Synthetic life

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

NaturalNews.com

WiSe 2020/21

Zbigniew Pianowski

Overview of the course

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

unnatural base pairing – expansion of the genetic alphabet;

Artificial genetic polymers and oligonucleotide analogues (XNA);

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

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

Artificial lipid vesicles as models for protocell multiplication;

design of artificial organisms

CHAPTER 1

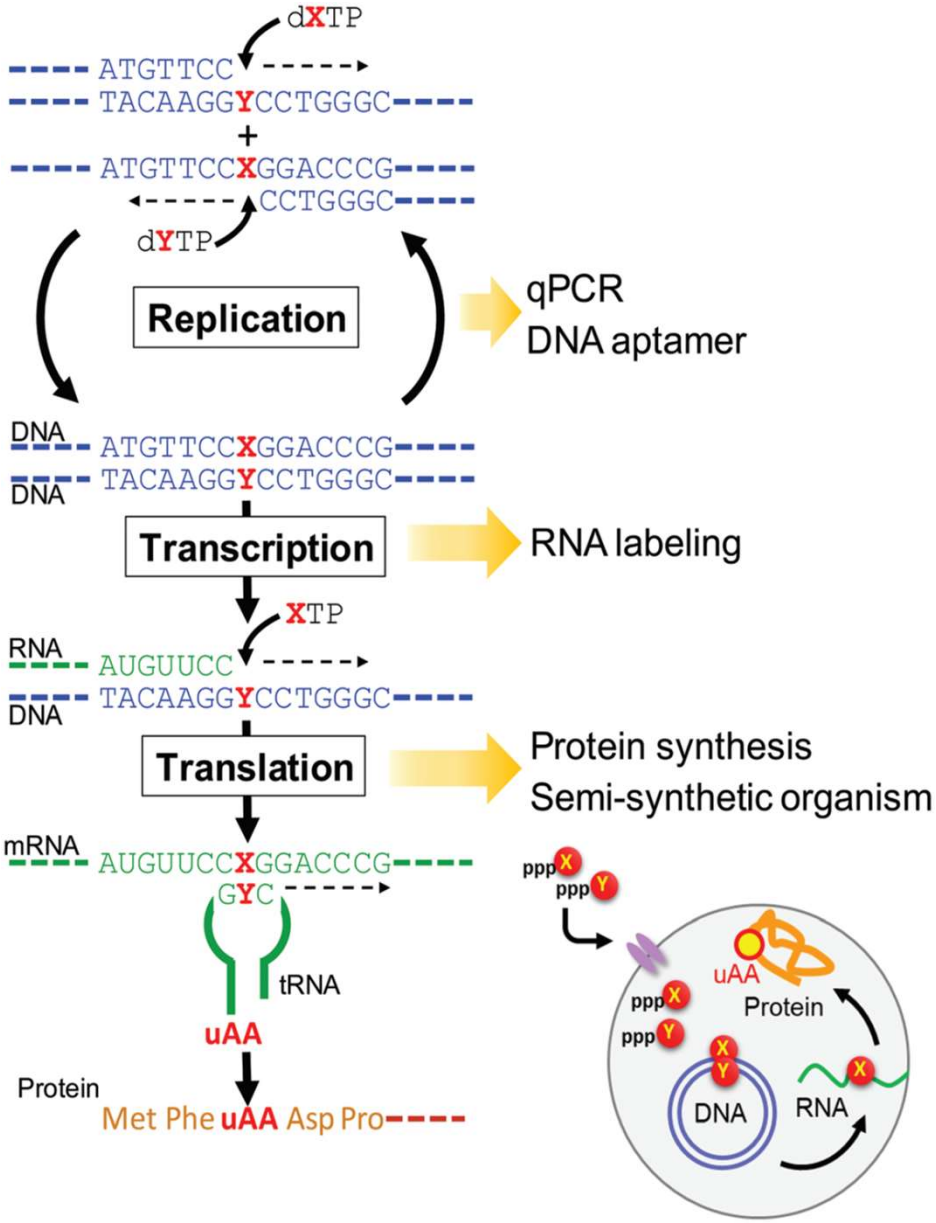
OLIGONUCLEOTIDES

Part 2 – noncanonical nucleobases

Expansion of the genetic alphabet

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

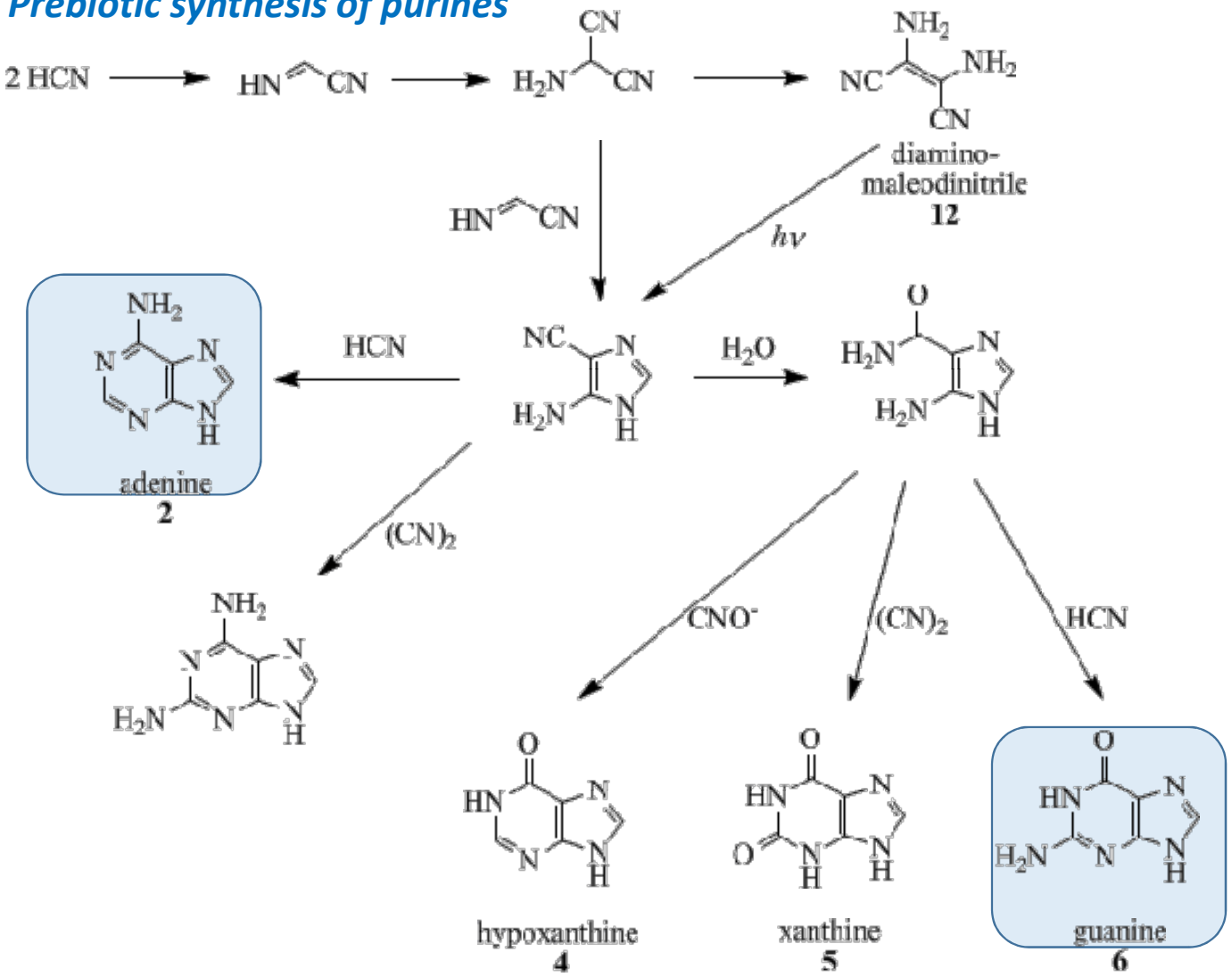
The creation of a UBP (i.e., X–Y) that functions in replication, transcription, and translation as a third base pair with the natural A–T(U) and G–C pairs allows the storage and retrieval of the expanded genetic information in vitro and in vivo, enabling a variety of applications using biopolymers with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



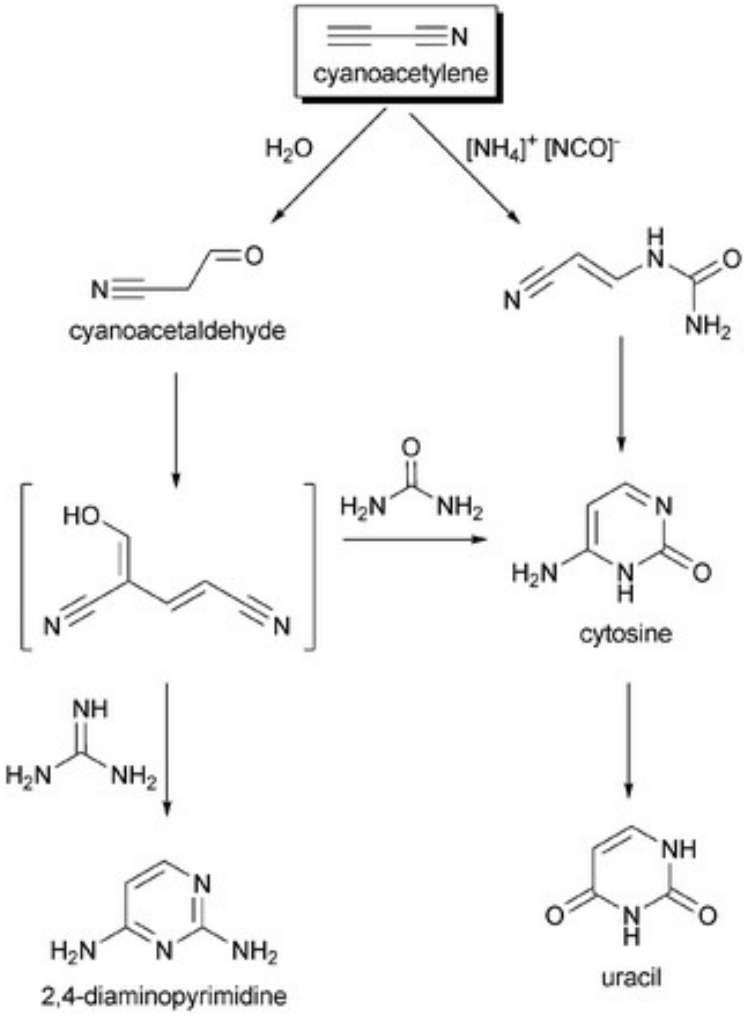
M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

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

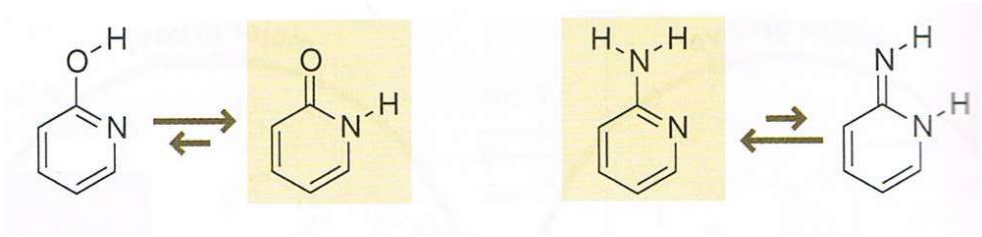
Prebiotic synthesis of purines

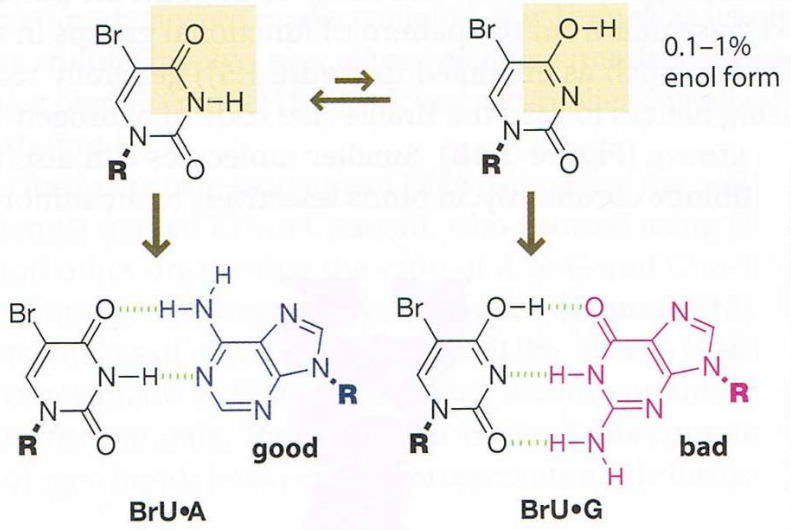


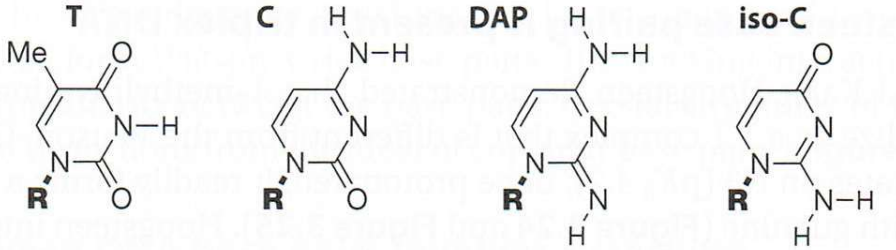
Prebiotic synthesis of pyrimidines



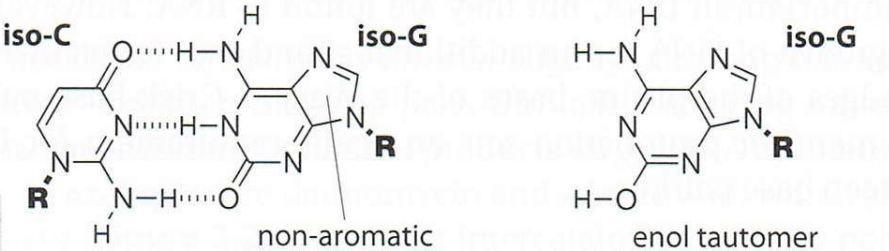
Biological consequences of nucleobase modifications







why not a third type of base pair?



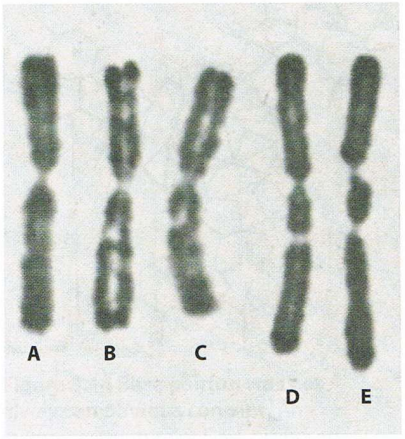
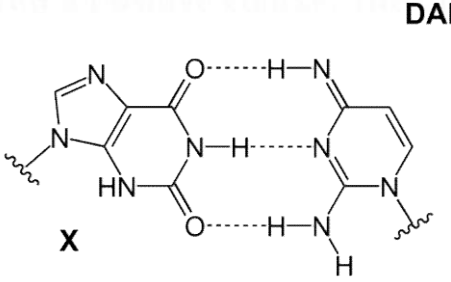
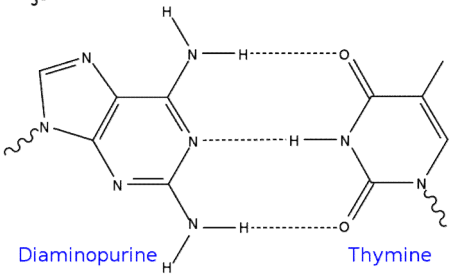


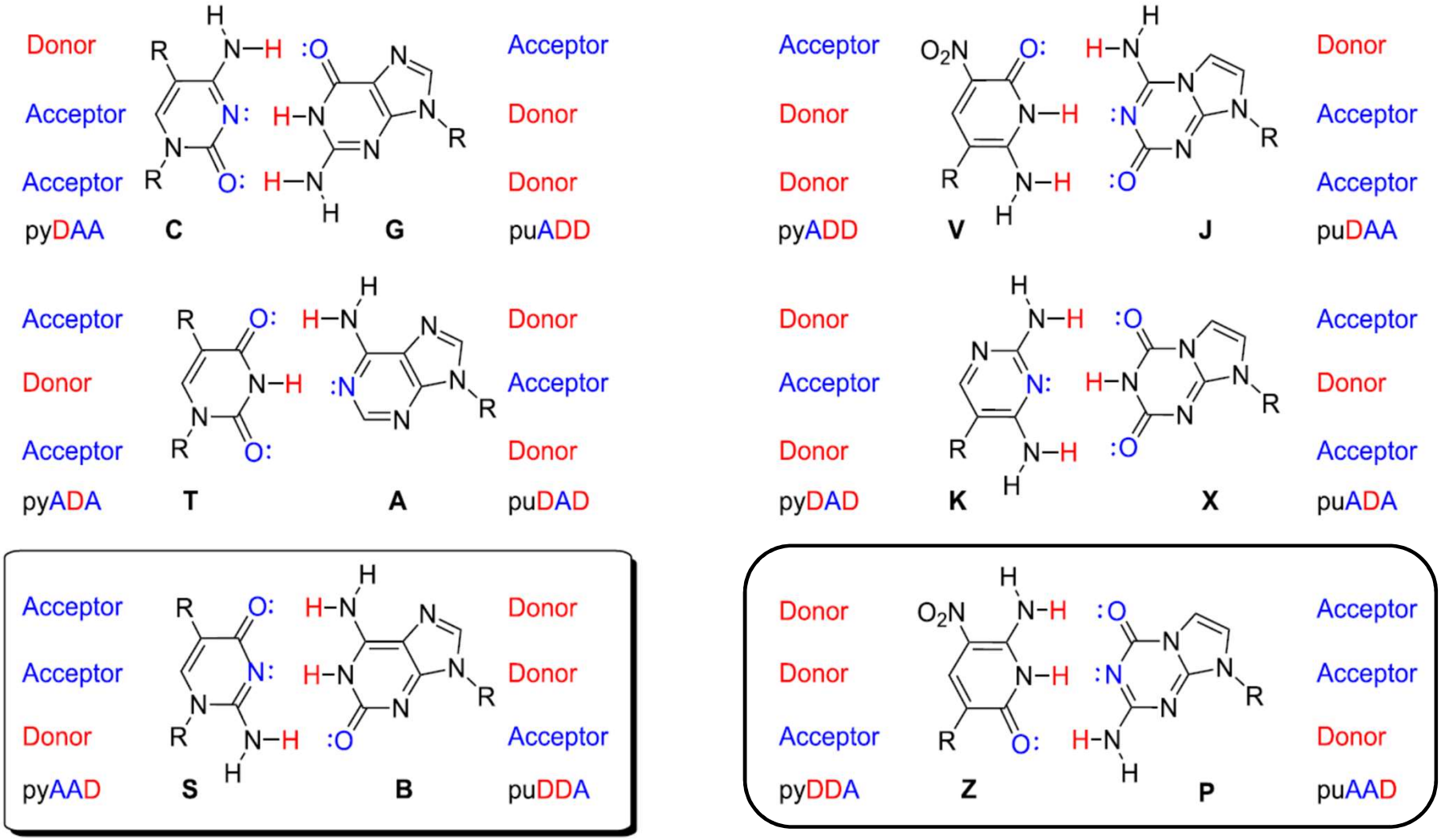
Figure 3.20 Chromosome 1 from hamster cells exposed to bromodeoxyuridine. (A) Normal chromosome. (B–E) Aberrant chromosomes. (From T.C. Hsu and C.E. Somers, *Proc. Natl. Acad. Sci. USA* 47: 396–403, 1961. With permission from the MD Anderson Cancer Center.)



Cyanophage S-2L:



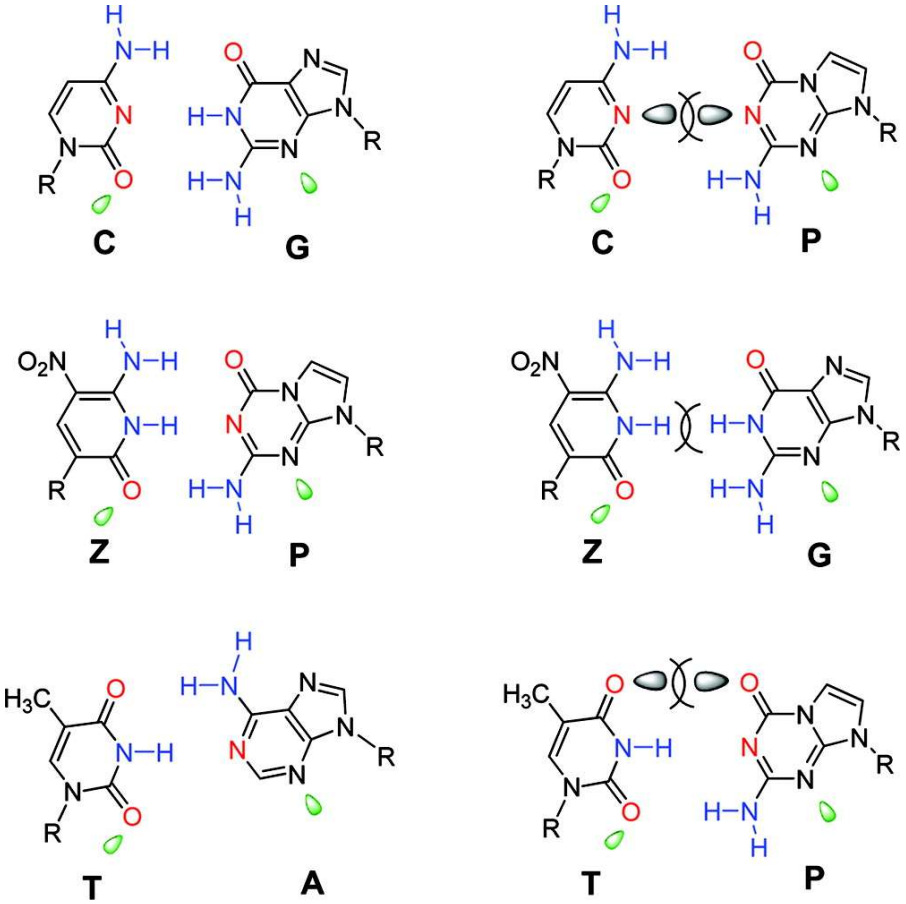
AEGIS – Artificially Expanded Genetic Information System



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

AEGIS – Artificially Expanded Genetic Information System

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



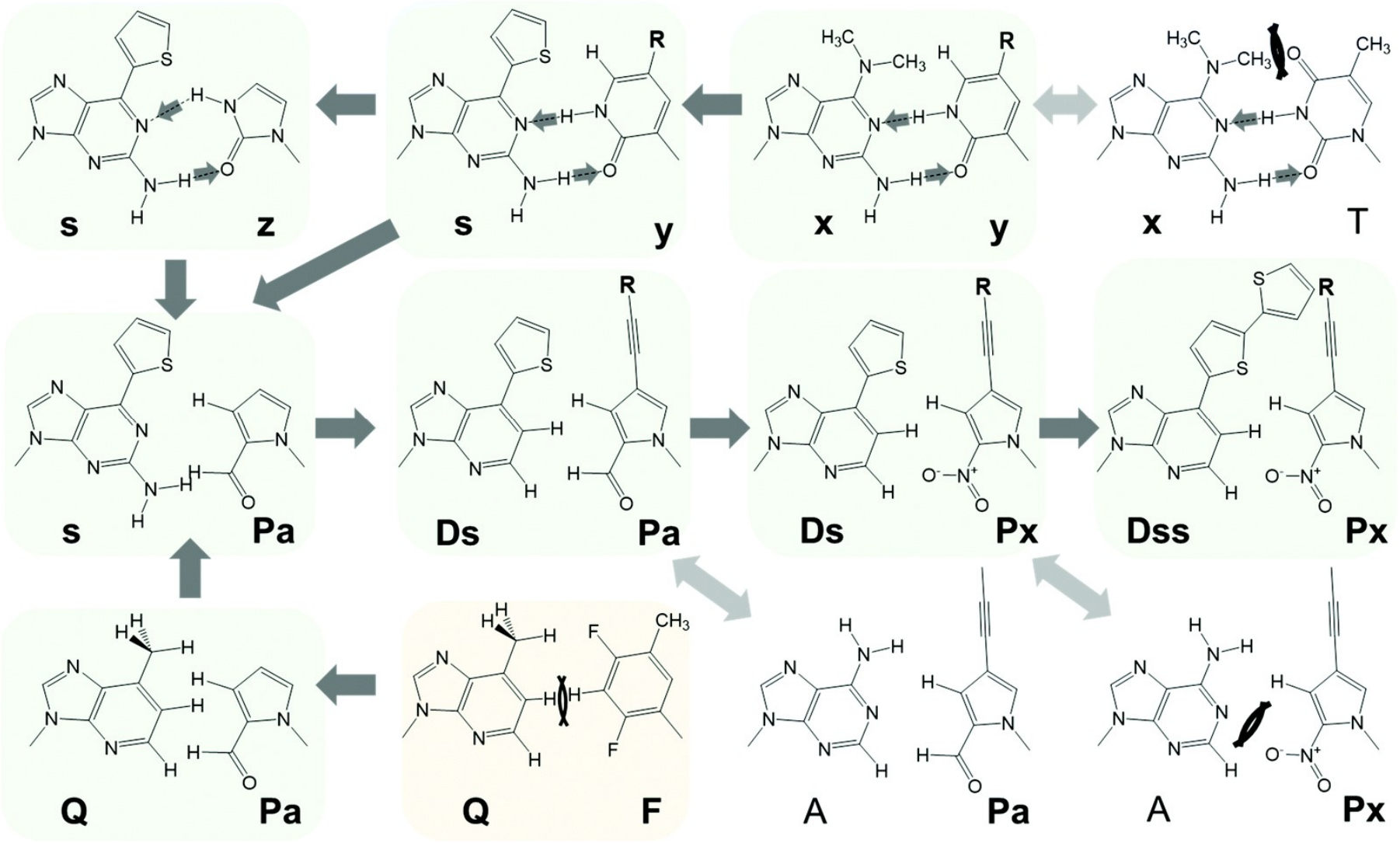
Applications:

- **artificial gene synthesis**
- **Hachimoji DNA/RNA ACGTPZSB genetic system (transcription of spinach RNA aptamer)**
- **AEGIS Cell-SELEX:**
  - **Aptamers against breast cancer cells**
  - **Positive/negative selection cancer/healthy liver cells**
  - **Aptamer + doxorubicin nanotrain**
  - **Aptamers against protein (cancer marker)**

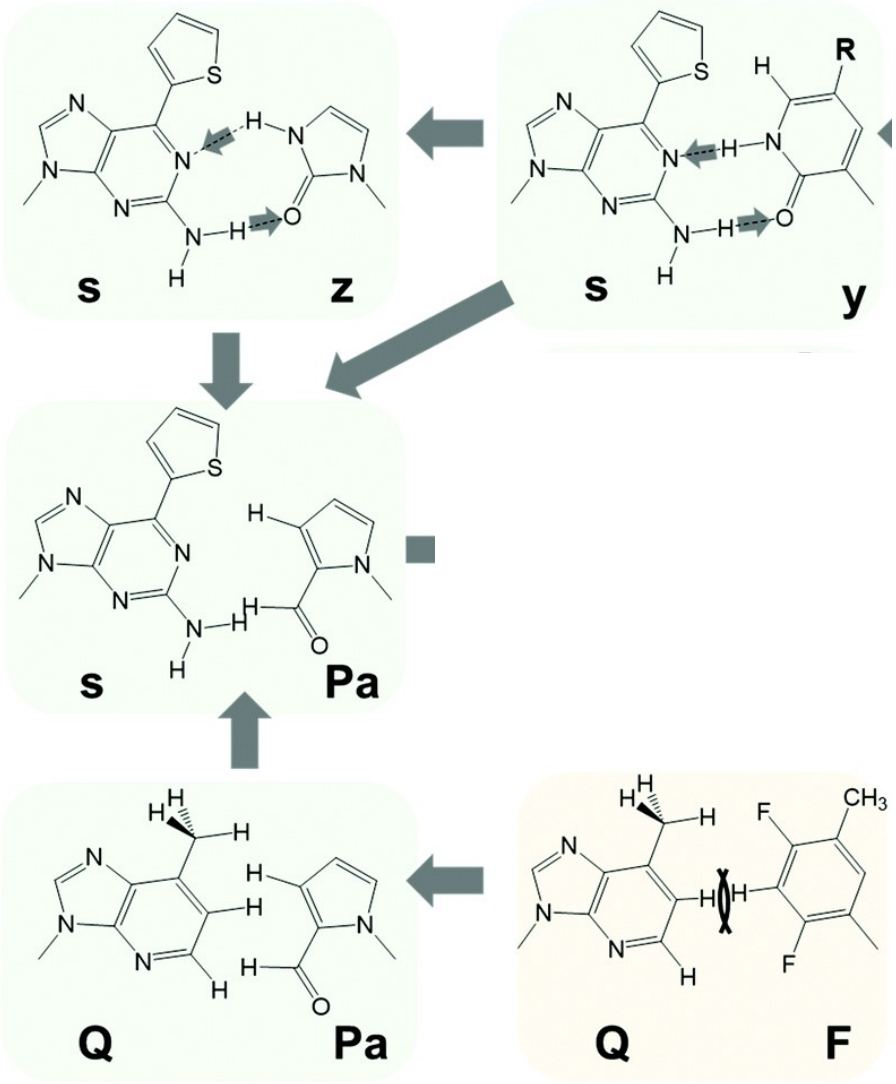
Error rate 0,2% per a PCR cycle – both removal and incorporation of **Z** and **P** ➔ the artificial genetic system capable to evolve.
T7 RNA Polymerase mutant – **ZPSB** transcription

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

Steric exclusion and hydrophobic non-natural base pairs



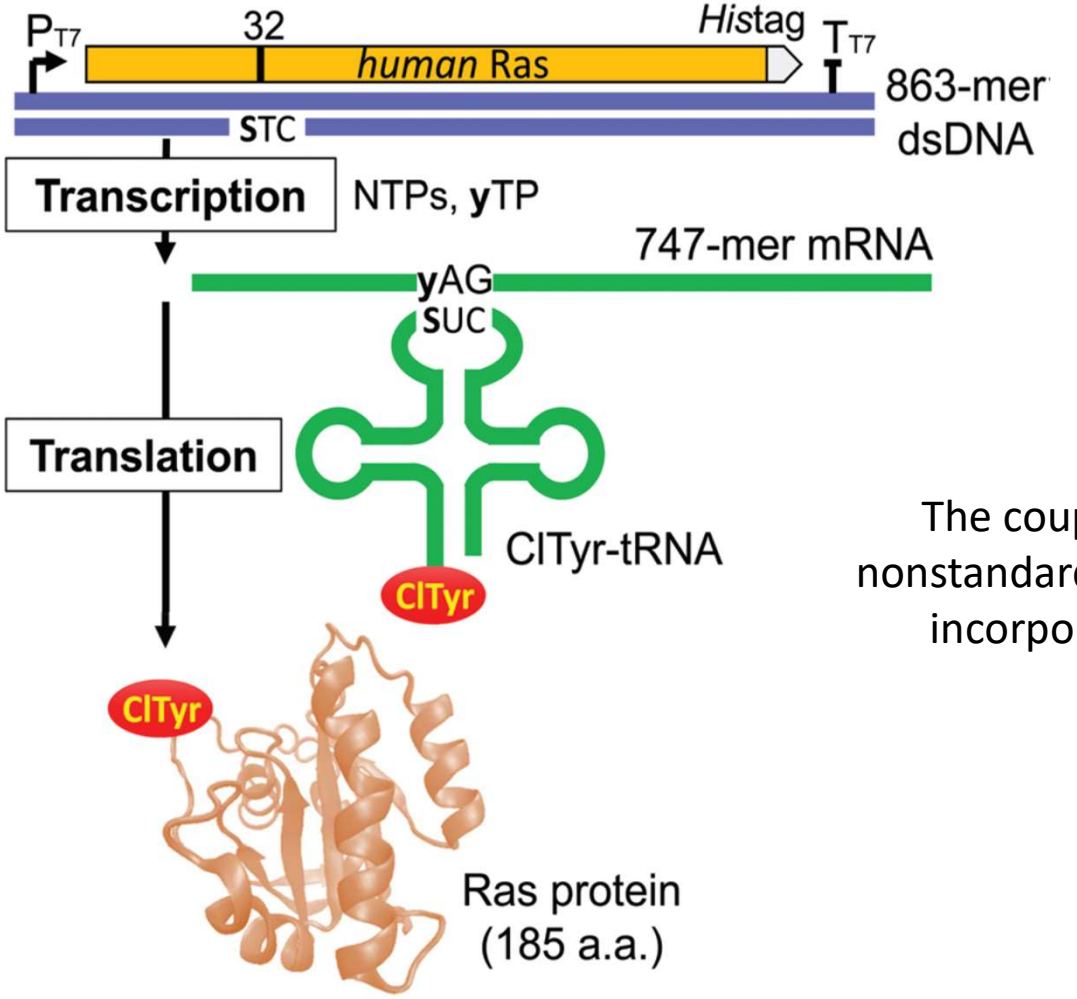
Steric exclusion and hydrophobic non-natural base pairs



Kool (1998): non-hydrogen-bonded **Z–F** pair (an isostere of A–T) (Z is 4-methylindole, F base is 2,4-difluorotoluene.), but lack of Z and F interactions with polymerases.
Kool (1999): Q-F pair - shape-complementarity rather than the hydrogen-bond interactions, ➔a new strategy of UBP development using hydrophobic UBs without any hydrogen-bond interactions between the pairing bases.

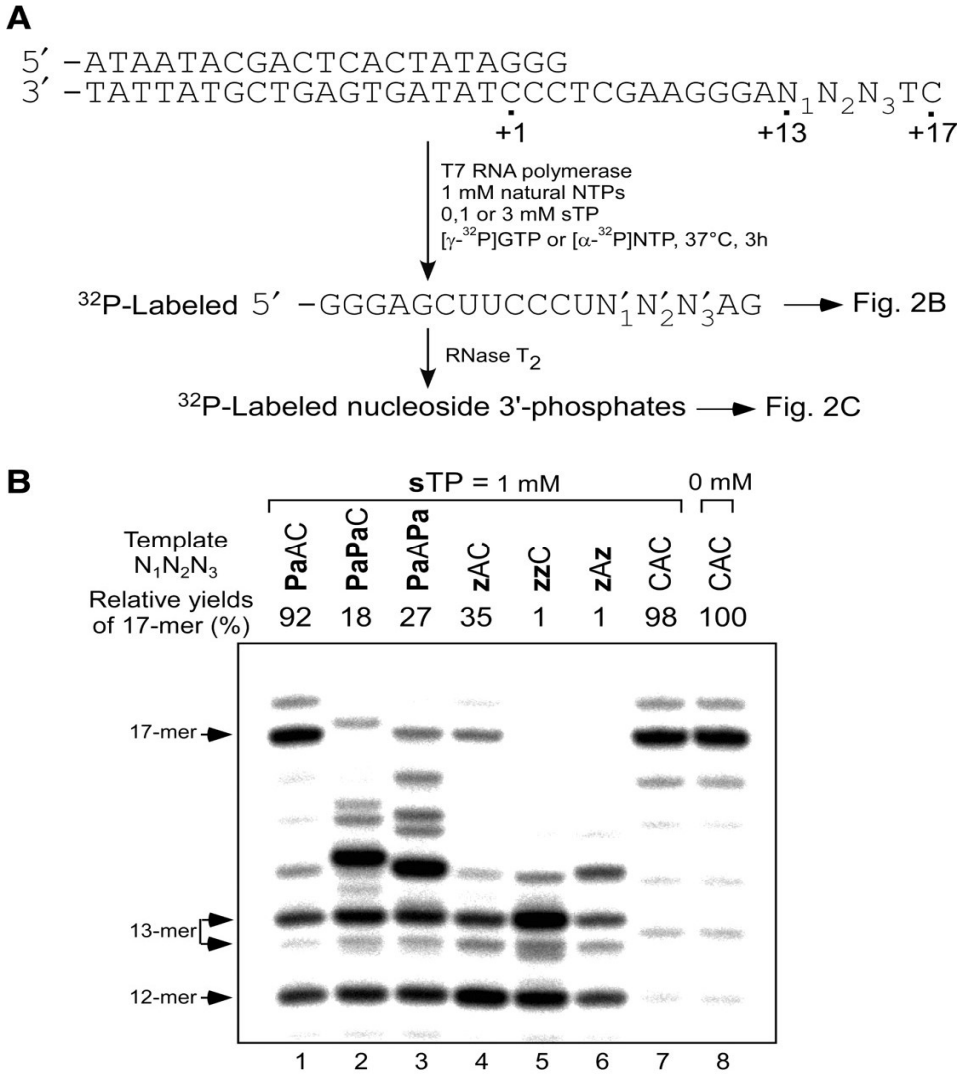
Hirao (2003): the Q–Pa pair to avoid steric clashes.
Hirao (2007): Pa could be used as a pairing partner of the s base, and the s substrate was incorporated specifically into RNA opposite Pa in templates by T7 transcription.
Hirao (2004): The s–z pair increased the s incorporation selectivity opposite z in T7 transcription, as compared to that opposite y in the s–y pair

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)



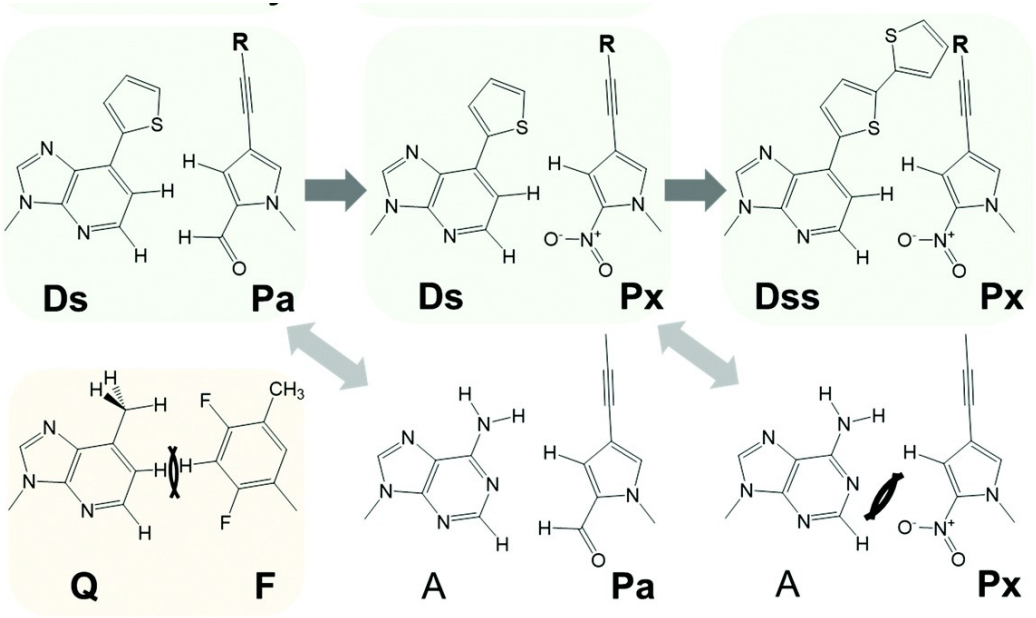
Efficient site-specific incorporation of a fluorescent base analog, 2-amino-6-(2-thienyl) purine (s), into RNA by transcription mediated by an unnatural base pair between s and pyrrole-2-carbaldehyde (Pa).

The ribonucleoside 5₀-triphosphate of s was site-specifically incorporated into RNA, by T7 RNA polymerase, opposite Pa in DNA templates.

The Pa-containing DNA templates can be amplified by PCR using 7-(2-thienyl)imidazo[4,5-b]pyridine (Ds), another pairing partner of Pa.

M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama and I. Hirao, *Nucleic Acids Res.*, **2007**, *35*, 5360–5369

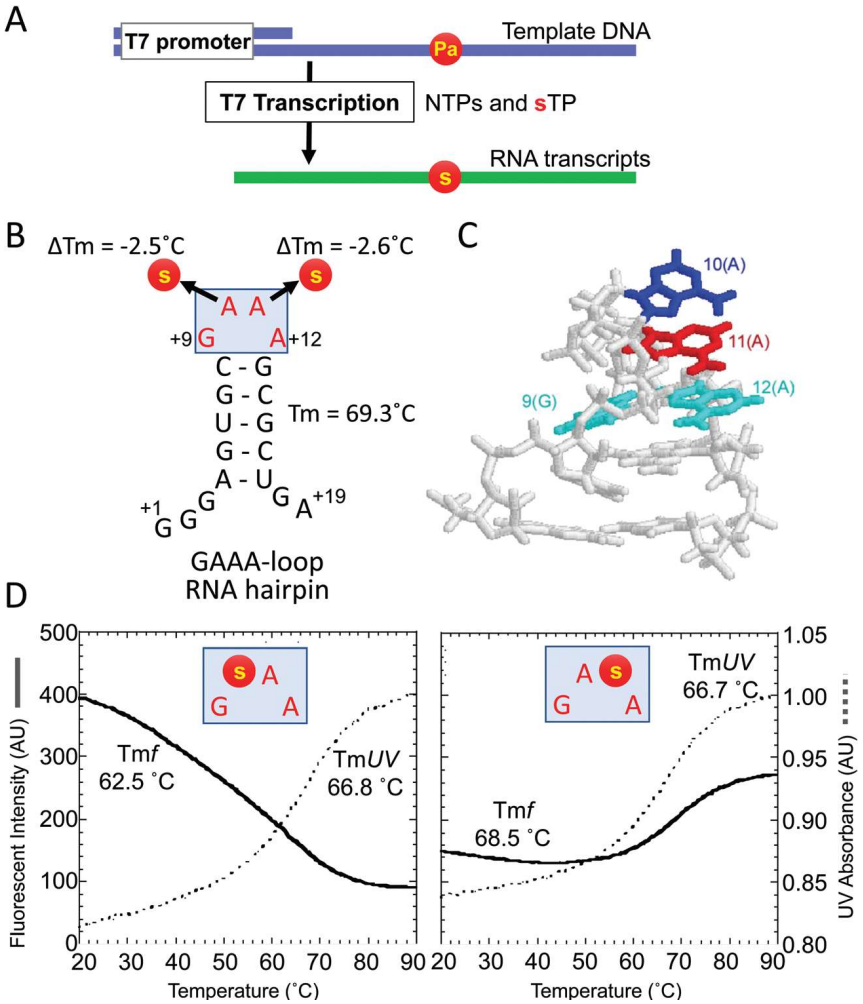
Steric exclusion and hydrophobic non-natural base pairs

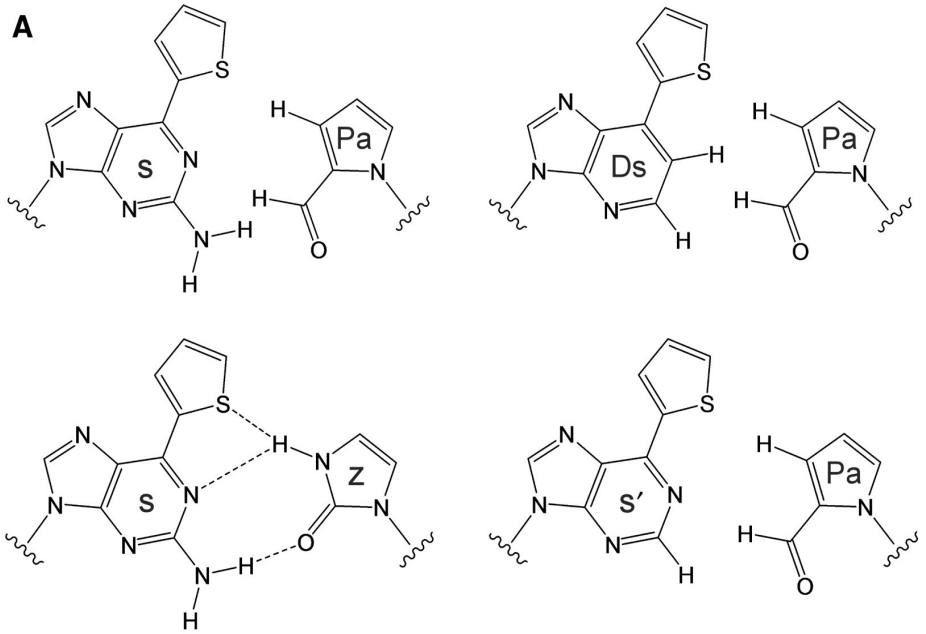


Hirao (2006) the hydrophobic **Ds** base (improved **s**). The **Ds–Pa** pair exhibited high incorporation efficiency in replication. **Problems**: the non-cognate Ds–Ds and A–Pa pairings (solved by using γ-amido-triphosphates of Ds and A, which significantly reduced their mispairing with Ds and Pa - ~99.9% selectivity per cycle in PCR).

Hirao (2007): the **Ds–Px** pair - electrostatic clash with the 1-nitrogen of A (no A–Px pairing), The additional propynyl group reduces the Ds–Ds self-pairing.
The Ds–Px pair exhibits high fidelity and high efficiency in PCR (>99.9% per cycle).

Fluorescent probing for RNA molecules by an unnatural base-pair system



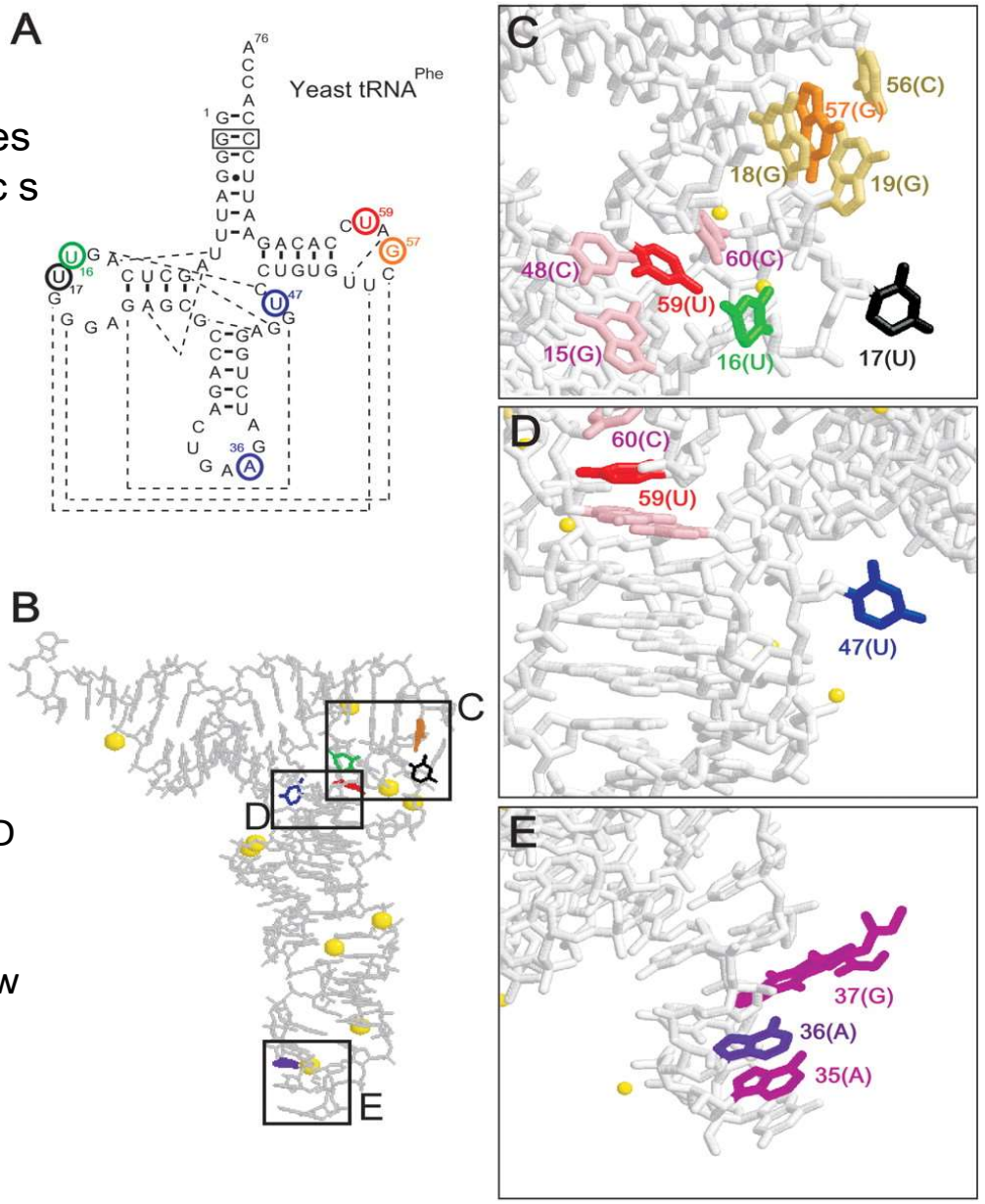


(B) Incorporation of the fluorescent **s** base into GNRA hairpins. (*left*) The secondary structure of the RNA hairpin with a GAAA loop. The second A and third A in the loop are shown in blue and red, respectively. *(right)* The 3D structure of the GAAA-loop hairpin

M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama and I. Hirao, *Nucleic Acids Res.*, **2007**, *35*, 5360–5369

The „s” incorporation sites in yeast tRNA^{Phe}

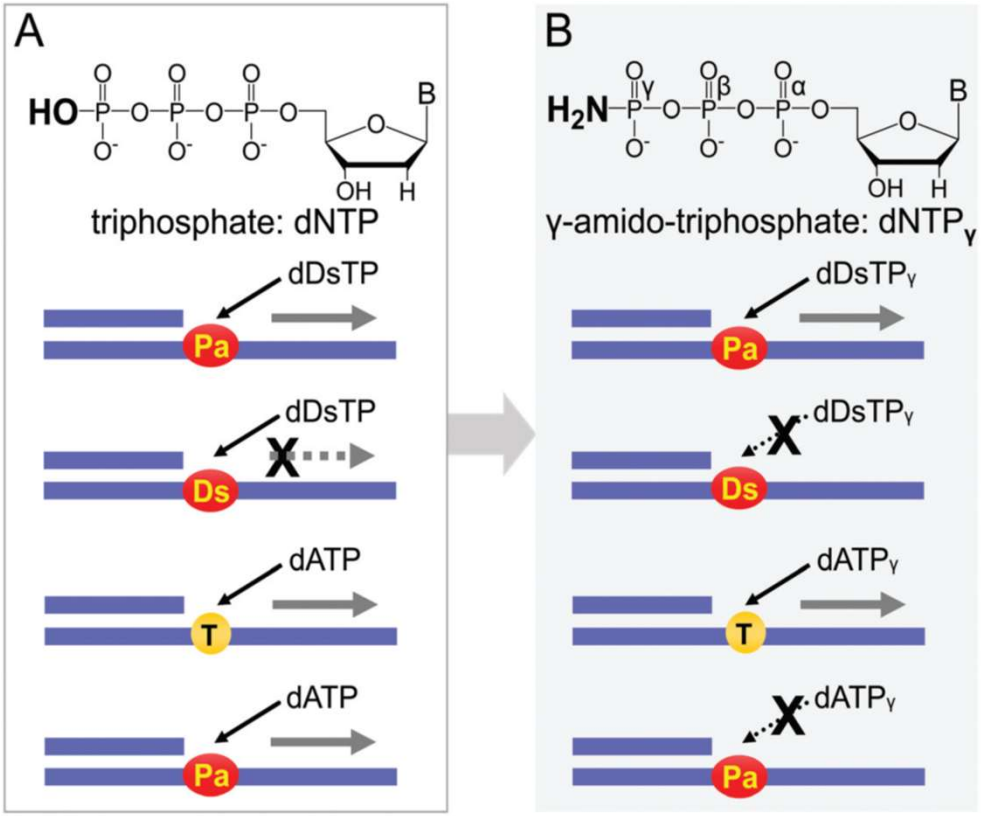
The fluorescent intensity of „s” in RNA molecules changes according to the structural environment. The site-specific s labeling of RNA hairpins and tRNA molecules provided characteristic fluorescent profiles, depending on the labeling sites, temperature and Mg_{2+} concentration.

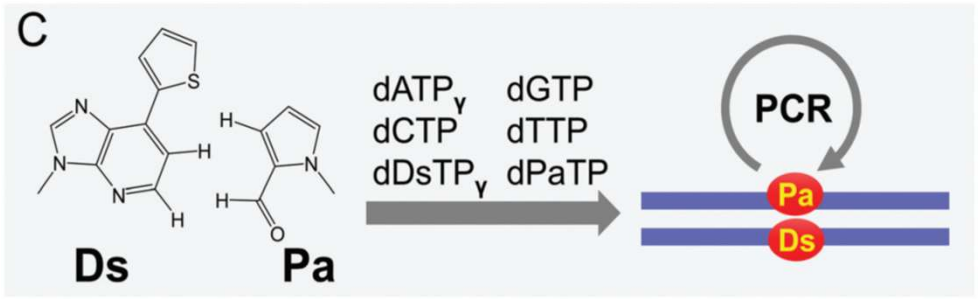


(**A**) The secondary structure of the original tRNA transcript. The positions substituted with **s** are circled. The broken lines show base–base interactions for the 3D structure.
(**B**–**E**) The deep-colored bases were substituted with **s**, which stacks with the light-colored bases, and the yellow spheres represent Mg^{2+}.

M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama and I. Hirao, *Nucleic Acids Res.*, **2007**, *35*, 5360–5369

Steric exclusion and hydrophobic non-natural base pairs





The Ds–Pa pair for faithful replication in combination with γ-amidotriphosphates.

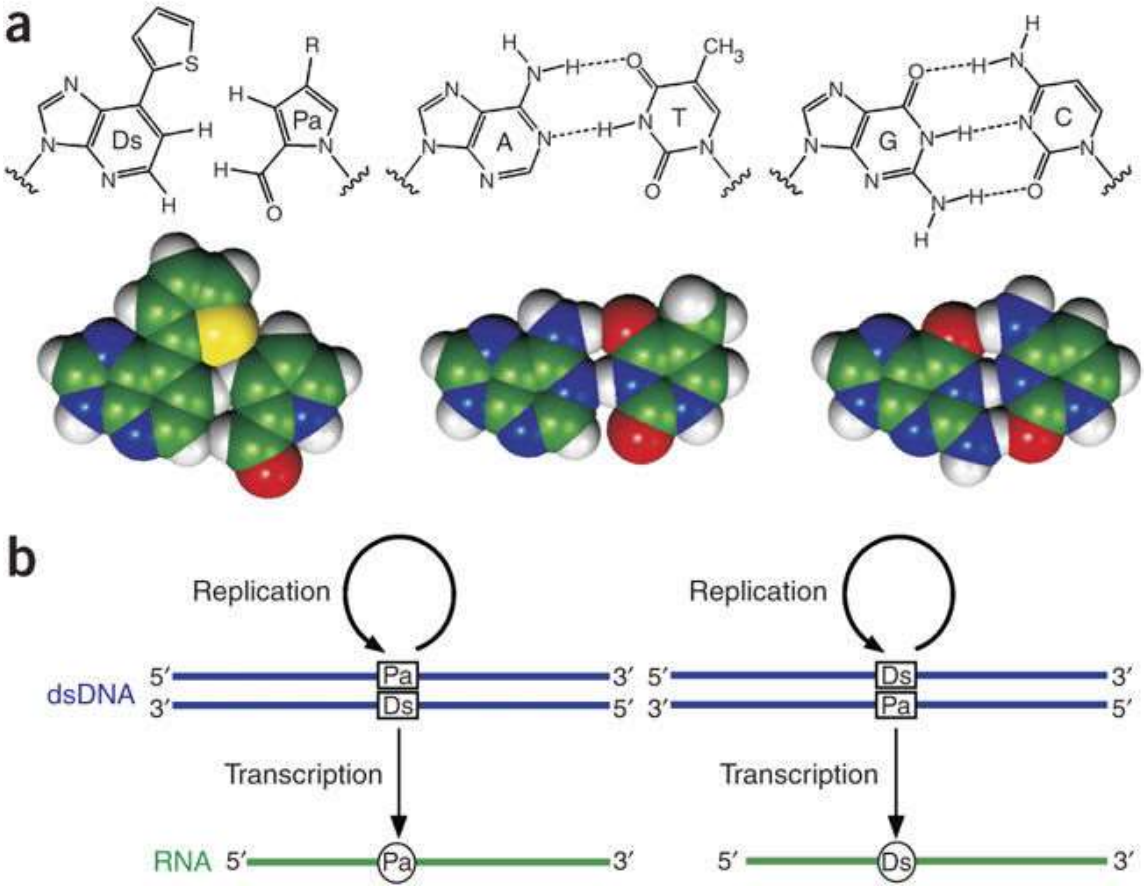
(A) The substrates, dDsTP and dATP, are incorporated opposite Pa and T in the templates, respectively. However, dDsTP and dATP are also misincorporated opposite Ds and Pa, respectively. The Ds incorporation opposite Ds inhibits further primer extension.

(B) The γ-amido-triphosphates, dDsTPγ and dATPγ, reduce the Ds–Ds and A–Pa mispairings in replication.

(C) The combination of the usual triphosphates (dG/C/T/PaTP) and the γ-amido-triphosphates (dDs/ATPγ) enables the faithful PCR amplification of 6-letter DNA

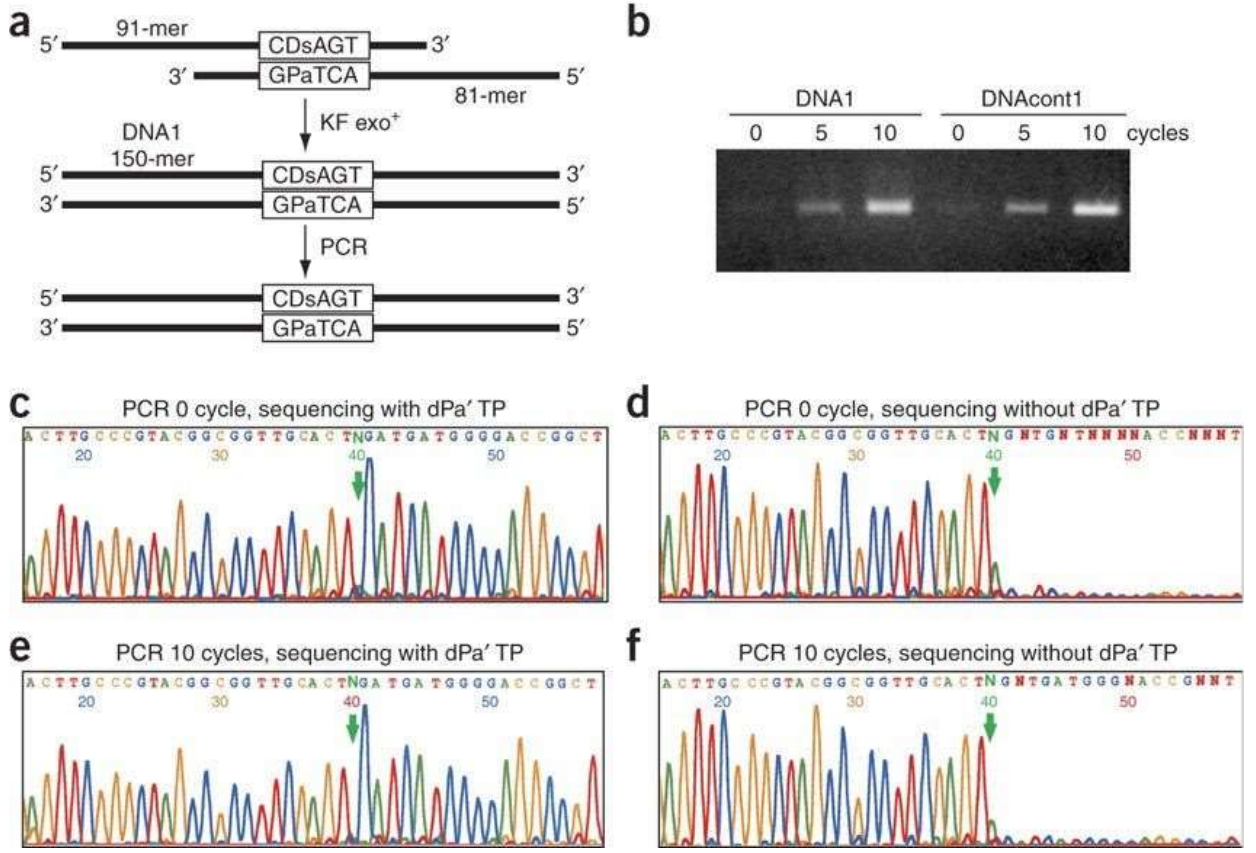
M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

Replication and transcription of the hydrophobic UBP



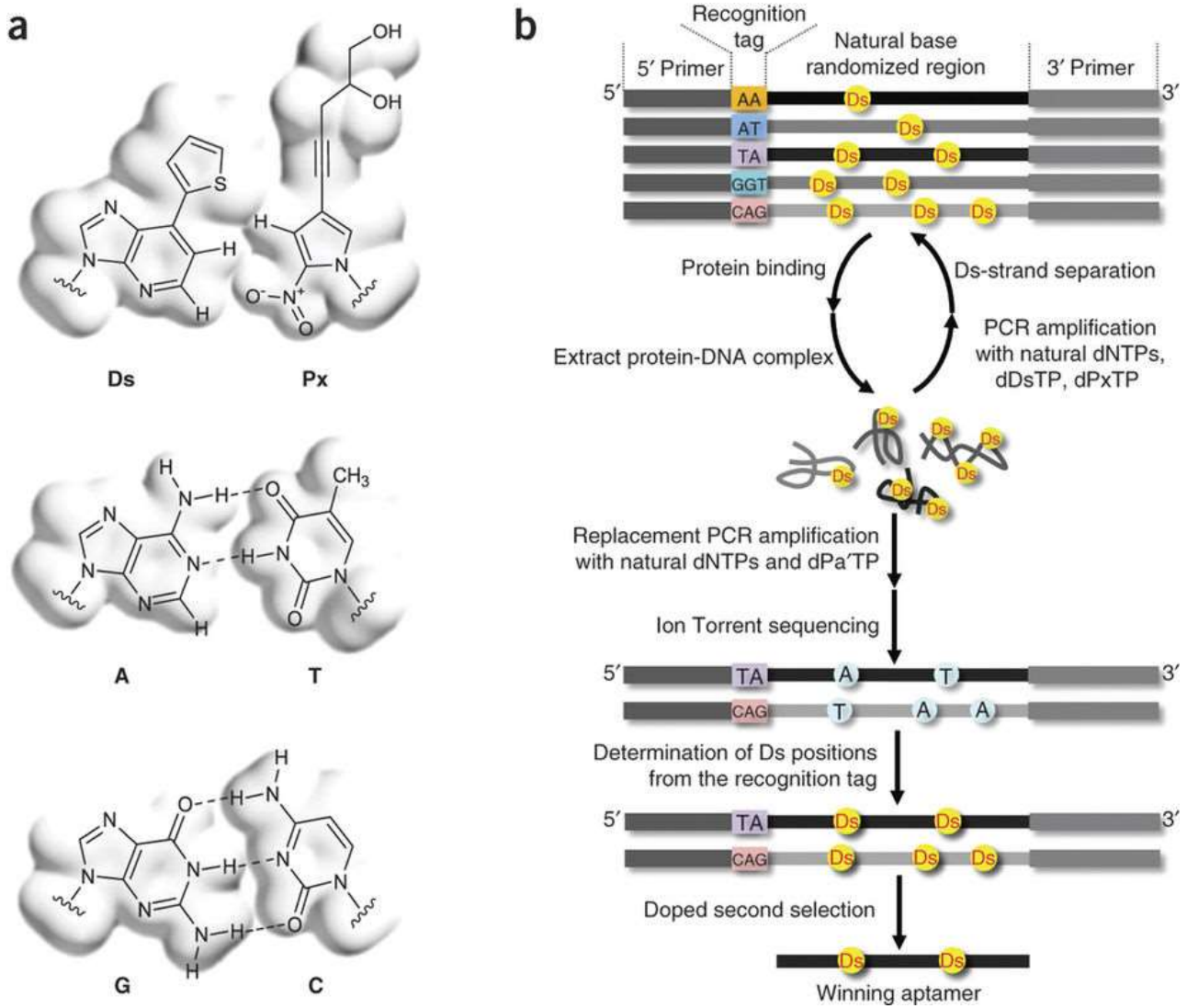
(**a**) The unnatural Ds-Pa and natural base pairs (Pa: R = H, Pa': R = C≡C–CH_3). (**b**) The unnatural base pair system that functions in PCR amplification, primer extension, DNA sequencing and T7 transcription. Original DNA templates were prepared by chemical synthesis and ligation, and were amplified by PCR with unmodified dNTPs (dPaTP, dGTP, dCTP and dTTP) and modified $dNTP_N$s (γ-amidotriphosphates, $dDsTP_N$ and $dATP_N$). RNA molecules containing Pa or modified Pa bases at specific positions were transcribed from DNA templates containing Ds in the template strands, by T7 RNA polymerase with PaTP (or modified PaTP) and the natural NTPs (left). RNA molecules containing Ds were transcribed from DNA templates containing Pa in the template strands, with DsTP and the natural NTPs (right).

I. Hirao *et al. Nature Methods* **2006**, *3,* 729-735



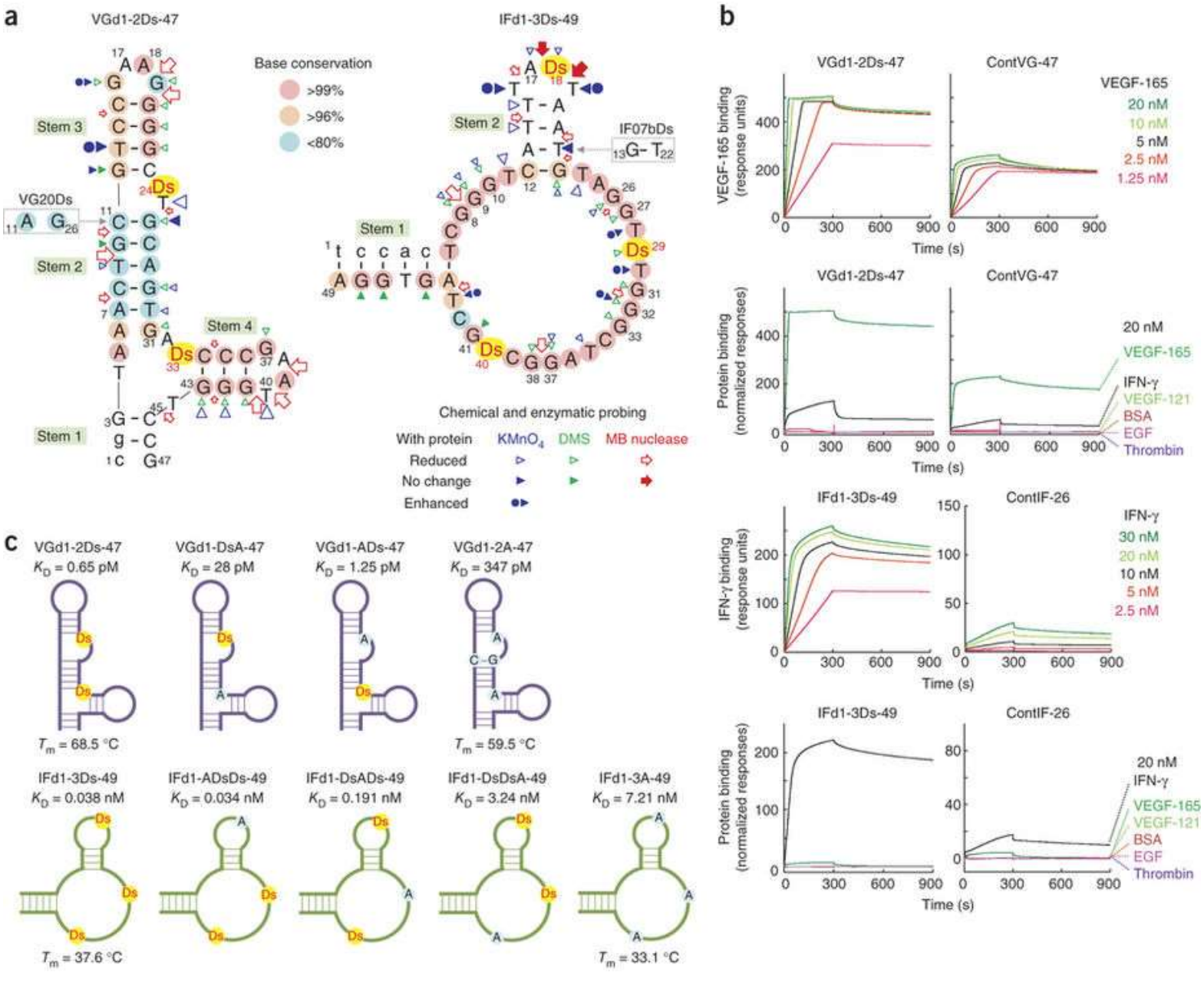
(**a**) The double stranded DNA fragment (150-mer, DNA1) was prepared by primer extension using chemically synthesized DNA fragments (91-mer and 81-mer) containing Ds and Pa. (**b**) Agarose-gel analysis of original DNA fragments (0 cycle) and PCR products after 5 and 10 cycles of amplification. For DNA1, PCR was performed with 0.04 unit/µl Vent DNA polymerase and the following cycle conditions: 0.5 min at 94 °C, 0.5 min at 45 °C and 4 min at 65 °C. For DNAcont1 consisting only of the natural bases, PCR was performed with 0.01 unit/µl Vent DNA polymerase with the following cycle conditions: 0.5 min at 94 °C, 0.5 min at 45 °C and 1 min at 72 °C. (**c**–**f**) DNA sequencing, in the presence (**c,e**) or absence (**d**,**f**) of dPa'TP, of the original DNA1 (**c**,**d**) and PCR-amplified DNA1 after 10 cycles using the unnatural base pair system (**e**,**f**).

Ds-Px noncanonical base pair – High-affinity DNA aptamers



Targets:
VEGF-165
and IFN-γ

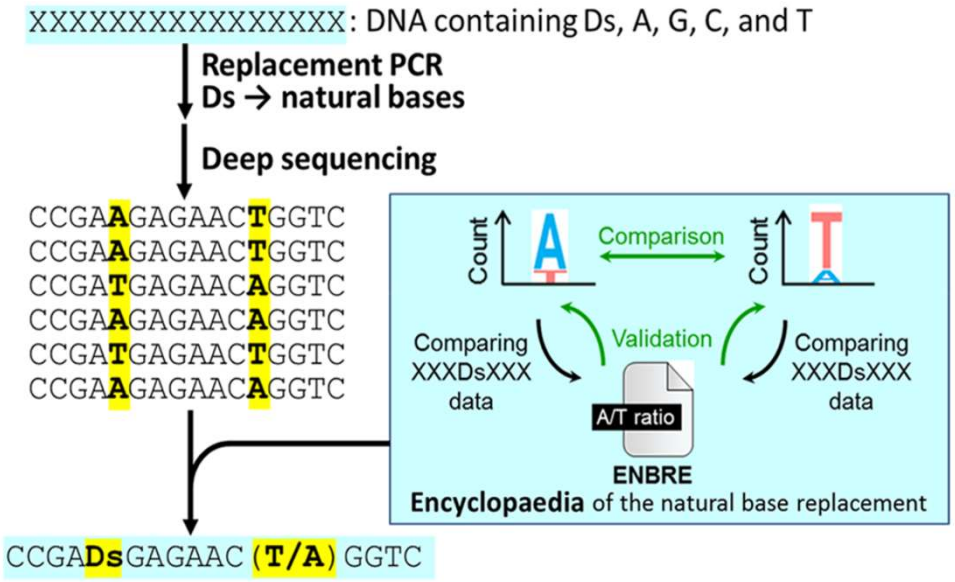
I. Hirao *et al. Nature Biotechnology* **2013**, *31,* 453-458

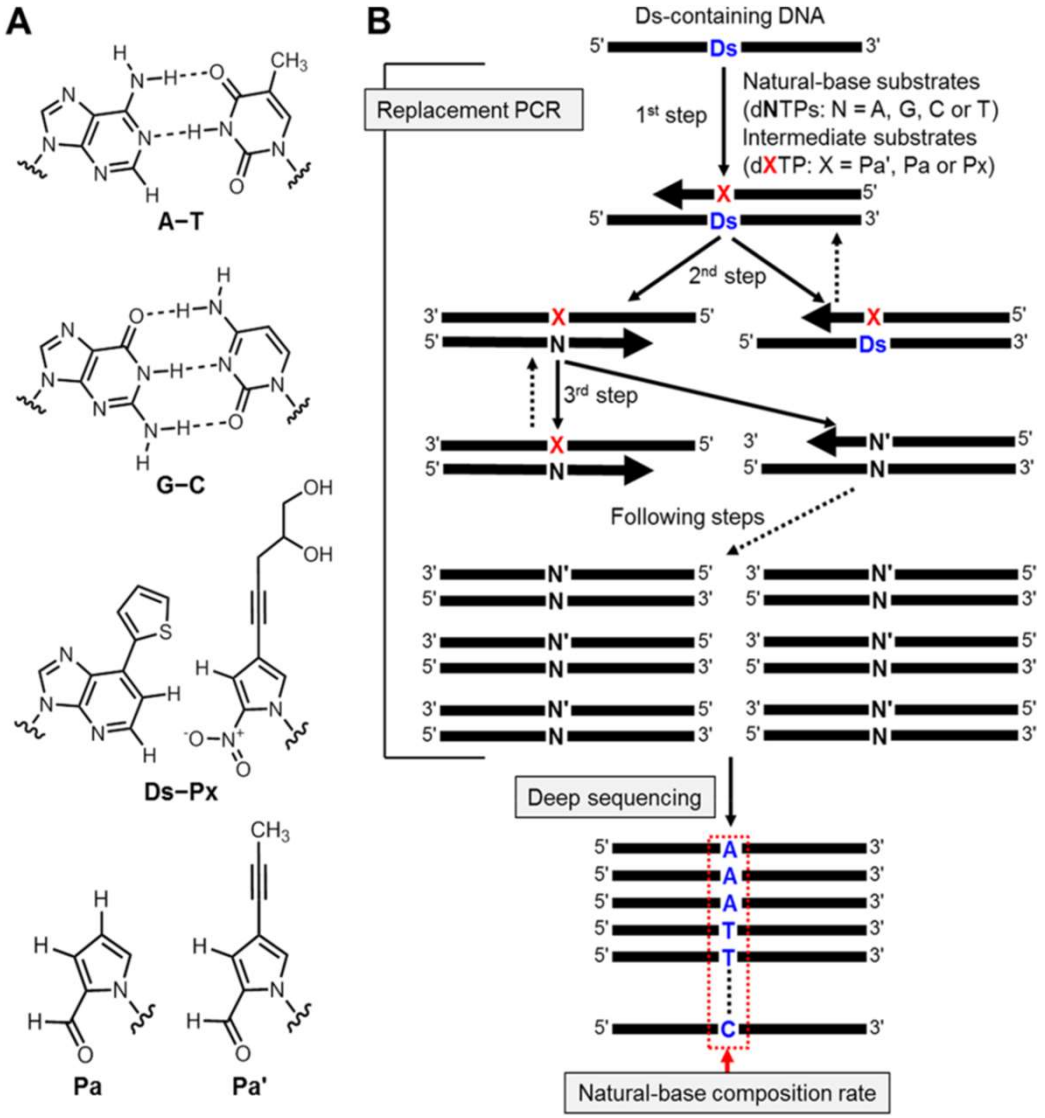


Characterizations and binding affinities of anti–VEGF-165 aptamer (VGd1-2Ds-47) and anti–IFN-γ aptamer (IFd1-3Ds-49)

I. Hirao *et al. Nature Biotechnology* **2013**, *31*, 453-458

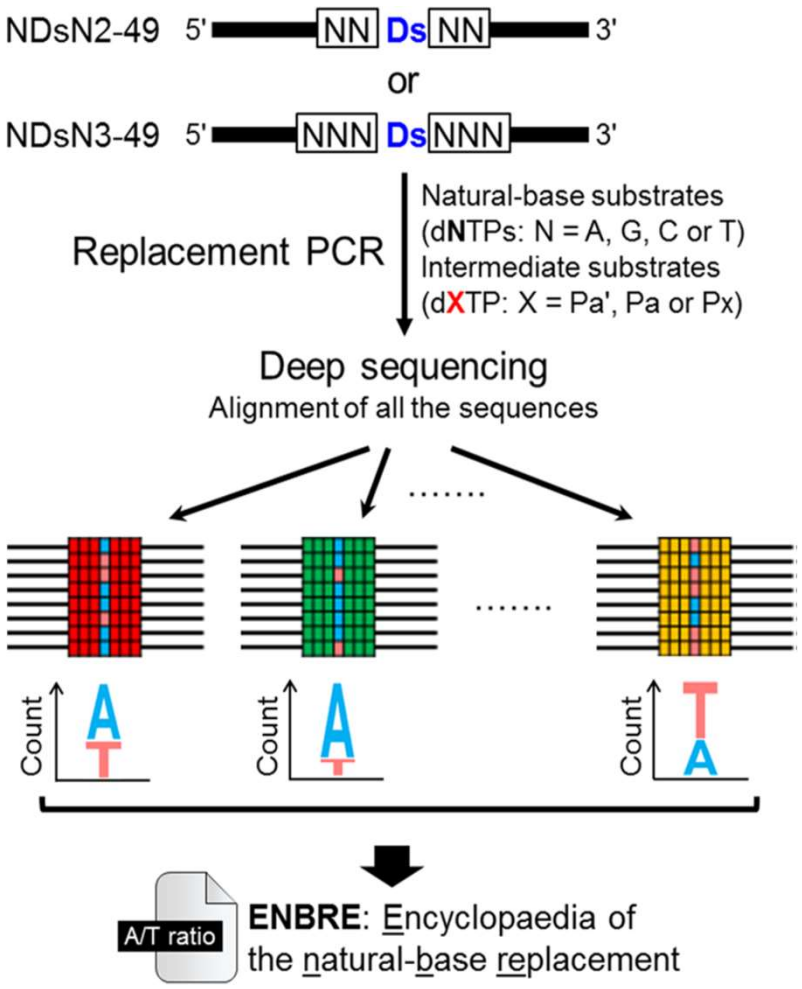
Ds-Pa – need for deep sequencing in the amplification process

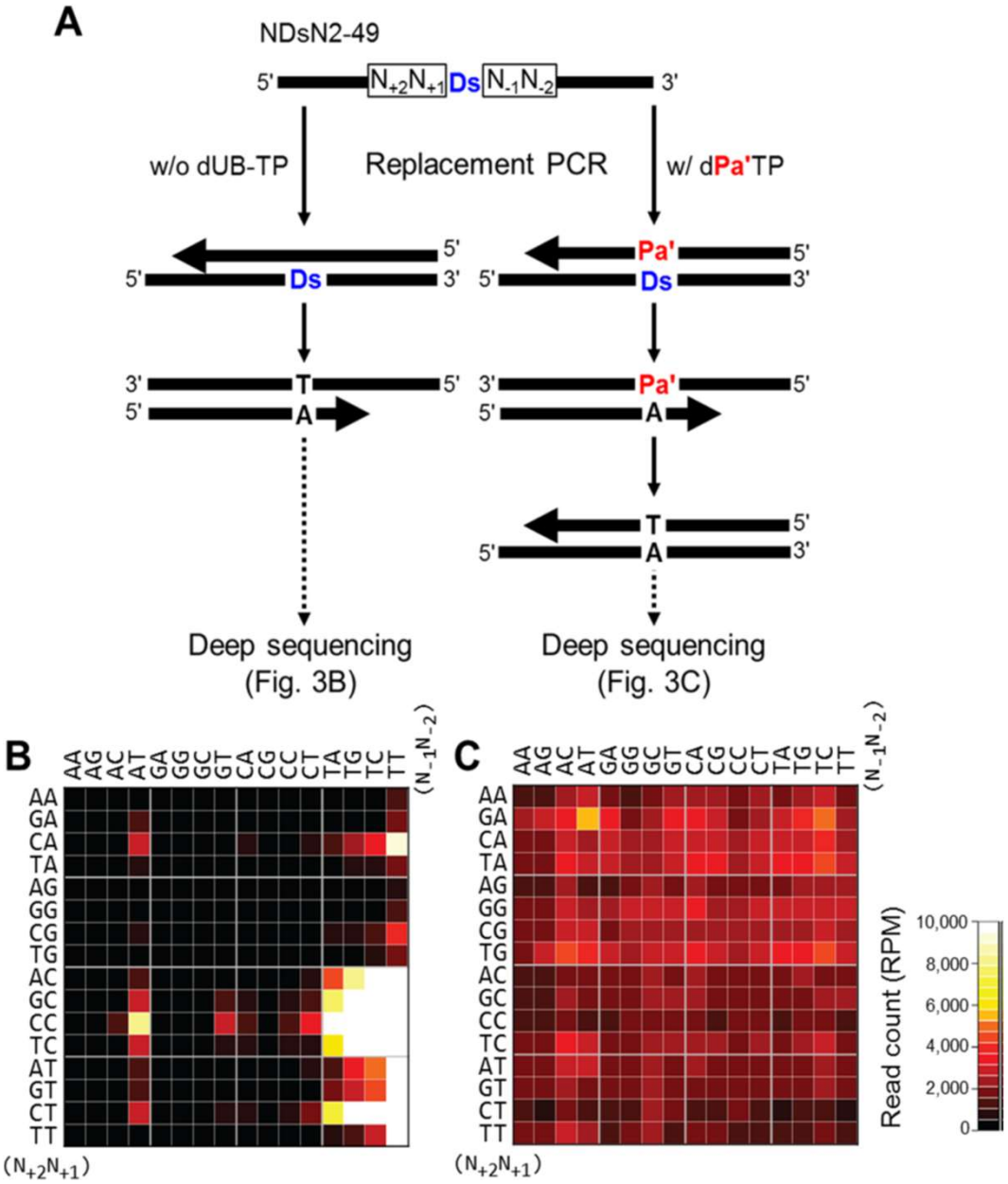




K. Hamashima, I. Hirao *et al. ACS Synth. Biol.* **2019**, *8*, 1401-1410

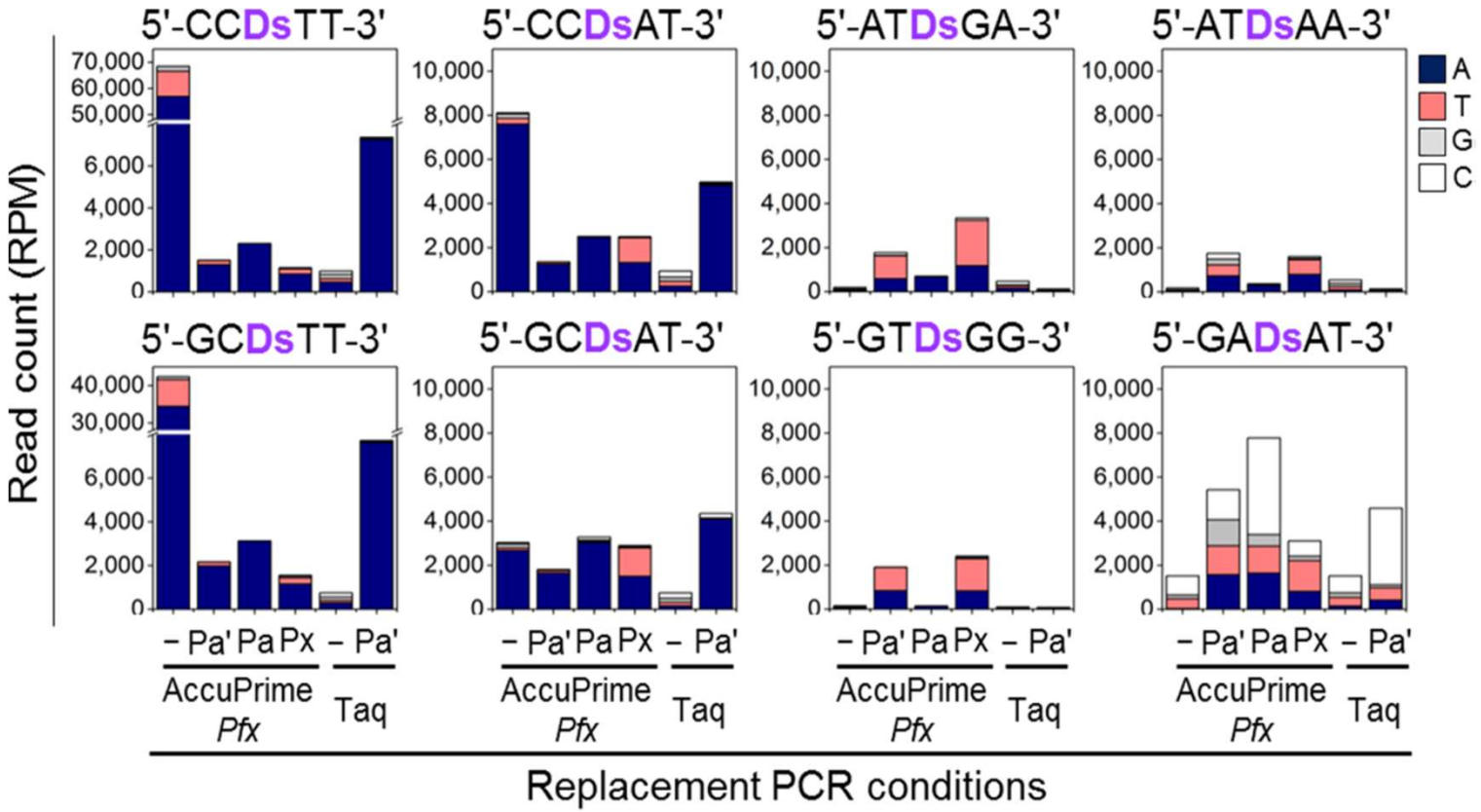
Ds-Pa – deep sequencing

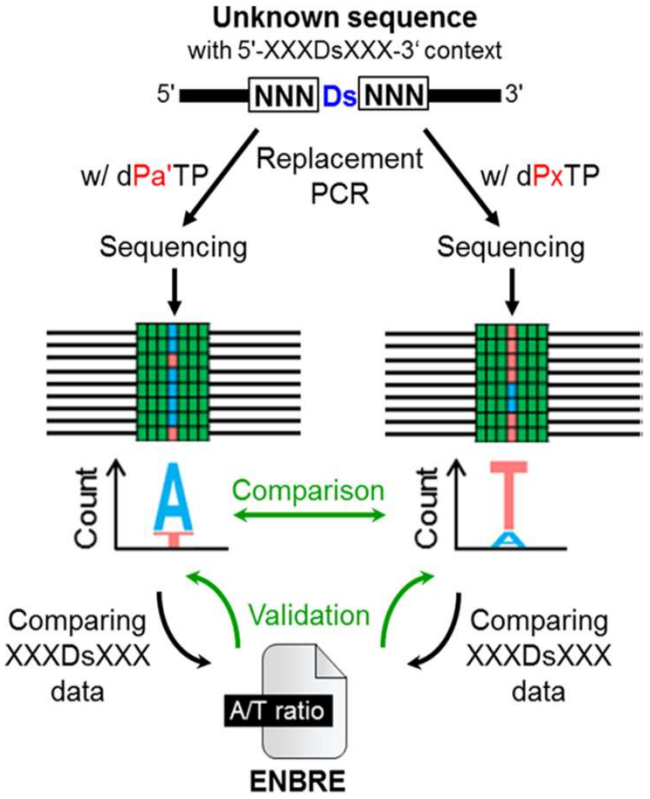




K. Hamashima, I. Hirao *et al. ACS Synth. Biol.* **2019**, *8*, 1401-1410

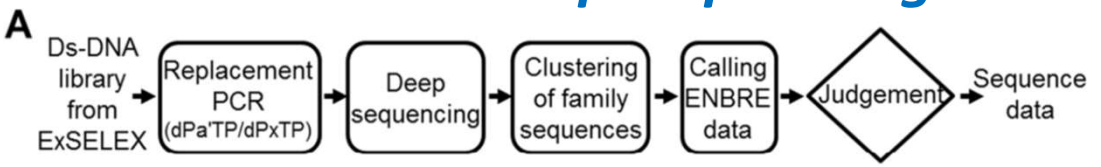
Ds-Pa – deep sequencing

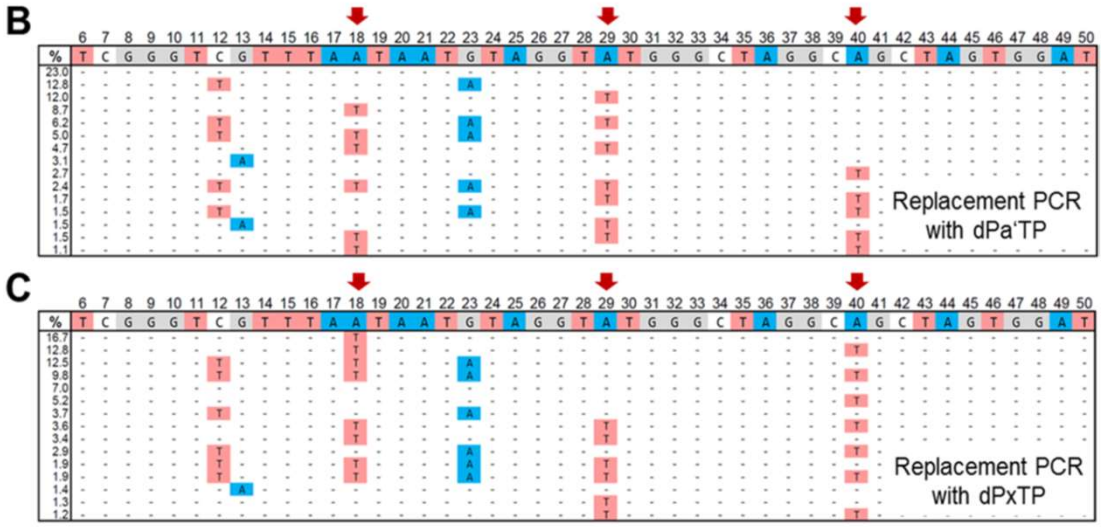


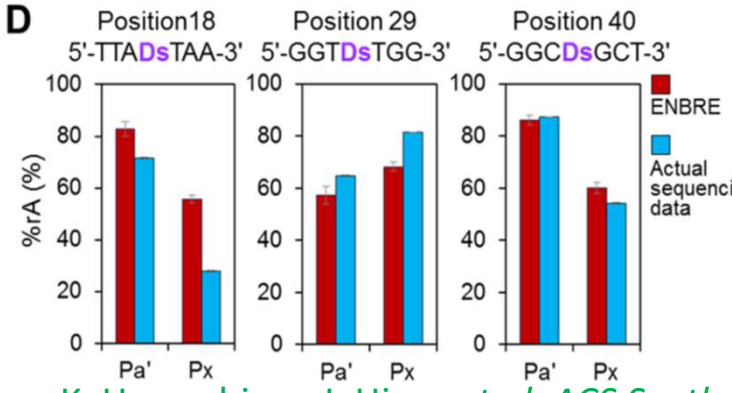


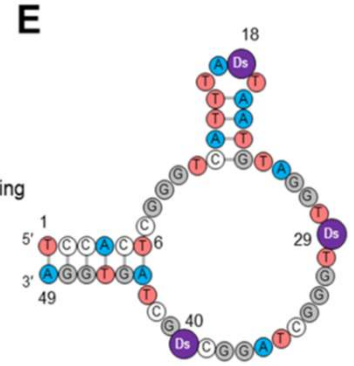
K. Hamashima, I. Hirao *et al. ACS Synth. Biol.* **2019**, *8,* 1401-1410

Ds-Pa – deep sequencing



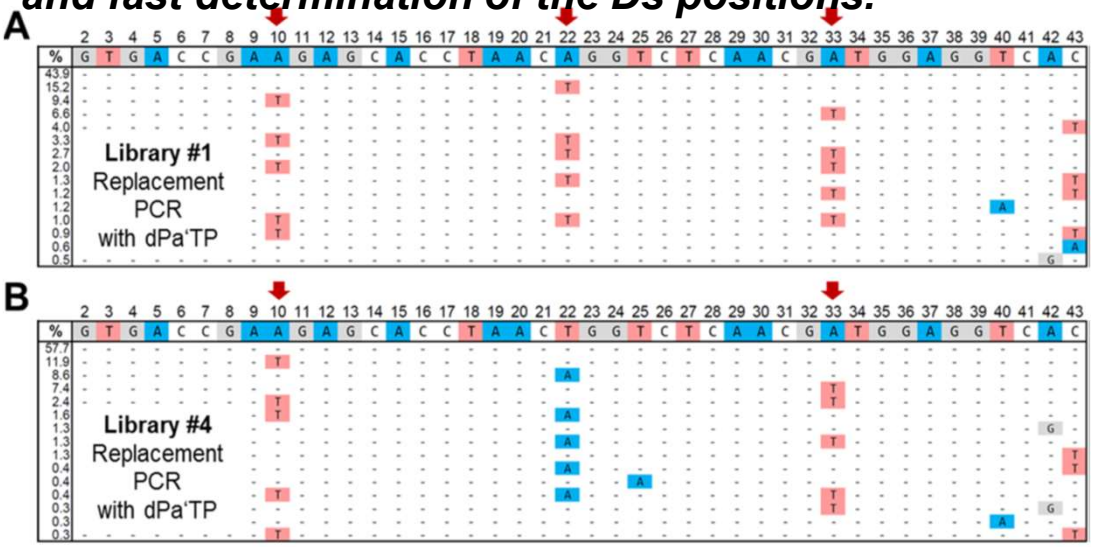


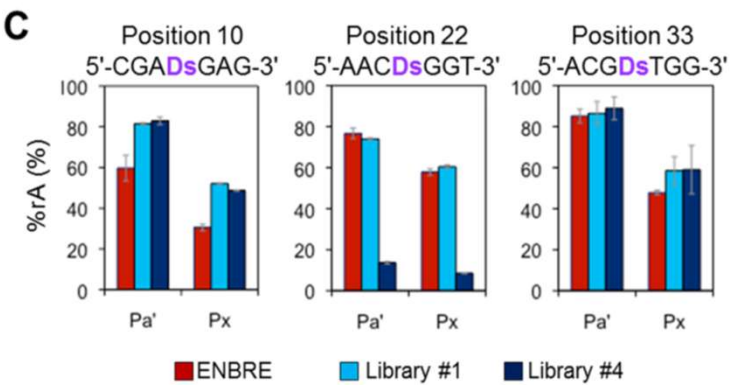


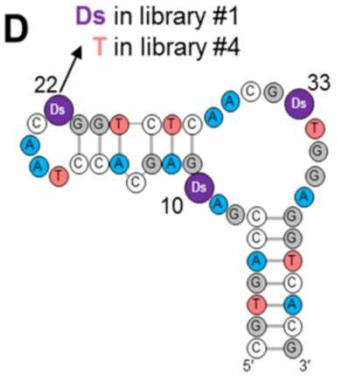


K. Hamashima, I. Hirao *et al. ACS Synth. Biol.* **2019**, *8*, 1401-1410

Referring to the encyclopedia data allows for simple and fast determination of the Ds positions.

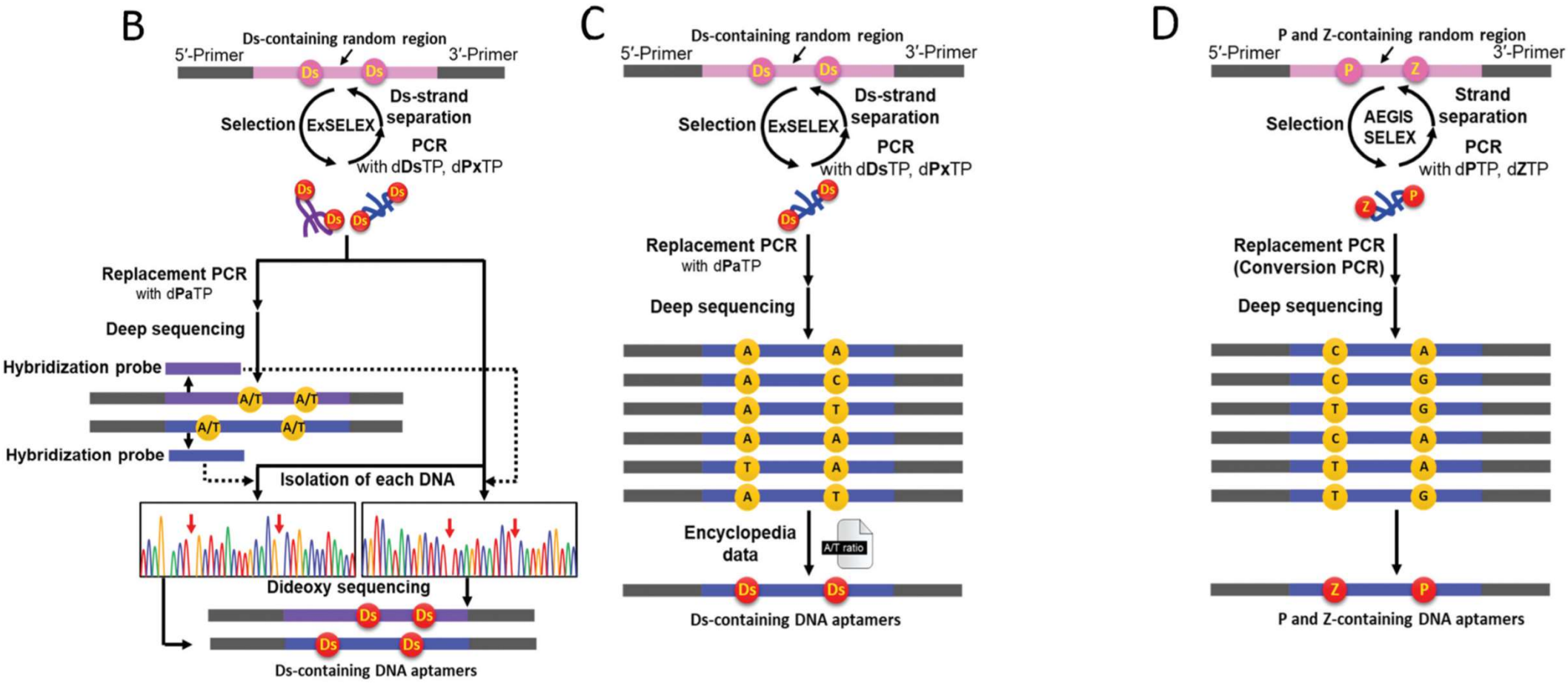






Comparison of the replacement patterns between two conditions enables the Ds positions to be distinguished from other natural-base positions.

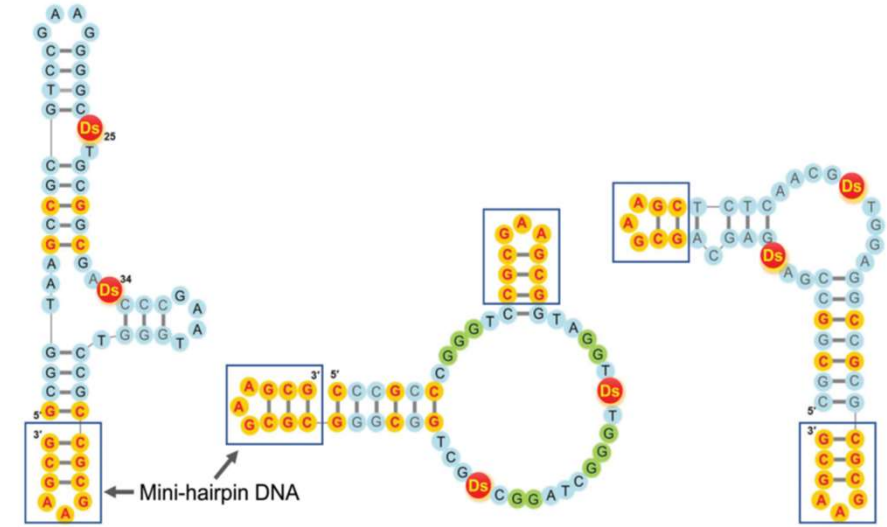
Ds-Pa aptamer selection using deep sequencing



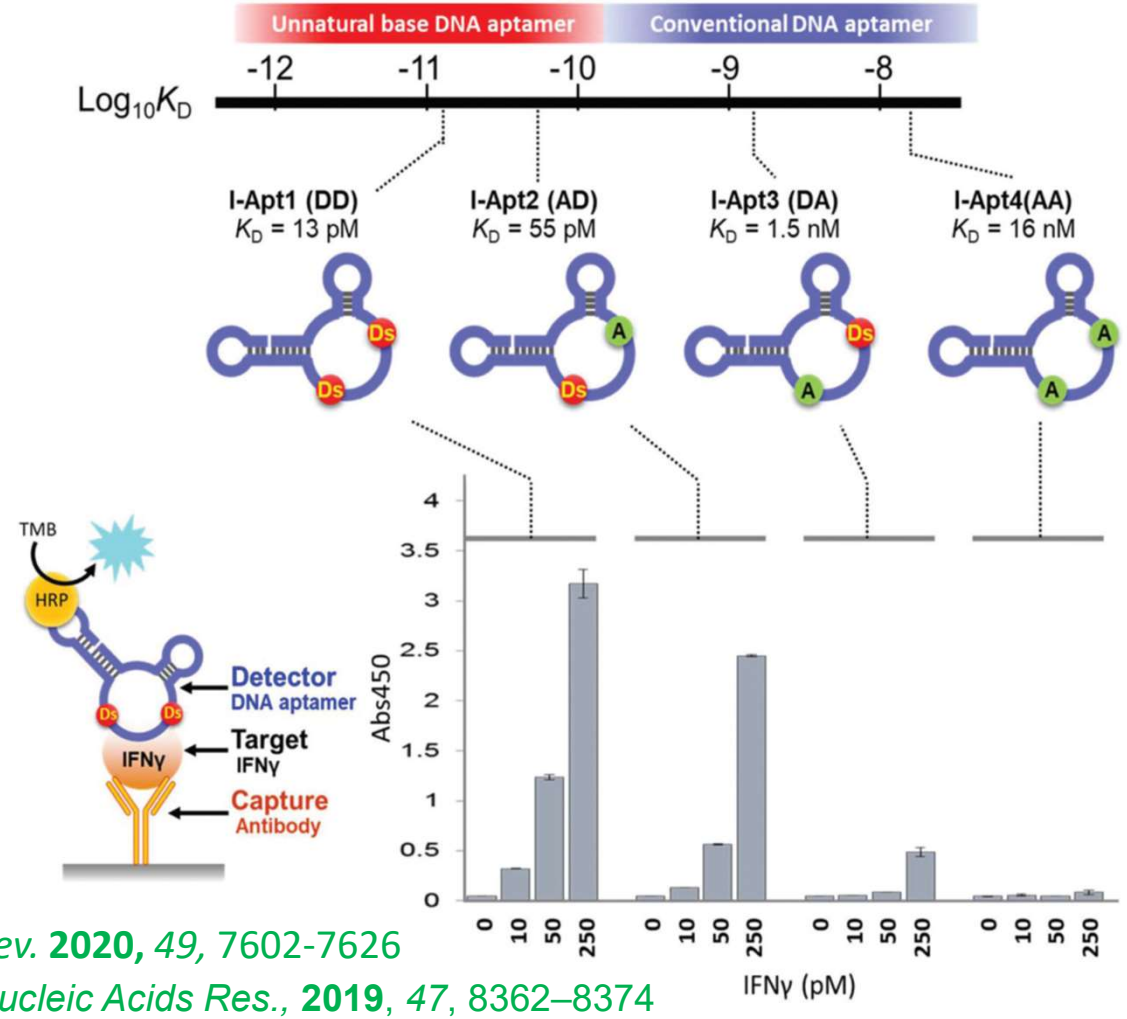
M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

Ds-Pa aptamer selection using deep sequencing

Secondary structures of the optimized Ds-DNA aptamers targeting VEGF165, IFNγ, and vWF.



Generation of a molecular affinity ruler using anti-IFNγ Ds-DNA aptamer variants, replacing the Ds base with the natural A base.

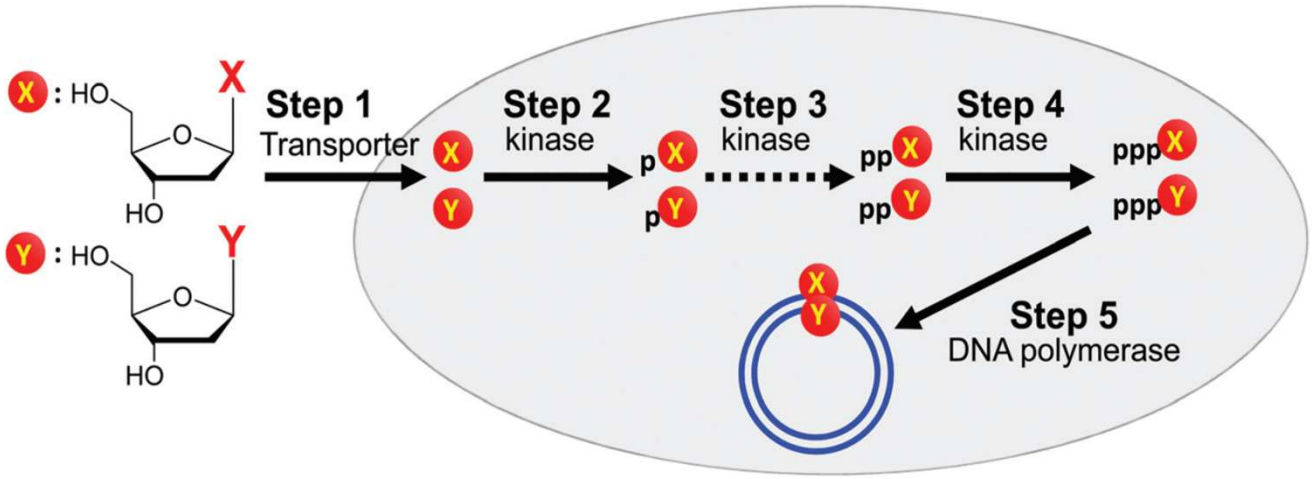


M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

M. Kimoto, Y. W. S. Lim and I. Hirao, *Nucleic Acids Res.*, **2019**, *47*, 8362–8374

A semi-synthetic organism with an expanded genetic alphabet

Nucleoside phosphorylation: a biosynthetic pathway to provide unnatural substrates (dX/dYTP) in a cell, using their unnatural nucleosides (dX/dY) as a source.



In Step 2, the nucleoside kinase from D. melanogaster (DmdNK) was explored for its ability to phosphorylate the unnatural nucleosides by three teams in vitro. Benner's team confirmed that the kinase in Step 4 was available for their Z and P bases in vitro. All UB triphosphates are, at least, accepted by the Klenow fragment of E. coli Pol I in vitro

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

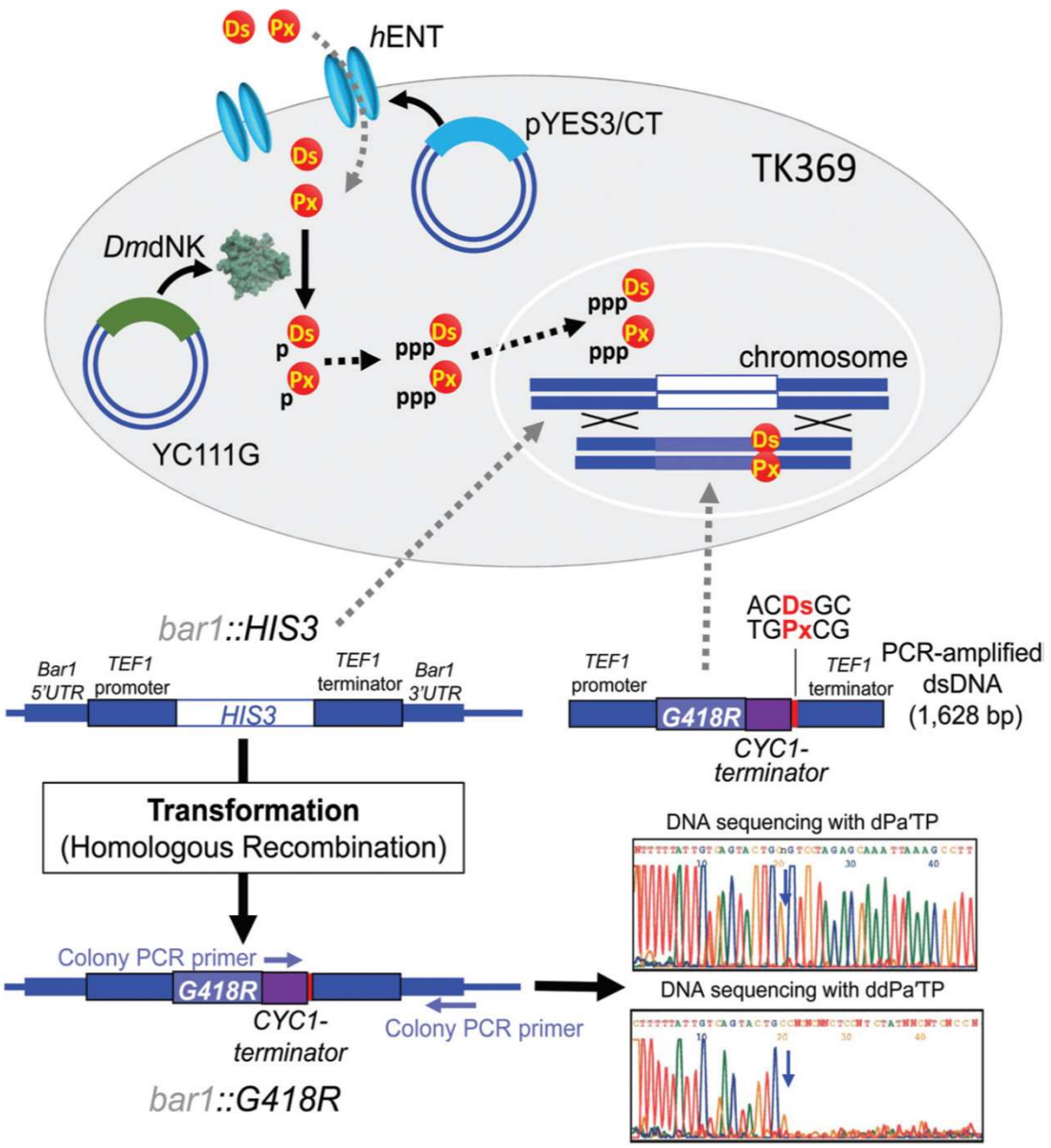
A semi-synthetic organism with an expanded genetic alphabet

Homologous recombination strategy using the yeast strain TK369, to integrate dsDNA with the Ds–Px pair.

DmdNK – nucleoside kinase
hENT - human equilibrium nucleoside transporter

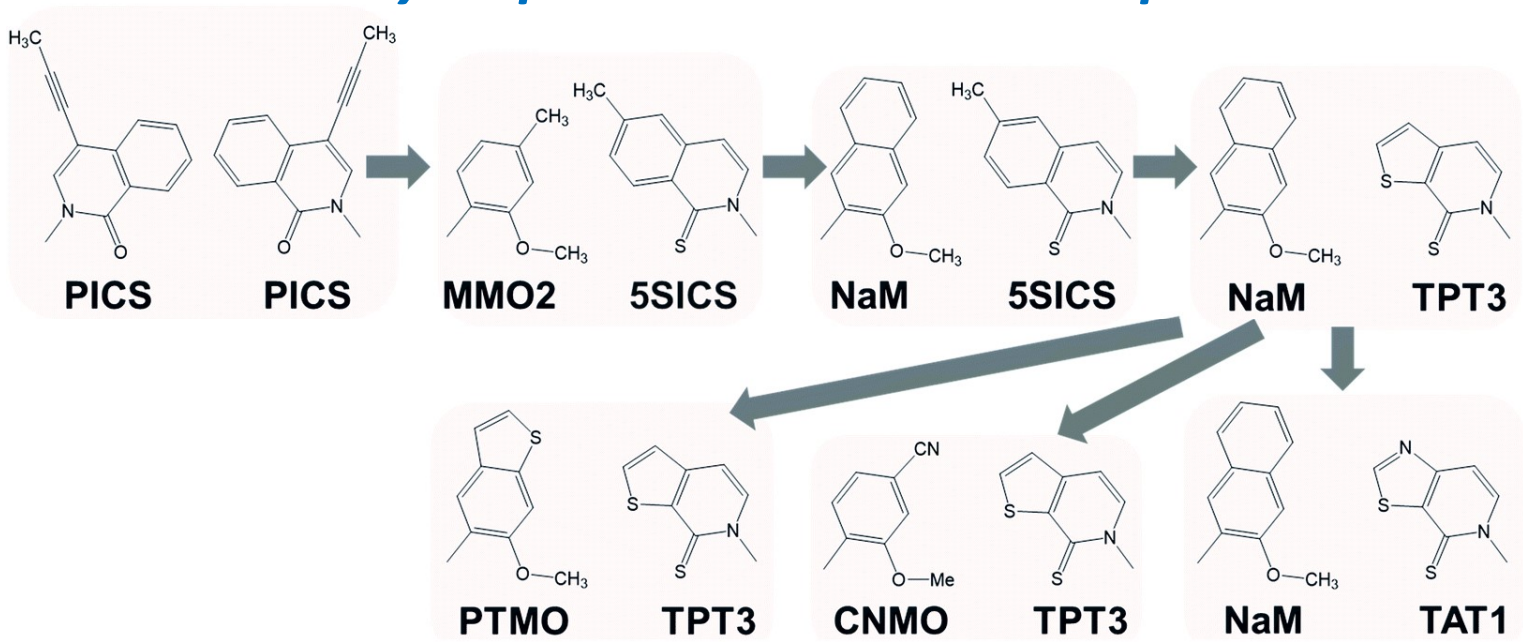
The Ds and Px nucleosides supplied in the media.
The long dsDNA with the Ds–Px pair was prepared by fusion/overlapping PCR,

Analysis: colony PCR with dDsTP and dPxTP, then modified Sanger sequencing.



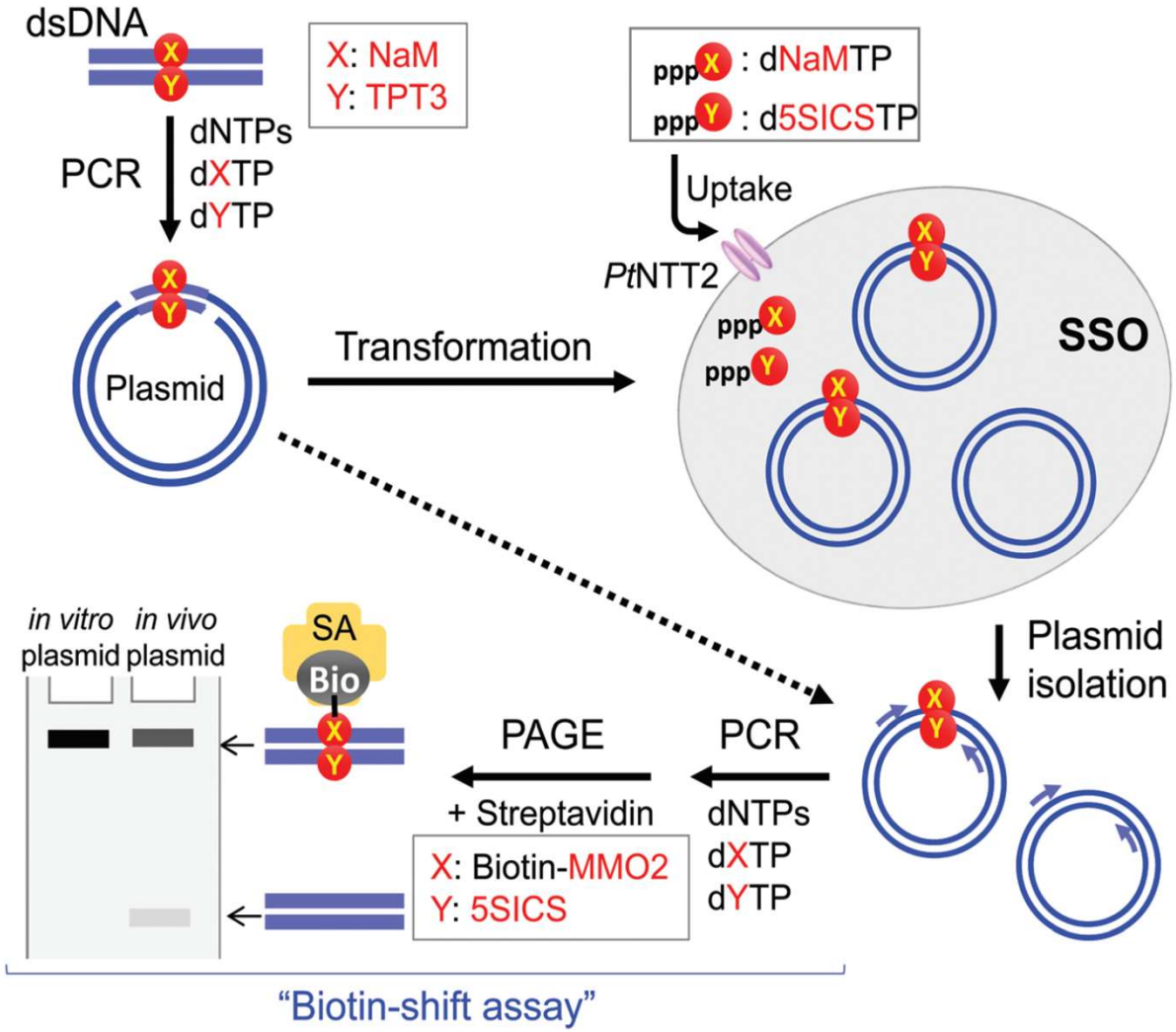
M. Kimoto, I. Hirao *Chem. Soc. Rev.* **2020,** *49,* 7602-7626

Hydrophobic unnatural base pairs



Romesberg (1999): **PICS–PICS**, but the polymerase reaction is paused at the PICS–PICS pairing position due to self-stacking.
Romesberg (2008): the **MMO2–5SICS** pair - the methyl group of 5SICS prevented the disfavored 5SICS–5SICS self-pairing.
Further optimized into the **NaM–5SICS** pair (>99% fidelity in PCR and transcription).
Romesberg (2014) the **NaM–TPT3** pair (systematic *in vitro* replication screening) ➔ semi-synthetic organism (SSO) of *E. coli* with six-letter DNA.
Optimizations: the **CNMO–TPT3** pair to increase the UBP retention on a plasmid within the SSO,
the **NaM–TAT1** pair for efficient transcription and translation in the SSO.
Using these UBP systems, they reported protein synthesis with the site-specific incorporation of unnatural amino acids into proteins in the SSO.

A semi-synthetic organism with an expanded genetic alphabet



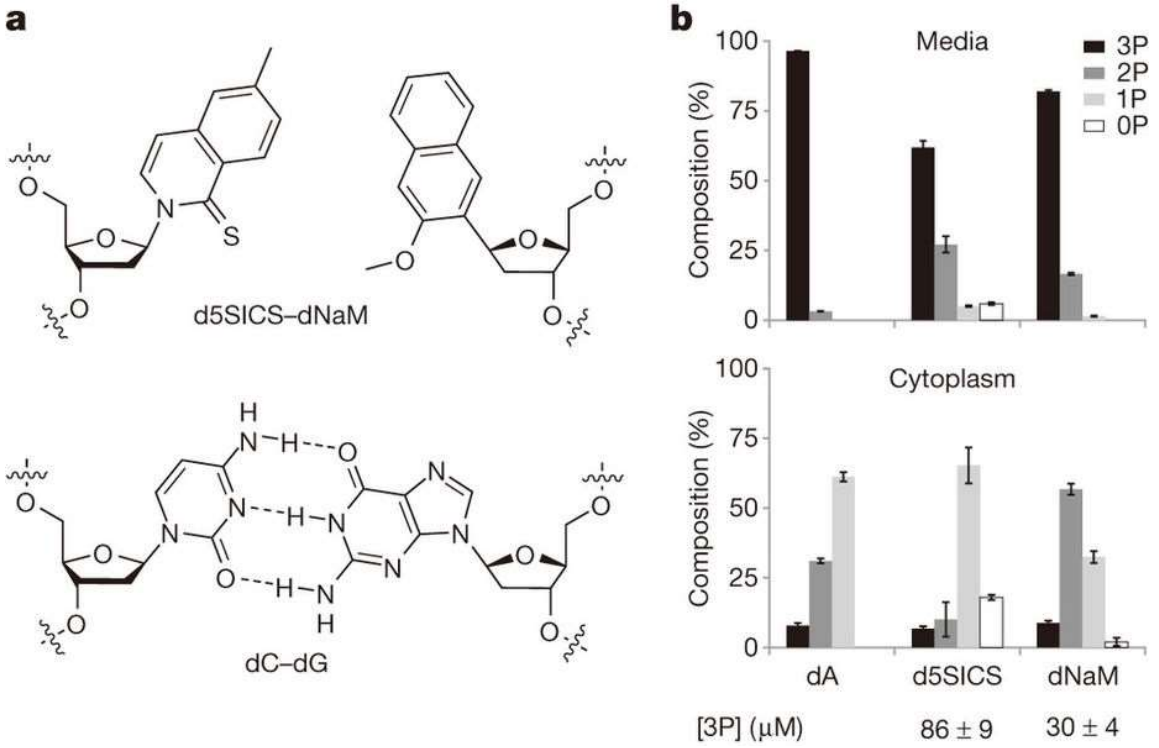
The first creation of an SSO, using the NaM–TPT3 and NaM–5SCIS pairs and E. coli C41(DE3) as the host strain.

The plasmid containing the NaM–TPT3 pair was prepared by PCR and transformed into E. coli C41(DE3) with a PtNTT2 overexpression system. The transformants were cultured in inorganic phosphate-rich growth media in the presence of the NaM and 5SCIS triphosphates.

The propagated plasmids were isolated and subjected to the analysis of the NaM–5SCIS pair retention by a biotin-shift assay

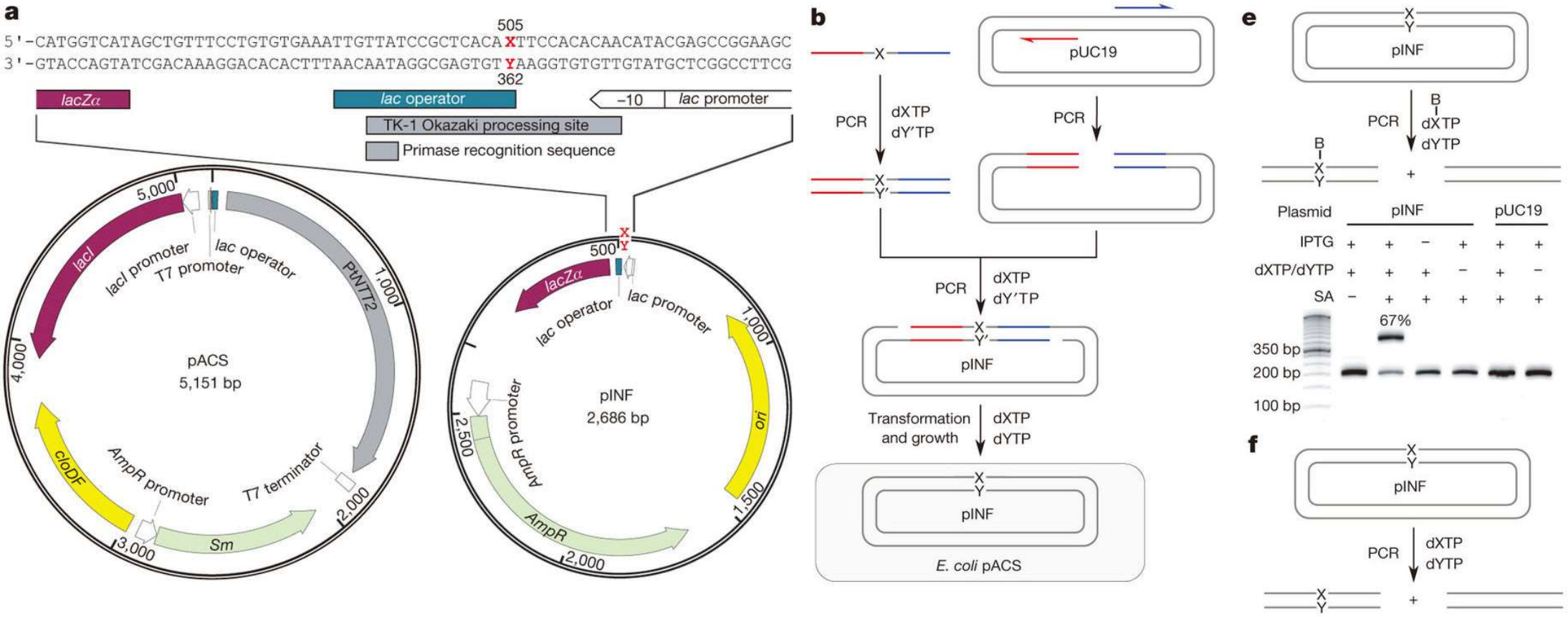
Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *Nature* **2014**, *509,* 385–388

A semi-synthetic organism with an expanded genetic alphabet



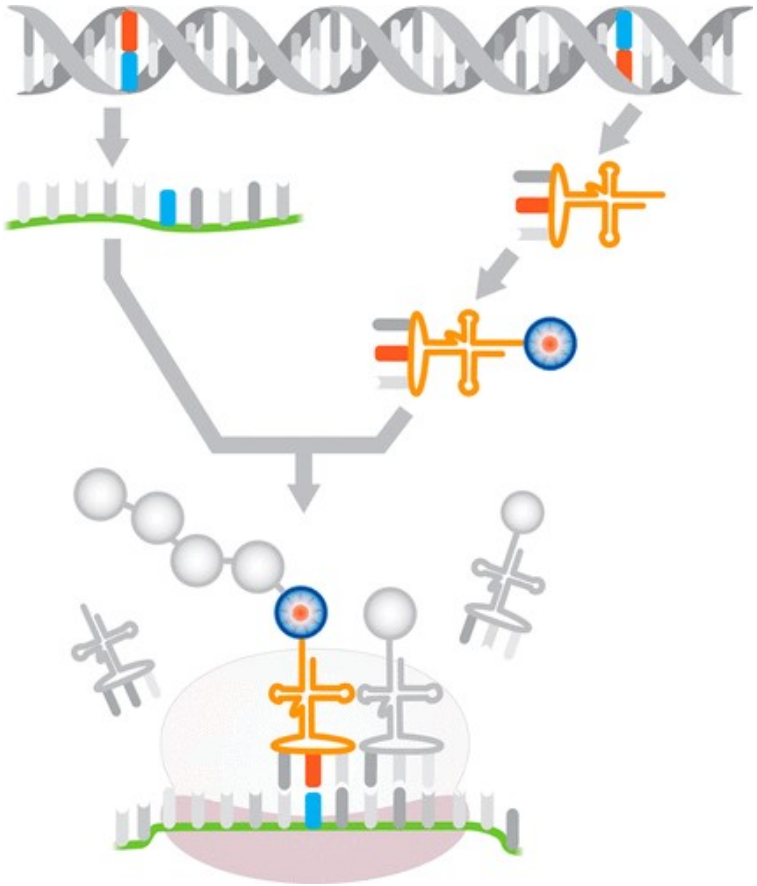
a, Chemical structure of the d5SICS–dNaM UBP compared to the natural dG–dC base pair. **b**, Composition analysis of d5SICS and dNaM in the media (top) and cytoplasmic (bottom) fractions of cells expressing *Pt*NTT2 after 30 min incubation; dA shown for comparison. 3P, 2P, 1P and 0P correspond to triphosphate, diphosphate, monophosphate and nucleoside, respectively; [3P] is the intracellular concentration of triphosphate.

Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *Nature* **2014**, *509,* 385–388



Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *Nature* **2014**, *509,* 385–388

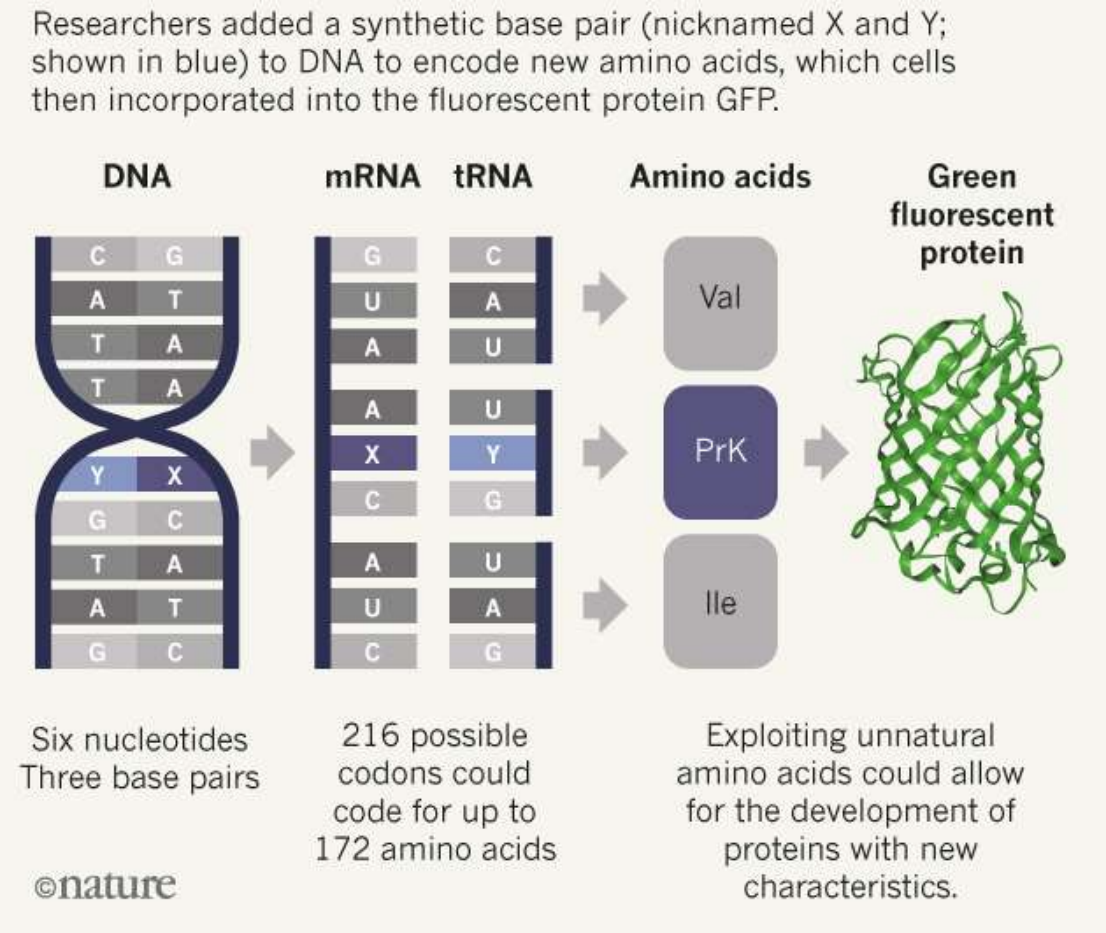
A semi-synthetic organism with an expanded genetic alphabet



The SSO replicates DNA containing the d**NaM**-d**TPT3** UBP (blue and red, respectively), transcribes mRNA and tRNA with complementary codons and anticodons containing **NaM** or **TPT3**, uses an orthogonal synthetase to charge the tRNA with an ncAA, and uses the charged tRNA to translate the mRNA into proteins containing ncAAs

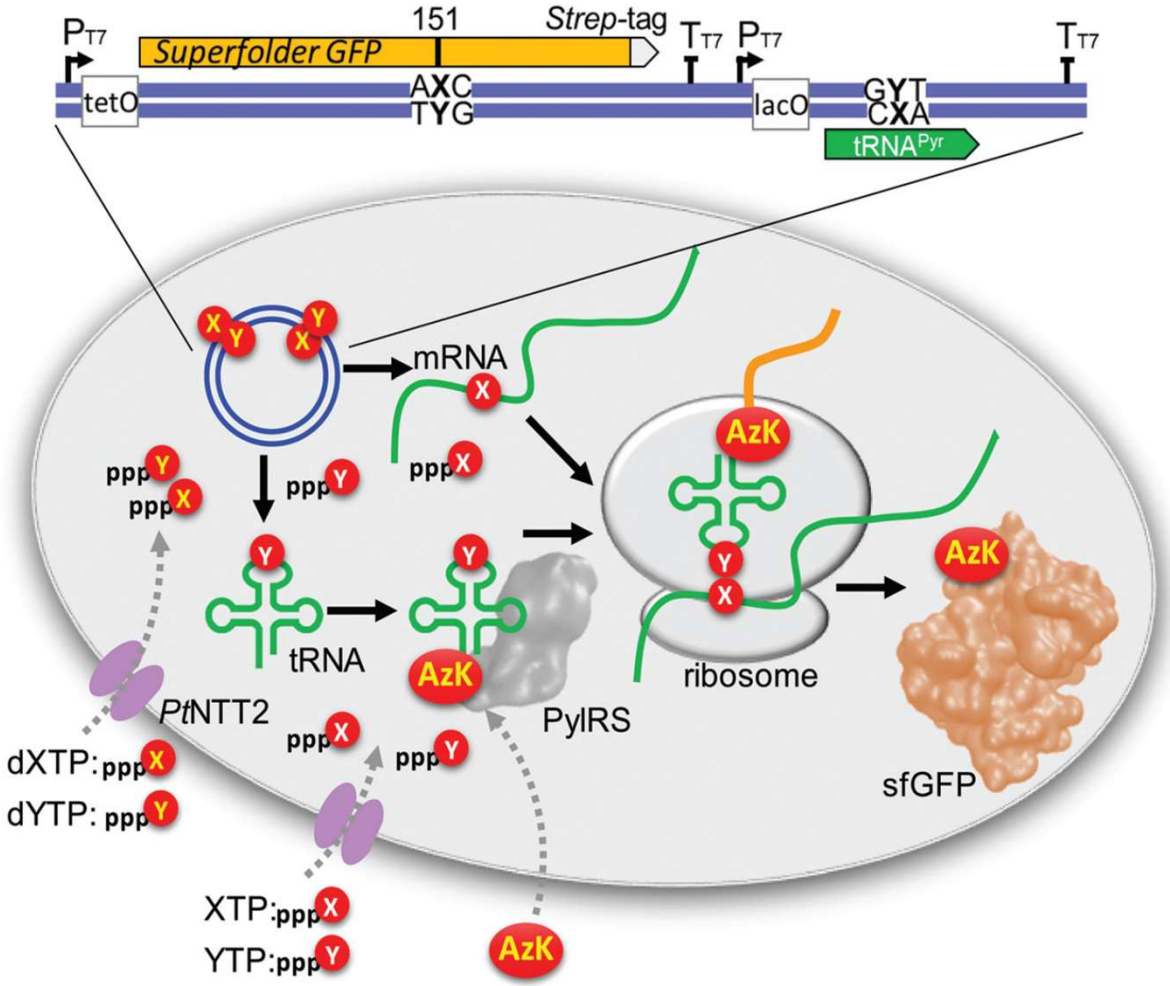
Y. Zhang, F. Romesberg, *Biochemistry* **2018**, *57, 15*, 2177–2178

A semi-synthetic organism with an expanded genetic alphabet



Zhang, Y.; Romesberg, Floyd E. *et al.* *Nature* **2017**, *551,* 644-647

A semi-synthetic organism with an expanded genetic alphabet

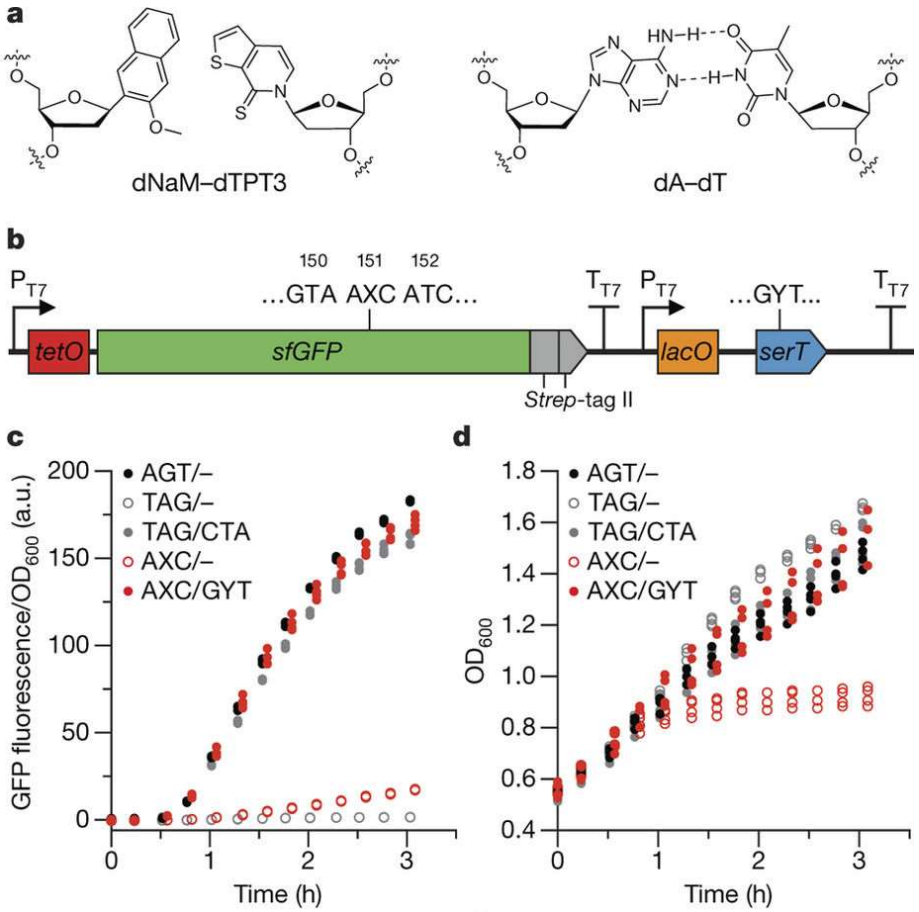


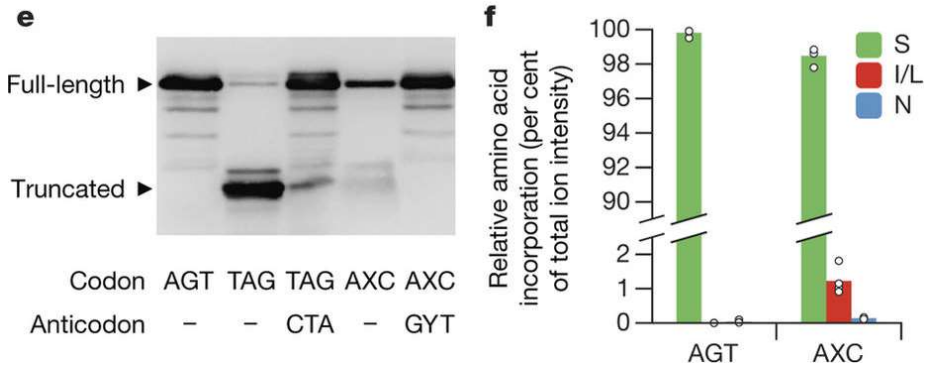
Zhang, Y.; Romesberg, Floyd E. *et al.* *Nature* **2017**, *551,* 644-647

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

A semi-synthetic organism with an expanded genetic alphabet

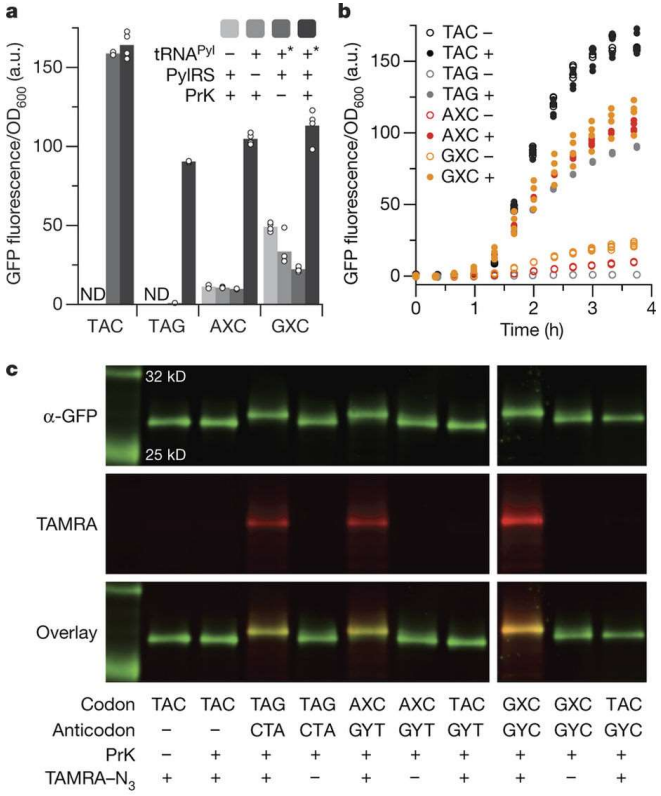
The *in vivo* transcription of DNA containing dNaM and dTPT3 into mRNAs with two different unnatural codons and tRNAs with cognate unnatural anticodons, and their efficient decoding at the ribosome to direct the site-specific incorporation of natural or non-canonical amino acids into superfolder green fluorescent protein. The resulting semi-synthetic organism both encodes and retrieves increased information and should serve as a platform for the creation of new life forms and functions.

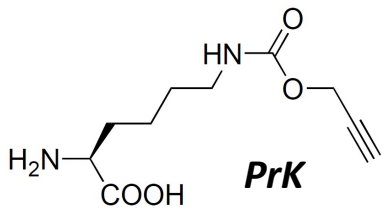




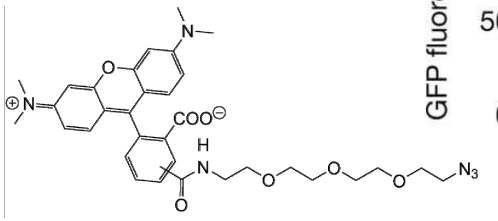
Zhang, Y.; Romesberg, Floyd E. *et al.* *Nature* **2017**, *551*, 644-647

A semi-synthetic organism with an expanded genetic alphabet

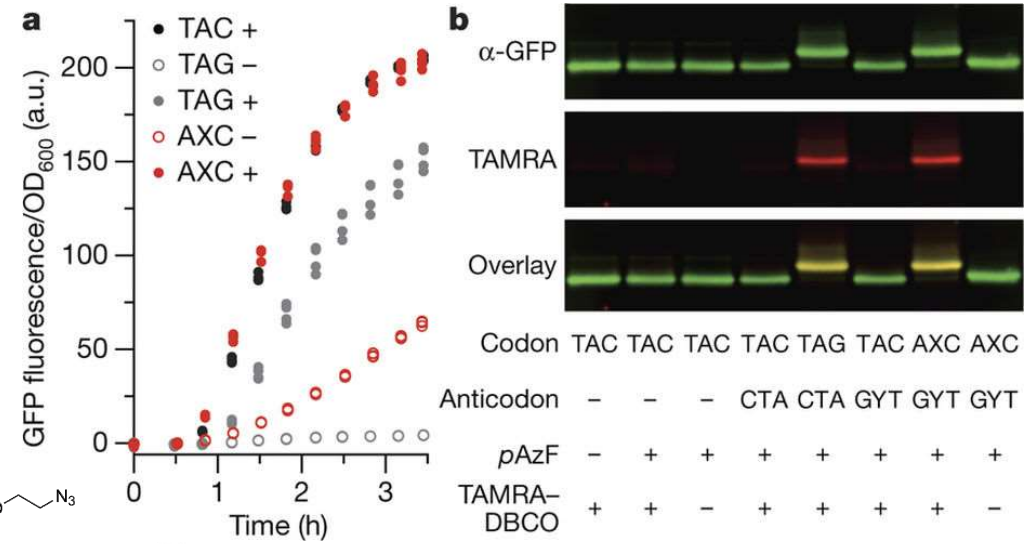


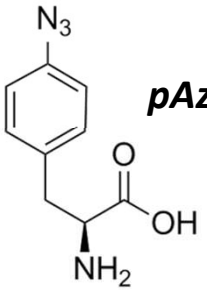


PrK



TAMRA-N_3





pAzF

TAMRA-DBCO

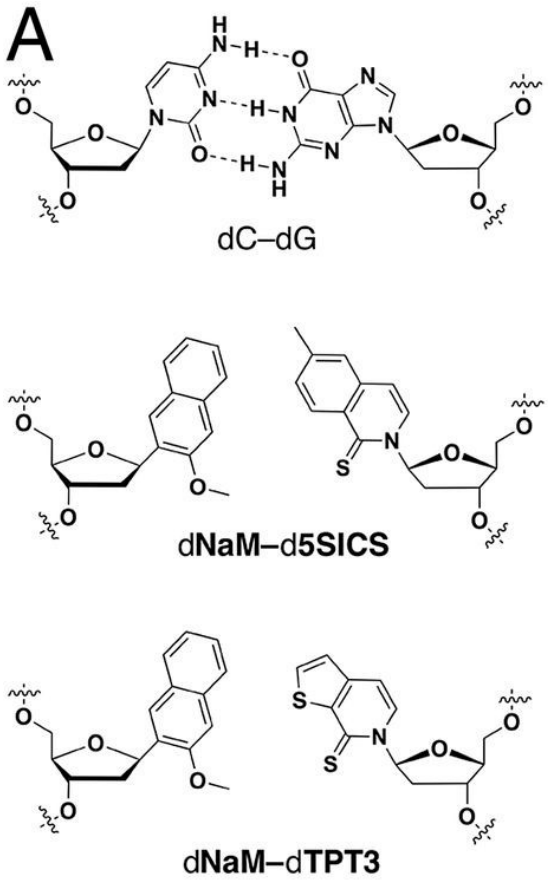
Zhang, Y.; Romesberg, Floyd E. *et al.* *Nature* **2017**, *551*, 644-647

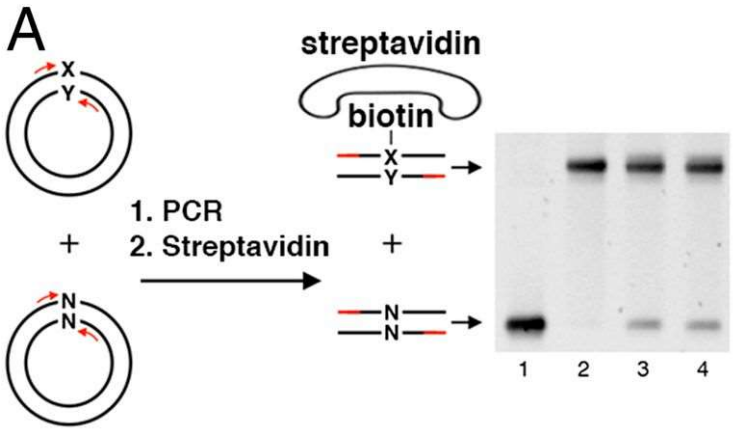
A semi-synthetic organism with an expanded genetic alphabet

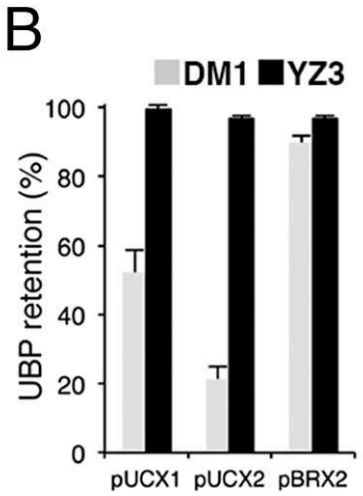
- An unnatural base pair (UBP) would increase the information storage potential of DNA
- and semisynthetic organisms (SSOs) that stably harbor this expanded alphabet would thereby have the potential to store and retrieve increased information,
- *Escherichia coli* grown in the presence of the unnatural nucleoside triphosphates dNaMTP and d5SICSTP, and provided with the means to import them via expression of a plasmid-borne nucleoside triphosphate transporter, replicates DNA containing a single dNaM-d5SICS UBP,
- to fortify and vivify the nascent SSO, a more chemically optimized UBP dTPT3 was used, and the power of the bacterial immune response was harnessed by using Cas9 to eliminate DNA that had lost the UBP.
- The optimized SSO grows robustly, constitutively imports the unnatural triphosphates, and is able to indefinitely retain multiple UBPs in virtually any sequence context. This SSO is thus a form of life that can stably store genetic information using a six-letter, three-base-pair alphabet

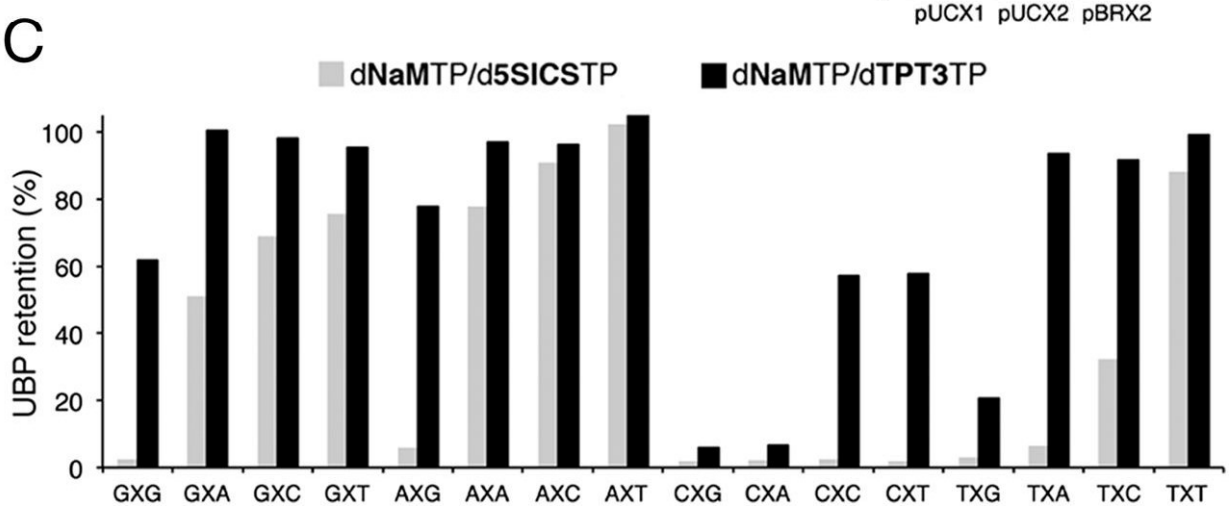
Romesberg, Floyd E. *et al.* *PNAS* **2017**, *114,* 1317-1322

A semi-synthetic organism with an expanded genetic alphabet



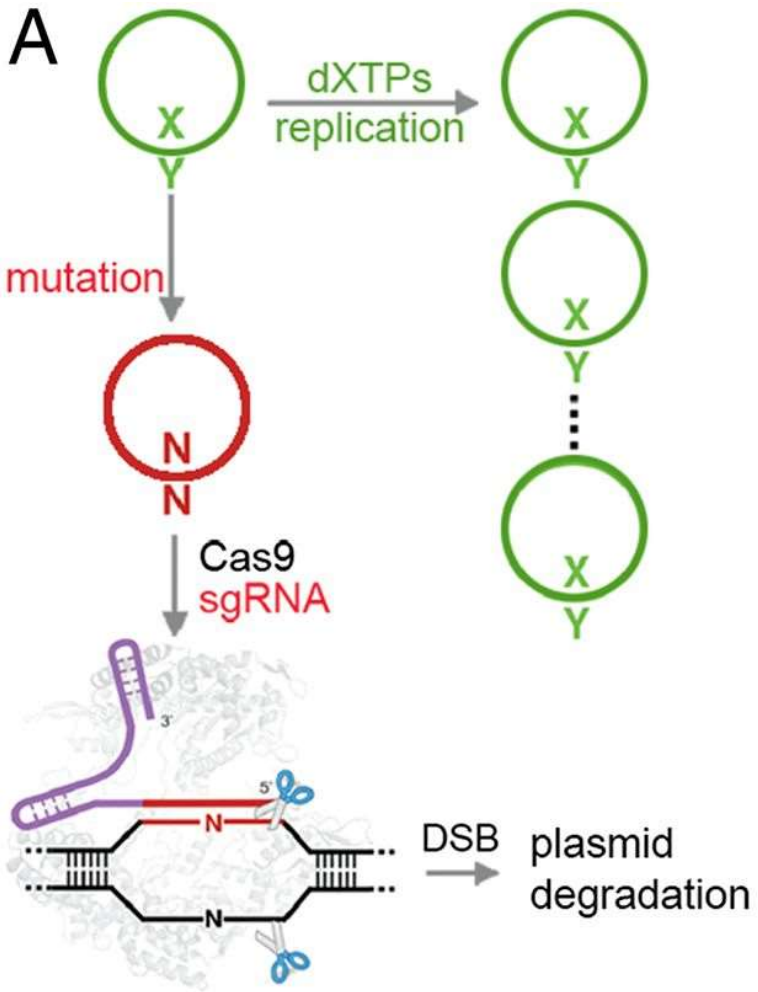


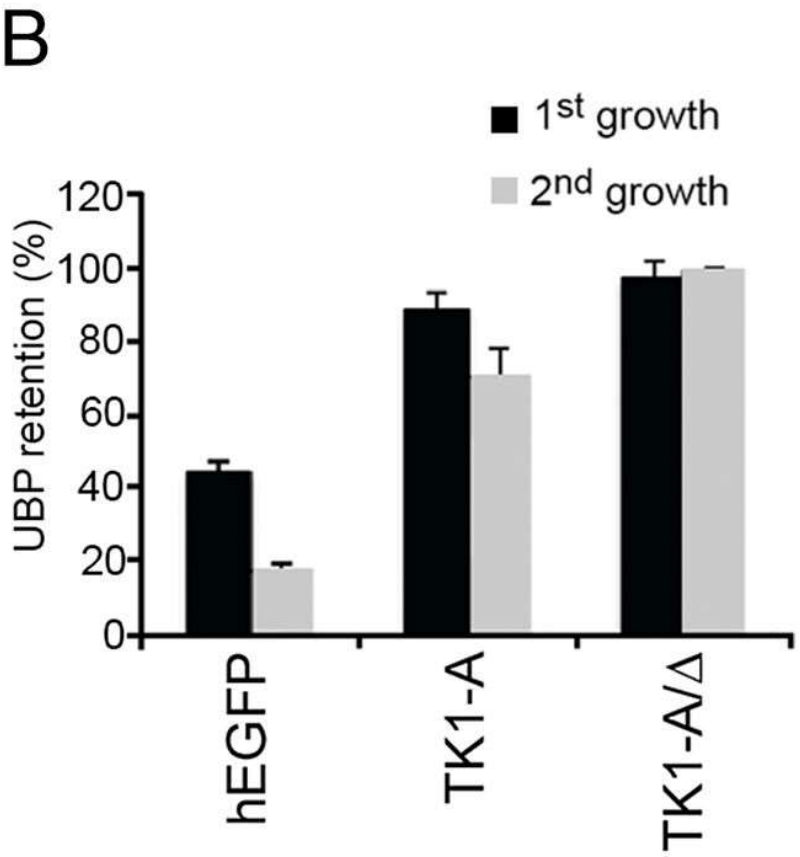




Romesberg, Floyd E. *et al.* *PNAS* **2017**, *114,* 1317-1322

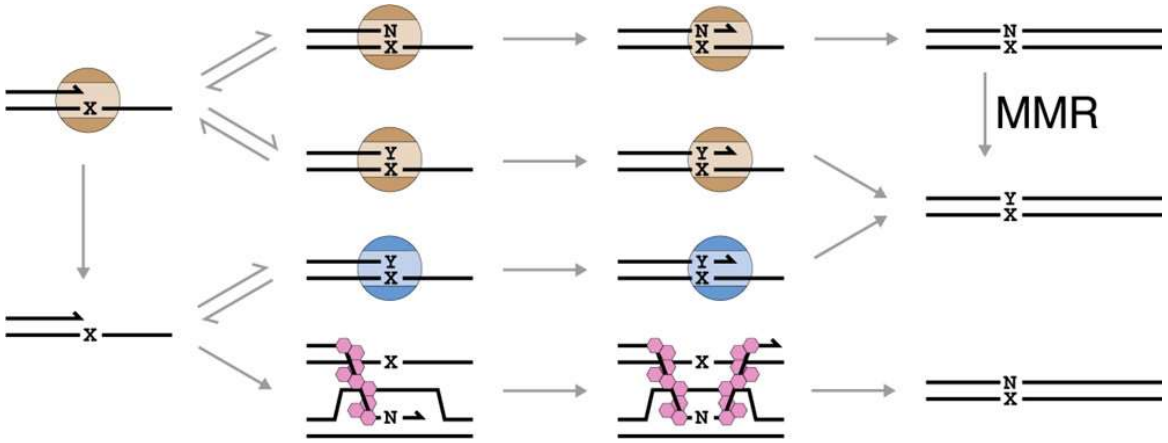
A semi-synthetic organism with an expanded genetic alphabet





Romesberg, Floyd E. *et al.* *PNAS* **2017**, *114,* 1317-1322

Reprograming the replisome of a semi-synthetic organism

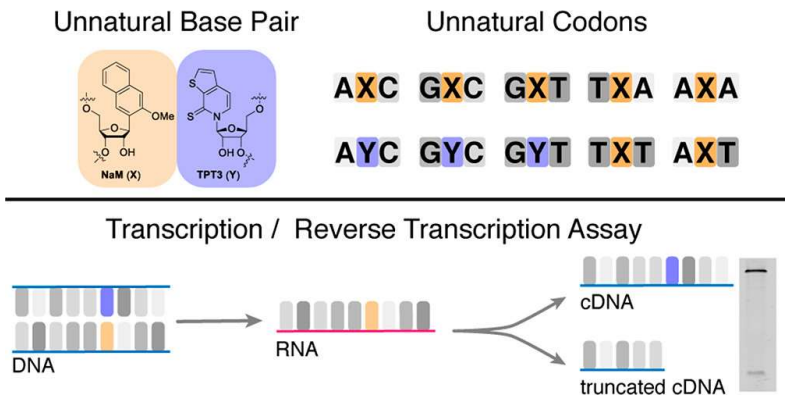


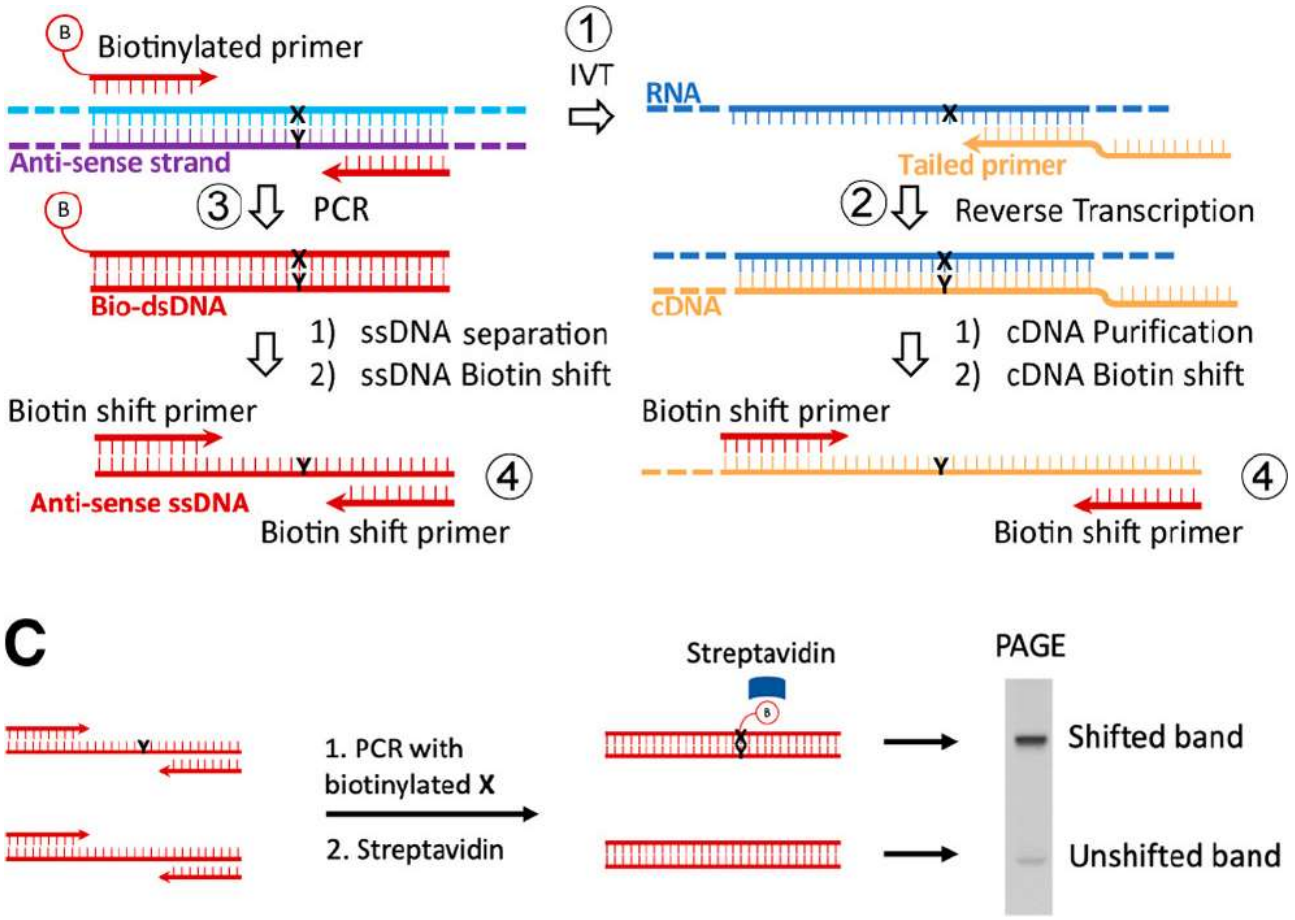
Proposed mechanism of UBP replication. *UBP retention is mediated by the activities of Pol III (tan), Pol II (blue), and MMR. UBP loss is mediated by polymerase replication errors or RecA (pink)-mediated RER.*

UBPs are not well retained in all sequences, limiting the information that can be encoded, and are invariably lost upon extended growth. The contributions of the *E. coli* DNA replication and repair machinery to the propagation of DNA containing d**NaM**-d**TPT3** have been explored and showed that replication by DNA polymerase III, supplemented with the activity of polymerase II and methyl-directed mismatch repair contribute to retention of the UBP and that recombinational repair of stalled forks is responsible for the majority of its loss.

M. P. Ledbetter, R. J. Karadeema, and F. E. Romesberg *J. Am. Chem. Soc.* **2018**, *140, 2,* 758-765

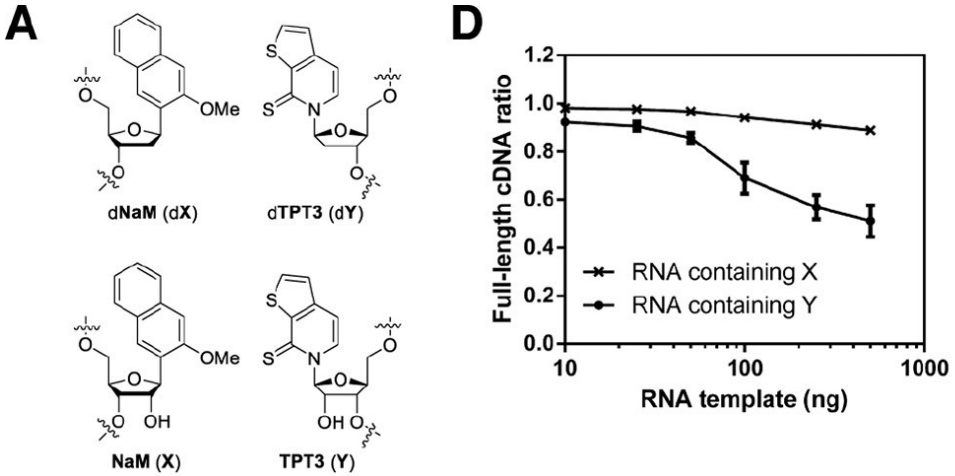
Transcription and reverse transcription in vitro and in SSO

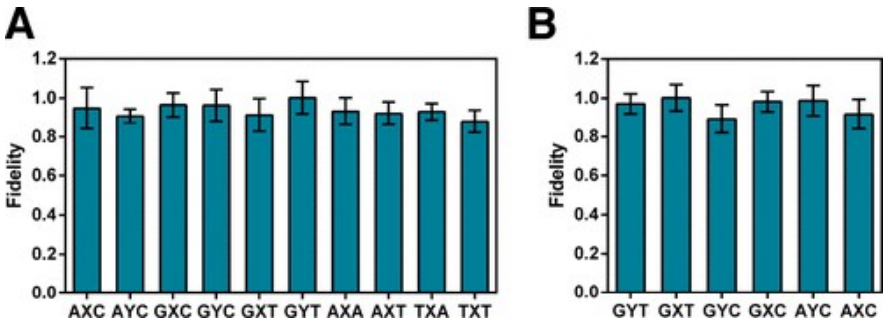




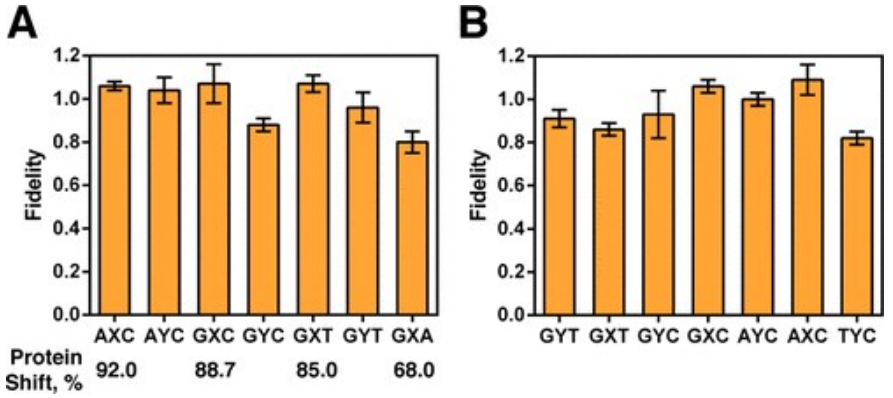
A. X.-Z. Zhou, X. Dong, F. Romesberg,
J. Am. Chem. Soc., **2020**, *142*, 19029-19032

Transcription and reverse transcription in vitro and in SSO





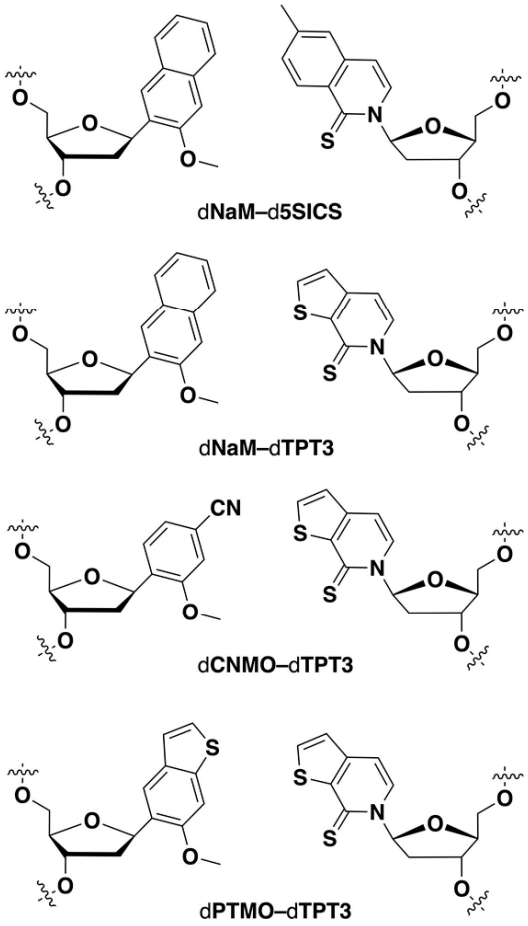
Transcription fidelity measured by T-RT assay of (A) mRNA and (B) tRNA made by in vitro transcription using T7 RNA polymerase.

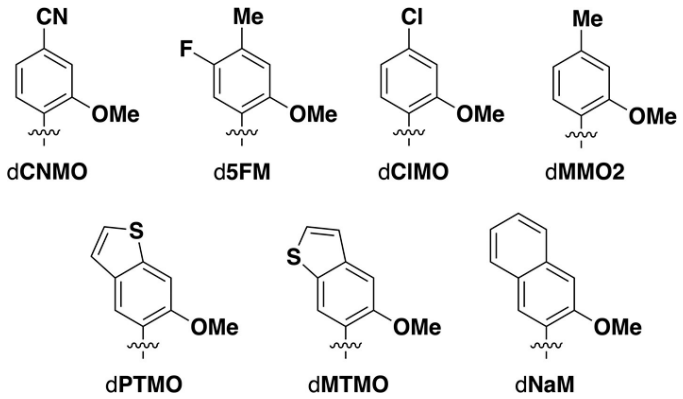


Transcription fidelity measured by T-RT assay of (A) mRNA and (B) tRNA extracted from SSO in vivo translation experiments

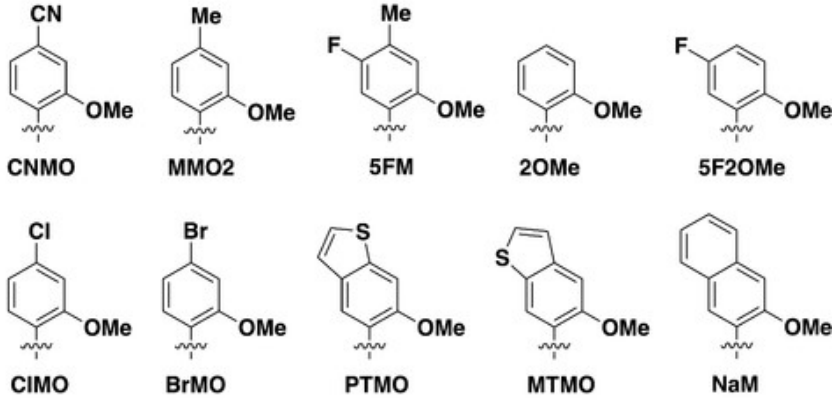
A. X.-Z. Zhou, X. Dong, F. Romesberg,
J. Am. Chem. Soc., **2020**, *142*, 19029-19032

Expanded genetic alphabet - 2019





dXTP analogues



XTP analogues

YTP analogues

Romesberg, Floyd E. *et al.*
J. Am. Chem. Soc. **2019**, *141, 27,* 10644–10653

Unnatural nucleobases - overview