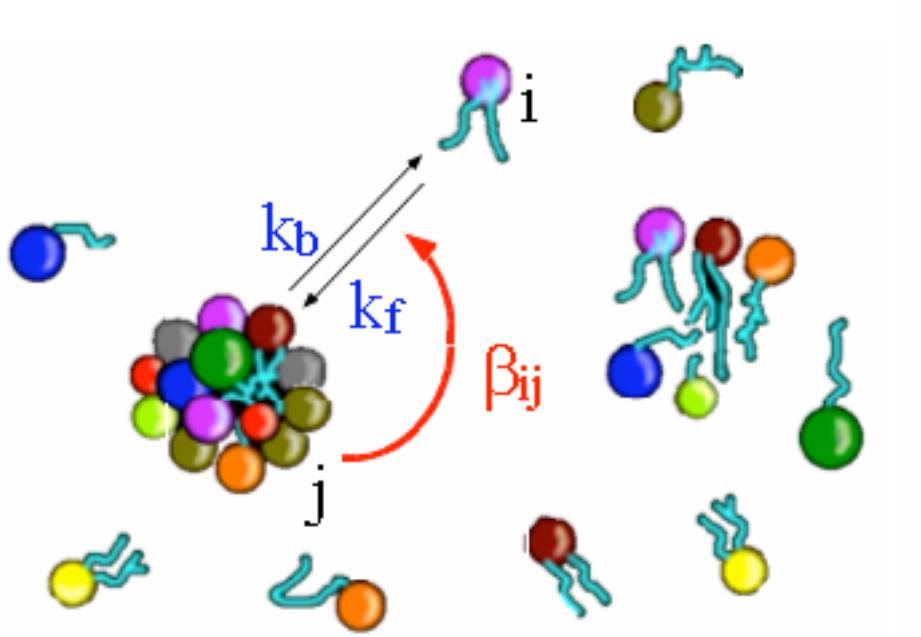
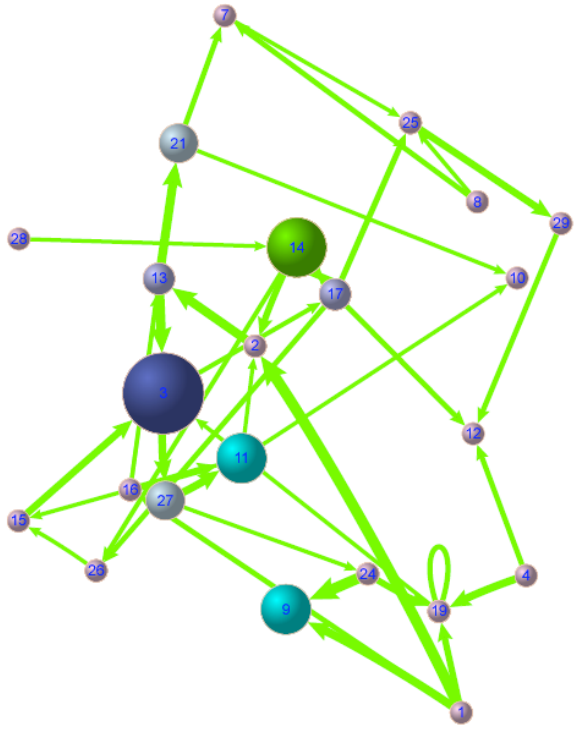
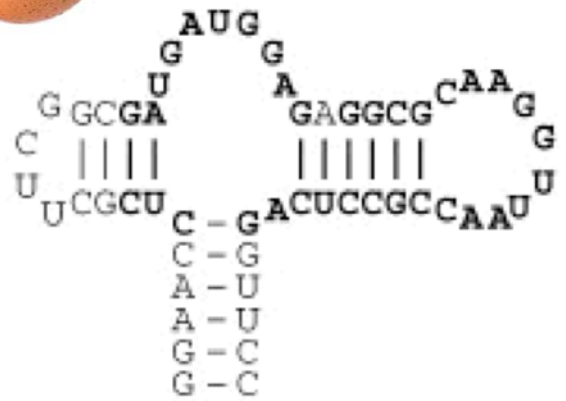
Lipid World and Systems Pre-Biology

Doron Lancet, Crown Human Genome Center, the Weizmann Institute of Science, Rehovot, Israel; and KITP

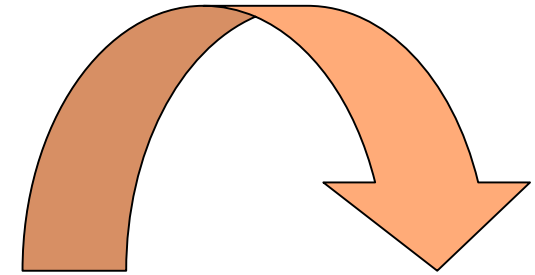


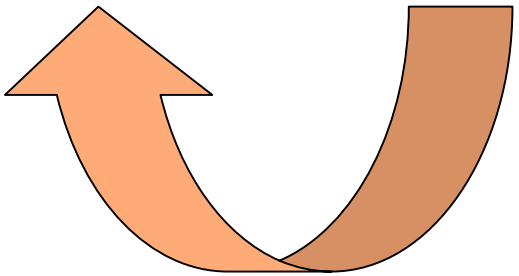


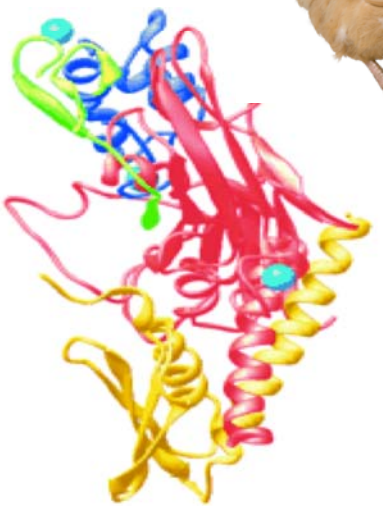
The molecular chicken and egg problem



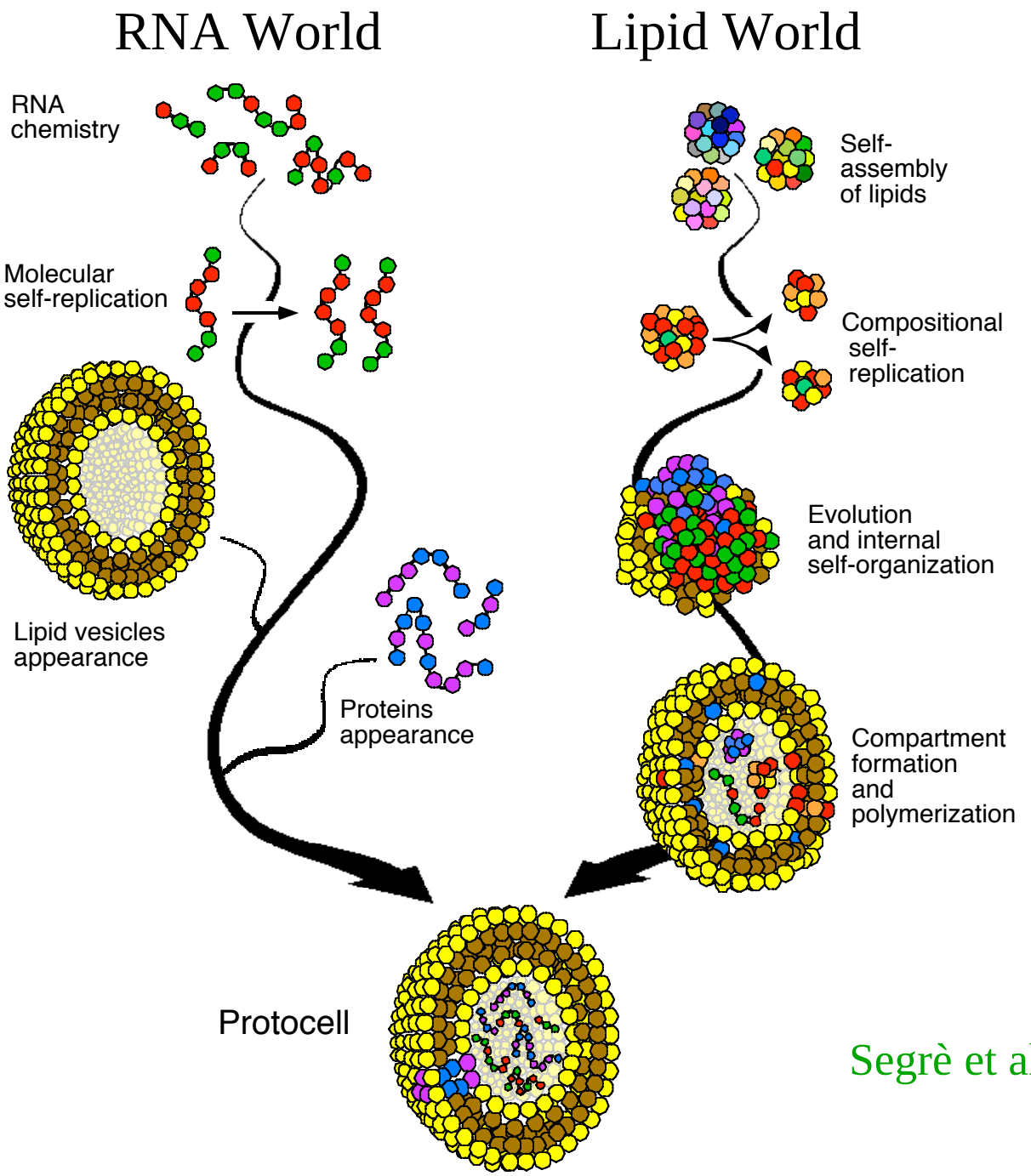
RNA





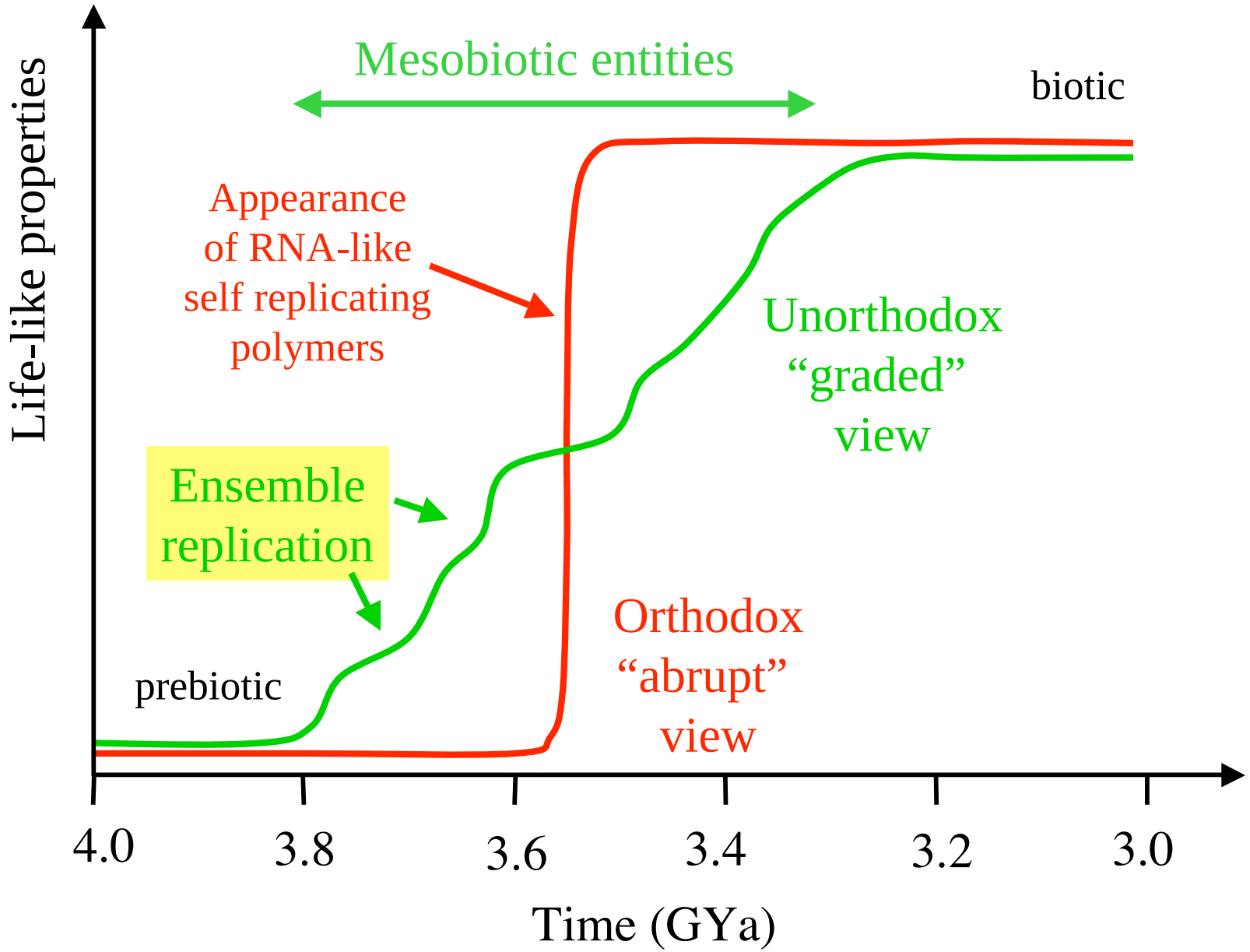


Protein

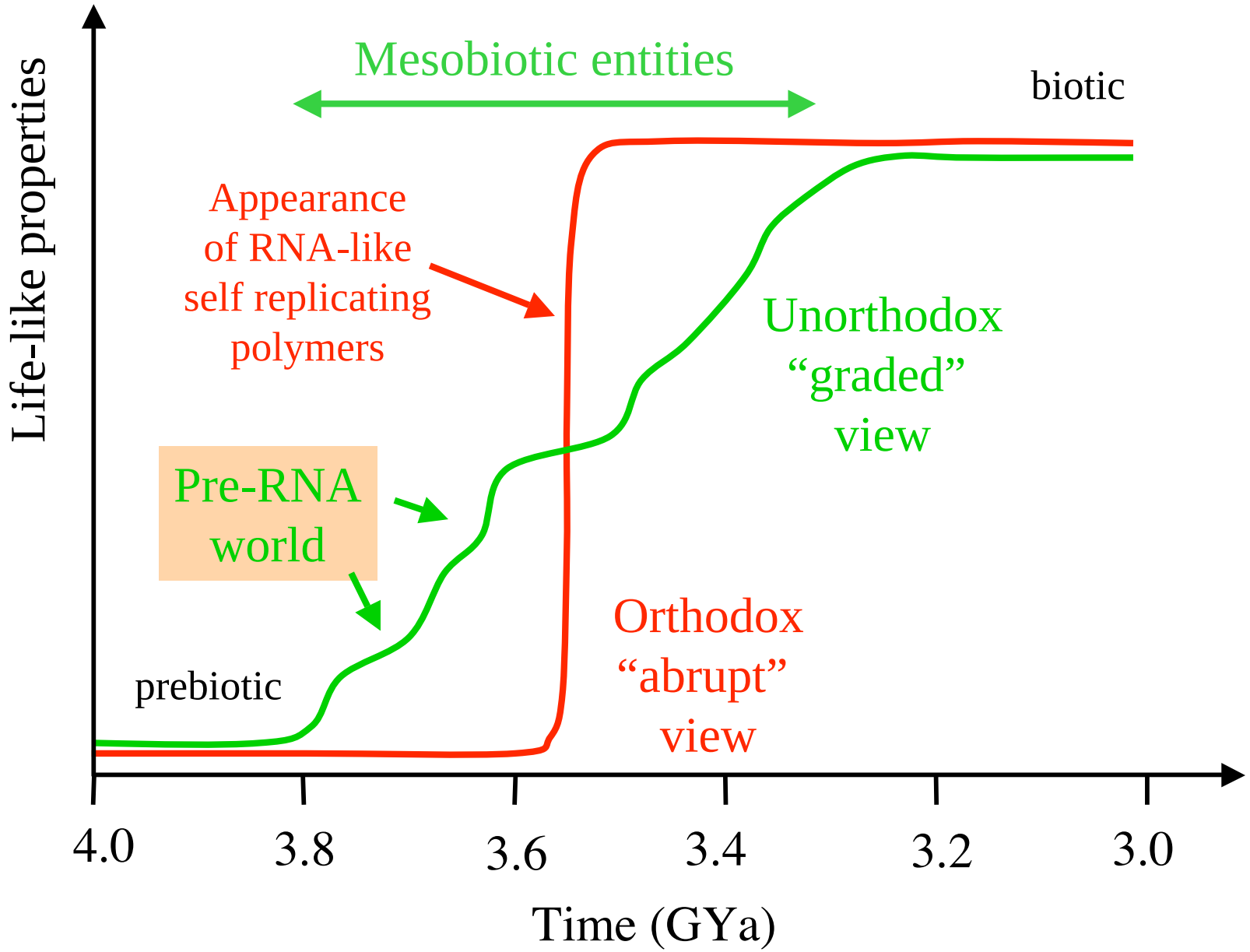


Segrè et al. EMBO Reports 2000

A graded appearance of life-like entities



A graded appearance of life-like entities



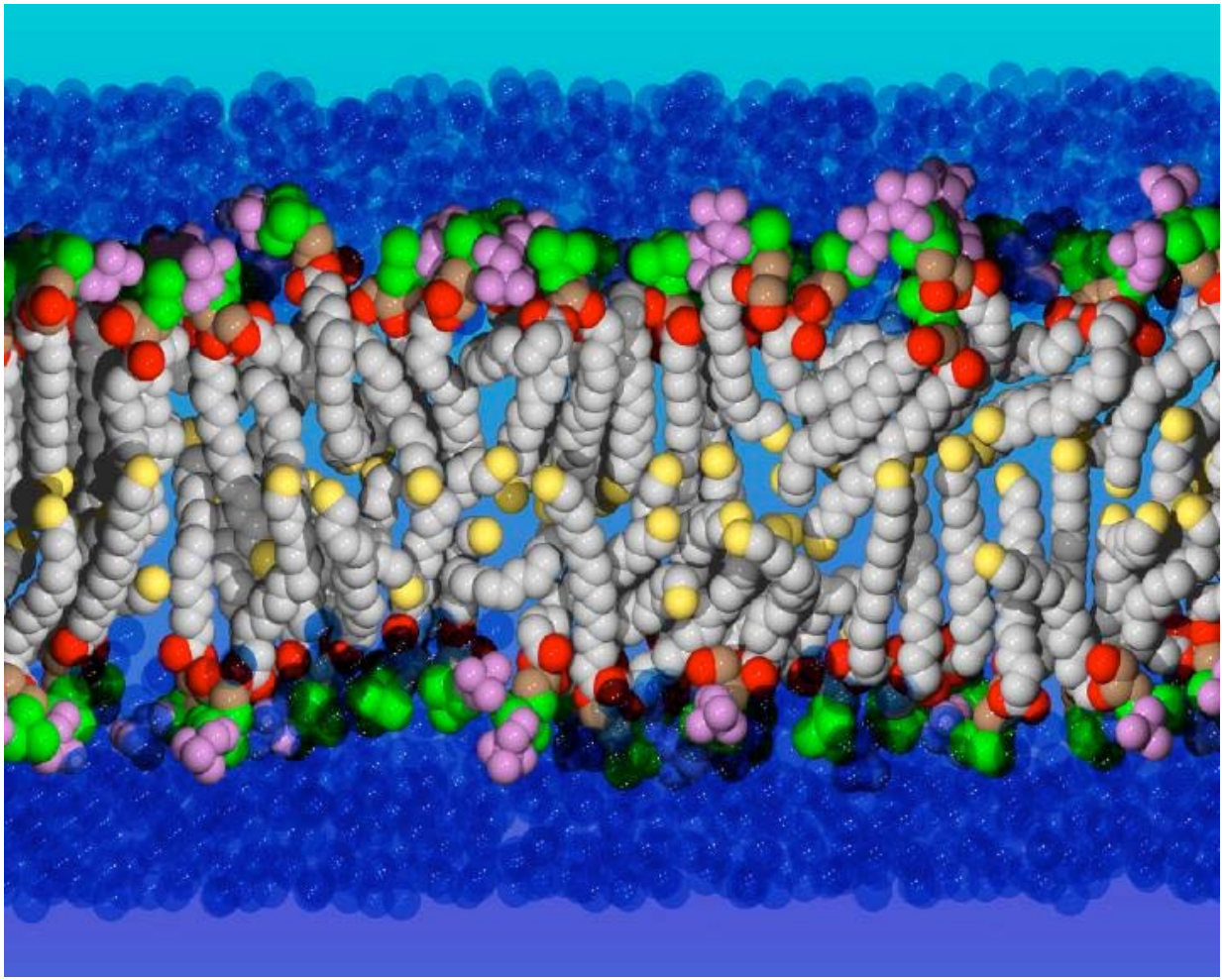
What is the real question?

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



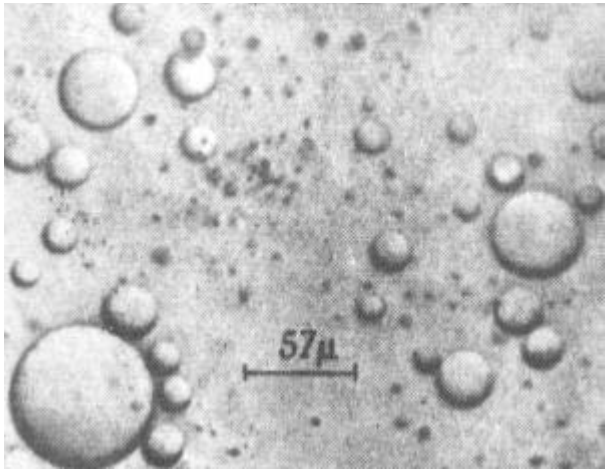
But could it have been different early on?

Alexander Oparin

"Origin of Life" 1924

Prebiotic "Soup"

Colloidal coacervates



"Metabolism first", not "genome first"

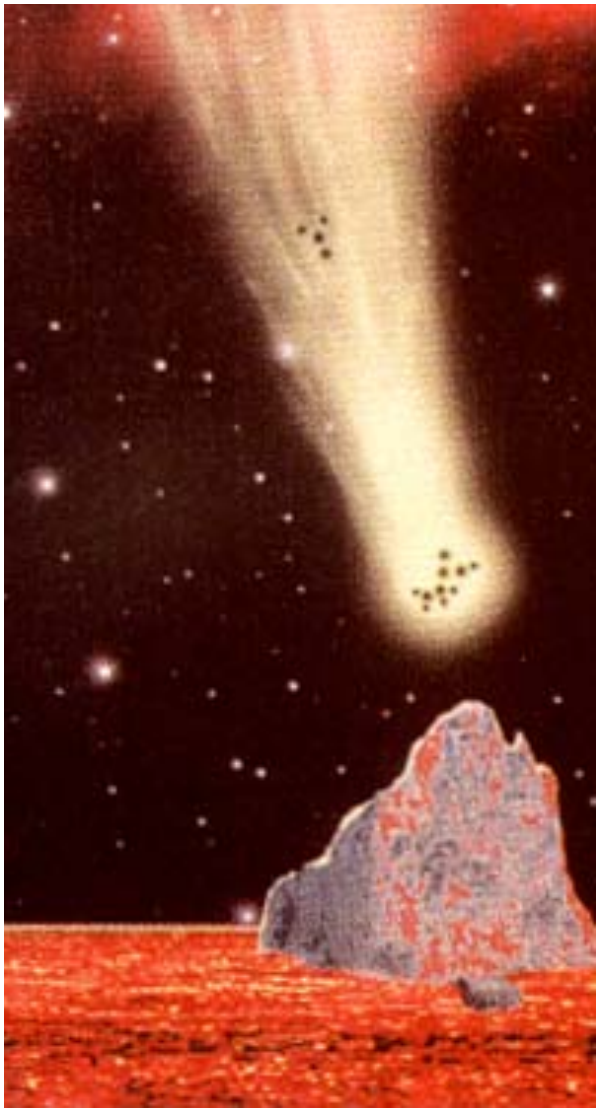
Replication without DNA

Miller SL, Schopf JW, Lazcano A.
Oparin's "Origin of Life": sixty years later.
J Mol Evol. 1997 Apr;44(4):351-3.

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

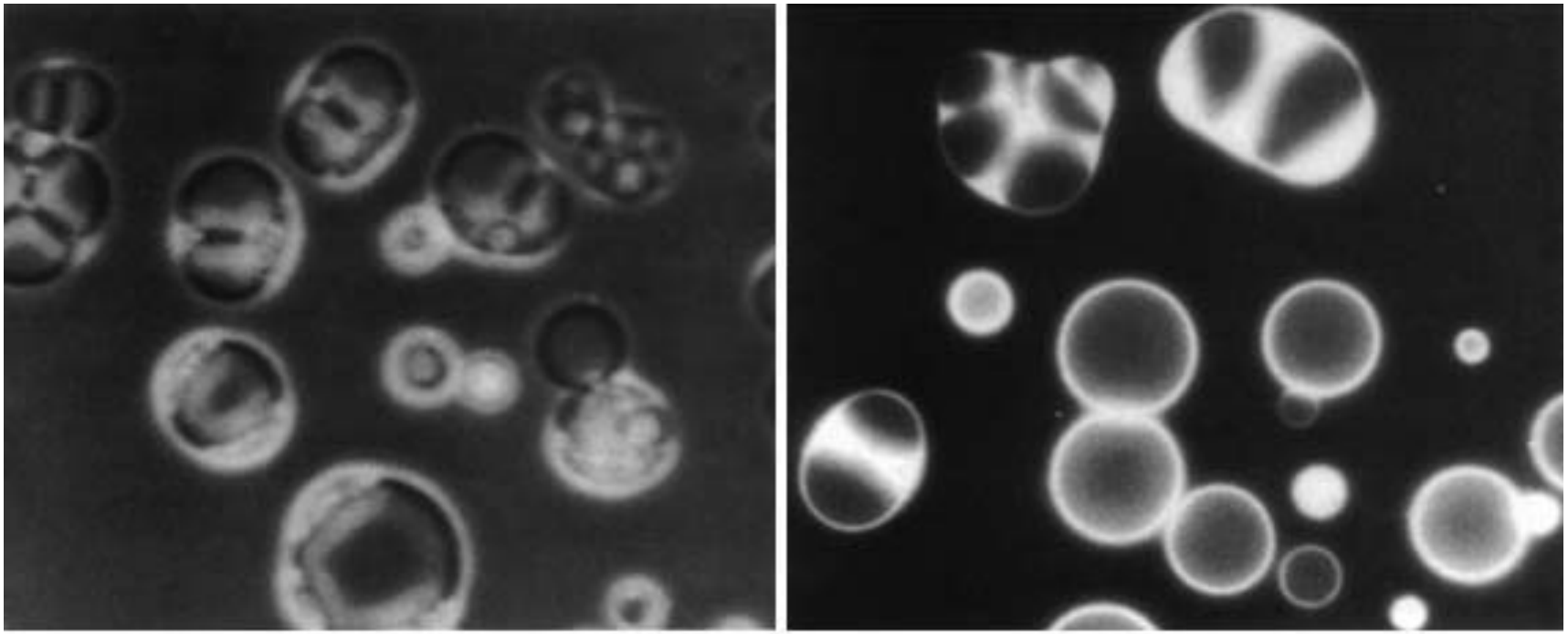
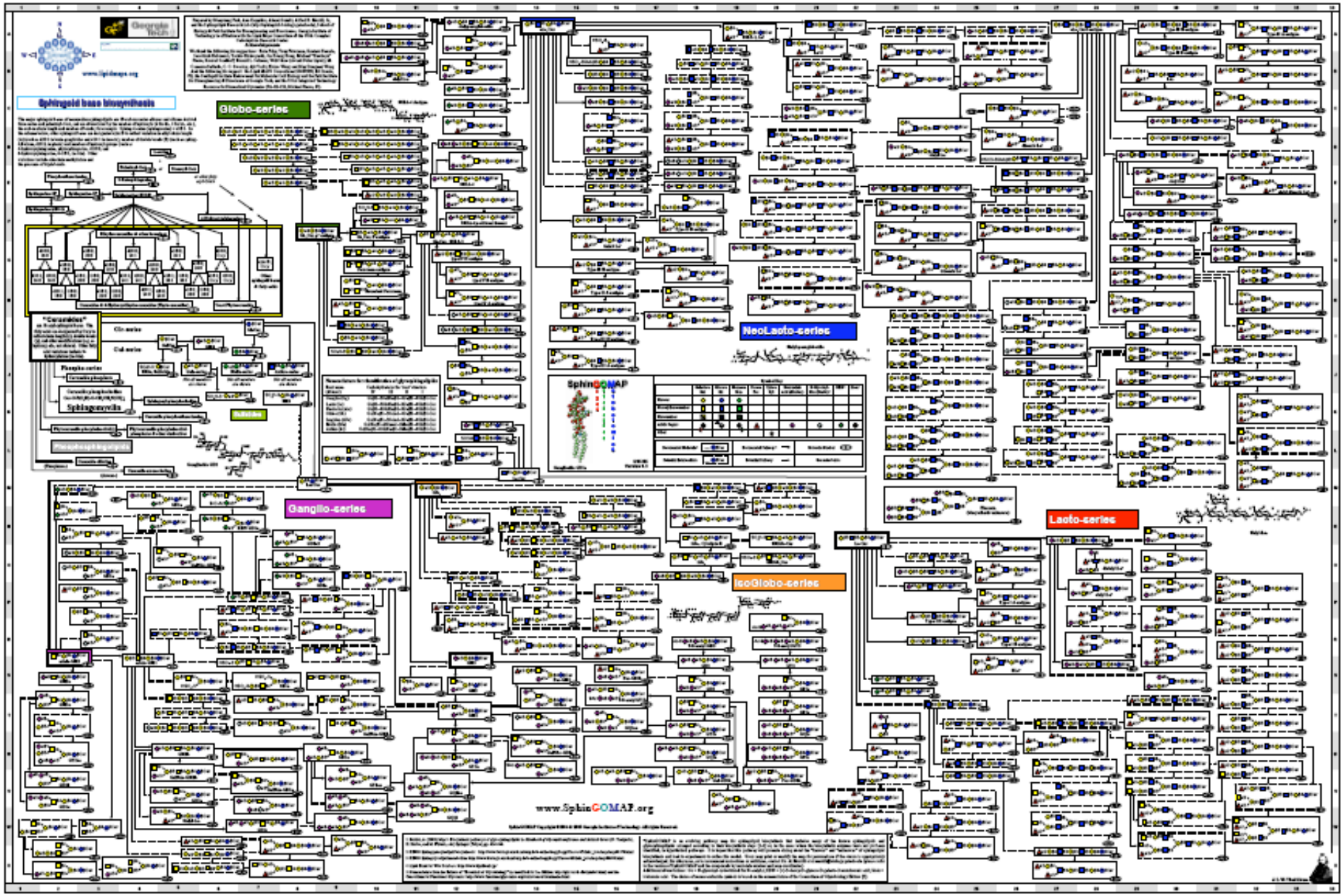


Figure 3. Self-assembled vesicular structures are produced by organic compounds extracted from the Murchison carbonaceous meteorite when they interact with water. The vesicles are 10-50 micrometers in diameter, and are bounded by bilayer membranes that can act as a diffusion barrier to ionic flux. Such relatively impermeable boundary structures are essential to the membranes that define all cellular life today. Left: phase micrograph. Right: light micrograph showing the natural fluorescence of the vesicles. The fluorescence is caused by polycyclic aromatic hydrocarbons that are abundant in carbonaceous meteorites. Original magnification: 400 X.

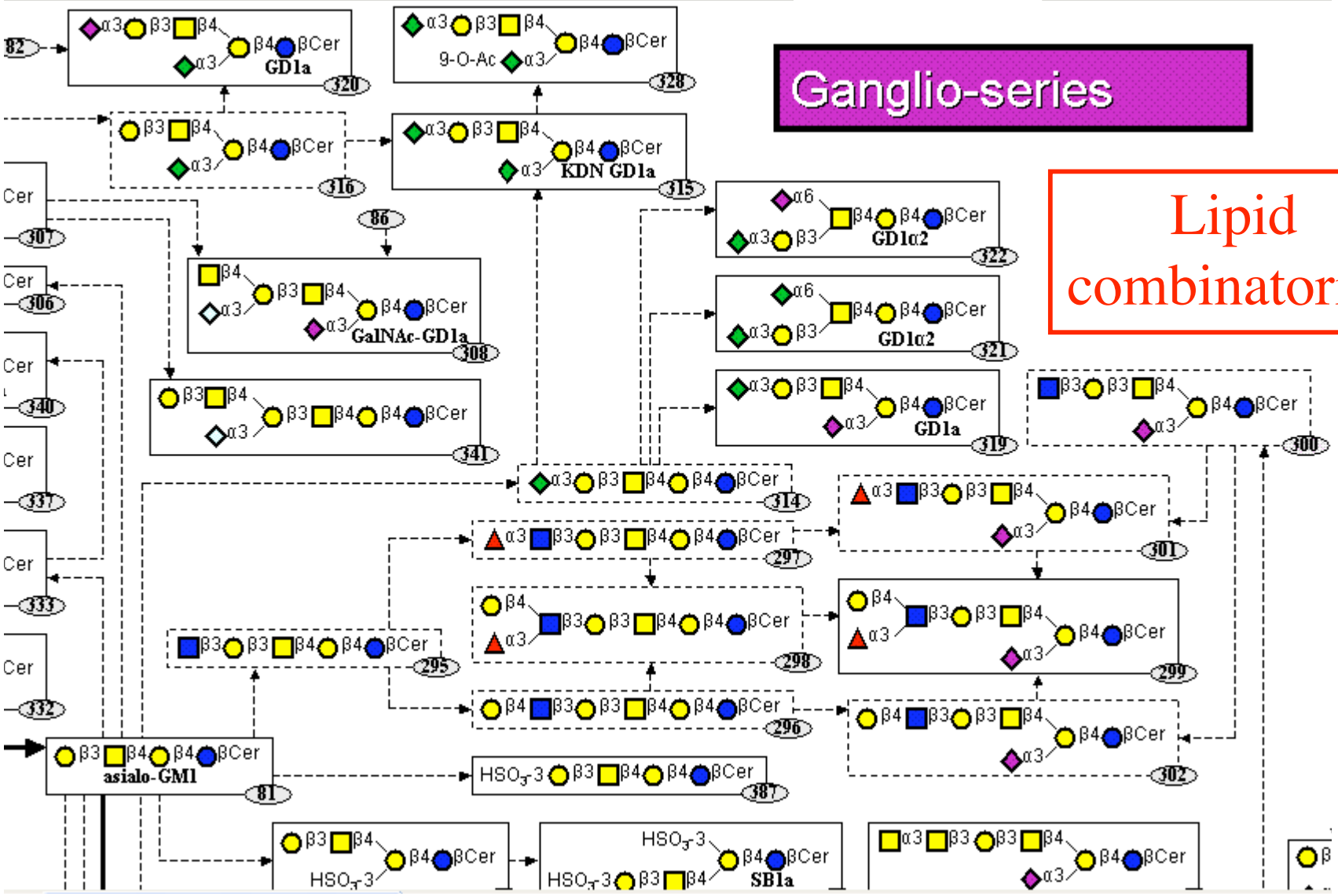
Meteorite material makes vesicles

Deamer and colleagues



Lipidomics: the diversity of present-day lipids

Alfred Merrill, Georgia Tech - SphingoMap



Alfred Merrill, Georgia Tech - SphingoMap - detail

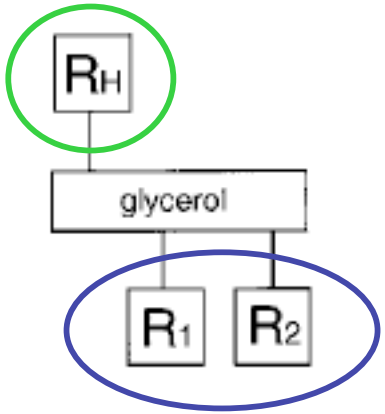
Lipid combinatorics

Lipid World

Head groups combinatorics

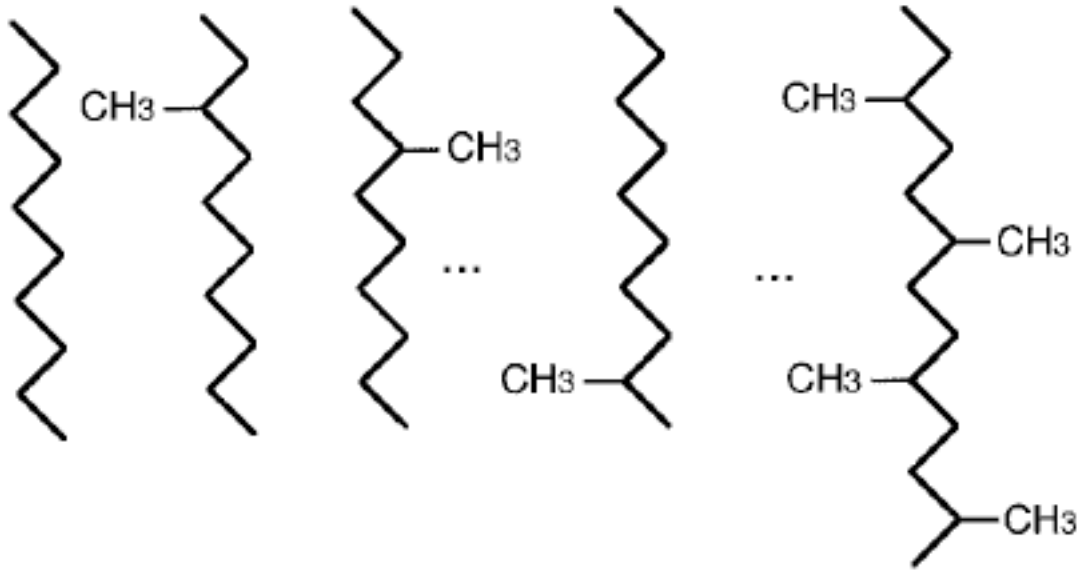
- peptides
- nucleotides
- phosphates, thiols
- metal chelators
- cofacotrs

etc

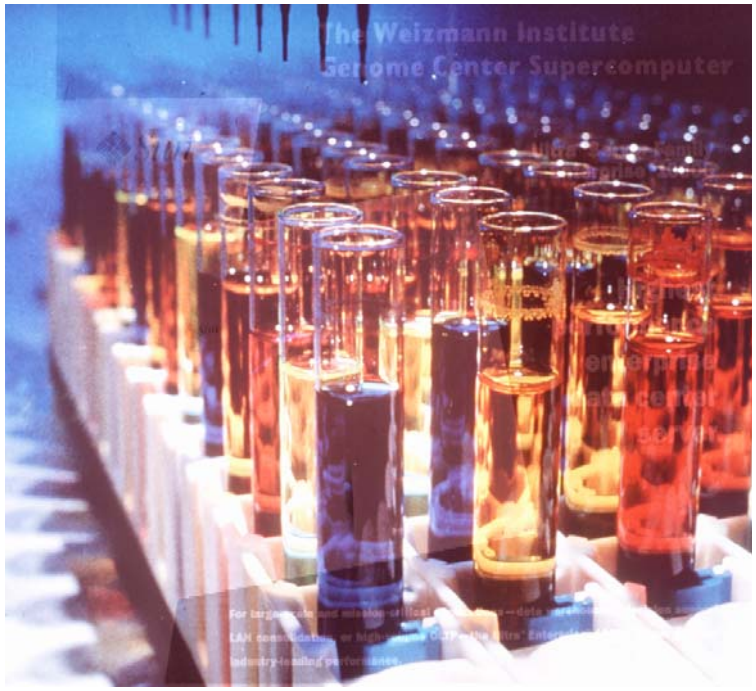


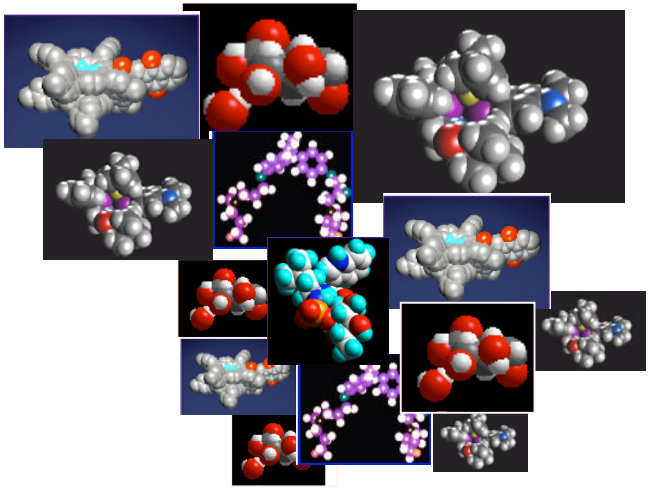
Tail combinatorics

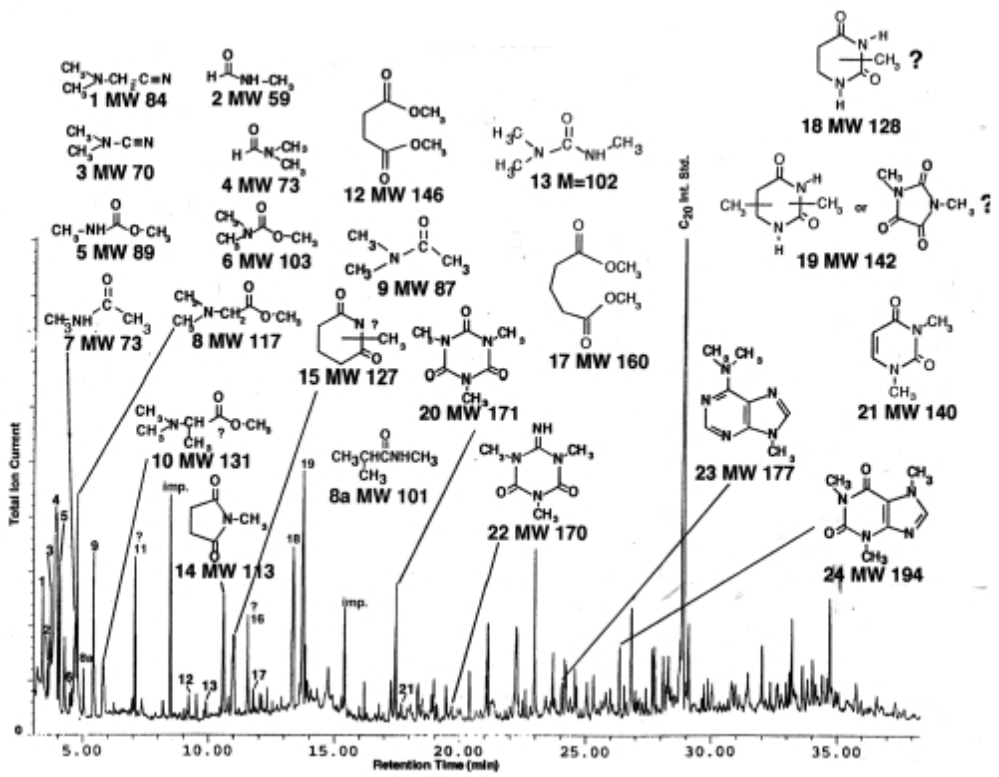
Planetary random chemistry



Life's origin: a planet-scale random chemistry "experiment"!





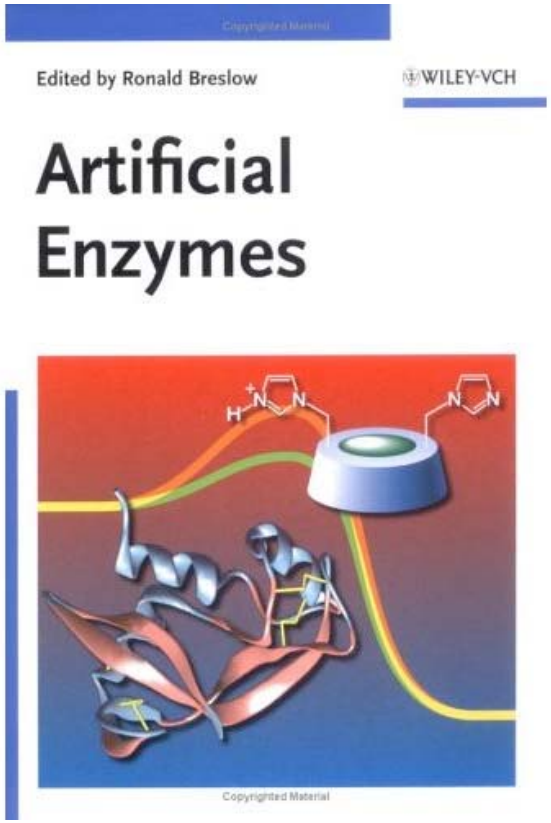


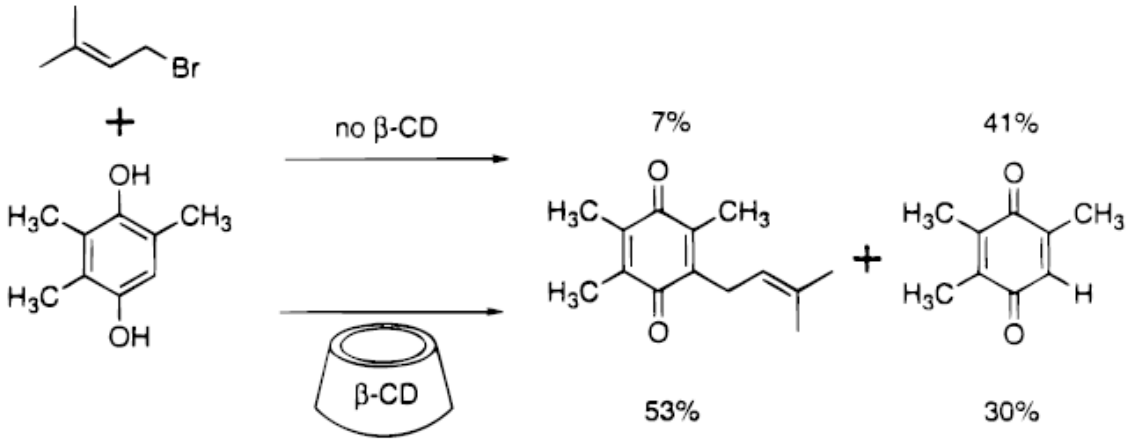
Diversity is cheap:
Billions of different organic compounds may form spontaneously!

What is the real question?

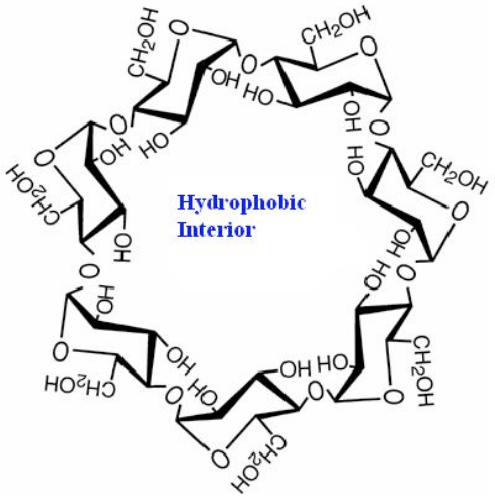
1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today's chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?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

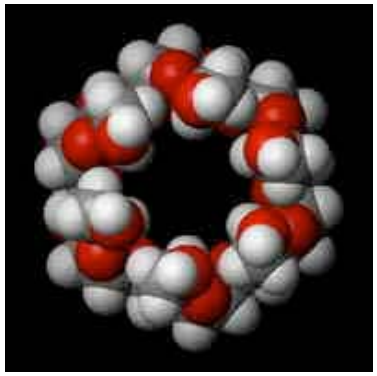
Cyclodextrins act as artificial enzymes





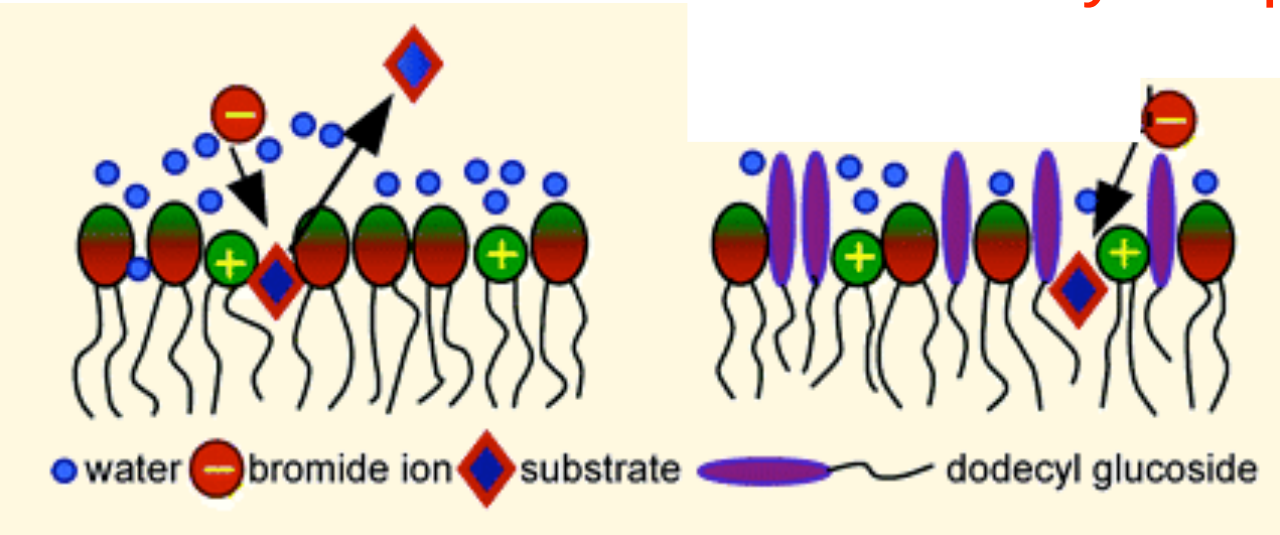
Cyclodextrin: cyclic oligosaccharide





Ronald Breslow
Department of Chemistry
Columbia University

Membrane mimetic chemistry – Lipid catalysis



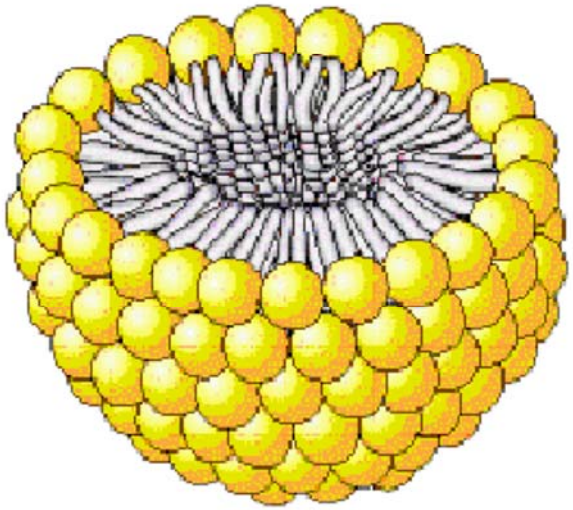
Vesicular Catalysis of an SN2 Reaction.
Jaap Klijn Jan Engberts
Langmuir; 2005; 21(22) pp 9809 - 9817;

Janos H. Fendler
Dept of Chemistry
Clarkson University,
Potsdam, New York

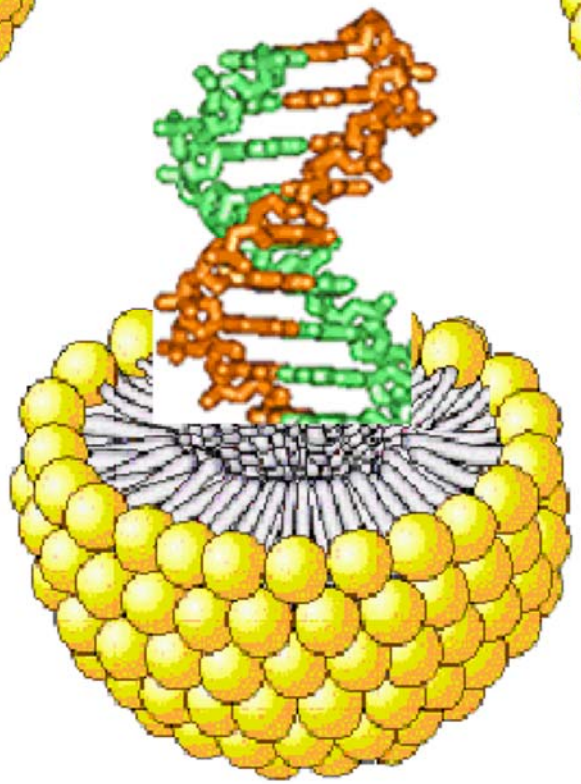
Catalysis in Micellar and Macromolecular Systems J. H. Fendler and E. J. Fendler, Academic Press, 1975

Zepik, H.H., Maurel, M.-C. & Deamer, D.W. (2004). Lipid catalysis of oligomerization of amino thioacids and thioesters
International Journal of Astrobiology, Supplement 1 (March): 105.

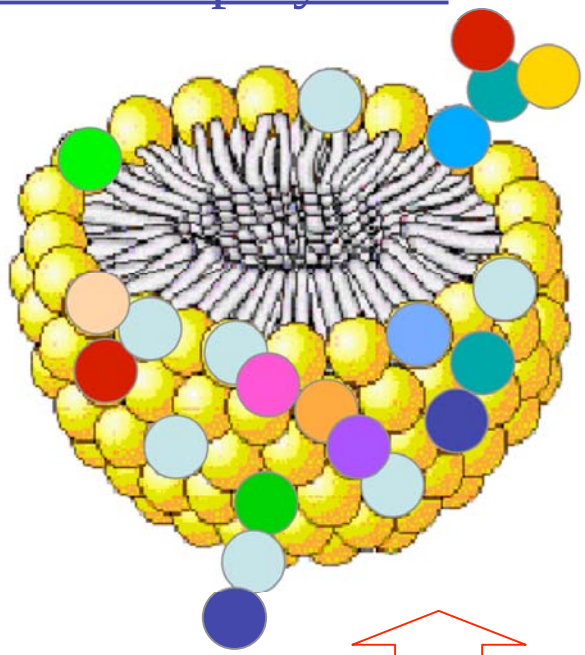
1) Lipid micelle:
Container only



2) Micelle
with trapped
polynucleotide:
Container +
sequence
information

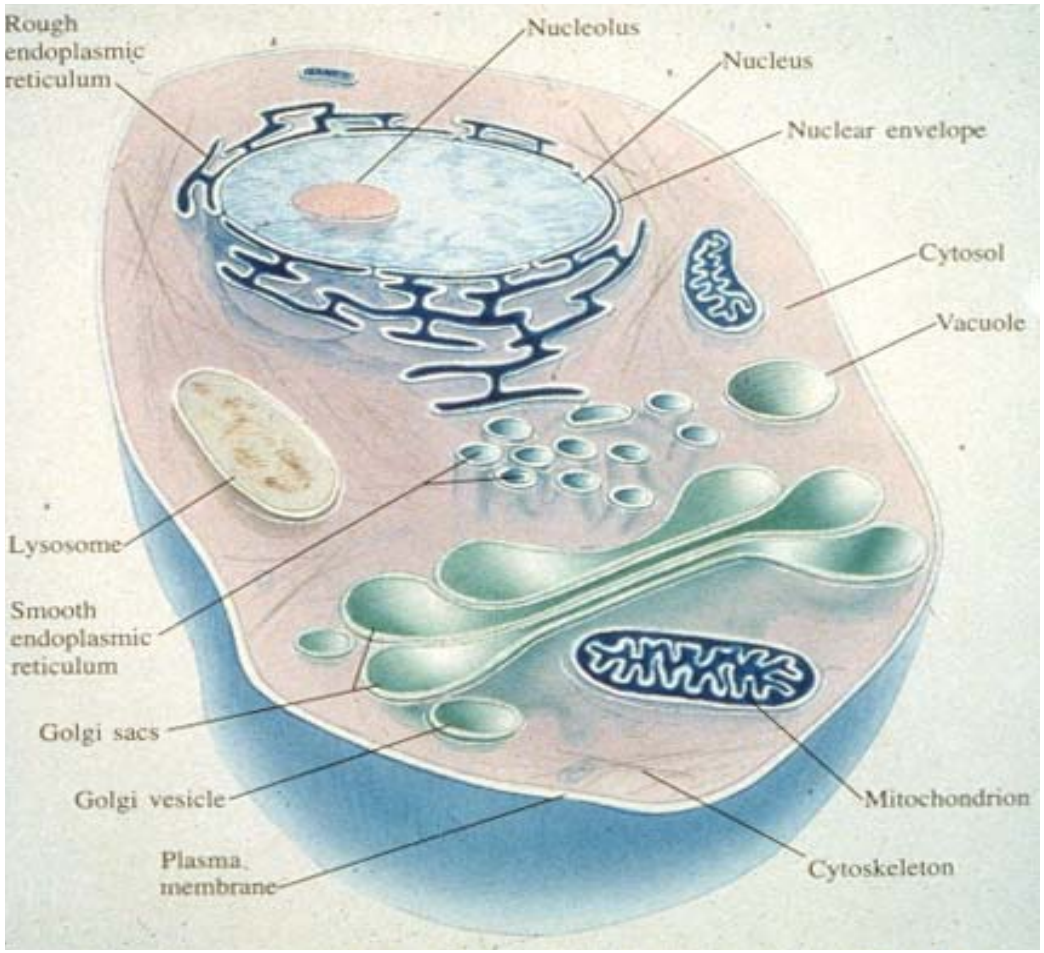


3) Mixed micelle: Assembly with compositional information and no informational biopolymers

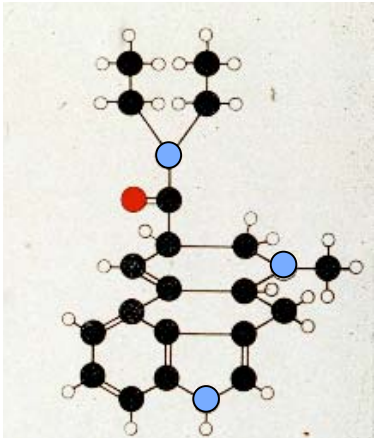


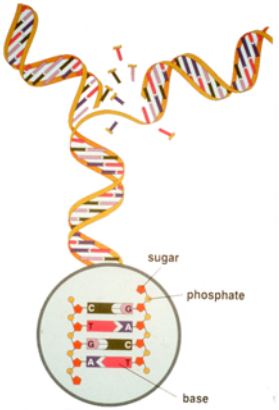
GARD
Lipid World

Two types of complexity in any living cell

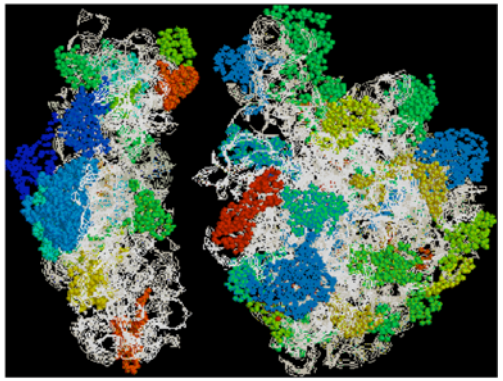


1) Molecular complexity



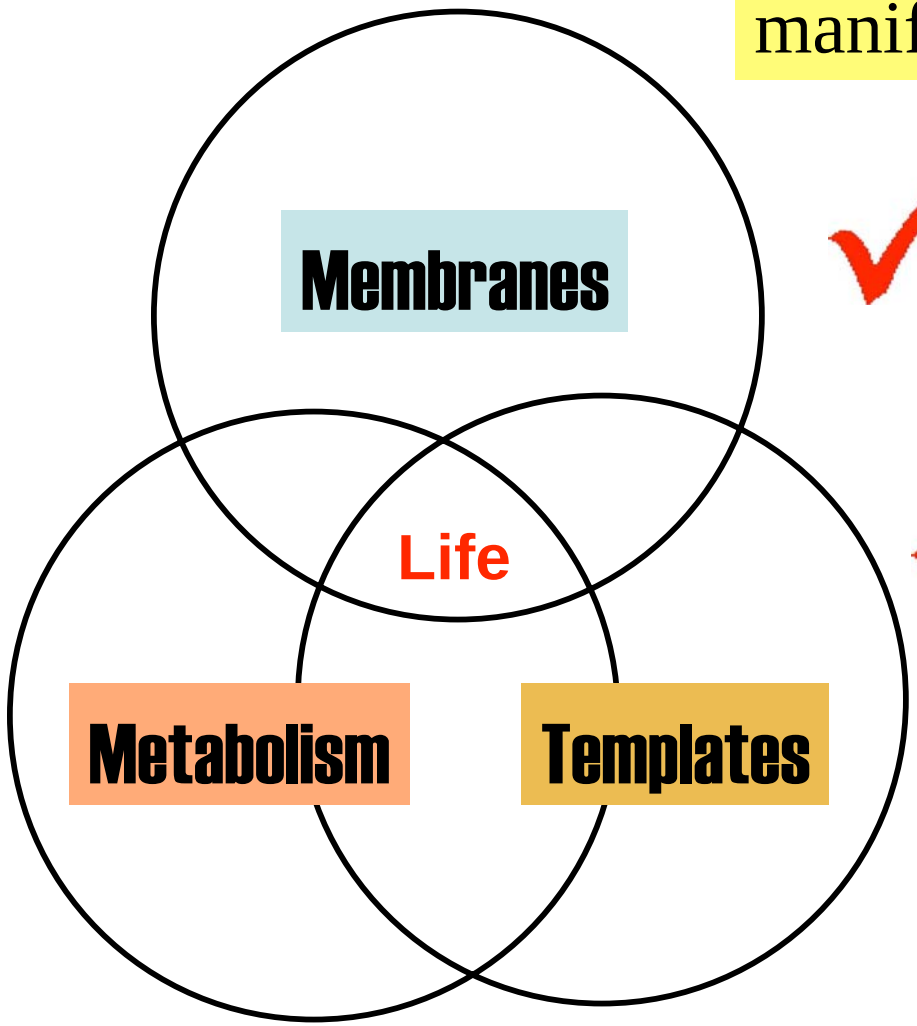


2) Ensemble complexity

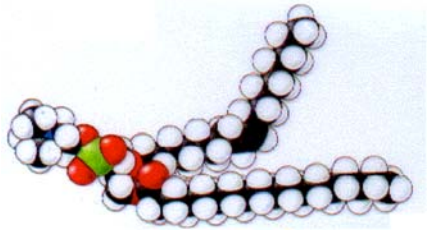


The GARD model is strongly based on ensemble complexity!

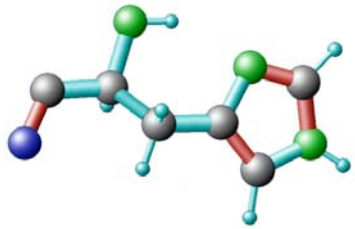
Can all three aspects of life be manifested by monomers only?



✓ enclosure



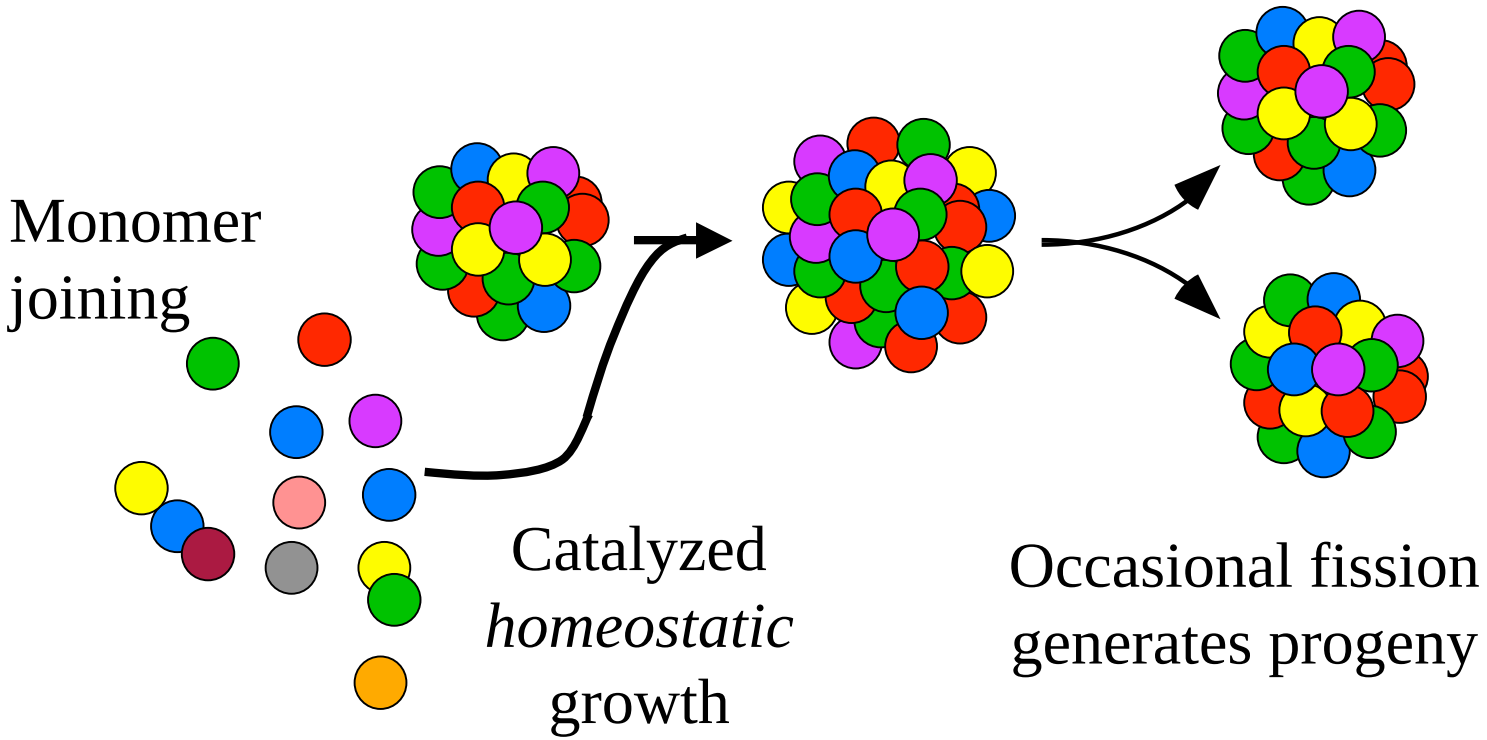
✓ Catalysis



Histidine

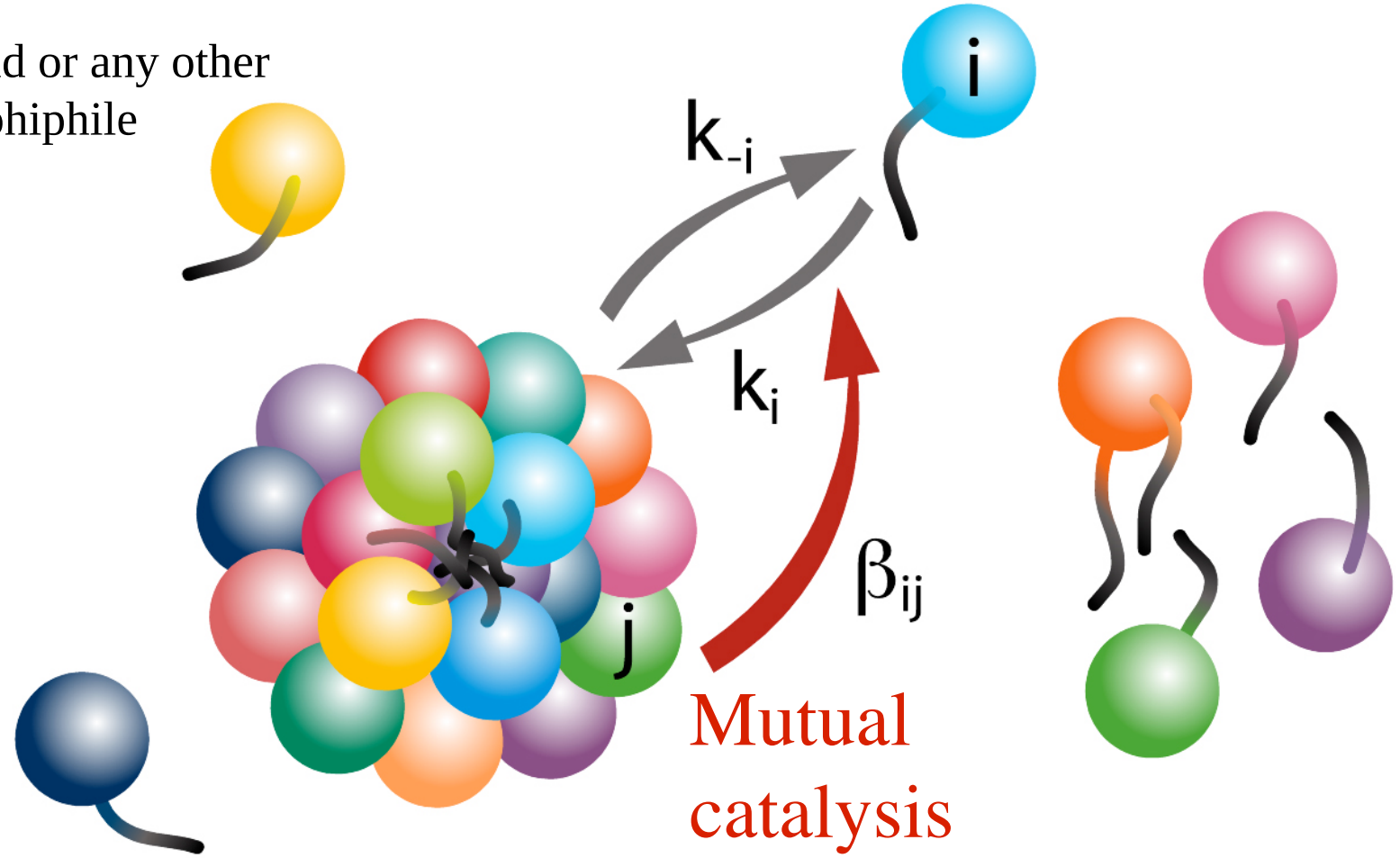
? Information

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



Lipid catalysis, Monomers only, Compositional information

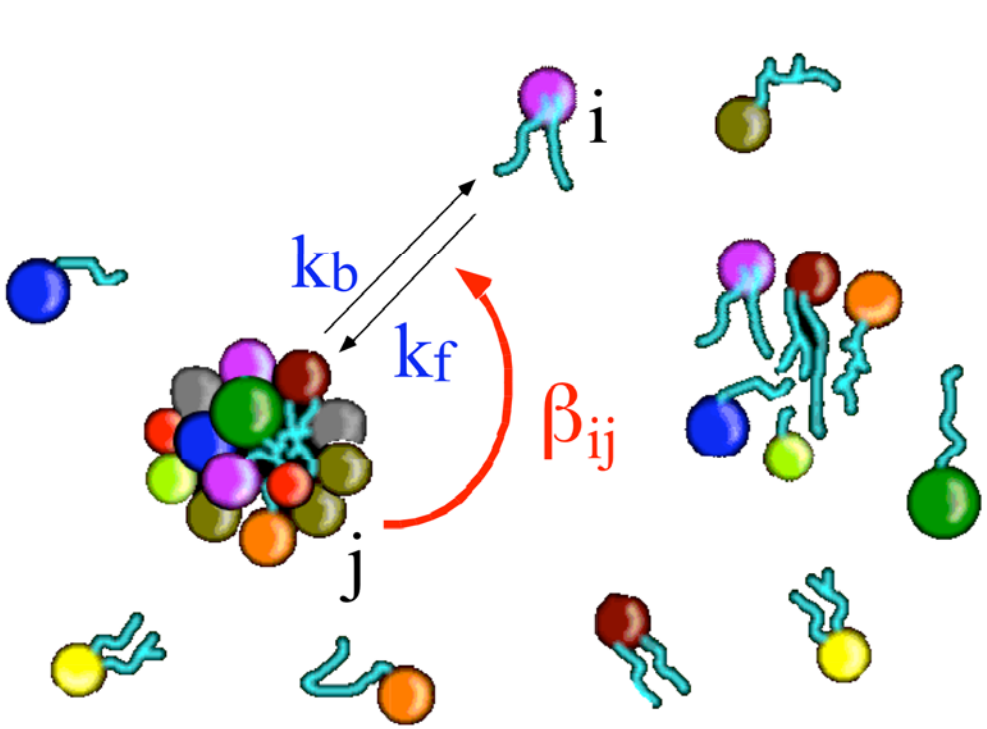
GARD - formal definition

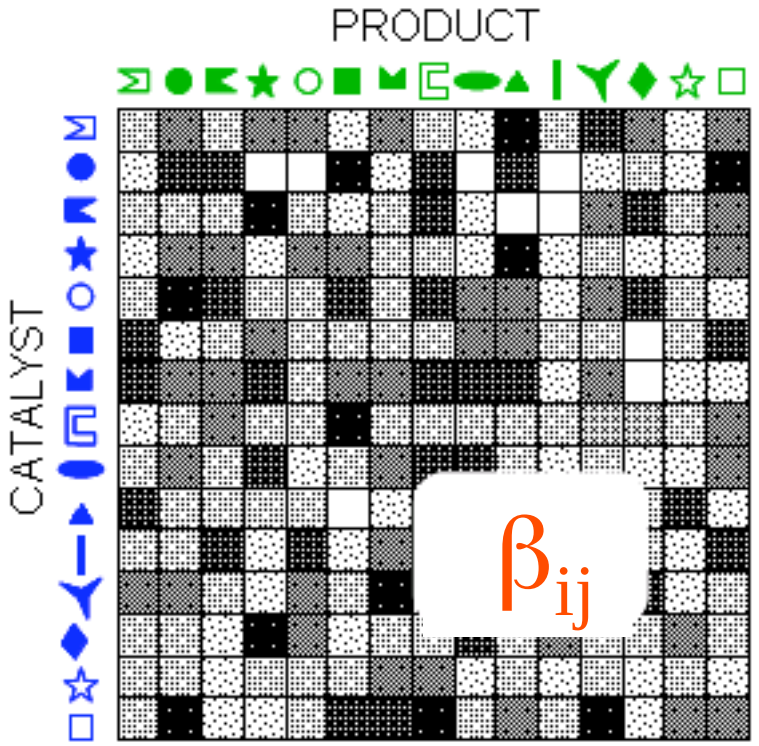


Differential equation set

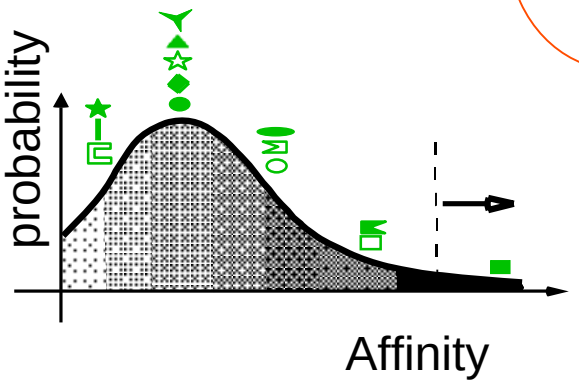
$$\frac{dn_i}{dt} = (k_i\rho_i - k_{-i}n_i)(1 + \Sigma\beta_{ji}n_j)$$

The GARD model: governed by drug-related statistics



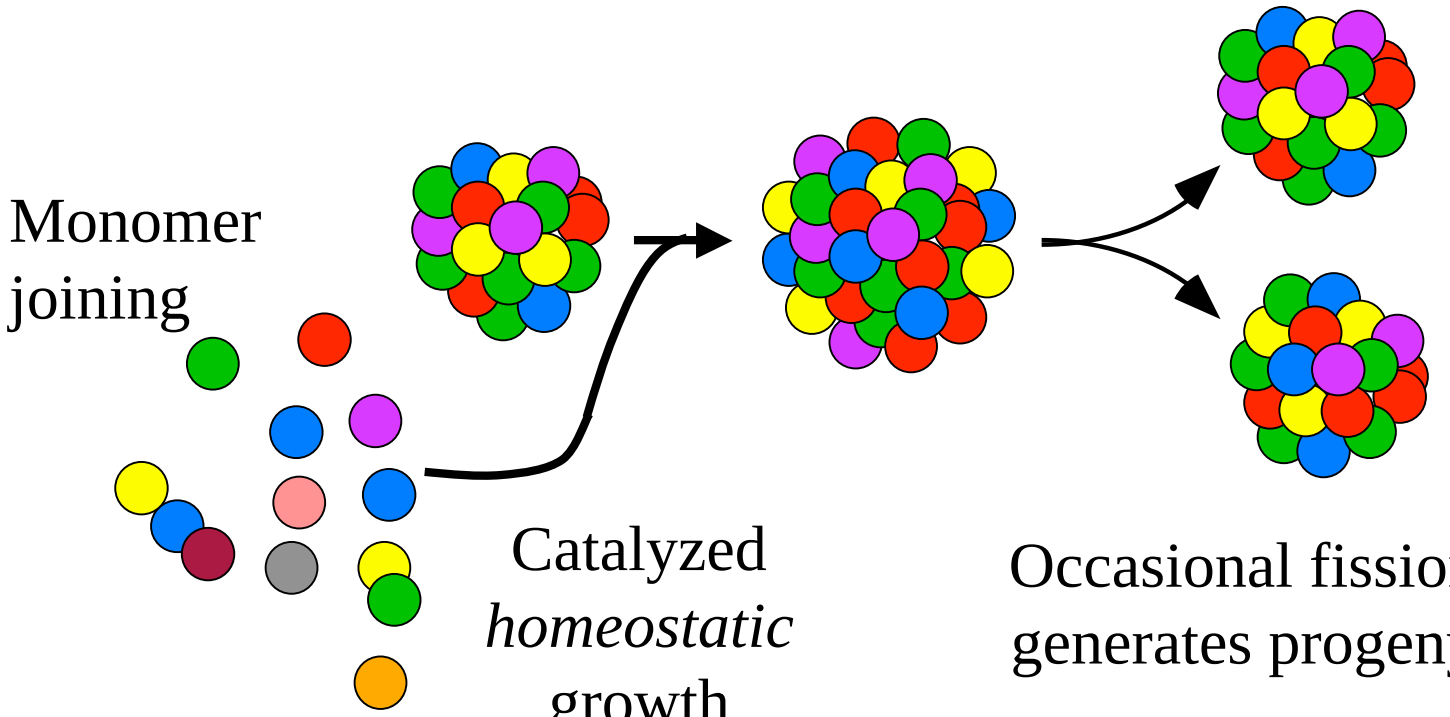


Mutual catalysis matrix

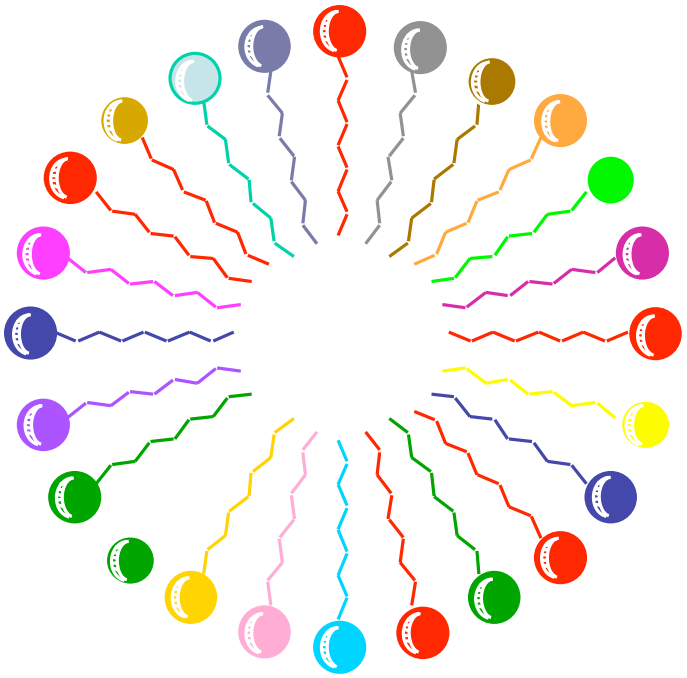


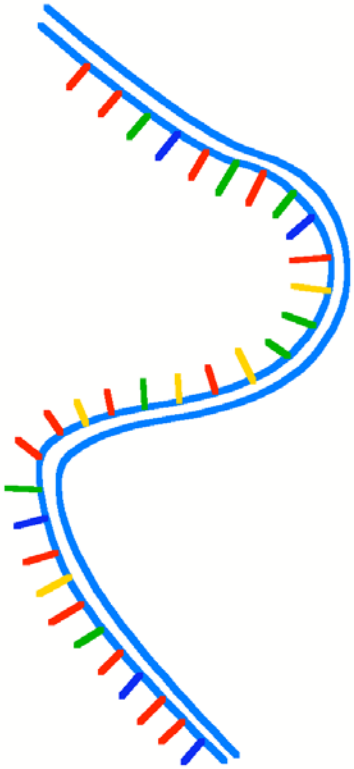
The Receptor Affinity Distribution model

GARD inheritance: what is propagated?



Compositional information

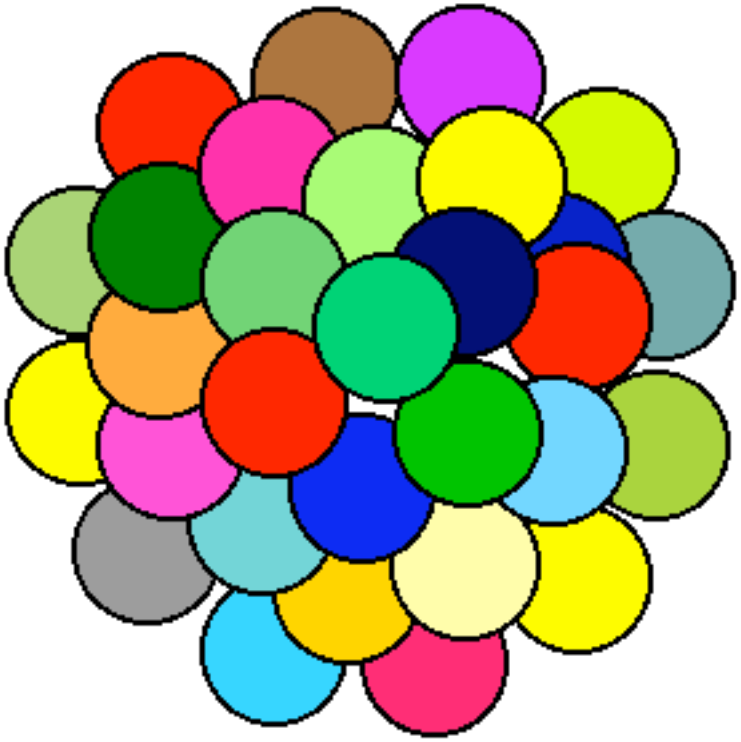




Lipid World
Combinatorial assemblies

RNA World:
Combinatorial Sequence

Compositional information



$n_{\bullet} = 3$

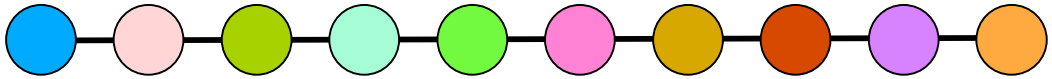
Consider a repertoire of N_G different molecule types

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

Comparing compositional to sequence information

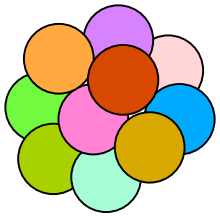
Alphabet of $N_G = 20$ monomers



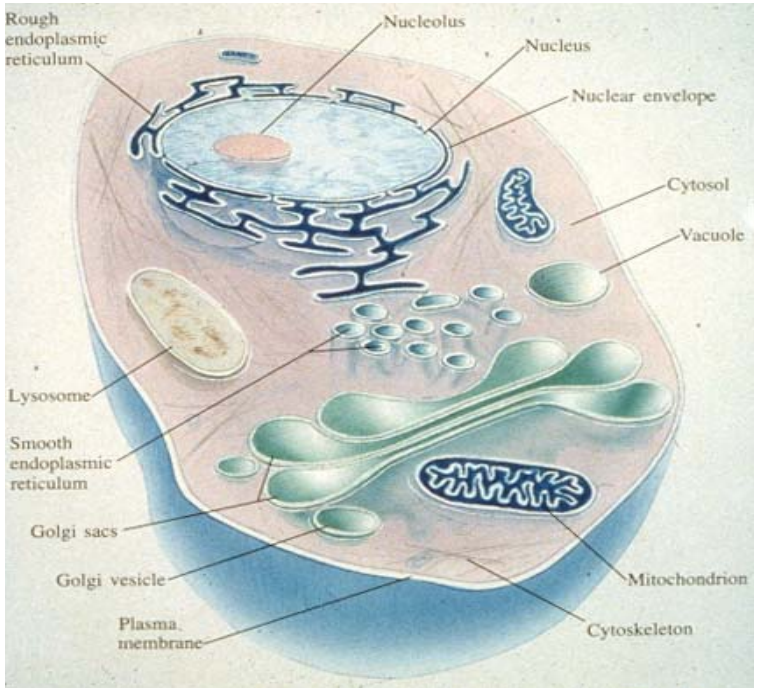
A sequential oligomer with N = 10 units can be constructed in 20^{10} different ways → has $\log_2(20^{10}) \approx 43$ bits of information

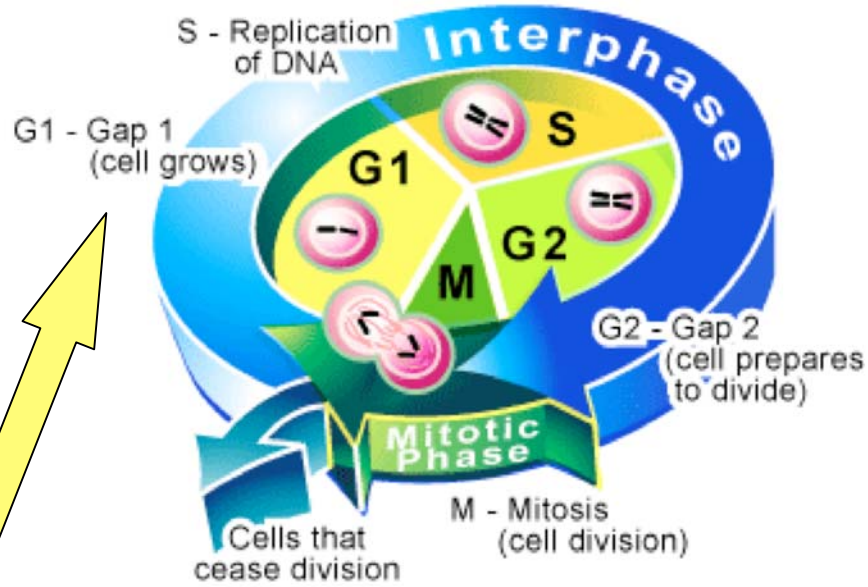
A compositional assembly with N = 10 units from the same alphabet can be constructed in $2X10^7$ different ways → has $\log_2(2X10^7) \approx 24$ bits of information

$$\log_2 \binom{N_G+N-1}{N}$$



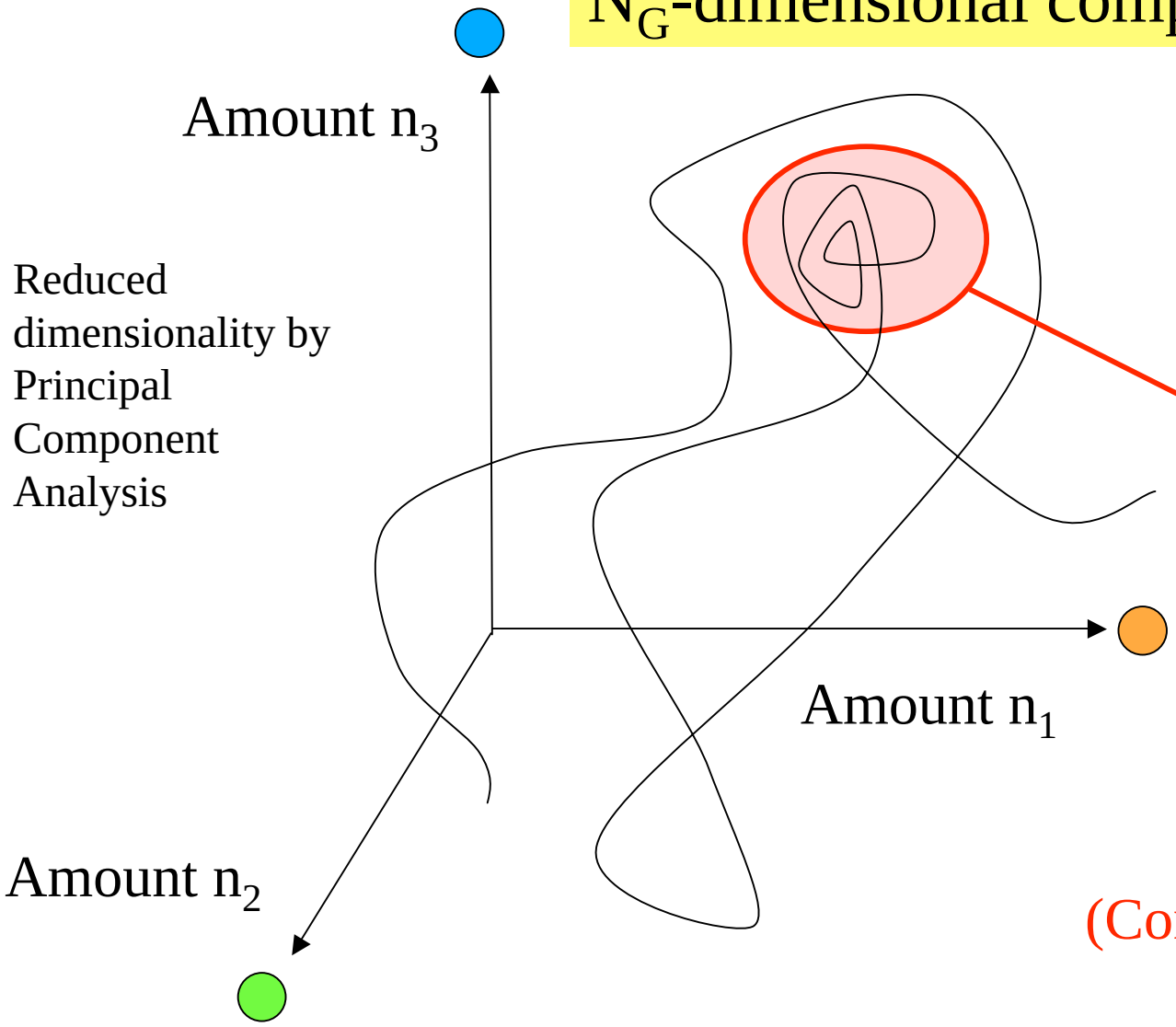
Compositional information is copied prior to cell division





To allow cell division, before DNA replicates, new copies of all the molecules in the cells need to be procuced.

GARD dynamics: Trajectory in a N_G-dimensional compositional space

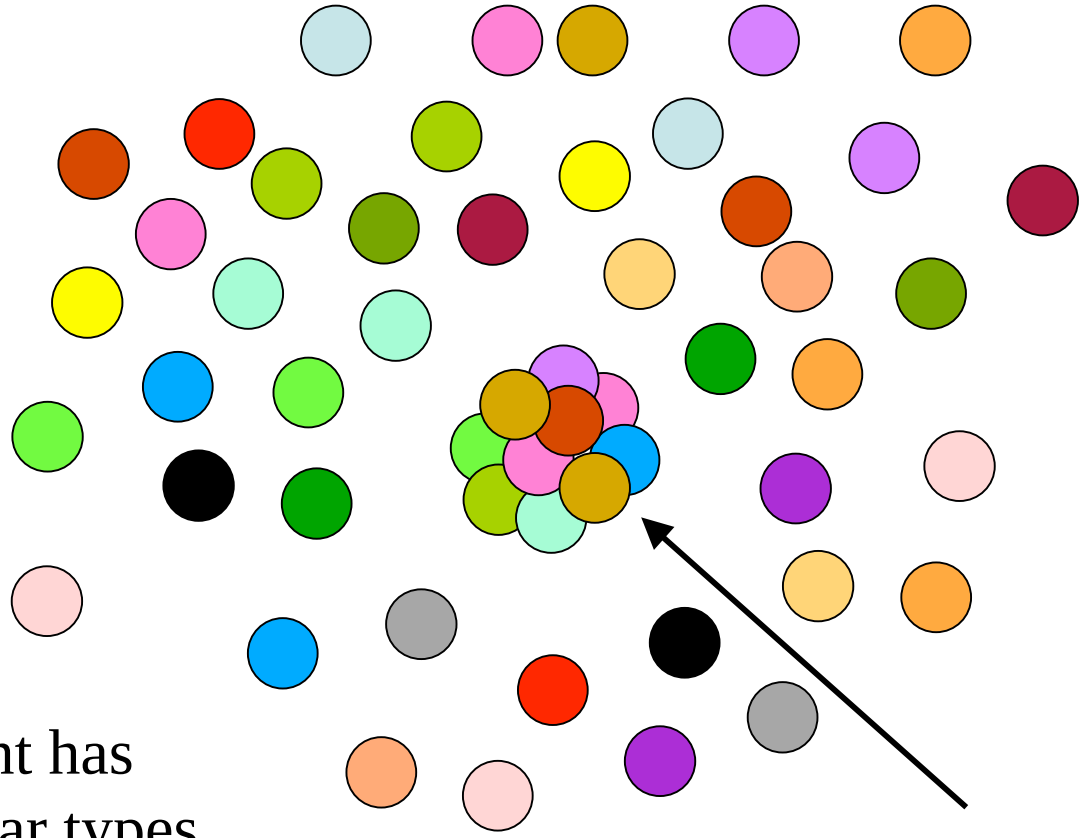


Fixed point or quasi-stationary state

Composome

(Compositional genome)

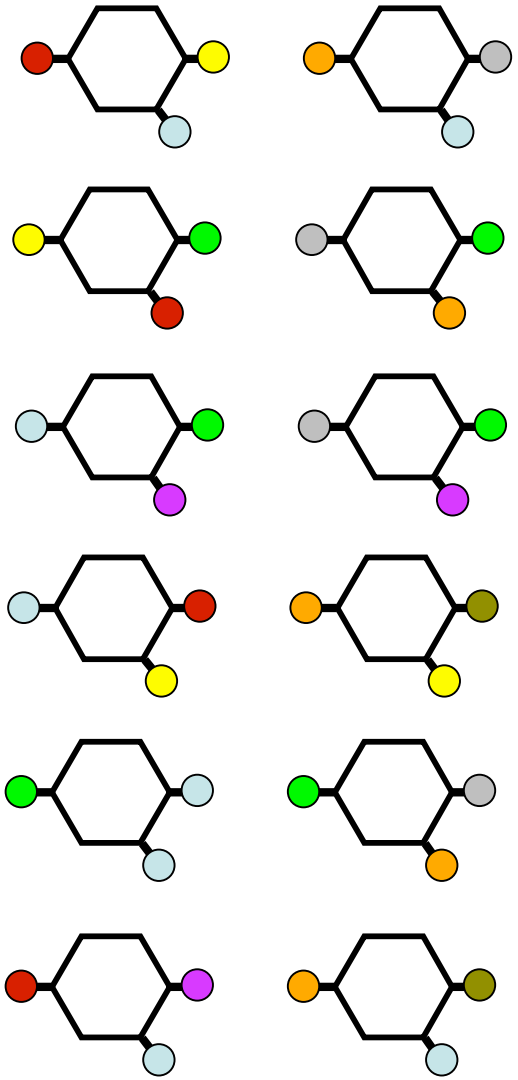
A composome has reduced repertoire



Environment has
N_G molecular types

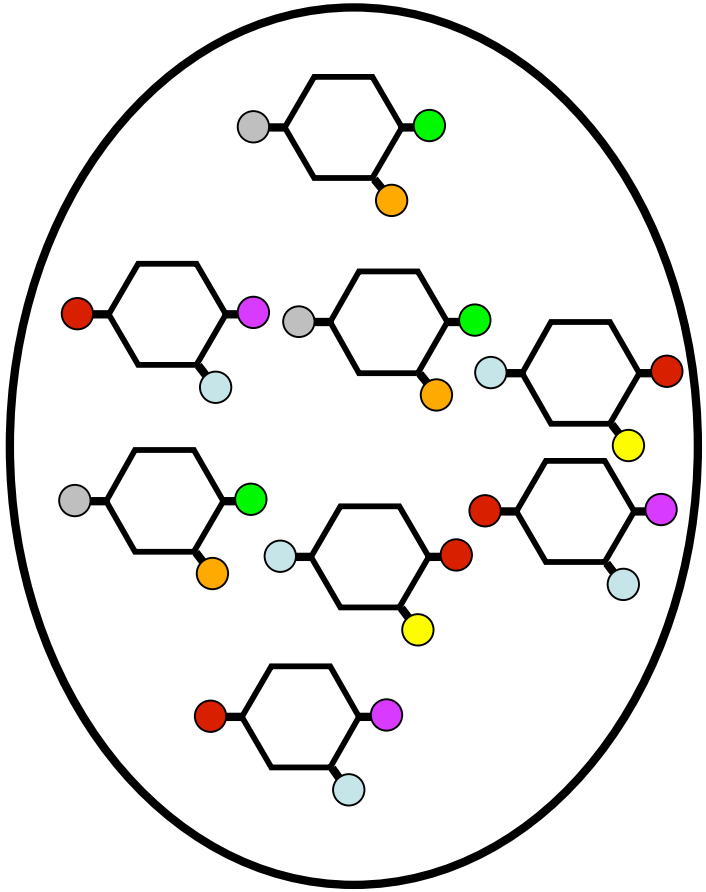
Composome has
$N_e << N_G$
molecular types

Prebiotics: a random library of chemicals

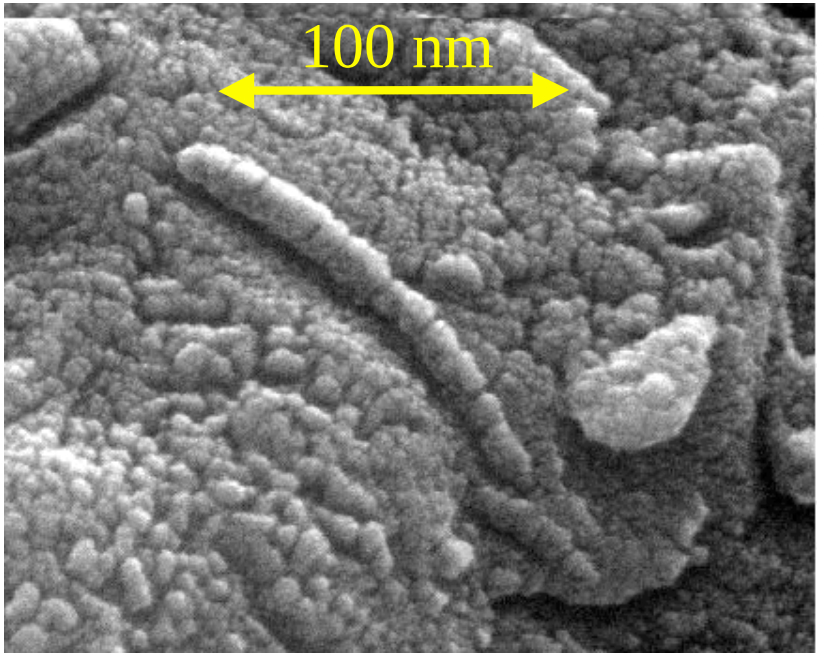


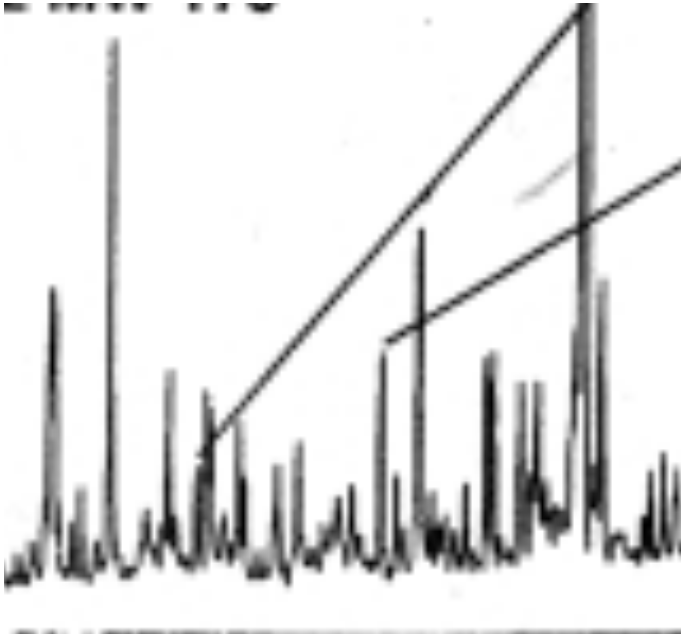
GARD may explain reduction of possibilities

Biotics:
A miniscule fraction of the possible molecules



“GARDobes” on Mars??



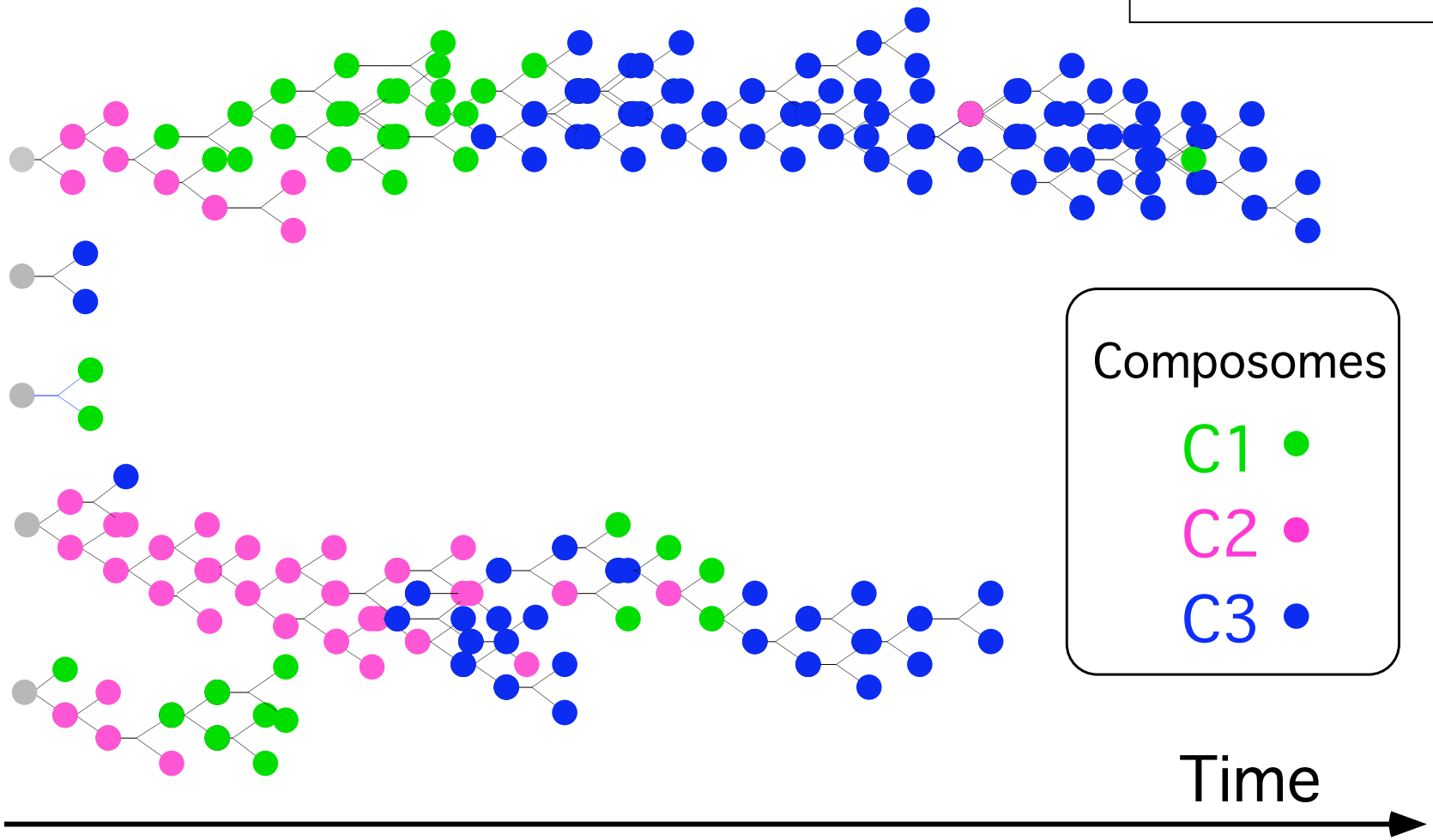


Not too small for early protocells!

Restricted PAH spectrum
Consistent with GARD concepts

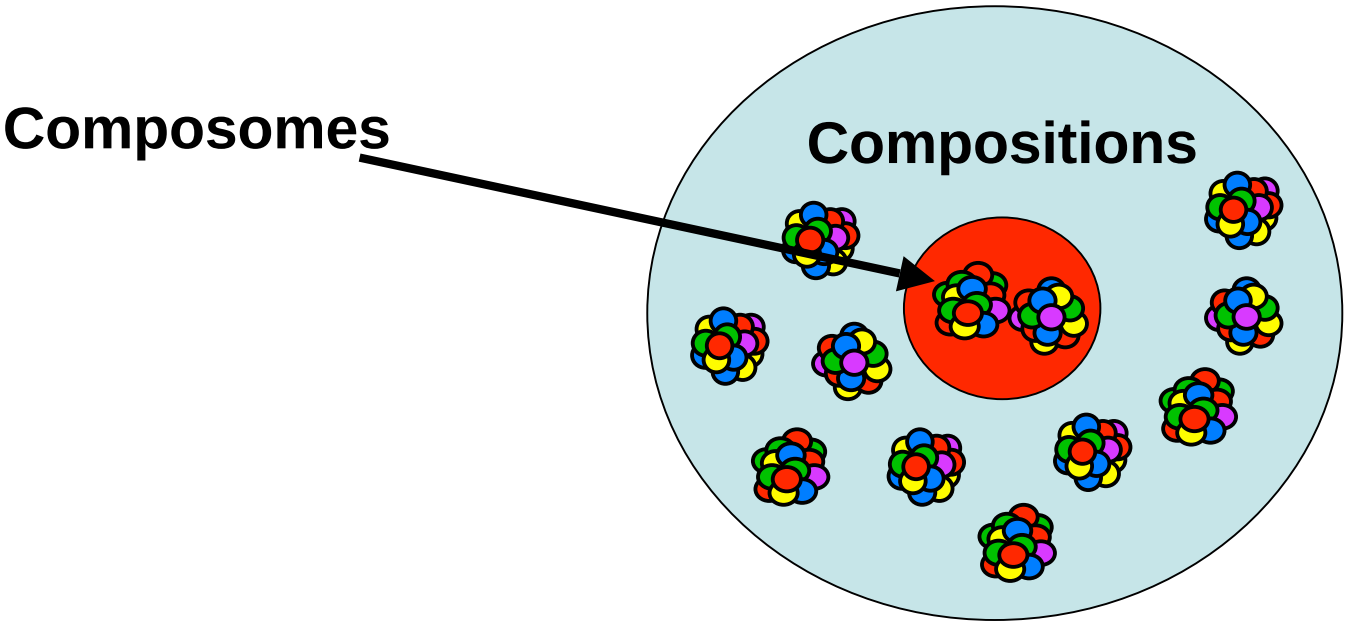
Composome evolution under a Constant Population (CP) constraint

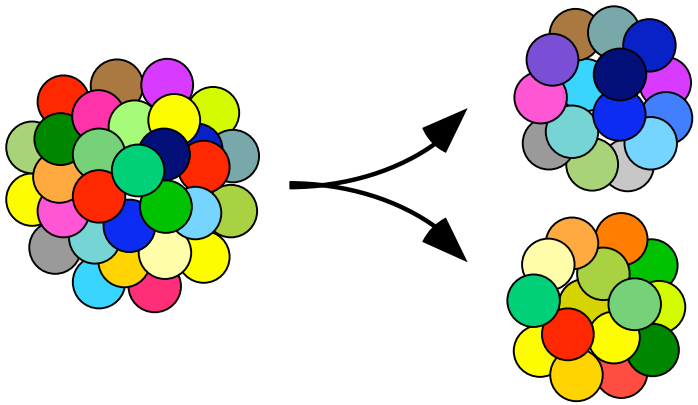
Selection
Evolution



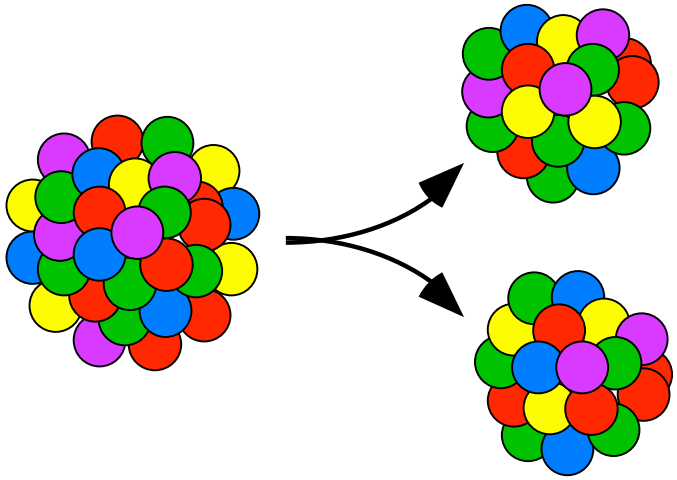
Segre et al., PNAS 2000

Only few compositions are 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?

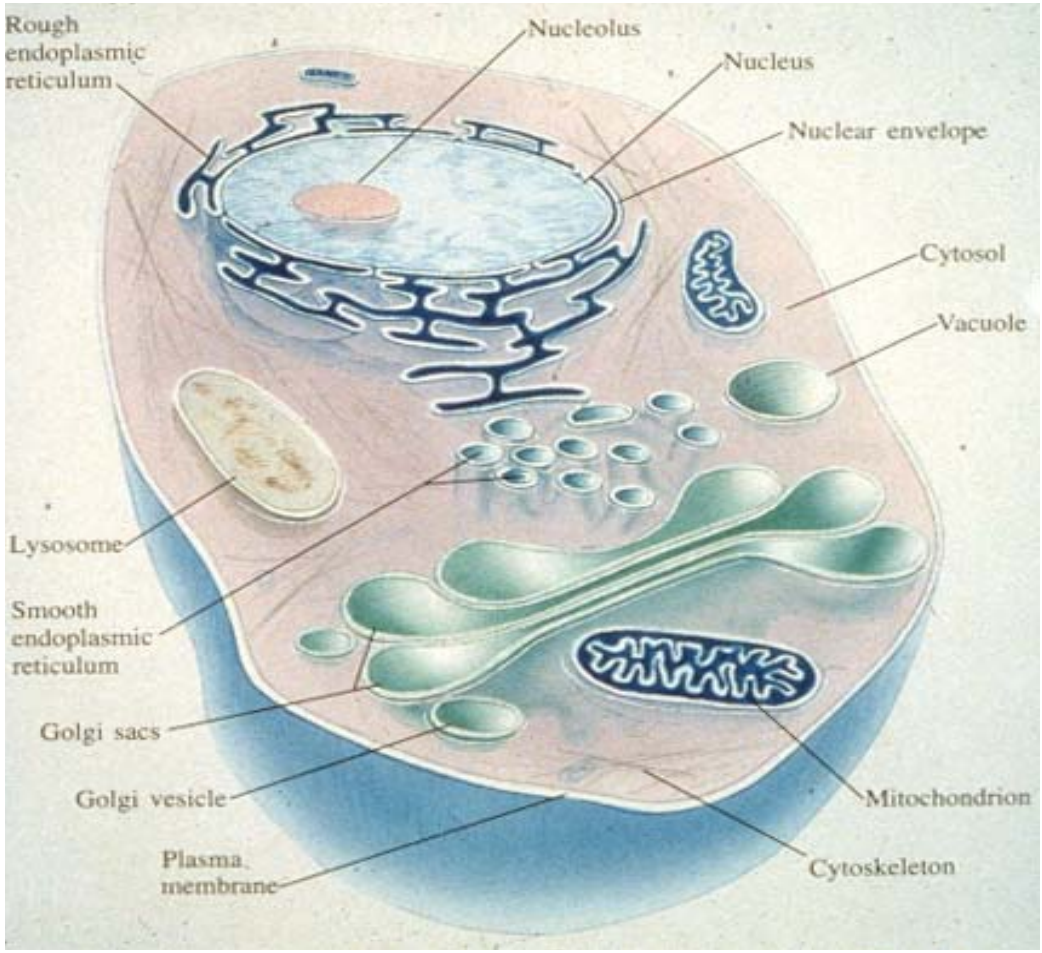
What is the real question?

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.

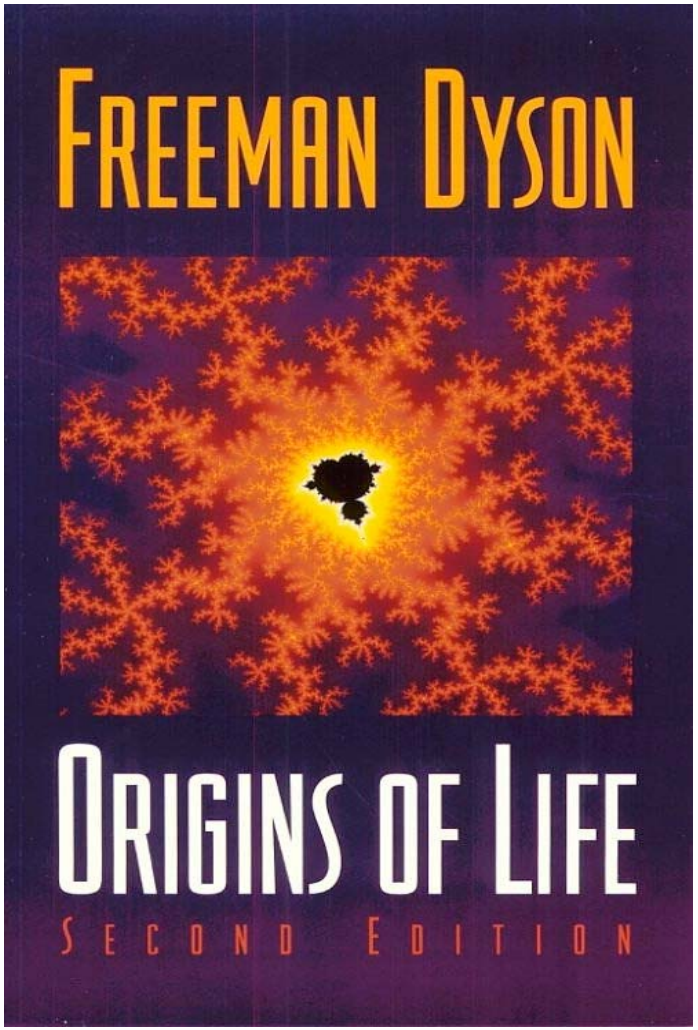
Prebiotic entities may have acquired Systems properties very early on

Freeman Dyson

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.

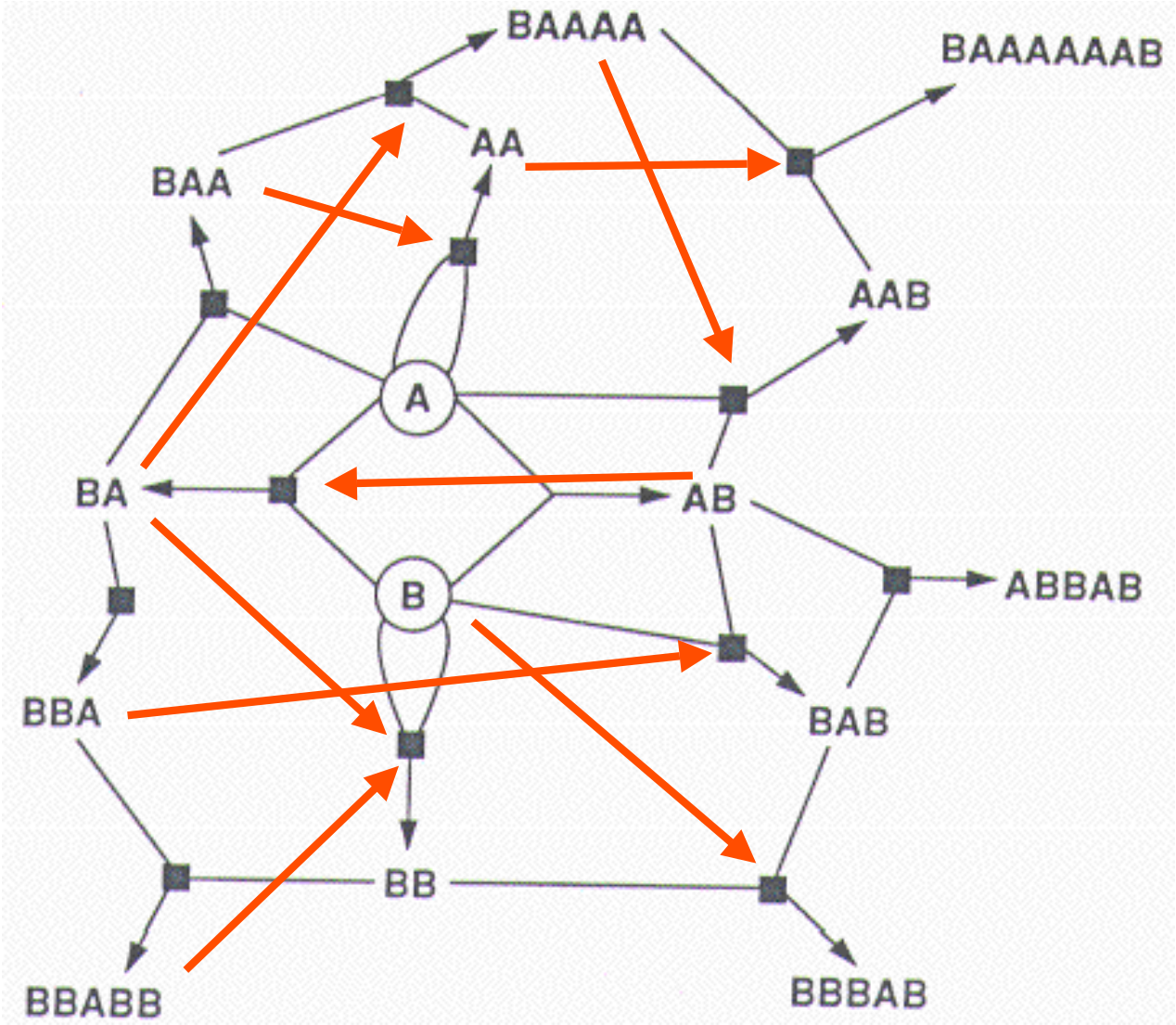
Darwinian selection then drives towards greater complexity.

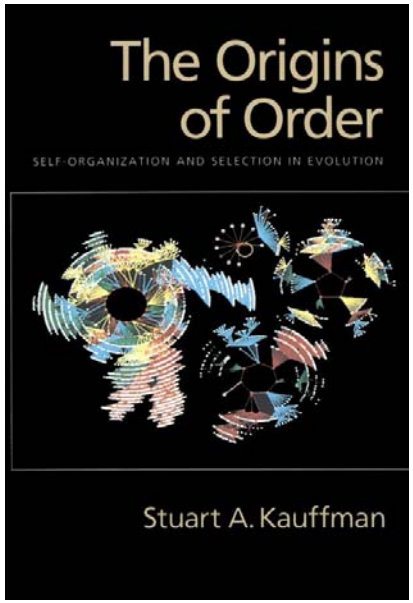


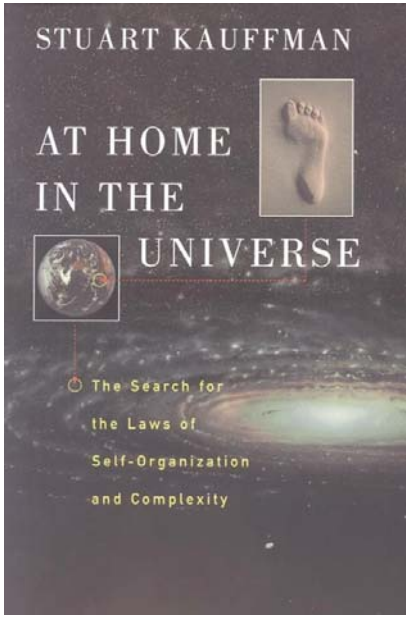
Cambridge University Press 1999

A network of mutually catalytic events

Stuart Kauffman: Mutually catalytic networks

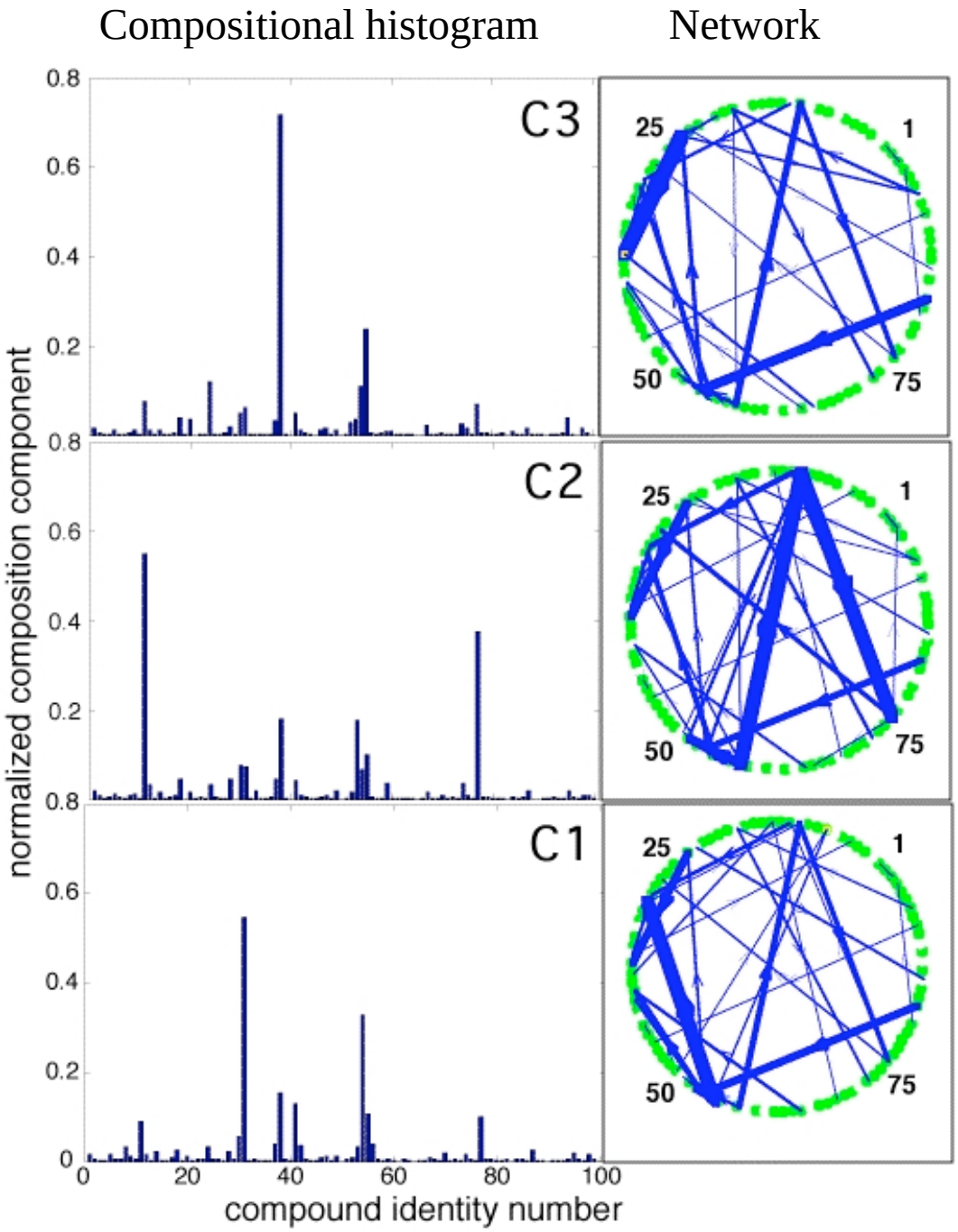




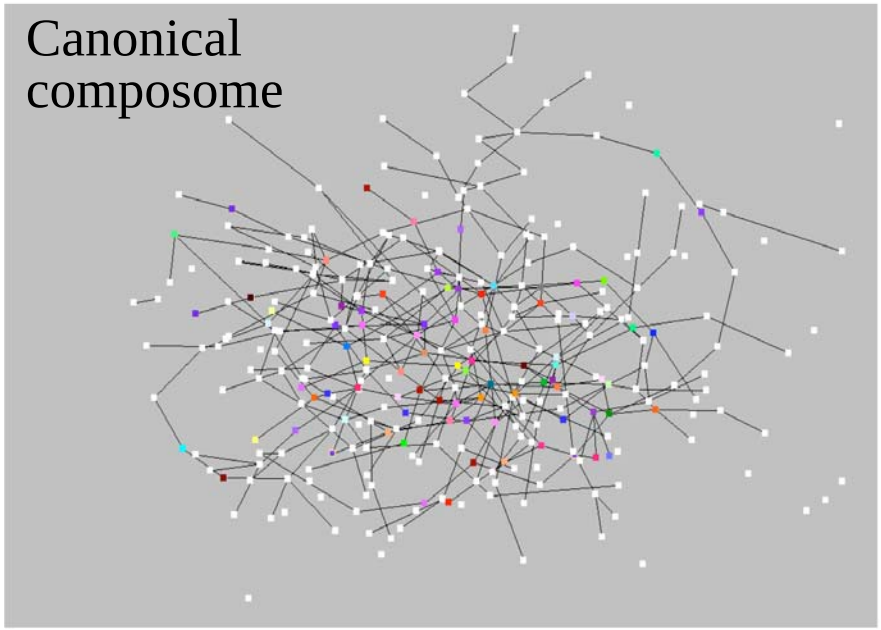


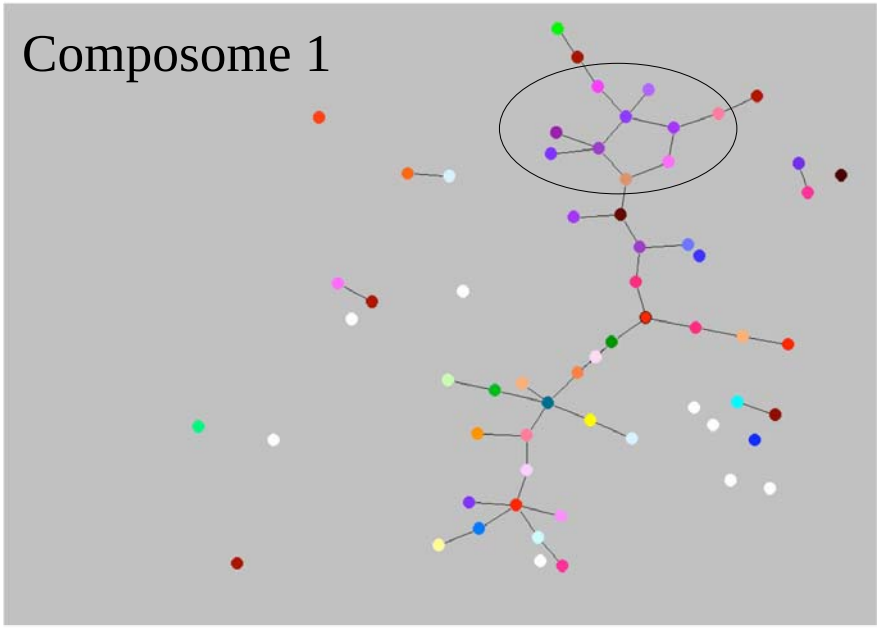
A, B are “foodstuff” (monomers)

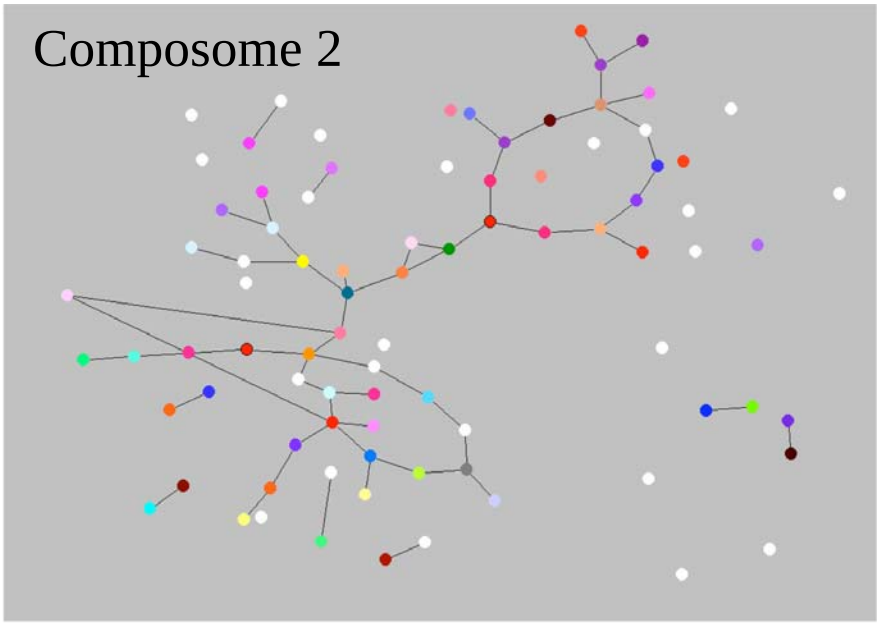
Each GARD composome is a network with a different molecular repertoire

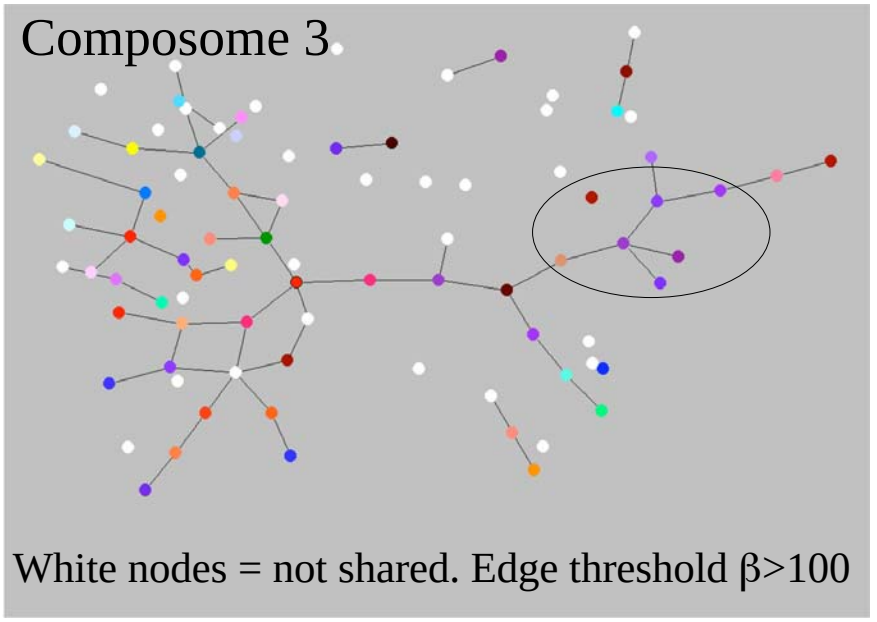


Diffderent composomes have different network motifs



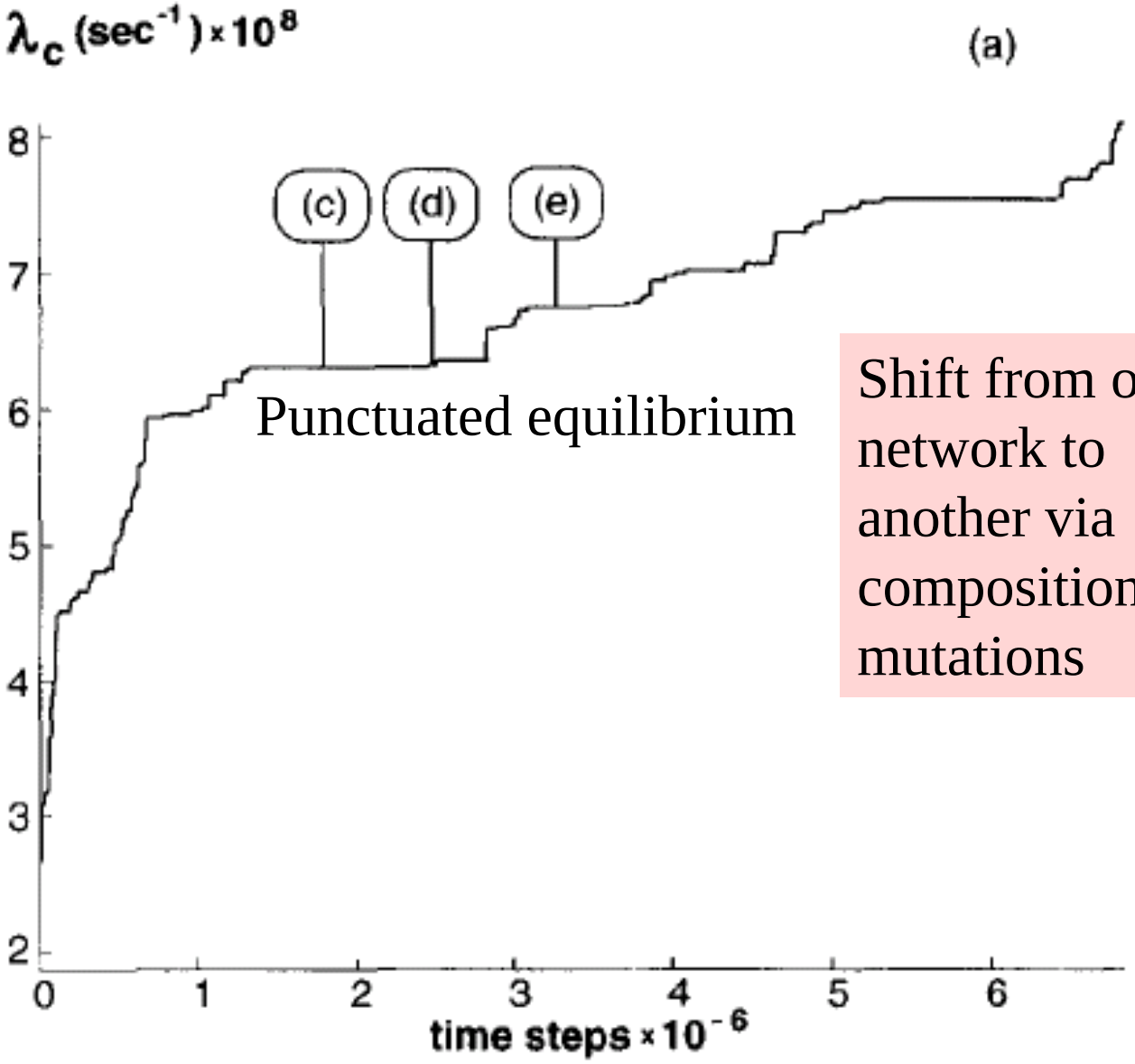






GARD evolutions

D. Seyré et al. / Physica A 249 (1998) 558–564



Shift from one network to another via compositional mutations

(c)

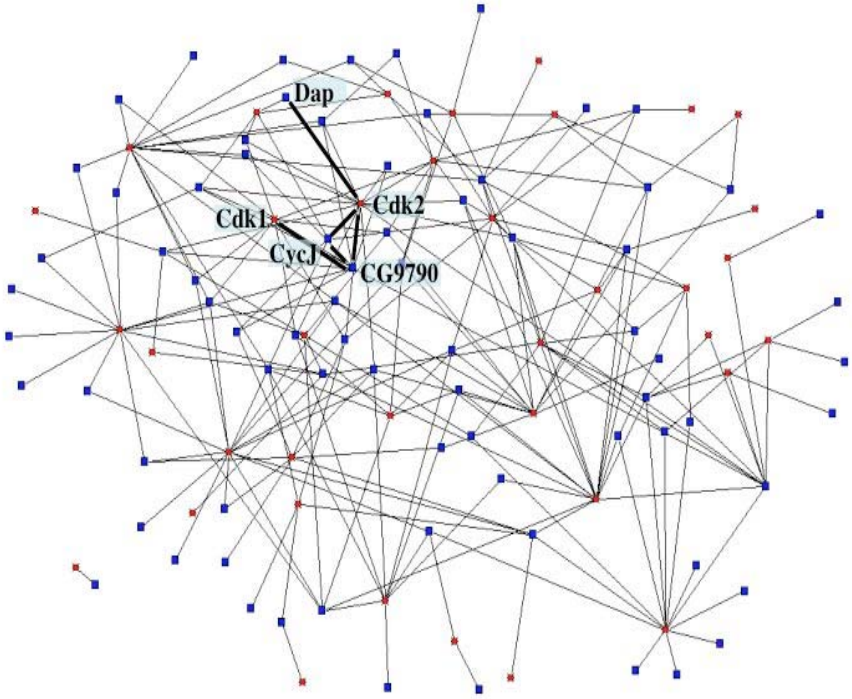
(d)

(e)

GARD network

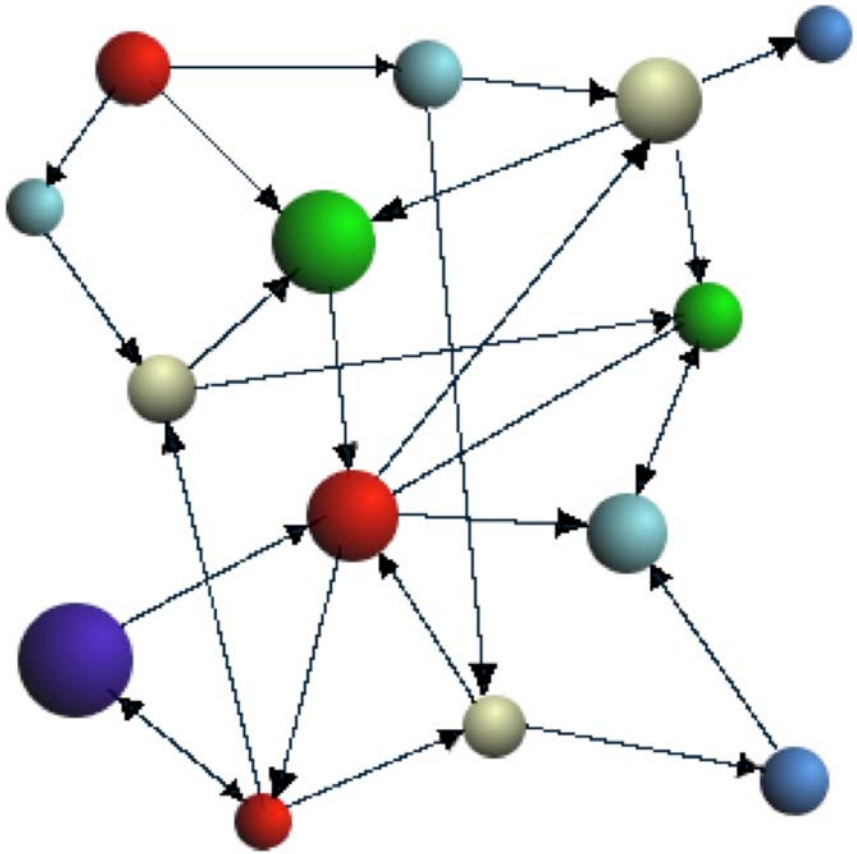
Canonic composome (eigenvector).
N_G=1000
N = 800
Catalytic potency (β) cutoff = 100

Protein interaction network

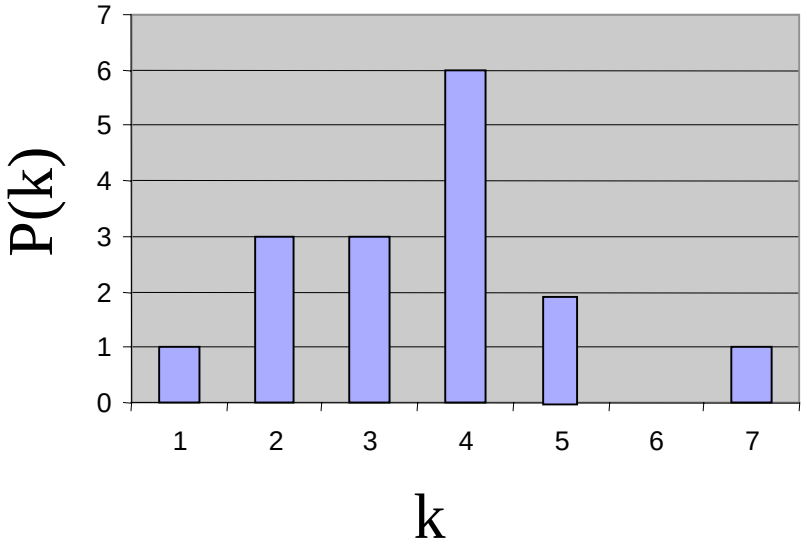


Protein interaction map (PIM) generated by using ~100 known or suspected cell cycle regulators in *Drosophila melanogaster*

Network degree (k) distribution analysis



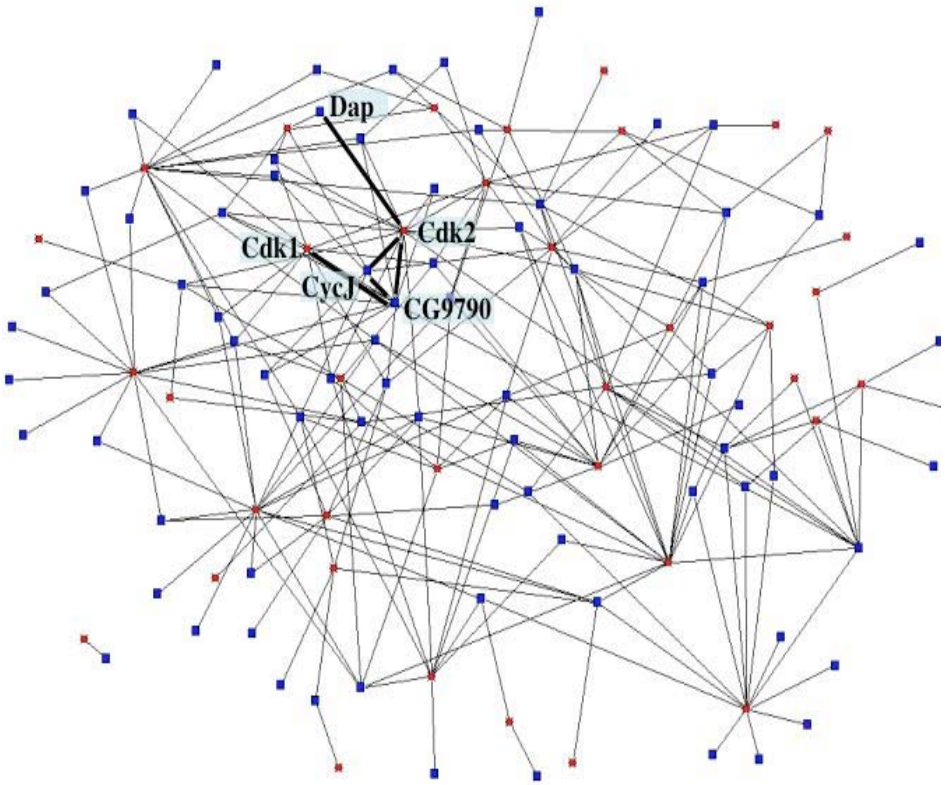
k = 4



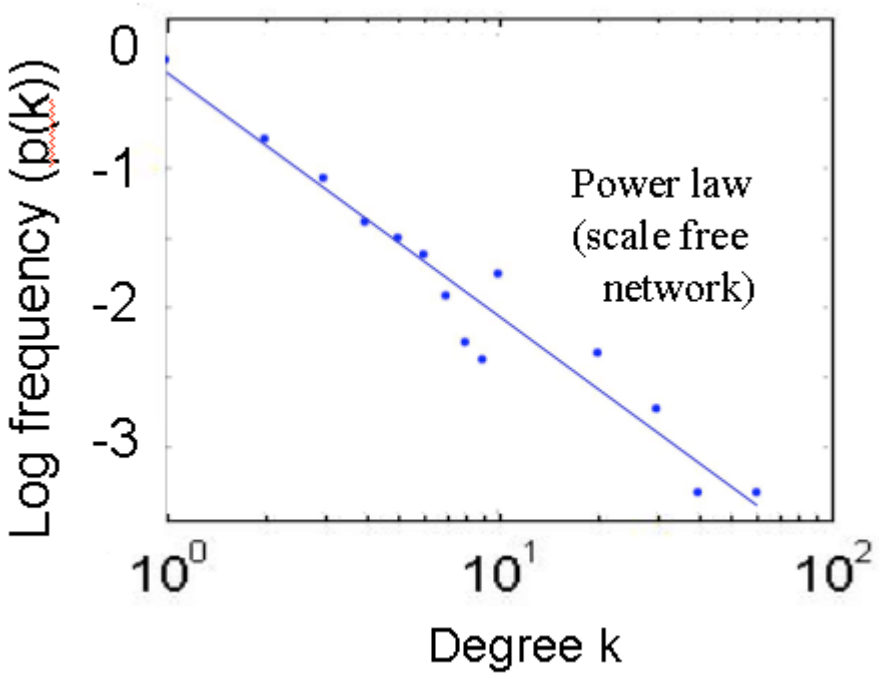
Random network

k = 2

Protein interaction network

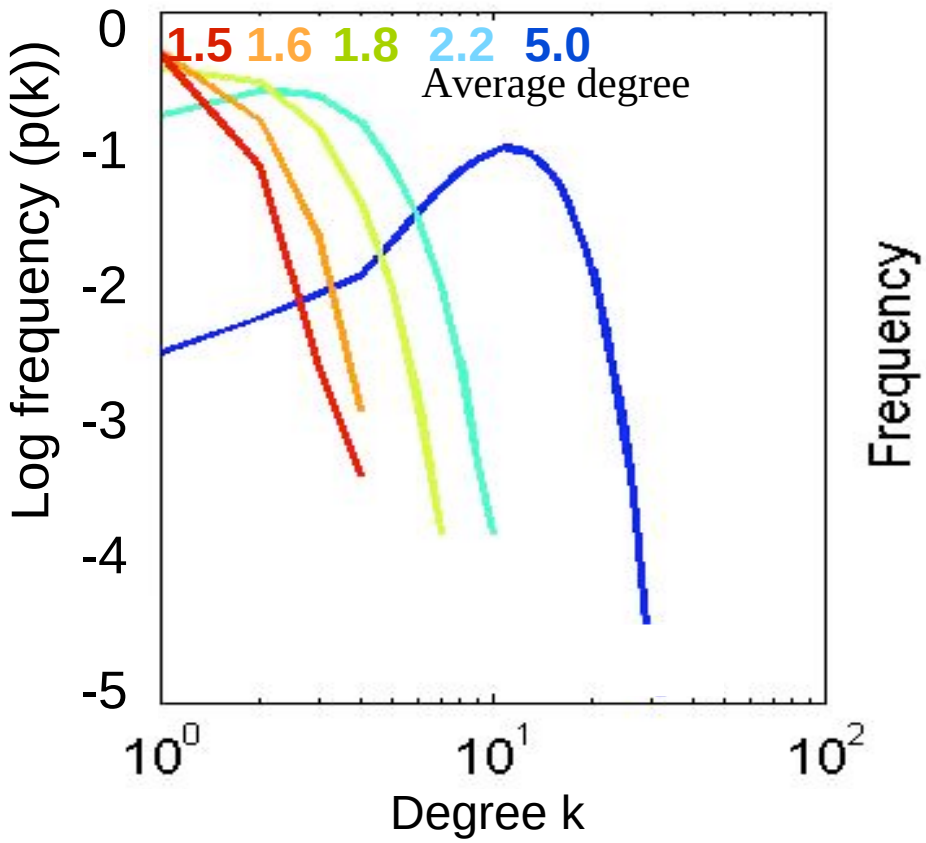


Power law!



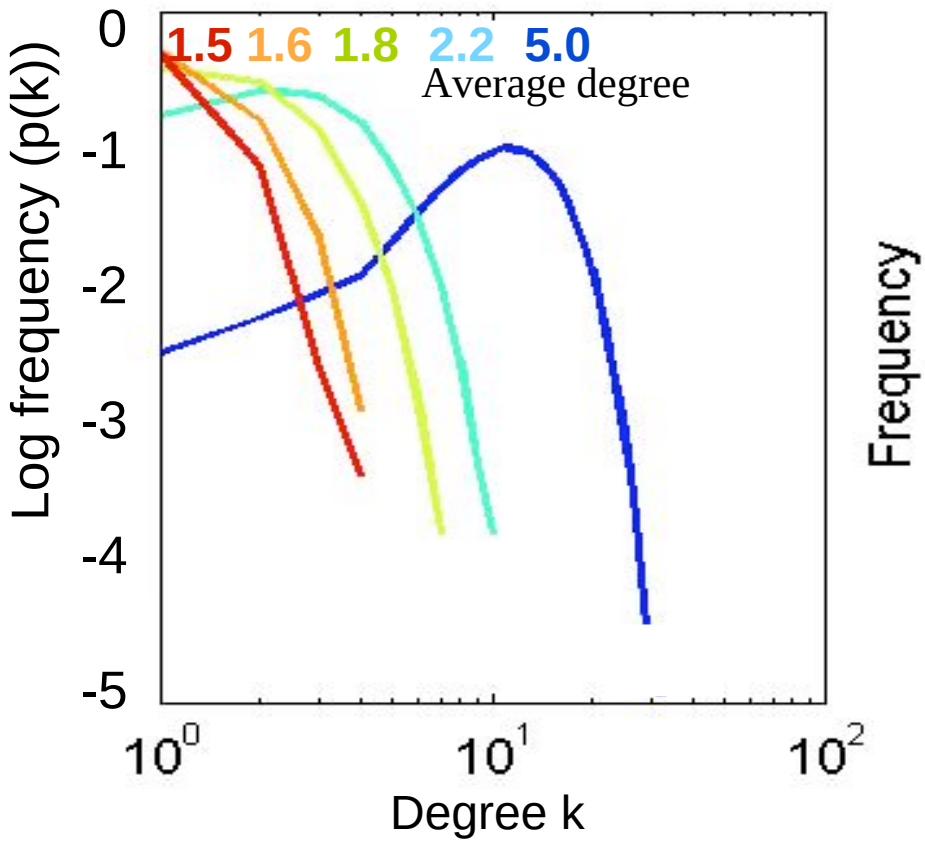
Degree distribution analysis

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



1000 canonical composomes with N_G = 1000
Different colors for different β cutoffs

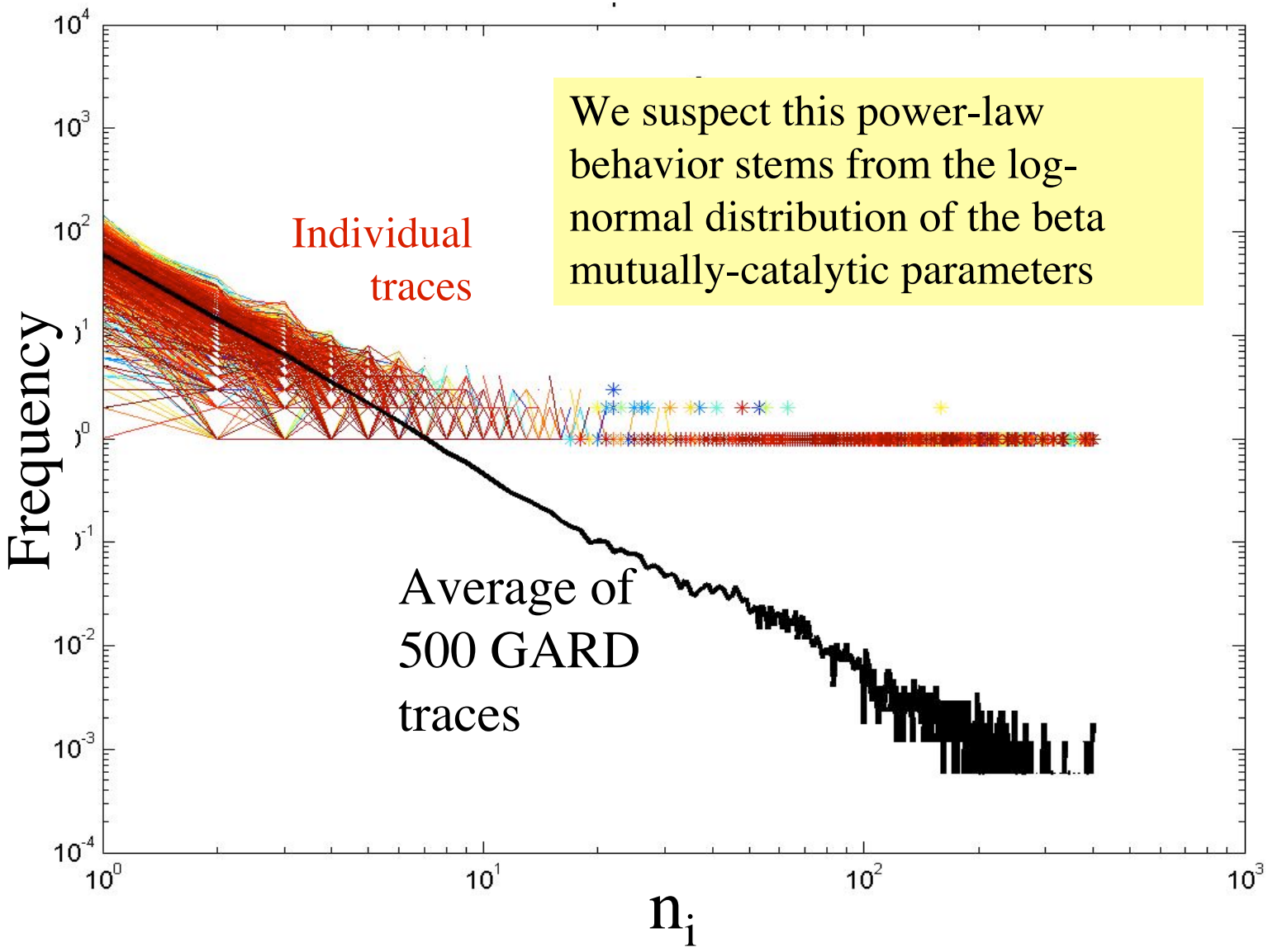
In the simplest embodiment, GARD network do *not* show a power law in their degree distribution

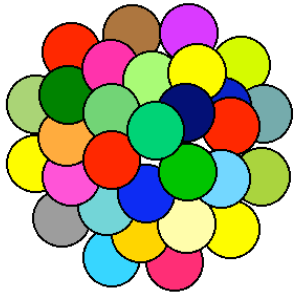


Open question:
What might render GARD networks more similar to present life's networks in showing a power-law?

1000 canonical composomes with N_G = 1000
Different colors for different β cutoffs

But there is a power law distribution for n_i values

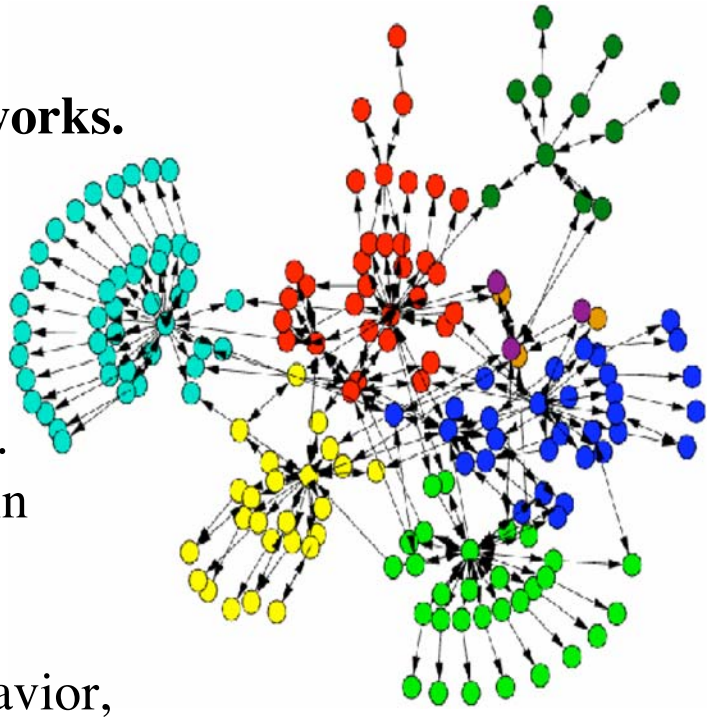




n_i is the count Of the n-th Type of molecule

Scale free networks arise due to a "rich gets richer" principle

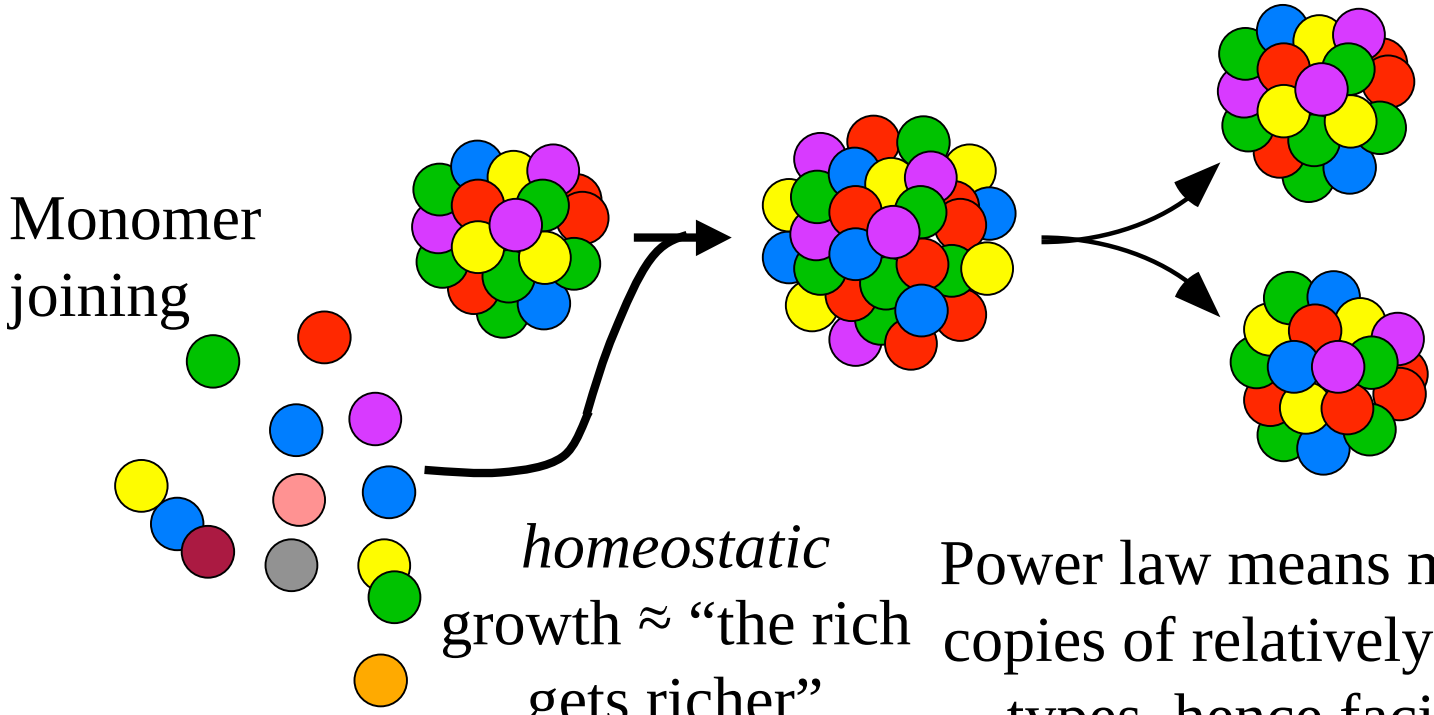
Bose-Einstein condensation in complex networks.
Bianconi G, Barabasi AL.
Phys Rev Lett. 2001 Jun 11;86(24):5632-5



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

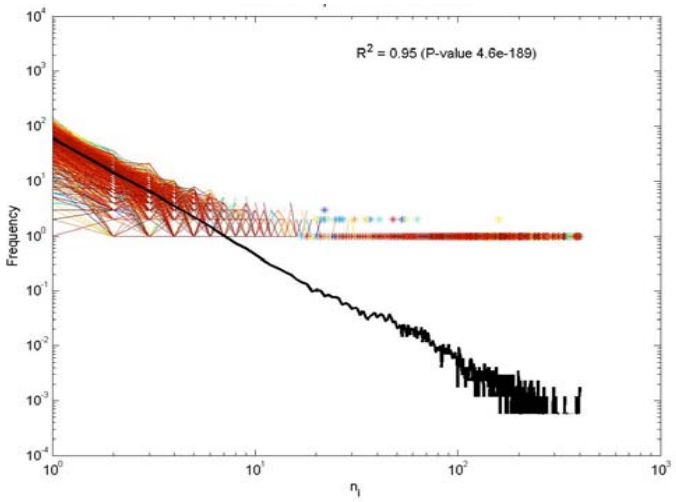
The model describes a "first-mover-wins" behavior, in which the oldest nodes acquire most links.

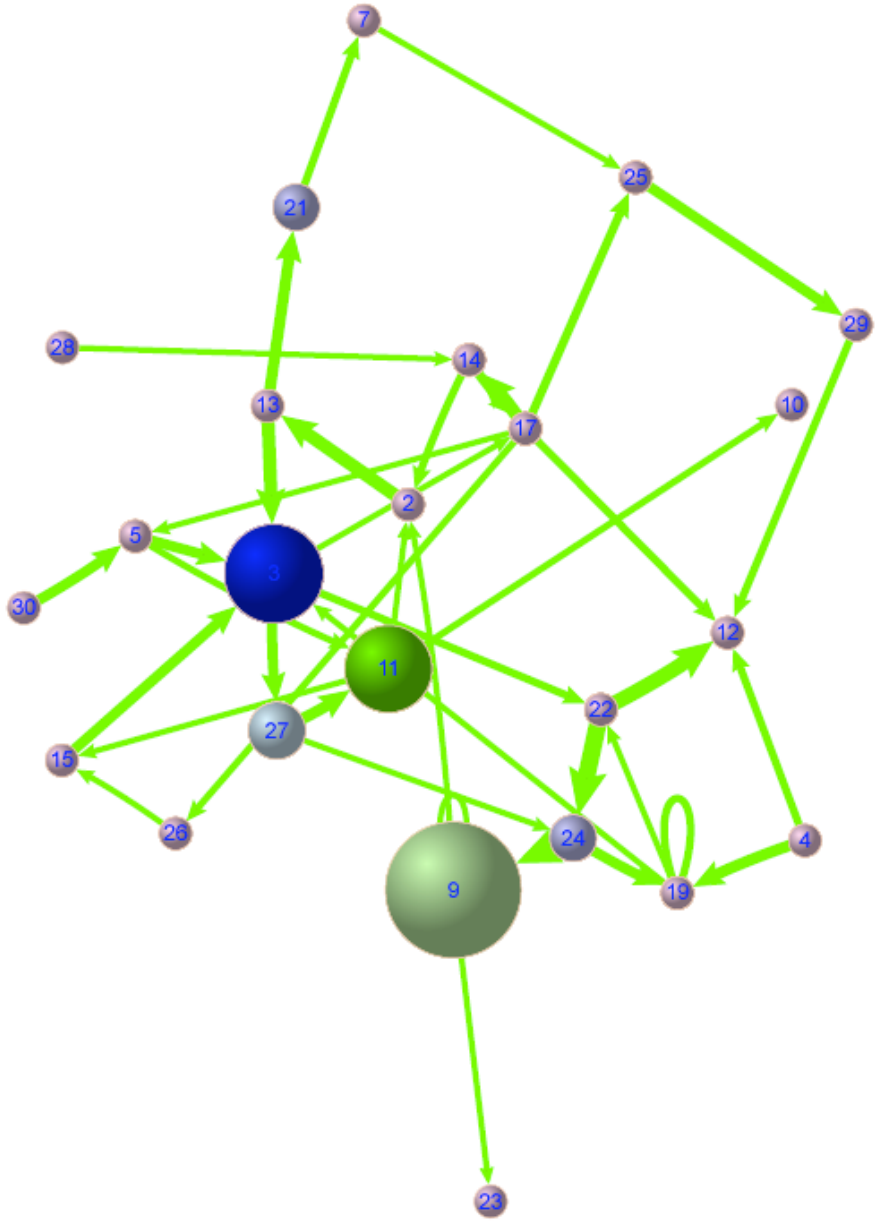
GARD dynamics and n_i power law



homeostatic growth ≈ "the rich gets richer"

Power law means many copies of relatively few types, hence facile "splittability"





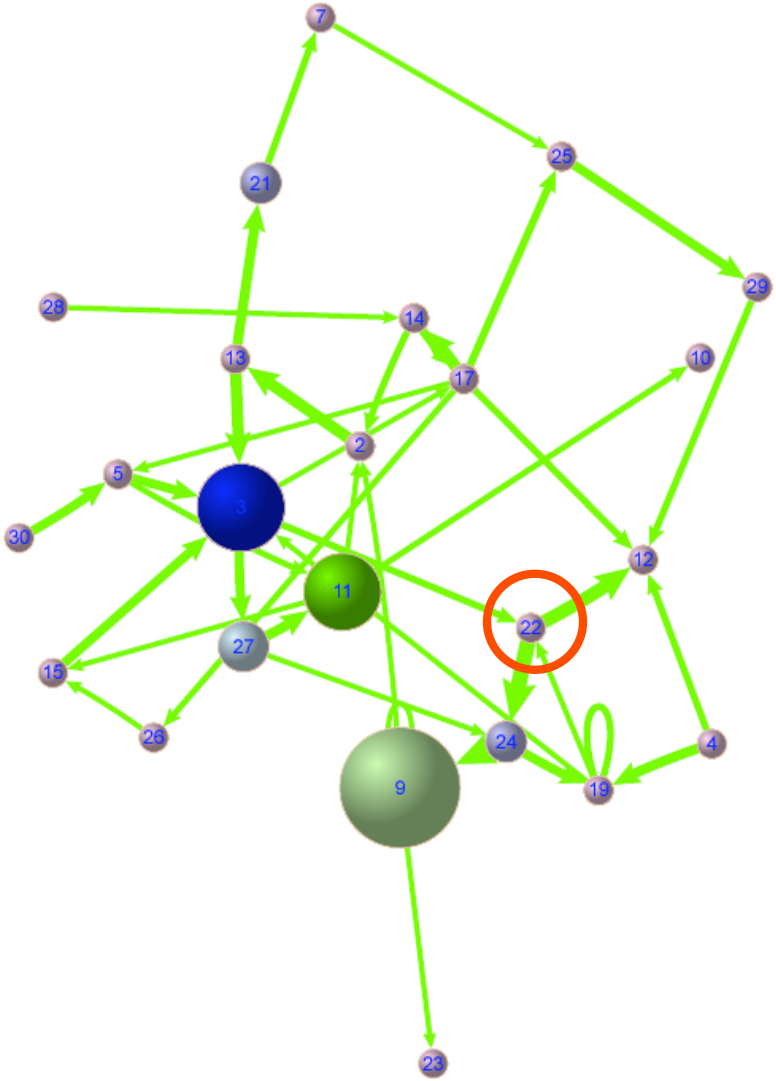
GARD networks:

* Weighted

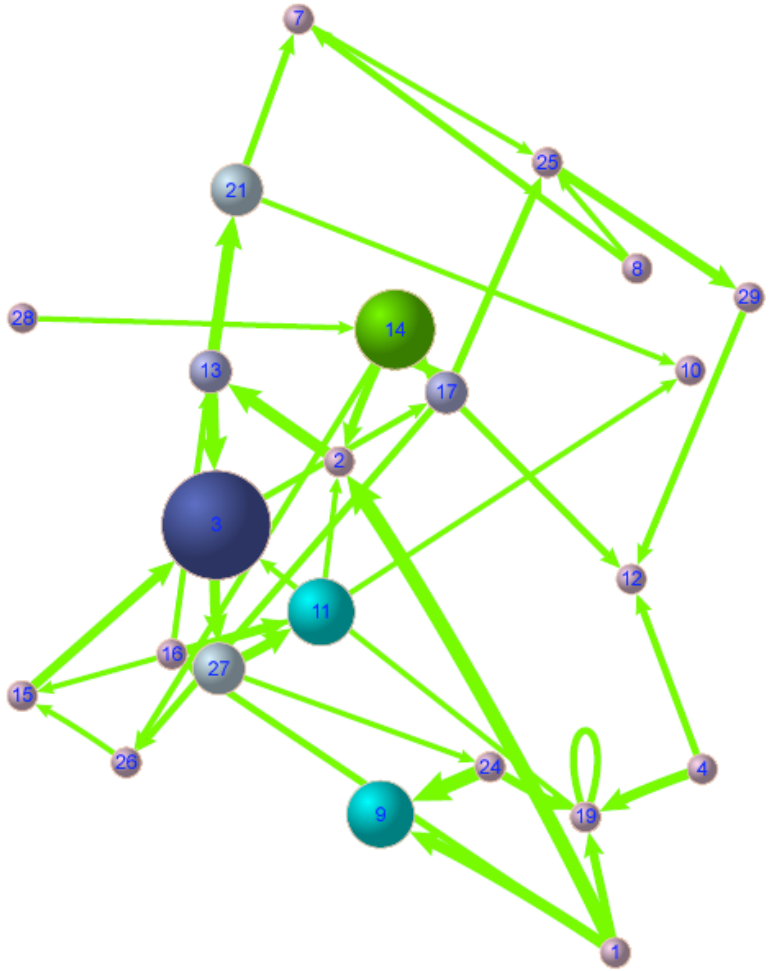
* Directed

•All nodes equivalent

GARD mutation analysis – sequentially delete every node



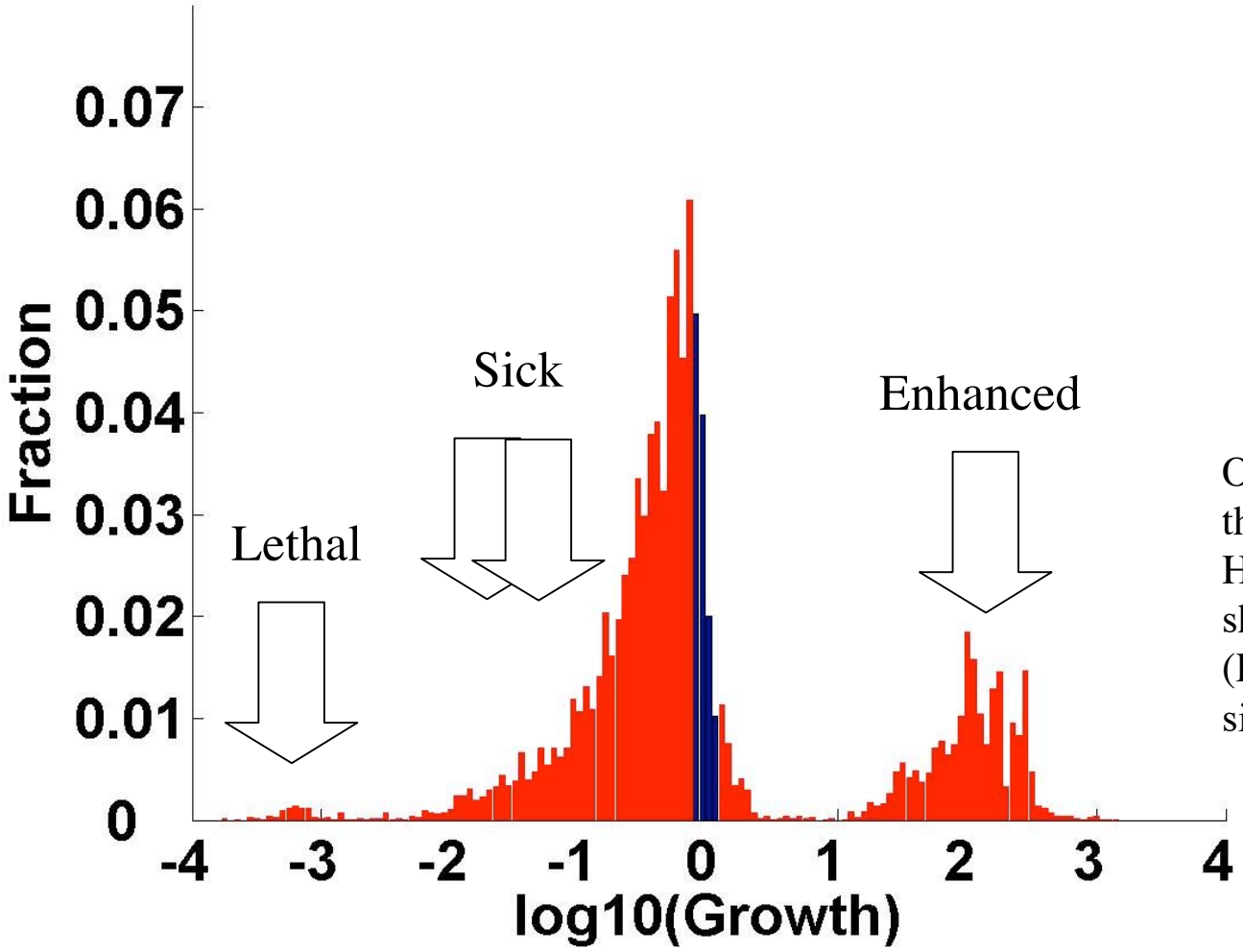
Wild type



Mutation in A22

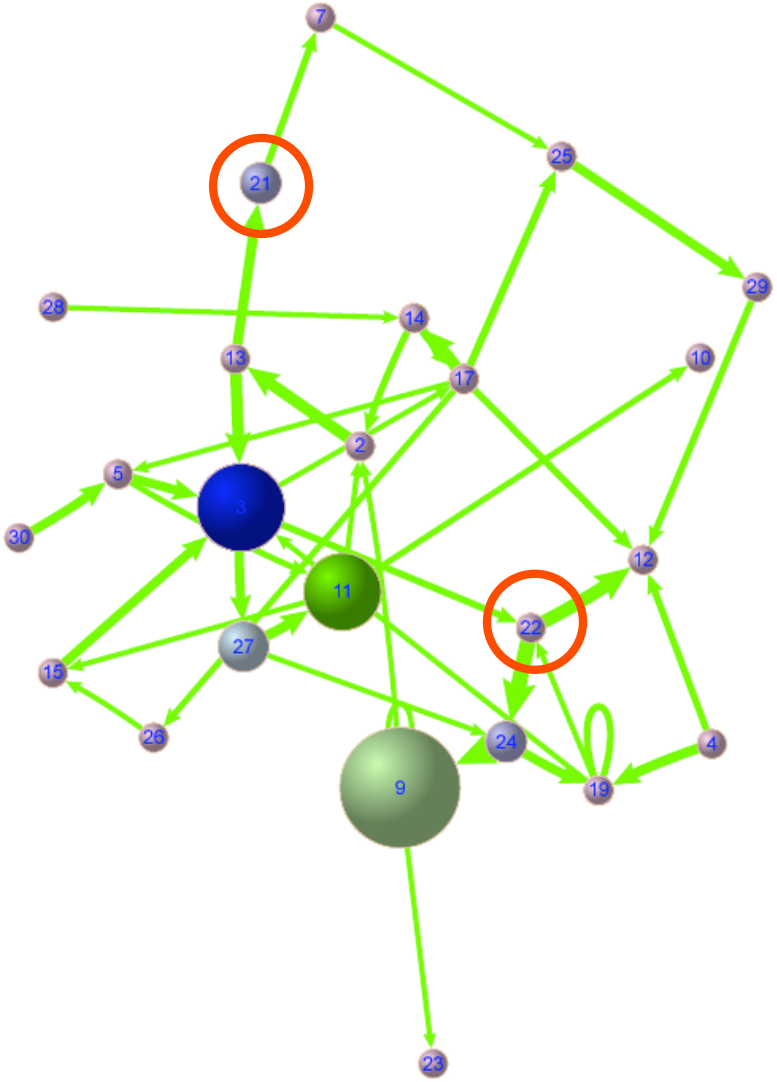
A statistical approach to GARD network mutations

1000 networks, each with different β matrix



Only the 3% of the networks with $H_{wt\text{-}Mut}$<0.9 are shown (H=compositional similarity)

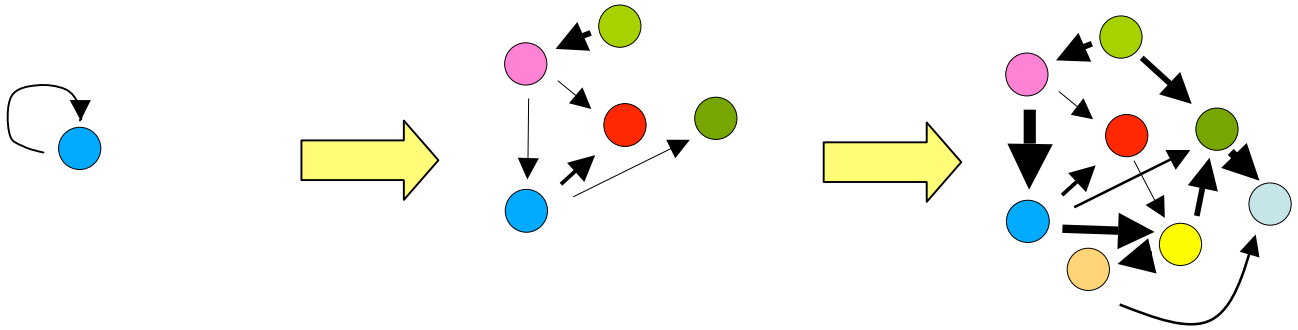
Future: GARD synthetic lethality – mutate two nodes at a time



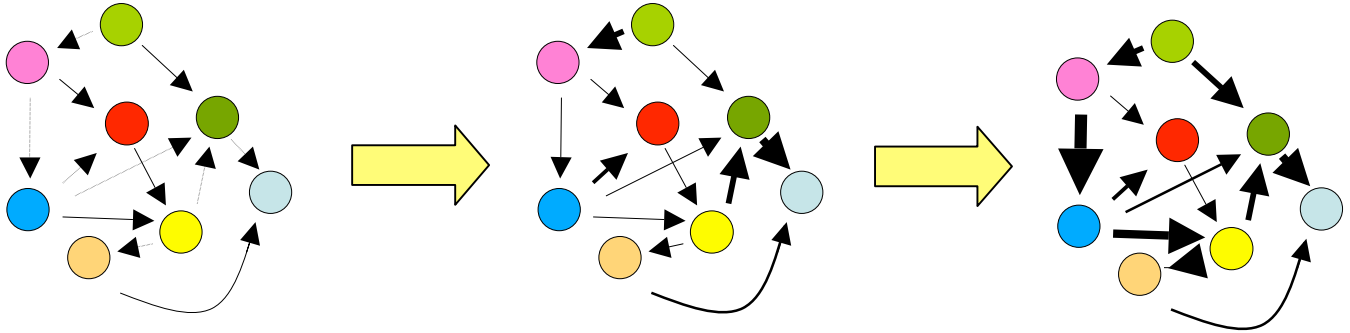
Plan to analyze:

- Sythetic lethality
- Synthetic sickness
- Extragnic suppression
- Robustness vs fragility
- Node addiction

Two scenarios for increasing network complexity



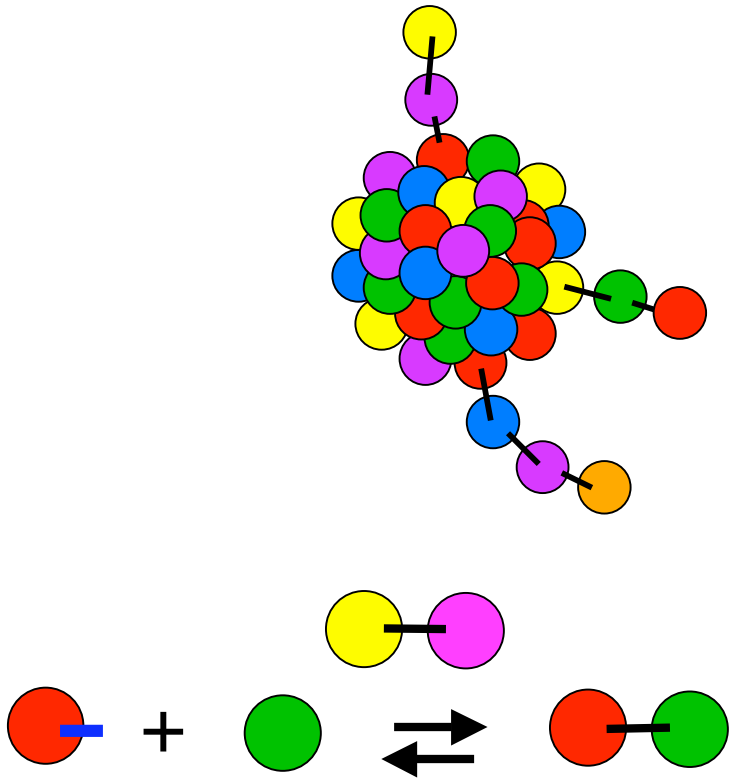
A: Increasing node count



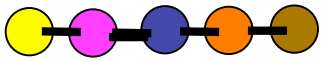
B: Increasing node fidelity

Current exploration - Polymer GARD

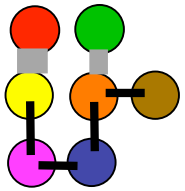
Introduce covalent oligomerization (endogenous synthesis) to GARD assemblies
(Shenhav et al, OLEB 2005)



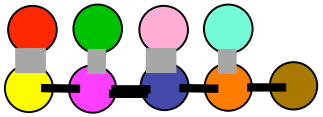
Beyond simple covalent oligomerization



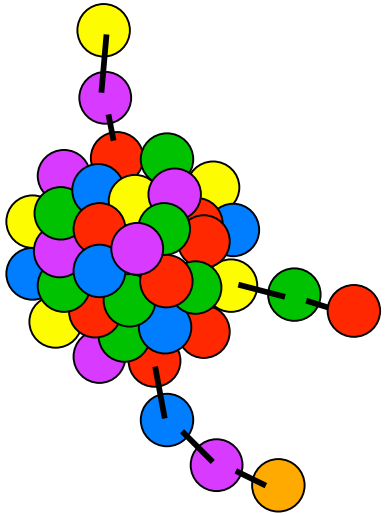
• Longer oligomers



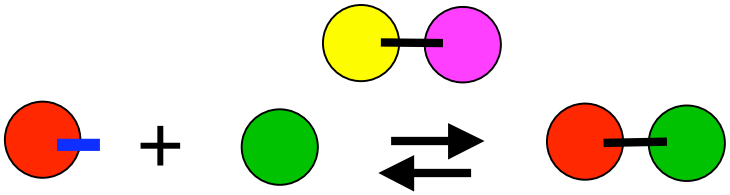
• “Folding” procedures

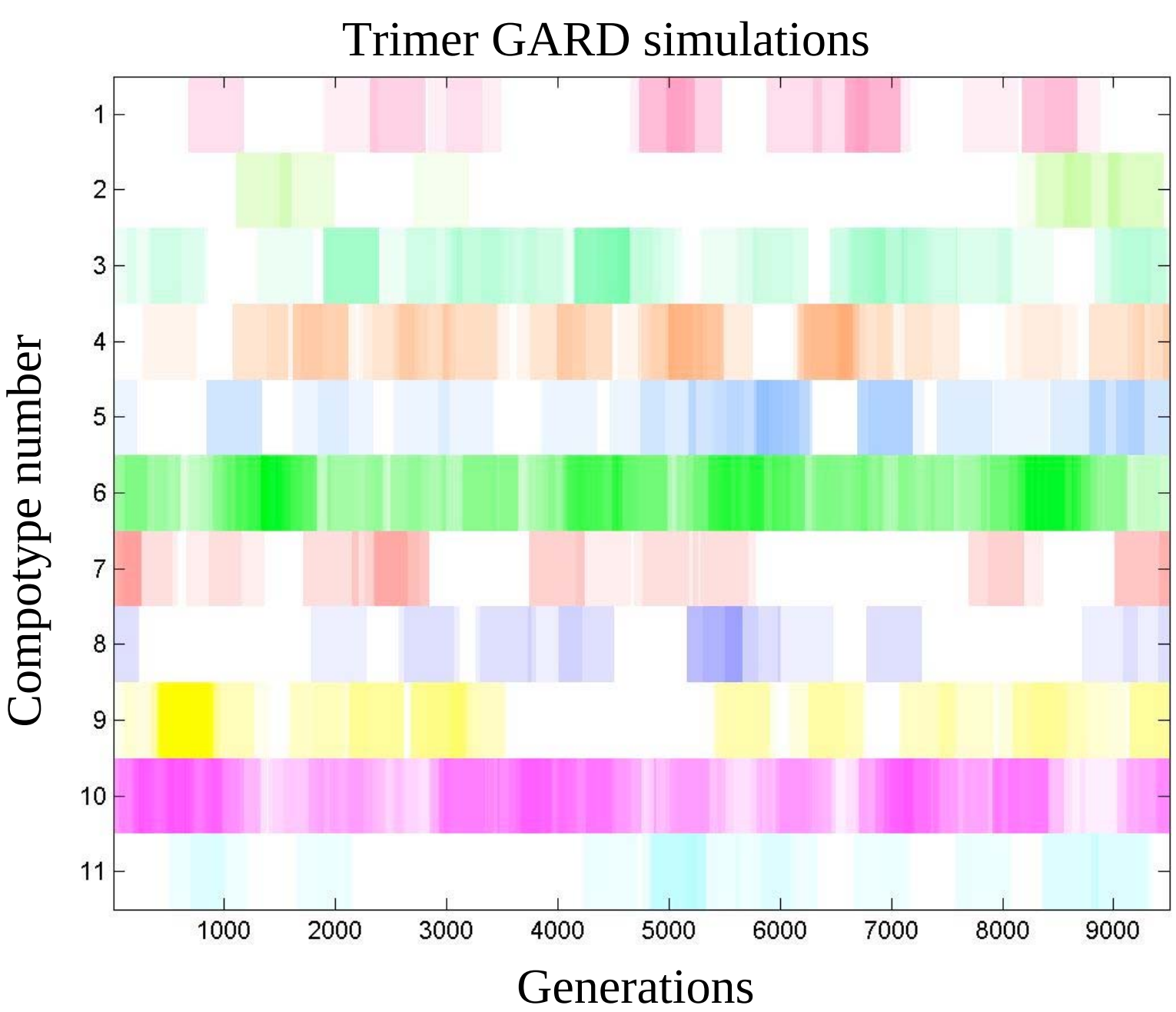


•RNA-like templating

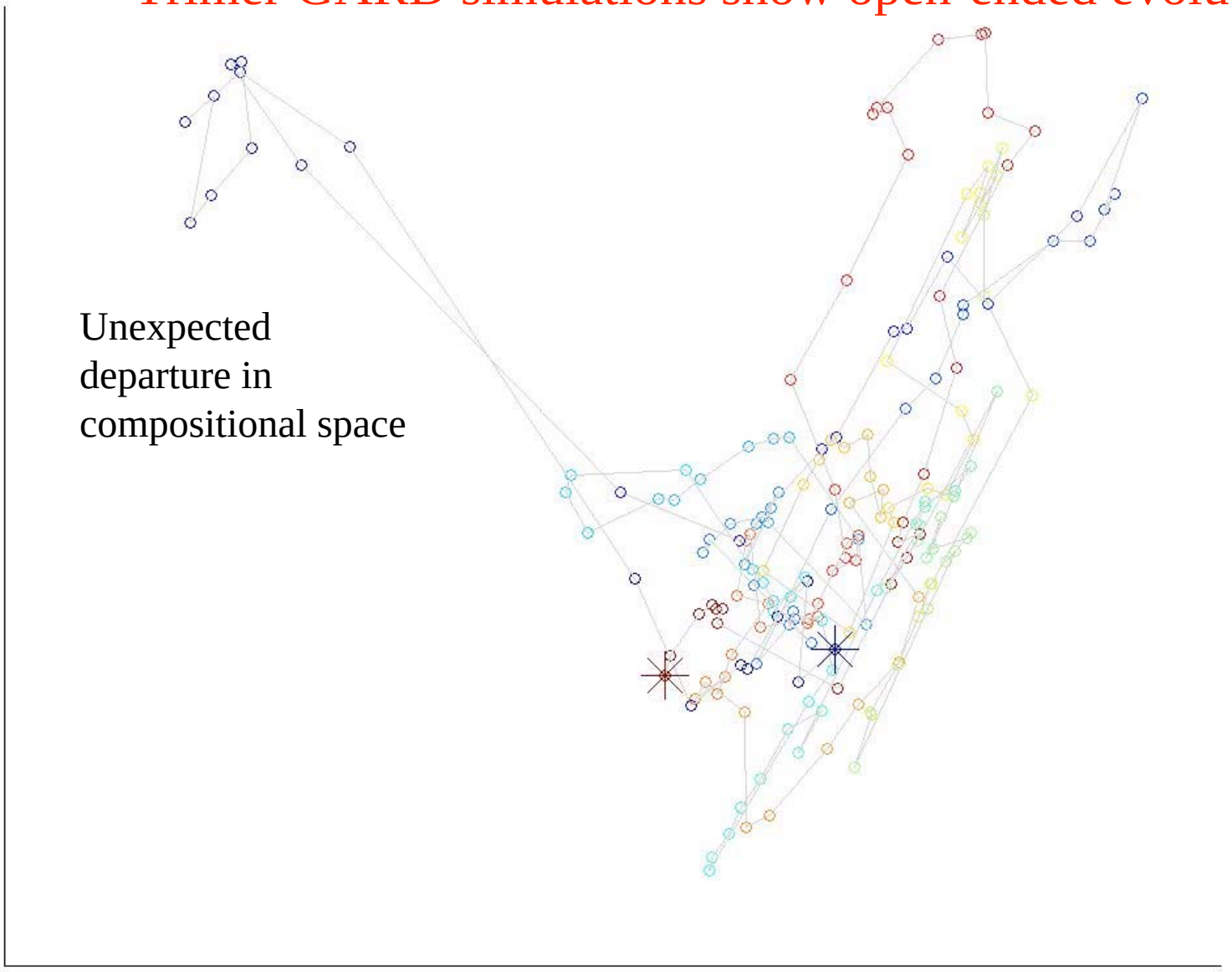


Genetic algorithms




Trimer GARD simulations
Compotype number
1
2
3
4
5
6
7
8
9
10
11
1000
2000
3000
4000
5000
6000
7000
8000
9000
Generations

Trimer GARD simulations show open-ended evolution



Principal component analysis of compotype combination

What is the real question?

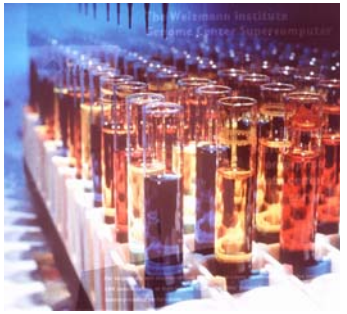
1) 3.8 or 3.5 billion years ago?
2) Here or elsewhere?
3) Probable or improbable?
4) Organic or inorganic?
5) Organics trivial or not?
6) Today's chemistry or not?
7) Catalysis by proteins only or not?
8) Large molecules or small?
9) Sequential or compositional information?
10) Single molecules or networks?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

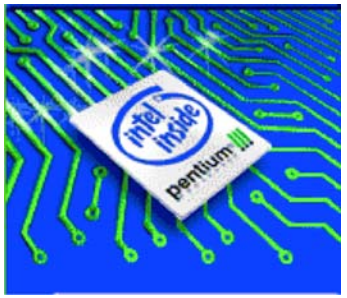
Approaches to the study of Life's origin

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

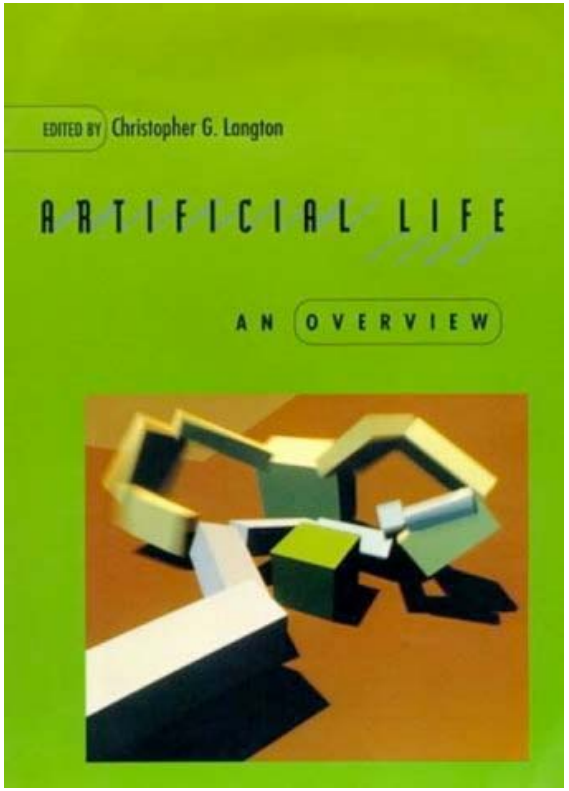
2) Galactic travel

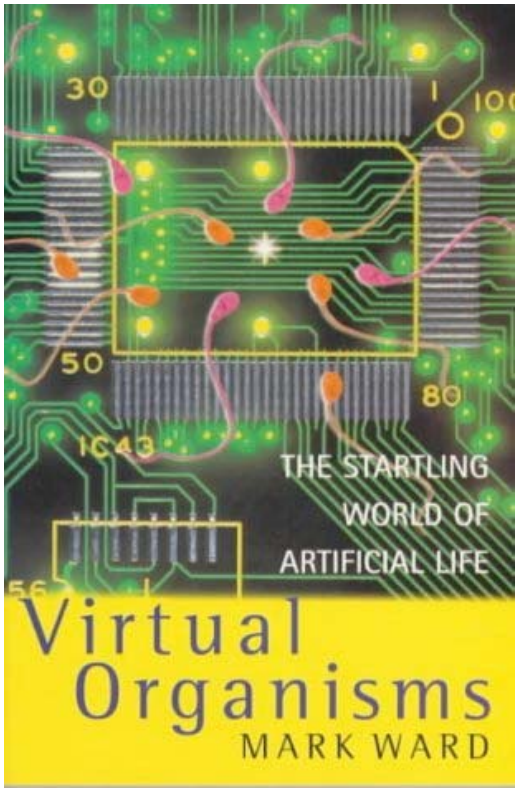
3) Large scale chemistry-realistic computer simulations.





2) Artificial Life (AL or Alife)

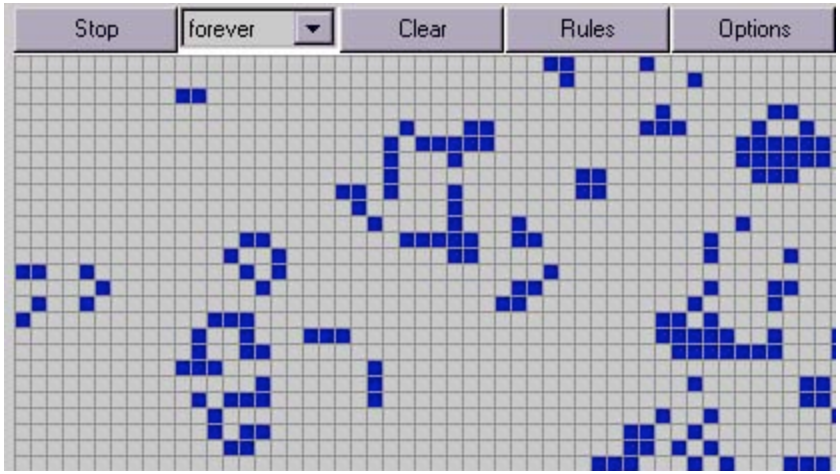




Helps understand principles but removed from chemical reality

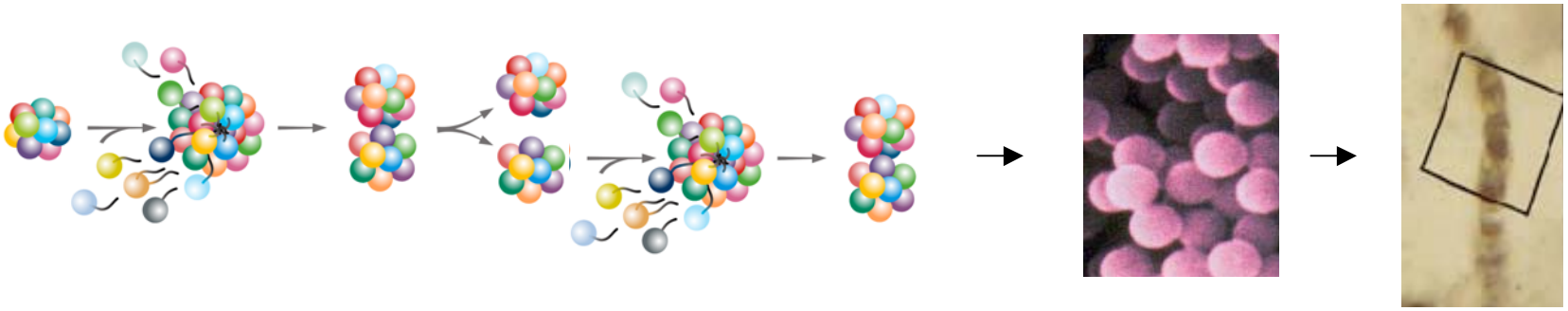
http://www.webslave.dircon.co.uk/alife/intro.html

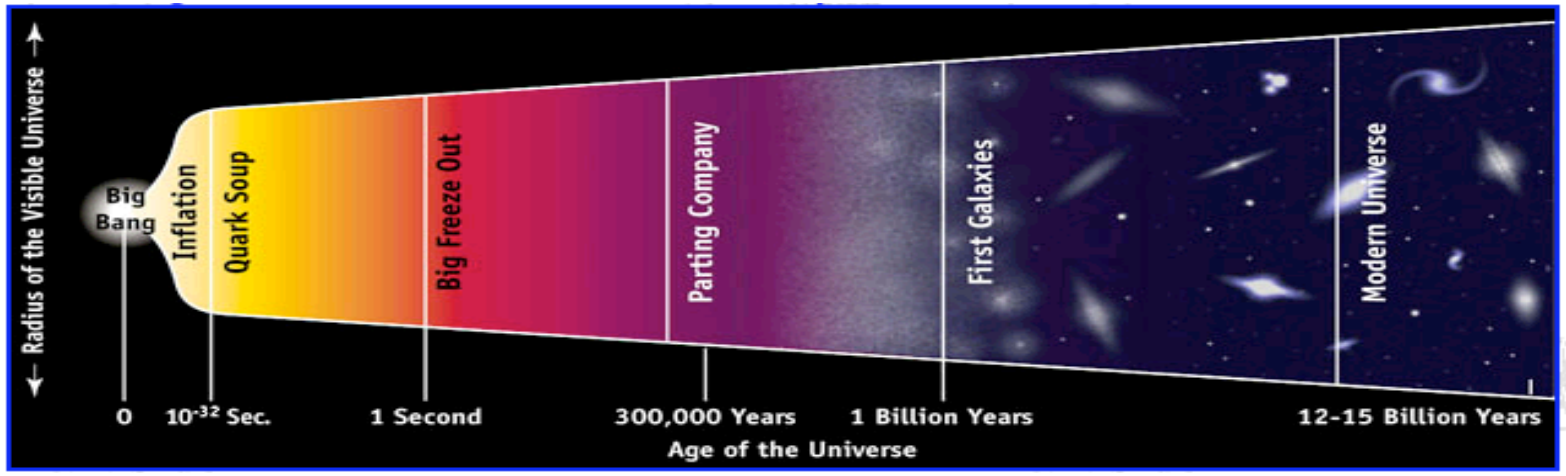
John Conway's Game of Life



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

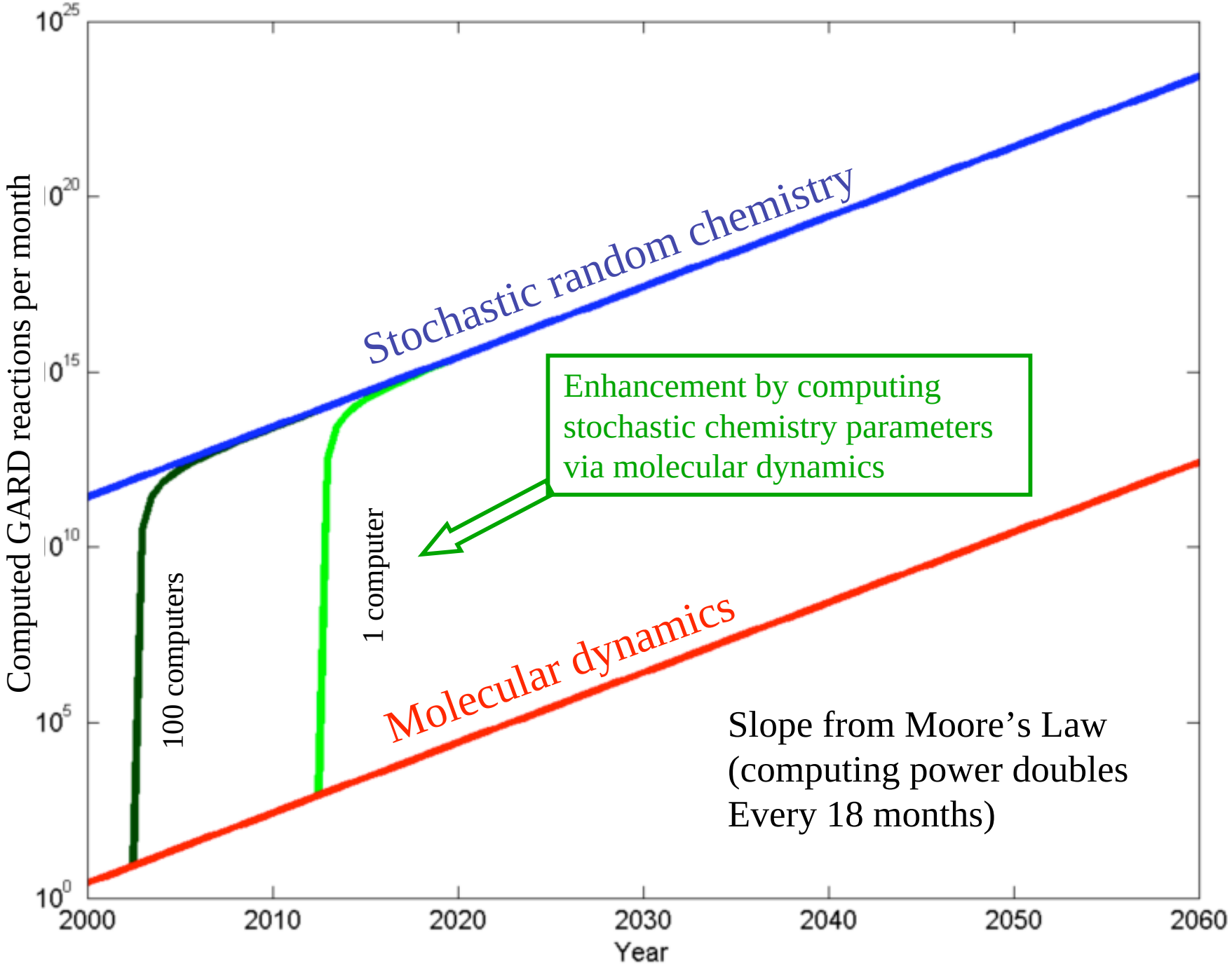




PROSPECTS OF A COMPUTATIONAL ORIGIN OF LIFE ENDEAVOR

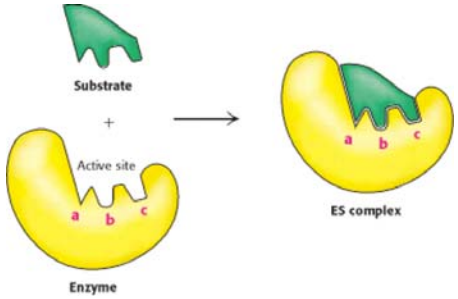
BARAK SHENHAV and DORON LANCET*
Department of Molecular Genetics and the Crown Human Genome Center, the Weizmann Institute of Science, Rehovot 76100, Israel

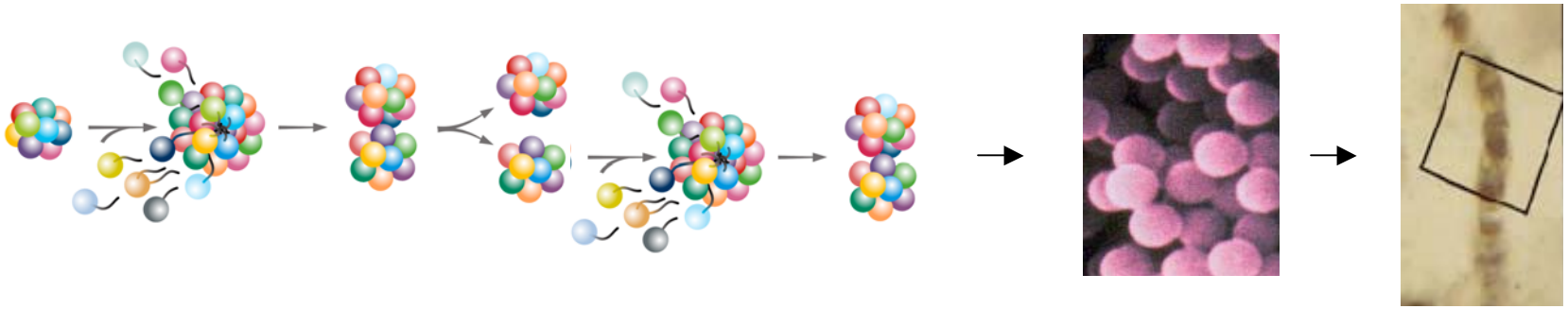
Origins of Life and Evolution of the Biosphere **34:** 181–194, 2004.

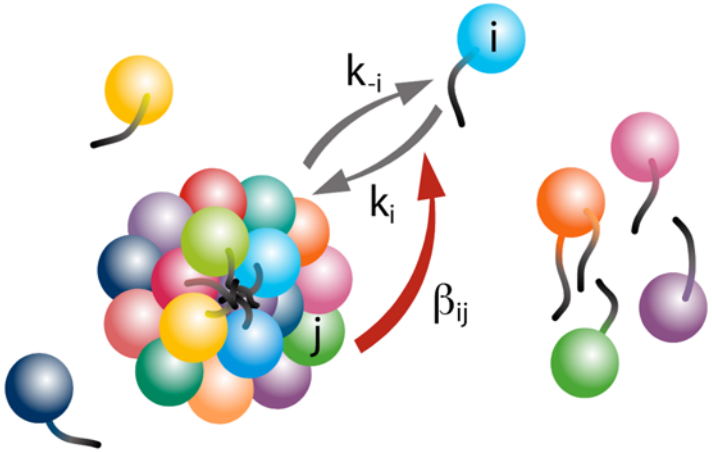

Computed GARD reactions per month
10^{25}
10^{20}
10^{15}
10^{10}
10^{5}
10^{0}
2000
2010
2020
2030
2040
2050
2060
Year
Stochastic random chemistry
Enhancement by computing
stochastic chemistry parameters
via molecular dynamics
100 computers
1 computer
Molecular dynamics
Slope from Moore's Law
(computing power doubles
Every 18 months)

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 specificty as well as prebiotic scenarios!





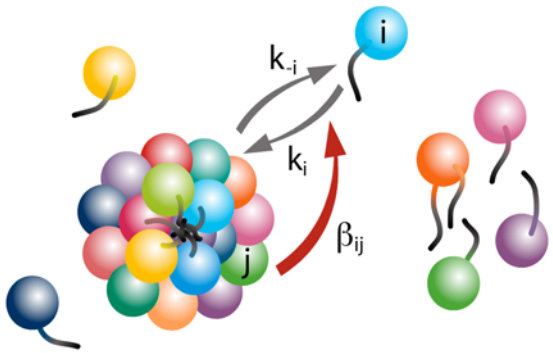


GARD – Lipid World

??

by
Karen Fincannon

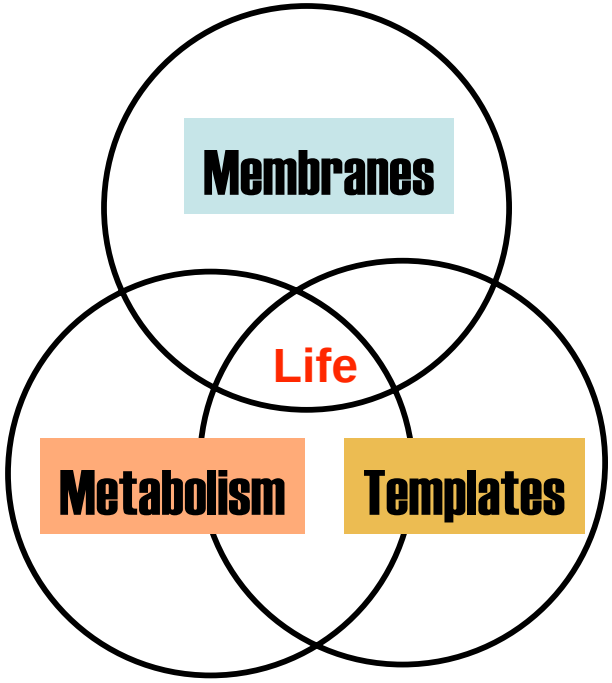
Chickegg



??

GARD – Lipid World

Chickegg



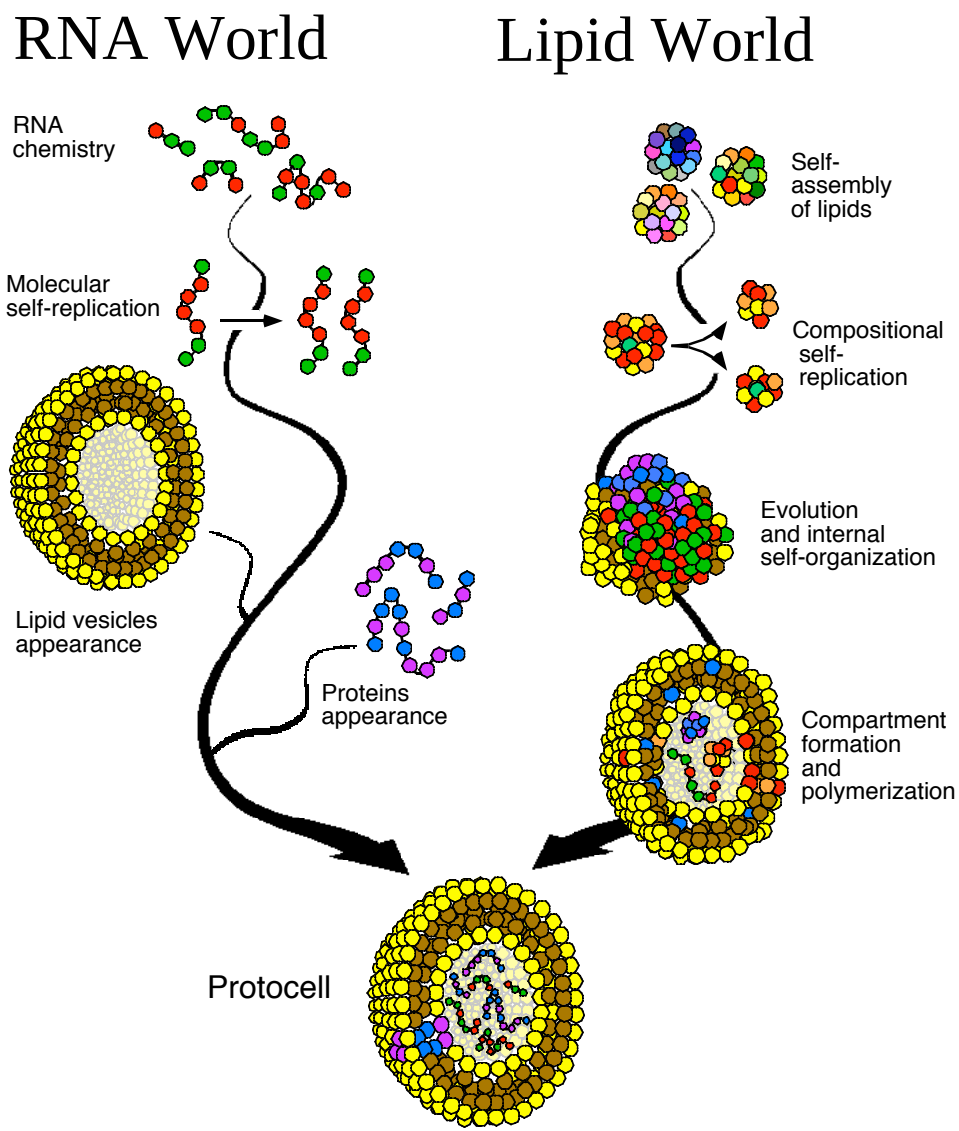
- Includes metabolism-like networks
- Contains compositional information
- Embodies an enclosed compartment
- Capable of rudimentary reproduction
- Transmits information with mutations
- Capable of primitive evolution
- Can be made gradually more elaborate and more life-like

OOL credits

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



Segrè et al. EMBO Reports 2000