Prebiotic evolution: replicator systems and information threshold

Course Computational Biology 2024; Paulien Hogeweg; Theoretical Biology and Bioinformatics Grp Utrecht University

Life as evolved (evolving) complex multilevel information processing system

how to model it?

(1) how did it get started?/ bootstrap itself

Origin of life

studied by

- phylogenetic reconstruction (LUCA)
  quite complex - what before that?....
- Sufficient chemistry / environment
- experimental studies of minimal 'living' systems
  (re)constructing/engineering/evolving) such systems
- modeling studies of minimal 'living' systems

different approaches (focus) dependent on:

what is life?

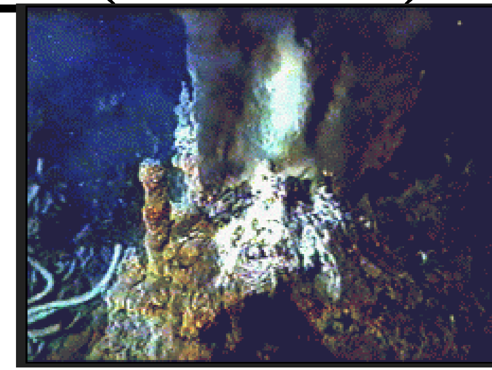
including alternative forms (e.g. extra terrestrial / lab.)

Life is ....
Unique properties of life not shared by technological systems

’In stark contrast with current computer technology, biological cells compute in construction using molecular and spatial information, in order to delimit, organize, power, sustain, repair, move, communicate, reproduce, protect and evolve themselves robustly from simple and scarce material and energy resources in their complex environments’
J. McCaskill and S. Rasmussen EU report (2012)

not good starting point...

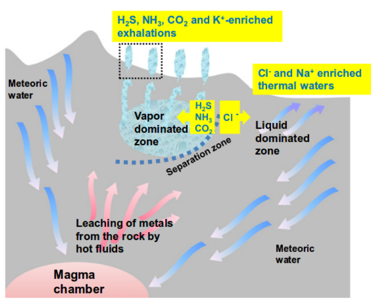
Hypothesized Environments of Prebiotic 'life' (metabolism)

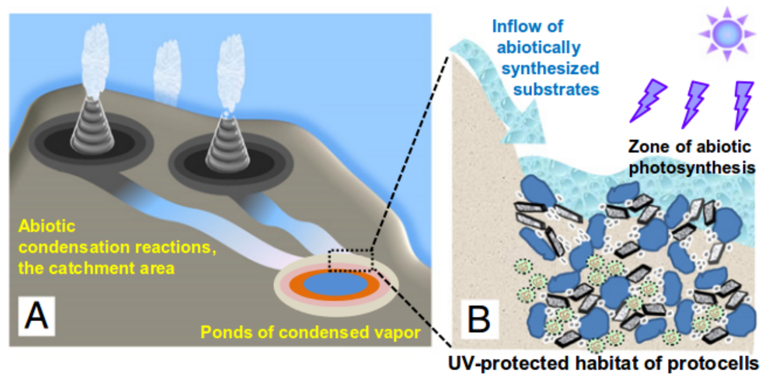


Hydrothermal vents: (black smokers)
energy/energy gradients for free
compartments (concentration of ingredients)
catalysis by metal sulphides; acetyl-coA pathway
abiotic aminoacid synthesis
Menez et al PNAS 2018

OR

Origin of first cells at terrestrial, anoxic geothermal fields Because of 'open' cell environment should match internal cell composition

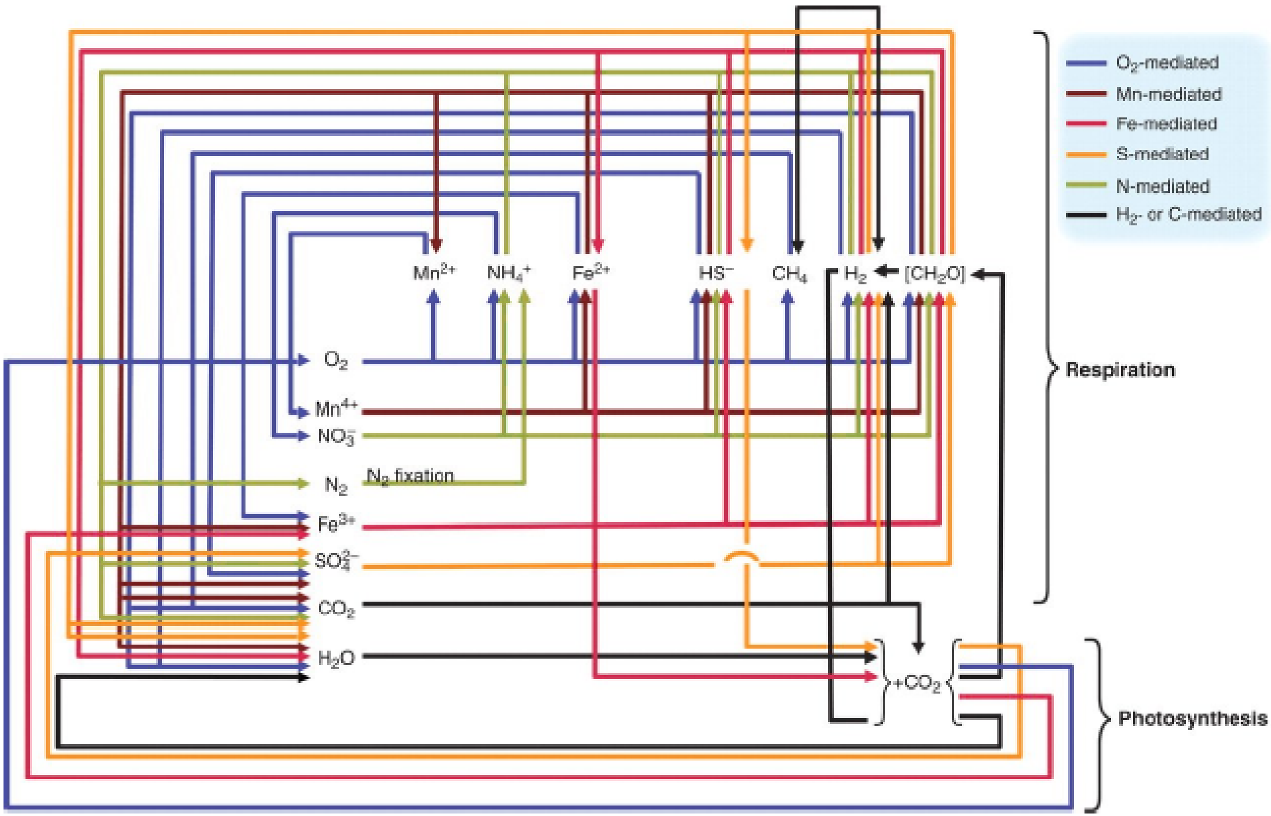




"shallow ponds of condensed and cooled geothermal vapor that were lined with porous silicate minerals mixed with metal (primarily Zn) sulfides and enriched in K+, Zn2+, and phosphorous compounds."

Armen Y......... Eugene V. Koonin, 2012 PNAS

(1) Life *is*...... energy/nutrient cycling



"The individual taxonomic units evolve and go extinct, yet the core machines survive surprisingly unperturbed."

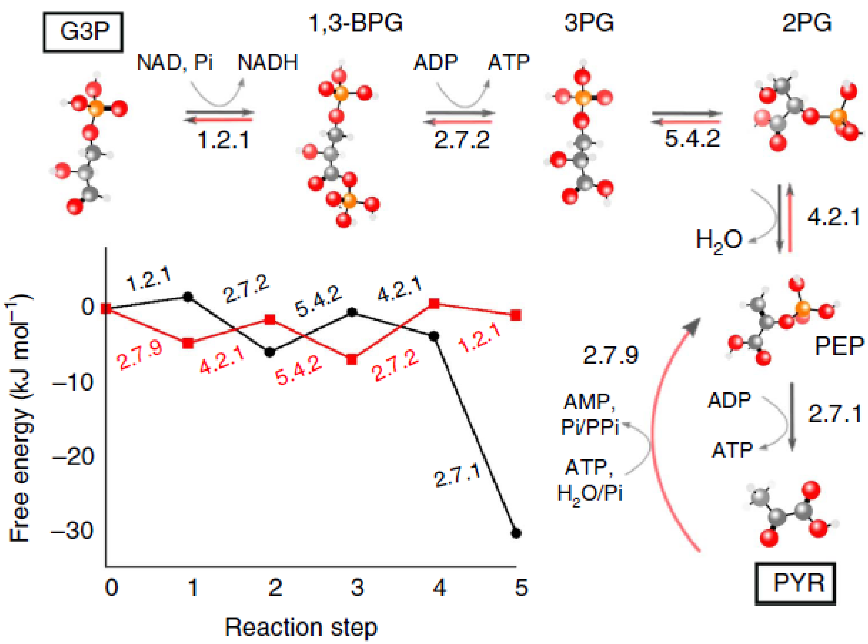
PG Falkowski et al, Science 2008

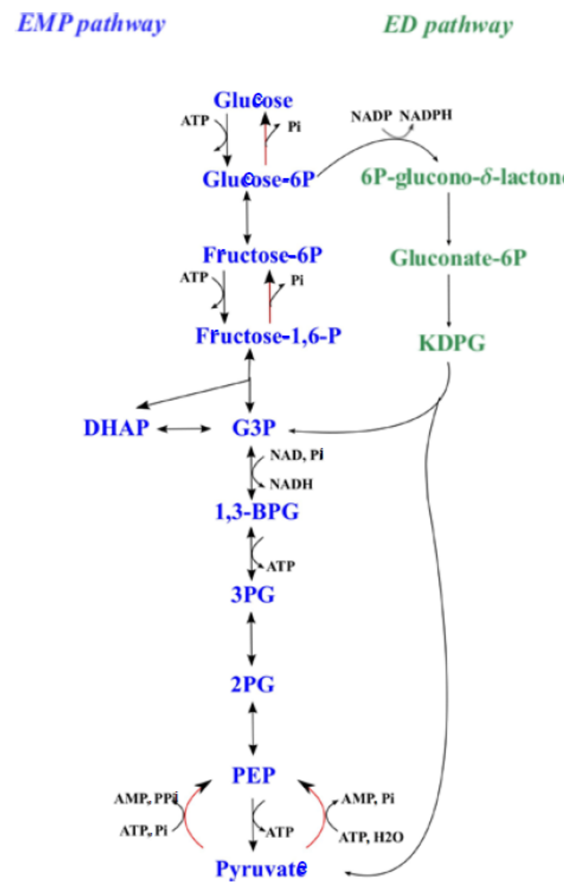
conserved metabolic pathway: glycolysis/gluconeogenesis

WHY??

“unique?” , “optimal?”

contingency?, (evolvability?)





Court, Waclaw & Allen 2015

Mapping all possible trunc pathways
Glyceraaldehyde 3 phosphate to pyruvate

All (Ĩ500)
unbranched aliphatic
CHOPN upto 4 carbon,
negatively charged

free energy for formation

All possible reaction

All possible pathways which

produce (at least) 2 ATP

length 4,5 or 6 $->$

1787 glycolysis pathways

6445 gluconeogenic pathways

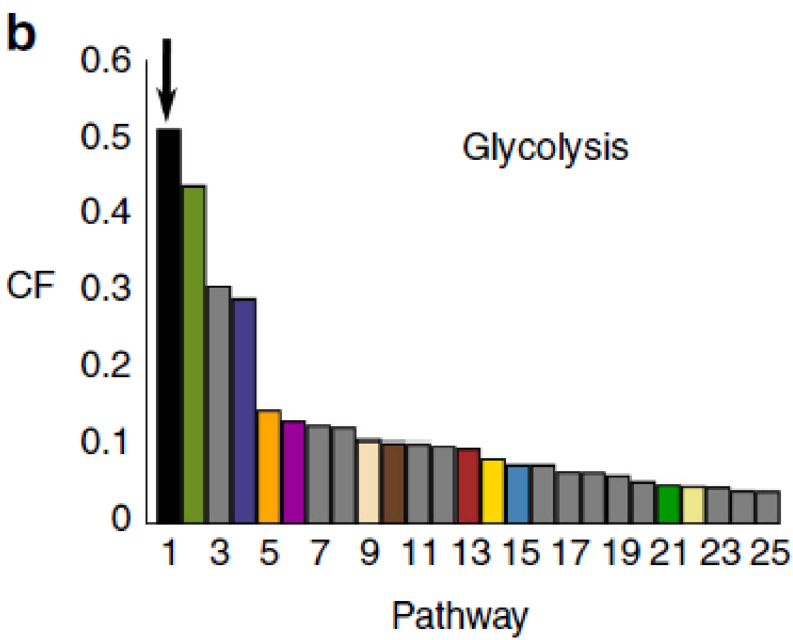
EC class	
1.1.1	Oxidation
1.2.1	phosphorylation
1.3.1	
1.4.1	Deamination
2.6.1	Transamination
2.7.1	Phosphate transfer
2.7.2	
2.7.9	
3.1.3	Hydrolysis
3.5.1	
3.6.1	
4.1.1	Decarboxylation
4.2.1	Dehydration
4.3.1	Ammonia-lyase
5.3.1	Isomerization
5.3.2	
5.4.2	
5.4.3	
6.3.1	ATP-driven amine ligase
6.4.1	ATP-driven carboxylation

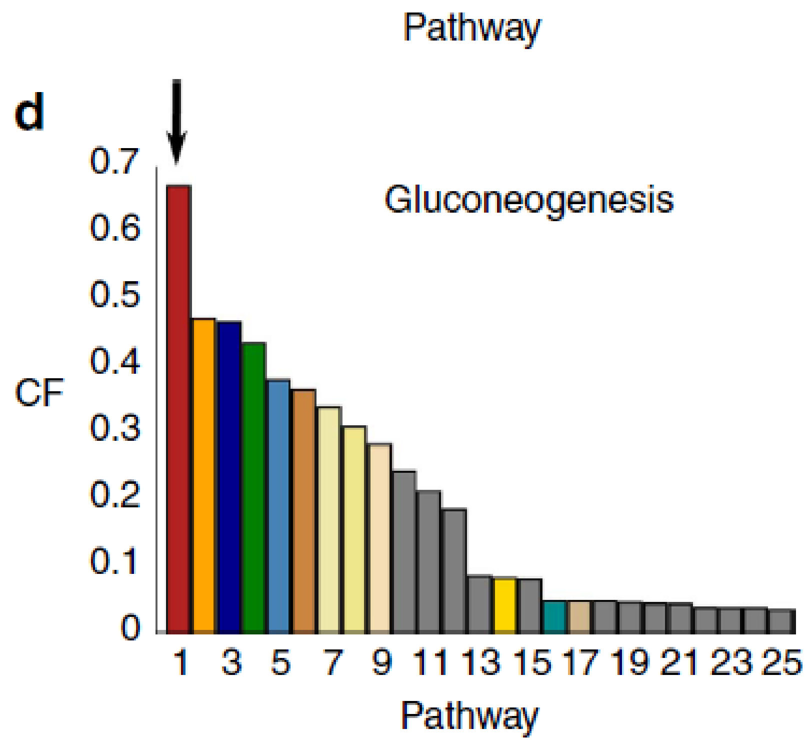
optimality of alternative pathways (maximal flux)

Sample 10000 conditions of
11 external metabolites + G3P and Pyruvate
(log sampling around typical existing levels)

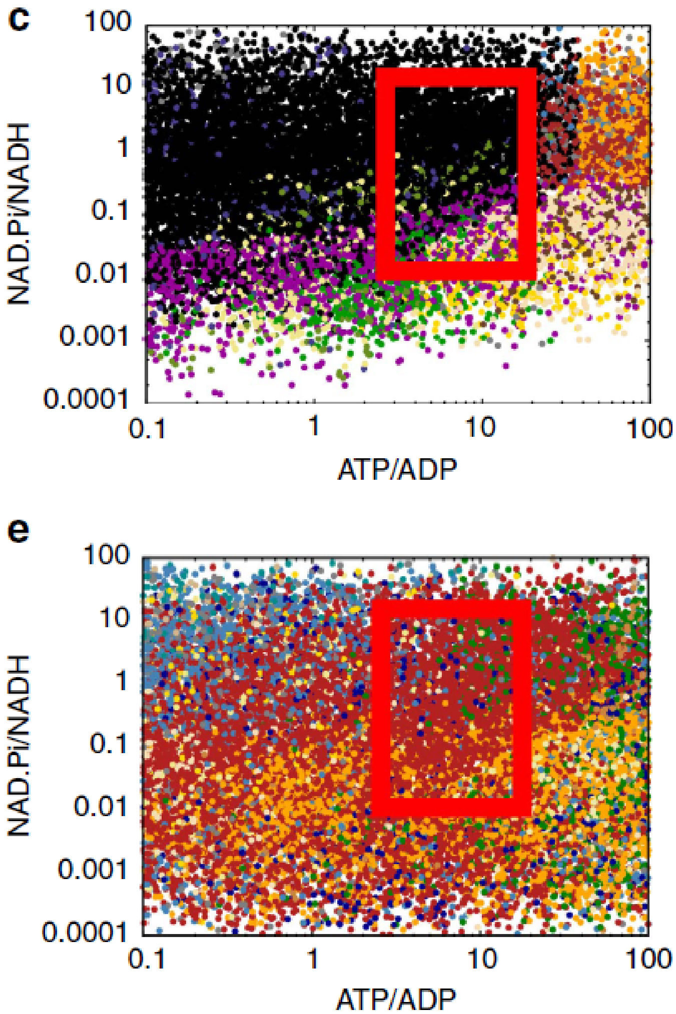
Limit internal metabolite conc. 0.1- 100mM

average relative flux in all samples





different optima for diffrent conditions



Parameter	Range sampled
[source]	1μM to 1 mM
[source]/[product]	0.01 to 100
[ATP]/[ADP]	0.1 to 100
[NAD]/[NADH]	0.1 to 1000
[AMP]	0.01 to 0.1 mM
[Pi]	1.0 to 100 mM
[PPi]	0.1 to 10 mM
[CO_2]	1.0 μM to 0.1 mM
[NH_3]	1.0 μM to 0.1 mM
[GLUT]	1.0 to 100 mM
[2-OXO]	0.1 to 10 mM

alternative 'bests'

sampling space

“life is evolution”

“Biological systems are distinguishable from chemical systems because they contain components that have many potential alternativ e compositions but adopt a particular composition based on the history of the system. In this sense biological systems have a molecular memory (genotype), which is shaped by experience (selection) and maintained by self-reproduction”

Joyce (2012) Bit by Bit: The Darwinian Basis of Life:

“How many heritable bits are involved, and where did they come from

Evolution-first scenario of the origin of life
RNA world

The RNA world hypothesis:

the worst theory of early evolution of life (except for all the others)
(Harold S Bernhardt, Biology Direct 2012)

- RNA for information storage and amplification
- (template and catalyst)
- potential 'generic' replication
- RNA in core processes of current biological systemsv
- New (old?) catalytic functions easily evolvable

HOWEVER
Many chemical caveats raised, and partially solved...
BUT

"The presumed RNA world should be viewed as a milestone, a plateau in the early history of life on Earth. So, too, the concept of an RNA world has been a milestone in the scientific study of life's origins. Although this concept does not fully explain how life originated, it has helped to guide scientific thinking and has served to focus experimental efforts"
Protocells and RNA Self-Replication
Gerald F. Joyce1 and Jack W. Szostak 2018
review of and towards RNA world from chemical point of view.

First polynycleotide synthesis without prior metabolism?

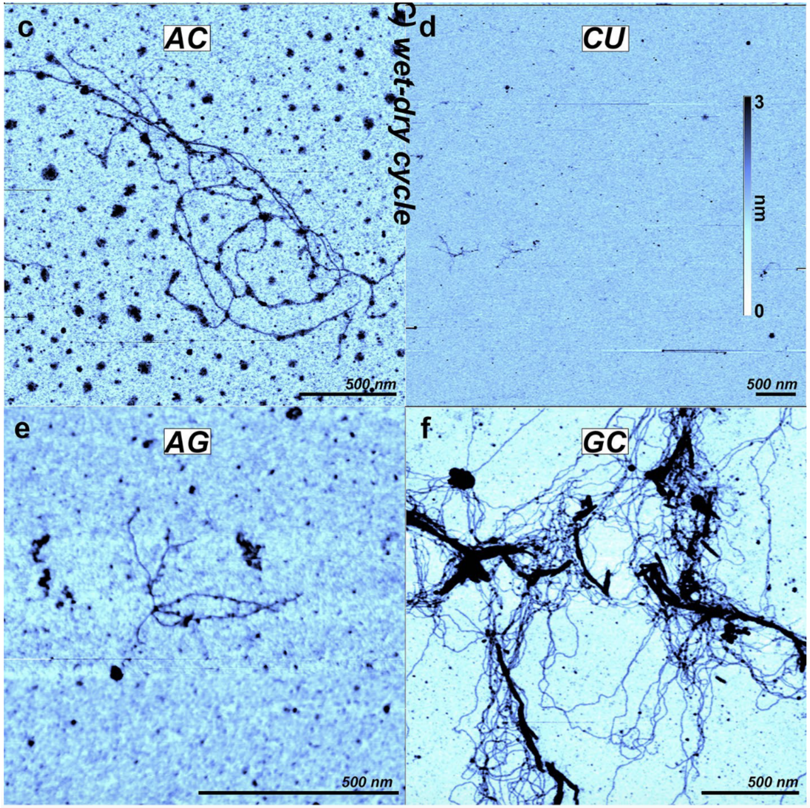
wet-dry cycling
as in hydrothermic ponds
enough energy for
RNA polymerization?

Visualizing RNA polymers

produced by hot dry-wet cycling

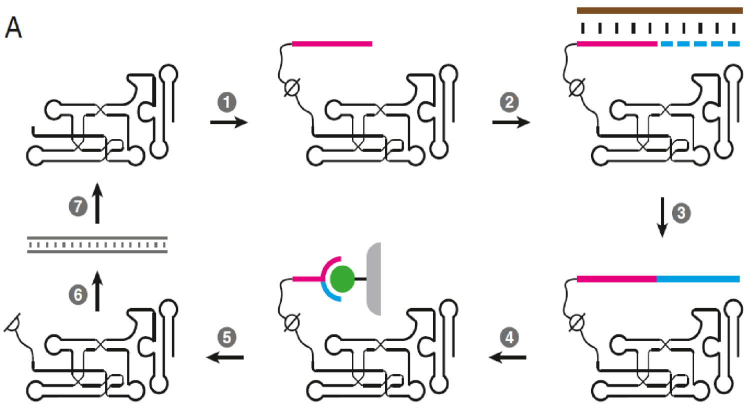
(Hassenkam & Deamer 2022

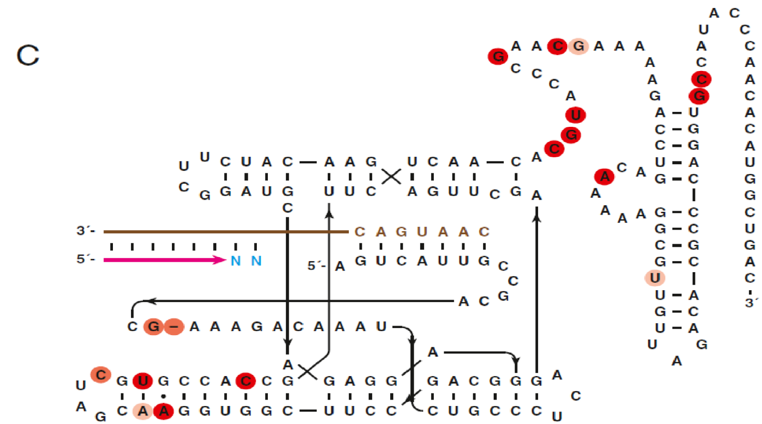
Scientific reports

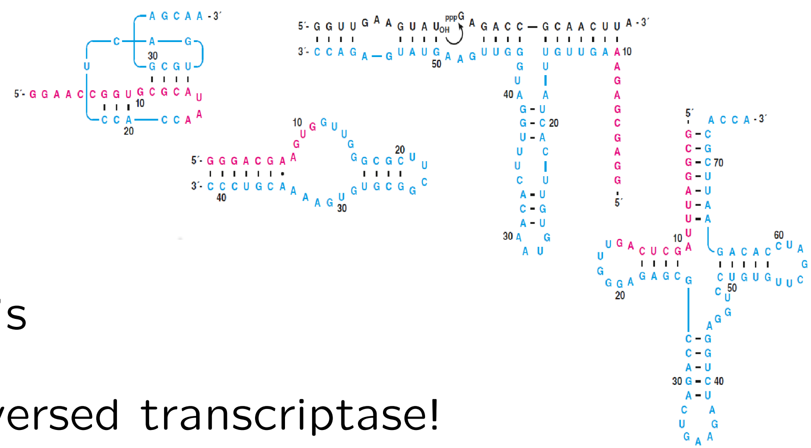


Continuing efforts to evolve improved RNA polymerases 24th round: Horning Joyce PNAS 2016

RNA dependent RNA polymerase ev0lved from a ligase (Bartel & Szostak 1993), and improved by design and evolution to current form:







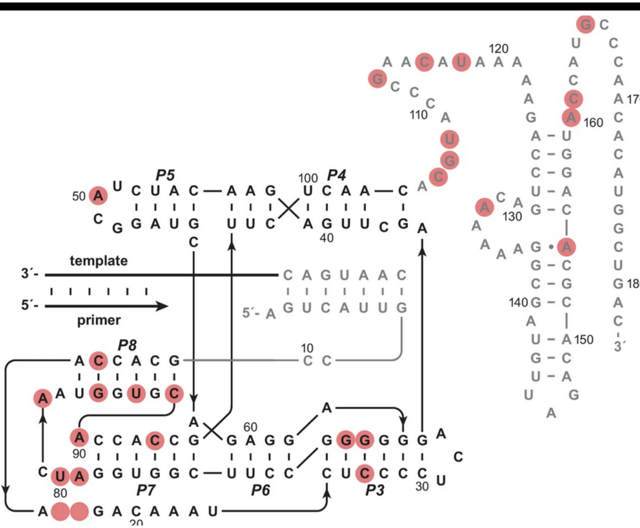
Replicted RNA's

Works also Reversed transcriptase!

tRNA

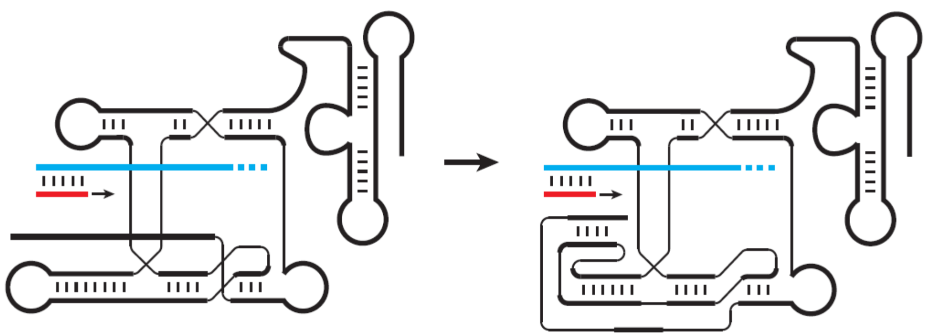
Current state after 56 rounds (26 mutations):

can replicate own ancestor (ligase) Portillo ...Joyce 2021 elife



”Now simulateous selection on yield and fidelity (– >length)”

Joyce 2024



RNA world hypothesis

here assumed as starting point for developing

bioinformatic theory prebiotic evolution

focusing on informatic rather than chemical constraints
AND as starting point for
modeling biotic systems as
evolving mutilevel information processing systems

Informatic potential and limitation of RNA world hyptothesis

limited evolvability?

minimal evolution system, replicator equation

- 'generic replicators'
- independent synthesis and decay
- mutation
- competition

Eigen: Replicator Equation (in chemostat)

$$dX_i/dt = A_iQ_iX_i - d_iX_i + \Sigma w_{ij}X_j - \Omega_i$$

$\Omega_i = (X_i/\Sigma X_j)\Sigma(A_j - d_j)X_j$ (constant populatio size)

$w_{i,i} = A_iQ_i - d_i$
$w_{ij} = \mu_{ij}A_j(1 - Q_j)$

Converges to eigenvector belonging to largest eigenvalue of W

== Quasispecies (wildtype)

growthrate Quasispecies: largest eigenvalue

HOW DOES COMPLEXITY ARISE THROUGH DARWINIAN EVOLUTION ?

V

Survival of the fittest does not imply increase of complexity

On the contrary:

(If we take genome size as meassure of complexity)

more complexity tends to decrease replication rate

MOREOVER

Classical caveat for the evolvability of complexity

Error catastrophe (Information threshold) (Eigen 1971)
limited replicator complexity

Error catastrophe
Survival of the fittest (quasi)species ?

Assume: $w_{ii} >> w_{jj} == C$ for all j

Simplify: lump all j into one quation (Y)

$$dX/dt = a_1 Q_1 X - d_1 X - X((a_1 - d_1)X + (a_2 - d_2)Y)$$

$$dY/dt = a_2 Y - d_2 Y + a_1(1 - Q_1)X - Y((a_1 - d_1)X + (a_2 - d_2)Y)$$

Is the fittest (X) selected? Does it "survive"

Can the fittest invade?

$X > 0$ iff dX/dt > 0 close to $X = 0$

$$a_1 Q - d_1 > a_2 - d_2$$

$Q > a_2/a_1 == 1/\sigma$ (assuming $d_1 = d_2$)

Darwinian evolution only when mutation rate small enough

(survival of the fittest NOT a tautology)

Similar for full model: N species with back mutation.

From error threshold to information threshold

If replicator is a information carrying polymer (e.g. RNA or DNA)

CCCCCCCCCCCCCGACACGGAAACGACGUGAGAGUCAUUAGAUAGGUGUC

Mutations can happen at each position (nucleotide) (with rate 1-q)

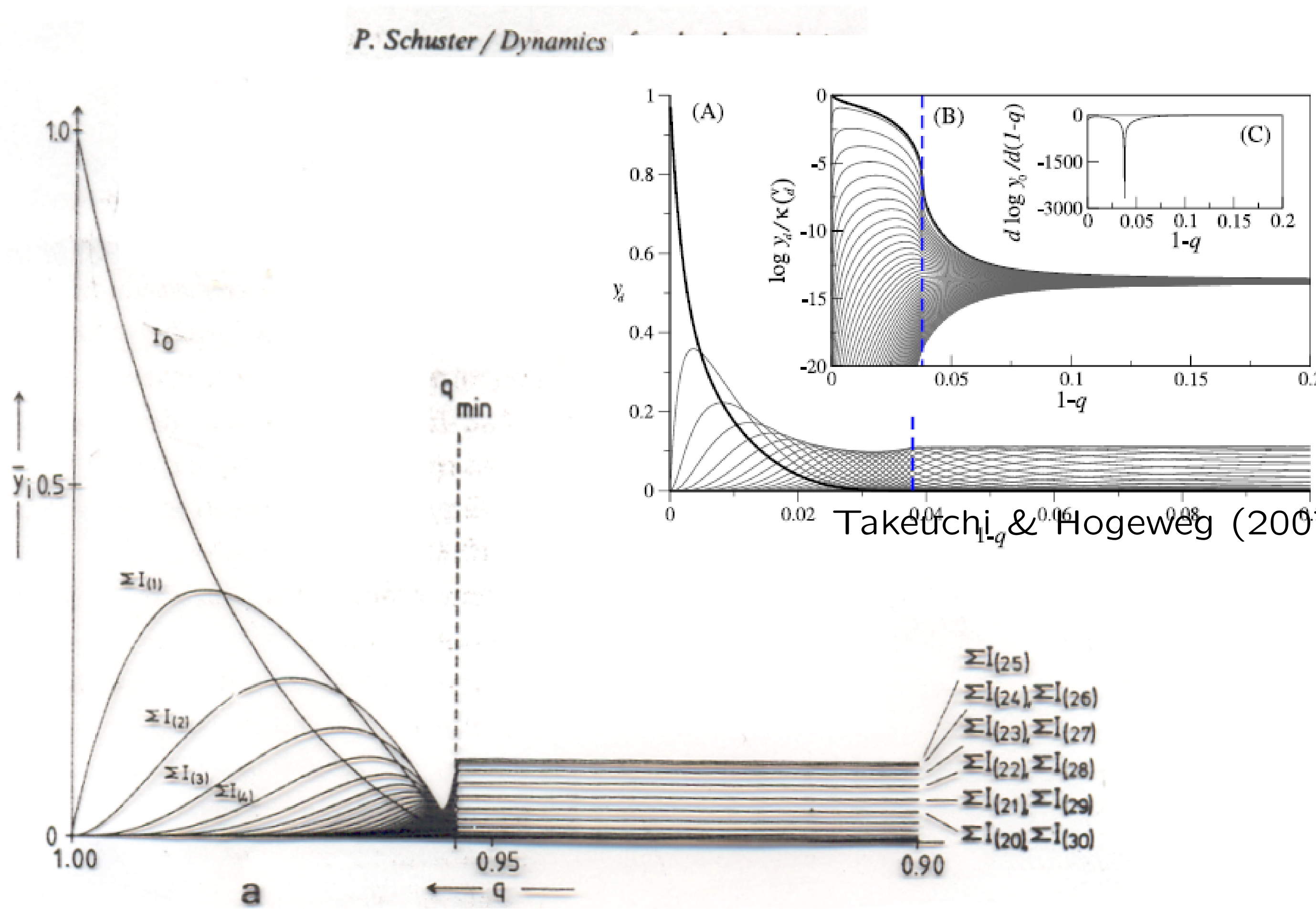
$$Q = q^L = e^{-L(1-q)}$$

$$L < ln\sigma/(1-q)$$

Given a mutation rate, given selction coefficient
Length of sequence limited $-- >$

Only limited information accumulation possible

Error threshold / Information threshold



v

Takeuchi & Hogeweg (2007)

this is 'best' case scenario

- infinite population size
  - always to 'best' quasispecies
  - no stochastic population dynamics
  - no extinction (everybody viable–replicatable)
- strong selection, single peak landscape
  - therefore sharp transition (threshold)
  - delocalization vs threshold
- fixed length - no other constraints
  - NO negative selection on length (rate, energy)