Prebiotic evolution:
replicator systems and information threshold

Course Computational Biology 2018/2019; Paulien Hogeweg;

Theoretical Biology and Bioinformatics Grp Utrecht University

Life as evolved (evolving) complex multilevel information processing system

how to model it?
how did it get started?/ bootstrap itself

studied by
- phylogenetic reconstruction (LUCA)
  quite complex - what before that?....
- 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...

(1) Life *is*...... energy/nutrient cycling
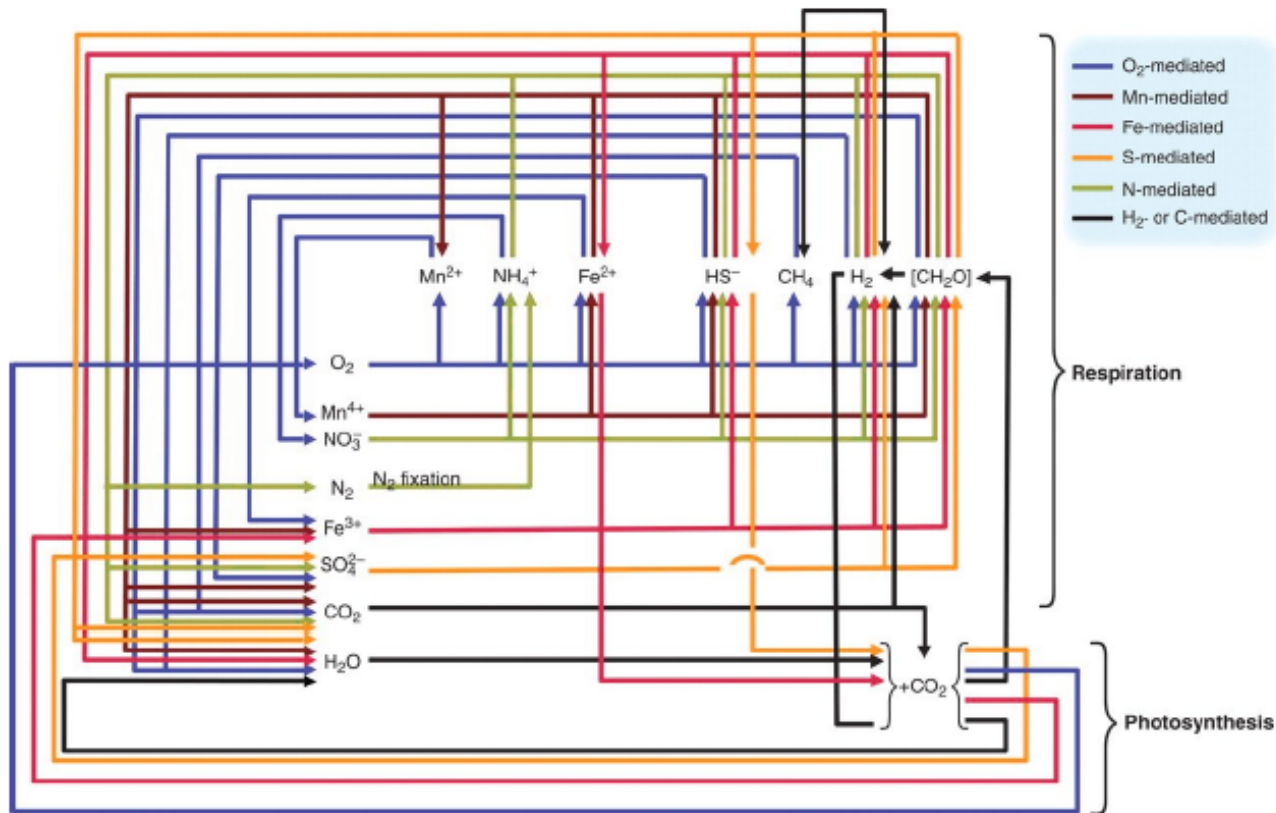
"The individual taxonomic units evolve and go extinct, yet the core machines survive surprisingly unperturbed."
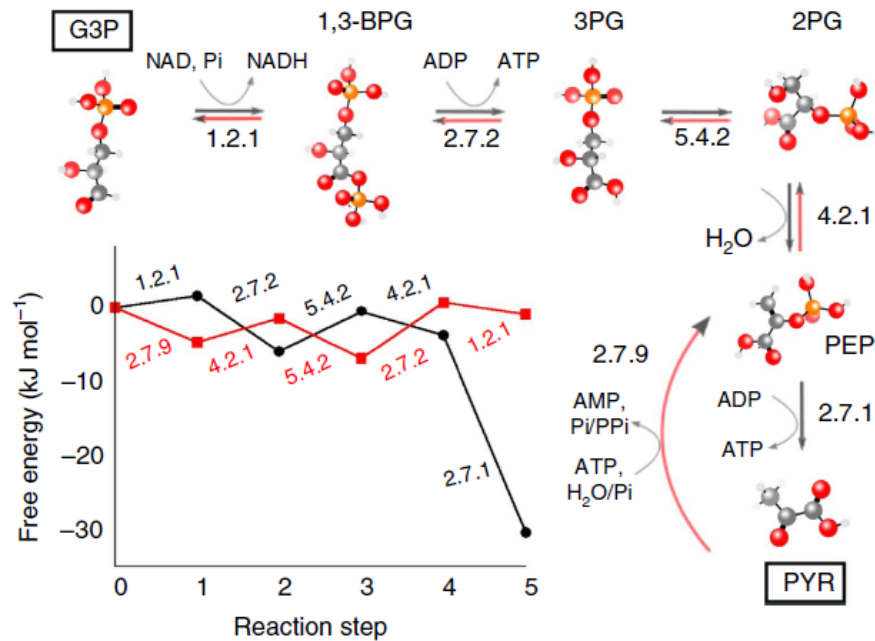
PG Falkowski et al, Science 2008

conserved metabolic pathway: glycolysis/gluconeogenesis
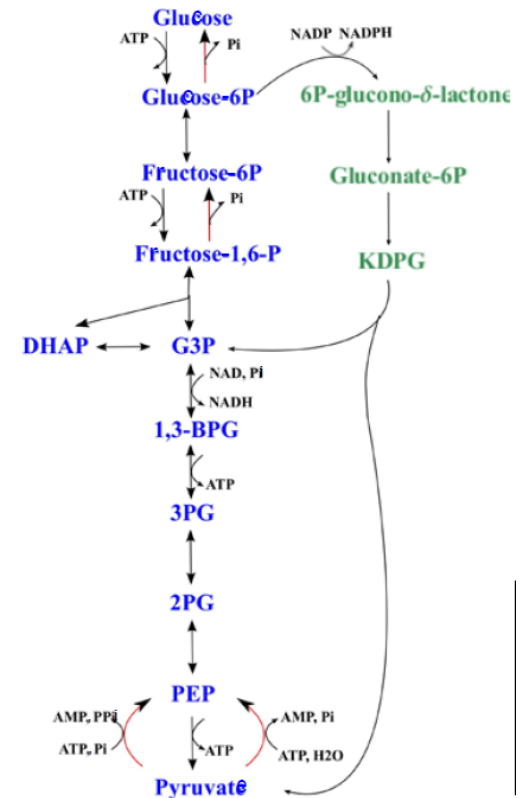
WHY??

"unique?" , "optimal?"

contingency?, (evolvability?)

Court, Waclaw & Allen 2015

Mapping all possible trunc pathways
Glyceraaldehyde 3 posphate to pyruvate

All (~1500)
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 gluconeogenicpathways

| 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

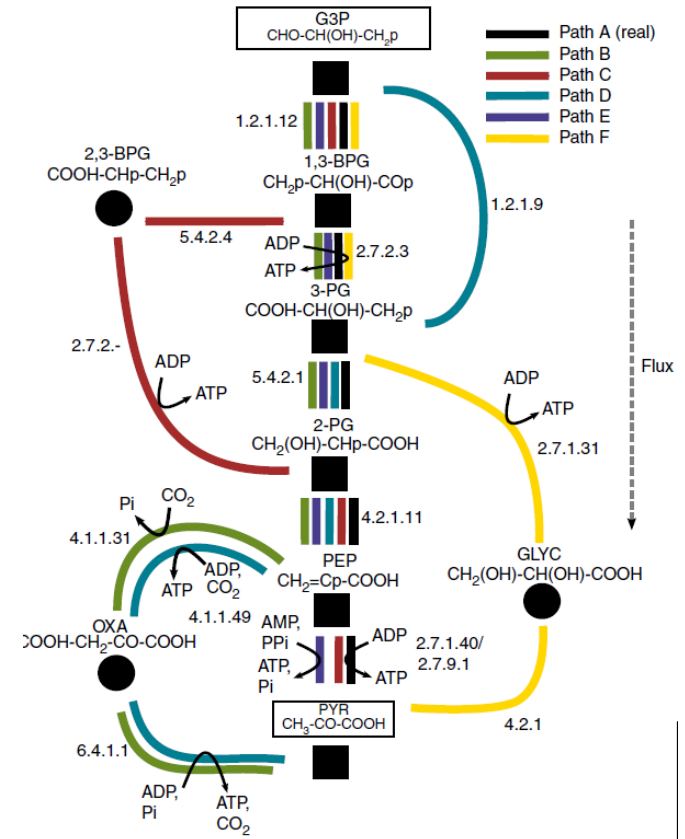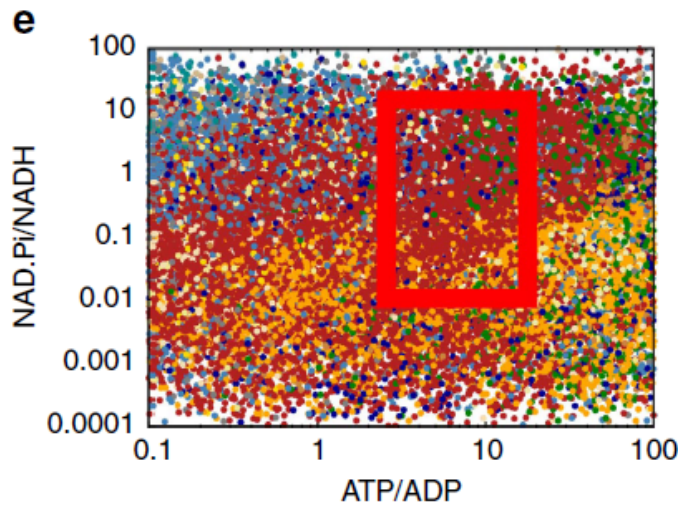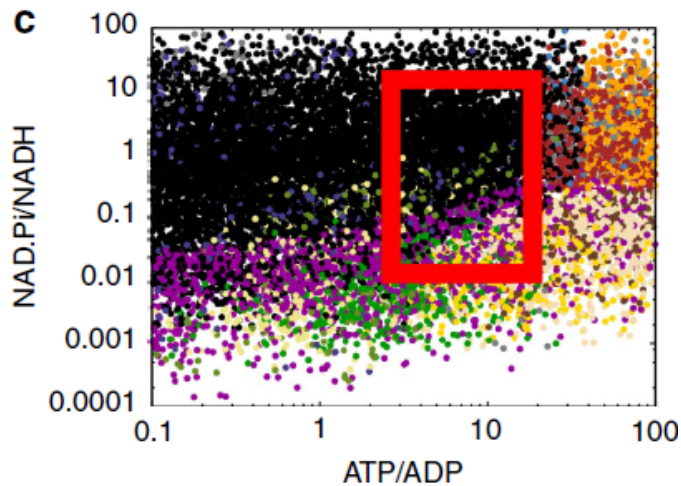# optimality of alternative pathways (maximal flux)

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
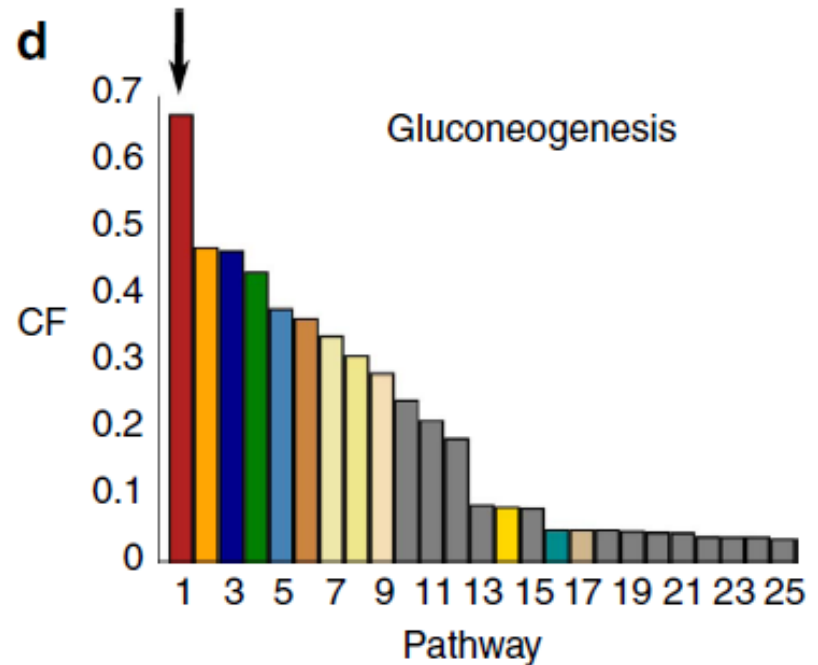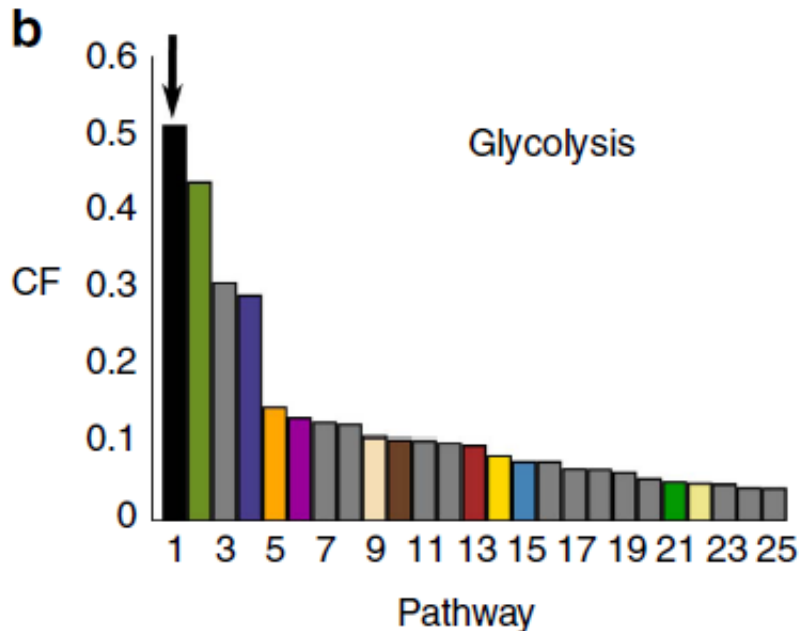
# different optima for diffrent conditions

alternative 'bests'   alternativ glycolysis (within KEGG)

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
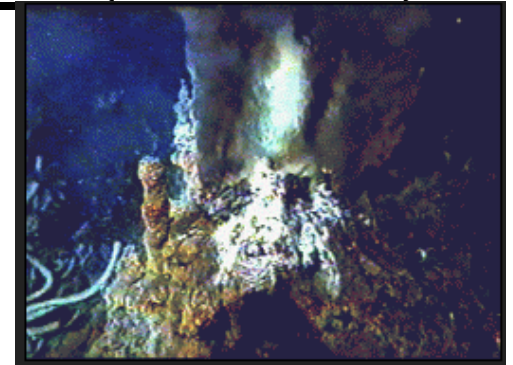
OR

Origin of first cells at terrestrial, anoxic geothermal fields Because of 'open' cell environment should match internal cell composition
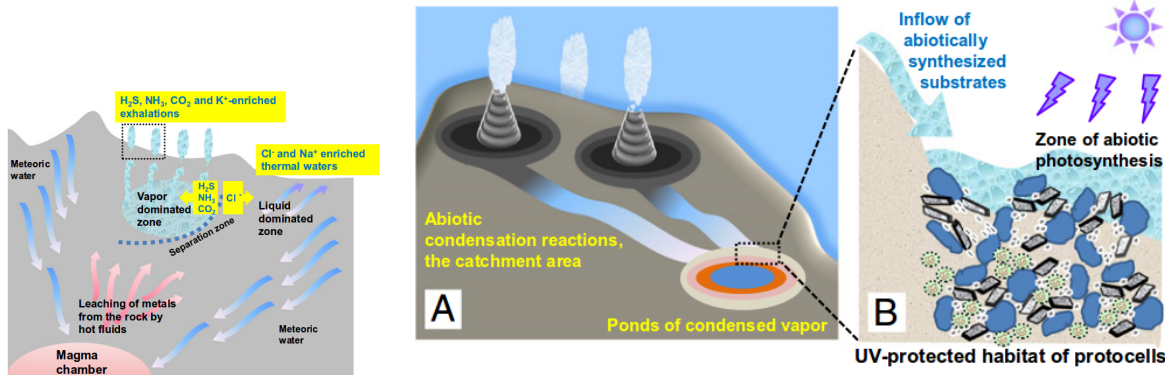
"shallow ponds of condensed and cooled geothermal vapor that were lined with porous silicate minerals mixed with metal (primarily Zn) sulfides and enriched in K+, Zn2+, and phosphorous compounds."

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

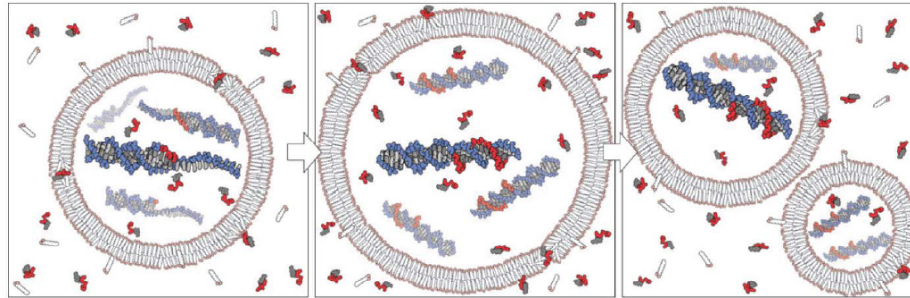# "Life is cells"

vescicle-first scenario's of origin of life

e.g. Szostak (2011, review) Life requires compartimalization (protocells)

Construct minimal chemical system
preferrable in heterogeneous environment

" optimal degree of physical and chemical heterogeneity for the origin of life?"

Heterotrophic protocells, Jack Szostak

Growth of the protocell fatty acid based membrane results from the incorporation of environmentally supplied amphiphiles, whereas division may be driven by intrinsic or extrinsic physical forces. Externally supplied activated nucleotides permeate across the protocell membrane and act as substrates for the non-enzymatic copying of internal templates. Complete template replication followed by random segregation of the replicated genetic material leads to the formation of daughter protocells.

Short, activated oligonucleotides them- selves plausibly generated by either templated or untemplated monomer polymerization are efficient catalysts of high fidelity primer extension with all four RNA monomers.

(e.g. Mansy ….Szostak Nature 2008 Szostak JW.2012 Journal of Systems Chemistry 2012, Prywes ..Szostak Elife 2016)

"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 the 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
systems
New (old?) catalytic functions easily evolvable

HOWEVER

Many chemical caveats Nucleotide synthesis hard/impossible (but Powder
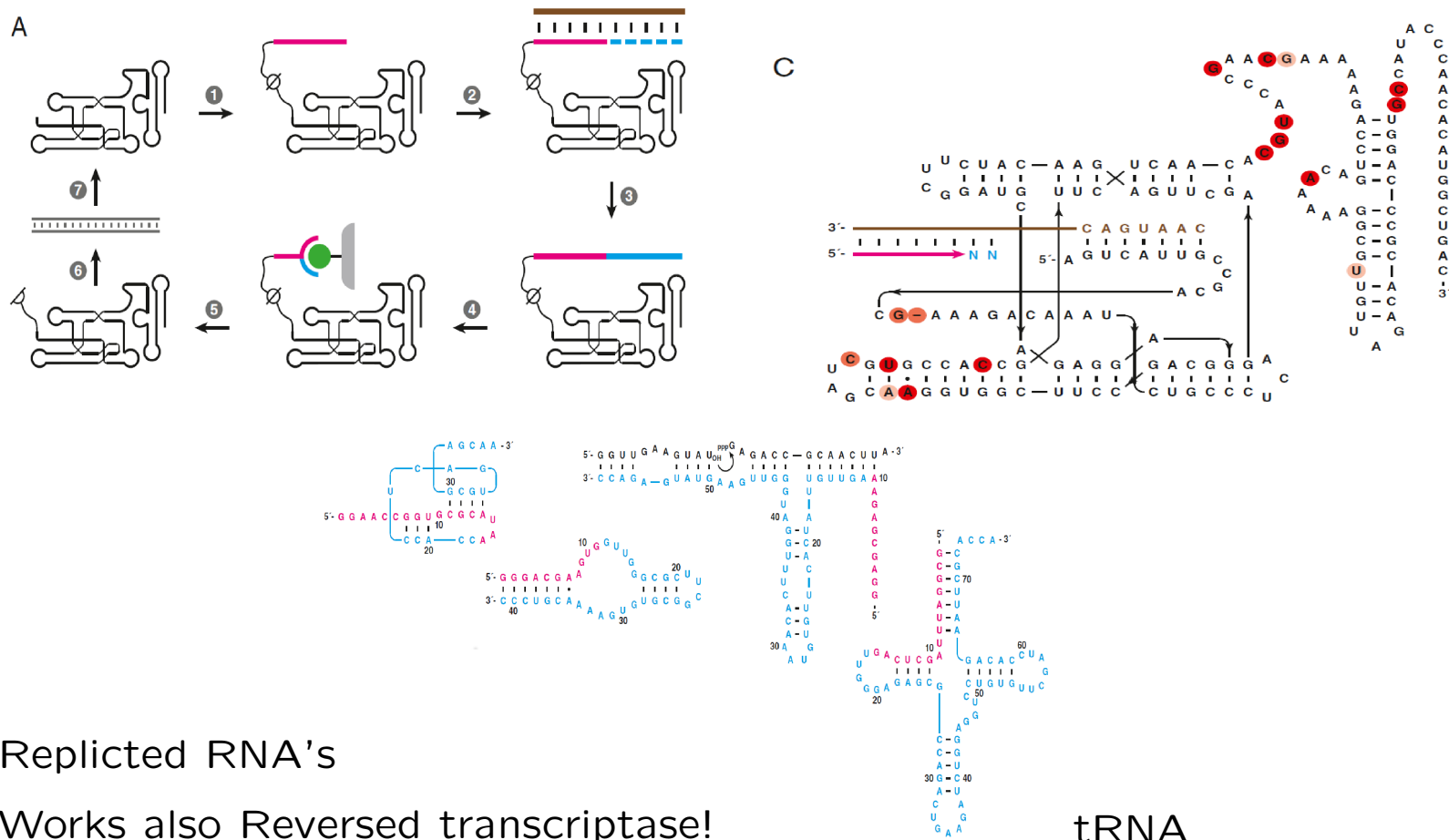2009)
RNA template replication slow - error prone (but see Szostak above)
specific catlysis/templates 'work'

BUT progress....

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

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 requirements for(Darwinian evolution
(cf definition of life of Joyce)
"RNA world without chemical constraints"

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

Sufficient to bootstrap, from small RNA's?

Eigen: Replicator Equation (in chemostat)

$$dX_i/dt = A_i Q_i X_i - d_i X_i + \Sigma w_{ij} X_j - \Omega_i$$

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

$$E(t) = (\Sigma(A_j - d_j)X_j)/\Sigma X_j$$

$-->$ **Quasispecies (='wildtype')**
=eigenvector of max eigenvalue of $\hat{\mathbf{W}}$
$-->$ **Information Threshold**

Error Threshold and Information Threshold

"Survival of the fittest" only if mutation rate small enough.

Illustrate by simplifying to 2 equations + No 'back-mutations'

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

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

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

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

$$Q > a_2/a_1 == 1/\sigma \qquad \text{(assuming } d_1 = d_2\text{)}$$
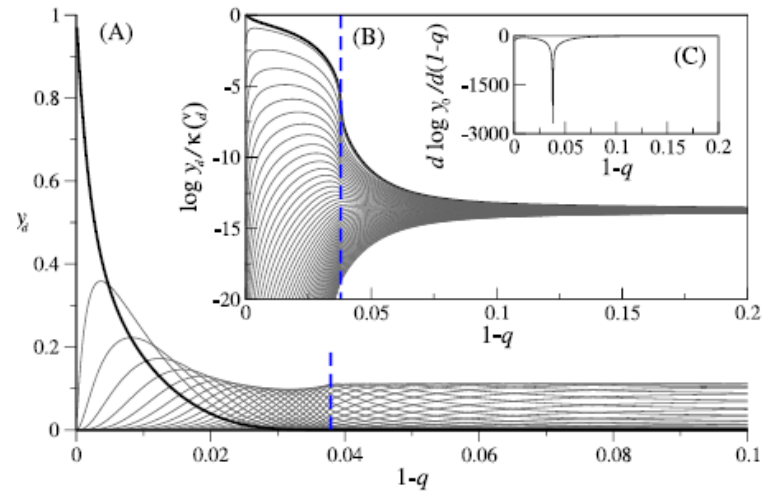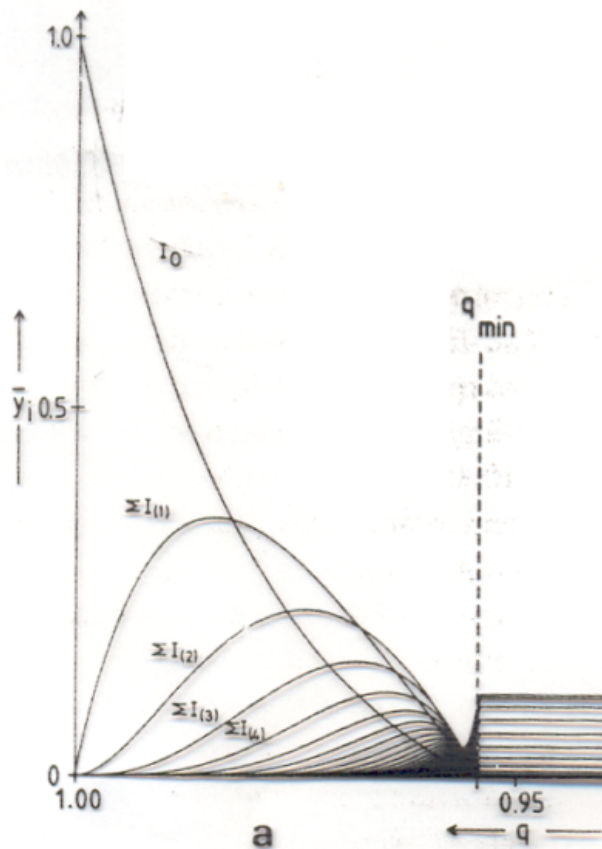
Error threshold $--> $ Information Threshold

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

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

Only limited information accumulation possible for given error rate

Error threshold / Information threshold

P. Schuster / Dynamics of molecular evolution

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)

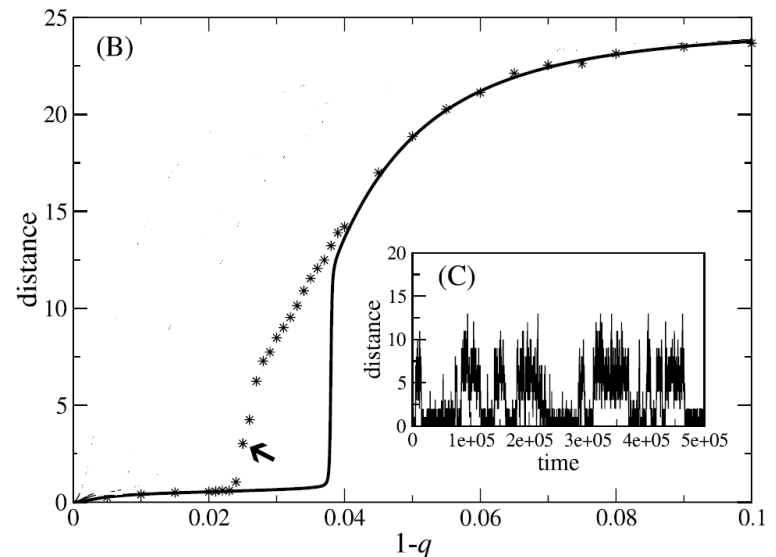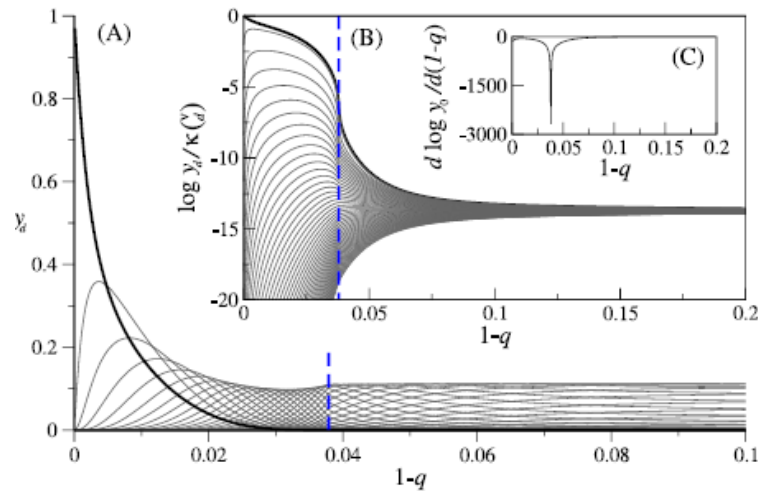# Delocalization but no threshold for exponential fitness landscape
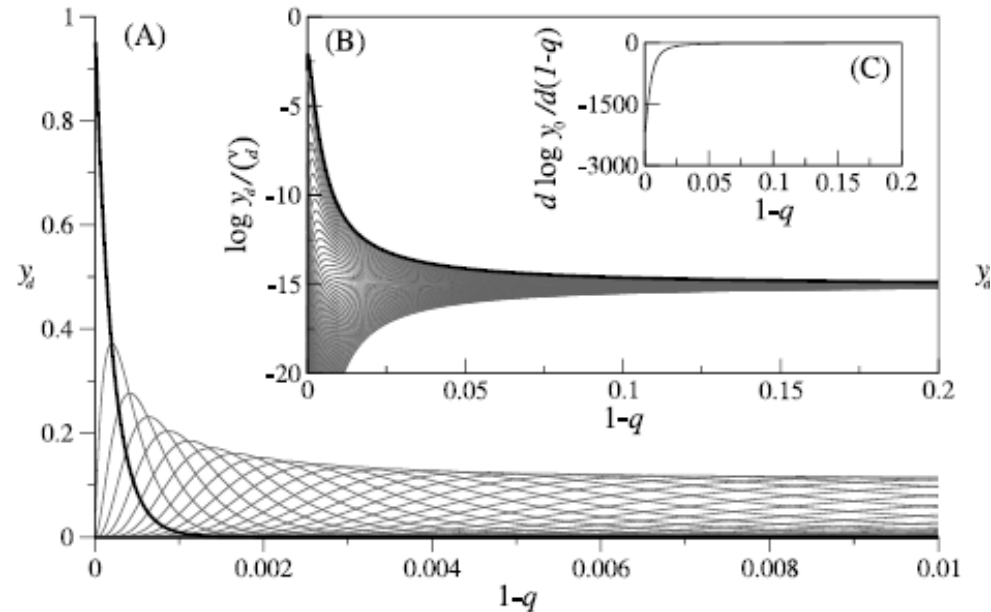
Takeuchi & Hogeweg (2007, BMC-evol)

However, if also lethal mutations - there is a sharp threshold

Before the error threshold common ancestor is master sequence, beyond the error threshold NOT

Common Ancestor: D to master seq.

cf Takeuchi & Hogeweg (2007); Hermisson J, Redner O, Wagner H, Baake E. 2002