Prebiotic Networks: From Molecules into Cells

Niles Lehman
Department of Chemistry
Portland State University
niles@pdx.edu
LIFE = “a self-sustaining chemical system capable of darwinian evolution” (Joyce/NASA)
The Seven Challenges to a Prebiotic Chemist

1. The origin/source of the elements
2. The origin/source of small molecule precursors
3. The origin/source of monomers
4. The condensation problem
5. The (self)-replication problem
6. The chirality problem
7. The compartmentalization problem
the origin of cells
“linking genotype with phenotype”

compartmentalization would offer life enormous advantages

- keeping water concentrations low
- creating gradients
- allowing genotypes to harvest “the fruits of their labor”
the “holy grail” of the RNA World: an RNA replicase ribozyme

the Bartel/Unrau/Holliger replicase ribozyme

Attwater et al. (2013) Nature Chemistry 5, 1011–1018

a 190-nt ribozyme that can polymerize a portion of itself

molecular self-replication
the world’s record

206 nt

alterations of cold (−7°C) and normal (17°C) temperatures used to select this RNA

the tC9Y ribozyme can perform template-directed replication to elongate RNA to greater than its own length (but it can’t replicate itself)

Attwater et al. (2013) Nature Chemistry 5, 1011–1018
autocatalysis
the chemical requirement for self-replication

\[ A + B \rightarrow C \]

the product of a reaction catalyzes its own formation
from selfishness to cooperation...

```
A + B → C

A' + B' → C'
```

“selfish”

“cooperative”
...extending cooperation to >2 “selves”...
... and from simple cycles to networks

an autocatalytic set

Kauffman (1993)
recombination

recombination, at the molecular level, is the breaking and re-formation of (phosphoester) bonds resulting in the swapping of ≥ 1 monomer units between two (nucleic-acid) strands.


recombination

“easy” chemistry

polymerization

“hard” chemistry
my claim...

recombination can provide a mechanism for the initial build-up of complex catalytic RNAs

\[
\begin{align*}
2\text{-mer} + 2\text{-mer} & \rightarrow 3\text{-mer} + 1\text{-mer} \\
3\text{-mer} + 3\text{-mer} & \rightarrow 5\text{-mer} + 1\text{-mer} \\
5\text{-mer} + 5\text{-mer} & \rightarrow 9\text{-mer} + 1\text{-mer} \\
9\text{-mer} + 9\text{-mer} & \rightarrow 17\text{-mer} + 1\text{-mer}
\end{align*}
\]

our goal: devise an all-RNA system that can exploit recombination to build up genetic information into a network of self-replication

Lehman (2003)  
analogy to “sexual” reproduction

By analogy to the Fisher-Muller argument, recombination can hasten the appearance of multiple beneficial “traits” in the same “genome”
getting RNAs to recombine RNAs: group I introns do this in Nature

**step 1**

- intron (ribozyme)
- G-OH
- left exon
- right exon

**step 2**

- reverse splicing = “pick-up-the-tail” (PUTT)
- spliced exons

self-splicing of rRNA and tRNA introns *in vivo*
the Azoarcus ribozyme as a recombinase

self-splicing intron from the isoleucine tRNA of the purple bacterium Azoarcus

L–8 ribozyme is 197 nt long, and has a 71% G+C content

active up to 70°C

internal guide sequence is GUG, its complement (i.e., “tag”) is CAU
recombination scheme by group I ribozymes

RNA-directed recombination of short oligomers

*Azoarcus* ribozyme: IGS = GUG; target = CAU

<table>
<thead>
<tr>
<th>Oligomer</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNL-1a</td>
<td>GGCAU•AAAAAUAUUAAUAACAAUA</td>
<td>22-mer</td>
</tr>
<tr>
<td>SNL-2a</td>
<td>GGAAAGGCAU•AAAUA</td>
<td>15-mer</td>
</tr>
<tr>
<td>SNL-4a</td>
<td>GGCAU•GGCCGAACACAGC</td>
<td>17-mer</td>
</tr>
<tr>
<td>SNL-5a</td>
<td>GGGAGUCUGAUGAGGCAU•AAAUA</td>
<td>23-mer</td>
</tr>
</tbody>
</table>

“head” • “tail”

SNL-1a X SNL-2a: 22-mer + *15-mer → *27-mer + 10-mer
recombining the recombinase itself

AZOARCUS RIBOZYME

no full-length Azoarcus RNA was added!

Azoarcus RCL6 "binary"

time (min): 0 2 5 10 20 30 40 60 120 180 240

1 μM each RNA
55°C
50 mM MgCl₂
four-piece (quad) self-assembly

no Azoarcus!

198-nt Azoarcus ribozyme

trans-catalysis first

trans-assembly

X(37)

Y(46)

Z(51)

covalent self-assembly

self-replication

W(63)
a small “selfish” autocatalytic network

Hayden, von Kiedrowski, Lehman (2008)
Here, the dot (+) represents a covalent bond.

IGS

\[ \begin{array}{c}
5' & G & M & N & 3' \\
\cdot & U & M' & N' & 5'
\end{array} \]

Tag

(invert Figure)

Tag

W X → W·X

W X Y Z

W X Y·Z

W·X Y Z

W X·Y Z

W X Y·Z

G C A U C G

3' OH

5'
a putative cooperative cycle
replicator yield is highest when all three components are present

only W in cycle $l_1$ is radiolabeled

l_1 + l_2 + l_3 = $H_1$

l_1 alone

“closed” reaction

% yield WXYZ (l_1)
a competitive advantage to cooperation

the cooperative cycle out-competes the selfish replicators...

mismatched guides & tags

matched guides & tags

... but only when in mixed in the same population

a mechanism by which networks “assimilate” autocatalysts?

inequality in rate constants for the subsystems (arrow thickness) leads to time lags
mathematical modeling supports empirical data

(Michael Manapat / Irene Chen)
moving beyond this single example: randomization experiment
randomization experiment

48 possible genotypes
(4 IGS choices x 4 IGS tag choices x 3 junctions)

e.g., C|U|x
randomization experiment

100 mM MgCl$_2$
48°C

200 pmol each (10$^{14}$ molecules):

- GNGWcn’U
- GNGWXcn’U
- GNGWXYcn’U
- hXYZ
- hYZ
- hZ

30 minutes

RT-PCR full-length WXYZ ribozymes

2 hours

RT-PCR full-length WXYZ ribozymes

4 hours

RT-PCR full-length WXYZ ribozymes

8 hours

RT-PCR full-length WXYZ ribozymes

high-throughput nucleotide sequence analysis (Illumina)

~3 million genotypes

~3 million genotypes

~3 million genotypes

~3 million genotypes
summarized results

global visualization

red: autocatalysts

green: “cooperators”

orange: both members of 2MCs increasing over time

thick green: $UG_x + AA_y + CU_z$

at 8 hours
serial transfer experiments:
emulating a steady-state flow reactor
what matters to prebiotic networks?

1. viable cores (clusters)
2. connectivity kinetics (who’s connected to whom)
3. information control (negative feedback)
4. scalability (scale-free networks)
5. resource availability (food supply)
6. compartmentalization (barriers to free flow)

Yields of **WXYZ** RNA are 10–20% higher in artificial water-in-oil (10 fL – 10 nL) droplets

we can test, for example, the Stochastic Corrector Model

with Philippe Nghe & Andrew Griffiths, ESPCI ParisTech
cooperate ... then be selfish!

cooperate ... then be selfish!

acknowledgements

Dr. Nilesh Vaidya (Portland State, now Princeton)
Dr. Eric Hayden (Portland State, now Boise State)
Dr. Prof. Günter von Kiedrowski (Bochum)
Dr. Irene Chen (Harvard, now UCSB)
Dr. Michael Manapat (Harvard, now Google)
Ms. Jessica Yeates (PSU Ph.D. student)

“NilesH”
autocatalytic rate constants ($k_a$, min$^{-1}$) for the 16 $\text{WXY}$ genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>$k_a$ (min$^{-1}$)</th>
<th>Std. error</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>0.0415</td>
<td>0.0066</td>
<td>0.98</td>
</tr>
<tr>
<td>AU</td>
<td>0.0319</td>
<td>0.0011</td>
<td>1.00</td>
</tr>
<tr>
<td>UA</td>
<td>0.0197</td>
<td>0.0004</td>
<td>1.00</td>
</tr>
<tr>
<td>GC</td>
<td>0.0125</td>
<td>0.0021</td>
<td>0.97</td>
</tr>
<tr>
<td>GU</td>
<td>0.0091</td>
<td>0.0007</td>
<td>0.99</td>
</tr>
<tr>
<td>AC</td>
<td>0.0069</td>
<td>0.0002</td>
<td>1.00</td>
</tr>
<tr>
<td>UG</td>
<td>0.0049</td>
<td>0.0004</td>
<td>0.99</td>
</tr>
<tr>
<td>UC</td>
<td>0.0038</td>
<td>0.0002</td>
<td>1.00</td>
</tr>
<tr>
<td>UU</td>
<td>0.0022</td>
<td>0.0001</td>
<td>1.00</td>
</tr>
<tr>
<td>CA</td>
<td>0.0020</td>
<td>0.0000</td>
<td>1.00</td>
</tr>
<tr>
<td>CC</td>
<td>0.0016</td>
<td>0.0001</td>
<td>1.00</td>
</tr>
<tr>
<td>GG</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.99</td>
</tr>
<tr>
<td>GA</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.98</td>
</tr>
<tr>
<td>AA</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.92</td>
</tr>
<tr>
<td>CU</td>
<td>0.0004</td>
<td>0.0000</td>
<td>1.00</td>
</tr>
<tr>
<td>AG</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.99</td>
</tr>
</tbody>
</table>