

# Synthetic life

*(continuation of „The molecular origins of life” SoSe 2021)*



[NaturalNews.com](http://NaturalNews.com)

WiSe 2021/22

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);

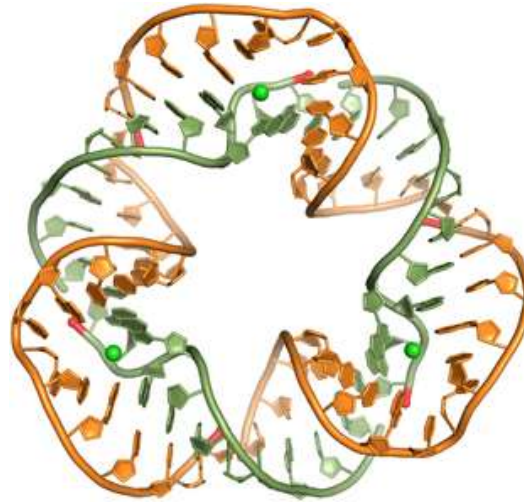
biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

**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**

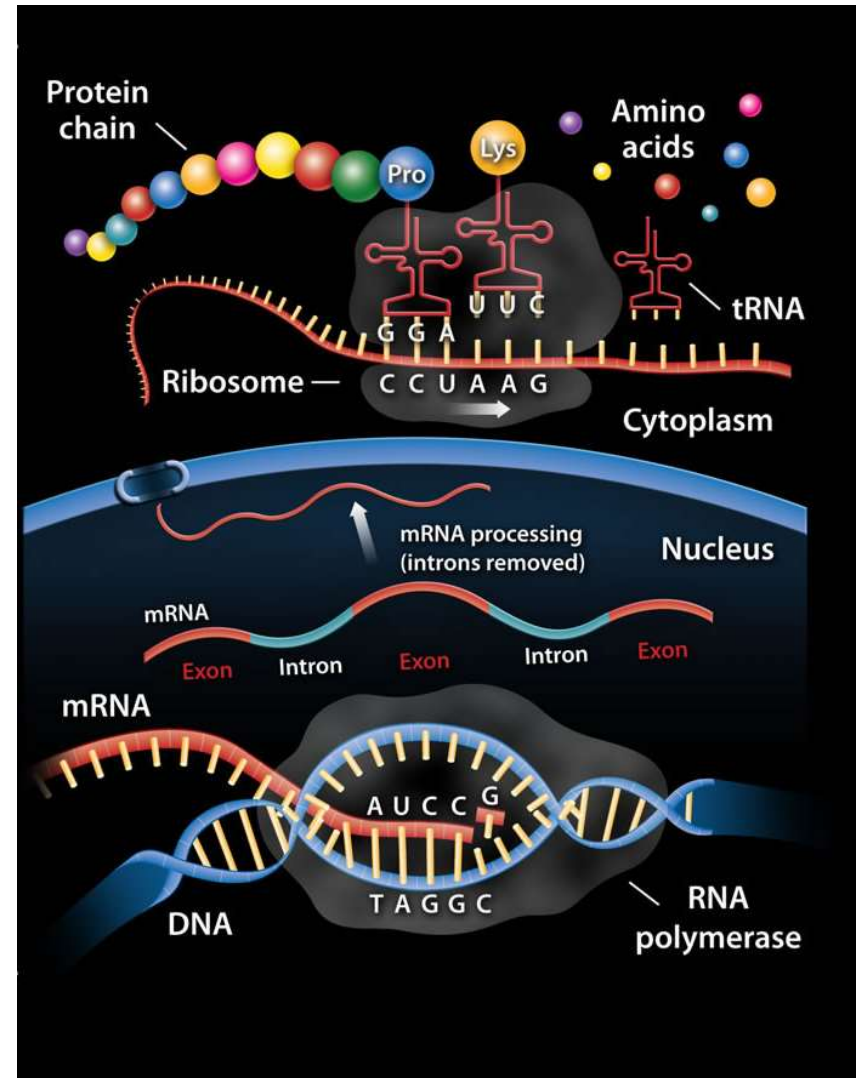
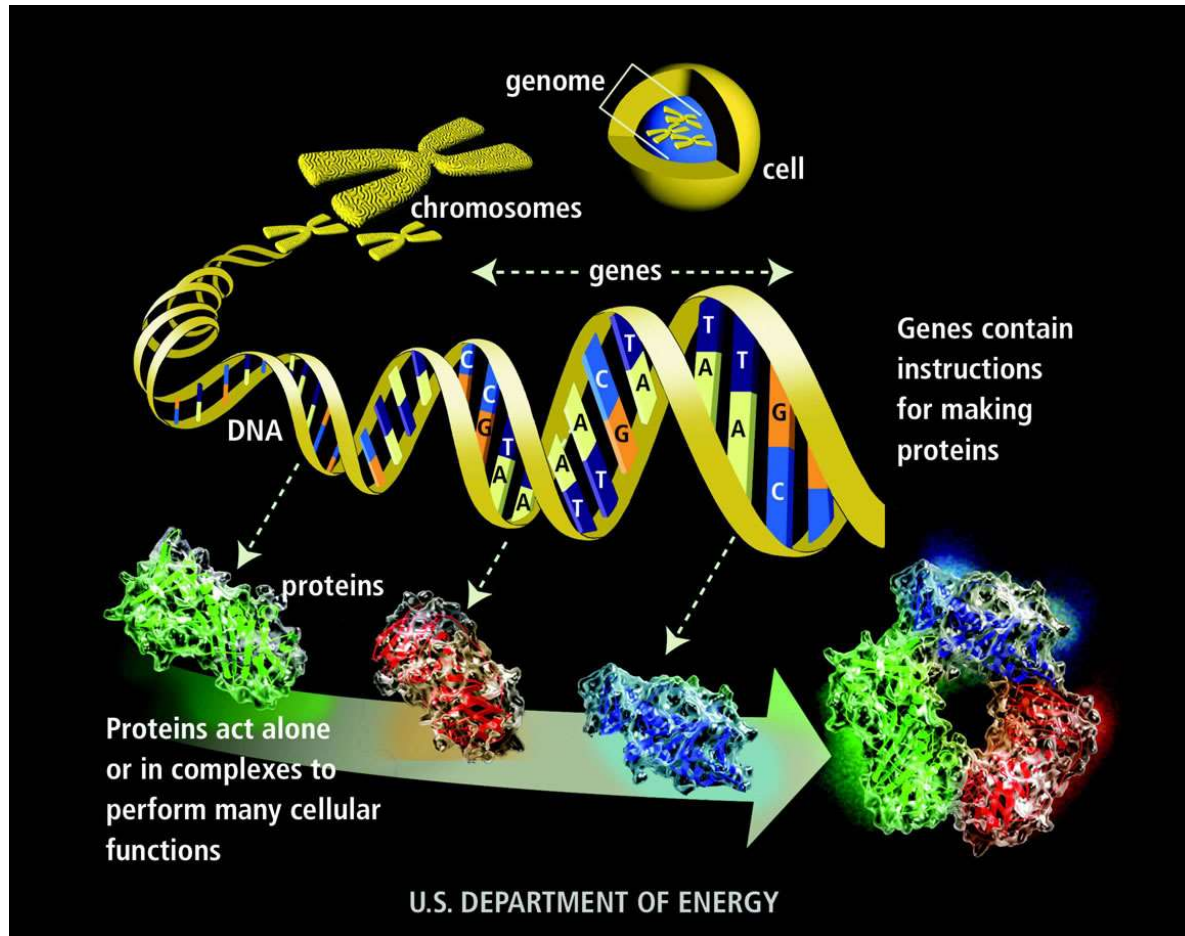
# CHAPTER 1



## ***OLIGONUCLEOTIDES***

*The less common side of RNA*

# From DNA to proteins

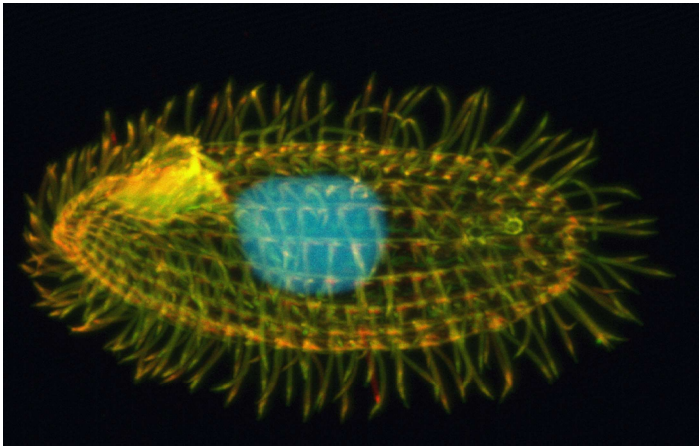


# Ribozymes

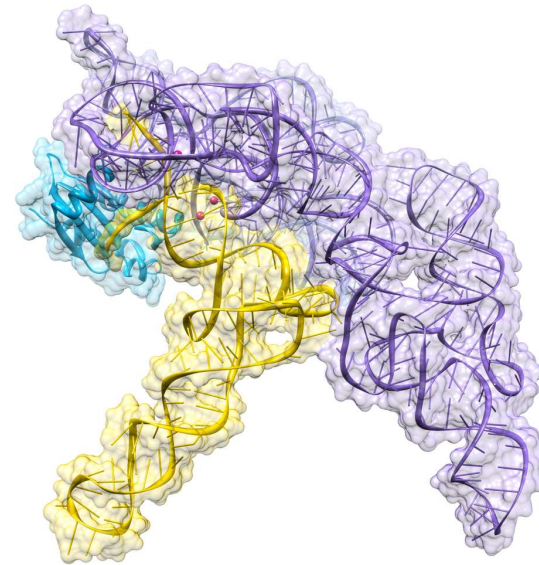
## Ribozymes – Ribonucleic acid enzymes

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.

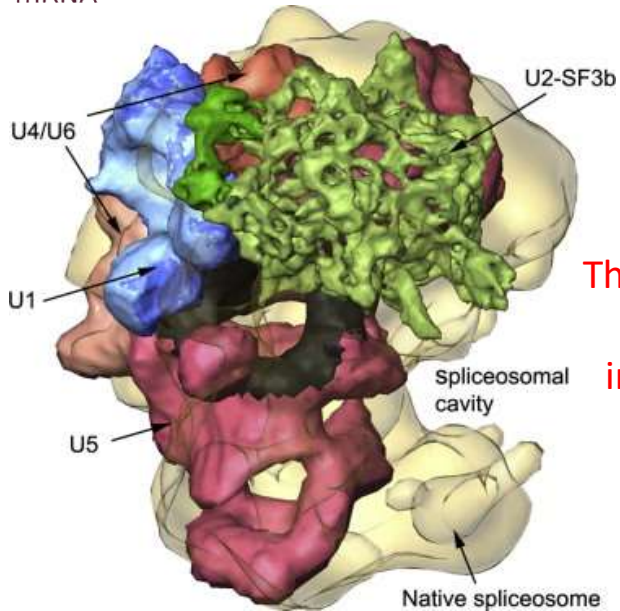
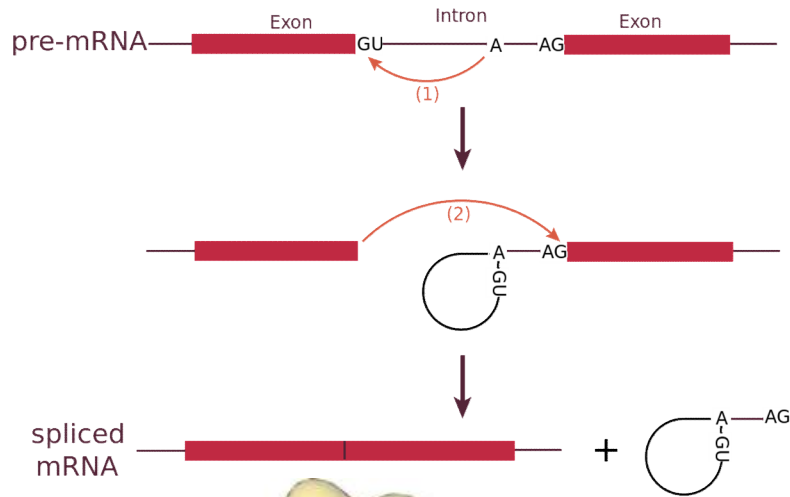


*Tetrahymena thermophila*



*Bacterial RNase P*

# RNA splicing



The spliceosome carries out the removal of introns as RNA lariats

Spliceosome – a complex of ribonucleoproteins

## 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**.

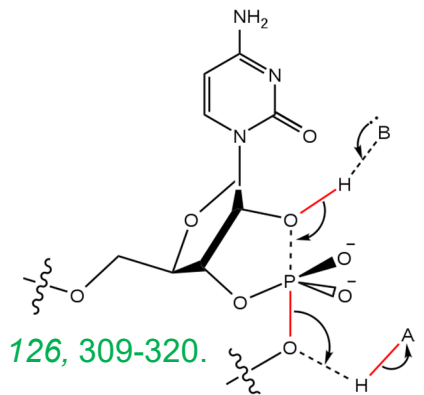
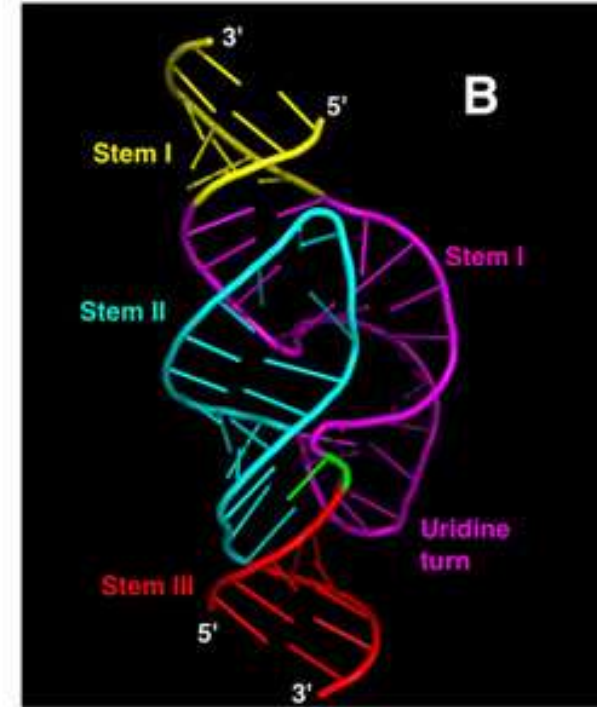
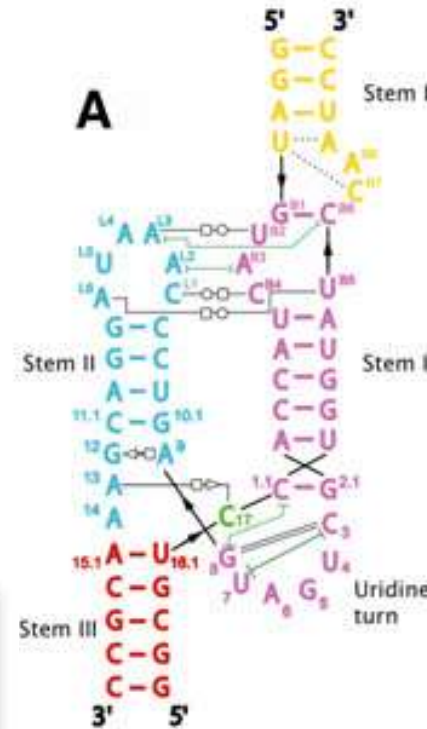
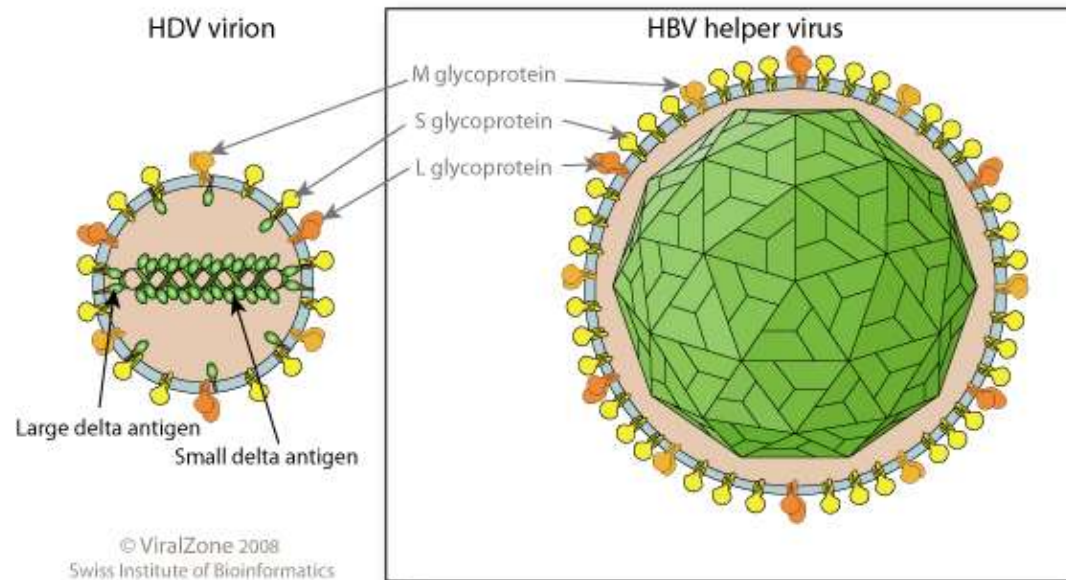
First known example of a **ribozyme** – ribonucleic acid-composed enzyme analogue.

# 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**.

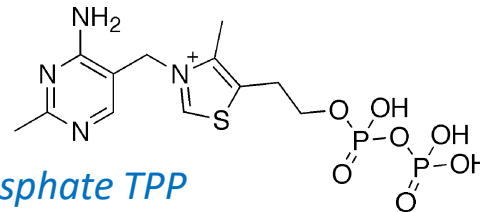
**Riboswitches** - naturally occurring regulatory segments of mRNA that bind small molecules specifically. The binding results in a change in production of the proteins encoded by the mRNA

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

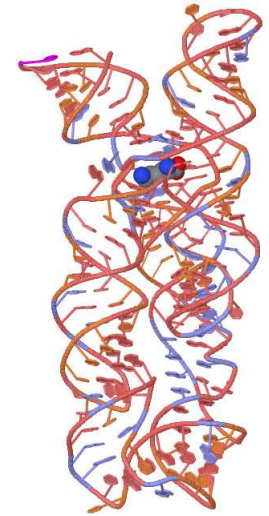
Most known **riboswitches** occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in archaea, plants and certain fungi.

**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

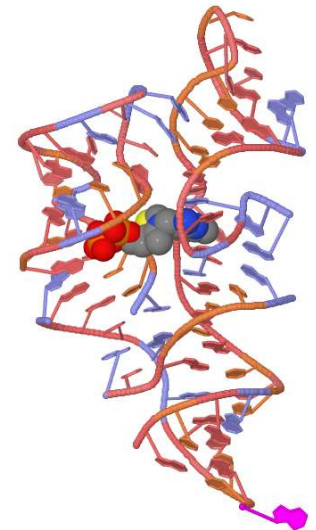
The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



*Thiamine pyrophosphate TPP*



*The lysine riboswitch*



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

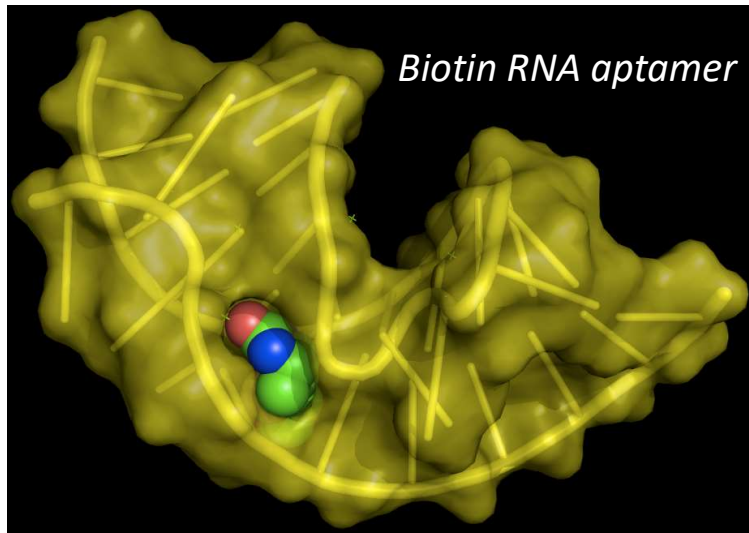


# Aptamers

**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.

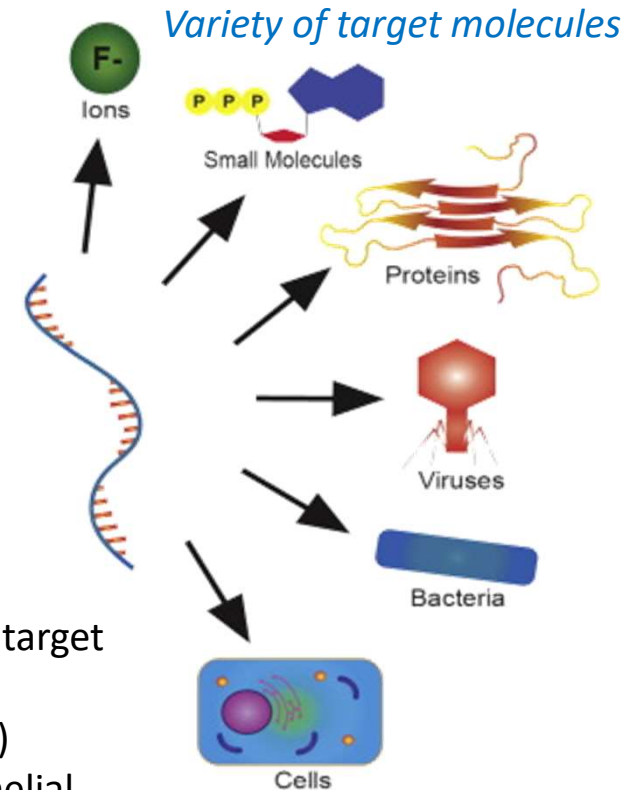


*Fdardel*

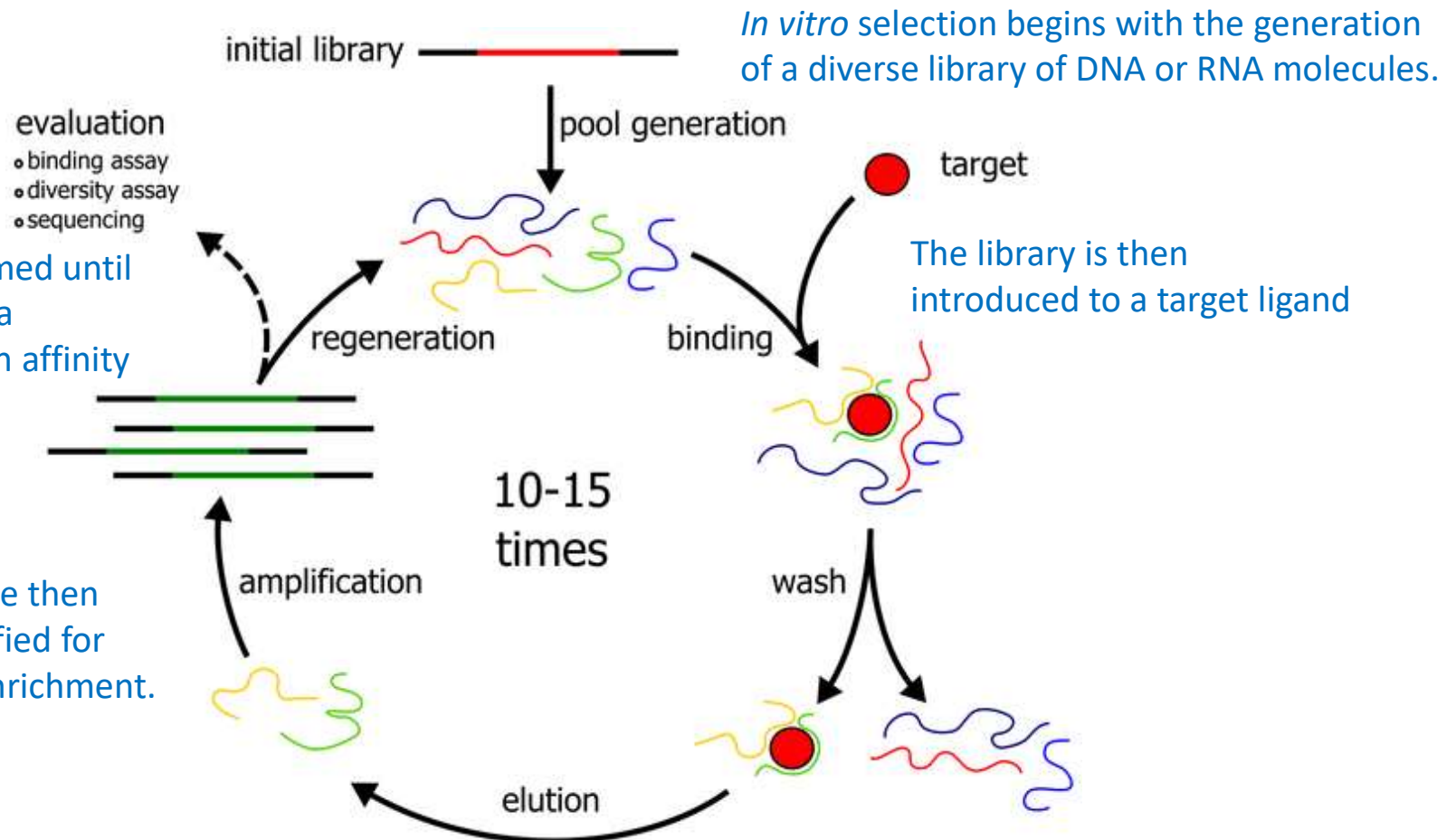
**Aptamers** were evolved for a variety of target ligands:

- 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



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

The library is then introduced to a target ligand

evaluation  
• binding assay  
• diversity assay  
• sequencing

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

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

10-15 times

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

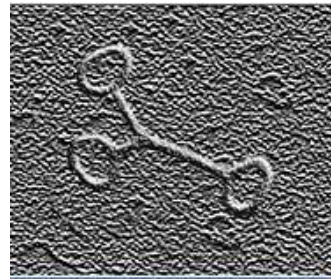
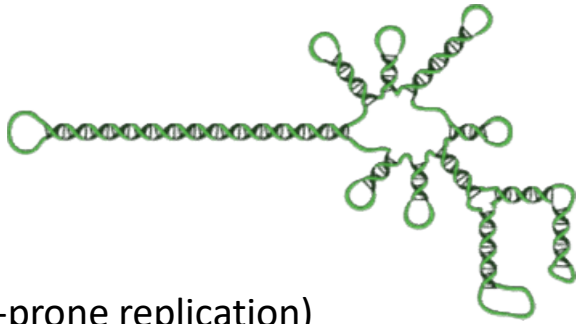
- 1990 – Gold *et al.* – selection of RNA ligands against T4 DNA polymerase
- 1990 – J. Szostak *et al.* – selecting RNA ligands towards organic dyes

# 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.

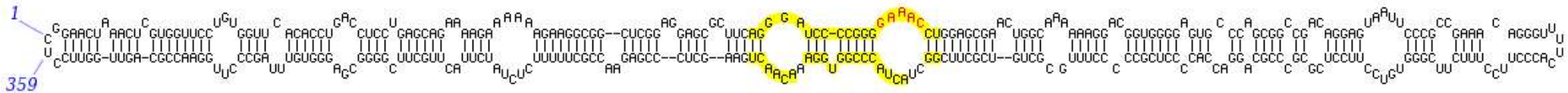
## Viroid properties:

- 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.



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings.

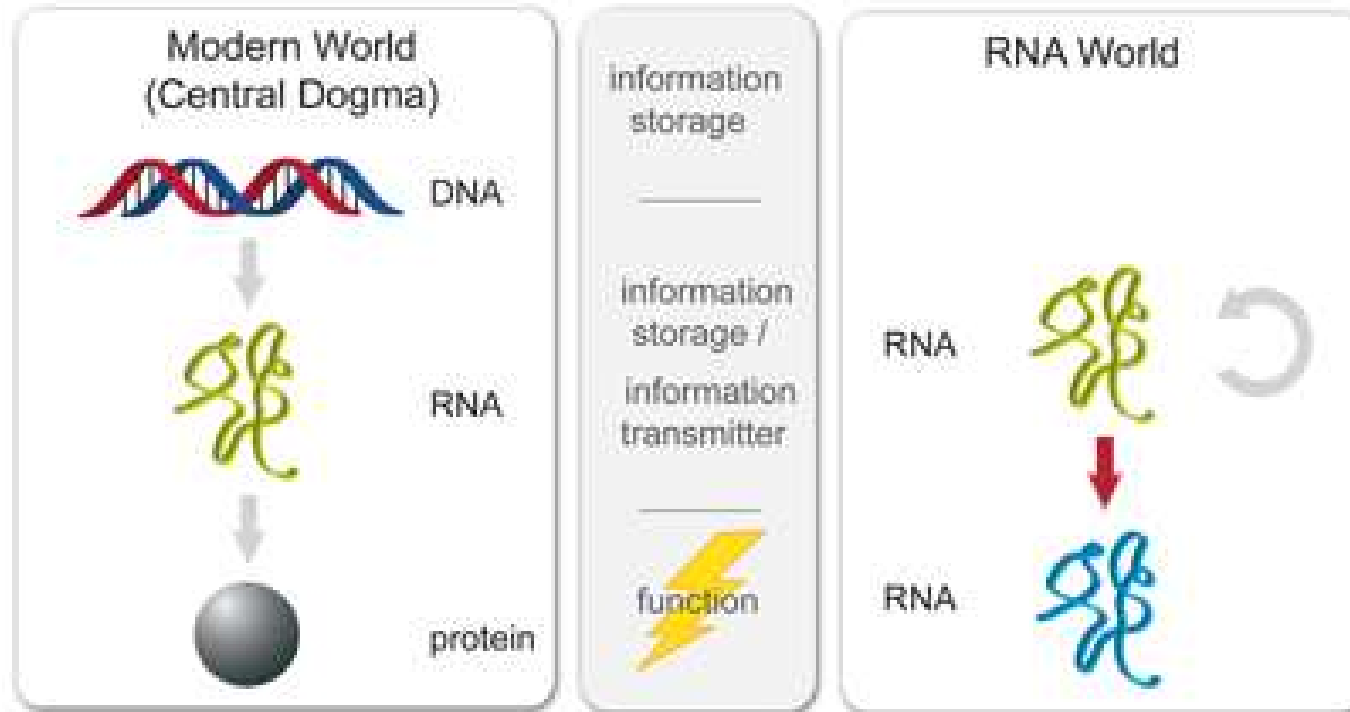
*PSTVd-infected potatoes (right)*



*Putative secondary structure of the PSTVd viroid*

**No virion reported. Viroids do not encode for proteins**

## The RNA world



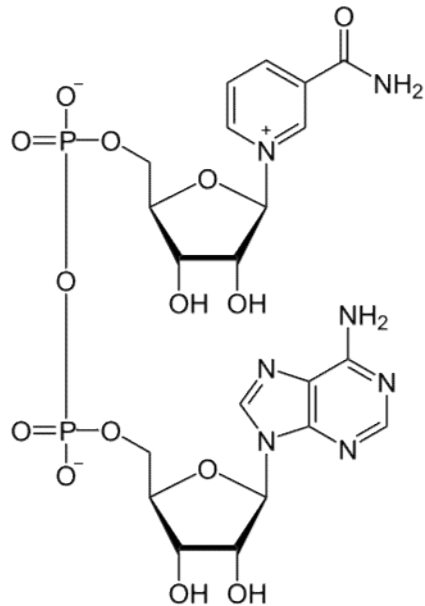
Conceptual idea that there was a period in the early history of life on Earth when RNA (or its structurally simplified analogue) carried out most of the information processing and metabolic transformations needed for biology to emerge from chemistry

## The RNA world

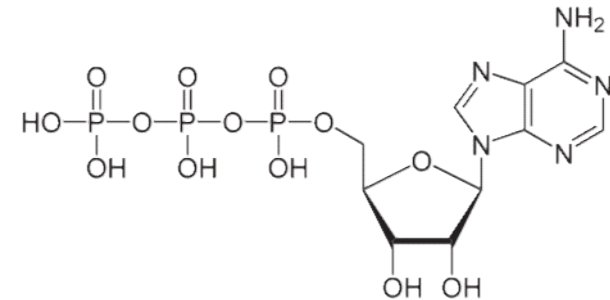
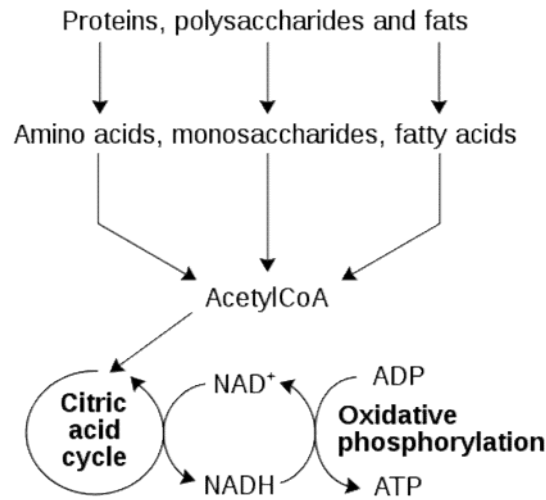
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

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

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



**Nicotinamide adenine dinucleotide (NAD<sup>+</sup>)**

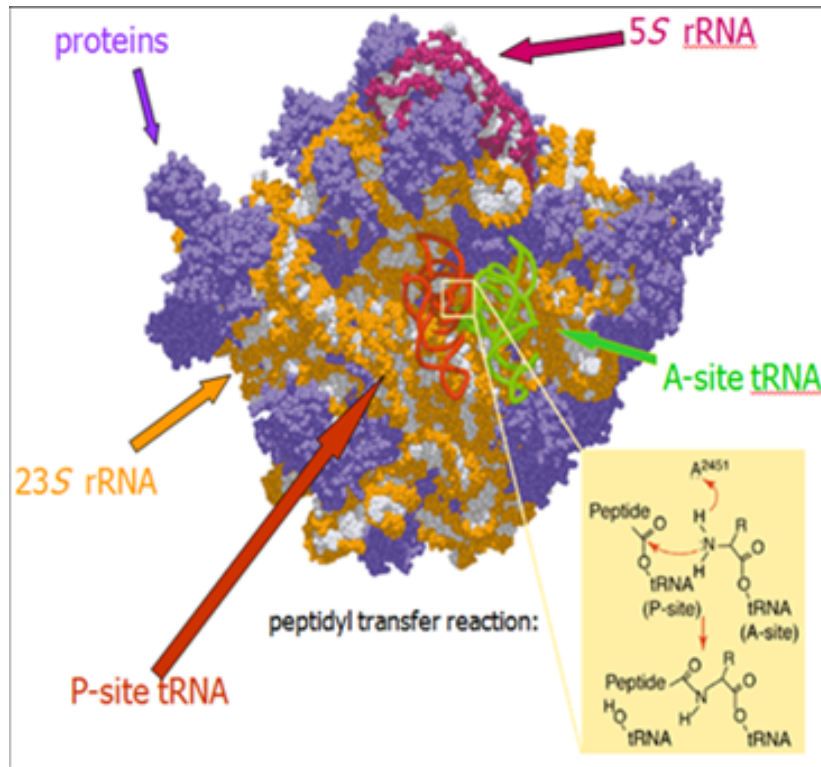


**Adenosine triphosphate (ATP)**

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

# 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

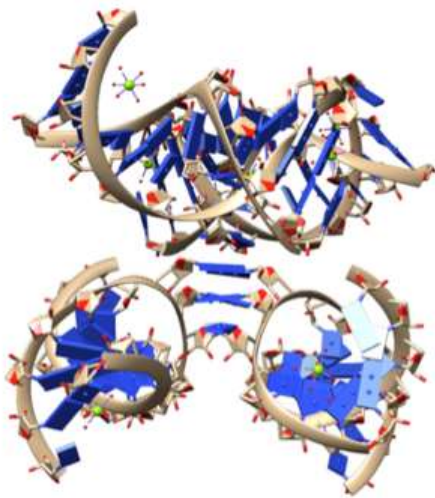
No protein is present within 18 Angstroms from the active site → **proteins** play a structural role, but **DO NOT CATALYZE THE ACYL TRANSFER PROCESS**

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

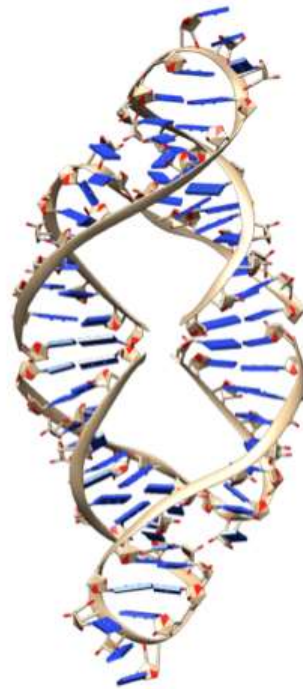
# Ribozymes

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

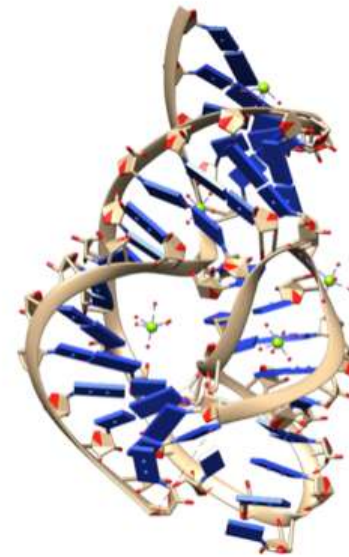
Ribozymes participate in a variety of RNA processing reactions, including RNA splicing, viral replication, and transfer RNA biosynthesis. **Examples** of ribozymes include the **hammerhead ribozyme**, the **VS ribozyme**, **Leadzyme** and the **hairpin ribozyme**.



**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.

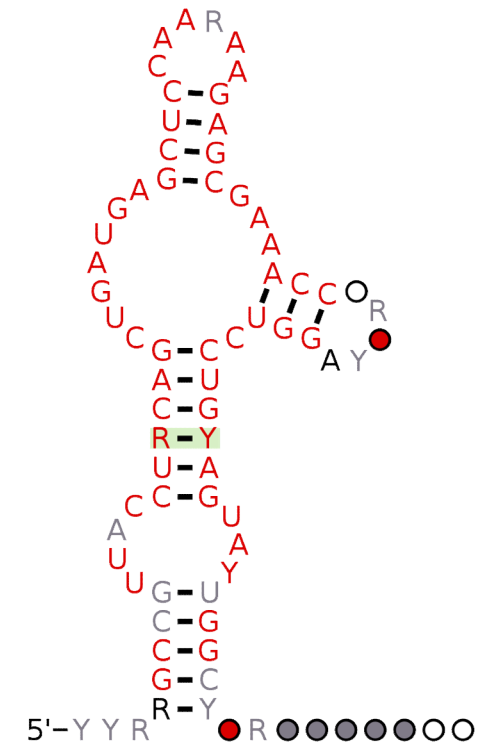
In its natural state, a hammerhead RNA motif is a self-cleaving single strand of RNA (not catalytic, it is consumed by the reaction → no multiple turnovers)

*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

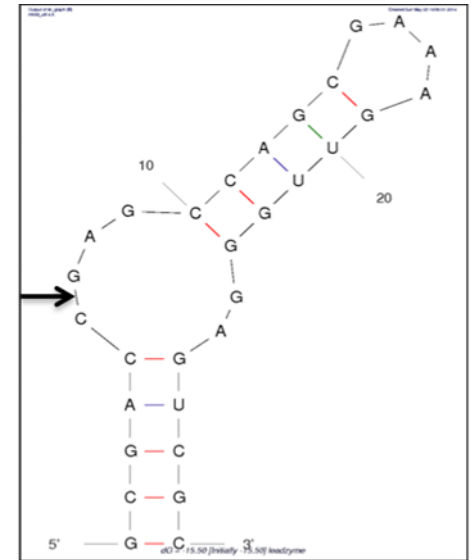


*Predicted secondary structure and sequence conservation of the HH9 ribozyme found conserved from lizard to human genomes*

## 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

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.



Anarkalimahmood

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.



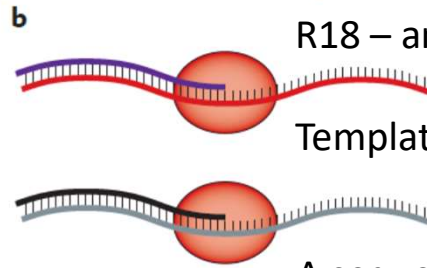
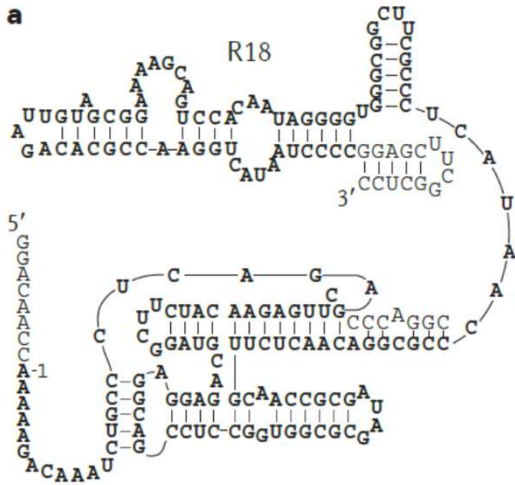
David Schechner

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

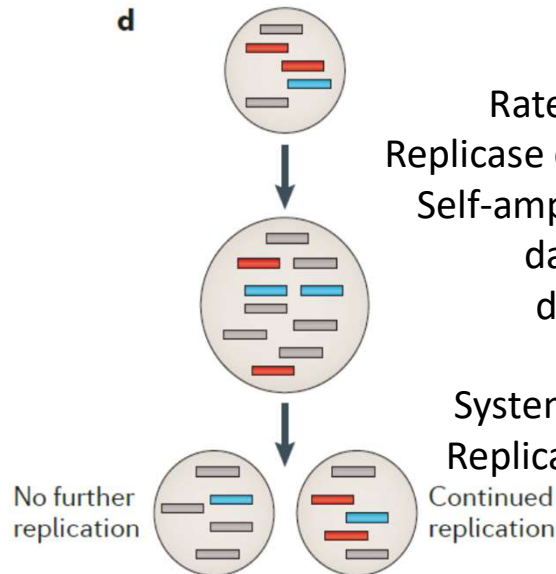
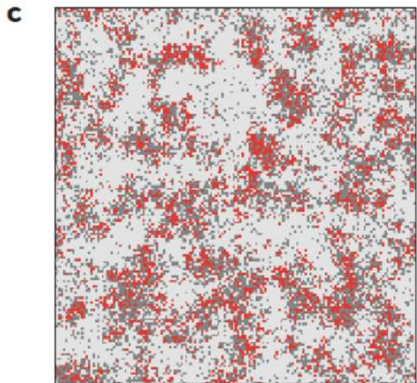
Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. *Science* **2001**, 292, 1319–1325.

# The RNA world

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



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 parasites (templates that copy themselves but do not contribute to the replication of the polymerase).

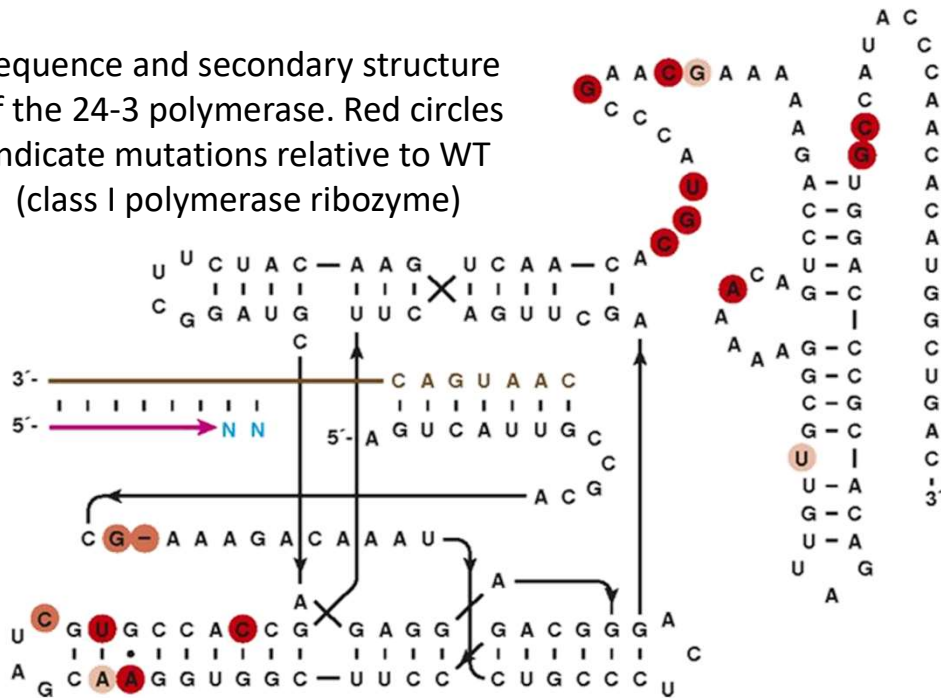
Systems of altruistic replicators are destroyed by parasites (grey).  
 Replicators (red) can survive e.g. by diffusion on 2D surfaces (**c**) or selection inside compartments (**d**)

## 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)



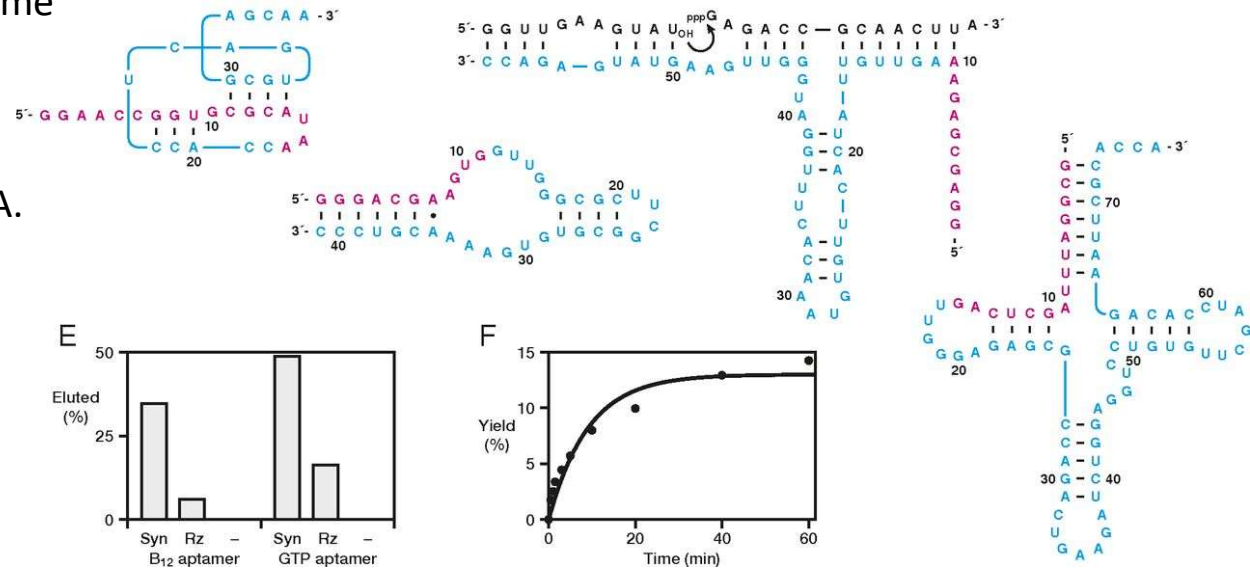
Thus, the replication of RNA and the expression of functional RNA can be accomplished with RNA alone. Combining and improving these activities may enable the self-sustained evolution of RNA and offers a potential route to a synthetic form of RNA life.

# 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.

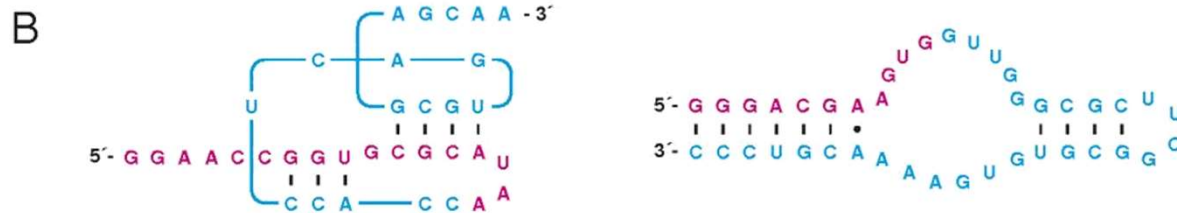
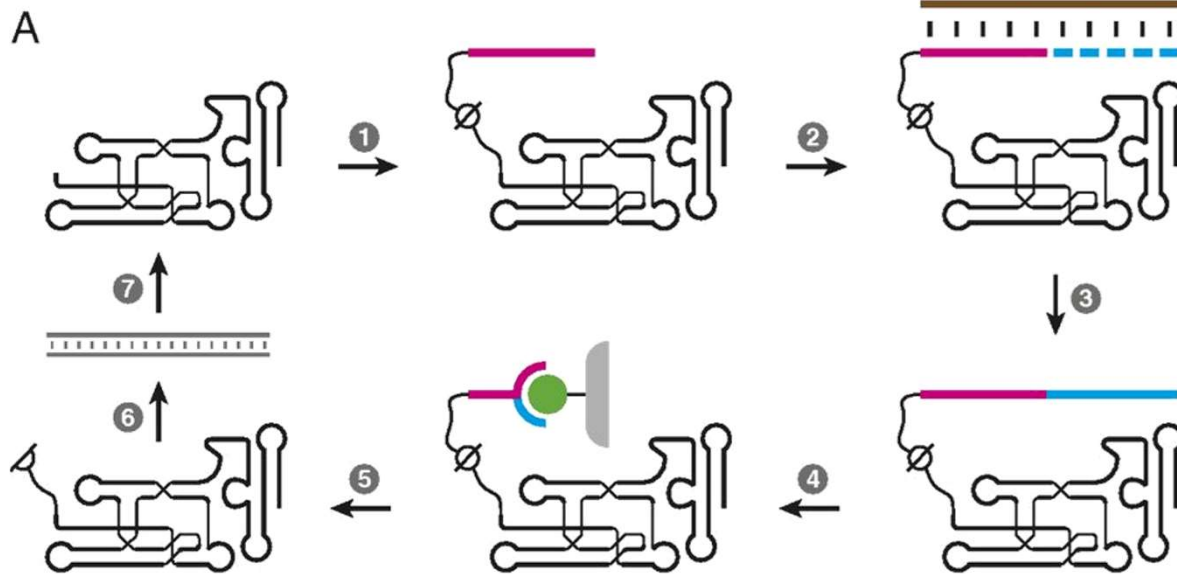


D. P. Horning, G. F. Joyce

*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<sub>12</sub> or GTP), enabling selection based on binding of the completed aptamer to its cognate ligand. Selection pressure for both sequence generality and accuracy.

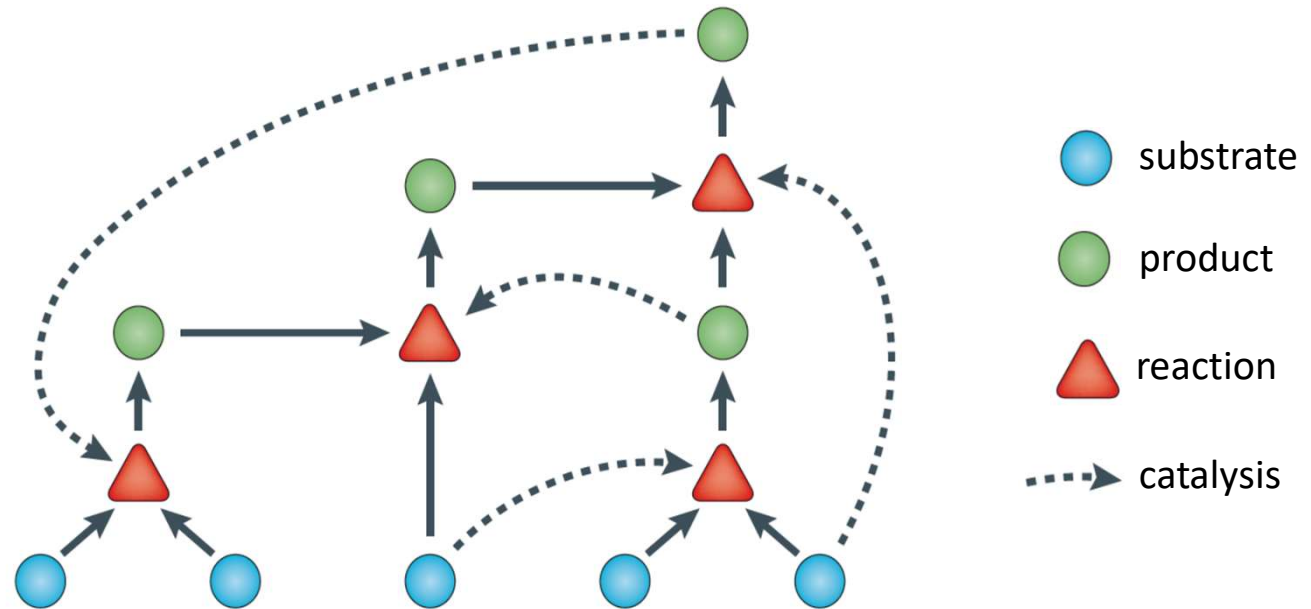
In vitro evolution of RNA polymerase ribozymes. (A) Selective amplification of ribozymes that extend a tethered RNA primer (magenta) on a separate RNA template (brown) to complete a 3'-truncated aptamer. (1) Attachment of the primer to the ribozyme via a photocleavable linker; (2) hybridization of the 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).

# The RNA world

## Replicase - problem

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

## Possible solution – autocatalytic networks

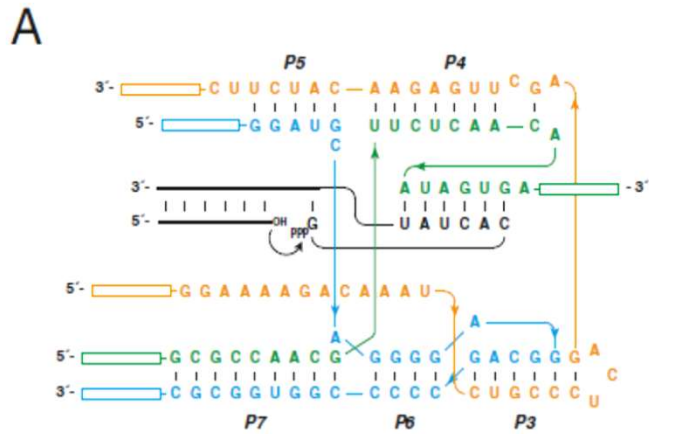


No component can replicate without all the others

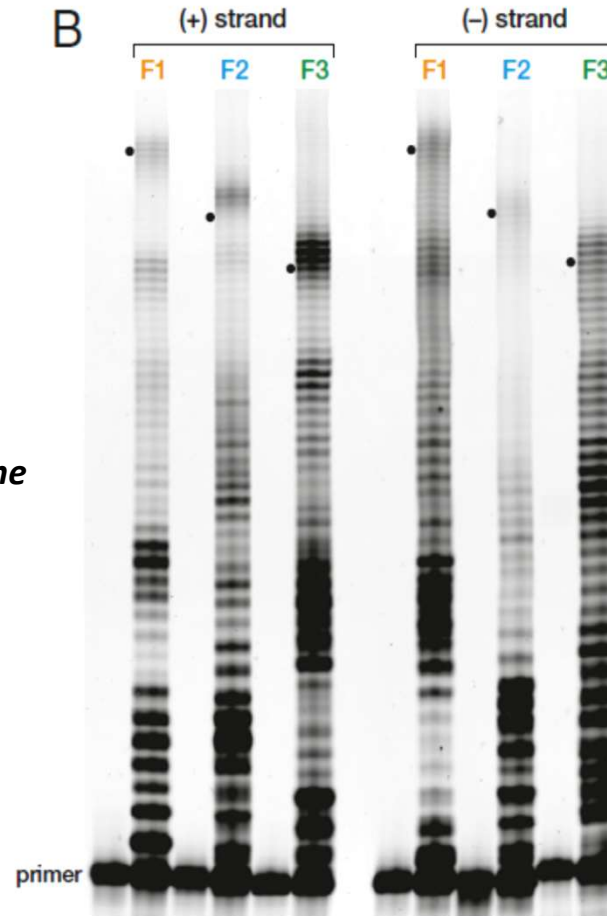
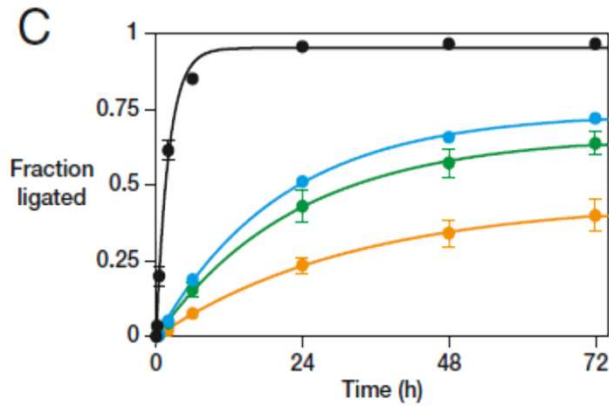


# The RNA world

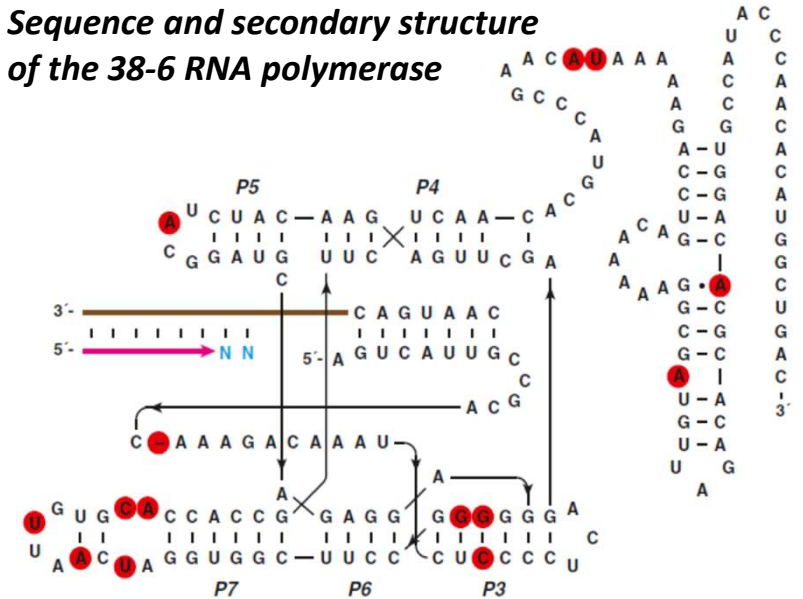
The polymerase „38-6” is able to synthesize its own evolutionary ancestor, an RNA ligase ribozyme, in the form of three fragments that assemble to give a functional complex



Three fragments of the class I ligase ribozyme



Sequence and secondary structure of the 38-6 RNA polymerase

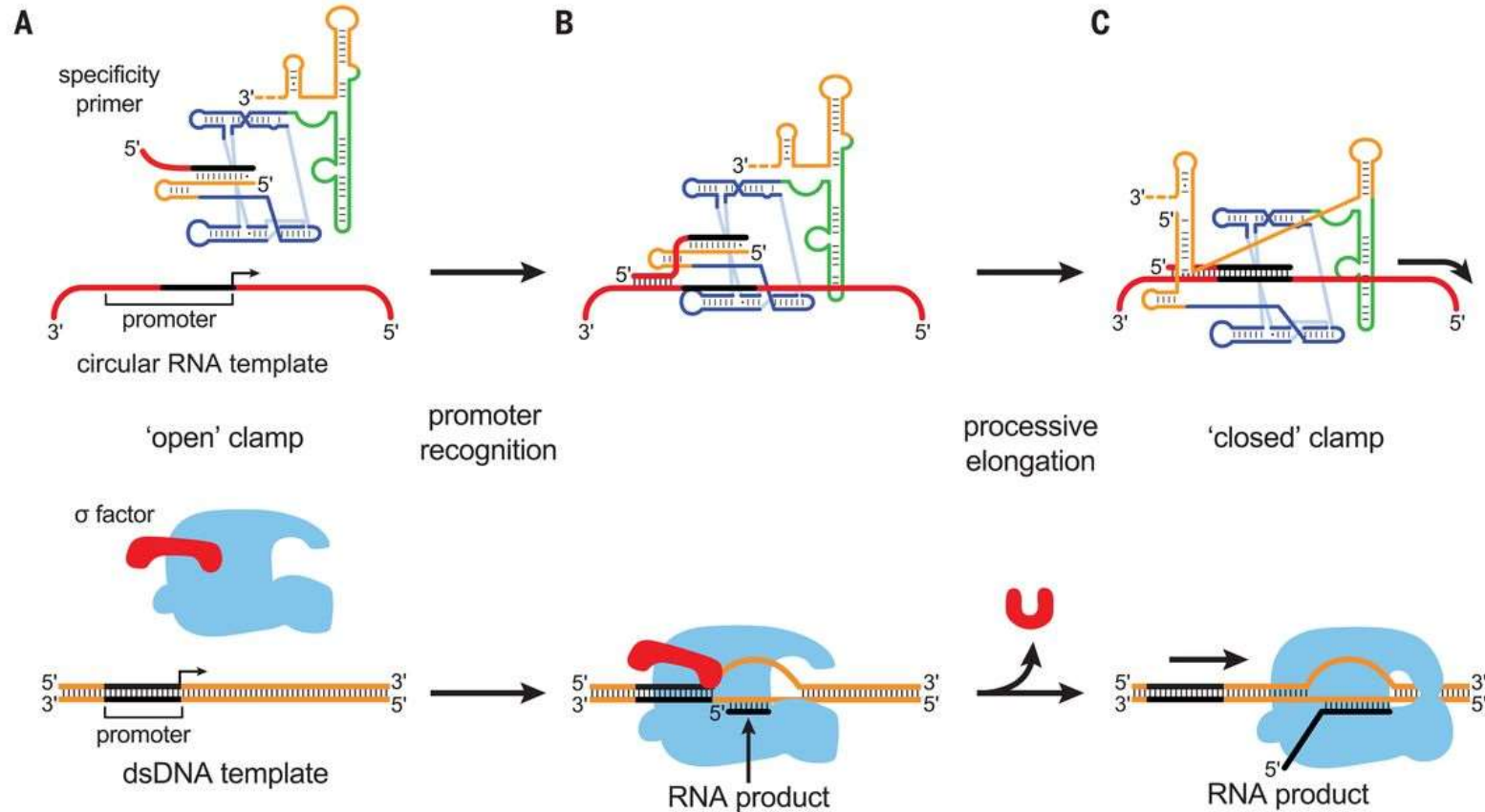


K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce  
*Proc. Natl. Acad. Sci. USA (PNAS)* **2020**, *117* (6), 2906-2913



# The RNA world

## Processive RNA polymerization and promoter recognition

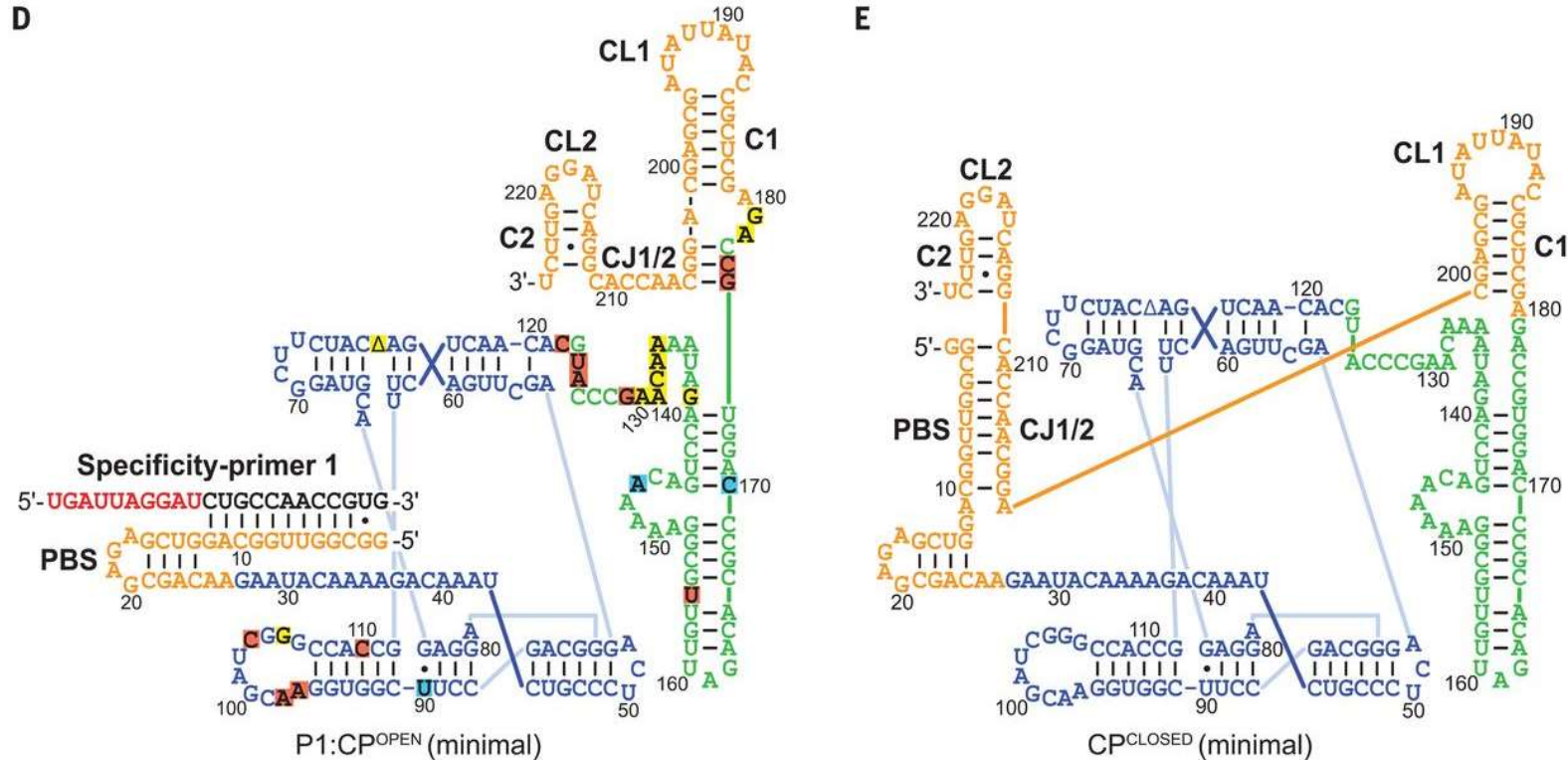


(A) RNA specificity primer–activated “open” form P1:CP<sup>OPEN</sup> (top) and DdRP holo-enzyme (bottom). (B) The specificity primer localizes CP to an ssRNA promoter (top), whereas a sigma factor localizes the DdRP to a DNA promoter (bottom). (C) In both cases, a clamped “closed” state forms, enhancing polymerization.

R. Cojocaru, P. J. Unrau *Science* 2021, 371, 1225-1232

# The RNA world

## Processive RNA polymerization and promoter recognition

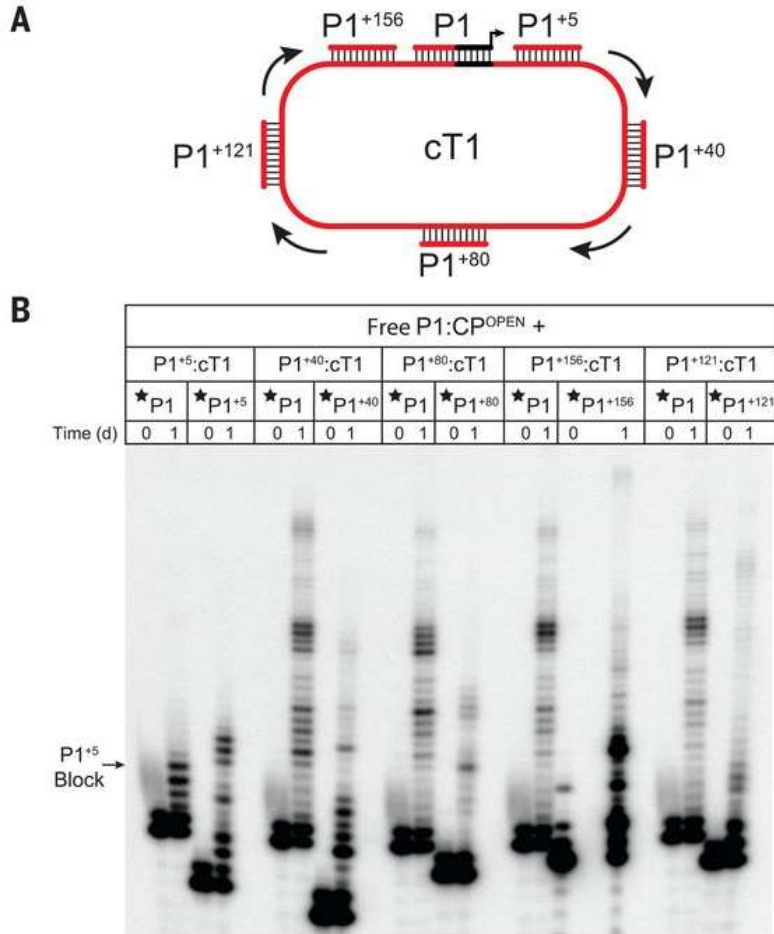


(**D** and **E**) Secondary structure of the minimal P1:CP<sup>OPEN</sup> form (**D**) and the CP<sup>CLOSED</sup> form (**E**). Colored lines indicate the ligase core (blue), accessory domain (green), and minimal clamping domain (orange). Up mutations designed into the selection are shown in red boxes, rediscovered up mutations in teal boxes, and newly discovered mutations in yellow boxes.

R. Cojocaru, P. J. Unrau *Science* **2021**, *371*, 1225-1232

# The RNA world

## Processive RNA polymerization and promoter recognition



CP processively extends multiple primers on the same promoter template.

(A) Distal primer locations relative to the P1:cT1 initiation site. Arrows indicate direction of polymerization.

(B) Extension assay of P1:CP<sup>OPEN</sup> and P1<sup>+n</sup> ( $n = 5, 40, 80, 121, \text{ and } 156$ ) primers added to cT1. Either P1 or the P1<sup>+n</sup> primers were radiolabeled as indicated (star).

## *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.

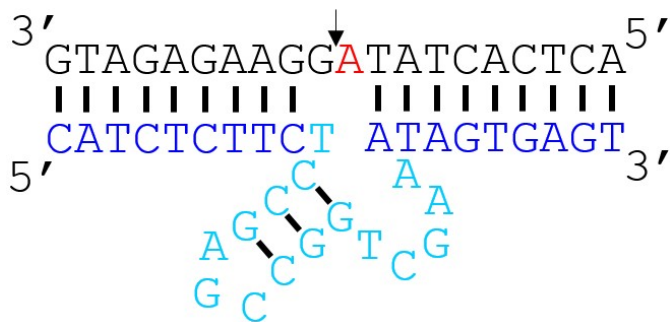
The apparent lack of naturally occurring deoxyribozymes may also be due to the primarily double-stranded conformation of DNA in biological systems → limited physical flexibility and ability to form tertiary structures (catalytic 3D-folds)

# 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<sup>2+</sup> 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?*

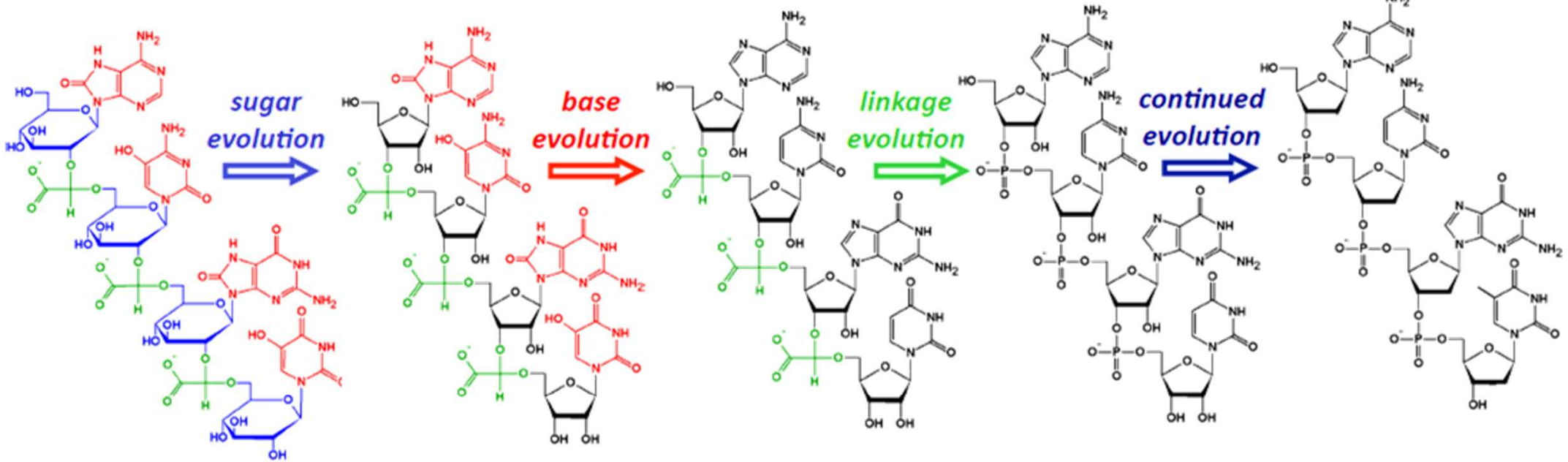


„RNA-second“

proto-RNA

RNA

DNA



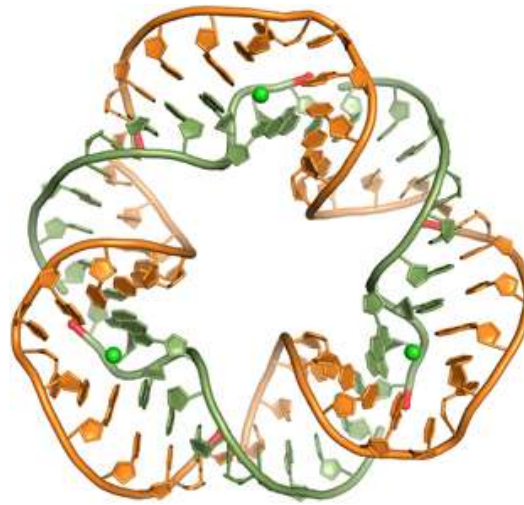
Easy to assemble



Functionally superior

*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.

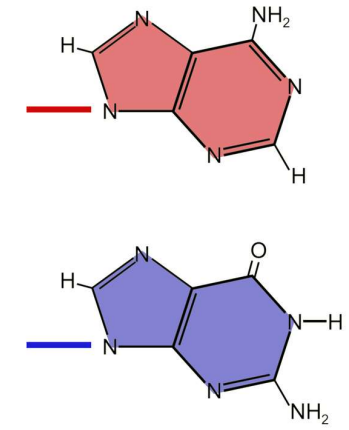
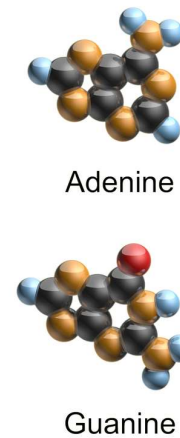
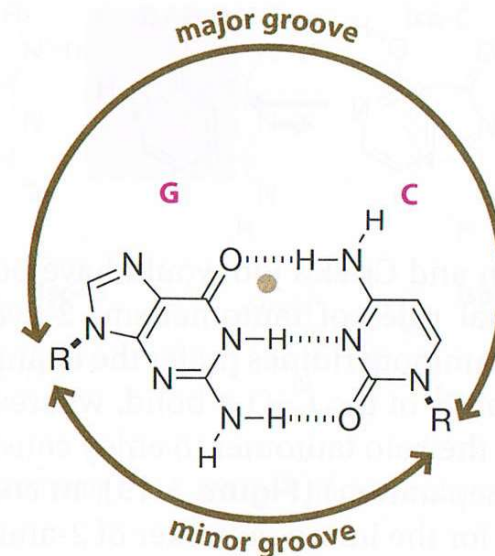
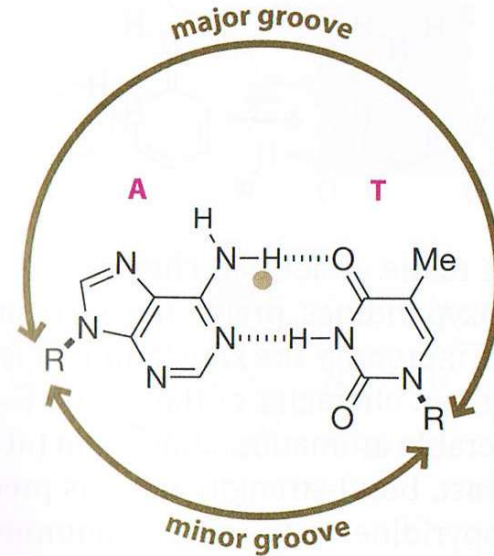
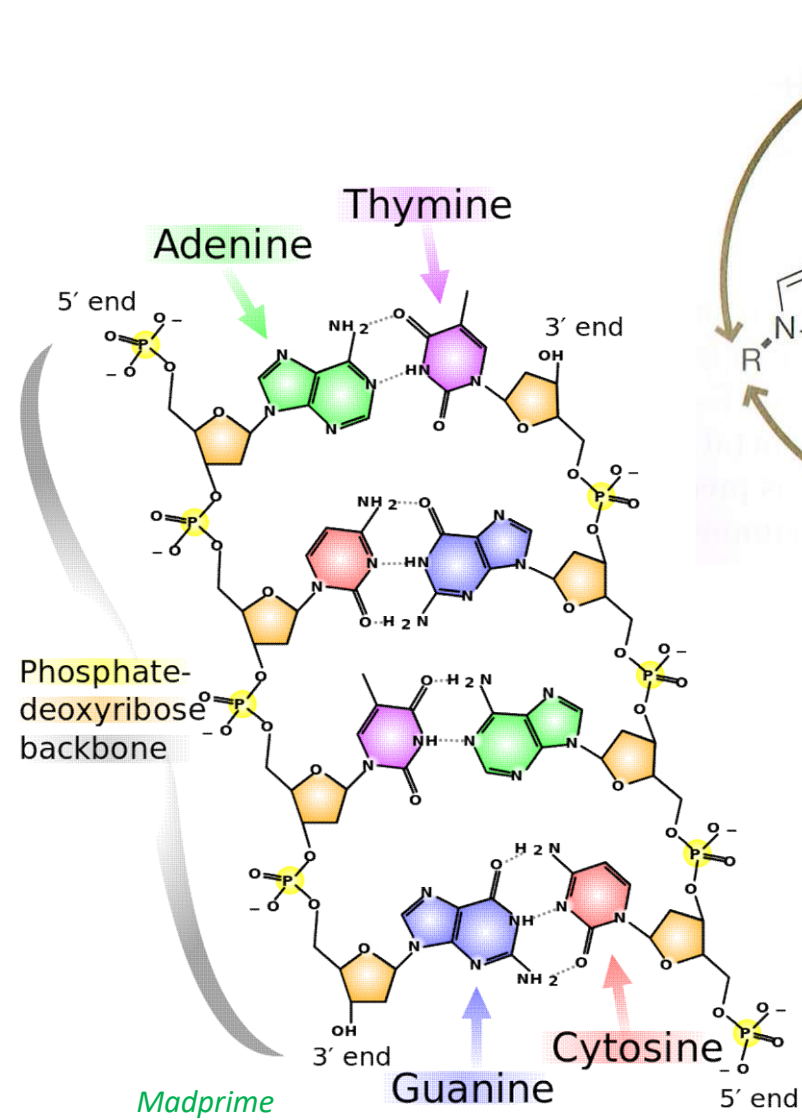
# CHAPTER 1



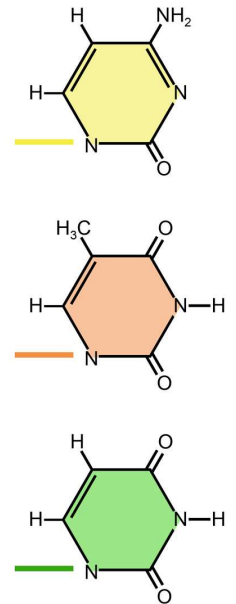
## ***OLIGONUCLEOTIDES***

*Part 1 – modified canonical nucleobases*

# Canonical nucleobases and Watson-Crick pairing in DNA



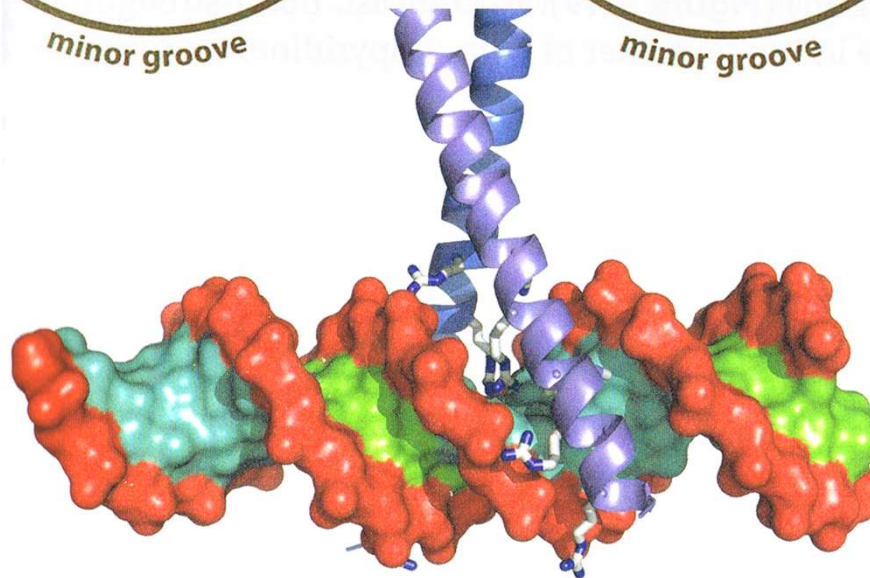
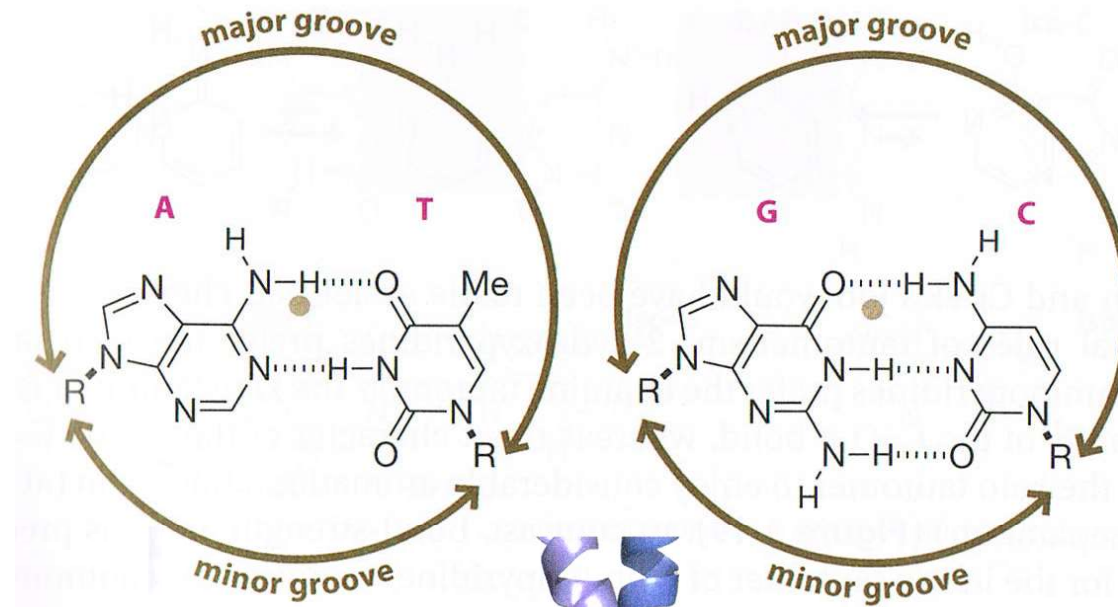
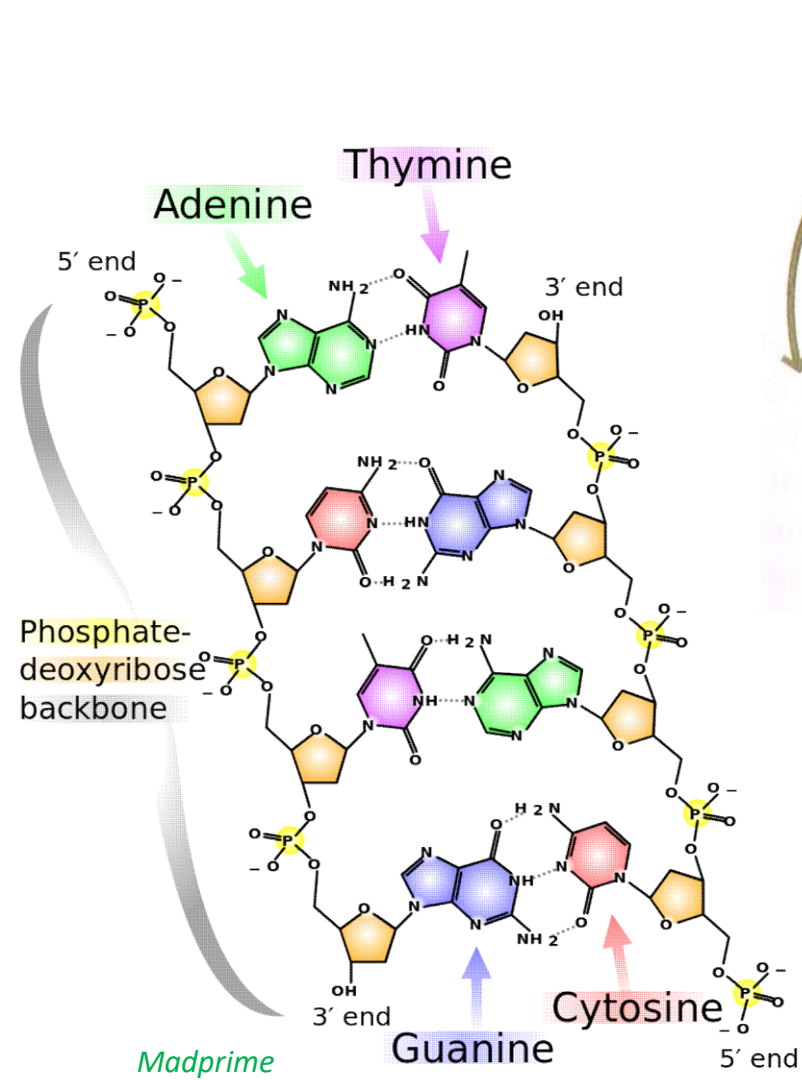
**Purines**



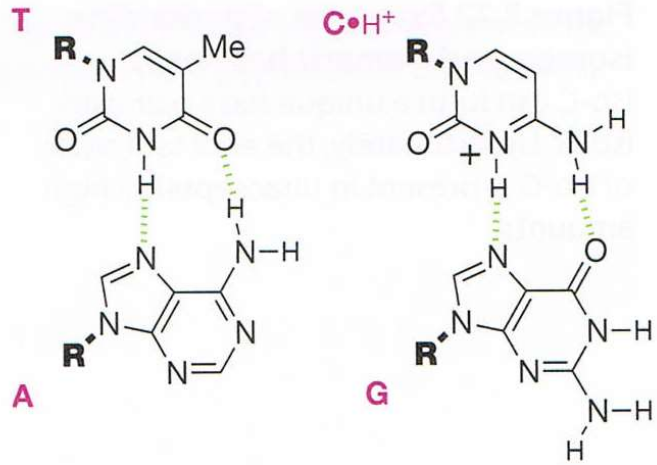
**Pyrimidines**

*Bruce Blaus*

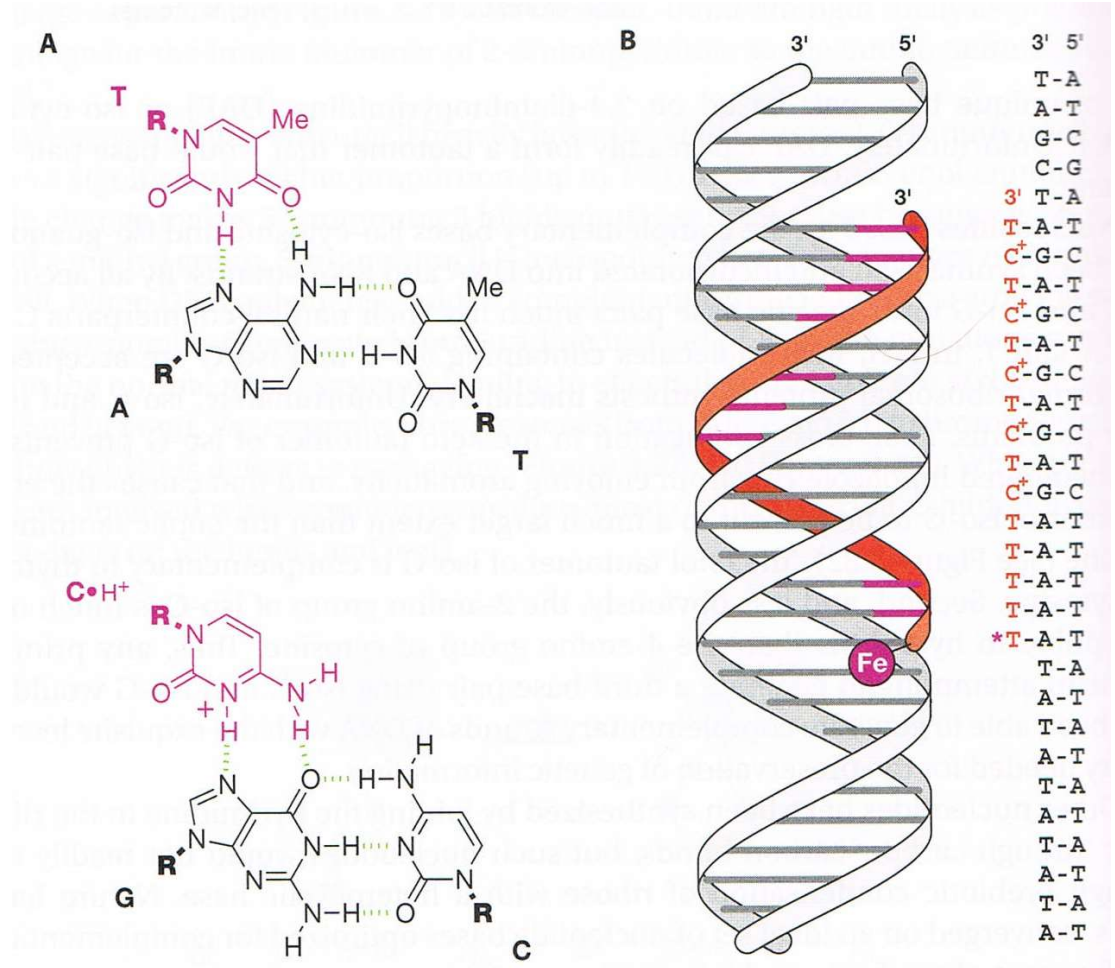
# Canonical nucleobases and Watson-Crick pairing in DNA



# Hoogsten base pairing of canonical DNA nucleobases

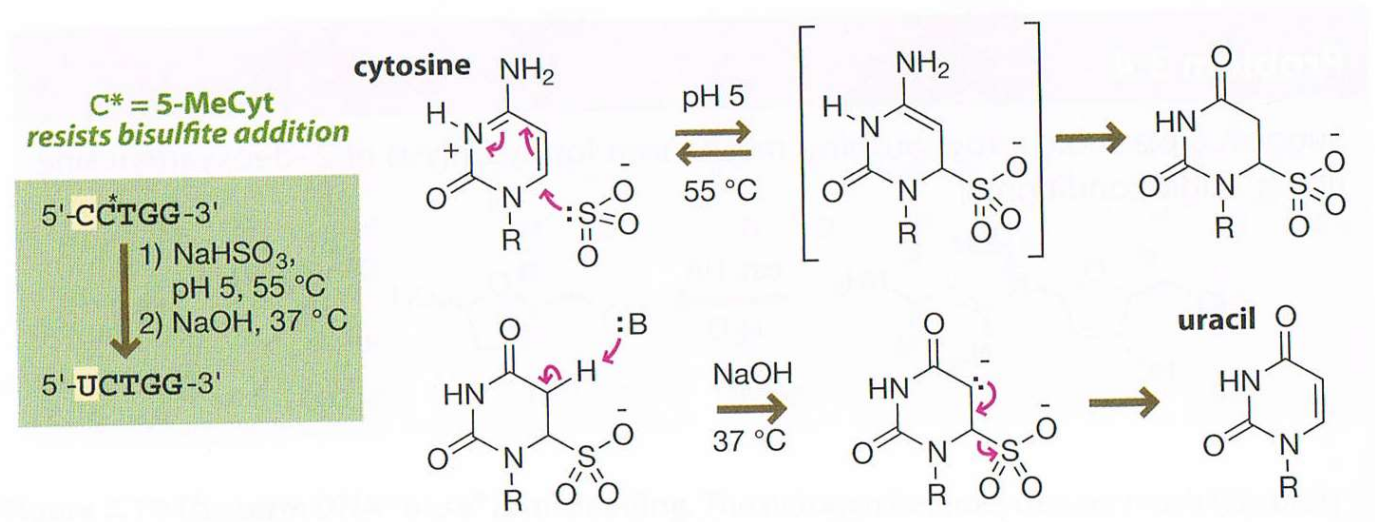
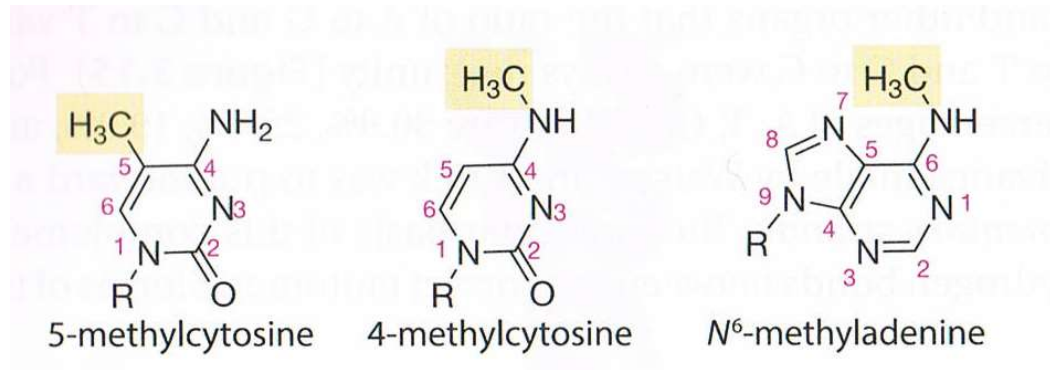


**Figure 3.24 Hoogsteen base pairs.** Hoogsteen base pairs use a different edge of the purine from a Watson–Crick base pair.

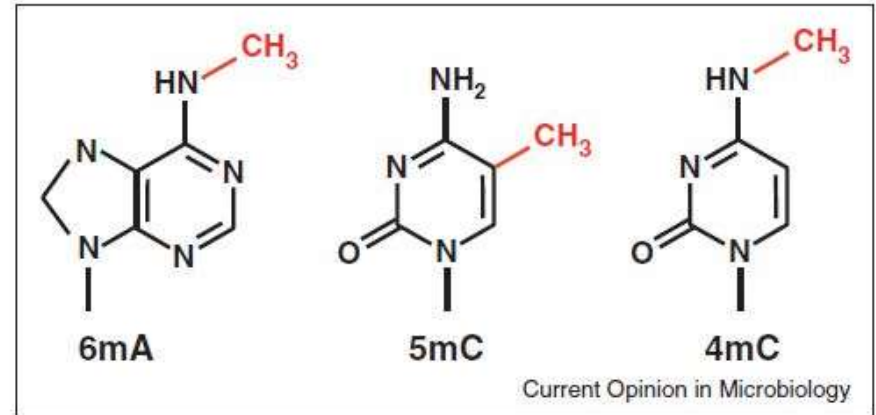
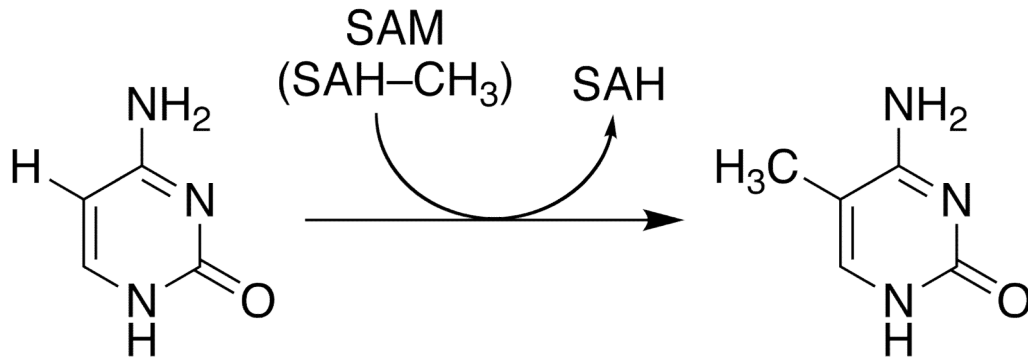


H. E. Moser, P. B. Dervan *Science* 1987, 238, 645-650

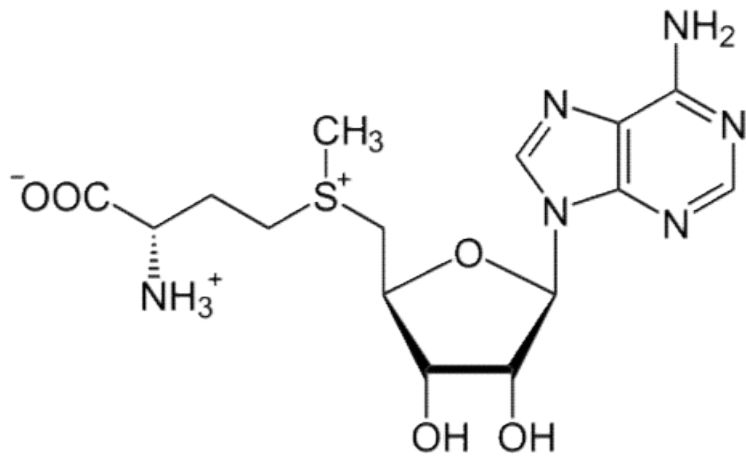
# Modifications of nucleobase structures tolerated by polymerases



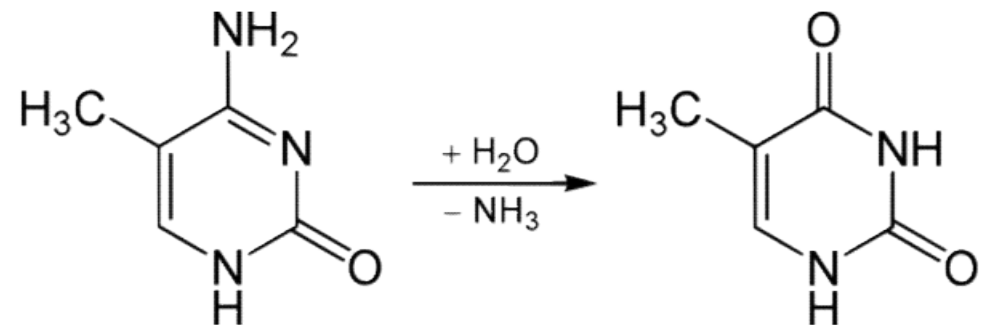
## Modifications of nucleobases



Chemical structures of common modified bases generated by DNA methyltransferases.

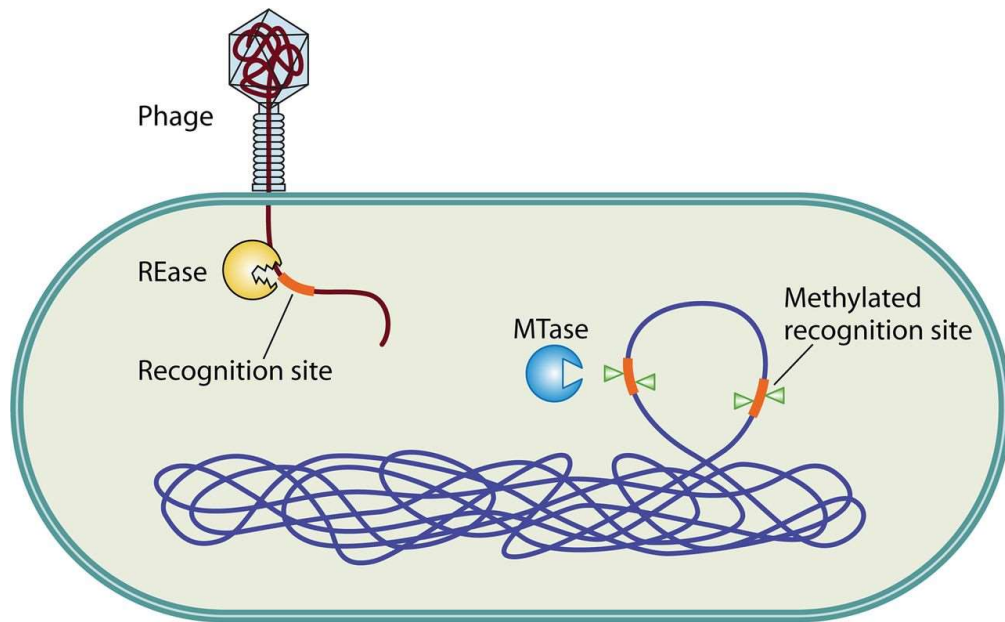


*S*-Adenosylmethionine (SAM)



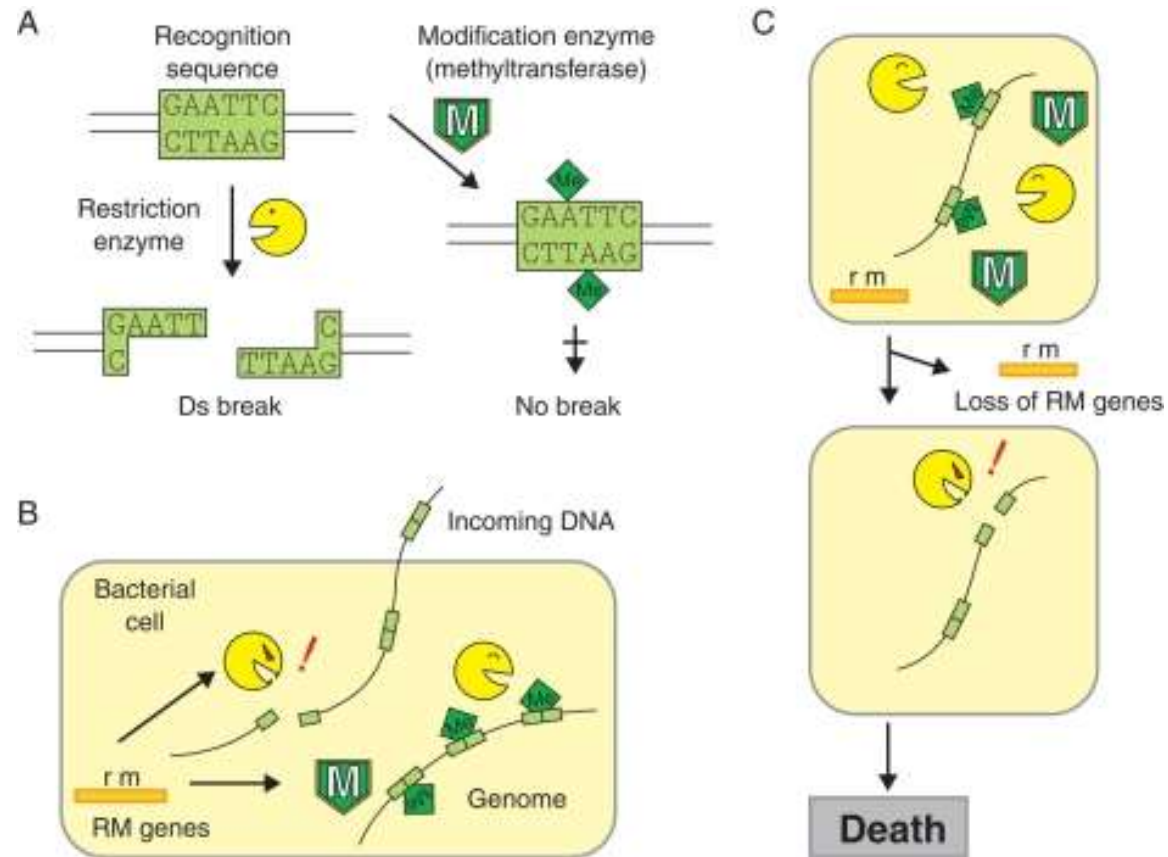
# Restriction modification system

„Immune system“ of bacteria and archaea against attacking viruses



K. Vasu, V. Nagaraja

*Microbiol. Mol. Biol. Rev.* **2013**, 77(1), 53-72

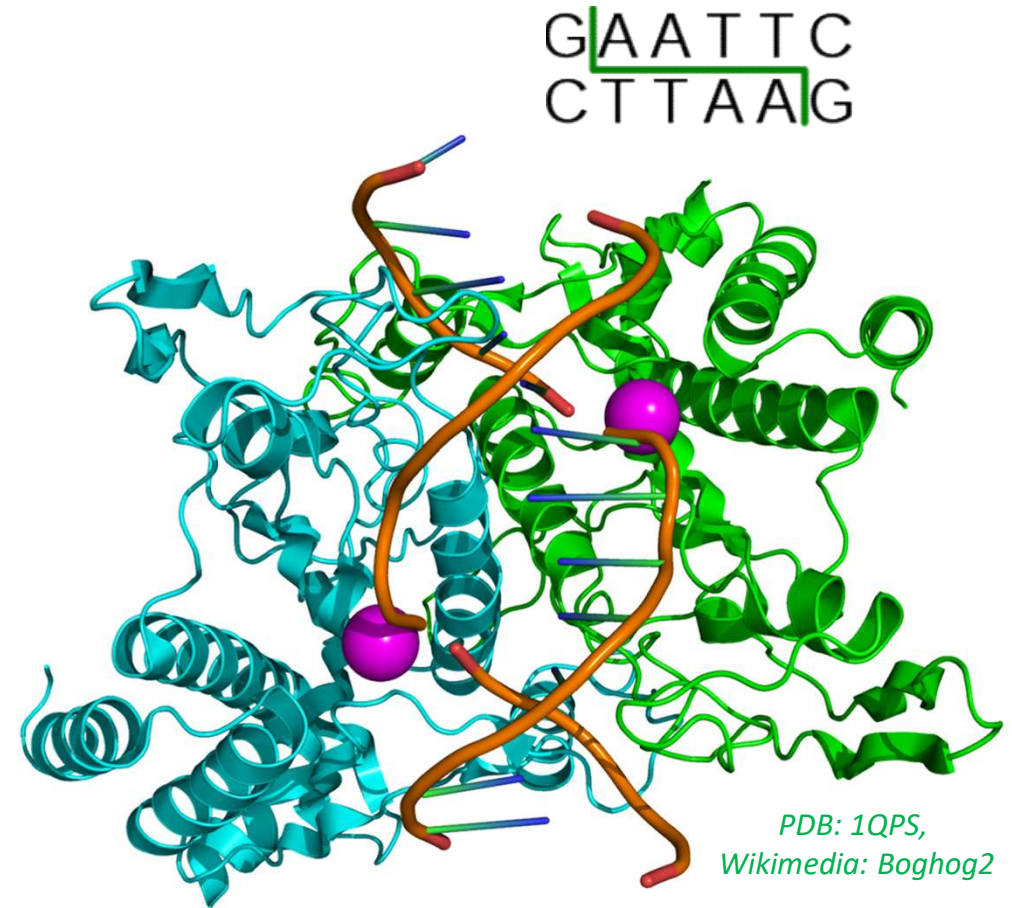
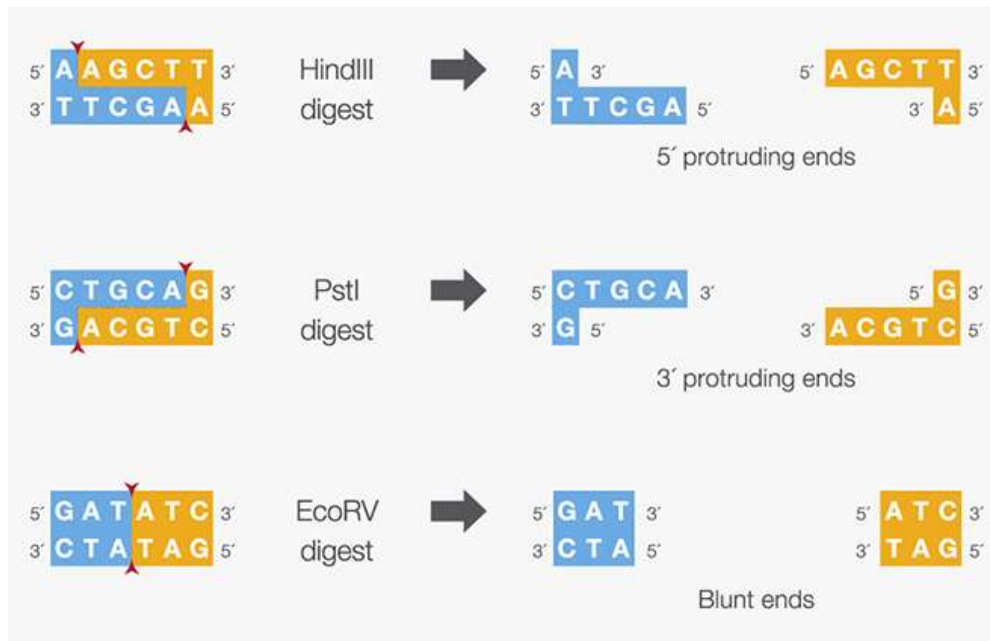


K. Ishikawa *et al.* DOI: 10.1093/dnares/dsq027



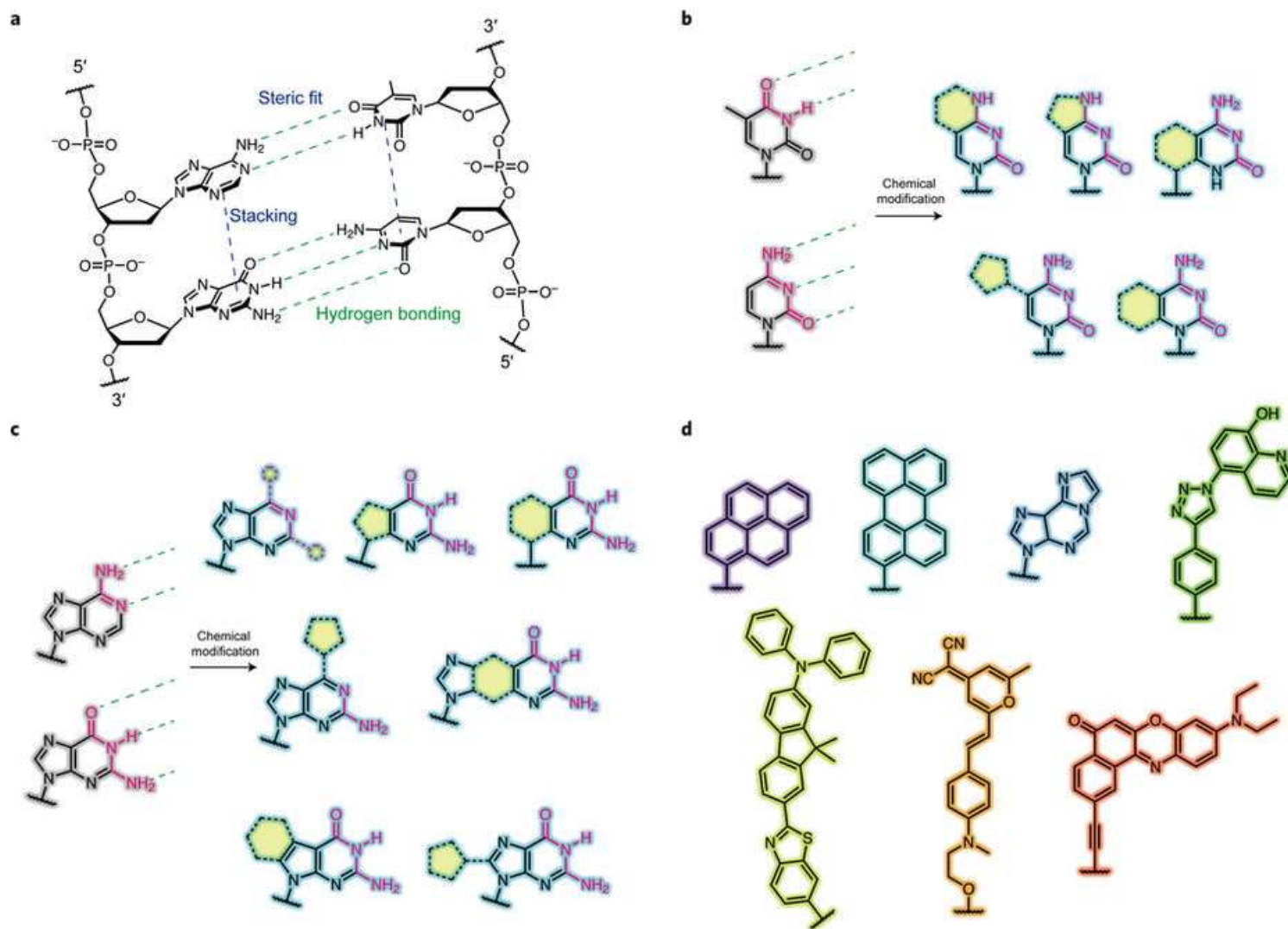
## *EcoI – a typical restriction enzyme*

### *Products of restriction enzymes*



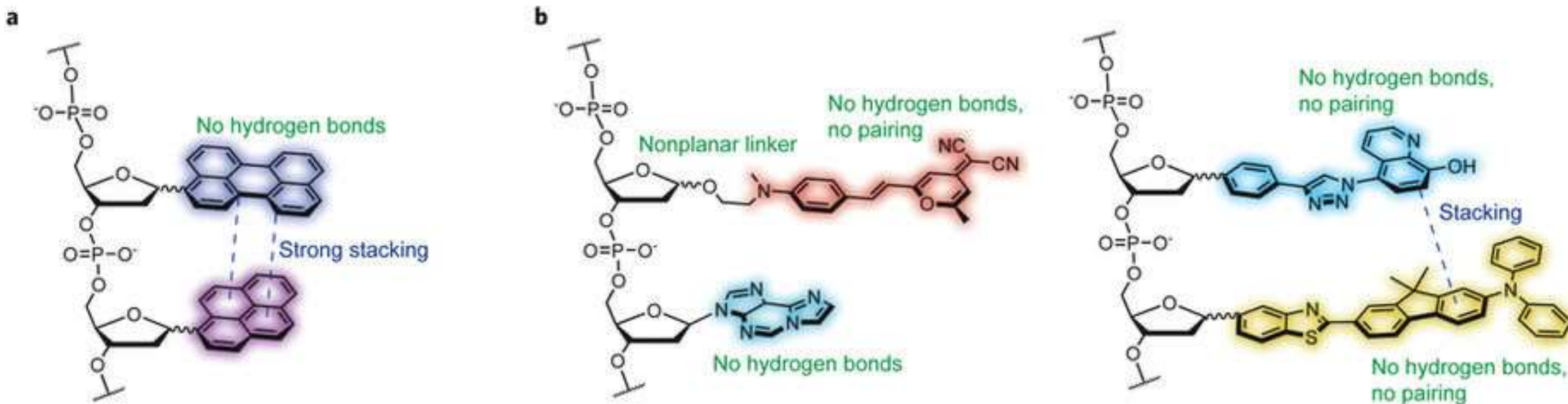
Structure of the homodimeric restriction enzyme EcoRI (cyan and green cartoon diagram) bound to double stranded DNA (brown tubes). Two catalytic magnesium ions (one from each monomer) are shown as magenta spheres and are adjacent to the cleaved sites in the DNA made by the enzyme (depicted as gaps in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

## Non-canonical fluorescent nucleobases



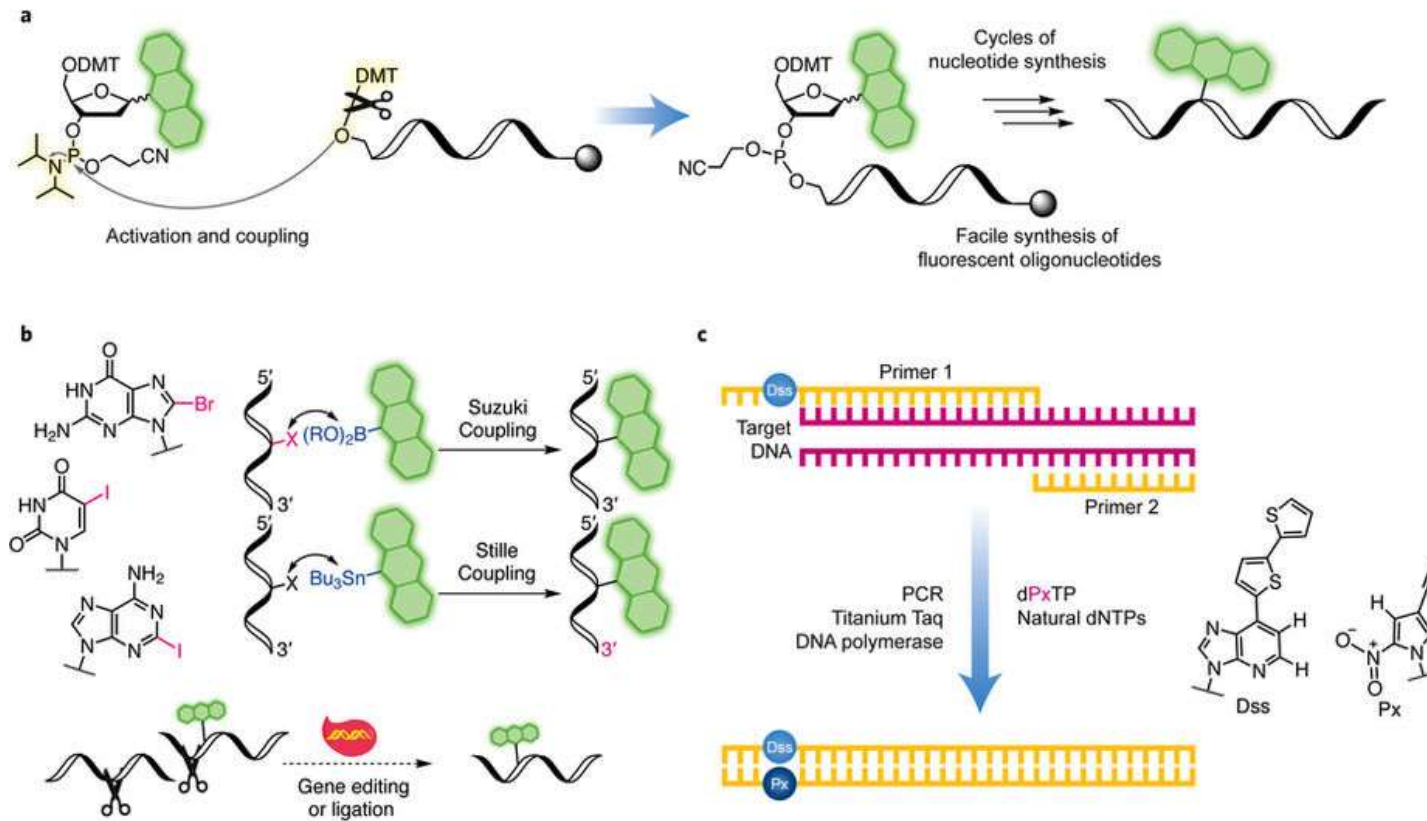
W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9*, 1043-1055

## Non-canonical fluorescent nucleobases



**a**, Nucleobases composed of aromatic hydrocarbons. **b**, Nucleobases composed of planar heterocyclic fluorophores. The lack of hydrogen bonding and weaker  $\pi$ -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.

# Strategies for incorporating fluorescent nucleobases

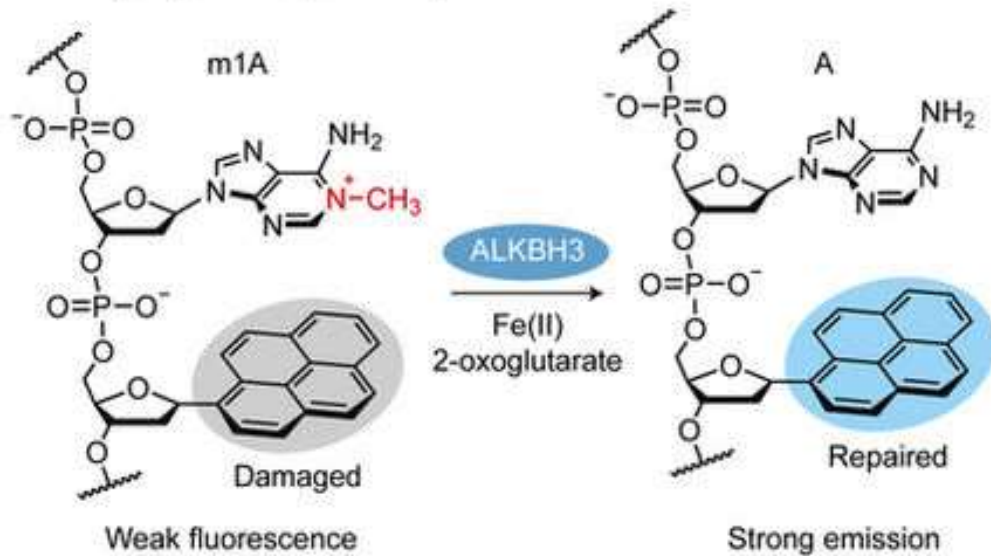


**a**, Direct oligonucleotide synthesis via synthesizer and phosphoramidite chemistry. **b**, Post-synthesis modification using mild coupling methods or gene-editing methods. Gene-editing and ligation methods enzymatically join smaller labelled strands to make longer ones. **c**, Direct enzymatic incorporation 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.

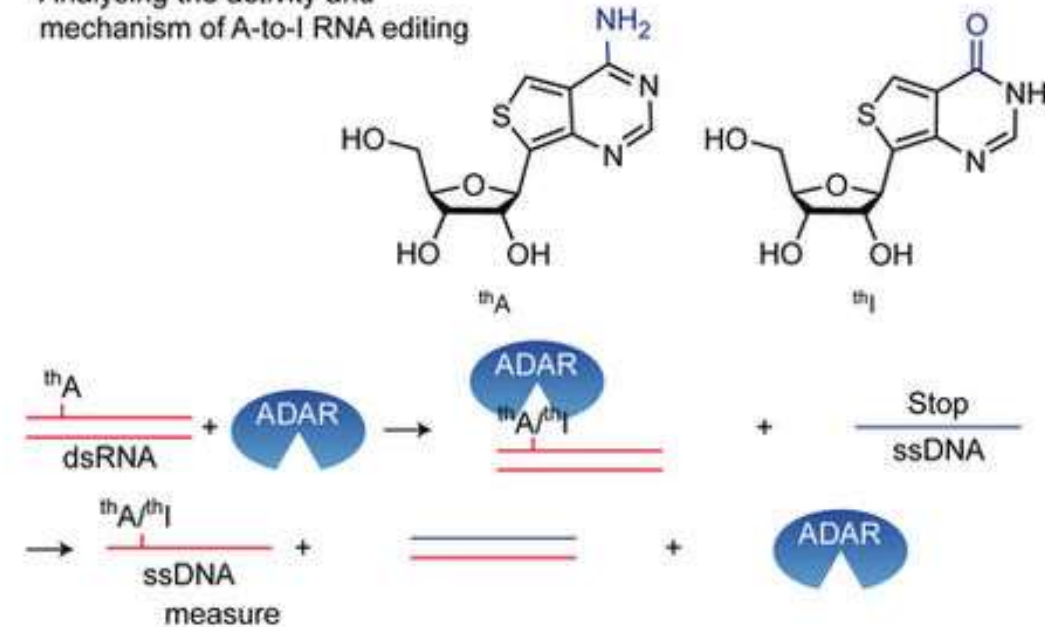
W. Xu, K. M. Chan, E. T. Kool *Nature Chem.* **2017**, *9*, 1043-1055

## Fluorescent nucleobases for studying DNA and RNA

**a** Probing alkylation repair activity



**b** Analysing the activity and mechanism of A-to-I RNA editing



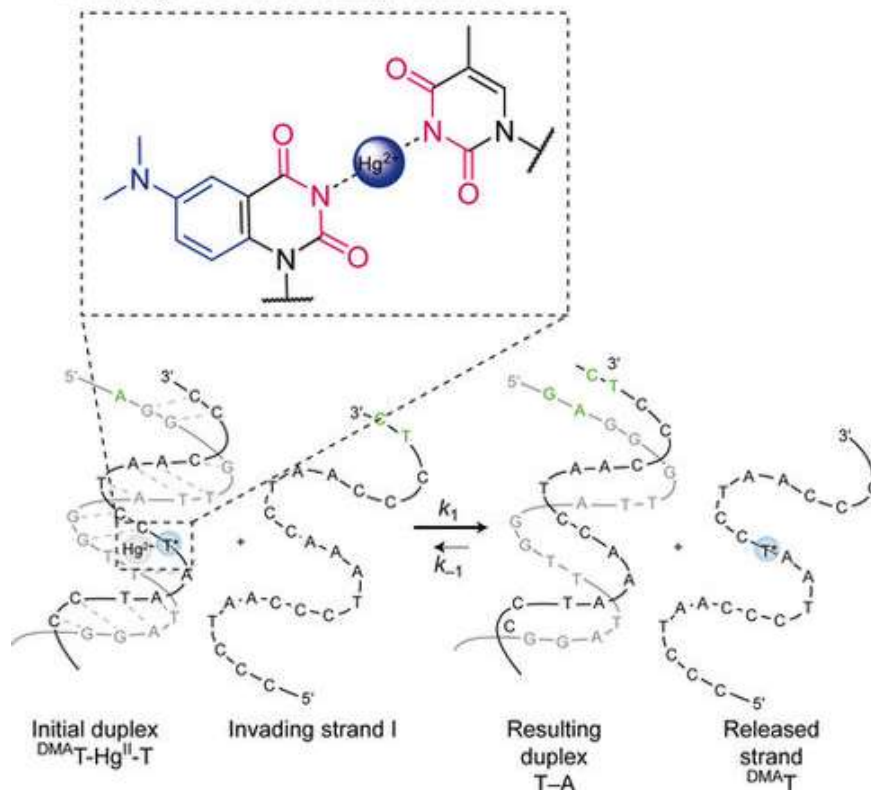
**a**, Fluorogenic sensing of a demethylation enzyme, ALKBH3. The emission of pyrene is initially quenched by the positive charge of 1-methylated adenine (m1A). When ALKBH3 demethylates m1A, the quenching effect is removed and a signal is generated.

**b**, Fluorogenic analysis of adenine-to-inosine RNA editing enzyme. The emission maxima of the thiolated adenine (<sup>th</sup>A) and inosine (<sup>th</sup>I) are different. Hence by measuring the intensity of <sup>th</sup>A and <sup>th</sup>I 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

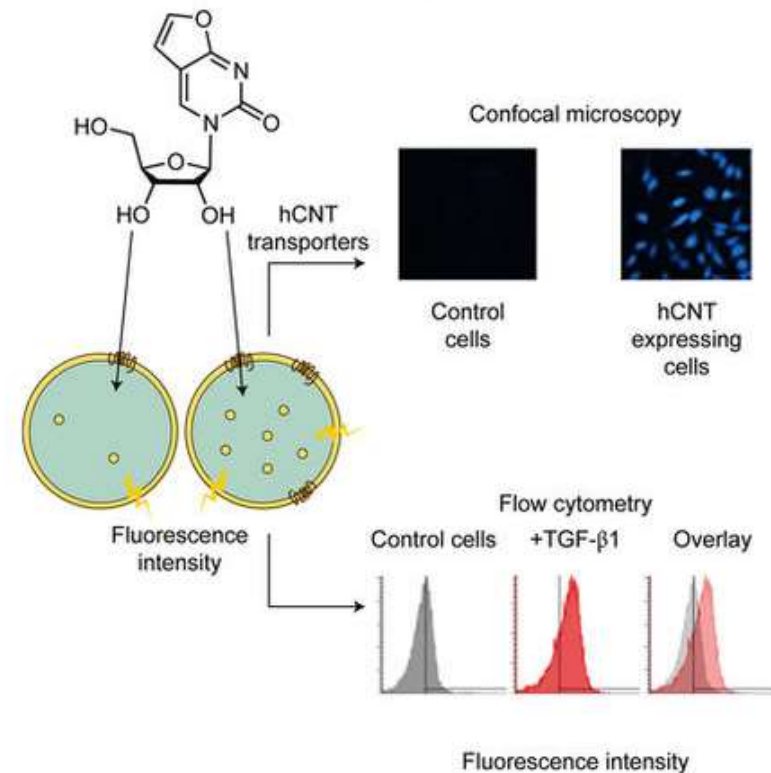
## Fluorescent nucleobases for studying DNA and RNA

**c** Analysing the effects of mercury metabolism



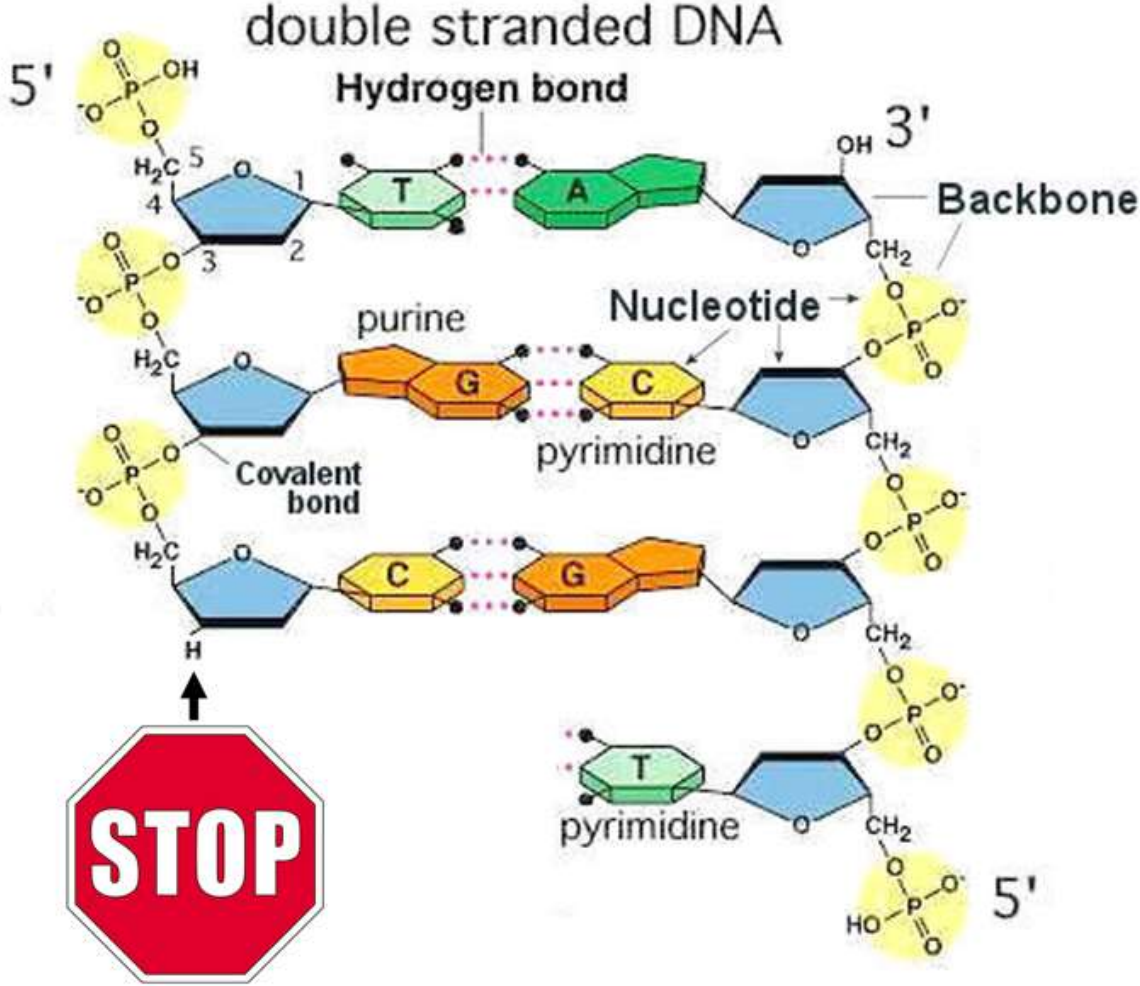
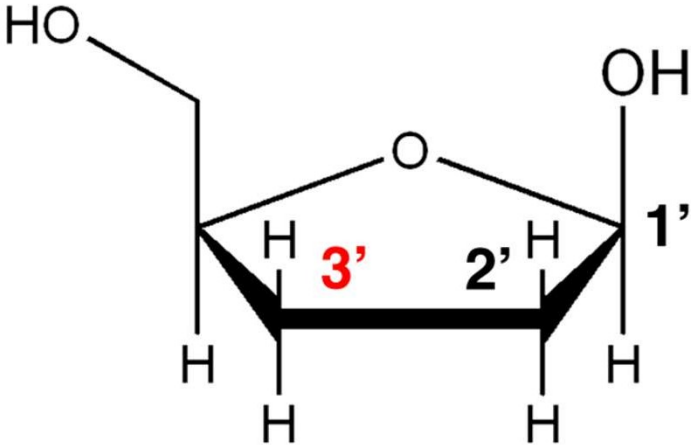
**c**, Kinetic and thermodynamic investigation of the effects of mercury on DNA metabolism. The fluorescent thymine can chelate mercury with another thymine ring and link DNA strands. This can be used to probe mercury metabolism *in vivo* and to study the effects of mercury on DNA status.

**d** Probing nucleoside transporter activity *in vivo*

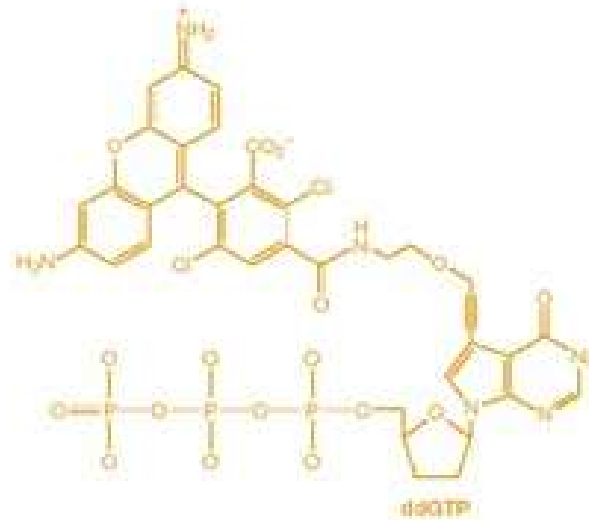
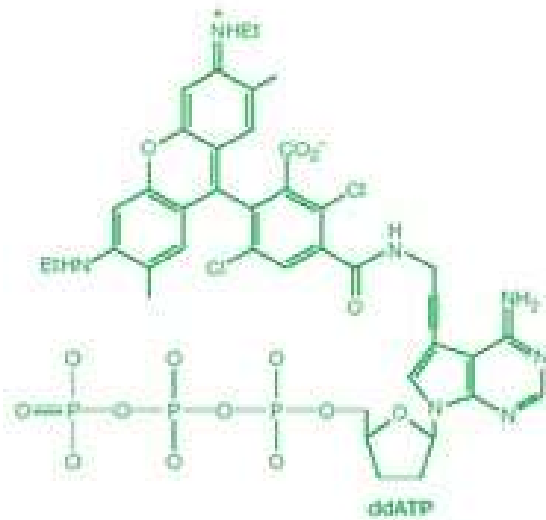
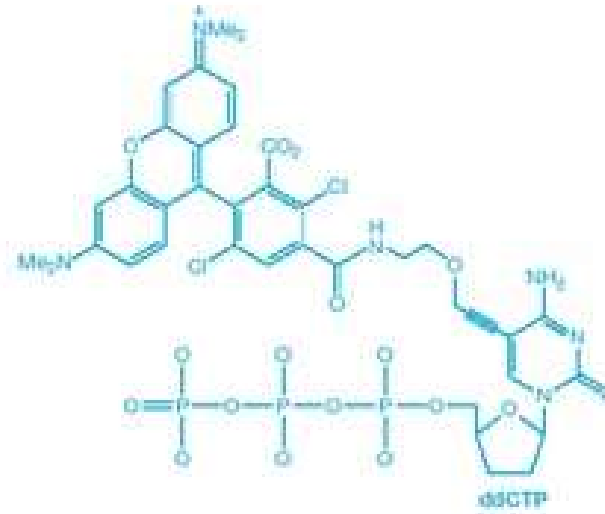
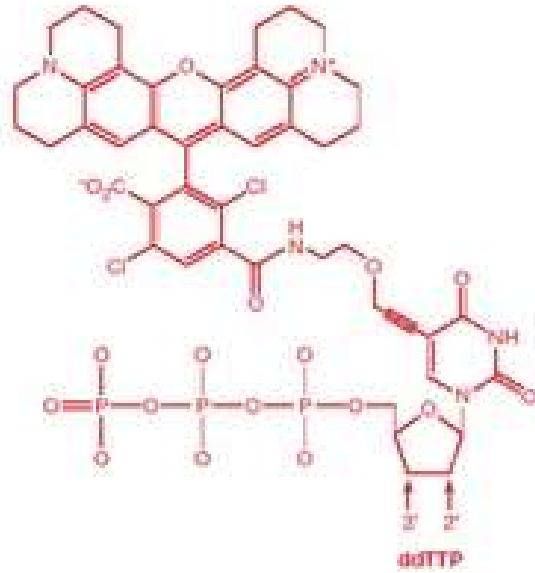


**d**, Visualization and analysis of human concentrative nucleoside transporters (hCNTs). The fluorescent nucleoside can enter the plasma membrane through the transporters, thus allowing the measurement of the transport activity. TGF- $\beta$ 1, transforming growth factor  $\beta$ 1. [W. Xu, K. M. Chan, E. T. Kool \*Nature Chem.\* 2017, 9, 1043-1055](#)

# Sanger sequencing



## Sanger sequencing





## Sanger sequencing

primer

5' 3'

TACGT

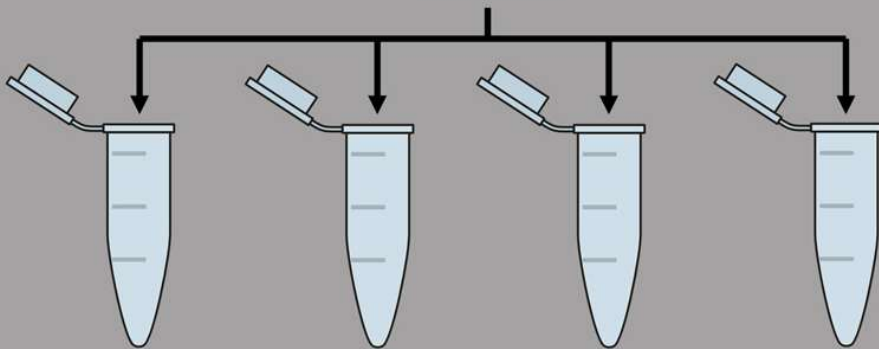
ATGCATTAGGGCCTGGCTCTTT

3'

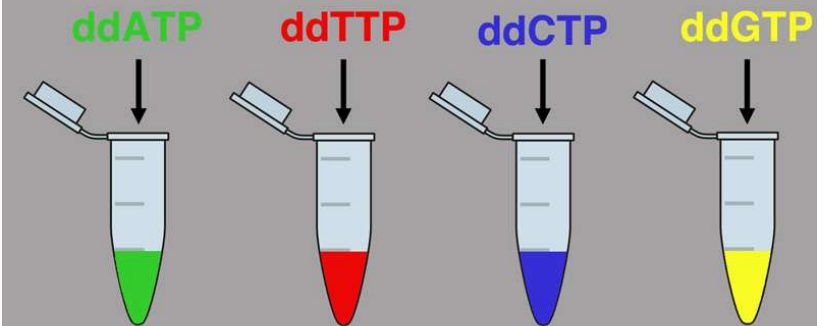
5'

template

DNA template + primer  
DNA polymerase  
dNTPs



Sanger sequencing



## Sanger sequencing

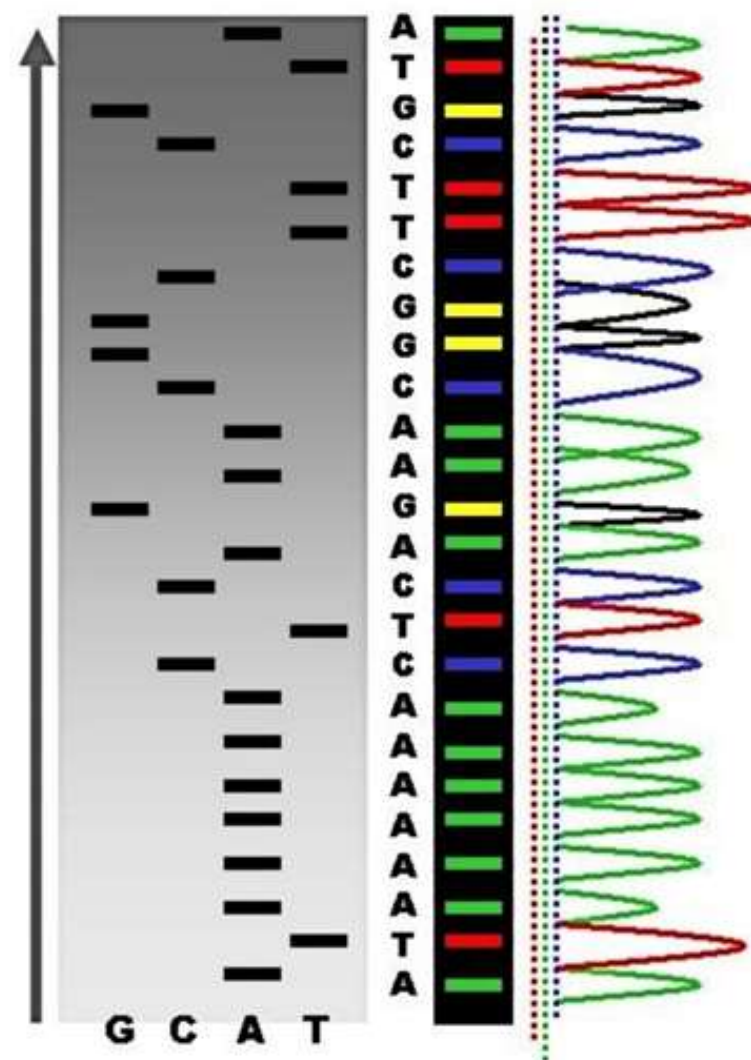
TACGTA**A**  
ATGCATTAGGGCCTGGCTCTTT

TACGTA**A**  
ATGCATTAGGGCCTGGCTCTTT

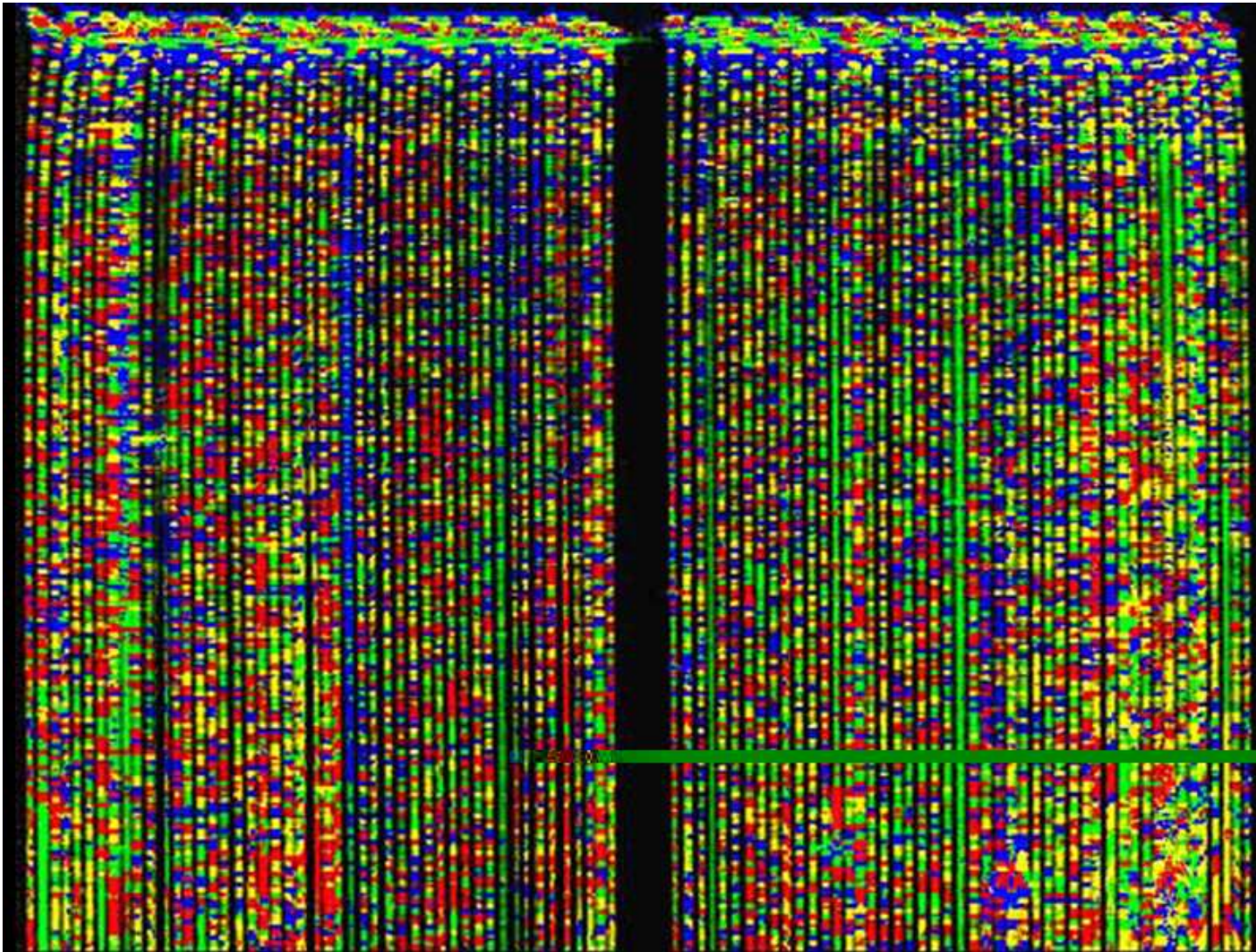
TACGTAATCCCGG**A**  
ATGCATTAGGGCCTGGCTCTTT

TACGTAATCCCGGACCG**A**  
ATGCATTAGGGCCTGGCTCTTT

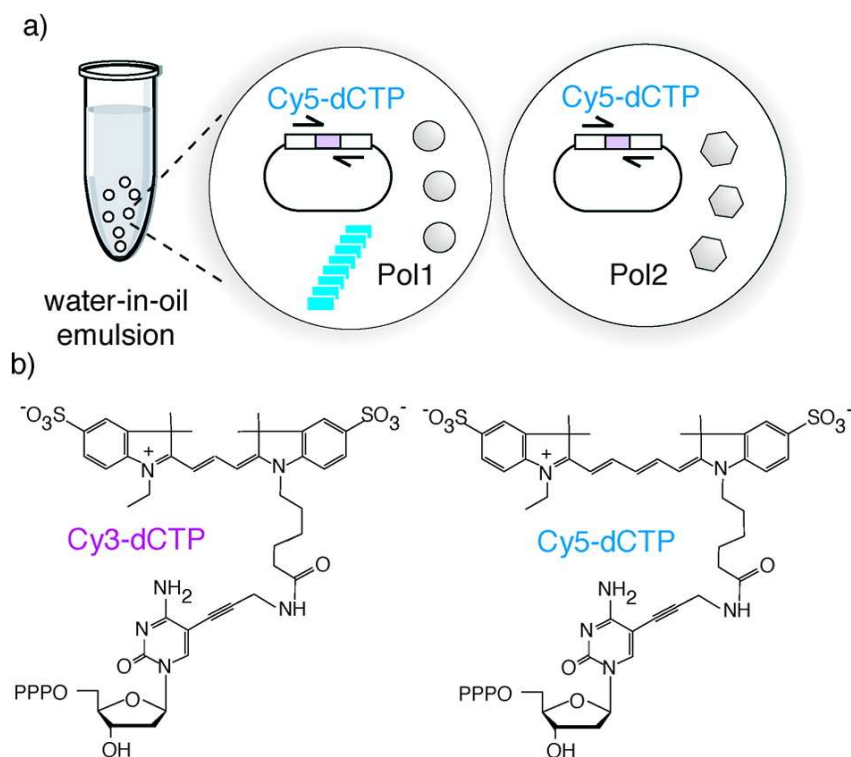
## Sequencing Gel



## *Sanger sequencing*



## CyDNA – synthesis and replication of highly fluorescently-labelled DNA



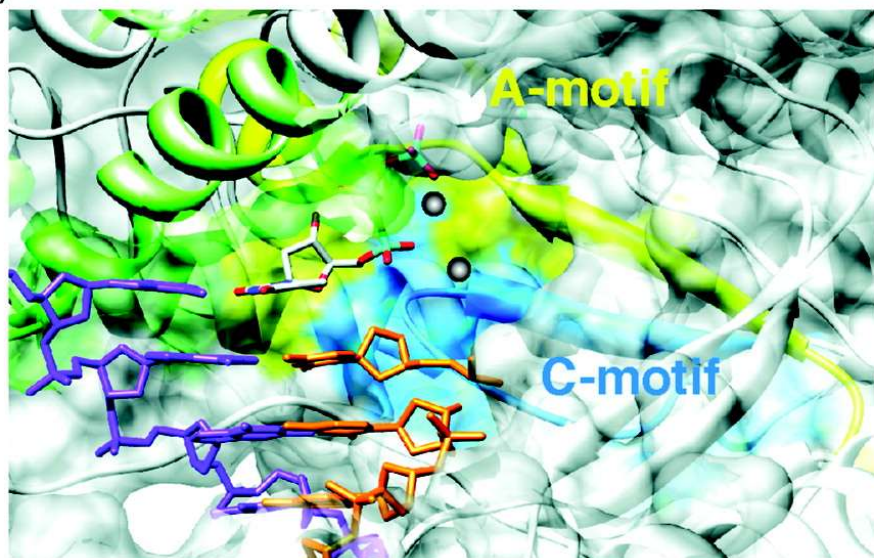
DNA Polymerase is evolutionarily optimized by mutagenesis to accept fluorescent Cy3 or Cy5 nucleotides instead of simple dCTP. Here, the polymerase replicates a short segment of its own encoding gene. Water/oil compartmentalization allows testing many independent mutations in parallel

*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.*

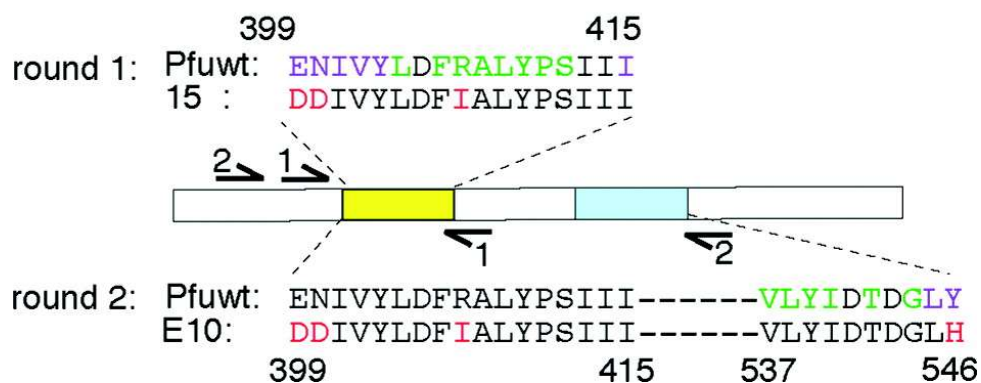
P. Holliger *et al.* *J. Am. Chem. Soc.* **2010**, *132*, 5096-5104

## 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)

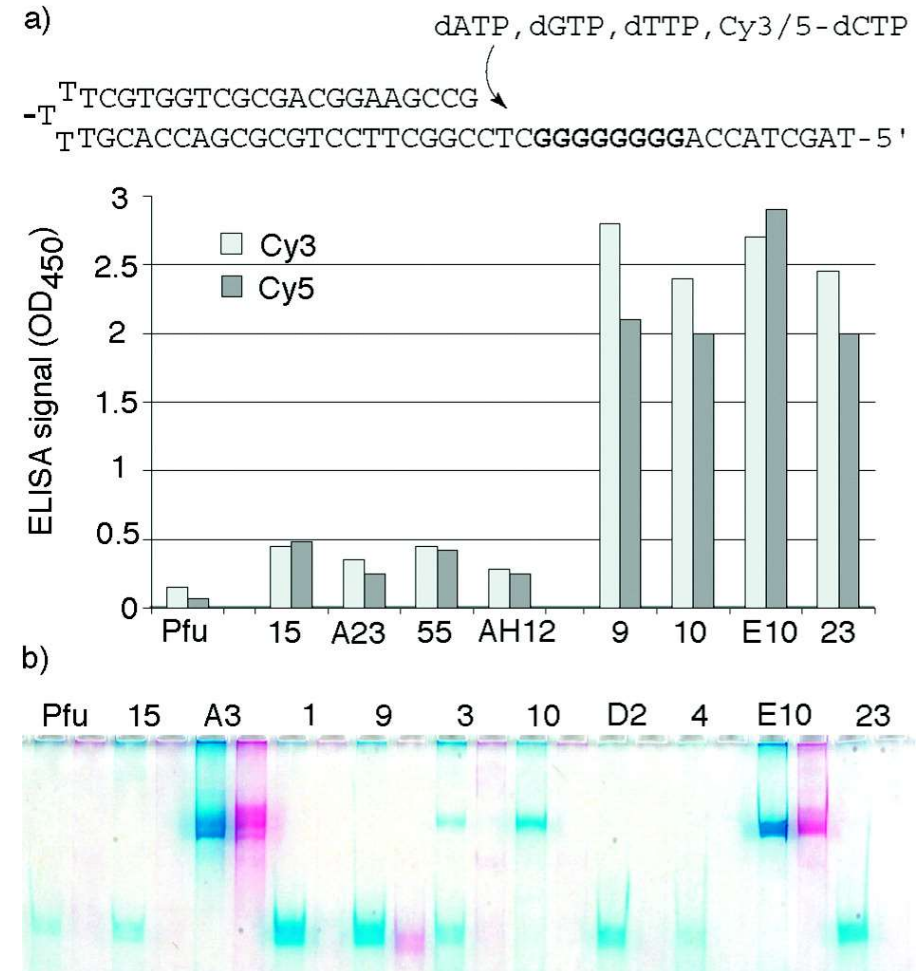
# CyDNA – synthesis and replication of highly fluorescently-labelled DNA

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).

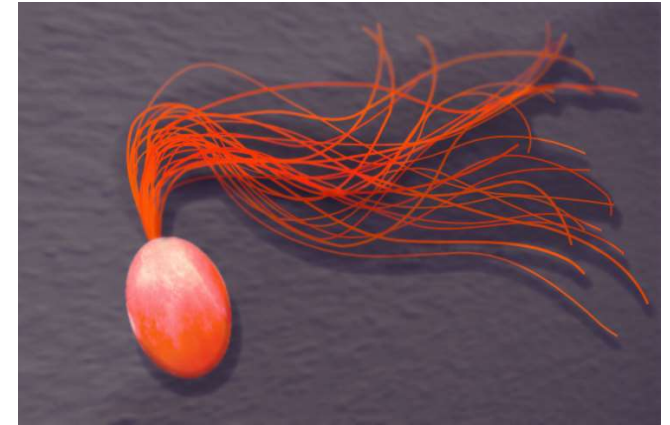
a) activities of round 1 clones (15, A23, 55, AH12) and round 2 clones (9, 10, E10, 23). Clones were chosen i.a. on the their ability to incorporate both Cy3- and Cy5-dCTPs with comparable efficiency.

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.



## 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°C

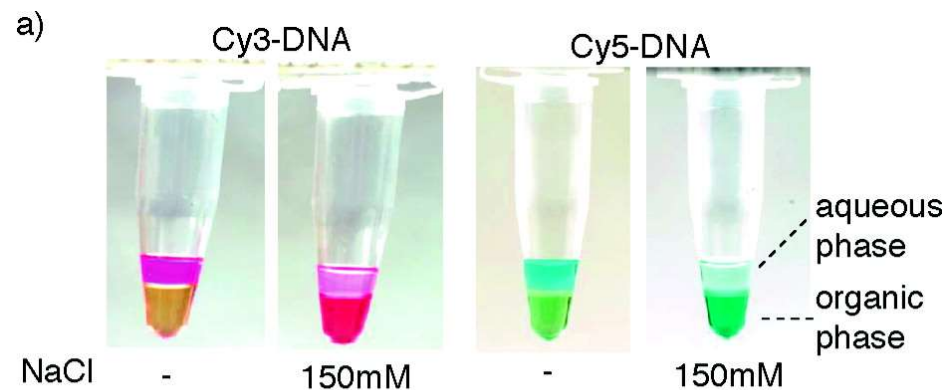
**Table 1.** Polymerase Fidelity

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

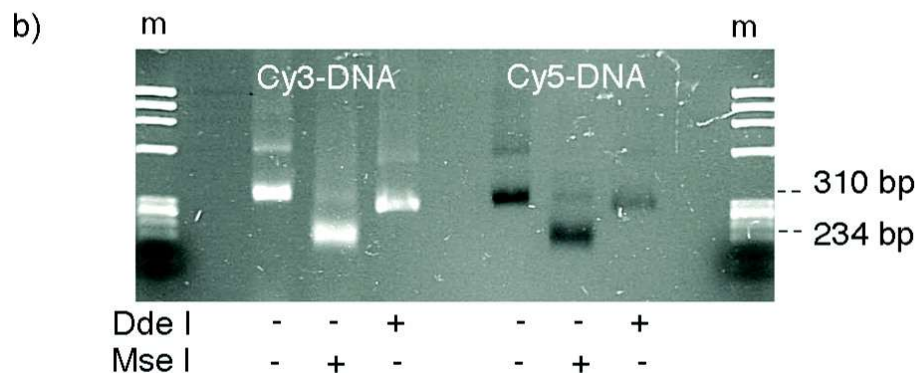
<sup>a</sup> Corrected for the number of doublings (PCR cycles). <sup>b</sup> As determined by a lacZ reversion assay.<sup>44</sup> <sup>c</sup> In the absence of additives (1% formamide, 10% glycerol, 10 µg/mL RNase, 1 mM DTT).

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% partitioning 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).

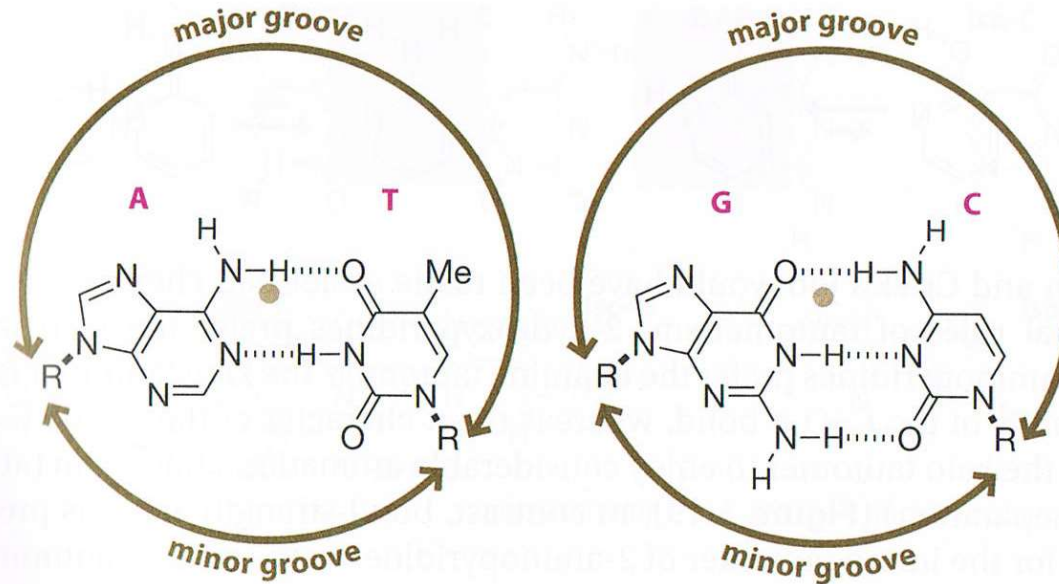


(b) Agarose gel electrophoresis of CyDNA restriction digests. Restriction endonucleases are sensitive probes of noncanonical DNA conformations such as those which occur under torsional strain. Both Cy3- and Cy5-DNA are resistant to cleavage by the restriction endonuclease DdeI (C'TNAG) but are cut by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

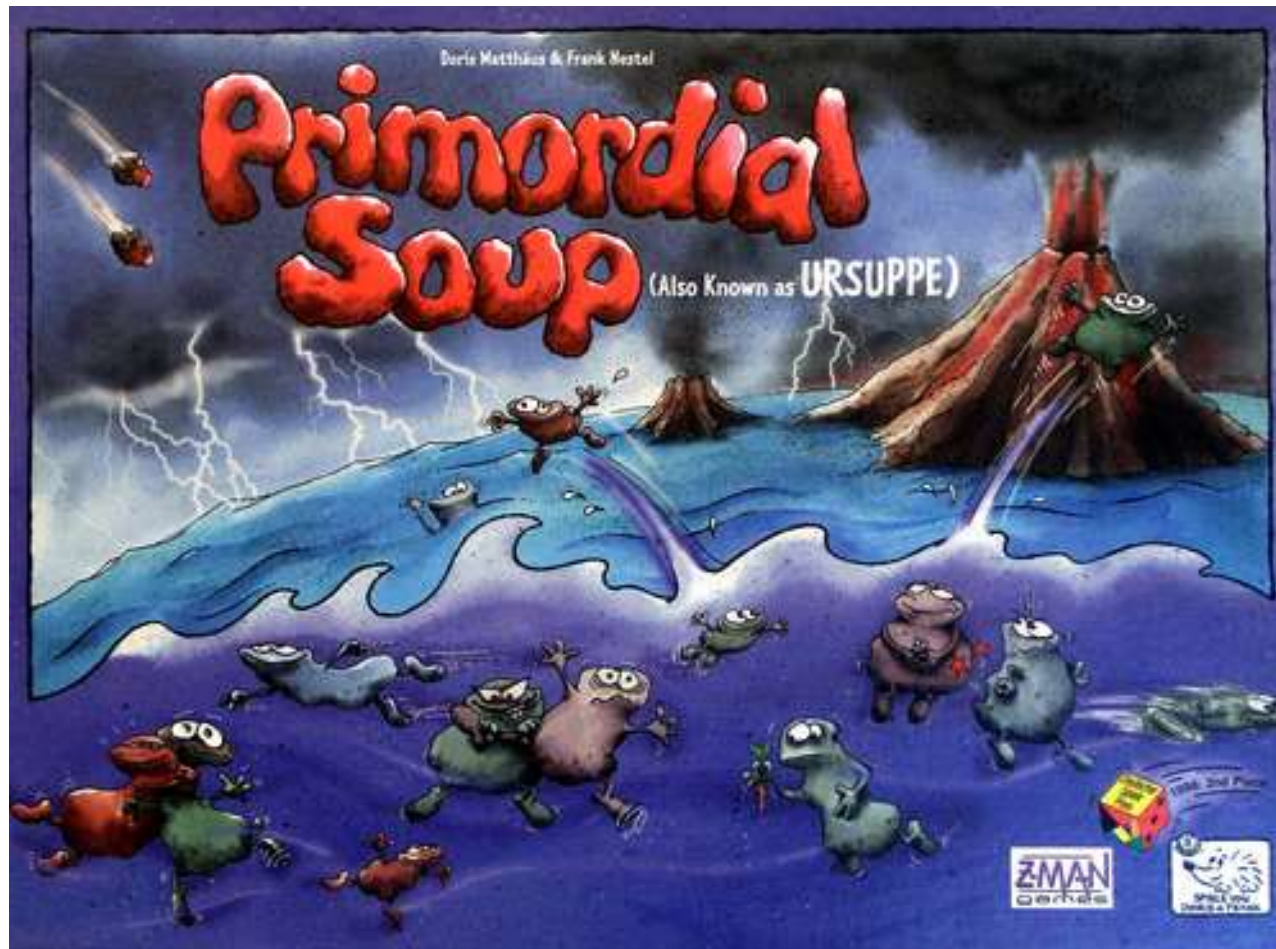


# Why are A, C, G and T the letters of genetic alphabet.

	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



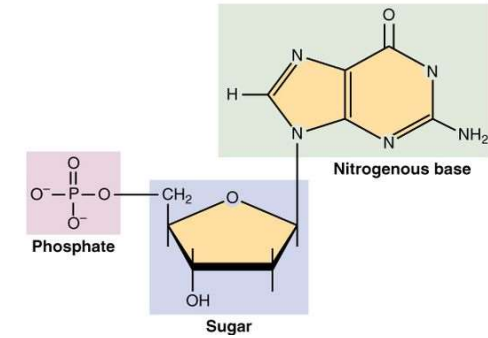
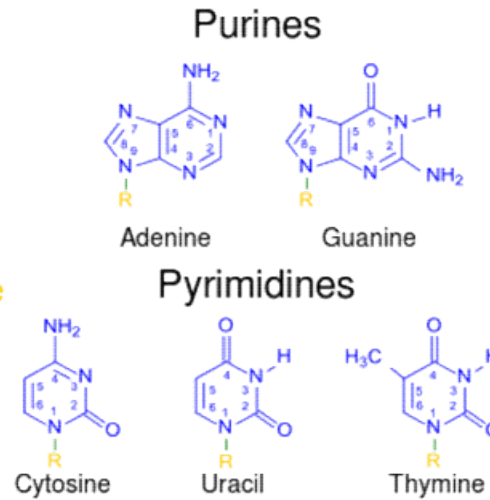
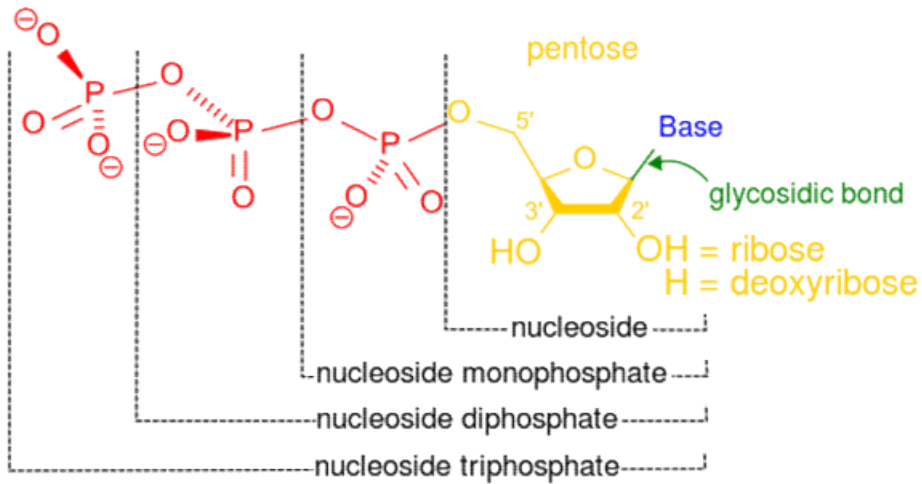
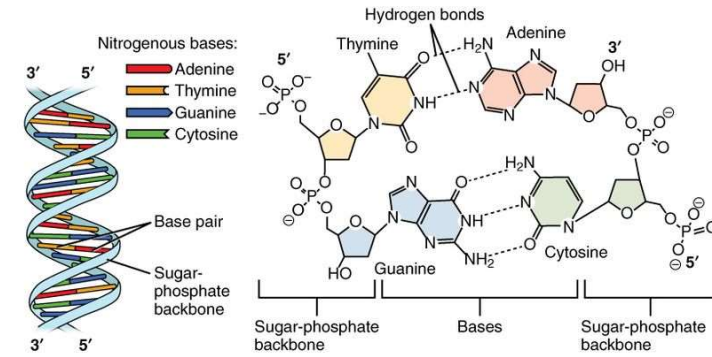
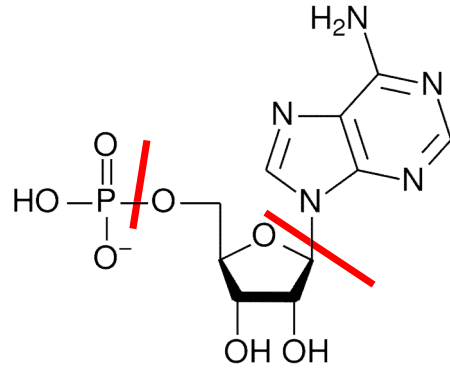
## *Prebiotic synthesis of nucleotides*



*State of the art*

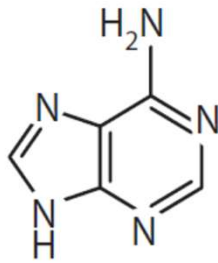
# Nucleotides - components

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

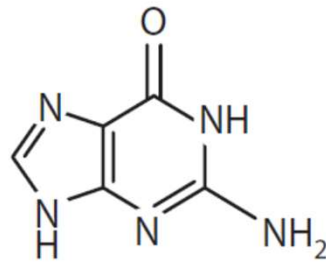


## *Prebiotic synthesis of nucleobases*

### Purines

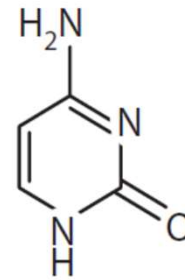


Adenine

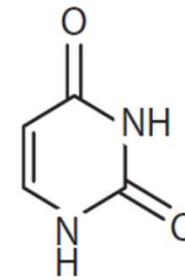


Guanine

### Pyrimidines

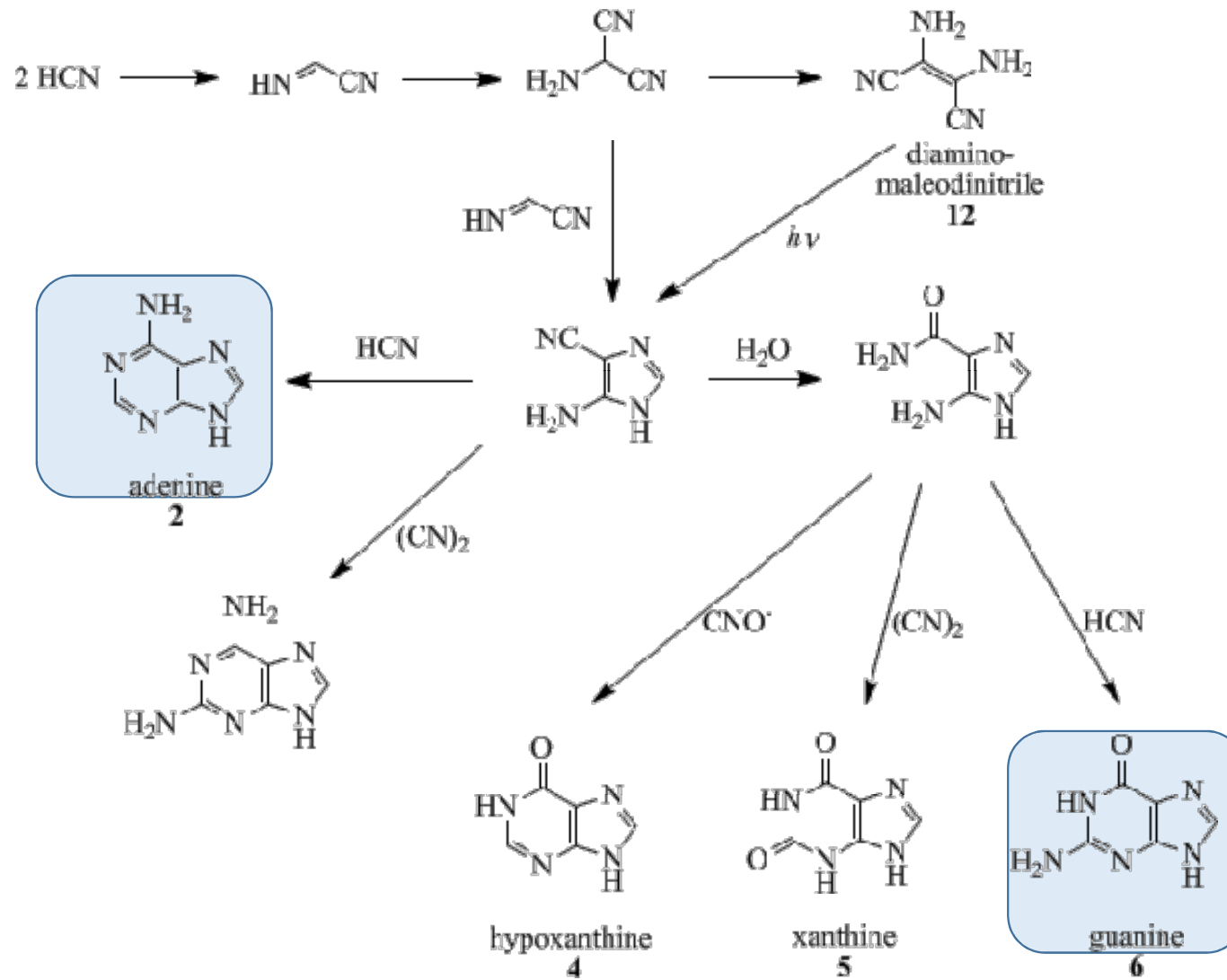


Cytosine

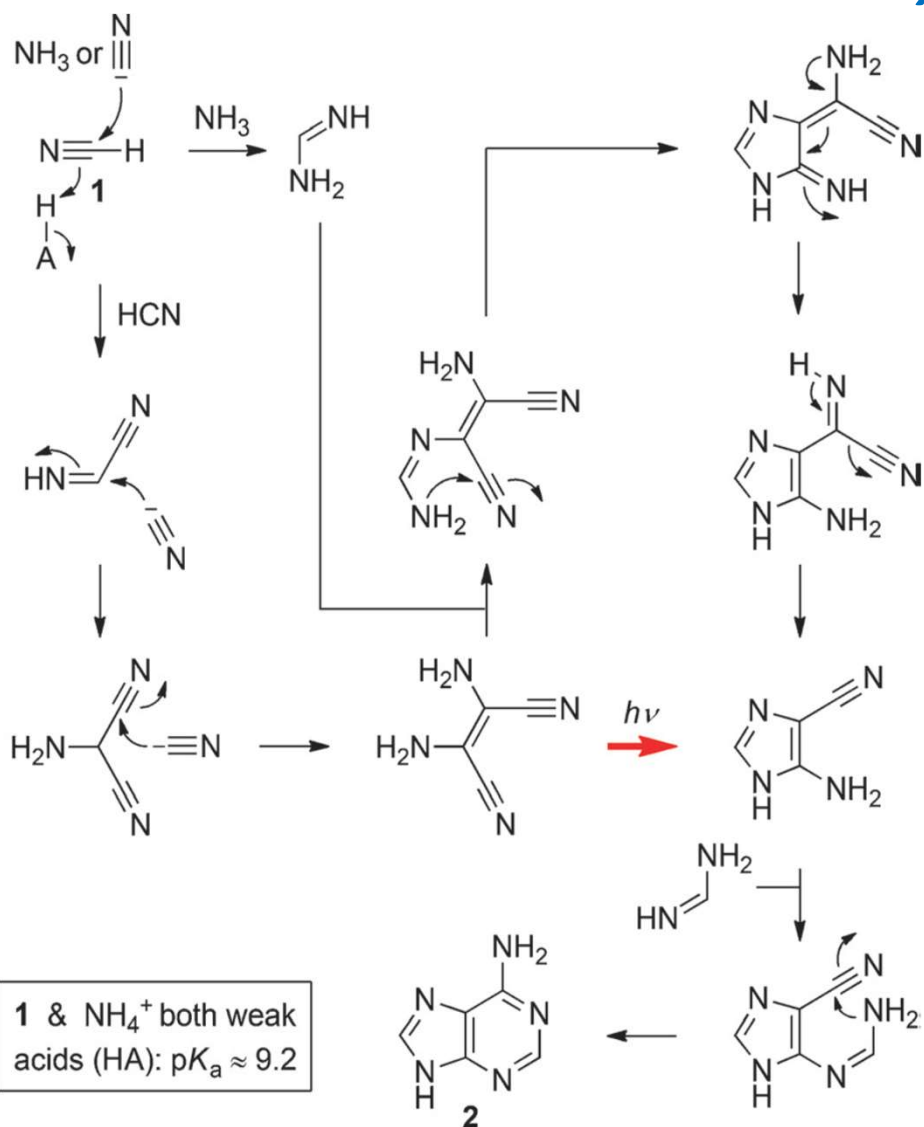


Uracil

## 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  $70^\circ\text{C}$  for a few days

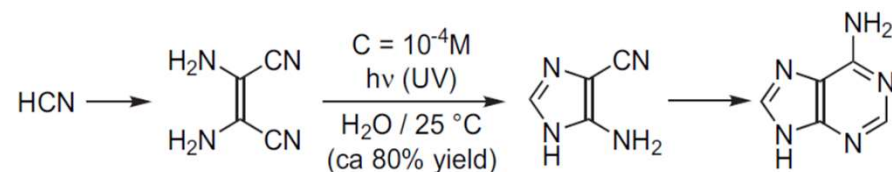
→ 0.5% adenine

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 ( $-20^\circ\text{C}$ ) increases the yield of DAMN formation by concentrating HCN between pure ice crystals

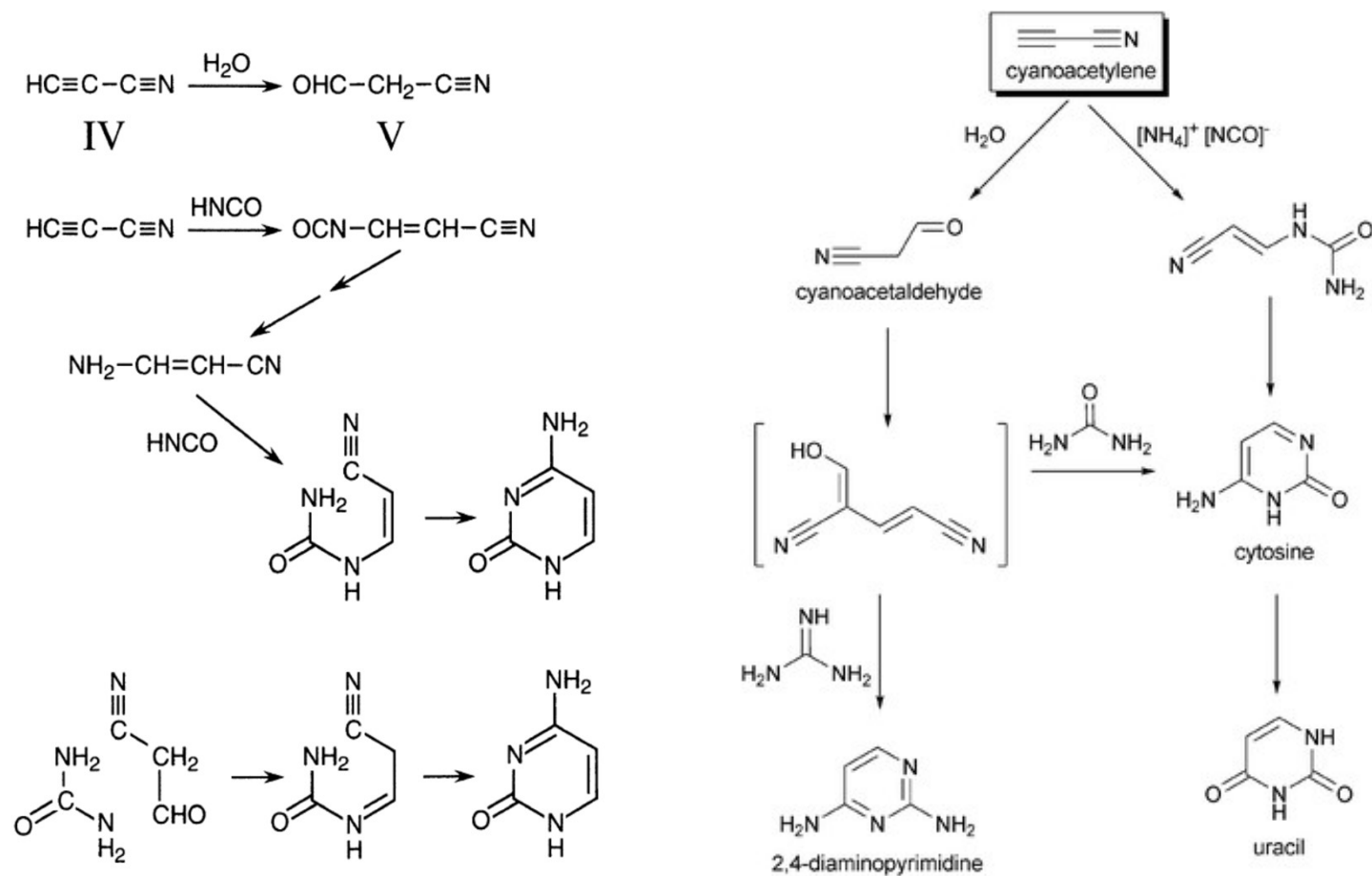


*J. Oro Biochem. Biophys. Res. Commun.* **1960**, *2*, 407.

*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.