The molecular origins of life

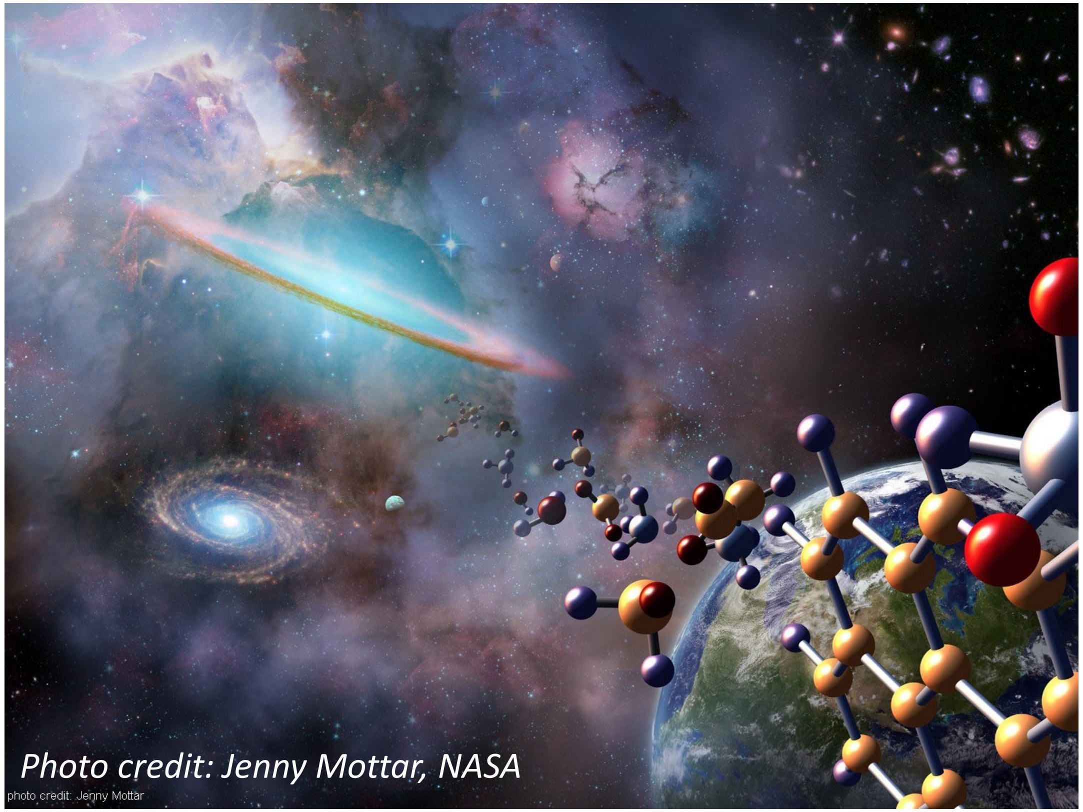
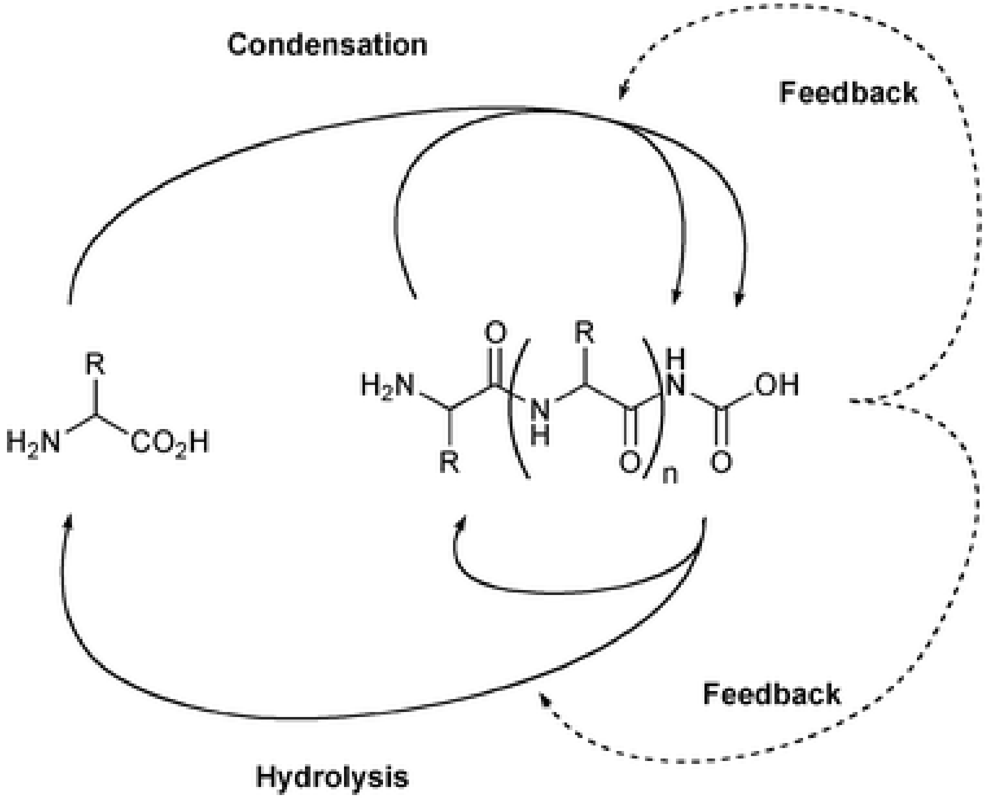


Photo credit: Jenny Mottar, NASA

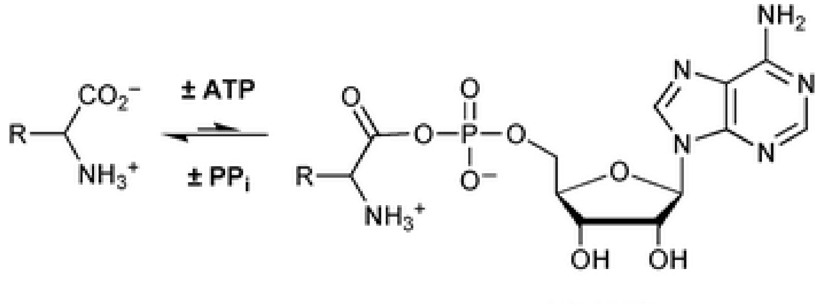
Lecture 5, SoSe 2019 KIT

Zbigniew Pianowski

Condensation of aminoacids into peptides

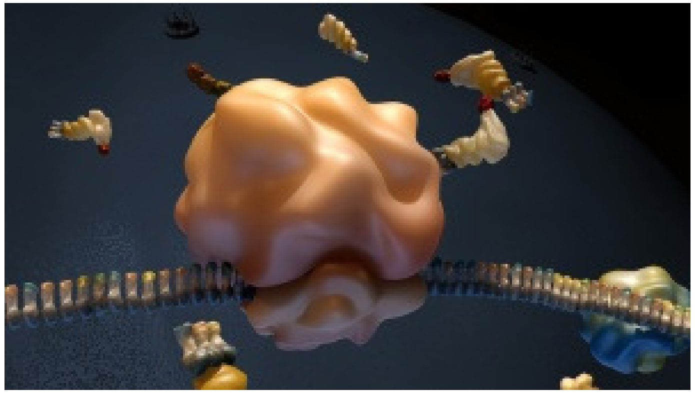


Biochemical condensation of aminoacids into peptides

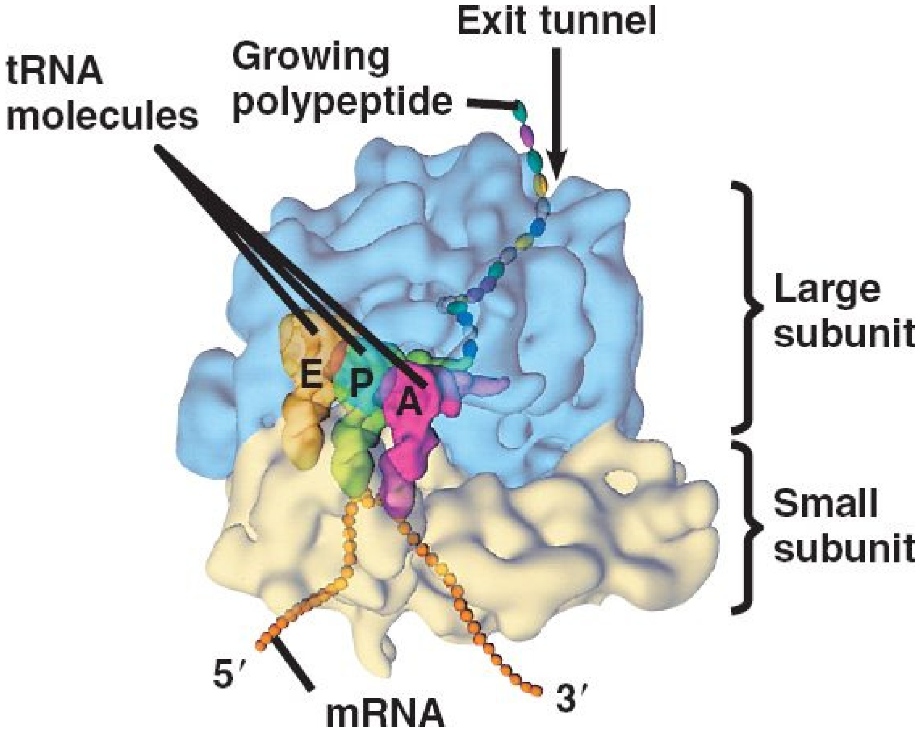


AA-AMP

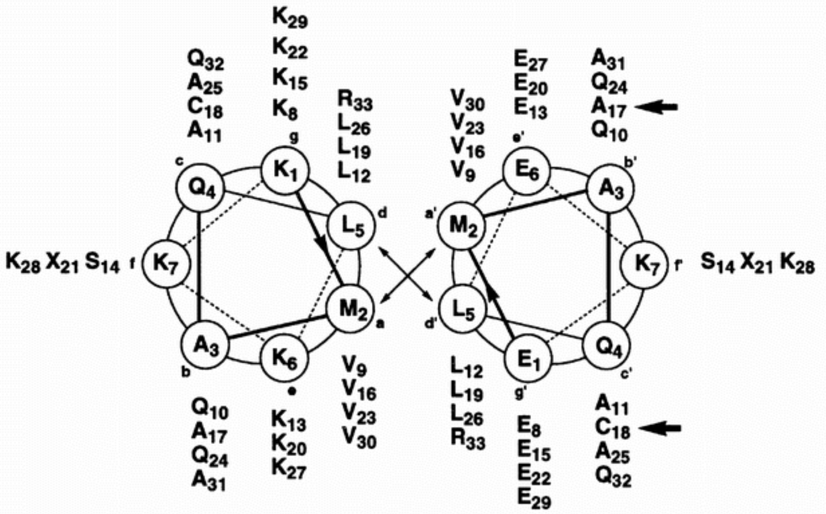
$$\text{AA-AMP} \xrightarrow{+\,\text{tRNA}} \text{AA-tRNA} \xrightarrow{\text{Ribosome}} \text{Peptides}$$

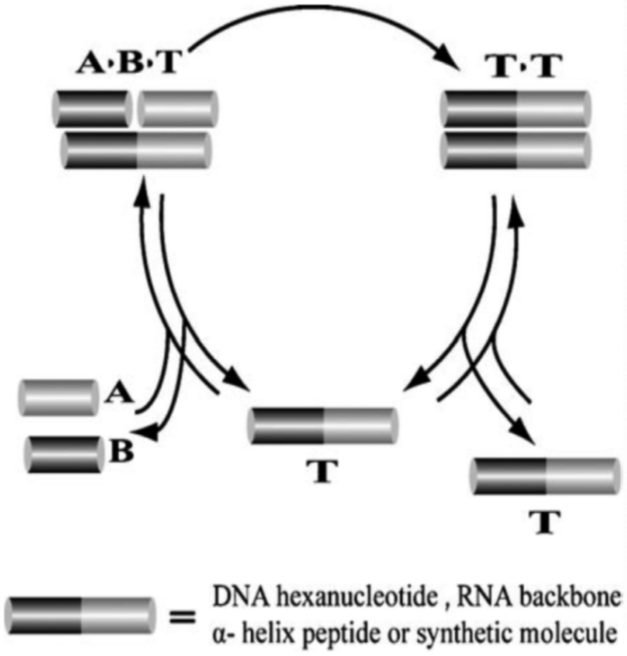


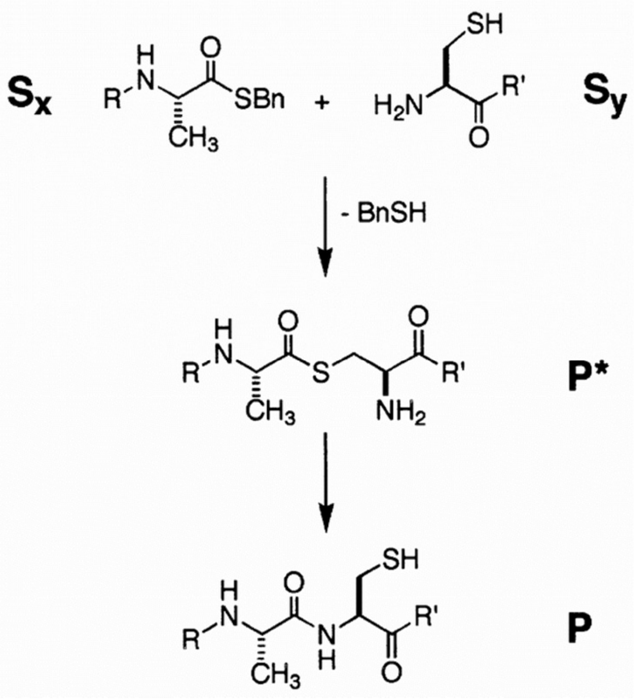
Nature Publishing Group, www.nature.com/nrg/multimedia



Peptide self-replication







E: Ar-**K**MAQL**K**K**K**VQAL**K**S**K**VA-CL**K**X**K**VQAL**K**K**K**VAQR-CONH$_2$

E*: Ar-**K**MAQL**K**K**K**VQAL**K**S**K**VA-SL**K**X**K**VQAL**K**K**K**VAQR-CONH$_2$

S$_1$: Ar-**E**MAQL**E**K**E**VQAL**E**S**E**VA-COSBn

S$_2$: Ar-**E**MAQL**E**K**E**VQAL**E**S**E**VA-COS(CH$_2$)$_2$CONH$_2$

S$_3$: Ar-**E**MAQL**E**K**E**VQAL**E**S**E**VA-CONHOH

S$_4$: Ar-**R**MKQL**E**E**K**VYELLS**K**VA-COSBn

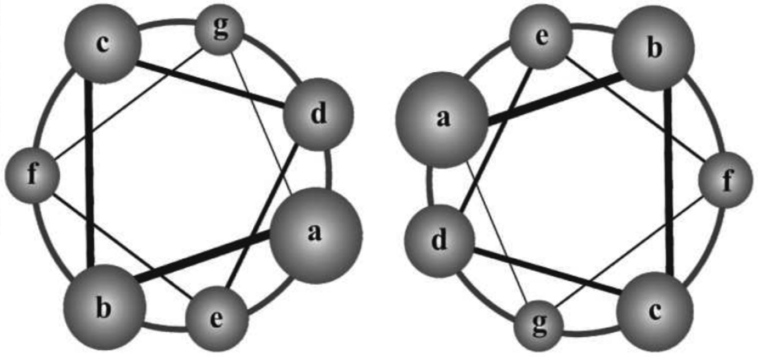
S$_5$: Ar-**K**MAQL**K**K**K**VQAL**K**S**K**VA-COSBn

S$_6$: H$_2$N-CL**E**X**E**VQAL**E**K**E**VAQR-CONH$_2$

S$_7$: H$_2$N-CL**E**X**E**VARL**K**KLVGE-CONH$_2$

S$_8$: H$_2$N-(*d*-C)L**E**X**E**VQAL**E**K**E**VAQR-CONH$_2$

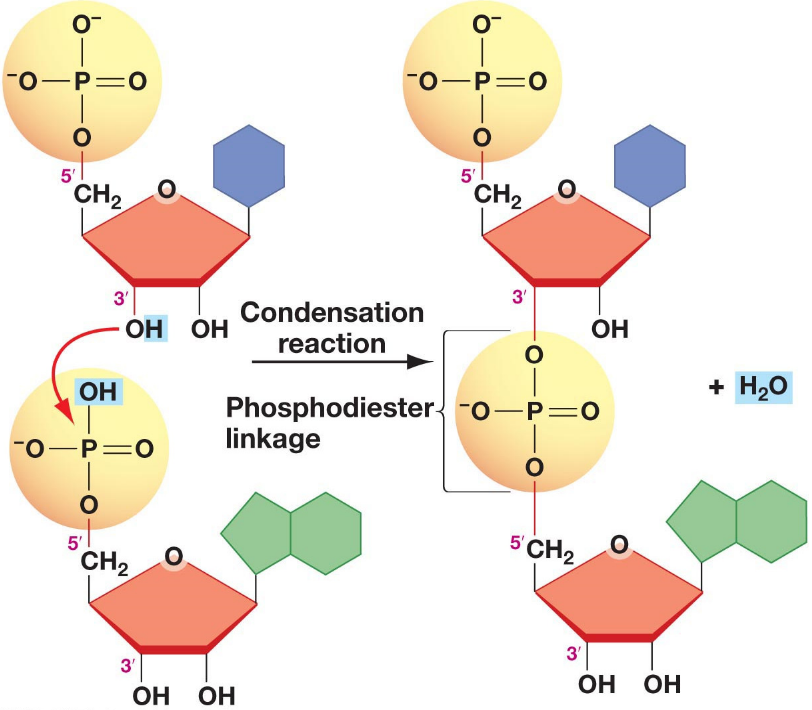
S$_9$: H$_2$N-GL**E**X**E**VARL**K**KLVGE-CONH$_2$



K. Severin, D. H. Lee, A. J. Kennan and M. Reza Ghadiri *Nature* **1997**, *389*, 706-709

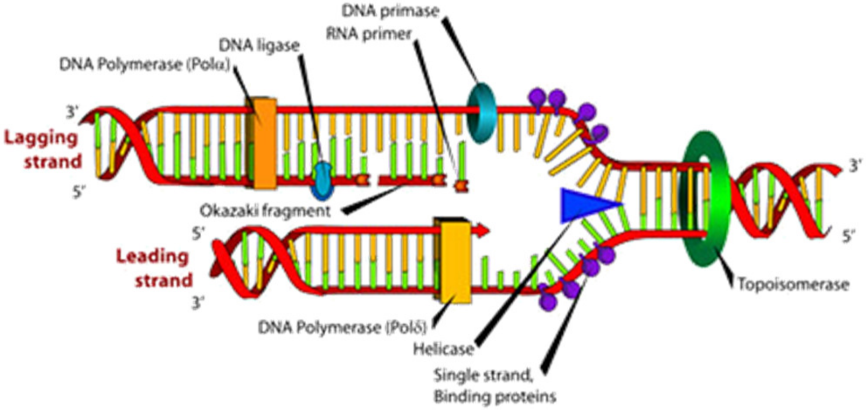
Nucleotide polymerization

Regioselective formation of 3’-5’ phosphodiester bonds between nucleotides

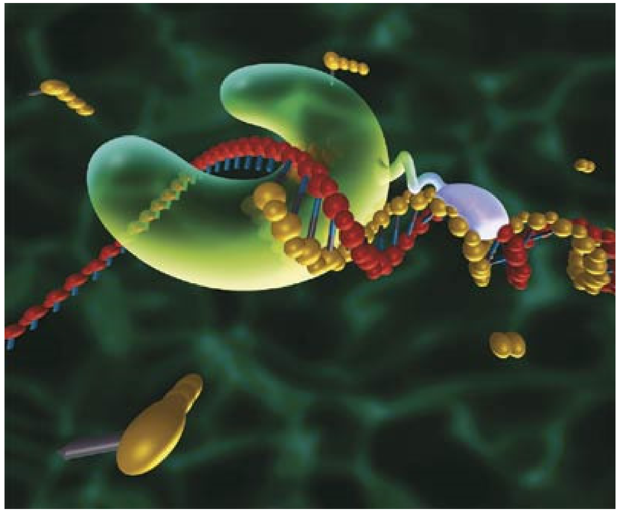


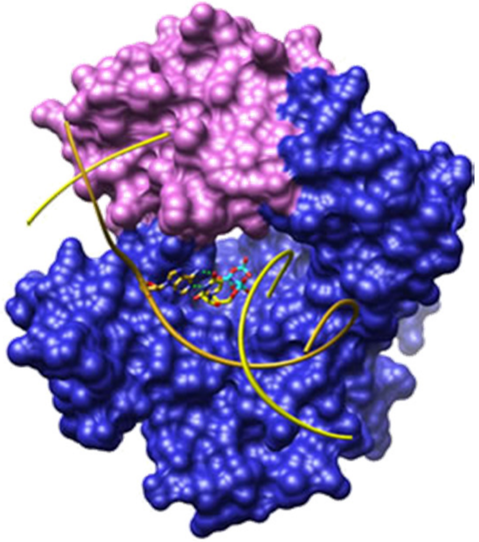
Vital chemical reactions

nucleotide polymerization → DNA/RNA polymerases

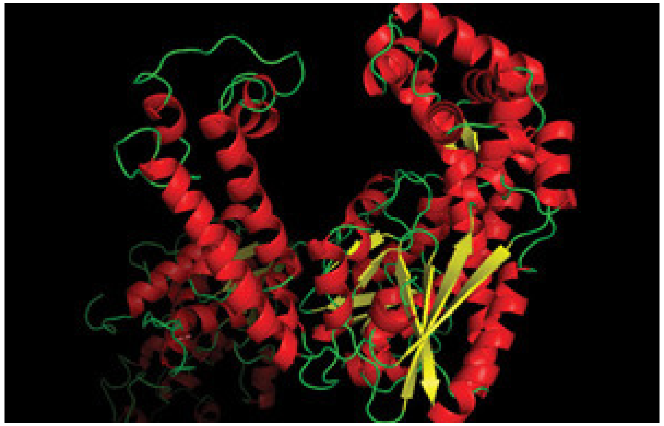


dxline.info/img/new_ail/dna-polymerase_1.jpg



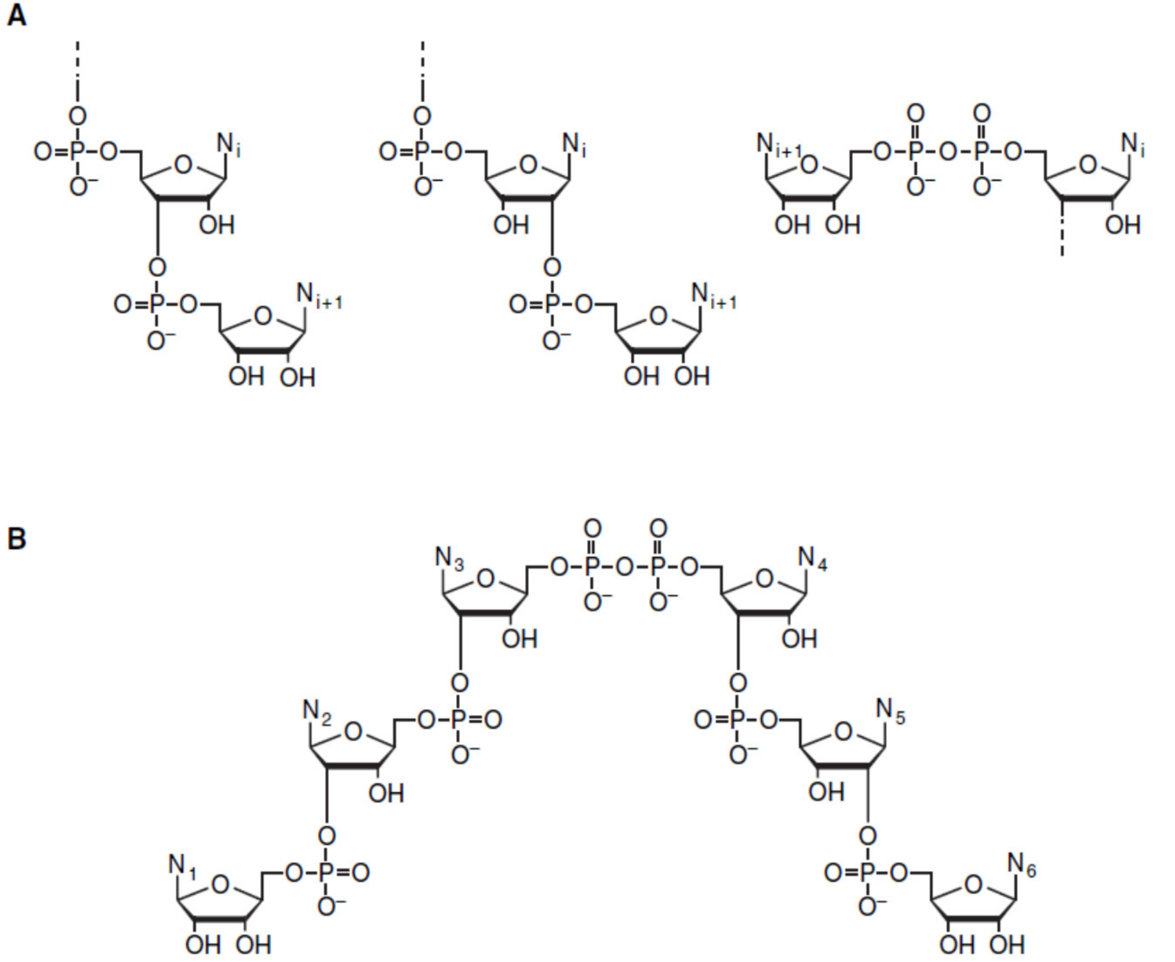


niehs.nih.gov



www.neb.com

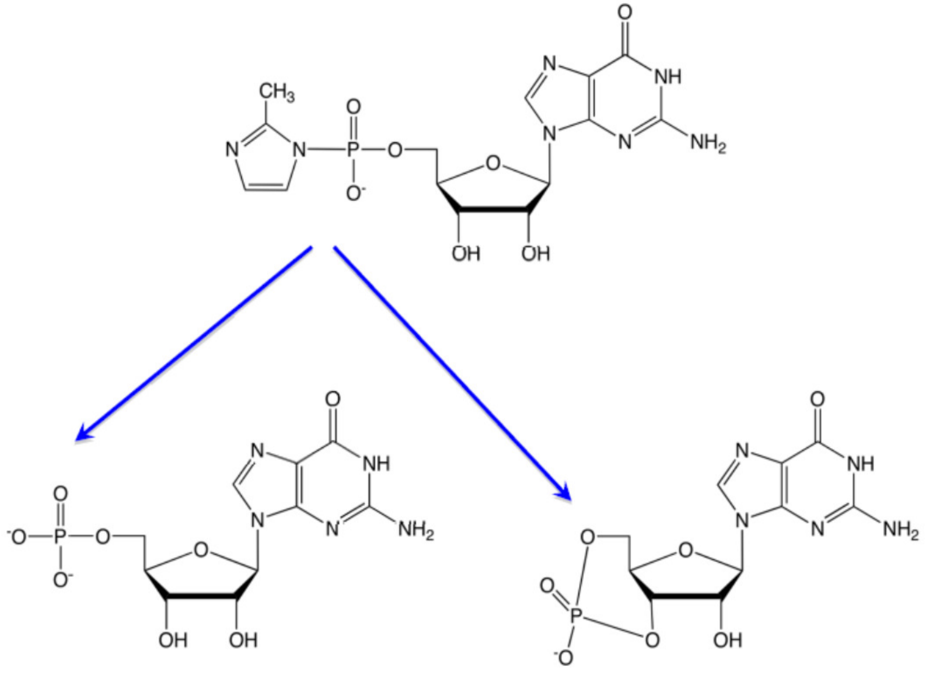
Products of chemical condensation of nucleotides



(**A**) Reaction of an activated mononucleotide (N$_{i+1}$) with an oligonucleotide (N$_1$–N$_i$) to form a 3’,5’-phosphodiester (left), 2’,5’-phosphodiester (middle), or 5’,5’-pyrophosphate linkage (right).

(**B**) Typical oligomeric product resulting from chemical condensation of activated mononucleotides

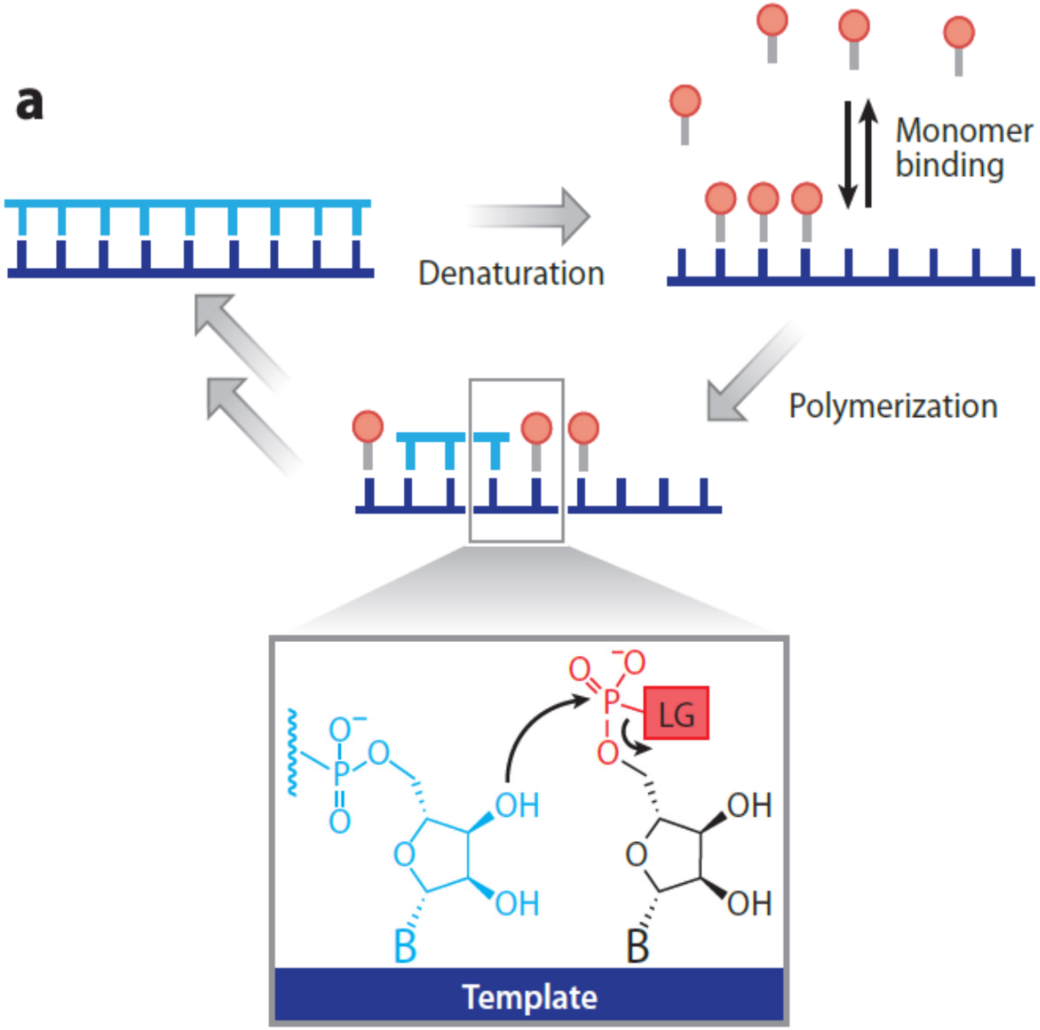
Degradation of activated nucleotides



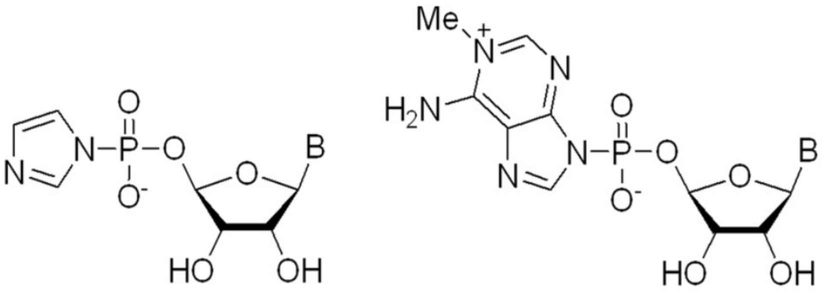
hydrolysis

3’,5’-cyclization

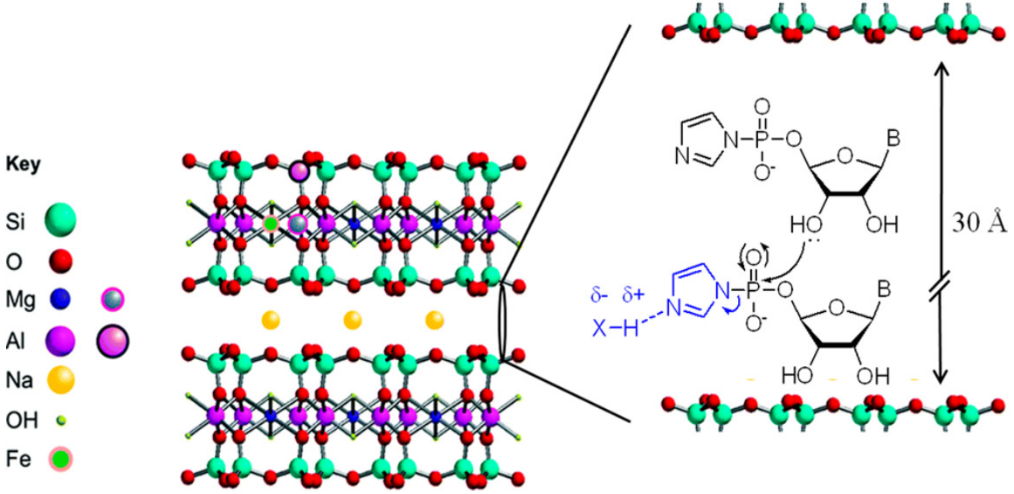
Template-directed synthesis



Montmorillonite



B = adenine, guanine, cytosine or uracil

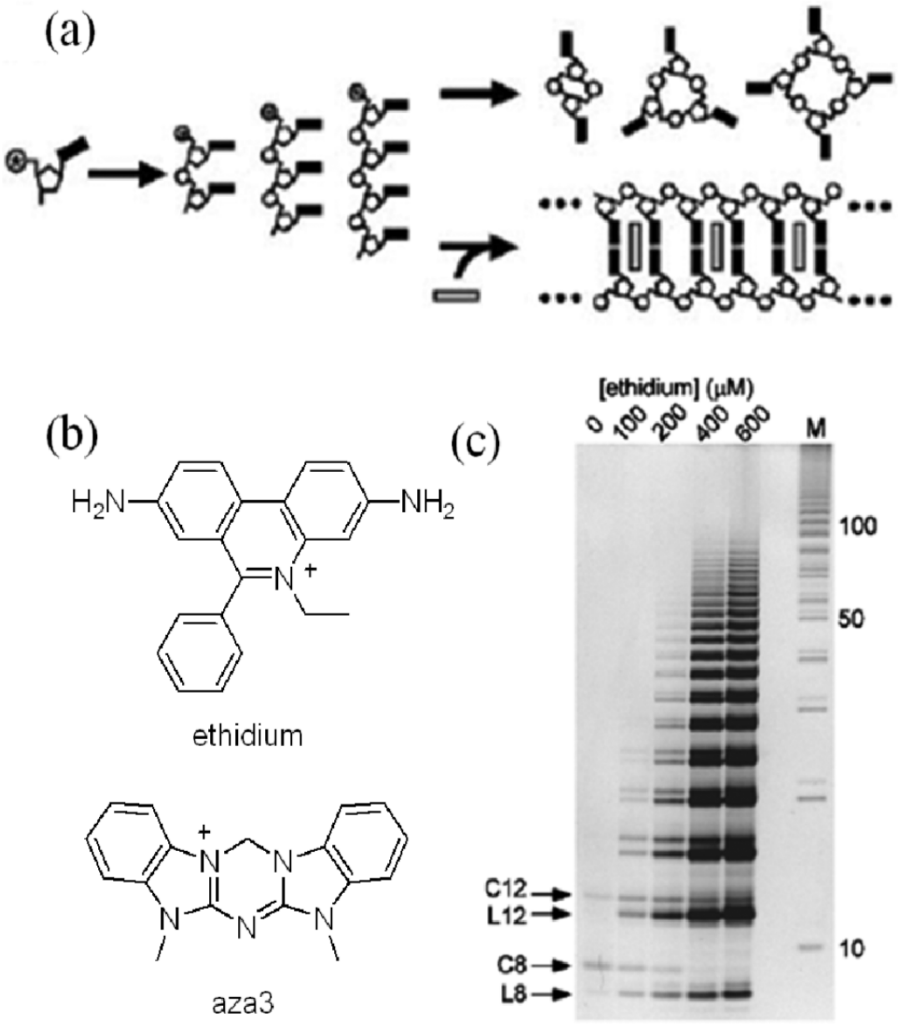


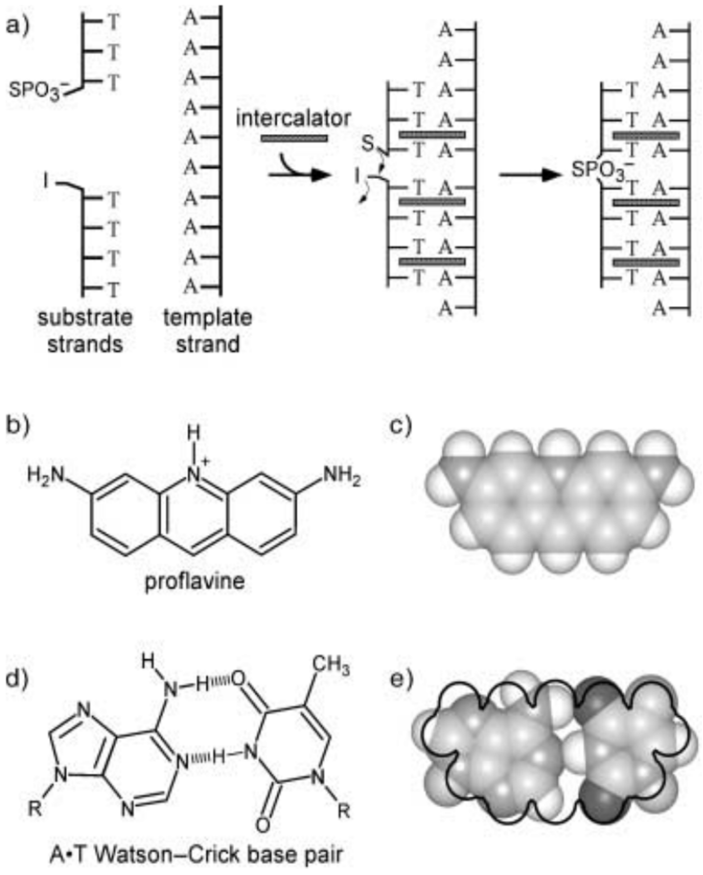
30-50 units successfully oligomerized

(Top) Structure of ribonucleotide 5′-phosphoimidazolides (left) and ribonucleotide 5′-phosphoro-1-methyladeninium (right). (Bottom) Unit cell of montmorillonite and phosphodiester bond formation within the clay interlayers, as proposed by Ferris and coworkers (right). XH, depicted in blue in the cartoon, is any undifferentiated protic species inside the clay galleries.

Joshi, P. C.; Aldersley, M. F.; Delano, J. W.; Ferris, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 13369

Intercalating agents



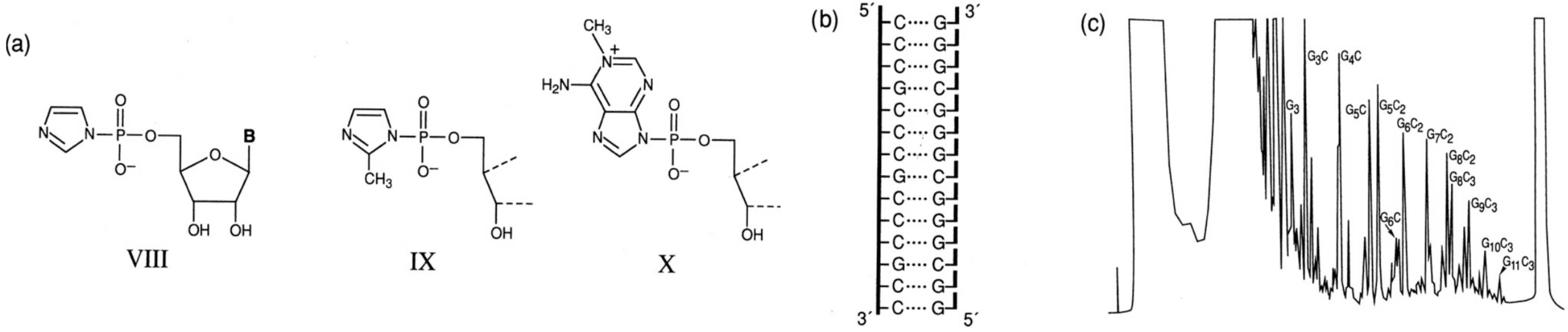


Rate increase by three orders of magnitude vs. ligation without proflavine

N. V. Hud *et al. Angew. Chem. Int. Ed.* **2004**, *43*, 2004 –2008

Template-directed synthesis

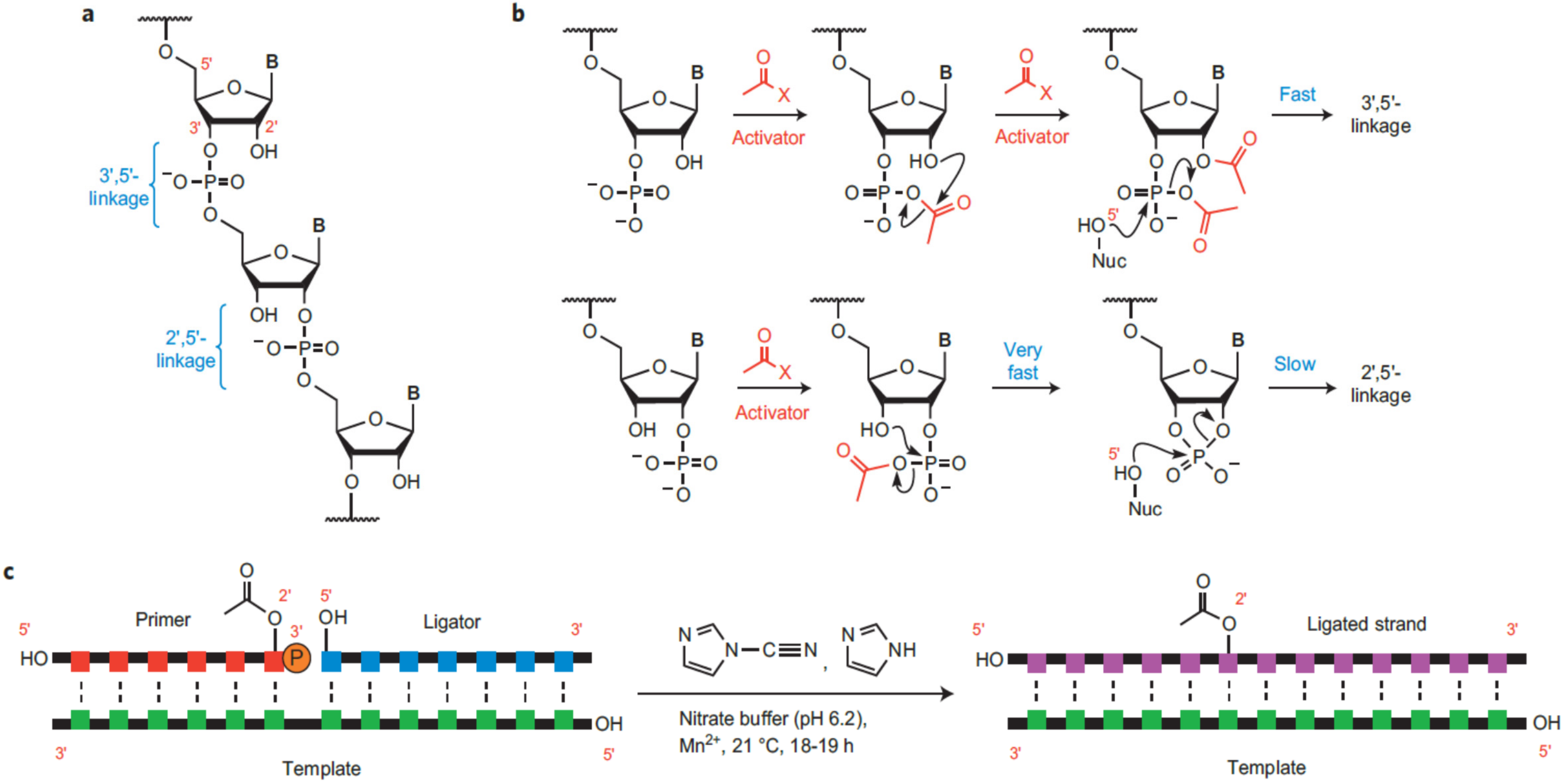
no example demonstrated yet, where single activated nucleotides would form a complementary strand on an RNA (or DNA) template without enzymatic support



Current experiments focus on ‚primer extention' or ‚filling abasic sites'– sequence-selective complementary nucleobase addition to a pre-existing strand (or between two pre-existing strands) already hybridized on a template. Here, pre-organization provided by the existing base-pairing network supports selection of the correct nucleoside to be joined.

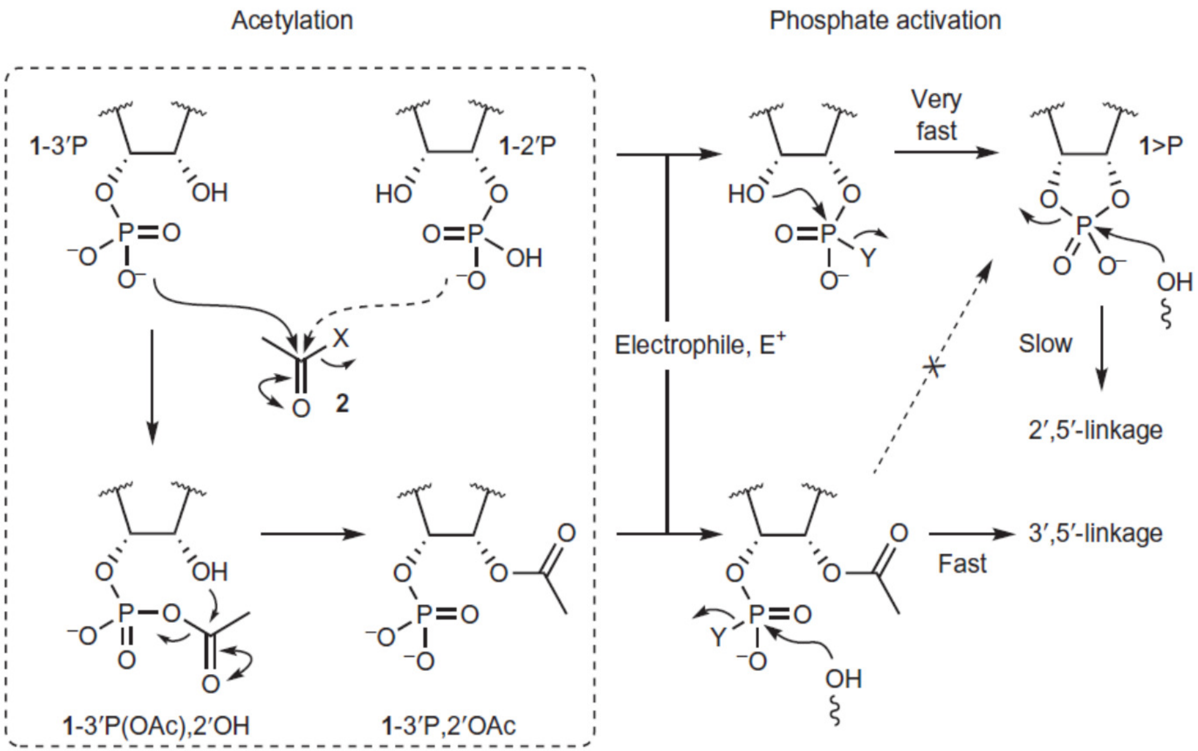
Complementary approaches are regioselective ligation reactions of short oligonucleotides on templates, or dynamic covalent chemistry, where nucleobase-containing components would be added sequence-specifically to a pre-existing ‚empty' backbone on a template

Regioselective ribonucleotide ligation



J. Sutherland *et al. Nature Chem.* **2013,** 383-389

Regioselective ribonucleotide ligation

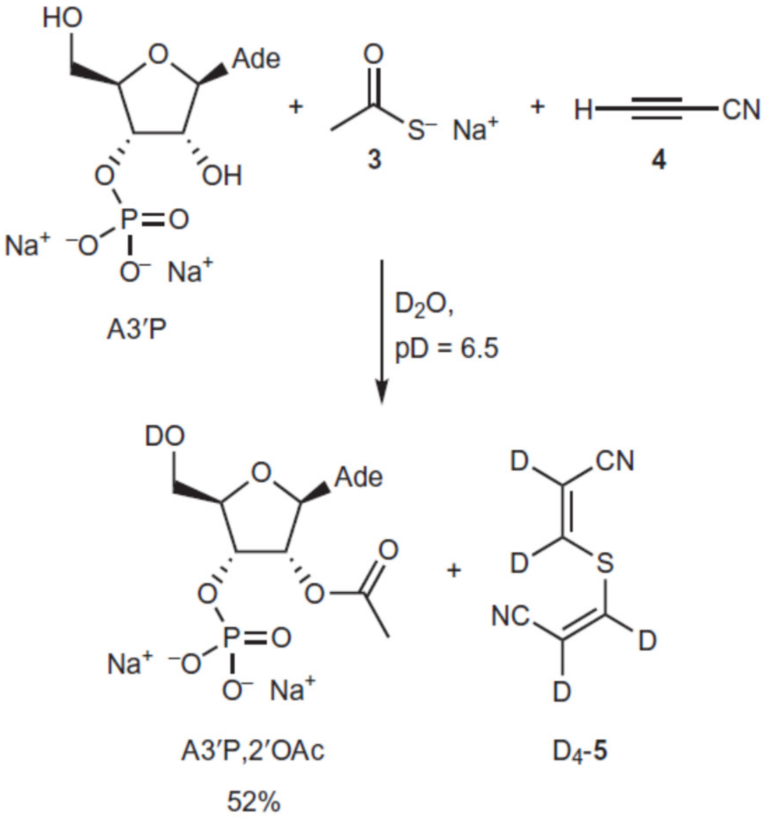


Protection of the 2’-OH group of **1**-3’P facilitates rapid template-directed 3’,5’-ligation after electrophilic phosphate activation. The 3’-OH group of **1**-2’P is protected to a lesser extent, such that **1**>P is the major product of phosphate activation and slow template-directed 2’,5’-ligation follows.

X = leaving group, Y = leaving group generated by electrophilic activation of phosphate oxygen with or without a subsequent nucleophilic displacement

J. Sutherland *et al. Nature Chem.* **2013,** 383-389

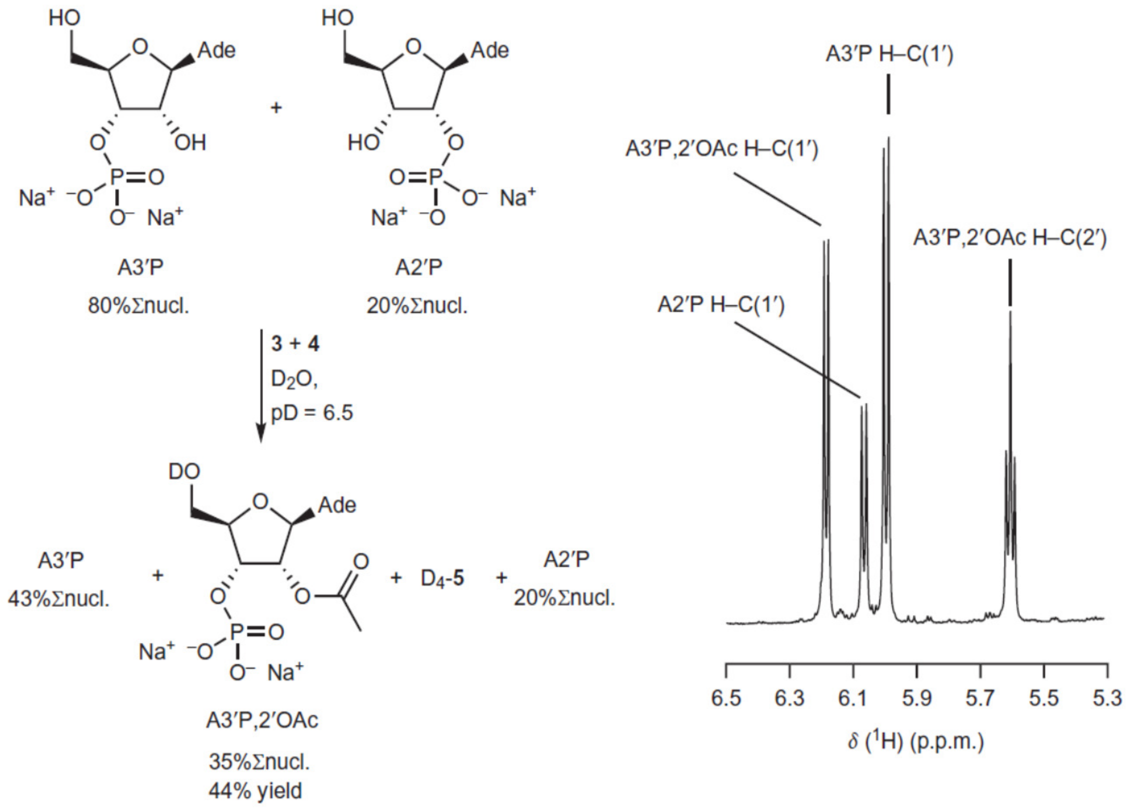
Regioselective ribonucleotide ligation



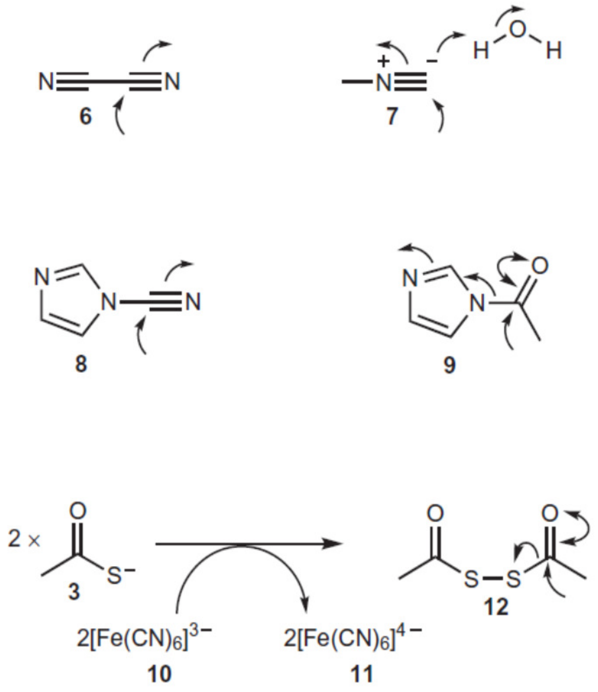
Treatment of adenosine-3’phosphate (A3’P) (100 mM) with sodium thioacetate 3 (100 mM) and cyanoacetylene 4 (200 mM) in D_2O at neutral pD for 24 hours results in selective acetylation of the 2-OH group.

J. Sutherland *et al. Nature Chem.* **2013,** 383-389

Regioselective ribonucleotide ligation



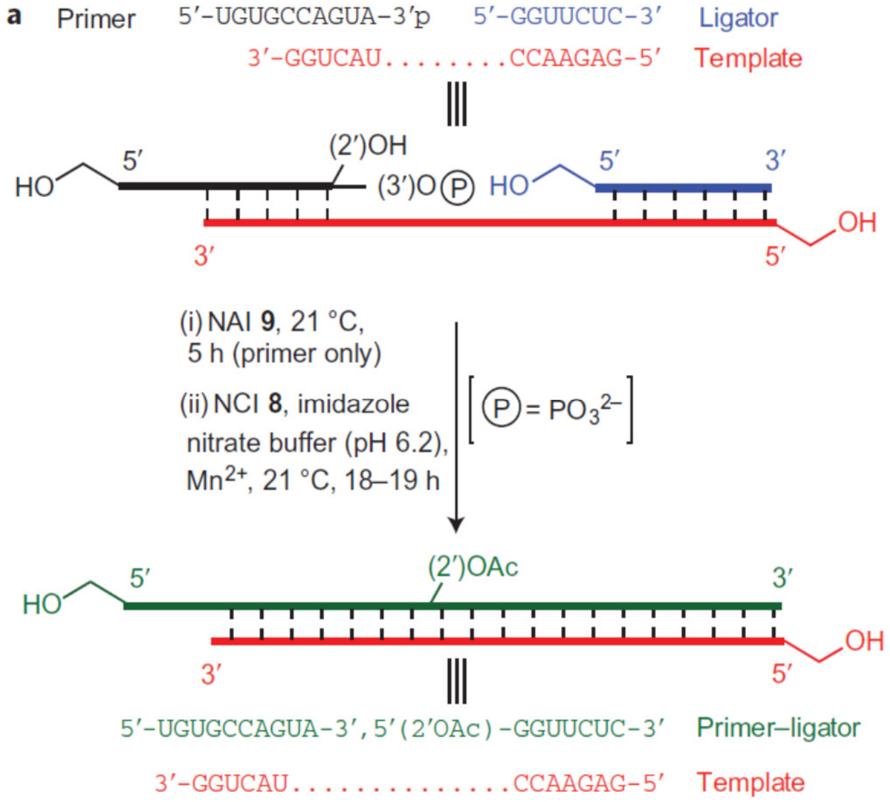
Treatment of A3’P (80 mM) and A2’P (20 mM) as given before results in the exclusive 2-acetylation of the former nucleotide. Partial ^{1}H NMR spectrum of the reaction products.



Additional electrophiles 6–8 shown to drive the acetylation of ribonucleotides with thioacetate 3. Direct acetylation with 9 is also possible, as is oxidative activation of 3 with ferricyanide 10 to afford ferrocyanide 11 and a dimeric acetylating agent 12. Curly arrows indicate electrophilic activation/acetylation steps.

Regioselective ribonucleotide ligation

Chemoselective acetylation of 3’P-oligoribonucleotides expedites templated ligation



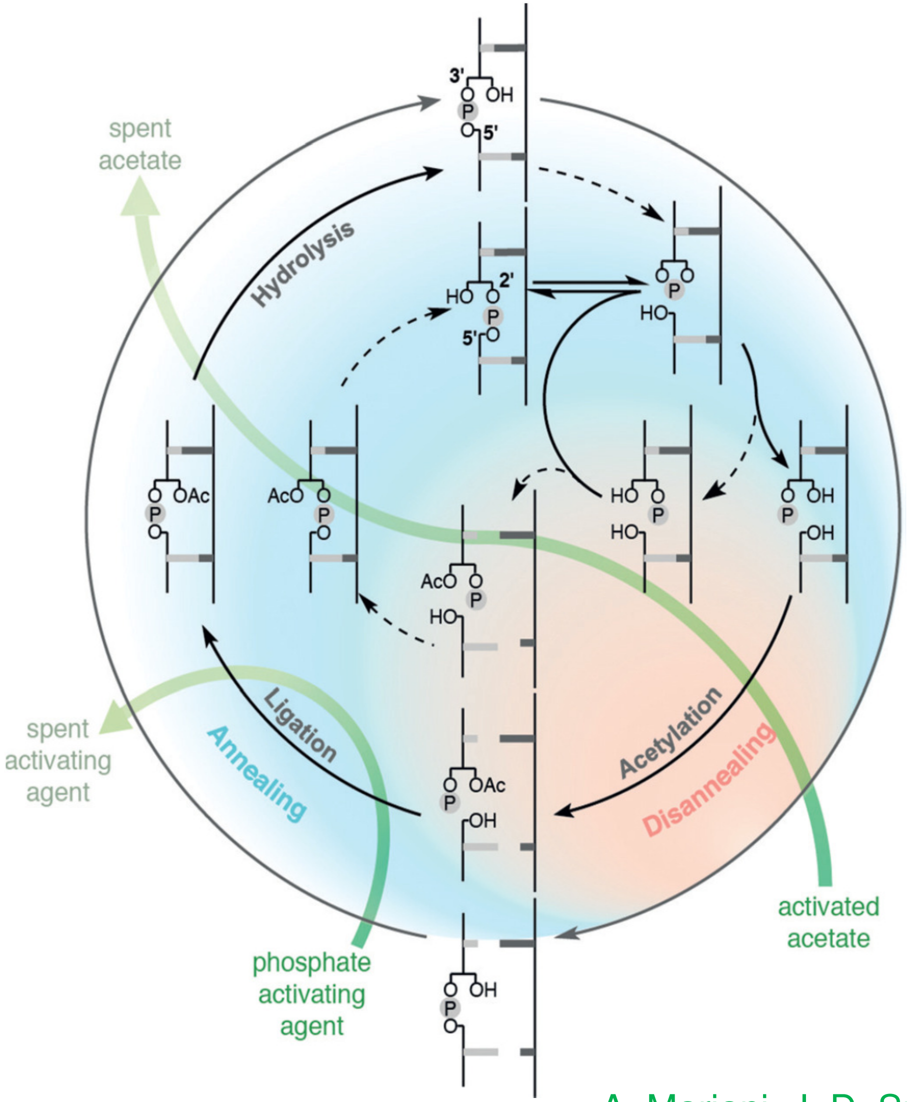
NAI – *N*-acetylimidazole

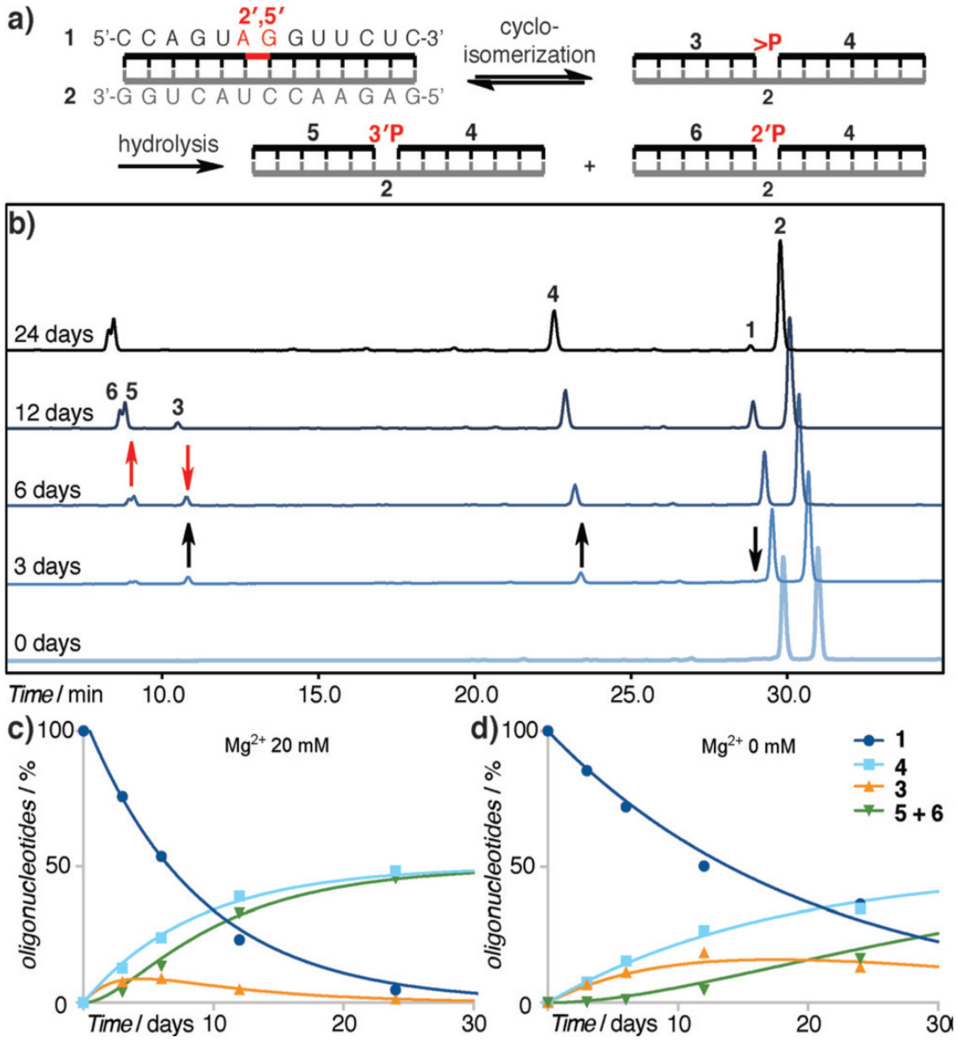
NCI – *N*-cyanoimidazole

Sequences and reaction conditions employed for acetylation (i) and subsequent templated ligation (ii).
The acetylation mixture contained 80 mM primer and 50 mM NAI; the ligation mixture contained 4 mM primer from the acetylation reaction, 25 mM template, 30 mM ligator, 200 mM imidazole nitrate buffer (pH 6.2), 10 mM $MnCl_2$ and 100 mM NCI.

J. Sutherland *et al. Nature Chem.* **2013,** 383-389

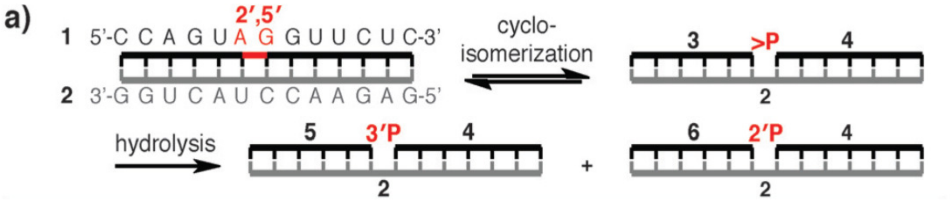
Correction mechanism 2'-5' → 3',5'



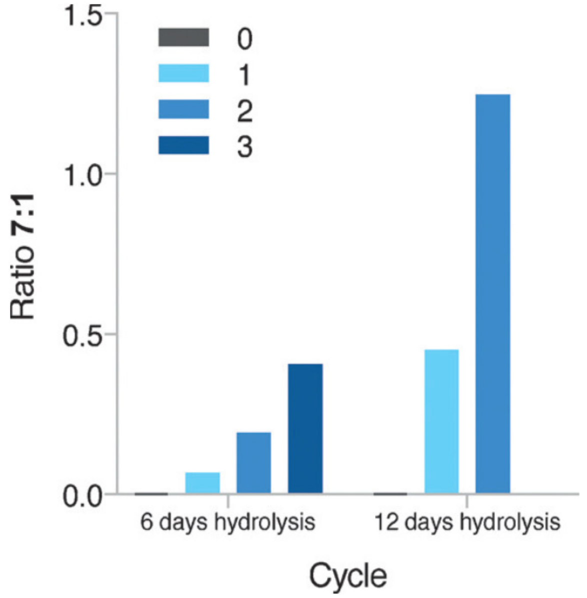


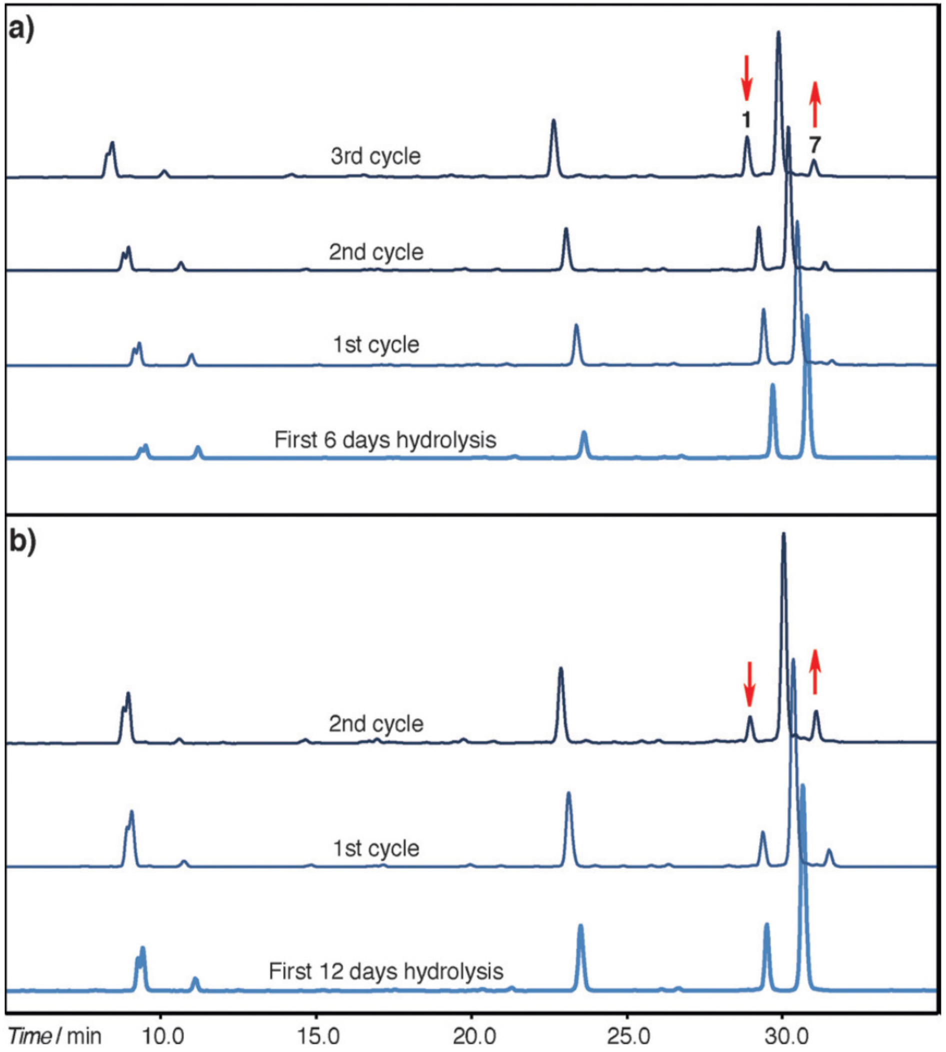
A. Mariani, J. D. Sutherland *Angew. Chem. Int. Ed.* **2017**, *56*, 6563-6566

Correction mechanism 2'-5' → 3',5'



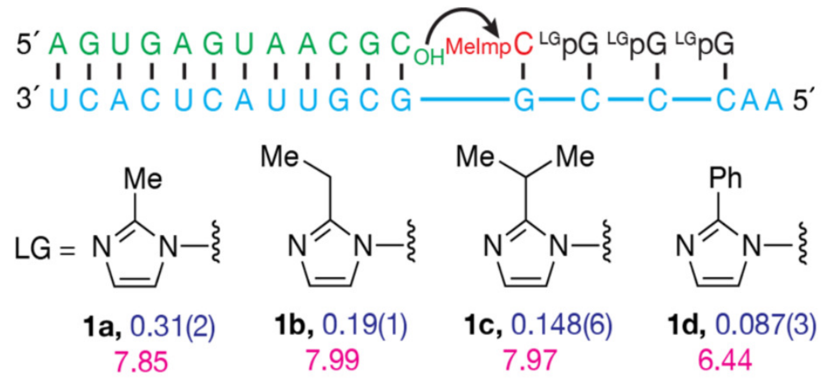
1: full 2',5' link
7: full 3'5' link





A. Mariani, J. D. Sutherland *Angew. Chem. Int. Ed.* **2017**, *56*, 6563-6566

Nonenzymatic primer extension

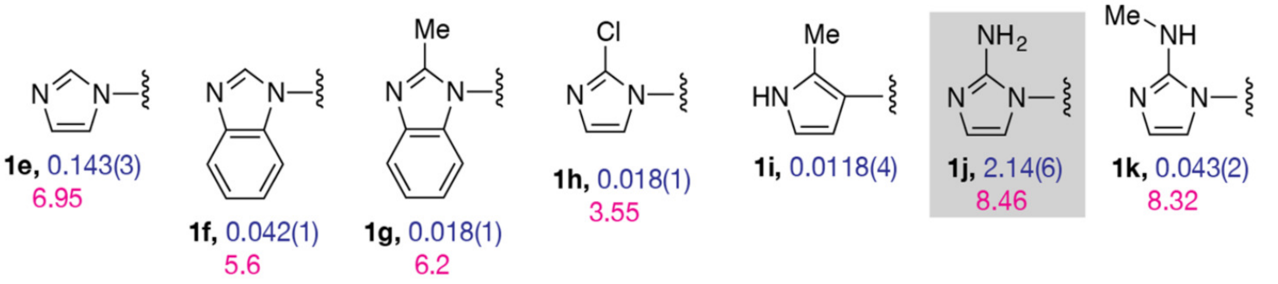


1a, 0.31(2)
7.85

1b, 0.19(1)
7.99

1c, 0.148(6)
7.97

1d, 0.087(3)
6.44



1e, 0.143(3)
6.95

1f, 0.042(1)
5.6

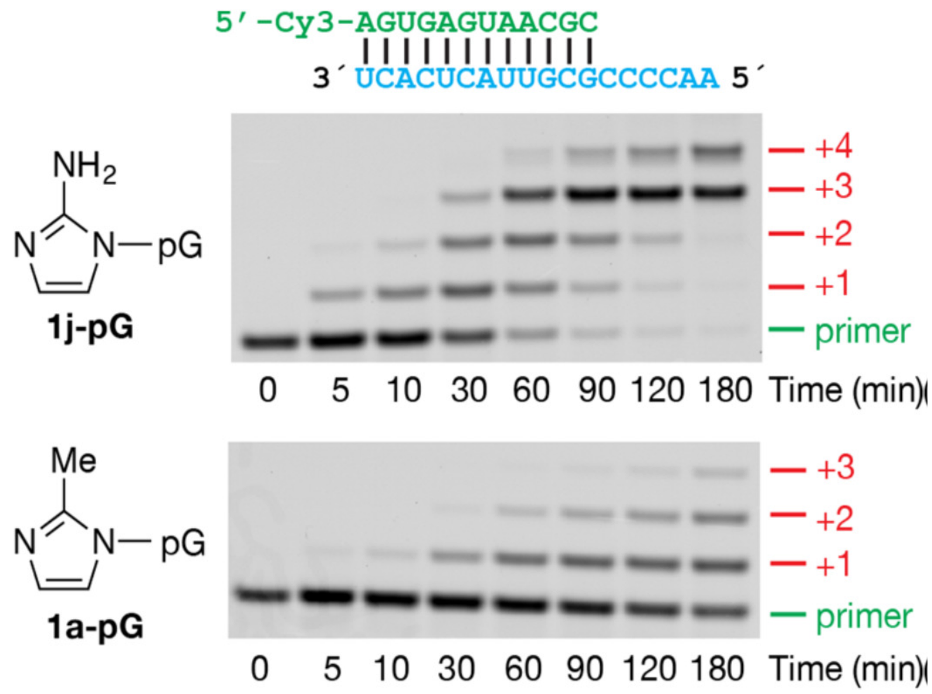
1g, 0.018(1)
6.2

1h, 0.018(1)
3.55

1i, 0.0118(4)

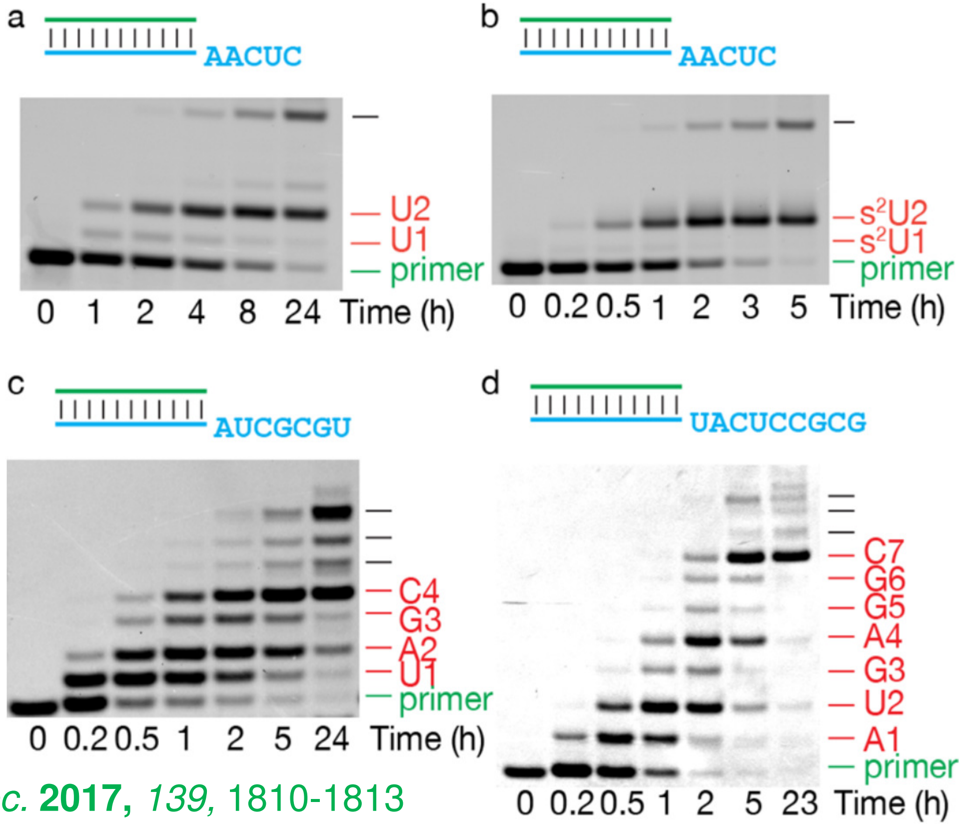
1j, 2.14(6)
8.46

1k, 0.043(2)
8.32



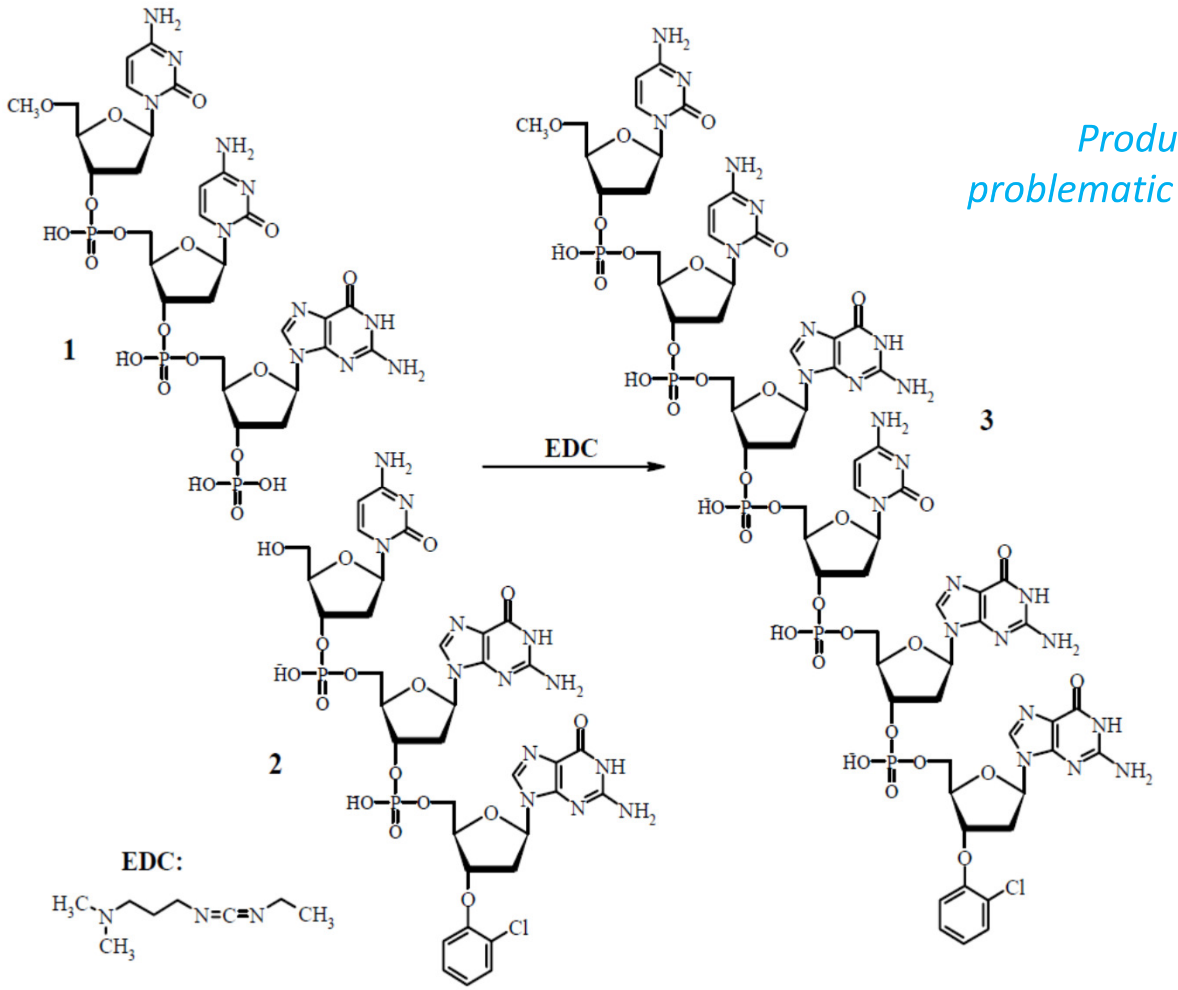
1j-pG

1a-pG



J. Szostak *et al. J. Am. Chem. Soc.* **2017,** *139,* 1810-1813

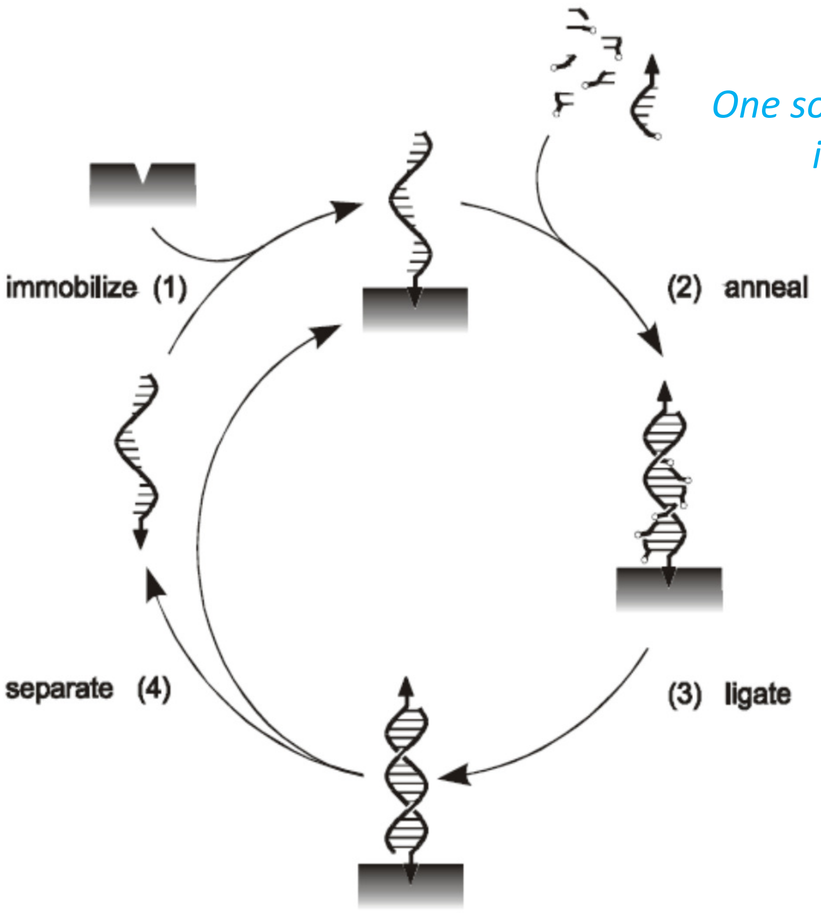
First non-enzymatic self-replicating system



Product inhibition –
problematic for longer templates

V. Patzke, G. von Kiedrowski *ARKIVOC* **2007** 293-310

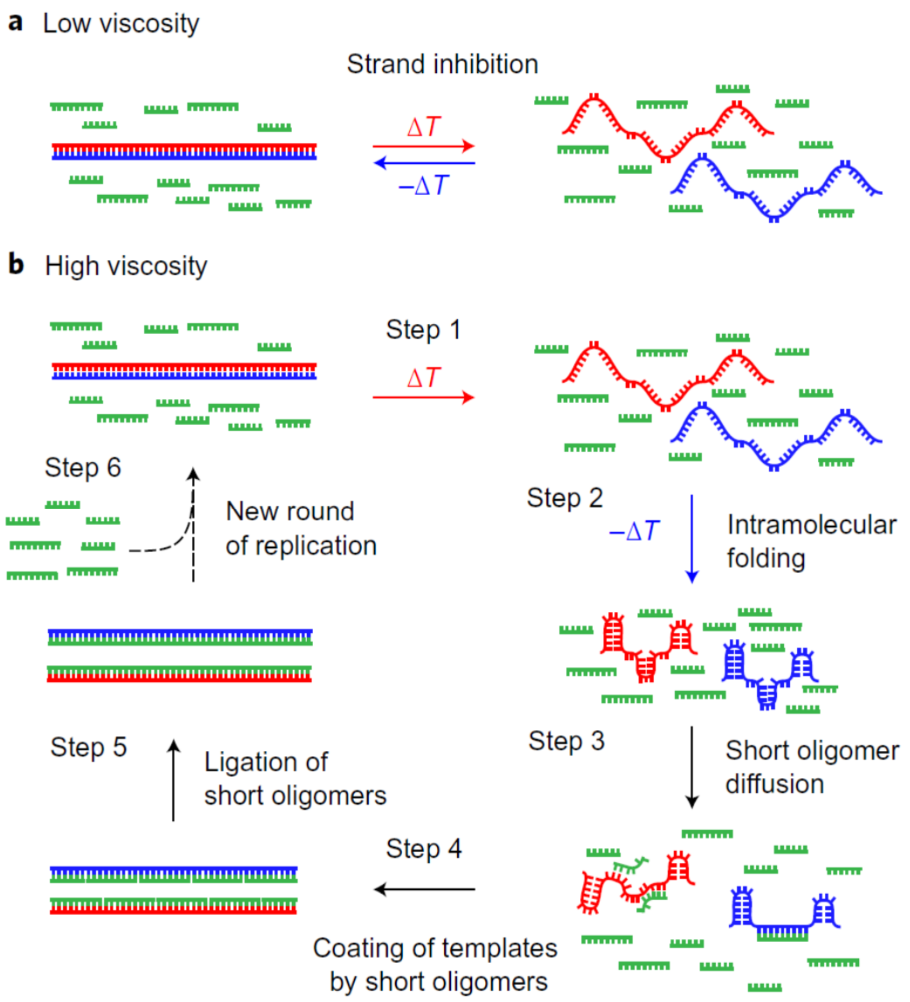
SPREAD – Surface-Promoted Replication and Exponential Amplification of DNA Analogues



One solution for the product inhibition – immobilization of one strand

V. Patzke, G. von Kiedrowski *ARKIVOC* **2007** 293-310

Prebiotic replication in a viscous solvent

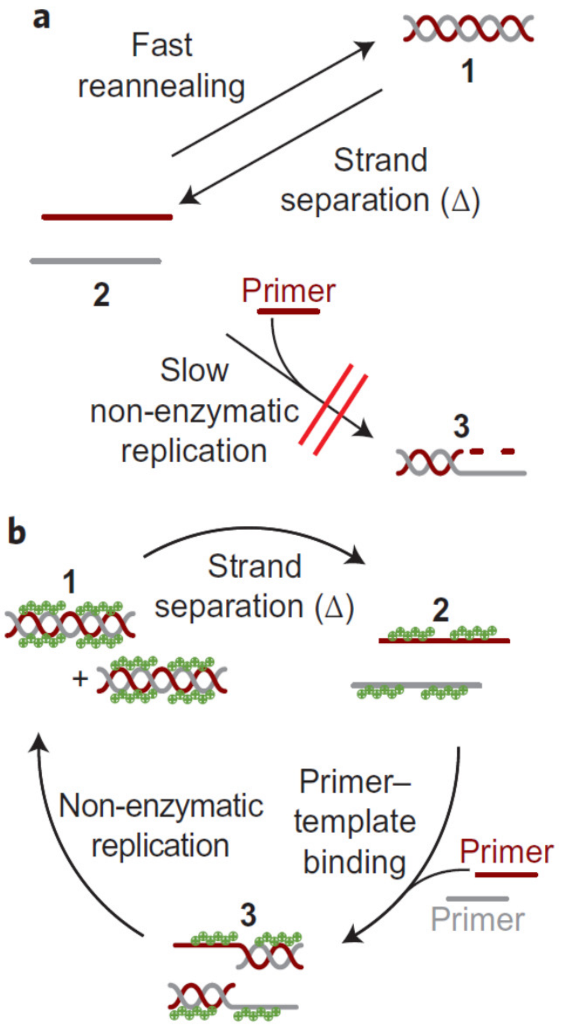


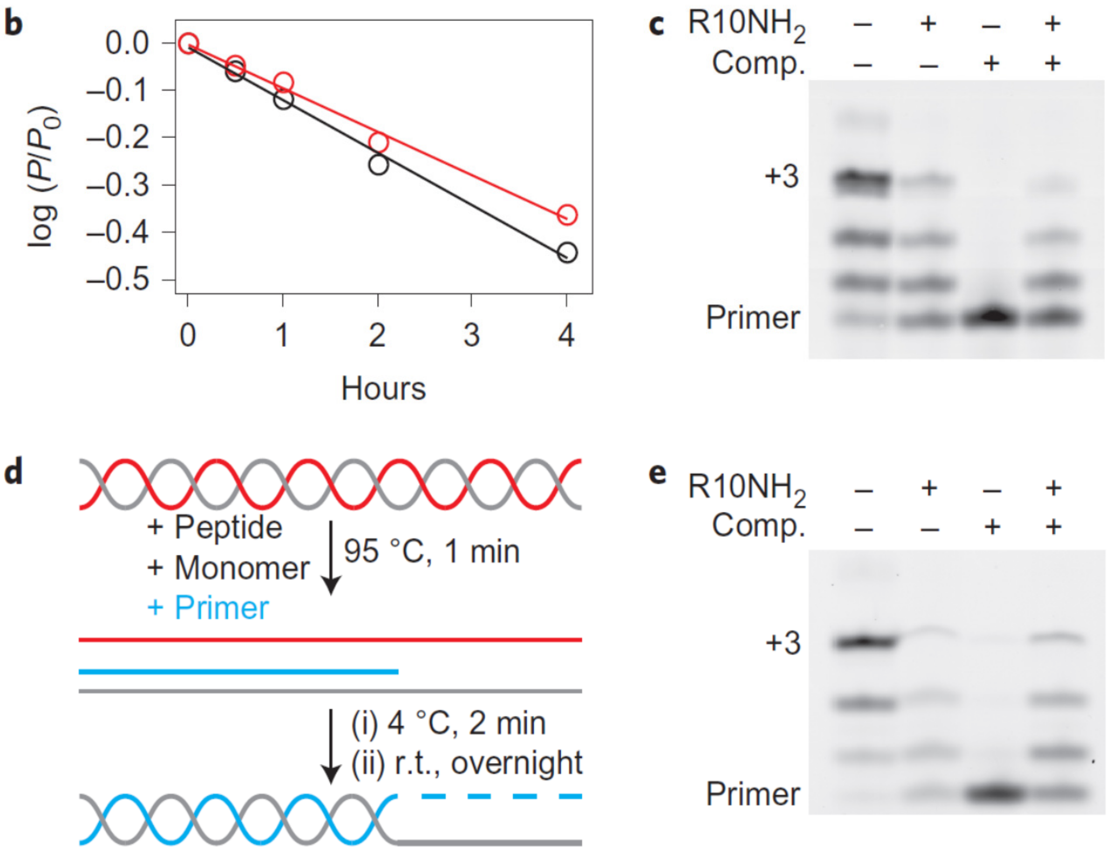
At low viscosity: warming causes duplex dissociation cooling – direct hybridization

At high viscosity: warming causes duplex dissociation cooling – intramolecular folding which prevents back-hybridization
Then slowly short oligomers diffuse into the folds and coat the templates preventing their re-annealing.
The following templated ligation of the short oligomers provides new generation of templates without product inhibition

N. V. Hud *et al. Nature Chem.* **2017**, *9*, 318-324

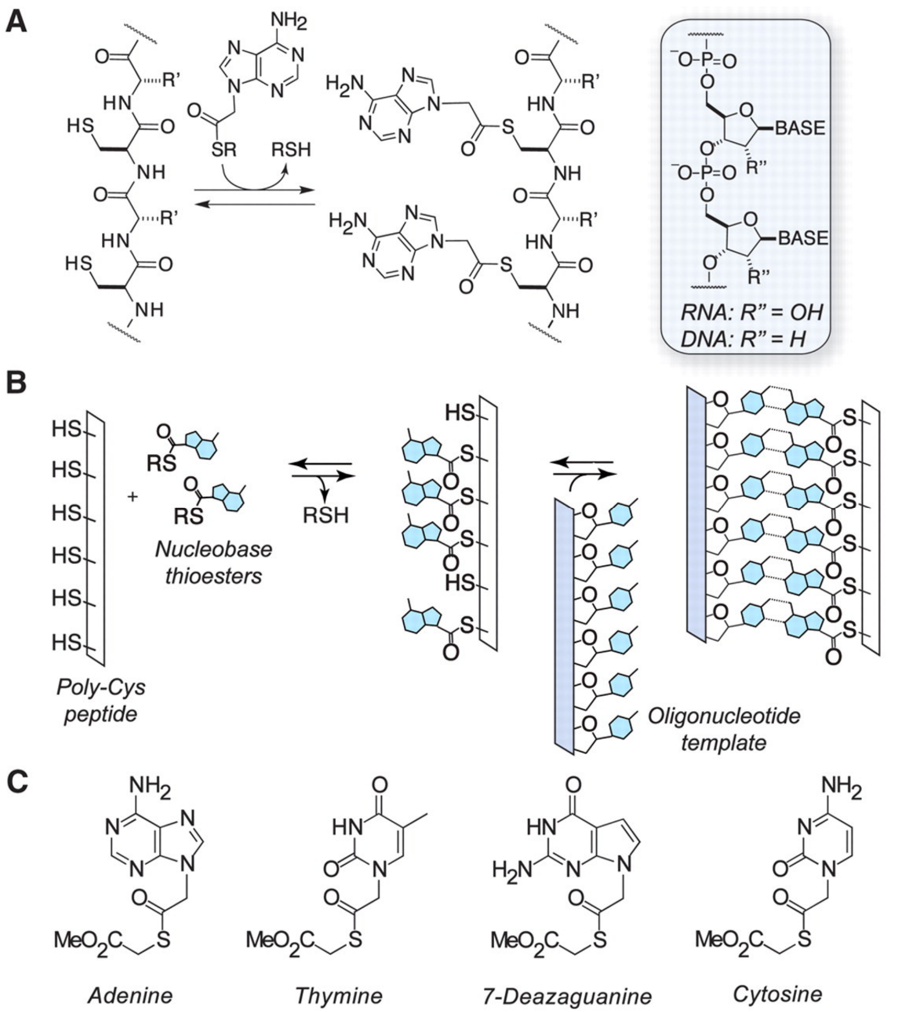
Nonenzymatic primer extension in presence of oligoarginine peptides

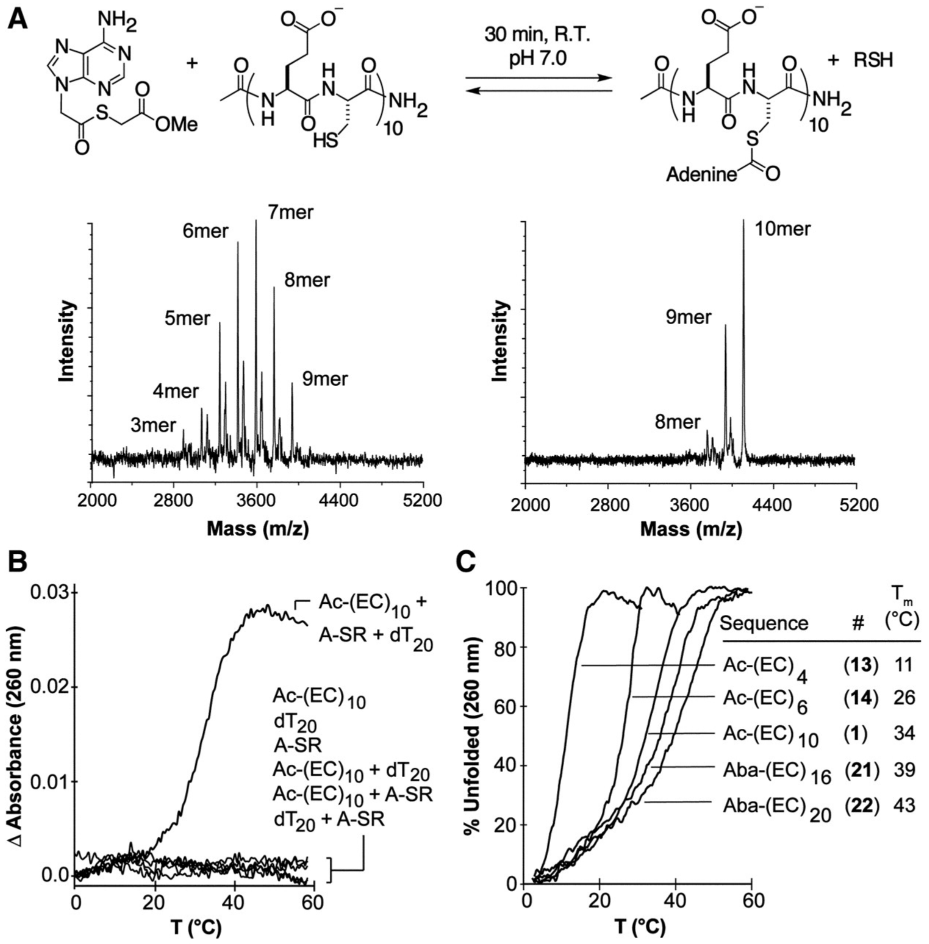




J. Szostak *et al. Nature Chem.* **2016,** *8,* 915-921

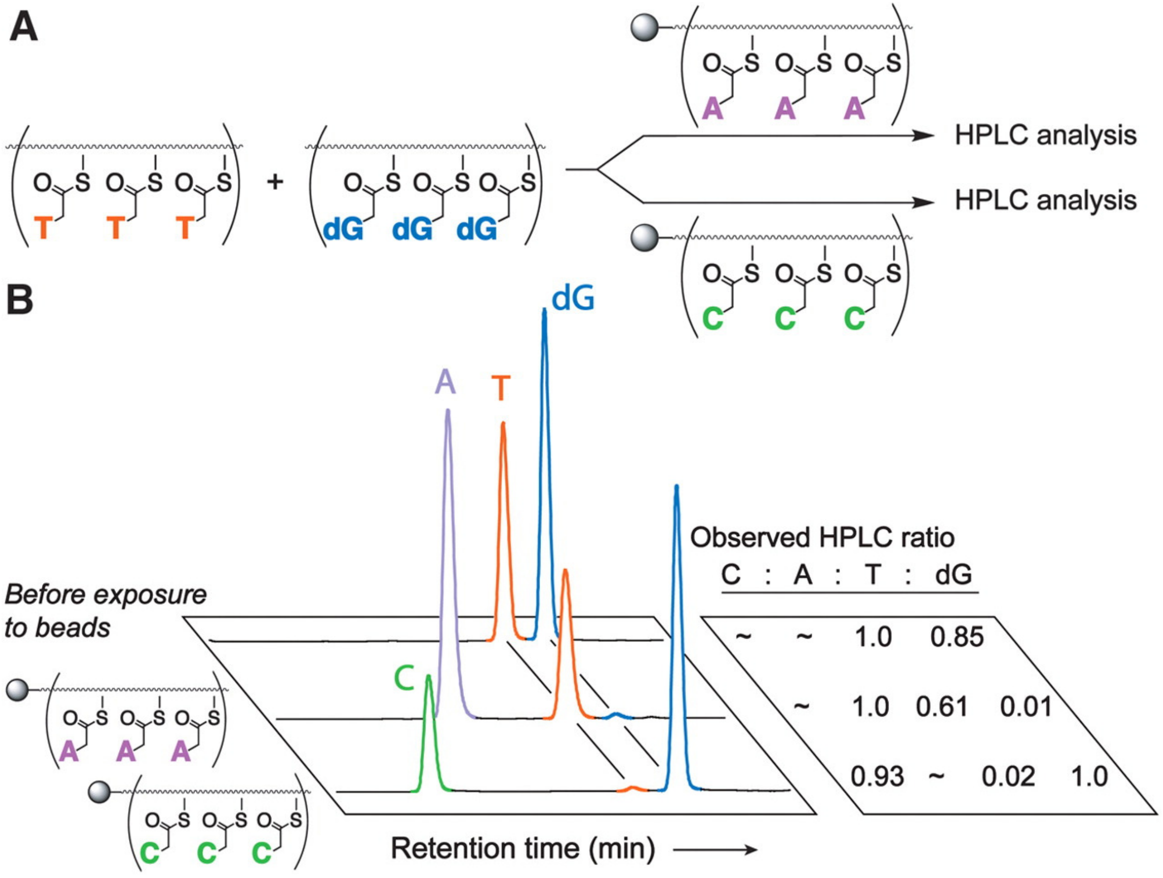
Dynamic oligonucleotide analogue sequence-specific assembly

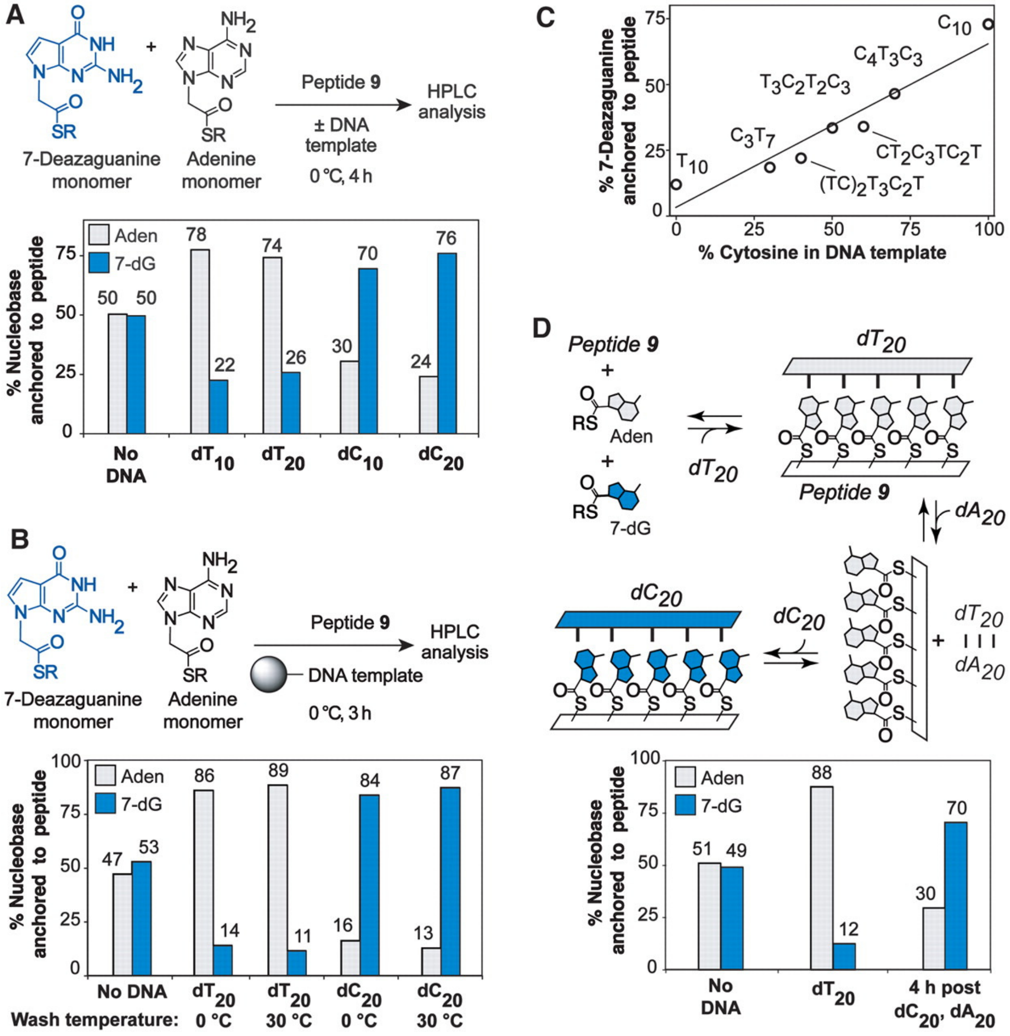




M. R. Ghadiri *et al. Science* **2009,** *325*, 73-77

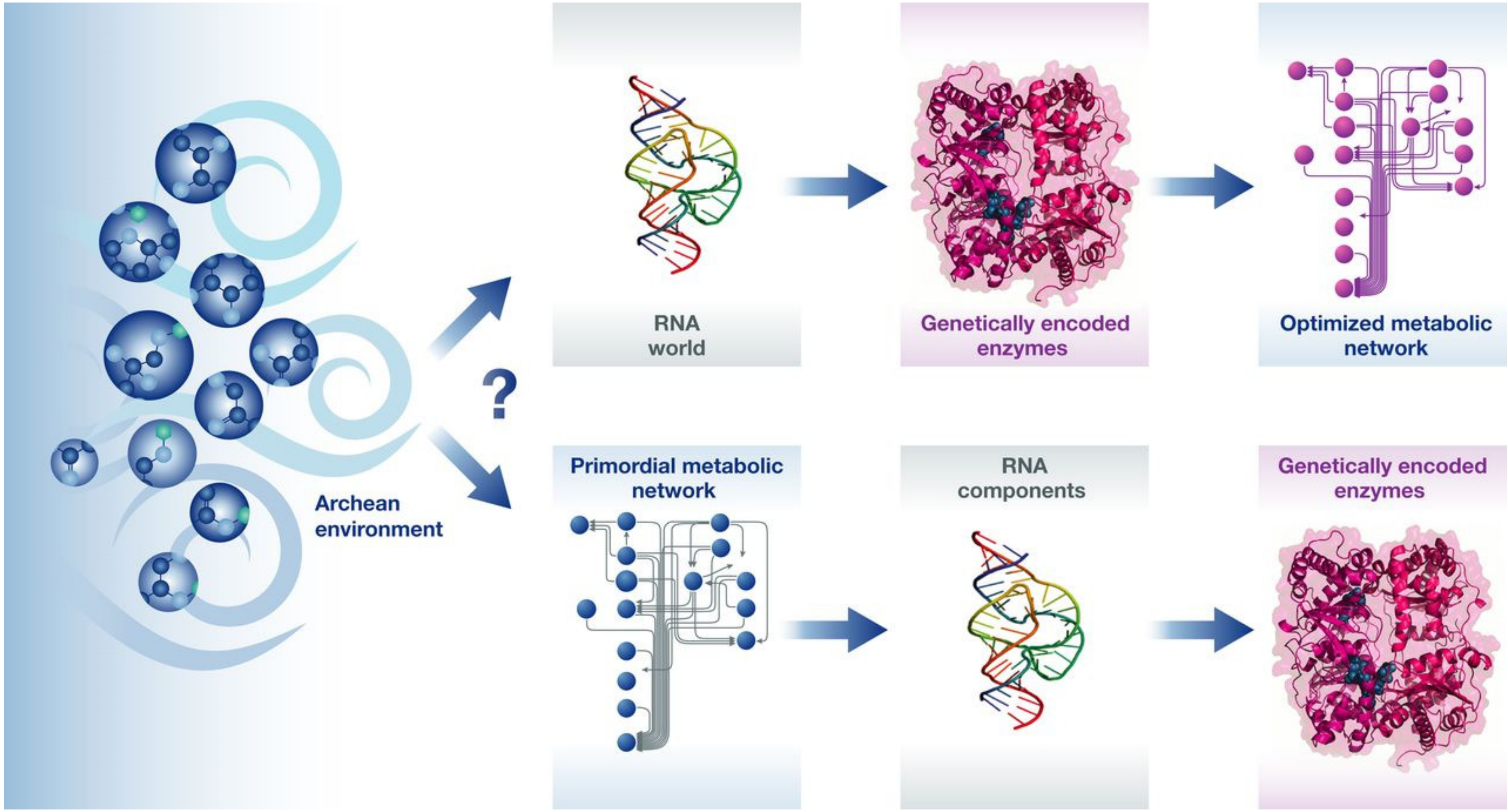
Dynamic oligonucleotide analogue sequence-specific assembly





M. R. Ghadiri *et al. Science* **2009,** *325*, 73-77

Route to life by chemical networks



P. L. Luisi *Mol Syst Biol.* **2014,** *10*, 729

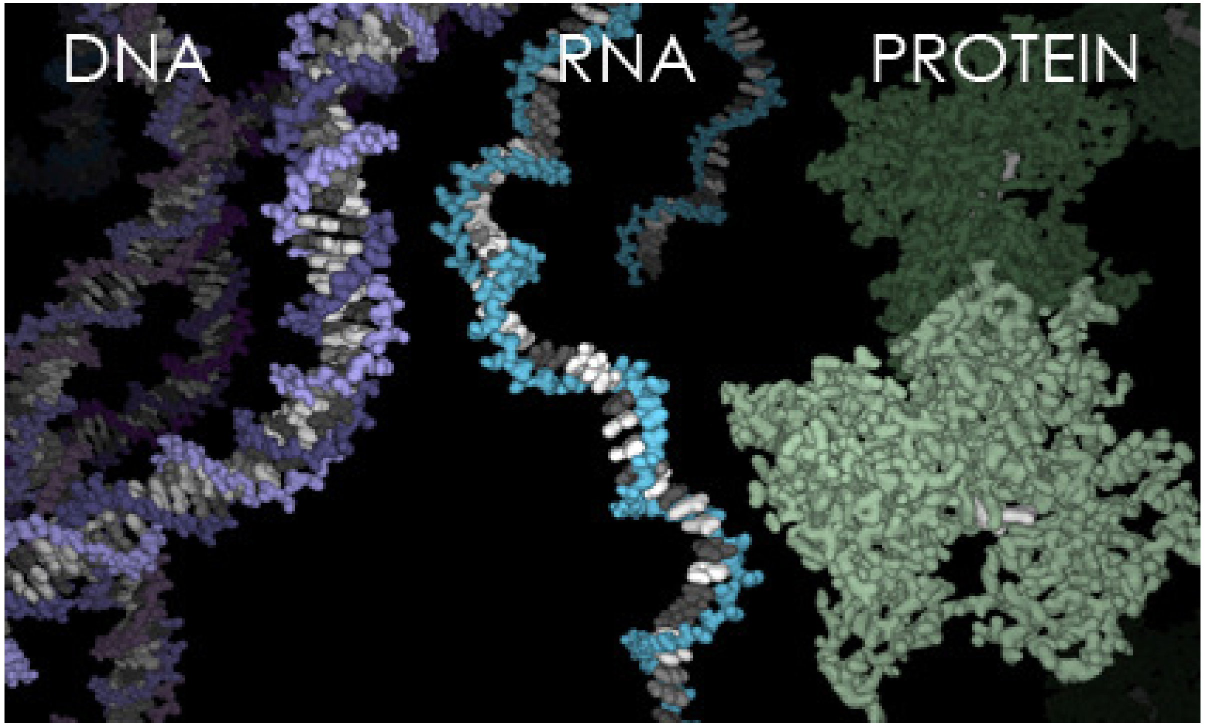
Metabolism-first vs. Genes-first

Genetics/replication-first: an information-carrying polymer capable of replication (RNA or something simpler) spontaneously arose from available prebiotic molecules available on early Earth. Metabolism incorporated later as a mean to receive energy from the surroundings in a controlled manner.

Metabolism-first: primitive metabolic cycles spontaneously assembled from simple prebiotic organic molecules or inorganic carbon sources as CO_2. And the cycles produced a set or more or less complex molecules needed for the replication process and construction of the genetic apparatus.

The supposed *proto-metabolism* would differ from the currently known one, because the chemical reactions were not catalysed by efficient enzymes, nor were aminoacid and peptide sequences determined by DNA.
The involved reactions were either spontaneous, or catalysed by inorganic catalysts or peptides.
Inorganic catalysts would be molecules, or ions, in solutions or on surfaces of solids such as clays or pyrites.
Peptides (or peptoids) formed either by random oligomerization or mutual catalysis.

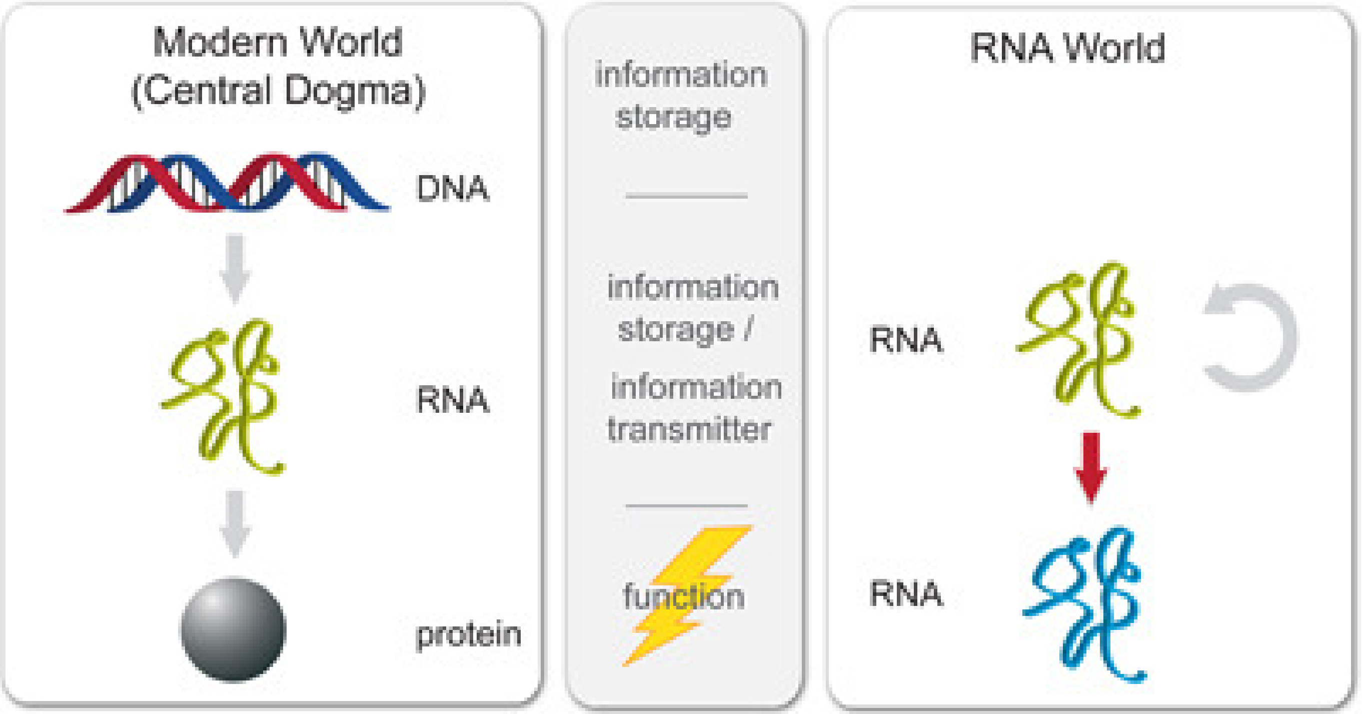
„Genes-first”



In modern cells, RNA (light blue, center) is made from a DNA template (purple, left) to create proteins (green, right).

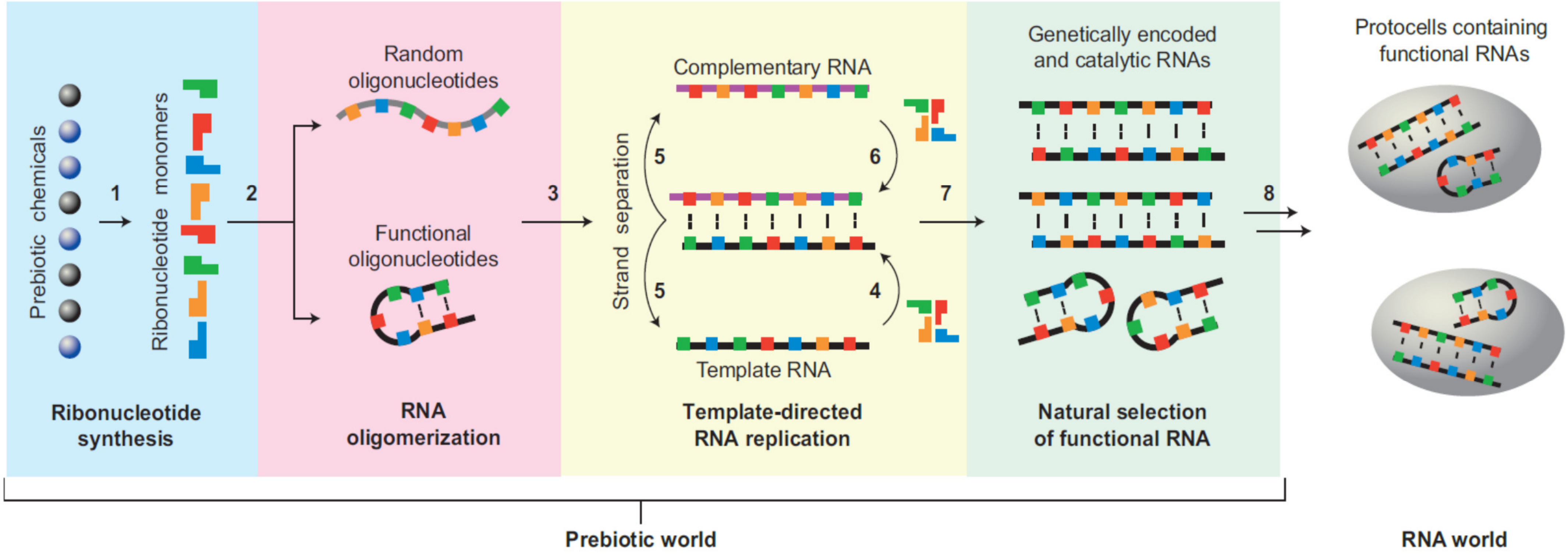
RNA folding is mediated by base-pairing interactions along different regions of a single-stranded RNA.

The RNA world

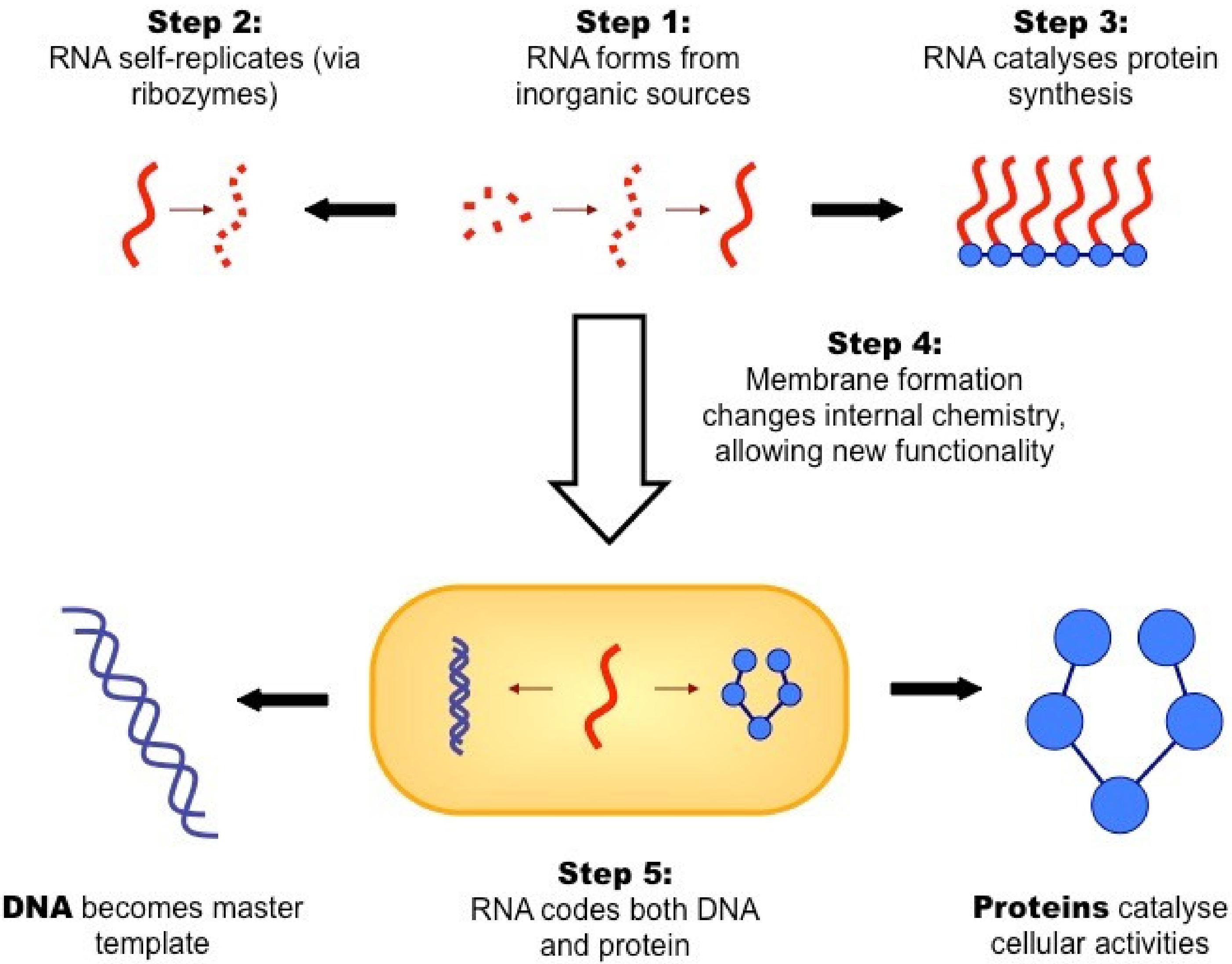


Conceptual idea that there was a period in the early history of life on Earth when RNA (or its structurally simplified analogue) carried out most of the information processing and metabolic transformations needed for biology to emerge from chemistry

The RNA world



The RNA world



Step 2:
RNA self-replicates (via ribozymes)

Step 1:
RNA forms from inorganic sources

Step 3:
RNA catalyses protein synthesis

Step 4:
Membrane formation changes internal chemistry, allowing new functionality

DNA becomes master template

Step 5:
RNA codes both DNA and protein

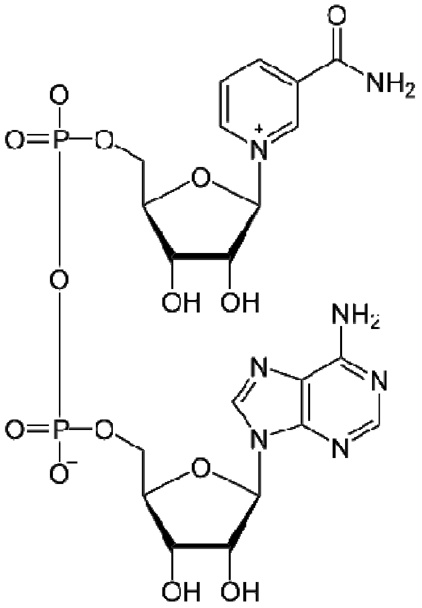
Proteins catalyse cellular activities

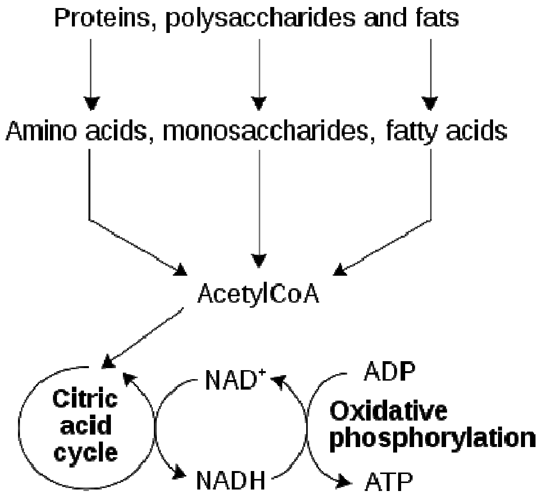
The RNA world

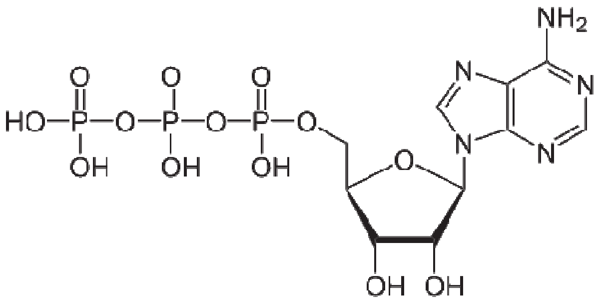
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

F. H. C. Crick *J. Mol. Biol.* **1968**, *38,* 367-379, L. E. Orgel *J. Mol. Biol.* **1968**, *38,* 381-393

Ribonucleotide coenzymes currently used by many proteins may be molecular „fossils” from the primordial RNA-based metabolism







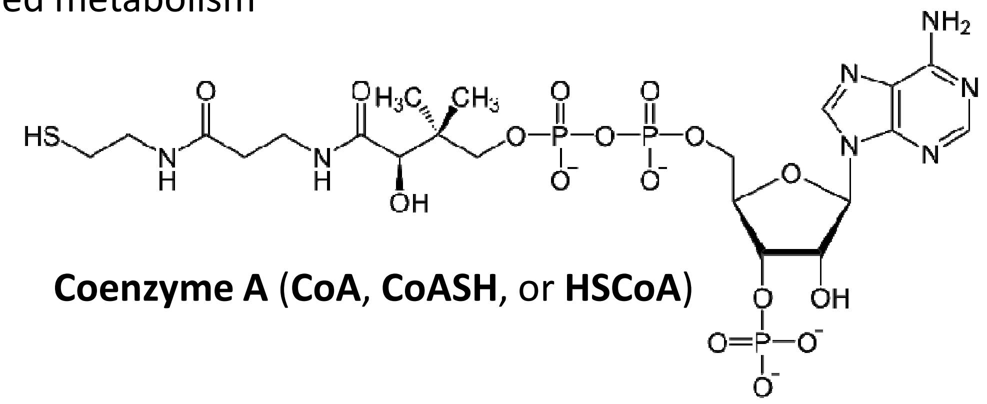
Nicotinamide adenine dinucleotide (**NAD$^+$**)

Adenosine triphosphate (**ATP**)

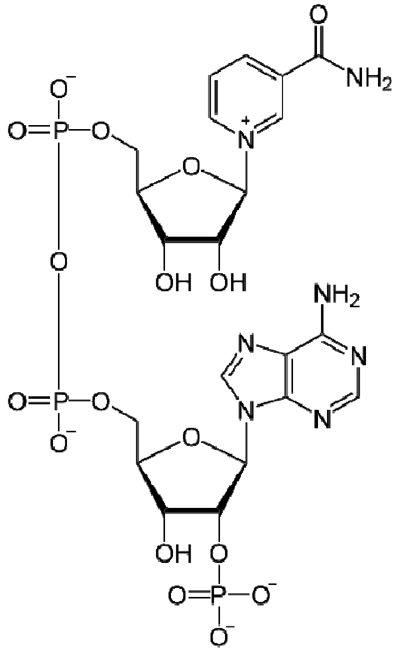
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

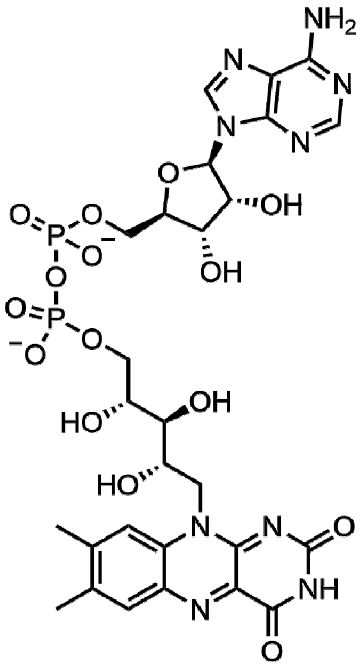
Ribonucleotide coenzymes now used by many proteins may be molecular „fossils” from the primoridal RNA-based metabolism



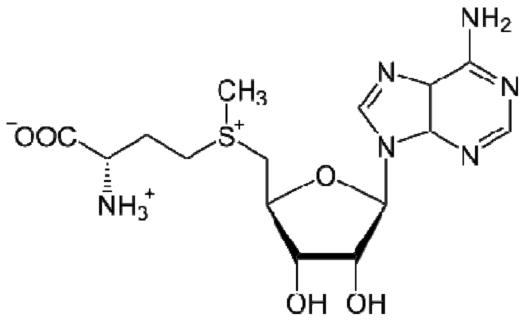
Coenzyme A (**CoA**, **CoASH**, or **HSCoA**)



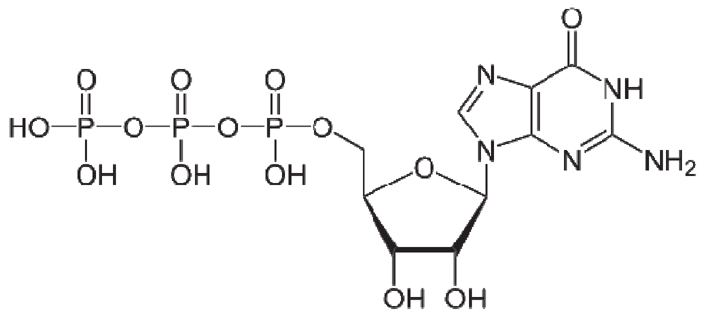
Nicotinamide adenine dinucleotide phosphate (NADP$^+$)



flavin adenine dinucleotide (**FAD**)



***S*-Adenosyl methionine**

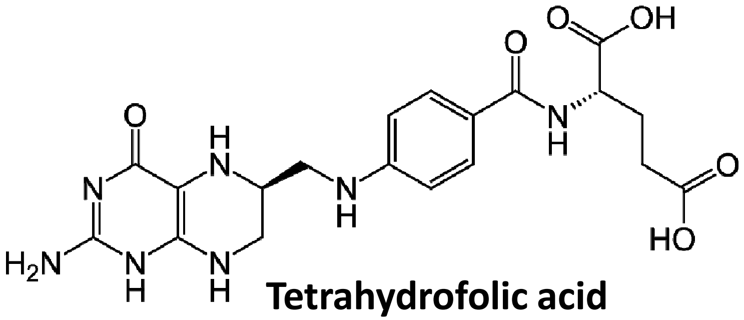


Guanosine-5'-triphosphate (**GTP**)

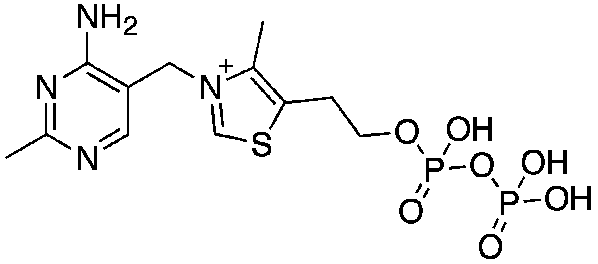
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

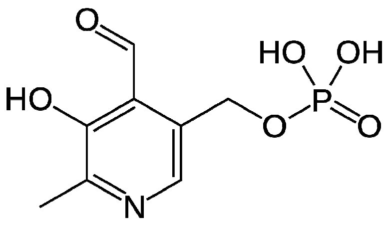
Other coenzymes contain cyclic nitrogen-containing bases that can also derive from nucleotides



Tetrahydrofolic acid



Thiamine pyrophosphate (TPP or ThPP) – Vit. B_1



Pyridoxal phosphate (PLP) – Vit. B_6

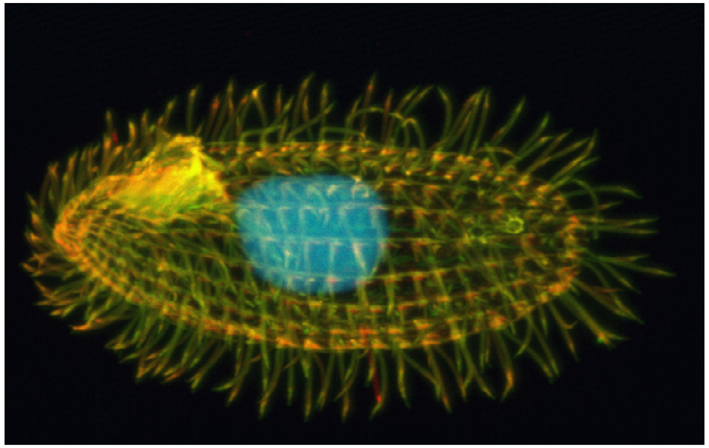
H. B. White III *J. Mol. Evol.* **1976**, *7*, 101-104

The RNA world

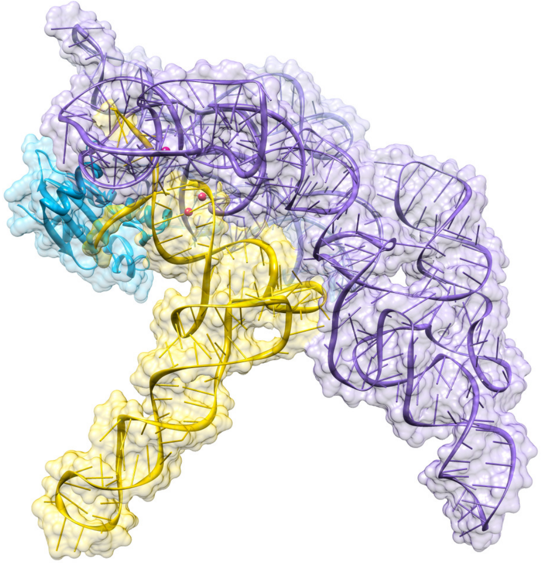
Ribozymes – Ribonucleic acid enzymes

1989 – Thomas Cech and Sidney Altman – Nobel Prize in chemistry for discovery of catalytic RNA

Thomas R. Cech was studying RNA splicing in the ciliated protozoan *Tetrahymena thermophila*
Sidney Altman and Norman Pace were studying the bacterial RNase P complex.



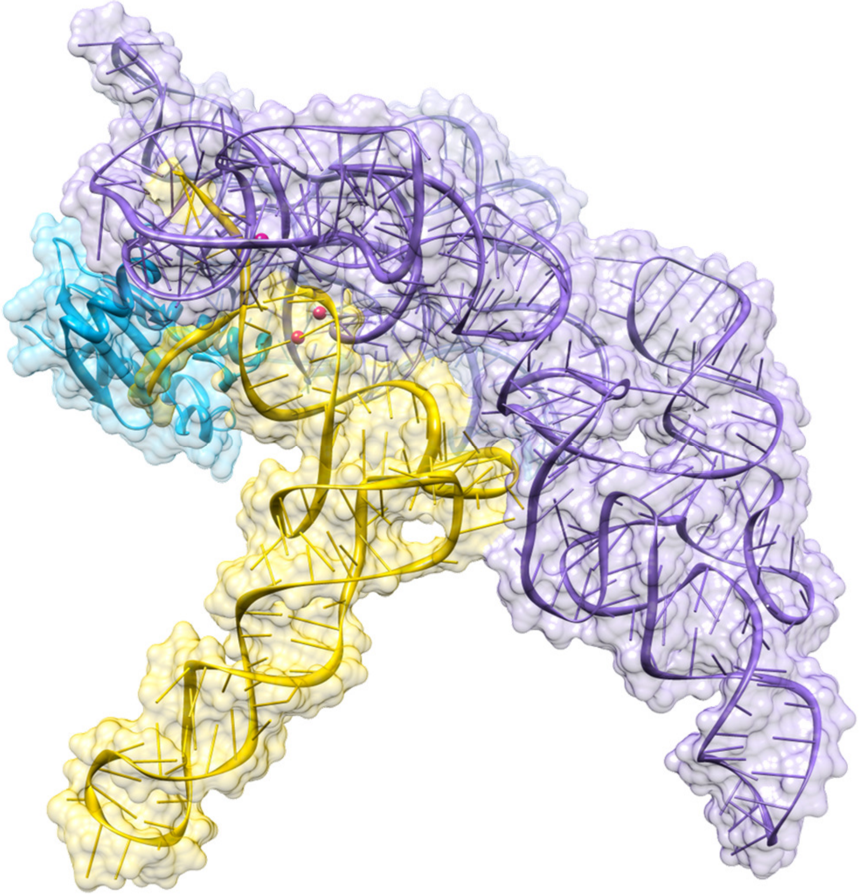
Tetrahymena thermophila



Bacterial RNAse P

The RNA world

Ribonuclease P

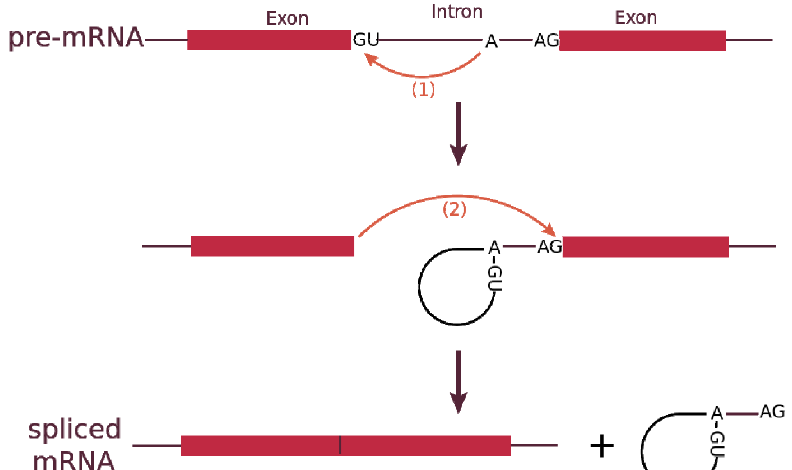


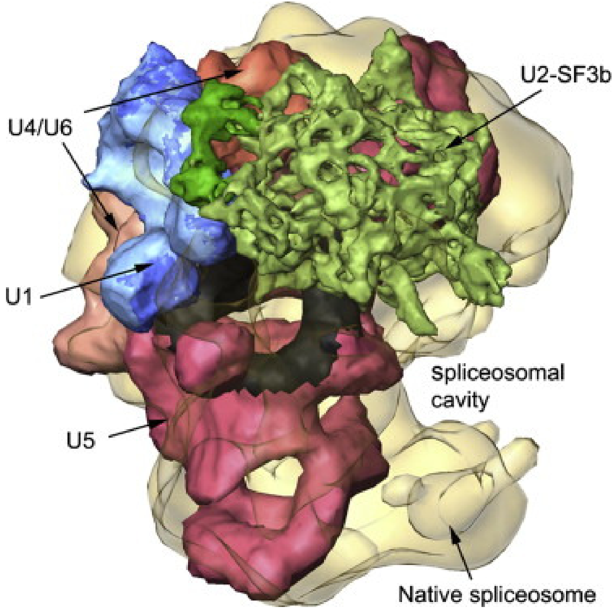
Ribonuclease P (RNase P) is a type of ribonuclease which cleaves RNA. RNase P is unique from other RNases in that it is a ribozyme – a ribonucleic acid that acts as a catalyst in the same way that a protein based enzyme would. Its function is to cleave off an extra, or precursor, sequence of RNA on tRNA molecules.

Bacterial RNase P has two components: an RNA chain, called M1 RNA, and a polypeptide chain, or protein, called C5 protein. *In vivo*, both components are necessary for the ribozyme to function properly, but *in vitro*, the M1 RNA can act alone as a catalyst. The primary role of the C5 protein is to enhance the substrate binding affinity and the catalytic rate of the M1 RNA enzyme probably by increasing the metal ion affinity in the active site.

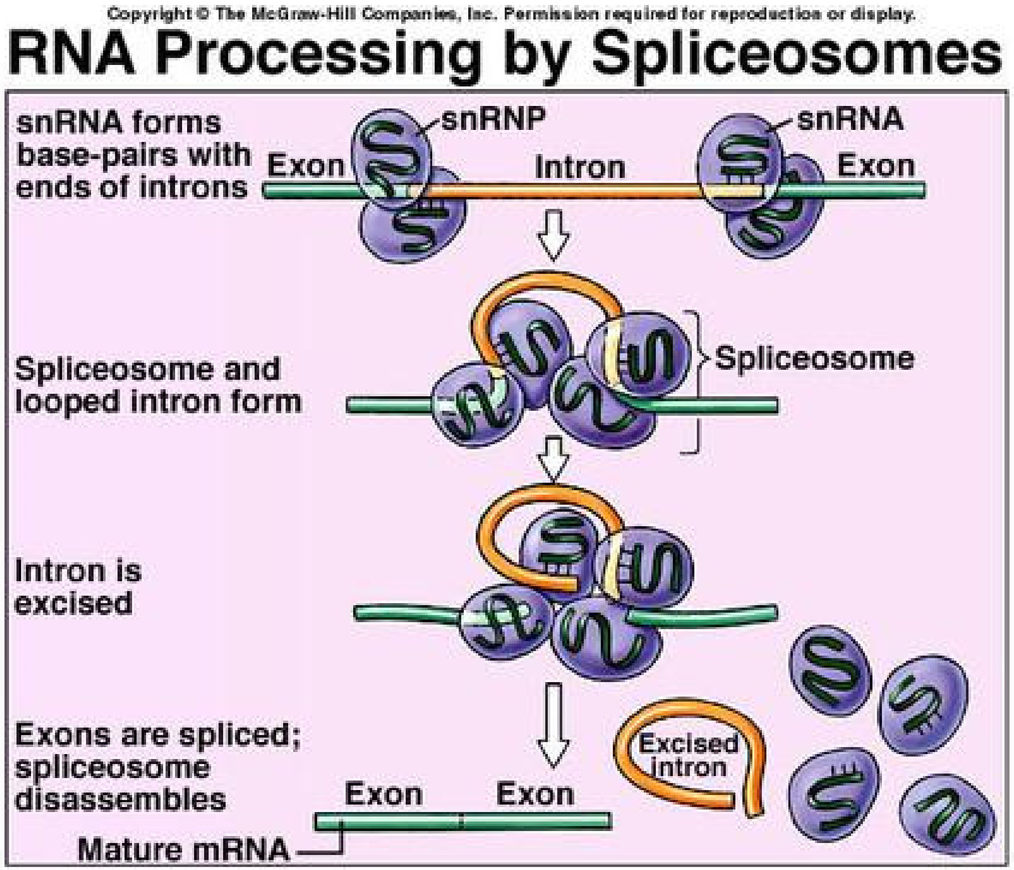
Crystal structure of a bacterial ribonuclease P holoenzyme in complex with tRNA (yellow), showing metal ions involved in catalysis (pink)

RNA splicing





Spliceosome – a complex of ribonucleoproteins

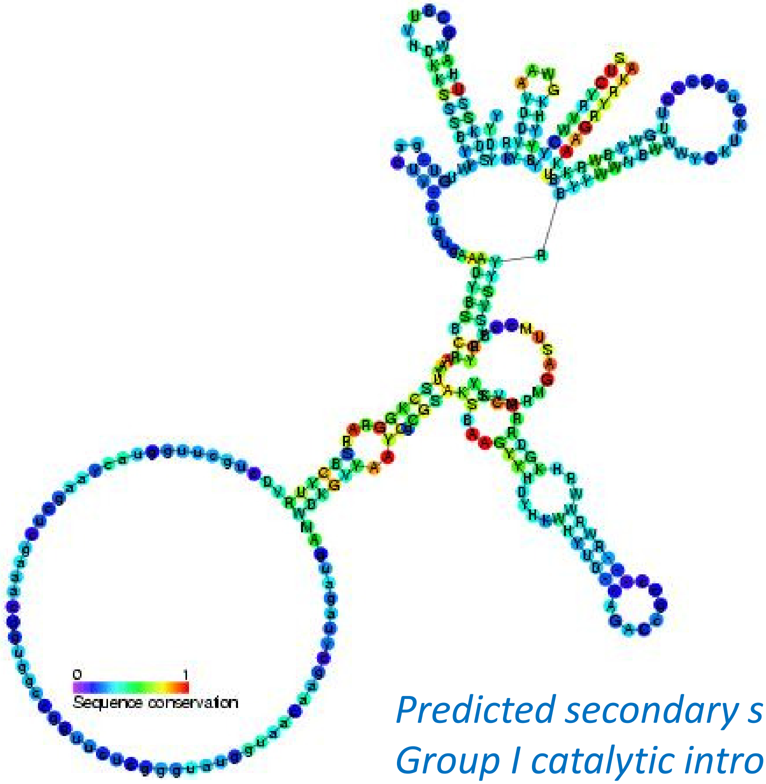


RNA splicing

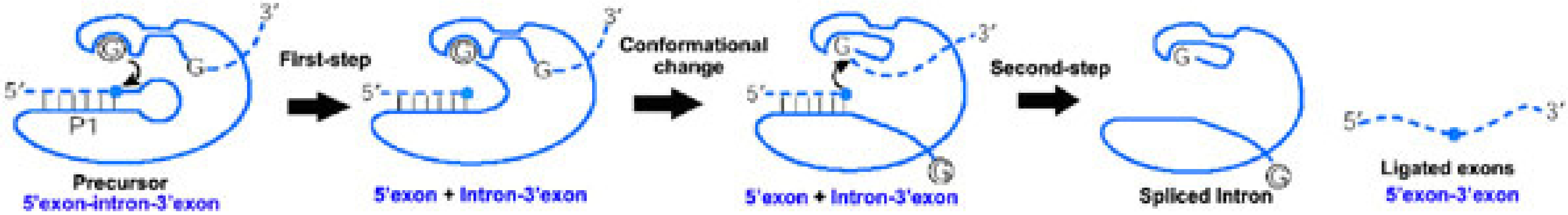
Self-splicing RNA introns

RNA splicing in *Tetrahymena* was taking place also in absence of the spliceosome - the ‚negative control’ obtained after protease digestion also spliced.

In contrary to the spliceosome, the ***catalytic*** motif ***does not*** contain protein part, ***only RNA***.
First known example of a ***ribozyme*** – ribonucleic acid-composed enzyme analogue.

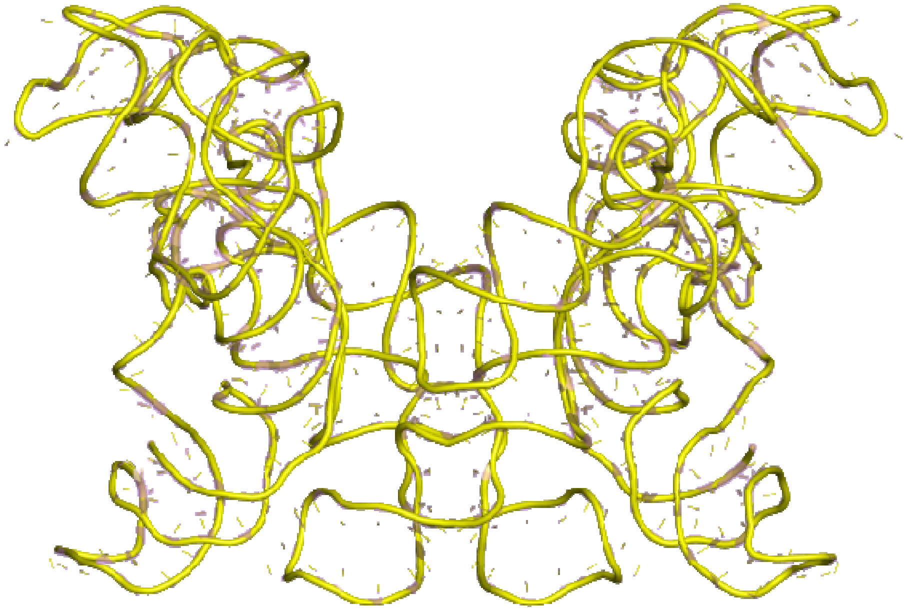


Predicted secondary structure and sequence conservation of Group I catalytic intron



RNA splicing

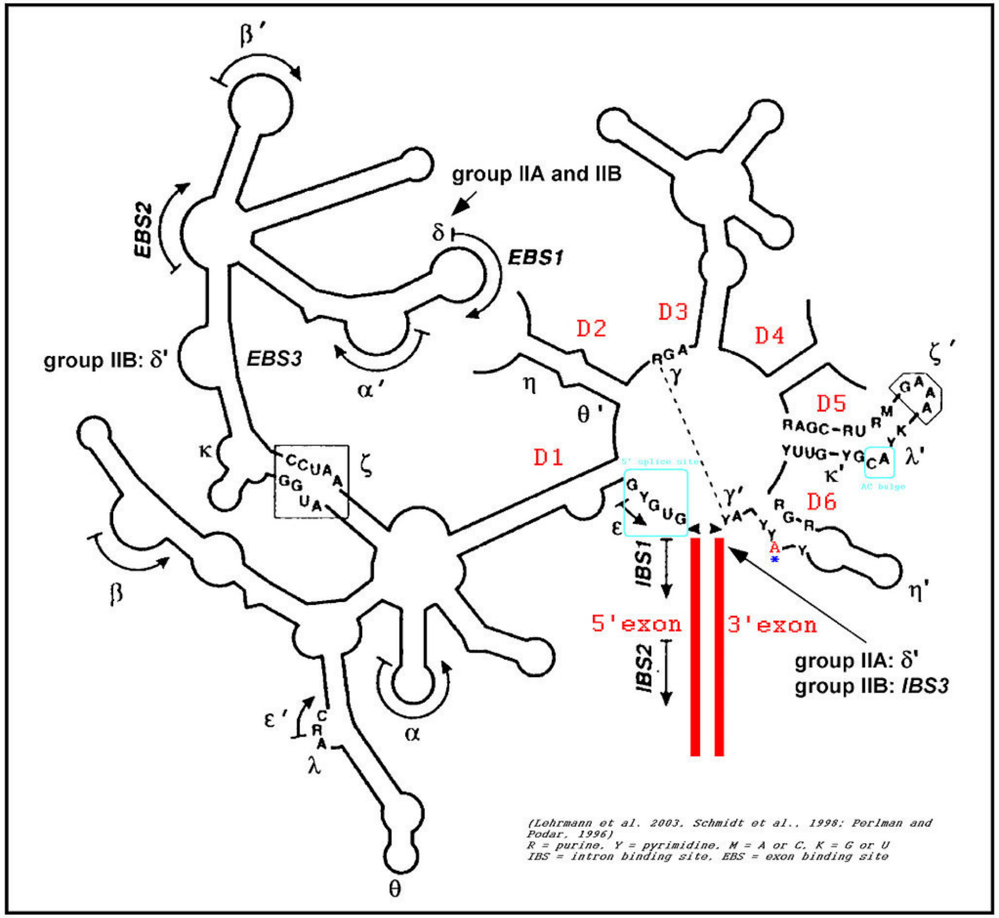
Group I catalytic introns



A 3D representation of the Group I catalytic intron.
This view shows the active site in the crystal structure of the Tetrahymena ribozyme

RNA splicing

Group II catalytic introns

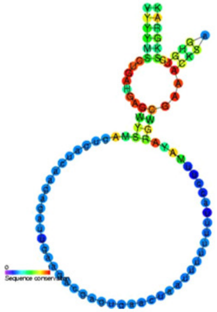


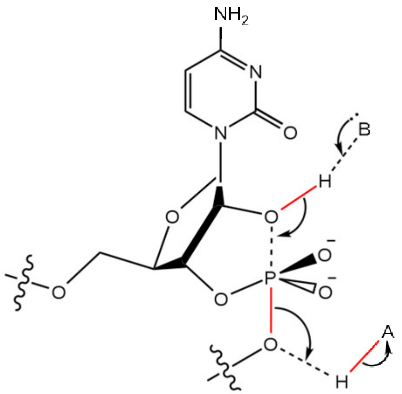
(Lehrmann et al. 2003. Schmidt et al., 1998; Perlman and Podar, 1996)
R = purine, Y = pyrimidine, M = A or C, K = G or U
IBS = intron binding site, EBS = exon binding site

Ribozyme activity (e.g., self-splicing) can occur under high-salt conditions in vitro. However, assistance from proteins is required for in vivo splicing

It is hypothesized that pre-mRNA splicing may have evolved from group II introns, due to the similar catalytic mechanism as well as the structural similarity of the Domain V substructure to the U6/U2 extended snRNA

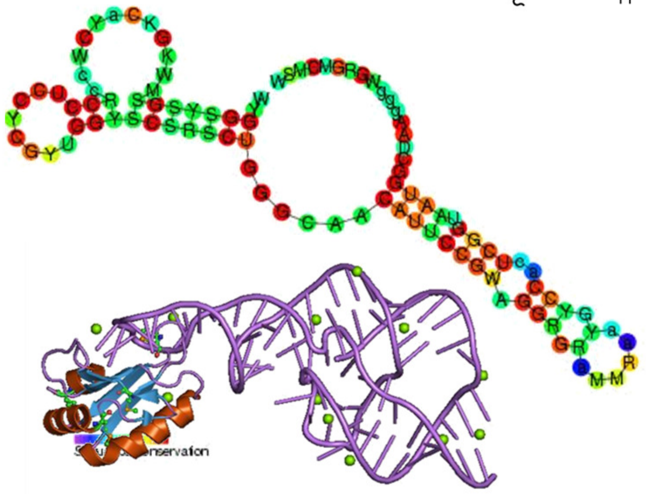
Ribozymes and riboswitches





Hammerhead ribozyme

The hammerhead ribozyme is a RNA molecule motif that catalyzes reversible cleavage and joining reactions at a specific site within an RNA molecule (model system; targeted RNA cleavage experiments)



HDV ribozyme

The hepatitis delta virus (HDV) ribozyme is a non-coding RNA found in the hepatitis delta virus that is necessary for viral replication and is thought to be the only catalytic RNA known to be required for viability of a human pathogen.

The ribozyme acts to process the RNA transcripts to unit lengths in a self-cleavage reaction. The ribozyme is found to be active in vivo in the absence of any protein factors and is the fastest known naturally occurring self-cleaving RNA.

Riboswitches

A riboswitch is a regulatory segment of a messenger RNA molecule that binds a small molecule, resulting in a change in production of the proteins encoded by the mRNA (bacteria, TPP riboswitch also in plants and funghi)

Riboswitches

2002 - (Breaker and Nudler) – discovery of a nucleic acid-based genetic regulatory element – ***riboswitch***.

The lysine riboswitch

Riboswitches - naturally occurring regulatory segments of mRNA that bind small molecules specifically. The binding results in a change in production of the proteins encoded by the mRNA

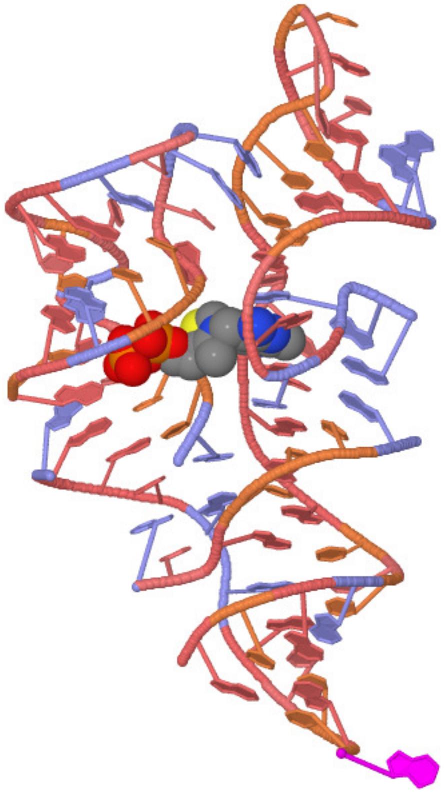
Before discovery of ***riboswitches*** only ***proteins*** were supposed to do so in the biological context.

Most known ***riboswitches*** occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in archaea, plants and certain fungi.

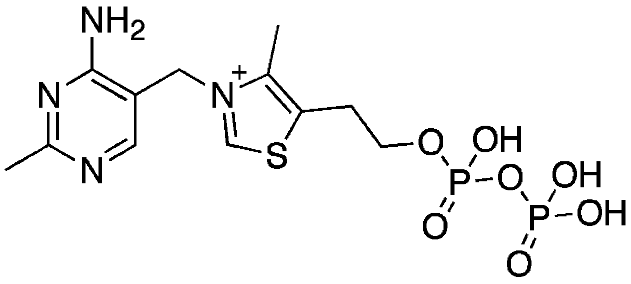
Riboswitches exist in all domains of life, and therefore are likely that they might represent ancient regulatory systems or fragments of ***RNA-world ribozymes*** whose binding domains remained conserved throughout the evolution

The TPP Riboswitch

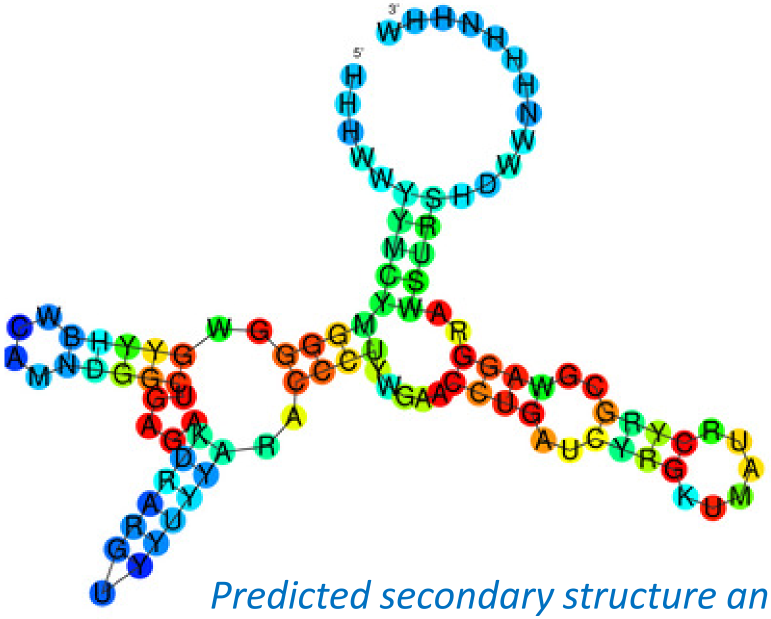
The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



The 3D structure of TPP riboswitch (by Benjamin Schuster-Böckler)



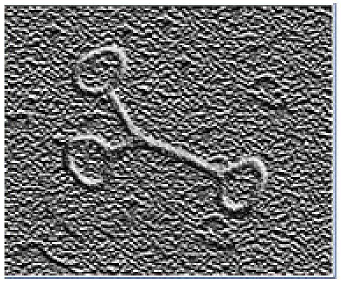
Thiamine pyrophosphate TPP



Predicted secondary structure and sequence conservation of TPP riboswitch

Viroids

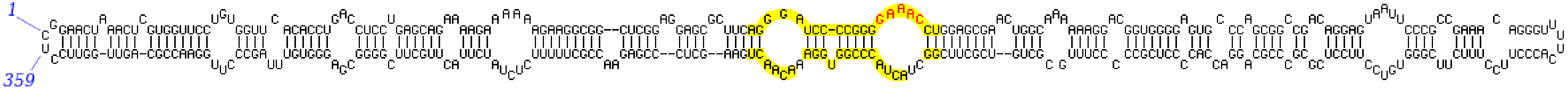
Viroids ("subviral pathogens„) are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Viroids are extremely small - 246 to 467 nucleobases (genomes of smallest viruses start from 2,000 nucleobases). Viroids are plausible "living relics" of the RNA world.



PSTVd-infected potatoes (right)

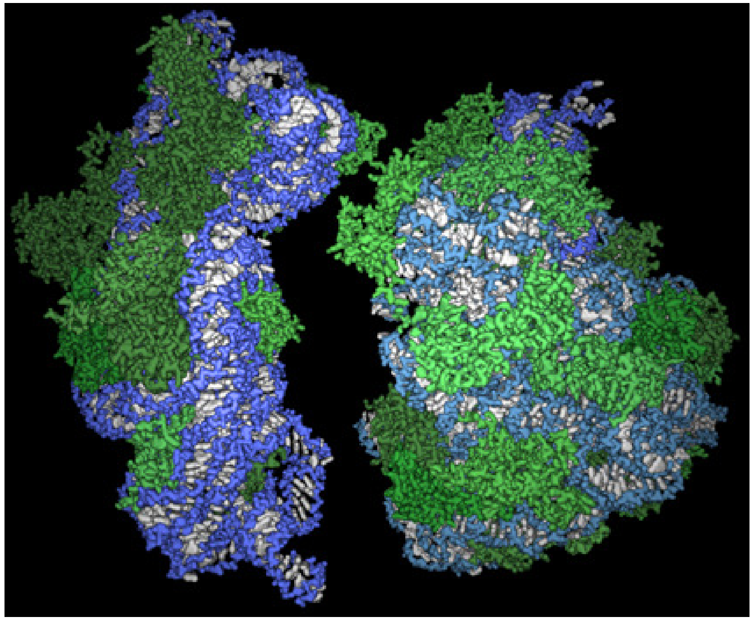
Viroid properties:

- small size (error-prone replication)
- high G-C content, (stability and replication fidelity)
- circular structure (complete replication without genomic tags)
- lack of protein-coding ability, consistent with a ribosome-free habitat; and replication mediated in some by ribozymes—the fingerprint of the RNA world.



Putative secondary structure of the PSTVd viroid

Ribosome – the ‚smoking gun'



Ribosome: green - proteins, blue and white - RNA

The ribosome is a molecular machine, found within all living cells, that serves as the site of biological protein synthesis (translation). Ribosomes link amino acids together in the order specified by messenger RNA (mRNA) molecules.

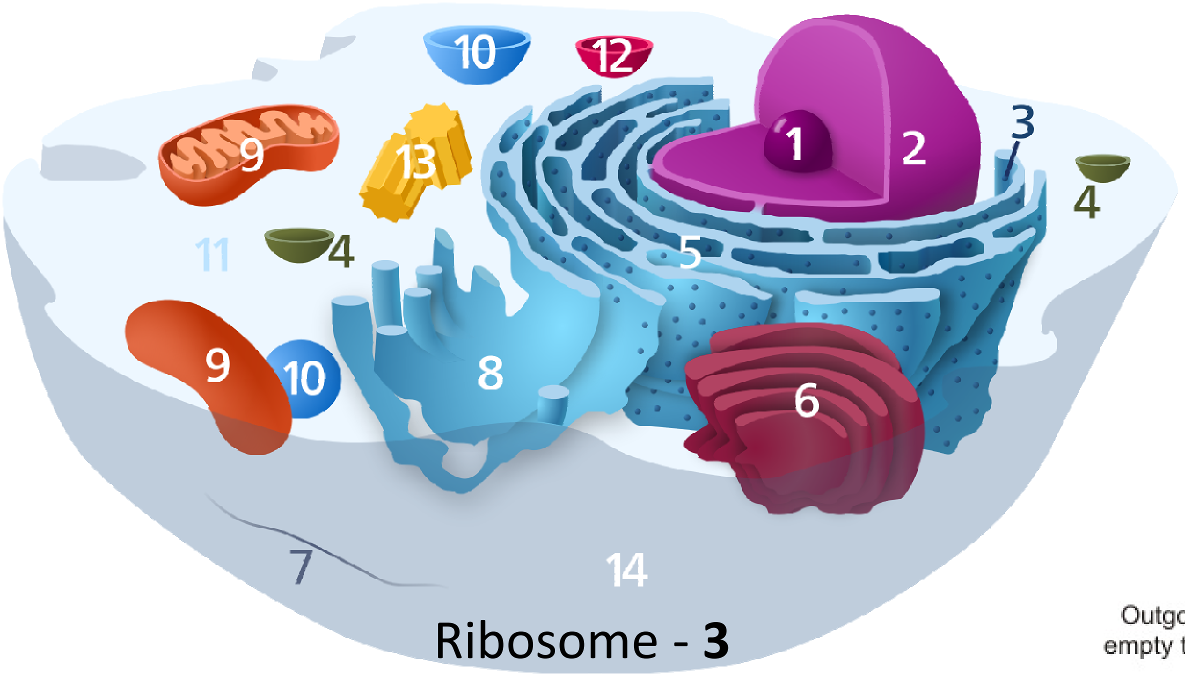
Ribosome is structurally highly conserved among all living species – most likely present in LUCA

Ribosomes:

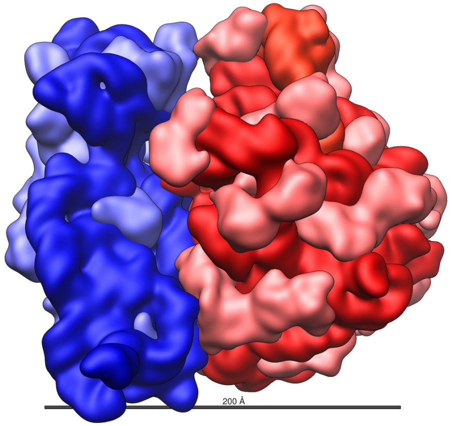
- the *small ribosomal subunit*, which reads the RNA
- the *large subunit*, which joins amino acids to form a polypeptide chain.

Each subunit is composed of one or more ribosomal RNA (rRNA) molecules and a variety of ribosomal proteins.

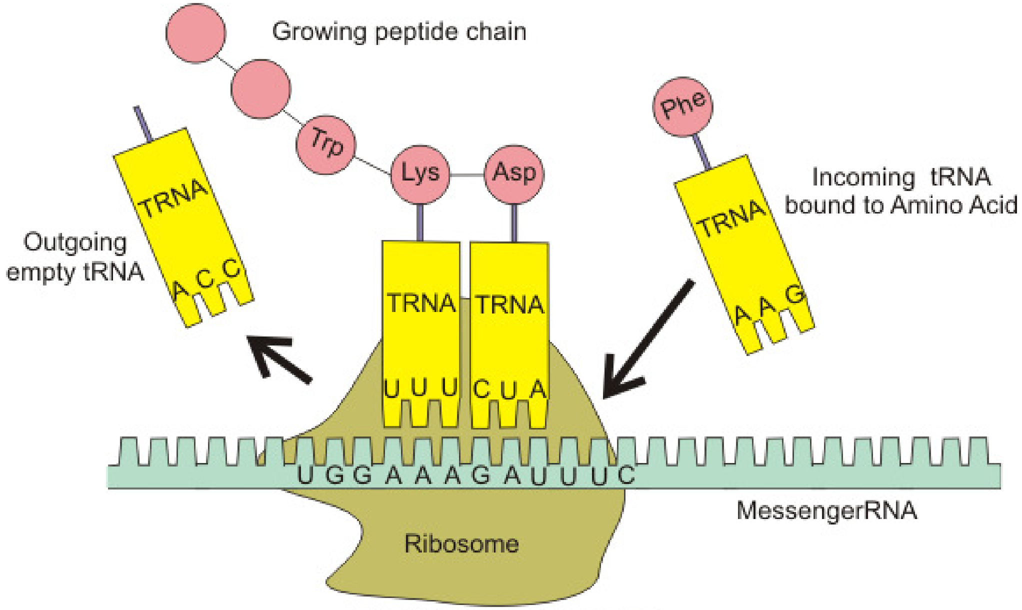
Ribosome – the ‚smoking gun’



Ribosome - **3**



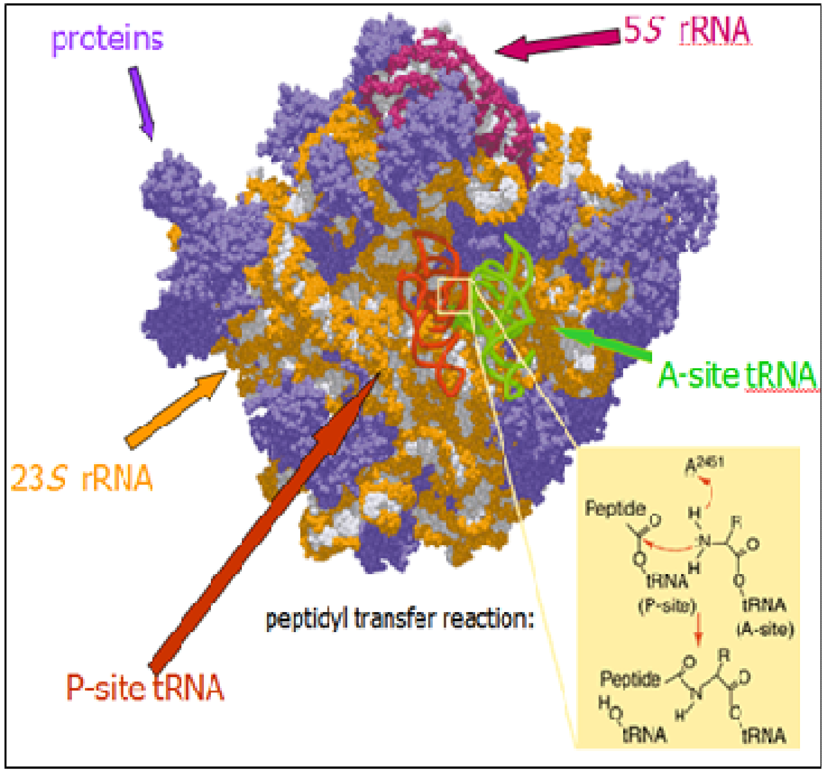
Large and small subunit



Peptide Synthesis

Ribosome – the ‚smoking gun'

Ribosome is a ribozyme!



No protein is present within 18 Angstroms from the active site → proteins play a structural role, but DO NOT CATALYZE THE ACYL TRANSFER PROCESS

T. Cech *Science.* **2000**, *289*, 878-879

Ribosome – the ‚smoking gun'

Ribosome is a ribozyme!

The ribosome may have first originated in an RNA world appearing as a self-replicating complex that only later evolved the ability to synthesize proteins when amino acids began to appear.

Studies suggest that ancient ribosomes constructed solely of rRNA could have developed the ability to synthesize peptide bonds.

In addition, evidence strongly points to ancient ribosomes as self-replicating complexes, where the rRNA in the ribosomes had informational, structural, and catalytic purposes because it could have coded for tRNAs and proteins needed for ribosomal self-replication.

As amino acids gradually appeared in the RNA world under prebiotic conditions, their interactions with catalytic RNA would increase both the range and efficiency of function of catalytic RNA molecules. Thus, the driving force for the evolution of the ribosome from an ancient self-replicating machine into its current form as a translational machine may have been the selective pressure to incorporate proteins into the ribosome's self-replicating mechanisms, so as to increase its capacity for self-replication

The RNA world

RNA as catalyst

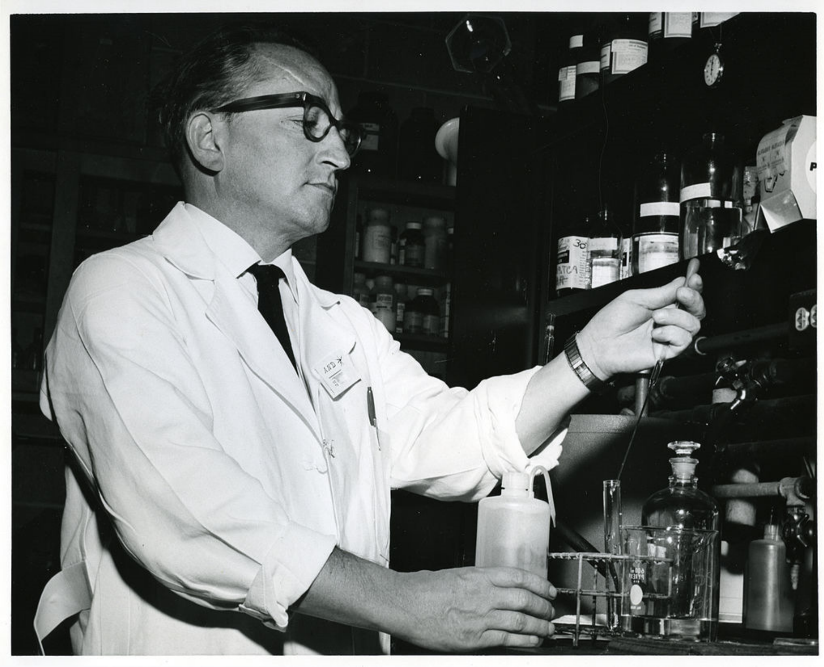
Currently known co-enzymes
Ribozymes
Ribosome

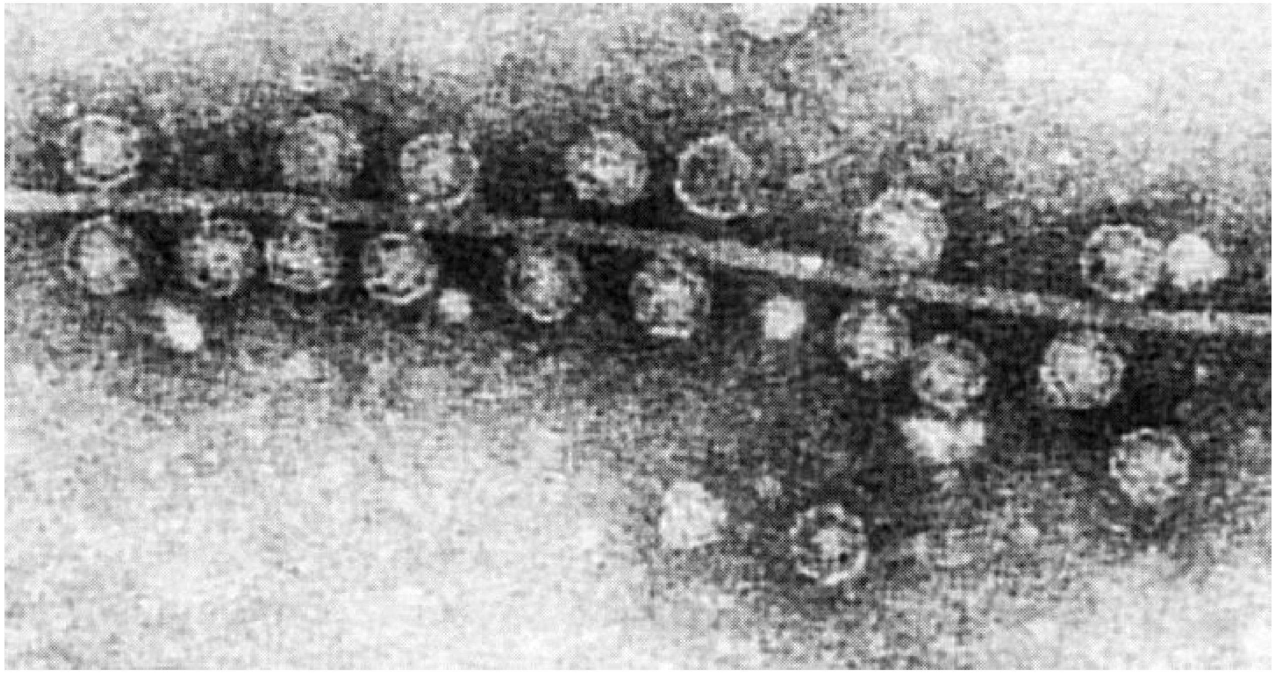
Can RNA evolve?

Can RNA replicate itself?

The RNA world

Can RNA evolve?



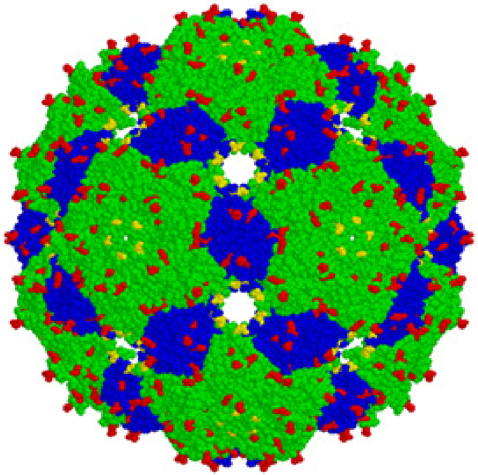


Spiegelman's monster

The RNA world

The bacteriophage Qβ – a virus containing RNA-dependent RNA polymerase (protein, enzymatic replicase)

Spiegelman's monster



Spiegelman mixed the Qβ RNA, the Qβ enzymatic replicase, mononucleotides
and some salts (buffer). RNA replication begun.
An aliquot was transferred several times to a fresh solution without template.

Shorter RNA chains replicate faster. The selection in this system favors speed.
And no evolutionary pressure on pathogenicity was present anymore.
So the RNA became shorter and shorter due to random mutations during copying.

After 74 passages, the original 4500 nt RNA strand was reduced to 218 nt.
Such a short RNA chain replicated very quickly under these unnatural circumstances.
Of course, it lost all its genes and was unable to produce any useful proteins anymore.
First example of *in vitro* RNA evolution

Kacian D. L., Mills D. R., Kramer F. R., Spiegelman S. *PNAS* **1972**, *69*, 3038-3042.

Spiegelman's monster can be also formed by simple mixing of activated RNA monoers and the Qβ enzymatic replicase,
in absence of any RNA template!

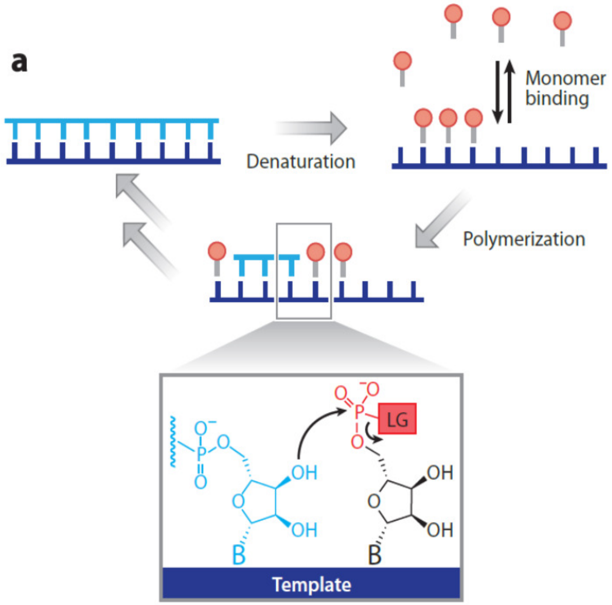
Sumper M., Luce R. *PNAS* **1975**, *72*, 162-166.

The RNA world

RNA self-replication

Nonenzymatic template-directed RNA polymerization

Maximally 30-50 nt extension, fidelity strongly sequence-dependent

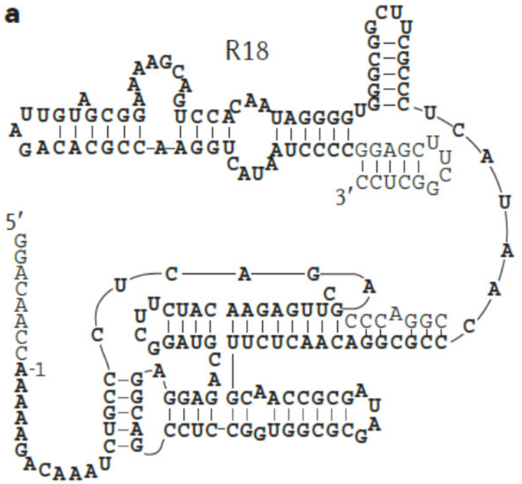


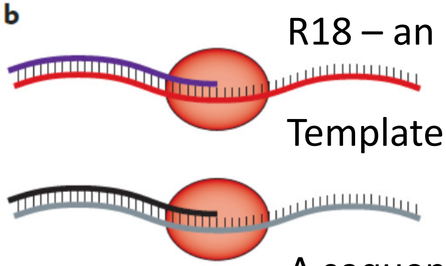
General RNA polymerase ribozyme (‚replicase')

Networks of RNA molecules that mutually catalyse their replication – autocatalytic replication of the whole network

The RNA world

RNA-dependent RNA polymerase ribozyme – Replicase - the ‚holy Grail' of the RNA world

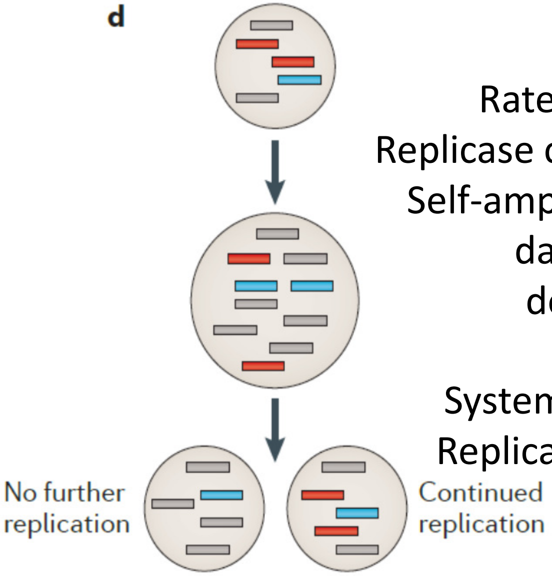




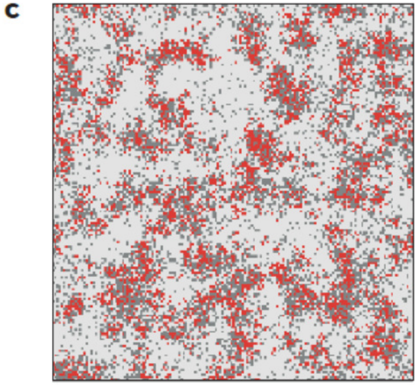
R18 – an artificial polymerase evolved from the class I ligase ribozyme.

Template: another copy of itself (red) or an unrelated sequence (grey).

A sequence of 206 nt was copied (fidelity 97.4%) at low temperatures by an engineered R18 mutant – first ribozyme capable to synthesize RNA oligomers longer than itself (though NO self-replication yet!)



Rate of replication not sensitive on the template's sequence.
Replicase could replicate other ribozymes (e.g. with metabolic functions).
Self-amplifying replicase needs a working complementary replicase – danger of paraistes (templates that copy themselves but do not contribute to the replication of the polymerase).



Systems of altruistic replicators are destroyed by parasites (grey).
Replicators (red) can survive e.g. by diffusion on 2D surfaces (**c**) or selection inside compartments (**d**)

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. *Science* **2001**, *292*, 1319–1325.

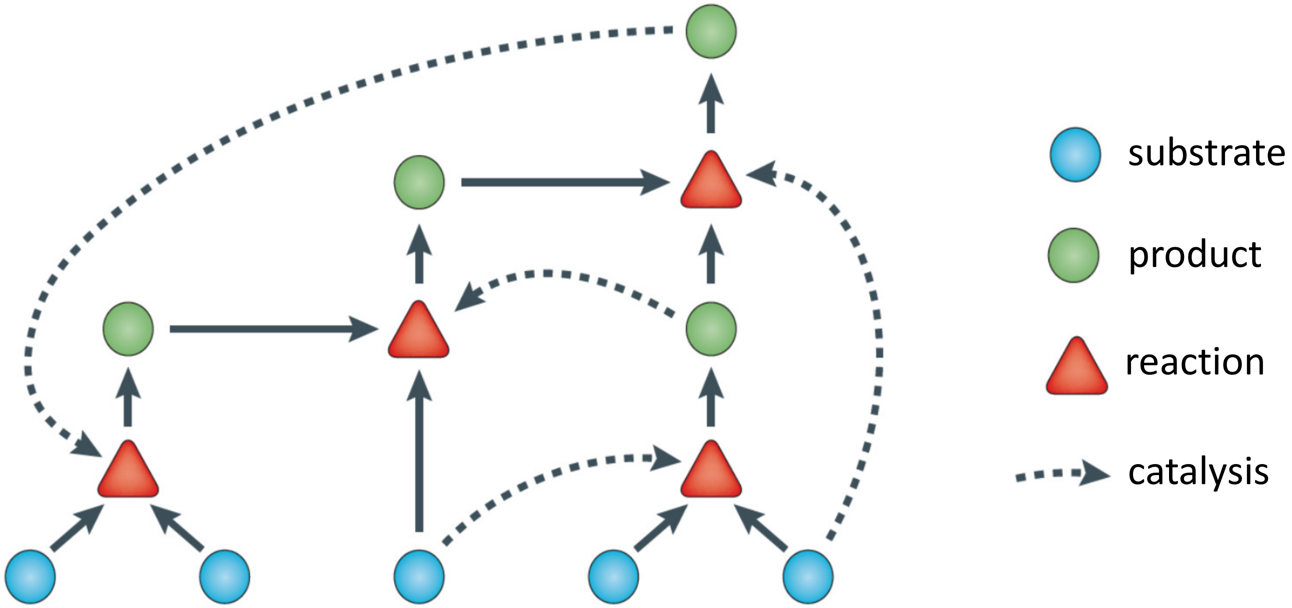
Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **2013**, *5*, 1011–1018.

The RNA world

Replicase - problem

The replicase most likely needs to be long (> 200 nt) for the efficient replication –
How could such long fucntional RNA be spontaneously generated?

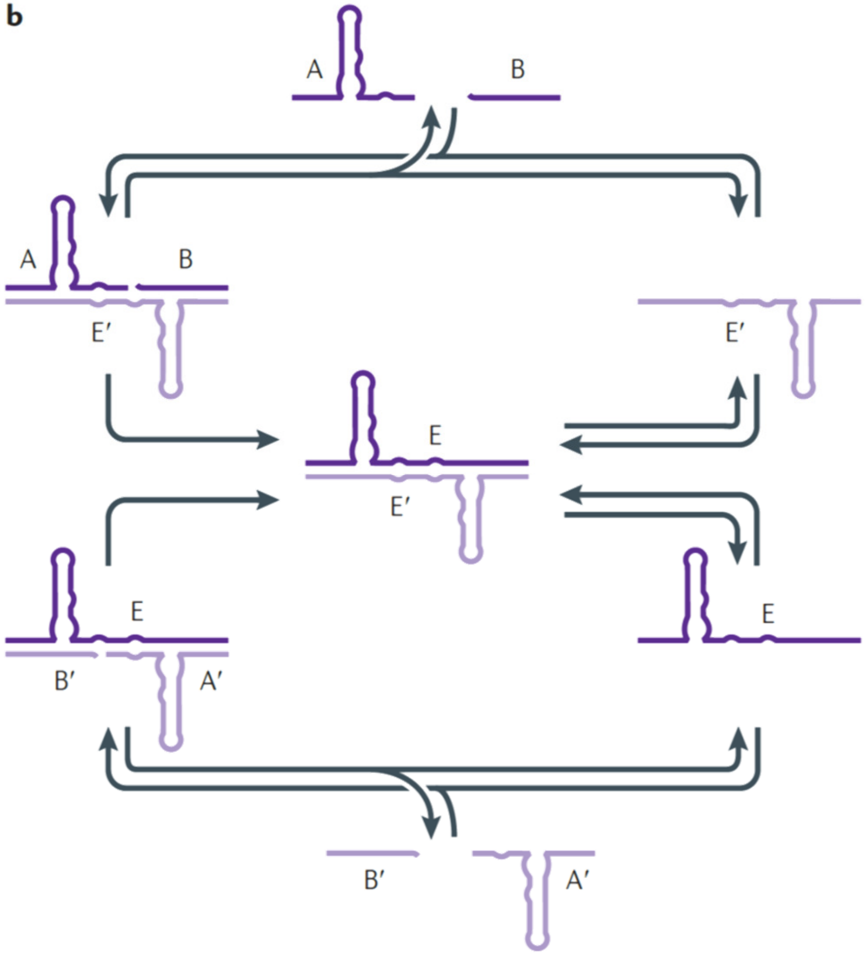
Possible solution – autocatalytic networks



No component can replicate without all the others

The RNA world

Mutually autocatalytic RNA networks

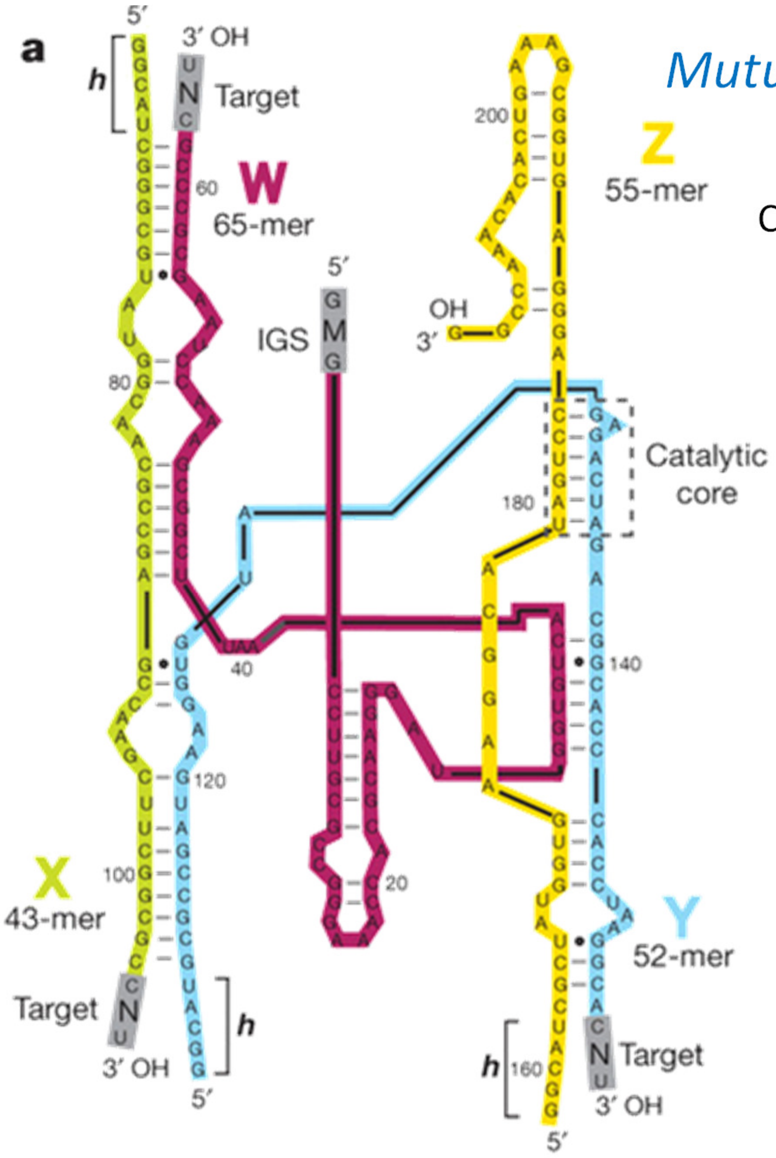


An autocatalytic set composed of two cross-catalytic ligases was demonstrated. RNA A and RNA B are ligated together by ribozyme E‘ to create ribozyme E, which can reciprocate and ligate RNA A‘ and RNA B‘ to create ribozyme E‘.

Lincoln, T. A. & Joyce, G. F. *Science* **2009**, *323*, 1229–1232.

The RNA world

Mutually autocatalytic RNA networks

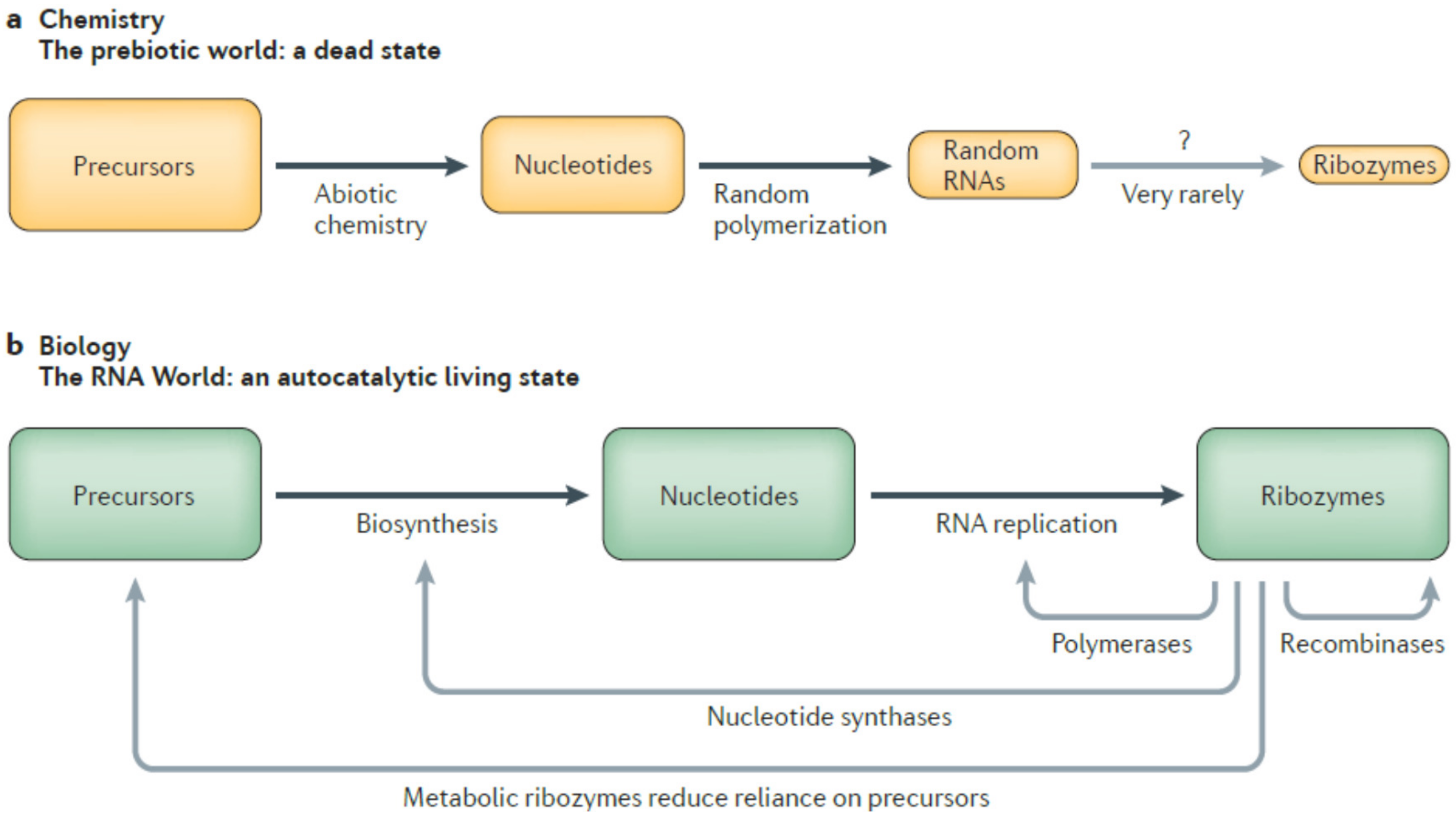


Cooperation between multiple strands that assemble to perform a single function.

Ribozymes, such as the *Azoarcus* recombinase, can be made from several short strands that assemble as a result of RNA secondary structure formation and information contained in internal guide sequences (IGSs) and complementary targets (grey).

Vadia, N. *et al. Nature* **2012**, *491*, 72-77.

The RNA world



a Chemistry
The prebiotic world: a dead state

Precursors → Abiotic chemistry → Nucleotides → Random polymerization → Random RNAs → ? Very rarely → Ribozymes

b Biology
The RNA World: an autocatalytic living state

Precursors → Biosynthesis → Nucleotides → RNA replication → Ribozymes

Polymerases

Recombinases

Nucleotide synthases

Metabolic ribozymes reduce reliance on precursors

Transition from chemistry to biology involves autocatalytic feedbacks from ribozymes to all stages of the prebiotic chemistry