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

Meeting of the Honda Foundation

Wien, 19.12.2008

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***.

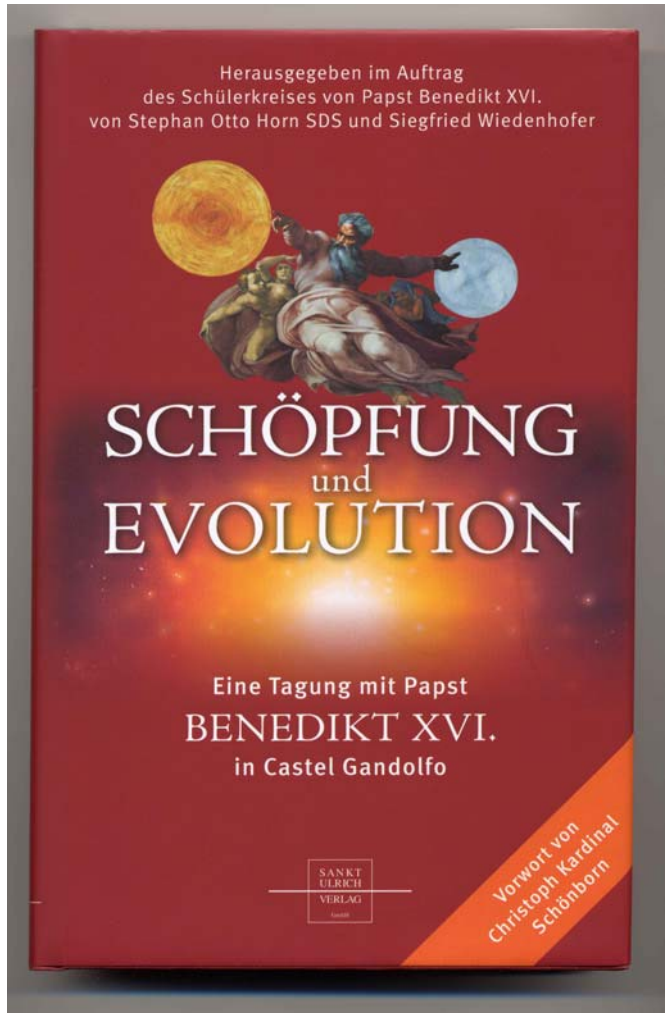
In: Stephan Otto Horn und Siegfried Wiedenhofer, Eds.

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



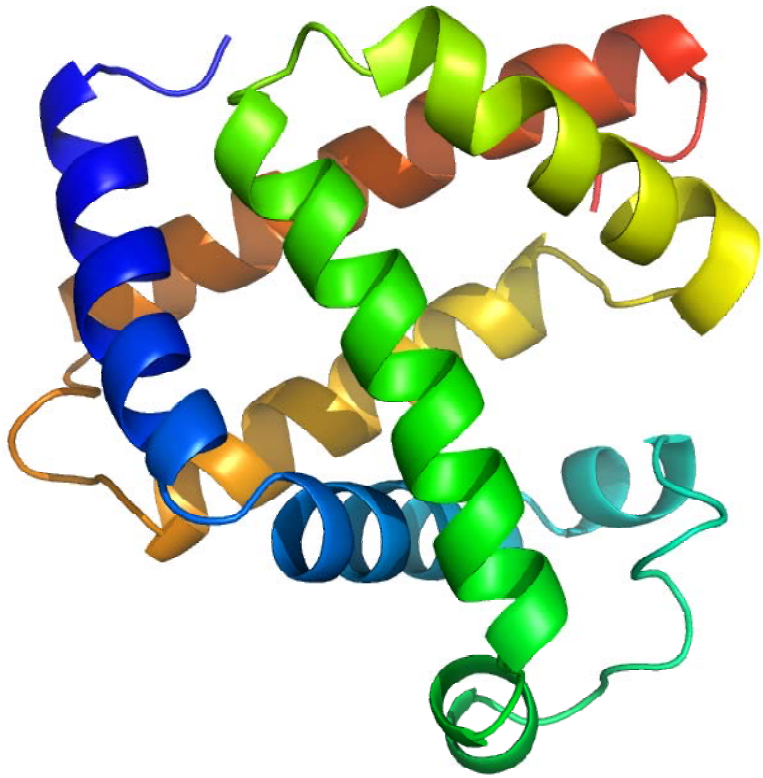
1. Biology and probabilities
2. Evolution – organismic and molecular
3. Multiplication, mutation, and selection
4. Rational design of molecules
5. Evolution and optimization of molecules
6. Origin of biological complexity

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

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: $20^{153} = \mathbf{0.11 \times 10^{200}}$

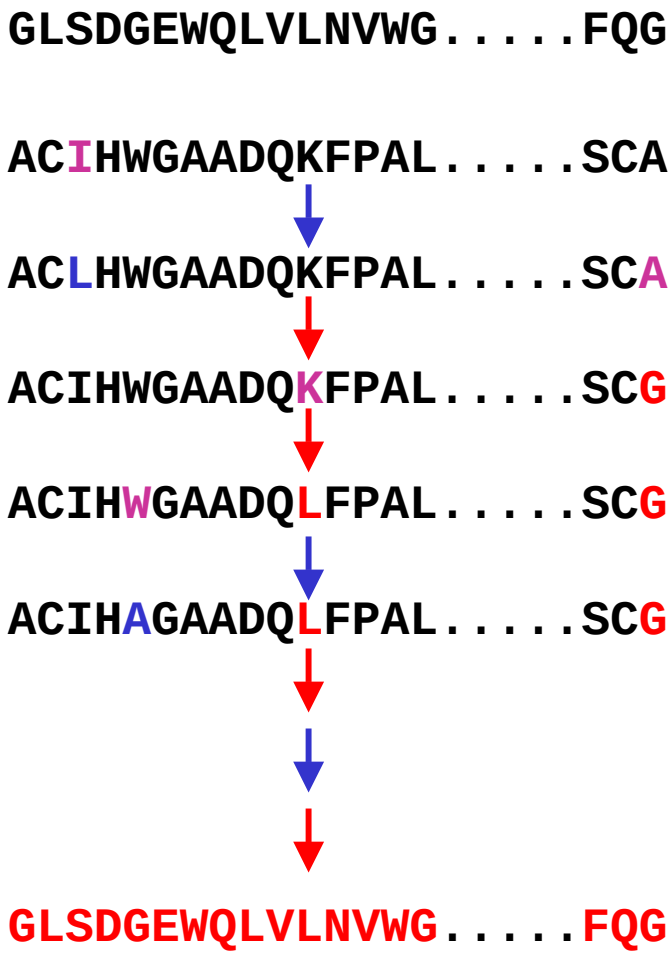
Probability to find the myoglobin sequence:

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

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: $19 \times 153 = 2907$

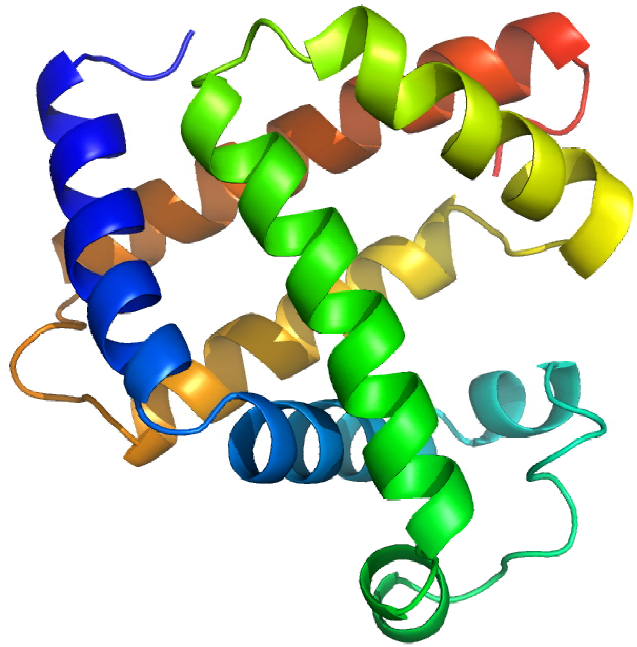
Probability to find the myoglobin sequence: **0.00034**

The folding problem of the myoglobin molecule:

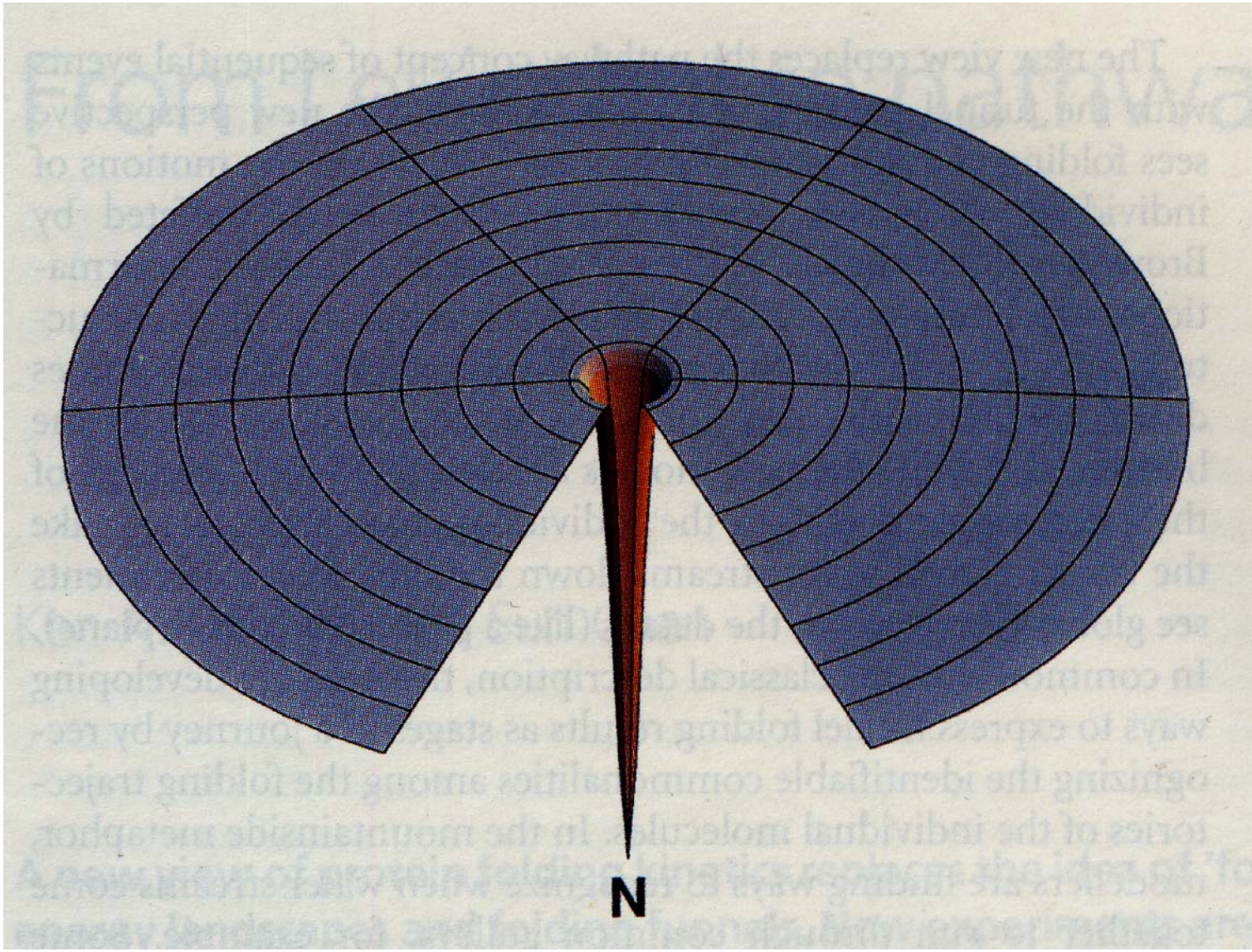
A chain of 153 amino acid residues, each of which can adopt about 15 different geometries, can exist in

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

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



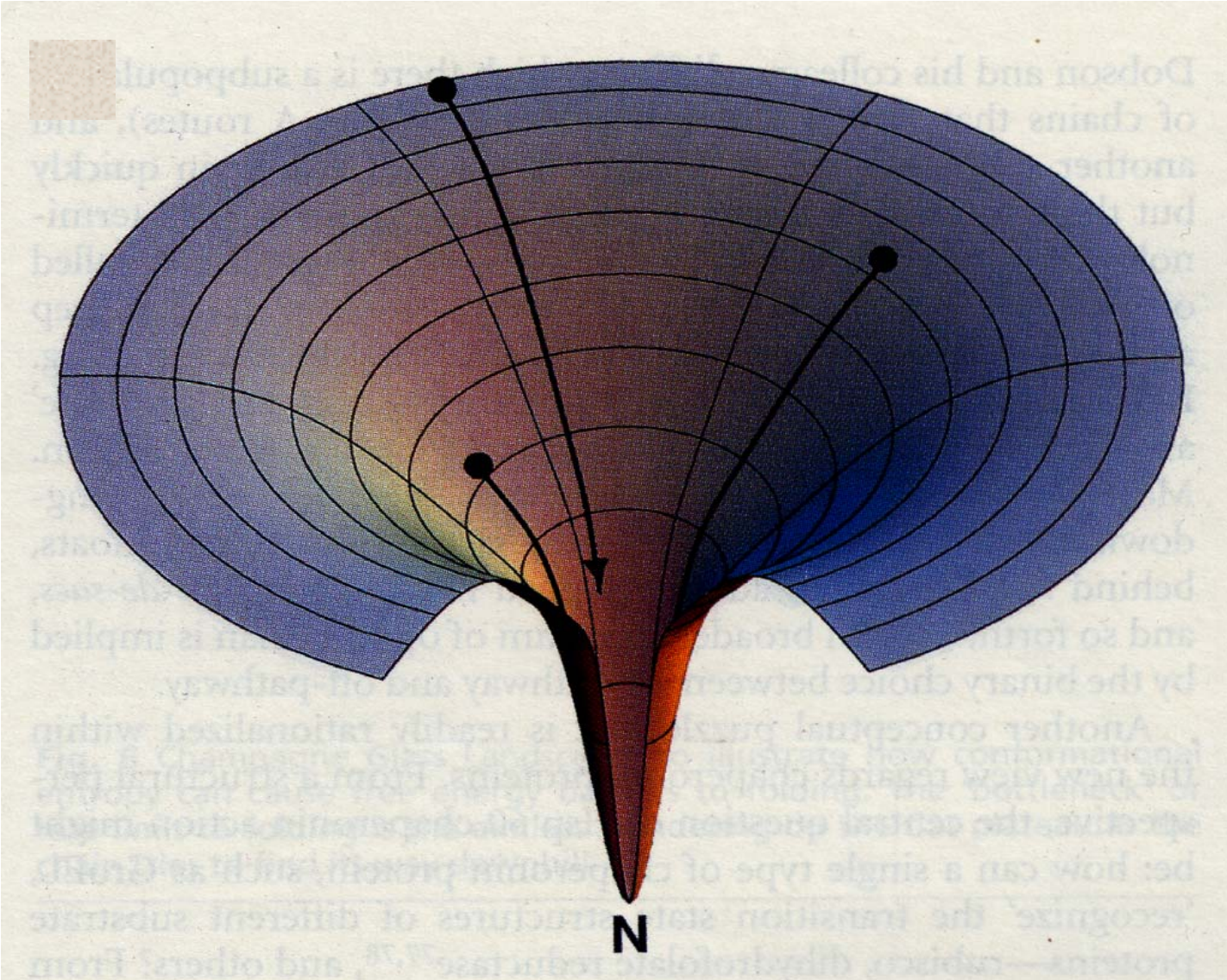
The Levinthal paradox of protein folding



The gulf course landscape

Solution to Levinthal's paradox

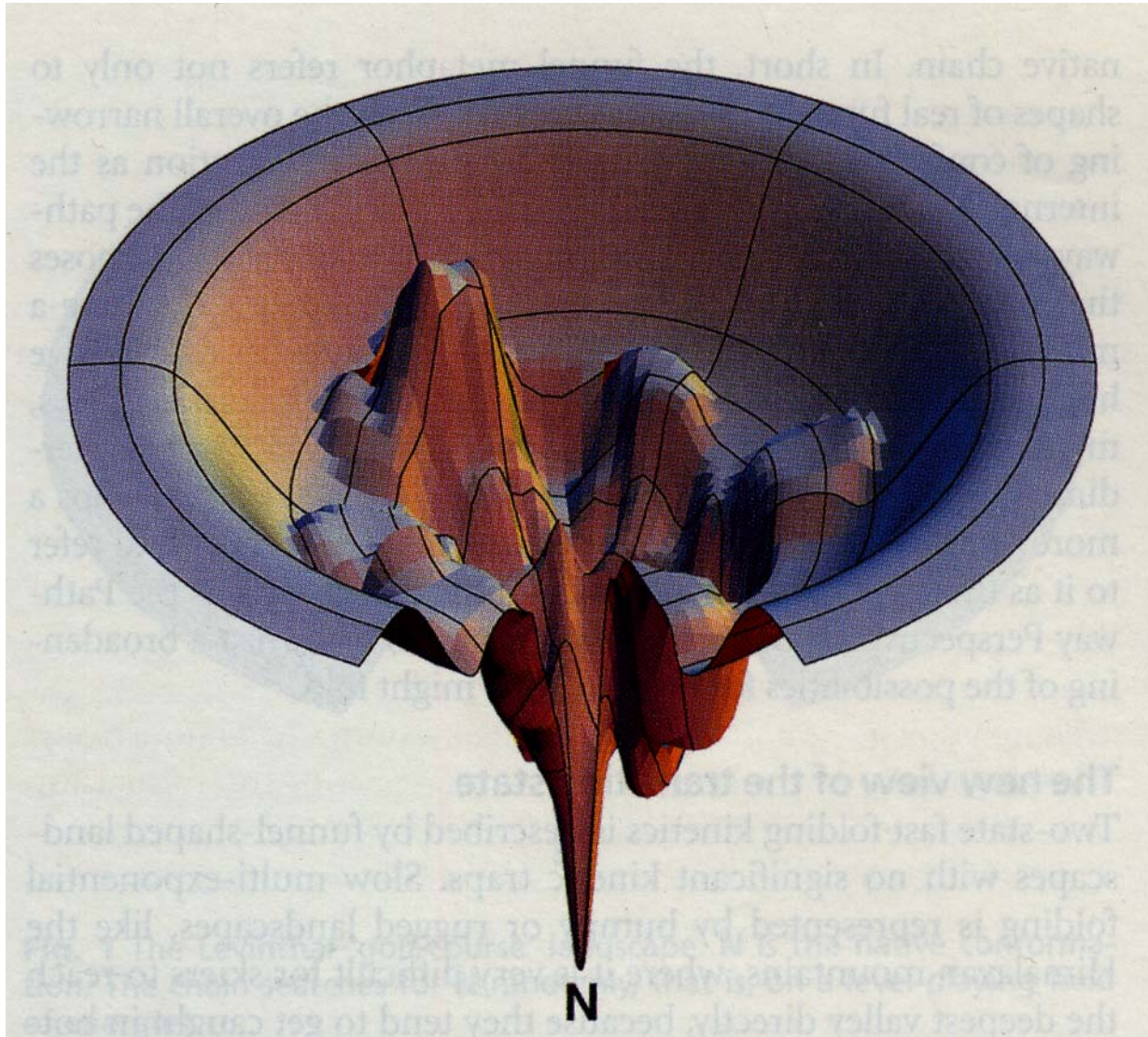
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19



The funnel landscape

Solution to Levinthal’s paradox

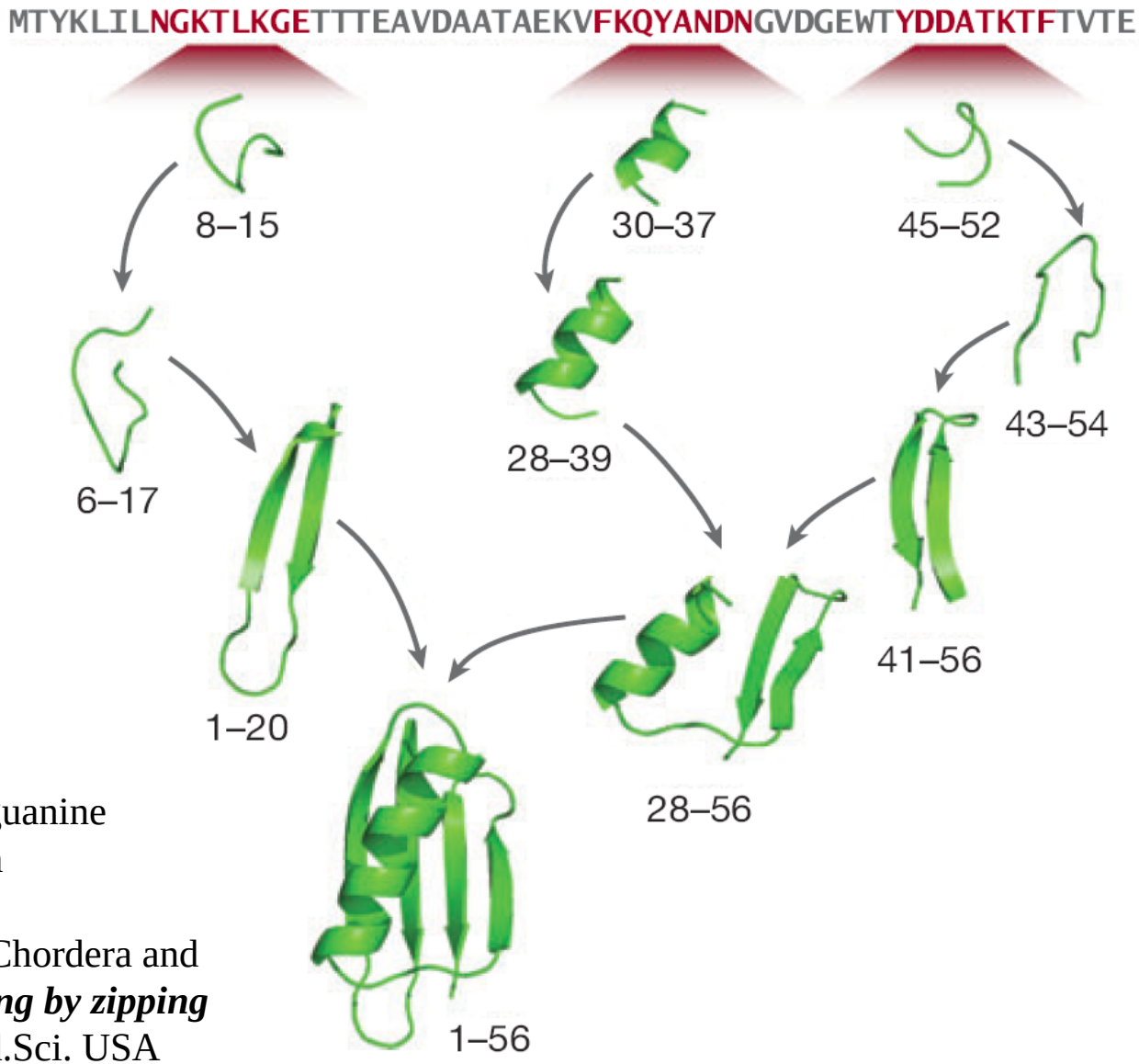
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19



The structured funnel landscape

Solution to Levinthal’s paradox

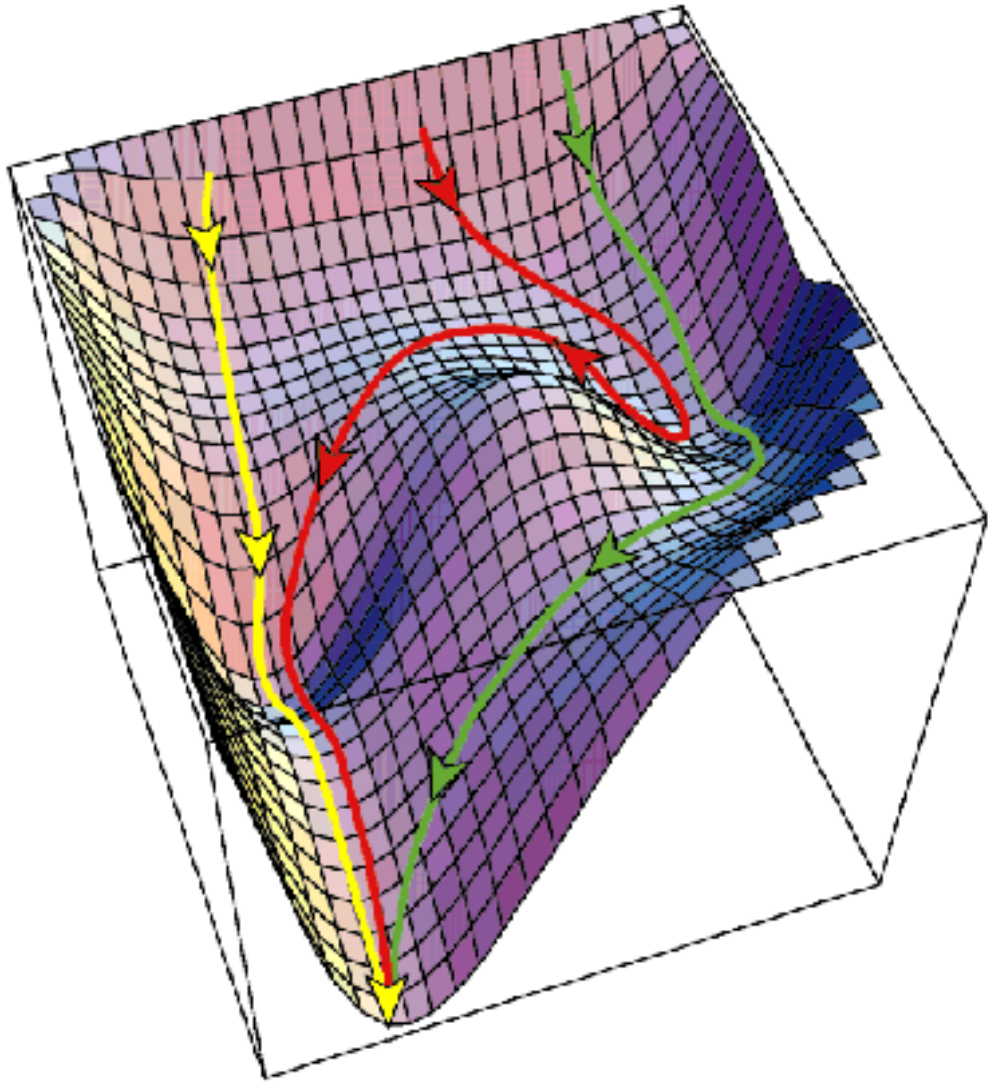
Picture: K.A. Dill, H.S. Chan, Nature Struct. Biol. 4:10-19



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.

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

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

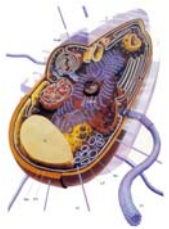
Genotype, Genome

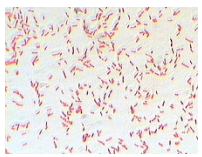
Collection of genes

Unfolding of the genotype

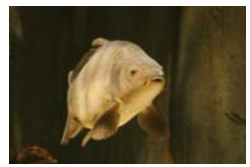
Developmental program

Highly specific environmental conditions












Evolution explains the origin of species and their interactions

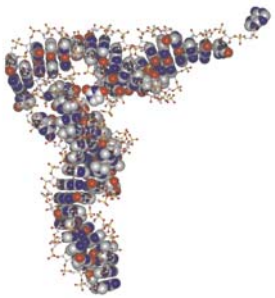



Genotype, Genome

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.

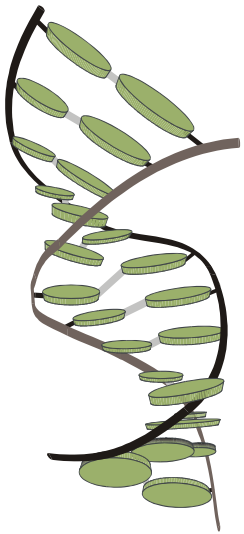
The exciting RNA story

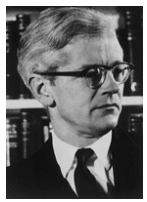
Biochemistry
molecular biology
structural biology
molecular evolution
molecular genetics
systems biology
bioinfomatics
epigenetics

Unfolding of the genotype

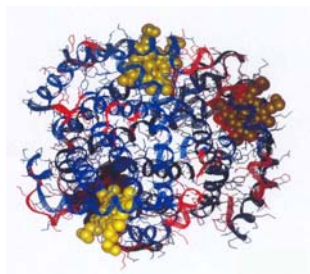
Highly specific environmental conditions

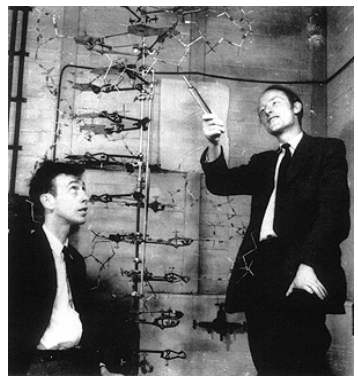
Phenotype





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

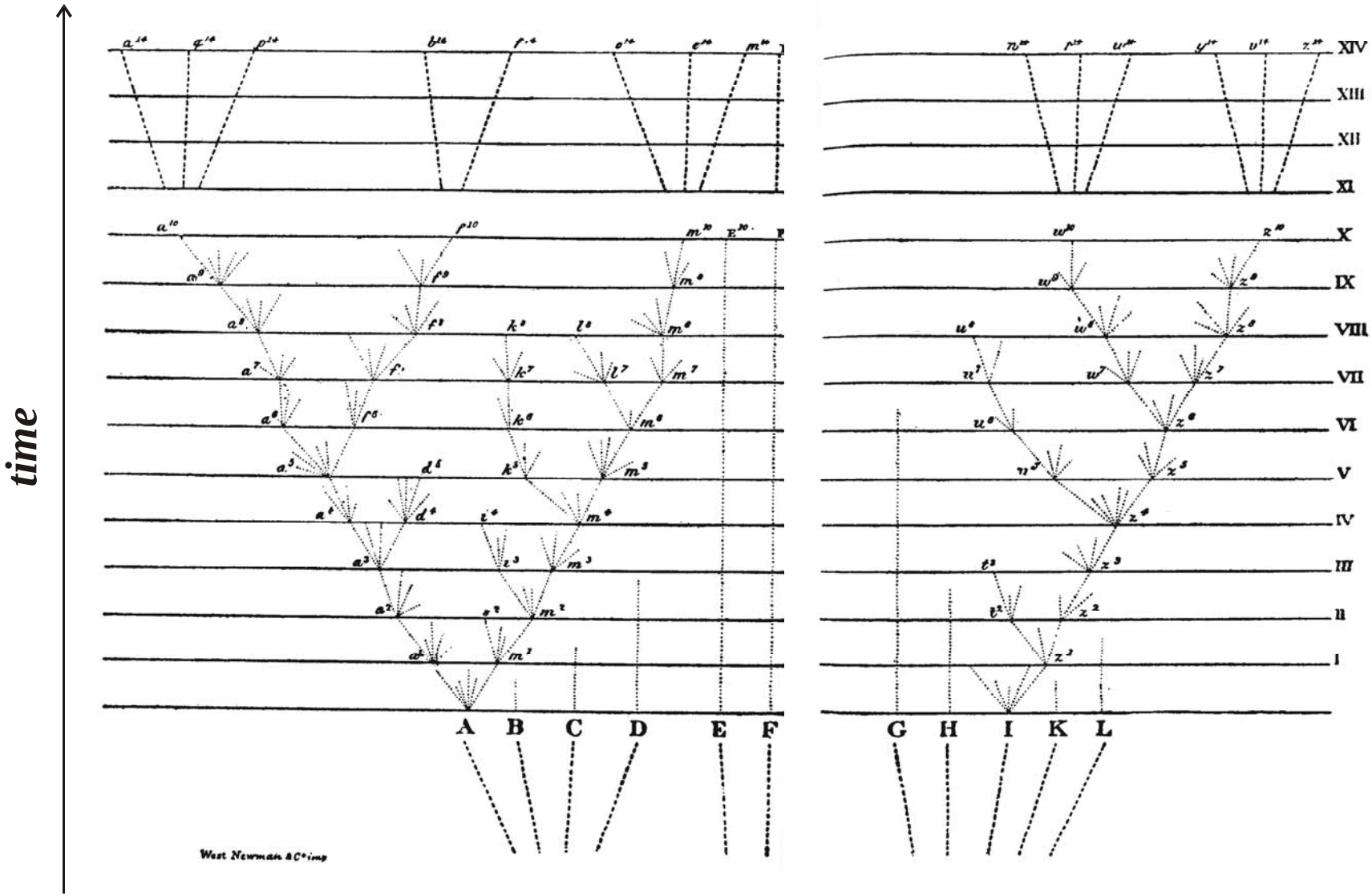
Three necessary conditions for Darwinian evolution are:

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

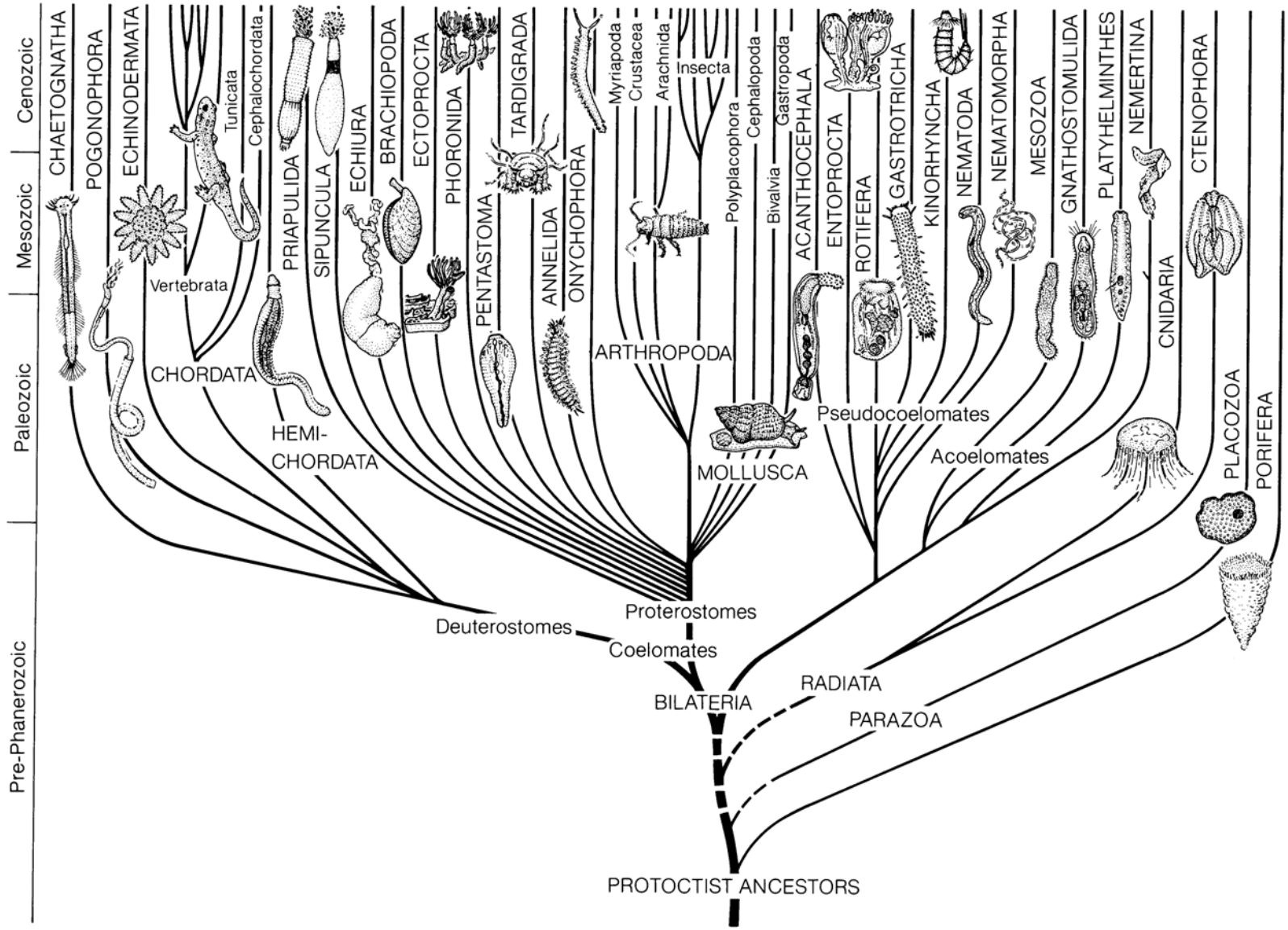
Variation through mutation and recombination operates on the **genotype** whereas the **phenotype** is the target of **selection**.

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**.

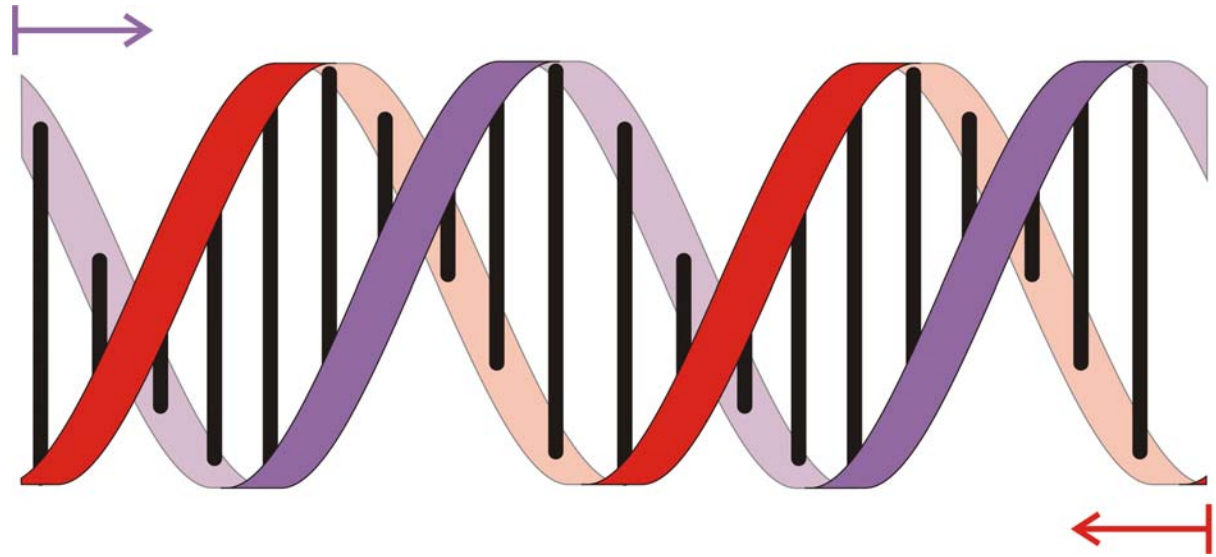
All conditions can be fulfilled not only by cellular organisms but also by **nucleic acid molecules** in suitable **cell-free experimental assays**.

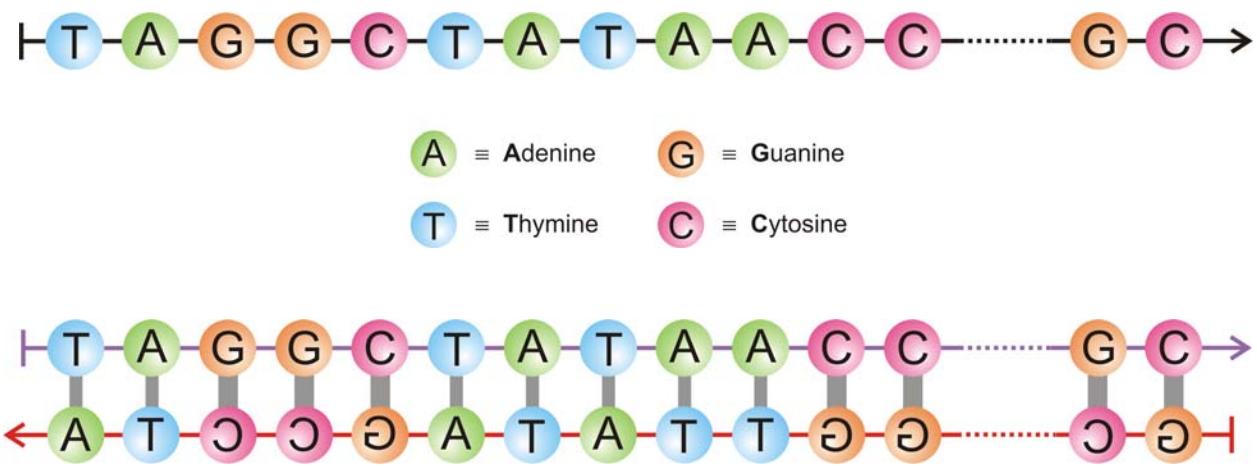


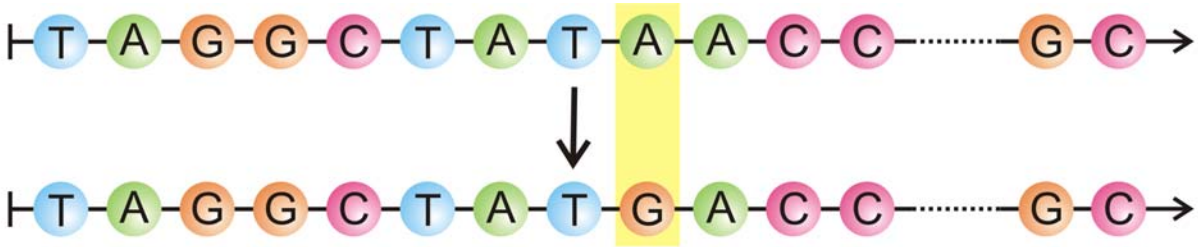
Charles Darwin, *The Origin of Species*, 6th edition.
Everyman‘s Library, Vol.811, Dent London, pp.121-122.



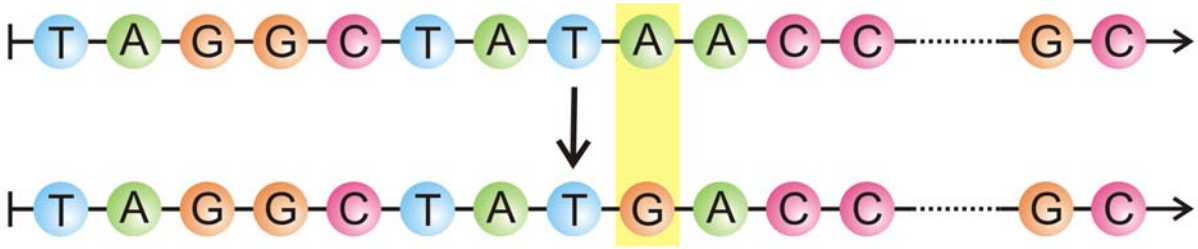
Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth*. W.H. Freeman, San Francisco, 1982.



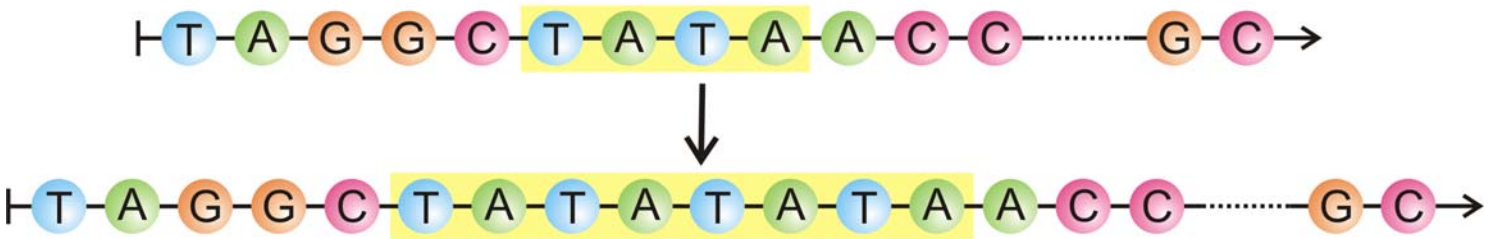

T A G G C T A T A A C C G C
A ≡ Adenine
G ≡ Guanine
T ≡ Thymine
C ≡ Cytosine
T A G G C T A T A A C C G C
A T C C G A T A T T G G C G



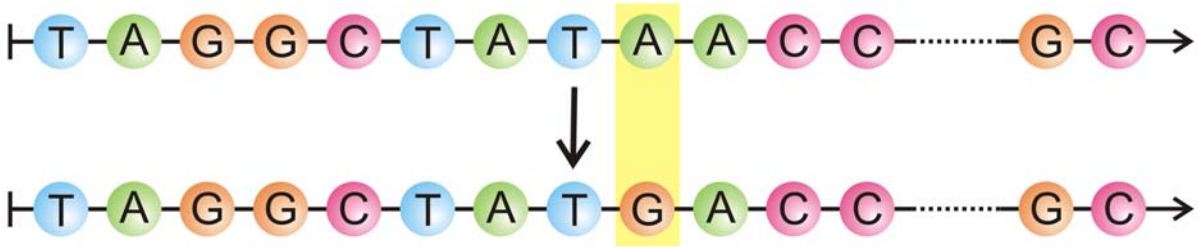
Point mutation



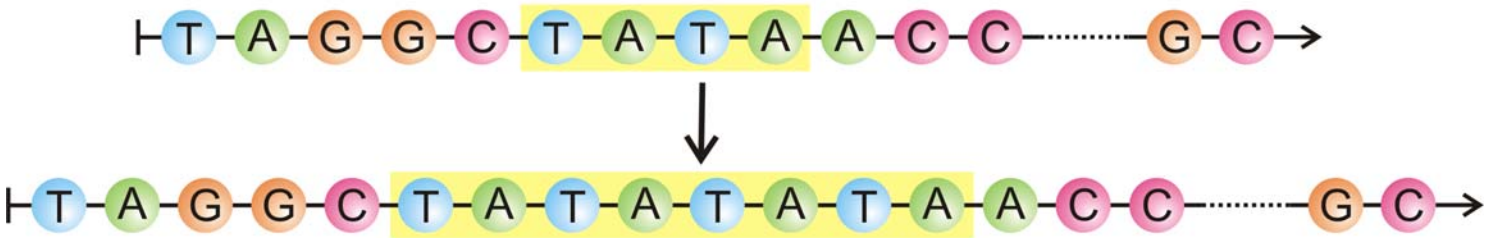
Point mutation



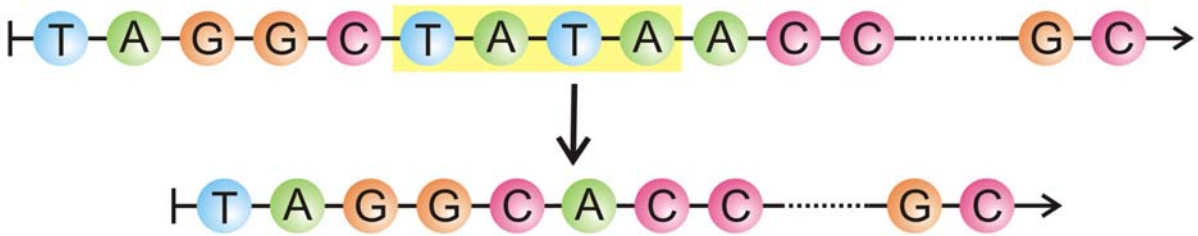
Insertion



Point mutation



Insertion



Deletion

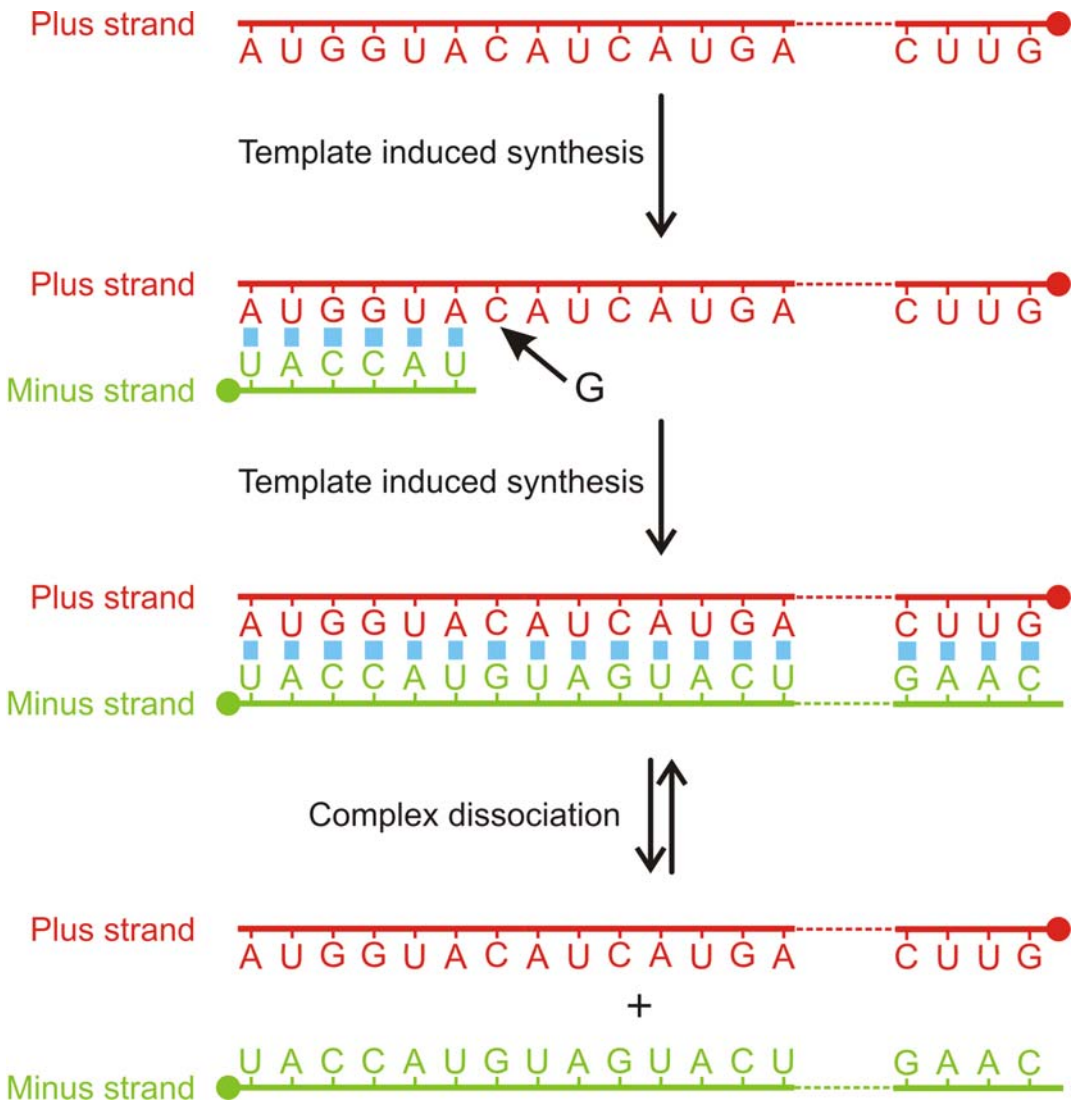
Reconstruction of phylogenies through comparison of molecular sequence data

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.

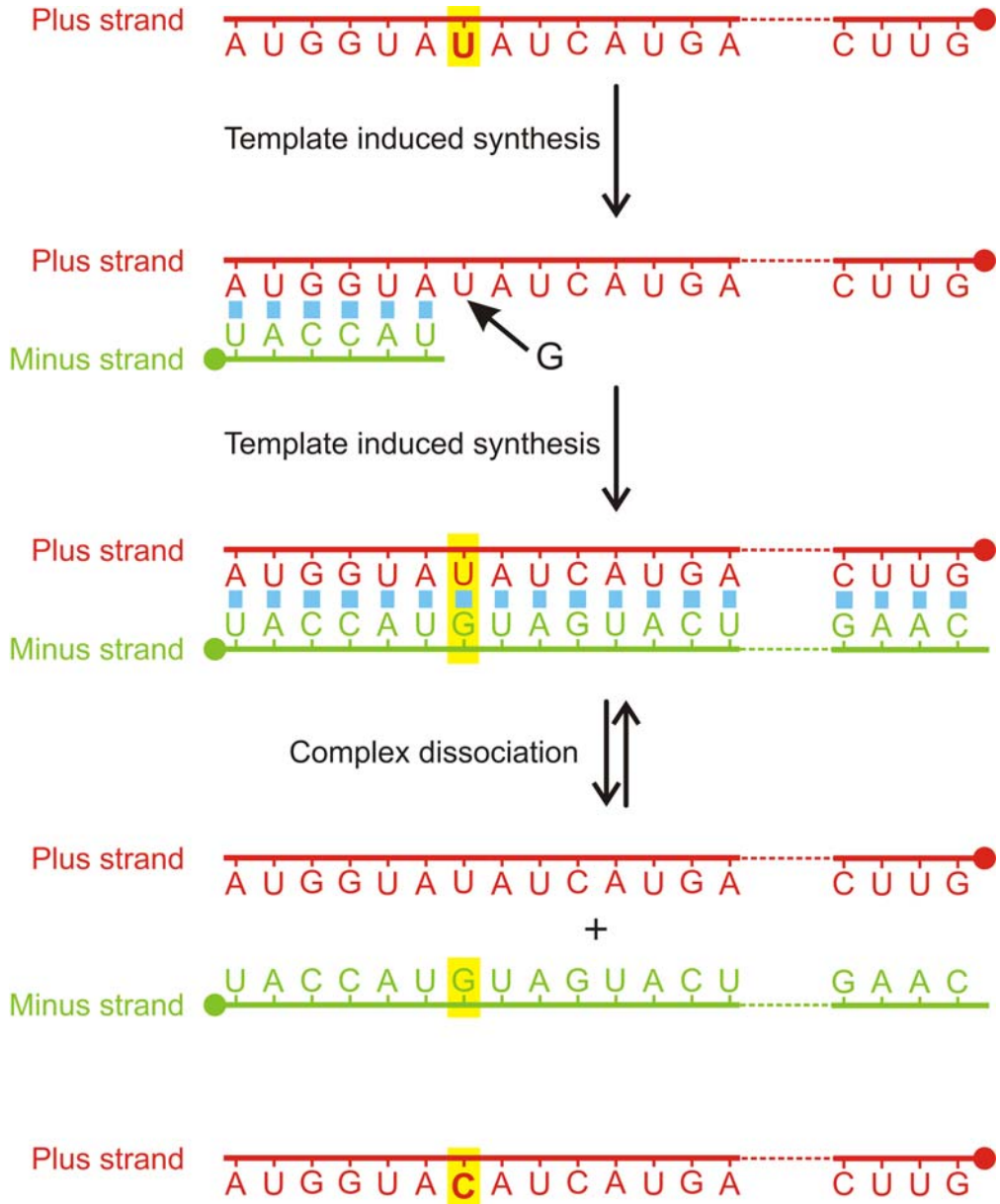
1. Biology and probabilities
2. Evolution – organismic and molecular
3. **Multiplication, mutation, and selection**
4. Rational design of molecules
5. Evolution and optimization of molecules
6. Origin of biological complexity

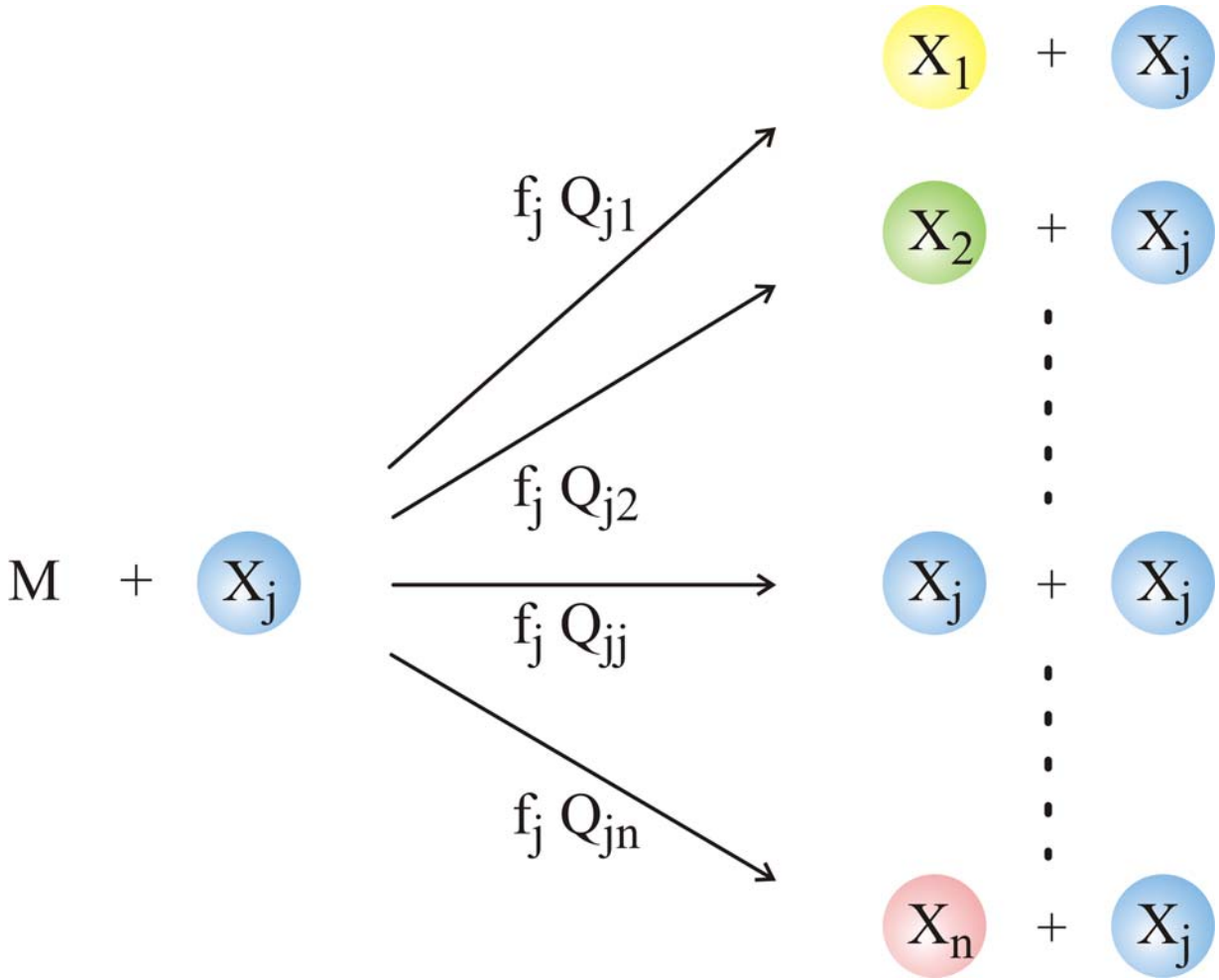


Complementary replication is the simplest copying mechanism of RNA.
Complementarity is determined by Watson-Crick base pairs:

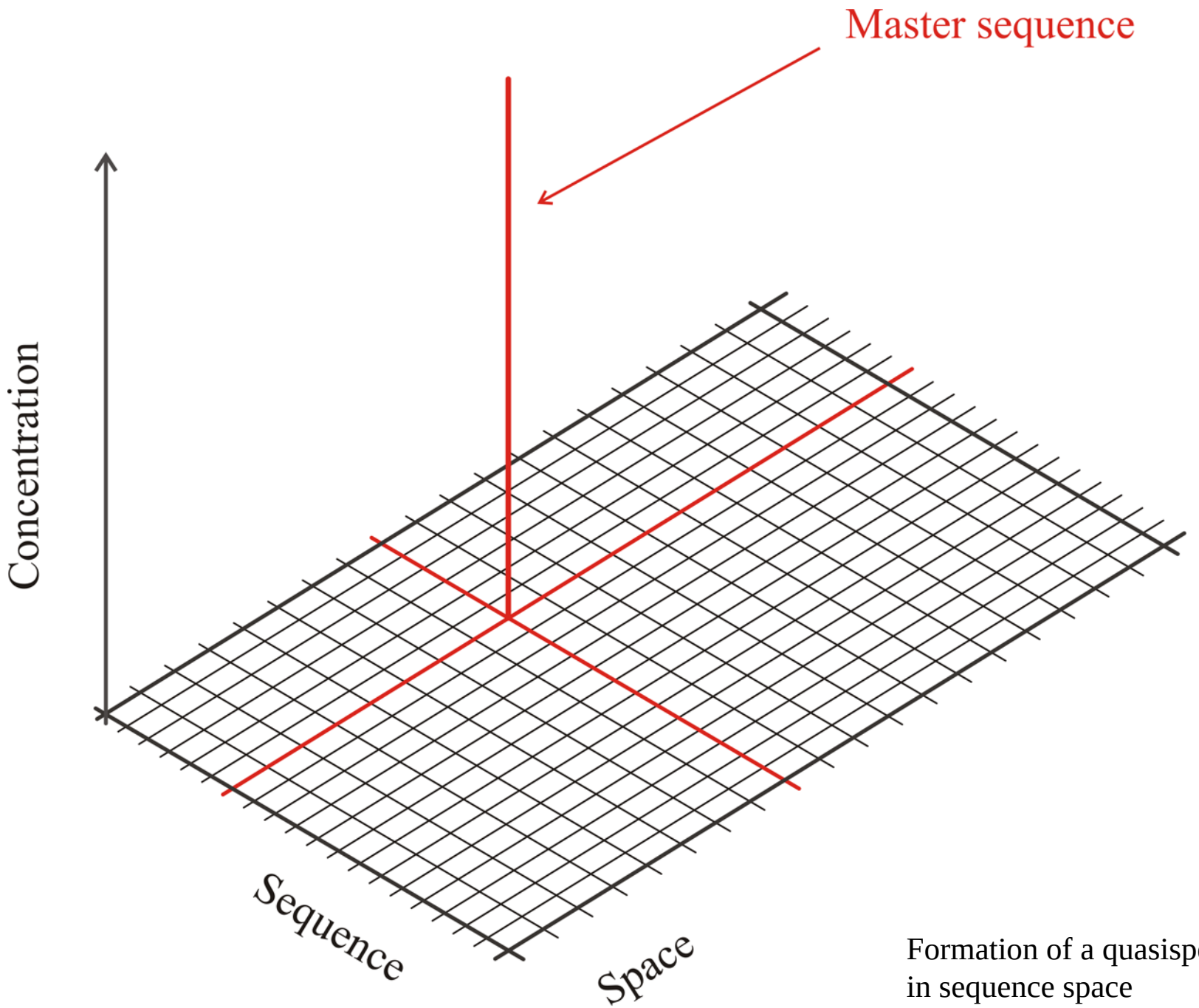
G≡C and **A=U**

Mutation as an error
in replication

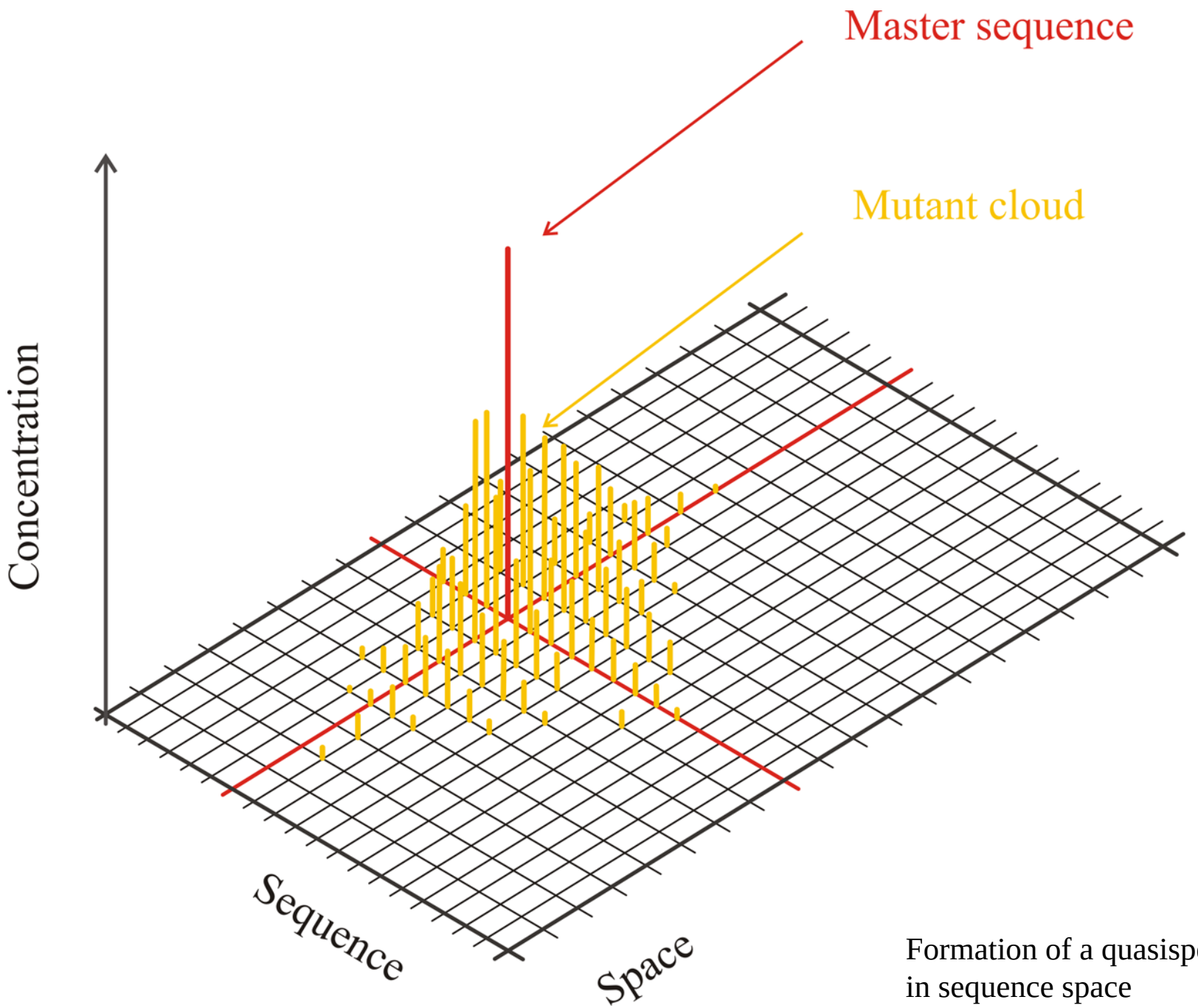




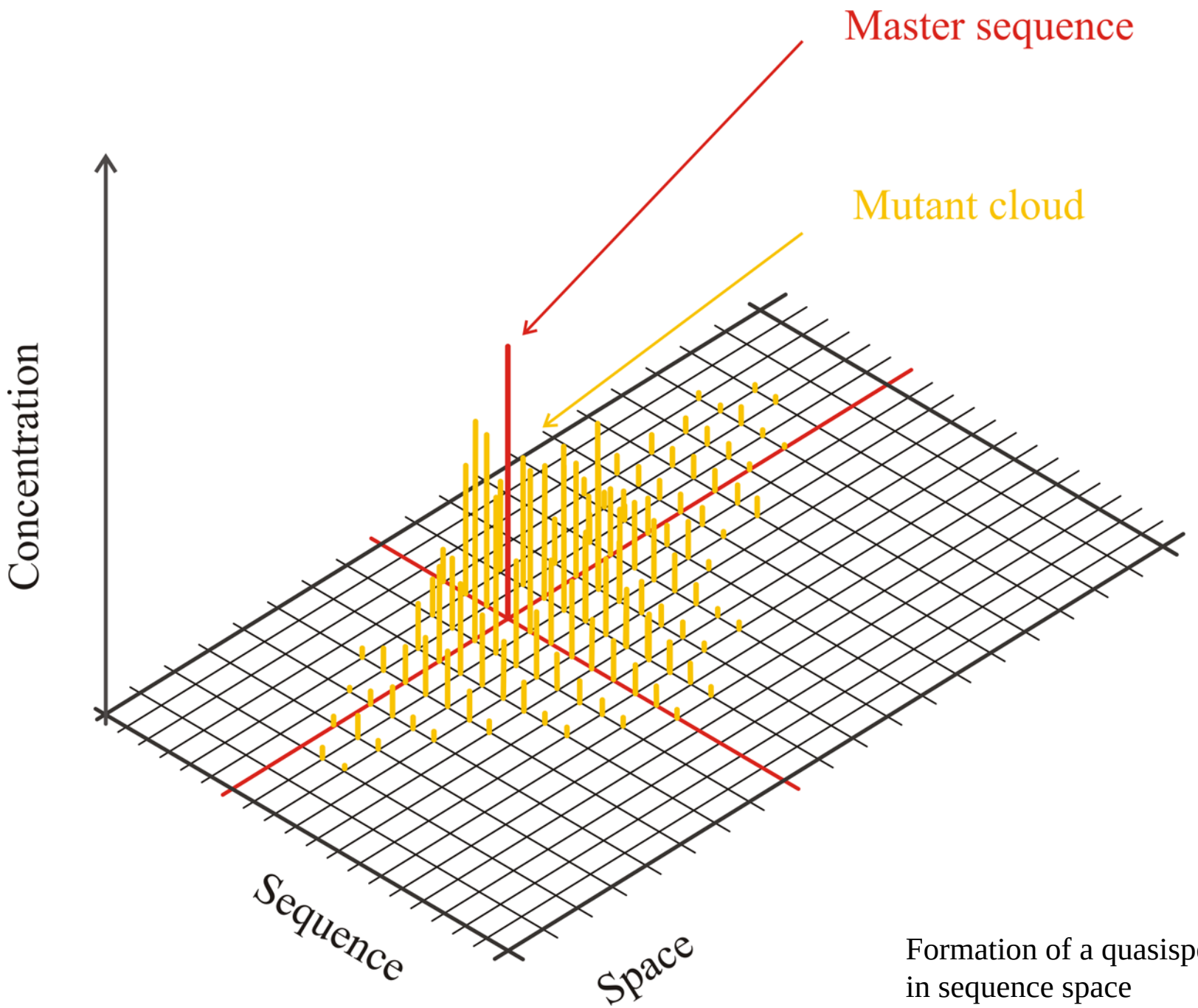
Chemical kinetics of replication and mutation as parallel reactions



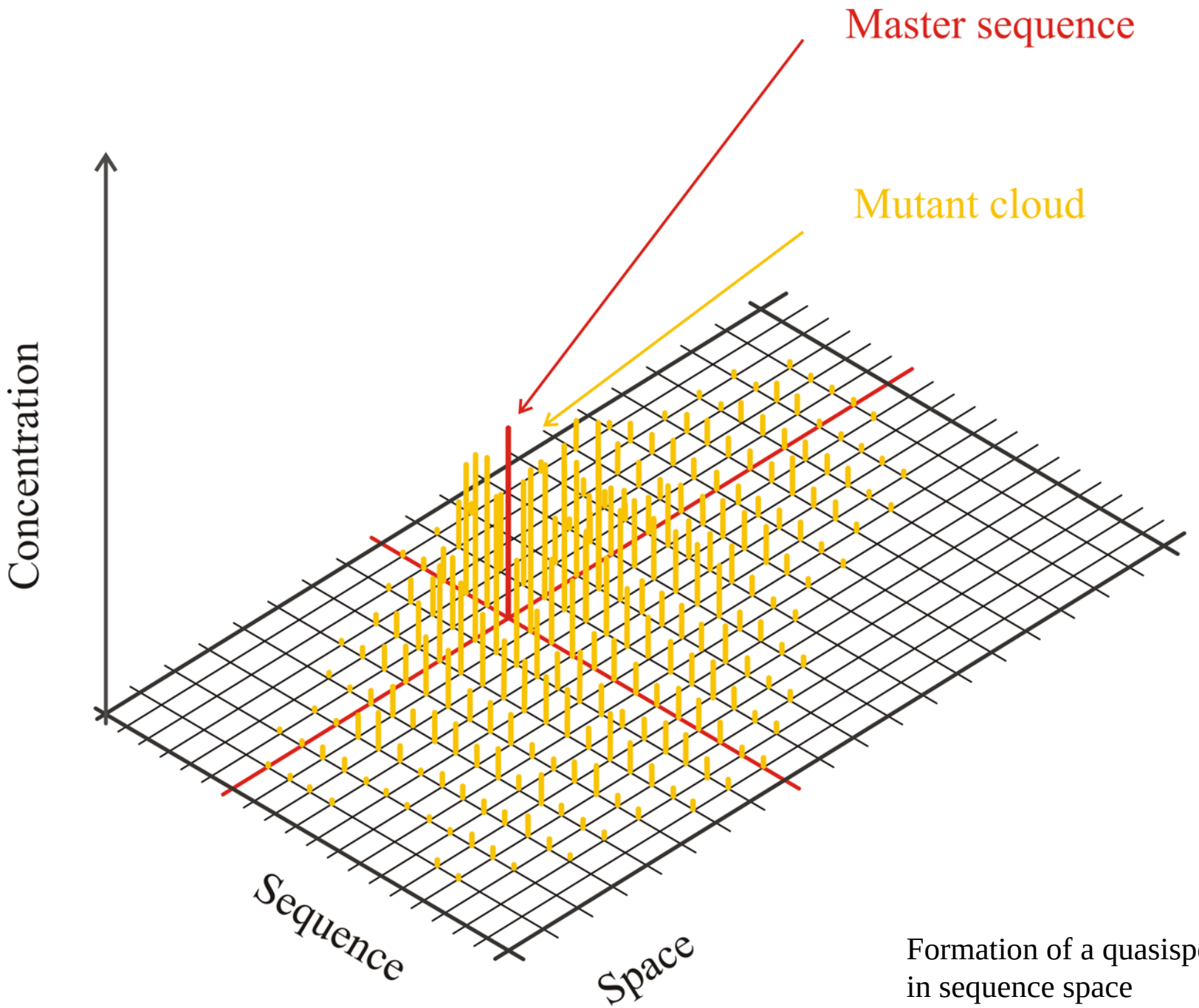
Formation of a quasispecies in sequence space



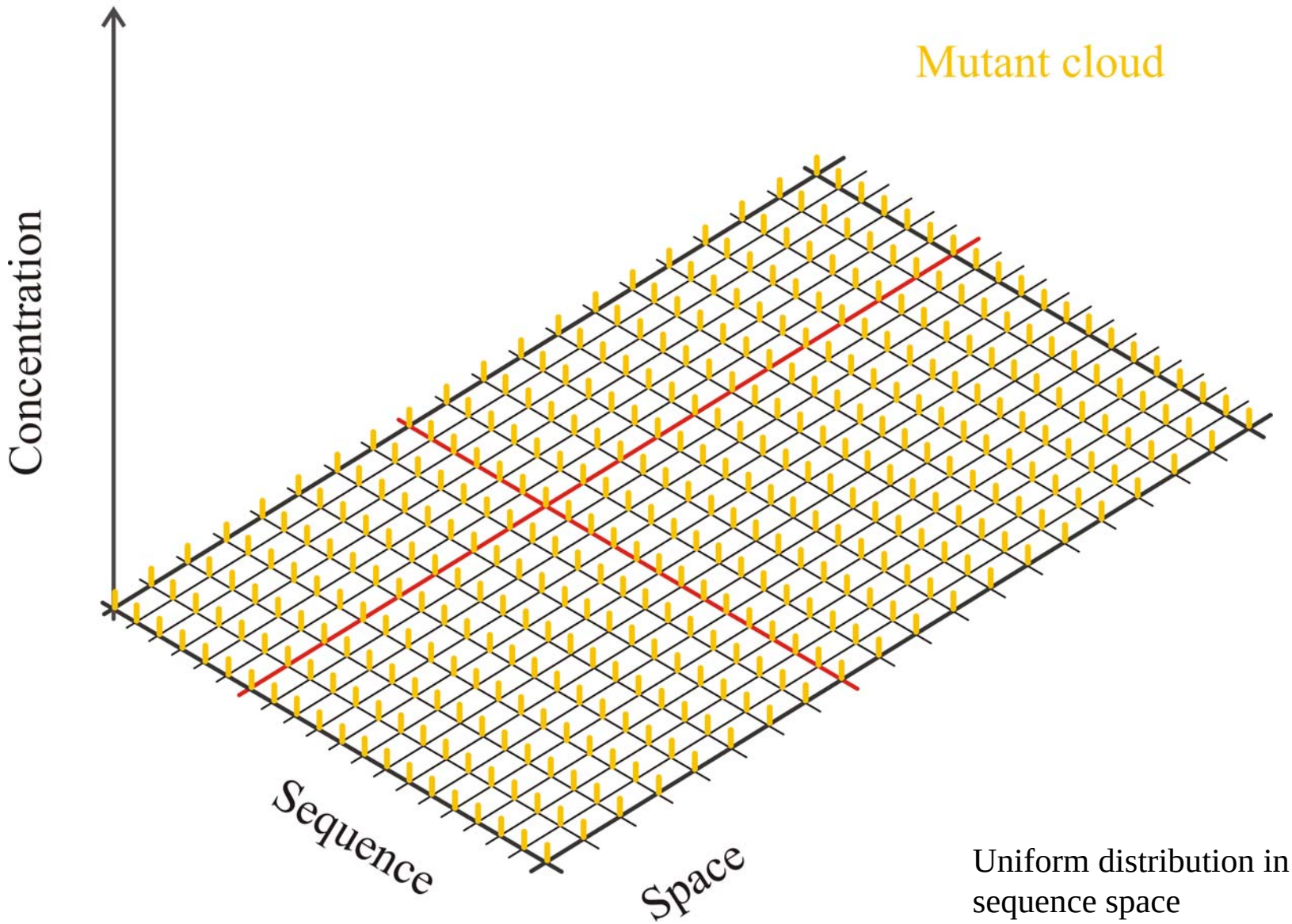
Formation of a quasispecies in sequence space



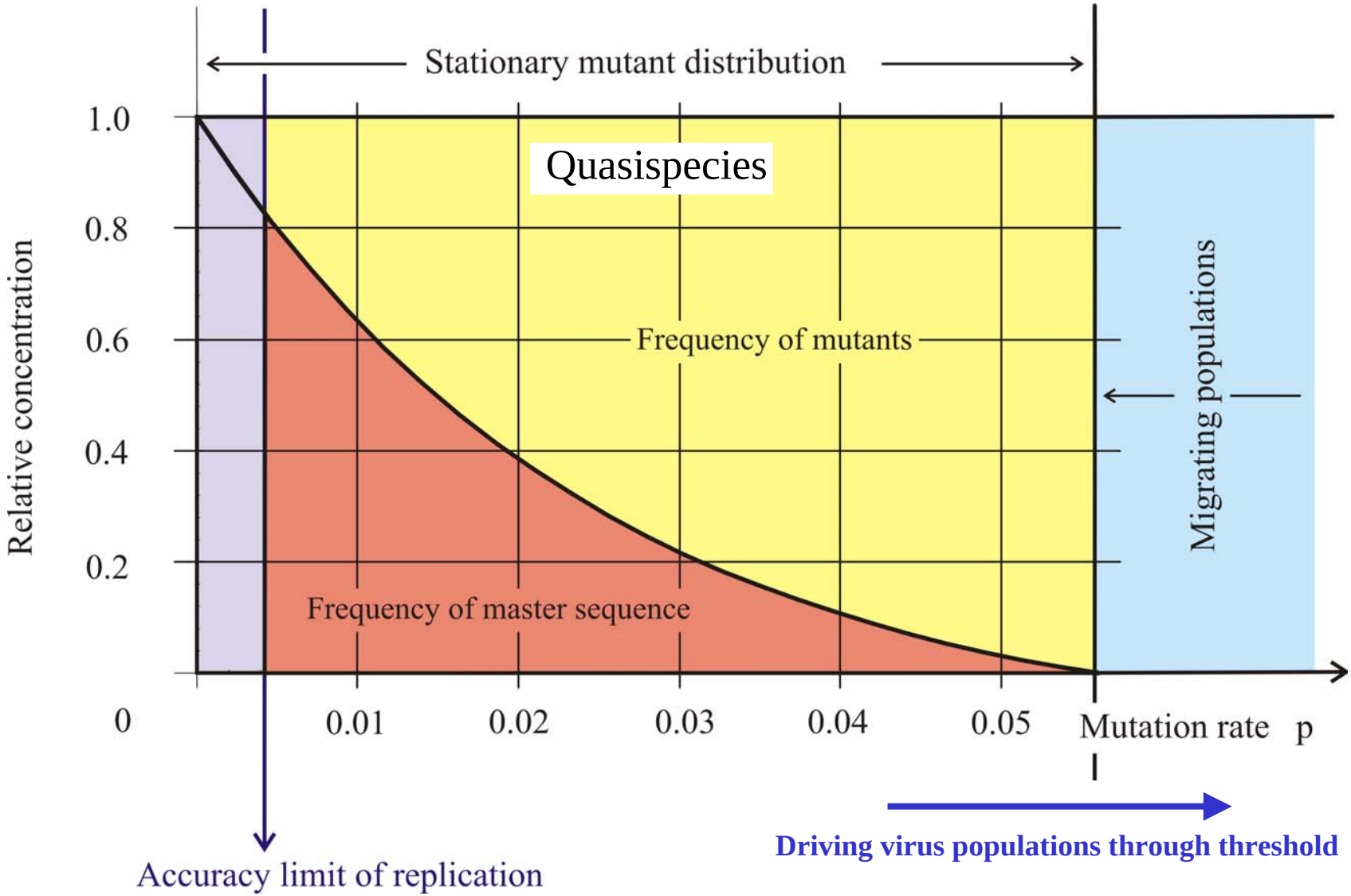
Formation of a quasispecies in sequence space



Formation of a quasispecies in sequence space



Uniform distribution in sequence space



The error threshold in replication

Results from the kinetic theory of molecular evolution:

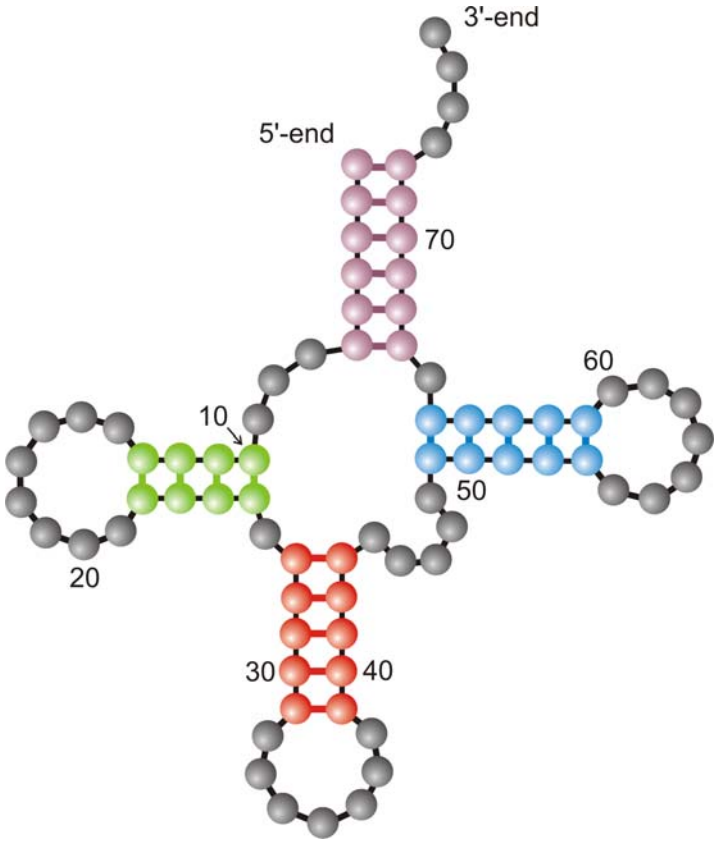
• Replicating ensembles of molecules form stationary populations called *quasispecies*, which represent the genetic reservoir of asexually reproducing species.

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

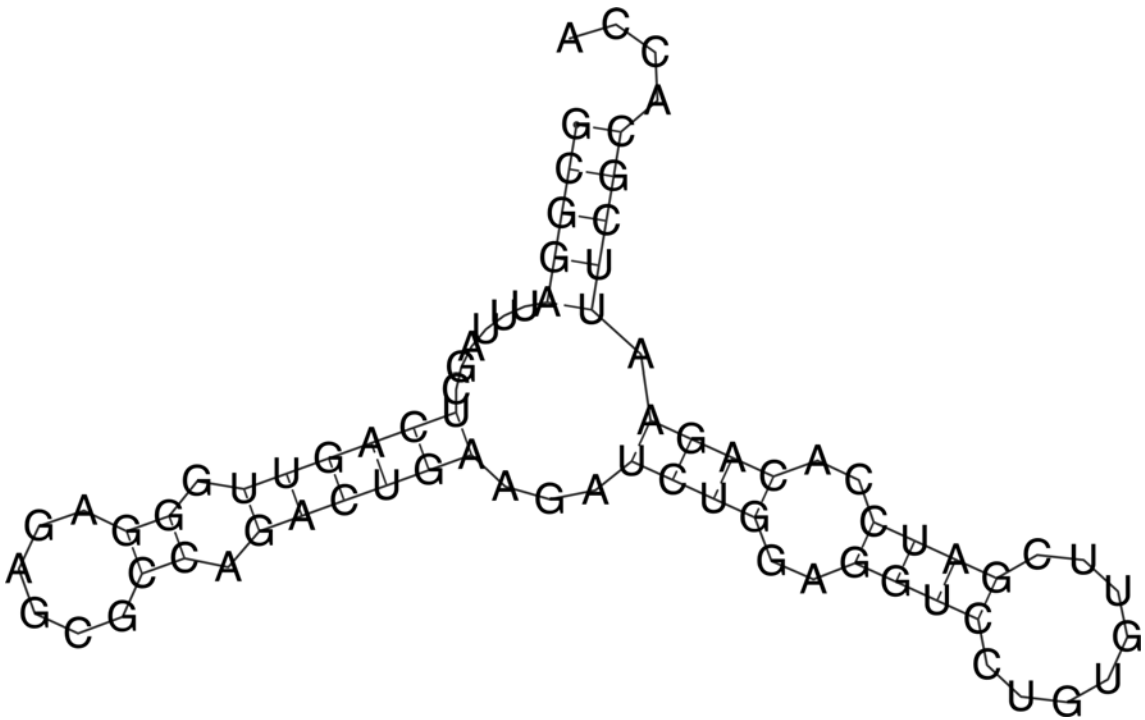
GCGGAUUUAGCUCAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCCUGUGTPCGAUCCACAGAAUUCGCACCA



ΔG = -20.20 kcal/mol

Sequence and structure of phenylalanyl-transfer-RNA

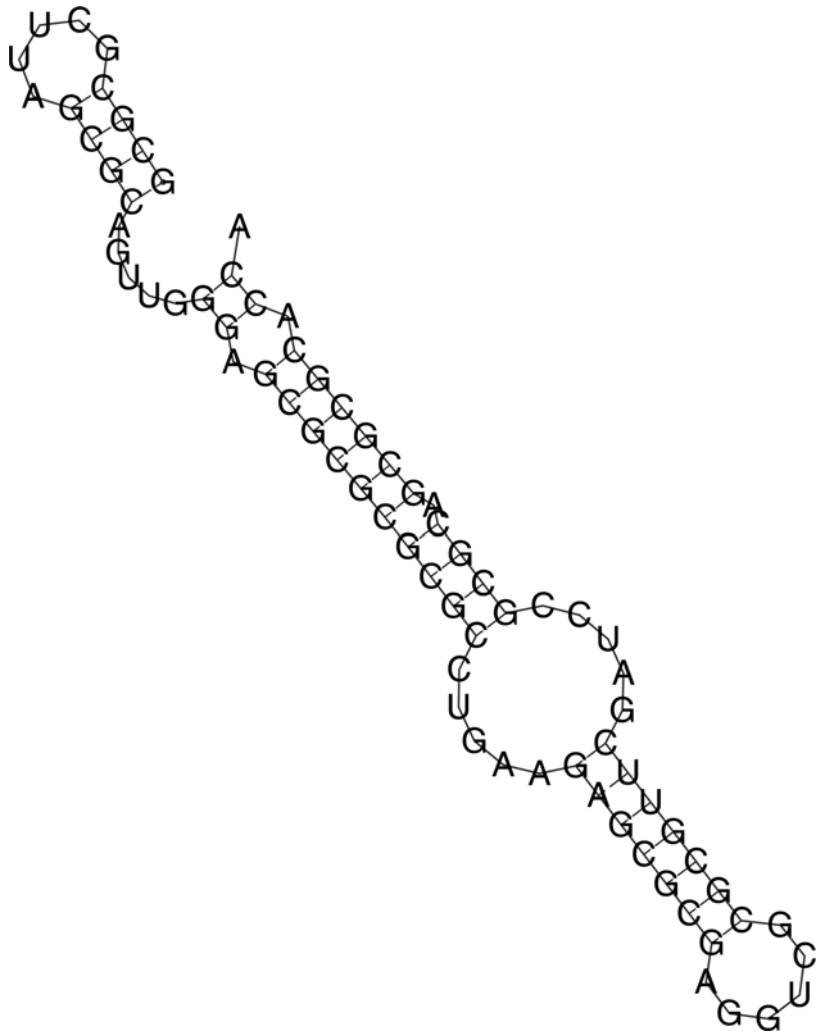
GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA



ΔG = -22.90 (-21.90) kcal/mol

GCGCGCUUAGCGCAGUUGGGAGCGCGCGCGCCUGAAGAGCGCGAGGUCGCGCGUUCGAUCCGCGCAGCGCGCACCA

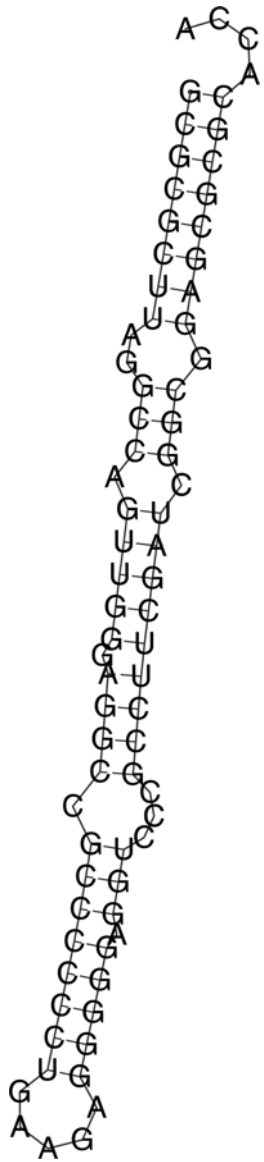
1. Trial



ΔG = -43.10 (-36.40) kcal/mol

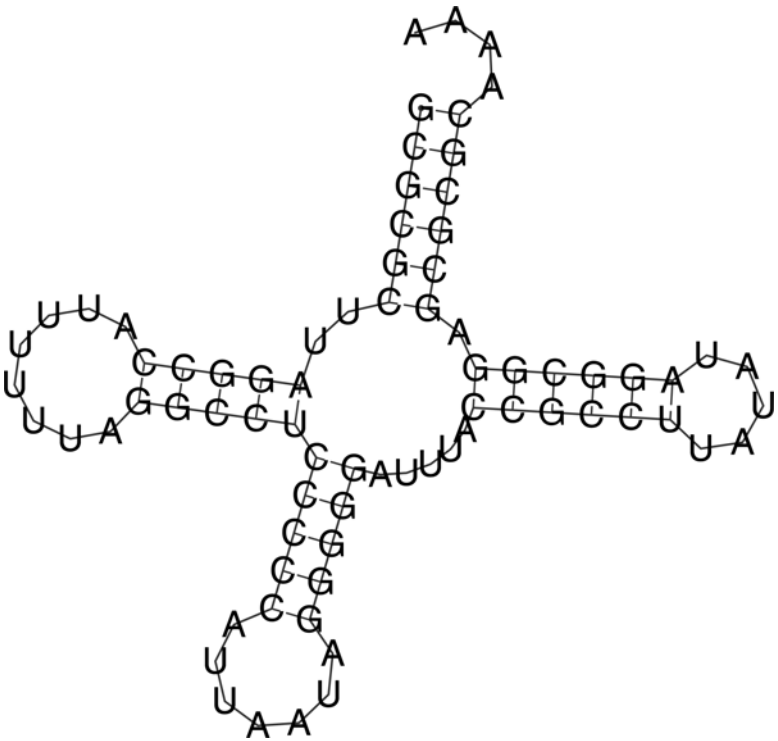
GCGCGCUUAGGCCAGUUGGGAGGCCGCCCCCCUGAAGAGGGGGAGGUCCCGCCUUCGAUCGGCGGAGCGCGCACCA

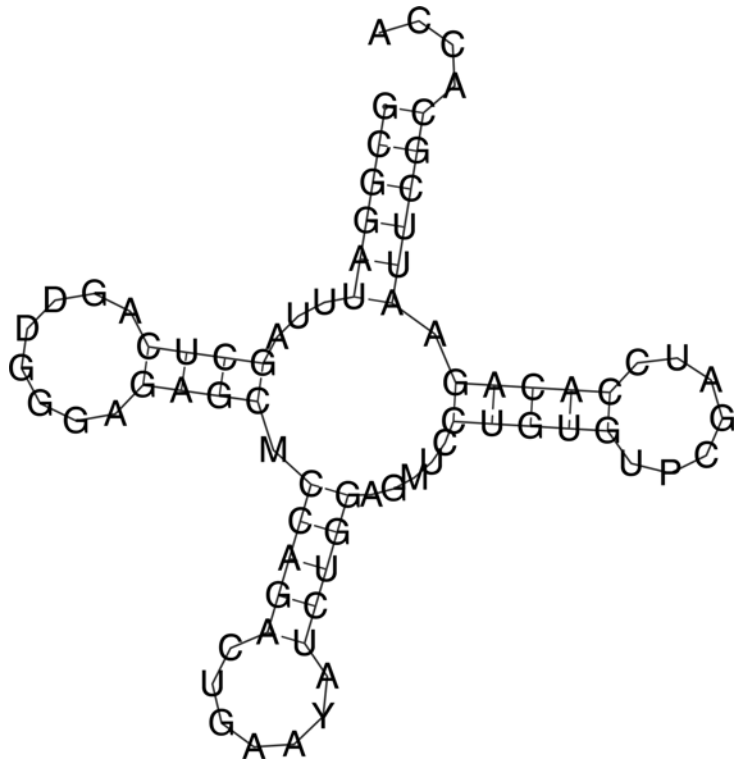
2. Trial



ΔG = -45.10 (-39.40) kcal/mol

GCGCGCUUAGGCCAUUUUUUAGGCCUCCCCCAUUAAUAGGGGGAUUUACCGCCUUAUAUAGGCGGAGCGCGCAAAA



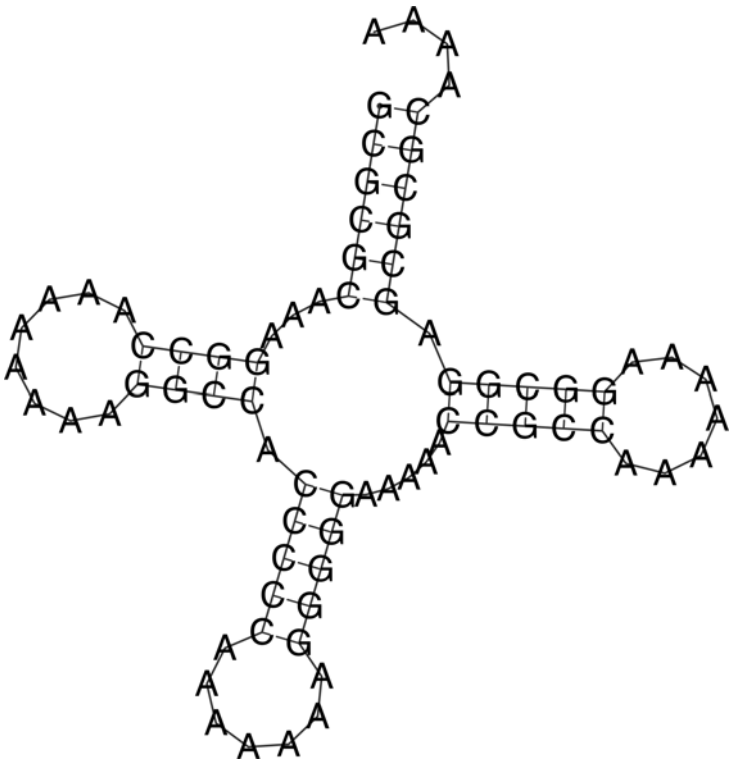


3. Trial

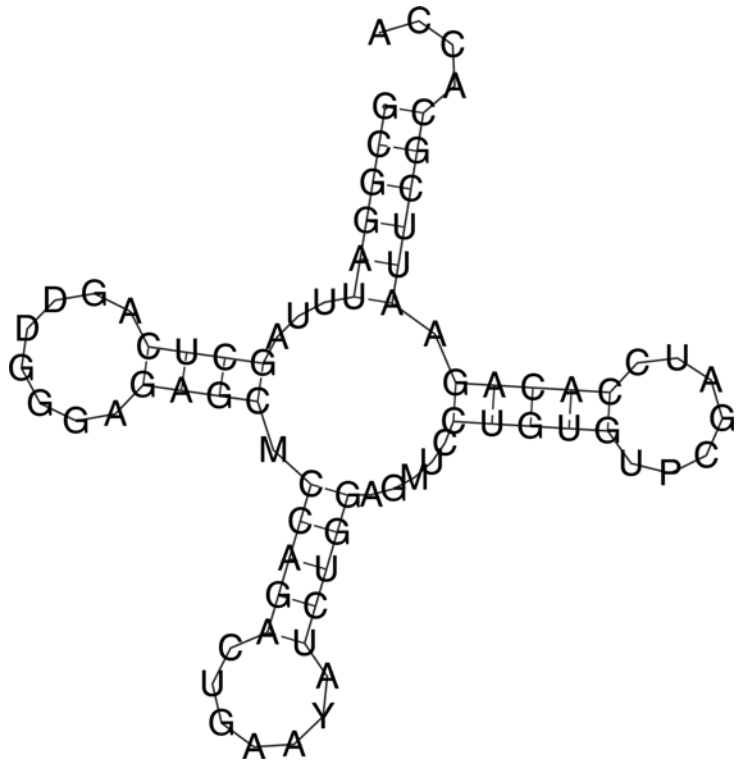
Target structure

ΔG = -41.80 (-39.90) kcal/mol

GCGCGCAAAGGCCAAAAAAAAGGCCACCCCCAAAAAAAGGGGGAAAAACCGCCAAAAAAAGGCGGAGCGCGCAAAA



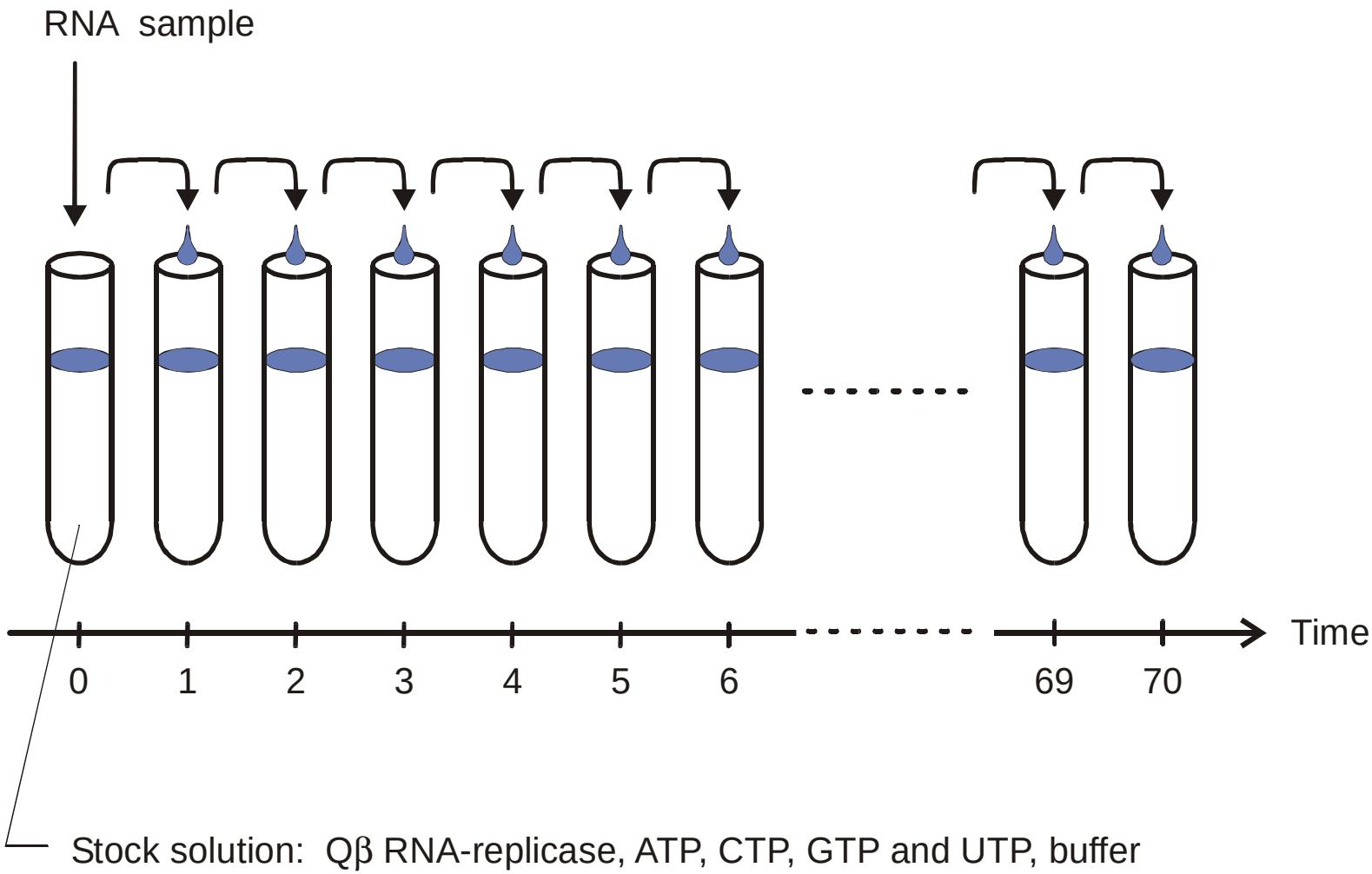
4. Trial



Target structure

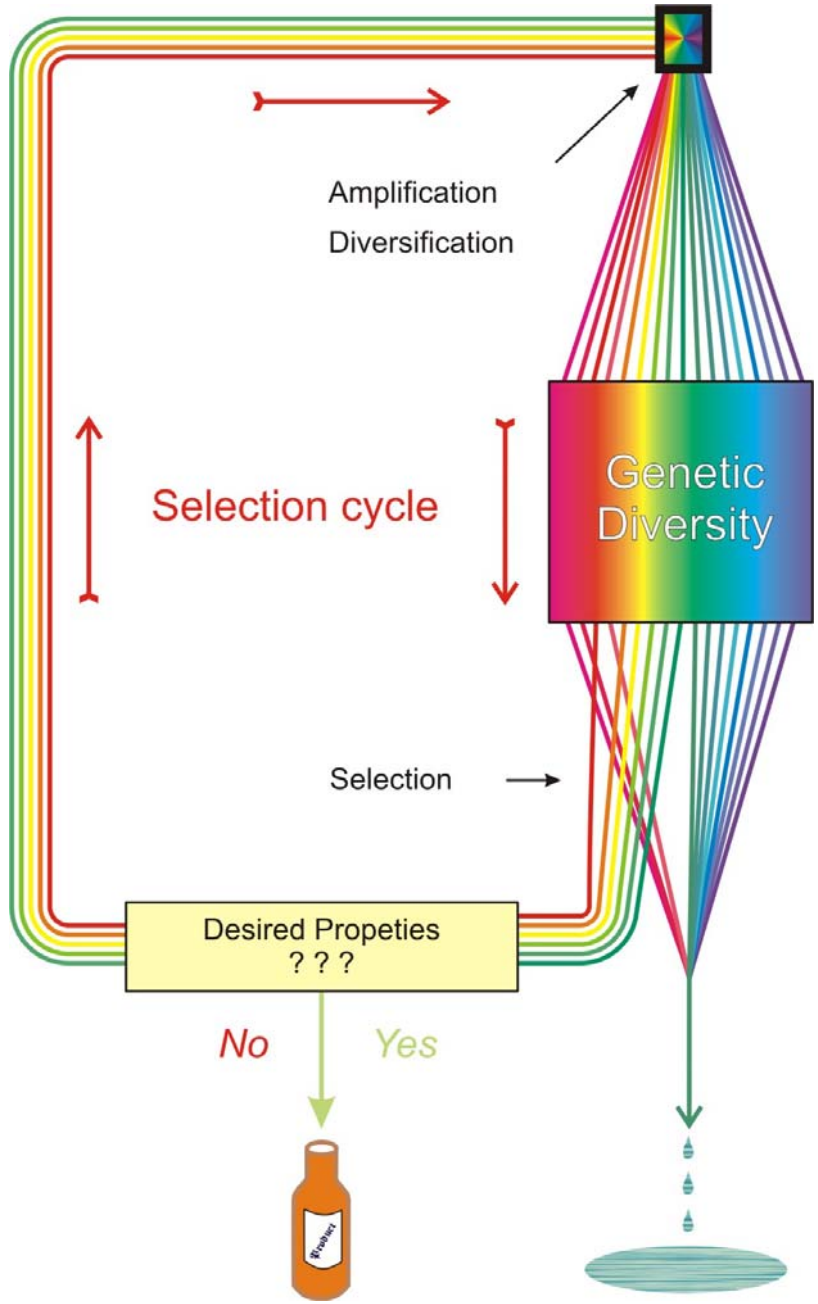
ΔG = -40.70 kcal/mol

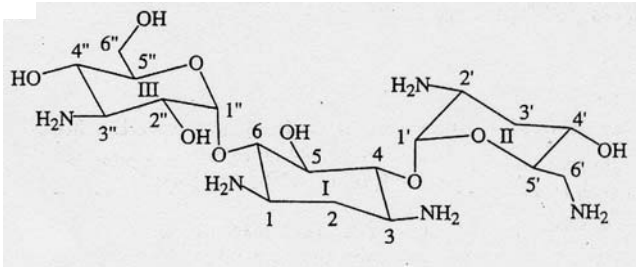
1. Biology and probabilities
2. Evolution – organismic and molecular
3. Multiplication, mutation, and selection
4. Rational design of molecules
5. **Evolution and optimization of molecules**
6. Origin of biological complexity



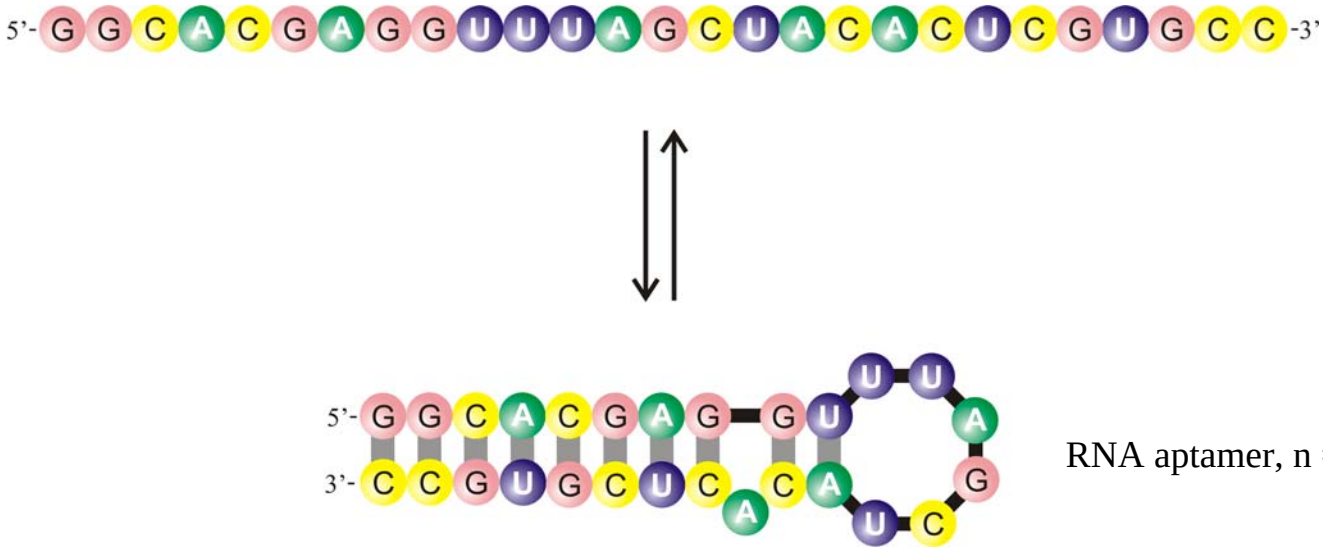
Application of the serial transfer technique to RNA evolution in the test tube

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





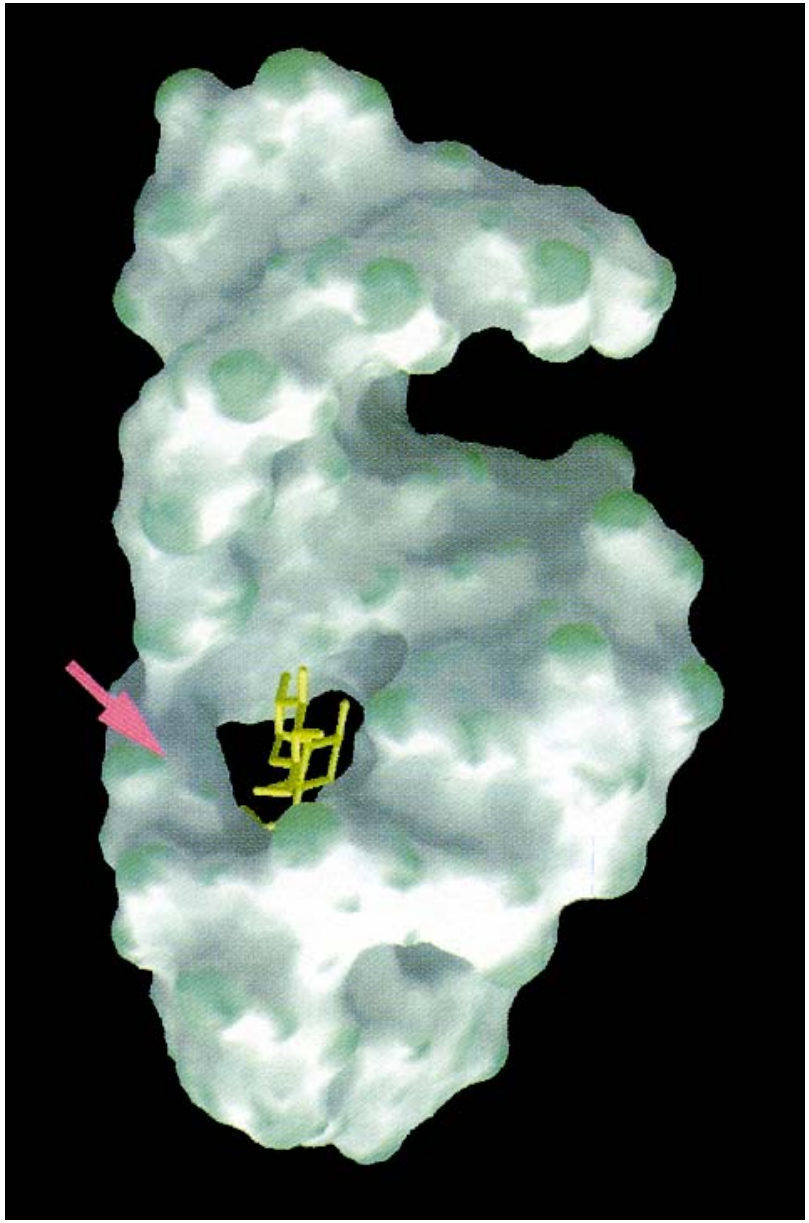
tobramycin



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)

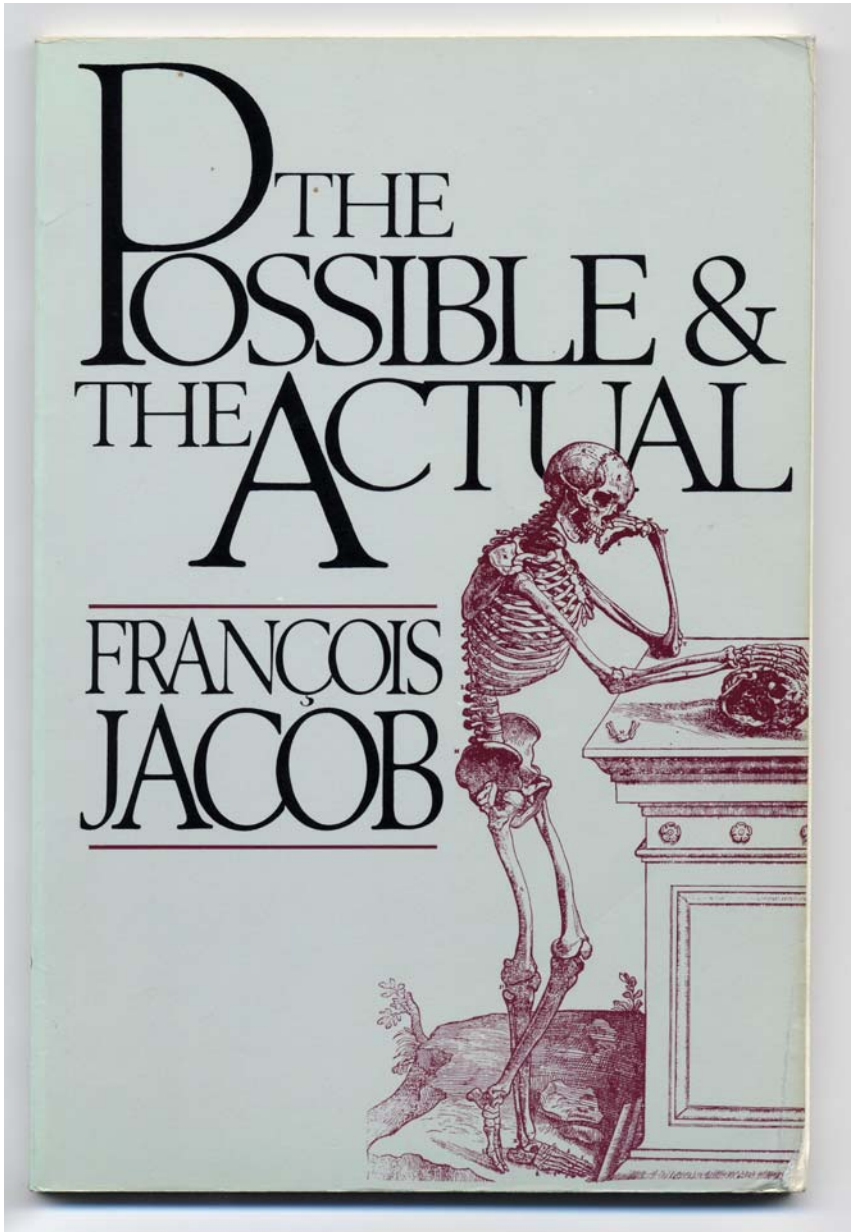
Results from laboratory experiments in molecular evolution:

- 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.

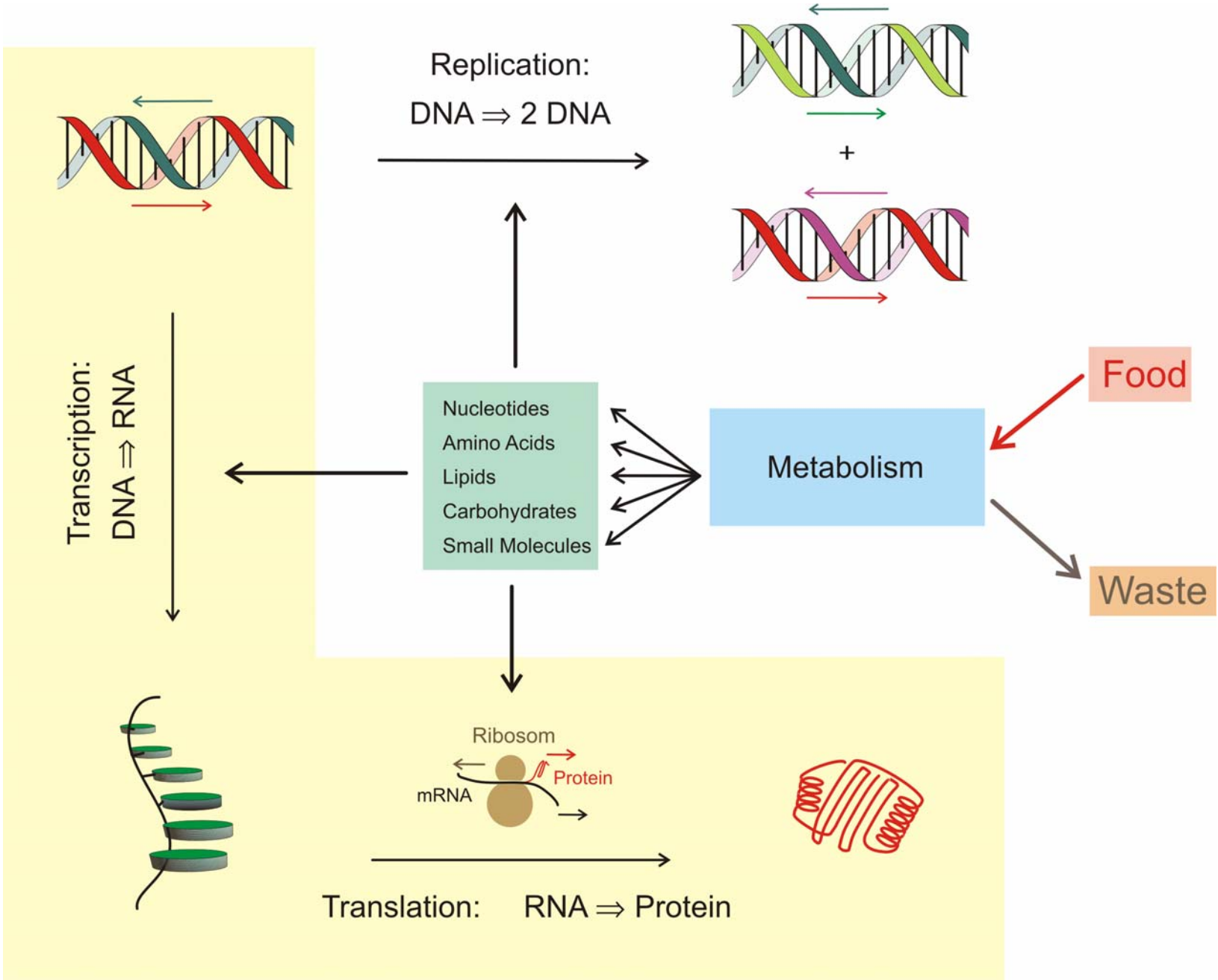
1. Biology and probabilities
2. Evolution – organismic and molecular
3. Multiplication, mutation, and selection
4. Rational design of molecules
5. Evolution and optimization of molecules
6. **Origin of biological complexity**

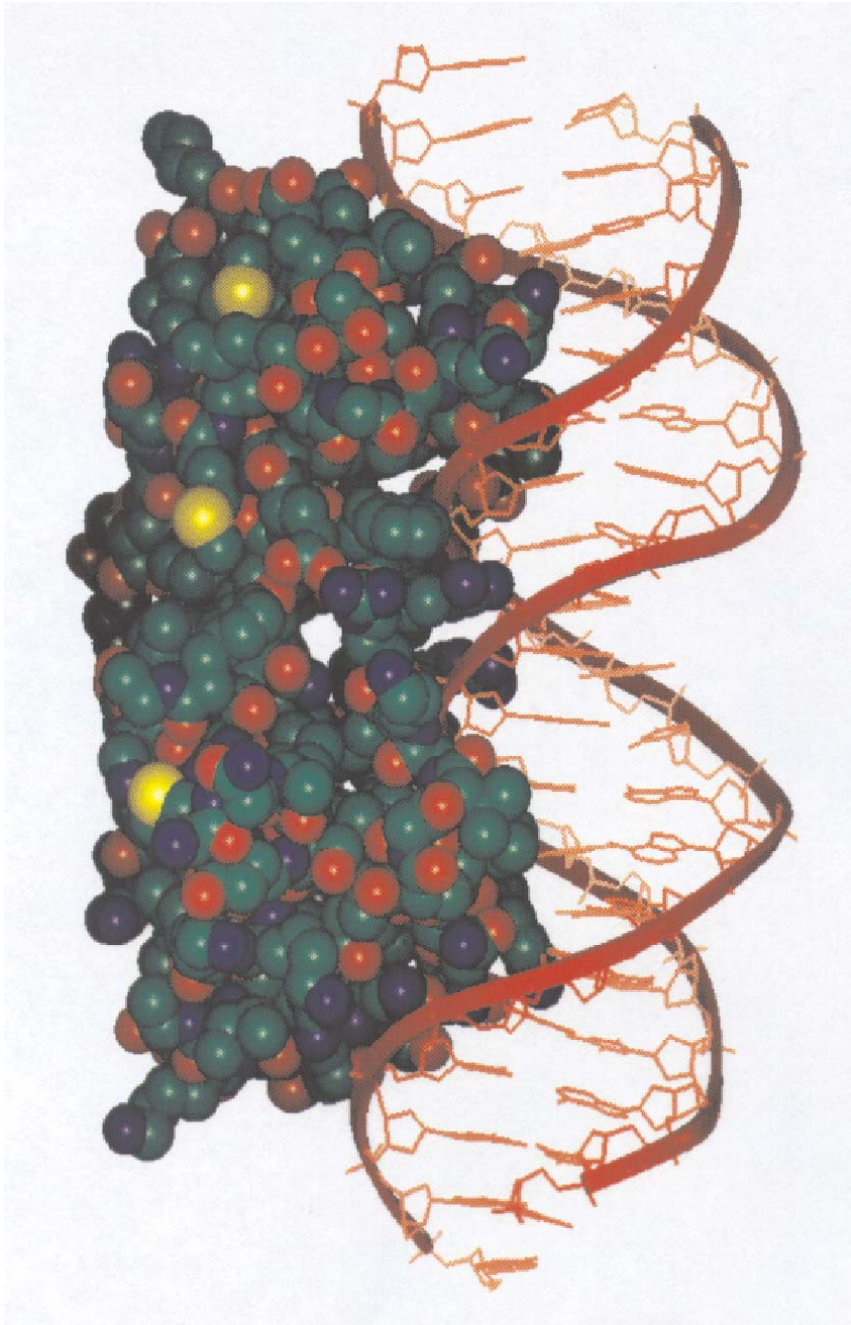


Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

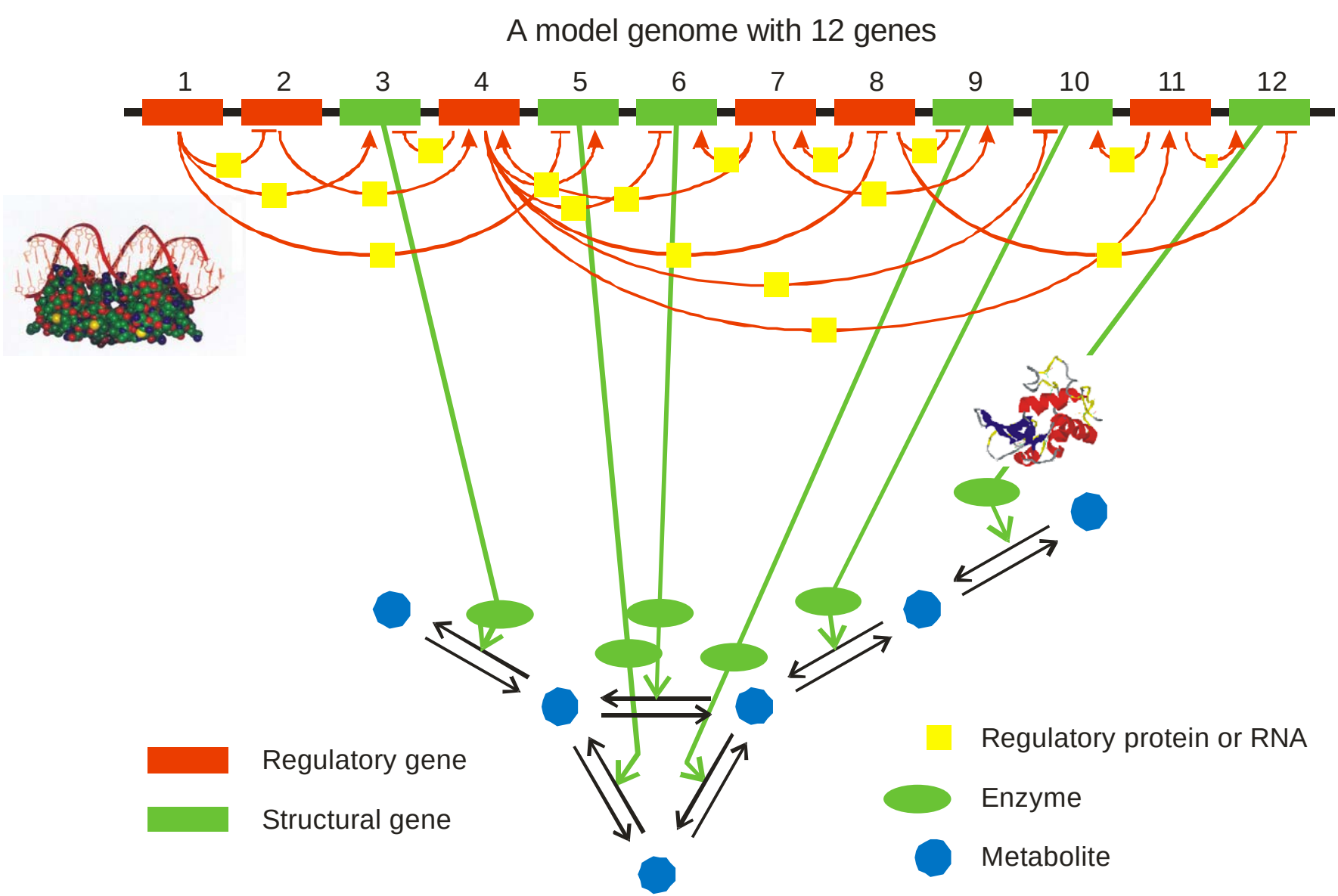
François Jacob. *The Possible and the Actual*. Pantheon Books, New York, 1982, and

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

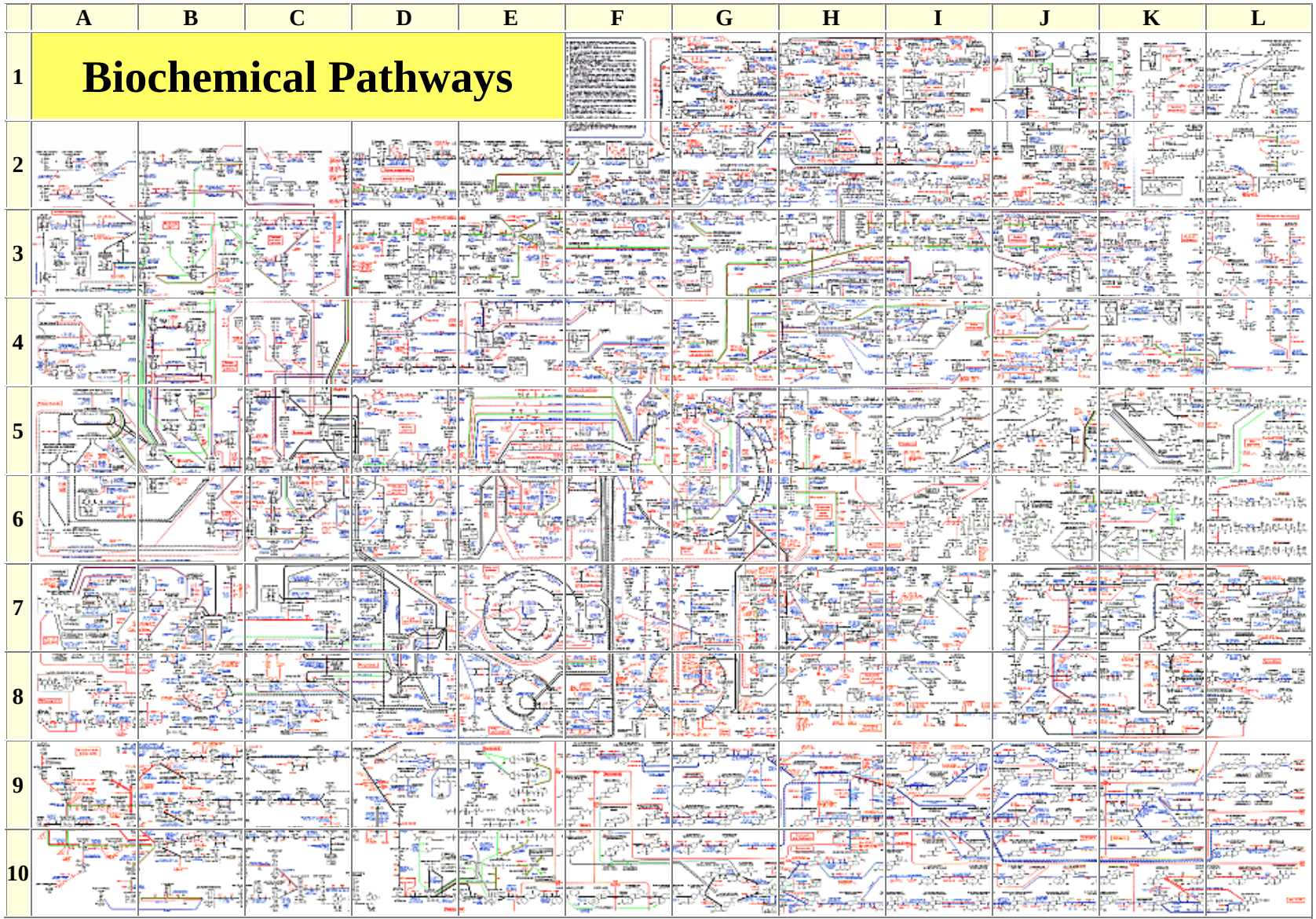

Replication:
DNA ⇒ 2 DNA
+
Transcription:
DNA ⇒ RNA
Nucleotides
Amino Acids
Lipids
Carbohydrates
Small Molecules
Metabolism
Food
Waste
Ribosom
Protein
mRNA
Translation: RNA ⇒ Protein



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

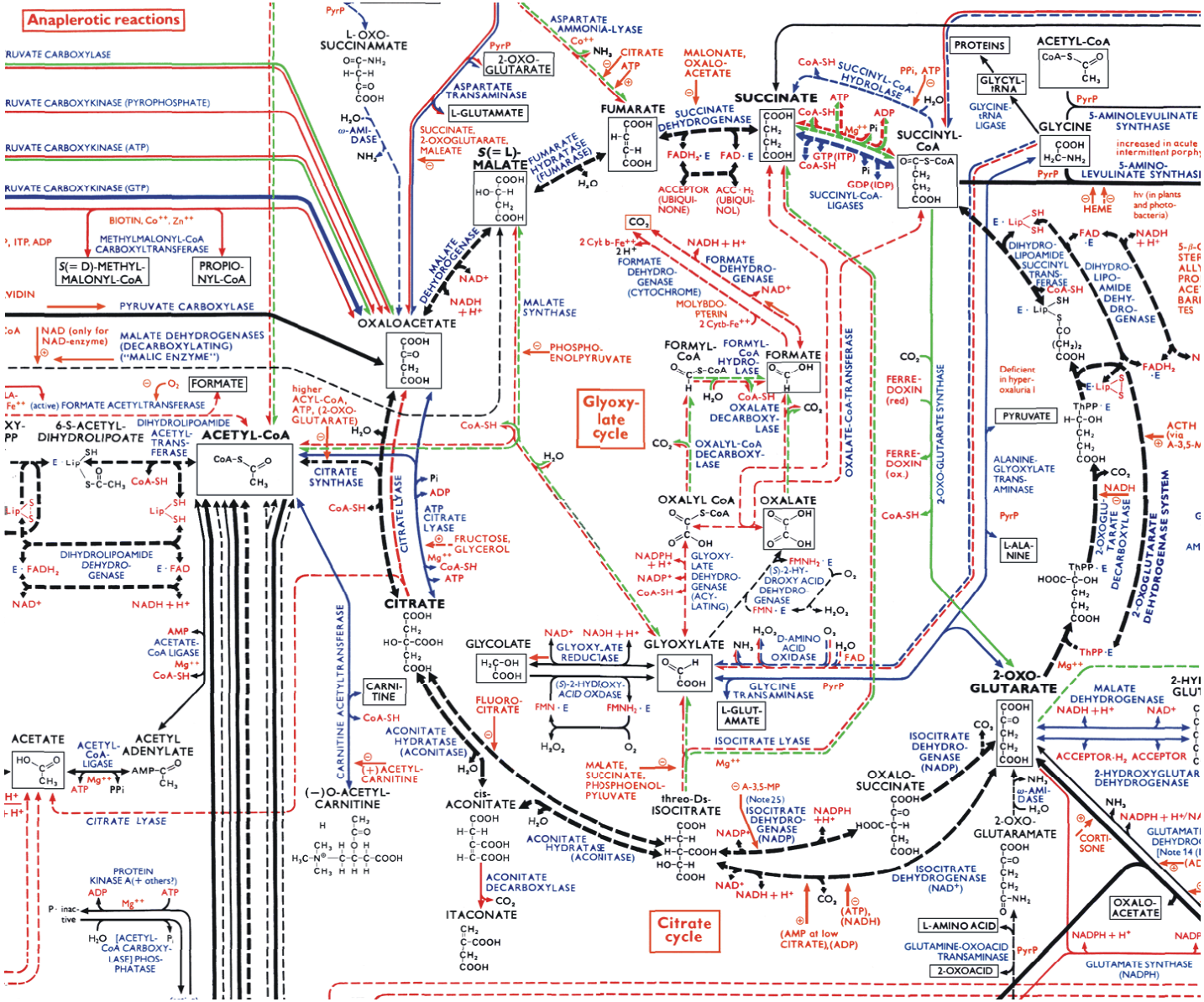


Sketch of a genetic and metabolic network



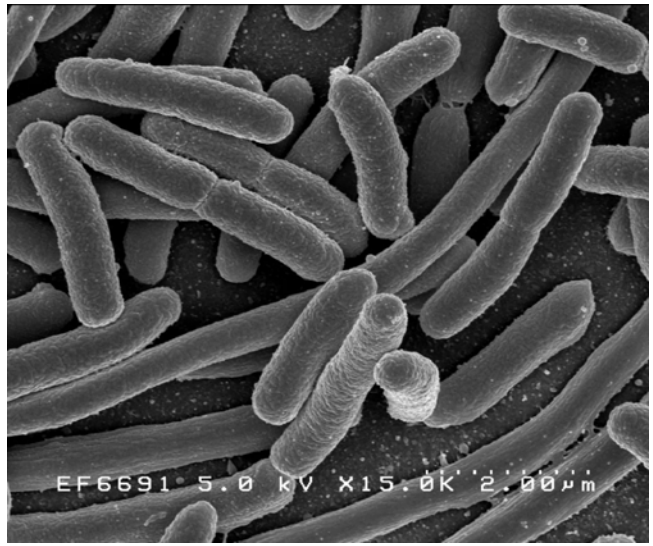
The reaction network of cellular metabolism published by Boehringer-Mannheim.

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



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

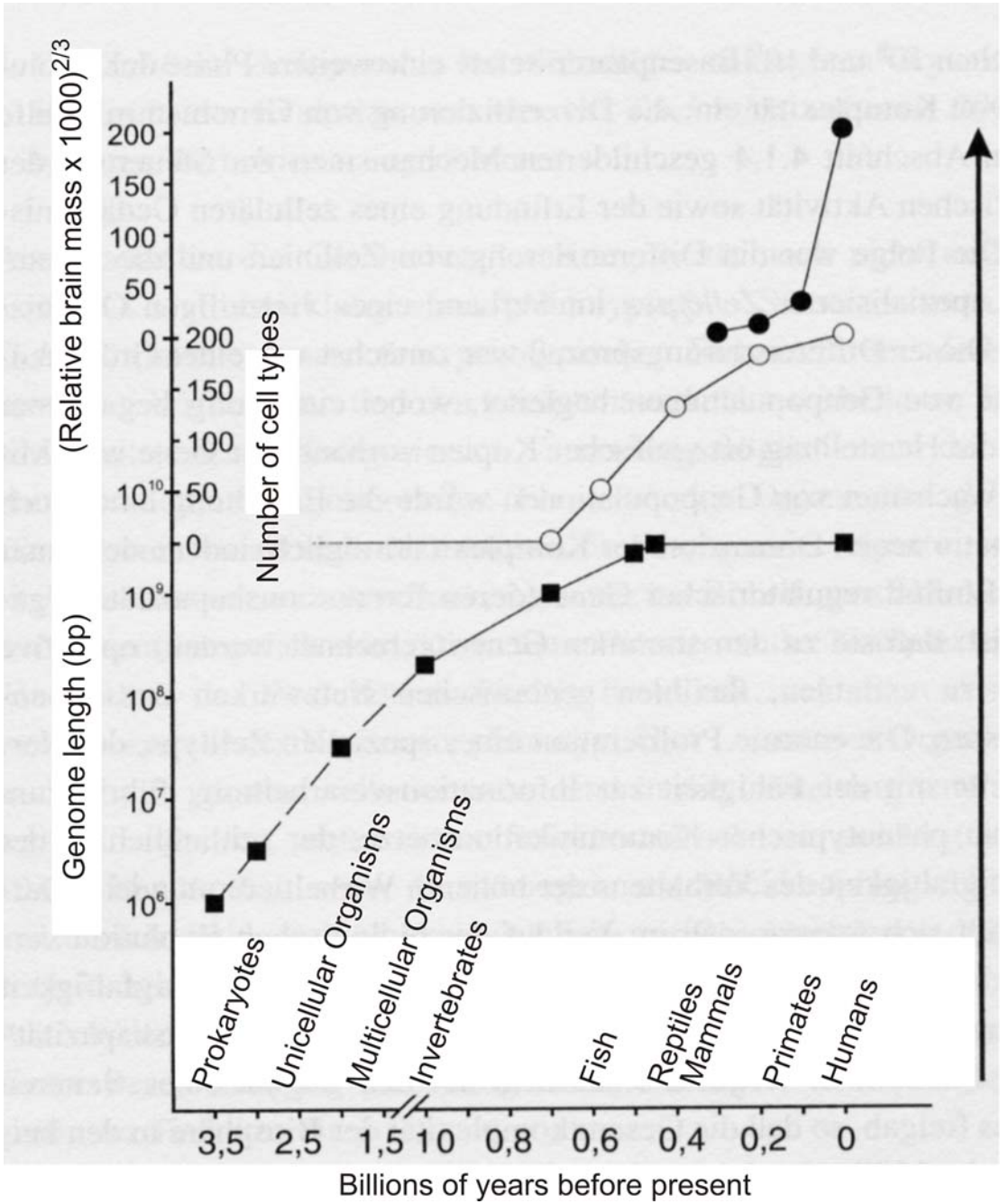
Four books, 300 pages each



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

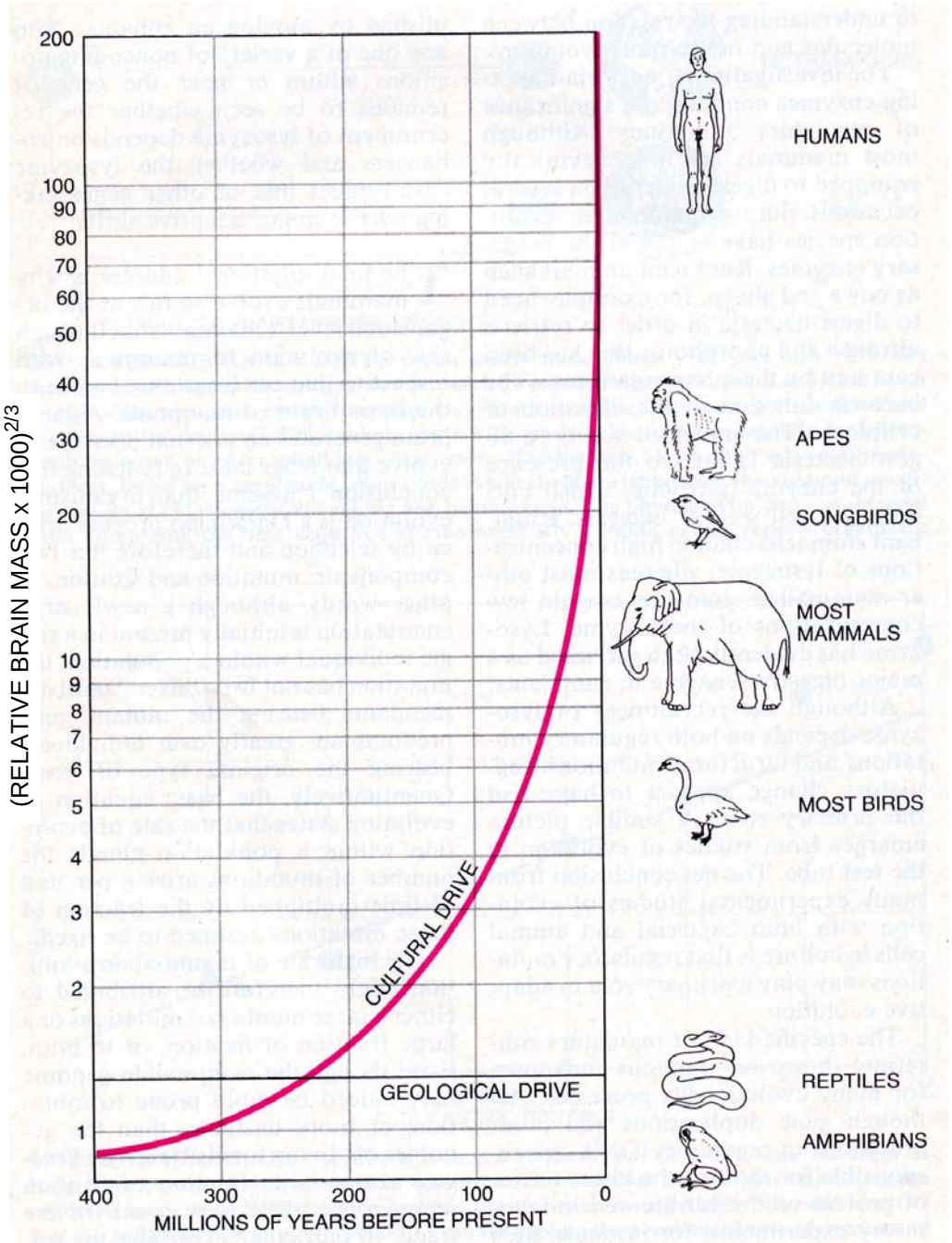
A library of 3000 volumes,
300 pages each

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.

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.

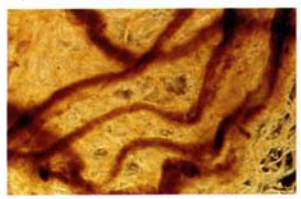
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia 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.

Web-Page for further information:

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