From Chemistry to Biology: Getting the RNA World started

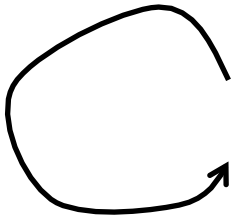
Paul Higgs
Andrew Tupper, Ye Eun Kim
Julie Shay, Chris Huynh, Meng Wu

McMaster University, Hamilton, Ontario.

Supported by:
• NSERC
• Harry Lonsdale (Origin of Life Research Award)

Definition of the RNA World

Modern Organisms.



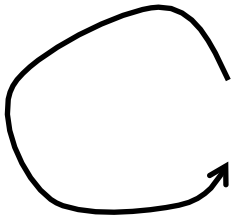
DNA genes → mRNAs → proteins

DNA replication
Catalyzed by
proteins

Transcription
Catalyzed by
proteins

Translation
Catalyzed by the
ribosome (RNA)

Hypothetical RNA World Organisms.



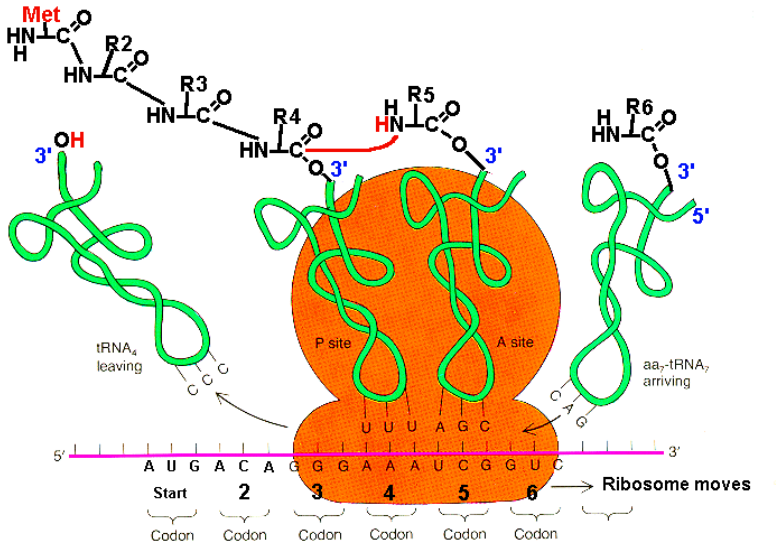
RNA genes

RNA replication
Catalyzed by RNAs

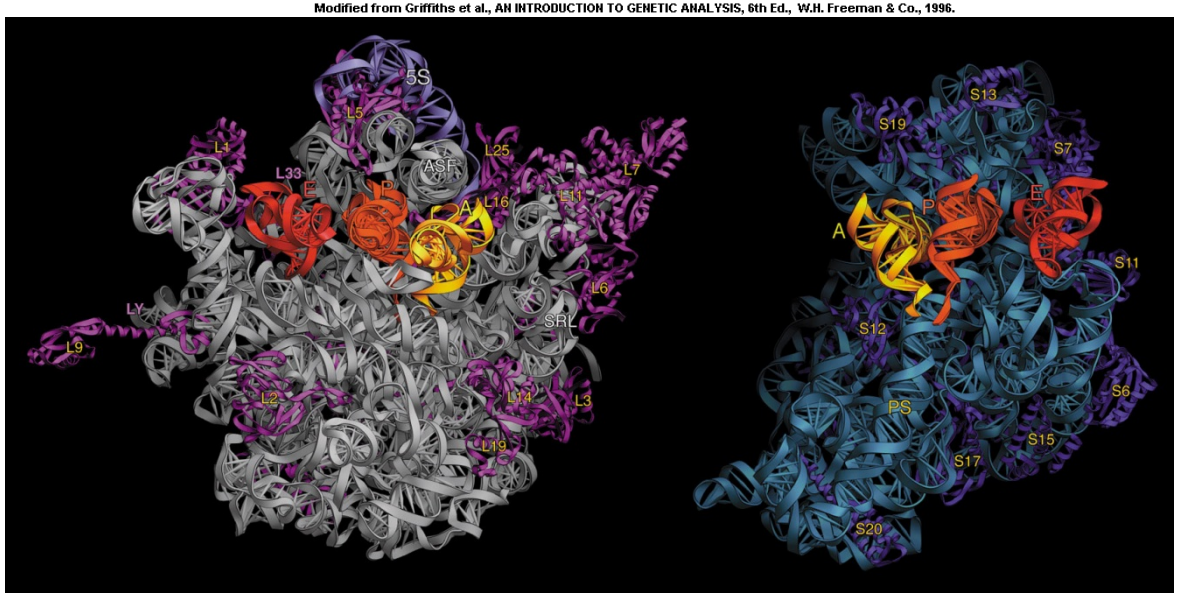
ribozyme =
catalytic RNA
strand

Evidence from current biology

Protein synthesis in modern cells relies on rRNA, tRNA, and mRNA

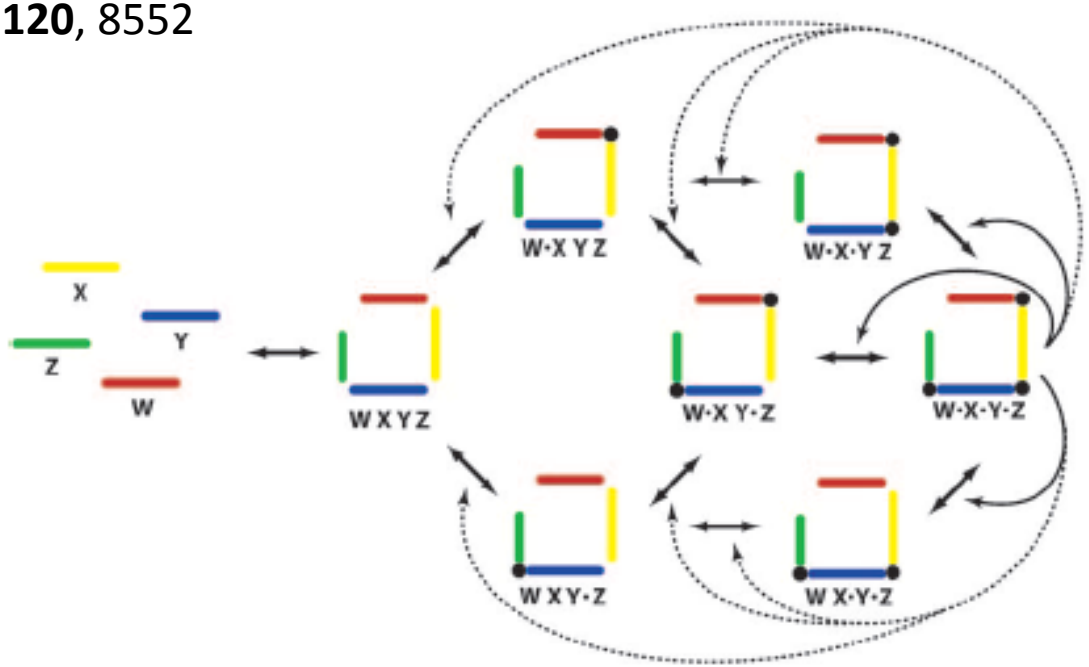


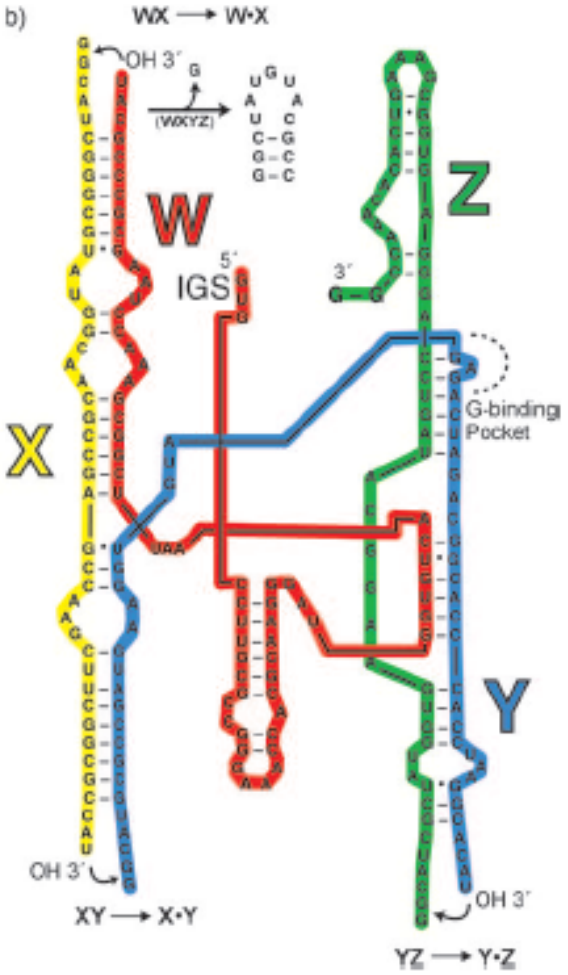
Modified from Griffiths et al., AN INTRODUCTION TO GENETIC ANALYSIS, 6th Ed., W.H. Freeman & Co., 1996.



Laboratory Ribozymes: Recombinases

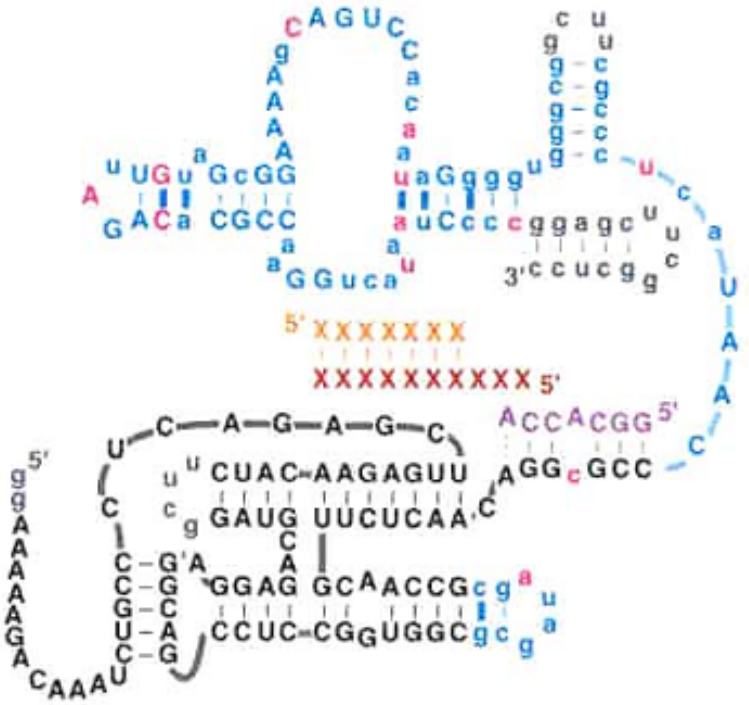
E.J. Hayden, G.v. Kiedrowski & N. Lehman, *Angew. Chem. Int. Edit.* (2008) **120**, 8552

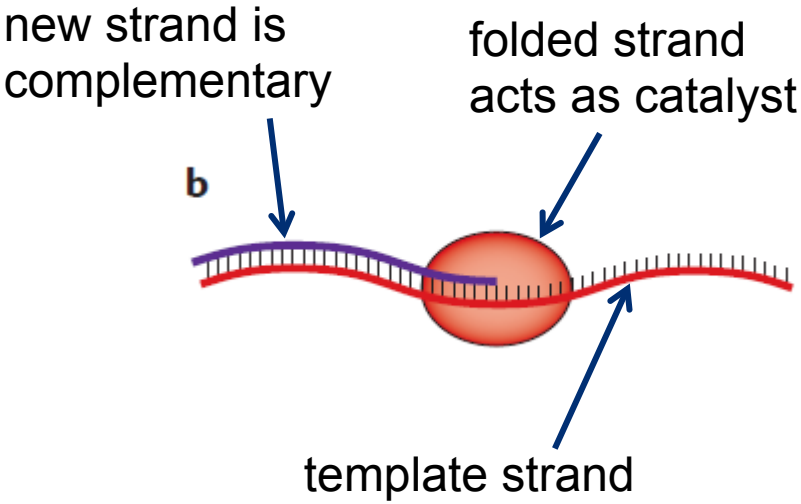




Catalyst is autocatalytic given a supply of W X Y Z.
The non-covalent assembly is also a catalyst.

Laboratory Ribozymes: Polymerases





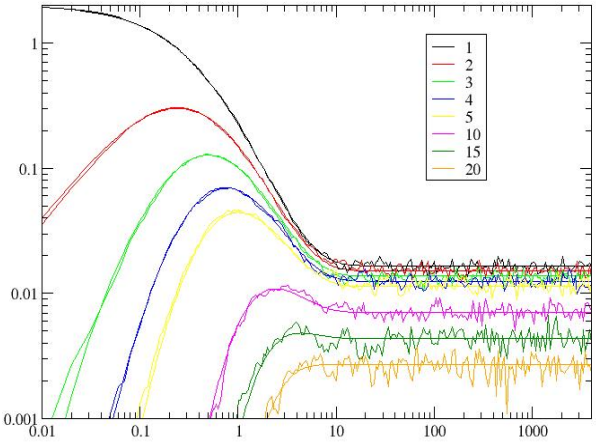
Johnstone et al. (2001) *Science* - Primer extended by up to 14 nucleotides
Wochner et al. (2011) *Science* - up to 95 nucleotides
Attwater et al. (2013) *Nature Chemistry* – up to 200 nucleotides

We're getting there ...

Two key questions for the RNA World (that a theoretical biophysicist can address)

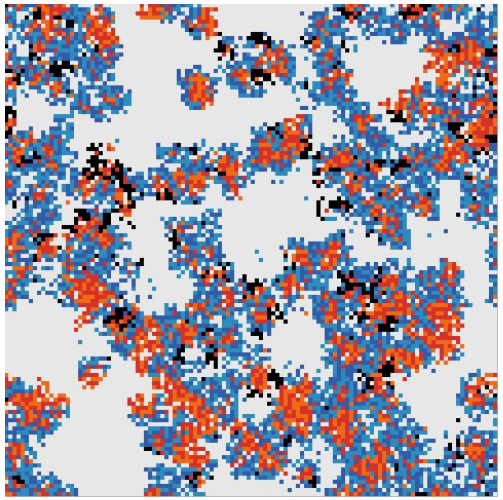
How was RNA synthesized abiotically?

- Wet-dry cycling drives formation of long strands

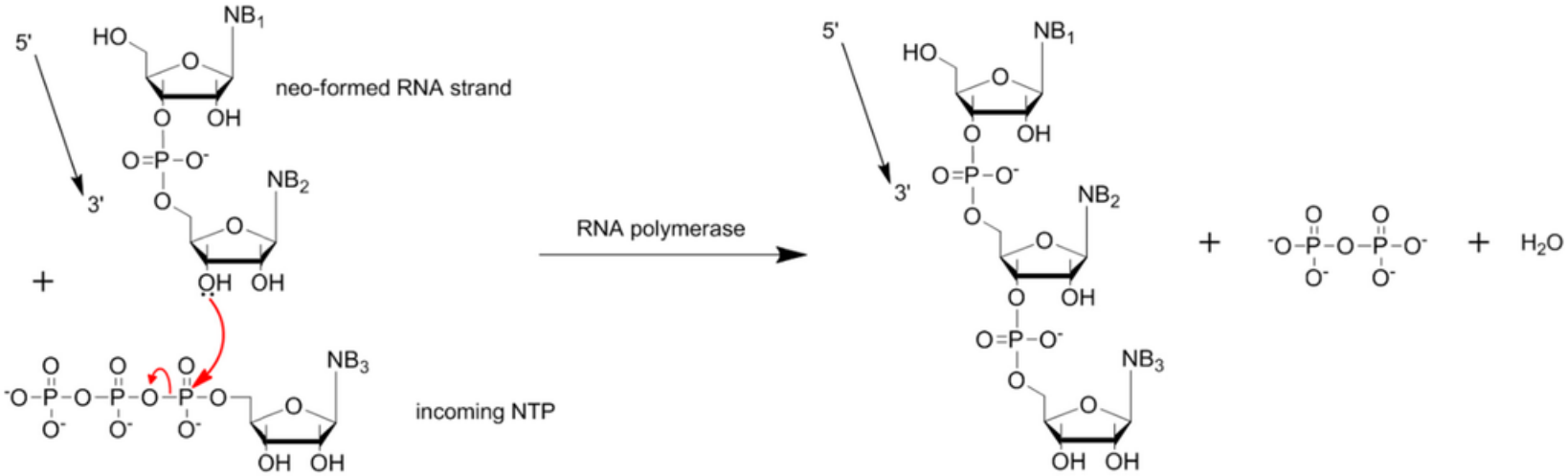


How did replicating sequences emerge from a mixture of random strands?

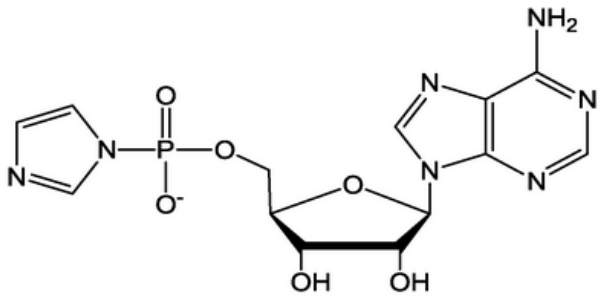
- Evolutionary models for replication in the RNA World.



How to make RNA polymers? – Activated nucleotides



Modern organisms do it using tri-phosphates and a protein catalyst

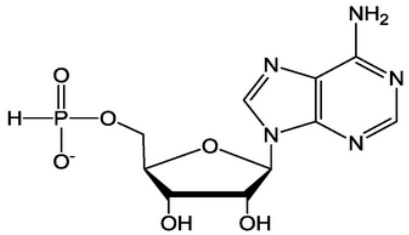


ImpA

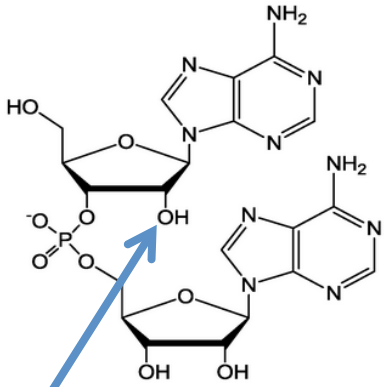
Ferris (2002) *Orig. Life Evol. Biosph.*
Montmorillonite catalyzed synthesis of RNA oligonucleotides (30-50 mers)

Uses imidazole-activated nucleotides

Can we make RNA polymers from mono-phosphates?



AMP



2’ OH initiates hydrolysis in RNA (but not DNA)

Polymerization tends to be up-hill thermodynamically.

Hydrolysis is faster than polymerization.

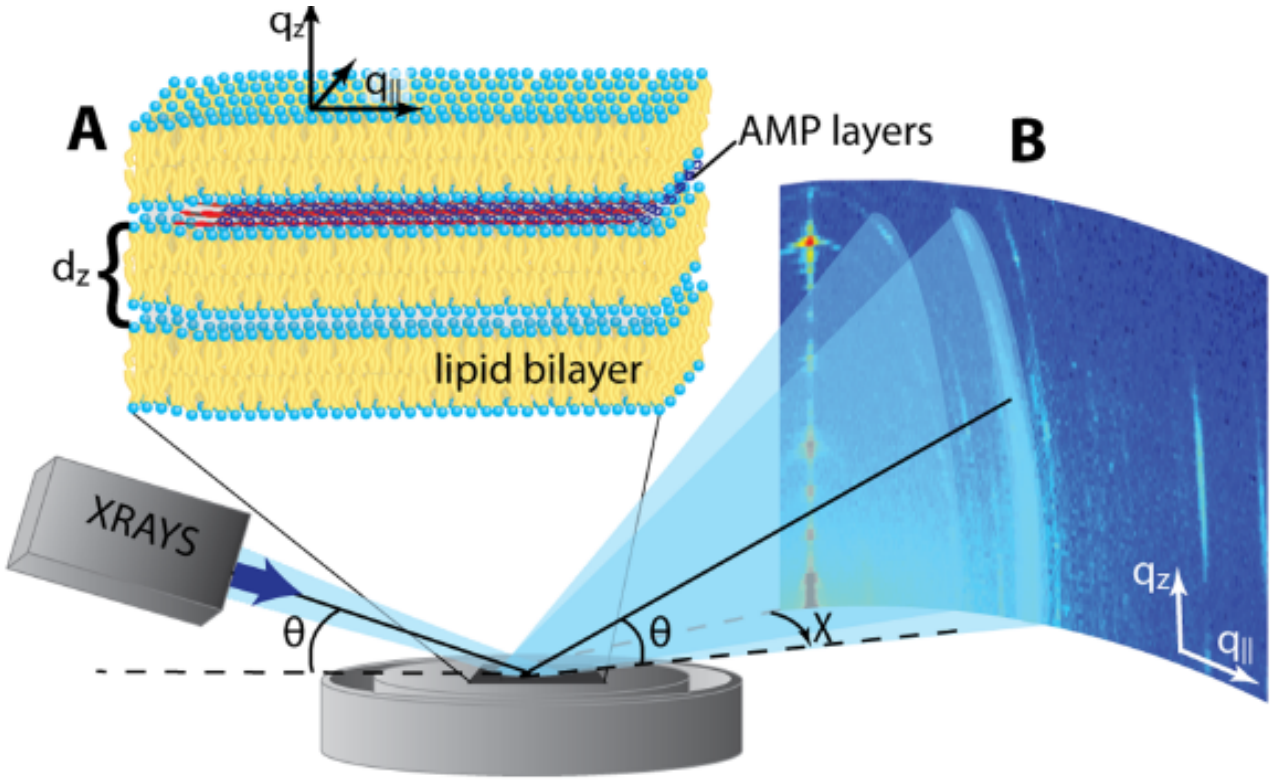
Equilibrium is dominated by monomers and very short oligomers.

Lipid assisted synthesis of RNA-like polymers from mononucleotides.

Drying AMPs between lipid bilayers creates favourable conditions for polymerization

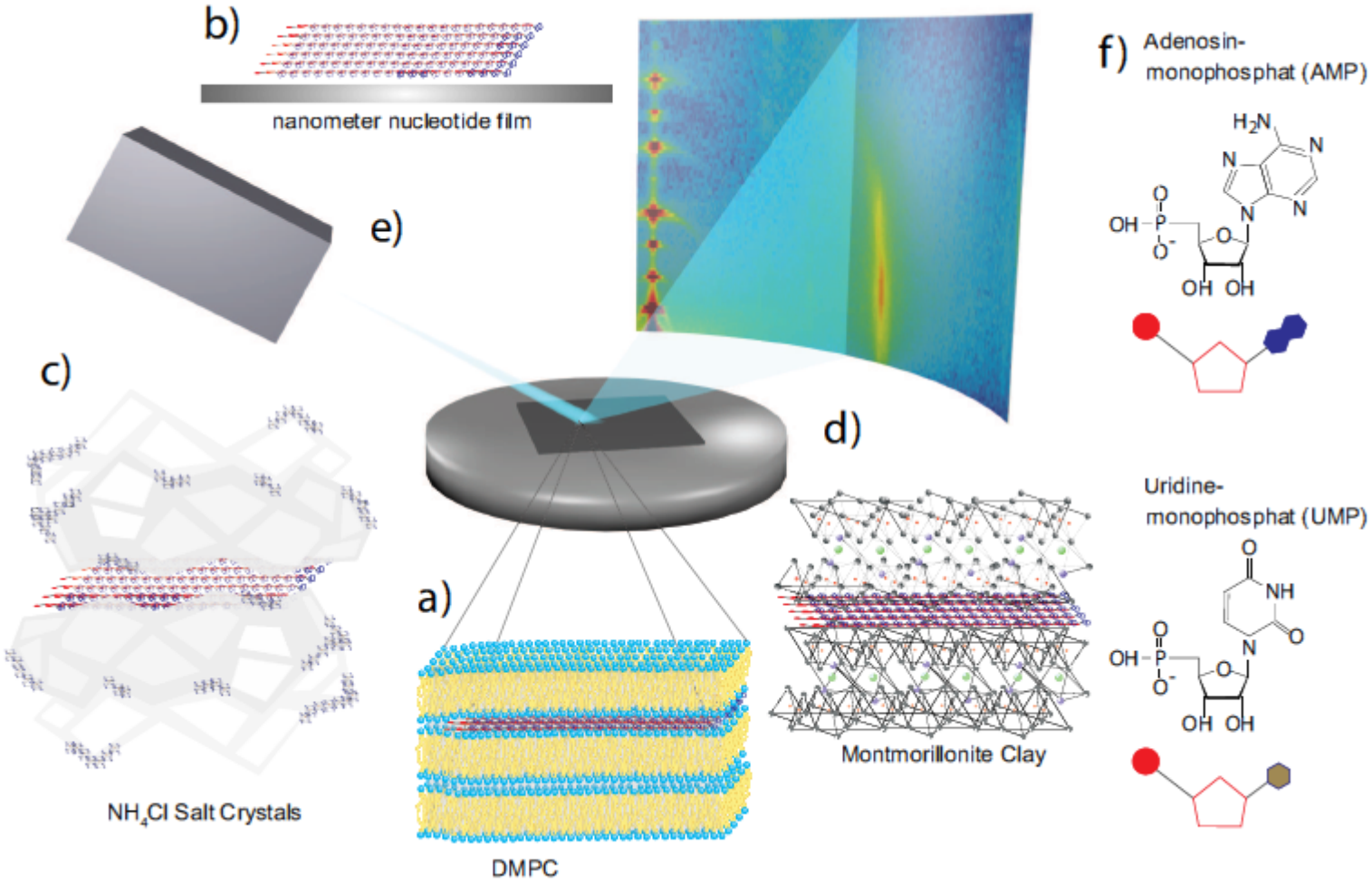
Rajamani ... Deamer et al. (2008) *Orig. Life Evol. Biosph.*

X-ray scattering shows ordered structures of nucleotides between lipid bilayers in lamellar phases



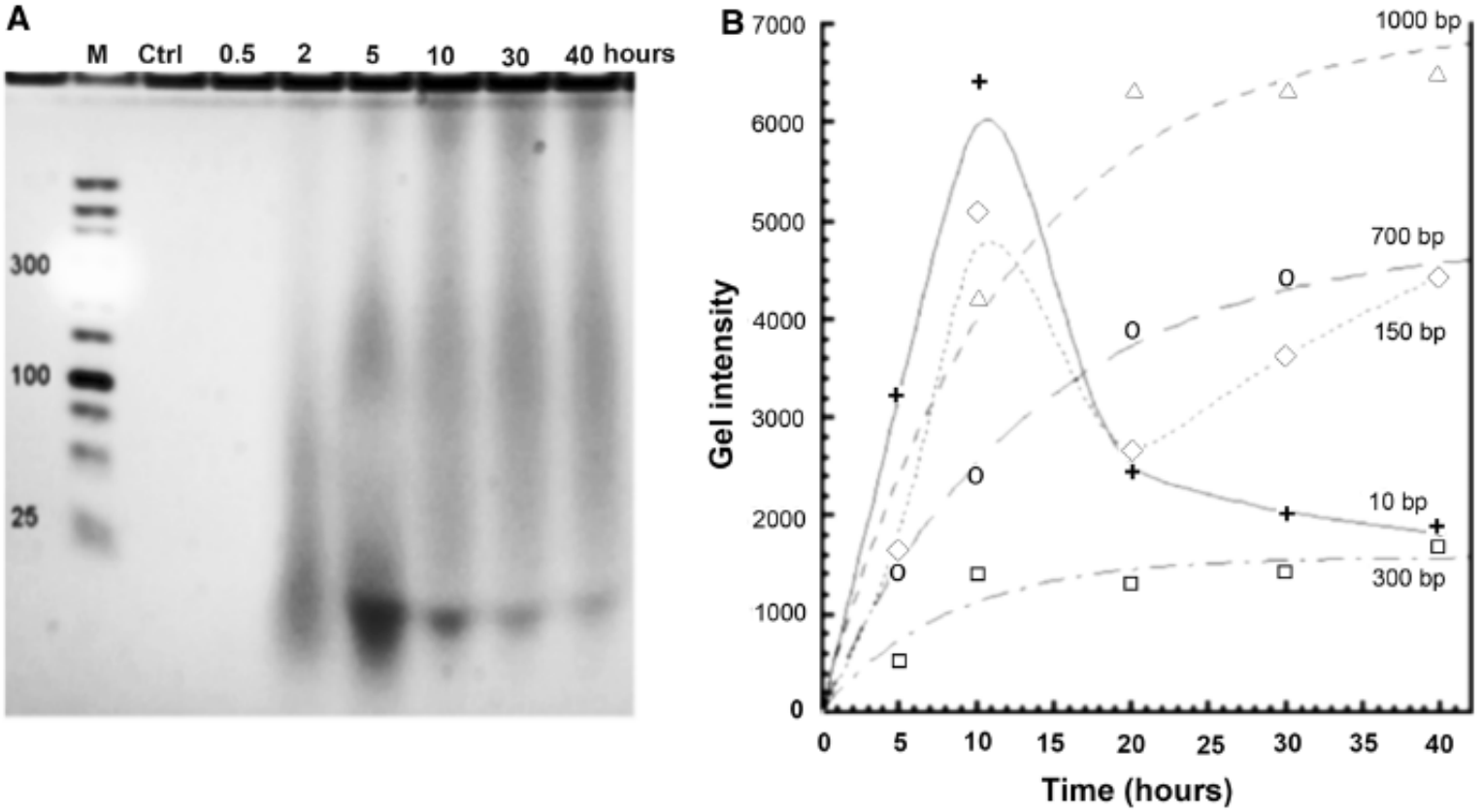
Toppozini, Dies, Deamer, Rheinstadter – PLoS ONE 2013

Drying nucleotides in different environments
Low water content favours polymerization.
Molecules may align in an arrangement favourable to bond formation.

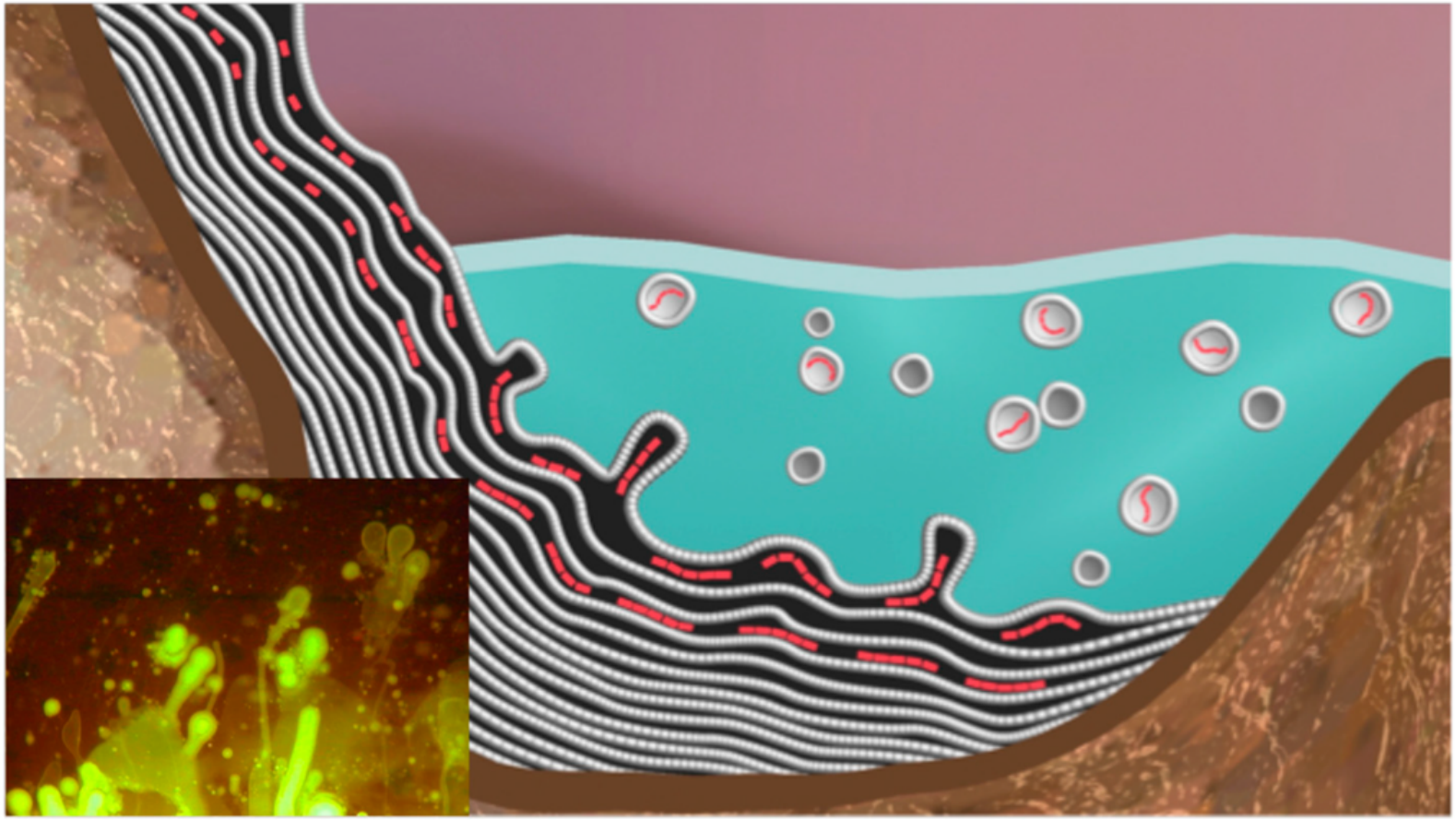


Himbert, Chapman, Deamer, Rheinstadter – 2016 submitted

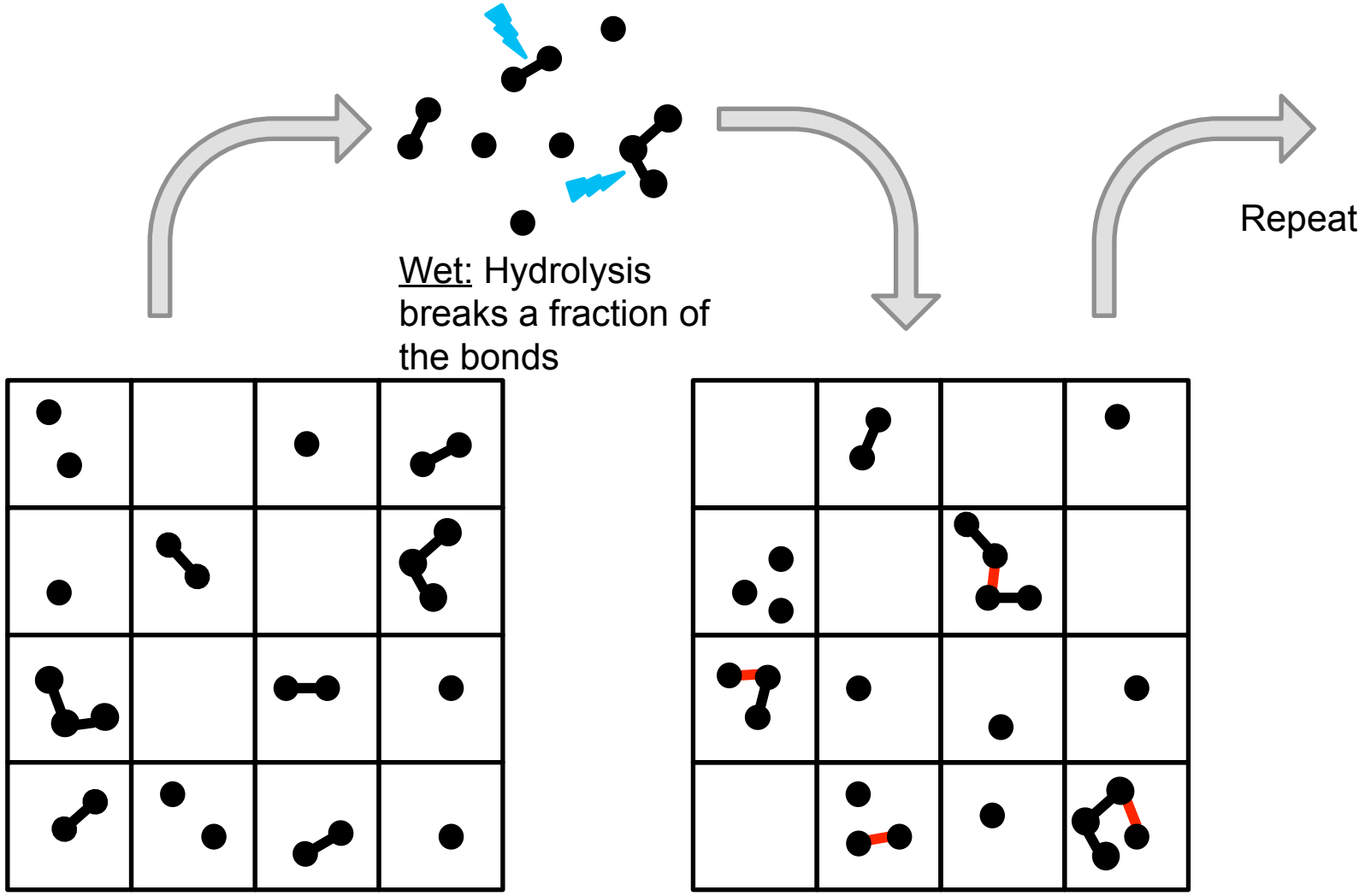
Wet-dry cycling in the presence of salt (e.g. ammonium chloride) generates a broad range of long strands



Da Silva, Maurel, Deamer – J. Mol. Evol. 2015



Bruce Damer and David Deamer envisage wetting and drying cycles leading to the Origin of Life in volcanic hydrothermal pools on land.


Wet: Hydrolysis breaks a fraction of the bonds
Repeat
Dry: Bond formation is favoured if monomers are close
Dry: Repositioning of molecules allows new bond formation and creates longer polymers.

Polymerization model – Reaction kinetics

$$\frac{dC_1}{dt} = -2k^+C_1^2 + 2k^-C_2 - 2k^+C_1\sum_{n=2}^{L-1}C_n + 2k^-\sum_{n=3}^{L}C_n$$

$$\frac{dC_n}{dt} = k^+\sum_{m=1}^{n-1}C_mC_{n-m} - (n-1)k^-C_n + 2k^-\sum_{m=n+1}^{L}C_m - 2k^+C_n\sum_{m=1}^{L-n}C_m$$

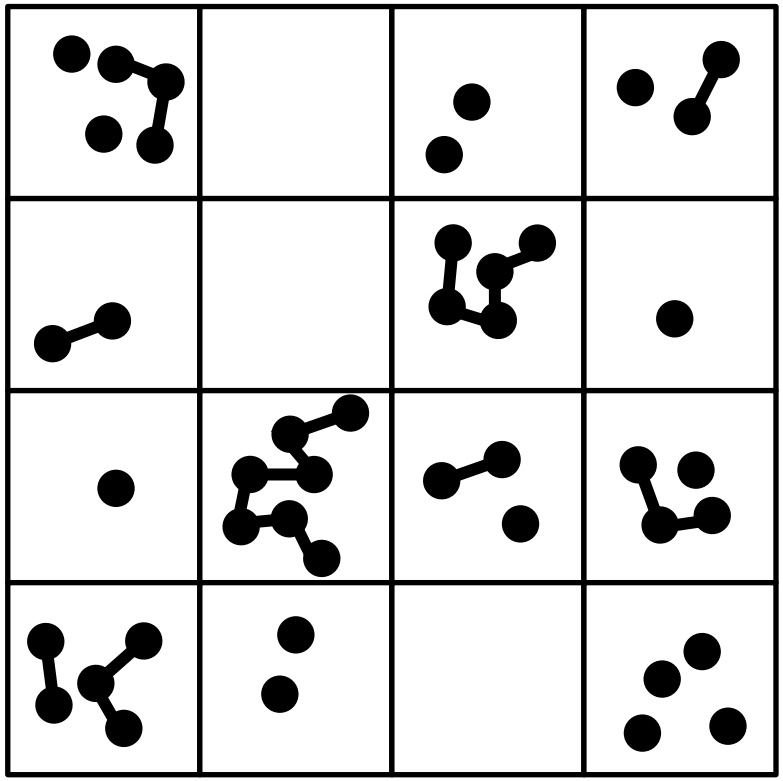
Equilibrium length distribution is exponential. Can be solved exactly.

$$C_n = ax^{n-1}$$

a and *x* are functions of the initial monomer concentration and the equil constant $K = k^+ / k^-$

mean length is $1/(1-x)$

Polymer model



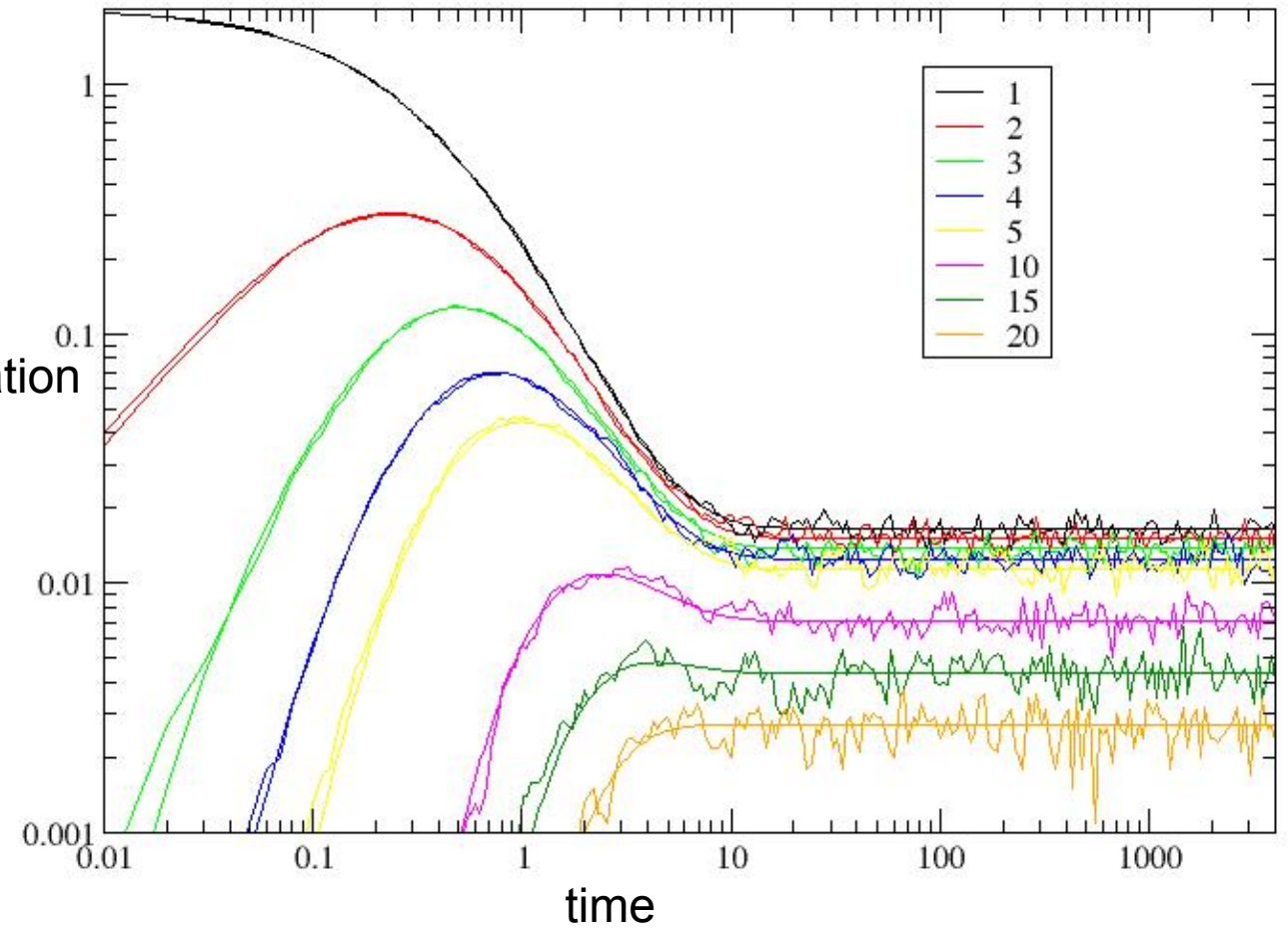
Can also solve this by Monte Carlo simulation

Any two strands on same site can bond at rate k^+

A strand of length m hydrolyzes at rate $(m-1)k^-$

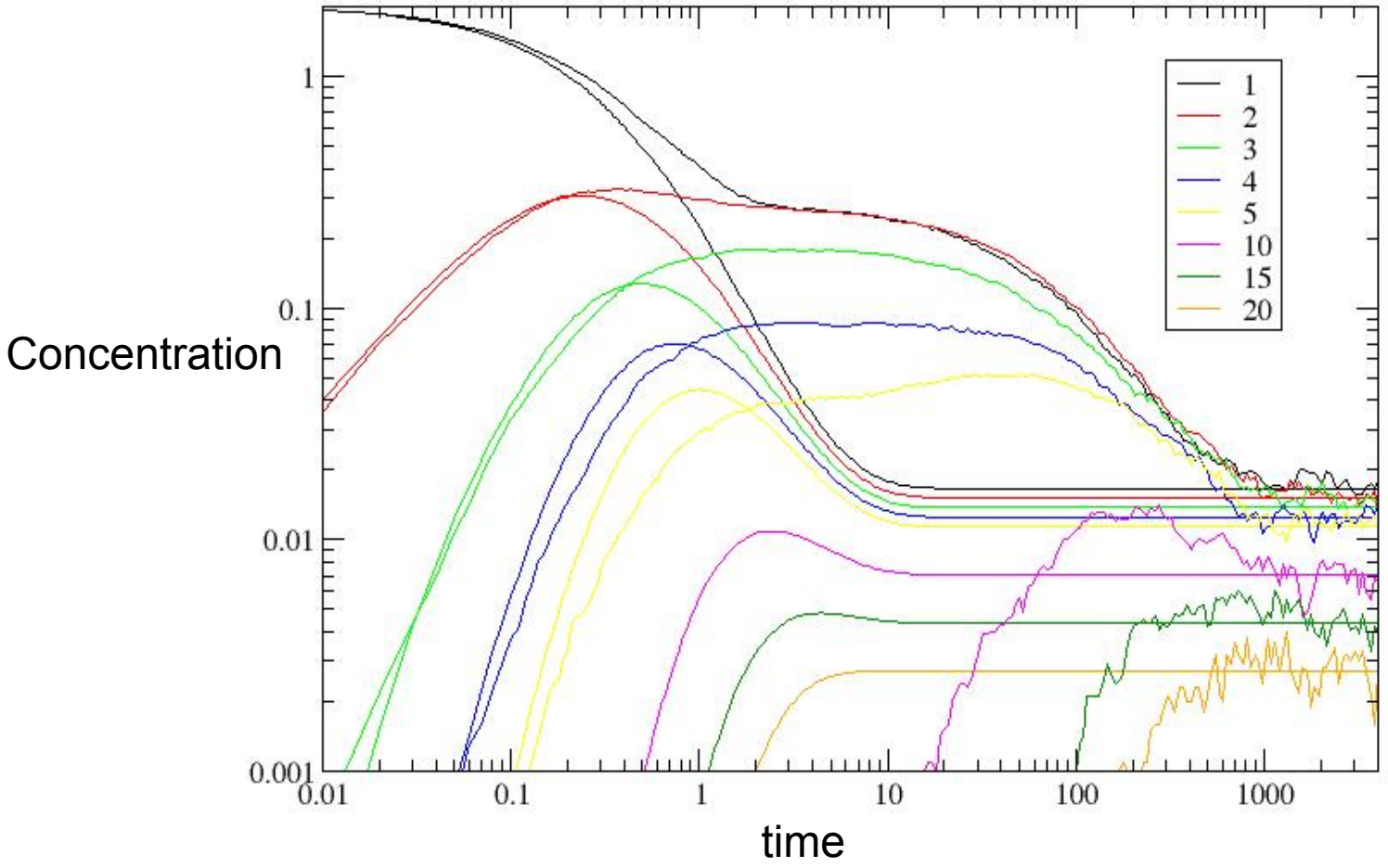
Strands can hop to neighbouring sites at rate h

Two methods are the same when diffusion is fast



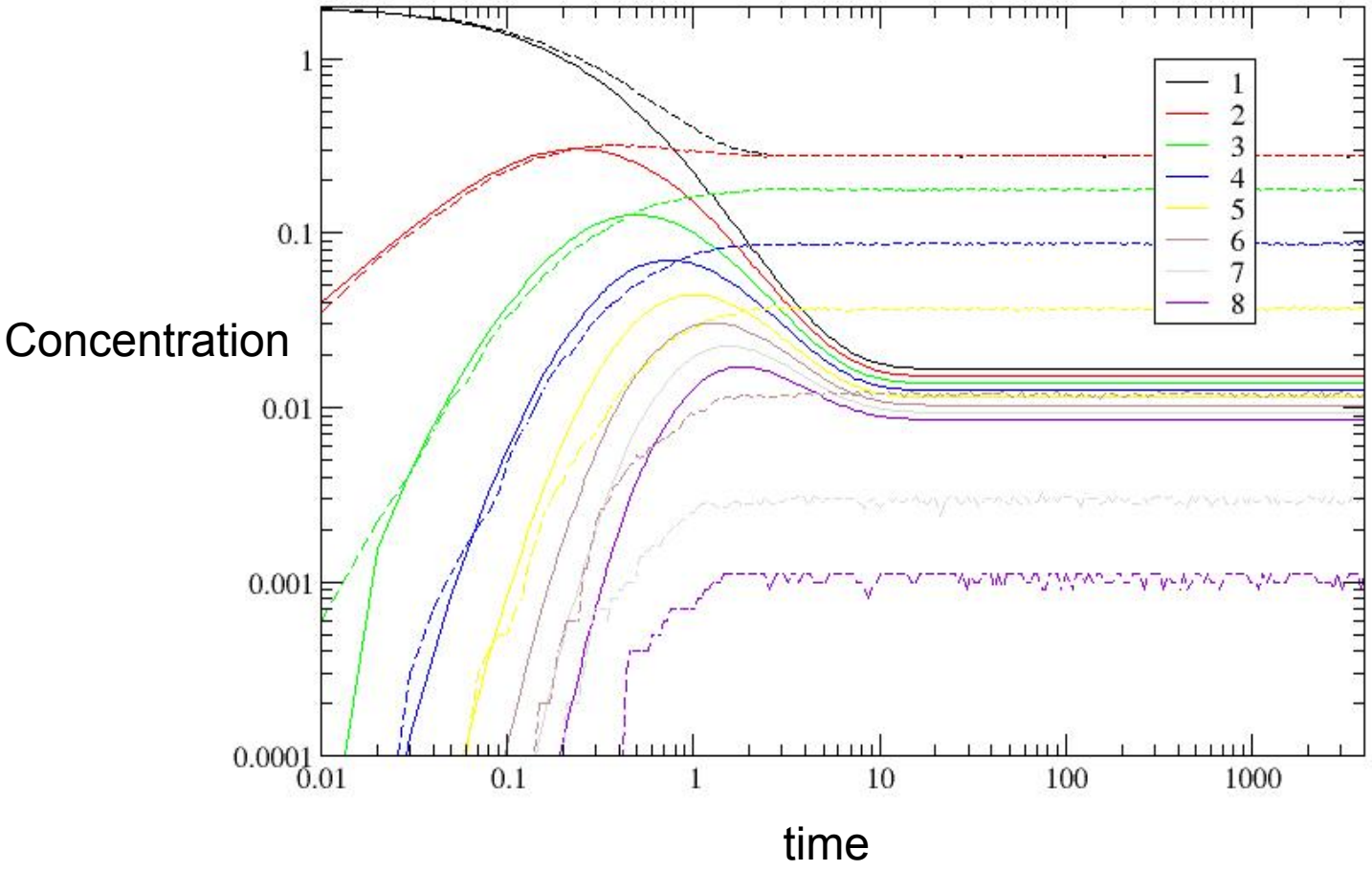
smooth = reaction kinetics
jagged = Monte Carlo with very fast diffusion

When diffusion is slow (*h = 0.01*) the Monte Carlo simulation is much slower to reach equilibrium, but the same equilibrium is reached.



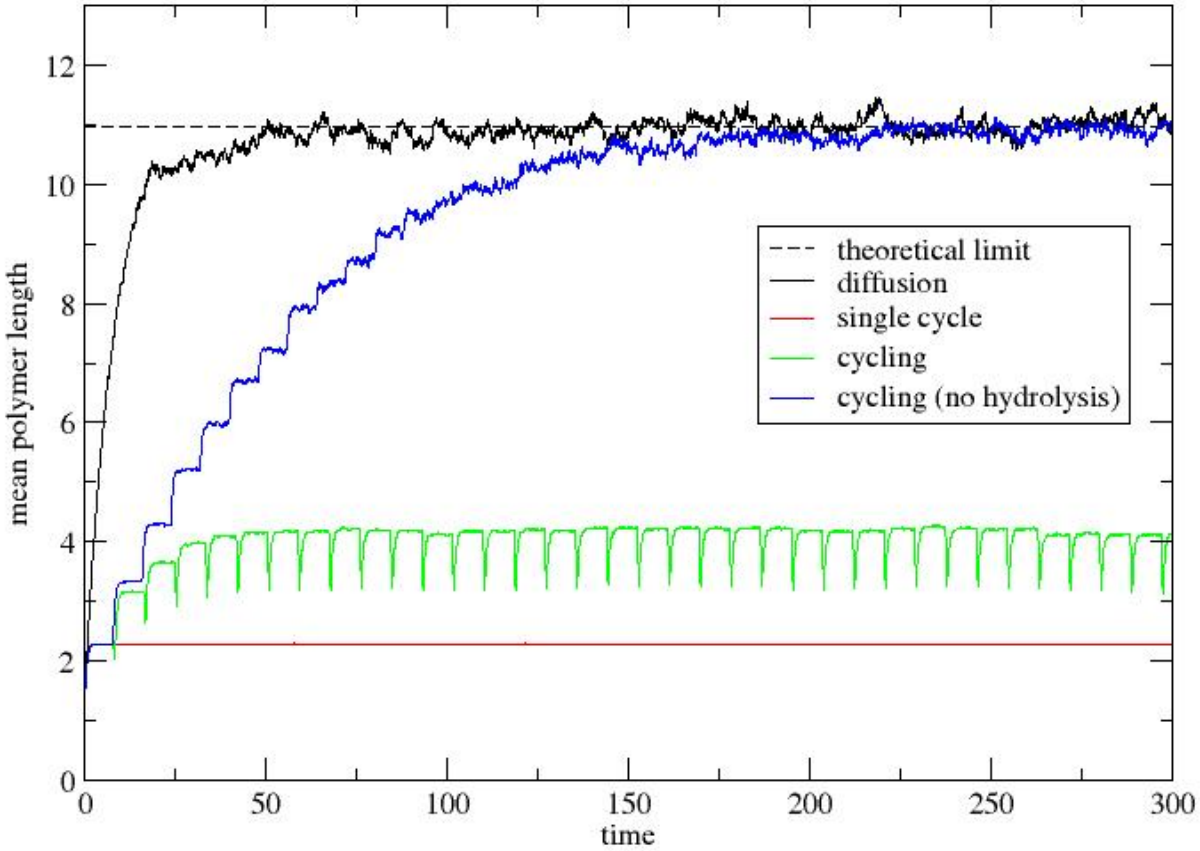
smooth = reaction kinetics
jagged = Monte Carlo with slow diffusion

When diffusion is zero, the Monte Carlo simulation cannot reach the equilibrium distribution. Only monomers initially on the same site can react.



smooth = reaction kinetics
jagged = Monte Carlo with no diffusion

The effects of wet-dry cycling on polymerization



- Dry phase gives long polymers when diffusion is possible
- Single dry phase without diffusion gives short mean length
- Cycling improves on single dry phase
- Cycling can approach the optimum case if $wt_{wet} << 1$

An Evolutionary Theory for the RNA World

Art: Julie Newdoll
"A brush with science"

Emergence from Prebiotic Chemistry

What were the first kind of ribozymes – a Polymerase or something else?

Stochastic – small number of molecules

Cooperation at the molecular level

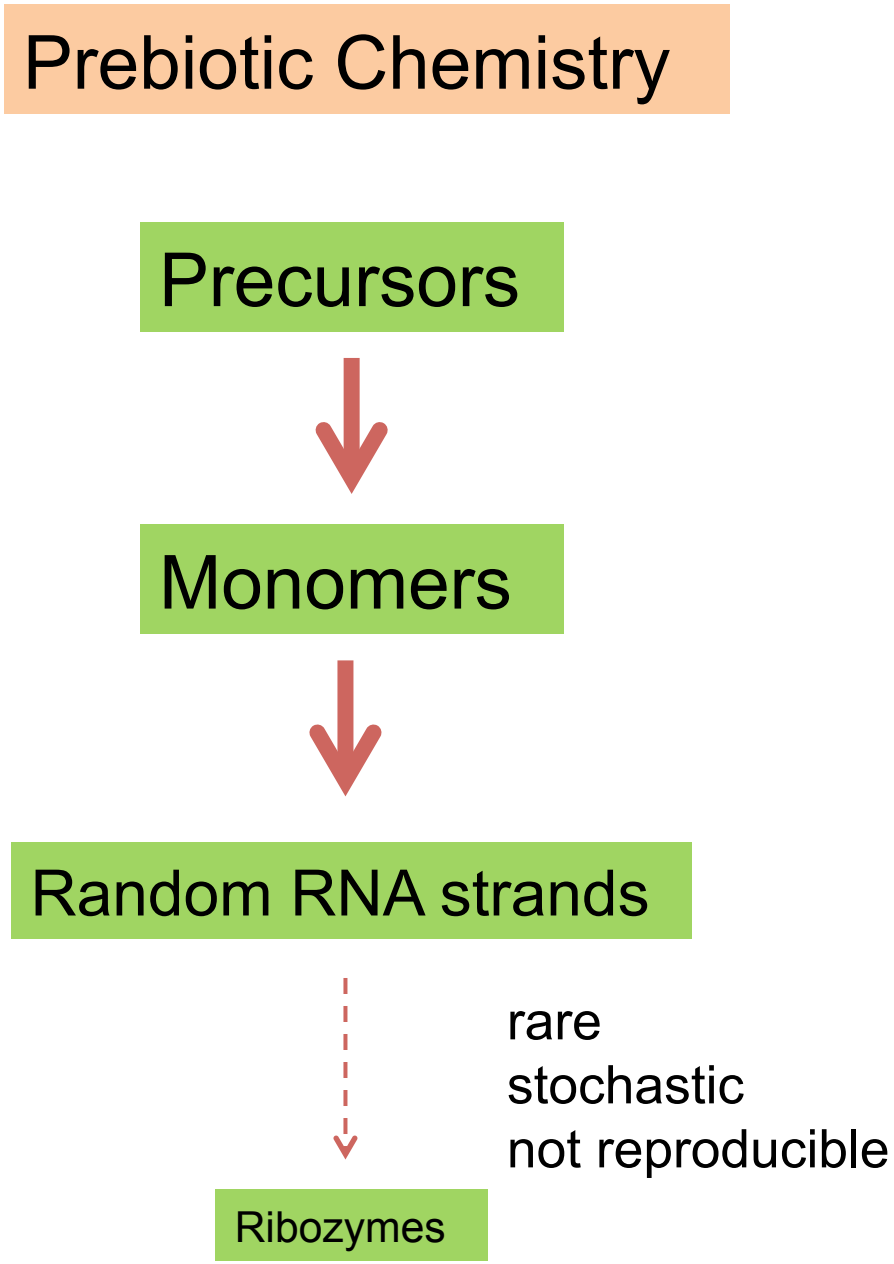
Parasitic templates

Spatial clustering

Was there a complex RNA metabolism?

Evolution of cells

Evolution of chromosomes (linkage)



Chemistry to make monomers and random strands must be reproducible:

- thermodynamics
- reaction kinetics

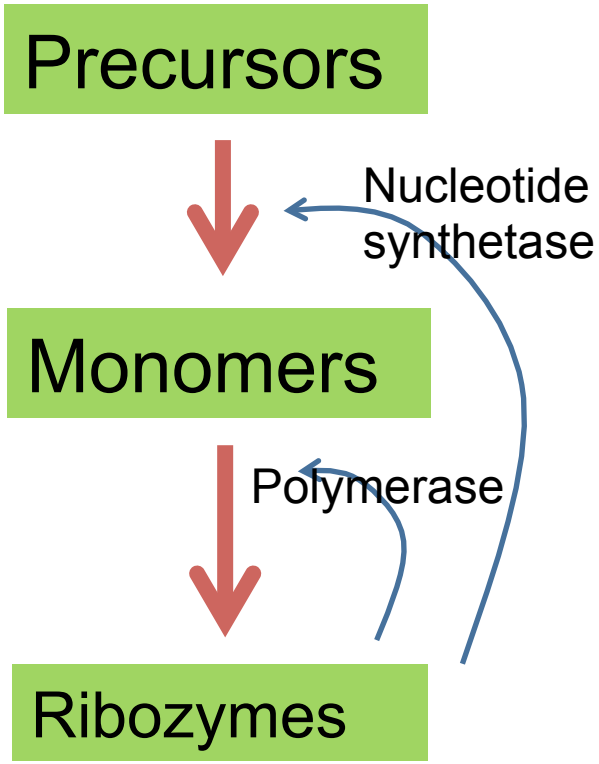
Does not need to be autocatalytic

Does not need to evolve

Not really a metabolism (?)

➔ Dead (non-living)

RNA World



Making ribozymes is now reproducible because of sequence replication.

- heredity
- autocatalysis

➔ Living

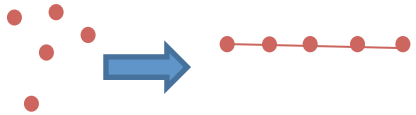
"Life is a self-sustained chemical system capable of evolution"

It counts as life if there are molecules that can only be made reproducibly because of replication/heredity

Three ways of making an RNA strand: s, r and k reactions

s

0 ➔ 1

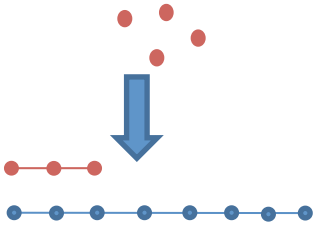


or

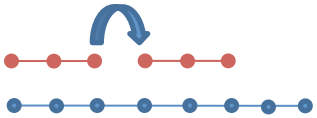
“Spontaneous”
No heredity.
Random Chemistry.

r

1 ➔ 2



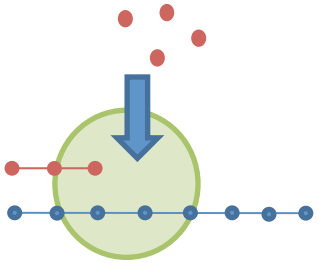
or



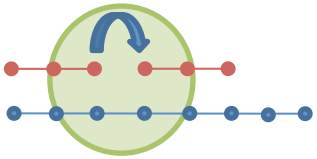
“Replication”
Non-enzymatic template directed synthesis.

k

2 ➔ 3



or



“Catalytic”
Ribozyme catalyzed polymerization or ligation.
Not necessarily processive.

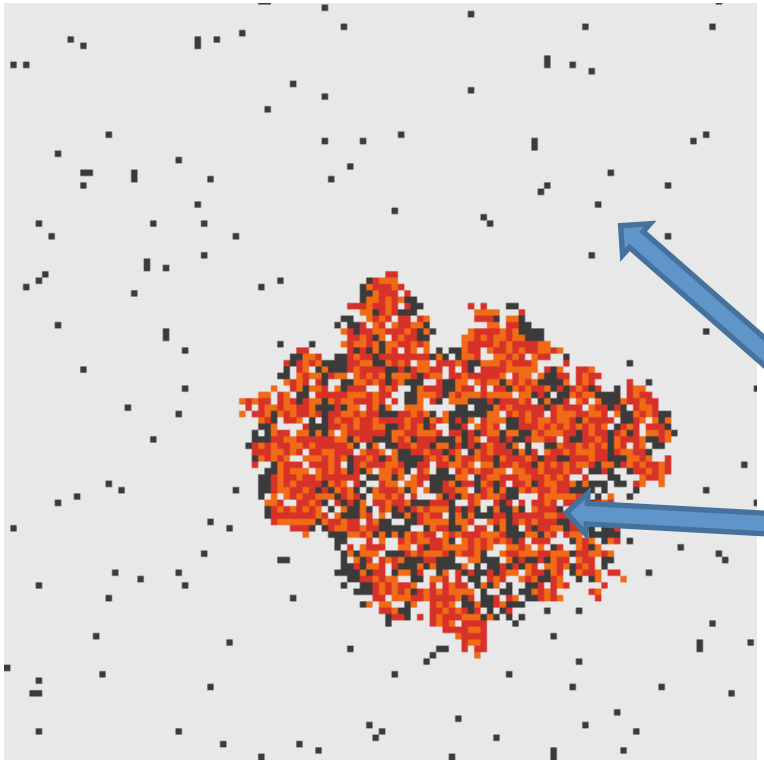
Spatial Model for Origin of Replication in the RNA World

0 = Empty site

1 = Polymerase

2 = Complementary to Polymerase

3 = Random strand or Mutant



The model in brief:

s reaction creates random strands
r reaction copies all strands
k reaction only performed by a polymerase.
Polymerases replicate strands on neighbouring sites.

New strands are placed on neighbouring vacancies.
Strands can hop to neighbouring sites.

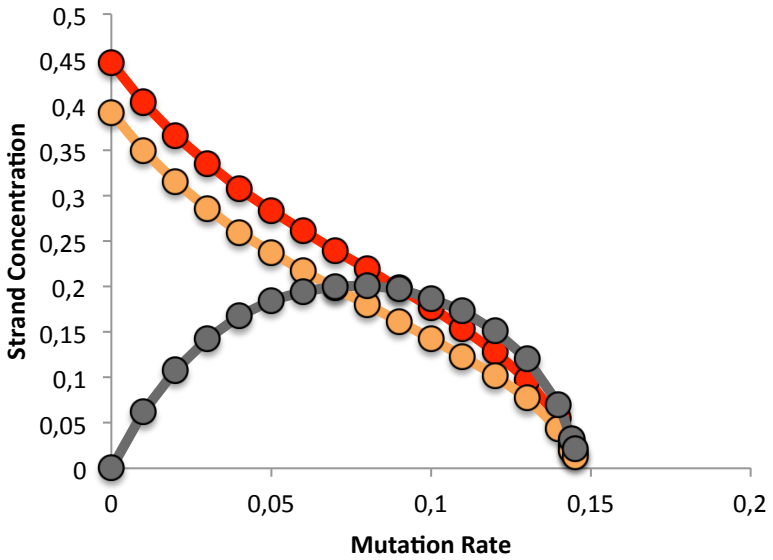
Strands breakdown to monomers again.

Successful replication creates a complimenatry strand.
Mutations create parasites.

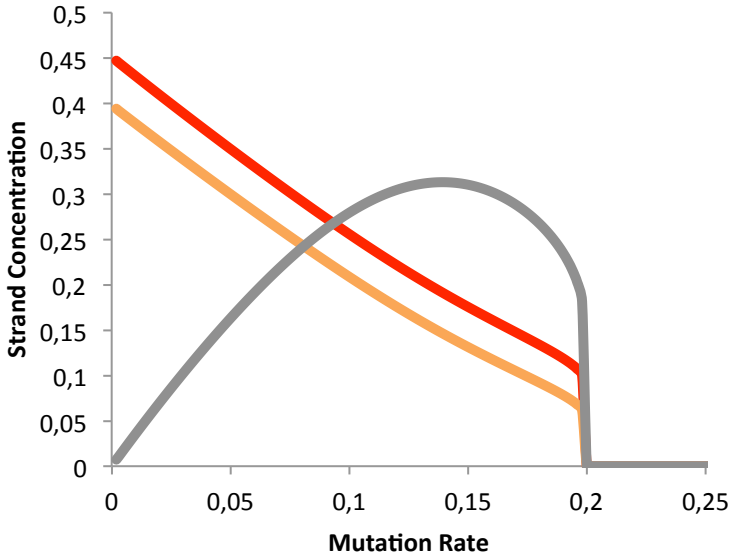
Dead state = low conc of random strands

Living state = high concentration of polymerases, complements, and mutants
The living patch has grown from the introduction of a single polymerase

Concentration as a function of mutation rate M



Simulations



Approximate theory based on paired site concentrations

The Error Threshold is the max mutation rate for which the system survives.

Spatial clustering allows for a fairly high error threshold.

Polymerases first

The importance of cooperation.
Spatial clustering allows survival against parasites

Shay, Huynh & Higgs (2015) J Theor Biol
Higgs & Lehman (2015) Nature Rev Genetics

Nucleotide synthases second

The synthase helps the polymerase by making more monomers.
The polymerase helps the synthase by replicating it.

When can the two coexist?

Can we find conditions where the two can survive together but the polymerase could not survive on its own?

Current work – Ye Eun Kim, Andrew Tupper

When can a Nucleotide Synthetase Cooperate with a Polymerase?

0 = Empty site

1 = Polymerase

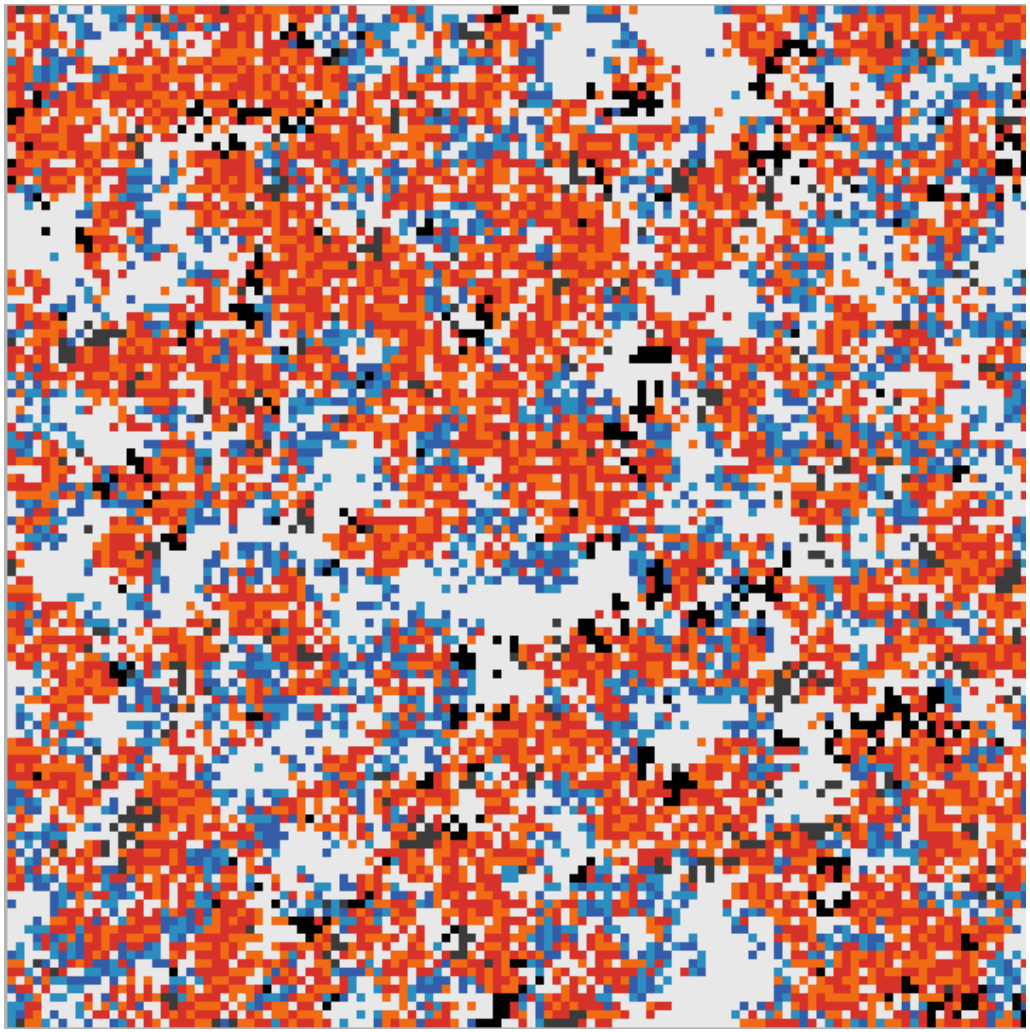
2 = Complementary strand to Polymerase

3 = Mutant polymerase (non-functional)

4 = Synthetase

5 = Complementary strand to Synthetase

6 = Mutant synthetase (non-functional)



Uni01

Polymerase + Independent short strands

0 = Empty site

1 = Polymerase (L_1 = 100)

2 = Complementary to Polymerase (L_1 = 100)

4 = Short strand (L_2 = 50)

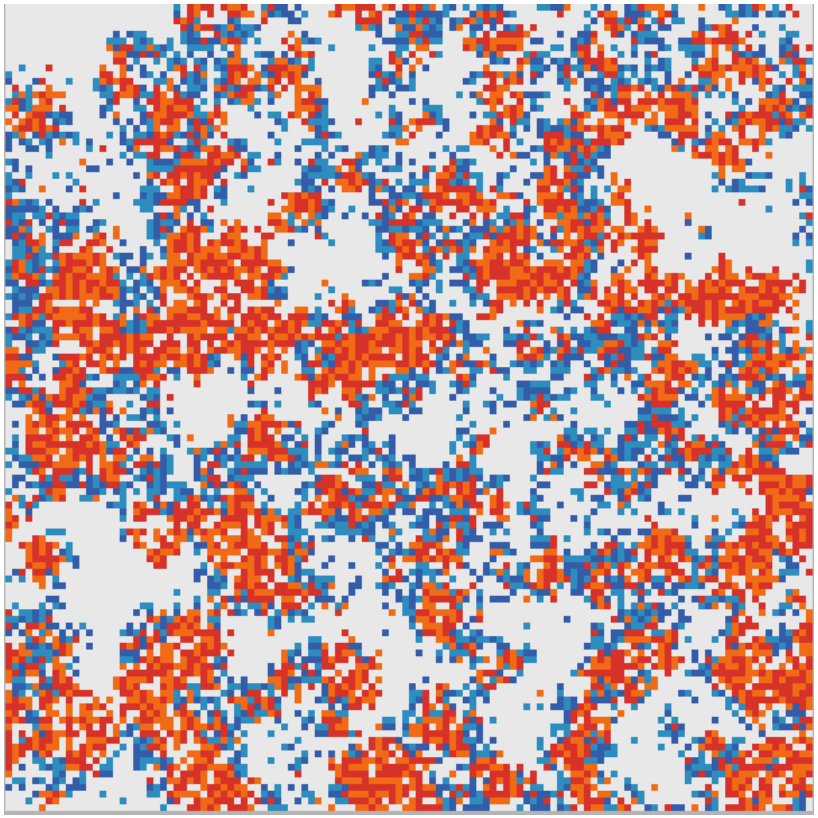
5 = Complementary short strand (L_2 = 50)

polymerization rate constant

monomer concentration

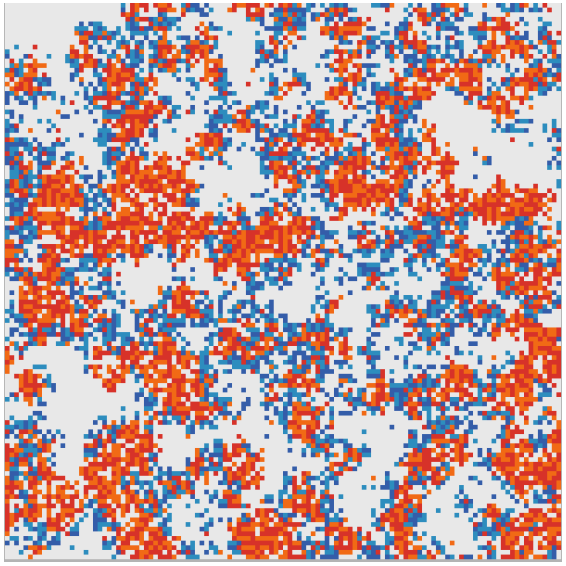
replication rate → $k = \frac{v_{pol} A}{L}$

template length

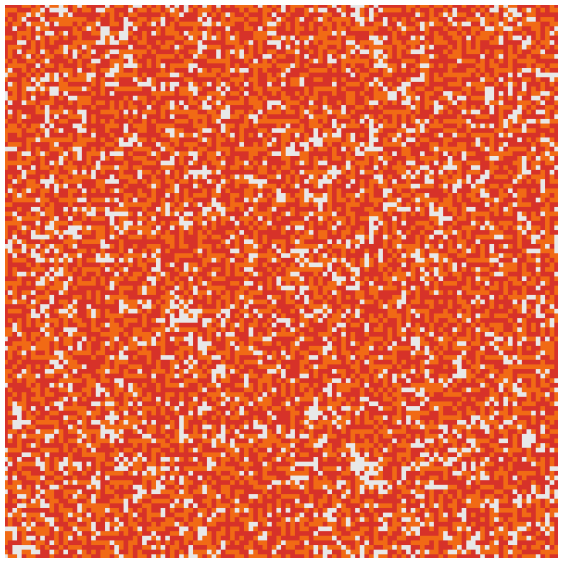


The short strands are non-functional parasites that multiply twice as fast as the polymerase.
The polymerase still survives because of spatial clustering.

Polymerase +
Independent short strands
$L_2 = 50$

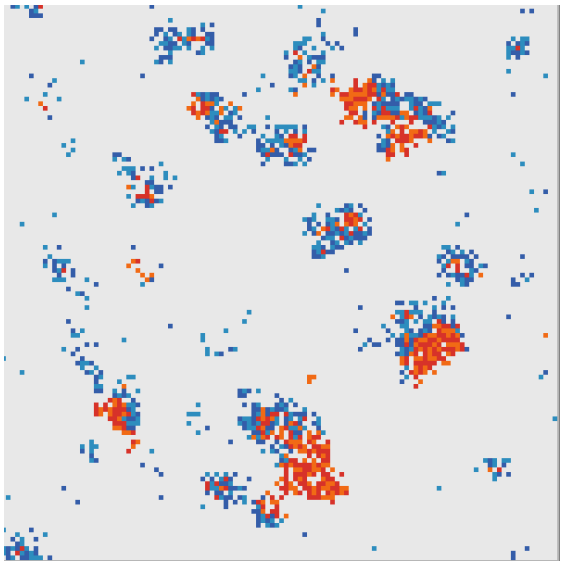


Increase L_2 to 90



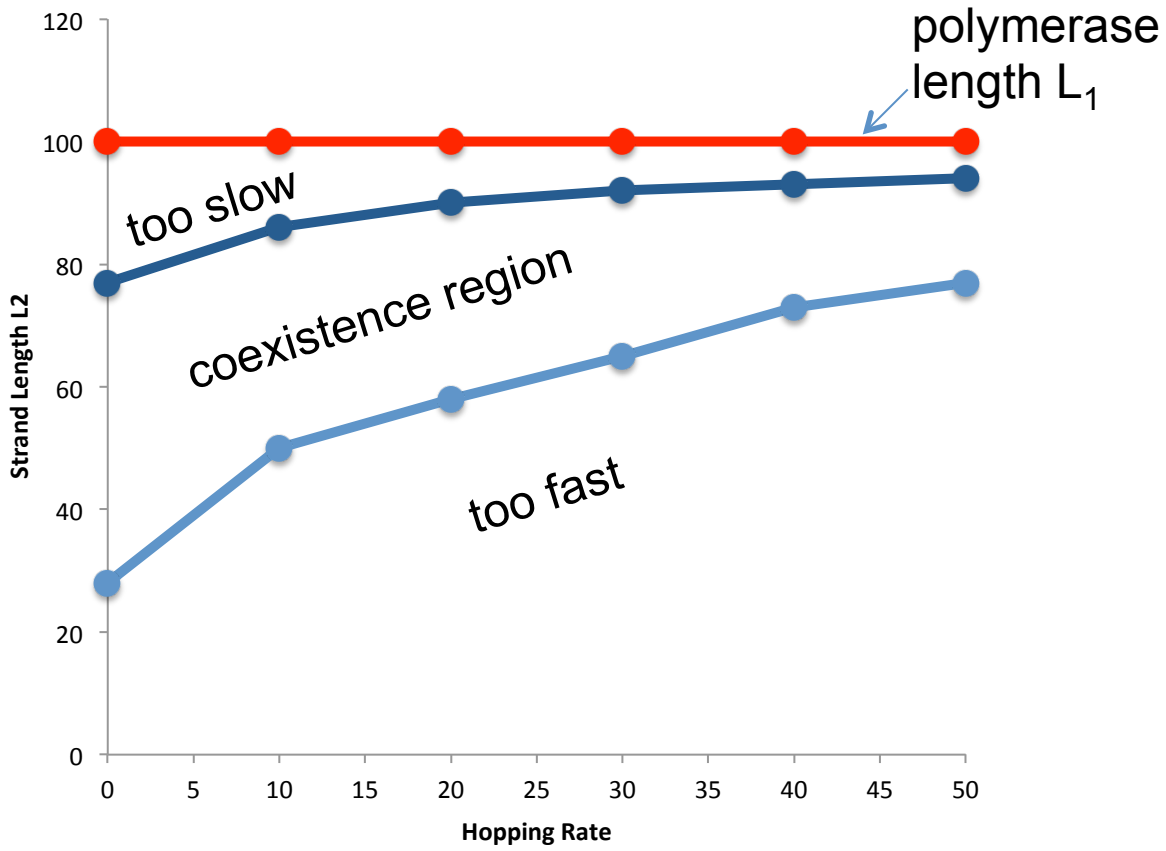
Blue strands multiply too slowly.
Blue strands die and
Polymerases survive alone

Decrease L_2 to 30



Blue strands multiply too fast.
Blue strands overrun the
Polymerases. Everything dies.

Coexistence region between polymerases and independent short strands



In the well-mixed case, coexistence is only possible if the two lengths are exactly equal.

Monomer synthesis and diffusion

Simple assumption that monomers diffuse rapidly everywhere does not work because synthetases are overrun by their own mutations.

Need a finite rate of monomer diffusion

$$\frac{\partial A}{\partial t} = a + bX_{syn} - A + D\nabla^2 A$$ - Polymerization + Breakdown

Polymerization - When a new strand is made, *L* monomers are used up from the site of the template.

Breakdown - When a strand is destroyed, *L* monomers are released to the site where the strand was.

Local concentration A_i on each site *i*

S_i = 1 if there is a synthetase on site *i*, or 0 otherwise

$$\delta A_i = \delta t(a + bS_i - A_i - D(A_i - \frac{1}{8}\sum_j A_j))$$ - Polymerization + Breakdown

Polymerase + Synthetase + Local monomer diffusion

0 = Empty site

1 = Polymerase (L_1 = 100)

2 = Complementary to Polymerase (L_1 = 100)

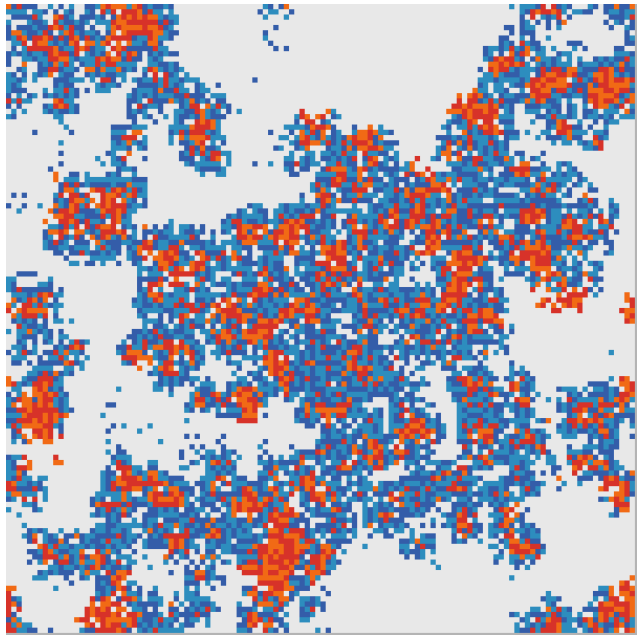
4 = Synthetase (L_2 = 60)

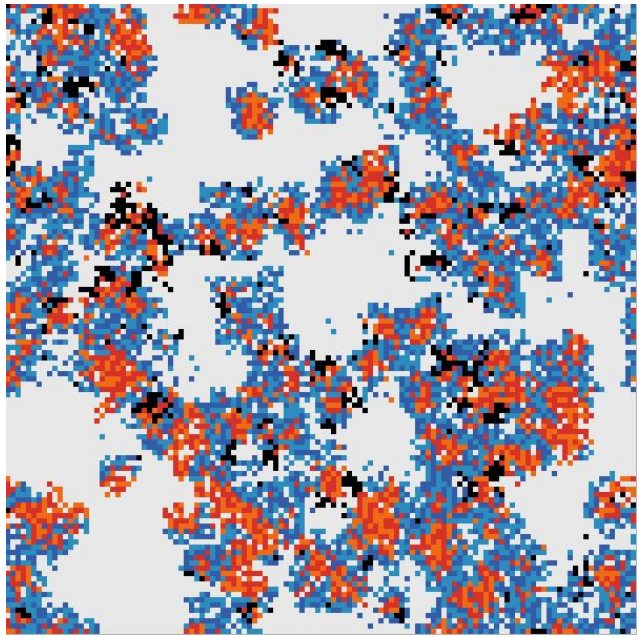
5 = Complementary to Synthase (L_2 = 60)

6 = Mutant Synthetase (L_2 = 60)

Monomer concentration controlled by diffusion

a = 0
b = 5000
D = 30
v_{pol} = 5





Two ribozymes survive in cooperation even when a = 0

Add mutation M_{syn} = 0.01

System remains stable

Spatial clustering is sufficient to allow cooperation of two unlinked ribozymes without membrane compartments

Synthetases can coexist with Polymerases....
... but only if conditions are right.

- Strand diffusion must be slow enough to allow joint clusters to form
- Synthases must replicate sufficiently fast, otherwise they die out
- Synthases must not replicate too fast, otherwise they overrun the polymerases
- Monomer diffusion must be fast enough to benefit the polymerases
- Monomer diffusion must not be too fast, otherwise the synthases are overrun by their own mutations

Is a multi-component ribozyme system possible?

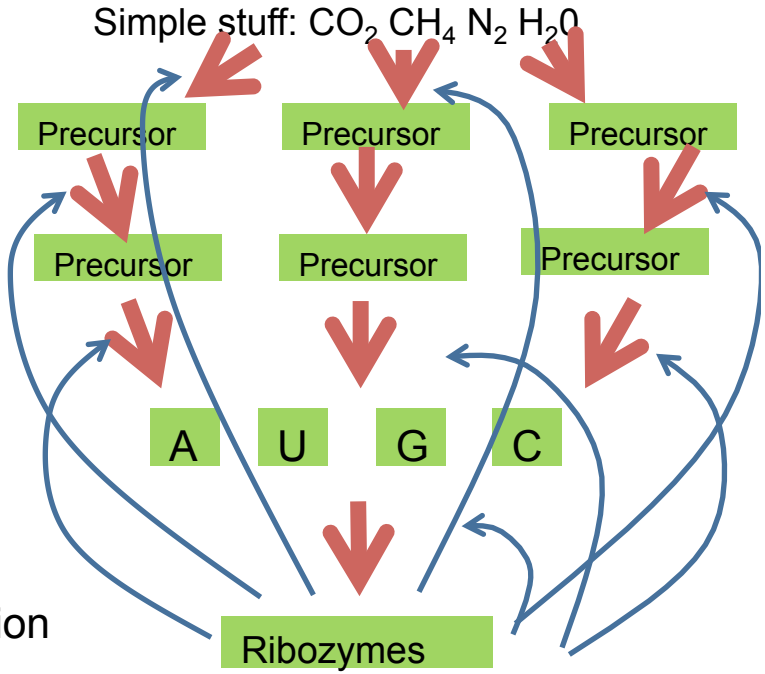
Multiple independent replicators in a surface based model will probably be difficult

From surface-bound replicators to cells

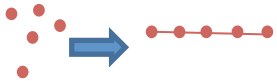
+ group selection
+ diffusion products kept within cell
- need to make lipids
- need to transport things across membranes
- need to control cell division

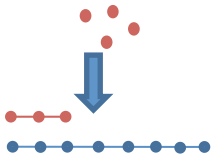
From independent genes to chromosomes

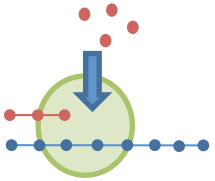
+ linked genes replicate together
- error threshold for long strands
- distinguish between replication and transcription



The importance of r

s  or  Non-living

r  or  Living or Non-living ?

k  or  Living

The r reaction could support life if it was fast and accurate

- It has heredity
- Possibility of template evolution

If r is too slow, it cannot support life even if replication is 100% accurate.

If r is fast - maybe you don't need a polymerase at all ....
Another ribozyme could be first