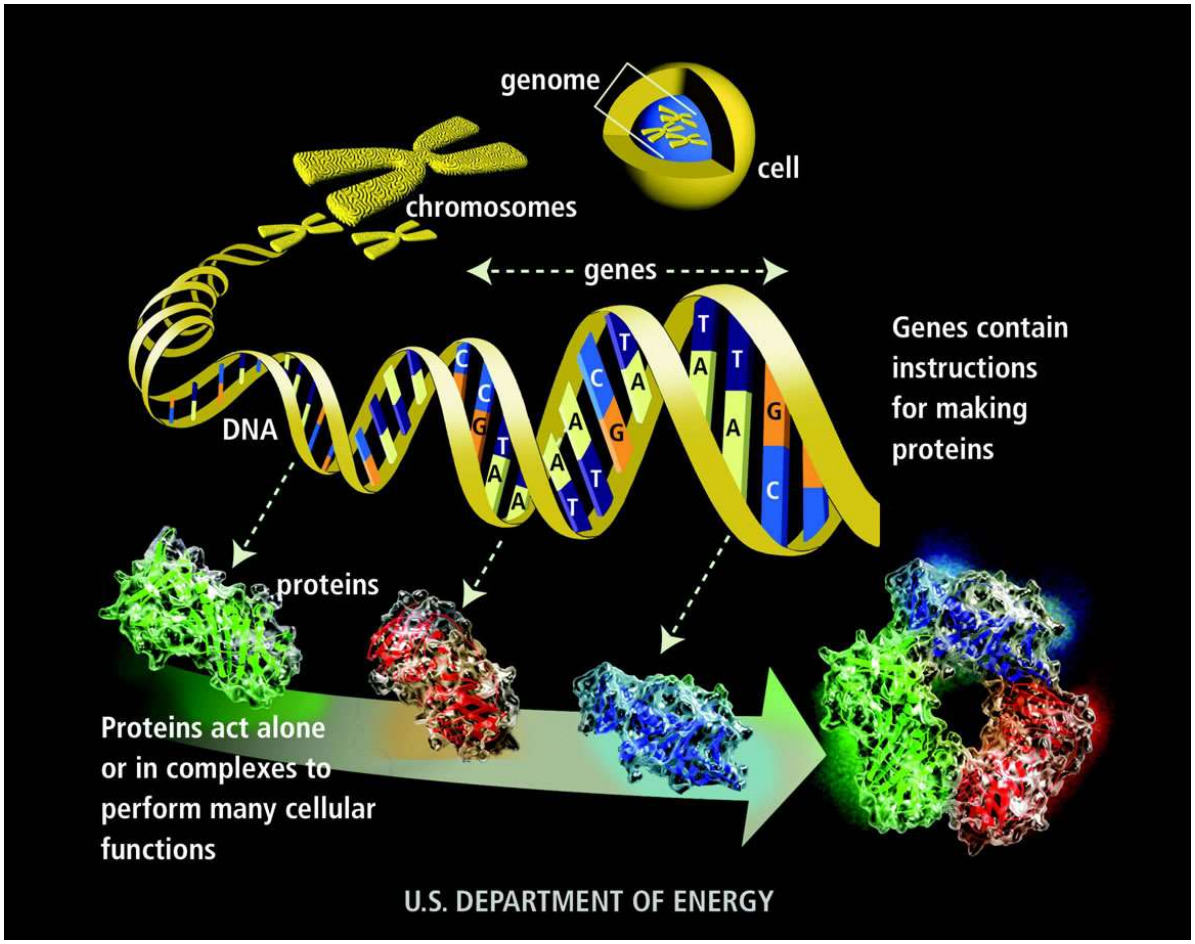
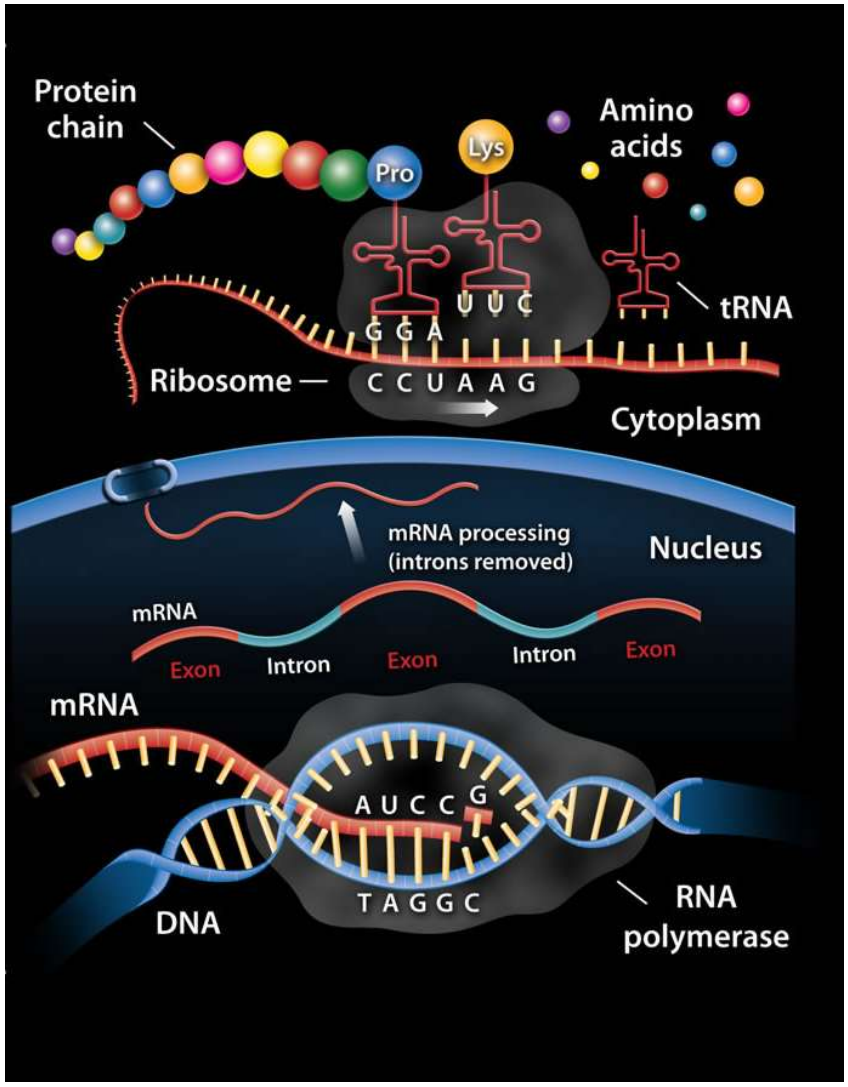
CHAPTER 1

OLIGONUCLEOTIDES

The less common side of RNA

From DNA to proteins



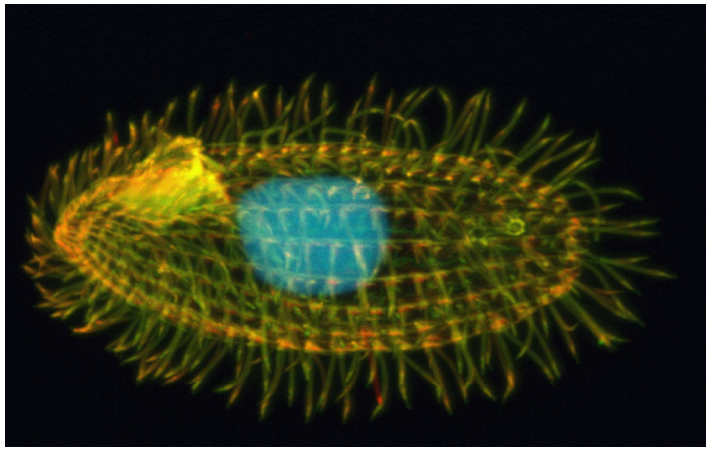


Ribozymes

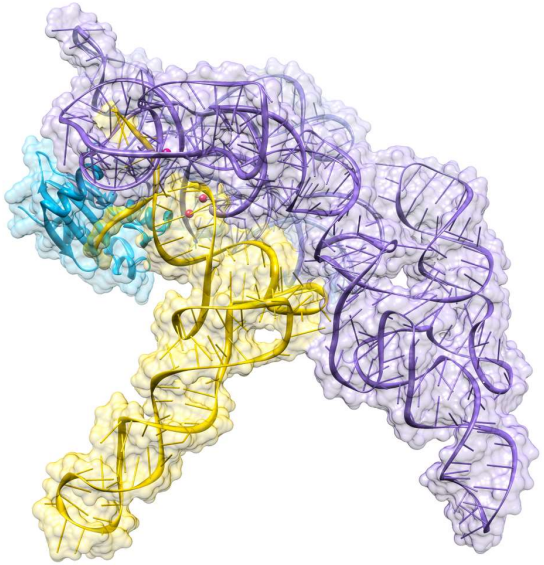
Ribozymes – Ribonucleic acid enzymes

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.

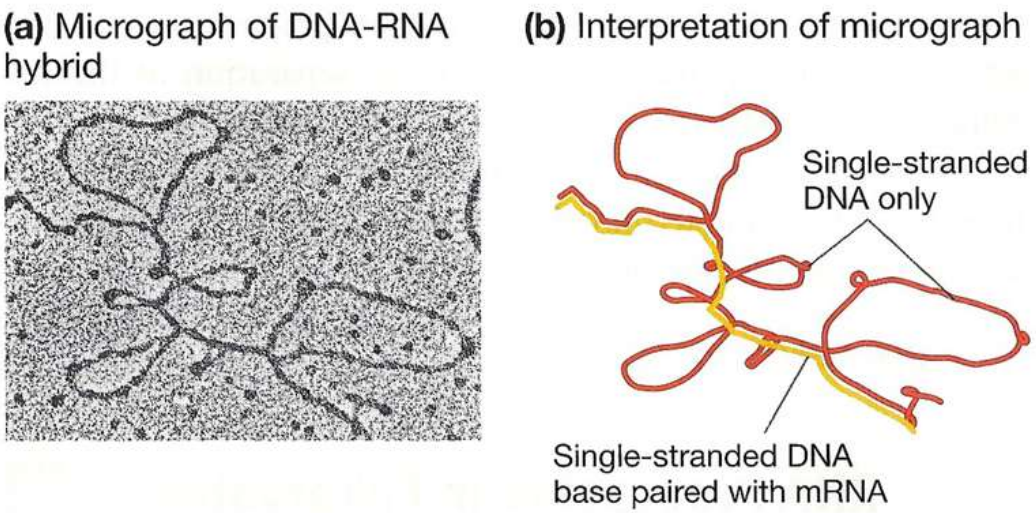


Tetrahymena thermophila

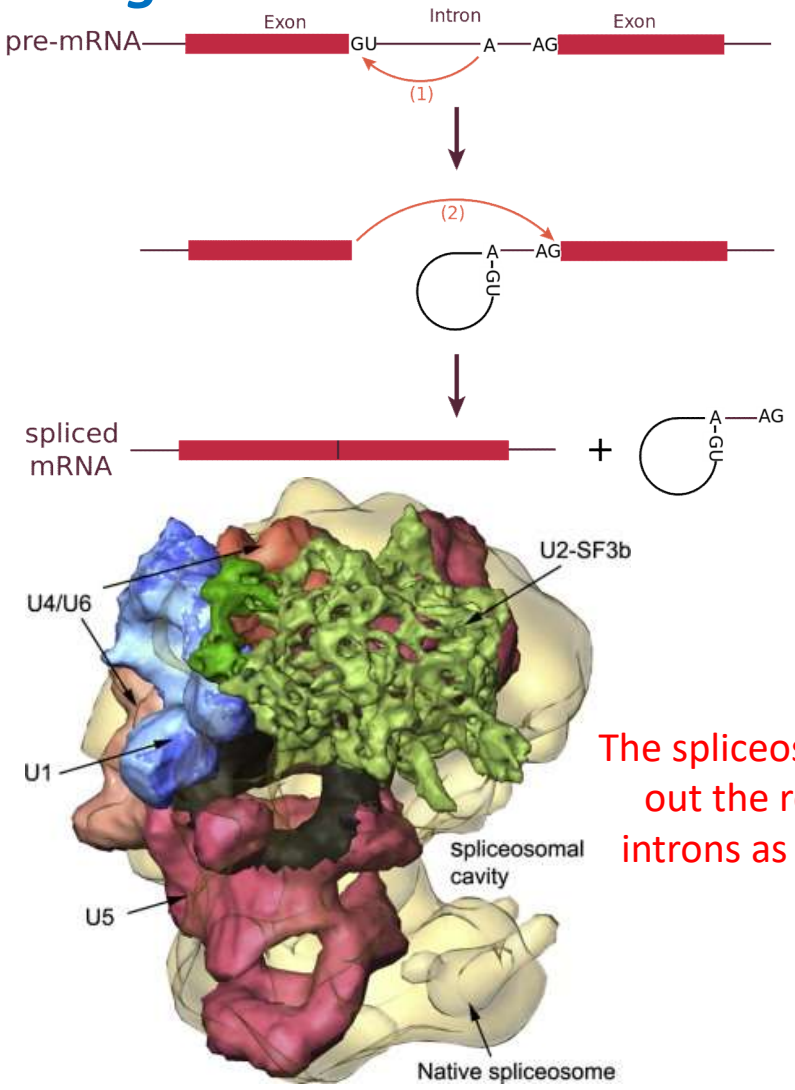


Bacterial RNAse P

mRNA processing



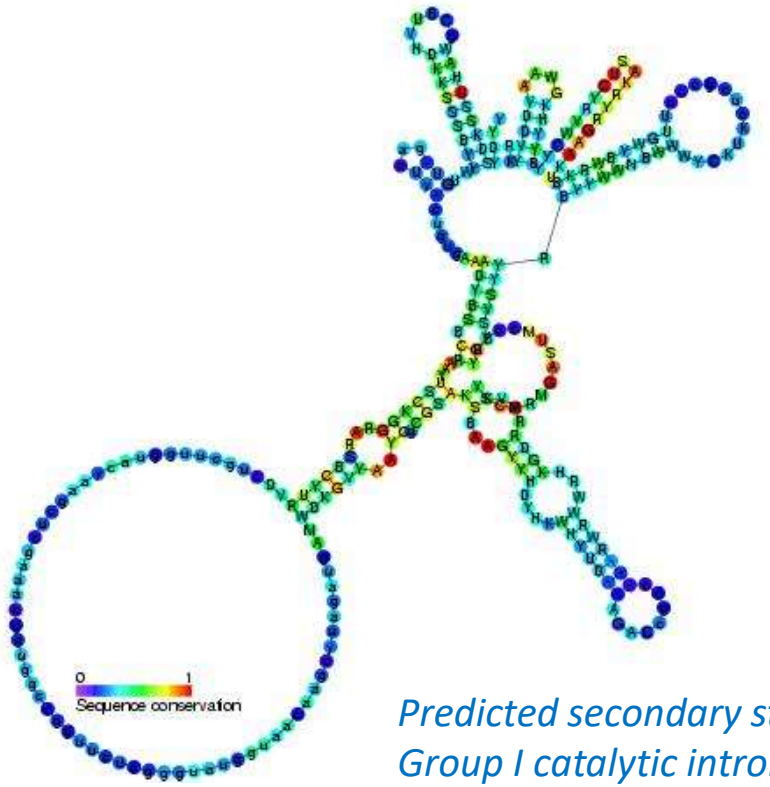
In 1977, Phil Sharp (Nobel Prize 1993) hybridized an mRNA to its DNA template and prepared the hybrid molecule for electron microscopy by coating the nucleic acid with a basic protein, then using rotary shadowing to coat the nucleic acid-protein complex.



The spliceosome carries out the removal of introns as RNA lariats

Spliceosome – a complex of ribonucleoproteins

RNA splicing

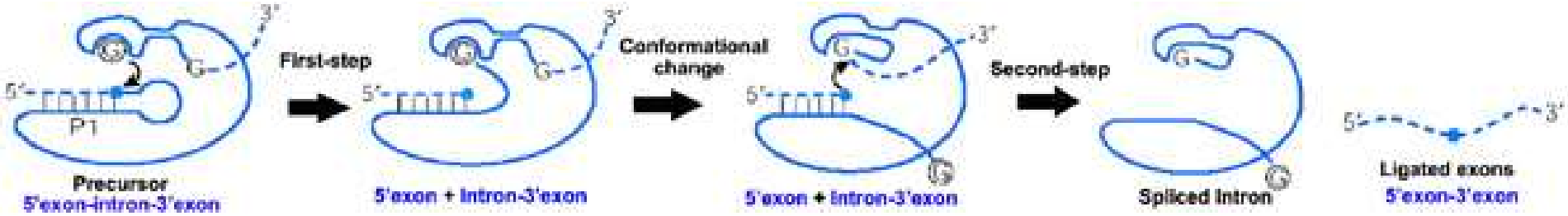


Predicted secondary structure and sequence conservation of Group I catalytic intron

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.

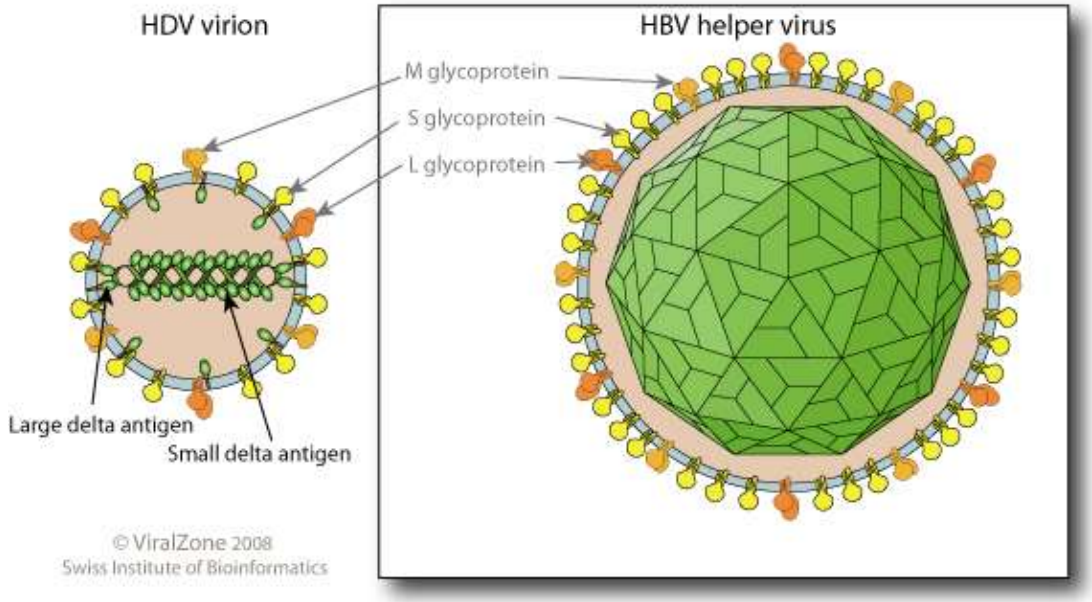


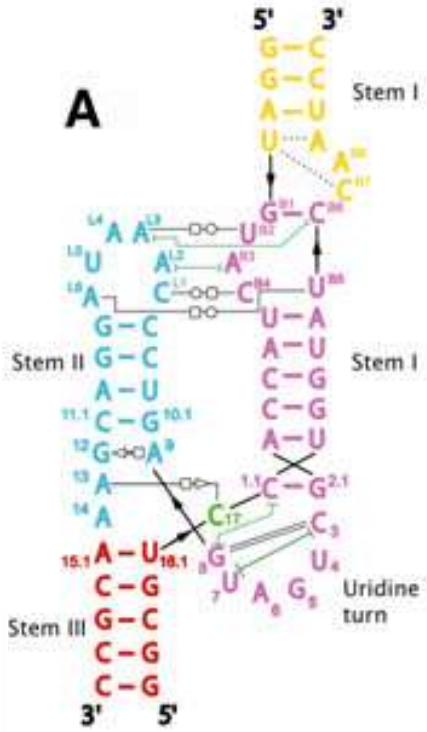
Ribozymes

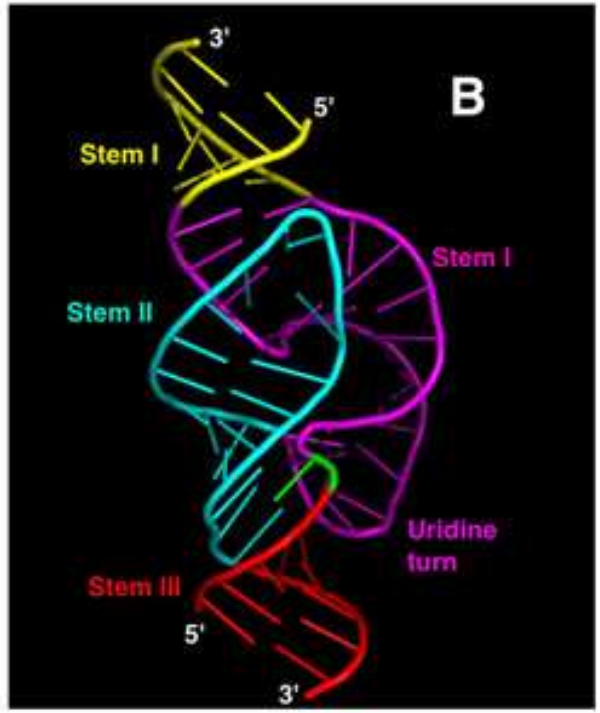
Hammerhead ribozyme

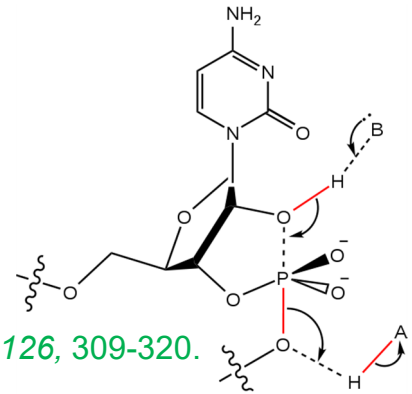
The hammerhead ribozyme is a RNA molecule motif that catalyzes reversible cleavage and joining reactions at a specific site within an RNA molecule (model system; targeted RNA cleavage experiments)

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.









M. Martick and W. G. Scott, *Cell* **2006**, *126*, 309-320.

Riboswitches

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

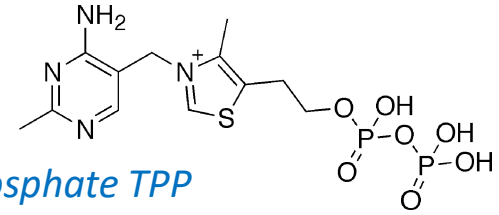
Riboswitches - naturally occurring regulatory segments of mRNA that bind small molecules specifically. The binding results in a change in production of the proteins encoded by the mRNA

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

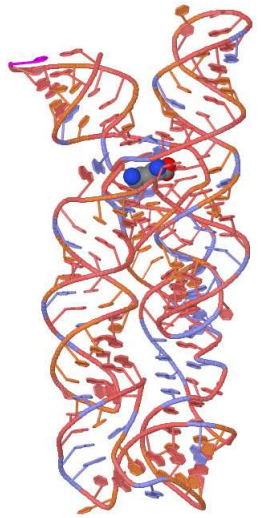
Most known ***riboswitches*** occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in archaea, plants and certain fungi.

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

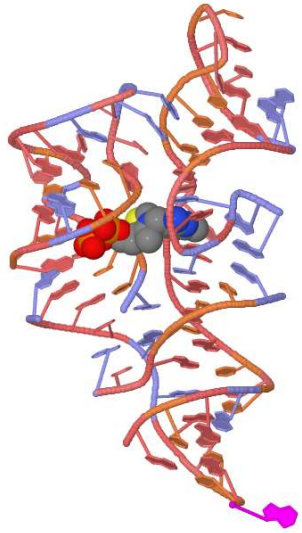
The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



Thiamine pyrophosphate TPP



The lysine riboswitch



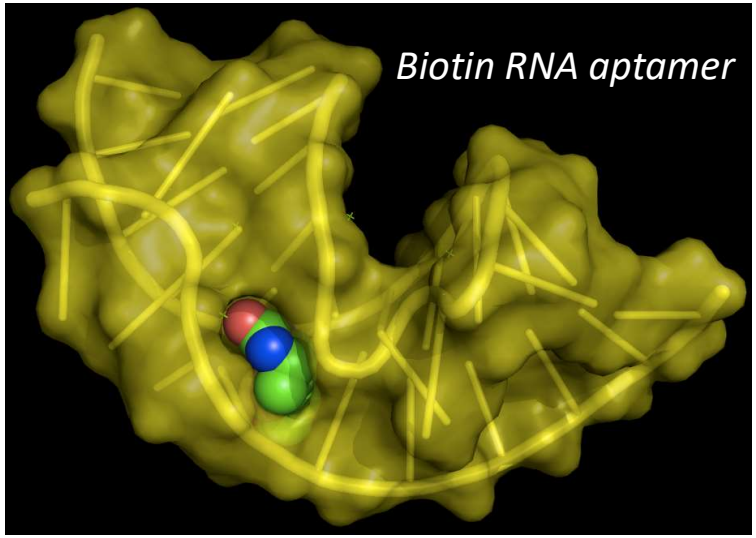
The 3D structure of TPP riboswitch (by Benjamin Schuster-Böckler)

Aptamers

Aptamers (from the Latin *aptus* – fit, and Greek *meros* – part) are ***oligonucleotide*** or ***peptide*** molecules that bind to a specific target molecule.

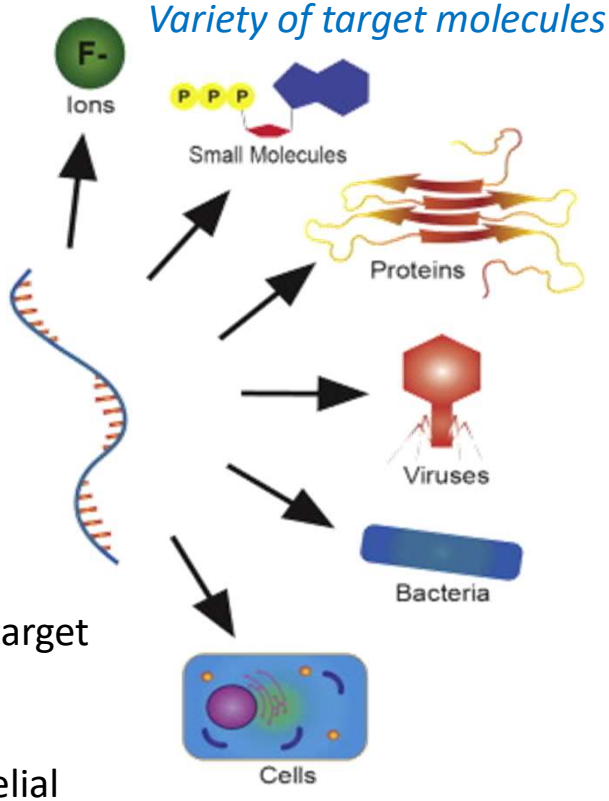
Aptamers are usually created by selecting them from a large random sequence pool, but natural aptamers also exist in riboswitches.

- ***DNA or RNA or XNA aptamers*** – oligonucleotide strands (usually short)
- ***Peptide aptamers*** - one (or more) short variable peptide domains, attached at both ends to a protein scaffold.



Fdardel

Variety of target molecules



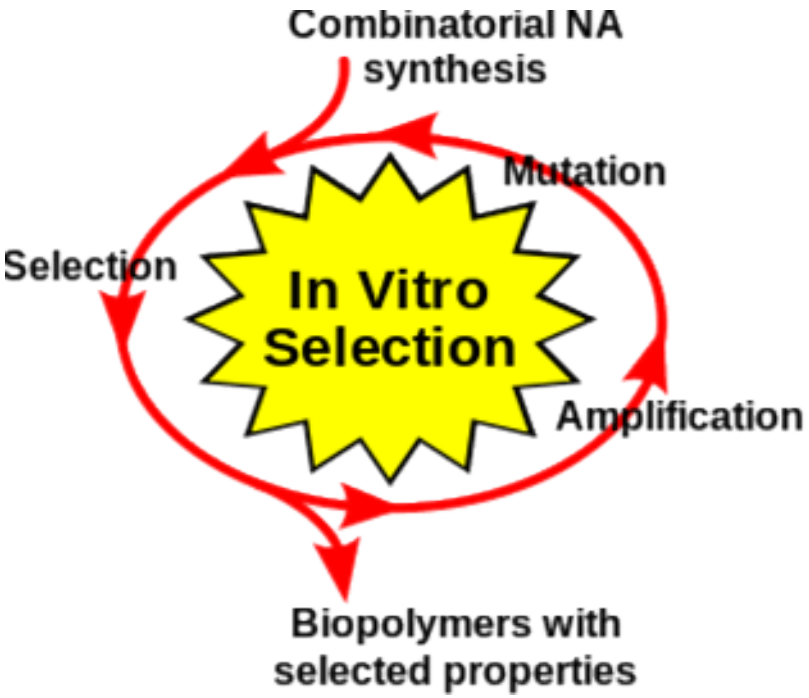
Aptamers were evolved for a variety of target ligands:

- small molecules (ATP and adenosine)
- proteins: prions and vascular endothelial growth factor (VEGF) - **MACUGEN**,
- tumor cells.

Additionally, SELEX has been utilized to obtain highly specific catalytic DNA or **DNAzymes**. Several metal-specific **DNAzymes** have been reported including the GR-5 DNAzyme (**lead-specific**), the CA1-3 DNAzymes (**copper-specific**), the 39E DNAzyme (**uranyl-specific**) and the NaA43 DNAzyme (**sodium-specific**).

Systematic evolution of ligands by exponential enrichment - SELEX

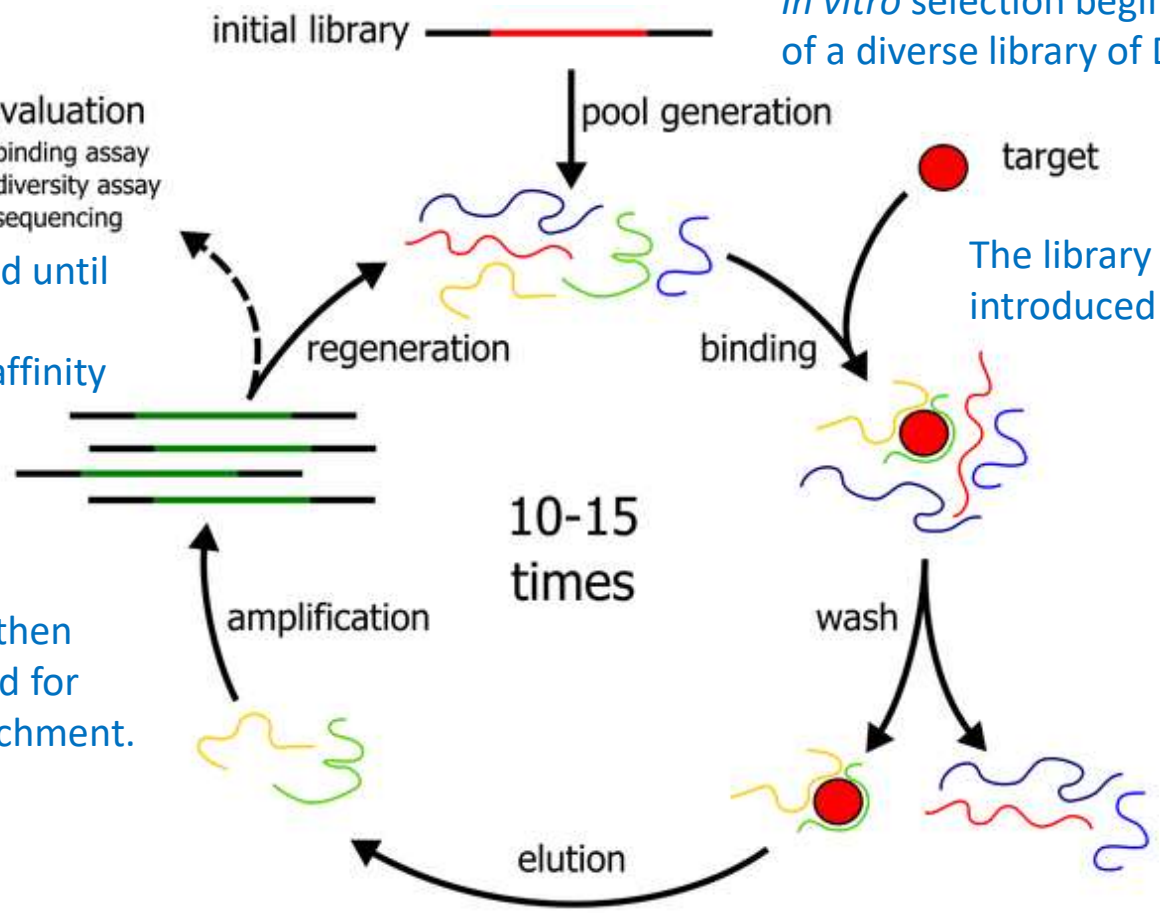
1990 – Gold *et al.* – selection of RNA ligands against T4 DNA polymerase
1990 – J. Szostak *et al.* – selecting RNA ligands towards organic dyes



A general overview of in vitro selection protocol. NA stands for Nucleic Acids (DNA, RNA) which start as a random pool, and are enriched through the selection process

Systematic evolution of ligands by exponential enrichment - SELEX

In vitro selection begins with the generation of a diverse library of DNA or RNA molecules.



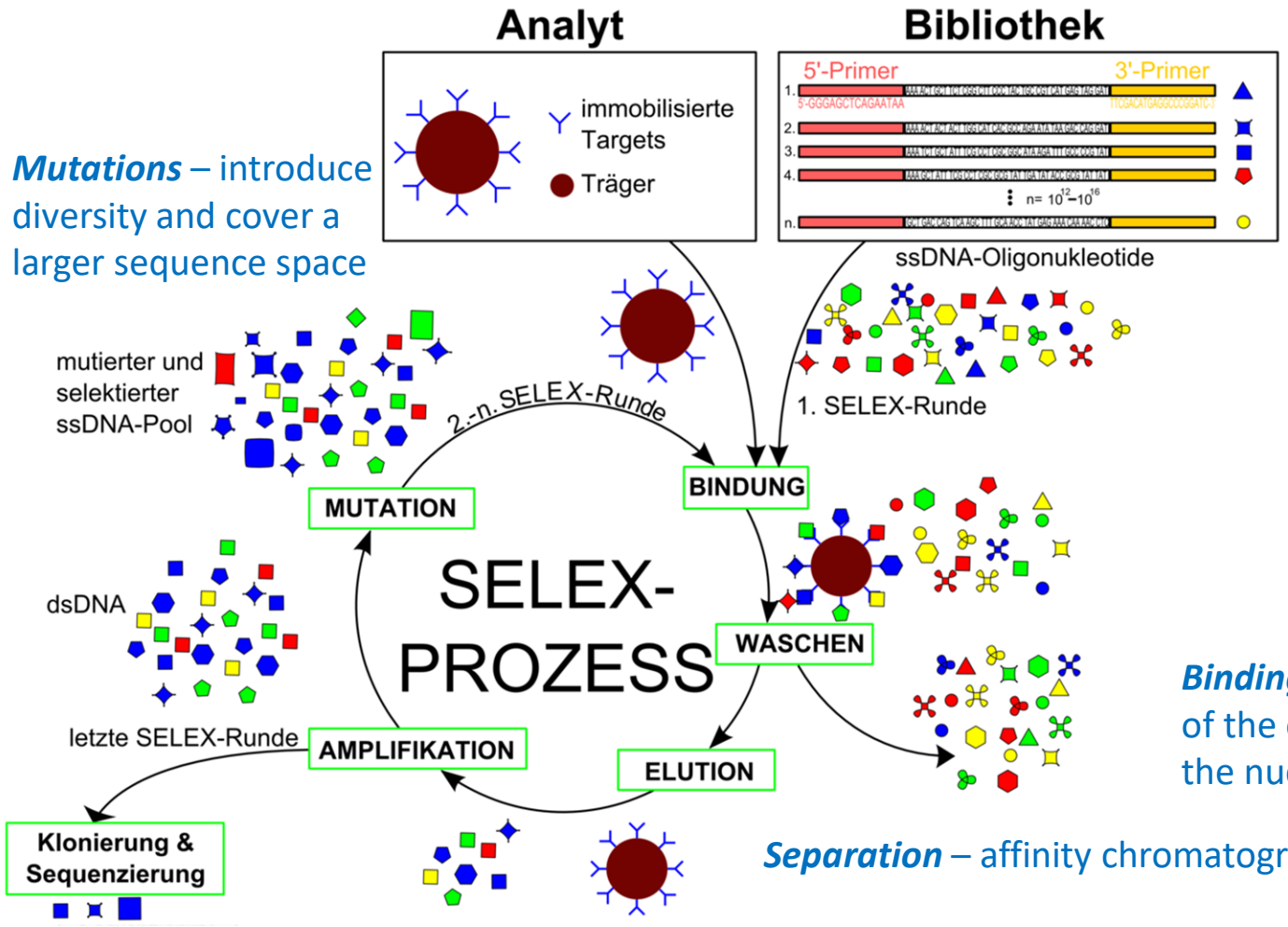
Multiple rounds are performed until the library converges on to a collection of sequences with affinity for the target molecule.

The library is then introduced to a target ligand

The bound sequences are then collected and PCR amplified for subsequent rounds of enrichment.

Sequences demonstrating affinity towards the target molecule are isolated from any unbound sequences.

Systematic evolution of ligands by exponential enrichment - SELEX



Mutations – introduce diversity and cover a larger sequence space

The initial library – usually contains sequences of 20-200 randomised nucleotides

The 5'- and 3'-termini are composed of constant primer sequences. Between them, the random sequences of *n* nucleotides result in 4^n combinations:

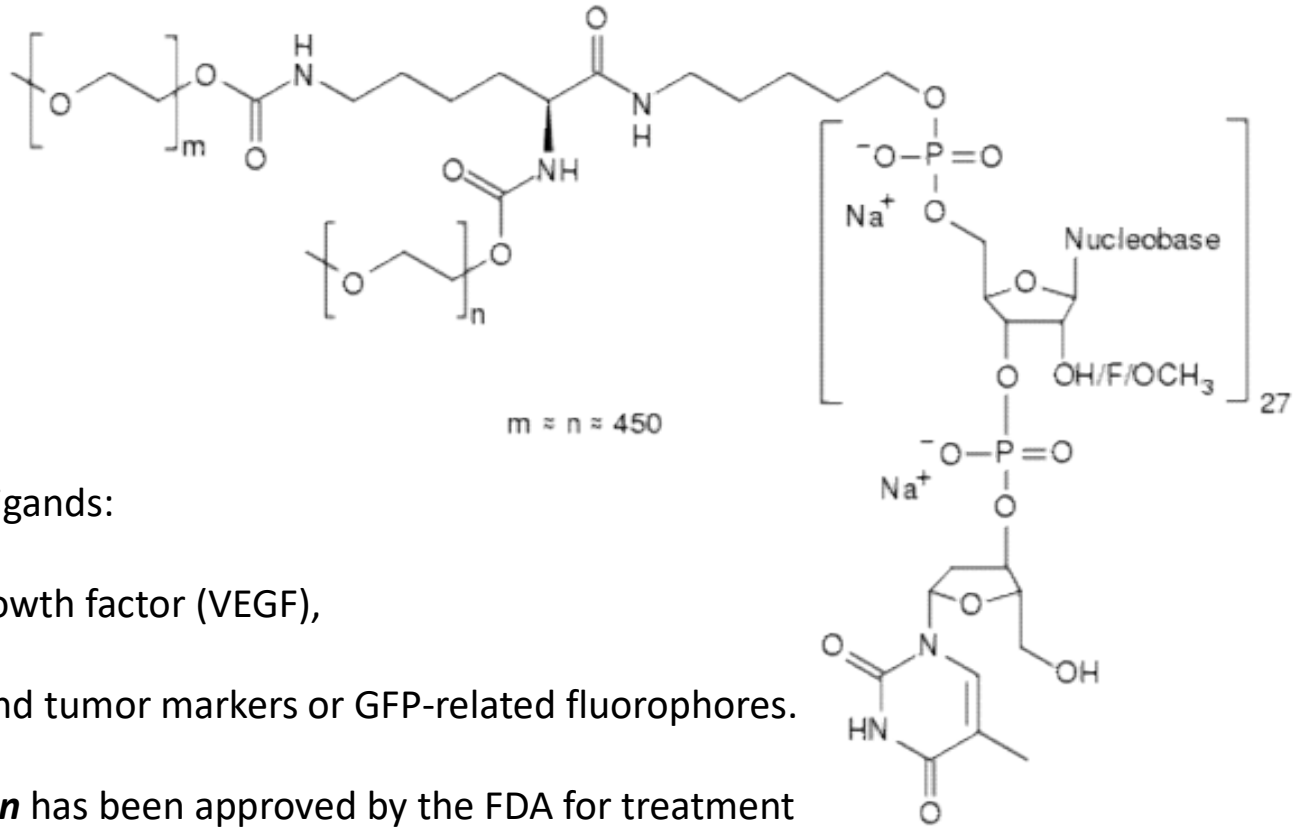
- n = 25 – a library of c.a. 10^{15} sequences,
- n = 50 – a library of c.a. 10^{30} sequences,
- n = 75 – a library of c.a. 10^{45} sequences,
- n = 100 – a library of c.a. 10^{60} sequences.

Binding strongly depends on the secondary structure of the oligonucleotides, which in turn depends on the nucleobase sequence

Separation – affinity chromatography

Amplification – reverse transcription RNA→DNA (only for RNA aptamers) + PCR (for RNA and DNA aptamers)

Systematic evolution of ligands by exponential enrichment - SELEX



Aptamers were evolved for a variety of target ligands:

- small molecules (ATP and adenosine)
- proteins: prions and vascular endothelial growth factor (VEGF),
- tumor cells.

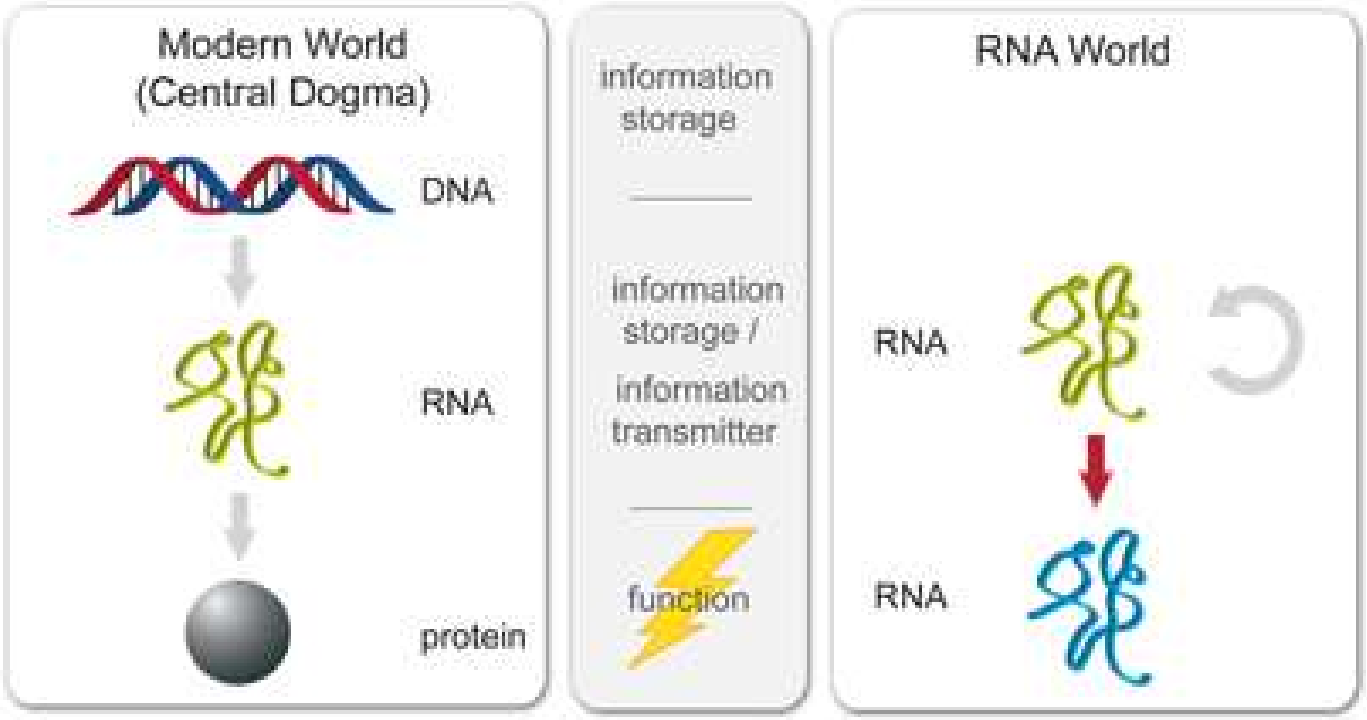
Clinical uses are suggested by aptamers that bind tumor markers or GFP-related fluorophores.

A VEGF-binding aptamer trade-named ***Macugen*** has been approved by the FDA for treatment of macular degeneration.

Macugen

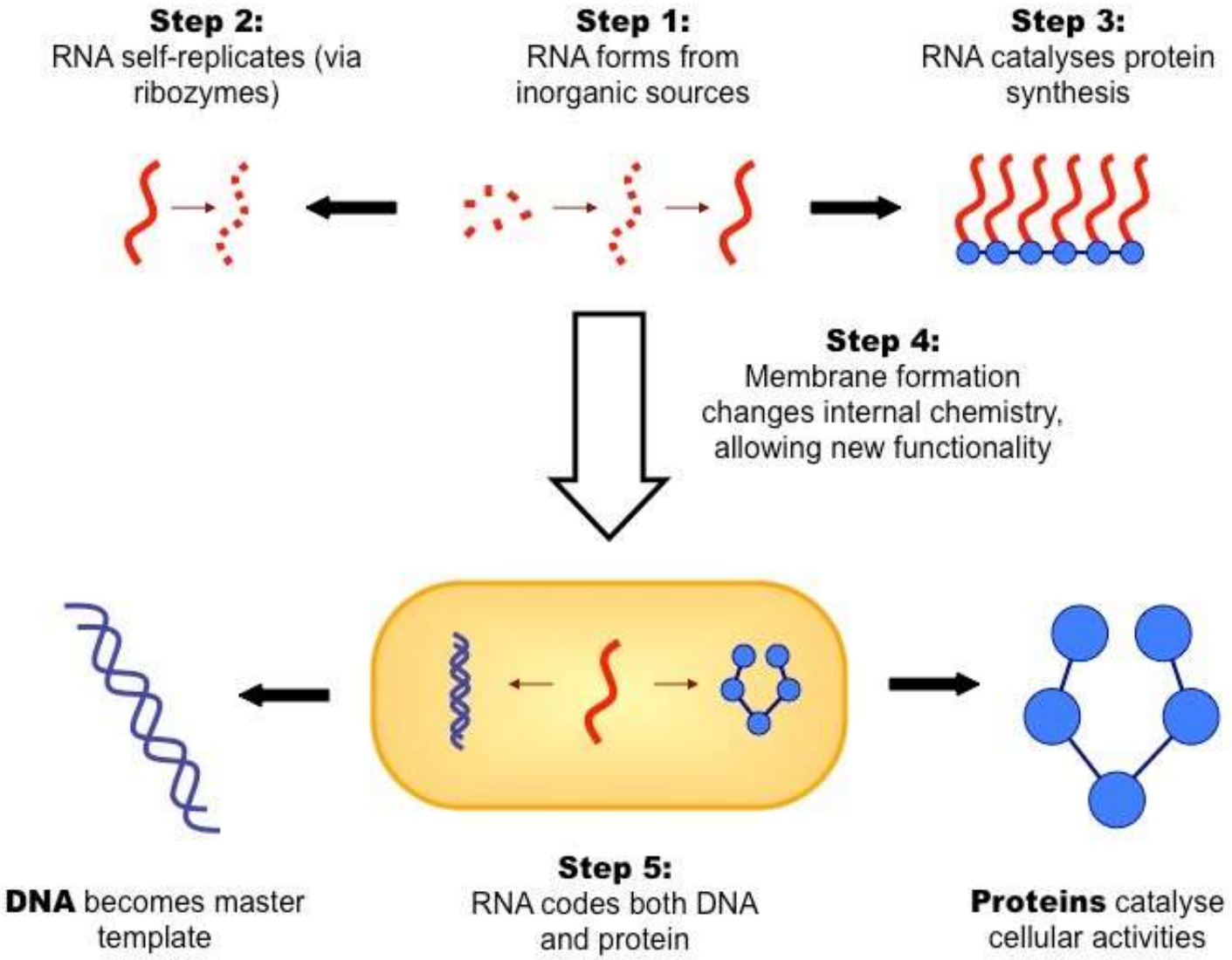
Additionally, SELEX has been utilized to obtain highly specific catalytic DNA or DNAzymes. Several metal-specific DNAzymes have been reported including the GR-5 DNAzyme (lead-specific), the CA1-3 DNAzymes (copper-specific), the 39E DNAzyme (uranyl-specific) and the NaA43 DNAzyme (sodium-specific).

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

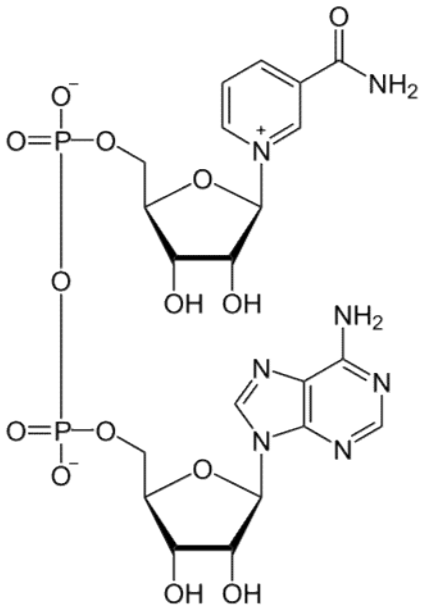


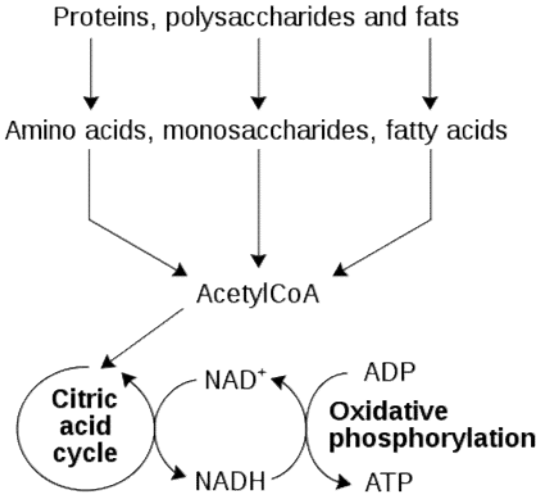
The RNA world

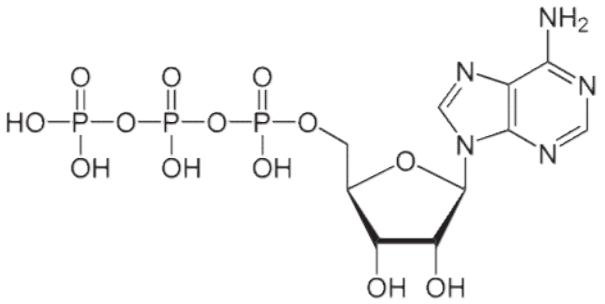
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

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 primoridal RNA-based metabolism







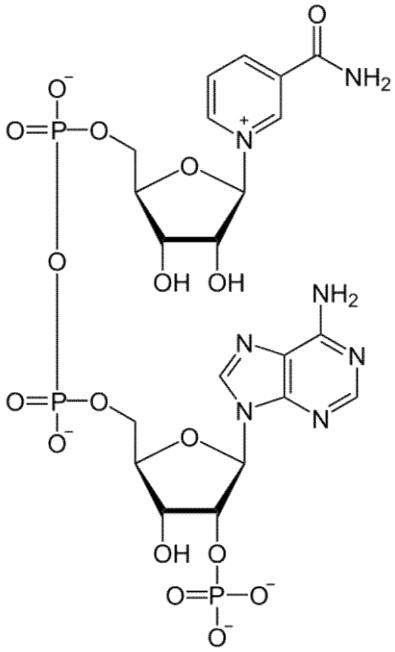
Nicotinamide adenine dinucleotide (NAD$^+$)

Adenosine triphosphate (ATP)

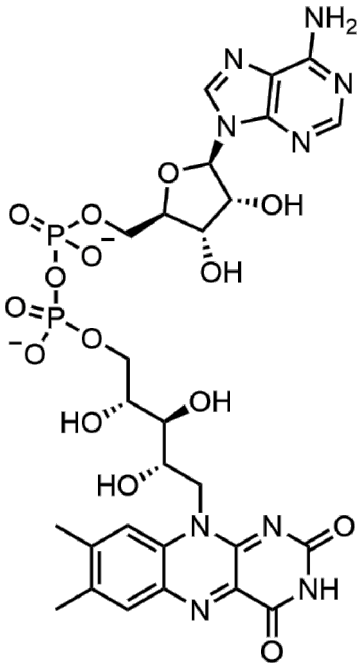
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

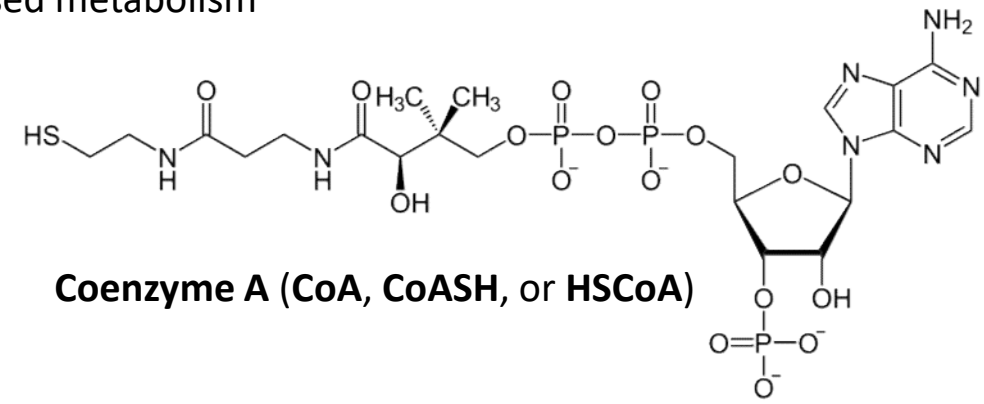
Ribonucleotide coenzymes now used by many proteins may be molecular „fossils” from the primoridal RNA-based metabolism



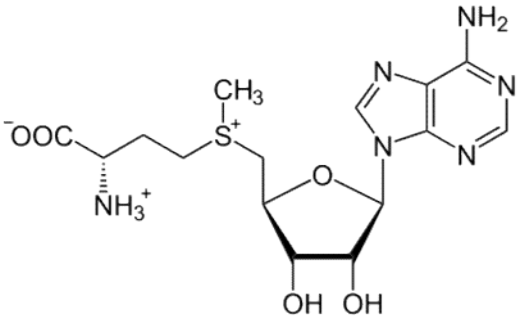
Nicotinamide adenine dinucleotide phosphate ($NADP^+$)



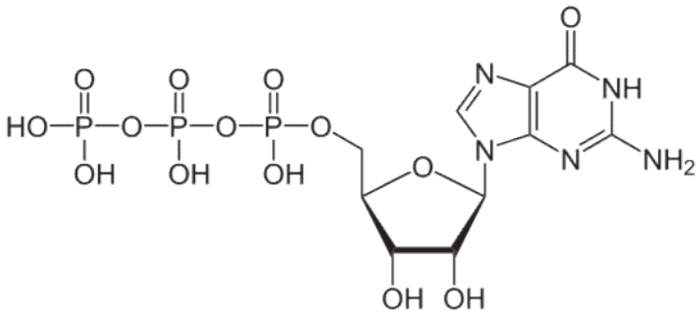
flavin adenine dinucleotide (FAD)



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



***S*-Adenosyl methionine**

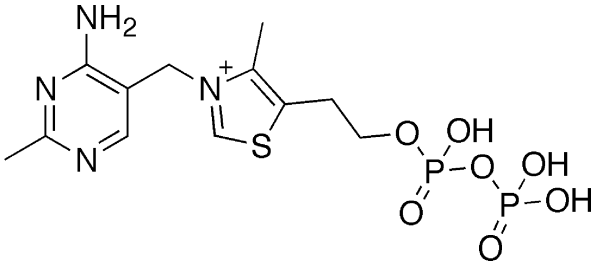


Guanosine-5'-triphosphate (**GTP**)

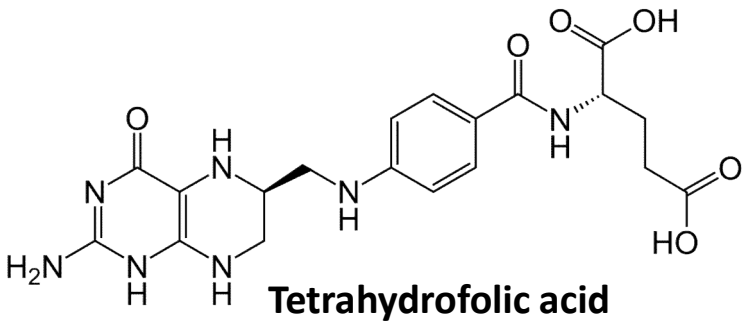
H. B. White III *J. Mol. Evol.* **1976**, *7*, 101-104

The RNA world

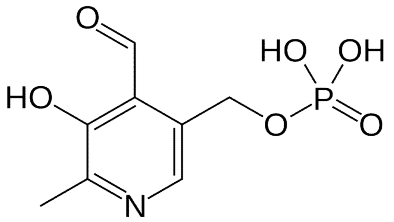
Other coenzymes contain cyclic nitrogen-containing bases that can also derive from nucleotides



Thiamine pyrophosphate (TPP or ThPP) – Vit. B_1



Tetrahydrofolic acid

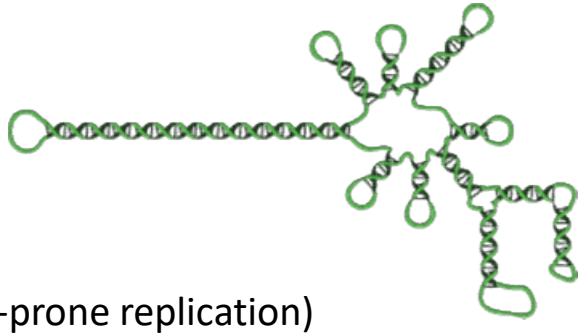


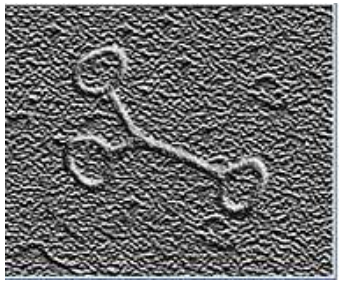
Pyridoxal phosphate (PLP) – Vit. B_6

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

Viroids

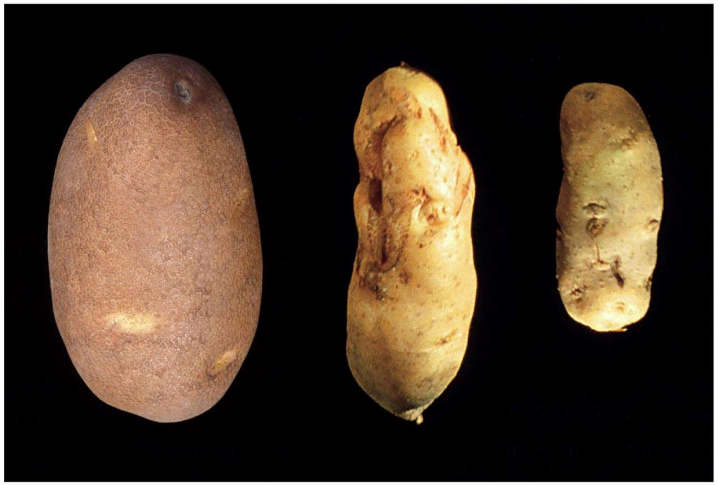
Viroids ("subviral pathogens„) are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Viroids are extremely small - 246 to 467 nucleobases (genomes of smallest viruses start from 2,000 nucleobases). Viroids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



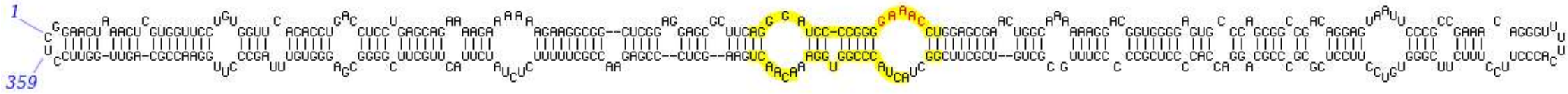


Viroid properties:

- 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.



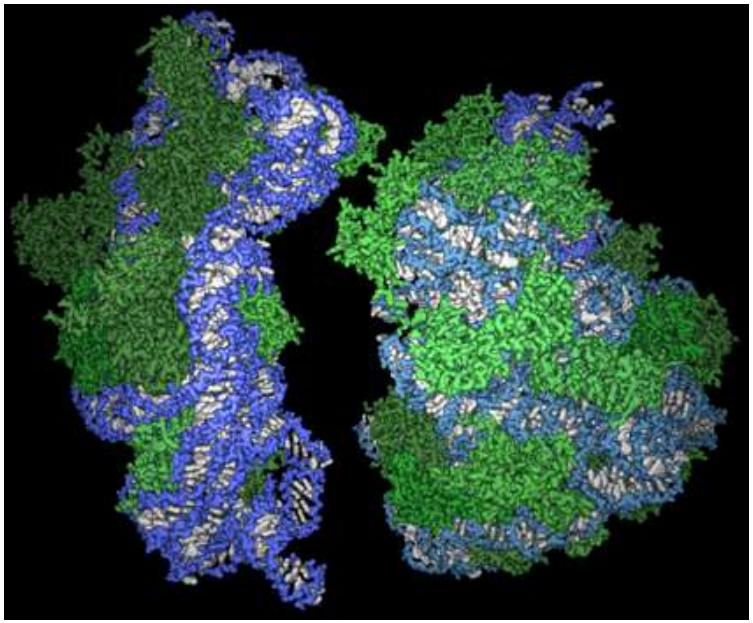
PSTVd-infected potatoes (right)



Putative secondary structure of the PSTVd viroid

No virion reported. Viroids do not encode for proteins

Ribosome



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

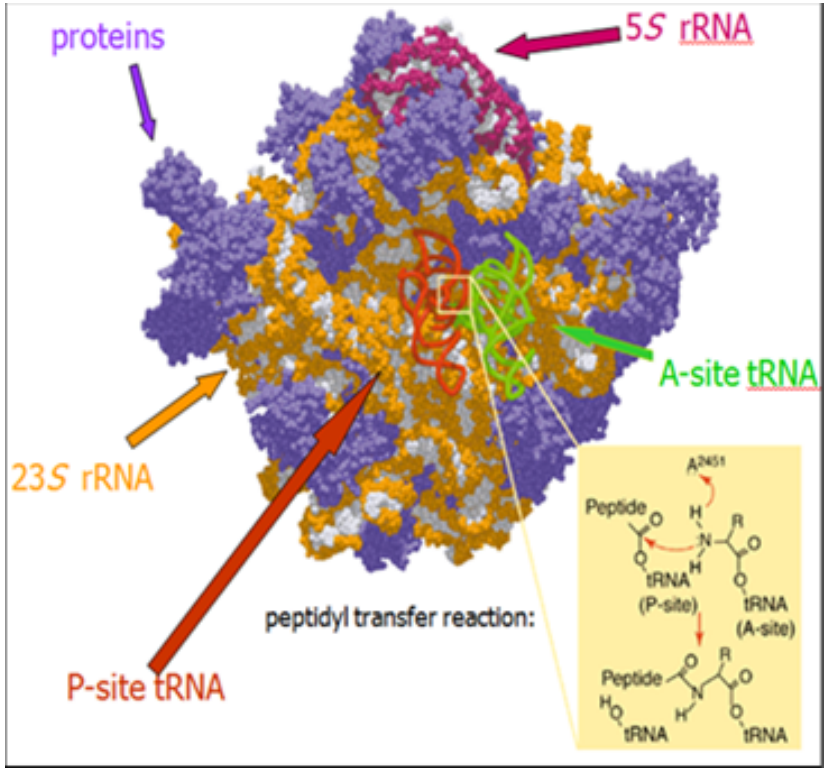
Ribosomes:

- 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 is a ribozyme!

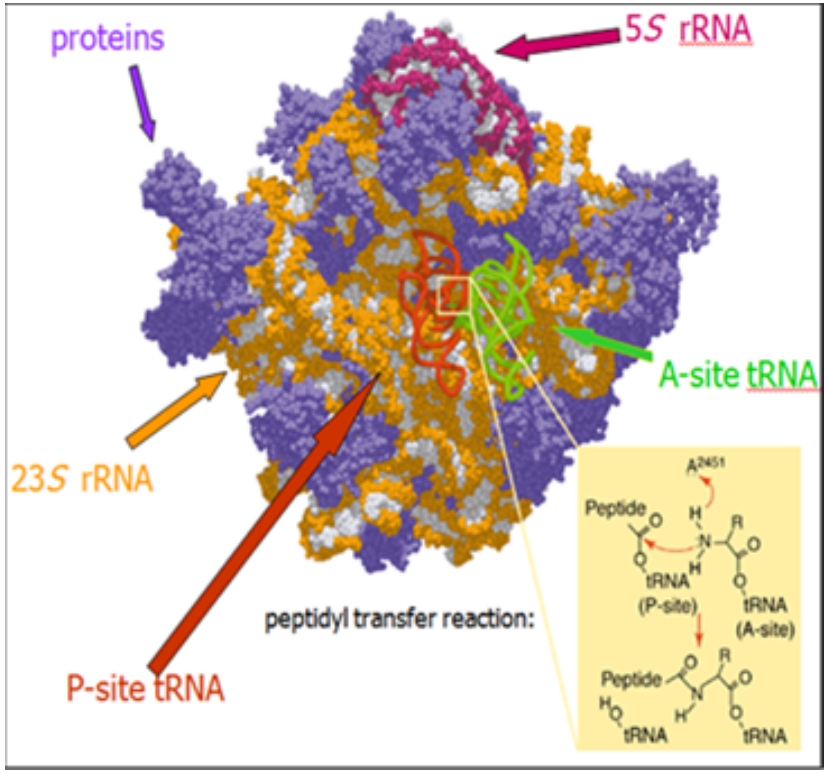


No protein is present within 18 Angstroms from the active site → proteins play a structural role, but DO NOT CATALYZE THE ACYL TRANSFER PROCESS

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

Ribosome – the ‚smoking gun'

Ribosome is a ribozyme!



The proto-ribosomes in the RNA world – first as a self-replicating complex, later evolved the ability to synthesize proteins with emerging amino acids.

Early proto-ribosomes were self-replicating complexes:
the rRNA had informational, structural, and catalytic purposes –
it coded for tRNAs and proteins needed for ribosomal self-replication.

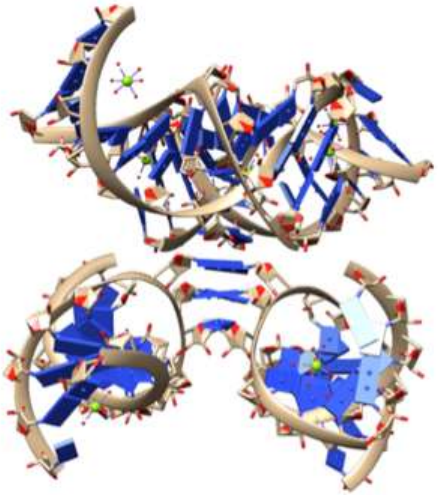
Emerging amino acids interacted with catalytic RNA:
increased scope and efficiency of catalytic RNA molecules.

→ Ability to synthesize peptide bonds was caused by the evolutionary pressure to increase its capacity for self-replication by incorporating proteins into the catalysis

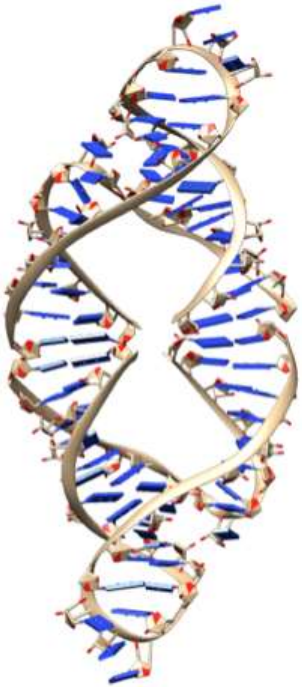
Ribozymes

The most common activities of natural or in vitro-evolved ribozymes are the cleavage or ligation of RNA and DNA and peptide bond formation.

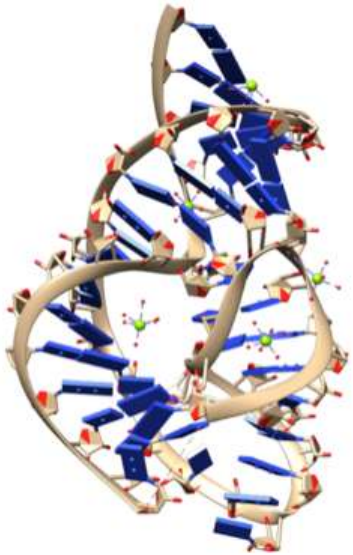
Ribozymes participate in a variety of RNA processing reactions, including RNA splicing, viral replication, and transfer RNA biosynthesis. **Examples** of ribozymes include the **hammerhead ribozyme**, the **VS ribozyme**, **Leadzyme** and the **hairpin ribozyme**.



Leadzyme



the **hammerhead ribozyme**,



the **twister ribozyme**.

by: Lucasharr

Ribozymes

The smallest ribozyme is UUU, which can promote the cleavage between G and A of the GAAA tetranucleotide via the S_{N2} mechanism in the presence of Mn^{2+}.

Attempts have been made to develop ribozymes as therapeutic agents, as enzymes which target defined RNA sequences for cleavage, as biosensors, and for applications in functional genomics and gene discovery

Many ribozymes have either a hairpin – or hammerhead – shaped active center and a unique secondary structure that allows them to cleave other RNA molecules at specific sequences.

Ribozymes that specifically cleave any RNA molecule may have **pharmaceutical applications**. For example, a ribozyme has been designed to **cleave the RNA of HIV**. If such a ribozyme were made by a cell, all incoming virus particles would have their RNA genome cleaved by the ribozyme, which would prevent infection.

The hammerhead ribozyme

The **hammerhead ribozyme** is an RNA motif that catalyzes reversible cleavage and ligation reactions at a specific site within an RNA molecule; the best-characterized ribozyme.

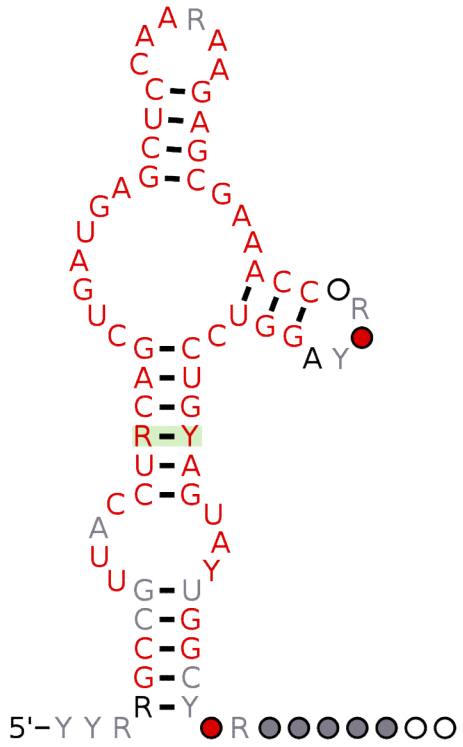
In its natural state, a hammerhead RNA motif is a self-cleaving single strand of RNA (not catalytic, it is consumed by the reaction ➔ no multiple turnovers)

Trans-acting hammerhead constructs: two interacting RNA strands - a hammerhead ribozyme that cleaves the other strand. The strand that gets cleaved can be supplied in excess, and multiple turnover can be demonstrated and shown to obey Michaelis-Menten kinetics, typical of protein enzyme kinetics.

The **minimal** *trans*-acting **hammerhead ribozyme** sequence: three base-paired stems flanking a central core of 15 conserved (mostly invariant) nucleotides.

In **eukaryotic genomes** (incl. **humans**), hammerhead ribozymes related e.g. to short interspersed retroelements (SINEs). These hammerhead ribozymes (the so-called HH9 and HH10) occur in the introns of a few specific genes and point to a preserved biological role during pre-mRNA biosynthesis

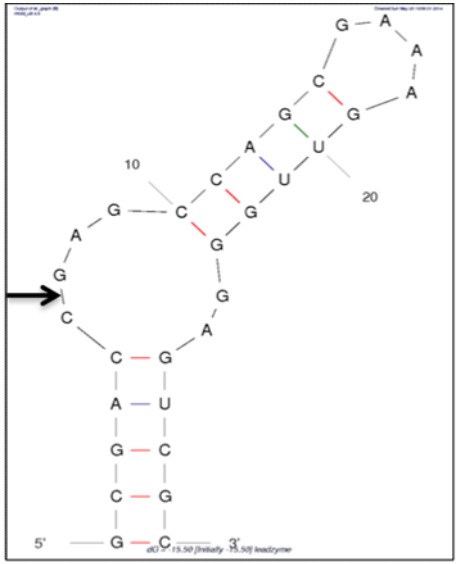
Forster AC, Symons RH, *Cell* **1987**, *49 (2),* 211–220



Predicted secondary structure and sequence conservation of the HH9 ribozyme found conserved from lizard to human genomes

Leadzyme

Leadzyme is a small ribonuclease ribozyme. It was discovered using an in-vitro evolution study - selection for RNAs that specifically cleaved themselves in the presence of lead. However, since then, it has been discovered in several natural systems



Anarkalimahmood

The **minimal secondary structure** of leadzyme is surprisingly simple: an asymmetric internal loop (6 nt) and a helical region on each side of the internal loop. The cleavage site of leadzyme is located within a four-nucleotide long asymmetric internal loop that also consists of RNA helices on its both sides.

Since leadzyme is a relatively simple motif, many sequences in the genomes of many natural systems which can potentially fold into a leadzyme structure. A simple search for this RNA motif in the genomes of humans, *D. melanogaster, C. elegans* and *A. thaliana* revealed that on average this motif is present with the frequency of 2-9 motifs for 1 Mbp of DNA sequence.

These transcripts could potentially self-cleave in the presence of lead ions. The targeting of these RNA motifs by lead in mRNAs and other RNAs may explain lead-mediated toxicity resulting in cell death

Pan, T.; Uhlenbeck, O. C., *Nature*. **1992,** *358 (6387)*, 560–563;
Pan, T.; Uhlenbeck, O. C., *Biochemistry* **1992**, *31 (16),* 3887–3895

Class I ligase ribozymes

No known RNA enzyme in biology catalyzes the polymerase-like joining of RNA. However, *in vitro* evolution have made it possible to generate such enzymes from scratch, starting from a large population of RNAs with random sequences
The **RNA Ligase ribozyme** was the first of several types of synthetic ribozymes produced by *in vitro* evolution and selection techniques.
It catalyzes the assembly of RNA fragments into phosphodiester RNA polymers.
Ligase ribozymes may have been significant part of the RNA world.

Bartel, D. P., and Szostak, J. W., *Science* **1993** *261*, 1411-1418.

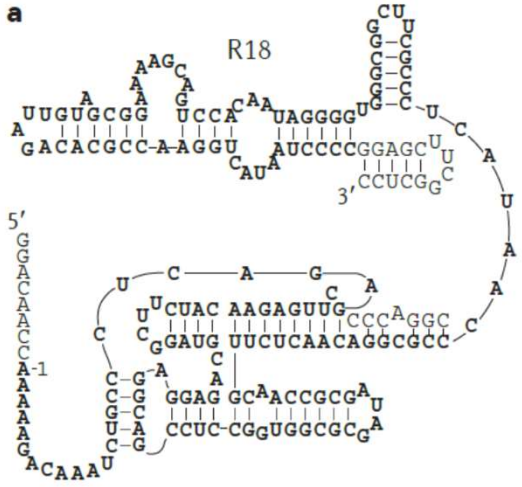
David Schechner

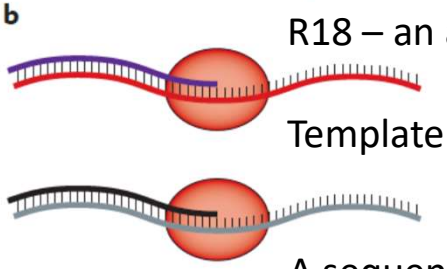
Then, through further evolution, the researcher attempts to coax the ligase to accept NTPs as substrates and to add multiple NTPs in succession. The class I ligase has been evolved further to polymerize as many as 14 successive NTPs with high fidelity

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

The RNA world

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

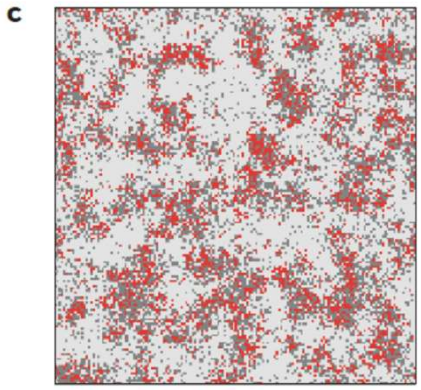


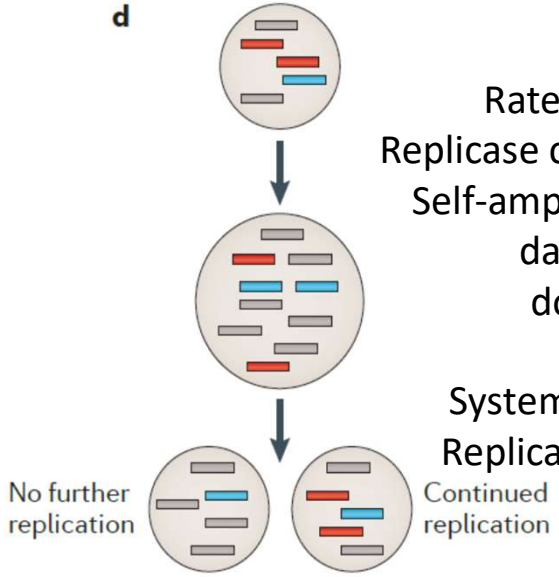


R18 – an artificial polymerase evolved from the class I ligase ribozyme.

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.

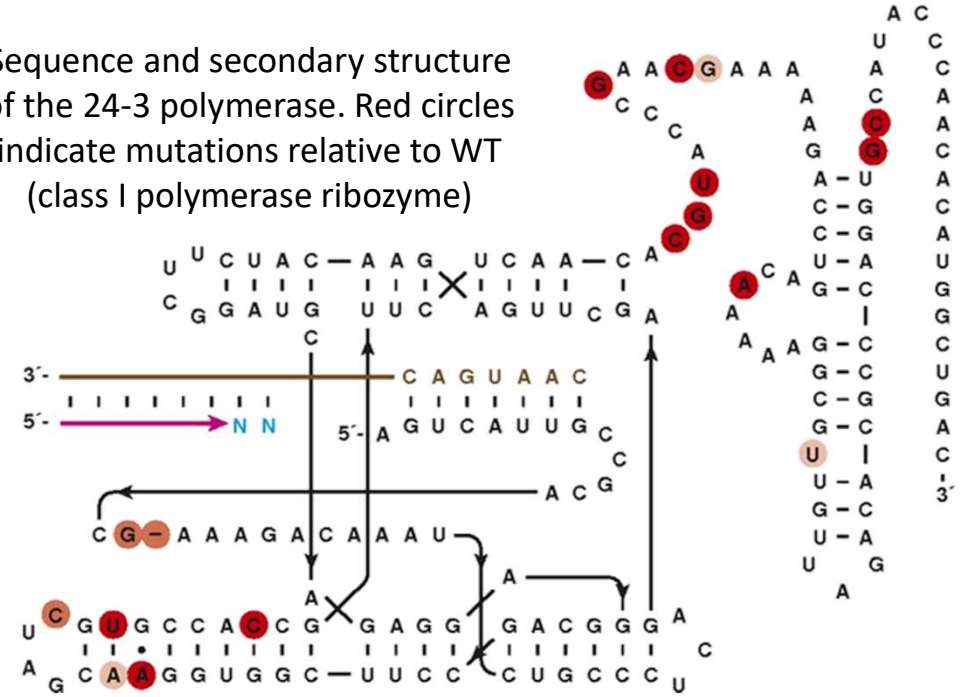
Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **2013**, *5*, 1011–1018.

The RNA world

RNA-dependent RNA polymerase ribozyme – Replicase

In vitro evolution of an improved RNA polymerase ribozyme that is able to synthesize structured functional RNAs, including aptamers and ribozymes, and replicate short RNA sequences in a protein-free form of the PCR.

Sequence and secondary structure of the 24-3 polymerase. Red circles indicate mutations relative to WT (class I polymerase ribozyme)

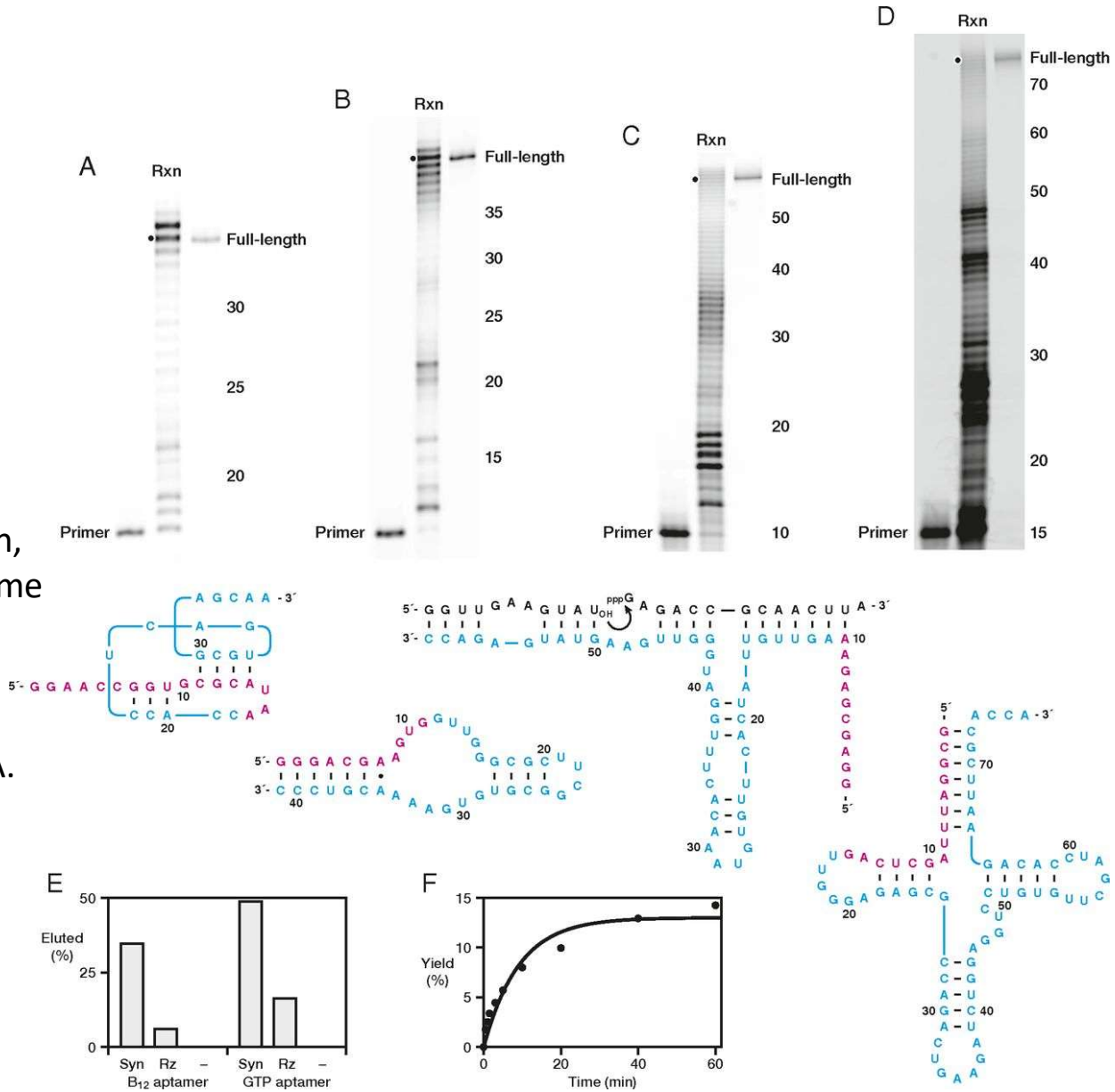


Thus, the replication of RNA and the expression of functional RNA can be accomplished with RNA alone. Combining and improving these activities may enable the self-sustained evolution of RNA and offers a potential route to a synthetic form of RNA life.

D. P. Horning, G. F. Joyce *Proc. Natl. Acad. Sci. USA (PNAS)* **2016**, *113 (35)*, 9786-9791

The RNA world

RNA-dependent
RNA polymerase ribozyme
– Replicase



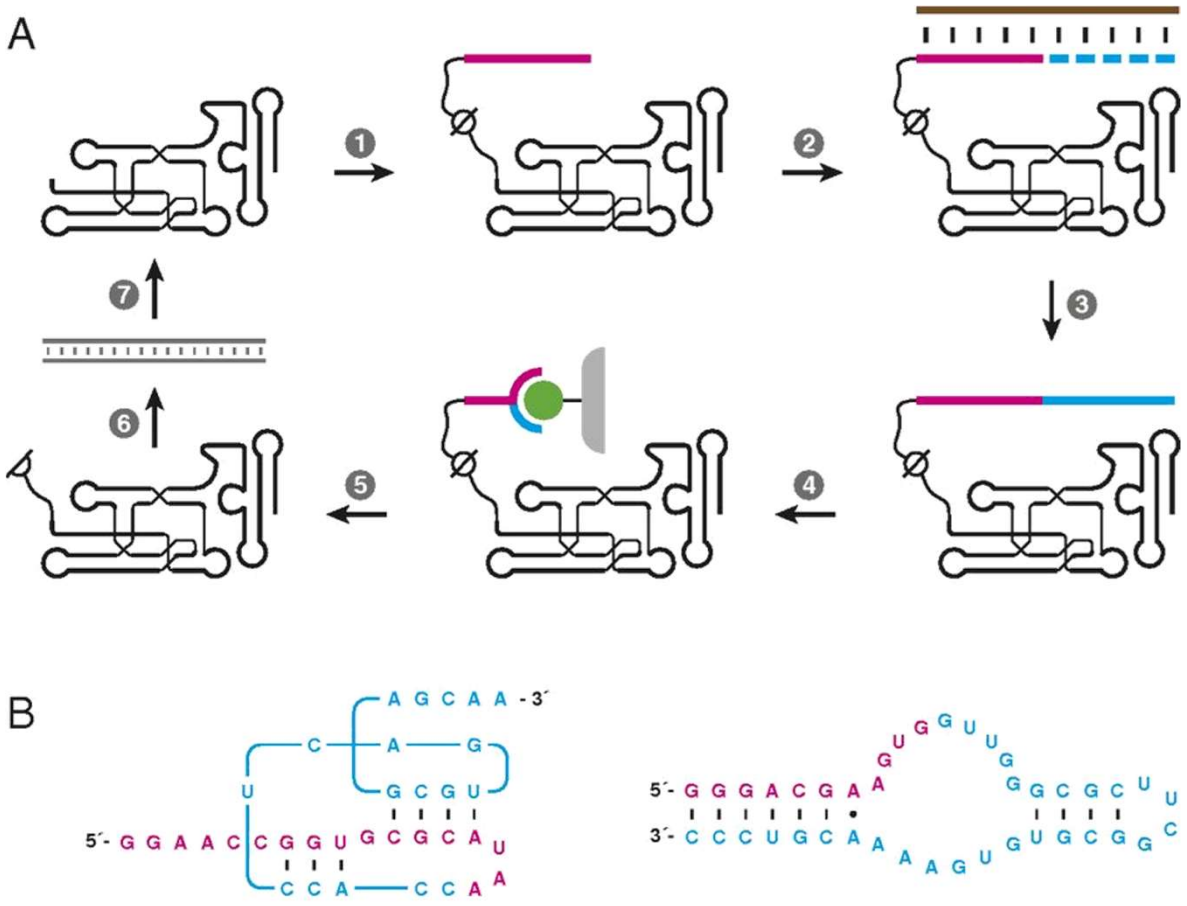
Synthesis of functional RNAs by the 24-3 polymerase. Synthesis of (*A*) the cyanocobalamin aptamer after 24 h, (*B*) the GTP aptamer after 24 h, (*C*) the F1 ligase ribozyme after 24 h, and (*D*) yeast phenylalanyl-tRNA after 72 h.

Sequence and secondary structure of the primer (magenta) and polymerized portion (cyan) of each RNA.

D. P. Horning, G. F. Joyce
Proc. Natl. Acad. Sci. USA (PNAS) **2016**, *113 (35)*, 9786-9791

The RNA world

RNA-dependent RNA polymerase ribozyme – Replicase



The WT polymerase was challenged to extend the attached primer to complete a 3′-truncated RNA aptamer (B_{12} or GTP), enabling selection based on binding of the completed aptamer to its cognate ligand. Selection pressure for both sequence generality and accuracy.

In vitro evolution of RNA polymerase ribozymes. (A) Selective amplification of ribozymes that extend a tethered RNA primer (magenta) on a separate RNA template (brown) to complete a 3′-truncated aptamer. (1) Attachment of the primer to the ribozyme via a photocleavable linker; (2) hybridization of the primer to the template; (3) extension of the primer by polymerization of NTPs (cyan); (4) capture of full-length materials by binding the aptamer portion to its immobilized ligand (green); (5) photocleavage to release the ribozyme portion; (6) reverse transcription and PCR amplification of the released ribozyme; and (7) transcription to generate progeny ribozymes. (B) Sequence and secondary structure of RNA aptamers that bind either cyanocobalamin (Left) or GTP (Right).

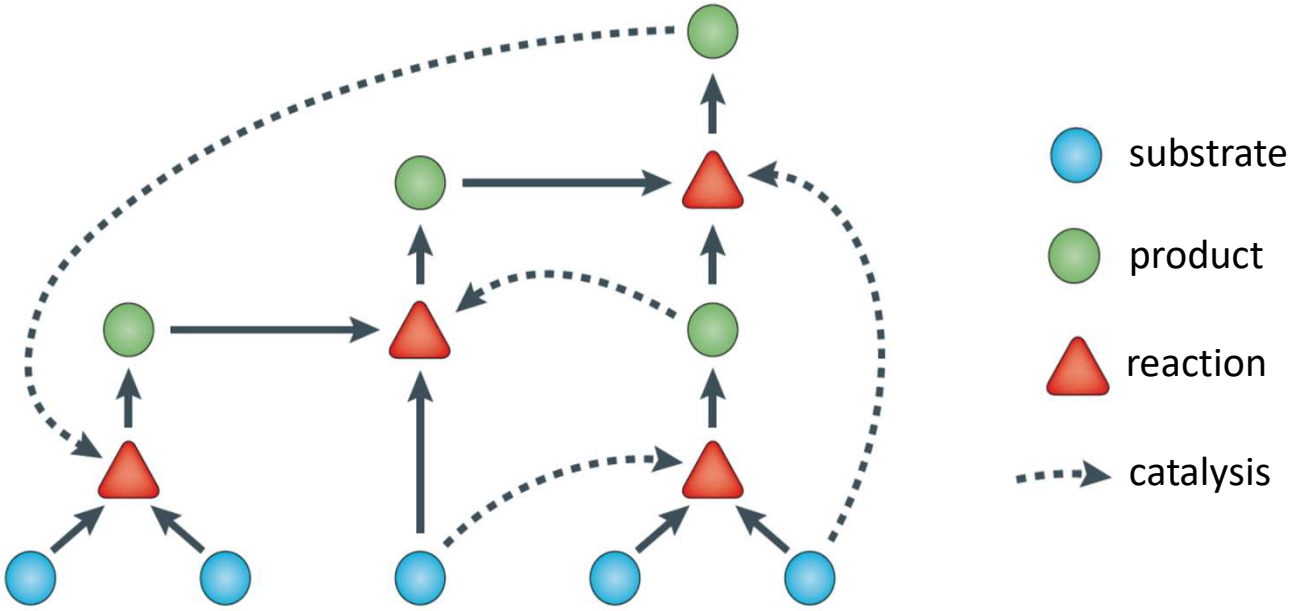
D. P. Horning, G. F. Joyce *Proc. Natl. Acad. Sci. USA (PNAS)* **2016**, *113 (35)*, 9786-9791

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?

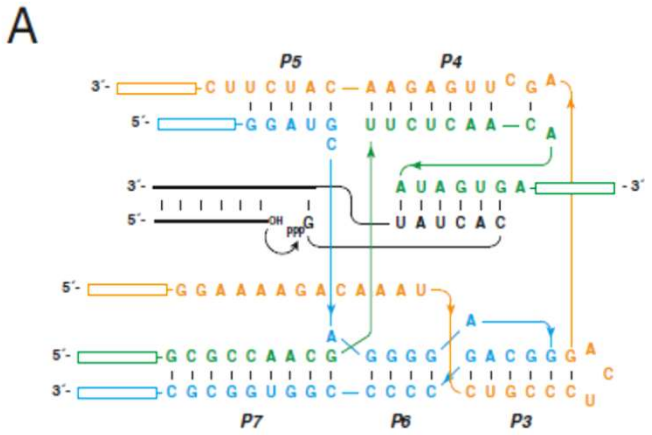
Possible solution – autocatalytic networks



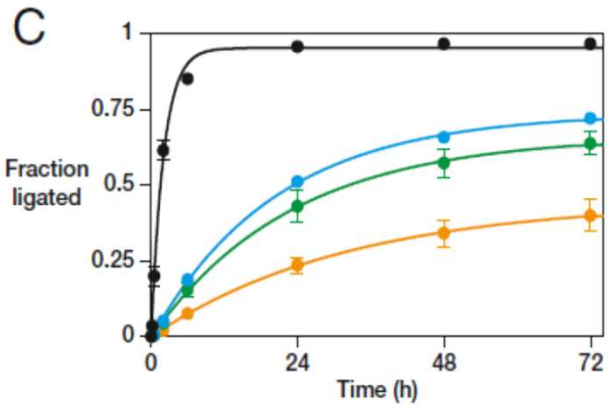
No component can replicate without all the others

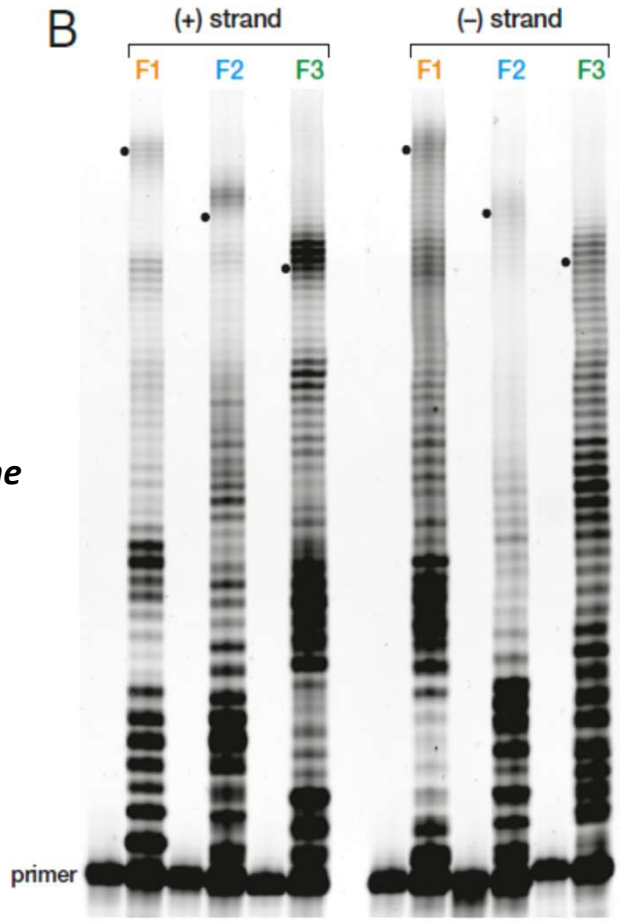
The RNA world

The polymerase „38-6” is able to synthesize its own evolutionary ancestor, an RNA ligase ribozyme, in the form of three fragments that assemble to give a functional complex

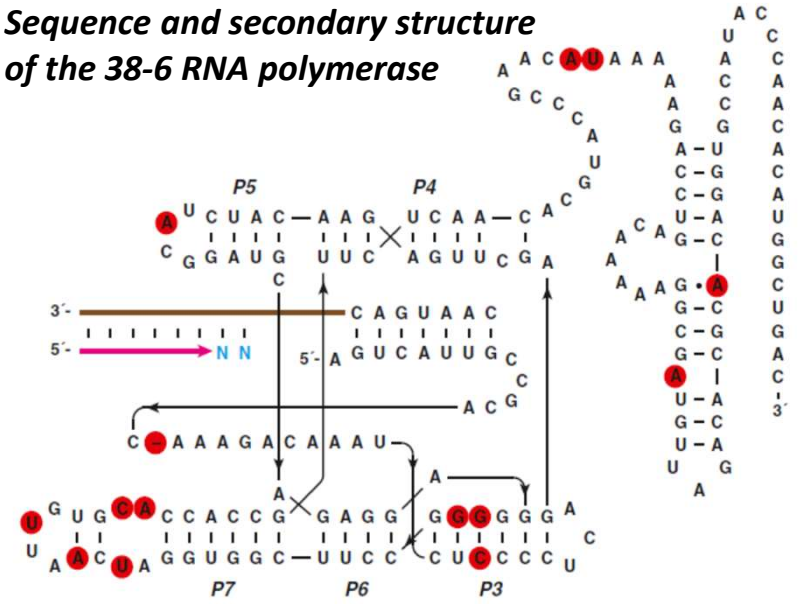


Three fragments of the class I ligase ribozyme





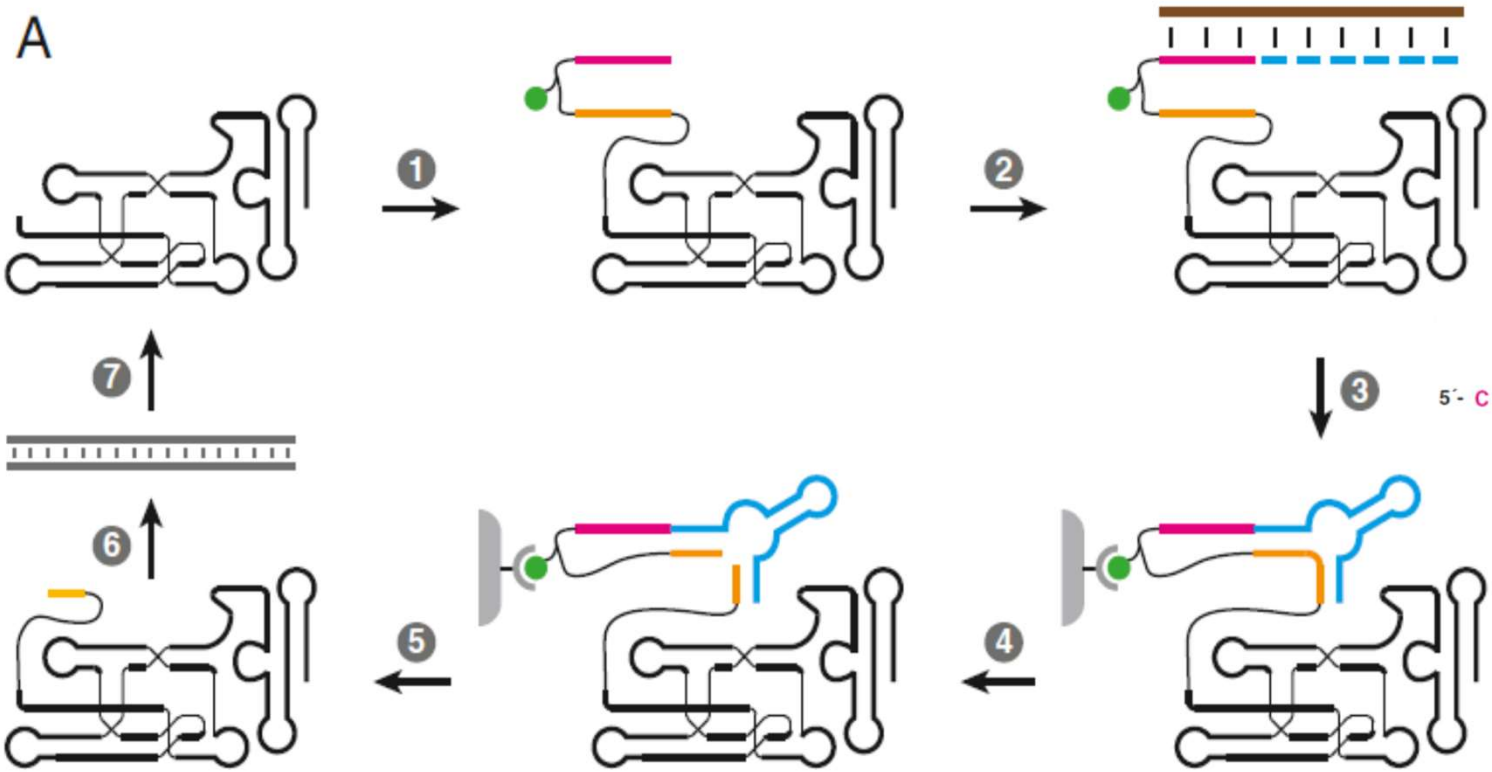
Sequence and secondary structure of the 38-6 RNA polymerase



K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce
Proc. Natl. Acad. Sci. USA (PNAS) **2020**, *117 (6)*, 2906-2913

The RNA world

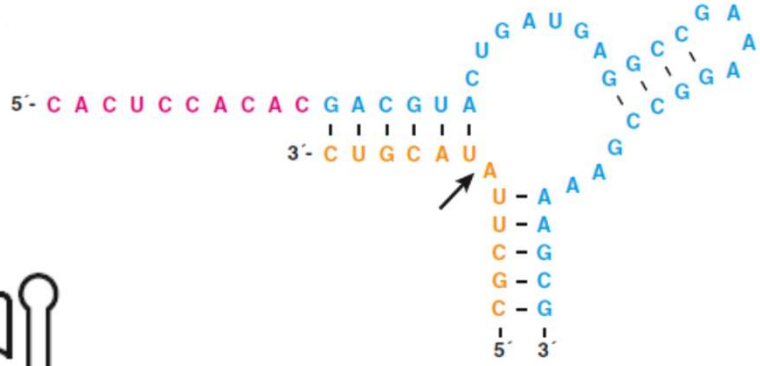
The polymerase „38-6” is able to synthesize its own evolutionary ancestor, an RNA ligase ribozyme, in the form of three fragments that assemble to give a functional complex



Selection of the polymerase „38-6”

Only full and accurate synthesis of the hammerhead ribozyme enables cleaving the yellow primer and further selection using green biotin

The hammerhead ribozyme:



K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce
Proc. Natl. Acad. Sci. USA (PNAS) **2020**, *117 (6)*, 2906-2913

DNAzymes

Nucleic acid molecules more limited in their catalytic ability in comparison to protein enzymes.

just three types of interactions: **hydrogen bonding**, **pi stacking**, and **metal-ion coordination**.

Reason: limited number of functional groups - nucleic acids are built from just **four chemically similar nucleobases** (proteins are built from up to **twenty** different amino acids with various functional groups)

In addition, DNA lacks the 2'-hydroxyl group found in RNA which limits the catalytic competency of deoxyribozymes even in comparison to ribozymes.

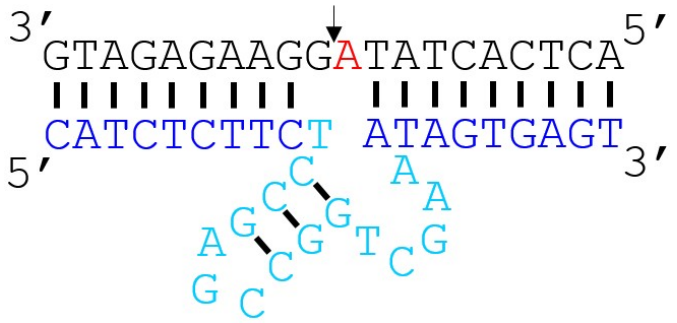
The apparent lack of naturally occurring deoxyribozymes may also be due to the primarily double-stranded conformation of DNA in biological systems → limited physical flexibility and ability to form tertiary structures (catalytic 3D-folds)

DNAzymes

Deoxyribozymes, also called ***DNA enzymes***, or catalytic DNA: DNA oligonucleotides that are capable of performing a specific chemical reaction, often but not always catalytic.

Although the working principle is similar to ***enzymes*** (and ***ribozymes***), there are no known naturally occurring ***deoxyribozymes***.

Deoxyribozymes should not be confused with ***DNA aptamers*** which are oligonucleotides that selectively bind a target ligand, but do not catalyze a subsequent chemical reaction.



The trans-form (two separate strands) of the 17E DNAzyme. Most ***ribonuclease DNAzymes*** have a similar form, consisting of a separate enzyme strand (blue/cyan) and substrate strand (black: all-RNA or a DNA with one RNA nucleotide). Two arms of complementary bases flank the catalytic core (cyan) on the enzyme strand and the single ribonucleotide (red) on the substrate strand. The arrow shows the ribonucleotide cleavage site.

1994 – the first DNAzyme (a ribonuclease) – R. Breaker, G. Joyce – Pb^{2+} GR-5

Currently known:

- Ribonucleases
- RNA ligases
- DNA phosphorylation, adenylation, deglycosylation
- DNA cleavage

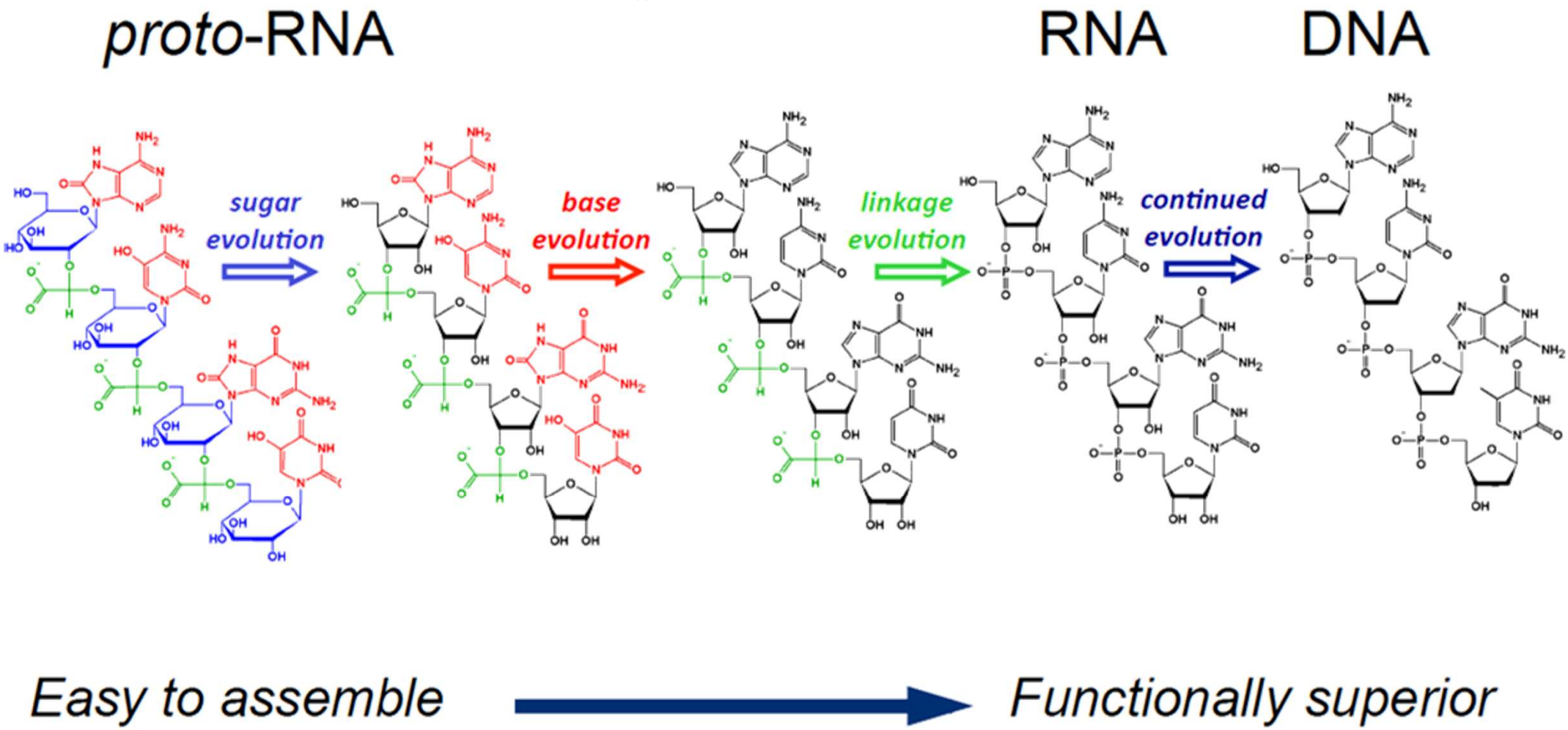
Problems: product inhibition, often single-turnover

The XNA world

Synthetic biology:

Can other genetic polymers act as catalysts?
Can they evolve and replicate themselves?

„RNA-second”



Proto-RNA evolution: According to the protoRNA theory, each of the components of RNA — sugar, base and phosphate backbone — may have originally taken different forms.