Evolution and Molecules
Basic questions of biology seen with physicists‘ eyes.

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Österreich
und
The Santa Fe Institute, Santa Fe, New Mexico, USA

401. Wilhelm and Else Heraeus Seminar
Evolution and Physics – Concepts, Models, and Applications
Bad Honnef, 21.– 23.01.2008
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. Simulation of stochastic phenomena
7. Biology in its full complexity
1. Replication and mutation

2. Quasispecies and error thresholds

3. Fitness landscapes and randomization

4. Lethal mutations

5. Ruggedness of natural landscapes

6. Simulation of stochastic phenomena

7. Biology in its full complexity
The three-dimensional structure of a short double helical stack of B-DNA


G \equiv C \text{ and } A = T
The mechanism of DNA replication is 'semi-conservative'.
Complementary replication is the simplest copying mechanism of RNA. Complementarity is determined by Watson-Crick base pairs:

\[ G = C \] and \[ A = U \]
Chemical kinetics of molecular evolution

Stock solution:
activated monomers, ATP, CTP, GTP, UTP (TTP);
a replicase, an enzyme that performs complementary replication;
buffer solution

Flow rate: \( r = \tau_R^{-1} \)

The population size \( N \), the number of polynucleotide molecules, is controlled by the flow \( r \)

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

The flowreactor is a device for studies of evolution in vitro and in silico.
Complementary replication as the simplest molecular mechanism of reproduction

\[
\frac{dx_1}{dt} = f_2 x_2 - x_1 \Phi \\
\frac{dx_2}{dt} = f_1 x_1 - x_2 \Phi
\]

\[
\Phi = \Sigma_i f_i x_i ; \quad \Sigma_i x_i = 1 ; \quad i = 1, 2
\]
Equation for complementary replication: \[ I_i = x_i \geq 0 , \quad f_i > 0 ; \quad i=1,2 \]

\[ \frac{dx_1}{dt} = f_2 x_2 - x_1 \phi , \quad \frac{dx_2}{dt} = f_1 x_1 - x_2 \phi , \quad \phi = f_1 x_1 + f_2 x_2 = f \]

Solutions are obtained by integrating factor transformation:

\[
x_{1,2}(t) = \frac{\sqrt{f_{2,1}}(\gamma_1(0) \cdot \exp(f t) + \gamma_2(0) \cdot \exp(-f t))}{(\sqrt{f_1} + \sqrt{f_2}) \gamma_1(0) \cdot \exp(f t) - (\sqrt{f_1} - \sqrt{f_2}) \gamma_1(0) \cdot \exp(-f t)}
\]

\[
\gamma_1(0) = \sqrt{f_1} x_1(0) + \sqrt{f_2} x_2(0) , \quad \gamma_2(0) = \sqrt{f_1} x_1(0) - \sqrt{f_2} x_2(0) , \quad f = \sqrt{f_1 f_2}
\]

\[ x_1(t) \rightarrow \frac{\sqrt{f_2}}{\sqrt{f_1} + \sqrt{f_2}} \quad \text{and} \quad x_2(t) \rightarrow \frac{\sqrt{f_1}}{\sqrt{f_1} + \sqrt{f_2}} \quad \text{as} \quad \exp(-f t) \rightarrow 0 \]
Reproduction of organisms or replication of molecules as the basis of selection

\[ \frac{dx_i}{dt} = f_i x_i - x_i \Phi = x_i (f_i - \Phi) \]

\[ \Phi = \sum_j f_j x_j; \quad \sum_j x_j = 1; \quad i,j = 1,2,\ldots,n \]

\[ [I_i] = x_i \geq 0; \quad i = 1,2,\ldots,n; \]

\[ [A] = a = \text{constant} \]

\[ f_m = \max \{f_j; j=1,2,\ldots,n\} \]

\[ x_m(t) \rightarrow 1 \text{ for } t \rightarrow \infty \]
Selection equation:  \[ [I_i] = x_i \geq 0, \ f_i > 0 \]

\[ \frac{dx_i}{dt} = x_i (f_i - \phi), \quad i=1,2,\cdots,n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \bar{f} \]

Mean fitness or dilution flux, \( \phi(t) \), is a non-decreasing function of time,

\[ \frac{d\phi}{dt} = \sum_{i=1}^{n} f_i \frac{dx_i}{dt} = \bar{f}^2 - \left( \frac{\bar{f}}{n} \right)^2 = \text{var}\{f\} \geq 0 \]

Solutions are obtained by integrating factor transformation

\[ x_i(t) = \frac{x_i(0) \cdot \exp(f_i t)}{\sum_{j=1}^{n} x_j(0) \cdot \exp(f_j t)}; \quad i = 1,2,\cdots,n \]
Selection between three species with $f_1 = 1$, $f_2 = 2$, and $f_3 = 3$
Variation of genotypes through mutation and recombination
* \rightarrow a_0 r \rightarrow A

A \quad + \quad X_j \quad \rightarrow \quad k_j Q_{jj} \rightarrow X_j \quad + \quad X_j

\vdots

A \quad + \quad X_j \quad \rightarrow \quad k_j Q_{nj} \rightarrow X_n \quad + \quad X_j

X_j \quad \rightarrow \quad r \rightarrow \emptyset

A \quad \rightarrow \quad r \rightarrow \emptyset

j = 1, 2, \ldots, n
\[ \begin{align*}
\ast & \xrightarrow{a_0 \cdot r} A \\
A + X_i & \xrightarrow{k_i Q_{ii}} 2X_i, \; i = 1, \ldots, n \quad \text{: replication} \\
A + X_i & \xrightarrow{k_i Q_{ji}} X_i + X_j, \; i, j = 1, \ldots, n; \; i \neq j \quad \text{: mutation} \\
A & \xrightarrow{r} \emptyset \quad \text{: outflux} \\
X_j & \xrightarrow{r} \emptyset, \; j = 1, \ldots, n \quad \text{: outflux}
\end{align*} \]

\[
\frac{da}{dt} = -a \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \cdot Q_{ji} \cdot x_i + r \left( a_0 - a \right) = -a \sum_{i=1}^{n} k_i \cdot x_i + r \left( a_0 - a \right)
\]

\[
\frac{dx_j}{dt} = a \sum_{i=1}^{n} k_i \cdot Q_{ji} \cdot x_i - r \cdot x_j
\]

Origin of the replication-mutation equation from the flowreactor
Stationary solutions of the flow reactor:

\[
\frac{da}{dt} = 0 = -\tilde{a} \left( \sum_{i=1}^{n} k_i \tilde{x}_i + r \right) + r \tilde{a}
\]

\[
\frac{dx_j}{dt} = 0 = \tilde{a} \sum_{i=1}^{n} k_i Q_{ji} \tilde{x}_i - r \tilde{x}_j; \quad c = \sum_{i=1}^{n} x_i; \quad \bar{k} = \frac{\sum_{i=1}^{n} k_i x_i}{c}
\]

\[
\frac{dc}{dt} = 0 = \tilde{c} \left( \bar{k} \tilde{a} - r \right)
\]

Stationary solutions: 1. active state

\[
r < \bar{k} a_0
\]

\[
\tilde{a} = \frac{r}{\bar{k}}
\]

\[
\tilde{c} = \frac{\bar{k} a_0 - r}{\bar{k}}
\]

Stationary solutions: 2. extinction

\[
r > \bar{k} a_0
\]

\[
\tilde{a} = a_0
\]

\[
\tilde{x}_j = 0; \quad j = 1, 2, \ldots, n
\]
\[ A + X \rightarrow 2X \]
Find \( r(t) \) such that \( a(t) = \bar{a} = \text{const.} \)

\[
\frac{da}{dt} = 0 = -\bar{a} \sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij} Q_{ji} x_i + r(t) (a_0 - \bar{a})
\]

\[
\begin{align*}
 r(t) &= \frac{\bar{a}}{a_0 - \bar{a}} \sum_{i=1}^{n} k_i x_i; \quad f_i = k_i \bar{a}
\end{align*}
\]

\[
\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i Q_{ji} x_i - x_j \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} x_i} = \sum_{i=1}^{n} f_i Q_{ji} x_i - x_j \bar{f}
\]

Origin of the replication-mutation equation from the flowreactor
1. Replication and mutation

2. Quasispecies and error thresholds

3. Fitness landscapes and randomization

4. Lethal mutations

5. Ruggedness of natural landscapes

6. Simulation of stochastic phenomena

7. Biology in its full complexity
Chemical kinetics of replication and mutation as parallel reactions
\[
\frac{dx_i}{dt} = \sum_{i=1}^{n} f_i Q_{ij} x_i - x_j \Phi \quad \text{with} \quad \Phi = \sum_{i=1}^{n} f_i x_i
\]

and \( \sum_{i=1}^{n} x_i = 1 \)

\[
Q_{ij} = (1 - p)^{n - d_H(X_i, X_j)} \cdot p^{d_H(X_i, X_j)}; \quad p \ldots \text{error rate per digit}
\]

\( d_H(X_i, X_j) \ldots \text{Hamming distance between } X_i \text{ and } X_j \)

\[
\sum_{j=1}^{n} Q_{ij} = 1
\]

The replication-mutation equation
Mutation-selection equation: 

\[
[I_i] = x_i \geq 0, \quad f_i > 0, \quad Q_{ij} \geq 0
\]

\[
\frac{dx_i}{dt} = \sum_{j=1}^{n} f_j Q_{ji} x_j - x_i \phi, \quad i = 1, 2, \ldots, n; \quad \sum_{i=1}^{n} x_i = 1; \quad \phi = \sum_{j=1}^{n} f_j x_j = \bar{f}
\]

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

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

\[
W \div \{ f_i Q_{ij}; \; i, j = 1, 2, \ldots, n \}; \quad L = \{ \ell_{ij}; \; i, j = 1, 2, \ldots, n \}; \quad L^{-1} = H = \{ h_{ij}; \; i, j = 1, 2, \ldots, n \}
\]

\[
L^{-1} \cdot W \cdot L = \Lambda = \{ \lambda_k; \; k = 0, 1, \ldots, n-1 \}
\]
Matrix $W$ and Frobenius theorem:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

Primitive matrix $W$:

A nonnegative square matrix $W = \{w_{ij}\}$ is said to be a primitive matrix if there exists $k$ such that $W^k \gg 0$, i.e., if there exists $k$ such that for all $i, j$, the $(i, j)$ entry of $W^k$ is positive.
Perron-Frobenius theorem applied to the value matrix \( W \)

\( W \) is primitive: (i) \( \lambda_0 \) is real and strictly positive

(ii) \( \lambda_0 > |\lambda_k| \) for all \( k \neq 0 \)

(iii) \( \lambda_0 \) is associated with strictly positive eigenvectors

(iv) \( \lambda_0 \) is a simple root of the characteristic equation of \( W \)

(v-vi) etc.

\( W \) is irreducible: (i), (iii), (iv), etc. as above

(ii) \( \lambda_0 \geq |\lambda_k| \) for all \( k \neq 0 \)
Decomposition of matrix $W$

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} = Q \cdot F \text{ with}$$

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nn} \end{pmatrix} \text{ and } F = \begin{pmatrix} f_1 & 0 & \cdots & 0 \\ 0 & f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_n \end{pmatrix}$$
Uniform error rate model:

\[ Q_{ij} = p^{d_H(x_i, x_j)} (1 - p)^{(n - d_H(x_i, x_j))} \]

\[ d_H(x_i, x_j) \ldots \text{Hamming distance} \]
Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space

Master sequence

Mutant cloud

Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space
Formation of a quasispecies in sequence space

Master sequence

Mutant cloud

Formation of a quasispecies in sequence space
Mutant cloud

Uniform distribution in sequence space
Quasispecies as a function of the replication accuracy $q$
Chain length and error threshold

\[ Q \cdot \sigma = (1 - p)^n \cdot \sigma \geq 1 \Rightarrow n \cdot \ln(1-p) \geq -\ln \sigma \]

\[ n \text{ ... constant} : \quad p_{\text{max}} \approx \frac{\ln \sigma}{n} \]

\[ p \text{ ... constant} : \quad n_{\text{max}} \approx \frac{\ln \sigma}{p} \]

\[ Q = (1 - p)^n \text{ ... replication accuracy} \]

\[ p \quad \text{ ... error rate} \]

\[ n \quad \text{ ... chain length} \]

\[ \sigma = \frac{f_m}{(1 - x_m) \sum_{j \neq m} f_j} \quad \text{... superiority of master sequence} \]
The error threshold in replication
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. Simulation of stochastic phenomena
7. Biology in its full complexity
Fitness landscapes showing error thresholds
Every point in sequence space is equivalent

Sequence space of binary sequences with chain length \( n = 5 \)

Binary sequences can be encoded by their decimal equivalents:

\[ C = 0 \quad \text{and} \quad G = 1, \] for example,

\[ "0" \equiv 00000 = \text{CCCCC}, \]

\[ "14" \equiv 01110 = \text{CGGGC}, \]

\[ "29" \equiv 11101 = \text{GGGCG}, \] etc.
Error threshold: Error classes and individual sequences

\[ n = 10 \text{ and } \sigma = 2 \]
Error threshold: Individual sequences

\( n = 10, \sigma = 2 \) and \( d = 0, 1.0, 1.85 \)
Error threshold: Error classes and individual sequences

\[ n = 10 \text{ and } \sigma = 1.1 \]
Error threshold: Individual sequences

\( n = 10, \sigma = 1.1, d = 1.95, 1.975, 2.00 \) and seed = 877
Error threshold: Individual sequences

\( n = 10, \sigma = 1.1, d = 1.975, \) and seed = 877, 637, 491
STATIONARY MUTANT DISTRIBUTIONS AND EVOLUTIONARY OPTIMIZATION

Peter Schuster and Jörg Swetina
Institut für theoretische Chemie
und Strahlenchemie der Universität Wien,
Währingerstraße 17,
A 1090 Wien,
Austria

Molecular evolution is modeled by erroneous replication of binary sequences. We show how the selection of two species of equal or almost equal selective value is influenced by its nearest neighbors in sequence space. In the case of perfect neutrality and sufficiently small error rates we find that the Hamming distance between the species determines selection. As the error rate increases the fitness parameters of neighboring species become more and more important. In the case of almost neutral sequences we observe a critical replication accuracy at which a drastic change in the "quasispecies", in the stationary mutant distribution occurs. Thus, in frequently mutating populations fitness turns out to be an ensemble property rather than an attribute of the individual.

In addition we investigate the time dependence of the mean excess production as a function of initial conditions. Although it is optimized under most conditions, cases can be found which are characterized by decrease or non-monotonic change in mean excess productions.

1. Introduction. Recent data from populations of RNA viruses provided direct evidence for vast sequence heterogeneity (Domingo et al., 1987). The origin of this diversity is not yet completely known. It may be caused by the low replication accuracy of the polymerizing enzyme, commonly a virus specific, RNA dependent RNA synthetase, or it may be the result of a high degree of selective neutrality of polynucleotide sequences. Eventually, both factors contribute to the heterogeneity observed. Indeed, mutations occur much more frequently than previously assumed in microbiology. They are by no means rare events and hence, neither the methods of conventional population genetics (Ewens, 1979) nor the neutral theory (Kimura, 1983) can be applied to these virus populations. Selectively neutral variants may be close with respect to Hamming distance and then the commonly made assumption that the mutation backflow from the mutants to the wild type is negligible does not apply.

A kinetic theory of polynucleotide evolution which was developed during the past 15 years (Eigen, 1971; 1985; Eigen and Schuster, 1979; Eigen et al., 1987; Schuster, 1986; Schuster and Sigmund, 1985) treats correct replication and mutation as parallel reactions within one and the same reaction network.
Fitness landscapes \textbf{not} showing error thresholds
Error thresholds and gradual transitions

$n = 20$ and $\sigma = 10$
(1) linear \[ f^1_{\text{scale}}(d) = 100(1 - d/l), \]
(2) exponential \[ f^2_{\text{scale}}(d) = 100^{1-d/l}, \]
(3) rational \[ f^3_{\text{scale}}(d) = \frac{1}{0.01 + d/l}, \]
(4) sigmoid \[ f^4_{\text{scale}}(d) = 100^{1-(d/l)^\sigma}, \]
(5) inverse \[ f^5_{\text{scale}}(d) = 100 - 100^{d/l} + 1. \]
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. Simulation of stochastic phenomena
7. Biology in its full complexity
\[ a_0 r \rightarrow A \]

\[ k_1 Q_{11} \]

\[ k_1 Q_{21} \]

\[ k_1 Q_{j1} \]

\[ k_1 Q_{n1} \]

\[ X_1 + X_1 \]

\[ X_2 + X_1 \]

\[ \cdots \]

\[ X_j + X_1 \]

\[ \cdots \]

\[ X_n + X_1 \]

\[ X_j \rightarrow r \rightarrow \emptyset \]

\[ A \rightarrow r \rightarrow \emptyset \]

\[ j = 1, 2, \ldots, n \]
Lethal mutants and Frobenius theorem:

\[
W = \begin{pmatrix}
  w_{11} & 0 & \ldots & 0 \\
  w_{21} & 0 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{n1} & 0 & \ldots & 0 \\
\end{pmatrix} = w_{11} \begin{pmatrix}
  1 & 0 & \ldots & 0 \\
  \frac{w_{21}}{w_{11}} & 0 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{w_{n1}}{w_{11}} & 0 & \ldots & 0 \\
\end{pmatrix}
\]

\[
W^k = w_{11}^k \begin{pmatrix}
  1 & 0 & \ldots & 0 \\
  \frac{w_{21}}{w_{11}} & 0 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{w_{n1}}{w_{11}} & 0 & \ldots & 0 \\
\end{pmatrix}
\]
\[
* \xrightarrow{a_0 \cdot r} A \quad : \text{influx}
\]

\[
A + X_1 \xrightarrow{k_1 Q_{11}} 2X_1 \quad : \text{replication}
\]

\[
A + X_1 \xrightarrow{k_1 Q_{j1}} X_1 + X_j \quad j = 2, \ldots, n \quad : \text{mutation}
\]

\[
A \xrightarrow{r} \emptyset \quad : \text{outflux}
\]

\[
X_j \xrightarrow{r} \emptyset \quad j = 1, \ldots, n \quad : \text{outflux}
\]

\[
\frac{da}{dt} = -a \sum_{j=1}^{n} k_1 Q_{j1} x_1 + r (a_0 - a) = -a k_1 x_1 + r (a_0 - a)
\]

\[
\frac{dx_j}{dt} = a Q_{j1} x_1 - r x_j
\]
Stationary solutions: 1. active state

\[ r < k_1 Q_{11} a_0 \]

\[ \tilde{a} = \frac{r}{k_1 Q_{11}} \]

\[ \tilde{x}_1 = Q_{11} (a_0 - \tilde{a}) = Q_{11} a_0 - \frac{r}{k_1} \]

\[ \tilde{x}_j = Q_{j1} (a_0 - \tilde{a}) = Q_{j1} \left( a_0 - \frac{r}{k_1 Q_{11}} \right); \quad j = 2, 3, \ldots, n \]

Stationary solutions: 2. extinction

\[ r > k_1 Q_{11} a_0 \]

\[ \tilde{a} = a_0 \]

\[ \tilde{x}_j = 0; \quad j = 1, 2, \ldots, n \]
Find \( r(t) \) such that \( a(t) = \bar{a} = \text{const.} \)

\[
\frac{da}{dt} = 0 = -\bar{a} \sum_{j=1}^{n} k_1 Q_{j1} x_1 + r(t) (a_0 - \bar{a})
\]

\[
 r(t) = \frac{\bar{a}}{a_0 - \bar{a}} k_1 x_1; \quad f_1 = k_1 \bar{a}; \quad \sum_{i=1}^{n} x_i = c = a_0 - \bar{a}
\]

\[
\frac{dx_j}{dt} = f_1 Q_{j1} x_1 - x_j \frac{f_1 x_1}{\sum_{i=1}^{n} x_i} = f_1 x_1 \left( Q_{j1} - \frac{x_j}{c} \right)
\]

Stationary solutions:

\[
\bar{x}_j = Q_{j1} \sum_{i=1}^{n} \bar{x}_i = Q_{ji} c
\]
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. Simulation of stochastic phenomena
7. Biology in its full complexity
Definition of RNA structure
A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs
The surrounding of **GUCAUAUCAG** in sequence space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space

GGCUAUCGUACGUUUACCCAAAGUCUACGUGGACCCAGGCAUUGGACG
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
One error neighborhood – Surrounding of an RNA molecule in sequence and shape space
<table>
<thead>
<tr>
<th>Number</th>
<th>Mean Value</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50125</td>
<td>0.334167</td>
<td>0.006961</td>
</tr>
<tr>
<td>2</td>
<td>2856</td>
<td>0.019040</td>
<td>0.006687</td>
</tr>
<tr>
<td>3</td>
<td>2799</td>
<td>0.018660</td>
<td>0.006687</td>
</tr>
<tr>
<td>4</td>
<td>2417</td>
<td>0.016113</td>
<td>0.005733</td>
</tr>
<tr>
<td>5</td>
<td>2265</td>
<td>0.015100</td>
<td>0.005333</td>
</tr>
<tr>
<td>6</td>
<td>2233</td>
<td>0.014887</td>
<td>0.004933</td>
</tr>
<tr>
<td>7</td>
<td>1442</td>
<td>0.009613</td>
<td>0.004633</td>
</tr>
<tr>
<td>8</td>
<td>1081</td>
<td>0.007207</td>
<td>0.003733</td>
</tr>
<tr>
<td>9</td>
<td>1025</td>
<td>0.006833</td>
<td>0.003733</td>
</tr>
<tr>
<td>10</td>
<td>1003</td>
<td>0.006687</td>
<td>0.003433</td>
</tr>
<tr>
<td>11</td>
<td>963</td>
<td>0.006420</td>
<td>0.003233</td>
</tr>
<tr>
<td>12</td>
<td>860</td>
<td>0.005733</td>
<td>0.003033</td>
</tr>
<tr>
<td>13</td>
<td>800</td>
<td>0.005333</td>
<td>0.002833</td>
</tr>
<tr>
<td>14</td>
<td>548</td>
<td>0.003653</td>
<td>0.002633</td>
</tr>
<tr>
<td>15</td>
<td>362</td>
<td>0.002413</td>
<td>0.002433</td>
</tr>
<tr>
<td>16</td>
<td>337</td>
<td>0.002247</td>
<td>0.002233</td>
</tr>
<tr>
<td>17</td>
<td>241</td>
<td>0.001607</td>
<td>0.002133</td>
</tr>
<tr>
<td>18</td>
<td>231</td>
<td>0.001540</td>
<td>0.002033</td>
</tr>
<tr>
<td>19</td>
<td>225</td>
<td>0.001500</td>
<td>0.001933</td>
</tr>
<tr>
<td>20</td>
<td>202</td>
<td>0.001347</td>
<td>0.001833</td>
</tr>
</tbody>
</table>

Shadow – Surrounding of an RNA structure in shape space – **AUGC** alphabet
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. **Simulation of stochastic phenomena**
7. Biology in its full complexity
Phenylalanyl-tRNA as target structure

Structure of randomly chosen initial sequence

Phenylalanyl-tRNA as target structure
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 the two concepts is used. Such a relation is based on the probability of one phenotype being accessible from another through changes in the genotype. This notion 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.

Evolution in silico
W. Fontana, P. Schuster,
Science 280 (1998), 1451-1455
Evolution of RNA molecules as a Markov process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markov process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markov process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series
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, \( N = \# \text{RNA molecules} \), 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
28 neutral point mutations during a long quasi-stationary epoch

Transition inducing point mutations change the molecular structure
Neutral point mutations leave the molecular structure unchanged

Neutral genotype evolution during phenotypic stasis
Randomly chosen initial structure

Phenylalanyl-tRNA as target structure
Evolutionary trajectory

Spreading of the population on neutral networks

Drift of the population center in sequence space
Spreading and evolution of a population on a neutral network: $t = 150$
Spreading and evolution of a population on a neutral network: $t = 170$
Spreading and evolution of a population on a neutral network: $t = 200$
Spreading and evolution of a population on a neutral network: $t = 350$
Spreading and evolution of a population on a neutral network: $t = 500$
Spreading and evolution of a population on a neutral network: $t = 650$
Spreading and evolution of a population on a neutral network: $t = 820$
Spreading and evolution of a population on a neutral network: $t = 825$
Spreading and evolution of a population on a neutral network: $t = 830$
Spreading and evolution of a population on a neutral network: $t = 835$
Spreading and evolution of a population on a neutral network: $t = 840$
Spreading and evolution of a population on a neutral network: $t = 845$
Spreading and evolution of a population on a neutral network: $t = 850$
Spreading and evolution of a population on a neutral network: $t = 855$
Anne Kupczok, Peter Dittrich, Determinants of simulated RNA evolution.
A sketch of optimization on neutral networks
Initial state

Replication, mutation and dilution

Target

Extinction
Application of molecular evolution to problems in biotechnology
1. Replication and mutation
2. Quasispecies and error thresholds
3. Fitness landscapes and randomization
4. Lethal mutations
5. Ruggedness of natural landscapes
6. Simulation of stochastic phenomena
7. Biology in its full complexity
Three-dimensional structure of the complex between the regulatory protein **cro-repressor** and the binding site on λ-phage **B-DNA**
Sketch of a genetic and metabolic network
Dynamic patterns of gene regulation I: Simple two-gene systems

Stefanie Widdera, Josef Schicho b, Peter Schustera, c, * 

a Institute für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria
b RICAM—Reichardt Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, Altenbergerstraße 99, A-4040 Linz, Austria
c Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Received 2 February 2006; revised in revised form 7 January 2007; accepted 8 January 2007
Available online 14 January 2007

Abstract

Regulation of gene activities is studied by means of computer assisted mathematical analysis of ordinary differential equations (ODEs) derived from binding equilibrium and chemical reaction kinetics. Here, we present results on cross-regulation of two genes through activator and/or repressor binding. Arbitrary (differentiable) binding functions can be used but systematic investigations are presented for gene-regulator complexes with integer valued Hill coefficients up to n = 4. The dynamics of gene regulation is derived from bifurcation patterns of the underlying systems of kinetic ODEs. In particular, we present analytical expressions for the parameter values at which one-dimensional (transcritical, saddle-node or pitchfork) and/or two-dimensional (Hopf) bifurcations occur. A classification of regulatory states is introduced, which makes use of the sign of a ‘regulatory determinant’ D (being the determinant of the block in the Jacobian matrix that contains the derivatives of the regulator binding functions) (6) systems with D < 0, observed, for example, if both proteins are activators or repressors, to give rise to one-dimensional bifurcations only and lead to bistability for n ≥ 2 and (6) systems with D > 0, found for combinations of activation and repression, sustain a Hopf bifurcation and undamped oscillations for n > 2. The influence of basal transcription activity on the bifurcation patterns is described. Binding of multiple subunits can lead to richer dynamics than pure activation or repression states if intermediates between the unbound state and the fully saturated DNA initiate transcription.

Then, the regulatory determinant D can adopt both signs, plus and minus.

Keywords: Basal transcription; Bifurcation analysis; Cooperative binding; Gene regulation; Hill coefficient; Hopf bifurcation

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Theoretical work on gene regulation goes back to the 1960s (Monod et al., 1963) soon after the first repressor protein had been discovered (Jacob and Monod, 1961). A little later the first paper on oscillatory states in gene regulation was published (Goodwin, 1965). The interest in gene regulation and its mathematical analysis never ceased (Tiwari et al., 1974; Tyson and Othmer, 1978; Smith, 1987) and saw a great variety of different attempts to design models of genetic regulatory networks that can be used in systems biology for computer simulation of genetic and metabolic networks. Most models in the literature aim at a minimalistic dynamic description which, nevertheless, tries to account for the basic regulatory functions of large networks in the cell in order to provide a better understanding of cellular dynamics. A classic in general regulatory dynamics is the monograph by Thomas and D’Ari (1990). The currently used mathematical methods comprise application of Boolean logic (Thomas and Kaufman, 2001b; Savageau, 2001; Albert and Othmer, 2003), stochastic processes (Hume, 2000) and deterministic dynamic models, examples are Cherry and Adler (2000), Bindschadler and Sneyd (2001) and Kobayashi et al. (2003) and the recent elegant analysis of bistability (Craciun et al., 2003)
The reaction network of cellular metabolism published by Boehringer-Ingelheim.
The citric acid or Krebs cycle (enlarged from previous slide).
The bacterial cell as an example for the simplest form of autonomous life

The human body:

$10^{14} \text{ cells} = 10^{13} \text{ eukaryotic cells} + \approx 9 \times 10^{13} \text{ bacterial (prokaryotic) cells, and } \approx 200 \text{ eukaryotic cell types}$

The spatial structure of the bacterium *Escherichia coli*
Cascades, A ⇒ B ⇒ C ⇒ ... , and networks of genetic control

Turing pattern resulting from reaction-diffusion equation?

Intercellular communication creating positional information

Development of the fruit fly *drosophila melanogaster*: Genetics, experiment, and imago
\[
\frac{dV}{dt} = \frac{1}{C_M} \left[ I - g_{Na} m^3 h (V - V_{Na}) - g_K n^4 (V - V_K) - g_l (V - V_l) \right]
\]

Hogdkin-Huxley OD equations

\[
\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m
\]

\[
\frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h
\]

\[
\frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n
\]

A single neuron signaling to a muscle fiber
Hodgkin-Huxley partial differential equations (PDE)

\[
\frac{1}{R} \frac{\partial^2 V}{\partial x^2} = C \frac{\partial V}{\partial t} + \left[ g_{Na} m^3 h (V - V_{Na}) + g_K n^4 (V - V_K) + g_l (V - V_l) \right] 2\pi r L
\]

\[
\frac{\partial m}{\partial t} = \alpha_m (1 - m) - \beta_m m
\]

\[
\frac{\partial h}{\partial t} = \alpha_h (1 - h) - \beta_h h
\]

\[
\frac{\partial n}{\partial t} = \alpha_n (1 - n) - \beta_n n
\]

Hodgkin-Huxley equations describing pulse propagation along nerve fibers
ANALYTICAL DYNAMICS OF NEURON PULSE PROPAGATION

PAUL E. PHILLIPSON* and PETER SCHUSTER
Institut für Theoretische Chemie, Universität Wien,
Währingerstraße 17, A-1090 Wien, Austria
*Austrian Academy of Sciences, Dr. Ignaz Seipel-Platz 2,
A-1010 Wien, Austria

Received July 12, 2005; Revised January 11, 2006

The four-dimensional Hodgkin-Huxley equations describe the propagation in space and time of the action potential $v(z)$ along a neural axon with $z = x + ct$ and $c$ being the pulse speed. The potential $v(z)$, which is parameterized by the temperature, is driven by three gating functions, $m(z)$, $h(z)$ and $n(z)$, each of which obeys formal first order kinetics with rate constants that are represented as nonlinear functions of the potential $v$. It is shown that this system can be analytically simplified (i) in the number of gating functions and (ii) in the form of associated rate functions while retaining to close approximation quantitative fidelity to computer solutions of the exact equations over the complete temperature range for which stable pulses exist. At a given temperature we record two solutions ($T < T_{max}$) corresponding to a high-speed and a low-speed branch in speed-temperature plots, $c(T)$, or no solution ($T > T_{max}$). The pulse is considered as composed of two contiguous parts: (i) a pulse front extending from $v(0) = 0$ to a pulse maximum $v = V_{max}$, and (ii) a pulse back extending from $V_{max}$ through a pulse minimum $V_{min}$ to a final regression back to $v(z \to \infty) = 0$. An approximate analytic solution is derived for the pulse front, which is predicted to propagate at a speed $c(T) = 1203.6(T/\degree C)\ cm/sec, \Theta = 3.518 \Theta^2$ in close agreement with computer solution of the exact Hodgkin-Huxley equations for the entire pulse. These results provide the basis for a derivation of two-dimensional differential equation systems for the pulse front and pulse back, which predict the pulse maximum and minimum over the operational temperature range $0 \leq T \leq 25^\circ C$, in close agreement with the exact equations. Most neuron dynamics studies have been based on voltage clamp experiments featuring external current injection in place of self-generating pulse propagation. Since the behaviors of the gating functions are similar, it is suggested that the present approximations might be applicable to such situations as well as to the dynamics of myelinated fibers.

Keywords: Hodgkin-Huxley equations; action potentials; neuron models; nonlinear dynamics; neuron pulse propagation.

1. Neuron Pulse Propagation and the Hodgkin-Huxley Equations

Conduction mechanisms for the propagation of a pulse along an unmyelinated neural axon were encapsulated within a predictive theory by the equations of Hodgkin and Huxley [1952]. These equations became the prototype for description of neural pulse propagation and provide the basis for all subsequent conduction models of neural behavior. The Hodgkin-Huxley equations relate the propagating action potential $v$ to sodium, potassium and leak conductances $g_{Na}$, $I_K$, $g_{leak}$ causing the
$T = 18.5 \, \text{C}; \ \theta = 1873.33 \, \text{cm/sec}$
The human brain

$10^{11}$ neurons connected by $\approx 10^{13}$ to $10^{14}$ synapses
Acknowledgement of support

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
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

Christian Reidys, Christian Forst, Los Alamos National Laboratory, NM

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, Universität Wien, AT

Jan Cupal, Stefan Bernhart, Lukas Endler, Ulrike Langhammer, Rainer Machine, Ulrike Mückstein, Hakim Tafer, Thomas Taylor, Universität Wien, AT
Web-Page for further information:

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