Life – A Result of Evolution or Design ?

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Österreich
und
The Santa Fe Institute, Santa Fe, New Mexico, USA

Conference on Knots and other Entanglement in Biopolymers

Trieste, ICTP, 15.– 19.09.2008

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

Kardinal Christoph Schönborn, ***Finding Design in Nature***, commentary in *The New York Times*, July 5, 2005

„ ... Evolution in the sense of common ancestry might be true, but evolution in the Neo-Darwinian sense – an unguided, unplanned process of random variation and natural selection – is not. Any system of thought that denies or seeks to explain away the overwhelming evidence for design in biology is ideology, not science.

... Scientific theories that try to explain away the appearance of design as the result of ‚chance and necessity' are not scientific at all, but ... an abdication of human intelligence."

Peter Schuster. ***Evolution and design. The Darwinian theory of evolution is a scientific fact and not an ideology.*** Complexity **11**(1):12-15, 2006

Peter Schuster. ***Evolution und Design. Versuch einer Bestandsaufnahme der Evolutionstheorie***.

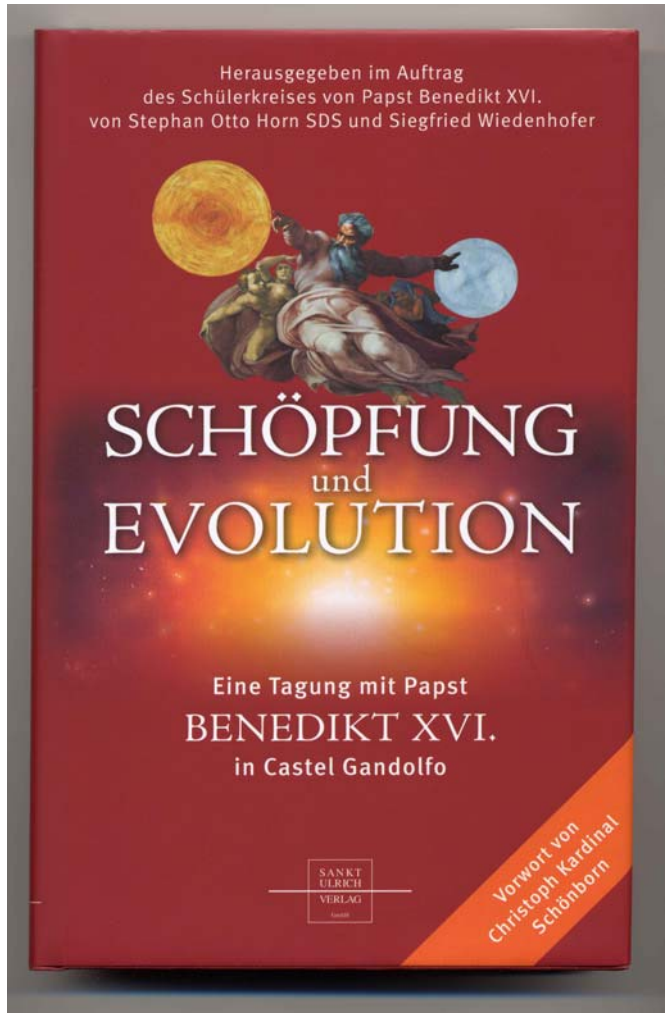
In: Stephan Otto Horn und Siegfried Wiedenhofer, Eds.

Schöpfung und Evolution. Eine Tagung mit Papst Benedikt XVI
in Castel Gandolfo. Sankt Ulrich Verlag, Augsburg 2007, pp.25-56.

English translation:

Creation and Evolution.

Ignatius Press, San Francisco, CA, 2008



1. Evolution – organismic and molecular
2. Multiplication, mutation, and selection
3. Rational design of molecules
4. Evolution and optimization of molecules
5. Origin of biological complexity
6. Biology and probabilities

1. **Evolution – organismic and molecular**
2. Multiplication, mutation, and selection
3. Rational design of molecules
4. Evolution and optimization of molecules
5. Origin of biological complexity
6. Biology and probabilities

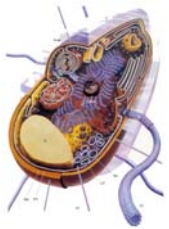
Genotype, Genome

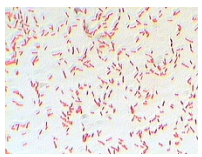
Collection of genes

Unfolding of the genotype

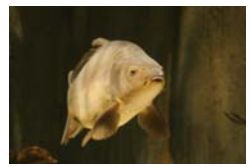
Developmental program

Highly specific environmental conditions












Evolution explains the origin of species and their interactions

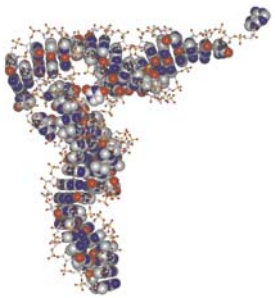



Genotype, Genome

GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

Quantitative biology

'the new biology is the chemistry of living matter'



evolution of RNA molecules, ribozymes and splicing, the idea of an RNA world, selection of RNA molecules, RNA editing, the ribosome is a ribozyme, small RNAs and RNA switches.

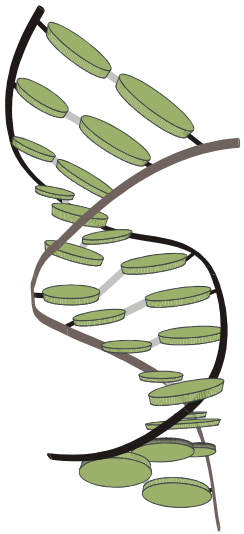
The exciting RNA story

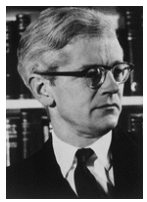
Biochemistry
molecular biology
structural biology
molecular evolution
molecular genetics
systems biology
bioinfomatics
epigenetics

Unfolding of the genotype

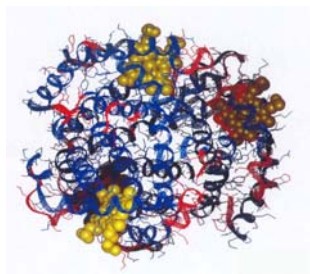
Highly specific environmental conditions

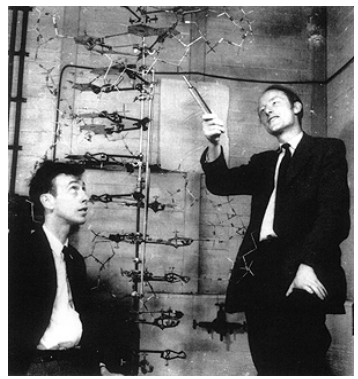
Phenotype





John Kendrew





James D. Watson und Francis H.C. Crick

Manfred Eigen

Molecular evolution
Linus Pauling and
Emile Zuckerkandl

Hemoglobin sequence
Gerhard Braunitzer

Max Perutz

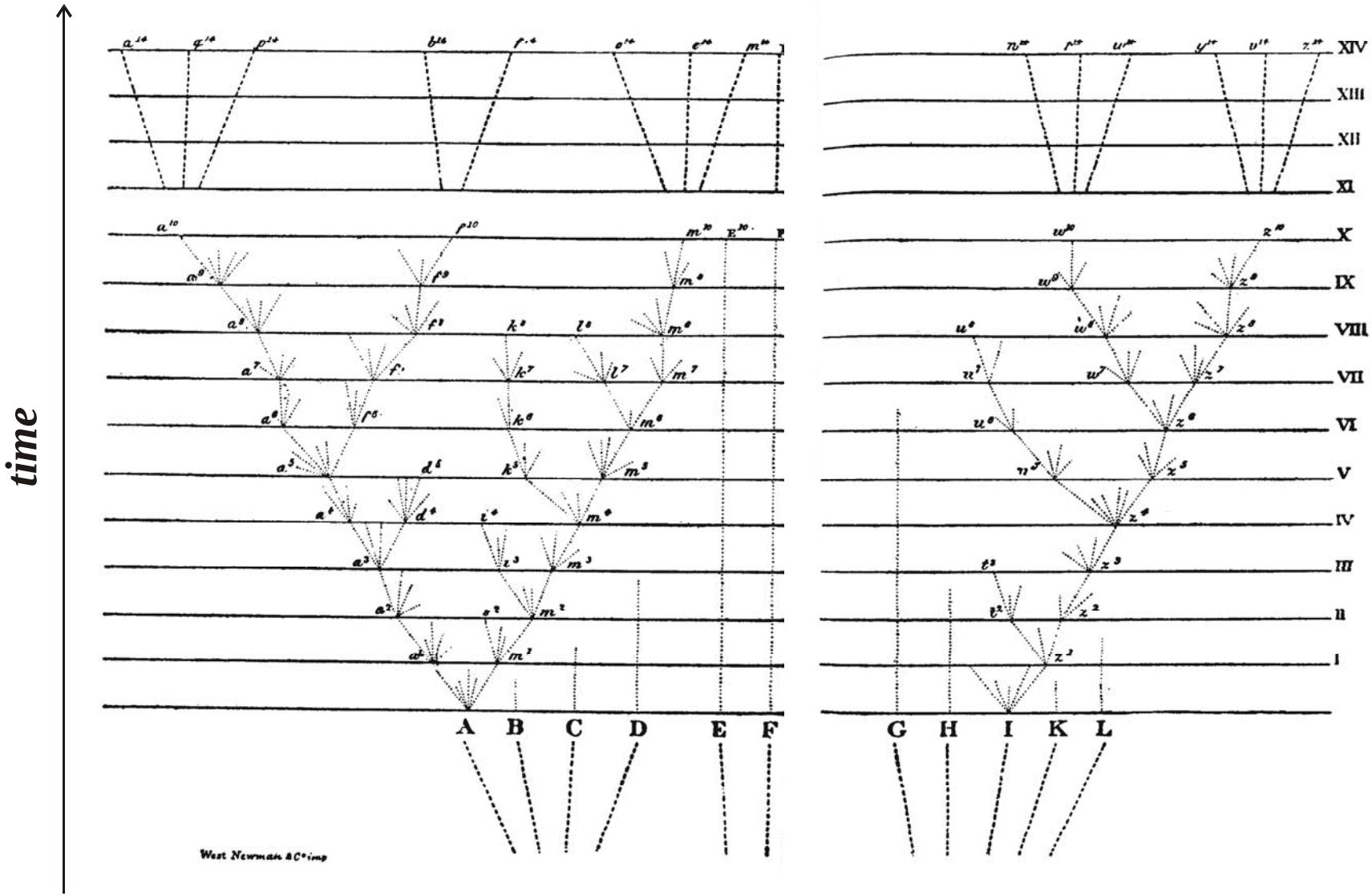
Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

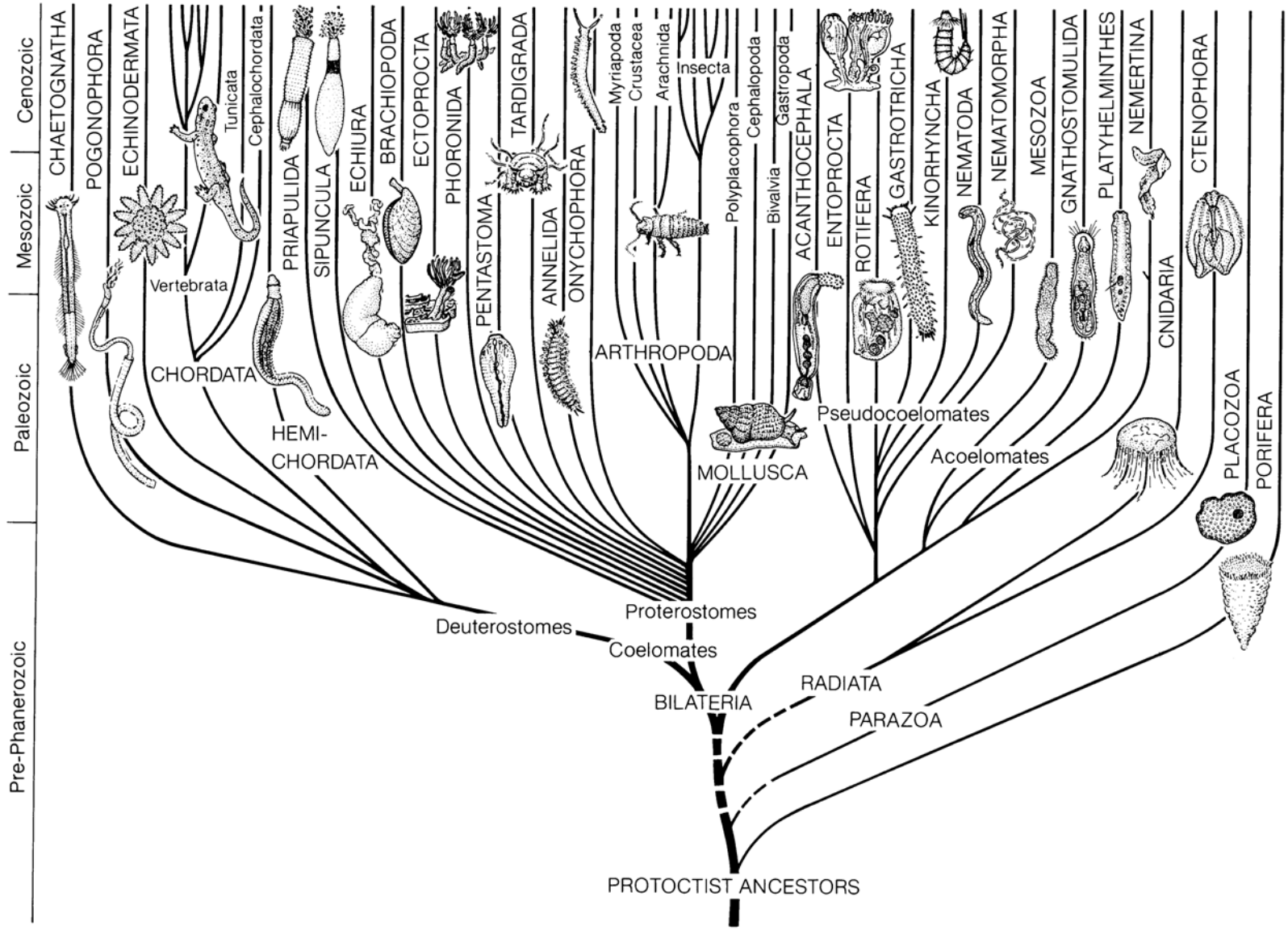
Variation through mutation and recombination operates on the **genotype** whereas the **phenotype** is the target of **selection**.

One important property of the Darwinian scenario is that **variations** in the form of mutations or recombination events occur **uncorrelated** with their **effects on the selection process**.

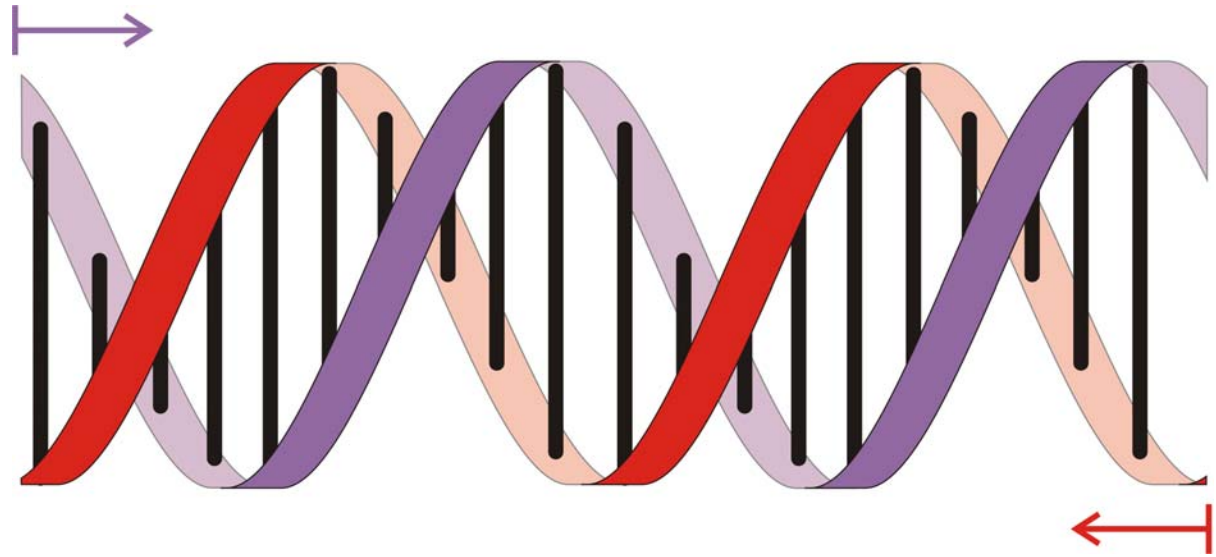
All conditions can be fulfilled not only by cellular organisms but also by **nucleic acid molecules** in suitable **cell-free experimental assays**.

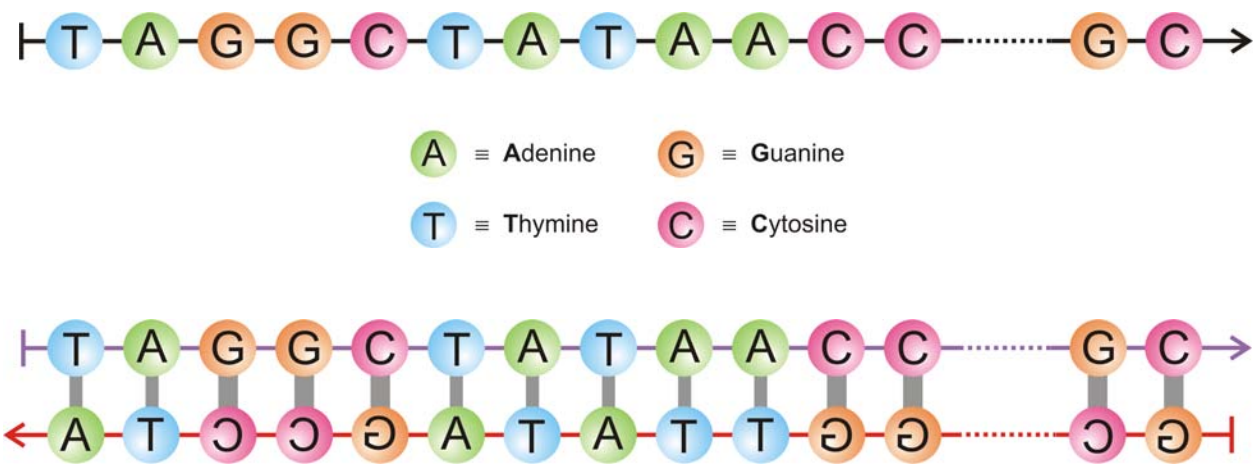


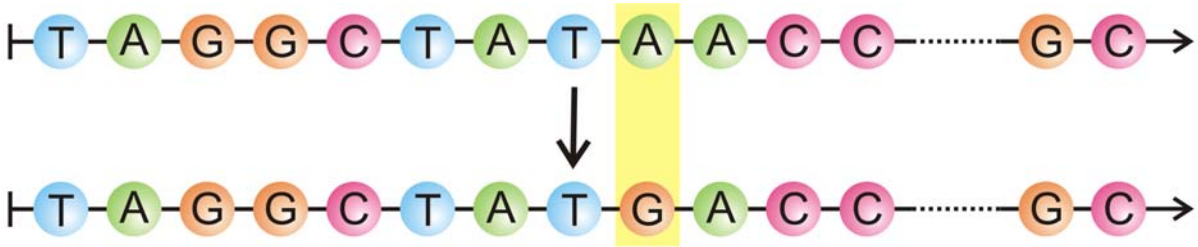
Charles Darwin, *The Origin of Species*, 6th edition.
Everyman‘s Library, Vol.811, Dent London, pp.121-122.



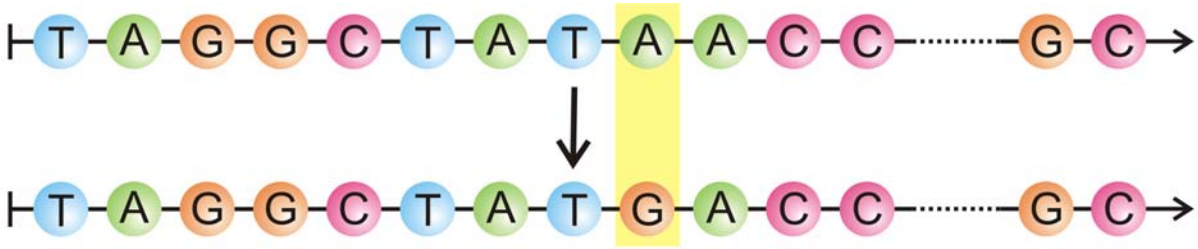
Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth*. W.H. Freeman, San Francisco, 1982.



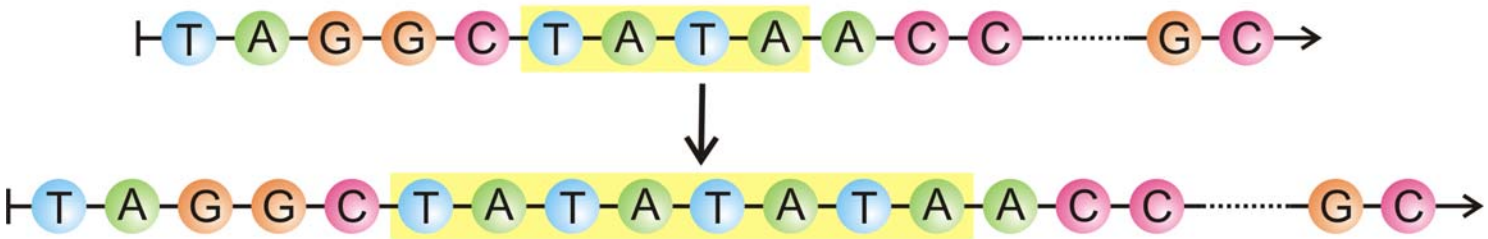

T A G G C T A T A A C C G C
A ≡ Adenine
G ≡ Guanine
T ≡ Thymine
C ≡ Cytosine
T A G G C T A T A A C C G C
A T C C G A T A T T G G C G



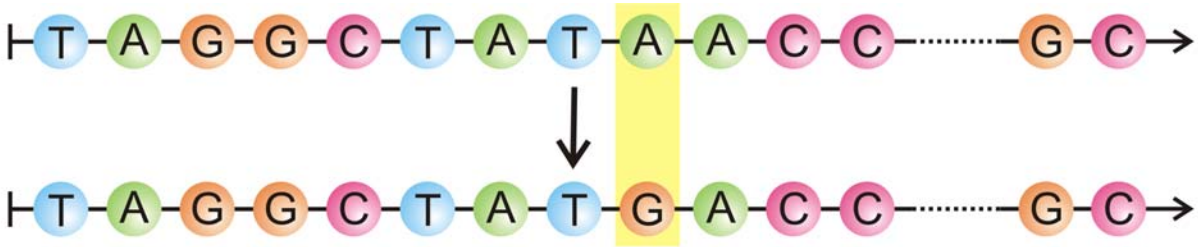
Point mutation



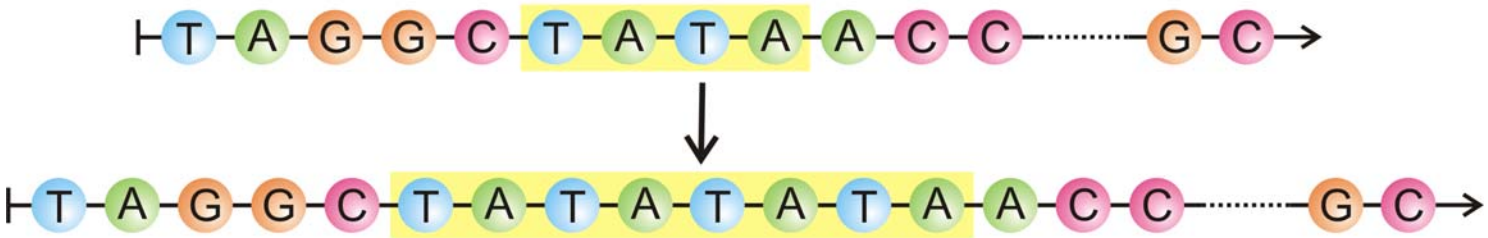
Point mutation



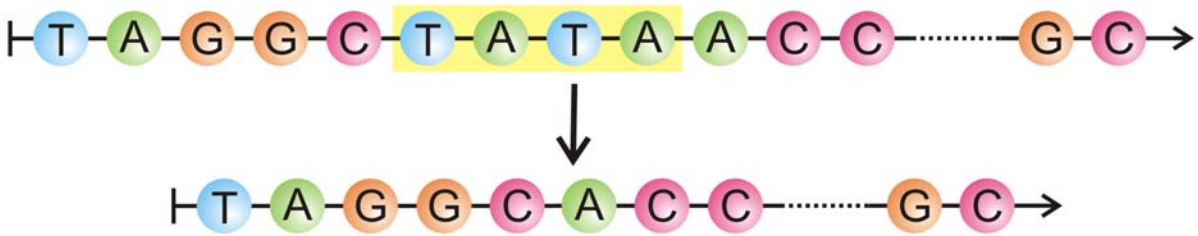
Insertion



Point mutation



Insertion



Deletion

Reconstruction of phylogenies through comparison of molecular sequence data

Results from molecular evolution:

• The molecular machineries of all present day cells are very similar and provide a strong hint that all life on Earth descended from one common ancestor (called „last universal common ancestor", **LUCA)**.

• Comparison of DNA sequences from present day organisms allows for a reconstruction of phylogenetic trees, which are (almost) identical with those derived from morphological comparison of species and the paleontologic record of fossils.

1. Evolution – organismic and molecular
2. **Multiplication, mutation, and selection**
3. Rational design of molecules
4. Evolution and optimization of molecules
5. Origin of biological complexity
6. Biology and probabilities

DIE NATURWISSENSCHAFTEN

58. Jahrgang, 1971 — Heft 10 Oktober

Selforganization of Matter and the Evolution of Biological Macromolecules

MANFRED EIGEN*

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

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, which even in its simplest forms always appears to be associated with complex macroscopic (i.e. multimolecular) systems, such as the living cell.

As a consequence of the exciting discoveries of "molecular biology", a common version of the above question is: *Which came first, the protein or the nucleic acid?* – a modern variant of the old "chicken-and-the-egg" problem. The term "first" is usually meant to define a causal rather than a temporal relationship, and the words "protein" and "nucleic acid" may be substituted by "function" and "information". The question in this form, when applied to the interplay of nucleic acids and proteins as presently encountered in the living cell, leads ad absurdum, because "function"

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

Die Naturwissenschaften 64. Jahrgang Heft 11 November 1977

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-replicative 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 (degenerate) master copies. External constraints enforce the selection of the best adapted distribution, commonly 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 deteriorate irreversibly leading to an error catastrophy. 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 functional linkages. A stable functional integration then will raise the system to a new level of organization and thereby enlarge its information capacity considerably. The hypercycle appears to be such a form of organization.

Preview on Part B: The Abstract Hypercycle

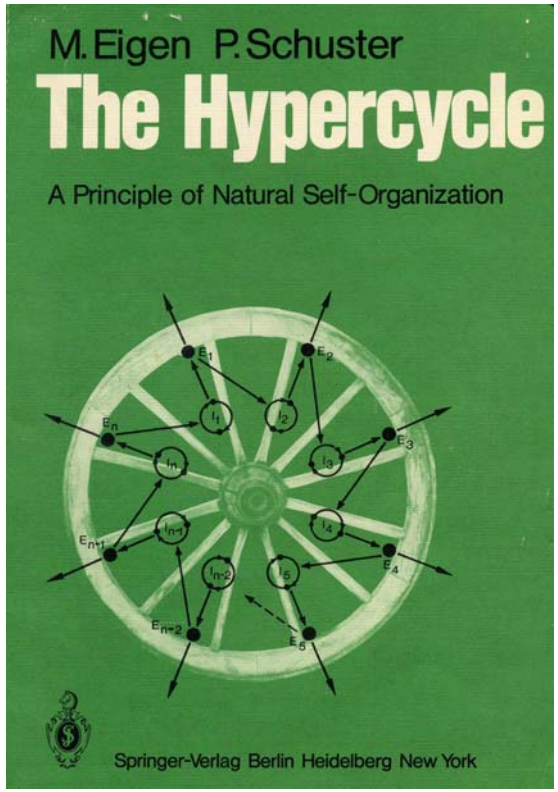
The mathematical analysis of dynamical systems using methods of differential topology, yields the result that there is only one type of mechanisms which fulfills the following requirements: The information stored in each single replicative unit (or reproductive cycle) must be maintained, i.e., the respective 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 linked ensemble which does not contribute to its integrated functions.
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 chains or branched, tree-like networks are devoid of such properties.
The mathematical methods used for proving these assertions are fixed-point, Lyapunov- and trajectorial analysis in higher-dimensional phase spaces, 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 Realistic 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 origination with finite probability under prebiotic conditions.
2) It permits a continuous emergence from closely interrelated (t-RNA-like) 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 still 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 answere 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-

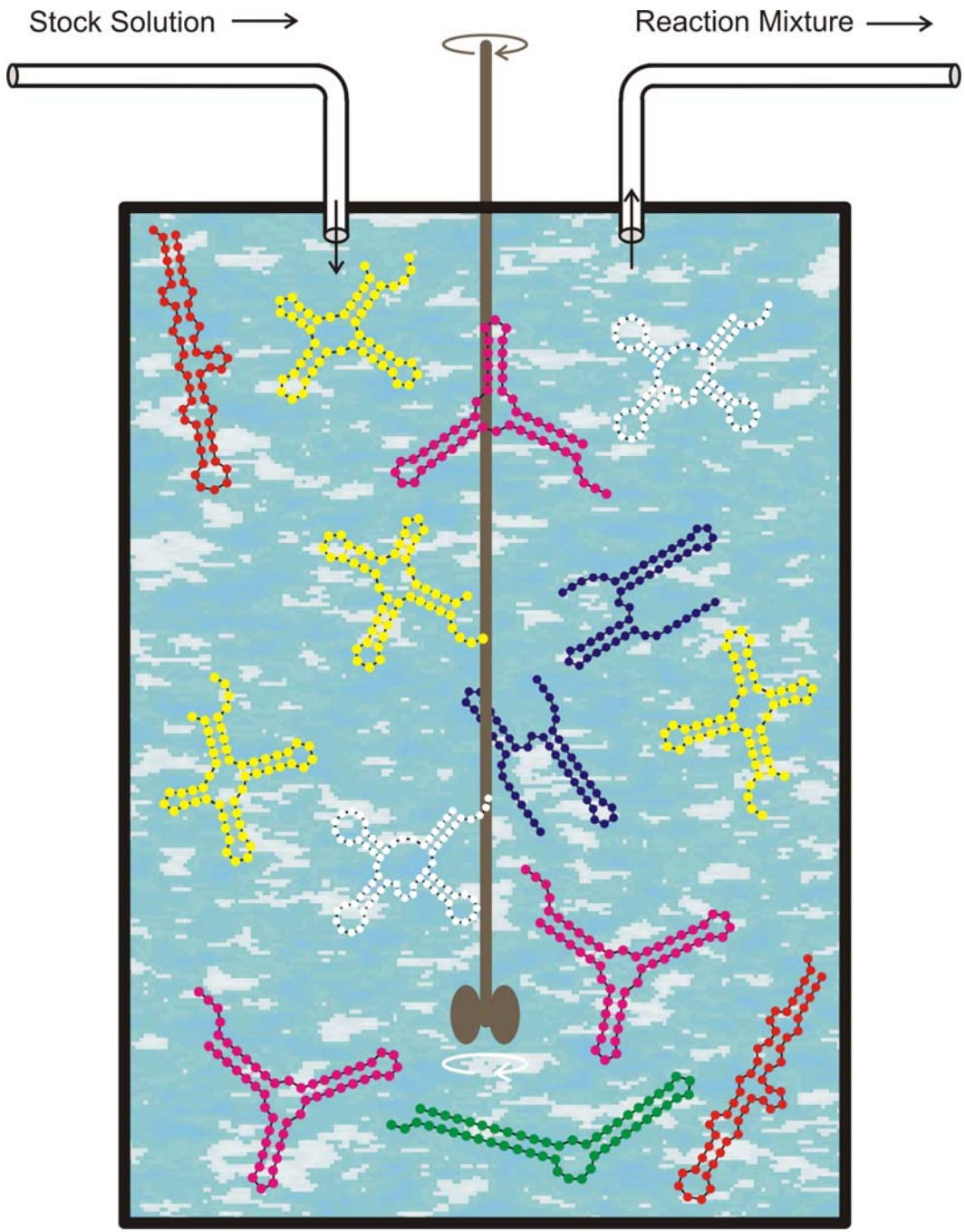


Chemical kinetics of molecular evolution

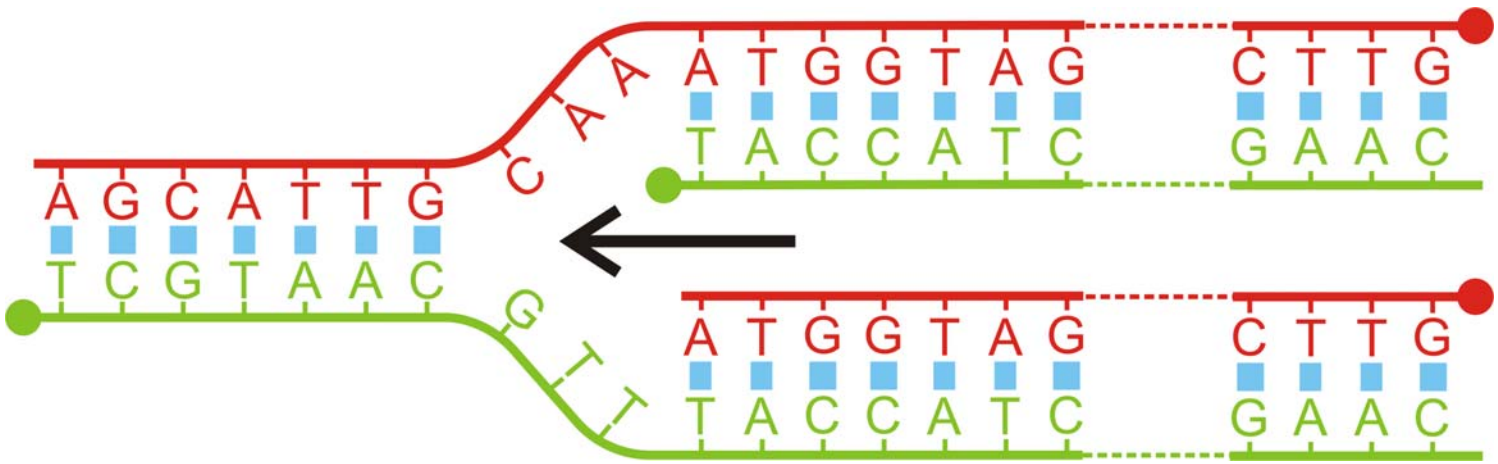
M. Eigen, P. Schuster, `The Hypercycle´, Springer-Verlag, Berlin 1979

Stock solution:

activated monomers, **ATP, CTP, GTP, UTP (TTP);**
a replicase, an enzyme that performs complemantary replication;
buffer solution

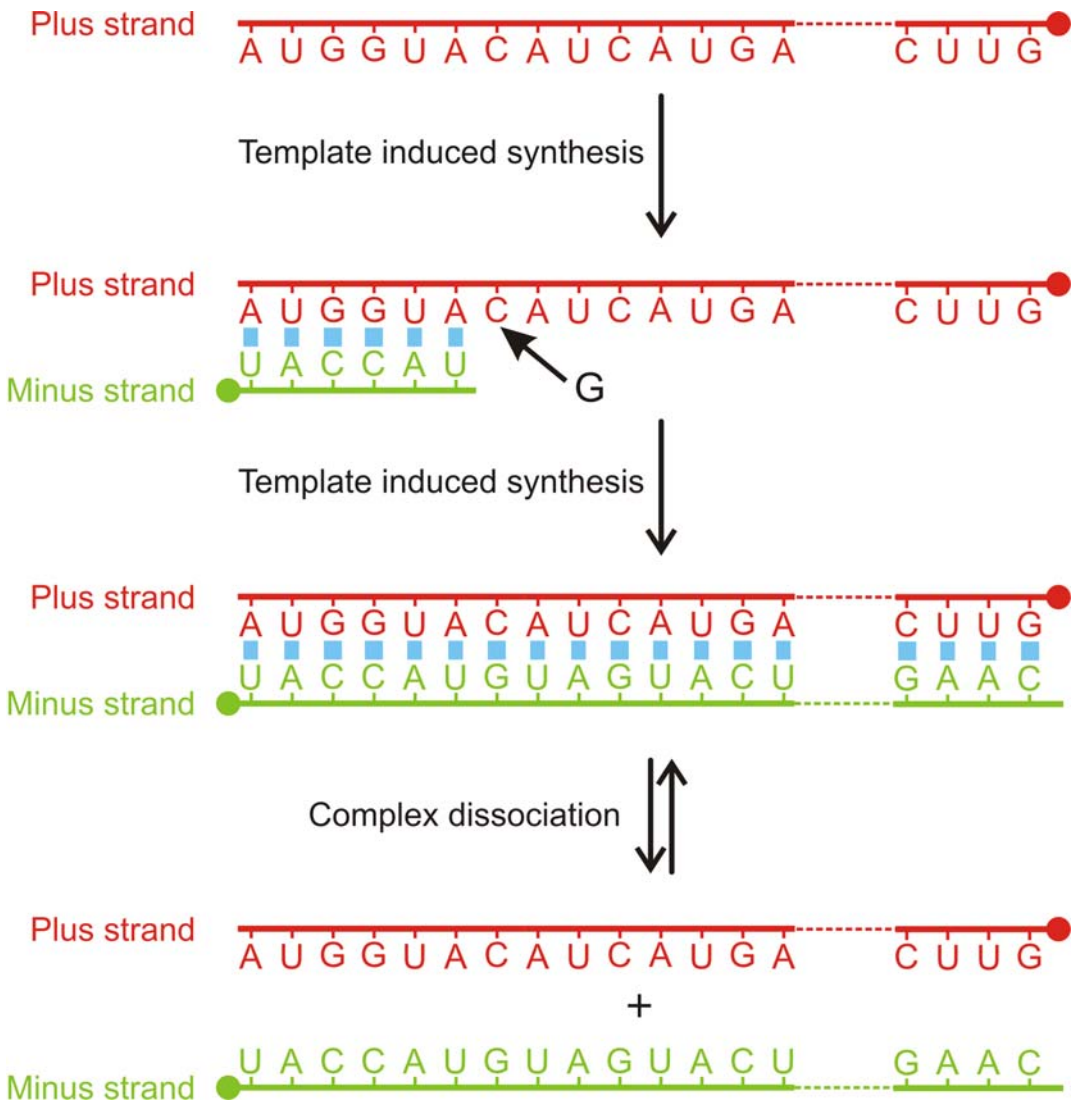


The flowreactor is a device for **studies** of evolution ***in vitro*** and ***in silico.***



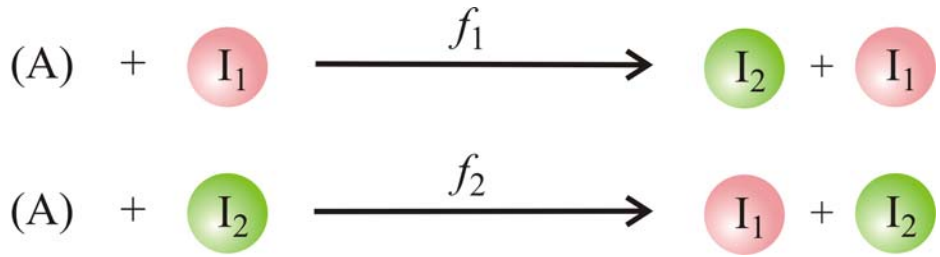
‚Replication fork' in DNA replication

The mechanism of DNA replication is ‚semi-conservative'



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

G≡C and **A=U**



$$\frac{dx_1}{dt} = f_2\, x_2 \quad \text{and} \quad \frac{dx_2}{dt} = f_1\, x_1$$

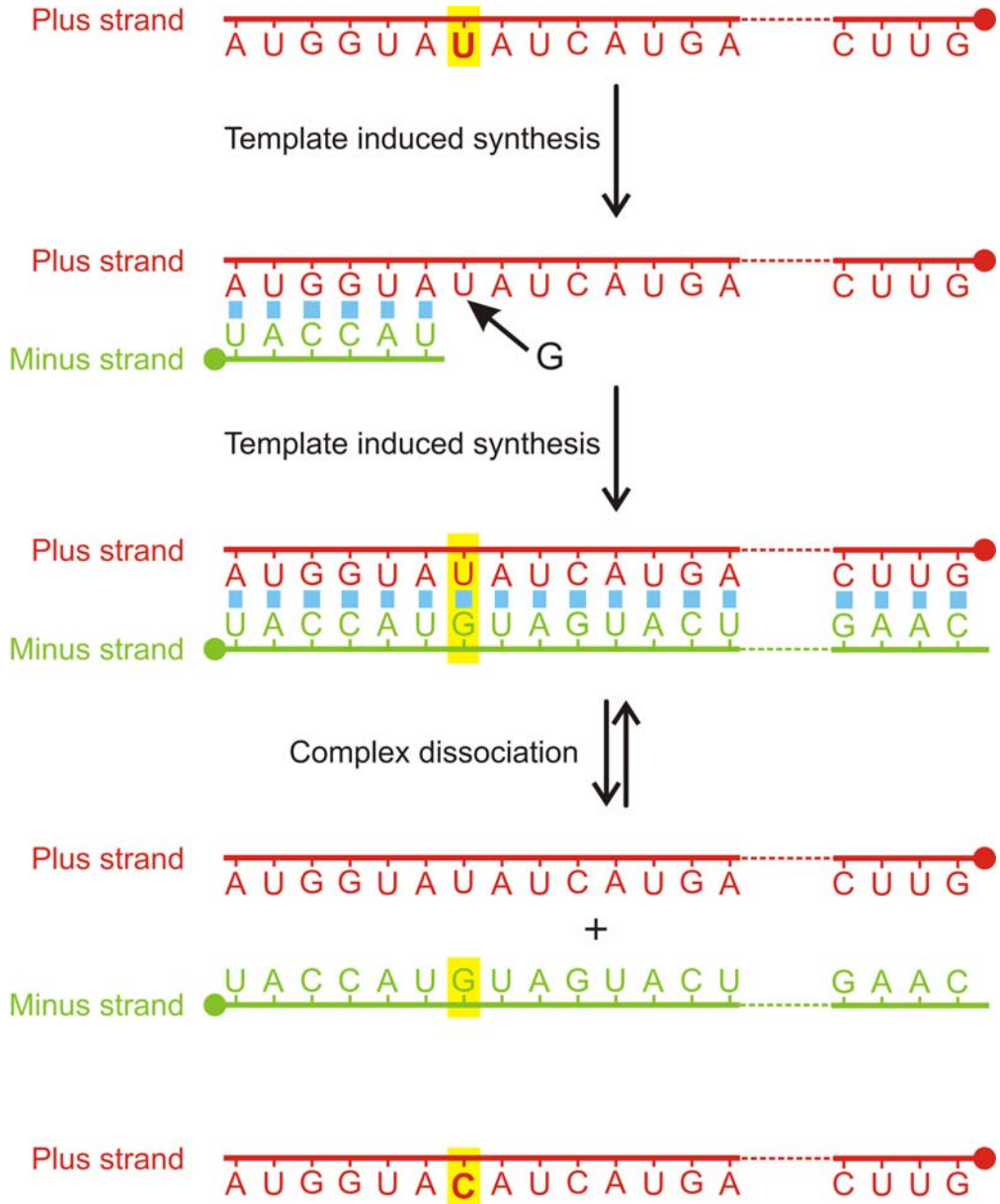
$$x_1 = \sqrt{f_2}\;\xi_1\,, \quad x_2 = \sqrt{f_1}\;\xi_2\,, \quad \zeta = \xi_1 + \xi_2\,, \quad \eta = \xi_1 - \xi_2\,, \quad f = \sqrt{f_1 f_2}$$

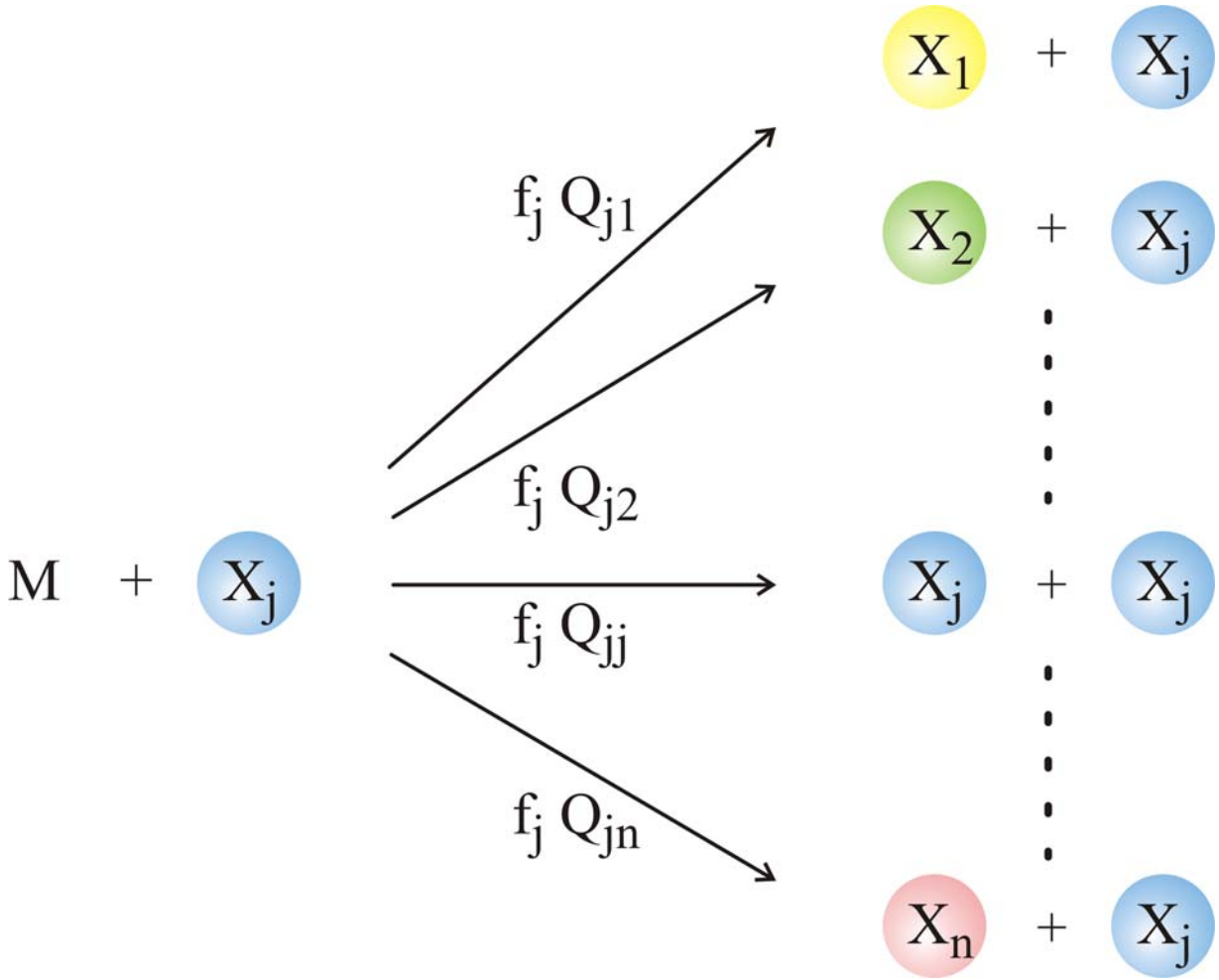
$$\eta(t) = \eta(0)\, e^{-ft}$$

$$\zeta(t) = \zeta(0)\, e^{ft}$$

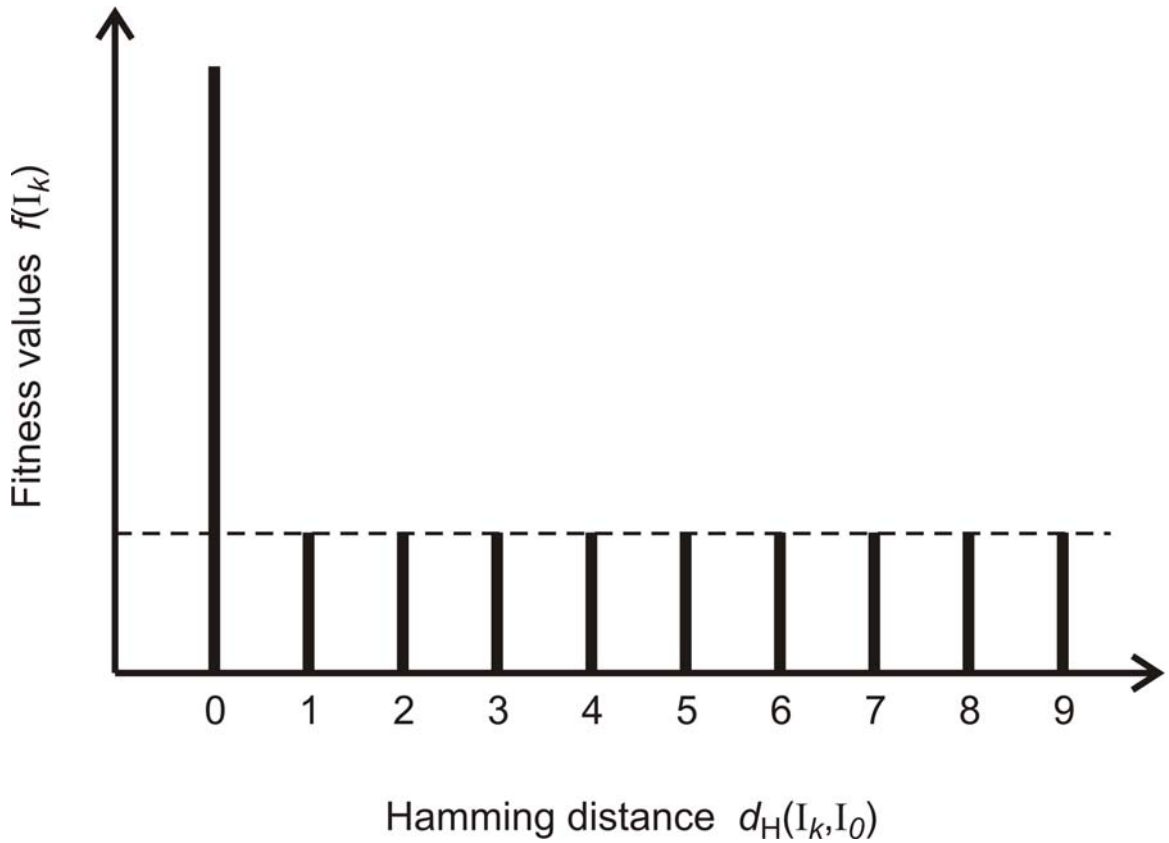
Complementary replication as the simplest molecular mechanism of reproduction

Mutation as an error
in replication





Chemical kinetics of replication and mutation as parallel reactions



A fitness landscape showing an error threshold

Biophysical Chemistry 16 (1982) 329–345
Elsevier Biomedical Press

329

SELF-REPLICATION WITH ERRORS

A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria

Received 4th June 1982
Revised manuscript received 23rd August 1982
Accepted 30th August 1982

Key words: Polynucleotide replication; Quasi-species; Point mutation; Mutant class; Stochastic replication

A model for polynucleotide replication is presented and analyzed by means of perturbation theory. Two basic assumptions allow handling of sequences up to a chain length of $\nu \approx 80$ explicitly: point mutations are restricted to a two-digit model and individual sequences are subsumed into mutant classes. Perturbation theory is in excellent agreement with the exact results for long enough sequences ($\nu > 20$).

1. Introduction

Eigen [8] proposed a formal kinetic equation (eq. 1) which describes self-replication under the constraint of constant total population size:

$$\frac{dx_i}{dt} = \dot{x}_i = \sum_j w_{ij}x_j - \frac{x_i}{c}\phi; \; i = 1,\ldots,n \; ^{\dagger} \qquad (1)$$

By x_i we denote the population number or concentration of the self-replicating element I_i, i.e., $x_i = [I_i]$. The total population size or total concentration $c = \sum_i x_i$ is kept constant by proper adjustment of the constraint ϕ: $\phi = \sum_i \sum_j w_{ij} x_j$. Characteristically, this constraint has been called 'constant organization'. The relative values of diagonal (w_{ii}) and off-diagonal (w_{ij}, $i \neq j$) rates, as we shall see in detail in section 2, are related to the accuracy of the replication process. The specific properties of eq. 1 are essentially based on the fact that it leads to exponential growth in the absence of constraints ($\phi = 0$) and competitors ($n = 1$).

The non-linear differential equation, eq. 1 – the non-linearity is introduced by the definition of ϕ at constant organization – shows a remarkable feature: it leads to selection of a defined ensemble of self-replicating elements above a certain accuracy threshold. This ensemble of a master and its most frequent mutants is a so-called 'quasi-species' [9]. Below this threshold, however, no selection takes place and the frequencies of the individual elements are determined exclusively by their statistical weights.

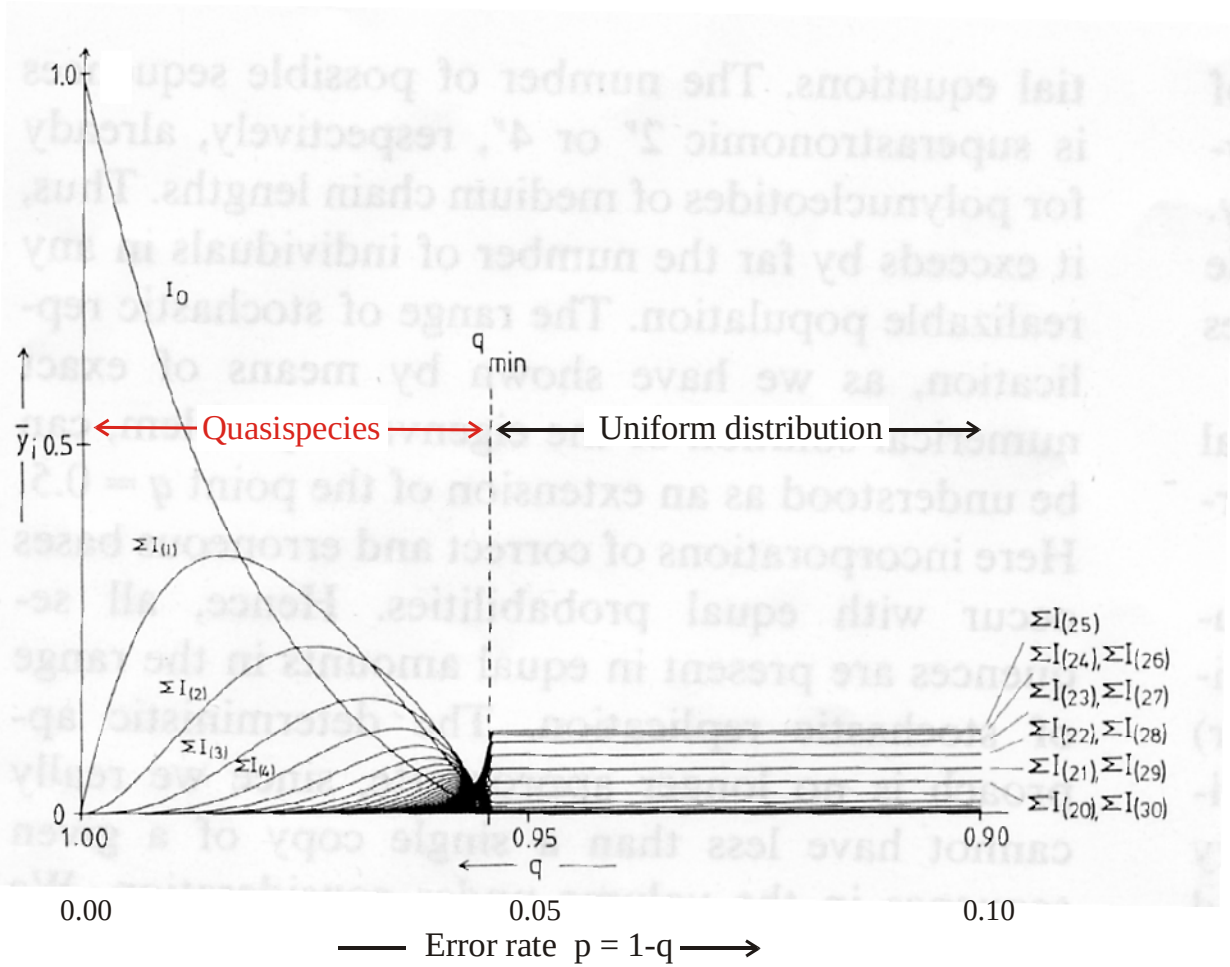
Rigorous mathematical analysis has been performed on eq. 1 [7,15,24,26]. In particular, it was shown that the non-linearity of eq. 1 can be removed by an appropriate transformation. The eigenvalue problem of the linear differential equation obtained thereby may be solved approximately by the conventional perturbation technique

* Dedicated to the late Professor B.L. Jones who was among the first to do rigorous mathematical analysis on the problems described here.

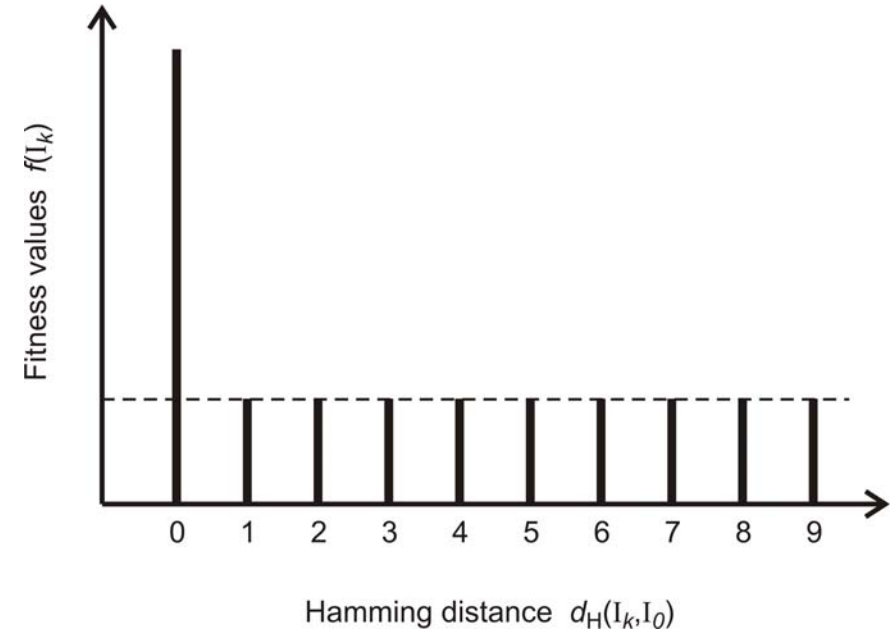
** This paper is considered as part II of Model Studies on RNA replication. Part I is by Gassner and Schuster [14].

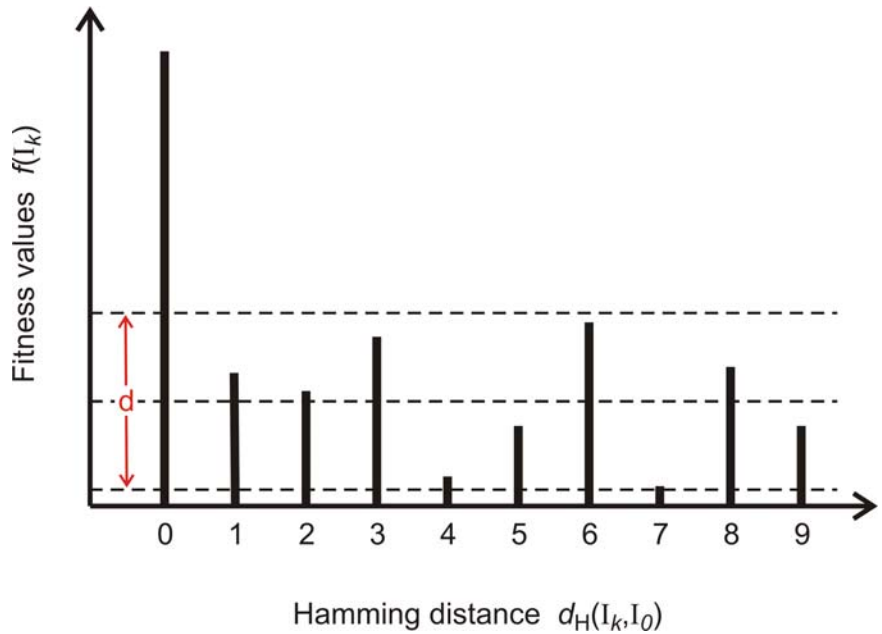
† All summations throughout this paper run from 1 to n unless specified differently: $\Sigma_i = \Sigma_{i=1}^{n}$ and $\Sigma_{i,i\neq j} = \Sigma_{i=1}^{j-1} + \Sigma_{i=j+1}^{n}$, respectively.

0301-4622/82/0000-0000/$02.75

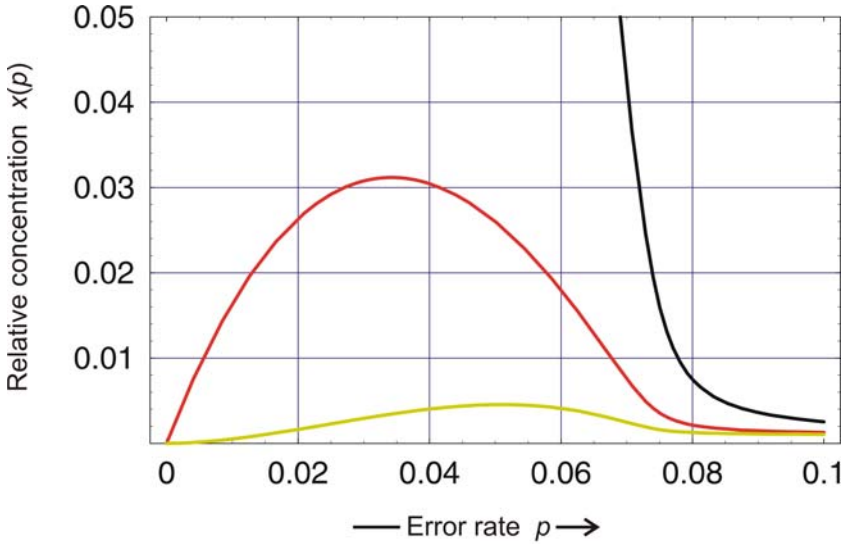


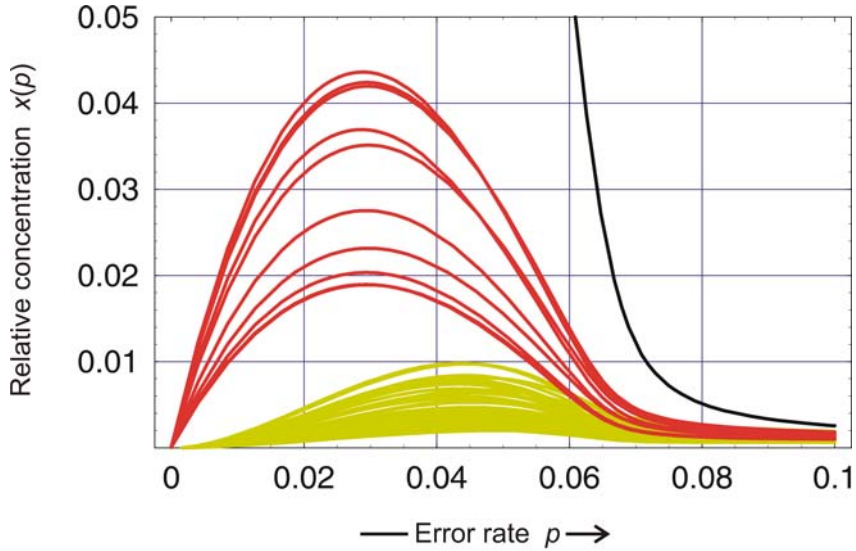
Stationary population or quasispecies as a function of the mutation or error rate *p*

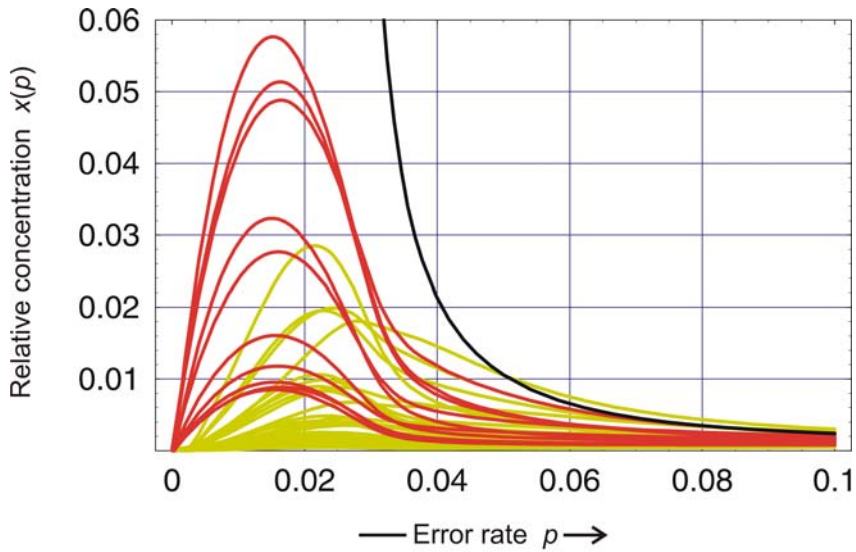




Fitness landscapes showing error thresholds

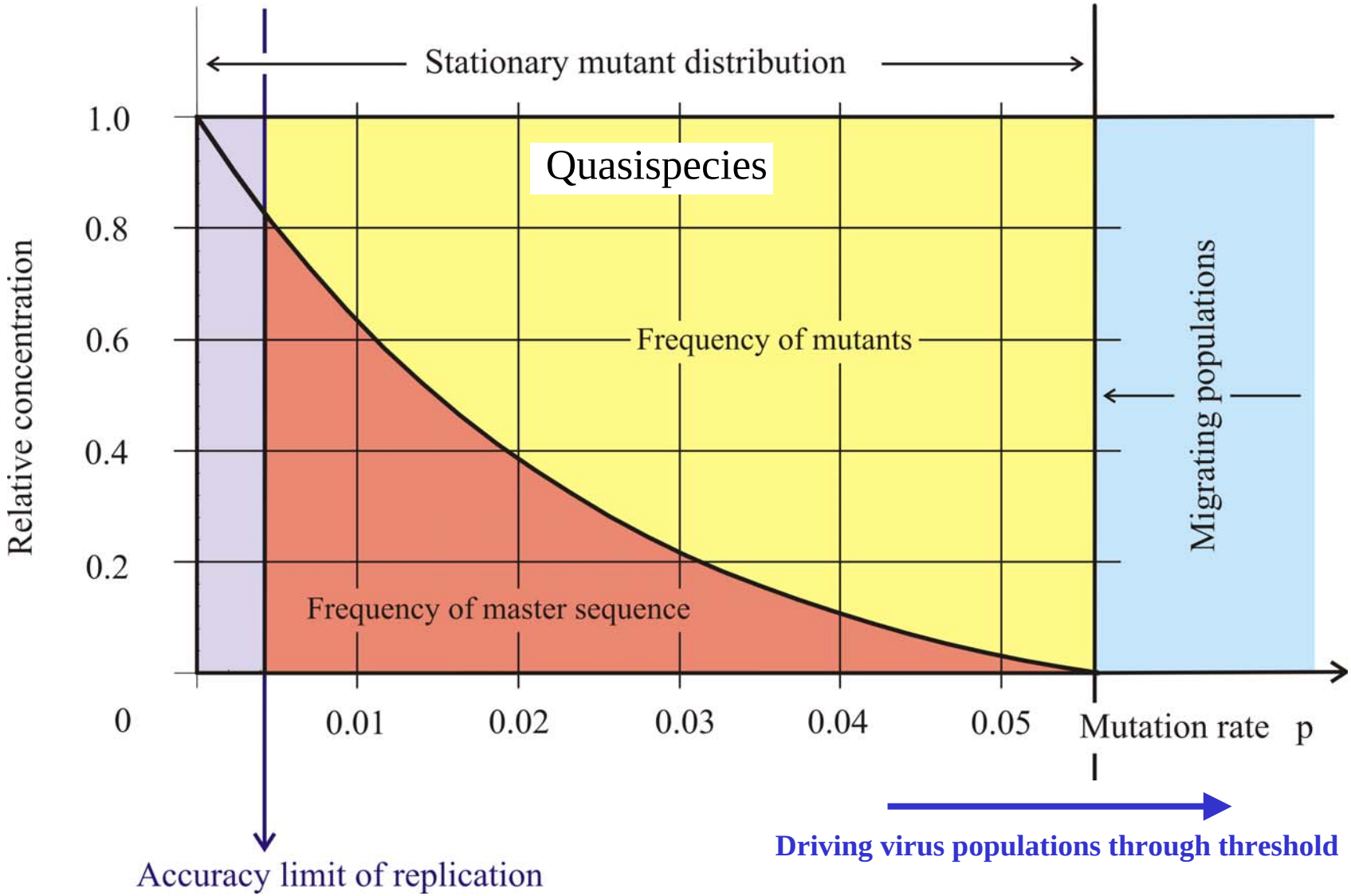






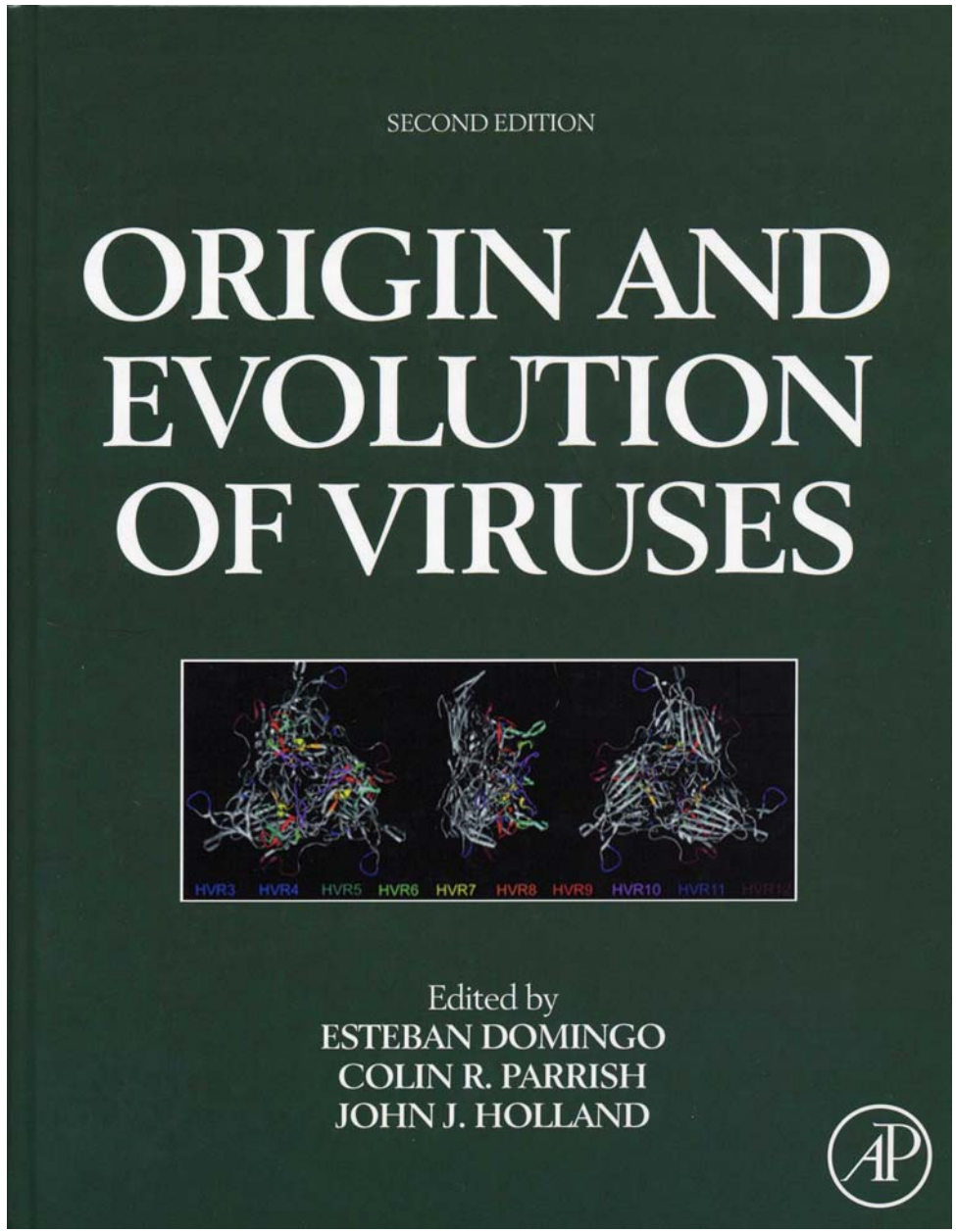
Error threshold: Individual sequences

n = 10, σ = 2 and d = 0, 1.0, 1.85



The error threshold in replication

Molecular evolution of viruses



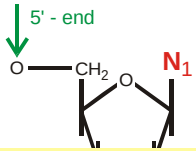
Results from the kinetic theory of molecular evolution:

• Replicating ensembles of molecules form stationary populations called *quasispecies*, which represent the genetic reservoir of asexually reproducing species.

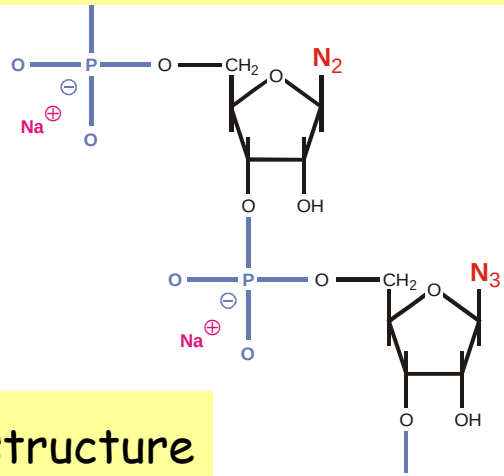
• For stable inheritance of genetic information mutation rates must not exceed a precisely defined and computable error-threshold.

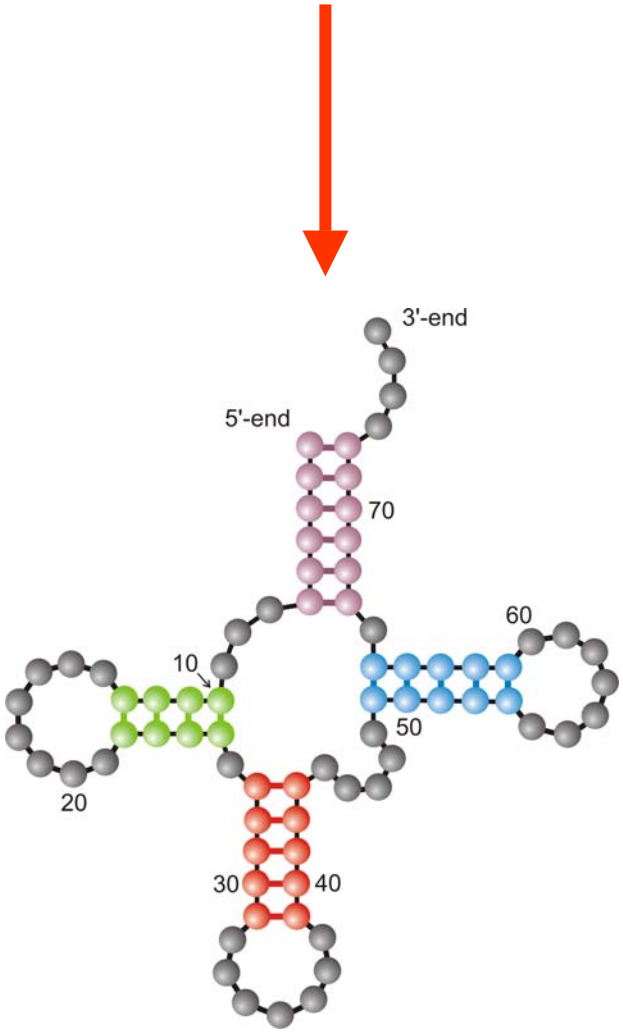
• The error-threshold can be exploited for the development of novel antiviral strategies.

1. Evolution – organismic and molecular
2. Multiplication, mutation, and selection
3. **Rational design of molecules**
4. Evolution and optimization of molecules
5. Origin of biological complexity
6. Biology and probabilities

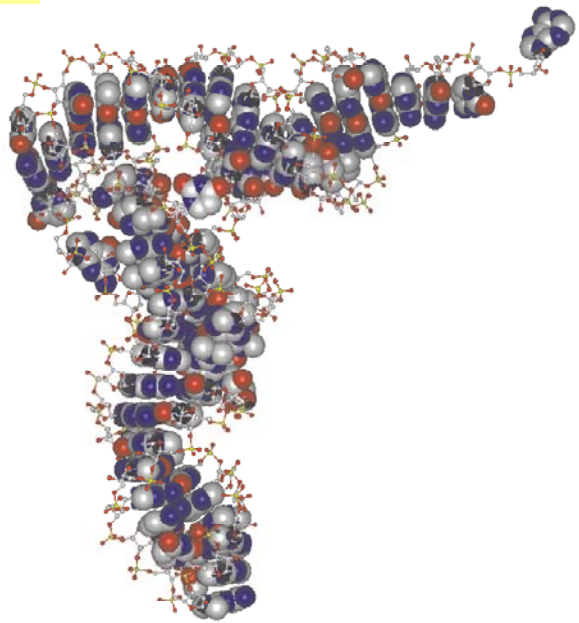


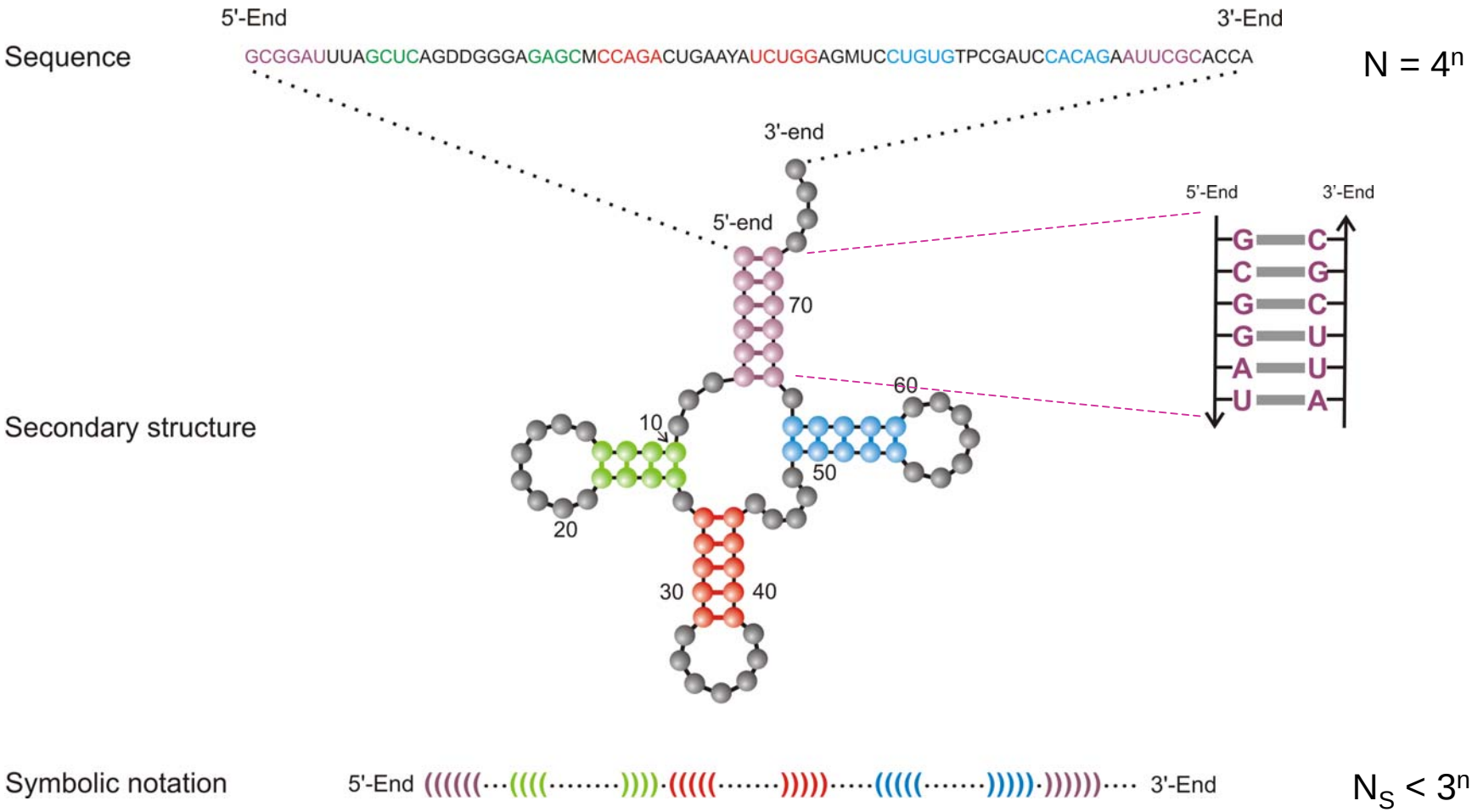
5'-end GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA 3'-end





Definition of RNA structure



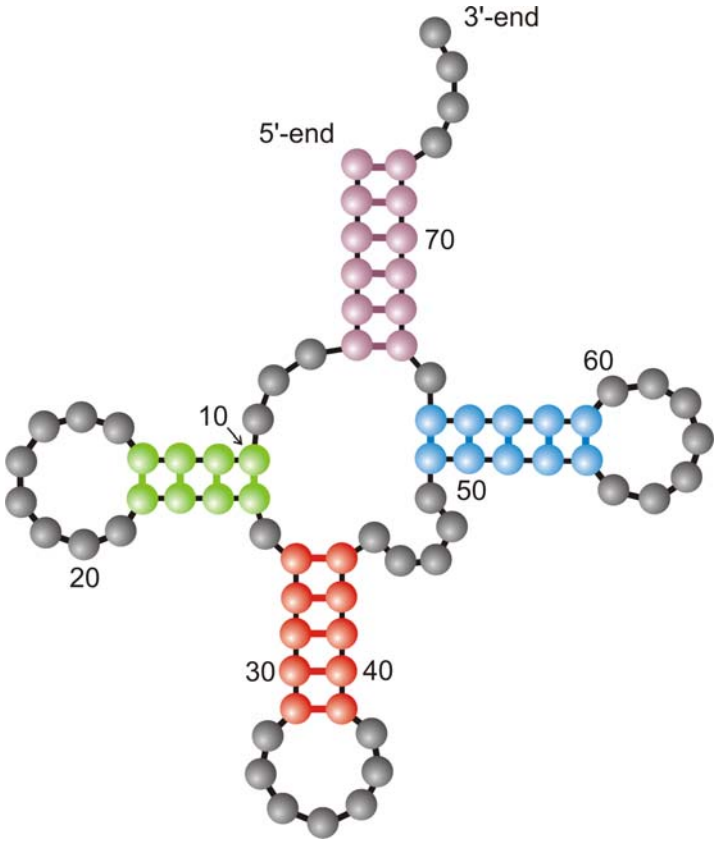


Criterion: Minimum free energy (mfe)

Rules: _ (_) _ ∈ {**AU**,**CG**,**GC**,**GU**,**UA**,**UG**}

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

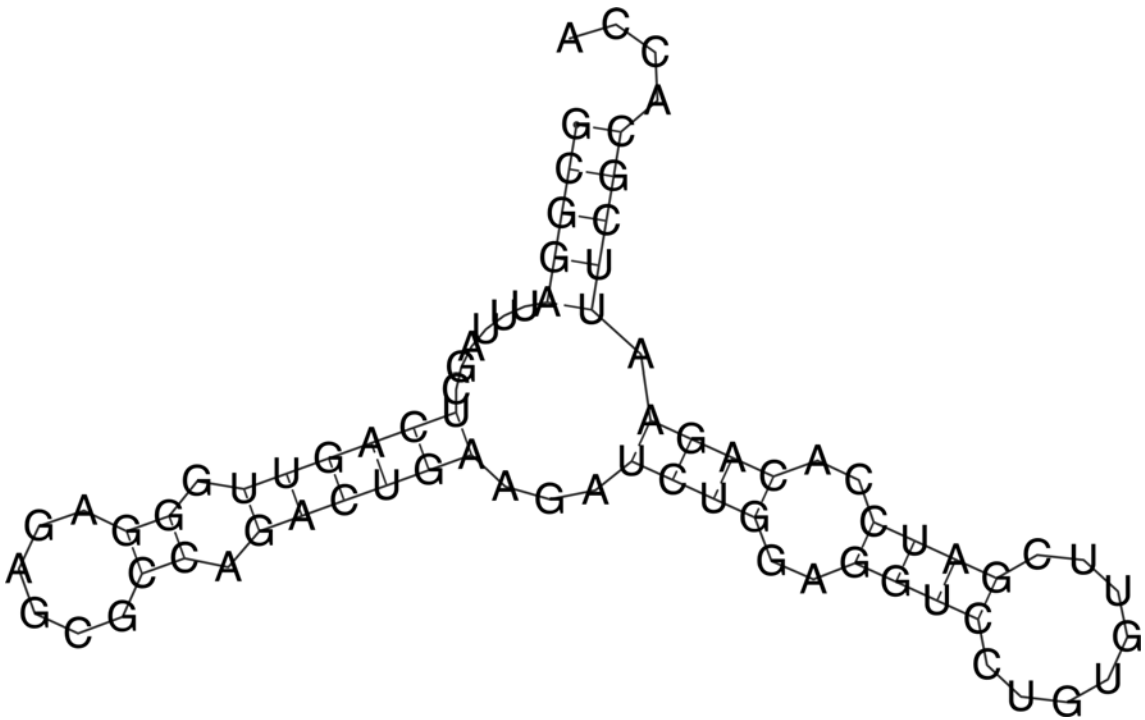
GCGGAUUUAGCUCAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCCUGUGTPCGAUCCACAGAAUUCGCACCA



ΔG = -20.20 kcal/mol

Sequence and structure of phenylalanyl-transfer-RNA

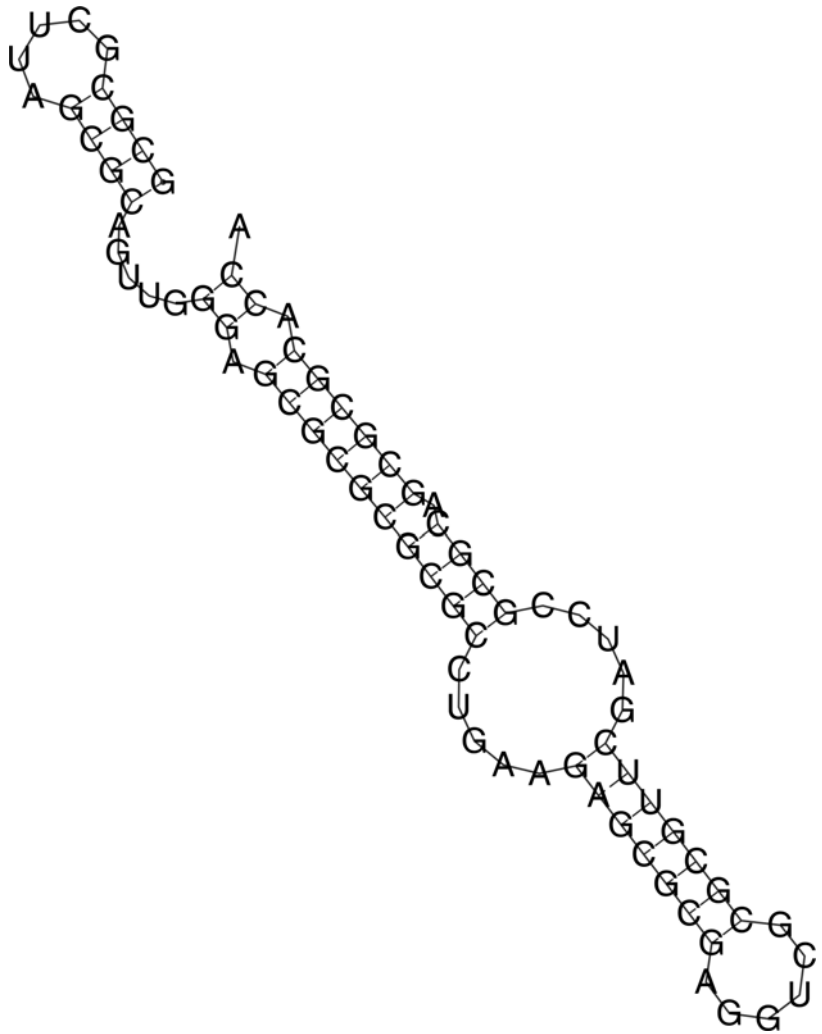
GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA



ΔG = -22.90 (-21.90) kcal/mol

GCGCGCUUAGCGCAGUUGGGAGCGCGCGCGCCUGAAGAGCGCGAGGUCGCGCGUUCGAUCCGCGCAGCGCGCACCA

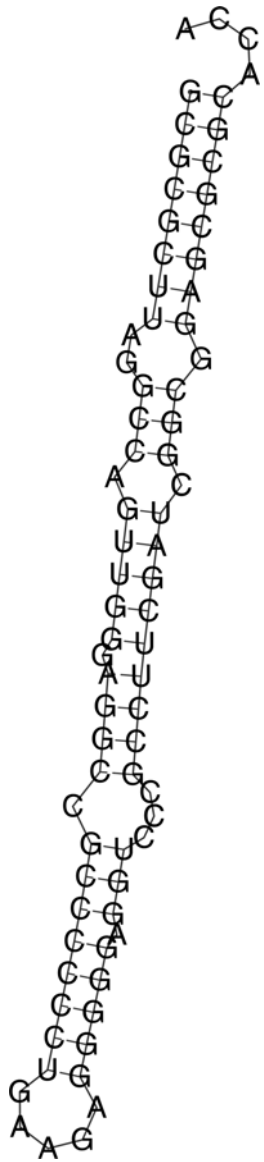
1. Trial



ΔG = -43.10 (-36.40) kcal/mol

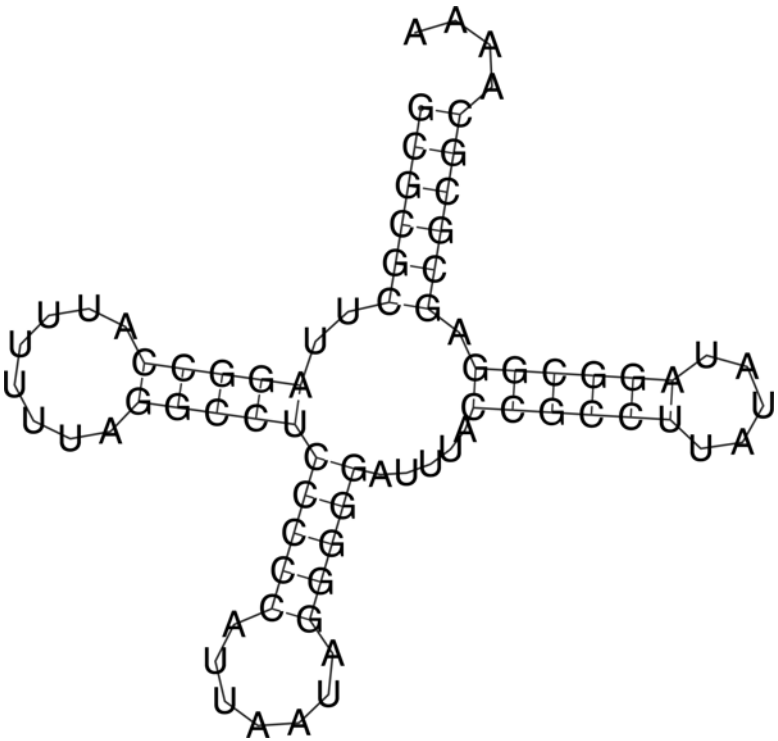
GCGCGCUUAGGCCAGUUGGGAGGCCGCCCCCCUGAAGAGGGGGAGGUCCCGCCUUCGAUCGGCGGAGCGCGCACCA

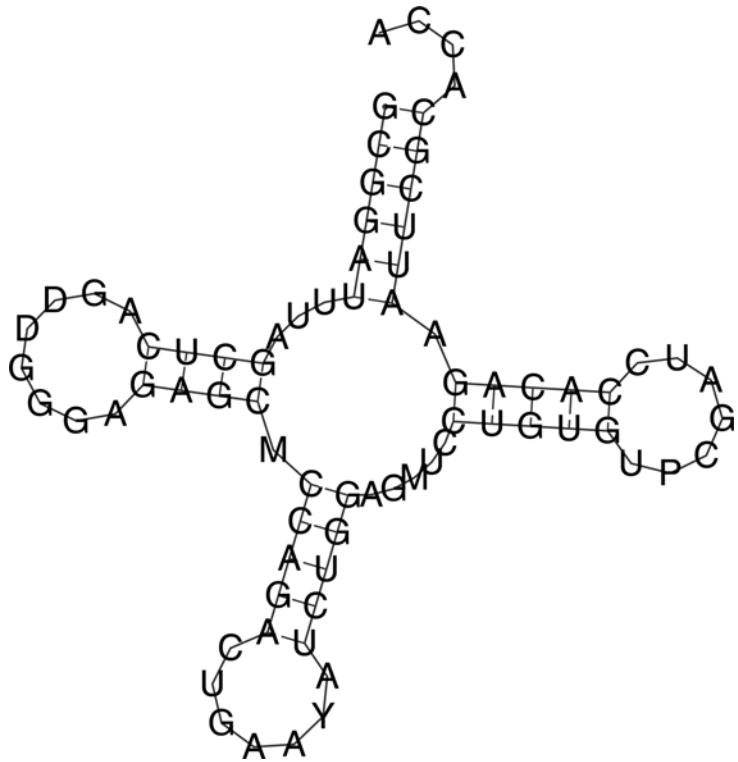
2. Trial



ΔG = -45.10 (-39.40) kcal/mol

GCGCGCUUAGGCCAUUUUUUAGGCCUCCCCCAUUAAUAGGGGGAUUUACCGCCUUAUAUAGGCGGAGCGCGCAAAA



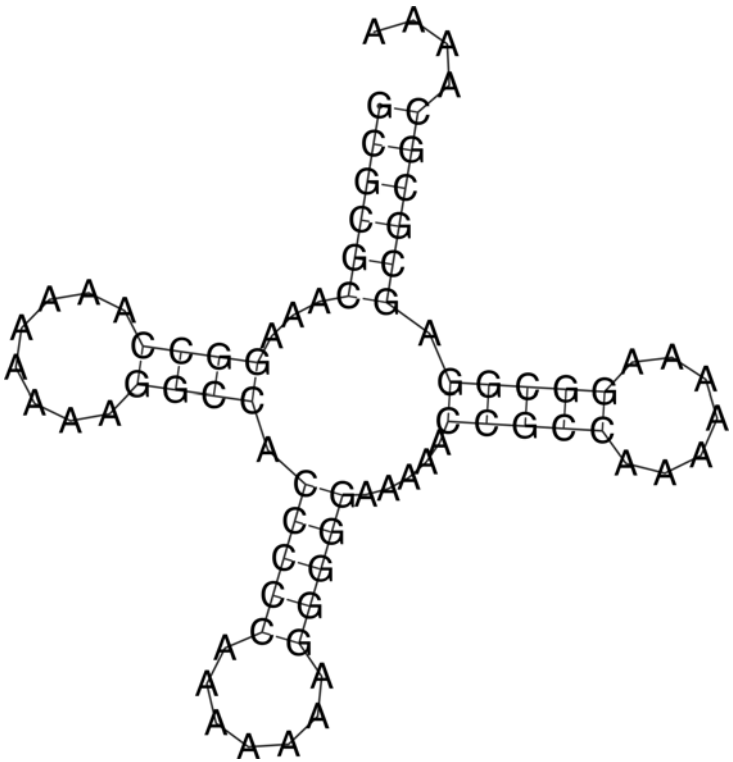


3. Trial

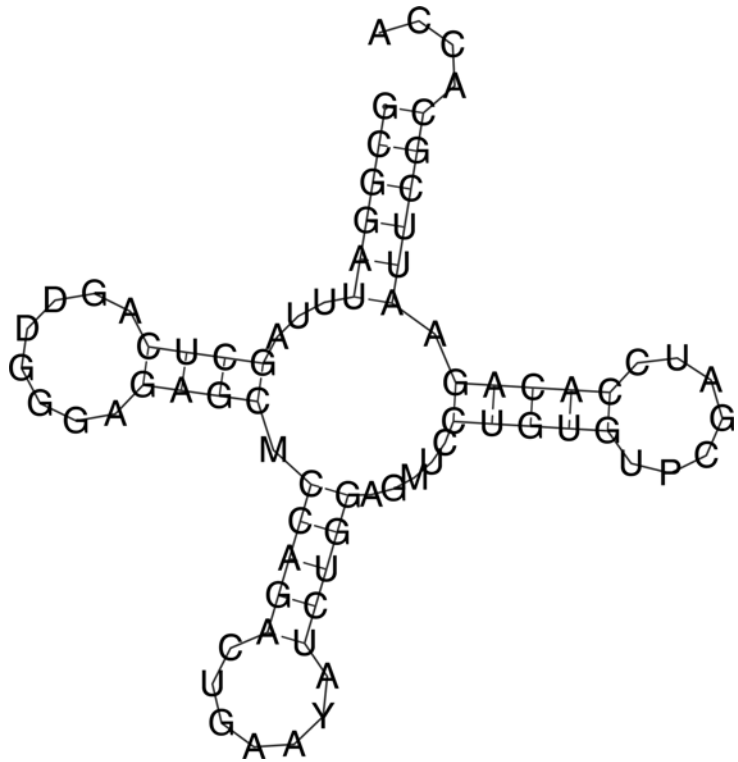
Target structure

ΔG = -41.80 (-39.90) kcal/mol

GCGCGCAAAGGCCAAAAAAAAGGCCACCCCCAAAAAAAGGGGGAAAAACCGCCAAAAAAAGGCGGAGCGCGCAAAA



4. Trial



Target structure

ΔG = -40.70 kcal/mol

RNA sequence: GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA folding:

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

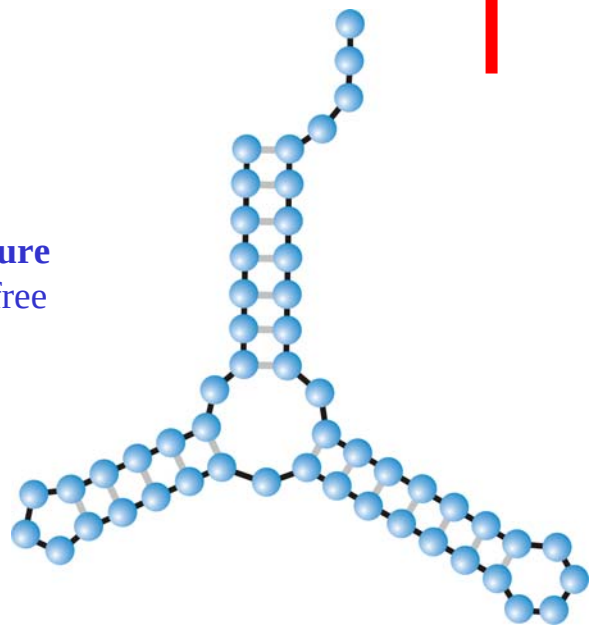
Iterative determination of a sequence for the given secondary structure

Inverse Folding Algorithm

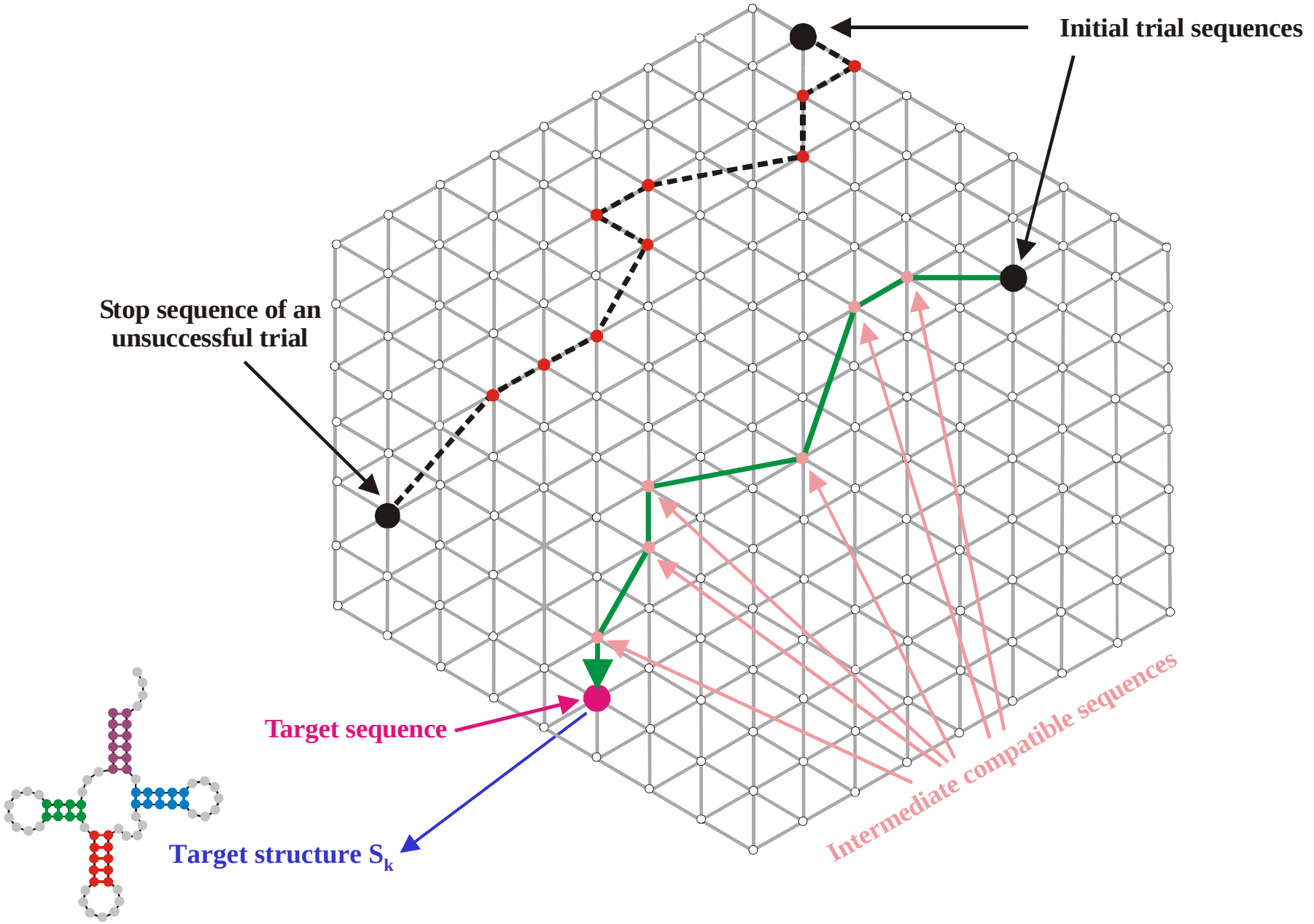
Inverse folding of RNA:

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

RNA structure of minimal free energy:



Sequence, structure, and design



Approach to the target structure S_k in the **inverse folding algorithm**

INSTITUTE OF PHYSICS PUBLISHING REPORTS ON PROGRESS IN PHYSICS

Rep. Prog. Phys. **69** (2006) 1419–1477 doi:10.1088/0034-4885/69/5/R04

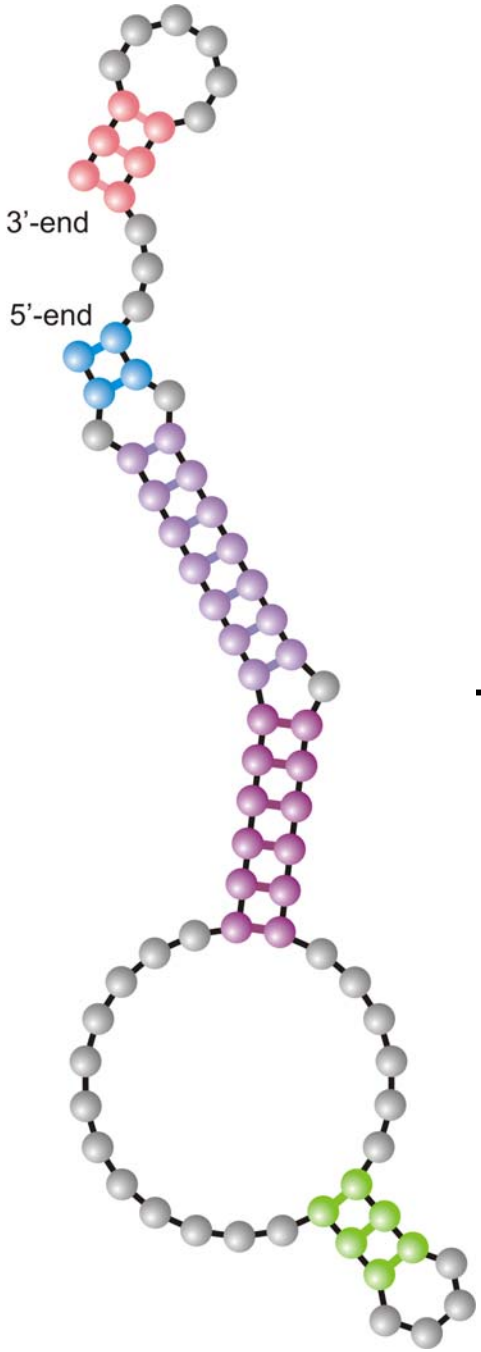
Prediction of RNA secondary structures: from theory to models and real molecules

Peter Schuster[1,2]

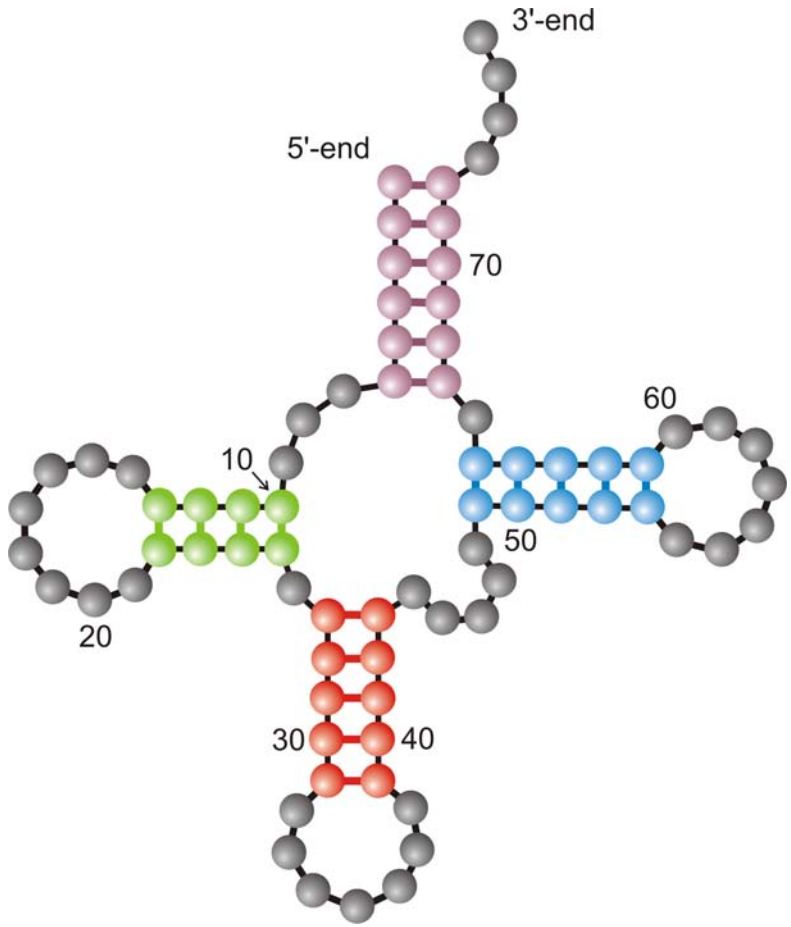
[1]Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria
[2]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

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

1. Evolution – organismic and molecular
2. Multiplication, mutation, and selection
3. Rational design of molecules
4. **Evolution and optimization of molecules**
5. Origin of biological complexity
6. Biology and probabilities



Structure of andomly chosen initial sequence



Phenylalanyl-tRNA as target structure

Evolution *in silico*

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

random individuals. The primer pair used for genomic DNA amplification is 5′-TCTCCCTGGATTCT-CATTTA-3′ (forward) and 5′-TCTTTGTCTTCTGT-TCCACC-3′ (reverse). Reactions were performed in 25 μl using 1 unit of Taq DNA polymerase with each primer at 0.4 μM; 200 μM each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM tris-HCl (pH 8.3), 50 mM KCl_2,1.5 mM $MgCl_2$] 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)$^+$ 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, *MRDD 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. Fridell, 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 labyrinths) 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)$^+$ 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′-GCATGACCTGCCGGCTAAT-GGG-3′; reverse, 5′-CTCACGGCTTCTGCATGGT-GCTCGGCTGGC-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. Fergusson, A. Gupta, E. Sorbello, R. Torkzadeh, 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 T. Barber, S. Sullivan, E. Green, D. Drayna, and J. 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.C.M.), the National Institute of Child Health and Human Development (R01 HD30428 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.

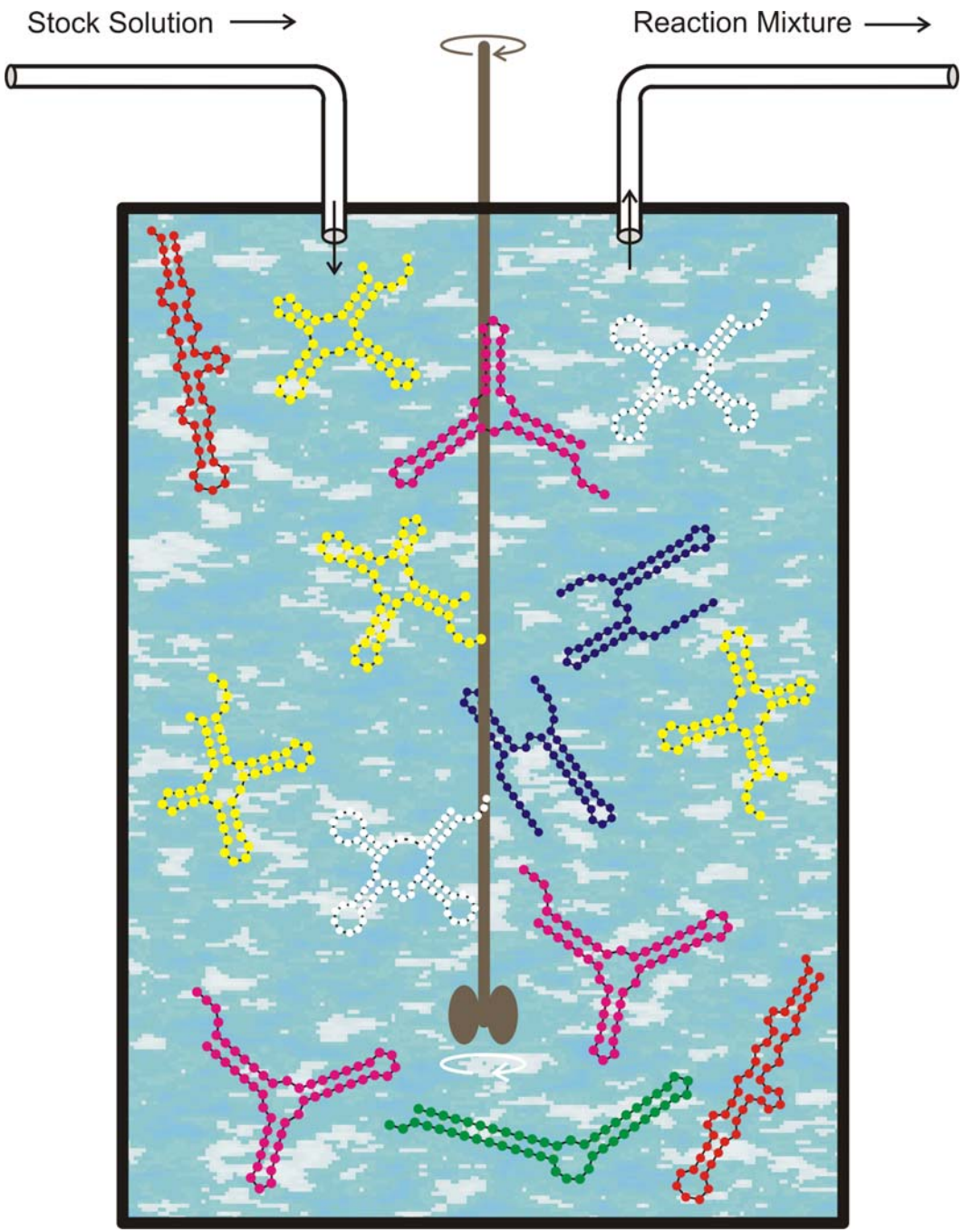
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level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

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:

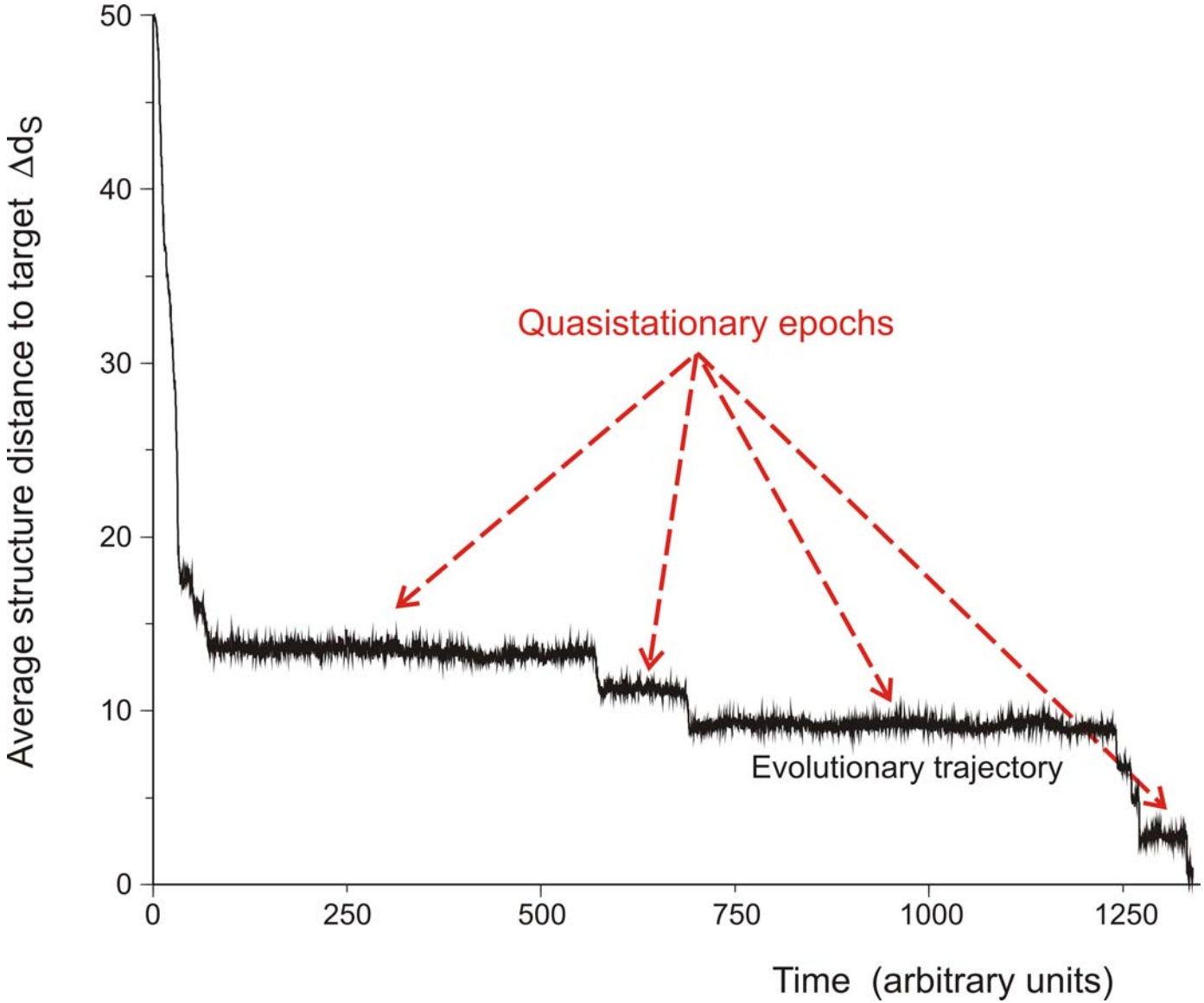
Population size, N = # RNA molecules, is controlled by the flow

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

Mutation rate:

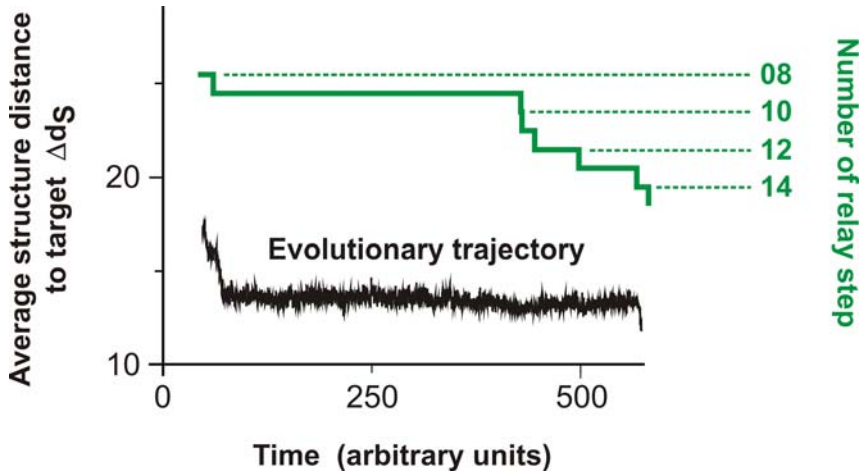
p = 0.001 / site × replication

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



In silico optimization in the flow reactor: Evolutionary Trajectory

28 neutral point mutations during a long quasi-stationary epoch



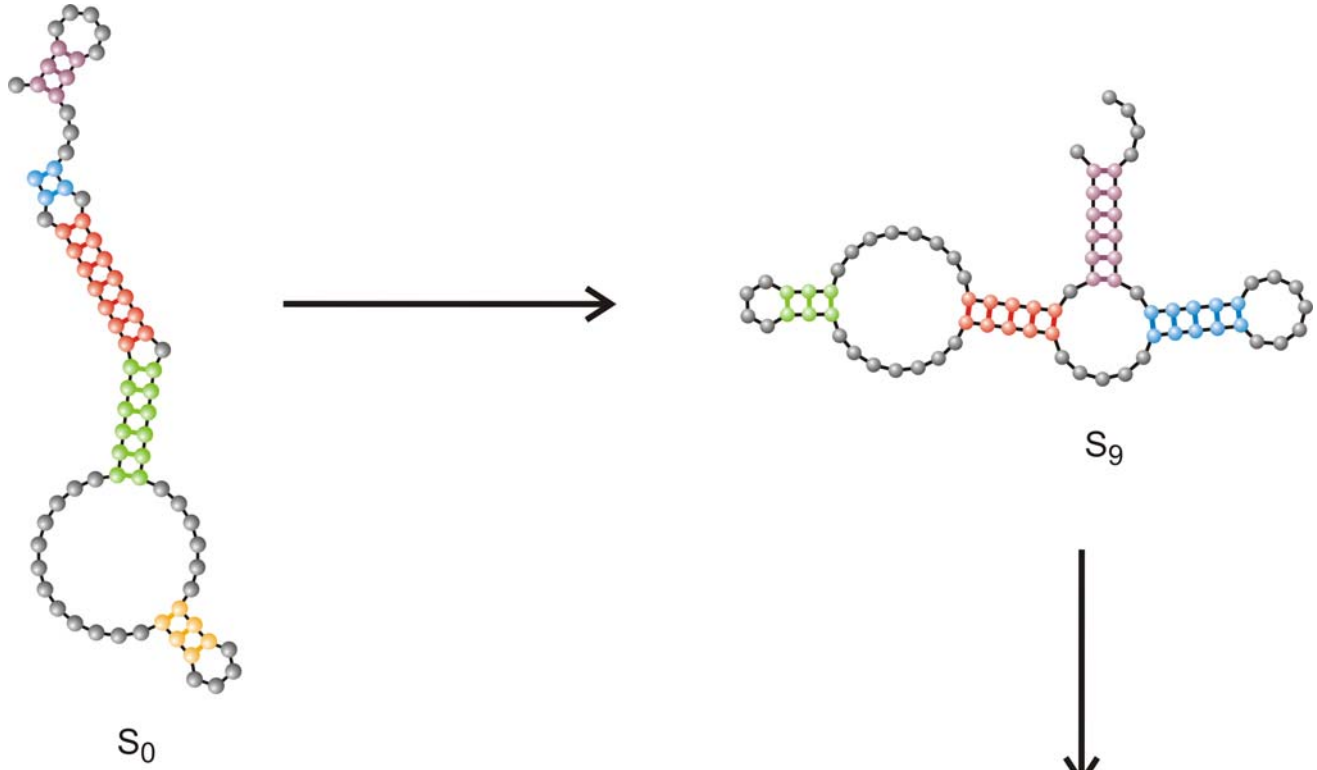

```
entry   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8       .((((((((((((.........(((....)))).......)))))).....((((.......))))))))))....
exit    GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCAUACAGAA
entry   GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUACCAUACAGAA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

Transition inducing point mutations change the molecular structure

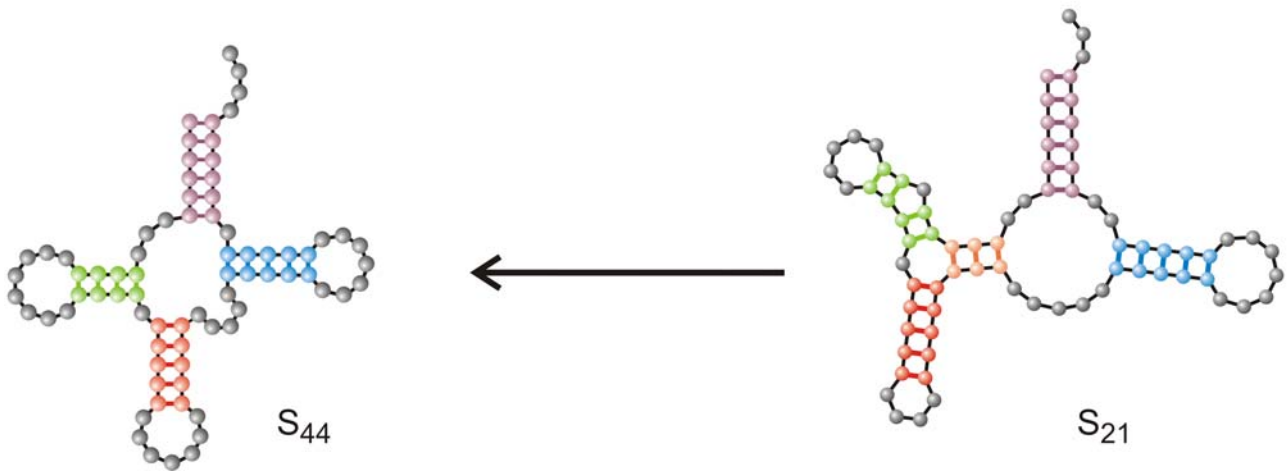
Neutral point mutations leave the molecular structure unchanged

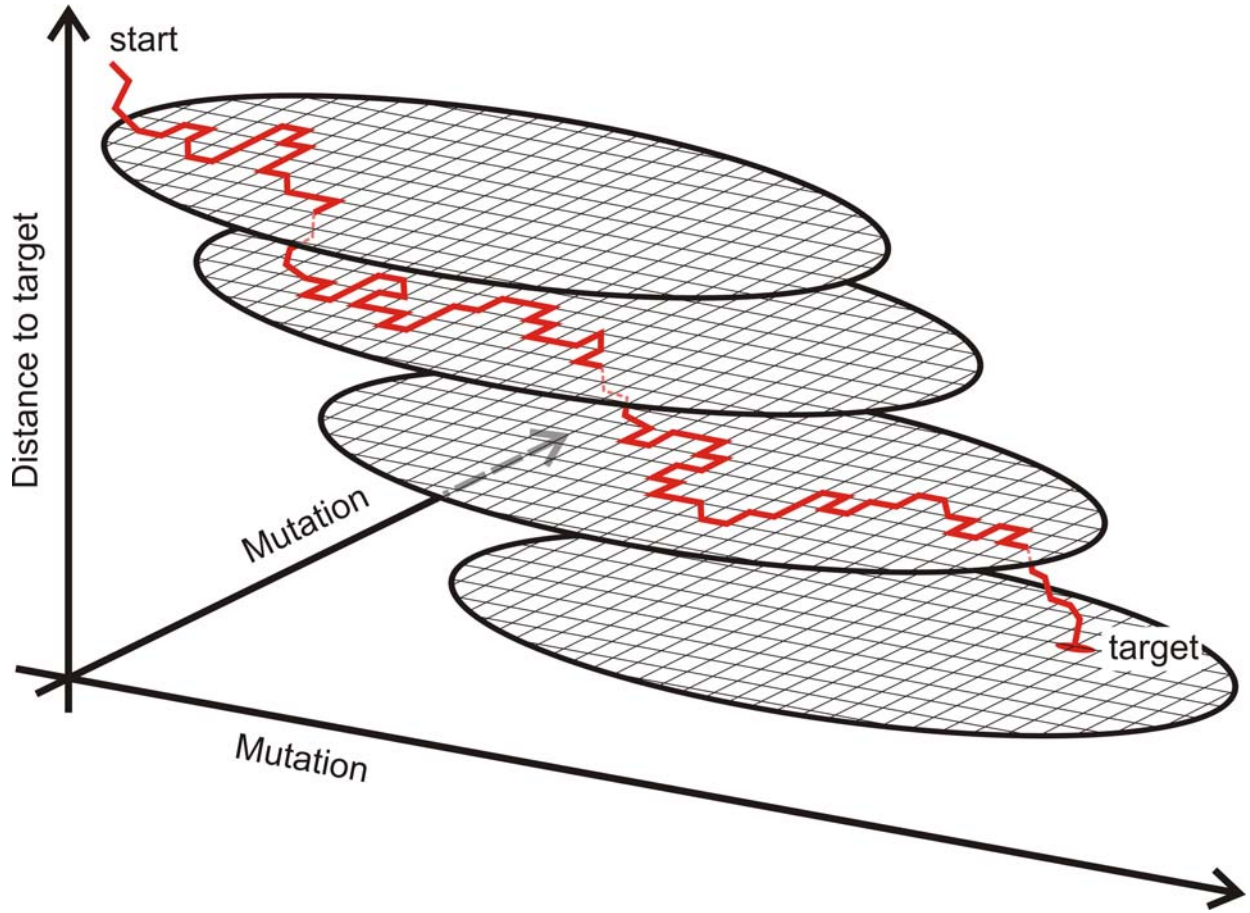
Neutral genotype evolution during phenotypic stasis

Randomly chosen
initial structure



Phenylalanyl-tRNA
as target structure





A sketch of optimization on neutral networks

Evolution of RNA molecules based on Qβ 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

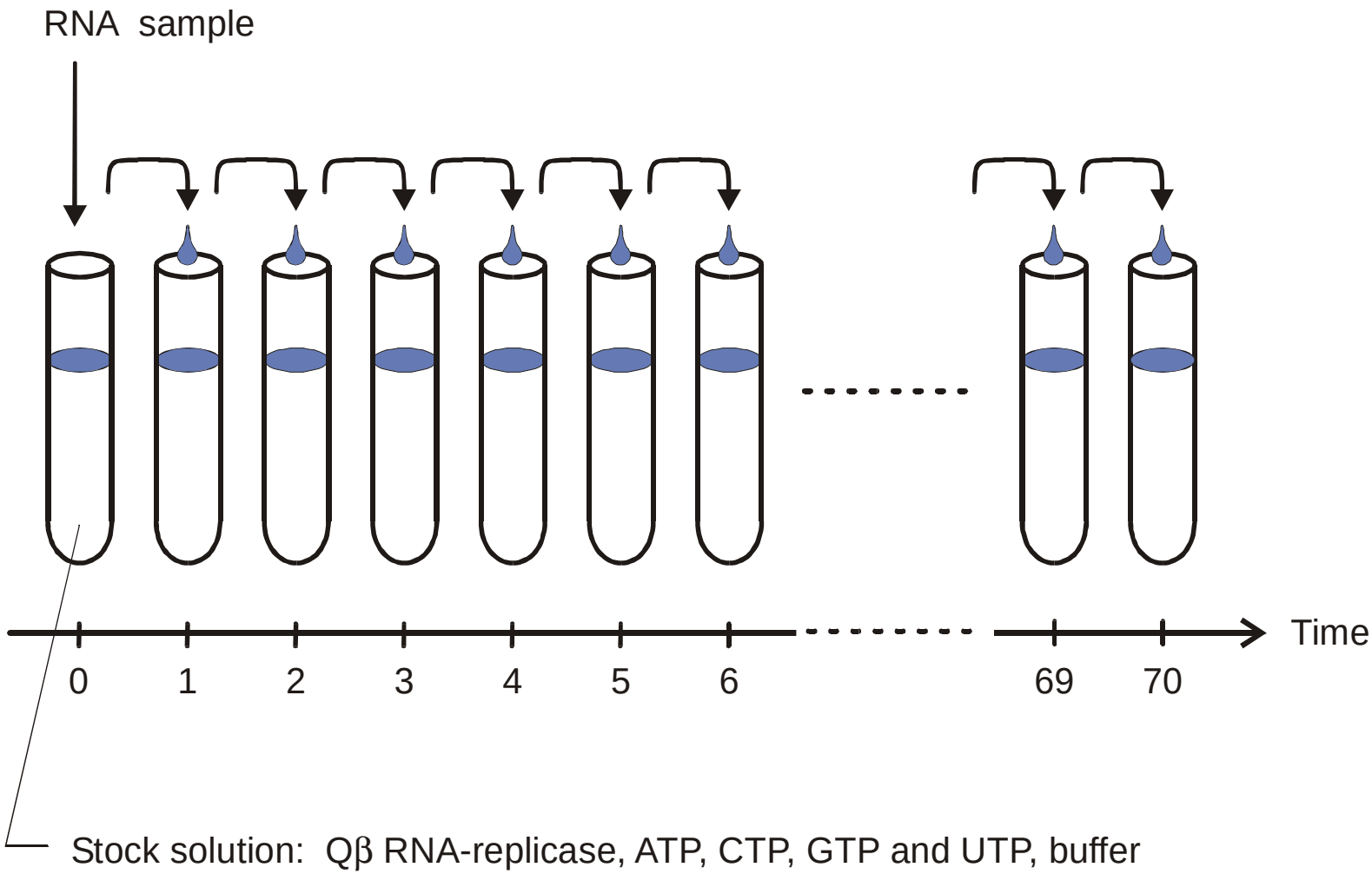
C.K.Biebricher, ***Darwinian selection of self-replicating RNA molecules***. Evolutionary Biology **16** (1983), 1-52

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.Öhlenschlager, 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



Anwendung der seriellen Überimpfungstechnik auf RNA-Evolution in Reagenzglas

Evolutionary design of RNA molecules

A.D. Ellington, 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

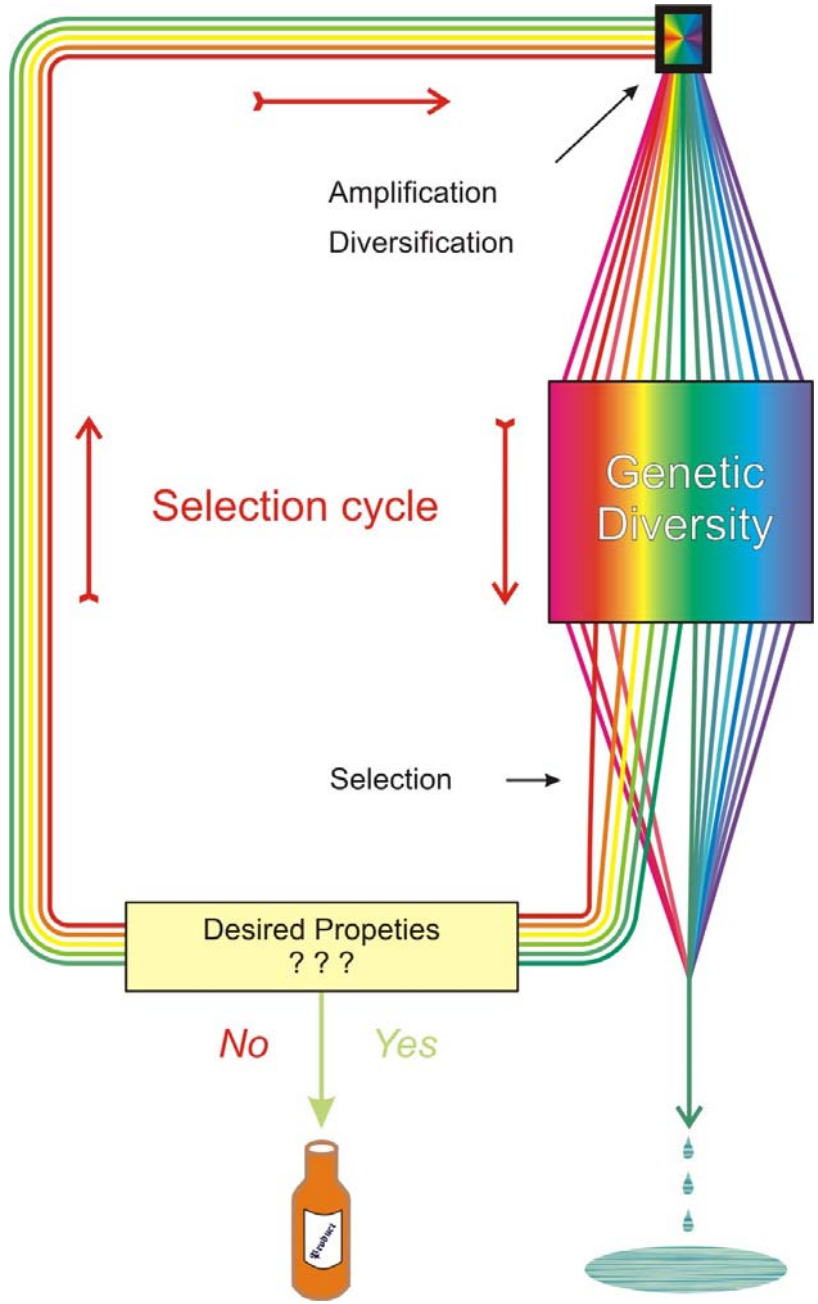
D.P. Bartel, J.W. Szostak, ***Isolation of new ribozymes from a large pool of random sequences***. Science **261** (1993), 1411-1418

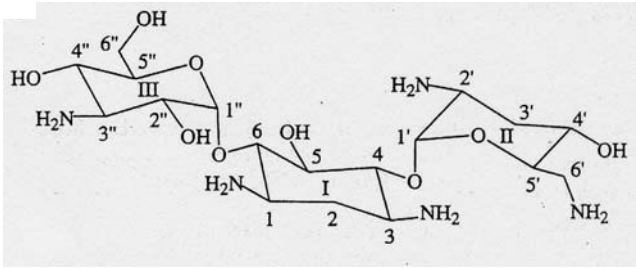
R.D. Jenison, S.C. Gill, A. Pardi, B. Poliski, ***High-resolution molecular discrimination by RNA***. Science **263** (1994), 1425-1429

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

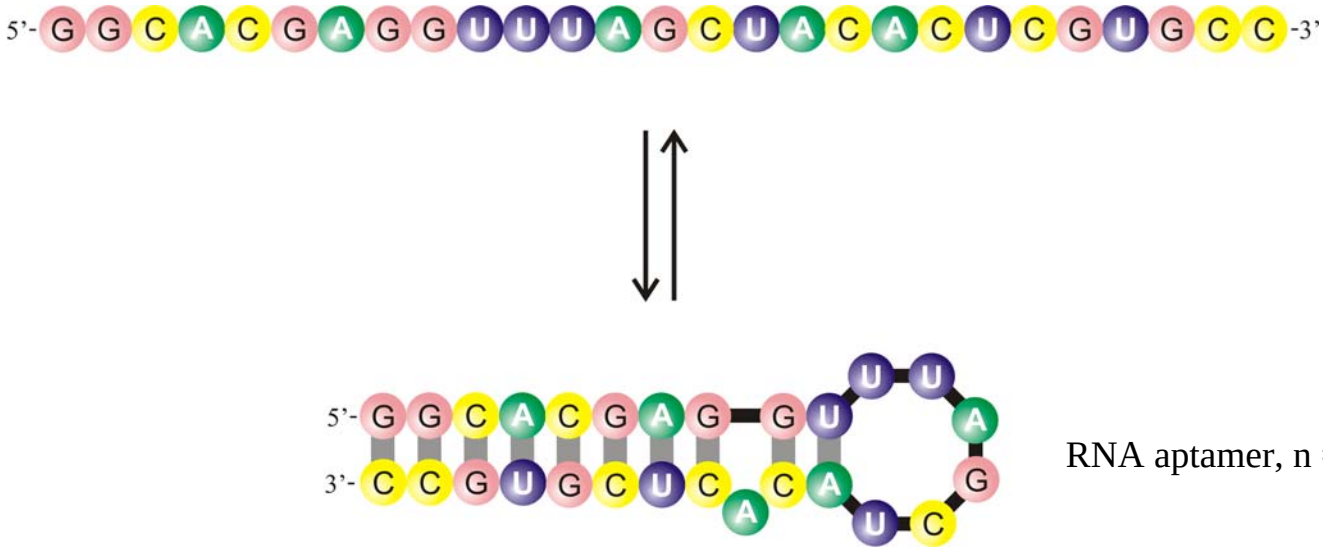
L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex***. Chemistry & Biology **4** (1997), 35-50

An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules





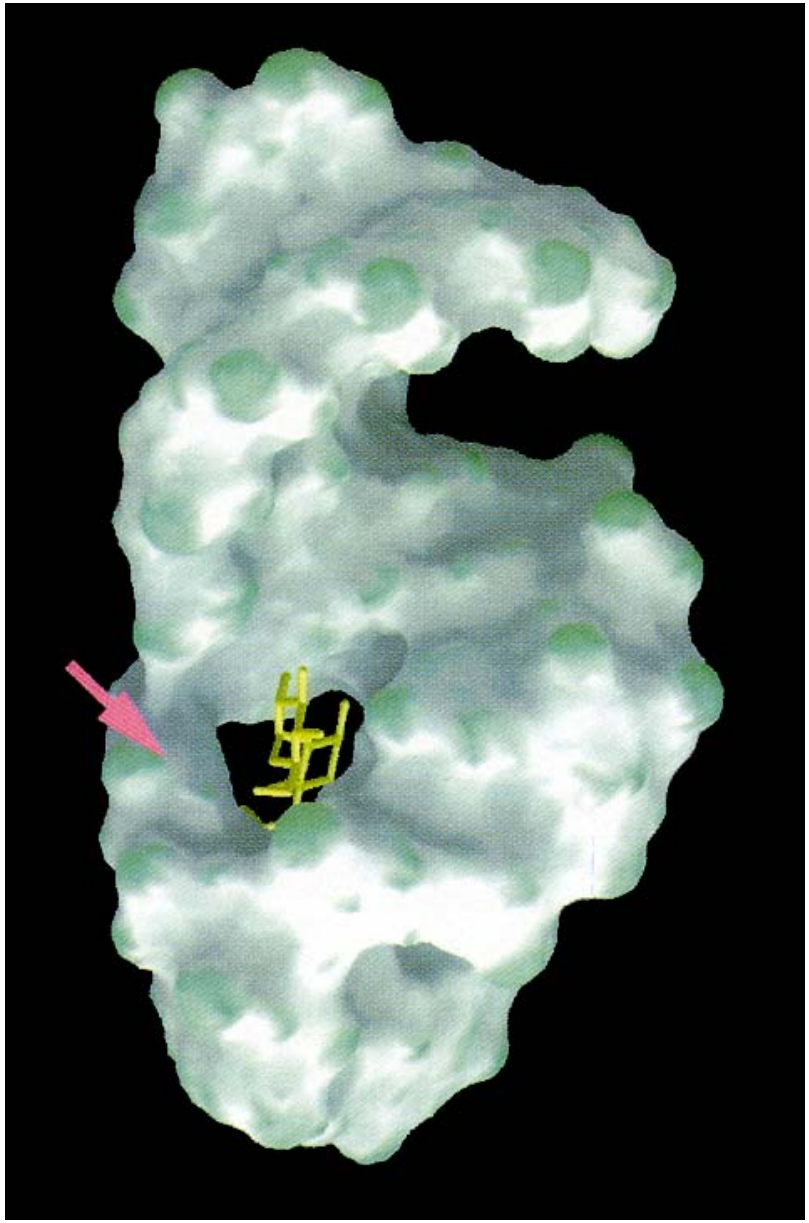
tobramycin



RNA aptamer, n = 27

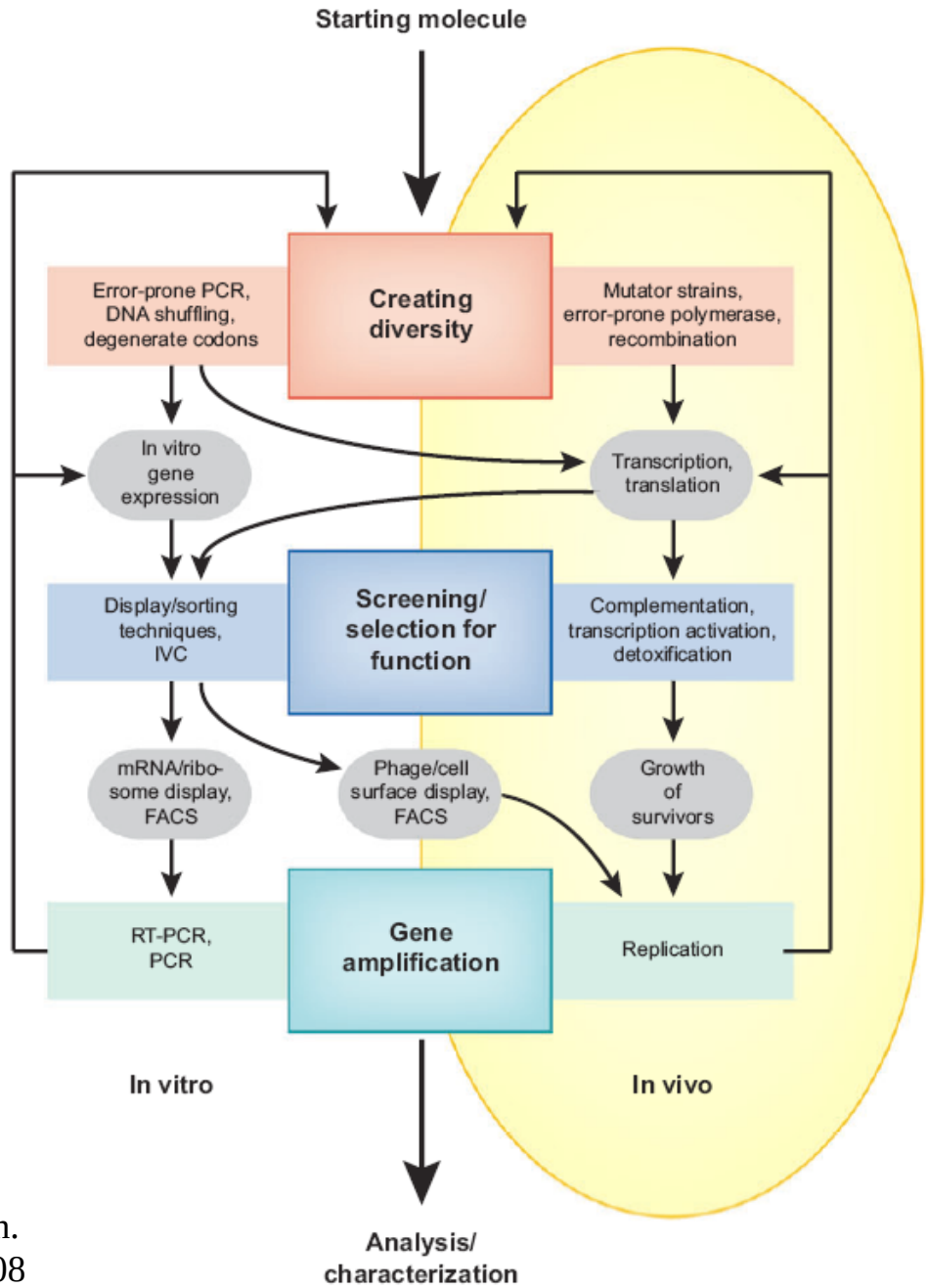
Formation of secondary structure of the tobramycin binding RNA aptamer with **K_D = 9 nM**

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)



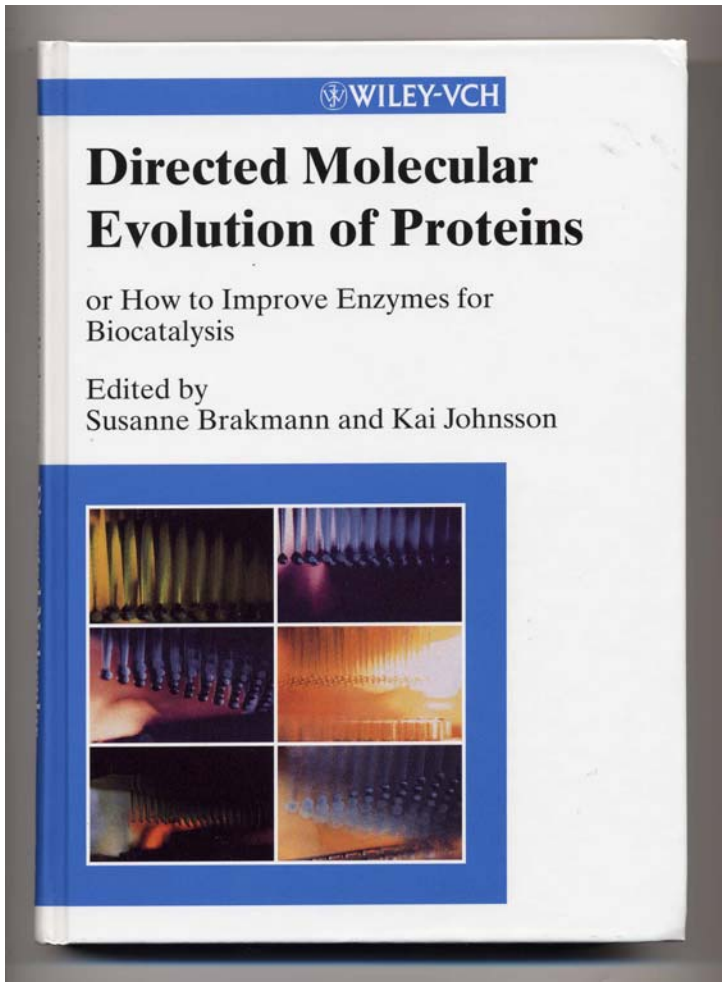
The three-dimensional structure of the tobramycin aptamer complex

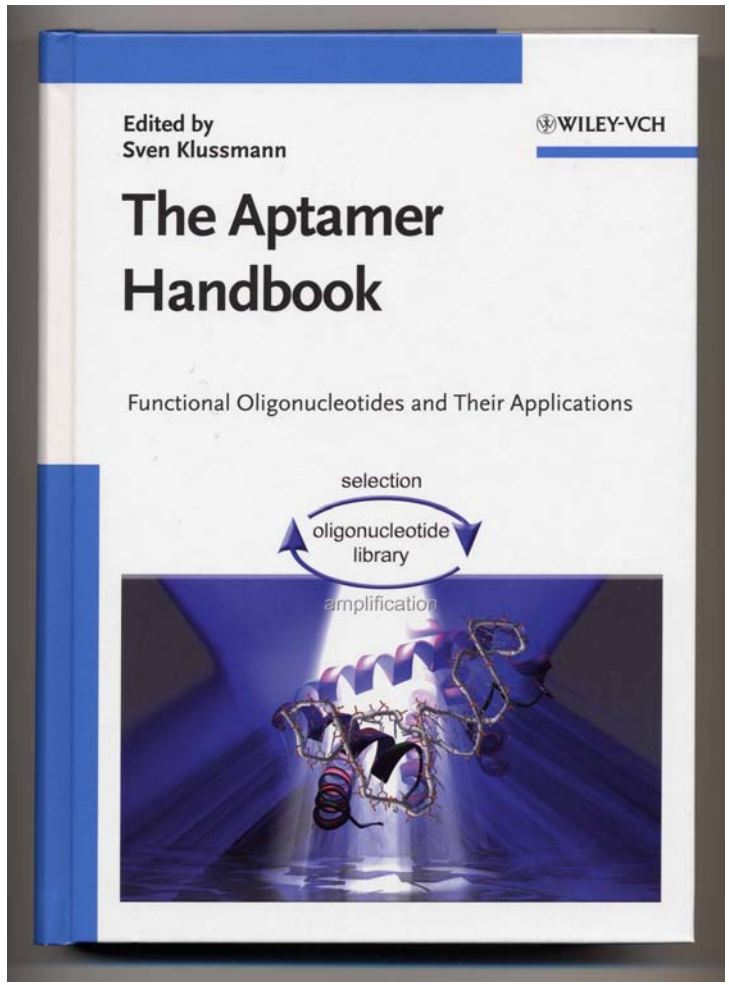
L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, Chemistry & Biology **4**:35-50 (1997)



Schematic overview of the principal processes, strategies, and techniques of directed evolution. Today, numerous experimental methods are available to perform the fundamental processes of true Darwinian evolution (*central boxes*) in the laboratory, either in vivo within microorganisms or entirely in vitro in the test tube. Arrows indicate possible routes for connecting individual evolutionary steps. Abbreviations: PCR, polymerase chain reaction; RT-PCR, reverse transcription PCR; IVC, in vitro compartmentalization; FACS, fluorescence-activated cell sorting.

Christian Jäckel, Peter Kast, and Donald Hilvert.
Protein design by directed evolution.
Annu.Rev.Biophys. **37**:153-173, 2008





Application of molecular evolution to problems in biotechnology

Artificial evolution in biotechnology and pharmacology

G.F. Joyce. 2004. Directed evolution of nucleic acid enzymes. *Annu.Rev.Biochem*. **73**:791-836.

C. Jäckel, P. Kast, and D. Hilvert. 2008. Protein design by directed evolution. *Annu.Rev.Biophys*. **37**:153-173.

S.J. Wrenn and P.B. Harbury. 2007. Chemical evolution as a tool for molecular discovery. *Annu.Rev.Biochem*. **76**:331-349.

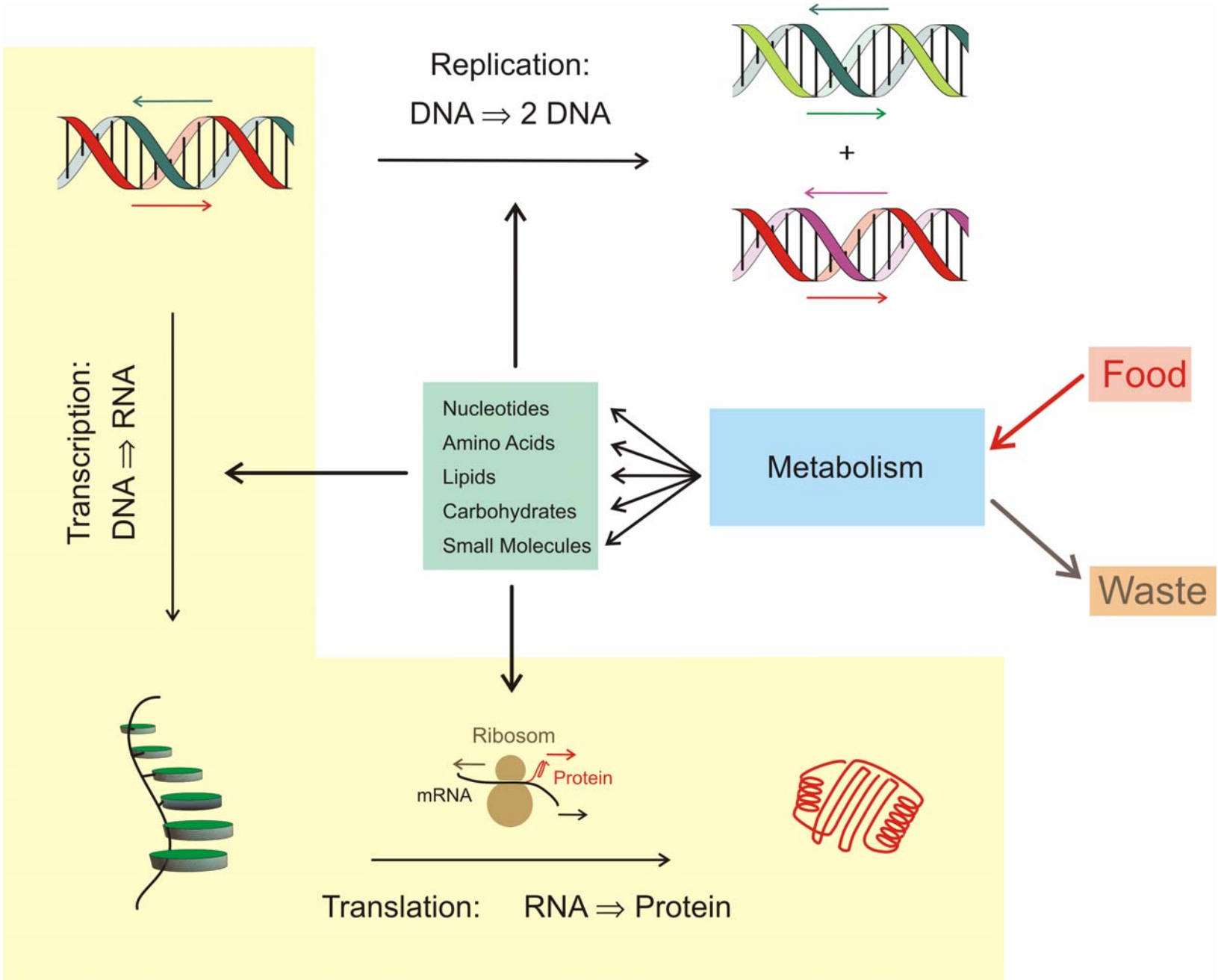
Results from laboratory experiments in molecular evolution:

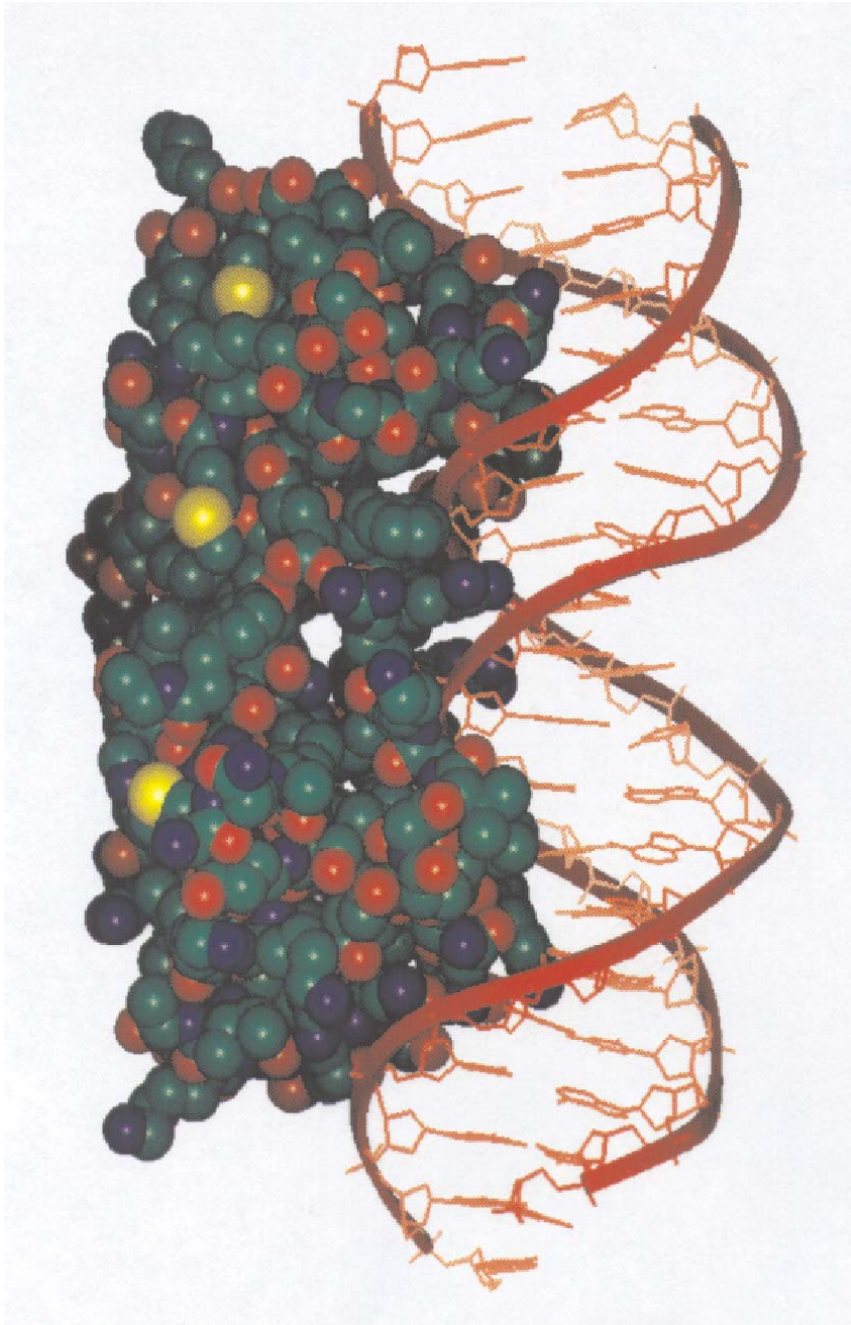
- Evolutionary optimization does not require cells and occurs in molecular systems too.

- *In vitro* evolution allows for production of molecules for predefined purposes and gave rise to a branch of biotechnology.

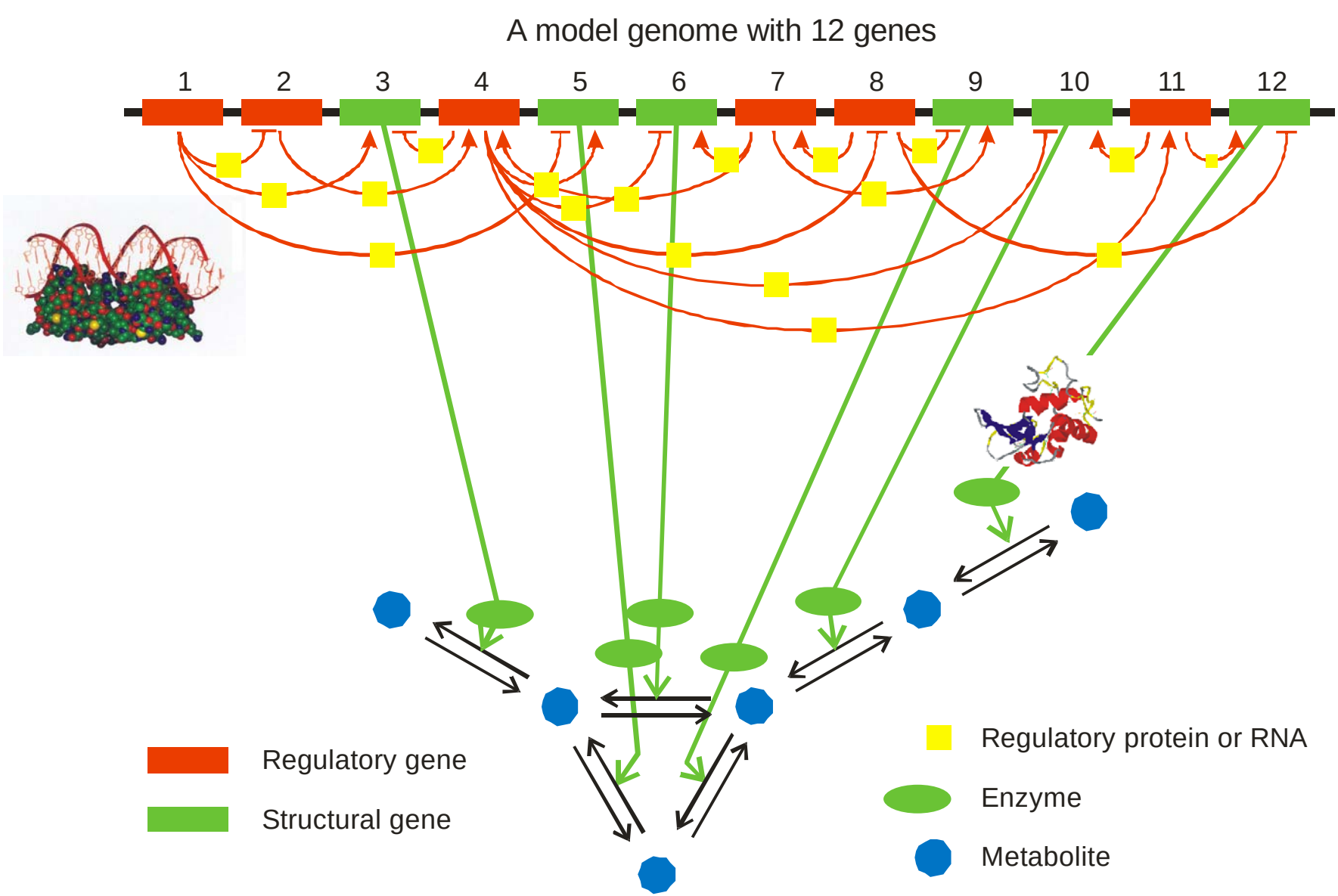
- Direct evidence that neutrality is a major factor for the success of evolution.

1. Evolution – organismic and molecular
2. Multiplication, mutation,and selection
3. Rational design of molecules
4. Evolution and optimization of molecules
5. **Origin of biological complexity**
6. Biology and probabilities


Replication:
DNA ⇒ 2 DNA
+
Transcription:
DNA ⇒ RNA
Nucleotides
Amino Acids
Lipids
Carbohydrates
Small Molecules
Metabolism
Food
Waste
Ribosom
Protein
mRNA
Translation: RNA ⇒ Protein



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



Sketch of a genetic and metabolic network

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 246 (2007) 395–419
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Dynamic patterns of gene regulation I: Simple two-gene systems

Stefanie Widder[a], Josef Schicho[b], Peter Schuster[a,c,*]

[a]*Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
[b]*RICAM—Johann Radon Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*
[c]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 24 February 2006; received in revised form 7 January 2007; accepted 8 January 2007
Available online 16 January 2007

Abstract

Regulation of gene activities is studied by means of computer assisted mathematical analysis of ordinary differential equations (ODEs) derived from binding equilibria and chemical reaction kinetics. Here, we present results on cross-regulation of two genes through activator and/or repressor binding. Arbitrary (differentiable) binding function can be used but systematic investigations are presented for gene–regulator complexes with integer valued Hill coefficients up to $n = 4$. The dynamics of gene regulation is derived from bifurcation patterns of the underlying systems of kinetic ODEs. In particular, we present analytical expressions for the parameter values at which one-dimensional (transcritical, saddle-node or pitchfork) and/or two-dimensional (Hopf) bifurcations occur. A classification of regulatory states is introduced, which makes use of the sign of a 'regulatory determinant' D (being the determinant of the block in the Jacobian matrix that contains the derivatives of the regulator binding functions): (i) systems with $D<0$, observed, for example, if both proteins are activators or repressors, to give rise to one-dimensional bifurcations only and lead to bistability for $n \geqslant 2$ and (ii) systems with $D>0$, found for combinations of activation and repression, sustain a Hopf bifurcation and undamped oscillations for $n>2$. The influence of basal transcription activity on the bifurcation patterns is described. Binding of multiple subunits can lead to richer dynamics than pure activation or repression states if intermediates between the unbound state and the fully saturated DNA initiate transcription. Then, the regulatory determinant D can adopt both signs, plus and minus.

Keywords: Basal transcription; Bifurcation analysis; Cooperative binding; Gene regulation; Hill coefficient; Hopf bifurcation

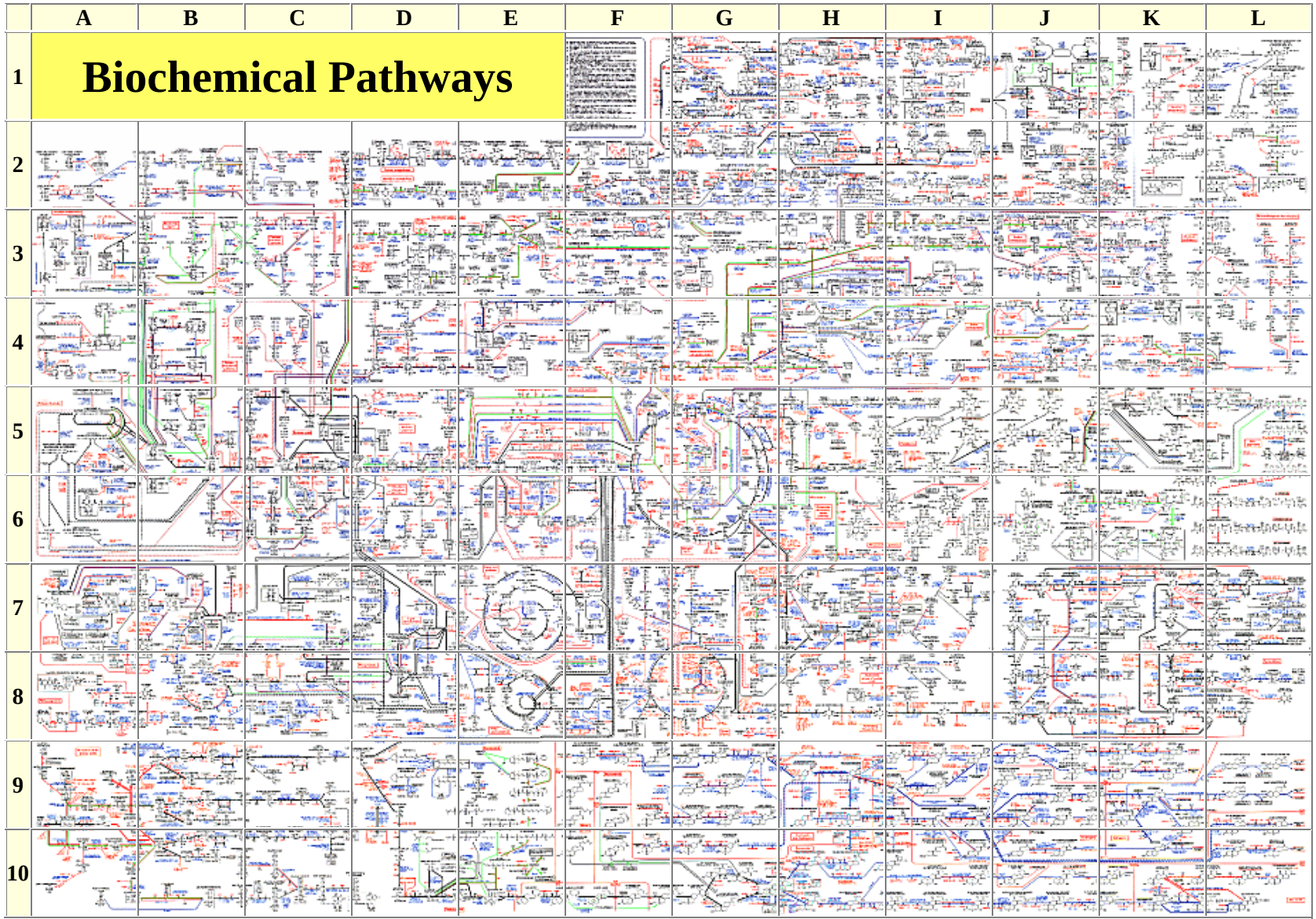
1. Introduction

Theoretical work on gene regulation goes back to the 1960s (Monod et al., 1963) soon after the first repressor protein had been discovered (Jacob and Monod, 1961). A little later the first paper on oscillatory states in gene regulation was published (Goodwin, 1965). The interest in gene regulation and its mathematical analysis never ceased (Tiwari et al., 1974; Tyson and Othmer, 1978; Smith, 1987) and saw a great variety of different attempts to design models of genetic regulatory networks that can be used in systems biology for computer simulation of *gen*(etic and met)*abolic* networks.[1] Most models in the literature aim at a minimalist dynamic description which, nevertheless, tries to account for the basic regulatory functions of large networks in the cell in order to provide a better understanding of cellular dynamics. A classic in general regulatory dynamics is the monograph by Thomas and D'Ari (1990). The currently used mathematical methods comprise application of Boolean logic (Thomas and Kaufman, 2001b; Savageau, 2001; Albert and Othmer, 2003), stochastic processes (Hume, 2000) and deterministic dynamic models, examples are Cherry and Adler (2000), Bindschadler and Sneyd (2001) and Kobayashi et al. (2003) and the recent elegant analysis of bistability (Craciun et al.,

*Corresponding author. Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.
Tel.: +43 1 4277 52743; fax: +43 1 4277 52793.
E-mail address: pks@tbi.univie.ac.at (P. Schuster).

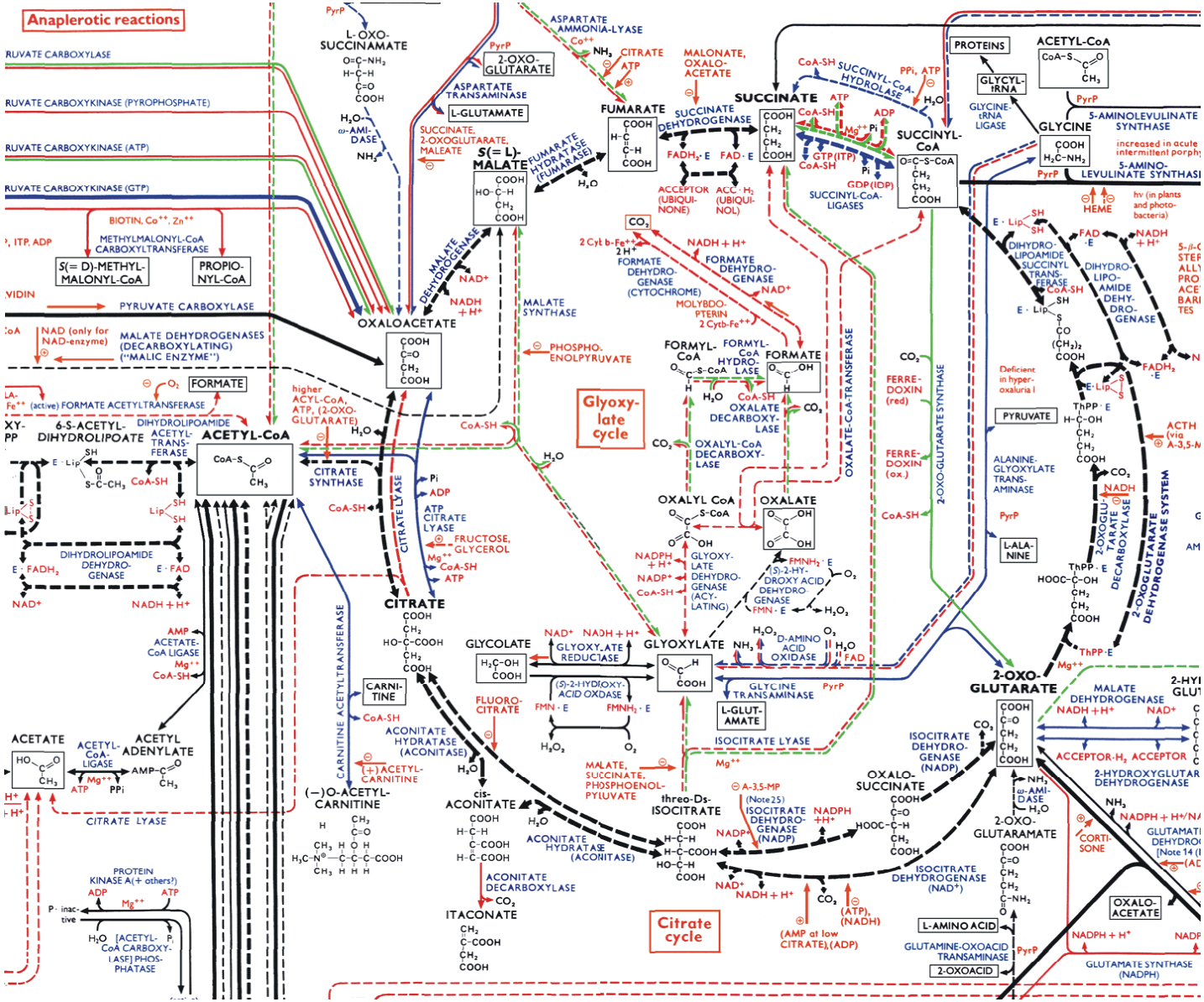
[1]Discussion and analysis of combined genetic and metabolic networks has become so frequent and intense that we suggest to use a separate term, *genabolic networks*, for this class of complex dynamical systems.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.01.004



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

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

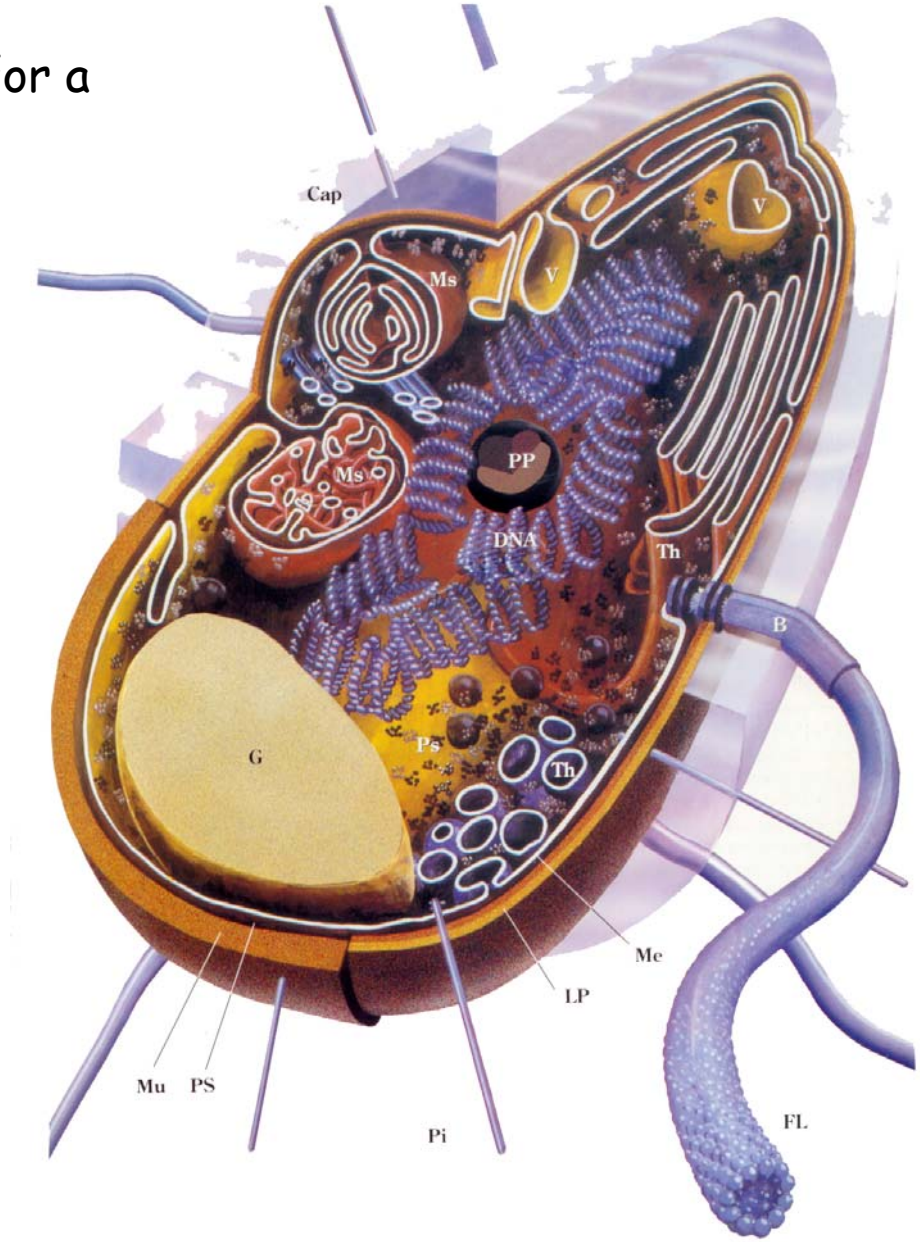


The bacterial cell as an example for a simple form of autonomous life

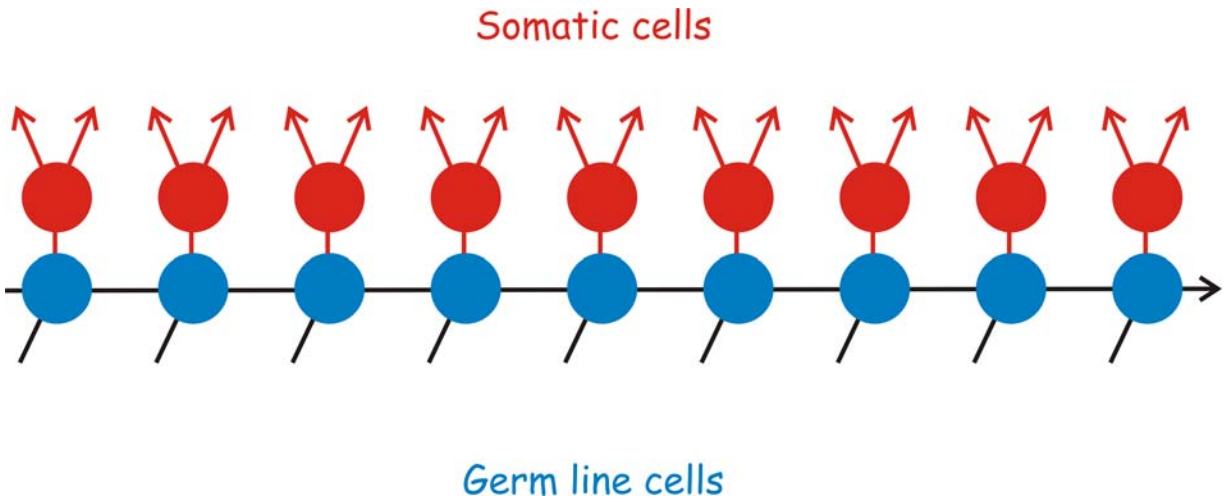
Escherichia coli genome:

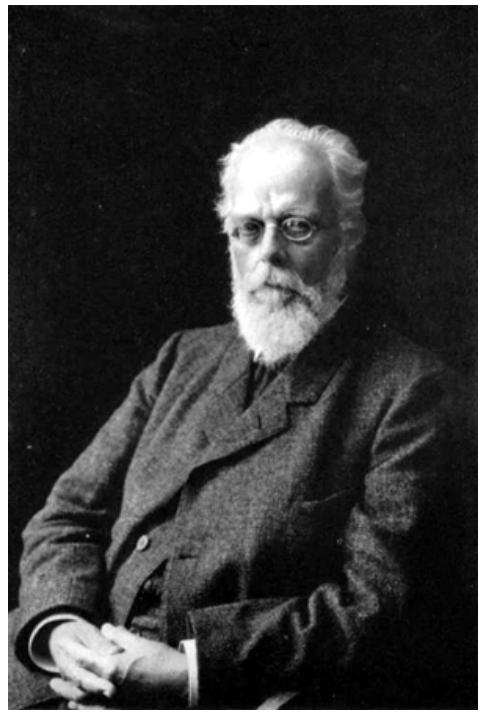
4 million nucleotides

4460 genes



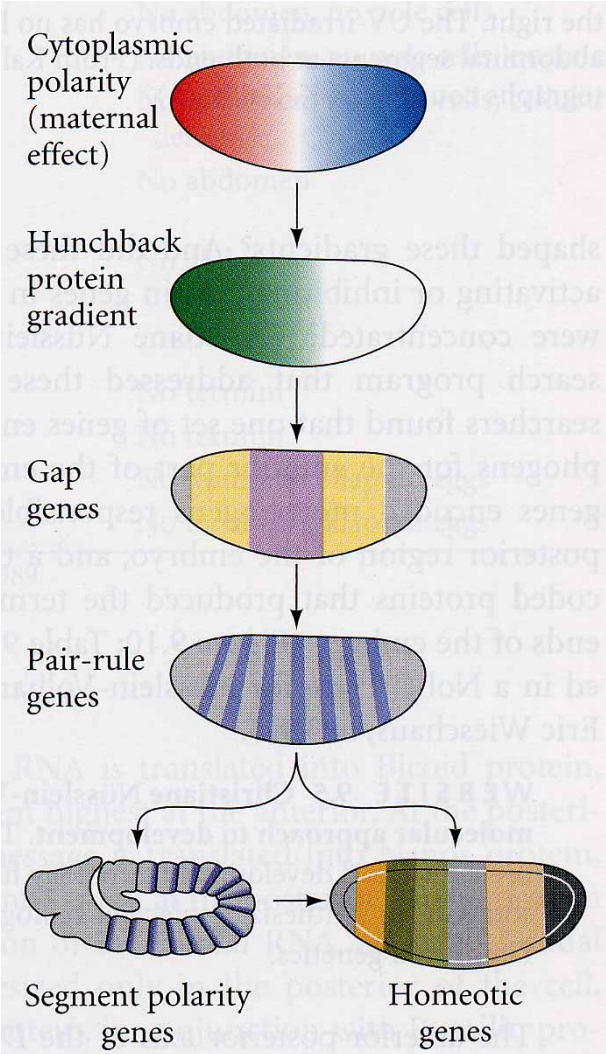
The structure of the bacterium *Escherichia coli*

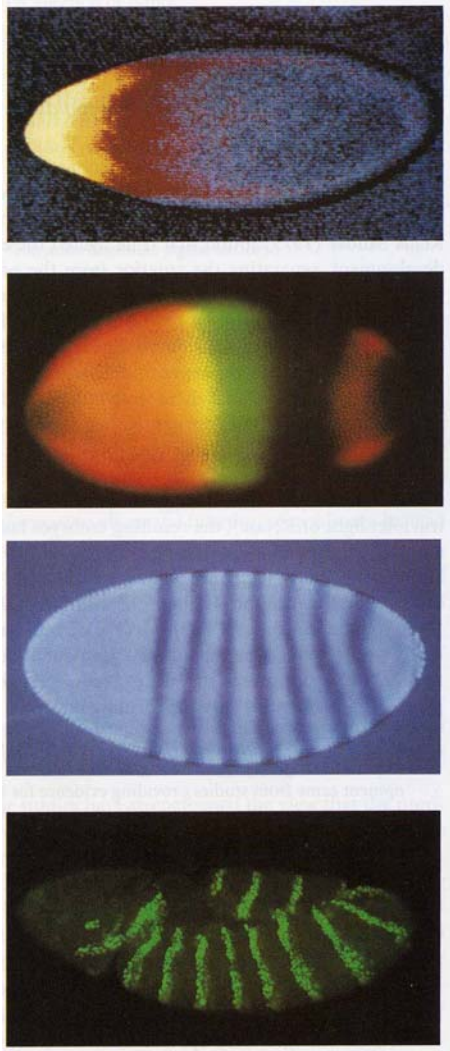


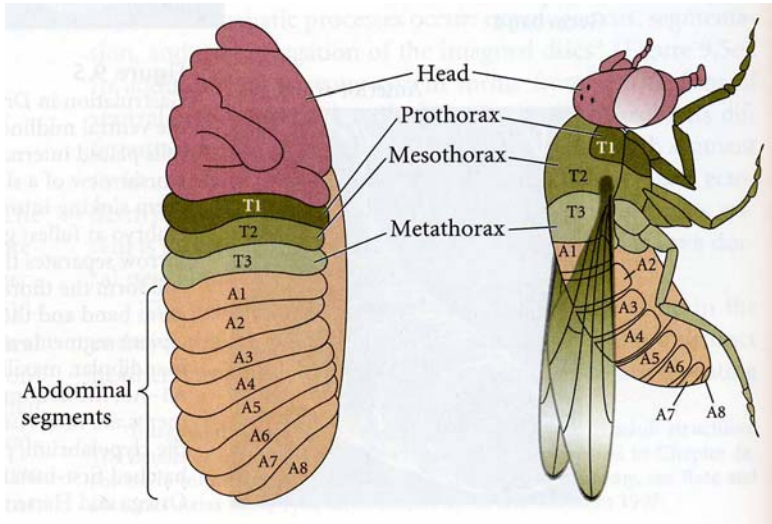


August Weismann, 1834-1914

Separation of germ line and soma







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

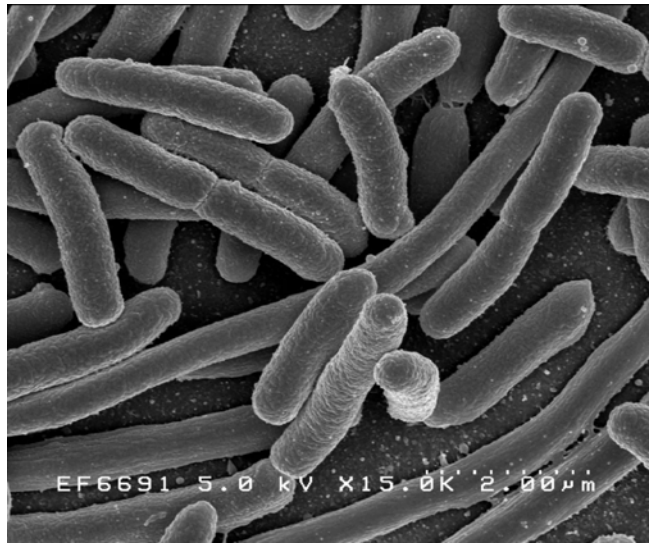
Turing pattern resulting from reaction-diffusion equation ?

Intercelluar communication creating positional information

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

E. coli:	Genome length	4×10^6 nucleotides
	Number of cell types	1
	Number of genes	4 460

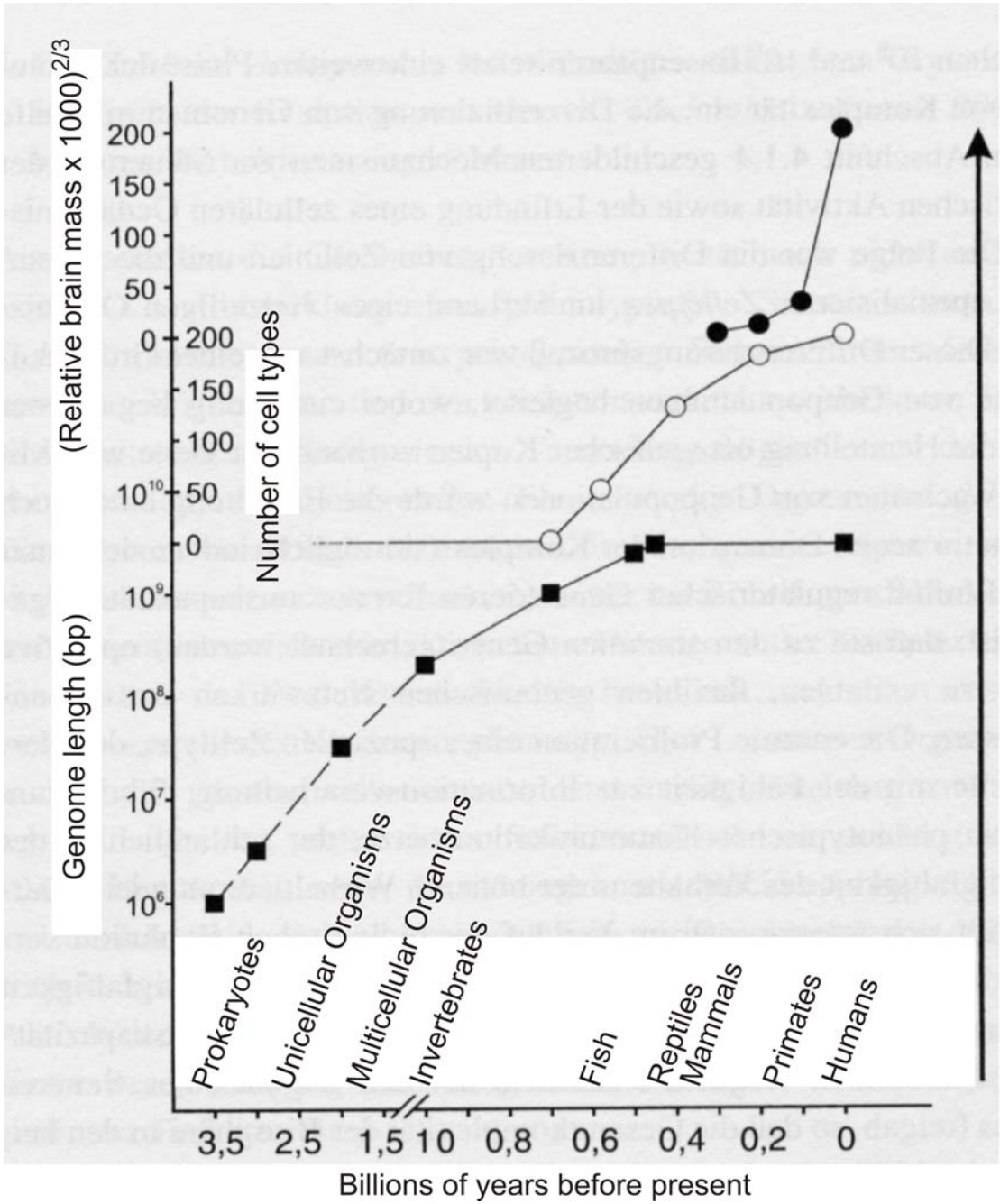
Four books, 300 pages each



Man:	Genome length	3×10^9 nucleotides
	Number of cell types	200
	Number of genes	≈ 30 000

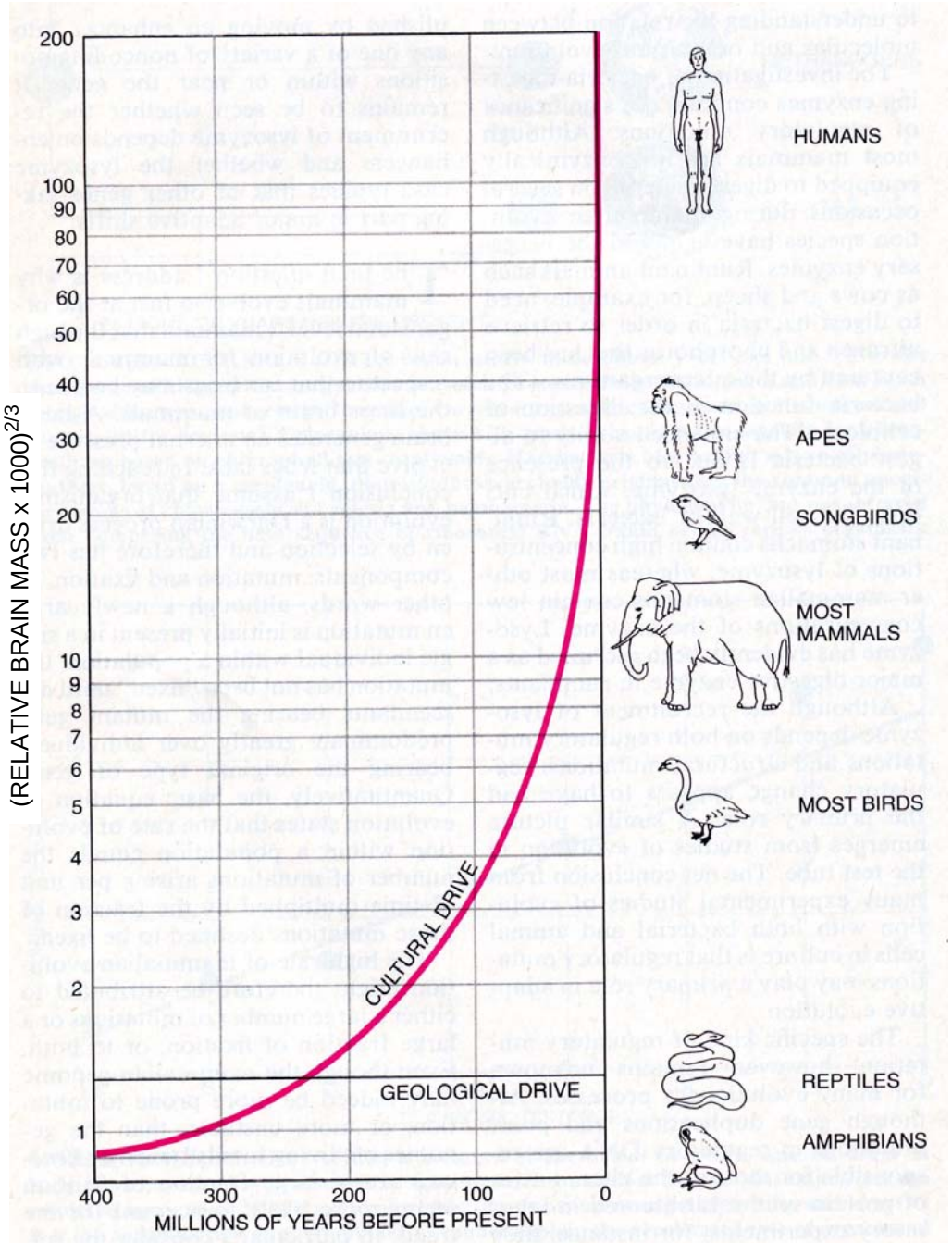
A library of 3000 volumes,
300 pages each

Complexity in biology

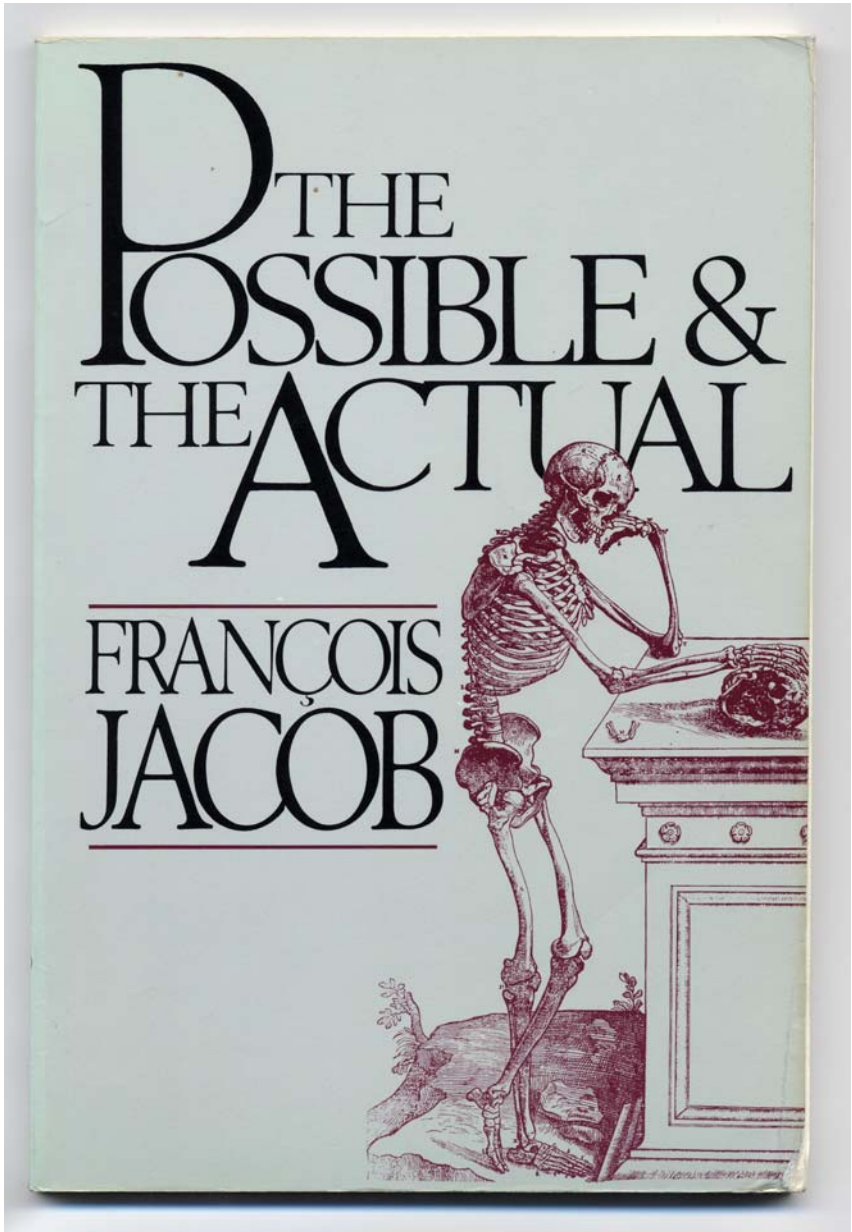


Wolfgang Wieser. 1998. ‚*Die Erfindung der Individualität*‘ oder ‚*Die zwei Gesichter der Evolution*‘. Spektrum Akademischer Verlag, Heidelberg 1998

BRITISH TIT



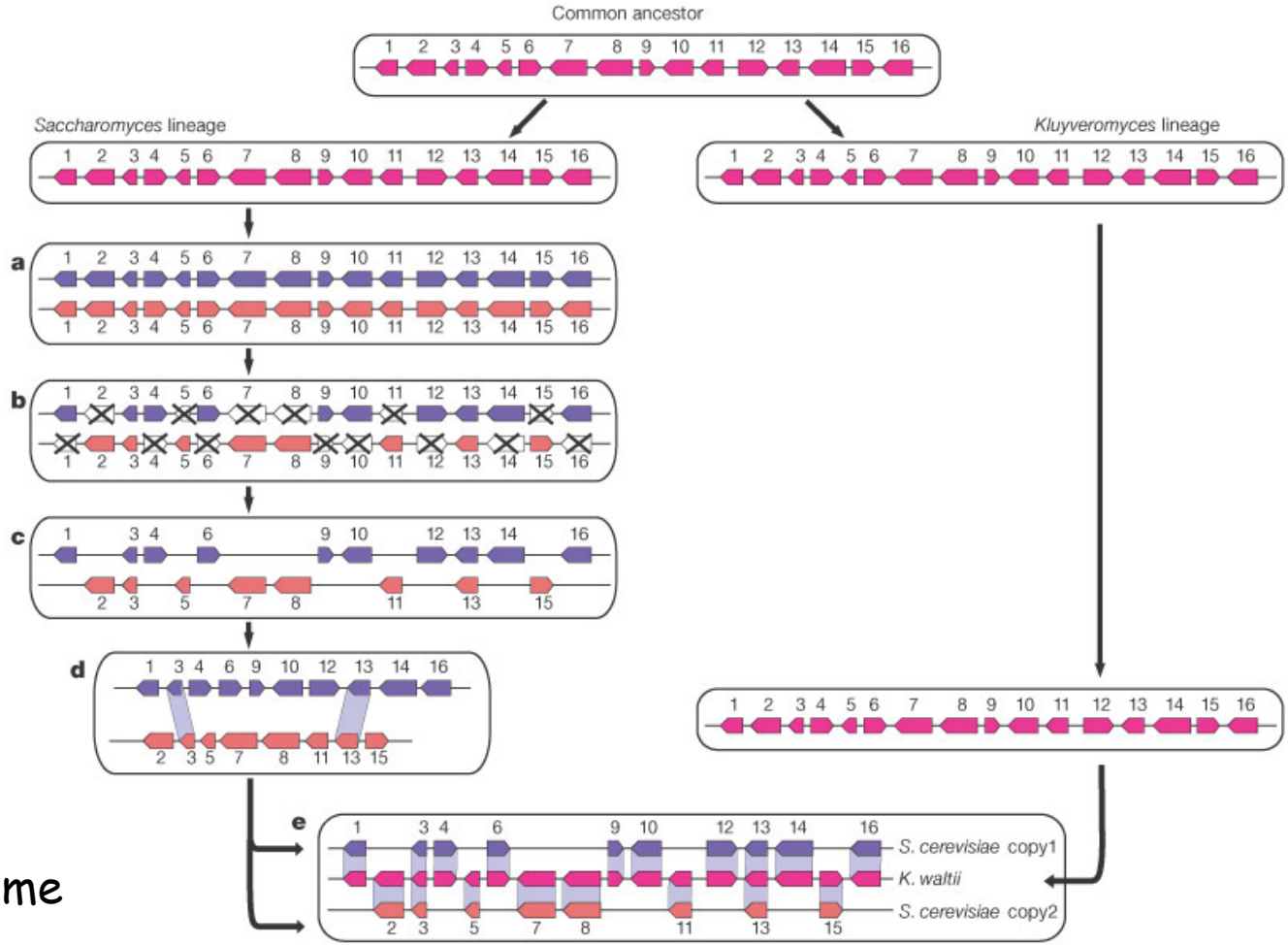
Alan C. Wilson.1985. The molecular basis of evolution.
Scientific American **253**(4):148-157.



Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

François Jacob. *The Possible and the Actual*. Pantheon Books, New York, 1982, and

Evolutionary tinkering. *Science* **196** (1977), 1161-1166.



A model for the genome duplication in yeast 100 million years ago

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004

The difficulty to define the notion of „gene".

Helen Pearson,
Nature **441**: 399-401, 2006

WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past[1]. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals[2]. If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist Laurence Hurst at the University of Bath, UK.

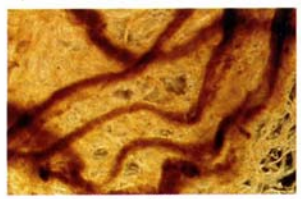
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

"We've come to the realization that the genome is full of overlapping transcripts."
— Phillip Kapranov

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the previously unimagined scope of RNA.



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

P. PLAILLY/SPL

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from[3].

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team[4], and one by geneticist Rotem Sorek[5], now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

ENCODE stands for
ENCyclopedia Of DNA Elements.

ENCODE Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
Nature **447**:799-816, 2007

Biology and complexity:

- Evolution does not design with the eyes of an engineer but uses available objects for new purposes.

- The tinkering or bricolage principle gives rise to new objects of increasing complexity.

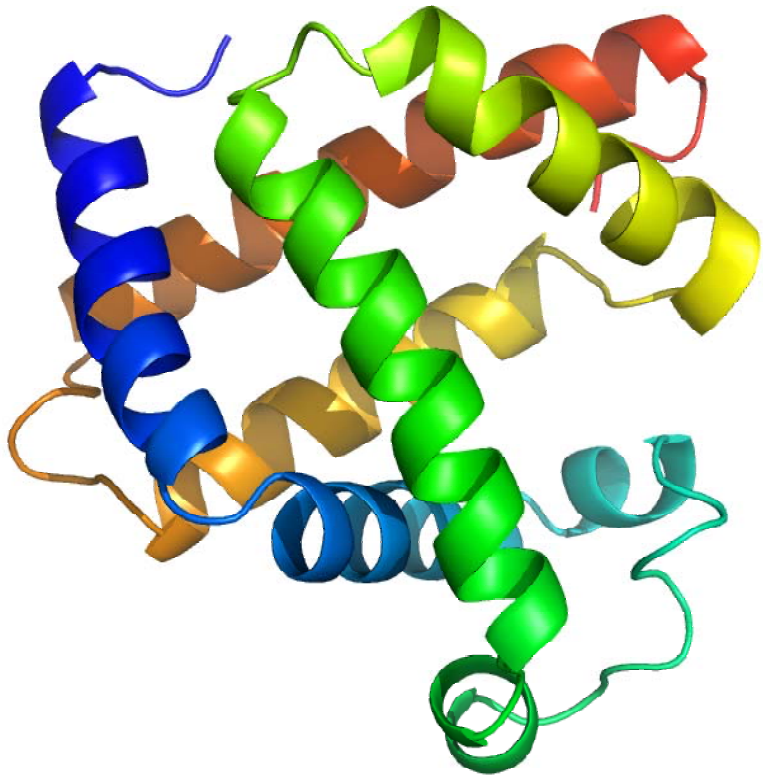
- The increase of complexity in biological evolution is an empirical fact.

1. Evolution – organismic and molecular
2. Multiplication, mutation, and selection
3. Rational design of molecules
4. Evolution and optimization of molecules
5. Origin of biological complexity
6. **Biology and probabilities**

Polymer chain of 153 amino acid residues with the sequence:

GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLK
SEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIP
VKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELG
FQG

The myglobin molecule



Eugene Wigner's or Fred Hoyle's argument applied to myoglobin:

All sequences have equal probability and all except the correct one have no survival value or are lethal

GLSDGEWQLVLNVWG.....FQG

Alphabet size: 20

Chain length: 153 amino acids

Number of possible sequences: $20^{153} = \mathbf{0.11 \times 10^{200}}$

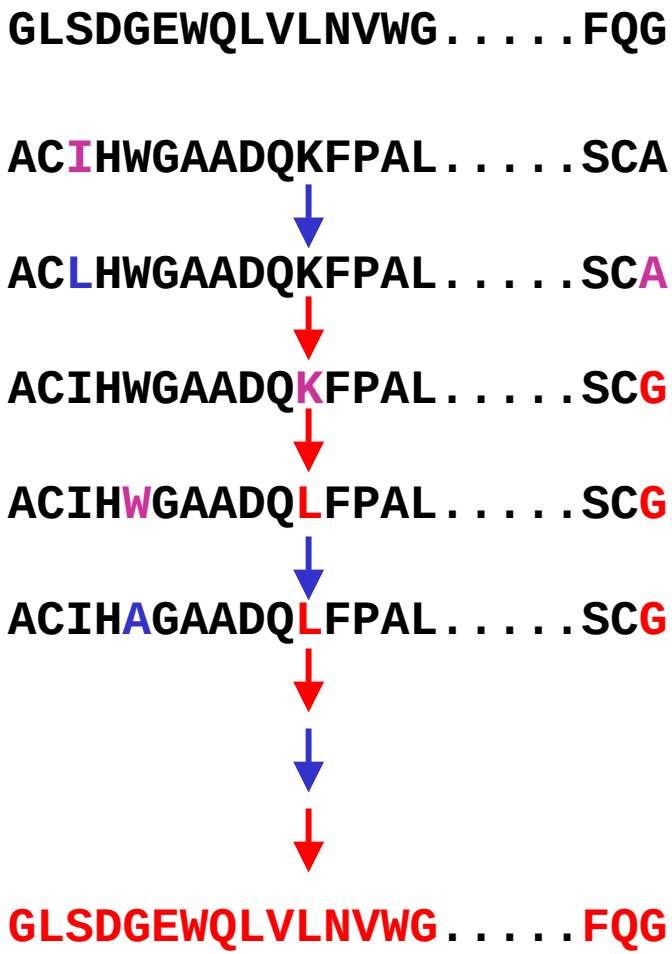
Probability to find the myoglobin sequence:

$20^{-153} = \mathbf{9 \times 10^{-200}} = 0.000\ldots\ldots009$

200

Eugene Wigner's and Fred Hoyle's arguments revisited:

Every single point mutation towards the target sequence leads to an improvement and is therefore selected



Alphabet size: 20

Chain length: 153 amino acids

Length of longest path to myoglobin sequence: $19 \times 153 = 2907$

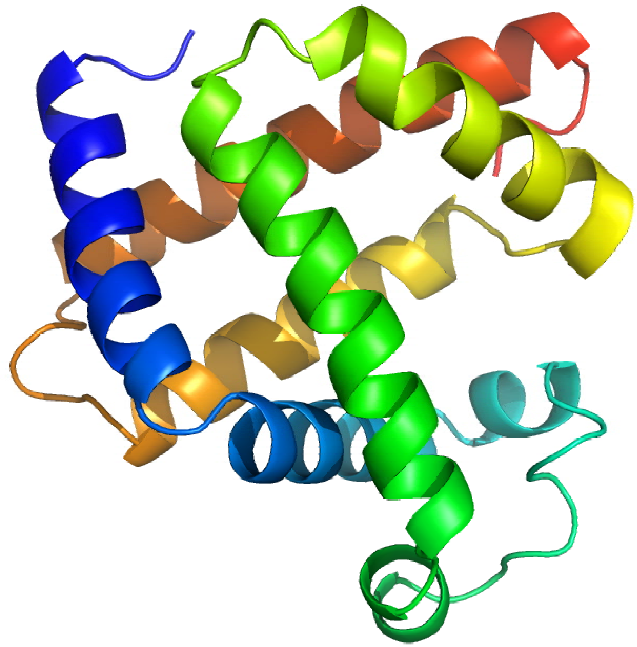
Probability to find the myoglobin sequence: **0.00034**

The folding problem of the myoglobin molecule:

A chain of 153 amino acid residues, each of which can adopt about 15 different geometries, can exist in

$15^{153} = 0.9 \times 10^{180}$ conformations.

One specific conformation – the most stable or minimum free energy conformation – has to be found in the folding process.



The Levinthal paradox of protein folding

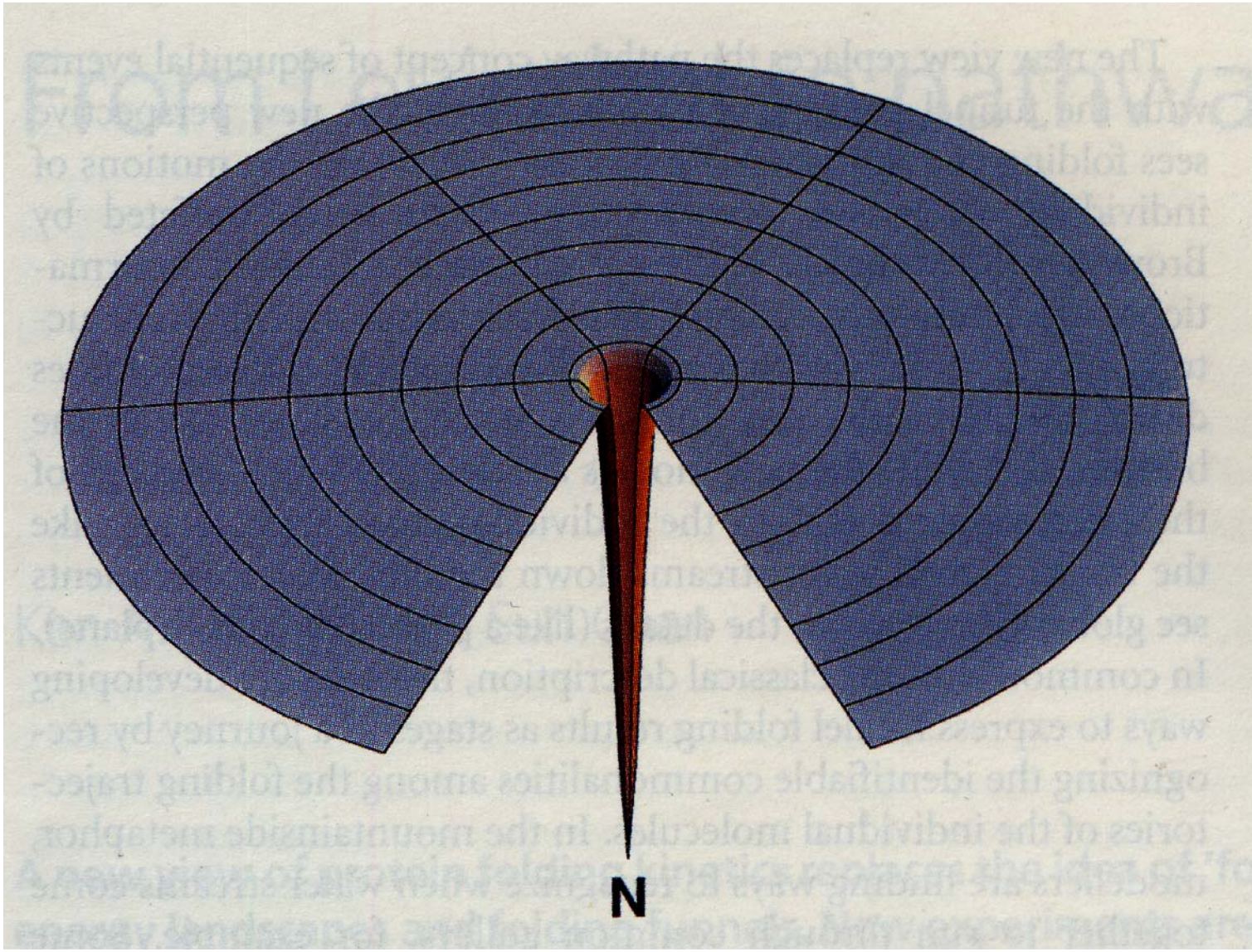
Three basic questions of the protein folding problem:

What is the folding code ?

What is the folding mechanism ?

Can we predict the native structure of a protein from its amino acid sequence?

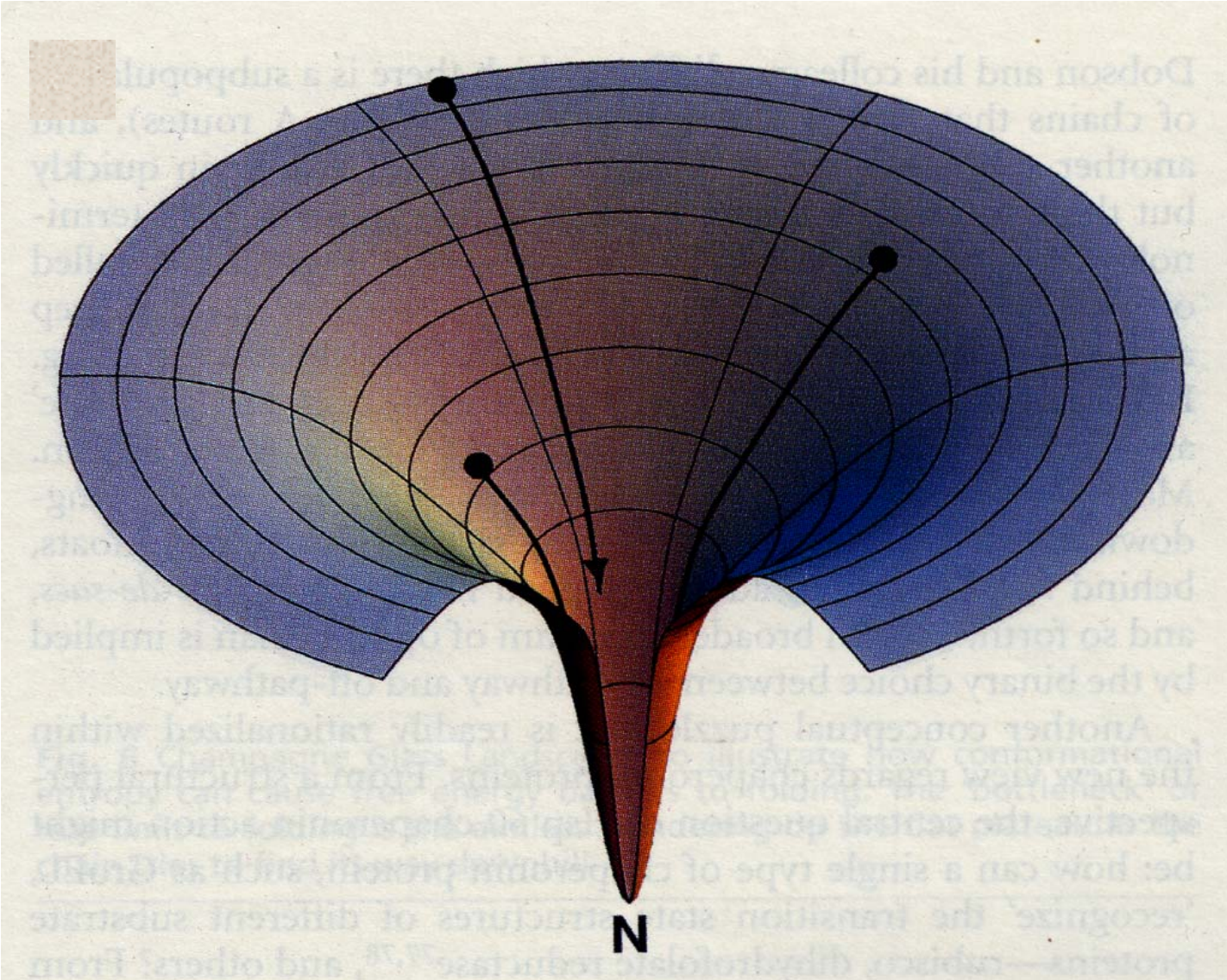
K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl. 2008. The protein folding problem. Annu.Rev.Biophys. **37**:289-316.



The gulf course landscape

Solution to Levinthal's paradox

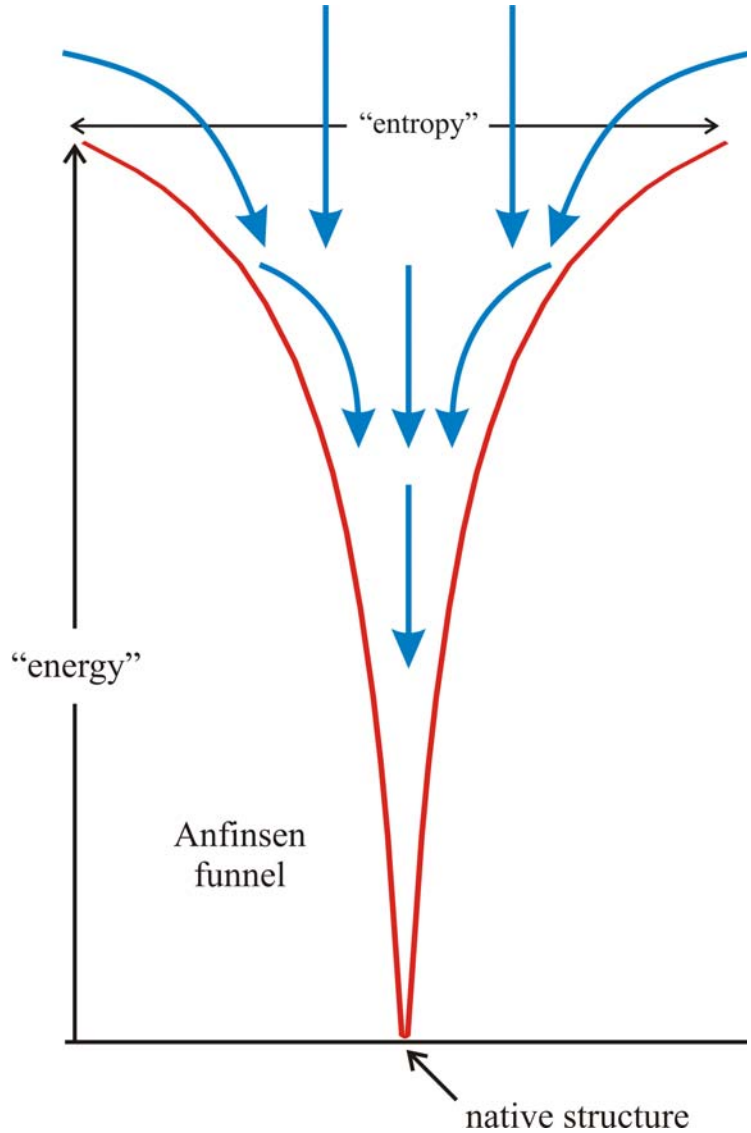
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19



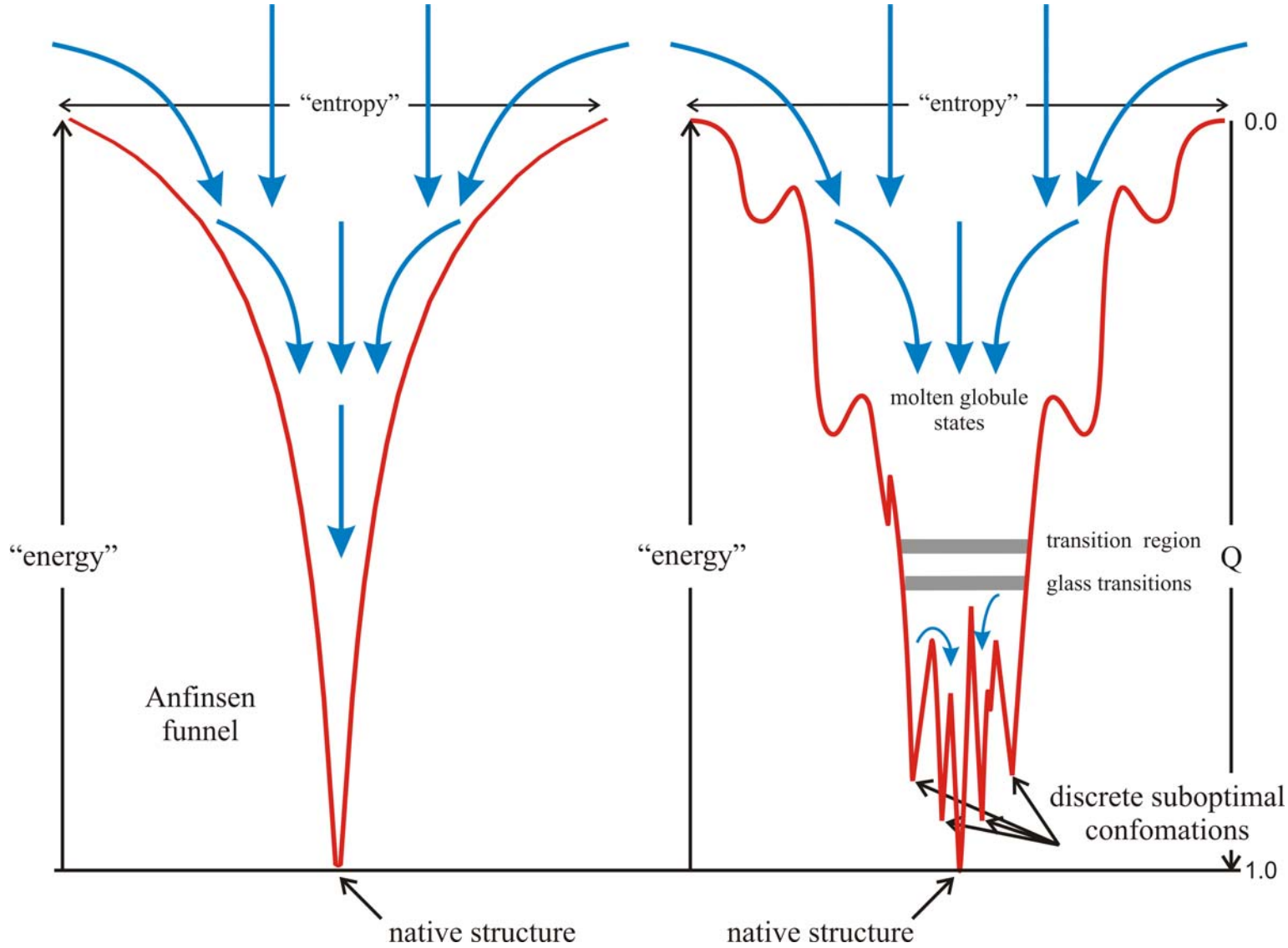
The funnel landscape

Solution to Levinthal’s paradox

Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19

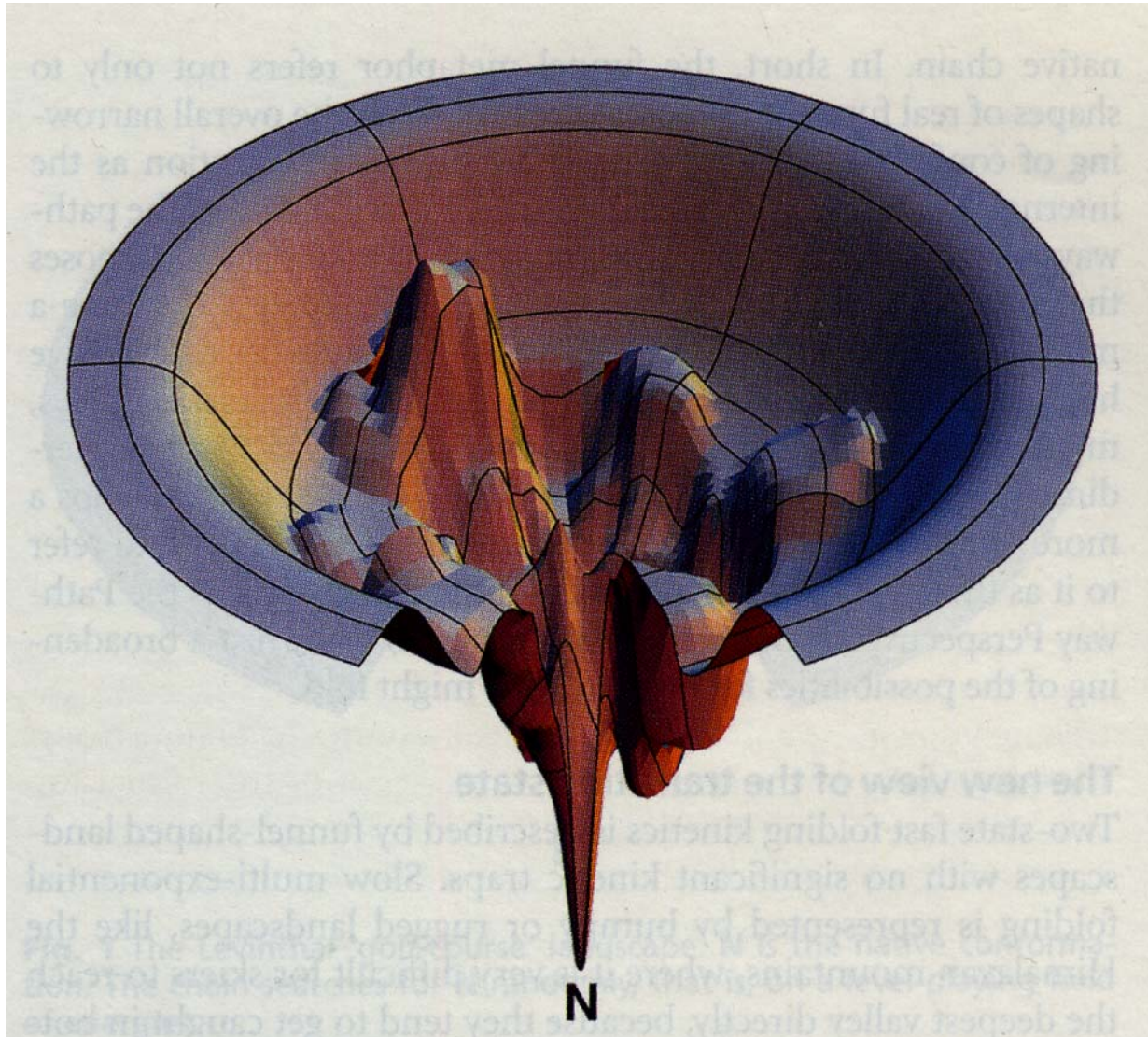


Folding funnel of ribonuclease A



Folding funnel of ribonuclease A

Folding funnel of an inefficiently folding protein

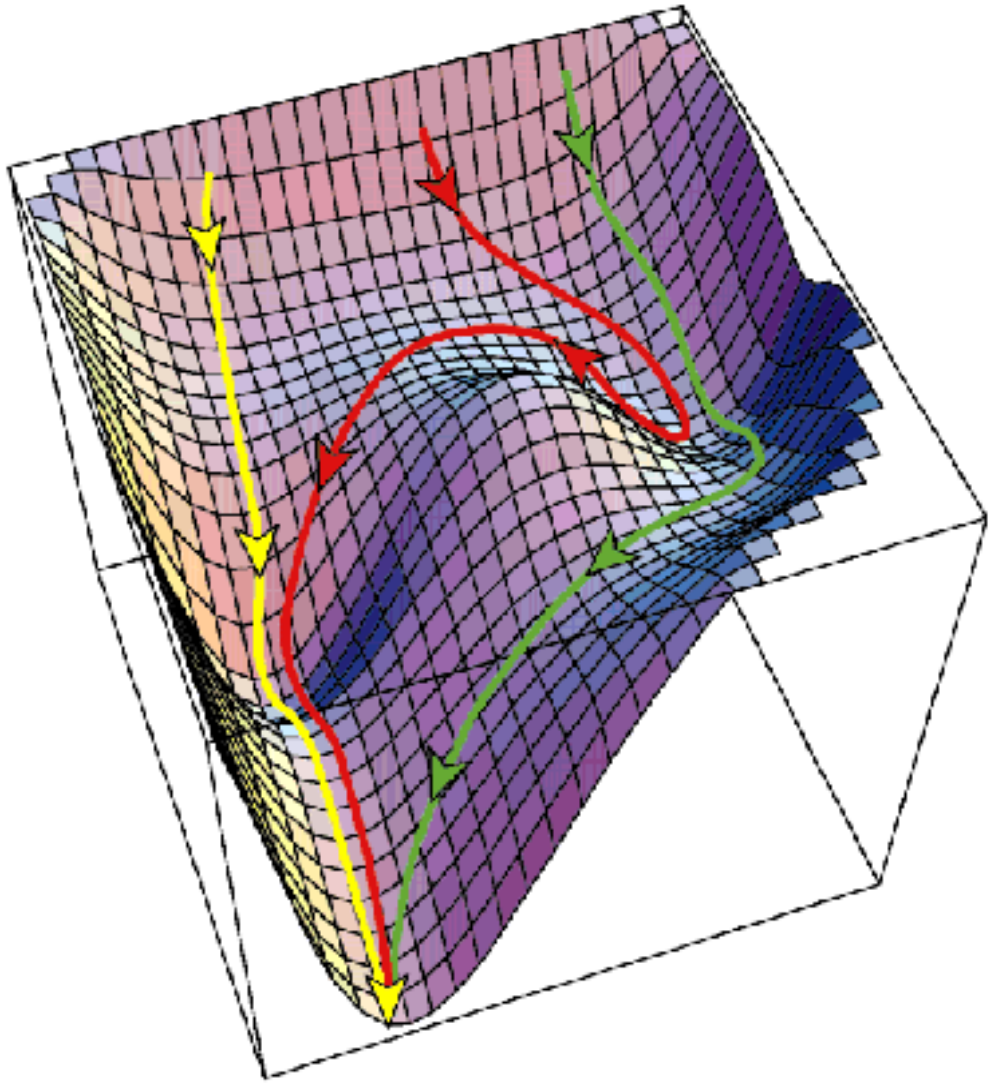


The structured funnel landscape

Solution to Levinthal’s paradox

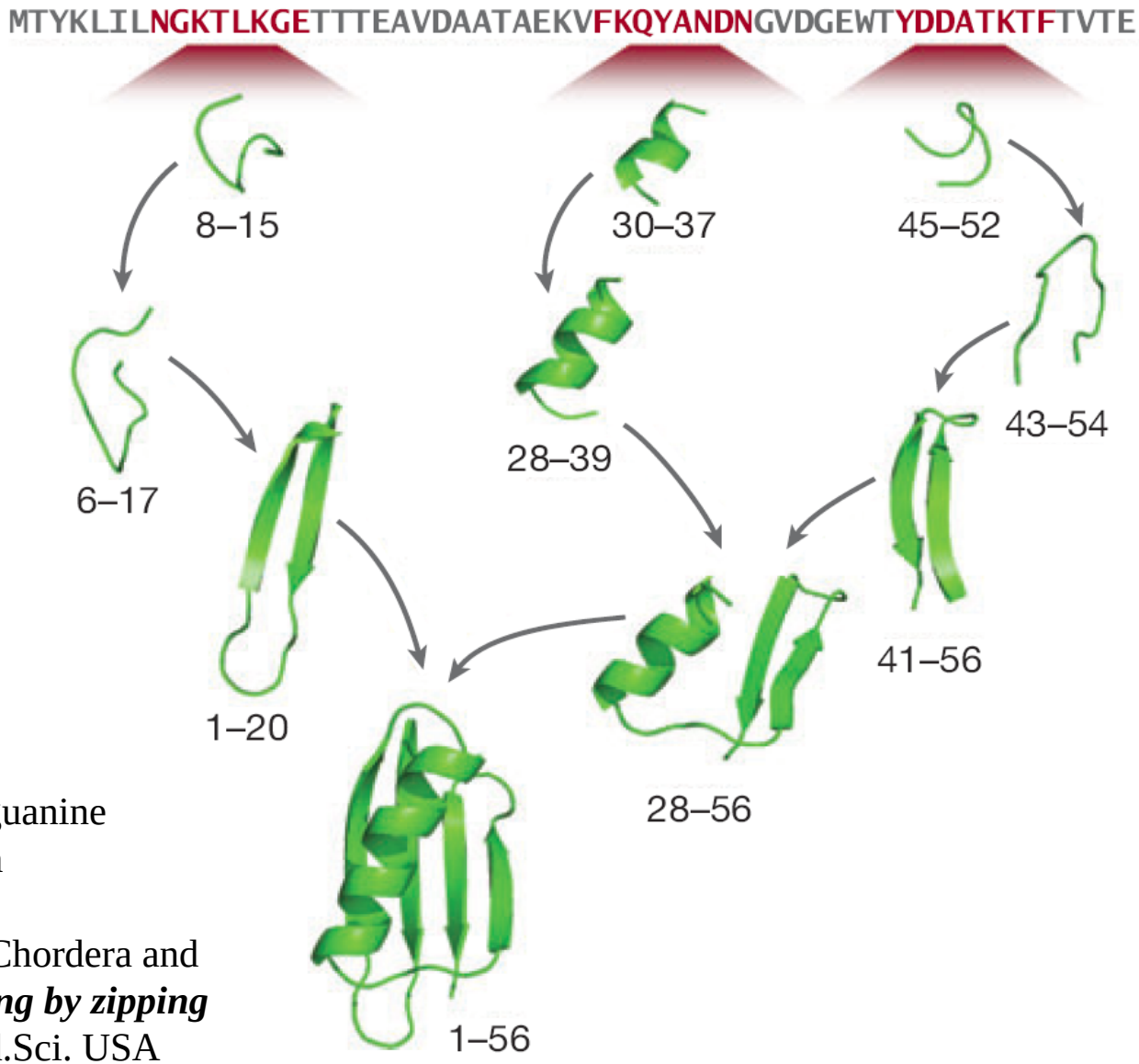
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19

An “all-roads-lead-to-Rome” landscape



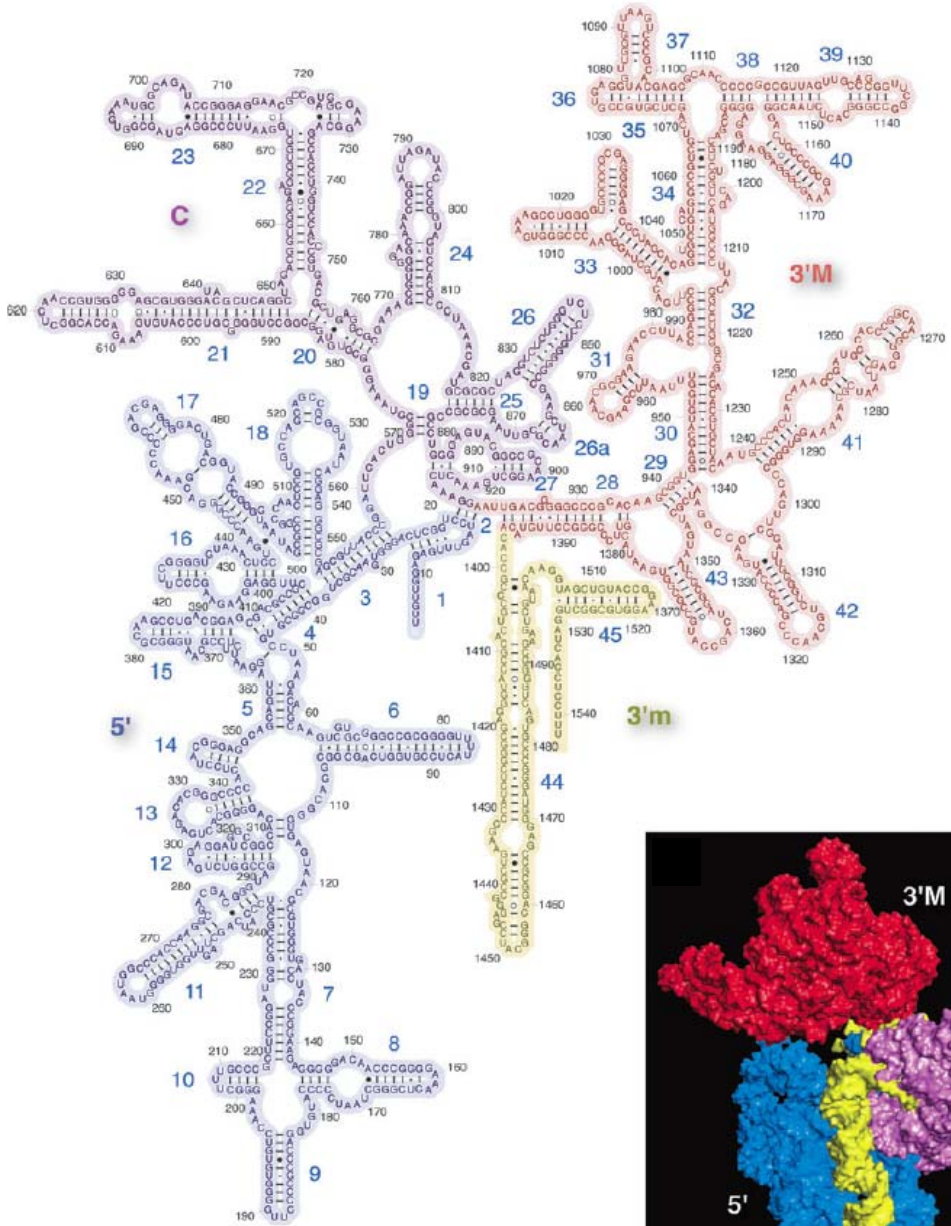
The reconstructed folding landscape
of a real biomolecule: “lysozyme”

Picture: C.M. Dobson, A. Šali, and M. Karplus, Angew.Chem.Internat.Ed. 37: 868-893, 1988



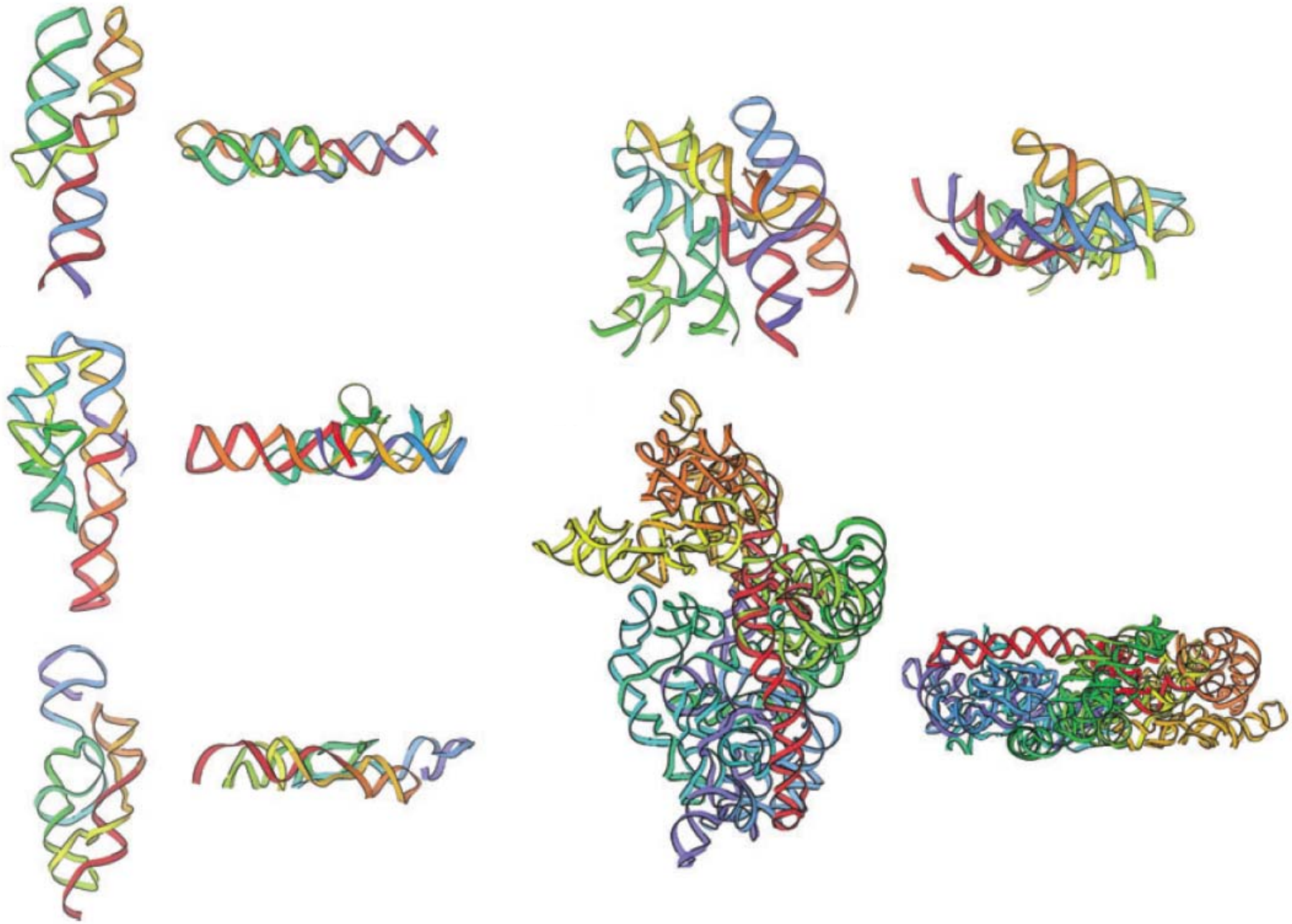
Computed folding routes for guanine nucleotide binding (G) protein

S.B. Ozkan, G.H.A. Wu, J.D.Chordera and K.A. Dill. 2007. ***Protein folding by zipping and assembly.*** Proc.Natl.Acad.Sci. USA **104**:11987-11992.



Structures of RNA molecules

S.R. Holbrook. 2008. Structural principles from large RNA molecules. Annu.Rev.Biophys. **37**:445-464.



Structures of RNA molecules

S.R. Holbrook. 2008. Structural principles from large RNA molecules. Annu.Rev.Biophys. **37**:445-464.

Web-Page for further information:

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