Evolution at the Molecular Level

150 Years after Darwin‘s Origin of Species

Peter Schuster

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

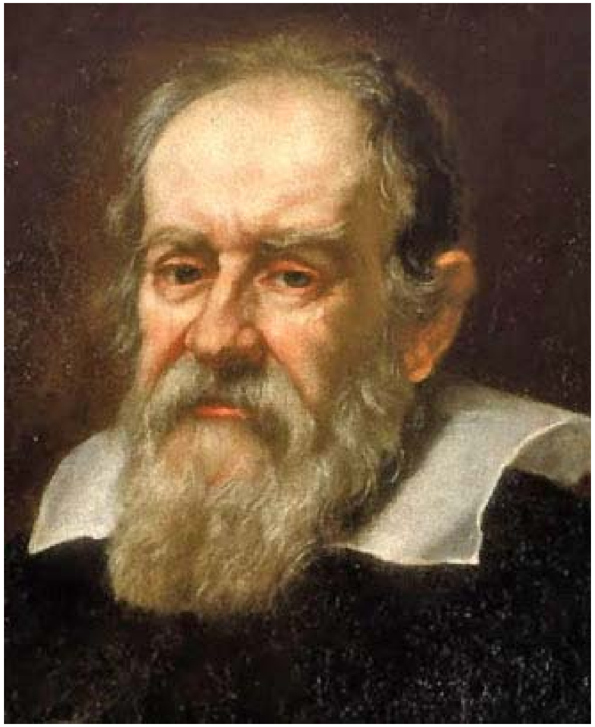
Systems Chemistry II: Evolution and Systems

Balatonfüred, 18.– 23.10.2009

Web-Page for further information:

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

"La Filosophia è scritta in questo grandissimo libro, que continuamente ci stà aperto innanzi à gli occhi (io dico l’universo) ma non si può intendere se prima non s’impara à intender la lingua, e conoscer i caratteri, nei quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi. & altre figure Geometriche ...",



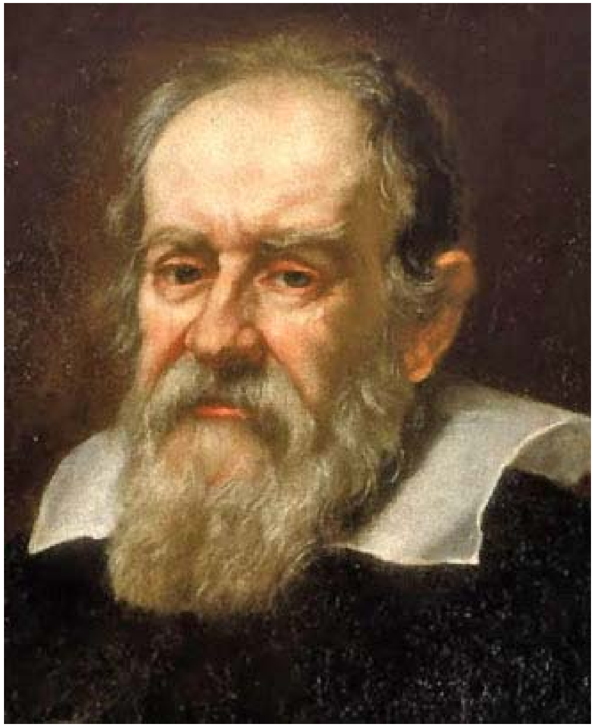
Galileo Galilei, 1564 - 1642

Galileo Galilei. 1632. ***Il Saggiatore***.
Edition Nationale, Bd.6, Florenz 1896, p.232.

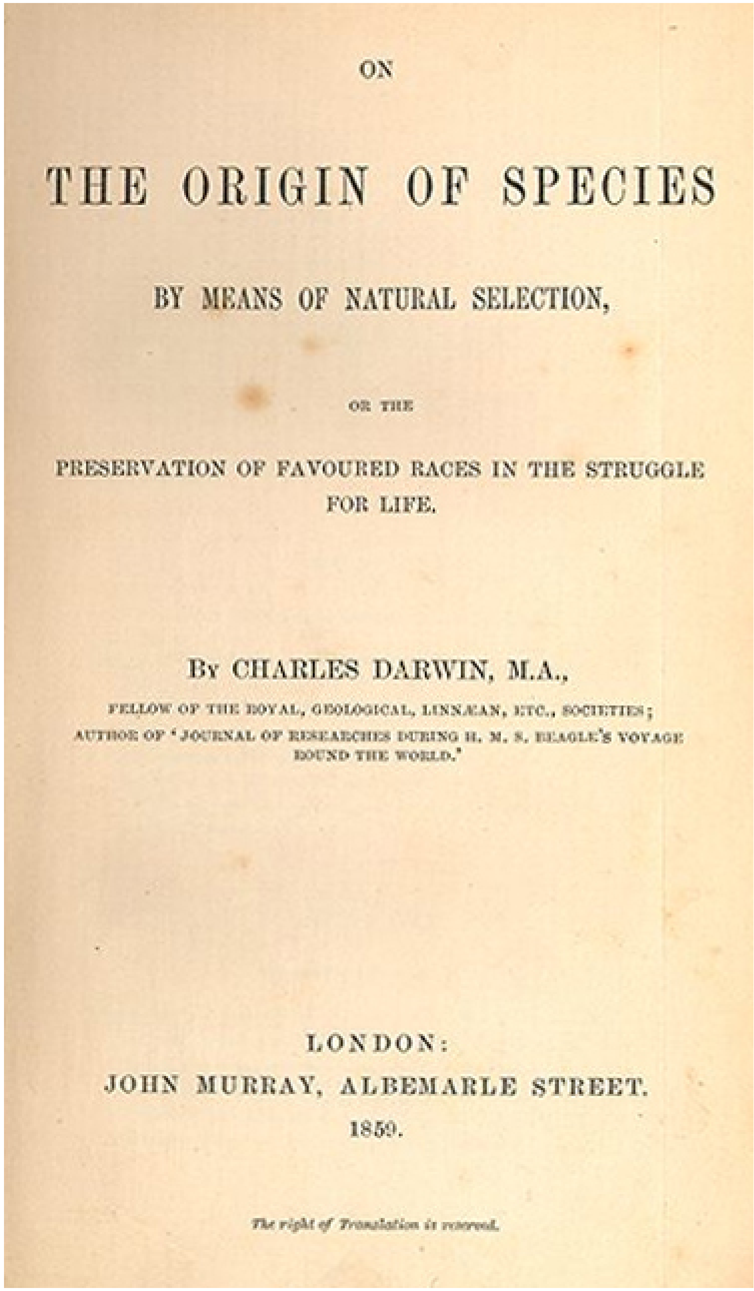
"La Filosophia è scritta in questo grandissimo libro, que continuamente ci stà aperto innanzi à gli occhi (io dico l'universo) ma non si può intendere se prima non s'impara à intender la lingua, e conoscer i caratteri, nei quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi. & altre figure Geometriche ...",

„Philosophy [science] is written in this grand book, the universe ... . It is written in the language of mathematics, and ist characters are triangles, circles and other geometric figures; …. „

Galileo Galilei. 1632. ***Il Saggiatore***.
Edition Nationale, Bd.6, Florenz 1896, p.232.



Galileo Galilei, 1564 - 1642



ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE
ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

If Charles Darwin would have written the „Origin“ in mathematical language, how would he have done it?

Molecular replicators – What do they need and how do they work?

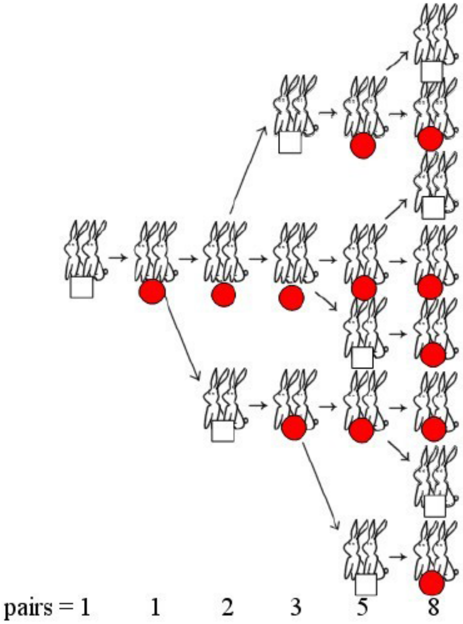
Can we have selection and coexistence simultaneously? Where does neutrality come from and what is it good for in evolution?

If Charles Darwin would have written the „Origin“ in mathematical language, how would he have done it?

Molecular replicators – What do they need and how do they work?

Can we have selection and coexistence simultaneously? Where does neutrality come from and what is it good for in evolution?

$$F_{n+1} = F_n + F_{n-1};\ F_0 = 0,\ F_1 = 1$$



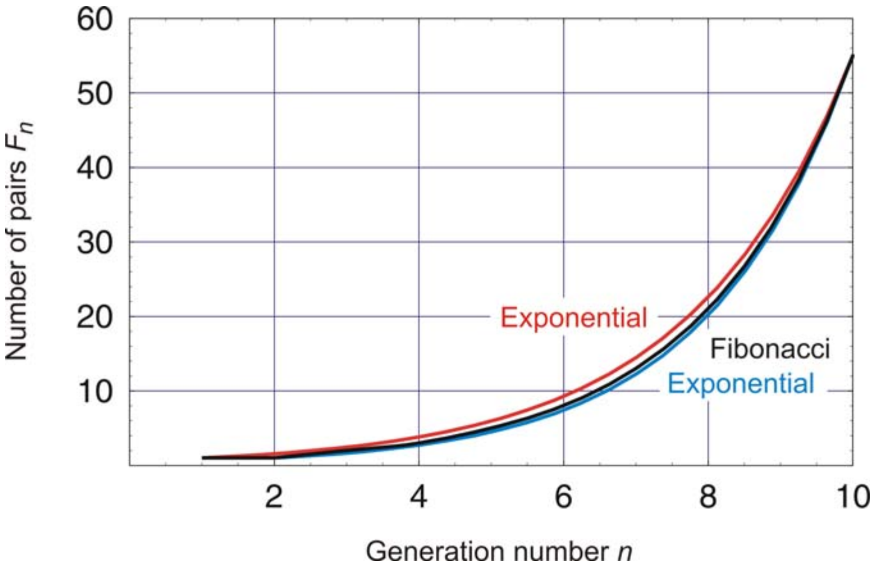
Thomas Robert Malthus
1766 – 1834

Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

1, 2 , 4 , 8 ,16 , 32 , 64, 128 , ...

geometric progression

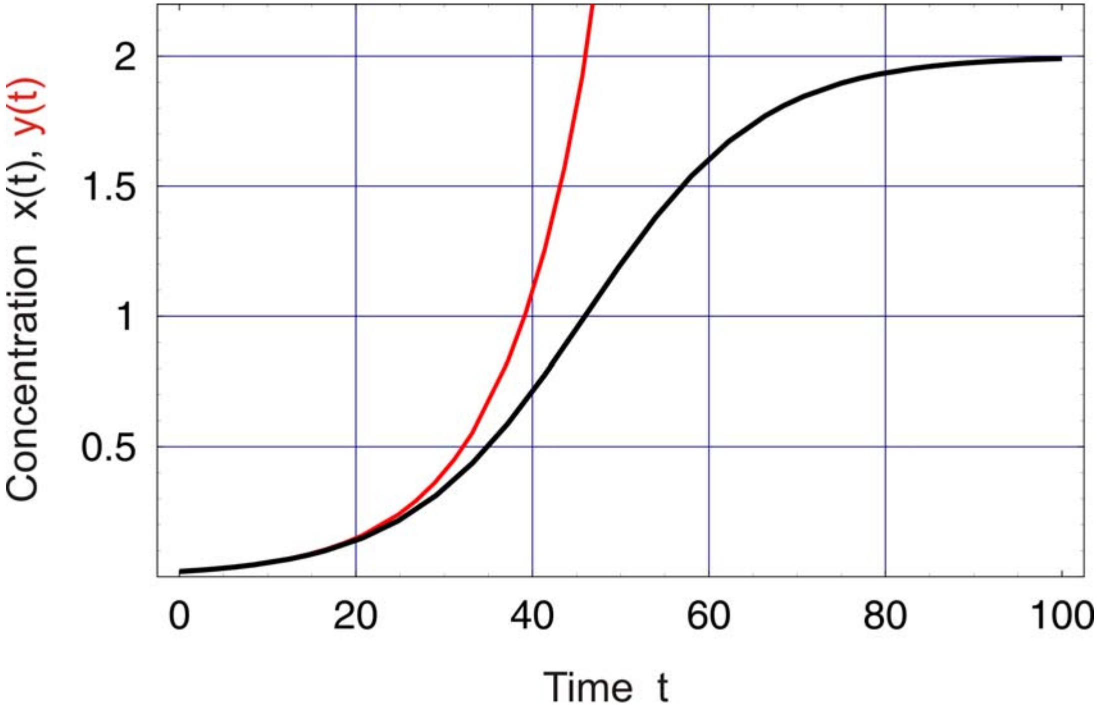
exponential growth



The history of exponential growth

Pierre-François Verhulst,
1804-1849

$$\frac{dx}{dt} = r\,x\left(1-\frac{x}{C}\right), \quad x(t) = \frac{x(0)\,C}{x(0)+\left(C-x(0)\right)e^{-rt}}$$



The logistic equation, 1828

$$\frac{\mathrm{d}x}{\mathrm{d}t} = r\,x\left(1 - \frac{x}{C}\right) \quad \Rightarrow \quad \frac{\mathrm{d}x}{\mathrm{d}t} = r\,x - \frac{x}{C}\,r\,x$$

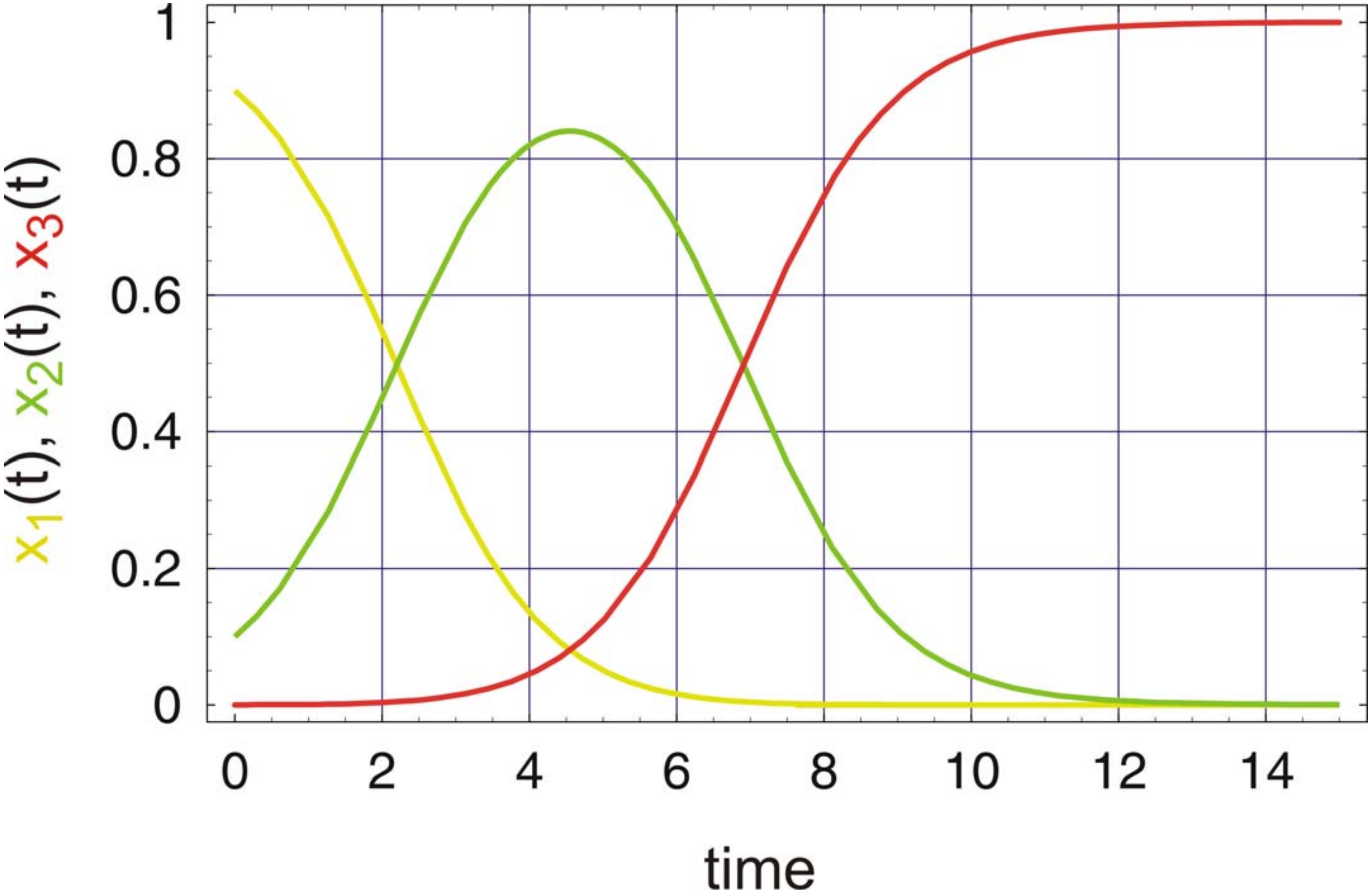
$$r\,x \equiv \Phi(\mathrm{t}),\; C = 1: \quad \frac{\mathrm{d}x}{\mathrm{d}t} = x\left(r - \Phi\right)$$

$$\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n: \quad \left[\mathrm{X}_i\right] = x_i\,; \quad \textstyle\sum_{i=1}^{n} x_i = C = 1$$

$$\boxed{\frac{\mathrm{d}x_j}{\mathrm{d}t} = x_j\left(f_j - \textstyle\sum_{i=1}^{n} f_i\,x_i\right) = x_j\left(f_j - \Phi\right)}\,; \quad \Phi = \textstyle\sum_{i=1}^{n} f_i\,x_i$$

Darwin

Generalization of the logistic equation to *n* variables yields selection



$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = x_j \left(f_j - \sum_{i=1}^{n} f_i \, x_i \right) = x_j \left(f_j - \Phi \right); \quad f_1 = 1,\ f_2 = 2,\ f_3 = 3$$

Ronald Fisher (1890-1962)

alleles: $A_1, A_2, \ldots\ldots, A_n$

frequencies: $x_i = [A_i]$; genotypes: $A_i \cdot A_j$

fitness values: $a_{ij} = f(A_i \cdot A_j),\ a_{ij} = a_{ji}$

Mendel

Darwin

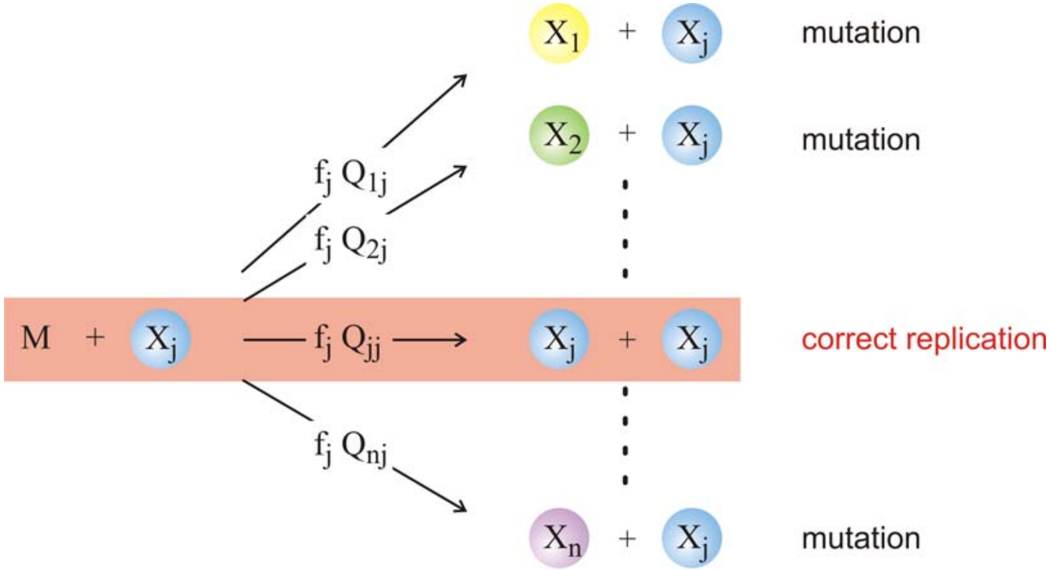
$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \sum_{i=1}^{n} a_{ji}\, x_i\, x_j - \Phi\, x_j = \boxed{x_j \left(\sum_{i=1}^{n} a_{ji}\, x_i - \Phi \right)},\quad j = 1, 2, \ldots, n$$

$$\text{mit}\quad \Phi(\mathrm{t}) = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji}\, x_i\, x_j \quad \text{und} \quad \sum_{j=1}^{n} x_j = 1$$

$$\boxed{\frac{\mathrm{d}\Phi}{\mathrm{d}t} = 2\left(< \bar{a}^2 > - < \bar{a} >^2 \right) = 2\ \mathrm{var}\left\{ \bar{a} \right\} \geq 0}$$

Ronald Fisher‘s selection equation: The genetical theory of natural selection.
Oxford, UK, Clarendon Press, 1930.

Manfred Eigen
1927 -



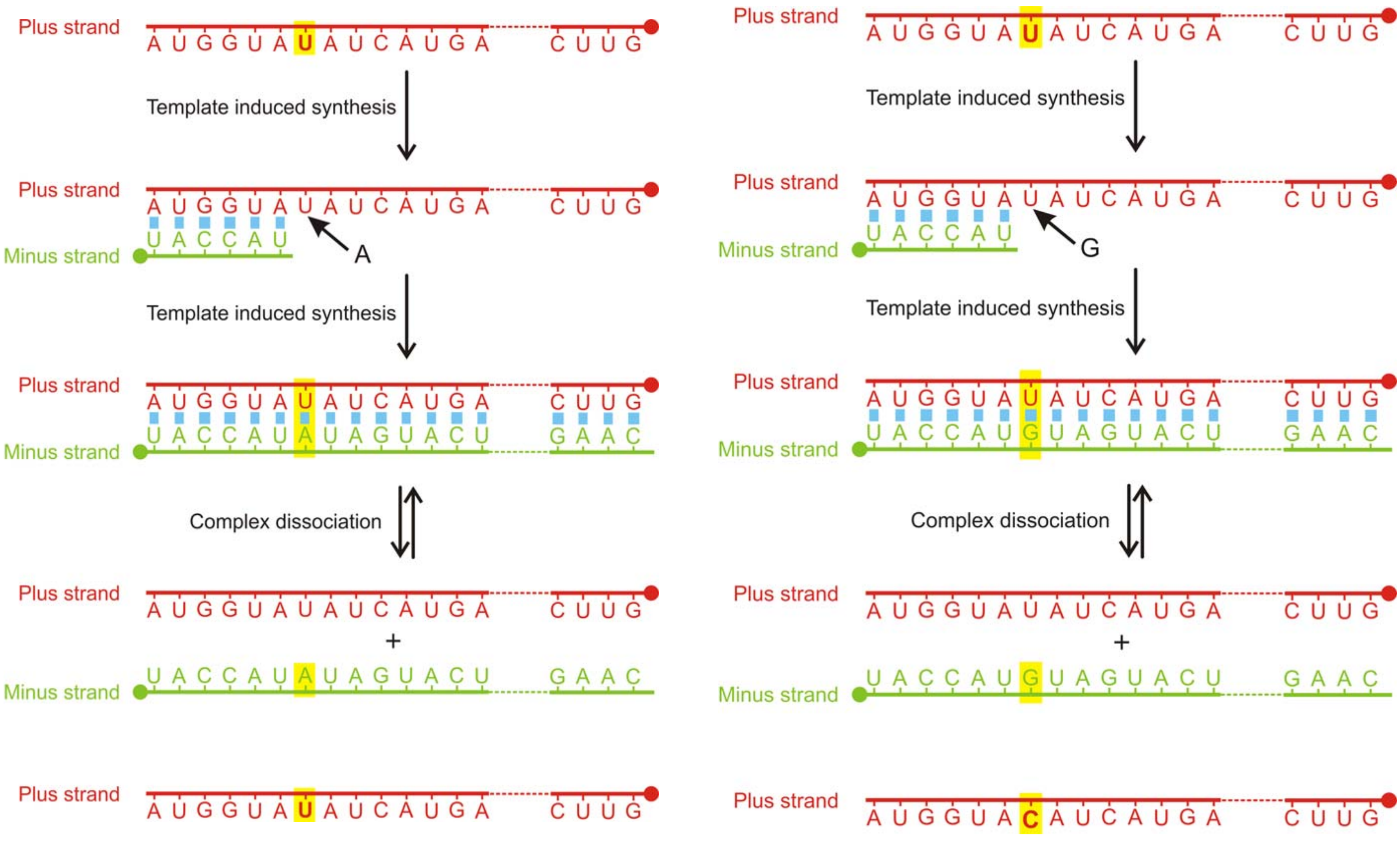
f_j ... replication rate function or fitness function

Q_{ij} ... mutation frequency: $X_j \rightarrow X_i$

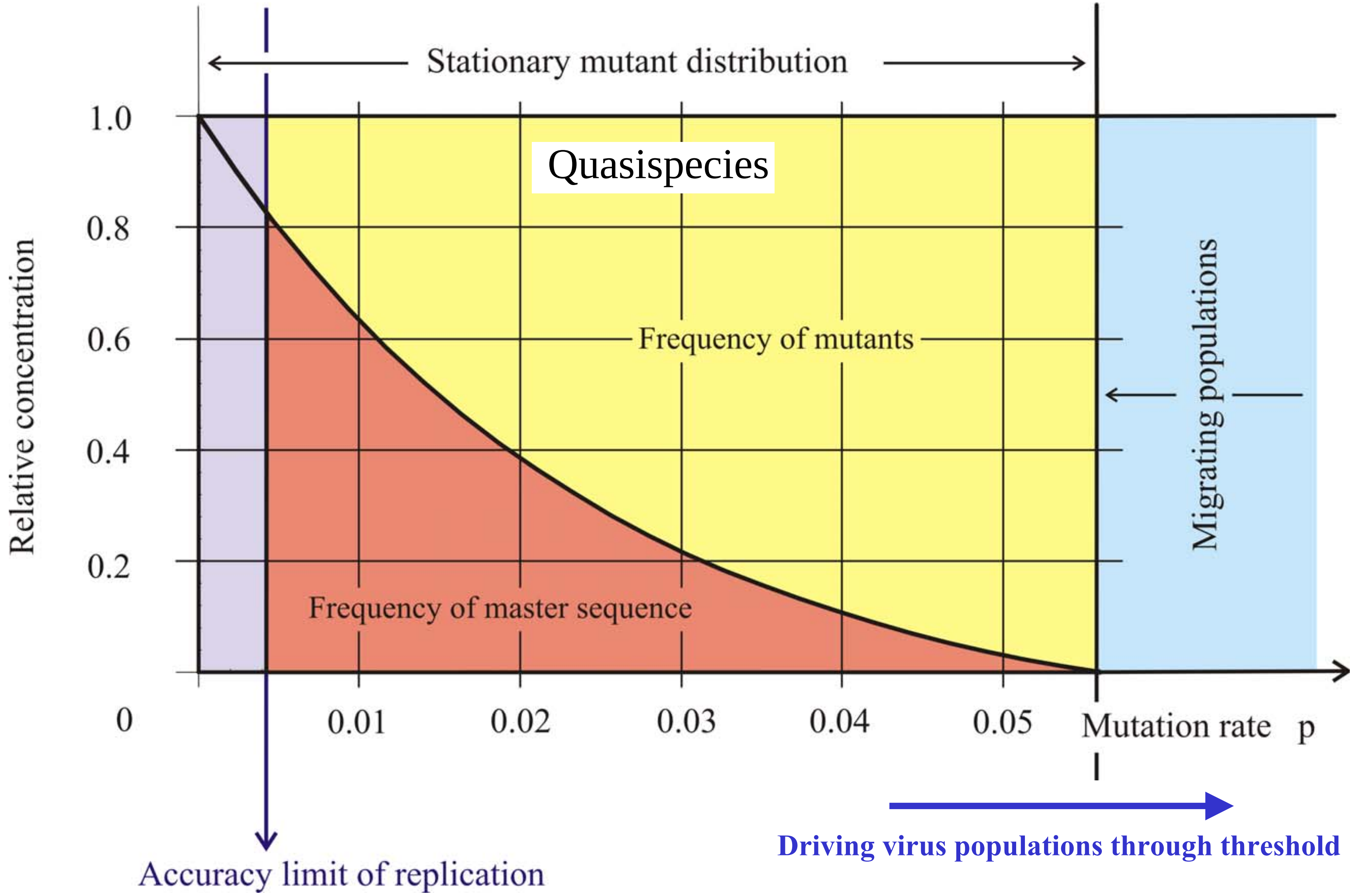
$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \sum_{i=1}^{n} Q_{ji}\, f_i\, x_i - x_j\, \Phi \,; \quad j = 1, 2, \ldots, n$$

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



Replication and mutation are parallel chemical reactions.



The error threshold in replication

Chain length and error threshold

$$Q \cdot \sigma \;=\; (1-p)^n \cdot \sigma \;\geq\; 1 \;\Rightarrow\; n \cdot \ln(1-p) \geq -\ln\sigma$$

$$p \;\ldots\; \text{constant}: \quad n_{\max} \approx \frac{\ln\sigma}{p}$$

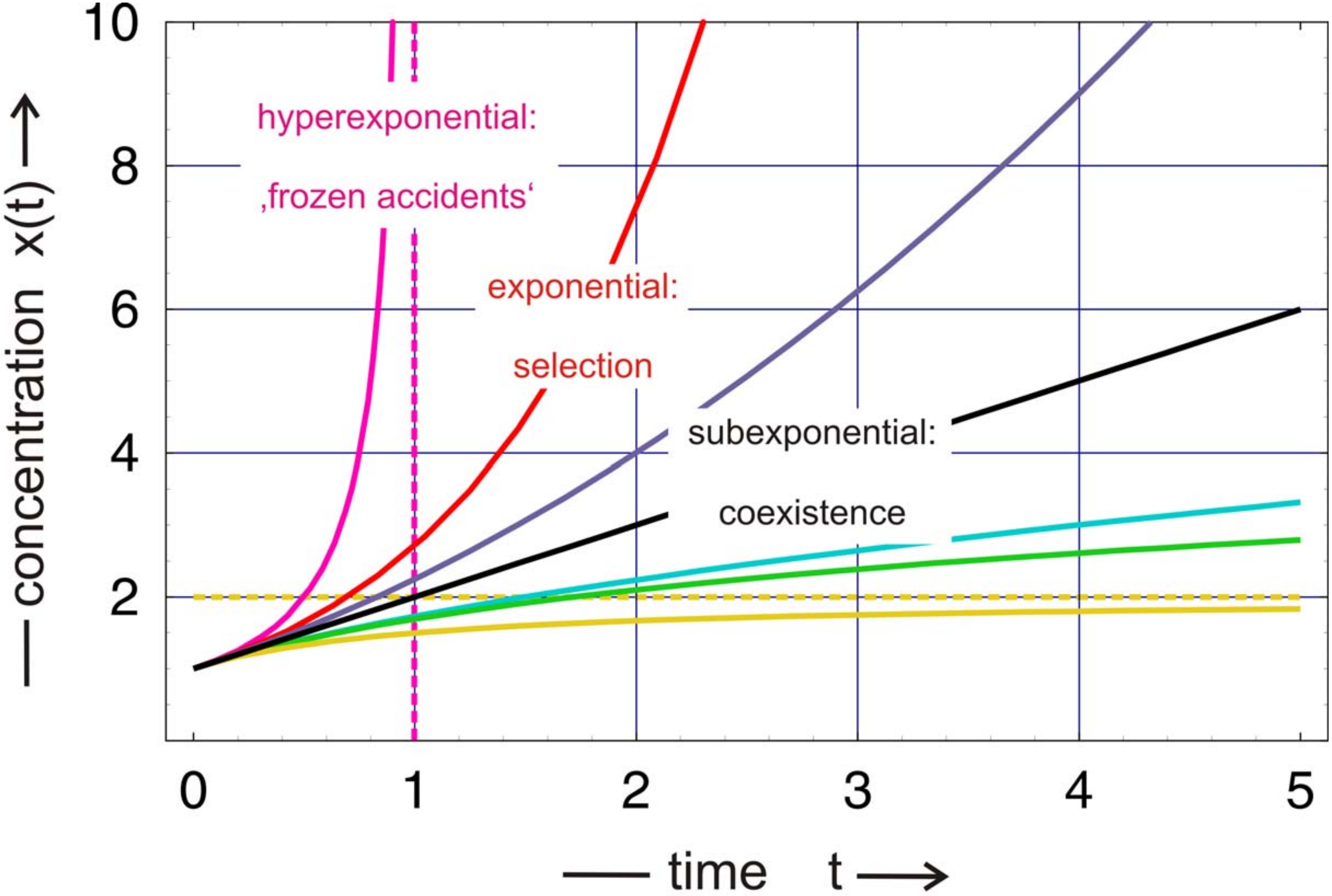
$$n \;\ldots\; \text{constant}: \quad p_{\max} \approx \frac{\ln\sigma}{n}$$

$Q=(1-p)^n$ … replication accuracy

p … error rate

n … chain length

$\sigma = \dfrac{f_m}{\sum_{j\neq m} f_j}$ … superiority of master sequence



The unique feature of exponential growth

Exponential growth and limited ressources give rise to selection.

There is no known working example of effective selection without exponential growth.

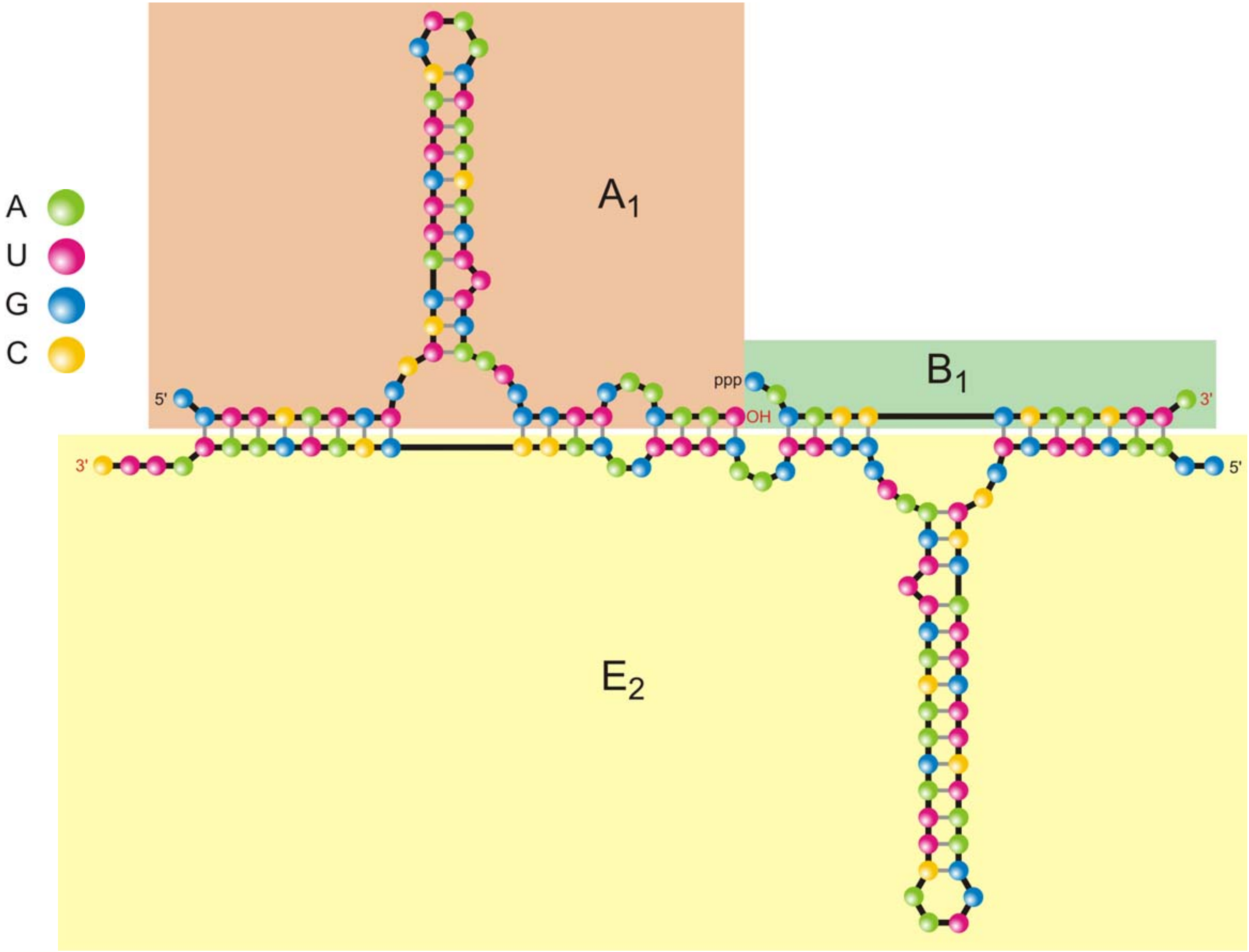
Copying digital (genetic) information gives rise to mutation.

There is no known working example of effective mutation without digital information.

If Charles Darwin would have written the „Origin“ in mathematical language, how would he have done it?

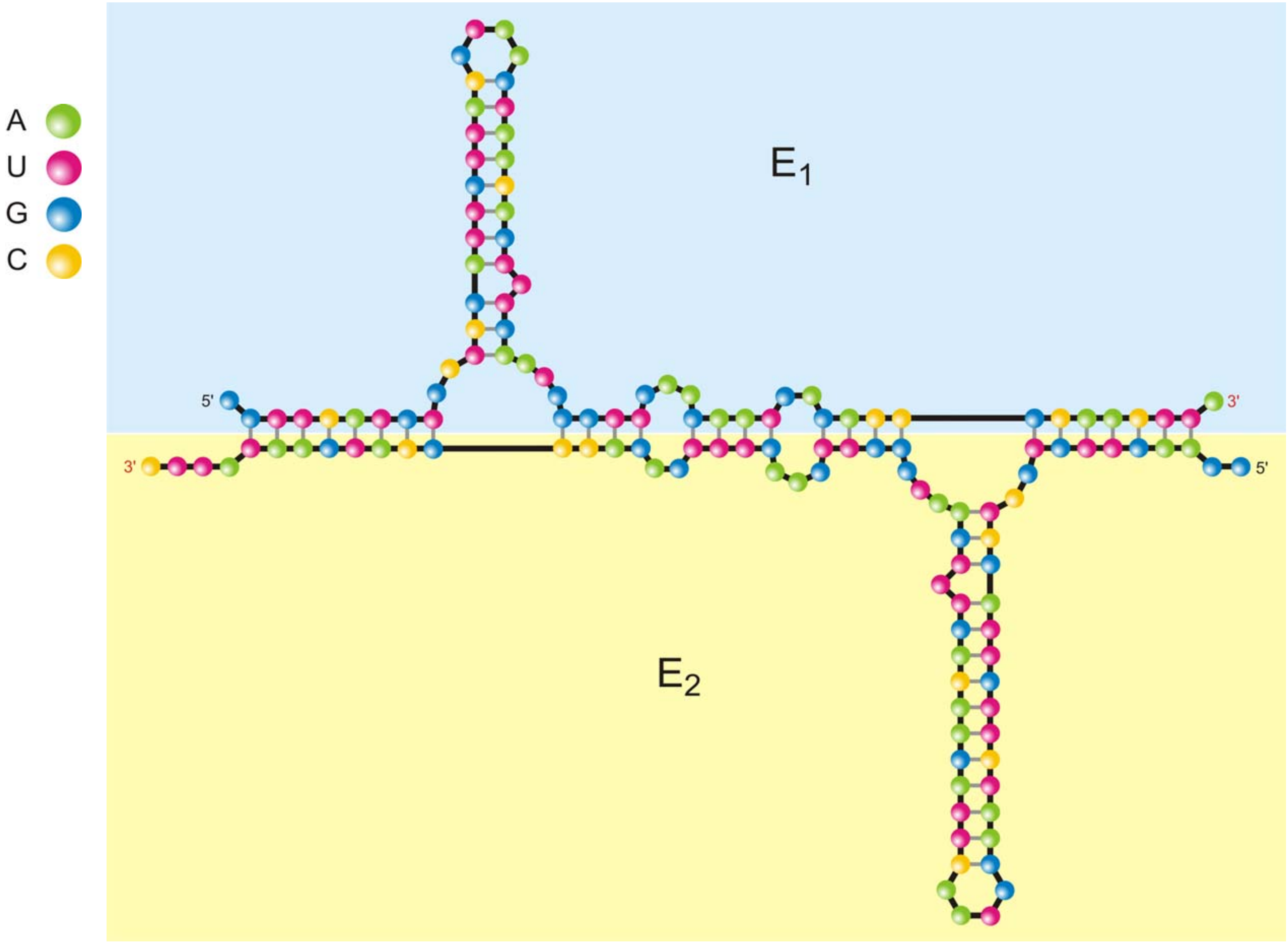
Molecular replicators – What do they need and how do they work?

Can we have selection and coexistence simultaneously? Where does neutrality come from and what is it good for in evolution?



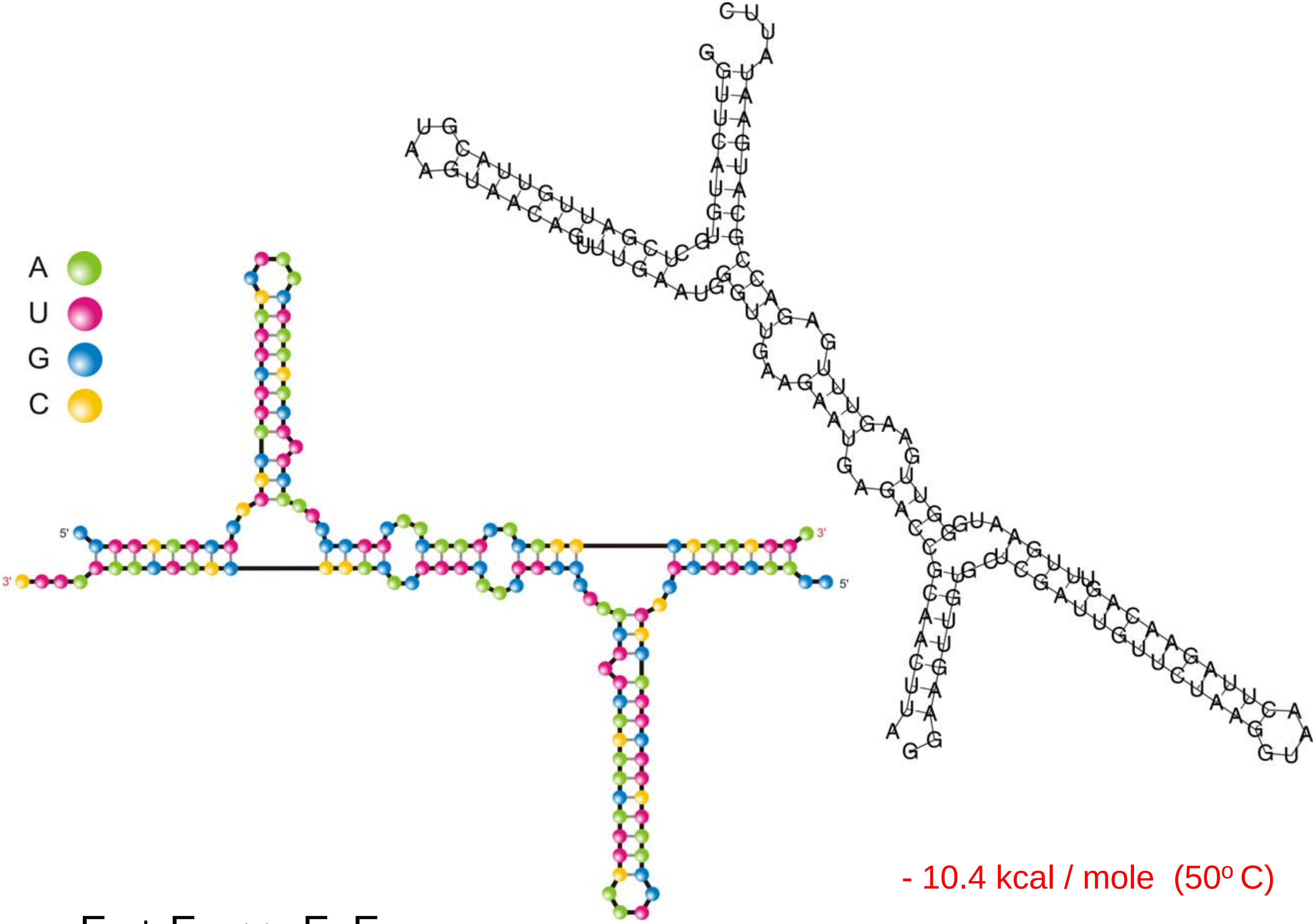
An example of two ribozymes growing exponentially by cross-catalysis.

T.A. Lincoln, G.F. Joyce. 2009. Self-sustained replication of an RNA enzyme. Science 323:1229-1232



An example of two ribozymes growing exponentially by cross-catalysis.

T.A. Lincoln, G.F. Joyce. 2009. Self-sustained replication of an RNA enzyme. Science 323:1229-1232



$E_1 + E_2 \leftrightarrow E_1 \cdot E_2$

Vienna RNA Package Version 1.8.2

- 10.4 kcal / mole (50^{o} C)

ΔG = - 12.7 kcal / mole (37^{o} C)

- 15.4 kcal / mole (20^{o} C)

T = 37° C; concentrations in mole / l

Initial concentrations		Relative equilibrium concentrations		
E_1	**E_2**	**$E_1.E_2$**	**E_1**	**E_2**
1×10^{-6}	1×10^{-6}	0.4832	0.0168	0.0168
1×10^{-7}	1×10^{-7}	0.4489	0.0511	0.0511
1×10^{-8}	1×10^{-8}	0.3561	0.1439	0.1439
1×10^{-9}	1×10^{-9}	0.1781	0.3219	0.3219
1×10^{-10}	1×10^{-10}	0.0369	0.4631	0.4631

T = 20° C; concentrations in mole / l

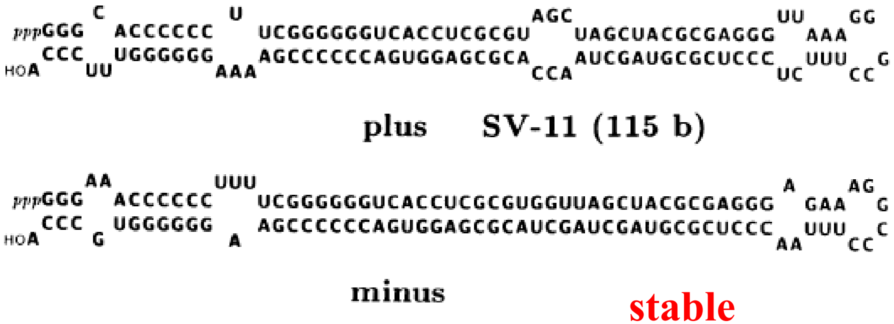
Initial concentrations		Relative equilibrium concentrations		
E_1	**E_2**	**$E_1.E_2$**	**E_1**	**E_2**
1×10^{-6}	1×10^{-6}	0.4991	0.0009	0.0009
1×10^{-7}	1×10^{-7}	0.4971	0.0029	0.0029
1×10^{-8}	1×10^{-8}	0.4908	0.0092	0.0092
1×10^{-9}	1×10^{-9}	0.4715	0.0285	0.0285
1×10^{-10}	1×10^{-10}	0.4153	0.0847	0.0847

T = 50° C; concentrations in mole / l

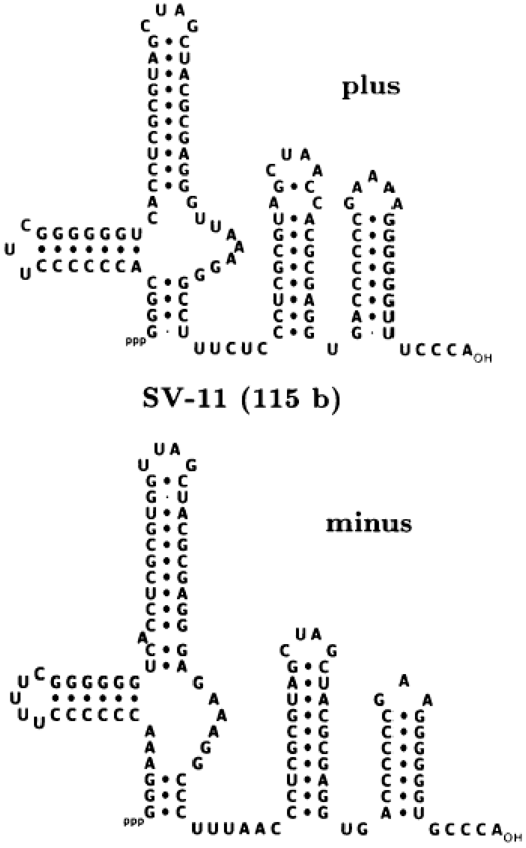
Initial concentrations		Relative equilibrium concentrations		
E_1	**E_2**	**$E_1.E_2$**	**E_1**	**E_2**
1×10^{-6}	1×10^{-6}	0.3655	0.1345	0.1341
1×10^{-7}	1×10^{-7}	0.1920	0.3080	0.3076
1×10^{-8}	1×10^{-8}	0.0424	0.4576	0.4575
1×10^{-9}	1×10^{-9}	0.0050	0.4950	0.4950
1×10^{-10}	1×10^{-10}	0.0005	0.4995	0.4995

Calculated concentrations of ribozyme monomers and dimers

Christof K. Biebricher 1941-2009



stable



metastable

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

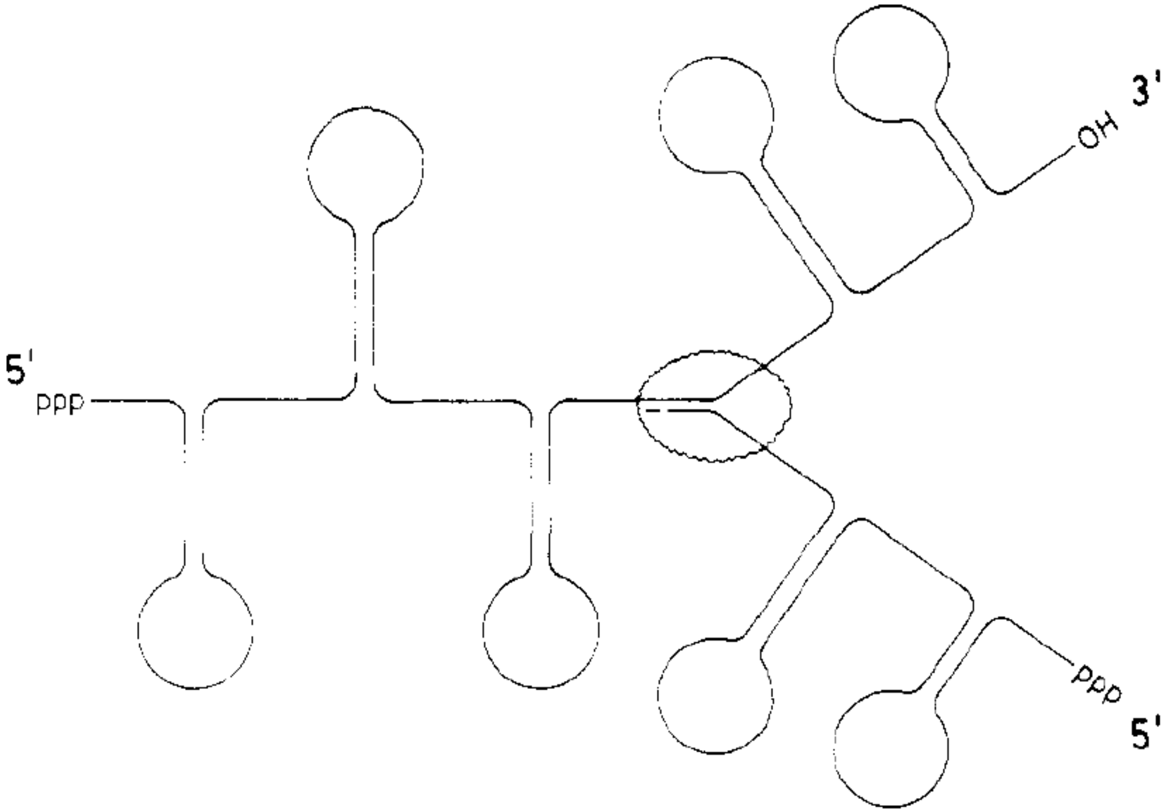
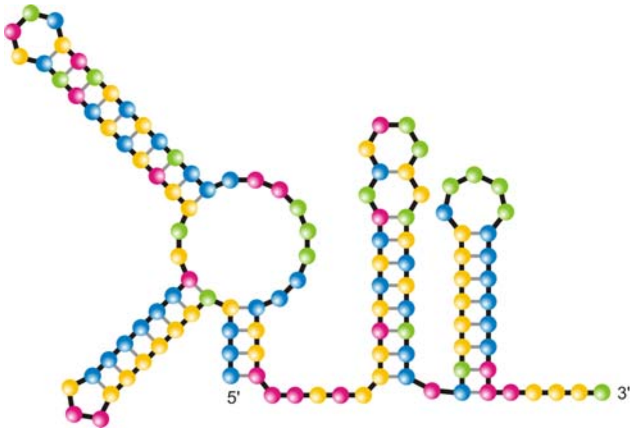


FIGURE 2: Elongation mechanism, adapted from models of Weissmann (1974) and Mills et al. (1978). At the replication fork, new substrate is added by base pairing with the template and pyrophosphate release. The short double helix at the replication fork is opened by the enzyme as replication continues.

A sketch of complementary replication by Qβ replicase



ΔG_{fold} = - 68.5 kcal / mole

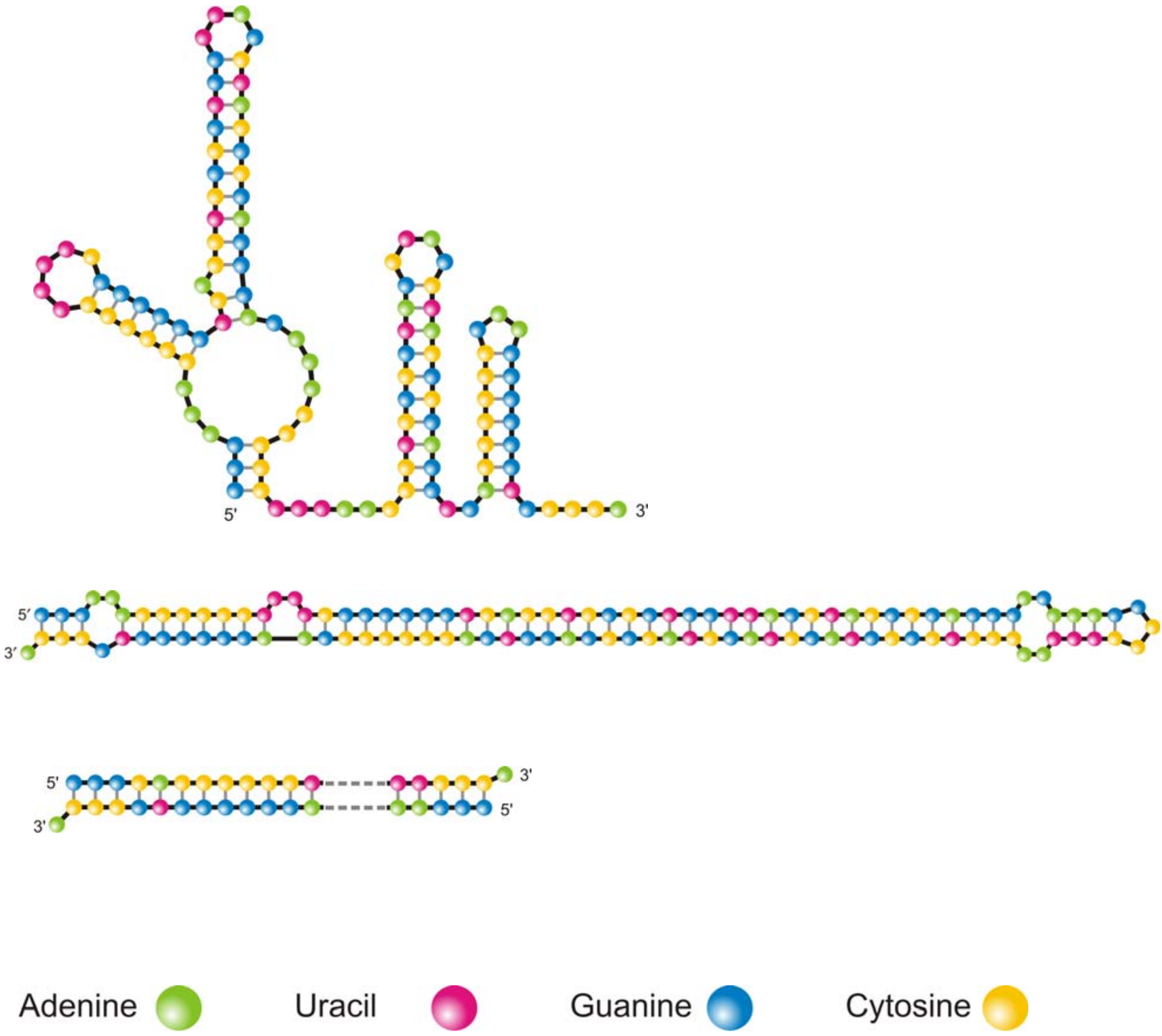
ΔG_{fold} = - 98.4 kcal / mole

ΔG_{fold} = - 277.4 kcal / mole

ΔG_{bind} = - 72.1 kcal / mole

Adenine 🟢 Uracil 🟣 Guanine 🔵 Cytosine 🟡

SV11 plus strand



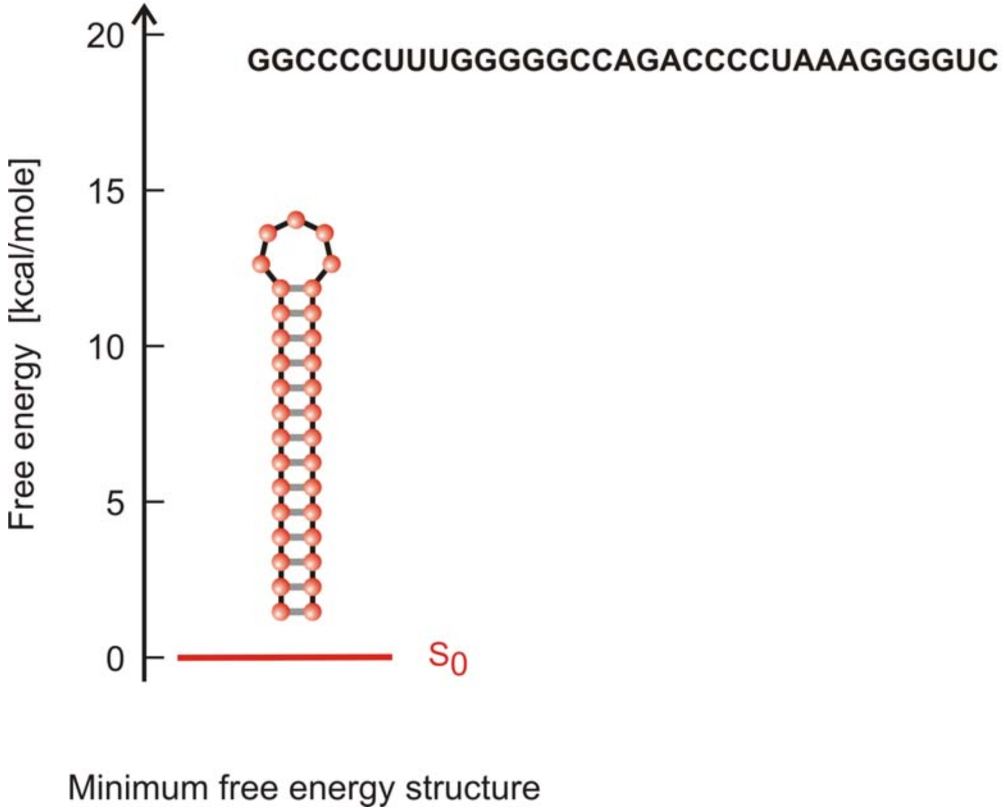
ΔG_{fold} = - 71.1 kcal / mole

ΔG_{fold} = - 101.9 kcal / mole

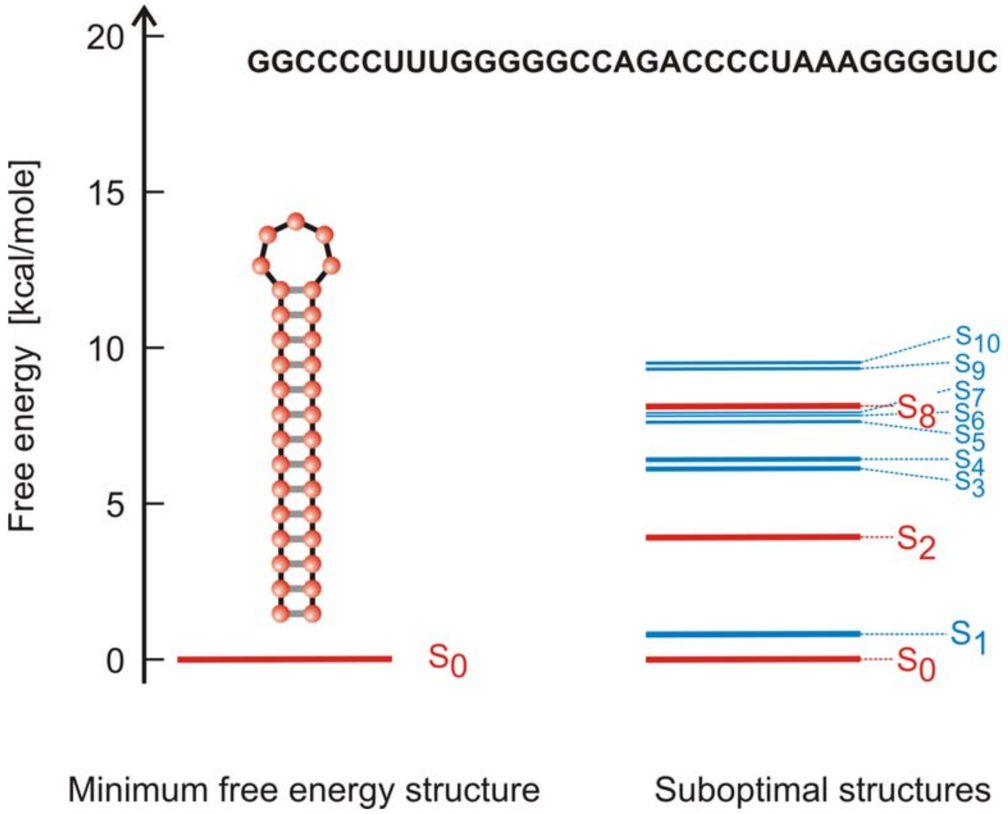
ΔG_{fold} = - 277.4 kcal / mole

ΔG_{bind} = - 72.1 kcal / mole

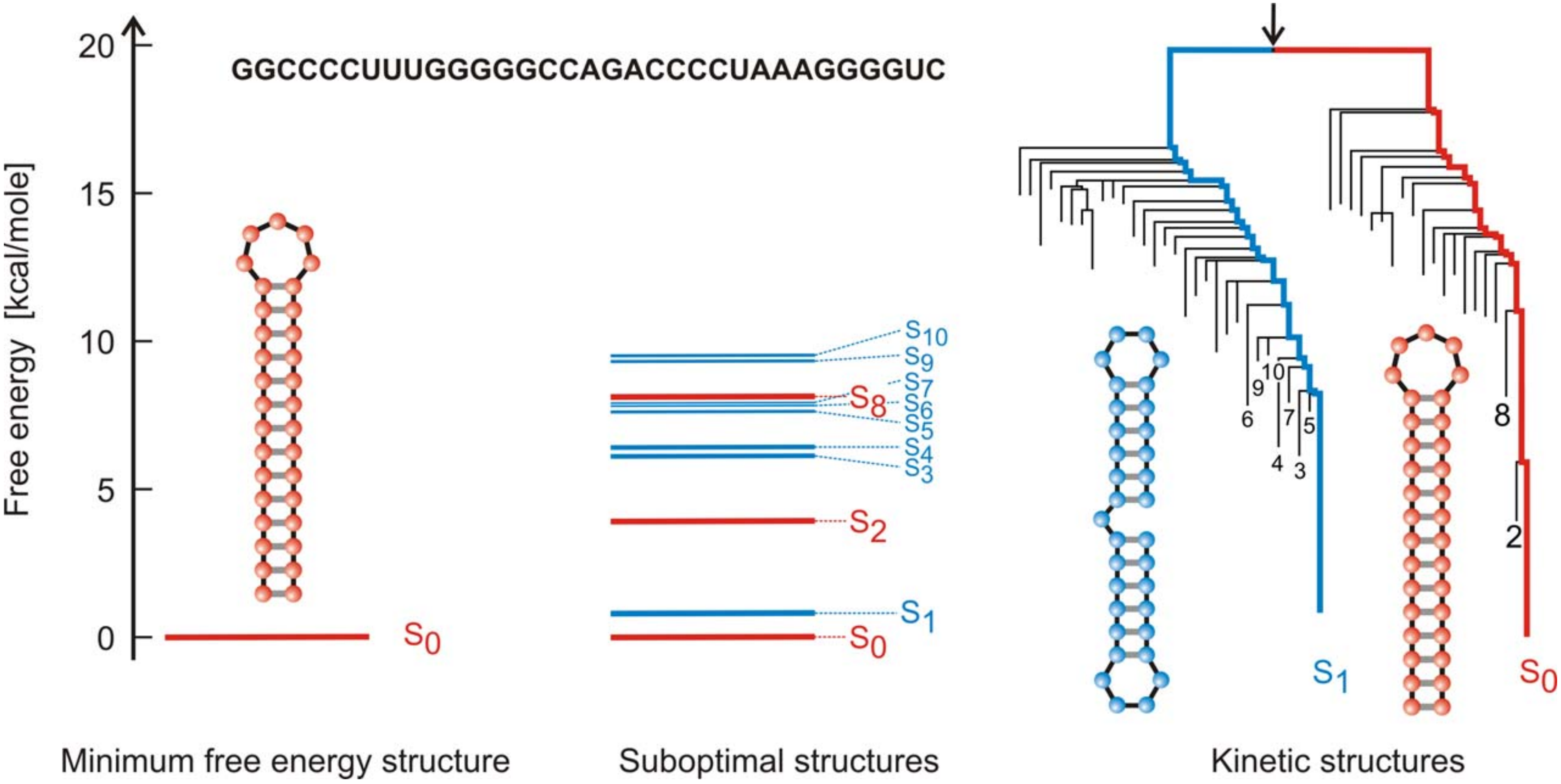
SV11 minus strand



Extension of the notion of molecular structure

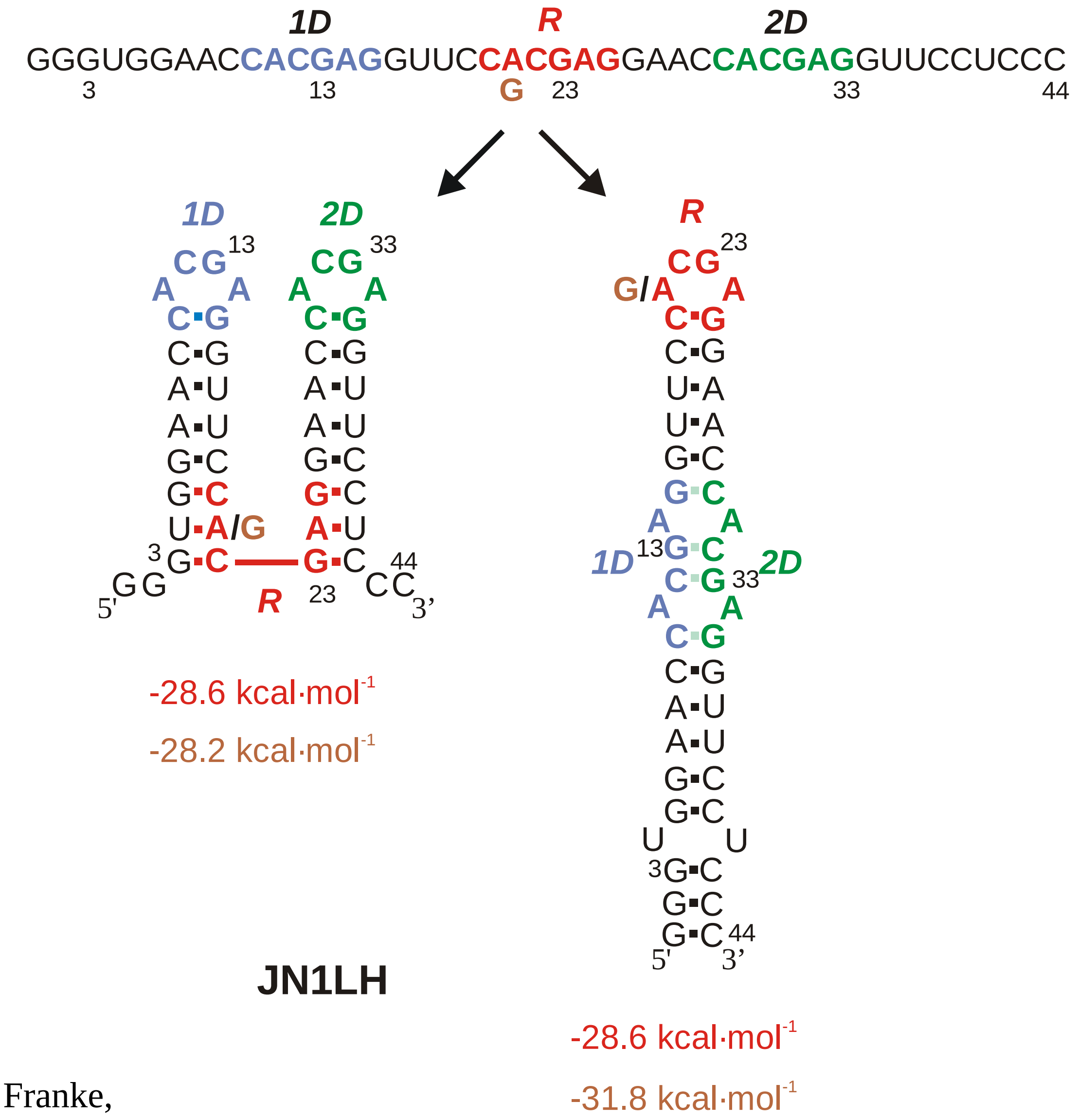


Extension of the notion of molecular structure



Extension of the notion of molecular structure

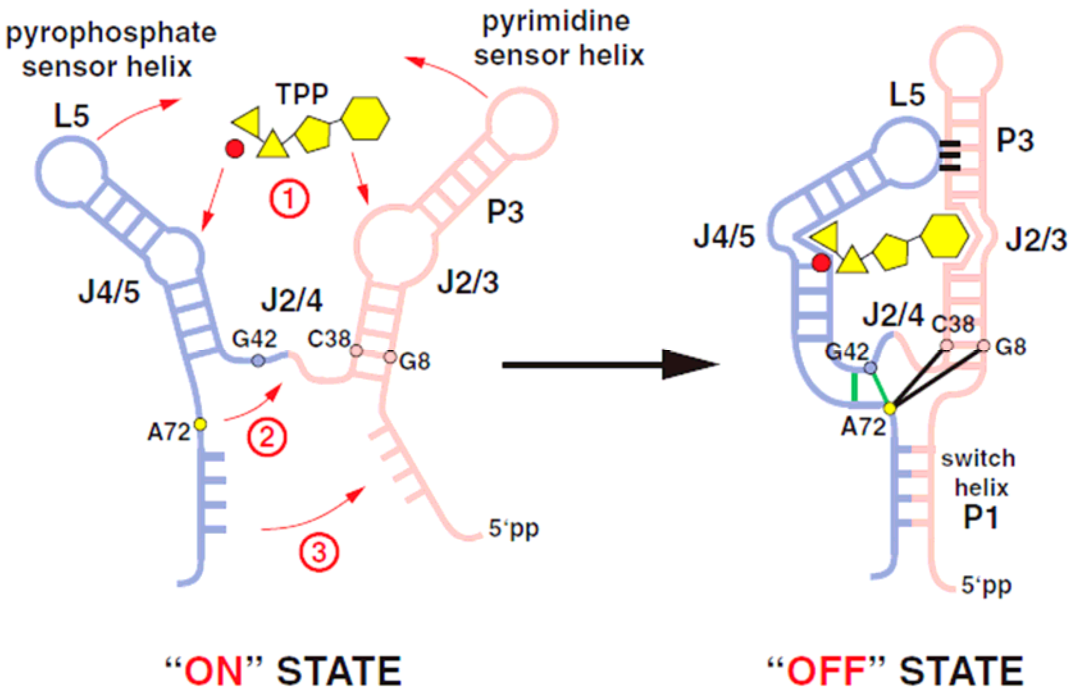
An RNA switch

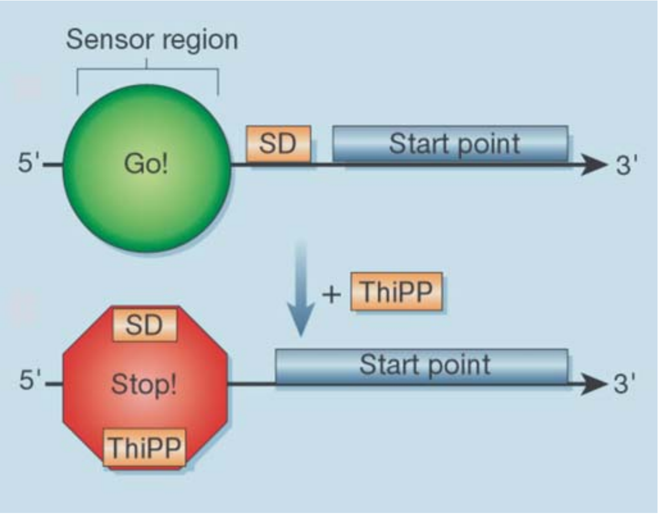


JN1LH

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

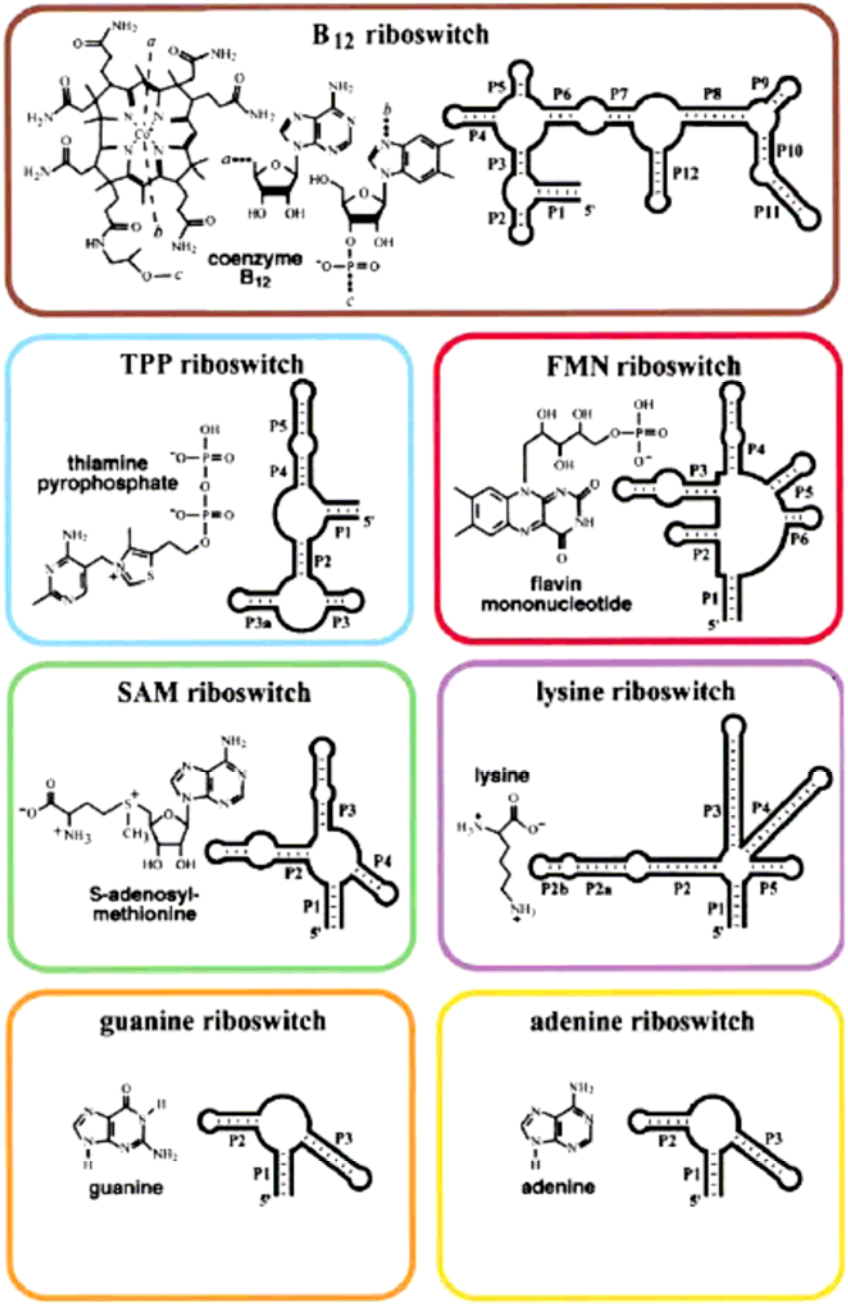
Structural parameters affecting the kinetic competition of RNA hairpin formation. *Nucleic Acids Res.* **34**:3568-3576, 2006.

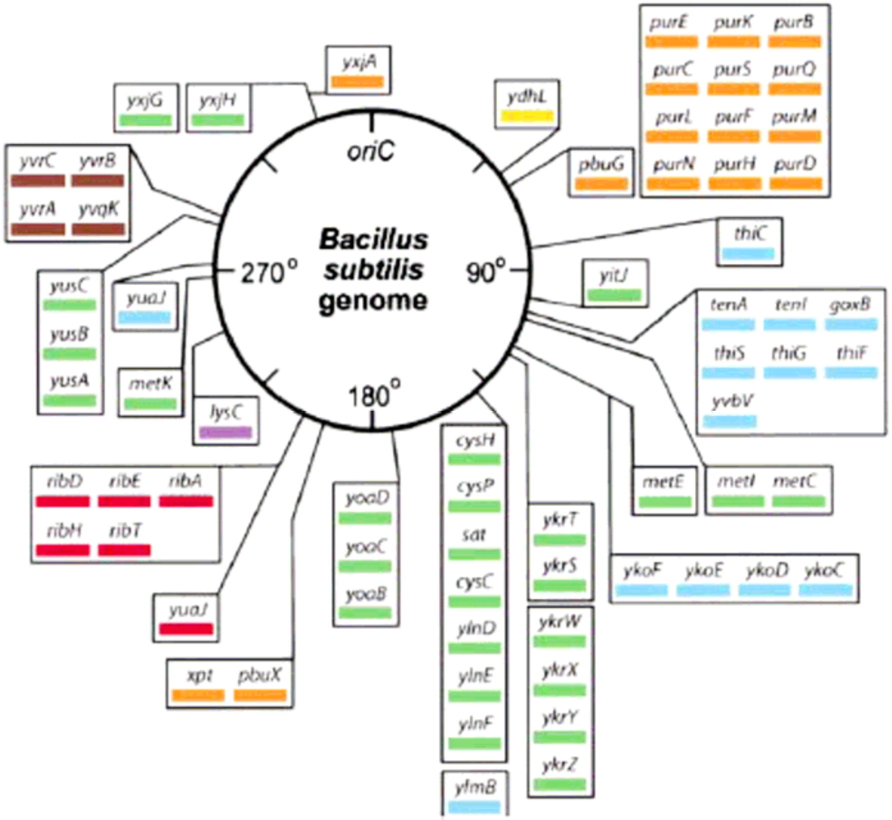




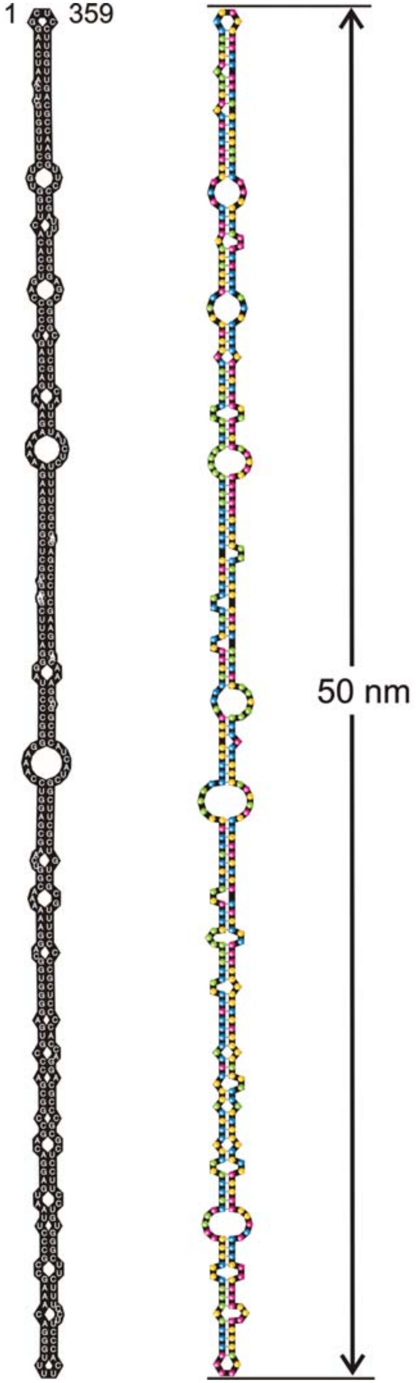
The natural thiamin-pyrophosphate RNA-switch

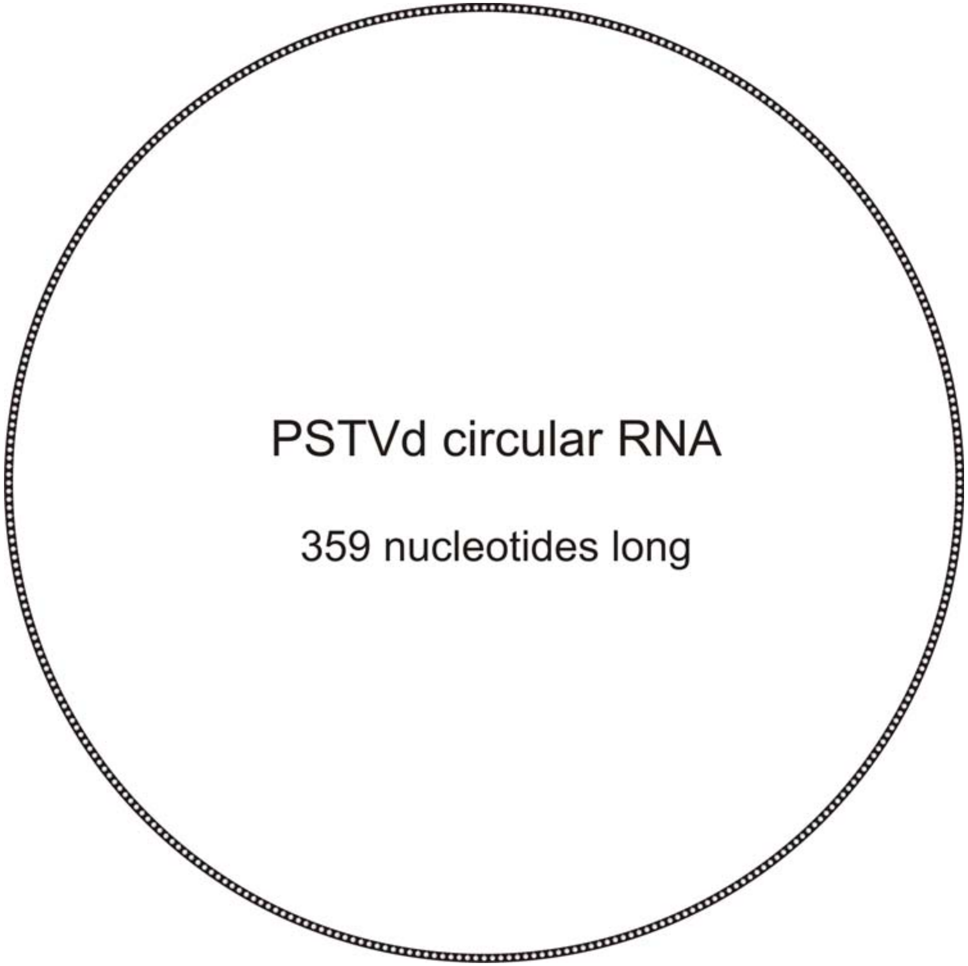
S. Thore, M. Leibundgut, N. Ban.
Science **312**:1208-1211, 2006.





M. Mandal, B. Boese, J.E. Barrick, W.C. Winkler, R.R, Breaker.
Cell 113:577-586 (2003)

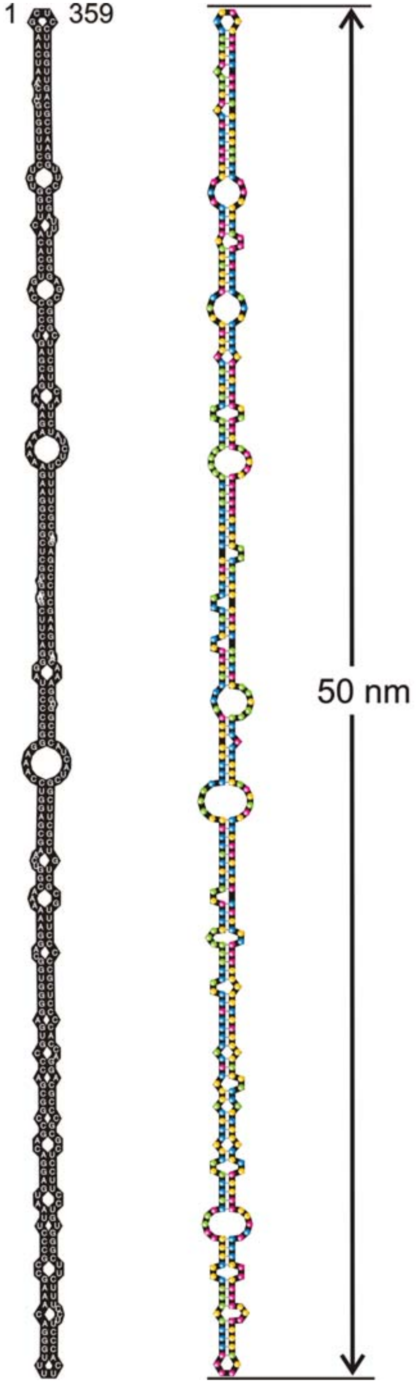


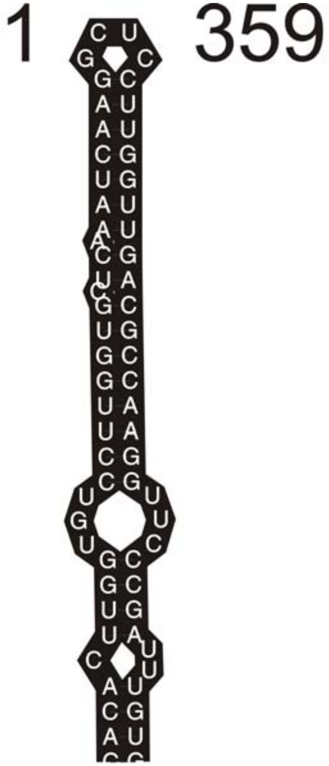


- ● **Adenine**
- ● **Uracil**
- ● **Guanine**
- ● **Cytosine**

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)





Vienna RNA Package 1.8.2

Biochemically supported structure

🟢 **Adenine**

🔴 **Uracil**

🔵 **Guanine**

🟡 **Cytosine**

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)

Exponential growth of RNA replicators requires structures with sufficient stability resulting from many but not too stable stacks.

Multiconformational RNA molecules can fulfill otherwise conflicting criteria.

If Charles Darwin would have written the „Origin“
in mathematical language, how would he have done it?

Molecular replicators – What do they need and how do they work?

Can we have selection and coexistence simultaneously?
Where does neutrality come from and what is it good for in evolution?

Symbioses, hypercycles and other dynamical mechanisms allow for the coexistence of replicators through

suppression of selection

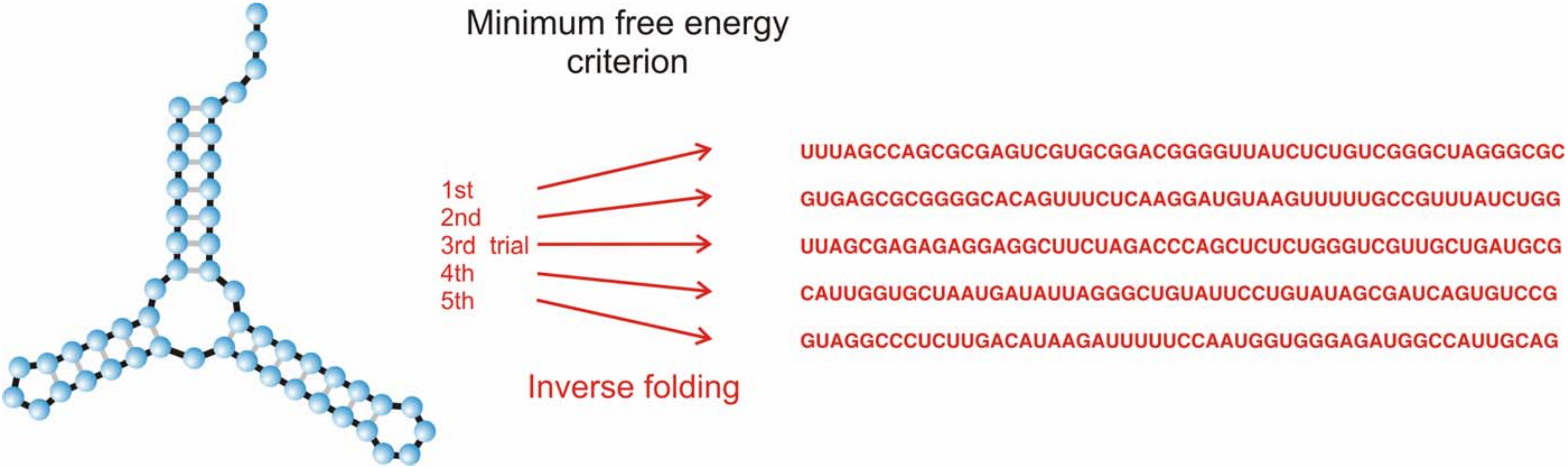
by functional coupling and interdependence.

Is there an alternative? Neutrality?

What is neutrality ?

Selective neutrality =
= several genotypes having the same fitness.

Structural neutrality =
= several genotypes forming molecules with the same structure.

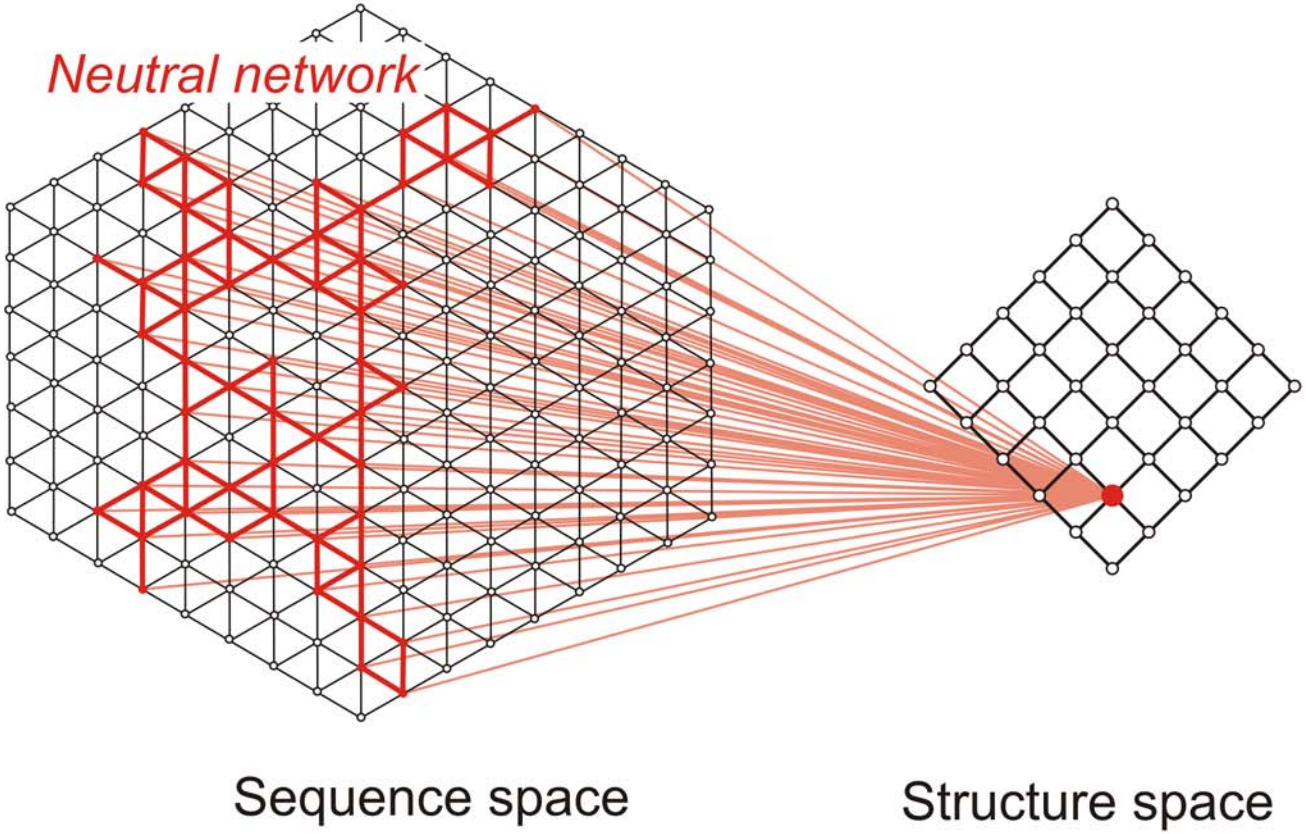


Vienna RNA-Package

Version 1.8.3

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

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



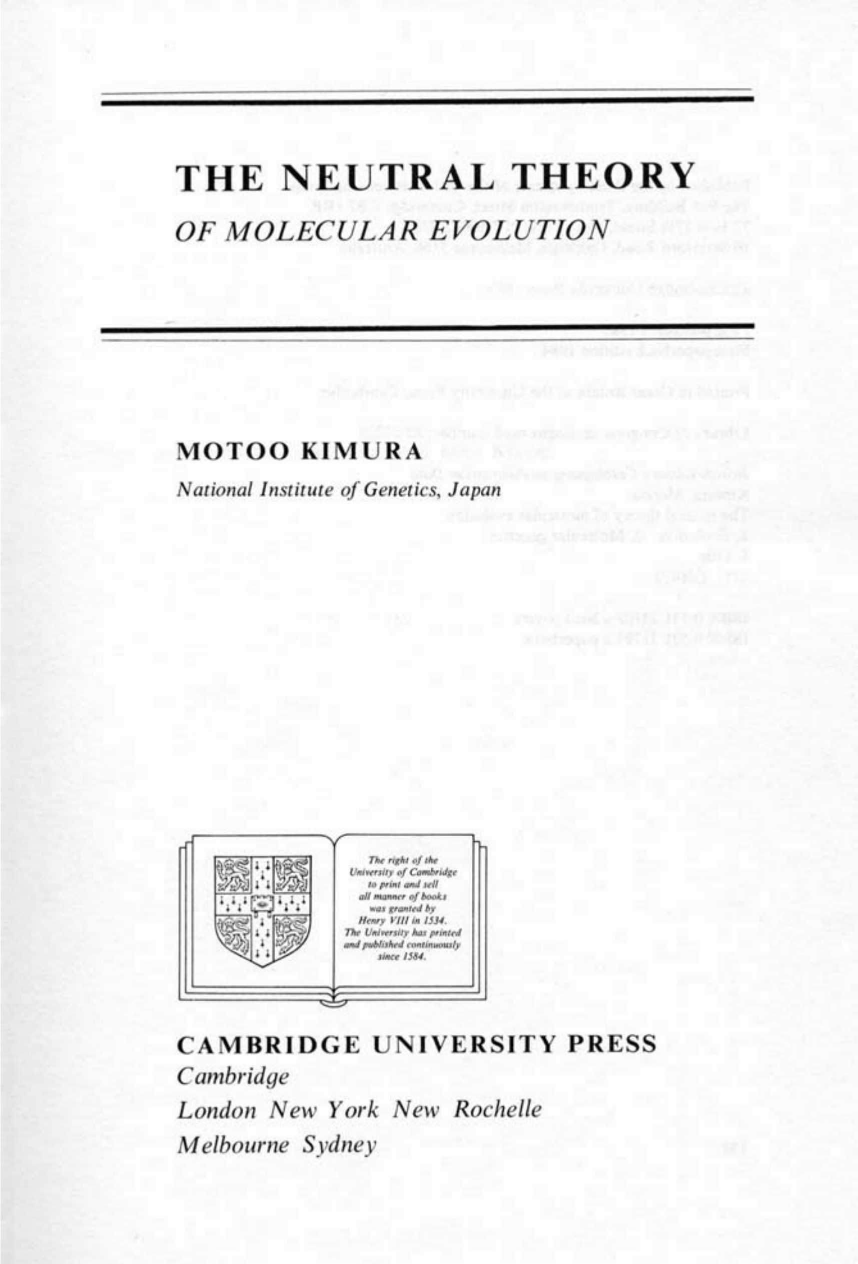
many genotypes ⇒ one phenotype

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.



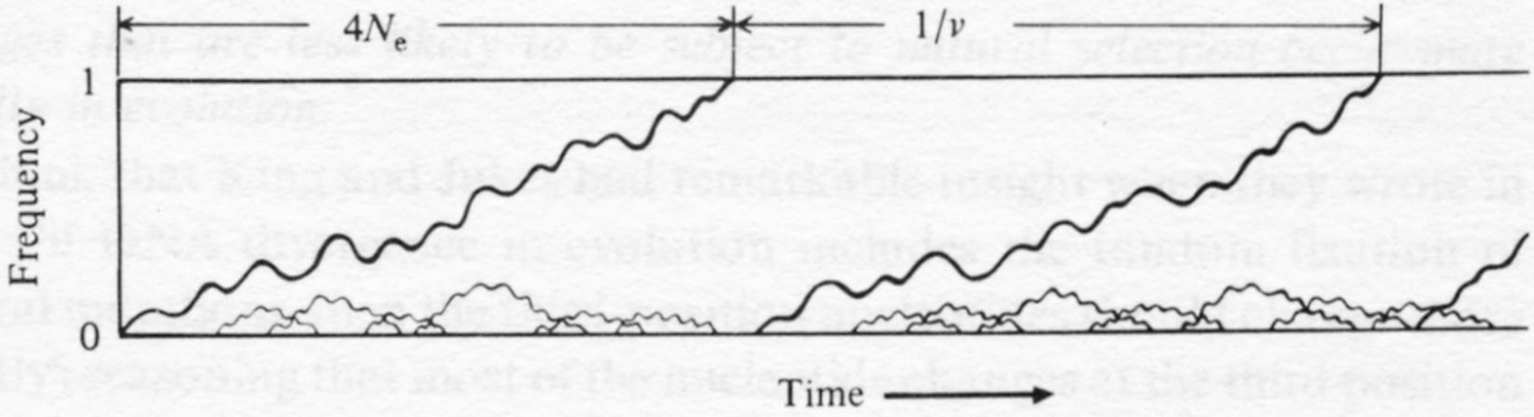
THE NEUTRAL THEORY
OF MOLECULAR EVOLUTION

MOTOO KIMURA
National Institute of Genetics, Japan

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

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.

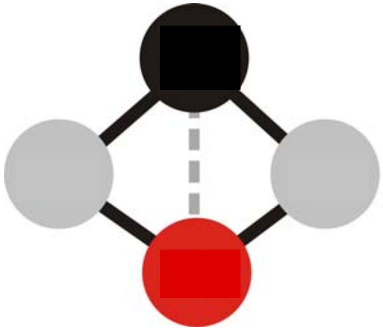


Motoo Kimura

Is the Kimura scenario correct for frequent mutations?

$d_H = 1$

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



$d_H = 2$

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

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

$d_H \geq 3$

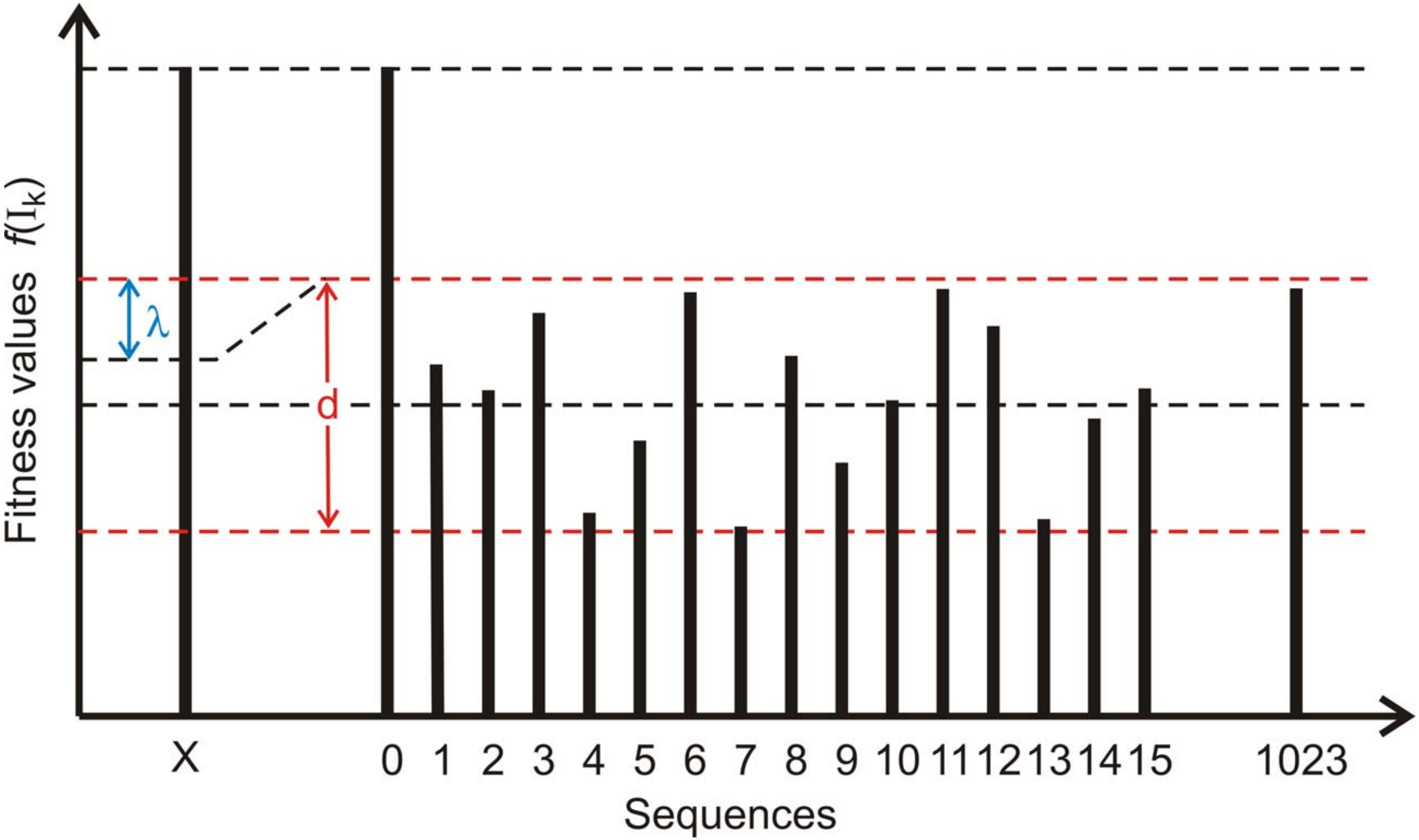
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

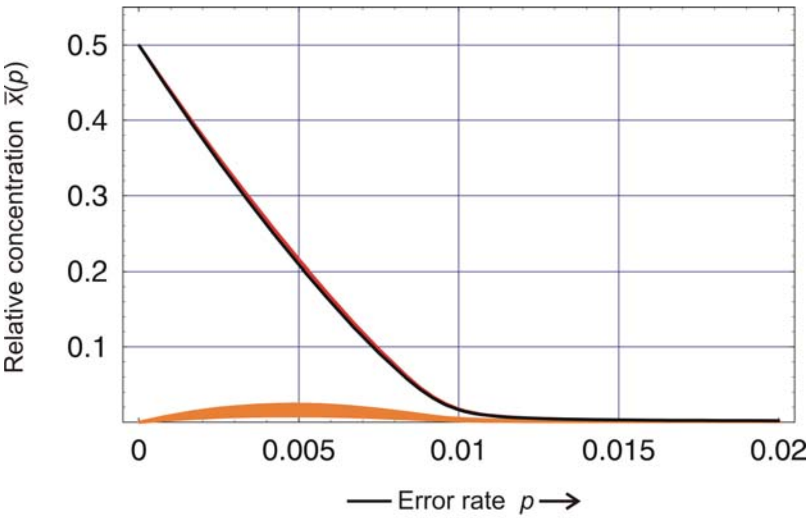
Pairs of neutral sequences in replication networks

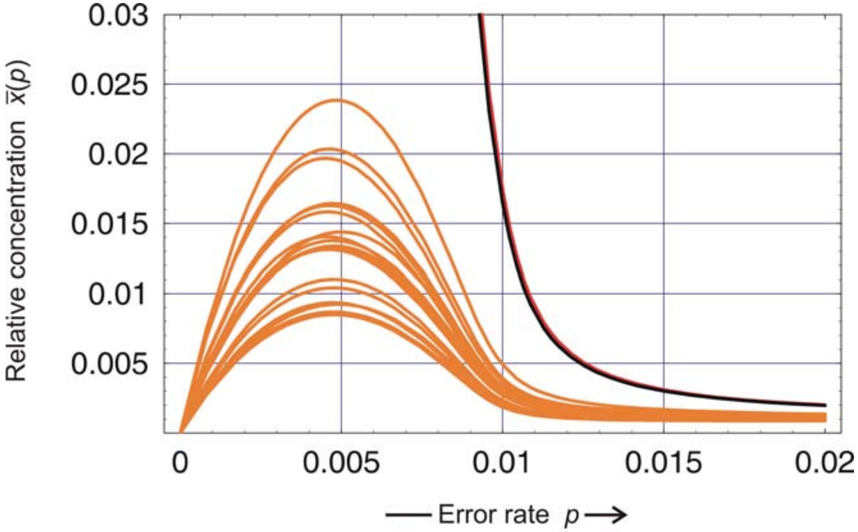
Random fixation in the sense of Motoo Kimura

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



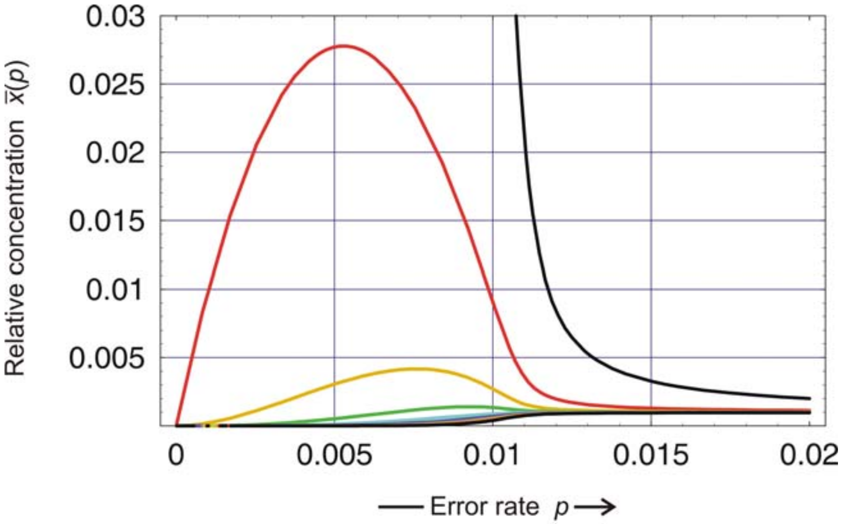
A fitness landscape including neutrality





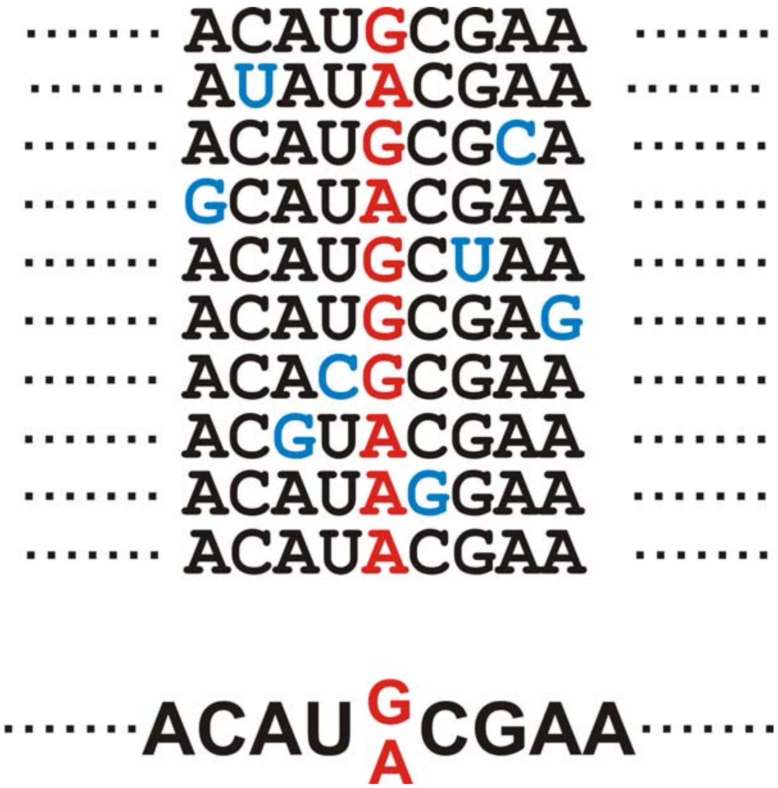
Neutral network

$\lambda = 0.01$, $s = 367$

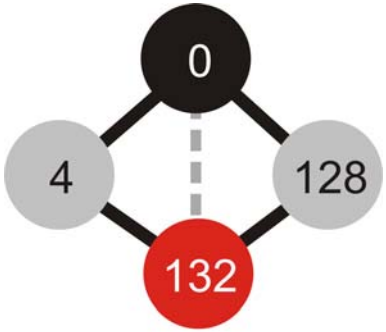


Neutral network: Individual sequences

$n = 10$, $\sigma = 1.1$, $d = 1.0$

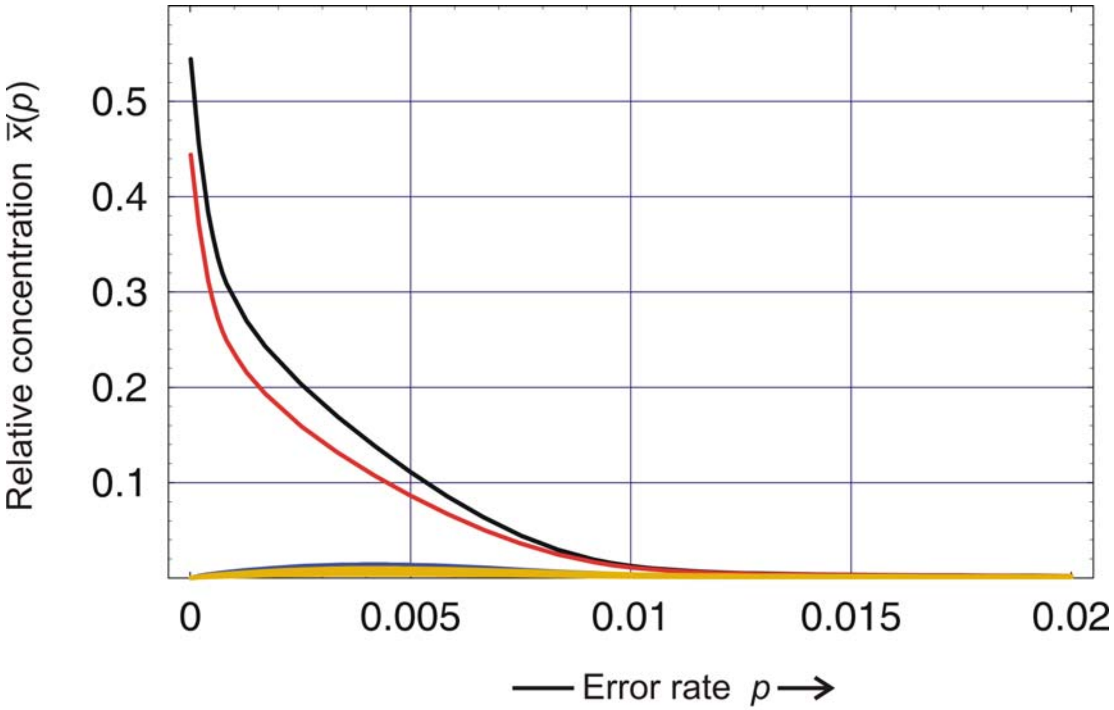


Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 1$.



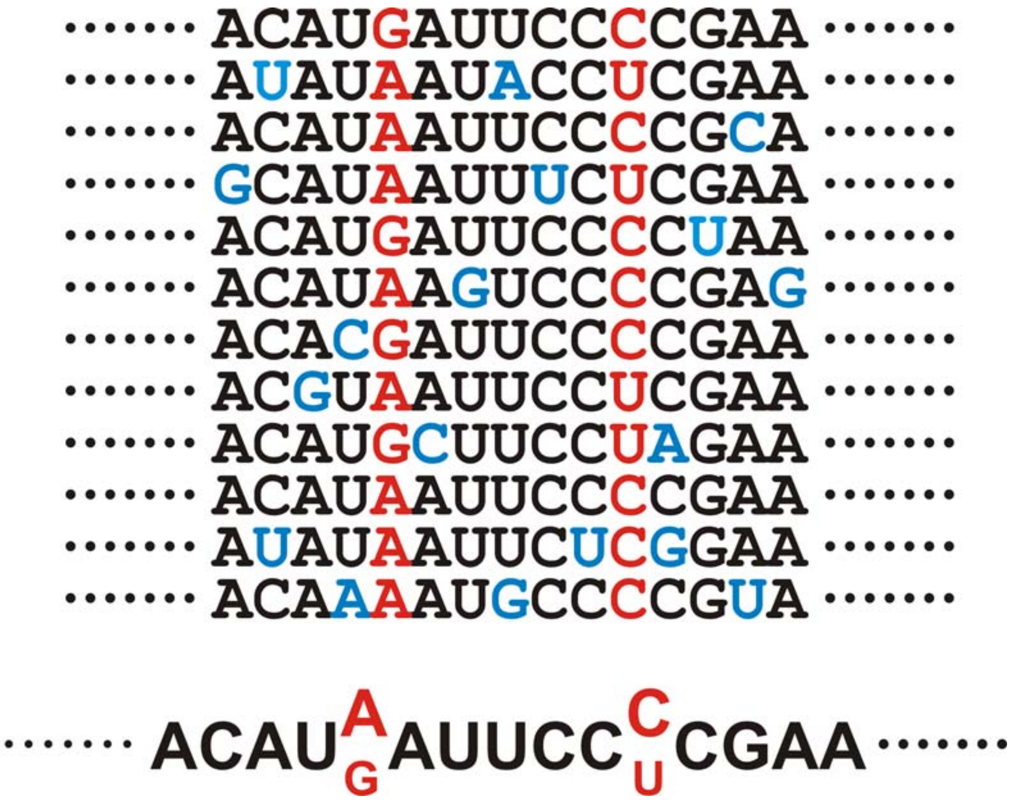
Neutral network

$\lambda = 0.01$, $s = 877$

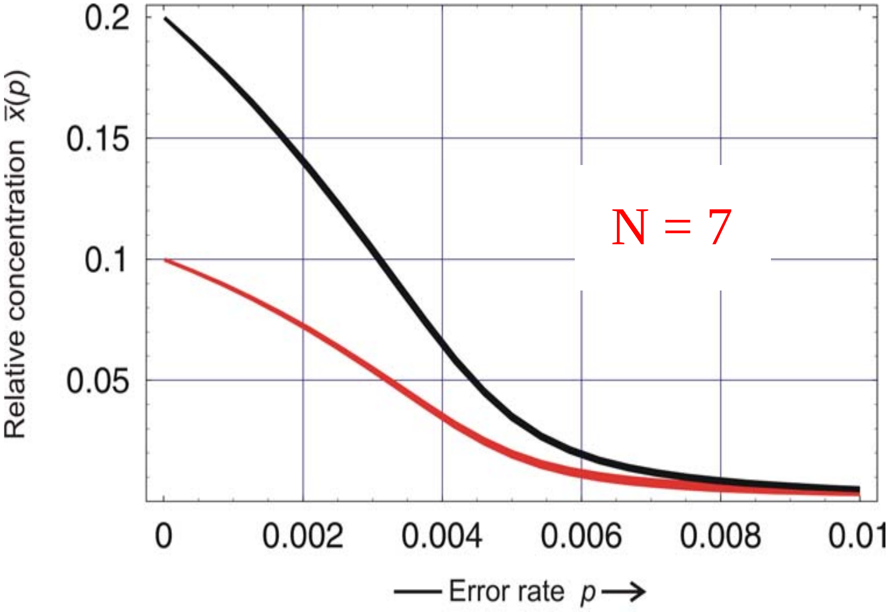


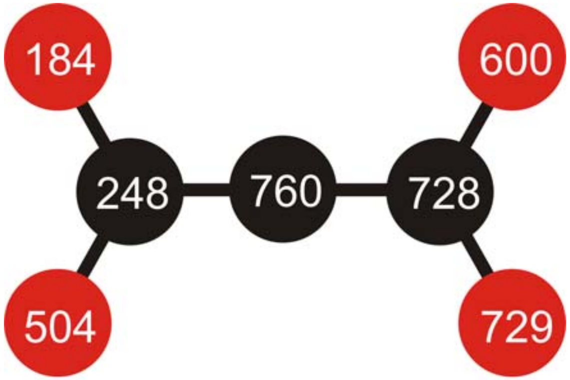
Neutral network: Individual sequences

$n = 10$, $\sigma = 1.1$, $d = 1.0$



Consensus sequence of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_{i,},X_j) = 2$.

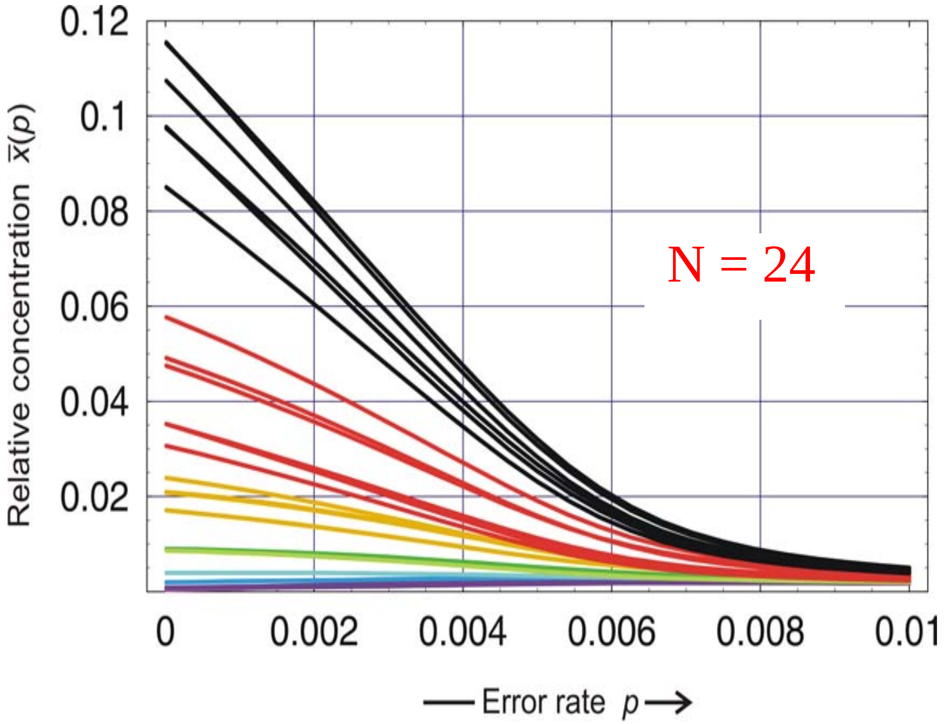


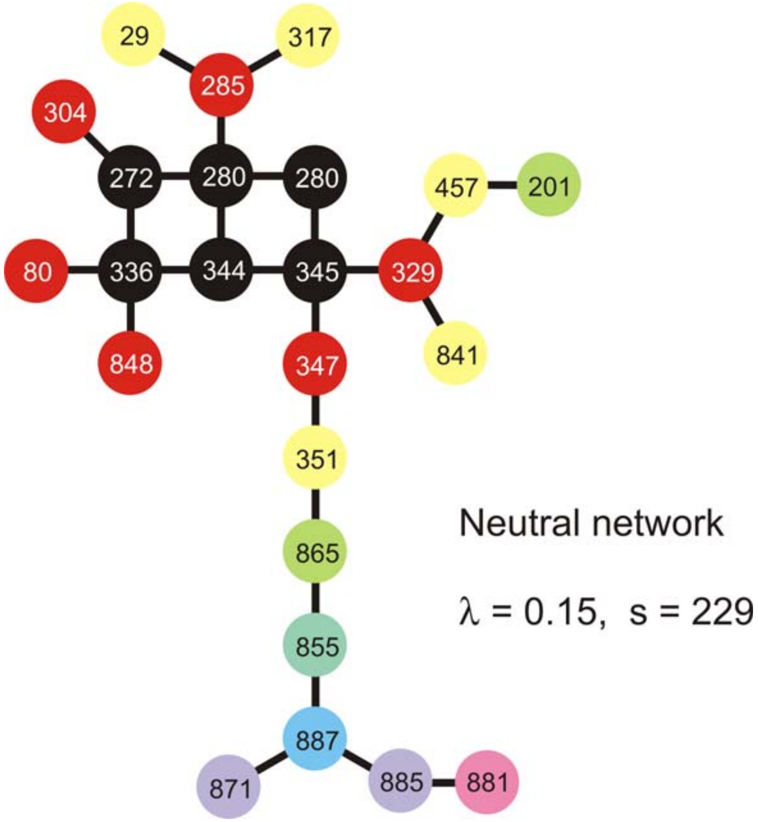


Neutral network

$\lambda = 0.10$, s = 229

Neutral networks with increasing λ: $\lambda = 0.10$, s = 229





Neutral networks with increasing λ: λ = 0.15, s = 229

Structural neutrality results from the enormous size of sequence space and the limited repertoire of structural motifs.

Neutrality in catalytic function is a result of the limited repertoire of transition state structures.

Neutrality in replication dynamics allows for simultaneous selection and coexistence.

Coworkers

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

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

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

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

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

Christian Forst, Los Alamos National Laboratory, NM

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

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

Kurt Grünberger, **Michael Kospach** , **Andreas Wernitznig**, **Stefanie Widder**,
Stefan Wuchty, **Jan Cupal**, **Stefan Bernhart**, **Lukas Endler**, **Ulrike Langhammer**,
Rainer Machne, **Ulrike Mückstein**, **Erich Bornberg-Bauer**,
Universität Wien, AT

Universität Wien

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: Bioinformatics Network (BIN)

Österreichische Akademie der Wissenschaften

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