Sequence Evolution

Carsten Kemena & Inken Wohlers

FU Berlin

May 14, 2007

Big Steps in Evolution

1. Molecular replication
2. Translation and genetic code
3. Procariotic cells
4. Eucariotic cells
5. Multicellular live

Question

- How do macromolecules evolve?
- How can information content of genes be maintained/increased?

Quasispecies und Hypercycles

- Model from M. Eigen and P. Schuster [1]
- Explains evolution from simple self-replicating units to complex forms of self-organization
- Based on a mathematical model

Necessary Properties for Darwinian Behaviour

1. Metabolism
2. Self-reproduction
3. Mutability

Metabolism

- Formation and degradation of molecular species
  $\Rightarrow$ independent and spontanious
- System away from equilibration
- Selection effective only for intermediate states

Self-reproduction

- Necessary for any selection process involving destabilization
- Constant degeneration
  $\Rightarrow$ Conservation of information

Mutability

- Physical associated with self-reproduction
- Main source of new information
- Has to be limited - danger of information loss

Modelling of Self-reproduction

$$A + X \xrightarrow{k} 2X, \quad X \xrightarrow{k'} F$$

- X = molecular species
- A = raw material
- F = final product
- k = rate of self-replication
- k'= rate of degeneration

Modelling with a Differential Equation

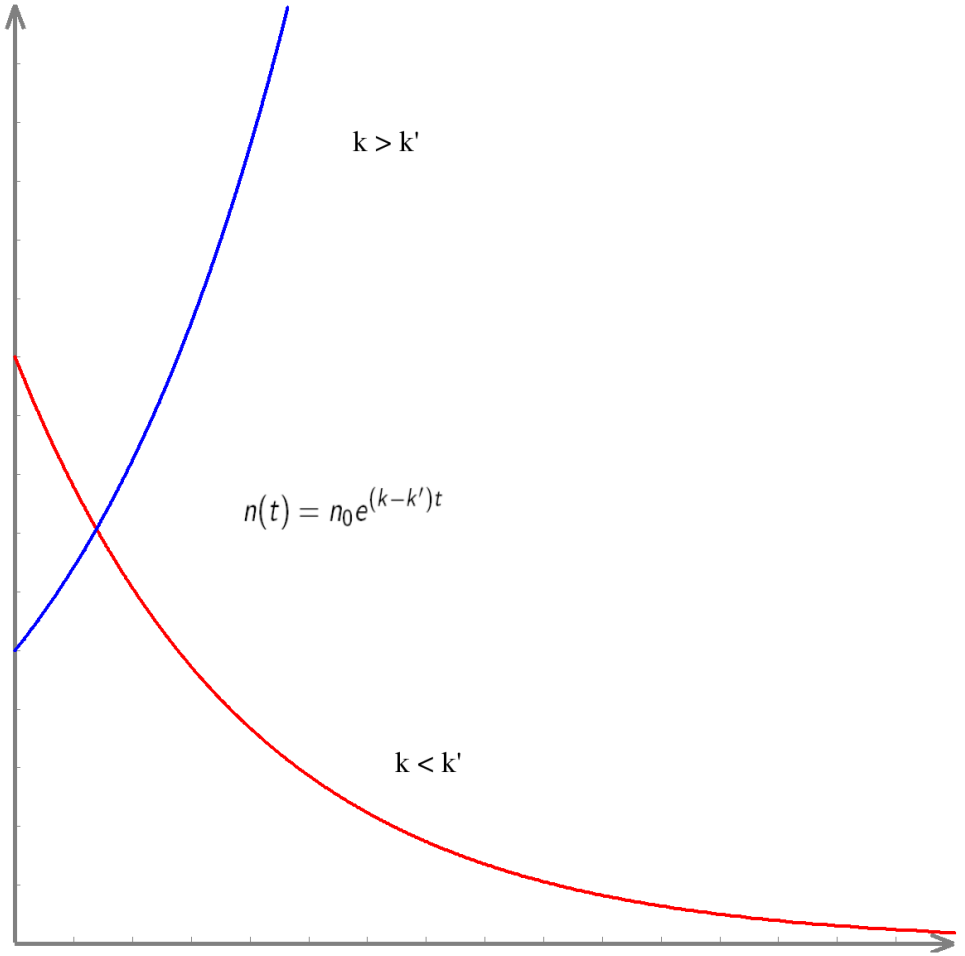
$$\frac{dn}{dt} = kn - k'n = (k - k')n$$

solution:

$$n(t) = n_0 e^{(k-k')t}$$

- n = start concentration
- t = time

Development of Concentrations



Modelling of Molecular Evolution

Usually there exist several different species X_i

$$A + X_1 \xrightarrow{k_1} 2X_1, \quad X_1 \xrightarrow{k_1'} F$$

$$A + X_2 \xrightarrow{k_2} 2X_2, \quad X_2 \xrightarrow{k_2'} F$$

$$\vdots$$

$$A + X_n \xrightarrow{k_n} 2X_n, \quad X_n \xrightarrow{k_n'} F$$

Generalized Differential Equation

$$\frac{dn_i}{dt} = k_i n_i - k_i' n_i = (k_i - k_i') n_i$$

solution:

$$n_i(t) = n_{i,0} e^{(k_i - k_i')t}$$

Including the Selection Process

- A selection process needs selection pressure
  1. limited raw material
  2. limited total number of macromolecules
     $\Rightarrow$ permanent dilution

Differential Equation

$$\frac{dn_i}{dt} = (k_i - k_i')n_i - \varphi n_i$$

Determining φ

$$0 = \sum_i \frac{dn_i}{dt} = \sum_i (k_i - k_i')n_i - \varphi \sum_i n_i$$

$$\Rightarrow \varphi = \frac{\sum_i (k_i - k_i')n_i}{\sum_i n_i} = \frac{\sum_i E_i n_i}{\sum_i n_i} = \bar{E}$$

- $\varphi =$ rate of dilution
- $E_i = k_i - k_i'$ (selection value)

Selection Equation

$$\frac{dn_i}{dt} = (E_i - \bar{E})n_i$$

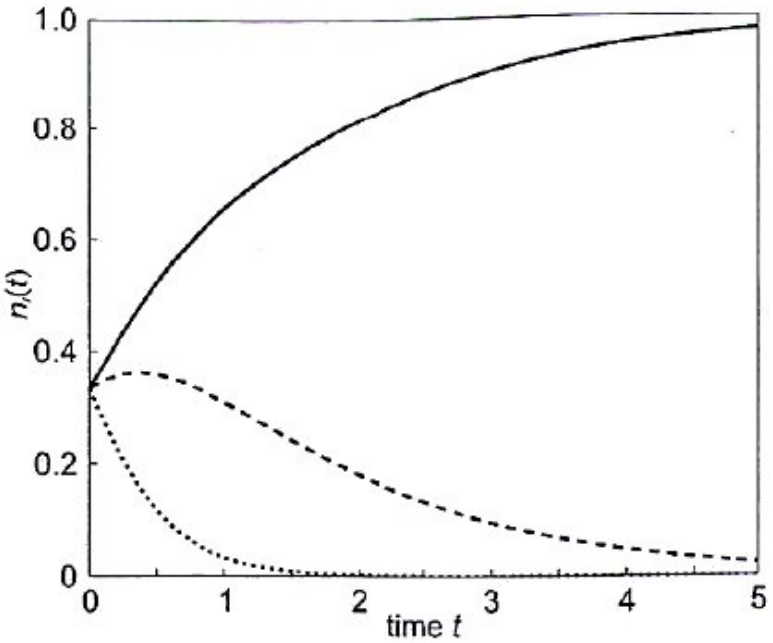
- $\bar{E}$ = mean excess productivity (selection threshold)

Observation

Species with

- $E_i > \bar{E} \Rightarrow \frac{dn_i}{dt} > 0$: concentration increases
- $E_i < \bar{E} \Rightarrow \frac{dn_i}{dt} < 0$: concentration decreases

Development of Concentration

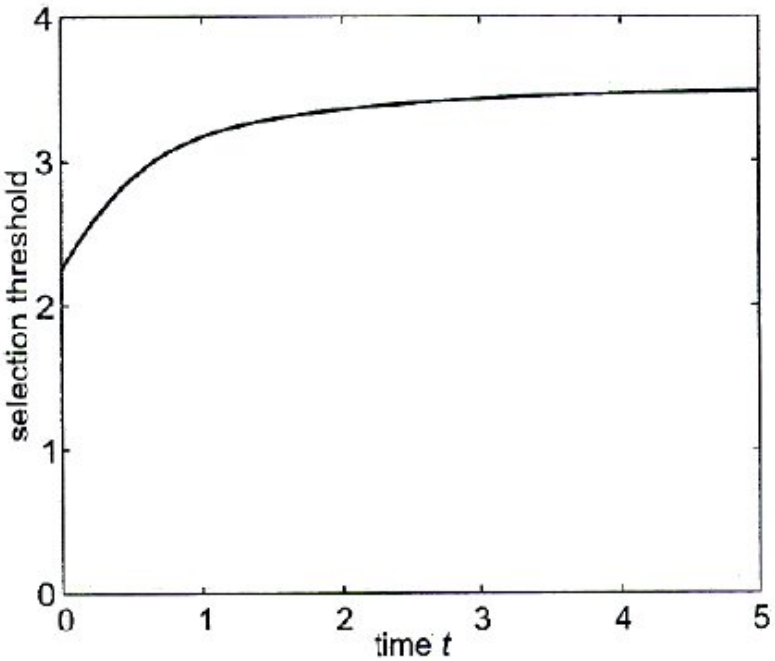


$$—E_1 = 3.5$$
$$\text{- -}E_2 = 2.75$$
$$\cdots E_3 = 0.5$$
$$\bar{E} = 2.25$$

$\Rightarrow$ Only one species survives

Ebeling et. al.

Development of the Selection Threshold



Ebeling et. al.

Mutation

$$\frac{dn_i}{dt} = F_i n_i - k_i' n_i + \sum_j m_{ij} n_j - \varphi n_i$$

Q_i = fraction of error-free replication
$F_i = k_i Q_i$ (error free replication rate)
m_{ij} = mutation rate from species j to species i

Selection Value W_i

- $W_i = k_i Q_i - k_i^{'}$
- $W_i > \bar{E} \Rightarrow$ concentration increases

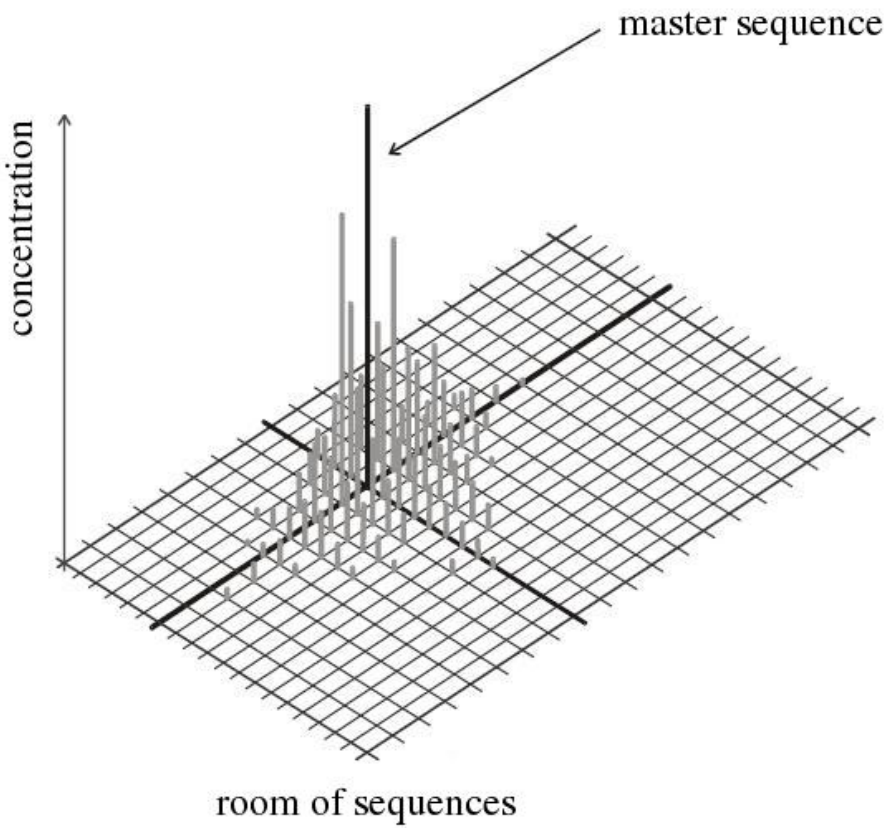
Quasi Species Distribution

- Mutations $\Rightarrow$ not a selection of a single species
- Case $W_m \gg W_j, \quad j \neq m$
  $\Rightarrow$ high concentration of master species X_m
  $\Rightarrow$ positive concentration of closely related species
  $\Rightarrow$ selection of one quasi species

The Quasi Species

- Average sequence is called wild-type
- Only a small fraction equals the wild-type

Distribution of the Quasi Sequence

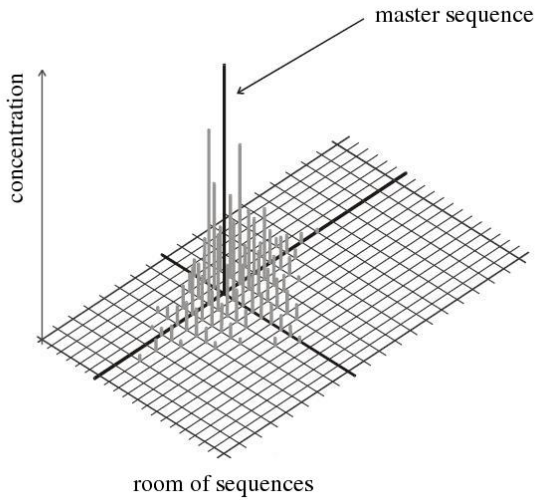


(Schuster)

Stable Conservation of Information

The quasi species is a steady distribution around the master sequence. If Mutation rate gets too high

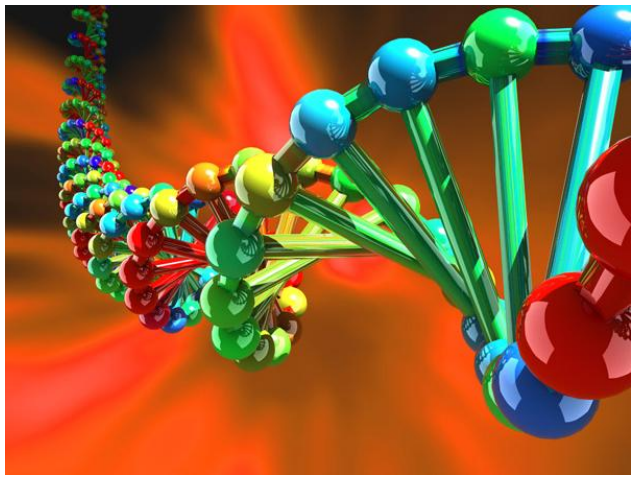
- Too many mutations are produced and the master sequence can not survive
- The quasi species distribution gets unsteady
- Process of inheritence of information collapses



[3]

Three Questions

1. What is the lower bound for the probability of an error-free replication?
2. Given an error rate how much information (number of nucleotides) can be conserved?
3. How can the information content (number of nucleotides) be increased?



[4]

Minimal Probability for Error-free Replication

$$Q_m = q_m^{N_m}$$

Q_m: Probability for error-free replication of the master sequence, Quality factor
q_m: Probability for error-free replication of one monomer in master sequence
N_m: Length of master sequence

- Correct master sequence copies must be able to compete with their error copies:

$$1 > Q_m > Q_{min}.$$

Minimal Probability for Error-free Replication

- Depends on superiority of the master sequence
- Minimal probability of error free replication of one monomer depends on sequence length

$$Q_m > Q_{min} = q_{min}^{N_m} = \frac{1}{\sigma_m}$$

$$\Leftrightarrow$$

$$q_{min} = \sqrt[N_m]{\frac{1}{\sigma_m}}$$

σ_m: Superiority of master sequence

Superiority of the Master Sequence

- Weighted Quotient of the master sequence's fitness divided by the rest of the population's fitness

$$\sigma_m = \frac{A_m}{D_m + \overline{E}_{k \neq m}}$$

σ_m: Superiority of the master sequence
A_m: Reproduction rate of master sequence
D_m: Degradation rate of master sequence
$\overline{E}_{k \neq m}$: average productivity of all competitors

Maximum Sequence Length

- Limit inversely proportional to the average error rate per symbol $1 - q_m$

$$N_{max} = \frac{\ln \sigma_m}{1 - q_m}$$

Expectation Value of an Error

- Expectation value of an error in the master sequence must always remain below a sharply defined threshold

$$\mathbb{E}(\varepsilon_m) = N_m(1 - q_m) \Leftrightarrow \exp(\mathbb{E}(\varepsilon_m)) < \sigma_m$$

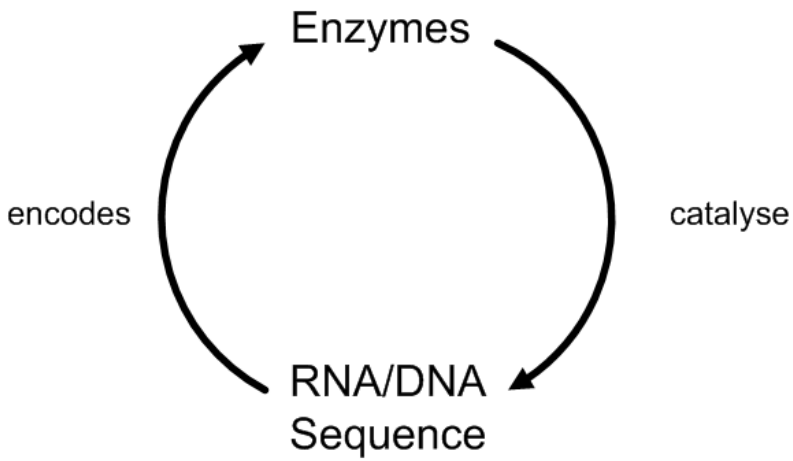
Error rate per monomer $1 - q_m$	**Superiority** σ_m	**Sequence length** N_{max}	**Molecular mechanism**	**Biological Example**
$5x10^{-2}$	2	14	enzyme-free RNA replication	t-RNA precursor, $N = 80$
	20	60		
	200	106		
$5x10^{-4}$	2	1386	single-stranded RNA replication via specific replicases	phage Q_β, $N = 4500$
	20	5991		
	200	10597		
$1x10^{-6}$	2	$0.7x10^6$	DNA replication via polymerases including proofreading by exonuclease	E.coli, $N = 4x10^6$
	20	$3.0x10^6$		
	200	$5.3x10^6$		
$1x10^{-9}$	2	$0.7x10^9$	DNA replication and recombination in eucaryotic cells	vertebrates (man), $N = 3x10^9$
	20	$3.0x10^9$		
	200	$5.3x10^9$		

Cooperation in Sequence Evolution

Lower error rates and thus larger sequence lengths can be achieved if enzymes catalyse replication.

Enzymes

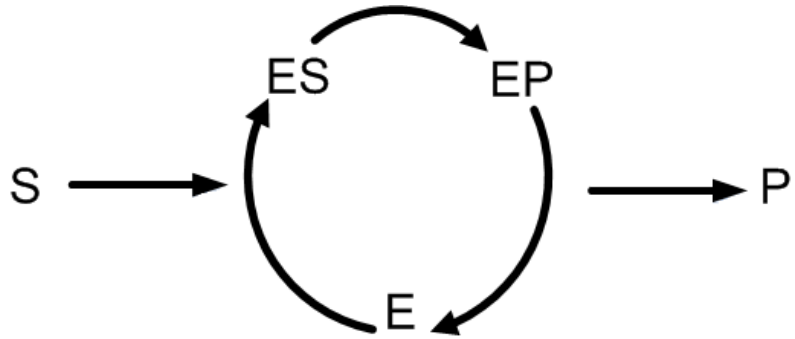
- Lower the error-rate
- Accelerate replication



Cooperation = Selective Advantage

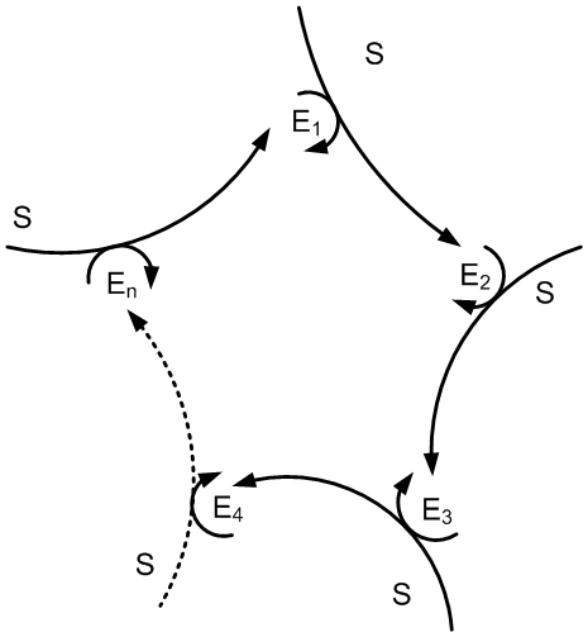
Catalyst

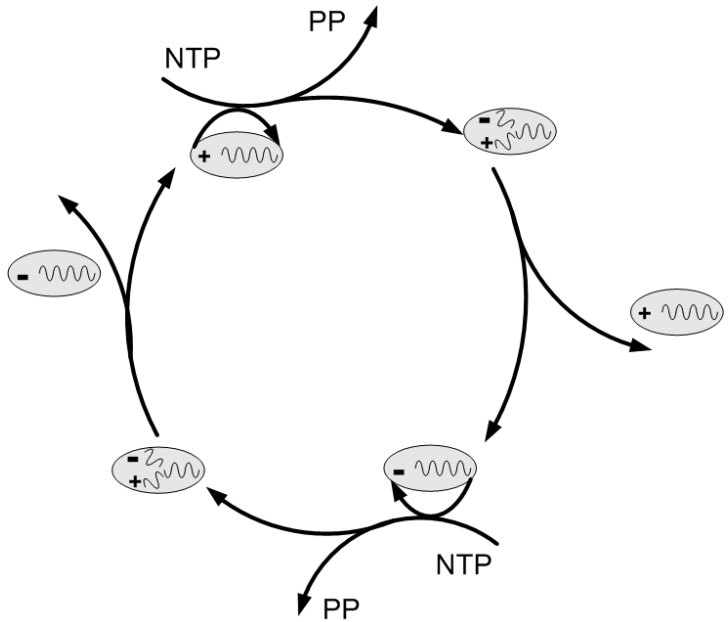
- Sequence of reactions where any product is identical with a rectant of a preceeding step
- Example: Catalytic mechanism of an enzyme



Catalytic Cycle

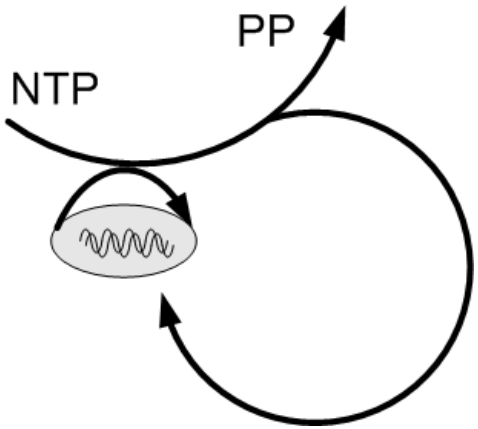
- One up to all intermediates are catalysts
- Example: Replication of single stranded RNA (linear growth)

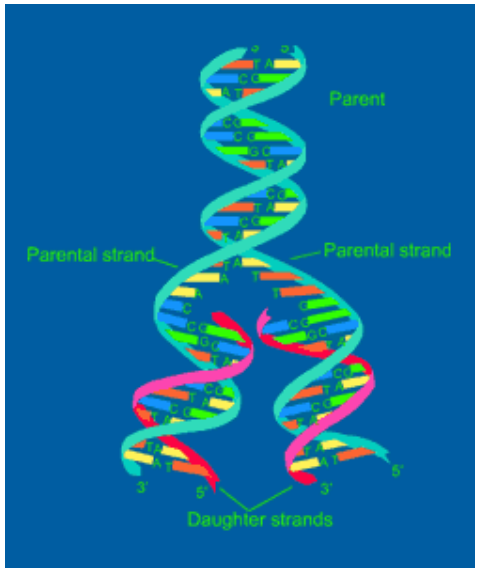




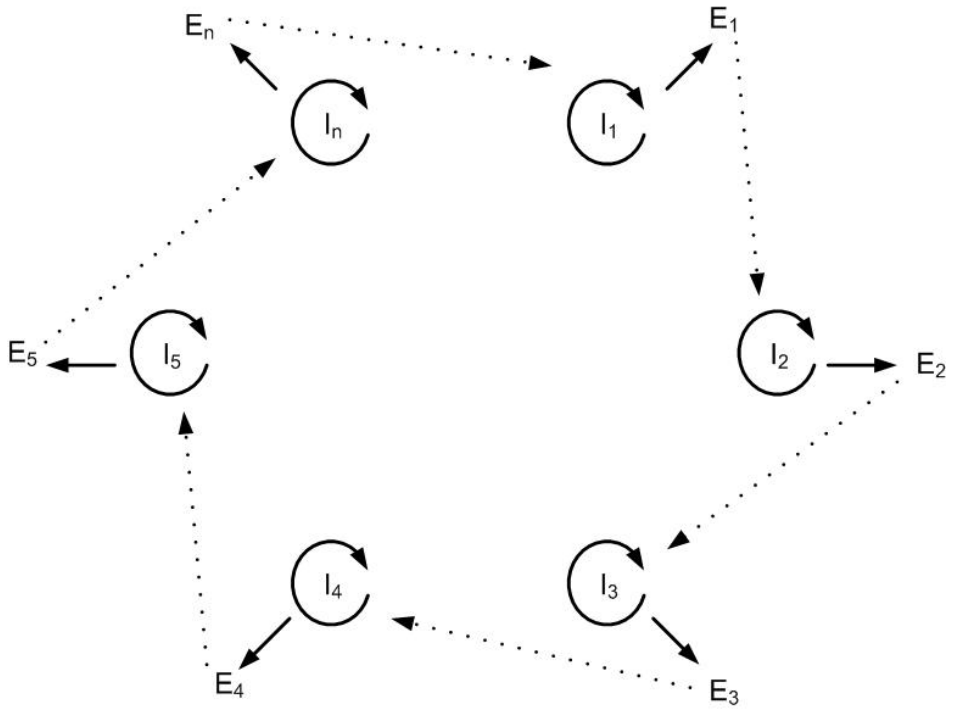
Autocatalyst

- Self-replicative Unit
- Example: DNA-Replication (exponential growth)





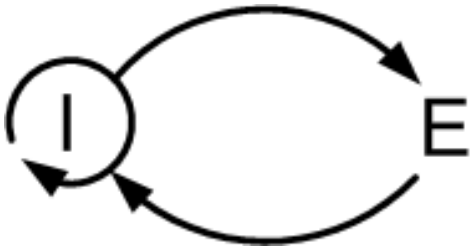
Catalytic Hypercycles

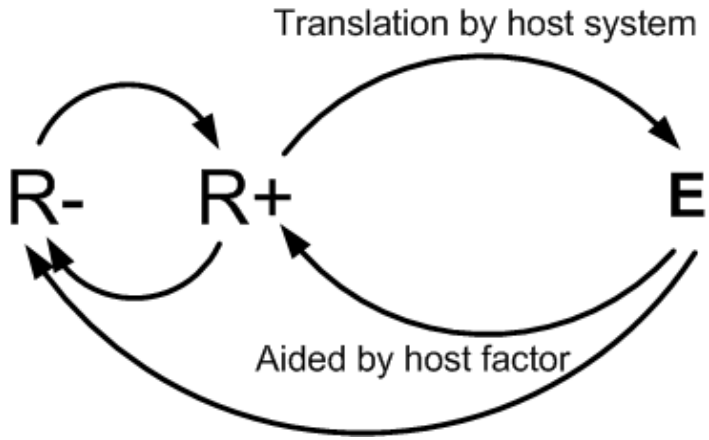


- Cycle of self-replicating or autocatalytic units
- RNA/DNA molecules catalyse synthesis of enzymes which in return catalyse Synthesis of RNA/DNA molecules

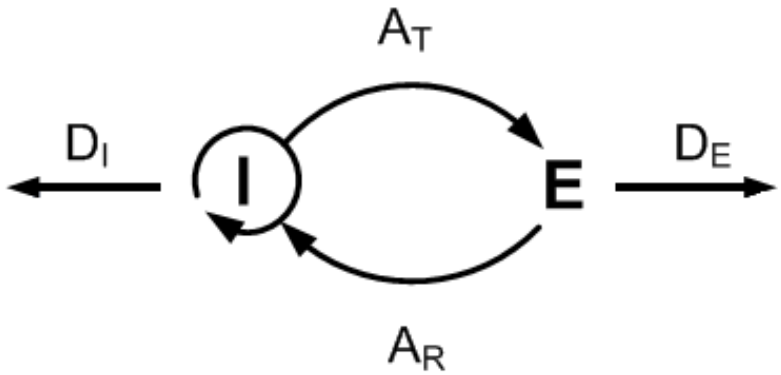
RNA-Phage Infection of a Bacterial Cell

- Translation of infectious strand instructs synthesis of a protein subunit (E) which, associated to host proteins, forms a phage-specific RNA-replicase
- This replicase complex exclusively recognizes phenotypic features of the phage-RNA
- Results in burst of phage-specific RNA-replicase





Simple Hypercycle



$$\frac{dn_E}{dt} = A_T n_I - D_E n_E$$
$$\frac{dn_I}{dt} = A_R n_I n_E - D_I n_I$$

Timecourse

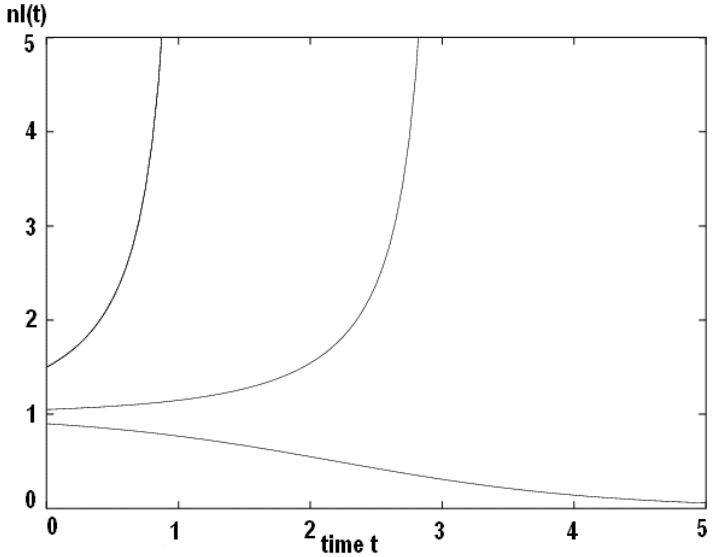
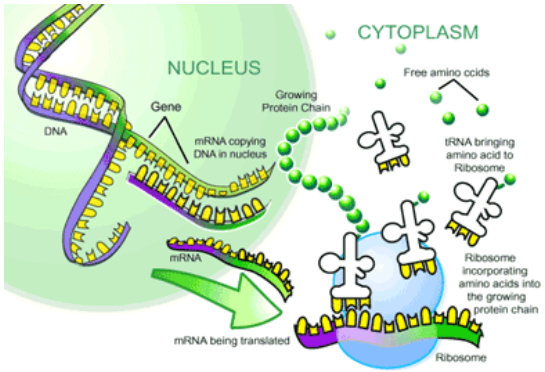


Figure: $A_R = A_T = D_I = D_E = 1$

Decay for $n_I(0) > 1$

Growth for $n_I(0) < 1$

Competition of Hypercycles



[5]

Why has only one basic molecular maschinery of the cell evolved, common for all species?

- Hypercycles compete with each other if they need the same chemical building blocks and if they do not cooperate in higher order linkage

Competition of Hypercycles

$$\frac{dn_{E,i}}{dt} = A_{T,i} n_{I,i} - \varphi_E(t) n_{E,i}$$
$$\frac{dn_{I,i}}{dt} = A_{R,i} n_{I,i} n_{E,i} - \varphi_I(t) n_{I,i}$$

From

$$\sum_i \frac{dn_{E,i}}{dt} = 0 \ , \ \sum_i \frac{dn_{I,i}}{dt} = 0$$

follows

$$\varphi_E = \frac{\sum_i A_{T,i} n_{I,i}}{\sum_k n_{E,k}} \ , \ \varphi_I = \frac{\sum_i A_{R,i} n_{I,i} n_{E,i}}{\sum_k n_{I,k}}$$

Steady States for Competition of Two Hypercycles

There are two possible stable steady states:

$$(n_{E,1}, n_{I,1}, n_{E,2}, n_{I,2}) = (0, 0, C_E, C_I)$$
$$(n_{E,1}, n_{I,1}, n_{E,2}, n_{I,2}) = (C_E, C_I, 0, 0)$$

Depending on the initial value the number of molecules tends to one of the steady states.

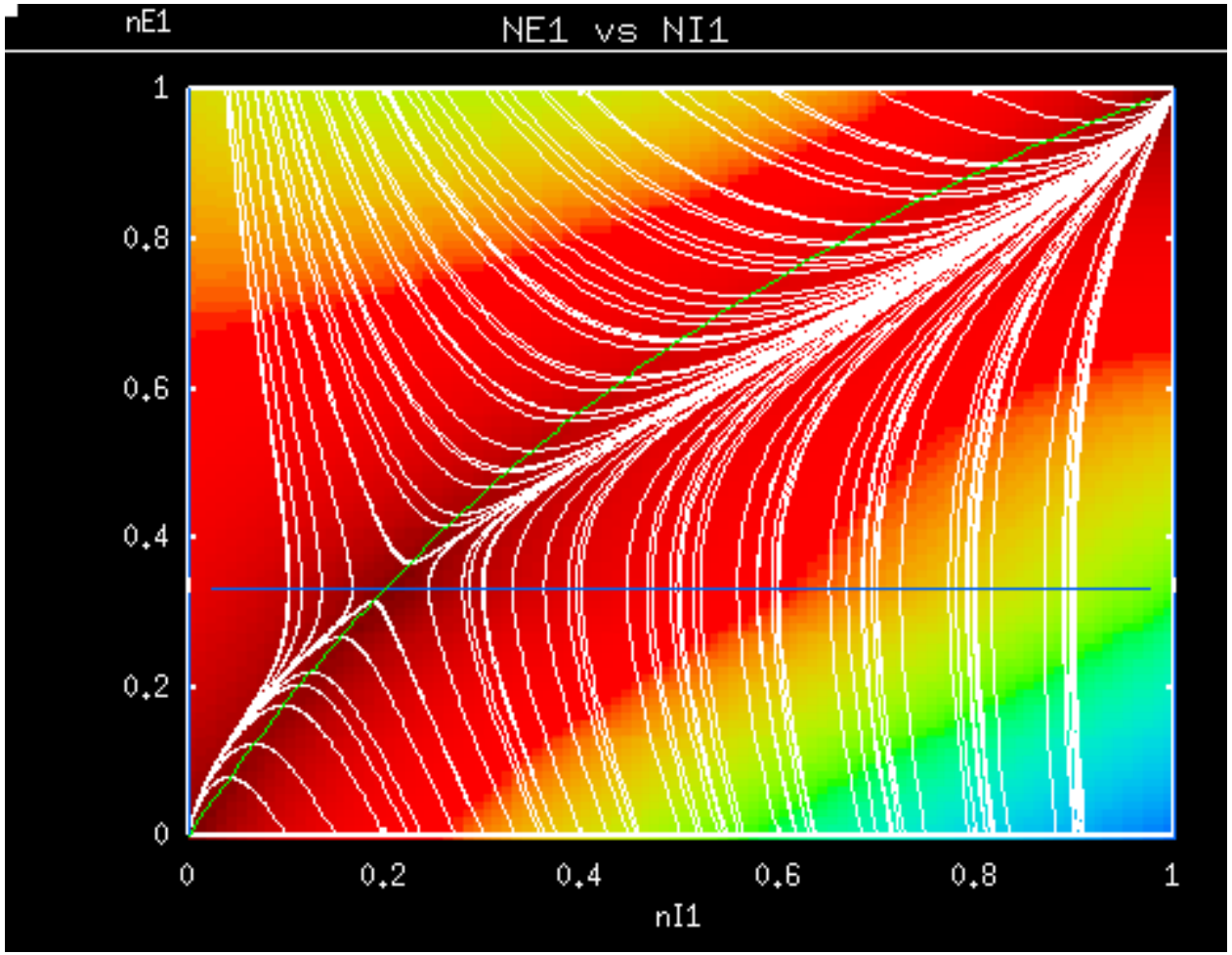


Figure: $A_{T,1} = A_{R,1} = 1, A_{T,2} = A_{R,2} = 0.5$

Once-forever-selection

- If two hypercycles compete with each other only one of them can survive.
- The lower the number of molecules belonging to a competing hypercycle the smaller the chance it survives.
- With (coincidentally) good initial values also non-optimal systems could have prevailed

Only one hypercycle survived and developed to become the molecular machinery of the cell.

Summary

Quasi species

- Explain the inner diversity of species

Hypercycles

- Explain how to increase the information content of self-replicating units
- Are a possible explanation why there is no diversity of the molecular maschinery of the cell.

Open Questions

- How were more and more specific enzymes added during evolution?
- How did a genetic code emerge?

Thanks for your attention!

Literature


Eigen & Schuster (1977) The Hypercycle. A Principle of Natural Self-Organisation. Part A: Emergence of the Hypercycle. Naturwissenschaften 64:541-565.


Competition and Selection in Biological Systems: aus dem Fachkurs 'Modellierung biologischer Systeme', HU 2000.


Schuster (1999) http://www.tbi.univie.ac.at/ pks/PublicLectures/ursprung-1999.pdf


http://www.carlgoodman.co.uk/scientific/scientific/Image%20(24).jp


http://www.scq.ubc.ca/wp-content/translation1.gif