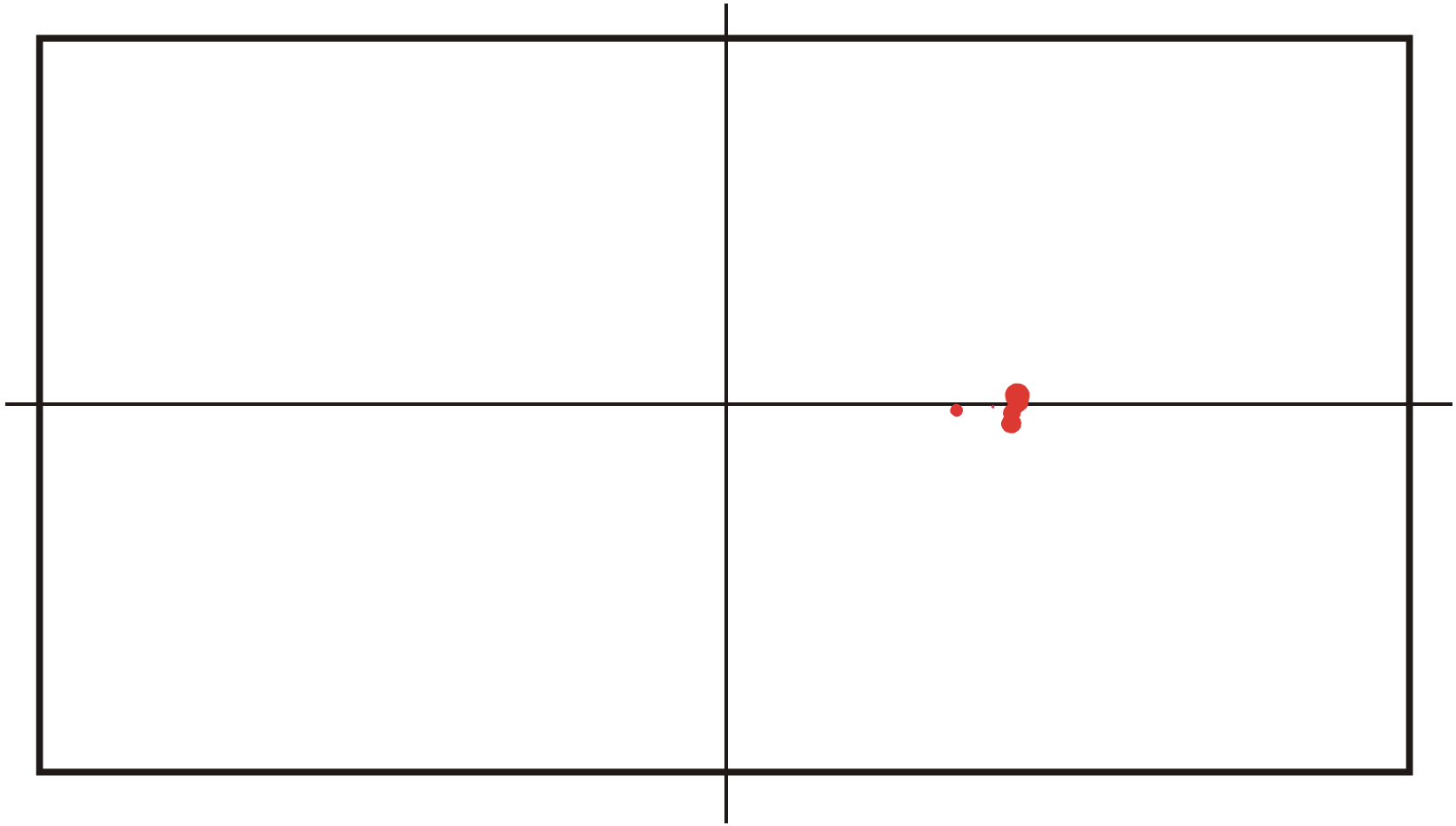
Spread of population in sequence space during a quasistationary epoch: t = 835



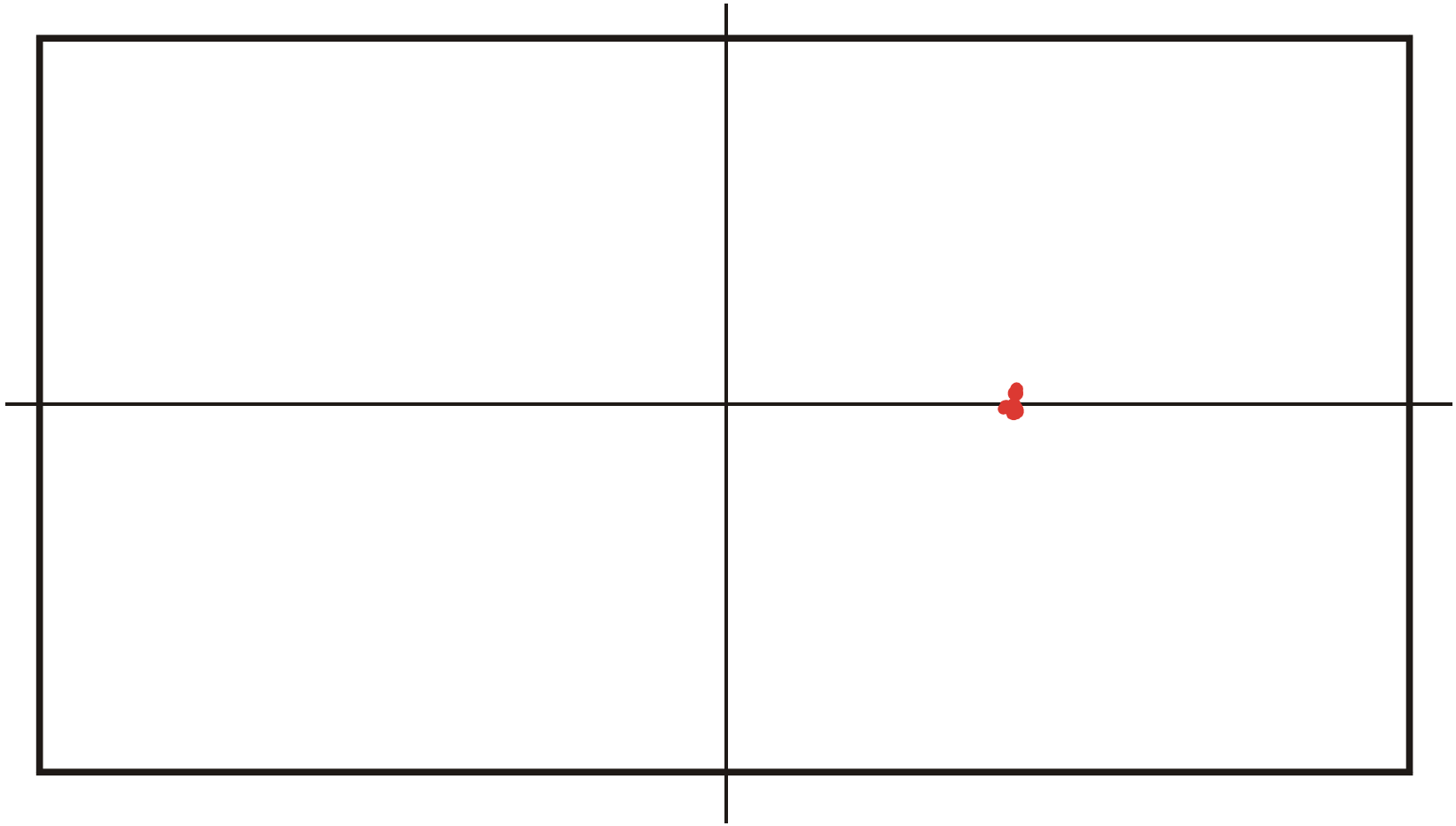
Spread of population in sequence space during a quasistationary epoch: t = 840



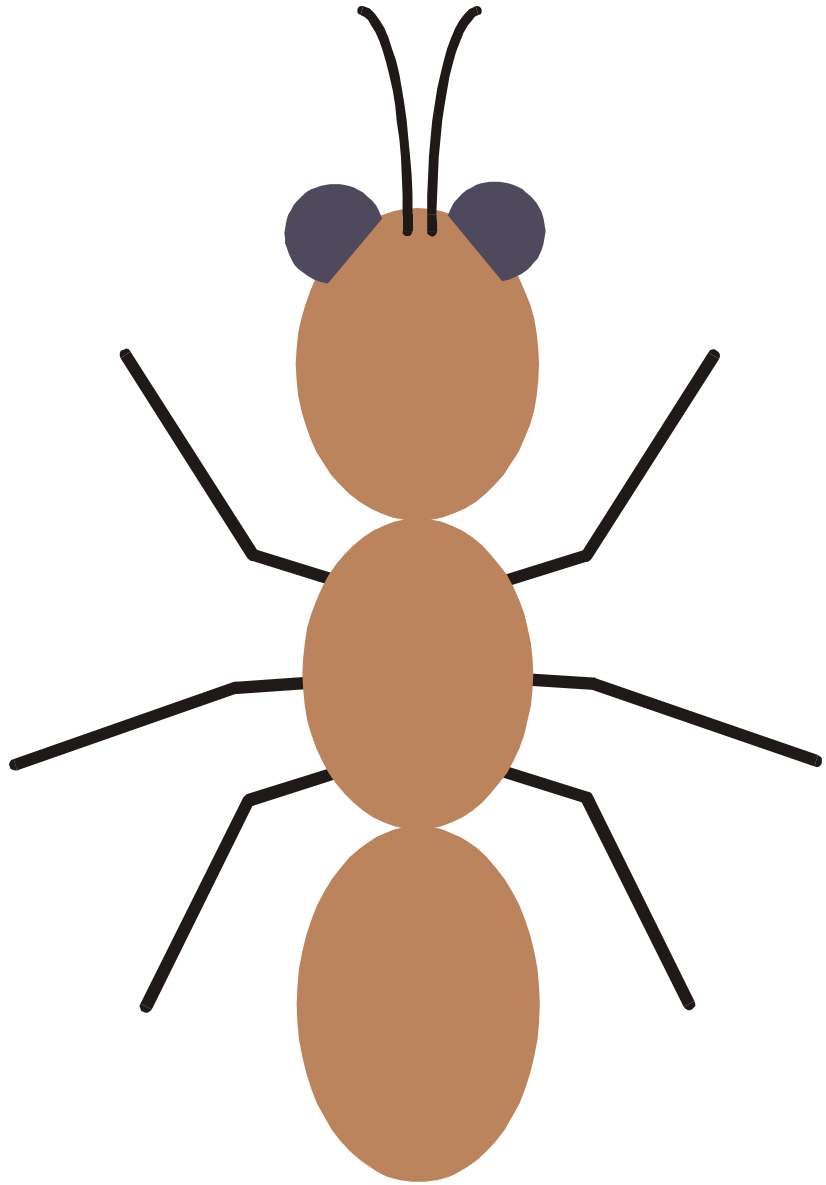
Spread of population in sequence space during a quasistationary epoch: t = 845



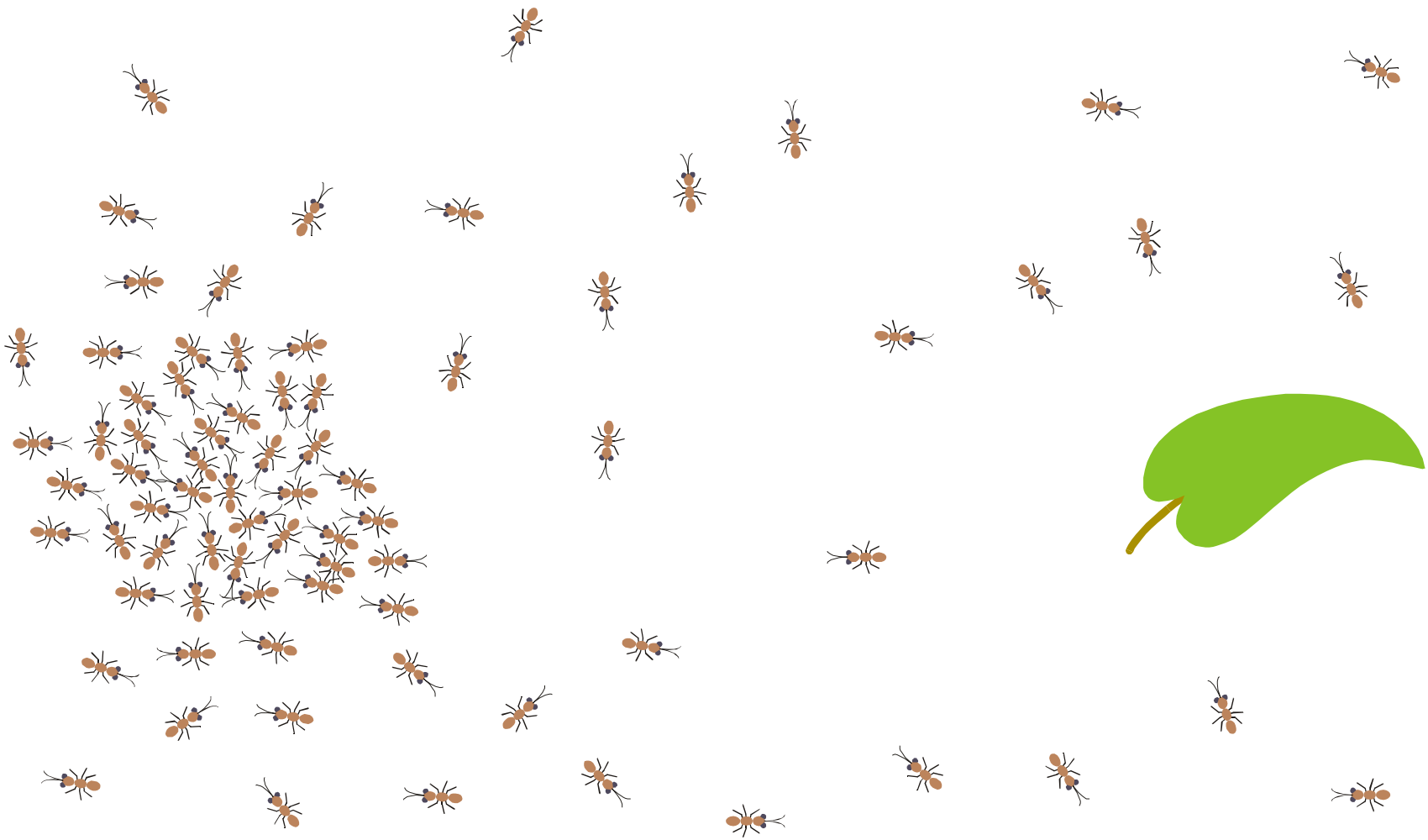
Spread of population in sequence space during a quasistationary epoch: t = 850



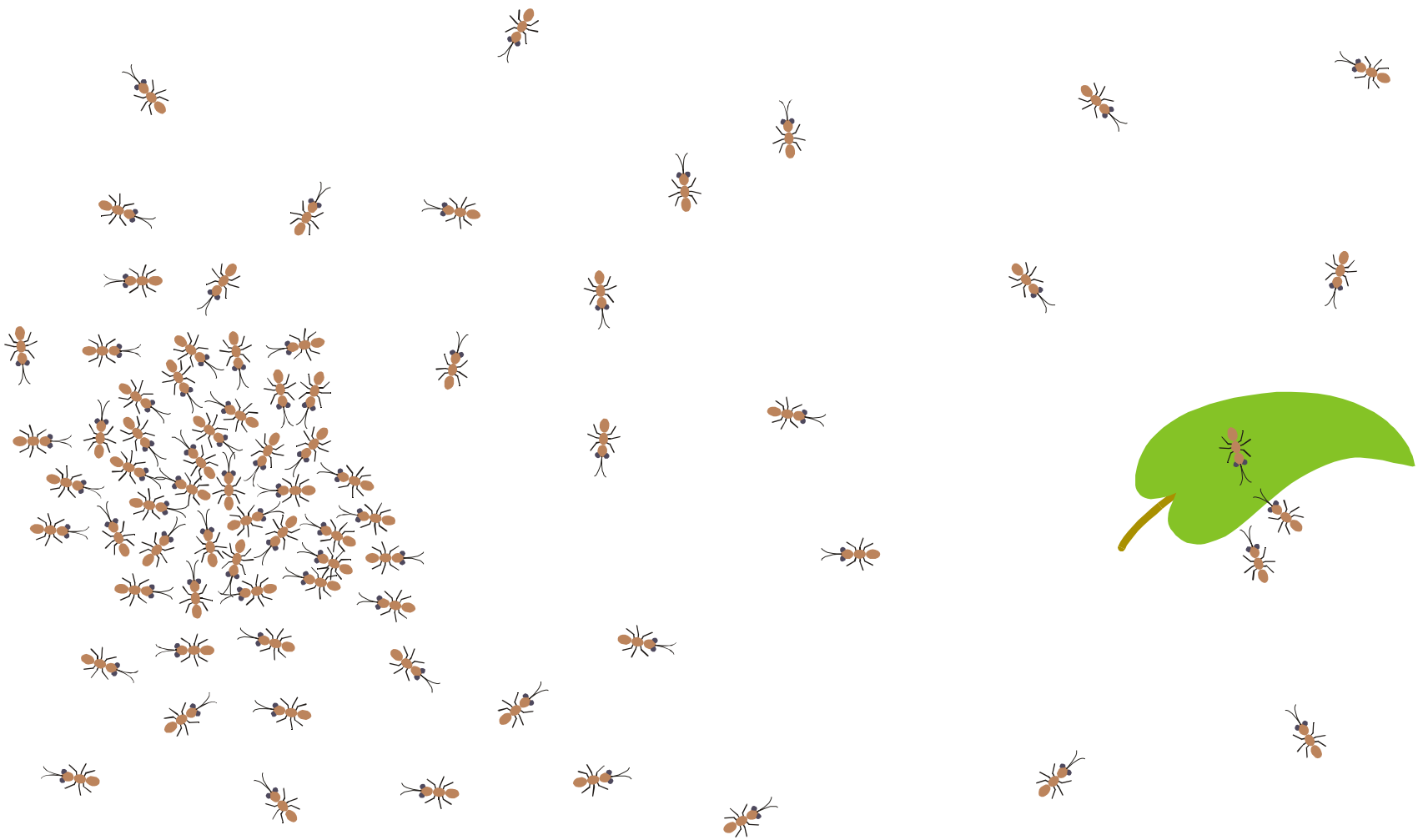
Spread of population in sequence space during a quasistationary epoch: t = 855



Element of class 2: The ant worker



Foraging behavior of ant colonies

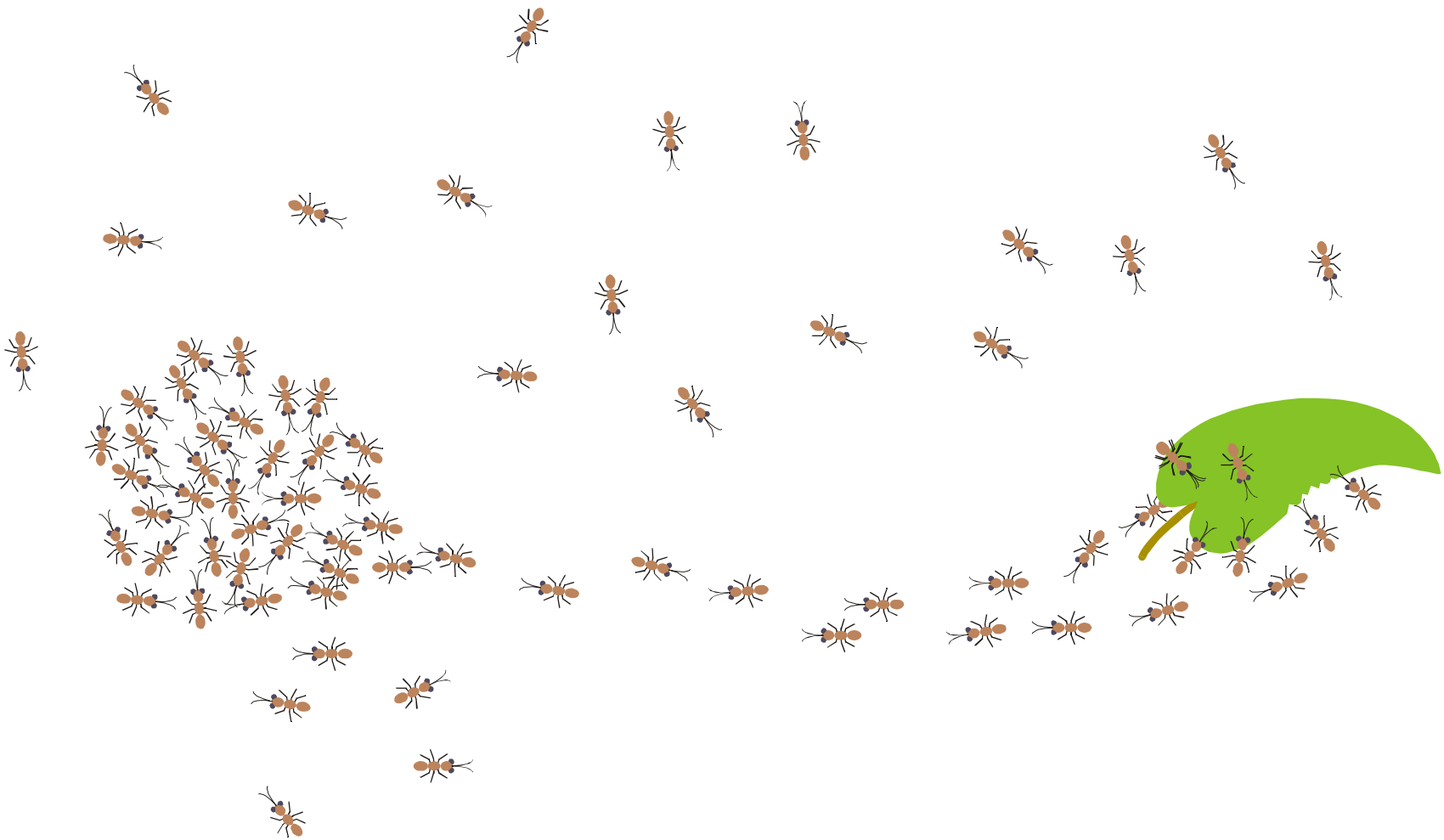


Ant colony

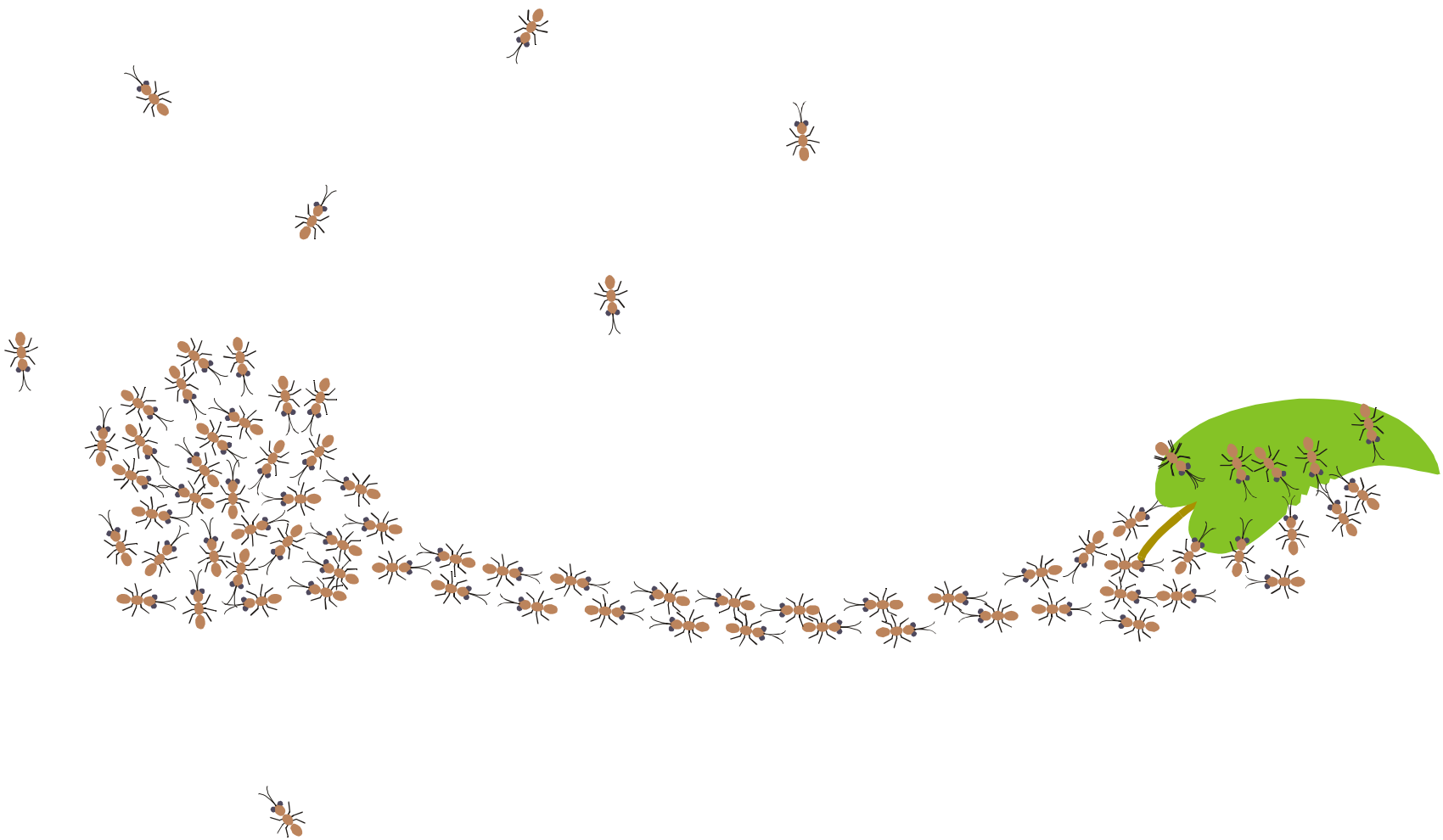
Food source detected

Food source

Foraging behavior of ant colonies



Foraging behavior of ant colonies



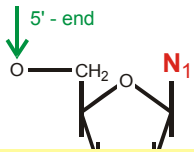
Foraging behavior of ant colonies

	RNA model	Foraging behavior of ant colonies
Element	**RNA molecule**	**Individual worker ant**
Mechanism relating elements	Mutation in quasi-species	Genetics of kinship
Search process	Optimization of RNA structure	Recruiting of food
Search space	Sequence space	Three-dimensional space
Random step	Mutation	Element of ant walk
Self-enhancing process	Replication	Secretion of pheromone
Interaction between elements	Mean replication rate	Mean pheromone concentration
Goal of the search	Target structure	Food source
Temporary memory	RNA sequences in population	Pheromone trail
'Learning' entity	**Population of molecules**	**Ant colony**

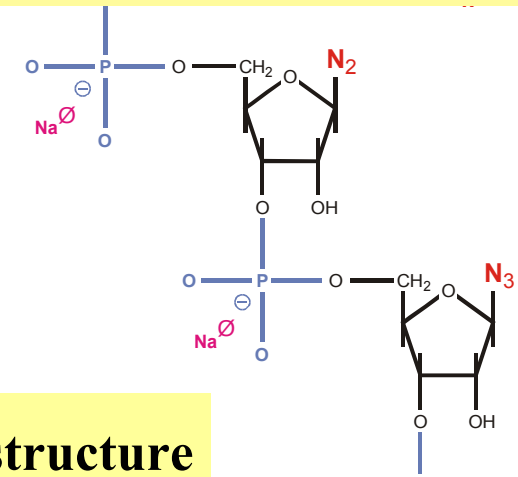
Learning at population or colony level by trial and error

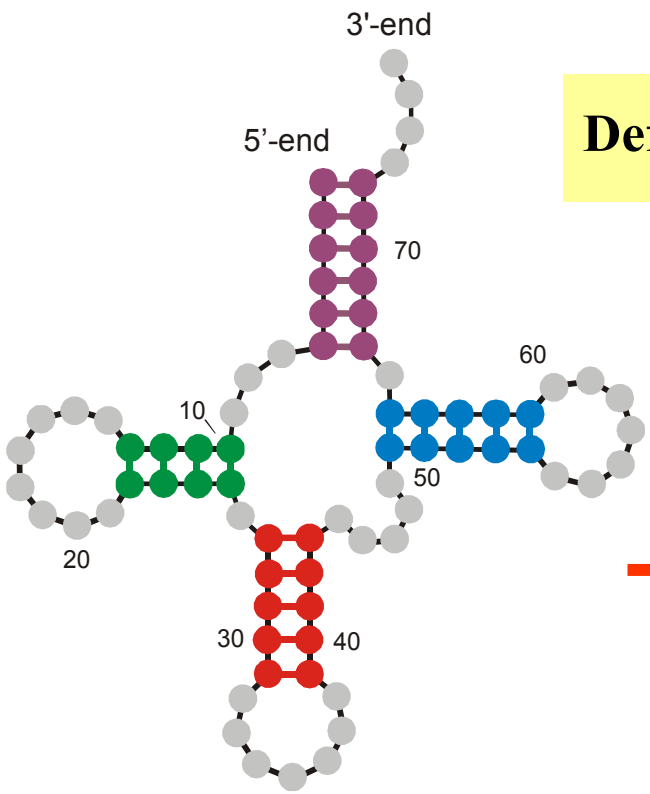
Two examples: (i) RNA model and (ii) ant colony

1. Komplexität in der Biologie
2. Evolutionäre Optimierung und Lernen im Ensemble
3. **Strukturbildung von Biomolekülen als kombinatorisches Problem**
4. Modellbildung in der Neurobiologie

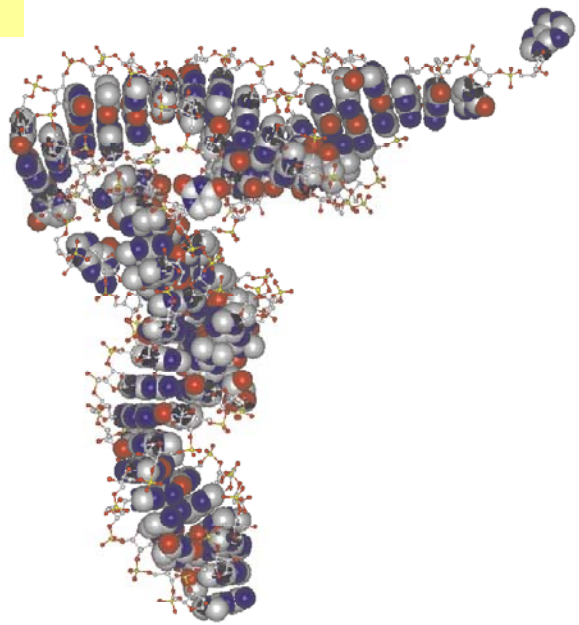


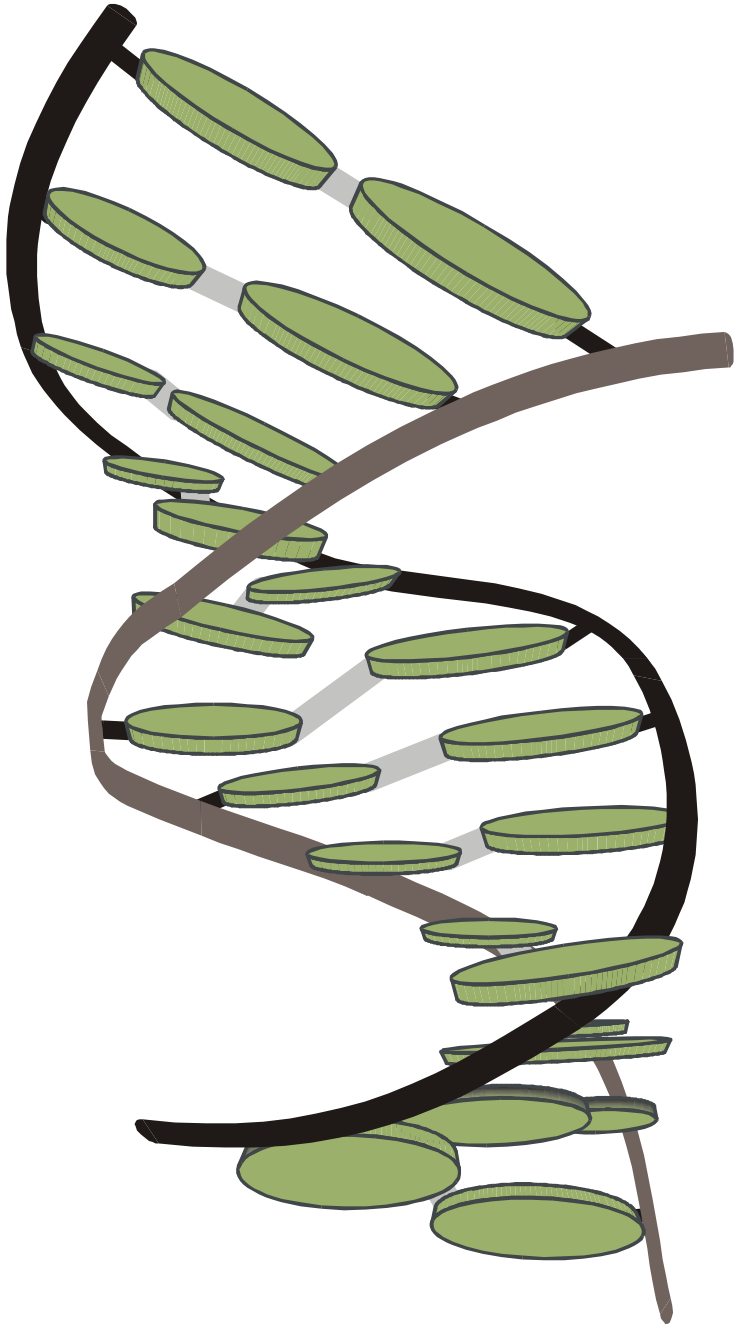
5'-end GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA 3'-end

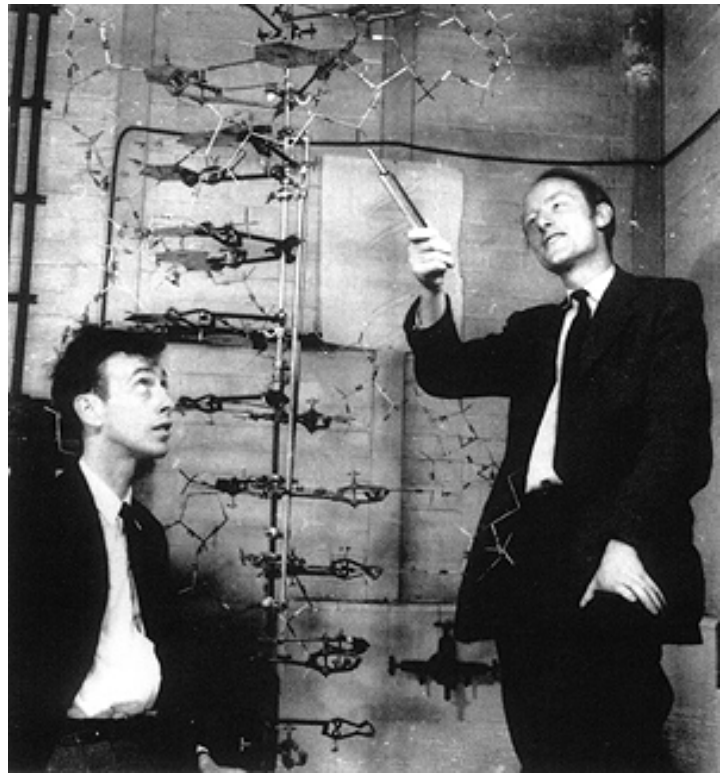




Definition of RNA structure







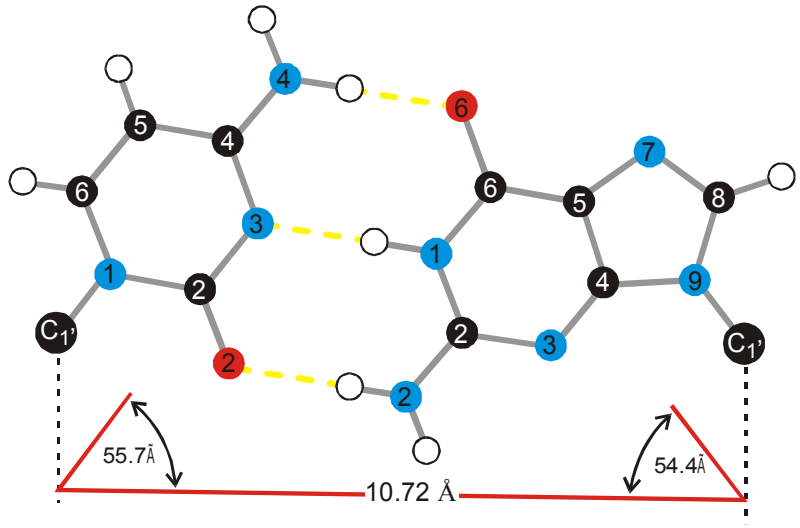
James D. Watson and Francis H.C. Crick

Nobel prize 1962

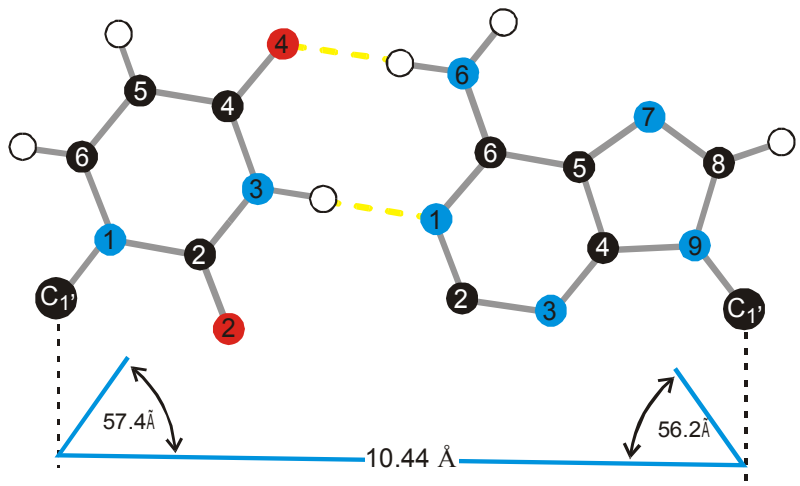
1953 – 2003 fifty years double helix

Stacking of base pairs in nucleic acid double helices (B-DNA)

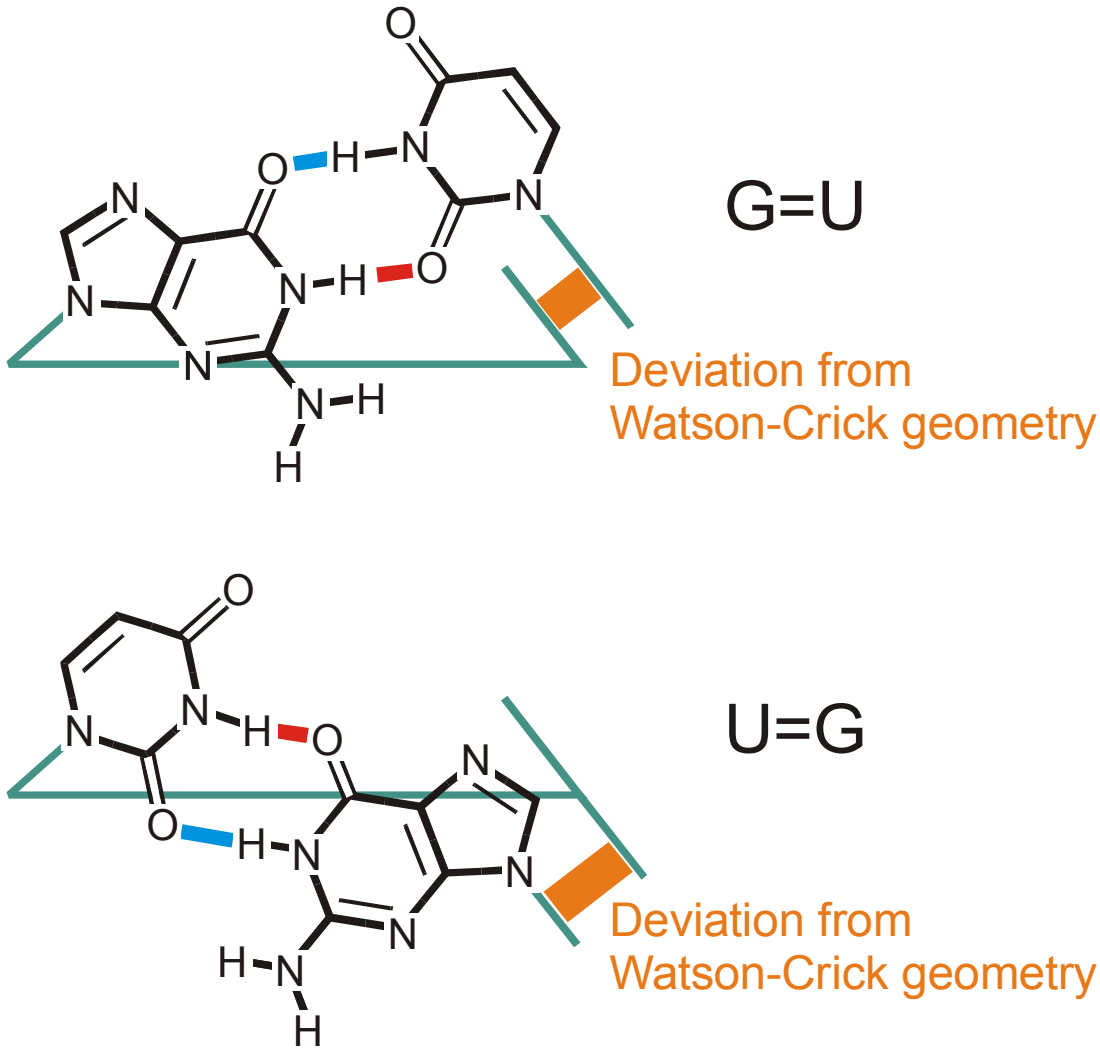
C Í G



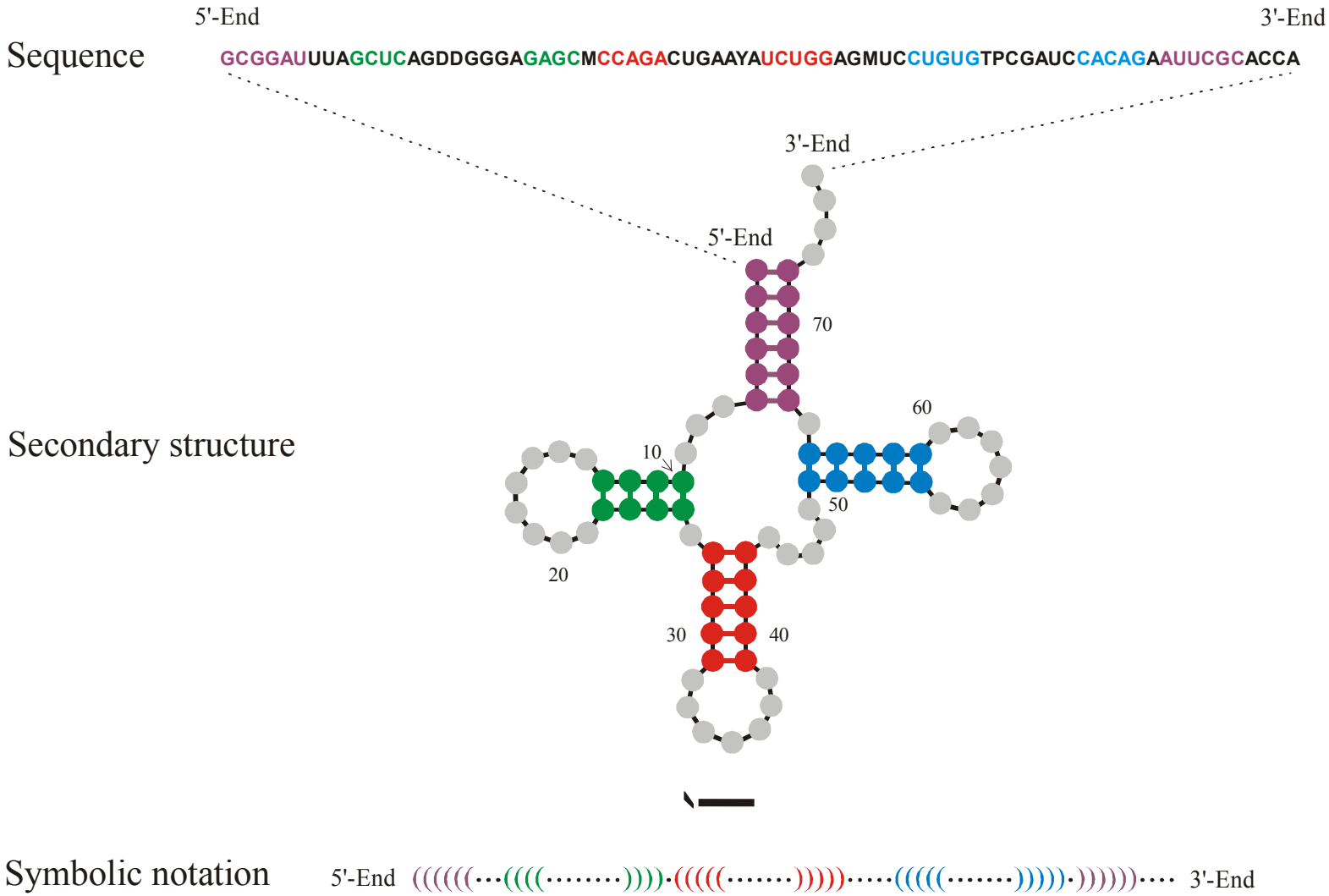
U = A



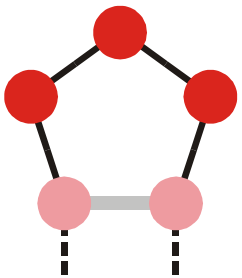
Watson-Crick type base pairs



Wobble base pairs

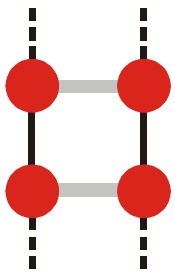


A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs



Minimal hairpin loop size:

$n_{lp} \geq 3$



Minimal stack length:

$n_{st} \geq 2$

TABLE 2 A recursion to calculate the numbers of acceptable RNA secondary structures, $N_S(\ell) = S_\ell^{(\min[n_{lp}],\min[n_{st}])}$ [49]. A structure is acceptable if all its hairpin loops contain three or more nucleotides (loopsize: $n_{lp} \geq 3$) and if it has no isolated base pairs (stacksize: $n_{st} \geq 2$). The recursion $m+1 \Longrightarrow m$ yields the desired results in the array Ψ_m and uses two auxiliary arrays with the elements Φ_m and Ξ_m, which represent the numbers of structures with or without a closing base pair $(1, m)$. One array, e.g., Φ_m, is dispensible, but then the formula contains a double sum that is harder to interpret.

Recursion formula:

$$\Xi_{m+1} = \Psi_m + \sum_{k=5}^{m-2} \Phi_k \cdot \Psi_{m-k-1}$$

$$\Phi_{m+1} = \sum_{k=1}^{\lfloor (m-2)/2 \rfloor} \Xi_{m-2k+1}$$

$$\Psi_{m+1} = \Xi_{m+1} + \Phi_{m-1}$$

Recursion: $m+1 \Longrightarrow m$

Initial conditions:

$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = \Psi_4 = \Psi_5 = \Psi_6 = 1$$

$$\Phi_0 = \Phi_1 = \Phi_2 = \Phi_3 = \Phi_4 = 0$$

$$\Xi_0 = \Xi_1 = \Xi_2 = \Xi_3 = \Xi_4 = \Xi_5 = \Xi_6 = \Xi_7 = 1$$

Solution: $S_\ell^{(3,2)} = \Psi_{m=\ell}$

Recursion formula for the number of acceptable RNA secondary structures

ℓ	Number of Sequences 2^ℓ	4^ℓ	Number of Structures $S_\ell^{(3,2)}$	GC	UGC	AUGC	AUG	AU
7	128	1.64×10^4	2	1	1	1	1	1
8	256	6.55×10^4	4	3	3	3	1	1
9	512	2.62×10^5	8	7	7	7	1	1
10	1 024	1.05×10^6	14	13	13	13	1	1
15	3.28×10^4	1.07×10^9	174	130	145	152	37	15
16	6.55×10^4	4.29×10^9	304	214	245	257	55	25
19	5.24×10^5	2.75×10^{11}	1 587	972	1 235		220	84
20	1.05×10^6	1.10×10^{12}	2 741	1 599	2 112		374	128
29	5.37×10^8	2.88×10^{17}	430 370	132 875				8 690
30	1.07×10^9	1.15×10^{18}	760 983	218 318				13 726

Computed numbers of minimum free energy structures over different nucleotide alphabets

P. Schuster, ***Molecular insights into evolution of phenotypes***. In: J. Crutchfield & P.Schuster, Evolutionary Dynamics. Oxford University Press, New York 2003, pp.163-215.

RNA sequence **GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA**

RNA folding:

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

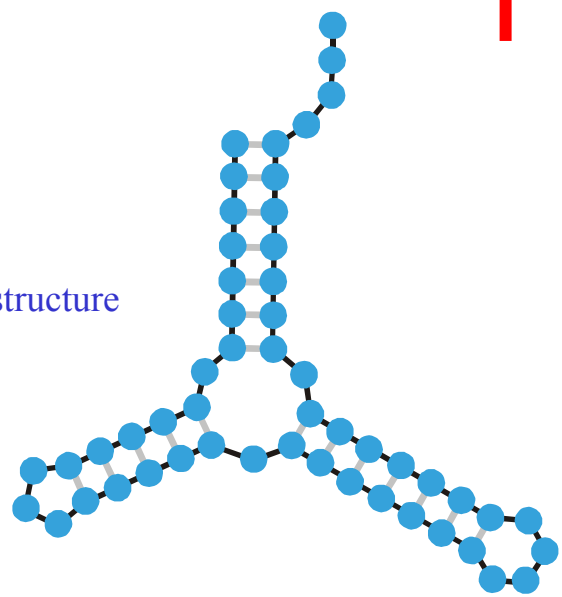
Biophysical chemistry: thermodynamics and kinetics

Empirical parameters

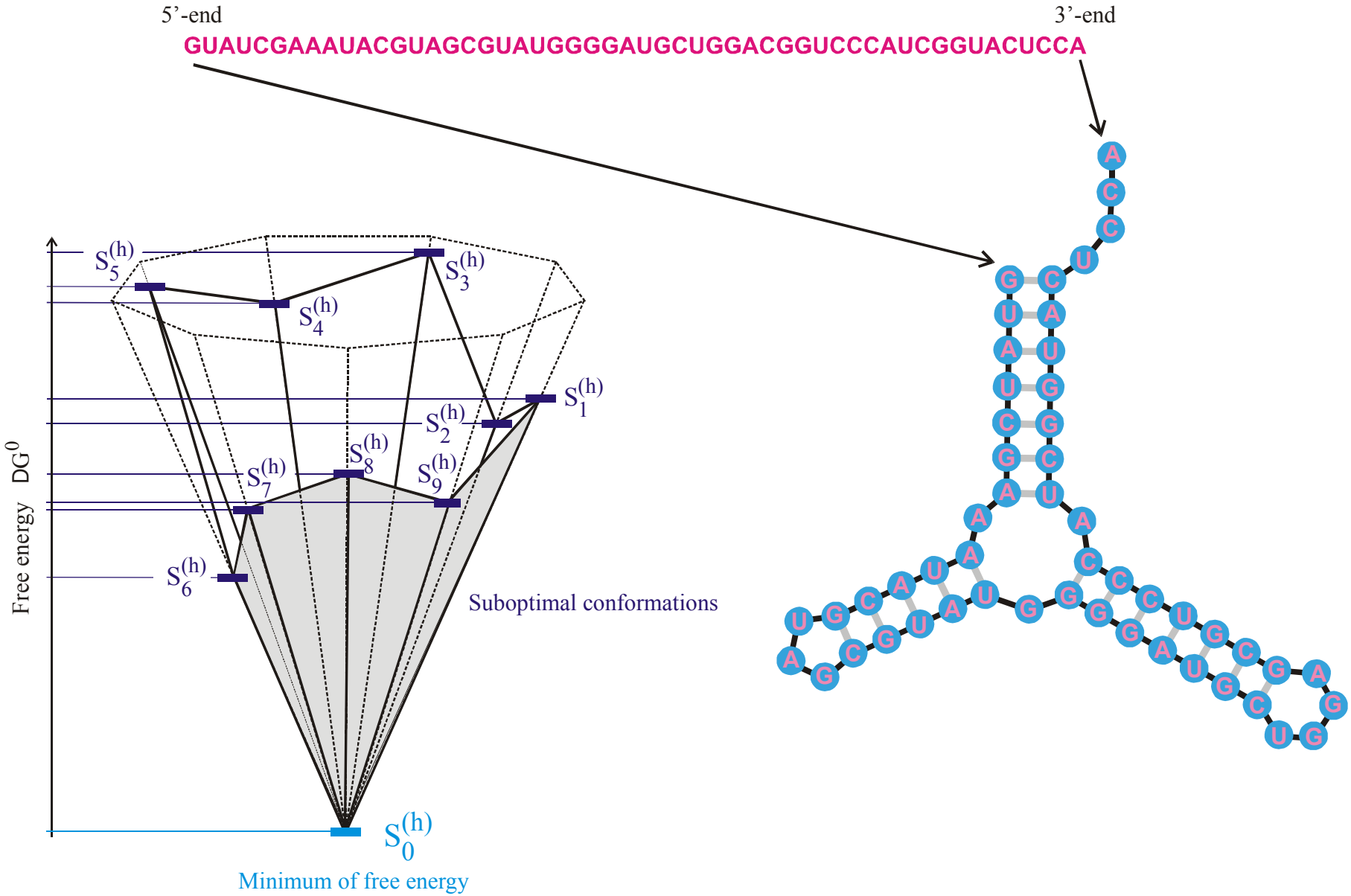
Inverse folding of RNA:

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

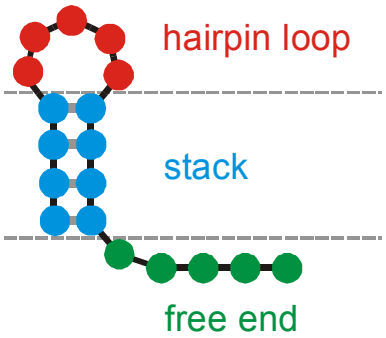
RNA structure

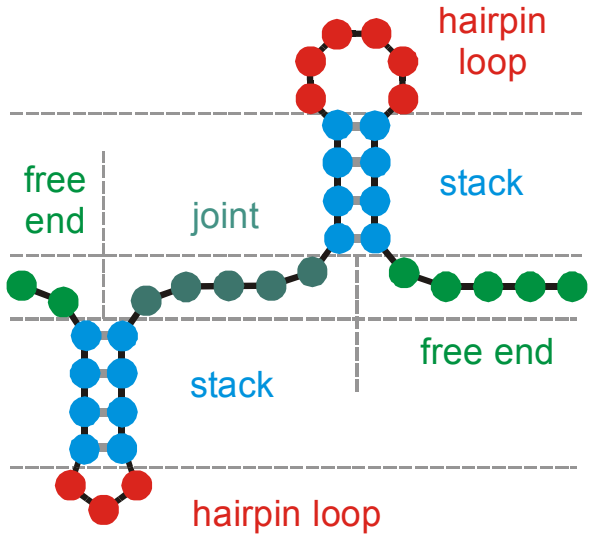


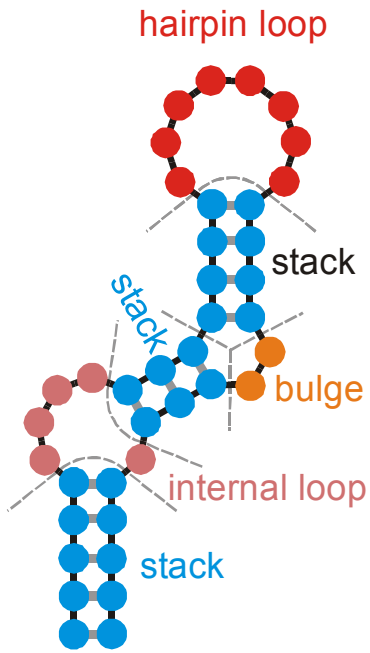
Sequence, structure, and function

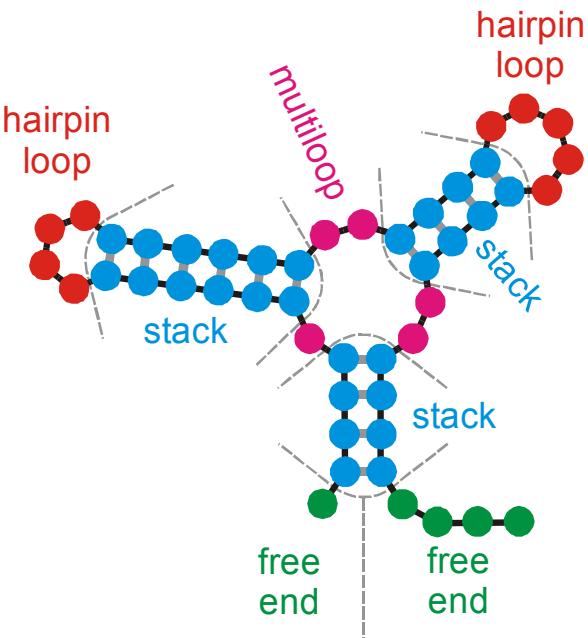


The minimum free energy structures on a discrete space of conformations

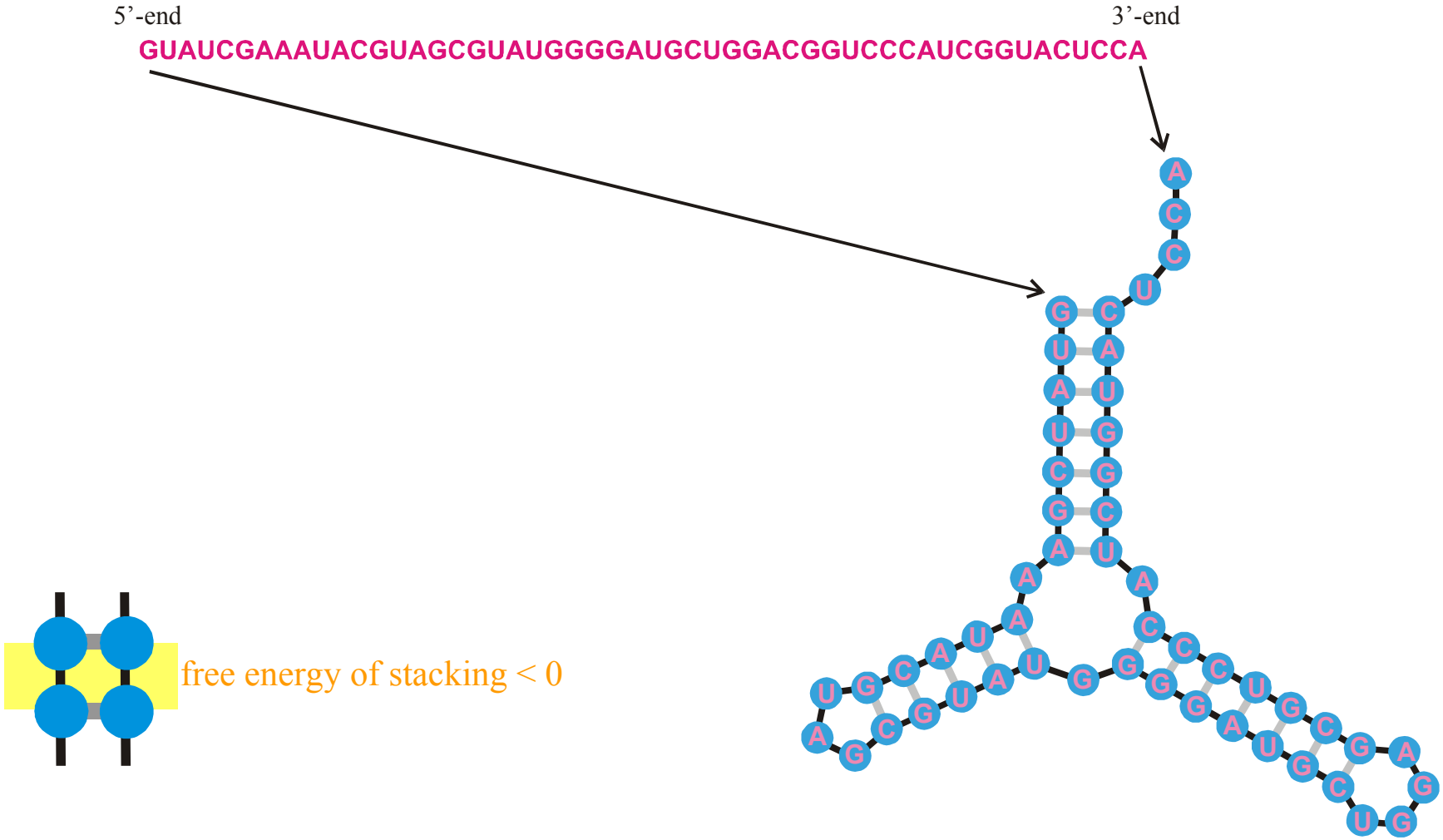








Elements of RNA secondary structures
as used in free energy calculations



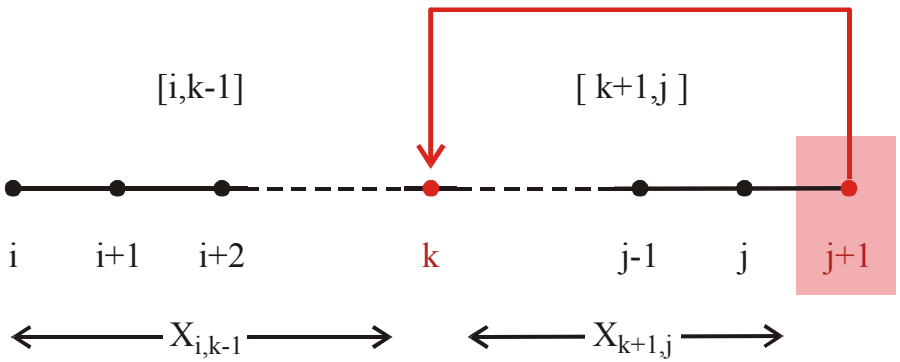
$$\Delta G_0^{300} = \underbrace{\sum g_{ij,kl}}_{\text{stacks of base pairs}} + \underbrace{\sum h(n_l)}_{\text{hairpin loops}} + \underbrace{\sum b(n_b)}_{\text{bulges}} + \underbrace{\sum i(n_i)}_{\text{internal loops}} + \cdots$$

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

Maximum matching

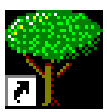
An example of a **dynamic programming** computation of the maximum number of base pairs

Back tracking yields the structure(s).



$$X_{i,j+1} = \max\left\{ X_{i,j}, \max_{i \le k \le j-1}\left((X_{i,k-1} + 1 + X_{k+1,j})\,\rho_{k,j+1} \right) \right\}$$

Minimum free energy computations are based on empirical energies



RNAStudio.lnk

GGCGCGCCCGGCGCC

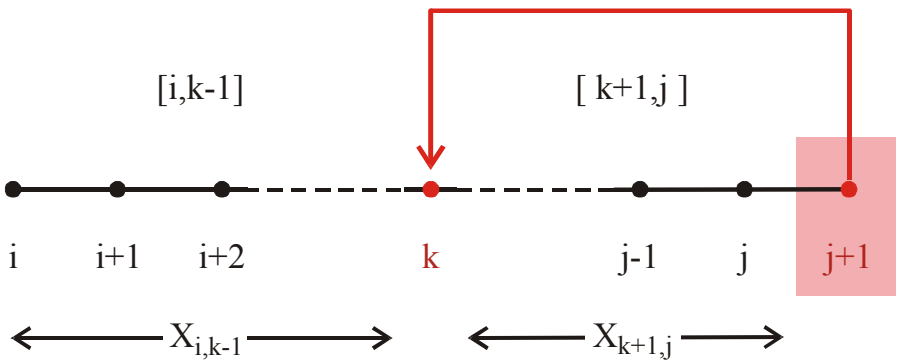
GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

UGGUUACGCGUUGGGGUAACGAAGAUUCCGAGAGGAGUUUAGUGACUAGAGG

Maximum matching

An example of a **dynamic programming** computation of the maximum number of base pairs

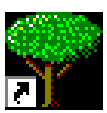
Back tracking yields the structure(s).



i		j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
			G	G	C	G	C	G	C	C	C	G	G	C	G	C	C
1	G		*	*	1	1	1	1	2	3	3	3	4	4	5	6	6
2	G			*	*	0	1	1	2	2	2	3	3	4	4	5	6
3	C				*	*	0	1	1	1	2	3	3	3	4	5	5
4	G					*	*	0	1	1	2	2	2	3	4	5	5
5	C						*	*	0	1	1	2	2	3	4	4	4
6	G							*	*	1	1	1	2	3	3	3	4
7	C								*	*	0	1	2	2	2	2	3
8	C									*	*	1	1	1	2	2	2
9	C										*	*	1	1	2	2	2
10	G											*	*	1	1	1	2
11	G												*	*	0	1	1
12	C													*	*	0	1
13	G														*	*	1
14	C															*	*
15	C																*

$$X_{i,j+1} = \max\left\{ X_{i,j}, \max_{i \le k \le j-1}\left((X_{i,k-1} + 1 + X_{k+1,j})\,\rho_{k,j+1} \right) \right\}$$

Minimum free energy computations are based on empirical energies



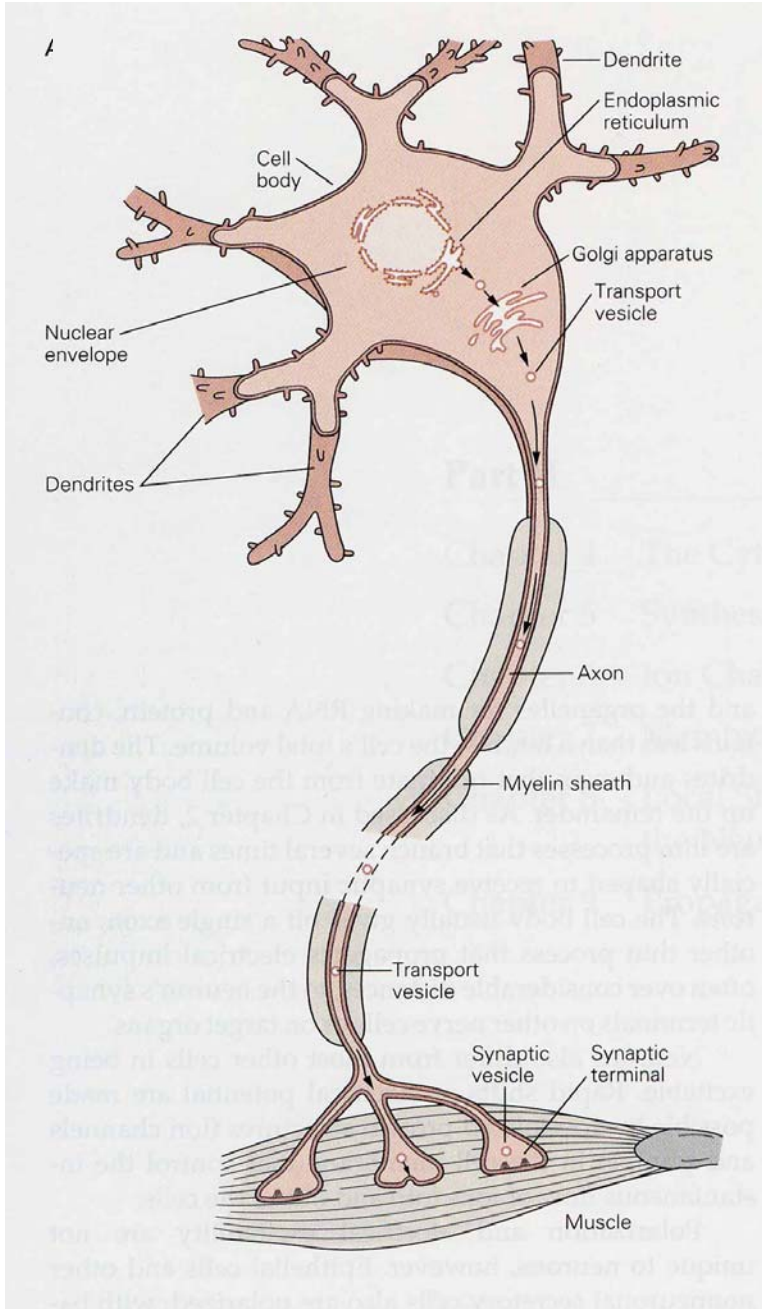
RNAStudio.lnk

GGCGCGCCCGGCGCC

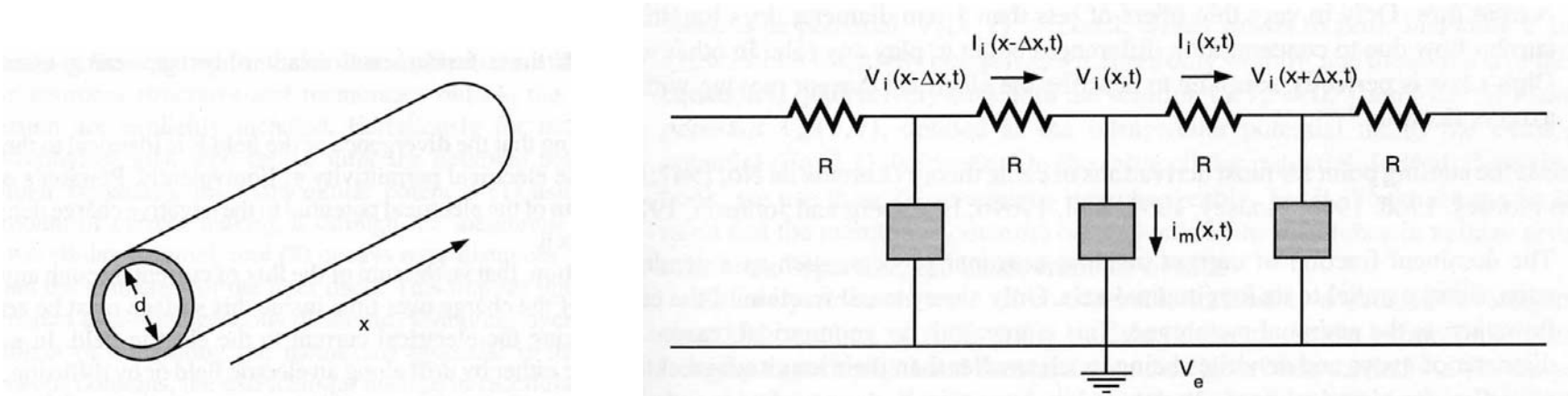
GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

UGGUUACGCGUUGGGGUAACGAAGAUUCCGAGAGGAGUUUAGUGACUAGAGG

1. Komplexität in der Biologie
2. Evolutionäre Optimierung und Lernen im Ensemble
3. Strukturbildung von Biomolekülen als kombinatorisches Problem
4. **Modellbildung in der Neurobiologie**



A single neuron signaling to a muscle fiber



A **B**

Fig. 2.2 ELECTRICAL STRUCTURE OF A CABLE **(A)** Idealized cylindrical axon or dendrite at the heart of one-dimensional cable theory. Almost all of the current inside the cylinder is longitudional due to geometrical (the radius is much smaller than the length of the cable) and electrical factors (the membrane covering the axon or dendrite possesses a very high resistivity compared to the intracellular cytoplasm). As a consequence, the radial and angular components of the current can be neglected, and the problem of determining the potential in these structures can be reduced from three spatial dimensions to a single one. On the basis of the bidomain approximation, gradients in the extracellular potentials are neglected and the cable problem is expressed in terms of the transmembrane potential $V_m(x,t) = V_i(x,t) - V_e$. **(B)** Equivalent electrical structure of an arbitrary neuronal process. The intracellular cytoplasm is modeled by the purely ohmic resistance R. This tacitly assumes that movement of carriers is exclusively due to drift along the voltage gradient and not to diffusion. Here and in the following the extracellular resistance is assumed to be negligible and V_e is set to zero. The current per unit length across the membrane, whether it is passive or contains voltage-dependent elements, is described by i_m and the system is characterized by the second-order differential equation, Eq. 2.5.

Christof Koch, Biophysics of Computation. Information Processing in single neurons.
Oxford University Press, New York 1999.

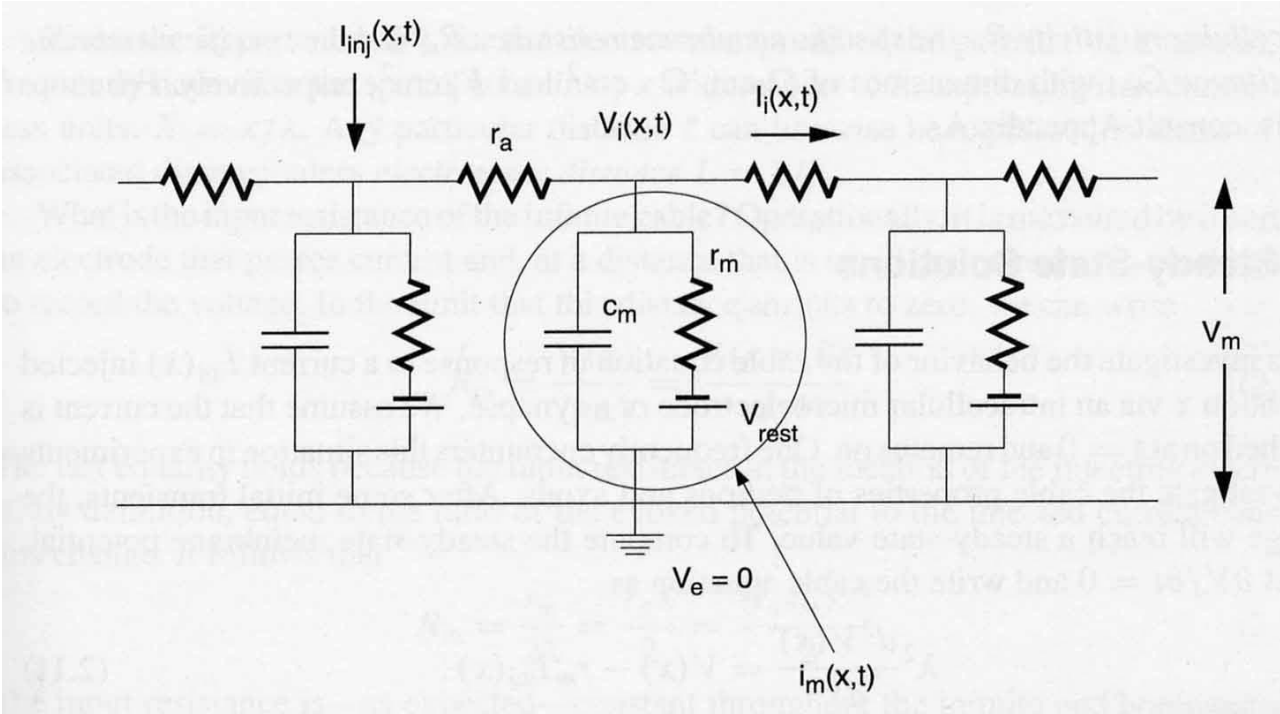


Fig. 2.3 A SINGLE PASSIVE CABLE Equivalent lumped electrical circuit of an elongated neuronal fiber with passive membrane. The intracellular cytoplasm is described by an ohmic resistance per unit length r_a and the membrane by a capacitance c_m in parallel with a passive membrane resistance r_m and a battery V_{rest}. The latter two components are frequently referred to as *leak resistance* and *leak battery*. An external current $I_{\text{inj}}(x,t)$ is injected into the cable. The associated linear cable equation (Eq. 2.7) describes the dynamics of the electrical potential $V_m = V_i - V_e$ along the cable.

Christof Koch, Biophysics of Computation. Information Processing in single neurons. Oxford University Press, New York 1999.

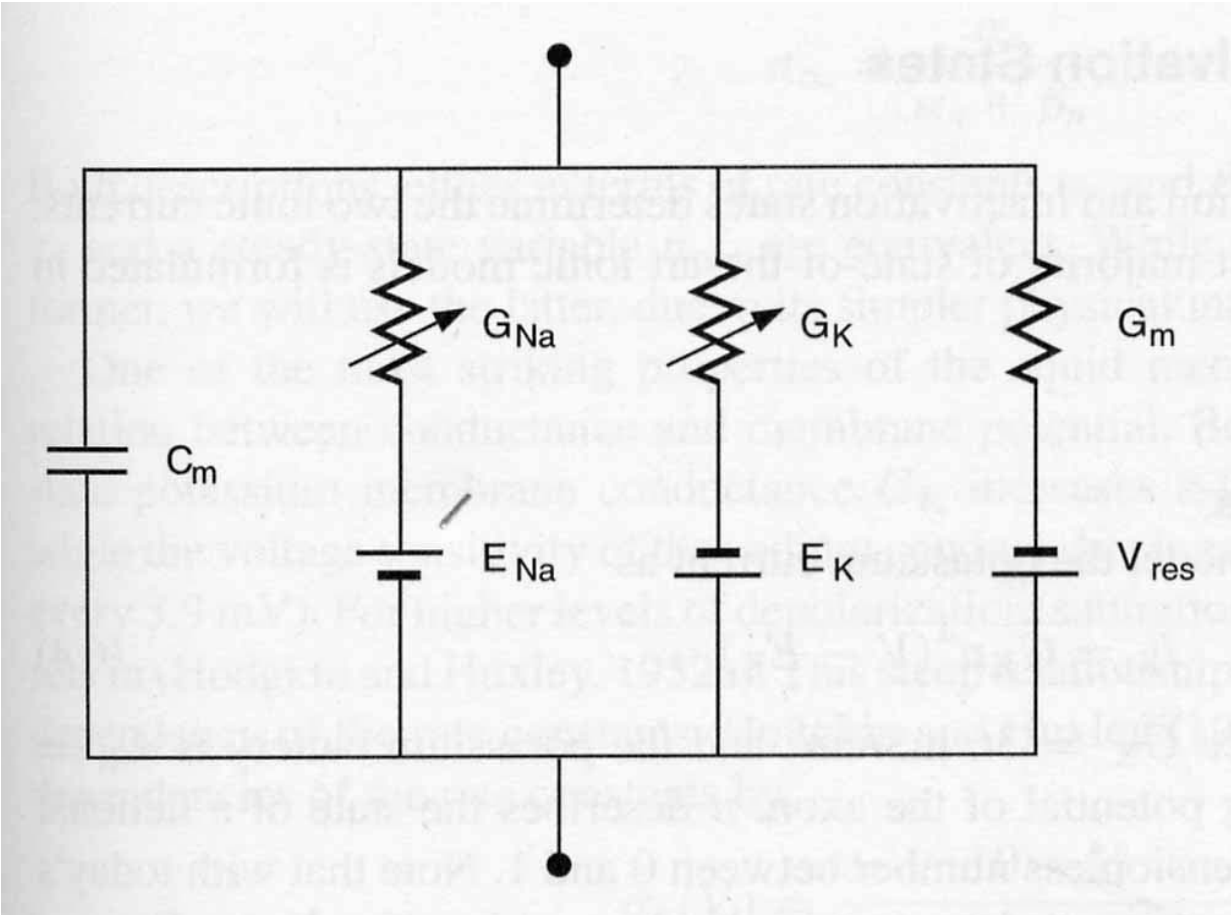


Fig. 6.2 Electrical Circuit for a Patch of Squid Axon Hodgkin and Huxley modeled the membrane of the squid axon using four parallel branches: two passive ones (membrane capacitance C_m and the leak conductance $G_m = 1/R_m$) and two time- and voltage-dependent ones representing the sodium and potassium conductances.

Christof Koch, Biophysics of Computation. Information Processing in single neurons. Oxford University Press, New York 1999.

$$\frac{dV}{dt} = \frac{1}{C_M}\left[I - \bar{g}_{Na}\, m^3\, h\,(V - V_{Na}) - \bar{g}_K\, n^4\,(V - V_K) - \bar{g}_l\,(V - V_l)\right]$$

$$\frac{dm}{dt} = \alpha_m\,(1-m) - \beta_m\, m$$

$$\frac{dh}{dt} = \alpha_h\,(1-h) - \beta_h\, h$$

$$\frac{dn}{dt} = \alpha_n\,(1-n) - \beta_n\, n$$

Hogdkin-Huxley OD equations

Hhsim.lnk

Simulation of space independent Hodgkin-Huxley equations:
Voltage clamp and constant current

$$\frac{1}{R}\frac{\partial^2 V}{\partial x^2} = C\frac{\partial V}{\partial t} + \left[g_{Na}\, m^3\, h (V - V_{Na}) + g_K\, n^4\, (V - V_K) + g_l\, (V - V_l) \right] 2\pi\, r\, L$$

$$\frac{\partial m}{\partial t} = \alpha_m (1 - m) - \beta_m\, m$$

$$\frac{\partial h}{\partial t} = \alpha_h (1 - h) - \beta_h\, h$$

$$\frac{\partial n}{\partial t} = \alpha_n (1 - n) - \beta_n\, n$$

Hodgkin-Huxley PDEquations

Travelling pulse solution: $V(x,t) = V(\mathsf{x})$ with

$$\mathsf{x} = x + \mathsf{q}\, t$$

Hodgkin-Huxley equations describing pulse propagation along nerve fibers

$$\frac{1}{R}\frac{d^2V}{d\xi^2} = C_M\,\theta\frac{dV}{d\xi} + \left[\,\bar{g}_{Na}\,m^3\,h(V-V_{Na}) + \bar{g}_K\,n^4\,(V-V_K) + \bar{g}_l\,(V-V_l)\right]\,2\pi r L$$

$$\theta\,\frac{dm}{d\xi} = \alpha_m\,(1-m) - \beta_m\,m$$

$$\theta\,\frac{dh}{d\xi} = \alpha_h\,(1-h) - \beta_h\,h$$

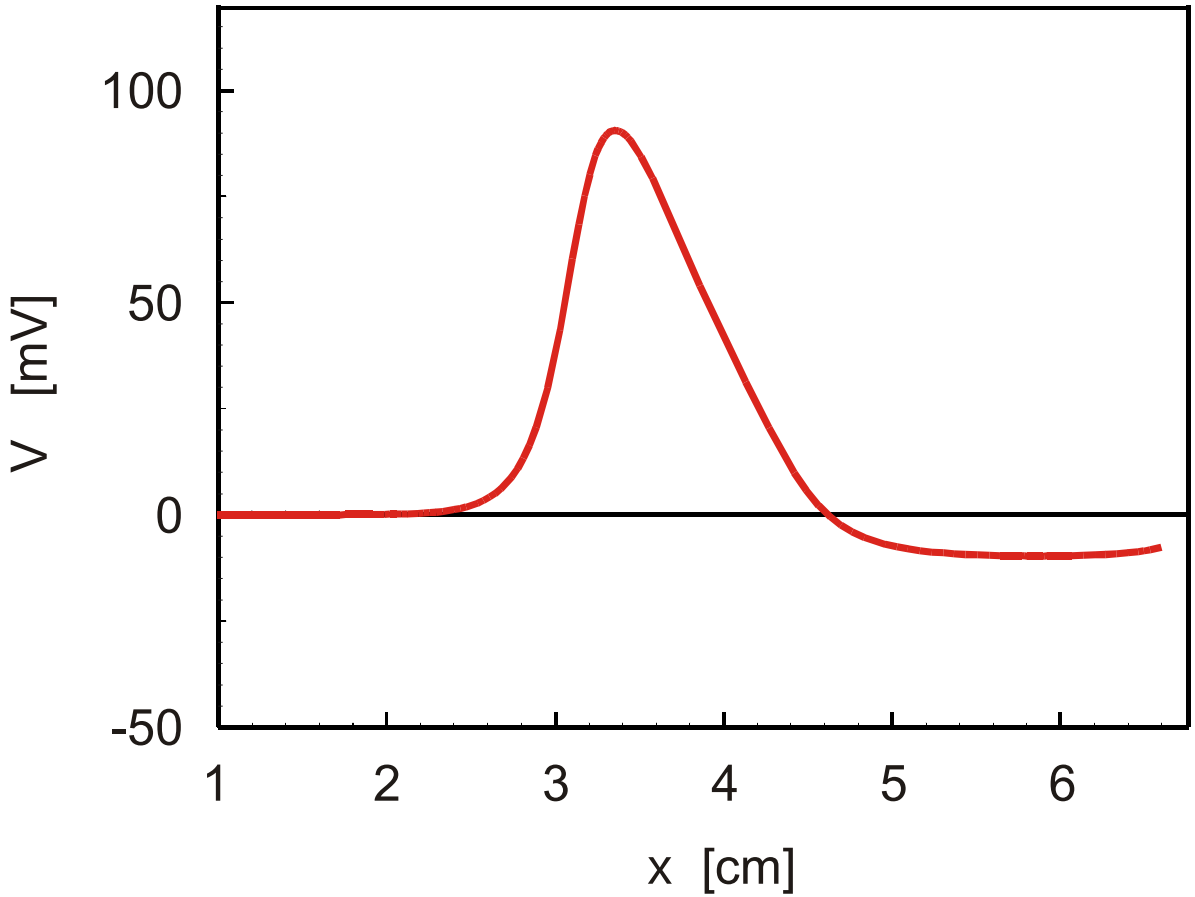
$$\theta\,\frac{dn}{d\xi} = \alpha_n\,(1-n) - \beta_n\,n$$

Hodgkin-Huxley PDEquations

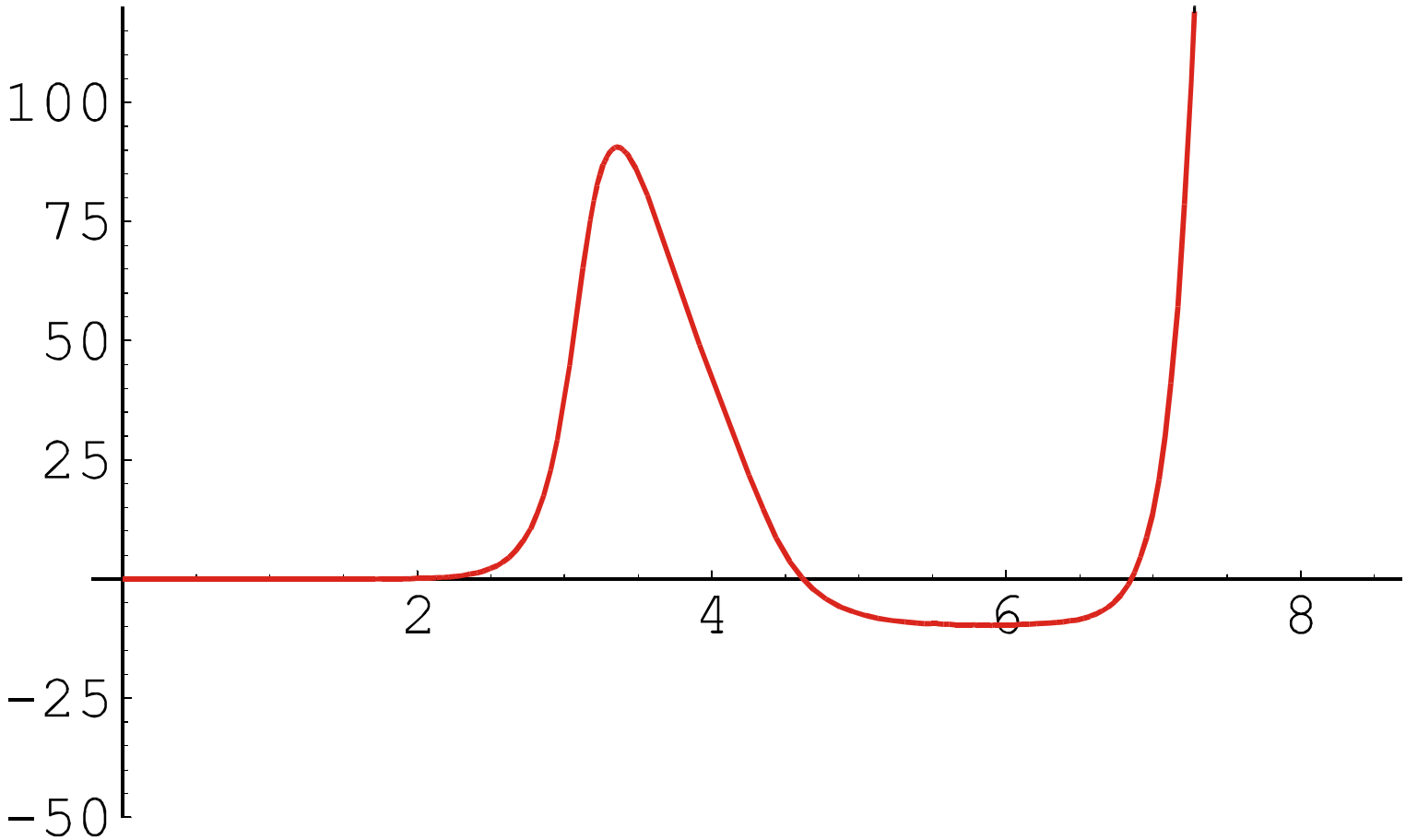
Travelling pulse solution: $V(x,t) = V(\mathsf{x})$ with

$$\mathsf{x} = x + \mathsf{q}\,t$$

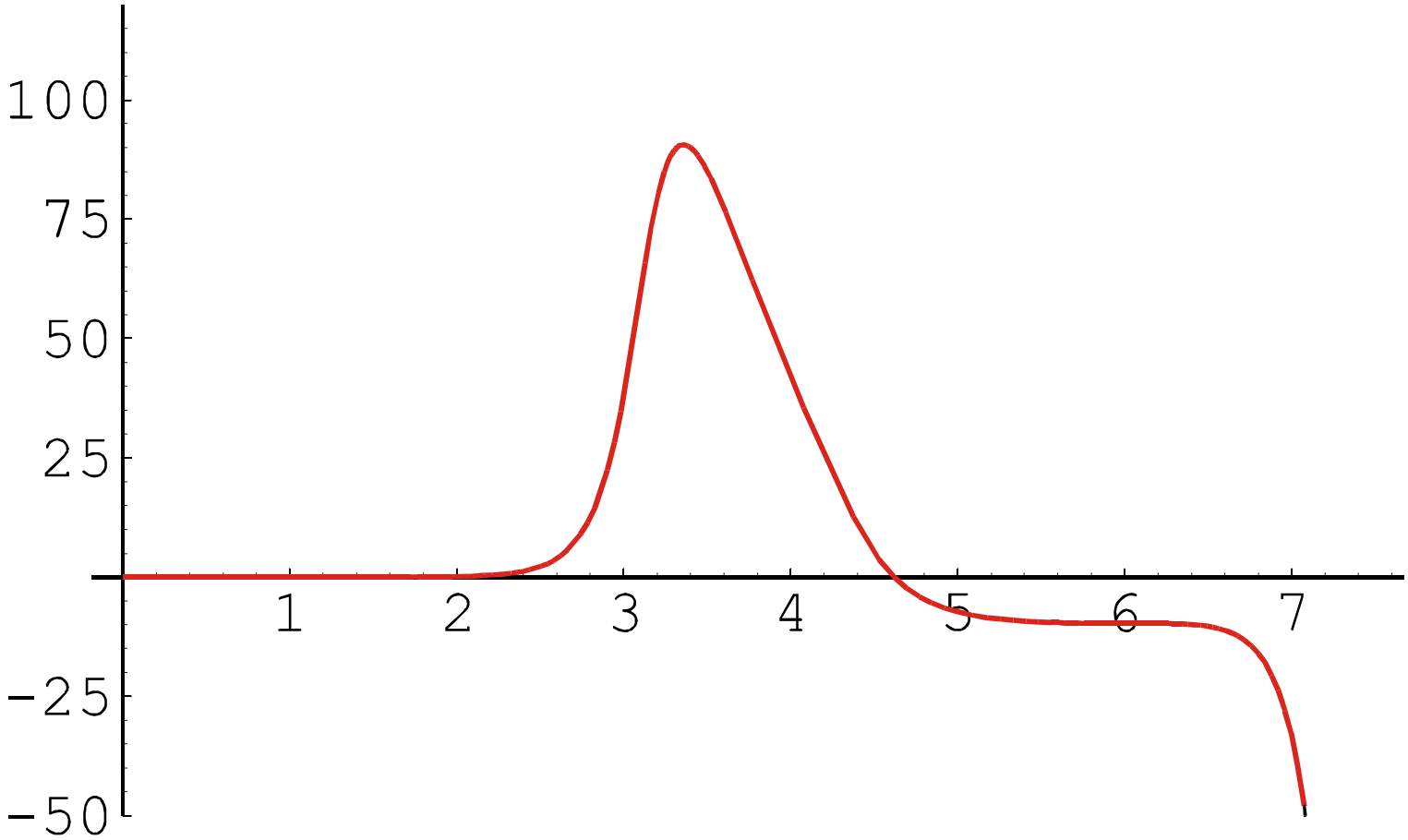
Hodgkin-Huxley equations describing pulse propagation along nerve fibers



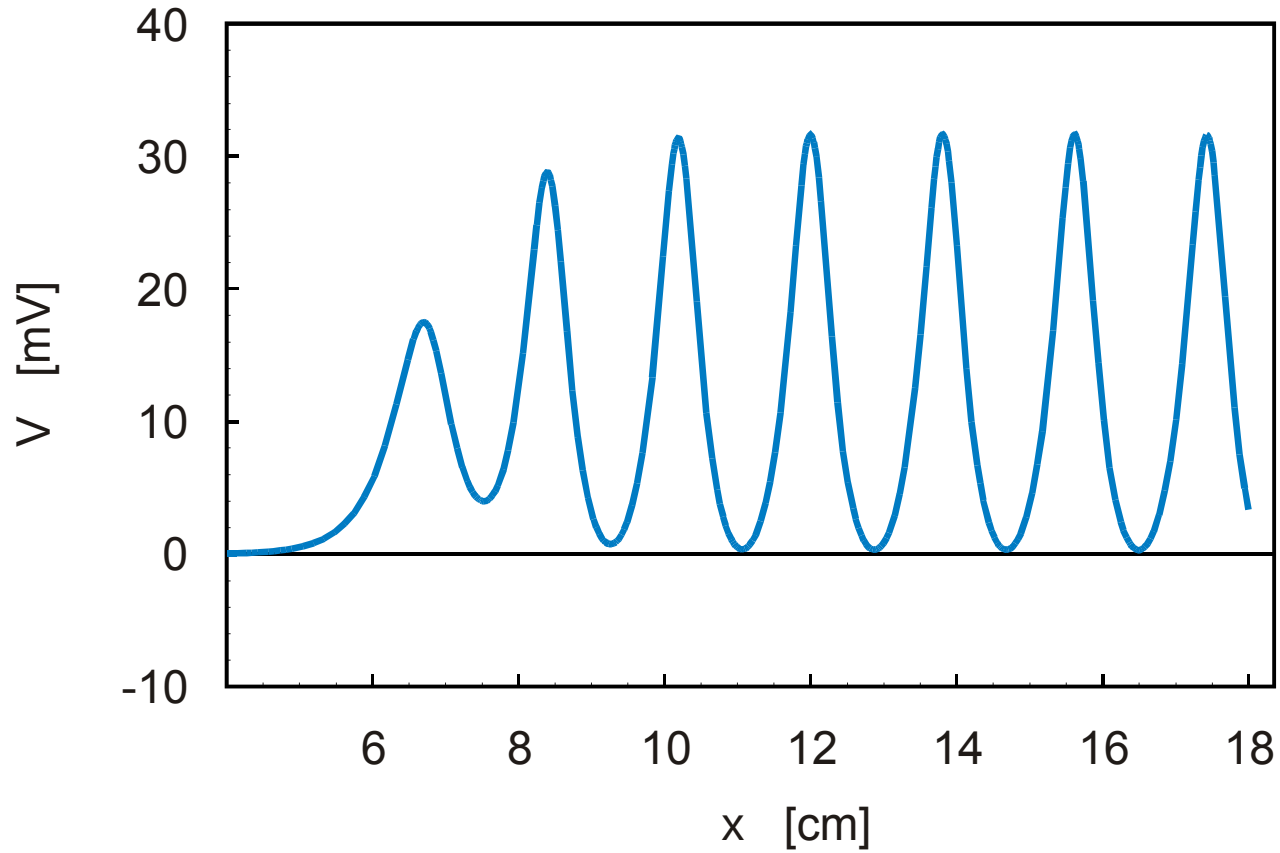
T = 18.5 C; θ = 1873.33 cm / sec



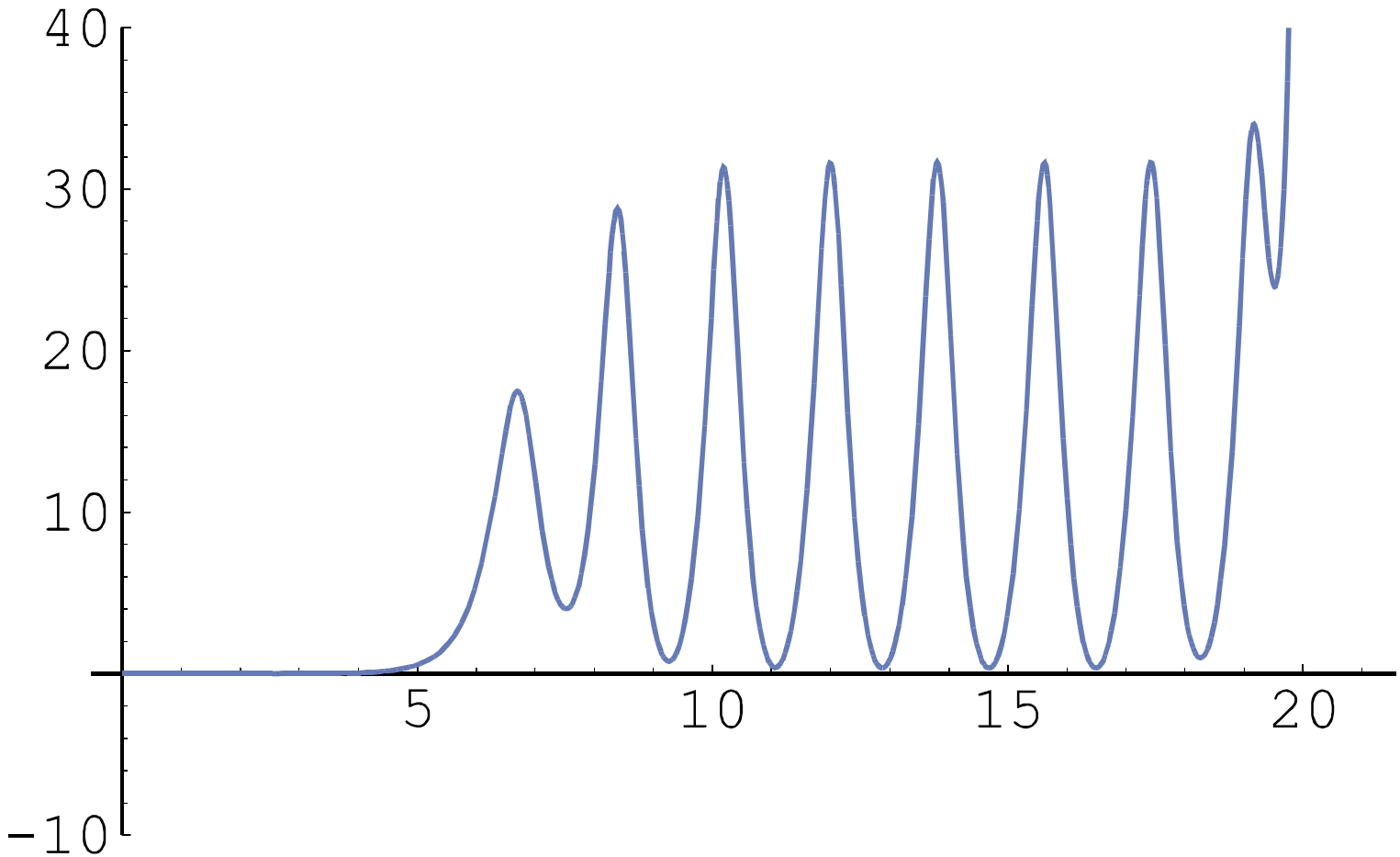
T = 18.5 C; θ = 1873.3324514717698 cm / sec



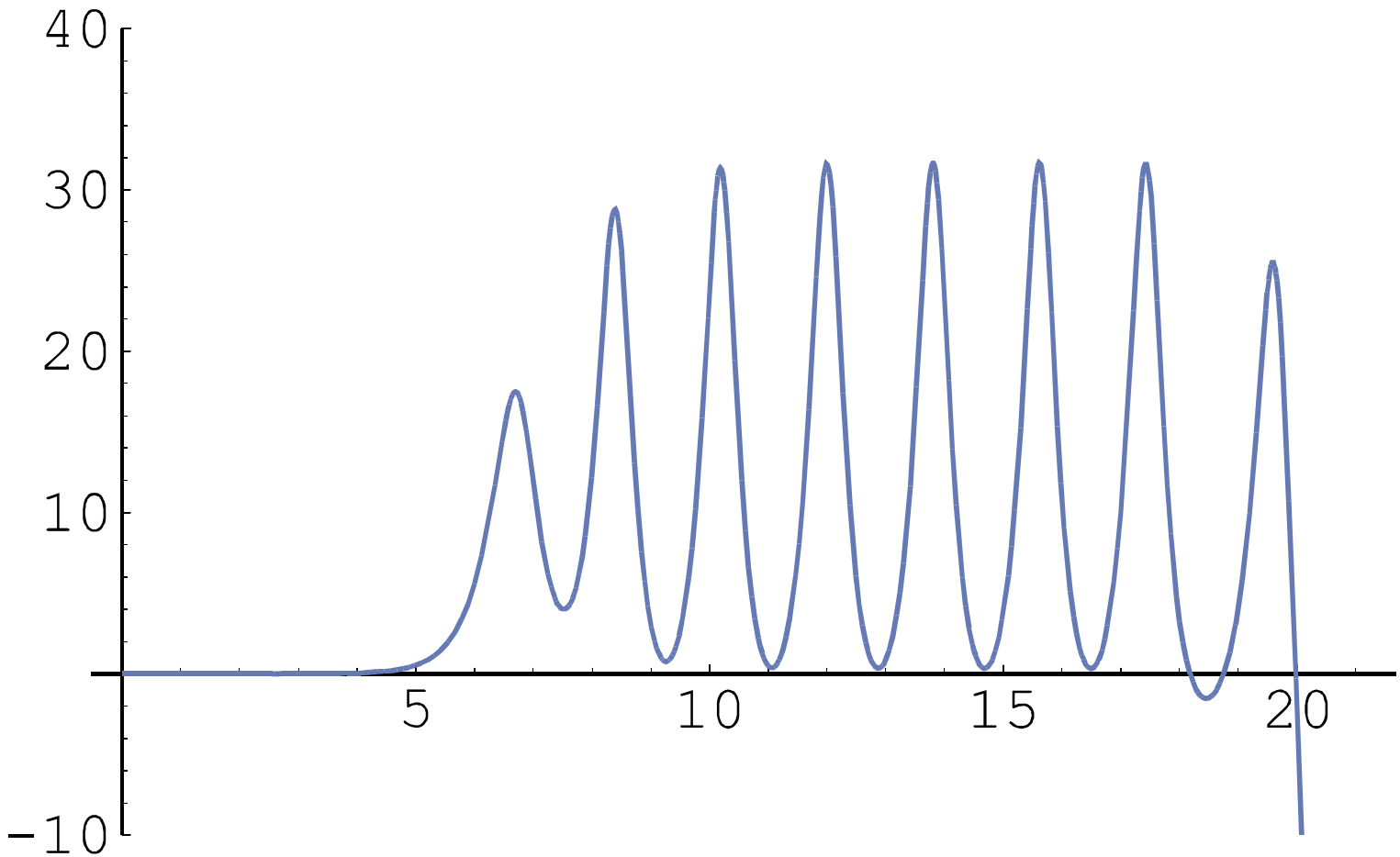
T = 18.5 C; θ = 1873.3324514717697 cm / sec



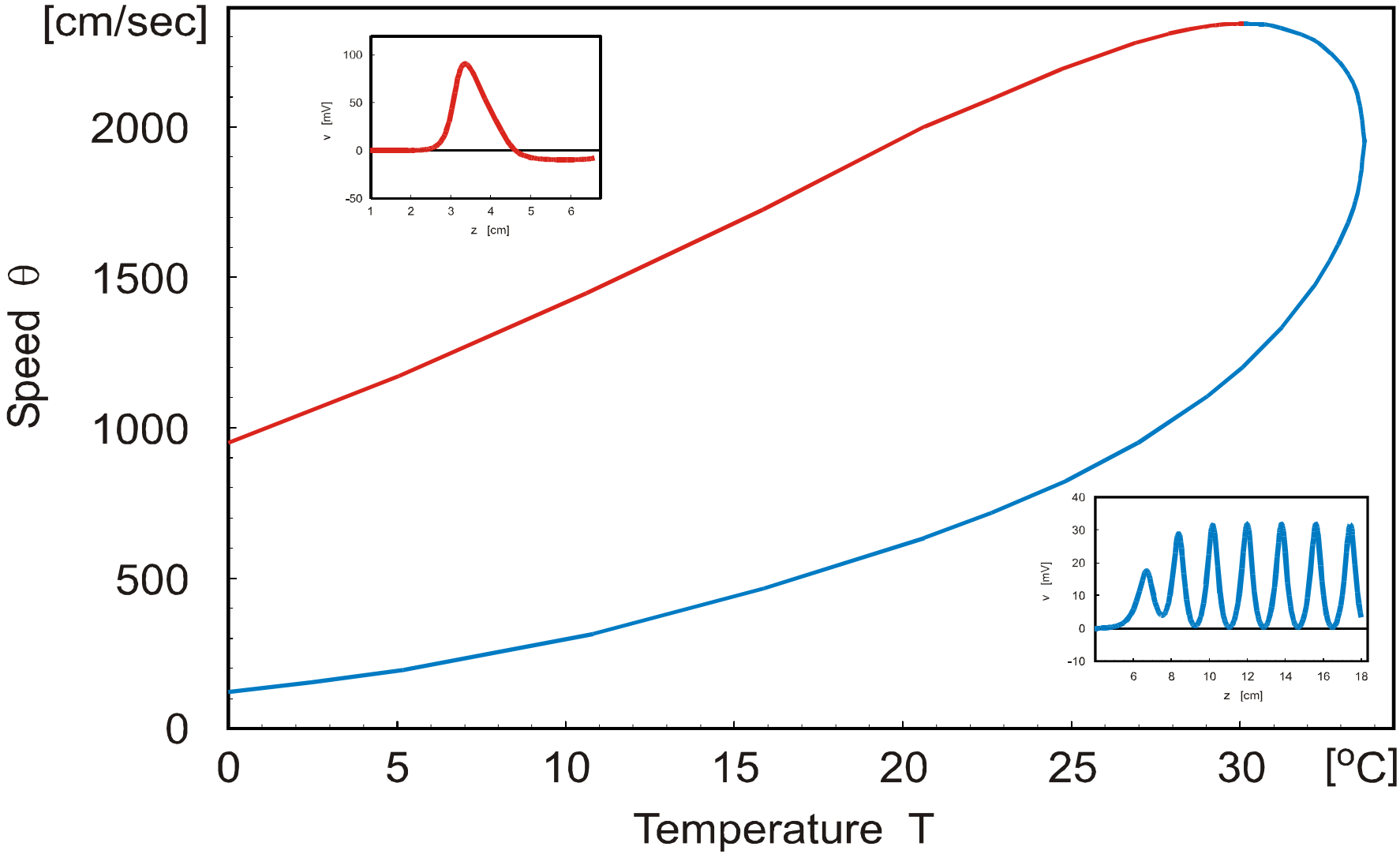
T = 18.5 C; θ = 544.070 cm / sec



T = 18.5 C; θ = 554.070286919319 cm/sec



T = 18.5 C; θ = 554.070286919320 cm/sec



Propagating wave solutions of the Hodgkin-Huxley equations

FitzHugh-Nagumo Equations

$$\frac{\partial V'}{\partial t'} = -V'(V'-a)(V'-1) - Y - \frac{\partial^2 V'}{\partial x'^2}$$

$$\frac{\partial Y'}{\partial t'} = bV' \text{ all quantities dimensionaless}$$

$$\frac{d}{d\zeta}\left[\frac{d^2V'}{d\zeta^2} - c\frac{dV'}{d\zeta} - V'(V'-a)(V'-1)\right] - \frac{b}{c}V' = 0, \text{ with } \zeta = x' + ct'$$

$$c(a, b=0) = \frac{1}{\sqrt{2}}(1-2a) - \text{Propagation of the Pulse Front}$$

$c(a,b) < c(a,0)$ Front slows down to shape the back of the pulse

The good news: The pulse, stationary in the frame of ζ, is now described by an ordinary third order differential equation. The bad news is that the result is a structurally unstable homoclinic orbit, which is why calculations take a bit of care. We have found an analytic expression for $c(a,b)$.

Connection to Hodgkin-Huxley: $[h,n]$ are relatively slow variables, so keep them at resting values. This results in a contracted two-dimensional Hodgkin-Huxley $[V,m]$ system which describes the pulse front at the full speed θ to very close approximation. The two descriptions are now equivalent under the transformations

$$D = \frac{r}{2R_2C_m}\ [336\ \frac{\text{cm}^2}{\text{sec}}] \quad \Gamma = \sqrt{\frac{G_{\text{Na}}h(0)}{C_mD}}\ [14.5876\ \text{cm}^{-}1]$$

$$V' = \frac{V}{E_{\text{Na}}}, E_{\text{Na}} = 115\ \text{mv},\ \zeta = \Gamma z,\ c = \frac{\theta}{\Gamma D}\ [\Gamma D = 4903\ \frac{cm}{sec}$$

Advantage: Provides a bridge between between detailed Hodgkin-Huxley based conductance models and formal spiking models which dispense with such details [a pulse is regarded as a delta-function-like spike].

Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)

Projects No. 09942, 10578, 11065, 13093, 13887, and 14898

Jubiläumsfonds der Österreichischen Nationalbank

Project No. Nat-7813

European Commission: Project No. EU-980189

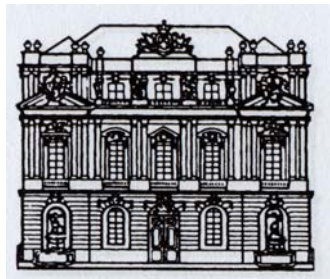
Siemens AG, Austria

Österreichische Akademie der Wissenschaften

The Santa Fe Institute and the Universität Wien

The software for producing RNA movies was developed by
Robert Giegerich and coworkers at the Universität Bielefeld

Universität Wien



Österreichische Akademie der Wissenschaften

Universität Wien

Coworkers

Walter Fontana, Santa Fe Institute, NM

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

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

Ivo L.Hofacker, Christoph Flamm, Universität Wien, AT

Andreas Wernitznig, **Michael Kospach,** Universität Wien, AT
Ulrike Langhammer, Ulrike Mückstein, Stefanie Widder
Jan Cupal, Kurt Grünberger, Andreas Svrček-Seiler, Stefan Wuchty

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

Web-Page for further information:

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