



**HAL**  
open science

# Mechanisms supporting aminoadenine-based viral DNA genomes

Pierre-Alexandre Kaminski

► **To cite this version:**

Pierre-Alexandre Kaminski. Mechanisms supporting aminoadenine-based viral DNA genomes. Cellular and Molecular Life Sciences, 2022, 79 (1), pp.51. 10.1007/s00018-021-04055-7 . pasteur-03549987

**HAL Id: pasteur-03549987**

**<https://pasteur.hal.science/pasteur-03549987>**

Submitted on 24 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Mechanisms Supporting Aminoadenine-based Viral DNA Genomes

P.A. Kaminski

<https://orcid.org/0000-0001-9841-6837>

Biologie des Bactéries Pathogènes à Gram-Positif, Institut Pasteur, CNRS-UMR 2001, Paris, France.

[pierre-alexandre.kaminski@pasteur.fr](mailto:pierre-alexandre.kaminski@pasteur.fr)

running title: ZTGC-DNA

keywords: Aminoadenine, bacteriophage, DNA, adenylosuccinate synthase, DNA polymerase, nucleotide phosphohydrolase

### Abstract

Bacteriophage genomes are the richest source of modified nucleobases of any life form. Of these, 2,6 diaminopurine, which pairs with thymine by forming three hydrogen bonds violates Watson and Crick's base pairing. 2,6 diaminopurine initially found in the cyanophage S-2L is more widespread than expected and has also been detected in phage infecting Gram-negative and Gram-positive bacteria. The biosynthetic pathway for aminoadenine containing DNA as well as the exclusion of adenine are now elucidated. This example of a natural deviation from the genetic code represents only one of the possibilities explored by nature and provides a proof of concept for the synthetic biology of non-canonical nucleic acids.

## Introduction

Genetic information is generally encoded by four bases: adenine (A), guanine (G), cytosine (C) and thymine (T). In addition to the canonical ACGT bases, modified bases are found in large numbers in RNA, reflecting the wide range of functions that RNA molecules perform within cells. DNA also contains a smaller number of modified bases and we will mainly mention the methylated bases m4dC, m5dC, m6dA, hm5dU/hmU, and hm5dC/hmC to which we can add the epigenetic bases m5C, hm5C, f5C and ca5C identified in the DNA of higher organisms[1]. In prokaryotes, DNA methylation at the C-5 or N-4 cytosine (m5C, m4C) and at the N-6 adenine (m6A) also plays an important role, as these modifications discriminate between endogenous and exogenous DNA and help to maintain genetic integrity. DNA base methylation is also involved in the initiation of replication, gene expression which is cell cycle dependent for *Caulobacter crescentus* and other proteobacteria [2].

Bacteriophages, viruses that infect bacteria, contain methylated bases but also other modified bases that are not found in any other organism. These modified bases play a role in resistance strategies to bacterial restriction systems. These base modifications can be pre-replicative, i.e. by the enzymatic modification of DNA precursors: deoxynucleotide monophosphates (dNMPs). The modified dNMPs are then converted into modified deoxynucleotide triphosphates (dNTPs) and incorporated into the bacteriophage DNA by a DNA polymerase. Some bacteriophages that have already incorporated modified nucleotides into their DNA can modify them a second time, known as post-replicative modification or "hypermodification". Purines and modified pyrimidines have been identified in bacteriophage DNA, but it is the pyrimidines that show the greatest diversity: uracil (U), 5-hydroxymethyluracil(hm5U), 5-hydroxymethyldeoxyuracil (hm5dU),  $\alpha$ -glutamylthymine,  $\alpha$ -putrescinyllthymine, 5-dihydroxypentyluracil replace thymine in the DNA of Bacillus SP10 and Delftia W-14 bacteriophages, and 5-(2-aminoethoxy)methyluridine (NeOm5dU) and 5-(2-aminoethyl)uridine (Ne5dU) in the DNA of Vil and M6 phages [3], 5-methylcytosine (m5C), 5-hydroxycytosine, 5-hydroxymethylcytosine (hm5C) and glucosylhydroxymethylcytosine completely replace cytosine in the genomes of *Escherichia coli* T bacteriophages [4]. Biosynthetic pathways have been described for some of these bases: dU, hm5dU, NeOm5dU, Ne5dU, hm5C and glucosyl-m5C [3, 5].

Eight modified purines have been described so far: N-carbamoyl-methyladenine (*E. coli* bacteriophage Mu)[6] , deoxyarchaeosine (*E. coli* phage 9g) [7], 7-methylguanine (*Shigella sonnei* phage) [8], N6-methyladenine (*E. coli* lambda phage) [9], deoxyinosine and 2'-deoxy-7-amido-7-deazaguanosine (*Campylobacter* viruses) [10], 7-cyano-7-deazaguanine (*E. coli* phage CAjan [11] and 2-aminoadenine (*Synechococcus* phage S-2L) [12, 13]. These bases are present in variable proportions depending on the genome, up to 100%, i.e. a total replacement of the corresponding canonical base, which is the case for 2-aminoadenine, which substitutes for adenine [4, 12].

### **2-aminoadenine and cyanophage S-2L**

2-aminoadenine, or 2,6-diaminopurine, was originally synthesized and used as an adenine analogue to study the structural properties of nucleic acids. The 2-aminoadenine-thymine pair affects the local flexibility of DNA and prevents interaction with helix-bending proteins. Interaction with small molecules such as antibiotics or anticancer drugs is also impaired [14, 15]. 2-aminoadenine has been detected in the DNA of S-2L cyanophage [12, 13] indicating that its presence in DNA is consistent with normal DNA function.

Cyanophage S-2L was isolated from water samples near Leningrad in 1976. It has an icosahedral capsid 56 nm in diameter and a flexible non-contractile tail 120 nm in length. It lyses a limited number of cyanobacteria of the genus *Synechococcus*: *S. sp. 698*, *S. elongatus* 58 and *S. elongatus* 6907. Analysis of its DNA composition by acid or enzymatic hydrolysis showed that adenine was absent and completely replaced by 2-aminoadenine. The presence of this base stabilizes the secondary structure of the DNA because it forms a third hydrogen bond with thymine through its amine group (Figure 1) which enhances the melting temperature by 3.6 degrees as compared to adenine-DNA of equivalent base composition. In addition, 2-aminoadenine confers resistance to restriction enzymes that have an adenine in their recognition site [16].

The genome of this bacteriophage was sequenced on the initiative of Philippe Marlière (then at the Institut Pasteur), in collaboration with the Genoscope (GenBank AX955019.1). Analysis of this genome revealed the presence of a gene coding a protein with similarities to adenylosuccinate synthase (PurA). In all living organisms, PurA catalyzes the conversion of

inosine 5'-monophosphate (IMP) to adenylosuccinate (SMP) in the presence of GTP, Mg<sup>2+</sup> and L-aspartate, which is the first step in the de novo biosynthesis of adenosine 5'-monophosphate (AMP) (Figure 2) [17].

The adenylosuccinate synthase-like of S-2L which we have named PurZ, could be the first enzyme of the 2-aminoadenine biosynthetic pathway. Indeed, if 2-aminoadenine is formed after replication it would require a break in the hydrogen bonds between bases and an opening of the double helix, which seems unlikely.

### **2-aminoadenine biosynthetic pathway**

The cyanophage has long remained a somewhat exotic scientific curiosity, but search into the databases using the PurZ sequence of S-2L as a query identified recently proteins with approximately 40% identity encoded by the genomes of the phages PhiVC8, J2, JSF33, VP5, QH and JF15. These phages are all lytic for *Vibrio cholerae*. PhiVC8 was isolated from water samples and belongs to the order Caudovirales and the family Podoviridae. It has a hexagonal capsid and a short tail [18].

Its genome is a double-stranded linear DNA of 39,422 base pairs (bp) containing a gene encoding PurZ and a DNA polymerase belonging to the PolA or PolI DNA polymerase family. These two observations suggest that the PhiVC8 DNA might contain 2-aminoadenine, the biosynthesis of which must be encoded by the phage, and that a dedicated DNA polymerase can incorporate the corresponding deoxyribonucleotide. In *Vibrio cholerae*, unlike *Synechococcus*, the genetic tools exist for the engineering of the strain which is why we decided to focus on the phiVC8 *V. cholerae* system. We started by characterizing the activity of PurZ enzymes.

The PurZs of S-2L and PhiVC8 are not functional equivalents of adenylosuccinate synthases because their expression does not restore the auxotrophy for adenine of an *E. coli* purA mutant, unlike the PurAs of the hyperthermophilic archaeobacterium *Pyrococcus* sp. strain ST700 [19], the parasite *Plasmodium falciparum* [20] or even human [21].

As *P. horikoshii* PurA and PhiVC8 PurZ are not functional equivalents, we hypothesized that PurZ from PhiVC8 must catalyse the same reaction as PurA but with different substrates. PurA catalyzes the transfer of the  $\gamma$ -phosphate of GTP to O6 of IMP, which is displaced by aspartate to form [22].

In the presence of deoxyguanosine 5'-monophosphate (dGMP), adenosine 5'-triphosphate (ATP) and L-aspartate as an amine group donor, PurZ from PhiVC8 and PurZ from S-2L catalyzes the formation of a product with a m/z of 462 as analyzed by LC-MS. This product was purified at the mg-scale quantities, the product was subjected to NMR determination, and formally demonstrated to correspond to N6-succino-2-amino-2' deoxyadenylate (dSMP), a compound that has never been described before. Therefore, PurZ was named N6-succino-2-amino-2' deoxyadenylate synthase [23]. PurZ was identified in 60 bacteriophage isolates, 13 of which were from phage contigs in metagenomes suggesting that Z is more widely distributed than previously thought [24].

Crystallization of PurZ from PhiVC8 in its apo form shows identical folding to that of PurA from *Pyrococcus horikoshii* (Fig3A). This structural similarity is reflected in the phylogeny of adenylosuccinate synthetases. A clear distinction can be made between a branch containing archaeal PurAs and phage PurZs of eukaryotic and bacterial PurAs (Fig3B). These observations are in favour of an ancient origin of PurZs.

The structures of PurZ in complex with the various ligands allowed us to identify determinants of substrate specificity. In particular, Serine 14 and Isoleucine 235 are involved in the specificity of PurZ for dGMP. Interestingly, the aminoacid corresponding to Serine 14 is an aspartate (at position13) residue in *Escherichia coli* PurA which has been shown to be key in the recognition of the substrate IMP, Figure 4 [16]. Furthermore, the Serine-Aspartate change in the *Sinobacteraceae* bacterium phage PurZ (*SbPurZ*) completely abolish the activity confirming the importance of this aminoacid in the binding to dGMP [24].

However, no residues could be identified regarding the specificity of PurZ for dGMP versus GMP and for ATP versus GTP [19]. This could be explained by a cavitation mechanism as described by Iancu *et al.*, [25]. The preference for ATP vs GTP may be due to the presence of an asparagine or glutamine in place of an aspartate, Figure 4, which at position 333 of *E. coli* PurA alters the nucleoside triphosphate specificity [26]. Mutants of the three residues in *SbPurZ* interacting with the adenine base of ATP (Asn306Thr, Phe307Lys, and Asn309Asp) resulted in an increase of the  $K_m$  for ATP ranging from 89- to 450-fold [24]. The aspartate substrate interacts mainly with the protein backbone (Thr262, Thr263, Val264 and Arg269) and is ideally positioned to react with the dGMP molecule. The Thr27Gly mutant of *SbPurZ*

(corresponding to Thr262 in PhiVC8 PurZ) displayed a 271-fold increase in the  $K_m$  for Asp, consistent with its role in Asp binding [24].

Following the AMP reaction scheme, dZMP can only be obtained by hydrolysis of dSMP by a lyase. Adenylosuccinate lyase (PurB) catalyzes two distinct but chemically similar reactions: - the cleavage of the simplified purine ribonucleotide 5-aminoimidazole-4-(N-succinylcarboxamide) ribonucleotide (SAICAR) to 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) and fumarate and -the cleavage of succinyladenosine monophosphate (SAMP) to AMP and fumarate. It was therefore reasonable to assume that PurB could hydrolyse dSMP.

As the PhiVC8 genome does not contain a lyase capable of hydrolyzing N6-succino-2-amino-2' deoxyadenylate to 2-amino-2' deoxyadenylate (dZMP) and fumarate, nor a kinase capable of phosphorylating dZMP, we hypothesized that these activities must be encoded by the host i.e. *V. cholerae*. We therefore cloned, expressed and purified adenylosuccinate lyase (PurB), guanylate kinase (Gmk) and nucleoside diphosphate kinase (Ndk) from *V. cholerae*.

PurB converts dSMP to dZMP as shown by the disappearance of dSMP and the appearance of a peak with a  $m/z$  of 347 when the reaction was followed by LC-MS. Gmk and Ndk provide the diphosphate form, dZDP ( $m/z$  426) and then the triphosphate form, dZTP ( $m/z$  505.9) respectively [23]. Similar conclusions were drawn by using *E. coli* PurB, and *Salmonella enterica* GMP kinase confirming that the host enzymes tolerate the presence of the amine group at position 2 of adenine [24].

### **dZTP polymerization**

dZTP is a substrate for a number of RNA or DNA polymerases and Taq DNA polymerase accept dZTP as well as dATP [15]. While the sequencing of S-2L genome did not reveal any DNA polymerase, a *polA* gene homolog designated *dpoZ* was found to occur in synteny with *purZ* in all the other phage genomes having *purZ*. DpoZ corresponds to the Klenow fragment of *Escherichia coli* DNA polymerase I, lacking the 5'-exonuclease domain but with the 3'-exonuclease domain. Four His-tagged DpoZ from Vibrio phage PhiVC8, *Acinetobacter* phage SH-Ab 15497, *Arthrobacter* phage Wayne and *Gordonia* phage Ghobes were purified and tested for their ability to incorporate and/or to copy dZTP and dATP in primer extension assays. Full length products were observed with dZTP whatever the DNA polymerase used in

contrast to dATP that is incorporated less efficiently by the four phage polymerases as compared to *E. coli* the Klenow fragment. Similarly, dTTP is more efficiently incorporated by the phage polymerases in front of a dZ matrix than a dA matrix. The opposite is found with the Klenow fragment of *E. coli*. Valérie Pezo and colleagues show that *Vibrio* phage PhiVC8, *Acinetobacter* phage SH-Ab 15497 DNA polymerase (DpoZ) mutated in the active site of the 3'-exonuclease domain have a better affinity for dZTP than for dATP (0.8 vs 4.5  $\mu$ M, and 1.3 vs 21  $\mu$ M respectively) which favours its incorporation into DNA [27]. They also propose that excision of incorporated dA may be promoted by the 3'-exonuclease activity of DpoZ [27]. However, a close investigation of the exonuclease domain of PhiVC8 DpoZ suggests that the selectivity towards dZTP is not directly encoded in its active site [28].

Further structural studies are needed to explain the specificity of DpoZ towards 2-aminoadenine.

Together, the results describe the dZTP biosynthetic pathway and its incorporation into DNA. It remained to be understood how dATP is excluded.

### **Exclusion of adenine**

The genomic context of PurZ was the source of the adenine-excluding enzymes. Indeed, close to PurZ was found an HD domain-containing hydrolase-like enzyme. This protein designated DatZ was further studied both structurally and biochemically.

DatZ belongs to a family of metal-dependent enzymes that catalyze the hydrolysis of phosphoester bond of canonical deoxyribonucleoside 5'-monophosphate (dNMP). A few cases of deoxyribonucleoside triphosphate hydrolyzing enzymes were reported such as *B. subtilis* YpgQ [29] and the *Enterococcus faecalis* EF1143 protein [30]. The S-2L DatZ was purified after its overexpression in *E. coli* and inserted in a polymerization assay. No polymerization was observed in the presence of the four dNTPs in contrast to the dGTP, dCTP, dTTP and dZTP. This suggested that DatZ was active on dATP which was further confirmed by incubating the enzyme with each of the dNTP and following the reaction by HPLC. dATP was rapidly converted to dA as well as dADP and dAMP but no other dNTP was substrate of DatZ. The triphosphate hydrolysis of dATP does not appear to be sequential since no intermediate was observed during the reaction which contrast to OxsA phosphohydrolase [31]. Comparable results were obtained with *Acinetobacter* phage SH-Ab 15497 and *Salmonella* phage PMBT28

DatZ. These latter hydrolyze dATP into dA with the highest catalytic activity using  $\text{Co}^{2+}$  as the divalent metal cofactor [24]. At the structural level, DatZ is a hexamer in which each monomer has a globular form. The aminoacids participating to the metal ions binding His34, His66, Asp67, Glu70, Asp75 and Asp119 are conserved in all phage DatZ. Arg19, Lys81 and Lys116 interacting with  $\alpha$ -,  $\beta$ - and  $\gamma$ -phosphates of dATP, Trp20, Ile22 and Pro79, interacting with the base, are also conserved or involve conservative substitutions. Overlay of the DatZ structures obtained in the presence of divalent ions or substrate ie dATP allow to propose a model of catalysis where DatZ uses a typical two- metal-ion mechanism to dephosphorylate dATP. First ion  $\text{B}^{2+}$  stabilizes the leaving  $\text{O}5'$  atom and one oxygen of the  $\alpha$ - phosphate, ion  $\text{A}^{2+}$  positions a hydroxide ( $\text{OH}^-$ ) in an attacking position opposite to  $\text{O}5'$ . Then, by interacting with  $\text{OH}^-$ , the  $\alpha$ -phosphate passes through a penta-coordinate intermediate, forming an unstable oxyanion stabilized by Arg19. Finally, the bond  $\text{O}5'-\text{P}_\alpha$  is broken and a new one,  $\text{P}_\alpha\text{-OH}$ , is created (Figure 5) [31]

Thus, elimination of dATP occurs at two levels: before replication by the action of DatZ which is a dATP-specific triphosphohydrolase [31] and after replication by the 3' exonuclease activity of the DNA polymerase DpoZ in most bacteriophages except S-2L [25].

### **Origin of dGMP, substrate of PurZ**

Between PurZ and DatZ a DUF550 domain-containing protein called MazZ was found in all phage genomes containing PurZ. Recombinant Acinetobacter phage SH-Ab 15497 MazZ exhibits dATP and 2'-deoxyguanosine 5'-triphosphate (dGTP) pyrophosphohydrolase activity, catalyzing the hydrolysis of dATP/dGTP to pyrophosphate and dAMP/dGMP, respectively, with the highest activity obtained using  $\text{Co}^{2+}$  as the divalent metal cofactor and little or no activities for NTP or dYTP (Y:pyrimidine). This specificity for purines is confirmed with S-2L MazZ that removes the two terminal phosphates of both dGTP and GTP and no other (deoxy) nucleotide. Structure of S-2L MazZ bound to the dephosphorylation product of dGTP and catalytic  $\text{Mn}^{2+}$  was obtained. No particular aminoacid are involved in the specificity of MazZ; rather the cavity volume matching guanine's shape and its charge compatibility. There is also no steric hindrance for the 2'OH group which explains the activity of MazZ on dGTP and GTP. The electron density of dGDP is completely buried in the enzyme except the  $\beta$ -phosphate

group near the three catalytic  $Mn^{2+}$  ions. Ion A is coordinated by residues E34, E35 and E38; ion B by E50 and D53; and ion C by E38 and E50 (Figure). Since these three  $Mn^{2+}$  ions are strictly equivalent to  $Mg^{2+}$  ions found in *Campylobacter jejuni* dUTPase [32], the presence of dGDP in the crystal suggested that the phosphate groups were removed successively rather than in a single step (Figure 6).

The origin of the dGMP substrate N6-succino-2-amino-2' deoxyadenylate was then demonstrated by characterizing the activity of the MazZ enzyme which is a dGTP-specific diphosphohydrolase [33]. The characterization of PurZ, DpoZ, DatZ and MazZ allows us to propose a biosynthetic pathway for dZTP and also explain how dATP is excluded (Figure 7).

### **Occurrence of 2-aminoadenine in DNA from phages infecting Gram-negative and Gram-positive bacteria**

The presence of 2,6-diaminopurine is not limited to phages S-2L and PhiVC8. Indeed, PurZ is present in many phages infecting Gram-negative bacteria such as *Acinetobacter* phage SH-Ab 15497(MG674163), *Salmonella* phage PMBT 28 (MG641885), *Alteromonas* phage ZP6 (MK203850) but also in lytic bacteriophages for Gram-positive bacteria such as *Arthrobacter* phage Wayne (KU160672), *Ghordonia* phage Ghobes (KX557278) or *Streptomyces* phage Hiyaa (MK279841) (Figure 8). The correlation between the presence of PurZ and 2-aminoadenine seems good as this base is detected in the genomes of *Acinetobacter* phage SH-Ab 15497, Wayne and Ghobes [22,25]. Most phages possess a DpoZ DNA polymerase with the exception of S-2L and the *Bacillus* phage vB\_BpsS-140 which encodes an alpha subunit of DNA polymerase III.

As with PhiVC8, Wayne and Ghobes DpoZ DNA polymerases have a preference for dZTP over dATP and remarkably so whether or not the template DNA contains Z for DpoZ from *Acinetobacter* phage SH-Ab 15497 [27].

The exclusion of dATP by DatZ is conserved in Gram-negative phages and presumably must be achieved by a "dUTPase" in Gram-positive phages. However, MazZ is found in two different forms in Gram-negative phages but not in Gram+ phages, which must use an enzyme that catalyzes the same reaction as MazZ, which has yet to be identified. It is interesting to note that the *Gordonia* phage Ghobes and the *Streptomyces* phage Hiyaa have their own adenylosuccinate lyase which could be specific for dZMP which, in the *Bacillus* phage vB\_BpsS-140, could be phosphorylated into dZDP by the phage adenylate kinase.

Each phage therefore has, in addition to PurZ, different genes to synthesise dZTP and eliminate dATP, which most likely originate from horizontal transfer between phages.

This Z biosynthetic pathway was validated *in vivo* in *V. cholerae* [23] but also in *E. coli* by expressing PurZ, PurZ MazZ, PurZ DatZ or PurZ DatZ MazZ. *In vivo*, the inactivation of *purB* results in a 5 log diminution of plating efficiency as compared with the wild type while no phage was obtained with the *ndk* mutant indicating that no other enzymes are able to convert dZDP into dZTP. This result also shows that dZTP synthesis is absolutely required for the phage DNA replication.

*In vitro*, while the expression of PurZ alone hijacks the metabolism of *E. coli* and incorporates 2-aminoadenine into plasmid DNA (0.6%) and chromosomal DNA (0.1%) in low proportions, the expression of PurZ DatZ and MazZ results in higher substitution rates of 4% and 12% respectively [33]. Above this percentage, the incorporation of 2-aminoadenine becomes toxic.

## Conclusion

The biosynthesis of dZTP and the exclusion of dATP are now better understood and it is remarkable that the production of dZTP and its incorporation into DNA results from a combination of phage and host enzymes, illustrating how a bacteriophage hijacks a host pathway (here AMP) to synthesize its own deoxyribonucleotide.

Several questions about the Z-base remain open, such as:

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?

The body of work is an illustration of what nature has been exploring and is certainly only one of the possibilities for the storage of genetic information [34]

## Declarations

Ethics approval and consent to participate 'Not applicable'

Consent for publication Yes

Availability of data and material

The atomic coordinates of PhiVC8 PurZ , S-2L DatZ and S-2L MazZ have been deposited in the Protein Data Bank with the accession codes 6FLF, 6FM0, 6FM1, 6TNH ; 6ZPB and 6ZPC; 7ODY

Competing interests

The author declares no competing financial interests.

Funding

This study was funded from the French government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant no. ANR-10-LABX-62-IBEID).

Authors' contributions 'Not applicable'

Acknowledgements

I thank Professor Patrick Trieu-Cuot for his support.

Authors' information

Biologie des Bactéries Pathogènes à Gram-Positif, Institut Pasteur, CNRS-UMR 2001, Paris, France.

## References

1. Carell T, Brandmayr C, Hienzsch A, et al (2012) Structure and Function of Noncanonical Nucleobases. *Angew Chem Int Ed* 51:7110–7131. <https://doi.org/10.1002/anie.201201193>
2. Adhikari S, Curtis PD (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiology Reviews* 40:575–591. <https://doi.org/10.1093/femsre/fuw023>
3. Lee Y-J, Dai N, Walsh SE, et al (2018) Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. *Proc Natl Acad Sci USA* 115:E3116–E3125. <https://doi.org/10.1073/pnas.1714812115>
4. Weigele P, Raleigh EA (2016) Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem Rev* 116:12655–12687. <https://doi.org/10.1021/acs.chemrev.6b00114>
5. Lee Y-J, Dai N, Müller SI, et al (2021) Pathways of thymidine hypermodification. *Nucleic Acids Res* gkab781. <https://doi.org/10.1093/nar/gkab781>

6. Swinton D, Hattman S, Crain PF, et al (1983) Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosyl)purin-6-ylglycinamide, specified by the phage Mu modification function. *Proc Natl Acad Sci USA* 80:7400–7404. <https://doi.org/10.1073/pnas.80.24.7400>
7. Thiaville JJ, Kellner SM, Yuan Y, et al (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc Natl Acad Sci USA* 113:E1452-1459. <https://doi.org/10.1073/pnas.1518570113>
8. Nikolskaya II, Lopatina NG, Debov SS (1976) Methylated guanine derivative as a minor base in the DNA of phage DDVI *Shigella dysenteriae*. *Biochim Biophys Acta* 435:206–210. [https://doi.org/10.1016/0005-2787\(76\)90251-3](https://doi.org/10.1016/0005-2787(76)90251-3)
9. Dunn DB, Smith JD (1958) The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem J* 68:627–636. <https://doi.org/10.1042/bj0680627>
10. Crippen CS, Lee Y-J, Hutinet G, et al (2019) Deoxyinosine and 7-Deaza-2-Deoxyguanosine as Carriers of Genetic Information in the DNA of *Campylobacter* Viruses. *J Virol* 93:. <https://doi.org/10.1128/JVI.01111-19>
11. Hutinet G, Lee Y, de Crécy-Lagard V, Weigele P Hypermodified DNA in Viruses of *E. coli* and *Salmonella*. *EcoSal Plus* 9:eESP-0028-2019
12. Kirnos MD, Khudyakov IY, Alexandrushkina NI, Vanyushin BF (1977) 2-Amino adenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* 270:369–370. <https://doi.org/10.1038/270369a0>
13. Khudyakov IY, Kirnos MD, Alexandrushkina NI, Vanyushin BF (1978) Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. *Virology* 88:8–18. [https://doi.org/10.1016/0042-6822\(78\)90104-6](https://doi.org/10.1016/0042-6822(78)90104-6)
14. Bailly C, Suh D, Waring MJ, Chaires JB (1998) Binding of daunomycin to diaminopurine- and/or inosine-substituted DNA. *Biochemistry* 37:1033–1045. <https://doi.org/10.1021/bi9716128>
15. Bailly C, Waring MJ (1998) The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA. *Nucleic Acids Res* 26:4309–4314. <https://doi.org/10.1093/nar/26.19.4309>
16. Szekeres M, Matveyev AV (1987) Cleavage and sequence recognition of 2,6-diaminopurine-containing DNA by site-specific endonucleases. *FEBS Lett* 222:89–94. [https://doi.org/10.1016/0014-5793\(87\)80197-7](https://doi.org/10.1016/0014-5793(87)80197-7)
17. Honzatko RB, Stayton MM, Fromm HJ (1999) Adenylosuccinate synthetase: recent developments. *Adv Enzymol Relat Areas Mol Biol* 73:57–102, ix–x
18. Solís-Sánchez A, Hernández-Chiñas U, Navarro-Ocaña A, et al (2016) Genetic characterization of ØVC8 lytic phage for *Vibrio cholerae* O1. *Virol J* 13:47. <https://doi.org/10.1186/s12985-016-0490-x>

19. Bouyoub A, Barbier G, Forterre P, Labedan B (1996) The adenylosuccinate synthetase from the hyperthermophilic archaeon *Pyrococcus* species displays unusual structural features. *J Mol Biol* 261:144–154. <https://doi.org/10.1006/jmbi.1996.0448>
20. Jayalakshmi R, Sumathy K, Balaram H (2002) Purification and Characterization of Recombinant Plasmodium falciparum Adenylosuccinate Synthetase Expressed in *Escherichia coli*. *Protein Expression and Purification* 25:65–72. <https://doi.org/10.1006/prep.2001.1610>
21. Powell SM, Zalkin H, Dixon JE (1992) Cloning and characterization of the cDNA encoding human adenylosuccinate synthetase. *FEBS Letters* 303:4–10. [https://doi.org/10.1016/0014-5793\(92\)80465-S](https://doi.org/10.1016/0014-5793(92)80465-S)
22. Honzatko RB, Fromm HJ (1999) Structure-function studies of adenylosuccinate synthetase from *Escherichia coli*. *Arch Biochem Biophys* 370:1–8. <https://doi.org/10.1006/abbi.1999.1383>
23. Sleiman D, Garcia PS, Lagune M, et al (2021) A third purine biosynthetic pathway encoded by aminoadenine-based viral DNA genomes. *Science* 372:516–520. <https://doi.org/10.1126/science.abe6494>
24. Zhou Y, Xu X, Wei Y, et al (2021) A widespread pathway for substitution of adenine by diaminopurine in phage genomes. *Science* 372:512–516. <https://doi.org/10.1126/science.abe4882>
25. Iancu CV, Zhou Y, Borza T, et al (2006) Cavitation as a mechanism of substrate discrimination by adenylosuccinate synthetases. *Biochemistry* 45:11703–11711. <https://doi.org/10.1021/bi0607498>
26. Kang C, Sun N, Honzatko RB, Fromm HJ (1994) Replacement of Asp333 with Asn by site-directed mutagenesis changes the substrate specificity of *Escherichia coli* adenylosuccinate synthetase from guanosine 5'-triphosphate to xanthosine 5'-triphosphate. *J Biol Chem* 269:24046–24049
27. Pezo V, Jaziri F, Bourguignon P-Y, et al (2021) Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science* 372:520–524. <https://doi.org/10.1126/science.abe6542>
28. Czernecki D, Hu H, Romoli F, Delarue M (2021) Structural dynamics and determinants of 2-aminoadenine specificity in DNA polymerase DpoZ of vibriophage  $\phi$ VC8. *Nucleic Acids Res* gkab955. <https://doi.org/10.1093/nar/gkab955>
29. Jeon YJ, Park SC, Song WS, et al (2016) Structural and biochemical characterization of bacterial YpgQ protein reveals a metal-dependent nucleotide pyrophosphohydrolase. *Journal of Structural Biology* 195:113–122. <https://doi.org/10.1016/j.jsb.2016.04.002>
30. Vorontsov II, Minasov G, Kiryukhina O, et al (2011) Characterization of the Deoxynucleotide Triphosphate Triphosphohydrolase (dNTPase) Activity of the EF1143 Protein from *Enterococcus faecalis* and Crystal Structure of the Activator-Substrate

Complex. Journal of Biological Chemistry 286:33158–33166.  
<https://doi.org/10.1074/jbc.M111.250456>

31. Czernecki D, Legrand P, Tekpinar M, et al (2021) How cyanophage S-2L rejects adenine and incorporates 2-aminoadenine to saturate hydrogen bonding in its DNA. Nat Commun 12:2420. <https://doi.org/10.1038/s41467-021-22626-x>
32. Moroz OV, Harkiolaki M, Galperin MY, et al (2004) The Crystal Structure of a Complex of *Campylobacter jejuni* dUTPase with Substrate Analogue Sheds Light on the Mechanism and Suggests the “Basic Module” for Dimeric d(C/U)TPases. Journal of Molecular Biology 342:1583–1597. <https://doi.org/10.1016/j.jmb.2004.07.050>
33. Czernecki D, Bonhomme F, Kaminski P-A, Delarue M (2021) Characterization of a triad of genes in cyanophage S-2L sufficient to replace adenine by 2-aminoadenine in bacterial DNA. Nat Commun 12:4710. <https://doi.org/10.1038/s41467-021-25064-x>
34. Cleaves HJ, Butch C, Burger PB, et al (2019) One Among Millions: The Chemical Space of Nucleic Acid-Like Molecules. J Chem Inf Model 59:4266–4277. <https://doi.org/10.1021/acs.jcim.9b00632>

## Figure legends

**Figure 1: Hydrogen bonding of different base pairs:** adenine:thymine A:T, guanine:cytosine G:C and 2,6 diaminopurine:thymine Z:T

## **Figure 2: AMP biosynthetic pathway.**

IMP Inosine 5'-monophosphate, SMP Adenylosuccinate, AMP Adenosine 5'-monophosphate, GTP Guanosine 5'-triphosphate, GDP Guanosine 5'-diphosphate, PurA adenylosuccinate synthase, PurB adenylosuccinate lyase

## **Figure 3: PhiVC8 PurZ resembles archaeal PurAs**

**A** Structures of PhiVC8 PurZ apo (blue) top, PhiVC8 PurZ overlaid to *P.horikoshii* PurA (grey), middle and PhiVC8 PurZ overlaid to *E. coli* PurA (grey) bottom. **B** Phylogenetic tree of a few representative adenylosuccinate synthases. The bacteriophage branch is in green, those from archaea in brown, those from eukaryotes in yellow and those from bacteria in blue.

## **Figure 4: Diagrams of interaction of PhiVC8 PurZ and *E. coli* PurA with their substrates**

**A** PhiVC8 PurZ -dGMP **B** PurZ-ATP **C** PurA -IMP **D** PurA-GTP

**Figure 5: Structure of DatZ a triphosphodeoxyadenosinetriphosphate hydrolase**

**A** Hexamer of DatZ **B** Diagram of interaction DatZ-dATP

**Figure 6: Structure of MazZ diphosphodeoxyguanosinetriphosphate hydrolase**

**A** Tetramer of MazZ **B** Diagram of interaction MazZ-dGTP

**Figure 7: Biosynthesis of 2,6-diaminopurine and adenine exclusion in PhiVC8/ *Vibrio cholerae***

dGMP deoxyguanosine 5'-monophosphate, dSMP N6-succino-2-amino-2' deoxyadenylate, dZMP 2-amino-2' deoxyadenylate, dZDP 2-amino-2' deoxyadenosine 5'-diphosphate, dZTP 2-amino-2' deoxyadenosine 5'-triphosphate, dATP 2' deoxyadenosine 5'-triphosphate, dCTP 2' deoxycytidine 5'-triphosphate, dGTP 2' desoxyguanosine 5'-triphosphate, dTTP, 2' deoxythymidine 5'-triphosphate dA deoxyadenosine, ATP Adenosine 5'-triphosphate, ADP Adenosine 5'-diphosphate.

PurZ N6-succino-2-amino-2' deoxyadenylate synthase, PurB Adenylosuccinate lyase, Gmk Guanylate kinase, Ndk Nucléoside diphosphate kinase DpoZ DNA polymerase Z, DatZ deoxyadenosine 5'-triphosphohydrolase, MazZ deoxyguanosine 5'-diphosphohydrolase.

**Figure 8: Genomic context around PurZ in different bacteriophages**

PurZ N6-succino-2-amino-2' deoxyadenylate synthase, PurB Adenylosuccinate lyase, DatZ deoxyadenosine 5'-triphosphohydrolase, MazZ deoxyguanosine 5'-diphosphohydrolase, DpoZ DNA polymerase, dUTPase deoxyuridine triphosphatase, Adk adenylate kinase.

Fig1

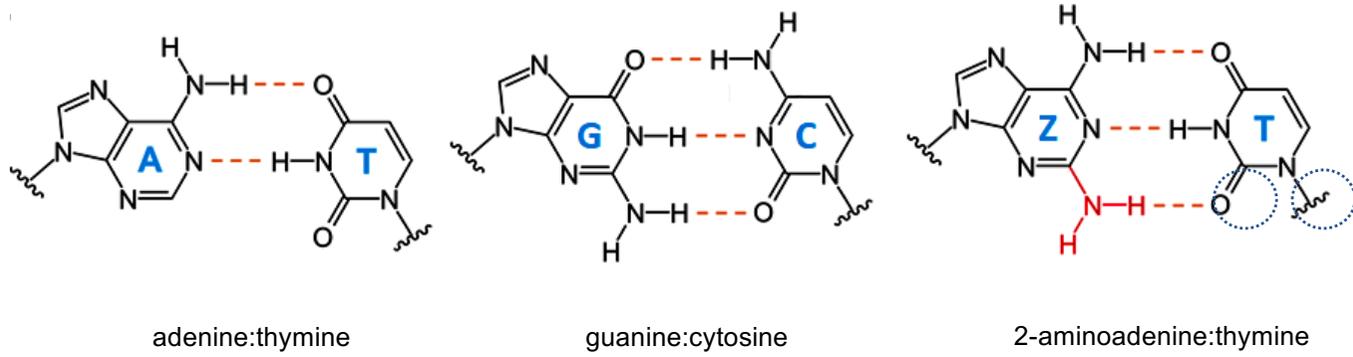


Fig2

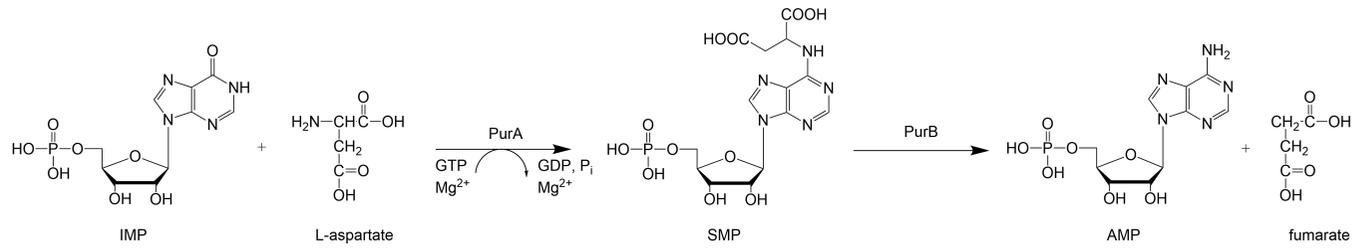


Fig3

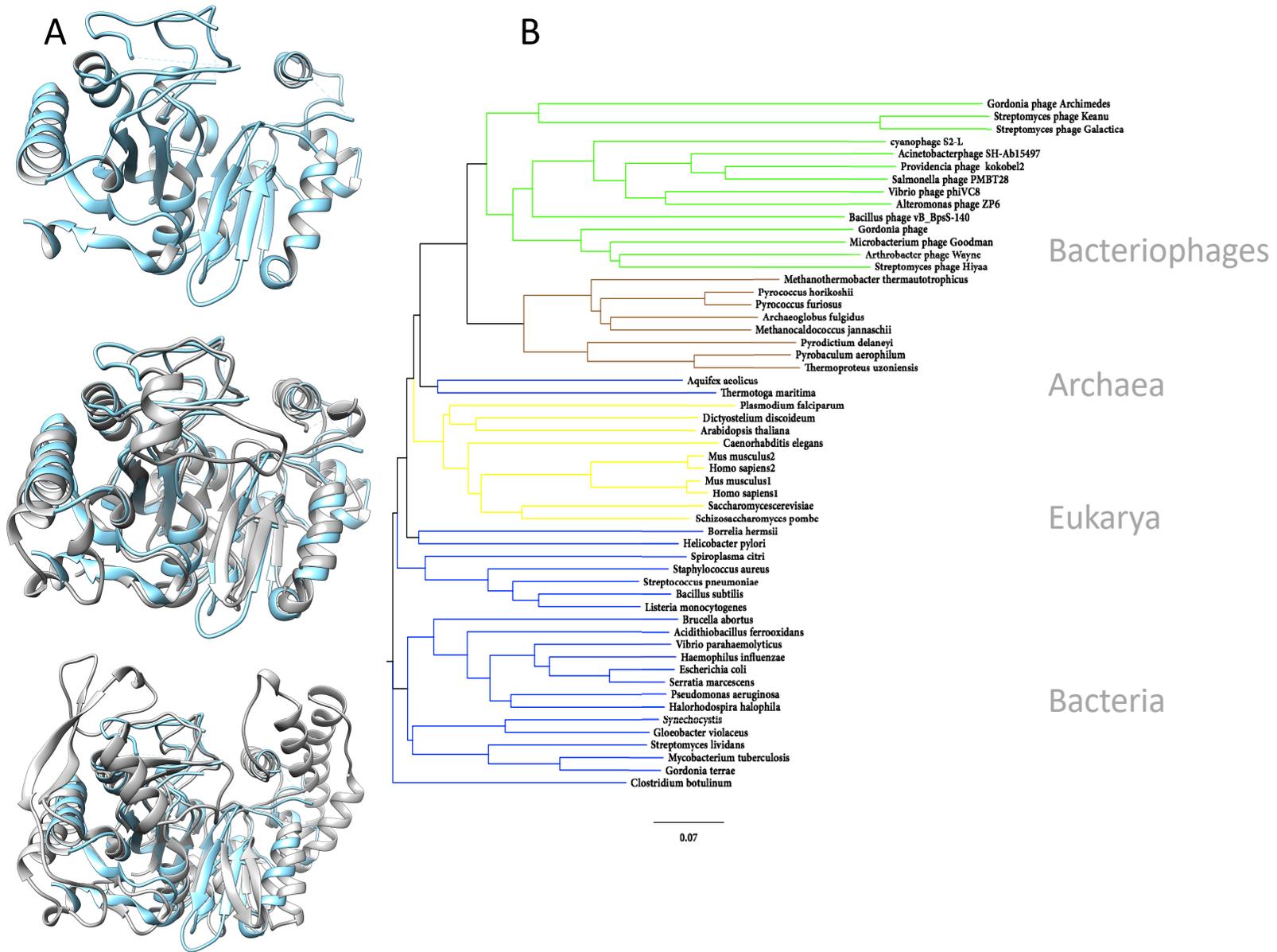


Fig4

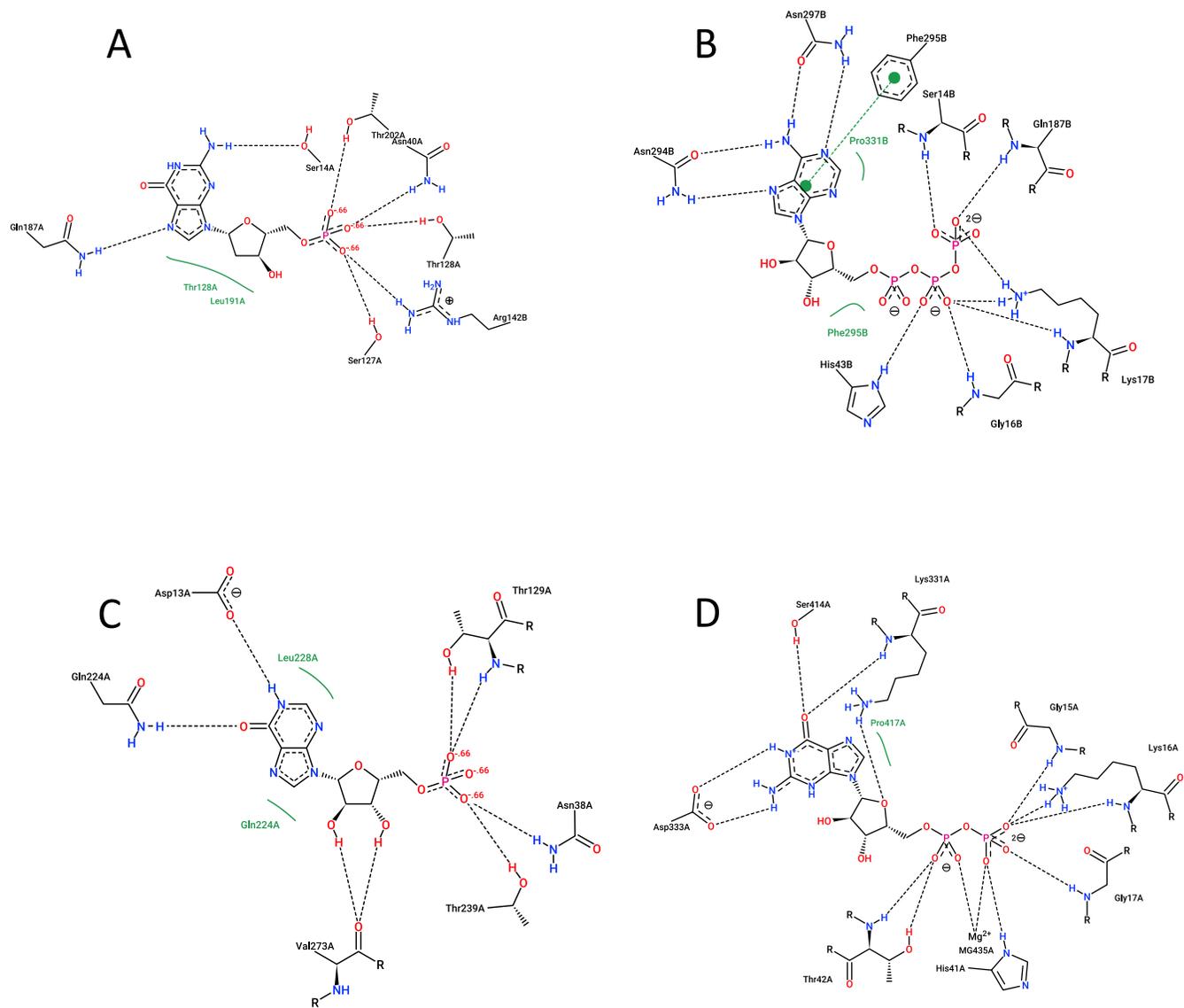
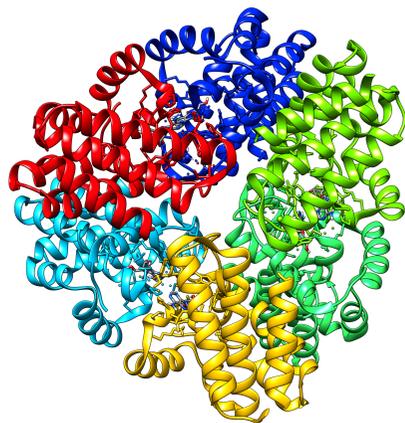


Fig5

A



B

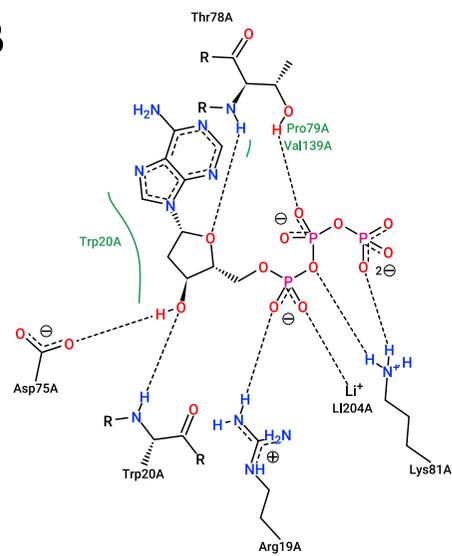




Fig7

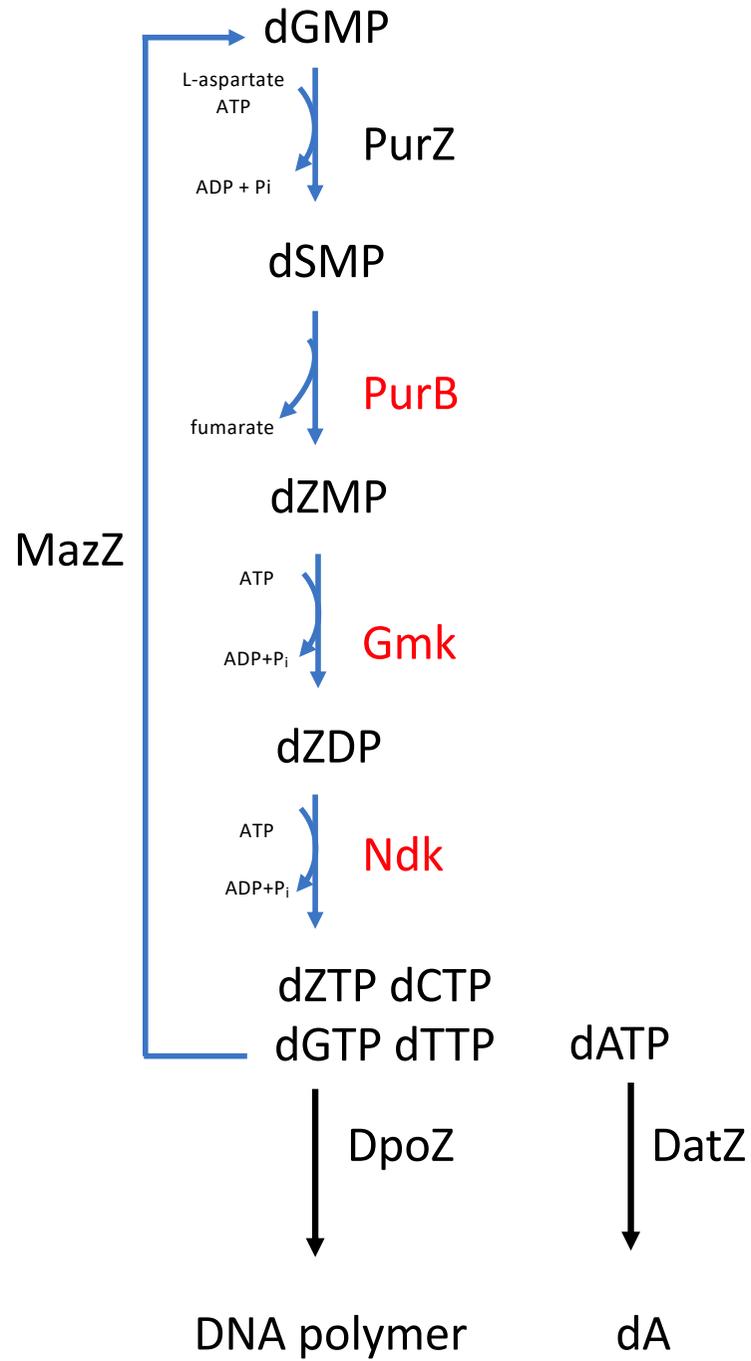


Fig8

