



# Selektive Neutralität und Effizienz der Evolution

Was wir aus Evolutionsexperimenten lernen können

Peter Schuster

Institut für Theoretische Chemie und Molekulare  
Strukturbiologie der Universität Wien

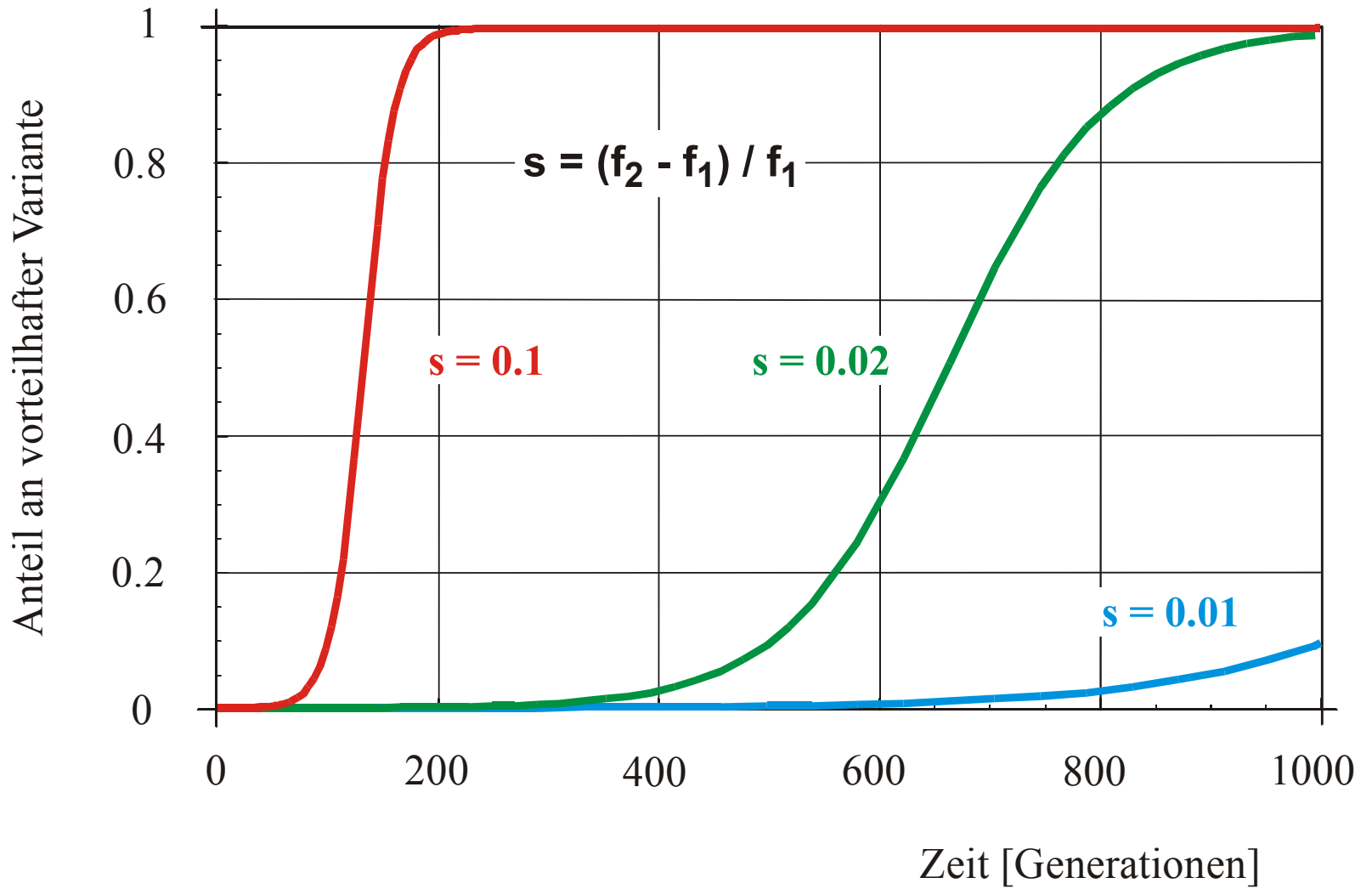


Seminar des Naturhistorischen Museums

Wien, 28.04.2004

Web-Page for further information:

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



Selektion vorteilhafter Varianten in einer Population von  $N = 10\,000$  Individuen



Massif Central



Mount Fuji

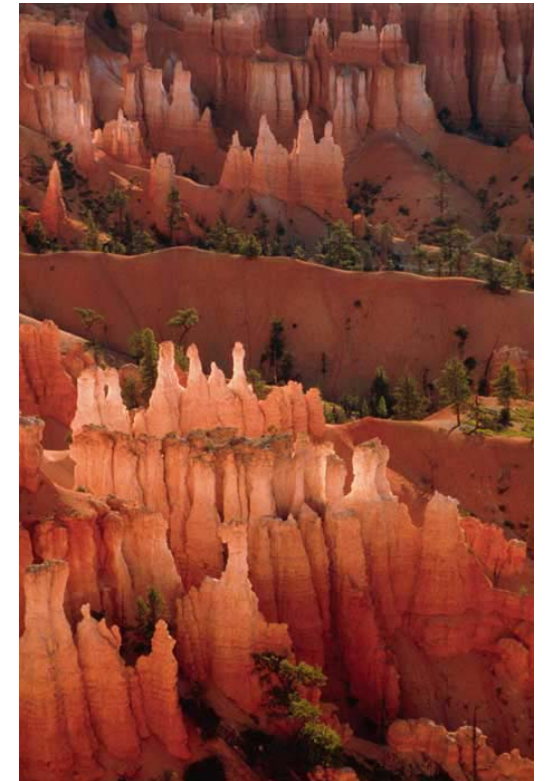
Beispiele glatter Landschaften



Dolomiten

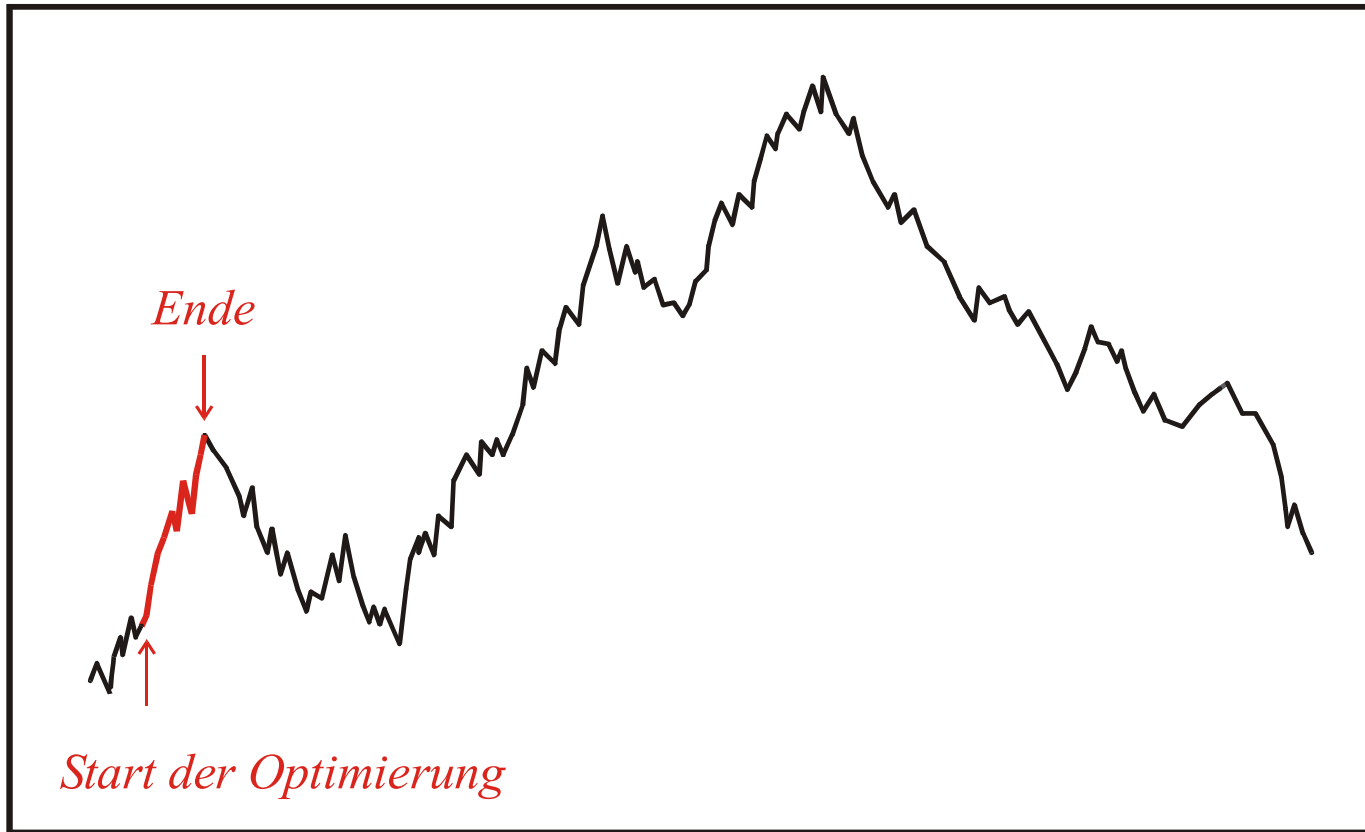


Beispiele zerklüfteter Landschaften



Bryce Canyon

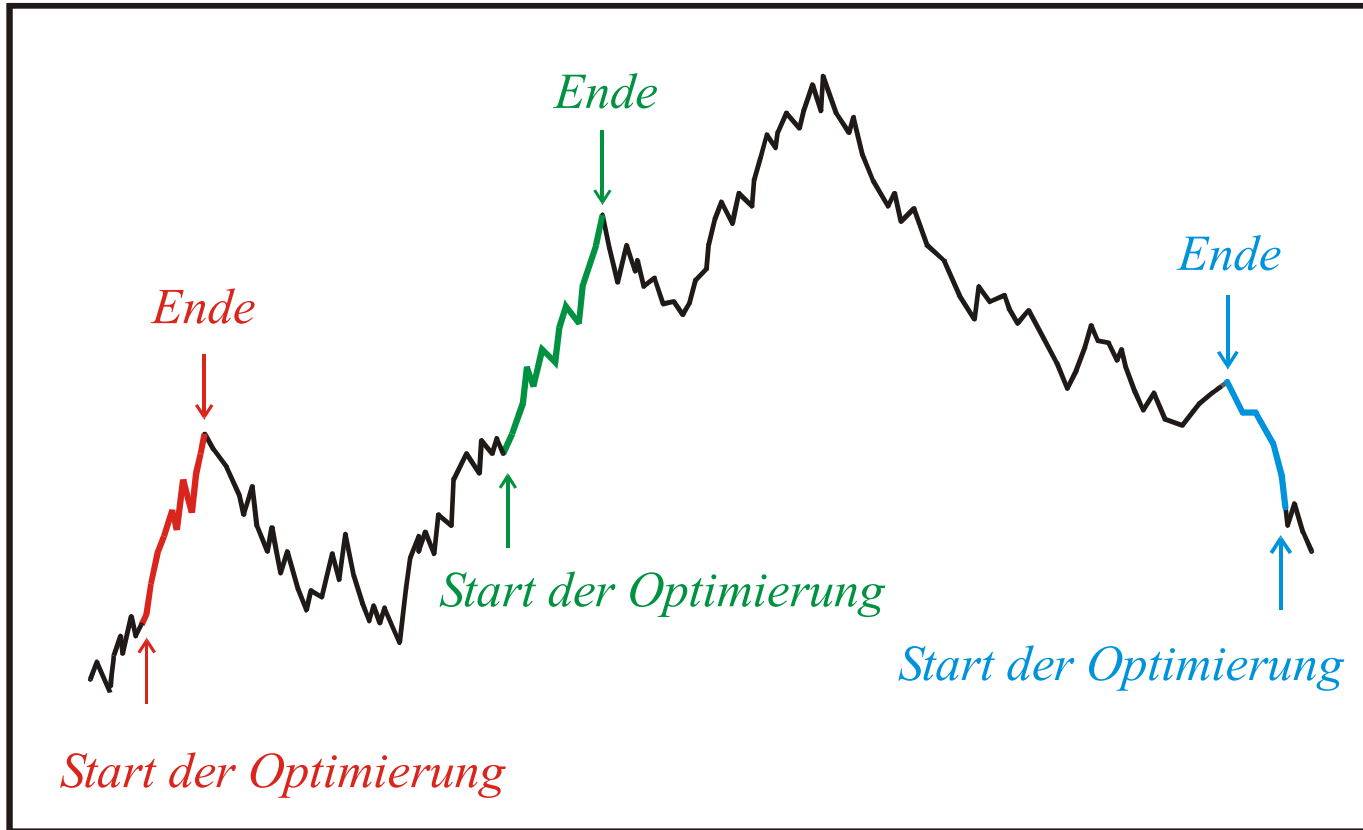
Mittlere Fitness



Sequenzraum

Optimierung auf einer Fitnesslandschaft ohne selektive Neutralität

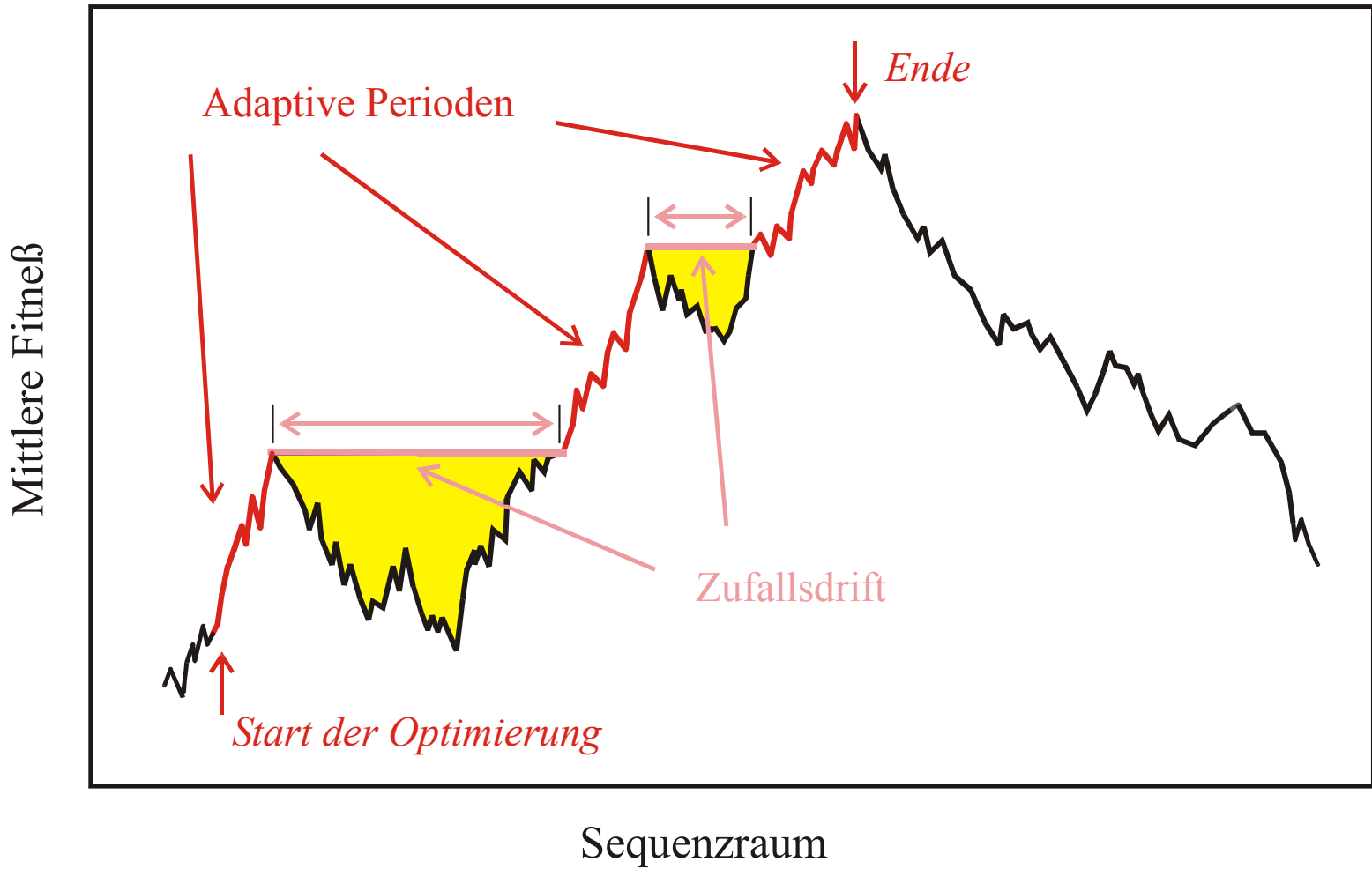
Mittlere Fitness



Sequenzraum

Optimierung auf einer Fitnesslandschaft ohne selektive Neutralität



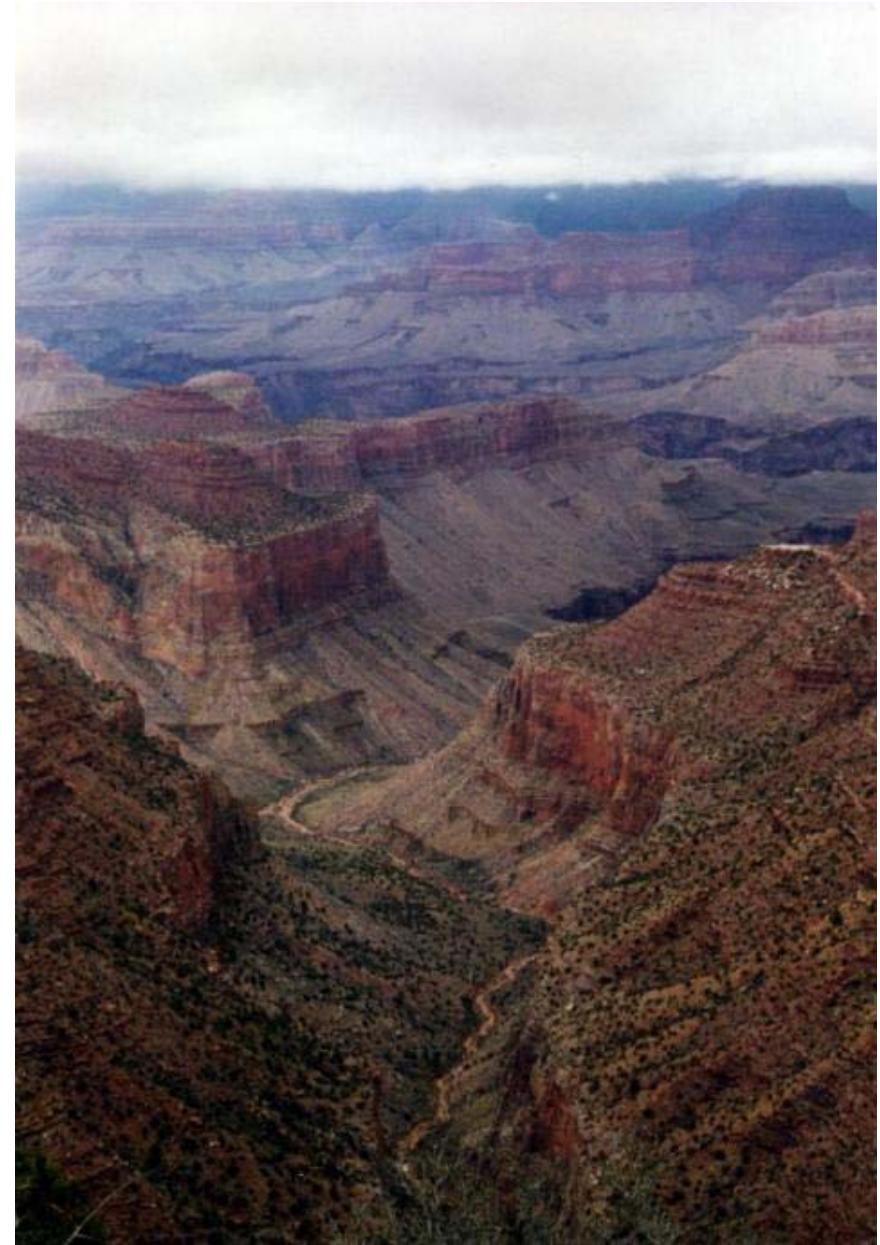


Evolutionäre Optimierung auf einer Landschaft mit neutralen Zonen



Grand Canyon

Beispiel einer Landschaft mit neutralen Graten  
und Plateaus



Neutrale Grate und Plateaus



*„... Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed, owing to the nature of the organism and the nature of the conditions. ...“*

Charles Darwin, Origin of species (1859)



## The molecular clock of evolution

Motoo Kimura's population genetics of neutral evolution.

Evolutionary rate at the molecular level.  
*Nature* **217**: 624-626, 1955.

*The Neutral Theory of Molecular Evolution.*  
Cambridge University Press. Cambridge, UK,  
1983.

---

---

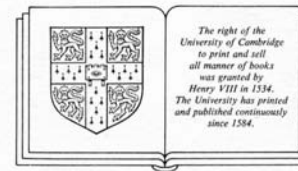
# THE NEUTRAL THEORY OF MOLECULAR EVOLUTION

---

---

MOTOO KIMURA

*National Institute of Genetics, Japan*



CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York New Rochelle

Melbourne Sydney

80

SONDERDRUCK

aus

Jahrbuch 1967 der Max-Planck-Gesellschaft  
zur Förderung der Wissenschaften e.V.

\*

Molekularbiologie und Evolution

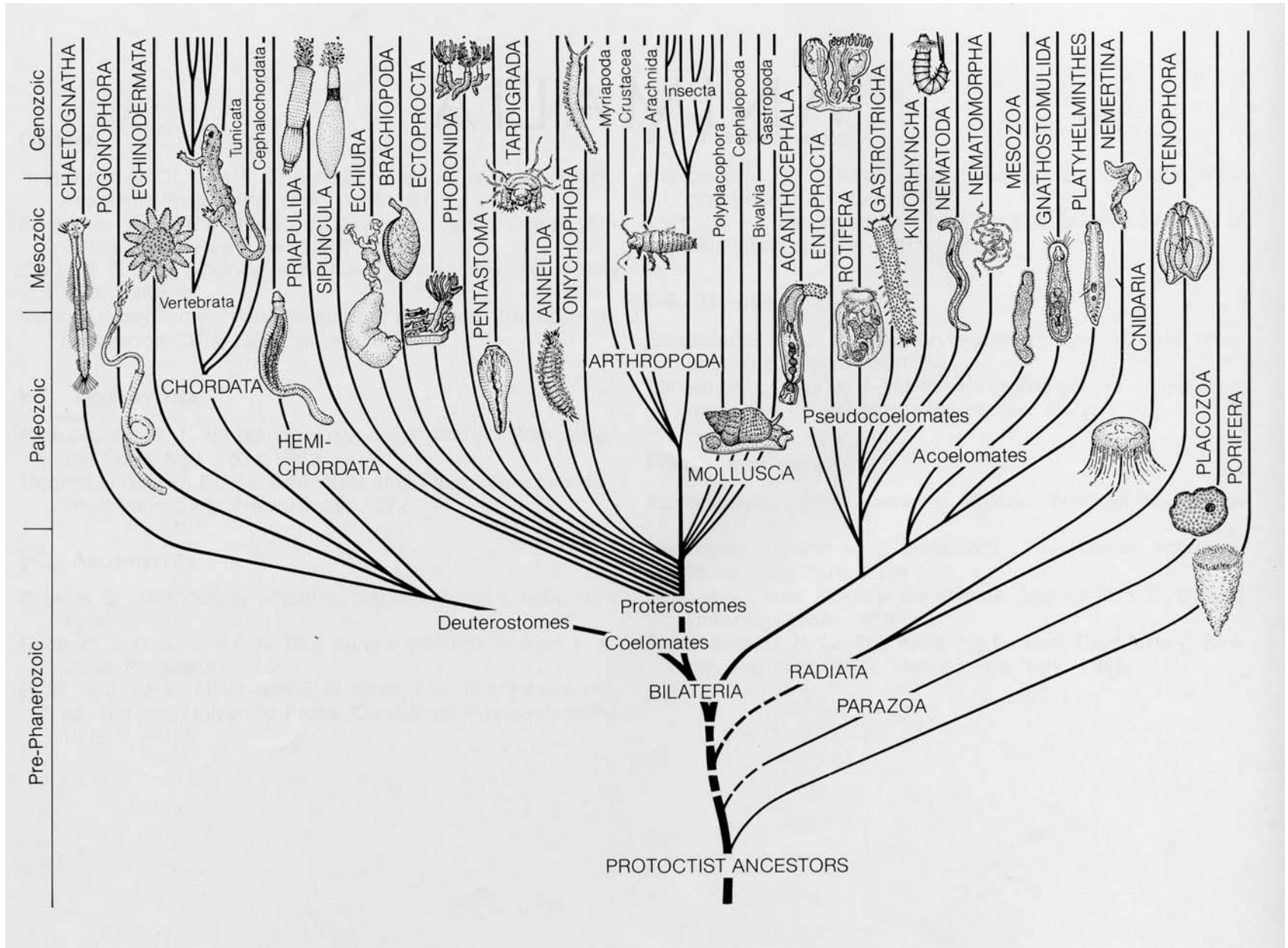
Von

Prof. Dr. GERHARD BRAUNITZER

Max-Planck-Institut für Biochemie, München

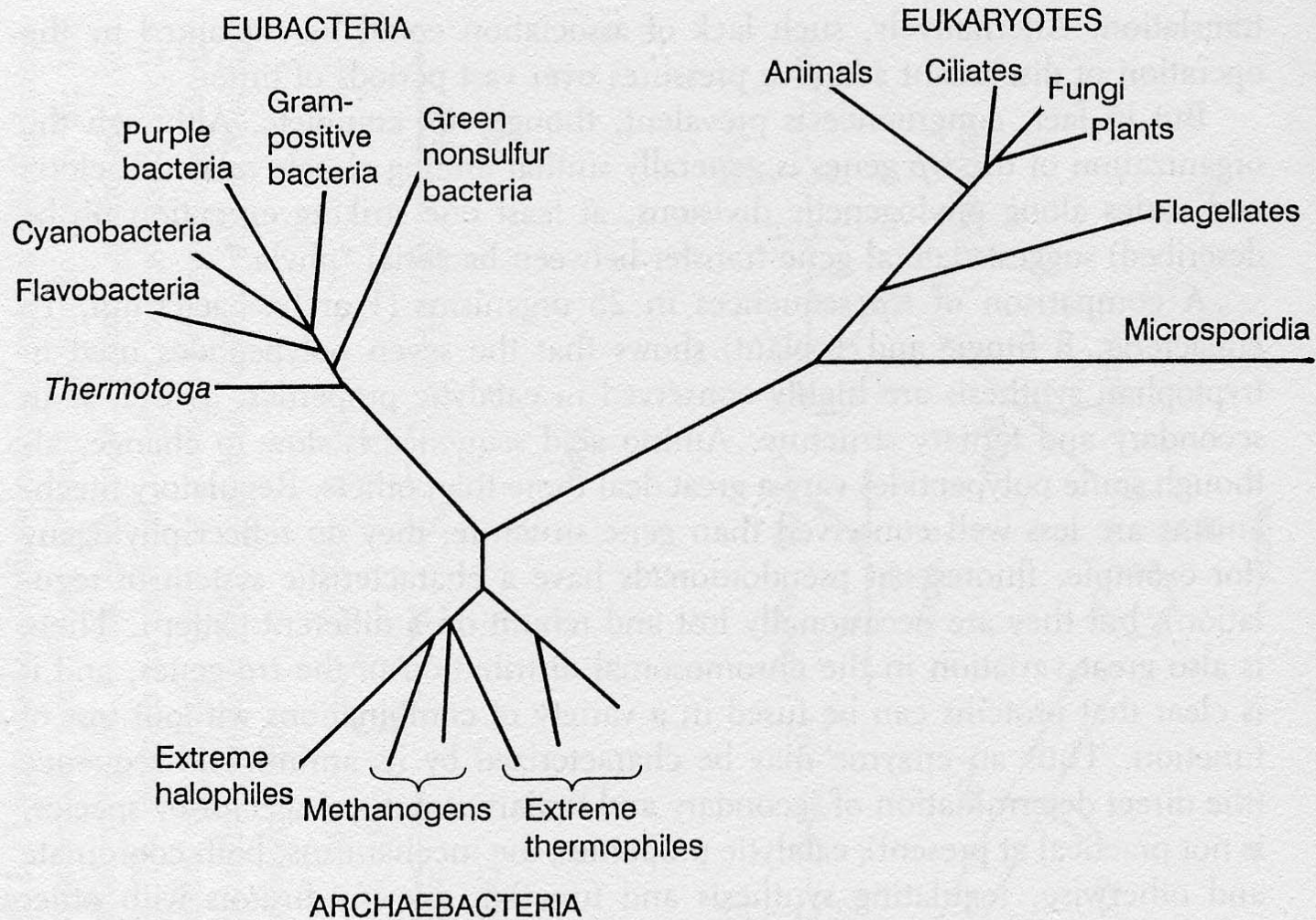


Molecular evolution through comparison  
of sequences from different organisms



## Five kingdoms.

L. Margulis, K.V. Schwartz, W.H. Freeman & Co., 1982



**FIGURE 2.** The two bacterial phylogenies, taken from the universal phylogenetic tree determined from rRNA sequence comparisons (Woese, 1987).

Evolution at the molecular level.

R.K. Selander, A.G. Clark, T.S. Whittam, eds. Sinauer Associates, 1991.



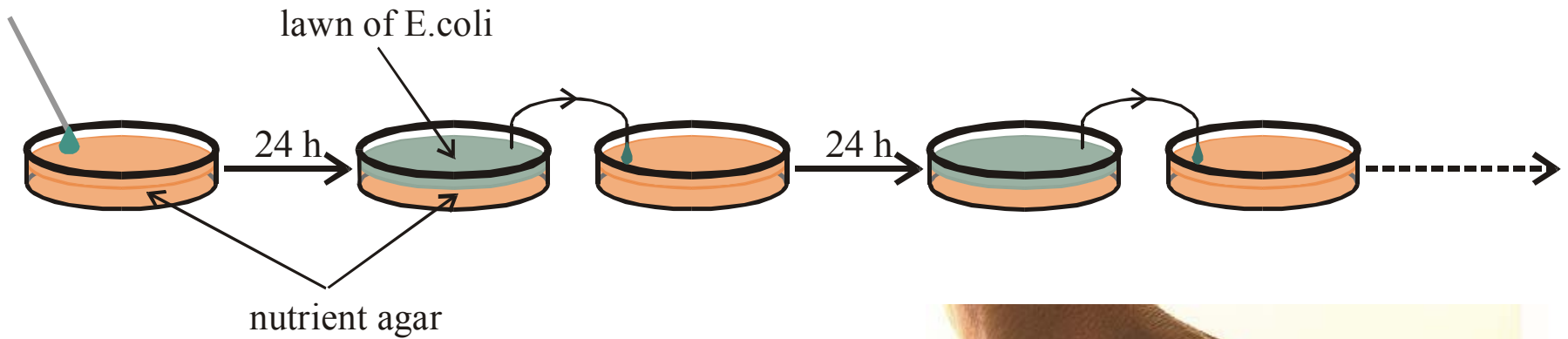
	Generation time	10 000 generations	10 <sup>6</sup> generations	10 <sup>7</sup> generations
RNA molecules	10 sec	27.8 h = 1.16 d	115.7 d	3.17 a
	1 min	6.94 d	1.90 a	19.01 a
Bacteria	20 min	138.9 d	38.03 a	380 a
	10 h	11.40 a	1 140 a	11 408 a
Higher multicellular organisms	10 d	274 a	27 380 a	273 800 a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

Time scales of evolutionary change

## **Bacterial Evolution**

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants*. Science **272** (1996), 1802-1804

D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, M. Blot. *Genomic evolution during a 10,000-generation experiment with bacteria*. Proc.Natl.Acad.Sci.USA **96** (1999), 3807-3812



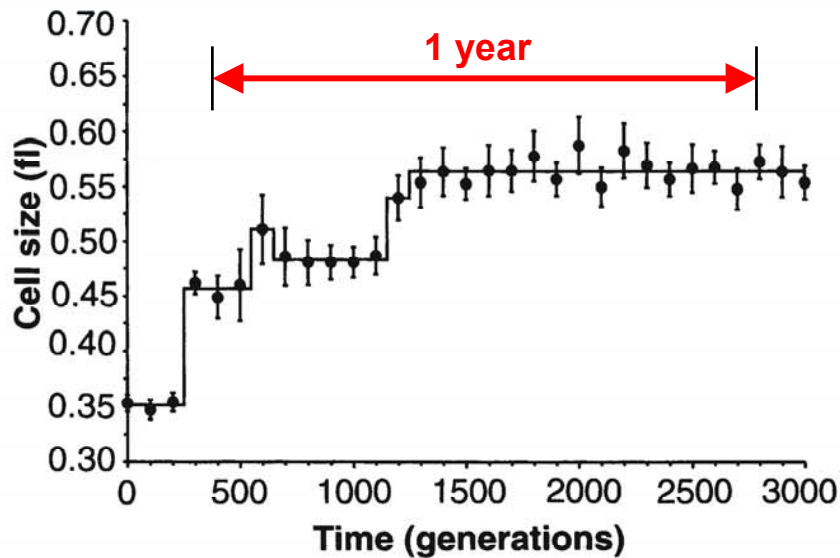
## Serial transfer of *Escherichia coli* cultures in Petri dishes

1 day <sup>a</sup> 6.67 generations

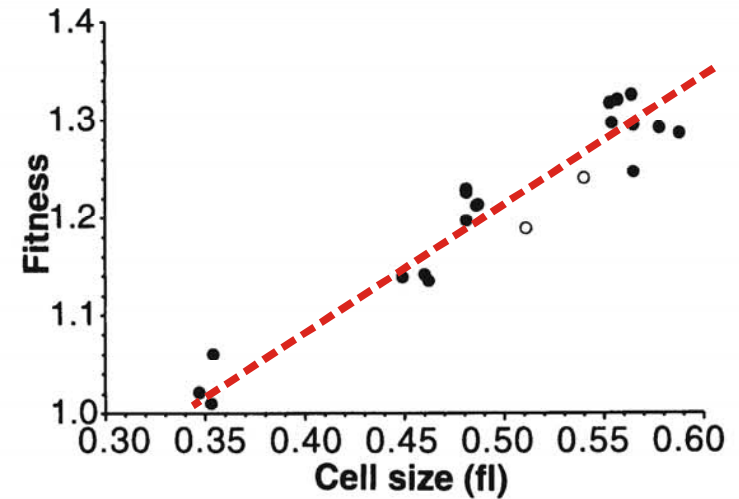
1 month <sup>a</sup> 200 generations

1 year <sup>a</sup> 2400 generations





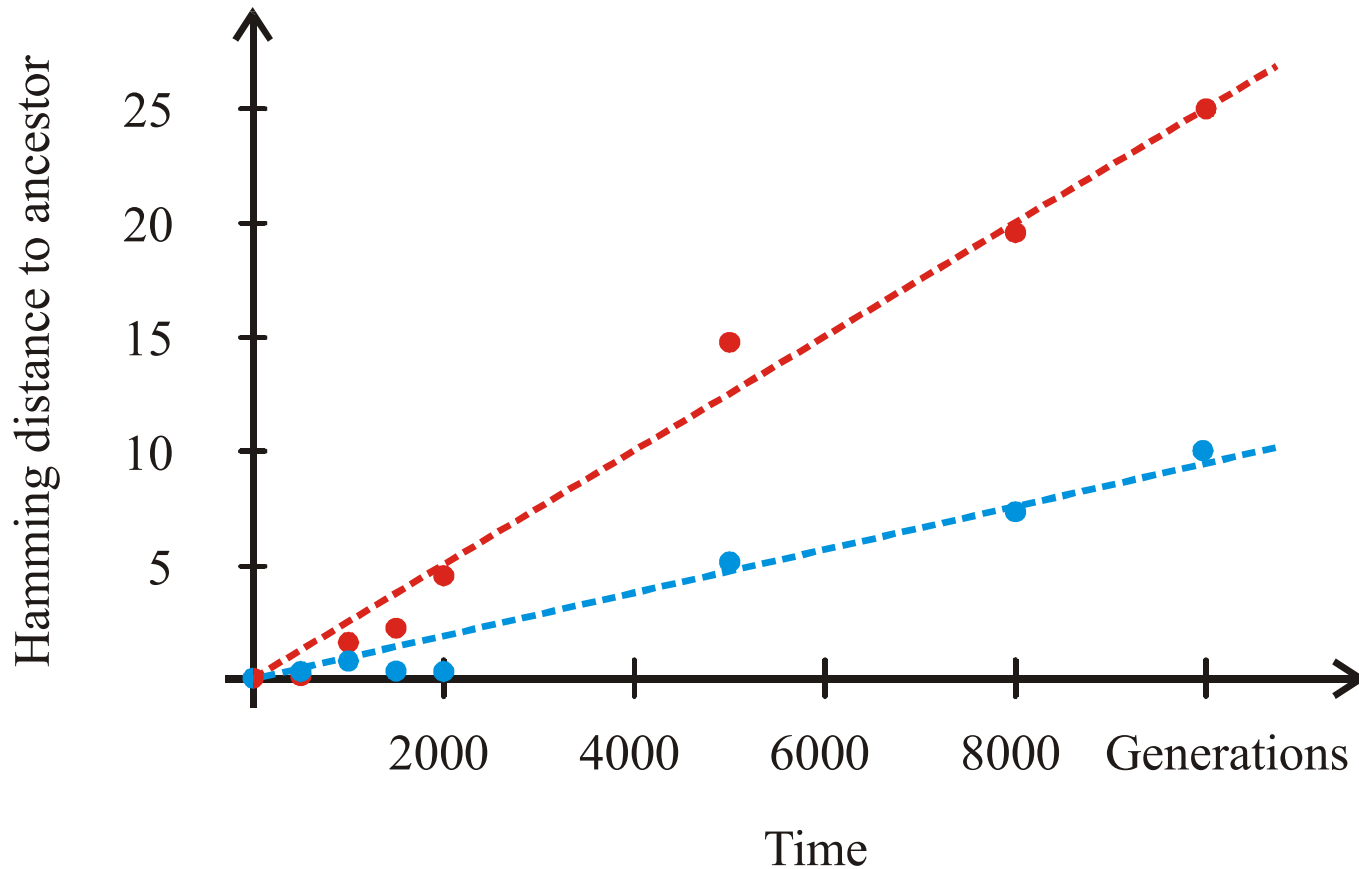
**Fig. 1.** Change in average cell size (1 fl =  $10^{-15}$  L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (22). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these data (Table 1).



**Fig. 2.** Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (6, 7). The open symbols indicate the only two samples assigned to different steps by the cell size and fitness data.

## Epochal evolution of bacteria in serial transfer experiments under constant conditions

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants.* *Science* **272** (1996), 1802-1804



## Variation of genotypes in a bacterial serial transfer experiment

D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, M. Blot. *Genomic evolution during a 10,000-generation experiment with bacteria*. Proc.Natl.Acad.Sci.USA **96** (1999), 3807-3812

## Evolution of RNA molecules based on Q $\beta$ phage

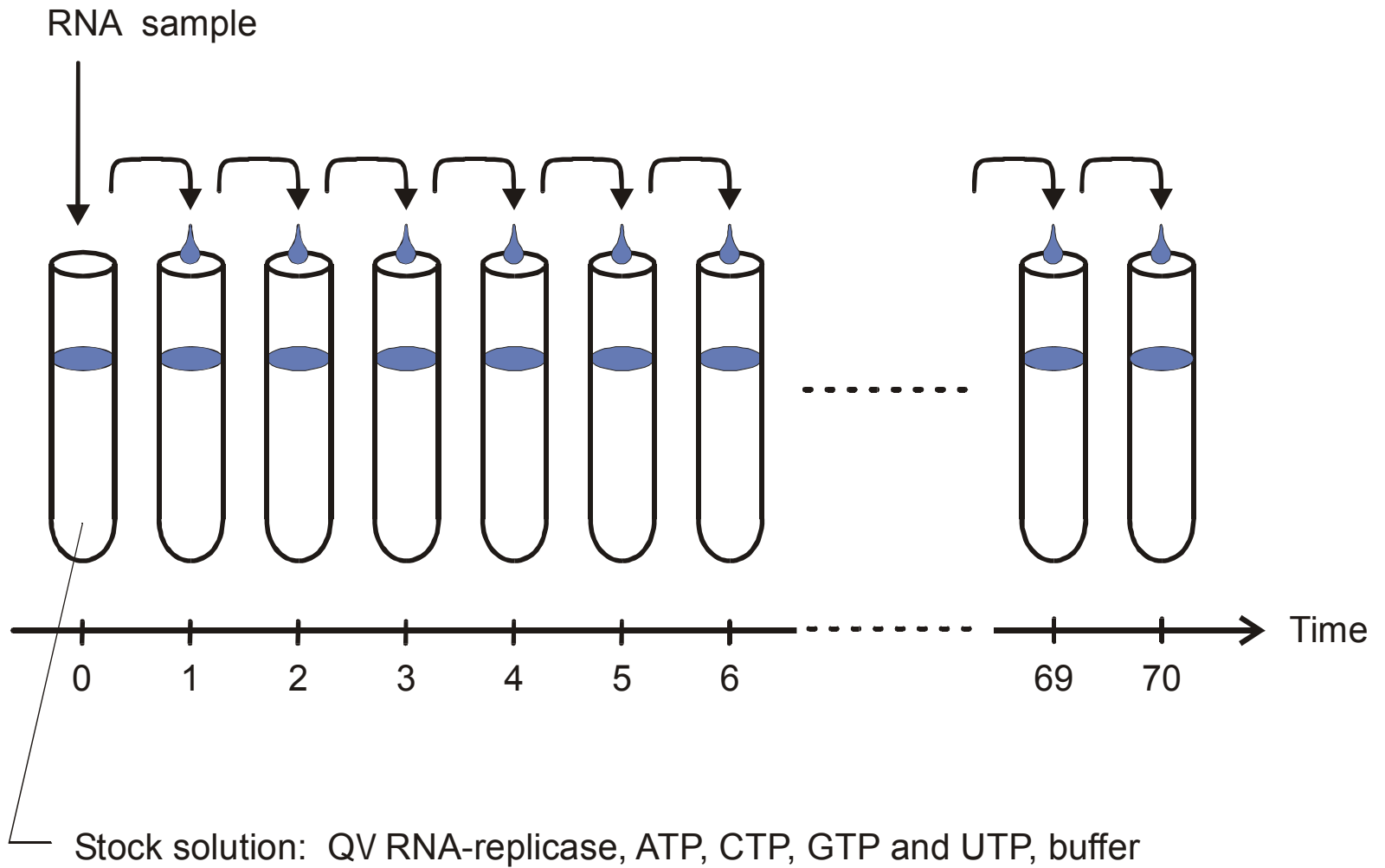
D.R.Mills, R.L.Peterson, S.Spiegelman, *An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule*. Proc.Natl.Acad.Sci.USA **58** (1967), 217-224

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

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

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

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



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

Reproduction of the original figure of the serial transfer experiment with Q $\beta$  RNA

D.R.Mills, R.L.Peterson, S.Spiegelman,  
*An extracellular Darwinian experiment  
 with a self-duplicating nucleic acid  
 molecule.* Proc.Natl.Acad.Sci.USA  
**58** (1967), 217-224

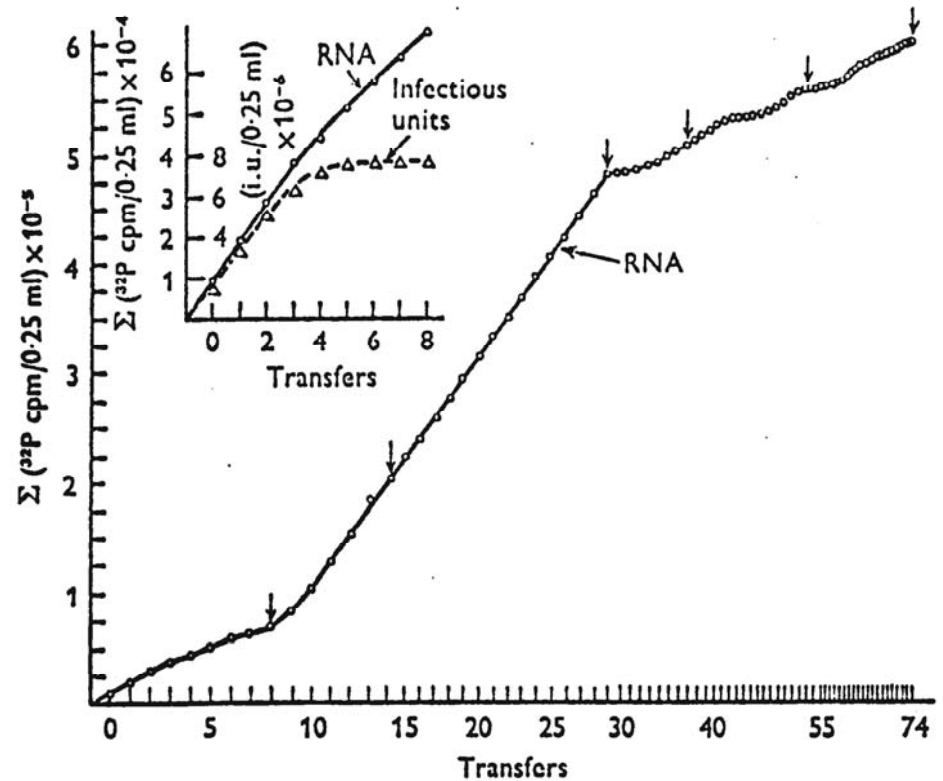
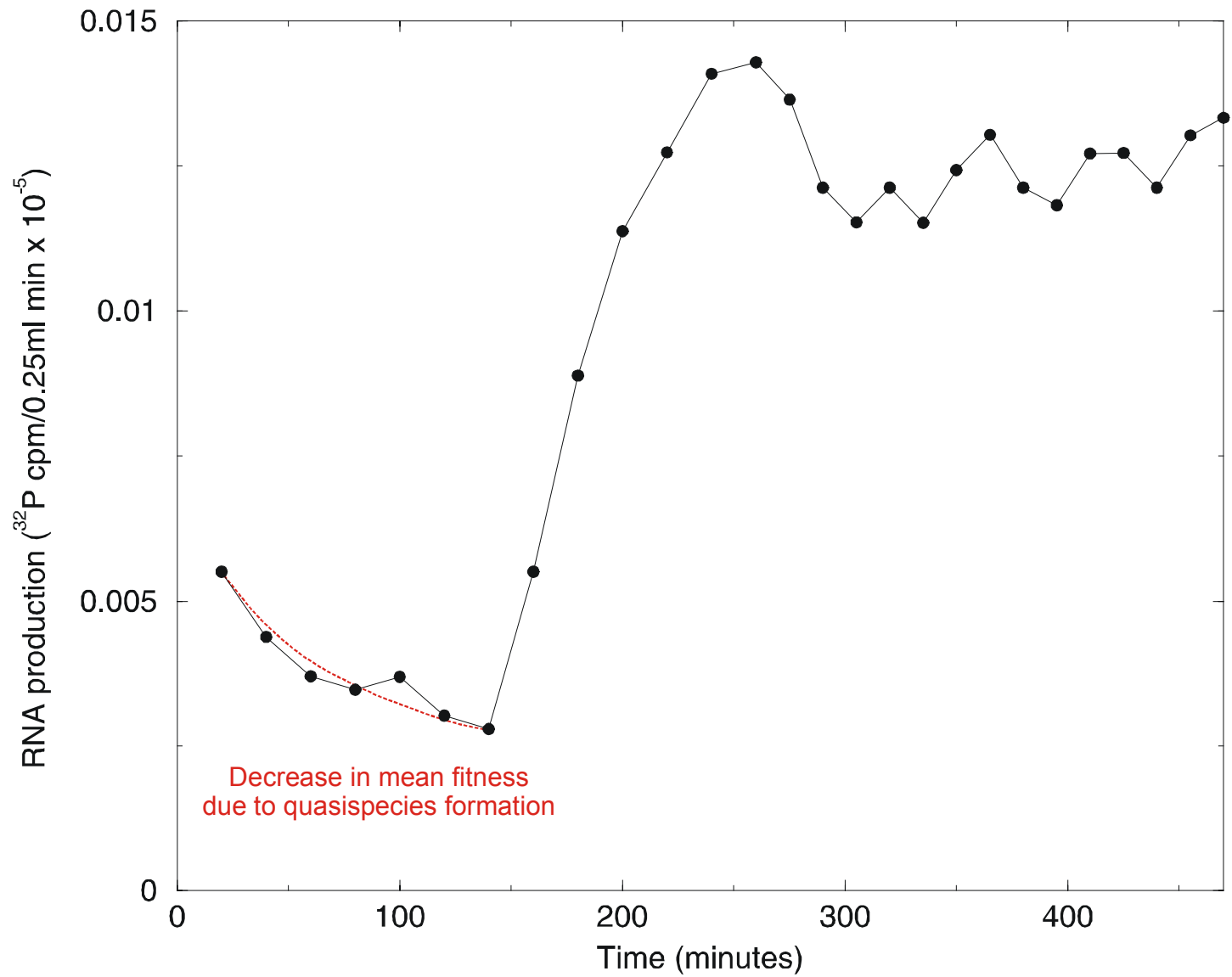


Fig. 9. Serial transfer experiment. Each 0.25 ml standard reaction mixture contained 40  $\mu\text{g}$  of Q $\beta$  replicase and  $^{32}\text{P}$ -UTP. The first reaction (0 transfer) was initiated by the addition of 0.2  $\mu\text{g}$  ts-1 (temperature-sensitive RNA) and incubated at 35  $^{\circ}\text{C}$  for 20 min, whereupon 0.02 ml was drawn for counting and 0.02 ml was used to prime the second reaction (first transfer), and so on. After the first 13 reactions, the incubation periods were reduced to 15 min (transfers 14-29). Transfers 30-38 were incubated for 10 min. Transfers 39-52 were incubated for 7 min, and transfers 53-74 were incubated for 5 min. The arrows above certain transfers (0, 8, 14, 29, 37, 53, and 73) indicate where 0.001-0.1 ml of product was removed and used to prime reactions for sedimentation analysis on sucrose. The inset examines both infectious and total RNA. The results show that biologically competent RNA ceases to appear after the 4th transfer (Mills *et al.* 1967).





The increase in RNA production rate during a serial transfer experiment

## Evolutionary design of RNA molecules

D.B.Bartel, J.W.Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature **346** (1990), 818-822

C.Tuerk, L.Gold, *SELEX - Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science **249** (1990), 505-510

D.P.Bartel, J.W.Szostak, *Isolation of new ribozymes from a large pool of random sequences.* Science **261** (1993), 1411-1418

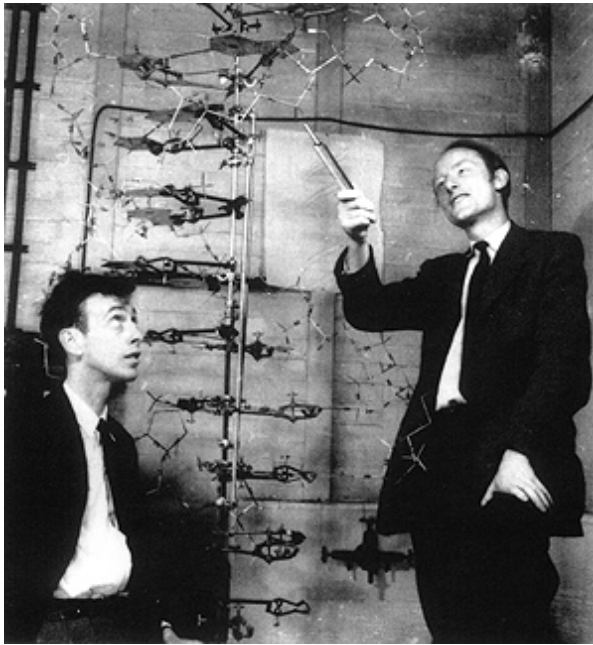
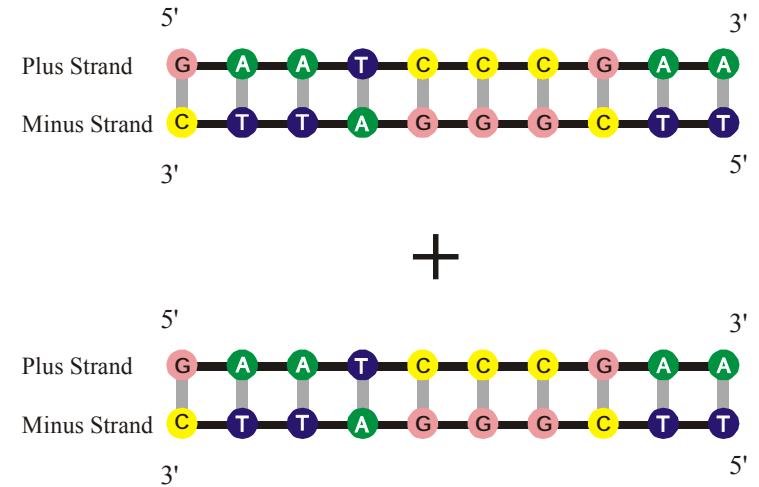
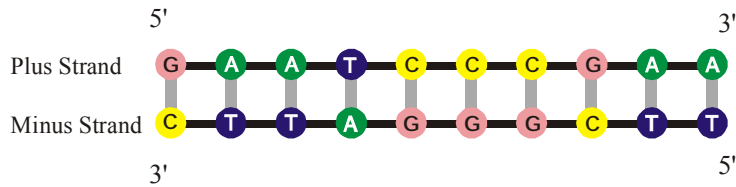
R.D.Jenison, S.C.Gill, A.Pardi, B.Poliski, *High-resolution molecular discrimination by RNA.* Science **263** (1994), 1425-1429

Y. Wang, R.R.Rando, *Specific binding of aminoglycoside antibiotics to RNA.* Chemistry & Biology **2** (1995), 281-290

Jiang, A. K. Suri, R. Fiala, D. J. Patel, *Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.* Chemistry & Biology **4** (1997), 35-50

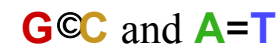
*No new principle will declare itself  
from below a heap of facts.*

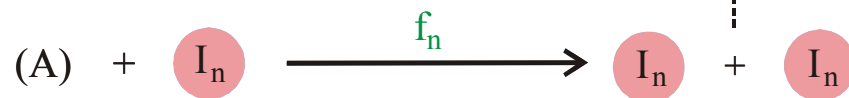
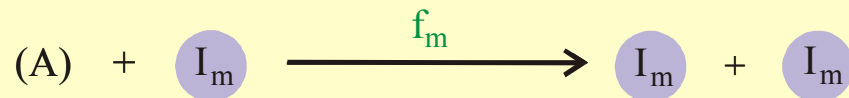
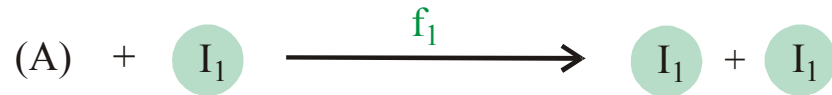
Sir Peter Medawar, 1985



James Watson and Francis Crick, 1953

**Replication of DNA** is a highly complex copying mechanism involving more than ten different protein molecules. Complementarity is determined by Watson-Crick base pairs:





$$\frac{dx_i}{dt} = f_i x_i - x_i \Phi = x_i (f_i - \Phi)$$

$$\Phi = \sum_j f_j x_j ; \quad \sum_j x_j = 1 ; \quad i, j = 1, 2, \dots, n$$

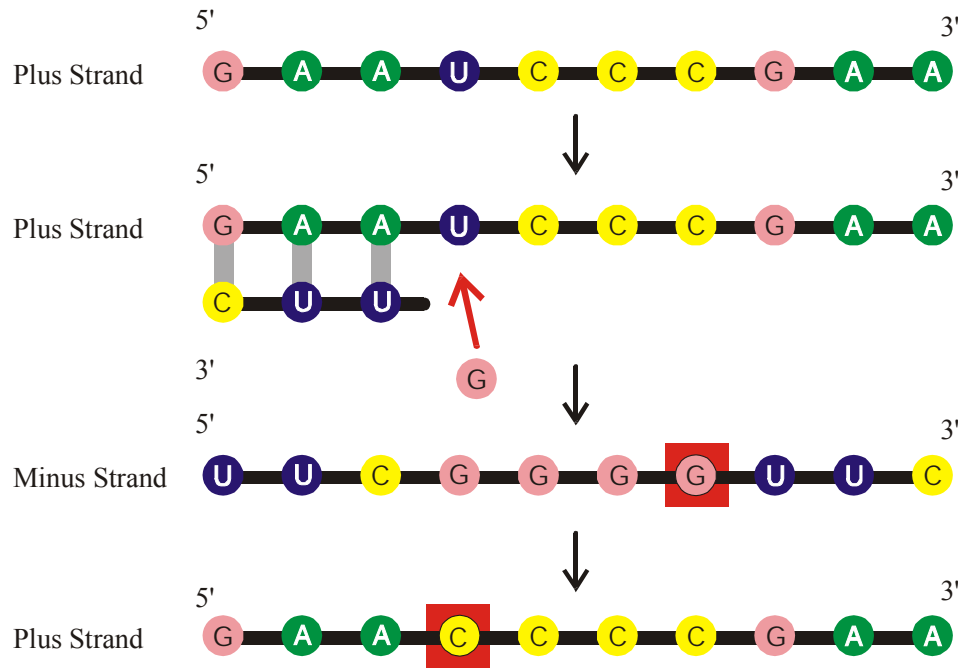
$$[I_i] = x_i \geq 0 ; \quad i = 1, 2, \dots, n ;$$

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

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

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

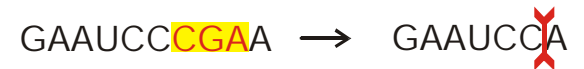
**Reproduction of organisms or replication of molecules as the basis of selection**



**Point Mutation**



**Insertion**



**Deletion**

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

# Theory of molecular evolution

M.Eigen, *Self-organization of matter and the evolution of biological macromolecules*.

Naturwissenschaften **58** (1971), 465-526

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

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

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

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

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

J. Swetina, P. Schuster, *Self-replication with errors - A model for polynucleotide replication*.

Biophys.Chem. **16** (1982), 329-345

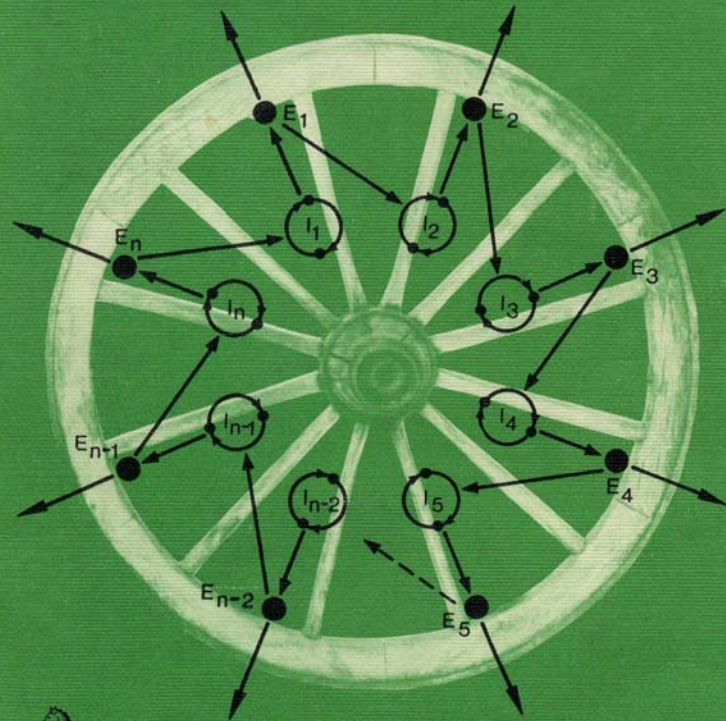
J.S. McCaskill, *A localization threshold for macromolecular quasispecies from continuously distributed replication rates*. J.Chem.Phys. **80** (1984), 5194-5202

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

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

M. Eigen P. Schuster  
**The Hypercycle**

A Principle of Natural Self-Organization



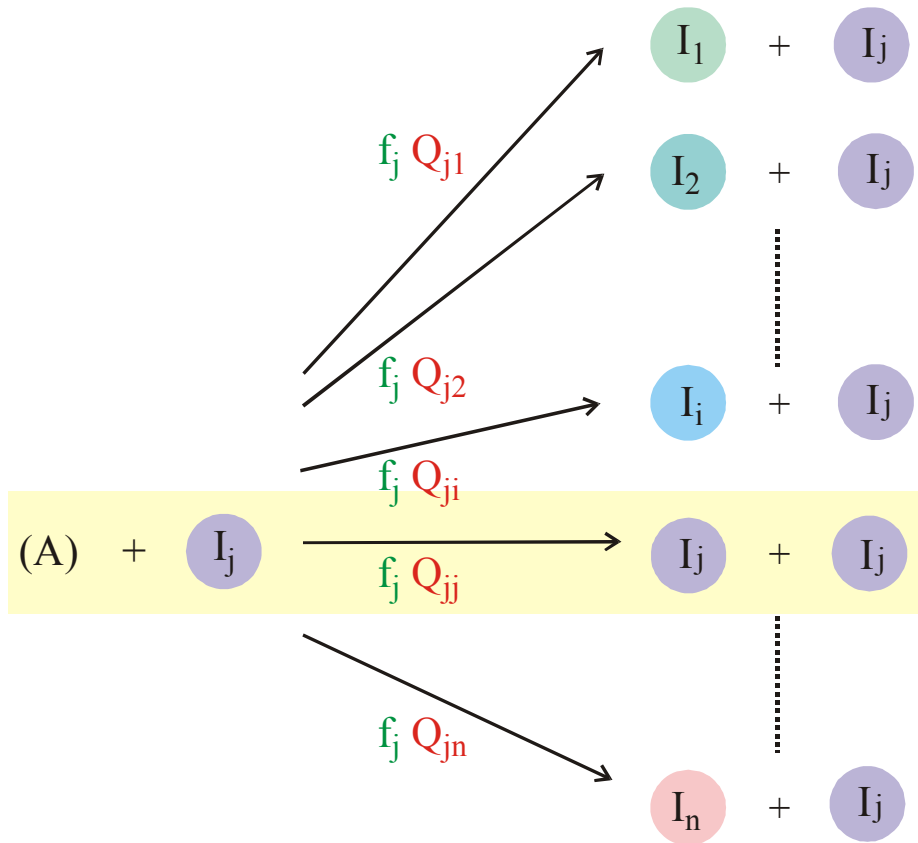
**Chemical kinetics of molecular evolution**

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



Springer-Verlag Berlin Heidelberg New York





$$\frac{dx_i}{dt} = \sum_j f_j Q_{ji} x_j - x_i \Phi$$

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

$$[I_i] = x_i \ll 1 ; \quad i = 1, 2, \dots, n ;$$

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

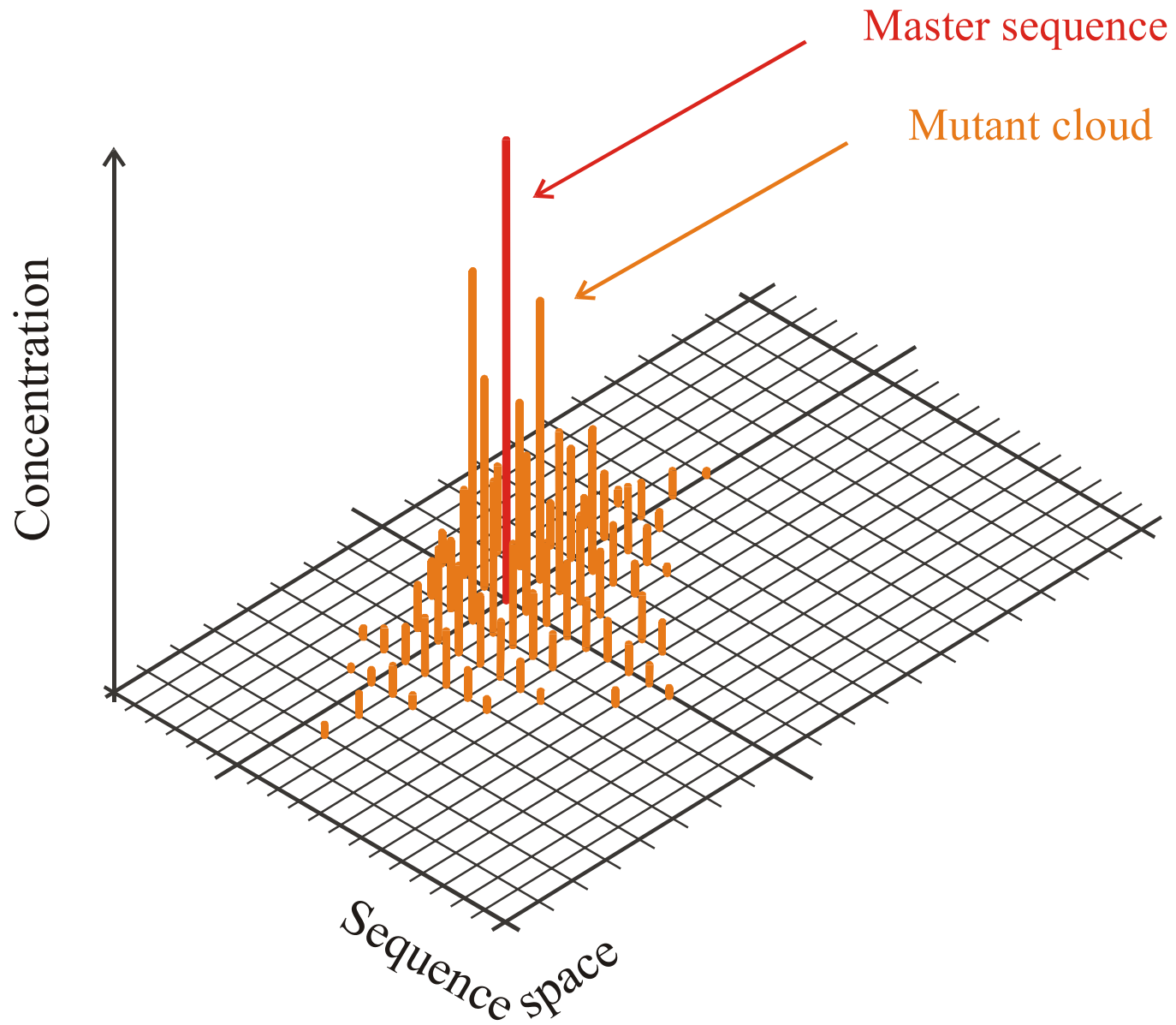
$$Q_{ij} = (1-p)^{\ell-d(i,j)} p^{d(i,j)}$$

$p$  ..... Error rate per digit

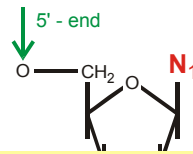
$\ell$  ..... Chain length of the polynucleotide

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

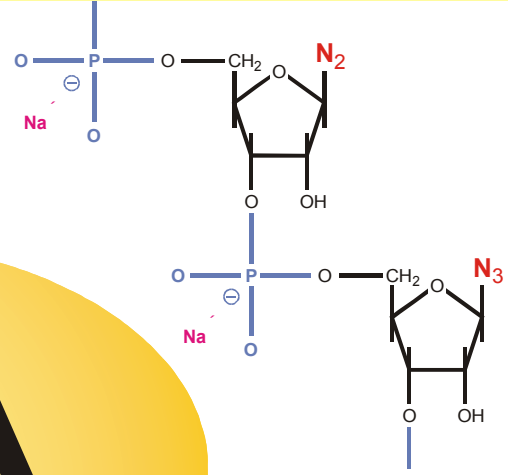
Chemical kinetics of replication and mutation as parallel reactions



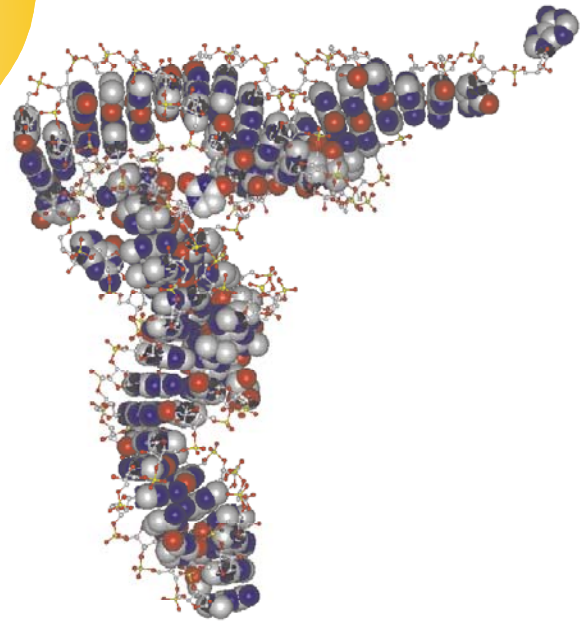
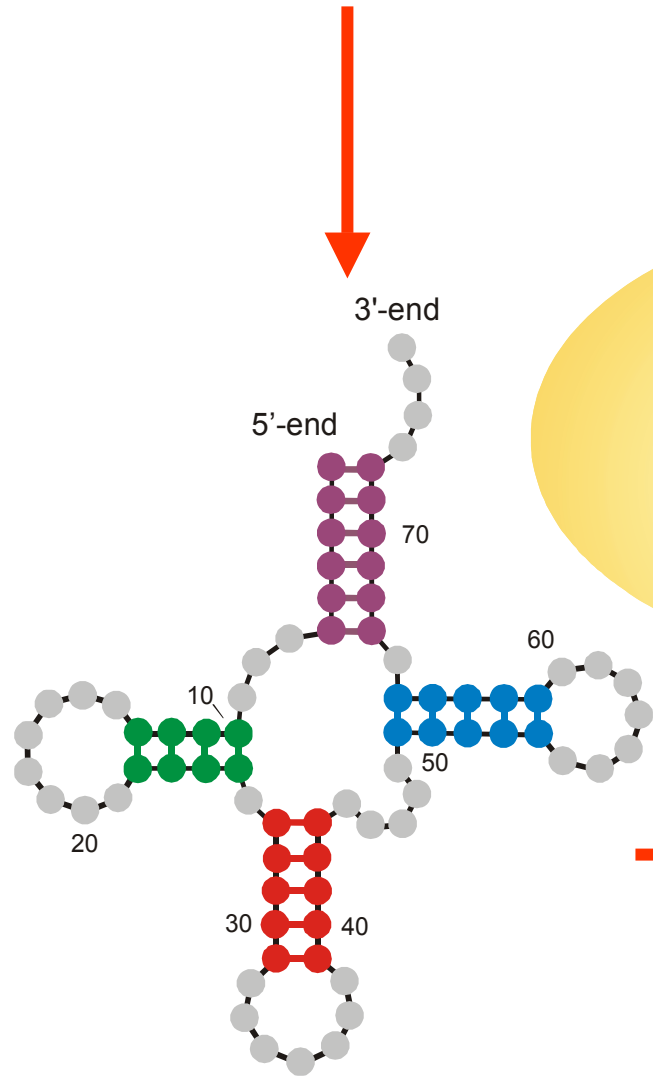
The molecular quasispecies in sequence space



5'-end **GCGGAUUUAGCUC**AGUUGGGAGAG**CGCCAGACUGAAGAUCUGG**AGGUC**CUGUGUUCGAUCCACAGAAUUCGCACCA** 3'-end



**RNA**



Definition of RNA structure

## **Optimization of RNA molecules *in silico***

W.Fontana, P.Schuster, *A computer model of evolutionary optimization*. Biophysical Chemistry **26** (1987), 123-147

W.Fontana, W.Schnabl, P.Schuster, *Physical aspects of evolutionary optimization and adaptation*. Phys.Rev.A **40** (1989), 3301-3321

M.A.Huynen, W.Fontana, P.F.Stadler, *Smoothness within ruggedness. The role of neutrality in adaptation*. Proc.Natl.Acad.Sci.USA **93** (1996), 397-401

W.Fontana, P.Schuster, *Continuity in evolution. On the nature of transitions*. Science **280** (1998), 1451-1455

W.Fontana, P.Schuster, *Shaping space. The possible and the attainable in RNA genotype-phenotype mapping*. J.Theor.Biol. **194** (1998), 491-515

B.M.R. Stadler, P.F. Stadler, G.P. Wagner, W. Fontana, *The topology of the possible: Formal spaces underlying patterns of evolutionary change*. J.Theor.Biol. **213** (2001), 241-274

random individuals. The primer pair used for genomic DNA amplification is 5'-TCTCCCTGGATTCT-CATTAA-3' (forward) and 5'-TCTTTGTCTTCTGT-TGCACC-3' (reverse). Reactions were performed in 25  $\mu$ l using 1 unit of Taq DNA polymerase with each primer at 0.4  $\mu$ M, 200  $\mu$ M each dATP, dTTP, dCTP, and dGTP; and PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>] in a cycle condition of 94°C for 1 min and then 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by 72°C for 6 min. PCR products were purified (Qiagen), digested with Xmn I, and separated in a 2% agarose gel.

32. A nonsense mutation may affect mRNA stability and result in degradation of the transcript [L. Maquat, *Am. J. Hum. Genet.* **59**, 279 (1996)].

33. Data not shown; a dot blot with poly (A)<sup>+</sup> RNA from 50 human tissues (The Human RNA Master Blot, 7770-1, Clontech Laboratories) was hybridized with a probe from exons 29 to 47 of *MYO15* using the same condition as Northern blot analysis (13).

34. Smith-Magenis syndrome (SMS) is due to deletions of 17p11.2 of various sizes, the smallest of which includes *MYO15* and perhaps 20 other genes [6]; K-S Chen, L. Potocki, J. R. Lupski, *MROD Res. Rev.* **2**, 122 (1996)]. *MYO15* expression is easily detected in the pituitary gland (data not shown). Haploinsufficiency for *MYO15* may explain a portion of the SMS

phenotype such as short stature. Moreover, a few SMS patients have sensorineural hearing loss, possibly because of a point mutation in *MYO15* in trans to the SMS 17p11.2 deletion.

35. R. A. Fiedel, data not shown.

36. K. B. Avraham *et al.*, *Nature Genet.* **11**, 369 (1995); X-Z. Liu *et al.*, *ibid.* **17**, 268 (1997); F. Gibson *et al.*, *Nature* **374**, 62 (1995); D. Weil *et al.*, *ibid.*, p. 60.

37. RNA was extracted from cochlea (membranous labyrinth) obtained from human fetuses at 18 to 22 weeks of development in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Only samples without evidence of degradation were pooled for poly (A)<sup>+</sup> selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-for-PCR kit (Clontech Laboratories). A portion of the first-strand cDNA (4%) was amplified by PCR with Advantage cDNA polymerase mix (Clontech Laboratories) using human *MYO15*-specific oligonucleotide primers (forward, 5'-GCATGACCTGCGGGTAAT-GCG-3'; reverse, 5'-CTCAGCGGCTTCTGGCATGGT-GCTCGCTGCG-3'). Cycling conditions were 40 s at 94°C, 40 s at 66°C (3 cycles), 60°C (5 cycles), and 55°C (29 cycles); and 45 s at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 688-bp PCR

product is expected from amplification of the human *MYO15* cDNA. Amplification of human genomic DNA with this primer pair would result in a 2903-bp fragment.

38. We are grateful to the people of Bengkala, Bali, and the two families from India. We thank J. R. Lupski and K.-S. Chen for providing the human chromosome 17 cosmid library. For technical and computational assistance, we thank N. Dietrich, M. Ferguson-S. A. Gupta, E. Sorbello, R. Torzkadsh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Stenberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and J. Barber, S. Sullivan, E. Green, D. Drayna, and T. Battey for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (Z01 DC 00035-01 and Z01 DC 00038-01 to T.B.F. and E.R.W. and R01 DC 03402 to C.G.M.), the National Institute of Child Health and Human Development (R01 HD00428 to S.A.C.) and a National Science Foundation Graduate Research Fellowship to F.J.P. This paper is dedicated to J. B. Snow Jr. on his retirement as the Director of the NIDCD.

9 March 1998; accepted 17 April 1998

## Continuity in Evolution: On the Nature of Transitions

Walter Fontana and Peter Schuster

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

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

We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (2). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level of analysis compared with the three-dimensional structure at atomic resolution. Yet, secondary structures are empir-

ically well defined and obtain their biophysical and biochemical importance from being a scaffold for the tertiary structure. For the sake of brevity, we shall refer to secondary structures as "shapes." RNA combines in a single molecule both genotype (replicable sequence) and phenotype (selectable shape), making it ideally suited for in vitro evolution experiments (3, 4).

To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (5, 6). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (4). Although in the experiment the evolutionary end product was unknown, we thought of its shape as being specified implicitly by the imposed selection criterion. Because our intent is to study evolutionary histories rather than end products, we defined a target shape in advance and assumed the replication rate of a sequence to be a function of

the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

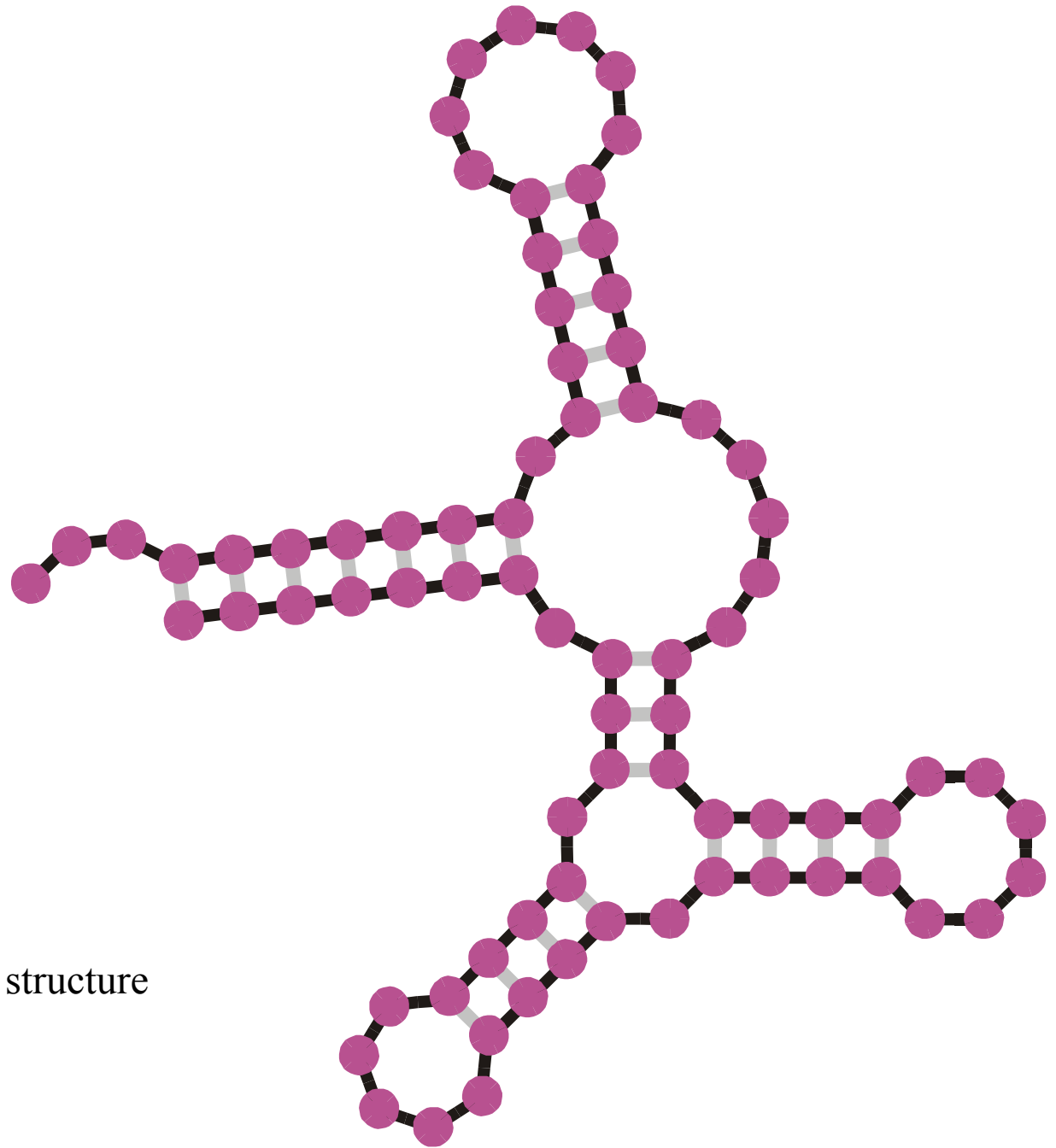
An instance representing in its qualitative features all the simulations we performed is shown in Fig. 1A. Starting with identical sequences folding into a random shape, the simulation was stopped when the population became dominated by the target, here a canonical tRNA shape. The black curve traces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

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

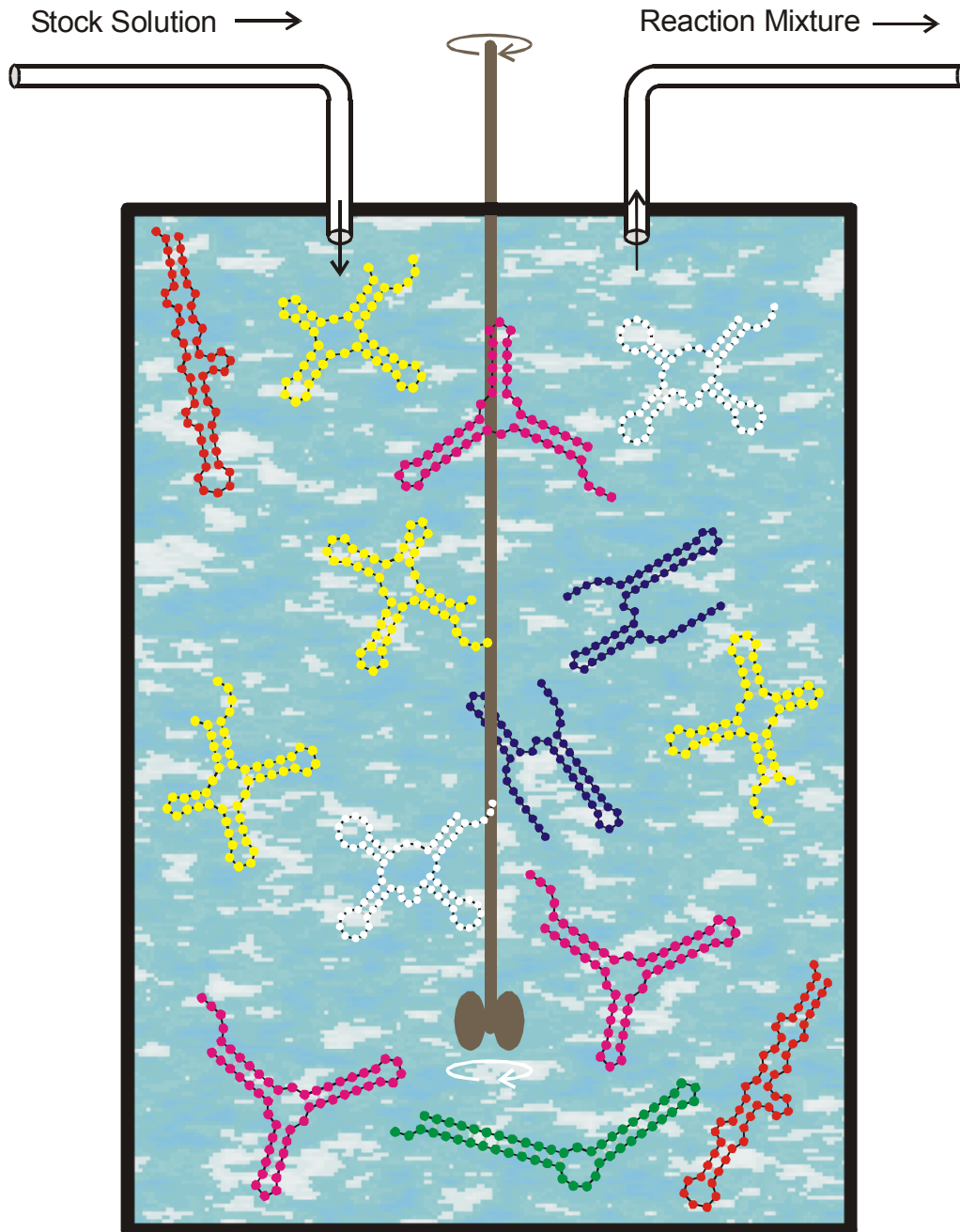
## Evolution *in silico*

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

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



Optimized element: RNA structure



Replication rate constant:

$$f_k = [ / [U + \delta d_S^{(k)}]$$

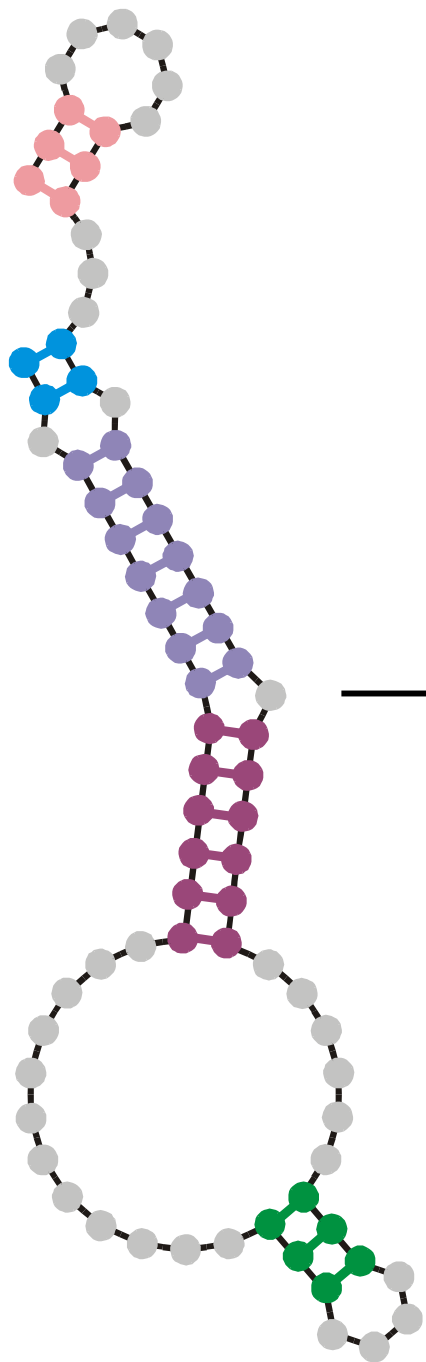
$$\delta d_S^{(k)} = d_H(S_k, S_H)$$

Selection constraint:

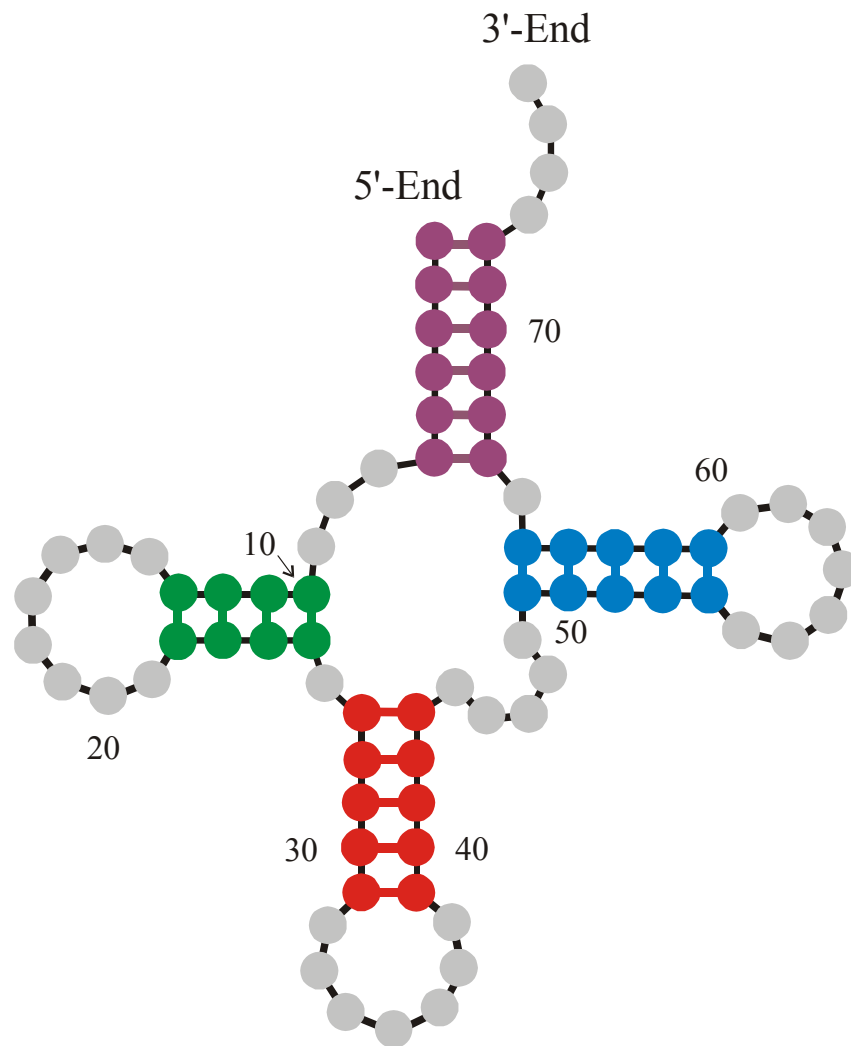
# RNA molecules is controlled by the flow

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

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

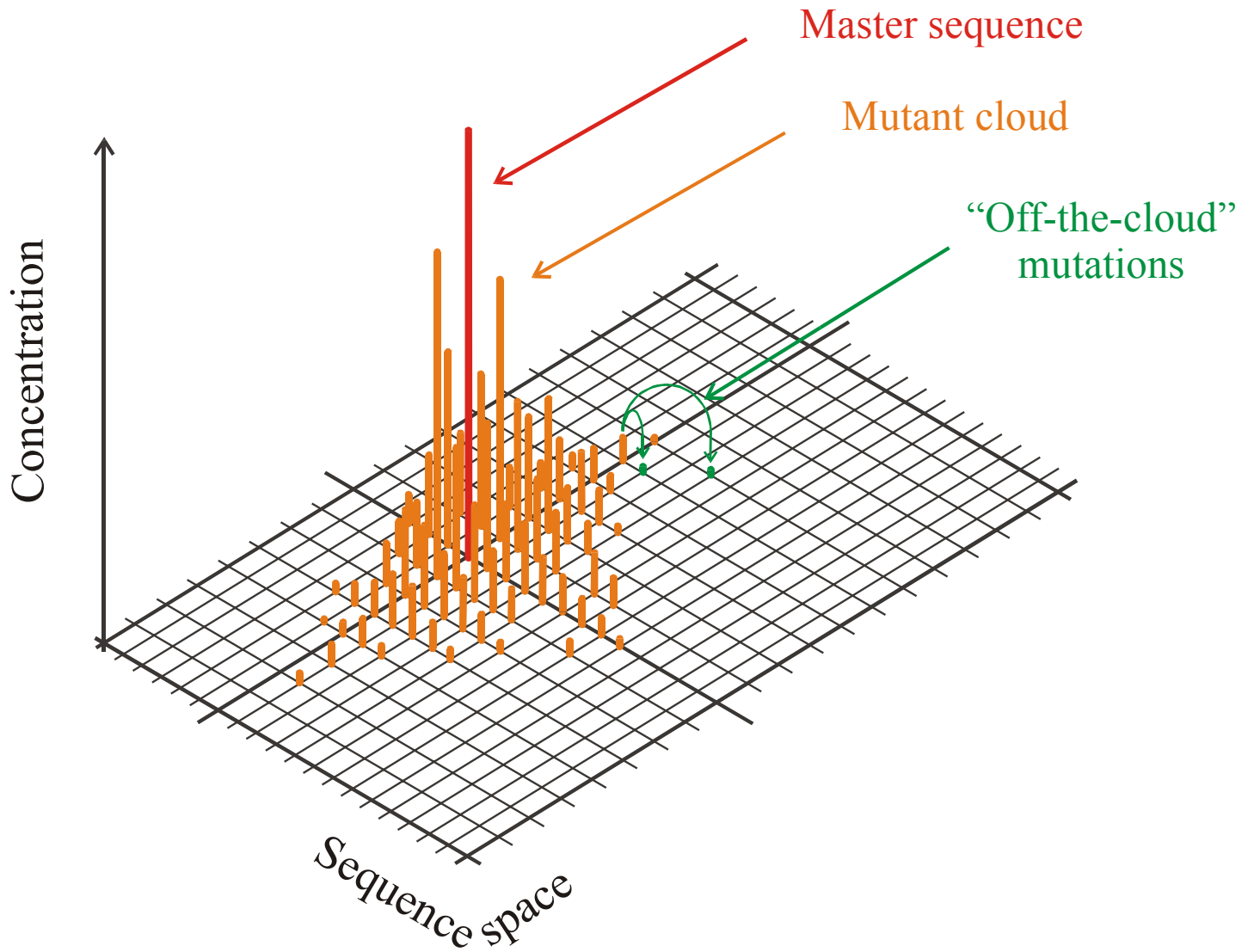


Randomly chosen  
initial structure



Phenylalanyl-tRNA as  
target structure



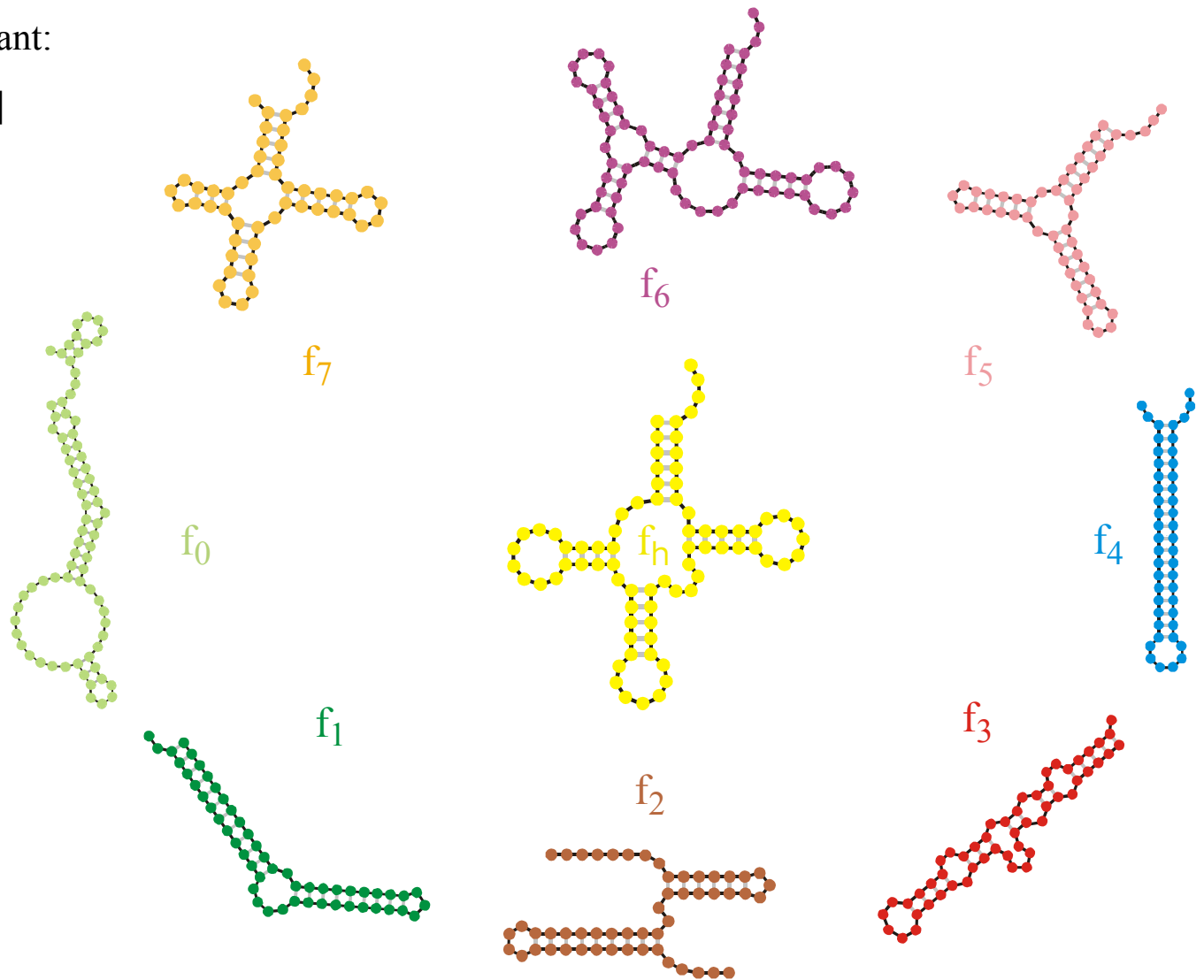


The molecular quasispecies  
in sequence space

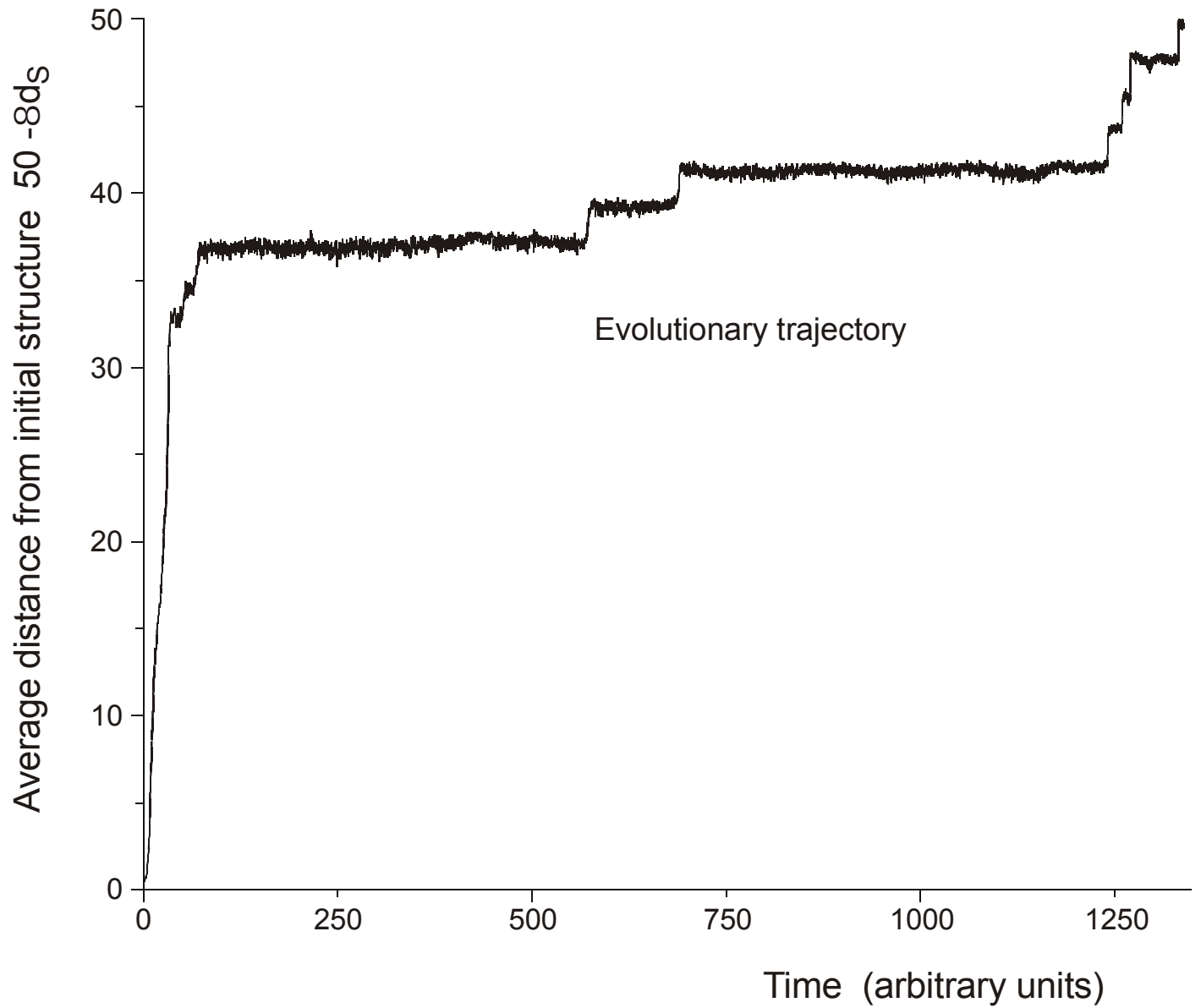
Replication rate constant:

$$f_k = \frac{[S_k]}{[U] + \sum d_S^{(k)}} \quad \text{or} \quad f_k = \frac{[S_k]}{[U] + \sum d_S^{(k)}}$$

$$d_S^{(k)} = d_H(S_k, S_h)$$

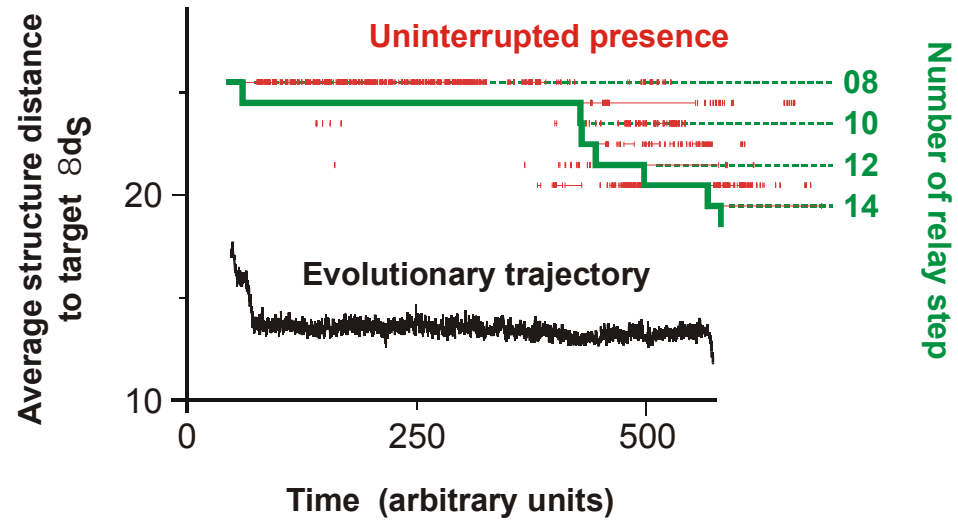


Evaluation of RNA secondary structures yields replication rate constants



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

**28 neutral point mutations** during a long quasi-stationary epoch

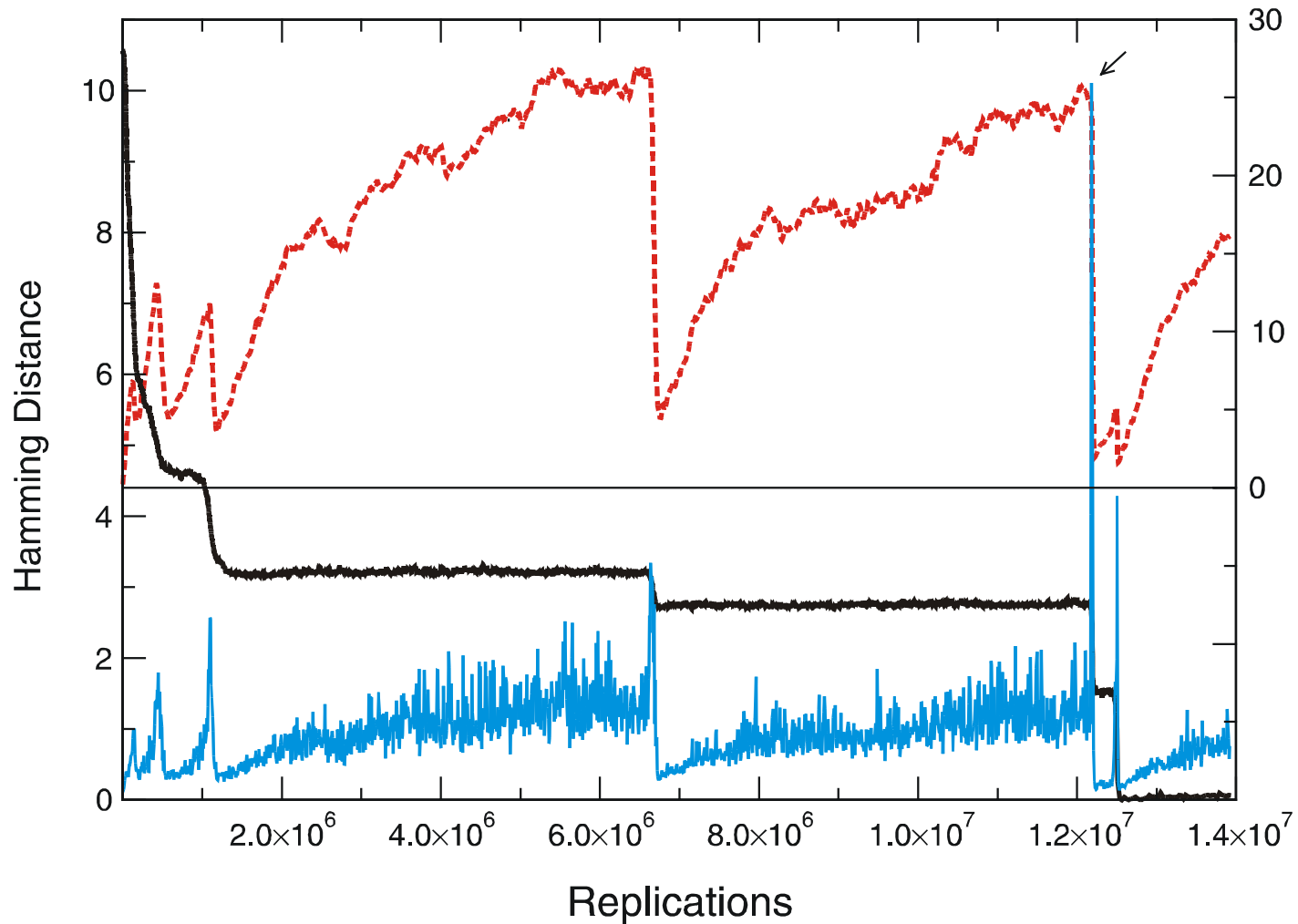


entry	GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGG	CAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8	.(((((((((((((. . . . . (((. . . . .)) . . . . .)))) . . . . .(((((. . . . .))))) . . . . .	
exit	GGUAUGGGCGUUGAAUA	AJAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAU
entry	GGUAUGGGCGUUGAAUA	AUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAU
9	.((((((.( . . . . .(((((. . . . .)) . . . . .)))) . . . . .(((((. . . . .)) . . . . .)) . . . . .	
exit	UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUA	CGCCCCACACACCGUCCCAAG
entry	UGGAUGGACGUUGAAUAACAAGGUAUCGACCAAACAACCAACGAGUAAGUGUGUA	CGCCCCACACACCGUCCCAAG
10	.(((((. . . . .(((((. . . . .)) . . . . .)))) . . . . .(((((. . . . .)) . . . . .)) . . . . .	
exit	UGGAUGGACGUUGAAUAACAAGGUAUCG	ACCAAACAACCAACGAGUAAGUGUGUA

**Transition inducing point mutations**

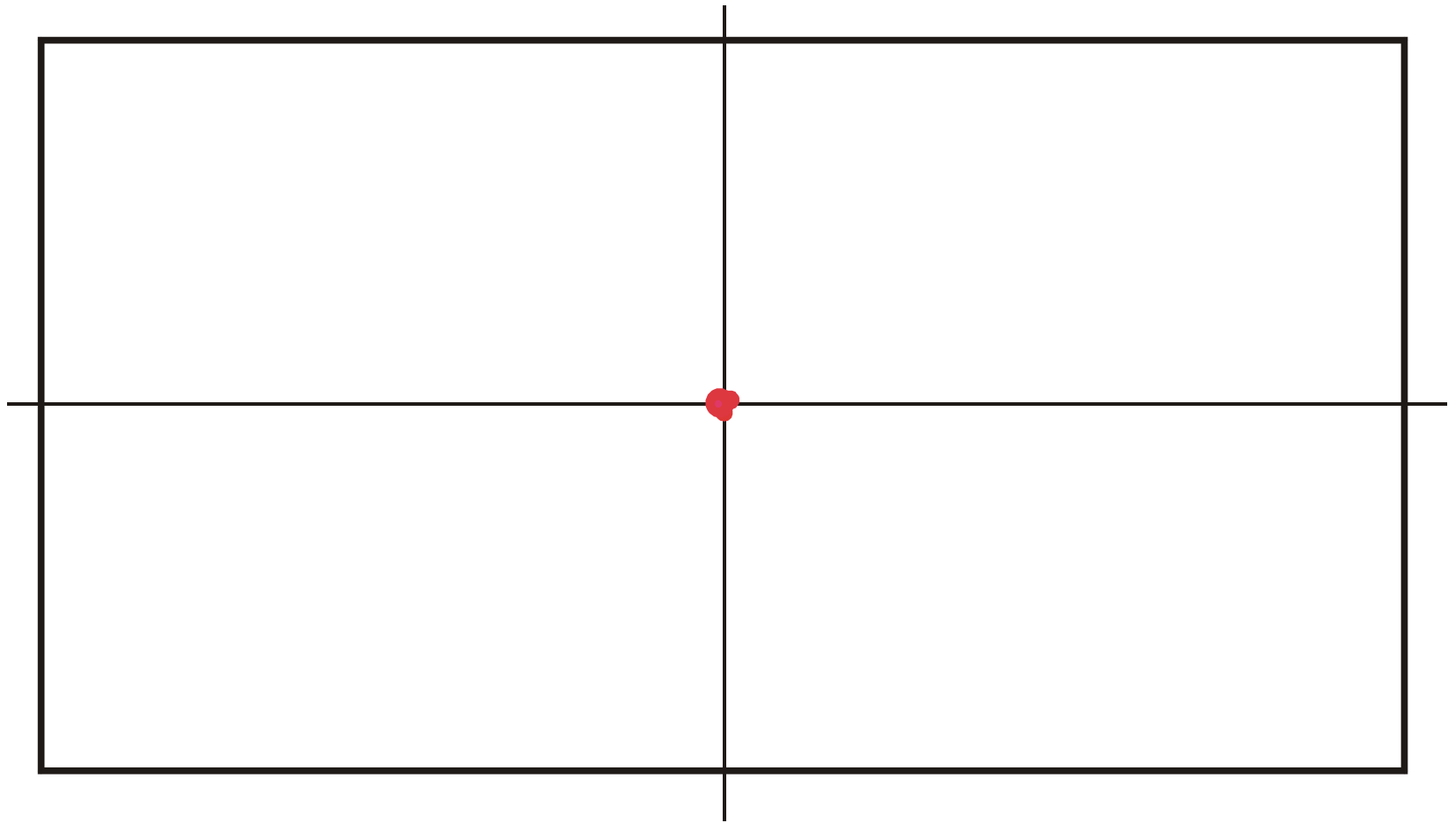
**Neutral point mutations**

**Neutral genotype evolution** during phenotypic stasis

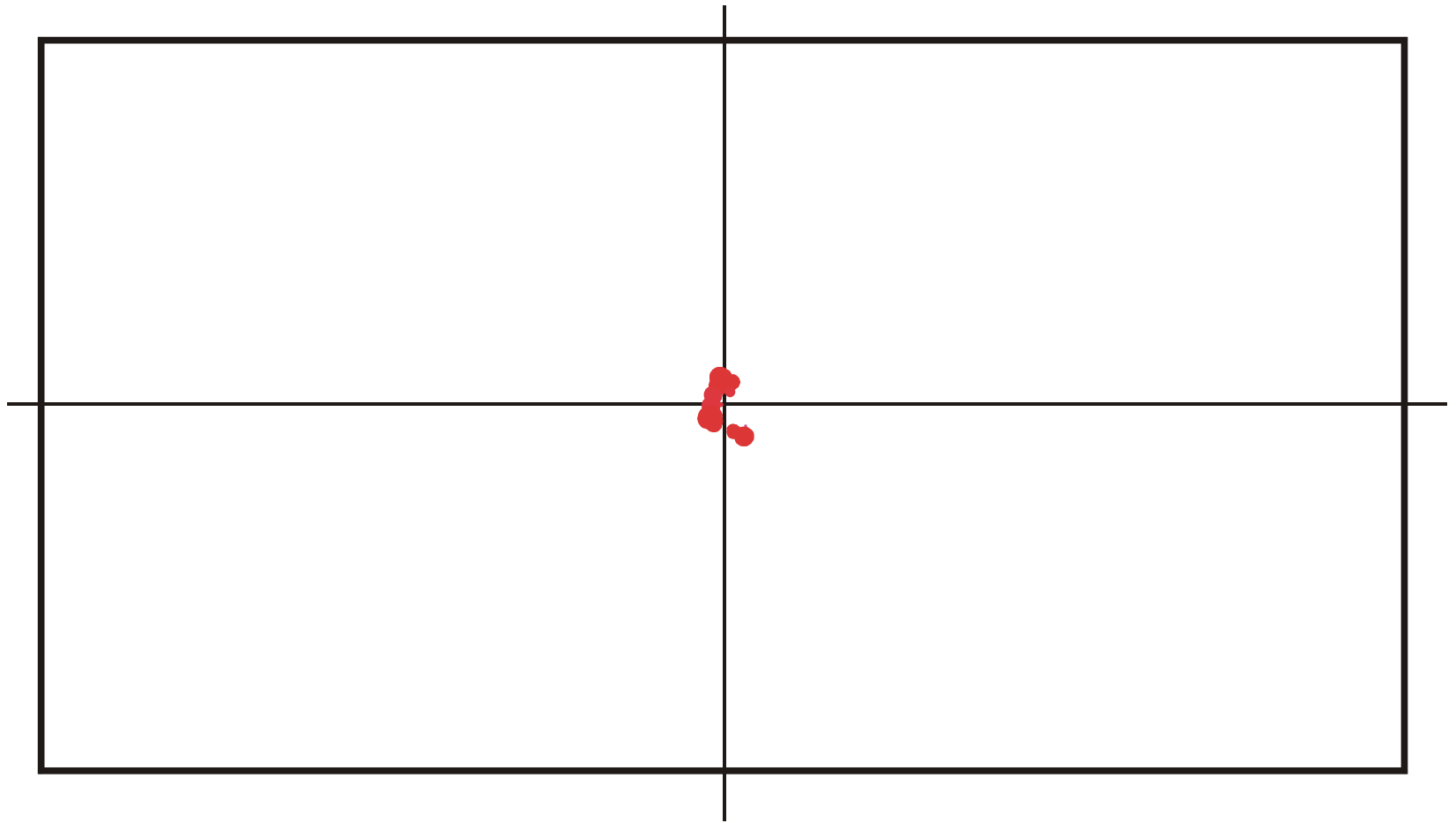


Variation in genotype space during optimization of phenotypes

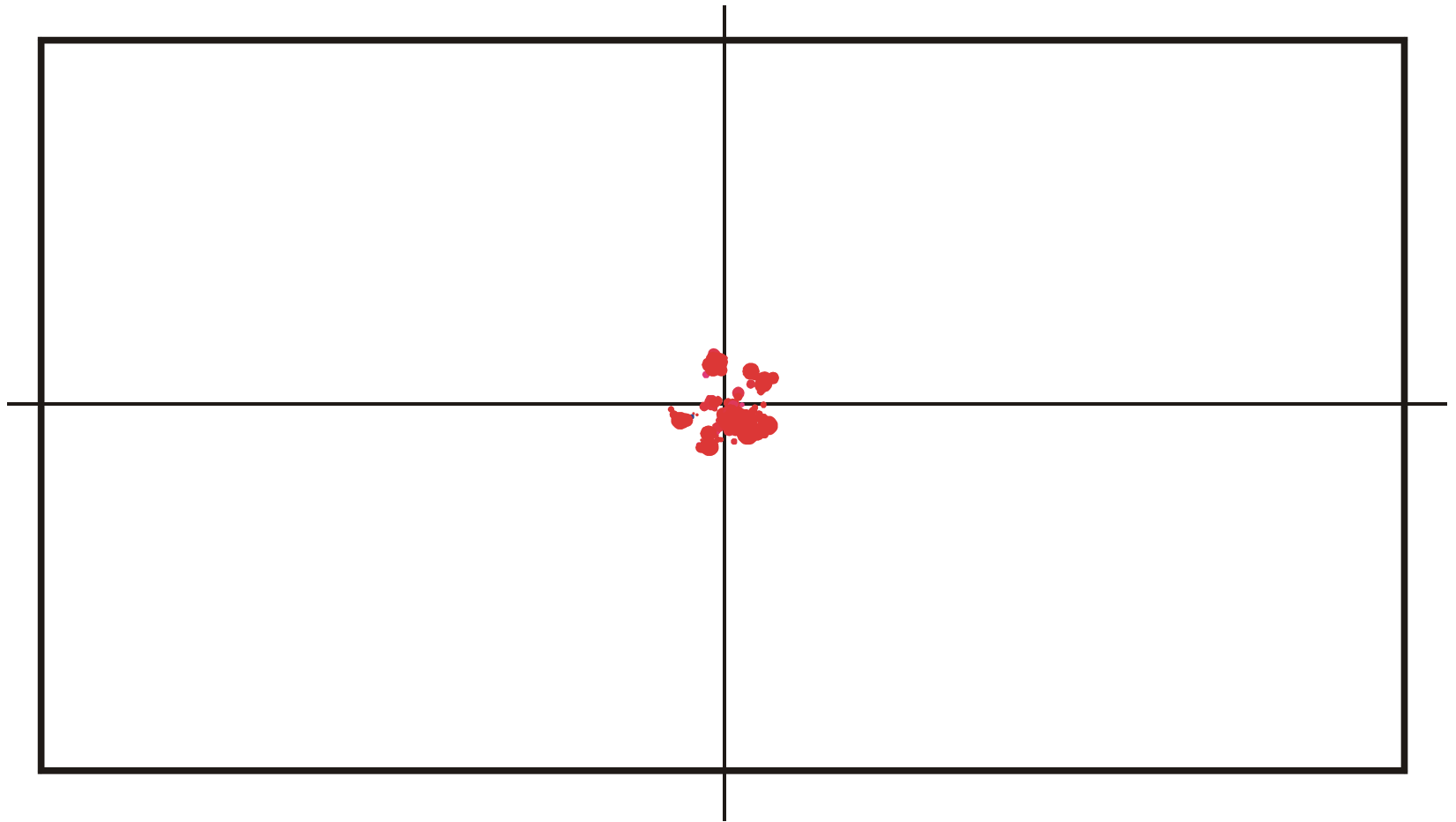
**Mean Hamming distance** within the population and **drift velocity of the population center** in sequence space.



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

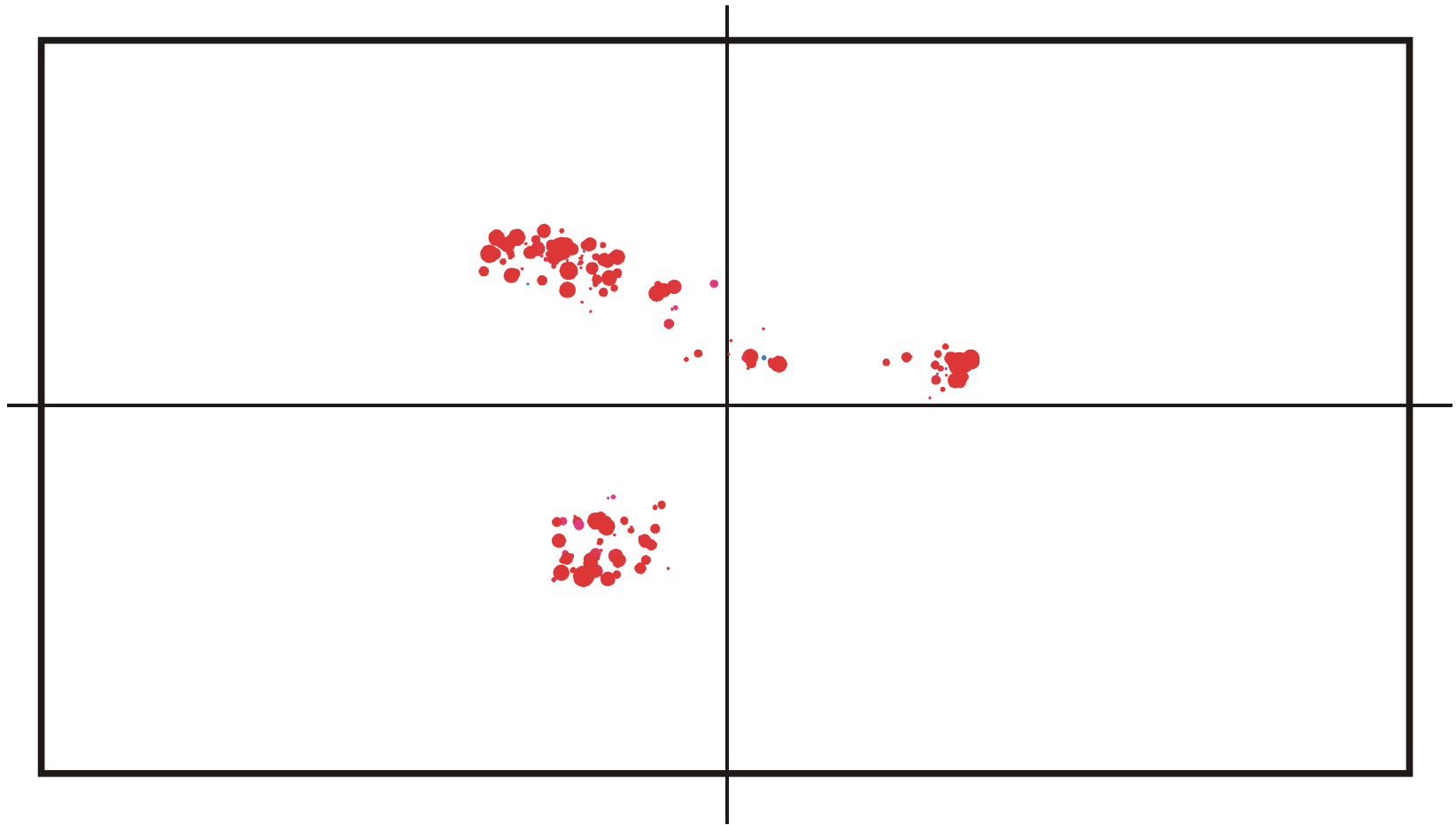


Spread of population in sequence space during a quasistationary epoch:  $t = 170$

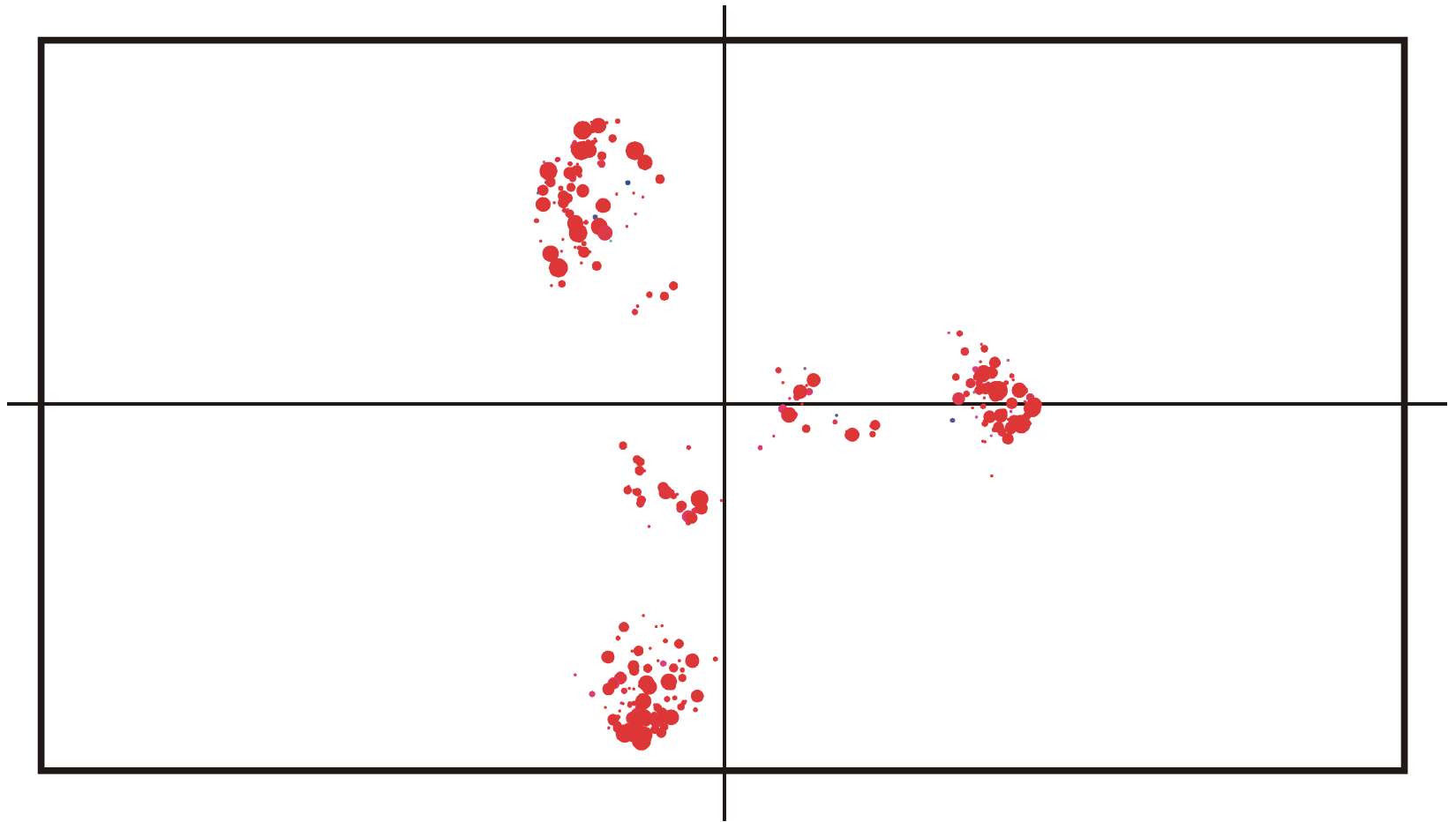


Spread of population in sequence space during a quasistationary epoch:  $t = 200$

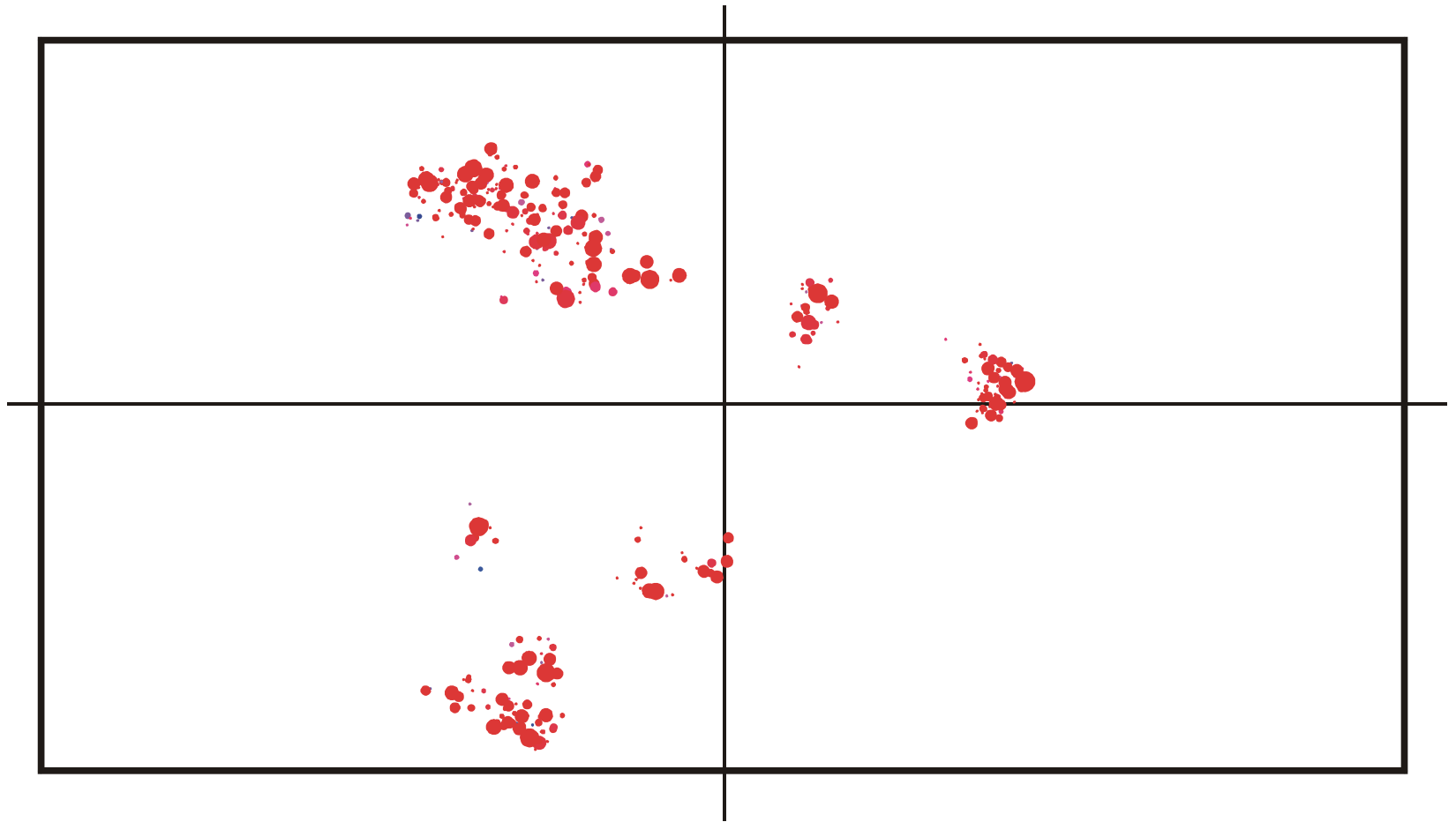




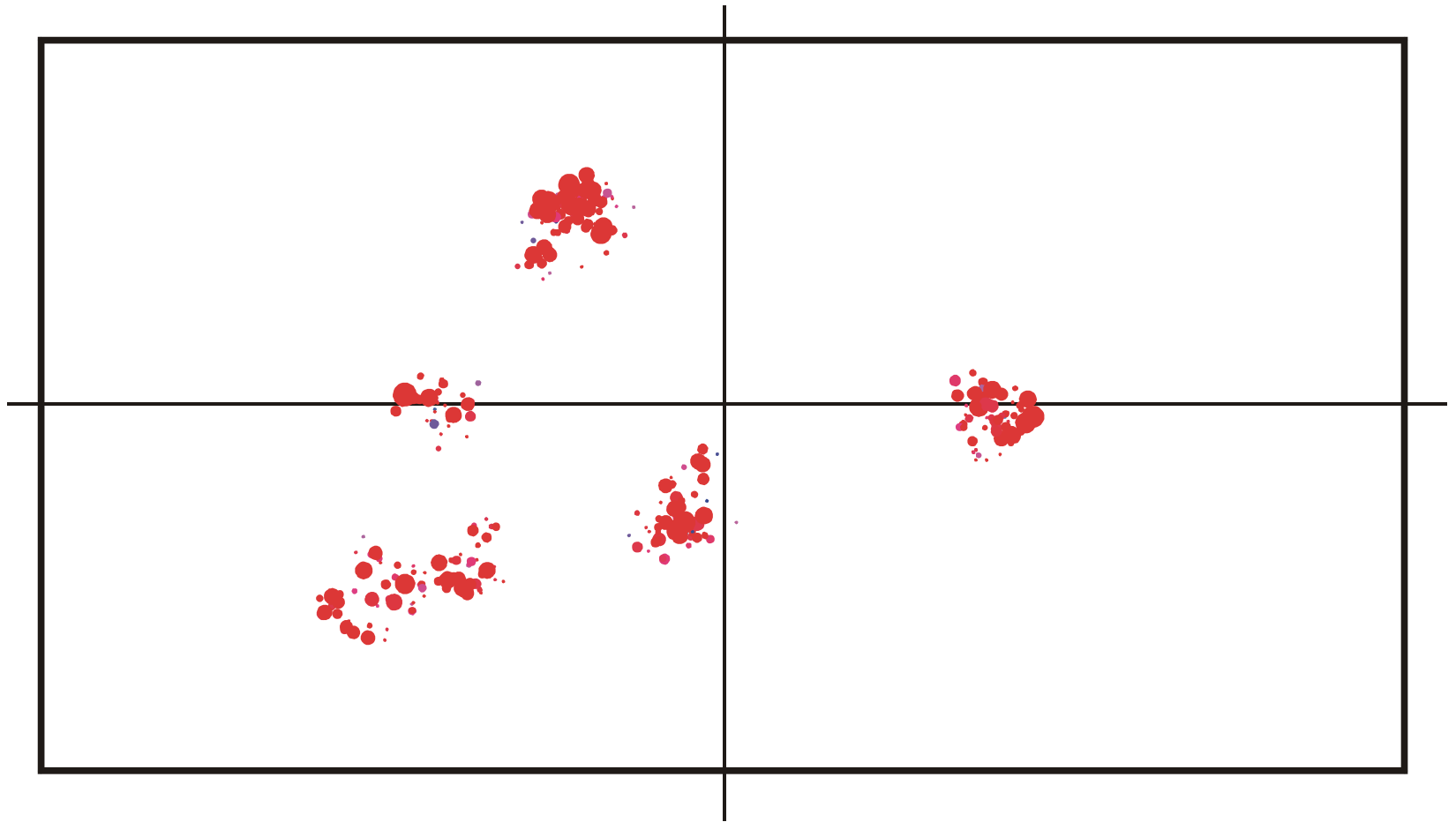
Spread of population in sequence space during a quasistationary epoch:  $t = 350$



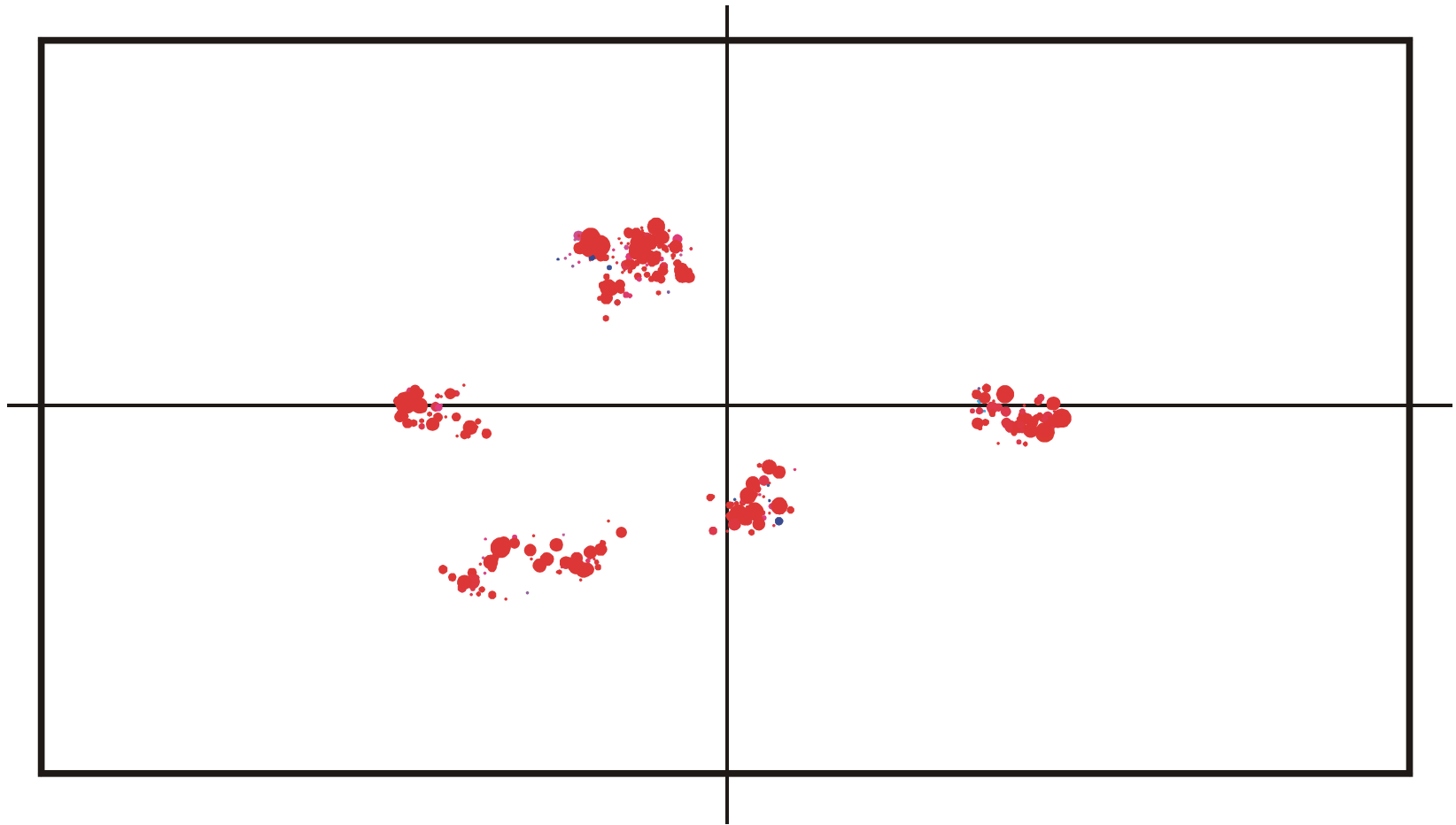
Spread of population in sequence space during a quasistationary epoch:  $t = 500$



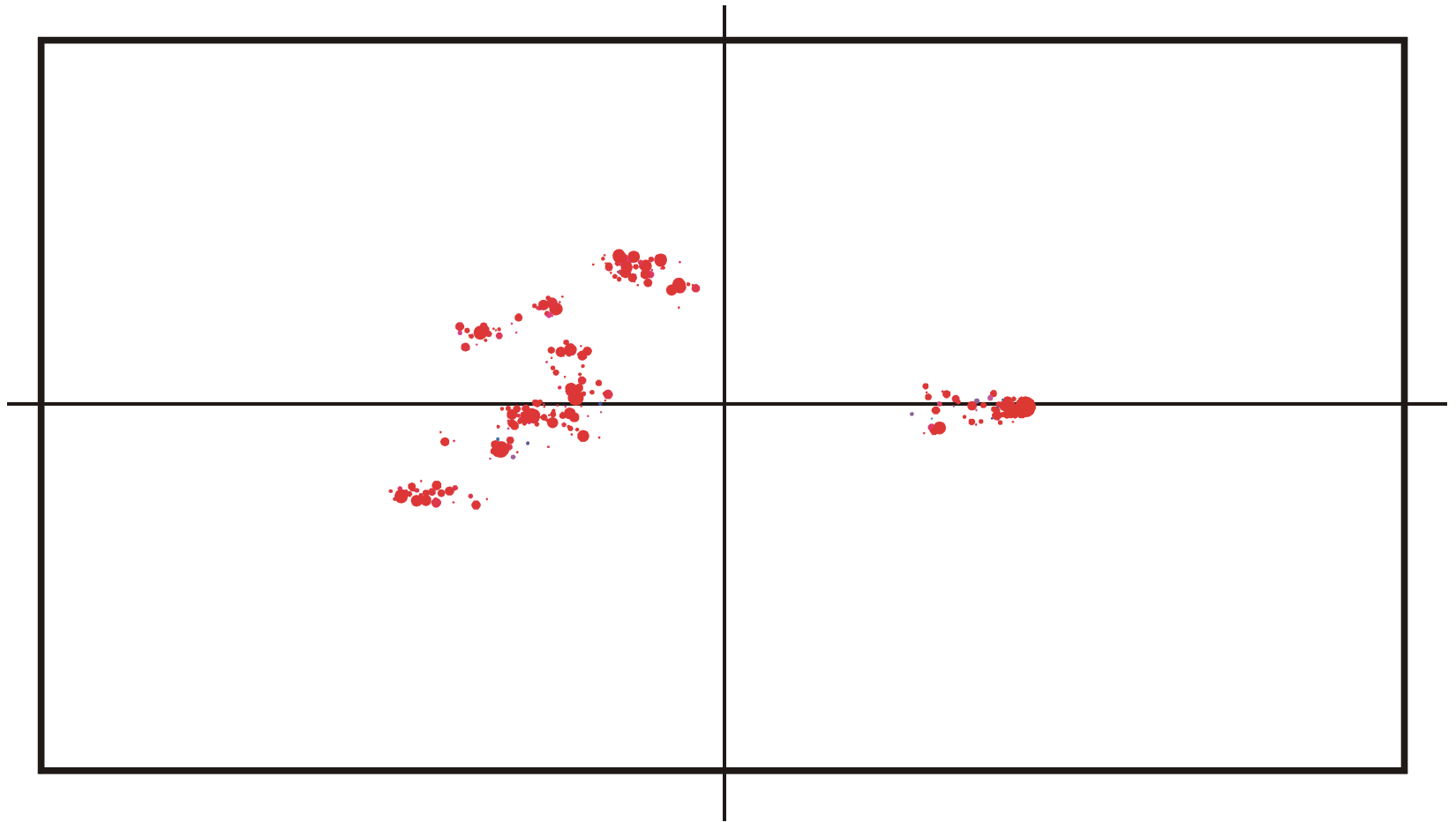
Spread of population in sequence space during a quasistationary epoch:  $t = 650$



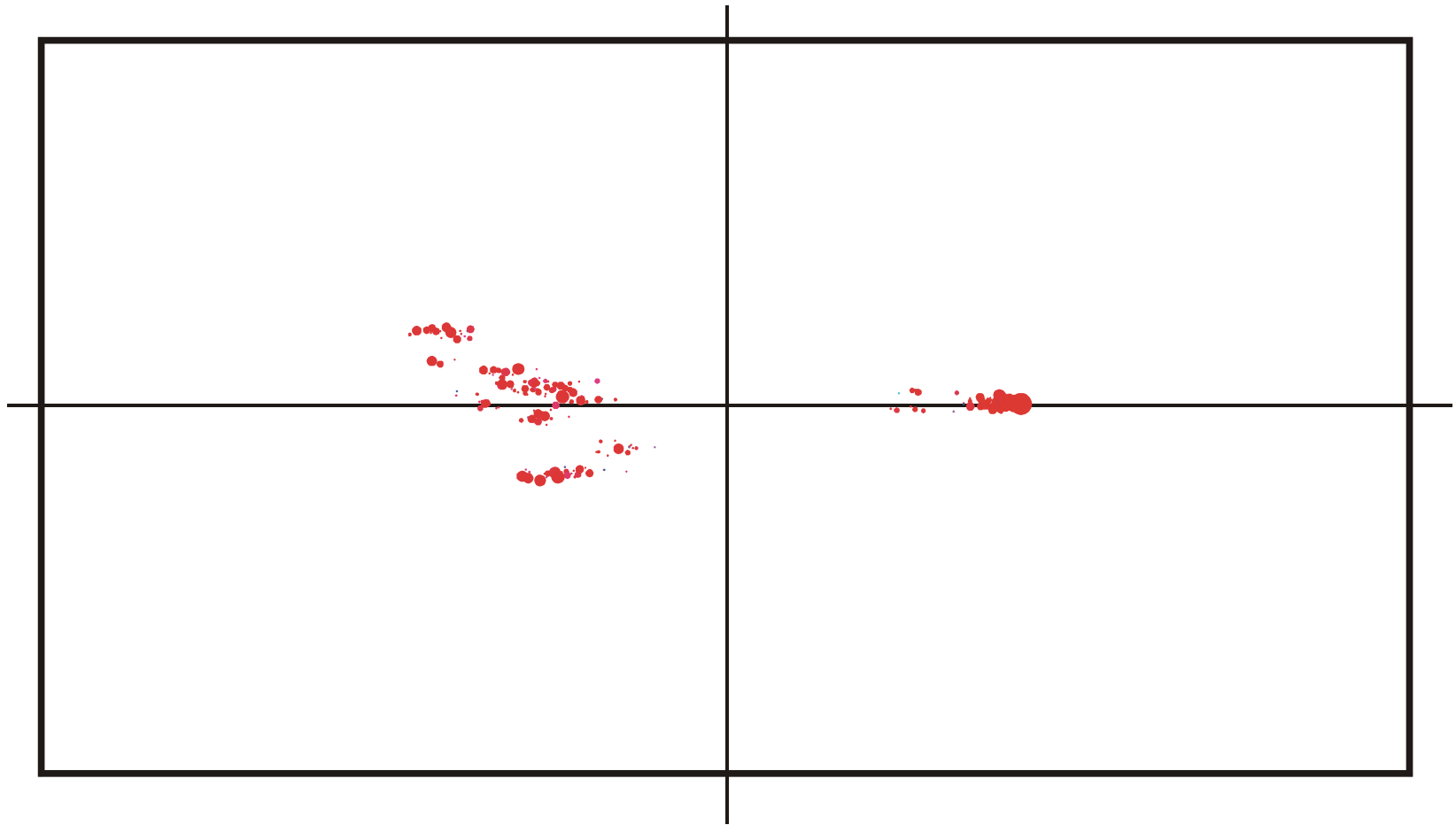
Spread of population in sequence space during a quasistationary epoch:  $t = 820$



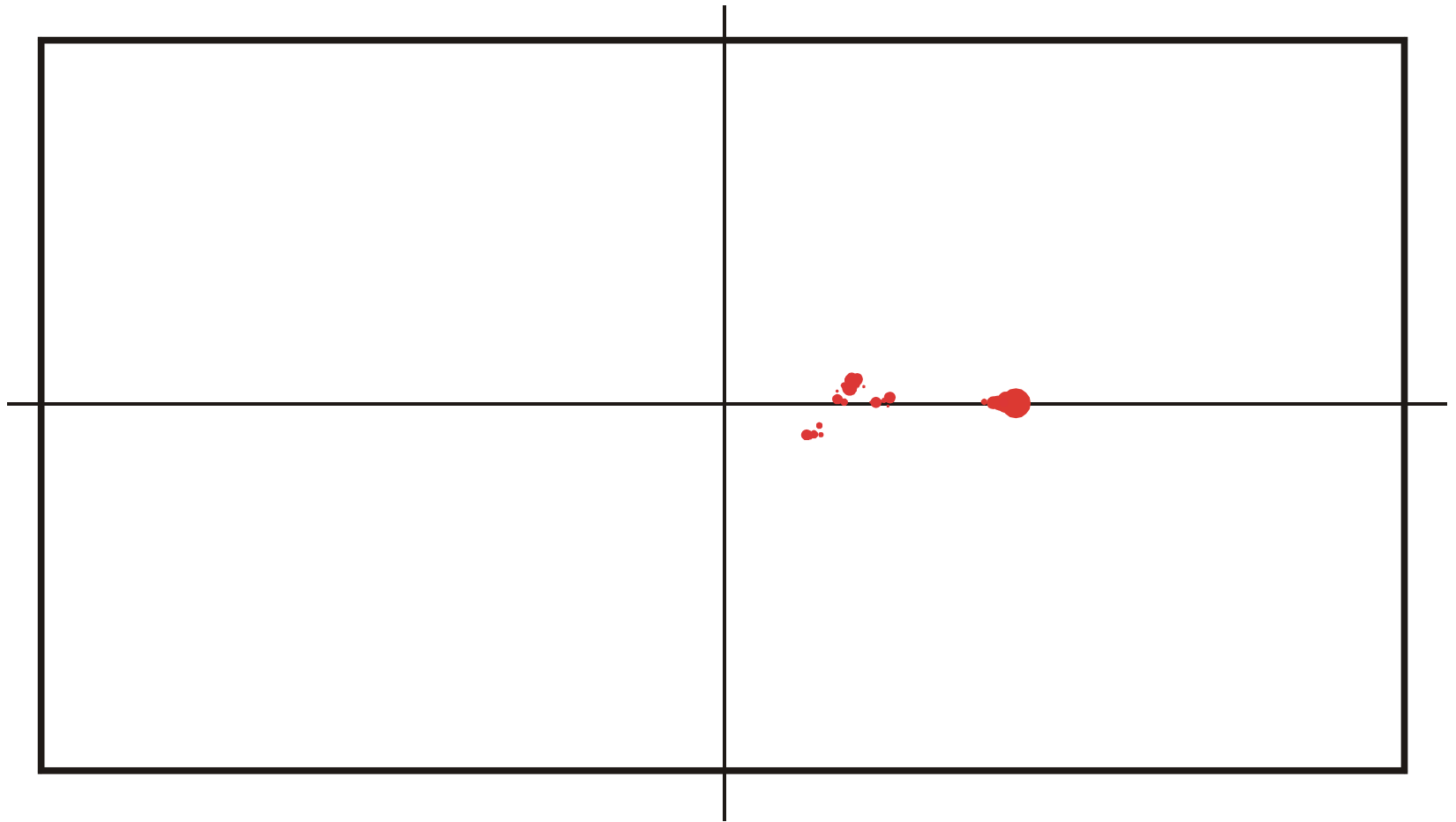
Spread of population in sequence space during a quasistationary epoch:  $t = 825$



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

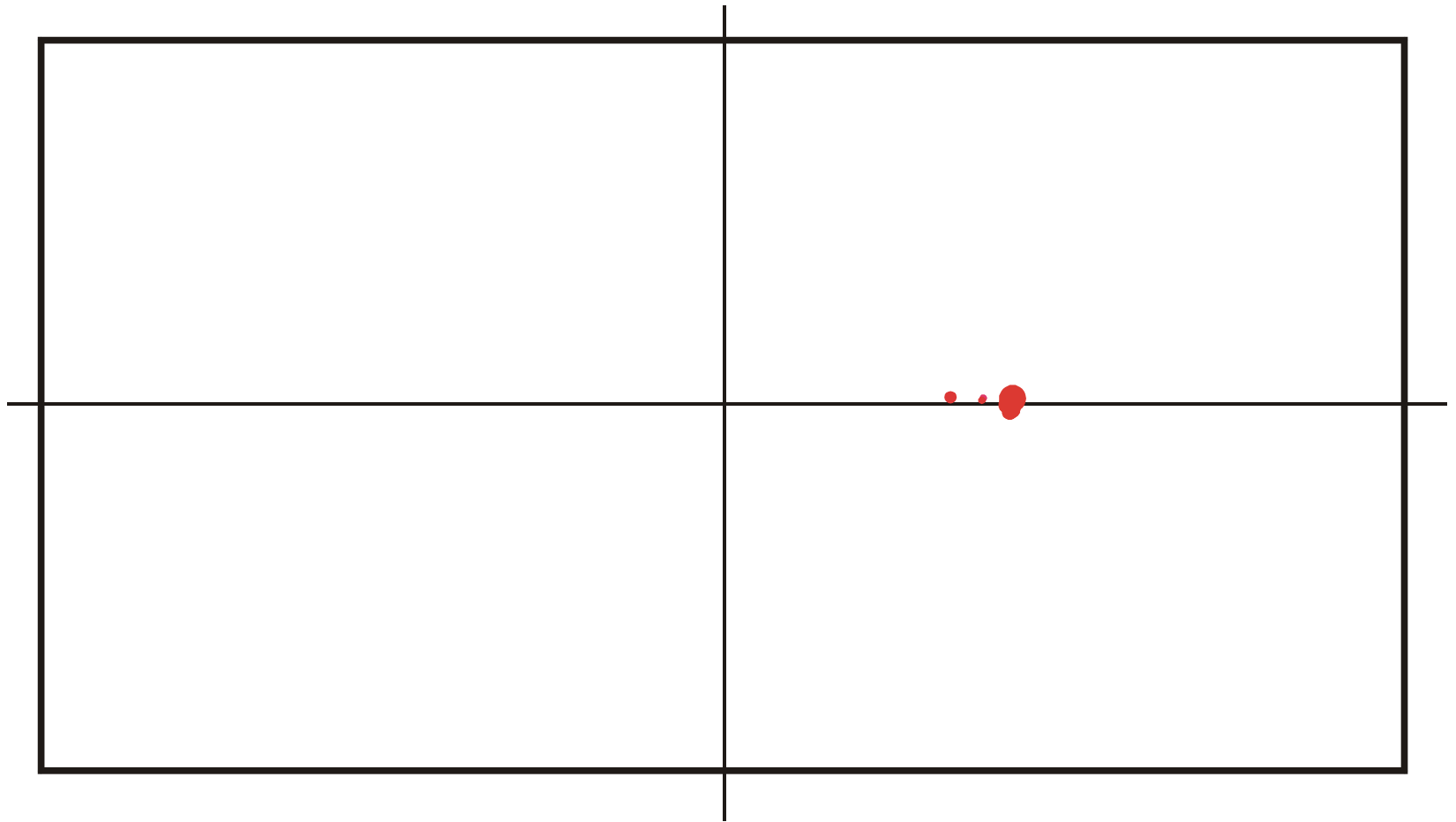


Spread of population in sequence space during a quasistationary epoch:  $t = 835$

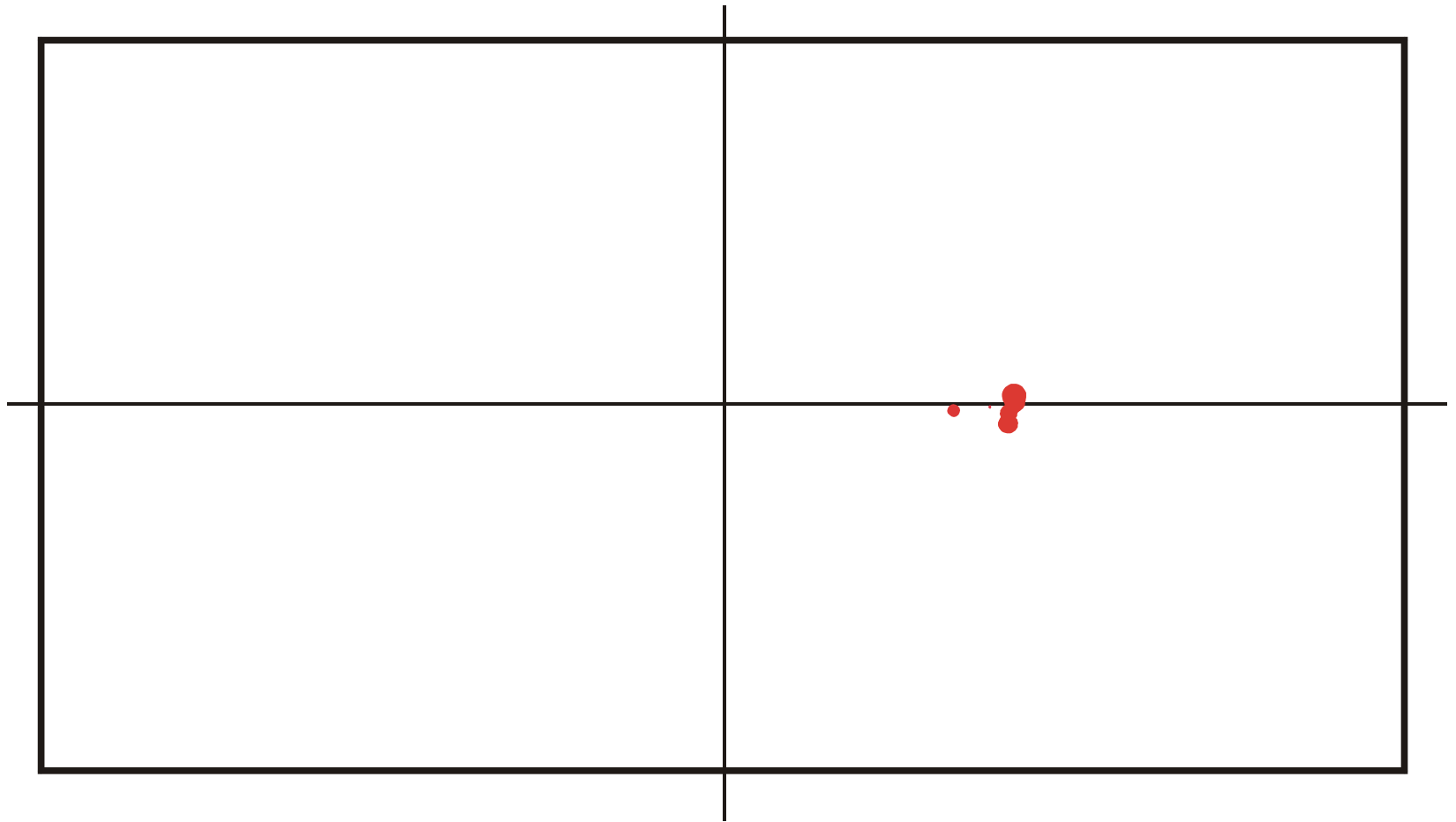


Spread of population in sequence space during a quasistationary epoch:  $t = 840$

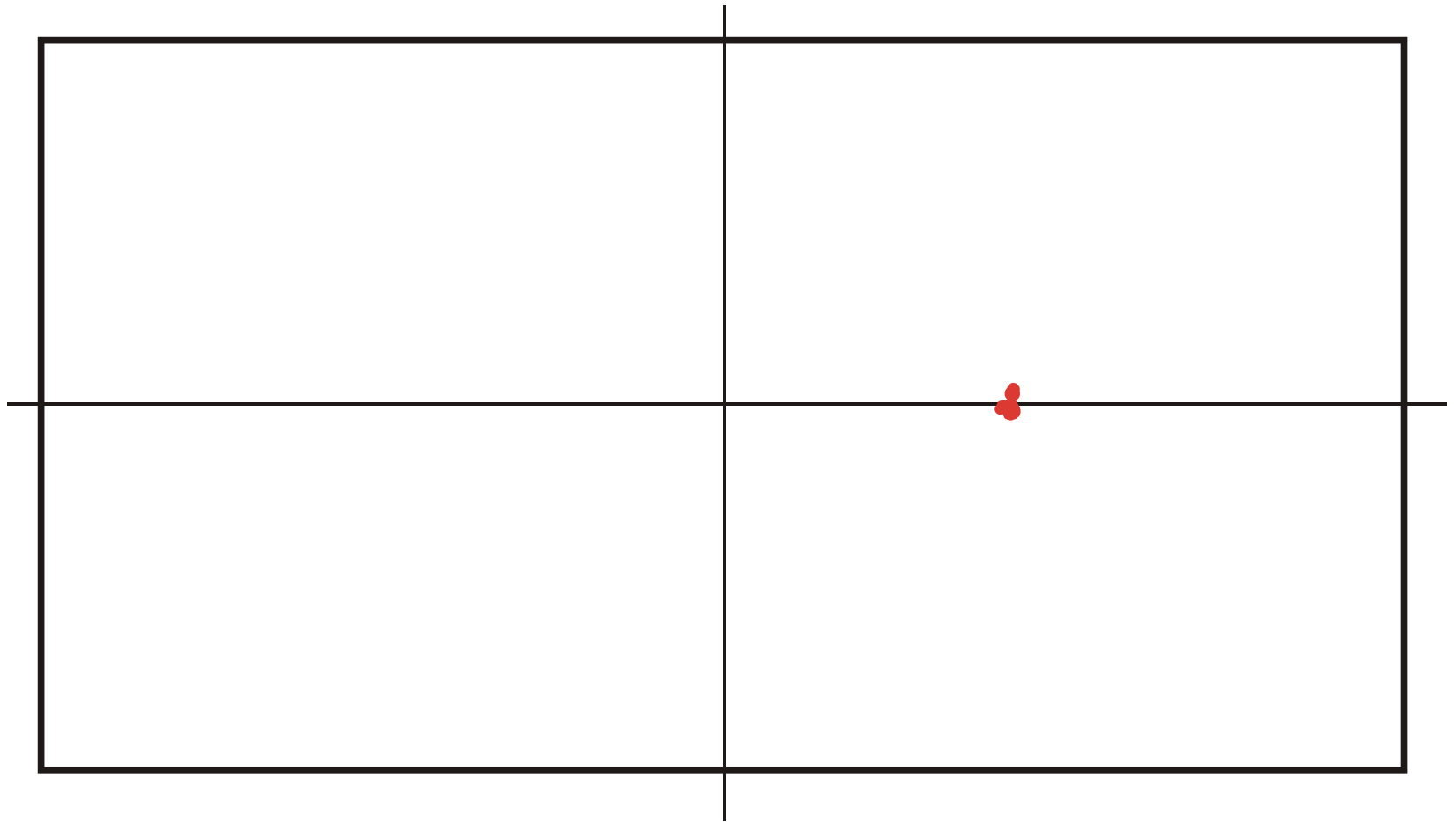




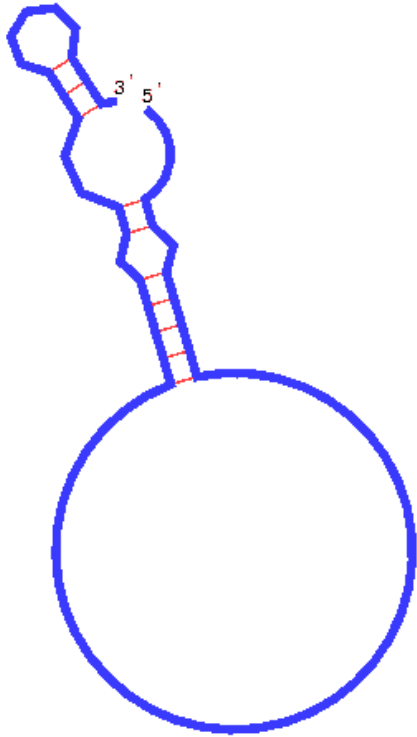
Spread of population in sequence space during a quasistationary epoch:  $t = 845$



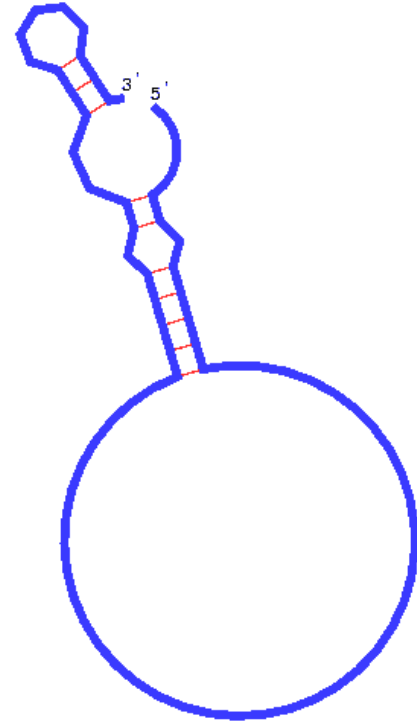
Spread of population in sequence space during a quasistationary epoch:  $t = 850$



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



**AUGC**



**GC**

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

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

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

## Theory of sequence – structure mappings

P. Schuster, W.Fontana, P.F.Stadler, I.L.Hofacker, *From sequences to shapes and back: A case study in RNA secondary structures*. Proc.Roy.Soc.London **B 255** (1994), 279-284

W.Grüner, R.Giegerich, D.Strothmann, C.Reidys, I.L.Hofacker, P.Schuster, *Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks*. Mh.Chem. **127** (1996), 355-374

W.Grüner, R.Giegerich, D.Strothmann, C.Reidys, I.L.Hofacker, P.Schuster, *Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structure of neutral networks and shape space covering*. Mh.Chem. **127** (1996), 375-389

C.M.Reidys, P.F.Stadler, P.Schuster, *Generic properties of combinatory maps*. Bull.Math.Biol. **59** (1997), 339-397

I.L.Hofacker, P. Schuster, P.F.Stadler, *Combinatorics of RNA secondary structures*. Discr.Appl.Math. **89** (1998), 177-207

C.M.Reidys, P.F.Stadler, *Combinatory landscapes*. SIAM Review **44** (2002), 3-54

# From sequences to shapes and back: a case study in RNA secondary structures

PETER SCHUSTER<sup>1,2,3</sup>, WALTER FONTANA<sup>3</sup>, PETER F. STADLER<sup>2,3</sup>  
AND IVO L. HOFACKER<sup>2</sup>

<sup>1</sup> Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, PF 100813, D-07708 Jena, Germany

<sup>2</sup> Institut für Theoretische Chemie, Universität Wien, Austria

<sup>3</sup> Santa Fe Institute, Santa Fe, U.S.A.

## SUMMARY

RNA folding is viewed here as a map assigning secondary structures to sequences. At fixed chain length the number of sequences far exceeds the number of structures. Frequencies of structures are highly non-uniform and follow a generalized form of Zipf's law: we find relatively few common and many rare ones. By using an algorithm for inverse folding, we show that sequences sharing the same structure are distributed randomly over sequence space. All common structures can be accessed from an arbitrary sequence by a number of mutations much smaller than the chain length. The sequence space is percolated by extensive neutral networks connecting nearest neighbours folding into identical structures. Implications for evolutionary adaptation and for applied molecular evolution are evident: finding a particular structure by mutation and selection is much simpler than expected and, even if catalytic activity should turn out to be sparse in the space of RNA structures, it can hardly be missed by evolutionary processes.

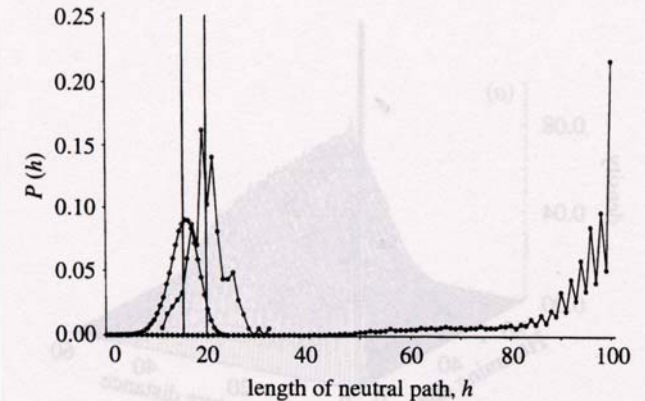
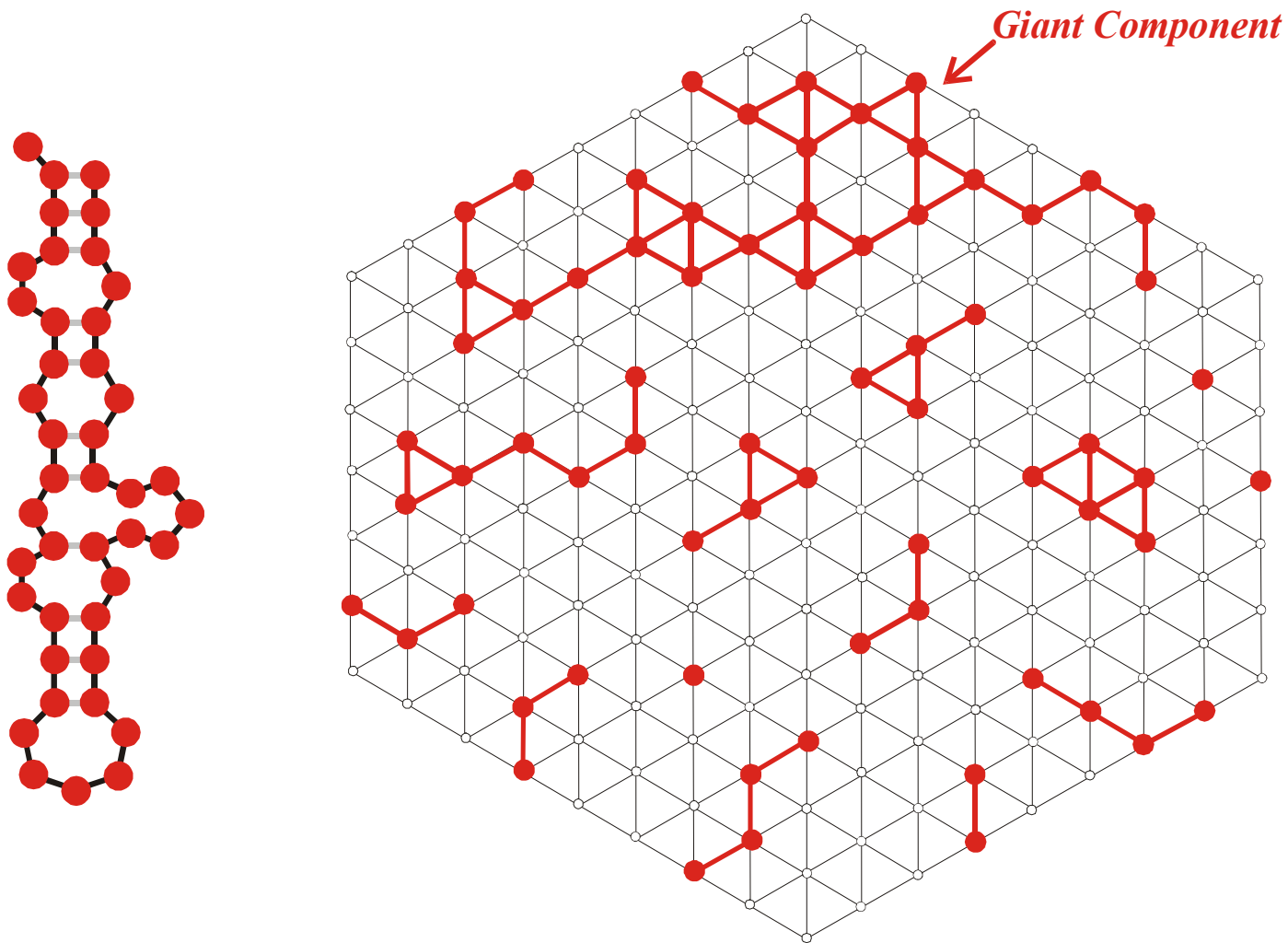


Figure 4. Neutral paths. A neutral path is defined by a series of nearest neighbour sequences that fold into identical structures. Two classes of nearest neighbours are admitted: neighbours of Hamming distance 1, which are obtained by single base exchanges in unpaired stretches of the structure, and neighbours of Hamming distance 2, resulting from base pair exchanges in stacks. Two probability densities of Hamming distances are shown that were obtained by searching for neutral paths in sequence space: (i) an upper bound for the closest approach of trial and target sequences (open circles) obtained as endpoints of neutral paths approaching the target from a random trial sequence (185 targets and 100 trials for each were used); (ii) a lower bound for the closest approach of trial and target sequences (open diamonds) derived from secondary structure statistics (Fontana *et al.* 1993a; see this paper, §4); and (iii) longest distances between the reference and the endpoints of monotonously diverging neutral paths (filled circles) (500 reference sequences were used).







A multi-component neutral network

---

# Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer

---

ZHEN HUANG<sup>1</sup> and JACK W. SZOSTAK<sup>2</sup>

<sup>1</sup>Department of Chemistry, Brooklyn College, Ph.D. Programs of Chemistry and Biochemistry, The Graduate School of CUNY, Brooklyn, New York 11210, USA

<sup>2</sup>Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

## ABSTRACT

Small changes in target specificity can sometimes be achieved, without changing aptamer structure, through mutation of a few bases. Larger changes in target geometry or chemistry may require more radical changes in an aptamer. In the latter case, it is unknown whether structural and functional solutions can still be found in the region of sequence space close to the original aptamer. To investigate these questions, we designed an *in vitro* selection experiment aimed at evolving specificity of an ATP aptamer. The ATP aptamer makes contacts with both the nucleobase and the sugar. We used an affinity matrix in which GTP was immobilized through the sugar, thus requiring extensive changes in or loss of sugar contact, as well as changes in recognition of the nucleobase. After just five rounds of selection, the pool was dominated by new aptamers falling into three major classes, each with secondary structures distinct from that of the ATP aptamer. The average sequence identity between the original aptamer and new aptamers is 76%. Most of the mutations appear to play roles either in disrupting the original secondary structure or in forming the new secondary structure or the new recognition loops. Our results show that there are novel structures that recognize a significantly different ligand in the region of sequence space close to the ATP aptamer. These examples of the emergence of novel functions and structures from an RNA molecule with a defined specificity and fold provide a new perspective on the evolutionary flexibility and adaptability of RNA.

**Keywords:** Aptamer; specificity; fold; selection; RNA evolution

*RNA* 9:1456-1463, 2003

Evidence for **neutral networks** and **shape space covering**

## Evolutionary Landscapes for the Acquisition of New Ligand Recognition by RNA Aptamers

Daniel M. Held, S. Travis Greathouse, Amit Agrawal, Donald H. Burke

Department of Chemistry, Indiana University, Bloomington, IN 47405-7102, USA

Received: 15 November 2002 / Accepted: 8 April 2003

**Abstract.** The evolution of ligand specificity underlies many important problems in biology, from the appearance of drug resistant pathogens to the re-engineering of substrate specificity in enzymes. In studying biomolecules, however, the contributions of macromolecular sequence to binding specificity can be obscured by other selection pressures critical to bioactivity. Evolution of ligand specificity *in vitro*—unconstrained by confounding biological factors—is addressed here using variants of three flavin-binding RNA aptamers. Mutagenized pools based on the three aptamers were combined and allowed to compete during *in vitro* selection for GMP-binding activity. The sequences of the resulting selection isolates were diverse, even though most were derived from the same flavin-binding parent. Individual GMP aptamers differed from the parental flavin aptamers by 7 to 26 mutations (20 to 57% overall change). Acquisition of GMP recognition coincided with the loss of FAD (flavin-adenine dinucleotide) recognition in all isolates, despite the absence of a counter-selection to remove FAD-binding RNAs. To examine more precisely the proximity of these two activities within a defined sequence space, the complete set of all intermediate sequences between an FAD-binding aptamer and a GMP-binding aptamer were synthesized and assayed for activity. For this set of sequences, we observe a portion of a neutral network for FAD-binding function separated from GMP-binding function by a distance of three muta-

tions. Furthermore, enzymatic probing of these aptamers revealed gross structural remodeling of the RNA coincident with the switch in ligand recognition. The capacity for neutral drift along an FAD-binding network in such close approach to RNAs with GMP-binding activity illustrates the degree of phenotypic buffering available to a set of closely related RNA sequences—defined as the set's functional tolerance for point mutations—and supports neutral evolutionary theory by demonstrating the facility with which a new phenotype becomes accessible as that buffering threshold is crossed.

**Key words:** Aptamers — RNA structure — Phenotypic buffering — Fitness landscapes — Neutral evolutionary theory — Flavin — GMP

### Introduction

RNA aptamers targeting small molecules serve as useful model systems for the study of the evolution and biophysics of macromolecular binding interactions. Because of their small sizes, the structures of several such complexes have been determined to atomic resolution by NMR spectrometry or X-ray crystallography (reviewed by Herman and Patel 2000). Moreover, aptamers can be subjected to mutational and evolutionary pressures for which survival is based entirely on ligand binding, without the complicating effects of simultaneous selection pressures for bioactivity, thus allowing the relative contributions of each activity to be evaluated separately.

Evidence for **neutral networks** and **intersection** of aptamer functions



S0092-8240(96)00089-4

## GENERIC PROPERTIES OF COMBINATORY MAPS: NEUTRAL NETWORKS OF RNA SECONDARY STRUCTURES<sup>1</sup>

■ CHRISTIAN REIDYS\*, †, PETER F. STADLER\*, ‡  
and PETER SCHUSTER\*, ‡, §, ¶<sup>2</sup>

\*Santa Fe Institute,  
Santa Fe, NM 87501, U.S.A.

†Los Alamos National Laboratory,  
Los Alamos, NM 87545, U.S.A.

‡Institut für Theoretische Chemie der Universität Wien,  
A-1090 Wien, Austria

§Institut für Molekulare Biotechnologie,  
D-07708 Jena, Germany

(E-mail: pks@tbi.univie.ac.at)

Random graph theory is used to model and analyse the relationships between sequences and secondary structures of RNA molecules, which are understood as mappings from sequence space into shape space. These maps are non-invertible since there are always many orders of magnitude more sequences than structures. Sequences folding into identical structures form *neutral networks*. A neutral network is embedded in the set of sequences that are *compatible* with the given structure. Networks are modeled as graphs and constructed by random choice of vertices from the space of compatible sequences. The theory characterizes neutral networks by the mean fraction of neutral neighbors ( $\lambda$ ). The networks are connected and percolate sequence space if the fraction of neutral nearest neighbors exceeds a threshold value ( $\lambda > \lambda^*$ ). Below threshold ( $\lambda < \lambda^*$ ), the networks are partitioned into a largest “giant” component and several smaller components. Structures are classified as “common” or “rare” according to the sizes of their pre-images, i.e. according to the fractions of sequences folding into them. The neutral networks of any pair of two different common structures almost touch each other, and, as expressed by the conjecture of *shape space covering* sequences folding into almost all common structures, can be found in a small ball of an arbitrary location in sequence space. The results from random graph theory are compared to data obtained by folding large samples of RNA sequences. Differences are explained in terms of specific features of RNA molecular structures. © 1997 Society for Mathematical Biology

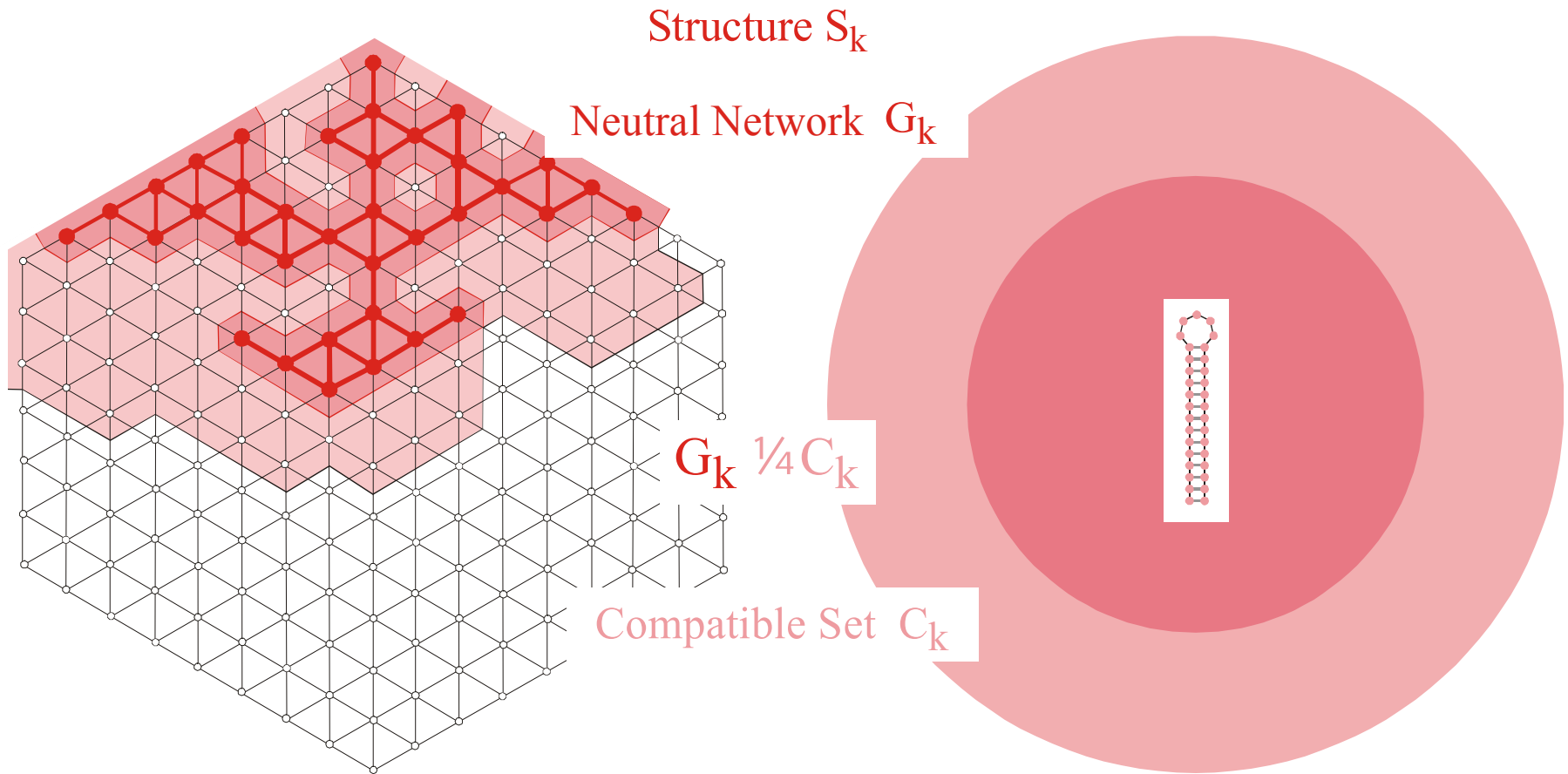
**THEOREM 5. INTERSECTION-THEOREM.** *Let  $s$  and  $s'$  be arbitrary secondary structures and  $C[s], C[s']$  their corresponding compatible sequences. Then,*

$$C[s] \cap C[s'] \neq \emptyset.$$

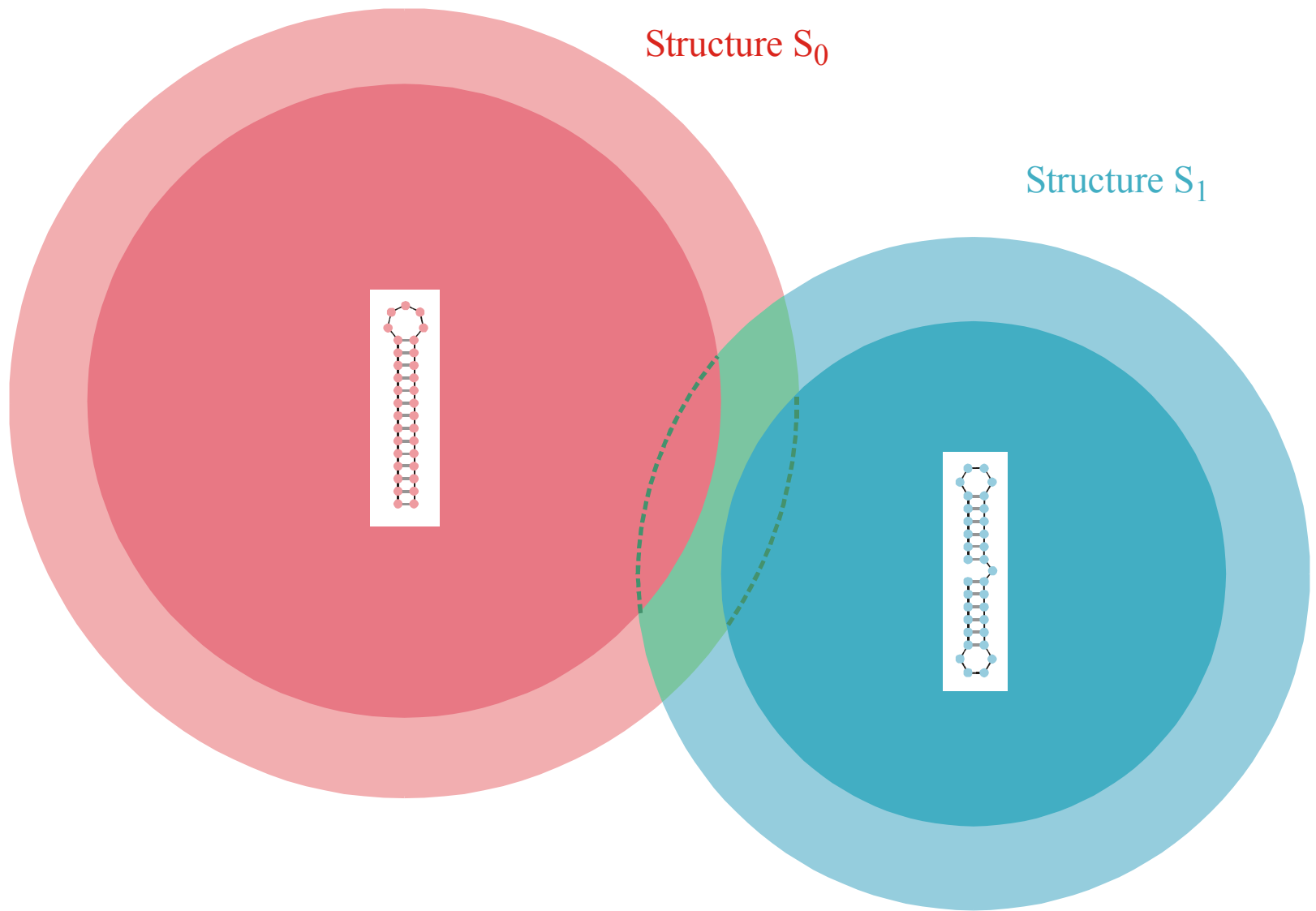
*Proof.* Suppose that the alphabet admits only the complementary base pair  $[XY]$  and we ask for a sequence  $x$  compatible to both  $s$  and  $s'$ . Then  $f(s, s') \cong D_m$  operates on the set of all positions  $\{x_1, \dots, x_n\}$ . Since we have the operation of a dihedral group, the orbits are either cycles or chains and the cycles have even order. A constraint for the sequence compatible to both structures appears only in the cycles where the choice of bases is not independent. It remains to be shown that there is a valid choice of bases for each cycle, which is obvious since these have even order. Therefore, it suffices to choose an alternating sequence of the pairing partners  $X$  and  $Y$ . Thus, there are at least two different choices for the first base in the orbit. ■

*Remark.* A generalization of the statement of theorem 5 to three different structures is false.

Reference for the definition of the intersection  
and the proof of the **intersection theorem**



The **compatible set**  $C_k$  of a structure  $S_k$  consists of all sequences which form  $S_k$  as its minimum free energy structure (the **neutral network**  $G_k$ ) or one of its suboptimal structures.



**Intersection** of two compatible sets:  $C_0 \cap C_1$

The intersection of two compatible sets is always non empty:  $C_0 \cap C_1 \neq \emptyset$



- minus the background levels observed in the HSP in the control (Sar1-GDP-containing) incubation that prevents COPII vesicle formation. In the microsome control, the level of p115-SNARE associations was less than 0.1%.
46. C. M. Carr, E. Grote, M. Munson, F. M. Hughson, P. J. Novick, *J. Cell Biol.* **146**, 333 (1999).
  47. C. Ungermann, B. J. Nichols, H. R. Pelham, W. Wickner, *J. Cell Biol.* **140**, 61 (1998).
  48. E. Grote and P. J. Novick, *Mol. Biol. Cell* **10**, 4149 (1999).
  49. P. Uetz et al., *Nature* **403**, 623 (2000).
  50. GST-SNARE proteins were expressed in bacteria and purified on glutathione-Sepharose beads using standard methods. Immobilized GST-SNARE protein (0.5  $\mu$ M) was incubated with rat liver cytosol (20 mg) or purified recombinant p115 (0.5  $\mu$ M) in 1 ml of NS buffer containing 1% BSA for 2 hours at 4°C with rotation. Beads were briefly spun (3000 rpm for 10 s) and sequentially washed three times with NS buffer and three times with NS buffer supplemented with 150 mM NaCl. Bound proteins were eluted three times in 50  $\mu$ l of 50 mM tris-HCl (pH 8.5), 50 mM reduced glutathione, 150 mM NaCl, and 0.1% Triton X-100 for 15 min at 4°C with intermittent mixing, and elutes were pooled. Proteins were precipitated by MeOH/CH<sub>2</sub>Cl<sub>2</sub> and separated by SDS-polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting using p115 mAb 13F12.
  51. V. Rybin et al., *Nature* **383**, 266 (1996).
  52. K. G. Hardwick and H. R. Pelham, *J. Cell Biol.* **119**, 513 (1992).
  53. A. P. Newman, M. E. Groesch, S. Ferro-Novick, *EMBO J.* **11**, 3609 (1992).
  54. A. Spang and R. Schekman, *J. Cell Biol.* **143**, 589 (1998).
  55. M. F. Rexach, M. Latterich, R. W. Schekman, *J. Cell Biol.* **126**, 1133 (1994).
  56. A. Mayer and W. Wickner, *J. Cell Biol.* **136**, 307 (1997).
  57. M. D. Turner, H. Plutner, W. E. Balch, *J. Biol. Chem.* **272**, 13479 (1997).
  58. A. Price, D. Seals, W. Wickner, C. Ungermann, *J. Cell Biol.* **148**, 1231 (2000).
  59. X. Cao and C. Barlowe, *J. Cell Biol.* **149**, 55 (2000).
  60. G. G. Tall, H. Hama, D. B. DeWald, B. F. Horadzovsky, *Mol. Biol. Cell* **10**, 1873 (1999).
  61. C. G. Burd, M. Peterson, C. R. Cowles, S. D. Emr, *Mol. Biol. Cell* **8**, 1089 (1997).
  62. M. R. Peterson, C. G. Burd, S. D. Emr, *Curr. Biol.* **9**, 159 (1999).
  63. M. G. Waters, D. O. Clary, J. E. Rothman, *J. Cell Biol.* **118**, 1015 (1992).
  64. D. M. Walter, K. S. Paul, M. G. Waters, *J. Biol. Chem.* **273**, 29565 (1998).
  65. N. Hui et al., *Mol. Biol. Cell* **8**, 1777 (1997).
  66. T. E. Kreis, *EMBO J.* **5**, 931 (1986).
  67. H. Plutner, H. W. Davidson, J. Saraste, W. E. Balch, *J. Cell Biol.* **119**, 1097 (1992).
  68. D. S. Nelson et al., *J. Cell Biol.* **143**, 319 (1998).
  69. We thank G. Waters for p115 cDNA and p115 mAbs; G. Warren for p97 and p47 antibodies; R. Scheller for rbt1, membrin, and sec22 cDNAs; H. Plutner for excellent technical assistance; and P. Tan for help during the initial phase of this work. Supported by NIH grants GM 33301 and GM42336 and National Cancer Institute grant CA58689 (W.E.B.), a NIH National Research Service Award (B.D.M.), and a Wellcome Trust International Traveling Fellowship (B.B.A.).

20 March 2000; accepted 22 May 2000

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

Erik A. Schultes and David P. Bartel\*

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

Related protein or RNA sequences with the same folded conformation can often perform very different biochemical functions, indicating that new biochemical functions can arise from preexisting folds. But what evolutionary mechanisms give rise to sequences with new macromolecular folds? When considering the origin of new folds, it is useful to picture, among all sequence possibilities, the distribution of sequences with a particular fold and function. This distribution can range very far in sequence space (1). For example, only seven nucleotides are strictly conserved among the group I self-splicing introns, yet secondary (and presumably tertiary) structure within the core of the ribozyme is preserved (2). Because these dis-

parate isolates have the same fold and function, it is thought that they descended from a common ancestor through a series of mutational variants that were each functional. Hence, sequence heterogeneity among divergent isolates implies the existence of paths through sequence space that have allowed neutral drift from the ancestral sequence to each isolate. The set of all possible neutral paths composes a "neutral network," connecting in sequence space those widely dispersed sequences sharing a particular fold and activity, such that any sequence on the network can potentially access very distant sequences by neutral mutations (3-5).

Theoretical analyses using algorithms for predicting RNA secondary structure have suggested that different neutral networks are interwoven and can approach each other very closely (3, 5-8). Of particular interest is whether ribozyme neutral networks approach each other so closely that they intersect. If so, a single sequence would be capable of folding into two different conformations, would

have two different catalytic activities, and could access by neutral drift every sequence on both networks. With intersecting networks, RNAs with novel structures and activities could arise from previously existing ribozymes, without the need to carry non-functional sequences as evolutionary intermediates. Here, we explore the proximity of neutral networks experimentally, at the level of RNA function. We describe a close apposition of the neutral networks for the hepatitis delta virus (HDV) self-cleaving ribozyme and the class III self-ligating ribozyme.

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

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

## A ribozyme switch

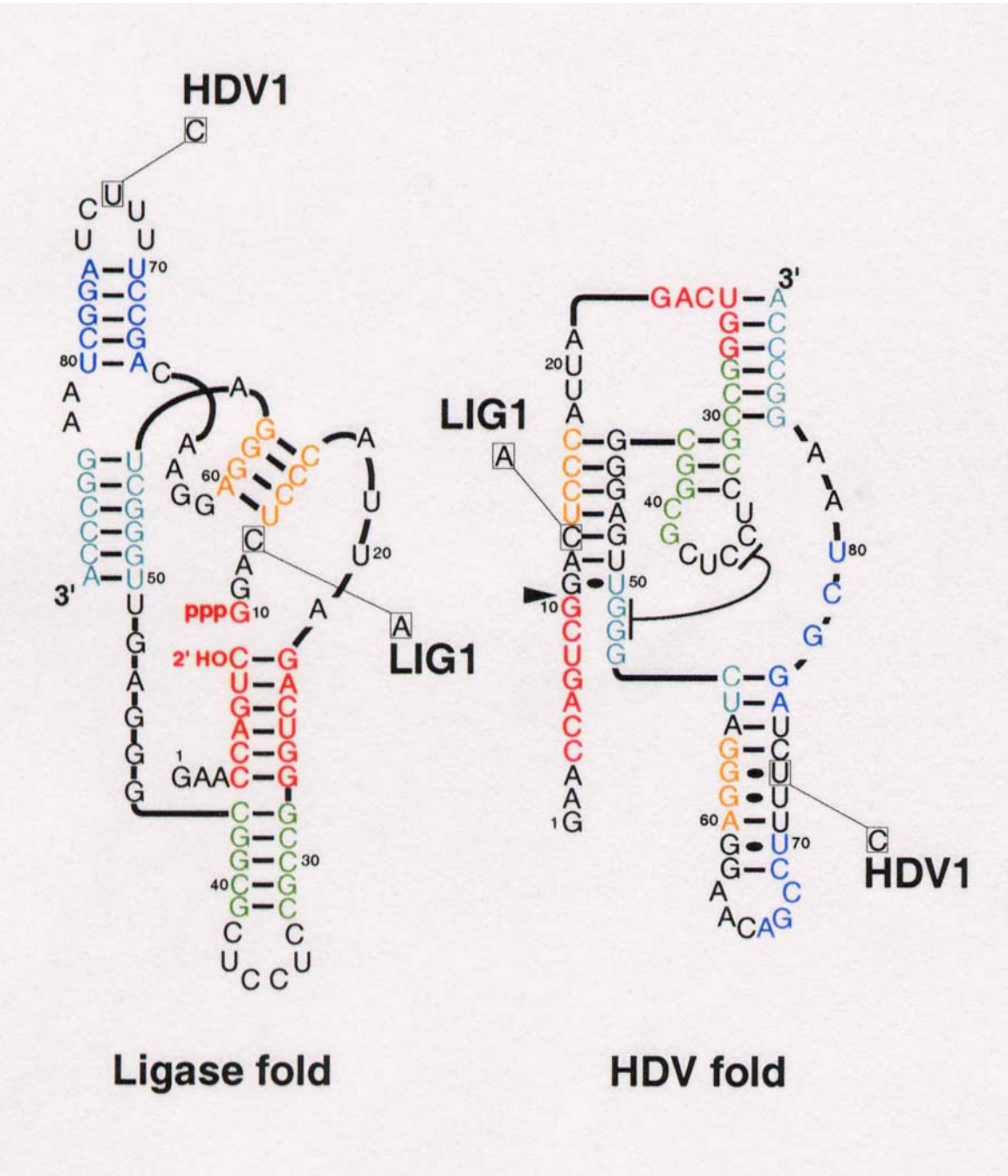
E.A.Schultes, D.B.Bartel, *Science*  
**289** (2000), 448-452

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

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

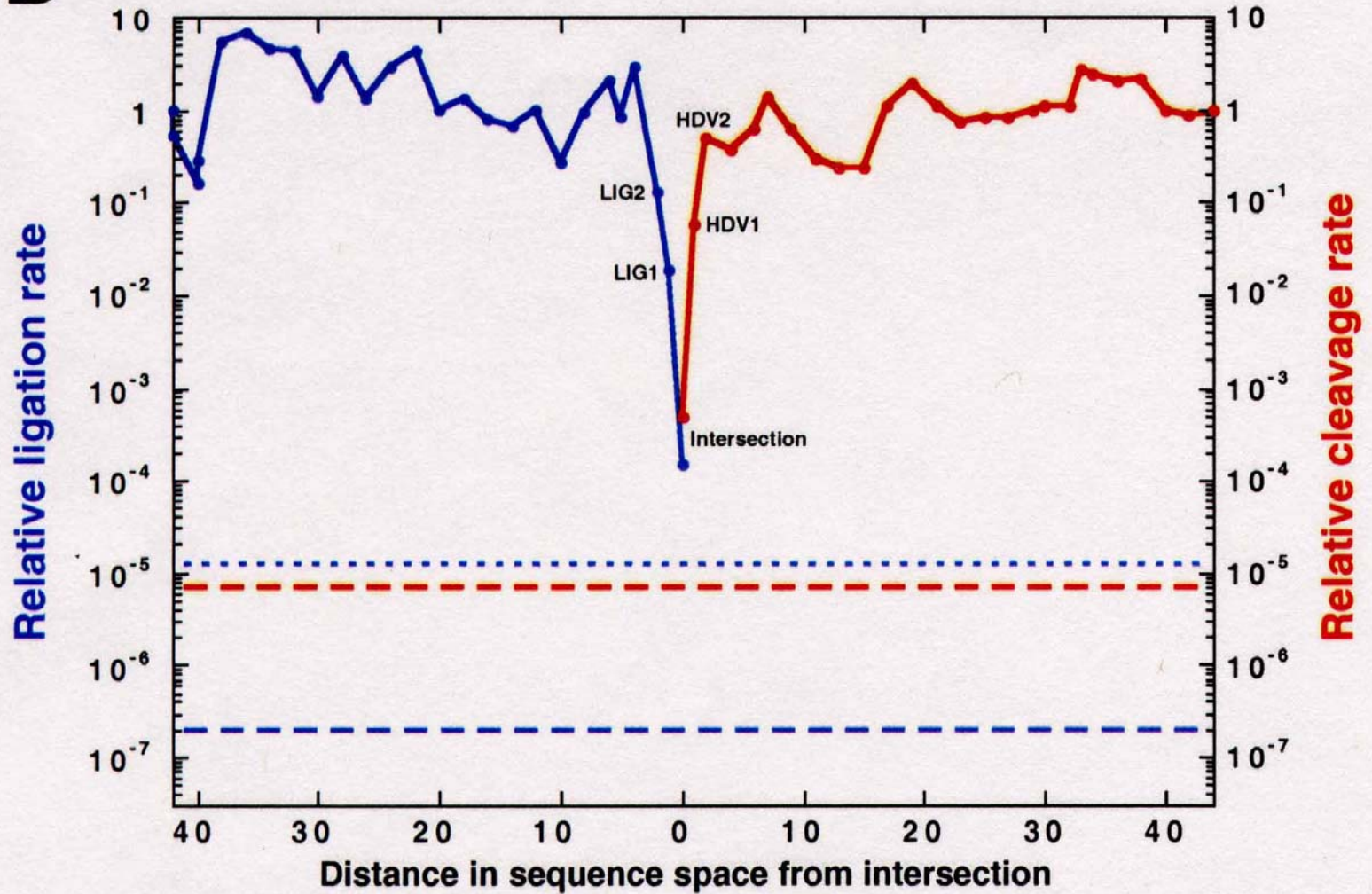




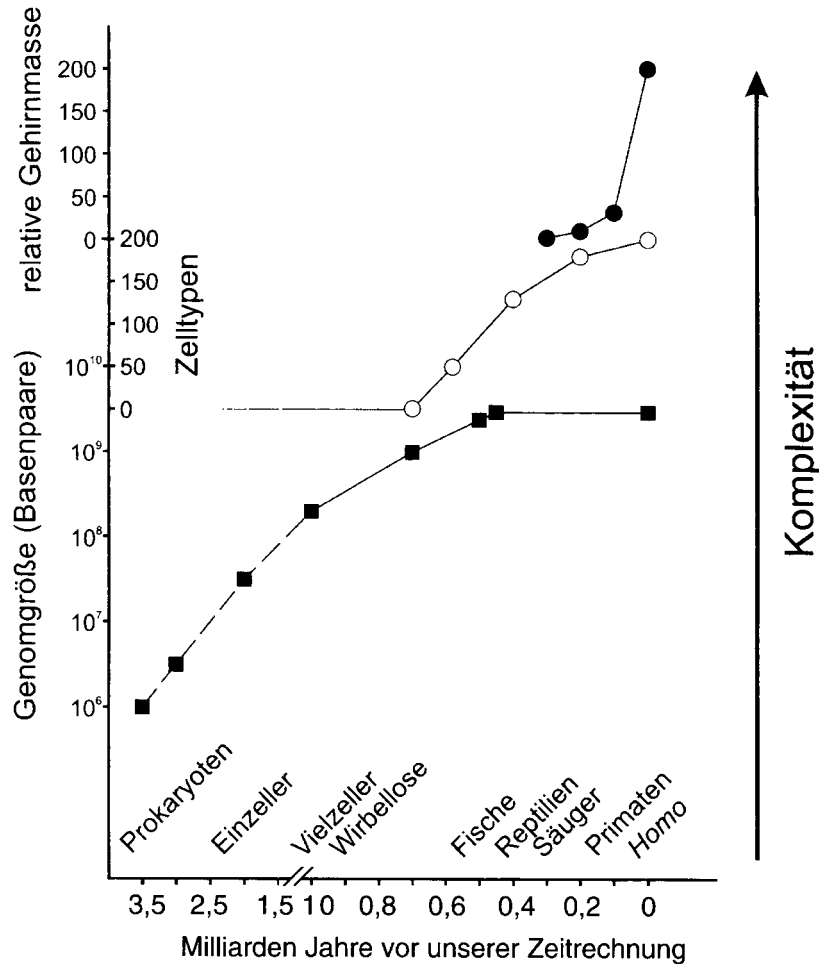


The sequence at the *intersection*:

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

**B**

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



**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<sup>9</sup> Jahre, eine andere für den Zeitraum <10<sup>9</sup> 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.





Evolution (Cartoon<sup>a</sup> 1980)

2004

## 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

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



Web-Page for further information:

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

