Synthetic life

(continuation of „The molecular origins of life" SoSe 2021)

NaturalNews.com

WiSe 2022/23

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;

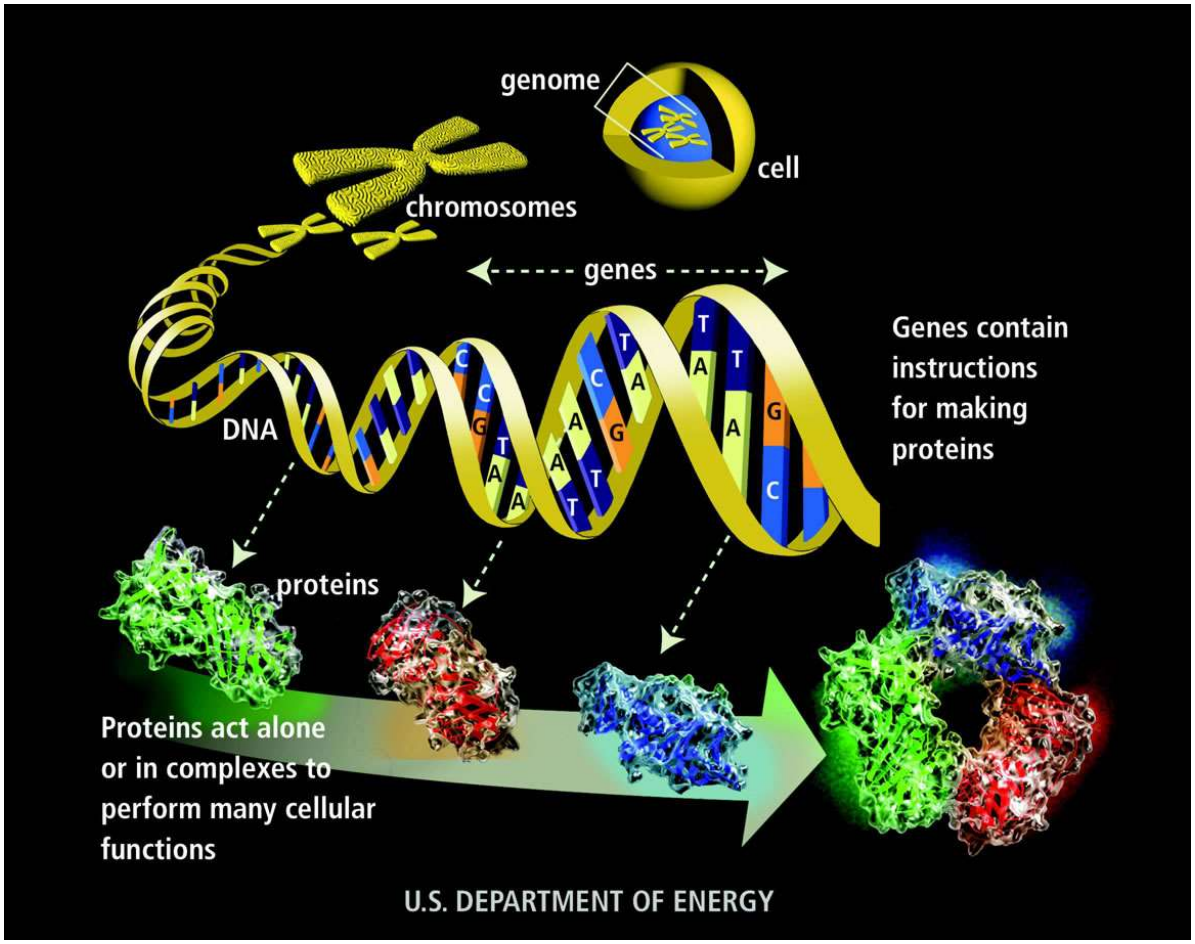
design of artificial organisms

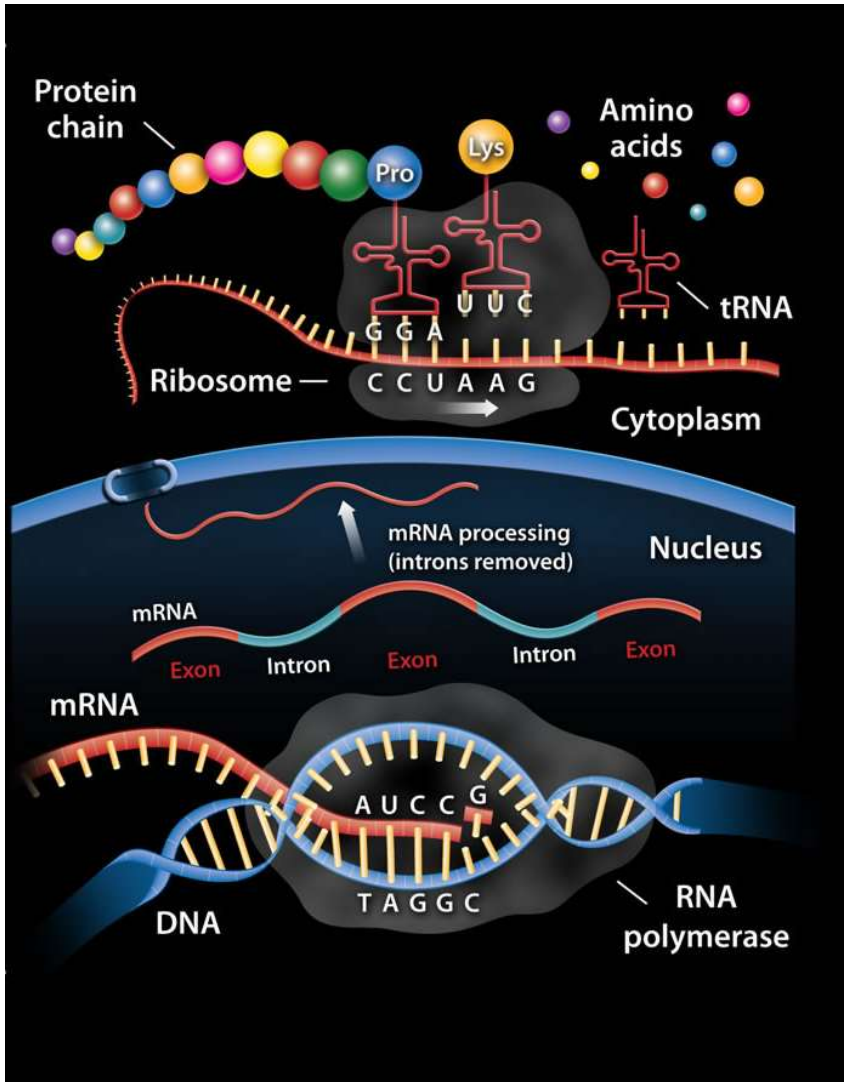
CHAPTER 1

OLIGONUCLEOTIDES

The less common side of RNA

From DNA to proteins



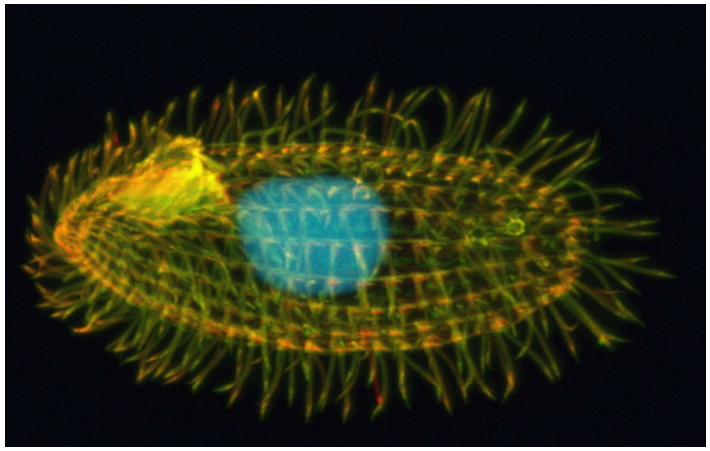


Ribozymes

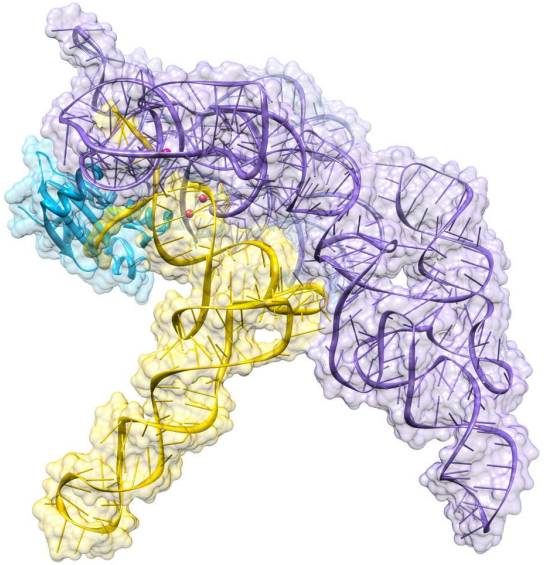
Ribozymes – Ribonucleic acid enzymes

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.

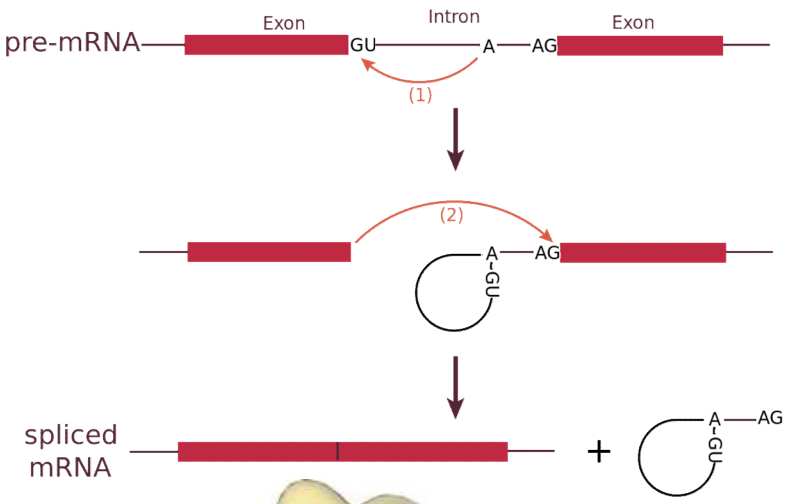


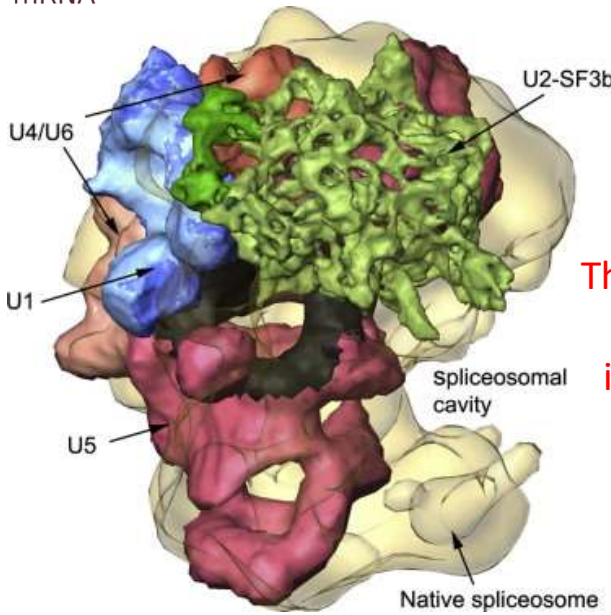
Tetrahymena thermophila



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.

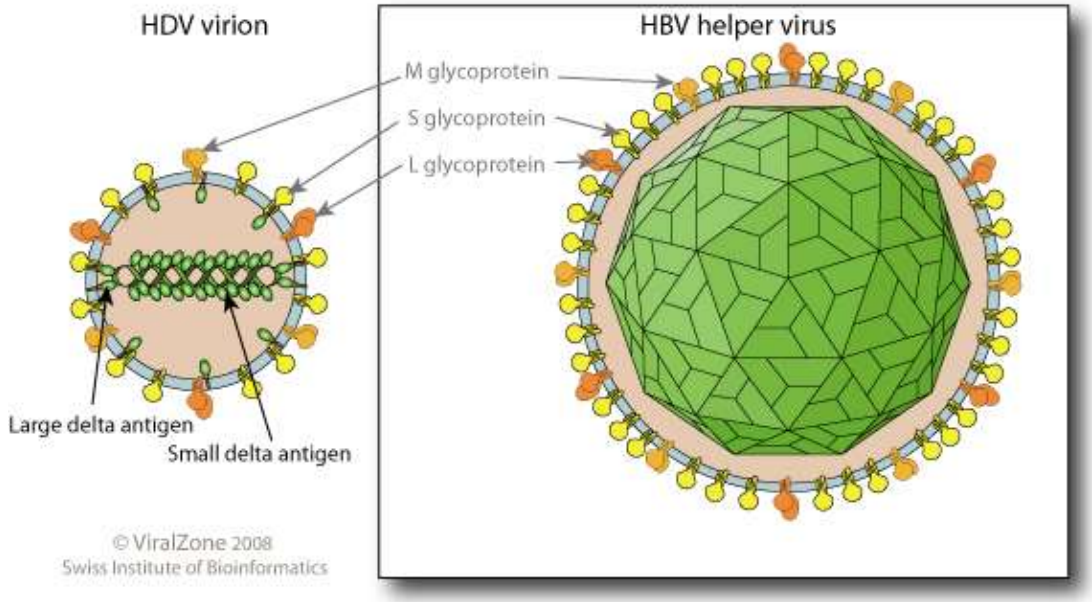
In contrary to the spliceosome, the ***catalytic*** motif ***does not*** contain protein part, ***only RNA***.
First known example of a ***ribozyme*** – ribonucleic acid-composed enzyme analogue.

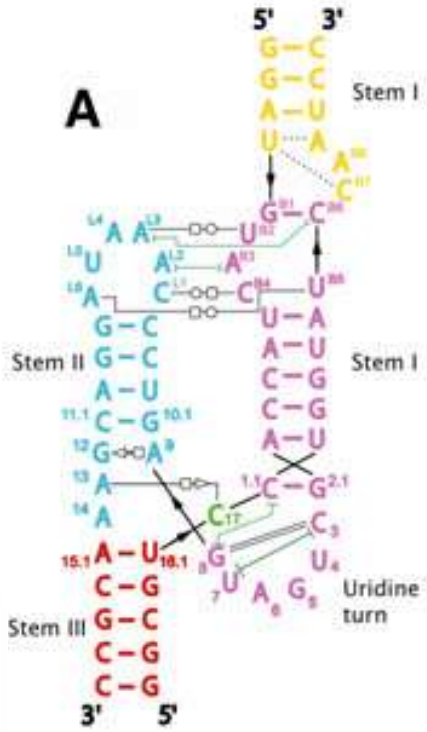
Ribozymes

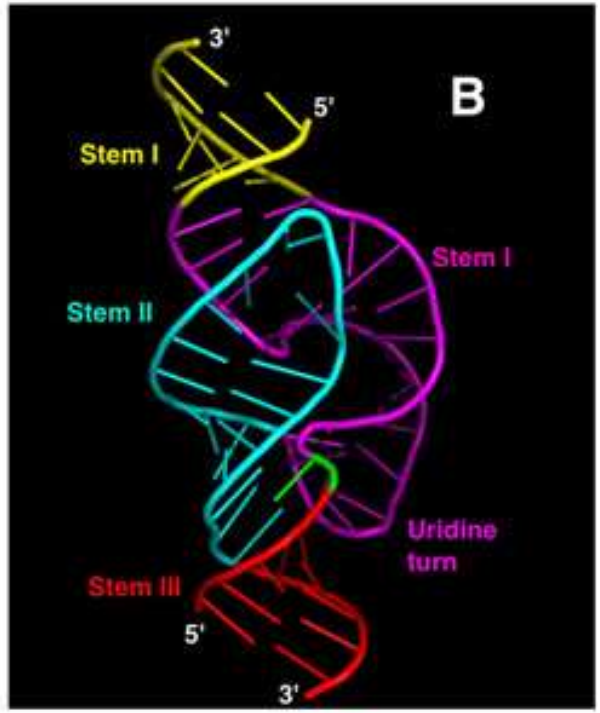
Hammerhead ribozyme

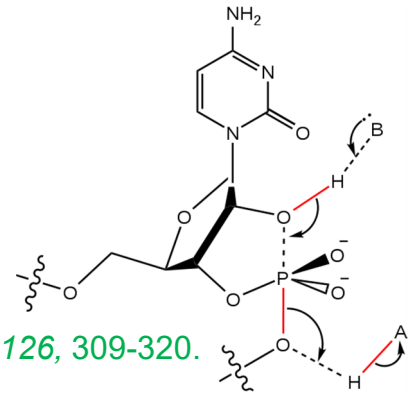
The hammerhead ribozyme is a RNA molecule motif that catalyzes reversible cleavage and joining reactions at a specific site within an RNA molecule (model system; targeted RNA cleavage experiments)

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***.

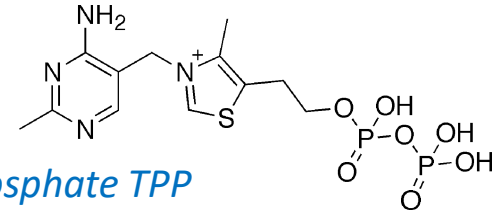
Riboswitches - naturally occurring regulatory segments of mRNA that bind small molecules specifically. The binding results in a change in production of the proteins encoded by the mRNA

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

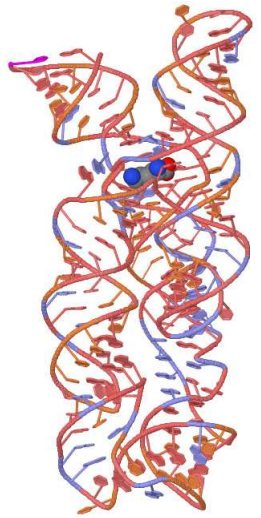
Most known ***riboswitches*** occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in archaea, plants and certain fungi.

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

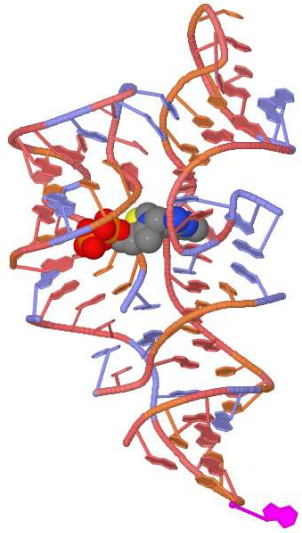
The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



Thiamine pyrophosphate TPP



The lysine riboswitch



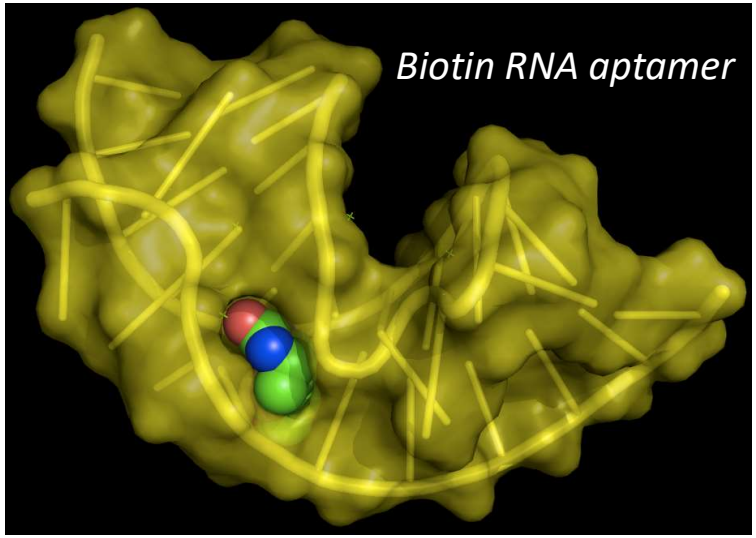
The 3D structure of TPP riboswitch (by Benjamin Schuster-Böckler)

Aptamers

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

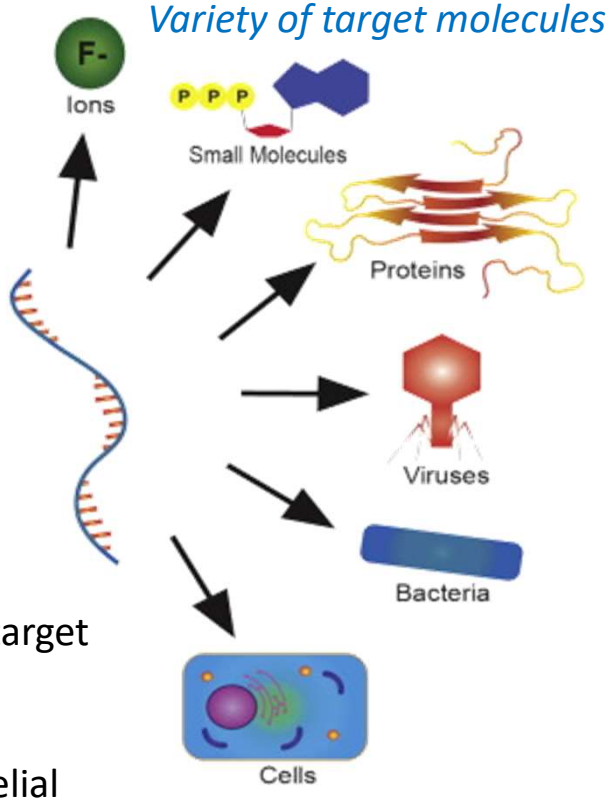
Aptamers are usually created by selecting them from a large random sequence pool, but natural aptamers also exist in riboswitches.

- ***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

Variety of target molecules



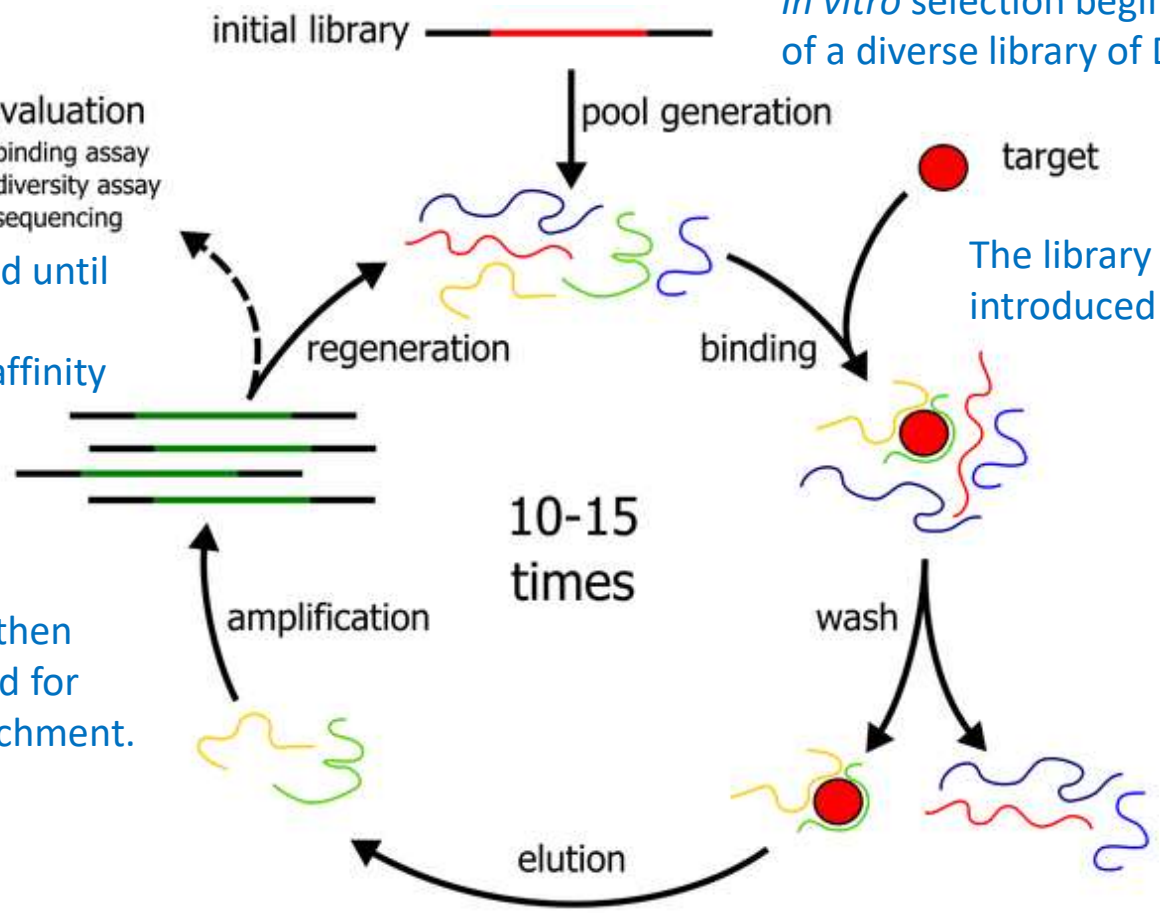
Aptamers were evolved for a variety of target ligands:

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

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

Systematic evolution of ligands by exponential enrichment - SELEX

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

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

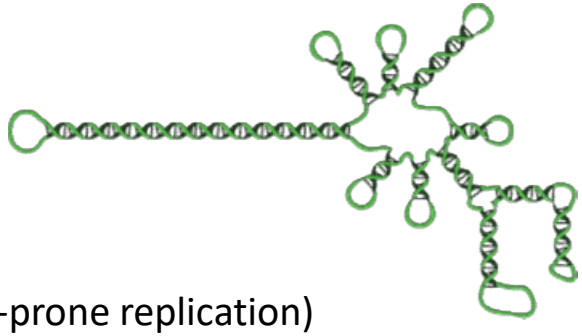
The bound sequences are then collected and PCR amplified for subsequent rounds of enrichment.

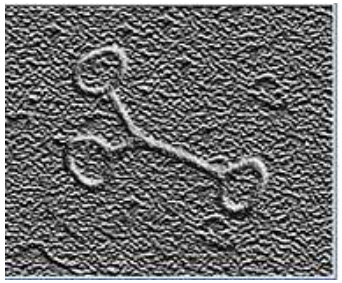
Sequences demonstrating affinity towards the target molecule are isolated from any unbound sequences.

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

Viroids

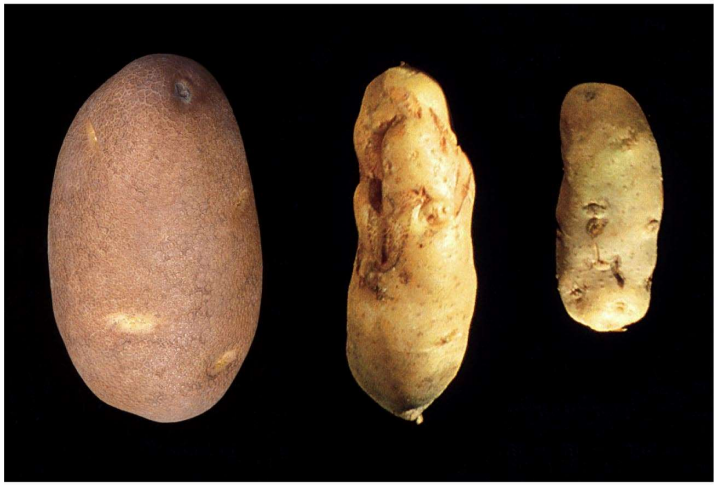
Viroids ("subviral pathogens„) are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Viroids are extremely small - 246 to 467 nucleobases (genomes of smallest viruses start from 2,000 nucleobases). Viroids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



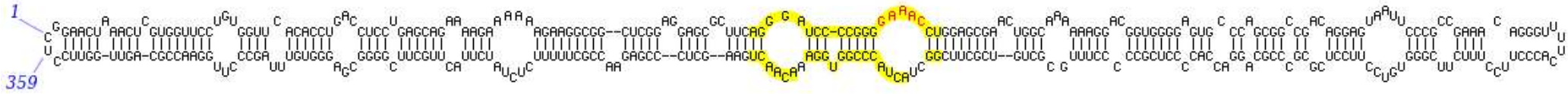


Viroid properties:

- 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.



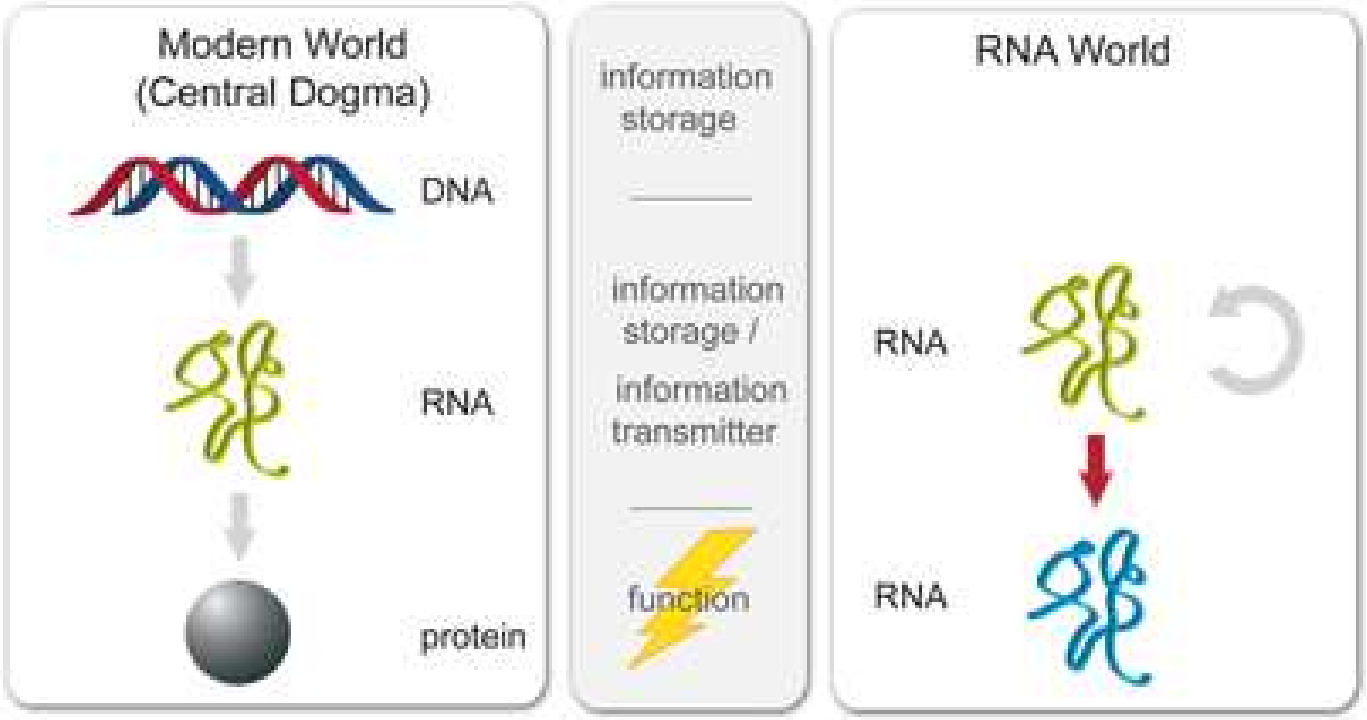
PSTVd-infected potatoes (right)



Putative secondary structure of the PSTVd viroid

No virion reported. Viroids do not encode for proteins

The RNA world



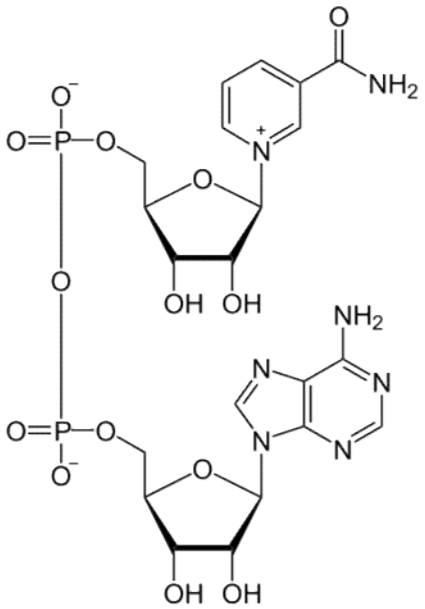
Conceptual idea that there was a period in the early history of life on Earth when RNA (or its structurally simplified analogue) carried out most of the information processing and metabolic transformations needed for biology to emerge from chemistry

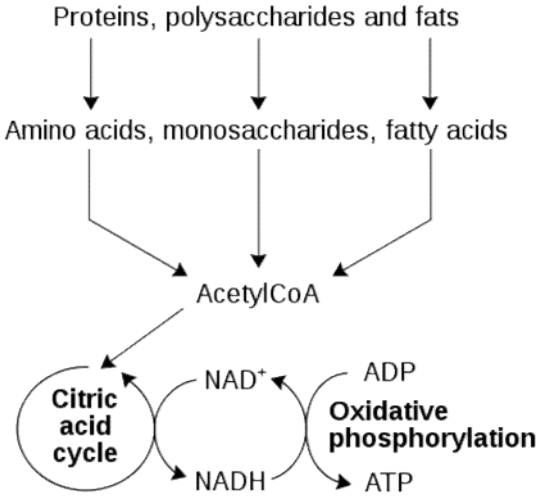
The RNA world

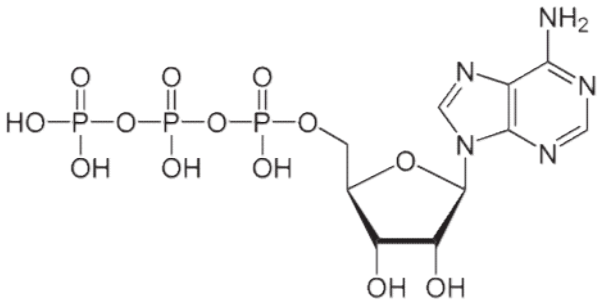
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

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 primoridal RNA-based metabolism







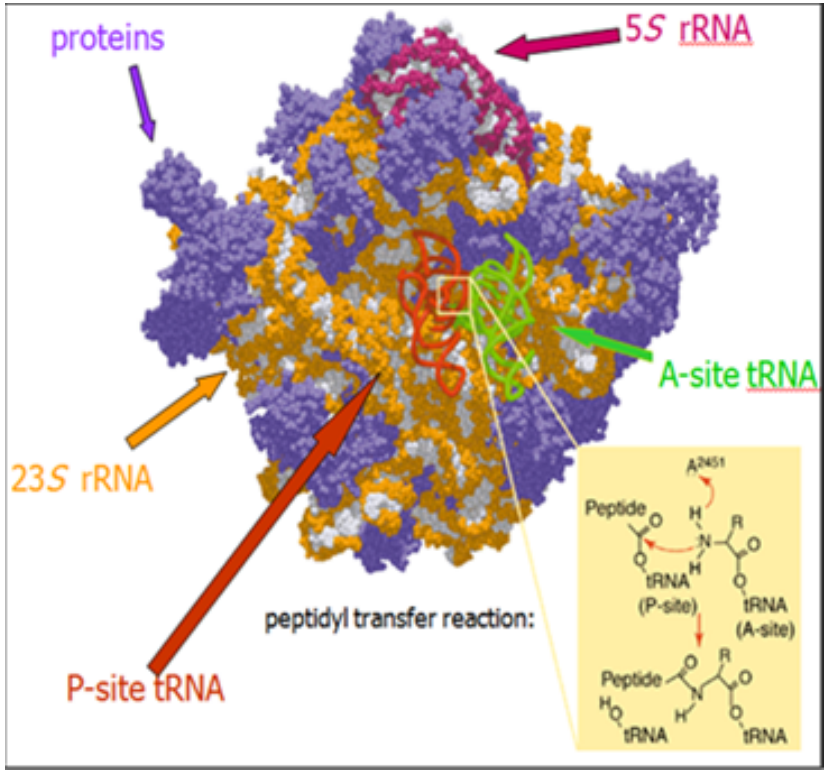
Nicotinamide adenine dinucleotide (NAD$^+$)

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

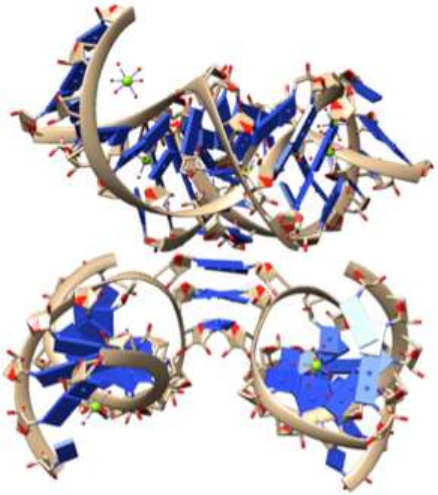
No protein is present within 18 Angstroms from the active site → proteins play a structural role, but
DO NOT CATALYZE THE ACYL TRANSFER PROCESS

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

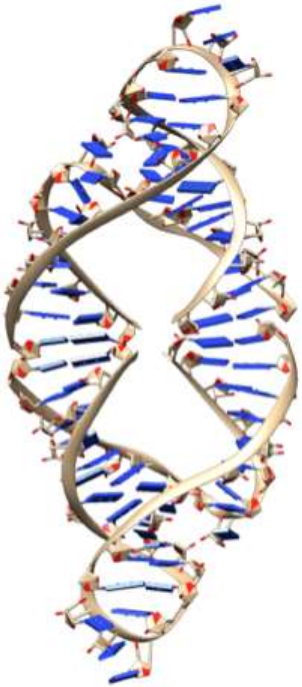
Ribozymes

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

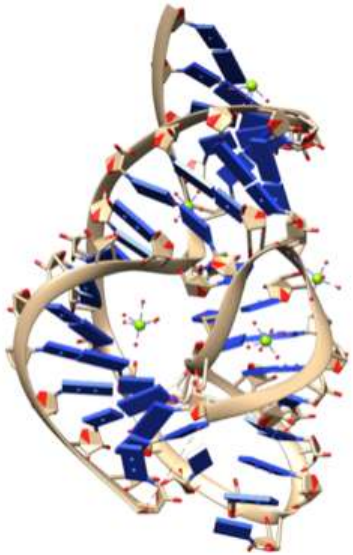
Ribozymes participate in a variety of RNA processing reactions, including RNA splicing, viral replication, and transfer RNA biosynthesis. **Examples** of ribozymes include the **hammerhead ribozyme**, the **VS ribozyme**, **Leadzyme** and the **hairpin ribozyme**.



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.

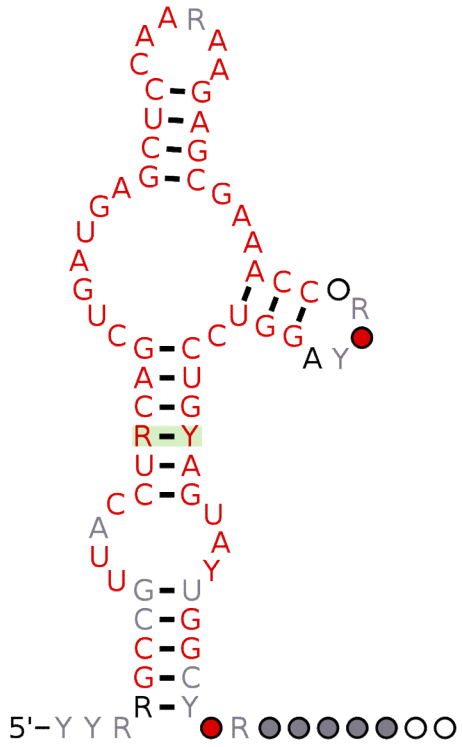
In its natural state, a hammerhead RNA motif is a self-cleaving single strand of RNA (not catalytic, it is consumed by the reaction ➔ no multiple turnovers)

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

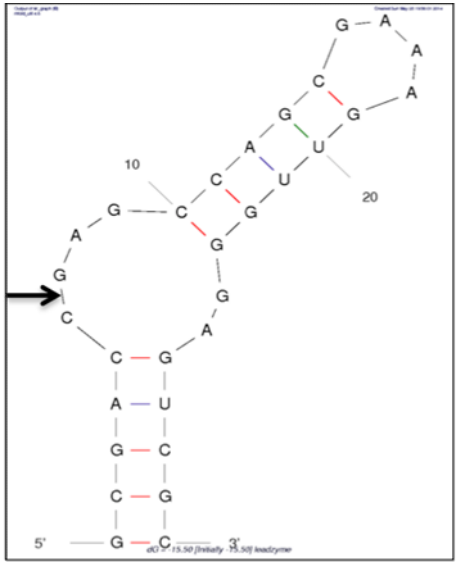
Forster AC, Symons RH, *Cell* **1987**, *49 (2),* 211–220



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



Anarkalimahmood

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

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

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

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

Class I ligase ribozymes

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

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

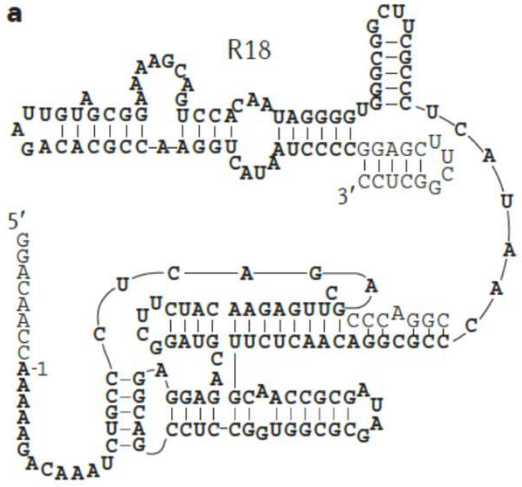
David Schechner

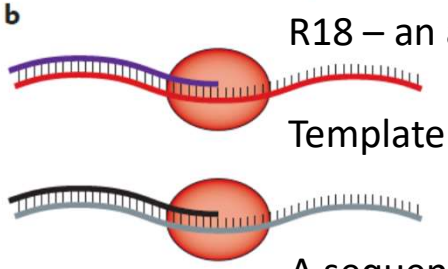
Then, through further evolution, the researcher attempts to coax the ligase to accept NTPs as substrates and to add multiple NTPs in succession. The class I ligase has been evolved further to polymerize as many as 14 successive NTPs with high fidelity

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

The RNA world

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

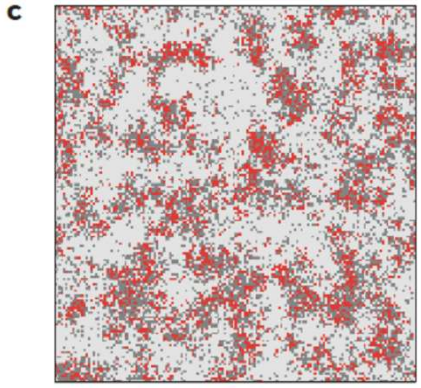


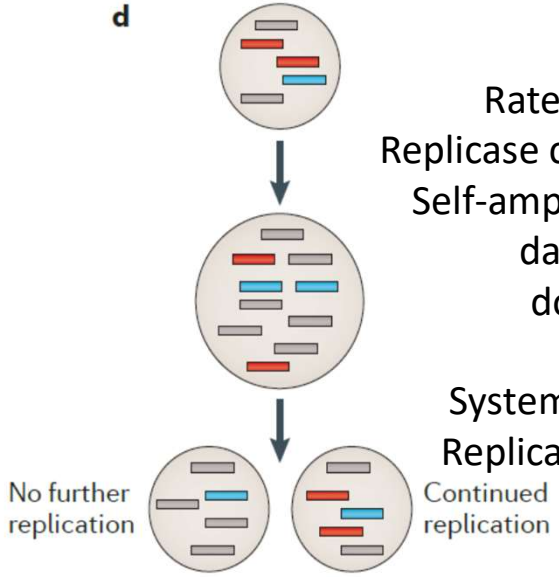


R18 – an artificial polymerase evolved from the class I ligase ribozyme.

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

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





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

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

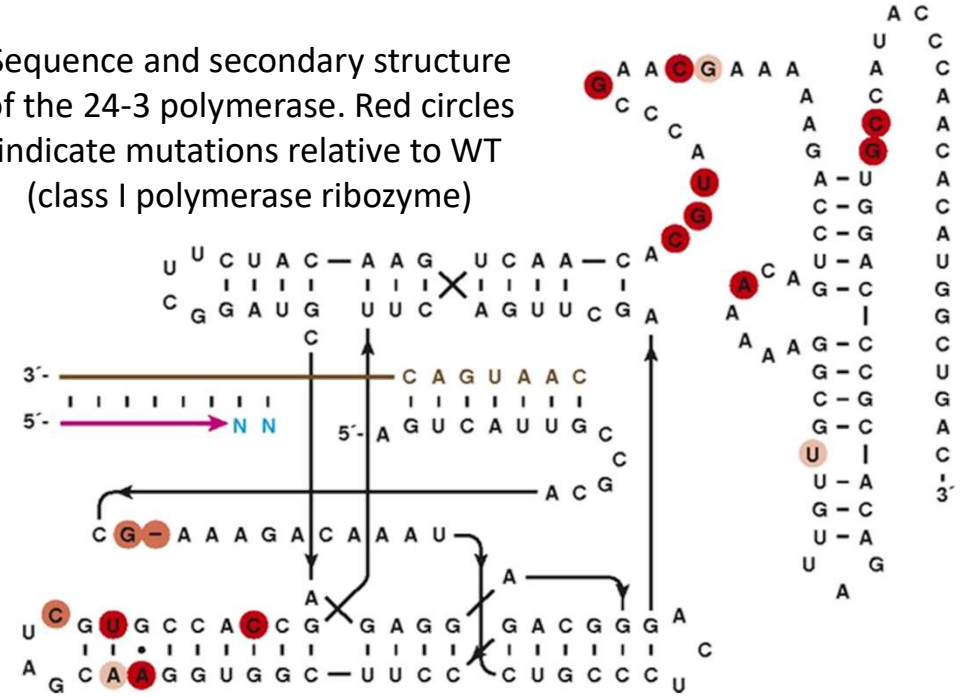
Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **2013**, *5*, 1011–1018.

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)

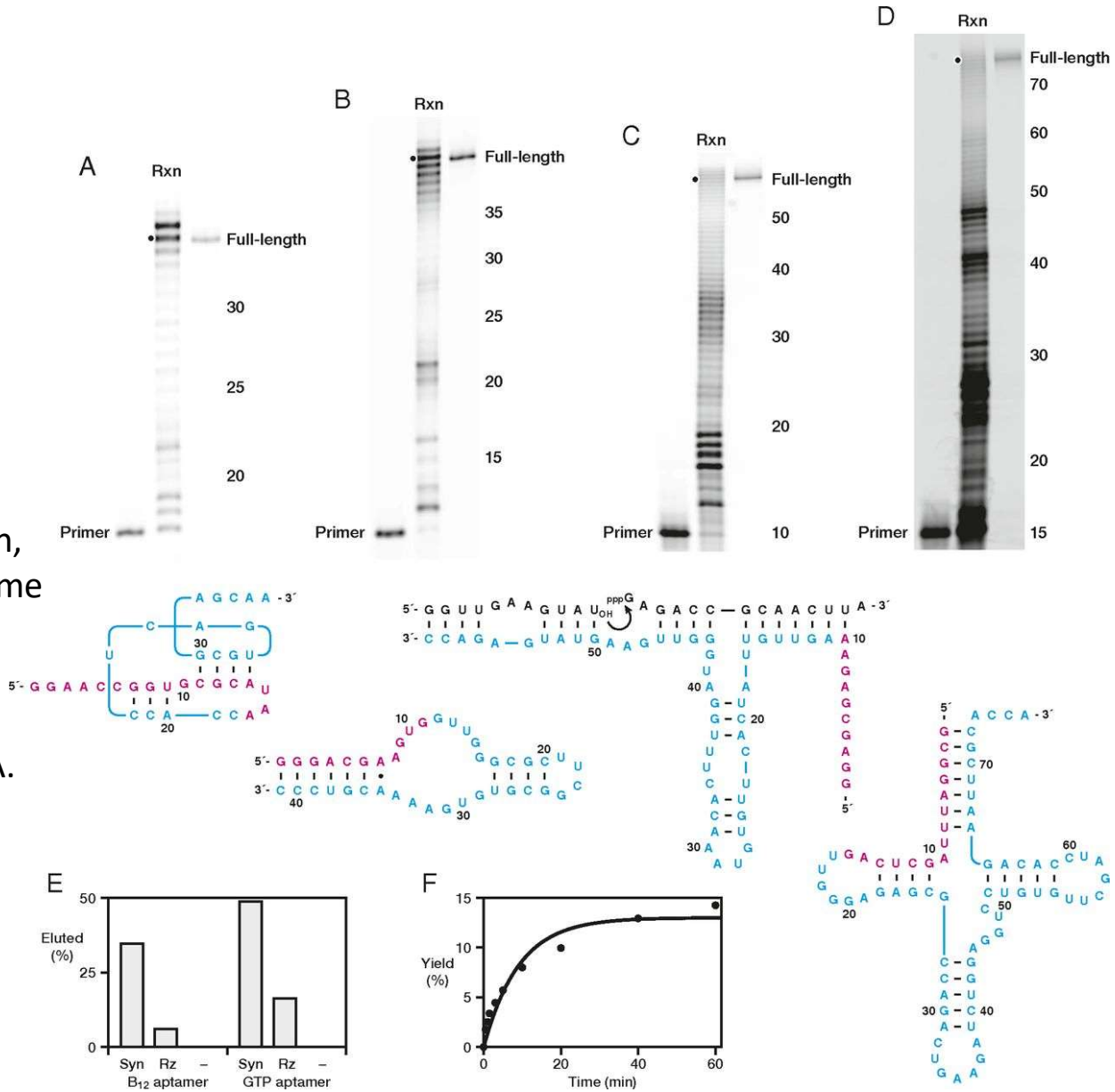


Thus, the replication of RNA and the expression of functional RNA can be accomplished with RNA alone. Combining and improving these activities may enable the self-sustained evolution of RNA and offers a potential route to a synthetic form of RNA life.

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



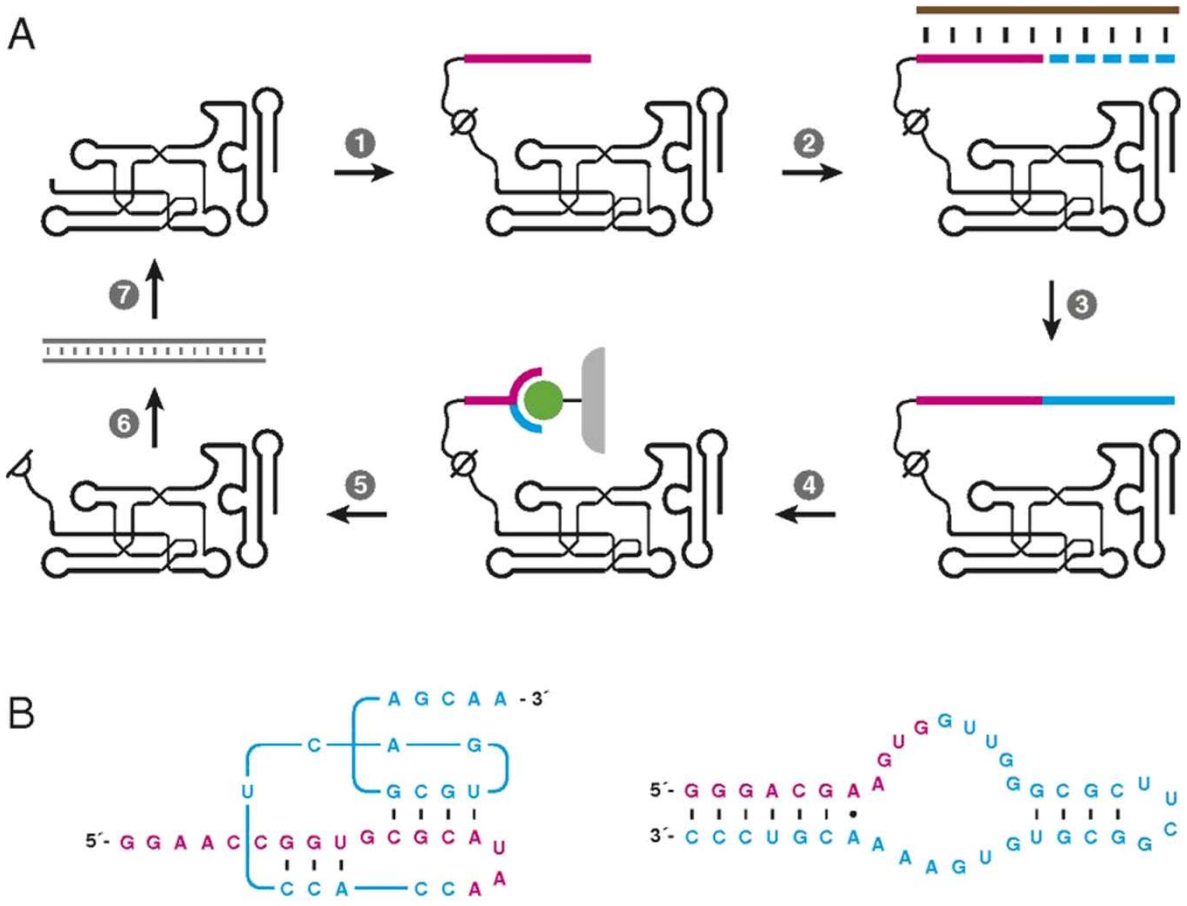
Synthesis of functional RNAs by the 24-3 polymerase. Synthesis of (*A*) the cyanocobalamin aptamer after 24 h, (*B*) the GTP aptamer after 24 h, (*C*) the F1 ligase ribozyme after 24 h, and (*D*) yeast phenylalanyl-tRNA after 72 h.

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).

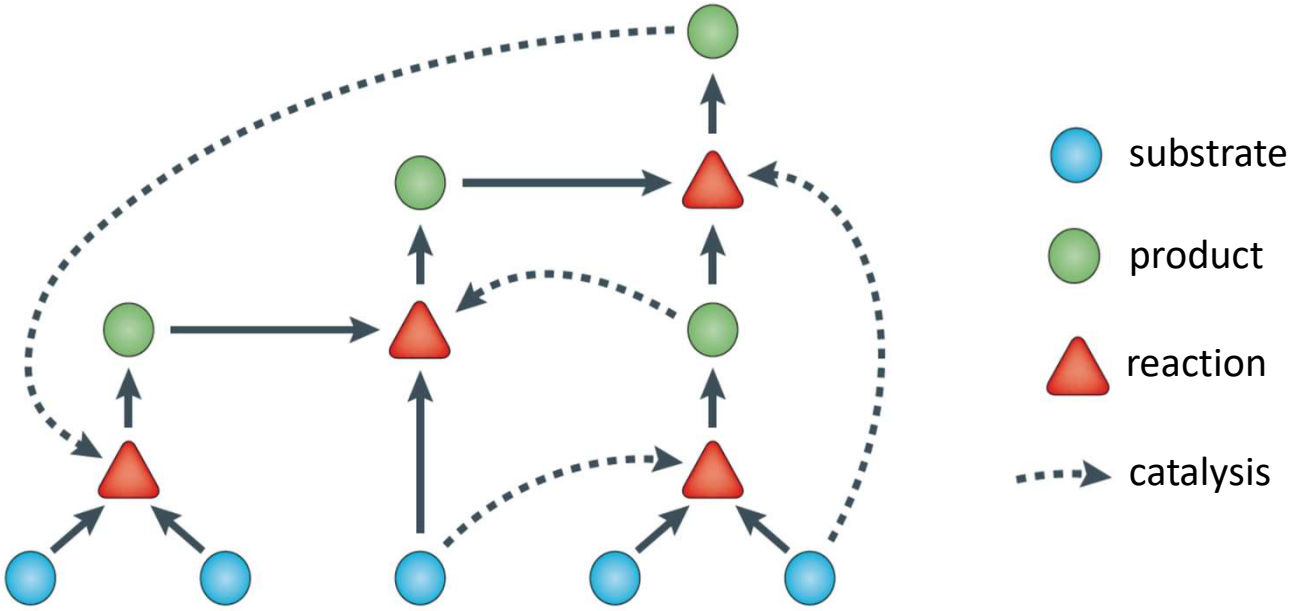
D. P. Horning, G. F. Joyce *Proc. Natl. Acad. Sci. USA (PNAS)* **2016**, *113 (35)*, 9786-9791

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?

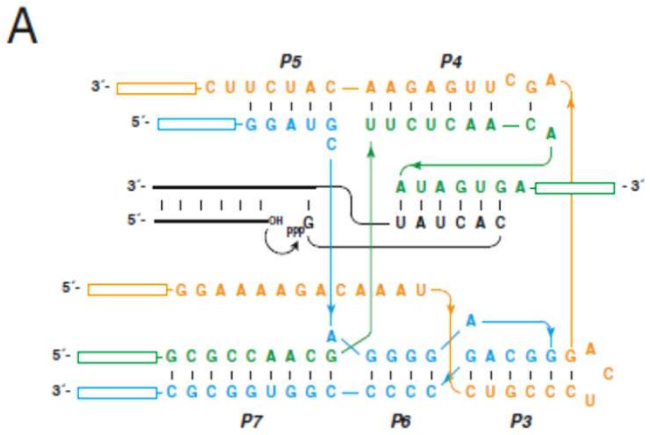
Possible solution – autocatalytic networks



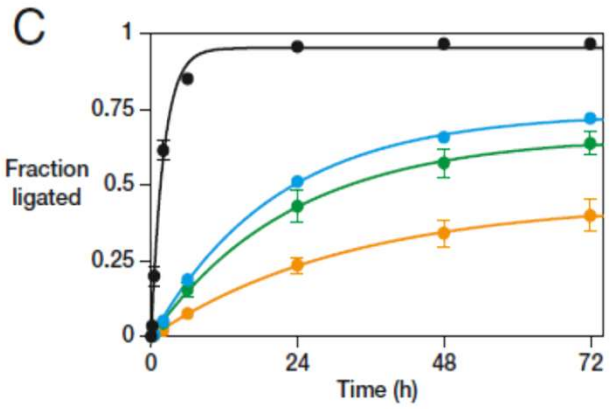
No component can replicate without all the others

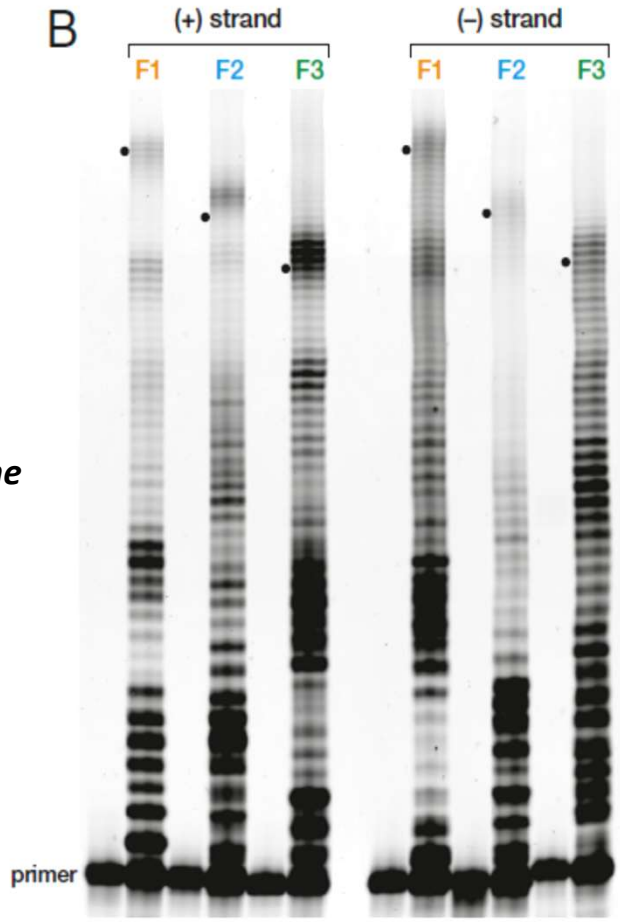
The RNA world

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

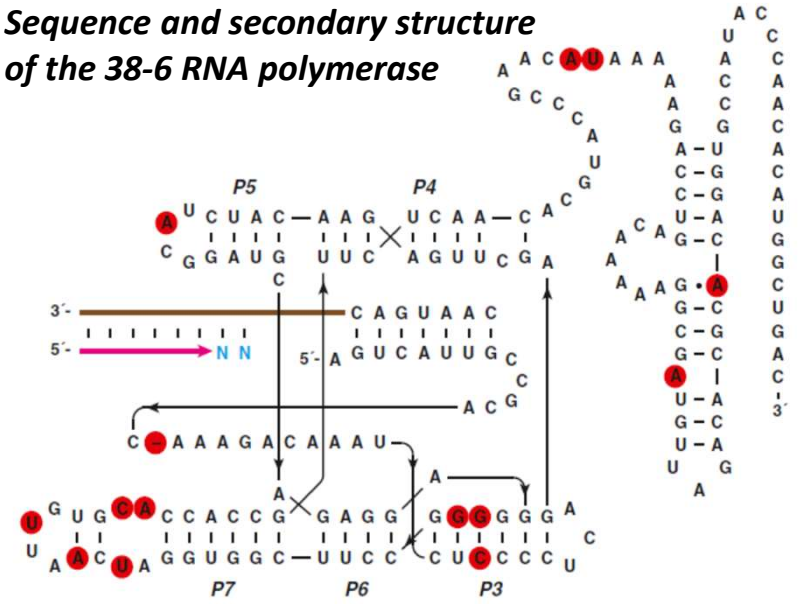


Three fragments of the class I ligase ribozyme





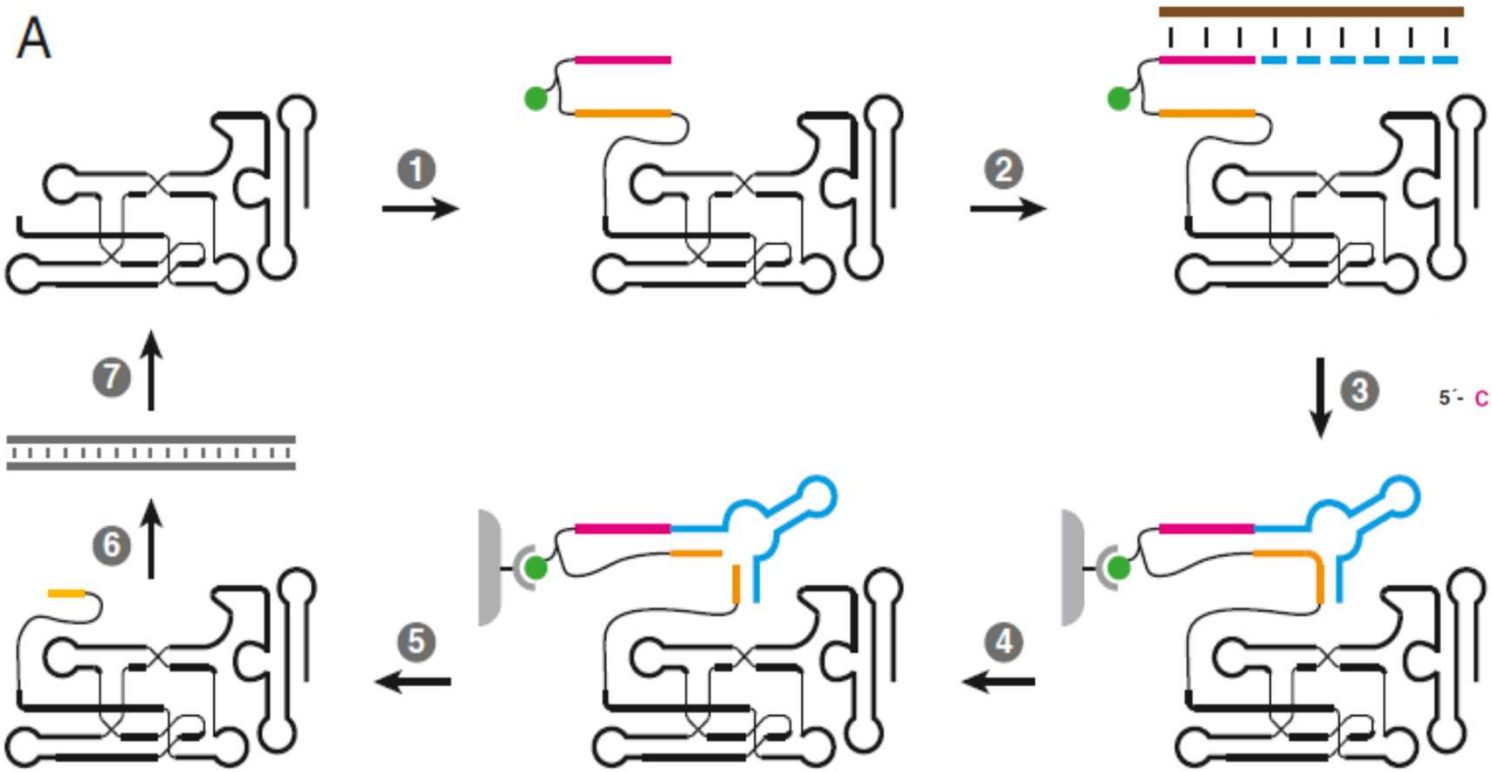
Sequence and secondary structure of the 38-6 RNA polymerase



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

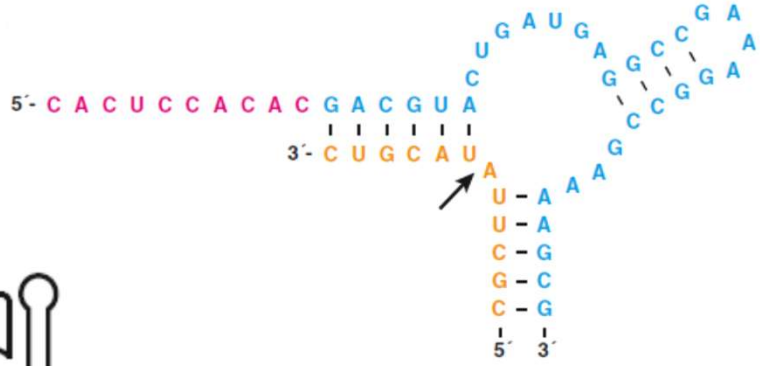
The polymerase „38-6” is able to synthesize its own evolutionary ancestor, an RNA ligase ribozyme, in the form of three fragments that assemble to give a functional complex



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

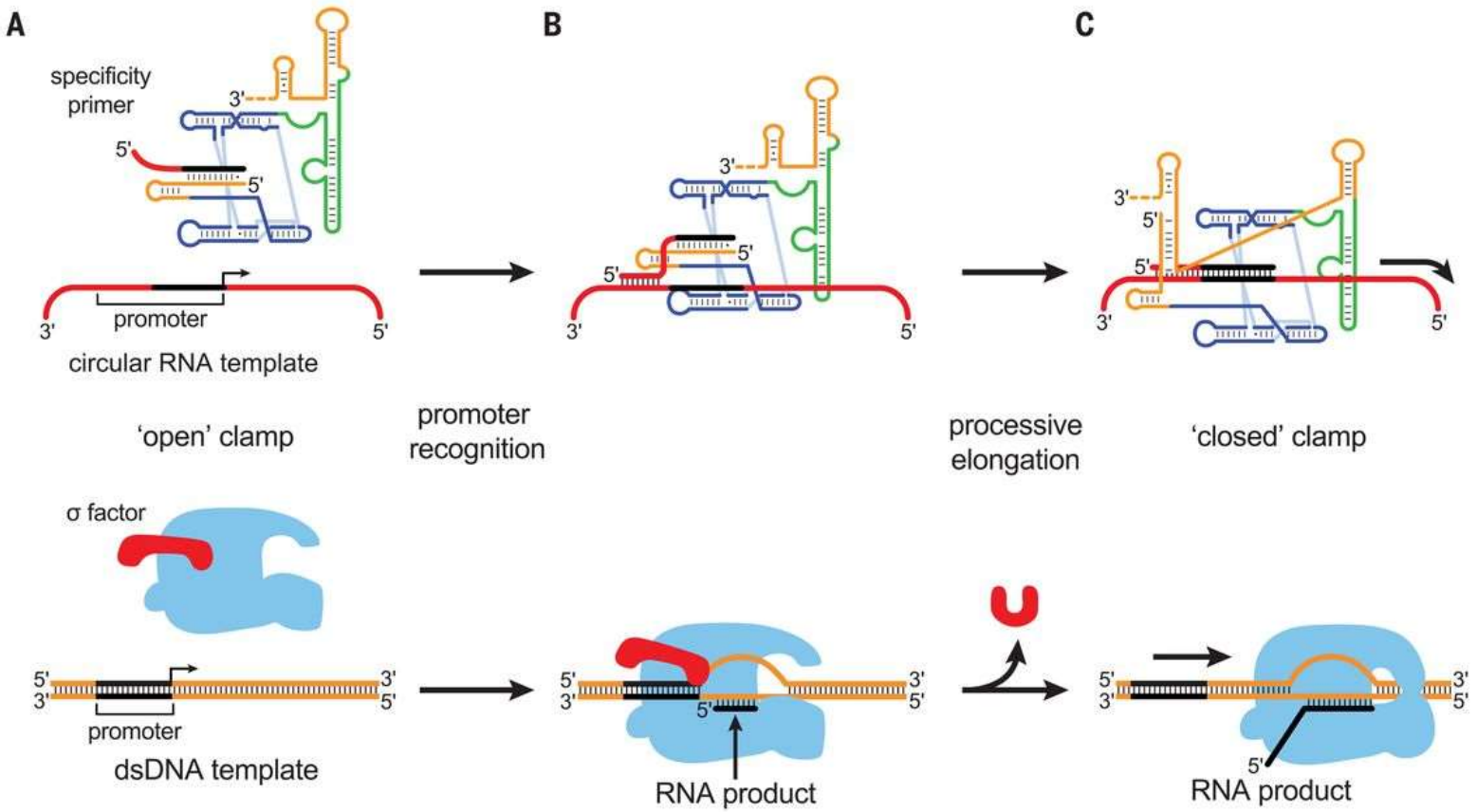
The hammerhead ribozyme:



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

The RNA world

Processive RNA polymerization and promoter recognition

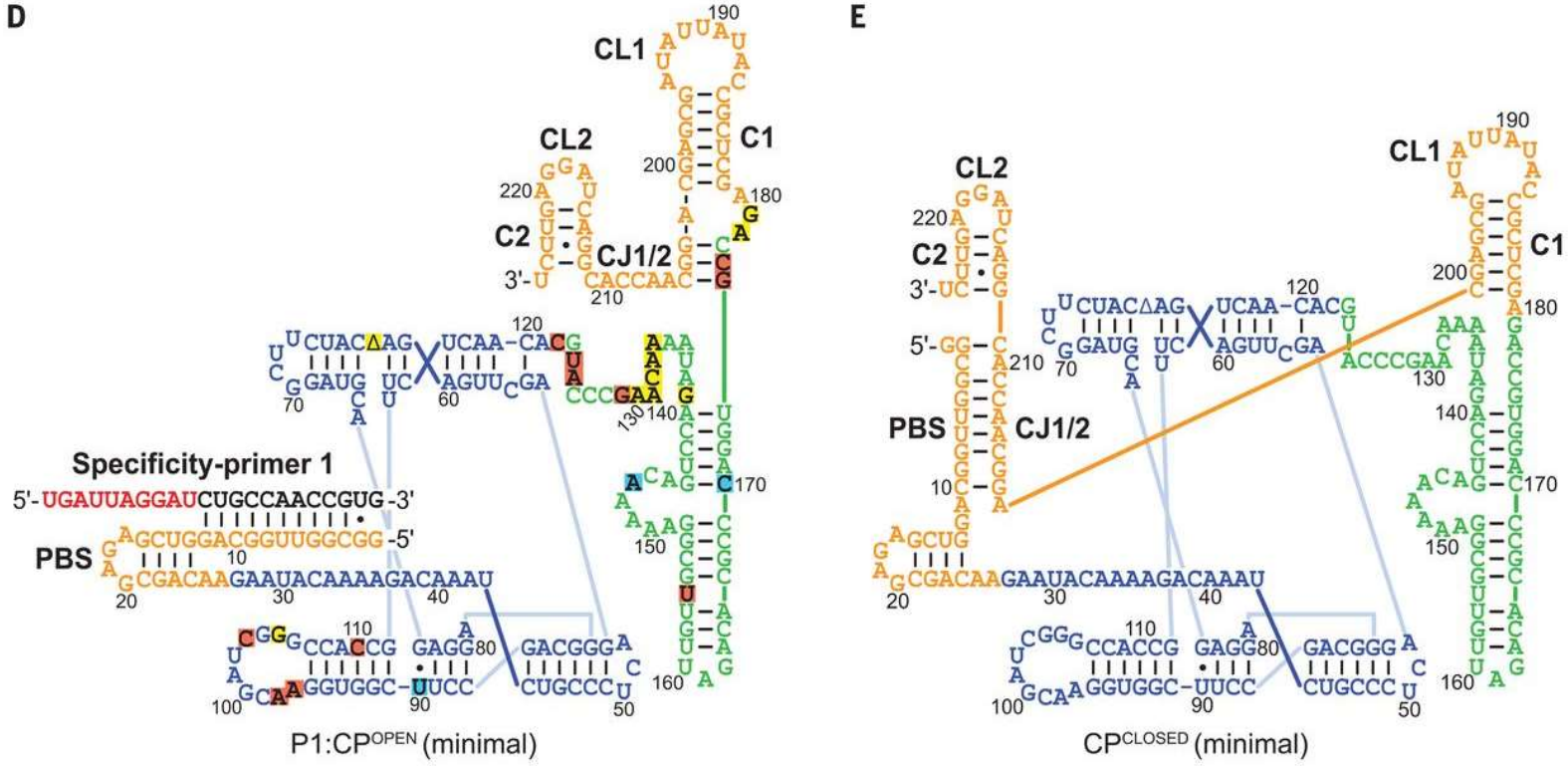


(**A**) RNA specificity primer–activated "open" form P1:CP^{OPEN} (top) and DdRP holo-enzyme (bottom). (**B**) The specificity primer localizes CP to an ssRNA promoter (top), whereas a sigma factor localizes the DdRP to a DNA promoter (bottom). (**C**) In both cases, a clamped "closed" state forms, enhancing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

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

The RNA world

Processive RNA polymerization and promoter recognition

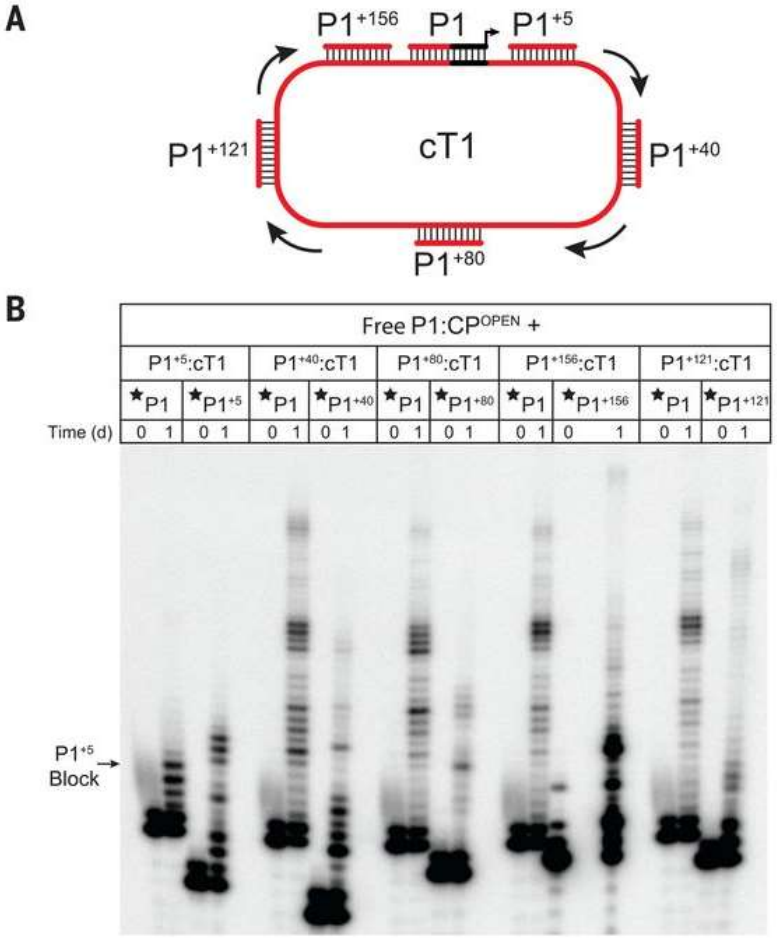


(**D** and **E**) Secondary structure of the minimal P1:CP^{OPEN} form (D) and the CP^{CLOSED} form (E). Colored lines indicate the ligase core (blue), accessory domain (green), and minimal clamping domain (orange). Up mutations designed into the selection are shown in red boxes, rediscovered up mutations 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^{OPEN} and P1^{+n} (n = 5, 40, 80, 121, and 156) primers added to cT1. Either P1 or the P1^{+n} primers were radiolabeled as indicated (star).

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

DNAzymes

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

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

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

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

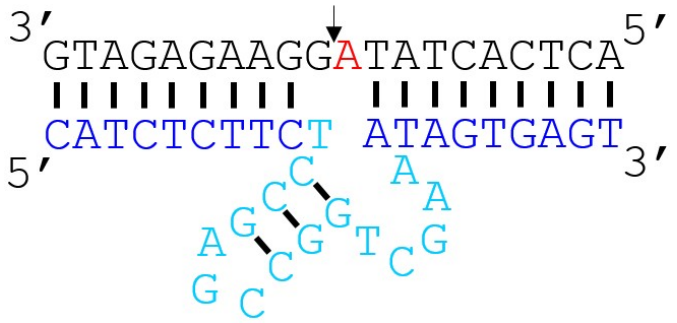
The apparent lack of naturally occurring deoxyribozymes may also be due to the primarily double-stranded conformation of DNA in biological systems → limited physical flexibility and ability to form tertiary structures (catalytic 3D-folds)

DNAzymes

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

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

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



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

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

Currently known:

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

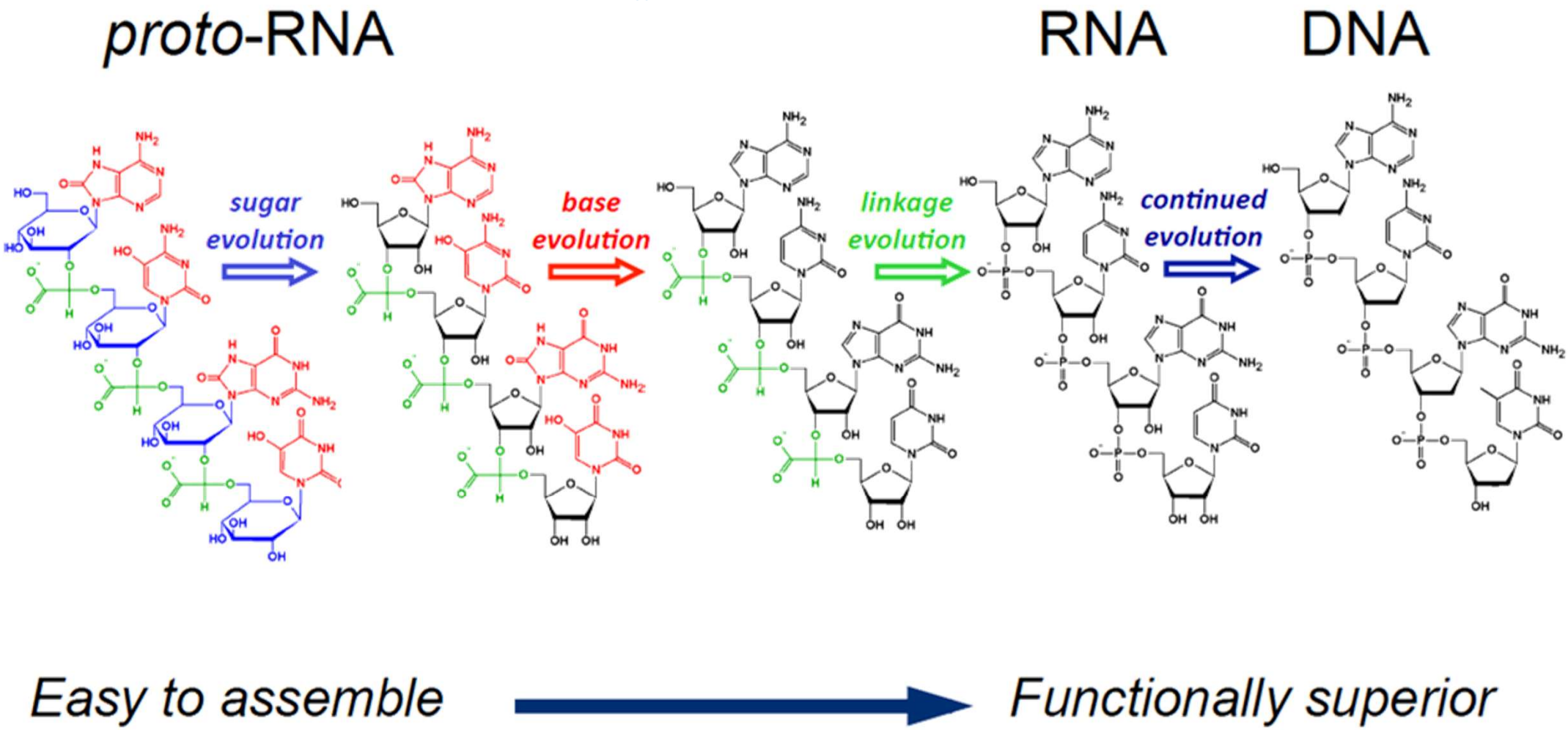
Problems: product inhibition, often single-turnover

The XNA world

Synthetic biology:

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

„RNA-second”



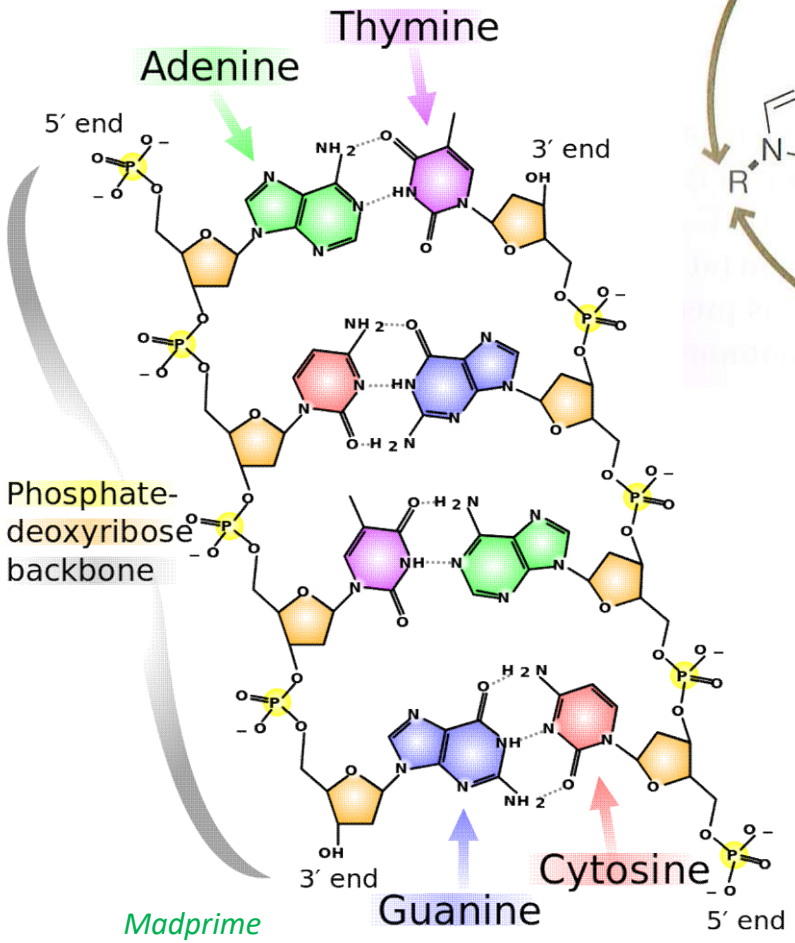
Proto-RNA evolution: According to the protoRNA theory, each of the components of RNA — sugar, base and phosphate backbone — may have originally taken different forms.

CHAPTER 1

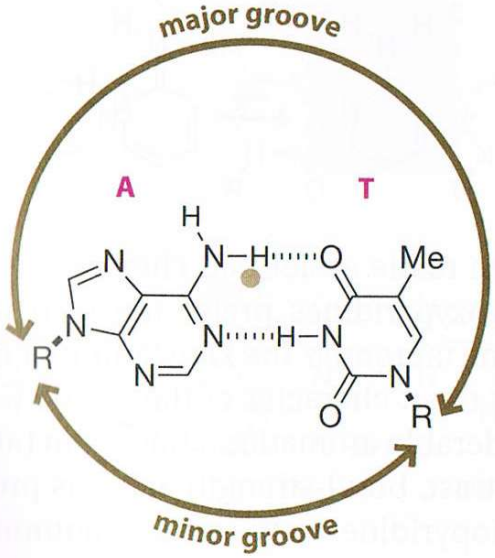
OLIGONUCLEOTIDES

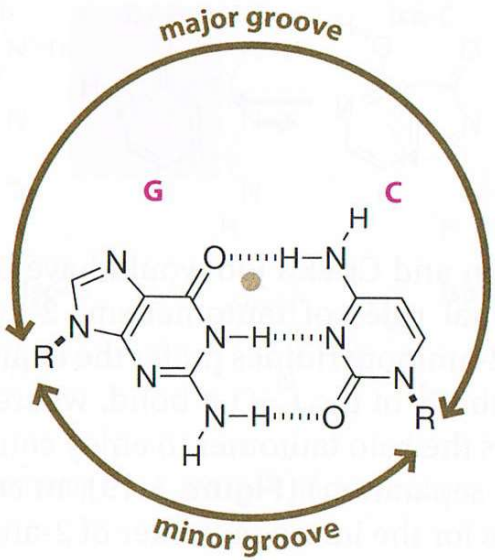
Part 1 – modified canonical nucleobases

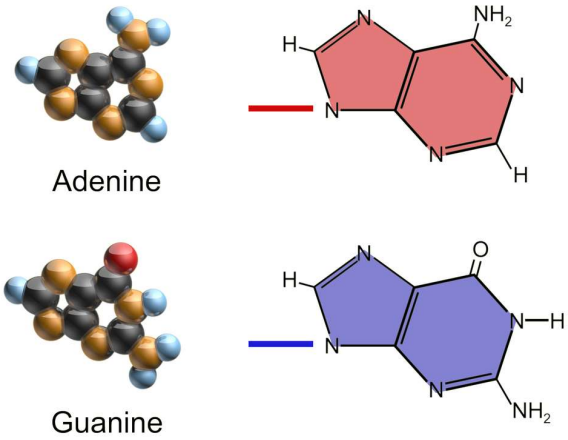
Canonical nucleobases and Watson-Crick pairing in DNA



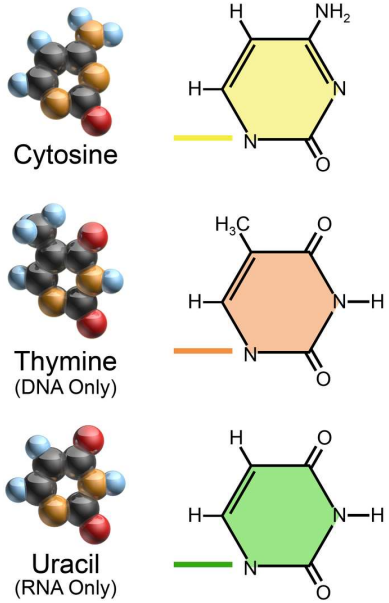
Madprime







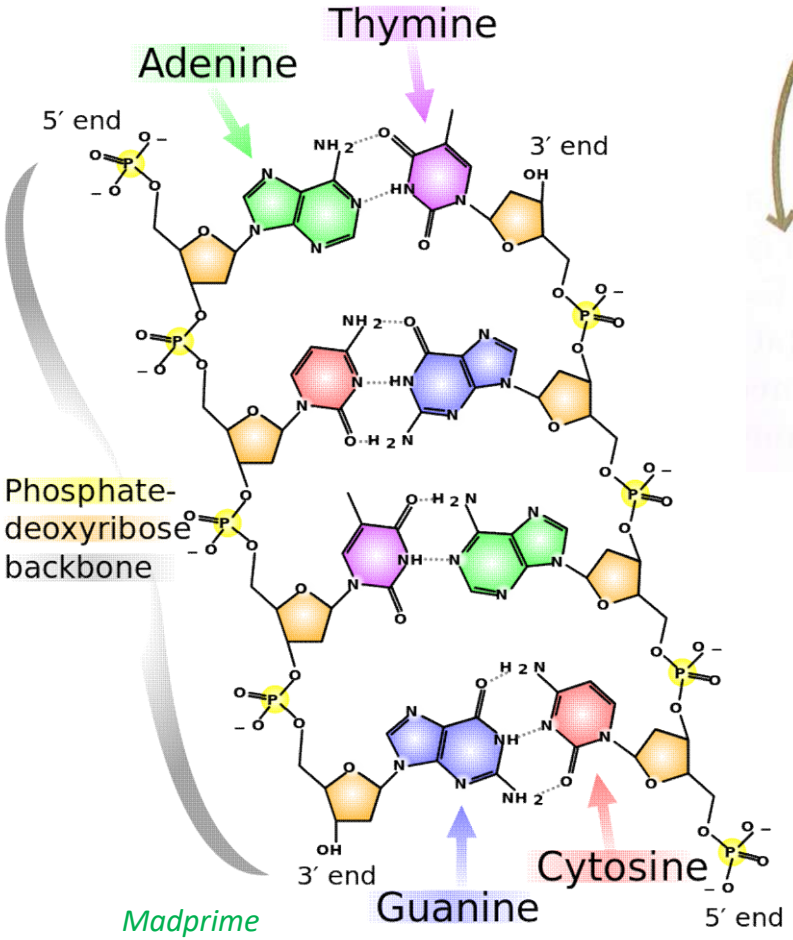
Purines

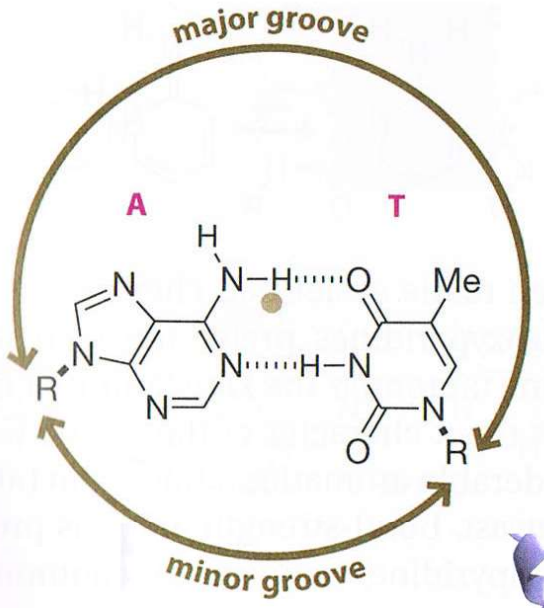


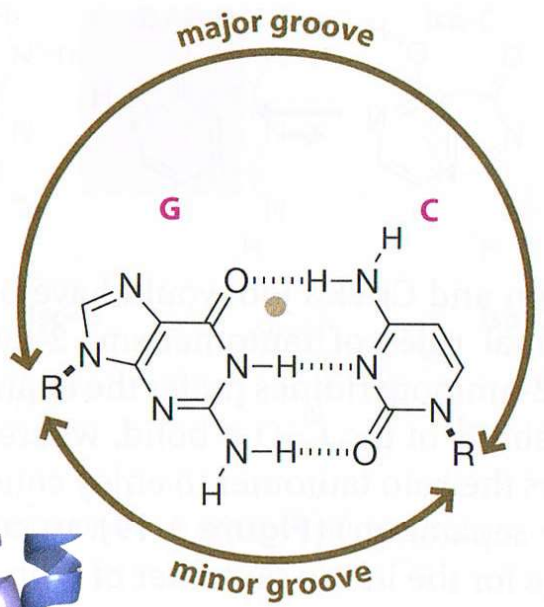
Pyrimidines

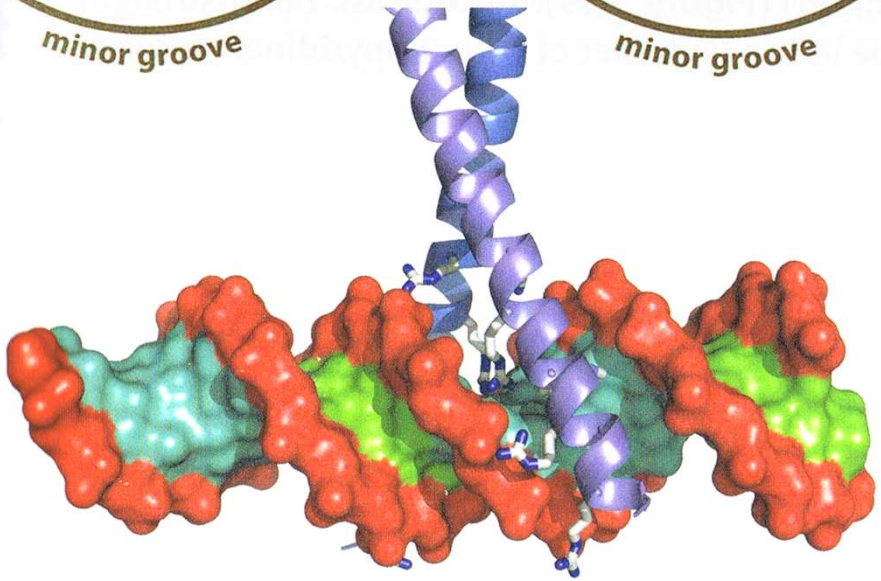
Bruce Blaus

Canonical nucleobases and Watson-Crick pairing in DNA









Hoogsten base pairing of canonical DNA nucleobases

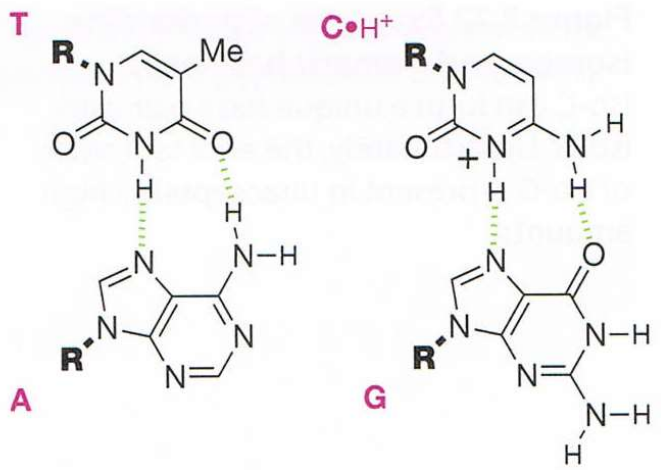
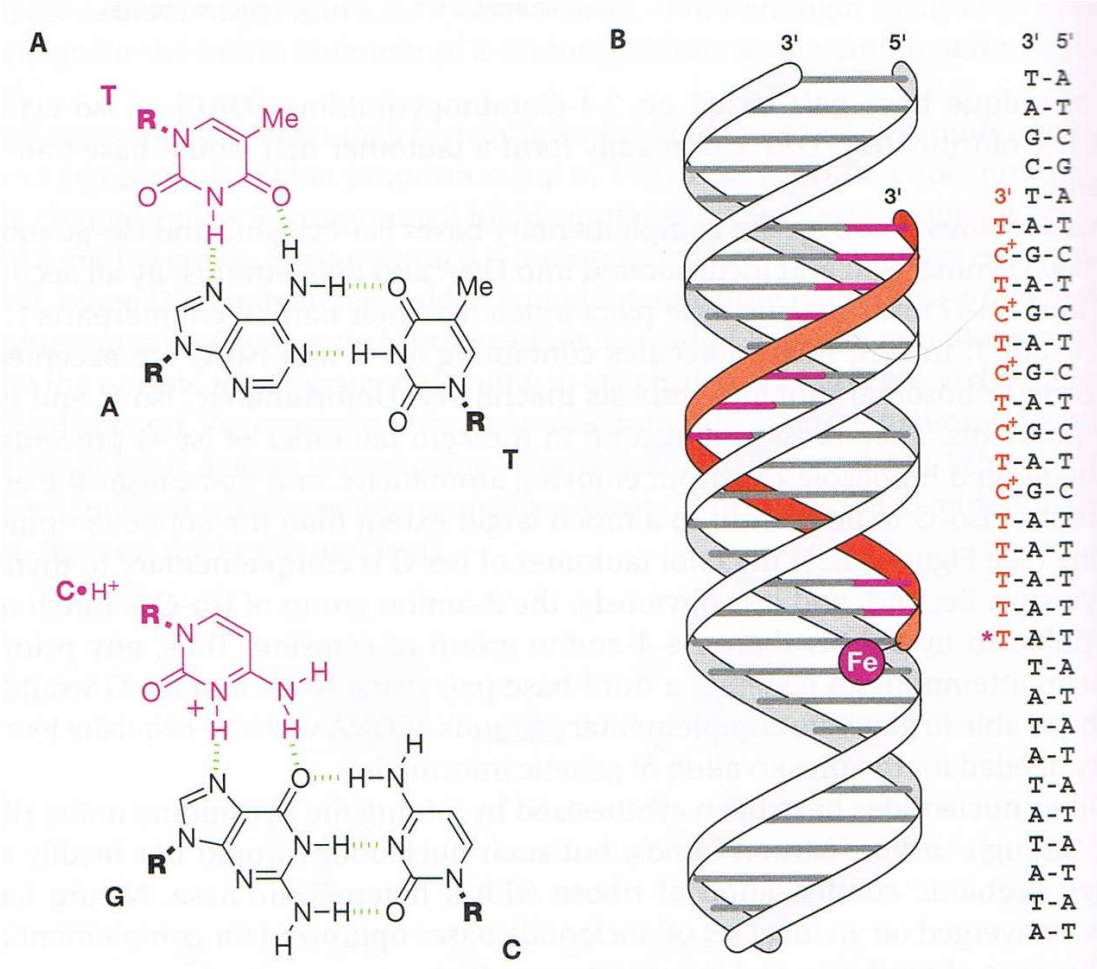
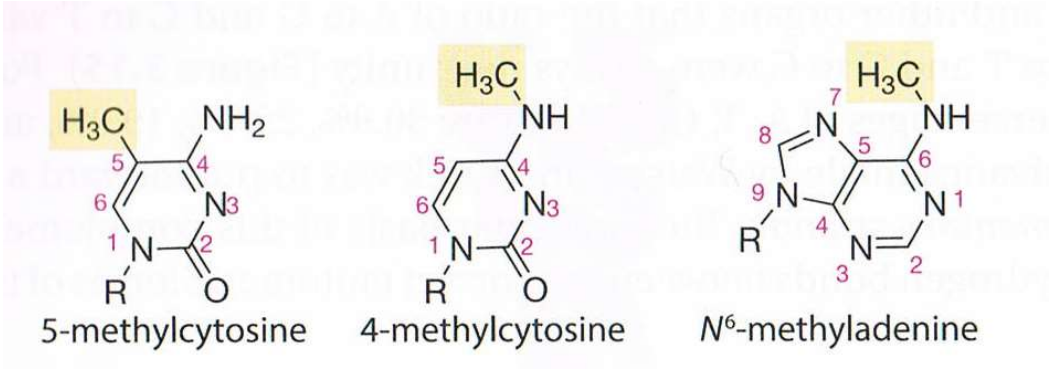


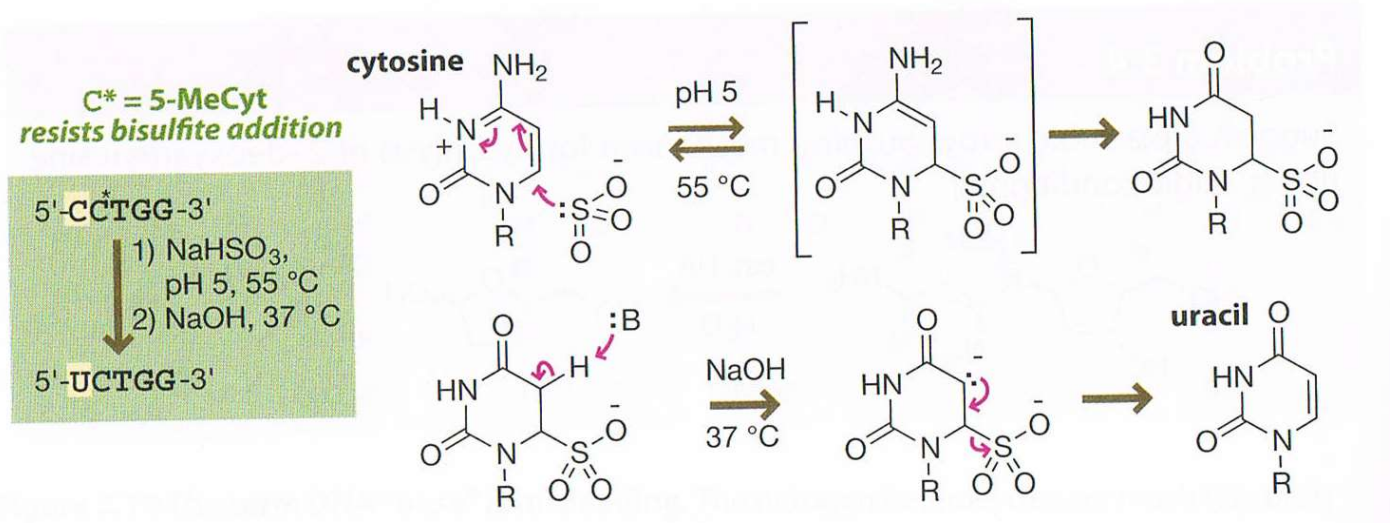
Figure 3.24 Hoogsteen base pairs. Hoogsteen base pairs use a different edge of the purine from a Watson–Crick base pair.



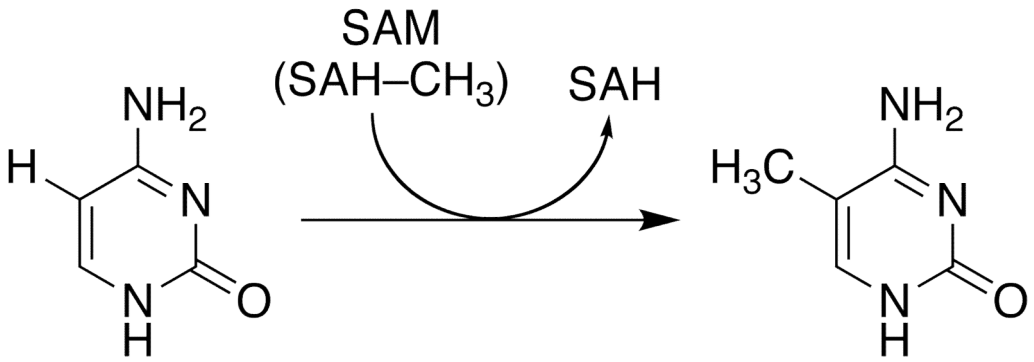
H. E. Moser, P. B. Dervan *Science* **1987**, *238,* 645-650

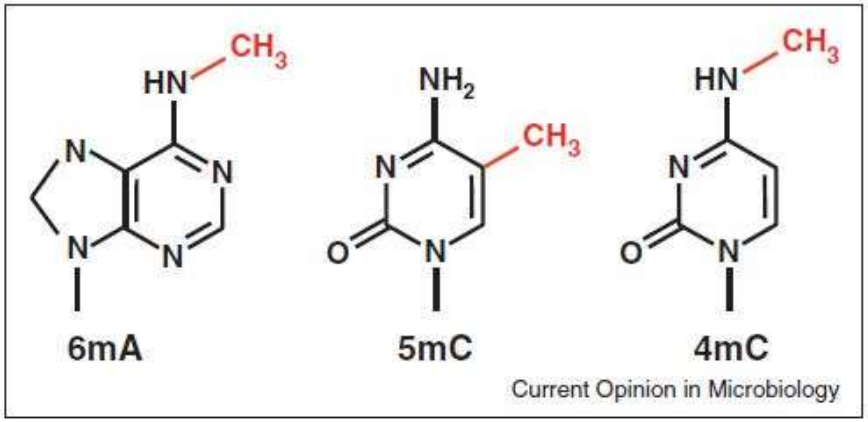
Modifications of nucleobase structures tolerated by polymerases



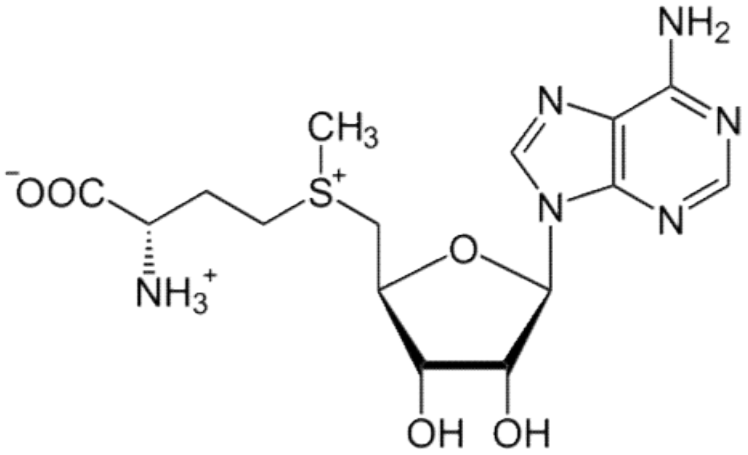


Modifications of nucleobases

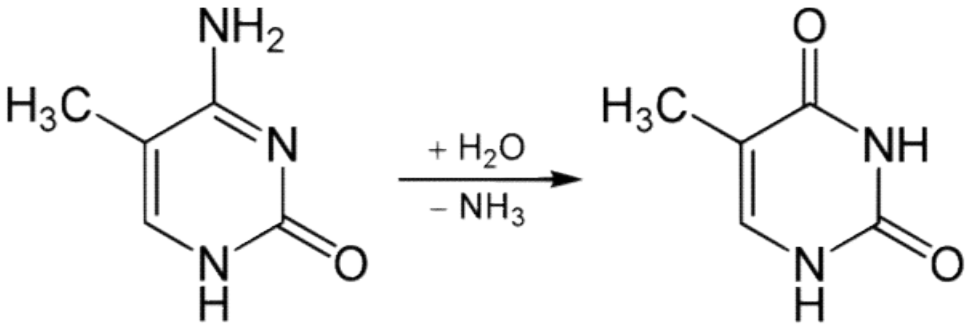




Chemical structures of common modified bases generated by DNA methyltransferases.

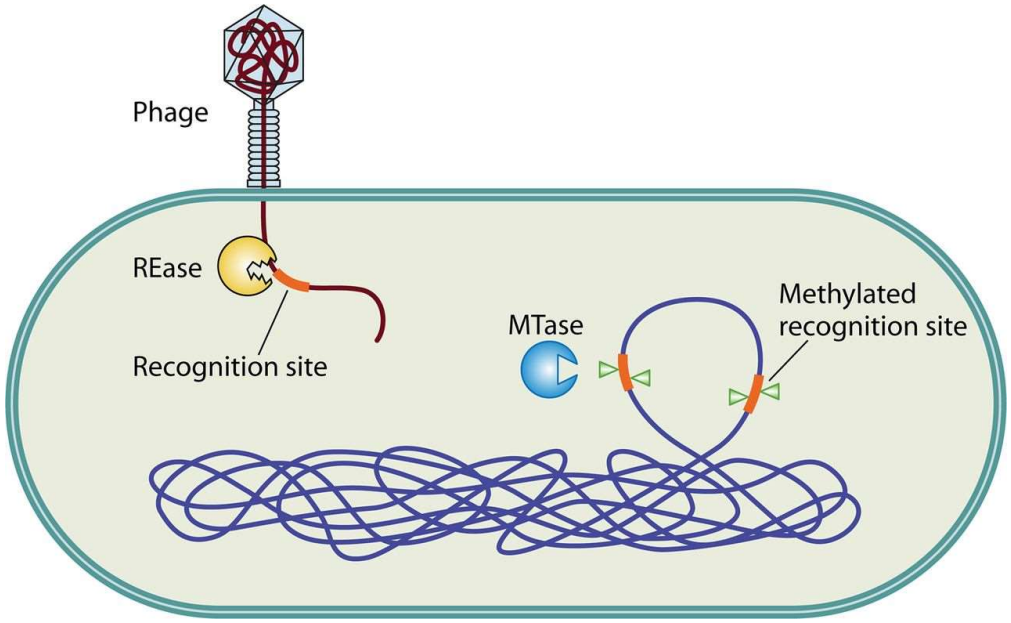


S-Adenosylmethionine (SAM)

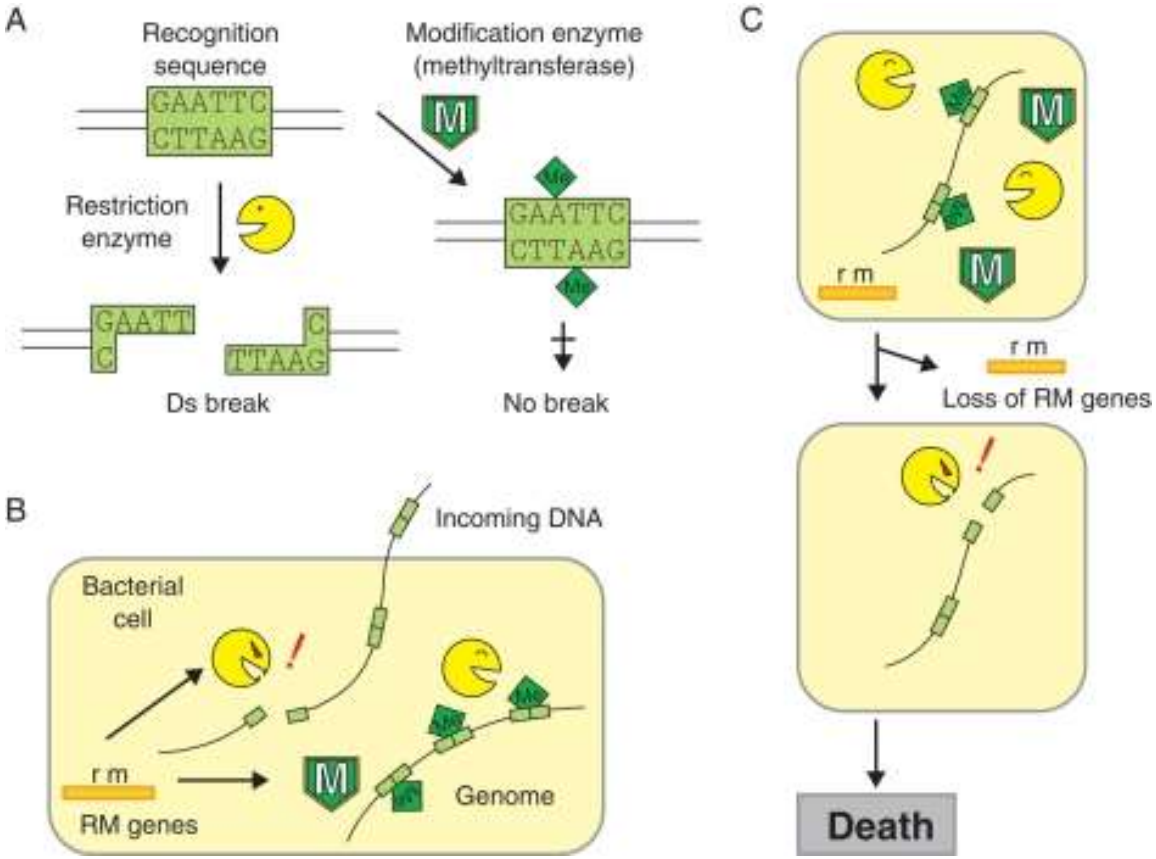


Restriction modification system

„Immune system" of bacteria and archaea against atacking viruses



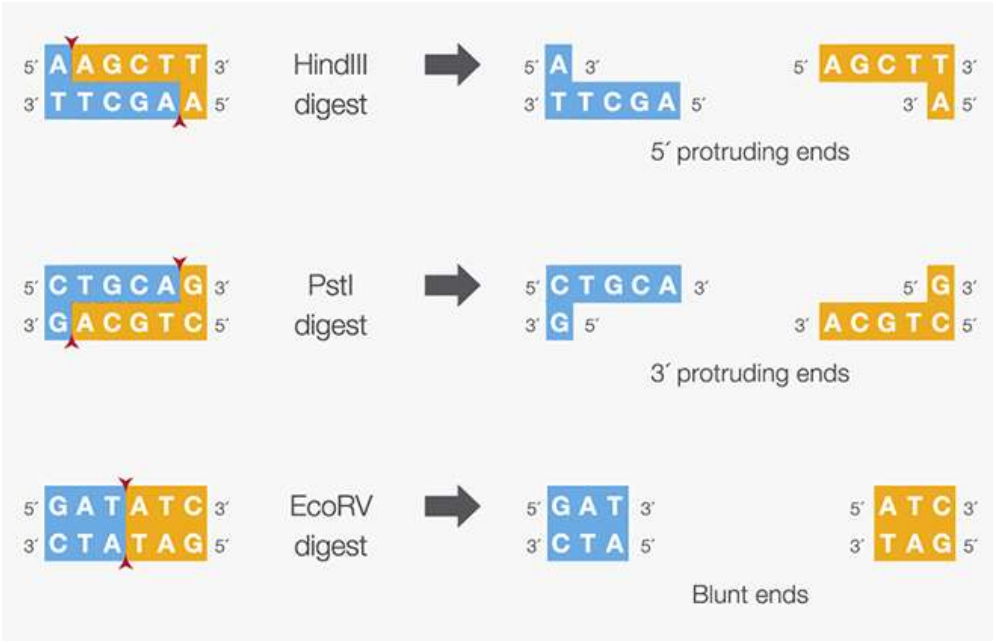
K. Vasu, V. Nagaraja
Microbiol. Mol. Biol. Rev. **2013**, *77(1)*, 53-72

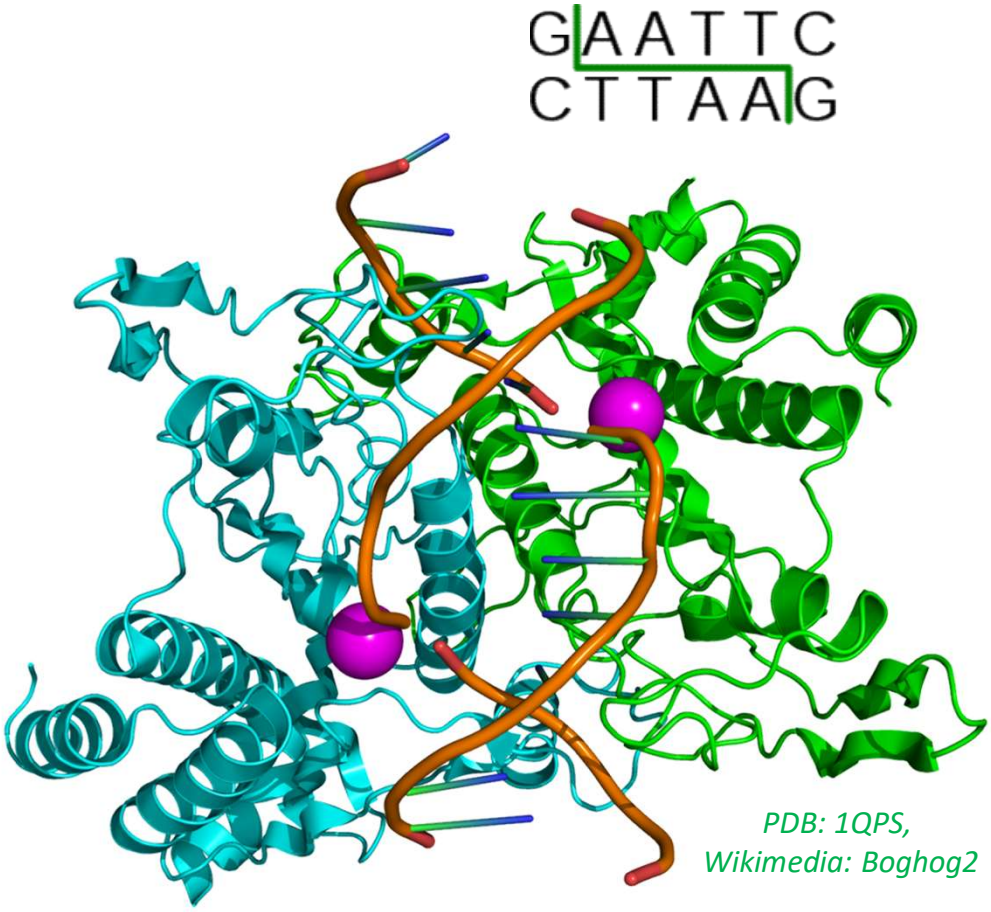


K. Ishikawa *et al. DOI: 10.1093/dnares/dsq027*

EcoI – a typical restriction enzyme

Products of restriction enzymes

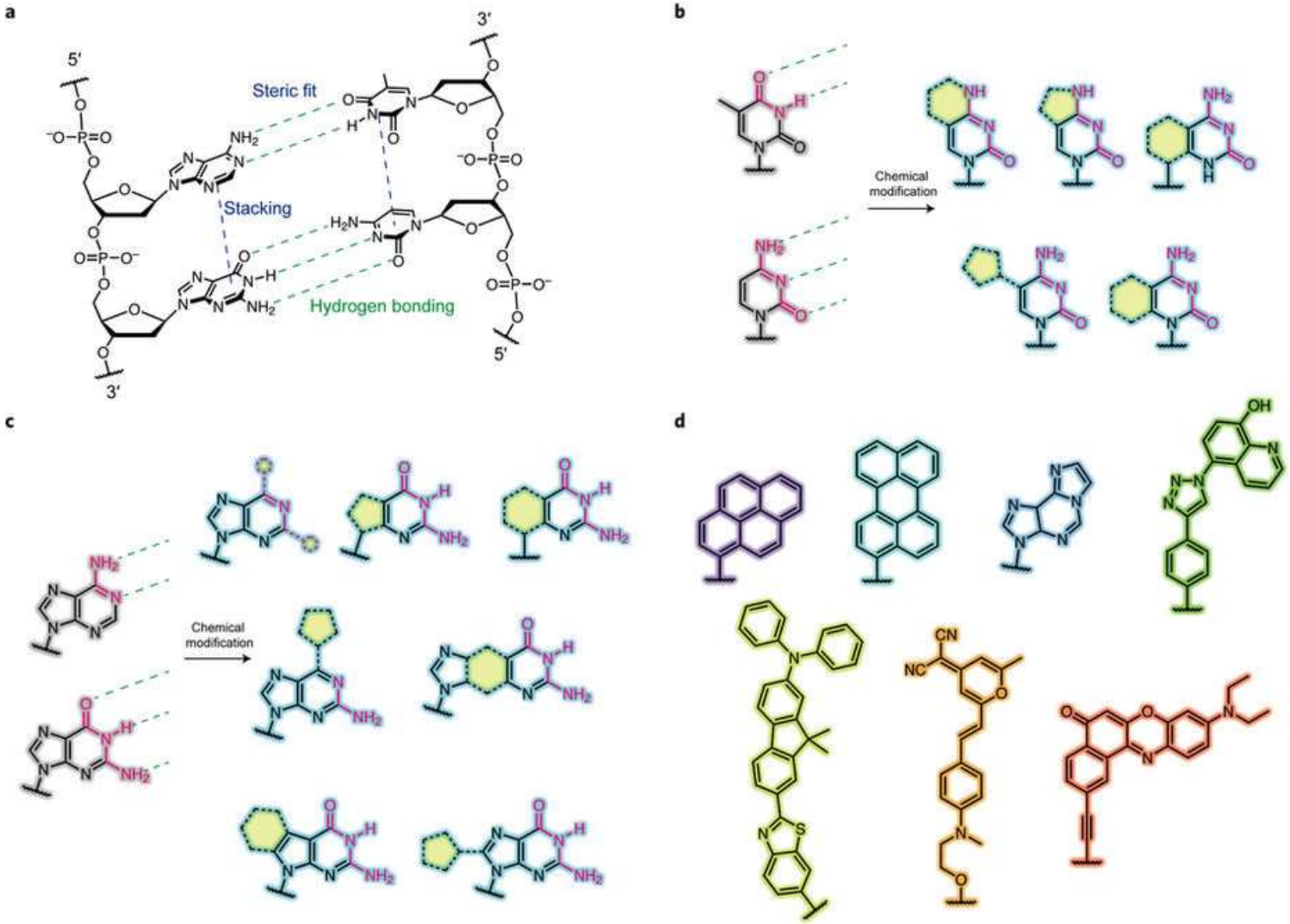




PDB: 1QPS, Wikimedia: Boghog2

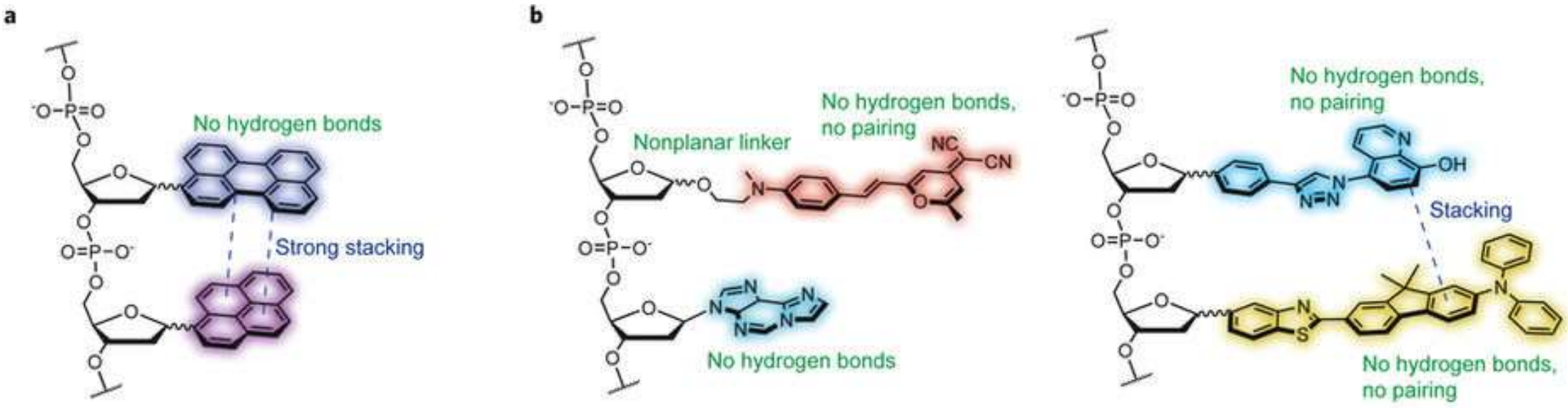
Structure of the homodimeric restriction enzyme EcoRI (cyan and green cartoon diagram) bound to double stranded DNA (brown tubes). Two catalytic magnesium ions (one from each monomer) are shown as magenta spheres and are adjacent to the cleaved sites in the DNA made by the enzyme (depicted as gaps in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

Non-canonical fluorescent nucleobases



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

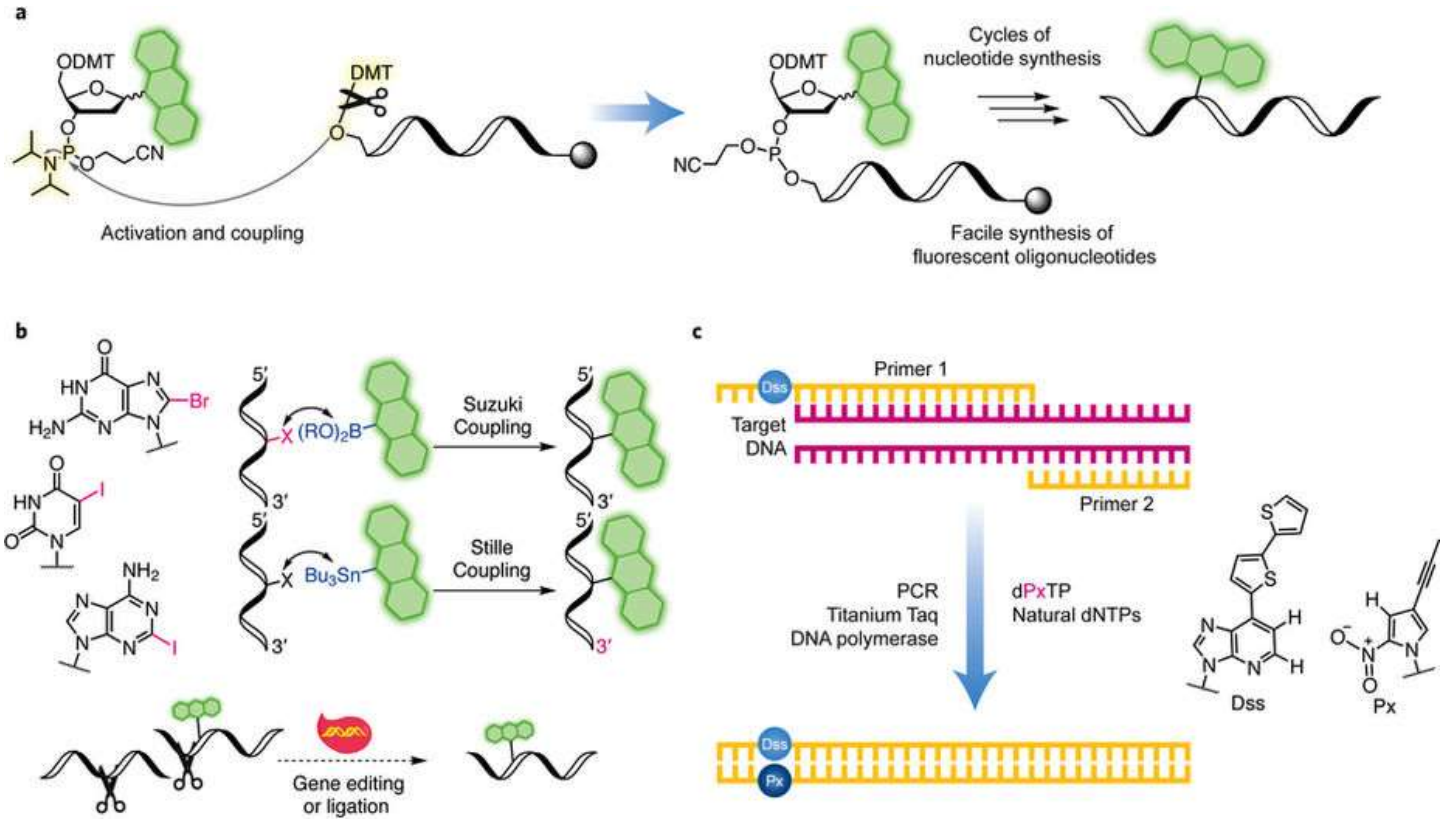
Non-canonical fluorescent nucleobases



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

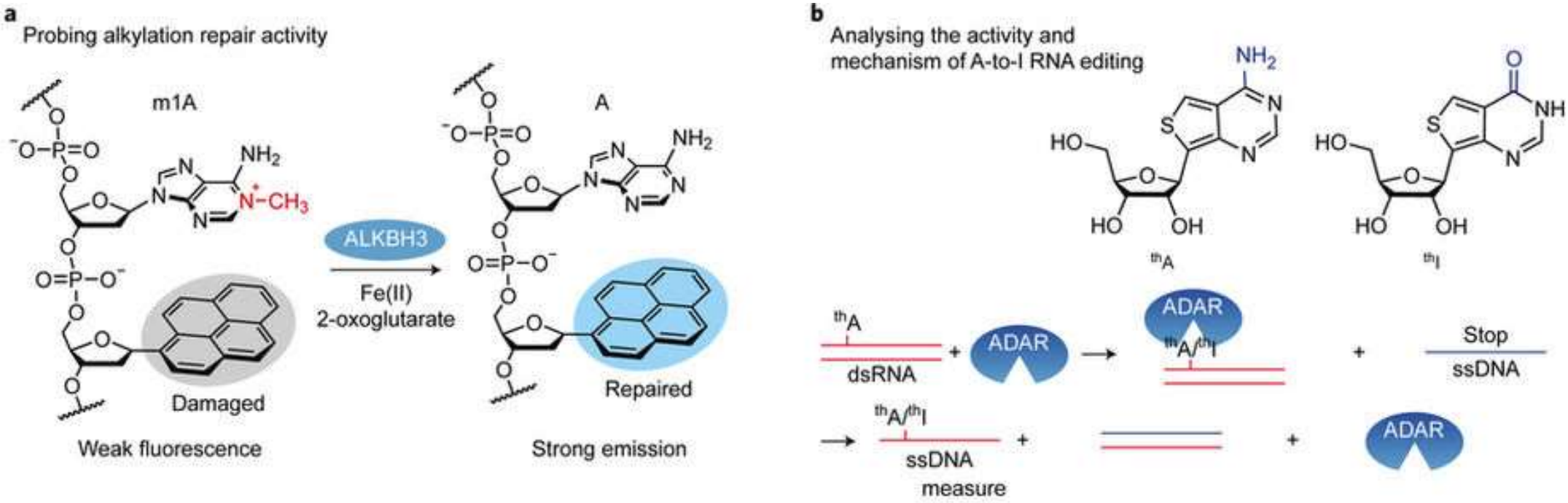
Strategies for incorporating fluorescent nucleobases



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

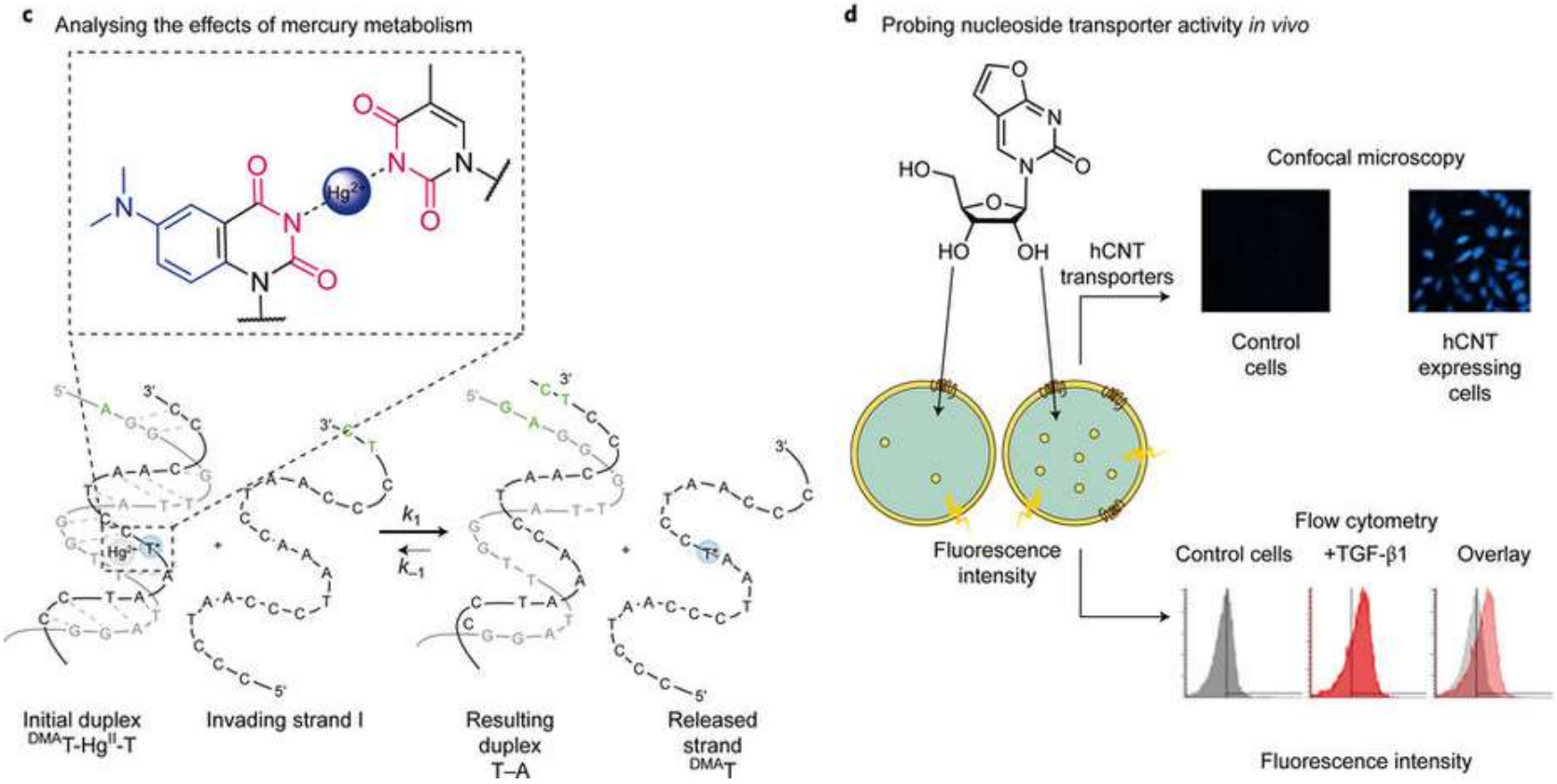


a, Fluorogenic sensing of a demethylation enzyme, ALKBH3. The emission of pyrene is initially quenched by the positive charge of 1-methylated adenine (m1A). When ALKBH3 demethylates m1A, the quenching effect is removed and a signal is generated.

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

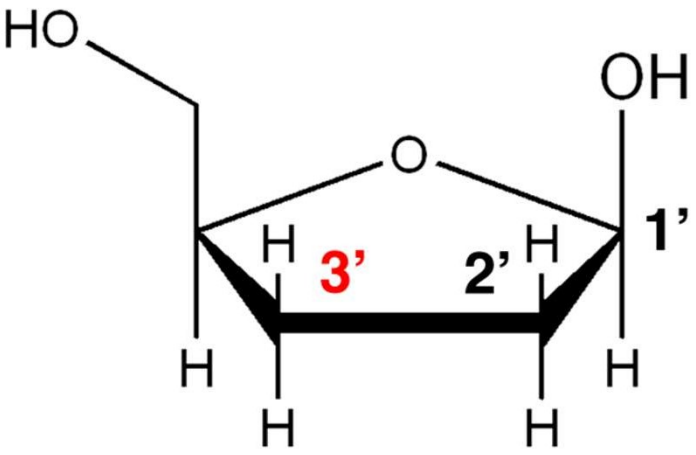
Fluorescent nucleobases for studying DNA and RNA

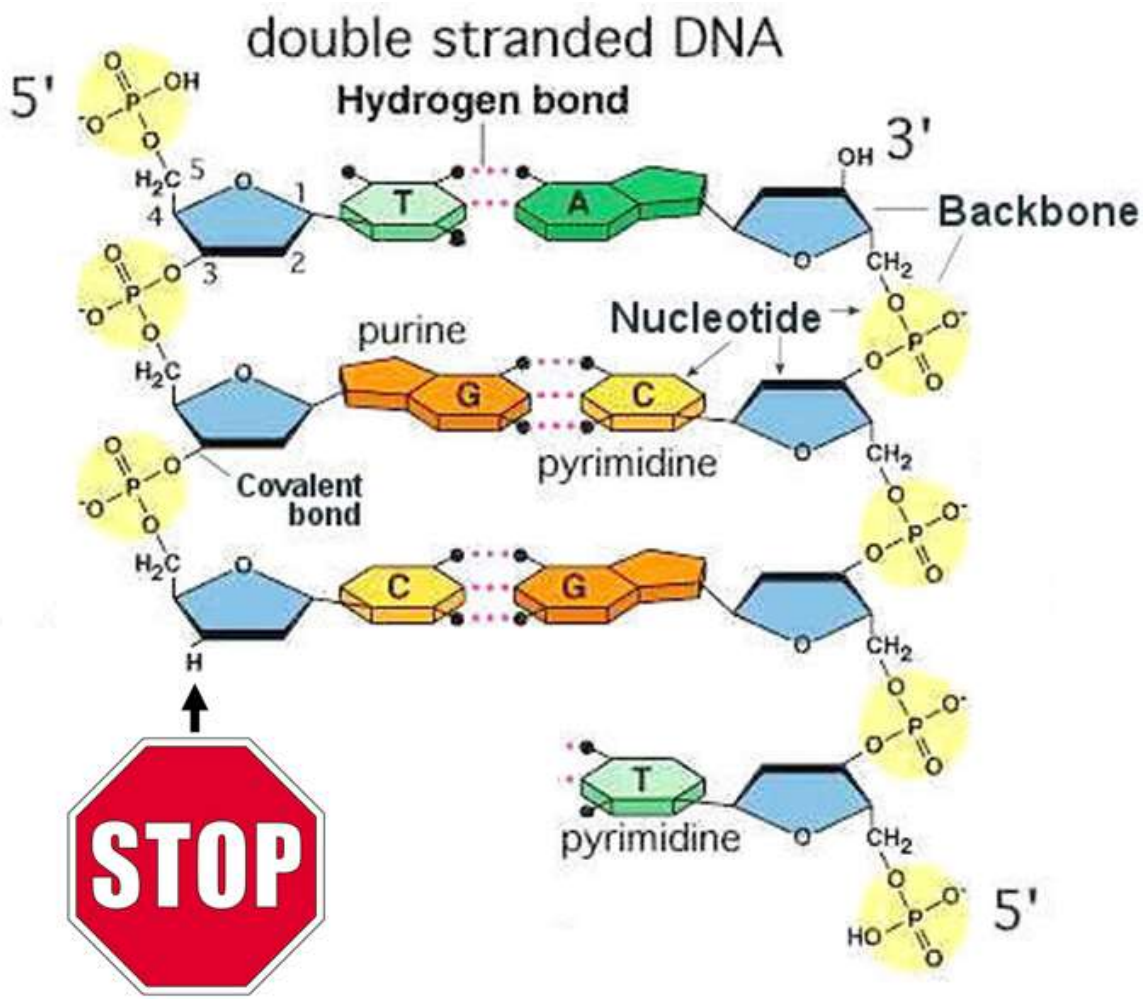


c, Kinetic and thermodynamic investigation of the effects of mercury on DNA metabolism. The fluorescent thymine can chelate mercury with another thymine ring and link DNA strands. This can be used to probe mercury metabolism in vivo and to study the effects of mercury on DNA status.

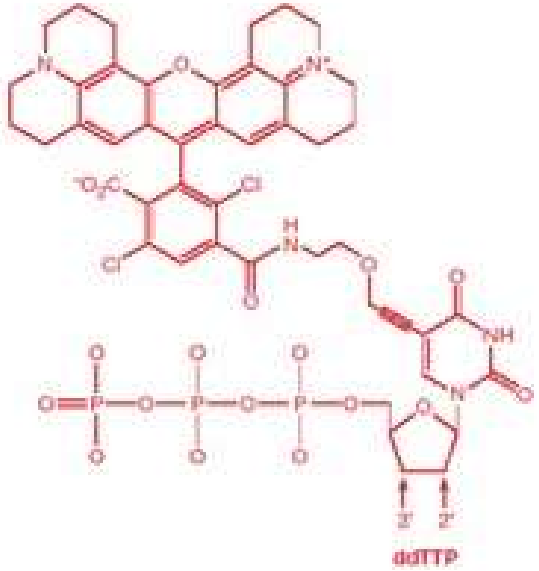
d, Visualization and analysis of human concentrative nucleoside transporters (hCNTs). The fluorescent nucleoside can enter the plasma membrane through the transporters, thus allowing the measurement of the transport activity. TGF-β1, transforming growth factor β1. W. Xu, K. M. Chan, E. T. Kool *Nature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

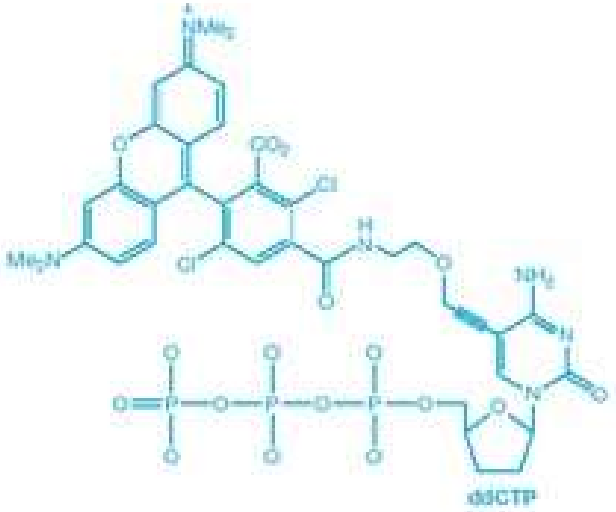
Sanger sequencing

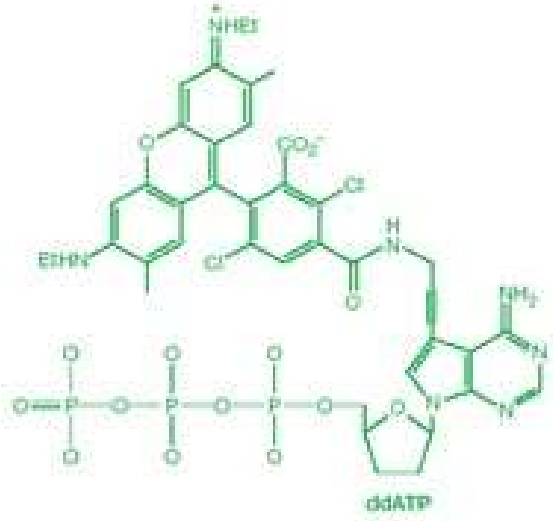


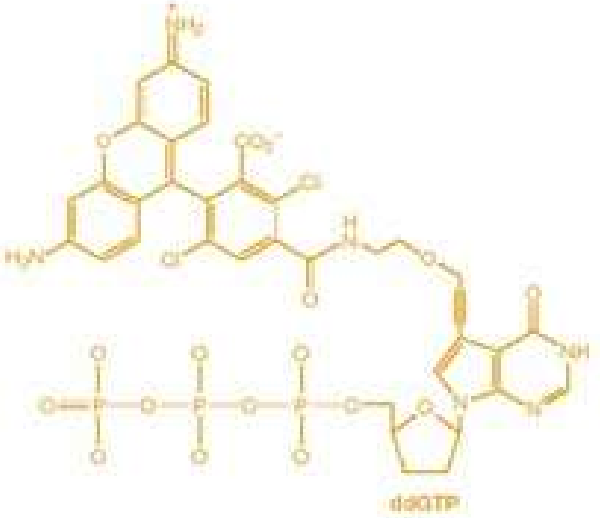


Sanger sequencing

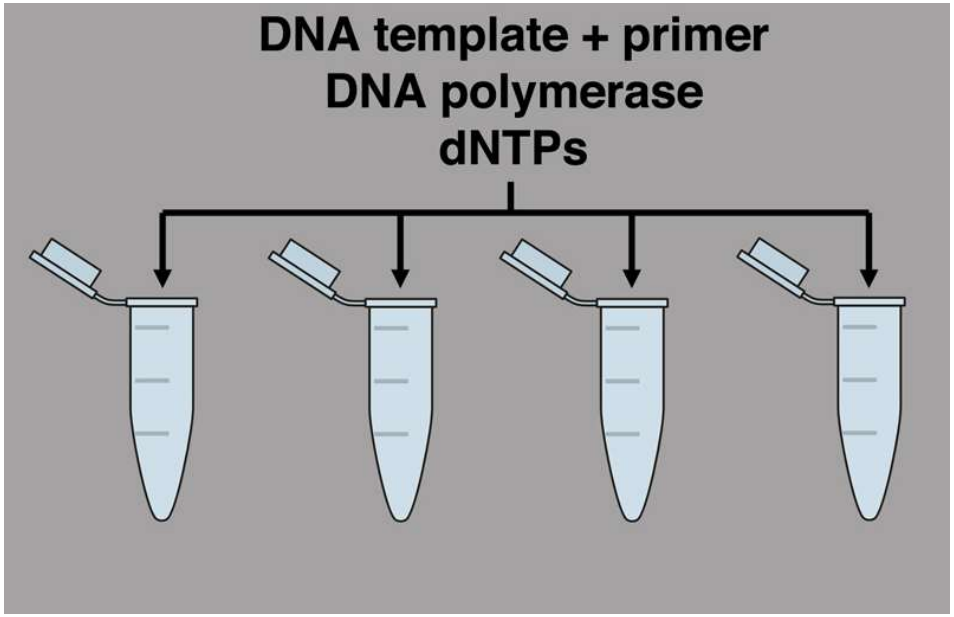


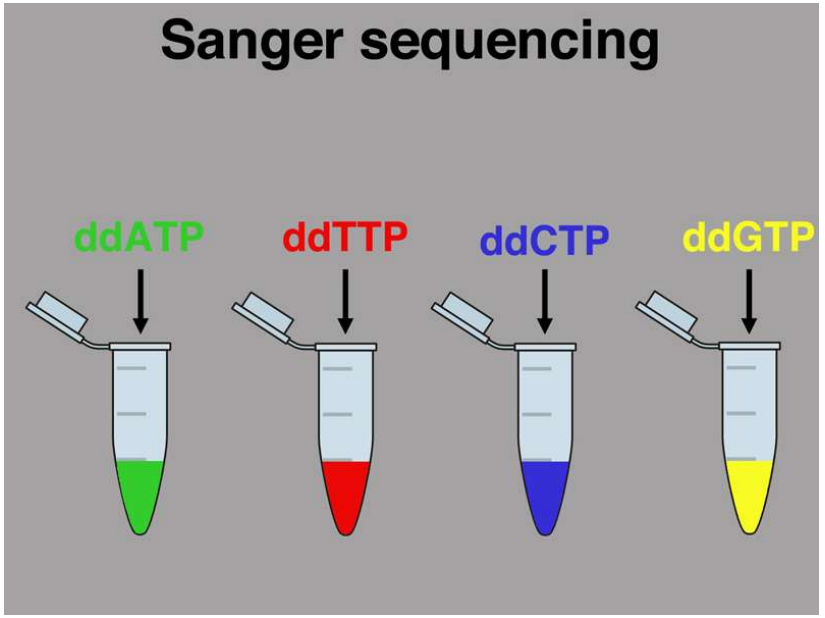






Sanger sequencing





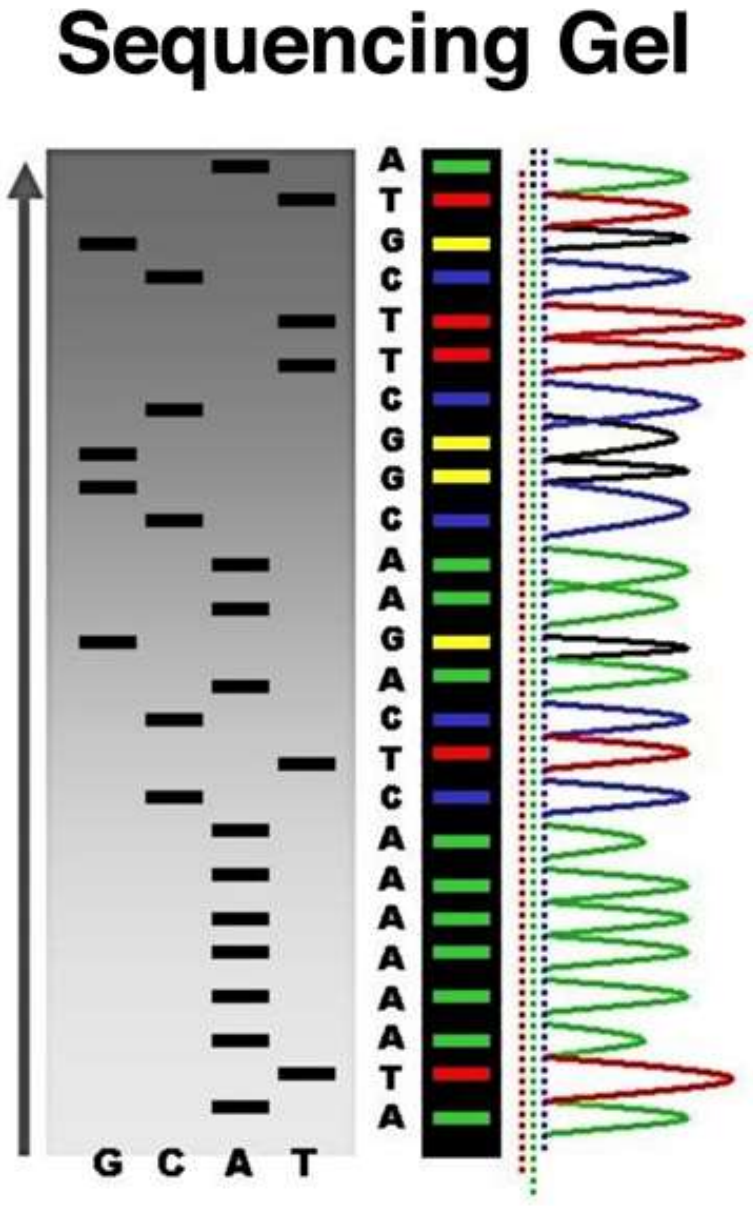
Sanger sequencing

TACGTA
ATGCATTAGGGCCTGGCTCTTT

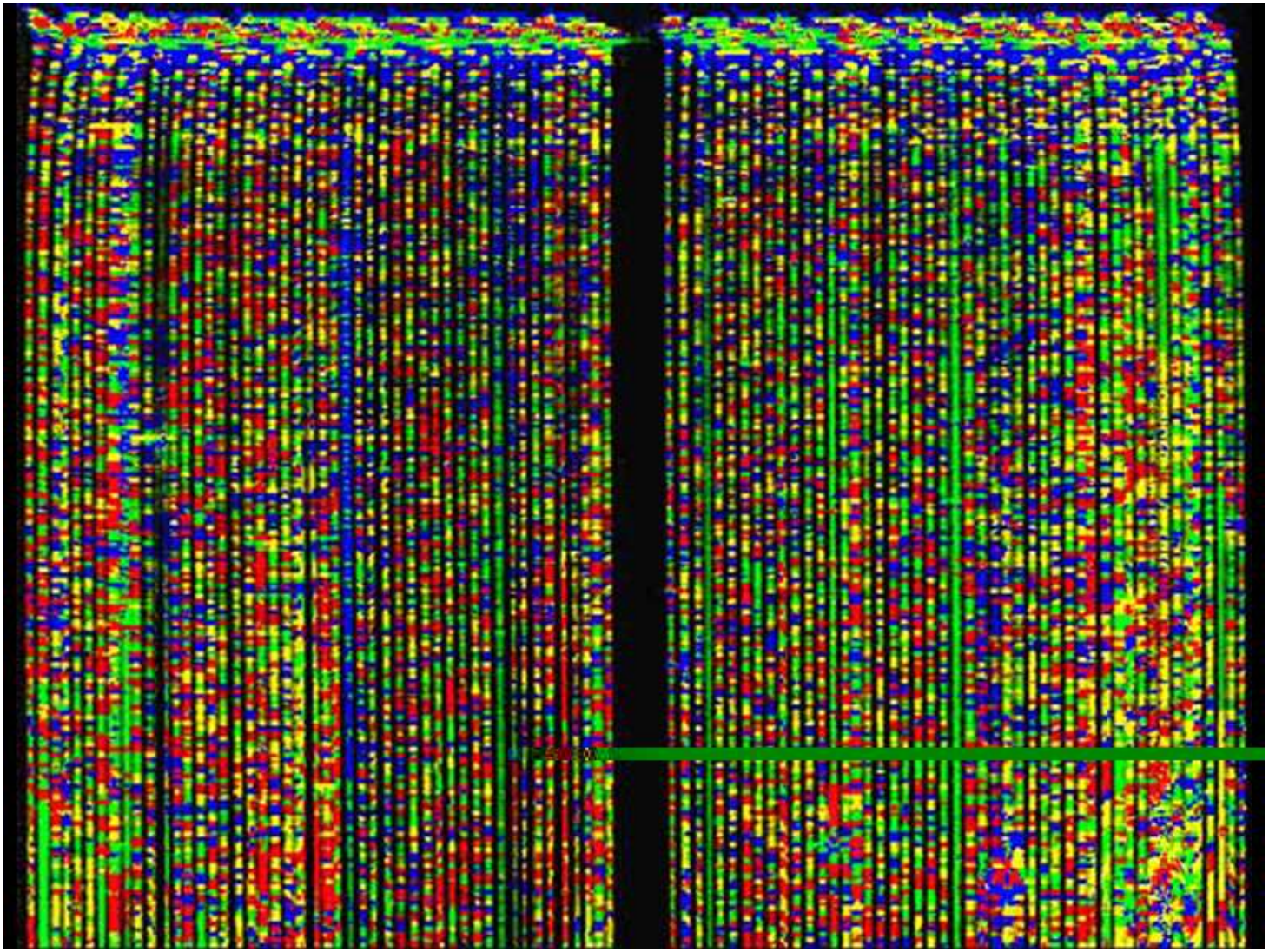
TACGTAA
ATGCATTAGGGCCTGGCTCTTT

TACGTAATCCCGGA
ATGCATTAGGGCCTGGCTCTTT

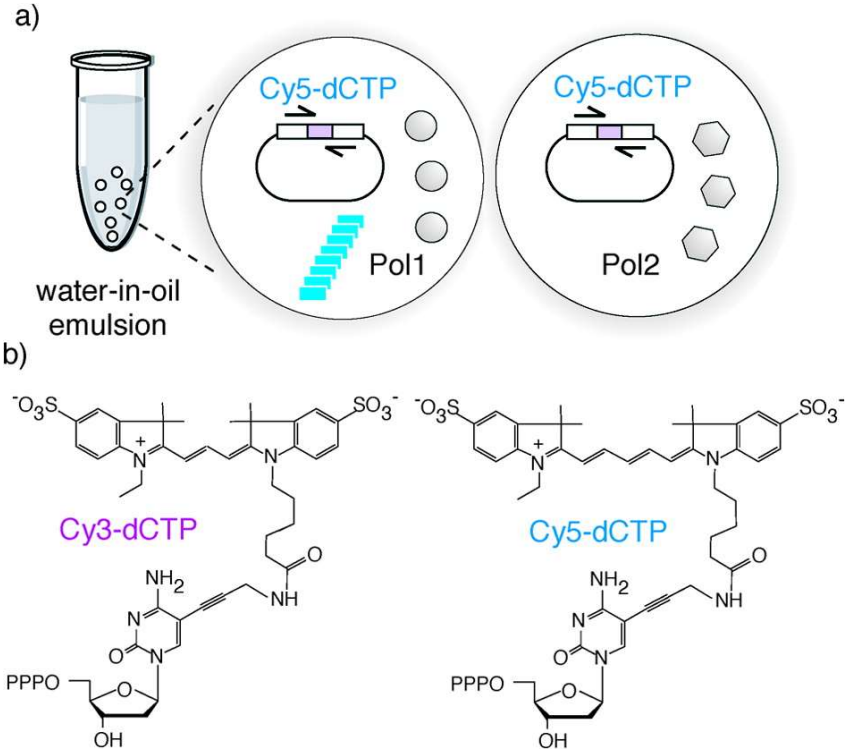
TACGTAATCCCGGACCGA
ATGCATTAGGGCCTGGCTCTTT



Sanger sequencing



CyDNA – synthesis and replication of highly fluorescently-labelled DNA



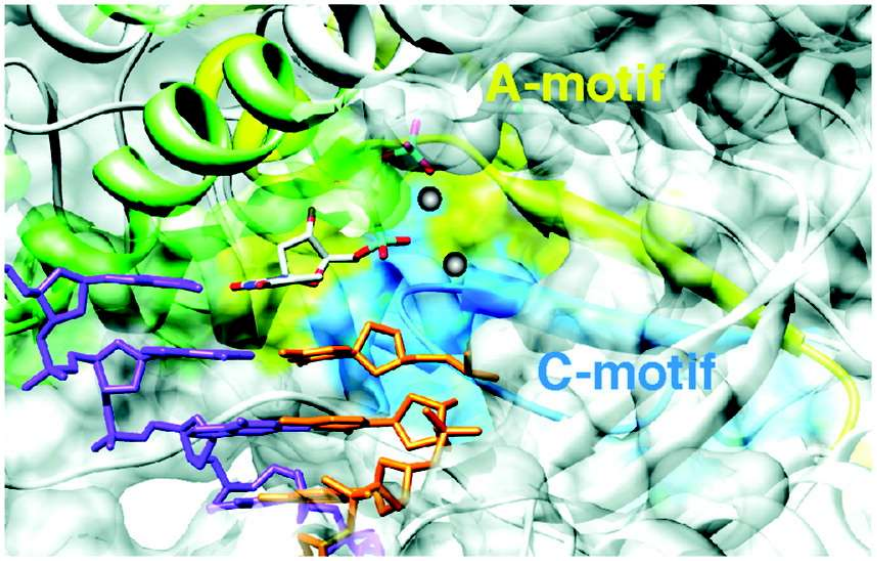
DNA Polymerase is evolutionarly optimized by mutagenesis to accept fluorescent Cy3 or Cy5 nucleotides instead of simple dCTP. Here, the polymerase replicates a short segment of its own encoding gene. Water/oil compartmentalization allows testing many independent mutations in paralell

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.

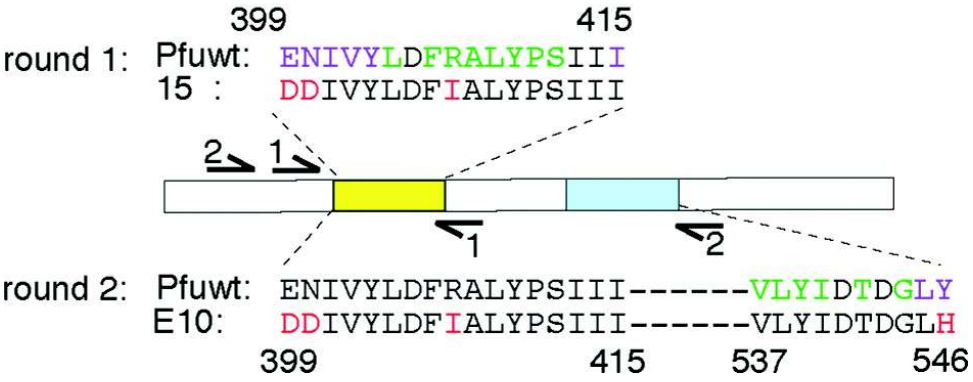
P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

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

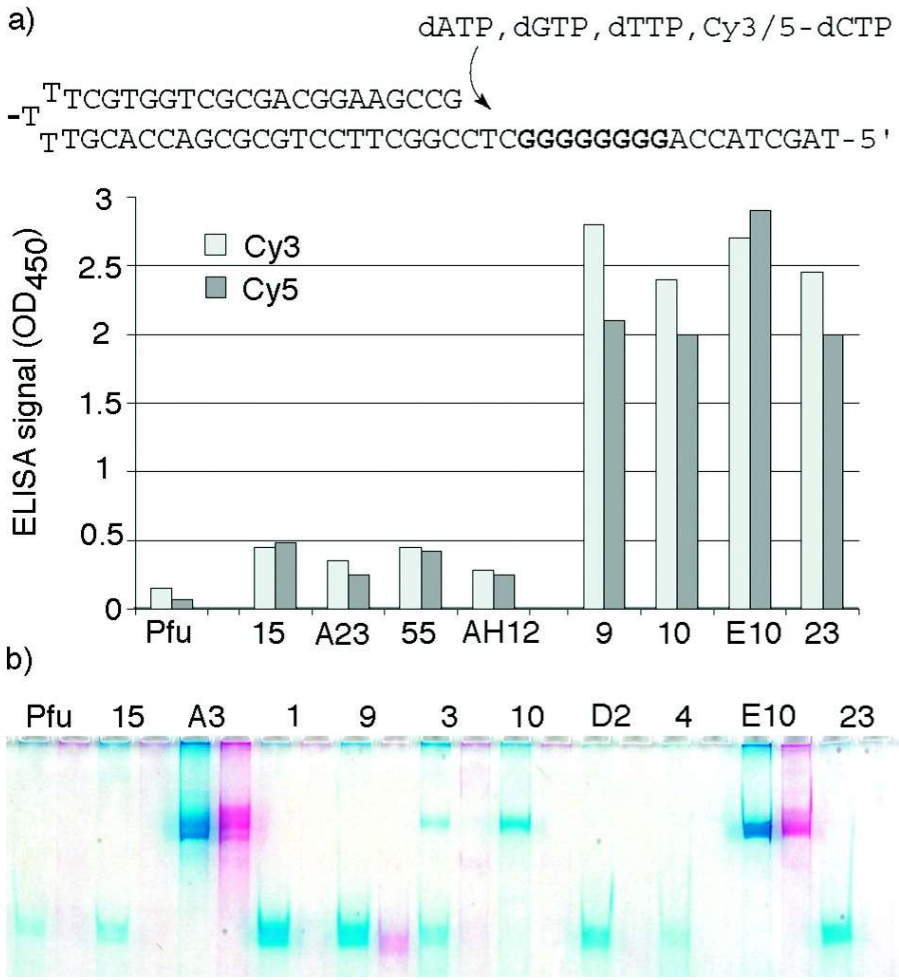
CyDNA – synthesis and replication of highly fluorescently-labelled DNA

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

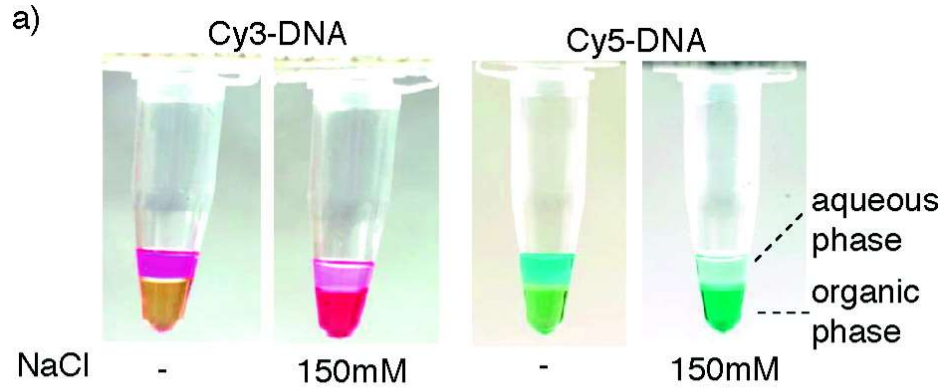
Table 1. Polymerase Fidelity

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

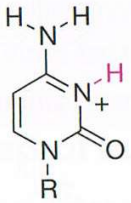
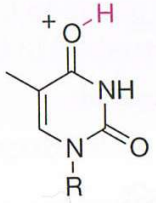
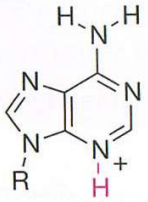
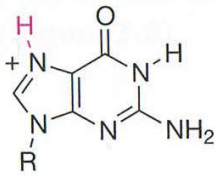


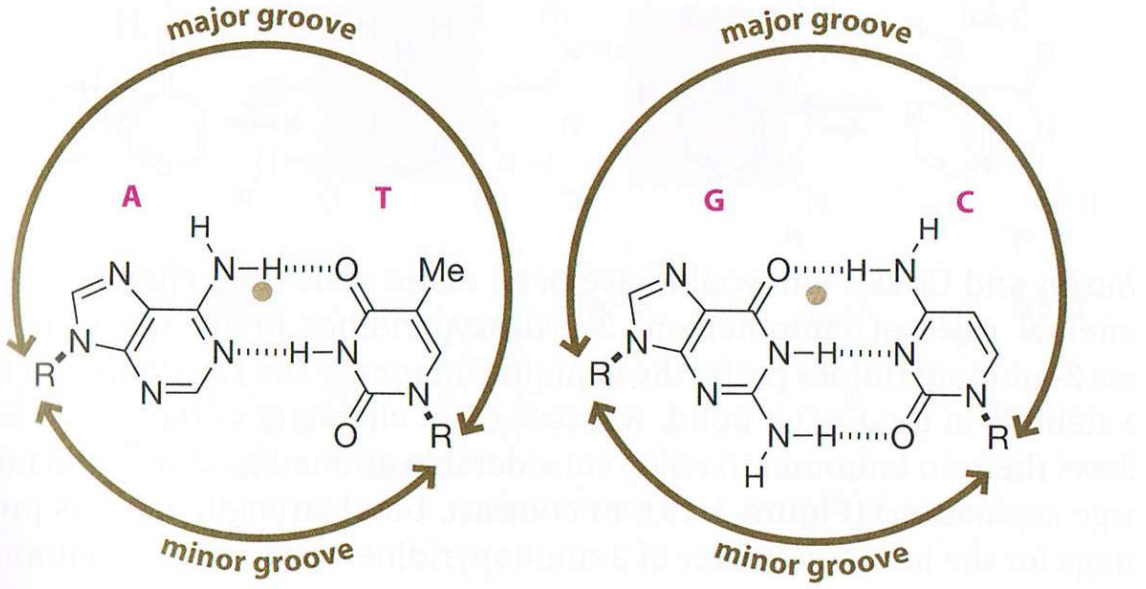
(a) Organic phase partitioning of CyDNA is shown for Cy3-DNA (left) and Cy5-DNA (right). Essentially 100% partitoning occurs in the presence of 150 mM NaCl (the yellow color of the phenol phase is due to addition of 8-hydroxyquinoline to prevent oxidation).

(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.

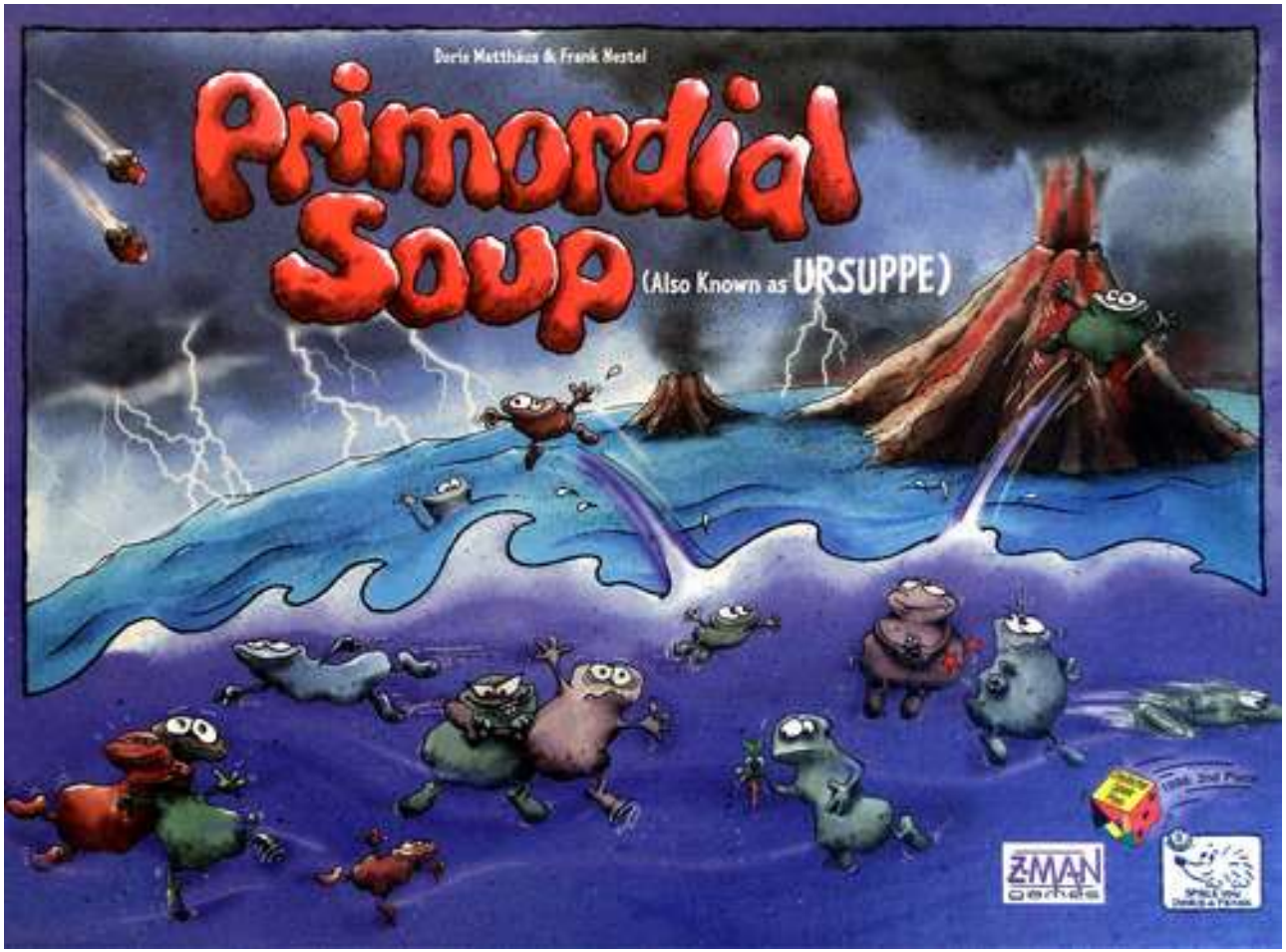
P. Holliger *et al. J. Am. Chem. Soc.* **2010**, *132,* 5096-5104

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



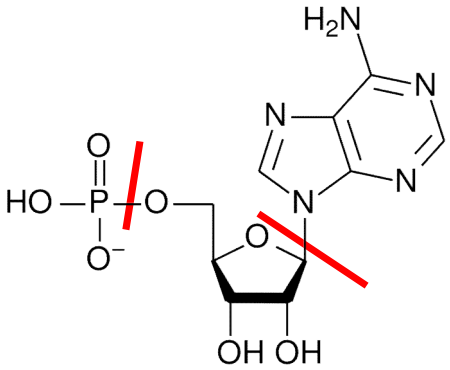
Prebiotic synthesis of nucleotides

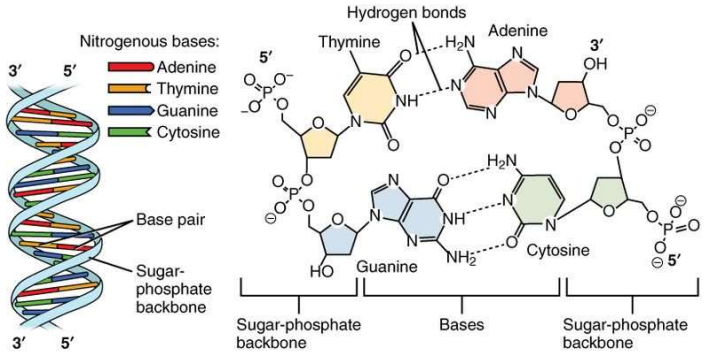


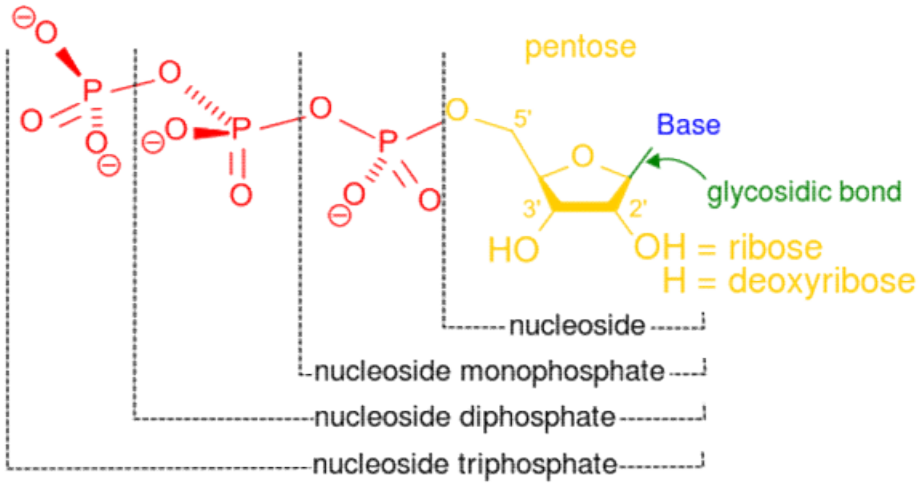
State of the art

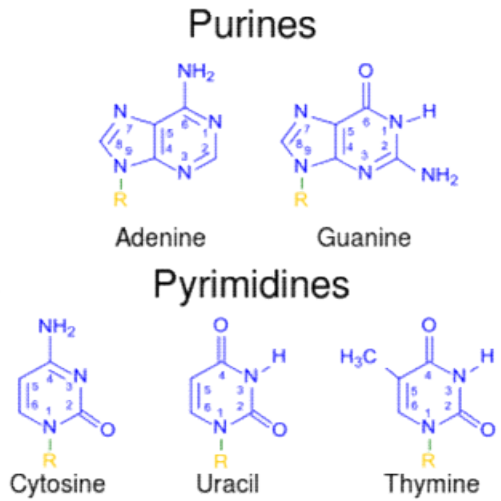
Nucleotides - components

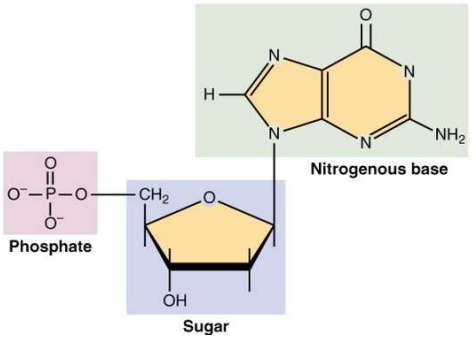
RNA – most likely evolutionarily older („RNA World") than DNA → prebiotic origin of ribose + A, C, G, and U nucleobases





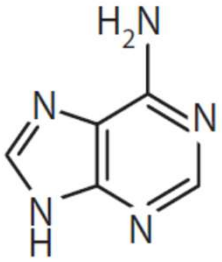




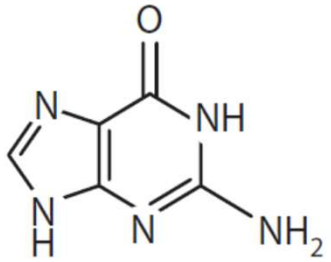


Prebiotic synthesis of nucleobases

Purines

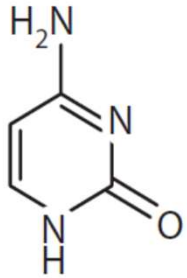


Adenine

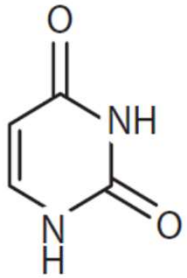


Guanine

Pyrimidines

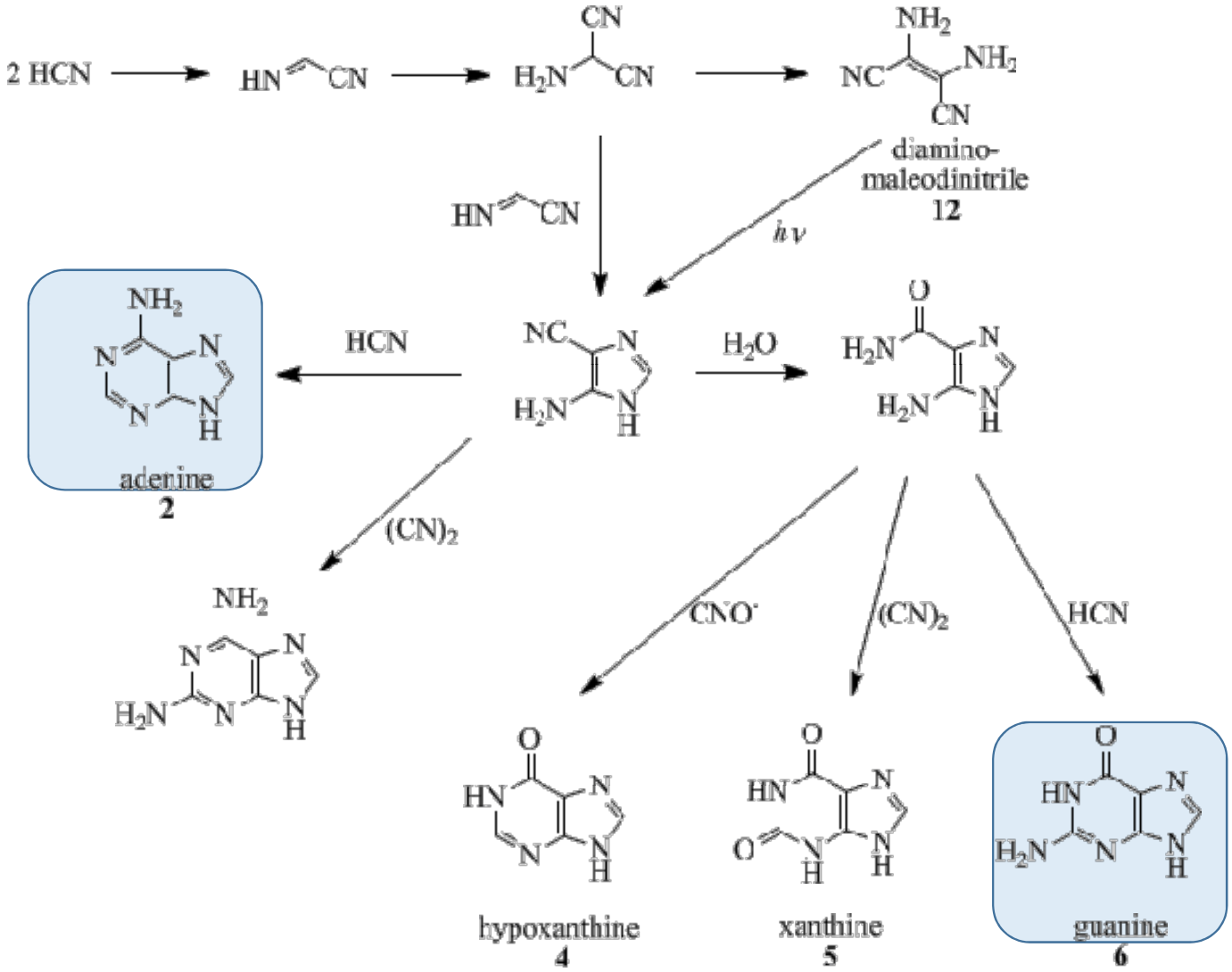


Cytosine

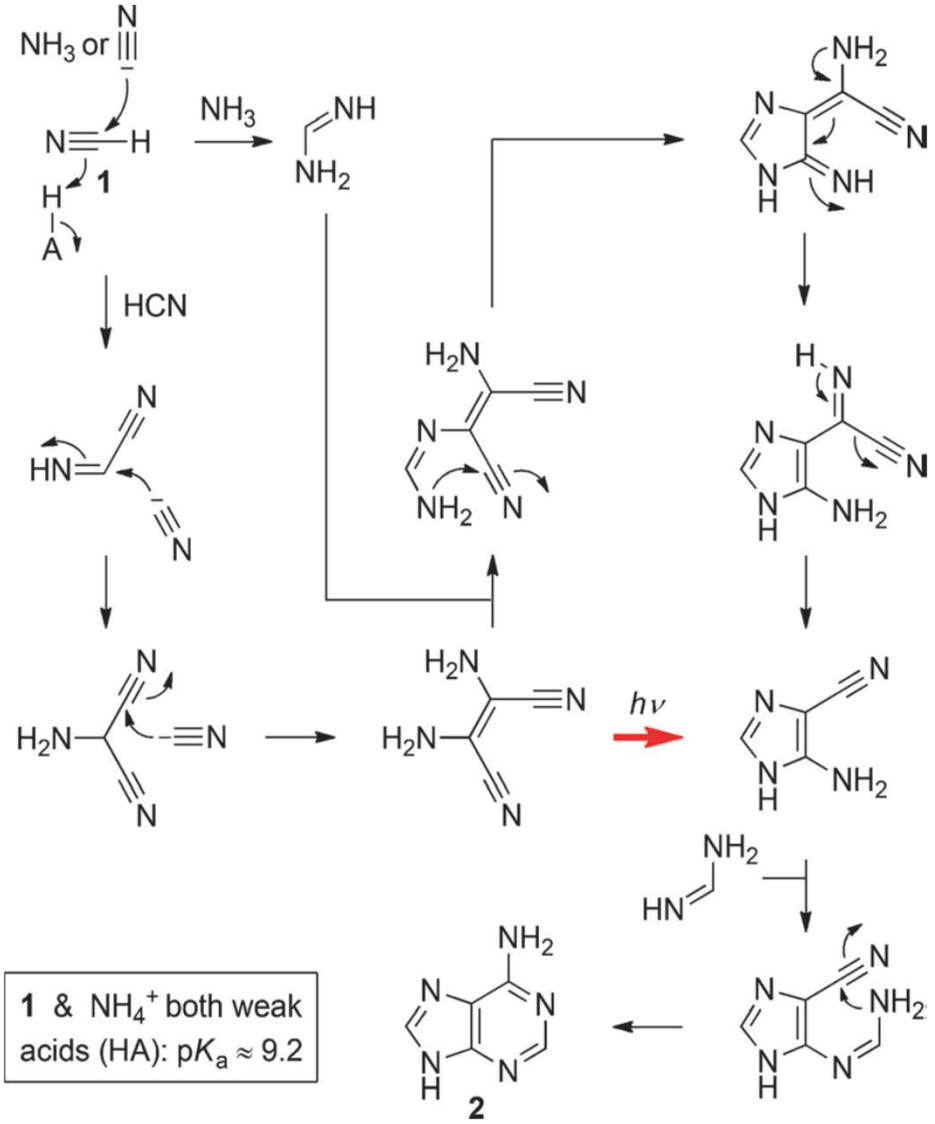


Uracil

Prebiotic synthesis of purines



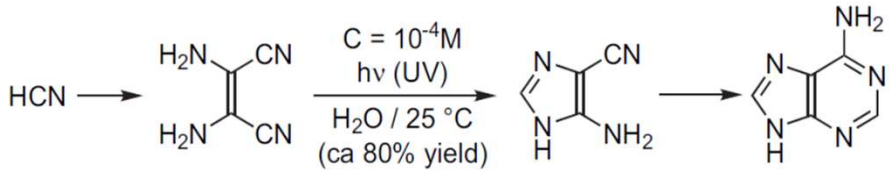
Prebiotic synthesis of adenine



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

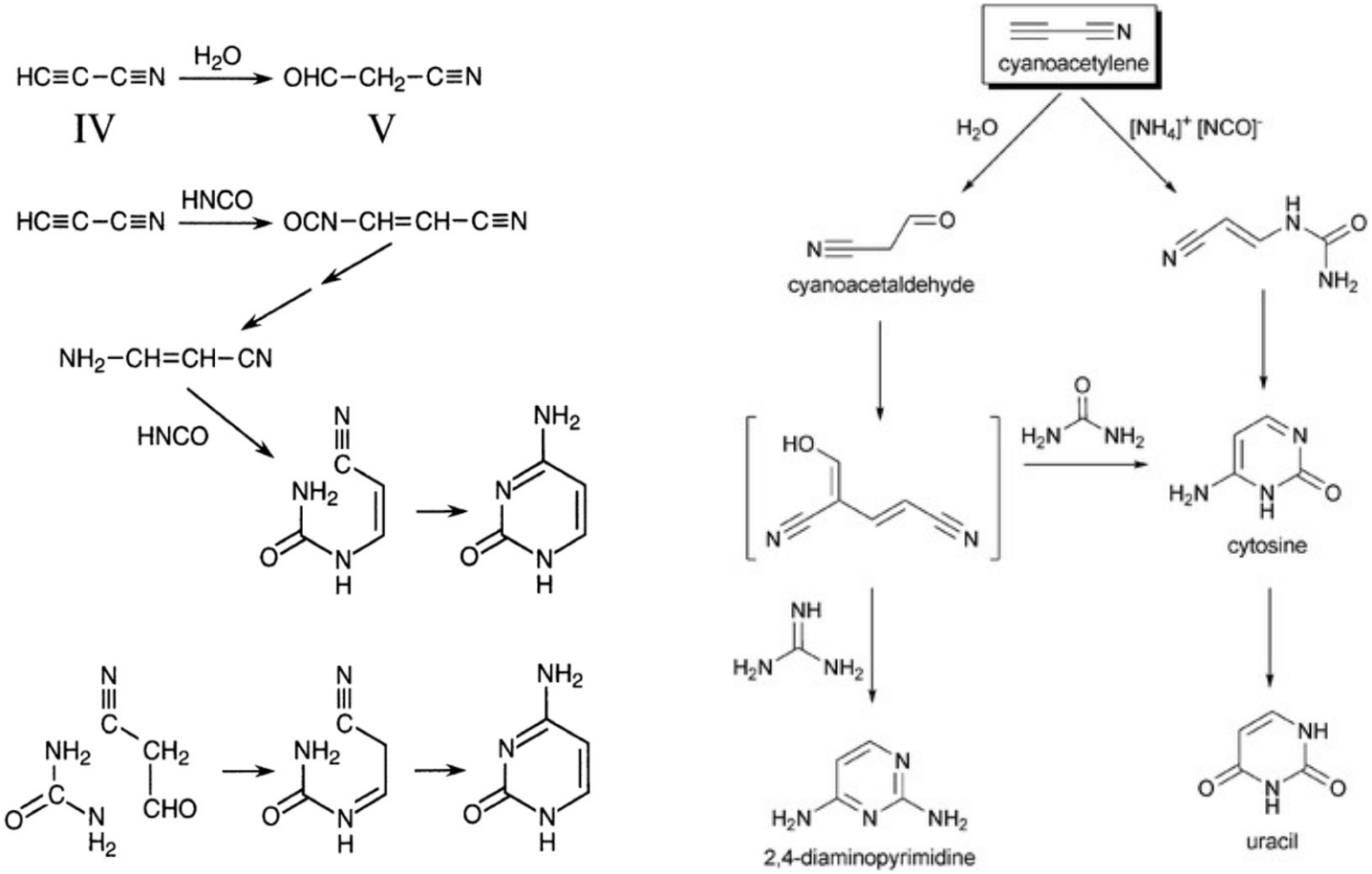


J. Oro *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407.

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

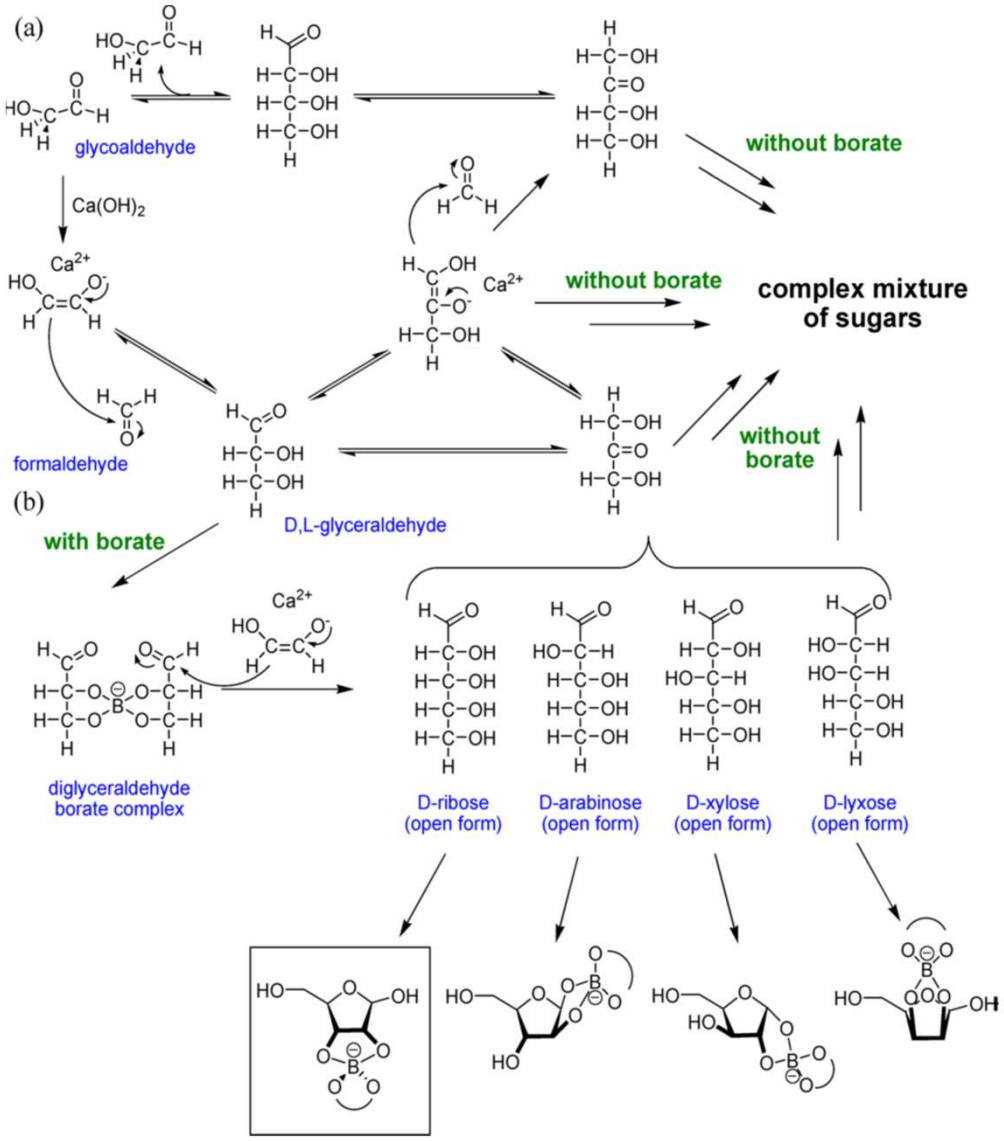
Prebiotic synthesis of pyrimidines

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.

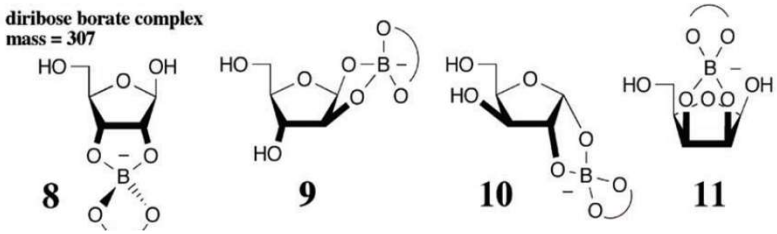
Formose reaction in presence of borates



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.

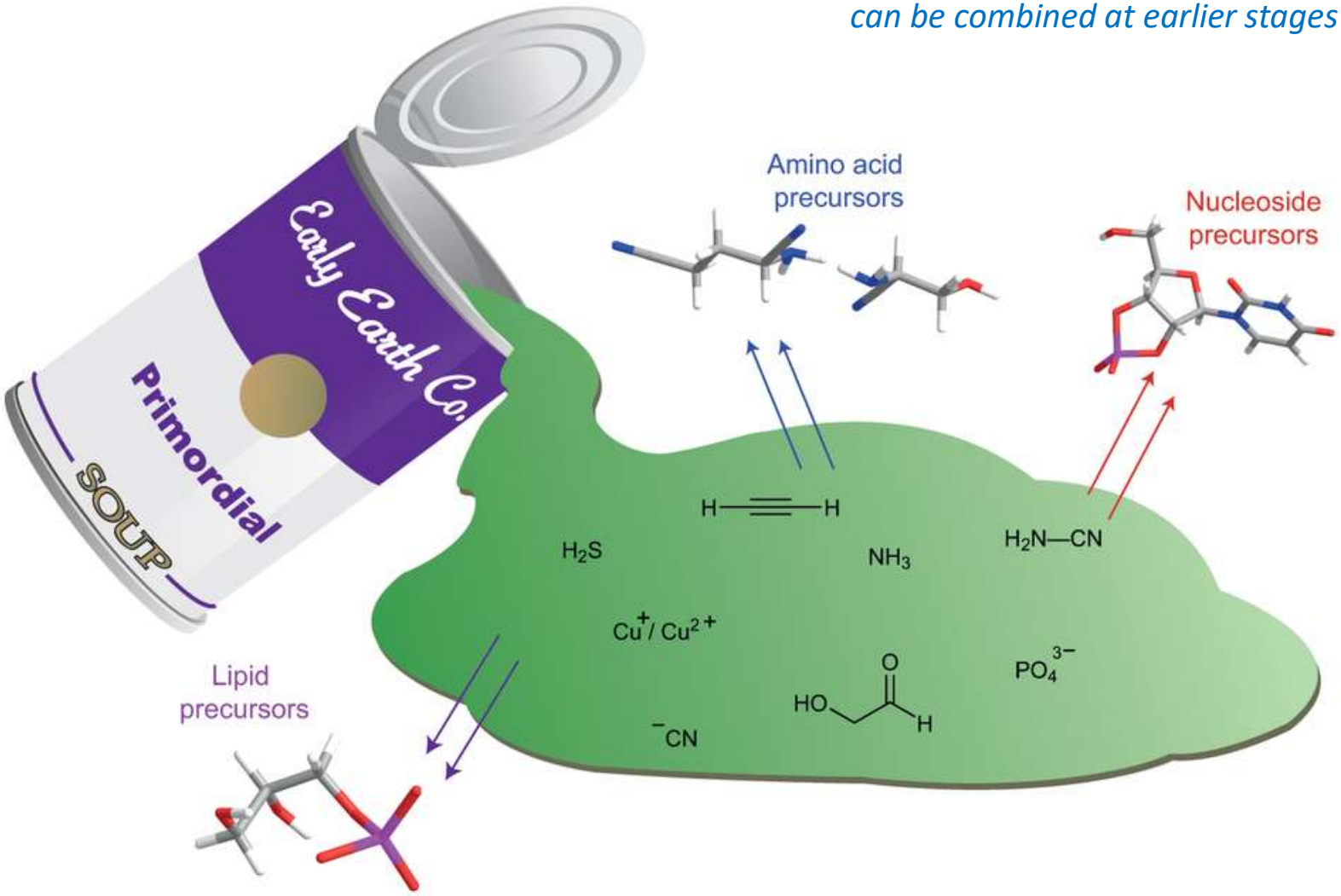


A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner
Science **2004**, *303*, 196

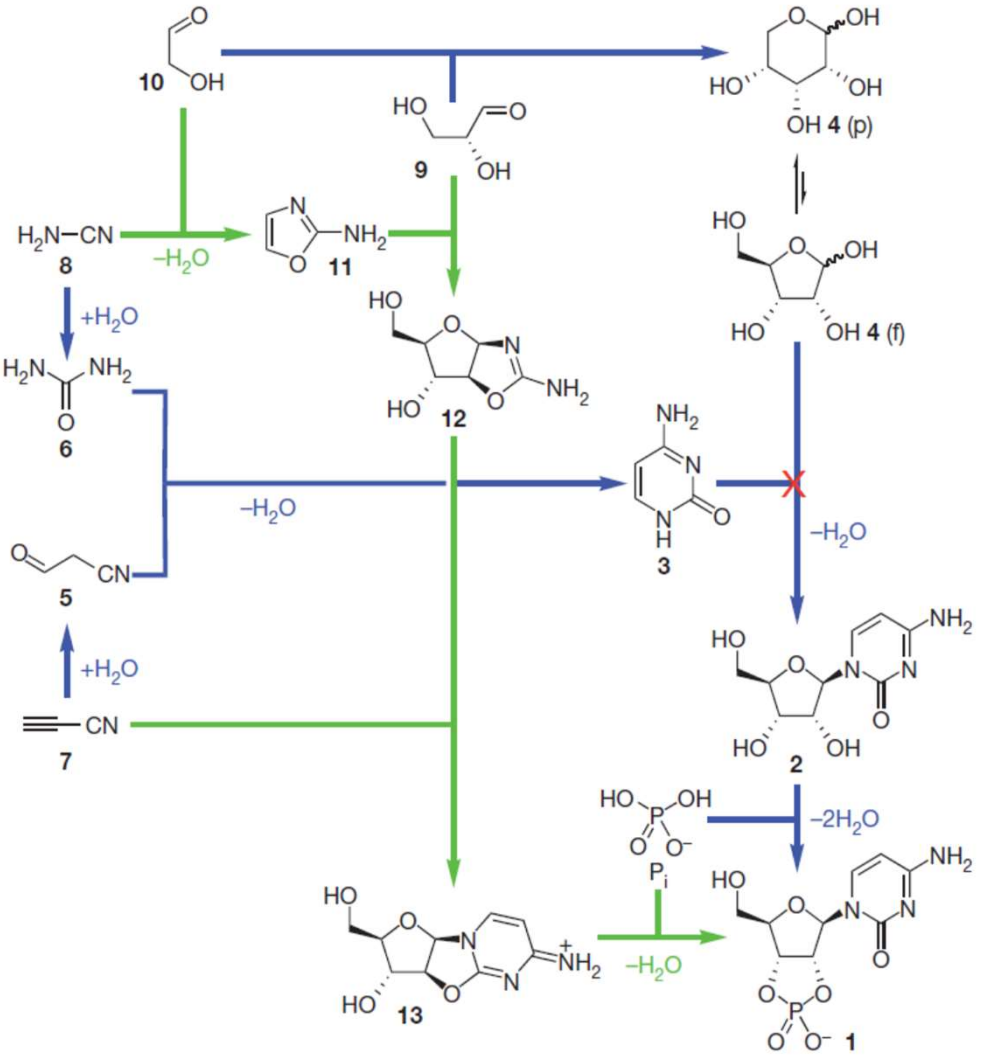
Escosura

Cyanosulfidic chemistry

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

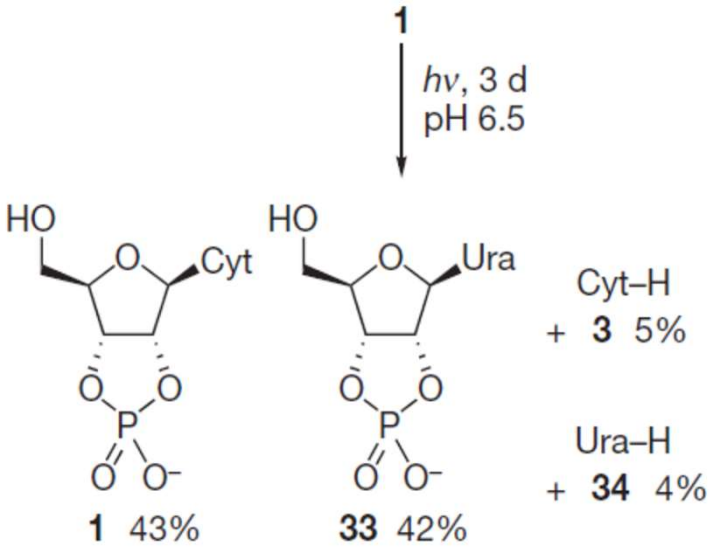


Cyanosulfidic chemistry



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.

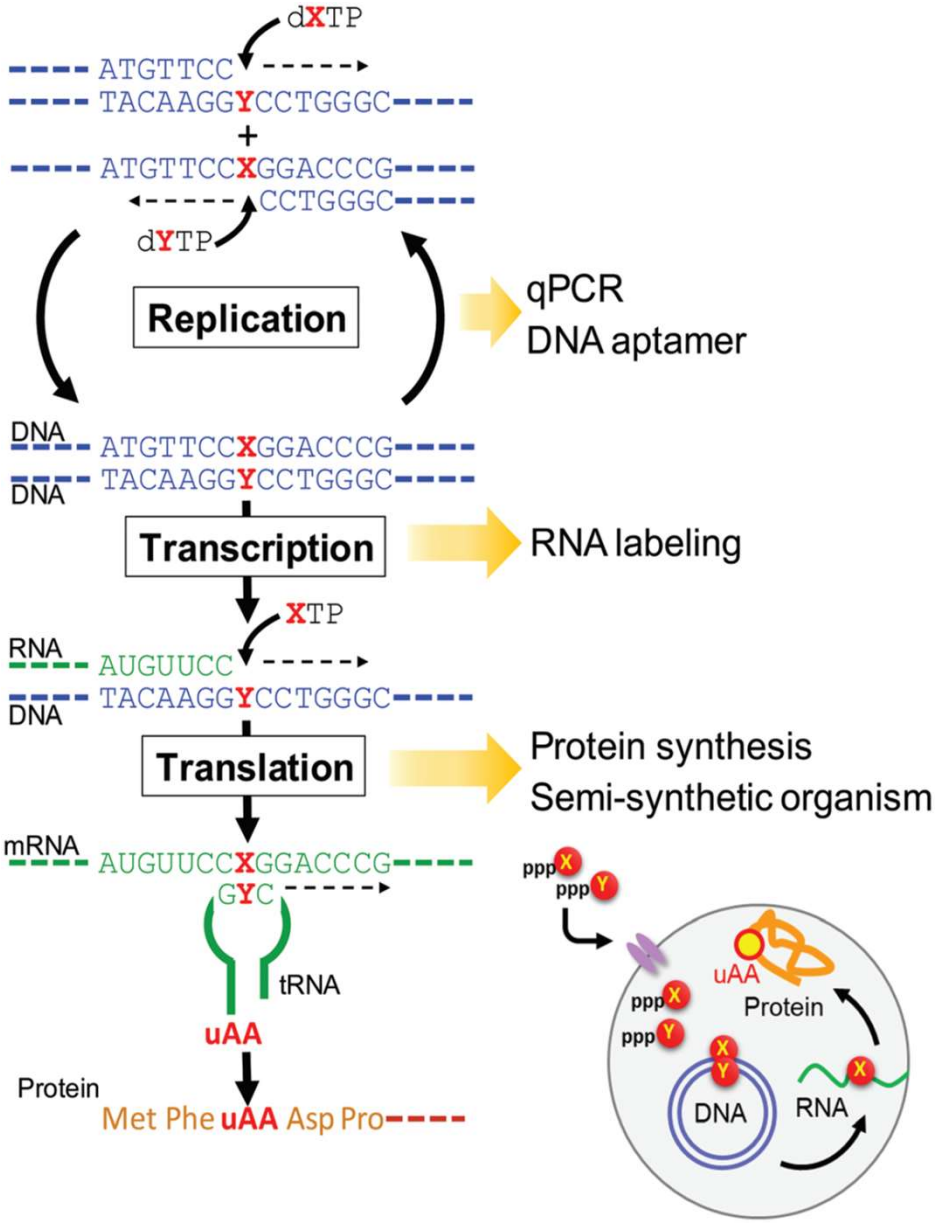
B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem*. **2015**, *7*, 301–307.

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

Expansion of the genetic alphabet

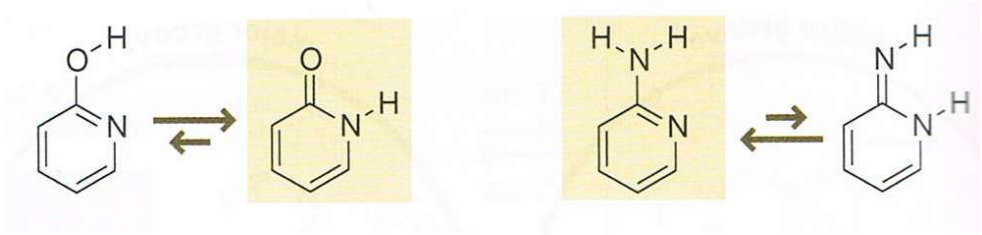
Expansion of the genetic alphabet and code by creating an unnatural base pair (UBP) as a third pair.

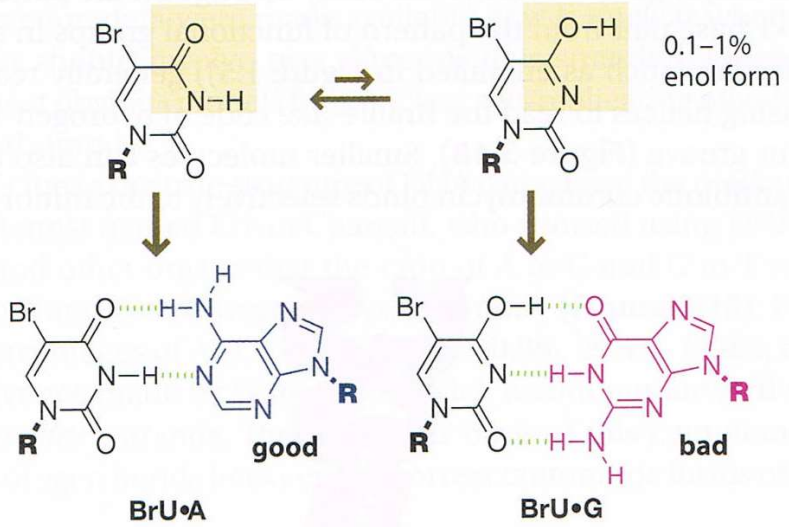
The creation of a UBP (i.e., X–Y) that functions in replication, transcription, and translation as a third base pair with the natural A–T(U) and G–C pairs allows the storage and retrieval of the expanded genetic information in vitro and in vivo, enabling a variety of applications using biopolymers with increased functionalities



M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

Tautomeric instability in non-canonical nucleobases





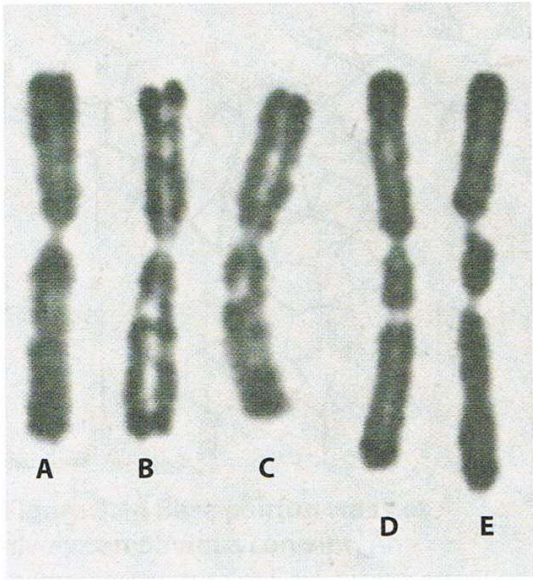
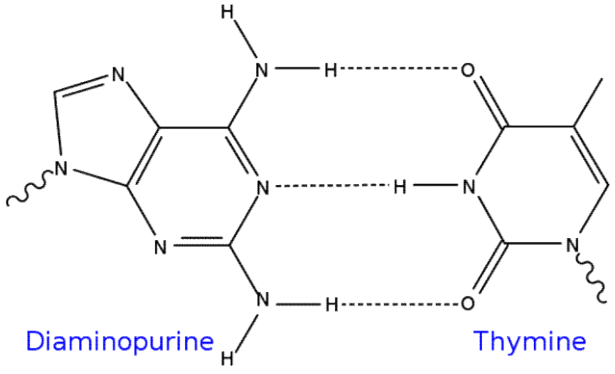
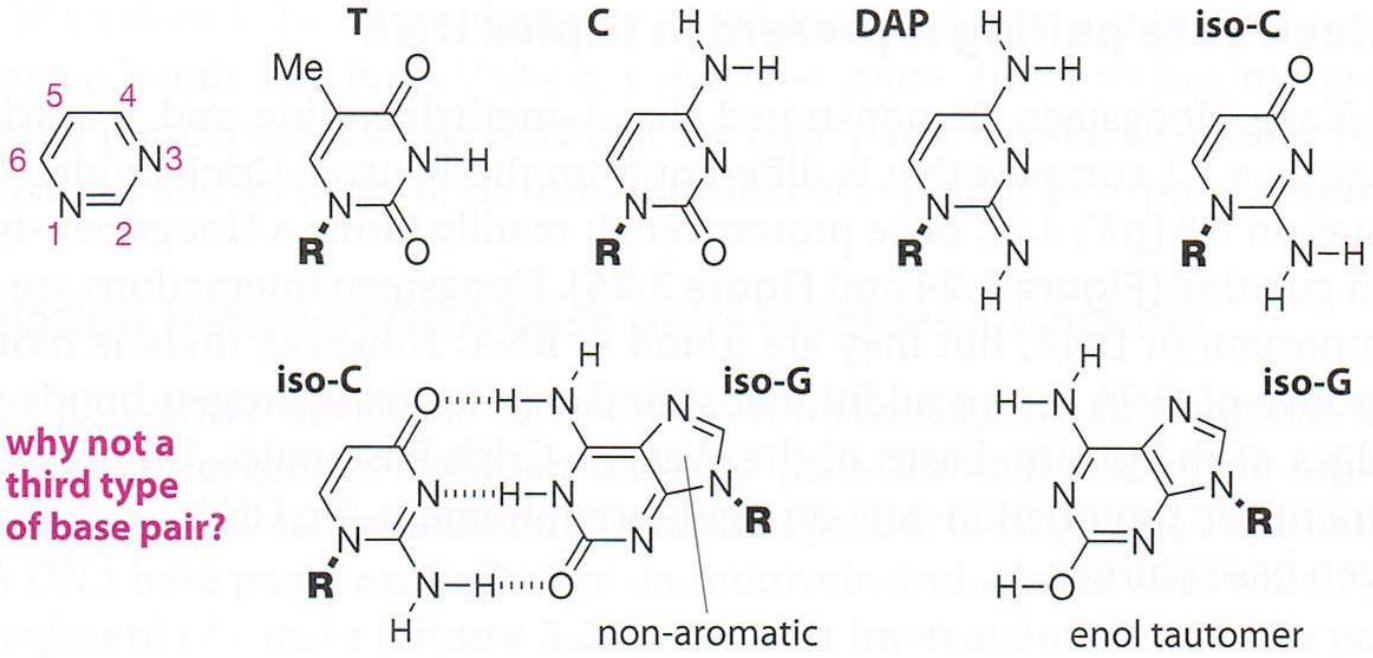


Figure 3.20 Chromosome 1 from hamster cells exposed to bromodeoxyuridine. (A) Normal chromosome. (B–E) Aberrant chromosomes. (From T.C. Hsu and C.E. Somers, *Proc. Natl. Acad. Sci. USA* 47: 396–403, 1961. With permission from the MD Anderson Cancer Center.)

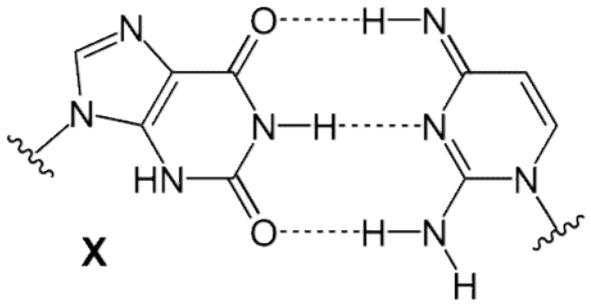
Alternative base pairs – synthetic biology

Cyanophage S-2L uses diaminopurine instead of adenine (3 H-bonds!)





DAP

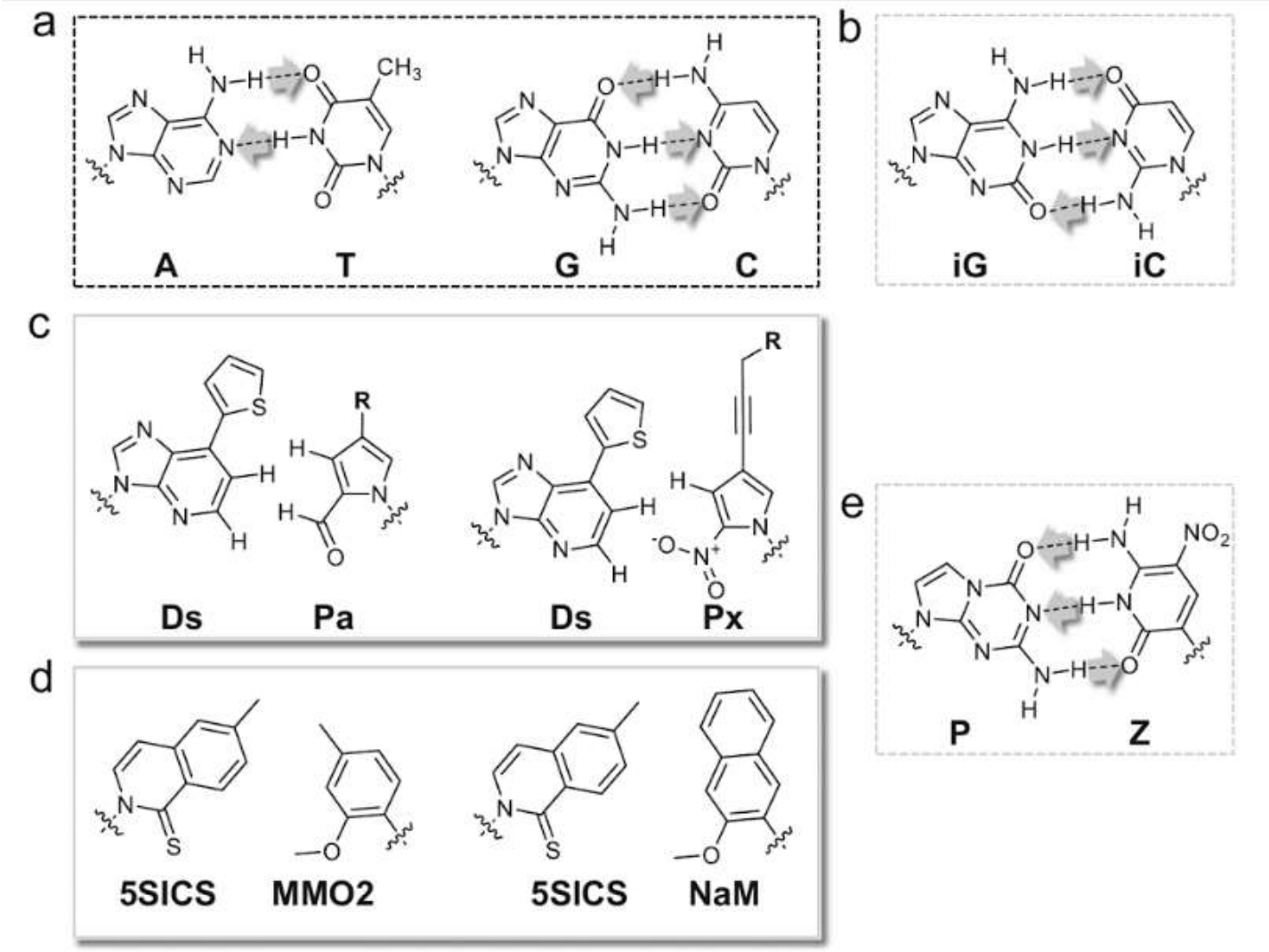


DAP – one tautomer forms a base pair with guanine

iso-C/iso-G

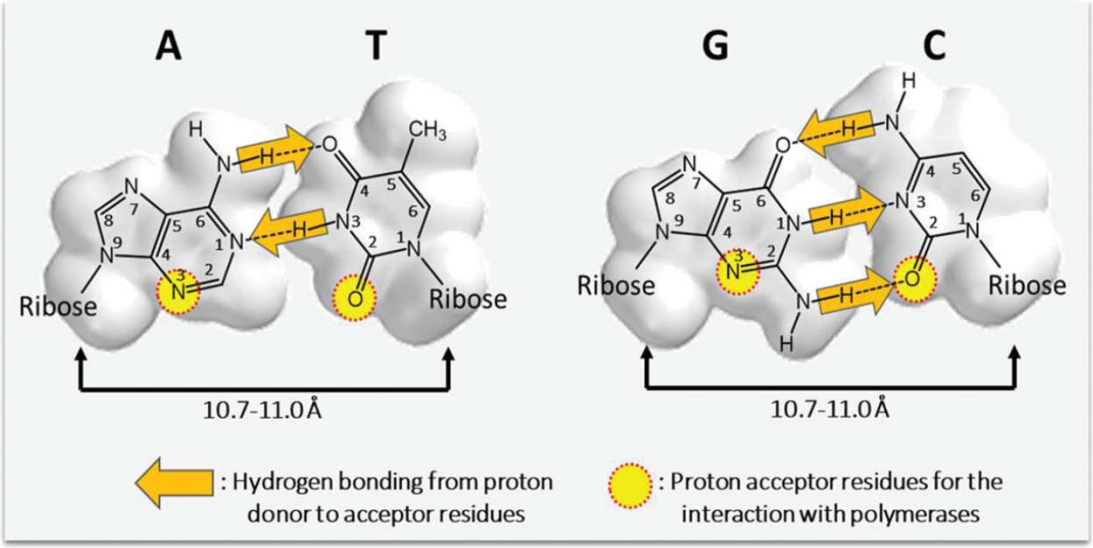
- specificity (the enol tautomer of iso-G, stabilized by aromatization, complementary to thymine)
- the 2-amino group of iso-C hydrolyses easily to uracil

Natural and non-natural base pairs that function in polymerase reactions



Unnatural base pair (UBP) design rules:

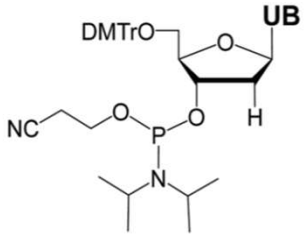
Design



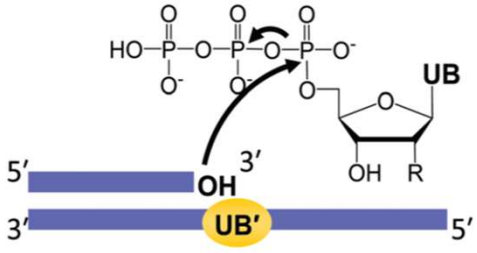
- Distance of 10.7-11.0 Å between the glycosidic bonds of the pair
- no mispairing with natural bases
- Reasonable hydrophobicity
- Chemical stability (phosphoramidite chemistry)
- Recognition by DNA and RNA polymerases (fidelity, efficiency)

Chemical synthesis

Phosphoramidite reagents
for DNA chemical synthesis

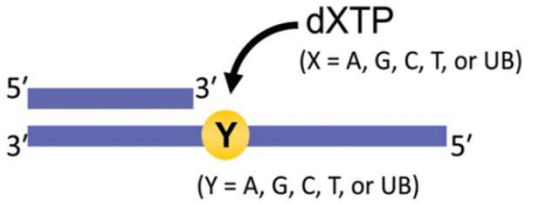


Nucleoside triphosphates
(R = H: d(UB)TP, R = OH: (UB)TP)

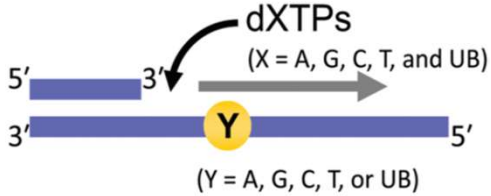


Polymerase reaction tests

Single-nucleotide insertion

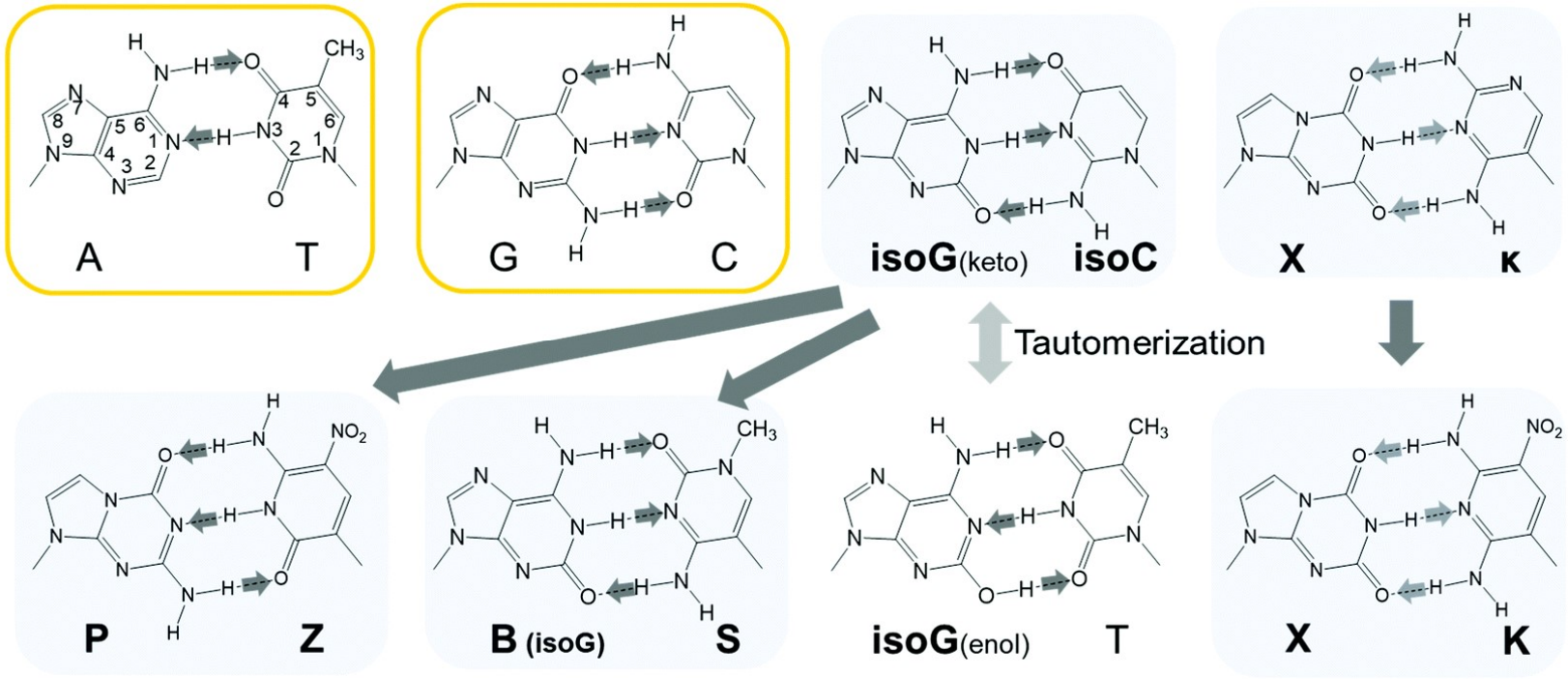


Primer extension



M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

Non-natural nucleobases compatible with polymerases



Alexander Rich (1962): isoG–isoC pair

Steven Benner (1989-95): the artificially expanded genetic information system (AEGIS) including the isoG–isoC and X–κ pairs → *in vitro* replication, transcription, and translation systems.

Benner and Prudent (2004): new quantitative PCR (qPCR) methods, such as Plexor, using the isoG–isoC pair

Benner (2007): Z-P pair - the P base by removing the hydrogen at position 1 of G to exclude the keto–enol tautomerism, by introducing the nitro group into Z, the chemical stability of the nucleoside was improved;

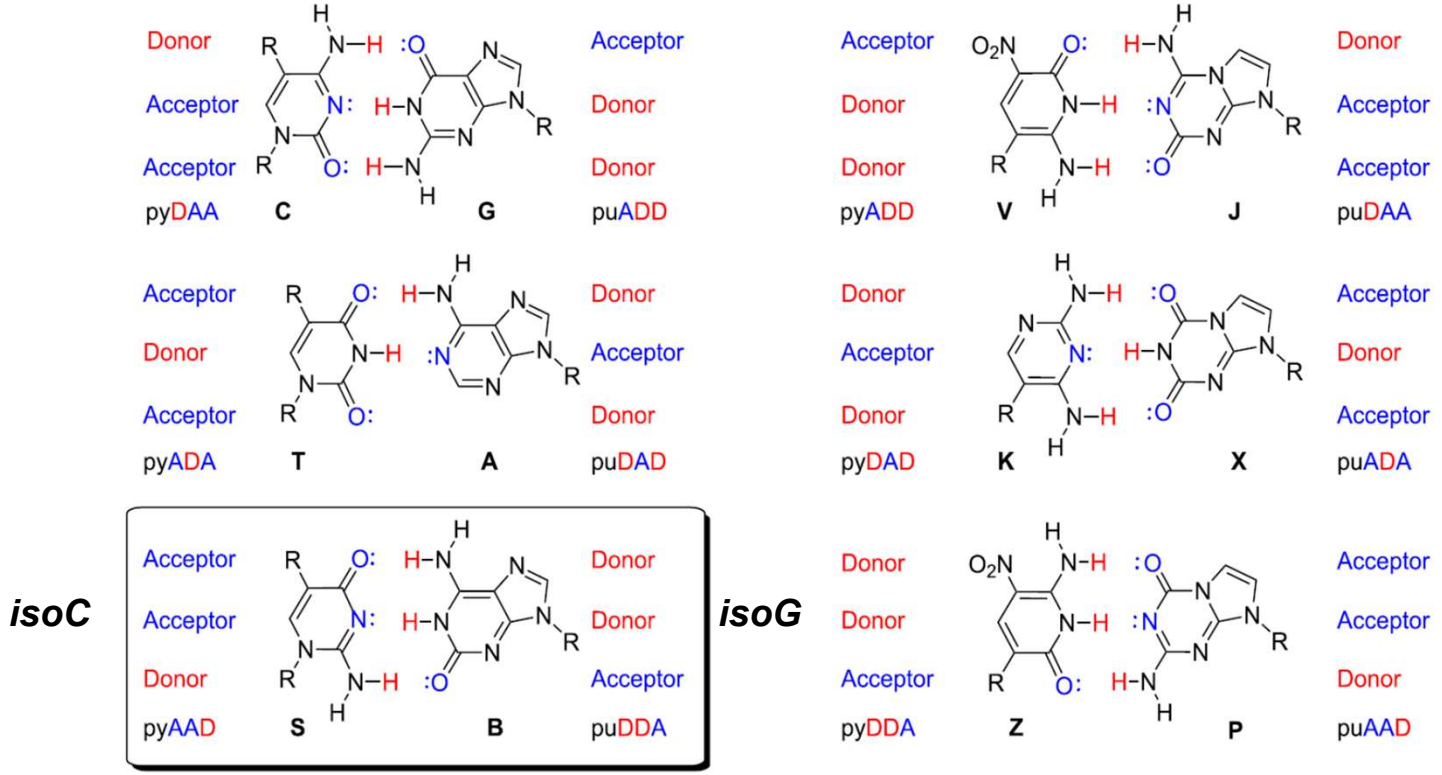
With the same strategy, they also developed the B–S and X–K pairs from isoG–isoC and X–κ pairs, respectively

AEGIS – Artificially Expanded Genetic Information System

Watson–Crick pairing rules:

(a) size complementarity - large purines pair with small pyrimidines

(b) hydrogen-bonding complementarity (hydrogen-bond acceptors, A, pair with hydrogen-bond donors, D).



Rearranging donor and acceptor groups on the nucleobases, while not changing the geometry of the Watson–Crick pair, creates an artificially expanded genetic information system (AEGIS). AEGIS components add information density to DNA strands built from them.

S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

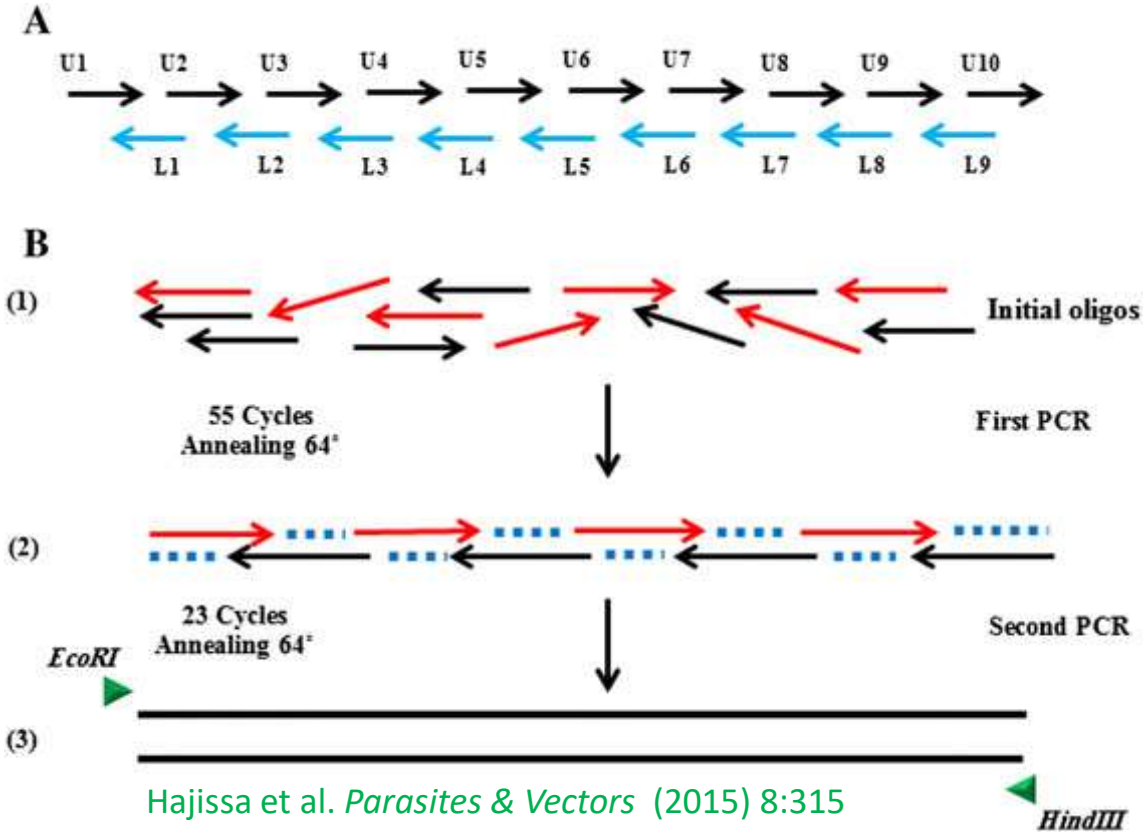
Artificial Gene Synthesis

Artificial gene synthesis (DNA printing) - method in synthetic biology to create artificial genes in the laboratory:

- *currently based on solid-phase DNA synthesis,*
- *the user does not have to begin with preexisting DNA sequences.*
- *Therefore, it is possible to make a completely synthetic double-stranded DNA molecule with no apparent limits on either nucleotide sequence or size.*

Applications:

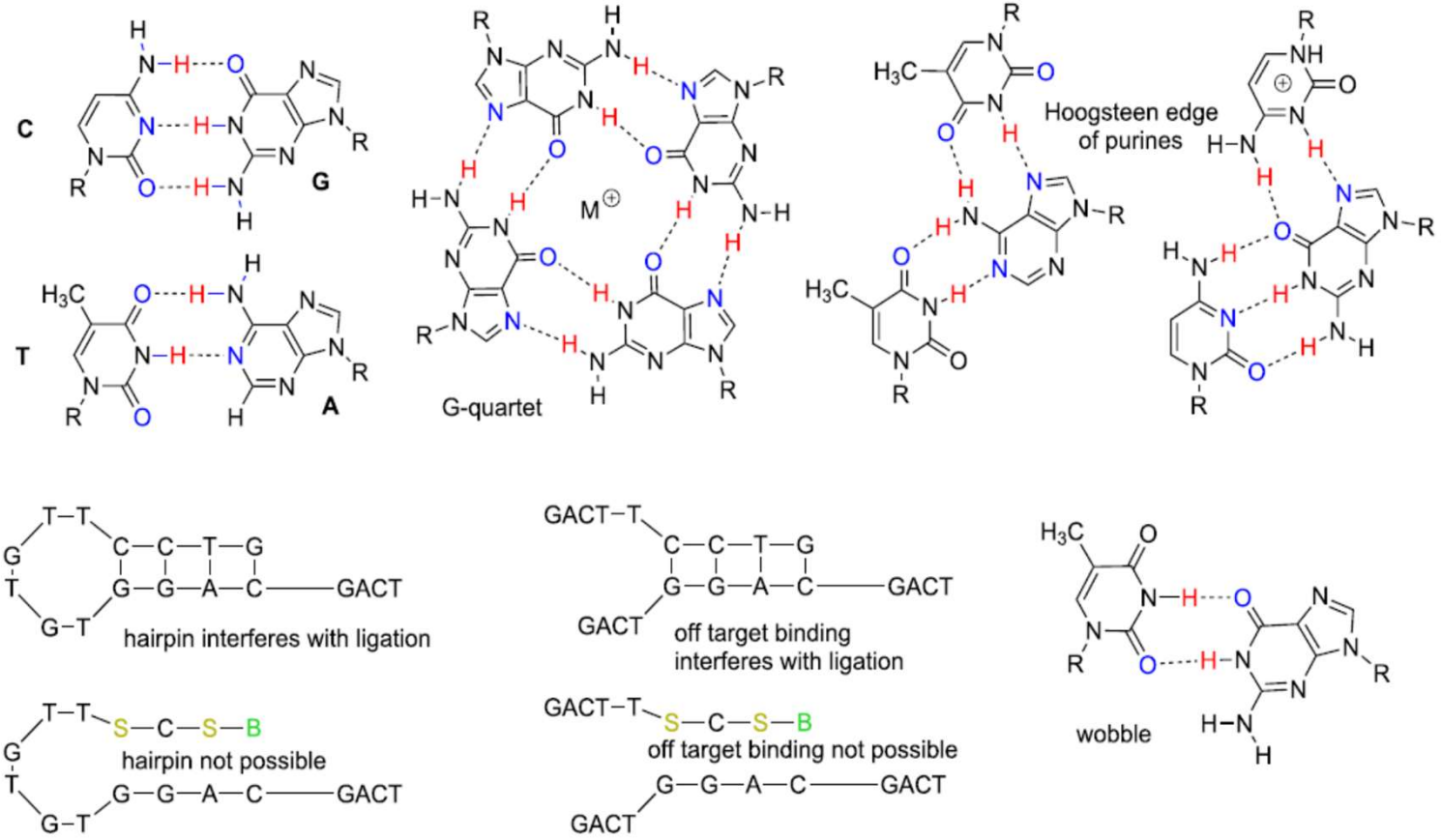
- recombinant DNA technology including heterologous gene expression, vaccine development, gene therapy and molecular engineering.
- The synthesis of nucleic acid sequences can be more economical than classical cloning and mutagenesis procedures
- the ability to safely obtain genes for vaccine research without the need to grow the full pathogens.
- to optimize protein expression in a particular host, or to remove non-functional DNA segments
- For DNA digital data storage and computing
- For synthetic biological circuits



Hajissa et al. *Parasites & Vectors* (2015) 8:315

Self-assembly of whole genes and DNA nanostructures

Limitations of DNA puzzle assembly: unequal A:T vs. G:C strength, insufficient ACGT information density, higher-order structures

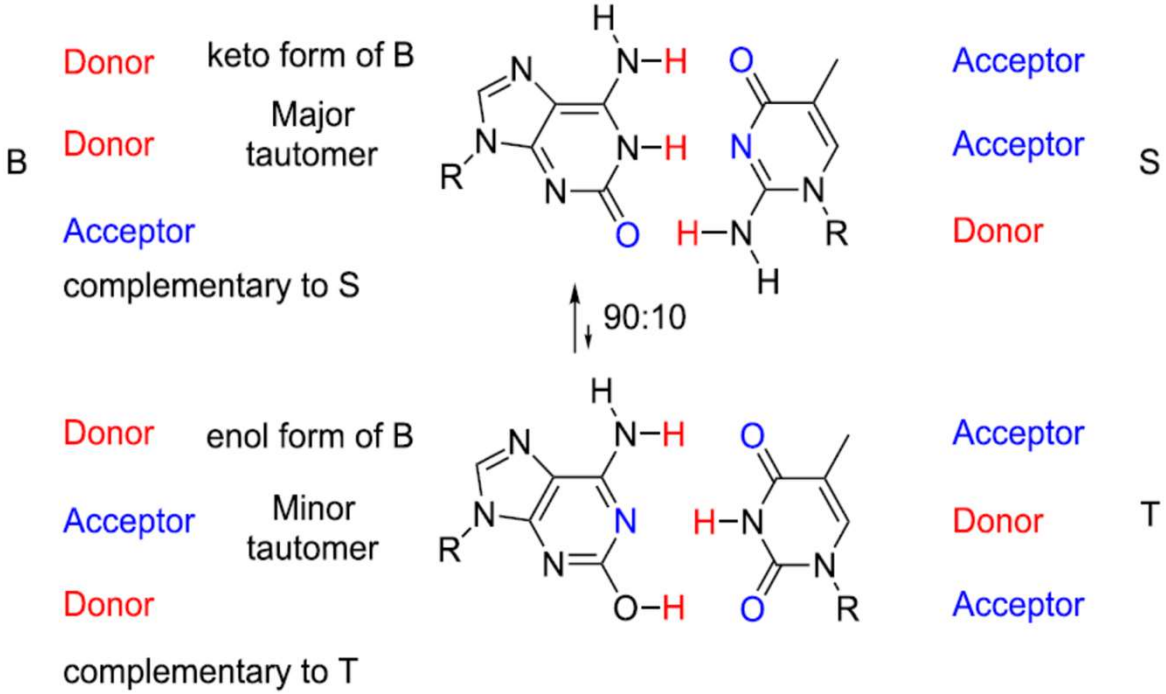


S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

Self-assembly of whole genes and DNA nanostructures

Solution: an orthogonal pair from the AEGIS system, that can be removed from the product, yielding native DNA structures

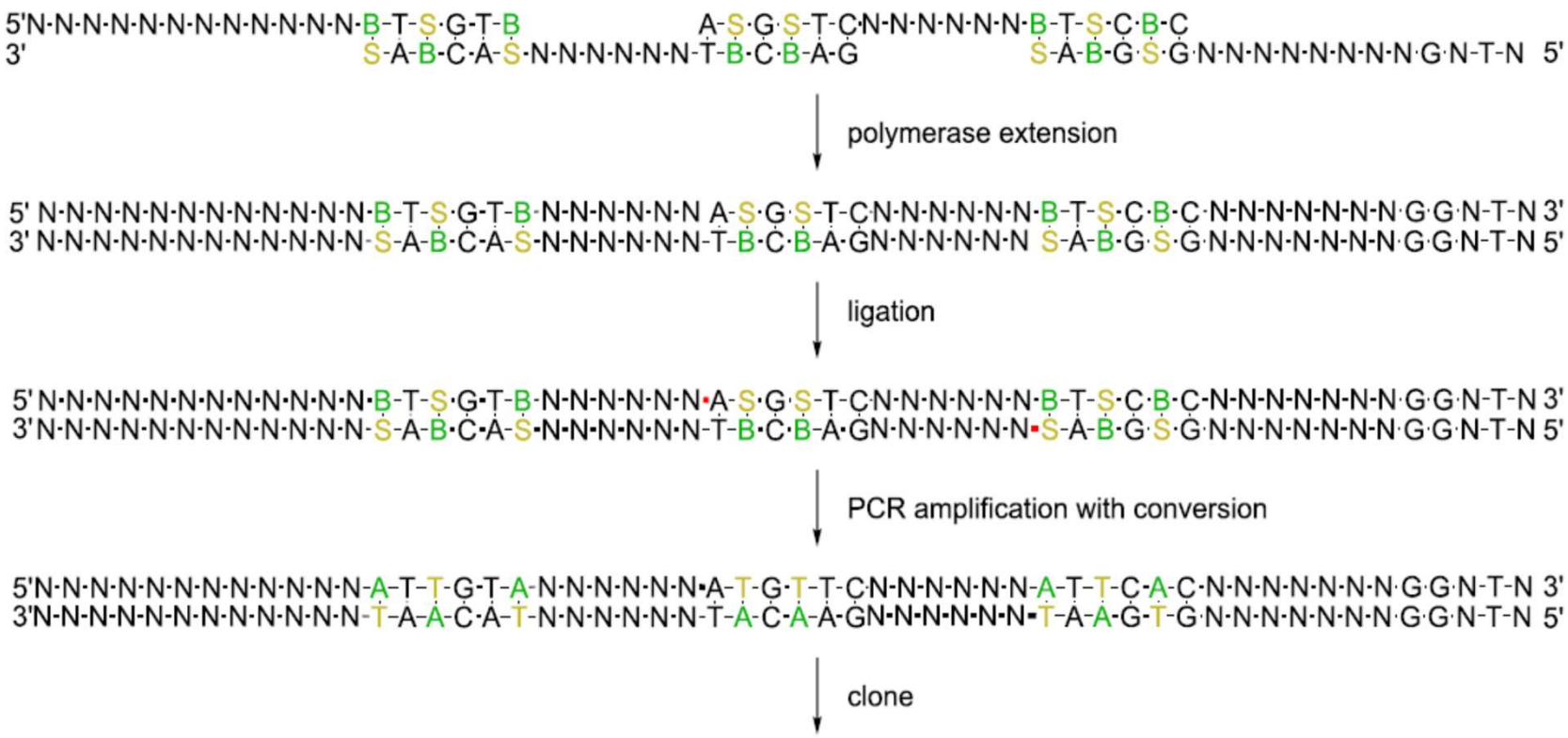
Conversion occurs when polymerases are forced to mismatch a standard nucleotide opposite an AEGIS nucleotide by
(a) not being provided the complementary AEGIS triphosphate and
(b) exploiting a chemical feature of the AEGIS nucleotide that directs a specific mismatch.



B in its major tautomeric form pairs with **S**; in its minor tautomeric form, **B** pairs with standard **T**. Asembly of the target gene/DNA nanostructure is followed by conversion of the **S:B** pairs to **T:A** pairs after two cycles of PCR:
B → **A** via an intermediate **B:T** mispairing, **S** → **T** (intermediate **S:B** followed by a second **B:T** mispairing).

S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

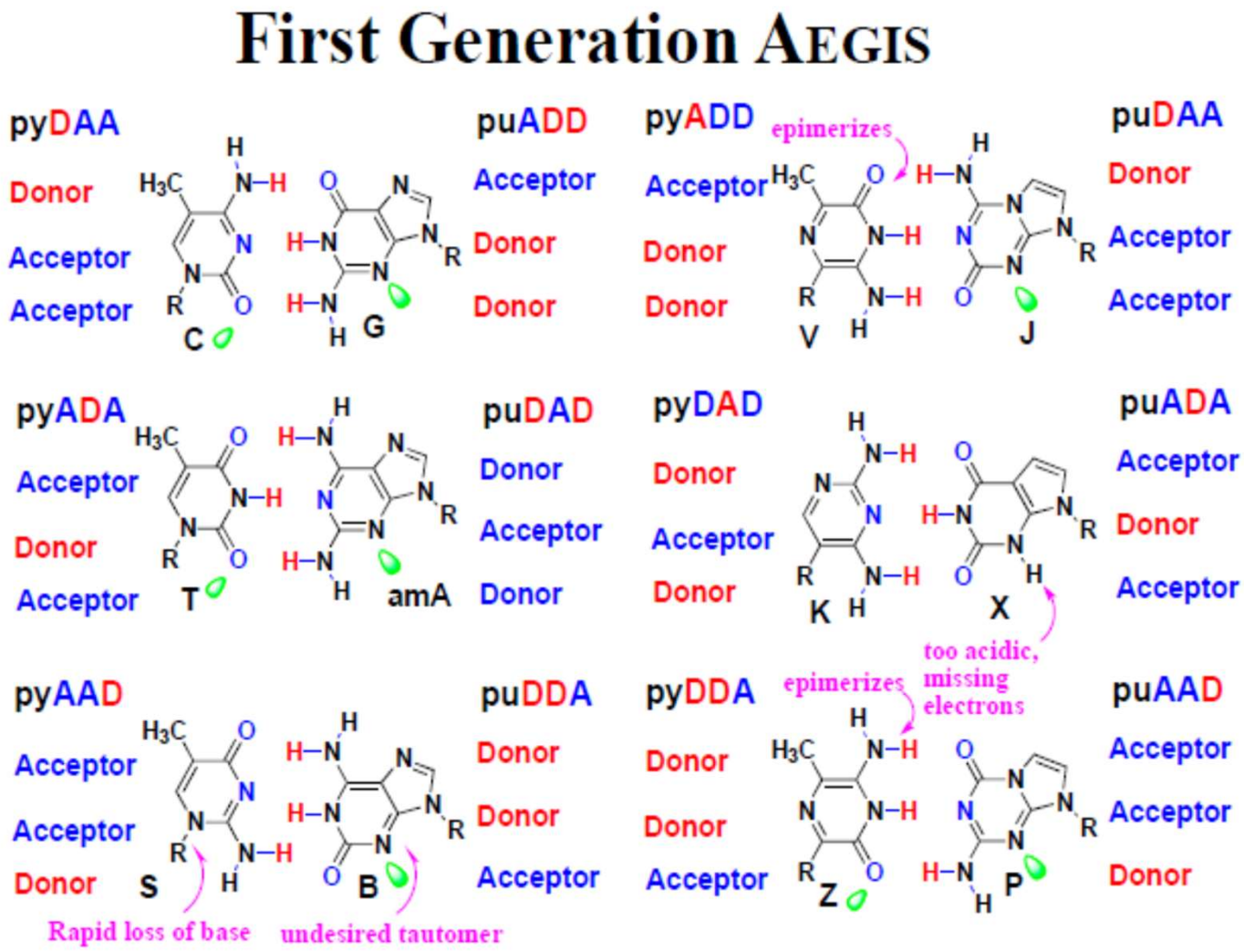
Self-assembly of whole genes and DNA nanostructures



The technology tested by assembly of the kanamycin-resistance gene and growing the bacteria in the environment containing kanamycin after assembly and conversion of that gene.

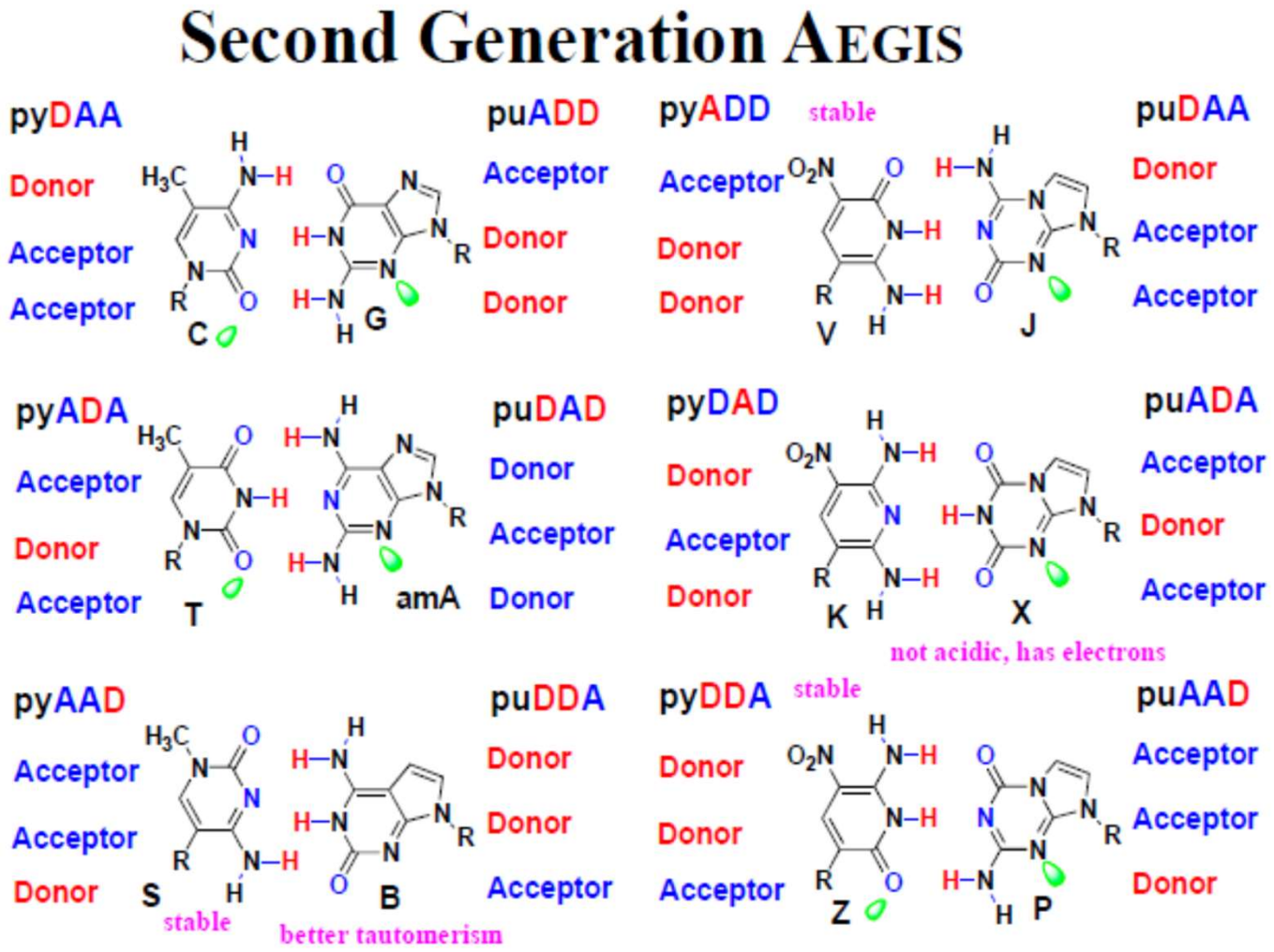
S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

AEGIS – Artificially Expanded Genetic Information System



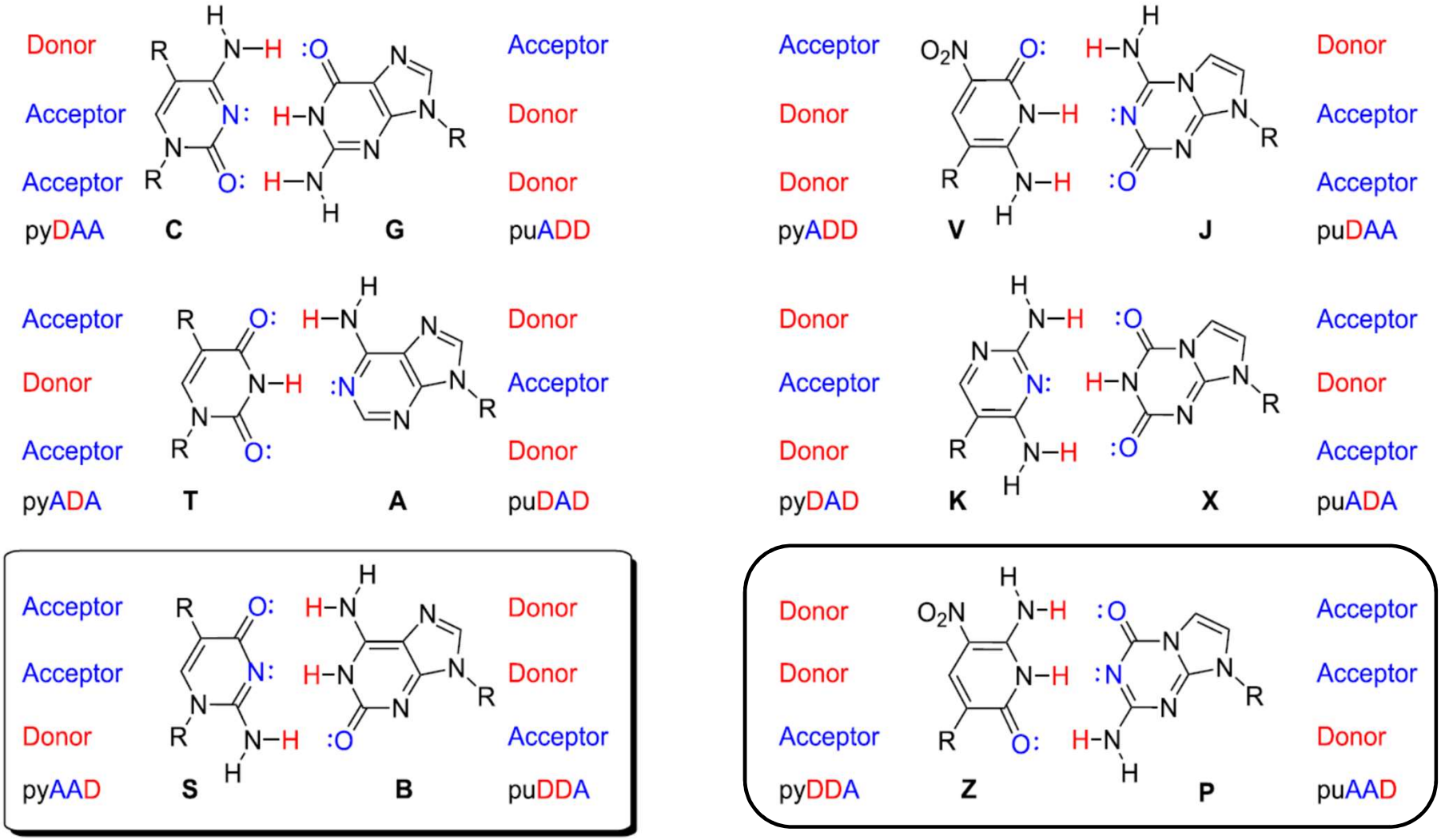
E. Biondi, S. Benner *Biomedicines* **2018**, *6*, 53; *doi:10.3390/biomedicines6020053*

AEGIS – Artificially Expanded Genetic Information System



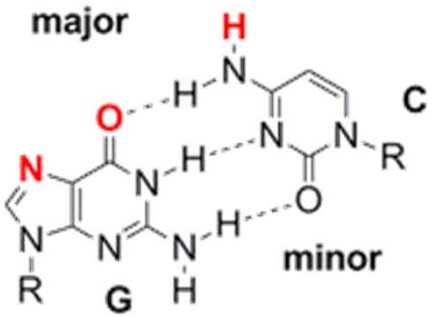
E. Biondi, S. Benner *Biomedicines* **2018**, *6*, 53; *doi:10.3390/biomedicines6020053*

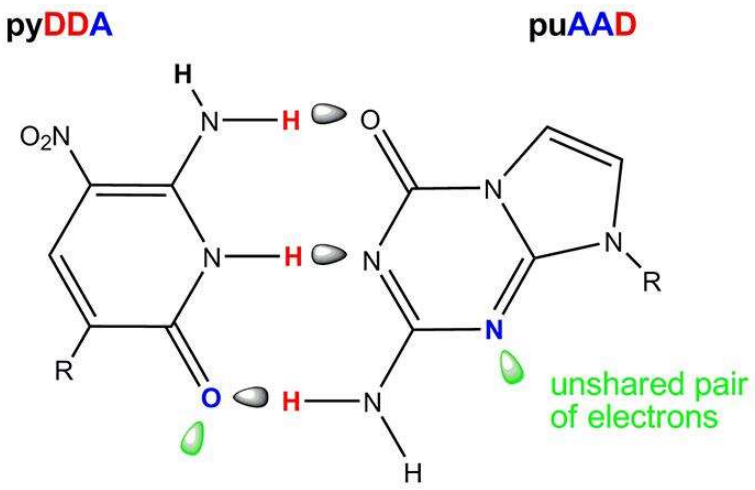
AEGIS – Artificially Expanded Genetic Information System

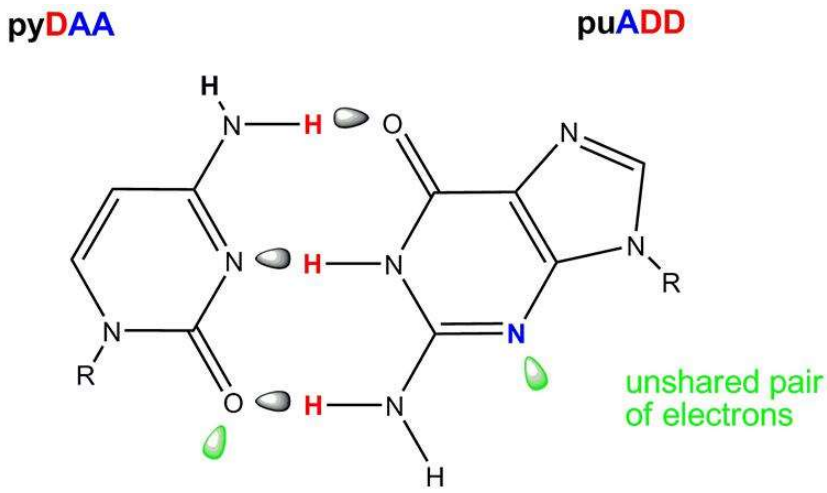


S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

AEGIS – Artificially Expanded Genetic Information System



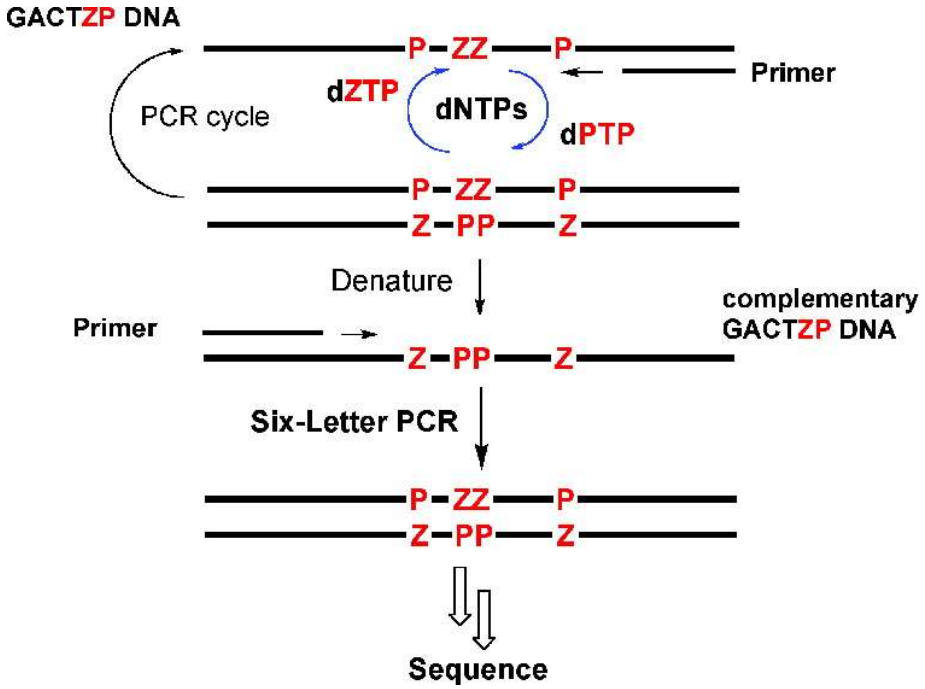




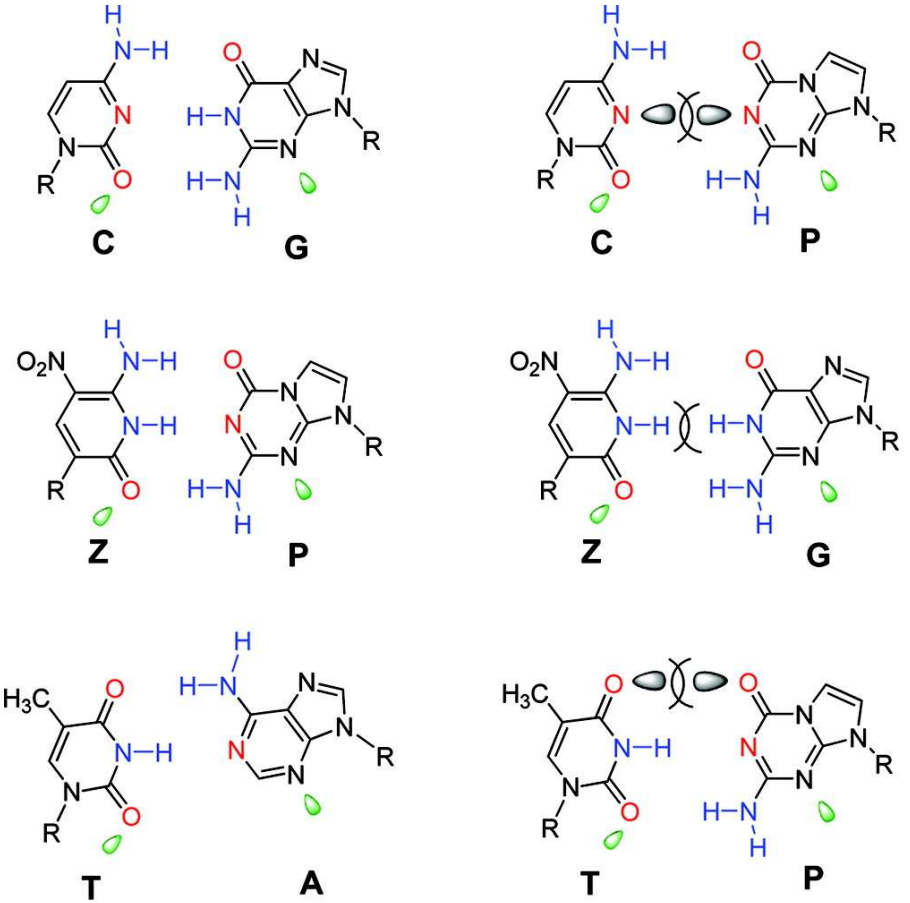
Z **P** **C** **G**

S. Benner *et al., J. Am. Chem. Soc.,* **2011**, *133 (38),* pp 15105–15112

AEGIS – Permanent orthogonal nucleobases surviving PCR



Electron density presented to the minor groove
→ recognition site by polymerases
„ minor groove scanning hypothesis"

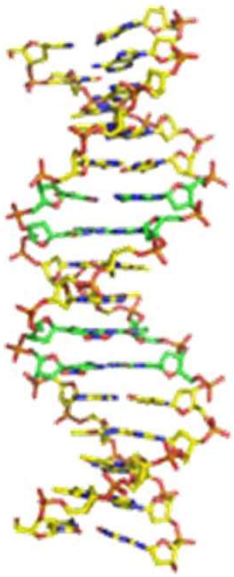


Error rate 0,2% per a PCR cycle – both removal and incorporation of **Z** and **P** ➔ the artificial genetic system capable to evolve.

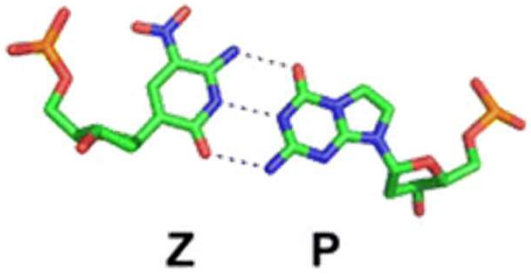
Pol: Deep Vent – 2 Z/P, Taq/Phu – 3-4 Z/P
dZTP (*deprotonated*) at higher pH pairs slightly with G
→ loss of some **Z**, but also gain of some new **Z** mutants.

S. Benner *et al., J. Am. Chem. Soc.,* **2011**, *133 (38),* pp 15105–15112

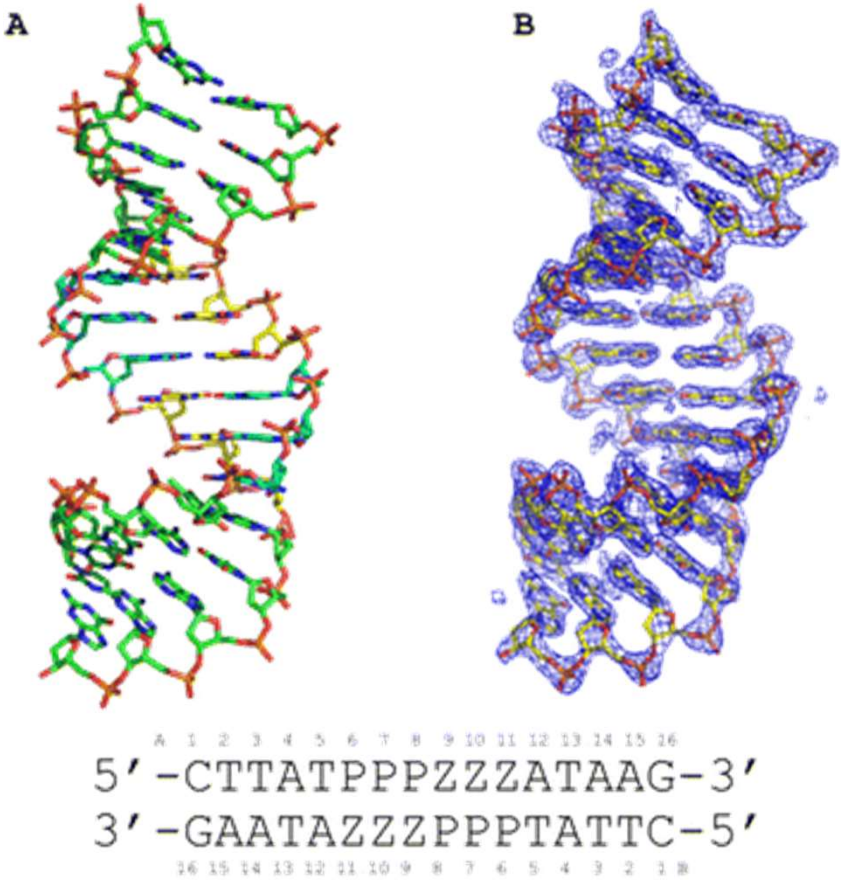
ACGTZP-DNA crystal structures



B-DNA



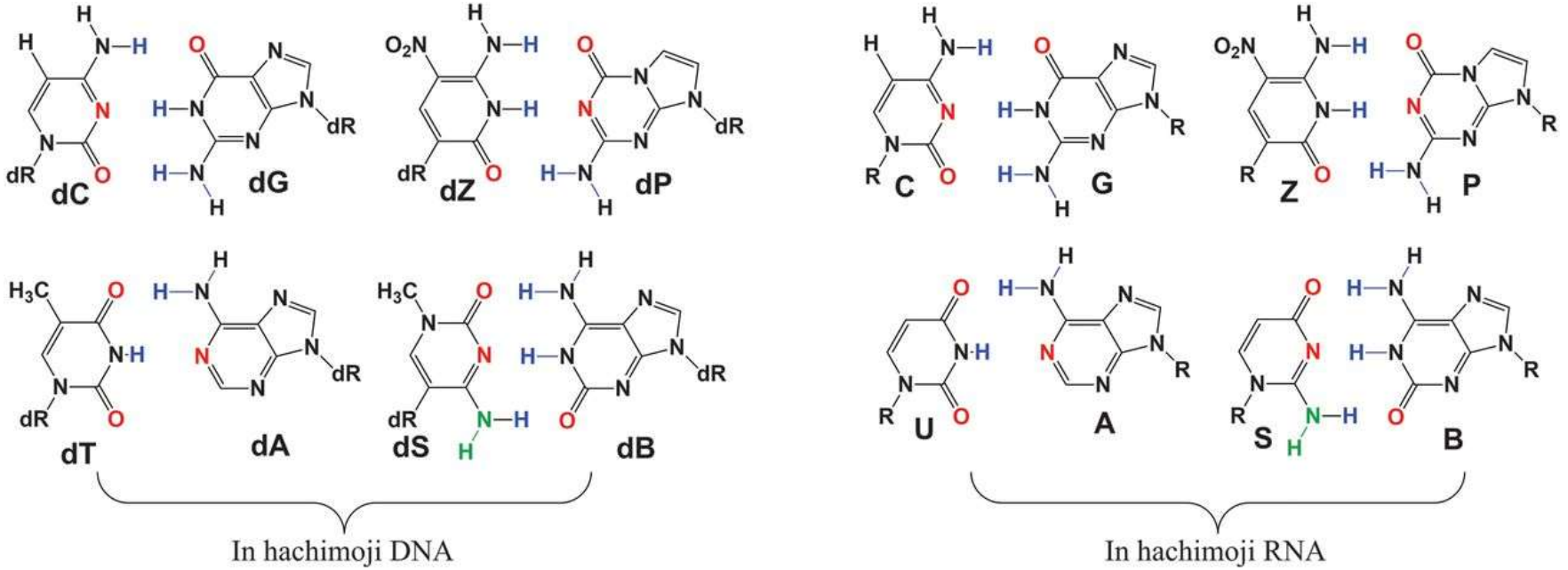
A-DNA

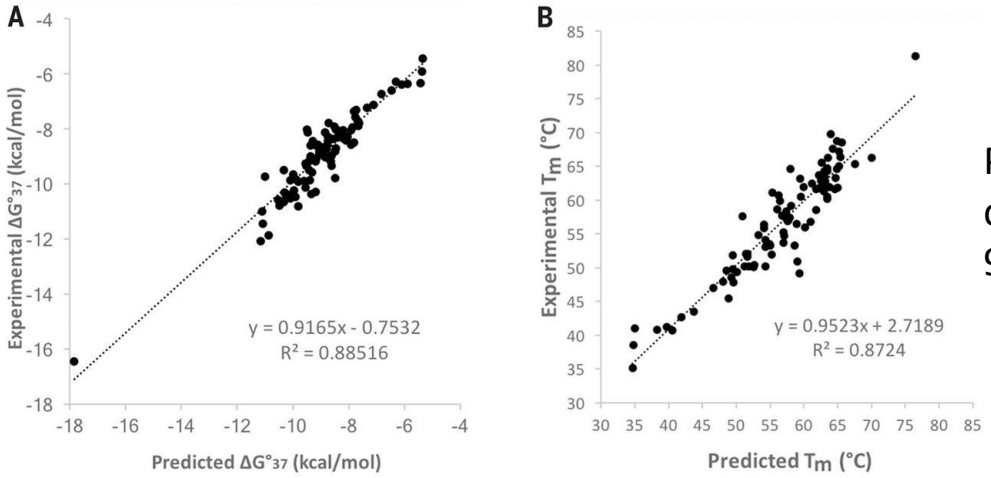


*18-mers: 2+2 **Z:P** pairs → B-DNA*
*6 consecutive **Z:P** → A-DNA*
*0,1 nm wider, but otherwise alike **G:C** pairs*

S. Benner *et al., J. Am. Chem. Soc.*, **2015**, *137*, pp 6947–6955

Hachimoji DNA and RNA – a genetic system with eight (Jap.- Hachi) letters

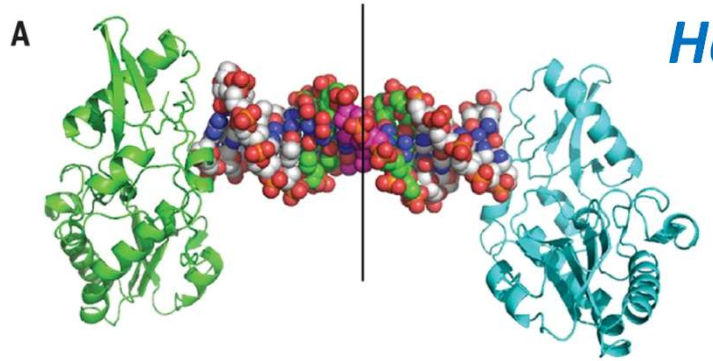




Plots of experimental versus predicted **(A)** free energy changes (ΔG°_{37}) and **(B)** melting temperatures T_m for 94 SBZP-containing hachimoji DNA duplexes.

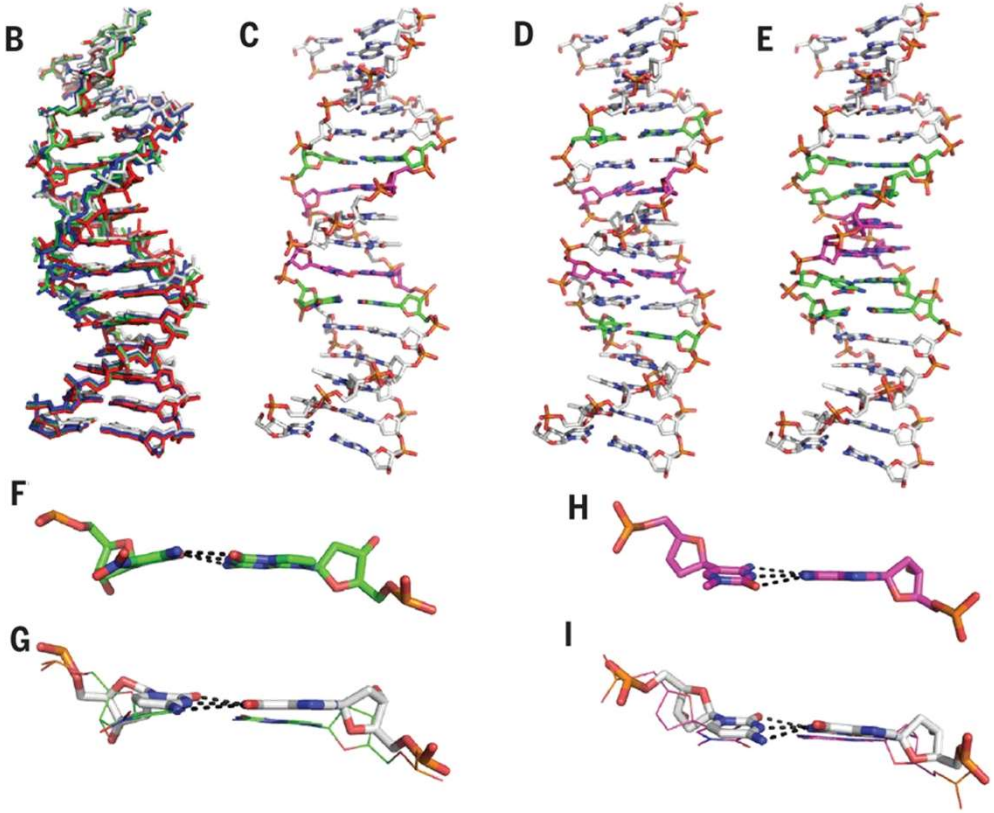
S. Hoshika *et al., Science* **2019**, *363,* 884-887

Hachimoji DNA and RNA



Crystal structures of hachimoji DNA.

(**A**) The host-guest complex with two N-terminal fragments from Moloney murine leukemia virus reverse transcriptase bound to a 16-mer PP hachimoji DNA; Z:P pairs are green and S:B pairs are magenta.



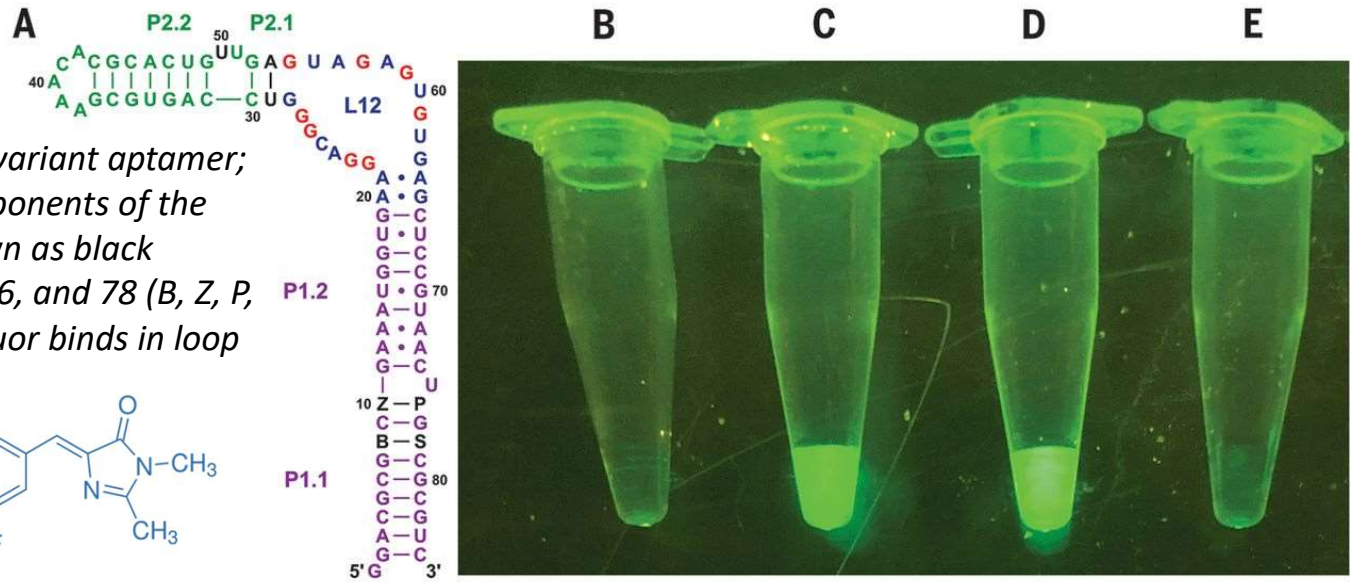
(**B**) Hachimoji DNA structures PB (green), PC (red), and PP (blue) are superimposed with GC DNA (gray).
(**C**) Structure of hachimoji DNA with self-complementary duplex 5′-CTTATPBTASZATAAG (“PB”).
(**D**) Structure of hachimoji DNA with self-complementary duplex 5′-CTTAPCBTASGZTAAG (“PC”).
(**E**) Structure of hachimoji DNA with self-complementary duplex with six consecutive nonstandard 5′-CTTATPPSBZZATAAG (PP) components.

(**F** to **I**) Examples of largest differences in detailed structures. The Z:P pair from the PB structure (F) is more buckled than the corresponding G:C pair (G). The S:B pair from the PB structure (H) exhibits a propeller angle similar to that in the corresponding G:C pair (I).

S. Hoshika *et al., Science* **2019**, *363,* 884-887

Hachimoji RNA aptamer

T7 RNA polymerase incorporates ZTP, PTP, and BTP, but not STP opposite to dP, dZ, dS, and dB, respectively. A mutant of T7 RNA Pol (Y639F H784A P266L, "**FAL**") incorporated also STP – full DNA→RNA conversion possible



the full hachimoji spinach variant aptamer; additional nucleotide components of the hachimoji system are shown as black letters at positions 8, 10, 76, and 78 (B, Z, P, and S, respectively). The fluor binds in loop L12

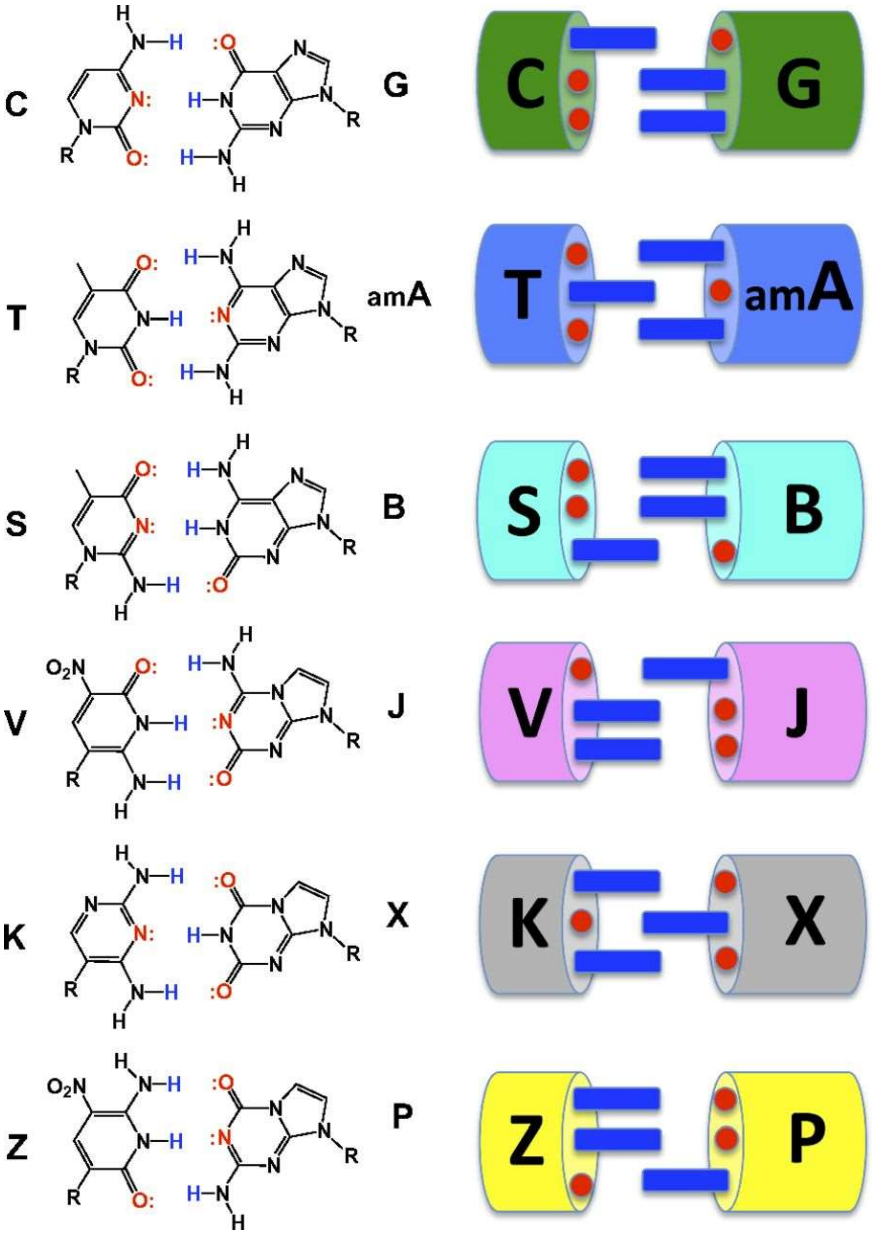
The hachimoji variant of the spinach fluorescent RNA aptamer. In its standard form, spinach folds and binds 3,5-difluoro-4-hydroxybenzylidene imidazolinone, which fluoresces green when bound.

(B) Control with fluor only, lacking RNA.
(C) Hachimoji spinach with the sequence shown in (A).
(D) Native spinach aptamer with fluor.
(E) Fluor and spinach aptamer containing Z at position 50, replacing the A:U pair at positions 53:29 with G:C to restore the triple observed in the crystal structure. This places the quenching Z chromophore near the fluor;

S. Hoshika *et al., Science* **2019**, *363,* 884-887

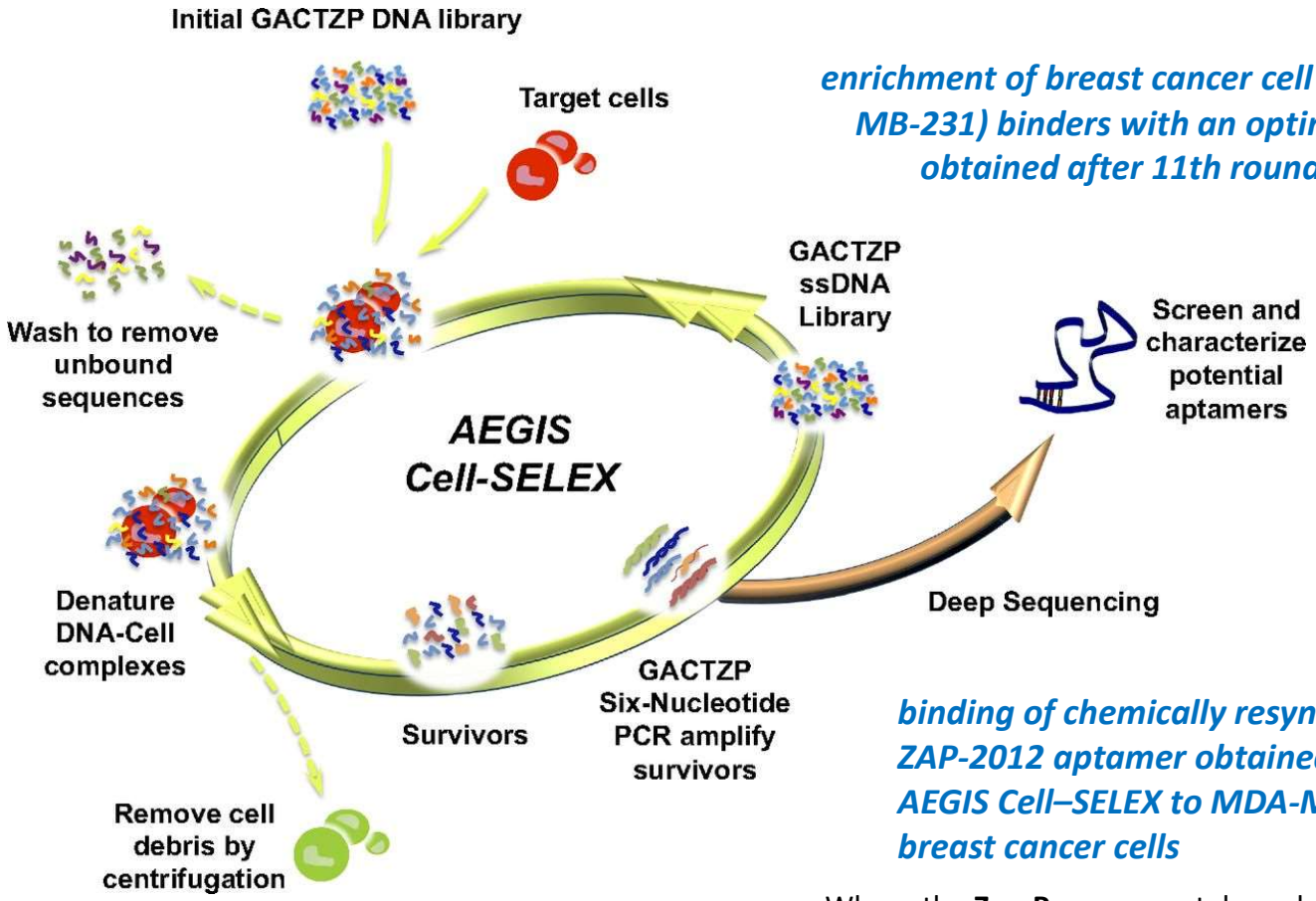
AEGIS – Artificially Expanded Genetic Information System

*An **xNA** biopolymer having functionalized AEGIS components may allow SELEX to yield protein-like aptamers better than the standard DNA and RNA biopolymers.*

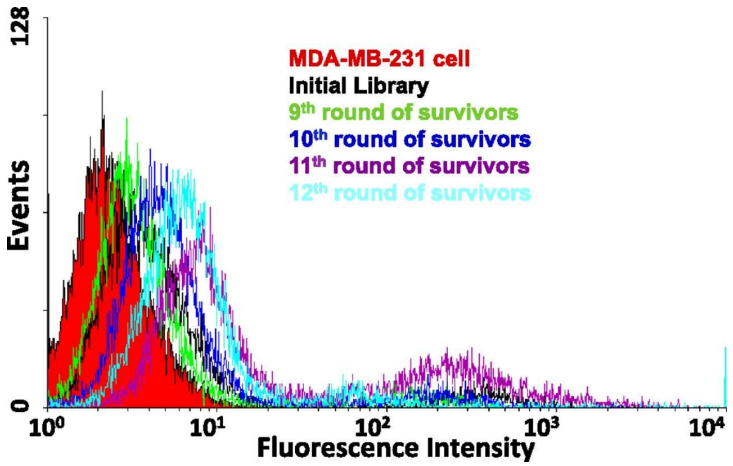


K. Sefah *et al., Proc. Natl. Acad. Sci. USA* **2014,** *111 (4),* 1449-1454.

ACGTZP-aptamers

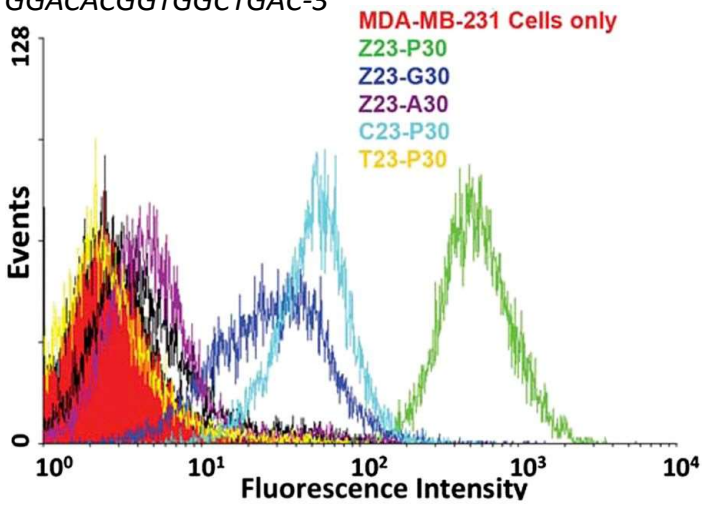


enrichment of breast cancer cell (MDA-MB-231) binders with an optimum obtained after 11th round



ZAP-2012 (Z23-P30) K_d = 30 nM

5′-biotin-TCCCGAGTGACGCAGC-
CCCCGGZGGGATTPATCGGT-
GGACACGGTGGCTGAC-3′

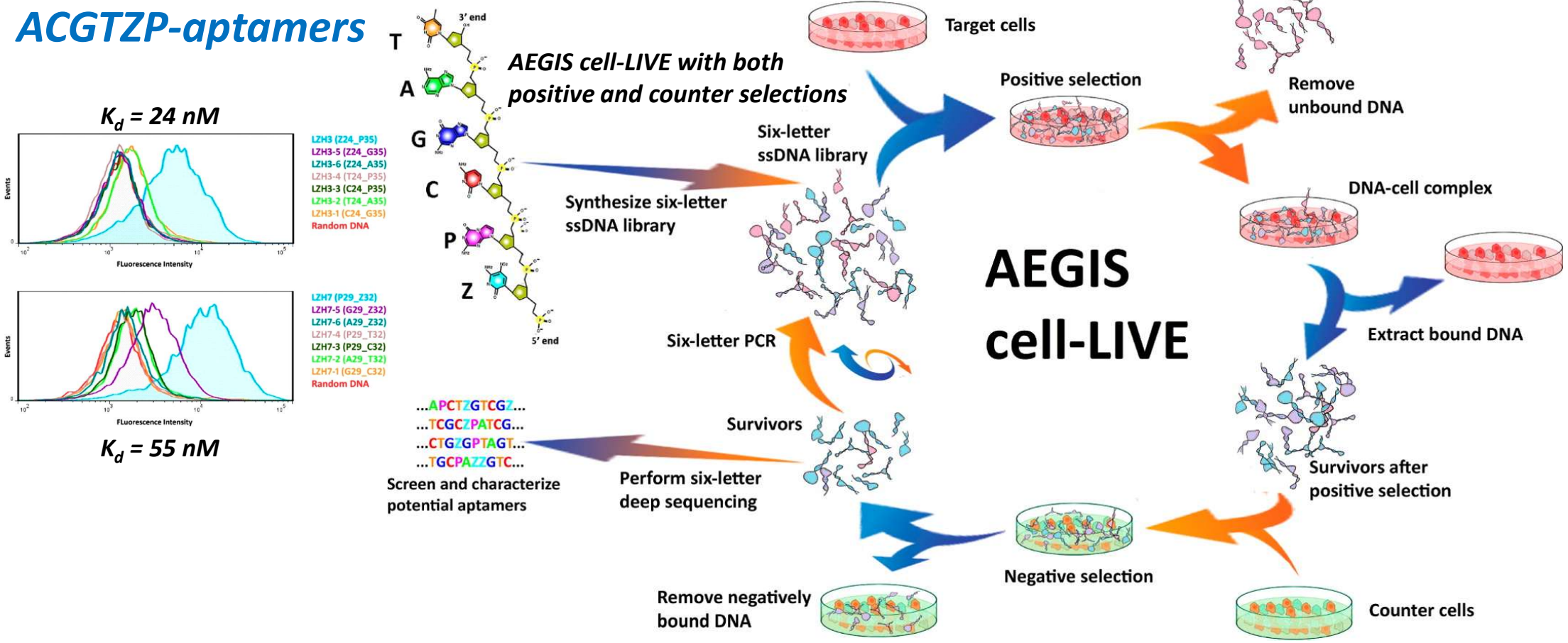


binding of chemically resynthesized ZAP-2012 aptamer obtained using AEGIS Cell–SELEX to MDA-MB-231 breast cancer cells

Where the **Z** or **P** are separately replaced with standard nucleotides, binding affinity is reduced. Where the **Z** and **P** are both replaced with standard nucleotides, binding affinity is lost (not shown).

K. Sefah *et al., Proc. Natl. Acad. Sci. USA* **2014,** *111 (4),* 1449-1454.

ACGTZP-aptamers



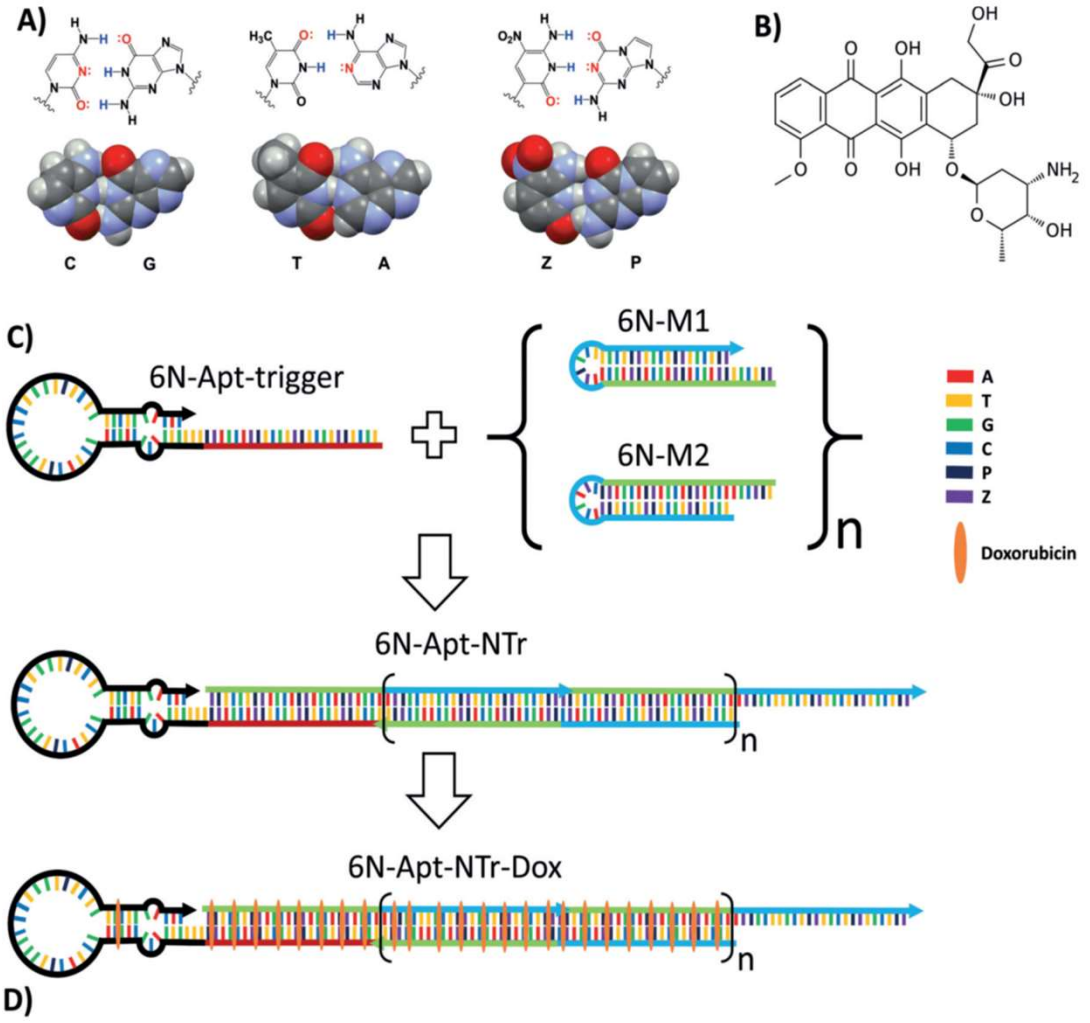
Nucleotides **Z** and **P** were added to a library of oligonucleotides used in a laboratory *in vitro* evolution (LIVE) experiment; the GACTZP library was challenged to deliver molecules that bind selectively to **liver cancer cells**, but not to untransformed **liver cells**. Unlike in classical *in vitro* selection, low levels of mutation allow this system to evolve to create binding molecules not necessarily present in the original library. Over a dozen binding species were recovered. The best had multiple Z and/or P in their sequences.

S. Benner *et al., J. Am. Chem. Soc.*, **2015**, *137,* pp 6734-6737

Aptamer-Nanotrain

The aptamer-nanotrain assembly, charged with doxorubicin, selectively kills liver cancer cells in culture, as the selectivity of the aptamer binding directs doxorubicin into the aptamer-targeted cells. The assembly does not kill untransformed cells that the aptamer does not bind.

The previously selected 6-letter aptamer which binds liver cancer cells

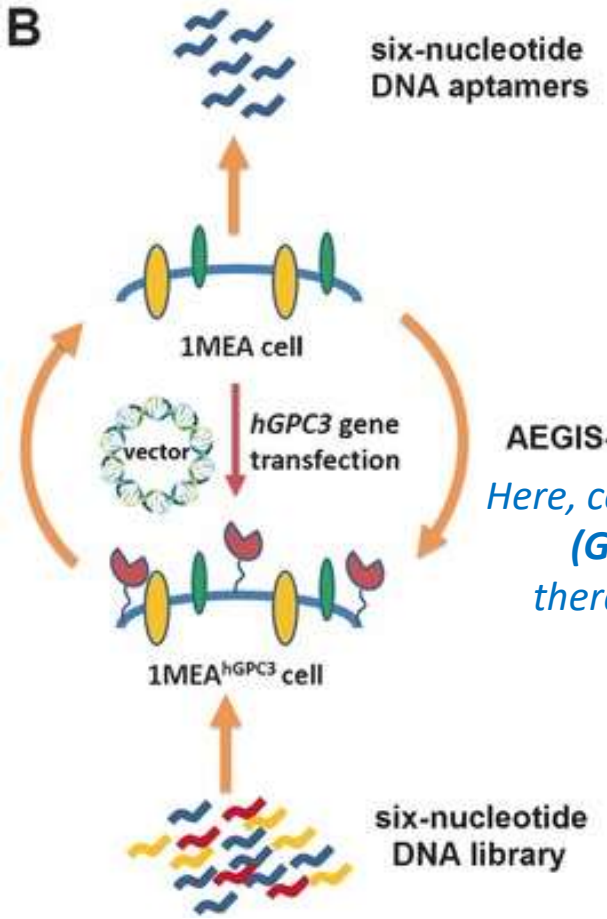


6N-Apt-trigger	5′-TGC TPZ TGC TPZ TGC TPZ ACG AZP TTT GTG ACG CAG CAG CTA CPT GGG CCC TGG TPT CTG TGC TGG ACA C-3′
6N-M1	5′-ZPT CGT PZA GCA PZA GCA PZA GCA AZP GCT TGC TPZ TGC TPZ TGC TPZ-3′
6N-M2	5′-TGC TPZ TGC TPZ TGC TPZ ACG AZP PZA GCA PZA GCA PZA GCA AGC ZPT-3′

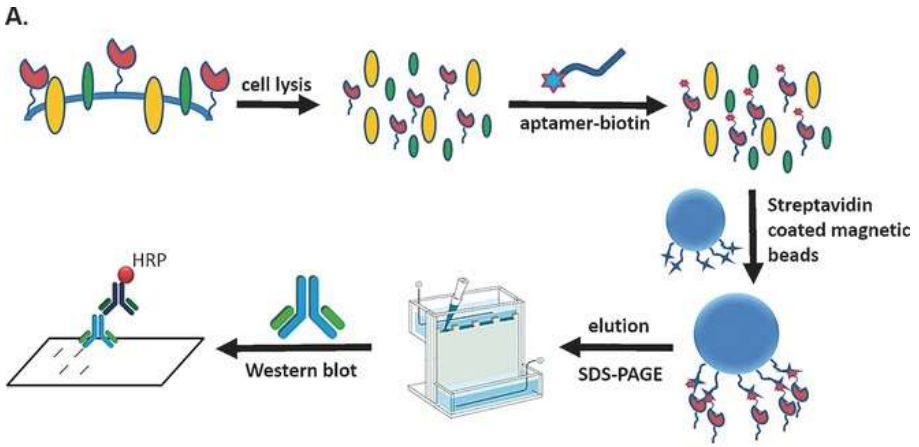
This architecture, built with an expanded genetic alphabet, is reminiscent of antibodies conjugated to drugs, which presumably act by this mechanism as well, but with the antibody replaced by an aptamer.

L. Zhang *et al., Angew. Chem., Int.Ed.*, **2020**, *59*, 663-668

ACGTZP-aptamers against a specific protein

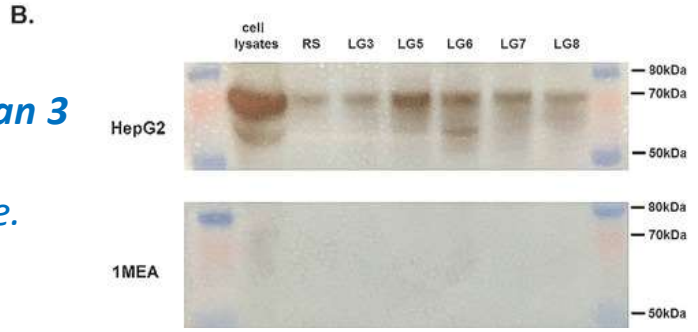


Laboratory in vitro evolution (LIVE) might deliver DNA aptamers that bind proteins expressed on the surface of cells.



AEGIS-LIVE

*Here, cell engineering was used to place **glypican 3 (GPC3),** a possible marker for liver cancer theranostics, on the surface of a liver cell line.*



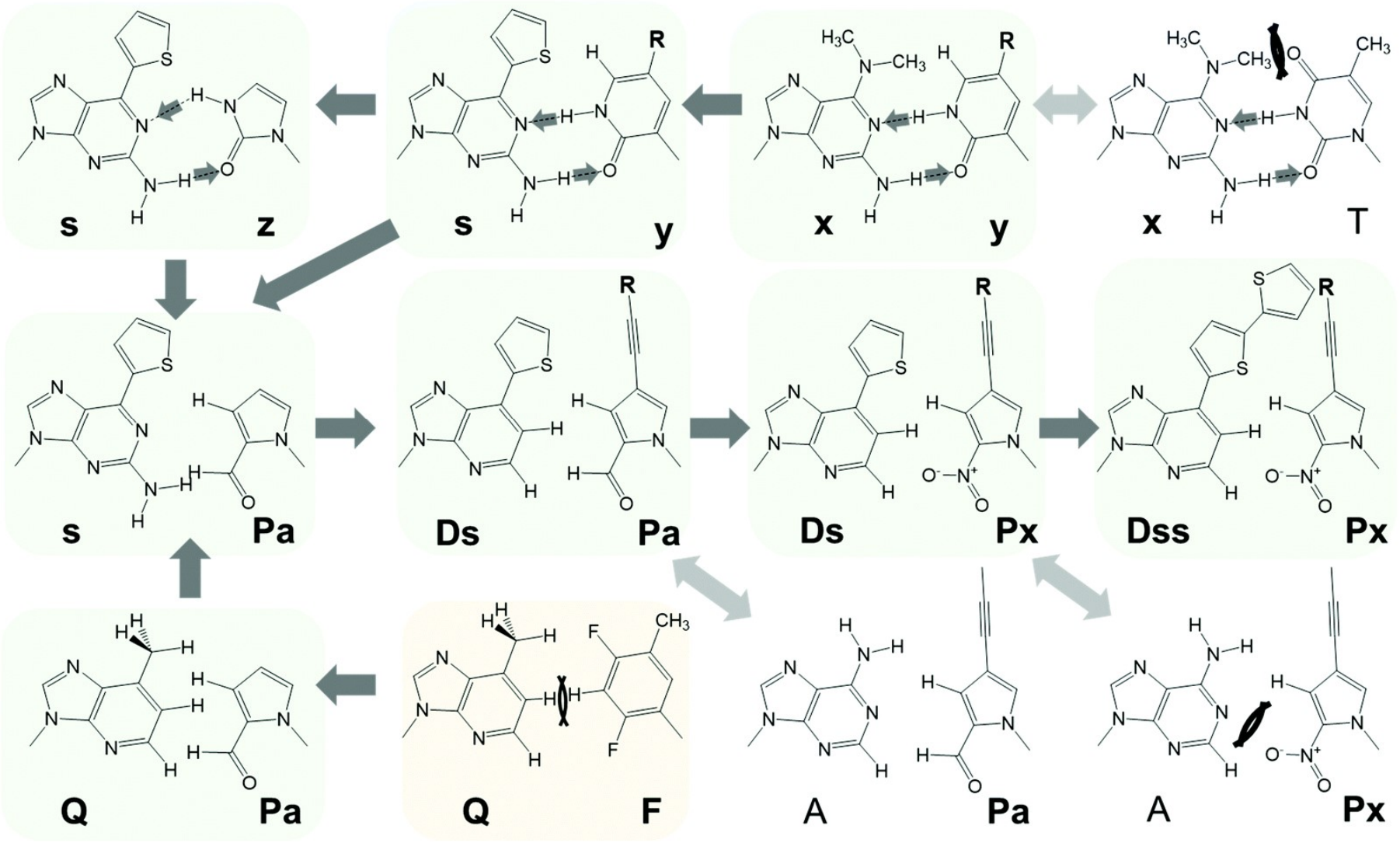
Libraries were then built from a six-letter genetic alphabet. With counterselection against non-engineered cells, eight AEGIS-containing aptamers were recovered. Five bound selectively to GPC3-overexpressing cells.

Name	Sequence	Percentage
LG1	~PGGTGGGCGGAGGTCTZGCTACAPGPTTTGGPGGC~	11.37%
LG2	~PGCCCGGGPTAPPGTGPTGGGTGTTCGCTATCCAG~	7.98%
LG3	~GGTAACTAGTAGTTGACCCTGPAGTGZTGTPTCTG~	6.01%
LG4	~GGCGGGGTZGPGTAAGGGGTCTAAGGCATTGGGTC~	4.48%
LG5	~GGAGGAAGTGGTCCTTGCTTTGCZTCGTATCTGGG~	2.57%
LG6	~GGTZGATTATTPGGTTCAATAACACPTCCTGGTGG~	1.96%
LG7	~PGCACAGTGTGZZCCATAGGTTGTAATGACPTZTG~	1.04%
LG8	~GGCAGCZCCTGPAGTPGAGTGTPATGGCTTATTCG~	0.91%

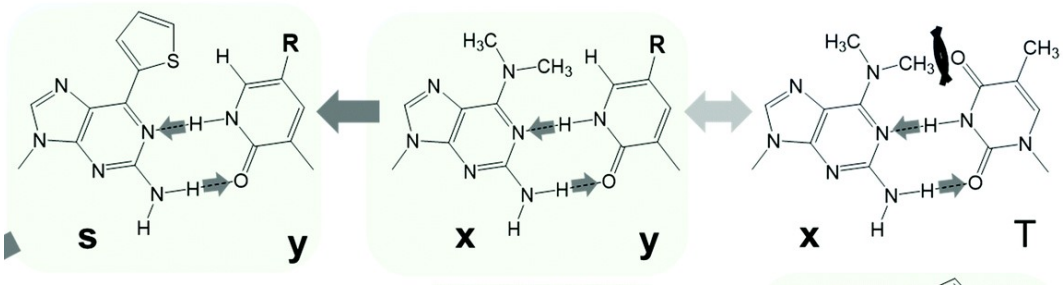
LG5 K_d = 6 nm (without Z – no binding)

L. Zhang *et al., Angew. Chem., Int.Ed.,* **2016**, *55,* 12372-12375

Steric exclusion and hydrophobic non-natural base pairs



Steric exclusion and hydrophobic non-natural base pairs



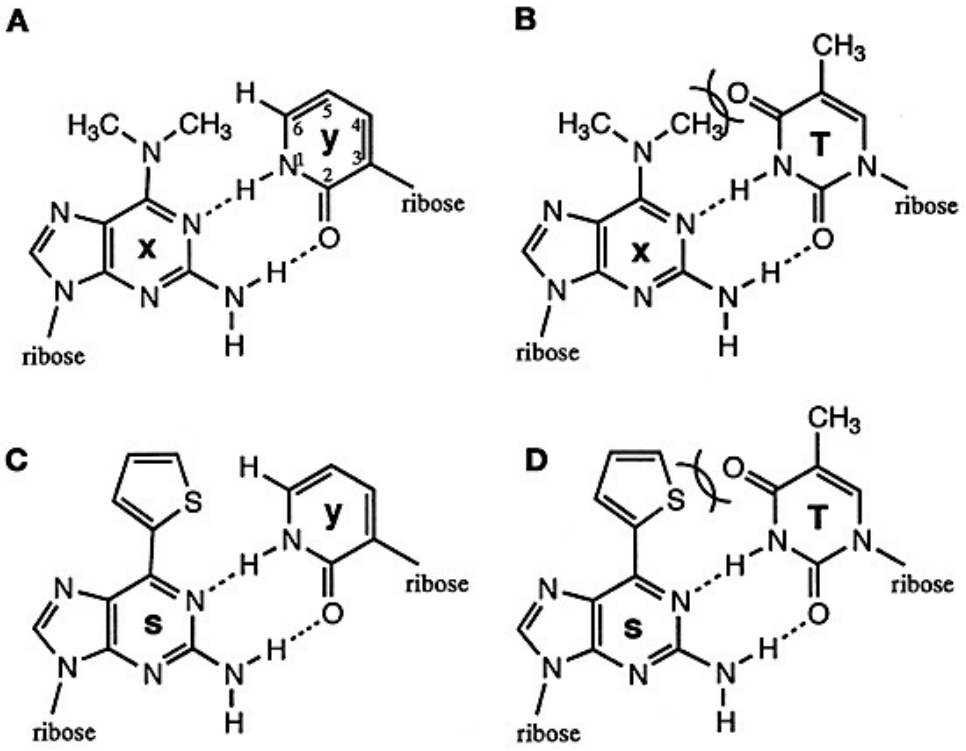
Hirao (2001): the steric hindrance concept to hydrogen-bonded UBPs to exclude the mispairing with natural bases → a series of hydrogen-bonded UBPs: **x–y** and **s–y** pairs

large residues at position 6 of x and s sterically and/or electrostatically clash with the 4-keto group of T, but not with hydrogen of the y base.

x–y and **s–y** pairs function in transcription with T7 RNA polymerase (T7 transcription), and the **y** substrate is incorporated (>96% selectivity **s–y**) site-specifically into RNA transcripts opposite **x** or **s** in DNA templates.

Problem: the **y** base cannot exclude the mispairing with A. Therefore, these UBPs cannot be used in replication

Unnatural aminoacid incorporation using a noncanonical base pair



Non-template strand 5'-d(ATAATACGACTCACTATAGGG)
Template strand 3'-d(TATTATGCTGAGTGATATCCCTTGCGCTCNTCACGCC)
+1 +12 +19
N = **s**, **x**, A, or G

↓ T7 RNA polymerase
1 mM N'TPs, [α-^{32}P]ATP, GMP

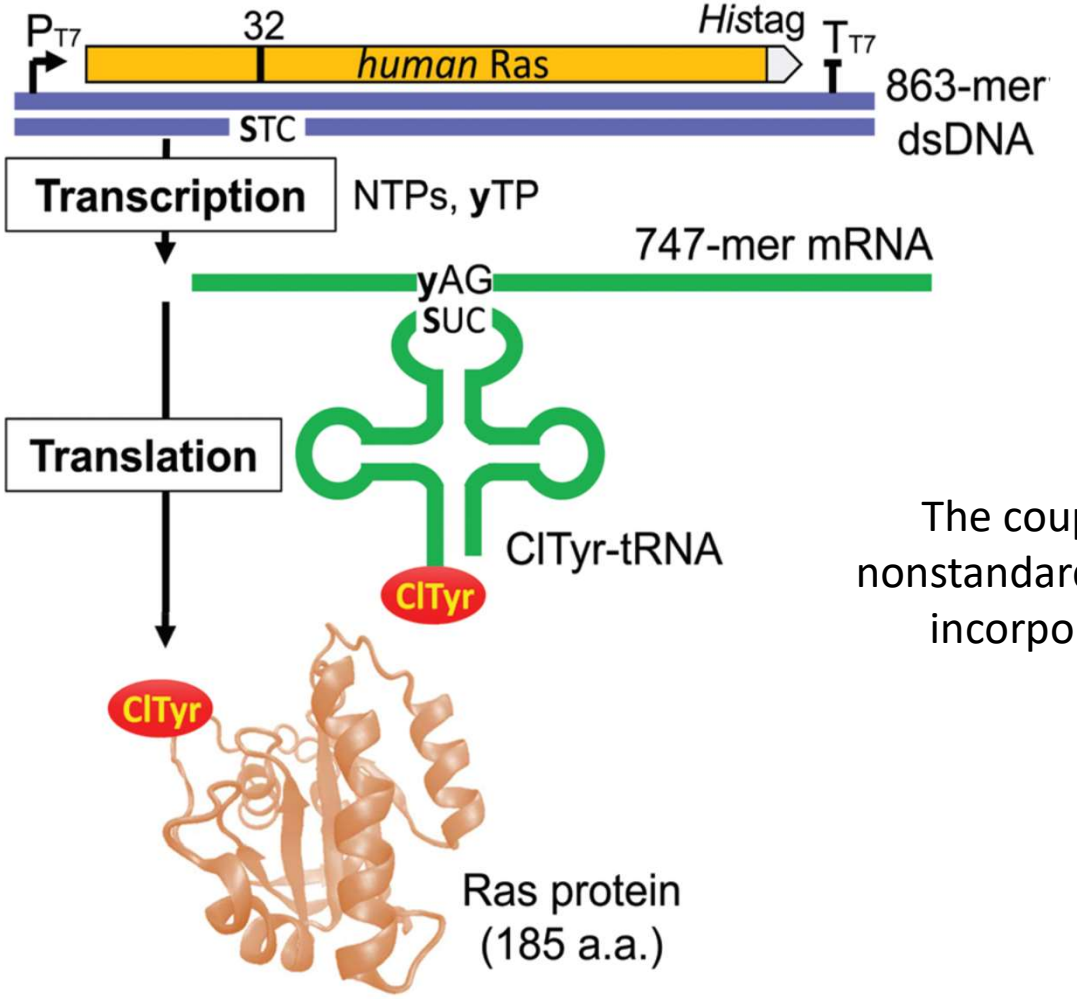
Full-length product (19-mer)
pGpGpGp*Ap*ApCpGpCpGp*ApGpN'p*ApGpTpGpCpGpG

↓ RNase T_2 digestion
2D-TLC

^{32}P-labeled nucleotides
Gp* x 2, Ap* x 1, **N'**p* x 1

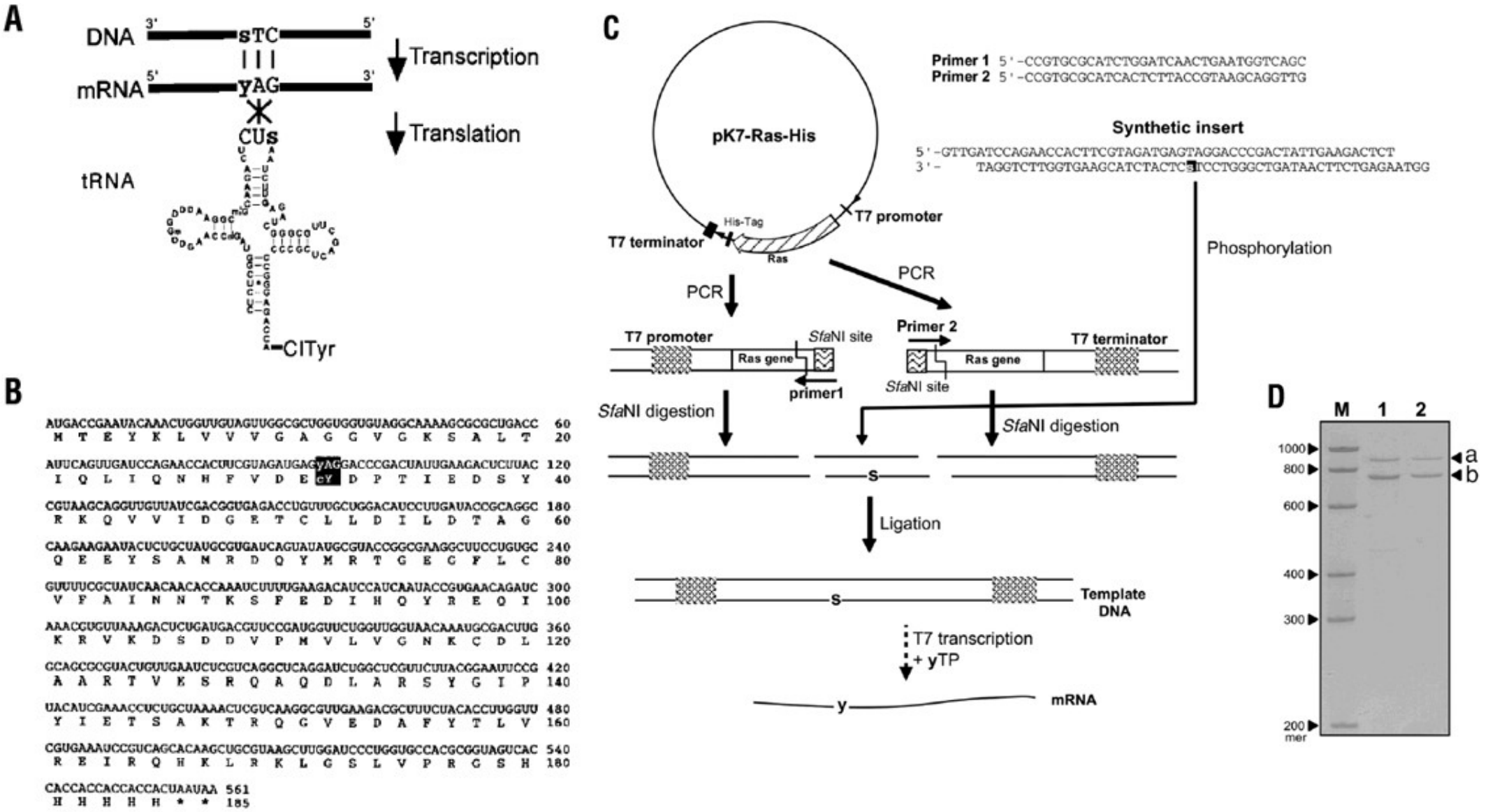
I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)

Unnatural aminoacid incorporation using a noncanonical base pair



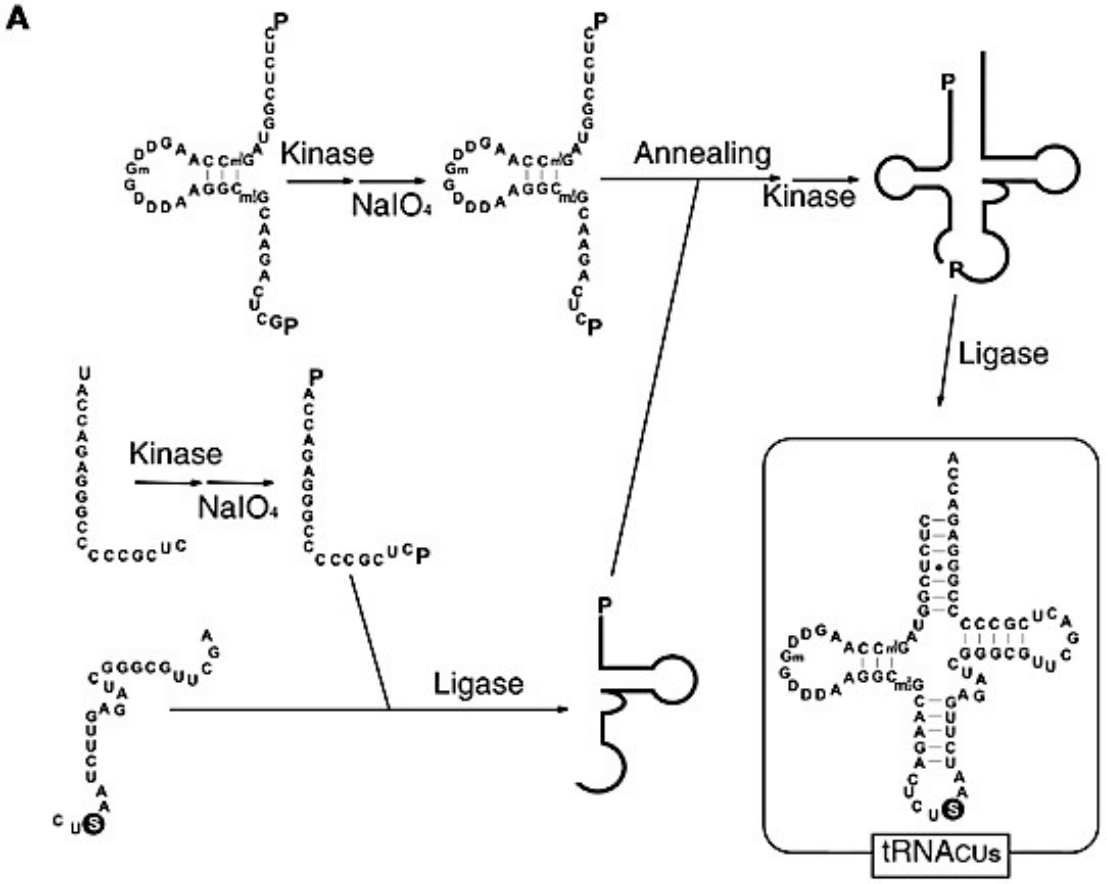
The coupled transcription–translation system using the nonstandard codon–anticodon interaction for the site-specific incorporation of 3-chlorotyrosine into the Ras protein.

I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)



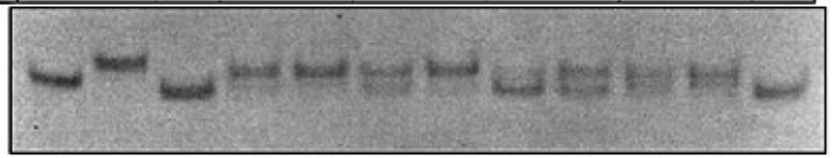
I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)

Unnatural aminoacid incorporation using a noncanonical base pair



B

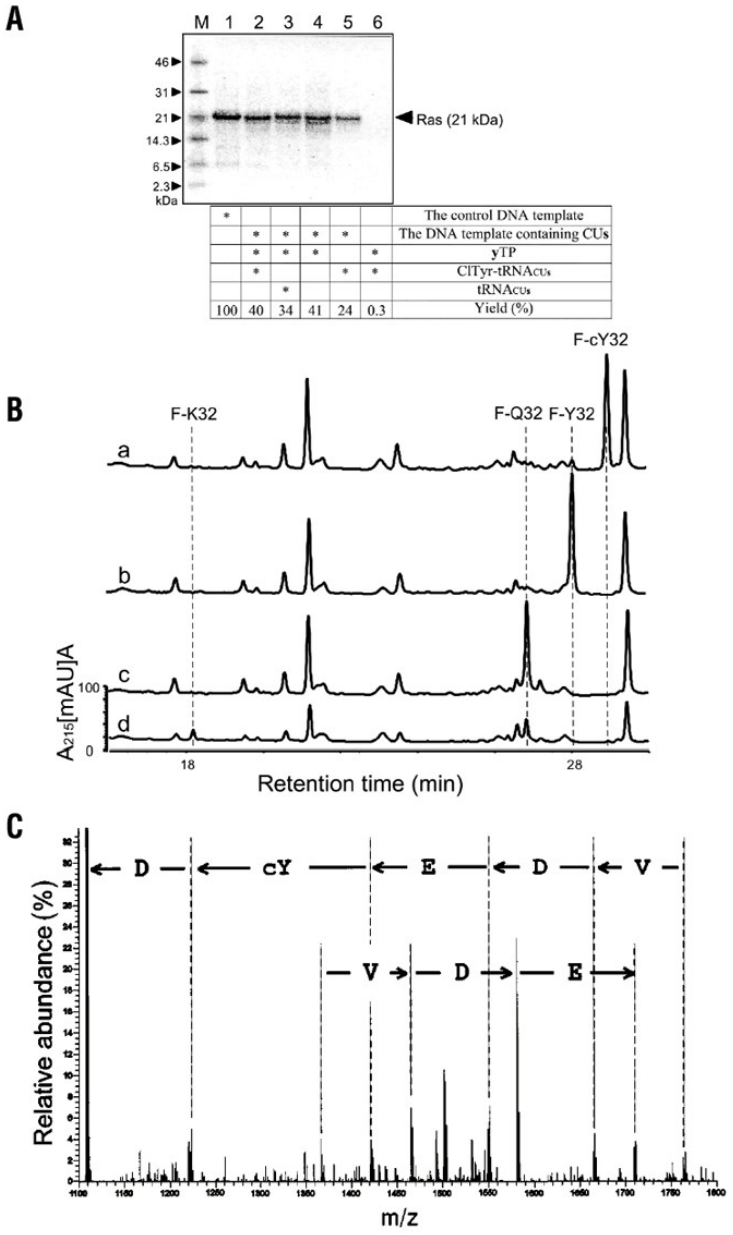
tRNA	Native tRNA		tRNA$_{CUs}$									
Amino acid	–	Tyr	–	FTyr		ClTyr		ITyr		DOPA		–
DMSO/Tween-20	–	–	+	–	+	–	+	–	+	–	+	+
Yeast TyrRS	–	+	–	+	+	+	+	+	+	+	+	–



1 2 3 4 5 6 7 8 9 10 11 12

(A) Construction of tRNA$_{CUs}$. (B) Acidic-gel electrophoresis of the products after aminoacylation of *S. cerevisiae* tRNA and tRNA$_{CUs}$ with tyrosine and the 3′-substituted analogs, in the absence or presence of 20% DMSO and 0.25% Tween-20. The upper bands are the aminoacylated tRNAs, and the lower bands are the noncharged tRNAs.

I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)



A) Autoradiogram of *in vitro* transcription–translation products labeled with l-[^{14}C]leucine. The conditions are noted at the bottom of each lane.

B) LC patterns of the products digested by Lys-C. Chart a shows the products obtained from the reaction in the presence of the DNA template containing the CT**s** sequence, **y**TP, and ClTyr-tRNA$_{CUs}$; chart b shows those obtained in the presence of the standard DNA template and Tyr; chart c shows those obtained in the presence of the DNA template containing CT**s** but in the absence of **y**TP and tRNA$_{CUs}$; and chart d shows those obtained in the presence of the DNA template containing CT**s** and **y**TP but in the absence of tRNA$_{CUs}$.

C) Tandem mass spectrum of the F-cY32 fragment. The partial sequence, VDE**cY**D, of F-cY32 was confirmed from the ion series

I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)

Unnatural nucleobases - overview