Synthetic life

(continuation of „The molecular origins of life" SoSe 2020)

NaturalNews.com

WiSe 2020/21

Zbigniew Pianowski

Overview of the course

artificial ribozymes and aptamers for efficient catalysis and recognition (SELEX, DNAzymes, foldamers);

unnatural base pairing – expansion of the genetic alphabet;

Artificial genetic polymers and oligonucleotide analogues (XNA);

biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

enzyme engineering – production of enzymes with unknown or unnatural properties, *ab initio* protein design, directed evolution, theozymes;

Artificial lipid vesicles as models for protocell multiplication;

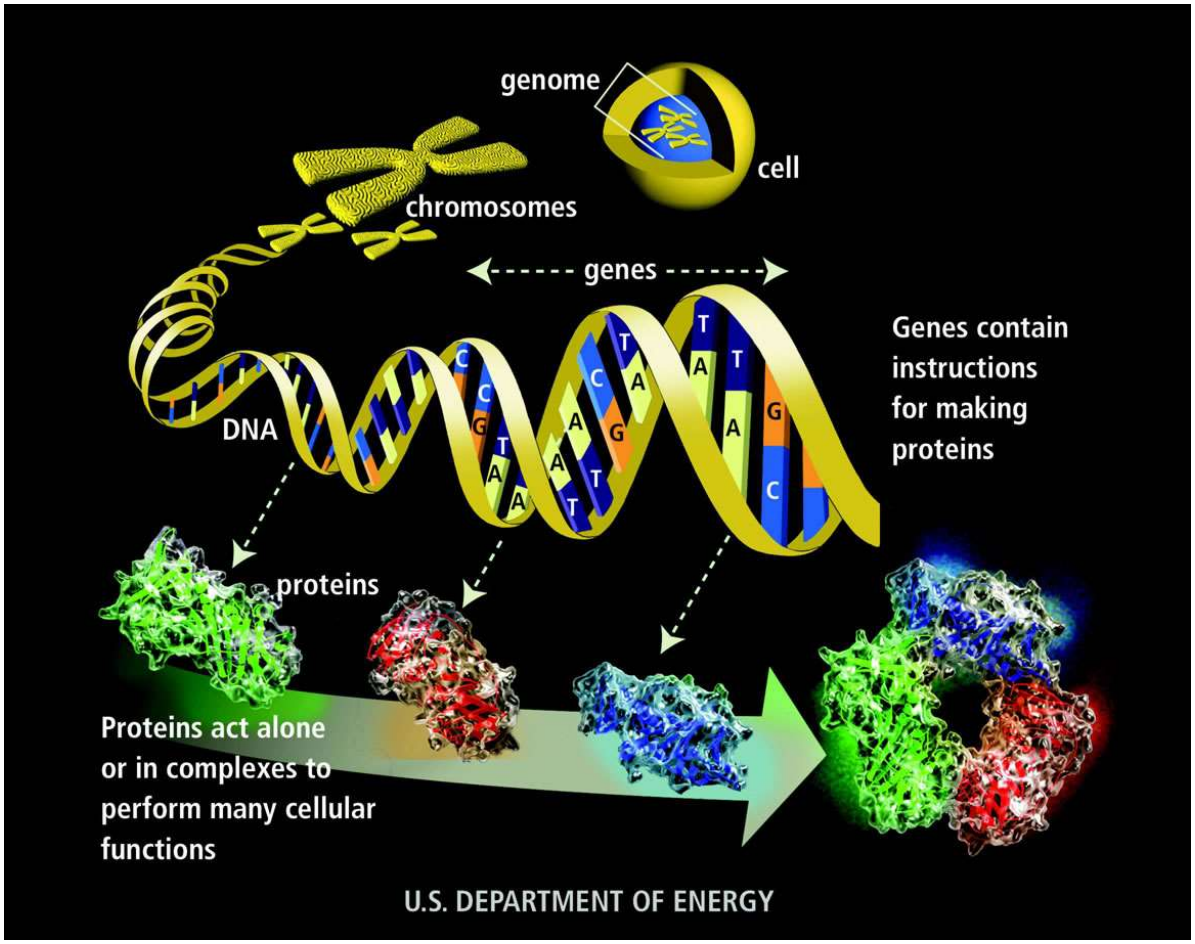
design of artificial organisms

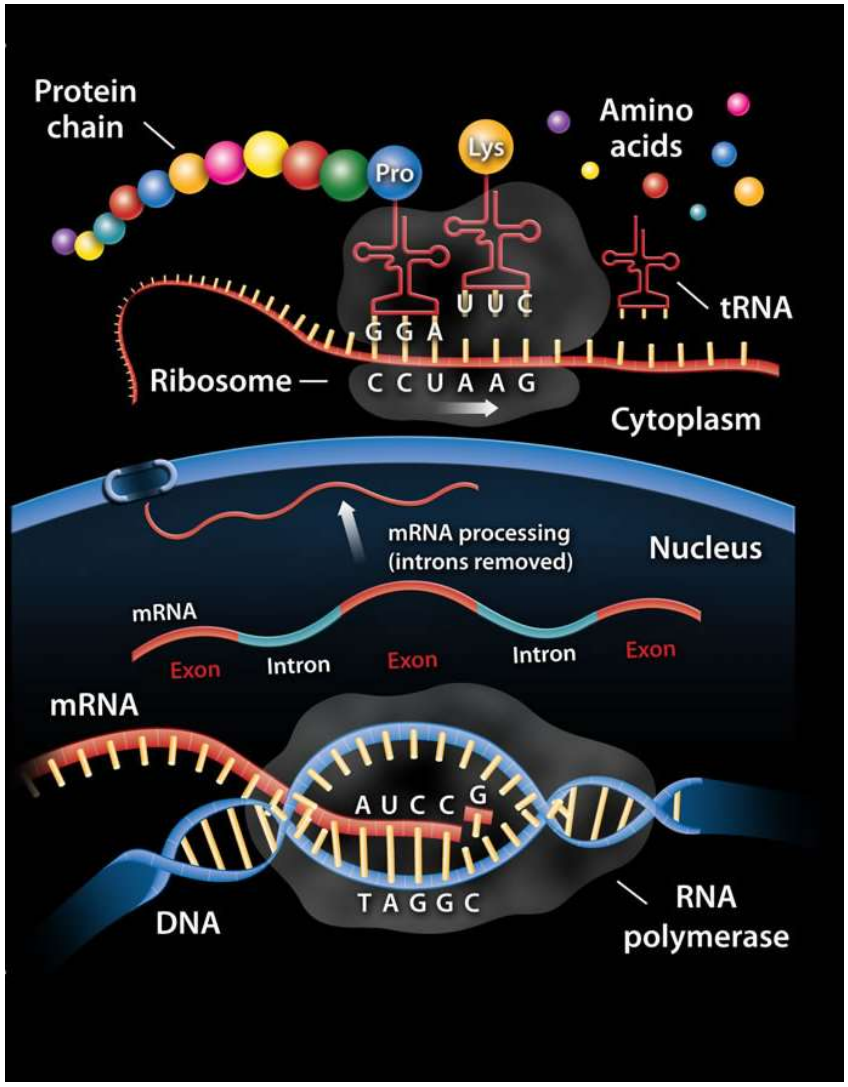
CHAPTER 1

OLIGONUCLEOTIDES

The less common side of RNA

From DNA to proteins



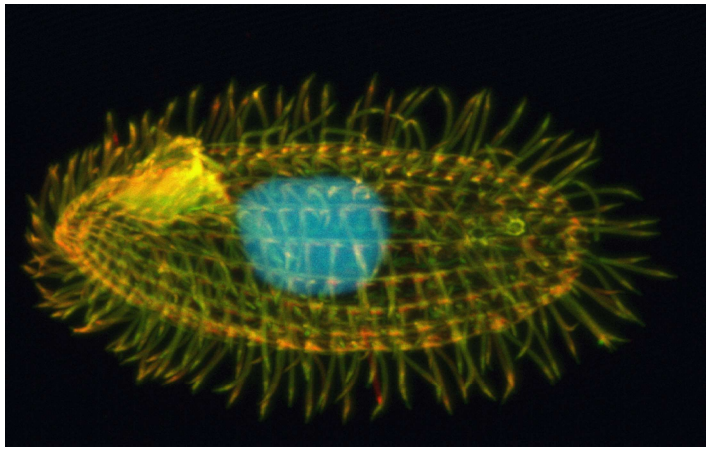


Ribozymes

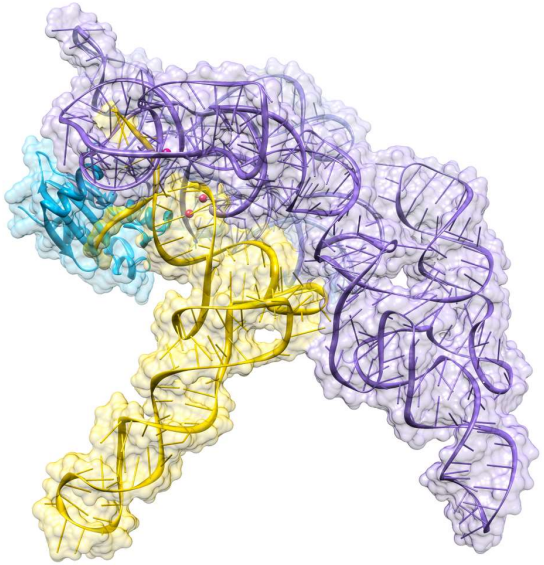
Ribozymes – Ribonucleic acid enzymes

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.

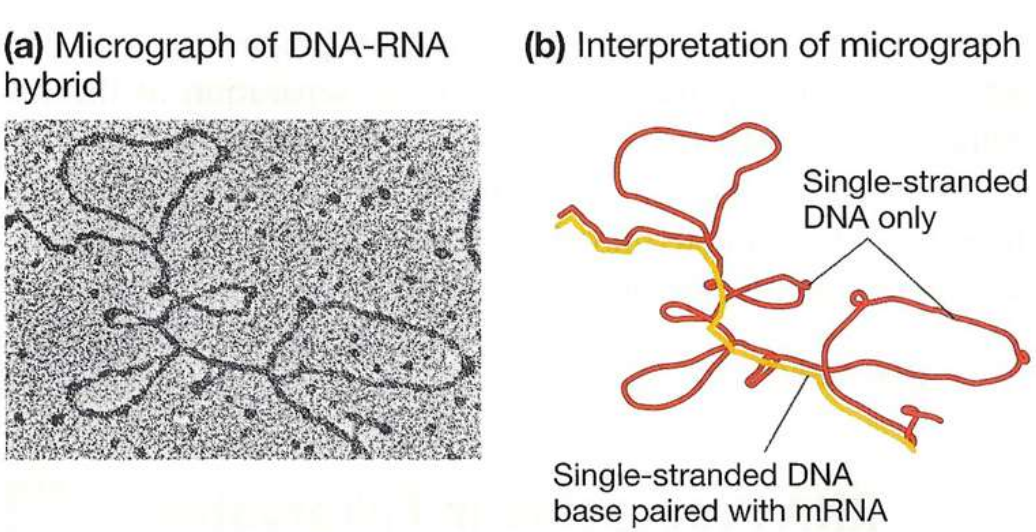


Tetrahymena thermophila

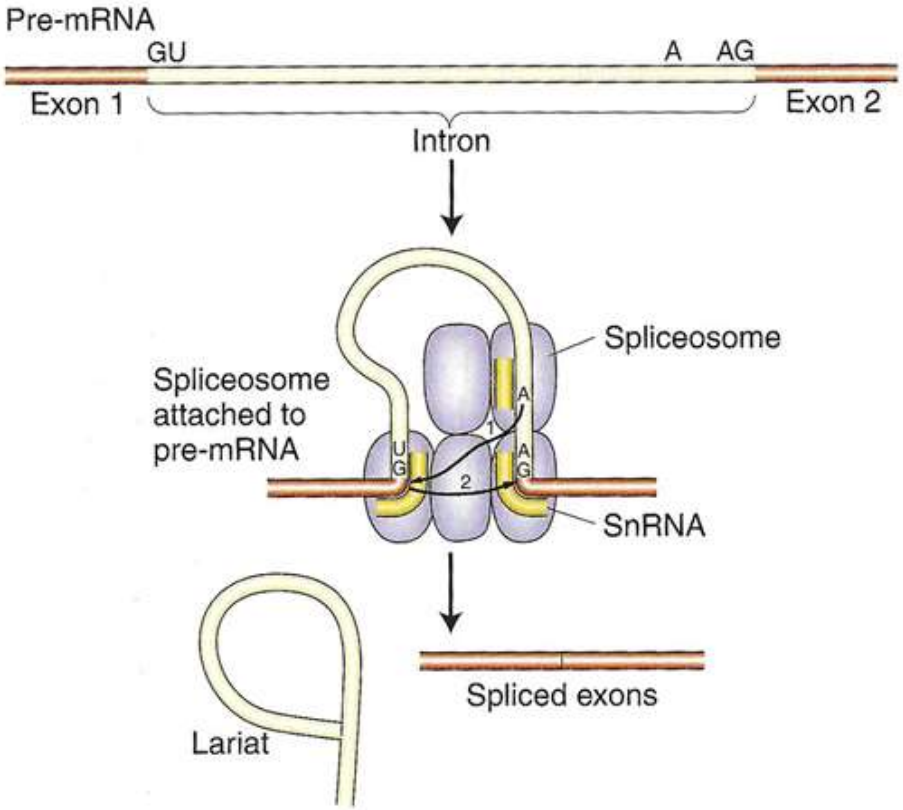


Bacterial RNAse P

mRNA processing

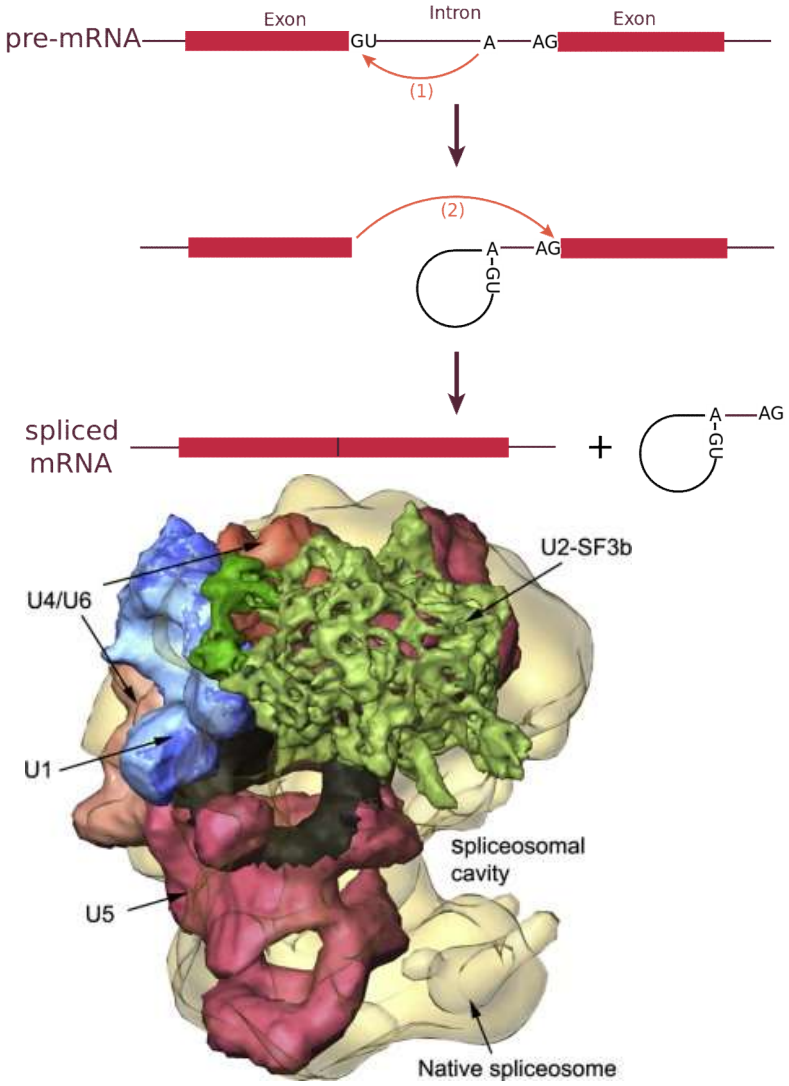


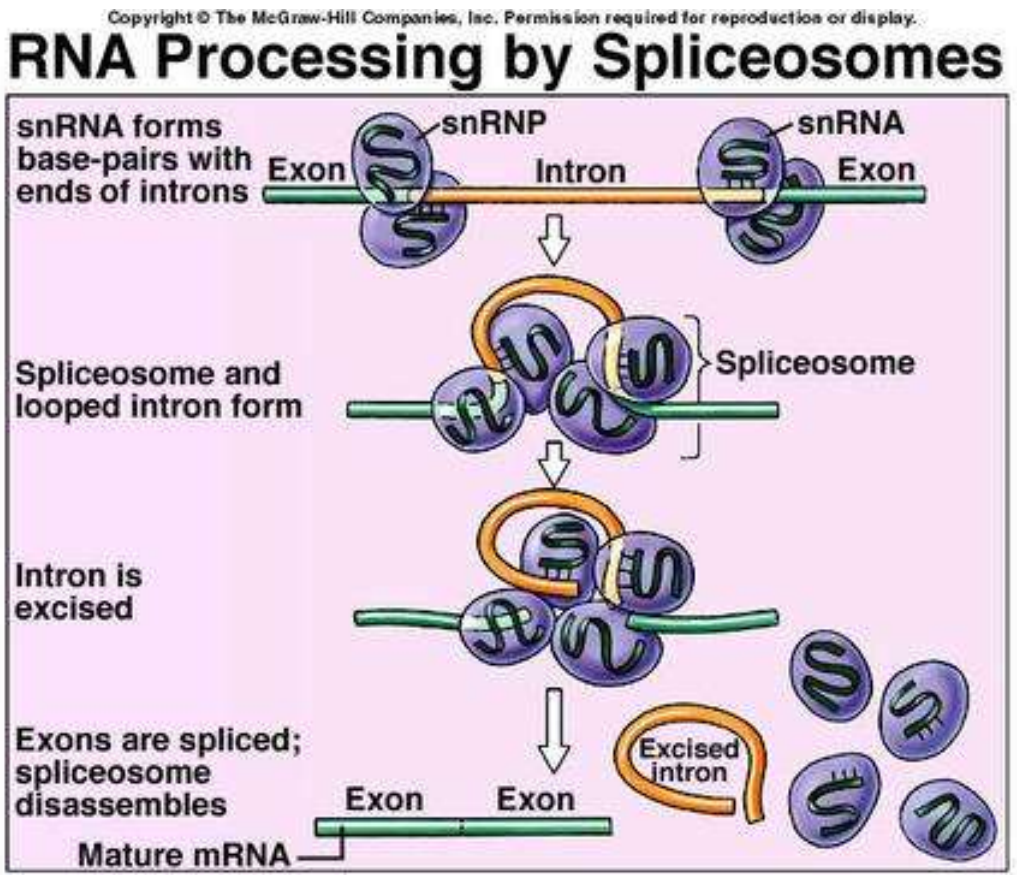
In 1977, Phil Sharp (Nobel Prize 1993) hybridized an mRNA to its DNA template and prepared the hybrid molecule for electron microscopy by coating the nucleic acid with a basic protein, then using rotary shadowing to coat the nucleic acid-protein complex.



The spliceosome carries out the removal of introns as RNA lariats

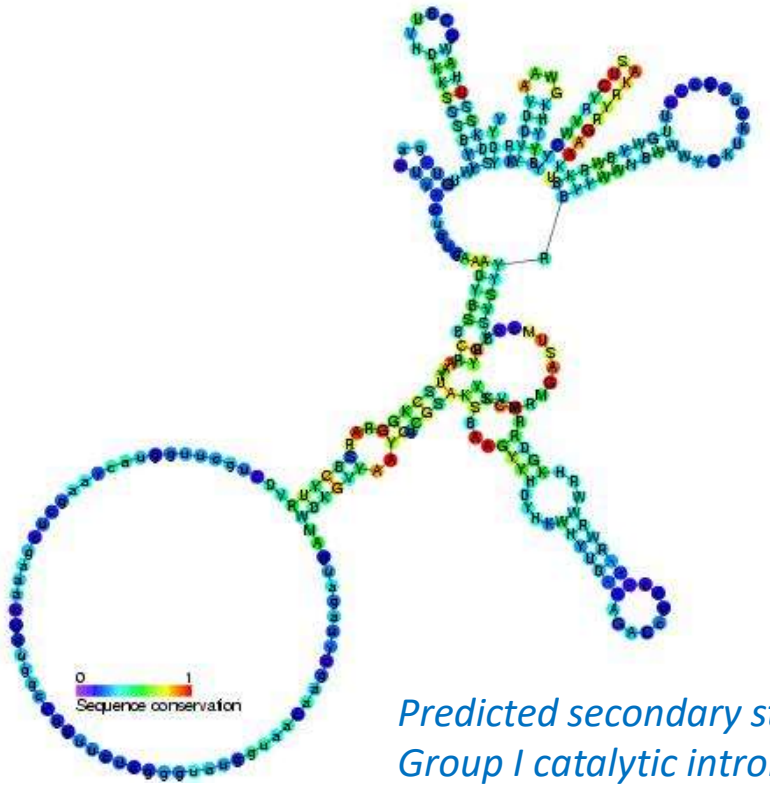
RNA splicing





Spliceosome – a complex of ribonucleoproteins

RNA splicing

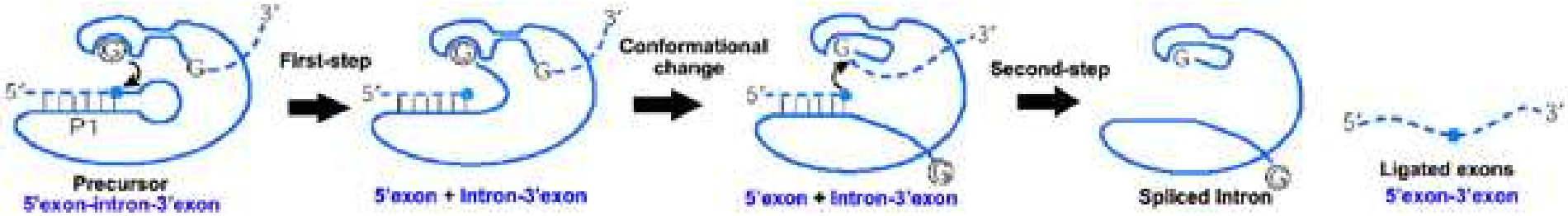


Predicted secondary structure and sequence conservation of Group I catalytic intron

Self-splicing RNA introns

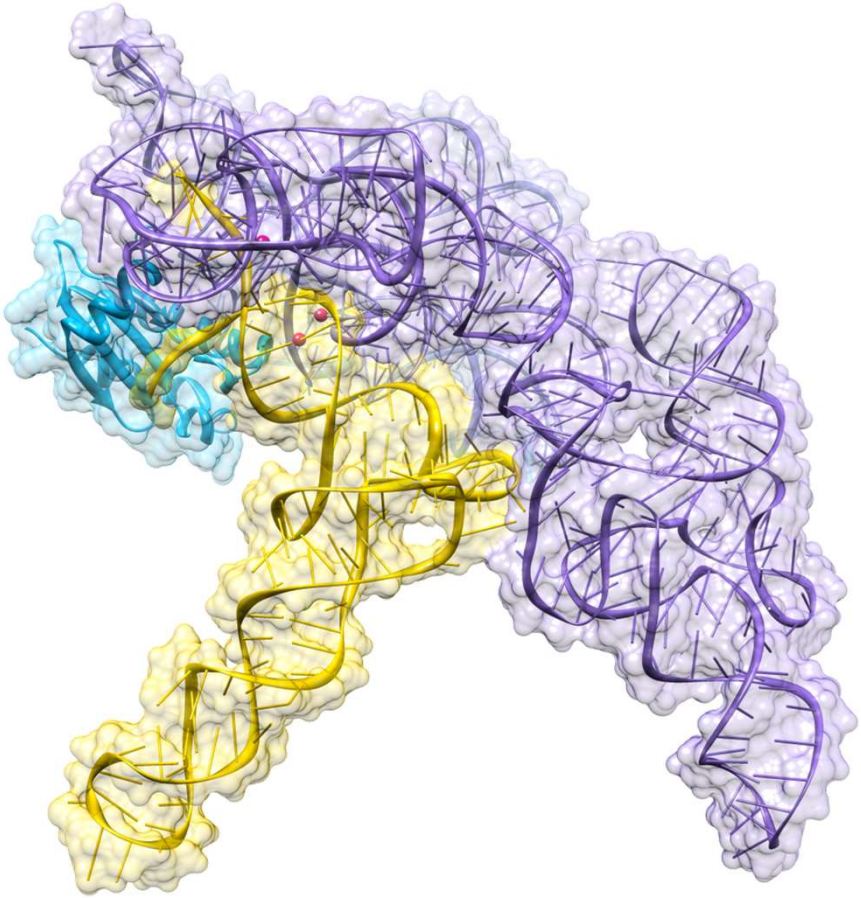
RNA splicing in *Tetrahymena* was taking place also in absence of the spliceosome - the ‚negative control' obtained after protease digestion also spliced.

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

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.

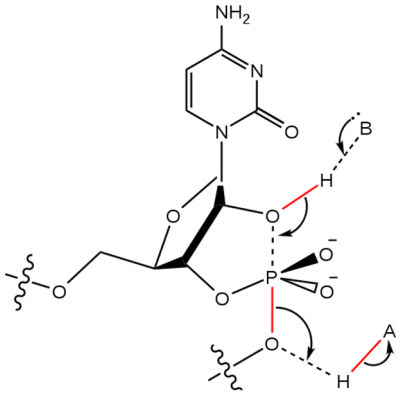
Bacterial RNase P has two components: an RNA chain, called M1 RNA, and a polypeptide chain, or protein, called C5 protein. *In vivo*, both components are necessary for the ribozyme to function properly, but *in vitro*, the M1 RNA can act alone as a catalyst. The primary role of the C5 protein is to enhance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

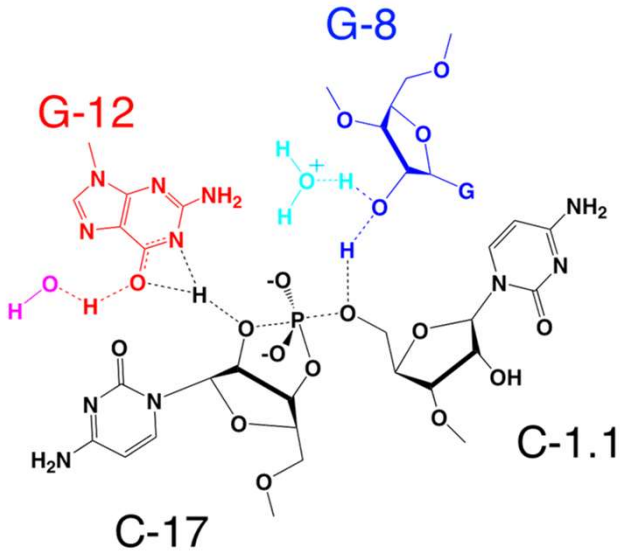
Crystal structure of a bacterial ribonuclease P holoenzyme in complex with tRNA (yellow), showing metal ions involved in catalysis (pink)

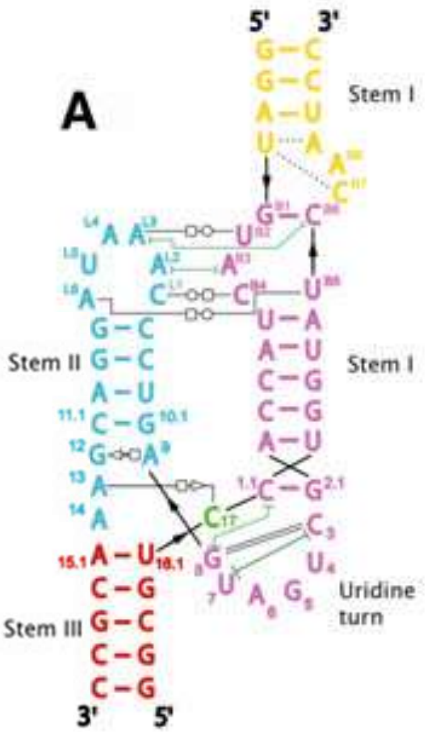
Ribozymes

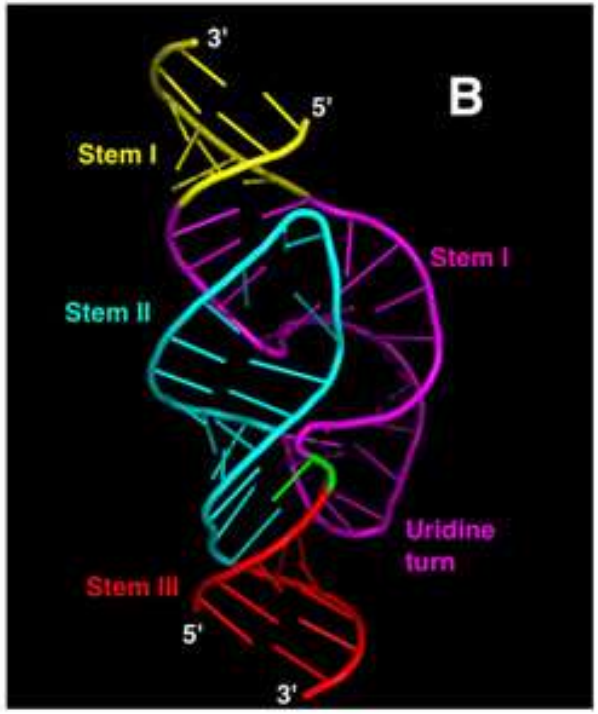
Hammerhead ribozyme

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)









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

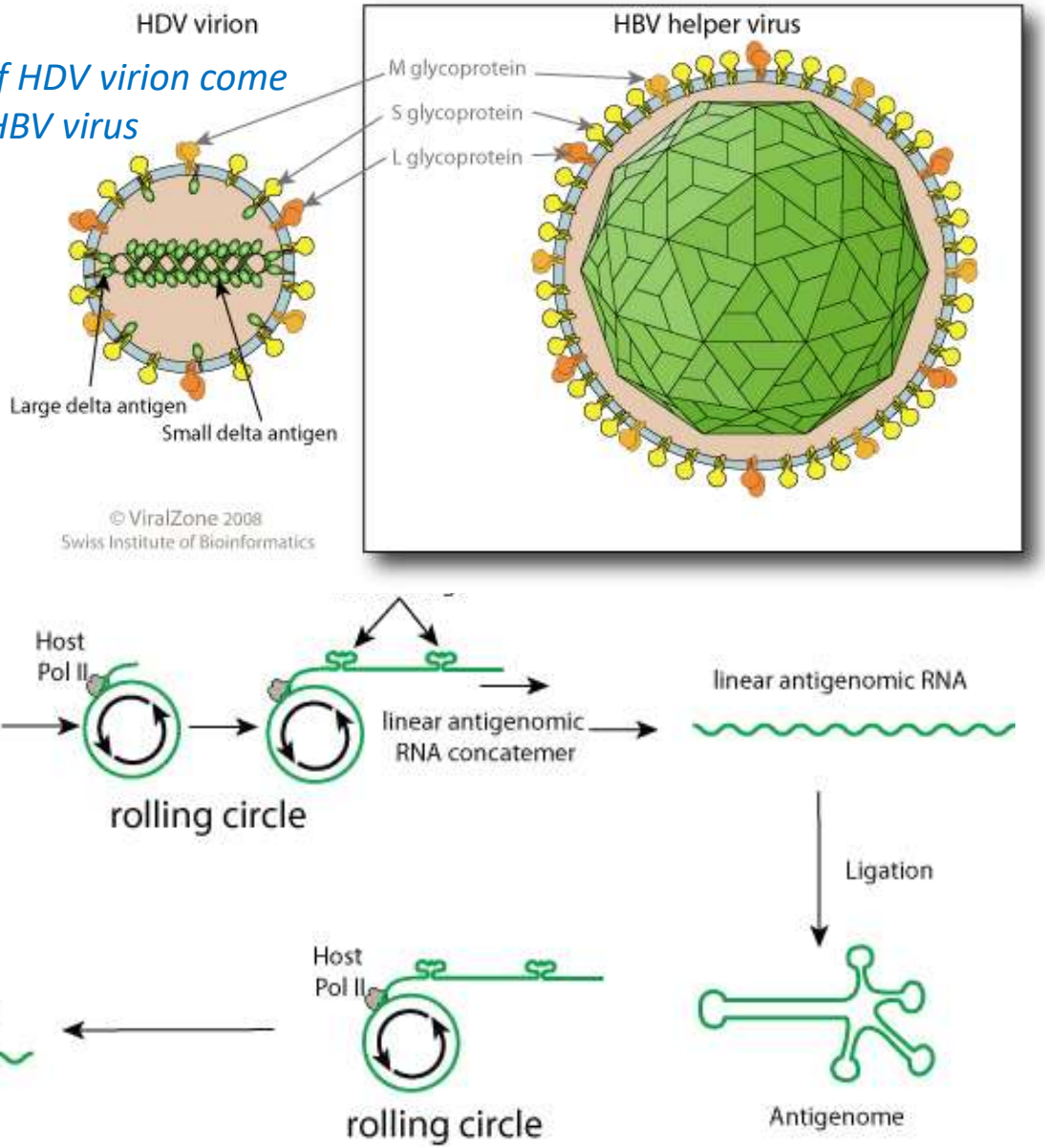
Ribozymes

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.

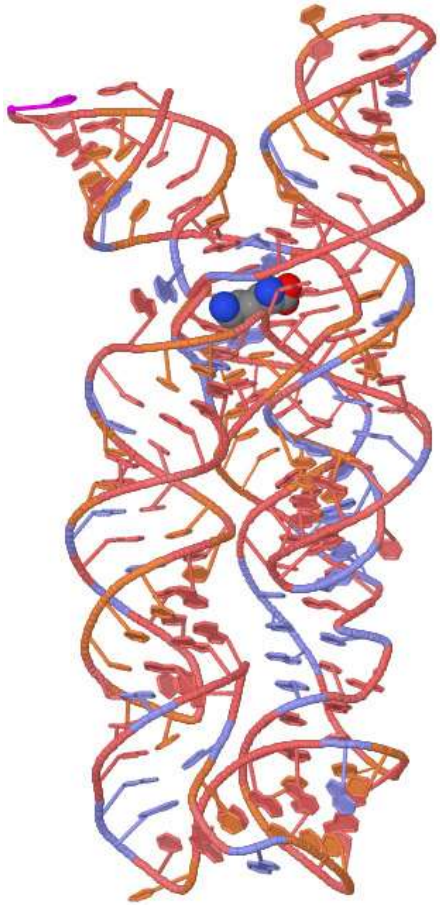
The ribozyme acts to process the RNA transcripts to unit lengths in a self-cleavage reaction. The ribozyme is found to be active in vivo in the absence of any protein factors and is the fastest known naturally occurring self-cleaving RNA.

Proteins of HDV virion come from the HBV virus



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

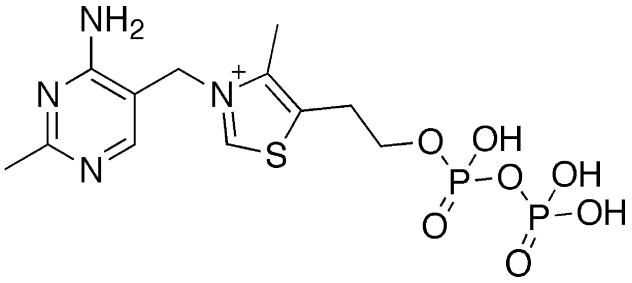
Before discovery of ***riboswitches*** only ***proteins*** were supposed to do so in the biological context.

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

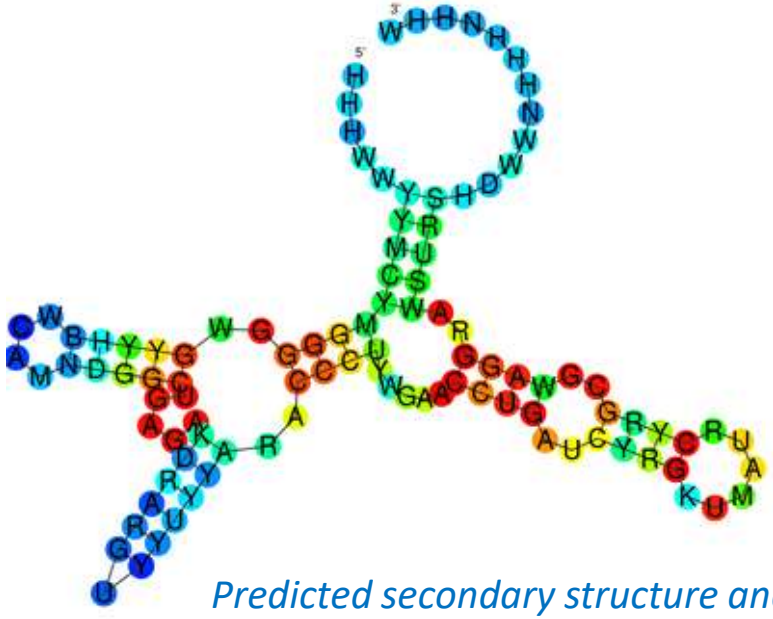
Riboswitches exist in all domains of life, and therefore are likely that they might represent ancient regulatory systems or fragments of ***RNA-world ribozymes*** whose binding domains remained conserved throughout the evolution

The TPP Riboswitch

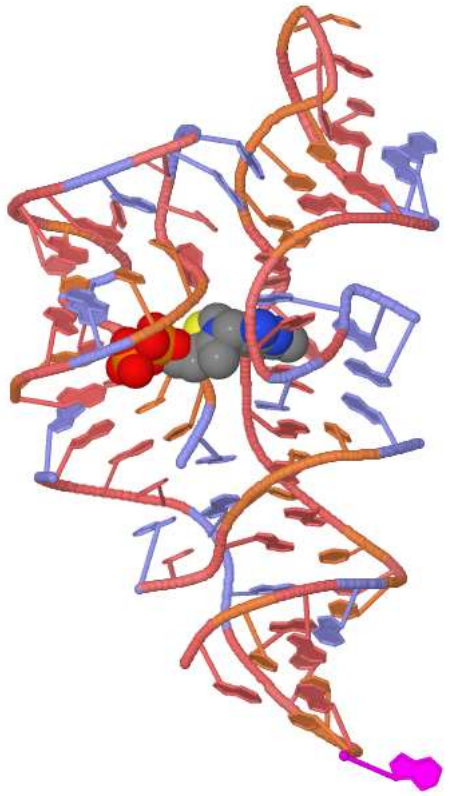
The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



Thiamine pyrophosphate TPP



Predicted secondary structure and sequence conservation of TPP riboswitch



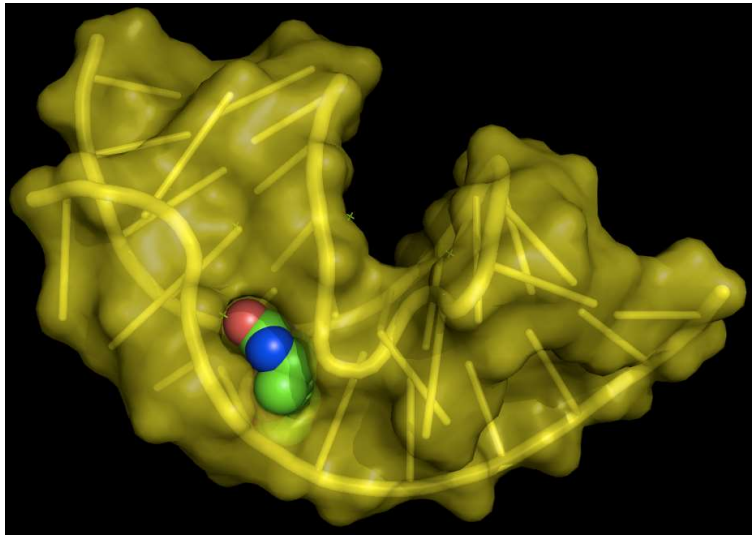
The 3D structure of TPP riboswitch (by Benjamin Schuster-Böckler)

Aptamers

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

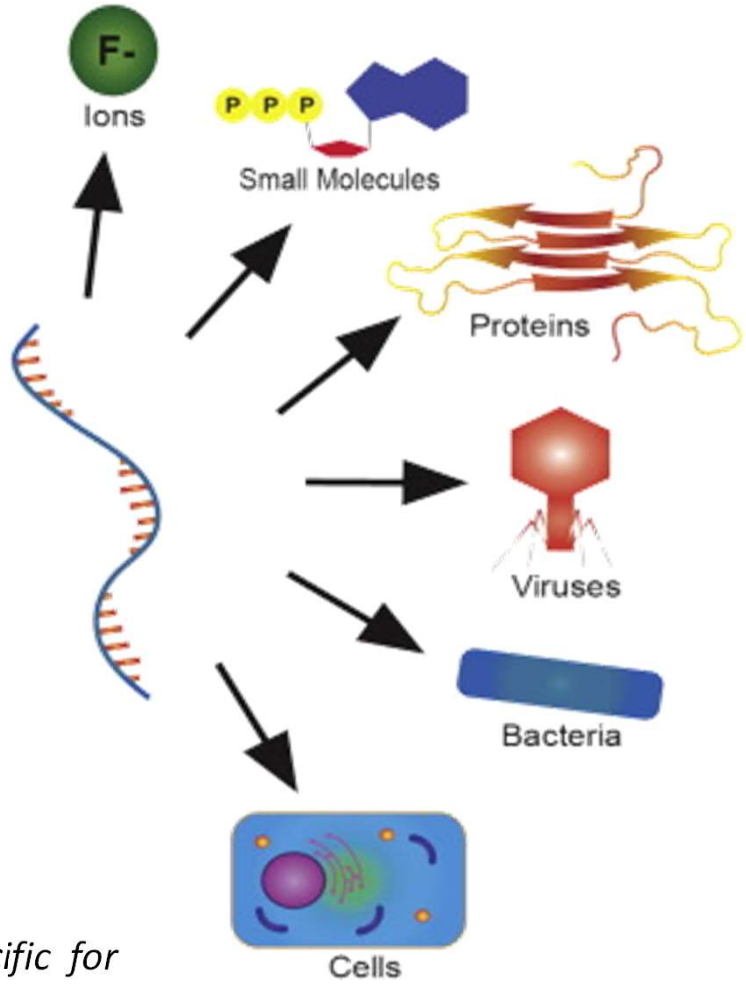
Aptamers are usually created by selecting them from a large random sequence pool, but natural aptamers also exist in riboswitches.

- ***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

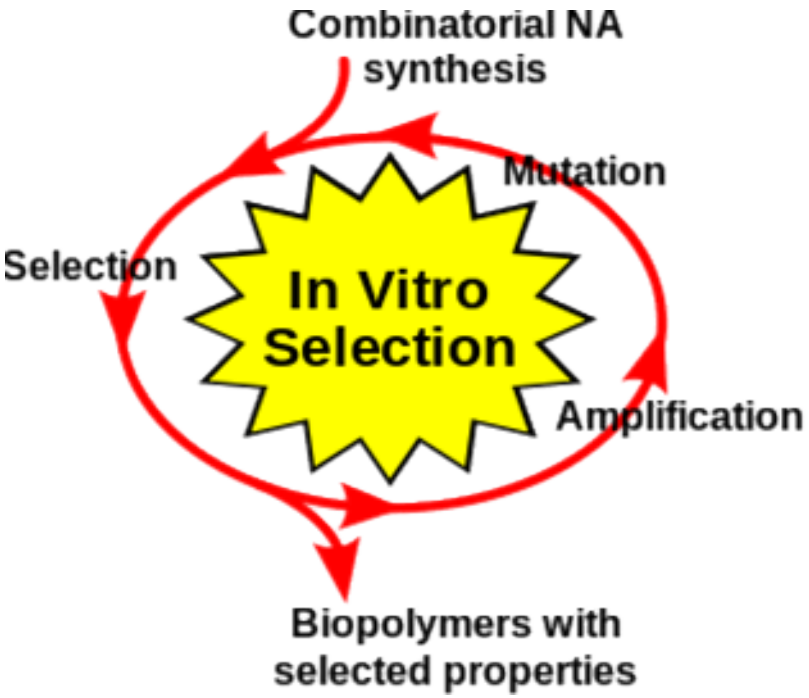
Structure of an RNA aptamer specific for biotin. The aptamer surface and backbone are shown in yellow. Biotin (spheres) fits snugly into a cavity of the RNA surface



Variety of target molecules

Systematic evolution of ligands by exponential enrichment - SELEX

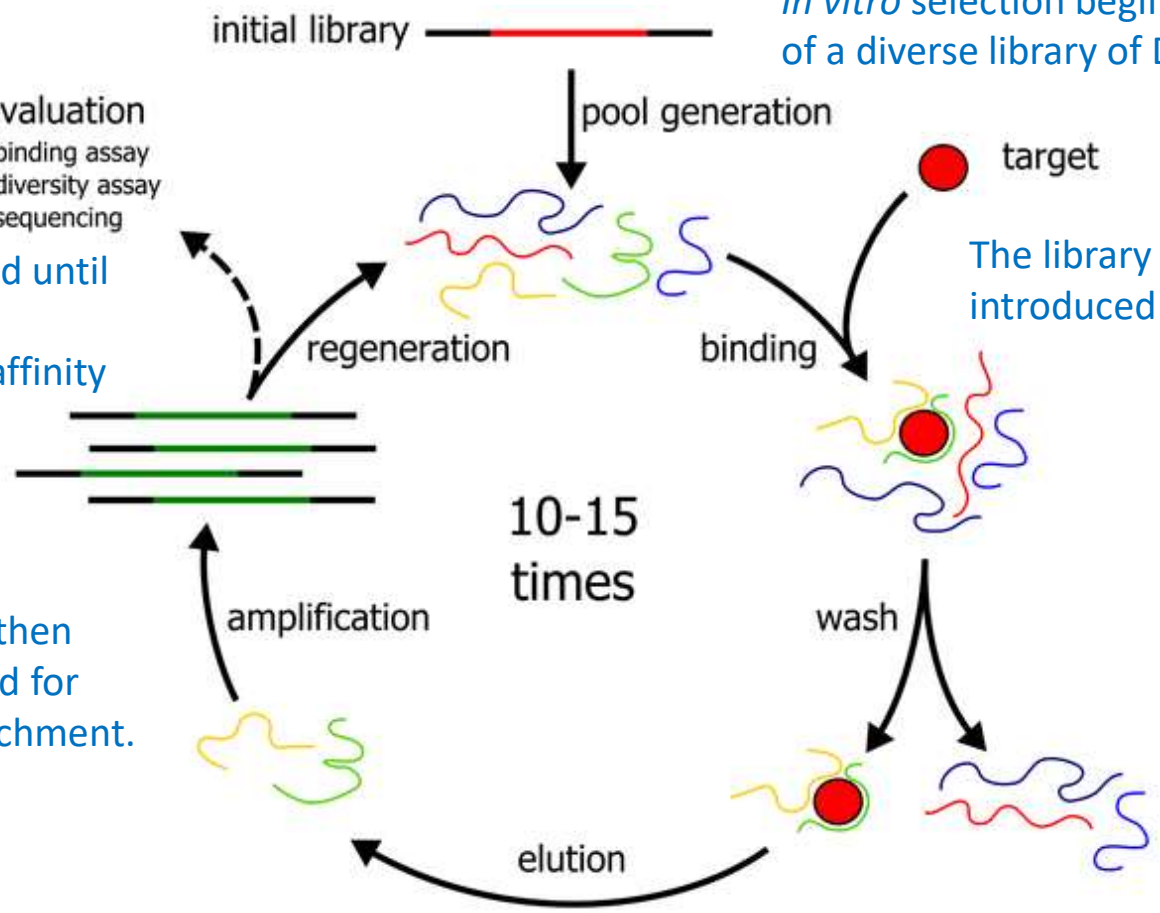
1990 – Gold *et al.* – selection of RNA ligands against T4 DNA polymerase
1990 – J. Szostak *et al.* – selecting RNA ligands towards organic dyes



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.



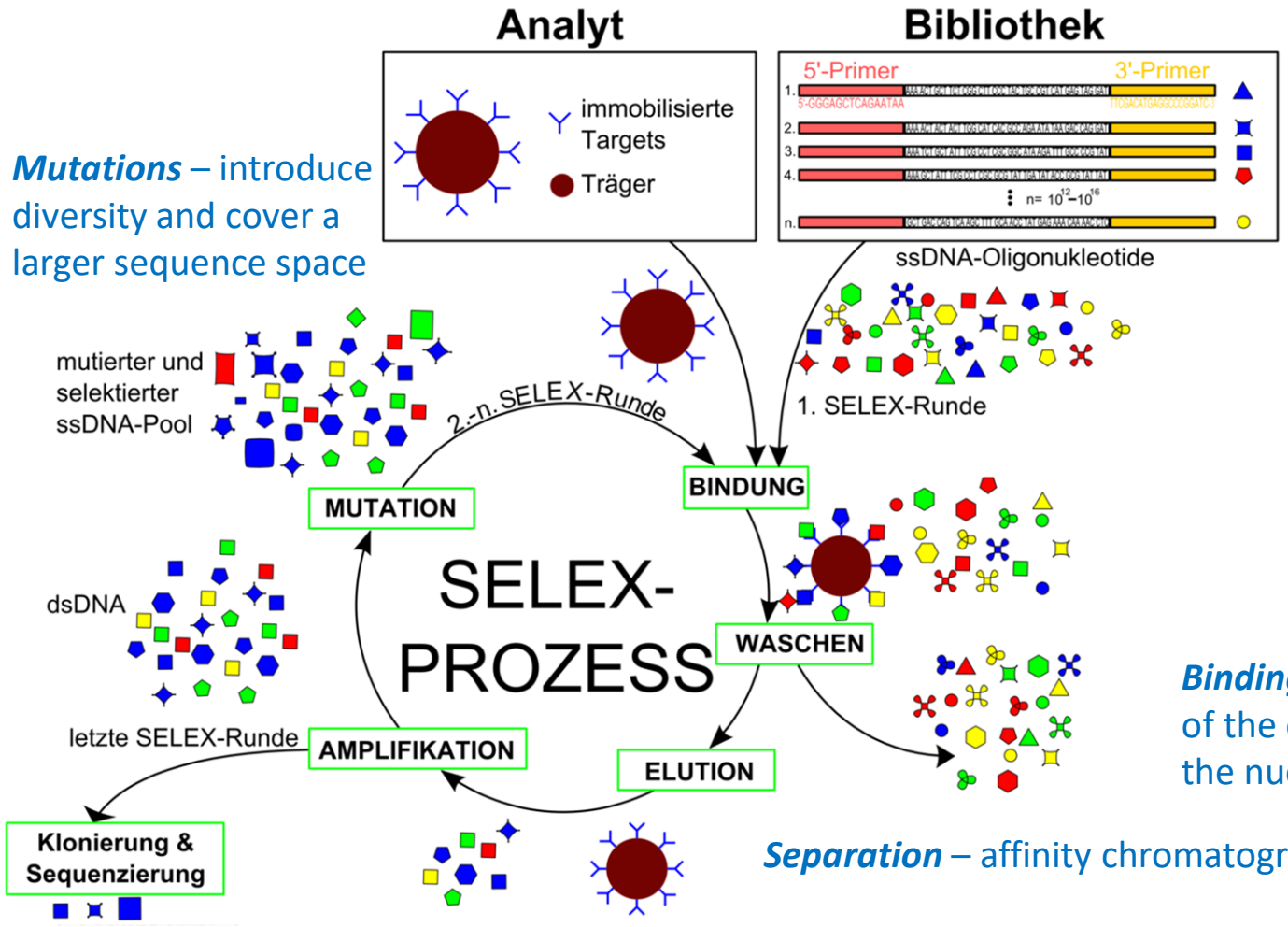
Multiple rounds are performed until the library converges on to a collection of sequences with affinity for the target molecule.

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:

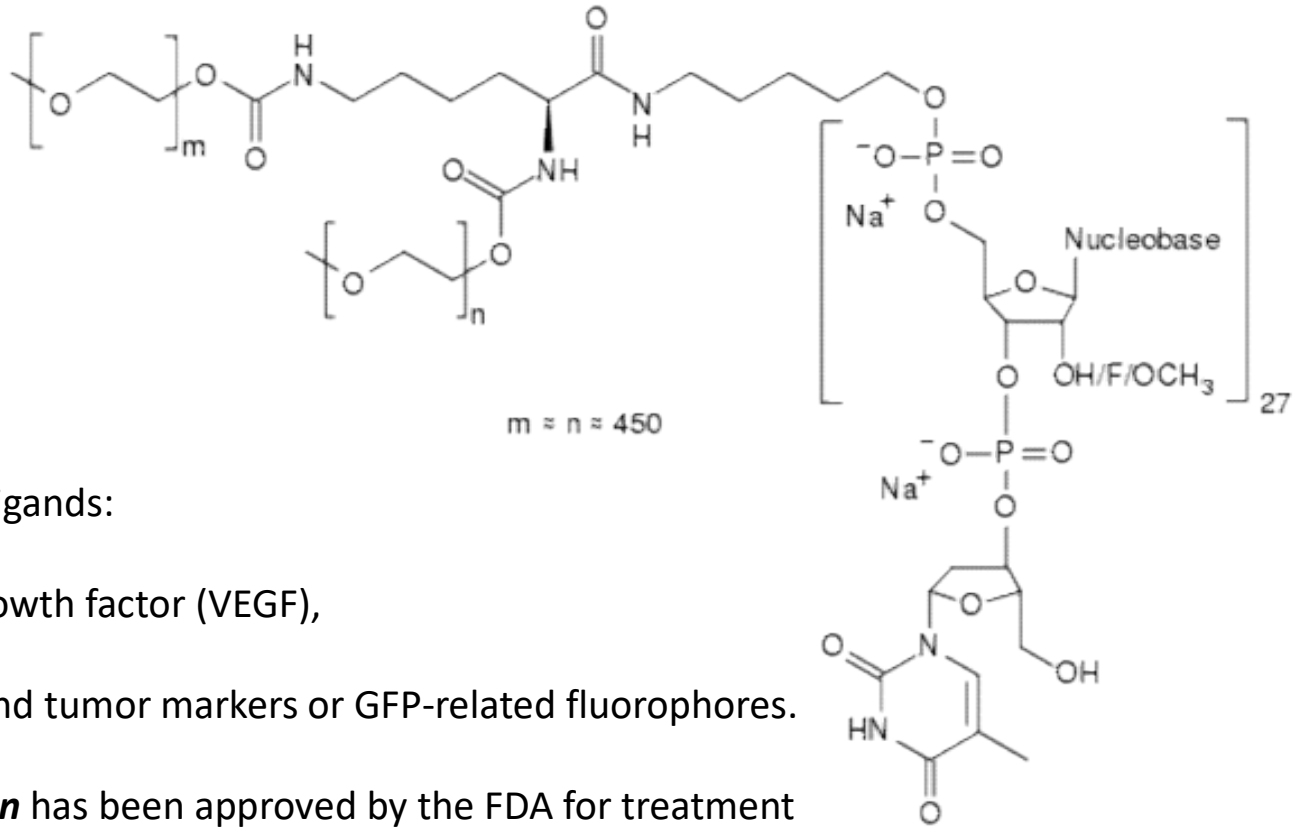
- n = 25 – a library of c.a. 10^{15} sequences,
- n = 50 – a library of c.a. 10^{30} sequences,
- n = 75 – a library of c.a. 10^{45} sequences,
- n = 100 – a library of c.a. 10^{60} sequences.

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.

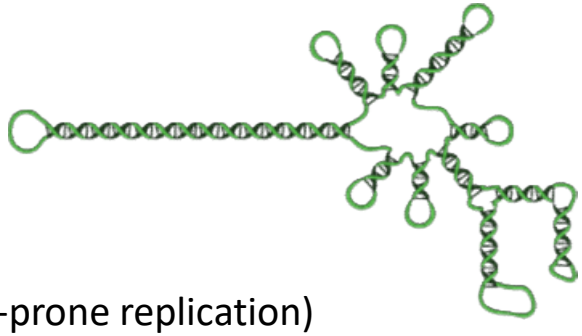
A VEGF-binding aptamer trade-named ***Macugen*** has been approved by the FDA for treatment of macular degeneration.

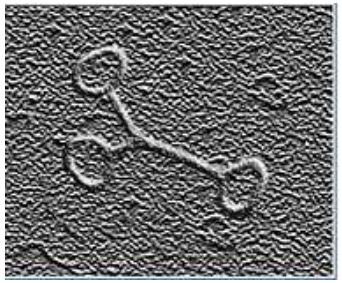
Macugen

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).

Viroids

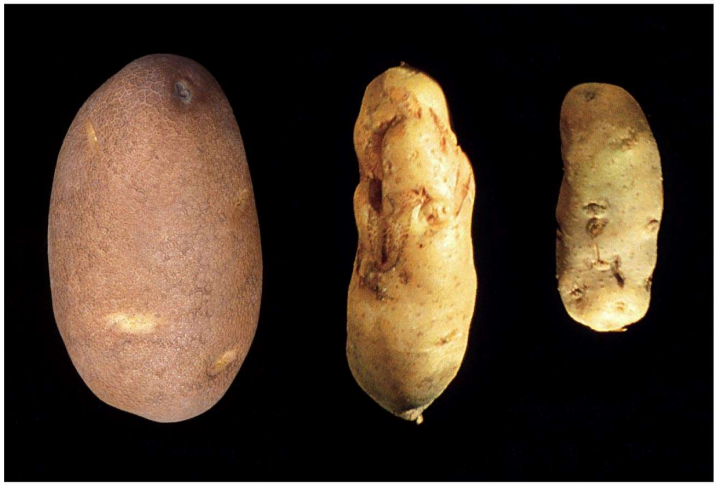
Viroids ("subviral pathogens„) are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Viroids are extremely small - 246 to 467 nucleobases (genomes of smallest viruses start from 2,000 nucleobases). Viroids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



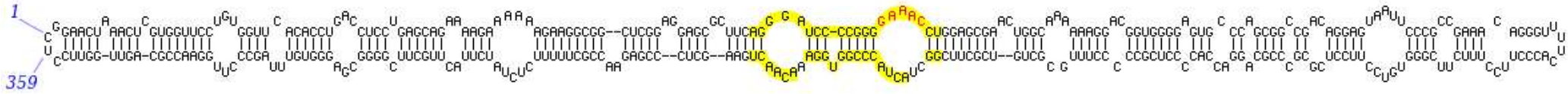


Viroid properties:

- 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.



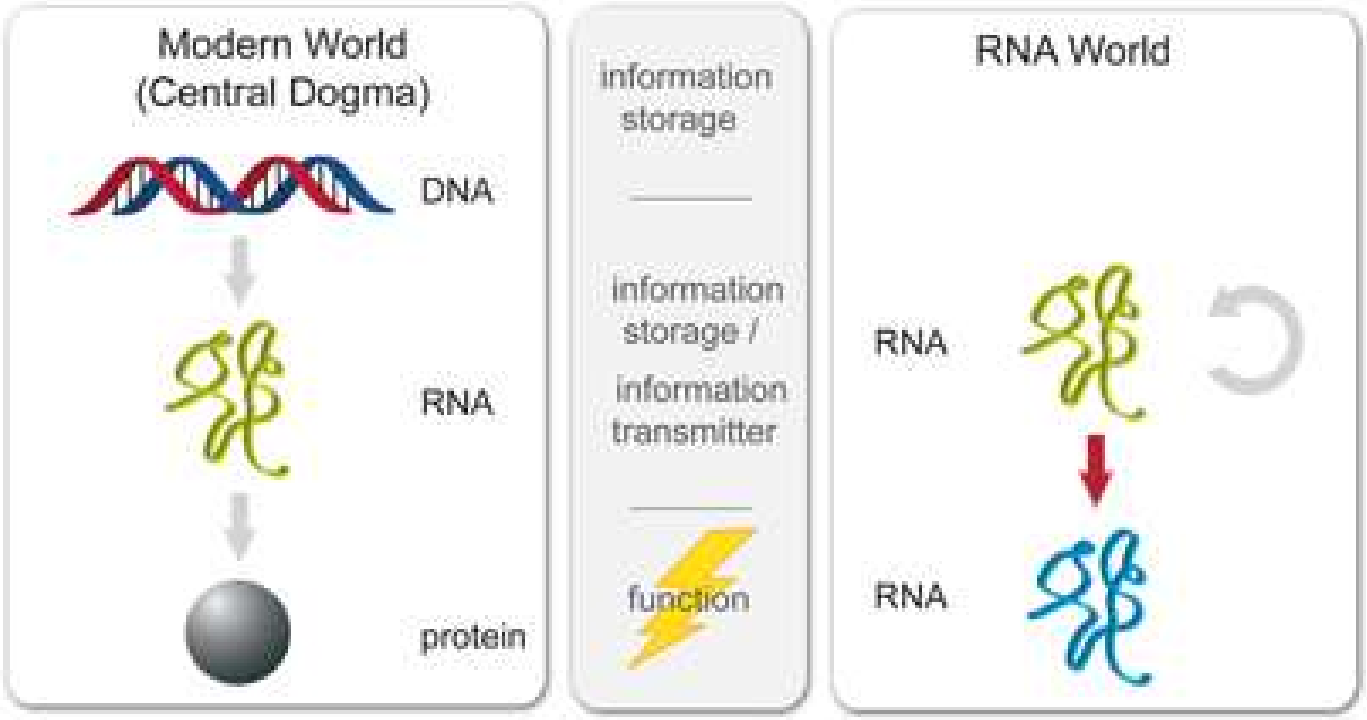
PSTVd-infected potatoes (right)



Putative secondary structure of the PSTVd viroid

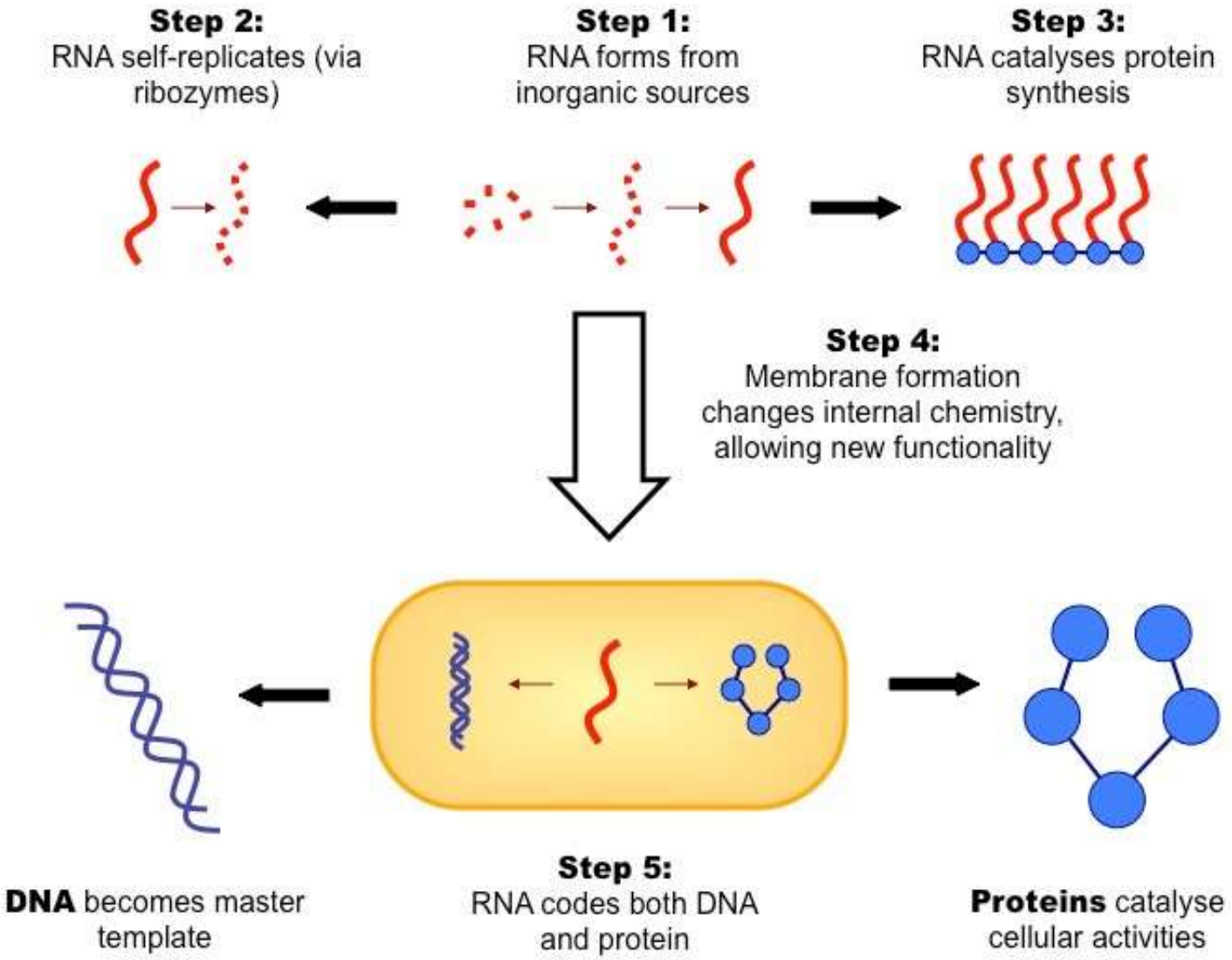
No virion reported. Viroids do not encode for proteins

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

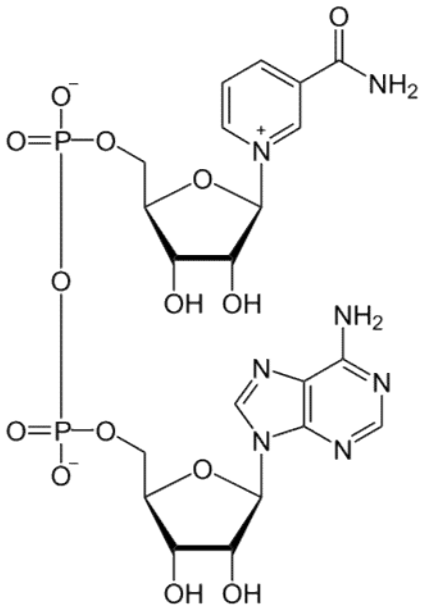


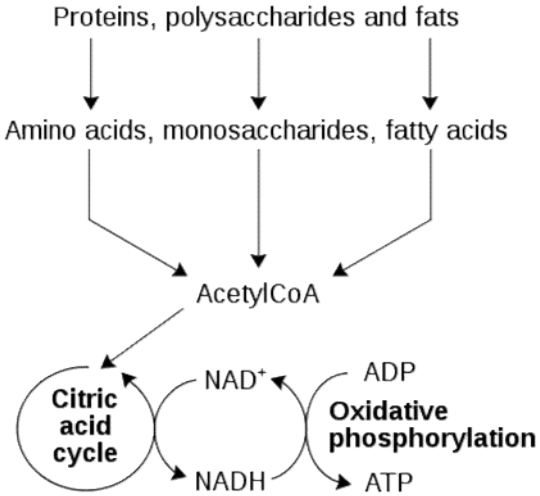
The RNA world

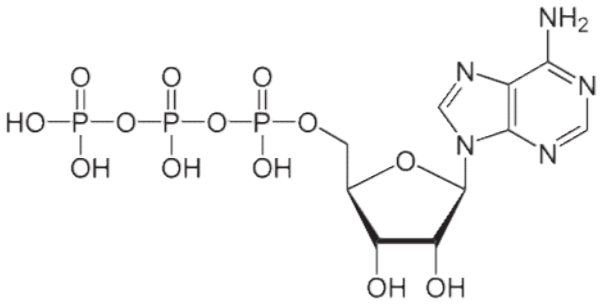
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

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







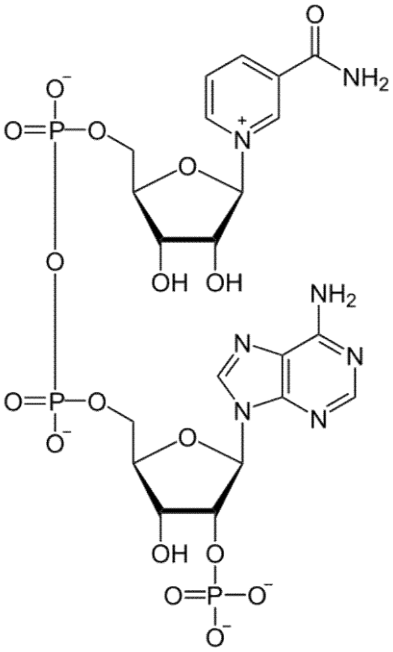
Nicotinamide adenine dinucleotide (NAD$^+$)

Adenosine triphosphate (ATP)

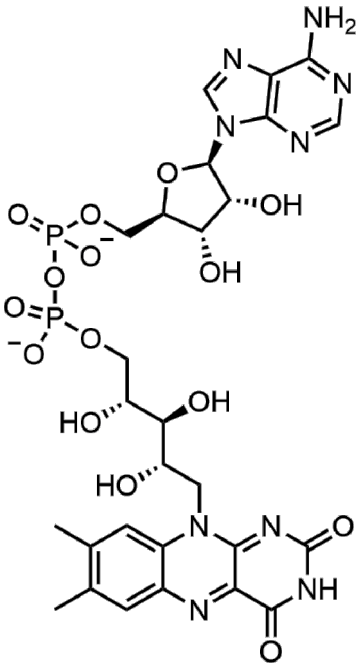
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

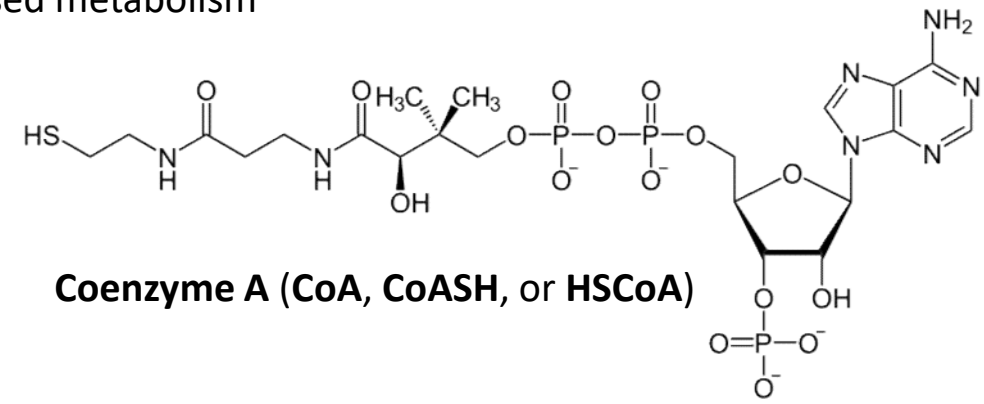
Ribonucleotide coenzymes now used by many proteins may be molecular „fossils” from the primoridal RNA-based metabolism



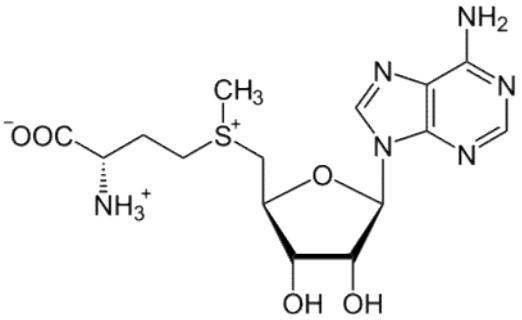
Nicotinamide adenine dinucleotide phosphate ($NADP^+$)



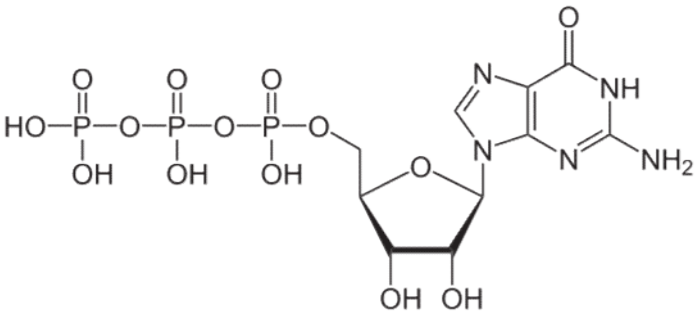
flavin adenine dinucleotide (FAD)



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



***S*-Adenosyl methionine**

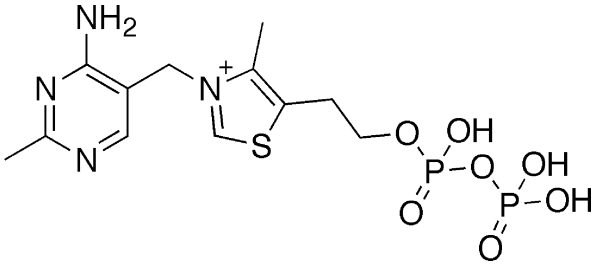


Guanosine-5'-triphosphate (**GTP**)

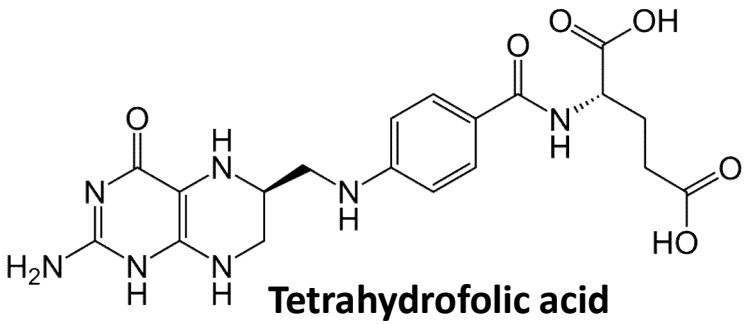
H. B. White III *J. Mol. Evol.* **1976**, *7*, 101-104

The RNA world

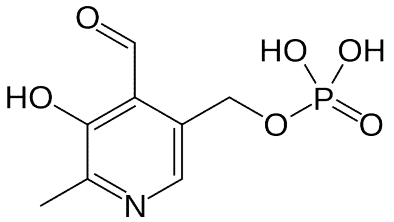
Other coenzymes contain cyclic nitrogen-containing bases that can also derive from nucleotides



Thiamine pyrophosphate (TPP or ThPP) – Vit. B_1



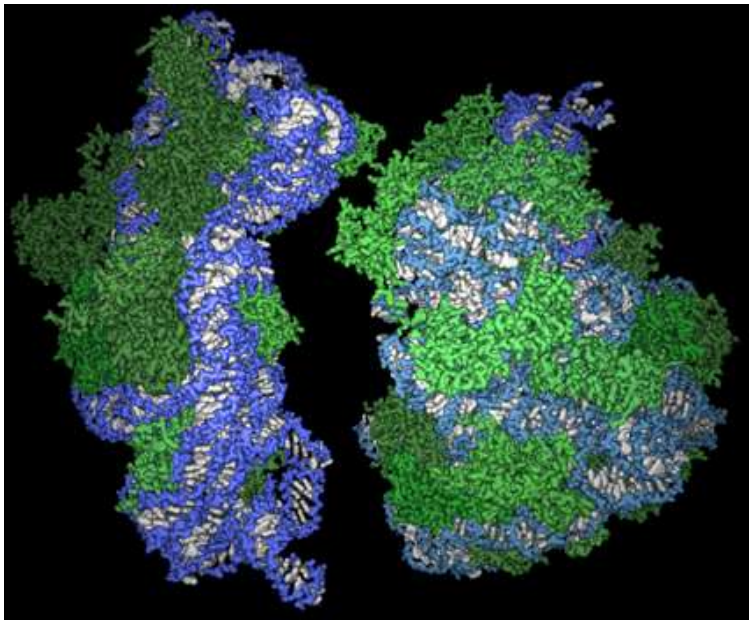
Tetrahydrofolic acid



Pyridoxal phosphate (PLP) – Vit. B_6

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

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

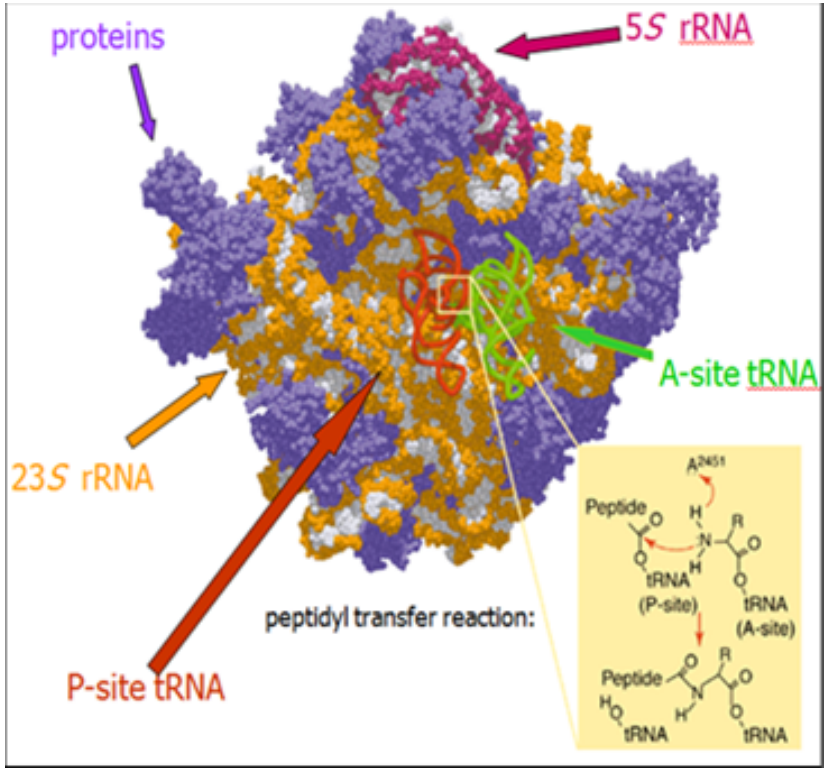
Ribosomes:

- 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!

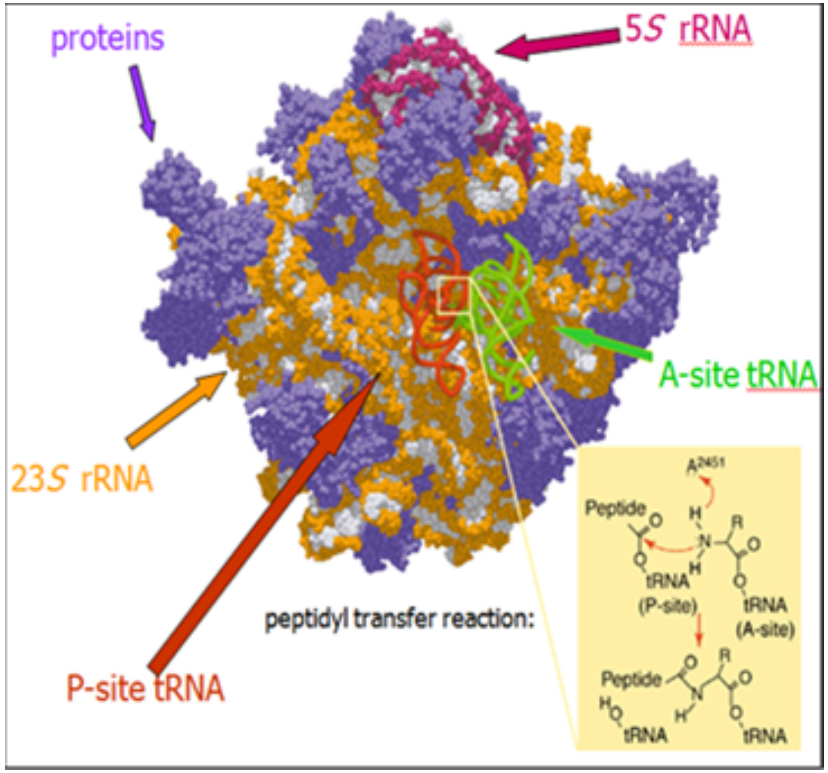


No protein is present within 18 Angstroms from the active site → proteins play a structural role, but DO NOT CATALYZE THE ACYL TRANSFER PROCESS

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.

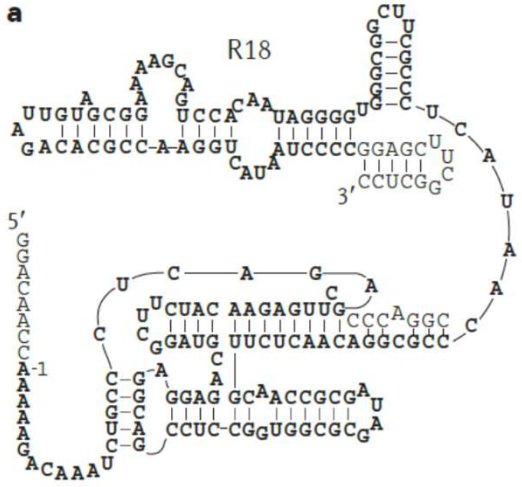
Early proto-ribosomes were self-replicating complexes:
the rRNA had informational, structural, and catalytic purposes –
it coded for tRNAs and proteins needed for ribosomal self-replication.

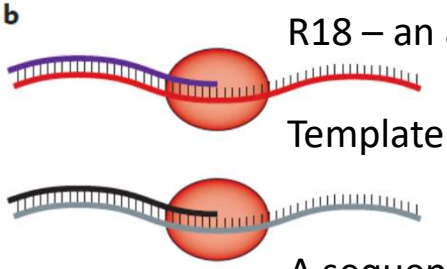
Emerging amino acids interacted with catalytic RNA:
increased scope and efficiency of catalytic RNA molecules.

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

The RNA world

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

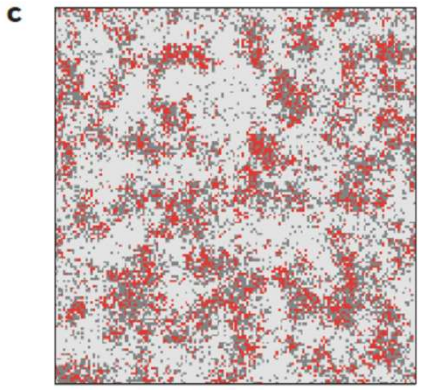


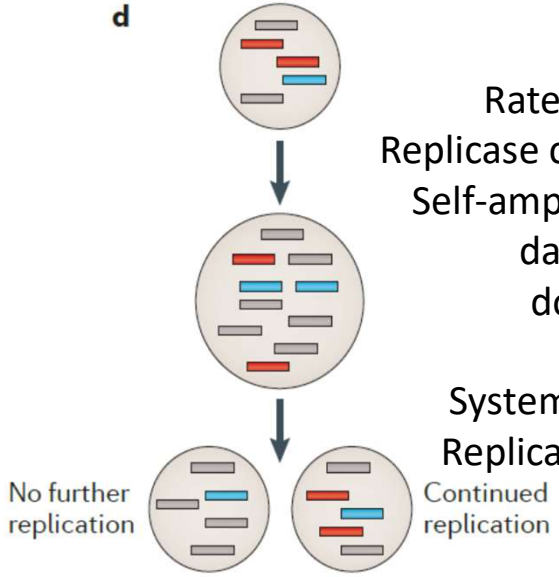


R18 – an artificial polymerase evolved from the class I ligase ribozyme.

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.

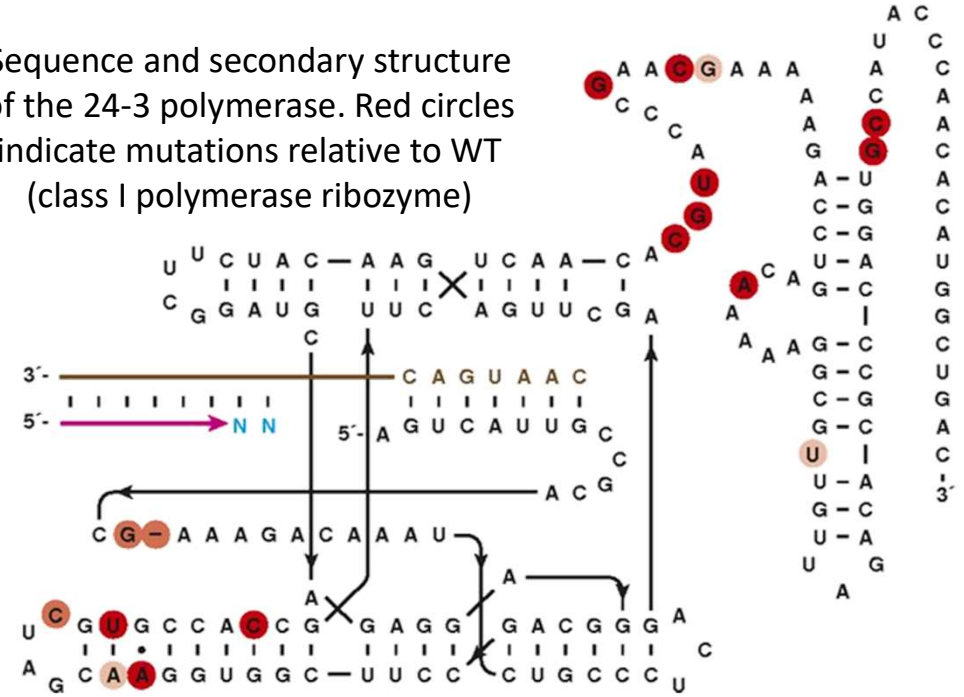
Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **2013**, *5*, 1011–1018.

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)

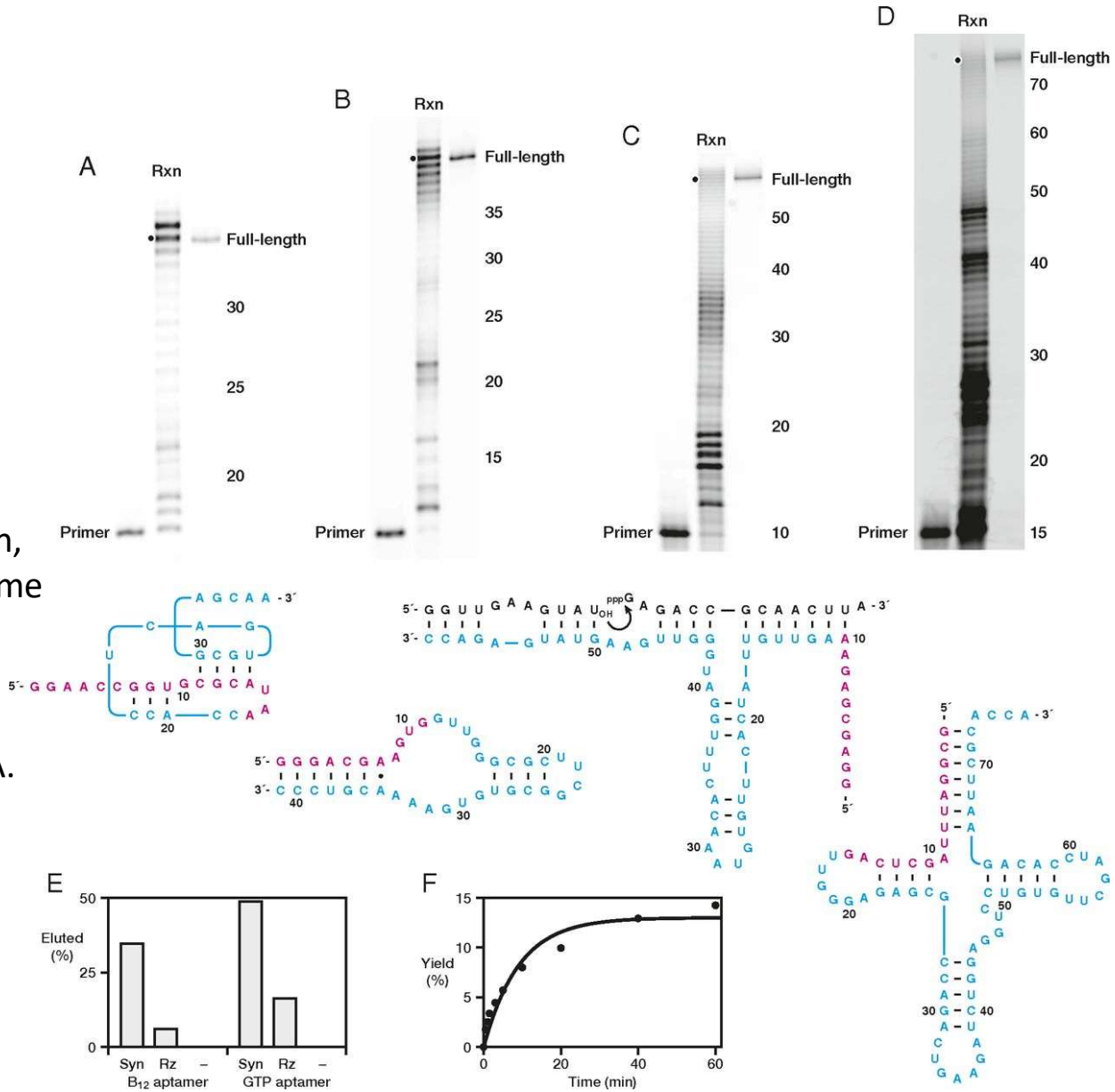


Thus, the replication of RNA and the expression of functional RNA can be accomplished with RNA alone. Combining and improving these activities may enable the self-sustained evolution of RNA and offers a potential route to a synthetic form of RNA life.

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



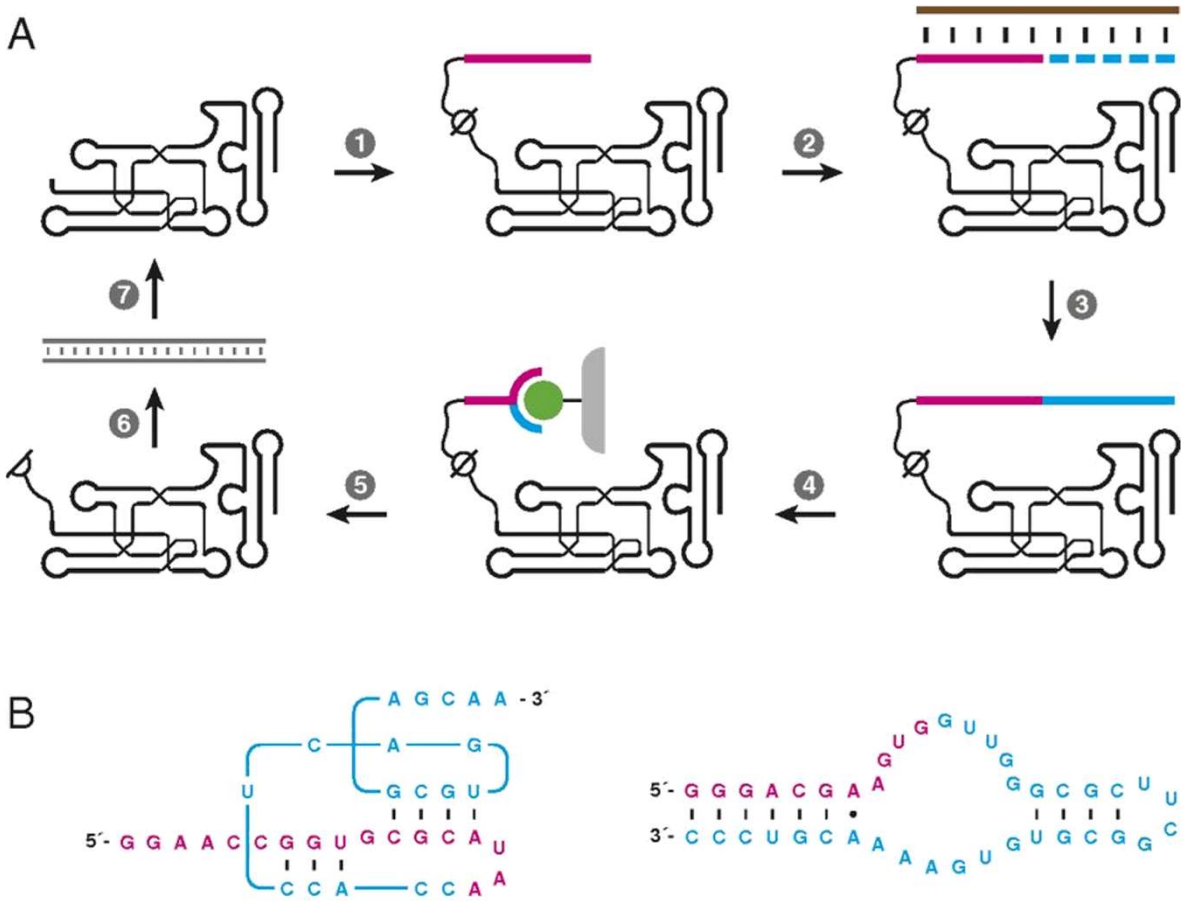
Synthesis of functional RNAs by the 24-3 polymerase. Synthesis of (*A*) the cyanocobalamin aptamer after 24 h, (*B*) the GTP aptamer after 24 h, (*C*) the F1 ligase ribozyme after 24 h, and (*D*) yeast phenylalanyl-tRNA after 72 h.

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).

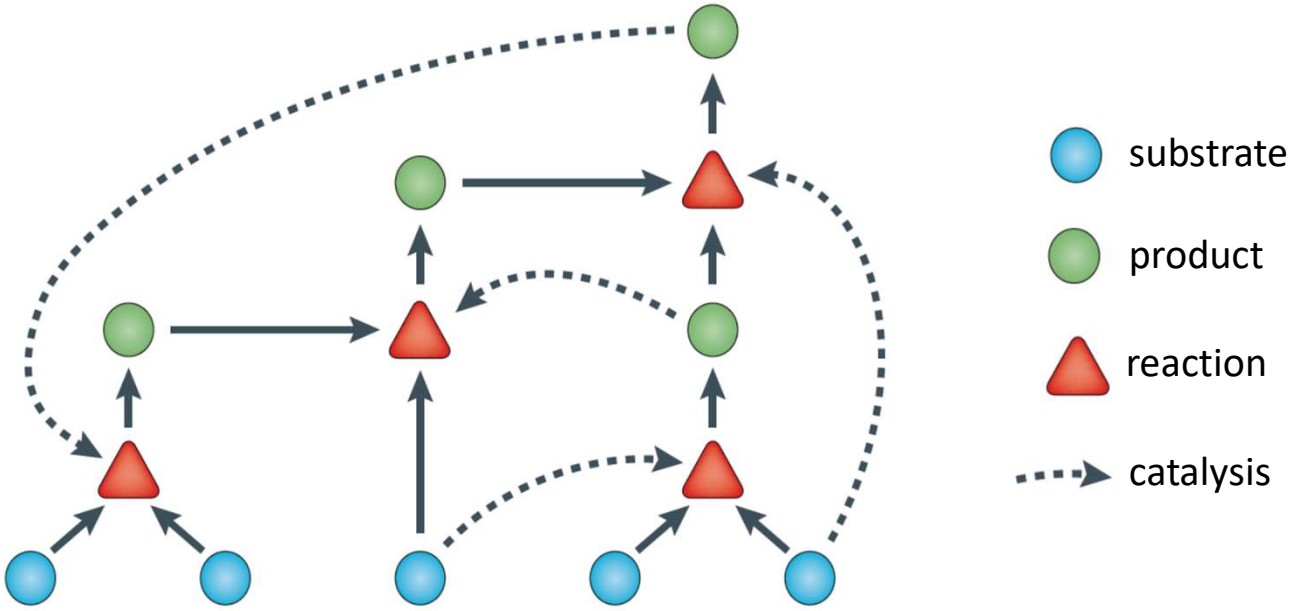
D. P. Horning, G. F. Joyce *Proc. Natl. Acad. Sci. USA (PNAS)* **2016**, *113 (35)*, 9786-9791

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?

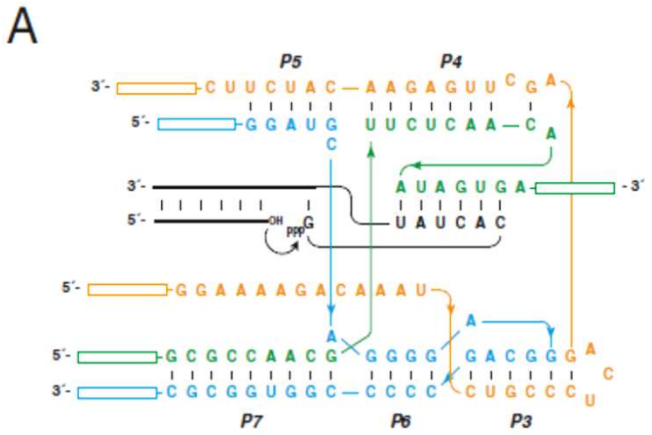
Possible solution – autocatalytic networks



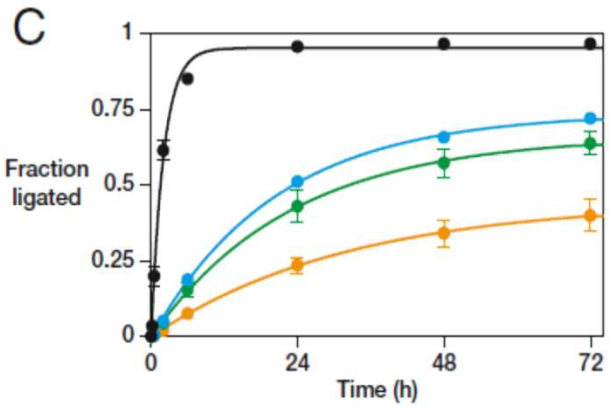
No component can replicate without all the others

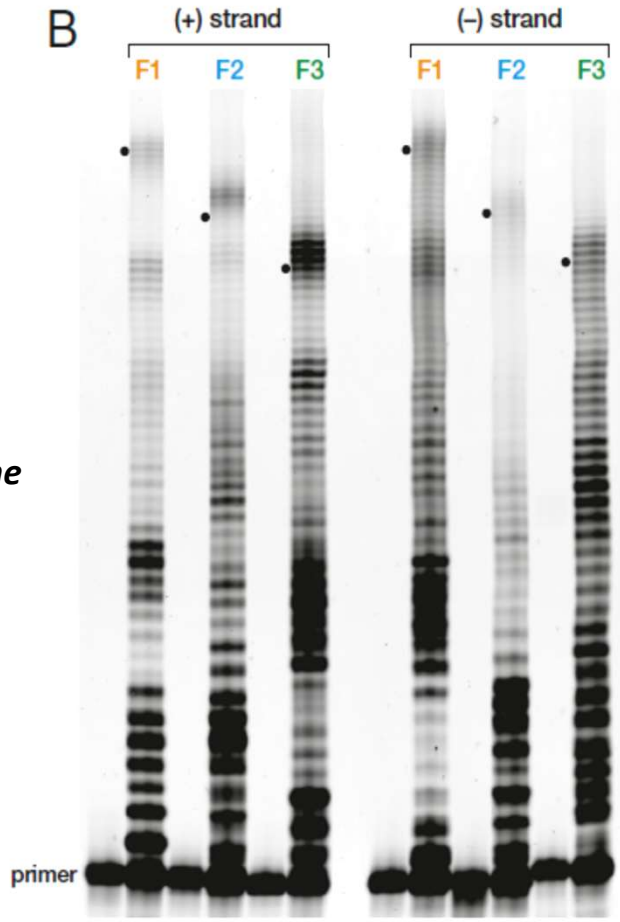
The RNA world

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

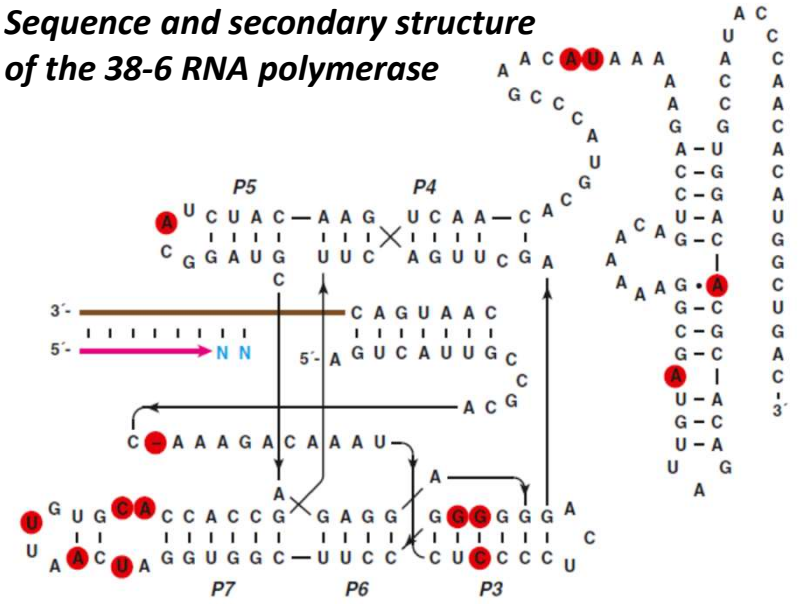


Three fragments of the class I ligase ribozyme





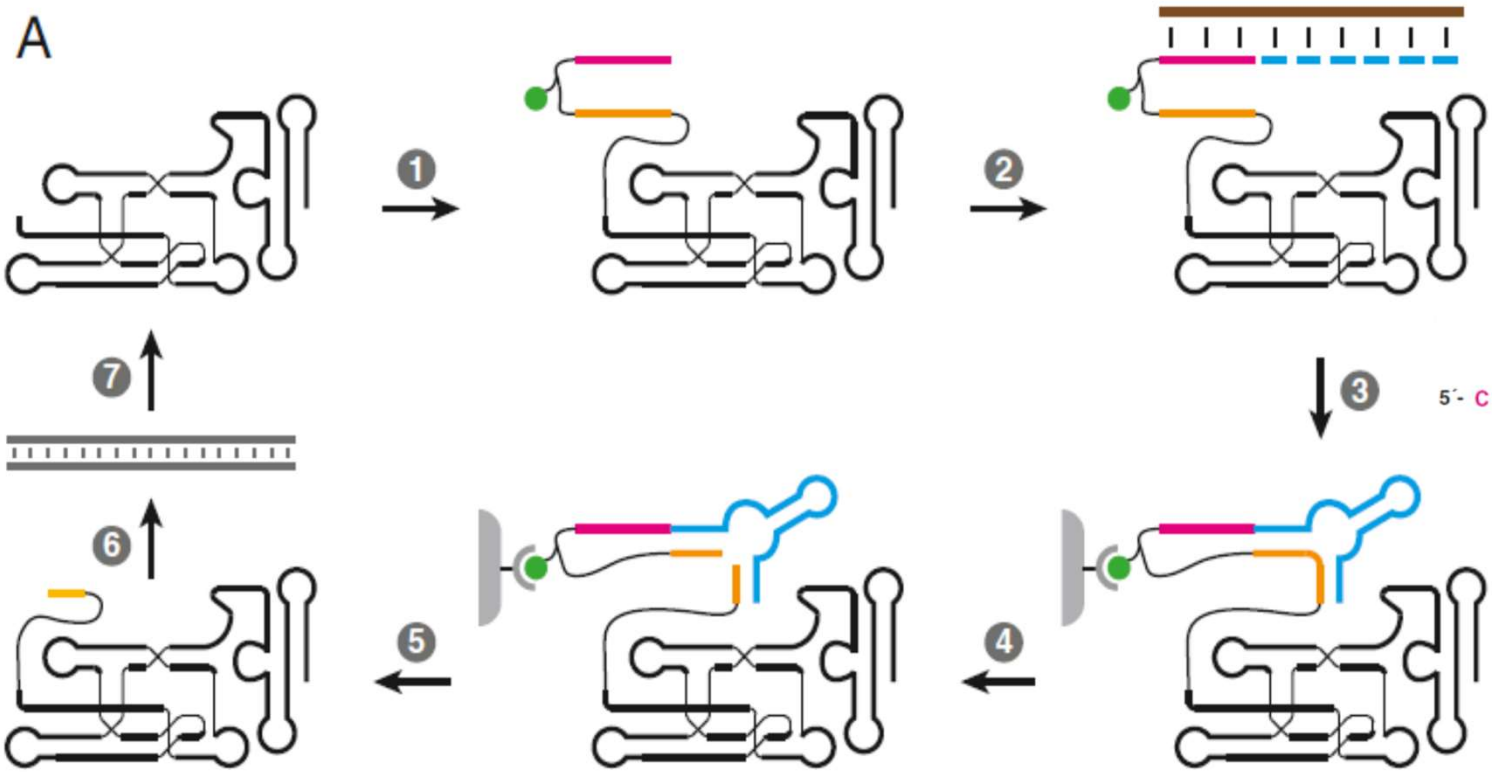
Sequence and secondary structure of the 38-6 RNA polymerase



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

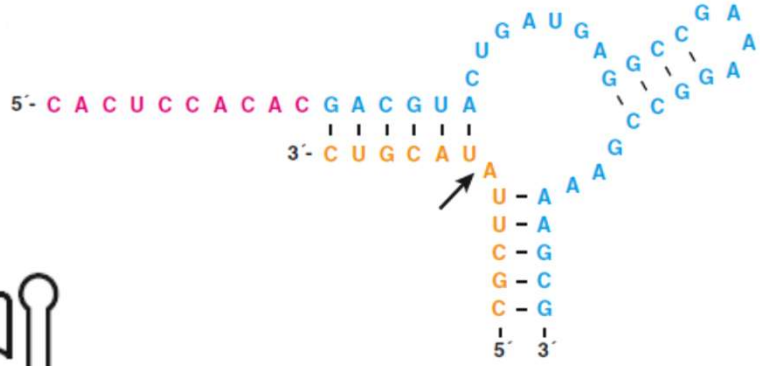
The polymerase „38-6” is able to synthesize its own evolutionary ancestor, an RNA ligase ribozyme, in the form of three fragments that assemble to give a functional complex



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:



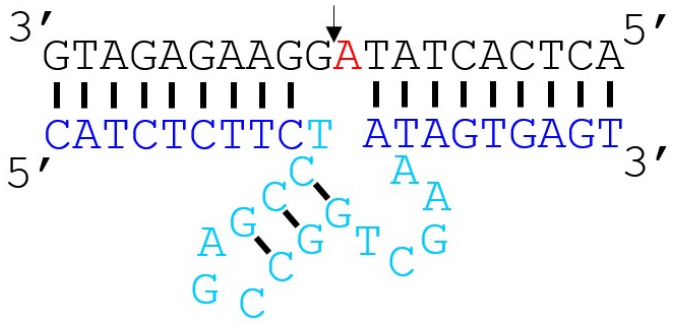
K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce
Proc. Natl. Acad. Sci. USA (PNAS) **2020**, *117 (6)*, 2906-2913

DNAzymes

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

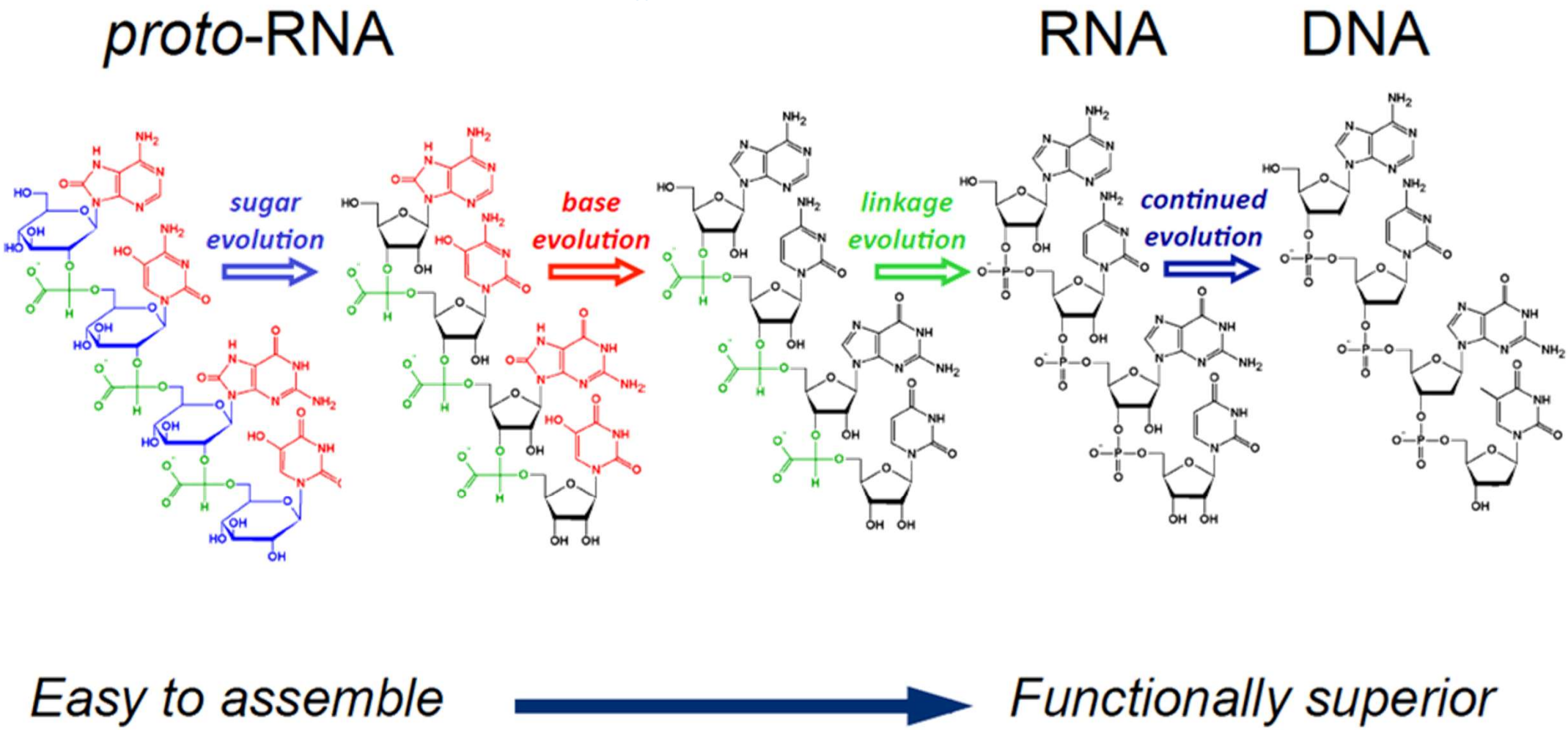
Problems: product inhibition, often single-turnover

The XNA world

Synthetic biology:

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

„RNA-second”



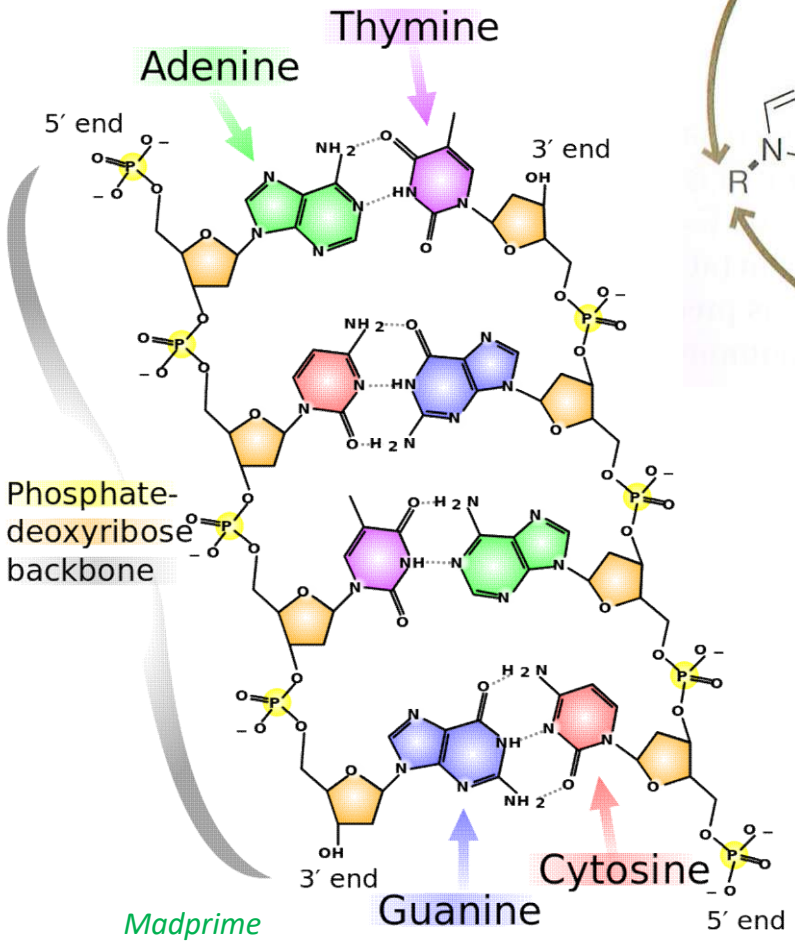
Proto-RNA evolution: According to the protoRNA theory, each of the components of RNA — sugar, base and phosphate backbone — may have originally taken different forms.

CHAPTER 1

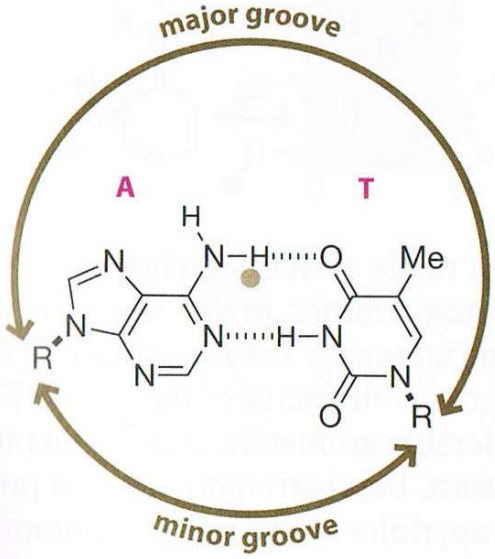
OLIGONUCLEOTIDES

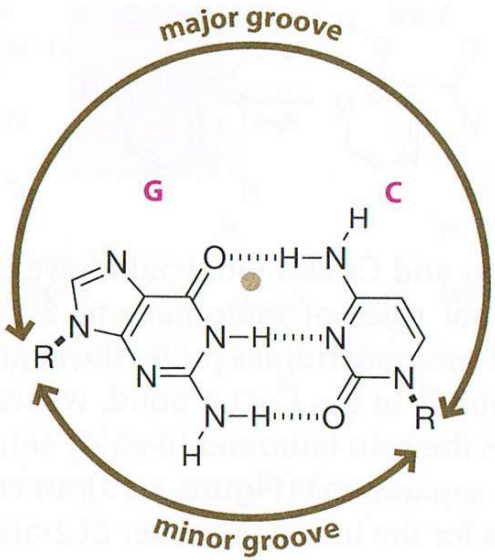
Part 1 – modified canonical nucleobases

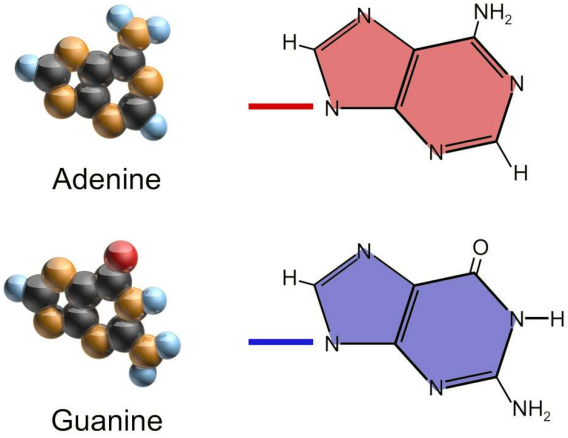
Canonical nucleobases and Watson-Crick pairing in DNA

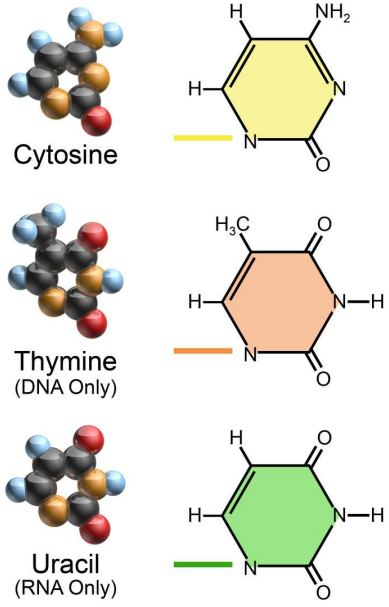


Madprime



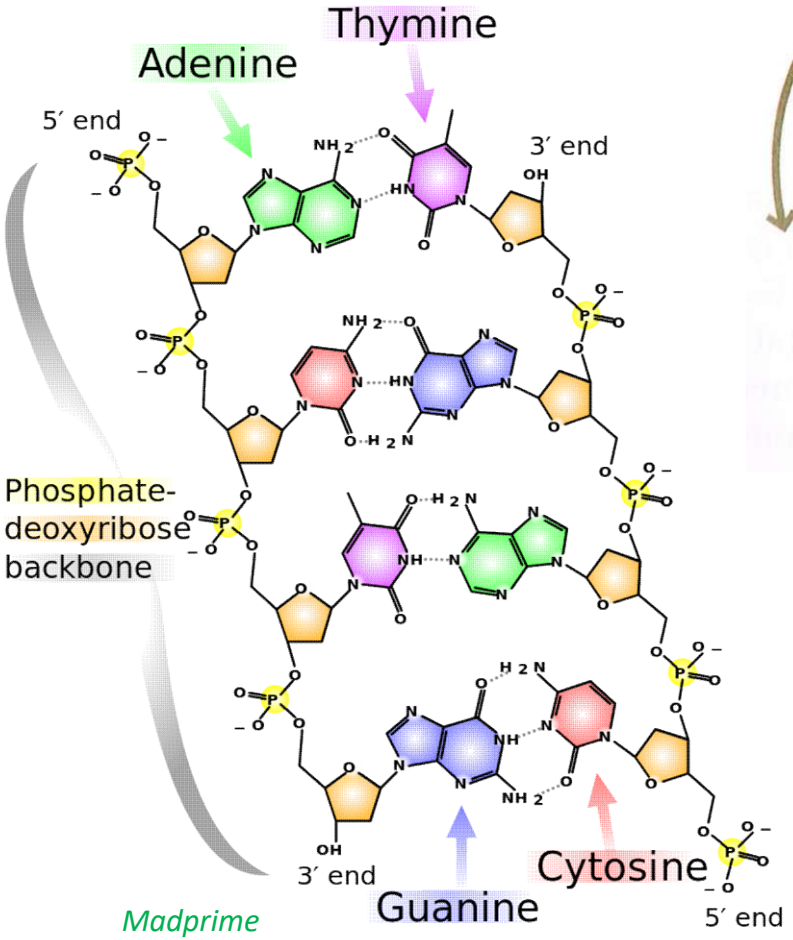


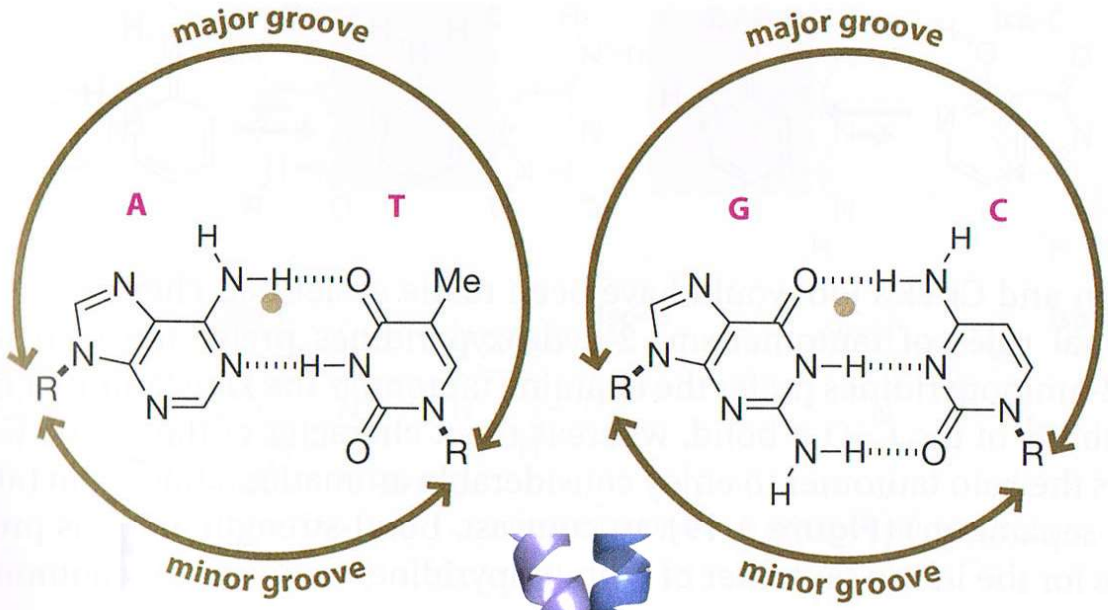


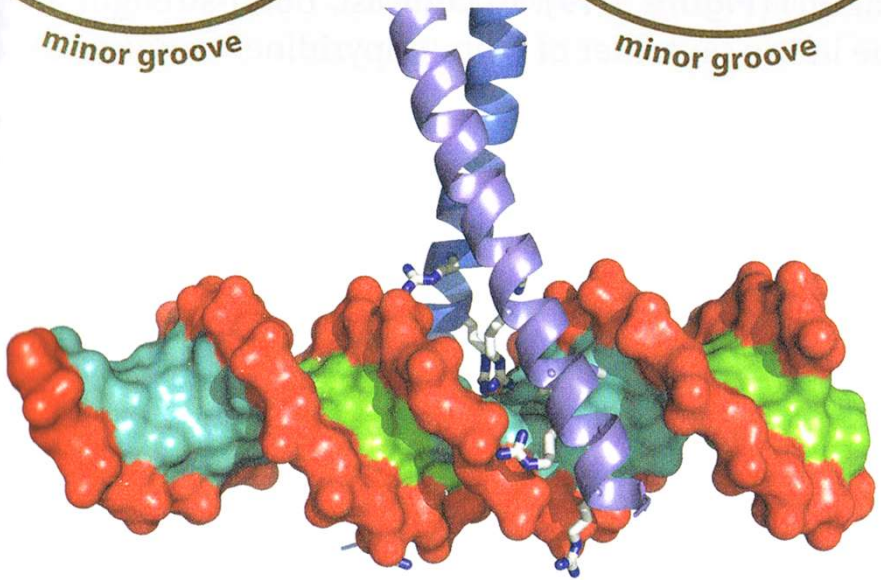


Bruce Blaus

Canonical nucleobases and Watson-Crick pairing in DNA







Hoogsten base pairing of canonical DNA nucleobases

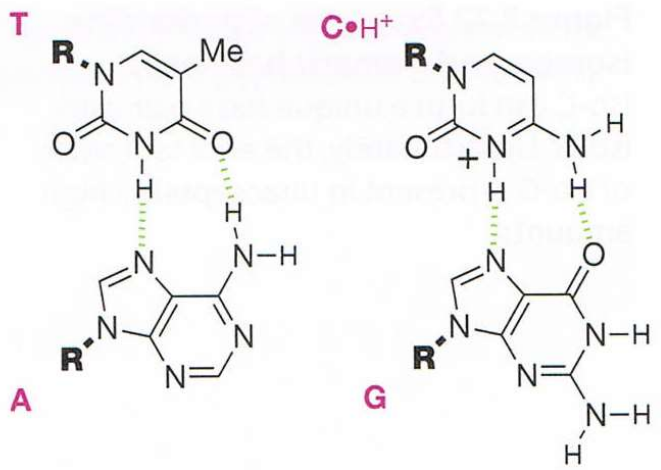
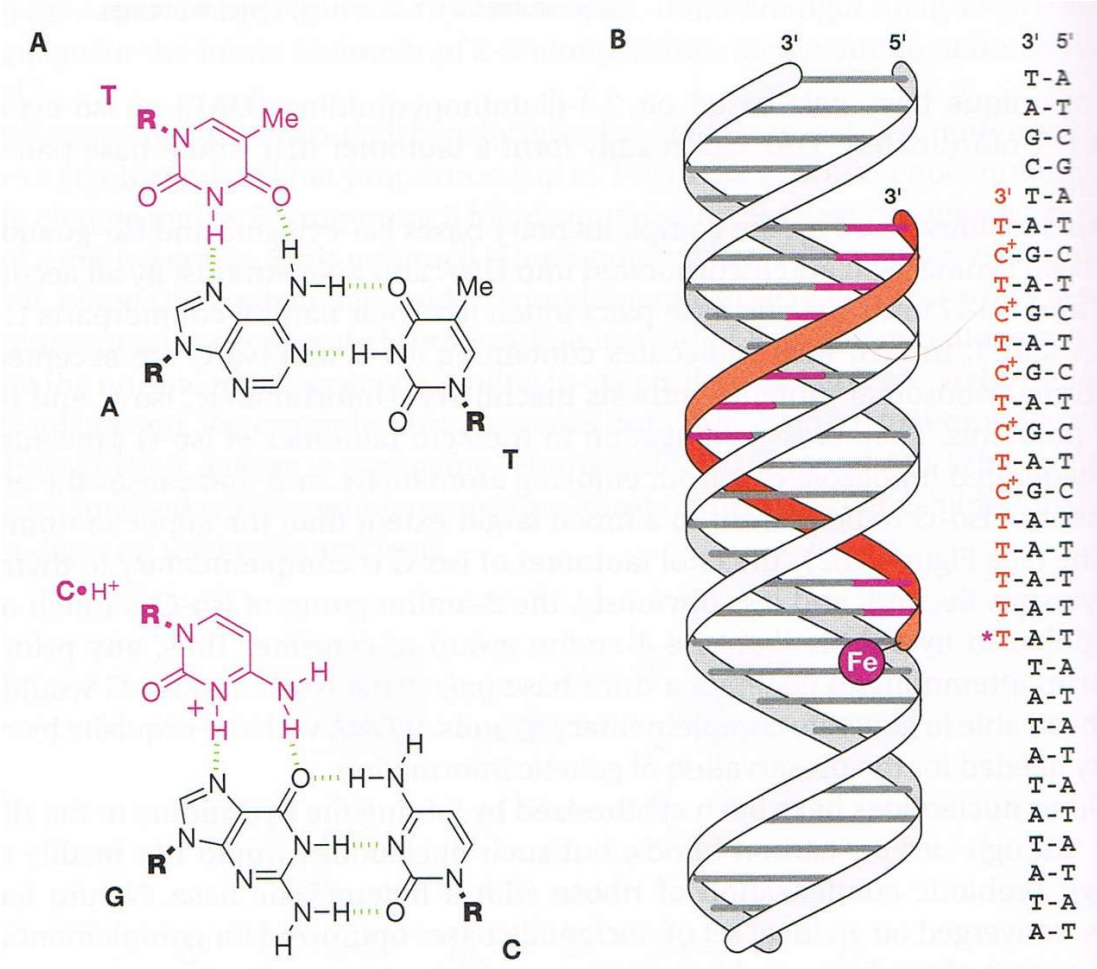
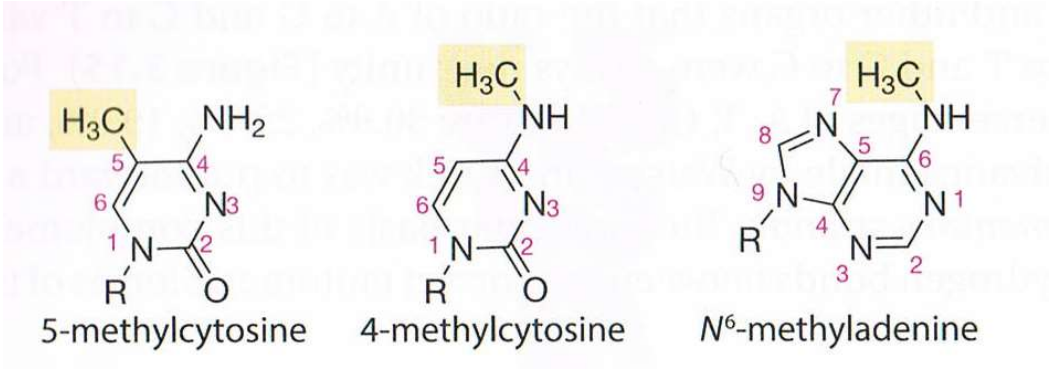


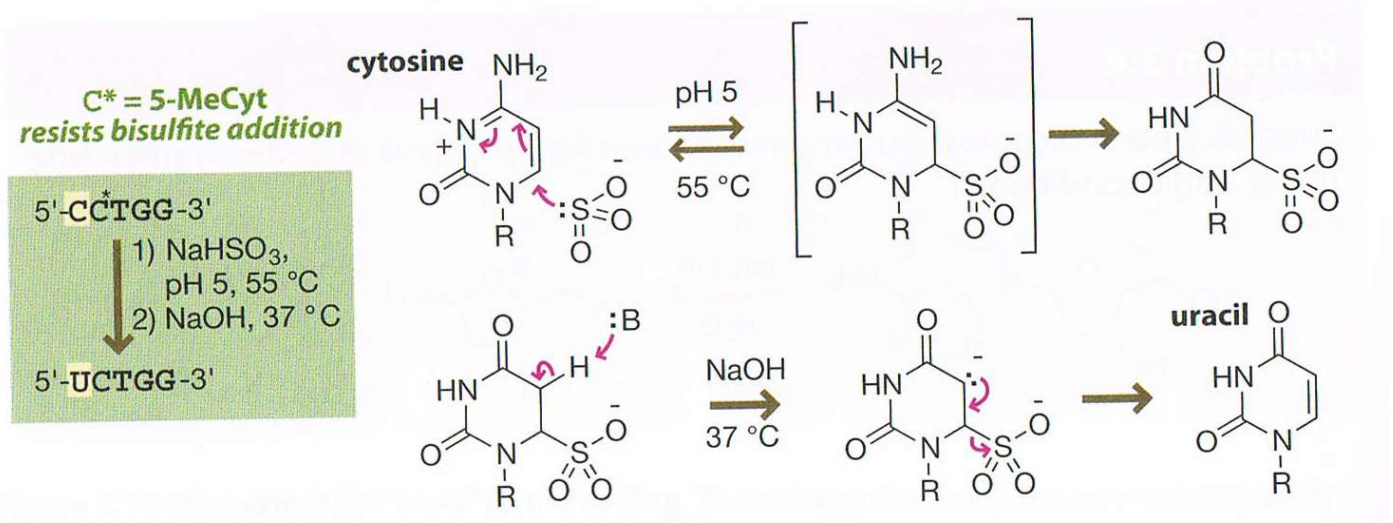
Figure 3.24 Hoogsteen base pairs. Hoogsteen base pairs use a different edge of the purine from a Watson–Crick base pair.



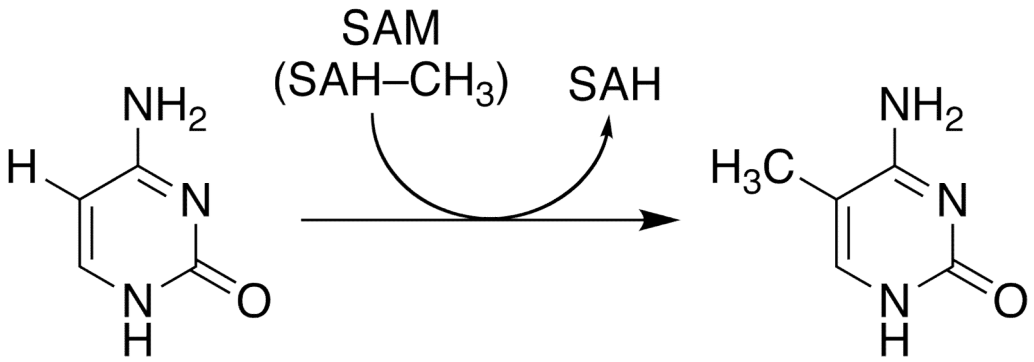
H. E. Moser, P. B. Dervan *Science* **1987**, *238,* 645-650

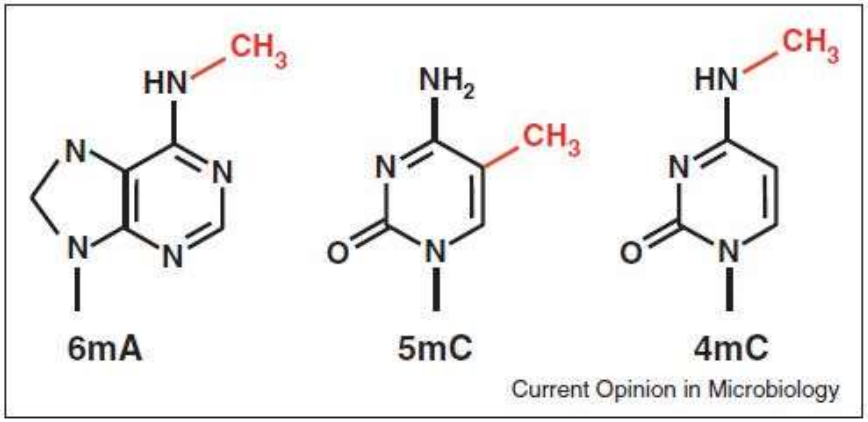
Modifications of nucleobase structures tolerated by polymerases



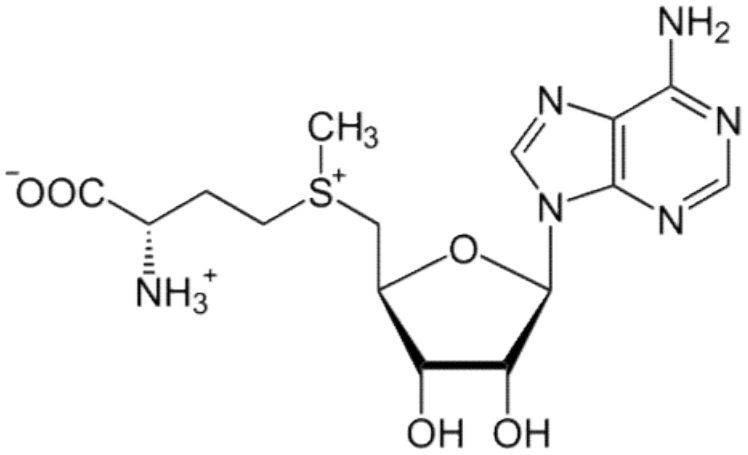


Modifications of nucleobases

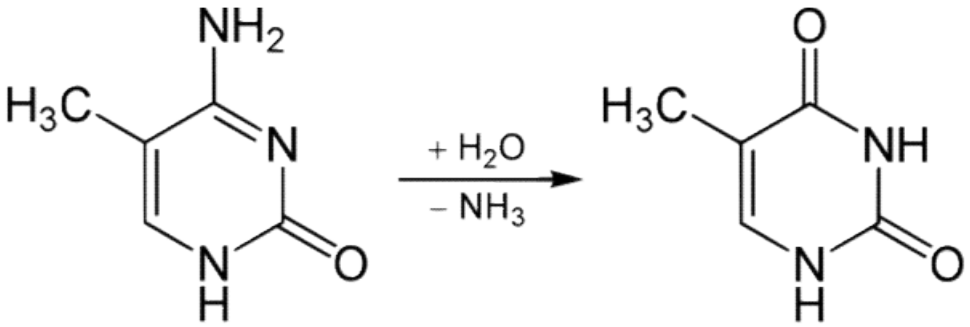




Chemical structures of common modified bases generated by DNA methyltransferases.

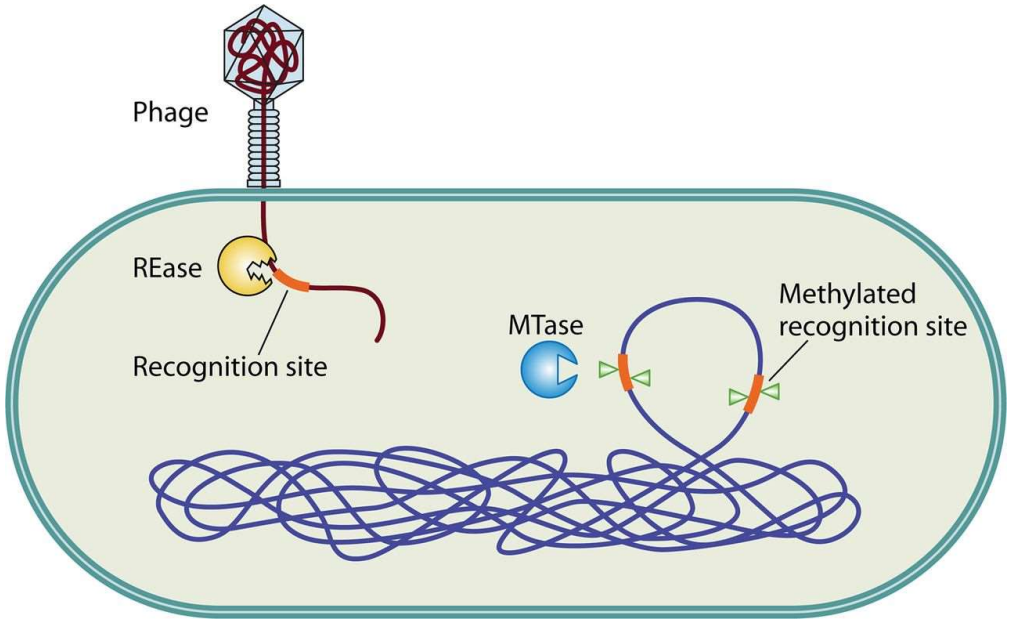


S-Adenosylmethionine (SAM)

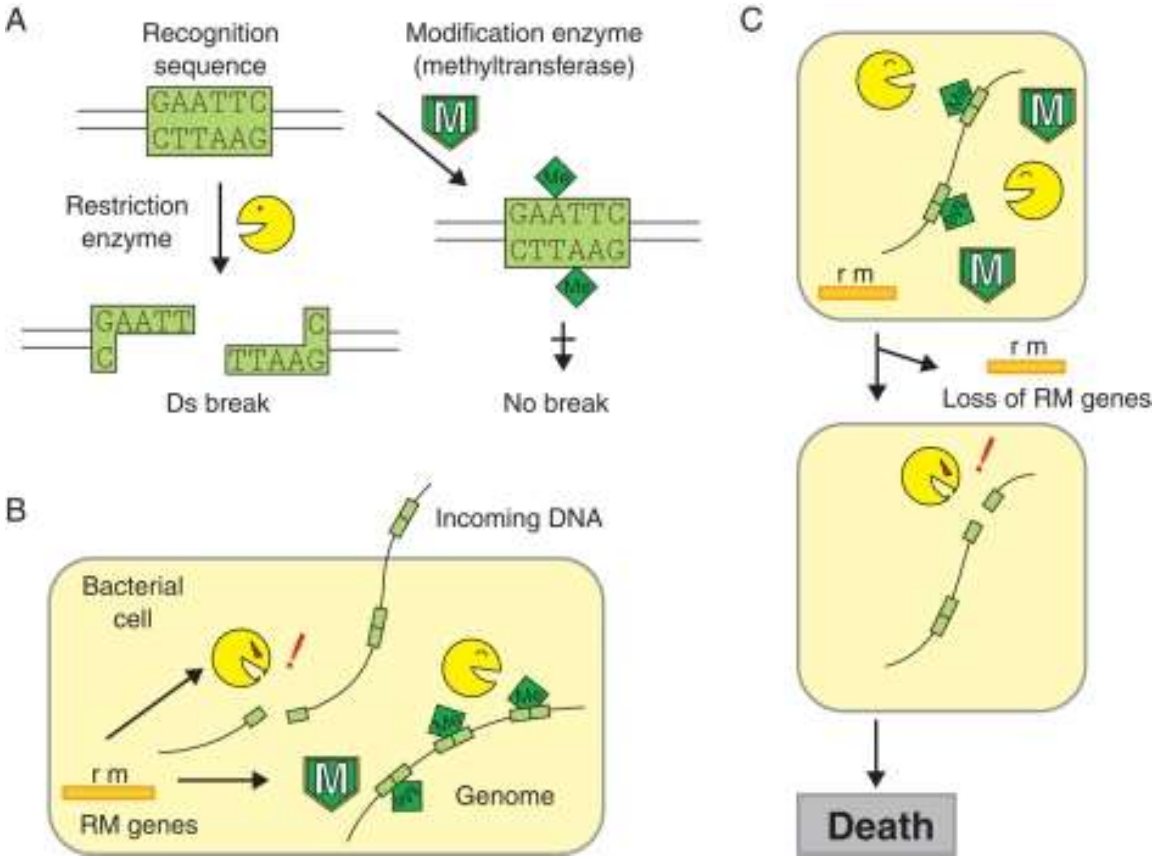


Restriction modification system

„Immune system" of bacteria and archaea against atacking viruses



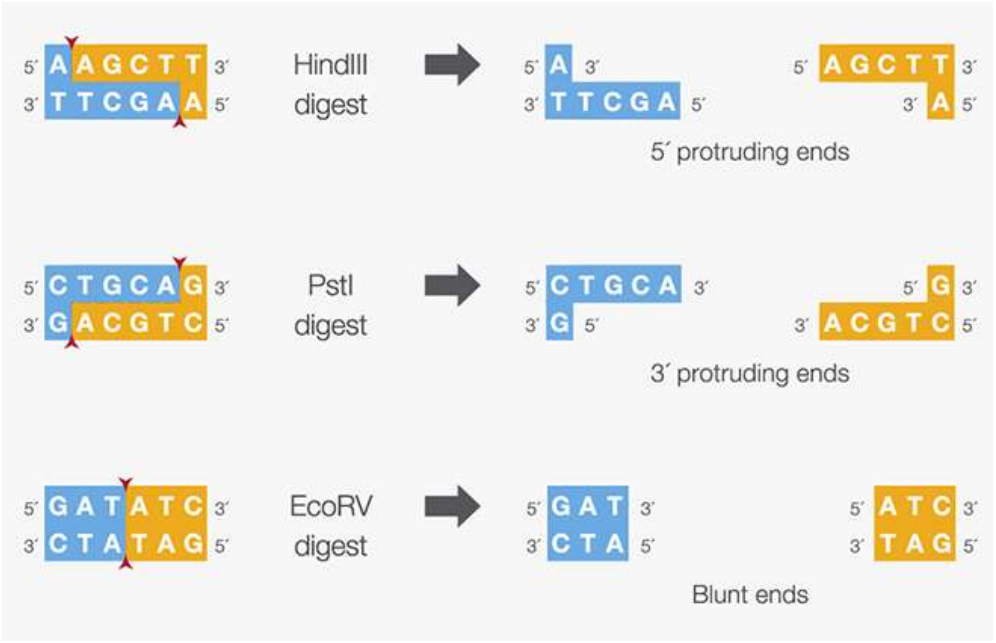
K. Vasu, V. Nagaraja
Microbiol. Mol. Biol. Rev. **2013**, *77(1)*, 53-72

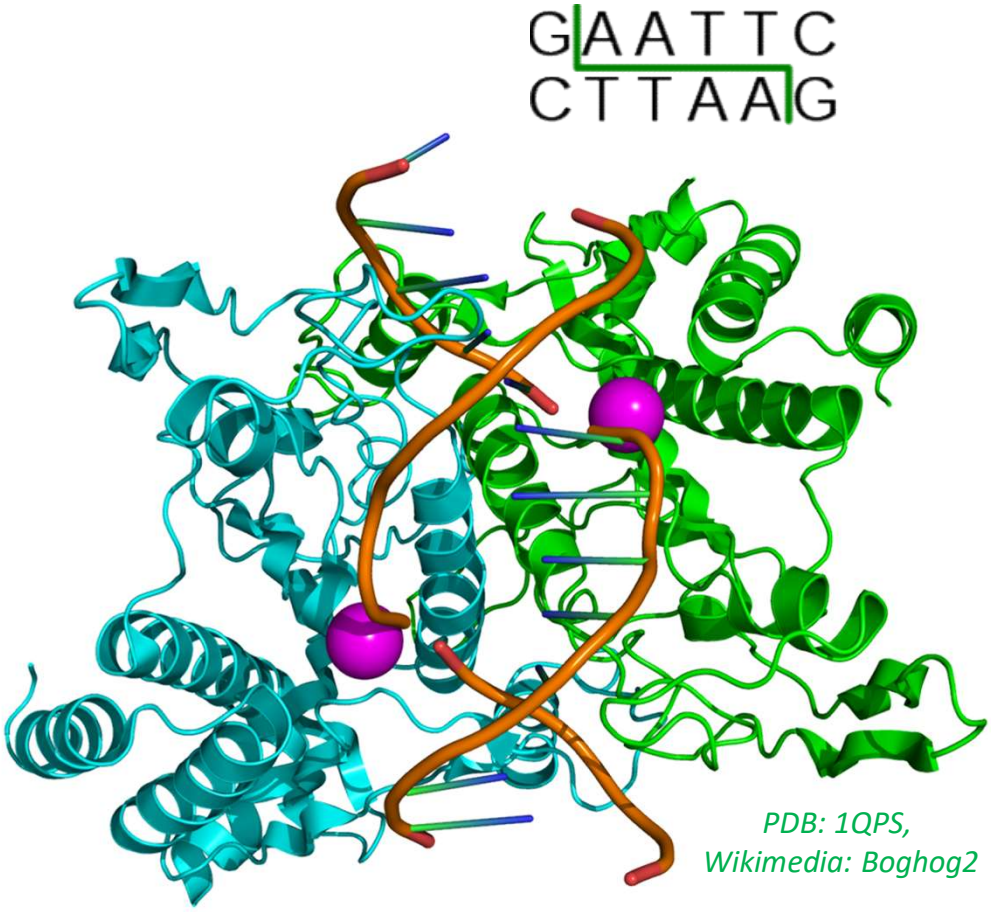


K. Ishikawa *et al. DOI: 10.1093/dnares/dsq027*

EcoI – a typical restriction enzyme

Products of restriction enzymes

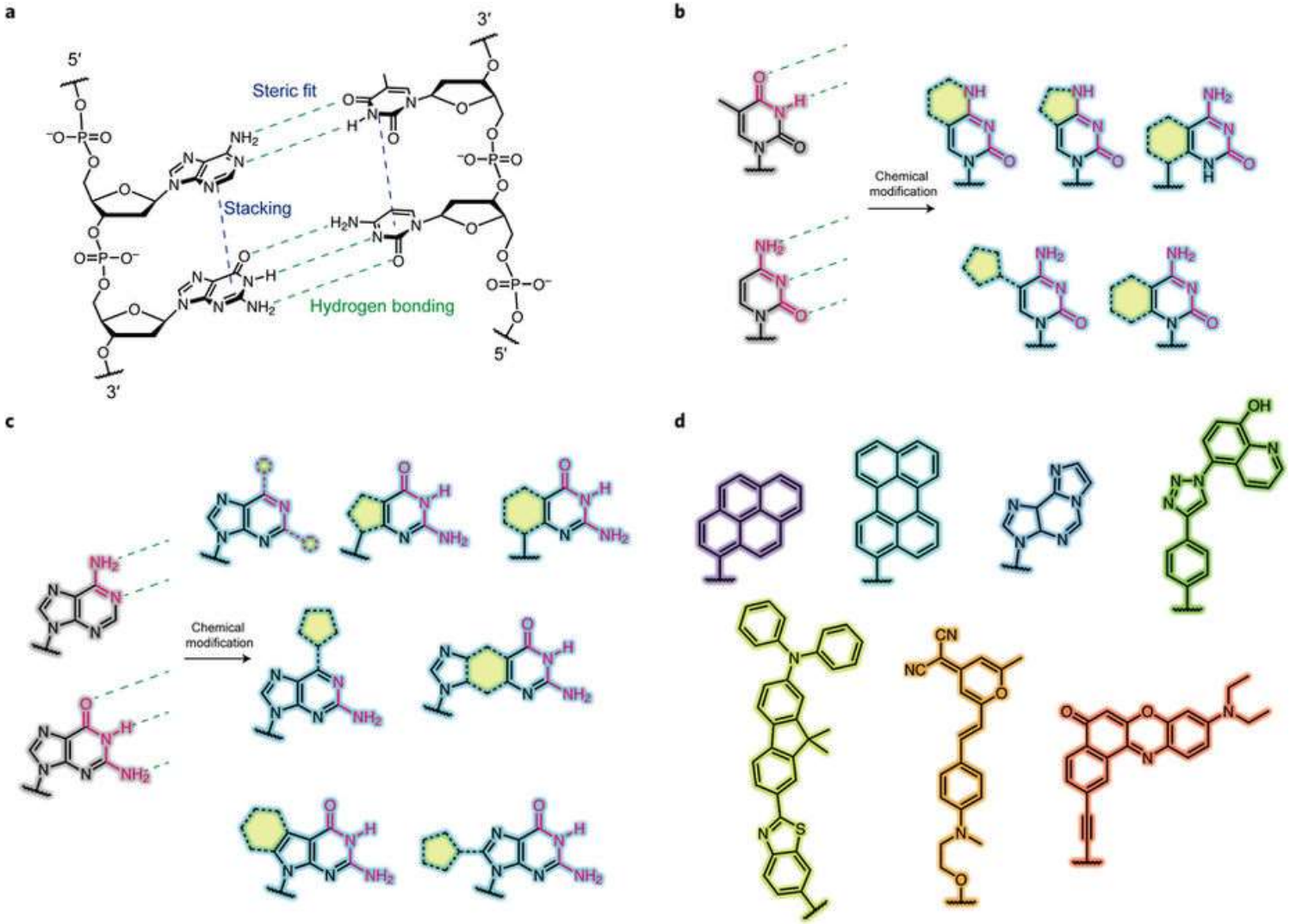




PDB: 1QPS, Wikimedia: Boghog2

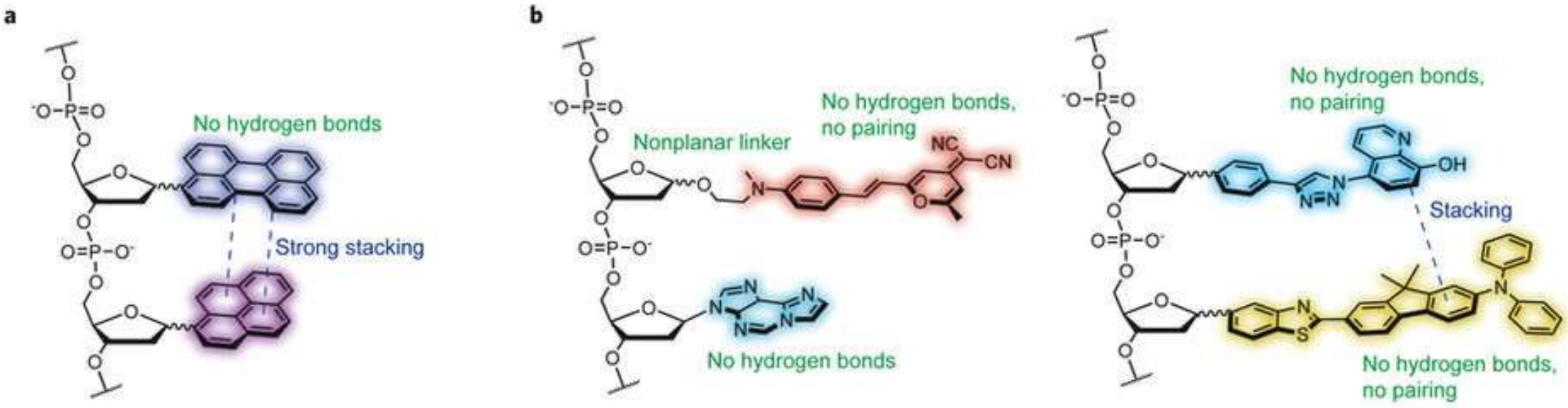
Structure of the homodimeric restriction enzyme EcoRI (cyan and green cartoon diagram) bound to double stranded DNA (brown tubes). Two catalytic magnesium ions (one from each monomer) are shown as magenta spheres and are adjacent to the cleaved sites in the DNA made by the enzyme (depicted as gaps in the DNA backbone).

Non-canonical fluorescent nucleobases



W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9,* 1043-1055

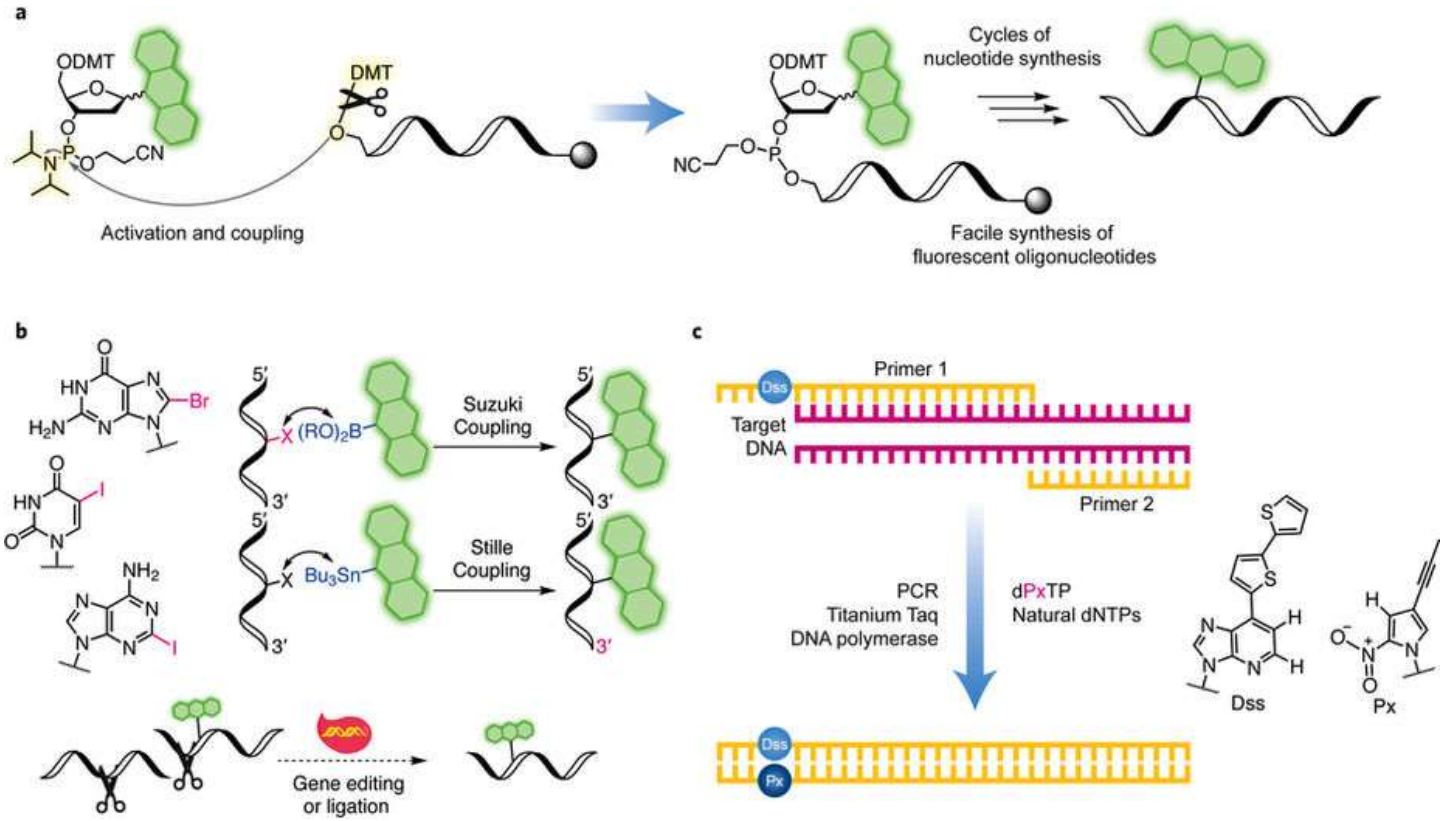
Non-canonical fluorescent nucleobases



a, Nucleobases composed of aromatic hydrocarbons. **b**, Nucleobases composed of planar heterocyclic fluorophores. The lack of hydrogen bonding and weaker π-stacking are compensated by versatile energy states brought from the heteroatoms. These fluorophores contribute to a broader spectrum of emission wavelengths. Functional groups can be added to expand functionality, such as metal binding.

W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9,* 1043-1055

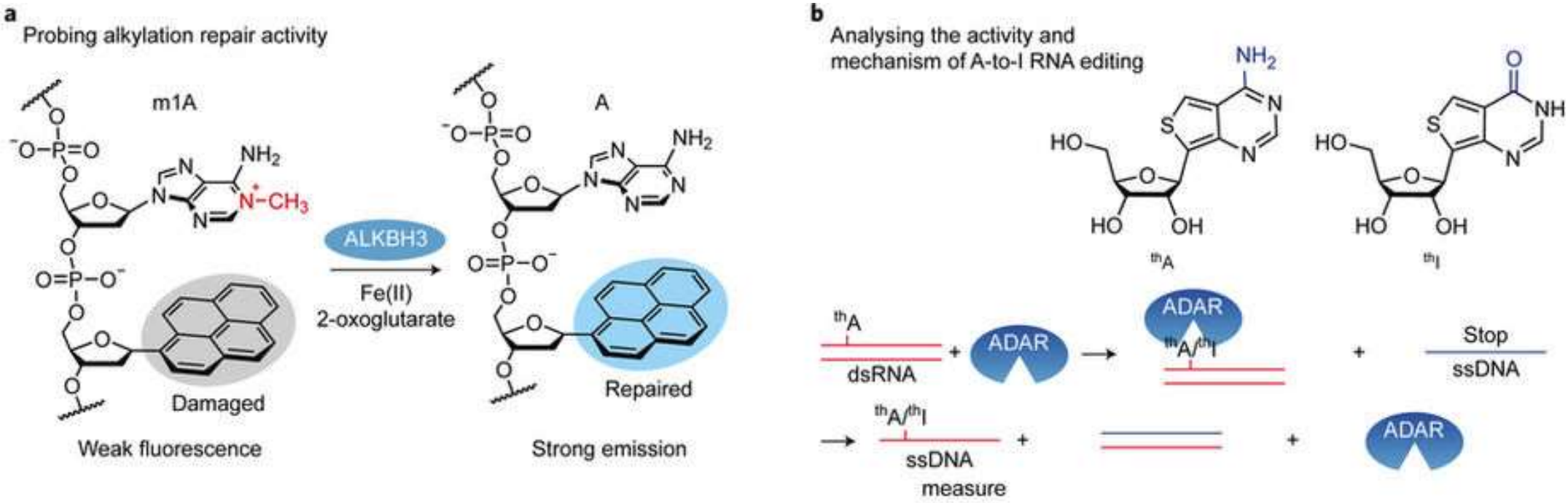
Strategies for incorporating fluorescent nucleobases



a, Direct oligonucleotide synthesis via synthesizer and phosphoramidite chemistry. **b**, Post-synthesis modification using mild coupling methods or gene-editing methods. Gene-editing and ligation methods enzymatically join smaller labelled strands to make longer ones. **c**, Direct enzymatic incorporation using fluorescent nucleoside triphosphate derivatives. When the fluorescent nucleobases are labelled in the primers or supplied as free nucleobases in the pool, polymerases that recognize them can incorporate the fluorescent nucleobases into DNA sequences.

W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9,* 1043-1055

Fluorescent nucleobases for studying DNA and RNA

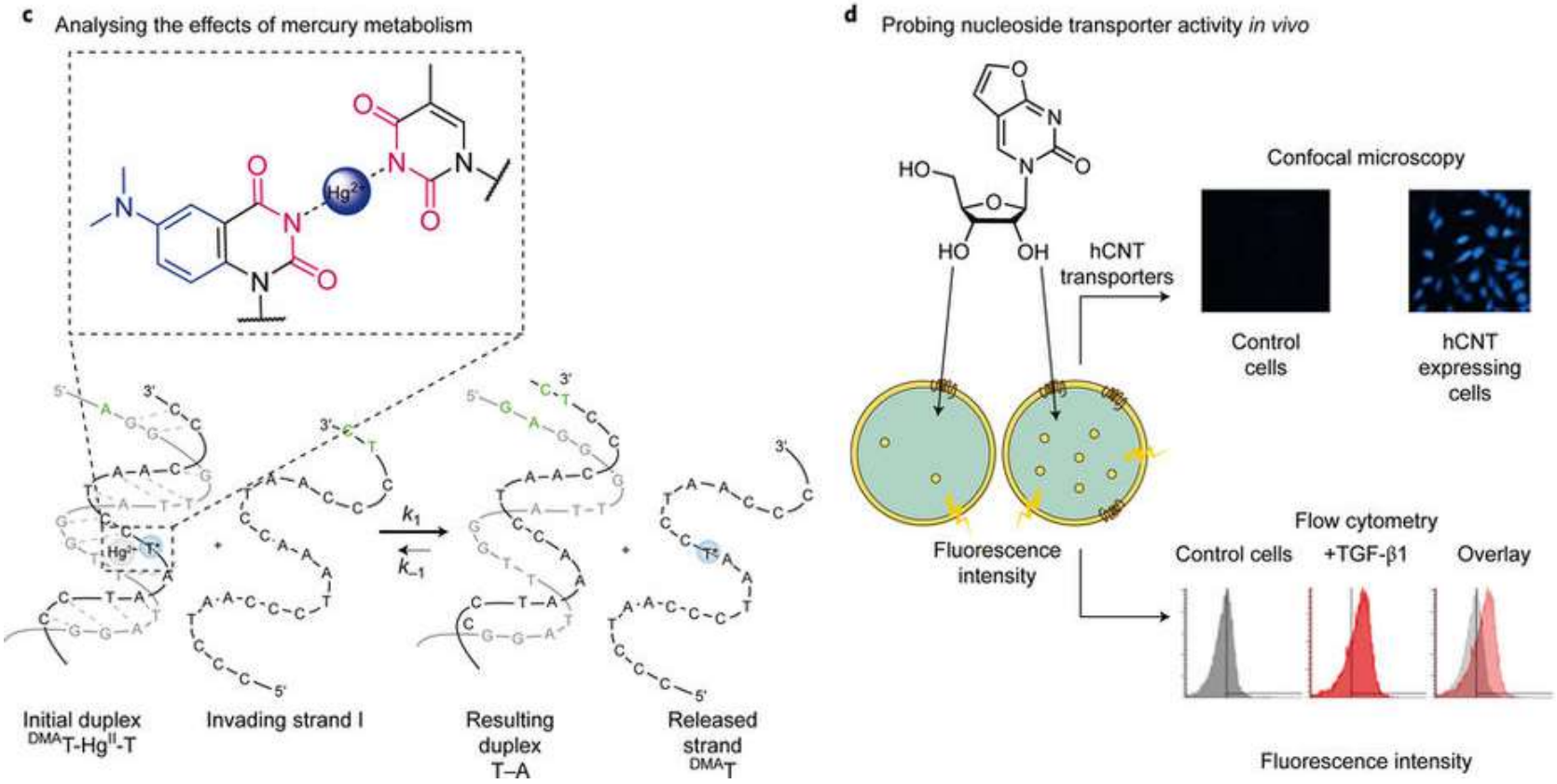


a, Fluorogenic sensing of a demethylation enzyme, ALKBH3. The emission of pyrene is initially quenched by the positive charge of 1-methylated adenine (m1A). When ALKBH3 demethylates m1A, the quenching effect is removed and a signal is generated.

b, Fluorogenic analysis of adenine-to-inosine RNA editing enzyme. The emission maxima of the thiolated adenine (thA) and inosine (thI) are different. Hence by measuring the intensity of thA and thI at their respective maximal wavelengths, the activity of the A-to-I enzyme can also be measured.

W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9,* 1043-1055

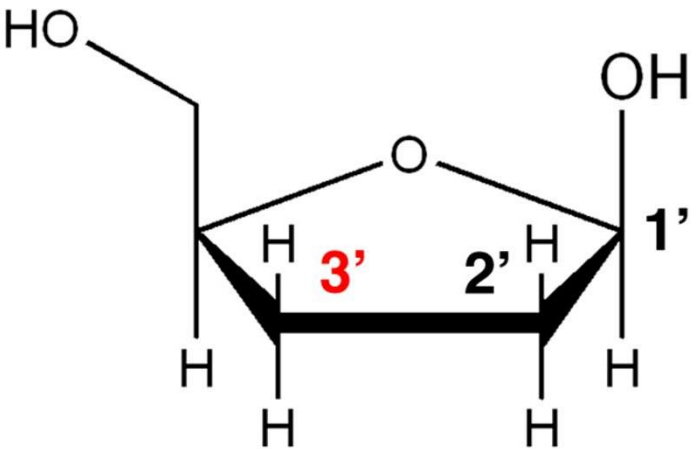
Fluorescent nucleobases for studying DNA and RNA

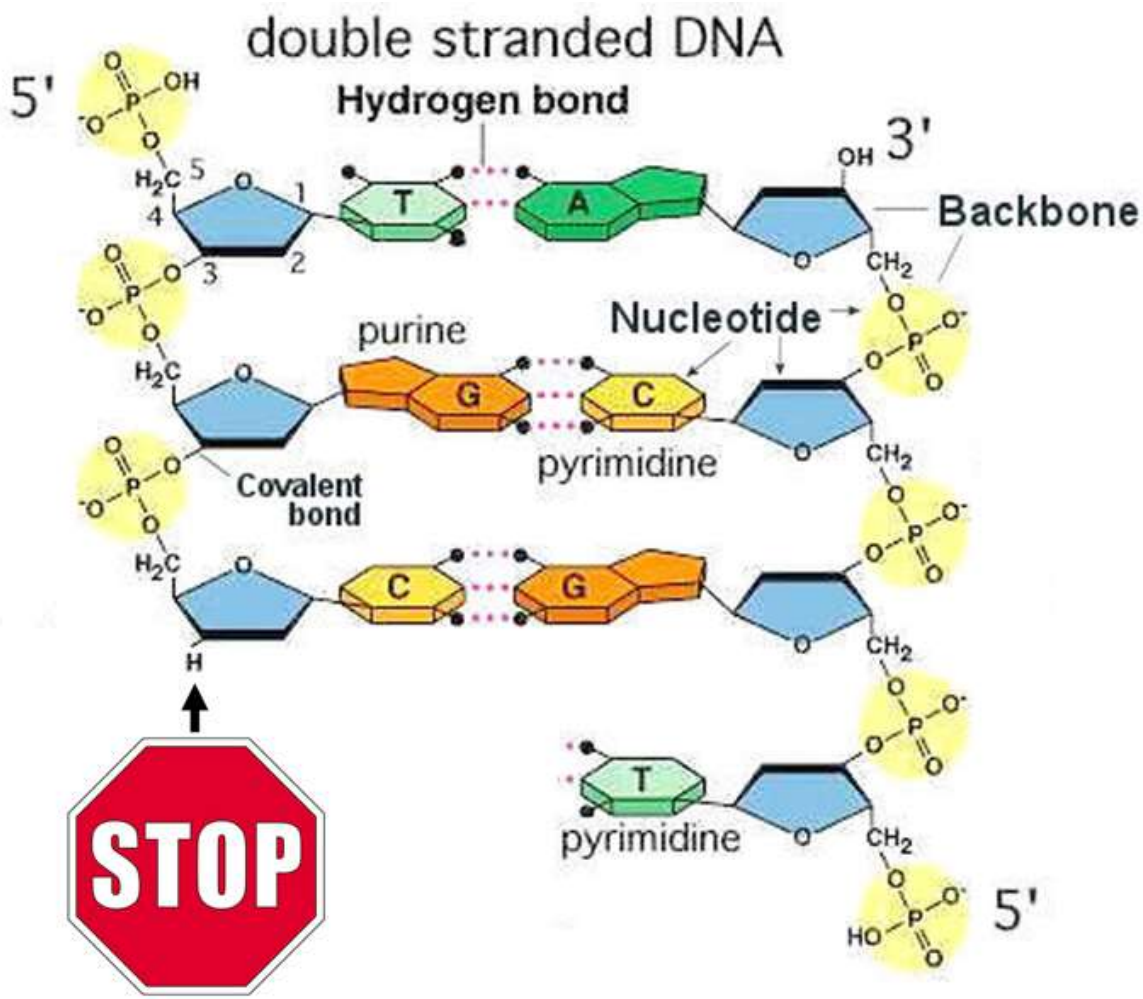


c, Kinetic and thermodynamic investigation of the effects of mercury on DNA metabolism. The fluorescent thymine can chelate mercury with another thymine ring and link DNA strands. This can be used to probe mercury metabolism in vivo and to study the effects of mercury on DNA status.

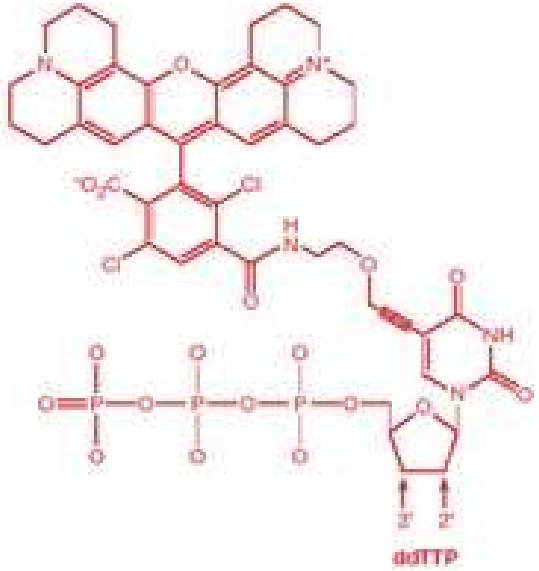
d, Visualization and analysis of human concentrative nucleoside transporters (hCNTs). The fluorescent nucleoside can enter the plasma membrane through the transporters, thus allowing the measurement of the transport activity. TGF-β1, transforming growth factor β1. W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9,* 1043-1055

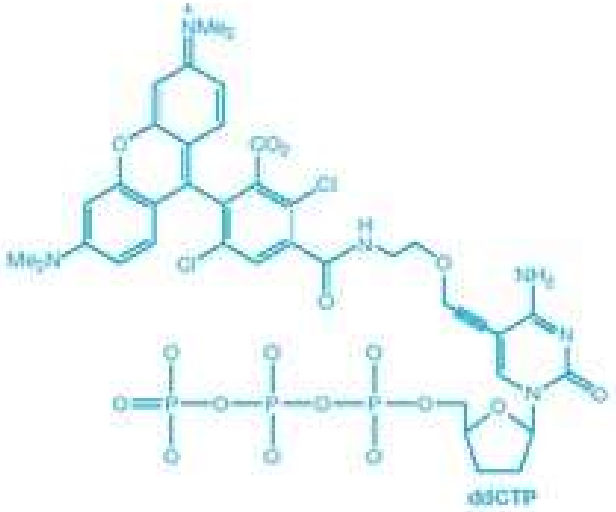
Sanger sequencing

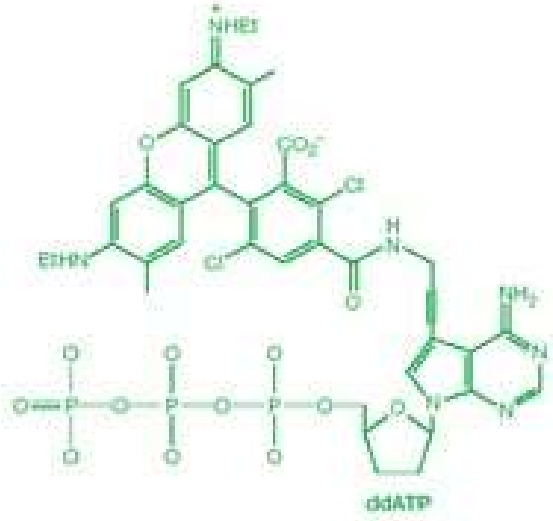


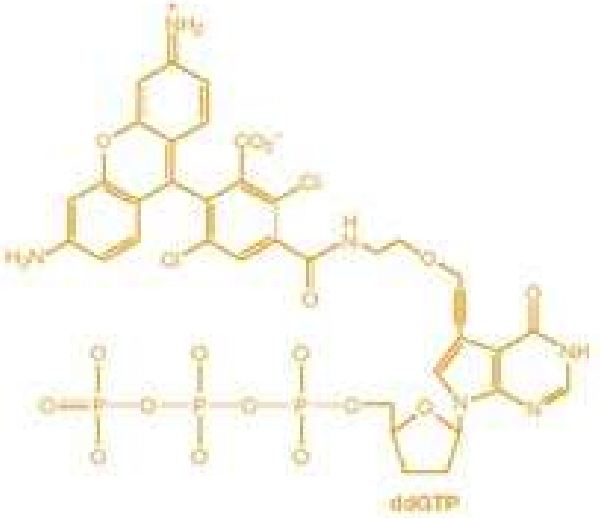


Sanger sequencing

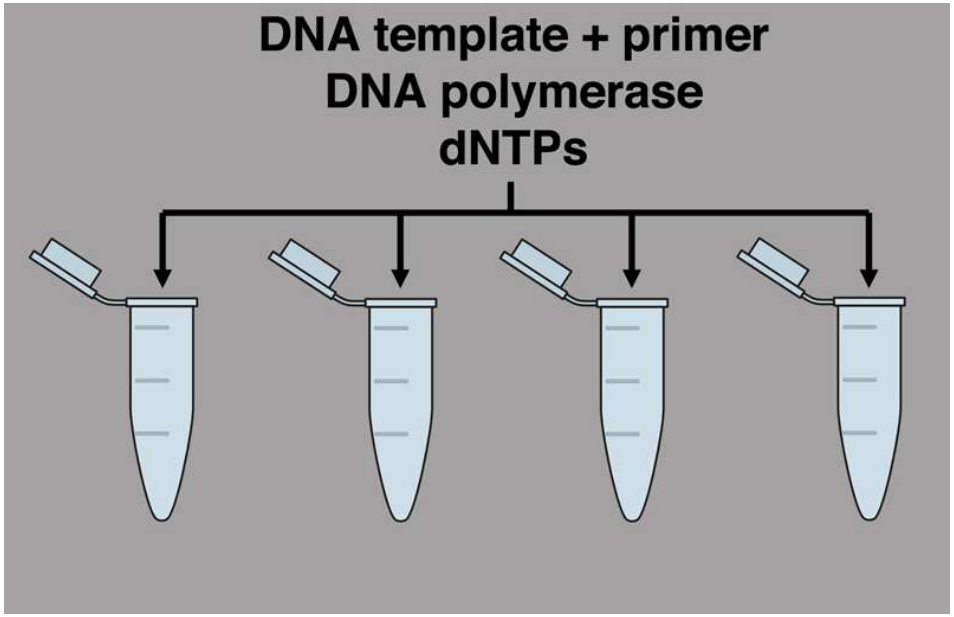


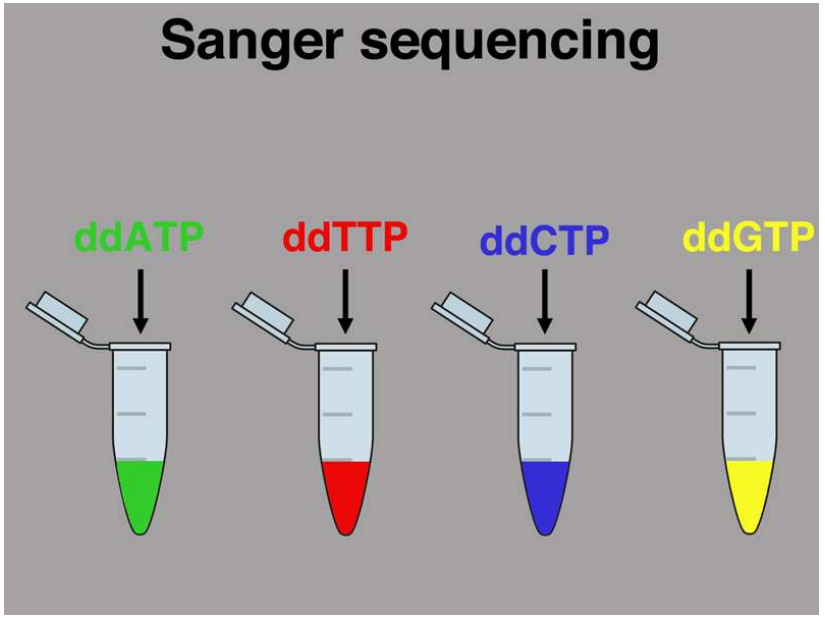






Sanger sequencing





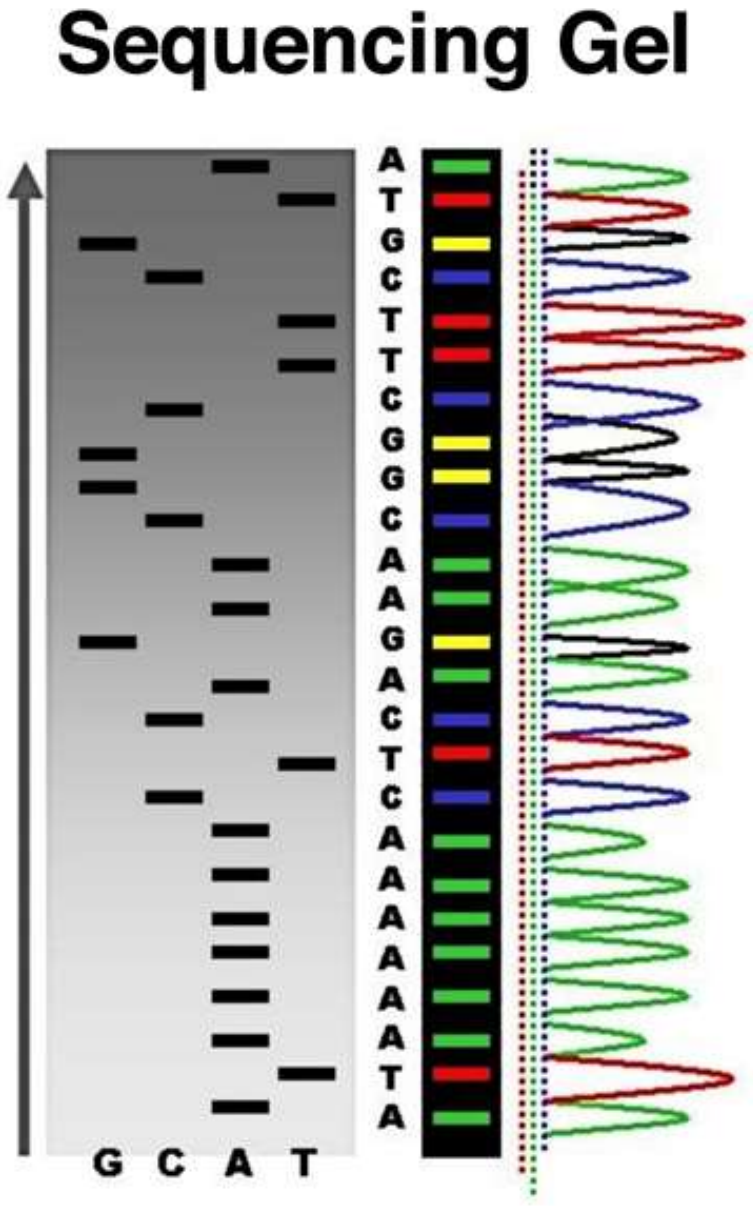
Sanger sequencing

TACGTA
ATGCATTAGGGCCTGGCTCTTT

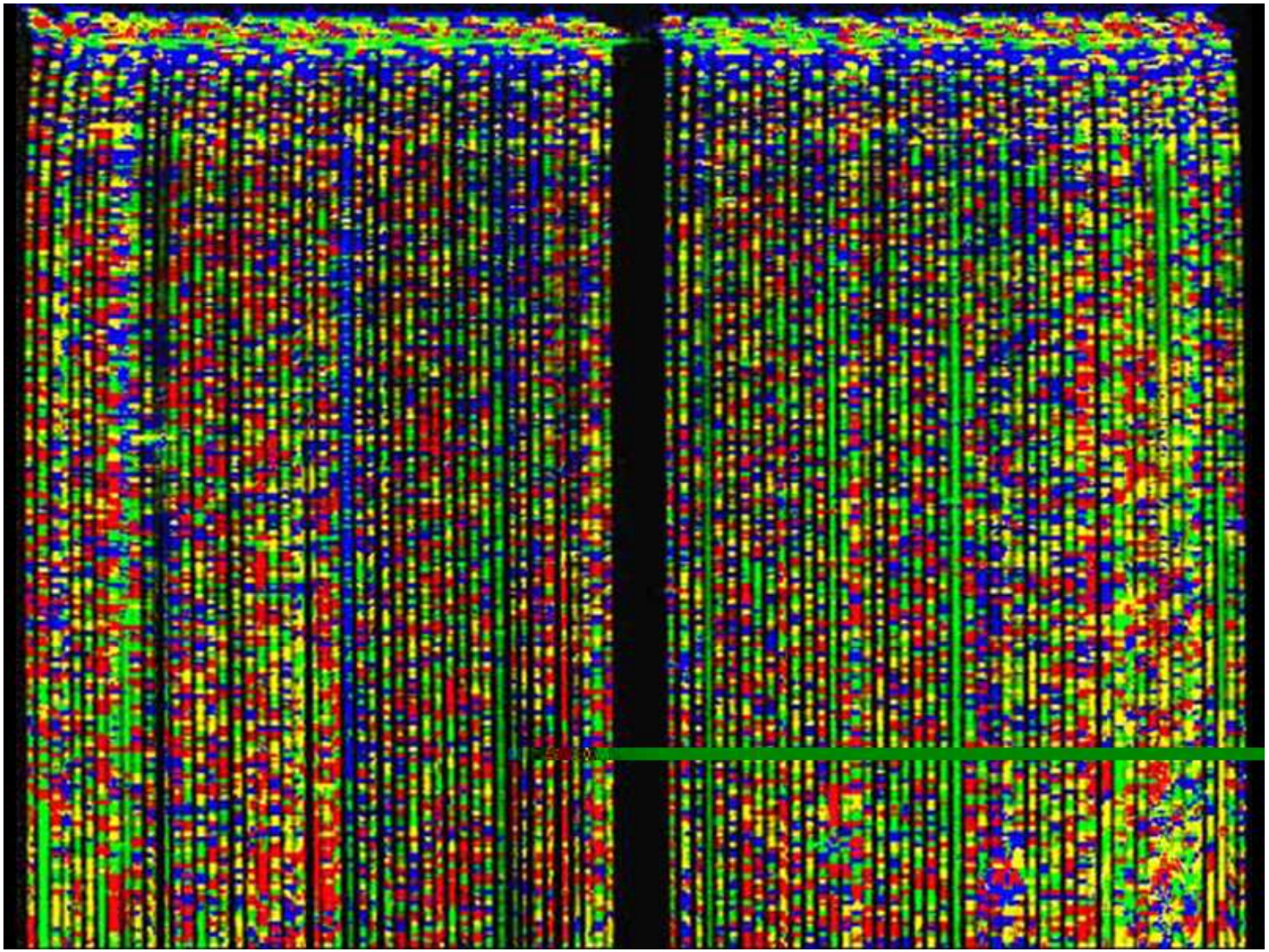
TACGTAA
ATGCATTAGGGCCTGGCTCTTT

TACGTAATCCCGGA
ATGCATTAGGGCCTGGCTCTTT

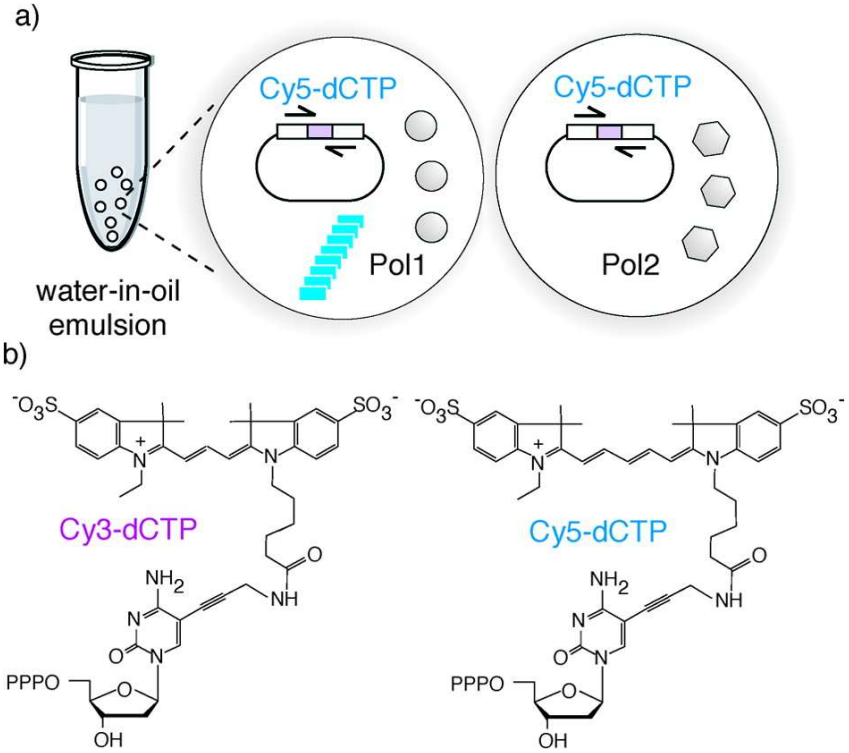
TACGTAATCCCGGACCGA
ATGCATTAGGGCCTGGCTCTTT



Sanger sequencing



CyDNA – synthesis and replication of highly fluorescently-labelled DNA



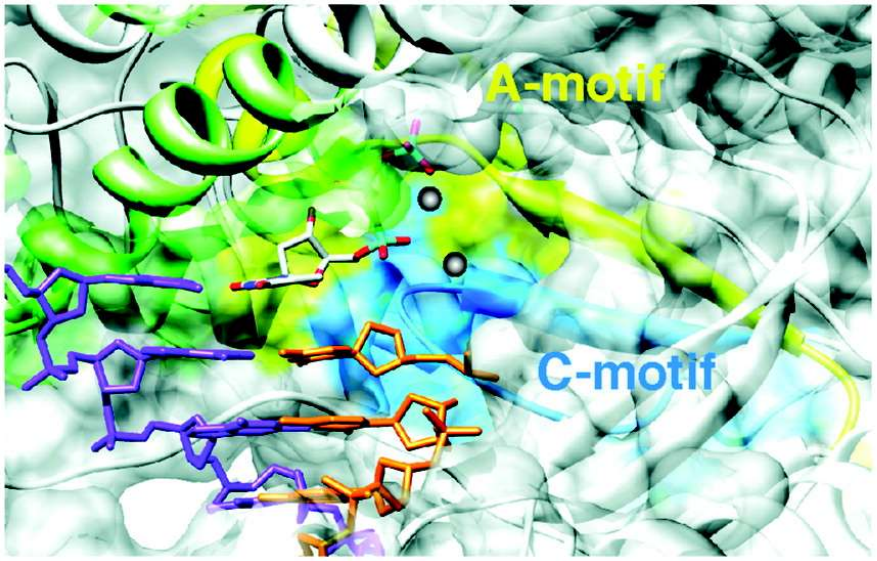
DNA Polymerase is evolutionarly optimized by mutagenesis to accept fluorescent Cy3 or Cy5 nucleotides instead of simple dCTP. Here, the polymerase replicates a short segment of its own encoding gene. Water/oil compartmentalization allows testing many independent mutations in paralell

Polymerases (Pol1 (left compartment)) that are capable of utilizing Cy5-dCTP are able to replicate, i.e., produce "offspring", while polymerases like Pol2 (right compartment) that are unable to utilize it disappear from the gene pool.

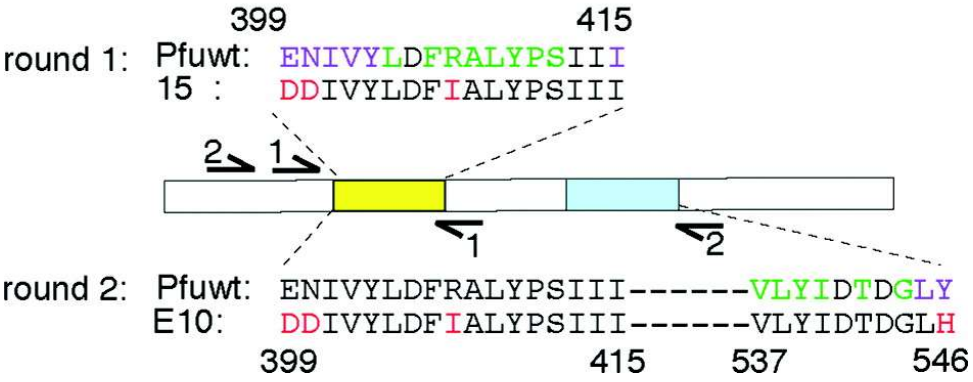
P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

CyDNA – synthesis and replication of highly fluorescently-labelled DNA

a)



b)



(a) Structural model of the active site of a polB-family polymerase (RB69, PDB: 1IG9). The polypeptide chain is shown as a ribbon diagram overlaid with a transparent surface model. Primer and template strands are shown in orange and purple, incoming nucleotide triphosphate in elemental colors, and gray spheres represent the two catalytic Mg^{2+} ions. Regions corresponding to the A-motif are colored yellow, and those for the C-motif cyan. Also shown is the conserved B-motif (green).

(b) For round 1 selection, diversity was focused on the A-motif and vicinity (399−415) comprising random mutation spike at the core of the A-motif (lime) as well as phylogenetic diversity in the adjacent sequences (purple). For round 2, successful clones from round 1 (e.g., 15) were diversified in the C-motif region (cyan), and selected for replication of A- and C-motif (399−546) yielding polymerase E10 (selected mutations in red)

P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

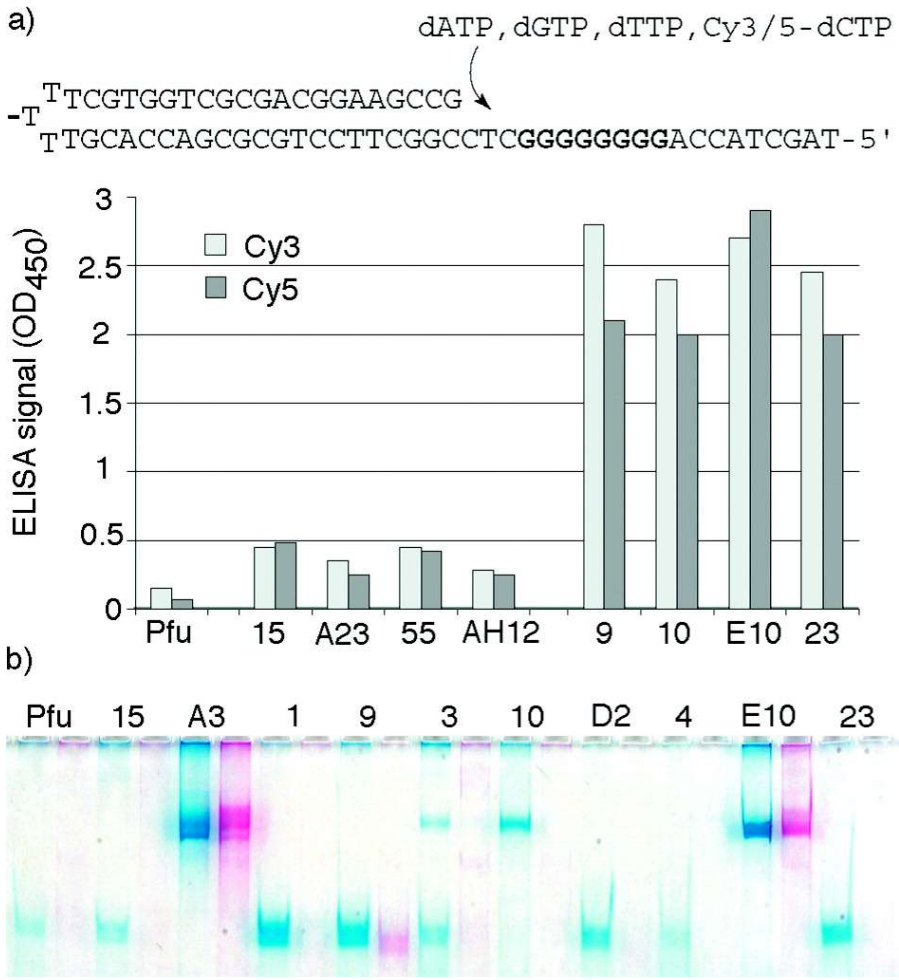
CyDNA – synthesis and replication of highly fluorescently-labelled DNA

380 selected variants from round 1 were screened by Polymerase-ELISA and ranked for their ability to incorporate 4 consecutive Cy5-dCTPs or Cy3-dCTPs. Polymerase-ELISA identified 4 mutant polymerases with significantly enhanced ability to incorporate either Cy3-dCTPs or Cy5-dCTPs compared with wild-type Pfuexo-:

- A23 (N400D, I401L, R407I),
- AH12 (E399D, N400G, I401L, V402A, R407I, Q572H),
- 55 (N400G, R407I), and in particular
- 15 (V337I, E399D, N400G, R407I).

a) activities of round 1 clones (15, A23, 55, AH12) and round 2 clones (9, 10, E10, 23). Clones were chosen i.a. on the their ability to incorporate both Cy3- and Cy5-dCTPs with comparable efficiency.

b) PAGE gel of a 0.4kb PCR amplification (70% GC) with complete replacement of dCTP with either Cy5- or Cy3-dCTP comparing selected clones to wild-type Pfuexo- (Pfu). Only E10 (and to a lesser extent A3) are able to perform PCR amplification with both dyes.



P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

CyDNA – synthesis and replication of highly fluorescently-labelled DNA

The polymerase fidelity in the selected mutant „E10” was not significantly compromised as compared to the starting Pfu DNA polymerase (*Pyrococcus furiosus*)

Fulvio314

Pyrococcus furiosus – an extremophilic *Archaeon* from marine sediments
Optimal life temperature 100^{0}C

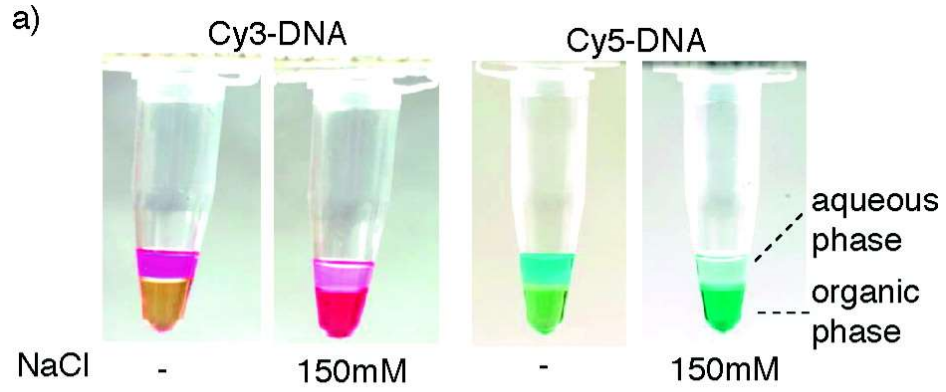
Table 1. Polymerase Fidelity

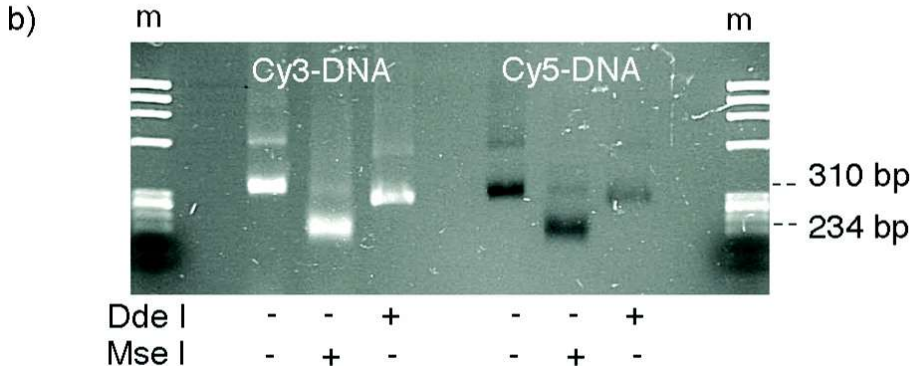
polymerase	PCR substrates	mutations/kb	mutation rate/bp/doubling[a]
Pfuexo-	dNTPs	1.1	4.4×10^{-5}
			$(4.7 \times 10^{-5})^b$
E10	dNTPs	0.4	1.6×10^{-5}
		(2.6 without additives[c])	$(1.04 \times 10^{-4})^c$
Pfuexo-	dNTPs[b]	2.7	6.0×10^{-5}
E10	dATP, dGTP, dTTP, Cy3-dCTP	4.3	9.6×10^{-5}
E10	dATP, dGTP, dTTP, Cy5-dCTP	4.9	1.1×10^{-4}

[a] Corrected for the number of doublings (PCR cycles). [b] As determined by a lacZ reversion assay.[44] [c] In the absence of additives (1% formamide, 10% glycerol, 10 μg/mL RNase, 1 mM DTT).

P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

CyDNA – synthesis and replication of highly fluorescently-labelled DNA



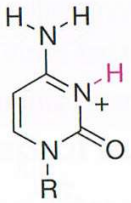
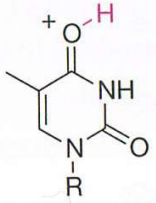
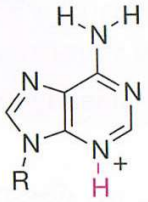
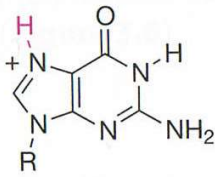


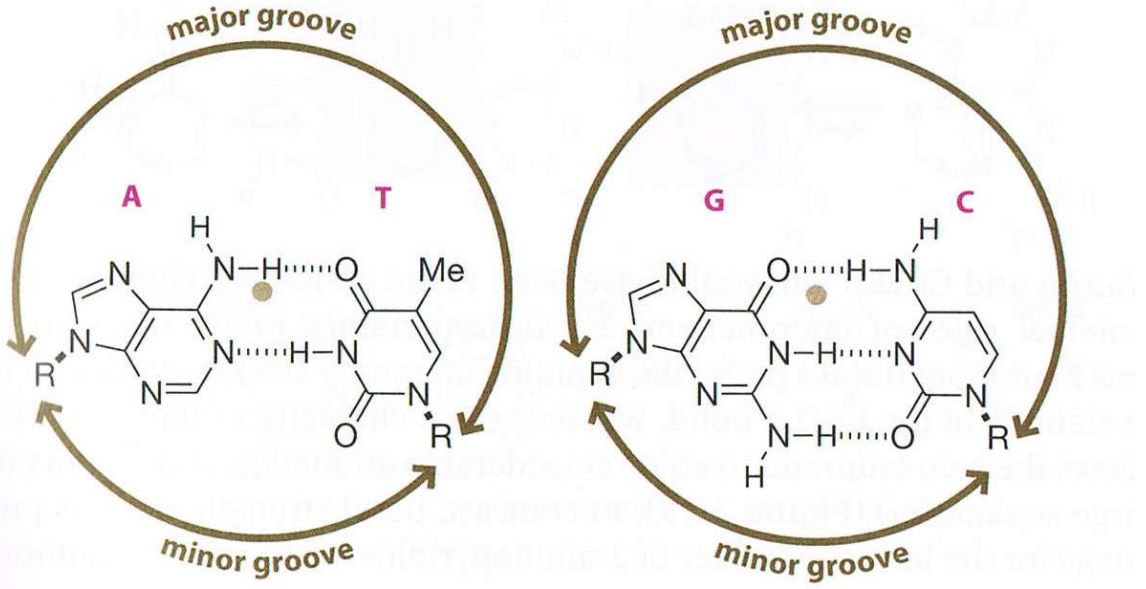
(a) Organic phase partitioning of CyDNA is shown for Cy3-DNA (left) and Cy5-DNA (right). Essentially 100% partitoning occurs in the presence of 150 mM NaCl (the yellow color of the phenol phase is due to addition of 8-hydroxyquinoline to prevent oxidation).

(b) Agarose gel electrophoresis of CyDNA restriction digests. Restriction endonucleases are sensitive probes of noncanonical DNA conformations such as those which occur under torsional strain. Both Cy3- and Cy5-DNA are resistant to cleavage by the restriction endonuclease DdeI (C'TNAG) but are cut by MseI (T'TAA). This indicates that at least the local regions of AT-sequence in Cy-DNA adopt a canonical B-form conformation.

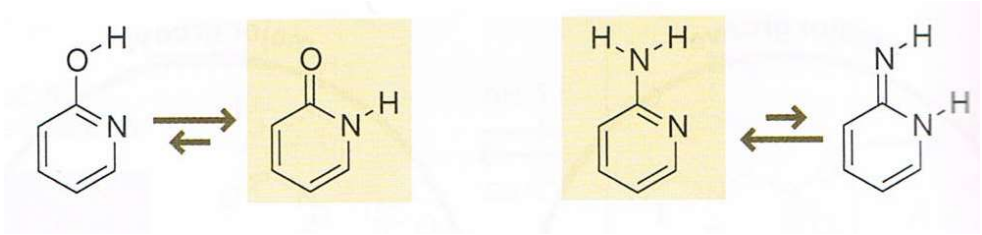
P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

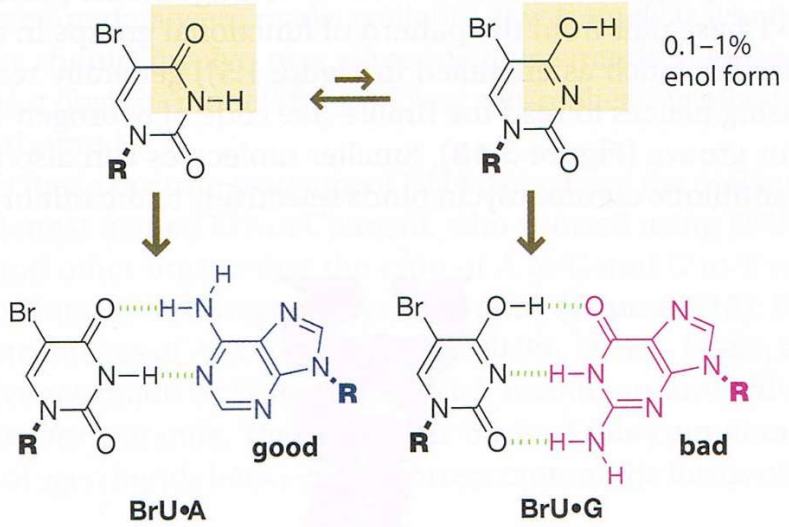
Why are A, C, G and T the letters of genetic alphabet.

	common amine base	cytosine	thymine	adenine	guanine
					
pK_a	10.8	4.2	0.5	4.2	3.3
relative basicity of conj. base	4,000,000	1	0.0002	1	0.1



Tautomeric instability in non-canonical nucleobases





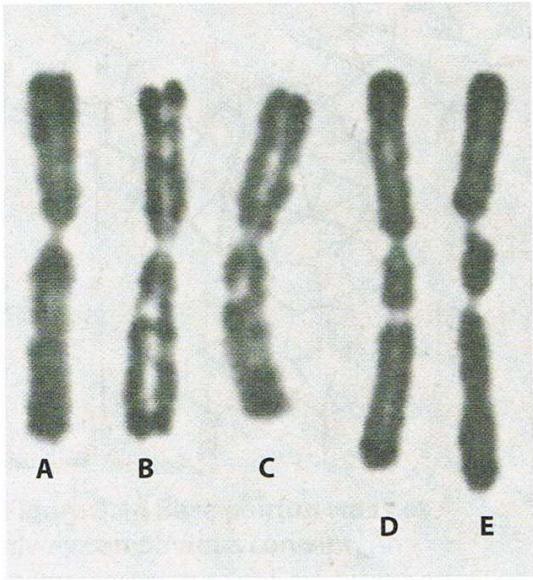
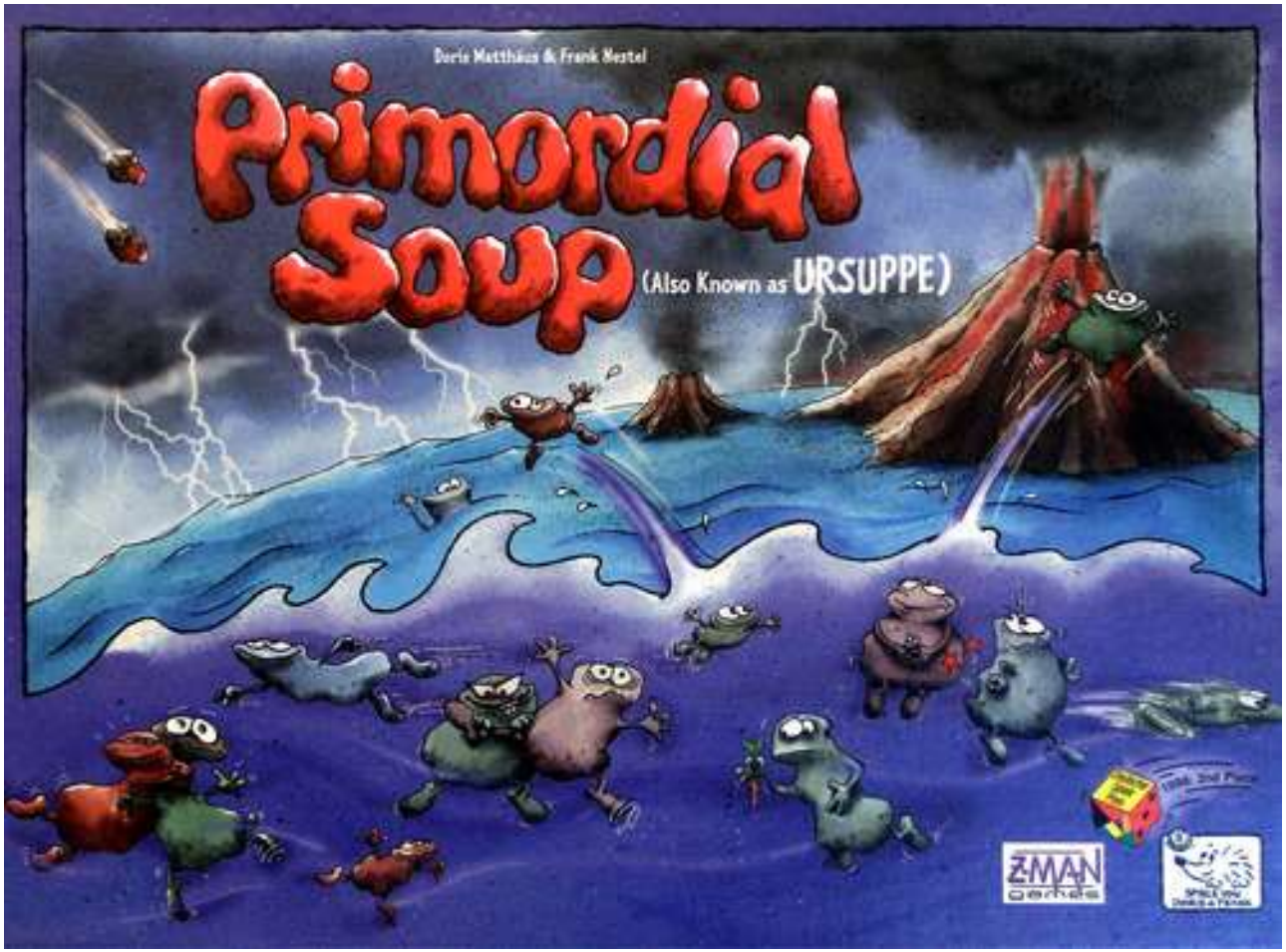


Figure 3.20 Chromosome 1 from hamster cells exposed to bromodeoxyuridine. (A) Normal chromosome. (B–E) Aberrant chromosomes. (From T.C. Hsu and C.E. Somers, *Proc. Natl. Acad. Sci. USA* 47: 396–403, 1961. With permission from the MD Anderson Cancer Center.)

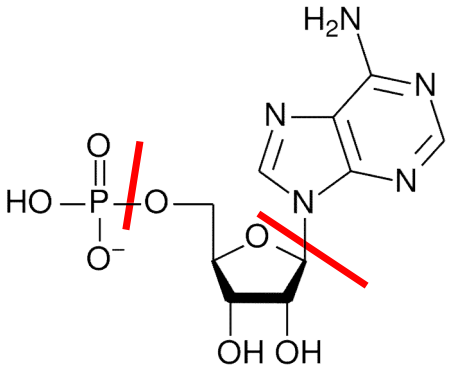
Prebiotic synthesis of nucleotides

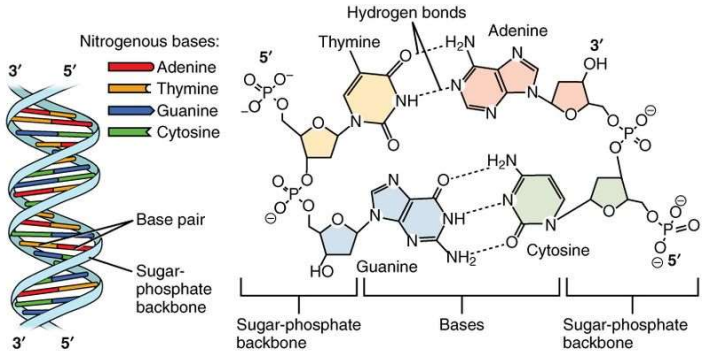


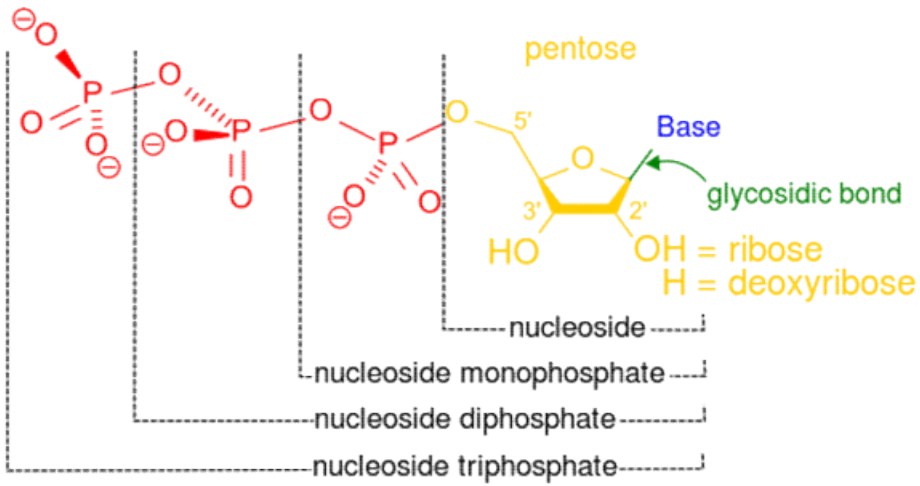
State of the art

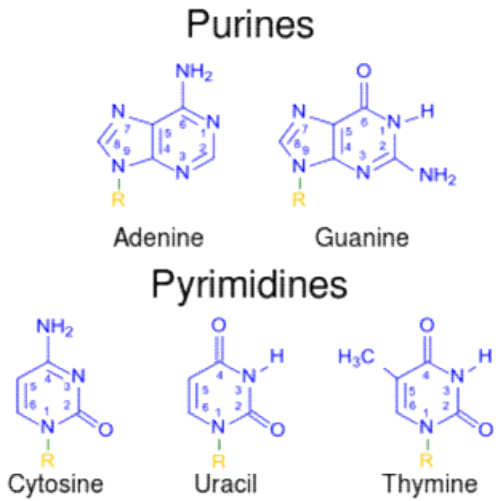
Nucleotides - components

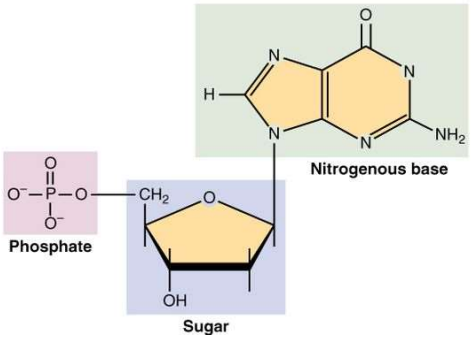
RNA – most likely evolutionarily older („RNA World") than DNA → prebiotic origin of ribose + A, C, G, and U nucleobases





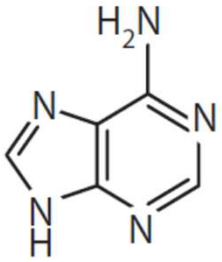




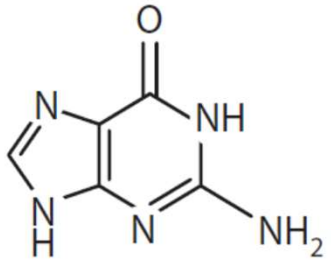


Prebiotic synthesis of nucleobases

Purines

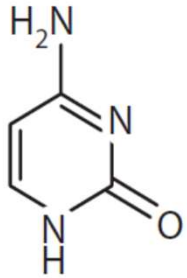


Adenine

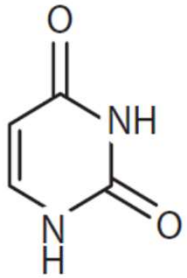


Guanine

Pyrimidines

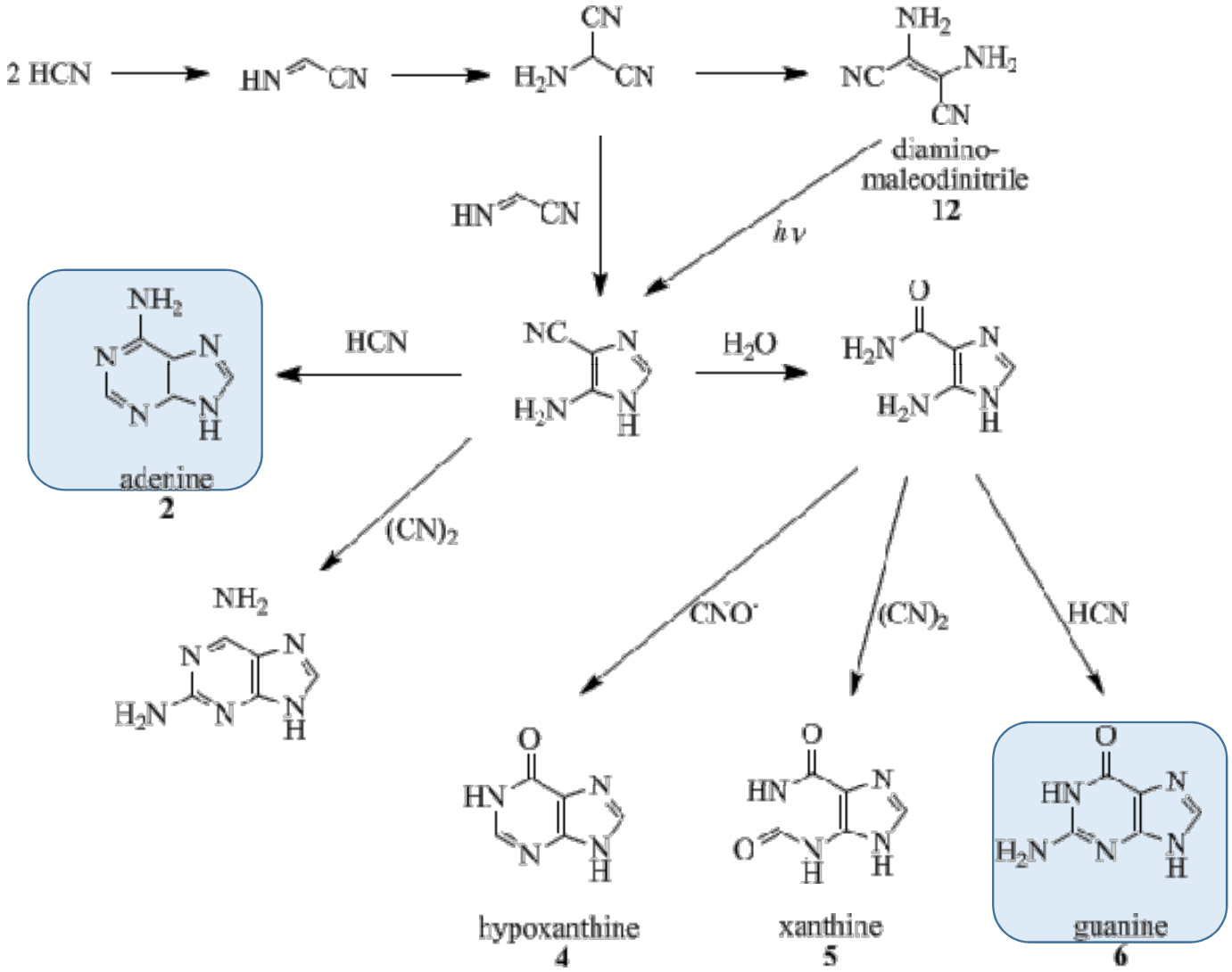


Cytosine

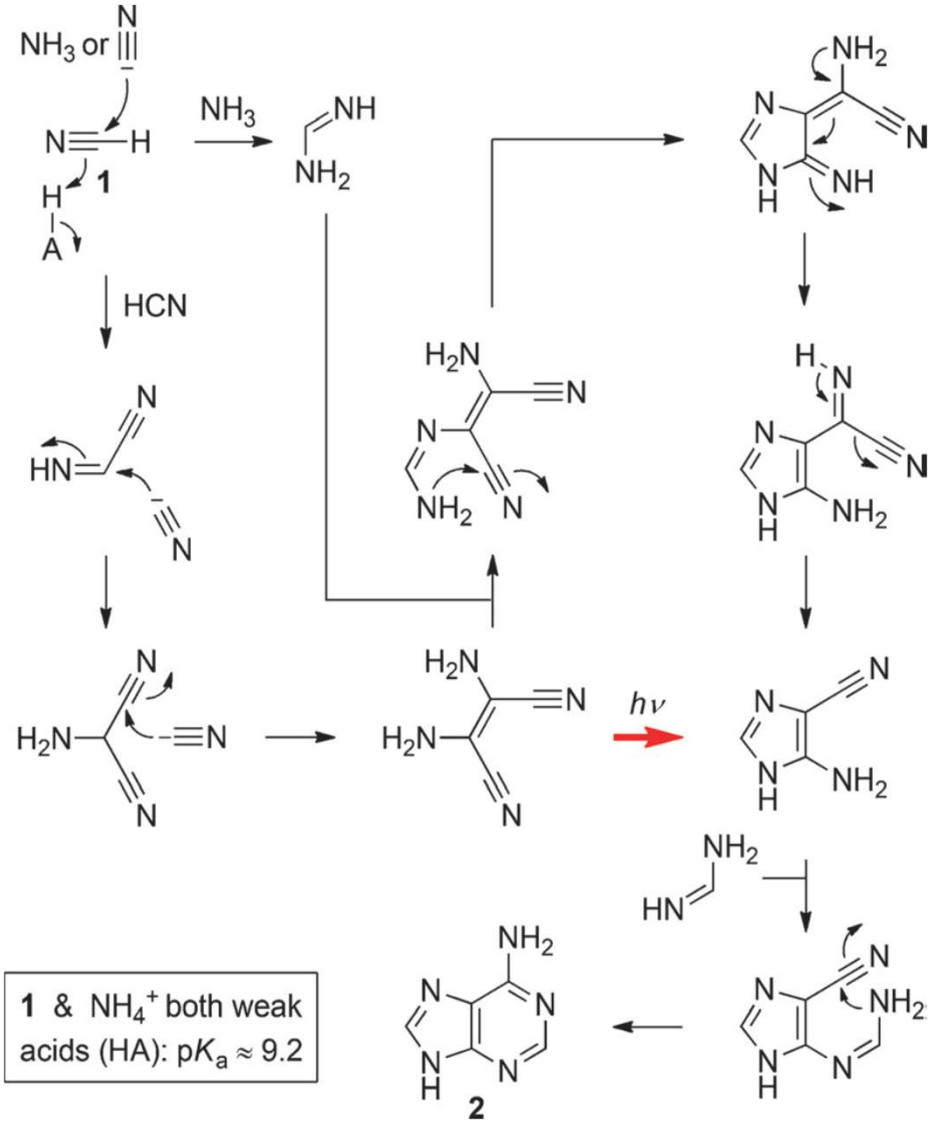


Uracil

Prebiotic synthesis of purines



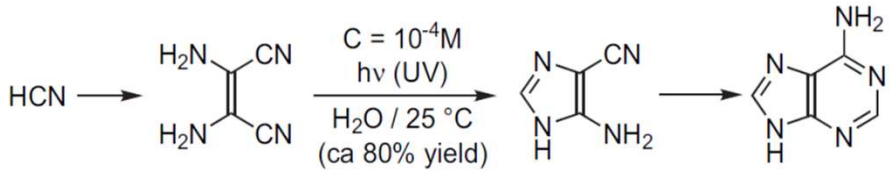
Prebiotic synthesis of adenine



1960 - Oró's synthesis of adenine **2** from hydrogen cyanide **1** and ammonia (general acid–base catalysis, presumed to operate in most steps, is only shown once).
Heating ammonium cyanide at 70^{0}C for a few days
→ 0.5% adenine
Heating HCN with liquid ammonia in a sealed tube → 20% adenine

The photochemical shortcut discovered by Ferris and Orgel is shown by the red arrow.
Optimized yields – up to 20% for adenine, 3% for guanine

Eutectic freezing (-20^{0}C) increases the yield of DAMN formation by concentrating HCN between pure ice crystals

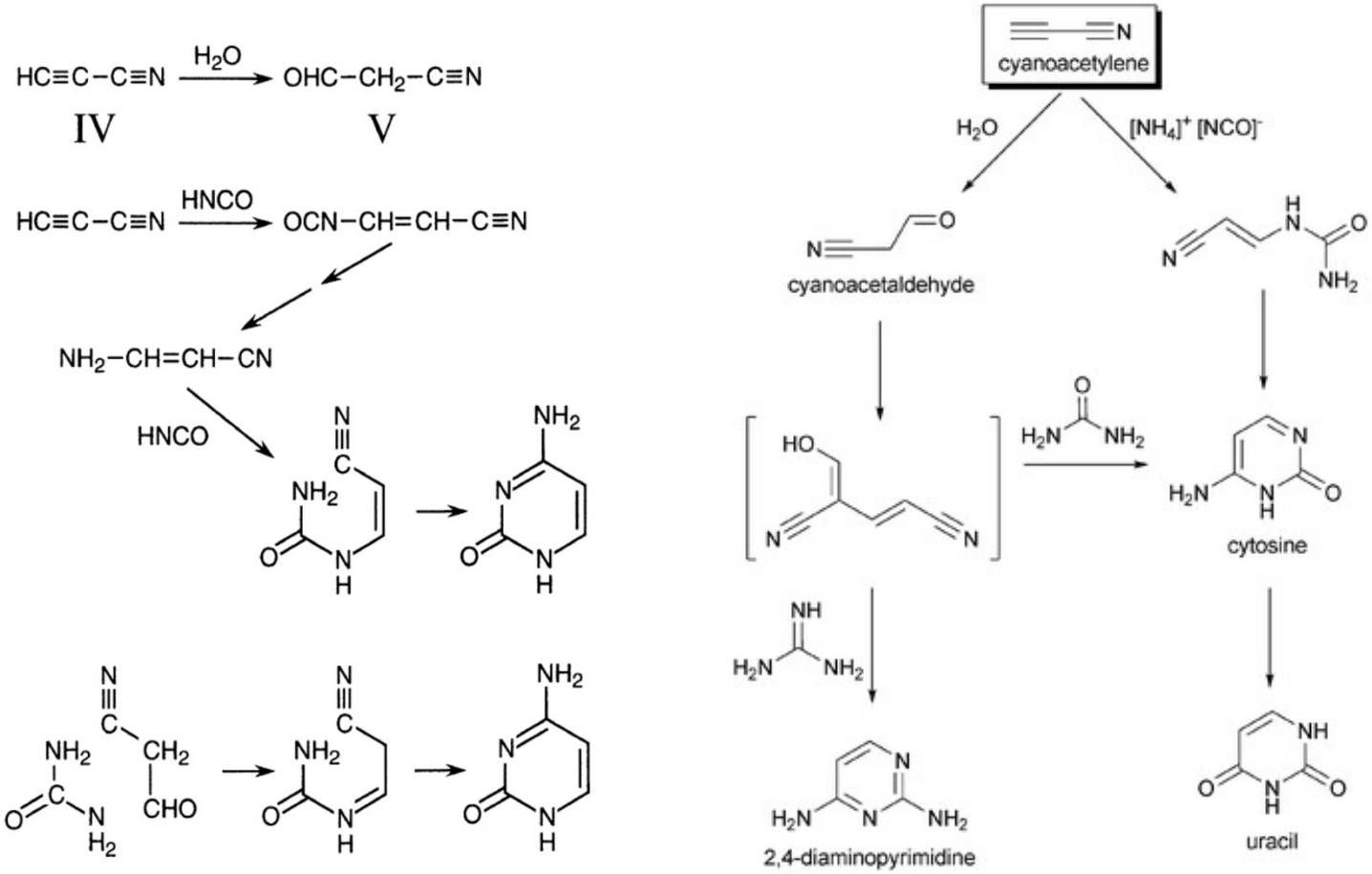


J. Oro *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407.

J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1966**, *88*, 1074

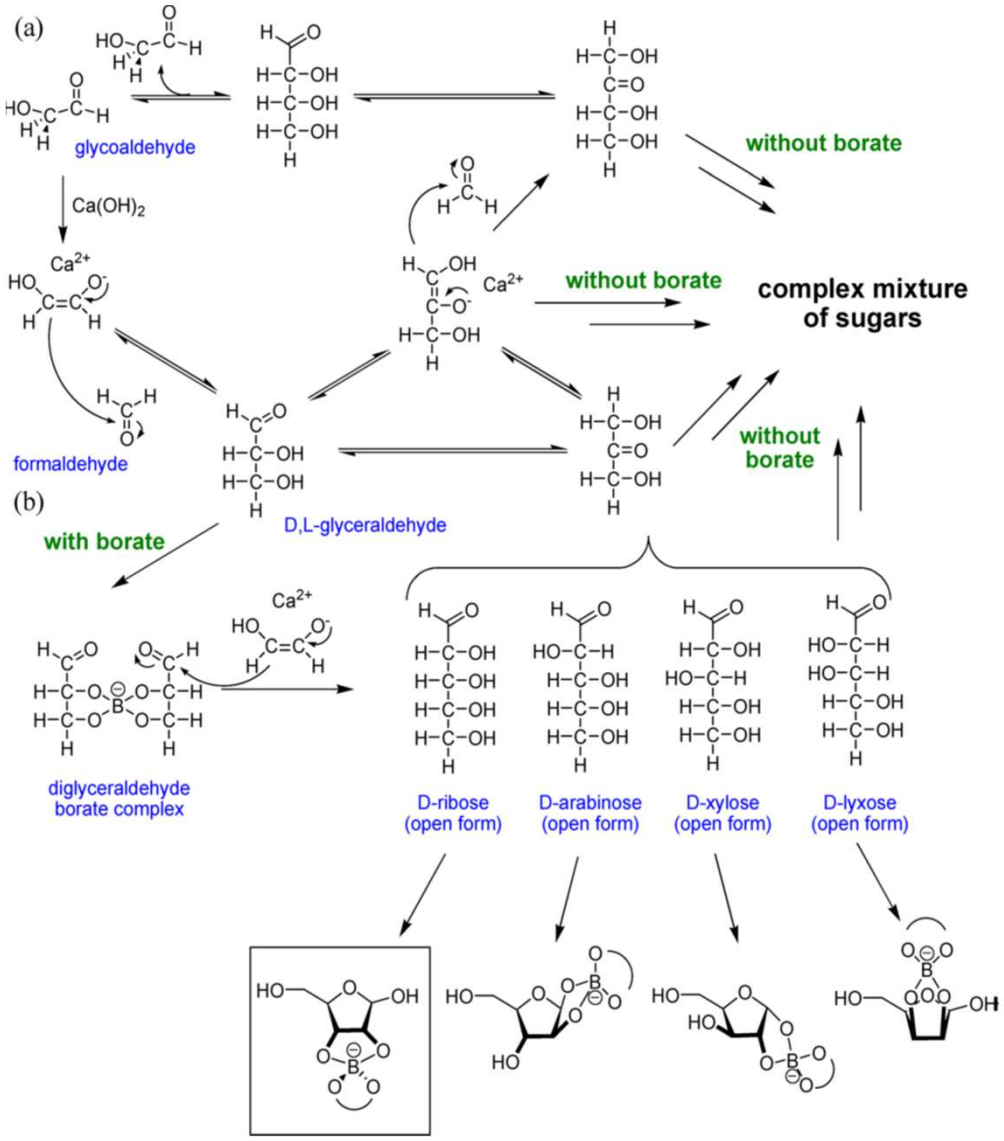
Prebiotic synthesis of pyrimidines

Cyanoacetylene is a major product of electric discharges in the mixture of nitrogen and methane



Cyanoacetylene incubated with saturated solution of urea yields up to 50% cytosine.
Other methods typically yield up to 5% cytosine.It is further converted to uracil by hydrolysis.

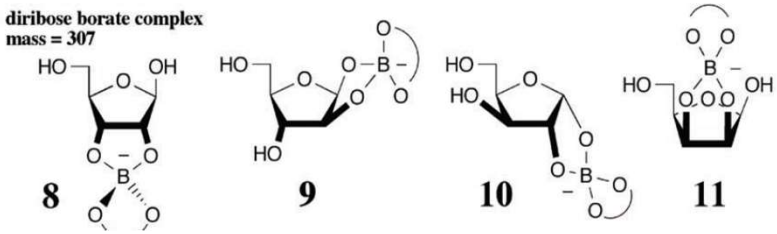
Formose reaction in presence of borates



Pentose formation in the presence of borate

With borate (left)
Without borate (right)
Colemanite (background)

By NMR, the ribose borate complex **8** has the structure shown; cyclic structures for other pentoses are speculative.

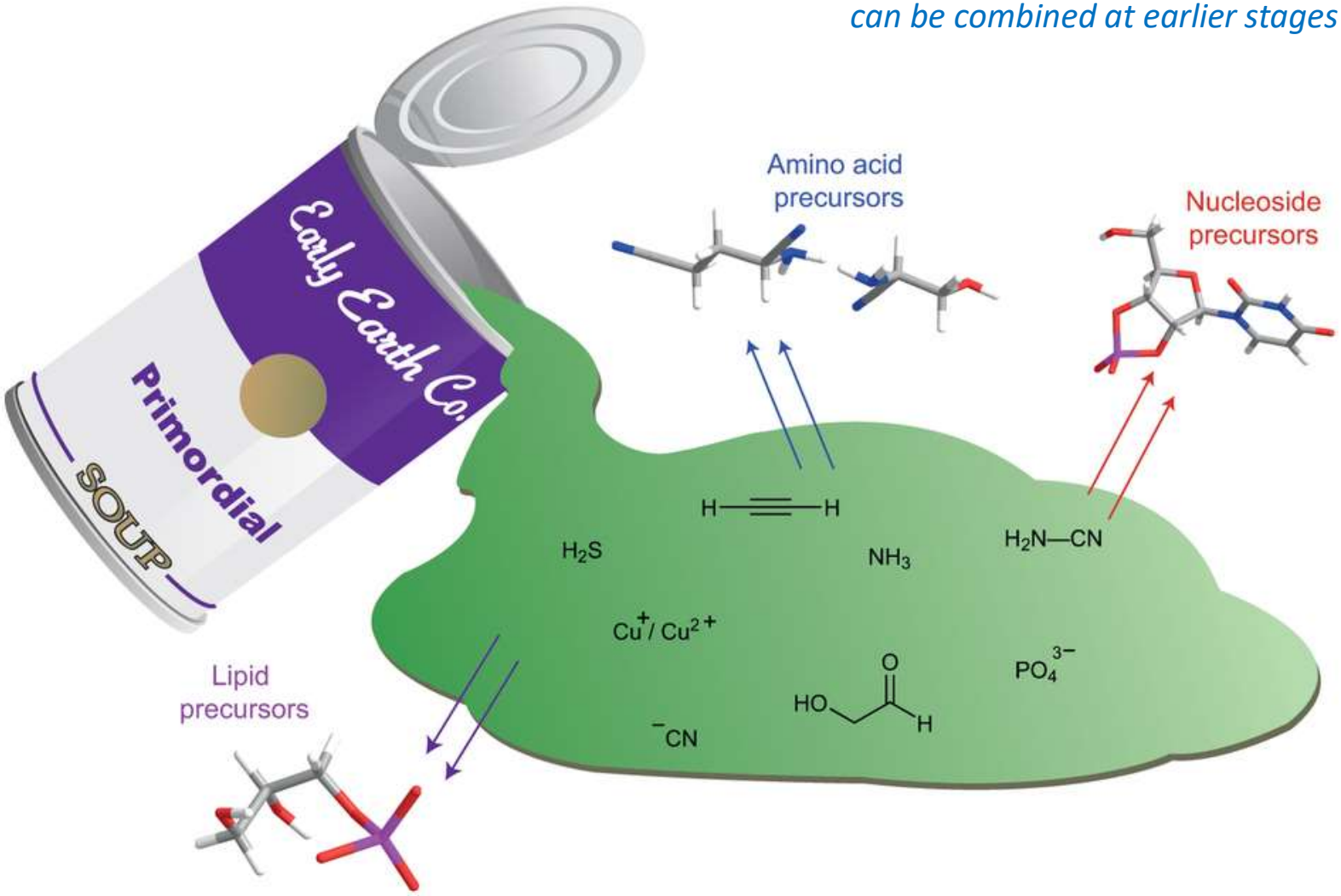


A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner
Science **2004**, *303*, 196

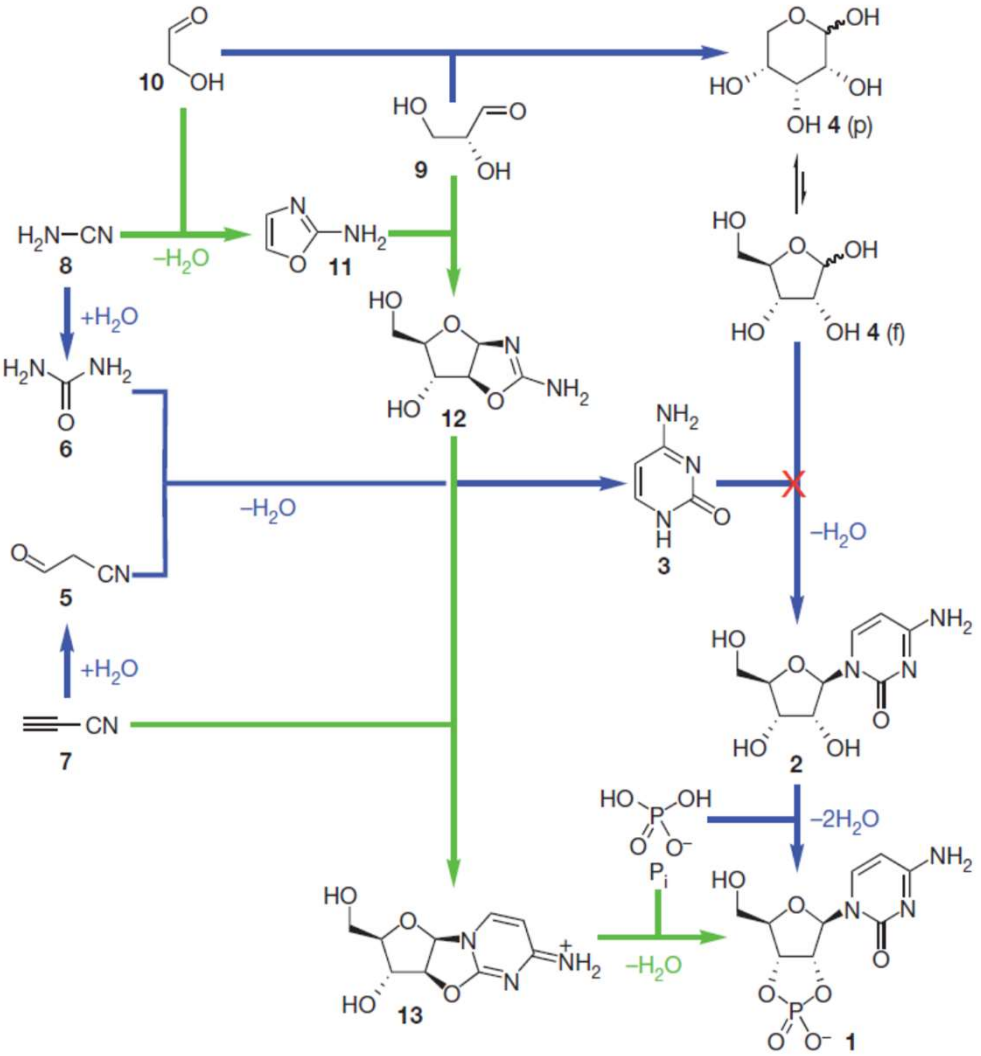
Escosura

Cyanosulfidic chemistry

The aldol chemistry of sugars and cyanide chemistry of nucleobases can be combined at earlier stages than glycosylation.

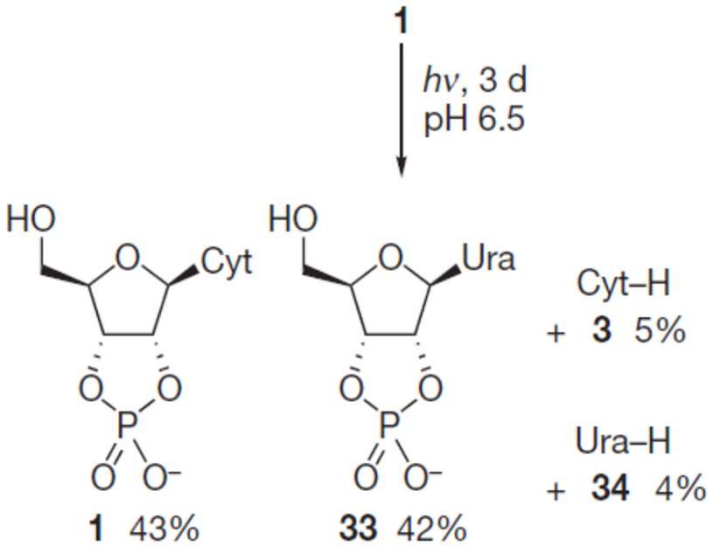


Cyanosulfidic chemistry



M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242

Cyanosulfidic chemistry



Photochemistry of *beta*-ribocytidine-2',3'-cyclic phosphate **1**. Under conditions of irradiation that destroy most other pyrimidine nucleosides and nucleotides, **1** undergoes partial hydrolysis and slight nucleobase loss. Ura, N1-linked uracil; Cyt–H, cytosine; Ura–H, uracil.

M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242

J. D. Sutherland, *Angew. Chem. Int. Ed.* **2016**, *55*, 104-121.

B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem*. **2015**, *7*, 301–307.

J. D. Sutherland, *et al. Nat. Chem*. **2013**, *5*, 383–389.