Lipid World and Systems Pre-Biology

Doron Lancet, Crown Human Genome Center, the Weizmann Institute of Science, Rehovot, Israel; and KITP
The molecular chicken and egg problem
RNA World

- RNA chemistry
- Molecular self-replication
- Lipid vesicles appearance
- Proteins appearance

Lipid World

- Self-assembly of lipids
- Compositional self-replication
- Evolution and internal self-organization
- Compartment formation and polymerization

Protocell

Segrè et al. EMBO Reports 2000
A graded appearance of life-like entities

- **Prebiotic**
- **Biocentric**
- **Mesobiotic entities**
  - Appearance of RNA-like self-replicating polymers
- **Ensemble replication**
- **Unorthodox “graded” view**
- **Orthodox “abrupt” view**
A graded appearance of life-like entities

Orthodox "abrupt" view

Pre-RNA world

Unorthodox "graded" view

Appearance of RNA-like self replicating polymers

Prebiotic

Biotic

Mesobiotic entities

Life-like properties

Time (GYa)
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable *in silico* or not?
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable *in silico* or not?
Lipids make barriers in present day life

But could it have been different early on?
Alexander Oparin

“Origin of Life” 1924

Prebiotic “Soup”

Colloidal coacervates

“Metabolism first”, not “genome first”

Replication without DNA

Organic compounds can come from space

The Murchison chondritic (carbonaceous) meteorite

Fragments fell on September 28th 1969 around the small town of Murchison, near Melbourne, Australia

Comets and interstellar dust particles
Meteorite material makes vesicles

Deamer and colleagues
Lipidomics: the diversity of present-day lipids
Alfred Merrill, Georgia Tech - SphingoMap
Lipid combinatorics

Alfred Merrill, Georgia Tech - SphingoMap - detail
Lipid combinatorics

Head groups combinatorics
• peptides
• nucleotides
• phosphates, thiols
• metal chelators
• cofactors
  etc

Planetary random chemistry
Life’s origin: a planet-scale random chemistry “experiment”!

Diversity is cheap: Billions of different organic compounds may form spontaneously!
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable *in silico* or not?
Cyclodextrins act as artificial enzymes

Ronald Breslow
Department of Chemistry
Columbia University

Cyclodextrin: cyclic oligosaccharide
Membrane mimetic chemistry – Lipid catalysis

Vesicular Catalysis of an SN2 Reaction.
Jaap Klijn Jan Engberts
Langmuir; 2005; 21(22) pp 9809 - 9817;


1) Lipid micelle: Container only

2) Micelle with trapped polynucleotide: Container + sequence information

3) Mixed micelle: Assembly with compositional information and no informational biopolymers

GARD Lipid World
Two types of complexity in any living cell

1) Molecular complexity
2) Ensemble complexity

The GARD model is strongly based on ensemble complexity!
Can all three aspects of life be manifested by monomers only? 

- Membranes: Enclosure
- Metabolism
- Templates
- Information

Histidine
A specific quantitative model for Lipid World: GARD – Graded Autocatalysis Replication Domain

Monomer joining

Catalyzed homeostatic growth

Occasional fission generates progeny

Lipid catalysis, Monomers only, Compositional information
Mutual catalysis

gard - formal definition

Lipid or any other amphiphile

Differential equation set

\[ \frac{d n_i}{dt} = (k_i \rho_i - k_{-i} n_i)(1 + \sum \beta_{ji} n_j) \]
Catalytic enhancement

\[ \frac{dn_i}{dt} = \left( \frac{k_f N}{k_b} \right) (1+\beta_{ij}) N \]

\( i, j = 1, 2, \ldots, N \)

Forward reaction

Backward reaction

(assembly size)

The GARD model: governed by drug-related statistics

Mutual catalysis matrix

The Receptor Affinity Distribution model
GARD inheritance: what is propagated?

- Monomer joining
- Catalyzed *homeostatic* growth
- Occasional fission generates progeny

Compositional information
Lipid World
Combinatorial assemblies

RNA World:
Combinatorial Sequence
Compositional information

Consider a repertoire of $N_G$ different molecule types

$n_i$ is the number of molecules of type $i$ in the assembly

The vector $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ defines the assembly’s composition

$n \equiv 3$
A sequential oligomer with $N = 10$ units can be constructed in $20^{10}$ different ways $\rightarrow$ has $\log_2(20^{10}) \approx 43$ bits of information.

A compositional assembly with $N = 10$ units from the same alphabet can be constructed in $2 \times 10^7$ different ways $\rightarrow$ has $\log_2(2 \times 10^7) \approx 24$ bits of information.
Compositional information is copied prior to cell division.

To allow cell division, before DNA replicates, new copies of all the molecules in the cells need to be produced.
GARD dynamics: Trajectory in a $N_G$-dimensional compositional space

Reduced dimensionality by Principal Component Analysis

Fixed point or quasi-stationary state

(Compositional genome)
A composome has reduced repertoire

Environment has $N_G$ molecular types

Composome has $N_e \ll N_G$ molecular types
Prebiotics: a random library of chemicals

Biotics: A miniscule fraction of the possible molecules

GARD may explain reduction of possibilities
“GARDobes” on Mars?

100 nm

Not too small for early protocells!

Restricted PAH spectrum
Consistent with GARD concepts
Composome evolution under a Constant Population (CP) constraint

Segre et al., PNAS 2000
Only few compositions are composomes

Compositions

Composomes
Bad replicator, $P = 10^{-4}$

Excellent replicator, $P = 10^{-40}$

GARD’s advantage: “Planetary Probability” computations

Can this $P$ be materialized given the ocean volume and time window?
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable in silico or not?
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable *in silico* or not?
Systems pre-biology

Systems Biology: Understanding cells and organisms as complete, highly complex entities.

Prebiotic entities may have acquired Systems properties very early on.
Dyson constructs a "toy" model a system of recombining monomers in which "alive" and "dead" can be defined.

With plausible parameter values, a jump to an organized state can happen.

Darwinian selection then drives towards greater complexity.

A network of mutually catalytic events
A, B are “foodstuff” (monomers)
Each GARD composome is a network with a different molecular repertoire.
Different composomes have different network motifs

Canonical composome

Composome 1

Composome 2

Composome 3

White nodes = not shared. Edge threshold $\beta > 100$
GARD evolutions

Punctuated equilibrium

Shift from one network to another via compositional mutations
Canonic composome (eigenvector).
\(N_G = 1000\)
\(N = 800\)
Catalytic potency (\(\beta\)) cutoff = 100

Protein interaction map (PIM) generated by using \(~100\) known or suspected cell cycle regulators in *Drosophila melanogaster*
Network degree (k) distribution analysis

Random network
Protein interaction network

Degree distribution analysis

Power law!
In the simplest embodiment, GARD network do not show a power law in their degree distribution.
In the simplest embodiment, GARD network do *not* show a power law in their degree distribution.

Open question: What might render GARD networks more similar to present life’s networks in showing a power-law?

1000 canonical composomes with $N_G = 1000$
Different colors for different $\beta$ cutoffs
But there is a power law distribution for $n_i$ values

We suspect this power-law behavior stems from the log-normal distribution of the beta mutually-catalytic parameters

$n_i$ is the count of the n-th type of molecule

Individual traces

Average of 500 GARD traces
Scale free networks arise due to a “rich gets richer” principle

Bose-Einstein condensation in complex networks.
Bianconi G, Barabasi AL.

… the model reduces to the scale-free model… power-law connectivity distribution observed in diverse systems…

The model describes a “first-mover-wins” behavior, in which the oldest nodes acquire most links.
GARD dynamics and $n_i$ power law

Monomer joining

Homeostatic growth $\approx$ “the rich gets richer”

Power law means many copies of relatively few types, hence facile “splittability”
GARD networks:

* Weighted
* Directed
  • All nodes equivalent
GARD mutation analysis – sequentially delete every node

Wild type

Mutation in A22
A statistical approach to GARD network mutations

1000 networks, each with different $\beta$ matrix

Only the 3% of the networks with $H_{wt-Mut} < 0.9$ are shown ($H$=compositional similarity)
Future: GARD synthetic lethality – mutate two nodes at a time

Plan to analyze:

- Synthetic lethality
- Synthetic sickness
- Extragnic suppression
- Robustness vs fragility
- Node addiction
Two scenarios for increasing network complexity

A: Increasing node count

B: Increasing node fidelity
Current exploration - Polymer GARD

Introduce covalent oligomerization (endogenous synthesis) to GARD assemblies (Shenhav et al, OLEG 2005)
Beyond simple covalent oligomerization

- Longer oligomers
- "Folding" procedures
- RNA-like templating

Genetic algorithms
Trimer GARD simulations
Trimer GARD simulations show open-ended evolution

Unexpected departure in compositional space

Principal component analysis of compotype combination
What is the real question?

1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today’s chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?
11) Understandable *in silico* or not?
Approaches to the study of Life’s origin

1) Test tube experiments, with a stress on microanalysis of individual entities, also in very large scale/duration experiments

2) Galactic travel

3) Large scale chemistry-realistic computer simulations.
2) Artificial Life (AL or Alife)

Helps understand principles but removed from chemical reality

http://www.webslave.dircon.co.uk/alife/intro.html

John Conway’s Game of Life
**In-silico** future of the GARD model:

Large scale computer simulations of *realistic chemistry*, similar to those used to study the origin of the universe or of galaxies and suns.
PROSPECTS OF A COMPUTATIONAL ORIGIN OF LIFE ENDEAVOR

BARAK SHENHAV and DORON LANCET*
Department of Molecular Genetics and the Crown Human Genome Center,
the Weizmann Institute of Science, Rehovot 76100, Israel

Computed GARD reactions per month

Slope from Moore’s Law (computing power doubles Every 18 months)

Stochastic random chemistry

Enhancement by computing stochastic chemistry parameters via molecular dynamics

Molecular dynamics
**In-silico** future of the GARD model:

**Conjecture:** The *In-silico* Chemistry of 2035 or 2055 may provide a highly accurate reenactment of protein folding, enzyme specificity as well as prebiotic scenarios!
GARD – Lipid World

Chickegg
GARD – Lipid World

- Includes *metabolism*-like networks
- Contains compositional *information*
- Embodies an enclosed *compartment*
- Capable of rudimentary reproduction
- Transmits information with mutations
- Capable of primitive evolution
- Can be made gradually more elaborate and more life-like
Barak Shenhav  
Ran Kafri  
Aia Oz  
Ariel Solomon  
Sagi Goldman

Past:  
Daniel Segrè  
Tzachi Pilpel  
David Deamer (UCSC)  
Tal Shai  
Arren Bar-Even  
Dafna Ben-Eli

**OOL credits**

**RNA World**
- RNA chemistry
- Molecular self-replication
- Lipid vesicles appearance
- Proteins appearance

**Lipid World**
- Self-assembly of lipids
- Compositional self-replication
- Evolution and internal self-organization
- Compartment formation and polymerization

Protocell

Segrè et al. EMBO Reports 2000