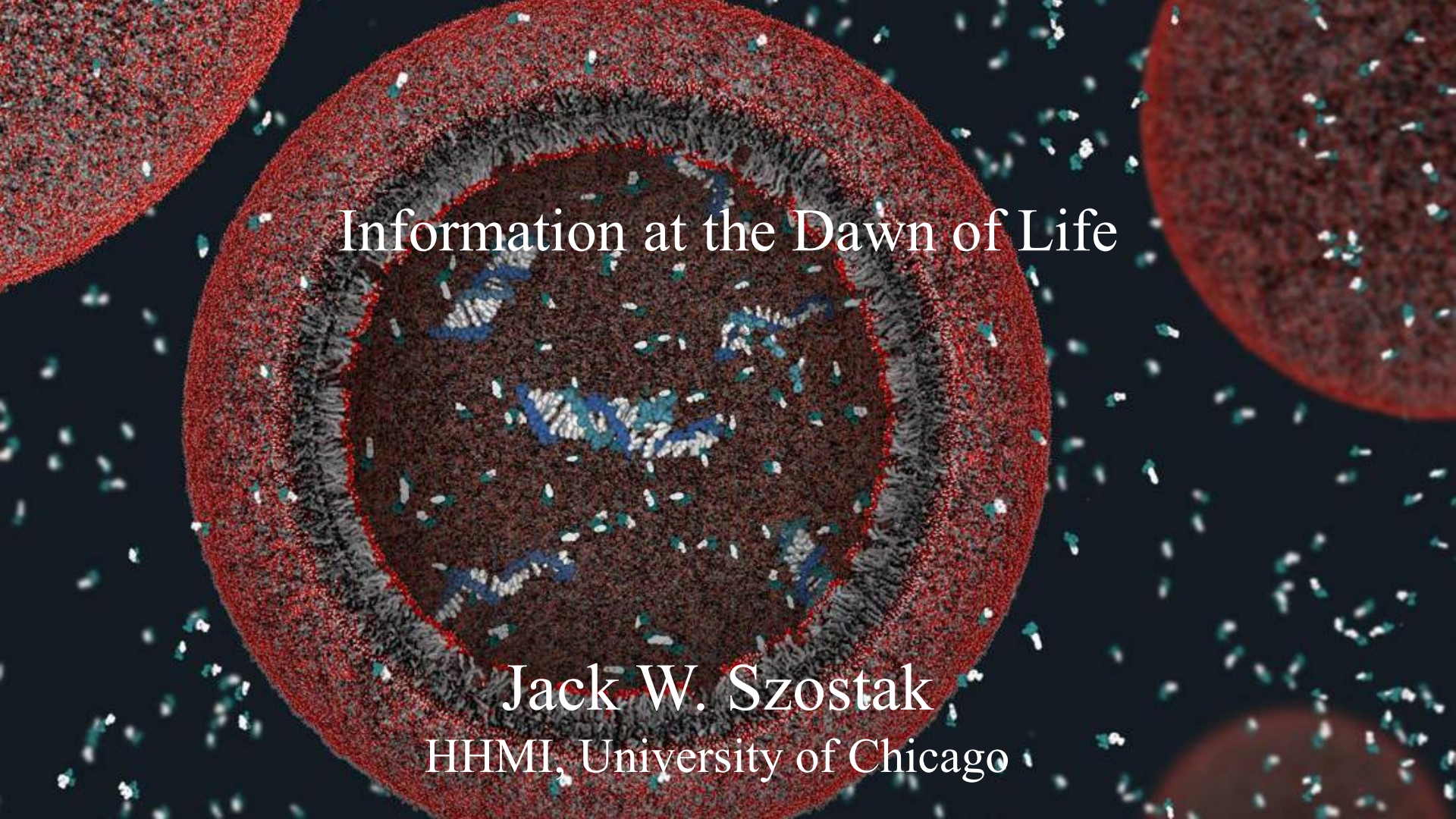


Information at the Dawn of Life

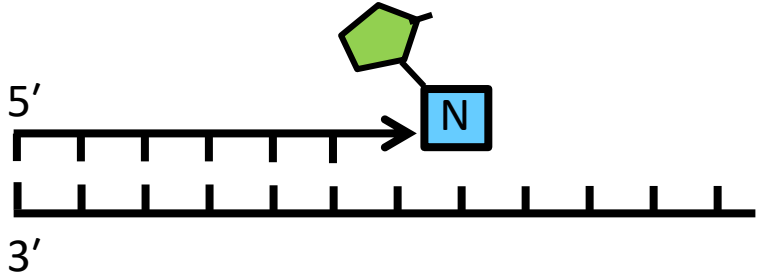
Jack W. Szostak
HHMI, University of Chicago

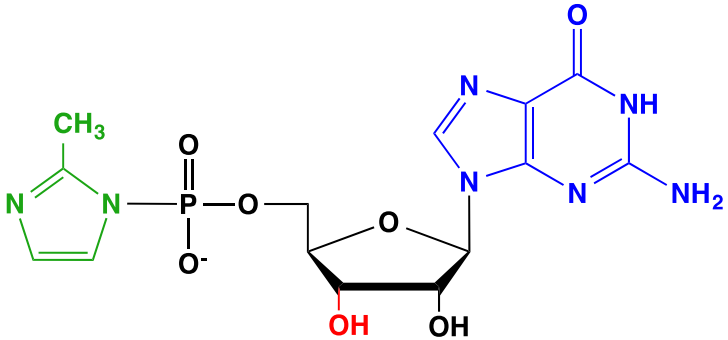
The setting for the first life?

RNA Template Copying Chemistry:

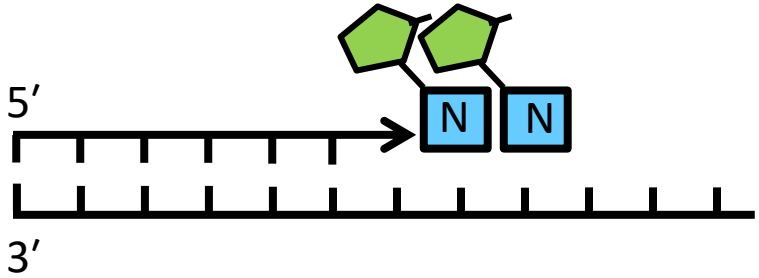
(nothing about this worked the way we thought it should)

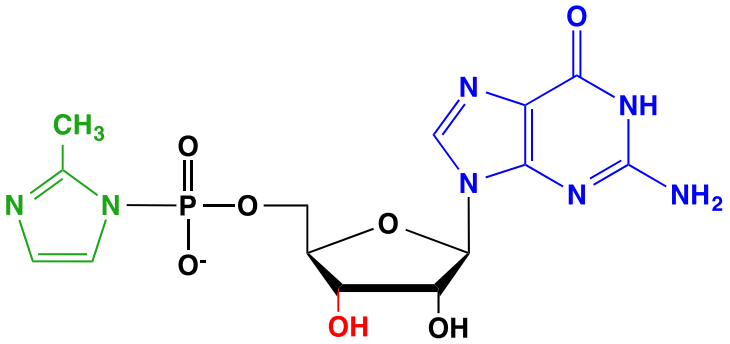
How we thought RNA copying worked: one monomer binds next to the primer and reacts



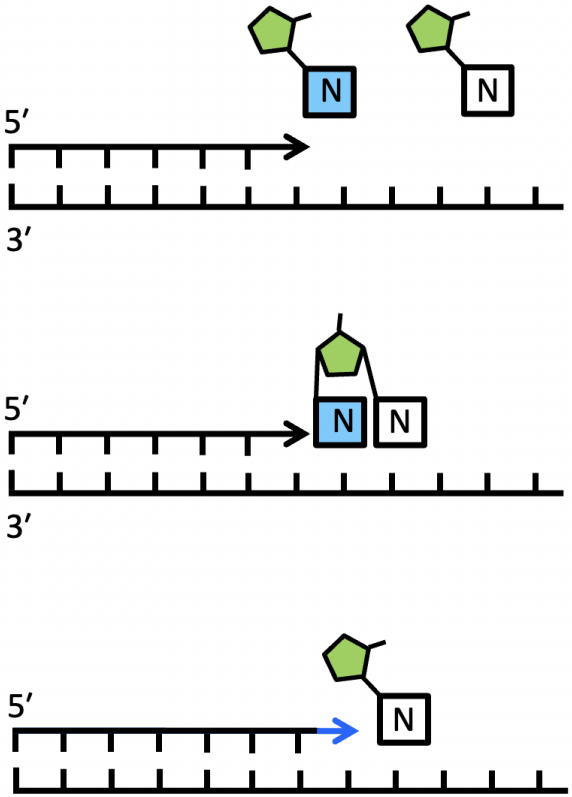


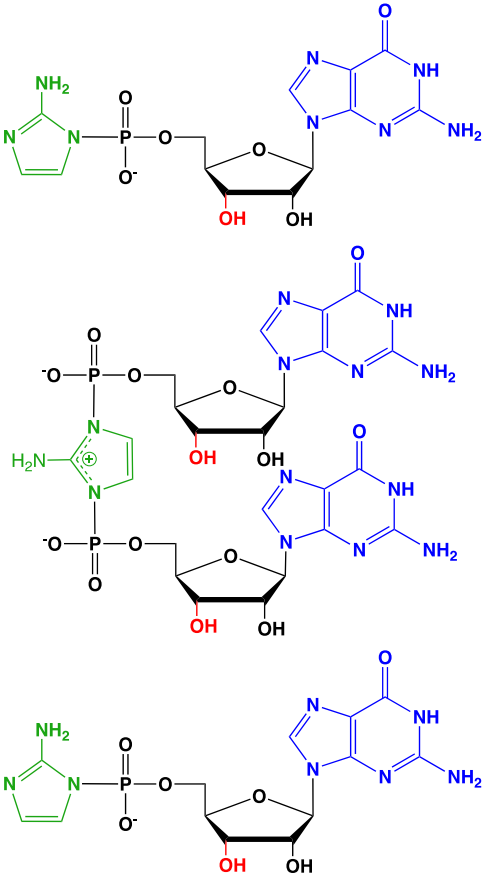
A downstream activated monomer catalyzes primer extension



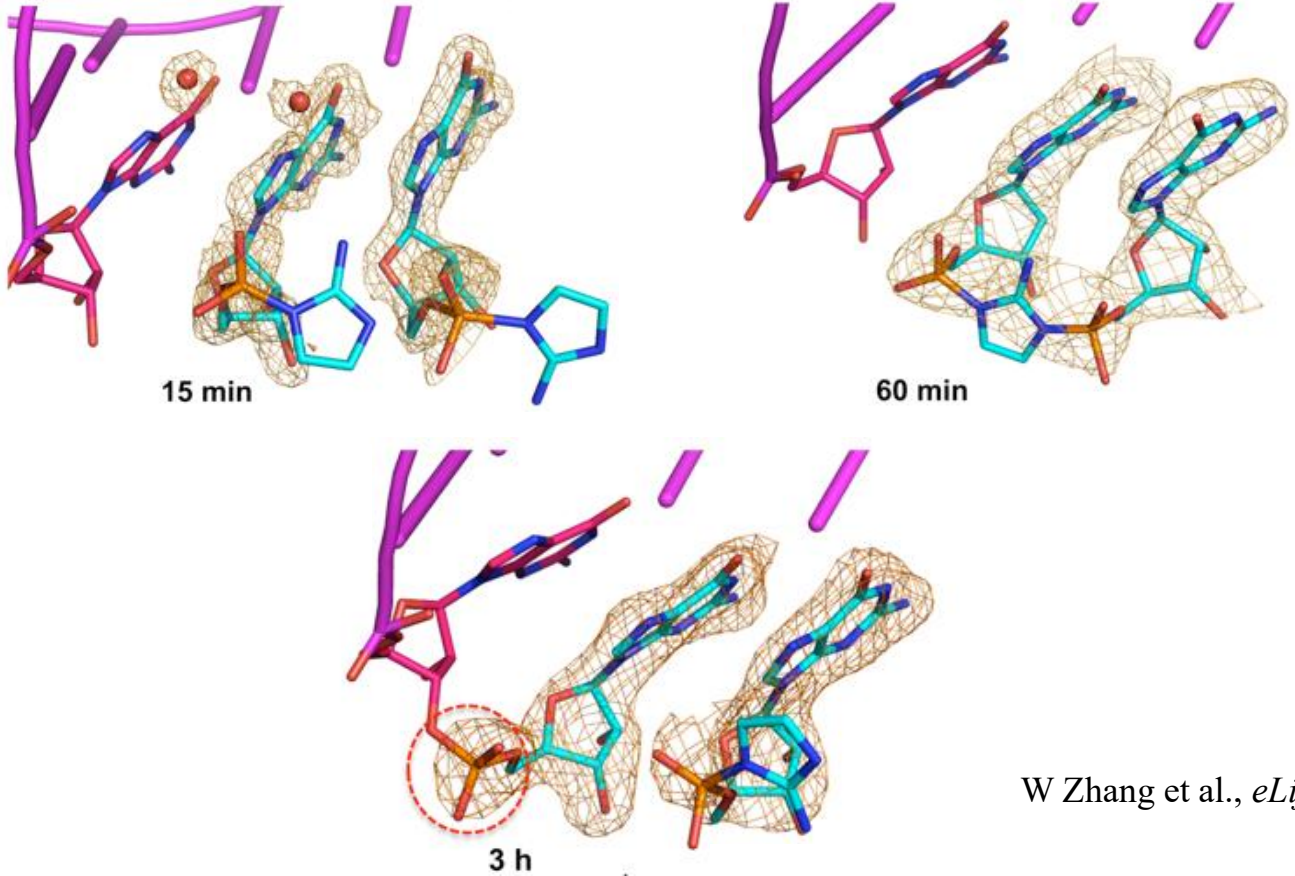


Mechanism of Nonenzymatic RNA Primer Extension



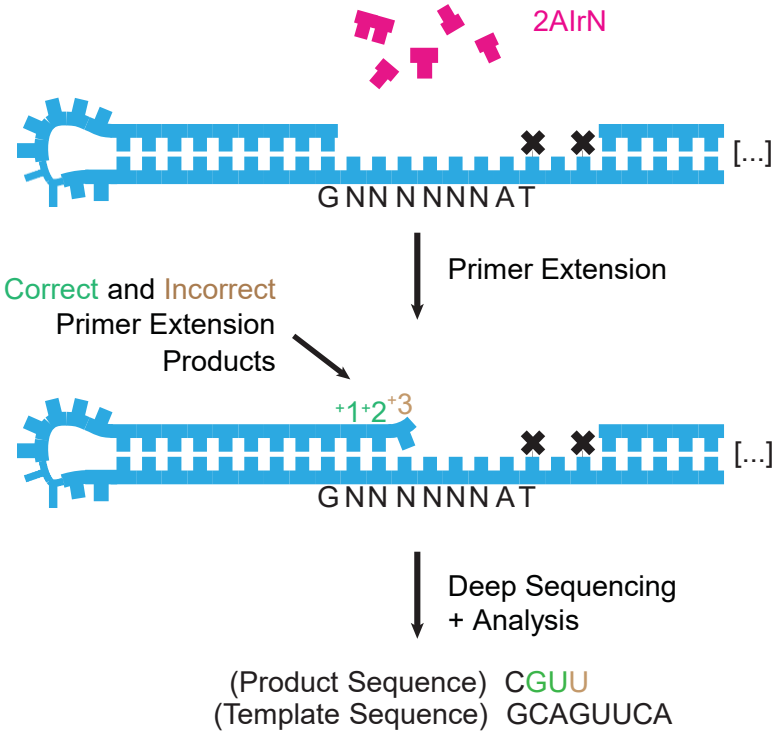


Primer extension in crystals: the 2-AI bridged intermediate forms first and then reacts with the primer



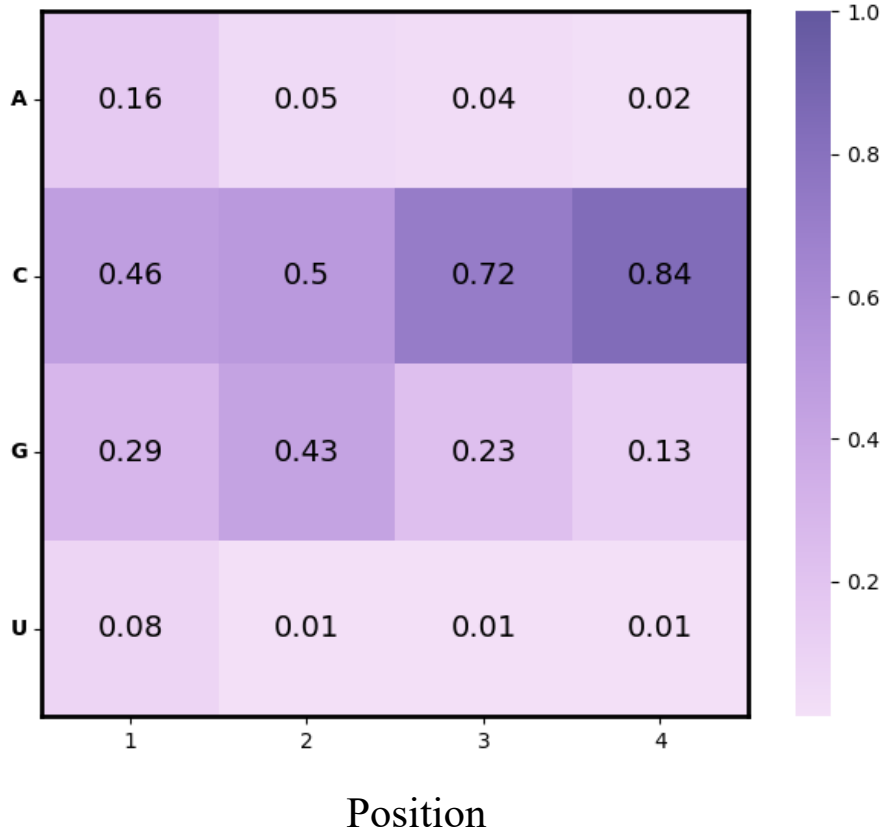
W Zhang et al., *eLife*, 2018, 7:e36422

Extent and Fidelity of RNA Copying Chemistry can be Analyzed by Deep Sequencing



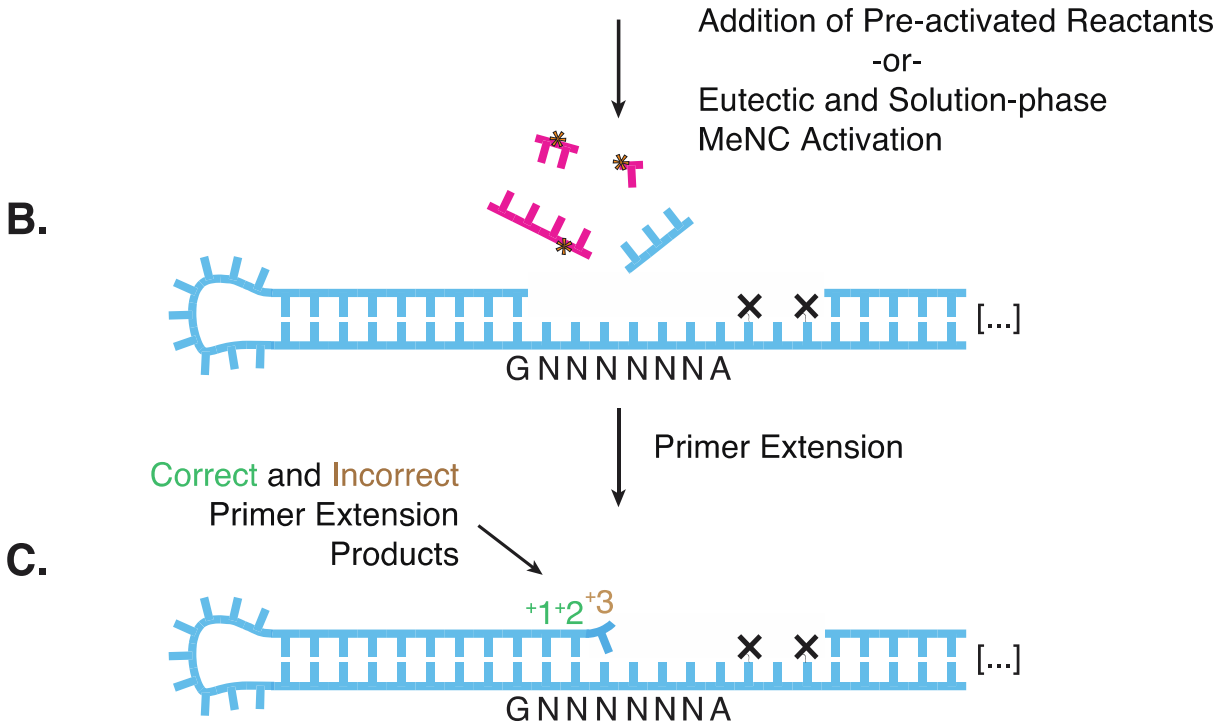
Duzdevich et al., *Nucleic Acids Res*., 2020, 49:3681-3691

RNA copying with activated monomers: extreme product sequence bias

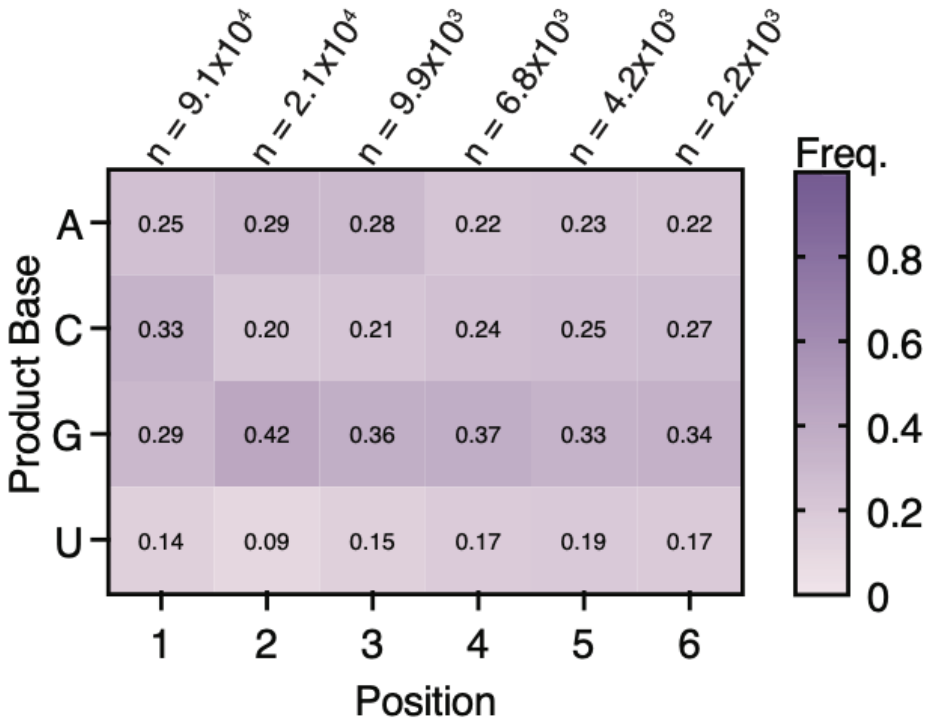


20 mM pre-activated 2AIrN
plus
MeNC activation chemistry

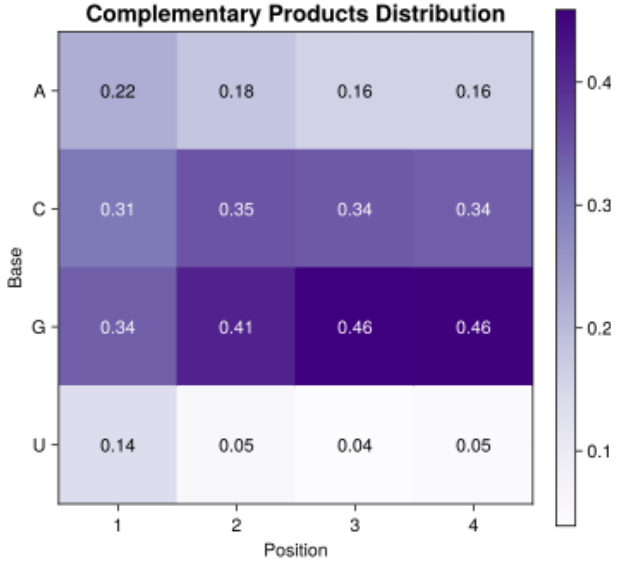
How can this bias be overcome? “helper oligos”



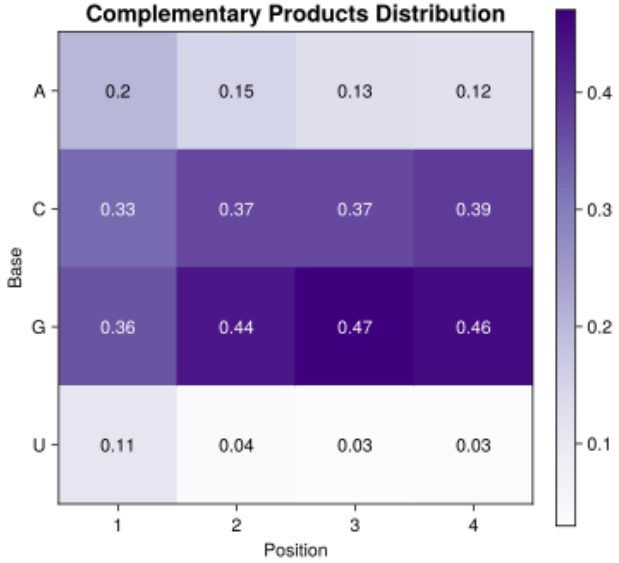
Helper oligos + activation chemistry reduces bias



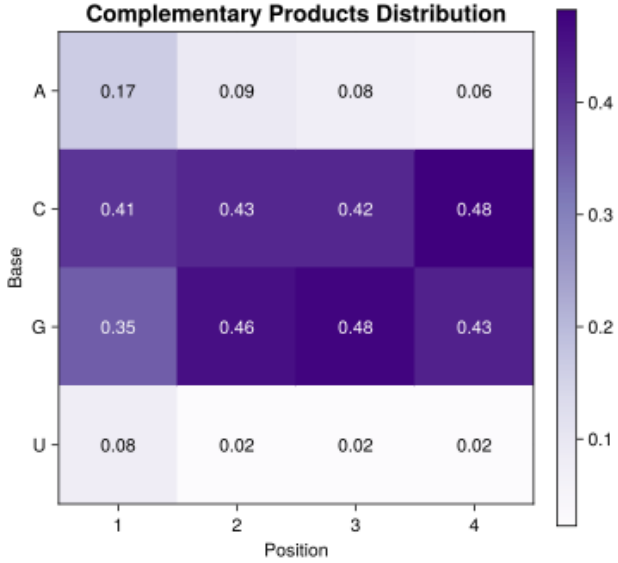
Lower temperature also reduces bias



4˚C



21˚C



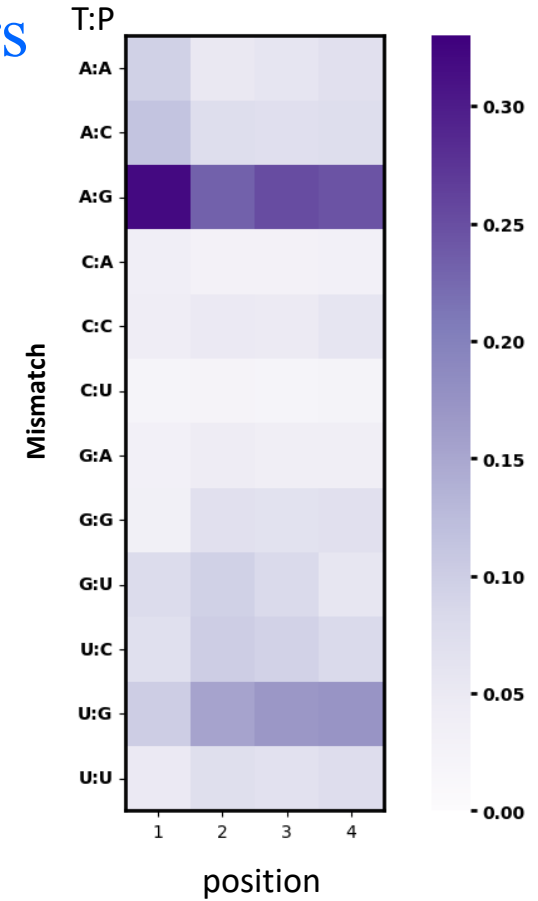
40 ˚C

Fidelity from sequencing: G:A and G:U mismatches dominate errors

Template Sequence: 6N

Error Frequency: 11%

10 mM 2AI-pN monomers, 24 h reaction



Duzdevich et al., Nucleic Acids Res., 2021, 49:3681-3691

The Fidelity of Nonenzymatic RNA Replication

A High Error Rate is Bad: The Eigen Error Catastrophe

An arbitrarily high error rate μ cannot be overcome simply by invoking an ever stronger selective pressure

The maximum number L of functionally important positions in a genome is commonly approximated as:

$$L = 1/\mu$$

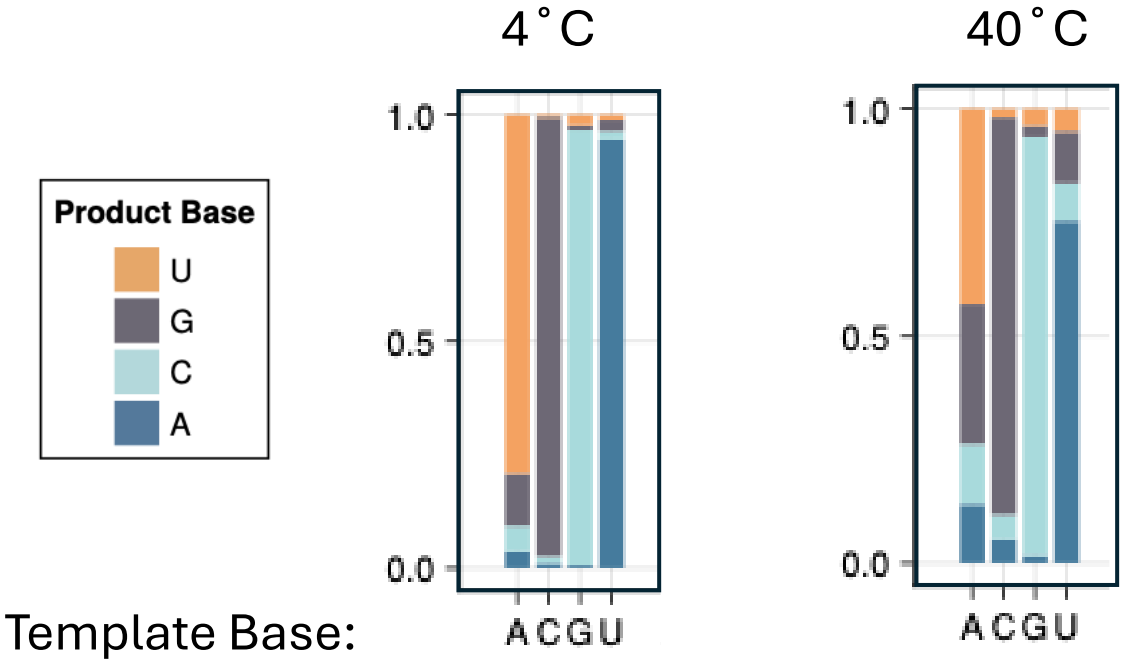
For μ = 6 to 10%, L = 10 to 16 nt

We would like μ to be less than 1-2%

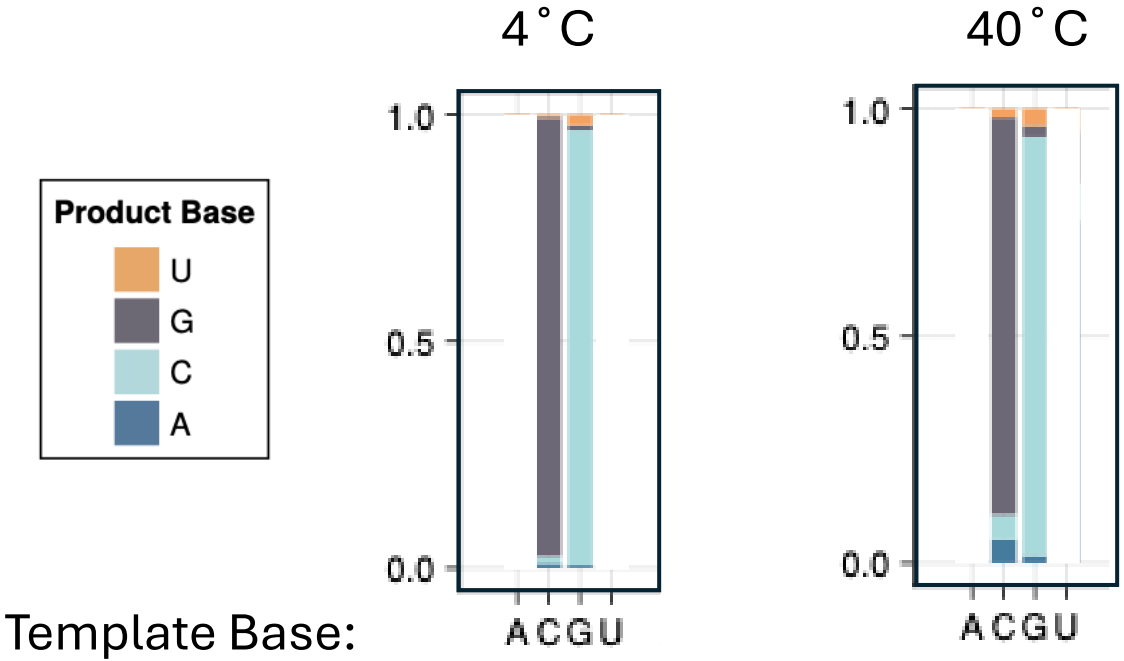
Reducing the Error Rate: Colder is Better

Temperature	Error Rate μ
4 °C	5.9%
21°C	7.9%
40°C	16%

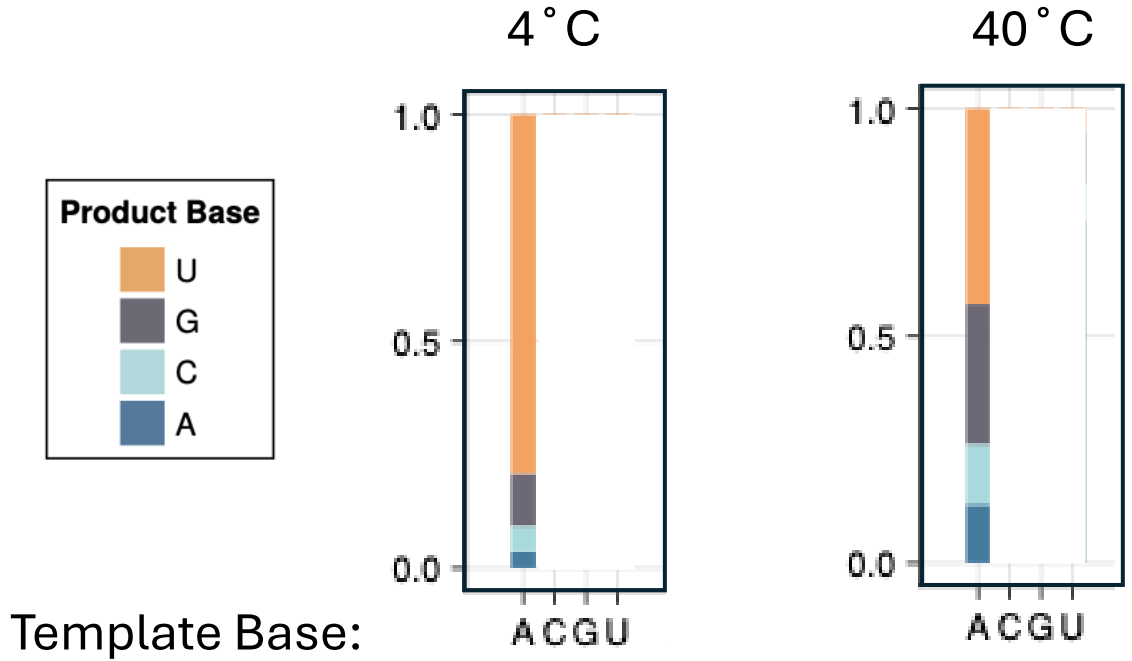
Cold is Better, but the Error Rate Depends on the Base being Copied



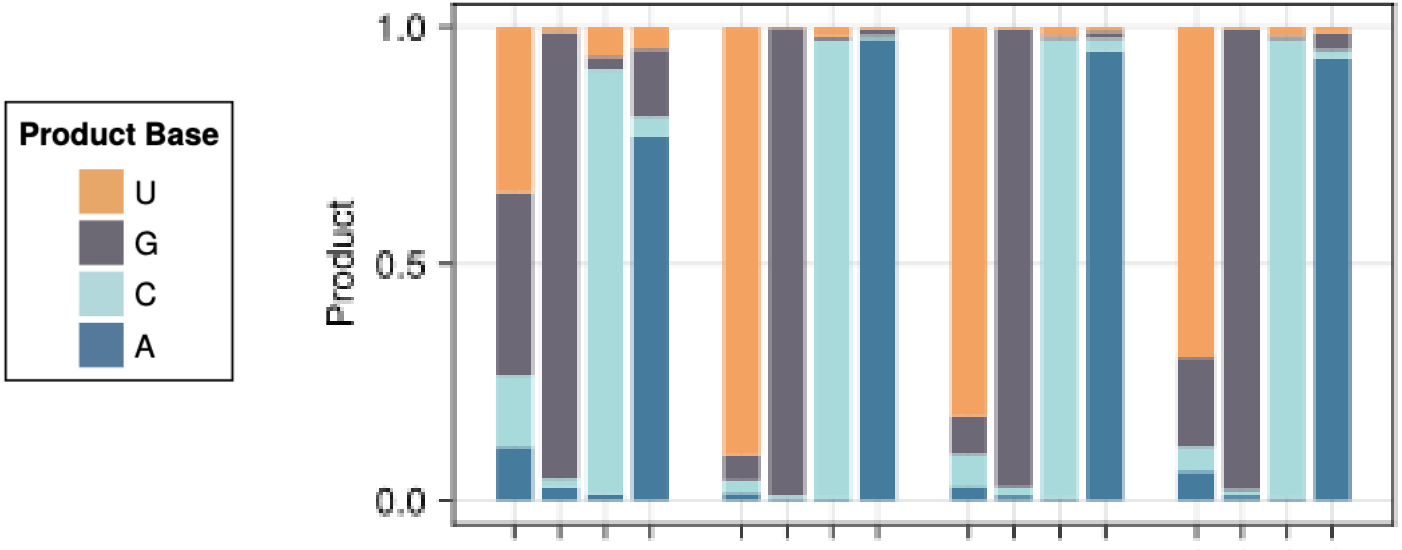
Cold is Better, but the Error Rate Depends on the Base being Copied



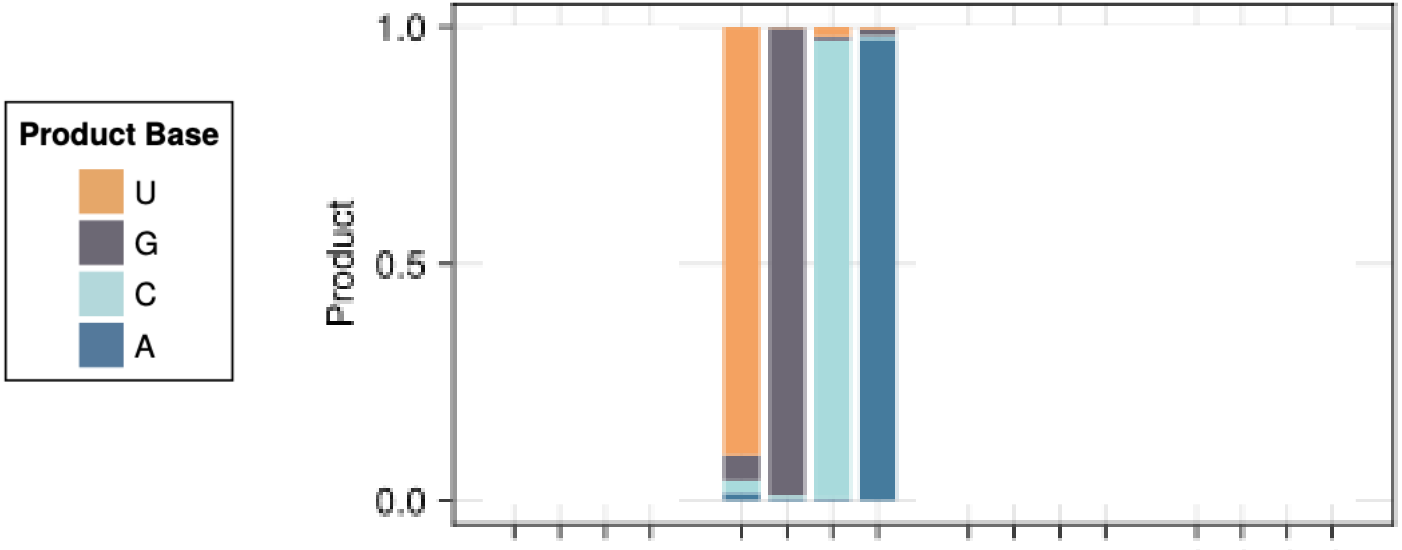
Cold is Better, but the Error Rate is really bad if the Base being Copied is an A



Context Dependence of the Error Rate: The subsequent template base has a huge effect!



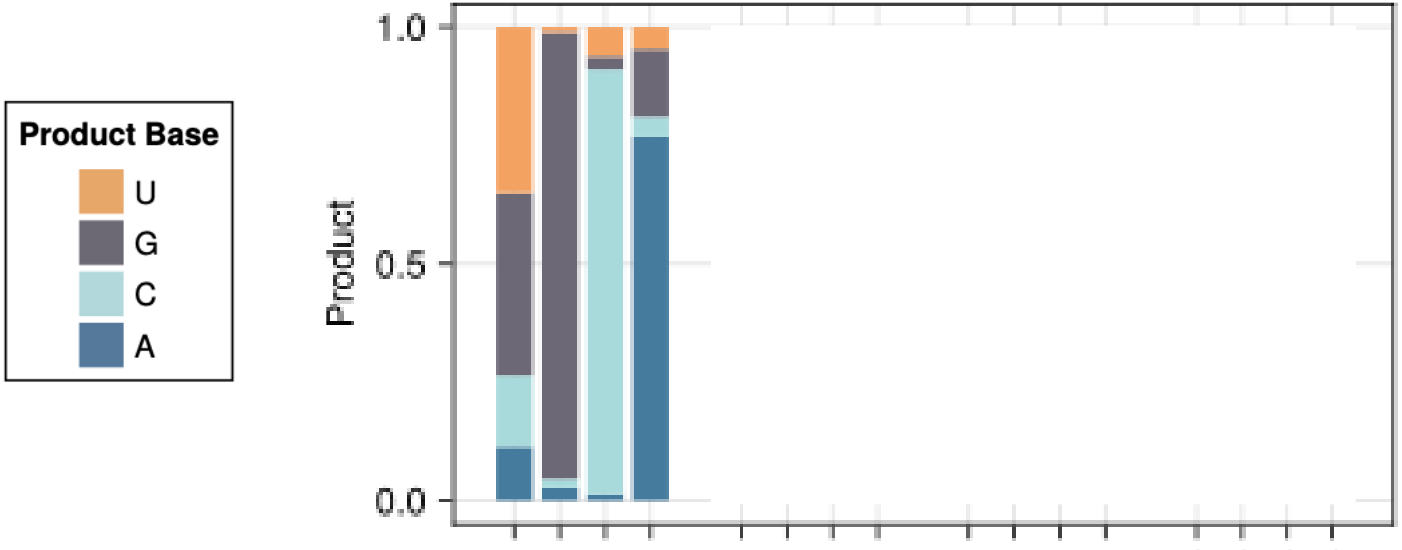
Context Dependence of the Error Rate: The subsequent template base has a huge effect!



Template Base: A C G U

Next Template Base is C

Context Dependence of the Error Rate: The subsequent template base has a huge effect!



Template Base: A C G U

Next Template Base is A

Is this telling us something about the composition of the primordial genome?

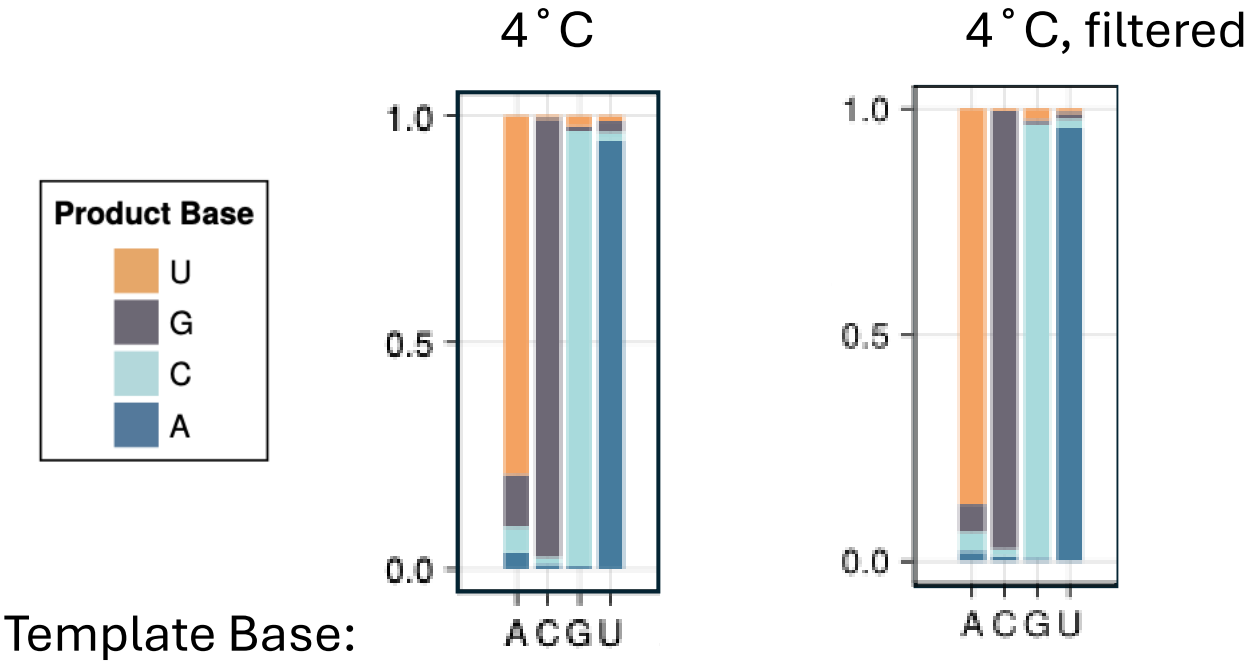
Copying through template sequences AA, AU, UA and UU is very inefficient.

When these sequences are copied, the copy is highly error prone.

Maybe primordial genomes that contained strings of 2 or more As and Us simply died out…

What happens if we computationally delete all templates that contain AA, AU, UA, and UU sequences?

Combining Cold with Template Restriction Decreases the Error Rate



Combining Cold with Template Restriction Decreases the Error Rate

Genome:	unrestricted μ	no AU strings μ
Temperature		
4 ˚C	5.9%	4.8%
21˚C	7.9%	6.3%
40˚C	16%	13%

Post-Mismatch Stalling

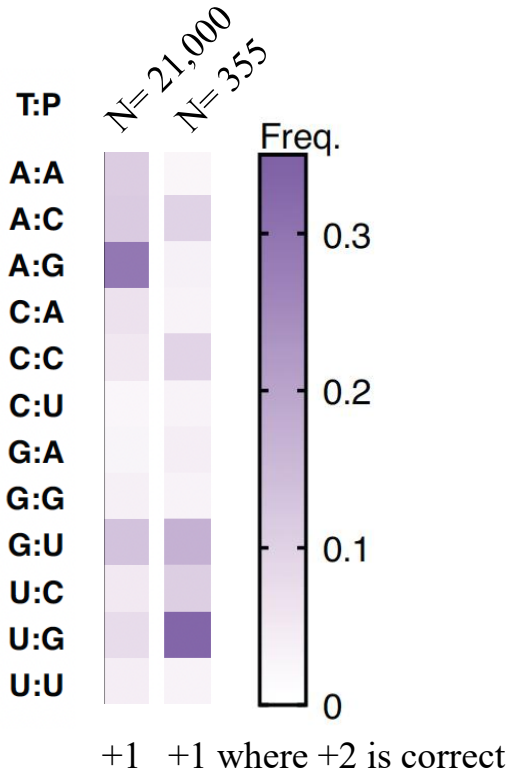
what it is, and why it helps

Post-mismatch Stalling Decreases the Internal Error Rate

Correct at +1, Correct at +2: 42,714
Mismatch at +1, Correct at +2: 355

Total Error Rate at +1: 12%
Effective (Internal) Error Rate at +1: 0.83%

Template: 6N
Substrates: 20 mM 2AIpN
Time: 24 h



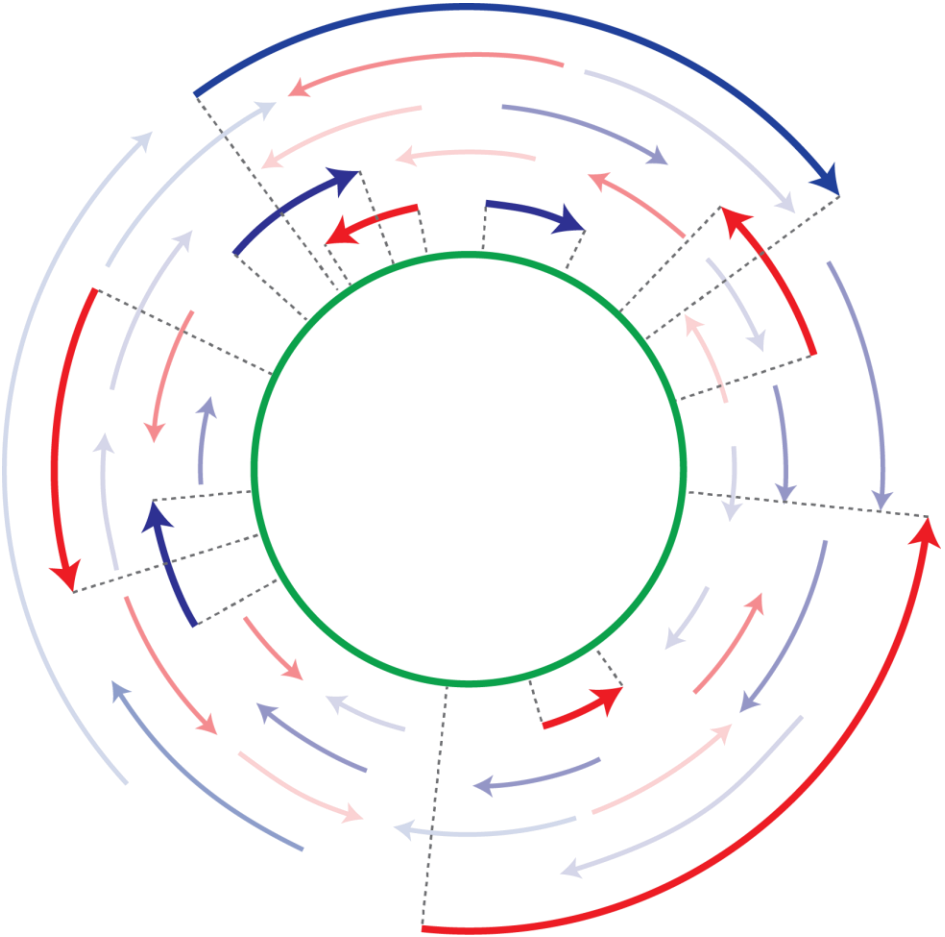
Duzdevich et al., Nucleic Acids Res., 2021, 49:3681-3691

note change in error distribution

From Copying to Replication

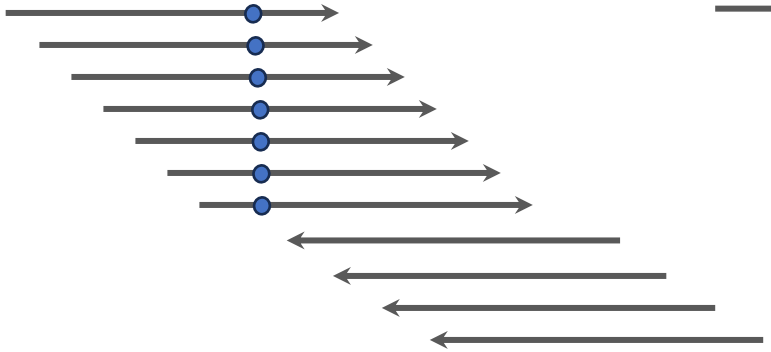
Virtual Circular Genome (VCG) Model

- No defined start or end
- Every oligo can act as the primer, template, and helper
- Extension of the entire population at the same time
- No need to copy long RNA templates within a single generation

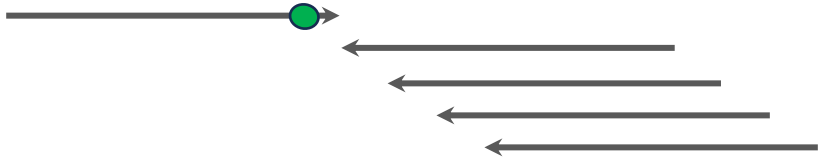


Zhou, L., Ding, D., & Szostak, J. W. (2021). *RNA*, *27*(1), 1-11.

VCG replication is biased against new mutations



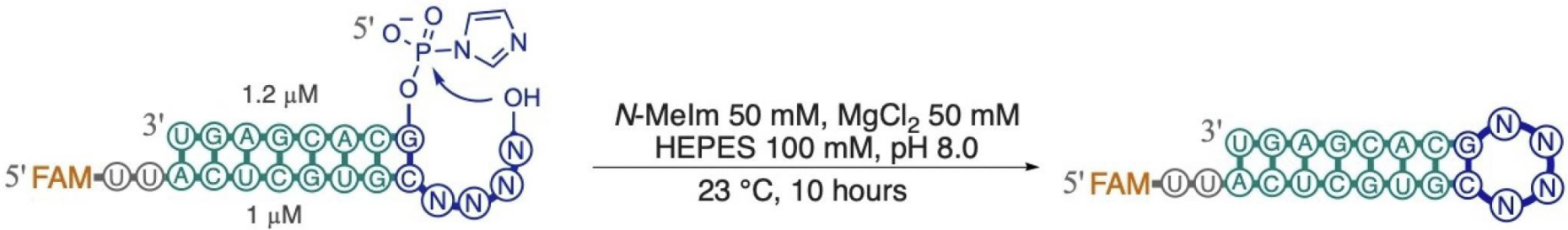
A wild type nucleotide is present on many different oligos, both close to and far from the 3′-end, so that many oligos from the opposite strand can copy that nucleotide



A new mutation is present at the 3′-end of an oligo, so it cannot be copied until further primer extension occurs, allowing a primer to bind downstream of the mutant site

Falk, Zhou, Matsubara, Husain, Szostak, Murugan, submitted

Building Structured RNAs from VCG Oligos by Loop-closing Ligation



Long-fei Wu et al., 2024

An Intriguing Possibility…

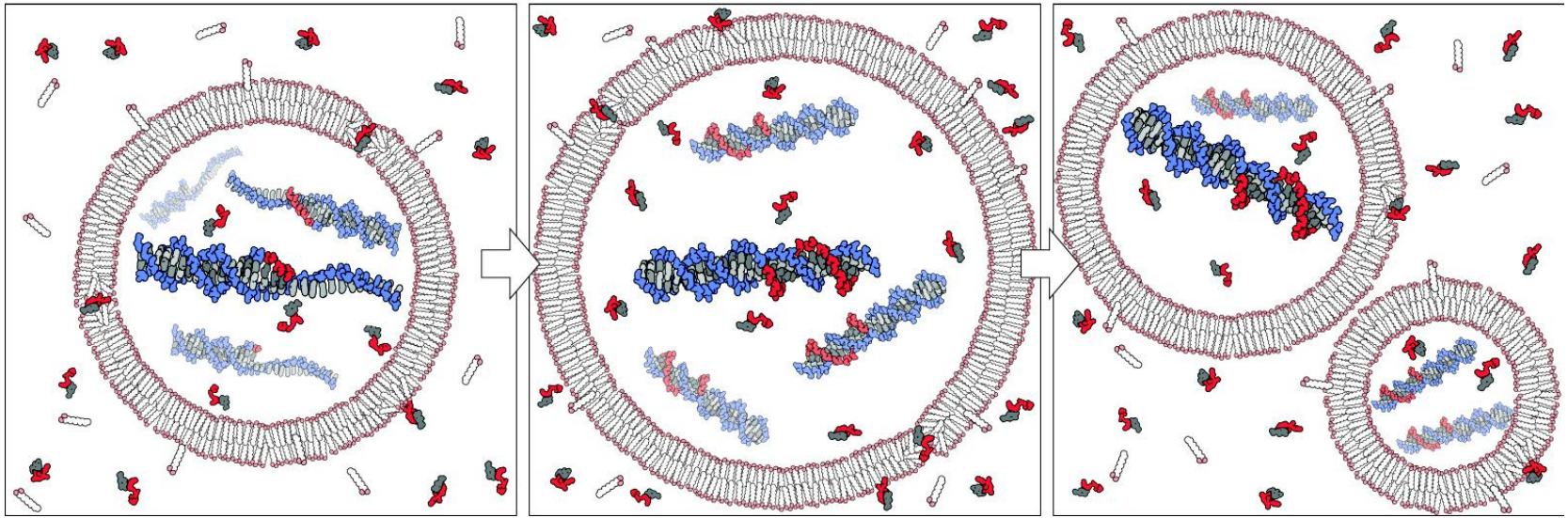
VCG oligos ending in an ‘incorrect’ G would be frequent.

Primer extension and Loop-closing ligation are competing processes

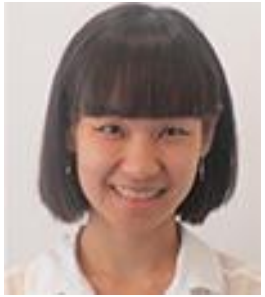
GNRA tetraloops are very stable and very common motifs in structured RNAs.

Might G:A mismatches be removed both from the genomic pool, and used to build up structured RNA, by loop-closing ligation?

The spontaneous emergence of Darwinian evolution in a purely physical-chemical system may be within reach



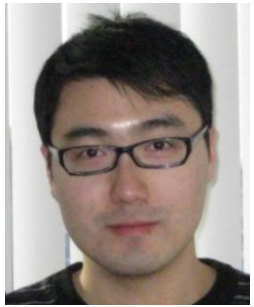
Acknowledgements



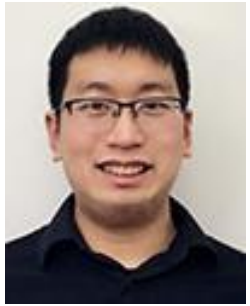
Lijun Zhou

Dian Ding

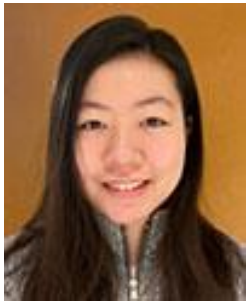
Stephanie Zhang



Wen Zhang



Ziyuan Fang



Xiwen Jia

Ben Colville

Yanfeng Xing

Collin Nisler

Daniel Duzdevich

Orhan Acikgoz

Andy Yu
Zhi Li

Collaboration with Murugan lab

Arvind Murugan

Martin Falk

Yoshiya Matsubara

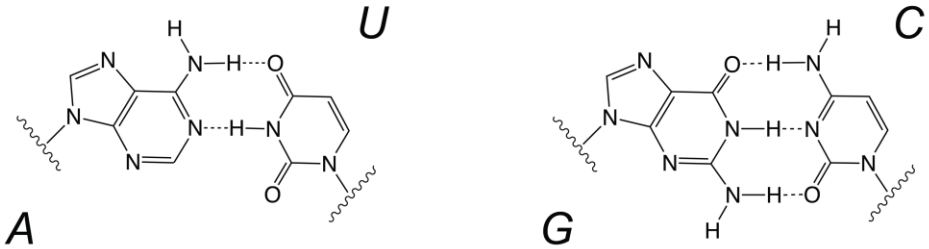
Leon Zhou (undergrad)

Ruby Wen

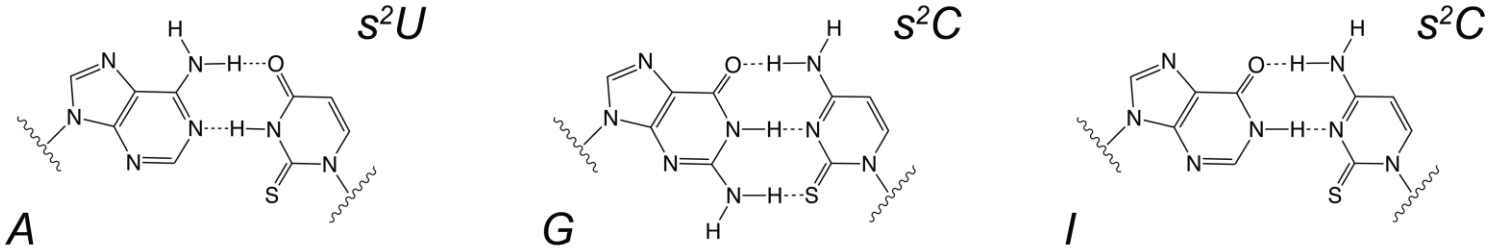
Kabir Husain
(Now faculty at UCL)

What was the Primordial Genetic Alphabet?

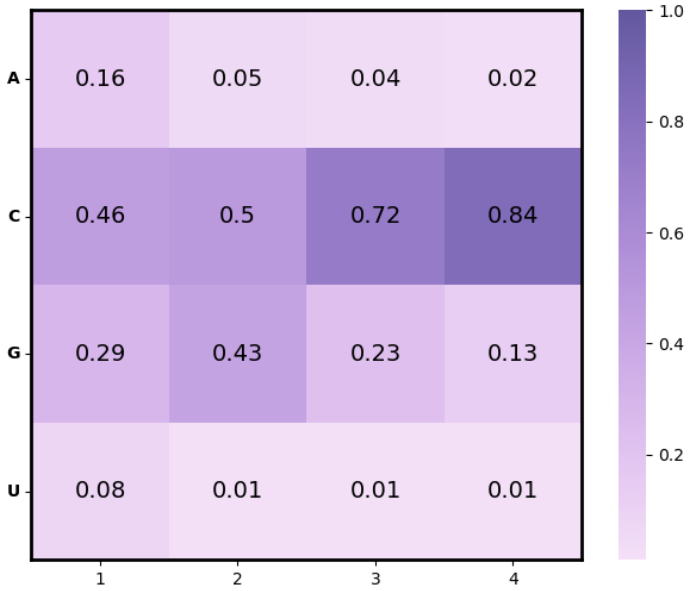
Canonical base pairs



Noncanonical base pairs

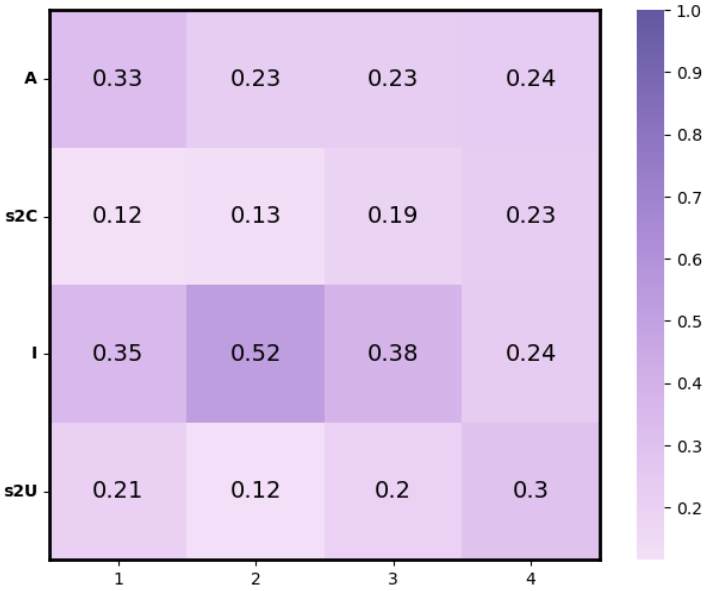


Copying RNA with a potentially primordial genetic alphabet reduces product bias



Position

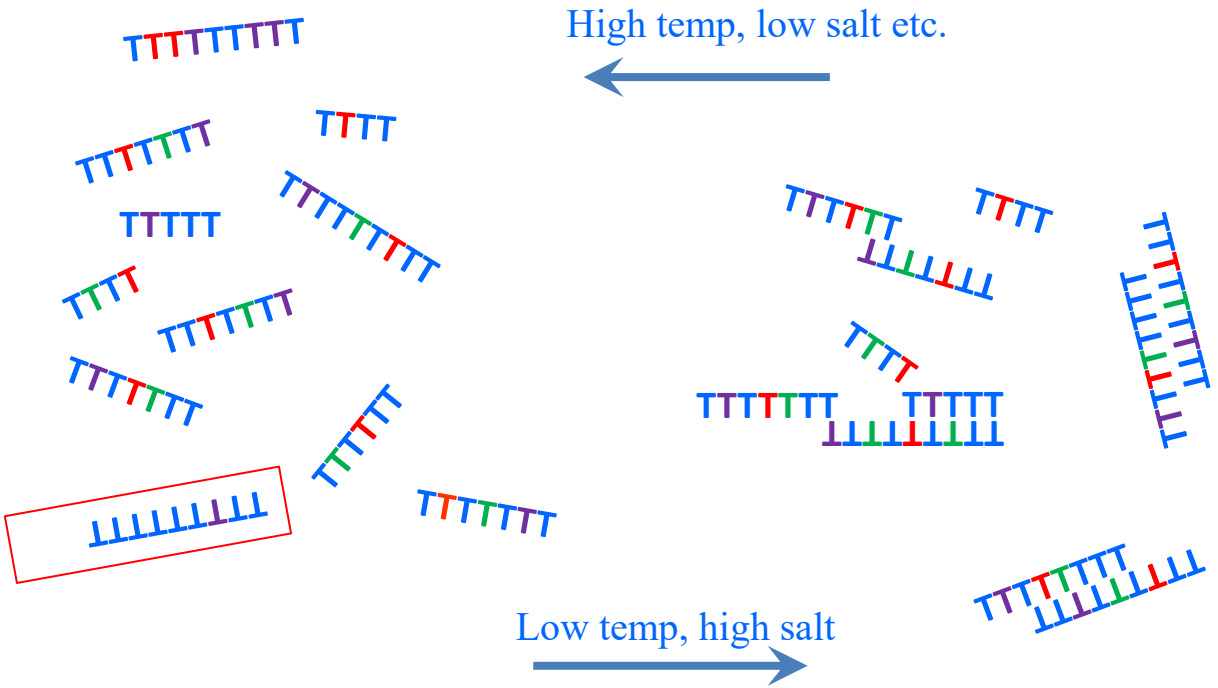
20 mM pre-activated 2AIrN
MeNC activation chemistry



Position

20 mM 2AIrN
N = A, I, s2U, s2C

Environmental fluctuations shuffle annealed configurations, allowing continued copying



Origin of the First Genomes

Nucleotide synthesis and activation chemistry

Formation of Oligonucleotides

Template Copying Chemistry

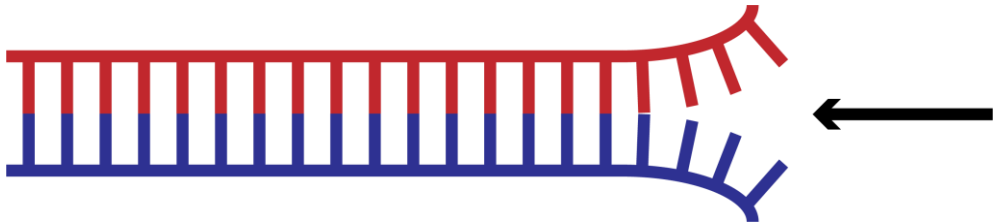
From Heterogeneity to Homogeneity

From Copying to Replication

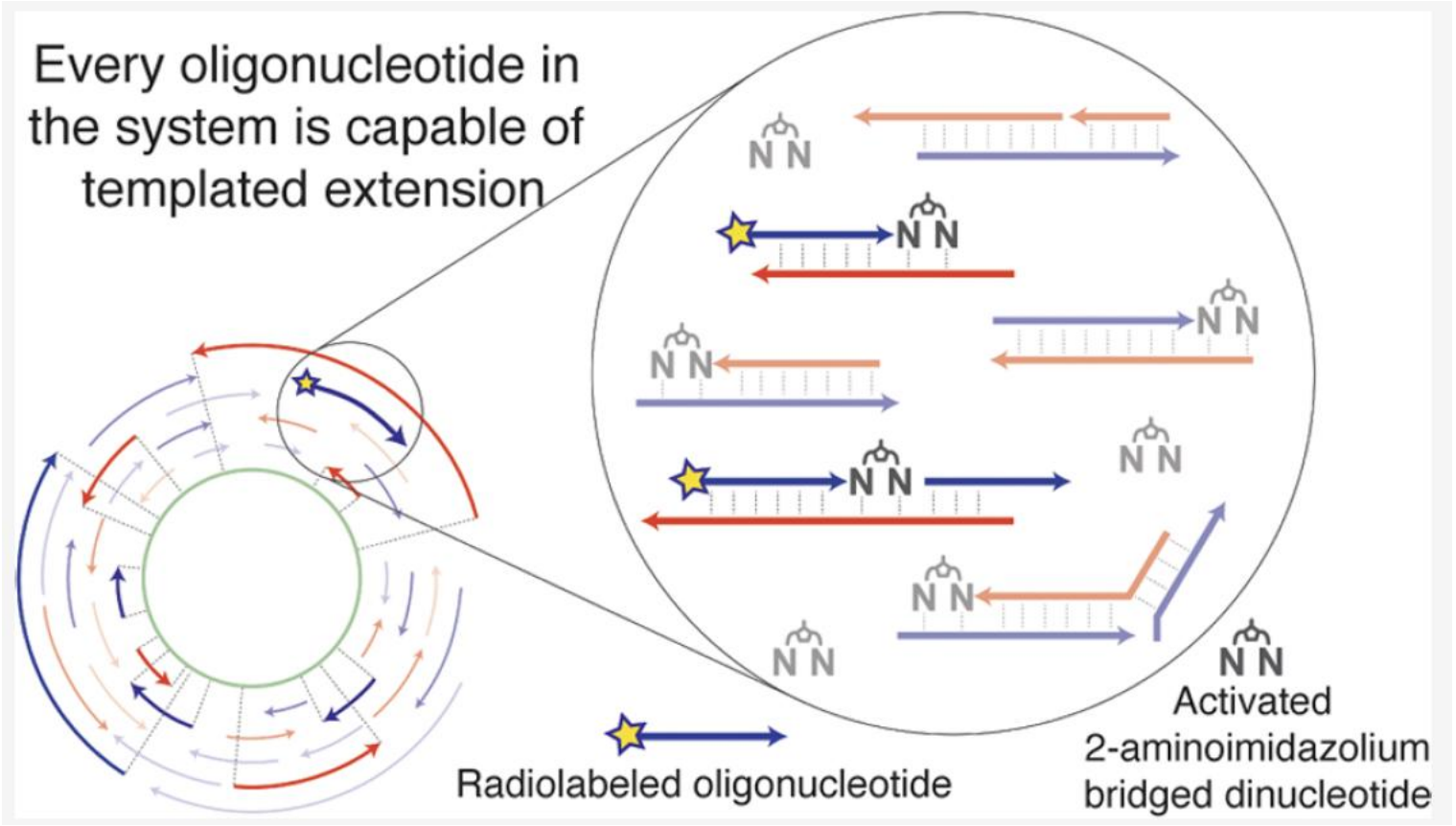
Problems with the Replication of Linear Sequences

- Defined primers required to copy complementary strands

- Strand separation of long RNA oligonucleotides is difficult

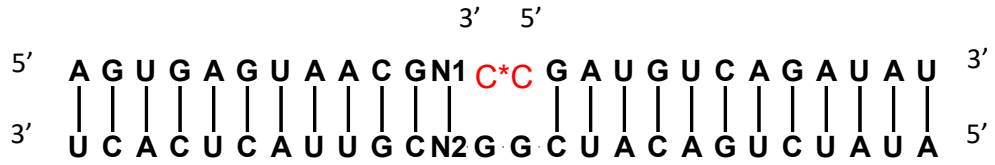


A path to replication by primer extension or ligation



Genome copying is distributive, not processive

Experimental Measurement of Stalling Factors in RNA: Rate of Primer Extension after a Mismatch

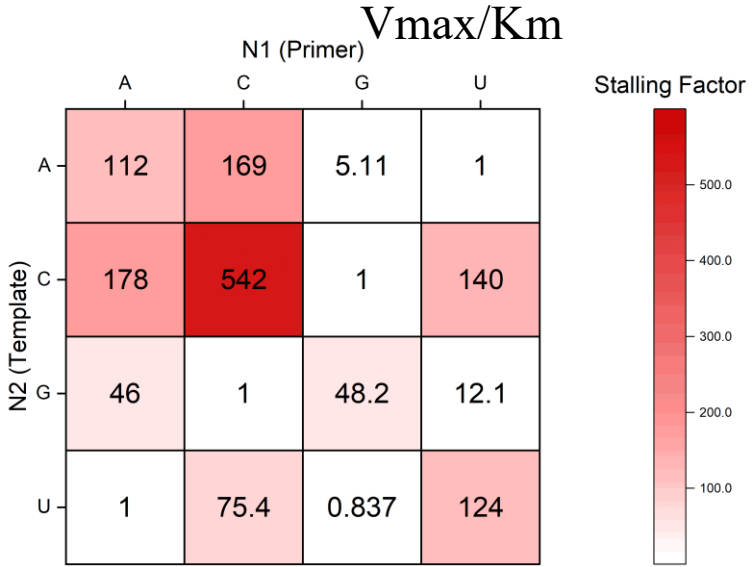


Maximum Stalling Factor:

$$\frac{\text{(Vmax/Km) correct}}{\text{(Vmax/Km) incorrect}}$$

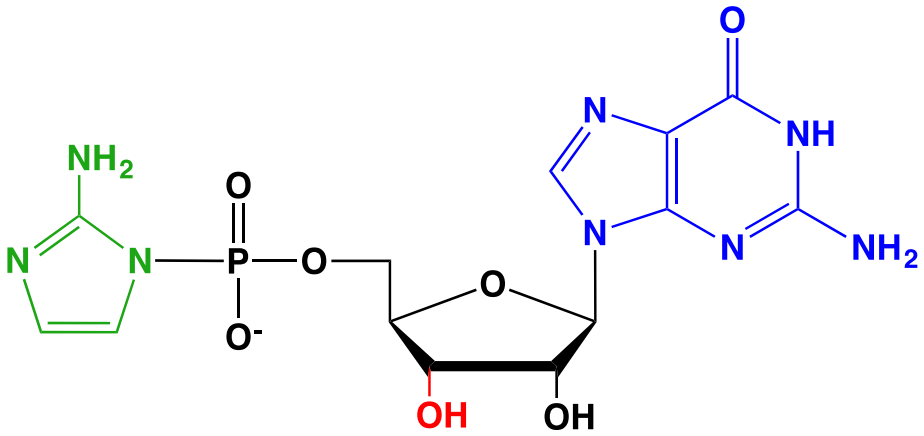
Minimum Stalling Factor:

$$\frac{\text{(Vmax) correct}}{\text{(Vmax) incorrect}}$$



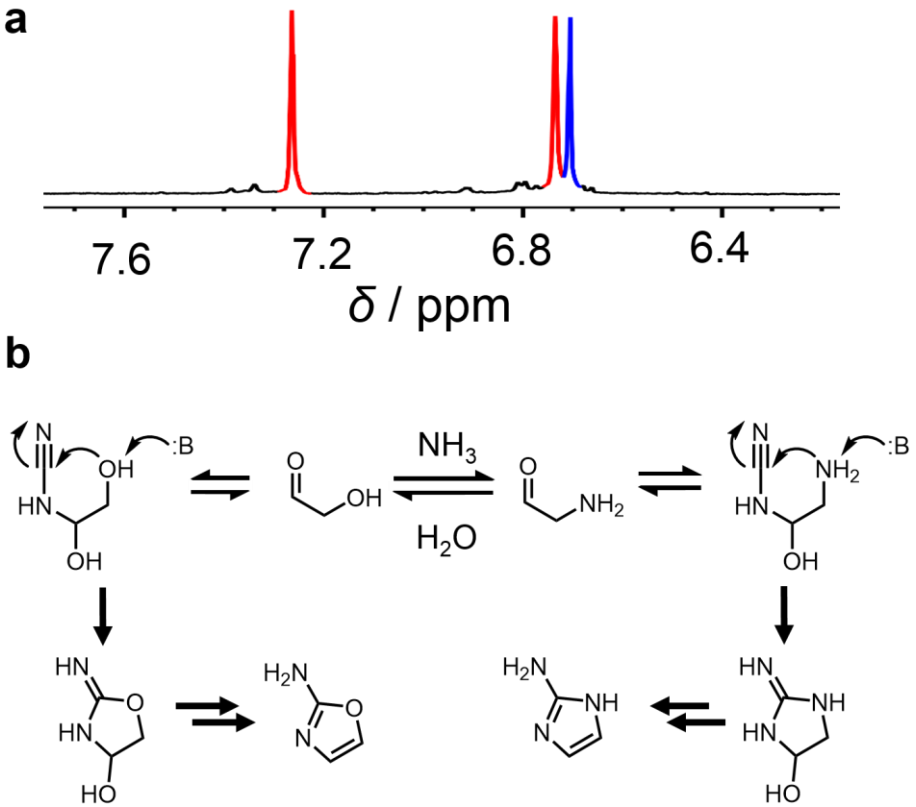
In preparation: Orhan Acikgoz

Monomers must be activated to polymerize: 2-aminoimidazole is our best activating group

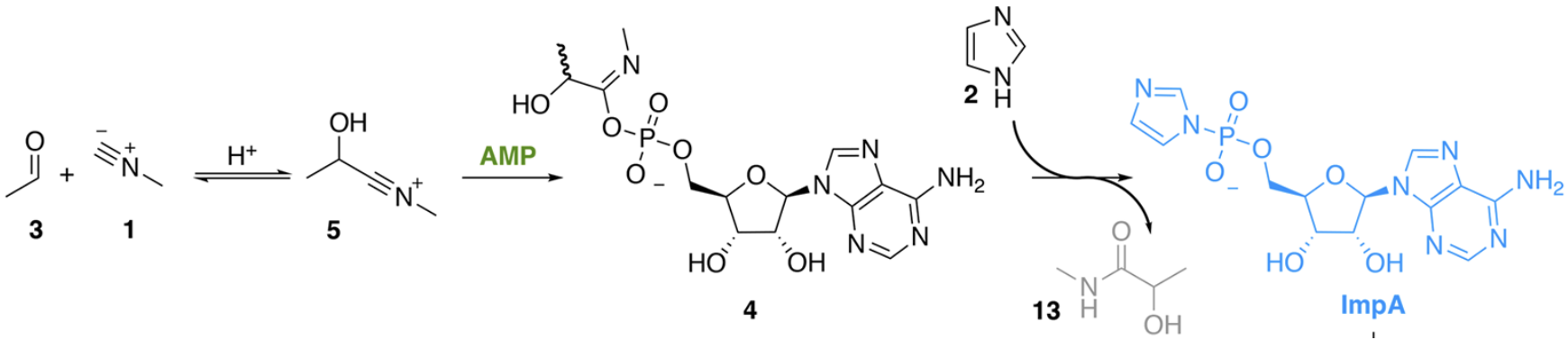


- higher pKa
- hydrolyzes slowly
- plausible prebiotic synthesis of 2AI

Potentially Prebiotic Synthesis of 2-aminooxazole and 2-aminoimidazole

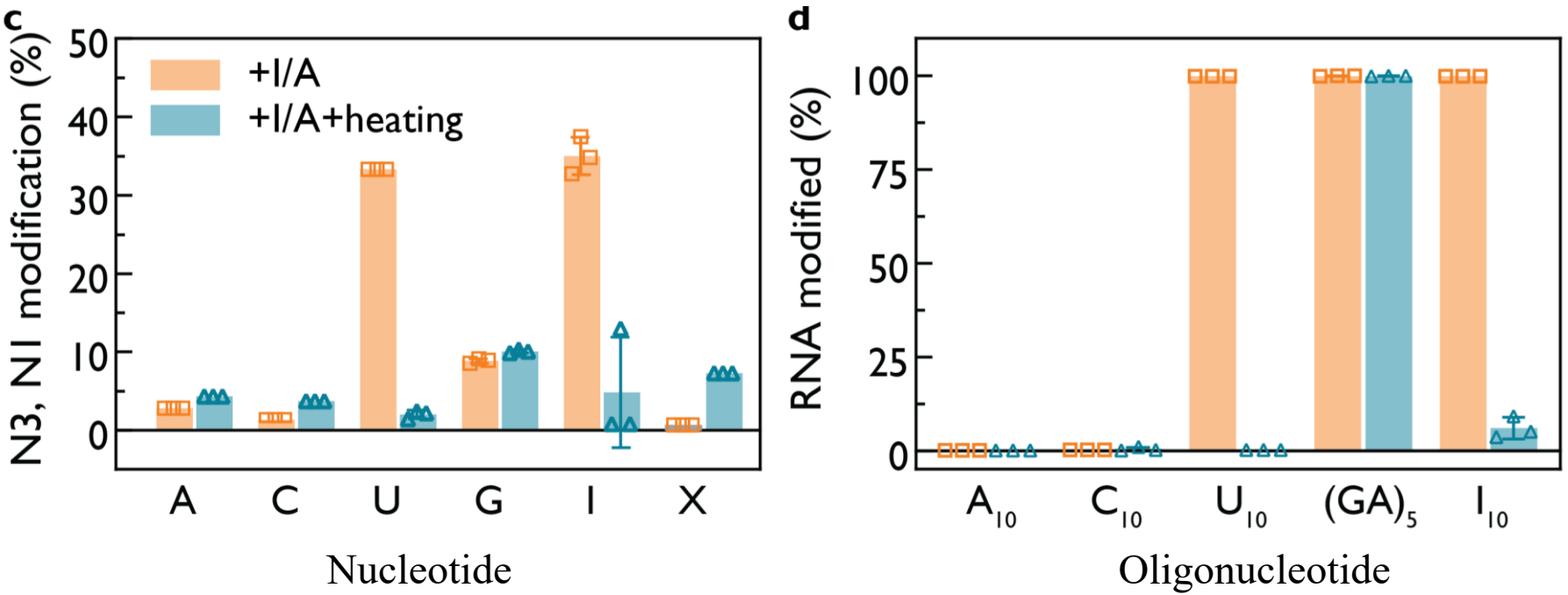


Isonitrile Activation Chemistry



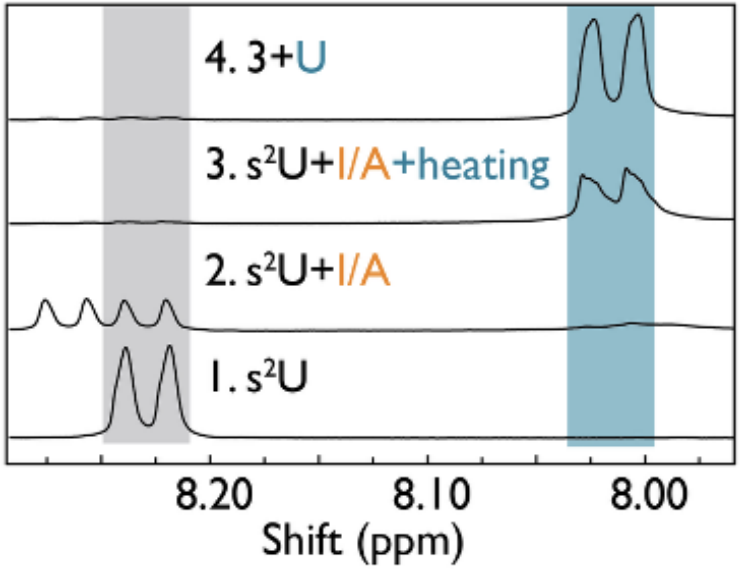
This works, but is there a better way?

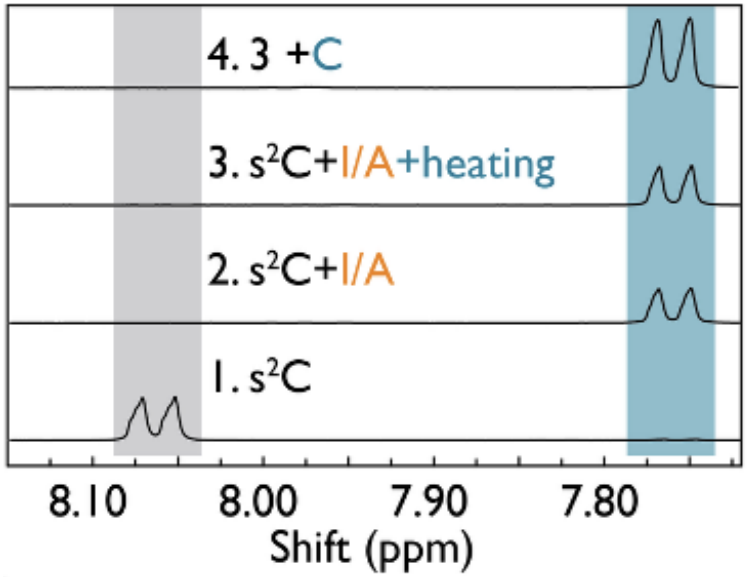
Nucleotide Modification by Activating Chemistry



Methyl isonitrile + 4-pentenal, before and after heating at 95˚ for 15 min

Conversion of 2-thio to Canonical Pyrimidines

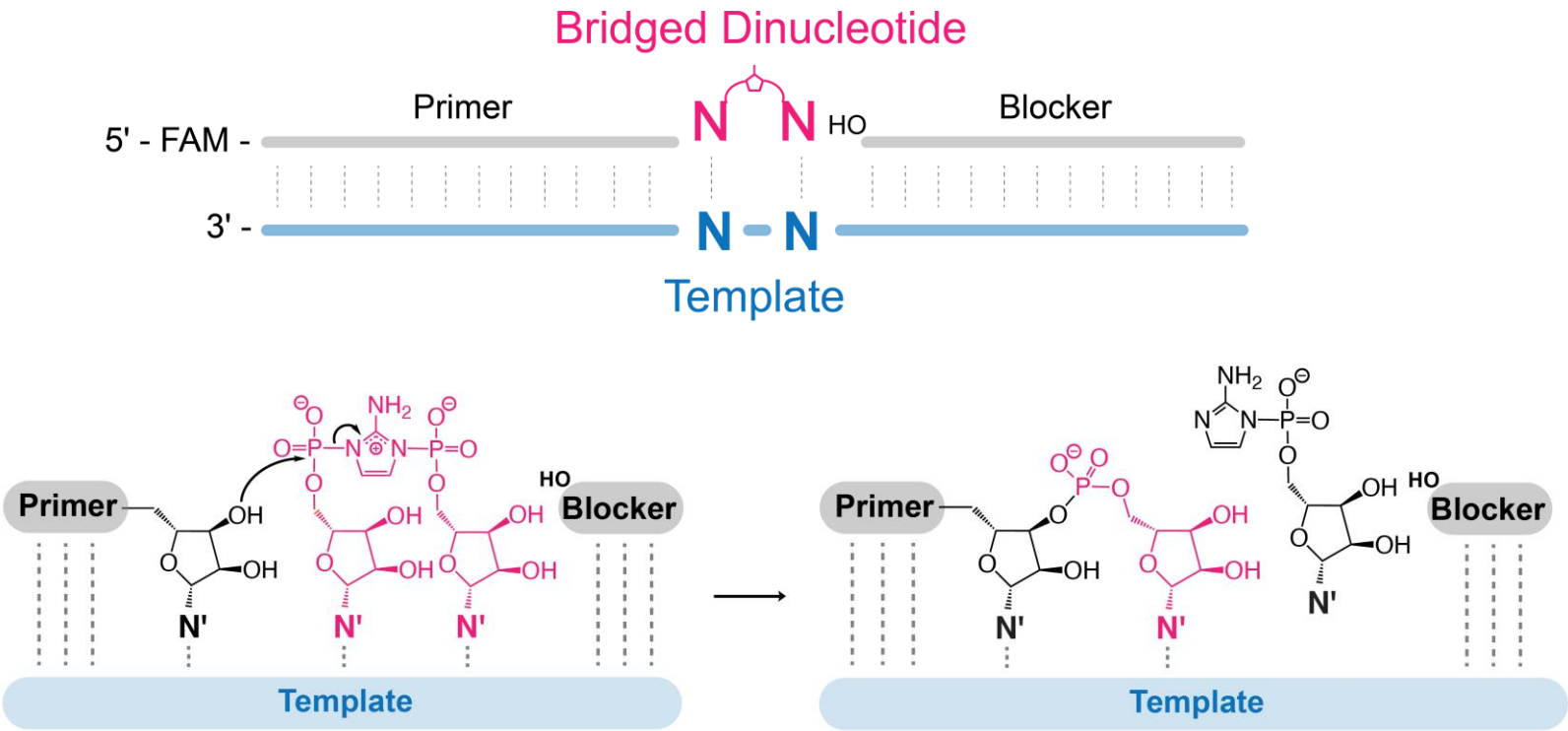




Modified 2sU to converted to U by heating at 95˚ for 15 min

2sC converted to C by exposure to MeNC and aldehyde

Kinetic Studies of Primer Extension



Jia, X. et al., *J. Am. Chem. Soc.* **2024**, 146, 23

Isonitrile Activating Chemistry would Favor the Modern RNA Alphabet

Alphabet selection	pH 6	pH 8
Isonitrile	A U C G ↑ ↑ A s^2U s^2C I	A U C G→G* ↑ ↑ A s^2U s^2C I
N-acylimidazoles	A U C G A s^2U s^2C I	A‡ U‡ C‡ G‡ A‡ s^2U‡ s^2C‡ I‡

*nucleobase modifications

‡hydroxyl (sugar) modifications