Tracing the Sources of Complexity in Evolution

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Springer Complexity Lecture

ICCS11

Wien, 12.09.2011

Web-Page for further information:

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

1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives

1. **Adaptation in biology**
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Optimization at different levels of complexity

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

Charles Darwin, 1809-1882

Three necessary conditions for Darwinian evolution are:

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

None of the three conditions involves specific properties of the evolving entity except for the capability of reproduction:

Darwinian evolution is universal for reproducing objects no matter whether they are molecules or societies.

Darwinian evolution, however, is not the only mechanism driving biological evolution.

system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

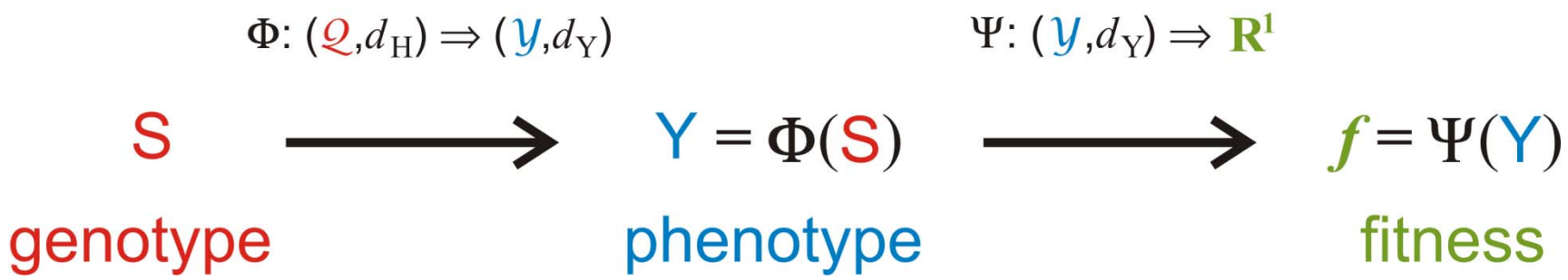
system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

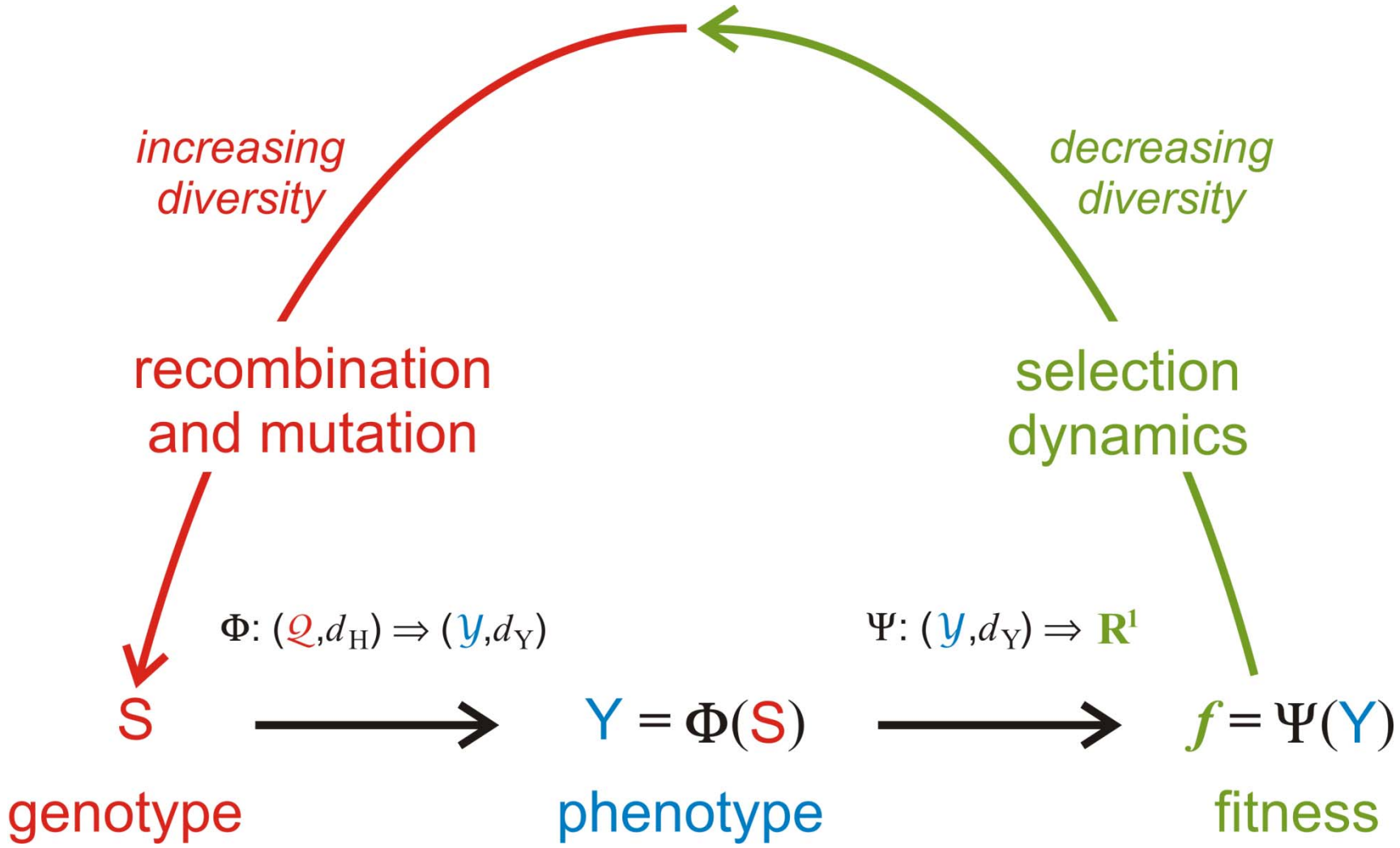
system	mechanism	driving force	external constraint	result
biomolecule	evolutionary design	better performance	thermodynamic and kinetic stability	functional biomolecules
cell	genetics and metabolism	better usage of resources	biochemical kinetics of metabolism	prokaryotic cell
cell / molecule	host/parasite coevolution	mutual „arms races“	availability and survival of host cells	viroids and viruses
organism	cell differentiation and development	improvment by task splitting	regulation and control of cell proliferation	multicellular organism
population	natural selection	maximization of progeny	resources limiting population size	optimized variants
ecosystem	competition and coevolution	adaptation to environments	total carrying capacity	speciation
global	climate changes singular events	survival in new environments	global sustainability	new classes and phyla

Adaptation at different levels of complexity

1. Adaptation in biology
2. **Cycles of evolution**
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives



Genotypes, phenotypes, and fitness



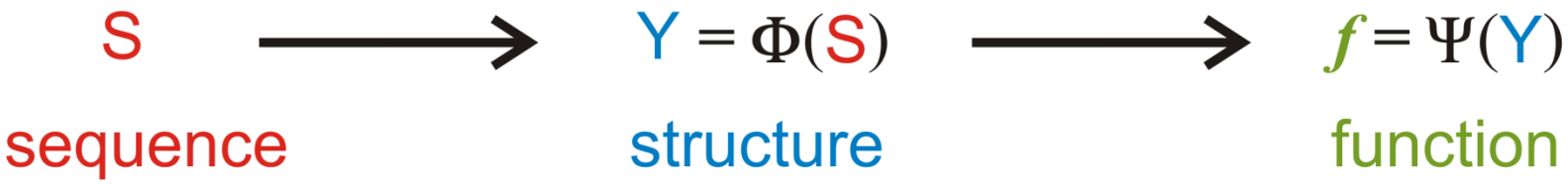
Genotypes, phenotypes, and fitness

1. Adaptation in biology
2. Cycles of evolution
3. **Molecules – from sequence to function**
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives

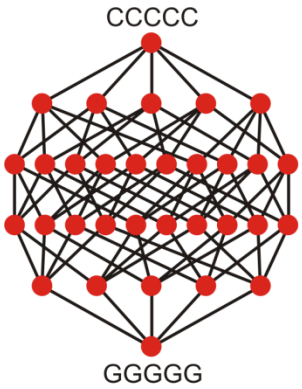
Make things as simple as possible,
but not simpler !

Albert Einstein

Albert Einstein‘s razor, precise refence is unknown.



The paradigm of structural biology

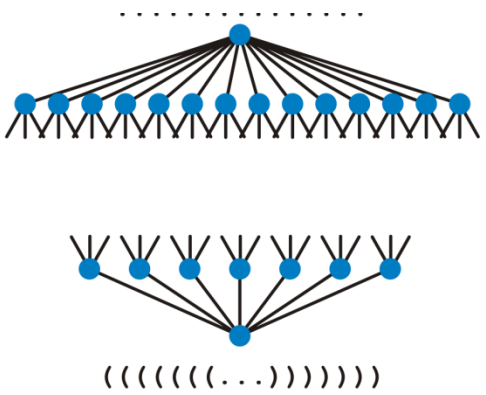


sequence space

S

sequence

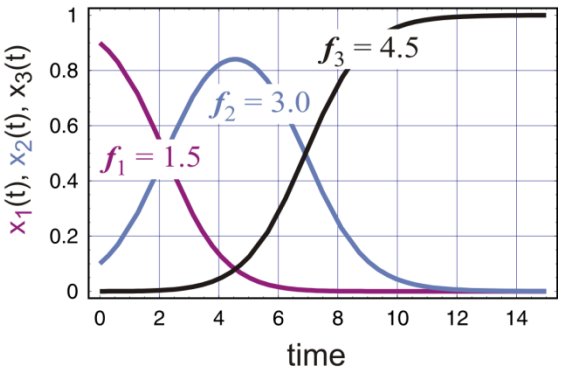
The paradigm of structural biology



Y

structure

The paradigm of structural biology

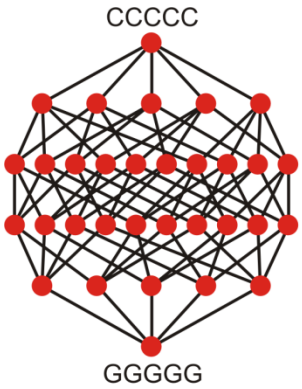


parameter space

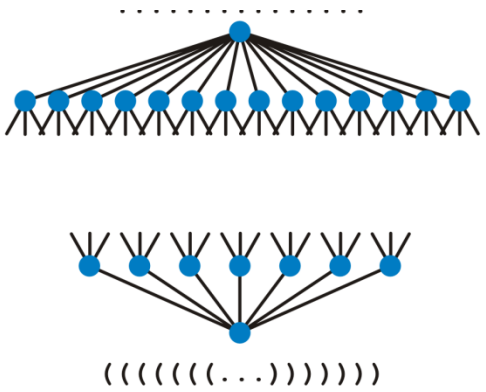
$$f = \Psi(Y)$$

function

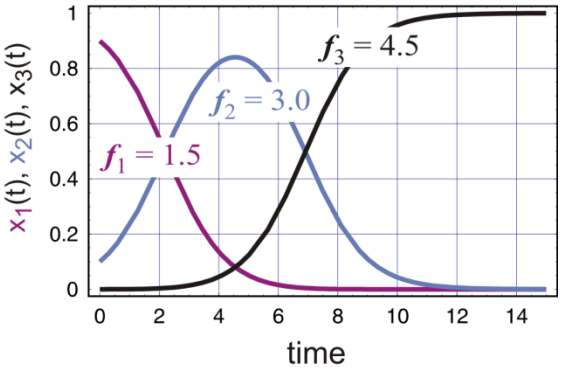
The paradigm of structural biology



sequence space



shape space



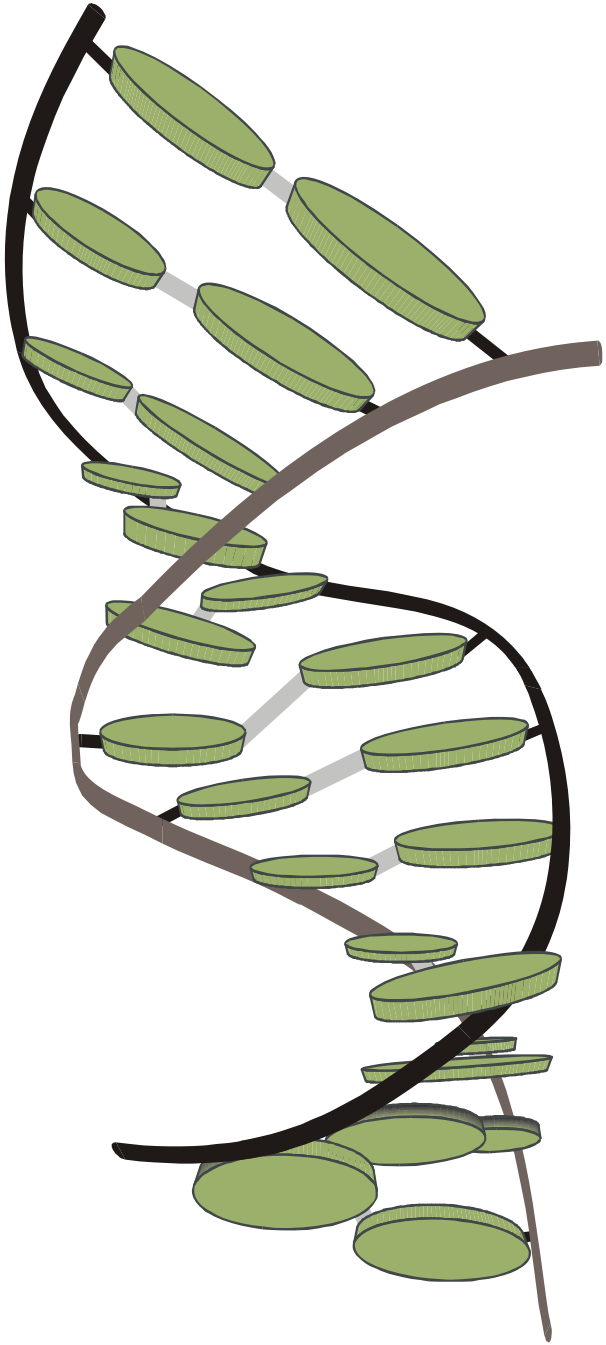
parameter space

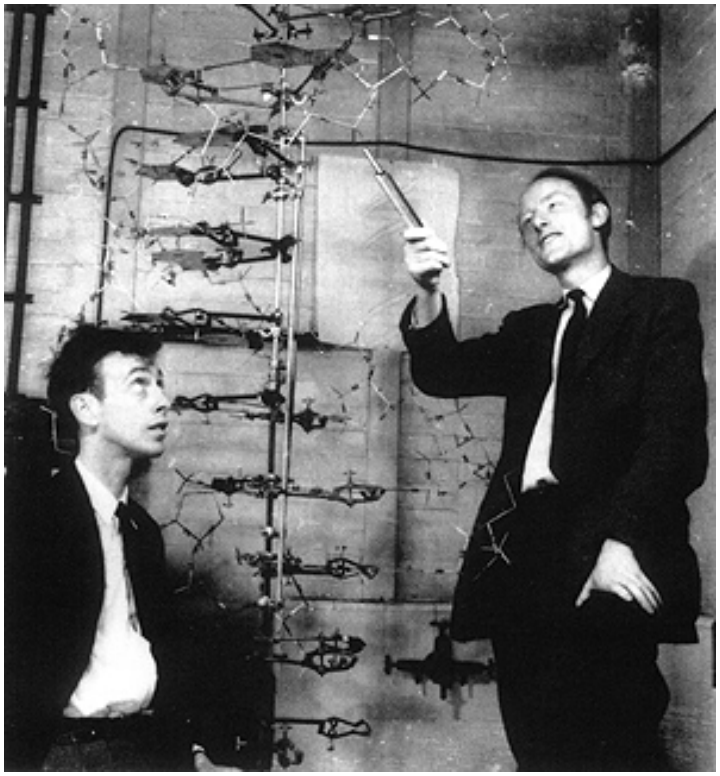
$\Phi: (\mathcal{Q}, d_H) \Rightarrow (\mathcal{Y}, d_Y)$ $\qquad$ $\Psi: (\mathcal{Y}, d_Y) \Rightarrow \mathbf{R}^1$

S $\longrightarrow$ Y = Φ(S) $\longrightarrow$ f = Ψ(Y)

sequence $\qquad$ structure $\qquad$ function

The paradigm of structural biology



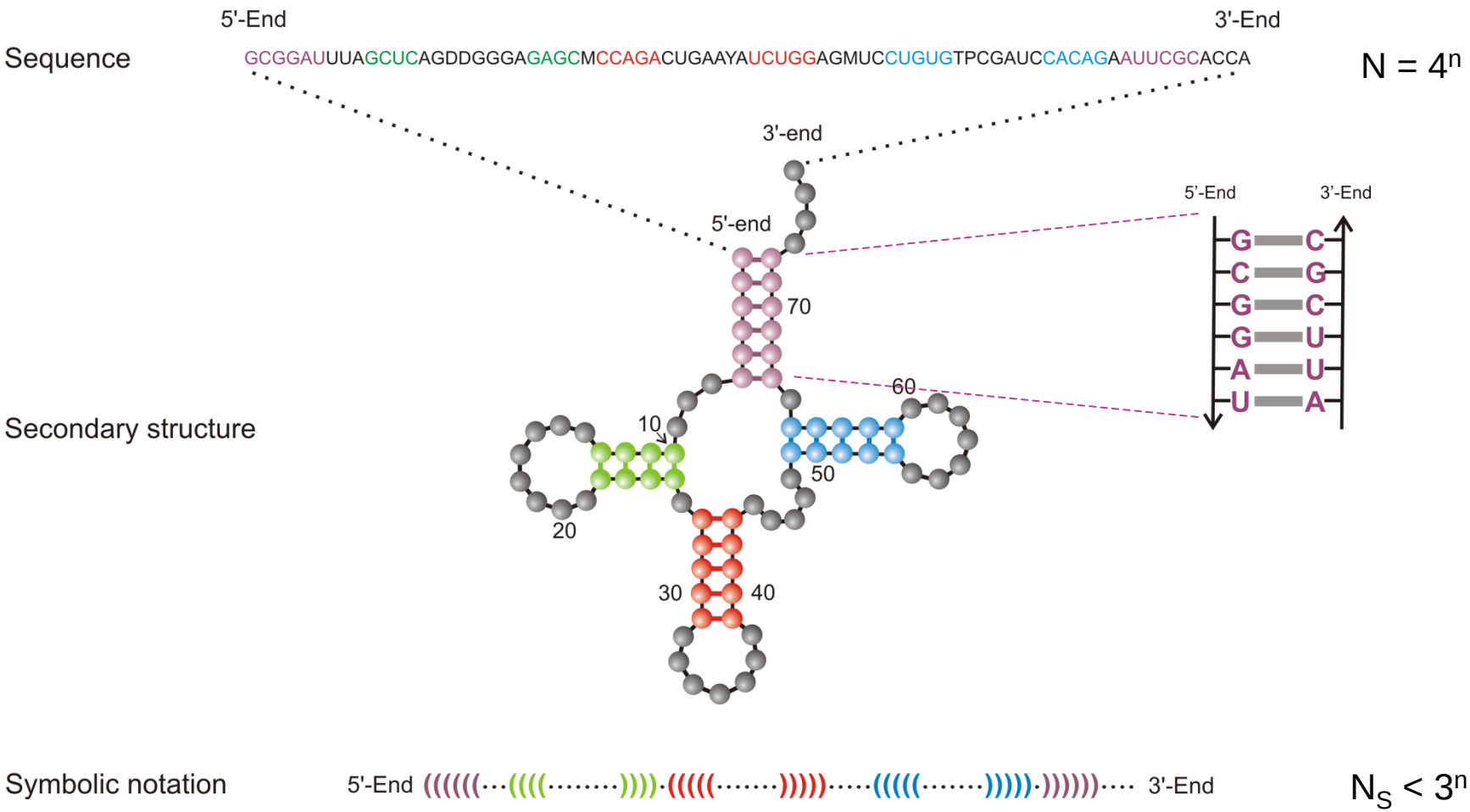


James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004

Nobel prize 1962

1953 – 2003 fifty years double helix

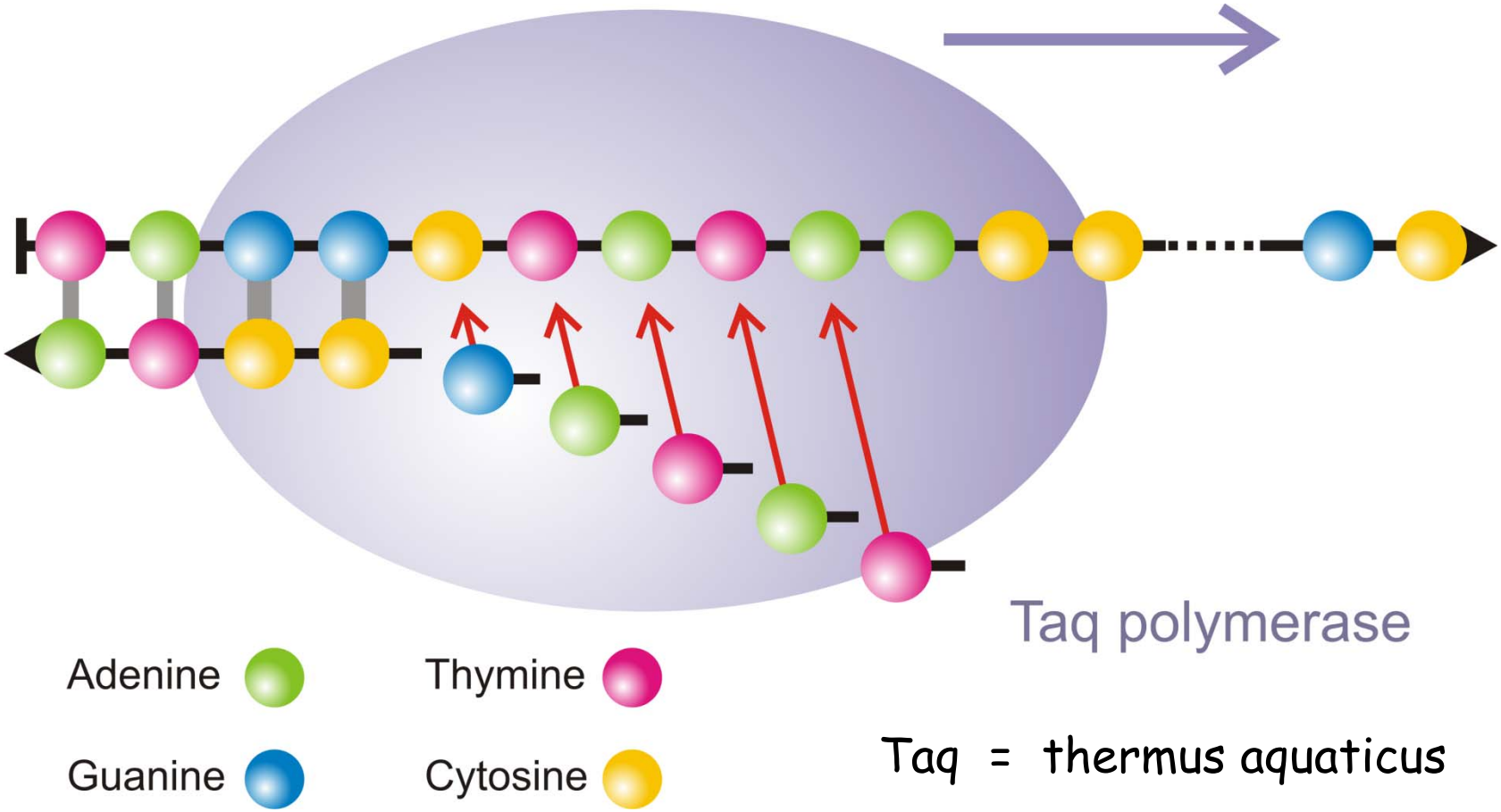
The three-dimensional structure of a short double helical stack of B-DNA



Criterion: Minimum free energy (mfe)

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

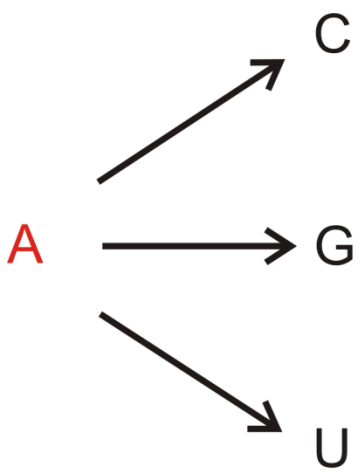
A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs



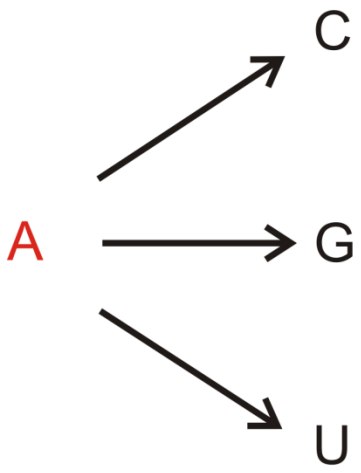
Taq = thermus aquaticus

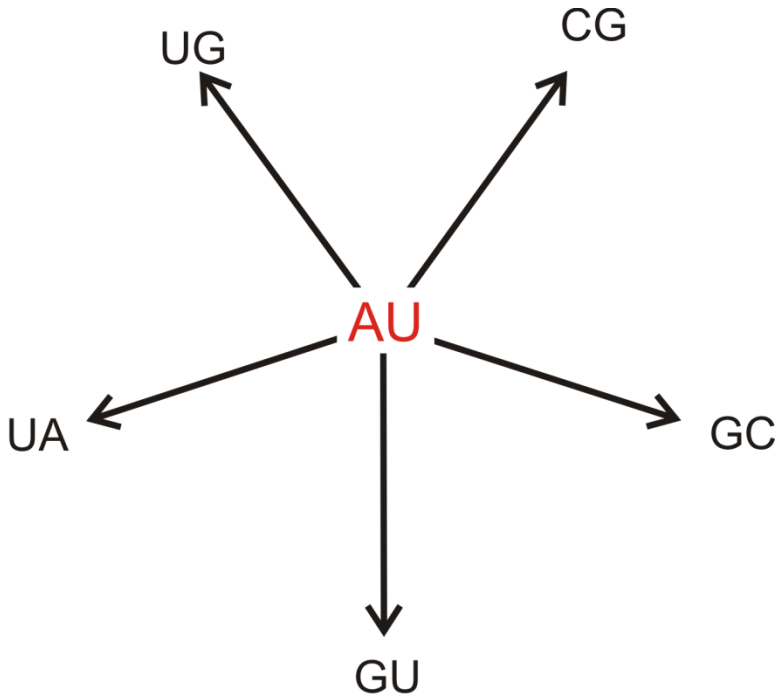
The logics of DNA replication

1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. **Mutation and structure**
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives

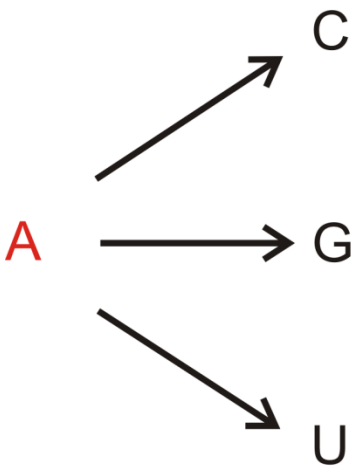


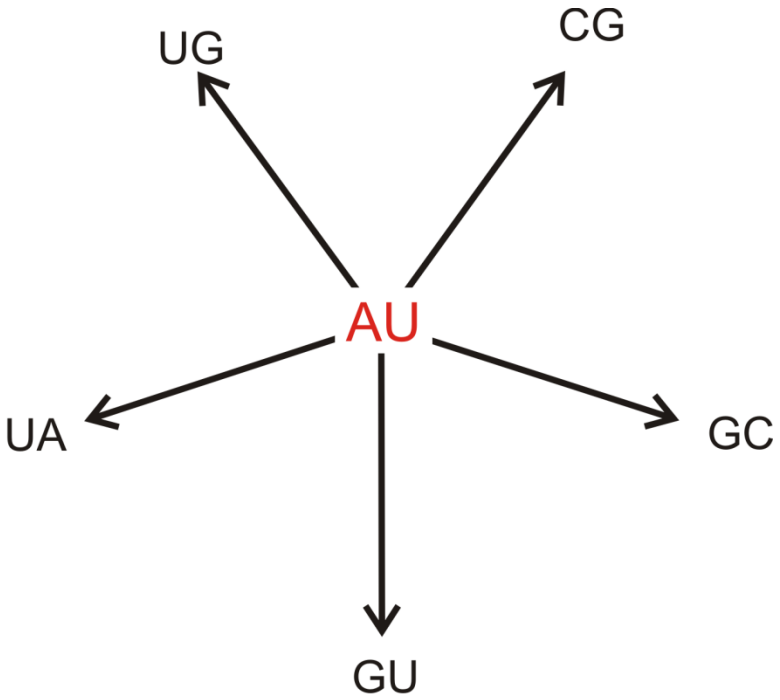
Nucleobase and base pair mutations

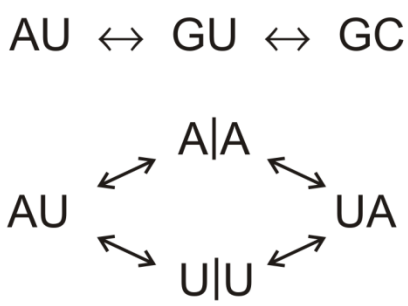




Nucleobase and base pair mutations



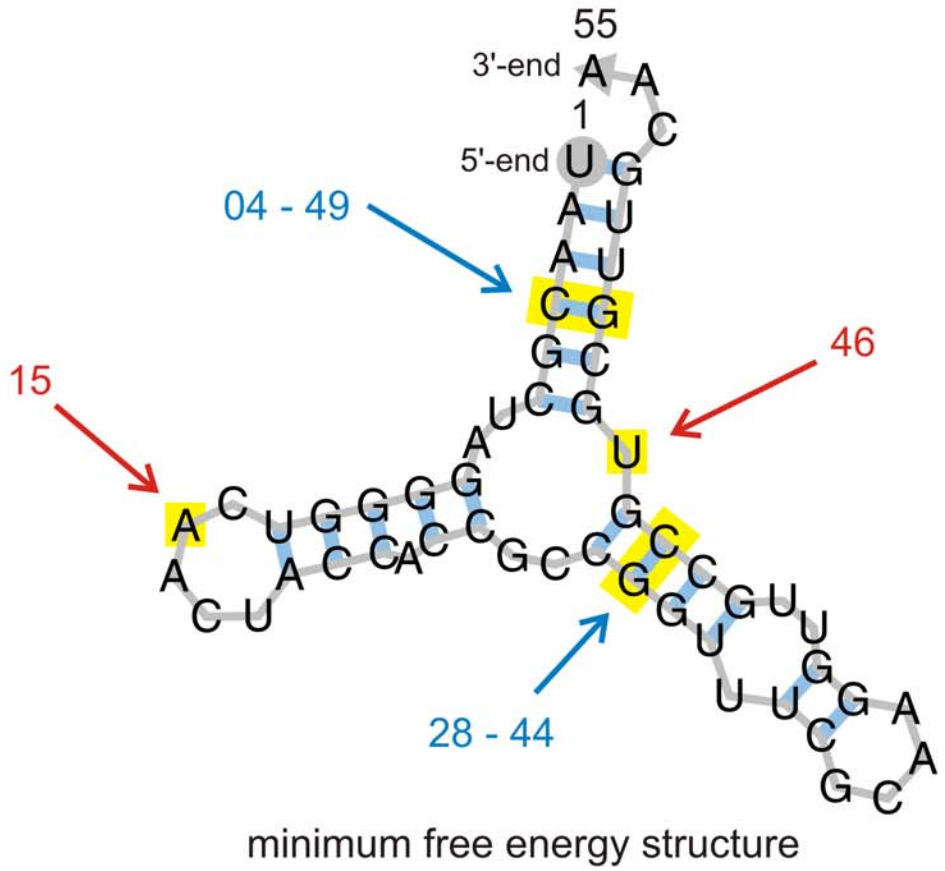




Nucleobase and base pair mutations

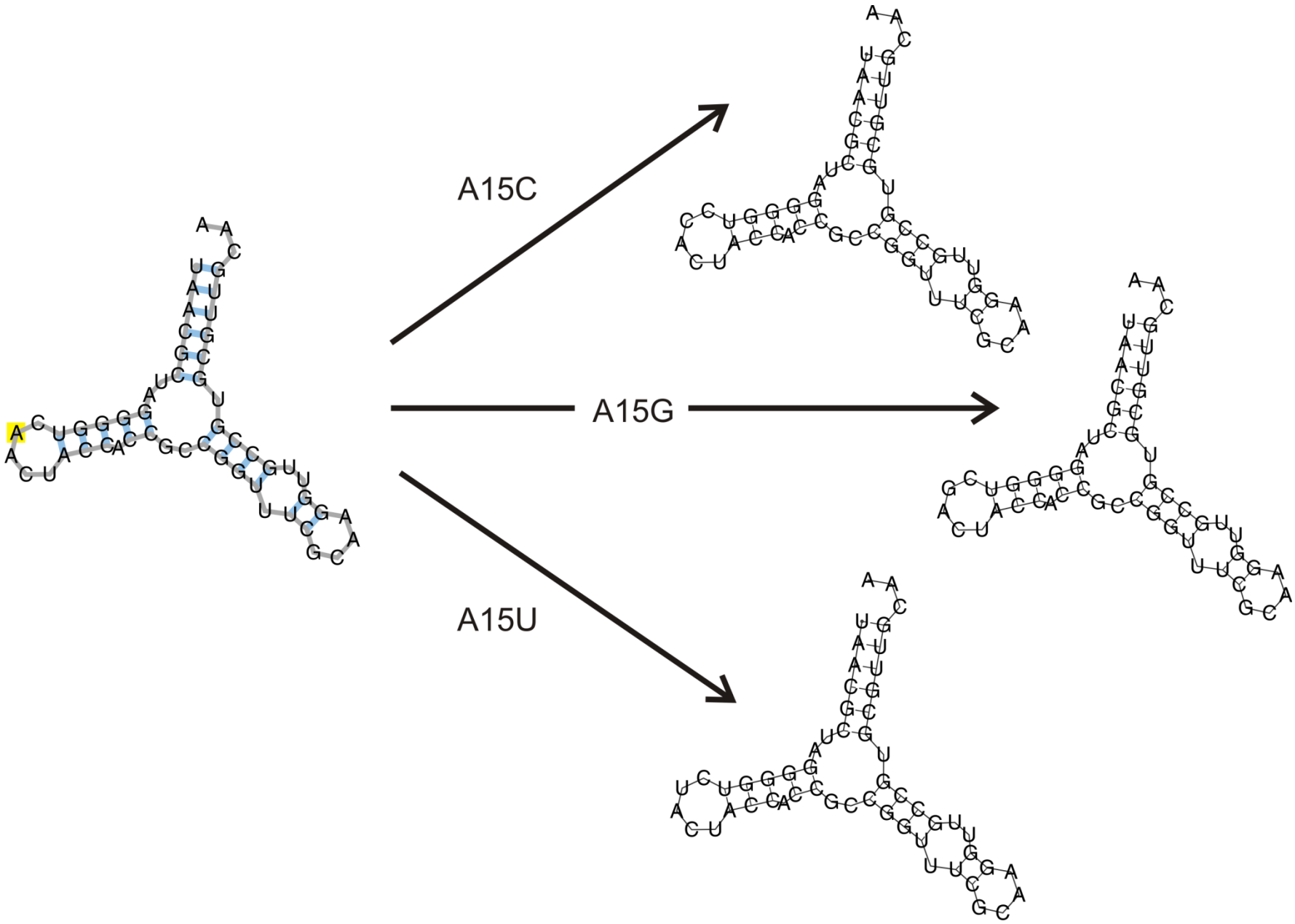
5'-end UAACGCUAGGGGUCAACUACCACCGCCGGUUUCGCAAGGUUGCCGUGCGUUGCAA 3'-end

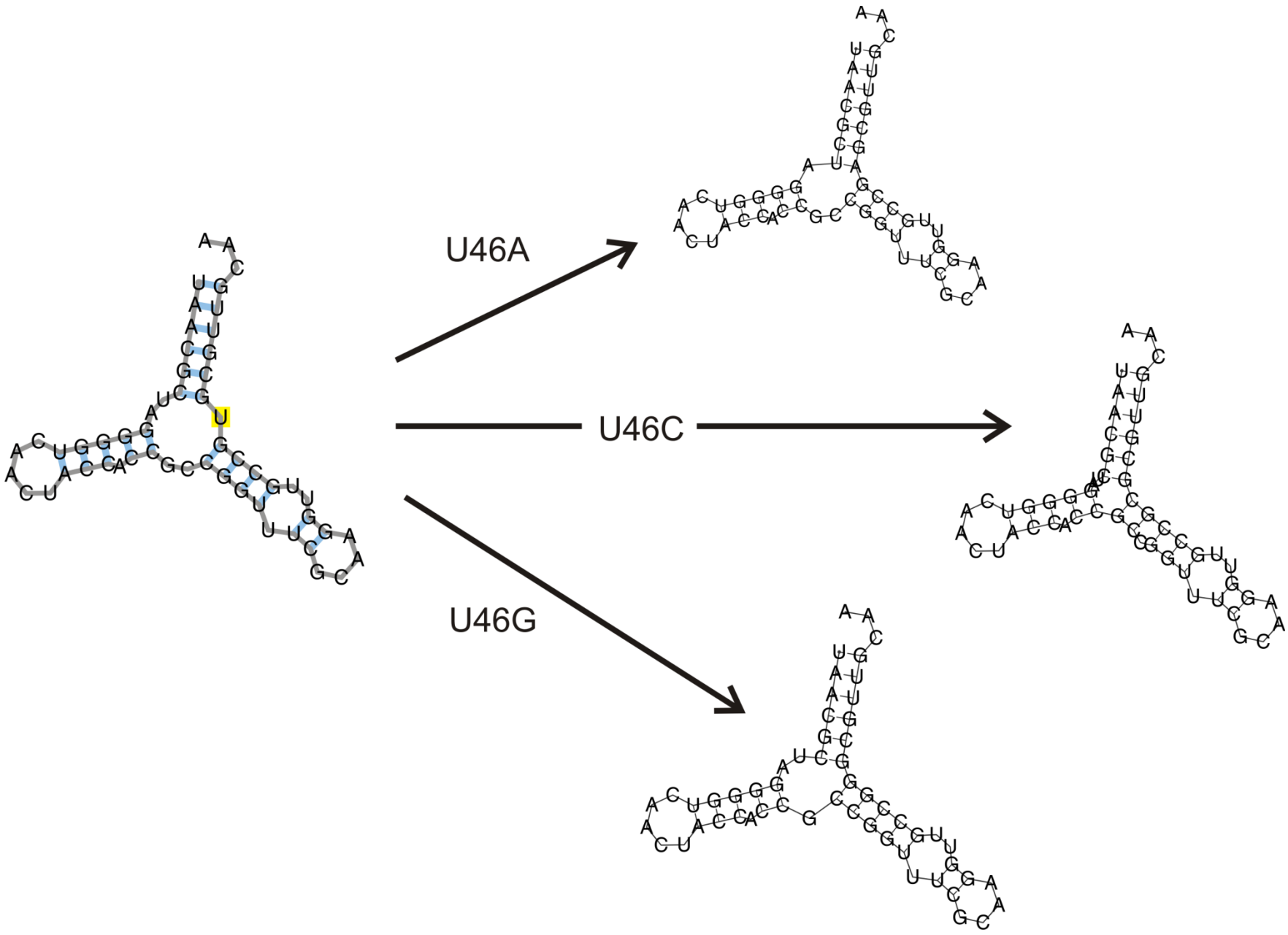
sequence

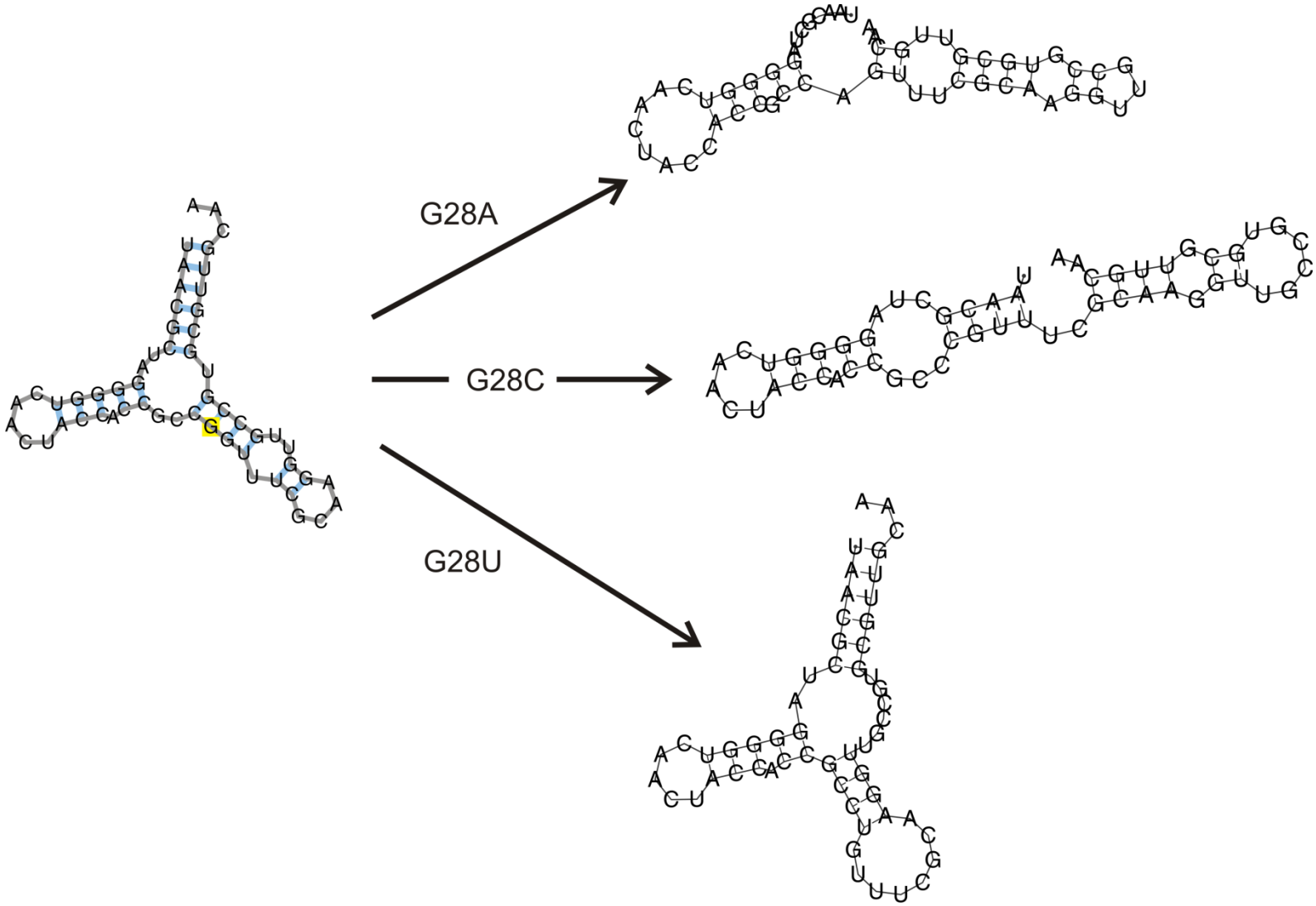


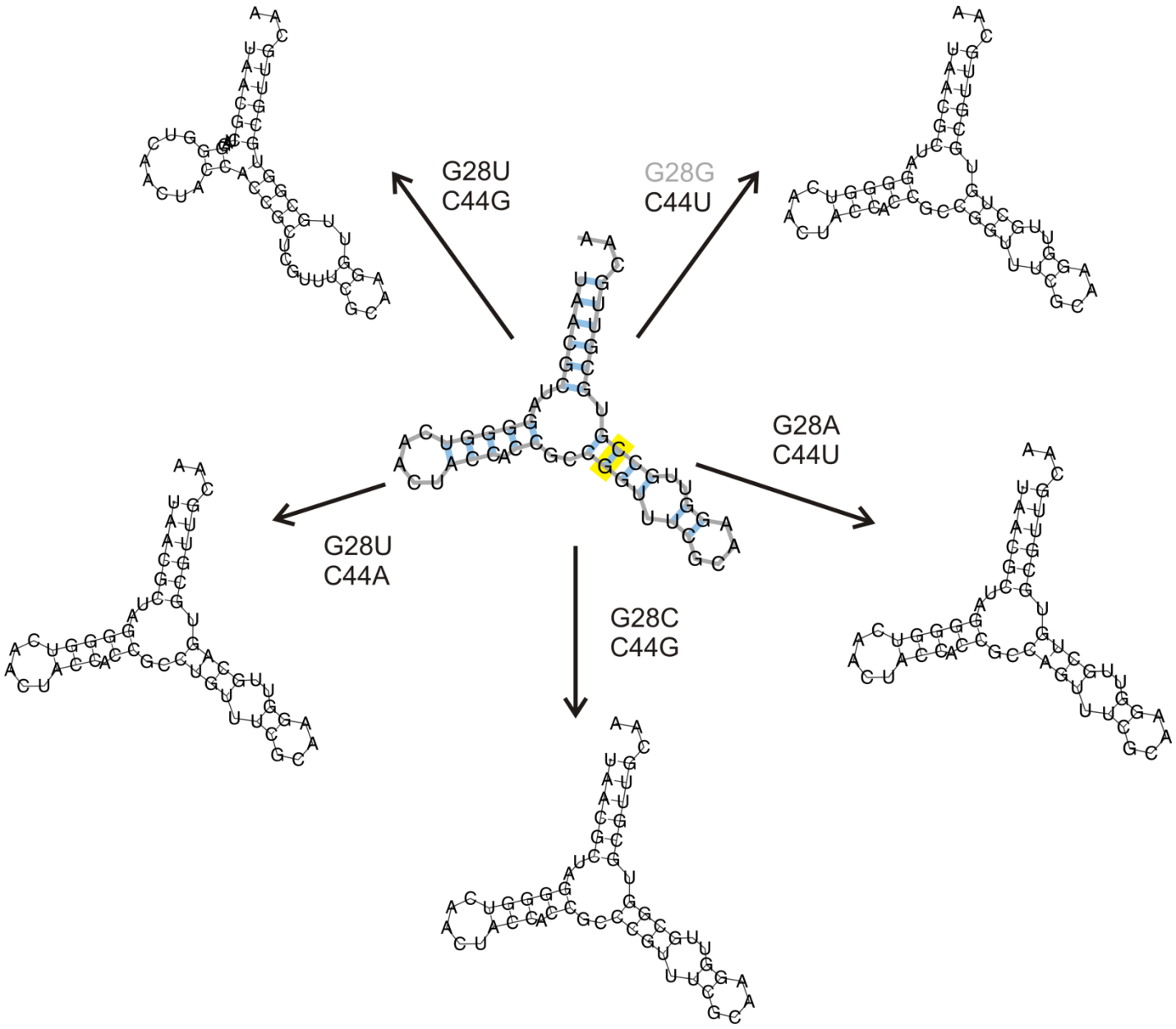
minimum free energy structure

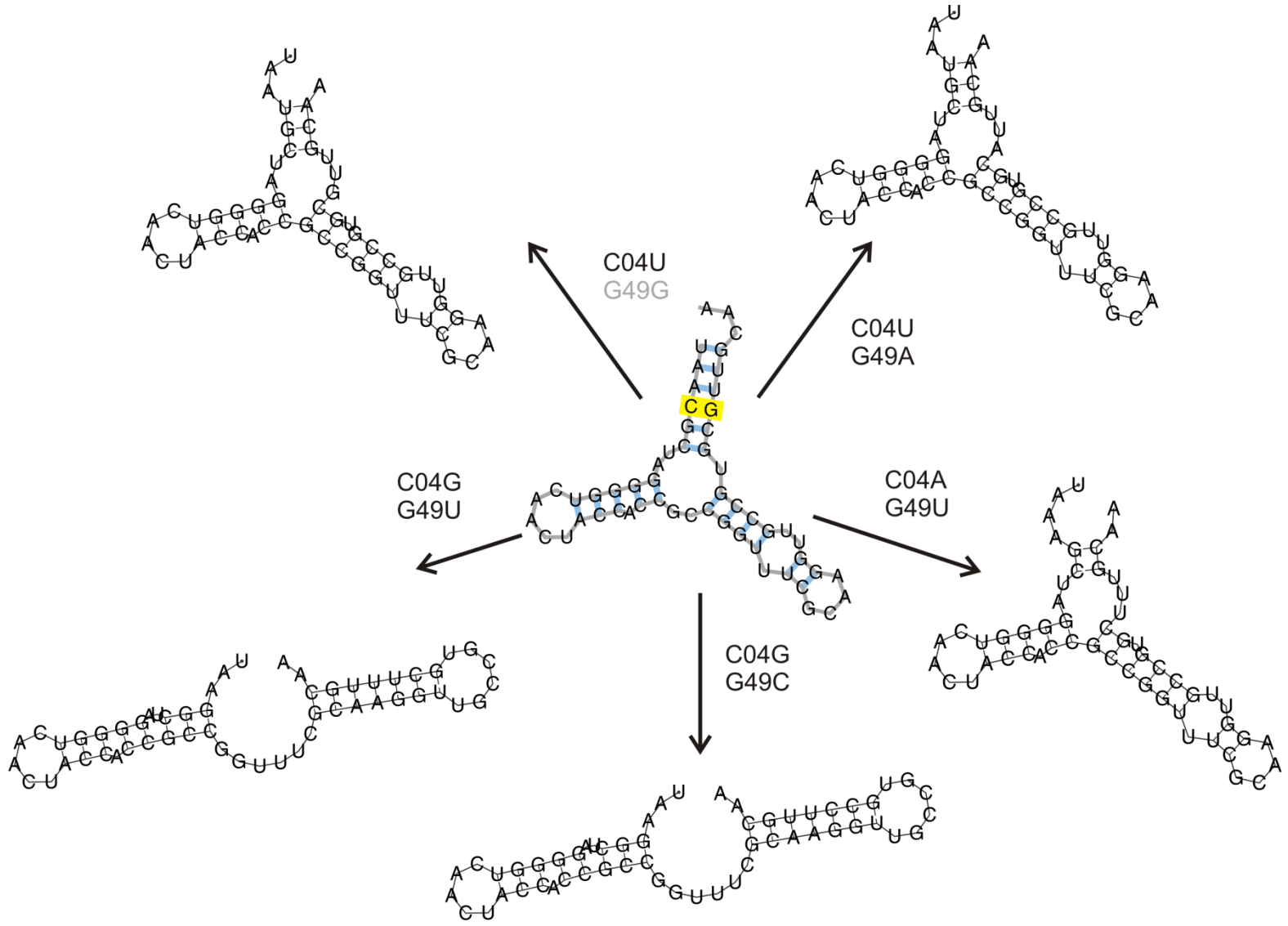
A case study: A simple RNA molecule

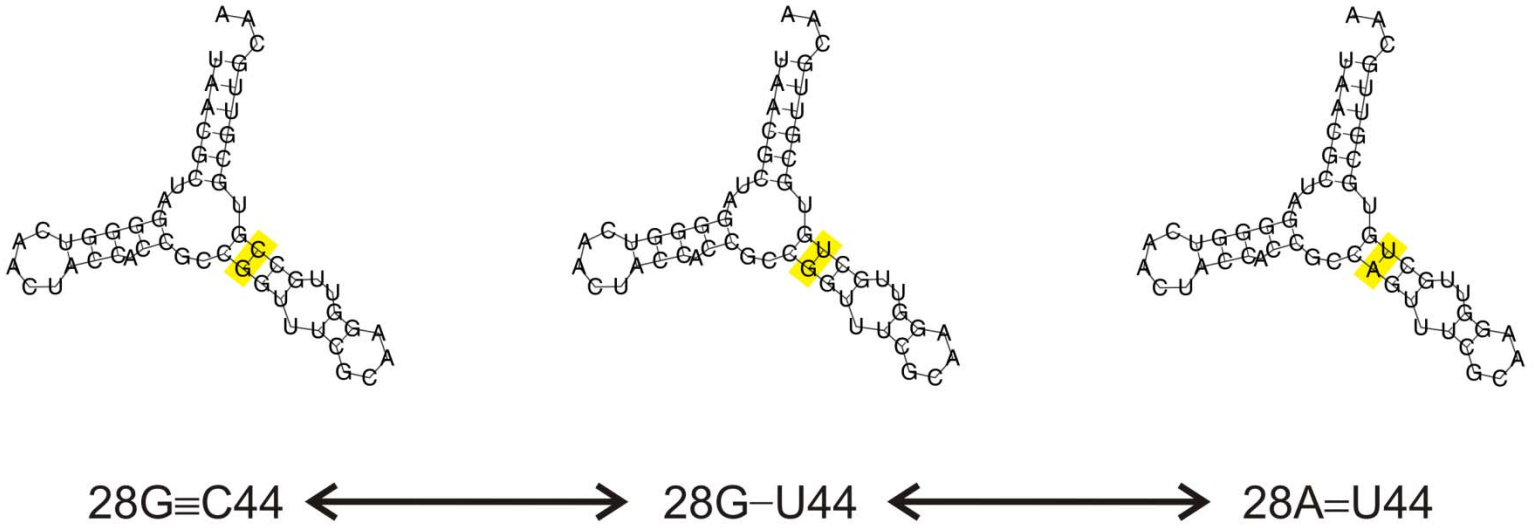

A15C
A15G
A15U

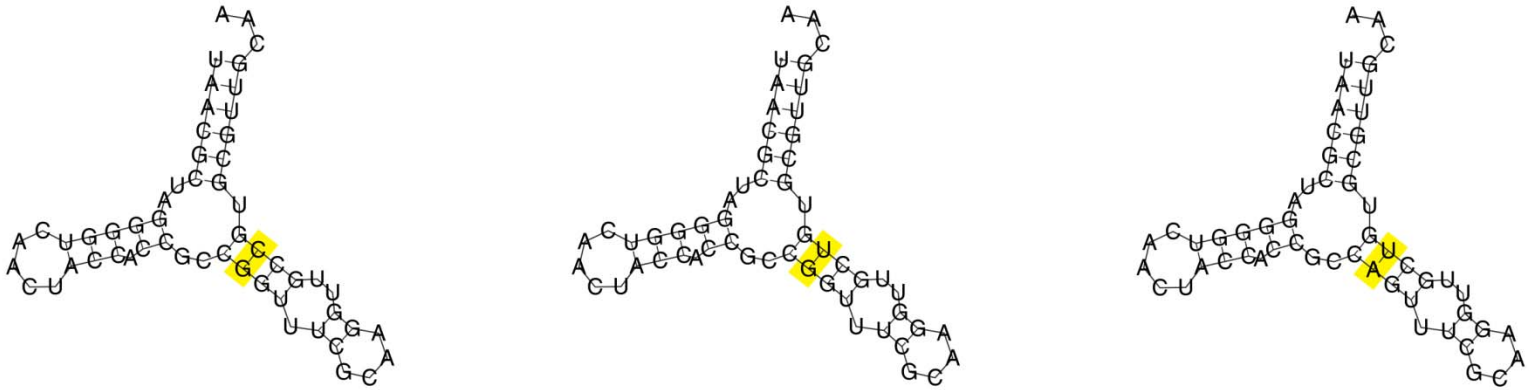

U46A
U46C
U46G


G28A
G28C
G28U


G28U
C44G
G28G
C44U
G28A
C44U
G28U
C44A
G28C
C44G

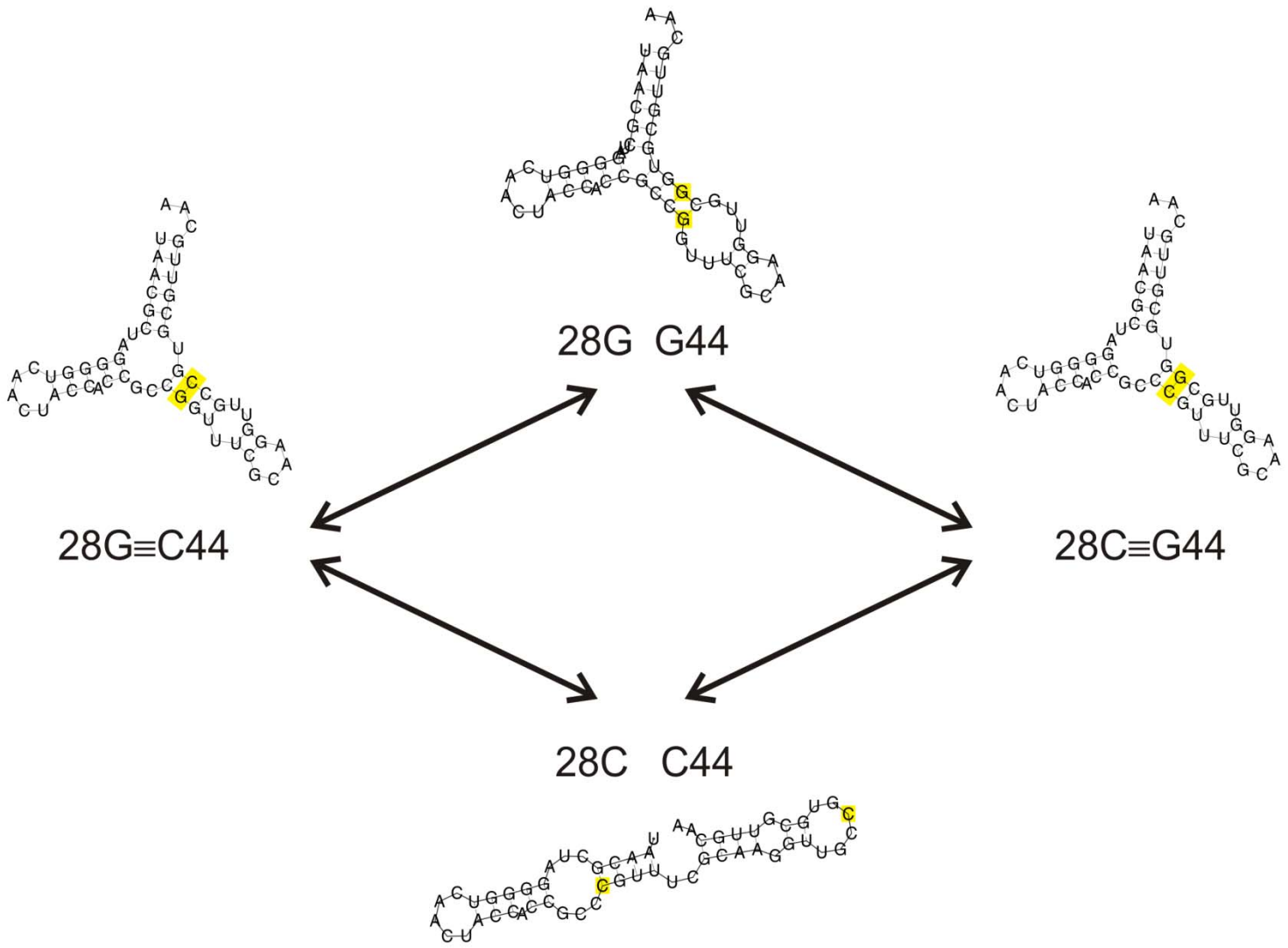

C04U
G49G
C04U
G49A
C04G
G49U
C04A
G49U
C04G
G49C

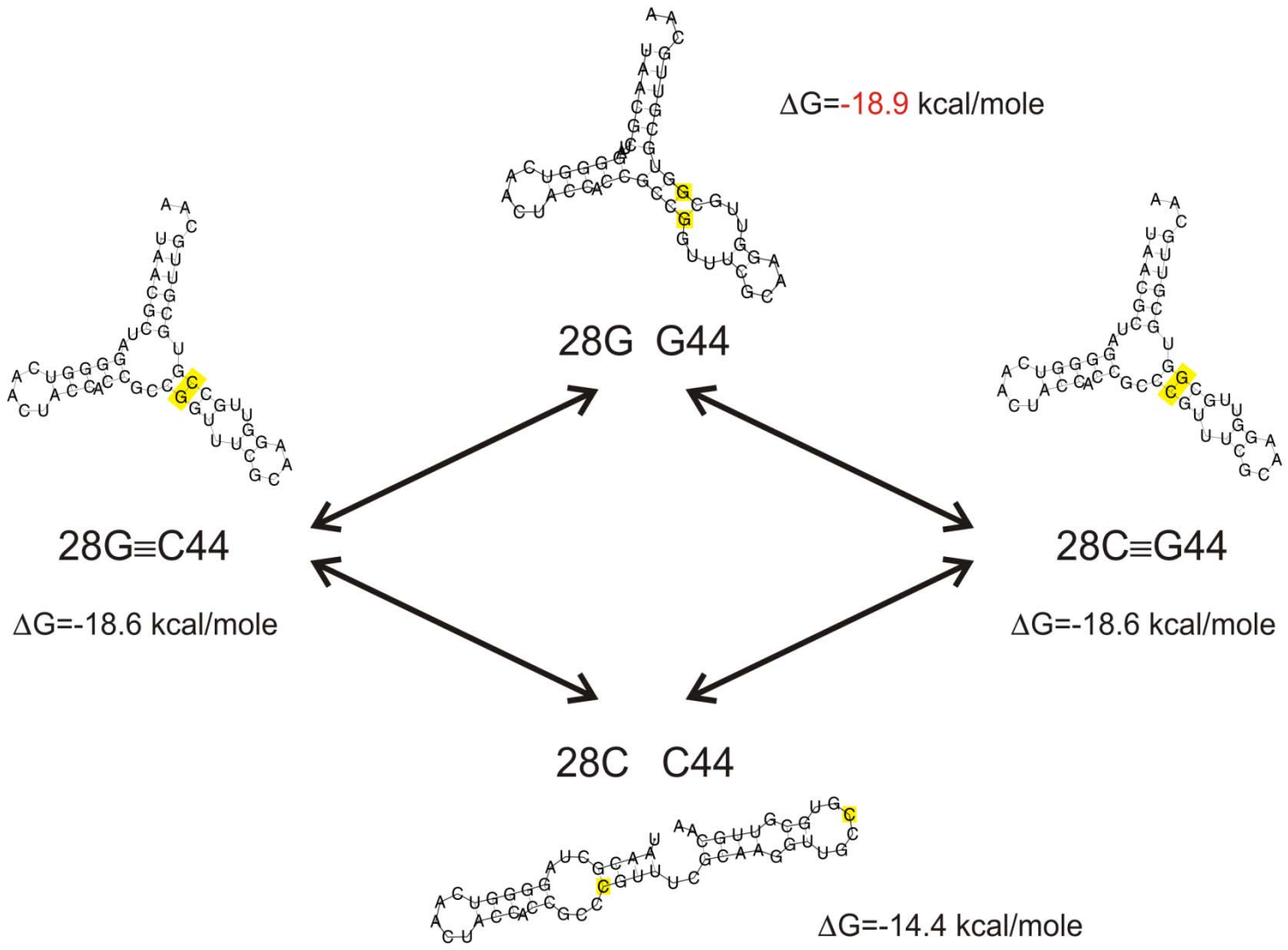

28G≡C44
28G–U44
28A=U44

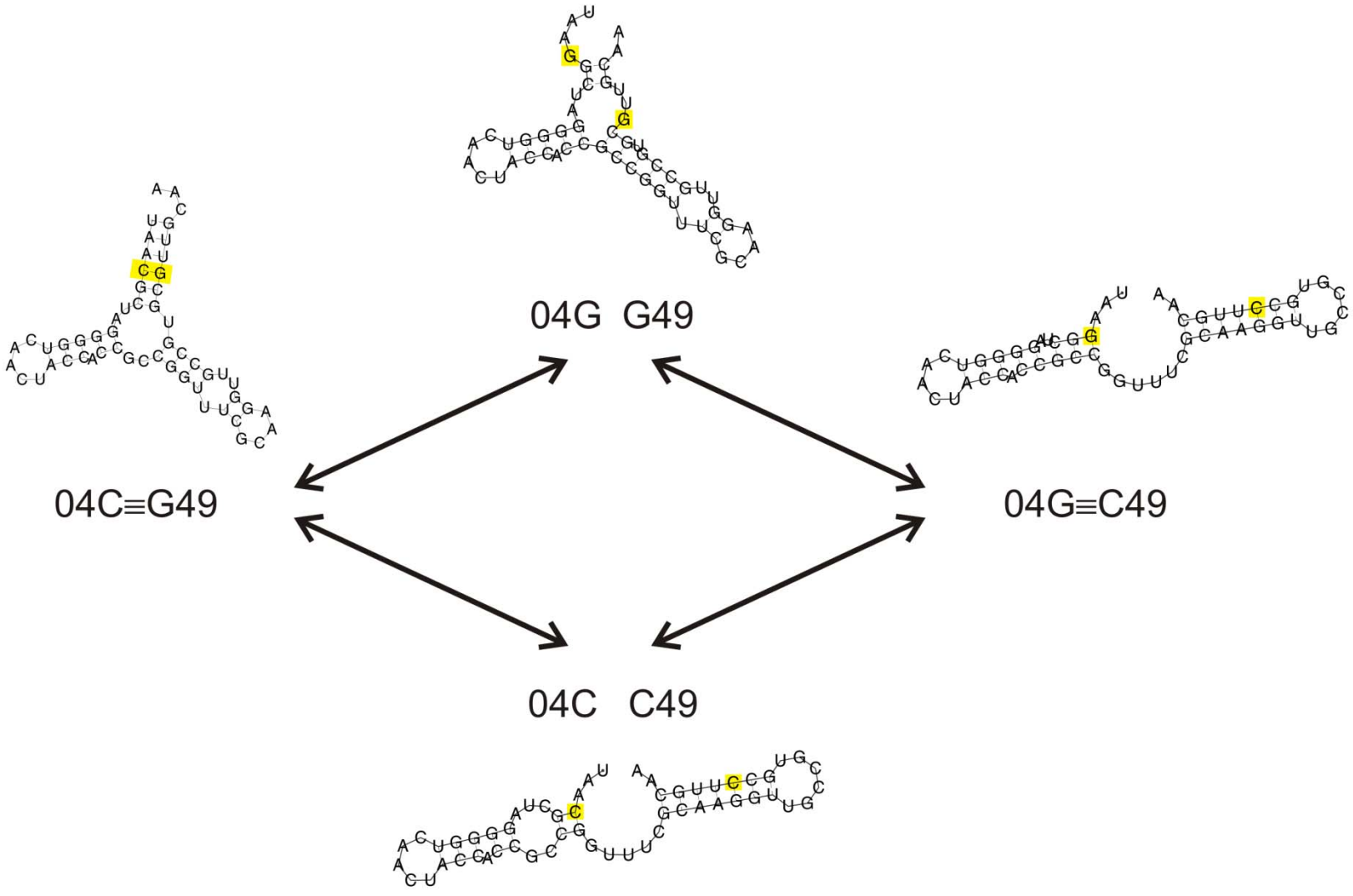


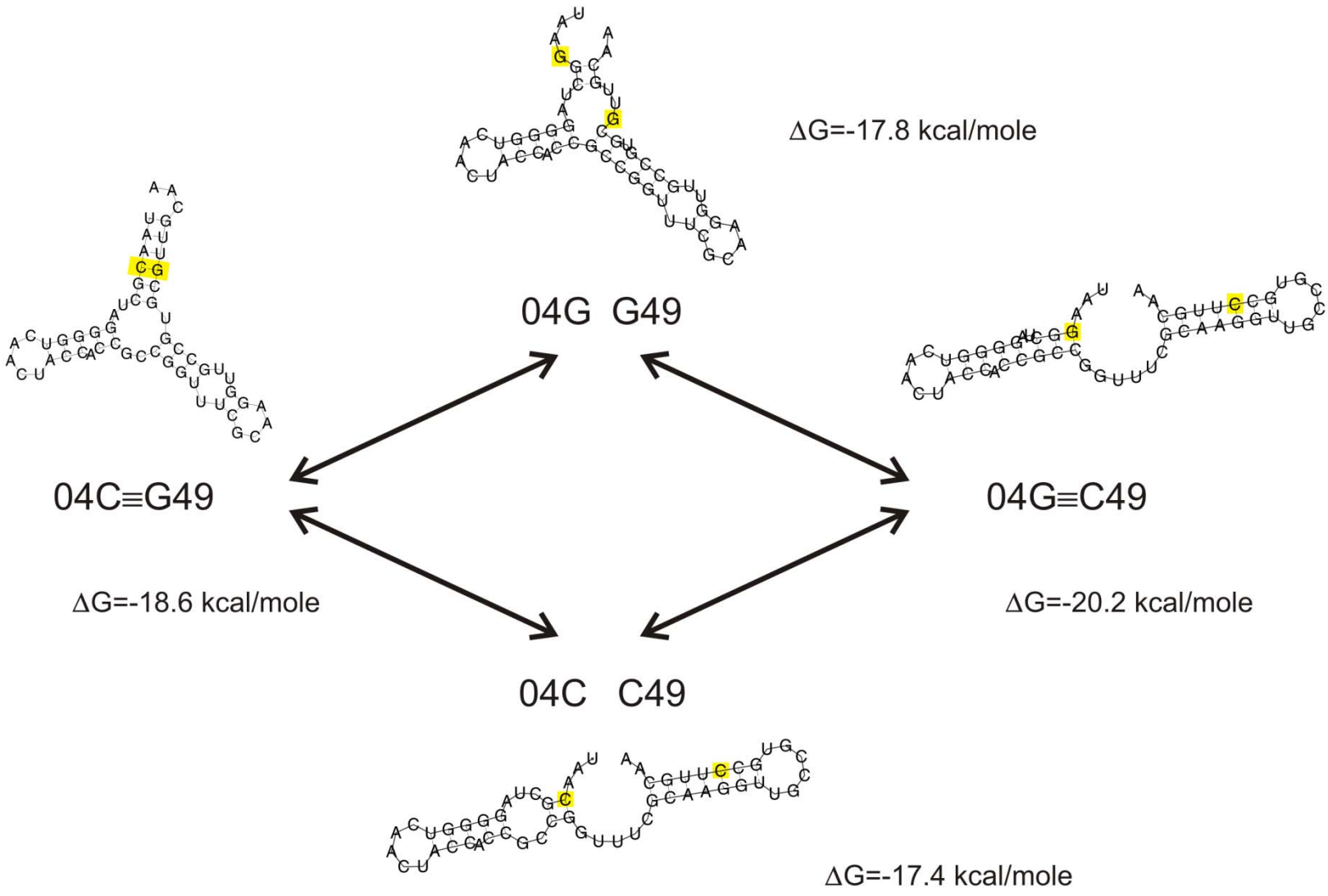
28G≡C44 ⟷ 28G–U44 ⟷ 28A=U44

ΔG=-18.6 kcal/mole ΔG=-14.4 kcal/mole ΔG=-17.1 kcal/mole

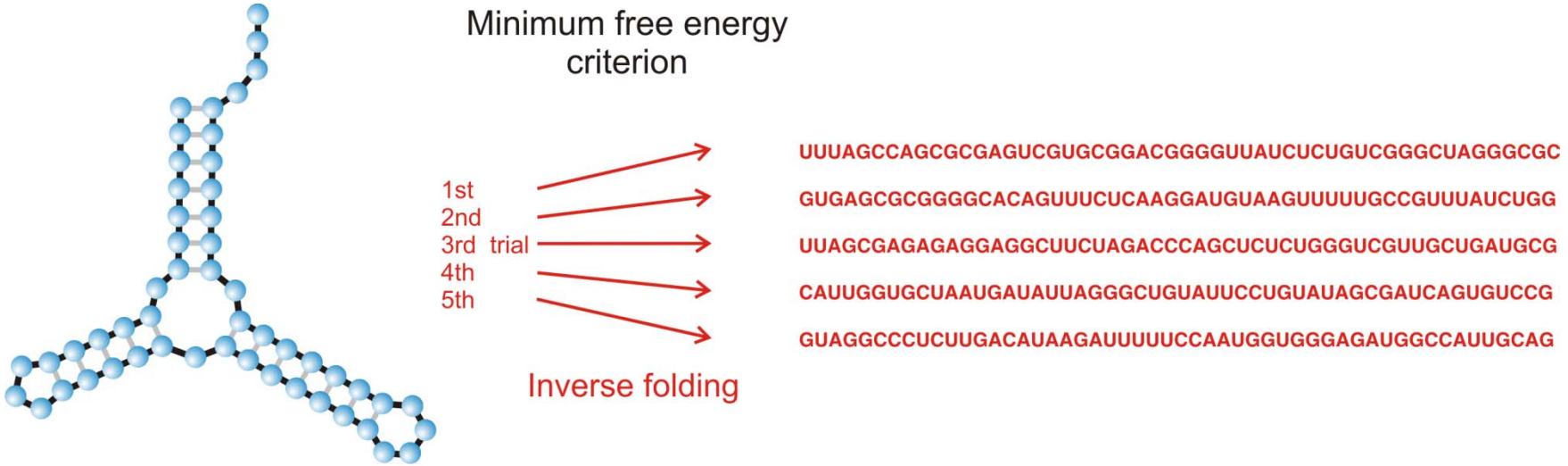

28G G44
28G≡C44
28C≡G44
28C C44


ΔG=-18.9 kcal/mole
28G G44
28G≡C44
ΔG=-18.6 kcal/mole
28C≡G44
ΔG=-18.6 kcal/mole
28C C44
ΔG=-14.4 kcal/mole


04G G49
04C≡G49
04G≡C49
04C C49


ΔG=-17.8 kcal/mole
04G G49
04C≡G49
04G≡C49
ΔG=-18.6 kcal/mole
ΔG=-20.2 kcal/mole
04C C49
ΔG=-17.4 kcal/mole

1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. **Spaces and mappings**
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives



The **inverse folding algorithm** searches for sequences that form a given RNA secondary structure under the minimum free energy criterion.

Monatshefte für Chemie 125, 167–188 (1994)

Monatshefte für Chemie
Chemical Monthly

Printed in Austria

Fast Folding and Comparison of RNA Secondary Structures

I. L. Hofacker[1,*], W. Fontana[3], P. F. Stadler[1,3], L. S. Bonhoeffer[4], M. Tacker[1] and P. Schuster[1,2,3]

[1] Institut für Theoretische Chemie, Universität Wien, A-1090 Wien, Austria
[2] Institut für Molekulare Biotechnologie, D-07745 Jena, Federal Republic of Germany
[3] Santa Fe Institute, Santa Fe, NM 87501, U.S.A.
[4] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

Summary. Computer codes for computation and comparison of RNA secondary structures, the Vienna RNA package, are presented, that are based on dynamic programming algorithms and aim at predictions of structures with minimum free energies as well as at computations of the equilibrium partition functions and base pairing probabilities.

An efficient heuristic for the inverse folding problem of RNA is introduced. In addition we present compact and efficient programs for the comparison of RNA secondary structures based on tree editing and alignment.

All computer codes are written in ANSI C. They include implementations of modified algorithms on parallel computers with distributed memory. Performance analysis carried out on an Intel Hypercube shows that parallel computing becomes gradually more and more efficient the longer the sequences are.

Keywords. Inverse folding; parallel computing; public domain software; RNA folding; RNA secondary structures; tree editing.

space of genotypes: $\mathcal{Q} = \{S_1, S_2, S_3, \ldots, S_N\}$

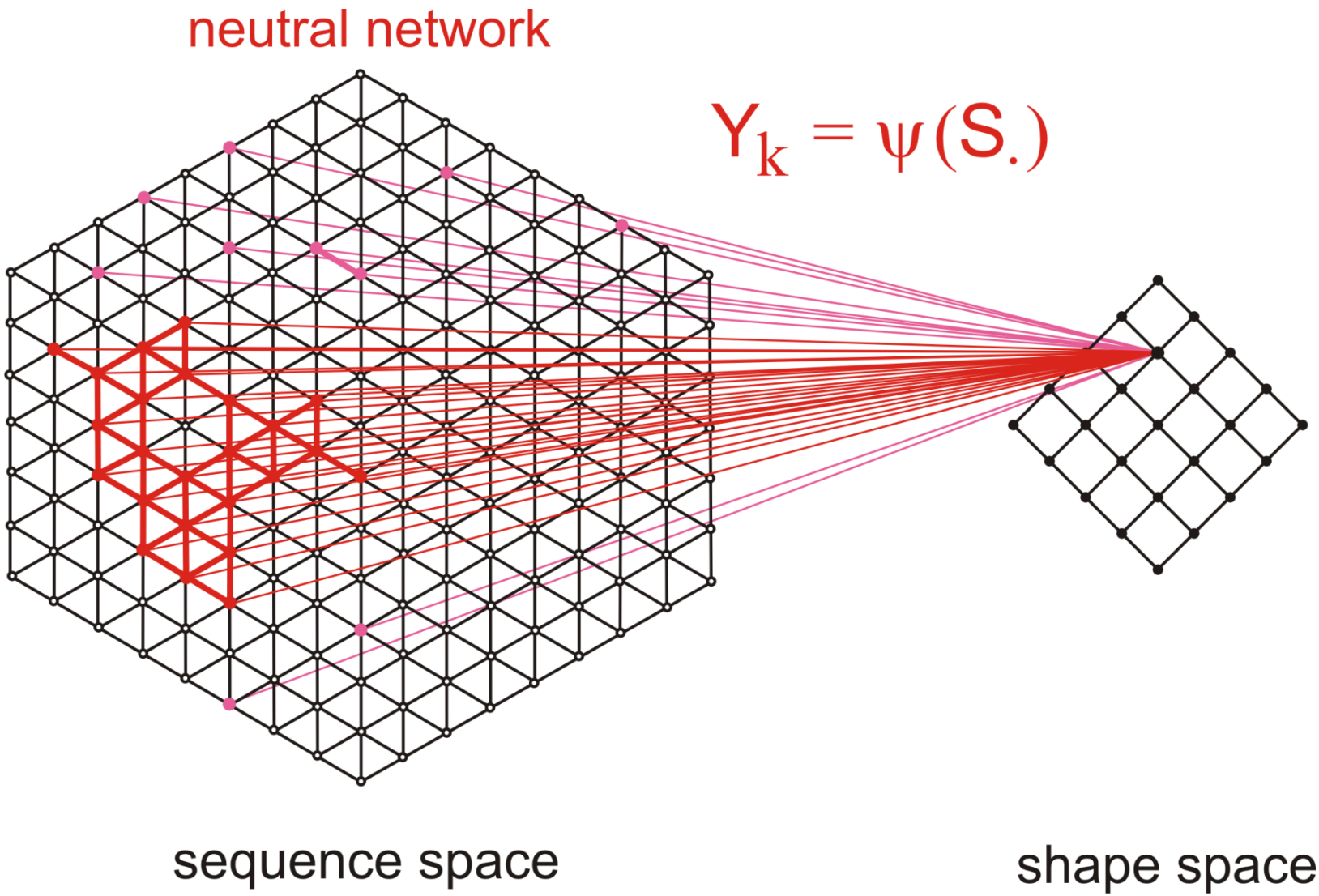
space of phenotypes: $\mathcal{Y} = \{Y_1, Y_2, Y_3, \ldots, Y_M\}$

$N >> M$

$$\Phi(S_j) = Y_k$$

$$G_k = \Phi^{-1}(Y_k) \equiv \{S_j \mid \Phi(S_j) = Y_k\}$$

Inversion of genotype-phenotype mapping



Neutral networks in sequence space

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

PETER SCHUSTER[1,2,3], WALTER FONTANA[3], PETER F. STADLER[2,3] AND IVO L. HOFACKER[2]

[1] *Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, PF 100813, D-07708 Jena, Germany*
[2] *Institut für Theoretische Chemie, Universität Wien, Austria*
[3] *Santa Fe Institute, Santa Fe, U.S.A.*

SUMMARY

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

Proc. R. Soc. Lond. B (1994) **255**, 279–284
Printed in Great Britain

Realistic fitness landscapes

1. Ruggedness: nearby lying genotypes may unfold into very different phenotypes

2. Neutrality: many different genotypes give rise to phenotypes with identical selection behavior

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. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. **Evolutionary dynamics on landscapes**
7. Neutrality
8. Stochasticity, contingency, and history
9. Perspectives

Charles Darwin, 1809-1882

Three necessary conditions for Darwinian evolution are:

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

All three conditions are fulfilled not only by cellular organisms but also by **nucleic acid molecules** – **DNA** or **RNA** – **in** suitable **cell-free experimental assays**:

Darwinian evolution in the test tube

Evolution in the test tube:

G.F. Joyce, *Angew.Chem.Int.Ed.* **46** (2007), 6420-6436

DOI: 10.1002/anie.200701369

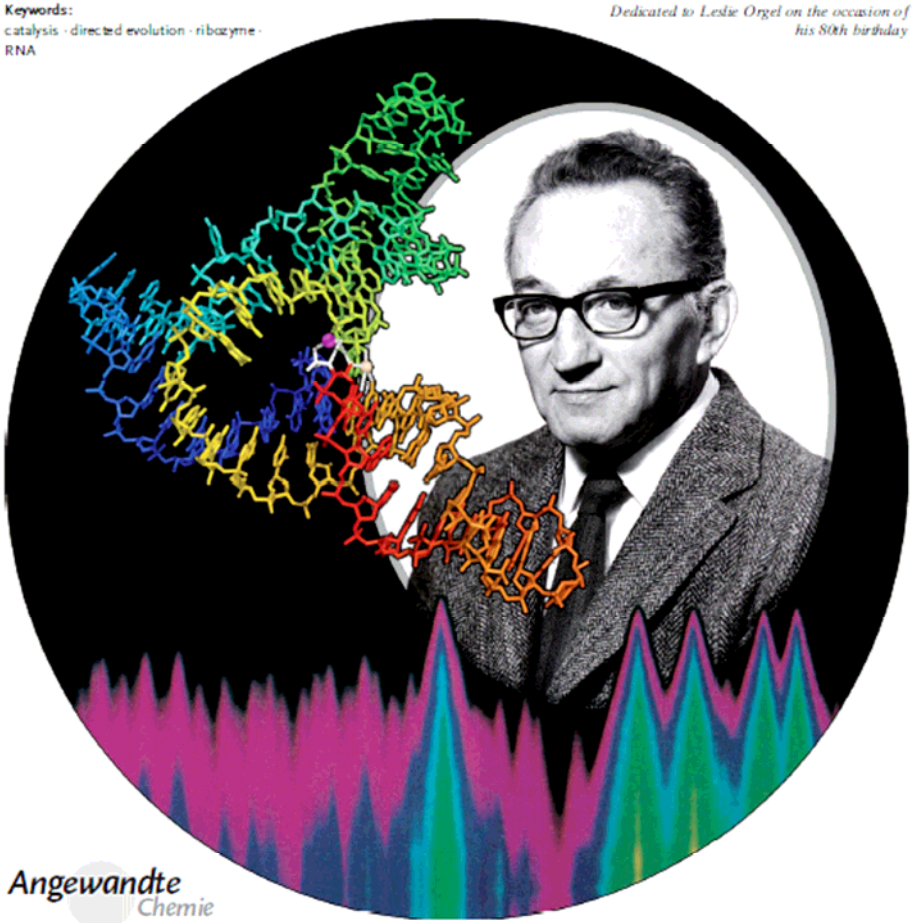
Molecular Evolution

Forty Years of In Vitro Evolution**

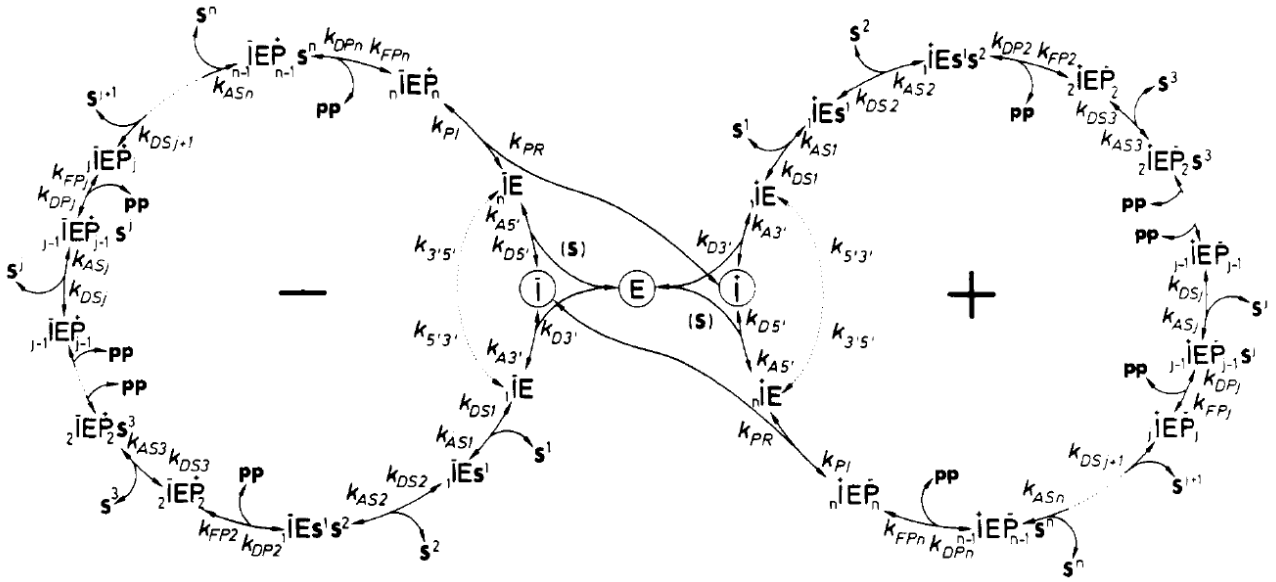
*Gerald F. Joyce**

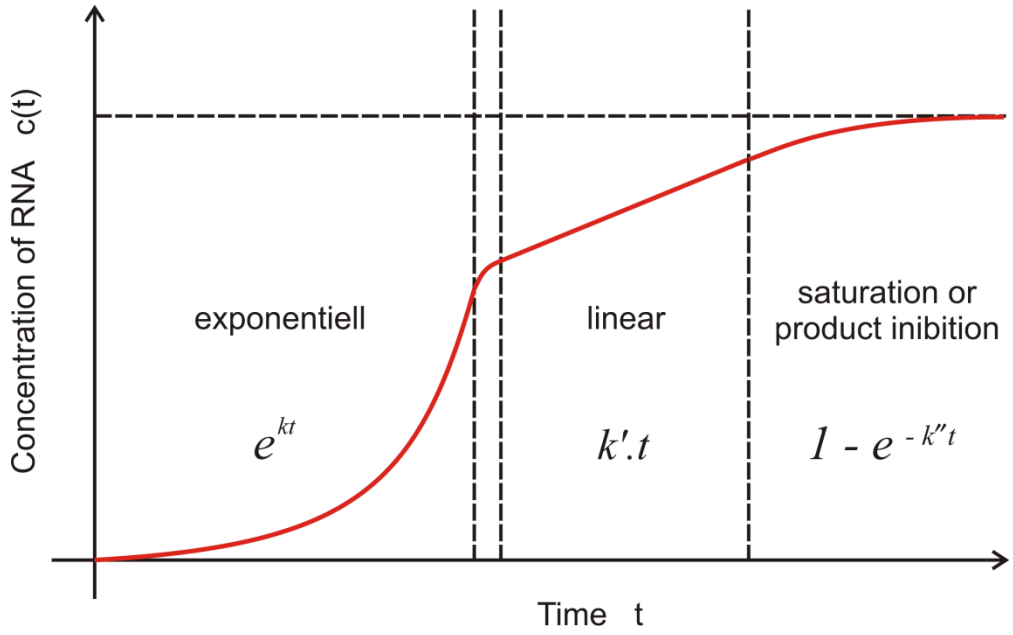
Keywords:
catalysis · directed evolution · ribozyme · RNA

Dedicated to Leslie Orgel on the occasion of his 80th birthday



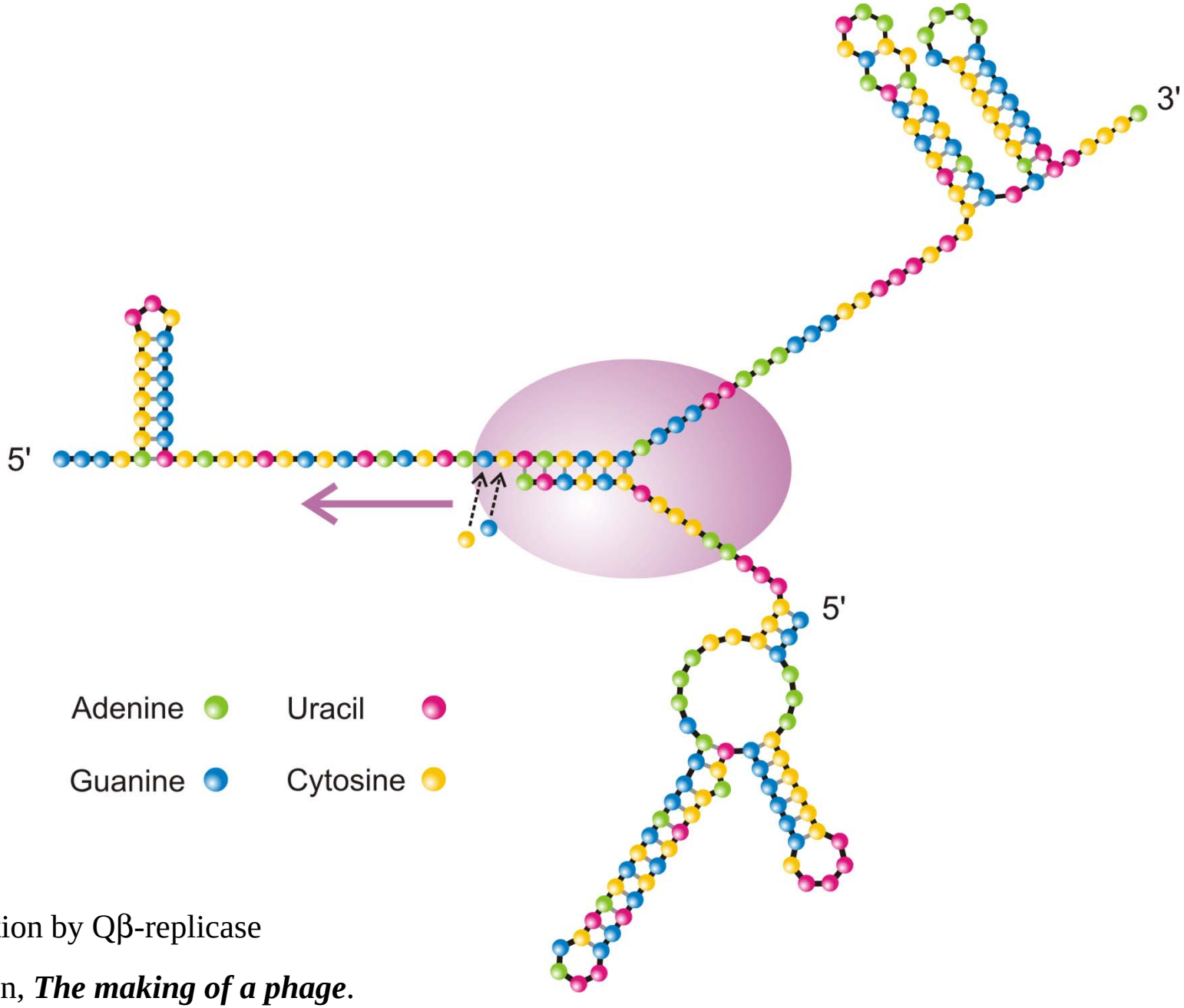
Christof K. Biebricher,
1941-2009





Kinetics of RNA replication

C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
Biochemistry **22**:2544-2559, 1983



RNA replication by Qβ-replicase

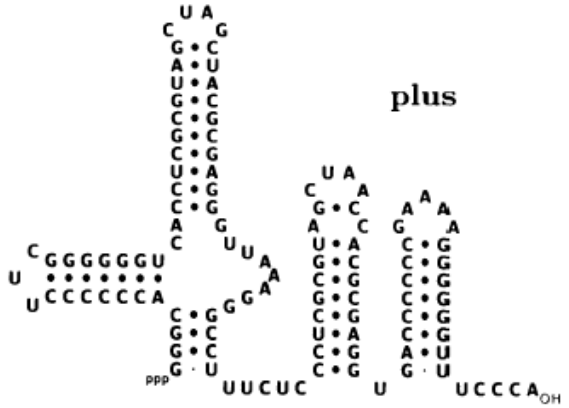
C. Weissmann, ***The making of a phage***.
FEBS Letters **40** (1974), S10-S18

plus SV-11 (115 b)

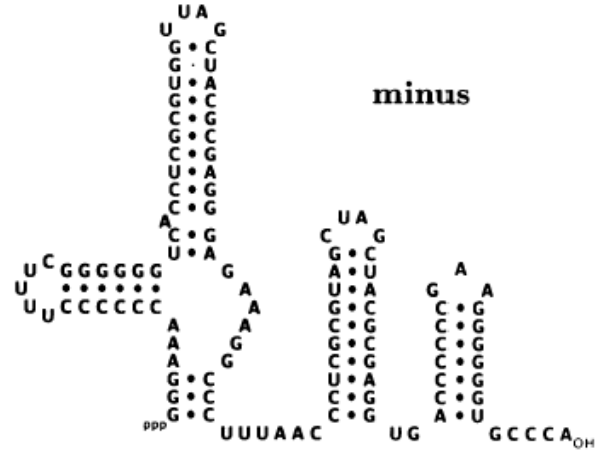
minus

stable

does not replicate!



plus SV-11 (115 b)



minus

metastable

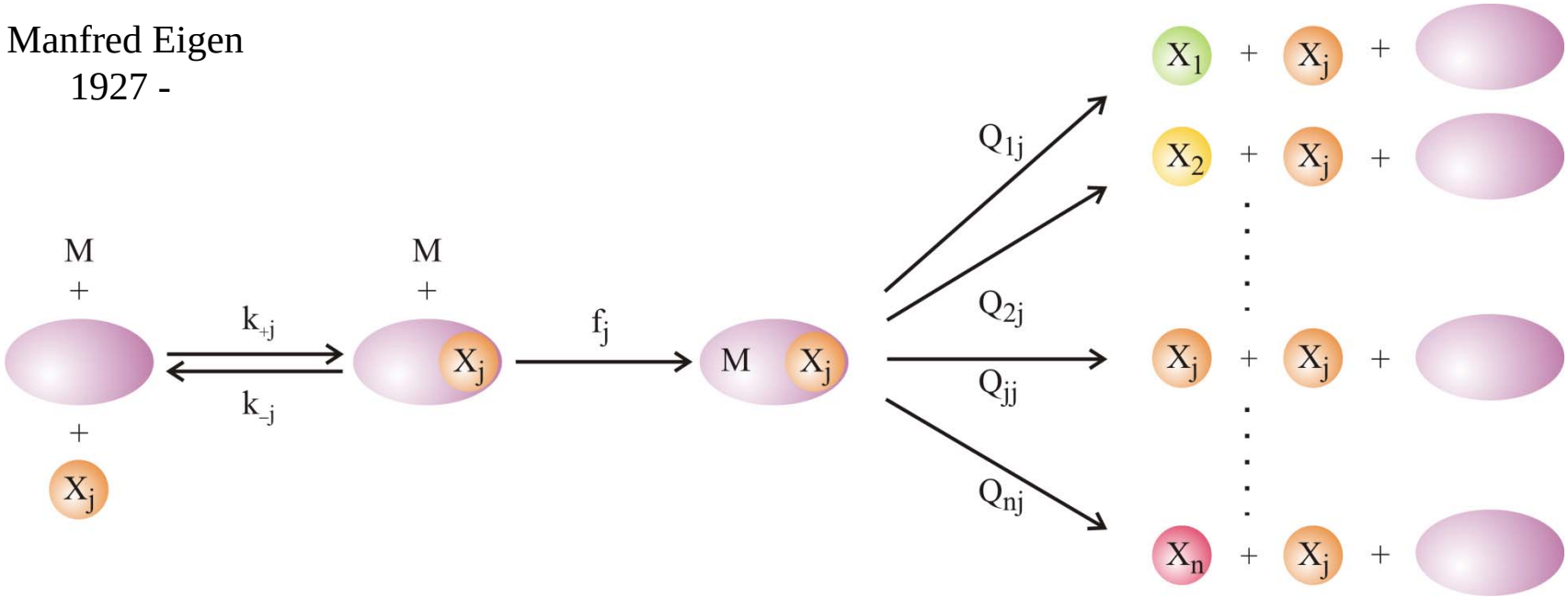
replicates!

C.K. Biebricher, R. Luce. 1992. *In vitro* recombination and terminal recombination of RNA by Qβ replicase. *The EMBO Journal* 11:5129-5135.

Manfred Eigen
1927 -

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$\Phi = \sum_{i=1}^{n} f_i\, x_i \Big/ \sum_{i=1}^{n} x_i$$



Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977. *Naturwissenschaften* 64:541, 65:7 und 65:341

Mutation-selection equation: $[\mathrm{I}_i] = x_i \geq 0,\ f_i > 0,\ Q_{ij} \geq 0$

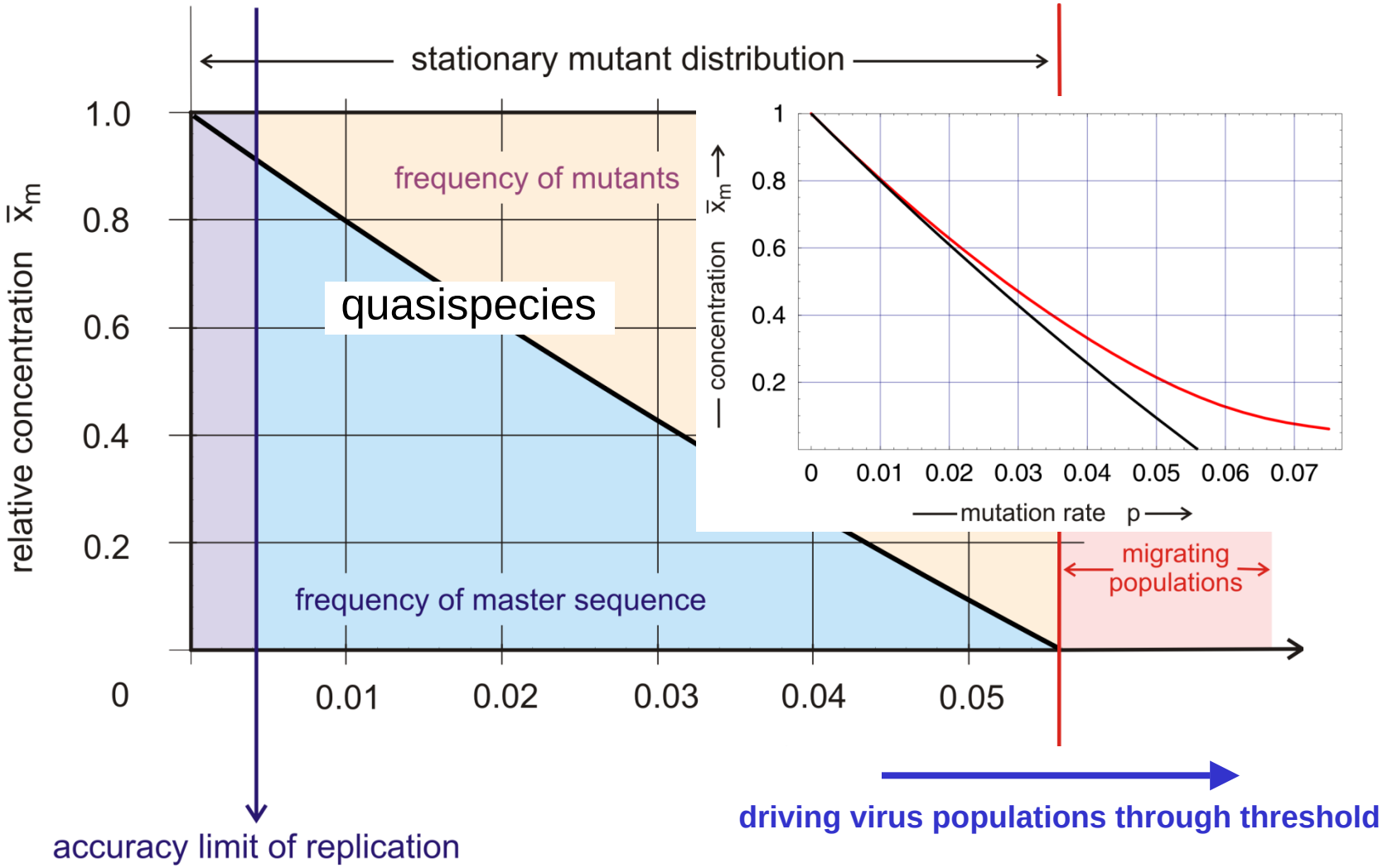
$$\frac{dx_i}{dt} = \sum_{j=1}^{n} f_j Q_{ji}\, x_j - x_i\, \phi, \quad i = 1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \overline{f}$$

Solutions are obtained after integrating factor transformation by means of an eigenvalue problem

$$x_i(t) = \frac{\sum_{k=0}^{n-1} \ell_{ik} \cdot c_k(0) \cdot \exp(\lambda_k t)}{\sum_{j=1}^{n} \sum_{k=0}^{n-1} \ell_{jk} \cdot c_k(0) \cdot \exp(\lambda_k t)}; \quad i = 1,2,\cdots,n; \quad c_k(0) = \sum_{i=1}^{n} h_{ki}\, x_i(0)$$

$$W \div \left\{ f_i Q_{ij};\ i, j = 1,2,\cdots,n \right\};\ L = \left\{ \ell_{ij};\ i, j = 1,2,\cdots,n \right\};\ L^{-1} = H = \left\{ h_{ij};\ i, j = 1,2,\cdots,n \right\}$$

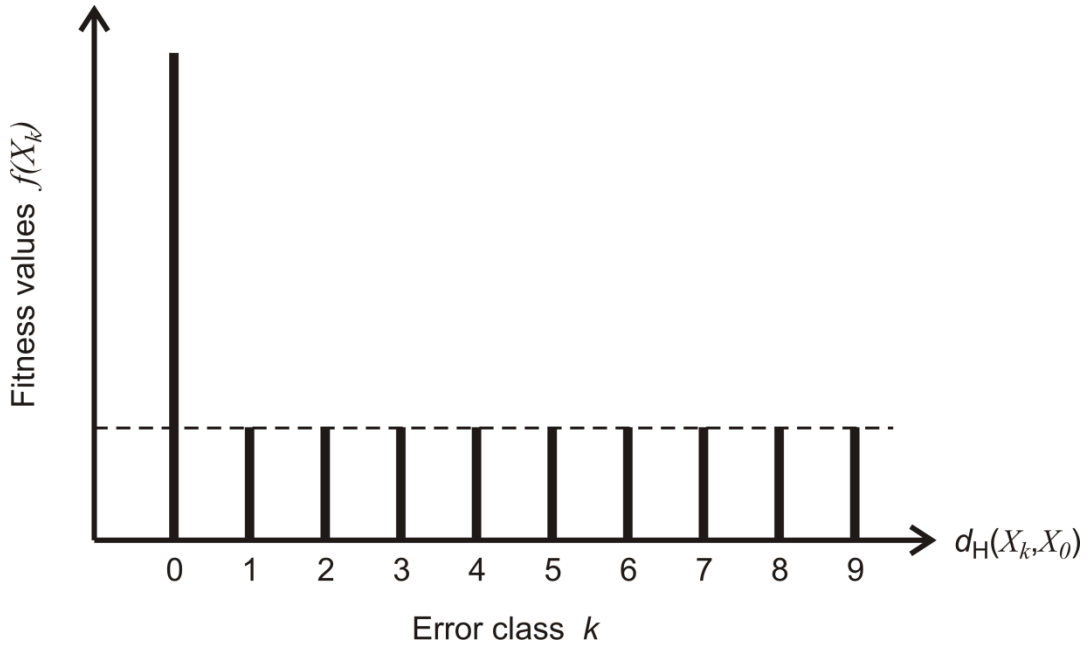
$$L^{-1} \cdot W \cdot L \quad = \quad \Lambda \quad = \quad \left\{ \lambda_k; k = 0,1,\cdots,n-1 \right\}$$



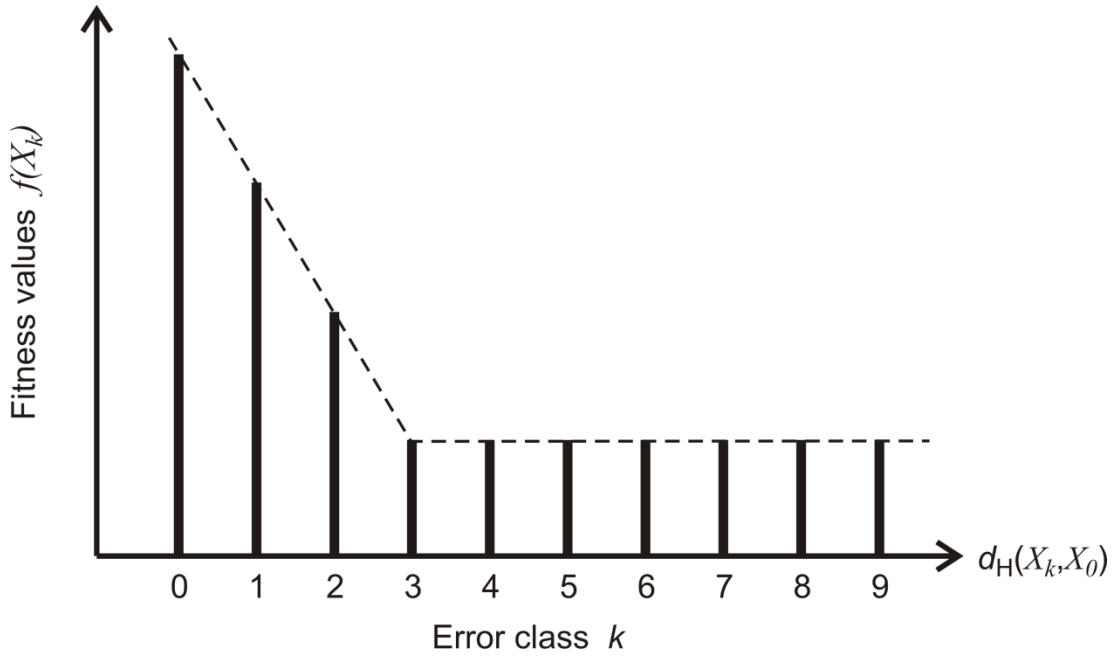
The error threshold in replication and mutation

Model fitness landscapes I

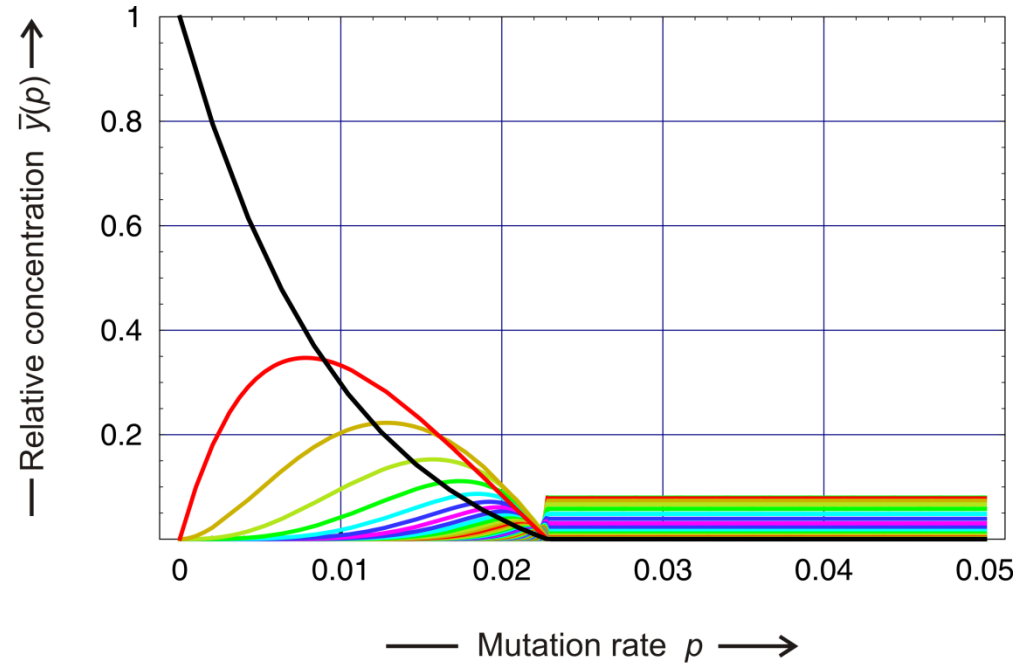
single peak landscape

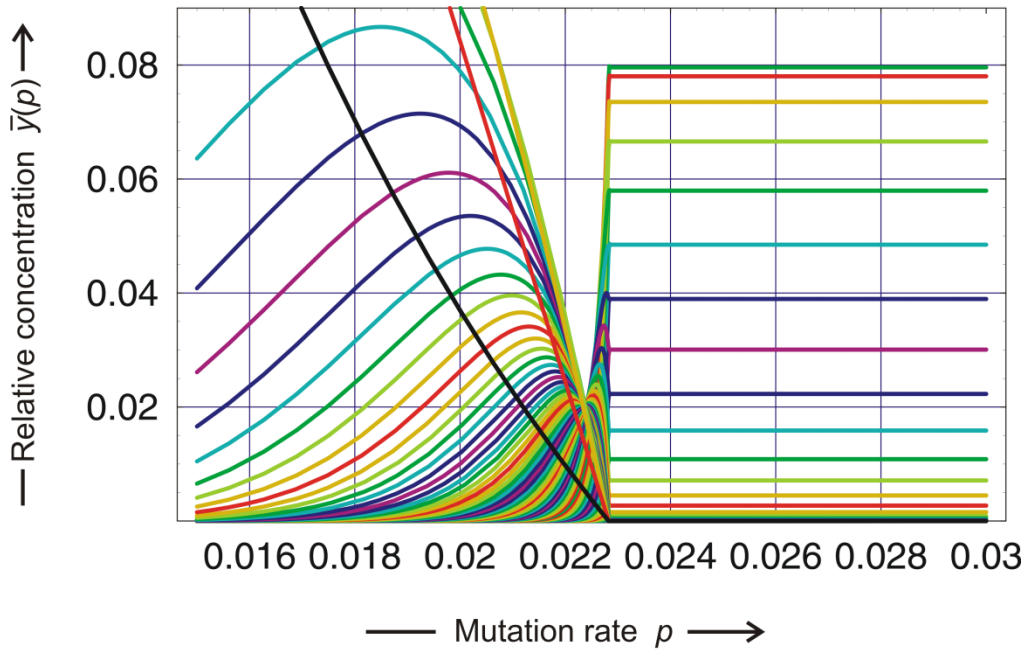


step linear landscape

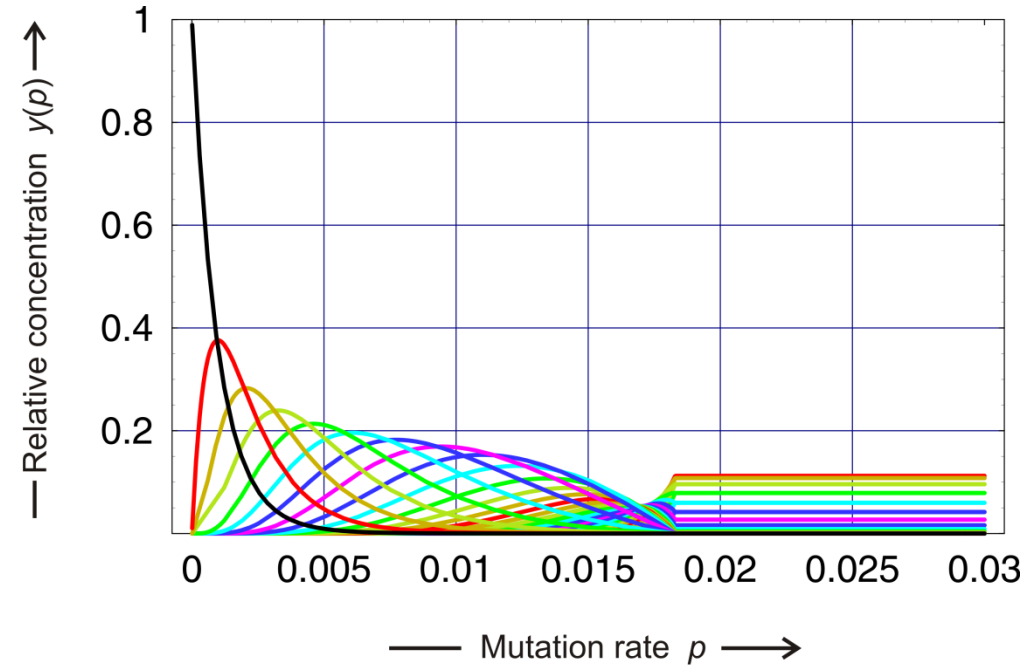


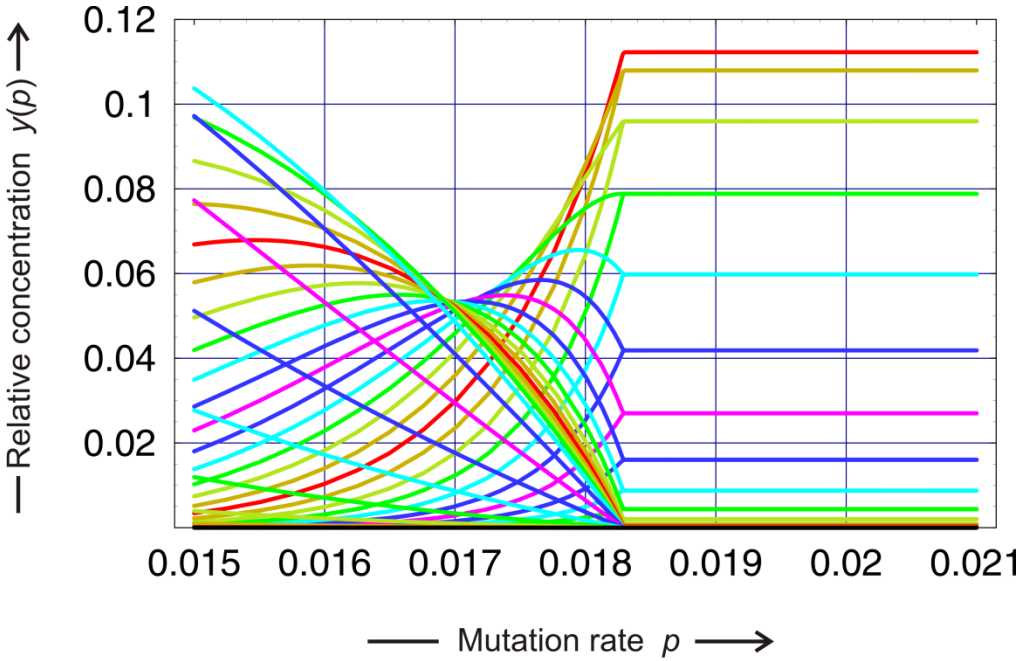
Error threshold on the single peak landscape



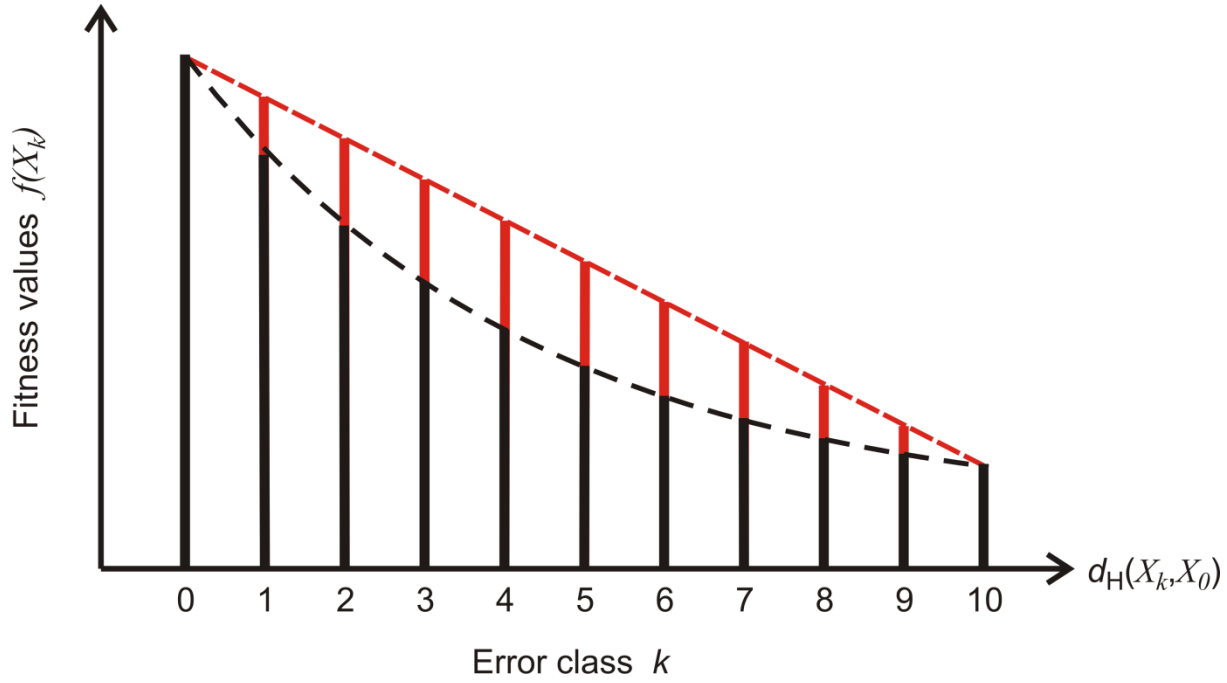


Error threshold on the step linear landscape

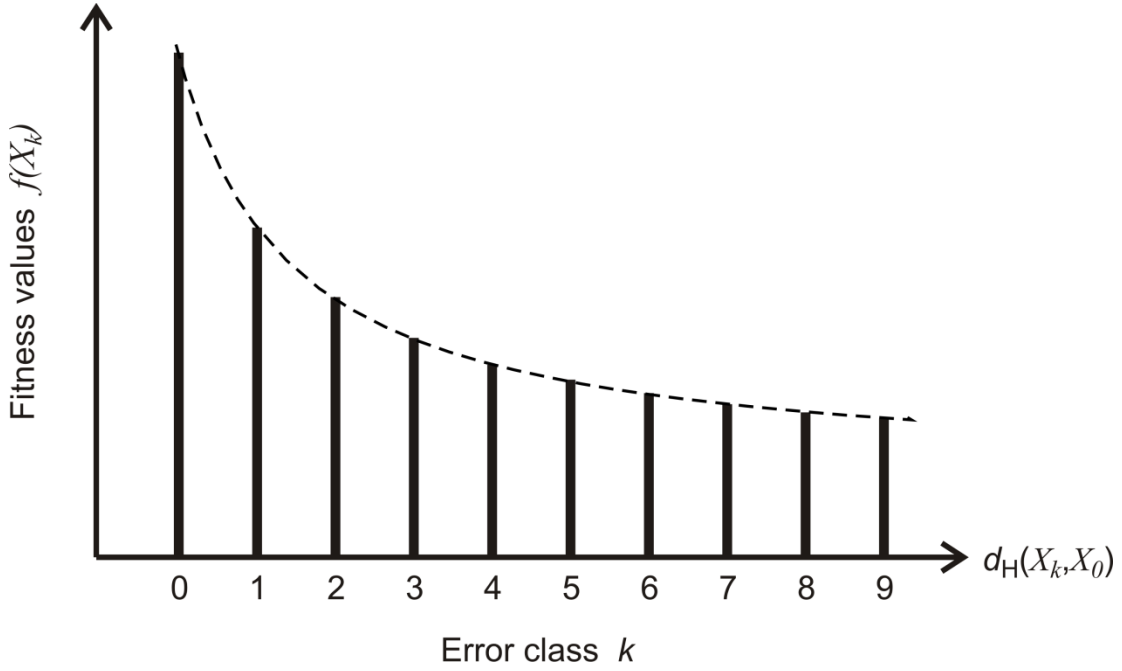




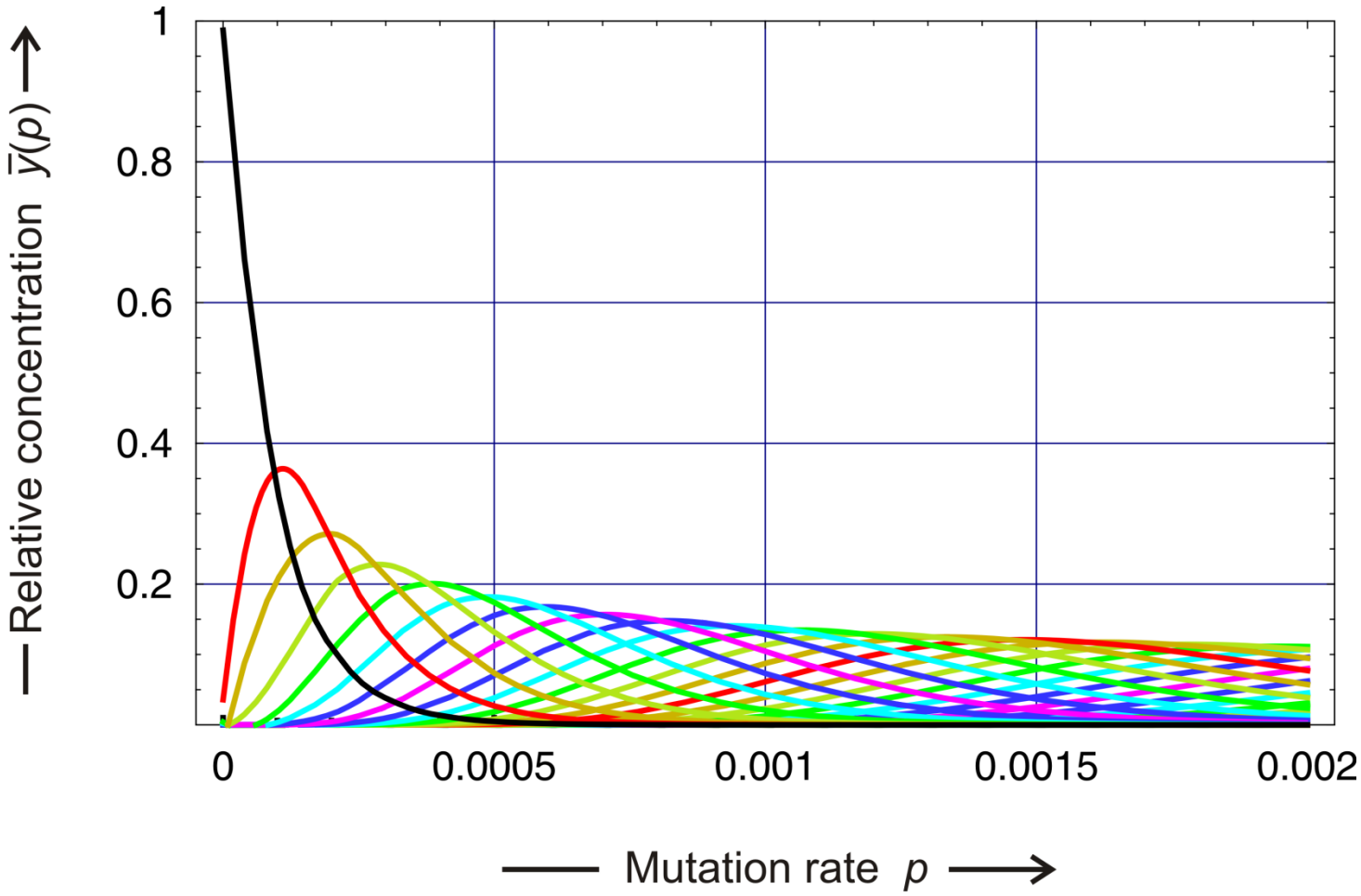
linear and
multiplicative



hyperbolic

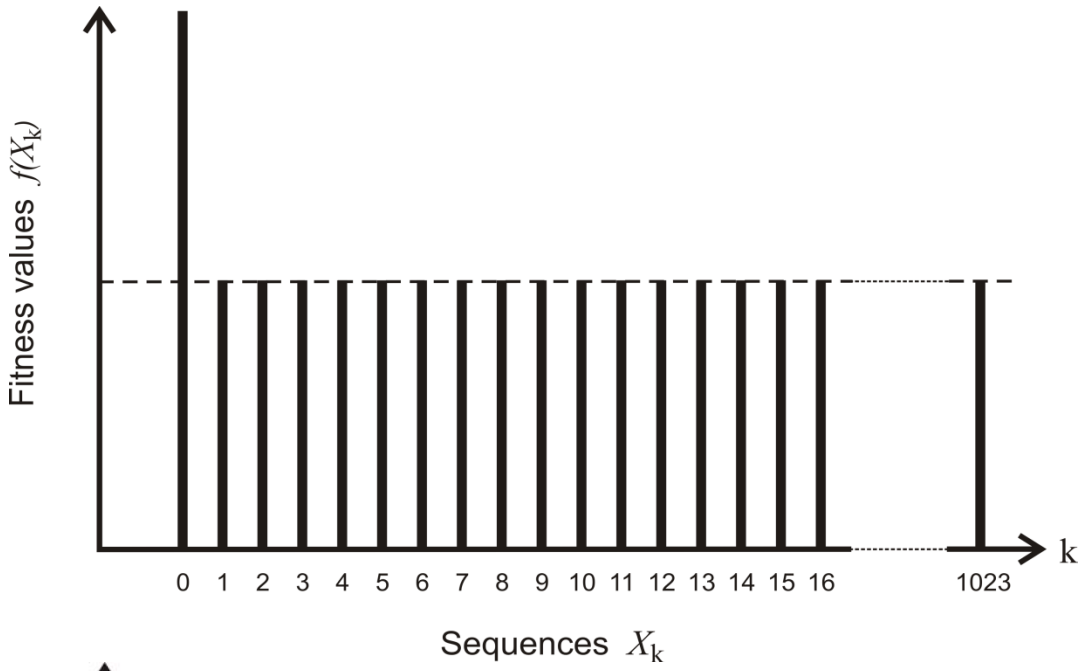


Model fitness landscapes II



The linear fitness landscape shows no error threshold

single peak landscape



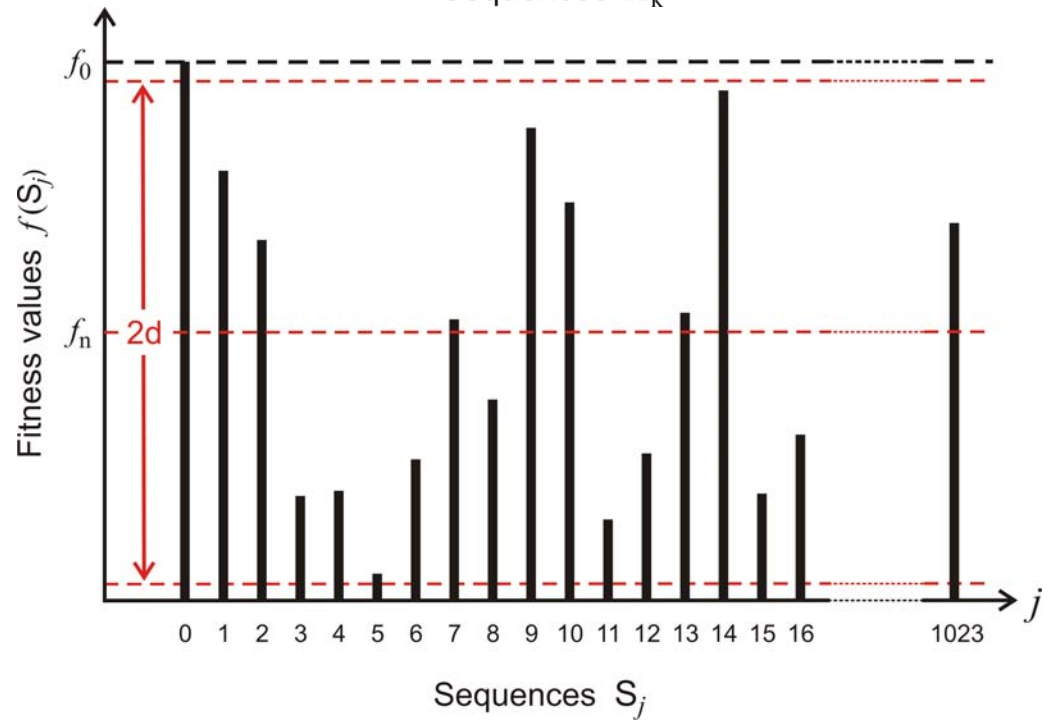
$$f(\mathsf{S}_j) = f_n + 2\,d\,(f_0 - f_n)\left(\eta_j^{(s)} - 0.5\right)$$

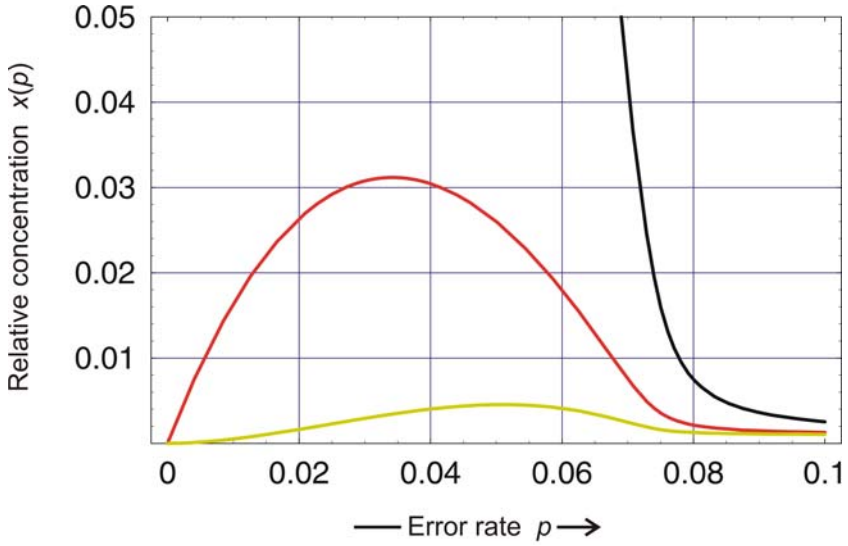
$$j = 1, 2, \ldots, N\,;\; j \neq m\,,$$

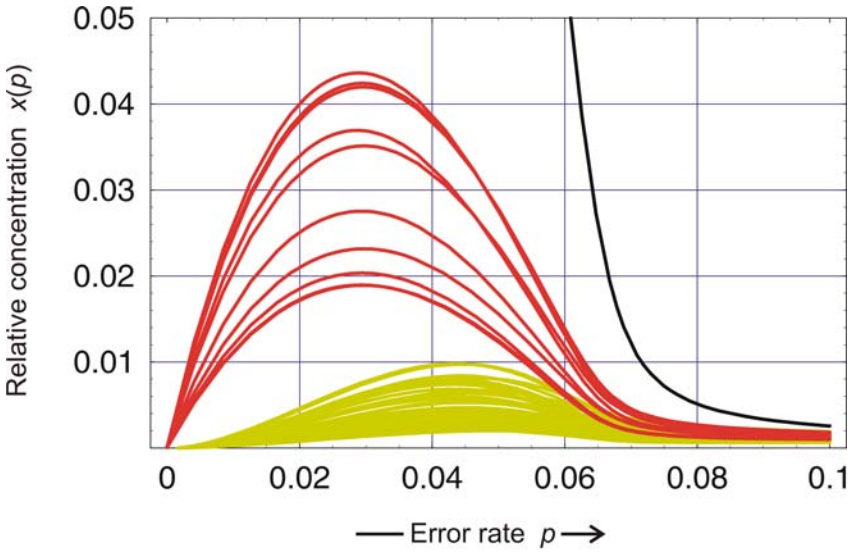
η ... random number ; s ... seeds

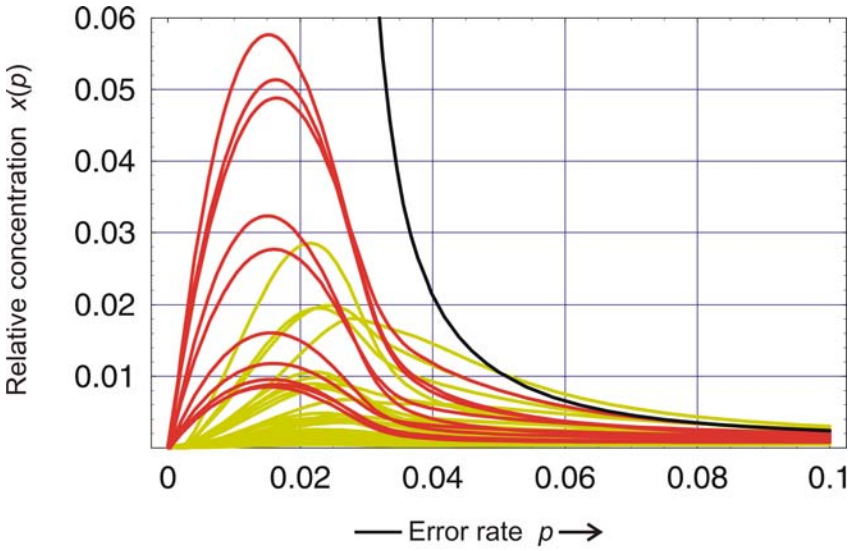
„realistic“ landscape

Rugged fitness landscapes
over individual binary sequences
with n = 10





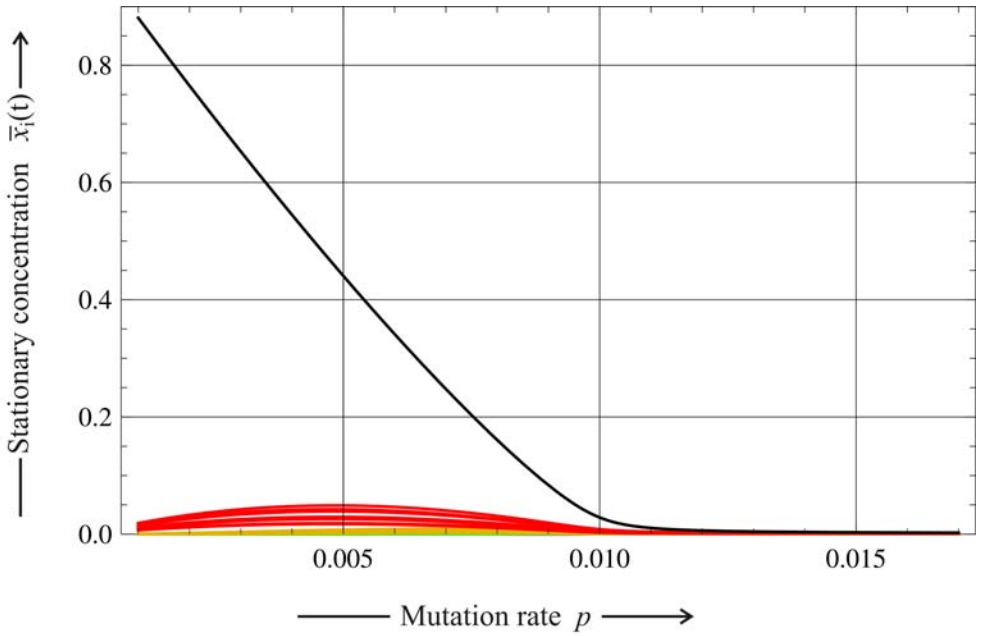




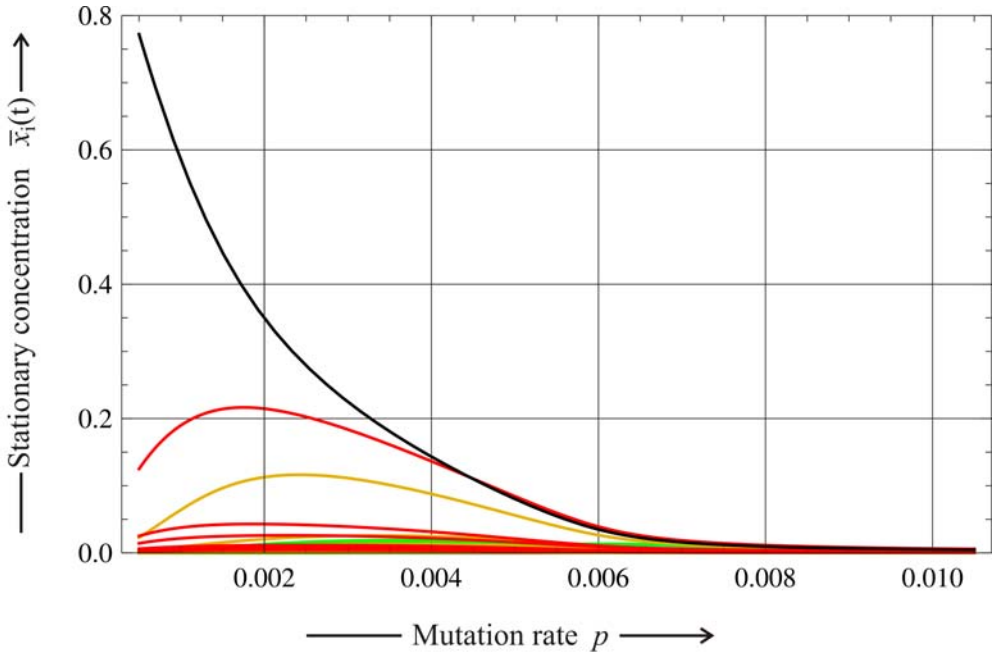
Error threshold: Individual sequences

n = 10, σ = 2, s = 491 and d = 0, 0.5, 0.9375

d = 0.5



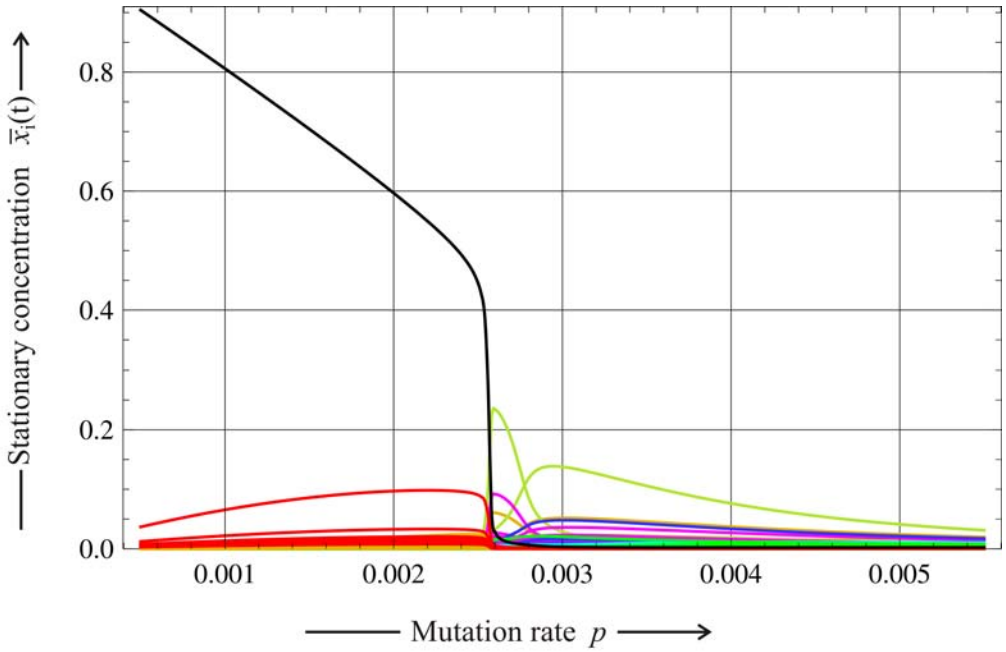
d = 1.0



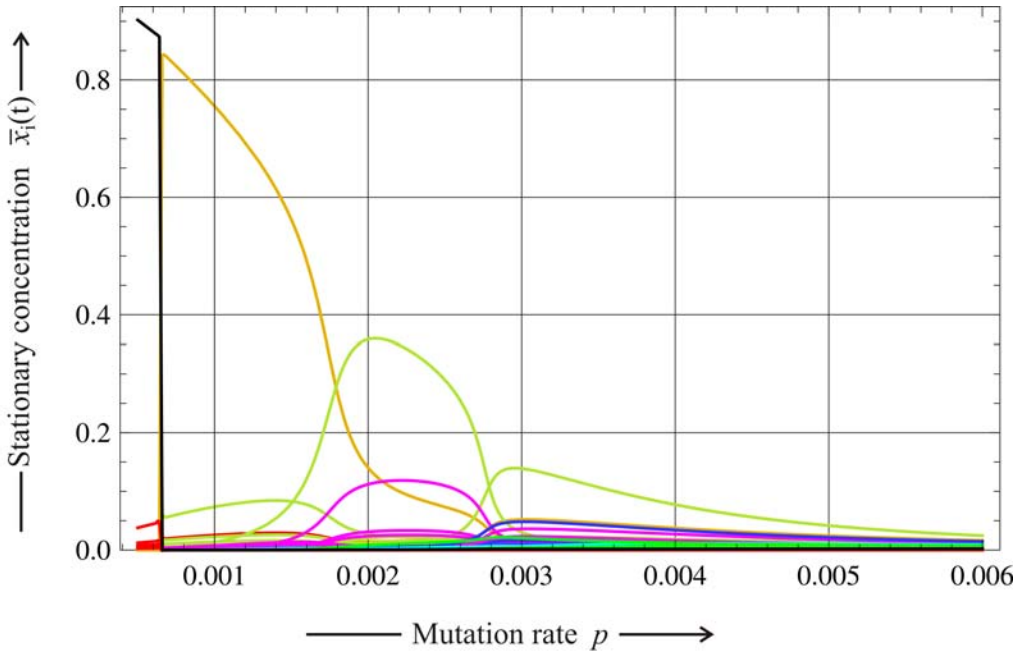
Case I: Strong quasispecies

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 919

d = 0.995



d = 1.0



Case III: multiple transitions

n = 10, $f_0 = 1.1, f_n = 1.0$, s = 637

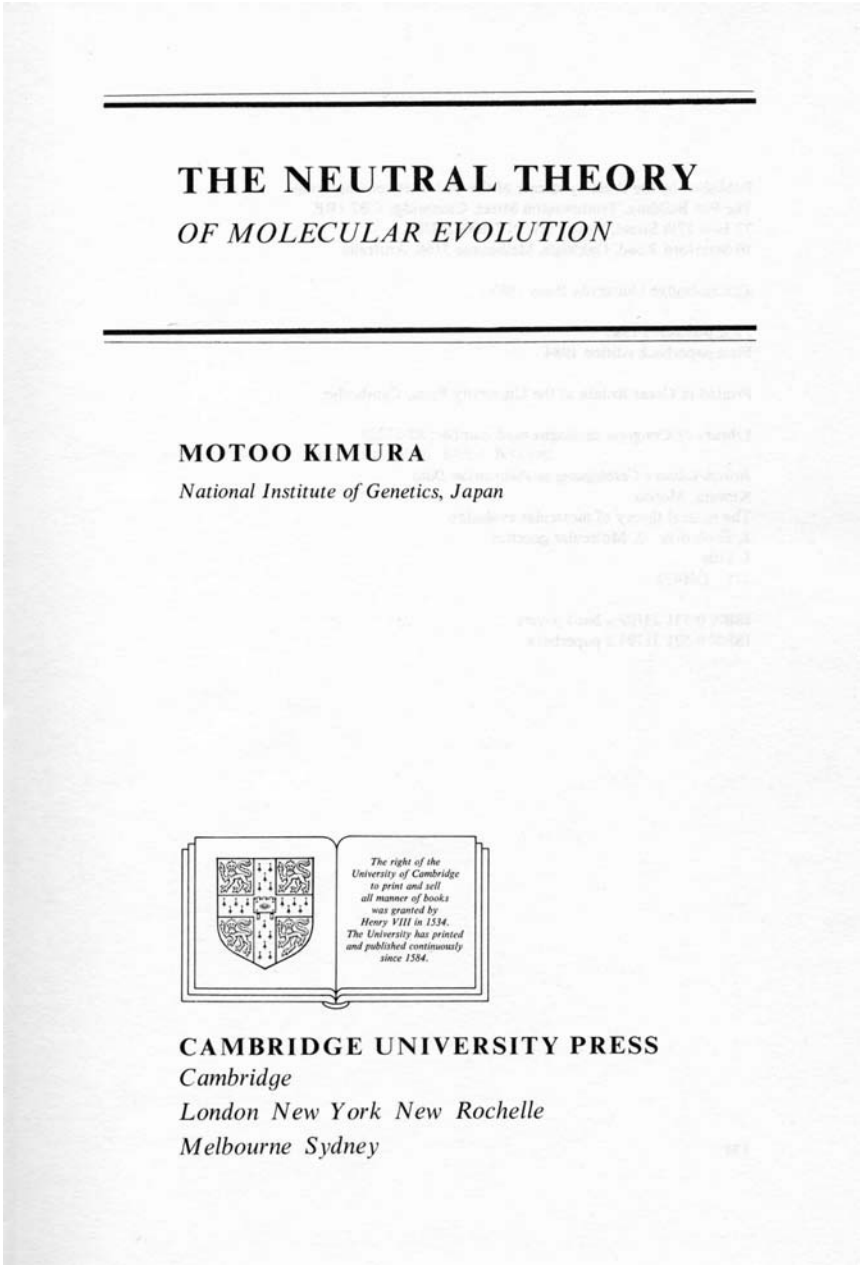
1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. **Neutrality**
8. Stochasticity, contingency, and history
9. Perspectives

MOTOO KIMURA

Motoo Kimuras population genetics of neutral evolution.

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

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



THE NEUTRAL THEORY

OF MOLECULAR EVOLUTION

MOTOO KIMURA

National Institute of Genetics, Japan

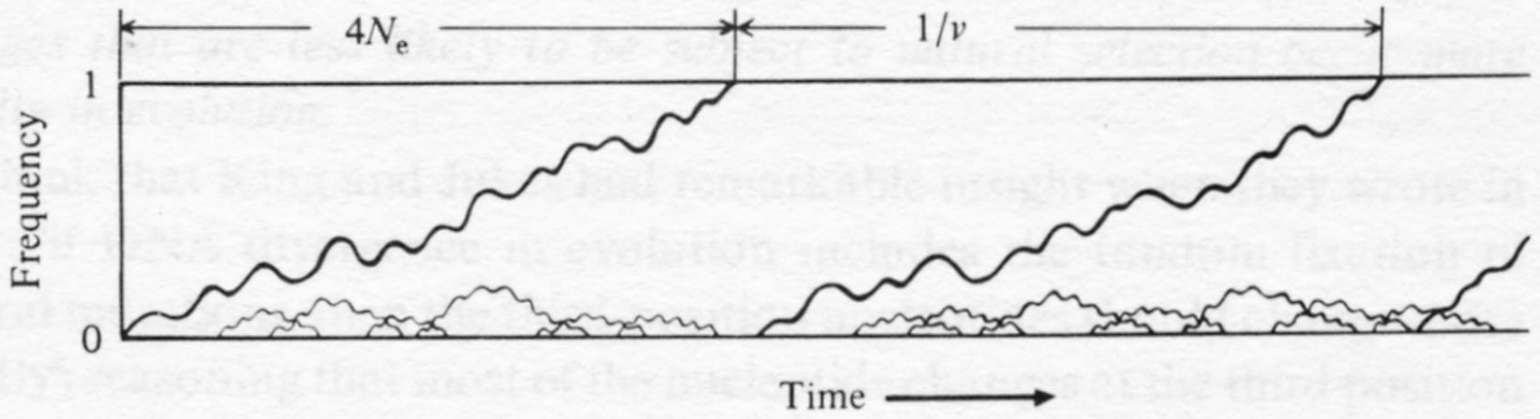
CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York New Rochelle

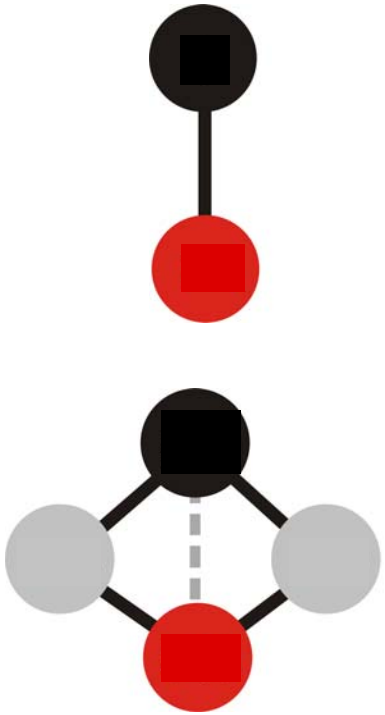
Melbourne Sydney

Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



Motoo Kimura

Is the Kimura scenario correct for frequent mutations?



$d_H = 1$

$$\lim_{p\to 0} x_1(p) = x_2(p) = 0.5$$

$d_H = 2$

$$\lim_{p\to 0} x_1(p) = a$$

$$\lim_{p\to 0} x_2(p) = 1-a$$

$d_H \geq 3$

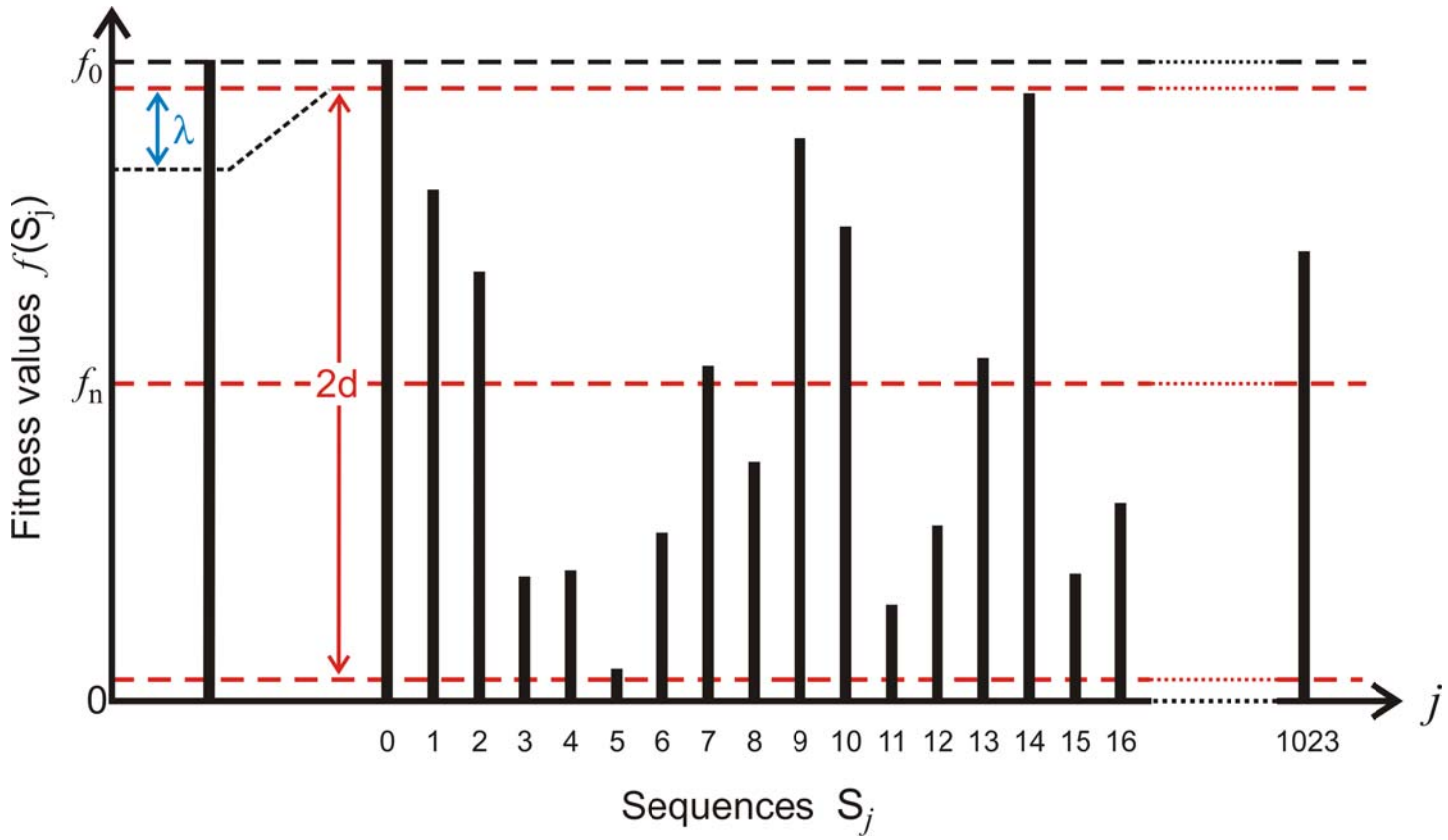
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \text{ or}$$

$$\lim_{p\to 0} x_1(p) = 0, \lim_{p\to 0} x_2(p) = 1$$

Pairs of neutral sequences in replication networks

Random fixation in the sense of Motoo Kimura

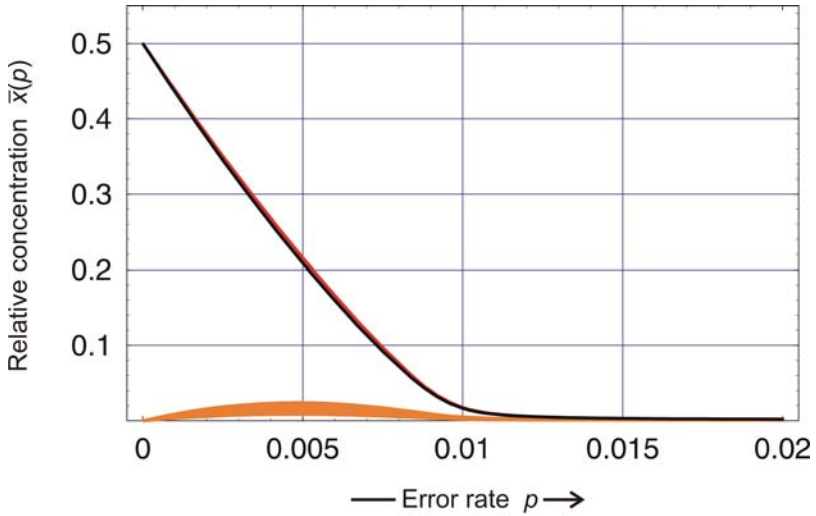
P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650

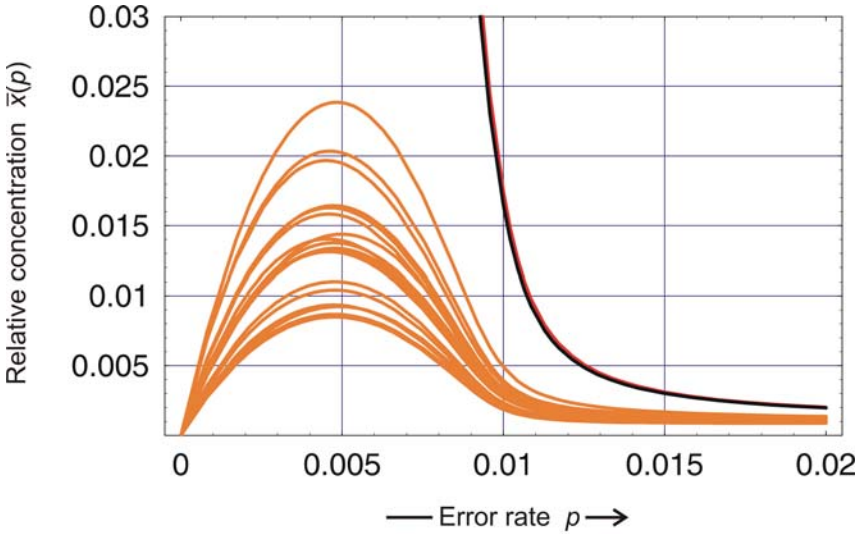


$$f(S_j) = \begin{cases} f_0 & \text{if } \eta_j(s) \geq 1-\lambda\,, \\ f_n + \frac{2d}{1-\lambda}\,(f_0 - f_n)\left(\eta_j^{(s)} - 0.5\right) & \text{if } \eta_j(s) \leq 1-\lambda\,, \end{cases}$$

$$j = 1, 2, \ldots, N\,;\ j \neq m\,;\ \eta \ldots \text{random number}\,;\ s \ldots \text{seeds}$$

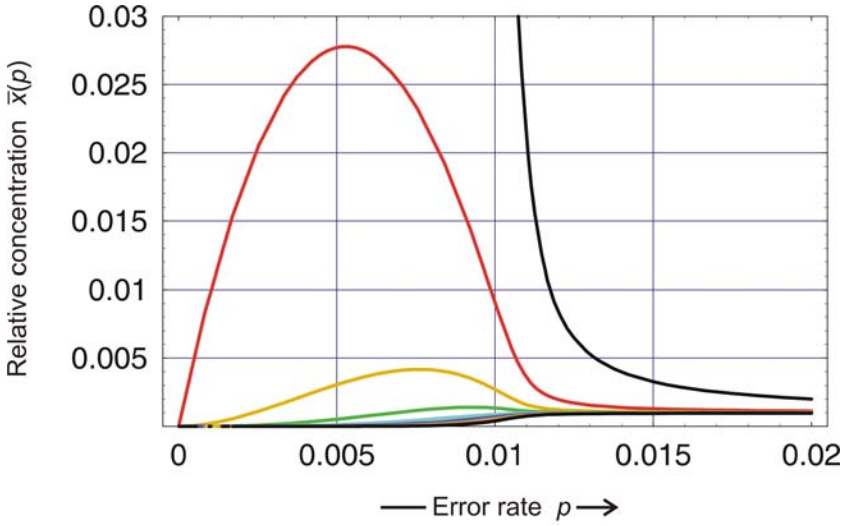
A fitness landscape including neutrality





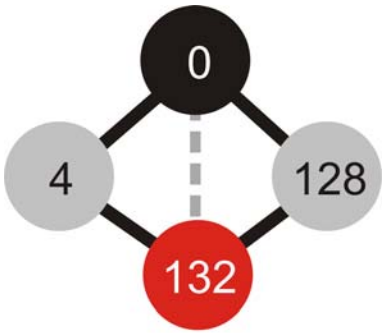
Neutral network

$\lambda = 0.01$, s = 367



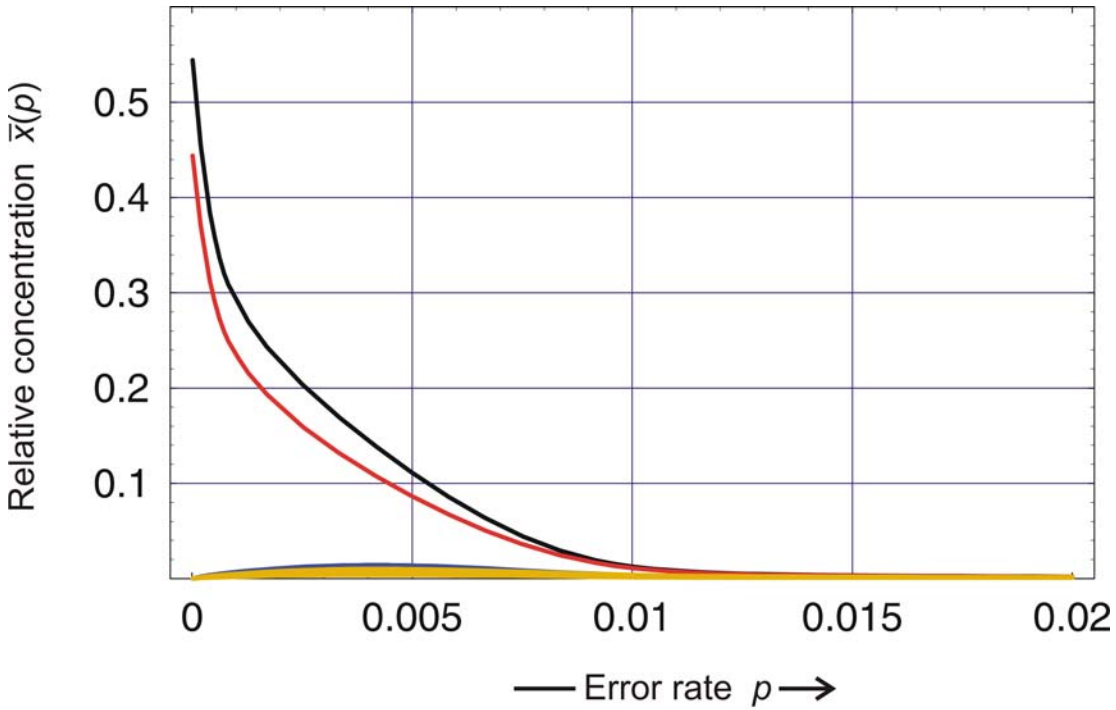
Neutral network: Individual sequences

n = 10, σ = 1.1, d = 0.5



Neutral network

$\lambda = 0.01,\ s = 877$



Neutral network: Individual sequences

$n = 10, \sigma = 1.1, d = 0.5$

```
..... ACAUGCGAA .....
..... AUAUACGAA .....
..... ACAUGCGCA .....
..... GCAUACGAA .....
..... ACAUGCUAA .....
..... ACAUGCGAG .....
..... ACACGCGAA .....
..... ACGUACGAA .....
..... ACAUAGGAA .....
..... ACAUACGAA .....

..... ACAU[G/A]CGAA .....
```

master sequence 1

master sequence 2

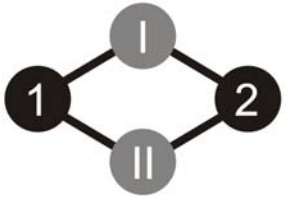
consensus sequence

```
..... ACAGUCAGAA .....
..... ACAGUCCGAA .....
..... AUAAUCCGAA .....
..... ACAGUCAGCA .....
..... GCAGUCAGAA .....
..... ACAGUCAUAA .....
..... ACAGUCAGAG .....
..... ACAACCCGAA .....
..... ACGGUCAGAA .....
..... ACAGUGAGAA .....
..... ACAAUCAGAA .....
..... ACAAUCCGAA .....

..... ACA[G/A]UC[A/C]GAA .....
```

master sequence 1

intermediate I

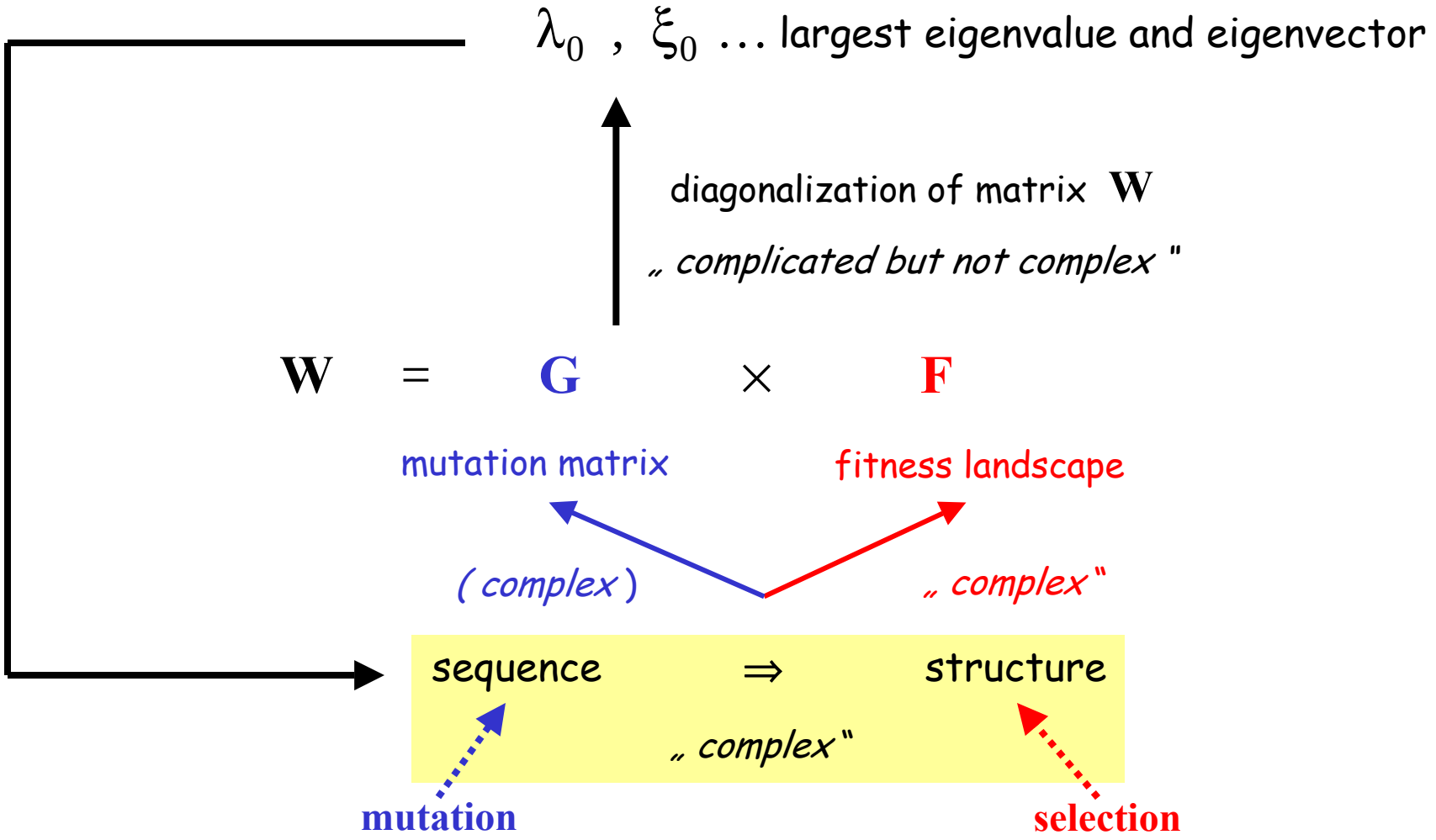


intermediate II

master sequence 2

consensus sequence

Consensus sequence of a quasispecies with strongly coupled sequences of Hamming distance $d_H(X_i,X_j) = 1$ and 2.



Complexity in molecular evolution

1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. **Stochasticity, contingency, and history**
9. Perspectives

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.

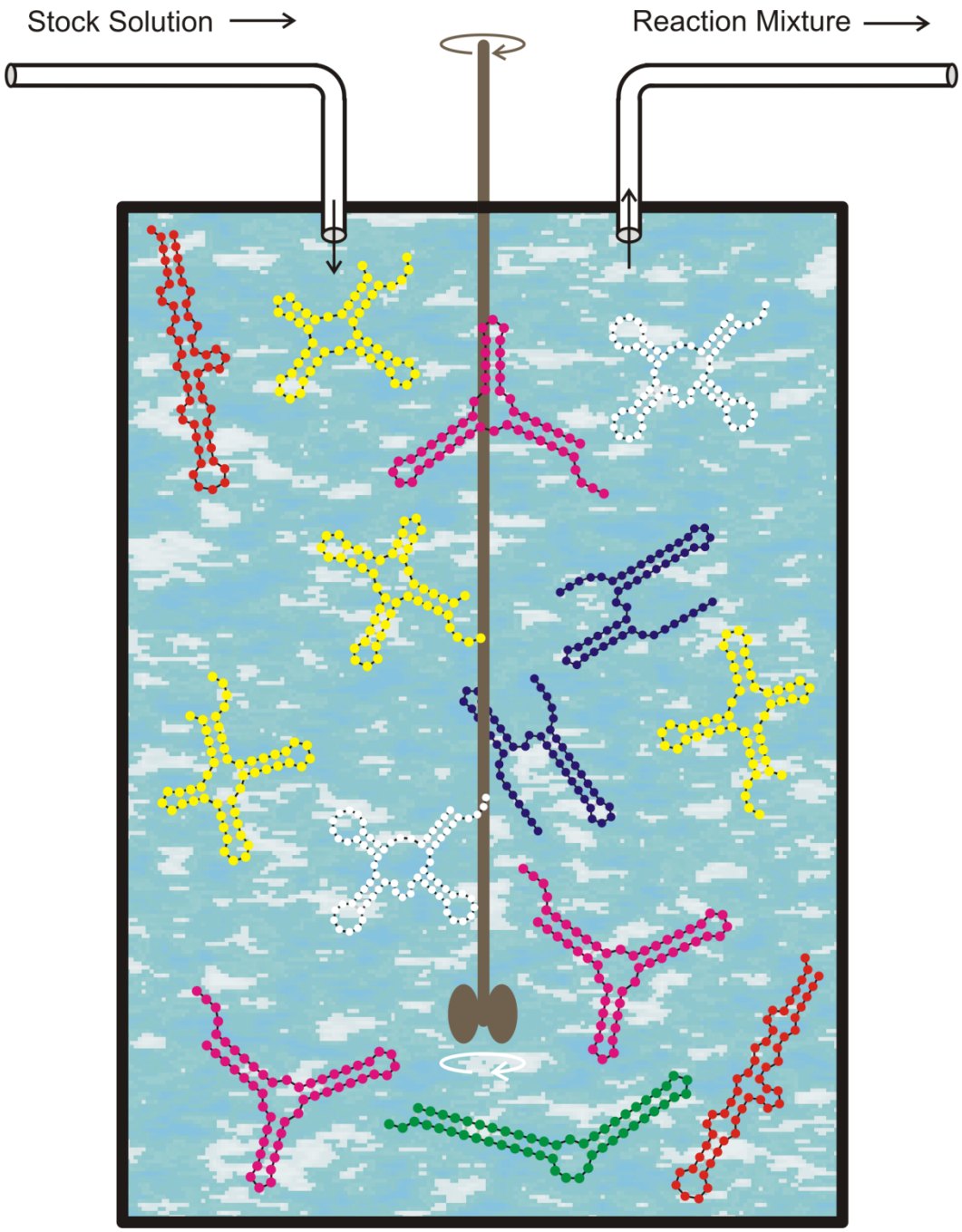
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level 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:

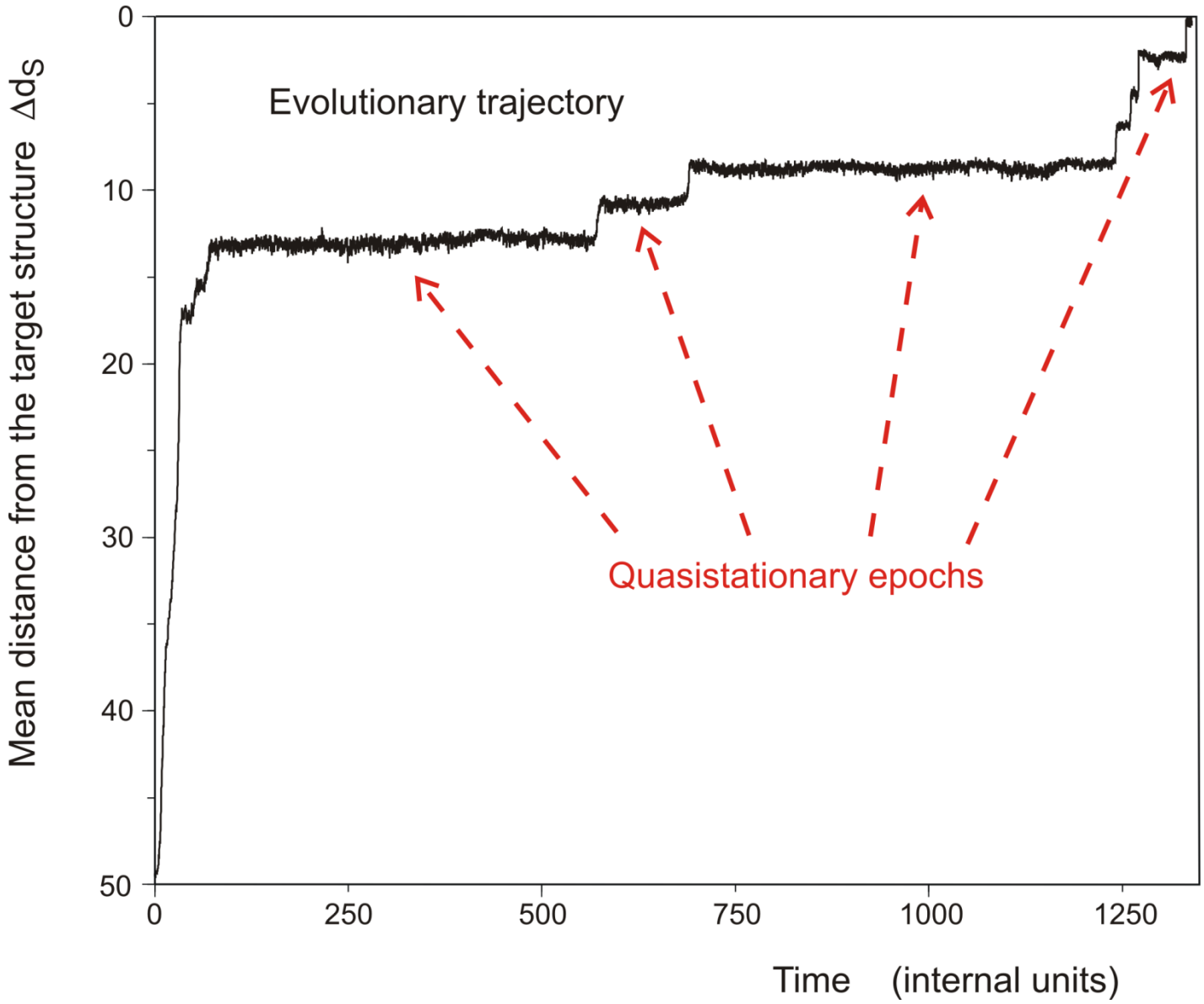
Population size, N = # RNA molecules, is controlled by the flow

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

Mutation rate:

p = 0.001 / site × replication

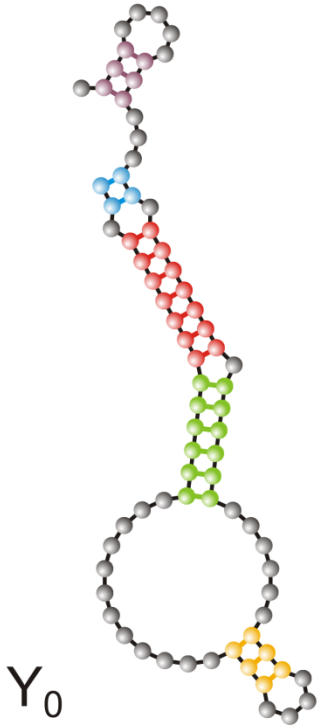
The flowreactor as a device for **studies** of evolution ***in vitro*** and ***in silico***



In silico optimization in the flow reactor: Evolutionary Trajectory

S_0: GUUAUGGGCGAUGAGGAGUAGAGAAAAAACCAAUCGGUGAAAGAUCUCGUGUGCCCAUUGCAUAGCCCGUACGGCA

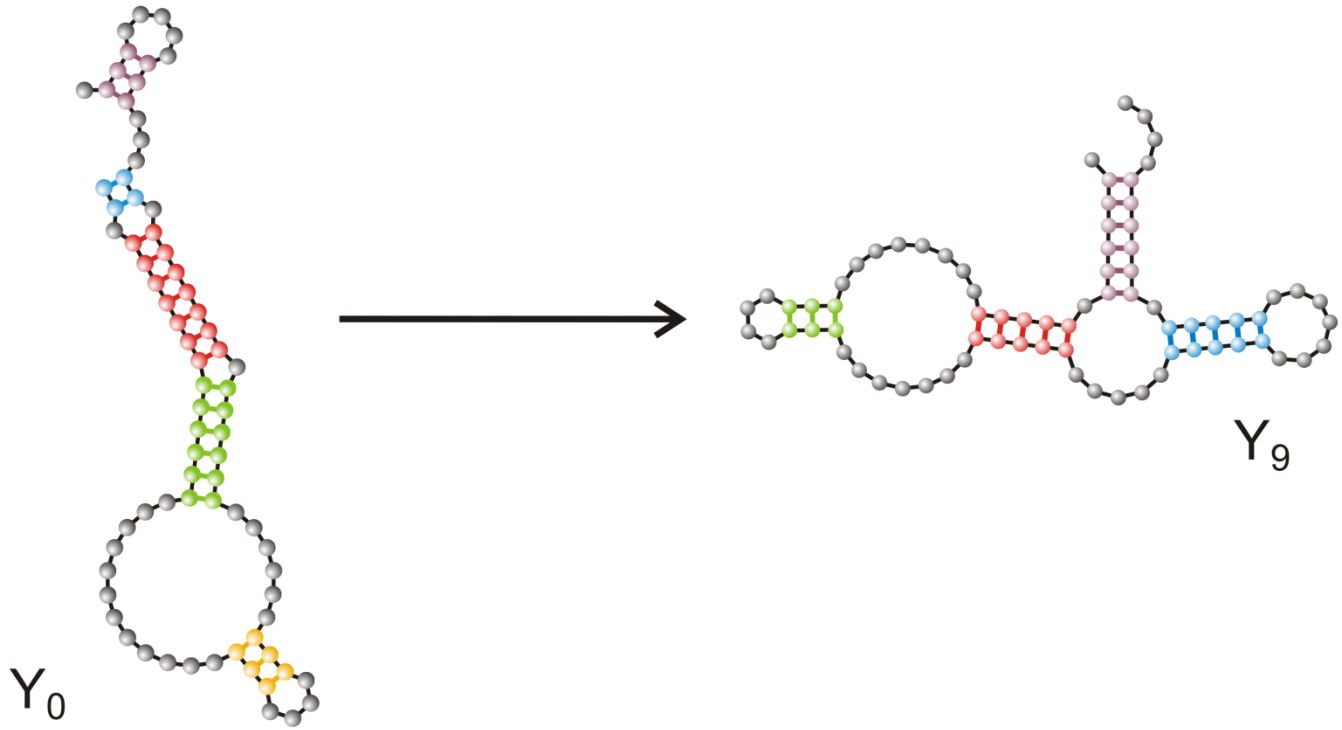
Randomly chosen initial structure



Y_0

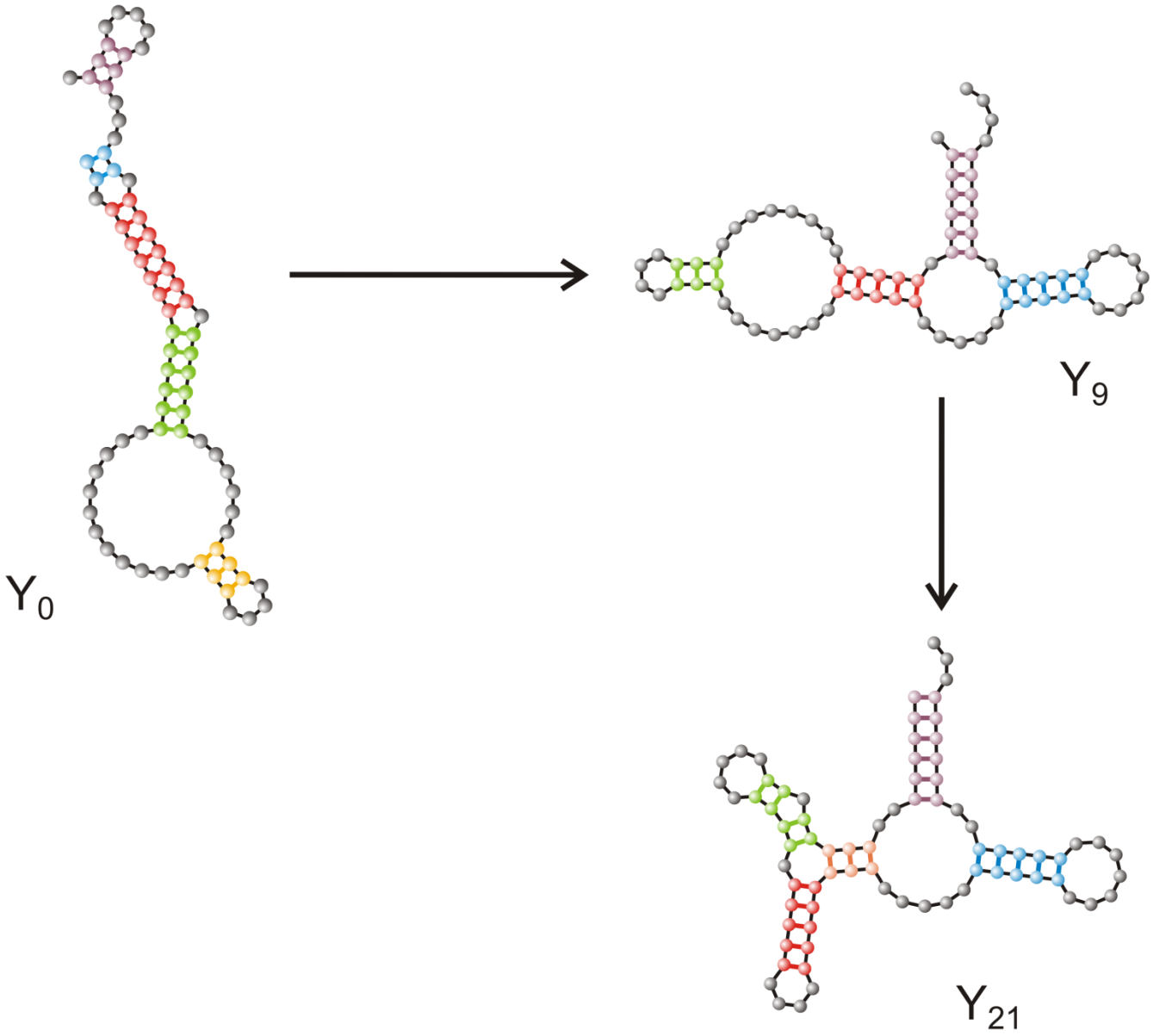
S_0: GUUAUGGGCGAUGAGGAGUAGAGAAAAAACCAAUCGGUGAAAGAUCUCGUGUGCCCAUUGCAUAGCCCGUACGGCA

Randomly chosen initial structure



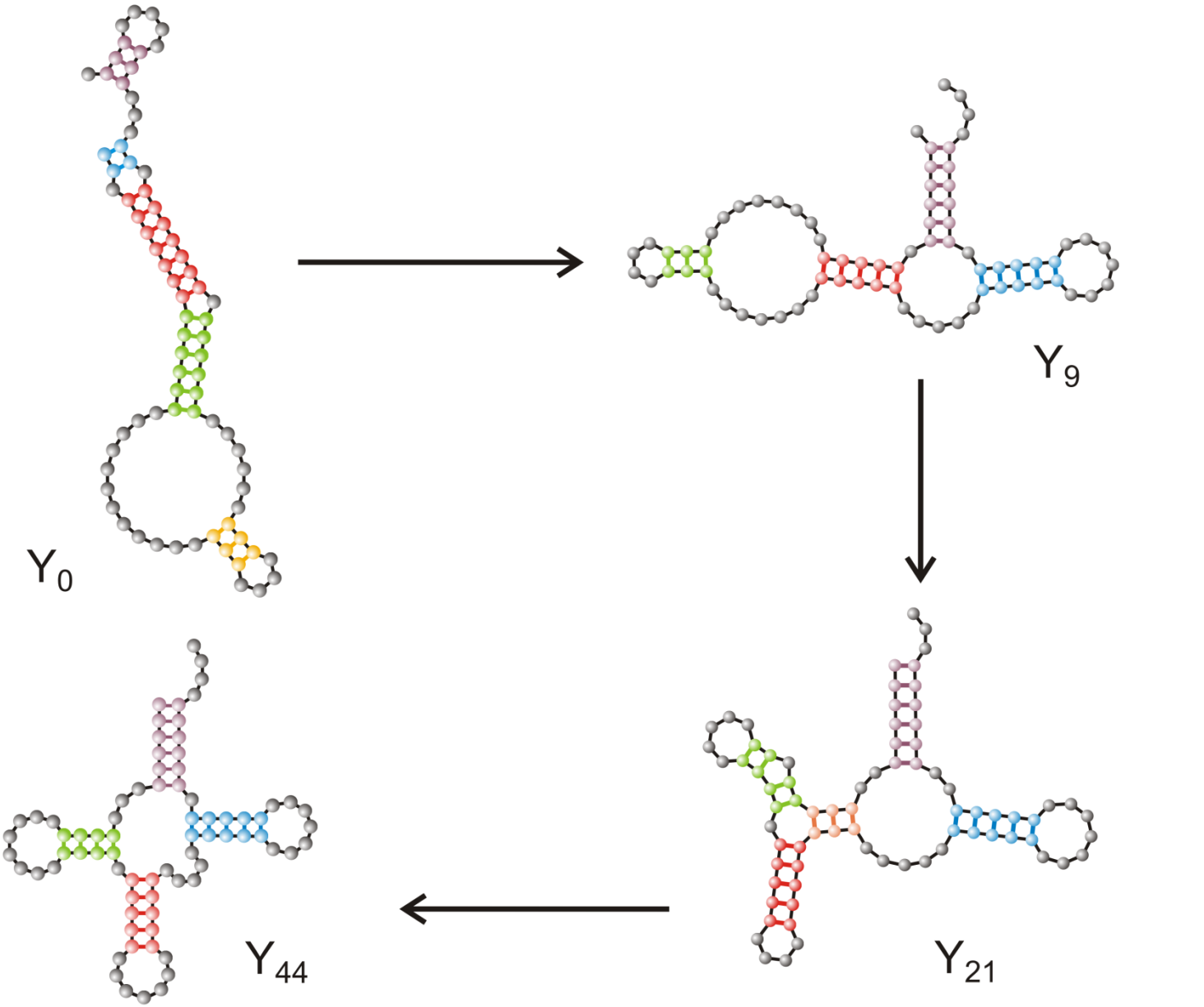
S_0: GUUAUGGGCGAUGAGGAGUAGAGAAAAAACCAAUCGGUGAAAGAUCUCGUGUGCCCAUUGCAUAGCCCGUACGGCA

Randomly chosen initial structure



S_0: GUUAUGGGCGAUGAGGAGUAGAGAAAAAACCAAUCGGUGAAAGAUCUCGUGUGCCCAUUGCAUAGCCCGUACGGCA

Randomly chosen initial structure



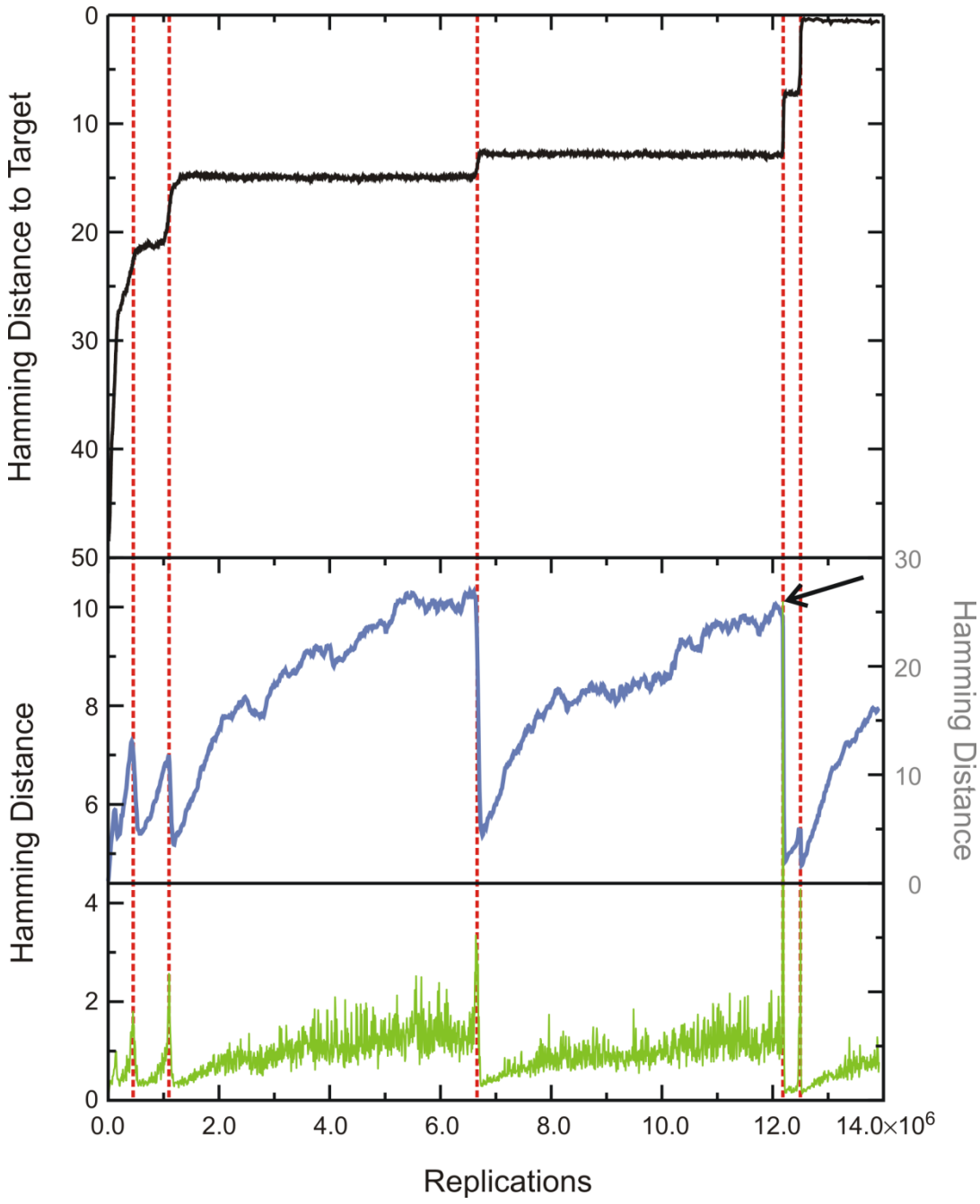
RNA^{phe} as target structure

S_{44}: GGGCAGAUAGGGCGUGUGAUAGCCCAUAGCGAACCCCCCGCUGAGCUUGUGCGACGUAUGCGCACCCUGUCCCGCU

Evolutionary trajectory

Spreading of the population on neutral networks

Drift of the population center in sequence space



Richard Lenski, 1956 -

Bacterial evolution under controlled conditions: A twenty years experiment.

Richard Lenski, University of Michigan, East Lansing

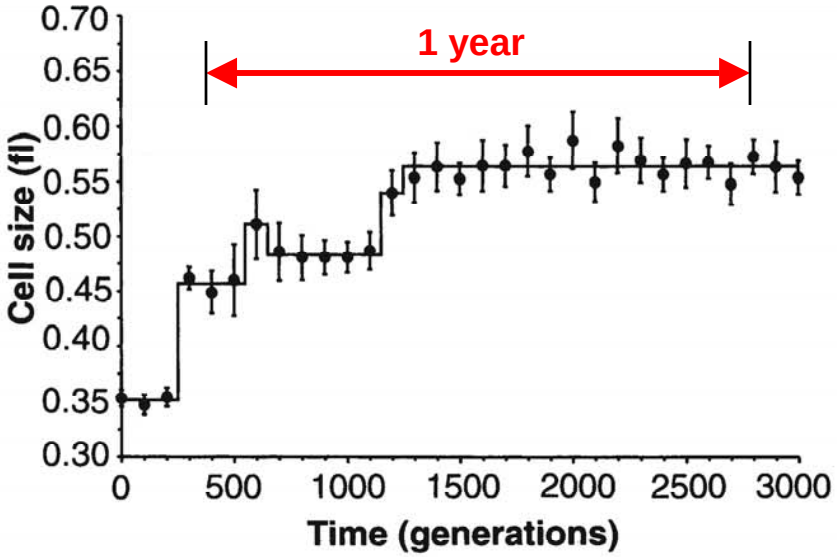


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

Epochal evolution of bacteria in serial transfer experiments under constant conditions

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

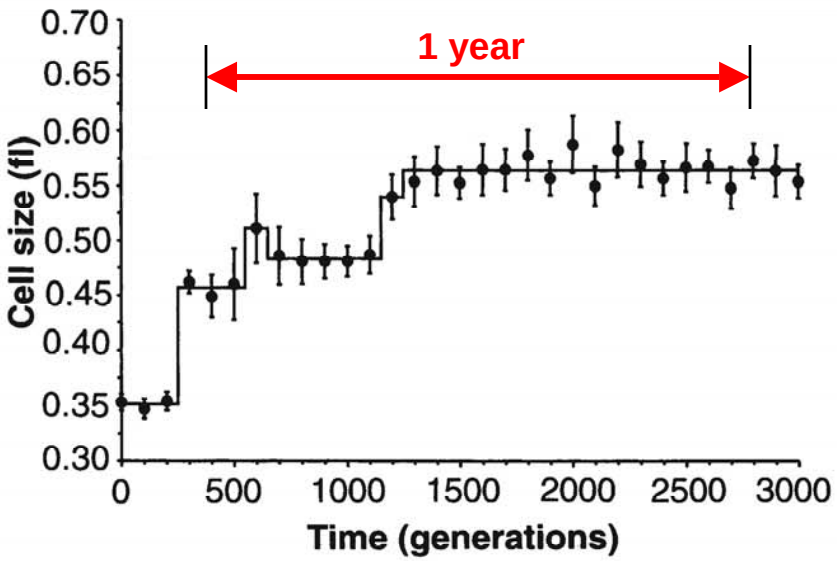


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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

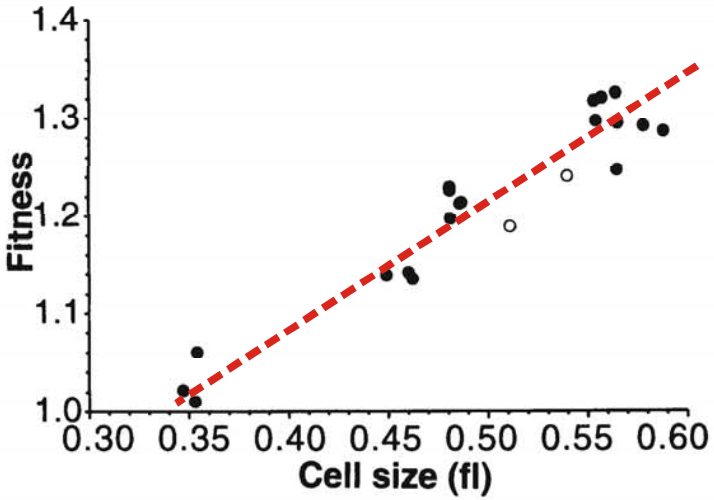


Fig. 2. Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (*6*, *7*). The open symbols indicate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

Epochal evolution of bacteria in serial transfer experiments under constant conditions

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

The twelve populations of Richard Lenski‘s long time evolution experiment
Enhanced turbidity in population A-3

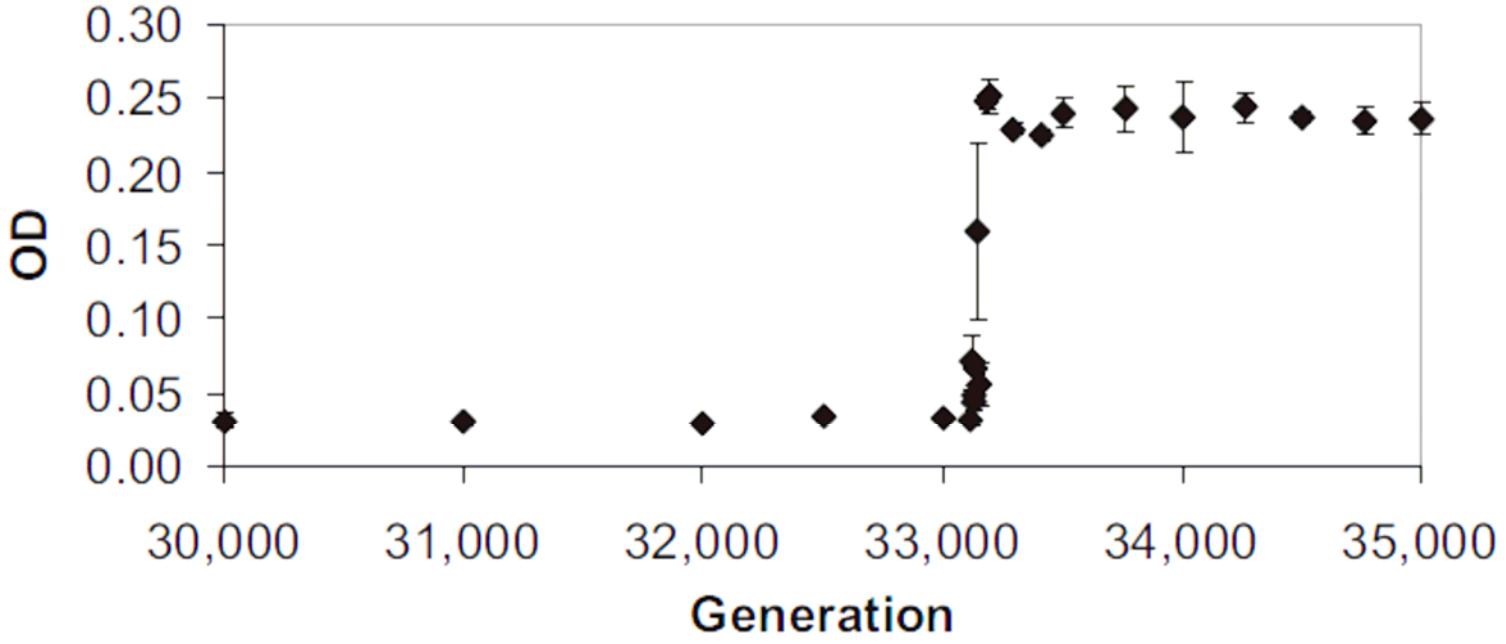


Fig. 1. Population expansion during evolution of the Cit$^+$ phenotype. Samples frozen at various times in the history of population Ara-3 were revived, and three DM25 cultures were established for each generation. Optical density (OD) at 420 nm was measured for each culture at 24 h. Error bars show the range of three values measured for each generation.

Innovation by mutation in long time evolution of Escherichia coli in constant environment

Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008. Proc.Natl.Acad.Sci.USA 105:7899-7906

Table 1. Summary of replay experiments

	First experiment		Second experiment		Third experiment	
Generation	Replicates	Independent Cit$^+$ mutants	Replicates	Independent Cit$^+$ mutants	Replicates	Independent Cit$^+$ mutants
Ancestor	6	0	10	0	200	0
5,000	—	—	—	—	200	0
10,000	6	0	30	0	200	0
15,000	—	—	—	—	200	0
20,000	6	0	30	0	200	2
25,000	6	0	30	0	200	0
27,000	—	—	—	—	200	2
27,500	6	0	30	0	—	—
28,000	—	—	—	—	200	0
29,000	6	0	30	0	200	0
30,000	6	0	30	0	200	0
30,500	6	1	30	0	—	—
31,000	6	0	30	0	200	1
31,500	6	1	30	0	200	1
32,000	6	0	30	4	200	2
32,500	6	2	30	1	200	0
Totals	72	4	340	5	2,800	8

Contingency of E. coli evolution experiments

1. Adaptation in biology
2. Cycles of evolution
3. Molecules – from sequence to function
4. Mutation and structure
5. Spaces and mappings
6. Evolutionary dynamics on landscapes
7. Neutrality
8. Stochasticity, contingency, and history
9. **Perspectives**

(i) Fitness landscapes for the evolution of molecules are obtainable by standard techniques of physics and chemistry.

(ii) Fitness landscapes for evolution of viroids and viruses under controlled conditions are accessible in principle.

(iii) Systems biology can be carried out for especially small bacteria and an extension to bacteria of normal size is to be expected for the near future.

(iv) The computational approach for selection on known fitness landscapes – ODEs or stochastic processes – is standard.

(v) The efficient description of migration and splitting of populations in sequence space requires new mathematical techniques.

Consideration of multistep and nonlinear replication mechanisms as well as accounting for epigenetic phenomena is readily possible within the molecular approach.

Coworkers

Universität Wien

Walter Fontana, Harvard Medical School, MA

Matin Nowak, Harvard University, MA

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

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

Sebastian Bonhoeffer, ETH Zürich, CH

Christian Reidys, University of Southern Denmark, Odense, DK

Christian Forst, University of Texas Southwestern Medical Center, TX

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

Jan Cupal, **Ulrike Langhammer**, **Ulrike Mückstein, Jörg Swetina,**
Universität Wien, AT

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

Acknowledgement of support

Universität Wien

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

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

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

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

Austrian Genome Research Program – GEN-AU

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

Thank you for your attention !

Web-Page for further information:

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