Evolving molecules, viroids, and viruses

Theory, models, and reality

Peter Schuster

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

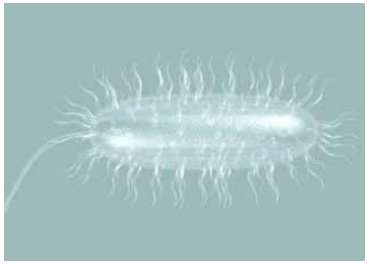
IUBMB & FEBS 2012

Sevilla, 04.– 09.09.2012

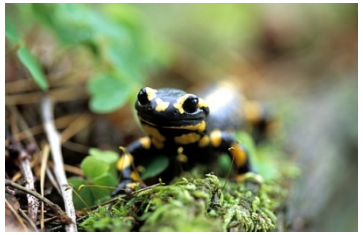
Prologue

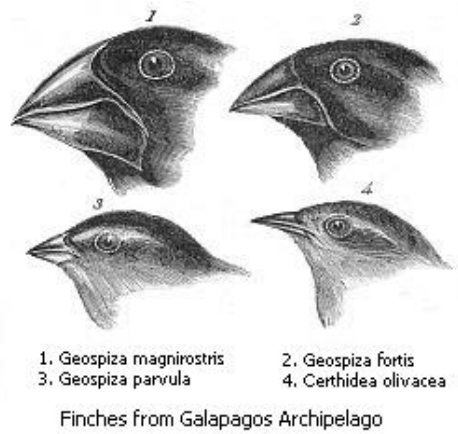
Charles Darwin, 1809 - 1882

Voyage on HMS Beagle, 1831 - 1836



Phenotypes





Finches from Galapagos Archipelago

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**.

Genotype, Genome

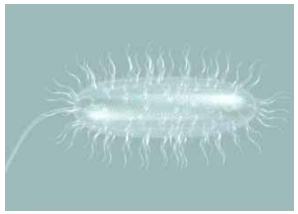
GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

Biochemistry
Structural Biology
Molecular Biology
Molecular Evolution
Molecular Genetics
Systems Biology
Bioinfomatics

Genetics
Epigenetics
Environment

Development

Cell Biology
Developmental Biology
Neurobiology
Microbiology
Botany and Zoology
Anthropology
Ecology



Phenotype

Biological evolution of higher organisms is an exceedingly complex process not because the mechanism of selection is complex but because cellular metabolism and control of organismic functions is highly sophisticated.

The Darwinian mechanism of selection does neither require organisms nor cells for its operation.

Make things as simple as possible,
but not simpler.

Albert Einstein, 1950 (?)

Occam's razor: Sir William Hamilton, 1852

1. Darwin and mathematics
2. Digitalizing chemistry
3. Evolution in the test tube
4. Viroids and viruses
5. Global genotype evolution

1. **Darwin and mathematics**
2. Digitalizing chemistry
3. Evolution in the test tube
4. Viroids and viruses
5. Global genotype evolution

There will never be a Newton of the blade of grass.

Immanuel Kant, 1790

Is it really impossible to cast the questions concerning evolution into a concise mathematical formulation?

Darwin‘s selection and Mendelian genetics have been first united in the mathematical model of population genetics.

Present day molecular life sciences urgently need a suitable theoretical basis – I call it theoretical biology new.

Sydney Brenner, 1999

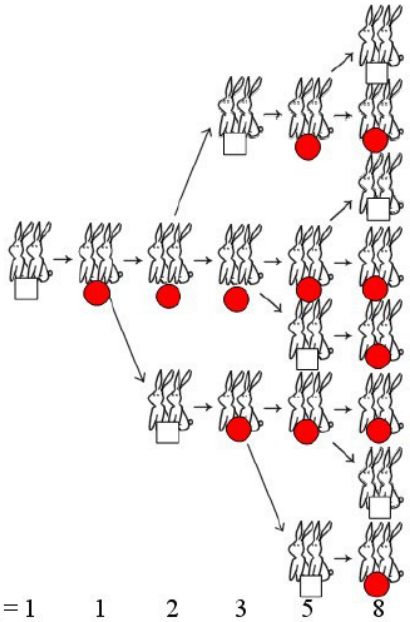
Theoretical biology in the third millenium. Phil.Trans.Roy.Soc.London B 354:1963-1965

Thomas Robert Malthus,
1766 – 1834

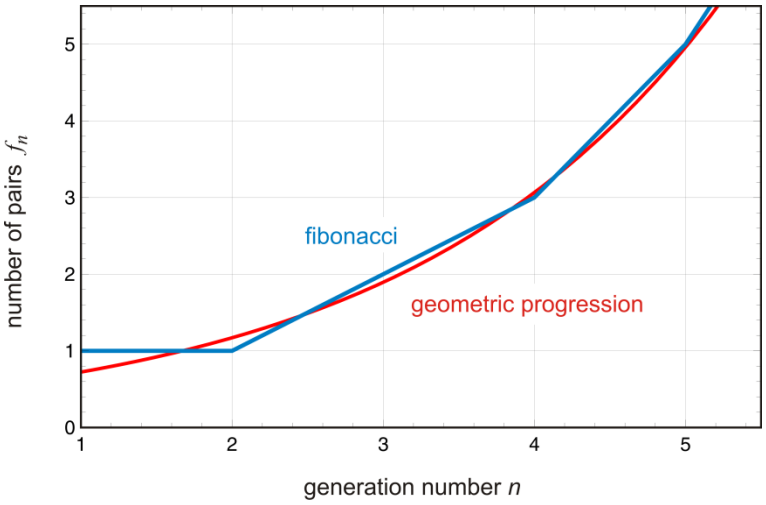
1, 2 , 4 , 8 ,16 , 32 , 64, 128 , ...

geometric progression

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



Leonardo da Pisa
„Fibonacci“
~1180 – ~1240



exponential function

Leonhard Euler, 1717 – 1783

The history of exponential growth

autocatalysis

$$\frac{dx}{dt} = f\, x(1-x)$$

competition

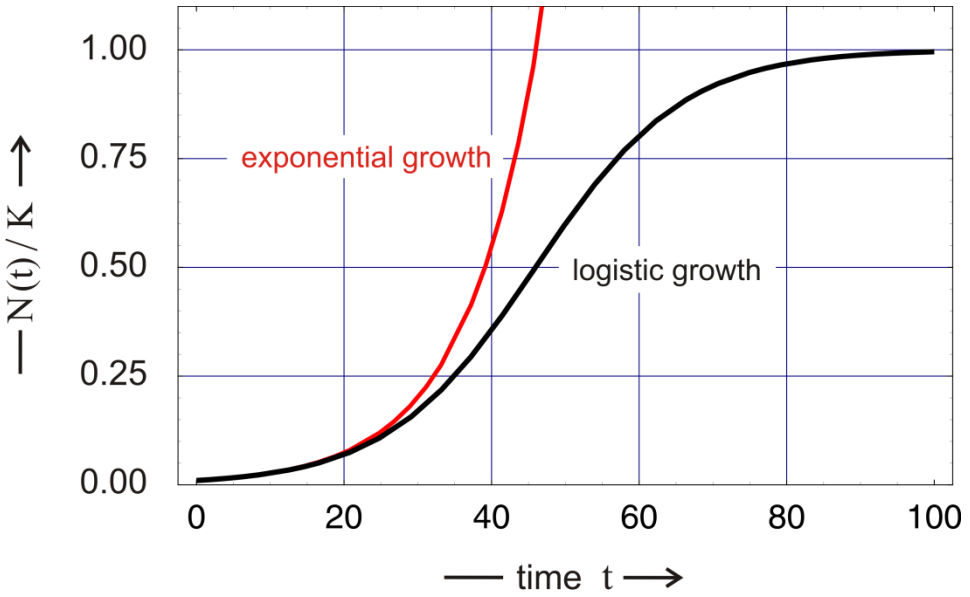
$$\frac{dx_k}{dt} = \left(f_k - \phi\right) x_k \,;\, k = 1,2,\ldots,n$$

$$\phi = \sum_{j=1}^{n} f_j\, x_j \,;\, \sum_{j=1}^{n} x_j = 1$$

The chemistry and the mathematics of reproduction

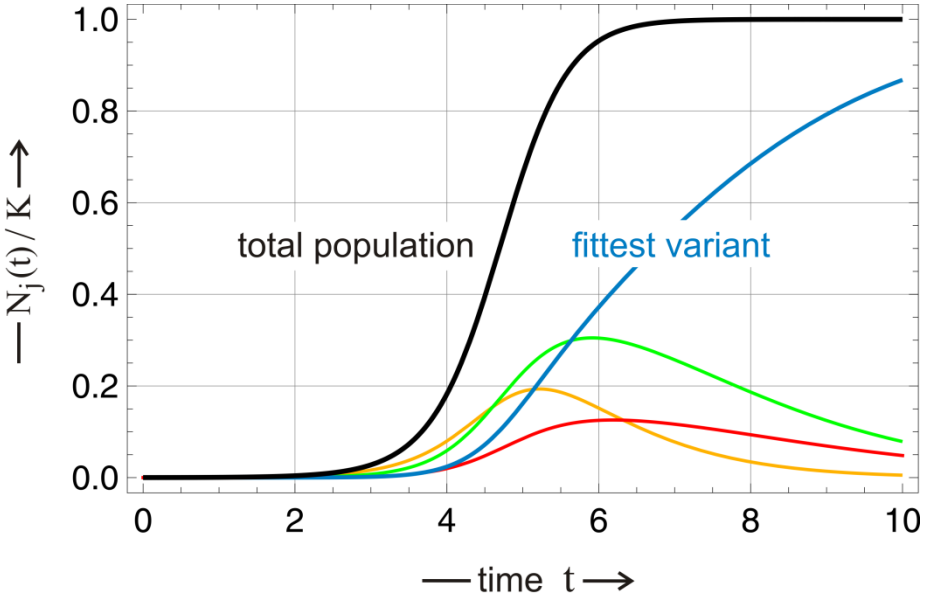
Pierre-François Verhulst,
1804-1849

the consequence of
finite resources



fitness values:

$f_1 = 2.80$, $f_2 = 2.35$, $f_3 = 2.25$,
and $f_4 = 1.75$

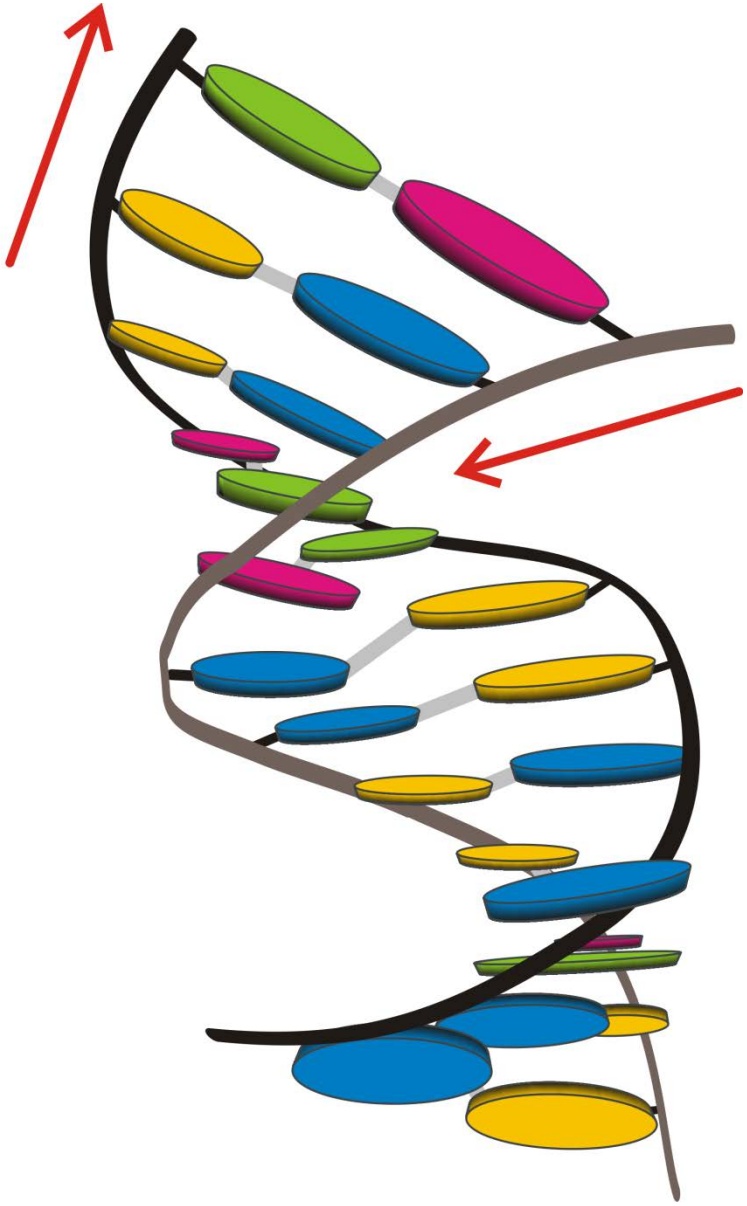


The logistic equation, 1828

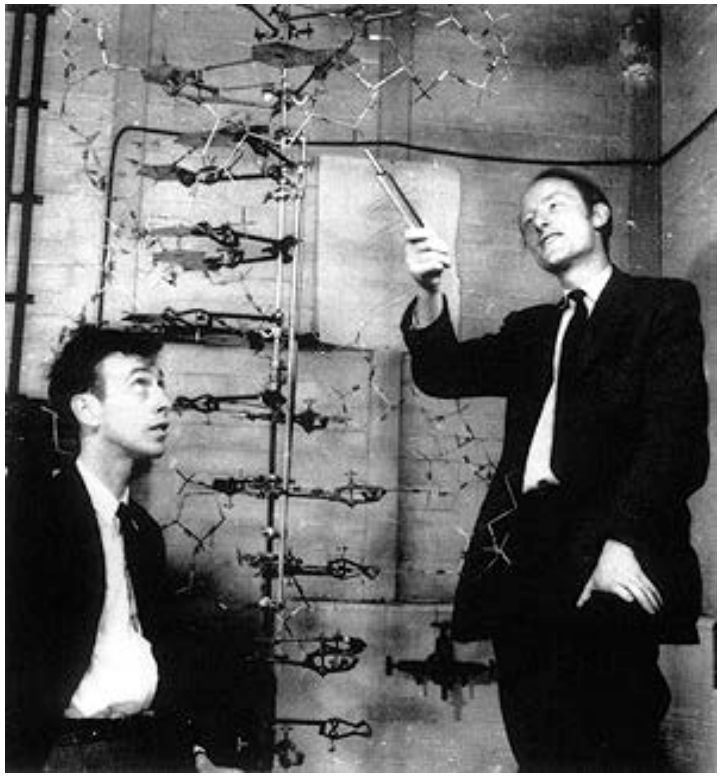
All mathematics required for modeling Darwin‘s principle of selection was readily available to his contemporary mathematicians.

It took about 70 years before selection has been cast into a mathematical model by the three great population geneticists Ronald A. Fisher, J.B.S. Haldane and Sewall Wright.

1. Darwin and mathematics
2. **Digitalizing chemistry**
3. Evolution in the test tube
4. Viroids and viruses
5. Global genotype evolution



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



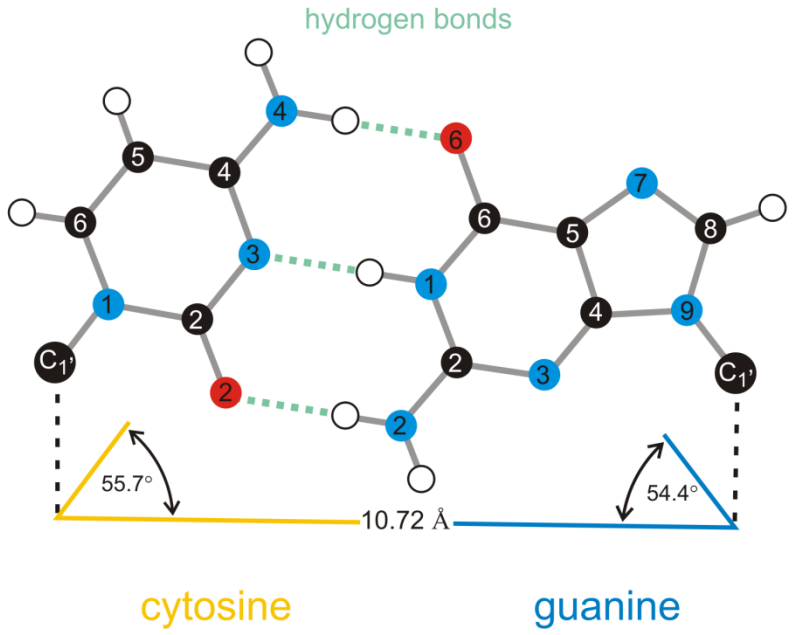
James D. Watson, 1928- , and Francis Crick, 1916-2004,
Nobel Prize 1962

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

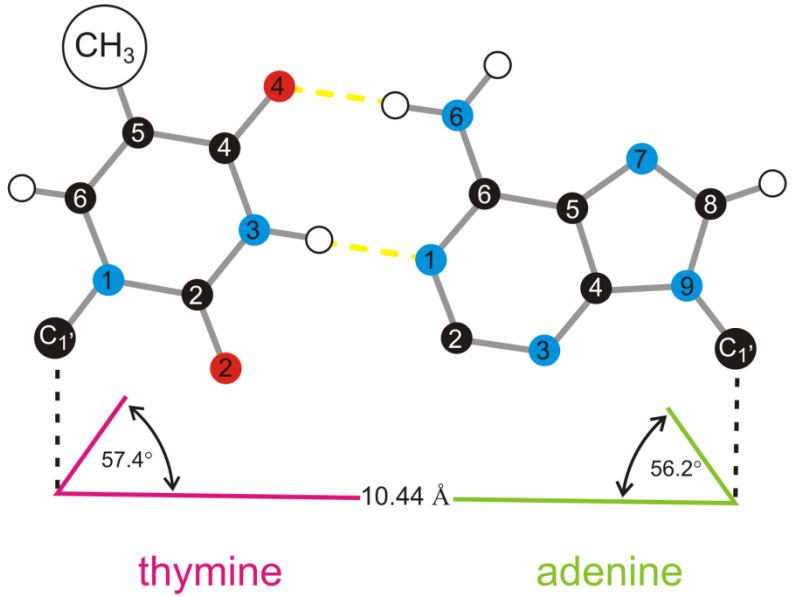
A Structure for Deoxyribose Nucleic Acid
Nature 171:737-738 (1953)

Although interactions involving G are much stronger than all other interactions between nucleotides, A=T and G≡C are base pairs on an equal footing.

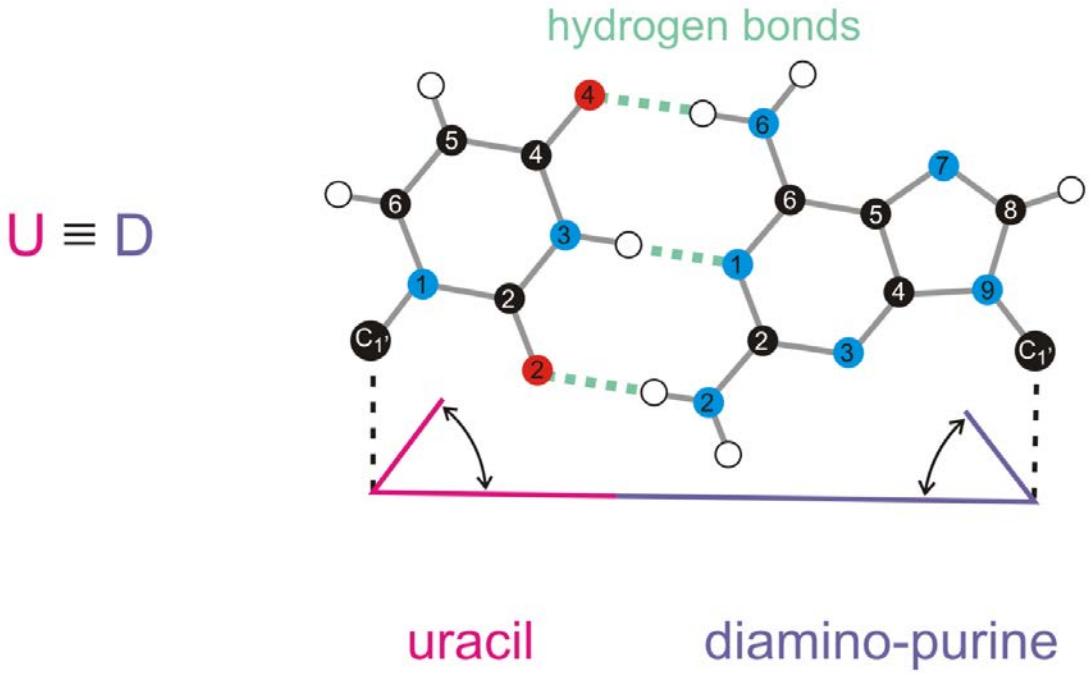
C ≡ G



T = A



Digitalization of chemistry:
The unique assigment of nucleotides in base pairs.

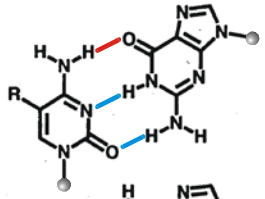
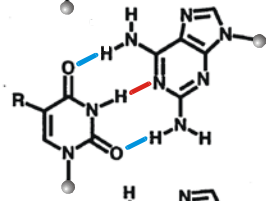
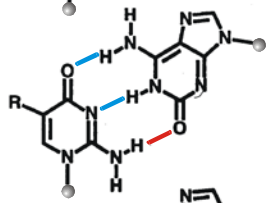
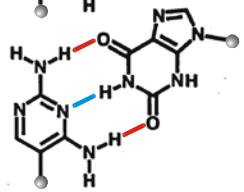
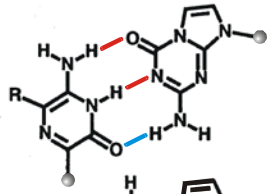
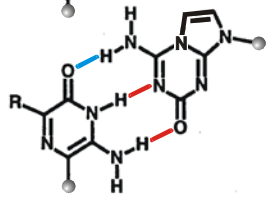


An example from synthetic biology: Introduction of a third hydrogen bond into the **U** = **A** base pair.

Color code:

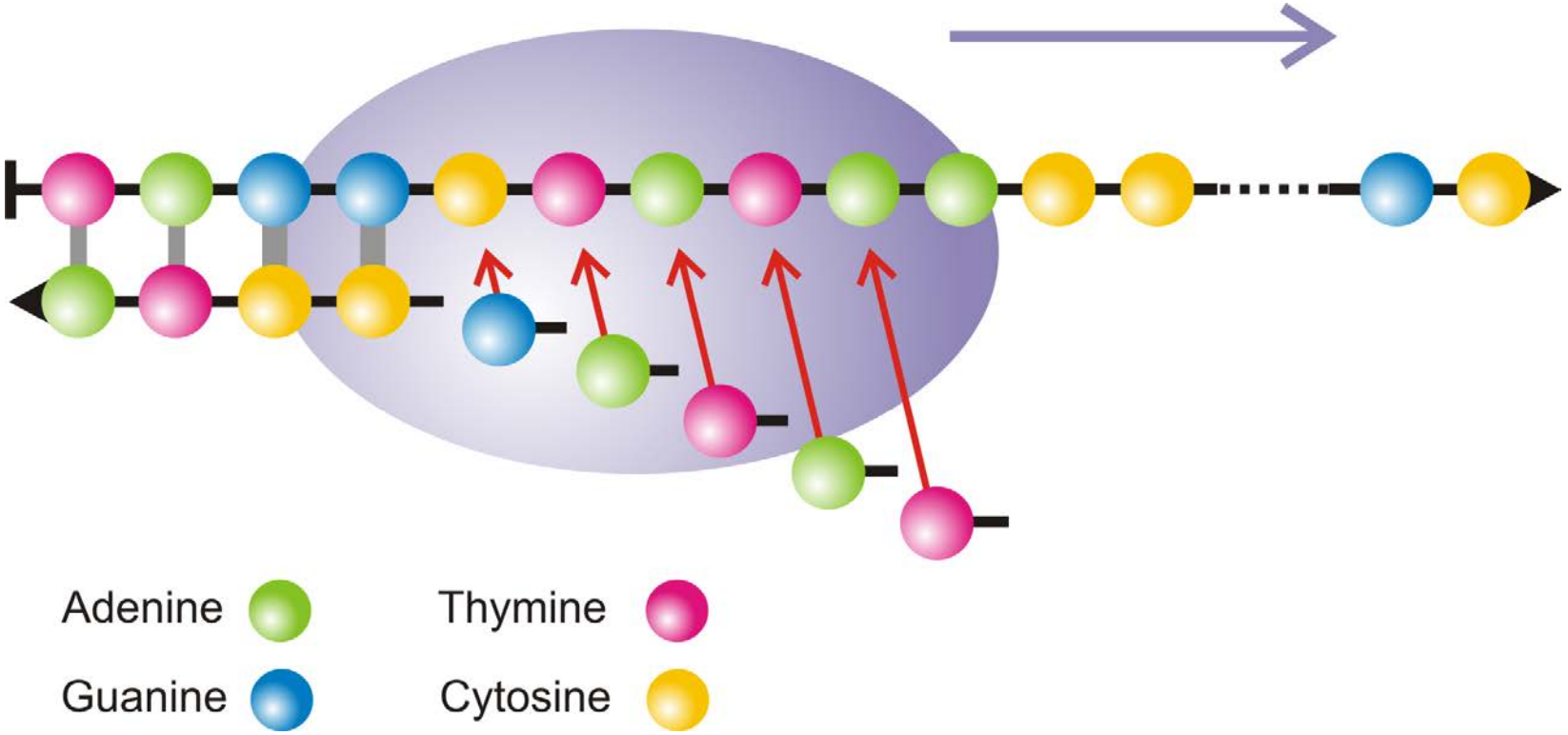
donor—acceptor

acceptor—donor

2-keto, 4-amino pyrimidine	Donor Acceptor Acceptor	C		G	Acceptor Donor Donor	**2-amino,6-keto purine**
2,4-diketo pyrimidine	Acceptor Donor Acceptor	U		``A´´	Donor Acceptor Donor	**2,6-diamino purine**
2-amino, 4-keto pyrimidine	Acceptor Acceptor Donor				Donor Donor Acceptor	**2-keto, 6-amino purine**
2,6-diamino pyrimidine	Donor Acceptor Donor				Acceptor Donor Acceptor	**2,6-diketo purine**
2-amino, 6-keto pyrazine	Donor Donor Acceptor				Acceptor Acceptor Donor	**5-keto, 7-amino, 1,6,8-triaza indolicine**
2-keto, 6-amino pyrazine	Acceptor Donor Donor				Donor Acceptor Acceptor	**5-amino, 7-keto, 1,6,8-triaza indolicine**

Hydrogen bonding patterns for Watson-Crick base pairs

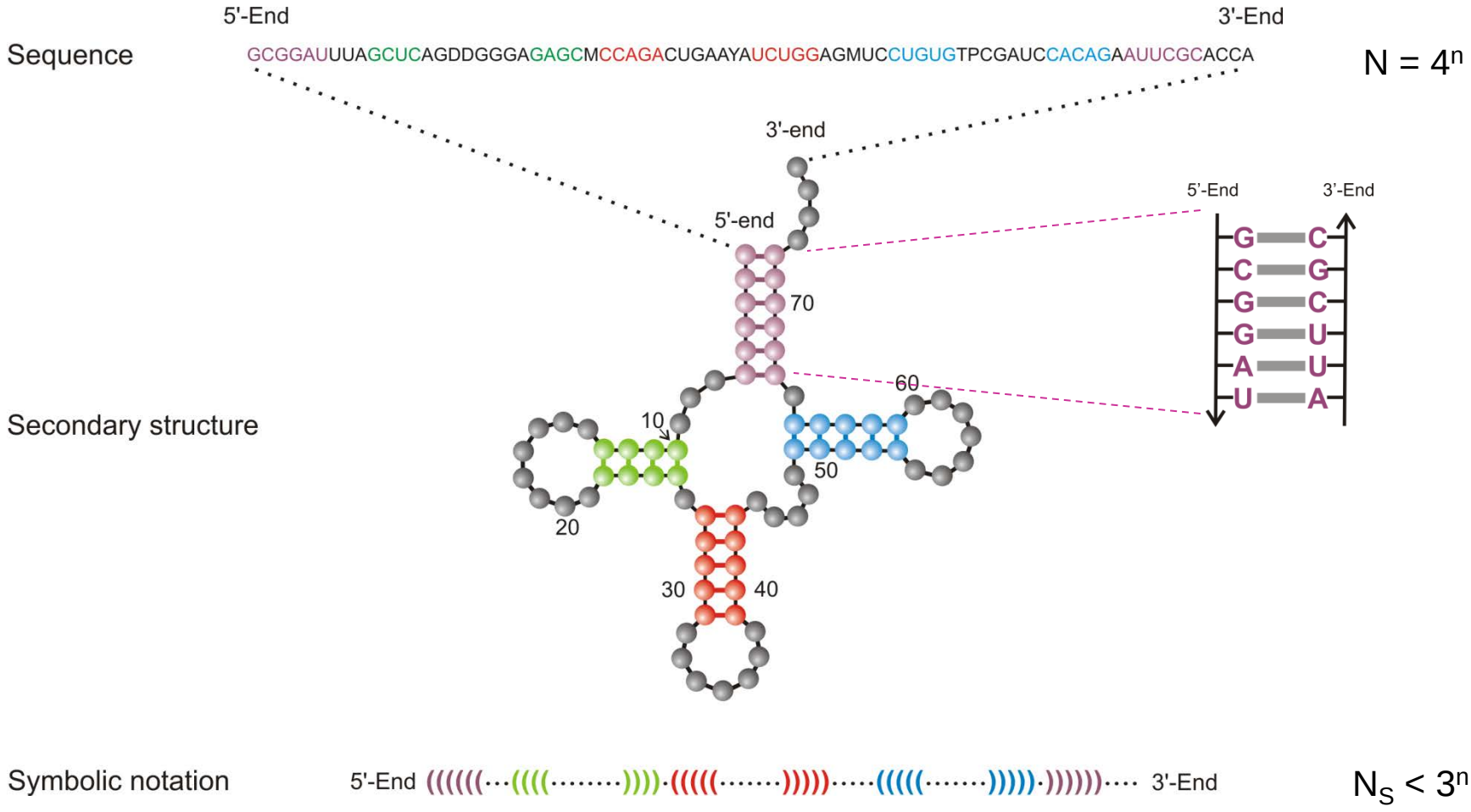
S.A. Benner *et al*., Reading the palimpsest: Contemporary biochemical data and the RNA world. In: R.F.Gesteland and J.F.Atkins, eds. The RNA World, pp.27-70. CSHL Press, 1993



The replication of DNA by Thermophilus aquaticus polymerase (PCR)

Accuracy of replication: $Q = q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot \ldots$

The logics of DNA (or RNA) replication



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

1. Darwin and mathematics
2. Digitalizing chemistry
3. **Evolution in the test tube**
4. Viroids and viruses
5. Global genotype evolution

Charles Darwin, 1809-1882

Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
1. **Variation**, and
1. **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

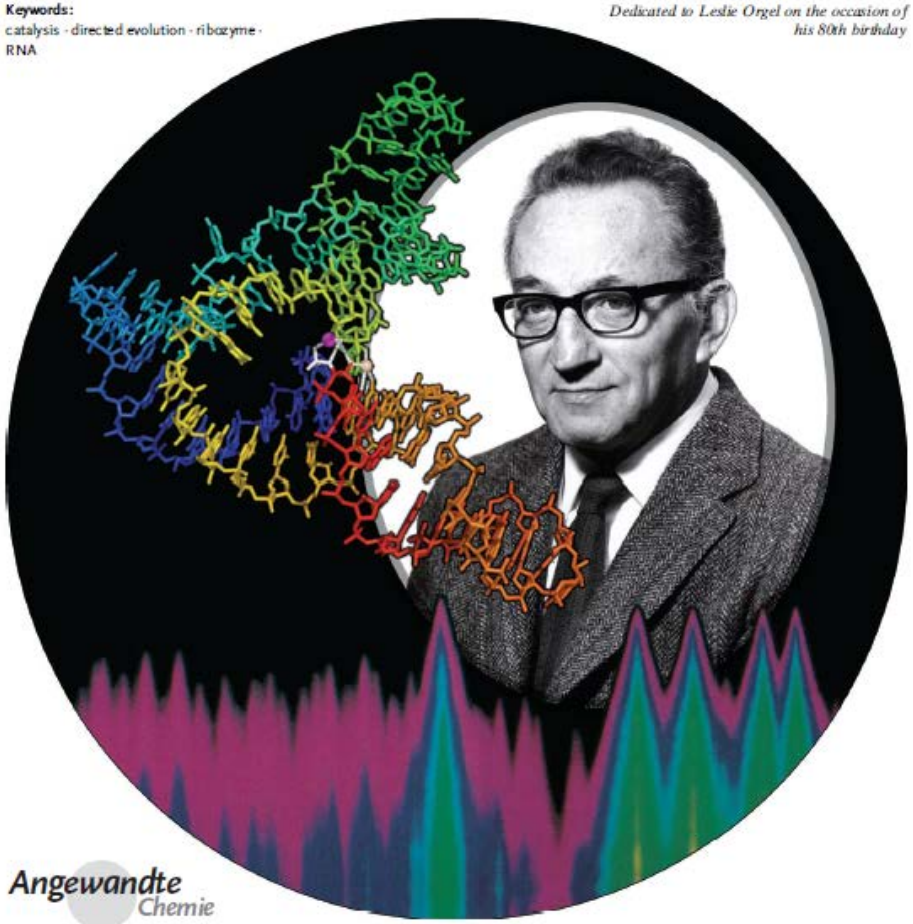
Molecular Evolution

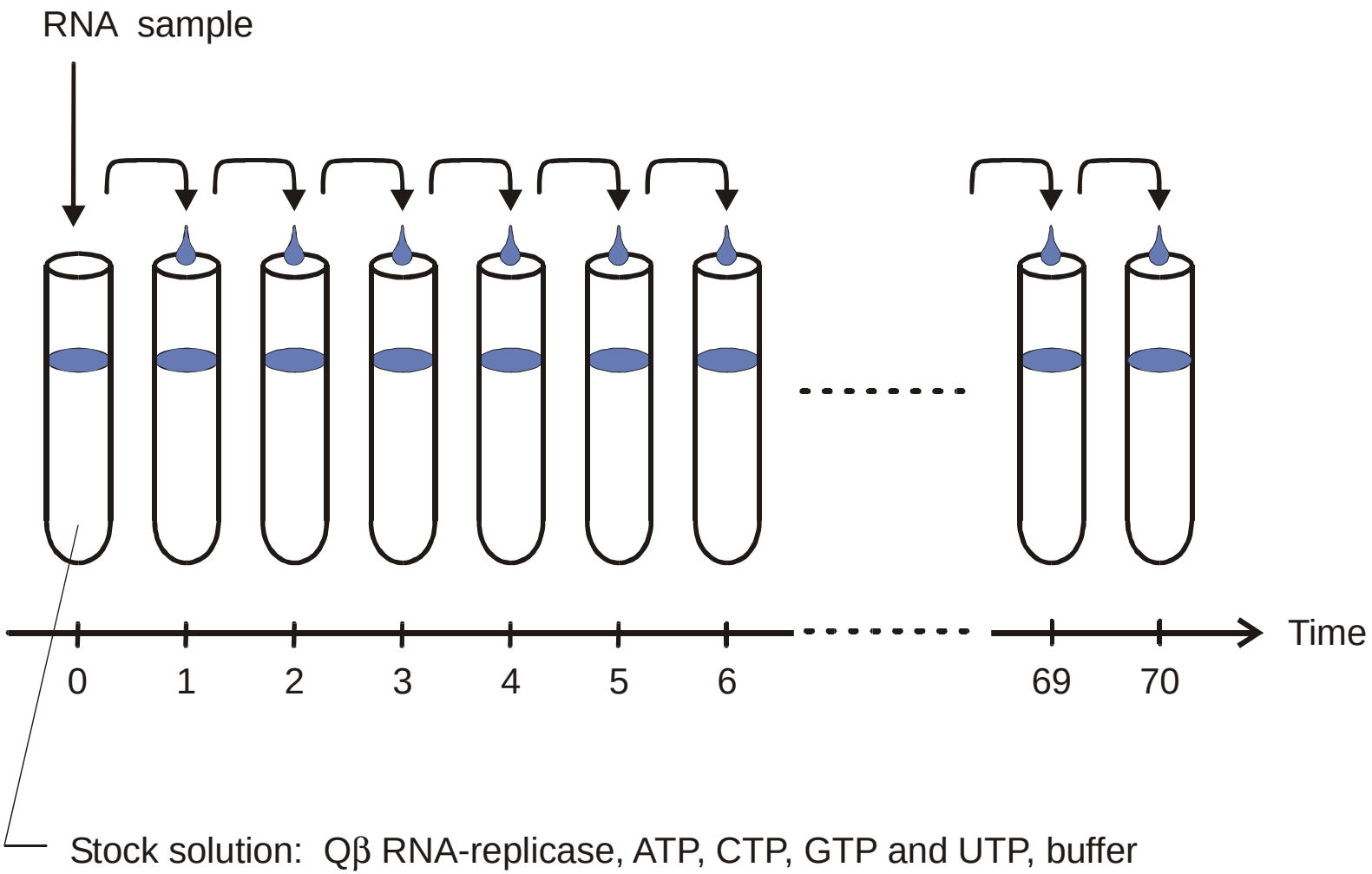
Forty Years of In Vitro Evolution**

*Gerald F. Joyce**

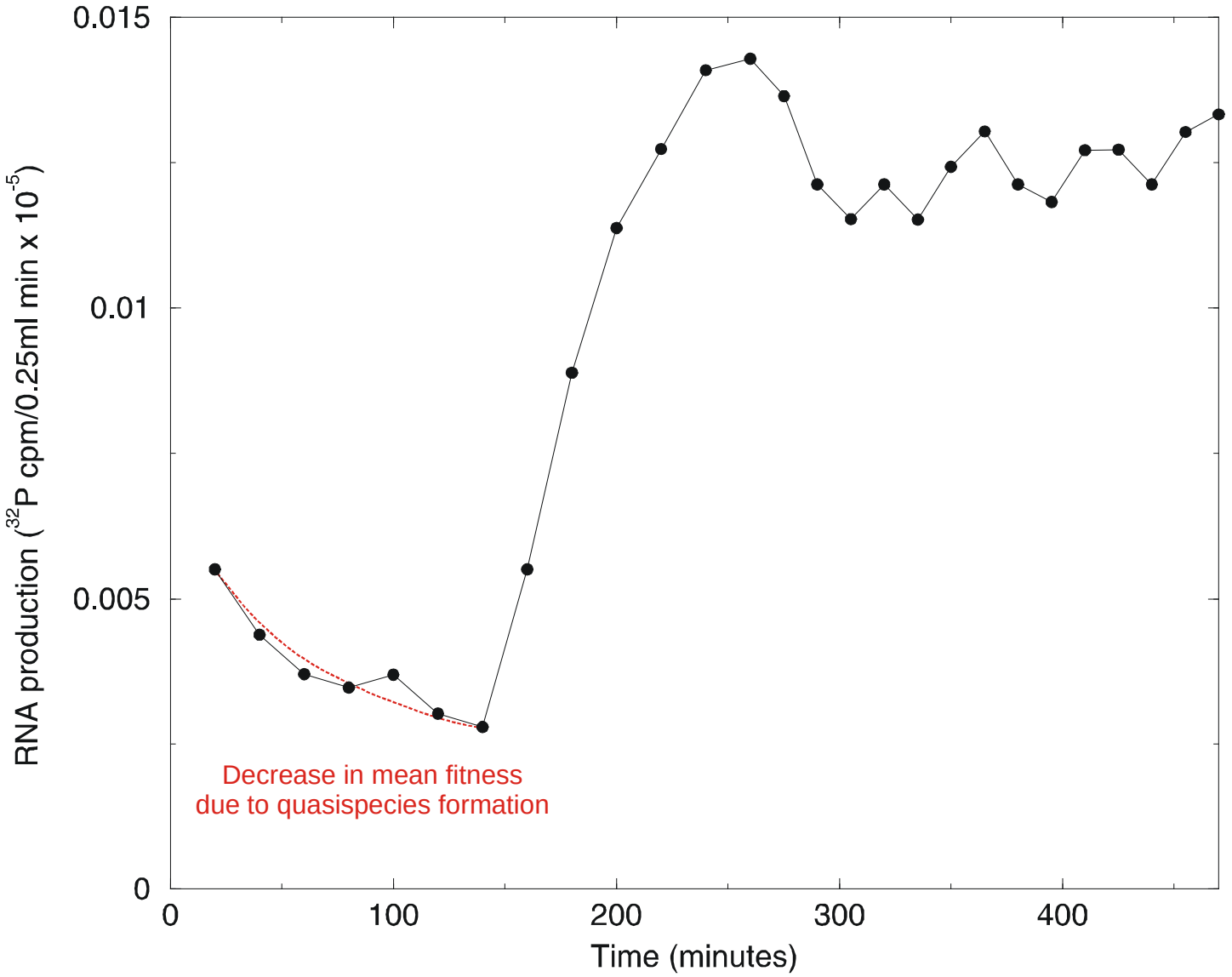
Keywords:
catalysis · directed evolution · ribozyme · RNA

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

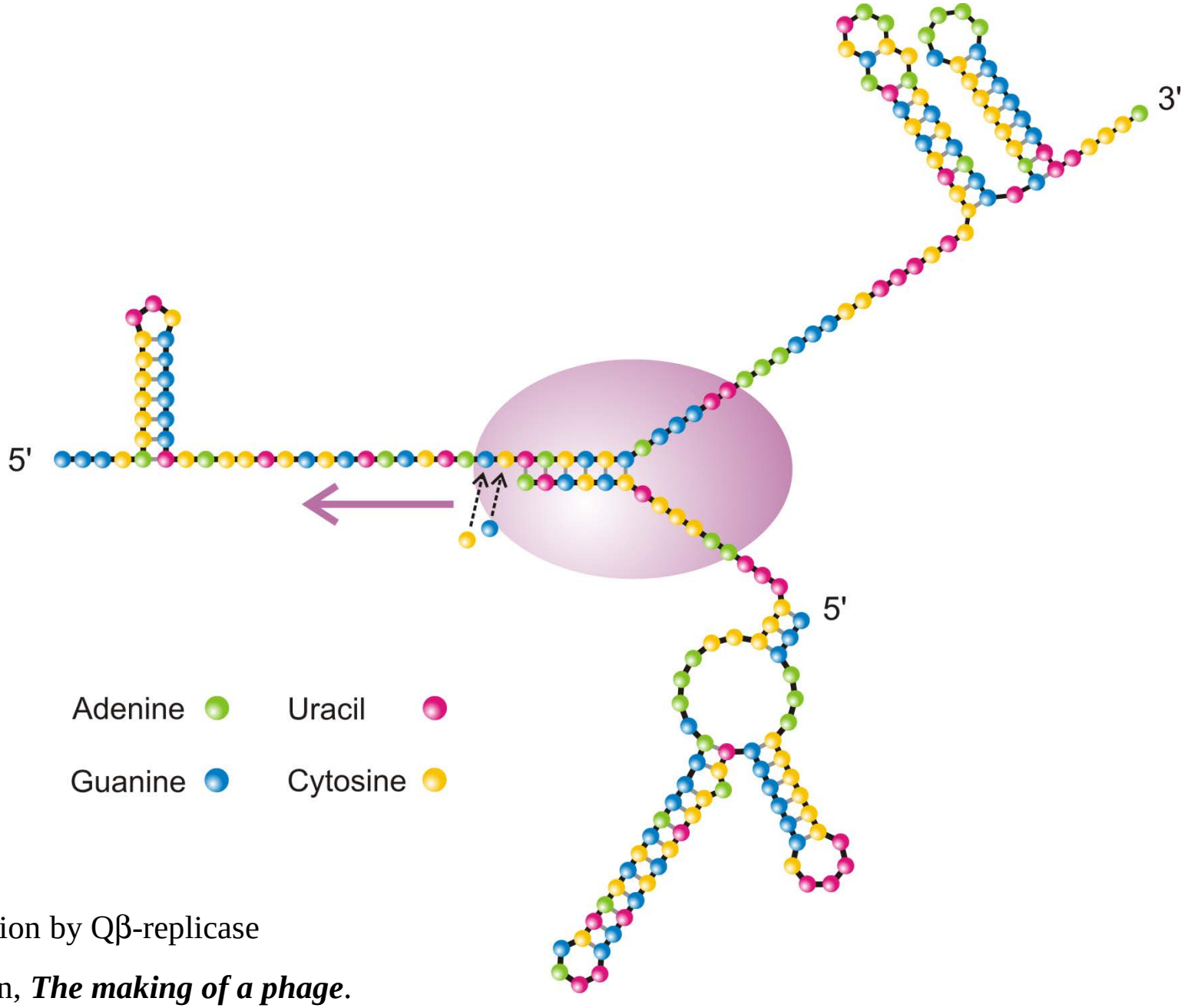




Application of serial transfer technique to evolution of RNA in the test tube



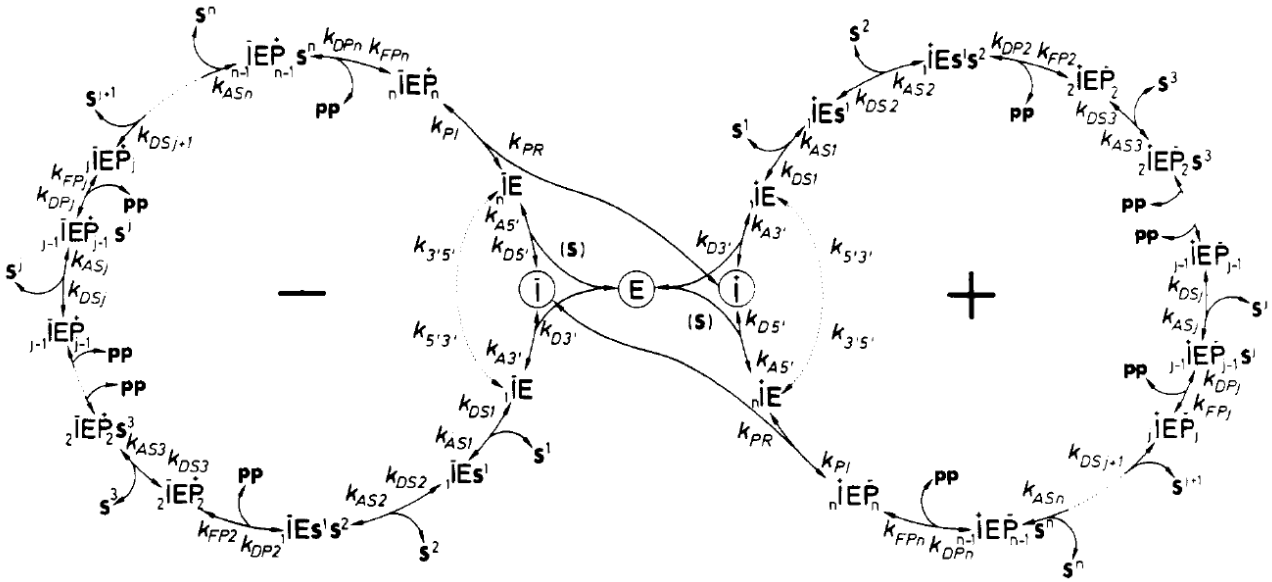
The increase in RNA production rate during a serial transfer experiment



RNA replication by Qβ-replicase

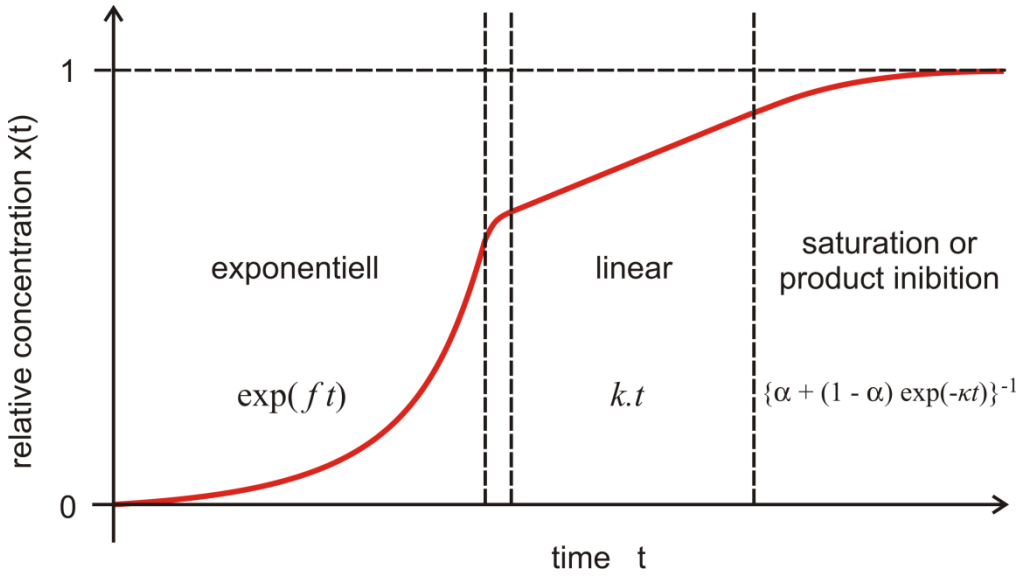
C. Weissmann, ***The making of a phage***.
FEBS Letters **40** (1974), S10-S18

Christof K. Biebricher,
1941-2009



Kinetics of RNA replication

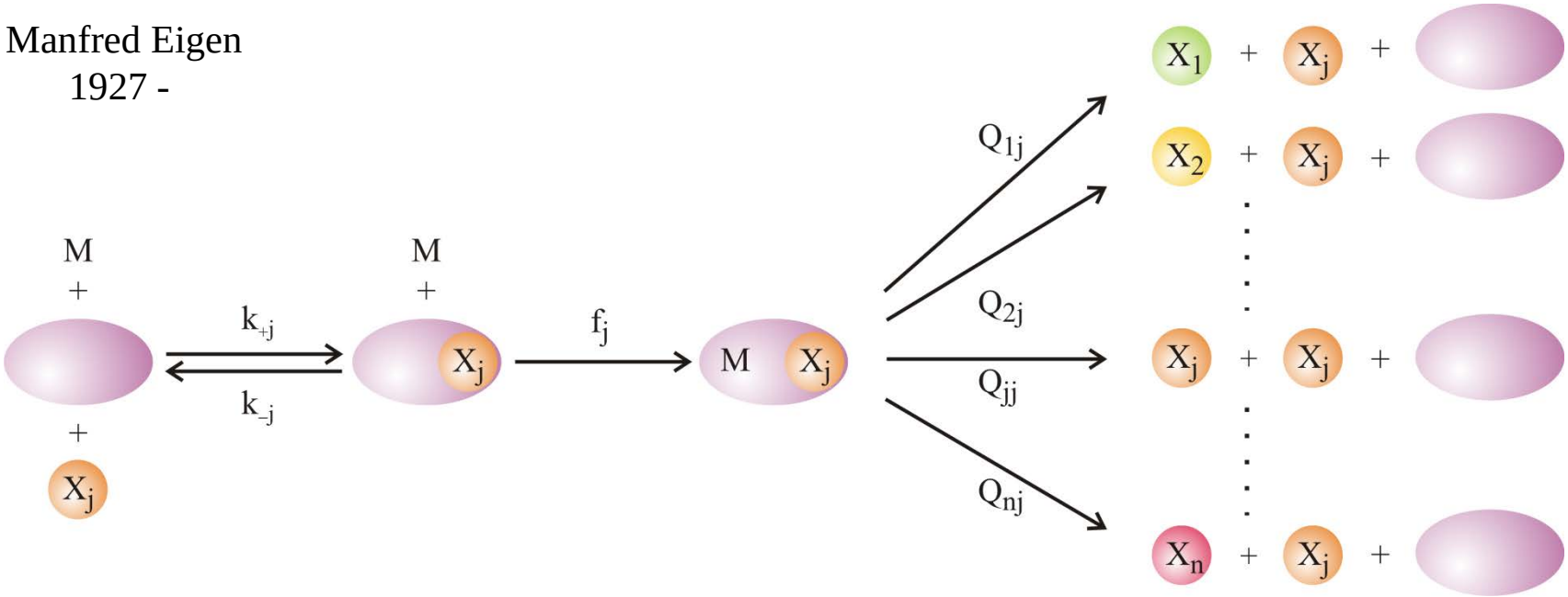
C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
Biochemistry **22**:2544-2559, 1983



Manfred Eigen
1927 -

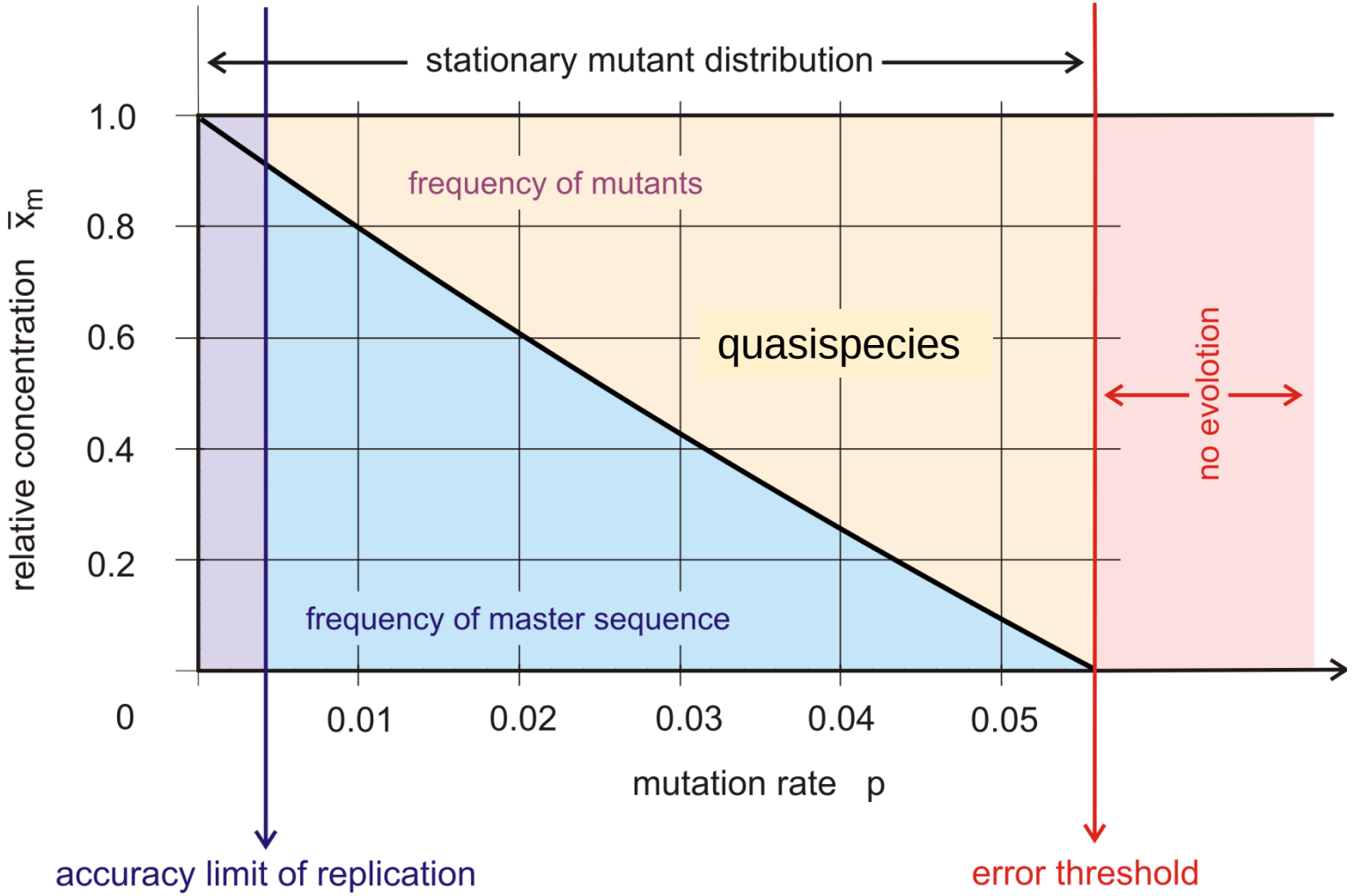
$$\frac{dx_i}{dt} = \sum_{j=1}^{n} Q_{ij} f_j x_j - x_i \Phi ; \; i = 1, 2, \ldots, n$$

$$\Phi = \sum_{j=1}^{n} f_j x_j ; \sum_{j=1}^{n} x_j = 1$$

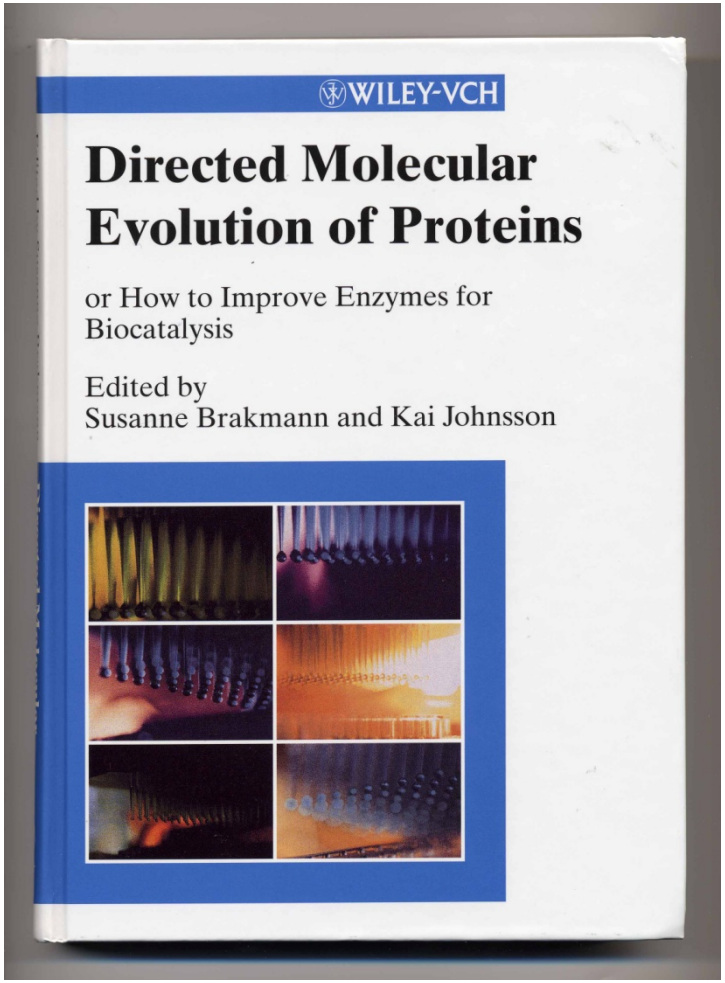


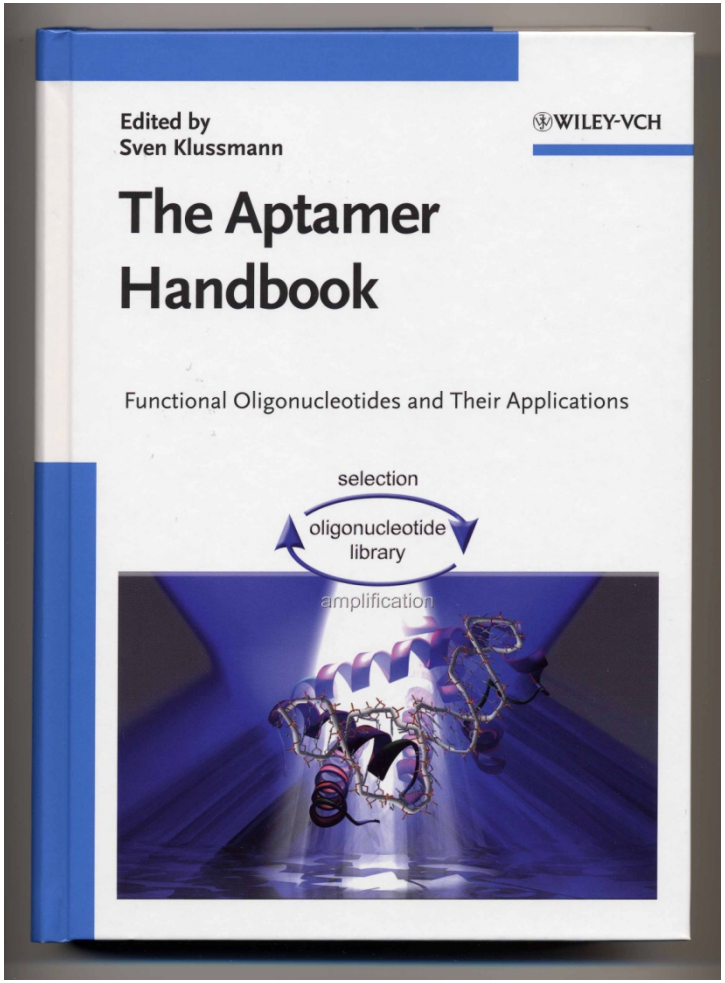
Mutation and (correct) replication as parallel chemical reactions

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



The error threshold in replication

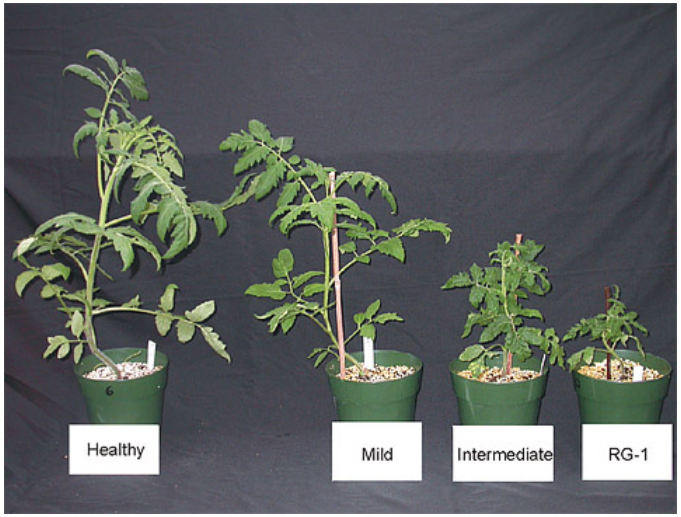




Application of molecular evolution to problems in biotechnology

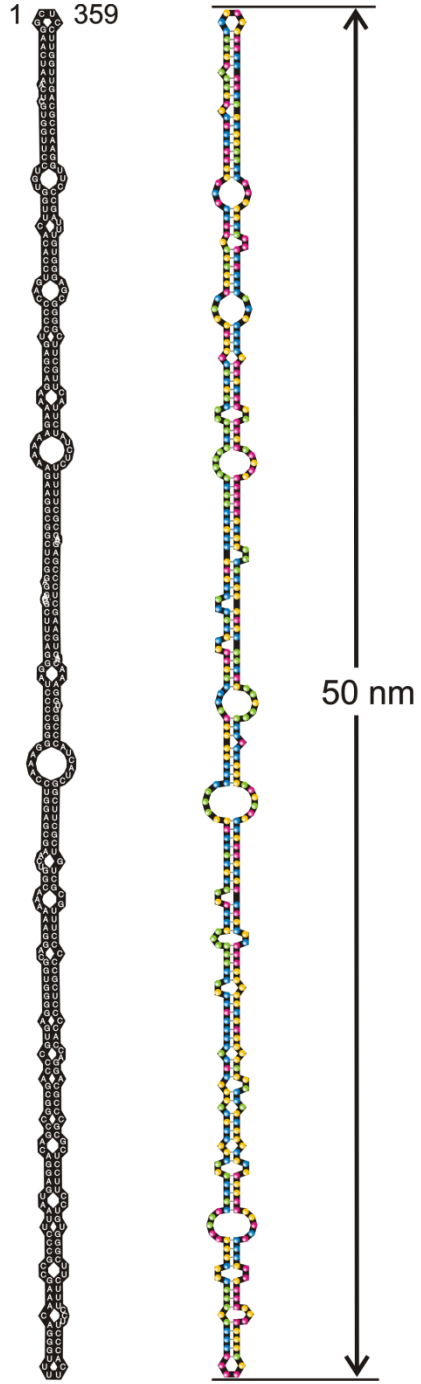
1. Darwin and mathematics
2. Digitalizing chemistry
3. Evolution in the test tube
4. **Viroids and viruses**
5. Global genotype evolution

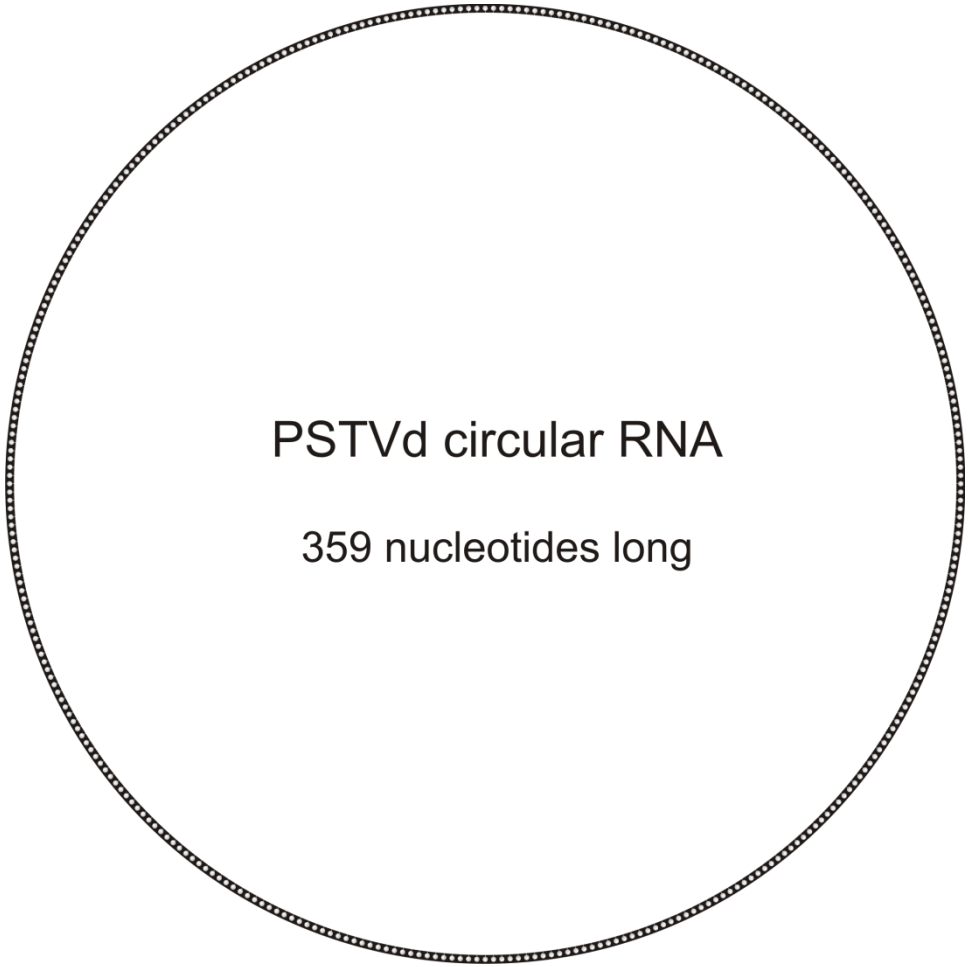
J. Demez. European and mediterranean plant protection organization archive. France



R.W. Hammond, R.A. Owens. Molecular Plant Pathology Laboratory, US Department of Agriculture

Plant damage by viroids





Adenine

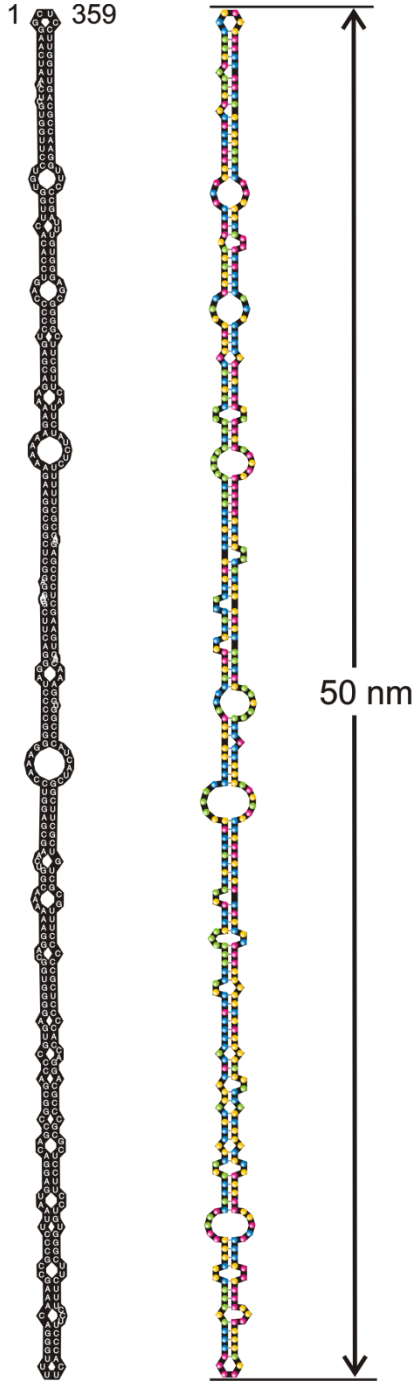
Uracil

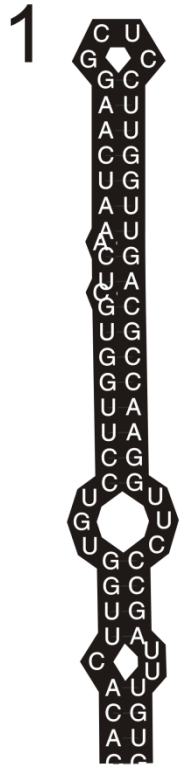
Guanine

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

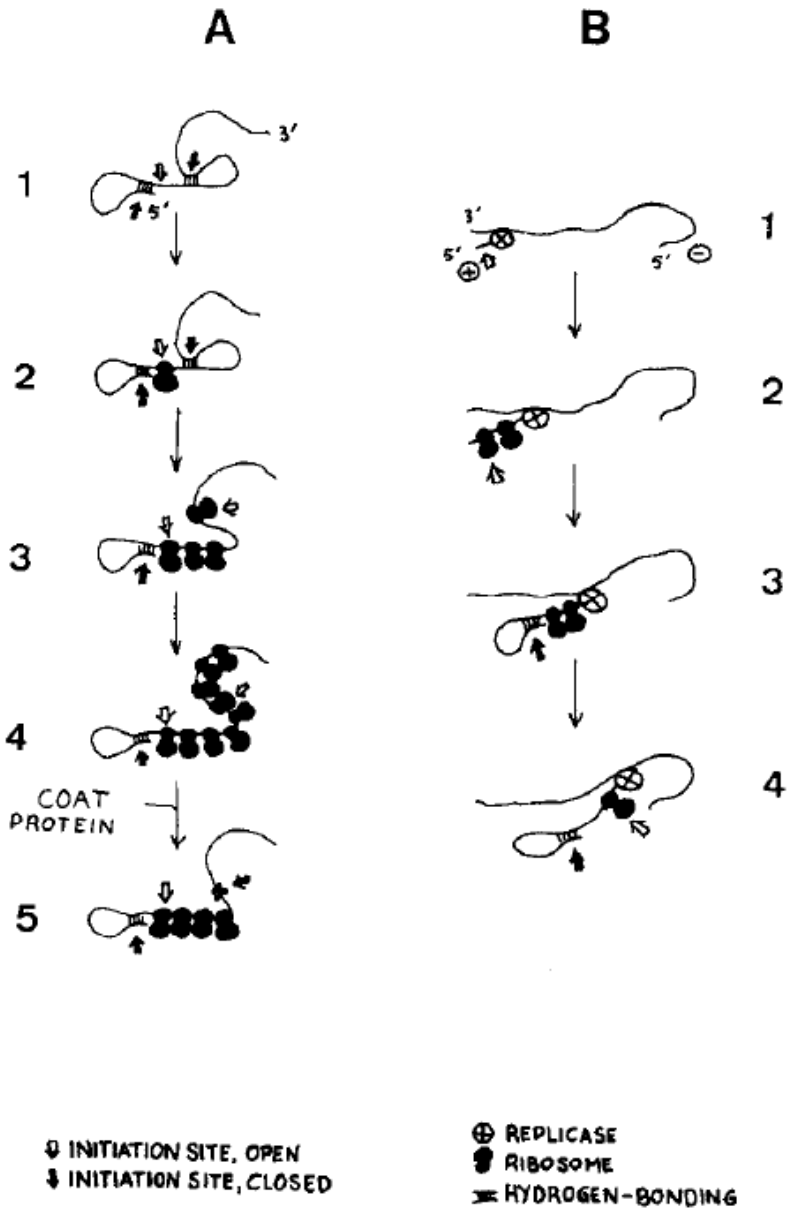
Biochemically supported structure

- 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)

Fig. 2. Translation on mature and nascent phage RNA. (A) Translation on mature RNA (1). Only the coat initiation site is accessible to ribosomes (2). As the coat cistron is translated, ribosomes can attach at the replicase cistron (3) giving rise to a polysome on which the coat and replicase, but not the maturation cistron are translated (4). During later stages of the infective cycle coat protein accumulates in the cell and binds to the RNA so as to block protein initiation at the replicase cistron (5). (B) Translation on nascent RNA. The viral replicase initiates synthesis of a plus strand at the 3′ end of a minus strand (1). When the ribosome binding site of the maturation (or A) protein has been formed, ribosomes attach and begin translation of this cistron (2). As plus strand synthesis progresses, the plus strand assumes a secondary structure which prevents access of ribosomes to the A cistron (3). At this point initiation of protein synthesis is now possible only at the coat cistron (4), as in the case of mature RNA (A). (See text for references).



Charles Weissmann. 1974. The Making of a Phage. FEBS Letters 40:S10 – S18.

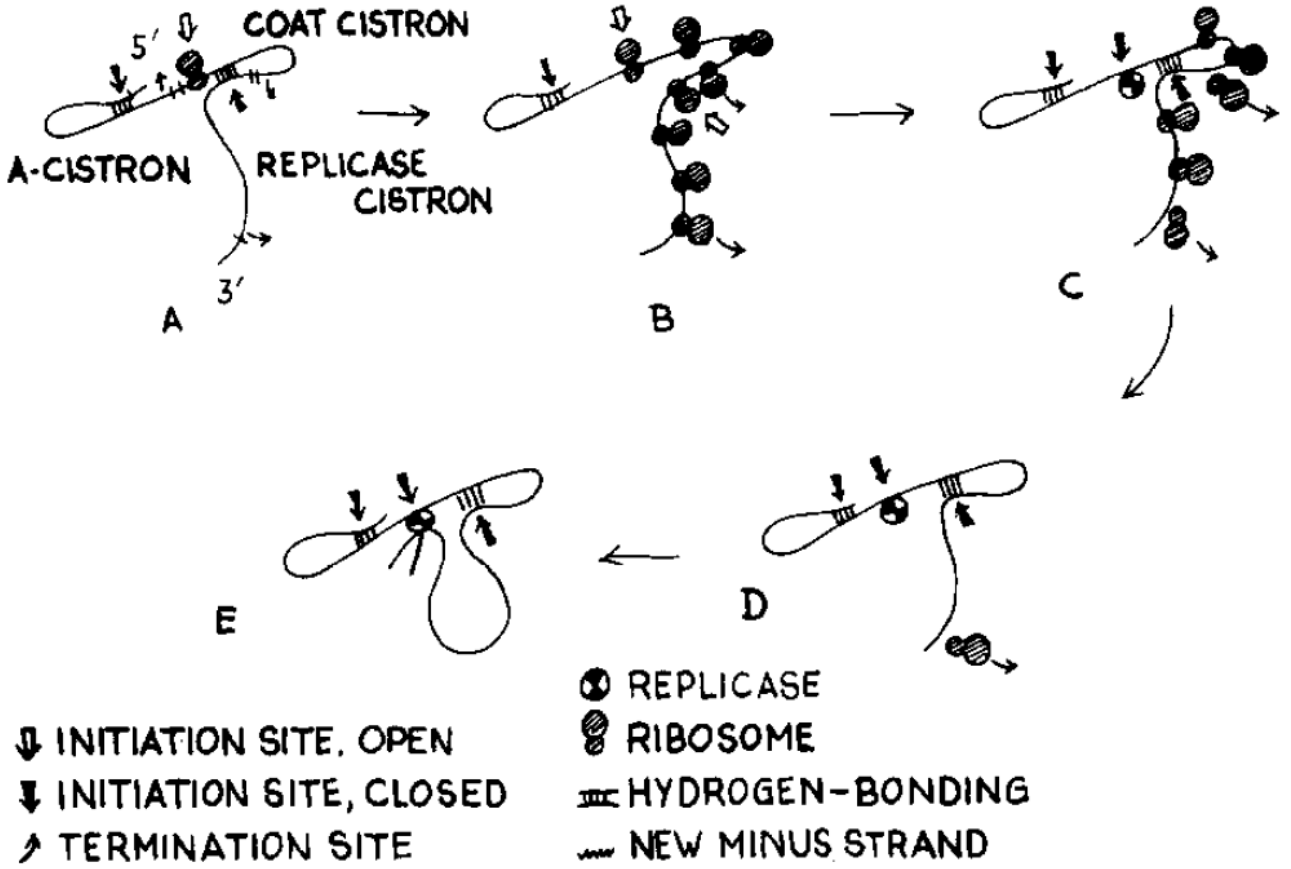


Fig. 3. Transition of phage RNA from polysome to replicating complex – repressor function of Qβ viral replicase.
(A) Ribosomes attach to the RNA at the coat initiation site. The initiation site of the replicase cistron is unavailable because of the secondary structure of the RNA. (B) Translation of the coat cistron ensues and the initiation site of the replicase cistron is exposed. The replicase cistron is translated. (C) When replicase becomes available, it attaches to the initiation site of the coat protein and blocks attachment of ribosomes in this position. The RNA refolds, preventing initiation at the replicase cistron. (D) The RNA is cleared of ribosomes. (E) Replicase can now attach to the 3′ terminus and initiate synthesis of the minus strand. The A cistron initiation site is at all times unaccessible to ribosomes because of the secondary structure of the mature RNA (cf. fig. 2) (from ref. [64]).

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 107 (2005) 115–116
Virus Research
www.elsevier.com/locate/virusres

Preface

Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations on enzymes involved in protein synthesis, as a theory of aging. In those times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be eliminated through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variation in differentiated organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to unveil the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Population geneticists rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparison of electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA recombination, and rapid nucleic acid sequencing techniques, molecular analyses of genomes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Most cellular processes, including such essential information-bearing and transferring events as genome replication, transcription and translation, are increasingly perceived as inherently inaccurate. Viruses, and in particular RNA viruses, are among the most extreme examples of exploitation of replication inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in quantitative terms as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Recently, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged as a possible antiviral strategy. This is the topic of the current special issue of ***Virus Research***.

Few would nowadays doubt that one of the major obstacles for the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principles that have shaped biological evolution over eons, that is, repeated rounds of reproduction with genetic variation, competition and selection, often perturbed by random events such as statistical fluctuations in population size. However, with viruses the consequences of the operation of these very same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be also observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens begs for alternative antiviral strategies, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect different lines of evidence (both theoretical and experimental) on which antiviral designs based on genetic deterioration inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted to progeny. Closely related to the theoretical developments have been numerous experimental studies on quasispecies dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of existing as mutant spectra rather than defined genetic entities, remarkably expand their potential to overcome selective pressures intended to limit their replication. Indeed, the use of antiviral inhibitors in clinical practice and the design of vaccines for a number of major RNA virus-associated diseases, are currently presided by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replicases, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy received an important impulse by the observation that ribavirin (a licensed antiviral nucleoside analogue) may be exerting, in some systems, its antiviral activity through enhanced mutagenesis. This has encouraged investigations on new mutagenic base analogues, some of them used in anticancer chemotherapy. Some chapters summarize these important biochemical studies on cell entry pathways and metabolism of mutagenic agents, that may find new applications as antiviral agents.

This volume intends to be basically a progress report, an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I can envisage (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific mutagens and inhibitors in an effective way, finding synergistic combinations and avoiding antagonistic ones as well as severe clinical side effects. (ii) Another on a deeper understanding of the metabolism of mutagenic agents, in particular base and nucleoside analogues. This includes identification of the transporters that carry them into cells, an understanding of their metabolic processing, intracellular stability and alterations of nucleotide pools, among other issues. (iii) Still another line of needed research is the development of new mutagenic agents specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with anticancer research, but others should result from the designs of new molecules, based on the structures of viral polymerases. I really hope that the reader finds this issue not only to be an interesting and useful review of the current situation in the field, but also a stimulating exposure to the major problems to be faced.

The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of ***Virus Research***, and then taken enthusiastically by Luis Enjuanes, recently appointed as Editor of ***Virus Research***. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of ***Virus Research***, Brian Mahy, for their continued interest and support to the research on virus evolution over the years.

My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
Universidad Autónoma de Madrid
Centro de Biología Molecular "Severo Ochoa"
Consejo Superior de Investigaciones Científicas
Cantoblanco and Valdeolmos
Madrid, Spain
Tel.: +34 91 497 84858/9; fax: +34 91 497 4799
E-mail address: edomingo@cbm.uam.es

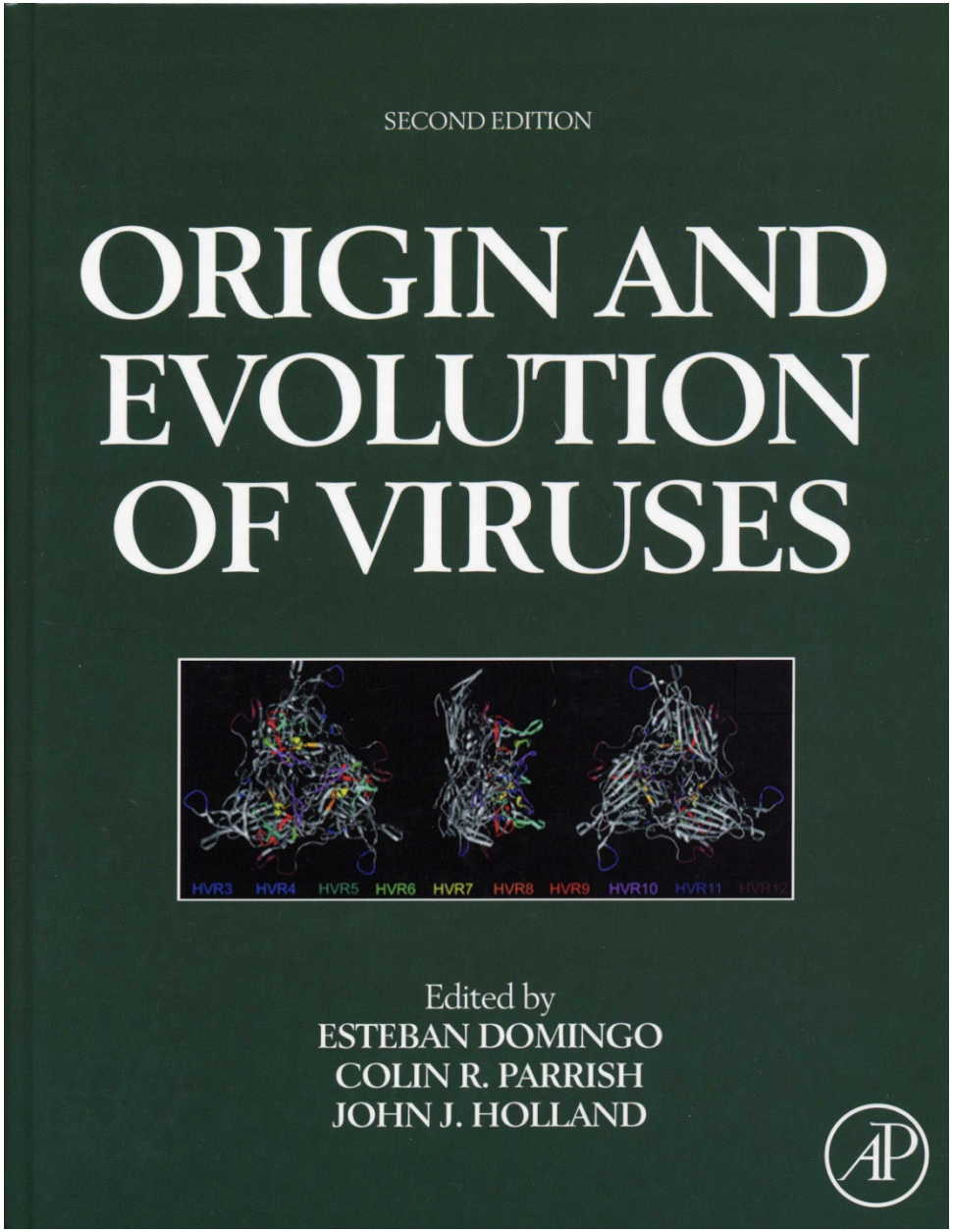
Available online 8 December 2004

doi:10.1016/j.virusres.2004.11.001

Esteban Domingo
1943 -

Application of quasispecies theory to the fight against viruses

Molecular evolution of viruses



Fitness landscapes are becoming accessible experimentally!

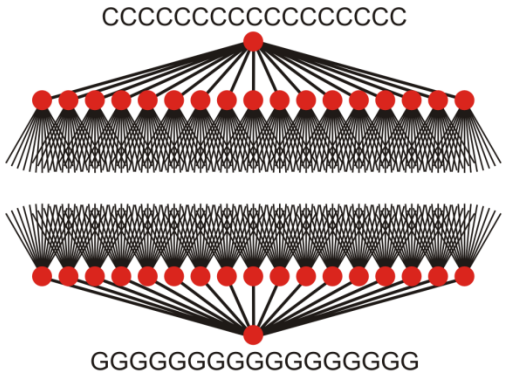
Protein landscapes: Yuuki Hayashi, Takuyo Aita, Hitoshi Toyota, Yuzuru Husimi, Itaru Urabe, Tetsuya Yomo. 2006. Experimental rugged fitness landscape in protein seqeunce space. *PLoS One* 1:e96.

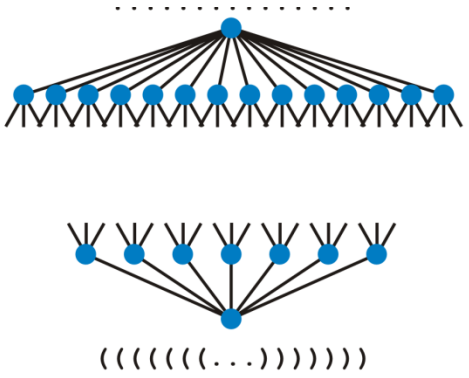
RNA landscapes: Sven Klussman, Ed. 2005. The aptamer handbook. Wiley-VCh, Weinheim (Bergstraße), DE.
Jason N. Pitt, Adrian Ferré-D'Amaré. 2010. Rapid construction of empirical RNA fitness landscapes. *Science* 330:376-379.

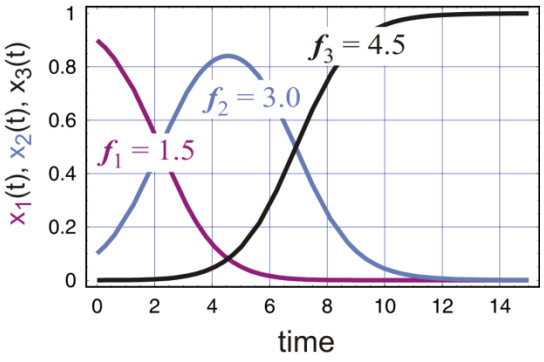
RNA viruses: Esteban Domingo, Colin R. Parrish, John J. Holland, Eds. 2007. Origin and evolution of viruses. Second edition. Elesvier, San Diego, CA.

Retroviruses: Roger D. Kouyos, Gabriel E. Leventhal, Trevor Hinkley, Mojgan Haddad, Jeannette M. Whitcomb, Christos J. Petropoulos, Sebastian Bonhoeffer. 2012. Exploring the complexity of the HIV-I fitness landscape. *PLoS Genetics* 8:e1002551

1. Darwin and mathematics
2. Digitalizing chemistry
3. Evolution in the test tube
4. Viroids and viruses
5. **Global genotype evolution**







sequence space | shape space | parameter space

$\mathcal{Q}$ | S | $\mathbb{R}_+$

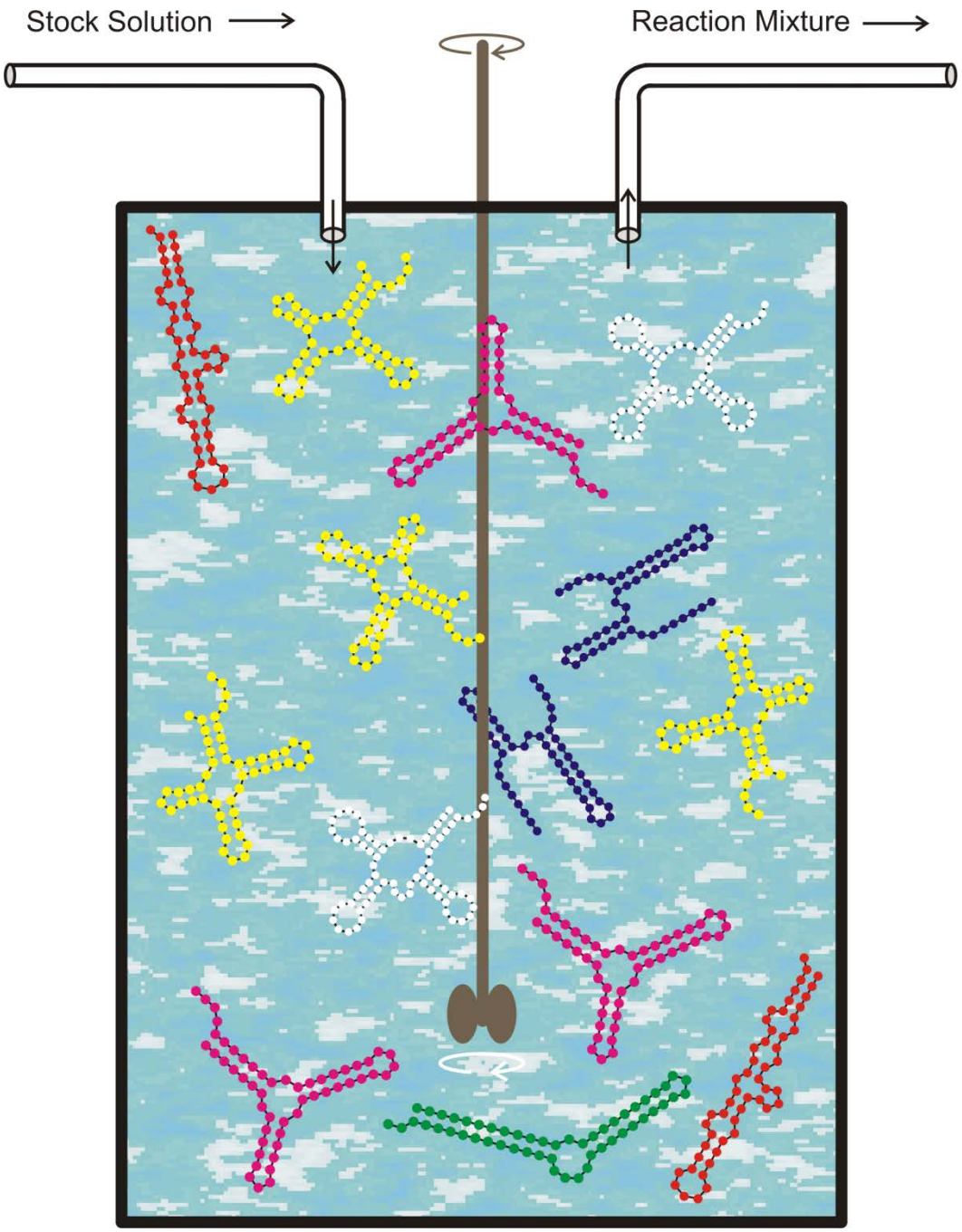
$\Phi: (\mathcal{Q}, d_H) \Rightarrow (S, d_S)$ $\quad$ $\Psi: (S, d_S) \Rightarrow \mathbb{R}_+$

$X \longrightarrow S = \Phi(X) \longrightarrow f = \Psi(S)$

sequence | structure | function

genotype $\longrightarrow$ **phenotype** $\longrightarrow$ **selection**

Evolution as a global phenomenon in genotype space



Replication rate constant

(Fitness):

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

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

Selection pressure:

The population size,

N = # RNA moleucles,

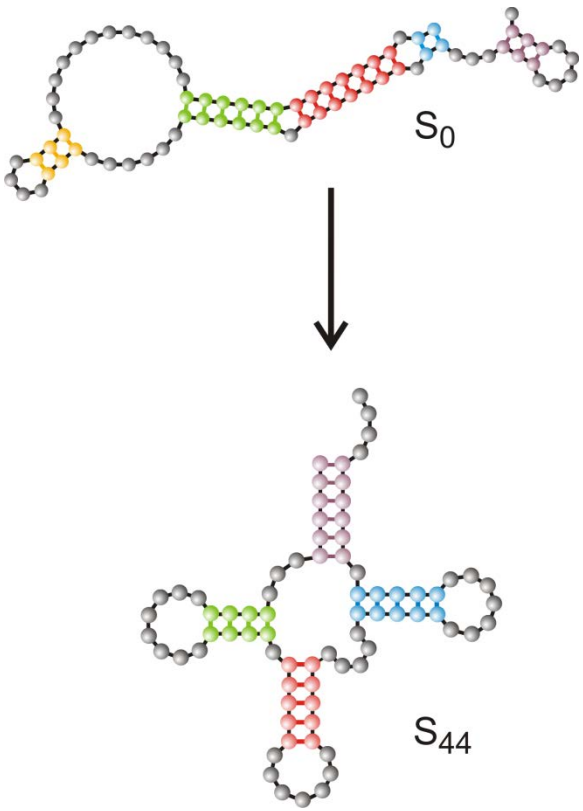
is determined by the flux:

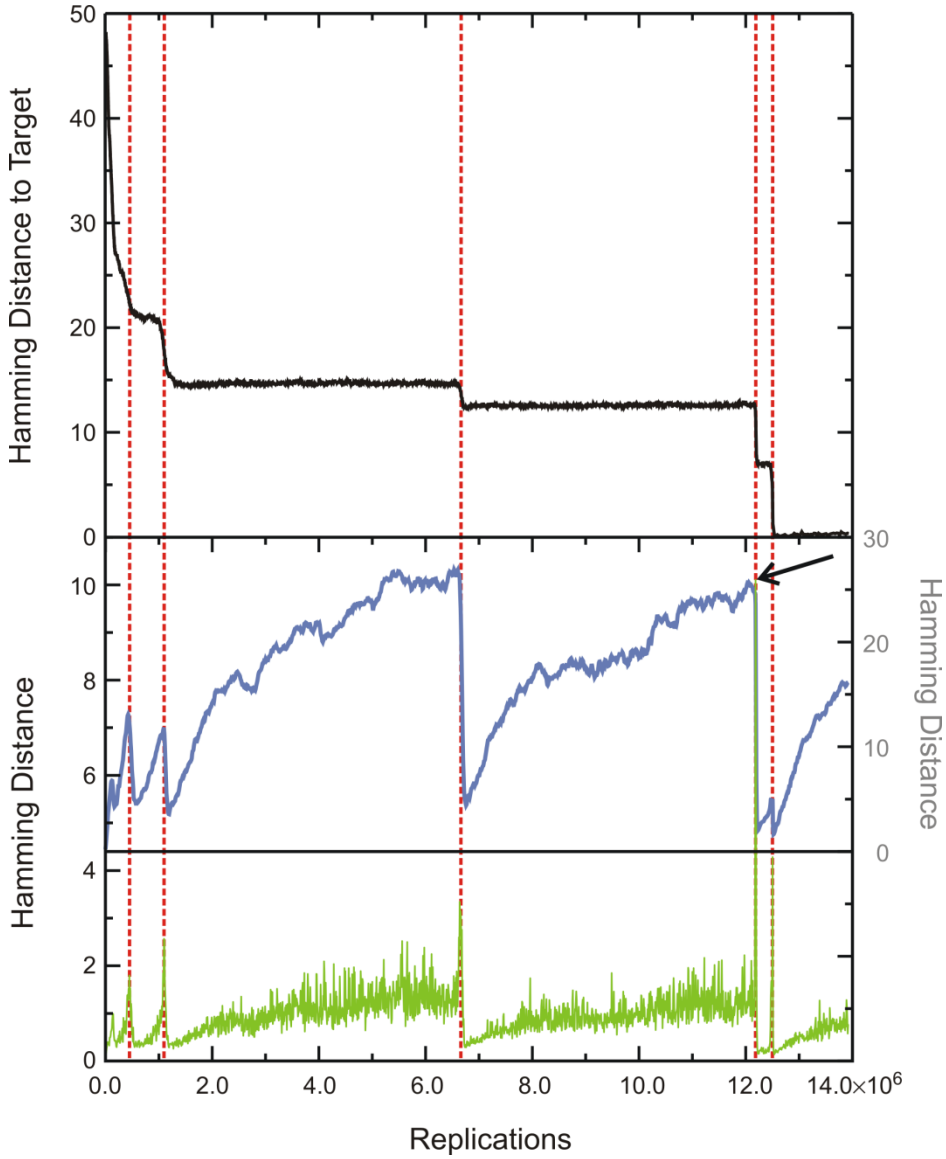
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

Mutation rate:

p = 0.001 / Nucleotide × Replication

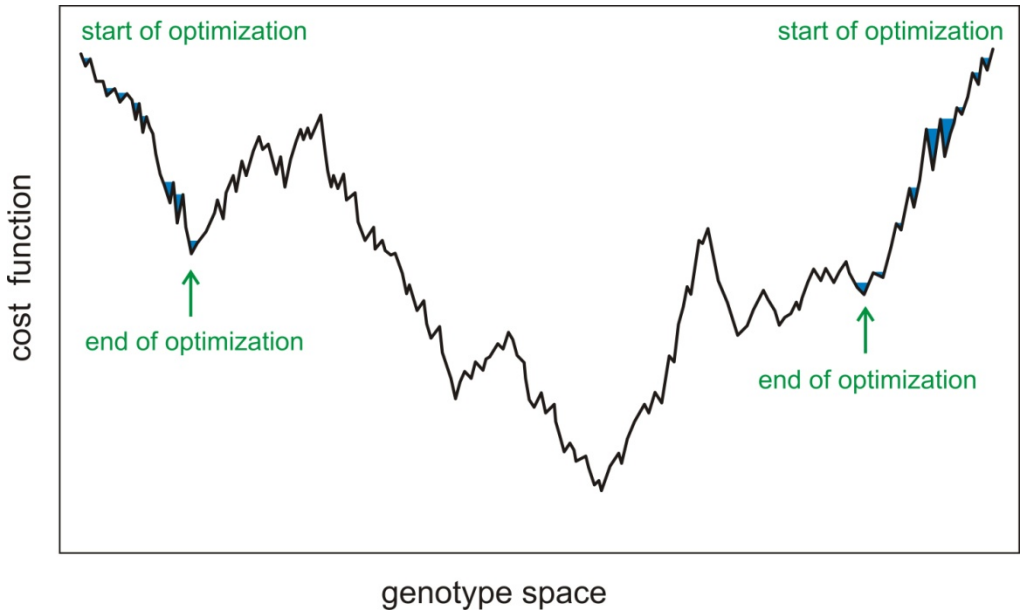
The flow reactor as a device for studying the evolution of molecules ***in vitro*** and ***in silico***.

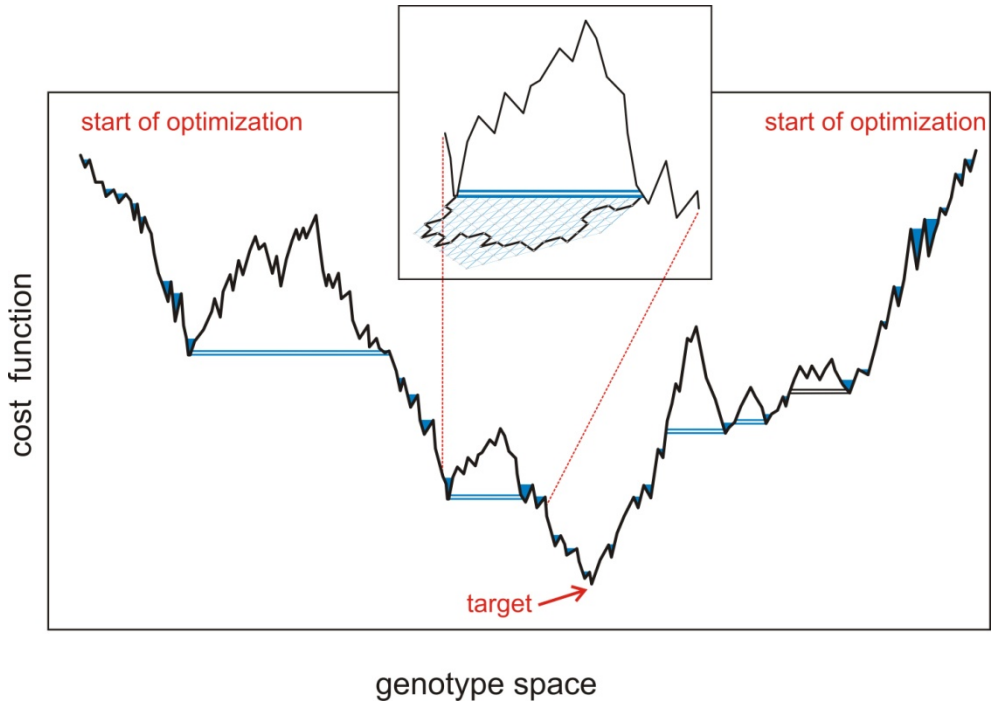




In silico optimization in the flow reactor: Evolutionary Trajectory

Optimization in populations living on rugged fitness landscapes



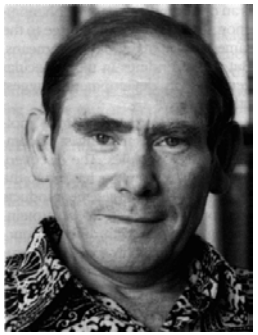


Advantages of the molecular approach

1. Complex reproduction mechanisms are readily included.

2. Gene regulation – DNA or RNA based – is chemical kinetics!

3. Accounting for epigenetic effects requires just the simultaneous consideration of several generations.



Sydney Brenner, 1927 -

What else is epigenetics than a funny form of enzymology ? Each protein, after all, comes from some piece of DNA.

Thank you for your attention!

Web-Page for further information:

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