EVOLUTION BEFORE GENES

Theoretical analysis of autocatalytic molecular networks and its implications to metabolism-first theories of the origin of life

Vera Vasas

Supervisors:
Dr. Mauro Santos
Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, Spain

Dr. Eörs Szathmáry
Institute of Biology, Eötvös University, Budapest, Hungary
1. Introduction
   1. Origin of life
   2. Motivation

2. Collectively autocatalytic chemical networks
   1. GARD model and its evolvability
   2. The autocatalytic set model and its evolvability
   3. Coopticive evolution

3. Conclusions
   1. Network motifs that enable selectability
   2. Summary
Introduction

Origin of life theories

Genetics-first theory
- Self-replicating oligonucleotids (RNA world)

First living cell?

Metabolism-first theory
- Self-reproducing catalytic networks
Origin of the unit of evolution

Random chemistry

Start of evolution
- Chemical system
- Information storage and transmission

What could it be...?

First protocell
- Template replication
- Metabolism
- Lipid membrane
Introduction

Units of evolution (JMS)

1. Multiplication
2. Heredity
3. Variation

hereditary traits affecting survival and/or reproduction
Template replicators

- Unlimited variation and heredity
- Experimental evidence for the RNA world
- BUT! Origin of template replicators?
Introduction

Autocatalytic cycles

- Direct autocatalysis (in multiple steps)
- Non-enzymatic cycles, e.g. formose reaction
- BUT! Variations are not heritable
Introduction

Ensemble replication of molecular networks

- Indirect autocatalysis (many more topologies)
- Compositional inheritance
- Models, e.g. GARD, Autocatalytis Sets
- BUT! Evolvable?
Molecules spontaneously aggregate and grow

\[ \beta_{ij} \text{ define the incorporation of molecular species } i \text{ in the growing assembly catalyzed by species } j \]

Randomly drawn from lognormal distribution

\[ N_G = \text{number of different kinds} \]
\[ n_i = \text{count of molecules of kind } i \text{ in the assembly} \]
\[ N = \sum n_i \text{ (assembly size)} \]


04/05/2012, Barcelona

Vera Vasas PhD viva
The GARD model

GARD model: ‘ensemble replicators’

Growth and splitting:

Hereditary trait: composition

The GARD model

Quasi-stationary states

Time-dependent progression of one assembly

Time correlation matrix of assemblies

QSSs: quasi-stationary states
A,B,C: compotyopes

Evolvability of the GARD model

Natural selection introduced

Mutation, variability, heredity... BUT!

*Selection included*

1. Analytical scenario 2. Stochastic simulations
Fitness (growth rate): similarity to target

→ Frequency distributions of composomes does not change

04/05/2012, Barcelona

Vera Vasas PhD viva
Evolvability of the GARD model

But why isn’t it evolvable?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>...</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0,00</td>
<td>0,00</td>
<td>0,23</td>
<td>0,00</td>
<td>0,01</td>
<td>0,00</td>
<td>1,76</td>
<td>0,00</td>
<td>0,38</td>
<td>0,11</td>
<td>0,32</td>
<td>...</td>
<td>351,29</td>
</tr>
<tr>
<td>2</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,03</td>
<td>0,06</td>
<td>0,21</td>
<td>9,10</td>
<td>0,00</td>
<td>0,00</td>
<td>0,18</td>
<td>10,96</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>0,03</td>
<td>0,95</td>
<td>0,17</td>
<td>0,01</td>
<td>4,45</td>
<td>0,00</td>
<td>4,64</td>
<td>0,06</td>
<td>0,28</td>
<td>0,00</td>
<td>0,00</td>
<td>...</td>
<td>0,89</td>
</tr>
<tr>
<td>4</td>
<td>0,06</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,04</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,51</td>
<td>0,00</td>
<td>...</td>
<td>16,92</td>
</tr>
<tr>
<td>5</td>
<td>0,00</td>
<td>0,07</td>
<td>0,03</td>
<td>2,09</td>
<td>0,00</td>
<td>0,01</td>
<td>0,11</td>
<td>0,00</td>
<td>0,67</td>
<td>0,00</td>
<td>0,04</td>
<td>...</td>
<td>0,74</td>
</tr>
<tr>
<td>6</td>
<td>2,15</td>
<td>0,05</td>
<td>0,00</td>
<td>0,02</td>
<td>12,64</td>
<td>0,05</td>
<td>0,70</td>
<td>58,13</td>
<td>0,00</td>
<td>1,48</td>
<td>0,15</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>7</td>
<td>2,13</td>
<td>0,02</td>
<td>0,00</td>
<td>0,16</td>
<td>0,50</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,01</td>
<td>0,06</td>
<td>0,02</td>
<td>...</td>
<td>1,03</td>
</tr>
<tr>
<td>8</td>
<td>0,02</td>
<td>0,00</td>
<td>0,12</td>
<td>0,00</td>
<td>0,05</td>
<td>19,16</td>
<td>0,04</td>
<td>675,62</td>
<td>0,00</td>
<td>0,00</td>
<td>0,06</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>9</td>
<td>0,07</td>
<td>0,00</td>
<td>0,01</td>
<td>0,00</td>
<td>0,27</td>
<td>12,04</td>
<td>0,39</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>...</td>
<td>0,16</td>
</tr>
<tr>
<td>10</td>
<td>0,04</td>
<td>0,00</td>
<td>2,58</td>
<td>0,06</td>
<td>0,00</td>
<td>0,22</td>
<td>0,00</td>
<td>0,04</td>
<td>0,04</td>
<td>0,01</td>
<td>0,01</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>11</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,03</td>
<td>0,56</td>
<td>0,03</td>
<td>0,00</td>
<td>0,74</td>
<td>0,00</td>
<td>0,04</td>
<td>0,01</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>100</td>
<td>0,01</td>
<td>0,00</td>
<td>2,94</td>
<td>0,00</td>
<td>0,01</td>
<td>0,33</td>
<td>0,04</td>
<td>0,07</td>
<td>4,86</td>
<td>0,07</td>
<td>0,04</td>
<td>...</td>
<td>0,00</td>
</tr>
</tbody>
</table>

Matrix of catalytic parameters (ß values)
Hidden compartmentalization

Compotytypes form where large $\beta$ values fall
= strongly catalytic and strongly catalyzed molecules

Evolvability of the GARD model

Analytical scenario

Molecular repertoire: $N_G = 10$

Assembly size: $N = 3-6$

Catalytic interactions: $N_G \times N_G$ matrix lognormal distribution

Number of possible assemblies: $\Omega = 220$

e.g.: $(0\ 1\ 0\ 0\ 0\ 0\ 0\ 2\ 0\ 0\ 0\ 0)$
Evolvability of the GARD model

Analytical scenario: Eigen equation

\[ X_k' = (r_k - E)X_k + \sum_{l=1}^{\Omega} \mu_{kl} X_l \]

\( k = 1,2,..., \Omega \) – all possible assemblies

\( X_k' \) - density of one assembly

\( r_k \) - self-replication factor

\( \mu_{kl} \) - mutation rate

\( E \) - overall excess productivity
Evolvability of the GARD model

Analytical scenario: fitness matrix

\[ X' = (W - E)X \]

\( W \) – fitness matrix

\( w_{kl} = \mu_{kl} \)

\( w_{kk} = r_k \)
Evolvability of the GARD model

Composition of a composome

• 0.1 % of the catalytic interactions govern the dynamics, while the rest is noise
Evolvability of the GARD model

Lack of strongly autocatalytic loops

- Autocatalytic loops are unlikely
  \[
P(\text{loop}) = 1 - \prod_{L=1}^{N_G} \left(1 - P^L\right)^{N_G \choose L}
\]

- Inherent kinetic instability of autocatalytic loops

One giant component

Full connectedness
Evolvability of the GARD model

Explanation for the lack of evolvability

- Strongly catalytic molecules give rise to ‘phenotypes’
- Strongly catalytic molecules are halved in each division
  \[\rightarrow\] The phenotype is eventually lost

\[\sum\] Informational molecules are not autocatalytic

\[\text{assembly} \quad \text{environment}\]

\[\text{‘heredity’} \quad \text{‘mutation’}\]
The autocatalytic set model

Kauffman’s model: polymer chemistry

- Catalysed ligation/cleavage reactions of peptides

- \( P \): probability that a given molecule catalyses a given reaction

- Food set

The autocatalytic set model

Autocatalytic sets

‘set of species in which the production of each species in the set is catalysed by at least one other species in the set’
= collection of heterocatalytic molecules

Propositions of the model:
• Self-organization without natural selection
• Natural selection acts on the autocatalytic set

Chemistry behind: peptides, or any other polymers

But what about hereditary variation...?
The autocatalytic set model

Reimplementation of Kauffman’s model

Mathematical model: supracritical growth

Chemical implementation in a flow reactor: logistic growth

The autocatalytic set model

Earlier criticisms

Lifson: $P$ is a composite probability
$P’$ = probability of being a catalyst
$P”’$ = probability of catalysing a particular reaction

$\Rightarrow$ with high $P’$ and $P”’$ there is supracritical growth

Szathmáry: ‘paradox of specificity’
$K$: probability that a given molecule inhibits a given reaction

$\Sigma$: only quantitative issues

$P = 0.0003, K = 0.001$
The autocatalytic set model

Framework for testing selectability

Compartmentalization
  Flow reactor

Multiple attractors
  Shuffling concentrations
  Growth - splitting in one lineage

Selectability in a population
  Artificial selection (N = 10)
  Natural selection (Moran process, N = 100)
Results: Kauffman’s original model

Only one attractor

⇒ Only one type

⇒ Selection is meaningless

Equilibrium mass after shuffling initial concentrations
The autocatalytic set model

Results: the inhibition model

Multiple attractors, but: Transitions between them are rare and random, no response to selection

Existence of multiple attractors does not guarantee selectability

Two examples of artificial selection for increased non-food mass
Cooptive evolution model

Growth by novel autocatalytic loops

Extra: rare uncatalysed reactions result in novel species
Rarely: non-food generated autocatalytic loops

tendency to increase in mass and complexity
Cooptive evolution model

Mutation - selection balance

Propagule size = 2000
Only the large network
~ higher growth rate

Propagule size = 800
Both networks:
selection not strong enough

Propagule size = 500
Only the small network
~ meltdown
Cooptive evolution model

Network motifs of heritable adaptation

Shared periphery of Cores 1 & 2
Periphery of Core 2
Cooptive evolution model

Results and conclusions

• Rare novel species generate reaction avalanches
• Novel extensions to the network (autocatalytic loops) ➔ bits of heritable information

• Properties:
  • Attractor-based evolution
  • Weakly correlated variation
  • Limited heredity

• Small catalytic networks are selectable in a population
• Any chemistry

## Summary

### Evolvability analyses

Selection is only possible in the cooptive evolution model.

<table>
<thead>
<tr>
<th></th>
<th>GARD model</th>
<th>Aut Sets: Standard model</th>
<th>Aut Sets: Inhibition model</th>
<th>Cooptive evolution model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compartments</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Multiple attractors</td>
<td>✔</td>
<td>✗</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Selectability</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
</tr>
</tbody>
</table>
Summary

Open questions

• Ecology of compartments
  • Waste products
  • Fusion -> Infection dynamics of cores

• Chemical basis for cooptive evolution?
  • Peptides – pattern matching, realistic reaction rates
  • RNA – catalytic properties of small RNA
  • Small metabolites – realistic simulations of organic chemistry

• Origin of compartments?
  • Autocatalytic lipophils (as in GARD)
  • Micropores in alkaline hydrothermal vents

• Transition to RNA world?
The GARD model

Autocatalytic loops in GARD

Catalyzed inclusion/exclusion chemistry

A

Reaction graph

assembly

environment

B

Catalytic graph

04/05/2012, Barcelona

Vera Vasas PhD viva
**Evolvability of the GARD model**

**Structural reasons for QSSs**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>...</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0,00</td>
<td>0,00</td>
<td>0,23</td>
<td>0,00</td>
<td>0,01</td>
<td>0,00</td>
<td>1,76</td>
<td>0,00</td>
<td>0,38</td>
<td>0,11</td>
<td>0,32</td>
<td>...</td>
<td>351,29</td>
</tr>
<tr>
<td>2</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,03</td>
<td>0,06</td>
<td>0,21</td>
<td>9,10</td>
<td>0,00</td>
<td>0,00</td>
<td>0,18</td>
<td>10,96</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>3</td>
<td>0,03</td>
<td>0,95</td>
<td>0,17</td>
<td>0,01</td>
<td>4,45</td>
<td>0,00</td>
<td>4,64</td>
<td>0,06</td>
<td>0,28</td>
<td>0,00</td>
<td>0,00</td>
<td>...</td>
<td>0,89</td>
</tr>
<tr>
<td>4</td>
<td>0,06</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,04</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>0,51</td>
<td>0,00</td>
<td>...</td>
<td>16,92</td>
</tr>
<tr>
<td>5</td>
<td>0,00</td>
<td>0,07</td>
<td>0,03</td>
<td>2,09</td>
<td>0,00</td>
<td>0,01</td>
<td>0,11</td>
<td>0,00</td>
<td>0,67</td>
<td>0,00</td>
<td>0,04</td>
<td>...</td>
<td>0,74</td>
</tr>
<tr>
<td>6</td>
<td>2,15</td>
<td>0,05</td>
<td>0,00</td>
<td>0,02</td>
<td>12,64</td>
<td>0,05</td>
<td>0,70</td>
<td>58,13</td>
<td>0,00</td>
<td>1,48</td>
<td>0,15</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>7</td>
<td>2,13</td>
<td>0,02</td>
<td>0,00</td>
<td>0,16</td>
<td>0,50</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,06</td>
<td>0,02</td>
<td>0,00</td>
<td>...</td>
<td>1,03</td>
</tr>
<tr>
<td>8</td>
<td>0,02</td>
<td>0,00</td>
<td>0,12</td>
<td>0,00</td>
<td>0,05</td>
<td>19,16</td>
<td>0,04</td>
<td>675,62</td>
<td>0,00</td>
<td>0,00</td>
<td>0,06</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>9</td>
<td>0,07</td>
<td>0,00</td>
<td>0,01</td>
<td>0,00</td>
<td>0,27</td>
<td>12,04</td>
<td>0,39</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
<td>...</td>
<td>0,16</td>
</tr>
<tr>
<td>10</td>
<td>0,04</td>
<td>0,00</td>
<td>2,58</td>
<td>0,06</td>
<td>0,00</td>
<td>0,22</td>
<td>0,00</td>
<td>0,04</td>
<td>0,03</td>
<td>0,04</td>
<td>0,01</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>11</td>
<td>0,01</td>
<td>0,00</td>
<td>0,00</td>
<td>0,03</td>
<td>0,56</td>
<td>0,03</td>
<td>0,00</td>
<td>0,74</td>
<td>0,00</td>
<td>0,04</td>
<td>0,01</td>
<td>...</td>
<td>0,00</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>100</td>
<td>0,01</td>
<td>0,00</td>
<td>2,94</td>
<td>0,00</td>
<td>0,01</td>
<td>0,33</td>
<td>0,04</td>
<td>0,07</td>
<td>4,86</td>
<td>0,07</td>
<td>0,04</td>
<td>...</td>
<td>0,00</td>
</tr>
</tbody>
</table>

**Matrix of catalytic parameters (β values)**

04/05/2012, Barcelona

Vera Vasas PhD viva
Evolvability of the GARD model

Structural reasons for QSSs

Compartmentalization analysis of the beta matrix

Interaction occurring

<table>
<thead>
<tr>
<th>Compartment membership</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Same</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Criterion = A + D – B – C
Evolvability of the GARD model

Structural reasons for QSSs

Compotytypes = quasi-compartmentalization of the beta matrix
Evolvability of the GARD model

Analytical scenario with natural selection

*Mutation, variability, heredity... BUT!*

Selection included: growth rate x fH

Equilibrium frequencies are the same
Evolvability of the GARD model

Simulations

Molecular repertoire: $N_G = 100$

Assembly size: $N = 40-80$

Catalytic interactions: $N_G \times N_G$ matrix
lognormal distribution
Evolvability of the GARD model

Simulations with natural selection

*Mutation, variability, heredity... BUT!*

Selection included: stochastic simulations: Moran population of 1500 assemblies fitness: similarity to target stationary distribution?

Frequency distributions of composites does not change Can NOT deviate from the background dynamics
Evolvability of the GARD model

Lack of strongly autocatalytic loops

- Autocatalytic loops are unlikely

\[
P(\text{loop}) = 1 - \prod_{L=1}^{N_G} (1 - P^L)^{N_G \choose L}
\]

- Inherent kinetic instability of autocatalytic loops

One giant component

Full connectedness
Evolvability of the GARD model

Disjoint autocatalytic loops are unlikely

Empirical frequency of autocatalytic loops
Evolvability of the GARD model

Kinetic instability of autocatalytic loops

• Equilibrium concentrations

\[
\begin{align*}
\dot{x} &= ky - dx(x + y + v + z) \\
\dot{y} &= gx - dy(x + y + v + z) \\
\dot{v} &= kx - dv(x + y + v + z) \\
\dot{z} &= gy - dz(x + y + v + z)
\end{align*}
\]

\[g = 100, k = 1 \text{ and } d = 0.1\]

\[(x, y, v, z) = (0.9 \ 9 \ 0.09 \ 90)\]

• General linear model

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>21.885(a)</td>
<td>3</td>
<td>7.295</td>
<td>15.958</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>16.000</td>
<td>1</td>
<td>16.000</td>
<td>35.001</td>
<td>.000</td>
</tr>
<tr>
<td>Catalyzed</td>
<td>21.334</td>
<td>1</td>
<td>21.334</td>
<td>46.670</td>
<td>.000</td>
</tr>
<tr>
<td>cycle_p0.010</td>
<td>.257</td>
<td>1</td>
<td>.257</td>
<td>.563</td>
<td>.455</td>
</tr>
<tr>
<td>catalyzed * cycle_p0.010</td>
<td>.293</td>
<td>1</td>
<td>.293</td>
<td>.641</td>
<td>.425</td>
</tr>
<tr>
<td>Error</td>
<td>43.885</td>
<td>96</td>
<td>.457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>81.769</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>65.769</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cooptive evolution model

Anatomy of autocatalytic sets

One-member autocatalytic loop

Two-member autocatalytic loop

Three-member autocatalytic loop

Only catalytic reaction dependencies in autocatalytic loop

Mixed reaction dependencies in autocatalytic loop
Summary

Main achievements

• Previous state-of-the-art
  • Lipid world (GARD) – one lineage, multiple attractors
  • Autocatalytic polymer networks – self-organization

• Population dynamics for existing models and suggestions
  • GARD  
  • Autocatalytic sets  
  • Autocatalytic sets with inhibition

• Theory of selectable chemical network organizations
• Proof of concept for their selectability