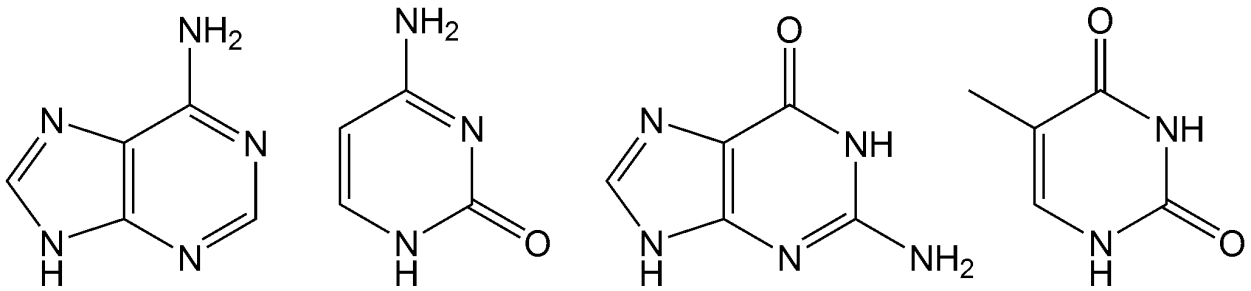
Bacteriophages with a ZTGC DNA alphabet a proof of concept for the synthetic biology of non-canonical nucleic acids

Diversification of nucleic acids Introduction.

All living organisms store genetic information in a four-letter code A, T, G, C

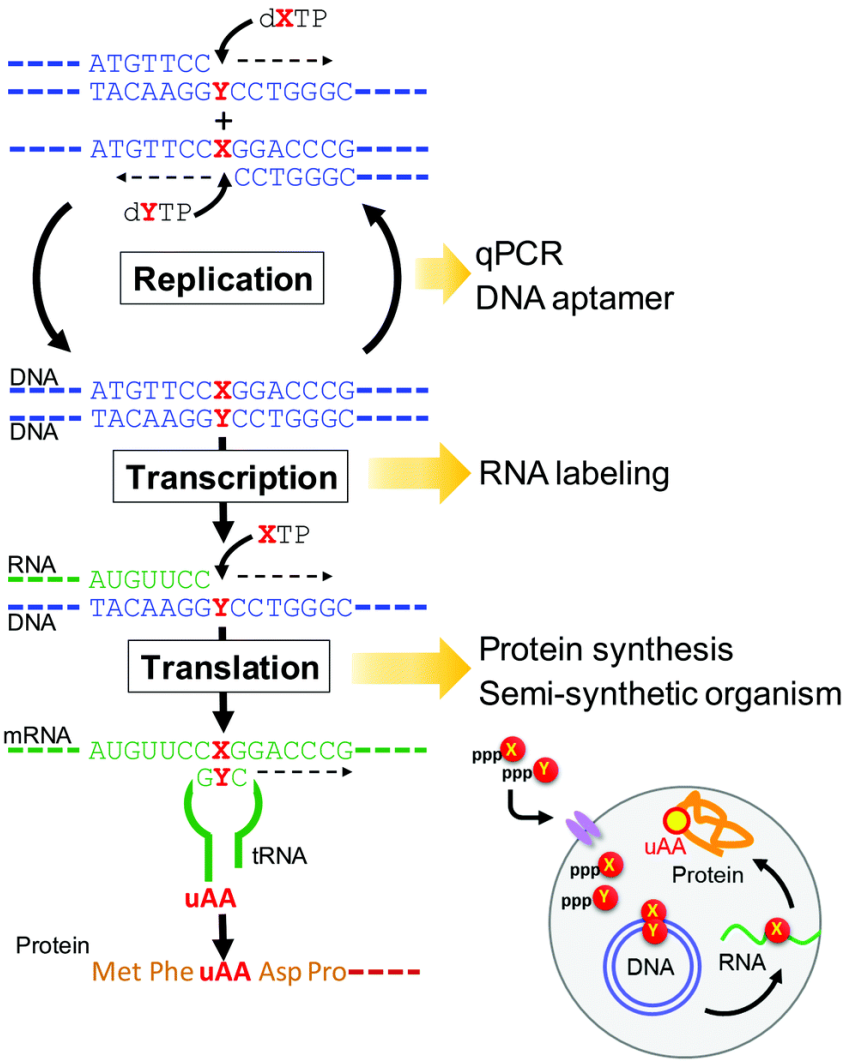


why nature chose the structure type of (deoxy)ribofuranosyl nucleic acids, rather than some other family of molecular structures, as the molecular basis of life's genetic system.

For several decades, chemists attempted to understand

- how prebiotic building blocks are formed
- if alternatives may exist

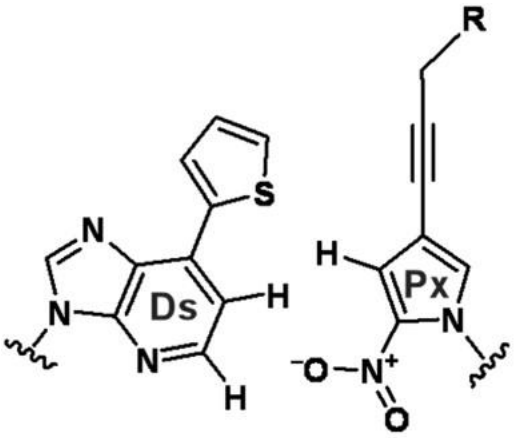
Expansion of the genetic alphabet and code by creating an unnatural base pair as a third pair

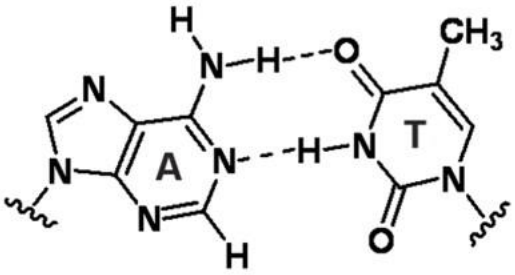


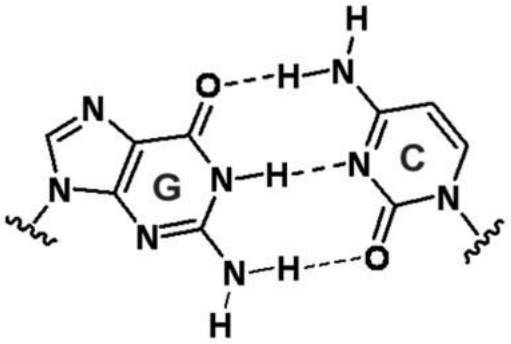
Genetic alphabet expansion technology by creating unnatural base pairs.
Kimoto M , **Hirao I .** Chem Soc Rev. 2020 Nov 7;49(21):7602-7626.

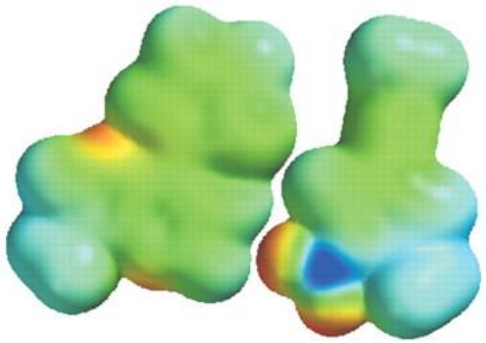
Expansion of the genetic alphabet 1

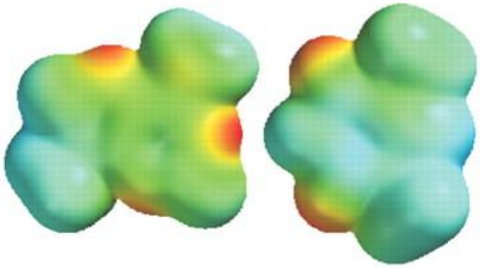
An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules

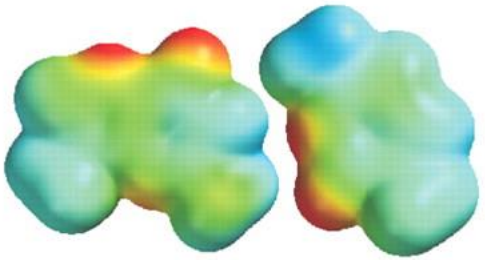










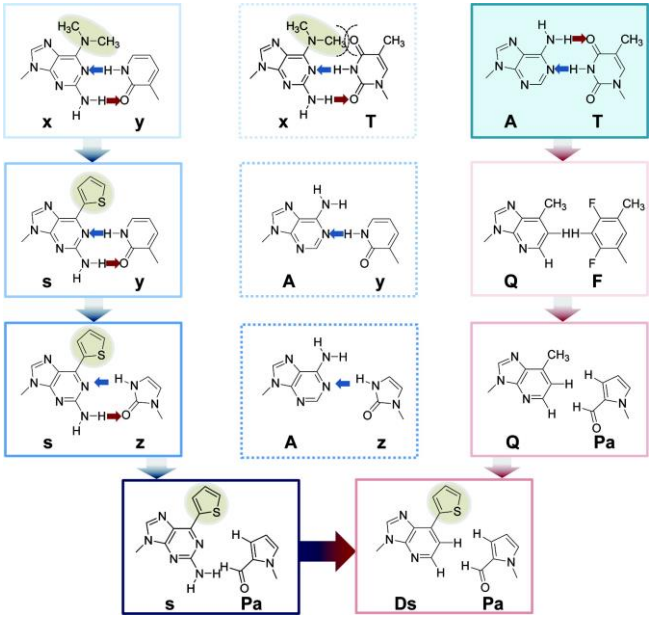


Michiko Kimoto, Rie Kawai, Tsuneo Mitsui, Shigeyuki Yokoyama, Ichiro Hirao

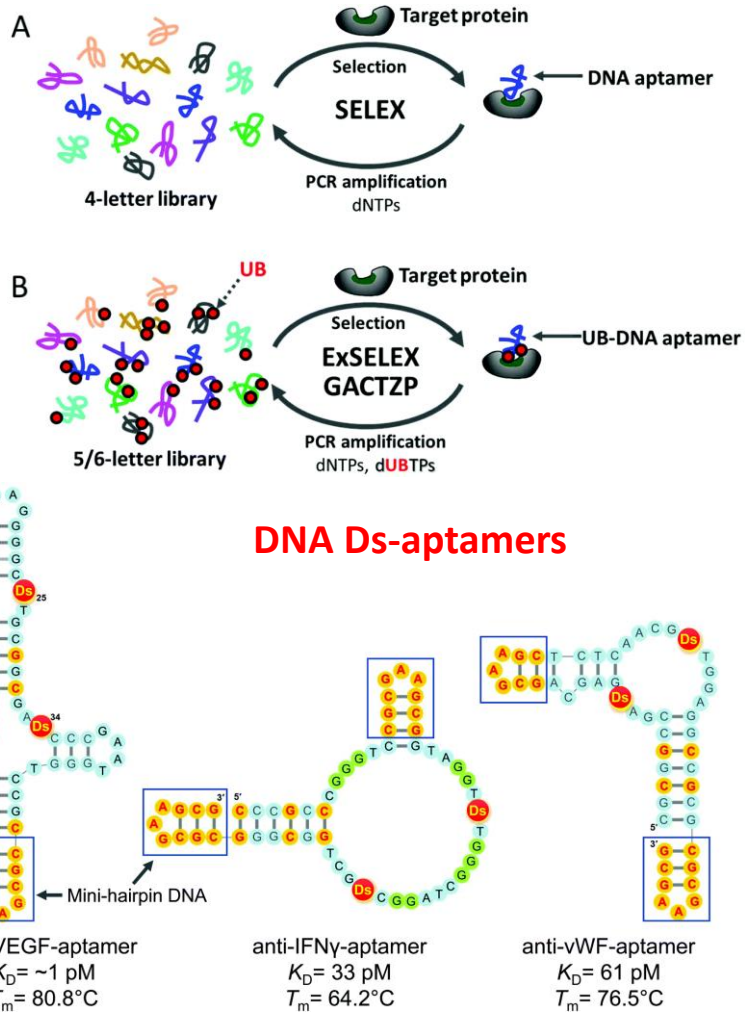
Nucleic Acids Research, Volume 37, Issue 2, 1 February 2009, Page e14

Expansion of the genetic alphabet 1

Non-hydrogen-bonded hydrophobic UBP

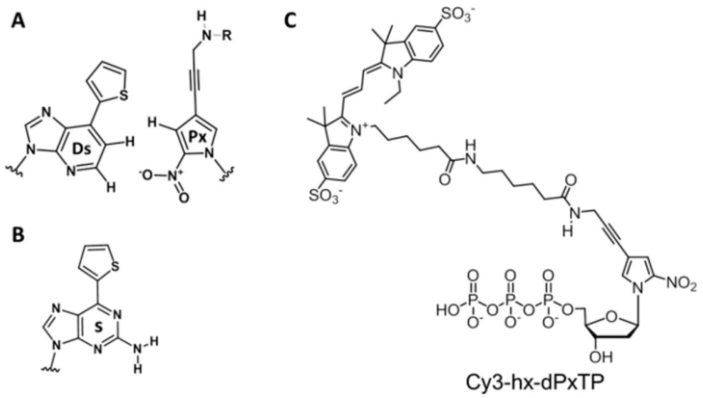


Hirao's unnatural base pair (**UBP)** development process: from the **x–y** and **Q–Pa** pairs to the **Ds–Pa** pair. The **Q–F** pair was synthesized by Kool's group as a hydrophobic A–T pair analogue. Ichiro Hirao; Michiko Kimoto; Rie Yamashige; *Acc. Chem. Res.* **2012,** 45, 2055-2065.



ExSELEX and secondary structures of high-affinity Ds-DNA aptamers
Better K_D T_m and nucelase resistant

Expansion of the genetic alphabet 1



Cy3-conjugated dPx substrate

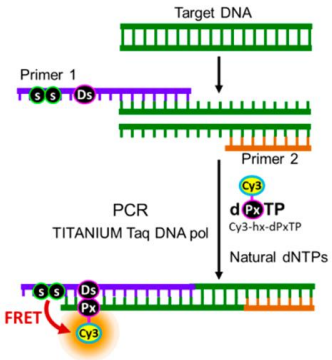
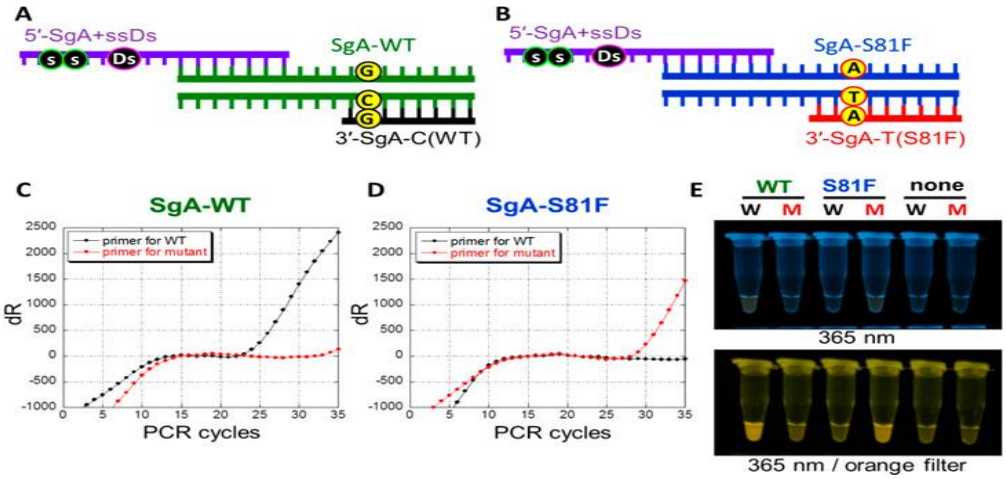


Figure 3. Visual PCR method using unnatural-base FRET system. The target DNA is amplified by PCR using a primer containing "Ds" and two adjacent s bases "ss", in the presence of natural dNTPs and Cy3-hx-dPxTP. During PCR, Cy3-hx-dPxTP is incorporated at the specific position opposite Ds, close to the "ss" part, which allows FRET between "ss" and Cy3.

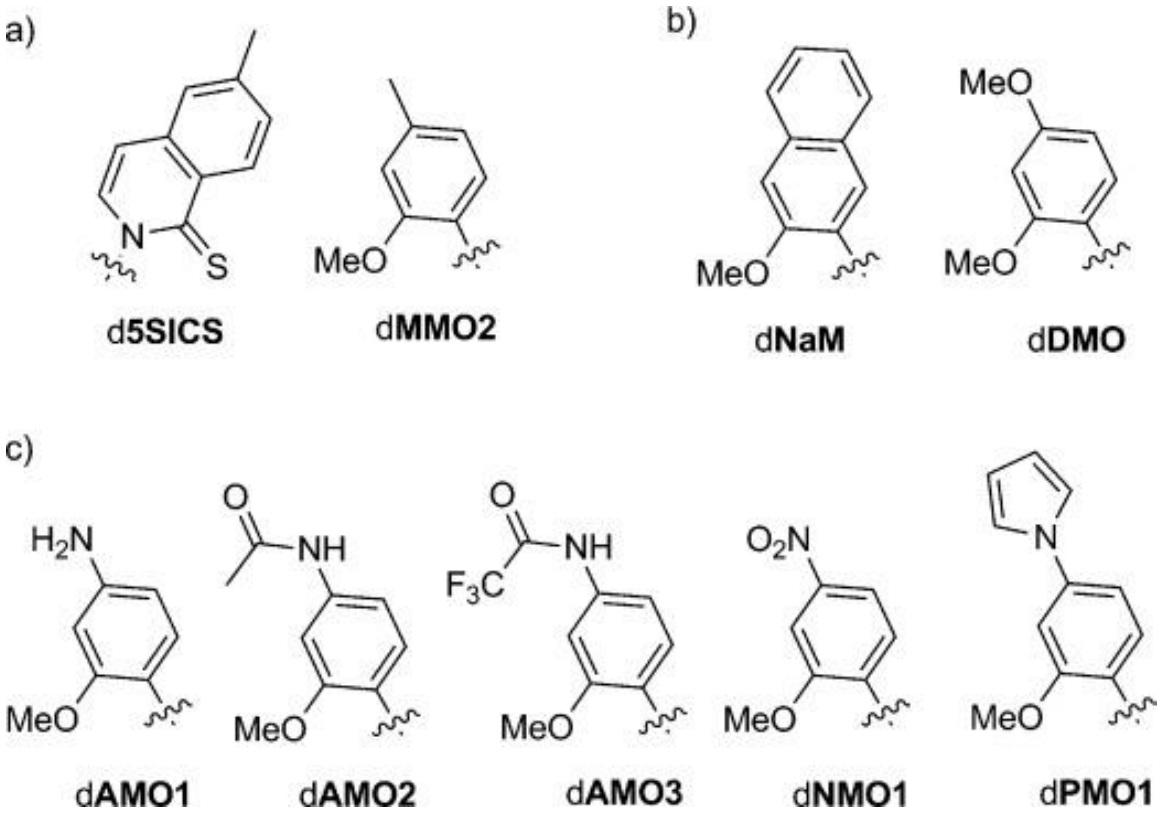
PCR detection of a single nucleotide polymorphism in quinolone-resistant *Streptococcus pneumoniae*.



PCR detection of a single nucleotide polymorphism in quinolone-resistant *Streptococcus pneumoniae*. Genomic DNA (5 ng) from a wild-type strain (SgA-WT) and a quinolone-resistant strain (SgA-S81F, a mutation in gyrA) was PCR-amplified by qPCR using two primer sets, respectively, with Cy3 fluorescence monitoring (A–D). After 35-cycle PCR, the end-point PCR solutions in the tubes were visualized by irradiation at 365 nm, with and without an orange filter (E).

J. Am. Chem. Soc. 2018, 140, 14038−14041

Expansion of the genetic alphabet 1



Major Groove Substituents and Polymerase Recognition of a Class of Predominantly Hydrophobic Unnatural Base Pairs

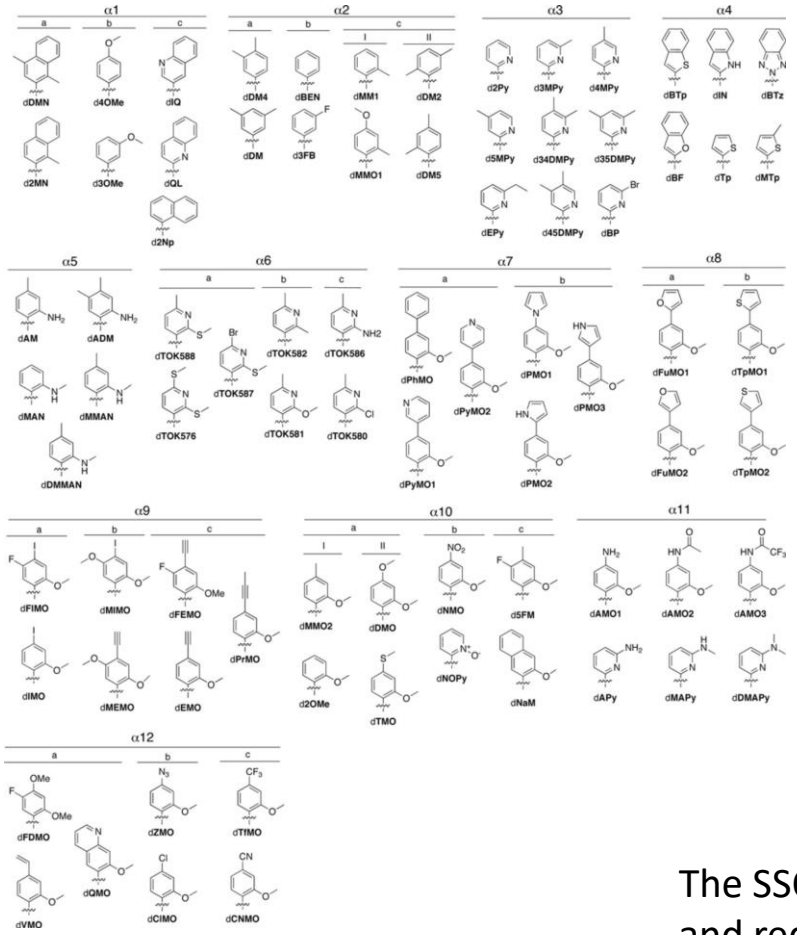
Dr. Thomas Lavergne, Denis A. Malyshev, Prof. Dr. Floyd E. Romesberg

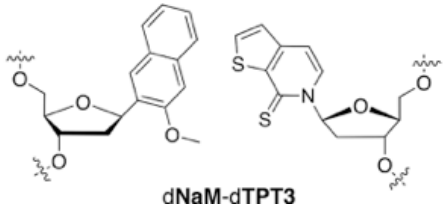
Chemistry A European J, Volume: 18, Issue: 4, Pages: 1231-1239, First published: 21 December 2011

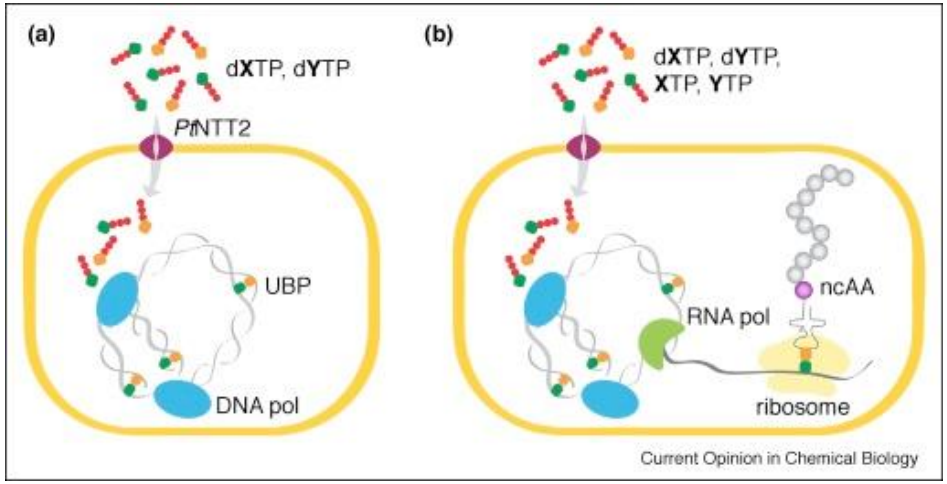
Expansion of the genetic alphabet 1

For several decades, chemists and synthetic biologists attempted to find alternatives

base modification





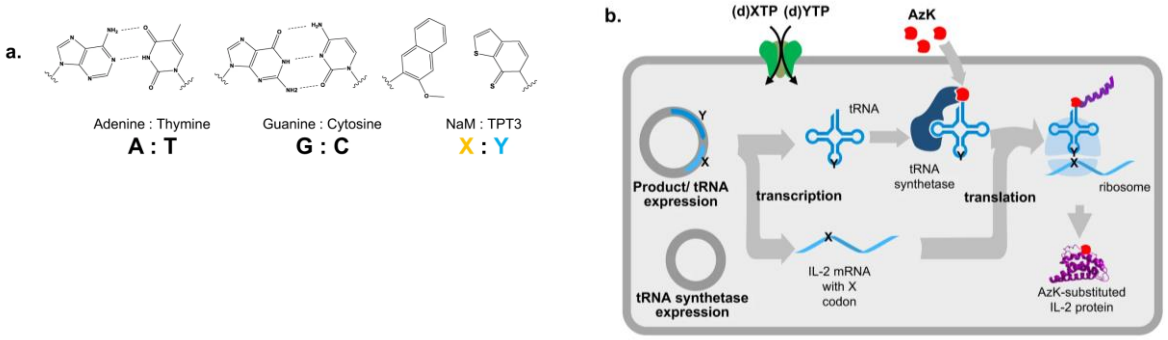


Systematic exploration of a class of hydrophobic unnatural base pairs
F. Romesberg Nucleic Acids Res. 2014;42(16):10235-10244

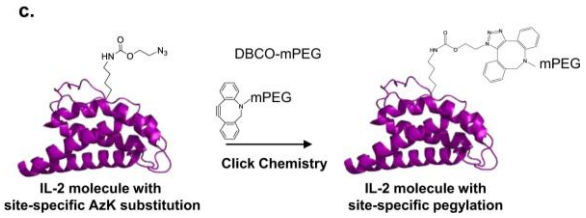
The SSO replicates DNA containing the dNaM-dTPT3 UBP (blue and red,) transcribes mRNA and tRNA with complementary codons and anticodons containing NaM or TPT3, uses an orthogonal synthetase to charge the tRNA with an ncAA, and uses the charged tRNA to translate the mRNA into proteins containing ncAAs.

Expansion of the genetic alphabet 1

An engineered IL-2 reprogrammed for anti-tumor therapy using a semi-synthetic organism
Nat Commun **12**, 4785 (2021).



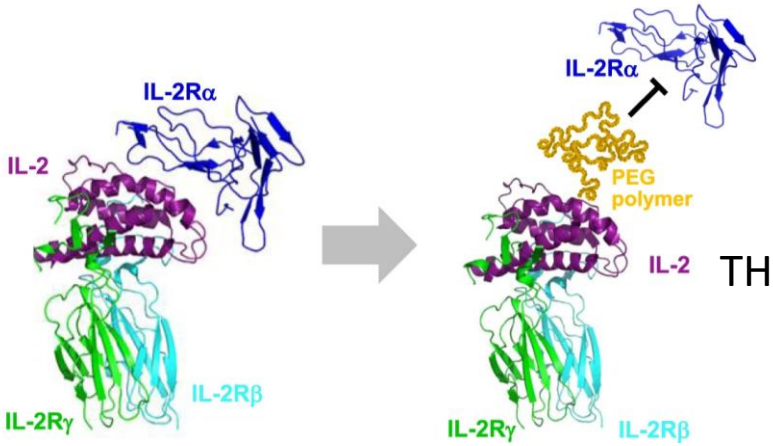
X–Y containing DNA sequences are transcribed into mRNA and tRNA that contain the codon (AXC) and anticodon sequences (GYU), respectively. The orthogonal tRNA synthetase PylRS specifically recognizes the AzK (red shapecharges this residue onto the corresponding GYU anticodon-containing orthogonal *pylT* tRNA. Finally, ribosomal translation machinery utilizes the resulting pools of AzK-charged tRNA (GYU) to specifically decode the AXC codon, allowing the production of site-specifically AzK-modified protein product



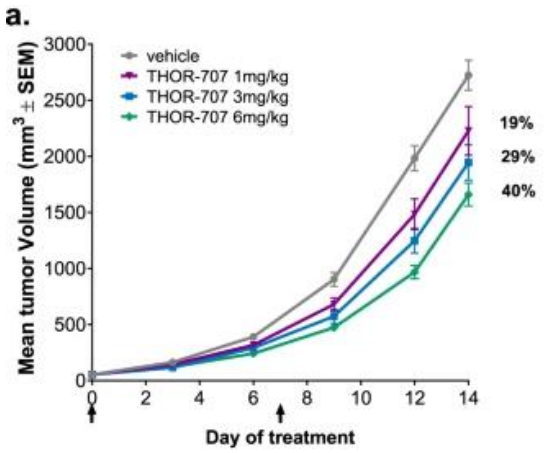
Reaction of the AzK-containing IL-2 proteins with DBCO-mPEG generates the site-specific and covalent attachment of the mPEG moiety to the AzK

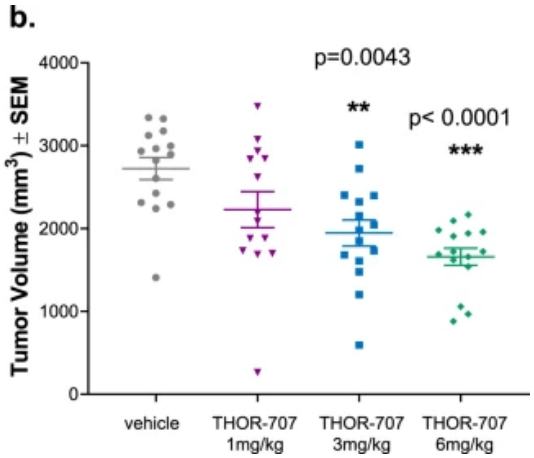
AzK: azide-containing non-natural amino acid (nAA) N6-(2-azidoethoxy)-carbonyl-l-lysine

Expansion of the genetic alphabet 1



THOR-707





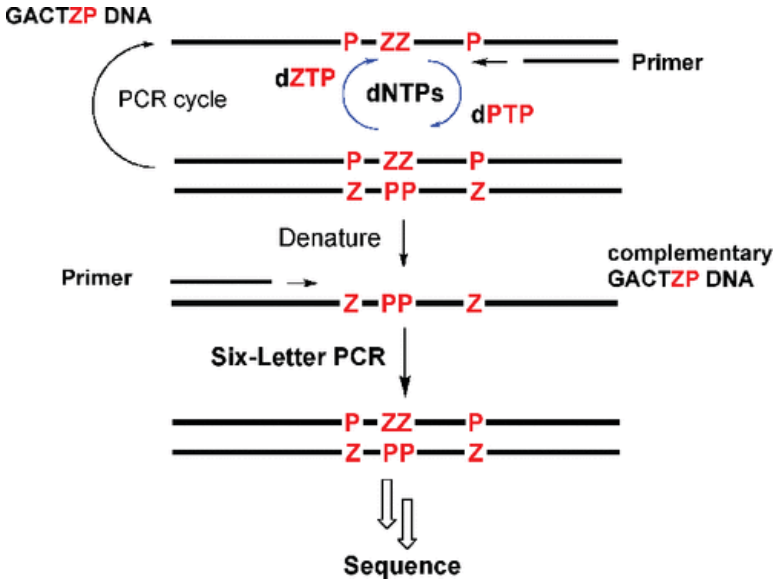
Mice implanted with B16-F10 tumors were administered vehicle or THOR-707 doses at 1, 3, and 6 mg/kg

Tumor volumes (mm^3) measured at day 14 post-dose initiation

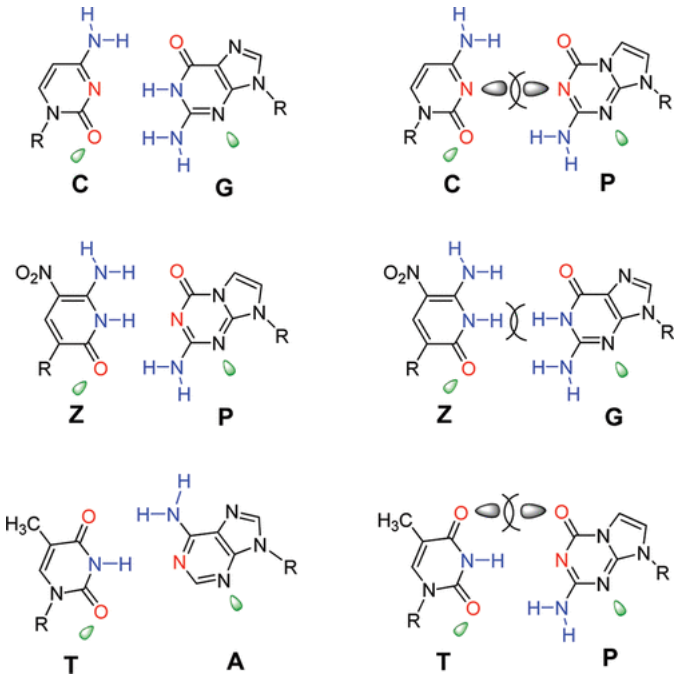
THOR-707 reduces B16-F10 tumor proliferation in C57BL/6 mice.

Expansion of the genetic alphabet 2

Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System



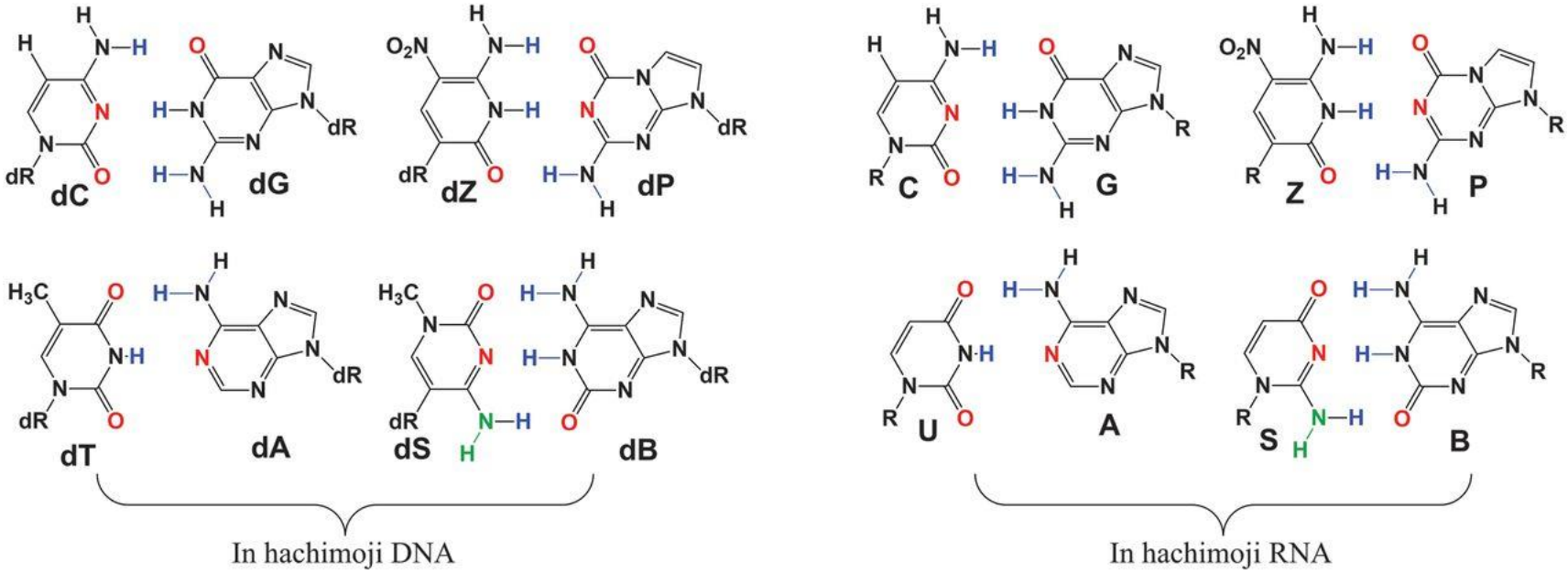
Hydrogen-bonded UBPs



Zunyi Yang, Fei Chen, J. Brian Alvarado, and Steven A. Benner
J. Am. Chem. Soc. 2011, 133, 15105–15112

Expansion of the genetic alphabet 2

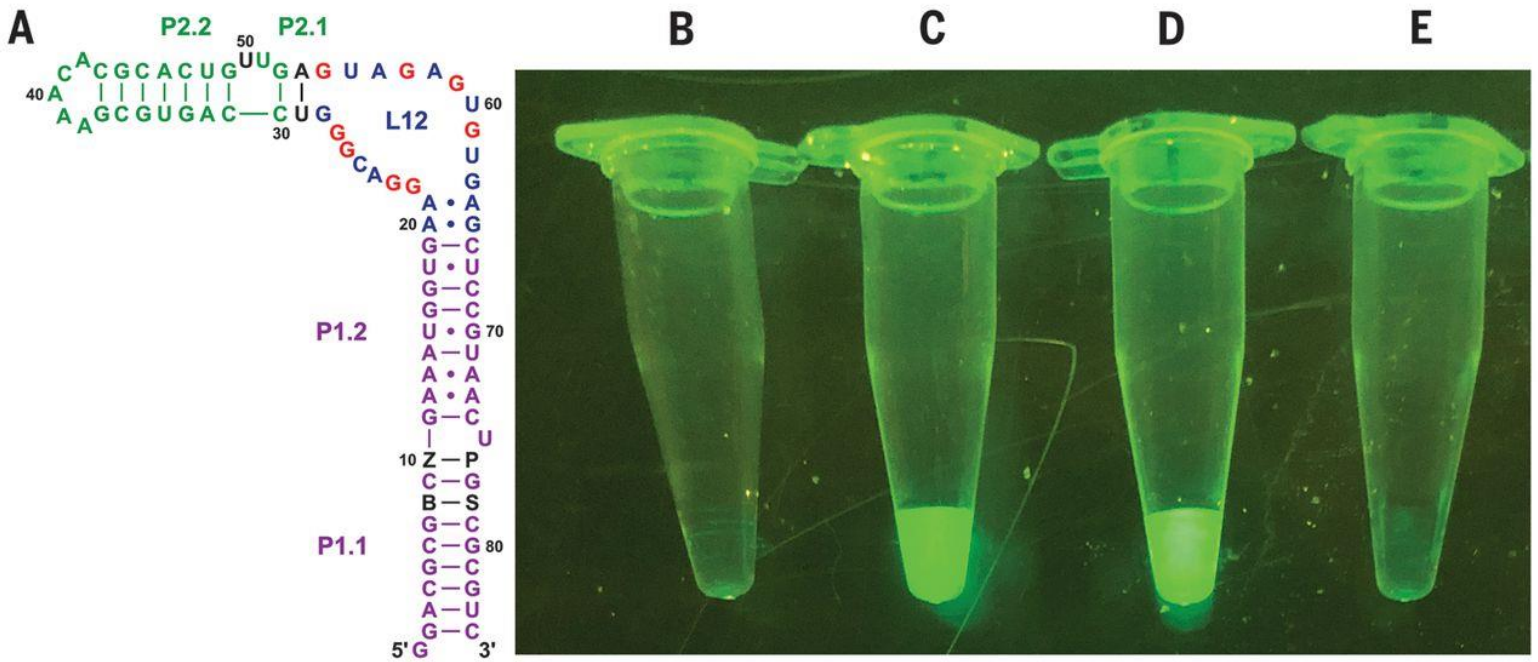
Hachimoji DNA and RNA: A genetic system with eight building blocks



Shuichi Hoshika, Nicole A. Leal, Myong-Jung Kim, Myong-Sang Kim,Nilesh B. Karalkar, Hyo-Joong Kim, Alison M. Bates, Norman E. Watkins Jr.,Holly A. SantaLucia, Adam J. Meyer, Saurja DasGupta, Joseph A. Piccirilli,Andrew D. Ellington, John SantaLucia Jr., Millie M. Georgiadis, Steven A. Benner

Science, Volume: 363, Issue: 6429, Pages: 884-887

Expansion of the genetic alphabet 2



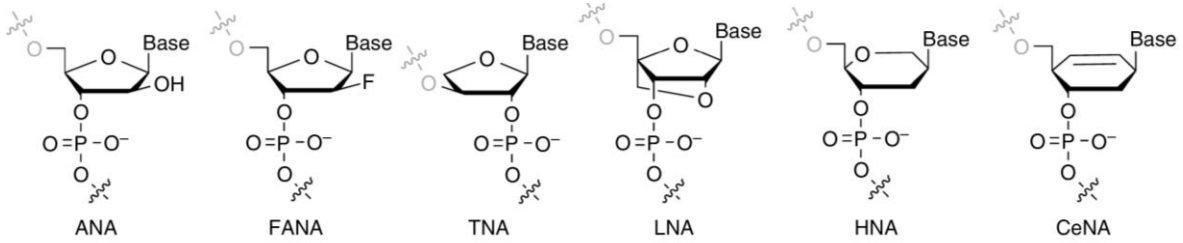
(A) the full hachimoji spinach variant aptamer (B) Control with fluor only, lacking RNA. (C) Hachimoji spinach with the sequence shown in (A). (D) Native spinach aptamer with fluor. (E) Fluor and spinach aptamer containing Z at position 50, replacing the A:U pair at positions 53:29 with G:C to restore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

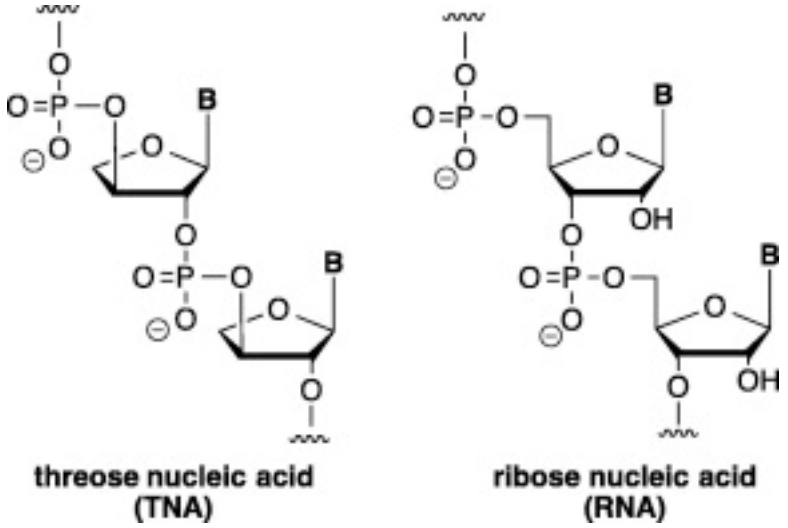
Science, Volume: 363, Issue: 6429, Pages: 884-887

Expansion of the genetic alphabet 3

Sugar modification

ex LNA, HNA, CeNA,……





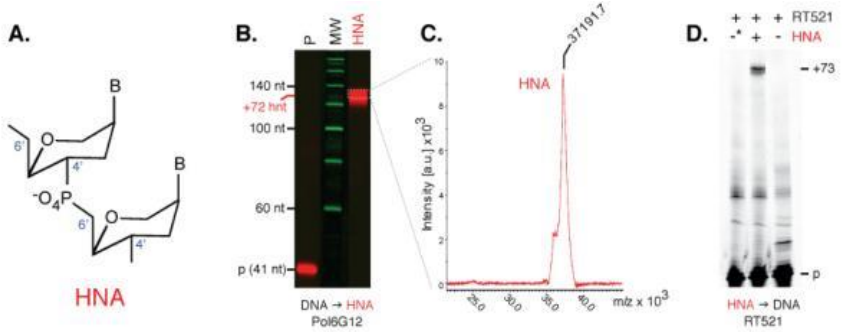
four-carbon threose sugar and phosphodiester linkages are connected at the 2′ and 3′ positions

TNA contains one less atom per backbone repeat unit than DNA

TNA adopts an A-like helical geometry, which is consistent with its ability to form stable antiparallel Watson–Crick duplexes with complementary strands of DNA and RNA

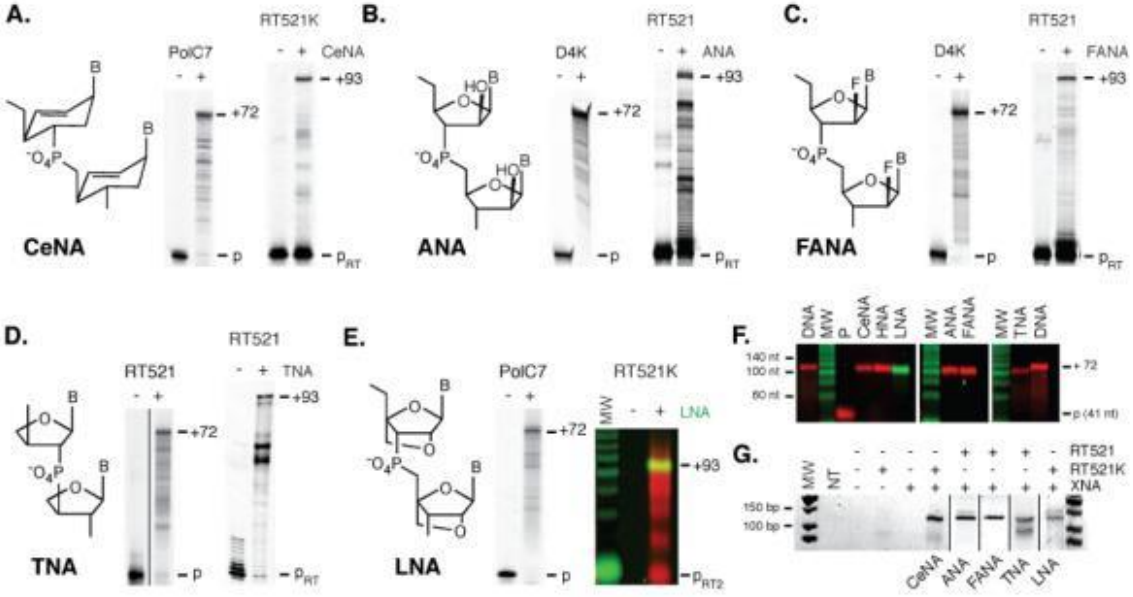
Expansion of the genetic alphabet 3

Synthetic genetic polymers capable of heredity and evolution



HNA synthesis,
mass spectrometry (MS) analysis
and reverse transcription

(A) Structure of 1,5-anhydrohexitol (HNA) nucleic acids (B: nucleobase).
(**B**) Pol6G12 extends the primer (p) incorporating 72 hNTP
(**C**) MS spectrum of full-length HNA molecule
showing a measured HNA mass of 37,190 ± 15 Da
(**D**) HNA reverse transcription (DNA synthesis from an HNA template)



XNA genetic polymers

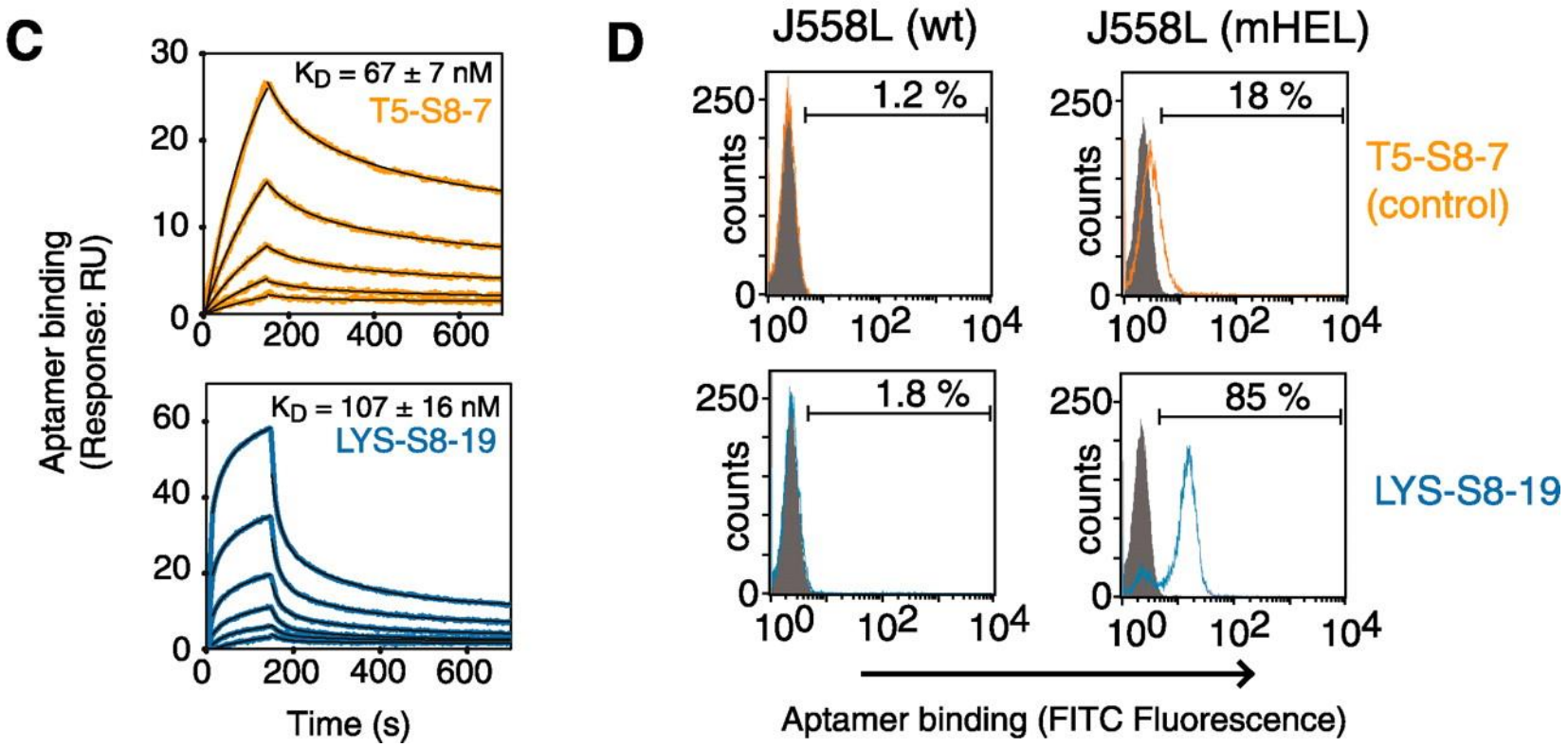
Structures (B: nucleobase), PAGE of synthesis (+72 xnt)
and reverse transcription (+93 nt)
of **(A)** CeNA, **(B)** ANA, **(C)** FANA, **(D)** and TNA.
(E) LNA synthesis (primer (41 nt) + 72 lnt) and LNA RT (red)
resolved by alkali agarose gel electrophoresis (AAGE).
LNA synthesis (green) migrates at its expected size (113 nt) and
co-migrates with reverse transcribed DNA (red)
synthesized from primer P_{RT2} (20 nt)
(F) AAGE of XNA and DNA polymers of identical sequence
(G) XNA RT-PCR)

Vitor B. Pinheiro,Alexander I. Taylor, Christopher Cozens, Mikhail Abramov, Marleen Renders, Su Zhang, John C. Chaput, Jesper Wengel, Sew-Yeu Peak-Chew, Stephen H. McLaughlin, Piet Herdewijn, Philipp Holliger

Science 20 Apr 2012 Vol 336, Issue 6079 pp. 341-344

Expansion of the genetic alphabet 3

Characterization of HNA aptamers



HIV trans-activating response RNA (TAR)T5-S8-7

HNA: 6'-AGGTAGTGCTGTTCGTTCATCTCAAATCTAGTTCGCTATCCAGTTGGC-4')

and anti-HEL aptamer LYS-S8-19

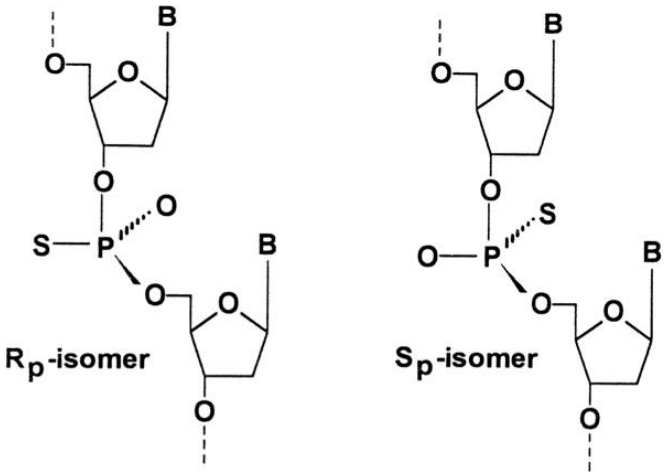
HNA: 6'-AGGTAGTGCTGTTCGTTTAAATGTGTGTCGTCGTTCGCTATCCAGTTGGC-4

FACS analysis of fluorescein isothiocyanate (FITC)–labeled aptamers binding to plasmacytoma line J558L with and without expression of membrane-bound HEL (mHEL)

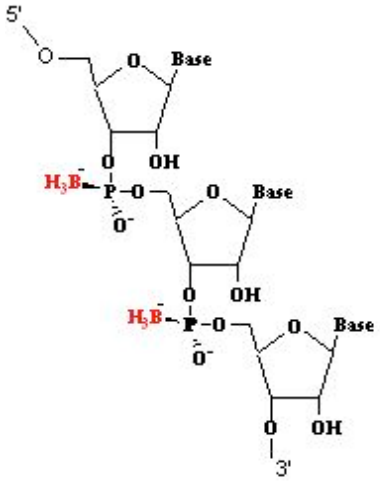
Expansion of the genetic alphabet 4

Phosphate modification

phosphorothioate



boranophosphate



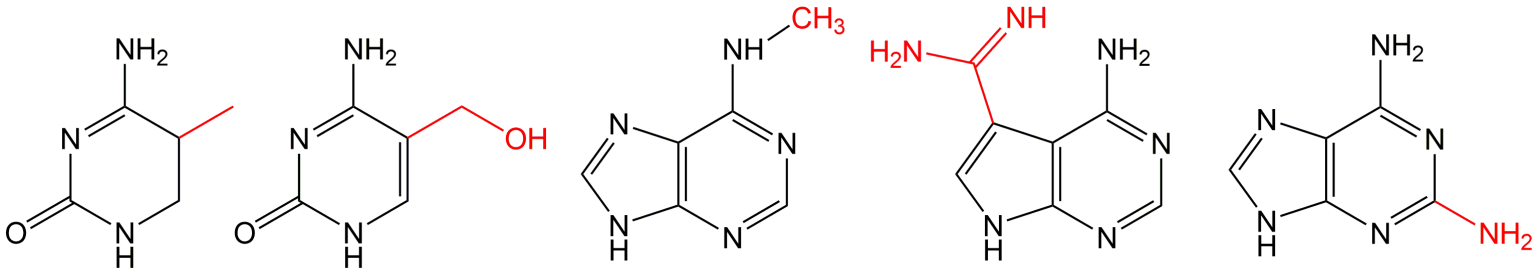
- Synthetic modification that stabilizes oligodeoxynucleotides against nuclease degradation
- Post replicative modification recenly found in DNA from bacteria harboring the five-gene *dnd* cluster

mimic of phosphate

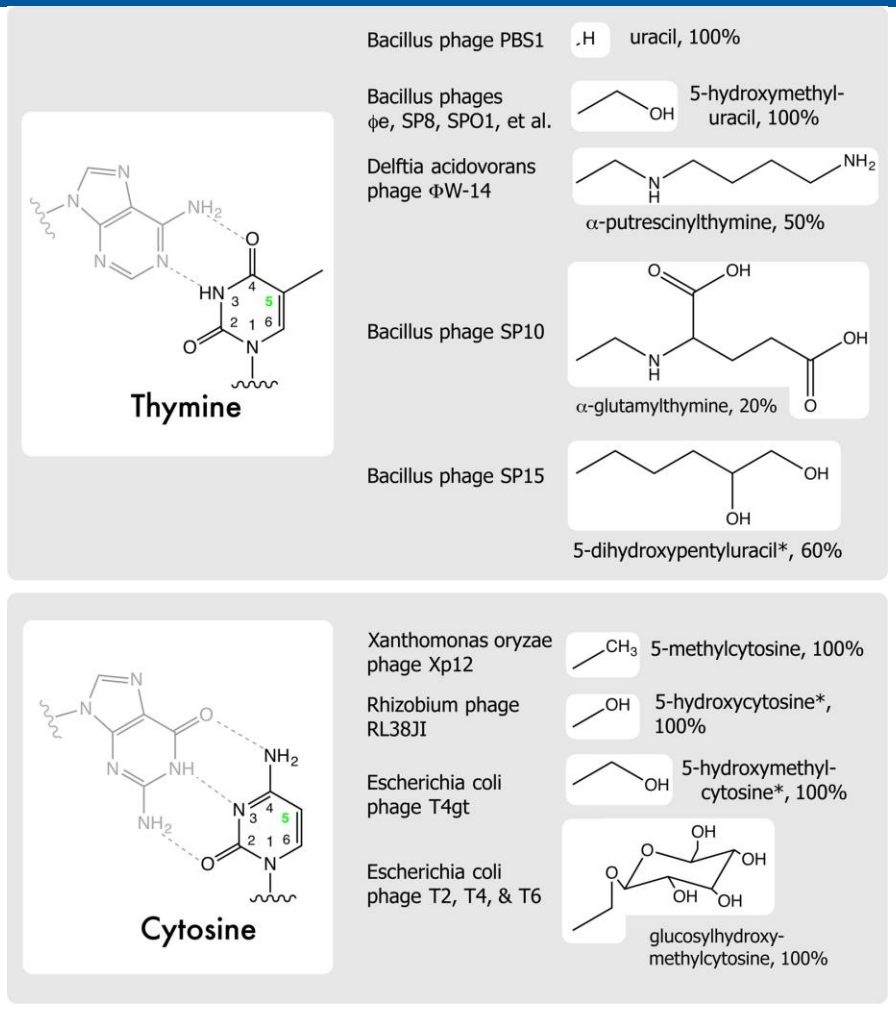
Bacteriophages contain the greatest diversity of modified bases so far observed in nature.

History

5-hydroxymethyl cytosine in T-even bacteriophages Wyatt and Cohen 1953
Glycosyl substituted 5-hydroxymethyl cytosine in T2r+ phage Sinsheimer 1954
5-methylcytosine in bacteriophage λ Ledinko, N. 1964



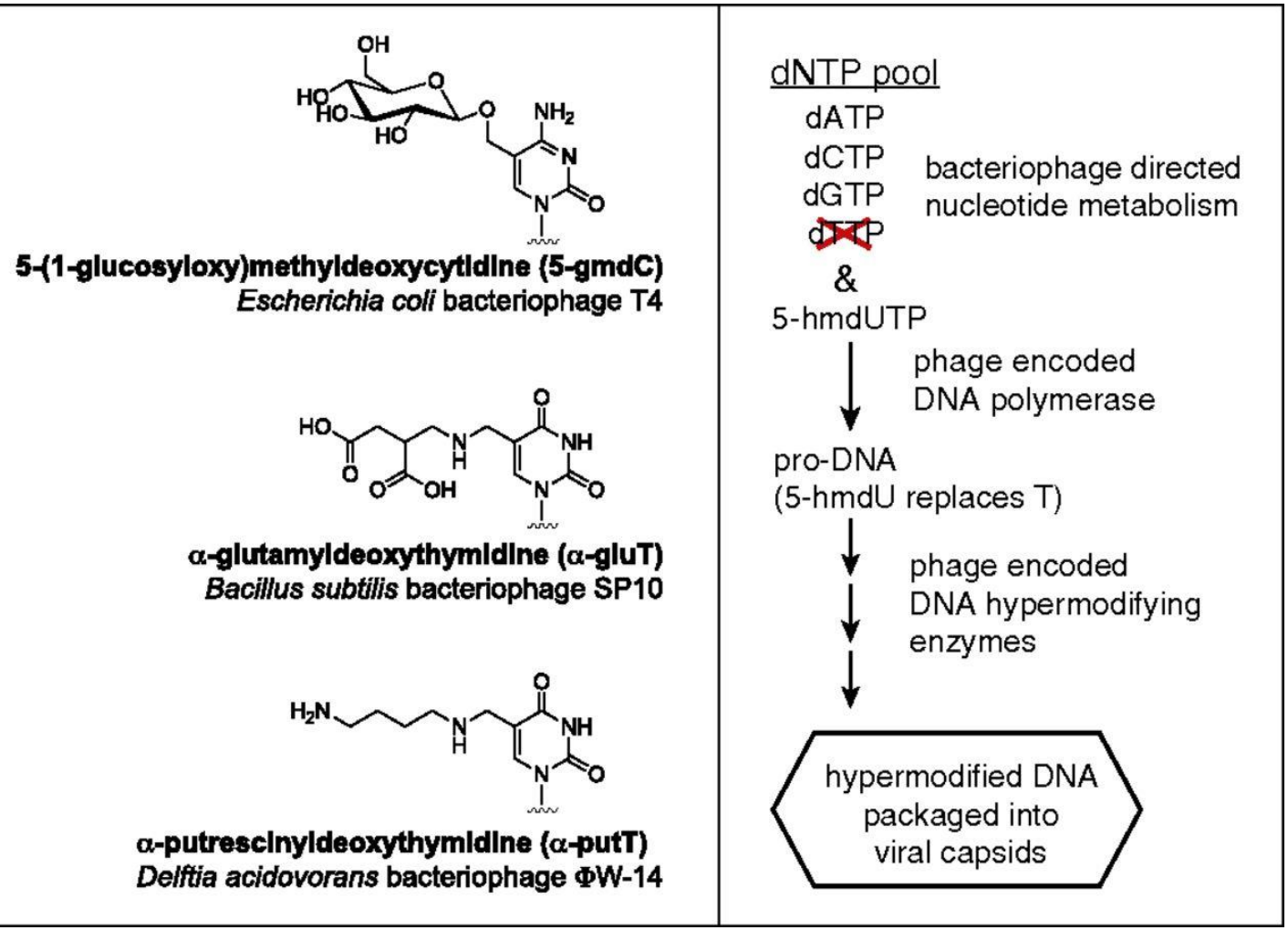
Modified pyrimidines in bacteriophages



Modified pyrimidines of bacteriophages. Adenosine and cytosine are shown in the context of their respective base pairs. Only the nucleobase portion is shown. The side groups illustrated are attached at those positions of the pyrimidine heterocycle indicated in green. The atoms of the pyrimidine heterocycles are numbered according to standard convention.

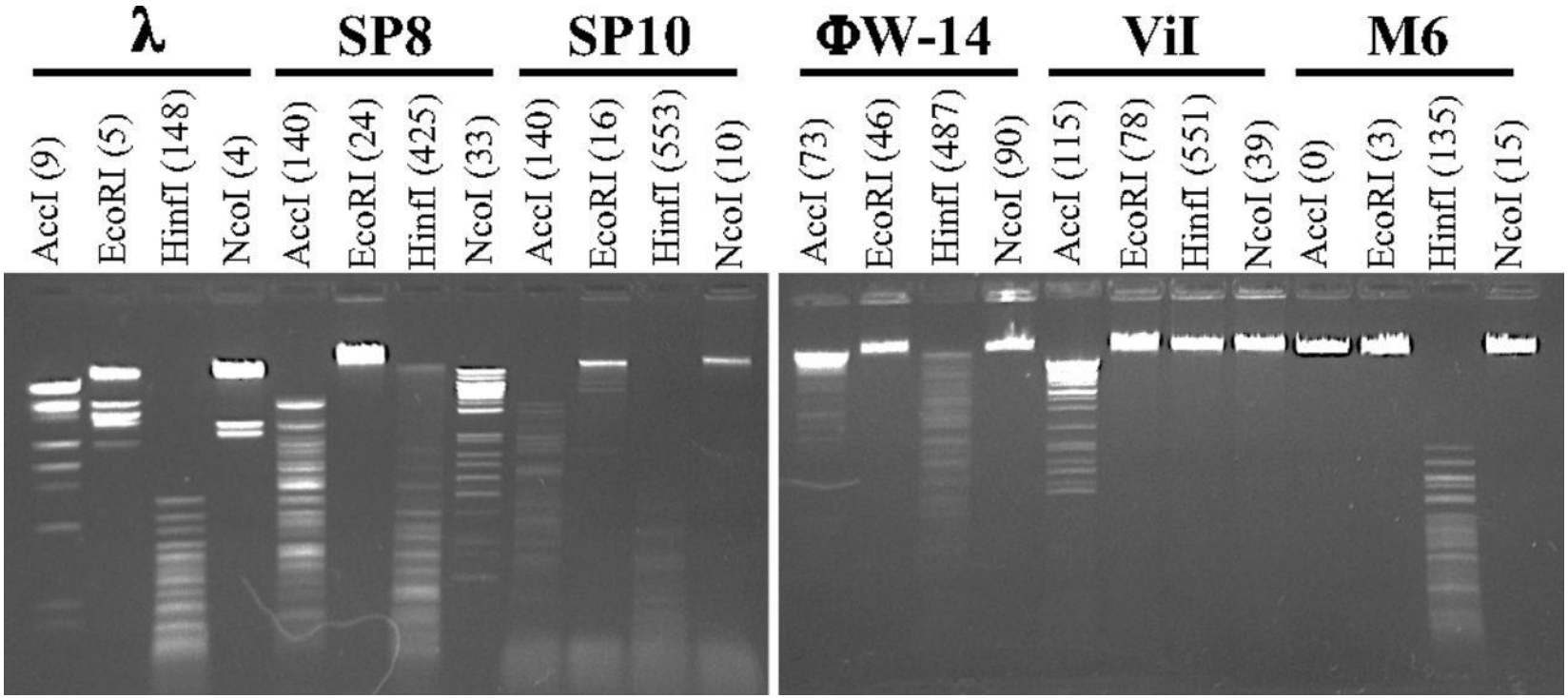
Published in: Peter Weigele; Published in: Elisabeth A. Raleigh; *Chem. Rev.* **2016,** 116, 12655-12687.
DOI: 10.1021/acs.chemrev.6b00114

Examples of hypermodified pyrimidines and the generalized DNA thymidine (T) hypermodification pathway of phages SP10 and ΦW-14



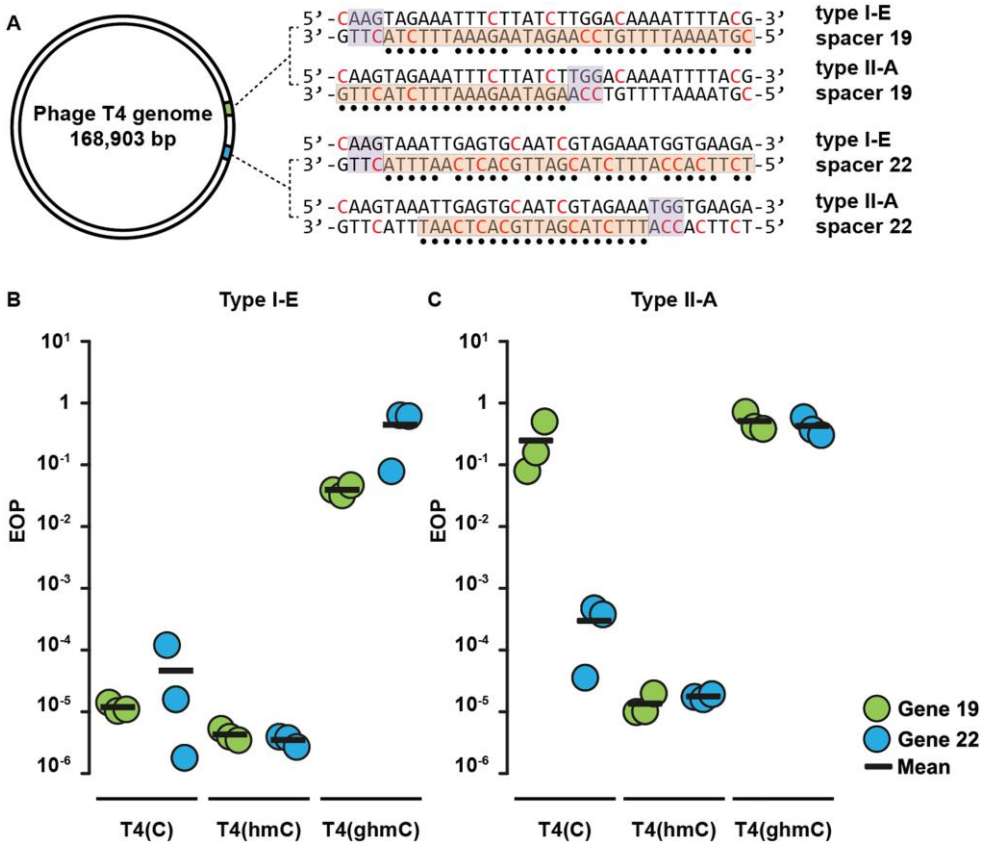
Yan-Jiun Lee et al. PNAS 2018;115:14:E3116-E3125

Modified and hypermodified DNAs are often resistant to a range of restriction endoucleases in vitro



Yan-Jiun Lee et al. PNAS 2018;115:14:E3116-E3125

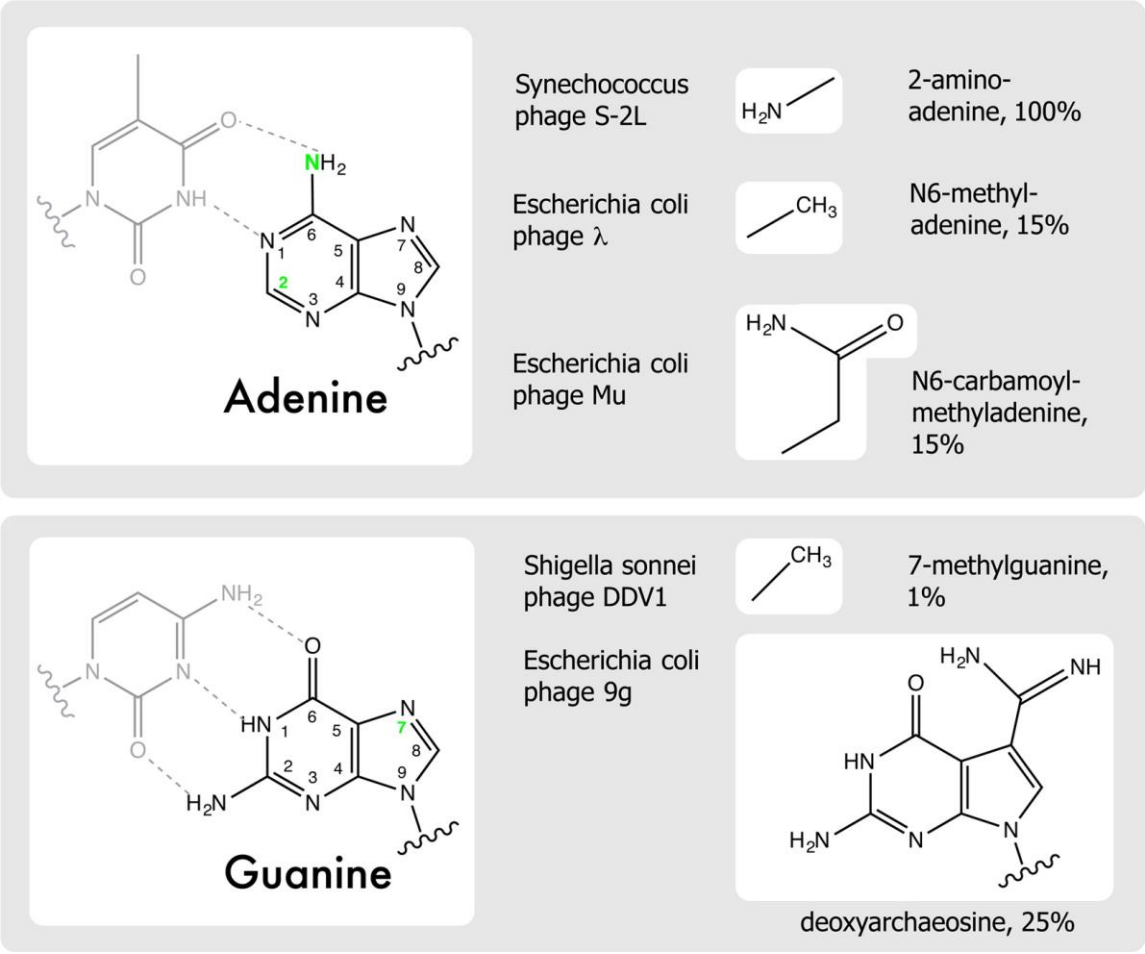
type I-E and type II-A CRISPR–Cas systems are severely impaired by phage DNA glucosylation



type I-E resistance is inhibited by 5-ghmC
but not by 5-hmC modifications

EOP of T4(C) and T4(hmC) are similar
strongly reduced for T4(ghmC)

Modified purines in bacteriophages



Modified purines of bacteriophages. Adenine and guanine are shown in the context of their respective base pairs. Only the nucleobase portion is shown; the side groups illustrated are attached at those positions of the purine heterocycle indicated in green. The atoms of the purine heterocycle are numbered according to standard convention.

Published in: Peter Weigele; Published in: Elisabeth A. Raleigh; *Chem. Rev.* **2016,** 116, 12655-12687.
DOI: 10.1021/acs.chemrev.6b00114

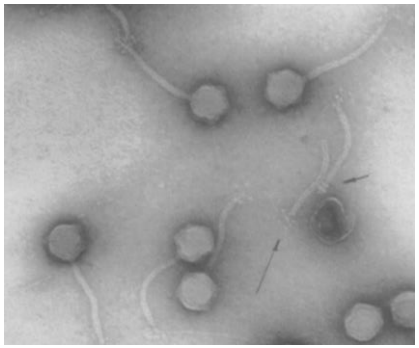
Cyanophage S-2L

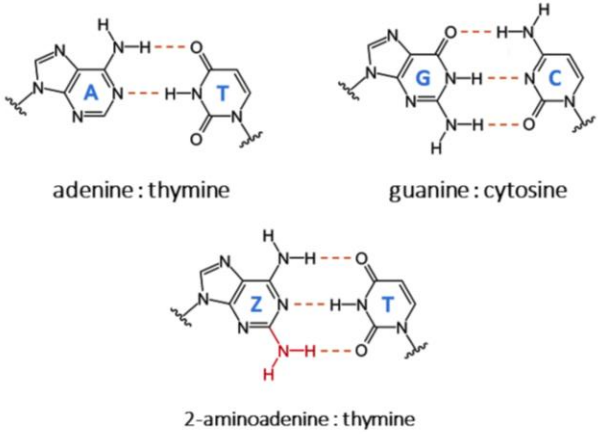
S-2L was isolated from water samples taken in the outskirts of Leningrad, USSR

In S-2L phage DNA, 2,6-diaminopurine (Z) completely replaces adenine

Kirnos MD IY *et al* Nature. 1977 Nov 24;270(5635):369-70

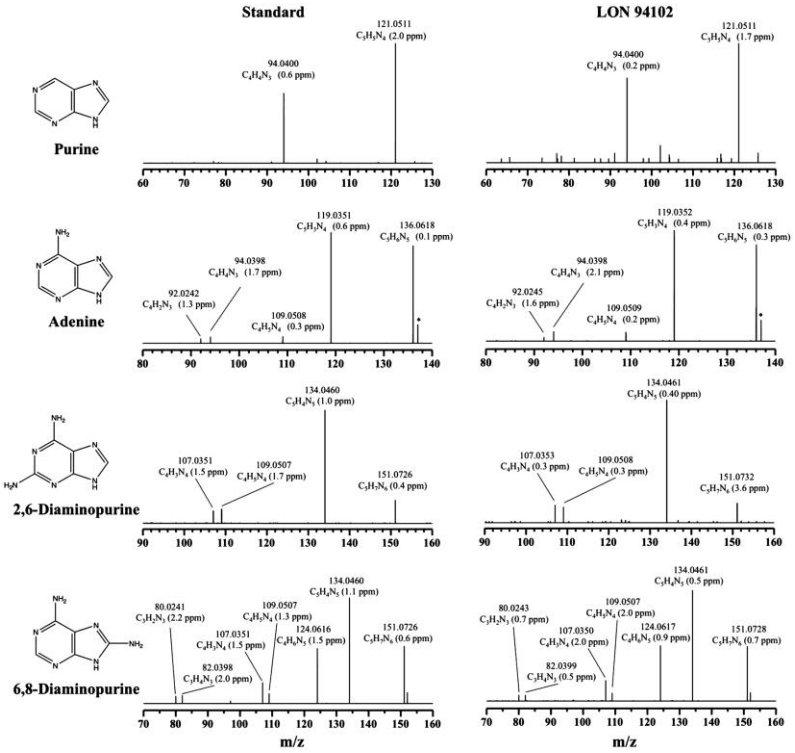
Khudyakov IY *et al Virology* 1978 (1):8-18.





augmentation of thermostability
modification of protein recognition

2,6-diaminopurine identification in carbonaceous meteorites

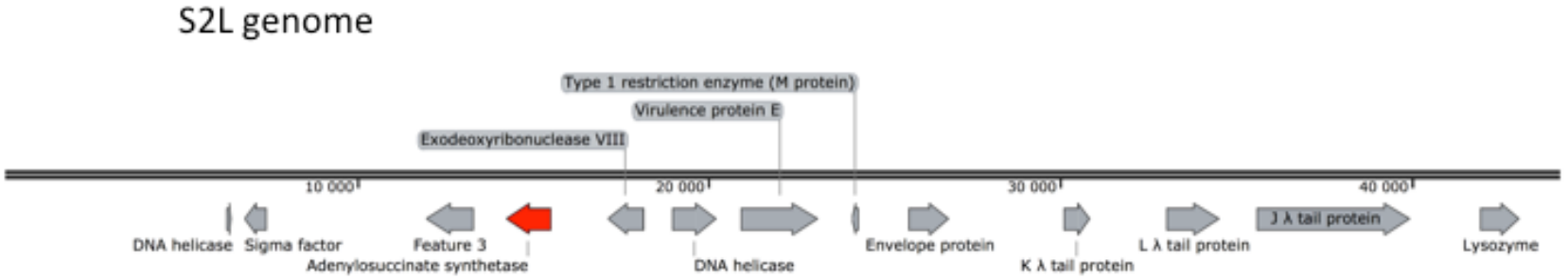


Mass-selected fragmentation spectra of reference standards (left spectra)
and compounds found in the meteorite LON 94102

The purines detected in meteorites are consistent with products of ammonium cyanide chemistry, which provides a plausible mechanism for their synthesis in the asteroid parent bodies, and strongly supports an extraterrestrial origin.

Michael P. Callahan et al. PNAS 2011;108:13995-13998

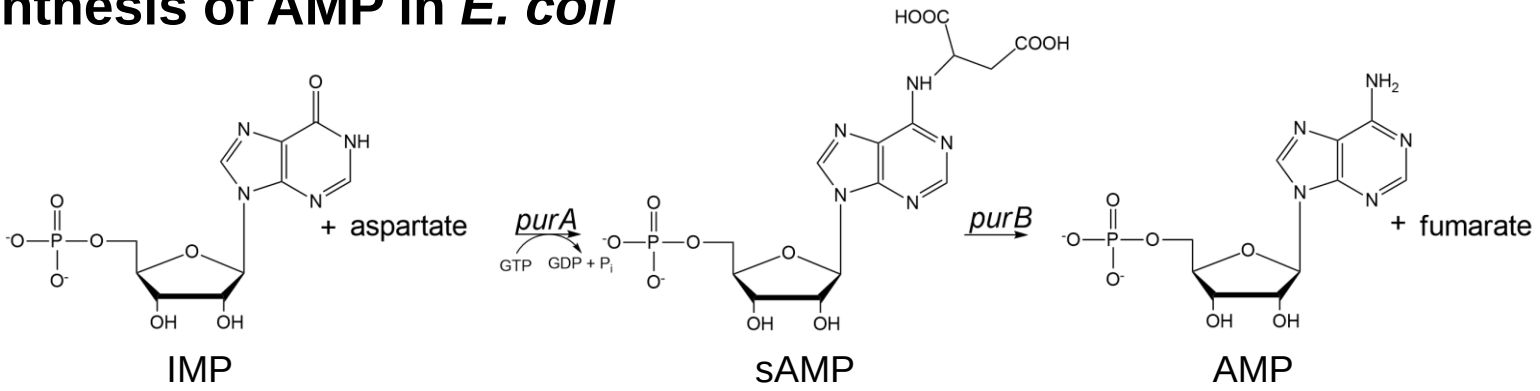
S-2L genome



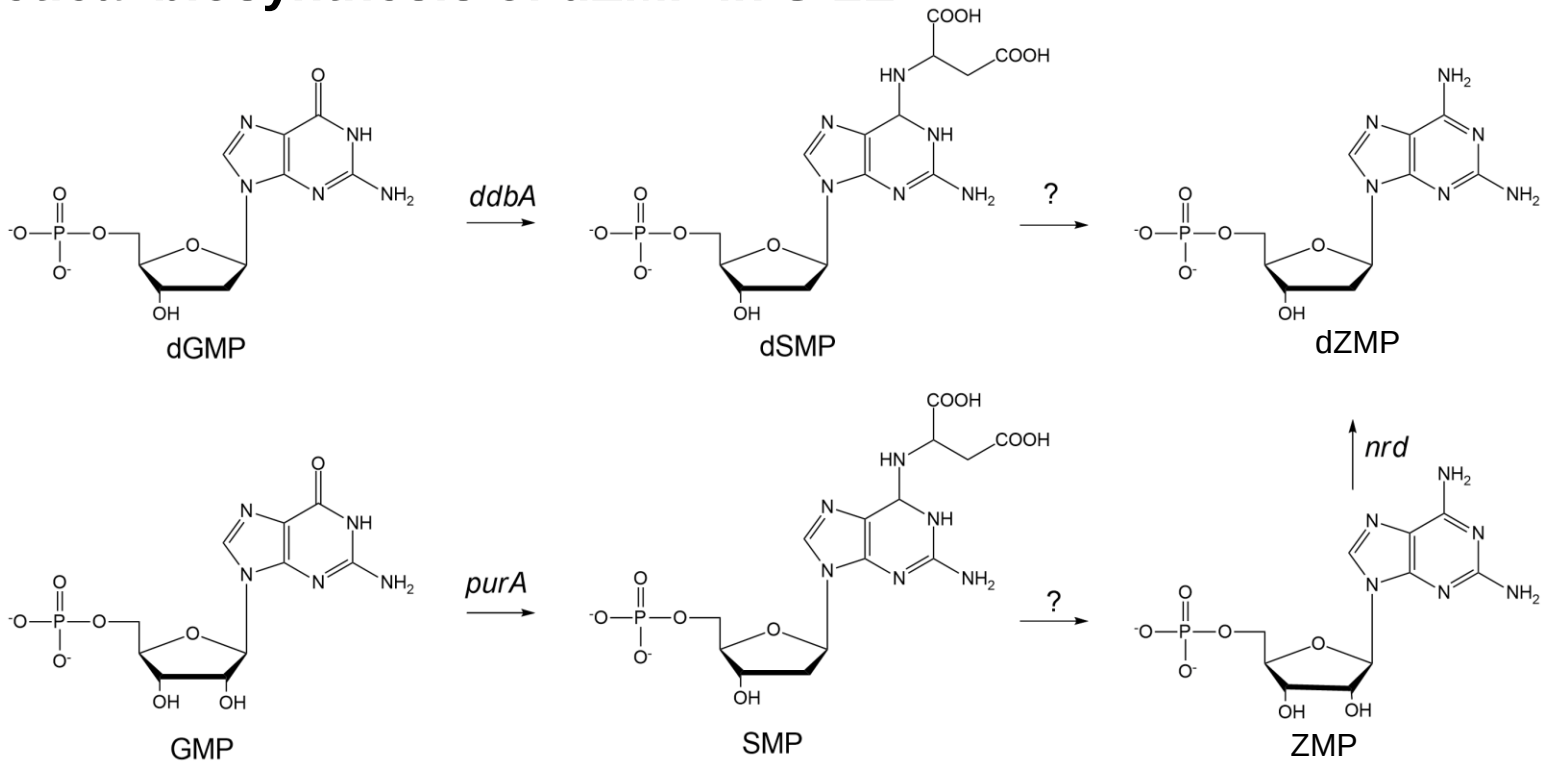
An adenylosuccinate synthase-like may participate to the biosynthesis of **2,6 diaminopurine** (Patent US 20060270005 A1)

Biological pathways for dZMP

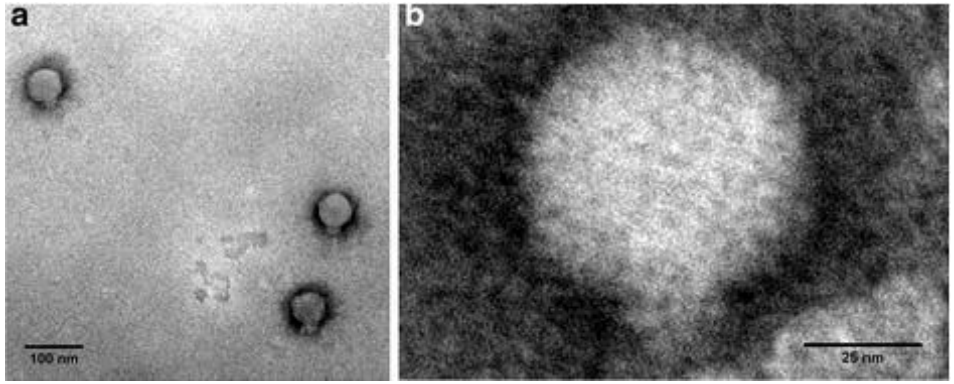
biosynthesis of AMP in *E. coli*



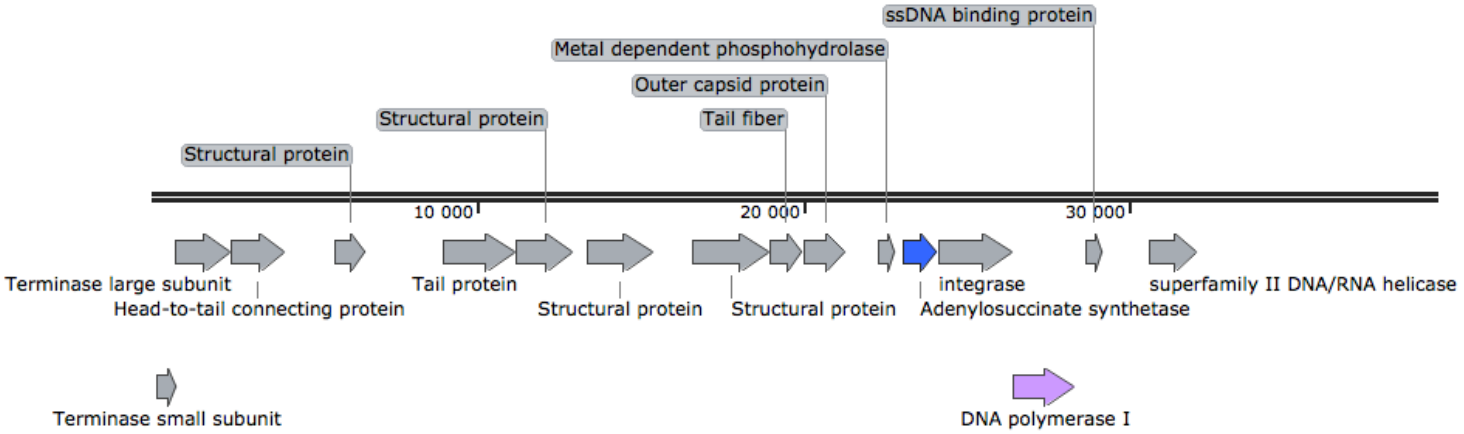
hypothetical biosynthesis of dZMP in S-2L



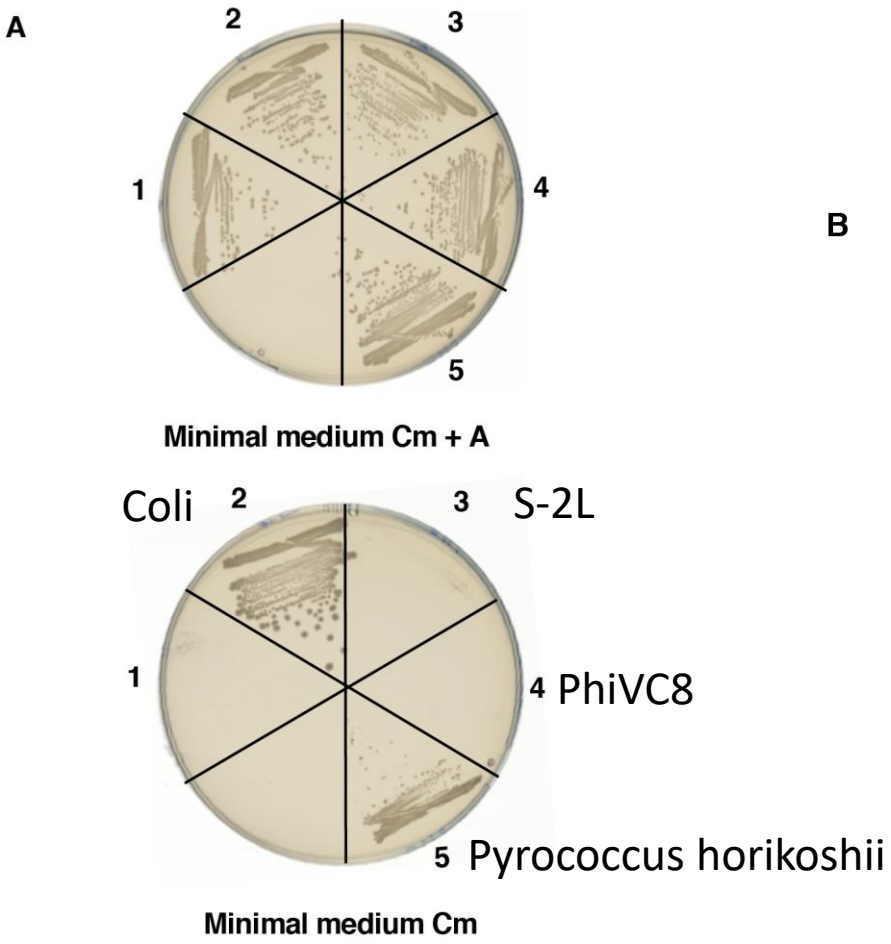
The ΦVC8 lytic phage for *Vibrio cholerae* O1



Electron microscopy of the ØVC8 phage isolated from water samples

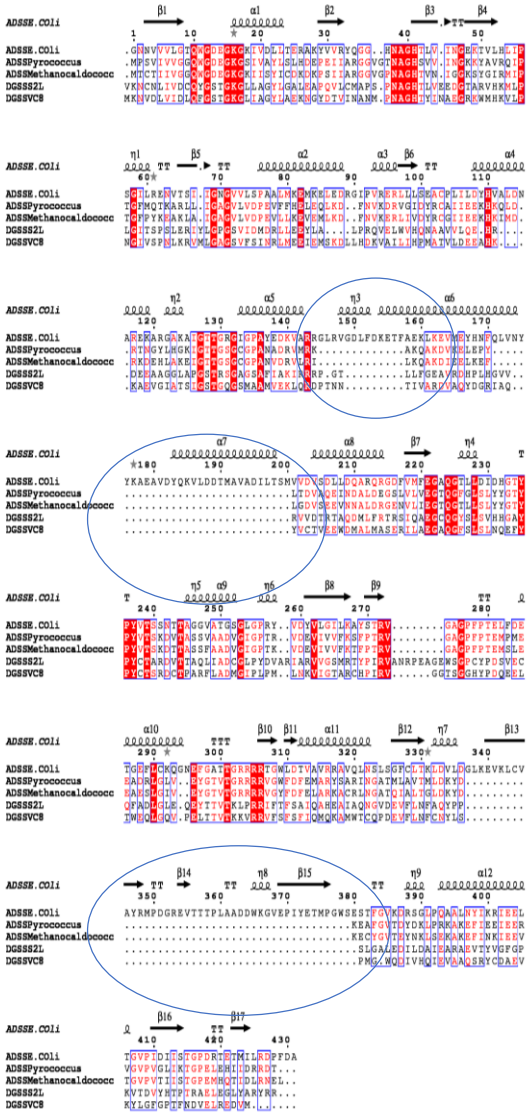


Functional compementation of an *E. coli purA* mutant

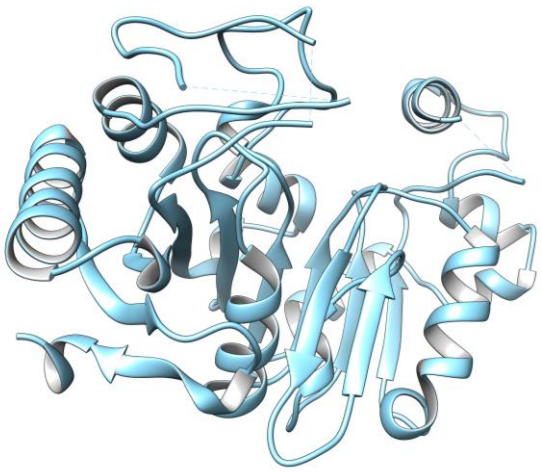


ΦVC8 and S-2L purA do not code an ADSS

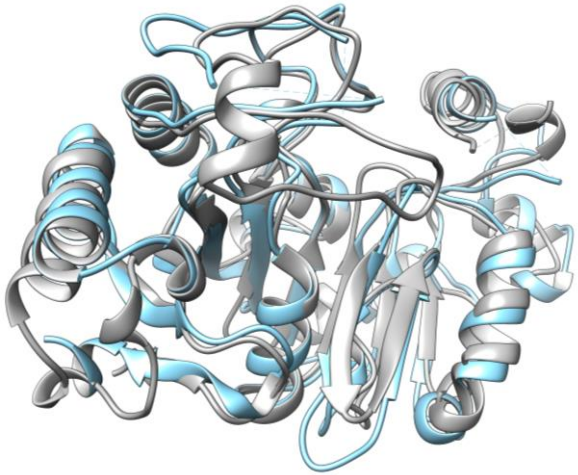
Comparaison of adenylosuccinate synthases



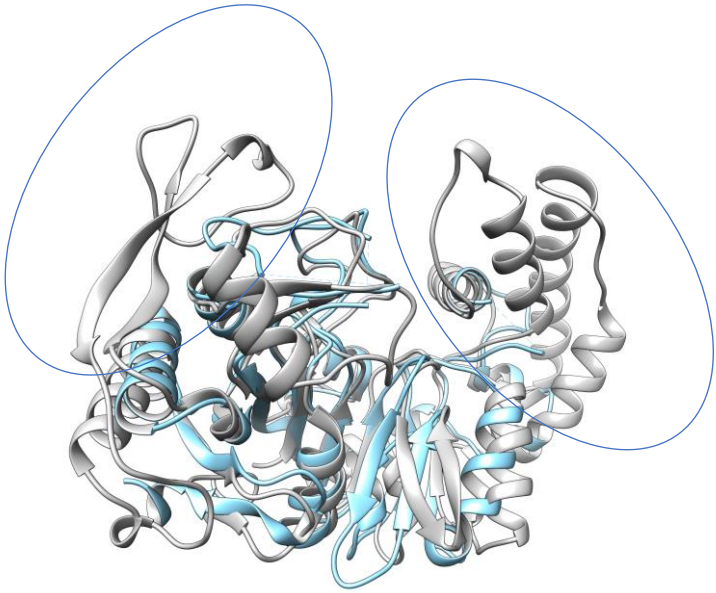
X-Ray structure of ΦVC8 PurZ



ΦVC8 PurZ Apo

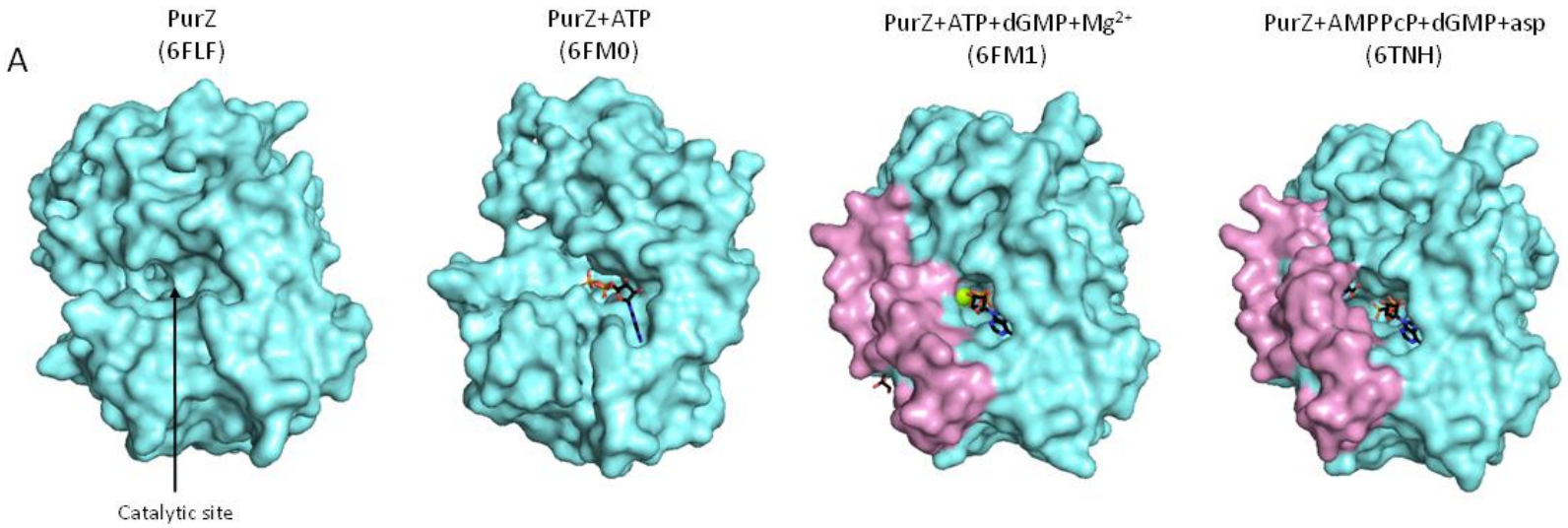


Overlay of ΦVC8 PurZ (blue) and *P. horikoshii* ADSS (grey)

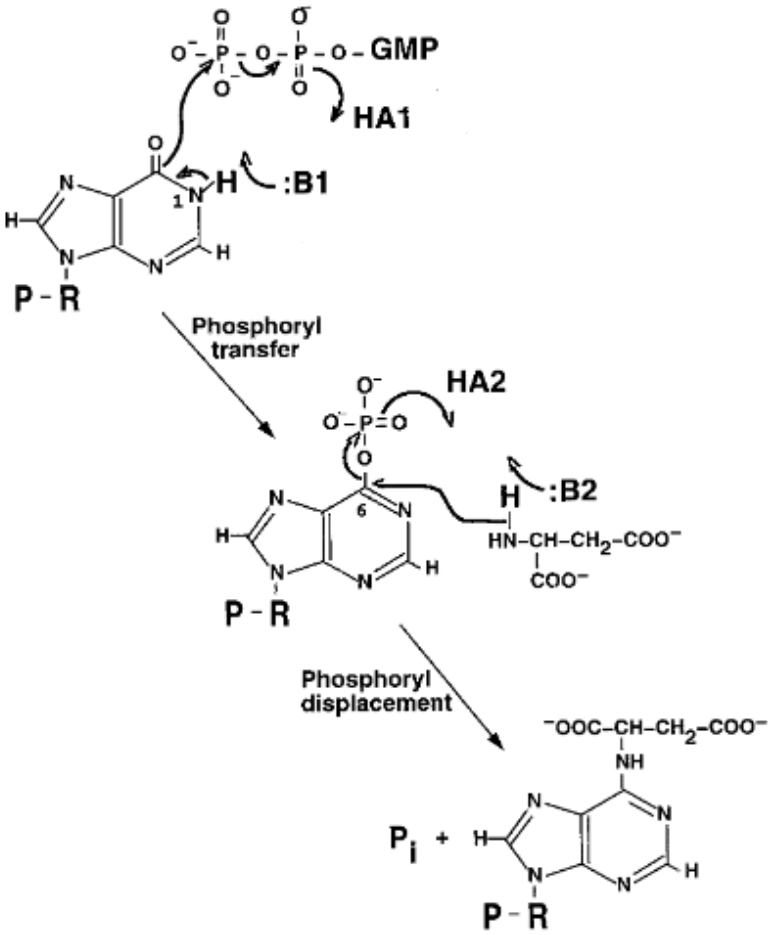


Overlay of ΦVC8 PurZ(blue) and *E.Coli* ADSS (grey)

X-Ray structures of ΦVC8 PurZ

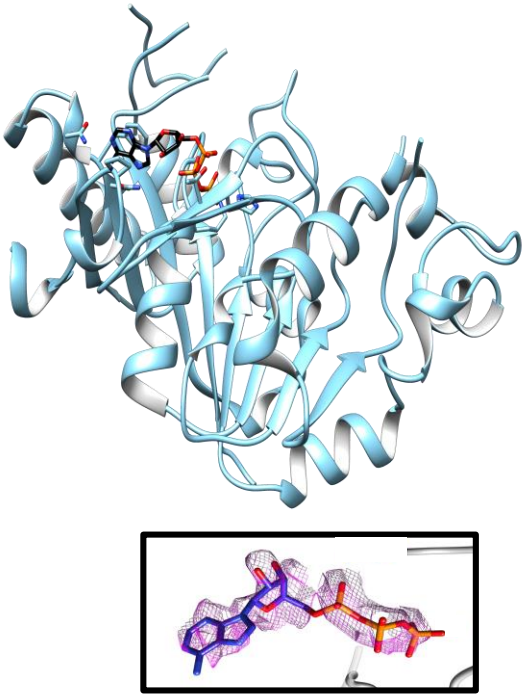


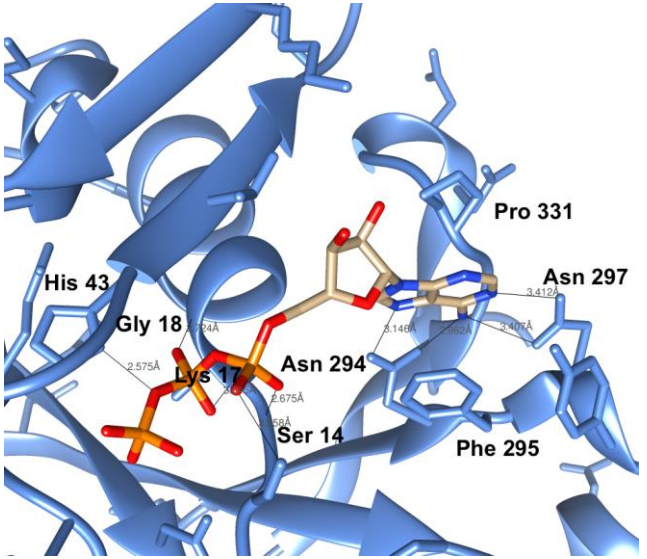
Two-step reaction mechanism proposed for adenylosuccinate synthetase

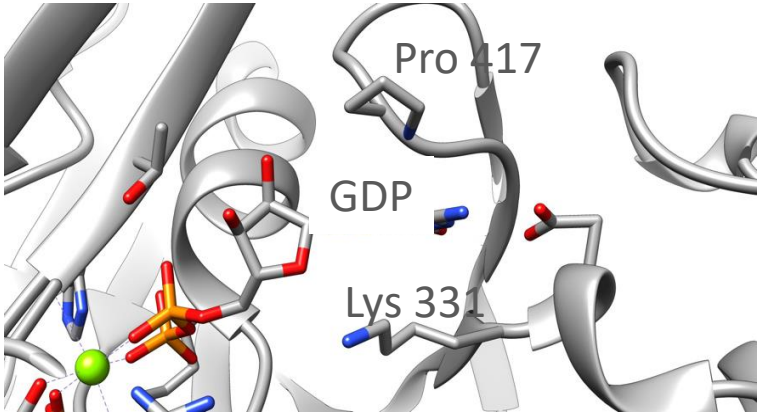


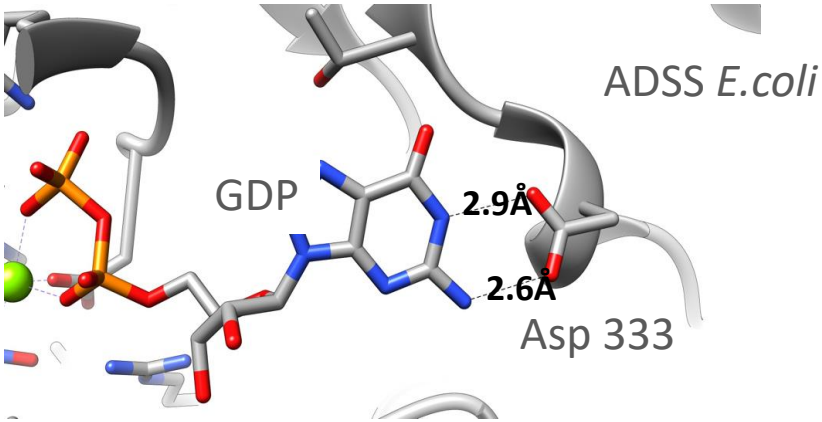
Hypothesis: ΦVC8 PurZ uses the same mechanism but with different substrates

X-Ray structure of ΦVC8 PurZ-ATP complex

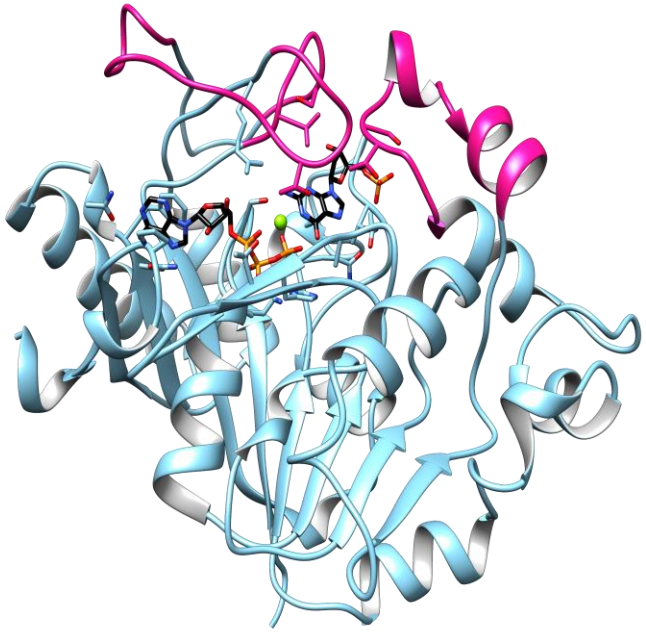


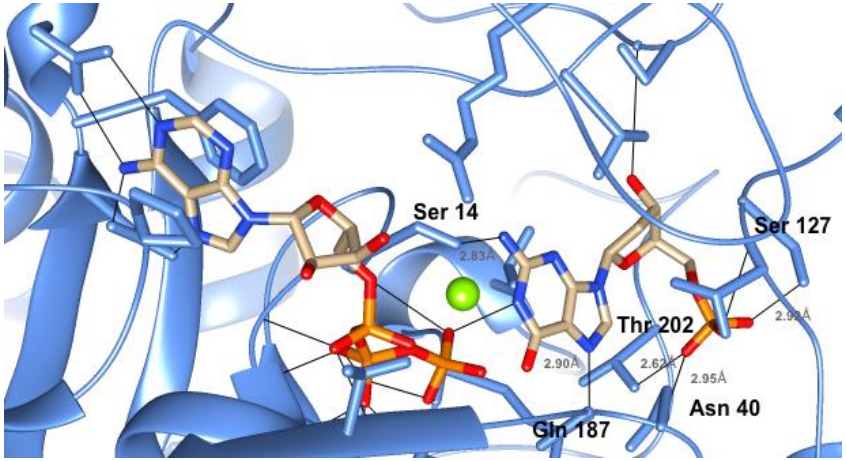




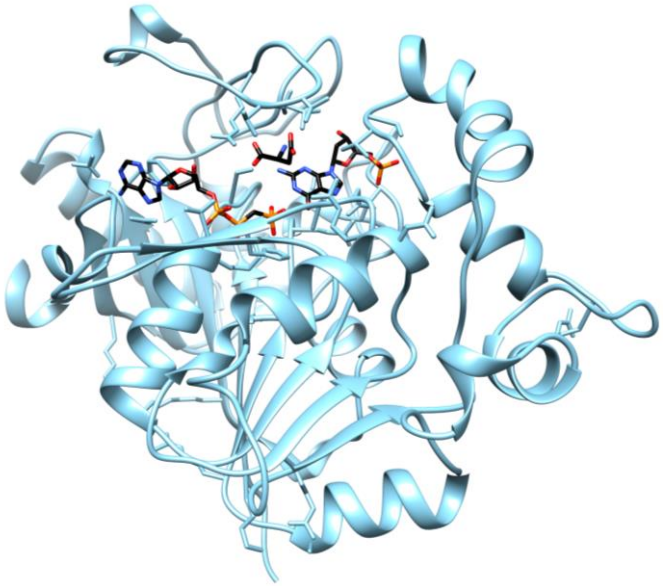


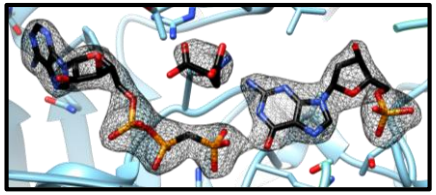
X-Ray structure of ΦVC8 PurZ in complex with ATP and dGMP



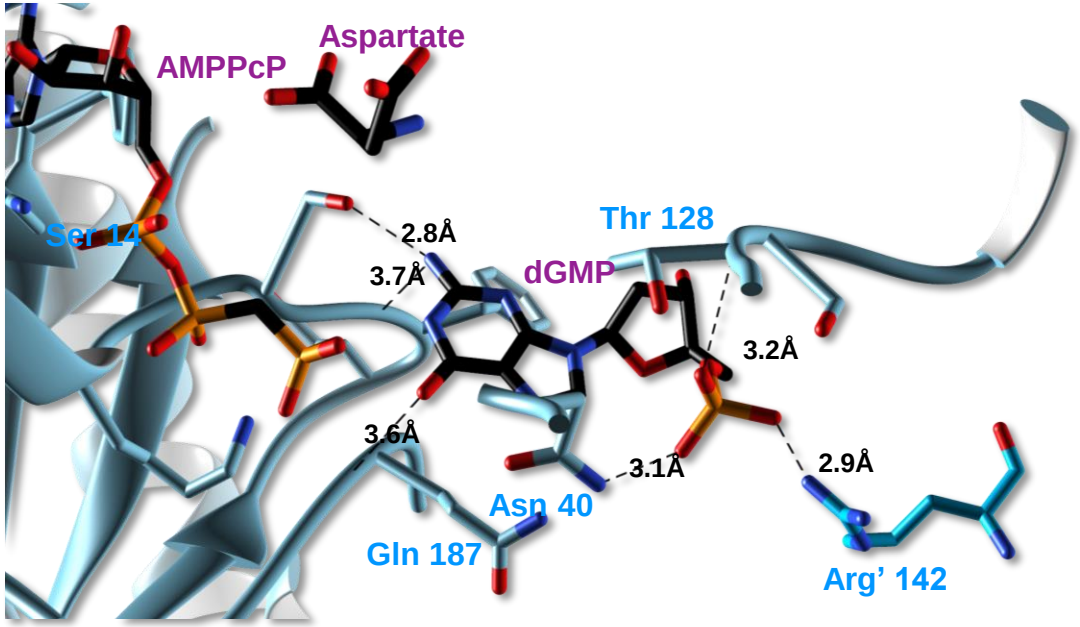


X-Ray structure of ΦVC8 PurZ with dGMP, aspartate and AMPPCP



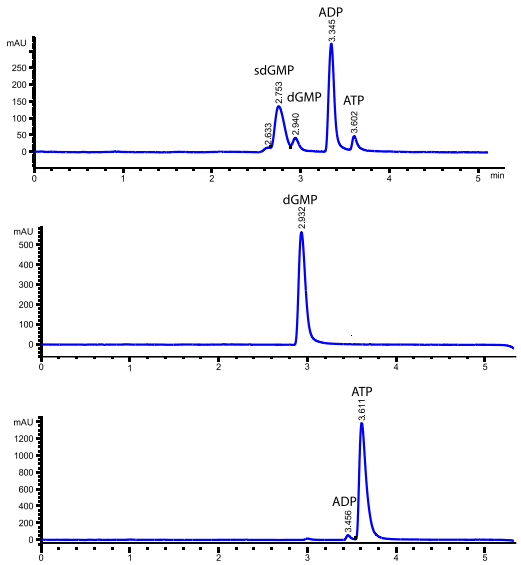


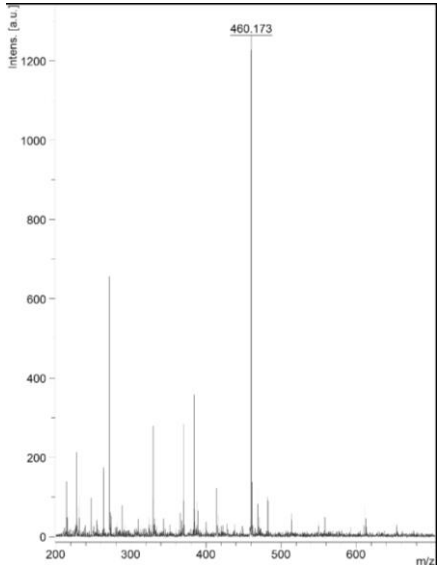
AMPPCP dGMP
L-aspartate

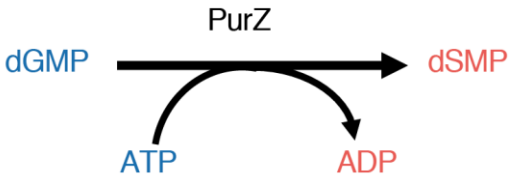


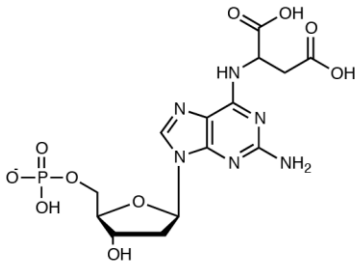
ΦVC8 PurZ codes a N6-succino-2-amino-2'-deoxyadenylate synthase

PurZ reacts only with L-aspartate, dGMP, ATP and Mg^{2+}





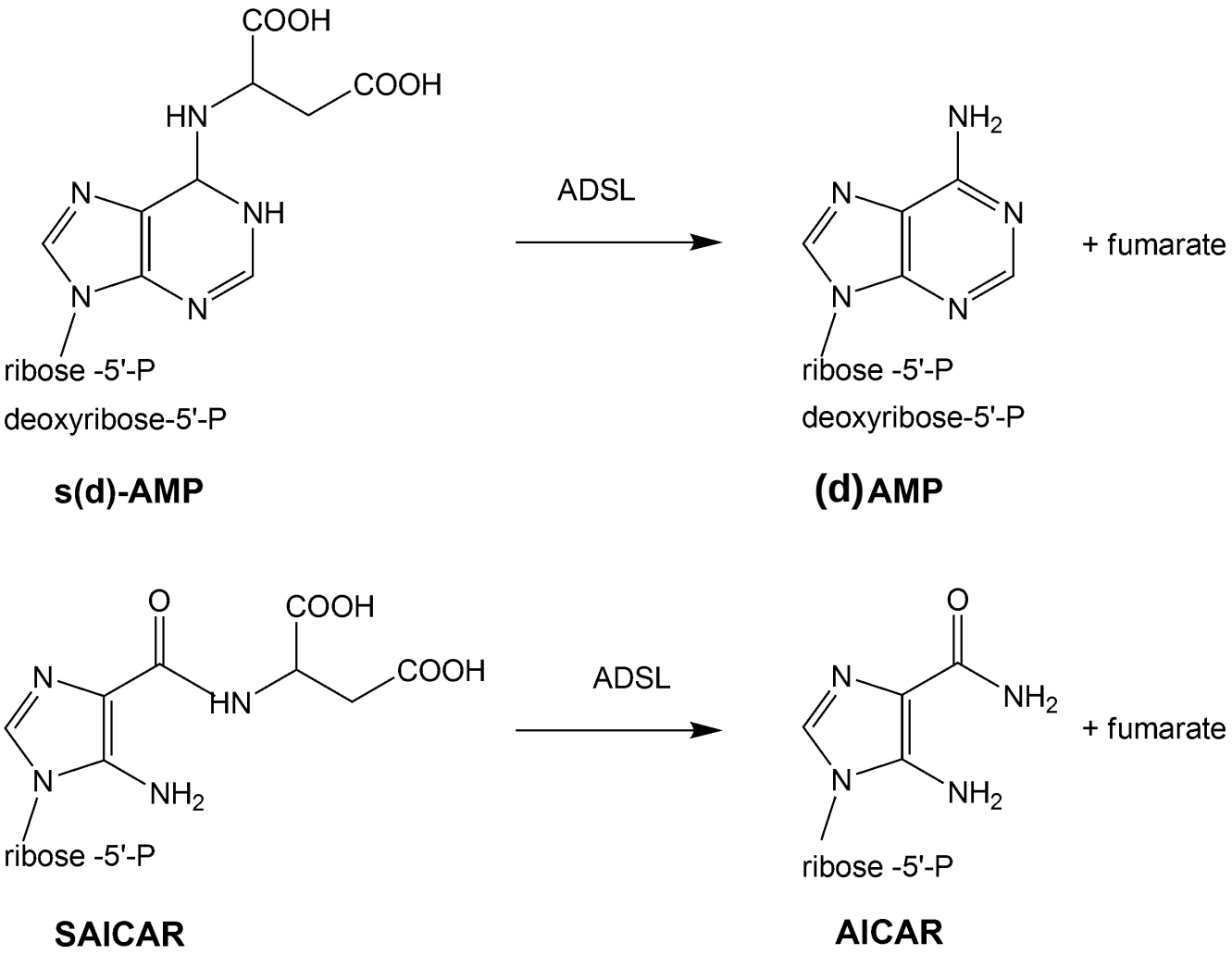




Chemical Formula: $C_{14}H_{18}N_6O_{10}P^-$
Exact Mass: 461.08

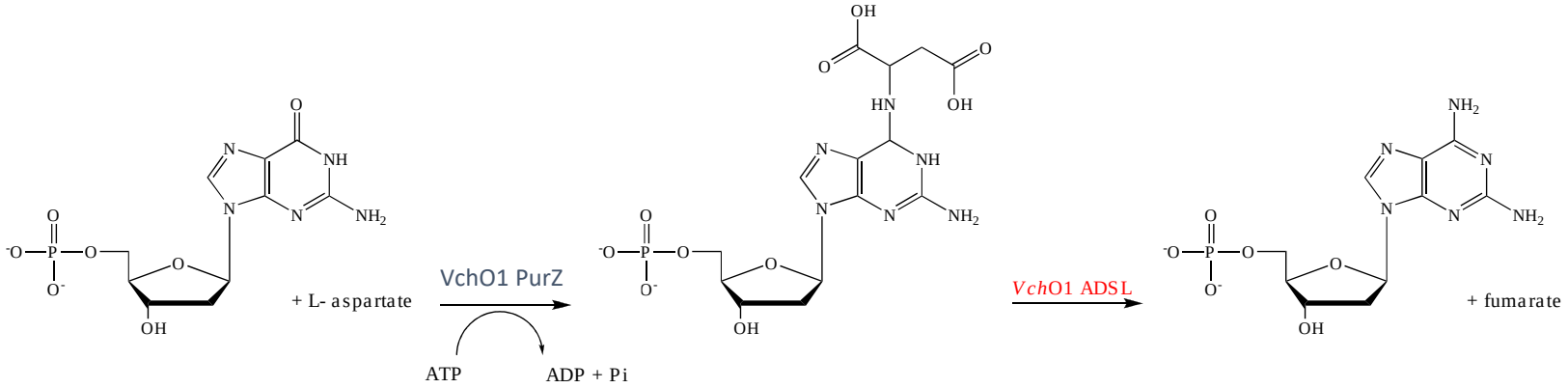
dSMP was also purified at the mg scale and confirmed by NMR

ADSL substrates

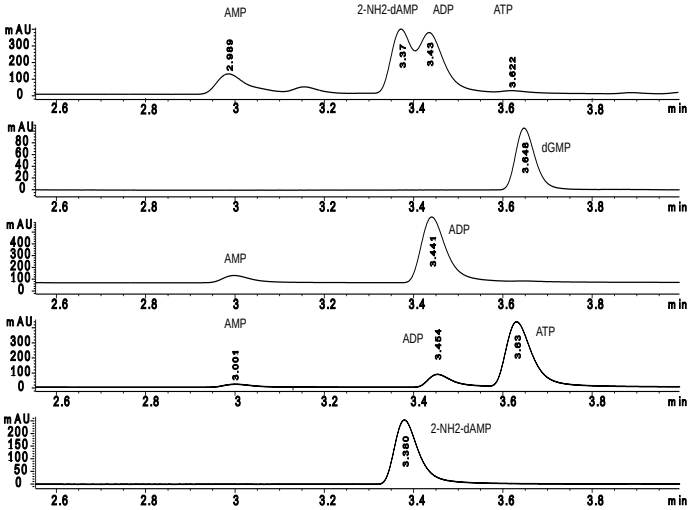


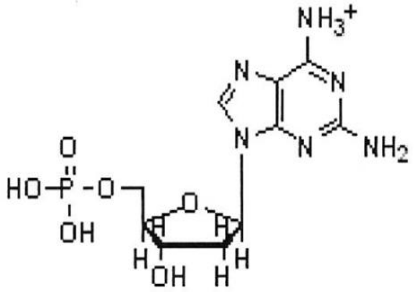
ΦVC8 PurZ and *V. cholerae* O1 ADSL convert dGMP to dZMP

dDMP identification



m/z 347.1

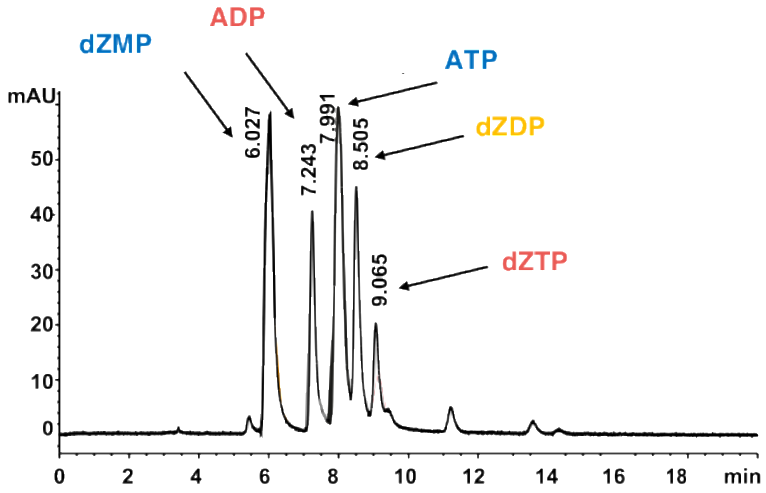


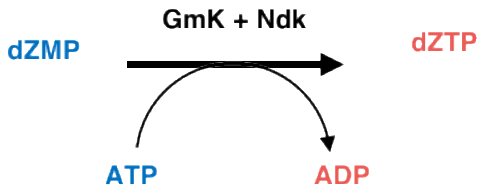


Exact Mass: 347

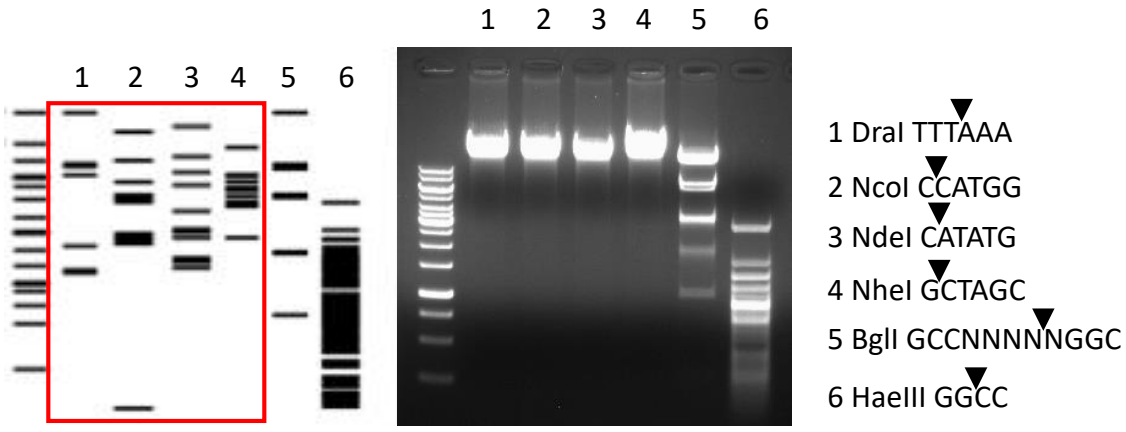
V. cholerae O1 Gmk and Ndk converts dZMP to dZTP

guanylate kinase reaction (d)GMP + ATP -> (d)GDP +ADP
Nucleoside diphosphate kinase reaction (d)NDP + ATP -> (d)NTP + ADP

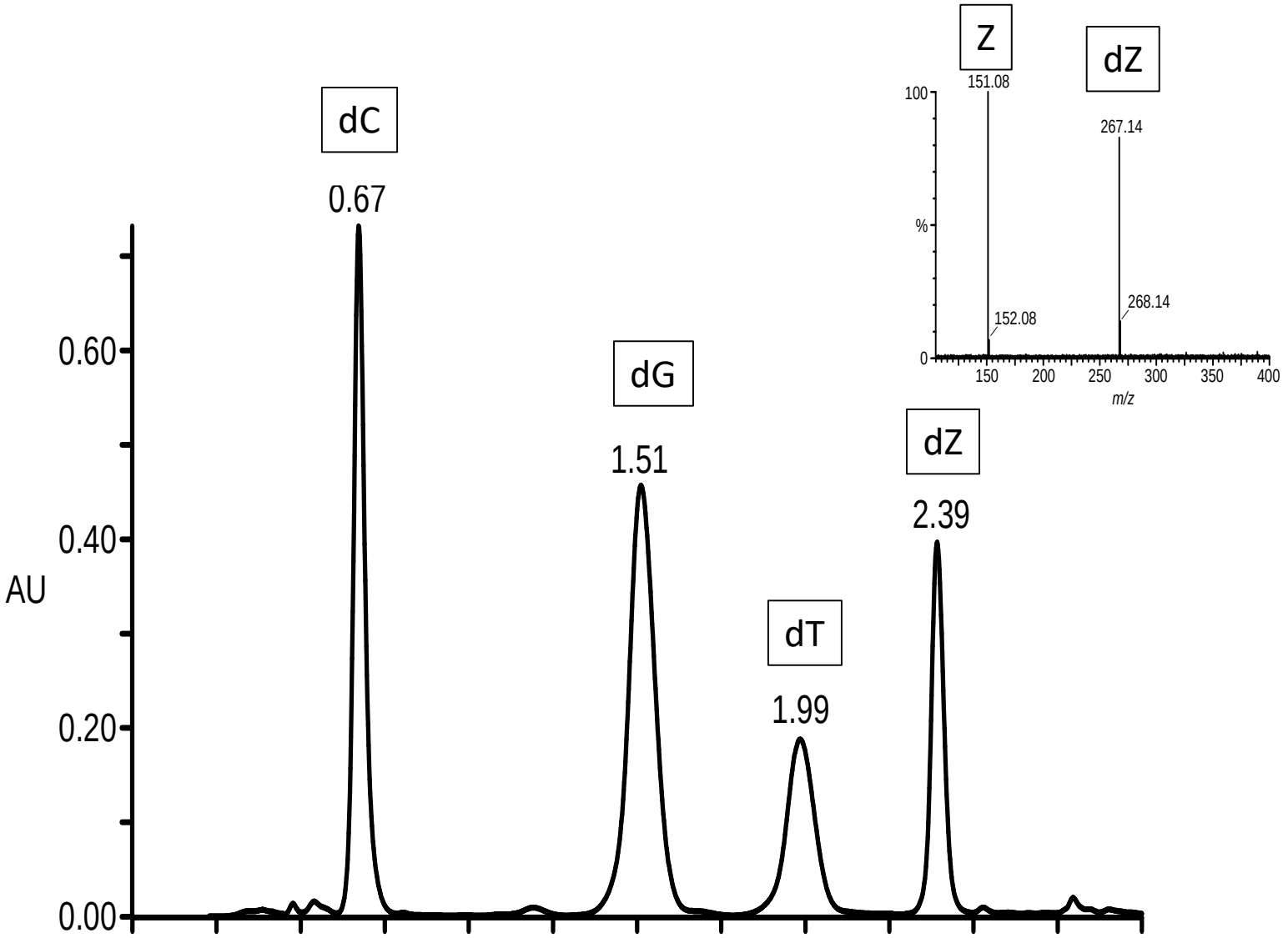




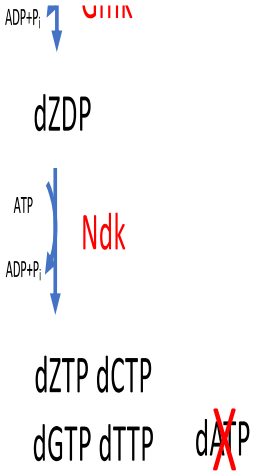
ΦVC8 DNA is resistant to restriction enzymes containing an A in their recognition site



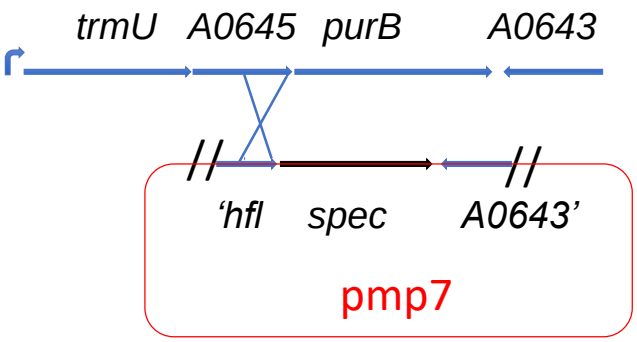
ΦVC8 DNA contains 2,6 diaminopurine

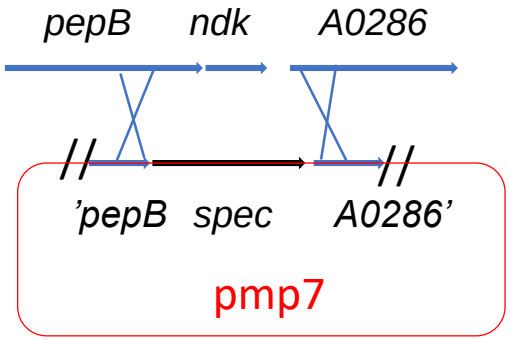


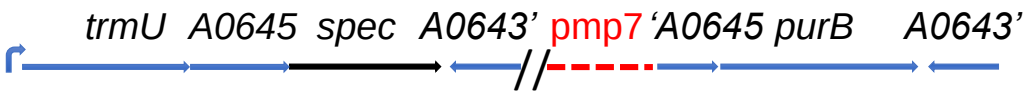
dZTP biosynthetic pathway

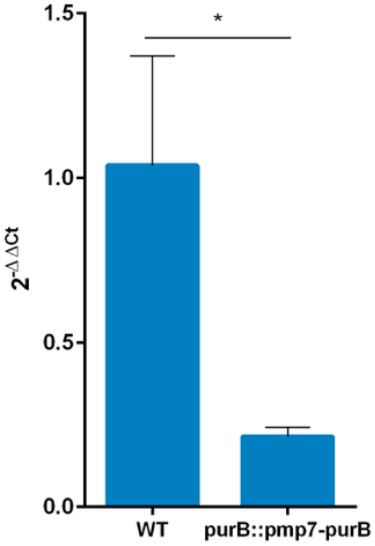


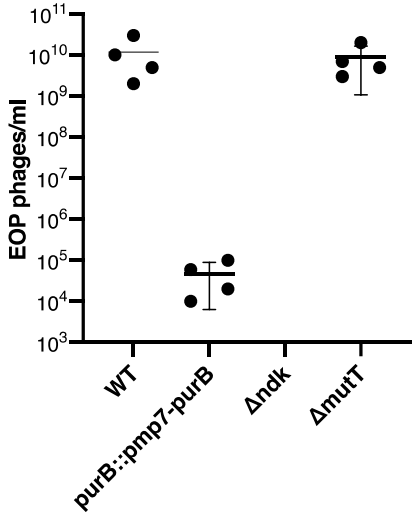
In vivo validation of the pathway









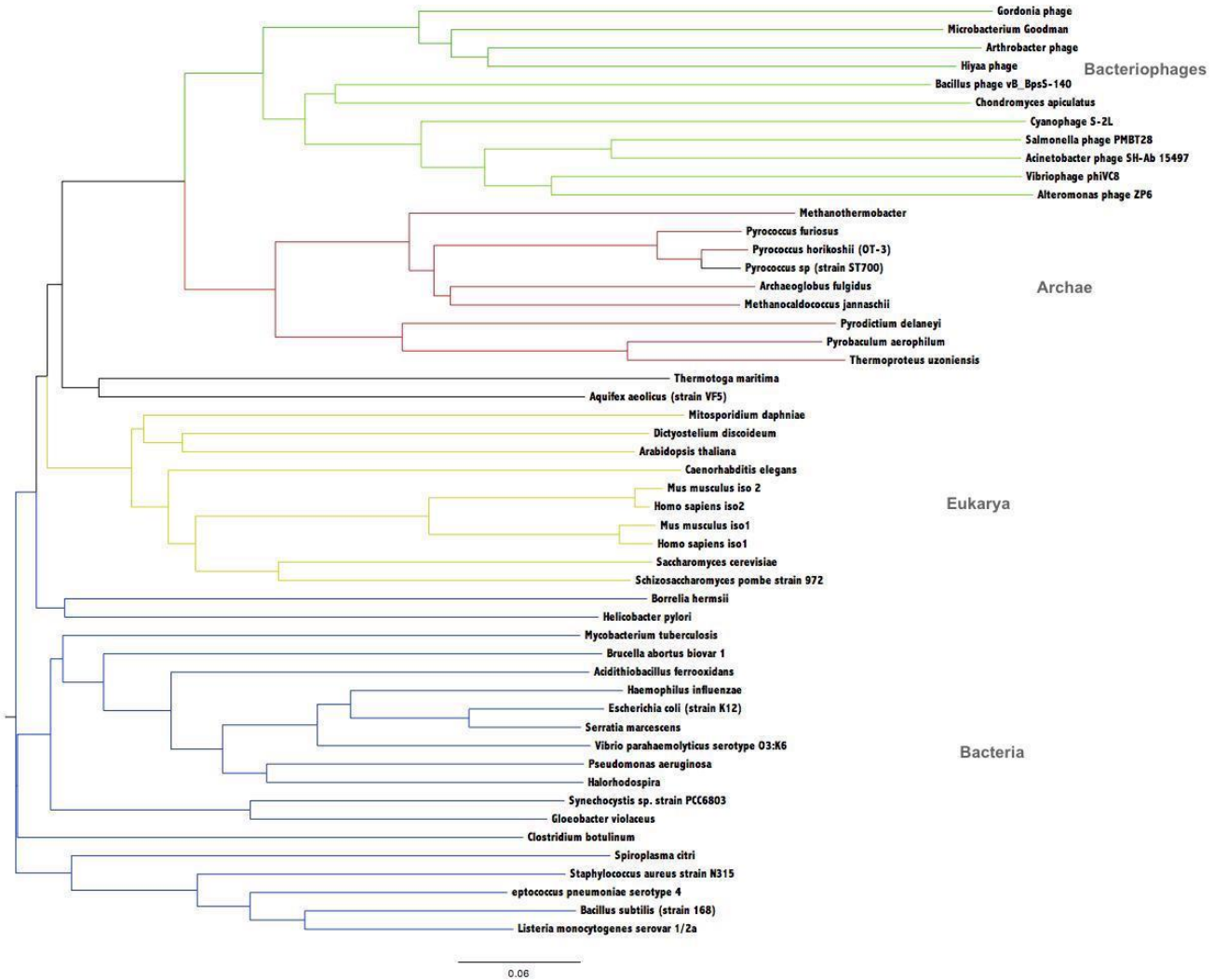


50 fold reduction of *purB* transcript -> 5 log reduction of phage infection

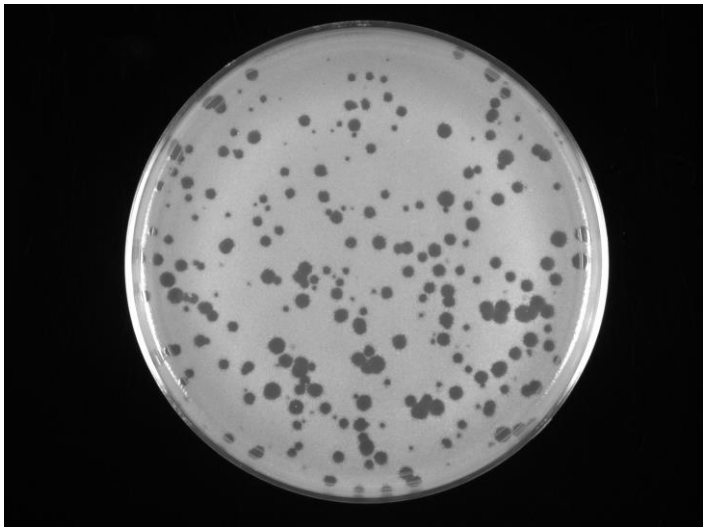
Ndk is essential for phage replication

dZTP biosynthetic pathway may be found in other phages

Phylogenetic tree of adenylosuccinate synthases



The Arthrobacteriophage Wayne

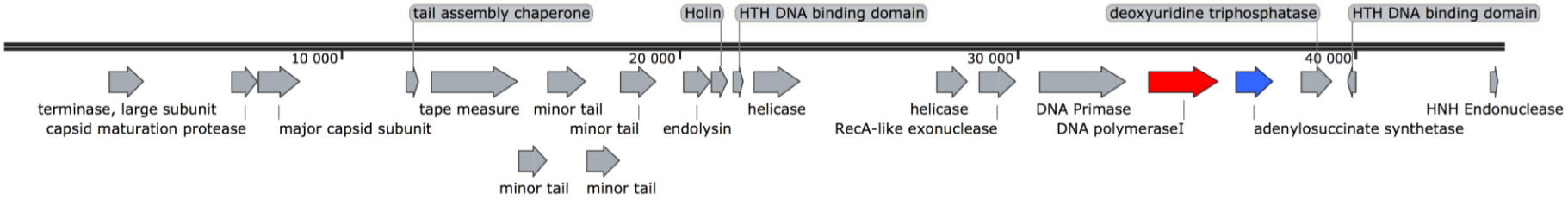


Wayne_48 DNA polymerase I
Wayne_50 Adenylosuccinate synthase

Host ***Arthrobacter sp.*** **ATCC 21022**
Arthrobacter spp. are common soil bacteria classified within the family Micrococcaceae in the order Actinomycetales,

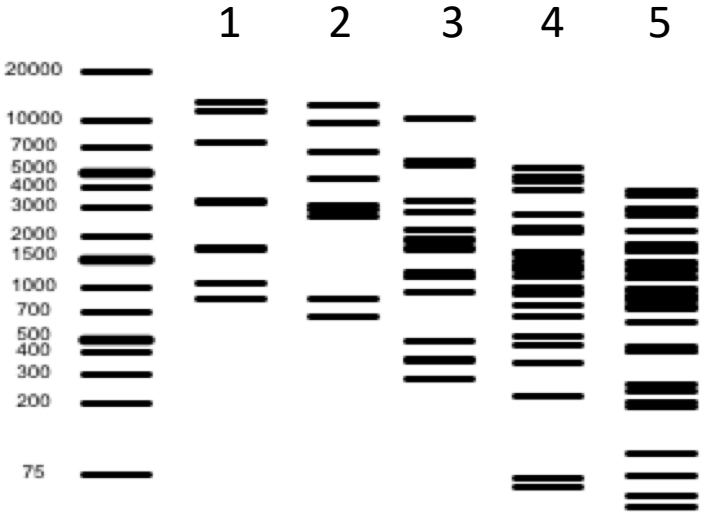
Coll with D. Jacobs-Sera and GF. Hatfull
PBI USA

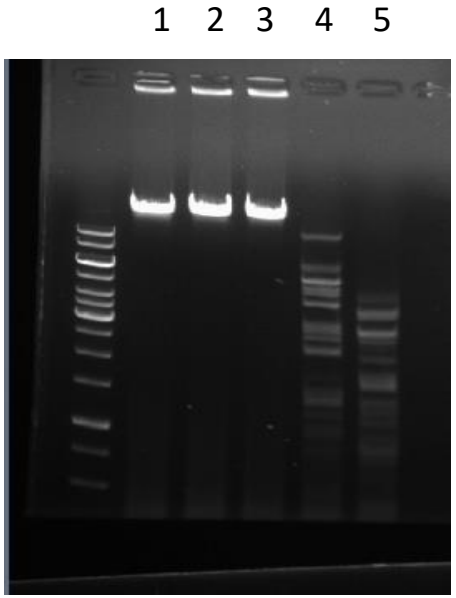
Genome map of Wayne



Wayne 44370 bp

Restriction of Wayne DNA





1 HindIII AAGCTT

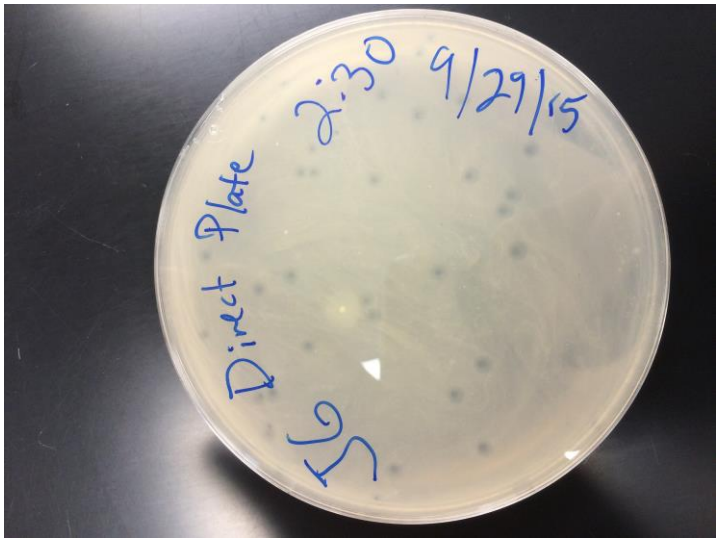
2 BglII AGATCT

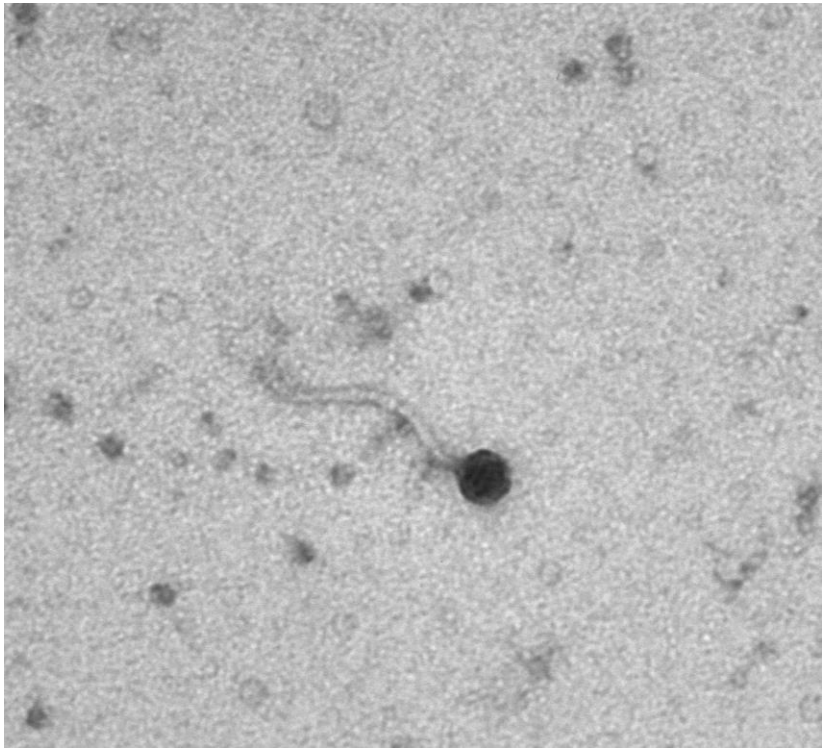
3 NcoI CCATGG

4 BglI GCCNNNNNGGC

5 XmaI CCCGGG

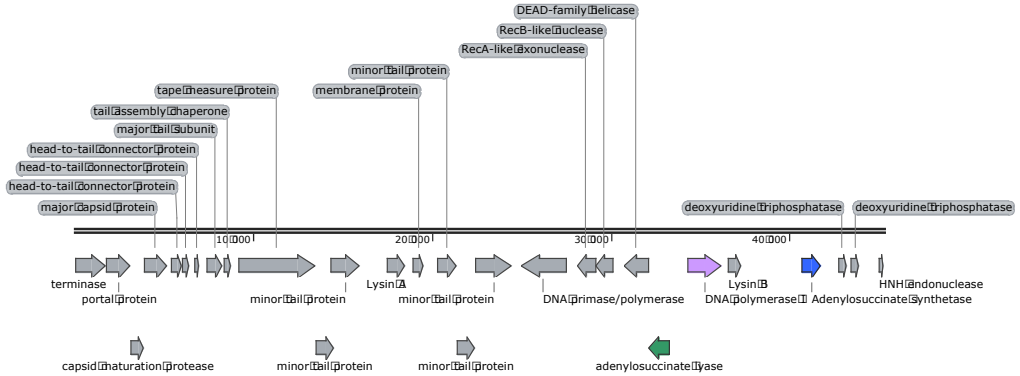
The Ghordoniaphage Ghobes





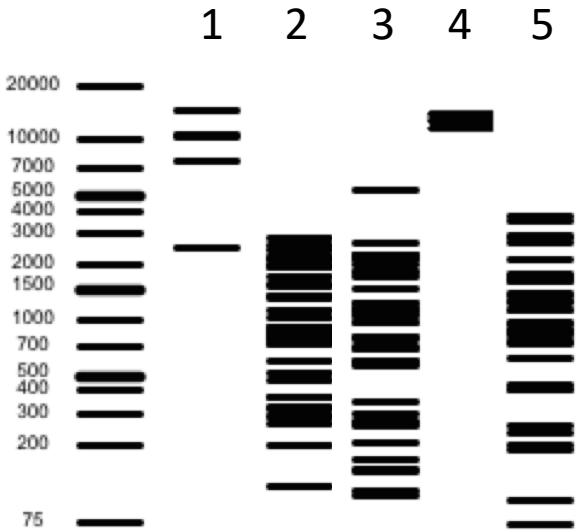
Host: *Gordonia terrae* 3612

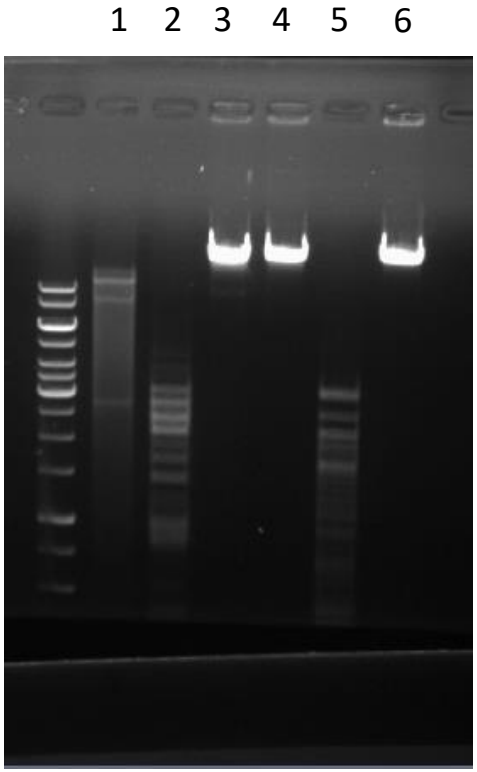
Genome map of Ghobes



Ghobes 45285 bp

Restriction of Ghobes DNA





1 BamHI GGATCC

2 XhoI CTCGAG

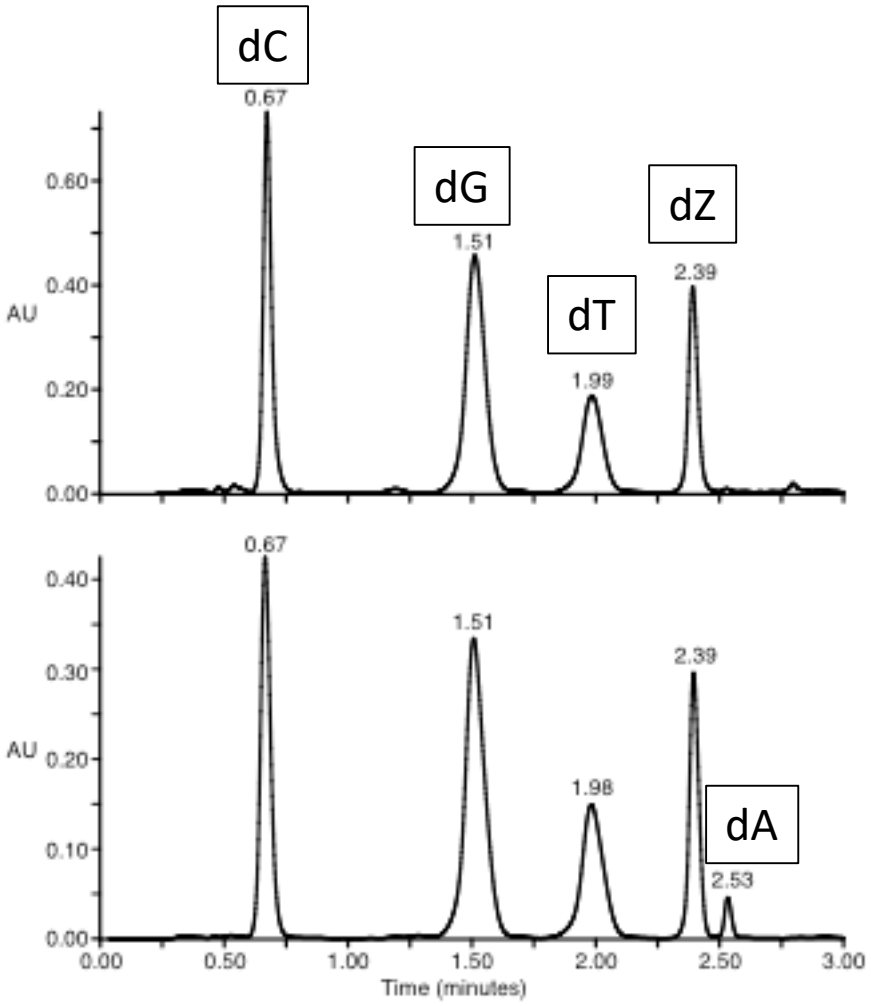
3 SalI GTCGAC

4 HindIII AAGCTT

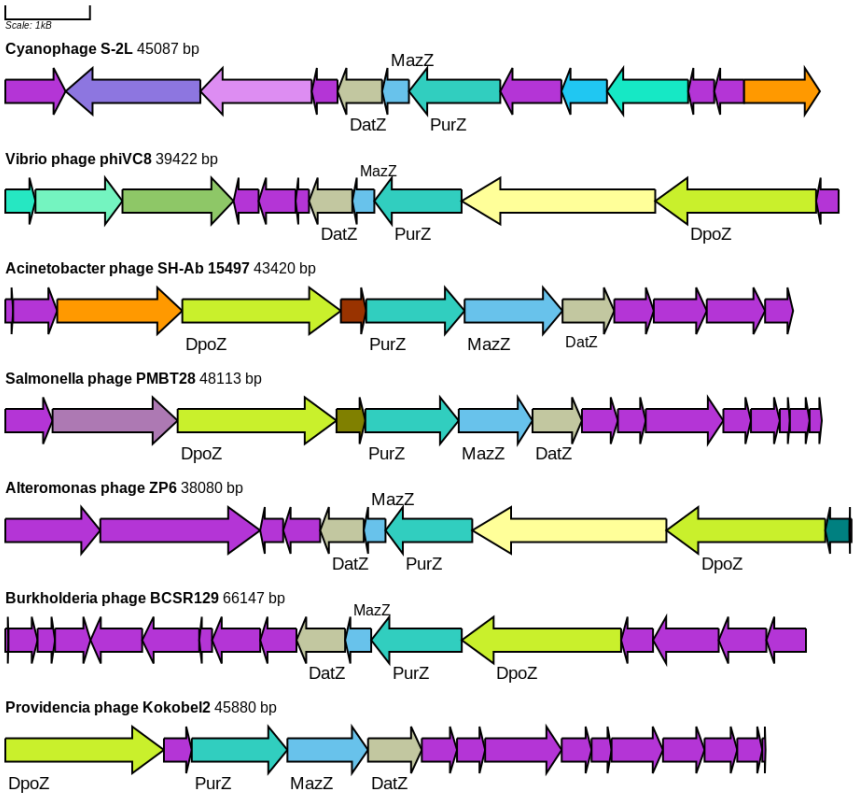
5 BglI GCCNNNNNGGC

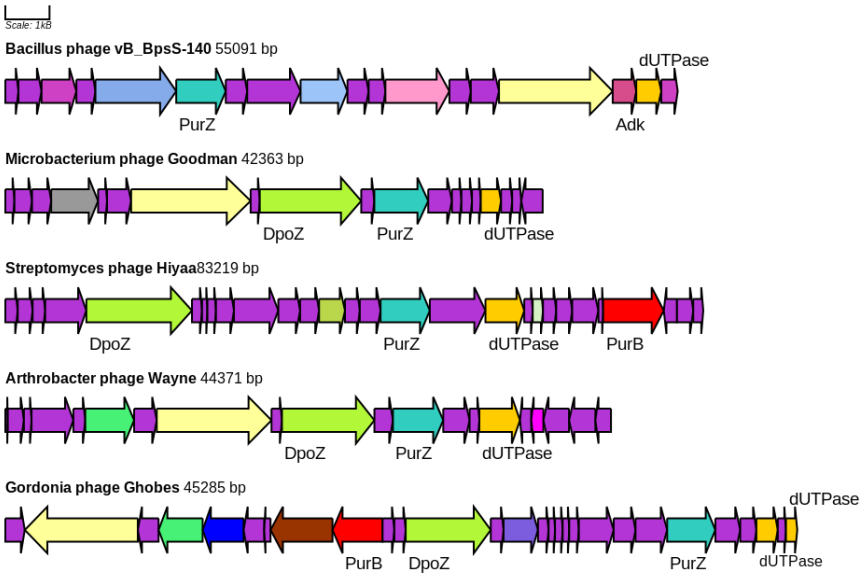
6 XmaI CCCGGG

Wayne and Ghobes DNA contain 2-amino-dATP



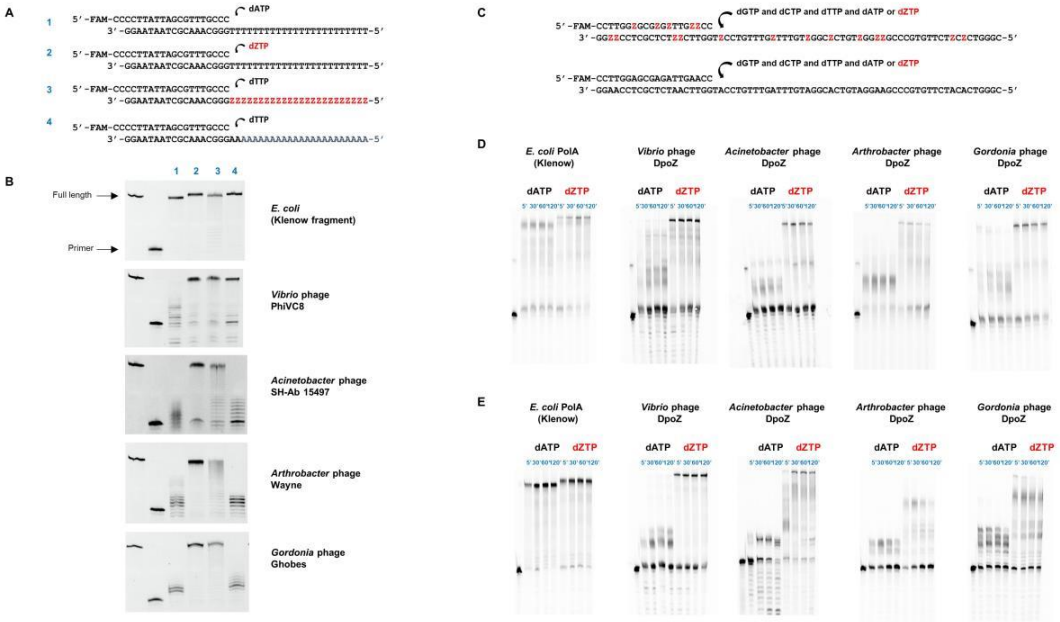
Genomic maps of the regions surrounding PurZ gene of bacteriophages





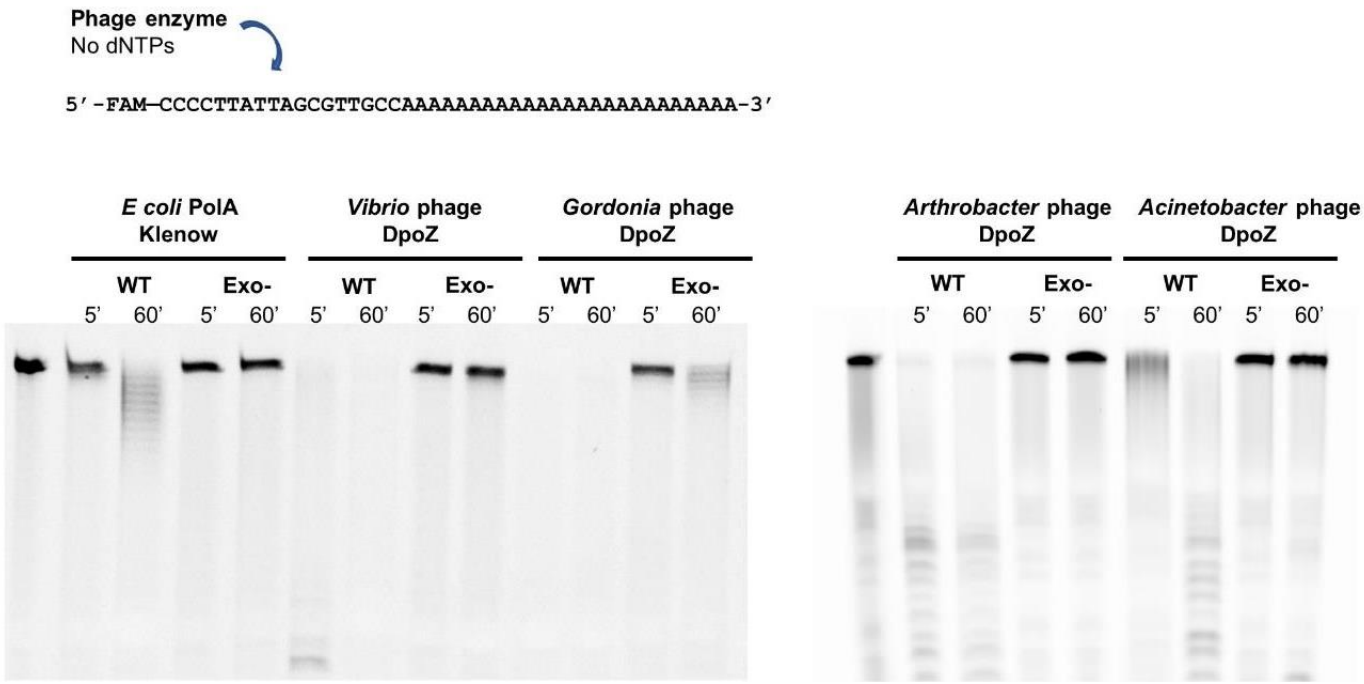
Phage DNA polymerase discriminates between dZTP and dATP

Discrimination between aminoadenine and adenine by DNA polymerases



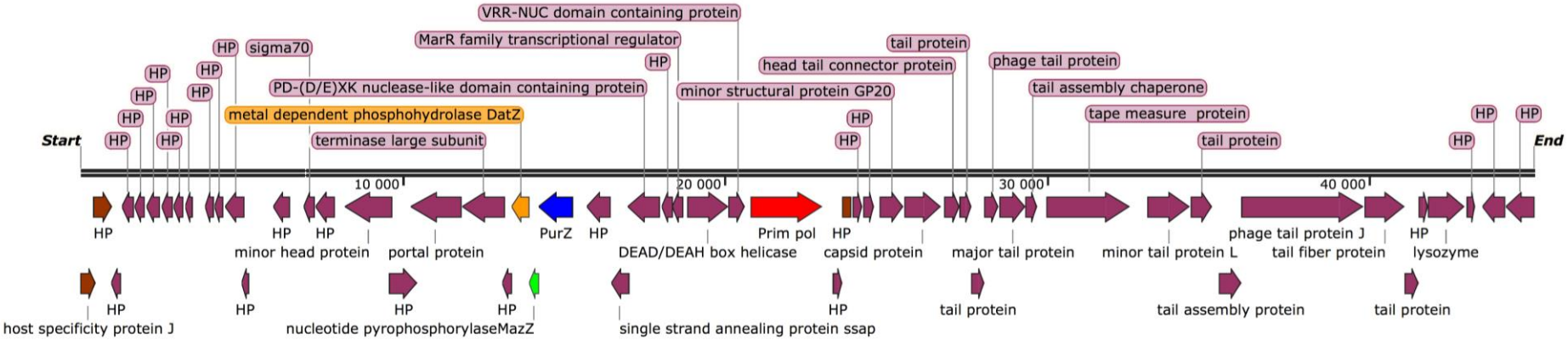
Primer extension assays were performed using purified His-tagged polymerases,
A/ Experimental setup showing nucleotide sequences of duplexes between a fluorescent (FAM) labeled primer (X1903) and a homopolymer template (X1904, (dT)24 ; X1930, (dA)24 ; X2009, (d Z)24).
B/ Elongation products polymerized during 30 min at 37°C were loaded on denaturing 17% polyacrylamide gels. Lane numbers refer to conditions indicated in A.
C/ Experimental setup using duplexes between a fluorescent primer and a heteropolymer template corresponding to a 50 nucleotide sequence from the SH-Ab 15497 phage genome.
D/ Elongation products polymerized from Z-containing template and primer
E/ Elongation products polymerized from A-containing template (X2364) and primer

The exo domain of ΦVC8 DNA polymerase discriminates A and Z

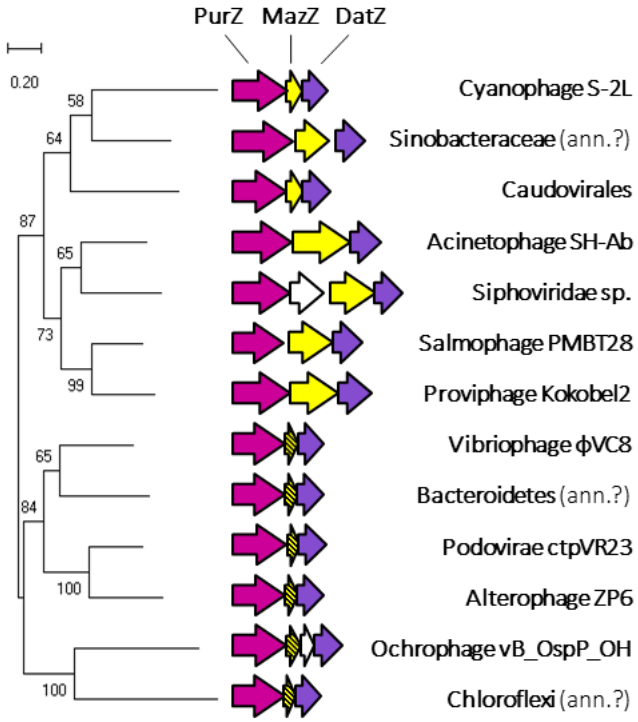


Exonuclease activity of the Klenow fragment, the phiVC8 and Gordonia and Arthrobacter phage DNA polymerase on single strand DNA

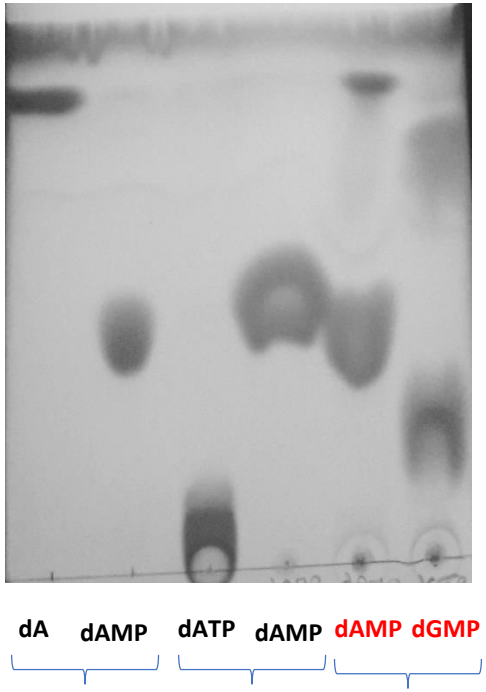
NGS of S-2L genome

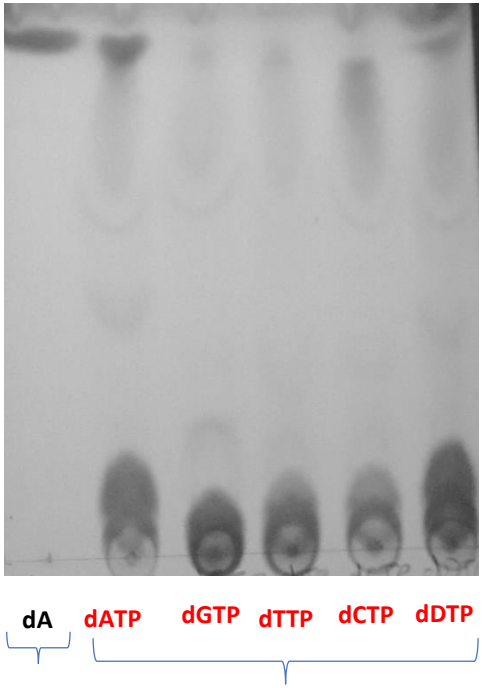


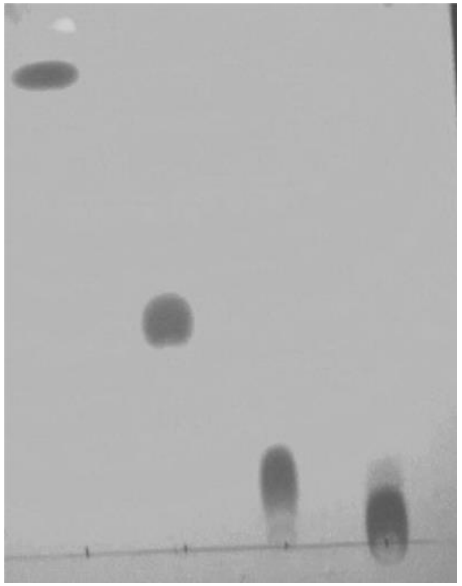
DatZ is a dATP triphosphohydrolase



ΦVC8 DatZ hydrolyzes dATP into dA

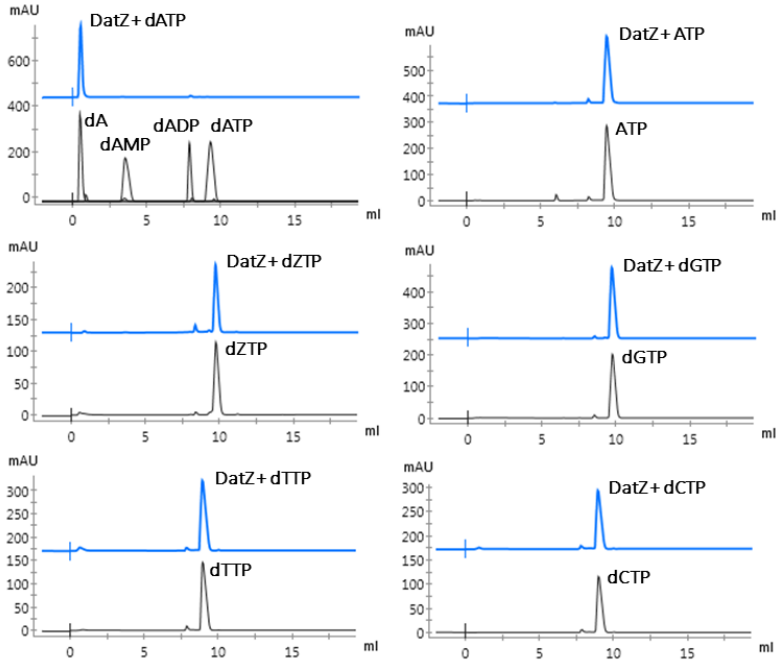






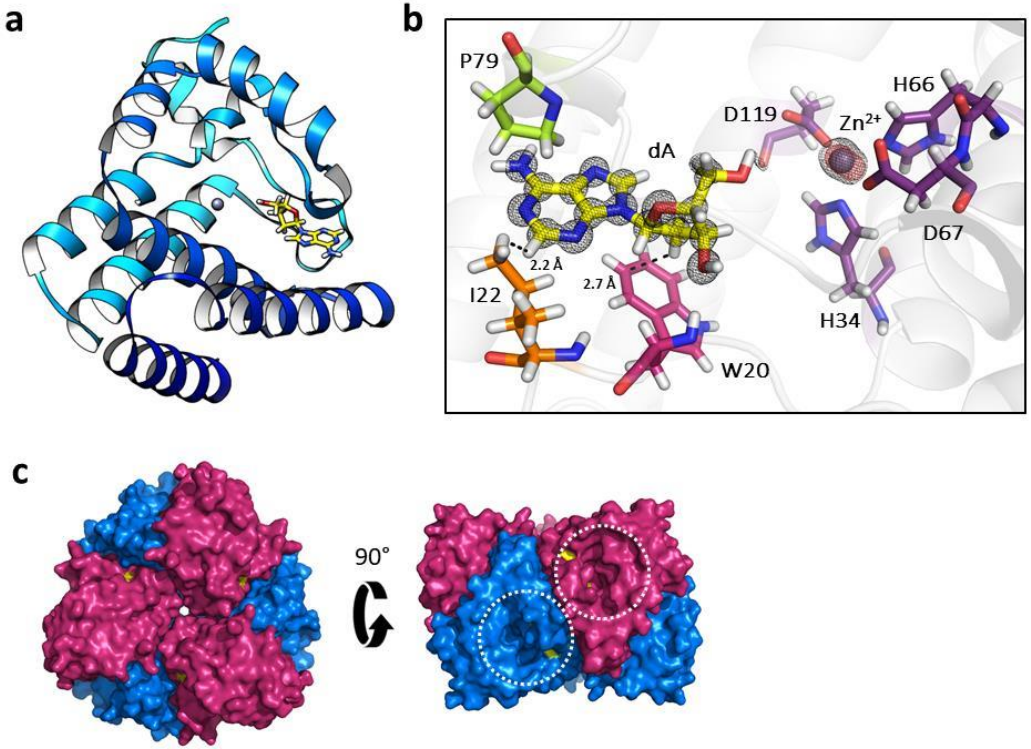
metal dependent phophohydrolase activity

S-2L DatZ



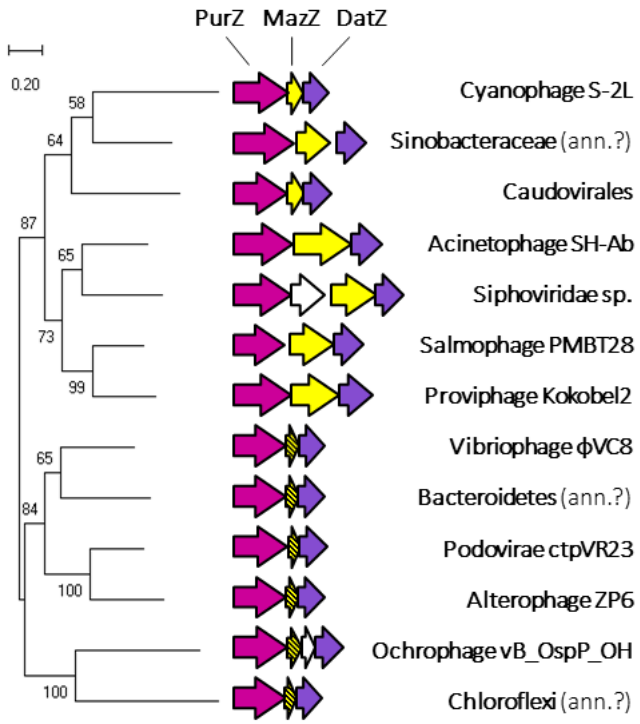
The enzyme is active exclusively with dATP and removes from it all phosphates
DatZ is a dATP triphosphohydrolase

Three-dimensional structure of S-2L DatZ

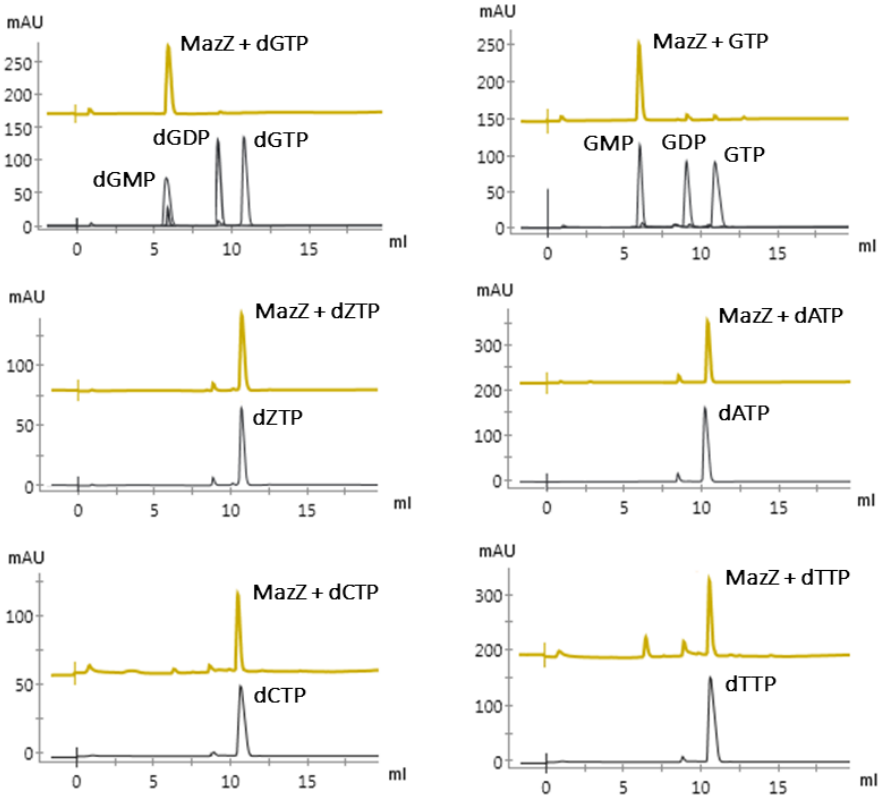


a. Ribbon representation of a DatZ monomer in a light blue-dark blue gradient, with bound dA in stick
b. catalytic pocket of DatZ with the experimental electron density contoured at 2.5 sigmas around bound ligands: dA and Zn2+
c. Structure of the full DatZ hexamer, top and side views, in surface representation.

PurZ MazZ DatZ conservation



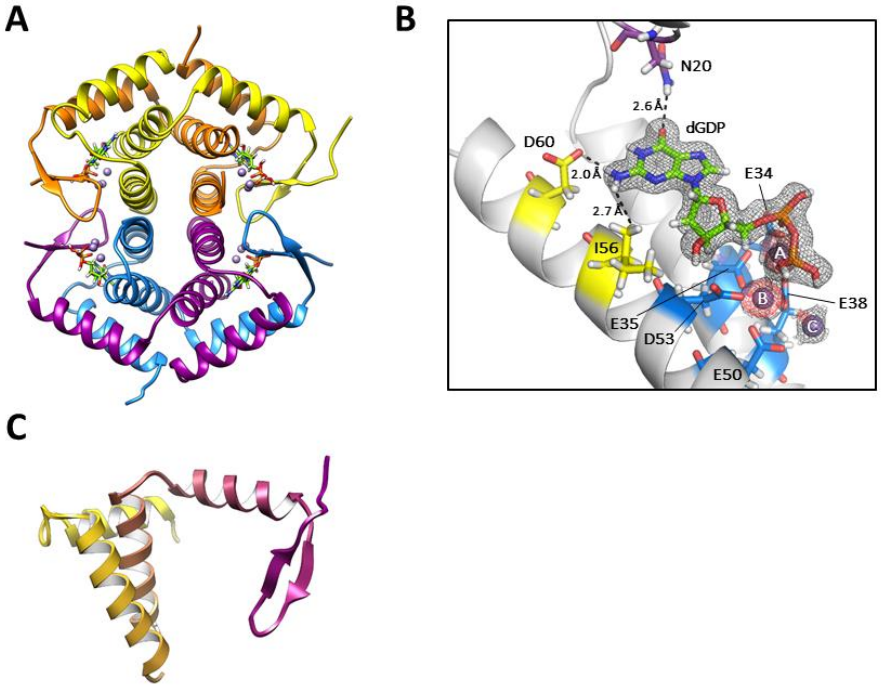
MazZ a dGTP pyrophosphorylase



MazZ is selective towards dGTP and GTP, removing their two terminal β- and γ-phosphates.

Structure of S-2L MazZ with bound dGDP and Mn^{2+} ions

Fold found in MazG(-like) and HisE enzymes

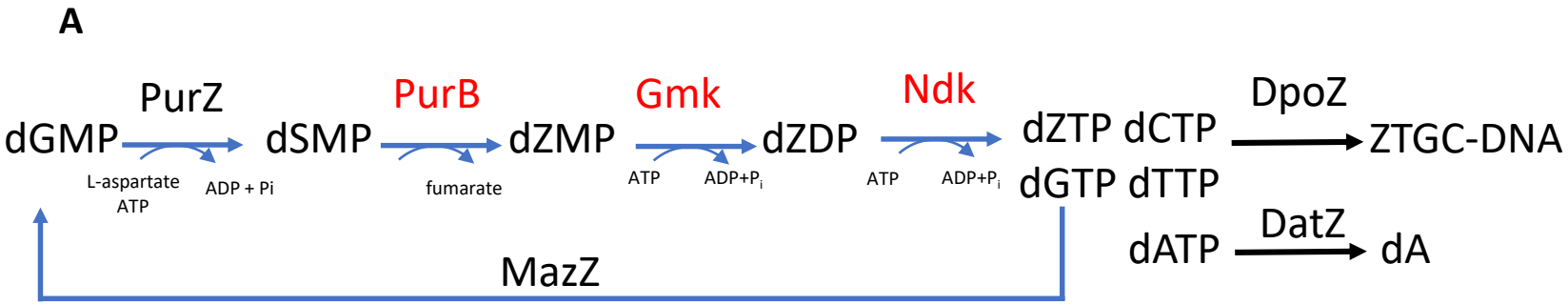


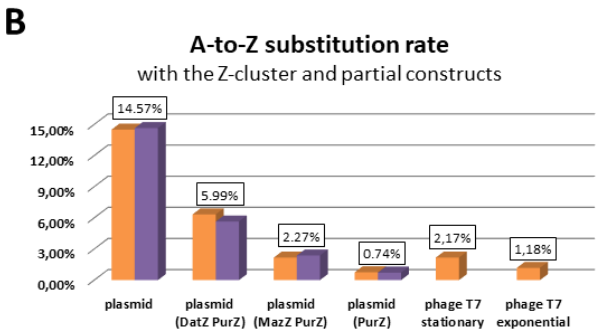
A Tetramer of MazZ. Two tight dimers (yellow:orange and blue:purple) further form a dimer;
each of the four catalytic pockets with the reactant (lime) and three catalytic ions is created from the two chains of a tight dimer.

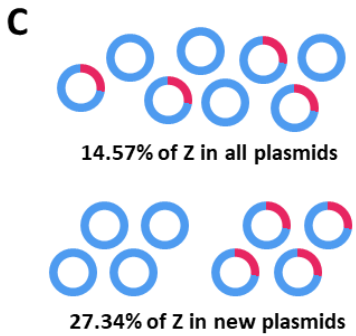
B. Close-up on the catalytic pocket. The product of dGTP dephosphorylation (lime) is identified as dGDP in the crystal, next to the catalytic Mn^{2+} ions.
The determinants of guanine specificity (yellow for N2 and purple for O6) and residues coordinating the ions (blue) are placed on one protein chain.
The three Mn^{2+} ions (lilac spheres), designated A, B and C, are hexa-coordinated by the negatively charged protein residues, deoxynucleotide phosphates and water molecules

C. A single protein chain of MazZ, coloured in yellow-purple gradient.

Biosynthesis of Z and Exclusion of A in Z containing phages







Open questions

- **Is Z-DNA able to resist antibacterial defence systems?**
- **What proteins are needed for Z-DNA replication?**
- **Can we obtain a genome from a cellular organism in which adenine is systematically replaced by 2-aminoadenine?**