ACS SyntheticBiology

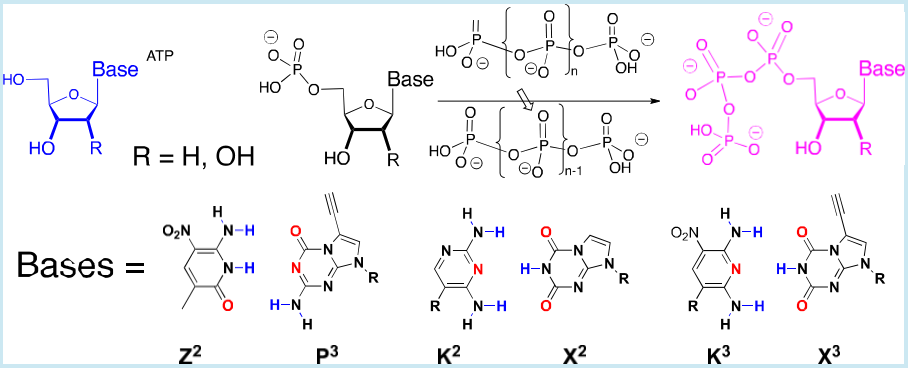
pubs.acs.org/synthbio Research Article

Synthetic Biology Pathway to Nucleoside Triphosphates for Expanded Genetic Alphabets

Yubing Li,* Clay Abraham, Oleg Suslov, Ozlem Yaren, Ryan W. Shaw, Myong-Jung Kim, Shuo Wan, Philippe Marliere, and Steven A. Benner

Cite This: *ACS Synth. Biol.* 2023, 12, 1772−1781 Read Online

ACCESS Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: One horizon in synthetic biology seeks alternative forms of DNA that store, transcribe, and support the evolution of biological information. Here, hydrogen bond donor and acceptor groups are rearranged within a Watson−Crick geometry to get 12 nucleotides that form 6 independently replicating pairs. Such artificially expanded genetic information systems (AEGIS) support Darwinian evolution *in vitro*. To move AEGIS into living cells, metabolic pathways are next required to make AEGIS triphosphates economically from their nucleosides, eliminating the need to feed these expensive compounds in growth media. We report that "polyphosphate kinases" can be recruited for such pathways, working with natural diphosphate kinases and engineered nucleoside kinases. This pathway *in vitro* makes AEGIS triphosphates, including third-generation triphosphates having improved ability to survive in living bacterial cells. In α-^{32}P-labeled forms, produced here for the first time, they were used to study DNA polymerases, finding cases where third-generation AEGIS triphosphates perform better with natural enzymes than second-generation AEGIS triphosphates.

KEYWORDS: *expanded DNA alphabets, polyphosphate kinase, engineered pathway, alpha-radiolabeled triphosphates, semisynthetic organisms*

INTRODUCTION

One horizon in synthetic biology[1] seeks nonstandard forms of DNA and RNA that store, replicate, and transcribe information, reminiscent of these processes in natural DNA and RNA, but on different scaffolds. Alternative forms of DNA may store information more stably and at higher densities to be retrieved long after magnetically stored information has decayed.[2,3] Different nucleic acids are better than standard DNA/RNA to support "laboratory *in vivo* evolution" (LIVE), where nucleic acid libraries are challenged to evolve receptors, ligands, and catalysts under scientist-generated selective pressures.[4]

One strategy to approach this horizon retains Watson−Crick complementarity, both size and hydrogen bonding, in the synthetic genetic system. Figure 1 shows a "second generation" form of a 12-letter genetic alphabet that has been developed in the Benner laboratory that does this.[5] These artificially expanded genetic information systems (AEGIS) support laboratory *in vitro* evolution,[6−9] surmounting limitations that have prevented LIVE in various forms from achieving its original promise.[10]

In other forays into this genre of synthetic biology, Kool,[14] Hirao,[15] and Romesberg[16,17] have explored the versatility of size complementarity (Figure 2). Kool has even created an eight-letter genetic system that supports duplex formation *in vitro* (Figure 2).[18] Romesberg has reported "semisynthetic" *Escherichia coli* that replicates, transcribes, and translates DNA

Received: January 28, 2023
Published: May 25, 2023

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

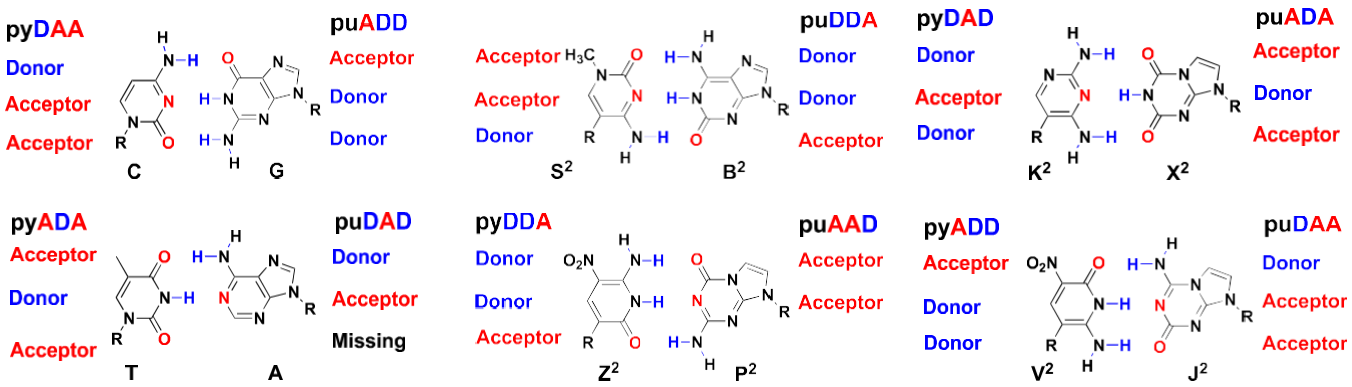


Figure 1. Second-generation components of an AEGIS, now commercially available,[11] were obtained by rearranging hydrogen bonding donor and acceptor groups on base pairs while retaining Watson−Crick size complementarity (large two-ring systems pair with small one-ring systems) and Watson−Crick hydrogen bonding complementarity (hydrogen bond donor groups pair with hydrogen bond acceptor groups). Polymerases have been developed to replicate AEGIS,[12] allowing it to support laboratory *in vitro* evolution (LIVE);[5−8] AEGIS-containing aptamers have potential in cancer therapy.[13]

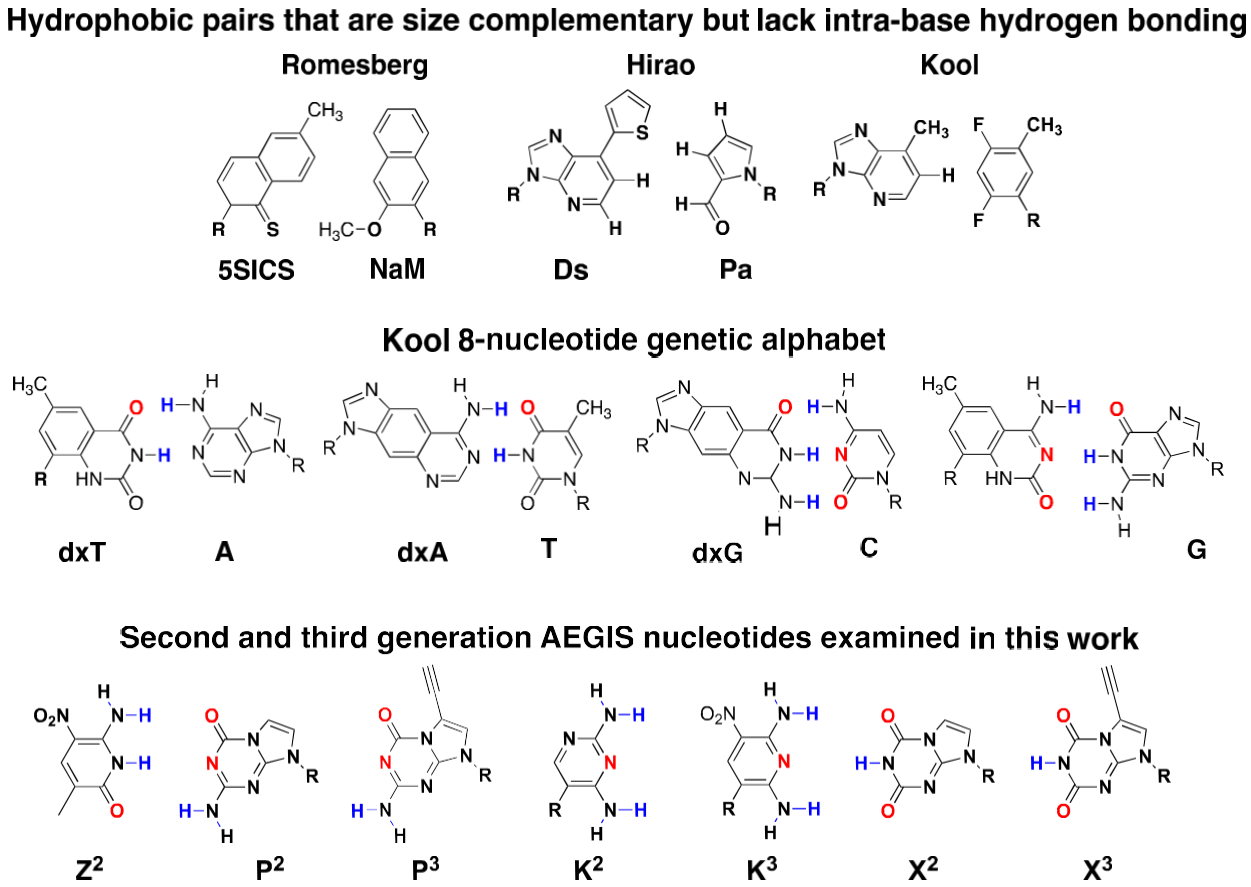


Figure 2. Structures of various nonstandard nucleobases used in various efforts to support synthetic biology that alters the biopolymers at the center of life. (Top panel) Three examples of hydrophobic pairs that approximate the geometry of standard Watson−Crick pairs but lack inter-nucleotide hydrogen bonding. (Middle panel) Eight-letter genetic system that retains inter-nucleotide hydrogen bonding complementarity but with geometrically larger pairs. (Bottom panel) AEGIS nucleobases examined here. Those with superscript "2" are second-generation AEGIS components that implement the indicated hydrogen bonding pattern (red, H-bond acceptors; blue, hydrogen bond donors). Those with superscript "3" are third-generation AEGIS species with properties better suited for intracellular performance.

and RNA that contained size-complementary hydrophobic base pairs not joined by any hydrogen bonds.[19,20] Unfortunately, Romesberg was required to feed expensive triphosphates to *E. coli* growth medium, after placing in the bacterium a gene for a plastid triphosphate transporter.[21,22] This is likely to expensive for any but research use.

This exemplifies a general problem with all synthetic biology: Elements of natural biology are almost obligatorily recruited to work with unnatural synthetic parts. Thus, enzymes that "phosphorylate up" standard nucleotides inside of cells (Figure 3) would deliver triphosphates less expensively. This, however, requires that the natural phosphorylating

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

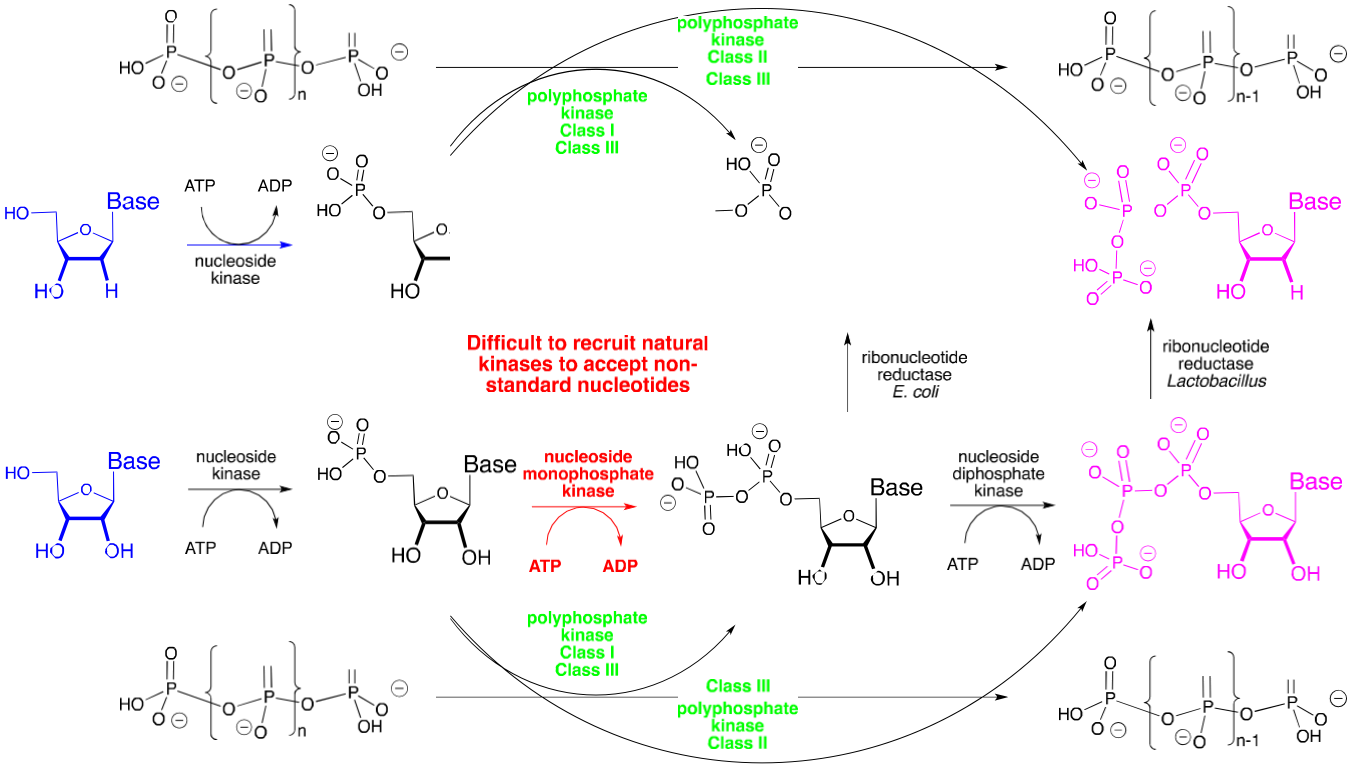


Figure 3. Enzymatic pathways to make nucleoside triphosphates (right, magenta) from nucleosides (left, blue). The standard pathway concept (center) adds single phosphate groups stepwise, one at a time, to convert a nucleoside to its nucleoside monophosphate, the nucleoside monophosphate to its nucleoside diphosphate, and the nucleoside diphosphate to its nucleoside triphosphate, with ATP providing the phosphate moiety at each step. It has proven difficult to find a natural nucleoside monophosphate kinase, or to engineer one, to accept monophosphates that have nonstandard nucleobases (Figures 1 and 2). Accordingly, polyphosphate kinases (green) were sought to convert monophosphates to di- and triphosphates using polyphosphate as a phosphate donor.

kinases directly, or in engineered form, accept nonstandard nucleoside and nucleotide substrates.

The more "alien" the nucleoside, the less likely that such recruitments might be possible. However, recruiting natural enzymes to phosphorylate even the structurally similar AEGIS nucleosides has been challenging. Many unnatural nucleosides are taken up from the growth medium by the NupC, NupG, and/or XapB (for xanthosine analogue) transporter,[23] and AEGIS nucleotides appear to be no different. Further, the base substrate specificities of the endogenous *E. coli* nucleoside diphosphate kinases are sufficiently broad that they convert most second-generation AEGIS nucleoside diphosphates to their triphosphates.[24]

However, mutation of the nucleoside kinase from *Drosophila melanogaster* was required to create an enzyme that converts various AEGIS nucleosides in Figure 1 to their monophosphates.[25] It proved even more difficult to engineer nucleoside monophosphate kinases to convert AEGIS monophosphates to their diphosphates. Rational design by ROSETTA and semi-random screening failed to generate enzymes that converted AEGIS monophosphates to diphosphates.

Accordingly, we sought to engineer a different path to generate AEGIS nucleoside triphosphates from AEGIS nucleosides. This began by surveying enzymes that use polyphosphate, rather than ATP, as a phosphate donor and act on nucleoside monophosphates. We asked whether such "polyphosphate kinases"[26] (named for the reverse reaction that they catalyze, $(P_i)_n$ + ADP → $(P_i)_{n+1}$ + AMP) might be more adaptable to accept AEGIS nucleotide monophosphates.

We report here experiments that show that polyphosphate kinases act on a series of second-generation AEGIS nucleoside monophosphates. This, together with other natural and engineered kinases, allowed us to engineer an *in vitro* metabolic path that delivered AEGIS nucleoside triphosphates from AEGIS nucleosides, ATP, and polyphosphate (Figure 3). This pathway was then used to solve an important problem in practical synthetic biology: making AEGIS triphosphates whose alpha-phosphorus atoms are ^{32}P radiolabeled.

We then applied this pathway to "third generation" AEGIS structures engineered to manage various challenges in developing AEGIS for intracellular use. These included variations of X carrying an ethynyl (acetylene) group (Figure 2). This third-generation heterocycle implementing the large acceptor−donor−acceptor had been found, in DNA, to resist intracellular endonuclease V degradation while still being transported by the NupG or XapB (Li et al., in preparation). This was extended to the third-generation version of P, also with an ethynyl group.

A third-generation form of the Watson−Crick complement of X, a pyrimidine analogue with a nitro group implementing a donor−acceptor−donor hydrogen bonding pattern, was designed (Figure 2). This group is analogous to the nitro group found on AEGIS Z (Figure 1) (Kim et al., in preparation). Interestingly, we were able to find cases where these third-generation "alien" DNA components perform

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

better with natural enzymes than second-generation AEGIS DNA components.

RESULTS AND DISCUSSION

Q81E Variant of *D. melanogaster* Nucleoside Kinase Accepts Third-Generation AEGIS Nucleosides. In separate literature, we found that that a single amino acid replacement in a nucleoside kinase from *D. melanogaster* (Q81E) allowed this kinase to accept many first- and second-generation AEGIS nucleosides.[25] With new variants in development, we asked whether this mutant kinase would also accept third-generation species. The data shown in Figure S1 shows that they do. Table 1 summarizes the kinetic parameters of this kinase with various AEGIS nucleosides.

Table 1. Summary of DmdNKE Activities with Selected AEGIS Nucleosides

		K_M (uM)[a]	k_{cat}/K_M (M^{-1} s^{-1})
	X^2	73.6 ± 26.35	3.19E+05
	K^2	20.0 ± 3.19	1.70E+06
	P^2	13.9 ± 0.68	2.32E+06
	Z^2	264 ± 12.7	2.09E+04
	G	922 ± 72.3	1.18E+04
	C	13.6 ± 3.24	1.27E+06
	X^3(Ethynyl X)	31.5 ± 11.58	9.08E+05
	K^3 (NitroK)	241 ± 24.2	1.11+04
	P^3(Ethynyl P)	61.3 ± 17.65	4.27E+05

[a]All values are means ± standard deviations from three repeats.

Interestingly, adding an ethynyl group to the puADA AEGIS component (converting X^2 to X^3) improved its k_{cat}/K_M performance with the nucleoside kinase. However, a slight decrease in performance was observed when attaching an ethynyl unit to the puAAD AEGIS component (converting P^2 to P^3). In stark contrast, the nitropyrimidone version of K (K^3), was a 100-fold poorer substrate than second generation K^2 (Figure 1).

Polyphosphate Kinases Phosphorylate AEGIS Nucleoside Monophosphates. We then asked whether polyphosphate kinases might convert a spectrum of AEGIS nucleoside monophosphates prepared by the Q81E mutant of nucleoside kinase to their diphosphates or triphosphates. Polyphosphate kinases come in various versions based on (i) an evolutionary analysis of their natural history and (ii) the products that they produce from the substrates.[27] Standard nomenclature divides these into classes: Class I enzymes generate triphosphates from the diphosphates; Class II enzymes generate diphosphates from monophosphates; Class III enzymes generate mixtures of diphosphates and triphosphates from monophosphates.

Several kinases (Class III) were comparatively examined (Figure S2), including those obtained from clones provided by the Matsuura Lab.[28] This comparison found that the ChuPPK2 (Class III, from *Cytophaga hutchinsonii*) catalyzes the best conversion of most AEGIS ribo- and deoxyribonucleoside monophosphates to mixtures of diphosphates and triphosphates (Figure S3). These AEGIS components included rZ^2MP, rK^2MP, rX^2MP, dZ^2MP, dK^2MP, dX^2MP, dP^2MP, dK^3MP, dX^3MP, and dP^3MP (Figure 2). Similarly broad activity was seen with the Class III PPK2 from Matsuura Lab and the Class III PPK2 from *Arthrobacter aurescens* (Figure S4).

Polyphosphate Kinases Support an *In Vitro* Path to Get α-^{32}P-Labeled AEGIS Triphosphates. Table 2 summarizes the data showing various product yields of the triphosphate formed. Since polyphosphate is a heterogeneous substrate, and is also a product, of this enzyme, and since mixtures of products are formed in this easily reversible reaction, standard kinetic studies are unrevealing. However, the data were sufficient to select the preferred enzyme for most AEGIS monophosphates (ChuPPK2) and to use this enzyme to solve a long-standing problem in this branch of synthetic biology: how to make alpha-radiolabeled AEGIS triphosphates.

Table 2. Summary of PPK2 Activities on AEGIS Nucleoside Monophosphates[a]

	substrate	DtCsPPK2	ChuPPK2	AaurPPK2
second-generation ribonucleotide	rP^2MP	di: 49%; tri: 3%	di: 21%; tri: 70%	di: 32%; tri: 61%
	rX^2MP	di: 53%; tri: 10%	di: 24%; tri: 71%	di: 21%; tri: 51%
	rZ^2MP	di: 32%; tri: 43%	di: 26%; tri: 59%	di: 37%; tri: 55%
	rK^2MP	di: 66%; tri: 23%	di: 19%; tri: 77%	di: 19%; tri: 76%
second-generation deoxyribonucleotide	dP^2MP	di: 5%; tri: 2%	di: 8%; tri: 3%	di: 3%; tri: 2%
	dX^2MP	di: 21%; tri: 3%	di: 18%; tri: 32%	di: 3%; tri: 6%
	dZ^2MP	di: 17%; tri: 26%	di: 22%; tri: 58%	di: 27%; tri: 24%
	dK^2MP	di: 26%; tri: 42%	di: 19%; tri: 72%	di: 13%; tri: 44%
third-generation deoxyribonucleotide	dP^3MP	di: 27%; tri: 4%	di: 19%; tri: 27%	di: 19%; tri: 47%
	dX^3MP	di: 16%; tri: none	di: 21%; tri: 27%	di: 7%; tri: 5%
	dK^3MP	di: 25%; tri: 45%	di: 19%; tri: 72%	di: 13%; tri: 44%

[a]By UV absorbance of the respective HPLC peaks. Addition of unreacted monophosphate makes the sum 100%.

The pathway in Figure 3 was implemented where the first step used γ-^{32}P-labeled ATP and the *D. melanogaster* nucleoside kinase Q81E mutant to make the AEGIS nucleoside monophosphate with a ^{32}P label. Here, it was expedient to remove leftover excess γ-^{32}P-labeled ATP by treating with sodium periodate ($NaIO_4$). This reagent cleaves the ribose ring but not the deoxyribose ring. Further with heat treatment at 85 °C for 10 min, both ATP and ADP were efficiently converted to adenosine dialdehyde (Figure 4A); the single step caused the elimination of the phosphate groups. The products did not

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

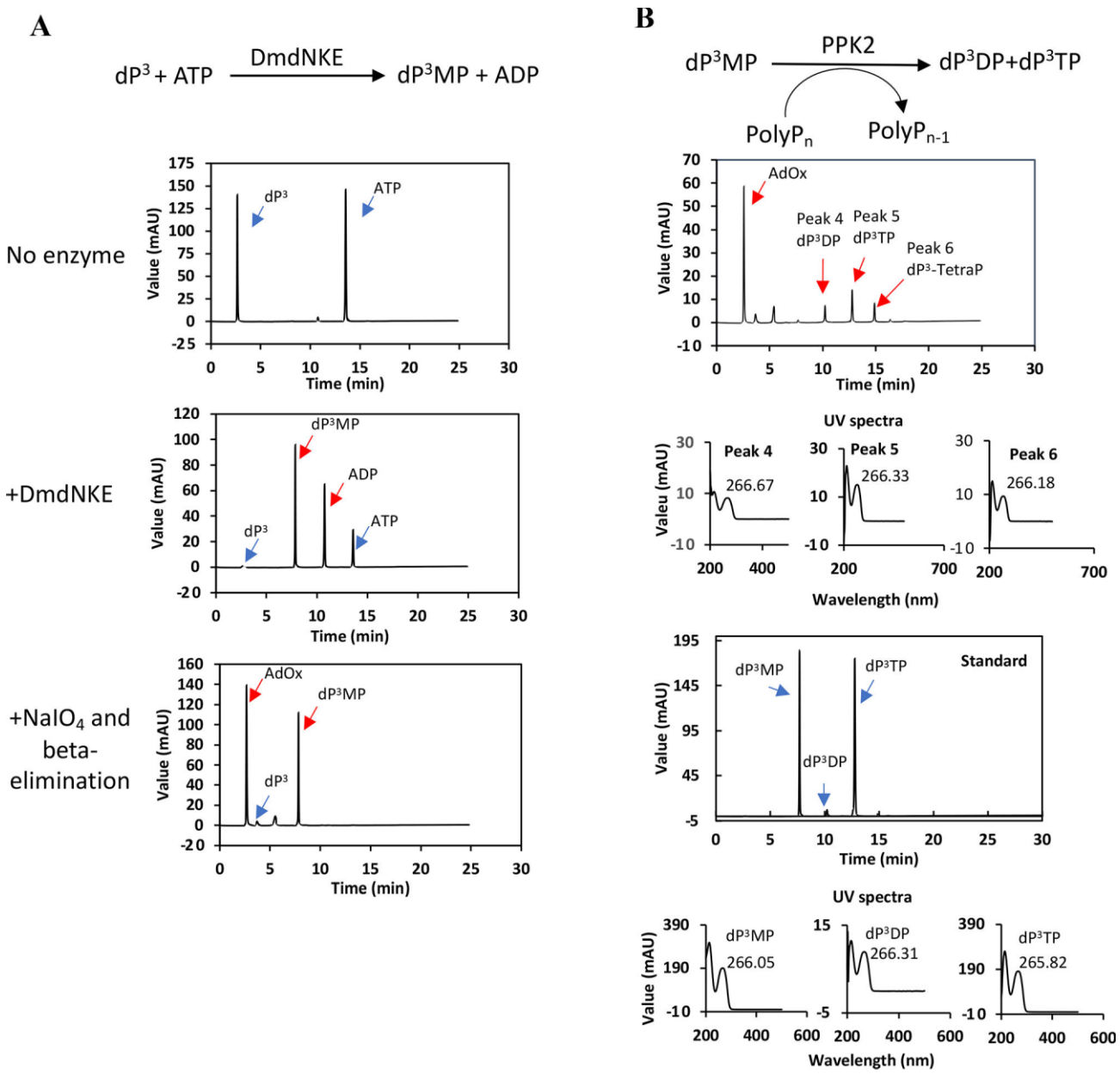


Figure 4. Two alternative pathways to make AEGIS nucleoside triphosphates. (A) Deoxynucleoside dP^3 was converted into monophosphate by DmdNKE with ATP as the phosphate donor. Excess ATP and ADP were destroyed by $NaIO_4$ and heating. Blue arrows, substrates; red arrows, products. AdOx: adenosine dialdehyde. (B) dP^3MP was converted into dP^3DP and dP^3TP by incubation with ChuPPK2 with $polyP_6$. Nucleotide products were confirmed by their UV spectra and retention time compared to authentic standards.

interfere with further phosphorylation reactions. Thus, nucleosides were smoothly converted to triphosphates, in one "pot", by sequentially adding ATP and the nucleoside kinase, periodate, heating, and then adding polyphosphate (commercial hexametaphosphate) and polyphosphate kinase (Figure 4B).

***α*-^{32}P-Labeled AEGIS Triphosphates Made by This Pathway Were Substrates for DNA Polymerases.** After the kinases were inactivated by heat, the *α*-^{32}P-labeled AEGIS triphosphates could be directly used without purification in primer extension reactions catalyzed by DNA polymerases. The following templates containing with $2K^1$ and $5Z^2$ (respectively 45 and 30 nucleotides long) were prepared by solid-phase synthesis from AEGIS phosphoramidites (Firebird Biomolecular Sciences LLC): KK template, 5′-CAT GTC TGA TCC TGC ACT GCT GGK^{1}K^{1}GG CCT TGA CTC TCG TAC CTG-3′; 5Z template, 5′-AGA GZ^{2}Z^{2}Z^{2}CZ^{2}Z^{2}CC ACC ACA CGC TGC TCC GAC-3′.

Gel electrophoresis showed that *α*-^{32}P-labeled AEGIS nucleoside triphosphates gave full-length radiolabeled oligonucleotides with various DNA polymerases (Figure 5A). Further after purification through ion-exchange HPLC, *α*-^{32}P-labeled AEGIS triphosphates were confirmed by TLC analysis (Figure 5B) and applied in the primer extension assay (Figure 5C).

Competition Assay Experiments to Measure the Fidelity of Incorporation. The *α*-^{32}P-labeled AEGIS triphosphates allowed us to do competition studies to compare second- and third-generation AEGIS deoxyribonucleoside triphosphates in polymerase extension reactions. With *α*-^{32}P-

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

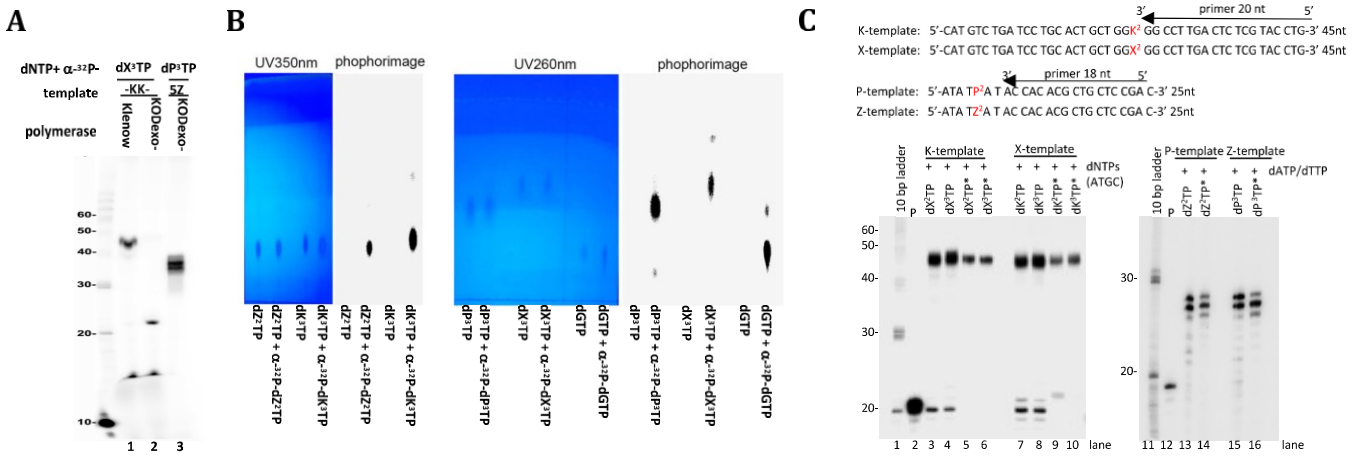


Figure 5. Analyses of α-^{32}P-labeled triphosphates by polymerase incorporation and TLC. (A) Primer extension with unpurified reaction mixtures that contain α-^{32}P-labeled dX^{3}TP or dP^{3}TP. Templates contain two K^{1}s or five Z^{2}s and reactions were carried out with a DNA polymerase Klenow fragment at 37 °C or KODexo- at 72 °C for 5 min. Samples were resolved on 18% urea denature PAGE. (B) TLC analysis of HPLC-purified α-^{32}P-labeled AEGIS triphosphates. 10 nmol of unlabeled authentic dZ^{2}TP, dK^{3}TP, dP^{3}TP, and dX^{3}TP alone or mixed with 50−80 pmol of HPLC-purified α-^{32}P-labeled (~1000 CPM) triphosphates was spotted on PEI cellulose plates and developed with 0.85 M KH_2PO_4 (pH 3.5) as the mobile phase. Both fluorescent images under UV exposure and phosphor images were recorded. (C) Primer extension assay with HPLC-purified α-^{32}P-labeled AEGIS triphosphates. The primer extension assay was carried out with Klenow large fragment at 37 °C for 90 s and samples were resolved on 18% PAGE with 7 M urea after being mixed with 1 vol of 90% formamide and 20 mM EDTA. In the left panel, lanes 3, 4, 7, 8, 13, and 15 are products from 5′-^{32}P-labeled primer and unlabeled AEGIS triphosphates, served as controls; lanes 5, 6, 9, 10, 14, and 16 are products from unlabeled primer and α-^{32}P-labeled AEGIS triphosphates (marked with asterisk). P: 5′-^{32}P-labeled primer only.

labeled AEGIS triphosphates, it is possible to compete hot and cold unlabeled materials in symmetric pairs of reactions to measure infidelity when both matching and mismatching triphosphates are present. Figure 6 shows an example of such experiments.

The results in Figure 6 show that the dX^{3}TP third-generation implementation of the puADA hydrogen bonding pattern is a better substrate for the Klenow fragment of DNA polymerase 1 than second-generation dX^{2}TP. This is unexpected, as the structure of dX^{2}TP is more "natural" than dX^{3}TP. Likewise, the dK^{3}TP third-generation implementation of the pyDAD hydrogen bonding pattern is a better substrate for this polymerase 1 than second-generation dK^{2}TP. This is also unexpected, as the structure of dK^{2}TP seems more "natural" than dK^{3}TP.

These results make available for the first time AEGIS nucleoside triphosphates where the α-phosphate is ^{32}P labeled. This *in vitro* pathway is a useful addition to the tools available to manipulate AEGIS, and it is made possible by the broad substrate specificity of the polyphosphate kinase. In particular, the labeled species are useful for analysis of the extent to which AEGIS nucleotides are incorporated into DNA, for example, inside of cells. This is applicable to the Romesberg strategy, where the triphosphate is imported from the growth medium by a triphosphate transporter to import it into the cell. Separately, data not reported here show that these second- and third-generation AEGIS triphosphates are also taken up into cells by the triphosphate importer originating from plastids and heterologously expressed in *E. coli*.

More importantly, this path offers a way of discarding the triphosphate importer architecture entirely when creating a semisynthetic organism. Since polyphosphate is abundant as a phosphate storage molecule, it should be possible to simply feed nucleosides to such semisynthetic organisms, have it enter the cell via nucleoside transporters, and be converted to triphosphates. These would then be available to replicate AEGIS DNA in plasmids and, eventually, chromosomes.

METHODS

Chemicals, Enzymes, and Other Reagents. If not mentioned elsewhere, chemicals were purchased from Sigma and enzymes for DNA manipulation such as restriction enzymes, DNA polymerases for PCR, and T4 polynucleotide kinase for 5′ labeling of primers were purchased from New England Biolabs (NEB). AEGIS nucleosides, nucleotides, and oligos were purchased from Firebird Biomolecular Sciences (Alachua, FL). Primers and oligos with natural nucleotides (ATGC) were purchased from Integrated DNA Technologies (IDT) (Coralville, Iowa). Plasmids and *E. coli* strains used in this study are listed in Table S3.

Expression and Purification of Recombinant His-Tagged PPK2s. Amino acid sequences of ChuPPK2 from *C. hutchinsonii* ATCC 33406 (ABG57400.1) and AaurPPK2 from *A. aurescens* TC1 (ABM08865.1)[26] were retrieved from NCBI and converted to their DNA coding sequences using *E. coli*-optimized codons with the online program provided by IDT (Table S4). DNA molecules encoding the kinase proteins were synthesized as gBlocks by IDT and cloned into a pASK-IBA43plus vector (IBA Lifesciences, Göttingen, Germany) using an In-Fusion HD cloning reagent (TaKaRa) following the manufacturer's instruction. The expressed proteins carried a 6× His-tag fused to their C-termini, respectively.

Both enzymes were expressed in a Novablue (DE3) (Novagen) strain with 100 ng/mL anhydrotetracycline (α-TC) induction in overnight cultures growing in LB media at 30 °C. Cells were lysed by sonication, and the supernatant following centrifugation was applied to TALON metal affinity resin (Clontech). Proteins were eluted in elution buffer (50 mM Na-phosphate buffer (pH 8.0), 300 mM NaCl, 150 mM imidazole, 0.1% Triton X-100) and then buffer-exchanged into 2× enzyme storage buffer (40 mM Tris−HCl (pH 7.5), 200 mM KCl, 2 mM DTT, 0.1% Tween-20) using Pierce Protein Concentrators PES, 10 K MWCO (Thermo Fisher) to the concentration of imidazole was less than 1 μM, and the purity

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

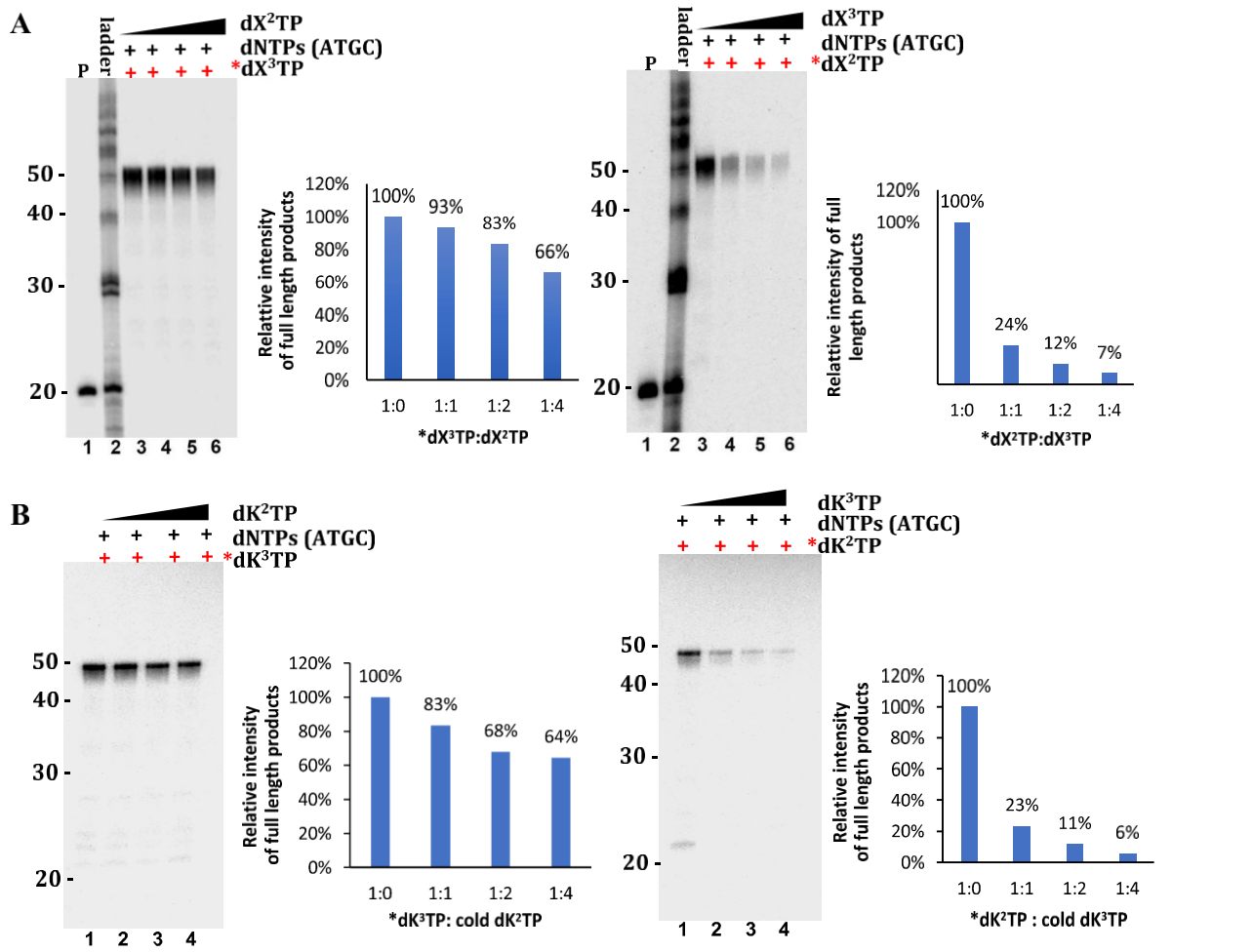


Figure 6. Competition experiments used various α-^{32}P-labeled AEGIS triphosphates to examine the preference of the Klenow fragment of *E. coli* DNA polymerase I for second- (Figure 1) versus third-generation (Figure 2) heterocycles implementing the puADA or pyDAD hydrogen bonding pattern. Primer extension assays were carried out with fixed amounts of α-^{32}P-labeled triphosphates and increasing amounts of unlabeled competitors. The full-length ^{32}P-labeled products were resolved on denaturing PAGE and quantified by phosphor imaging and densitometry. The relative amounts of products were plotted against the ratio of radiolabeled vs unlabeled triphosphates, where the intensity of the full-length product without unlabeled competitors was taken as 100%. (A) dX²TP vs dX³TP; (B) dK²TP vs dK³TP.

was checked by Coomassie blue-stained SDS-PAGE. Different batches of final products were combined and quantified with a QuickStart Bradford reagent (Bio-Rad) and then stored in storage buffer (20 mM Tris−HCl (pH 7.5), 100 mM KCl, 1 mM DTT, 0.05% Tween-20, 40% glycerol) at −20 °C.

For expression and purification of His-tagged DtCsPPK2 (*Delftia tsuruhatensis*), the DtCsPPK2 coding region from a construct pET28-DtCsPPK2 kindly provided by the Matsuura group[28] was PCR amplified and cloned into a home-constructed pRS vector derived from pASK-IBA43+. Protein expression and purification followed the procedure described above.

Phylogenetic Studies. The analysis involved a group of amino acid sequences of Class III PPK2 retrieved from NCBI based on the previously listed in the literature[27] with addition of recently characterized PPK2s.[26,28] Evolutionary analyses were conducted in MEGA11.[29] Multiple sequence alignment of the listed PPK2s was generated with Clustal Omega and used to construct the phylogenetic tree with the UPGMA method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 423 positions in the final data set. The full list of PPK2s is included in Table S2.

PPK2 Kinase Assay. In 20 μL reactions, 2 mM NMP or dNMP was mixed with 1× PPK2 buffer (50 mM MOPs (pH 7.0), 10 mM $MnCl_2$),[18] 1 mM polyP$_6$ (sodium hexametaphosphate, Sigma), and 1 μM PPK2 on ice. After incubation at 37 °C for 90 min, the mixtures were diluted to 100 μL by adding 80 μL of sterilized water and enzymes were removed by centrifugation through a Microcon 10KD-cutoff spin column (MilliporeSigma) at 14,000 × *g* for 20 min. The pass-through product (~90 μL) was transferred to a new tube and further centrifuged to move particulates at 17,000 × *g* for 10 min. Aliquots (80 μL) of each sample were transferred to HPLC sample vials and 10 μL of samples was injected and resolved by

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

ion-exchange HPLC (Waters 2695 separation module equipped with a Waters 2996 photodiode array detector and a DNAPac PA-100 BioLC 4× 250 mm analytical column (Thermo Fisher)) using water (A)/1 M NH_4HCO_3 (B) as the mobile phase with the gradient (0−20 min 0−20% B, 20−22 min 20−50% B, 22−25 min 50% B, and 25−27 min 0% B). The amounts of products (nucleoside diphosphate, triphosphate, etc.) were calculated from the overall integrated area of the resolved peaks, assuming that the extinction coefficient did not change as the level of phosphorylation changed.

Kinetics of the DmdNKE Kinase on AEGIS Nucleosides. DmdNKE kinase activities on AEGIS nucleosides were examined using the pyruvate kinase and lactate dehydrogenase-coupled enzyme assay.[30] The kinase reaction was carried out in solution containing 50 mM Tris−HCl (pH 7.5), 100 mM KCl and 2.5 mM $MgCl_2$, 0.18 mM NADH, 0.21 mM phosphoenolpyruvate, 1 mM ATP, 1 mM 1,4-dithio-DL-threitol, 30 U/mL pyruvate kinase, 33 unit/mL lactate dehydrogenase, and 250 nM DmdNKE. Substrate nucleoside ranges in mM: 0, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2, 2.5, and 3.0. Buffer components were prepared in 5× stock solution that was filter sterilized with a 0.22 μm PVDF filter. Substrate component stocks were stored at −20 °C, except for NADH, which was stored at −80 °C. All components were prepared in a Master Mix excluding the nucleoside substrate. Reaction mixtures were then loaded into the 96-well plate with a multichannel pipette in succession. Reactions were performed in a 96-well plate (Axygen), in 30 μL volumes, and read in the Omega reader at 340 nm to monitor the decrease in absorbance. For kinetic calculation, the fluorescent signal obtained was converted to velocity (V) and plotted against substrate concentrations [S]. The data were plotted as the Lineweaver−Burk plot ($\frac{1}{V} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$) to obtain Km and Vmax. The mean values from three replicates were used to calculate Kcat/Km.

The DmdNKE kinase assay with AEGIS nucleosides was also analyzed with ion-exchange HPLC. In brief, in total 20 μL reaction, 1 mM ATP was mixed with 0.8 mM AEGIS nucleosides in 1× DmdNKE buffer 2 (50 mM HEPES-KOH, pH 7.5, 20 mM KCl, 1 mM DTT) and 1 μM DmdNKE and incubated for 1 h at 37 °C, followed by centrifugation through a Microcon 10KD-cutoff spin column (MilliporeSigma) to remove kinases. The products were diluted with deionized water to 100 μL and 10 μL was injected into Thermo Fisher UHPLC (Dionex Ultimate 3000 system equipped with a diode array detector and a DNAPac PA-100 BioLC 4× 250 mm analytical column) and resolved with H2O (A):1 M NH4HCO3 (B) gradient (0−1 min 0−0.5% B, 1−3 min 0.5−10% B, 3−10 min 10−30% B, 10−25 min 30−70% B, 25−30 min 70−80% B, 30−33 min 80−90% B, and 33−36 min 0% B).

Synthesis of α-^{32}P-Labeled AEGIS Nucleoside Triphosphates in a Coupled Kinase Pathway. The coupled kinase reactions were carried out in 30 μL volume. AEGIS nucleosides (0.8 mM final concentration), including dX[2], dX[3], dP[3], dK[2], dK[3], and dZ[2], were individually mixed with 1× DmdNKE buffer 2, 1 mM ATP, 3 μL of γ-^{32}P-ATP (6000 Ci/mmol, 10 mCi/mL, PerkinElmer), and 1 μM glycerol-free DmdNKE. The reaction mixture was incubated at 37 °C for 2 h, followed by addition of 10 mM $NaIO_4$ and incubation at 25 °C for 10 min. After the excess of $NaIO_4$ was neutralized by addition of 20 mM pinacol at 25 °C for another 10 min, the reaction mixture was incubated at 85 °C for 10 min and then cooled on ice. The second step of the coupled reaction was continued by adding 1× PPK2 buffer, 0.5 mM $polyP_6$, and 1 μM ChuPPK2, followed by incubation at 37 °C for 90 min. The final volume of the reaction mixture was 60 μL. The reactions were stopped by heating at 85 °C for 10 min, and the mixtures were centrifuged at 17,000 × g for 10 min. The supernatant (~55 μL) was recovered, diluted to 80 μL by adding sterilized water, and manually injected into HPLC (Waters 600) equipped with a DNAPac PA-100 BioLC 4× 250 mm analytical column (Thermo Fisher), a 100 μL injection loop, and a dual λ UV detector (Waters 2487). The fractions of deoxynucleoside triphosphates were collected, frozen in liquid nitrogen, and dried in SpeedVac overnight. The dried samples were dissolved in 20 μL of sterilized water and the concentrations of purified nucleoside triphosphates were quantified using NanoDrop 2000 (Thermo Fisher) with their maximum UV absorbance, respectively. The total radioactivity of each purified nucleoside triphosphate was measured with a scintillation counter (Tri-Carb 2800TR, PerkinElmer) using the Cherenkov method.

To monitor the yields of products in each step of the reaction, the same set of reactions was carried out in parallel, except that no radiolabeled ATP was added. Aliquots (5 μL) of the mixtures were taken at the end of each kinase reaction and diluted to 50 μL with deionized water, followed by injecting 10 μL for HPLC analysis using Thermo Fisher UHPLC equipped with the same column using the same setting of the mobile phase gradient.

Primer Extension Assay. For the regular primer extension assay with 5′-^{32}P-labeled primers, reactions were carried out in 10 μL volume. DNA templates (10 pmol) were mixed with 7 pmol of the primer (Table S1), 0.1 pmol of the 5′-^{32}P-labeled primer, 100 μM natural dNTP (ATGC), 100 μM AEGIS deoxynucleoside triphosphates, and 1× KOD buffer (TOYOBO, Japan) or 1× Klenow buffer (NEB) on ice. The mixture was heated at 94 °C for 1 min and cooled to 25 °C at 0.1 C/s to anneal primers and templates. The primer extension was initiated by adding 0.02 U of KODexo- DNA polymerases (TOYOBO, Japan) into the prewarmed primer−template mixture at 72 °C or 0.1 U of Klenow at 37 °C. After 1 min incubation, the reaction was stopped by addition of one volume of sample buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF). All samples were resolved on 18 or 20% 7 M urea PAGE in 1× TBE buffer. Autoradiography was carried out by exposure of dried gels to the phosphor screen, which was then scanned with Personal Molecular Imager (Bio-Rad).

For primer extension with α-^{32}P-labeled AEGIS deoxynucleoside triphosphates, the assay followed the procedure above, except that non-radiolabeled primers were used and the unlabeled AEGIS nucleoside triphosphates were replaced with 100 μM α-^{32}P-labeled AEGIS deoxynucleoside triphosphates (0.1−0.2 μCi) purified from coupled kinase reactions.

For competition experiments, α-^{32}P-labeled AEGIS deoxynucleoside triphosphates at a fixed concentration of 50 μM were mixed with variable concentrations of unlabeled competitors at 0, 50, 100, and 200 μM and 50 μM dNTPs (ATGC) in the assay. Primer extension reactions were carried out, as mentioned above. The signals of full-length products were quantified from scanned phosphor images by densitometric analysis of bands with Quantity One software (Bio-Rad).

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00060.

Extended HPLC profiles for DmdNKE and three PPK2 kinase reactions, a phylogenetic tree of selected class III PPK2s, lists of oligos, plasmids, *E. coli* strains, and sequence files of synthesized PPK2 used in this study (PDF)

AUTHOR INFORMATION

Corresponding Author

Yubing Li − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States;* orcid.org/0000-0002-5850-7678; Email: yli@ffame.org

Authors

Clay Abraham − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States;* orcid.org/0000-0002-7126-5896

Oleg Suslov − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States;* orcid.org/0009-0003-6219-0997

Ozlem Yaren − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States;* orcid.org/0000-0001-8051-4527

Ryan W. Shaw − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States*

Myong-Jung Kim − *Firebird Biomolecular Sciences LLC, Alachua, Florida 32615, United States*

Shuo Wan − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States*

Philippe Marliere − *Institute of Systems & Synthetic Biology, 91030 Evry Cedex, France*

Steven A. Benner − *Foundation for Applied Molecular Evolution, Alachua, Florida 32615, United States; Firebird Biomolecular Sciences LLC, Alachua, Florida 32615, United States;* orcid.org/0000-0002-3318-9917

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.3c00060

Author Contributions

Y.L., P.M., and S.A.B. designed the research; Y.L., C.A., O.S., O.Y., R.W.S., M.-J.K., and S.W. performed the research; Y.L. and S.A.B. analyzed the data; Y.L. and S.A.B. wrote the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Research reported in this publication is based upon the work supported by the National Science Foundation under grant no. MCB-2123995. It was further supported in part by the National Institutes of Health under award number 5R01GM128186-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NIH.

REFERENCES

(1) Ellington, A. D. Defining an emerging field can be challenging. Nature Biotechnology asked 20 experts for their views on the term 'synthetic biology'. *Nat. Biotechnol.* 2009, *27*, 1071.

(2) Cox, J. P. L. Long-term data storage in DNA. *Trends Biotechnol.* 2001, *19*, 247−250.

(3) Matange, K.; Tuck, J. M.; Keung, A. J. DNA stability: a central design consideration for DNA data storage systems. *Nat. commun.* 2021, *12*, 1358.

(4) Lorsch, J. R.; Szostak, J. W. Chance and necessity in the selection of nucleic acid catalysts. *Acc. Chem. Res.* 1996, *29*, 103−110.

(5) Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Fajardo, D.; Moussatche, P. (2016) Alternative Watson-Crick synthetic genetic systems. *Synthetic Biology.*; Cold Spring Harbor Perspectives in Biology, Cold Spring Harbor Press. PMID: 27663774. doi: DOI: 10.1101/cshperspect.a023770

(6) Zhang, L.; Yang, Z.; Sefah, K.; Bradley, K. M.; Hoshika, S.; Kim, M.-J.; Kim, H.-J.; Zhu, G.; Jimenez, E.; Cansiz, S.; Teng, I.-T.; Champanhac, C.; McLendon, C.; Liu, C.; Zhang, W.; Gerloff, D. L.; Huang, Z.; Tan, W.; Benner, S. A. Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.* 2015, *137*, 6734−6737.

(7) Sefah, K.; Yang, Z.; Bradley, K. M.; Hoshika, S.; Jimenez, E.; Zhu, G.; Shanker, S.; Yu, F.; Tan, W.; Benner, S. A. In vitro selection with artificial expanded genetic information systems (AEGIS). *Proc. Natl. Acad. Sci. U. S. A.* 2014, *111*, 1449−1454.

(8) Biondi, E.; Lane, J. D.; Das, D.; Dasgupta, S.; Piccirilli, J. A.; Hoshika, S.; Bradley, K. M.; Krantz, B. A.; Benner, S. A. Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.* 2016, *44*, 9565−9577.

(9) Jerome, C. A.; Hoshika, S.; Bradley, K. M.; Benner, S. A.; Biondi, E. In vitro evolution of ribonucleases from expanded genetic alphabets. *Proc. Natl. Acad. Sci. U. S. A.* 2022, *119*, No. e2208261119.

(10) Jayasena, S. D. Aptamers. An emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* 1999, *45*, 1628−1650.

(11) www.firebirdbio.com

(12) Laos, R.; Shaw, R. W.; Leal, N. A.; Benner, S. A. Directed evolution of polymerases to accept nucleotides with non-standard hydrogen-bond patterns. *Biochemistry* 2013, *52*, 5288−5294.

(13) Zhang, L.; Wang, S.; Yang, Z.; Hoshika, S.; Xie, S.; Li, J.; Chen, X.; Wan, S.; Li, L.; Benner, S. A.; Tan, W. An aptamer-nanotrain assembled from six-letter DNA delivers doxorubicin selectively to liver cancer cells. *Angew. Chem., Int. Ed.* 2020, *59*, 663−668.

(14) Kool, E. T.; Morales, J. C.; Guckian, K. M. Mimicking the structure and function of DNA. Insights into DNA stability and replication. *Angew. Chem., Int. Ed.* 2000, *39*, 990−1009.

(15) Kimoto, M.; Yamashige, R.; Matsunaga, K. I.; Yokoyama, S.; Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* 2013, *31*, 453−457.

(16) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* 2008, *130*, 2336−2343.

(17) Malyshev, D. A.; Praff, D. A.; Ippoliti, S. I.; Hwang, G. T.; Dwyer, T. J.; Romesberg, F. E. Solution Structure, Mechanism of Replication, and Optimization of an Unnatural Base Pair. *Chem. - Eur. J.* 2010, *16*, 12650−12659.

(18) Gao, J.; Liu, H.; Kool, E. T. Assembly of the complete eight-base artificial genetic helix, xDNA, and its interaction with the natural genetic system. *Angew. Chem., Int. Ed.* 2005, *44*, 3118−3122.

(19) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Romesberg, F. E. A semi-synthetic organism with an expanded genetic alphabet. *Nature* 2014, *509*, 385−388.

(20) Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; San Jose, K.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 2017, *551*, 644−647.

(21) Feldman, A. W.; Fischer, E. C.; Ledbetter, M. P.; Liao, J. Y.; Chaput, J. C.; Romesberg, F. E. A Tool for the Import of Natural and Unnatural Nucleoside Triphosphates into Bacteria. *J. Am. Chem. Soc.* 2018, *140*, 1447−1454.

(22) Espinasse, A.; Lembke, H. K.; Cao, A. A.; Carlson, E. E. Modified nucleoside triphosphates in bacterial research for *in vitro* and live-cell applications. *RSC Chem. Biol.* 2020, *1*, 333−351.
(23) Patching, S. G.; Baldwin, S. A.; Baldwin, A. D.; Young, J. D.; Gallagher, M. P.; Henderson, P. J.; Herbert, R. B. The nucleoside transport proteins, NupC and NupG, from Escherichia coli: specific structural motifs necessary for the binding of ligands. *Org. Biomol. Chem.* 2005, *3*, 462−470.
(24) Matsuura, M. F.; Shaw, R. W.; Moses, J. D.; Kim, H.-J.; Kim, M.-J.; Kim, M.-S.; Benner, S. A. Assays to detect the formation of triphosphates of unnatural nucleotides. Application to Escherichia coli nucleoside diphosphate kinase (Ec-NDPK). *ACS Synth. Biol.* 2016, *5*, 234−240.
(25) Matsuura, M. F.; Winiger, C. B.; Shaw, R. W.; Kim, M.-J.; Kim, M.-S.; Daugherty, A. B.; Chen, F.; Moussatche, P.; Moses, J. D.; Lutz, S.; Benner, S. A. A single deoxynucleoside variant from *Drosophila melanogaster* synthesizes the monophosphates of nucleosides that are components of an expanded genetic system. *ACS Synth. Biol.* 2016, *6*, 388−394.
(26) Nocek, B. P.; Khusnutdinova, A. N.; Ruszkowski, M.; Flick, R.; Burda, M.; Batyrova, K.; Brown, G.; Mucha, A.; Joachimiak, A.; Berlicki, Ł.; Yakunin, A. F. Structural Insights into Substrate Selectivity and Activity of Bacterial Polyphosphate Kinases. *ACS Catal.* 2018, *8*, 10746−10760.
(27) Motomura, K.; Hirota, R.; Okada, M.; Ikeda, T.; Ishida, T.; Kuroda, A. A new subfamily of polyphosphate kinase 2 (class III PPK2) catalyzes both nucleoside monophosphate phosphorylation and nucleoside diphosphate phosphorylation. *Appl. Environ. Microbiol.* 2014, *80*, 2602−2608.
(28) Ogawa, M.; Uyeda, A.; Harada, K.; Sato, Y.; Kato, Y.; Watanabe, H.; Honda, K.; Matsuura, T. Class III Polyphosphate Kinase 2 Enzymes Catalyze the Pyrophosphorylation of Adenosine-5′-Monophosphate. *ChemBioChem* 2019, *20*, 2961−2967.
(29) Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 2021, *38*, 3022−3027.
(30) Kornberg, A.; Pricer, W. E., Jr. Di- and triphosphopyridine nucleotide isocitric dehydrogenases in yeast. *J. Biol. Chem.* 1951, *189*, 123−136.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on May 25, 2023. The presentation of Figure 4a has been corrected and was reposted on June 02, 2023.

https://doi.org/10.1021/acssynbio.3c00060
ACS Synth. Biol. 2023, 12, 1772−1781