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
Phenotypes

Charles Darwin, 1809 - 1882

Voyage on HMS Beagle, 1831 - 1836

1. Geospiza magnirostris
2. Geospiza fortis
3. Geospiza parvula
4. Certhidea olivacea

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.
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*

Leonardo da Pisa
~1180 – ~1240

Thomas Robert Malthus,
1766 – 1834

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

general progression

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

Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

Leonhard Euler, 1717 – 1783

The history of exponential growth
The chemistry and the mathematics of reproduction

Autocatalysis

\[ \frac{dx}{dt} = f(x)(1-x) \]

\[ A + X \rightarrow 2X \]

Competition

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

\[ \phi = \sum_{j=1}^{n} f_j x_j ; \sum_{j=1}^{n} x_j = 1 \]
Pierre-François Verhulst, 1804-1849

the consequence of finite resources

fitness values:
\[ f_1 = 2.80, \quad f_2 = 2.35, \quad f_3 = 2.25, \quad \text{and} \quad f_4 = 1.75 \]

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

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)
Digitalization of chemistry:
The unique assignment of nucleotides in base pairs.

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.

Digitalization of chemistry:
The unique assignment of nucleotides in base pairs.
An example from synthetic biology: Introduction of a third hydrogen bond into the $U = A$ base pair.
Hydrogen bonding patterns for Watson-Crick base pairs

The logics of DNA (or RNA) replication

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

The replication of DNA by Thermophilus aquaticus polymerase (PCR)

The logics of DNA (or RNA) replication
A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs

**Criterion:** Minimum free energy (mfe)

**Rules:** \( _\, (_\, _\, )\, _ \in \{AU,CG,GC,Gu,UA,UG\} \)

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
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:

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

RNA sample

Stock solution: Qβ RNA-replicase, ATP, CTP, GTP and UTP, buffer

Time

0 1 2 3 4 5 6 69 70
The increase in RNA production rate during a serial transfer experiment led to a decrease in mean fitness due to quasispecies formation.
RNA replication by Qβ-replicase

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

C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.

*Biochemistry* **22**:2544-2559, 1983
\[
\frac{dx_i}{dt} = \sum_{j=1}^{n} Q_{ij} f_j x_j - x_i \Phi; \quad i = 1, 2, \ldots, n
\]

\[
\Phi = \sum_{j=1}^{n} f_j x_j; \quad \sum_{j=1}^{n} x_j = 1
\]

Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
The error threshold in replication

The diagram illustrates the relationship between the relative concentration of the master sequence and the mutation rate, showing the transition from a quasispecies distribution to no evolution. The error threshold is indicated by the boundary where the frequency of mutants becomes significant enough to affect the population dynamics.
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
Plant damage by viroids

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

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

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).
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 inaccessible to ribosomes because of the secondary structure of the mature RNA (cf. fig. 2) (from ref. [64]).
Preface

Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations in enzymes involved in protein synthesis, as a theory of error. 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 evolution was 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 unravel 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 genetics, rare bacterial viruses or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species can be 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 reinforce this conclusion of genetic individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we use cellular RNAs, both pathogenic and exogenous, 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 replicase inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in 1968 as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Similarly, a conceptual extension of error catastrophe theory could be defined as "induced genetic determinant" has emerged as a possible antiviral strategy. This is the topic of the current special issue of *Viruses Research*.

Few would nowadays doubt that one of the major obstacles to the control of viral diseases is short-term adaptability of viral pathogens. Adaptability of viruses follows a similar Darwinian pattern as that described for RNA viruses. The consequences of this idea of a same Darwinian principle are seen as evolutionary changes in populations of such as statistical fluctuations in population size. However, with viruses the consequence of the operation of these same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens has made them targets for antiviral strategies, and focusing on the virus to cross the critical 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 crucial design based on genetic adaptation inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing mechanism for the adaptive 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 pressed by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replication, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy requires an important impact by the observation that ribonucleotides (a model of natural nucleoside analogues) may be not the ultimate adenylate or the reversal of some systems. in antiviral activity through enhanced expression.

This has encouraged investigations on new mutagenic base analogues, some of them in antiviral chemotherapies. 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 relate (not without many uncertainties) at least three lines of research (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific immunogens and inhibitors in an effective manner, 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 immunogens specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with antiviral research, but others should result from the development of new molecules, based on the structures of viral nucleosides. I truly 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 *Viruses Research*, and then taken enthusiastically by Luis Rejulanas, recently appointed as Editor of *Viruses Research*. I take this opportunity to thank Ulrich, Luis and the Editors-in-Chief of *Viruses Research*, Brian Milby, for their continued interest and support for the research on virus evolution over the years. My thanks 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. Alicia Hormillo from Centro de Biologia Molecular “Severo Ochoa” for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
Universidad Autonoma de Madrid
Centro de Biologia Molecular "Severo Ochoa"
Consejo Superior de Investigaciones Cientificas
Cantoblanco y Videolios
Madrid. Spain
Tel.: +34 91 497 4859; Fax: +34 91 497 4709
Email address: edomingo@ciba.uam.es
Available online 8 December 2004

Application of quasispecies theory to the fight against viruses
Molecular evolution of viruses
Fitness landscapes are becoming accessible experimentally!


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

sequence space $Q$

shape space $S$

parameter space $\mathbb{R}_+$

$\Phi: (Q, d_H) \Rightarrow (S, d_S)$

$\Psi: (S, d_S) \Rightarrow \mathbb{R}_+$

$f^* = \Psi(S)$

Evolution as a global phenomenon in genotype space
The flow reactor as a device for studying the evolution of molecules in vitro and in silico.

**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 molecules, is determined by the flux:

\[ N(t) \approx \bar{N} \pm \sqrt{N} \]

**Mutation rate:**

\( p = 0.001 / \text{Nucleotide} \times \text{Replication} \)

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.

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

Sydney Brenner, 1927 -
Thank you for your attention!
Web-Page for further information:

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