



# Design of Nucleic Acid Molecules for Predefined Purposes

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria

and

The Santa Fe Institute, Santa Fe, New Mexico, USA



Viennano 2007

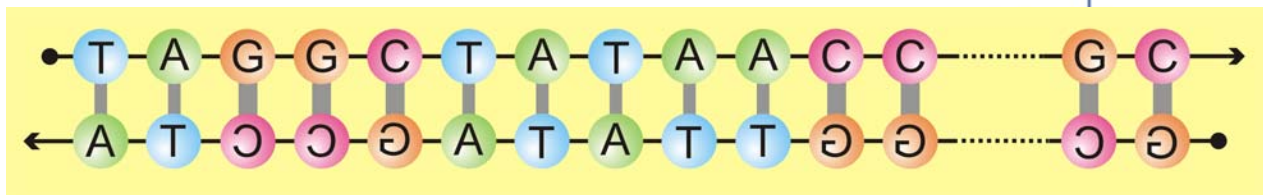
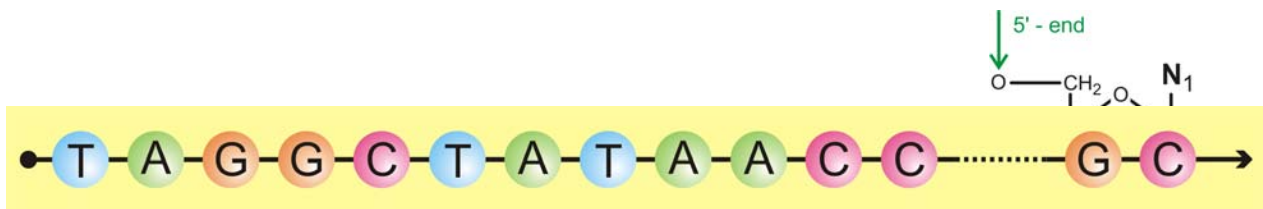
Wiener Neustadt, 14.03.2007

Web-Page for further information:

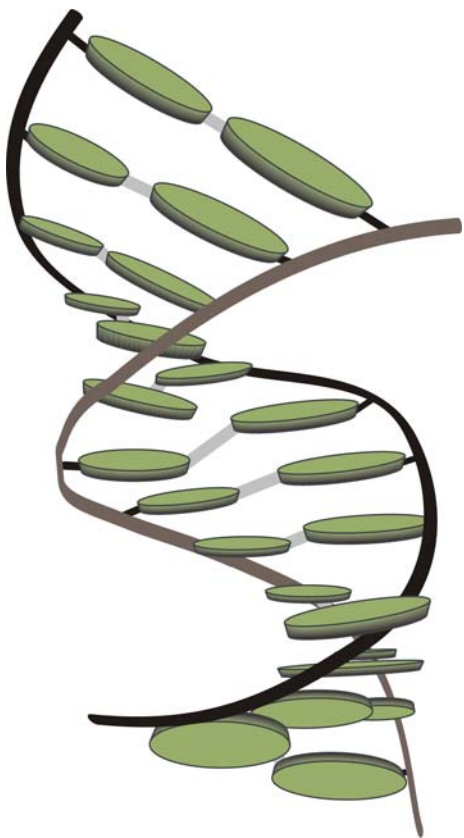
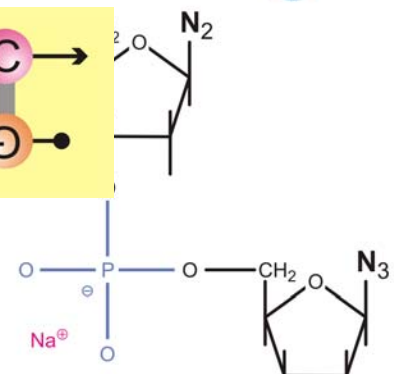
<http://www.tbi.univie.ac.at/~pks>

1. Nucleic acid structures
2. DNA nanotechnology
3. RNA - A magic molecule
4. Evolutionary optimization of structure
5. RNA design

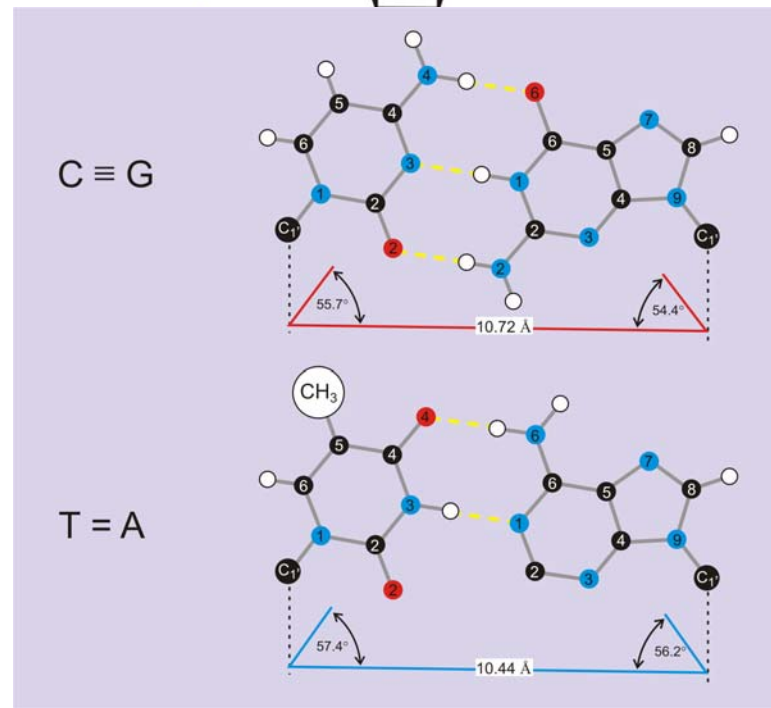
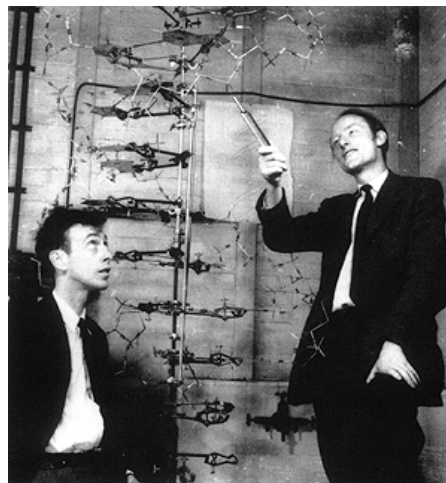
- 1. Nucleic acid structures**
2. DNA nanotechnology
3. RNA - A magic molecule
4. Evolutionary optimization of structure
5. RNA design

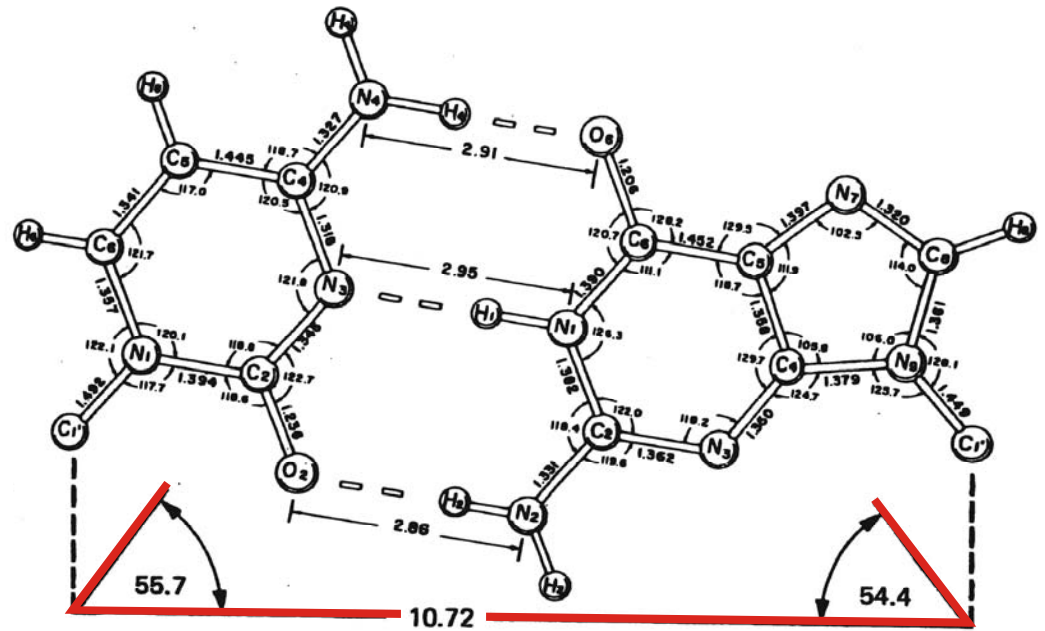


A ≡ Adenine      G ≡ Guanine  
 T ≡ Thymine      C ≡ Cytosine



# DNA



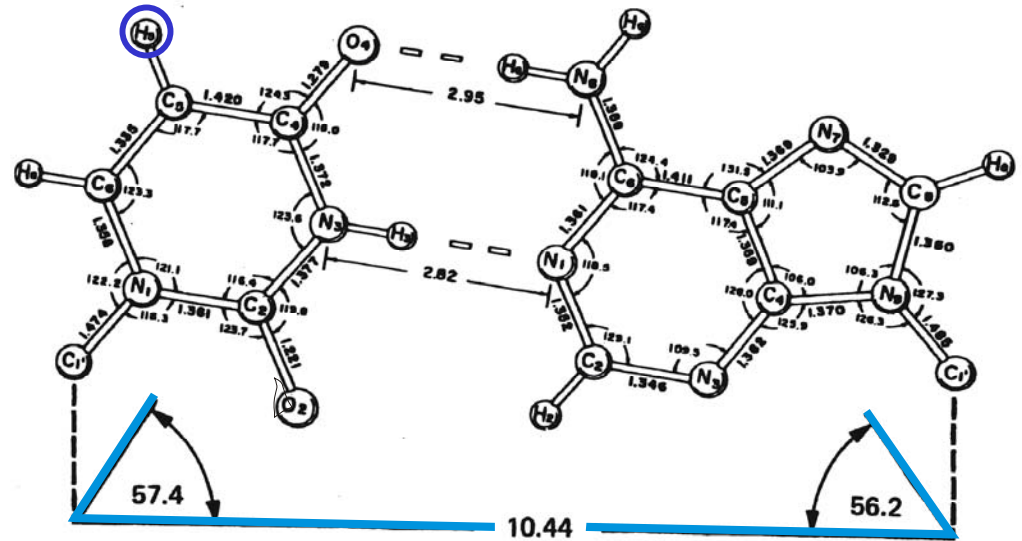


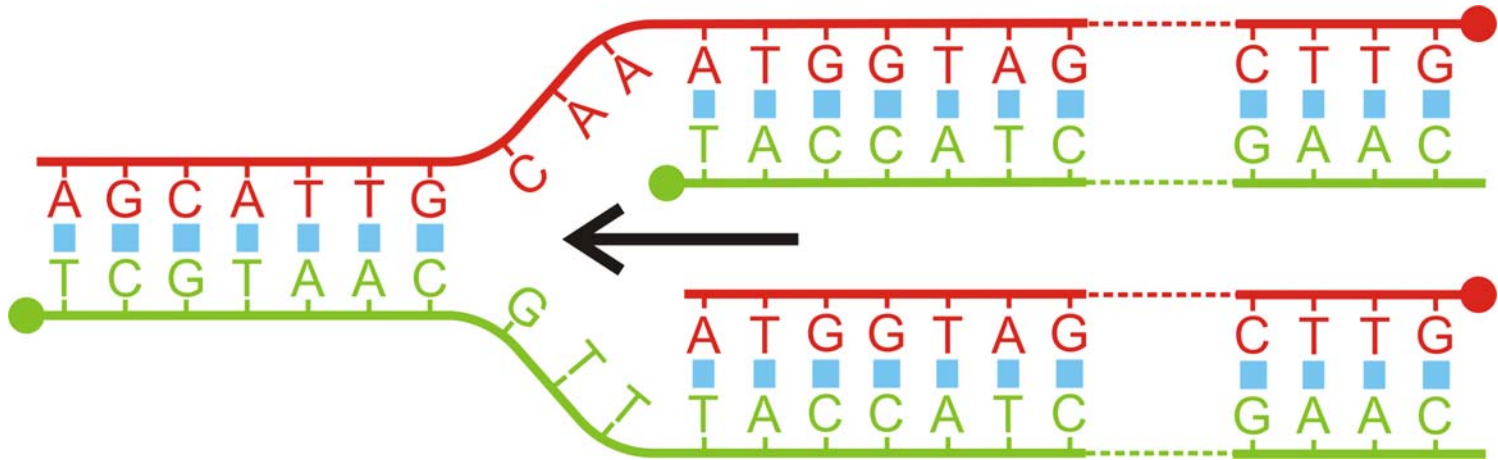
Canonical Watson-Crick  
base pairs:

cytosine – guanine

uracil – adenine (RNA)

thymine – adenine (DNA)

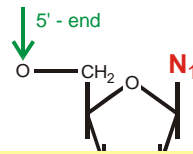




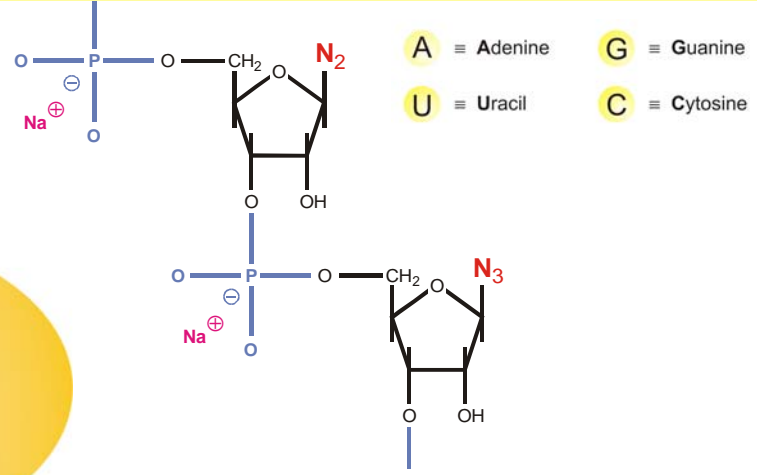
The ,replication fork' in DNA replication

The mechanism of DNA replication is ,semi-conservative'

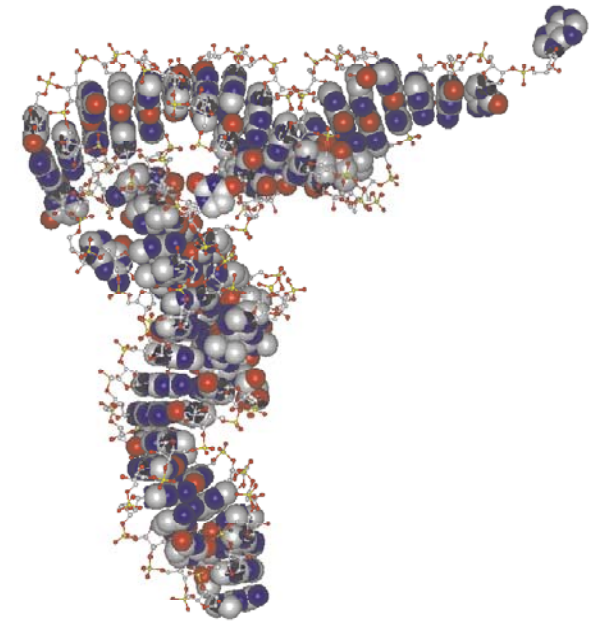
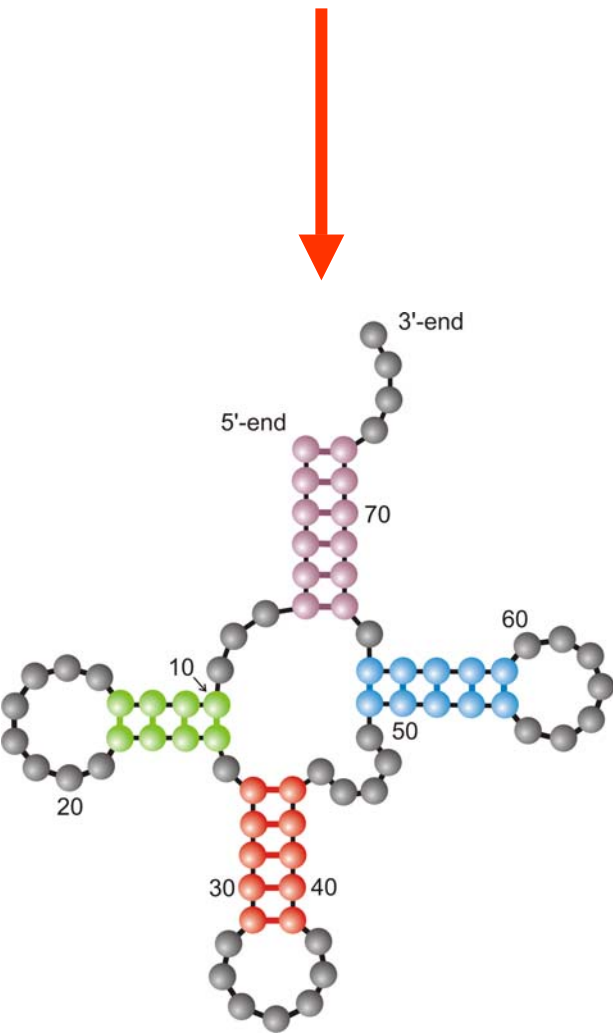




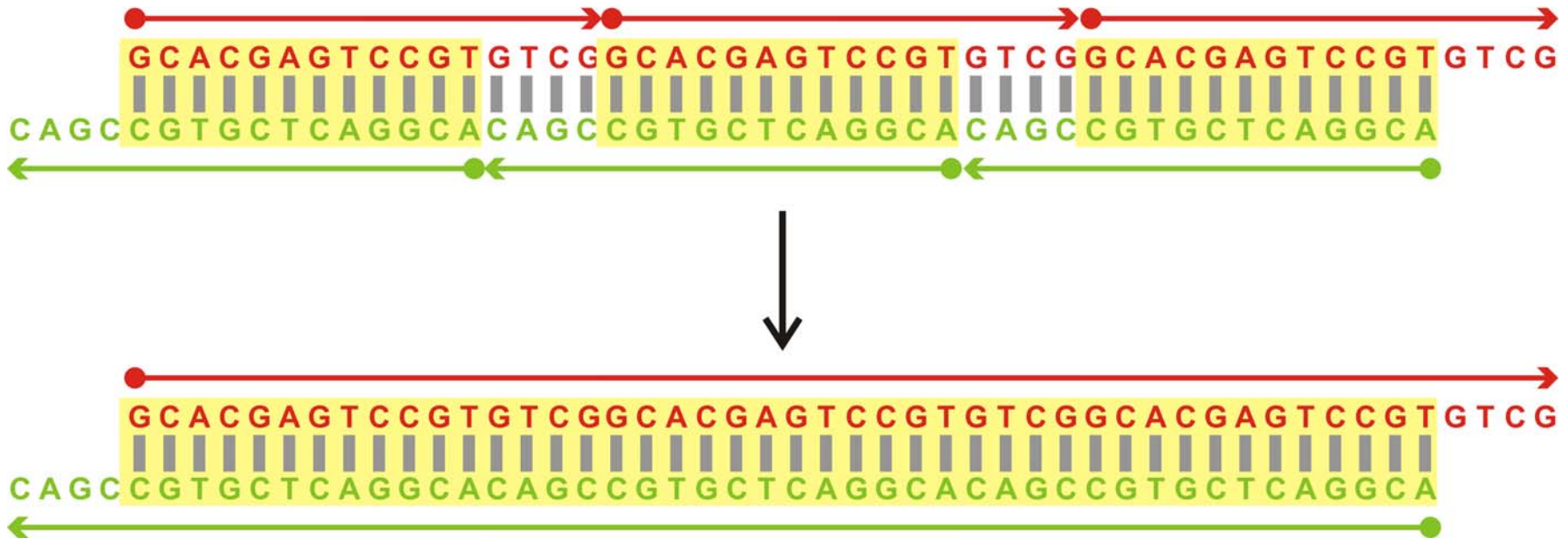
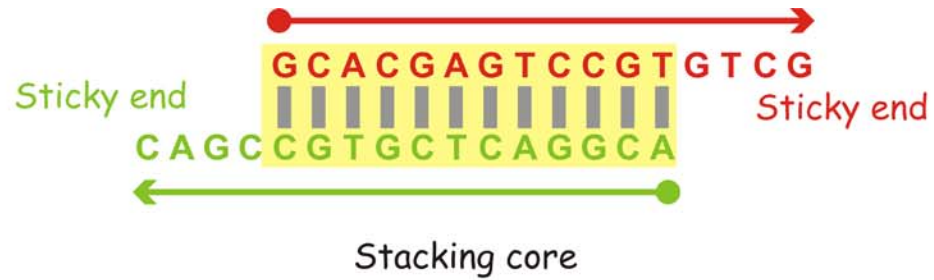
5'-end **GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA** 3'-end



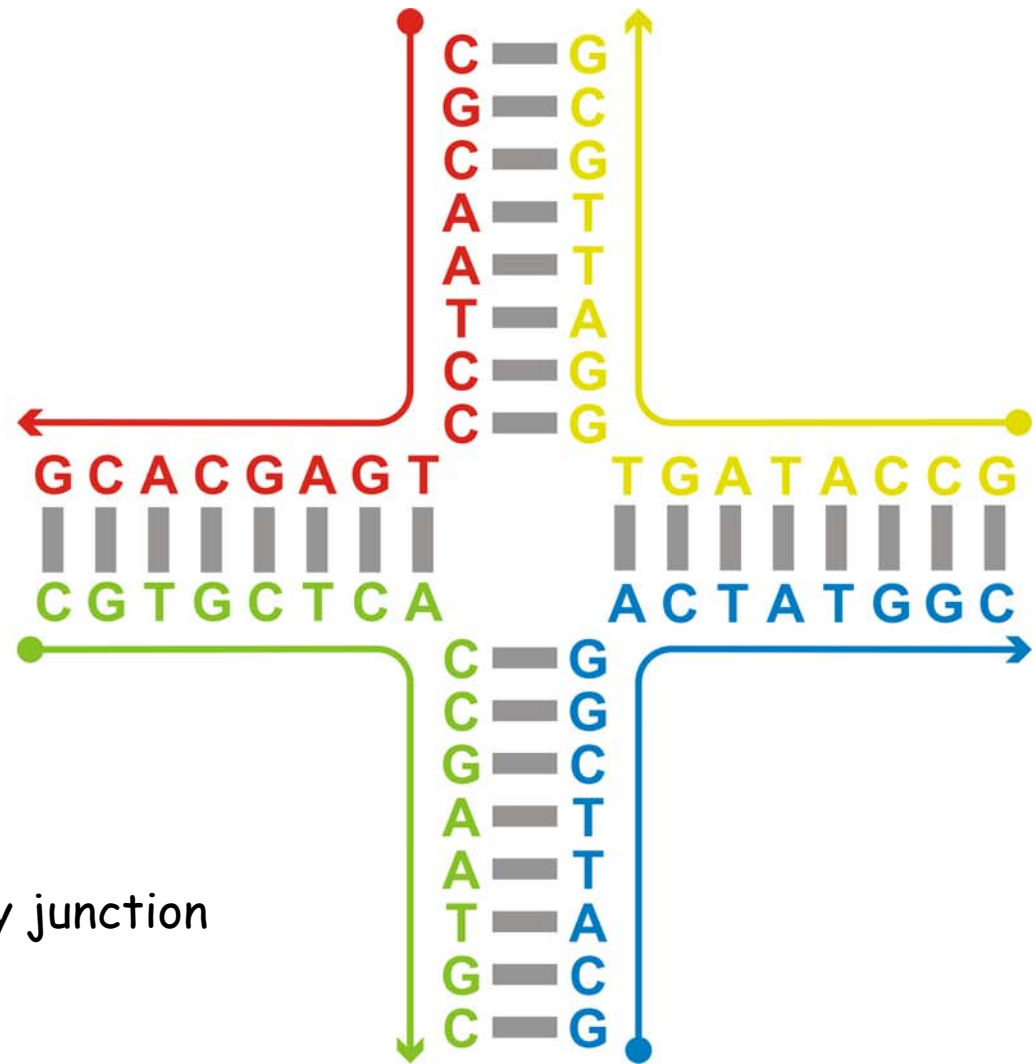
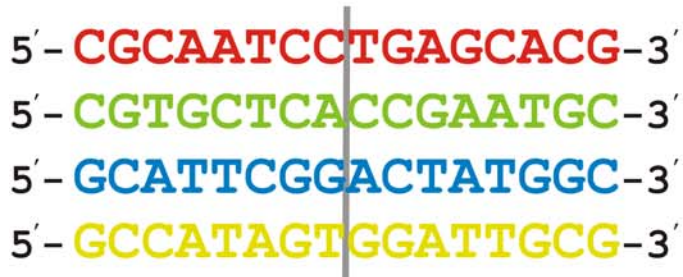
**RNA**



1. Nucleic acid structures
- 2. DNA nanotechnology**
3. RNA - A magic molecule
4. Evolutionary optimization of structure
5. RNA design

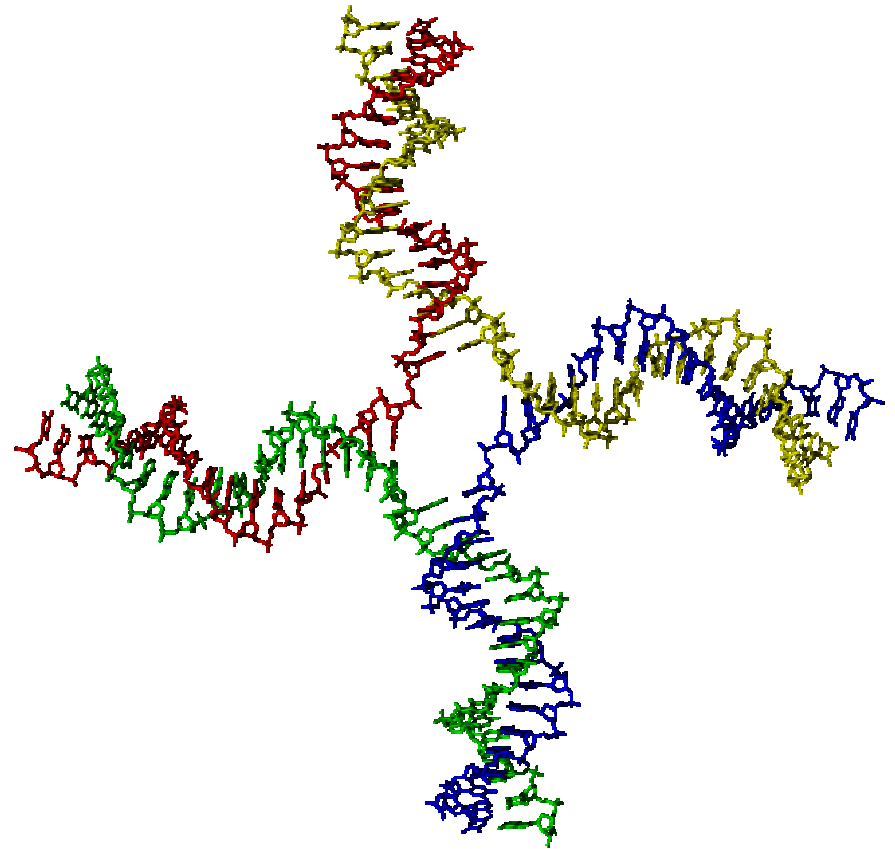
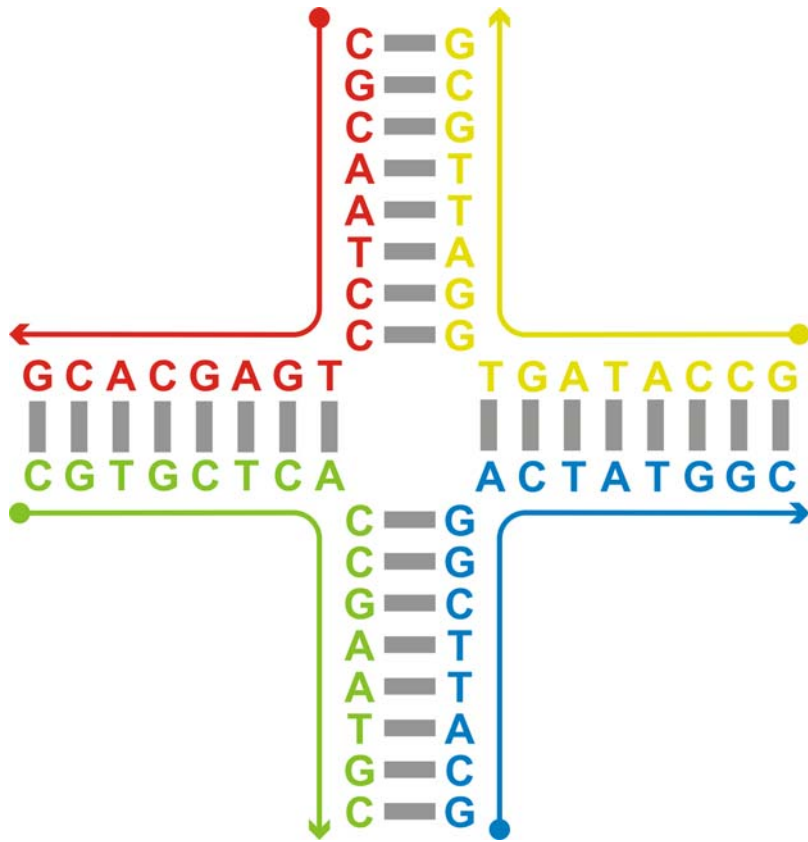


Principle of DNA design shown for DNA-rod formation



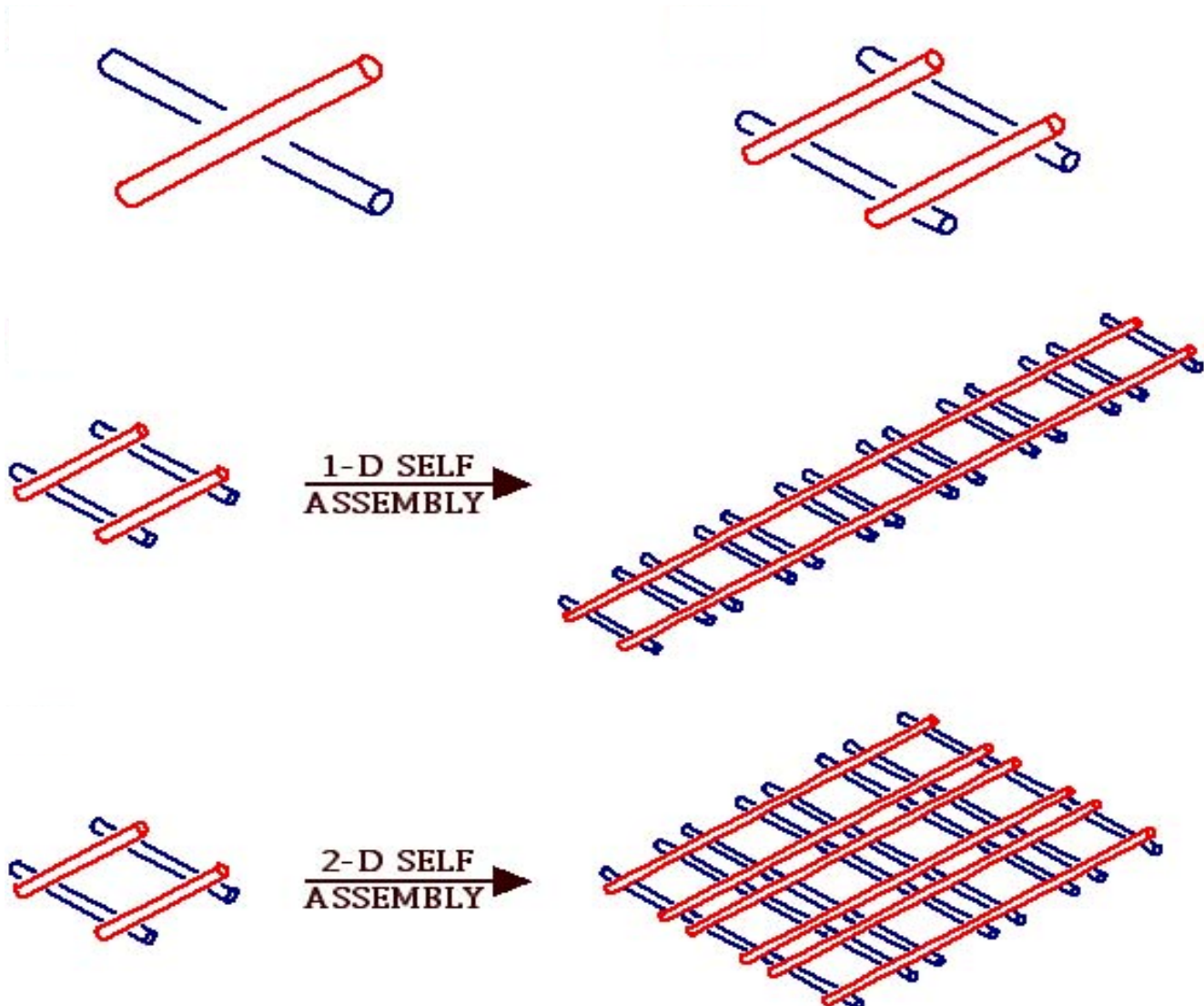
Formation of a stable Holliday junction

N.D. Seeman, P.S. Lukeman. Nucleic acid nanostructure. Bottom-up control of geometry on the nanoscale. *Rep.Prog.Phys.* **68**:237-270, 2005.

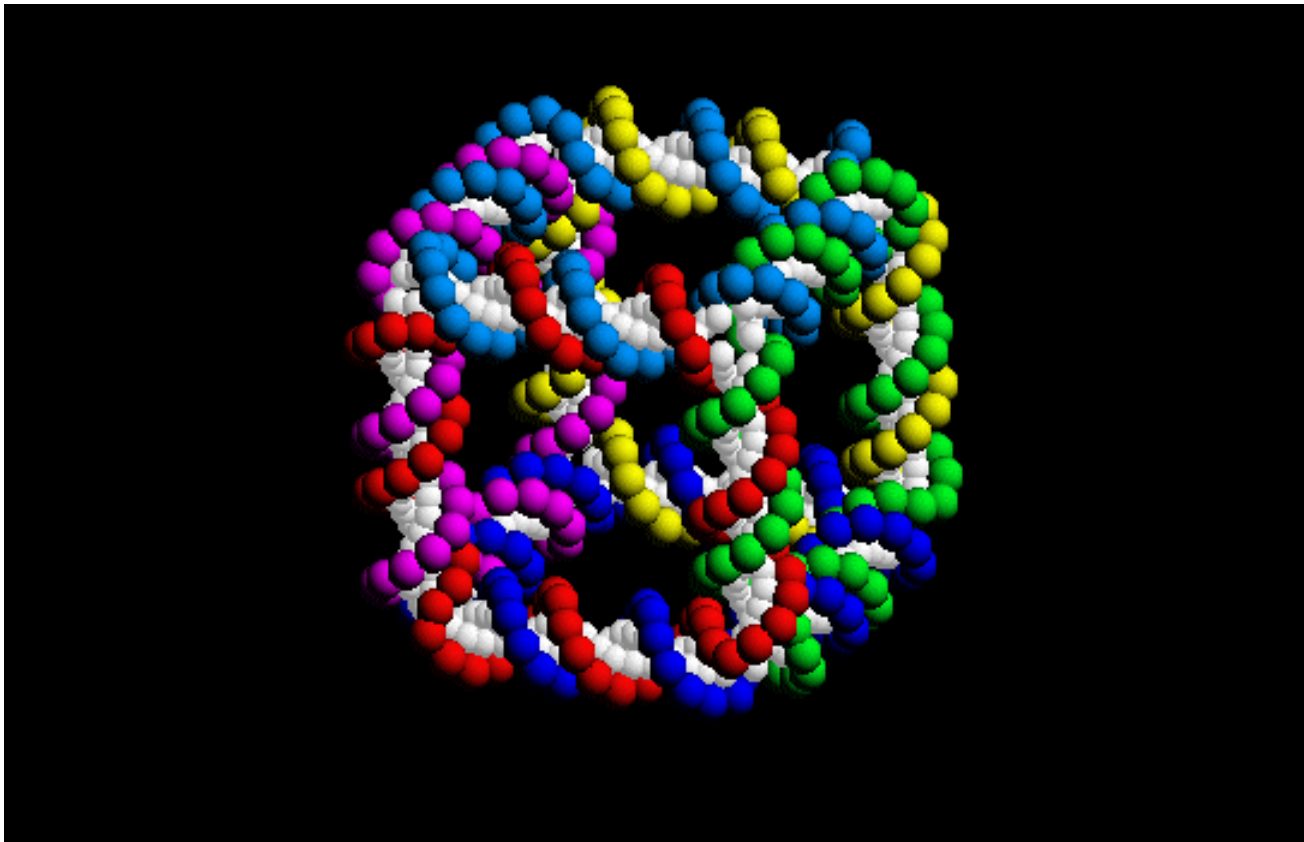


3D structure of a Holliday junction

N.D. Seeman, P.S. Lukeman. Nucleic acid nanostructure. Bottom-up control of geometry on the nanoscale. *Rep.Prog.Phys.* **68**:237-270, 2005.



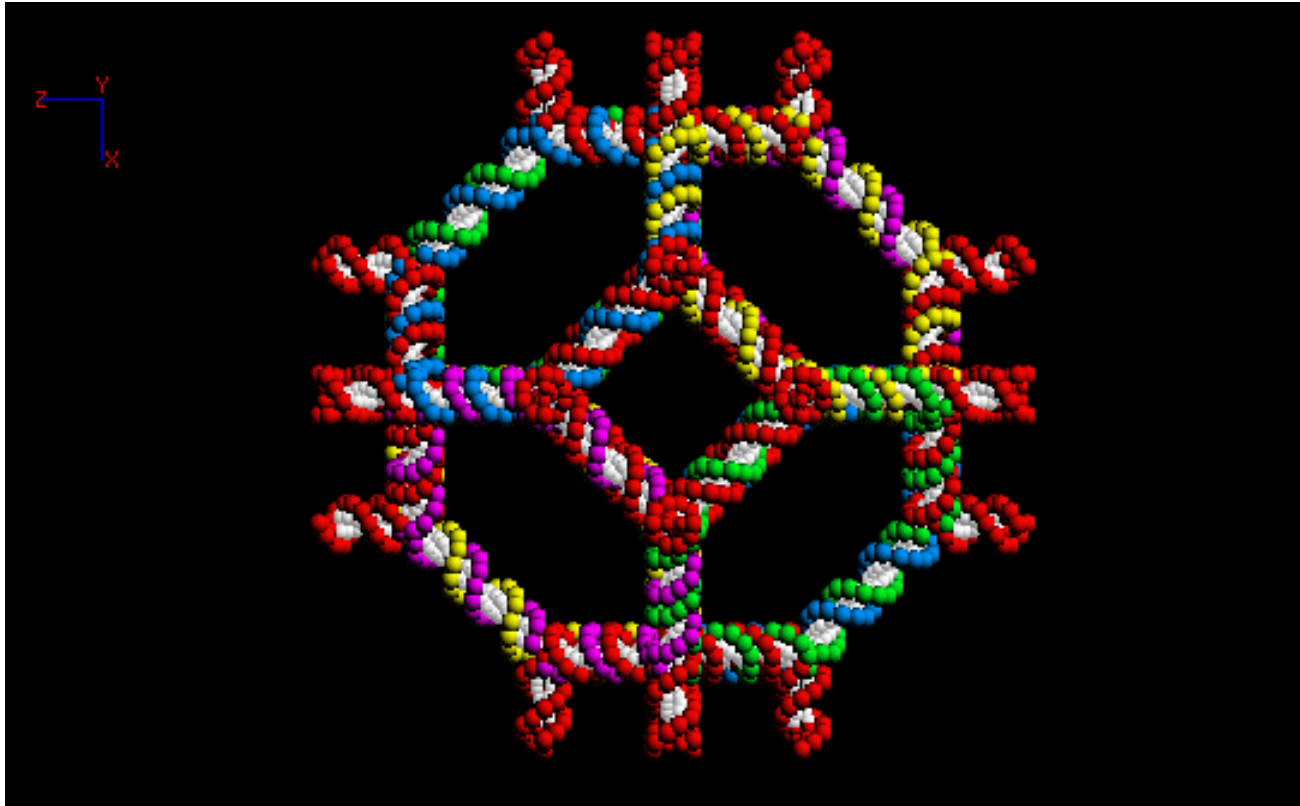
Usage of Holliday junctions to construct DNA lattices



Cube designed from DNA molecules

N.D. Seeman, P.S. Lukeman. Nucleic acid nanostructure. Bottom-up control of geometry on the nanoscale. *Rep.Prog.Phys.* **68**:237-270, 2005.





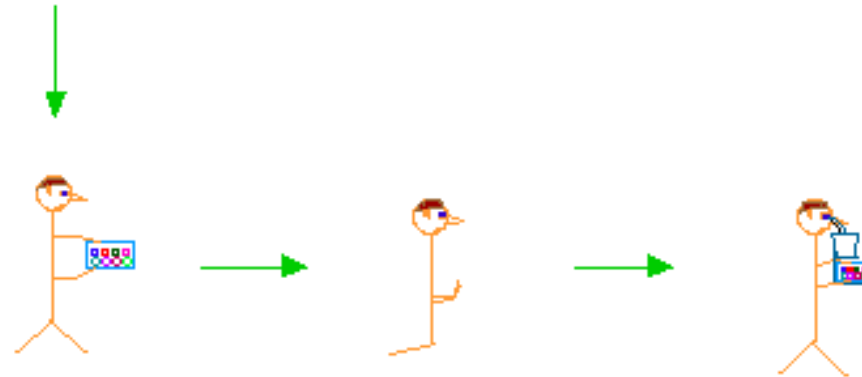
Truncated octahedron designed from DNA molecules

N.D. Seeman, P.S. Lukeman. Nucleic acid nanostructure. Bottom-up control of geometry on the nanoscale. *Rep.Prog.Phys.* **68**:237-270, 2005.



## CURRENT CRYSTALLIZATION PROTOCOL

GUESS CONDITIONS



SET UP CRYSTALS

PRAY FOR CRYSTALS

LOOK FOR CRYSTALS

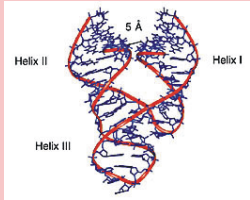
CHANGE DEITIES

GUESS NEW CONDITIONS

N.D. Seeman, P.S. Lukeman. Nucleic acid nanostructure. Bottom-up control of geometry on the nanoscale. *Rep.Prog.Phys.* **68**:237-270, 2005.

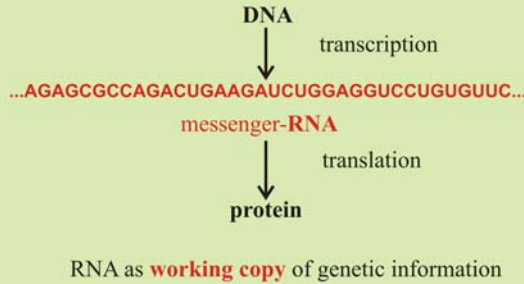
1. Nucleic acid structures
2. DNA nanotechnology
- 3. RNA - A magic molecule**
4. Evolutionary optimization of structure
5. RNA design

### RNA as catalyst

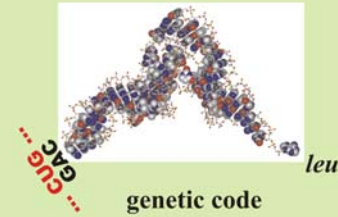


**Ribozyme**

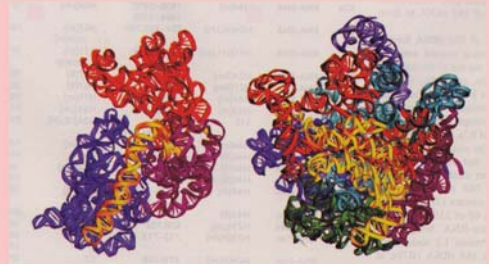
### RNA as transmitter of genetic information



### RNA as adapter molecule



### RNA is the catalytic subunit in supramolecular complexes



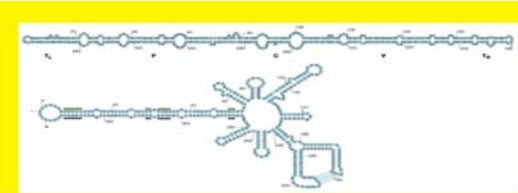
The **ribosome** is a **ribozyme** !

# RNA

*The RNA world as a precursor of the current DNA + protein biology*

RNA is modified by epigenetic control

RNA editing, alternative splicing



**Viroids**

### RNA as carrier of genetic information

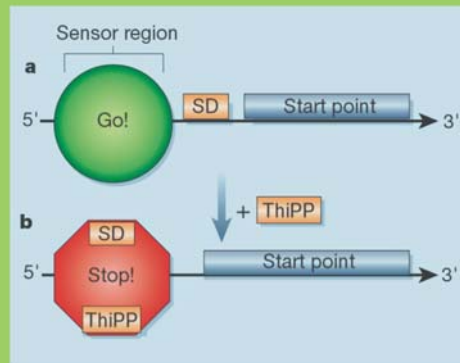
RNA viruses and retroviruses

RNA evolution *in vitro*

Evolutionary biotechnology

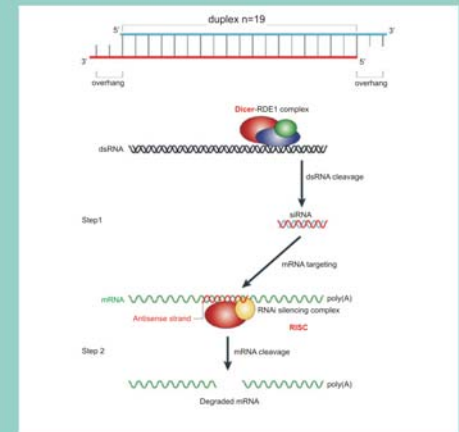
RNA aptamers, artificial ribozymes, allosteric ribozymes

### Allosteric control of transcribed RNA



Riboswitches controlled by metabolites

### RNA as regulator of gene expression



Gene silencing by siRNA

## Functions of RNA molecules

1. Nucleic acid structures
2. DNA nanotechnology
3. RNA - A magic molecule
- 4. Evolutionary optimization of structure**
5. RNA design

## 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

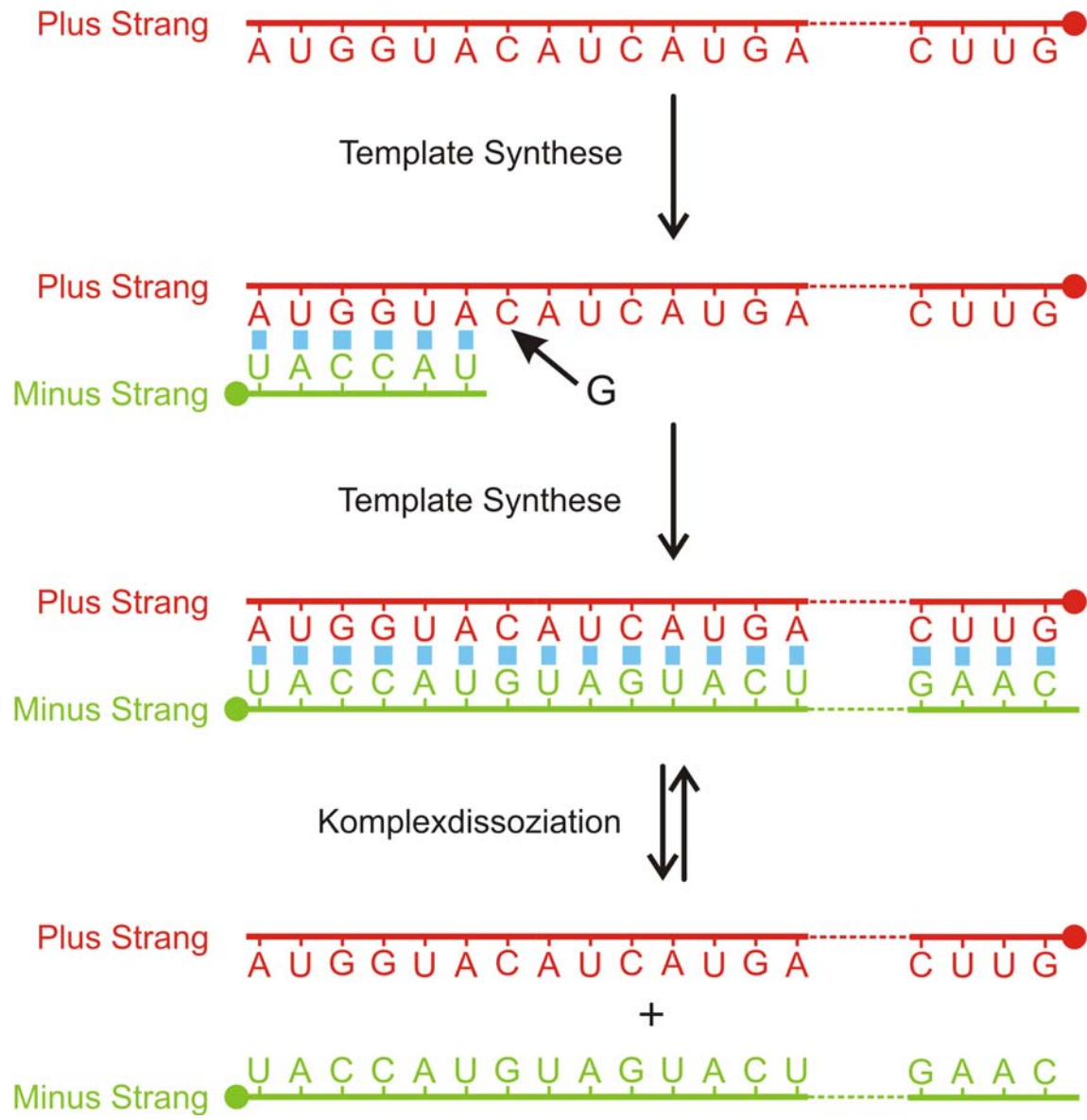
C.K.Biebricher, *Darwinian selection of self-replicating RNA molecules*. Evolutionary Biology **16** (1983), 1-52

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

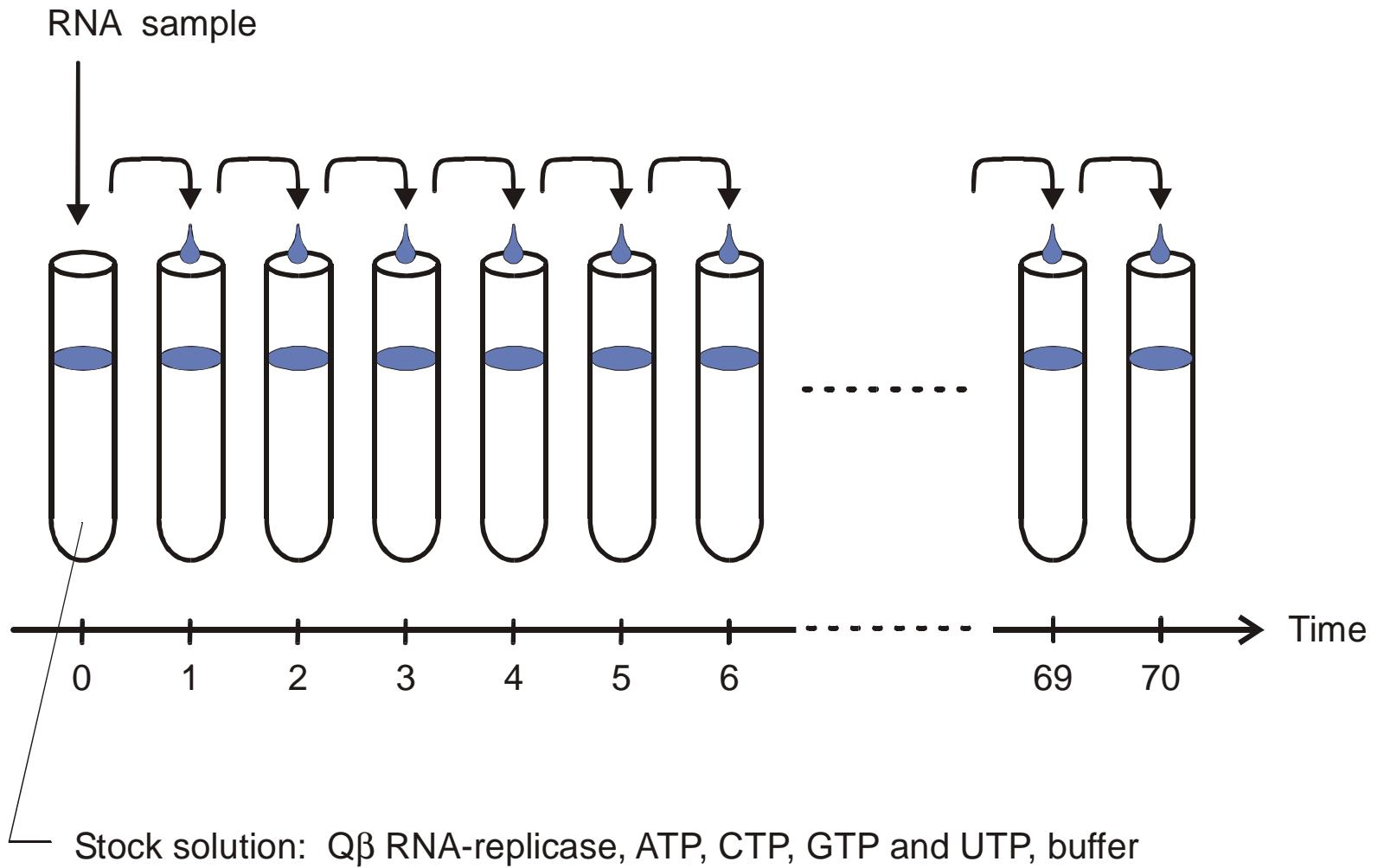
C.K.Biebricher, W.C.Gardiner, *Molecular evolution of RNA in vitro*. Biophysical Chemistry **66** (1997), 179-192

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

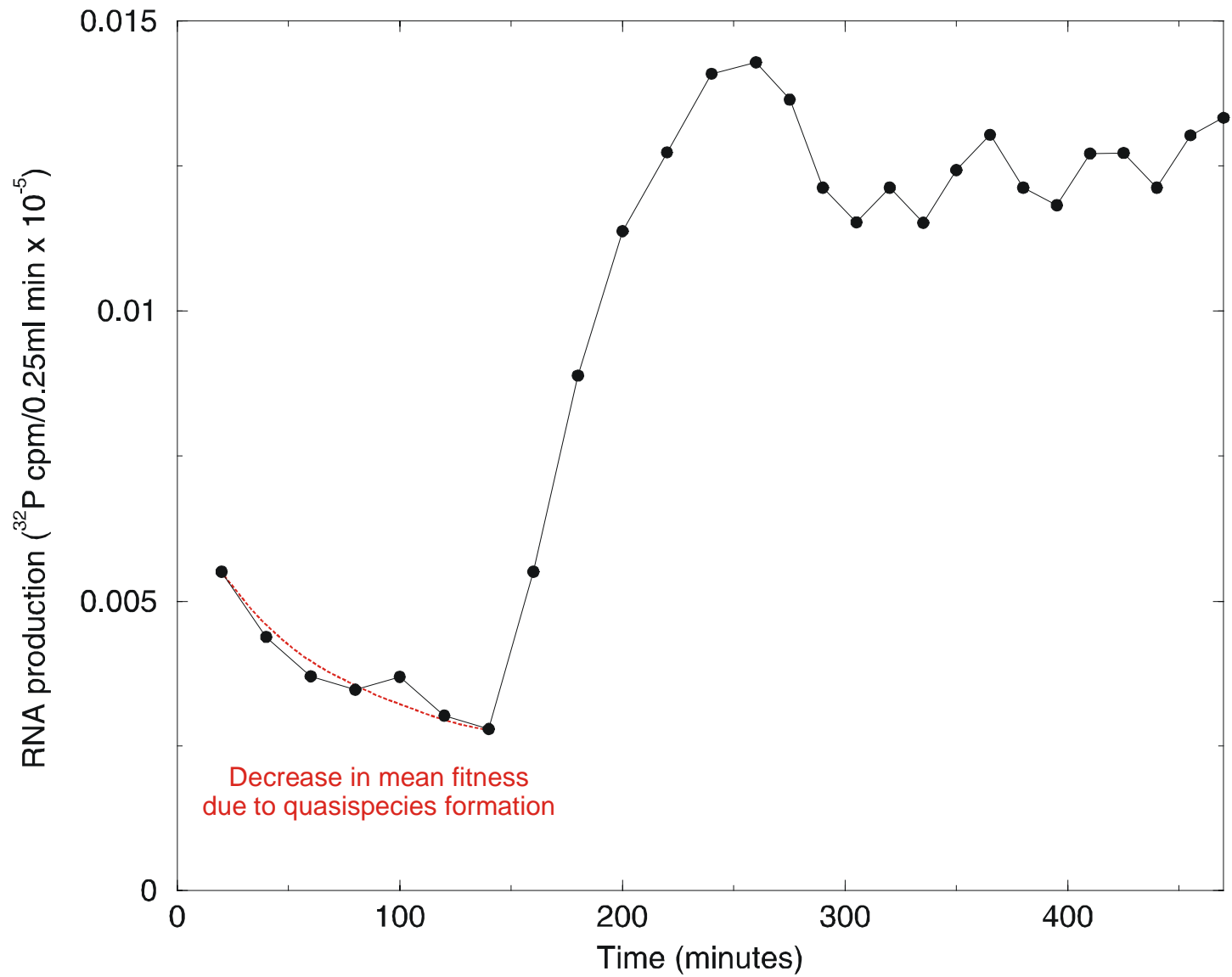
F.Öhlenschläger, M.Eigen, *30 years later – A new approach to Sol Spiegelman's and Leslie Orgel's in vitro evolutionary studies*. Orig.Life Evol.Biosph. **27** (1997), 437-457



The mechanism of single stranded RNA replication

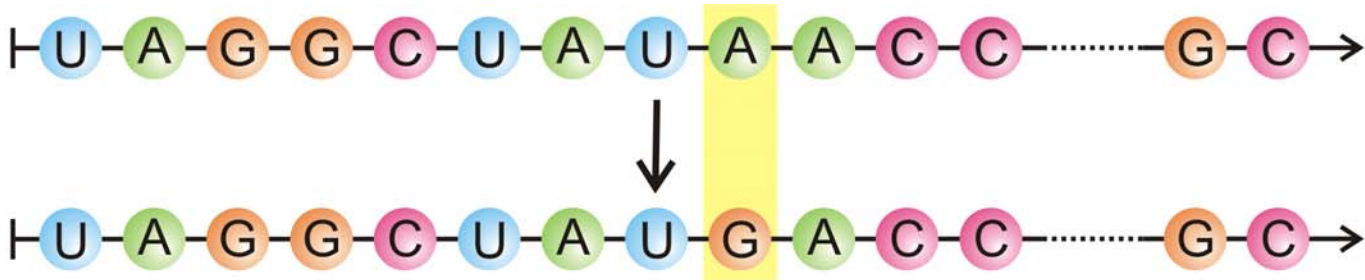


Serial transfer technique for RNA evolution in the test tube

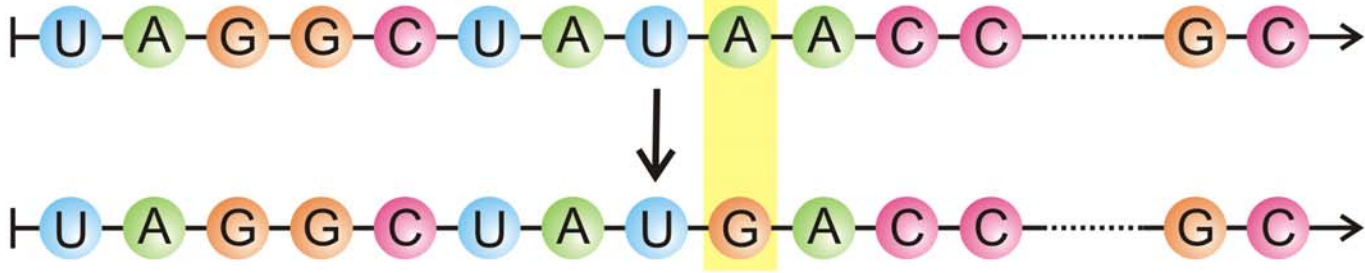


The increase in RNA production rate during a serial transfer experiment

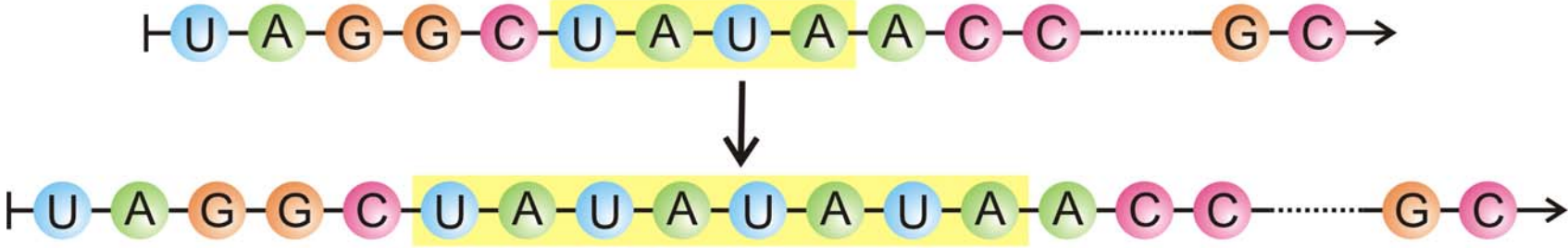




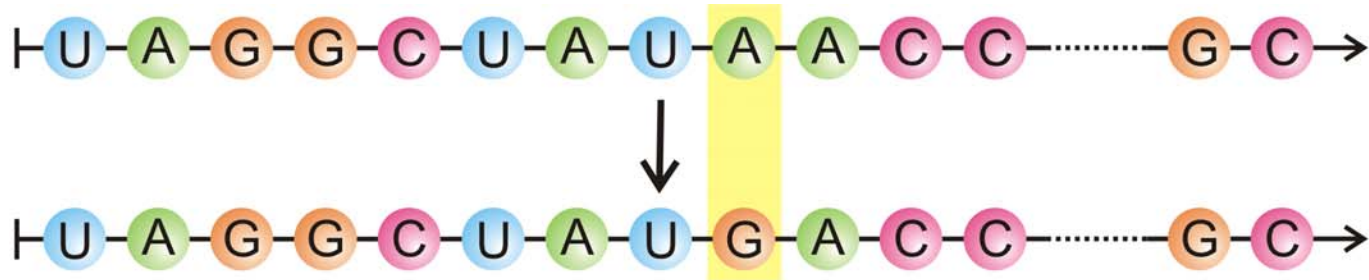
Punktmutation



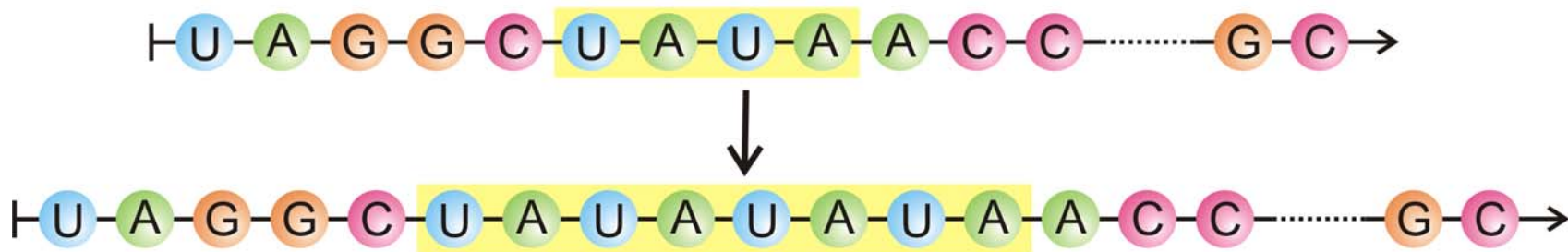
Punktmutation



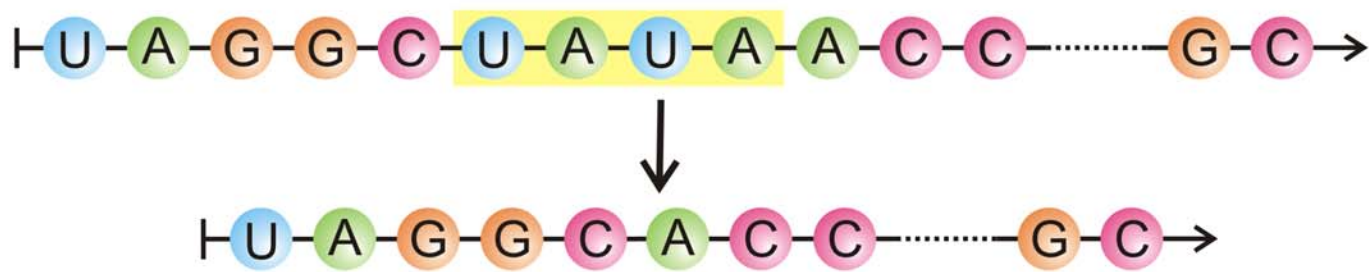
Insertion



Punktmutation



Insertion



Deletion

Selforganization of Matter and the Evolution of Biological Macromolecules

MANFRED EIGEN\*

Max-Planck-Institut für Biophysikalische Chemie, Karl-Friedrich-Bonhoefer-Institut, Göttingen-Nikolausberg

I. Introduction	462	F. Selforganization via Cyclic Catalysis: Proteins	498
I.1. Cause and Effect	465	V.1. Recognition and Catalysis by Enzymes	498
I.2. Penetration of Selforganization	467	V.2. Selforganizing Enzyme Cycles (Theory)	499
I.2.1. Evolution Must Start from Random Events	467	V.2.1. Catalytic Networks	499
I.2.2. Information Requires Information	467	V.2.2. The Selforganizing Loop and Its Variants	499
I.2.3. Information Originates or Gains Value by Selection	469	V.2.3. Cooperation between Different Cycles	501
I.2.4. Selection Occurs under Special Conditions	470	V.3. Can Protein Replicate Theoretically?	501
II. Phenomenological Theory of Selection	473	VI. Solvability by Enzymal Catalytic Functions	503
II.1. The Concept "Information"	473	VI.1. The Requirement of Cooperation between Nucleic Acids and Proteins	503
II.2. Phenomenological Equations	474	VI.2. A Selforganizing Hyper-Cycle	503
II.3. Selection Criteria	476	VI.2.1. The Model	503
II.4. Selection Equilibrium	479	VI.2.2. Theoretical Treatment	505
II.5. Quality Factor and Error Distribution	480	VI.3. On the Origin of the Code	508
II.6. Kinetics of Selection	481	VII. Radiation Experiments	511
III. Stochastic Approach to Selection	484	VII.1. The Q <sub>10</sub> -Replicase System	511
III.1. Limitations of a Deterministic Theory of Selection	484	VII.2. Darwinian Evolution in the Test Tube	512
III.2. Fluctuations around Equilibrium States	484	VII.3. Quantitative Selection Studies	513
III.3. Fluctuations in the Steady State	485	VII.4. "Mines One" Experiments	514
III.4. Stochastic Models in Markov Chains	487	VIII. Conclusion	515
III.5. Quantitative Discussion of Three Prototypes of Selection	487	VIII.1. Limits of Theory	515
IV. Selforganization Based on Complementary Interactions: Nucleic Acids	490	VIII.2. "Diagnosis" and the "Origin of Information"	516
IV.1. True "Selfinformation"	490	VIII.3. The Principles of Selection and Evolution	517
IV.2. Complementary Interaction and Selection (Theory)	492	VIII.4. "Indeterminate" but "Inevitable"	518
IV.3. Complementary Base Recognition (Experimental Data)	494	VIII.5. "Indeterminate" but "Inevitable"	520
IV.3.1. Single Pair Formation	494	VIII.6. Can the Phenomenon of Life be Explained by Our Present Concepts of Physics?	520
IV.3.2. Cooperative Interactions in Oligo- and Polynucleotides	495	IX. Deutsche Zusammenfassung	520
IV.3.3. Conclusions about Recognition	496	Acknowledgements	522
		Literature	522

I. Introduction

I.1. „Cause and Effect“

The question about the origin of life often appears as a question about "cause and effect". Physical theories of macroscopic processes usually involve answers to such questions, even if a statistical interpretation is given to the relation between "cause" and "effect". It is mainly due to the nature of this question that many scientists believe that our present physics does not offer any obvious explanation for the existence of life.

\* Partly presented at the "Robbins Lectures" at Pomona College, California, in spring 1970.

The Hypercycle

A Principle of Natural Self-Organization

Part A: Emergence of the Hypercycle

Manfred Eigen

Max-Planck-Institut für biophysikalische Chemie, D-3400 Göttingen

Peter Schuster

Institut für theoretische Chemie und Strahlenchemie der Universität, A-1090 Wien

This paper is the first part of a trilogy, which comprises a detailed study of a special type of functional organization and demonstrates its relevance with respect to the origin and evolution of life. Self-replicating macromolecules, such as RNA or DNA in a suitable environment exhibit a behavior, which we may call Darwinian and which can be formally represented by the concept of the quasi-species. A quasi-species is defined as a given distribution of macromolecular species with closely interrelated sequences, dominated by one or several (hypothesized) master copies. External conditions enforce the selection of the best adapted distribution, autocatalytically referred to as the wild-type. Most important for Darwinian behavior are the criteria for internal stability of the quasi-species. If these criteria are violated, the information stored in the nucleotide sequence of the master copy will disseminate irreversibly leading to an error catastrophe. As a consequence, selection and evolution of RNA or DNA molecules is limited with respect to the amount of information that can be stored in a single replicative unit. An analysis of experimental data regarding RNA and DNA replication at various levels of organization reveals, that a sufficient amount of information for the build up of a translation machinery can be gained only via integration of several different replicative units (or reproductive cycles) through reciprocal linkages. A stable functional organization then will arise if the system to a low level of organization and thereby enlarge its information capacity continuously. The Hypercycle appears to be such a form of organization.

Preview on Part B: The Abstract Hypercycle

The mathematical analysis of dynamical systems using methods of differential topology yields the result that there is only one type of mechanism which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective master copies must compete favorably with their error distributions. Despite their competitive behavior these units must establish a cooperation which includes all functionally integrated species. On the other hand, the cycle as a whole must continue to compete strongly with any other single entity or isolated ensemble which does not contribute to its sustained function. These requirements are crucial for a selection of the best adapted functionally linked ensemble and its evolutive optimization. Only

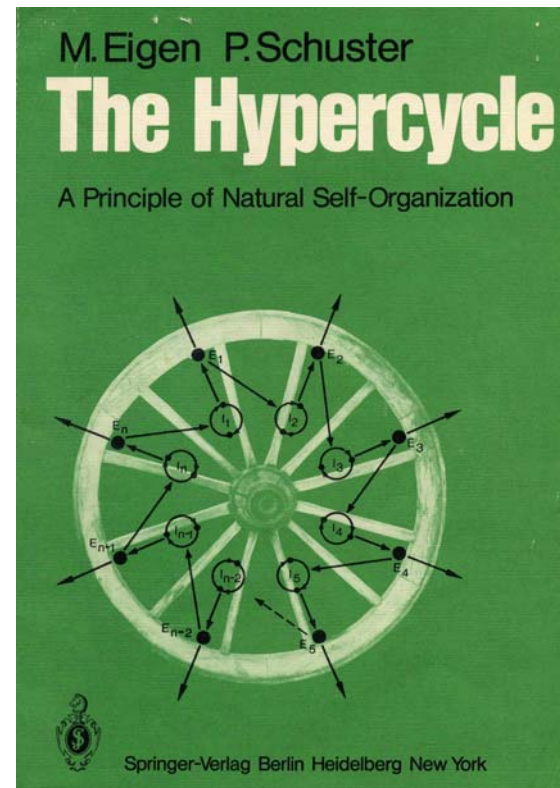
hypercyclic organizations are able to fulfil these requirements. Non-cyclic linkages among the autonomous reproduction cycles, such as chain or branched, one-like networks are devoid of such properties. The mathematical methods used for proving these assertions are fixed-point, Lyapunov and trajectory analysis in high-dimensional phase space, spanned by the concentration coordinates of the cooperating partners. The self-organizing properties of hypercycles are elucidated, using analytical as well as numerical techniques.

Preview on Part C: The Abstract Hypercycle

A realistic model of a hypercycle relevant with respect to the origin of the genetic code and the translation machinery is presented. It includes the following features referring to natural systems: 1) The hypercycle has a sufficiently simple structure to admit an organization with finite probability under prebiotic conditions. 2) It permits a continuous emergence from closely interrelated (in RNA-3-sets) precursors, originally being members of a stable RNA quasi-species and having been amplified to a level of higher abundance. 3) The organizational structure and the properties of single functional units of this hypercycle are well reflected in the present genetic code in the translation apparatus of the prokaryotic cell, as well as in certain bacterial viruses.

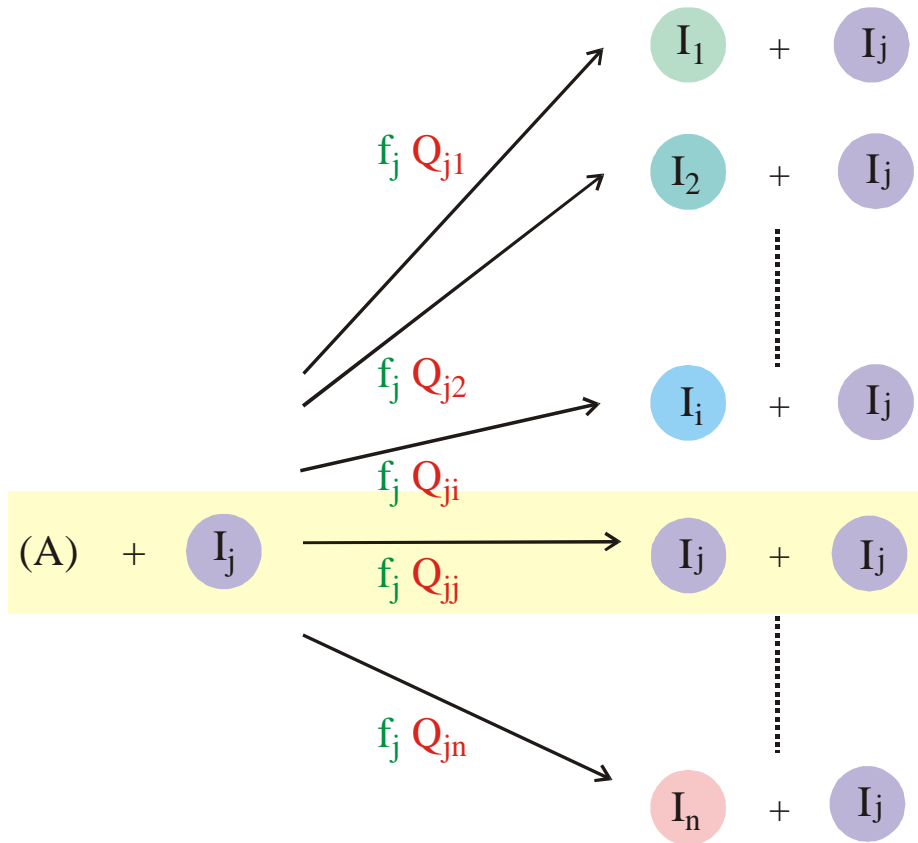
I. The Paradigm of Unity and Diversity in Evolution

Why do millions of species, plants and animals, exist, while there is only one basic molecular machinery of the cell: one universal genetic code and unique chiralities of the macromolecules? The geneticists of our day would not hesitate to give an immediate answer to the first part of this question. Diversity of species is the outcome of the tremendous branching process of evolution with its myriads of single steps of reproduction and mutation. It in-



Chemical kinetics of molecular evolution

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



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

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

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

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

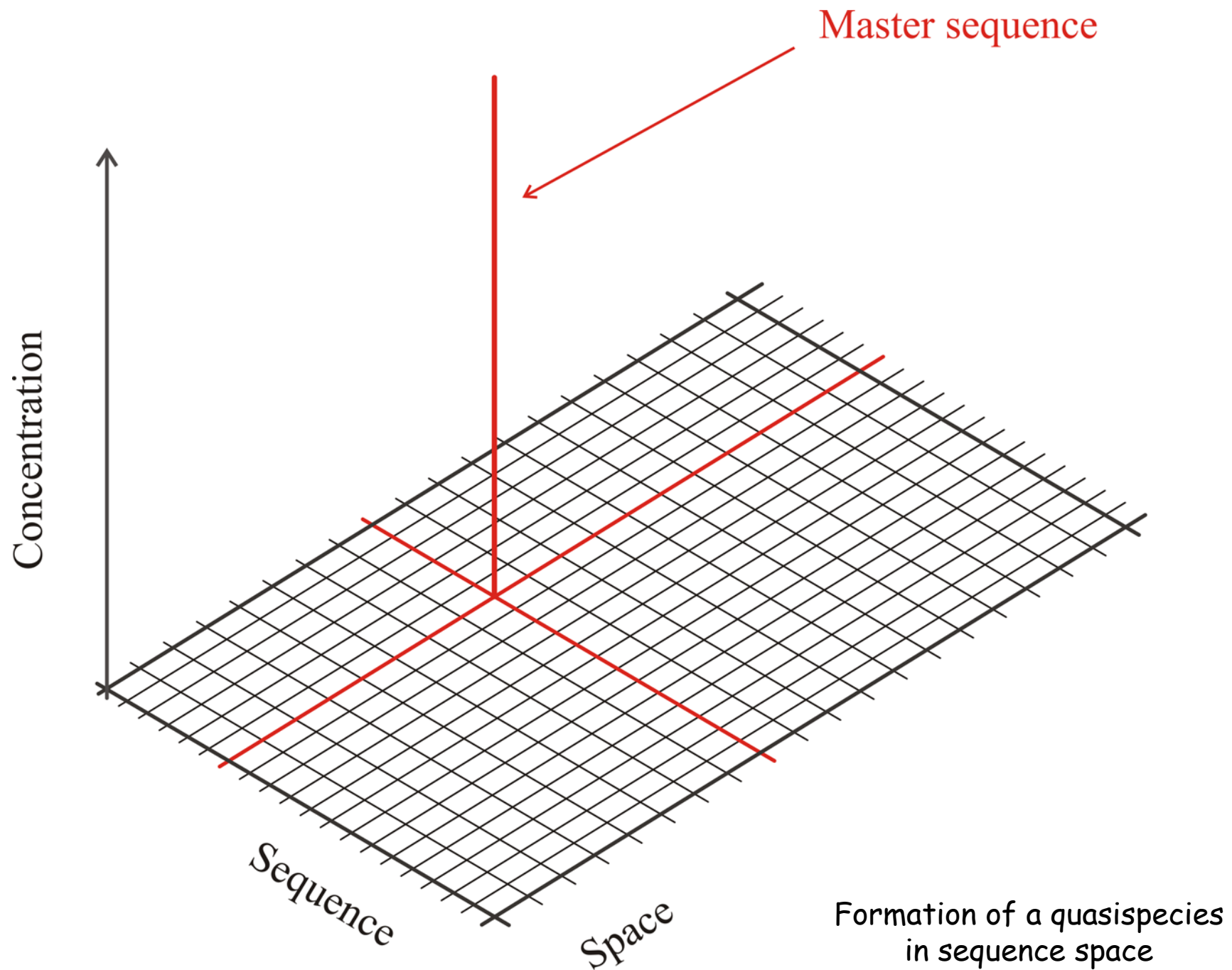
$$Q_{ij} = (1-p)^{\ell-d(i,j)} p^{d(i,j)}$$

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

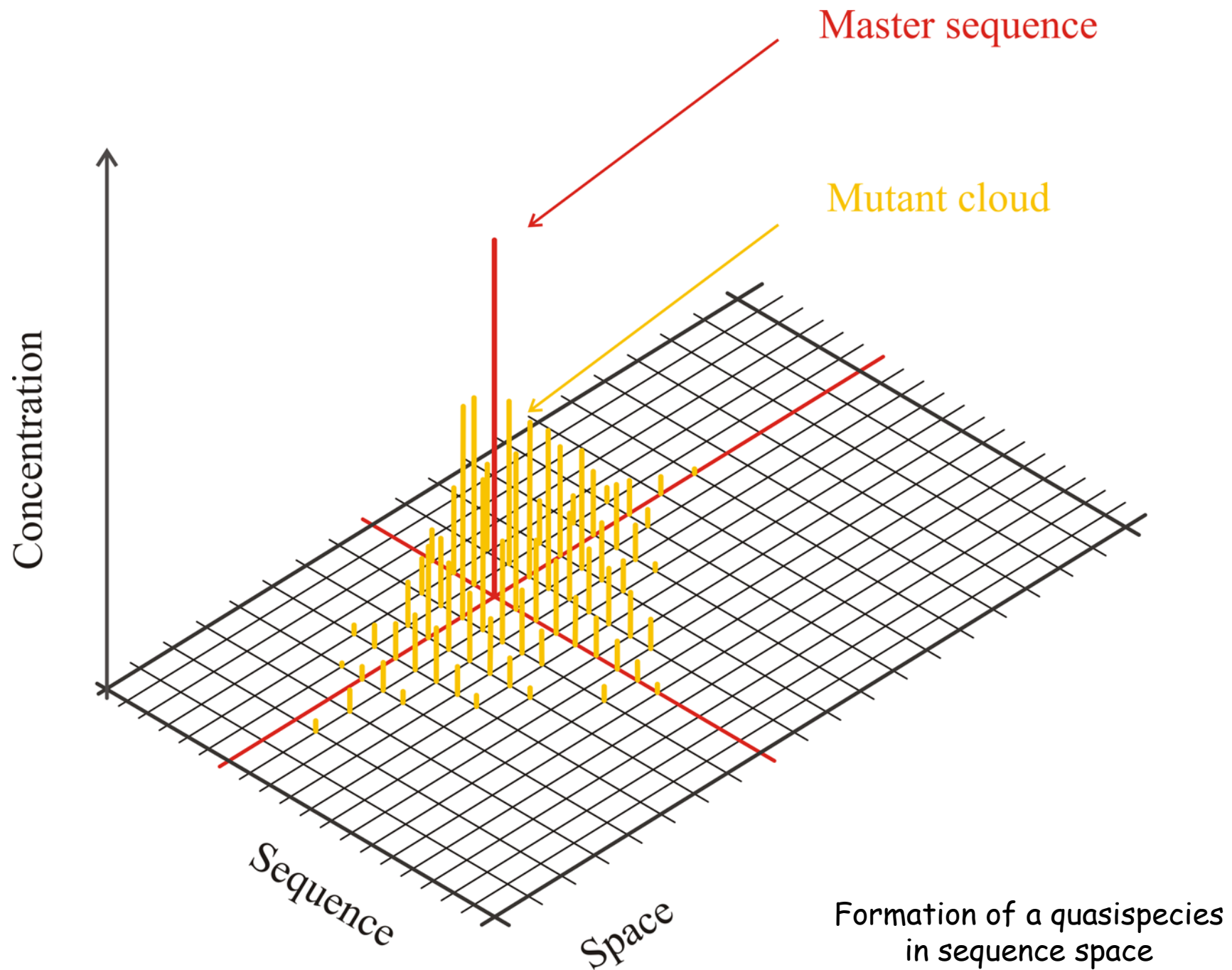
$\ell$  ..... Chain length of the polynucleotide

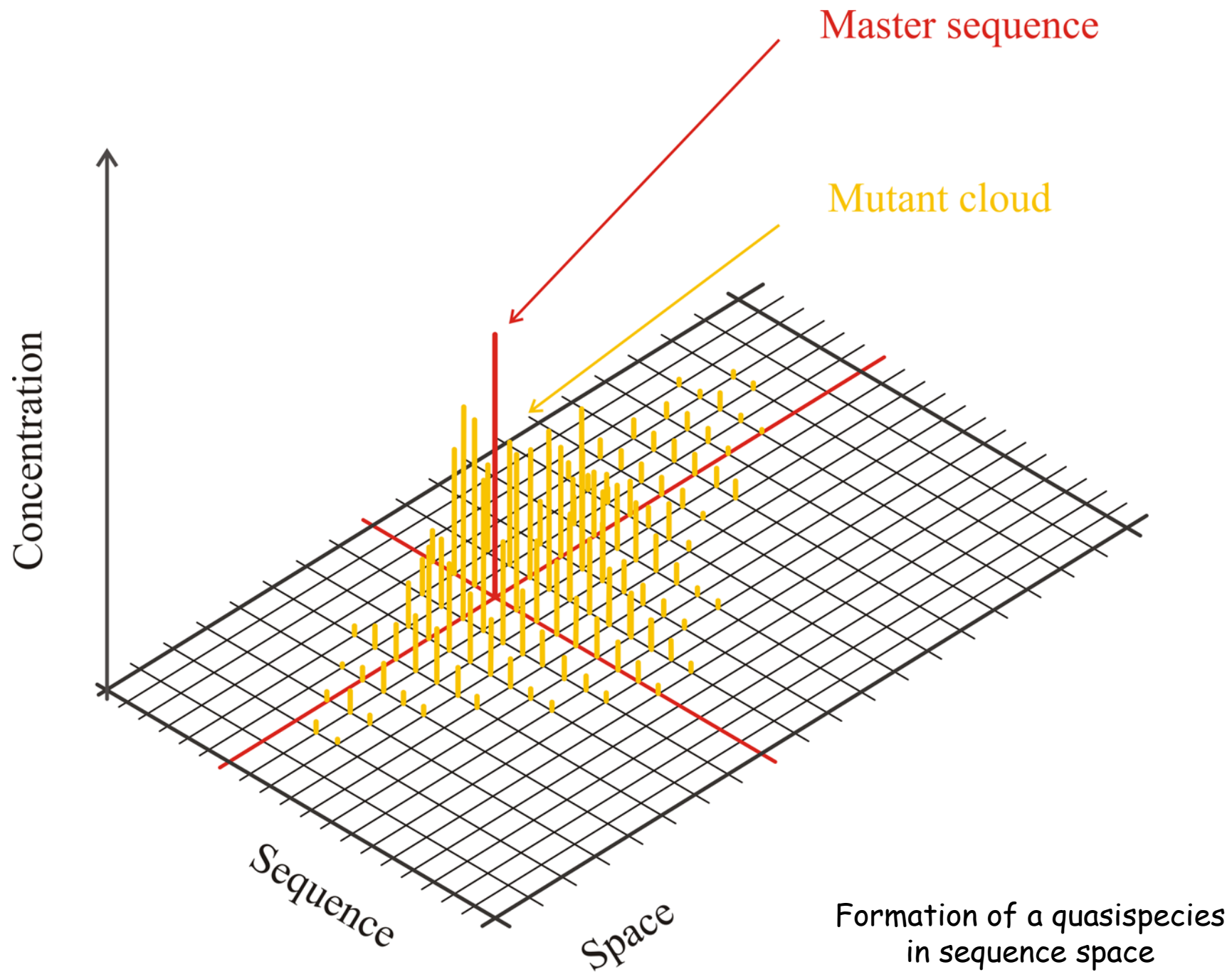
$d(i,j)$  .... Hamming distance between  $I_i$  and  $I_j$

Chemical kinetics of replication and mutation as parallel reactions

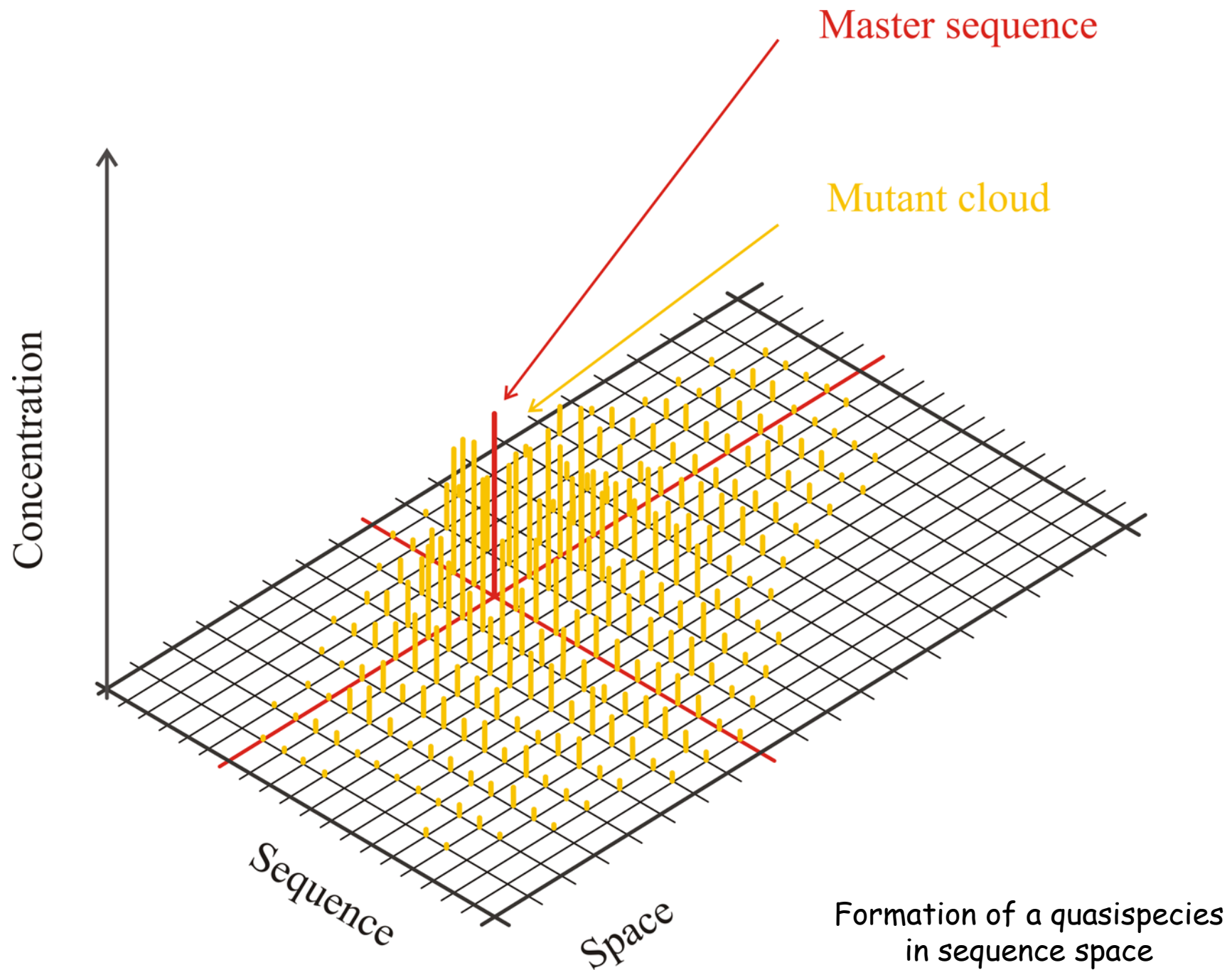


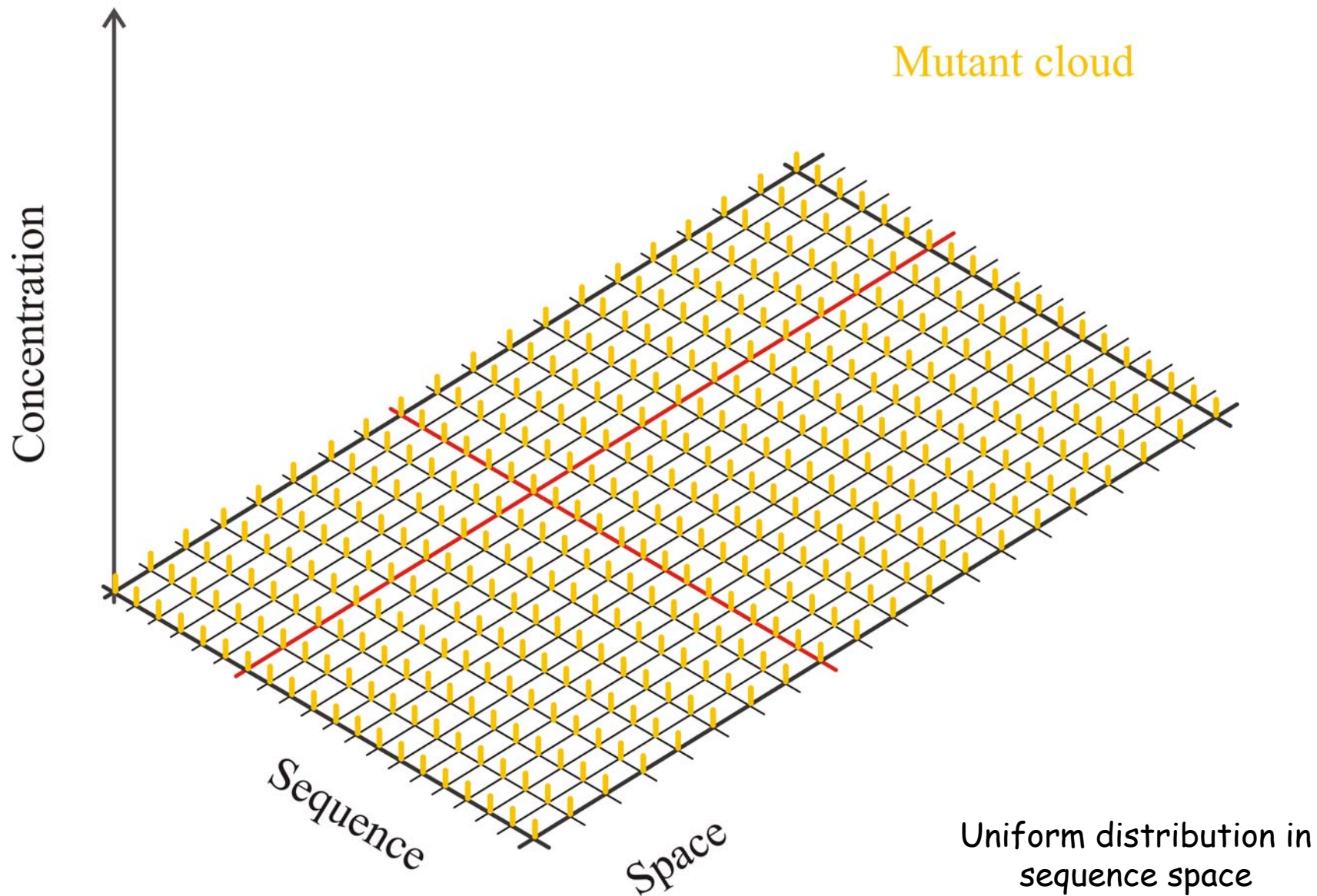


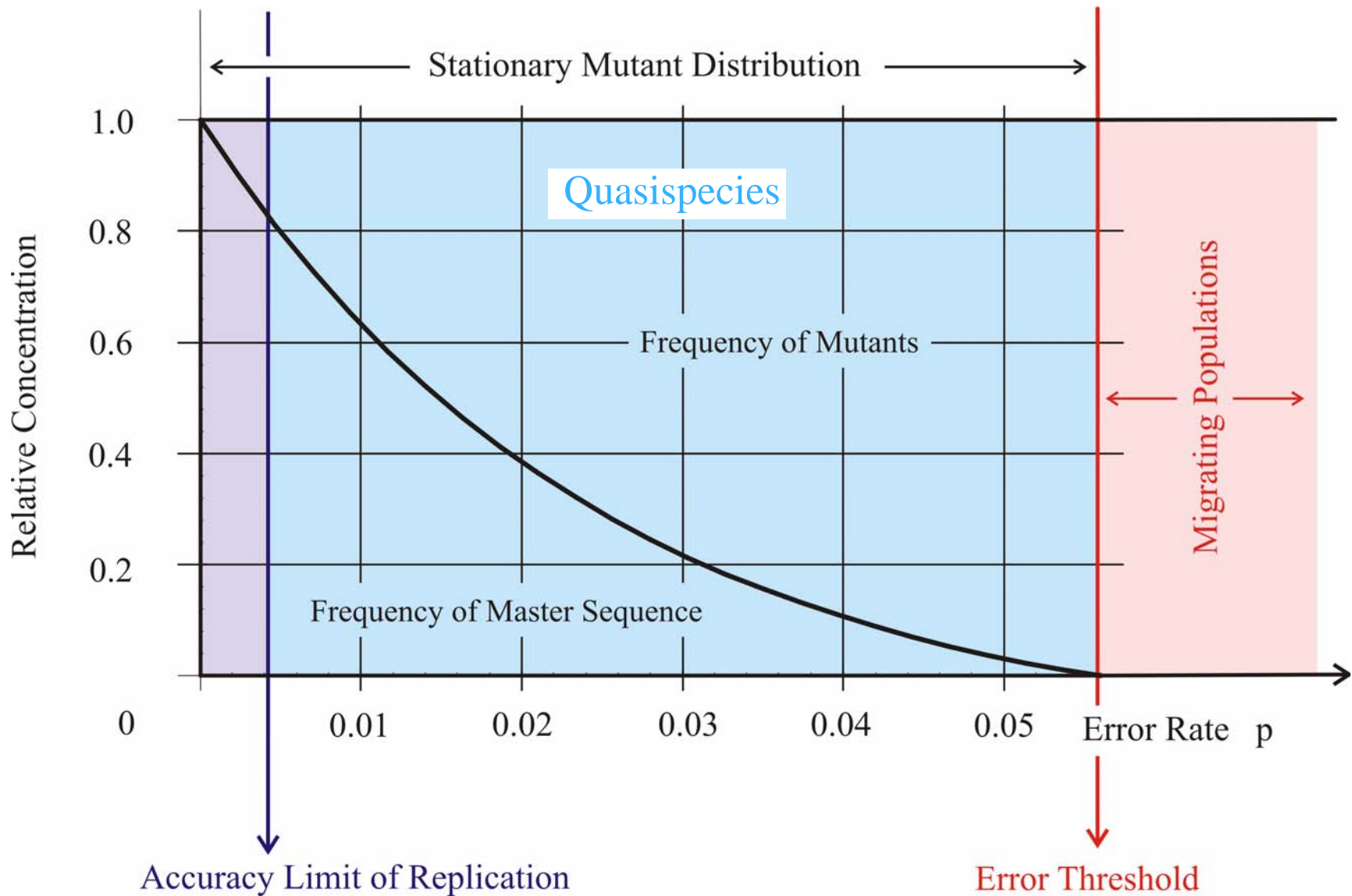












The error threshold in replication

random individuals. The primer pair used for genomic DNA amplification is 5'-TCTCCCTGGATTCT-CATTTA-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'-CTCAAGGCTTCTGGCATGGT-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, A. Gupta, E. Sorbello, R. Torzkadsh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Sternberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and 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.

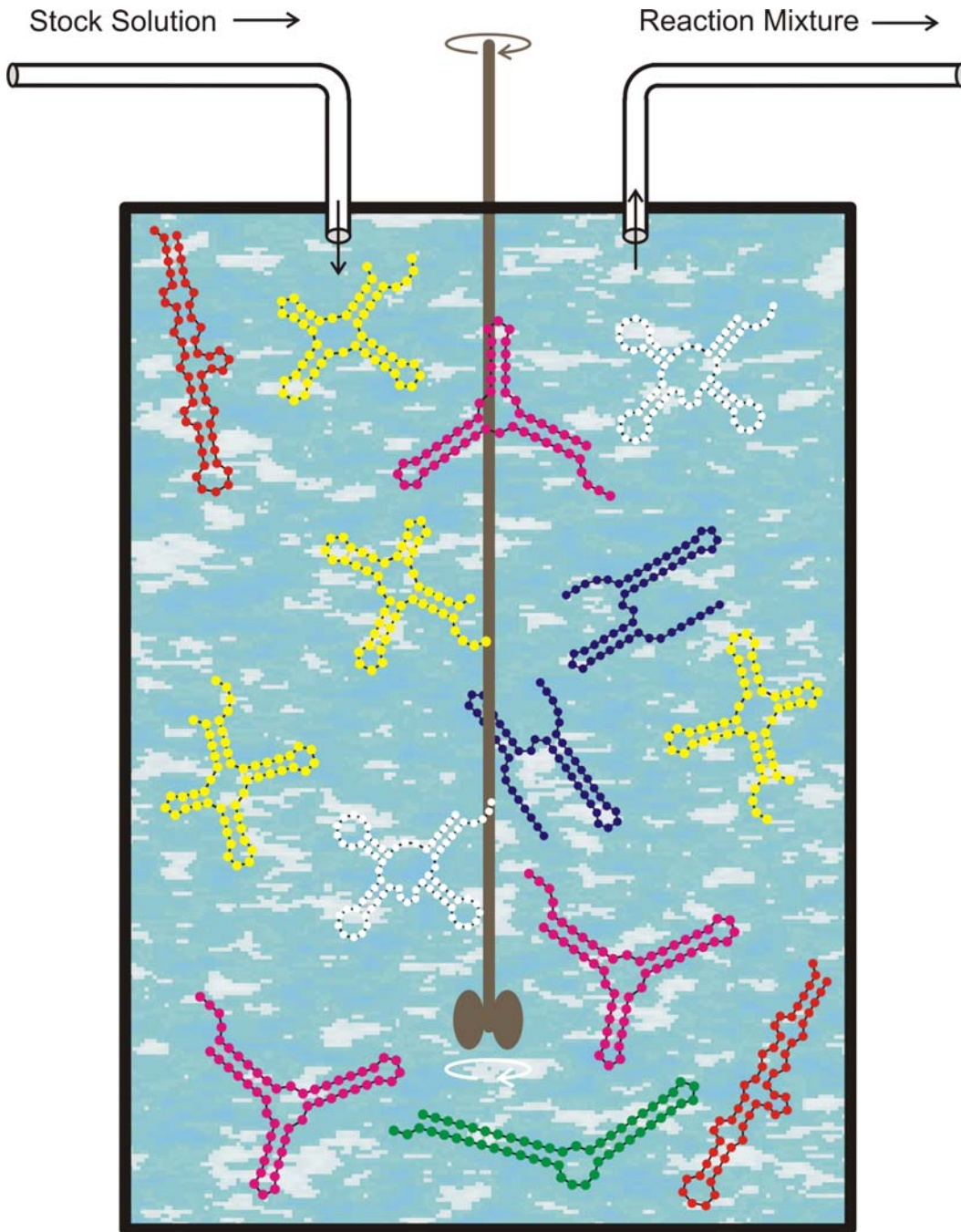
A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally 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.





**Replication rate constant:**

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

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

**Selection constraint:**

Population size,  $N = \#$  RNA molecules, is controlled by the flow

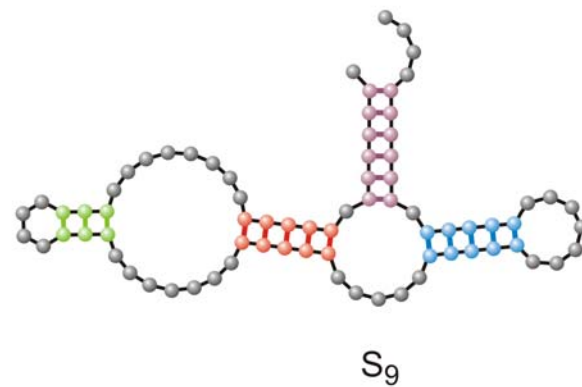
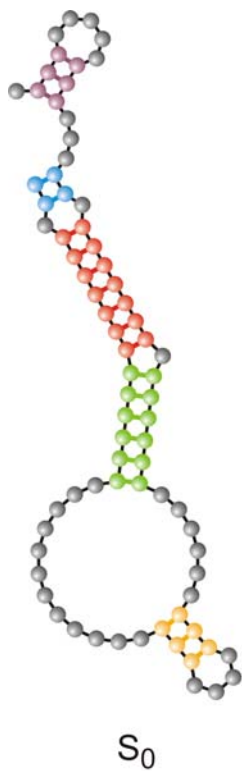
$$N(t) \approx \bar{N} \pm \sqrt{\bar{N}}$$

**Mutation rate:**

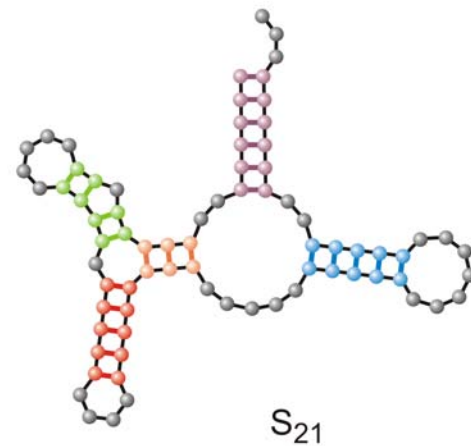
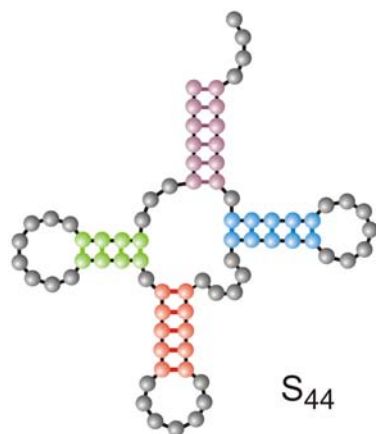
$$p = 0.001 / \text{site} \times \text{replication}$$

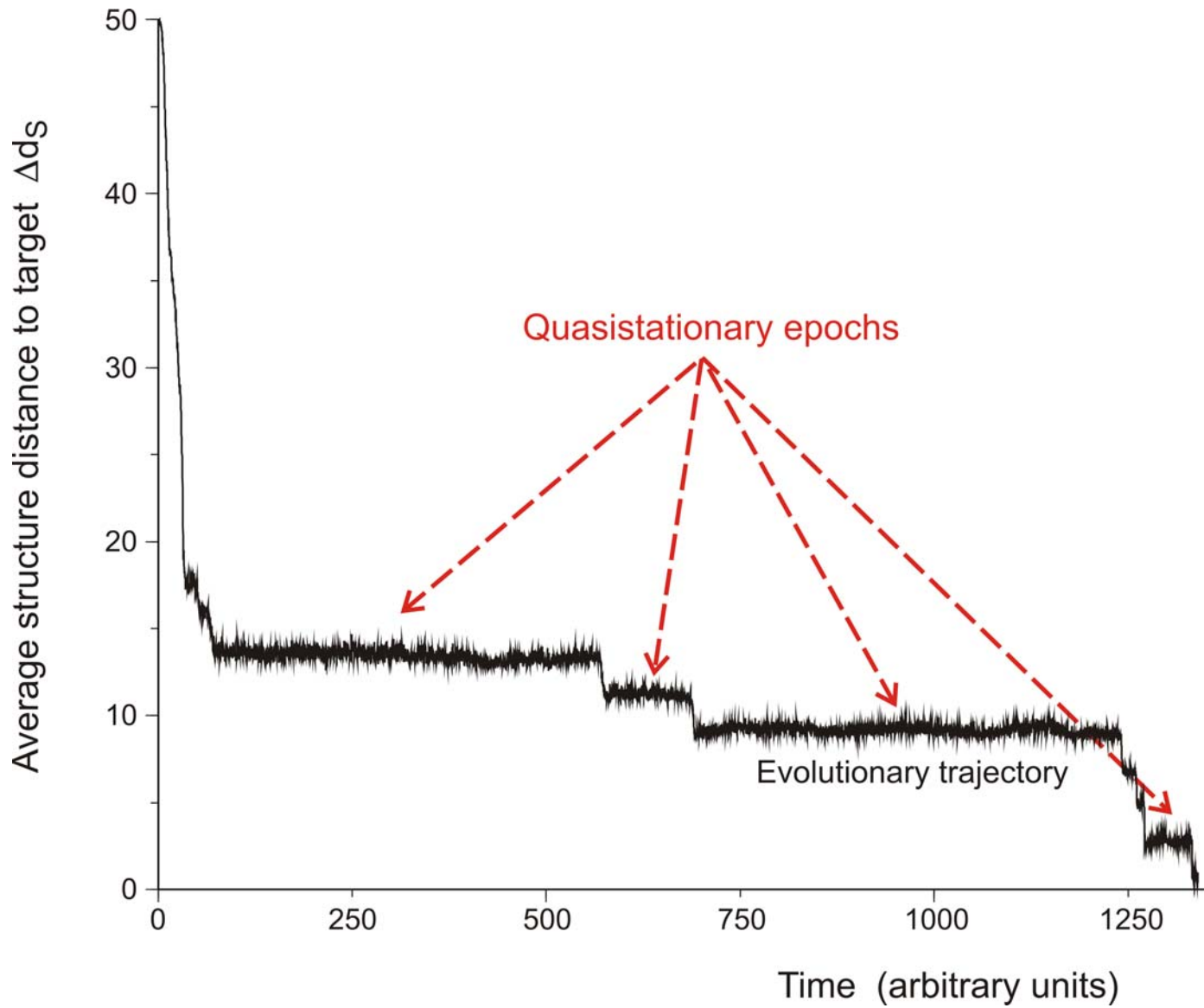
The flowreactor as a device for studies of evolution *in vitro* and *in silico*

Randomly chosen  
initial structure



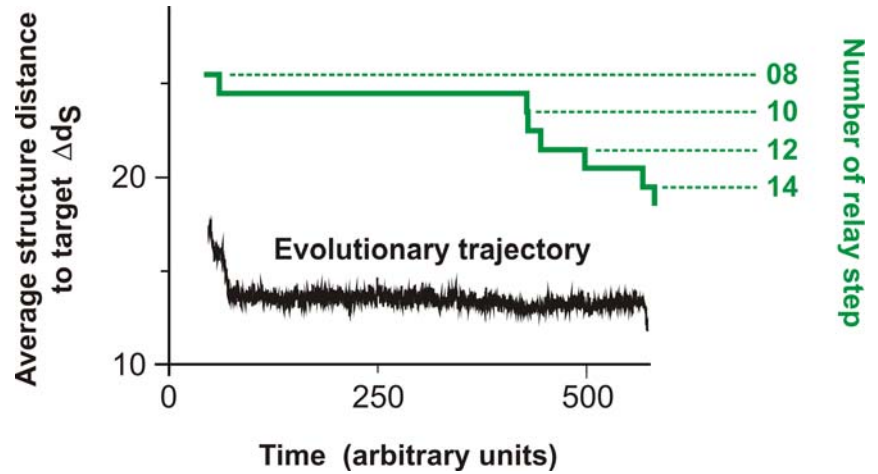
Phenylalanyl-tRNA  
as target structure





*In silico* optimization in the flow reactor: Evolutionary Trajectory

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



```

entry  GGUAUGGGCGUUGAAUAGG G U U U A A A C C A A U C G G C A A C G A U C U C G U G U G C G C A U U U C A U A U C C C G U A C A G A A
8      .(((((((((((((. . . . . (((. . . . .)))) . . . . .)))))) . . . . .(((((. . . . .))))))))) . . . . .
exit   GGUAUGGGCGUUGAAUA A U A G G G U U U A A A C C A A U C G G C C A A C G A U C U C G U G U G C G C A U U U C A U A U C C C A U A C A G A A
entry  GGUAUGGGCGUUGAAUA A U A G G G U U U A A A C C A A U C G G C C A A C G A U C U C G U G U G C G C A U U U C A U A U A C C A U A C A G A A
9      .(((((( . (((((. . . . . (((. . . . .)))) . . . . .)))) . . . . .(((((. . . . .)))) . ))))) . . . . .
exit   U G G A U G G A C G U U G A A U A A C A A G G U A U C G A C C A A A C A A C C A A C G A G U A A G U G U G U A C G C C C C A C A C A C G U C C C A A G
entry  U G G A U G G A C G U U G A A U A A C A A G G U A U C G A C C A A A C A A C C A A C G A G U A A G U G U G U A C G C C C C A C A C A C G U C C C A A G
10     .(((((. . (((((. . . . . (((. . . . .)))) . . . . .)))) . . . . .(((((. . . . .)))) . ))))) . . . . .
exit   U G G A U G G A C G U U G A A U A A C A A G G U A U C G A C C A A A C A A C C A A C G A G U A A G U G U G U A C G C C C C A C A C A C G U C C C A A G
  
```

**Transition inducing point mutations** change the molecular structure

**Neutral point mutations** leave the molecular structure unchanged

Neutral genotype evolution during phenotypic stasis



1. Nucleic acid structures
2. DNA nanotechnology
3. RNA - A magic molecule
4. Evolutionary optimization of structure
- 5. RNA design**

## 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

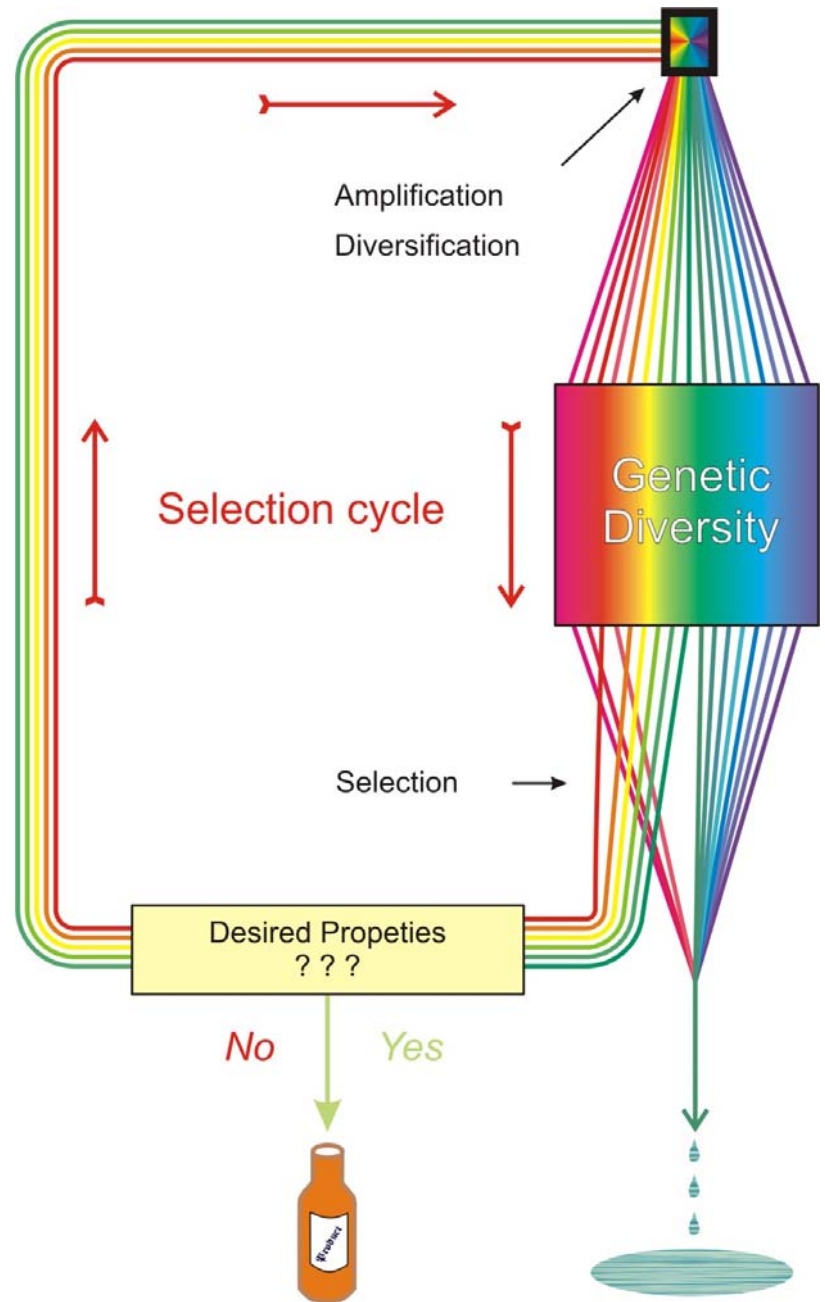
C.Tuerk, L.Gold, *SELEX - Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science **249** (1990), 505-510

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

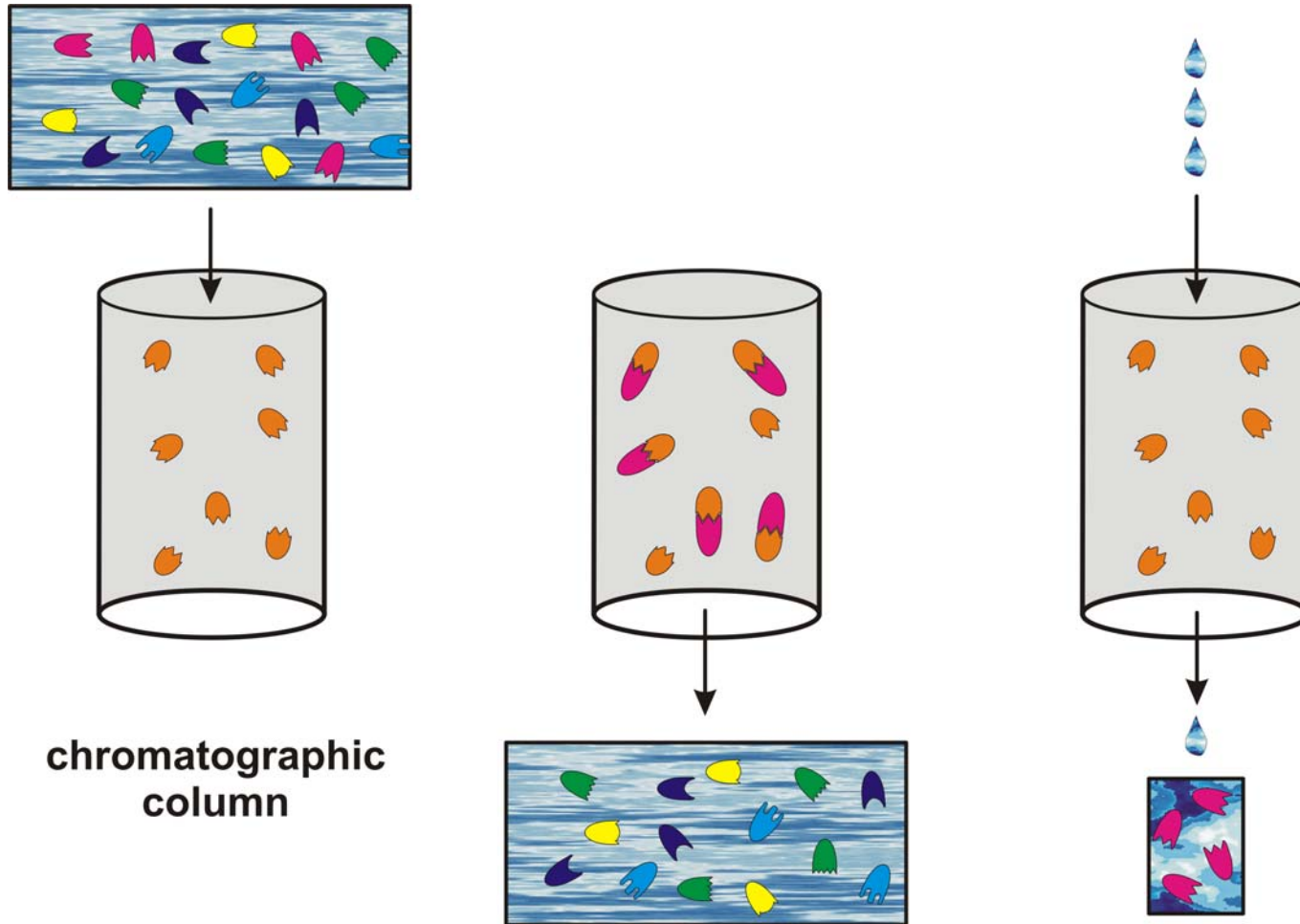
R.D.Jenison, S.C.Gill, A.Pardi, B.Poliski, *High-resolution molecular discrimination by RNA*. Science **263** (1994), 1425-1429

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

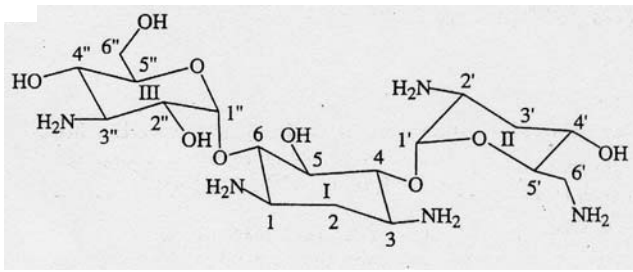
Jiang, A. K. Suri, R. Fiala, D. J. Patel, *Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex*. Chemistry & Biology **4** (1997), 35-50



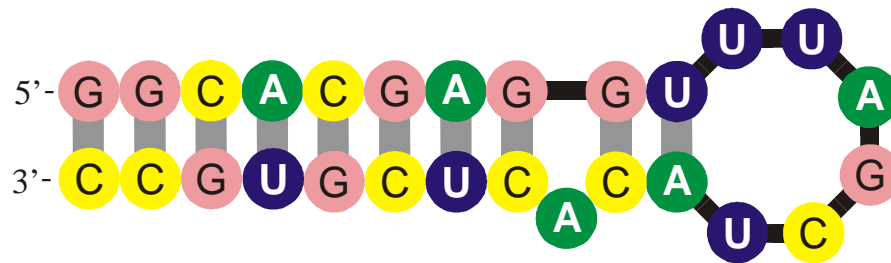
Selection of molecules with predefined properties in laboratory experiments



The SELEX technique for the evolutionary design of strong binders called aptamers



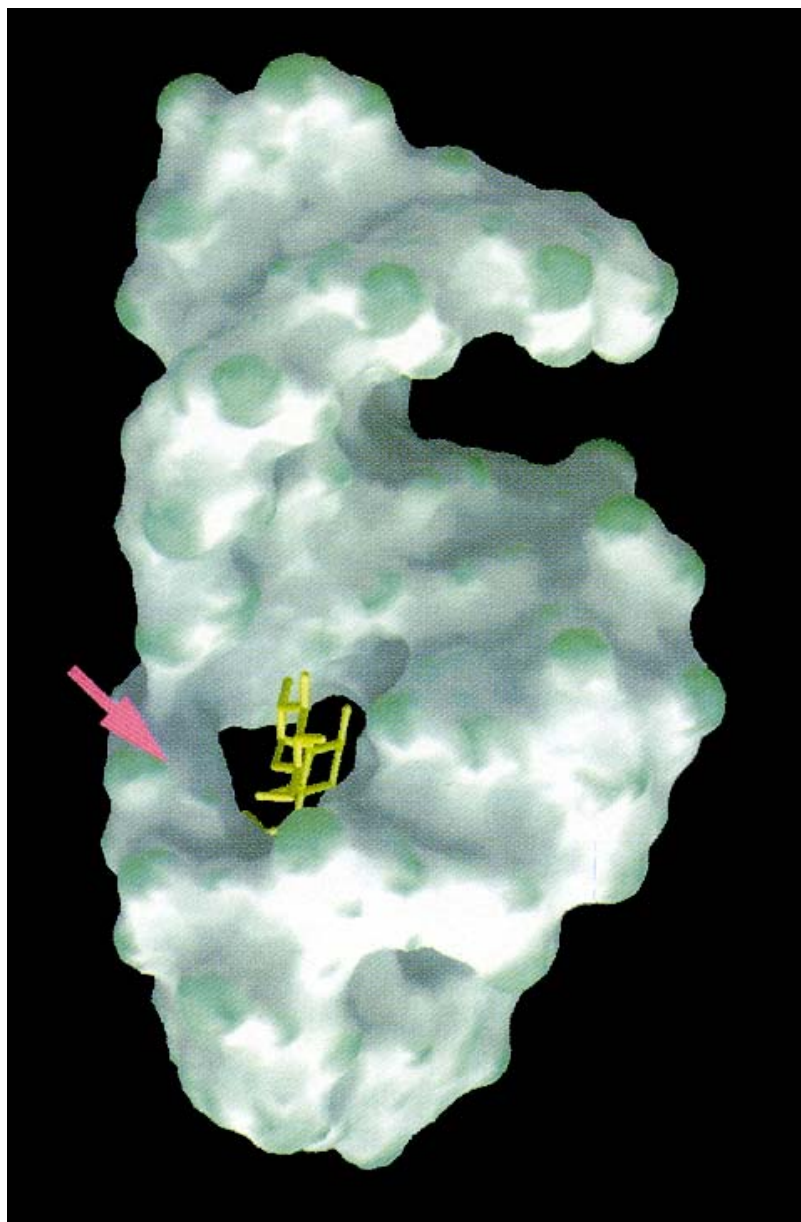
tobramycin



RNA aptamer

Secondary structure of the tobramycin binding RNA aptamer with  $K_D = 9 \text{ nM}$

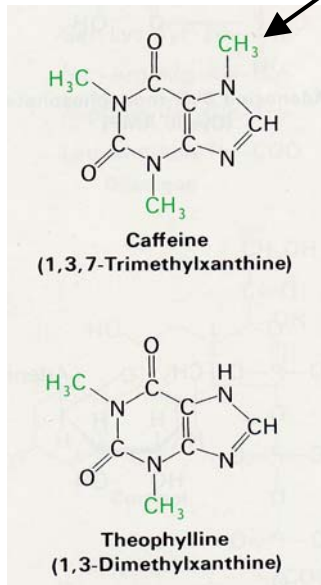
L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chemistry & Biology* 4:35-50 (1997)



## The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel,  
*Chemistry & Biology* 4:35-50 (1997)

additional methyl group

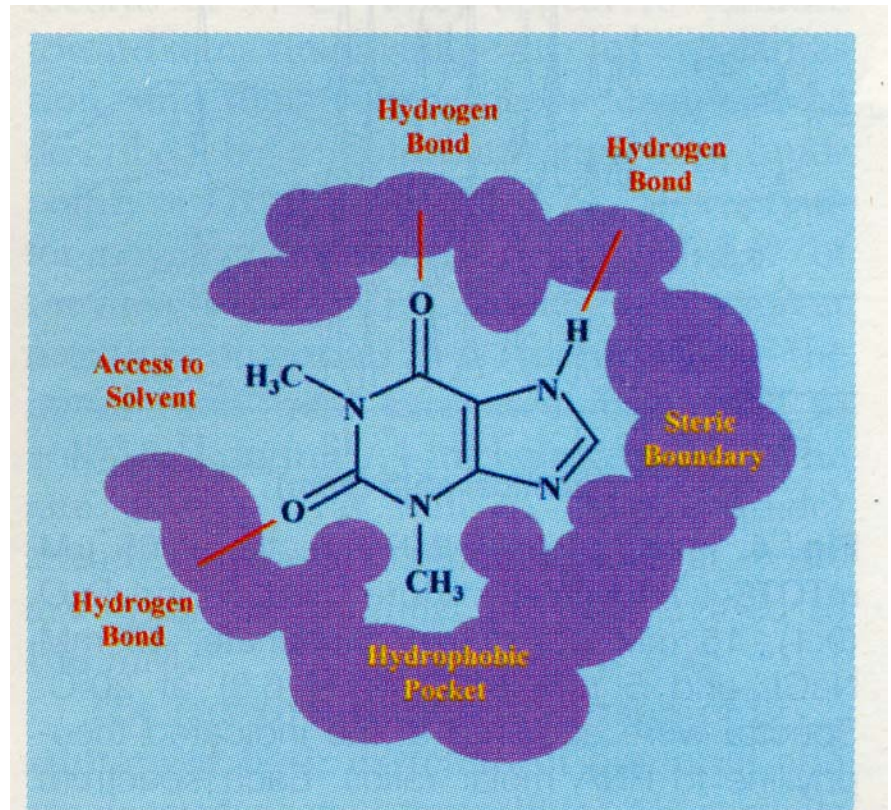


Dissociation constants and specificity of theophylline, caffeine, and related derivatives of uric acid for binding to a discriminating aptamer TCT8-4

**Table 1.** Competition binding analysis with TCT8-4 RNA. The chemical structures are shown for a series of derivatives used in competitive binding experiments with TCT8-4 RNA (Fig. 2) (20). The right column represents the affinity of the competitor relative to theophylline,  $K_d(c)/K_d(t)$ , where  $K_d(c)$  is the individual competitor dissociation constant and  $K_d(t)$  is the competitive dissociation constant of theophylline. Certain data (denoted by >) are minimum values that were limited by the solubility of the competitor. Each experiment was carried out in duplicate. The average error is shown.

Compound	Structure	$K_d(c)$ ( $\mu$ M)	$K_d(c)/K_d(t)$
Theophylline		$0.32 \pm 0.13$	1
CP-theophylline		$0.93 \pm 0.20$	2.9
Xanthine		$8.5 \pm 0.40$	27
1-Methylxanthine		$9.0 \pm 0.30$	28
3-Methylxanthine		$2.0 \pm 0.7$	6.3
7-Methylxanthine		> 500	>1500
3,7-Dimethylxanthine		> 500	> 1500
1,3-Dimethyluric acid		> 1000	>3100
Hypoxanthine		$49 \pm 10$	153
Caffeine		$3500 \pm 1500$	10,900





**Fig. 3.** Schematic representation of the RNA (purple) binding site for theophylline (blue).

Schematic drawing of the aptamer binding site for the theophylline molecule



## **Hammerhead ribozyme – The smallest RNA based catalyst**

H.W.Pley, K.M.Flaherty, D.B.McKay, *Three dimensional structure of a hammerhead ribozyme*. Nature **372** (1994), 68-74

W.G.Scott, J.T.Finch, A.Klug, *The crystal structures of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage*. Cell **81** (1995), 991-1002

J.E.Wedekind, D.B.McKay, *Crystallographic structures of the hammerhead ribozyme: Relationship to ribozyme folding and catalysis*. Annu.Rev.Biophys.Biomol.Struct. **27** (1998), 475-502

G.E.Soukup, R.R.Breaker, *Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization*. Structure **7** (1999), 783-791

Allosteric effectors:

FMN = flavine mononucleotide

H10 – H12

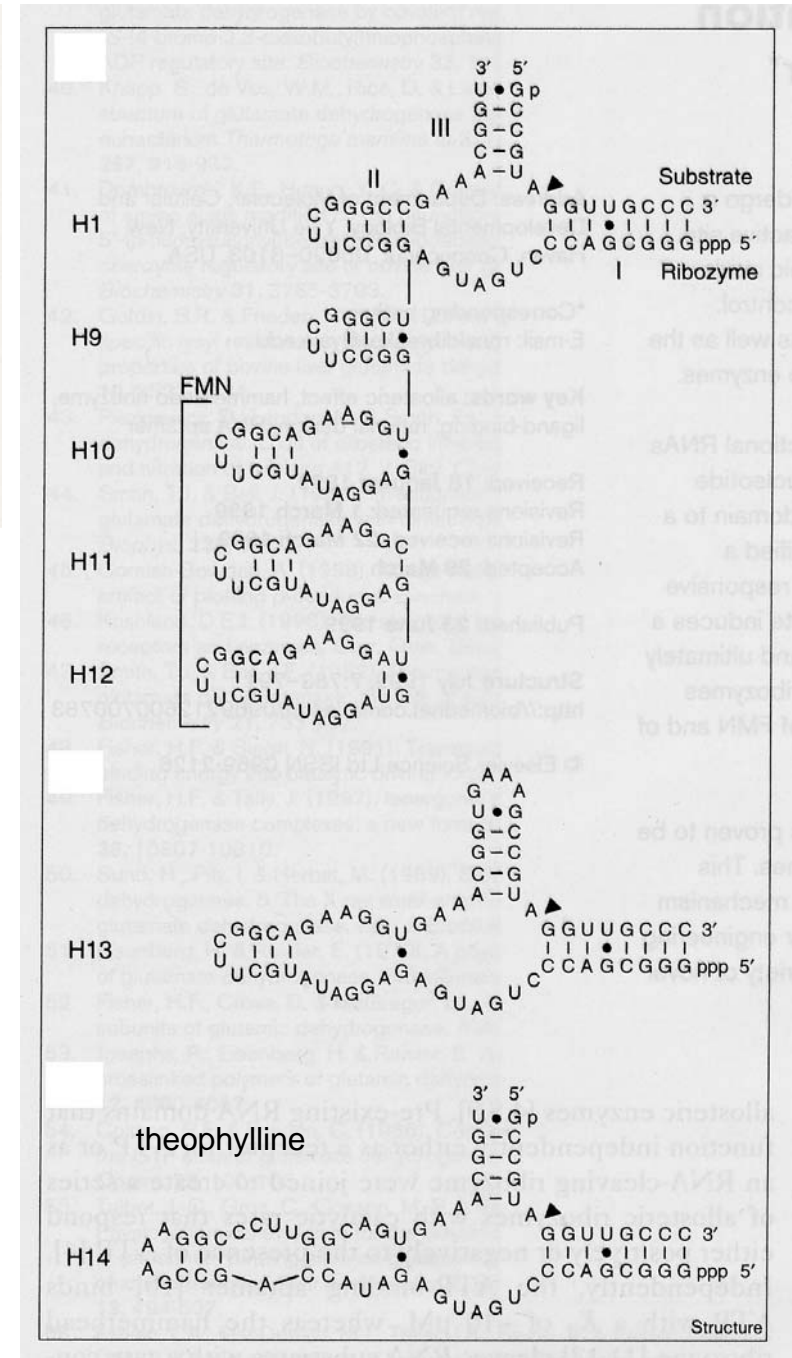
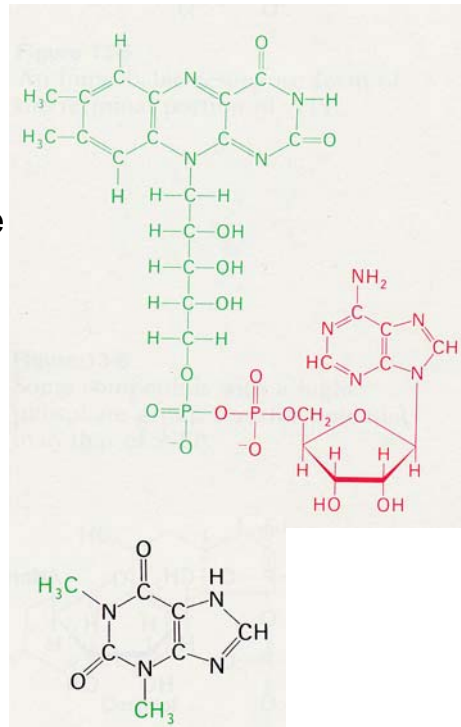
theophylline

H14

Self-splicing allosteric ribozyme

H13

Hammerhead ribozymes with  
allosteric effectors



- 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).

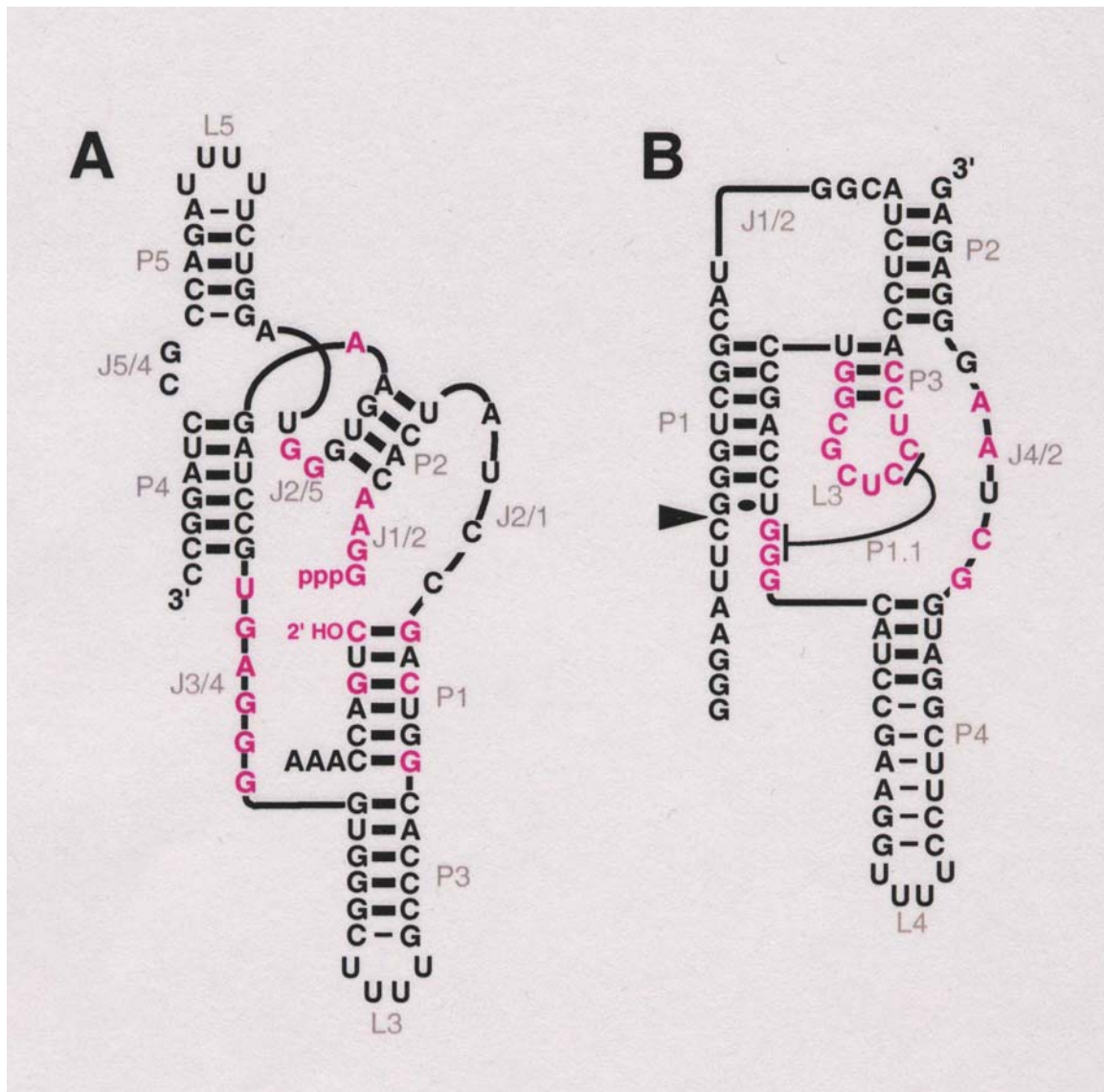
The prototype class III and HDV ribozymes have no more than the 25% sequence identity expected by chance and no fortuitous structural similarities that might favor an intersection of their two neutral networks. Nevertheless, sequences can be designed that simultaneously satisfy the base-pairing requirements

## 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

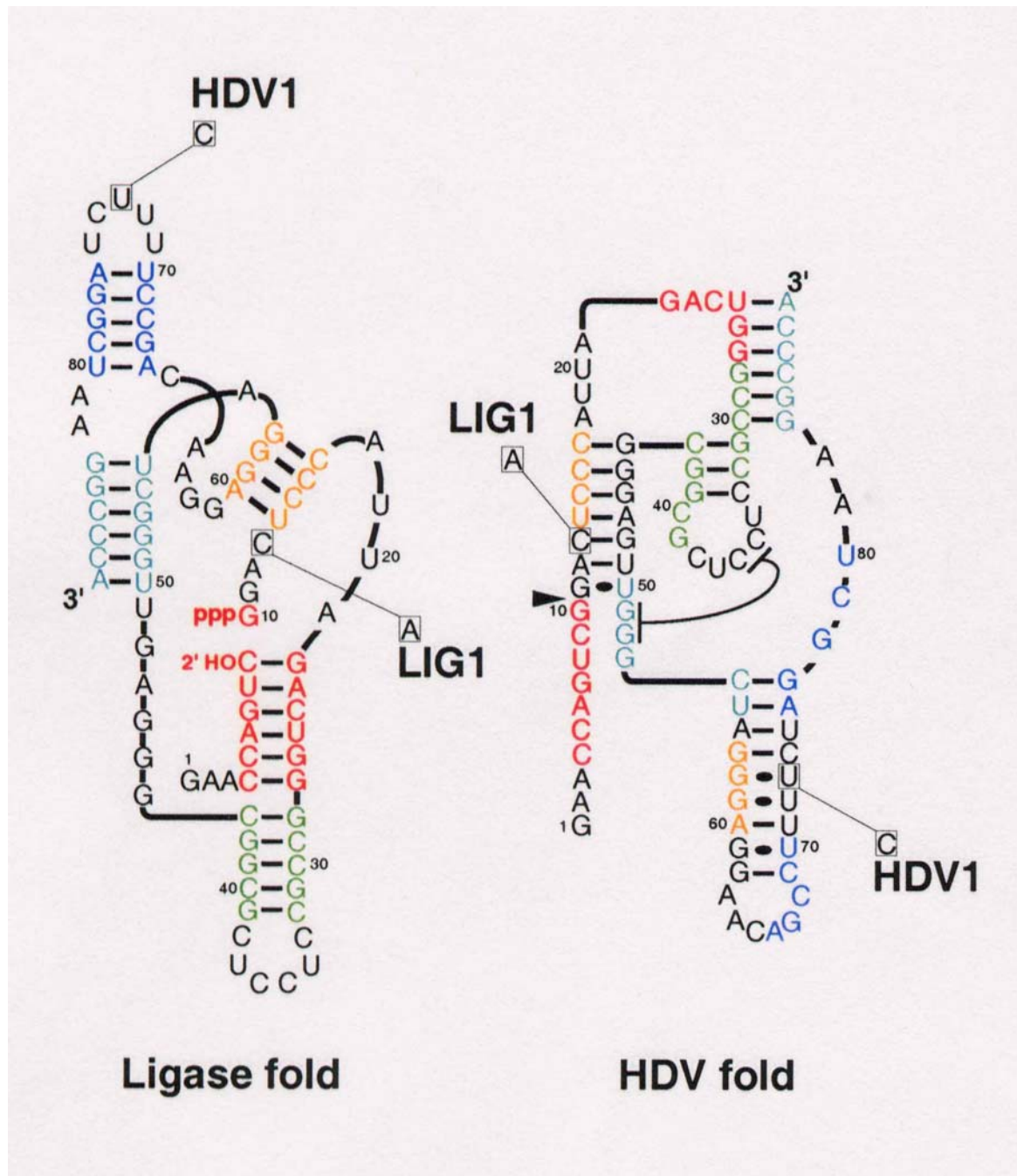


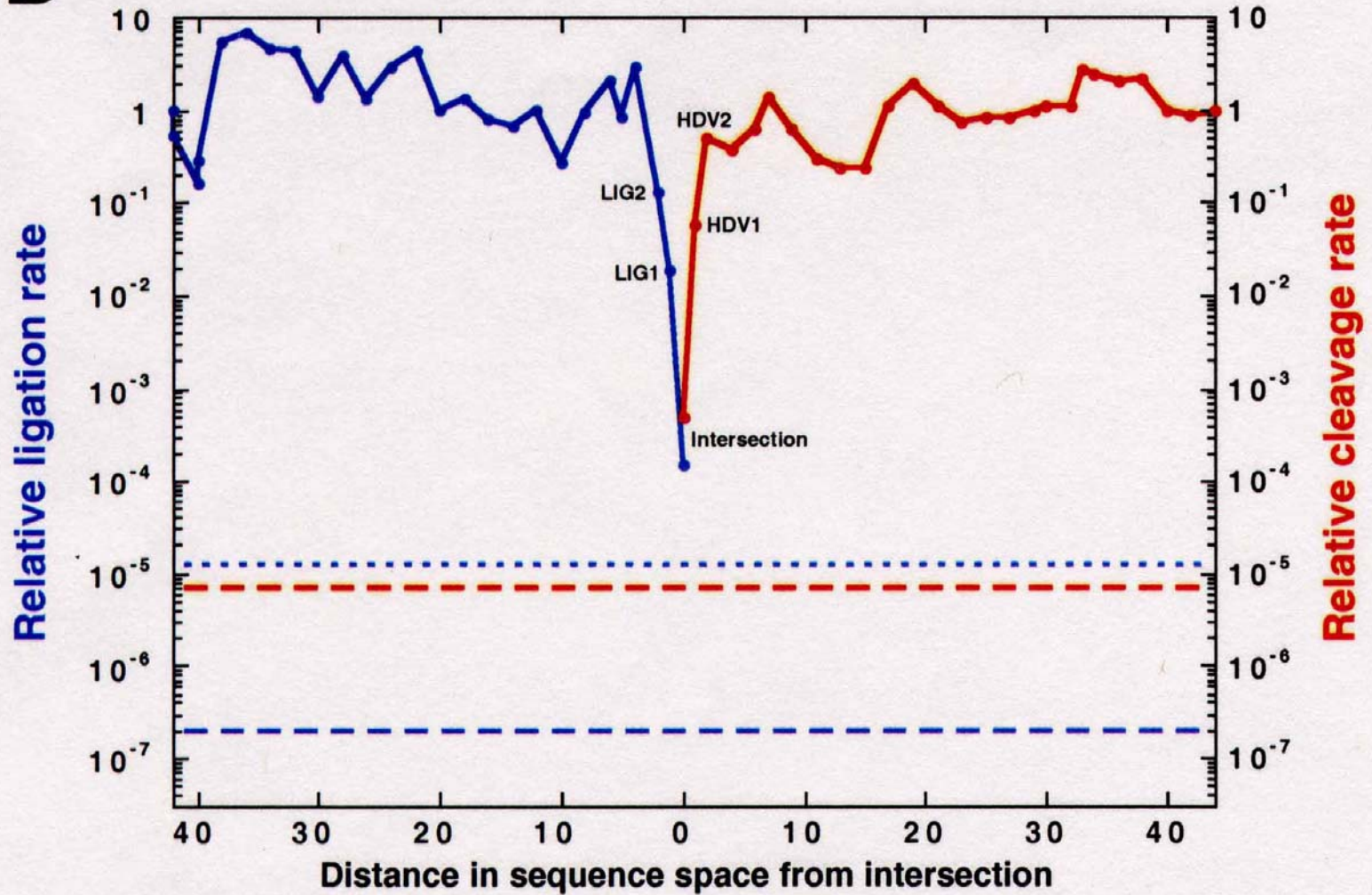
Two ribozymes of chain lengths  $n = 88$  nucleotides: An artificial ligase (A) and a natural cleavage ribozyme of hepatitis- $\delta$ -virus (B)



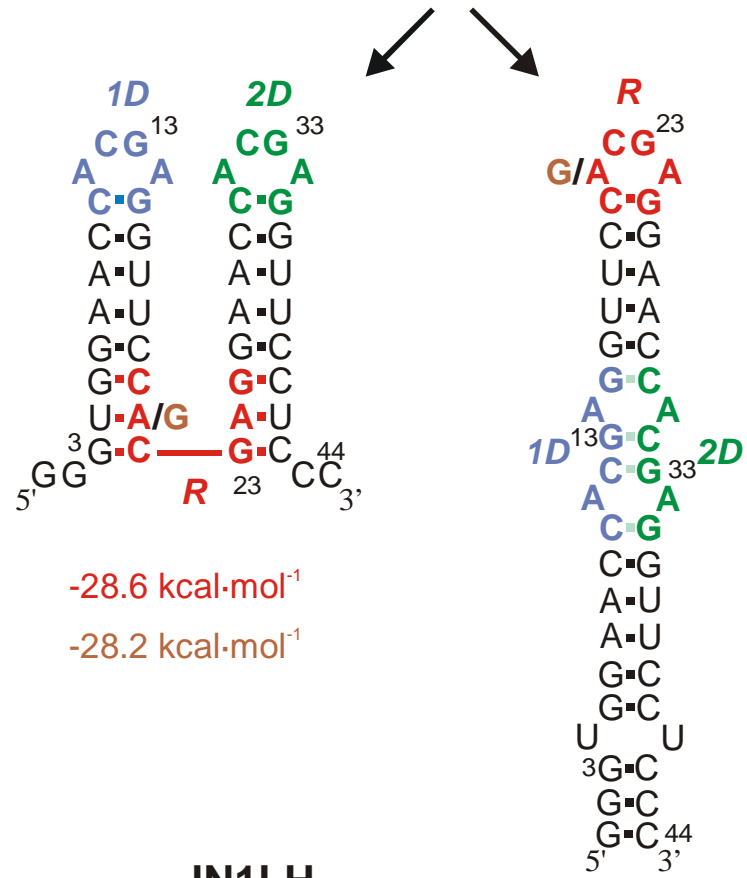
The sequence at the intersection:

An RNA molecule, which is 88 nucleotides long and which can form both structures.



**B**

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



-28.6 kcal·mol<sup>-1</sup>  
 -28.2 kcal·mol<sup>-1</sup>

-28.6 kcal·mol<sup>-1</sup>  
 -31.8 kcal·mol<sup>-1</sup>

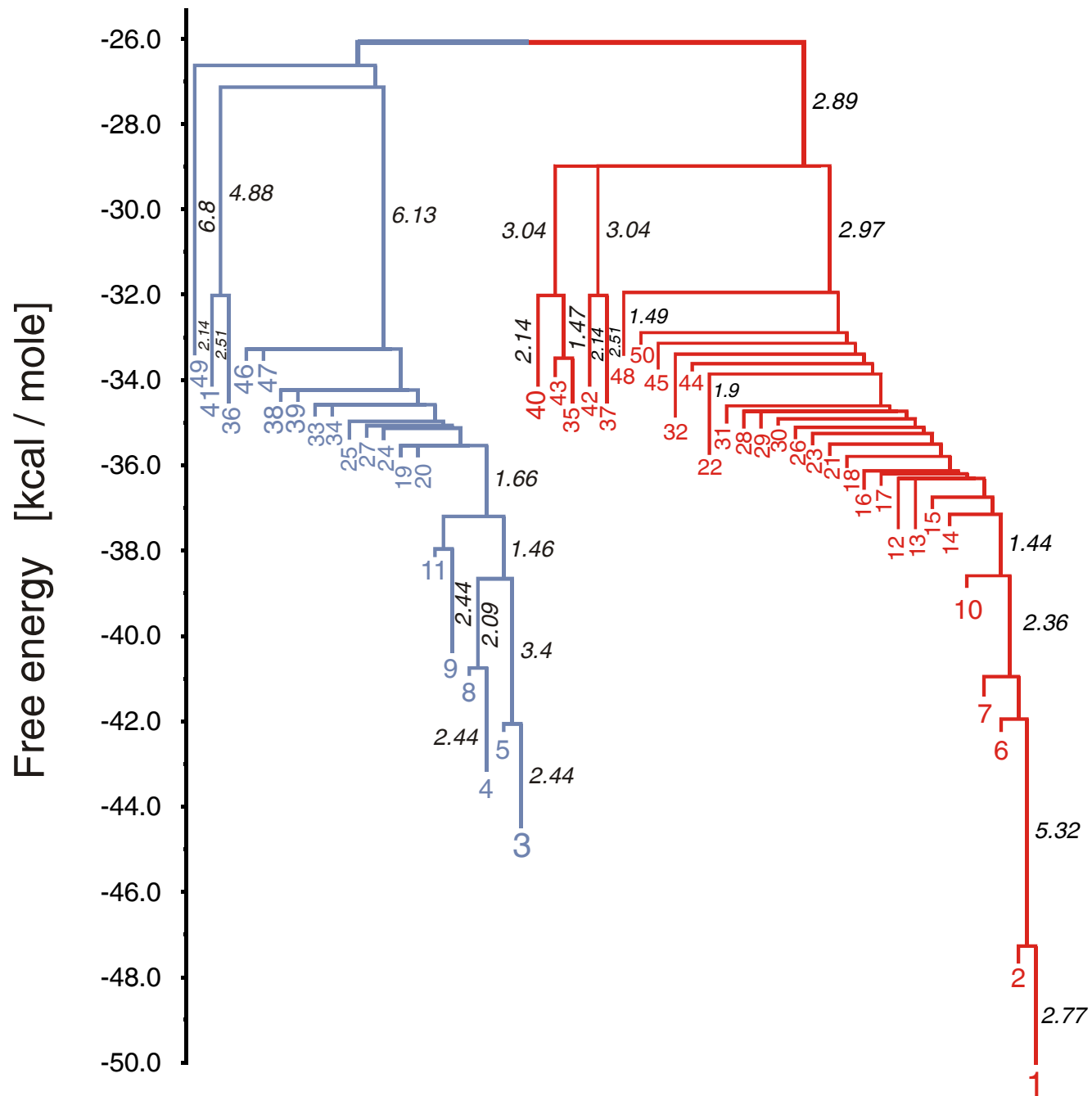
## An experimental RNA switch

JN1LH

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

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

J1LH barrier tree





## Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)  
Projects No. 09942, 10578, 11065, 13093  
13887, and 14898

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)  
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank  
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

Austrian Genome Research Program – GEN-AU: Bioinformatics  
Network (BIN)

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute



Universität Wien

# Coworkers

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

**Paul E. Phillipson**, University of Colorado at Boulder, CO

**Heinz Engl, Philipp Kügler, James Lu, Stefan Müller**, RICAM Linz, AT

**Jord Nagel, Kees Pleij**, Universiteit Leiden, NL

**Walter Fontana**, Harvard Medical School, MA

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

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

**Ivo L.Hofacker, Christoph Flamm, Andreas Svrček-Seiler**, Universität Wien, AT

**Kurt Grünberger, Michael Kospach, Andreas Wernitznig, Stefanie Widder,  
Stefan Wuchty**, Universität Wien, AT

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



Universität Wien

Web-Page for further information:

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

