The big quest: what is life?

The smaller quest: how do you make it?

The practical quest: how do you make money off of it?

In defining the origins of life, there is one major problem: What is life?

In 1944, the physicist Erwin Shrodinger defined living matter as that which "avoids the decay into equilibrium."

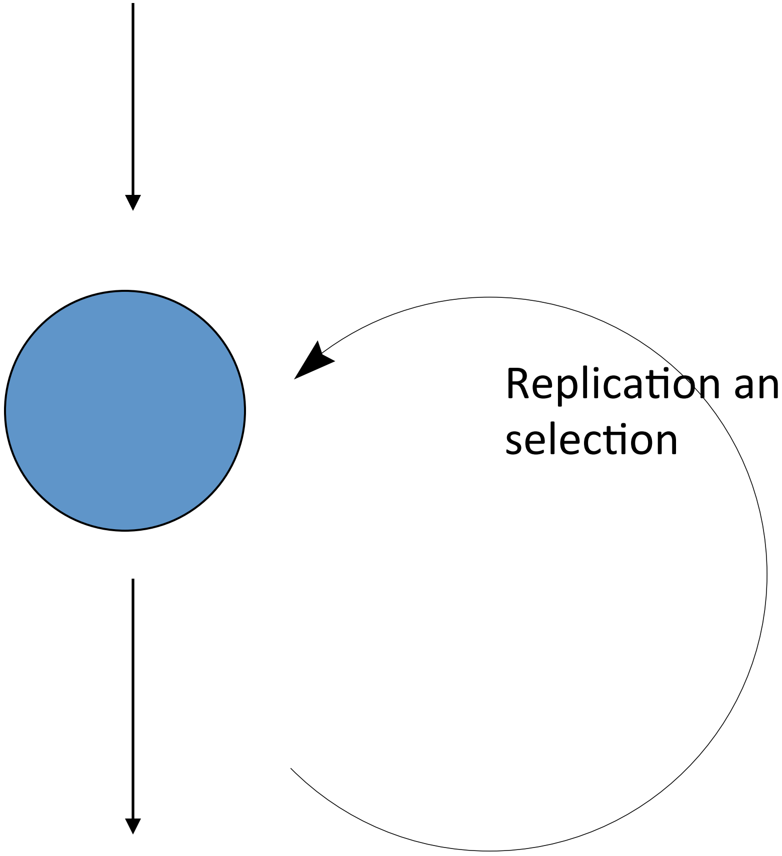
Professor, Departments of Chemistry and Molecular Biology, The Scripps Research Institute

Investigator, The Skaggs Institute for Chemical Biology

Another definition of life created by Gerald Joyce of the Scripps Research Institute doesn't mention disequilibrium per se. This definition says that life is "a self-sustaining system capable of Darwinian evolution."

This is a circular definition. ‘Life’ is a term similar to ‘pornography:’ we apparently know it when we see it, but we can’t actually define it with any precision. Because of this, ‘life’ is a term more suited to poets than scientists.

Here’s what fascinates us:



Simple components

Replication and selection

Complex function

- Nucleic acid imperialism

- Doppelgangers

- Looking backwards

We can also go the other way.
The Miller-Urey experiment
Is an ‘icon of evolution.’ While Miller originally used a reducing atmosphere and electricity to create amino acids, some work suggests that the Earth originally had a neutral atmosphere.

But … there were plenty of sites where a reducing environment was present (deep sea vents, volcanoes) and recent research suggests that a neutral atmosphere was not an impediment to amino acid formation.

But … no one knows whether and how much amino acid formation may have been necessary for abiogenesis.

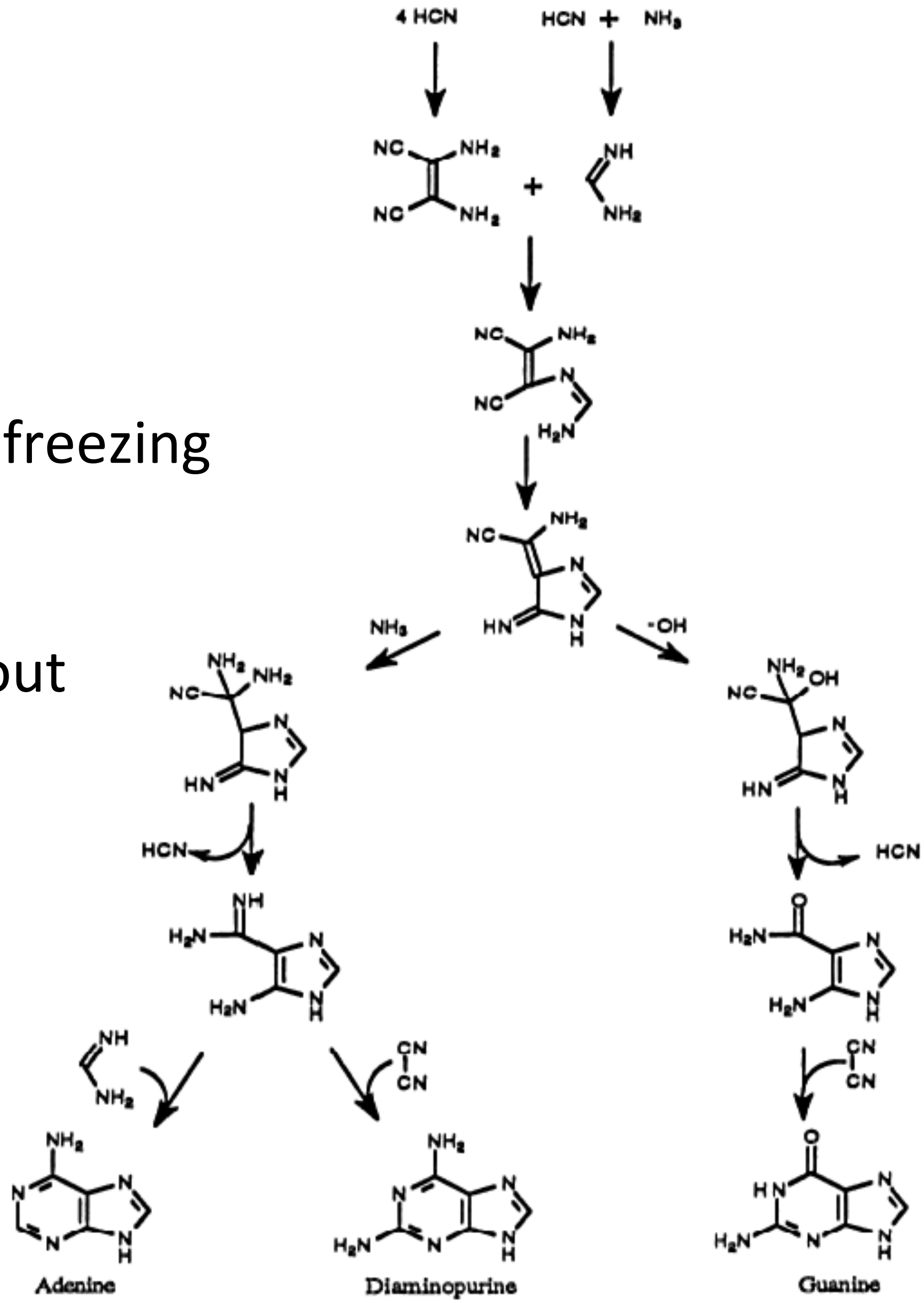
Purines from Poison!

-HCN polymerization to yield adenosine (Oró, 1960)

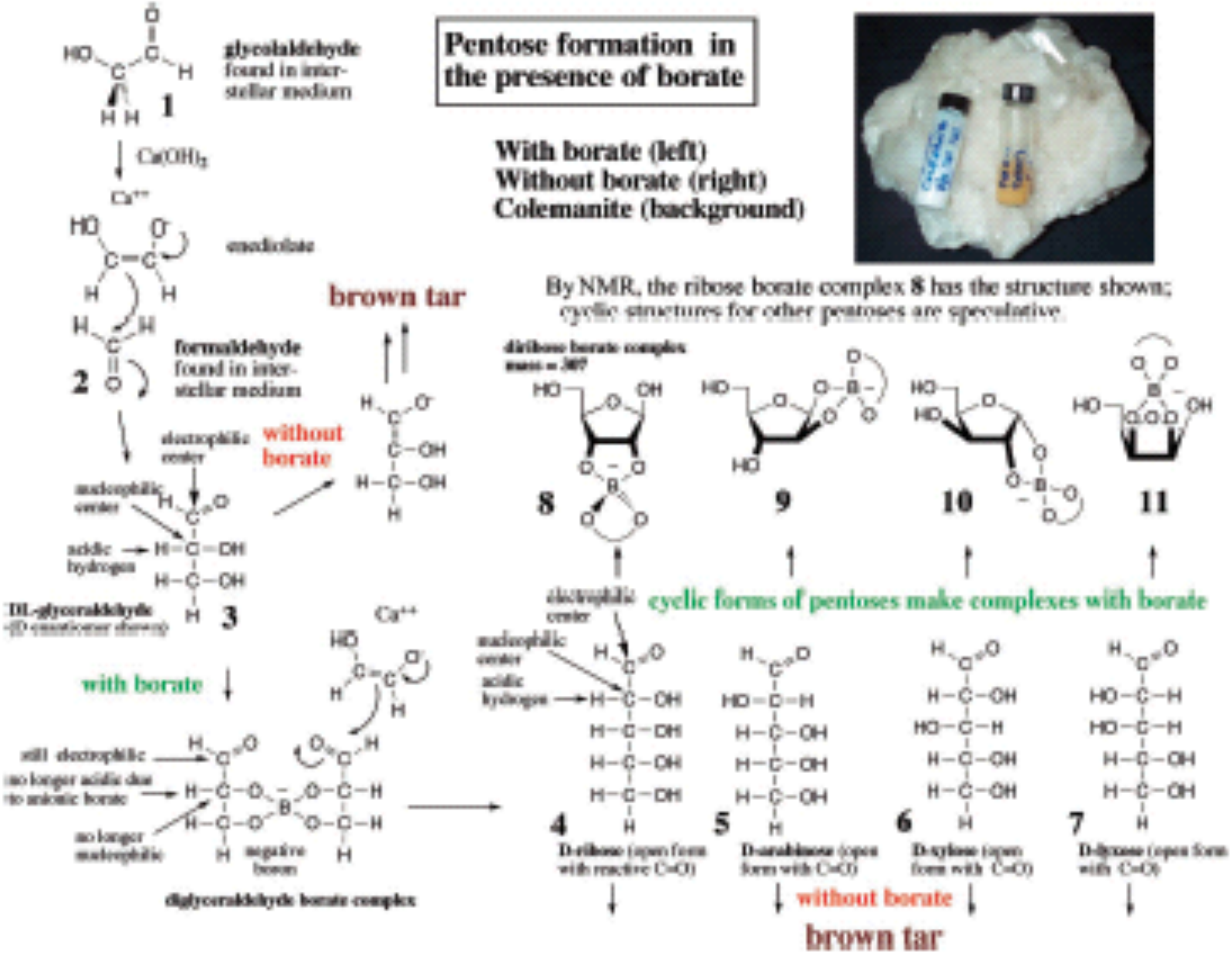
-HCN concentration by eutectic freezing (Schwartz, Joosten, Voet, 1982)

At first it was difficult to figure out where guanine came from …

-synthesis of guanine (Levy, Miller, Oró,1999)

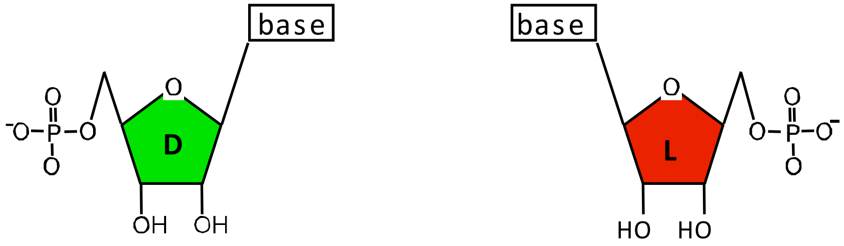


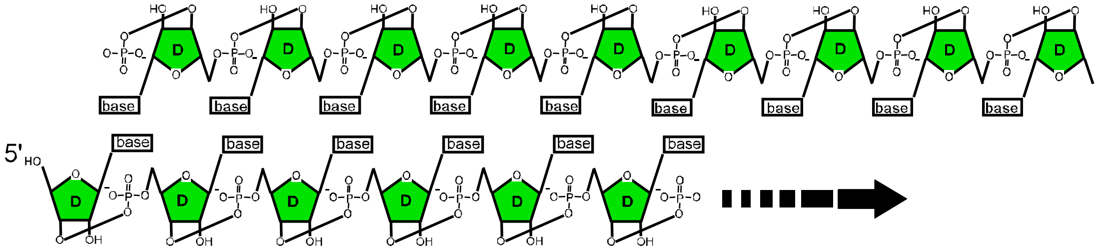
Another case in point: it was thought to be difficult to Prebiotically synthesize sugars for RNA; the formose reaction typically yields 'tar.'

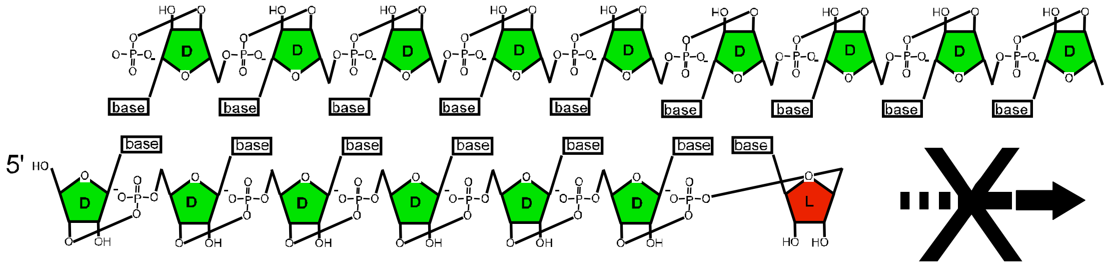


Until Steve Benner recently found that the addition of borate simplified the mixture considerably

So, how does a replicable material replicate?

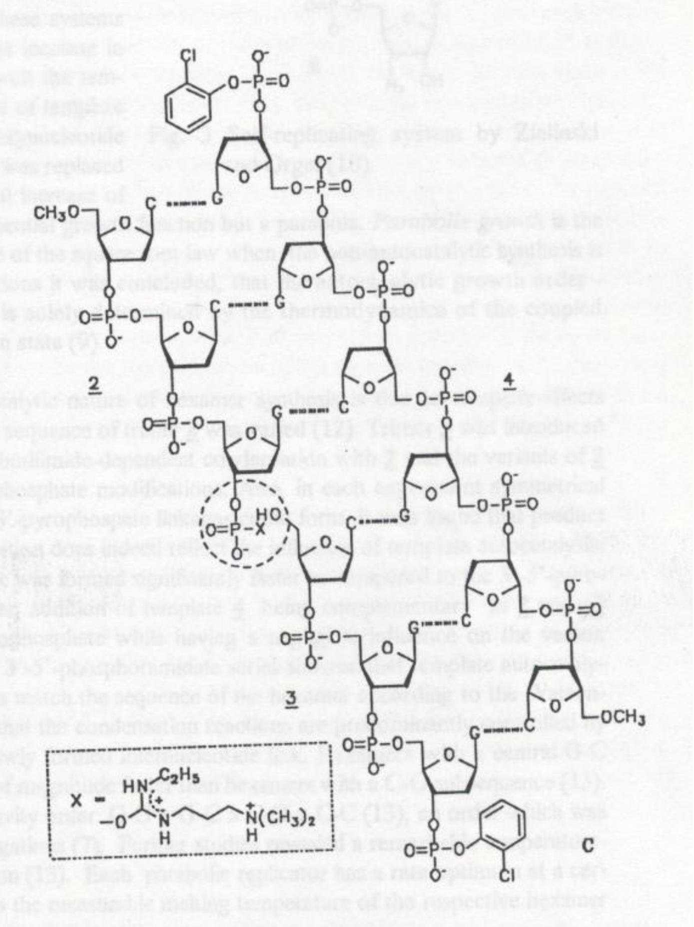






‘Enantiomeric poisoning’ can completely stop monomer replication

Oligonucleotide ligation rather than polymerization can overcome this barrier

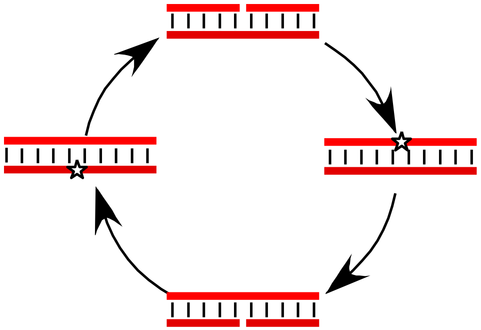


5’

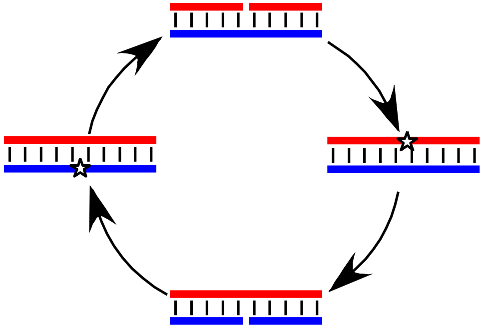
GCC GGC

CGG-CCG

5’

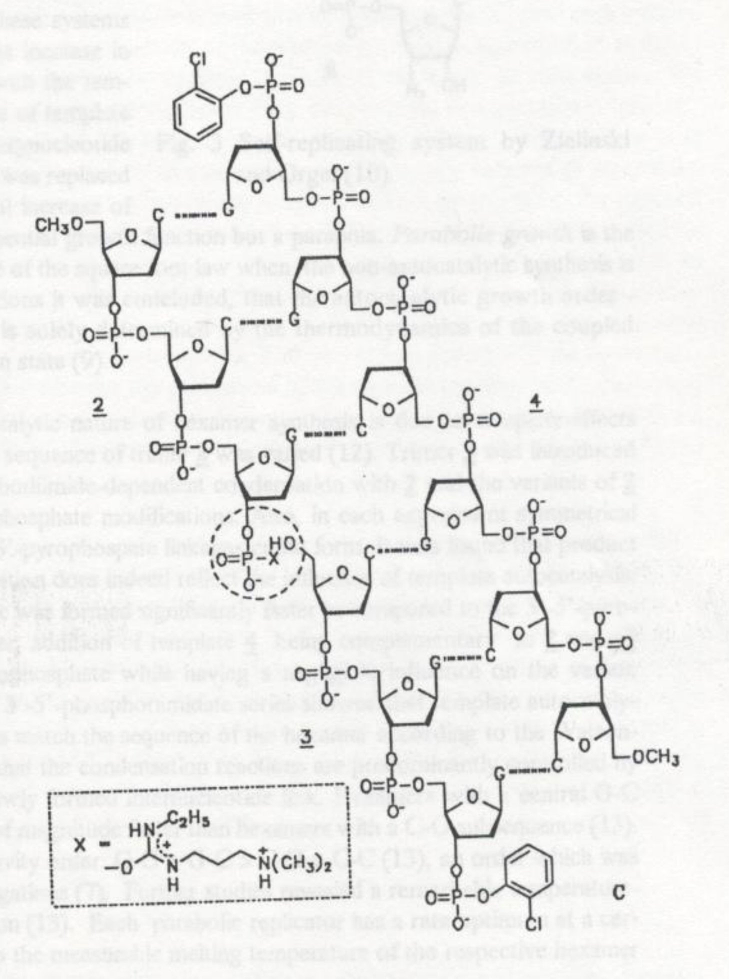


A System
(palindrome)



A/B System

Parabolic replication: the ‘survival of everyone’



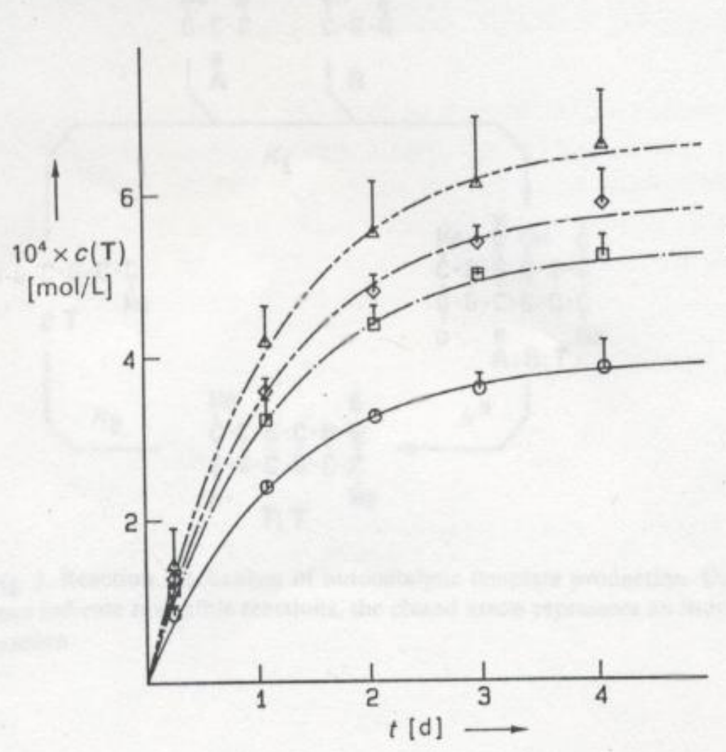
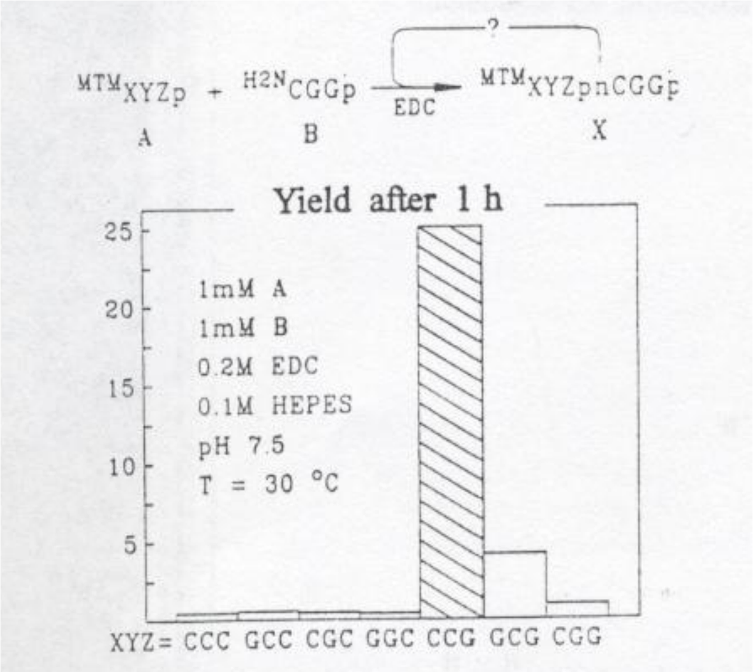
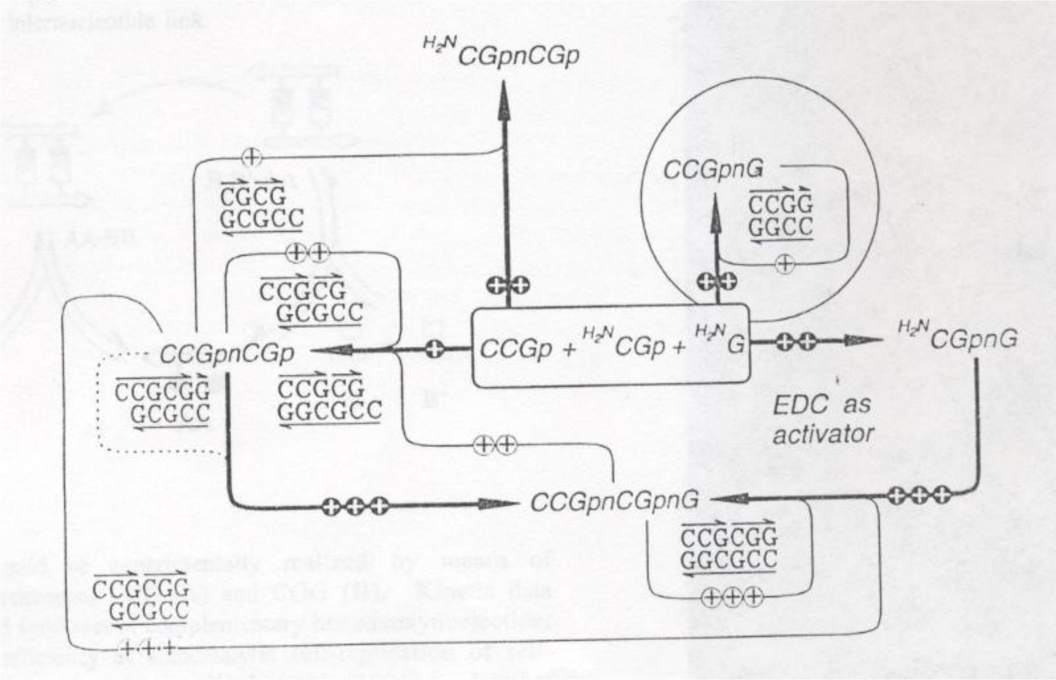


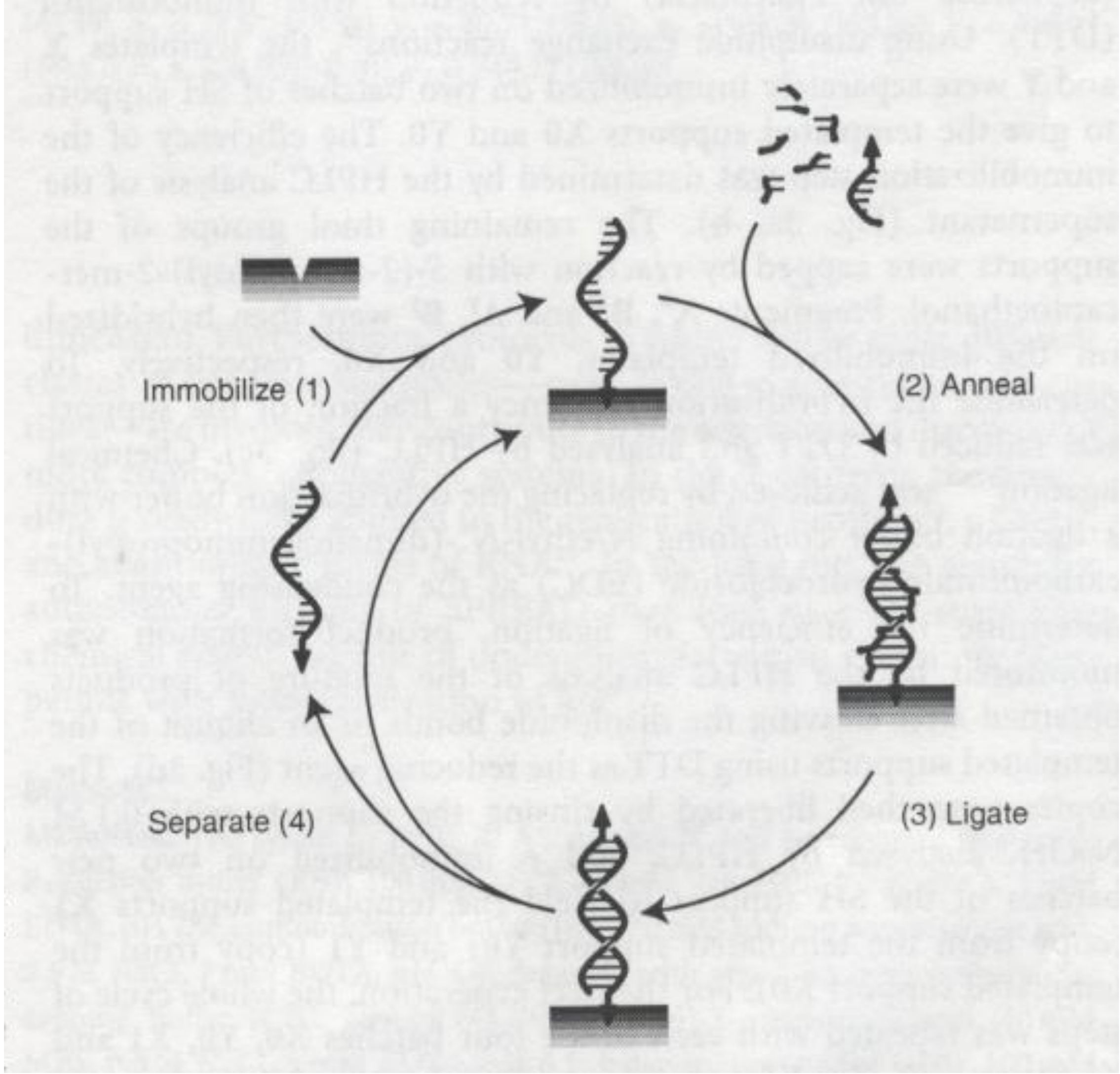
Figure from von Kiedrowski (1986), Angew Chem Intl Ed Engl 25:934

While there is sequence specificity, there is also a great deal of cross-catalysis



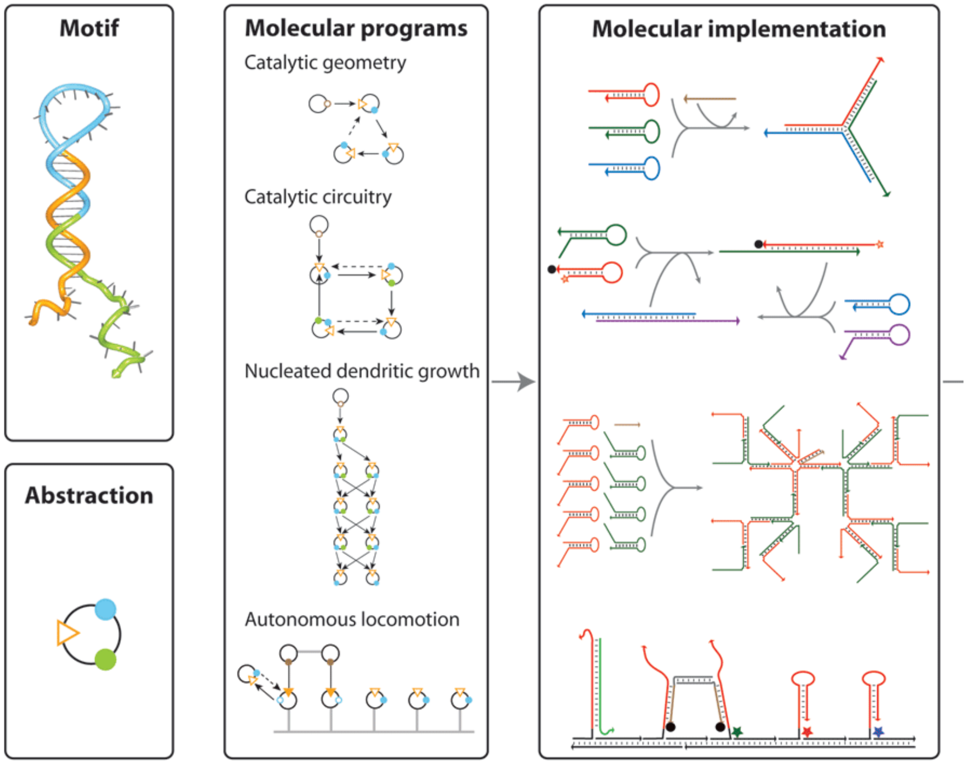


Exponential amplification requires strand separation

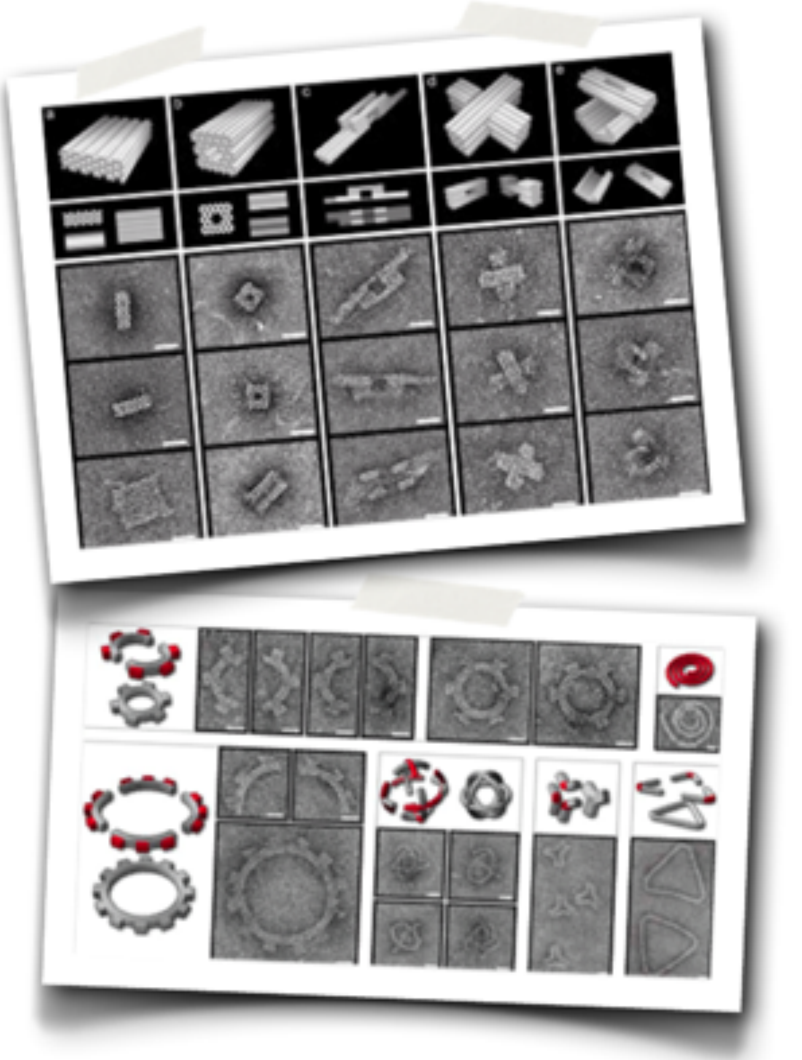


Picture taken from Luther et al. (1998), Nature 396:245

The possibilities inherent in DNA computation are nicely Illustrated by two young lions of the Wyss Institute:

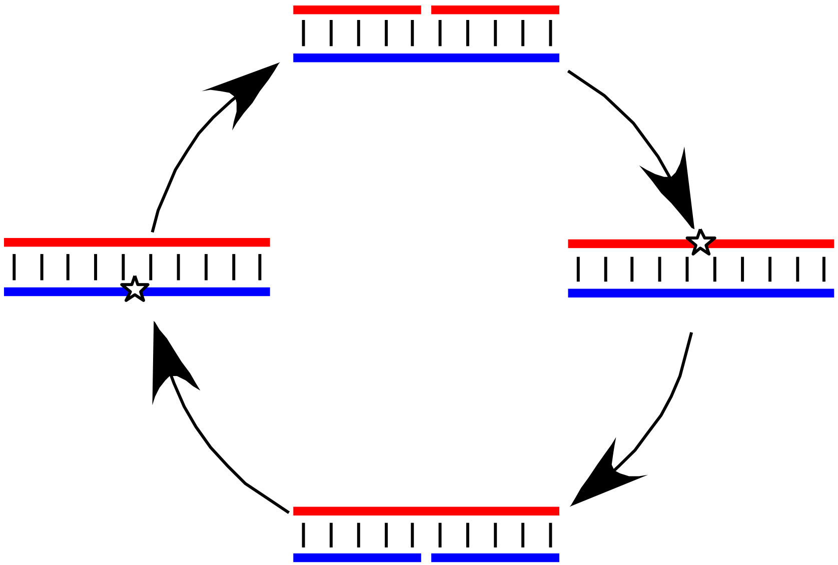


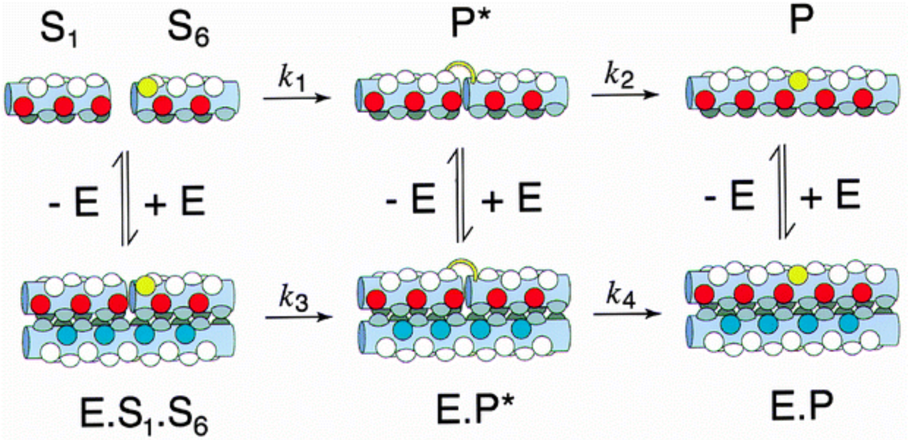
Peng Yin



William Shih

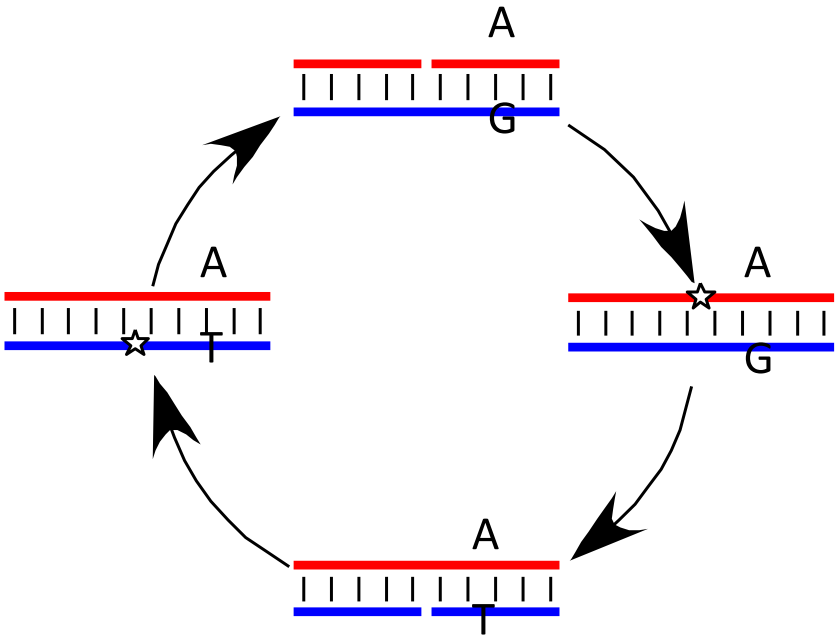
Key discoveries by Ghadiri and co-workers have provided alternatives to nucleic acid replicators

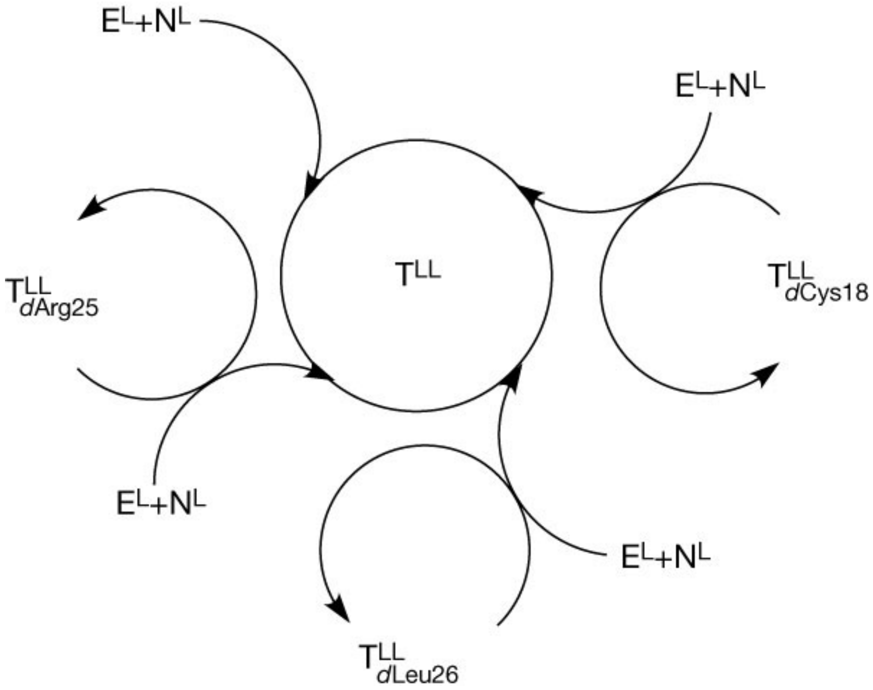




Severin et al. (1997), Nature 389:706

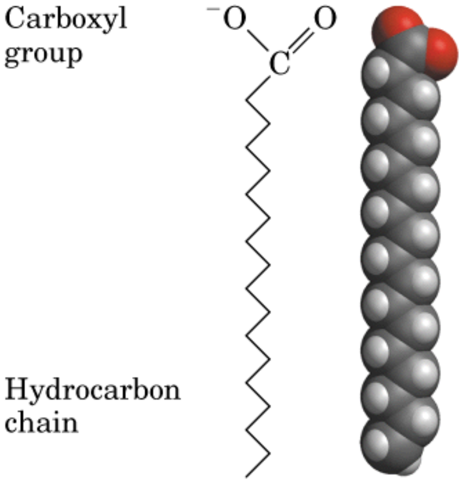
However, peptide replicators show tendencies for ‘error correction,’ either in terms of sequence or stereochemistry. While there are advantages to such features, they ultimately limit evolution.



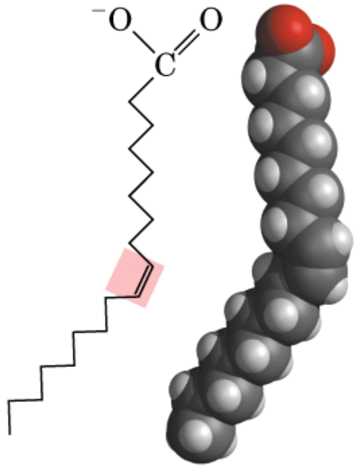


Saghatelian et al. (2001), Nature 409:797

Another contender: lipids



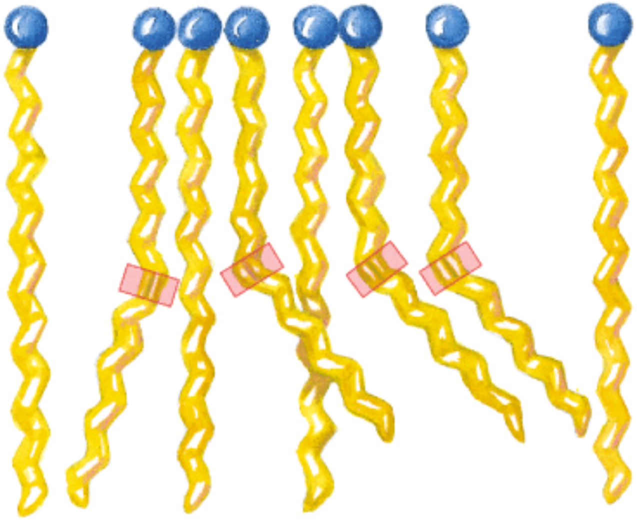
(a)



(b)

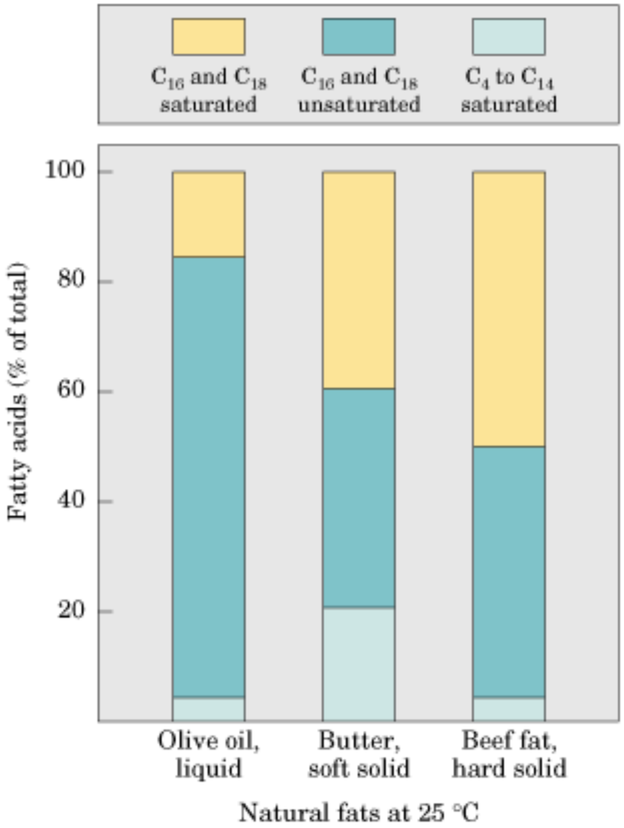
Saturated fatty acids

(c)

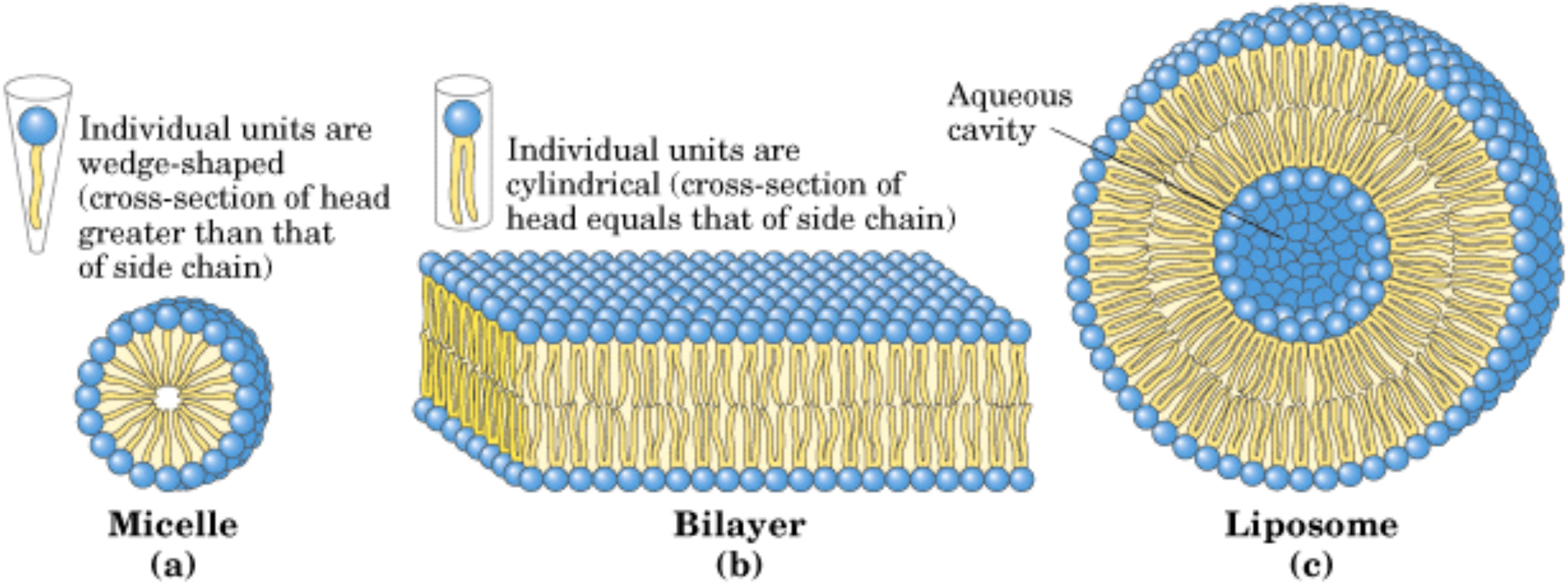


Mixture of saturated and unsaturated fatty acids

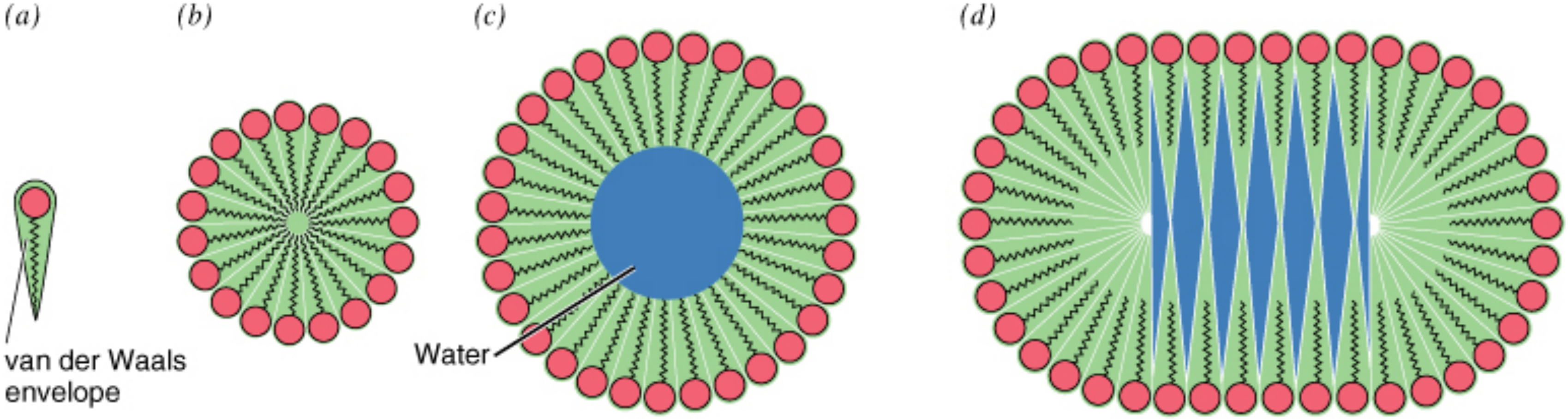
(d)



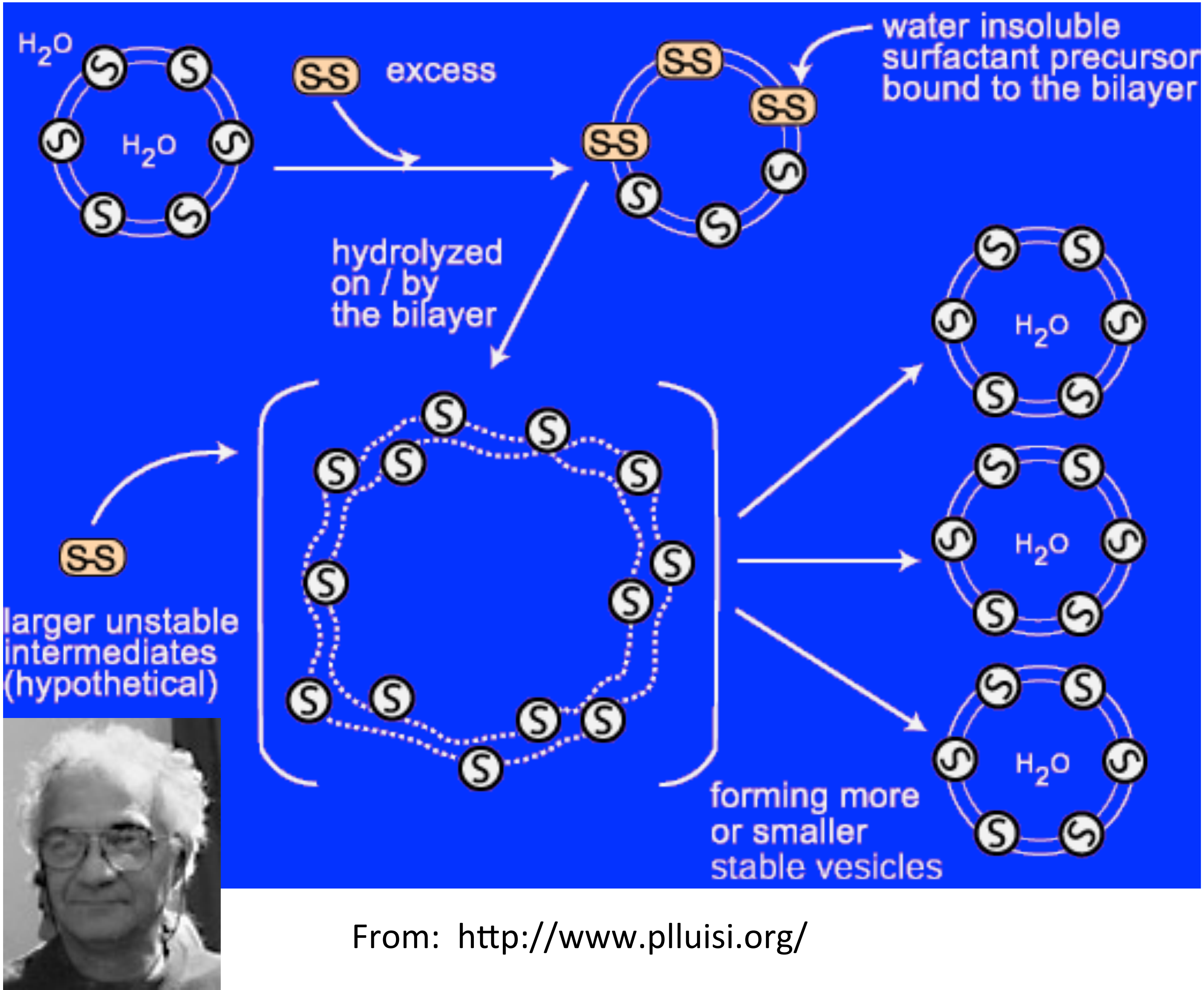
Different lipids can aggregate into different morphologies



These aggregates can grow and change

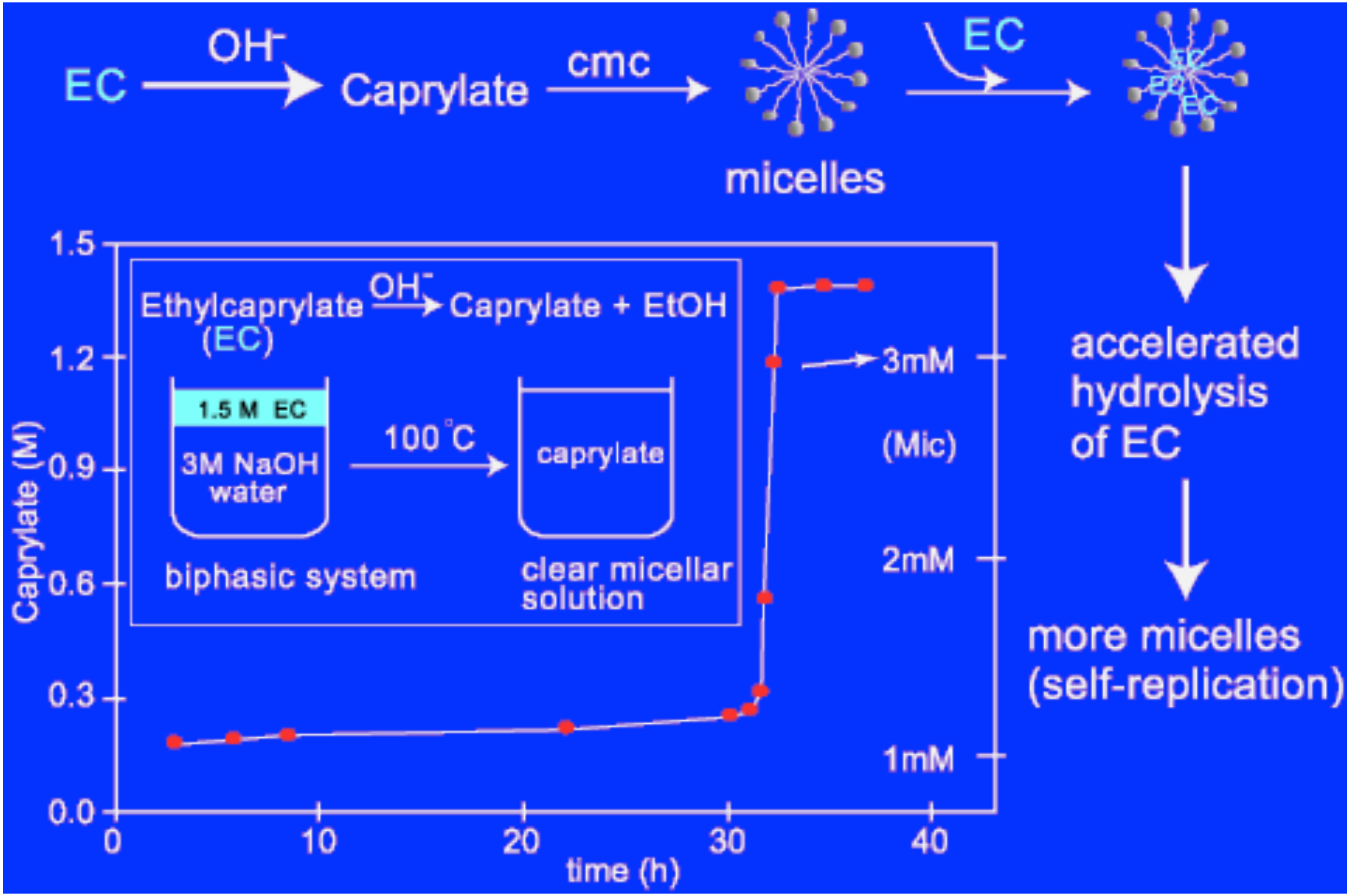


This includes the ability to grow and divide

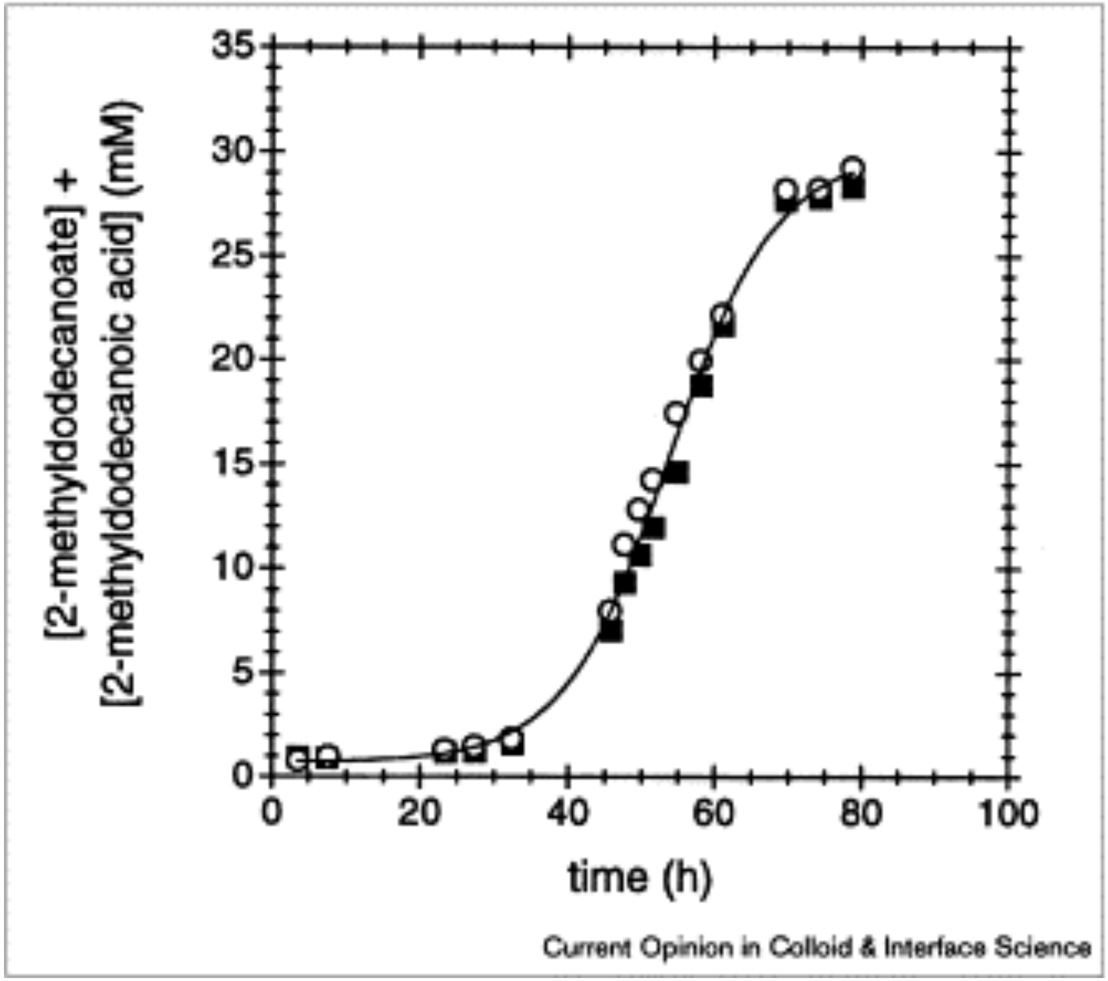


From: http://www.plluisi.org/

In fact, lipid aggregates are capable of self-replication



However, such replicative processes do not yet appear to be chiral (unlike nucleic acids or peptides)



Luisi et al. (1999), Curr Opin Colloid & Interface Sci 4:33

Nonetheless, such self-replication could be imagined to have occurred in a variety of ways in the prebiotic world

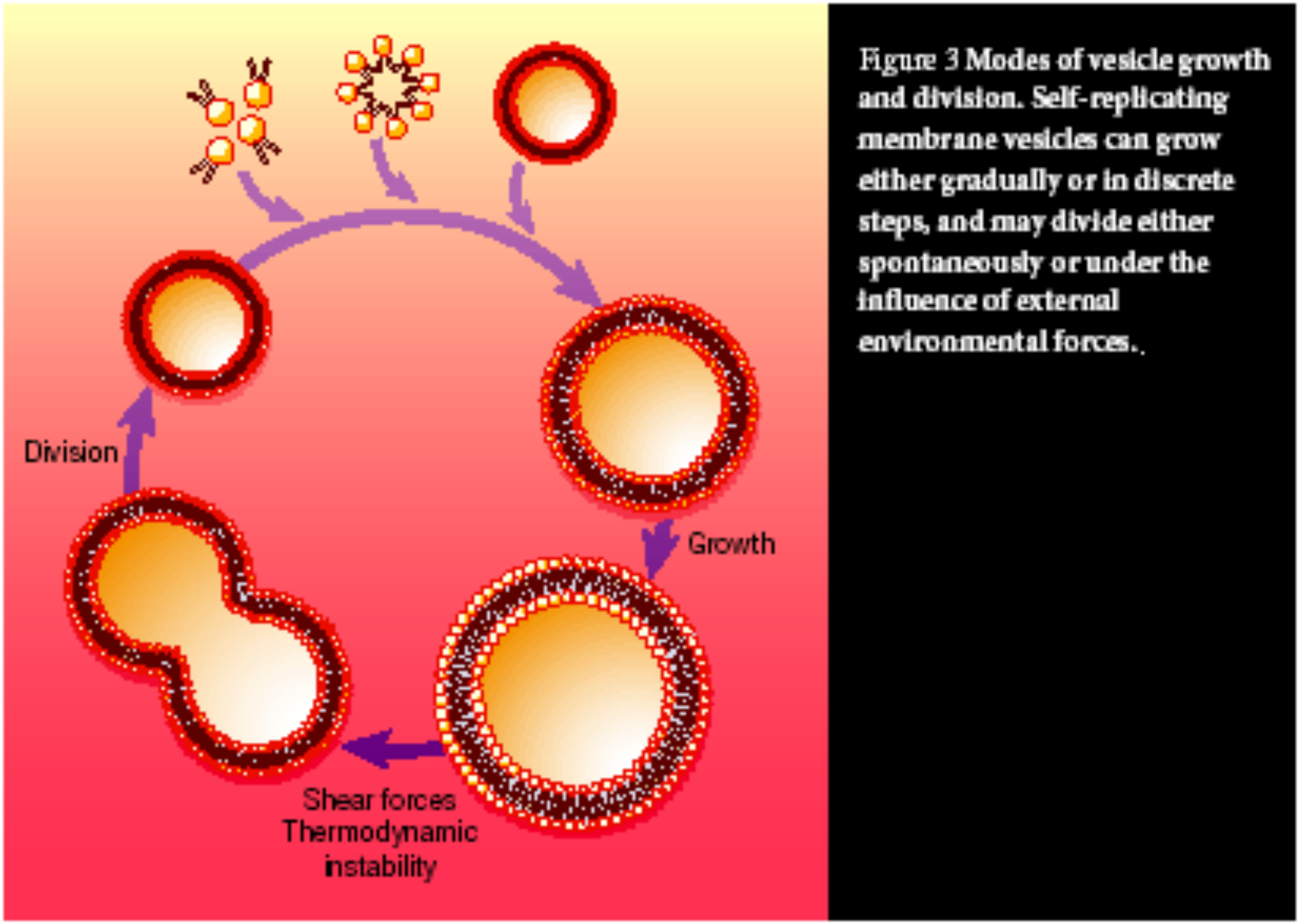
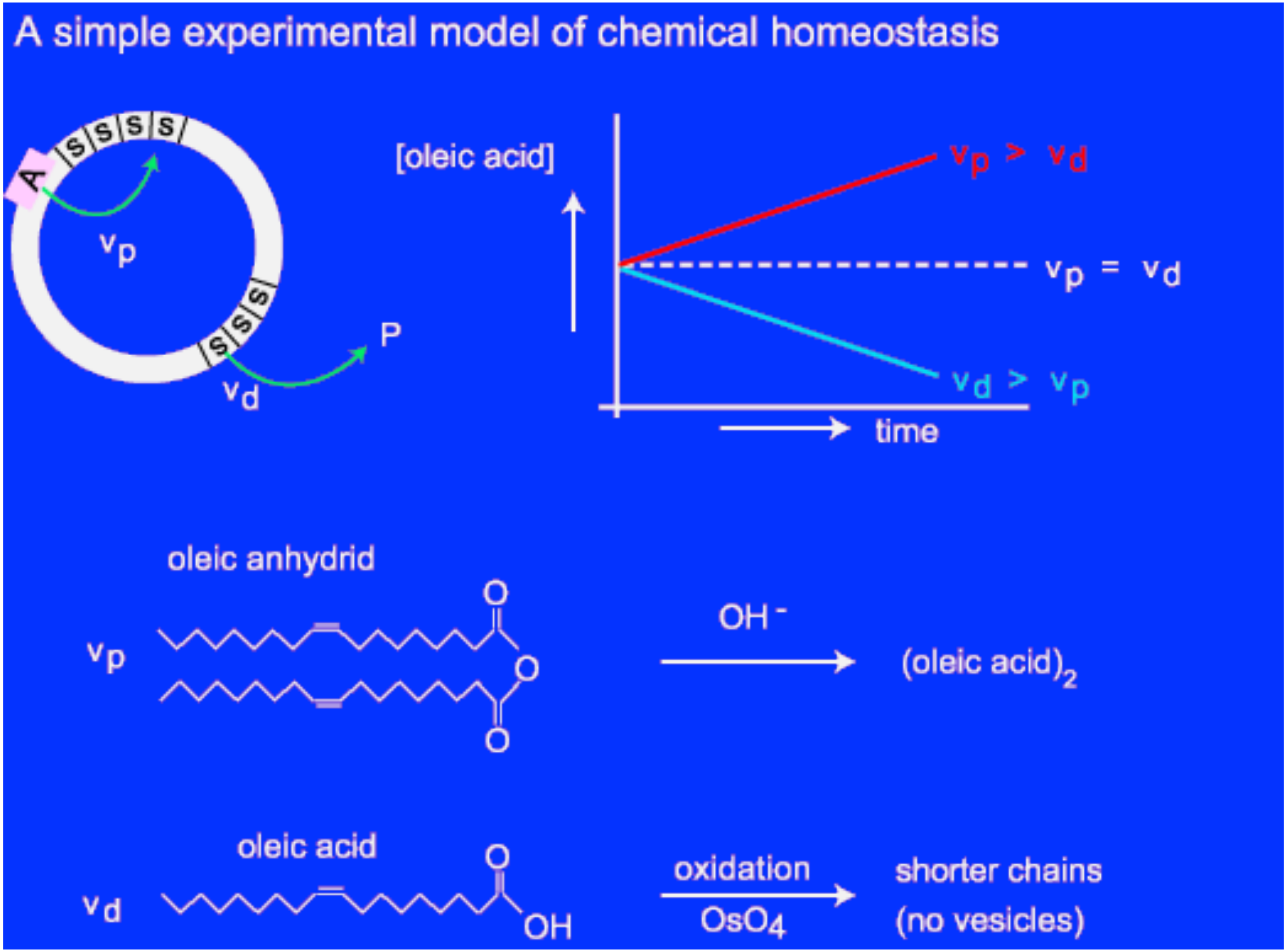


Figure 3 Modes of vesicle growth and division. Self-replicating membrane vesicles can grow either gradually or in discrete steps, and may divide either spontaneously or under the influence of external environmental forces.

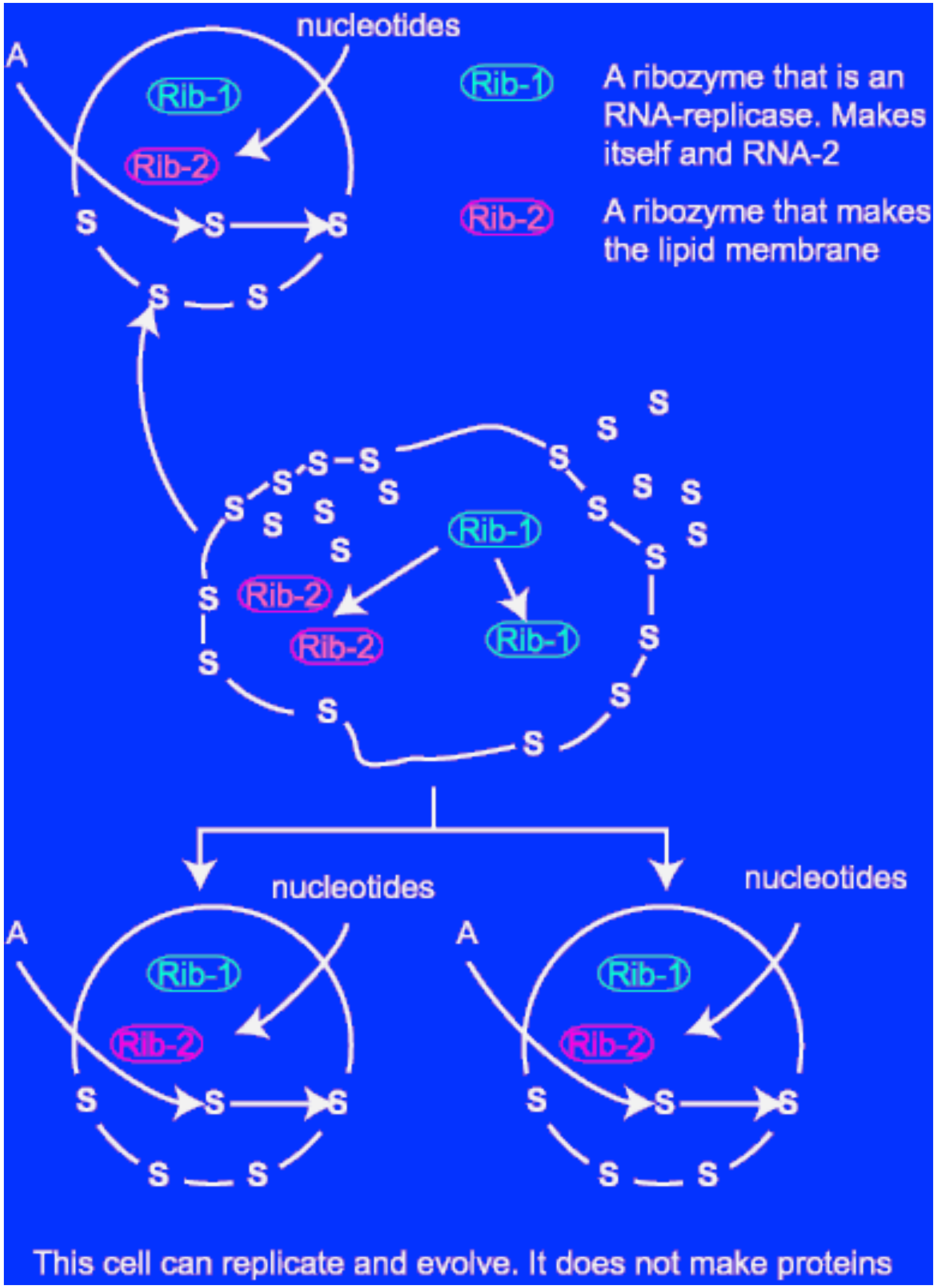
From Szostak et al. (2001), Nature 409:387

The balance of synthetic and destructive forces might have led to a type of (sustainable?) dynamic equilibrium



A simple experimental model of chemical homeostasis

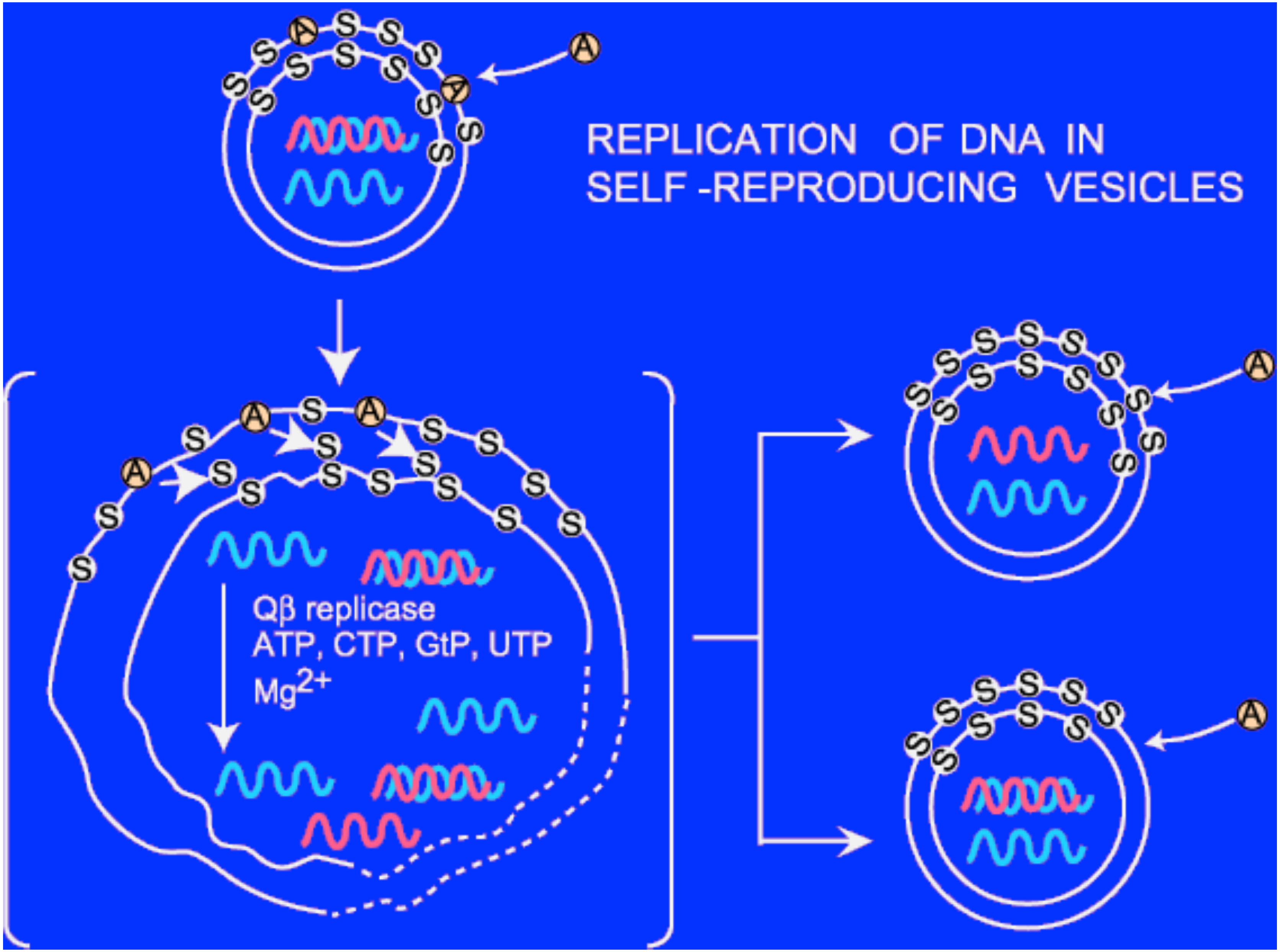
Nucleic acids and lipid aggregates could have interacted with one another



Rib-1	A ribozyme that is an RNA-replicase. Makes itself and RNA-2
Rib-2	A ribozyme that makes the lipid membrane

This cell can replicate and evolve. It does not make proteins

And in fact basal experiments of this type work



Cooperative interactions might have led to a ‘protocell’ in which genetic and lipid replicators were mutually interdependent

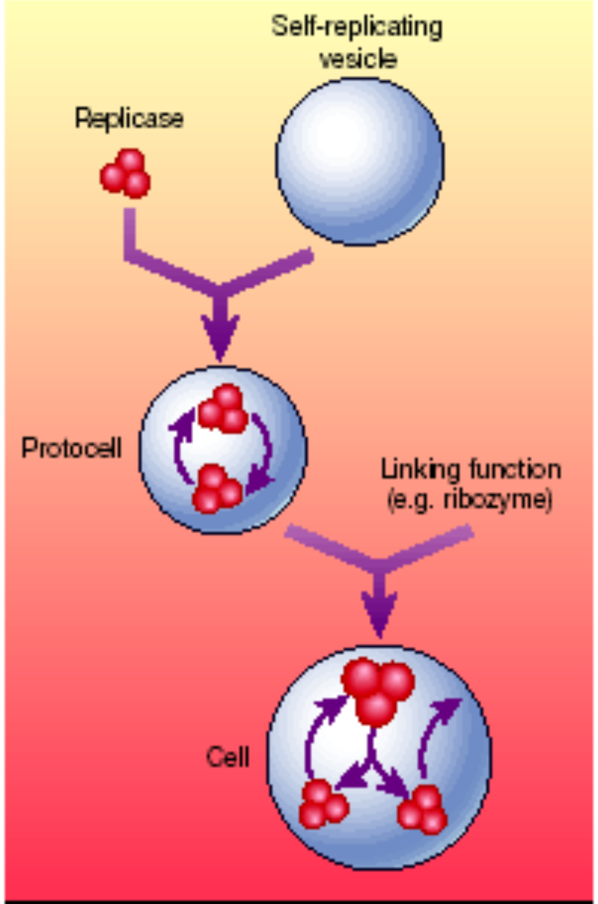


Figure 2 **Outline of proposed pathway for synthesis of a cell.** The first major synthetic intermediates are an RNA replicase and a self-replicating vesicle. These are combined into a protocell, enabling rapid evolutionary optimization of the replicase. Addition of an RNA-coded linking function, such as a lipid-synthesizing ribozyme, completes the cellular structure.

Such cooperative interactions would ultimately have enabled the true Darwinian evolution of self-replicases

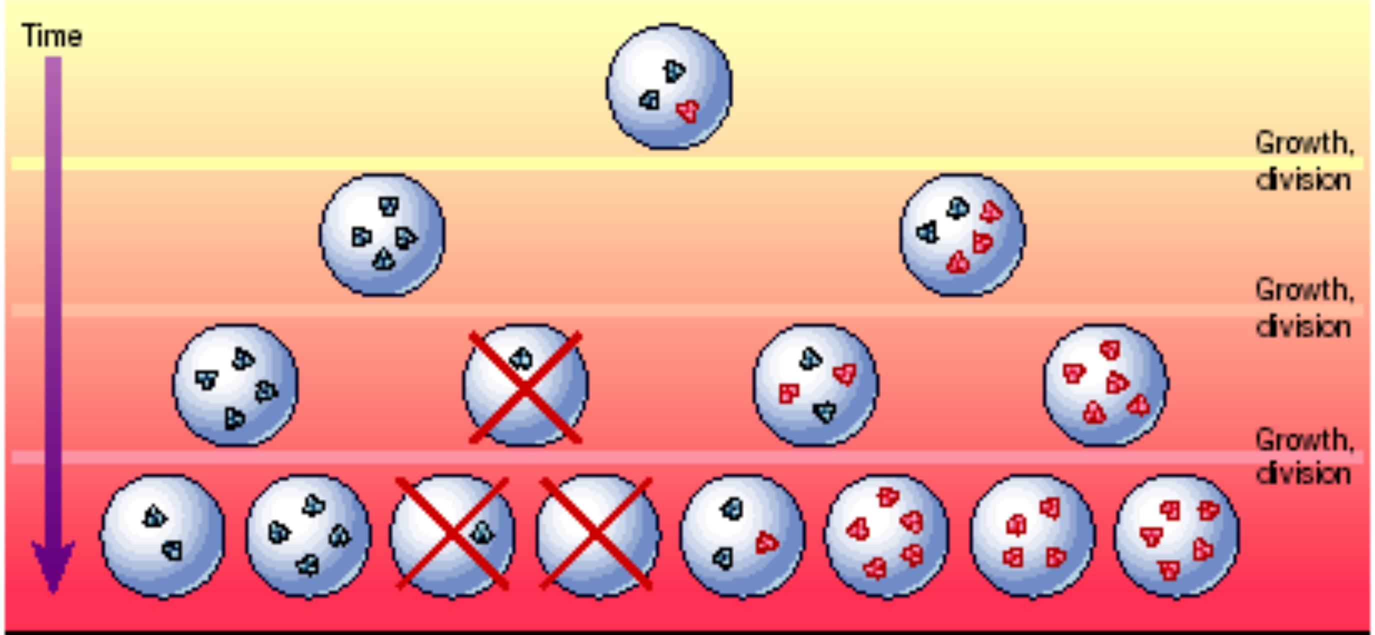


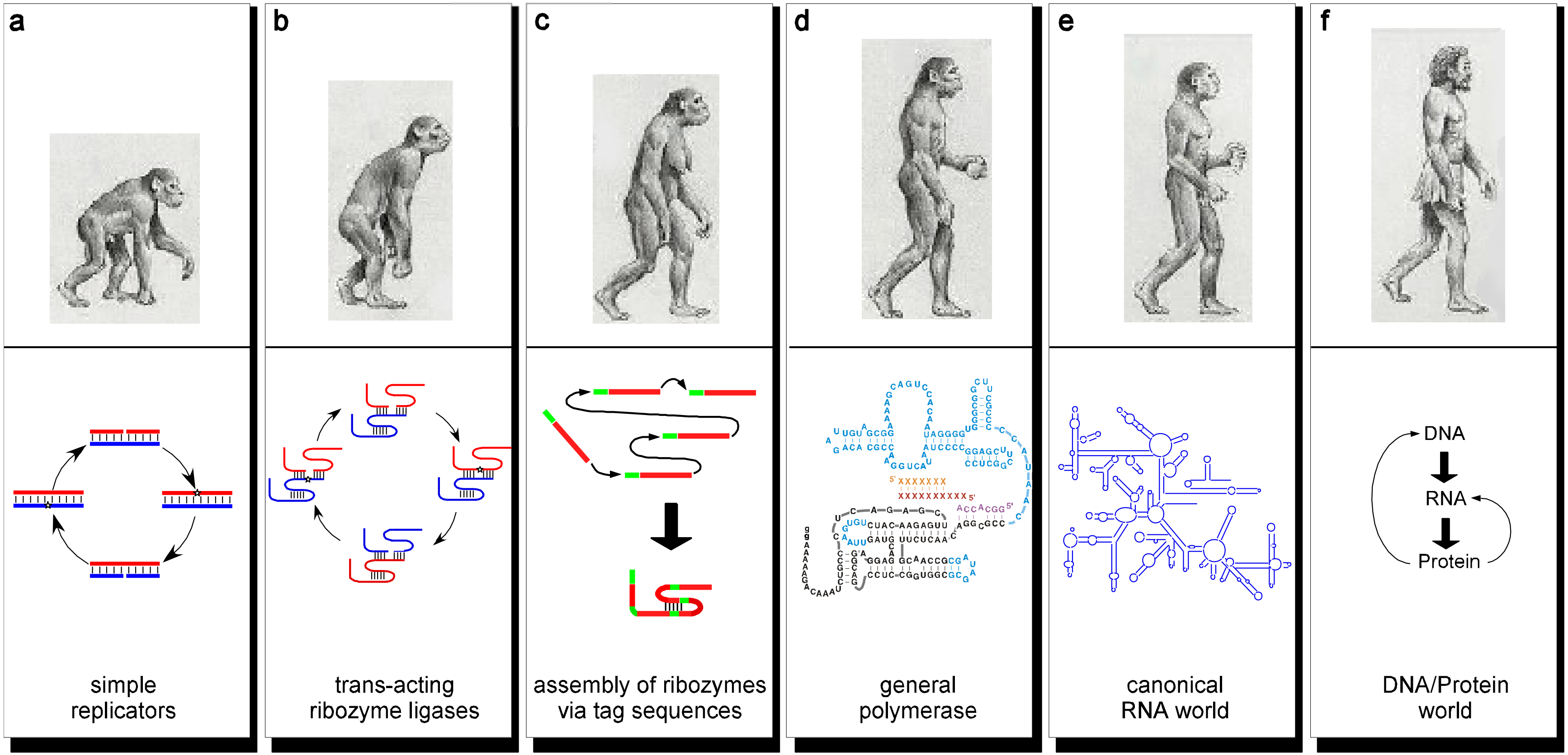
Figure 1 Role of vesicles in enabling darwinian evolution of a replicase. The vesicular compartment ensures that molecules related by descent are, on average, kept in physical proximity to each other, allowing a superior mutant replicase (red) to preferentially self-replicate, in comparison to the parental replicase (black). The evolutionary advantage of increased replication is amplified as vesicles with superior replicase molecules are more likely to give rise to vesicles with at least two replicase molecules (or a replicase and a template molecule). Vesicles with less than two replicase molecules (indicated by an X) and the progeny of these vesicles cannot continue RNA self-replication. In this way, vesicles with superior replicase molecules become an increasing fraction of the vesicles that maintain replicase activity.

- Nucleic acid imperialism
- Doppelgangers
- Looking backwards

The utility of doppelgangers

That which I can make money off of, I understand.

In some ways, what we are trying to do is show that it is pssible to demonstrate a path for early life. We can never fully recapitulate it.



simple replicators

trans-acting ribozyme ligases

assembly of ribozymes via tag sequences

general polymerase

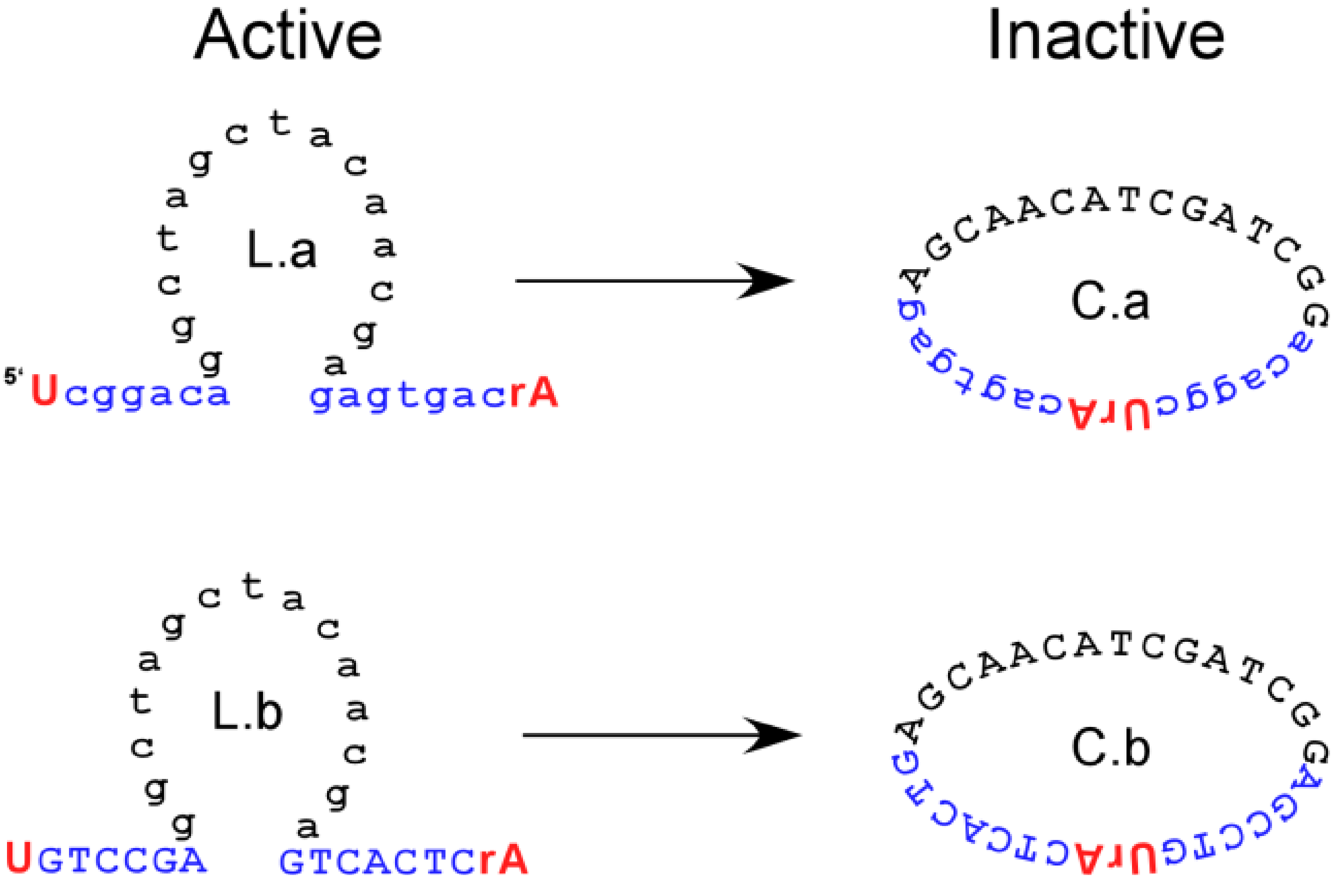
canonical RNA world

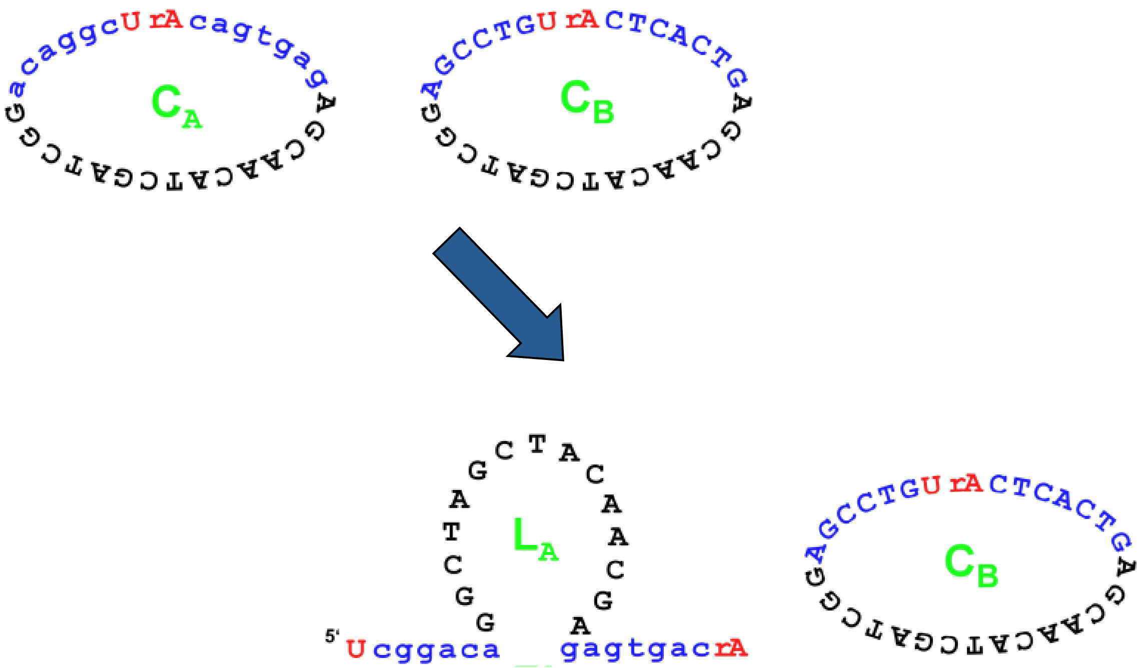
DNA/Protein world

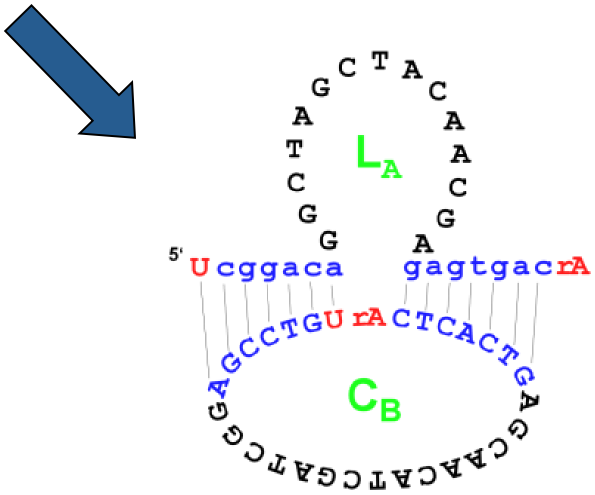
You are here

How do you get Here?

A conformational replicator based on the 10-23 deoxyribozyme



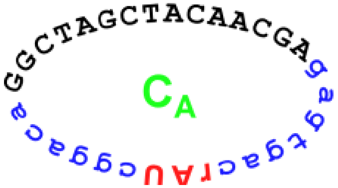


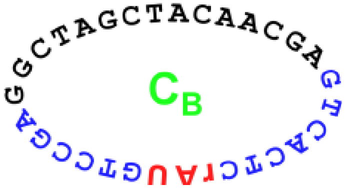


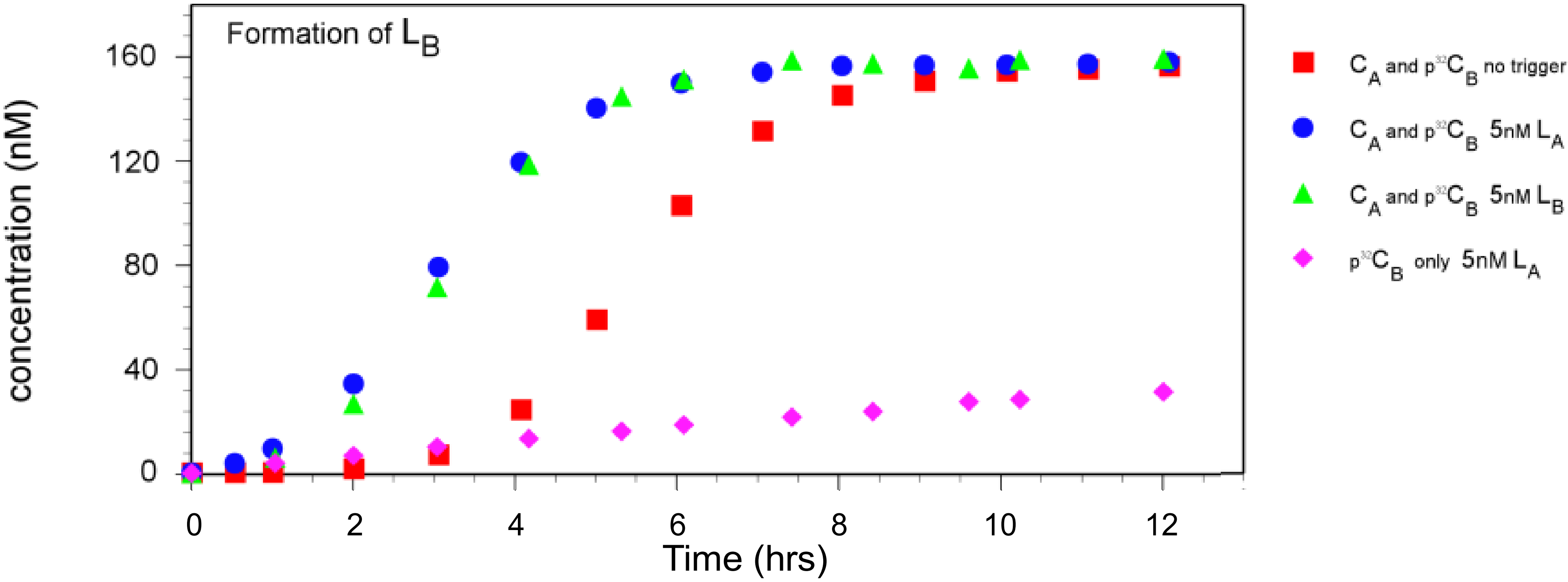
Matt Levy

Co-PI, now on faculty at Einstein

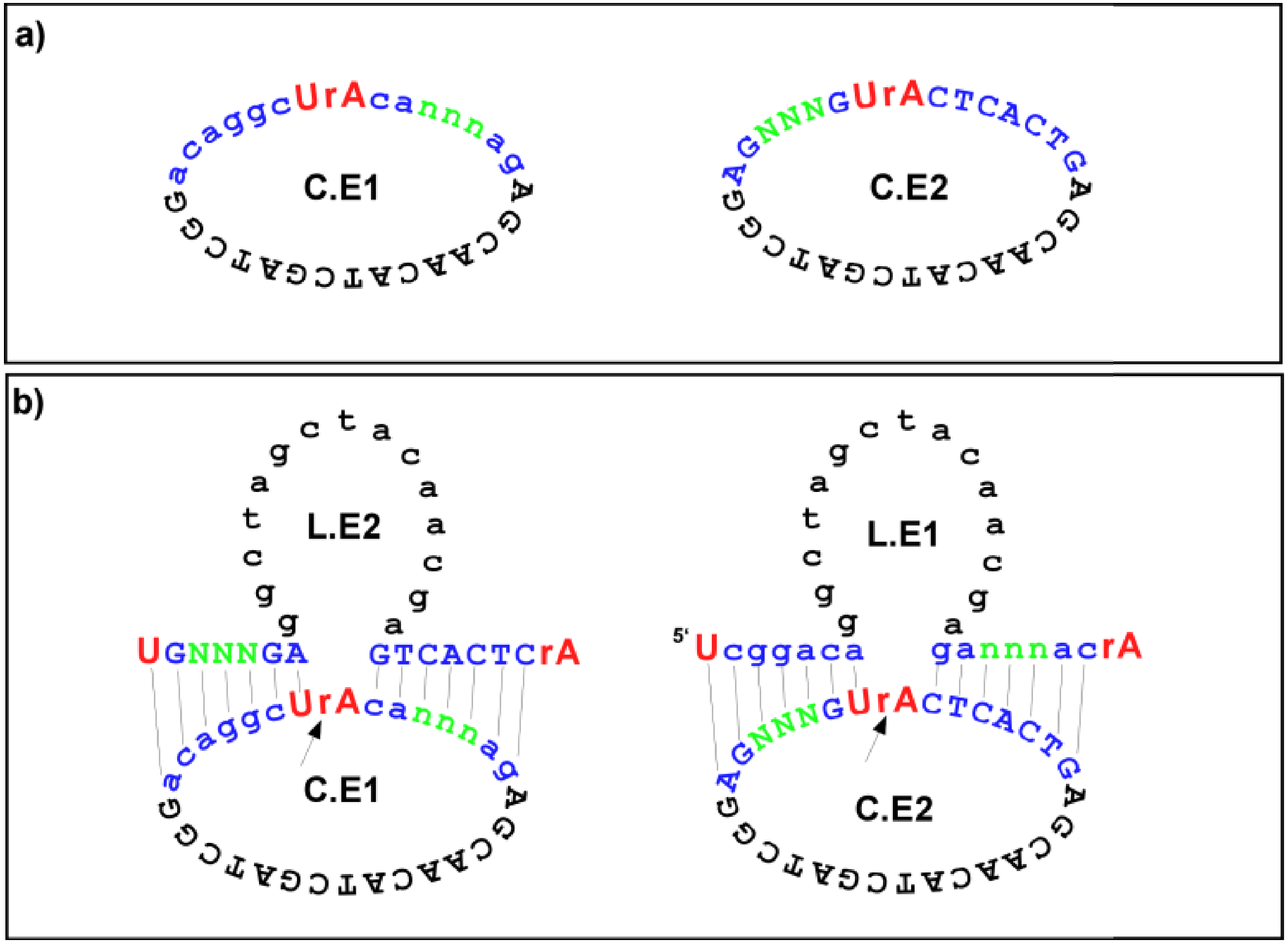
Remarkably, this works



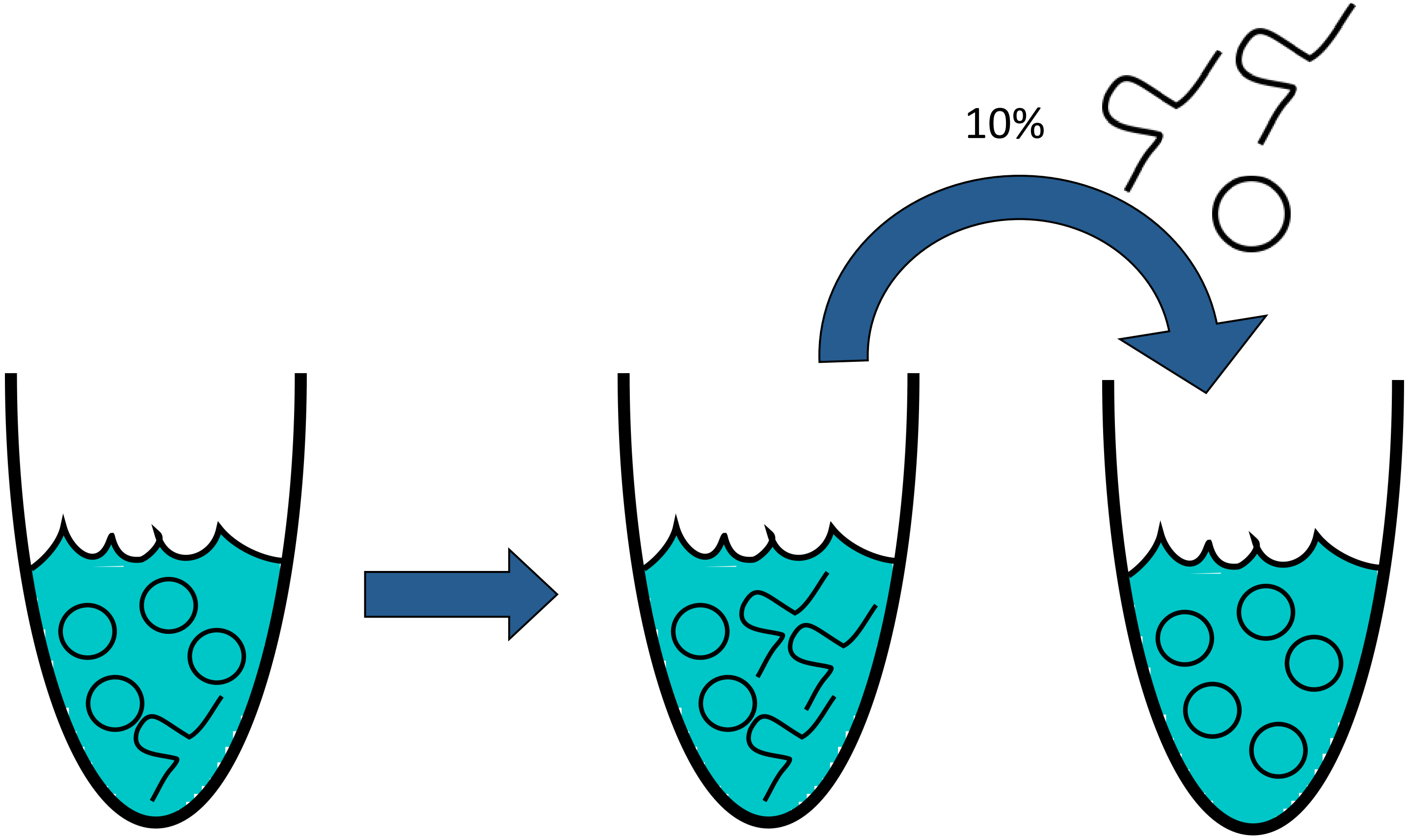




Can we evolve better replicators?

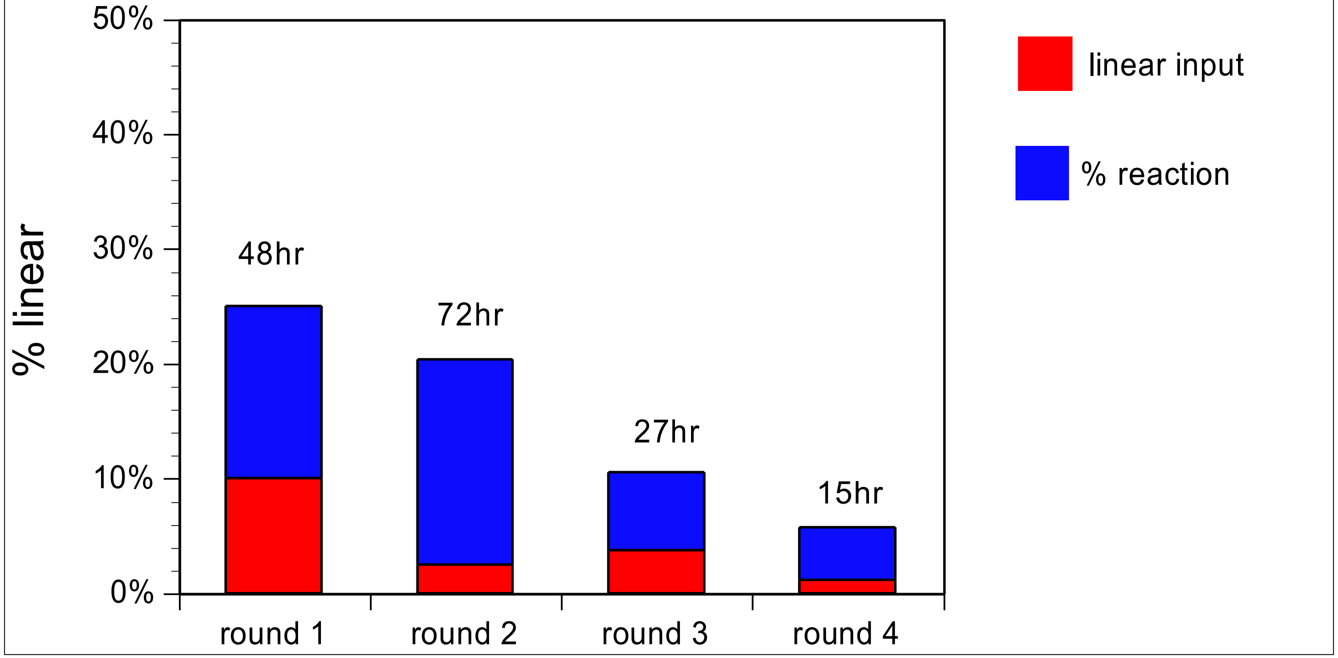


Selection Procedure



“Winners”, species that can cleave and be cleaved, seed the next round

Linears accumulated on serial transfer



The sequence pool collapsed to the imprint

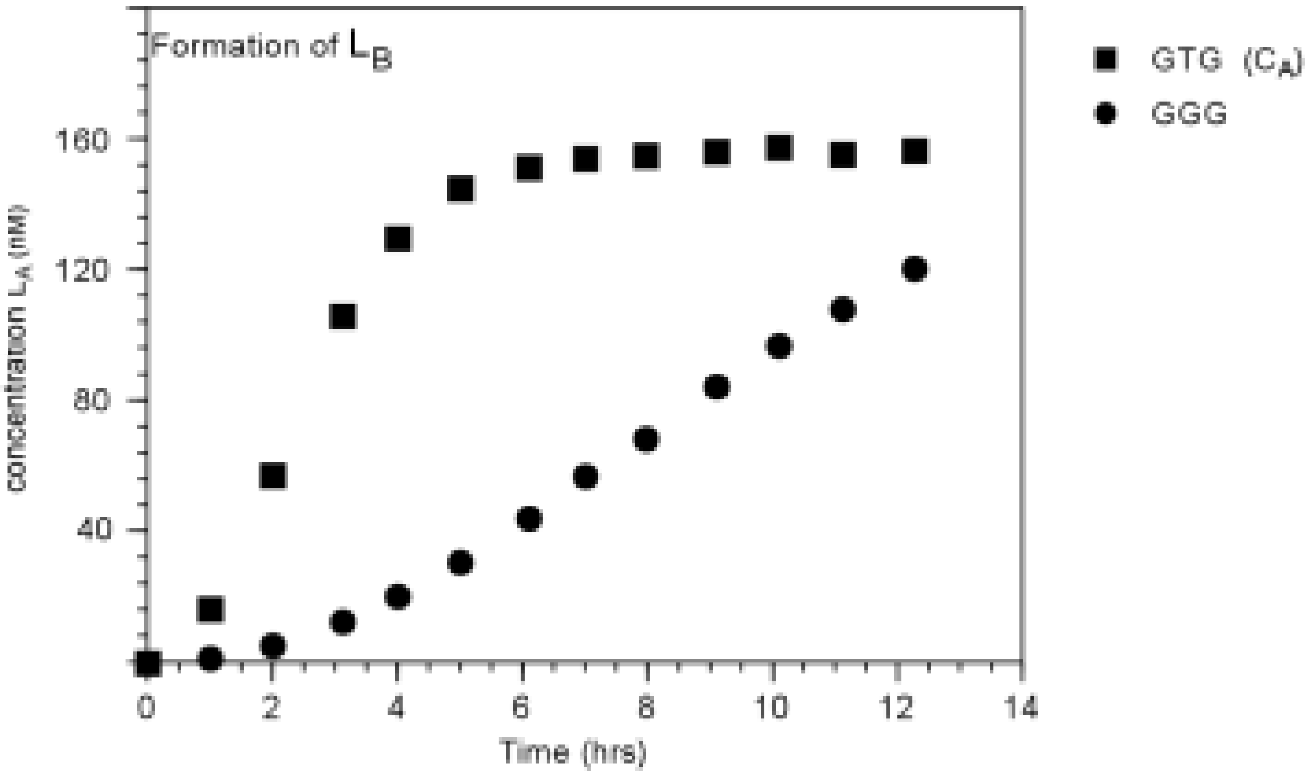
E1 round 0		E1 round 4		E2 round 0		E2 round 4	
aagga	3	agtga	13	GGTAG	2	GTCCG	20
acata	2	aggga	9	GGCCG	2	GGCCG	3
aggga	2			GTCTG	2	GTCAG	3
acaga	2			GTAGG	2	GCCCG	
agcaa	2			GCTCG		GATCG	
agaga	2			GAGCG		GCGGG	
agcca				GGTTG		GCAGG	
atgga				GACCG		GCGTG	
ataga				GATGG			
agtga				GTCAG			
accta				GCTAG			
atcga				GTTTG			
aagga				GTTGG			
accga				GCGAG			
aataa				GCATG			
agtca				GTATG			
				GCCCG			
				GGGTG			

Amongst competitors, the dominant species was the fittest species

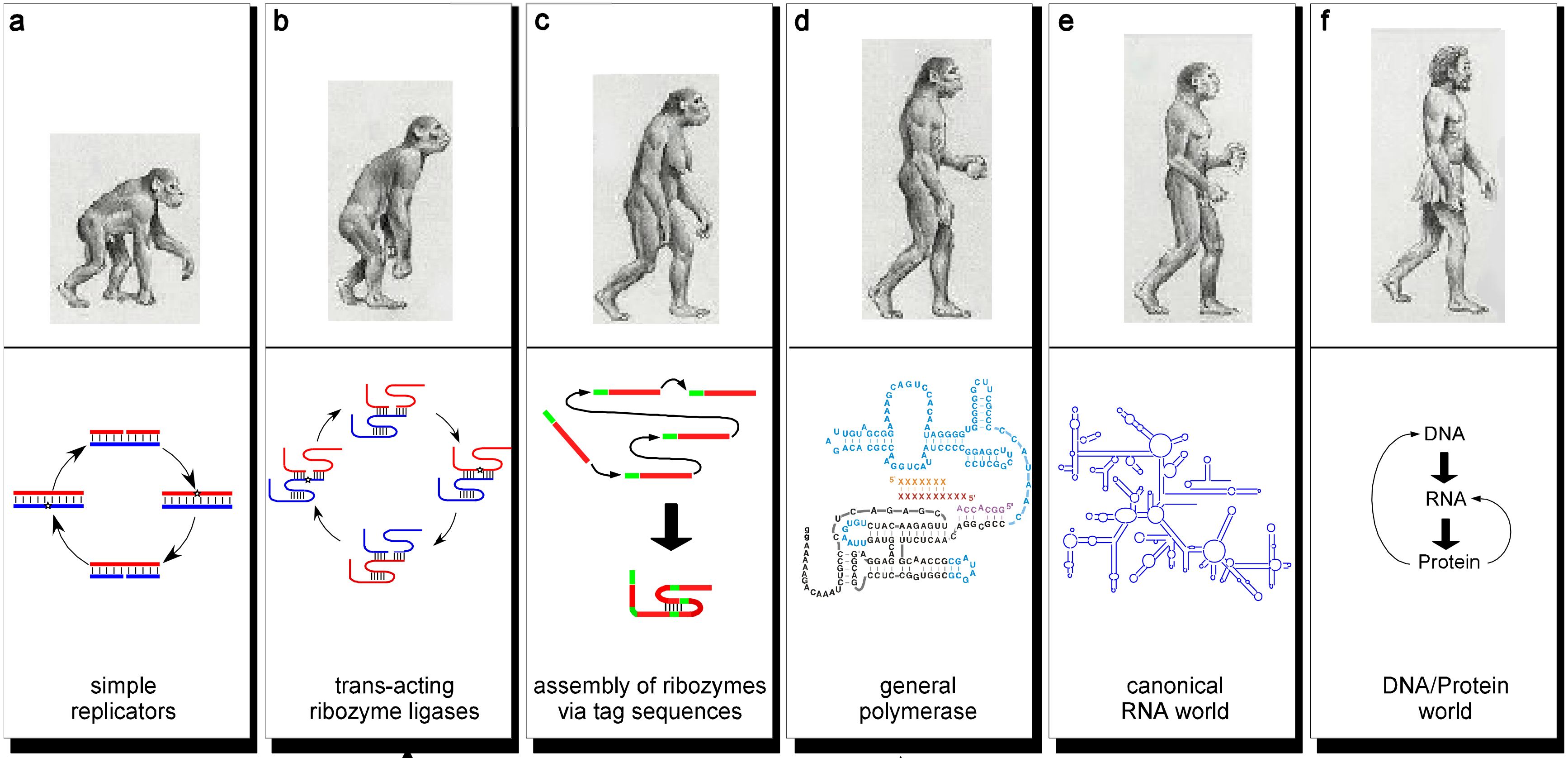
E1

round 4

agtga	13
aggga	9



In some ways, what we are trying to do is show that it is pssible to demonstrate a path for early life. We can never fully recapitulate it.



a simple replicators

b trans-acting ribozyme ligases

c assembly of ribozymes via tag sequences

d general polymerase

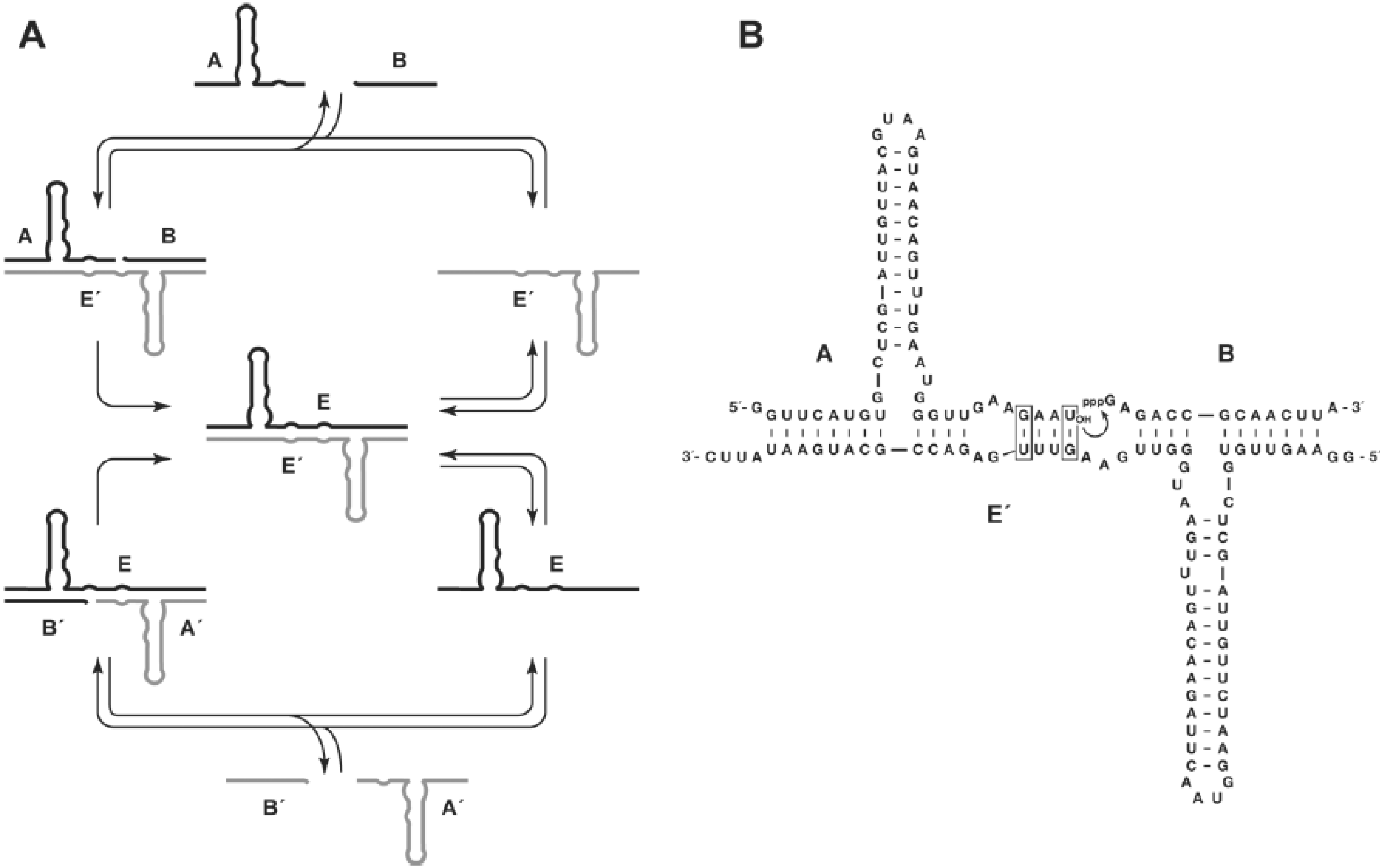
e canonical RNA world

f DNA/Protein world

You are here

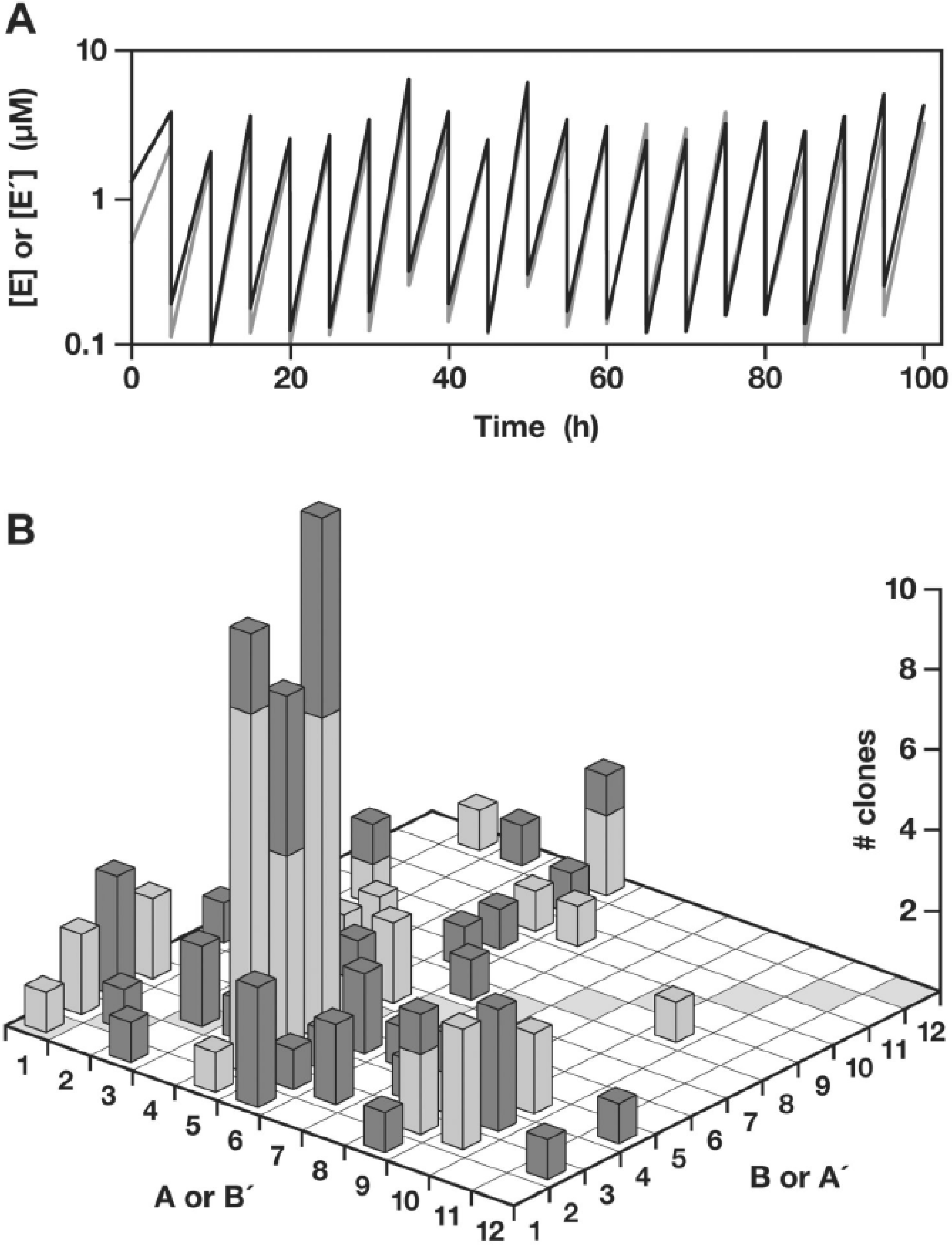
How do you get Here?

Replicases can assemble themselves from pieces



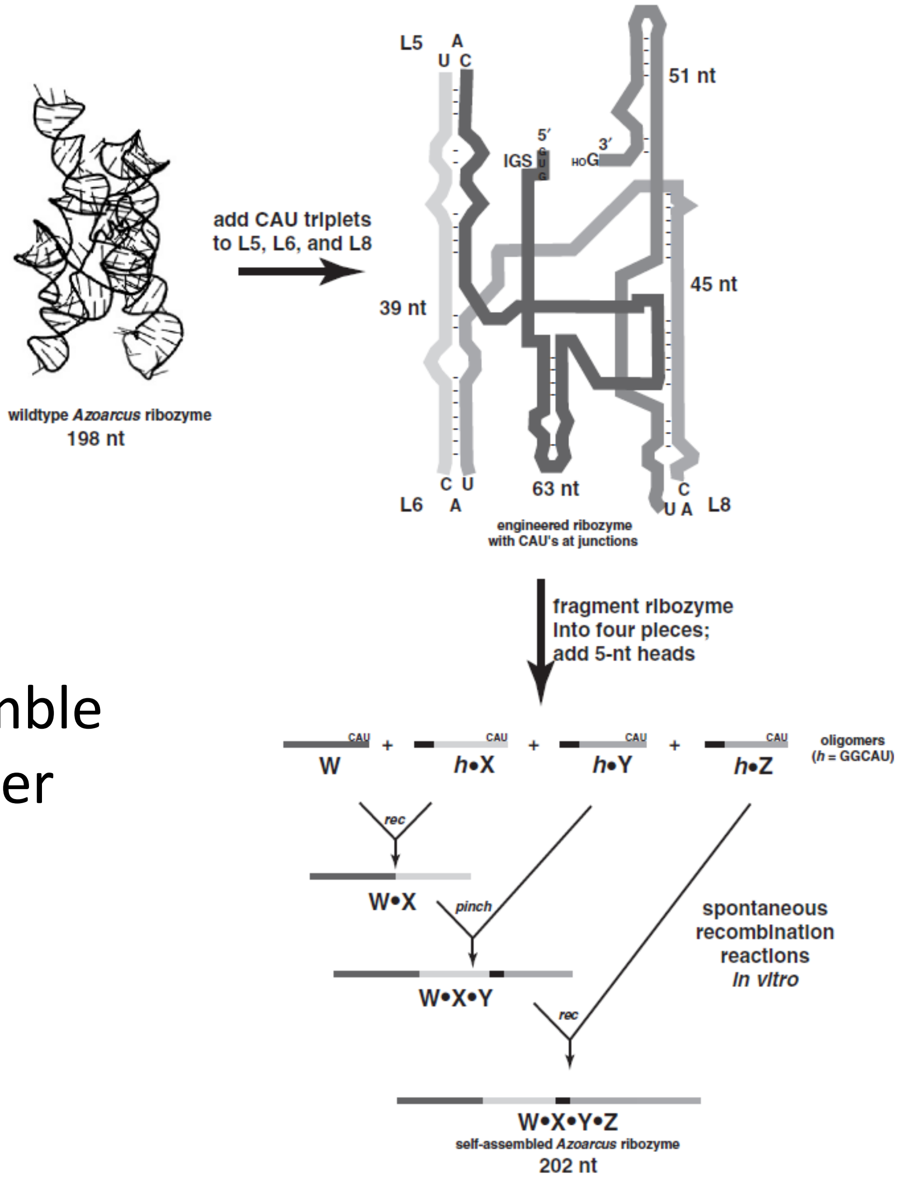
Lincoln and Joyce (2009), Science 323:1229

Selections for complex phenotypes can ensue

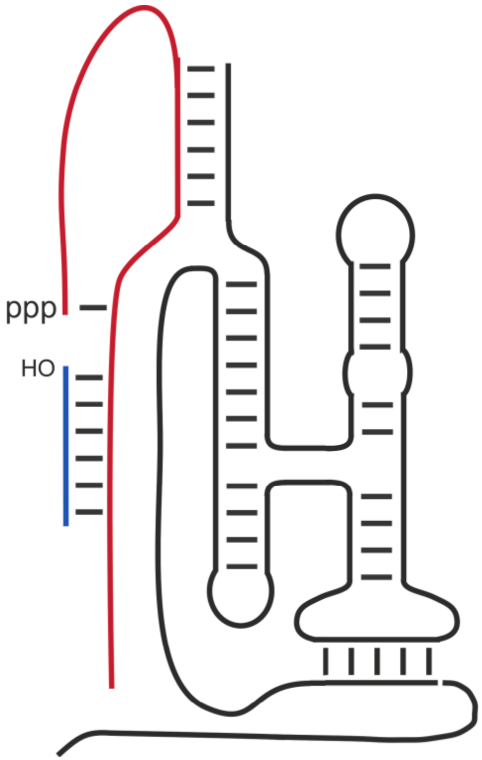


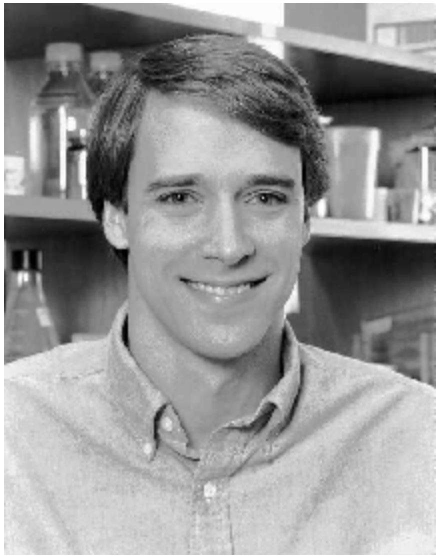
Ribozymes can assemble themselves into longer ‘genomes’

Draper et al. (2008)
NAR 36:520



Some researchers have tried to build a ‘ribozyme polymerase’ based on the Bartel Class I ligase: the ribozyme … that shouldn’t exist!





David Bartel,
Whitehead Institute

The miracle of the Bartel ligase

- The probability of finding the Class I ligase was ca. 4×10^{-19} (Ekland et al., 1995). In other words, Bartel should have found it once every 10,000 times he did the experiment. This is rightly interpreted to mean that there are multiple motifs out there with similar catalytic activities.

- It seems that the nature of the fitness landscape surrounding these extremely active ribozymes is very sparse. That is, movement away from the privileged structure produces much less active variants.

- Therefore, it may be that highly complex ribozymes do not evolve from simpler ribozymes by 'hill climbing' or by the accretion of additional information.

- It is possible that new complex ribozymes evolve from old complex ribozymes with different functions, or that recombination of simple ribozymes can lead to complex ribozymes.

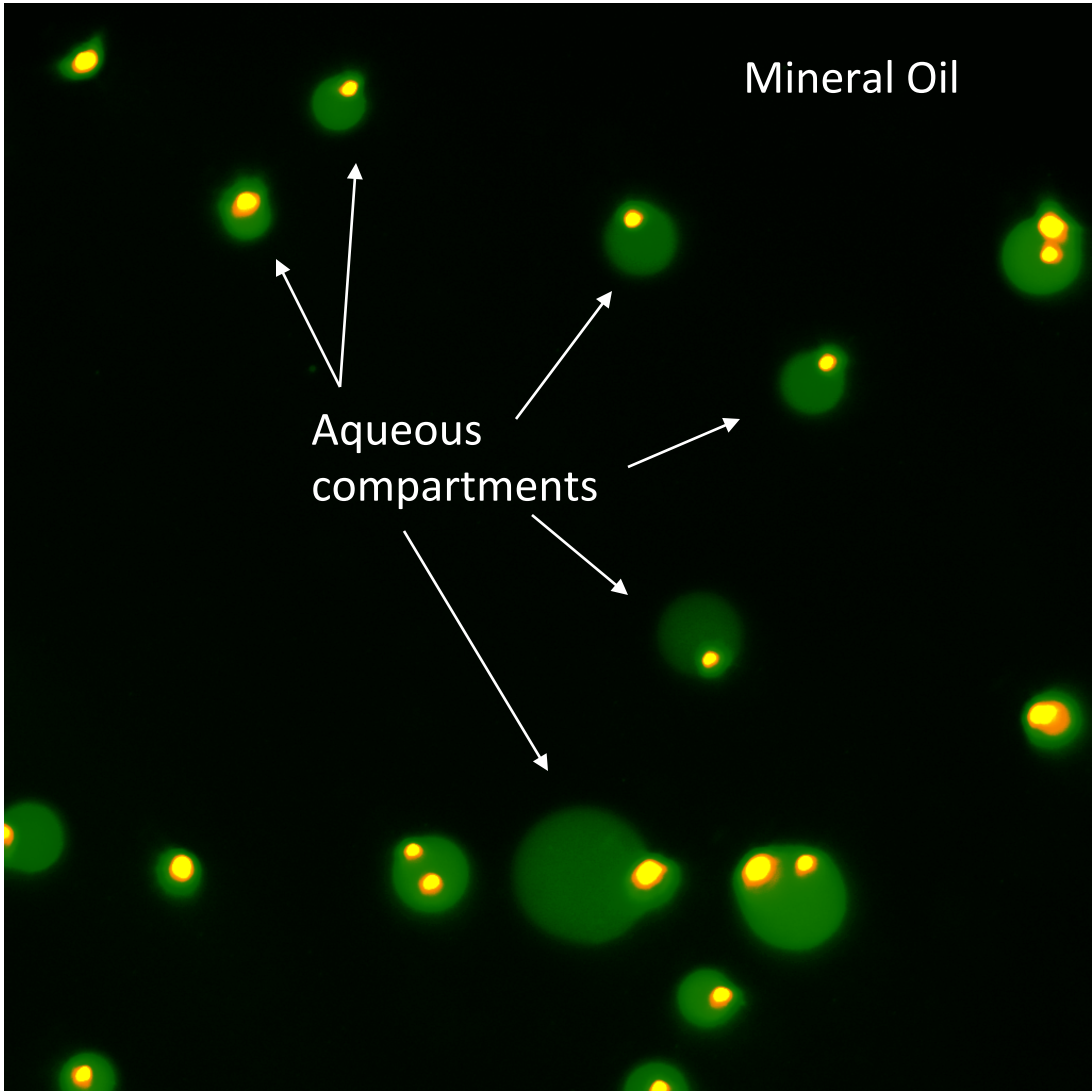
- But it is also possible that novel functions are miracles, in the sense of being extremely improbable events. In the absence of gradualism, origins becomes even harder to investigate.

In vitro compartmentalization (IVC) for acellular evolution

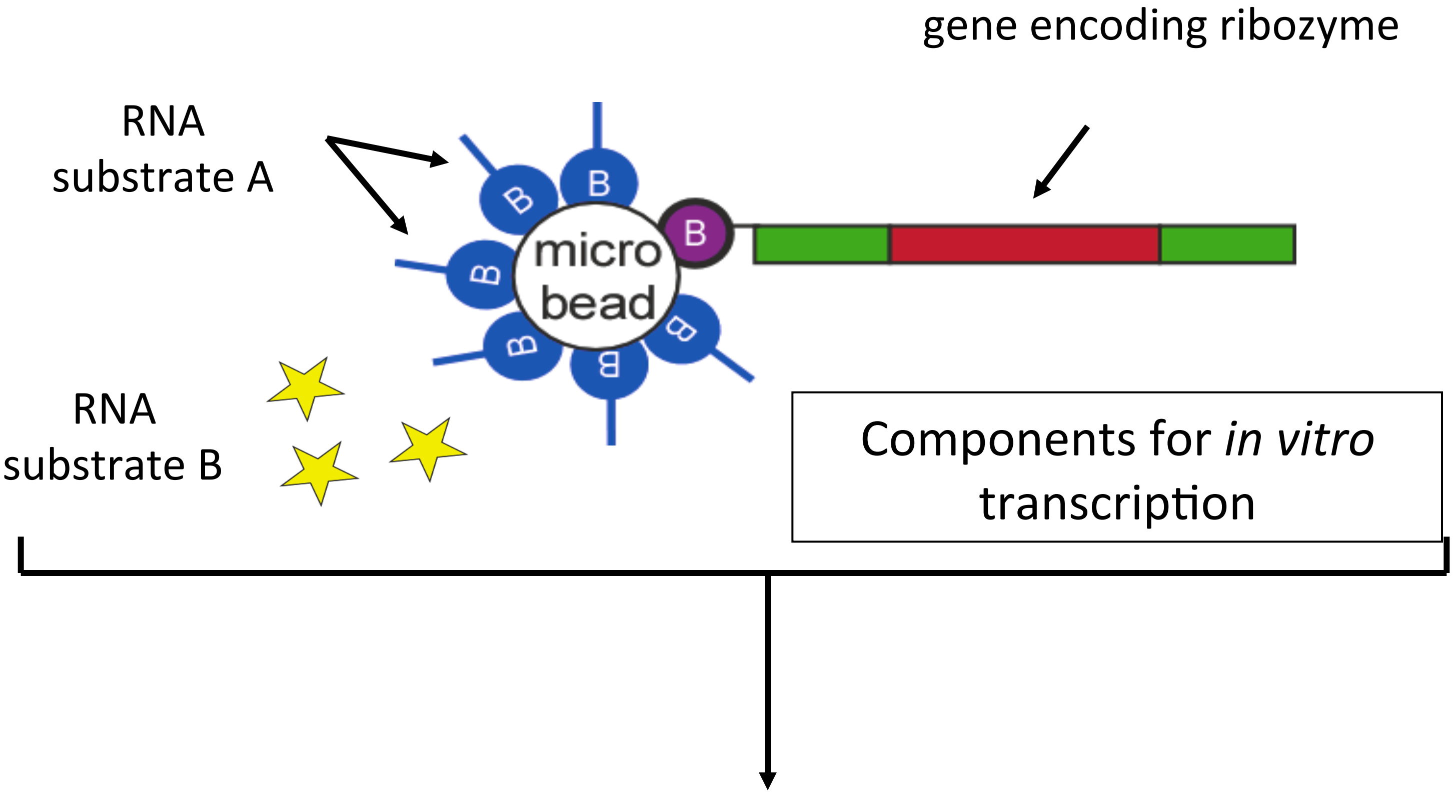
- Utilizes water-in-oil emulsion to generate discrete cell-like compartments

- Rapid screening of large libraries $\sim 10^{8-12}$ sequences/mL

- All *in vitro*



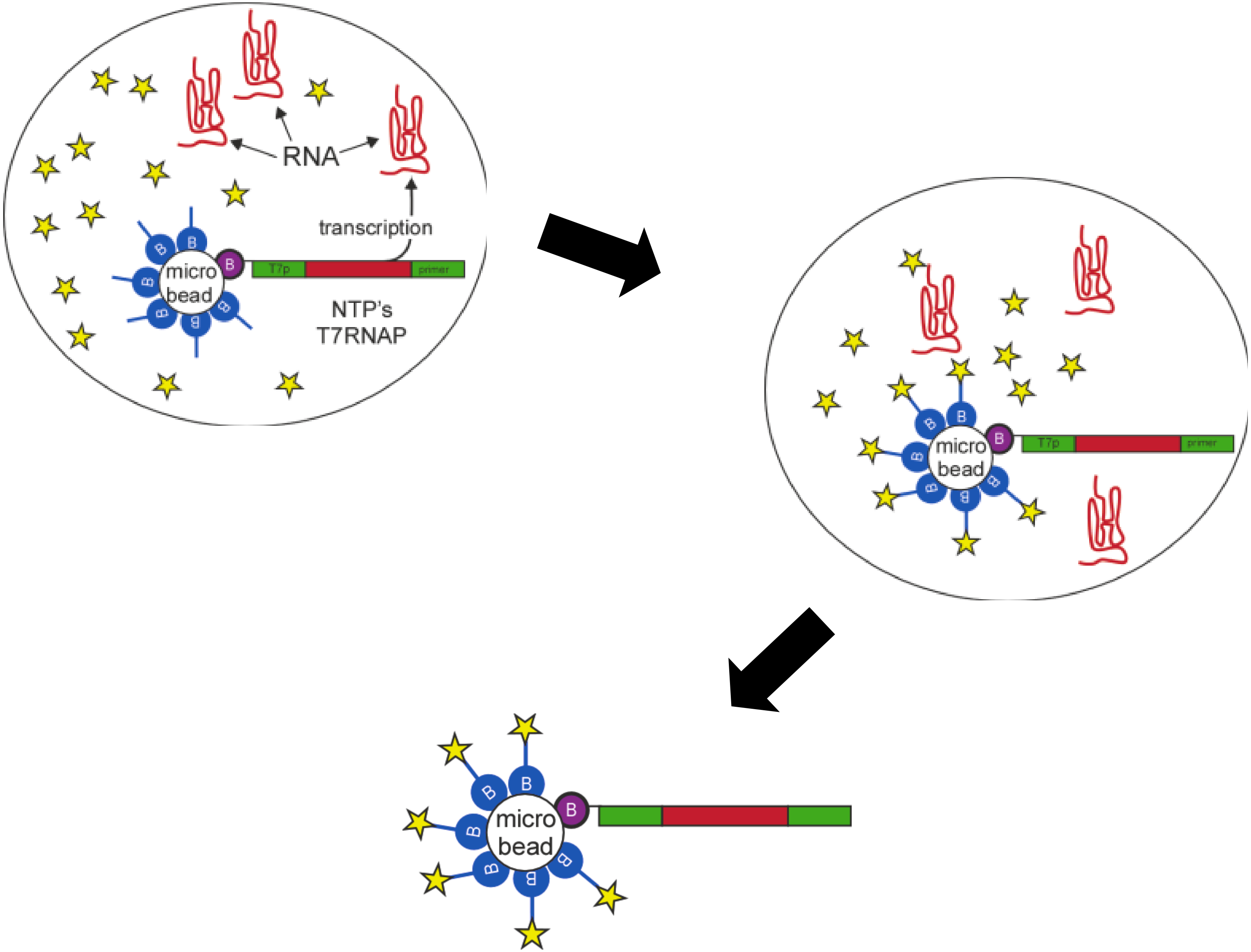
Our entry point was to use IVC to select for ribozymes capable of functioning in *trans*



Components for *in vitro* transcription

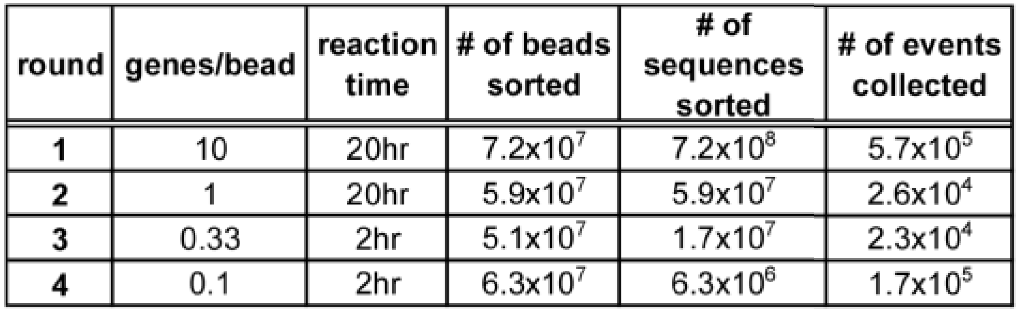
Emulsification in oil droplets
(*in vitro* compartmentalization)

Levy et al., RNA (2005)

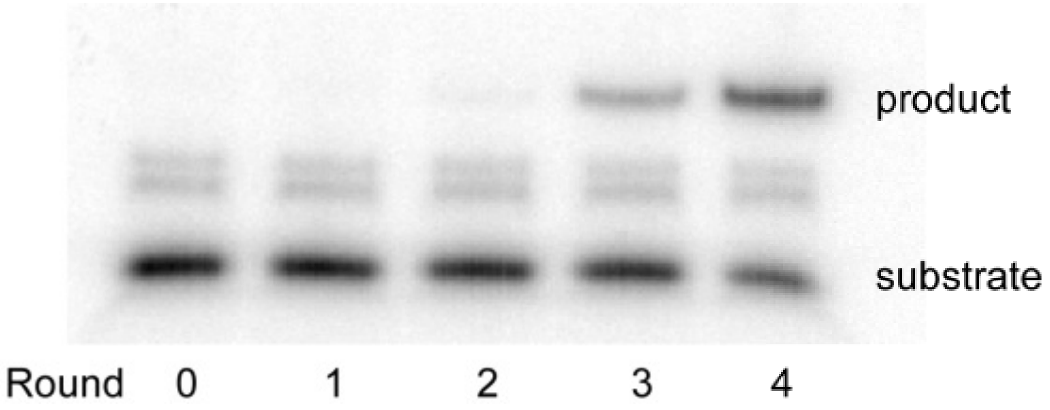


Fluorescent beads identified by FACS

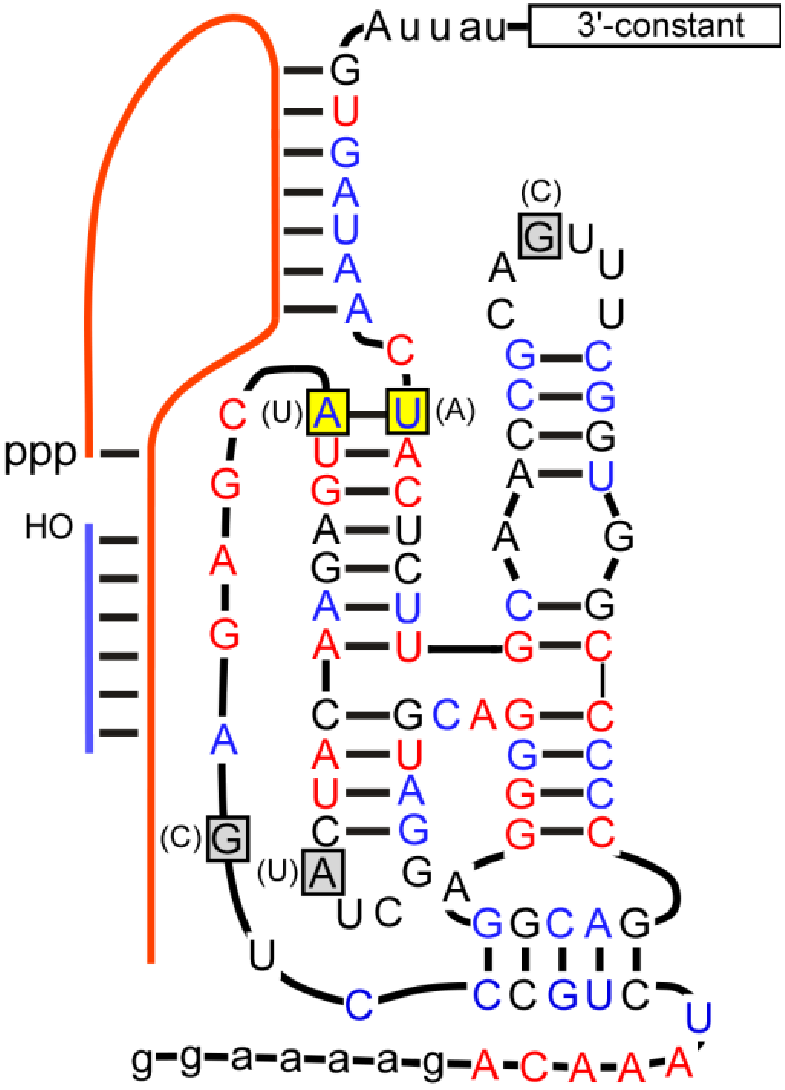
Progress of the Selection



round	genes/bead	reaction time	# of beads sorted	# of sequences sorted	# of events collected
1	10	20hr	7.2×10^{7}	7.2×10^{8}	5.7×10^{5}
2	1	20hr	5.9×10^{7}	5.9×10^{7}	2.6×10^{4}
3	0.33	2hr	5.1×10^{7}	1.7×10^{7}	2.3×10^{4}
4	0.1	2hr	6.3×10^{7}	6.3×10^{6}	1.7×10^{5}



Selected ligase



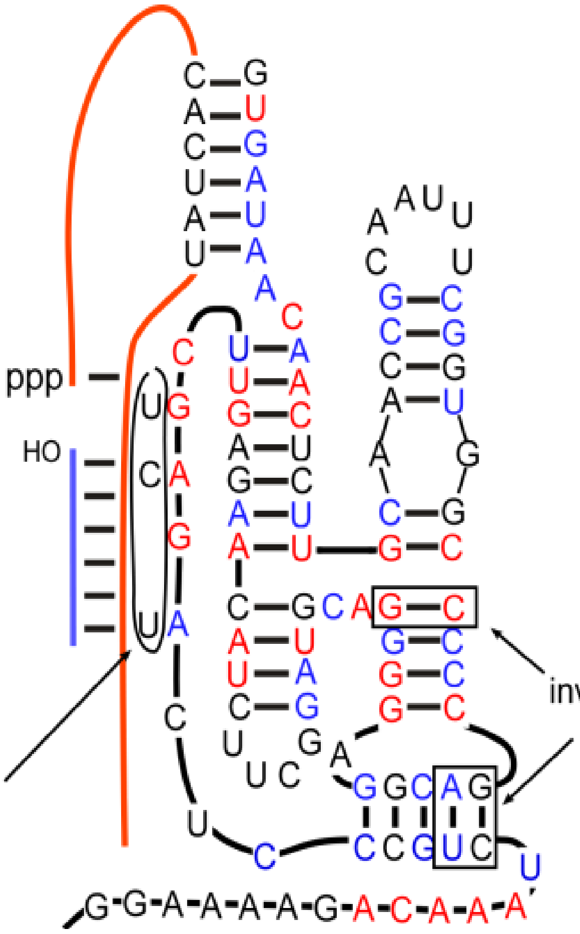
Selection statistics

- Pool size was ~7×10^8
- Wildtype ribozyme was expected ~1 in 6×10^{13}
- Mutant with any 5 mutations is expected ~1 in 1.4×10^8
- Upper estimate assumes for the selected ribozyme
    - 22 invariant and semi-conserved residues in single stranded regions
    - 26 basepairs conserved base pairs (includes wobbles)

$$0.7^{22} \times 0.6^{26} = 6.7 \times 10^{-10}$$

$$\sim 1 \text{ in } 1.5 \times 10^{9}$$

Other reselection studies also reveal the optimality of the Bartel ligase



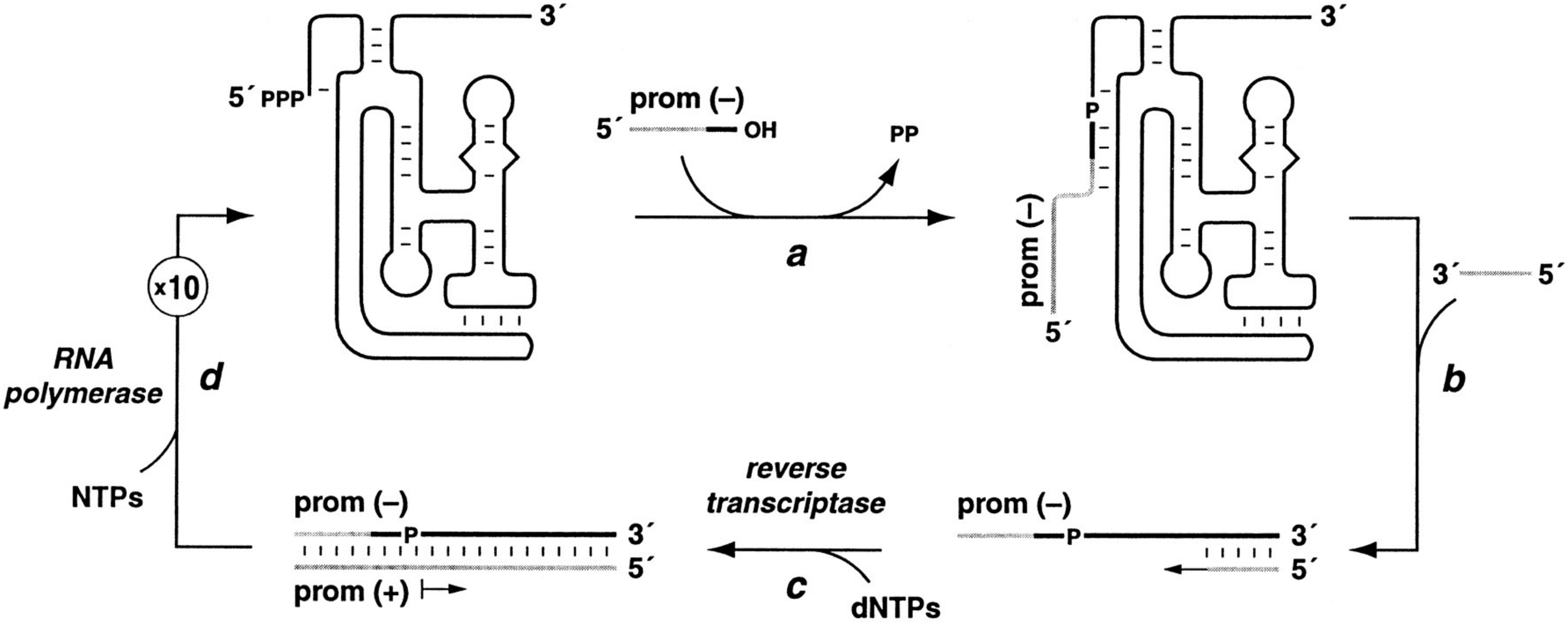
inverted, Kuhne and Joyce (2003)

“Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH”

Vaish et al. (2003)

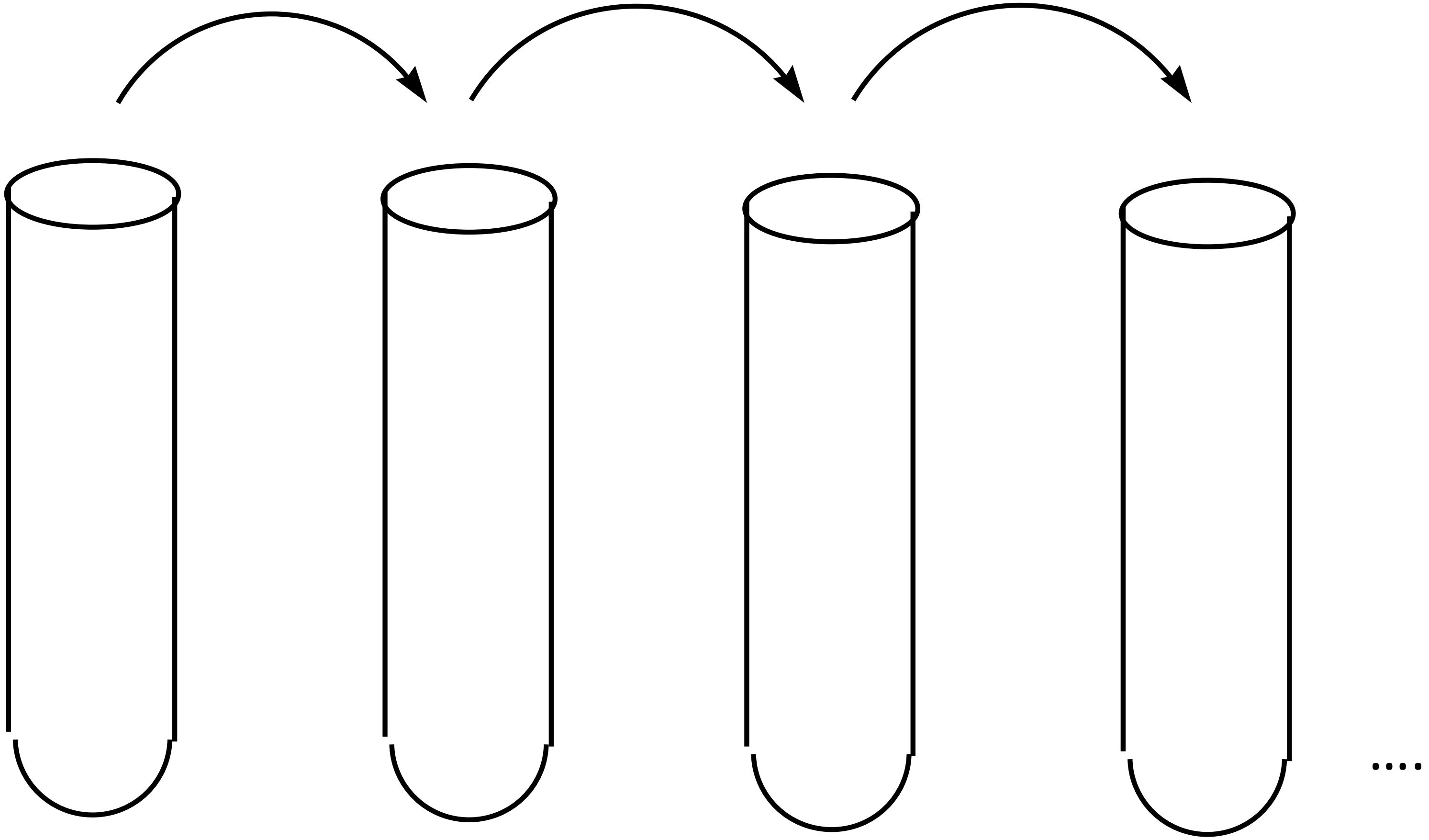
“Zeptomole detection of a viral nucleic acid using a target-activated ribozyme”

The Bartel ligase can even be adapted to function in a manner <u>similar to</u> a living organism

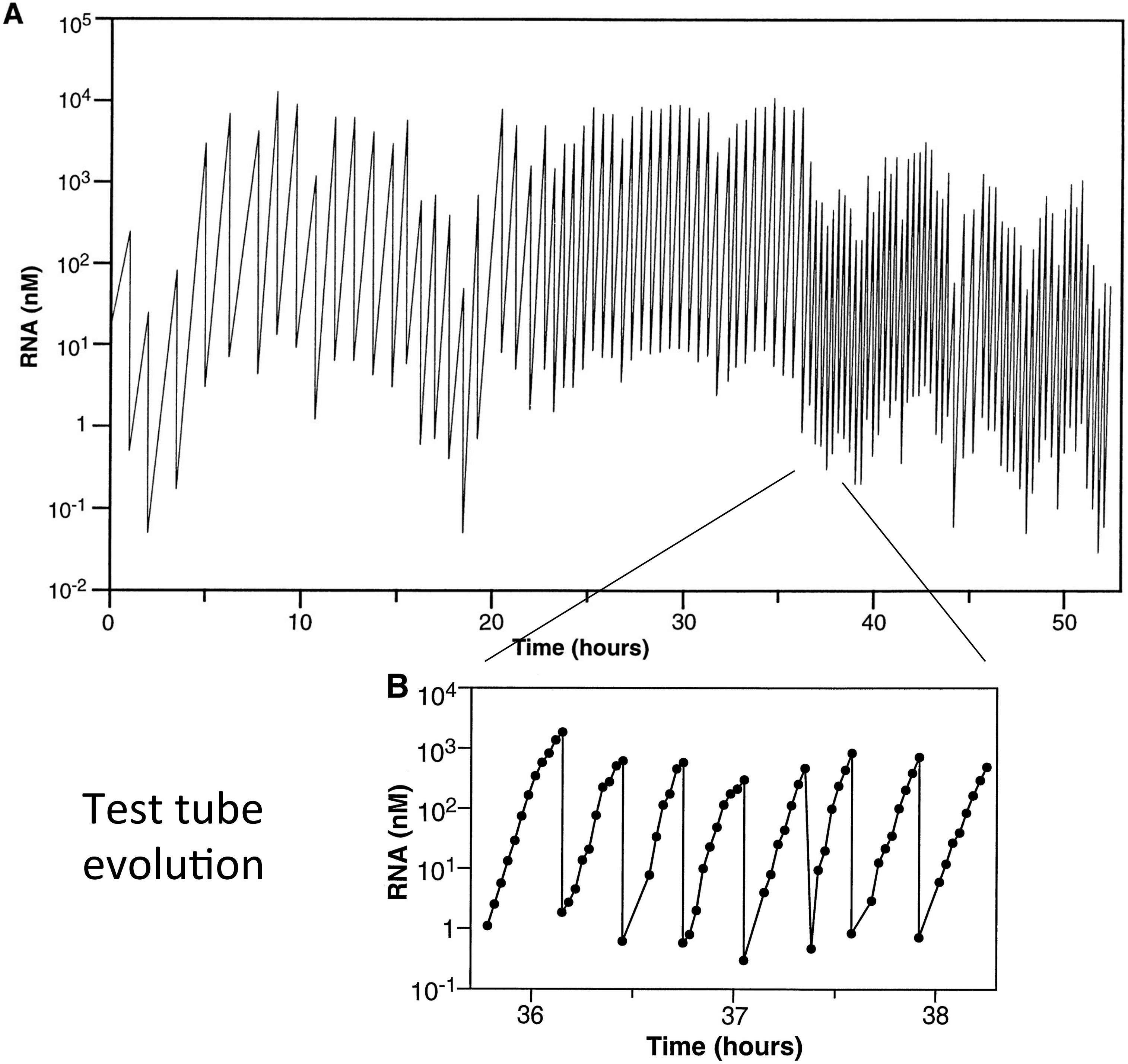


Wright and Joyce (1997), Science 276:614

Serial transfer of molecular ‘life’

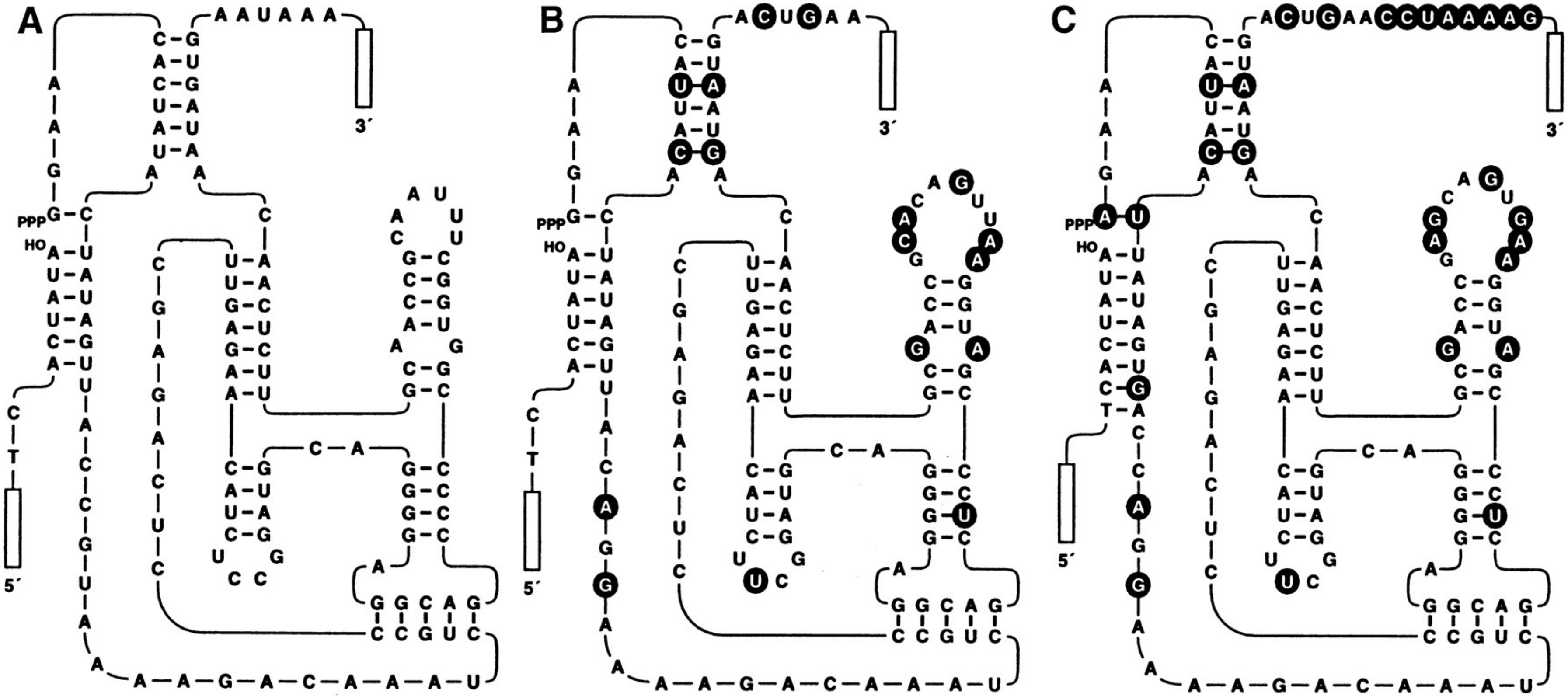


Food = RNA polymerase, reverse transcriptase, nucleotides

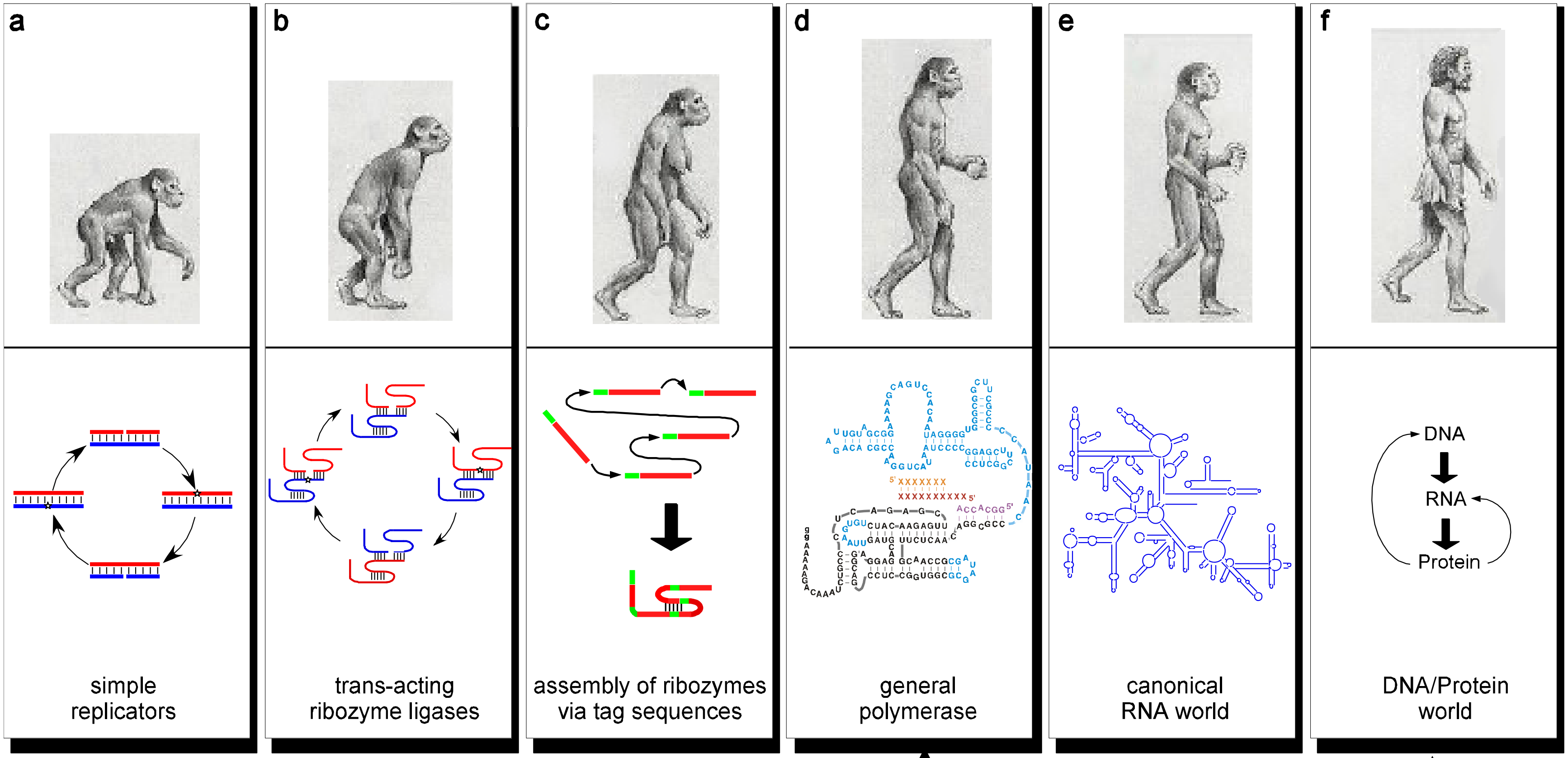


Test tube evolution

Mutational improvement



In some ways, what we are trying to do is show that it is pssible to demonstrate a path for early life. We can never fully recapitulate it.



simple replicators

trans-acting ribozyme ligases

assembly of ribozymes via tag sequences

general polymerase

canonical RNA world

DNA/Protein world

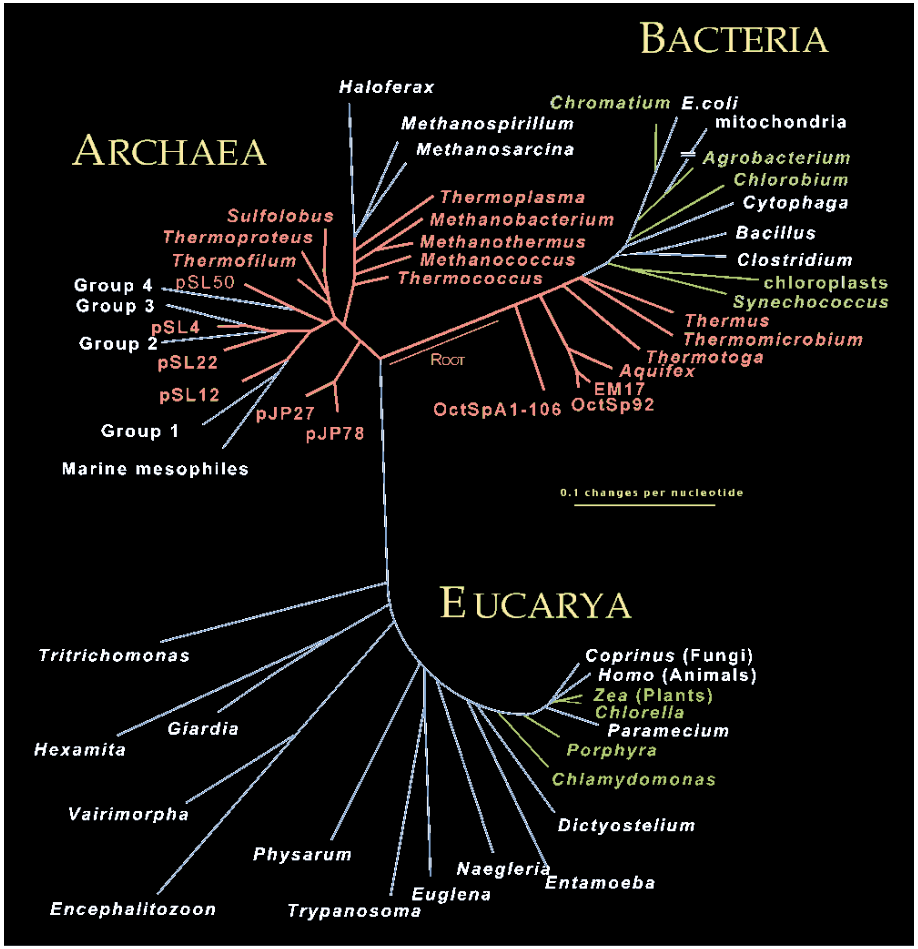
You are here

How do you get Here?

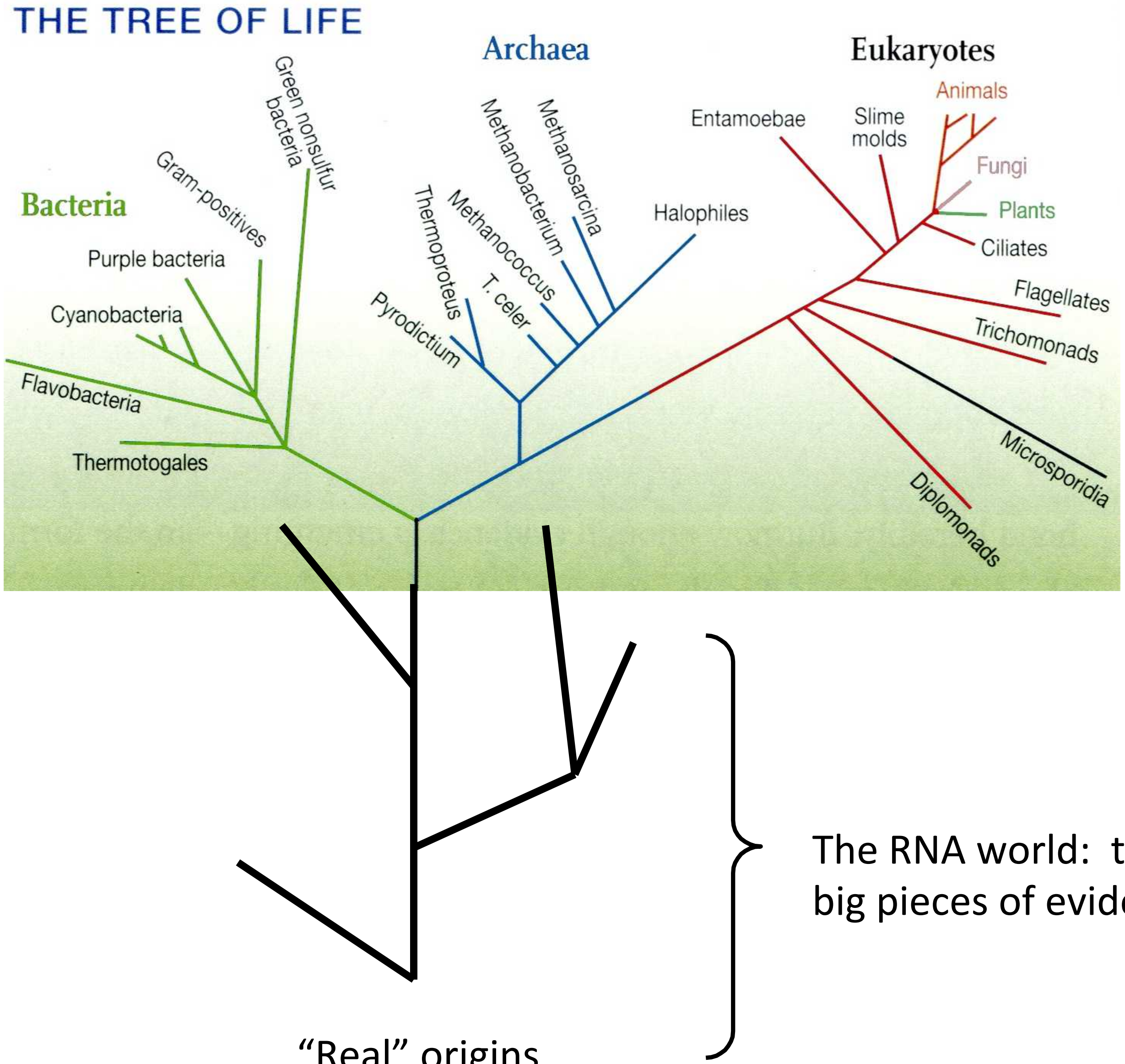
- Nucleic acid imperialism
- Doppelgangers
- Looking backwards

Origins is different from ‘evolution’ only in that the gaps are larger, not qualitatively different. We have virtually no record of the major molecular events that occurred, and must infer what is possible.

The last common ancestor is as removed from origins as we are from the LCA; maybe more so.



Examining life by looking backwards ….

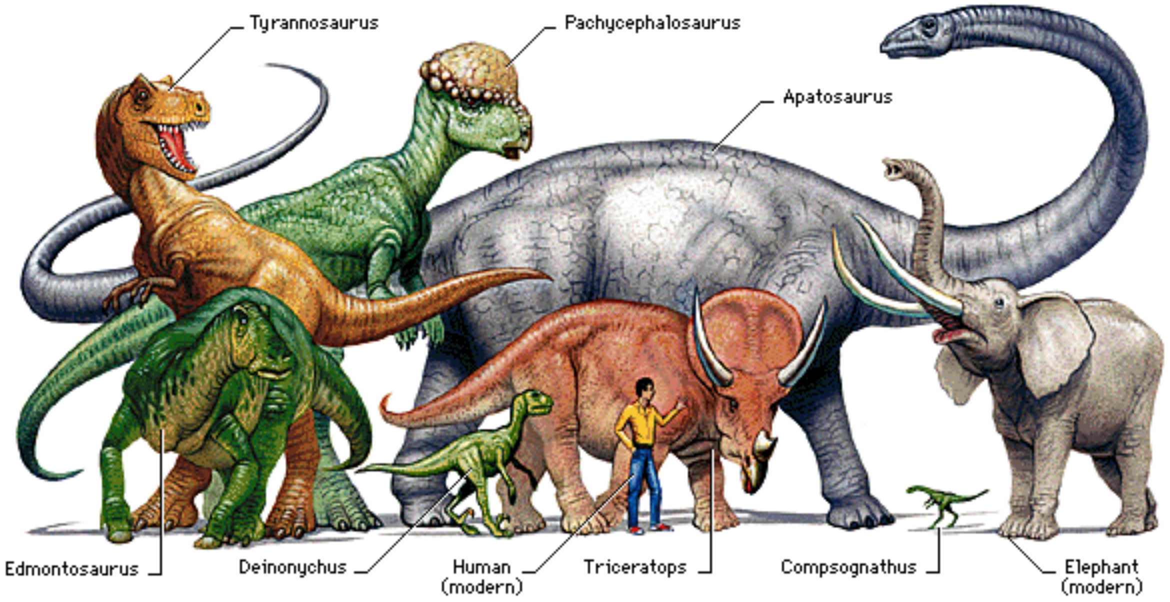


The RNA world: two big pieces of evidence

"Real" origins

This analysis in turn brings up the quite salient question:

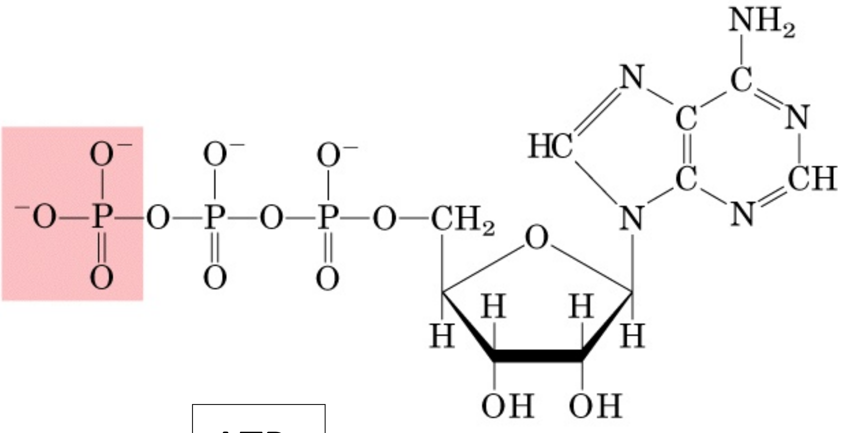
What color were the dinosaurs?



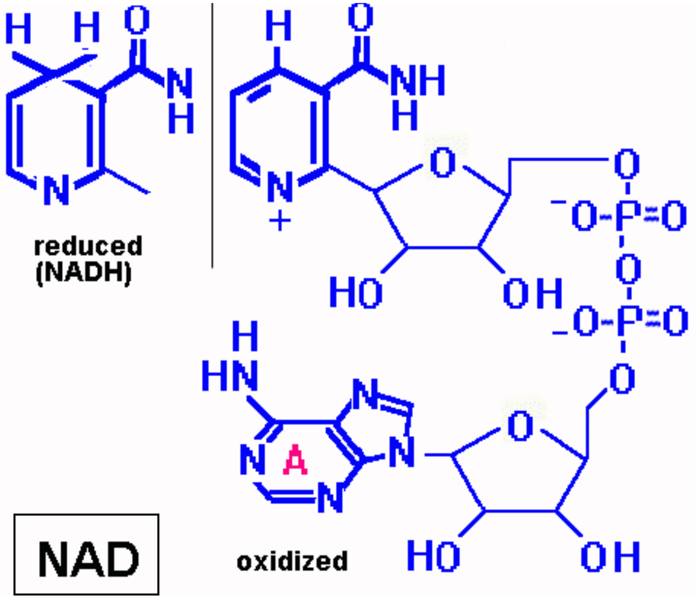
Minority view

There are only so many things we can know about origins. It is useful to remember that agnosticism is an intellectually viable position.

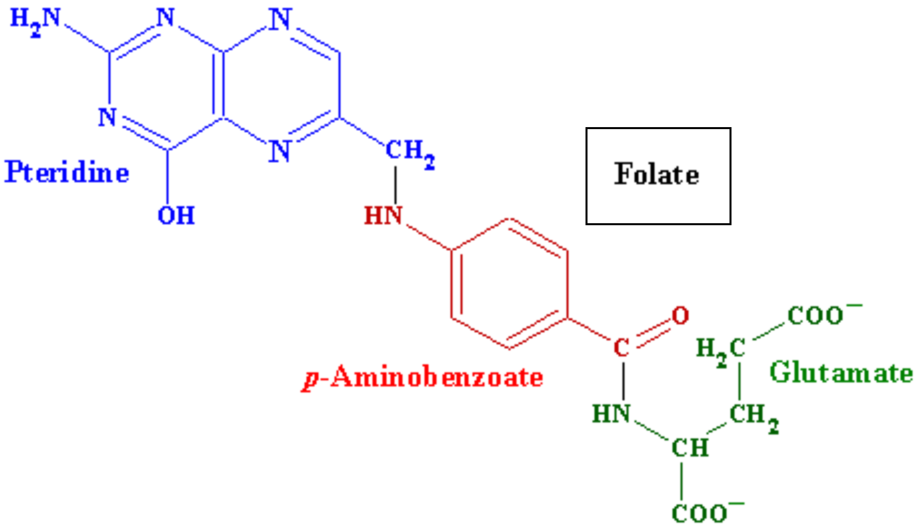
Most cofactors are nucleotide-based



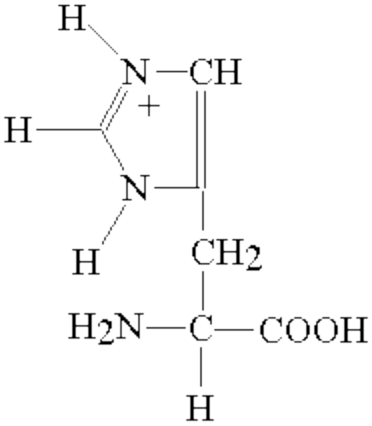
ATP



NAD



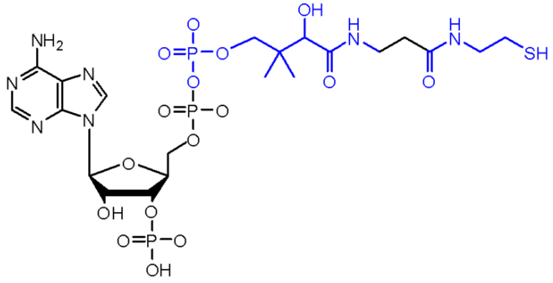
Folate



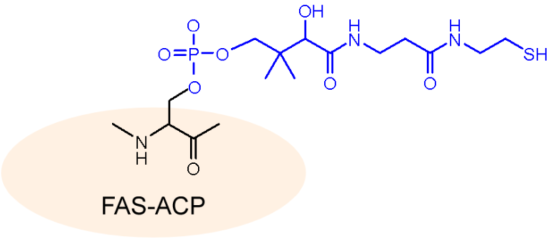
histidine

The remnants of nucleic acid ‘life’ can still be found in modern metabolism

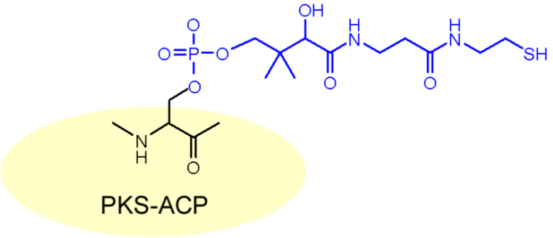
Coenzyme A



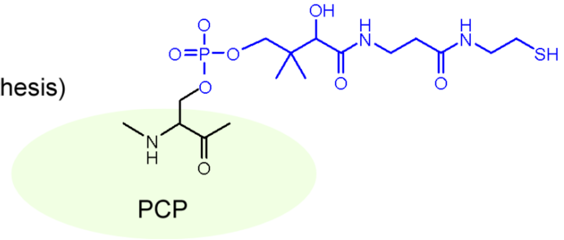
Acyl Carrier Protein
(Fatty Acid Synthesis)



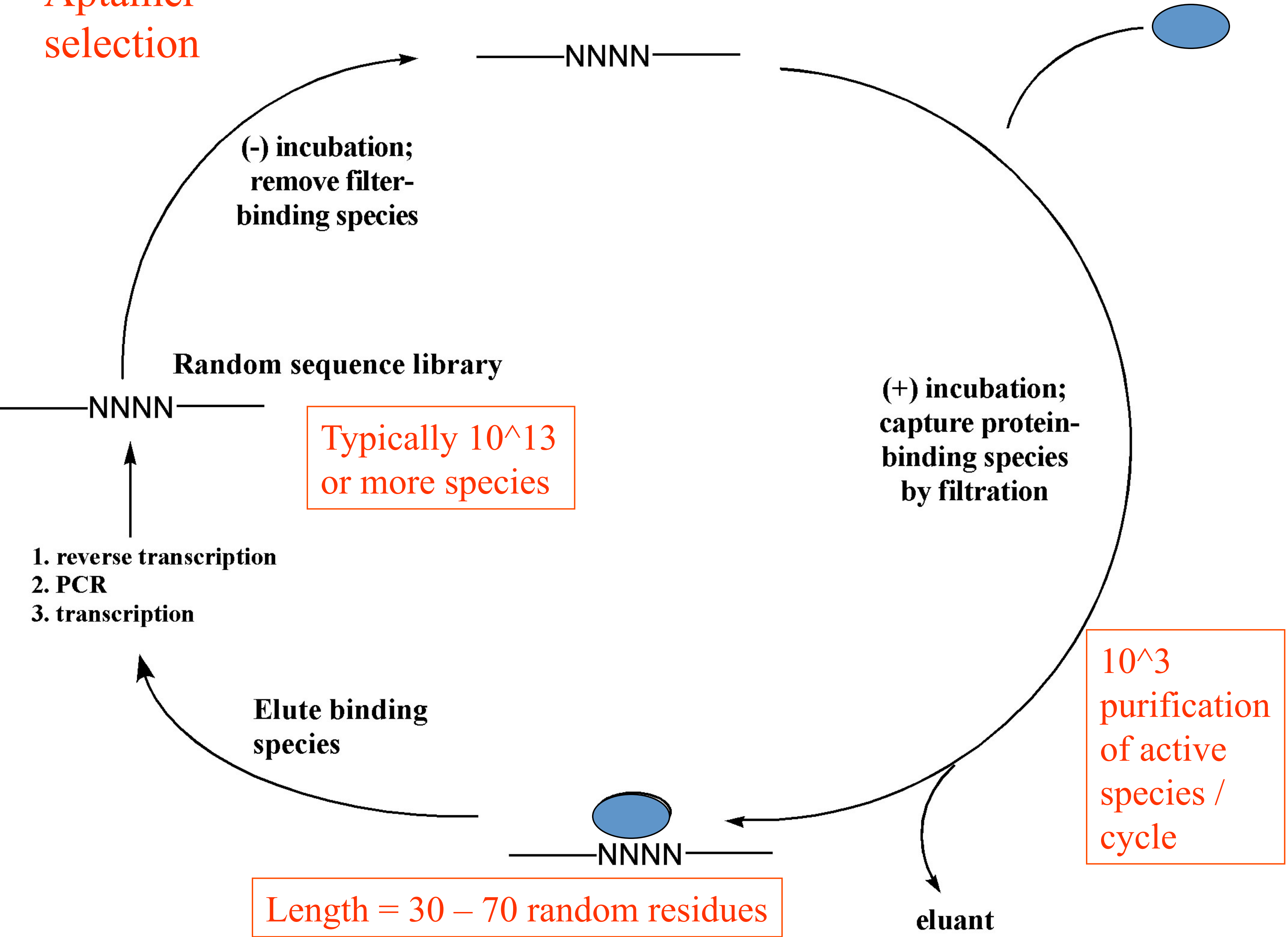
Acyl Carrier Protein
(Polyketide Synthesis)



Peptidyl Carrier Protein
(Nonribosomal peptide Synthesis)



Aptamer selection



Random sequence library

Typically 10^13 or more species

(-) incubation; remove filter-binding species

(+) incubation; capture protein-binding species by filtration

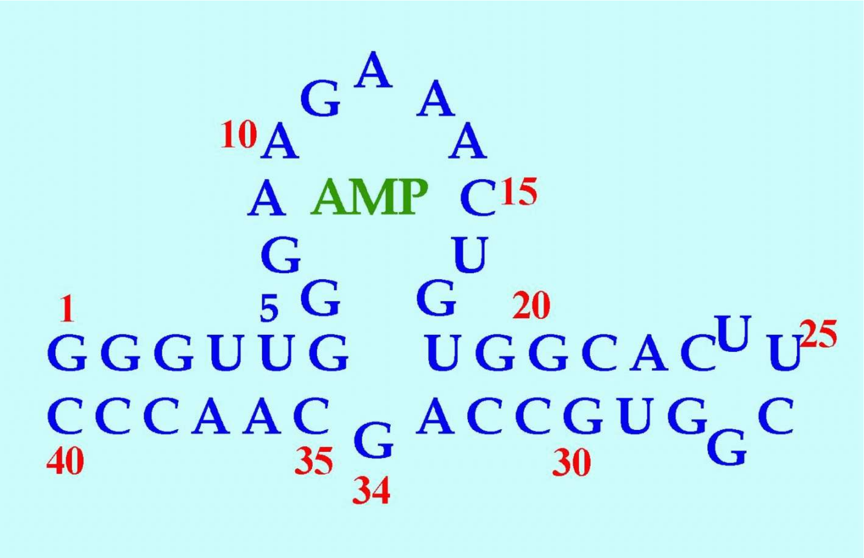
10^3 purification of active species / cycle

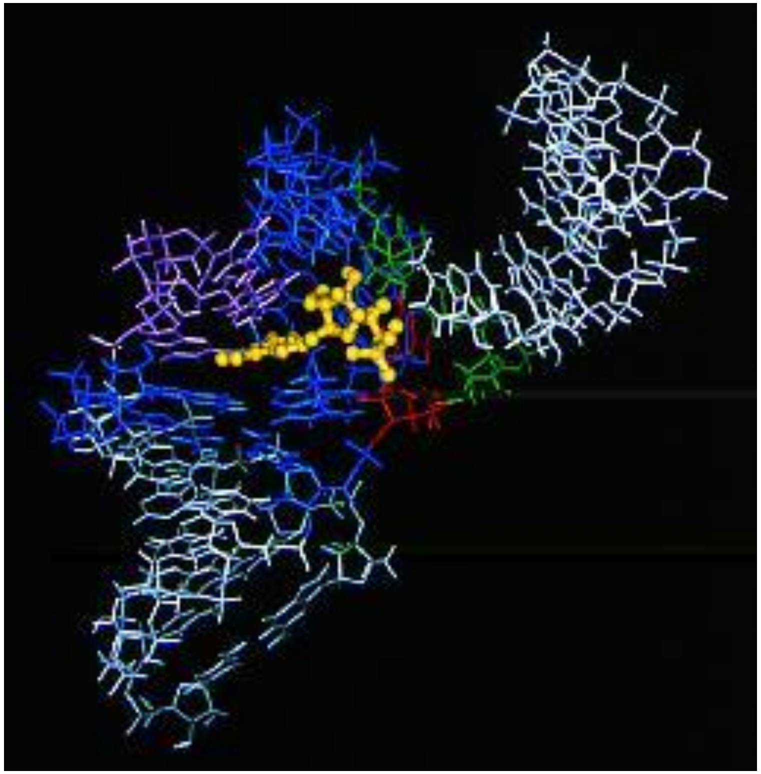
Elute binding species

1. reverse transcription
2. PCR
3. transcription

Length = 30 – 70 random residues

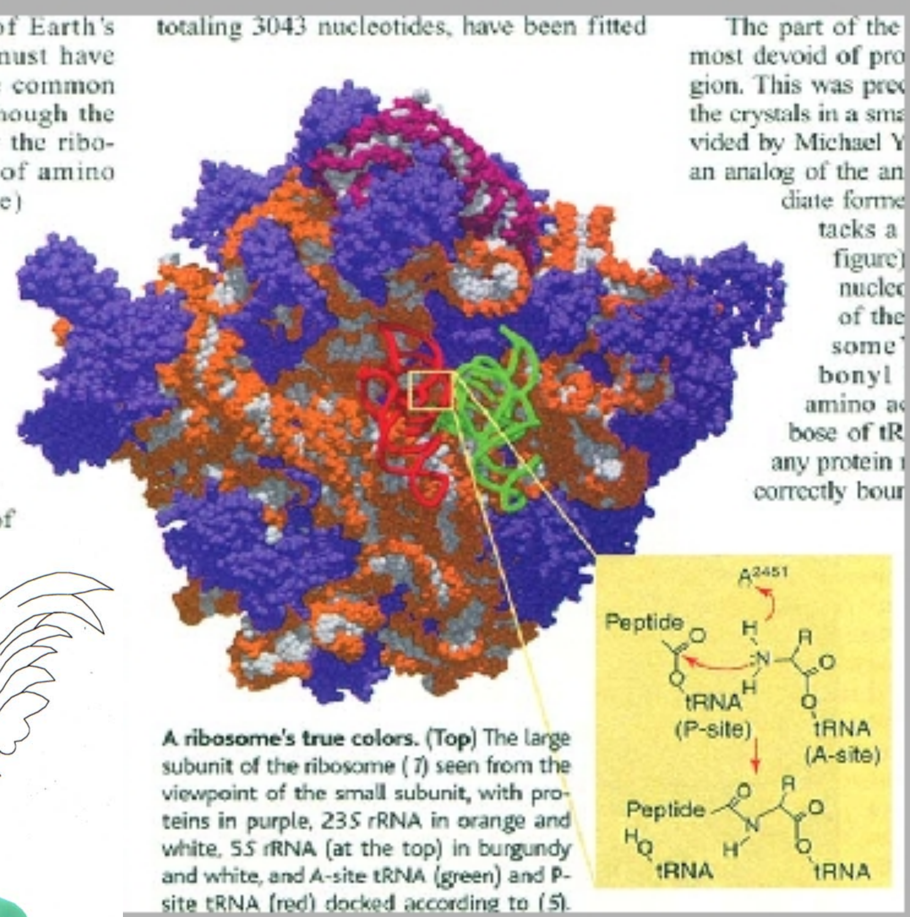
Nucleic acids can indeed form structured binding pockets, and could have interacted with 'ribo-cofactors'





The large, incredibly complex machine that makes proteins in a cell is known as the ribosome. The ribosome is composed of both proteins and RNA molecules, leading to a ‘chicken and egg’ problem. However, structural evidence reveals that:

The Ribosome is a Ribozyme !



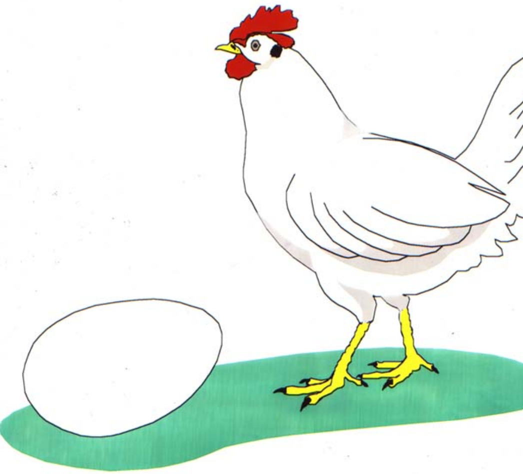
of Earth's
must have
he common
lthough the
ny the ribo-
g of amino
ide)

f
of

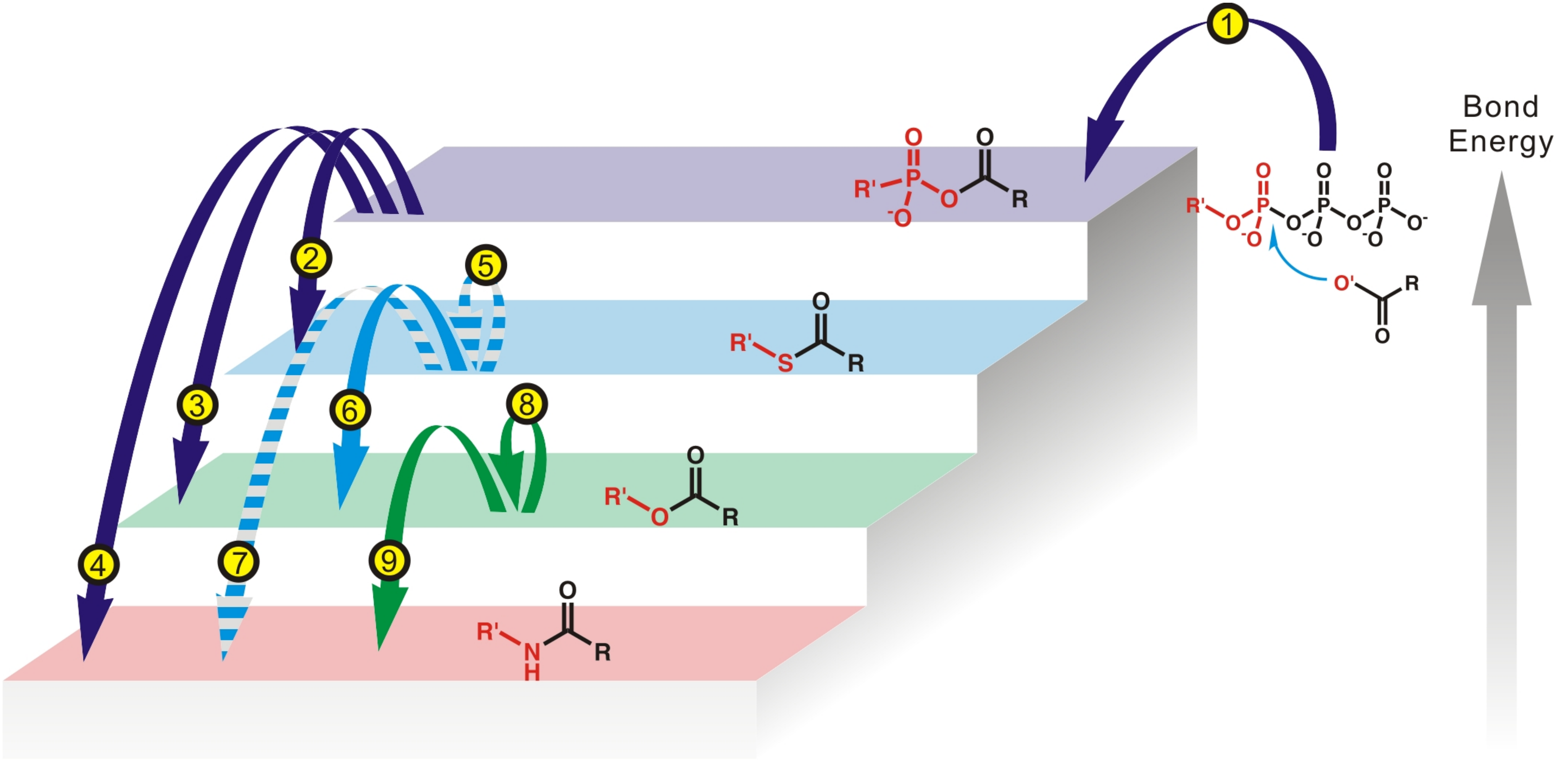
totaling 3043 nucleotides, have been fitted

The part of the
most devoid of pro
gion. This was prec
the crystals in a sma
vided by Michael Y
an analog of the an
diate forme
tacks a
figure)
nucle
of the
some
bonyl
amino a
bose of tR
any protein
correctly bour

A ribosome's true colors. (Top) The large subunit of the ribosome (1) seen from the viewpoint of the small subunit, with proteins in purple, 23S rRNA in orange and white, 5S rRNA (at the top) in burgundy and white, and A-site tRNA (green) and P-site tRNA (red) docked according to (5).



And almost all of the reactions leading to protein translation have proven to be amenable to ribozyme selections



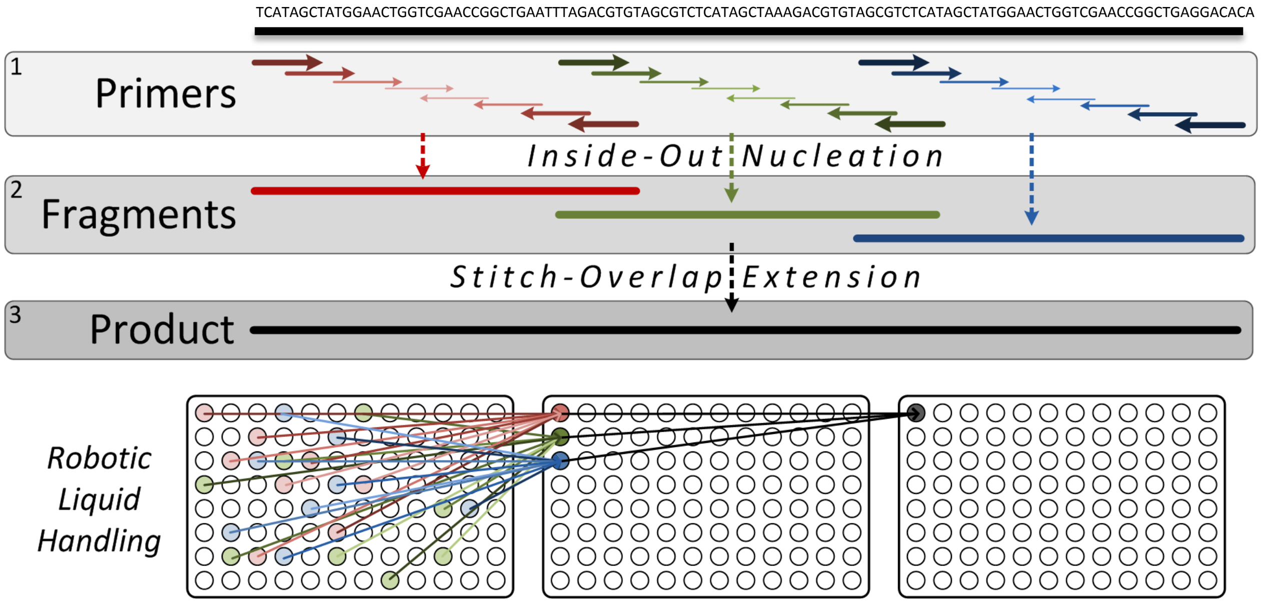
Fun and profit with synthetic biology

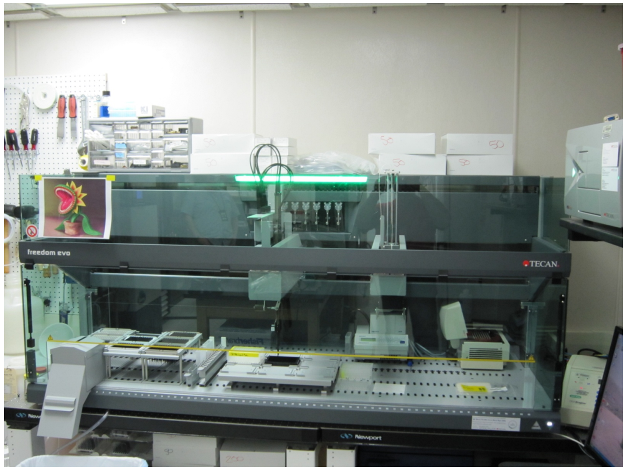
Gene Fabrication

Break down target sequences into overlaps; PCR assembly in two steps

Oligonucleotide databasing enables efficient manufacture of variants

100x 1 kb / week

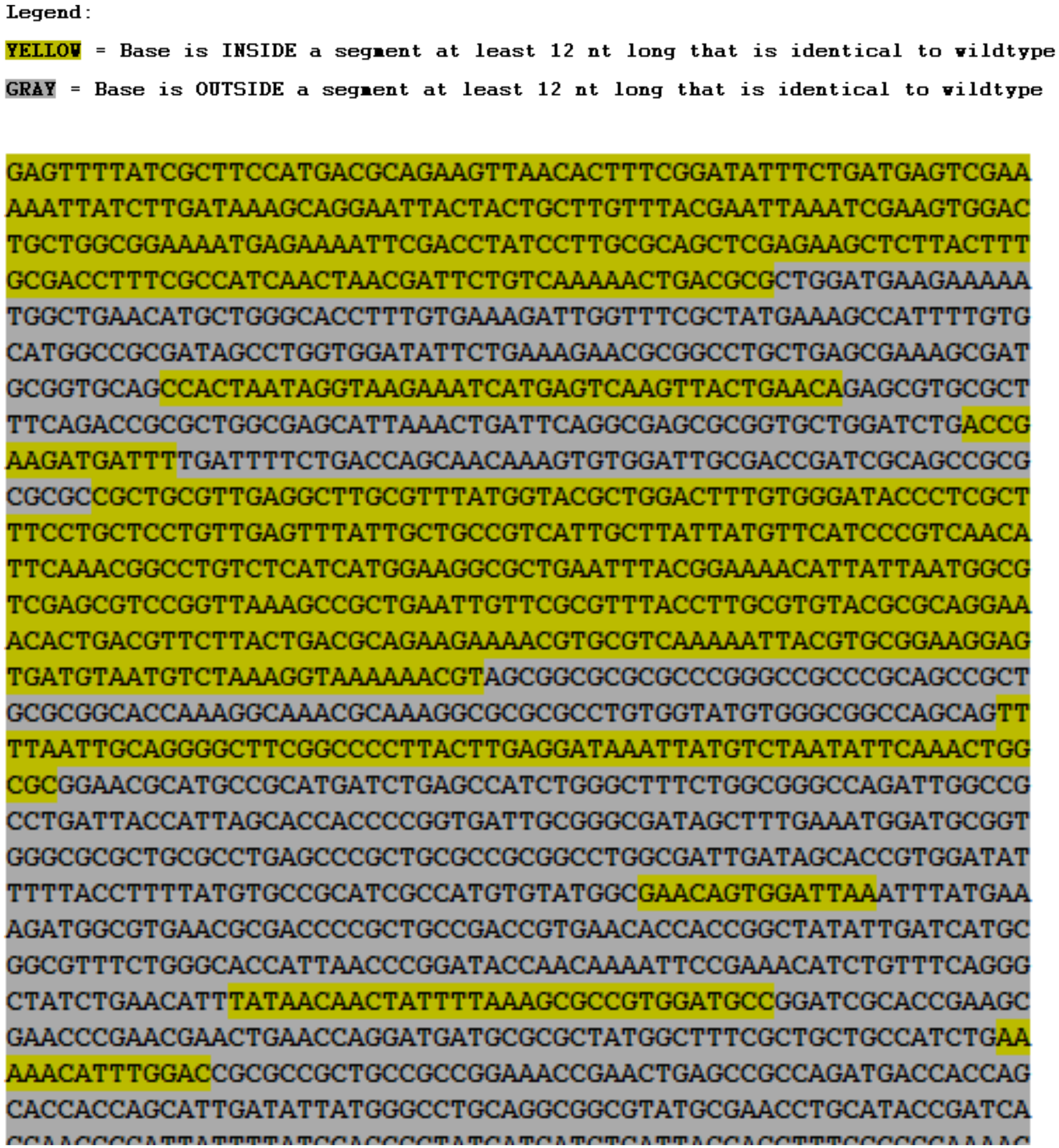




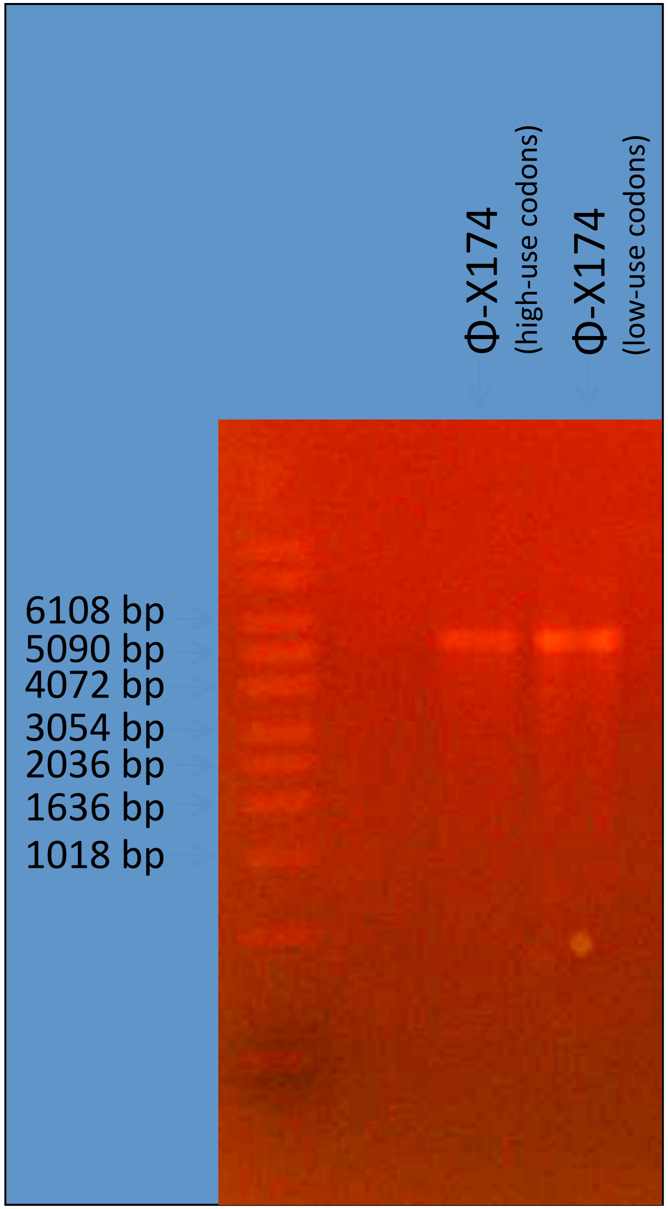

Gene fabrication facility (recently declassified)

Design of synthetic schemes, oligonucleotide synthesis and databasing, and generation of robotic operations scripts are all automated in custom software.

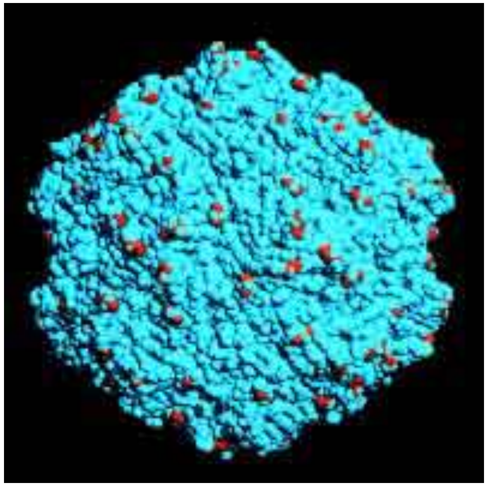
Effect of codon usage on viral fitness



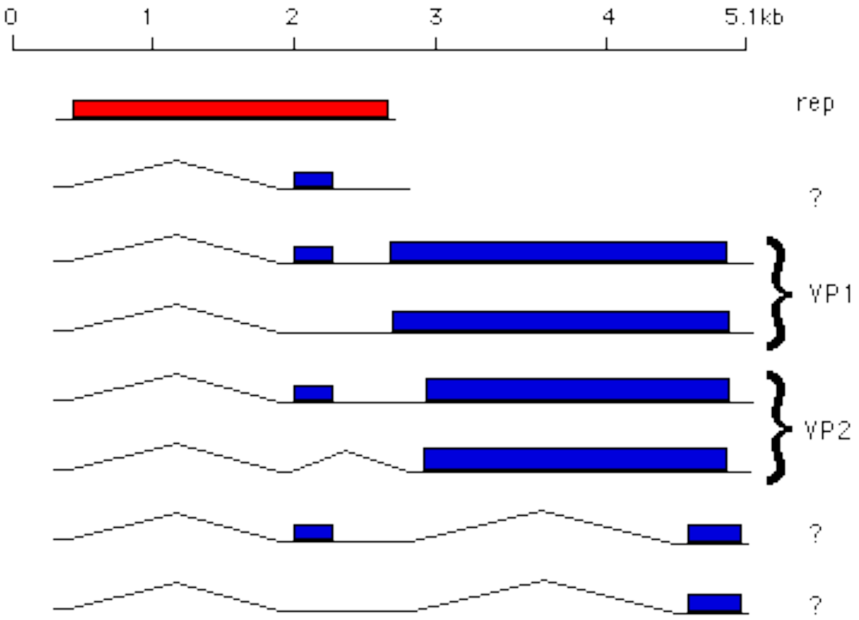
```
Legend:

YELLOW = Base is INSIDE a segment at least 12 nt long that is identical to wildtype

GRAY = Base is OUTSIDE a segment at least 12 nt long that is identical to wildtype

GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAA
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTTACGAATTAAATCGAAGTGGAC
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGCGCAGCTCGAGAAGCTCTTACTTT
GCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAACTGACGCGCTGGATGAAGAAAAA
TGGCTGAACATGCTGGGCACCTTTGTGAAAGATTGGTTTCGCTATGAAAGCCATTTTGTG
CATGGCCGCGATAGCCTGGTGGATATTCTGAAAGAACGCGGCCTGCTGAGCGAAAGCGAT
GCGGTGCAGCCACTAATAGGTAAGAAATCATGAGTCAAGTTACTGAACAGAGCGTGCGCT
TTCAGACCGCGCTGGCGAGCATTAAACTGATTCAGGCGAGCGCGGTGCTGGATCTGACCG
AAGATGATTTTGATTTTCTGACCAGCAACAAAGTGTGGATTGCGACCGATCGCAGCCGCG
CGCGCCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTGGGATACCCTCGCT
TTCCTGCTCCTGTTGAGTTTATTGCTGCCGTCATTGCTTATTATGTTCATCCCGTCAACA
TTCAAACGGCCTGTCTCATCATGGAAGGCGCTGAATTTACGGAAAACATTATTAATGGCG
TCGAGCGTCCGGTTAAAGCCGCTGAATTGTTCGCGTTTACCTTGCGTGTACGCGCAGGAA
ACACTGACGTTCTTACTGACGCAGAAGAAAACGTGCGTCAAAAATTACGTGCGGAAGGAG
TGATGTAATGTCTAAAGGTAAAAAACGTAGCGGCGCGCGCCCGGGCCGCCCGCAGCCGCT
GCGCGGCACCAAAGGCAAACGCAAAGGCGCGCGCCTGTGGTATGTGGGCGGCCAGCAGTT
TTAATTGCAGGGGCTTCGGCCCCTTACTTGAGGATAAATTATGTCTAATATTCAAACTGG
CGCGGAACGCATGCCGCATGATCTGAGCCATCTGGGCTTTCTGGCGGGCCAGATTGGCCG
CCTGATTACCATTAGCACCACCCCGGTGATTGCGGGCGATAGCTTTGAAATGGATGCGGT
GGGCGCGCTGCGCCTGAGCCCGCTGCGCCGCGGCCTGGCGATTGATAGCACCGTGGATAT
TTTTACCTTTTATGTGCCGCATCGCCATGTGTATGGCGAACAGTGGATTAAATTTATGAA
AGATGGCGTGAACGCGACCCCGCTGCCGACCGTGAACACCACCGGCTATATTGATCATGC
GGCGTTTCTGGGCACCATTAACCCGGATACCAACAAAATTCCGAAACATCTGTTTCAGGG
CTATCTGAACATTTATAACAACTATTTTAAAGCGCCGTGGATGCCGGATCGCACCGAAGC
GAACCCGAACGAACTGAACCAGGATGATGCGCGCTATGGCTTTCGCTGCTGCCATCTGAA
AAACATTTGGACCGCGCCGCTGCCGCCGGAAACCGAACTGAGCCGCCAGATGACCACCAG
CACCACCAGCATTGATATTATGGGCCTGCAGGCGGCGTATGCGAACCTGCATACCGATCA
```



There are many human viruses on the same scale (such as human parvoviruses: B19; Fifth's disease)



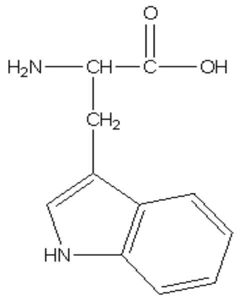
Canine parvovirus was derived from feline panleukopenia virus via a small number of point mutations in the viral capsid genes that expanded the host range to canine cells. Following its emergence in the late 1970s, canine parvovirus caused a pandemic that killed a large fraction of world's dogs



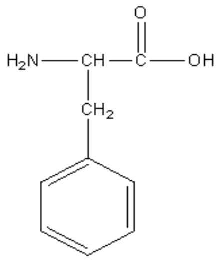
Parts Standardization via Nucleobase Amino Acids

Natural Amino Acids

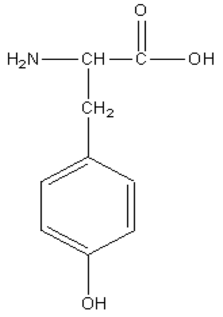
PNA-like amino acid surrogates



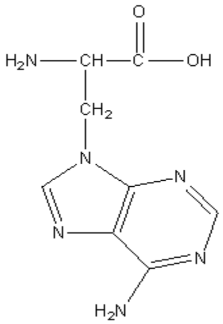
Tryptophan



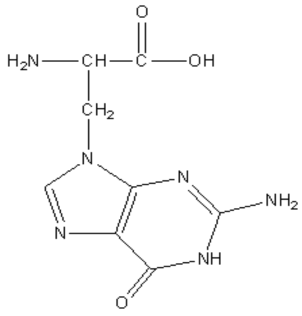
Phenylalanine



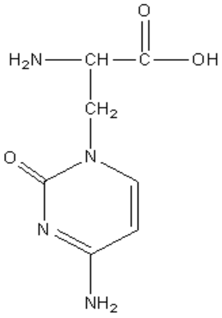
Tyrosine



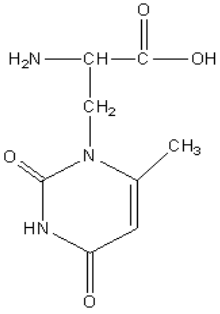
Adenylalanine



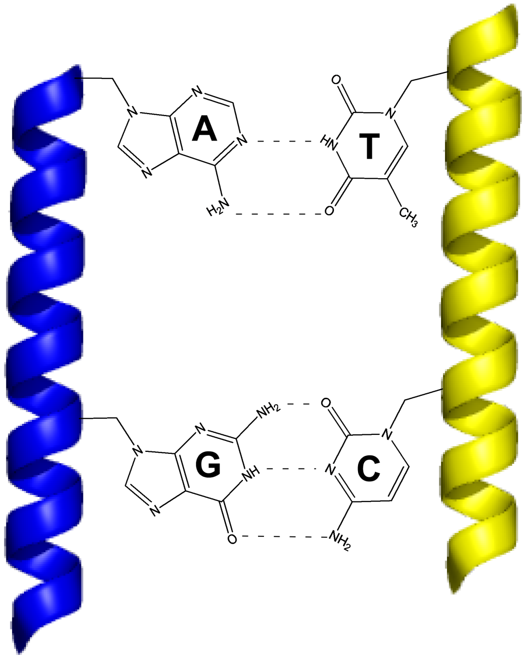
Guanylalanine



Cytidylalanine

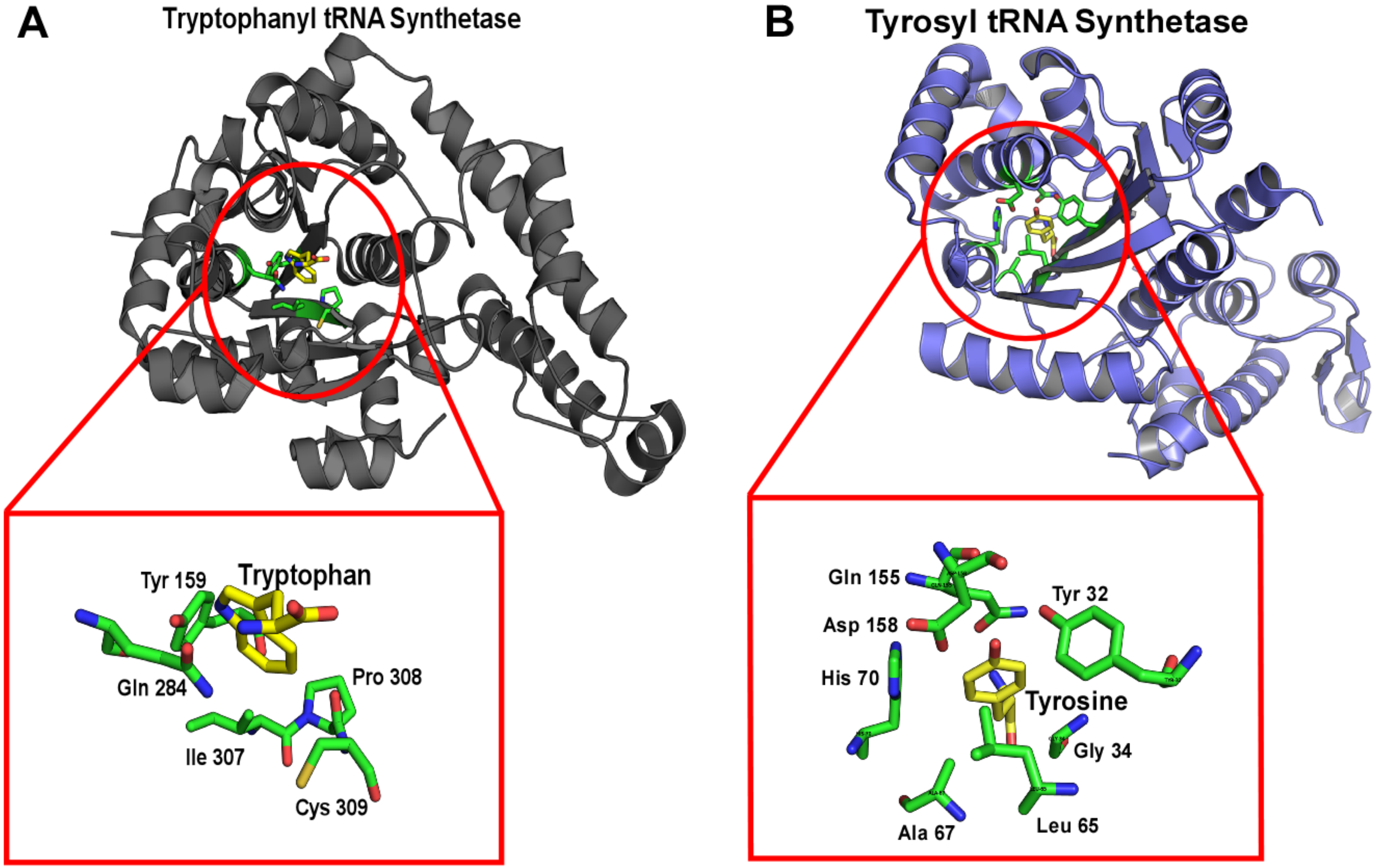


Thymidylalanine



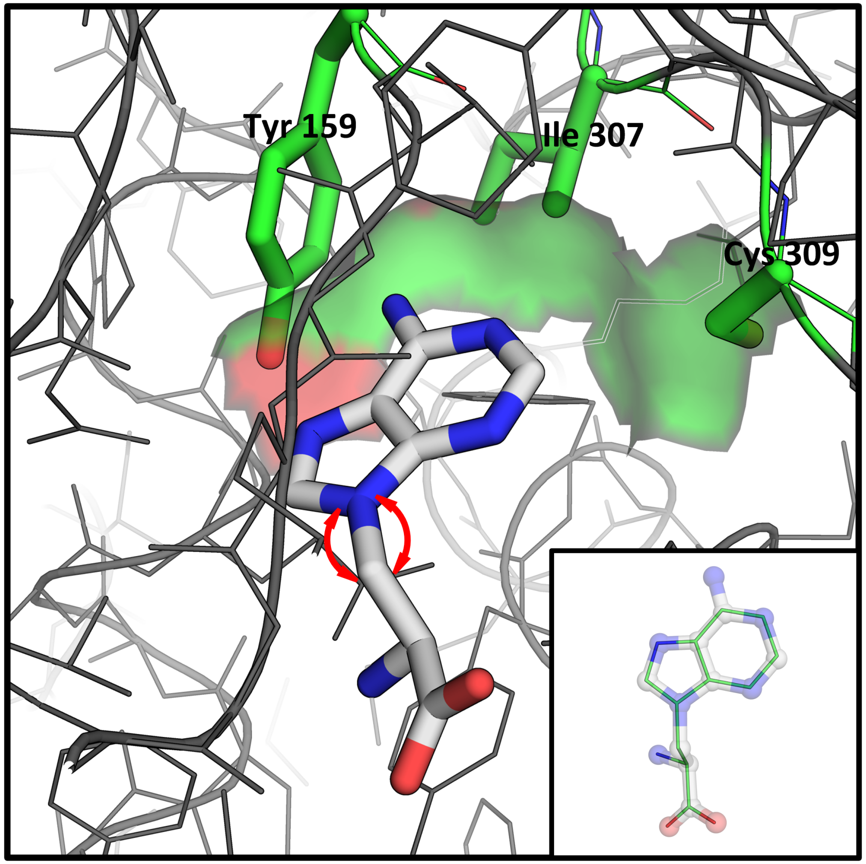
Protein ‘hybridization’

Choice of synthetase:tRNA pairs

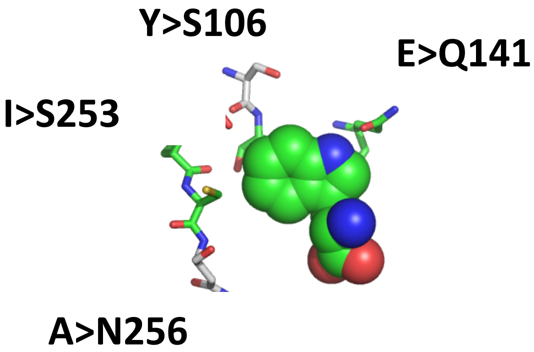


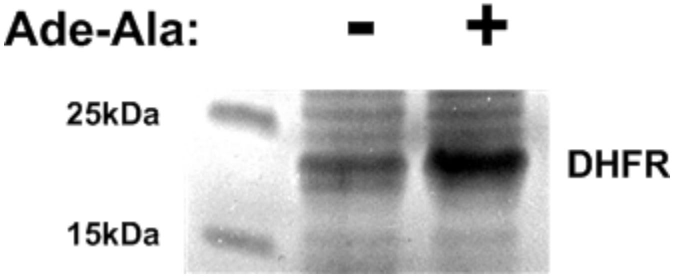
Ade-Ala RS rational designs

Adenyl alanine docked into the active site
of tryptophanyl tRNA synthetase; mutations
Introduced by rational design (eyeballing)



Raa2





Nucleobase amino acids → Proteins with nucleic acid-like properties → Programming