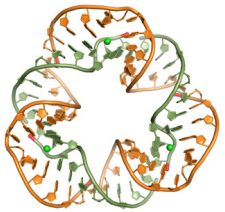
Synthetic life SL3

WiSe 2018/19

Zbigniew Pianowski

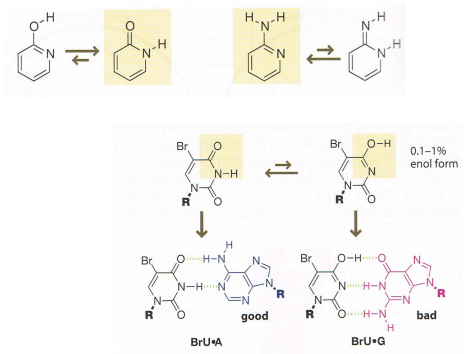
CHAPTER 1



OLIGONUCLEOTIDES

Part 2 – noncanonical nucleobases

Nucleobase modifications for biosynthetic



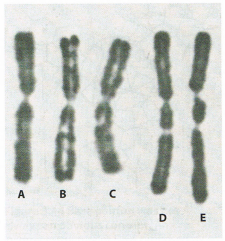
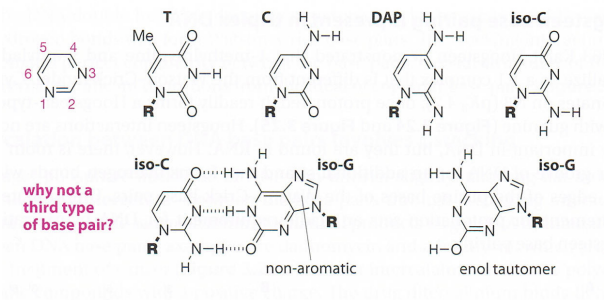


Figure 3.20 Chromosome 1 from hamster cells exposed to bromodeoxyuridine. (A) Normal chromosome. (B–E) Aberrant chromosomes. (From T.C. Hsu and C.E. Somers, *Proc. Natl. Acad. Sci. USA* 47: 396–403, 1961. With permission from the MD Anderson Cancer Center.)

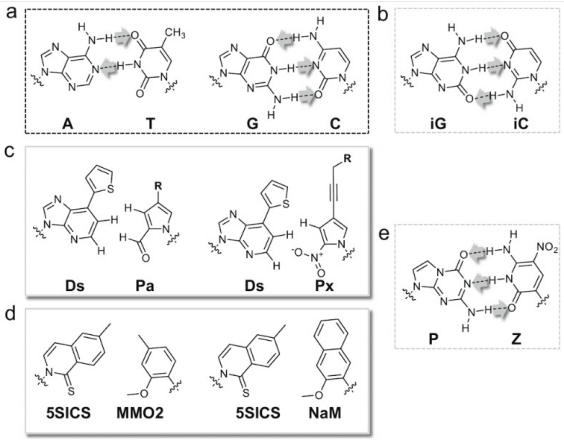
Alternative base pairs – synthetic biology



DAP – one tautomer forms a base pair with guanine

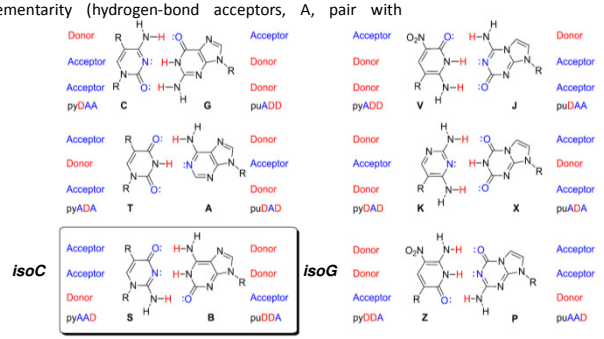
iso-C/iso-G
- specificity (the enol tautomer of iso-G, stabilized by aromatization, complementary to thymine)
- the 2-amino group of iso-C hydrolyses easily to uracil

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



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

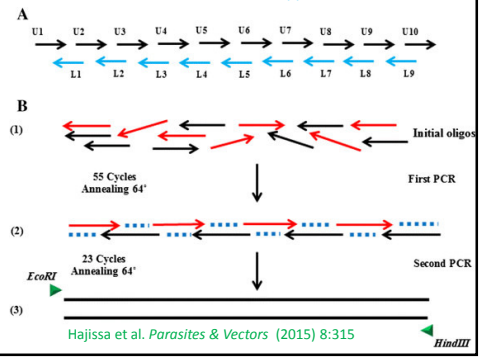
Artificial Gene Synthesis

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:

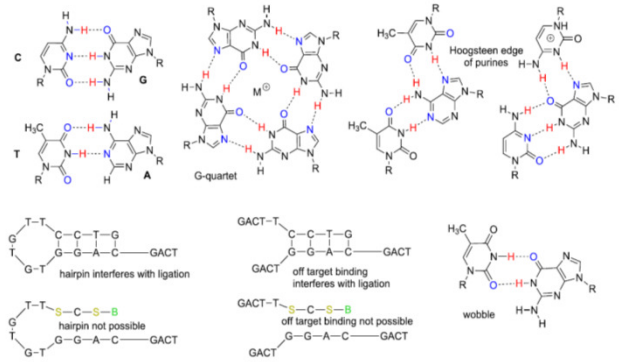
- recombinant DNA technology including heterologous gene expression, vaccine development, gene therapy and molecular engineering.
- The synthesis of nucleic acid sequences can be more economical than classical cloning and mutagenesis procedures
- the ability to safely obtain genes for vaccine research 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

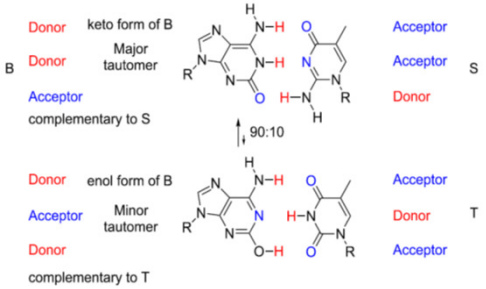


S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

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

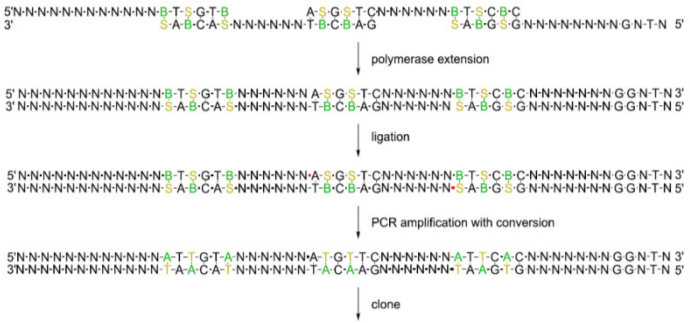
Conversion occurs when polymerases are forced to mismatch a standard nucleotide opposite an AEGIS nucleotide by
(a) not being provided the complementary AEGIS triphosphate and
(b) exploiting a chemical feature of the AEGIS nucleotide that directs a specific mismatch.



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

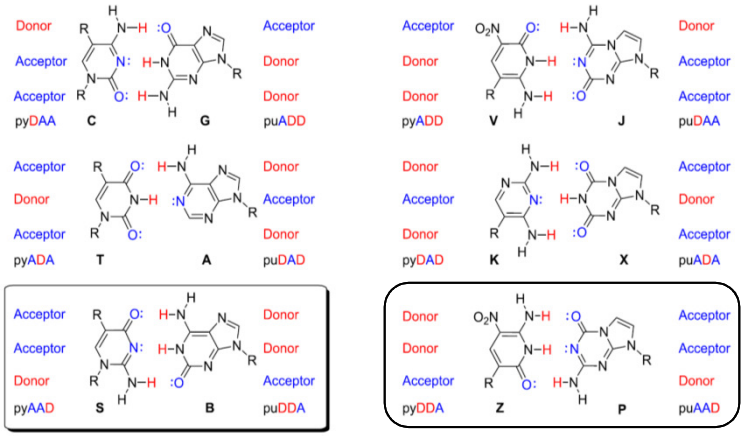
Self-assembly of whole genes and DNA nanostructures



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.

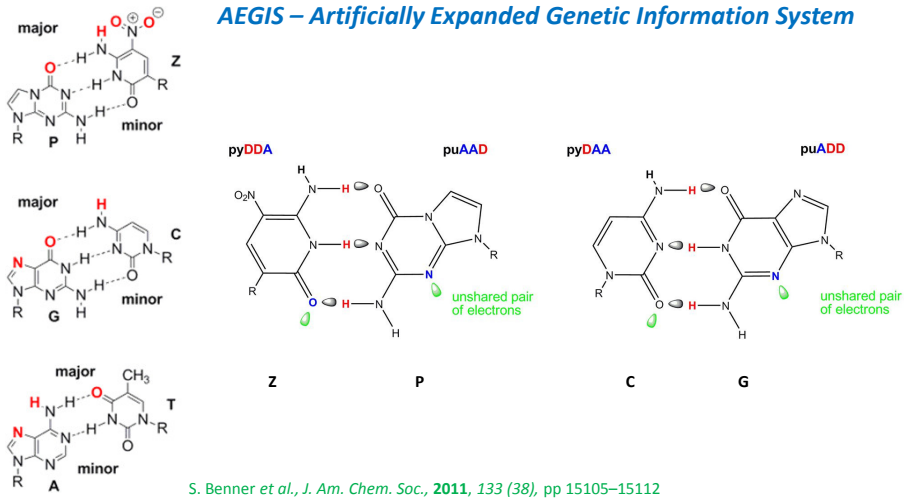
S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

AEGIS – Artificially Expanded Genetic Information System



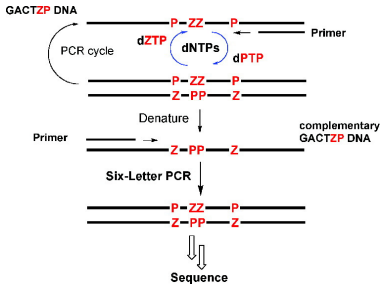
S. Benner *et al., Beilstein J. Org. Chem.* **2014,** *10,* 2348–2360. doi:10.3762/bjoc.10.245

AEGIS – Artificially Expanded Genetic Information System



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

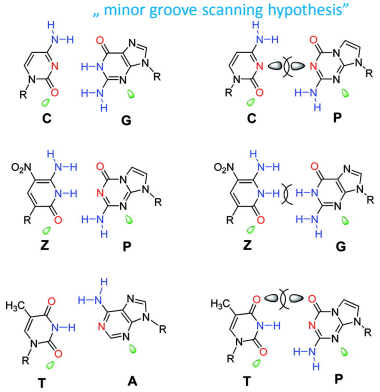
AEGIS – Permanent orthogonal nucleobases surviving PCR



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

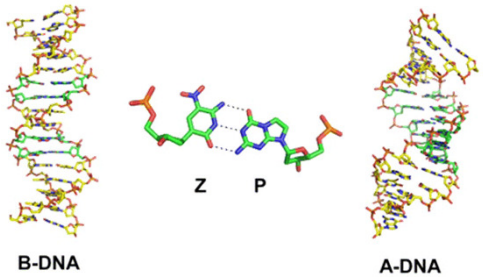
Pol: Deep Vent – 2 Z/P, Taq/Phu – 3-4 Z/P
dZTP (*deprotonated*) at higher pH pairs slightly with G
→ loss of some **Z**, but also gain of some new **Z** mutants.

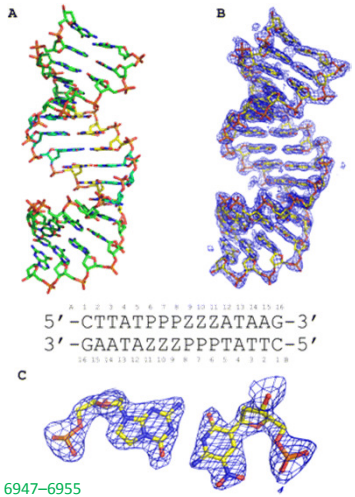
Electron density presented to the minor groove
→ recognition site by polymerases
„ minor groove scanning hypothesis"



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

ACGTZP-DNA crystal structures

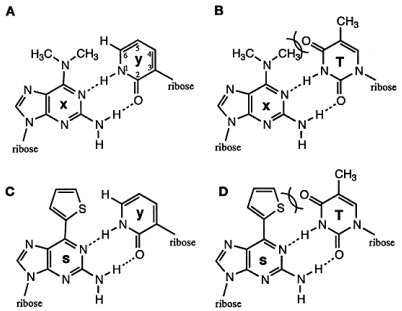




*18-mers: 2+2 **Z:P** pairs → B-DNA*
*6 consecutive **Z:P** → A-DNA*
*0,1 nm wider, but otherwise alike **G:C** pairs*

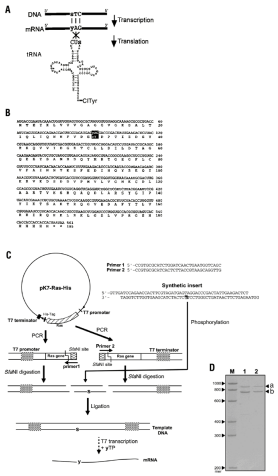
S. Benner *et al., J. Am. Chem. Soc.*, **2015**, *137*, pp 6947–6955

Unnatural aminoacid incorporation using a noncanonical base pair



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



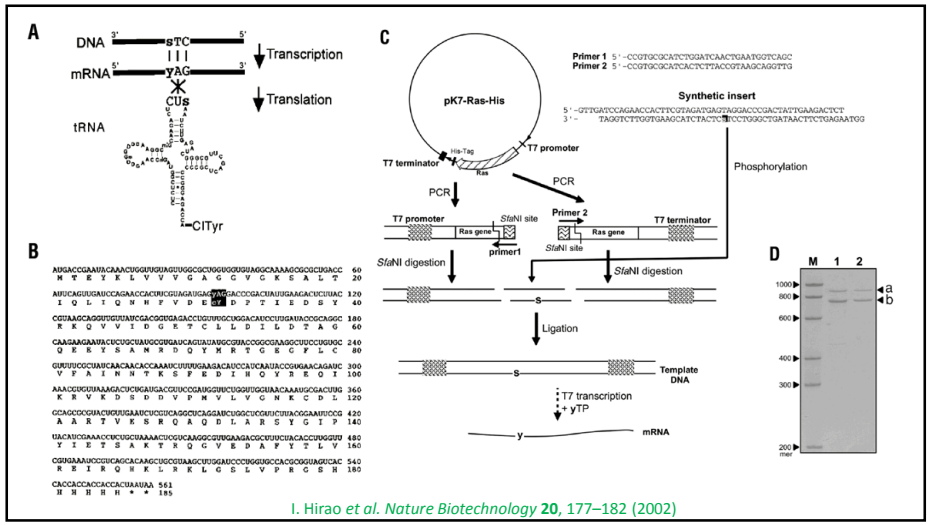
(A) The coupled transcription–translation system using the nonstandard codon–anticodon interaction for the site-specific incorporation of 3-chlorotyrosine into the Ras protein.

(B) The mRNA and amino acid sequences of human c-Ha-Ras replaced with the **y**AG codon for 3-chlorotyrosine at position 32.

(C) Construction of the DNA template containing the **s**TC sequence.

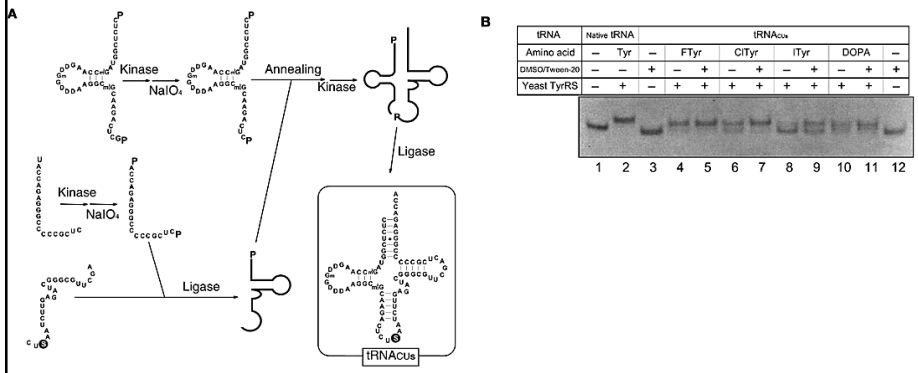
(D) Gel electrophoresis of the T7 transcripts. Lane 1 shows the transcripts from the DNA template containing the **s**TC sequence, and lane 2 shows those from the DNA template without **s** for the wild-type Ras protein. Bands a and b show the transcript terminated at the end of the template DNA (863-mer) and at the T7 terminator (747-mer), respectively.

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



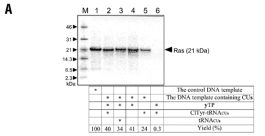
I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)

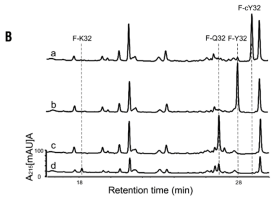
Unnatural aminoacid incorporation using a noncanonical base pair

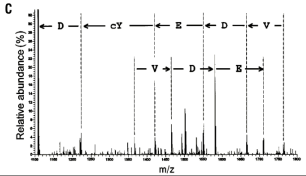


(A) Construction of $tRNA_{CUs}$. (B) Acidic-gel electrophoresis of the products after aminoacylation of *S. cerevisiae* tRNA and $tRNA_{CUs}$ with tyrosine and the 3′-substituted analogs, in the absence or presence of 20% DMSO and 0.25% Tween-20. The upper bands are the aminoacylated tRNAs, and the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

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







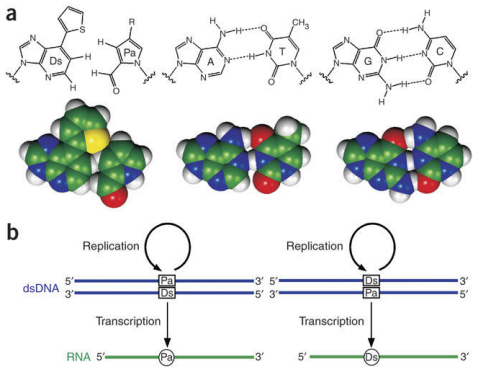
A) Autoradiogram of *in vitro* transcription–translation products labeled with l-[^{14}C]leucine. The conditions are noted at the bottom of each lane.

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

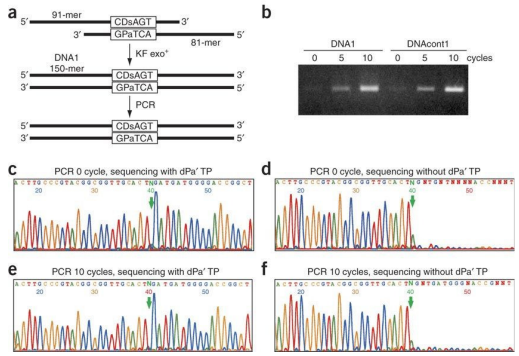
I. Hirao *et al. Nature Biotechnology* **20**, 177–182 (2002)

Unnatural aminoacid incorporation using a noncanonical base pair



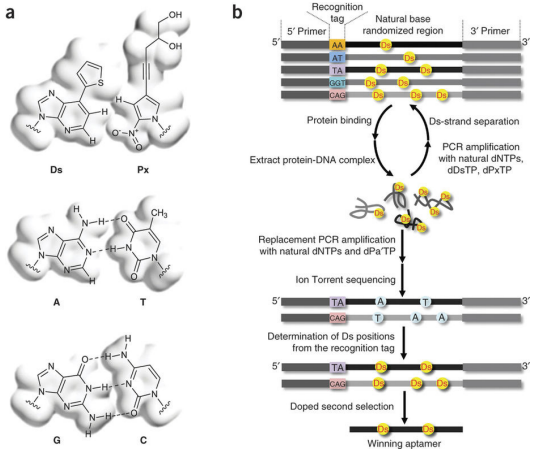
(**a**) The unnatural Ds-Pa and natural base pairs (Pa: R = H, Pa′: R = C≡C–CH_3). (**b**) The unnatural base pair system that functions in PCR amplification, primer extension, DNA sequencing and T7 transcription. Original DNA templates were prepared by chemical synthesis and ligation, and were amplified 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



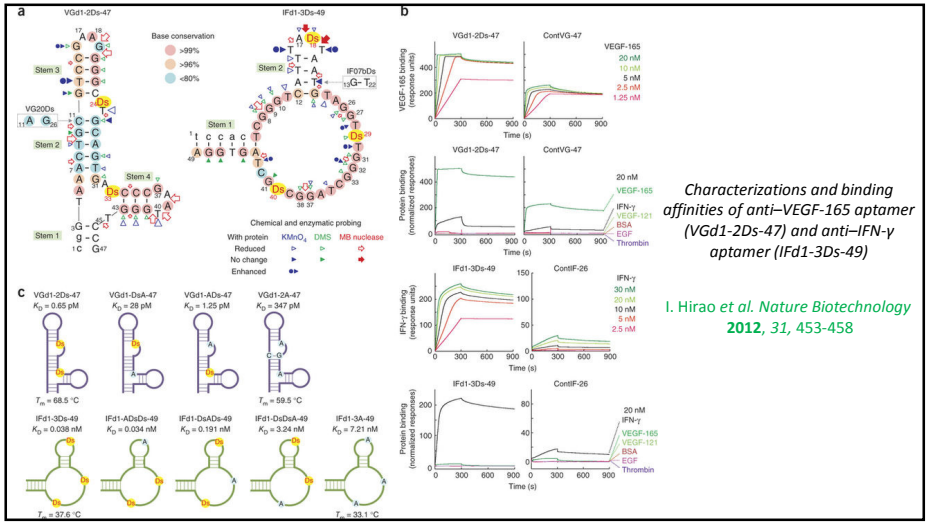
(**a**) The double stranded DNA fragment (150-mer, DNA1) was prepared by primer extension using chemically synthesized DNA fragments (91-mer and 81-mer) containing Ds and Pa. (**b**) Agarose-gel analysis of original DNA fragments (0 cycle) and PCR products after 5 and 10 cycles of amplification. For 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

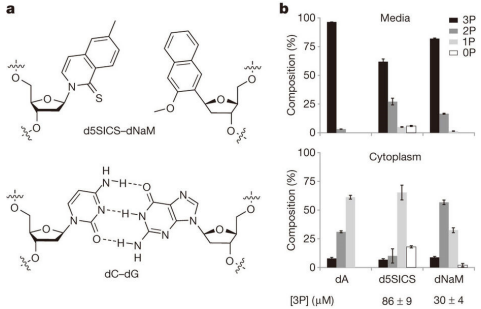
I. Hirao *et al. Nature Biotechnology* **2012**, *31*, 453-458



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

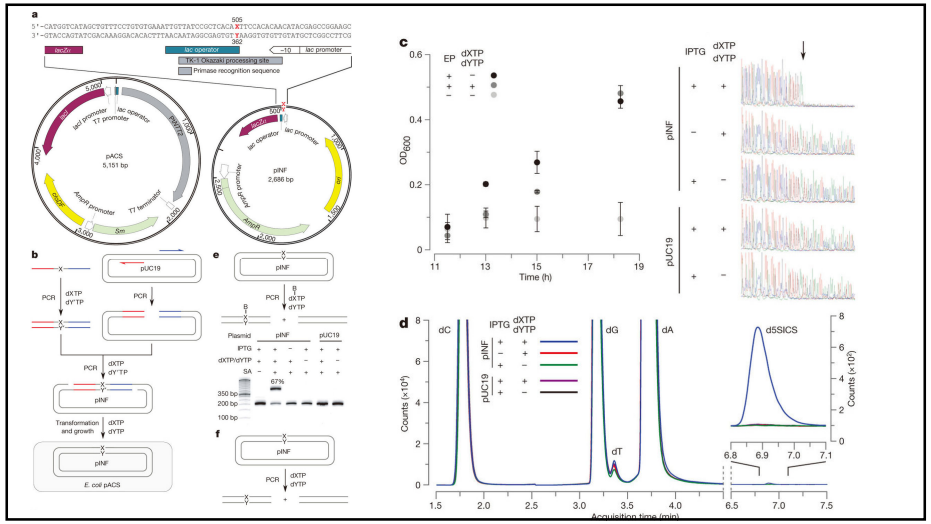
I. Hirao *et al. Nature Biotechnology* **2012**, *31*, 453-458

A semi-synthetic organism with an expanded genetic alphabet

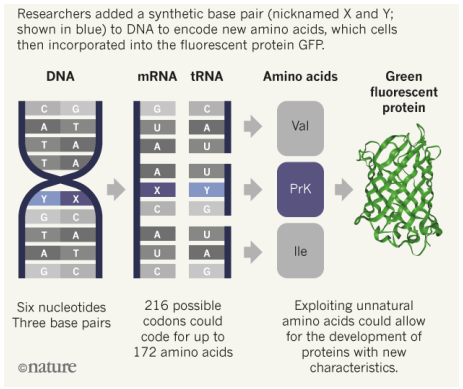


a, Chemical structure of the d5SICS–dNaM UBP compared to the natural dG–dC base pair. **b**, Composition analysis of d5SICS and dNaM in the media (top) and cytoplasmic (bottom) fractions of cells expressing *Pt*NTT2 after 30 min incubation; dA shown for comparison. 3P, 2P, 1P and 0P correspond to triphosphate,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



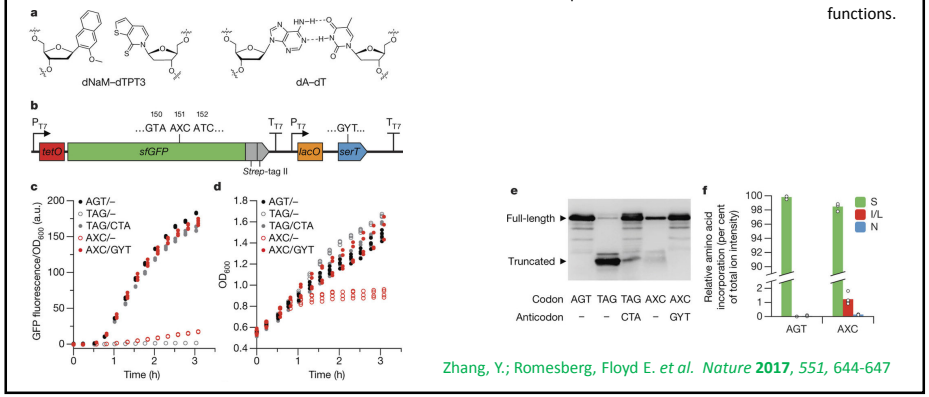
A semi-synthetic organism with an expanded genetic alphabet



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

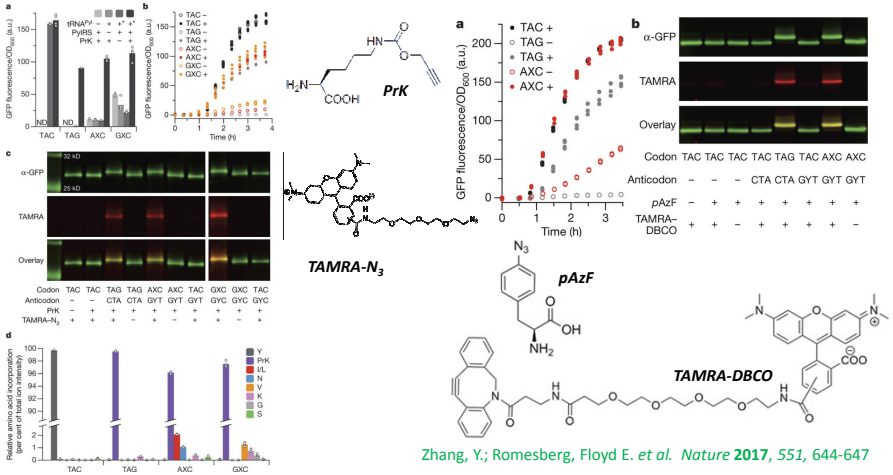
A semi-synthetic organism with an expanded genetic alphabet

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.



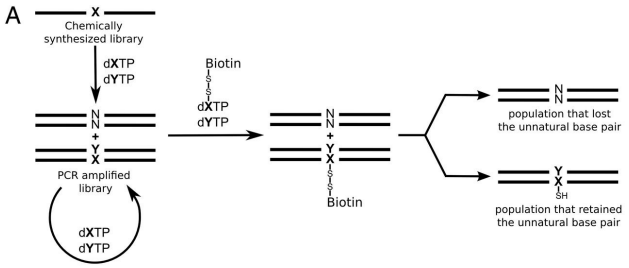
Zhang, Y.; Romesberg, Floyd E. *et al.* *Nature* **2017**, *551*, 644-647

A semi-synthetic organism with an expanded genetic alphabet



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

A semi-synthetic organism with an expanded genetic alphabet



(A) PCR selection scheme. X = NaM (or when biotinylated, its analog MMO2; see Fig. S5) and Y = 5SICS. (B) Library design. The regions proximal to the unnatural base pair that were analyzed for biases are shown in red, and the distal regions used as a control are shown in green. Sublibrary-specific two-nucleotide barcodes that indicate the position of the unnatural base pair flank the randomized regions and are shown in italics. Primer binding regions are denoted as PBR

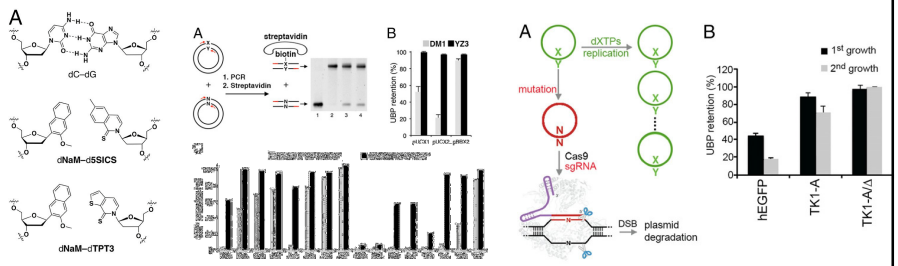
Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *PNAS* **2012**, *109 (30),* 12005-12010

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

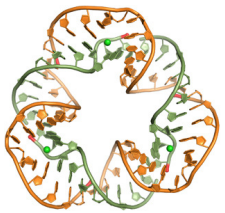
Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *PNAS* **2017**, *114,* 1317-1322

A semi-synthetic organism with an expanded genetic alphabet



Malyshev, Denis A.; Romesberg, Floyd E. *et al.* *PNAS* **2017**, *114,* 1317-1322

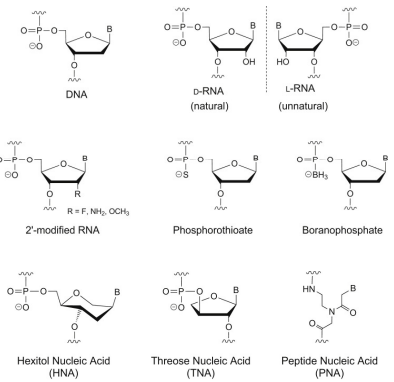
CHAPTER 1



OLIGONUCLEOTIDES

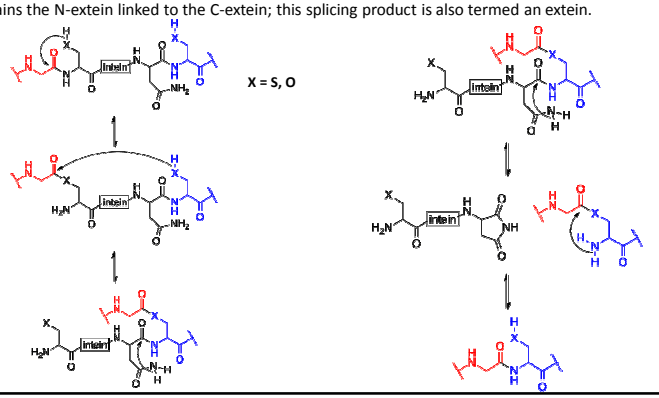
Part 3 – noncanonical backbone

Artificial genetic polymers



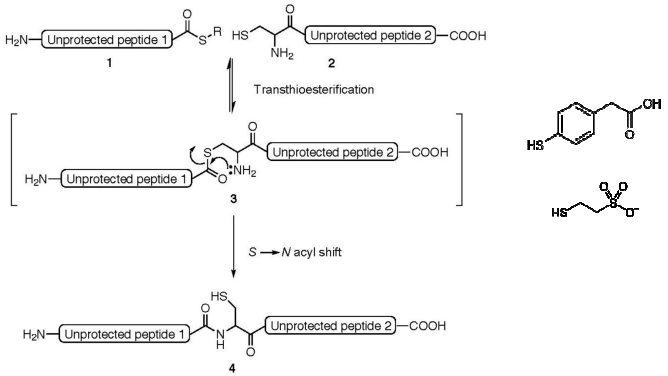
Intein splicing

An **intein** is a segment of a protein that is able to excise itself and join the remaining portions (the **exteins**) with a peptide bond in a process termed protein splicing. Inteins have also been called "protein introns".
Intein-mediated protein splicing occurs after the intein-containing mRNA has been translated into a protein. This precursor protein contains three segments—an **N-extein** followed by the intein followed by a **C-extein**. After splicing has taken place, the resulting protein contains the N-extein linked to the C-extein; this splicing product is also termed an extein.



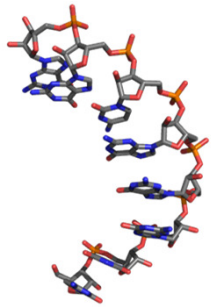
Native chemical ligation

Native chemical ligation or ***NCL*** is an important extension of the chemical ligation field, a concept for constructing a large polypeptide formed by the assembling of two or more unprotected peptides segments. Especially, NCL is the most powerful ligation method for synthesizing proteins (native or modified) of moderate size (i.e., small proteins< 200 AA).



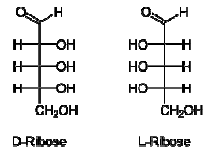
Spiegelmers: L-RNA

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.

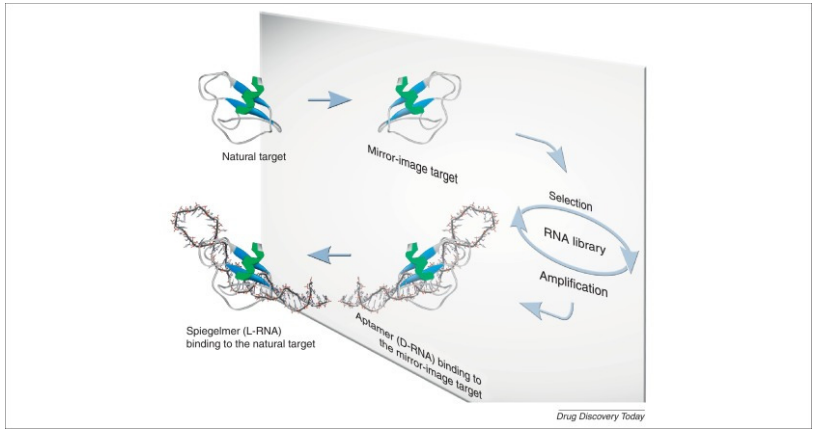


An ***L-ribonucleic acid aptamer*** (L-RNA aptamer, trade name ***Spiegelmer*** – from German Spiegel "mirror" – by Noxxon Pharma) is an RNA-like molecule built from L-ribose units.
It is an artificial oligonucleotide named for being a mirror image of natural oligonucleotides.

L-RNA aptamers are a form of aptamers. Due to their L-nucleotides, they are highly resistant to degradation by nucleases. ***Spiegelmers*** are considered potential drugs and are currently being tested in clinical trials.



A. Vater, S. Klussmann, *Drug Discovery Today* **2015**, *20*, 147-155



A. Vater, S. Klussmann, *Drug Discovery Today* **2015**, *20,* 147-155

Mirror-image RNA that binds D-Adenosine

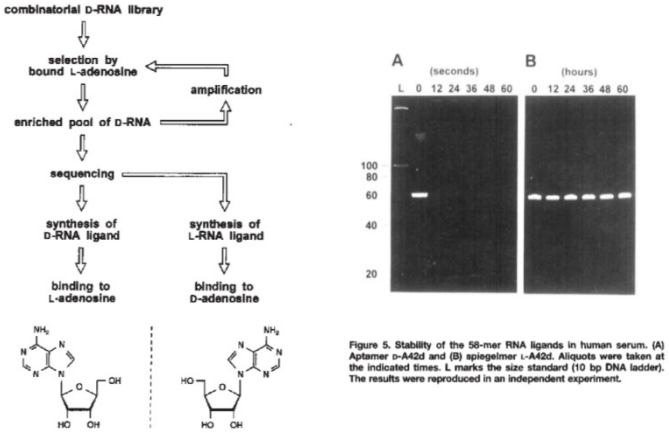
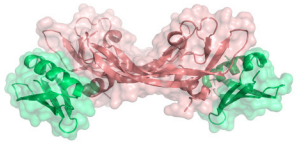


Figure 5. Stability of the 58-mer RNA ligands in human serum. (A) Aptamer D-A42d and (B) spiegelmer L-A42d. Aliquots were taken at the indicated times. L marks the size standard (10 bp DNA ladder). The results were reproduced in an independent experiment.

S. Klussmann, J. Furste, et al. *Nature Biotech.* **1996**, *14,* 1112-1115

D-proteins: almost ideal therapeutic agents

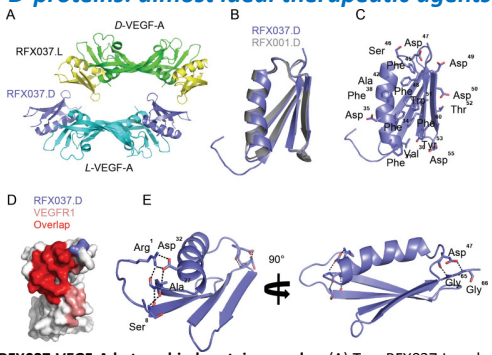
Polypeptides composed entirely of *D*-amino acids and the achiral amino acid glycine (*D* -proteins) inherently have *in vivo* properties that are proposed to be near-optimal for a large molecule therapeutic agent. Specifically, *D* -proteins are resistant to degradation by proteases and are anticipated to be nonimmunogenic. Furthermore, *D* -proteins are manufactured chemically and can be engineered to have other desirable properties, such as improved stability, affinity, and pharmacokinetics.



RFX037.D is a *D*-protein antagonist of natural vascular endothelial growth factor A (VEGF-A) that inhibited binding to its receptor, with extreme thermal stability (T_m > 95 °C) and high affinity for VEGF-A (K_d = 6 nM).
Comparison of the two enantiomeric forms of RFX037 revealed that the *D*-protein is more stable in mouse, monkey, and human plasma and has a longer half-life *in vivo* in mice. Significantly, RFX037.D was nonimmunogenic in mice, whereas the *L*-enantiomer generated a strong immune response. These results confirm the potential utility of synthetic *D*-proteins as alternatives to therapeutic antibodies.

S. Kent *et al.*, ACS Chem. Biol. **2016**, *11,* 1058-1065

D-proteins: almost ideal therapeutic agents



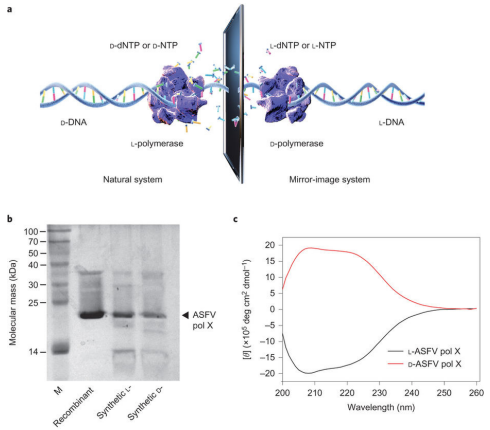
X-ray crystal structure of RFX037:VEGF-A heterochiral protein complex. (A) Two RFX037.L molecules (yellow) bound to one d-VEGF-A homodimer (green) and two RFX037.D molecules (blue) bound to one l-VEGF-A homodimer (cyan). (B) Superposition of RFX037.D (blue) and RFX001.D (gray, rcsb accession 4GLS). (C) RFX037.D side chains (shown as sticks) that contact l-VEGF-A. (D) The contact surfaces of l-VEGF-A to RFX037.D (blue), VEGFR1 (salmon), or both (red). (E) Hydrogen bond networks formed by intramolecular polar contacts originated from additional N- and C-terminal residues in RFX037.D.

S. Kent *et al.*, ACS Chem. Biol. **2016**, *11,* 1058-1065

Spiegelmers: L-DNA polymerase

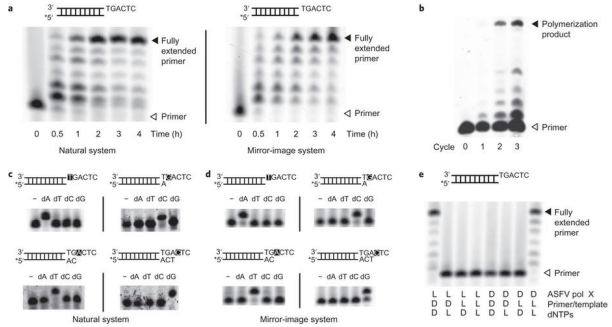
The mirror image configuration of polymerase X from African swine fever virus, the shortest known polymerase (174 amino acids), has recently been demonstrated to elongate an *L*-DNA primer with *L*-dNTPs; and a functional 56-mer *L*-DNAzyme was made within 36 hours.

This poses an important proof of concept, however, polymerase X is a thermo-labile repair enzyme and its catalytic activity does not meet the requirements for a standard PCR



Z. Wang, W. Xu, L. Liu, T. F. Zhu *Nature Chem.* **2016**, *8*, 698-704

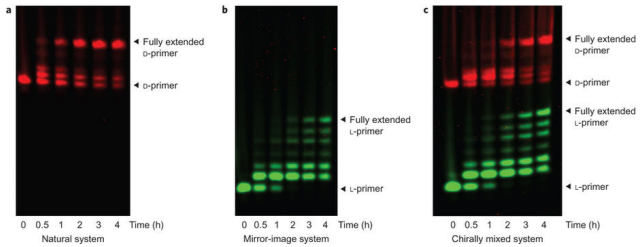
Spiegelmers: L-DNA polymerase



a, Template-directed primer extension by synthetic L-ASFV pol X (natural system) and D-ASFV pol X (mirror-image system) with the corresponding D- and L-DNA primers, templates and dNTPs. **b**, Repeated cycles of polymerization by D-ASFV pol X: **c**,**d**, The nucleotide substrate specificities of synthetic L- and D-ASFV pol X. **e**, Chiral specificity assay with different chiral combinations of polymerases, primer/template pairs and dNTPs.

Z. Wang, W. Xu, L. Liu, T. F. Zhu *Nature Chem.* **2016**, *8*, 698-704

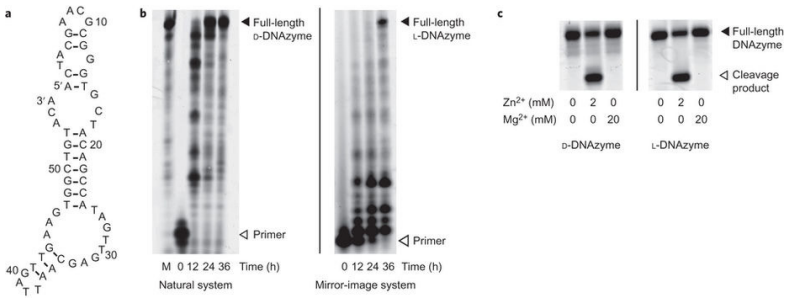
Spiegelmers: L-DNA polymerase



a,**b**, Primer extension by synthetic L- and D-ASFV pol X with the corresponding D-DNA primer (5′-Cy5 labelled) and L-DNA primer (5′-FAM labelled), templates and dNTPs. **c**, The above two polymerization reactions were carried out in a racemic mixture under the same conditions as described above, with the L- and D-ASFV pol X, D- and L-primers, D- and L-templates and D- and L-dNTPs added, incubated for up to 4 h at 37 °C.

Z. Wang, W. Xu, L. Liu, T. F. Zhu *Nature Chem.* **2016**, *8*, 698-704

Spiegelmers: L-DNAzyme

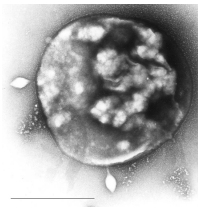


a, Sequence and predicted secondary structure of the previously reported Zn^{2+}-dependent self-cleaving DNAzyme. **b**, Primer extension on a 66 nt template to produce the Zn^{2+}-dependent self-cleaving DNAzyme. **c**, Self-cleavage of the enzymatically polymerized Zn^{2+}-dependent D- and L-DNAzymes.

Z. Wang, W. Xu, L. Liu, T. F. Zhu *Nature Chem.* **2016**, *8*, 698-704

Spiegelmers: A thermostable D-polymerase

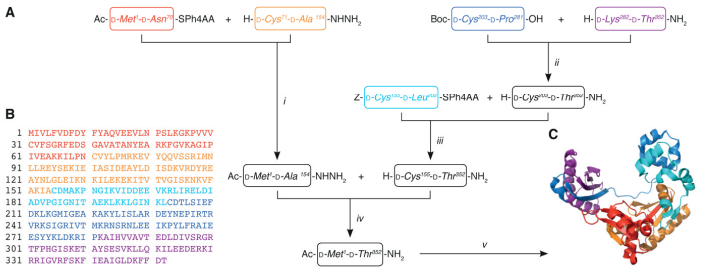
A thermostable mirror-image polymerase ***D*-Dpo4-3C** has been produced, that is able to amplify L-DNA in a classical PCR reaction and can even be used to assemble an L-DNA gene from L-DNA oligonucleotides. This artificial enzyme is a mutant of DNA polymerase IV from *Sulfolobus solfataricus*, a Y-family polymerase consisting of 352 amino acids, the longest protein made by chemical synthesis thus far.



Cell of *Sulfolobus* infected by virus STSV1 observed under microscopy. Two spindle-shaped viruses were being released from the host cell.

Furthermore, with an additional single point mutation (Tyr12Ala or Tyr12Ser), this DNA polymerase can be tuned to accept also ribonucleotides as substrates with reasonable efficiency. Thus, this enzyme may be hijacked to act as a DNA-dependent RNA polymerase to prepare longer stretches of *L*-RNA

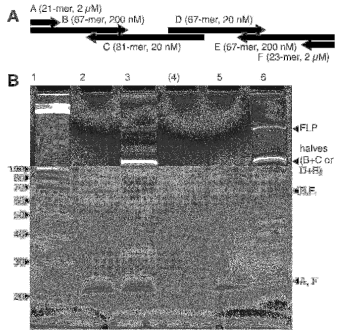
Spiegelmers: A thermostable D-polymerase



Synthesis strategy for d-Dpo4-3C. (**A**) five fragments were synthesized and assembled as follows: (i) native chemical ligation (NCL) of fragments 1 and 2. Isolated yield ≈ 18%. (ii) Segment condensation of fully protected fragments 4 and 5 followed by deprotection. Isolated yield ≈ 15%. (iii) NCL of fragments 3 and 4•5 followed by Z-deprotection. Isolated yield ≈ 25%. (iv) Thioester-conversion of fragment 1•2 and NCL with fragment 3•4•5. Isolated yield: 10%. (v) Folding. (**B**) sequence of d-Dpo4-3C; coloring as in panel A. (**C**) folded d-Dpo4-3C (artist impression based on PDB 3PR4 (31)).

S. Klussmann *Nucl. Acid Res.* **2017**, *45*, 3997-4005

Spiegelmers: A thermostable D-polymerase



Assembly of a mirror-image gene. (**A**) schematic of the oligonucleotide setup. (**B**) lane 1, 3 µl of 10 bp DNA ladder. Lane 2, mirror-image no-enzyme control. Lane 3, mirror-image gene assembly. Lane 4, empty. Lane 5, natural handedness no enzyme control. Lane 6, natural handedness gene assembly.

S. Klussmann *Nucl. Acid Res.* **2017**, *45*, 3997-4005

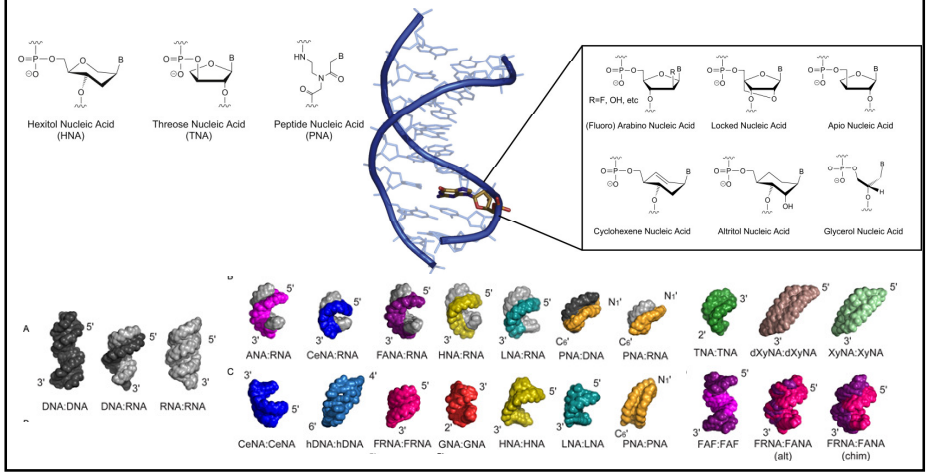
XNA – Xeno Nucleic Acids

XNA - synthetic alternative to DNA and RNA as information-storing biopolymers that differs in the sugar backbone.
- at least 6 XNAs can store and retrieve genetic information
- Ongoing research to create synthetic polymerases to transform XNA →

Xenobiology
- (XNA) as information carriers, expanded genetic code and, incorporation of non-proteinogenic amino acids into proteins
- the ***origin of life:*** *Primoridal soup → (XNA →) RNA → RNA(+DNA)+Proteins*
- development of industrial production systems with novel capabilities (pathogen resistance, biopolymer engineering)
- „genetic firewall" – excludes the risk of contaminating currently existing organisms (horizontal gene transfer)

The ***long-term goal*** - a cell that stores its genetic information on XNA, with different base pairs, using non-canonical amino acids and an altered genetic code.
So far cells have been constructed that incorporate only one or two of these features

XNA – Xeno Nucleic Acids



Synthetic genetic polymers capable of heredity and evolution

XNA are not recognized by natural polymerases.

One of the major challenges is to find or create novel types of polymerases that will be able to replicate these new-to-nature constructs. The method of polymerase evolution and design successfully led to the storage and recovery of genetic information (of less than 100bp length) from six alternative genetic polymers based on simple nucleic acid architectures not found in nature.

XNA aptamers, which bind their targets with high affinity and specificity, were also selected, demonstrating that beyond ***heredity***, specific XNAs have the capacity for ***Darwinian evolution*** and ***folding into defined structures***.

Thus, heredity and evolution, two hallmarks of life, are not limited to DNA and RNA but are likely to be emergent properties of polymers capable of information storage.

P. Herdewijn, P. Holliger, *et al. Science* **2012**, *336*, 341-344

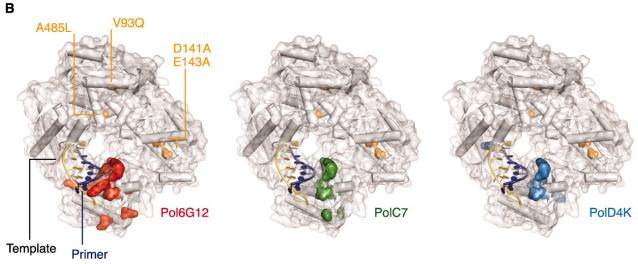
Engineering XNA polymerases

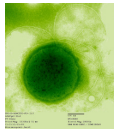
TgoT, a variant of the replicative polymerase of *Thermococcus gorgonarius*

A

```
                402  404  586  590  608   611  653                                  682  703          710  729  731
TgoT        YLD .. FVT .. LEIV .. YEVPPEKLVIYEQITRDLKDYKATGPHVAV .. VLKGSGRI .. AEY
Pol6G12     YLD .. FAT .. LKMV .. YEVPPEQLVIYQPITKQLHDYRARGPHVSV .. VPKGSGRI .. AGY
PolC7       YLD .. FVT .. LEIV .. YQVPPQQLAIYQPITRALQDYKAKGPHVAV .. VLKGSGKI .. AEY
PolD4K      YPD .. FVT .. LEIV .. YEVPTQHLVIHKQITRALNDYKAIGPHVAV .. VLKGSGRI .. AEY
```

B



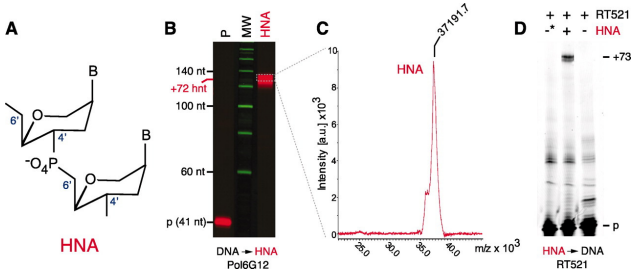


Thermococcus gorgonarius
(Angels Tapias)

(A) Sequence alignments showing mutations from wtTgo in polymerases Pol6G12 (red), PolC7 (green), and PolD4K (blue).
(B) Mutations are mapped on the structure of Pfu (PDB: 4AIL).
Yellow - template; dark blue - primer; orange - mutations present in the parent polymerase TgoT

P. Herdewijn, P. Holliger, *et al. Science* **2012**, *336*, 341-344

HNA synthesis



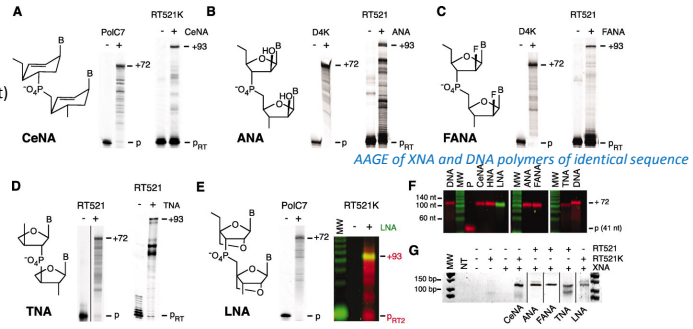
Pol6G12 extends the primer (p) incorporating 72 hNTPs against template T1 to generate a full-length hybrid molecule with a 37,215-dalton expected molecular mass.

HNA reverse transcription (DNA synthesis from an HNA template). Polymerase-synthesized HNA (from template YtHNA4) is used as template by RT521 for HNA-RT

P. Herdewijn, P. Holliger, *et al. Science* **2012**, *336*, 341-344

XNA genetic polymers.

Structures and PAGE of synthesis (+72 xnt), and reverse transcription (+93 nt)



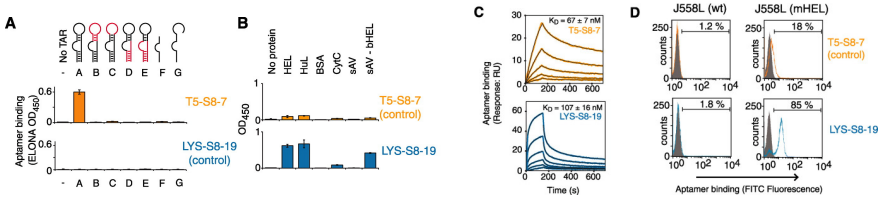
AAGE of XNA and DNA polymers of identical sequence

(E) PAGE of LNA synthesis [primer (41 nt) + 72 lnt] and LNA-RT (red). LNA synthesis (green) migrates at its expected size (113 nt) and comigrates with reverse transcribed DNA (red) synthesized from primer PRT2 (20 nt).

XNART–polymerase chain reaction. Amplification products of expected size (133 base pairs) are obtained only with both XNA forward synthesis and RT (RT521 or RT521K)

P. Herdewijn, P. Holliger, *et al. Science* **2012**, *336*, 341-344

HNA aptamers



Characterization of HNA aptamers. Anti-TAR aptamer T5-S8-7 and anti-HEL aptamer LYS-S8-19.

(A and B) Aptamer binding specificity against TAR variants (red, sequence randomized but with base-pairing patterns maintained) and different protein antigens (human lysozyme, HuL; cytochrome C, CytC; streptavidin, sAV; biotinylated-HEL bound to streptavidin, sAV-bHEL). OD, optical density.
(C) Affinity measurements of aptamer binding by SPR. RU, response units.
(D) FACS analysis of fluorescein isothiocyanate (FITC)–labeled aptamers binding to plasmacytoma line J558L with and without expression of membrane-bound HEL (mHEL). wt, wild type.

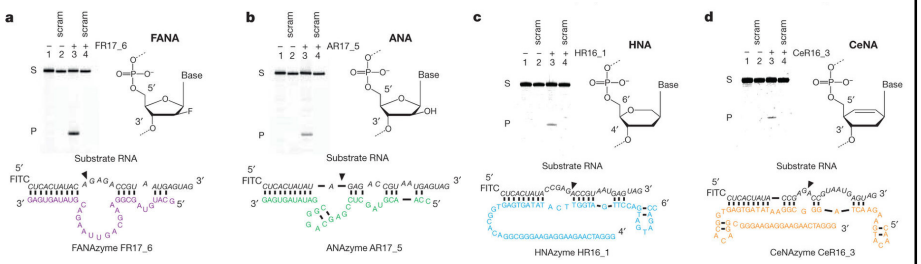
P. Herdewijn, P. Holliger, *et al. Science* **2012**, *336*, 341-344

XNA – Xeno Nucleic Acids

XNA – complementarity to DNA, also used as genetic catalysts.

FANA, HNA, CeNA and ANA - cleave RNA (***XNAzymes***).

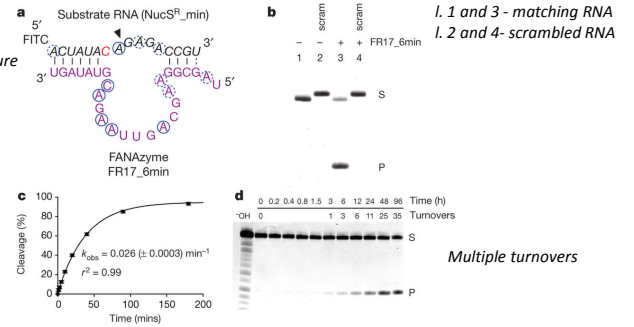
FANA XNAzymes can also ligate DNA, RNA and XNA substrates.



P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518*, 427-430

Chemical synthesis yields an active RNA endonuclease XNAzyme

Secondary FANAzyme structure



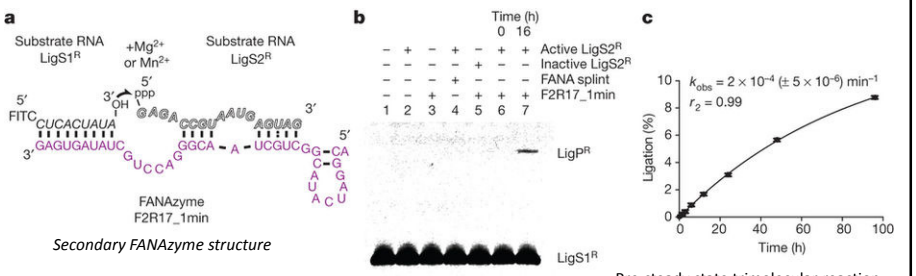
l. 1 and 3 - matching RNA
l. 2 and 4- scrambled RNA

Multiple turnovers

P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518*, 427-430

An RNA ligase XNAzyme (FANA)

FANA XNAzymes can also ligate DNA, RNA and XNA substrates.

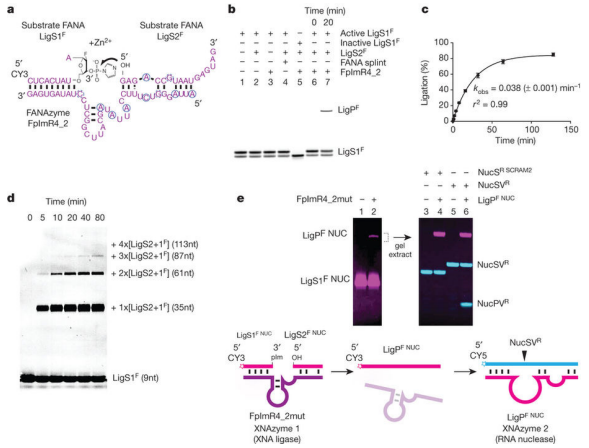


Secondary FANAzyme structure

Pre-steady state trimolecular reaction rate (k_{obs}) at 25 °C (n = 3; error bars, s.d.).

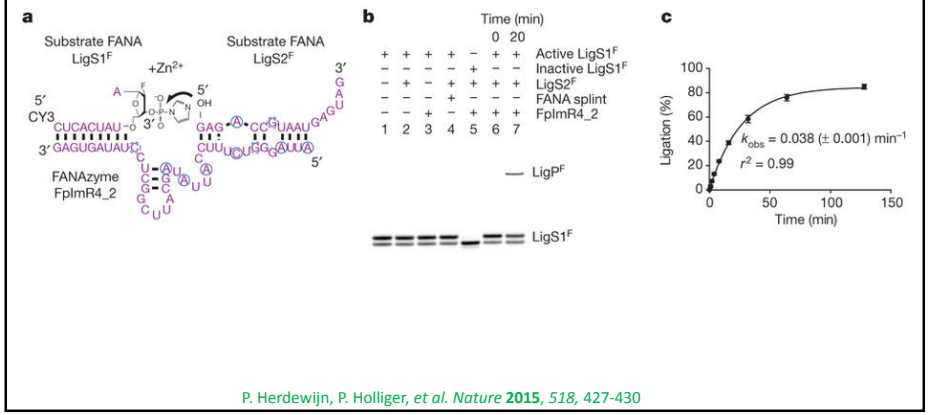
P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518,* 427-430

XNA–XNA ligase XNAzyme (FANA): catalysis without natural nucleic acids



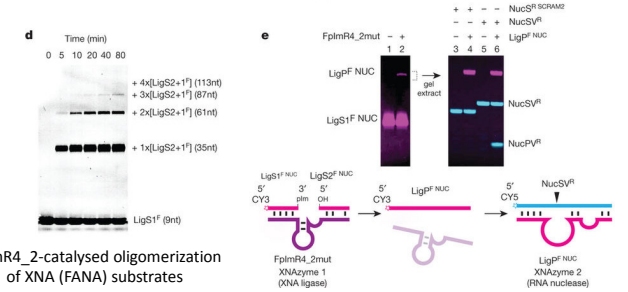
P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518,* 427-430

XNA–XNA ligase XNAzyme (FANA): catalysis without natural nucleic acids



P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518,* 427-430

XNA–XNA ligase XNAzyme (FANA): catalysis without natural nucleic acids



FplmR4_2-catalysed oligomerization of XNA (FANA) substrates

XNAzyme-catalysed assembly of an active XNAzyme. A variant XNA ligase (FplmR4_2mut) catalyses ligation (lane 2) of FANA substrates LigS1F NUC and LigS2F NUC. The product (LigPF NUC) is a variant of XNAzyme FR17_6 min (Fig. 2), which cleaves RNA substrate NucSVR (lanes 5 and 6), but not scrambled RNA (NucSR SCRAM2)(lanes 3 and 4).

P. Herdewijn, P. Holliger, *et al. Nature* **2015**, *518,* 427-430