Supporting Information

A universal polyphosphate kinase powers in vitro transcription

Ryusei Matsumoto[1,‡], Takayoshi Watanabe[2,‡,†], Eishin Yamazaki[1], Ako Kagawa[2], Liam M. Longo[2,3,*], and Tomoaki Matsuura[2,*]

[1]Department of Life Science and Technology, Institute of Science Tokyo, Tokyo, Japan

[2]Earth-Life Science Institute, Institute of Science Tokyo, Tokyo, Japan

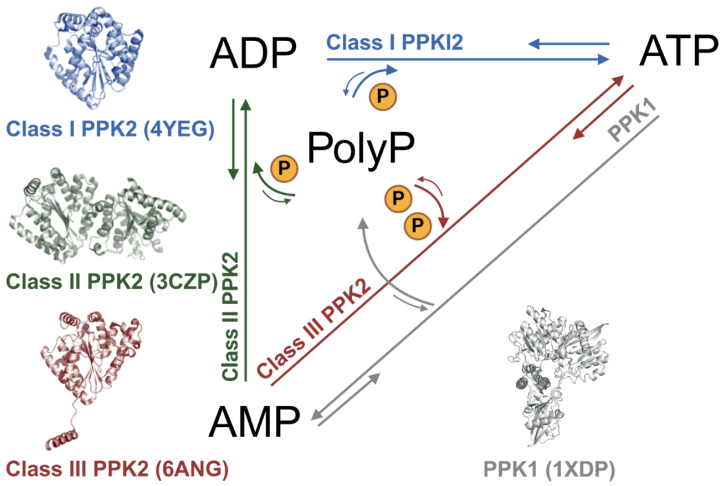
[3]Blue Marble Space Institute of Science, Seattle, USA

[‡]The authors contributed equally

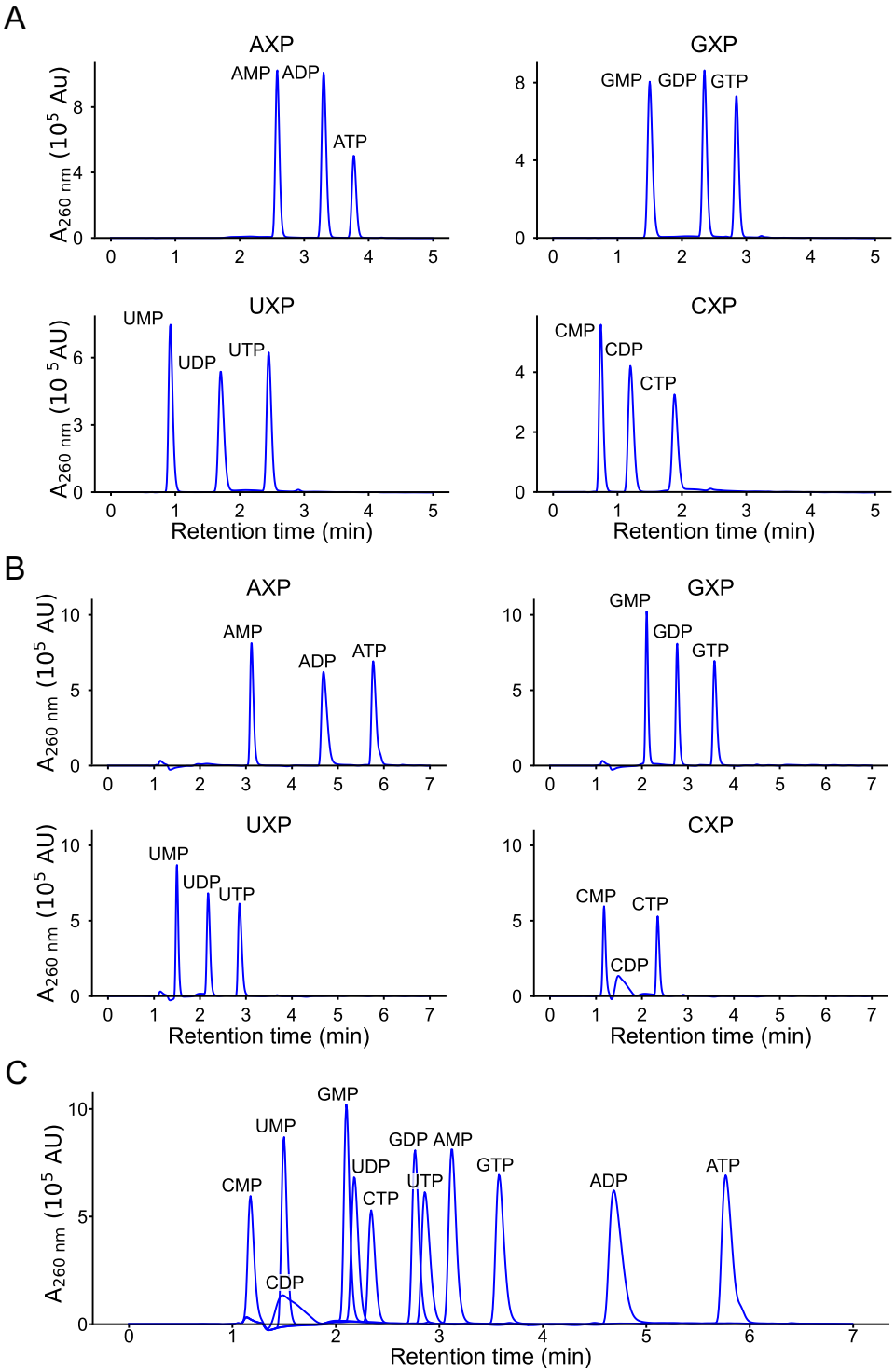
[†]Current address: MAQsys Inc., Kanagawa, Japan

*To whom correspondence should be addressed: llongo@elsi.jp, matsuura_tomoaki@elsi.jp

Supplementary Figures

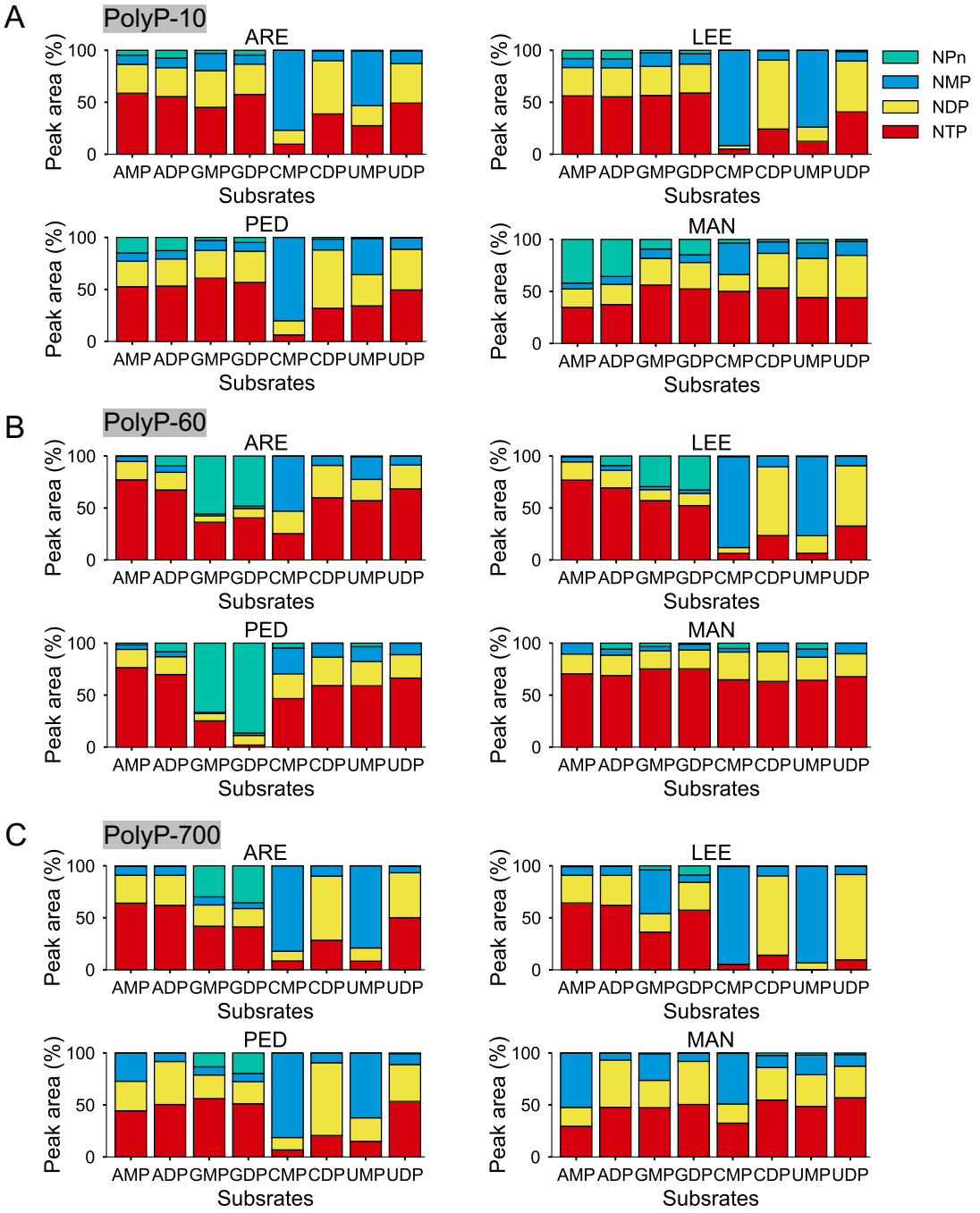


Supplementary Figure 1. Schematic showing PPK catalytic diversity. Both PPK1 and PPK2 use PolyP and nucleotides as substrates. PPK2 is divided into three classes in accordance with the phylogenetic tree. Catalytic profiles, however, may be not strictly distinguished by class.

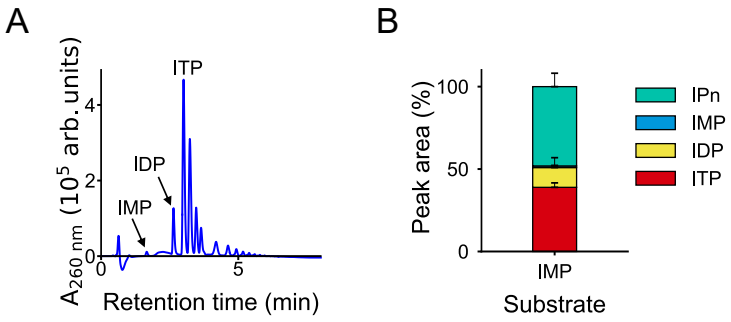


Supplementary Figure 2. HPLC analysis of nucleotides using the standard (A) and enhanced (B) protocols. Panel C shows overlaid chromatograms obtained with the enhanced protocol. "X" indicates a mixture of mono- (M), di- (D), and tri- (T) phosphate forms. For each base, nucleotides with different degrees of phosphorylation elute as distinct peaks with almost no overlap on a single HPLC column

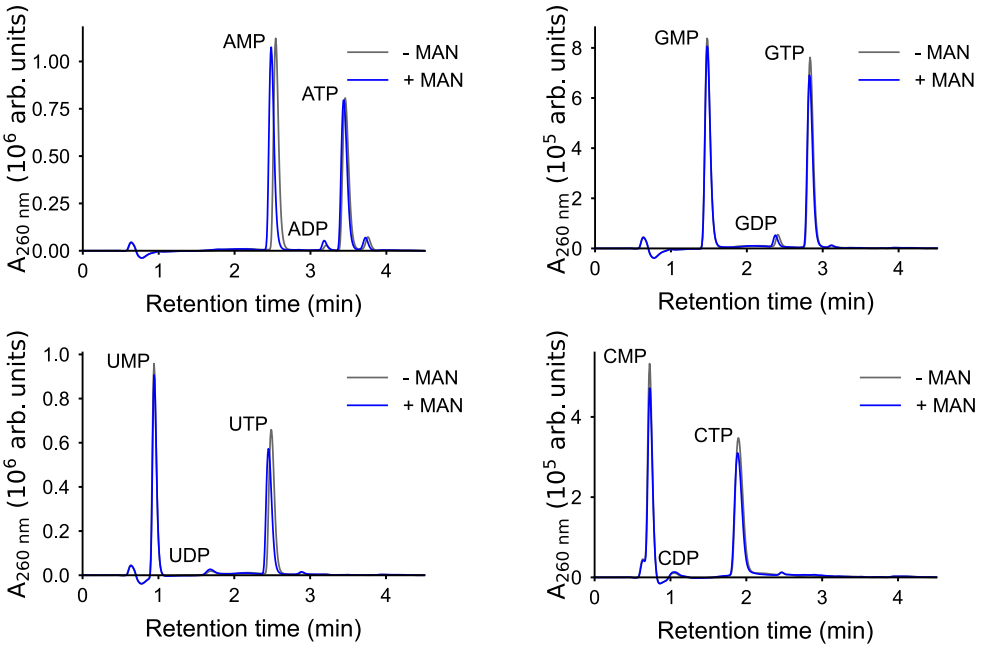
(standard protocol) or two columns in series (enhanced protocol). Authentic samples of 333 μM of each nucleotide were analyzed under identical conditions. Source data for this figure is available in the Source Data file.



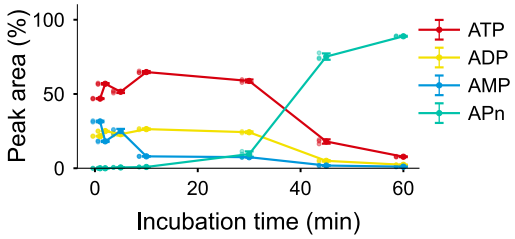
Supplementary Figure 3. NMP and NDP phosphorylation activity of Class III PPK enzymes (ARE, LEE, PED, and MAN) using PolyP of different lengths. Figure 2B reports these same data as NTP synthesis efficiency, while the bar plots here show the product distribution with (A) Poly10, (B) PolyP50, and (C) PolyP700. Source data for this figure is available in the Source Data file.



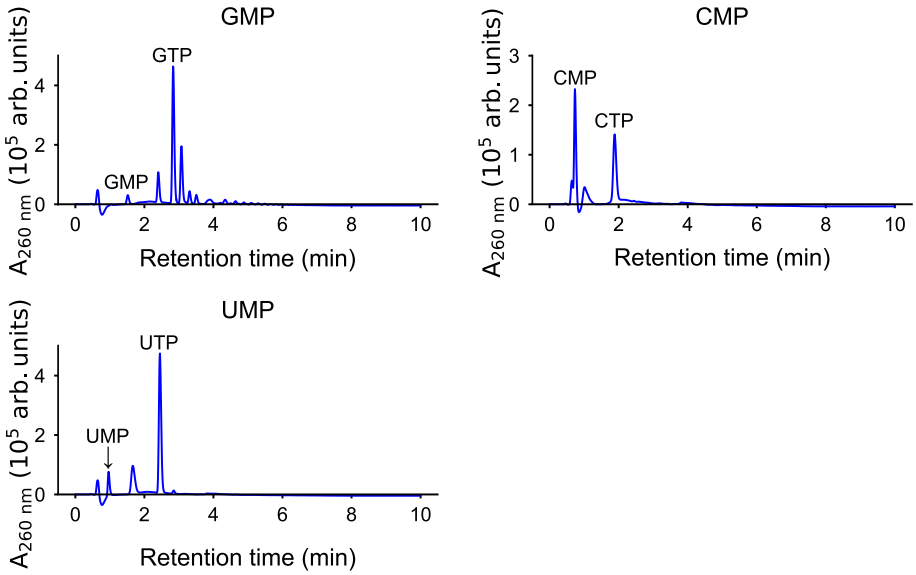
Supplementary Figure 4. Inosine monophosphate (IMP) phosphorylation by MAN. Reactions were carried out at 37 °C in the presence of 4 mM IMP, 20 mM PolyP-60, and 10 mM Mn^{2+}. (A) Representative HPLC chromatograms and (B) the average product distribution (right) are shown. Plotted values are the average of 3 independent runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.



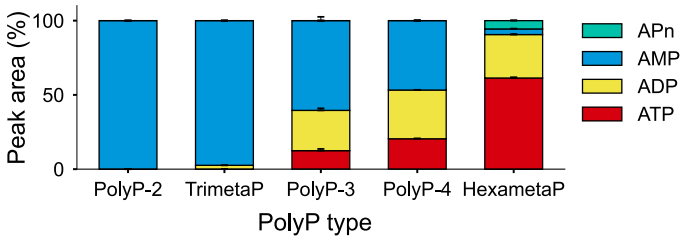
Supplementary Figure 5. MAN does not efficiently catalyze nucleoside diphosphate production from monophosphates and triphosphates. Reactions were performed at 37 °C with 4 mM NMP and 4 mM NTP of the same base, and 10 mM Mg^{2+}. The blue chromatogram represents reactions with MAN and the gray chromatogram represents reactions without MAN. Source data for this figure is available in the Source Data file.



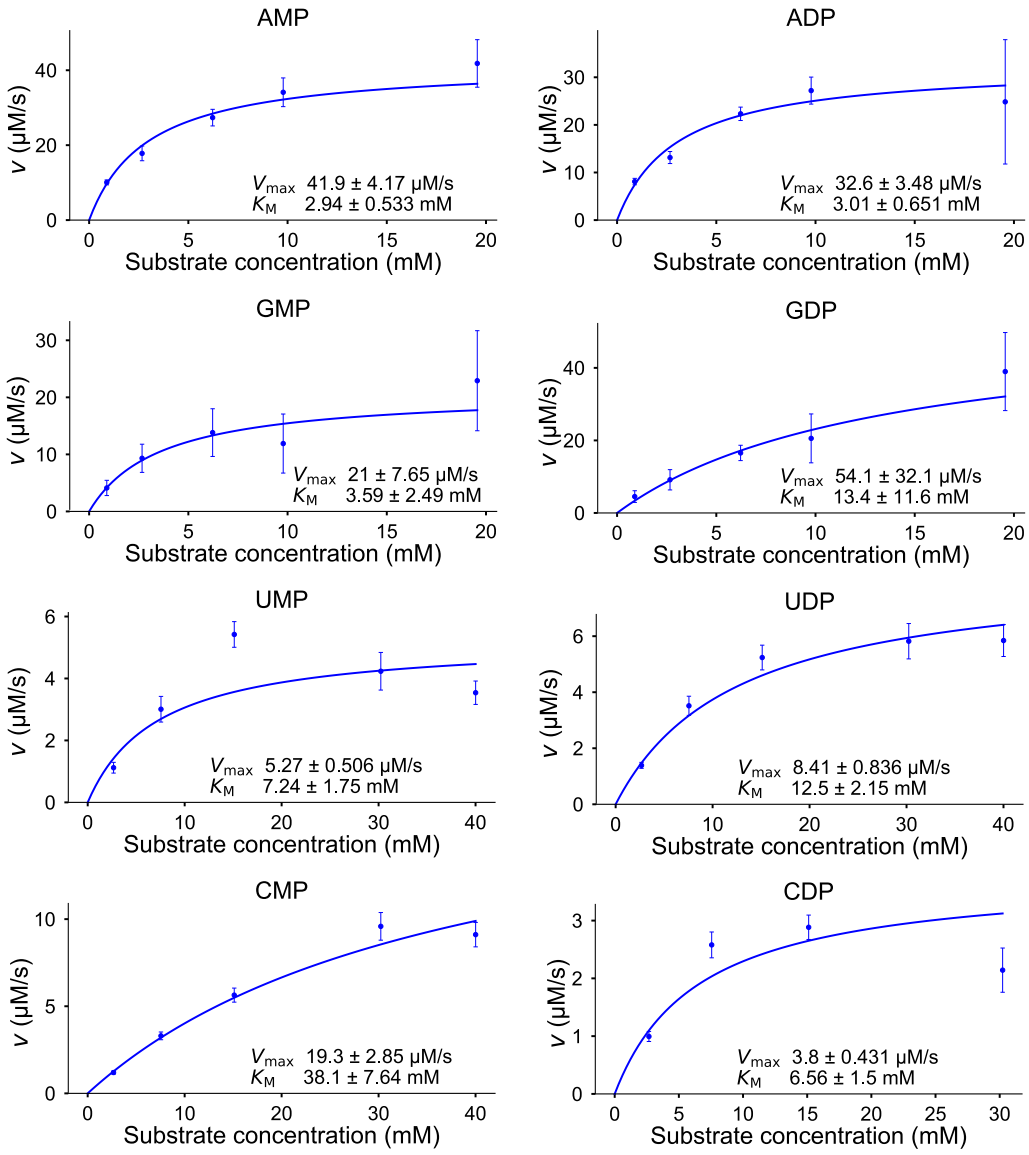
Supplementary Figure 6. Time-course of APn production by MAN at 55 °C. Reactions contained 4 mM AMP, 65 mM phosphate units of PolyP-60, and 10 mM Mn^{2+}. Solid lines and larger datapoints with error bars represent the average of 3 independent runs. Error bars are standard deviations. Smaller points to the left of the averaged data are the results from the individual runs. Source data for this figure is available in the Source Data file.



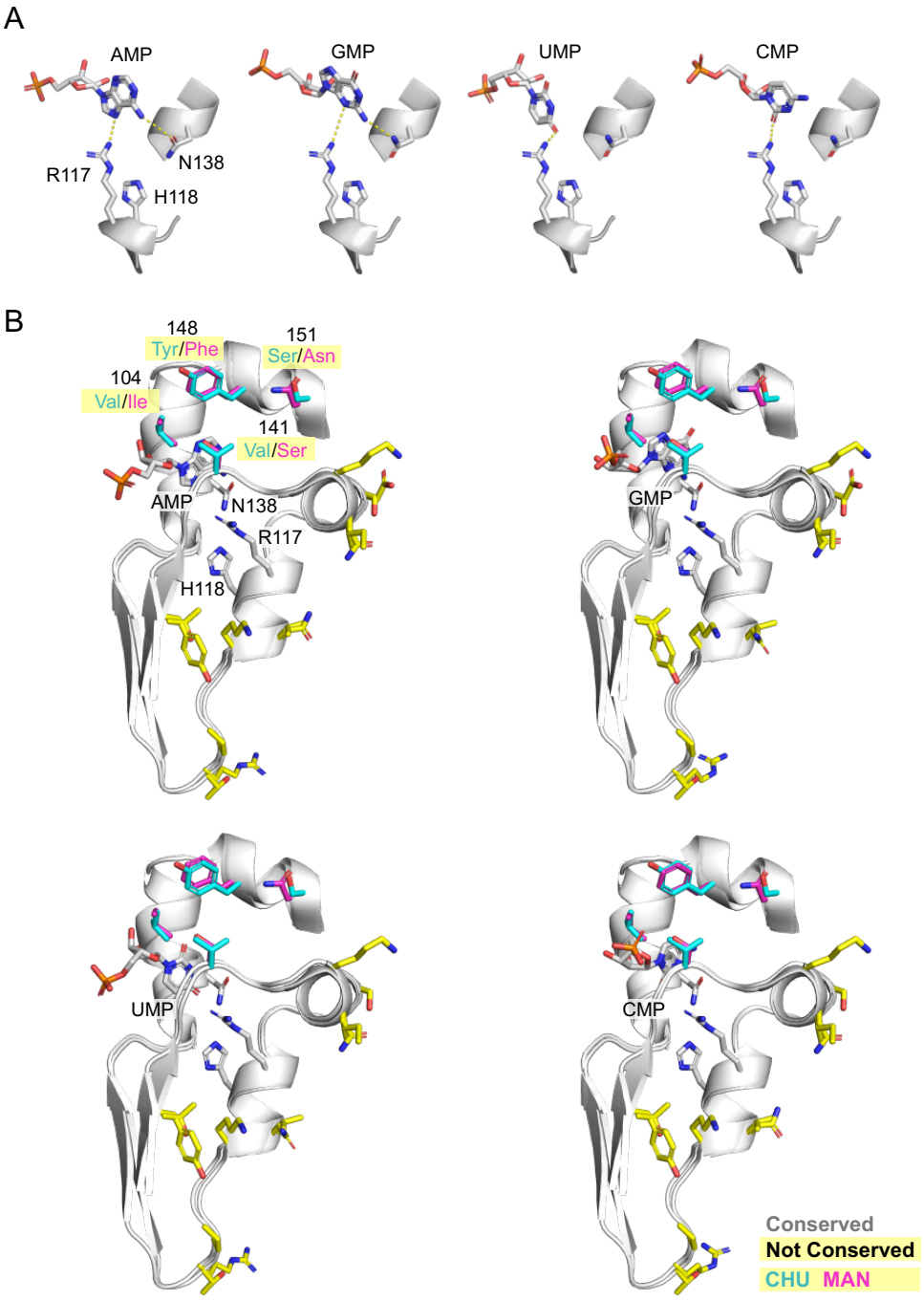
Supplementary Figure 7. Polyphosphate nucleotides generated by MAN. The reaction was carried out at 55 °C as in Figure 3A except that GMP, CMP, or UMP was used instead of AMP. While some minor over-phosphorylation of GMP was observed, neither of the pyrimidine nucleotides were significantly over-phosphorylated. Source data for this figure is available in the Source Data file.



Supplementary Figure 8. Activity of MAN with short-chain polyphosphates. Reactions were performed at 37 °C and contained 4 mM AMP, 20 mM phosphate units of diphosphate (PolyP-2), trimetaphosphate (TrimetaP), triphosphate (PolyP-3), tetraphosphate (PolyP-4), or hexametaphosphate (HexametaP), and 10 mM Mg^{2+}. Plotted values are the average of 3 independent runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.

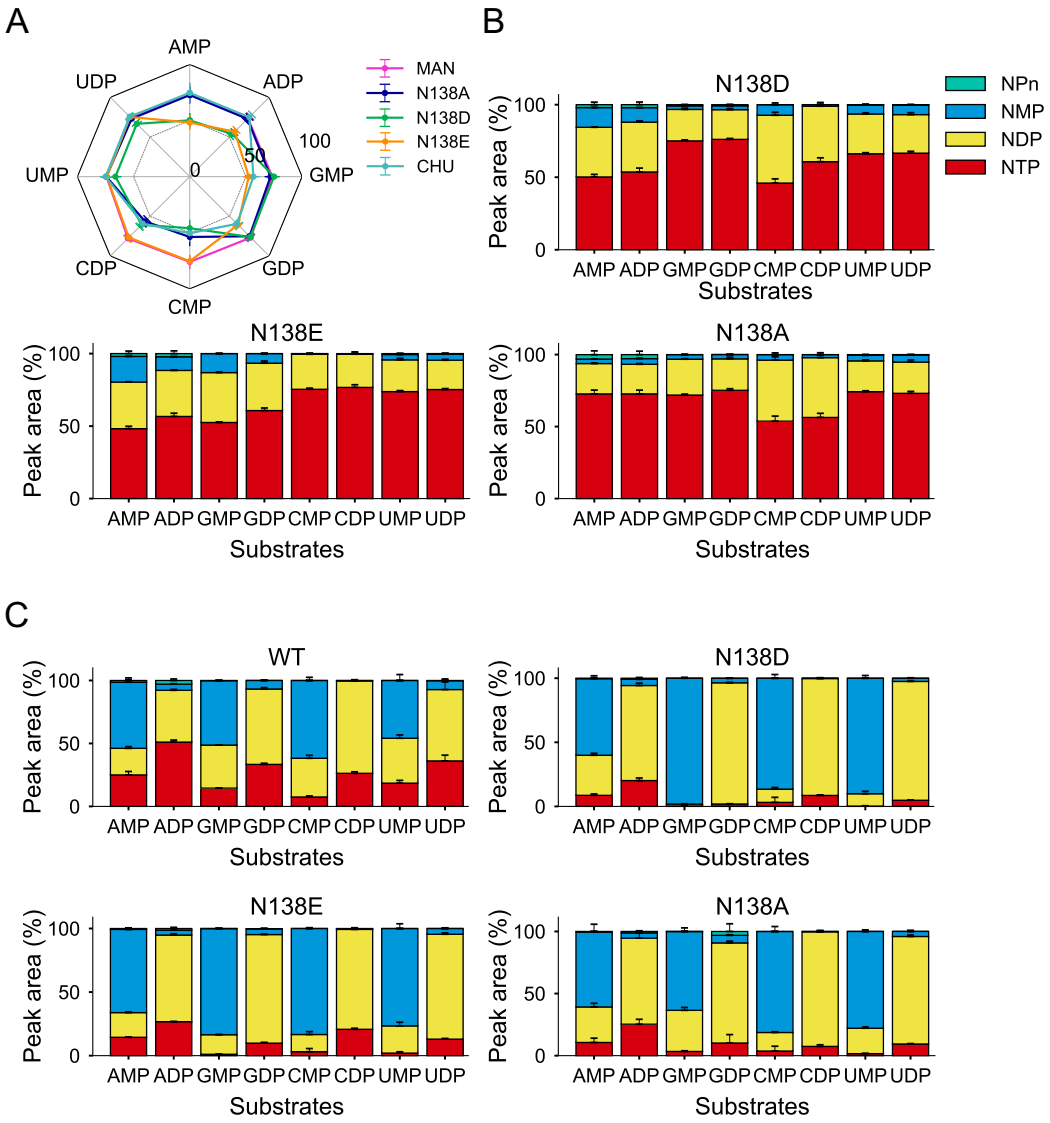


Supplementary Figure 9. Michaelis-Menten fits from which k_{cat} and K_M values were determined. Initial velocity (v) was determined from the rate of decrease of the initial substrate. Fitted kinetic parameters are inset and listed in Table 1. The error bars on each initial velocity measurement were calculated from the fit error of a line to the pooled (unaveraged) data from three independent experiments. Errors on the kinetic parameters were calculated from the fit to the Michaelis-Menten equation, taking into consideration the initial velocity errors. Note that GMP and GDP have significantly larger errors in initial reaction velocity at high concentrations, which may be due to the tendency for guanine nucleotides to assemble in solution. Consequently, the errors on k_{cat} and K_M for GMP and GDP are larger than for other nucleotides. Source data for this figure is available in the Source Data file.

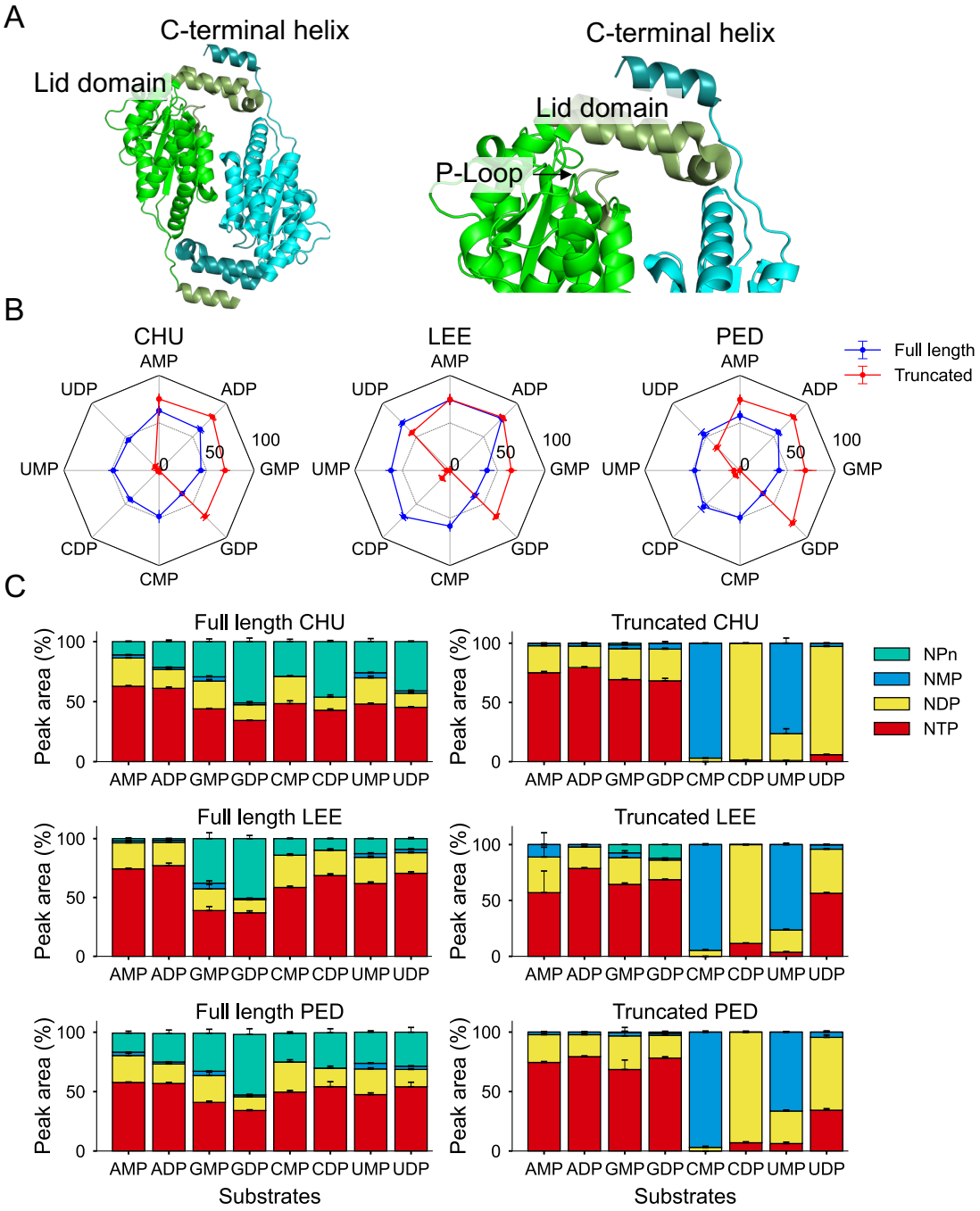


Supplementary Figure 10. Predicted binding of MAN with AMP, GMP, UMP, and CMP. (A) Hydrogen bonds formed with the base of MAN-bound nucleotides and (B) an overlay of the base binding pockets of MAN and CHU. Sequence differences in the primary base binding pocket are drawn as cyan and magenta sticks for CHU and MAN, respectively. Sequence differences that point away from the base binding pocket and were not characterized further are shown as yellow sticks. Additional information can be found in Figure 4. Structures were predicted using Protenix with the MAN sequence and substrates as inputs. Note that the guanidine moiety in the predicted GMP binding mode is flipped

relative to the crystal structure (Figure 1A). For the AMP-bound state, pTM = 0.93 and ipTM = 0.88; for the GMP-bound state, pTM = 0.92 and ipTM = 0.84; for the UMP-bound state, pTM = 0.92 and ipTM = 0.78; and for the CMP-bound state, pTM = 0.92 and ipTM = 0.77. Source data for this figure is available in the Source Data file.

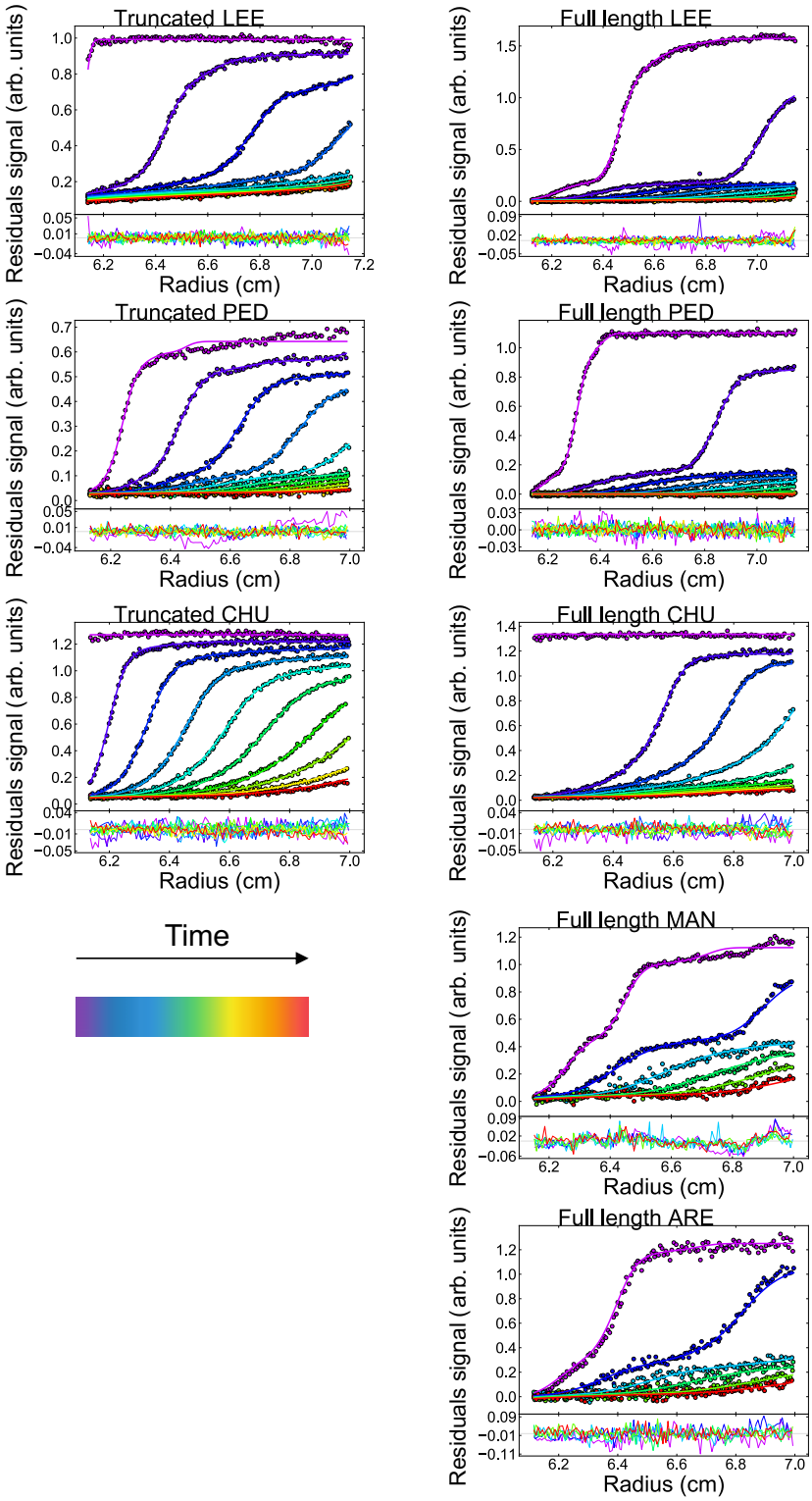


Supplementary Figure 11. N138 point mutant endpoint activity assay. See also Figure 4A. (A) Percent conversion to NTP for enzymes (10 µM) reacting with the indicated nucleotide (4 mM) in the presence of 10 mM Mn^{2+}. (B) Product distributions of mutants with 10 µM enzyme after 60 min incubation for the endpoint activity assay shown in panel A. (C) Product distributions of mutants with enzyme concentrations 0.4–6 µM after 5-10 min incubation. See Figure 4B for a plot of NTP synthesis efficiency. Plotted values are the average of 4 independent runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.



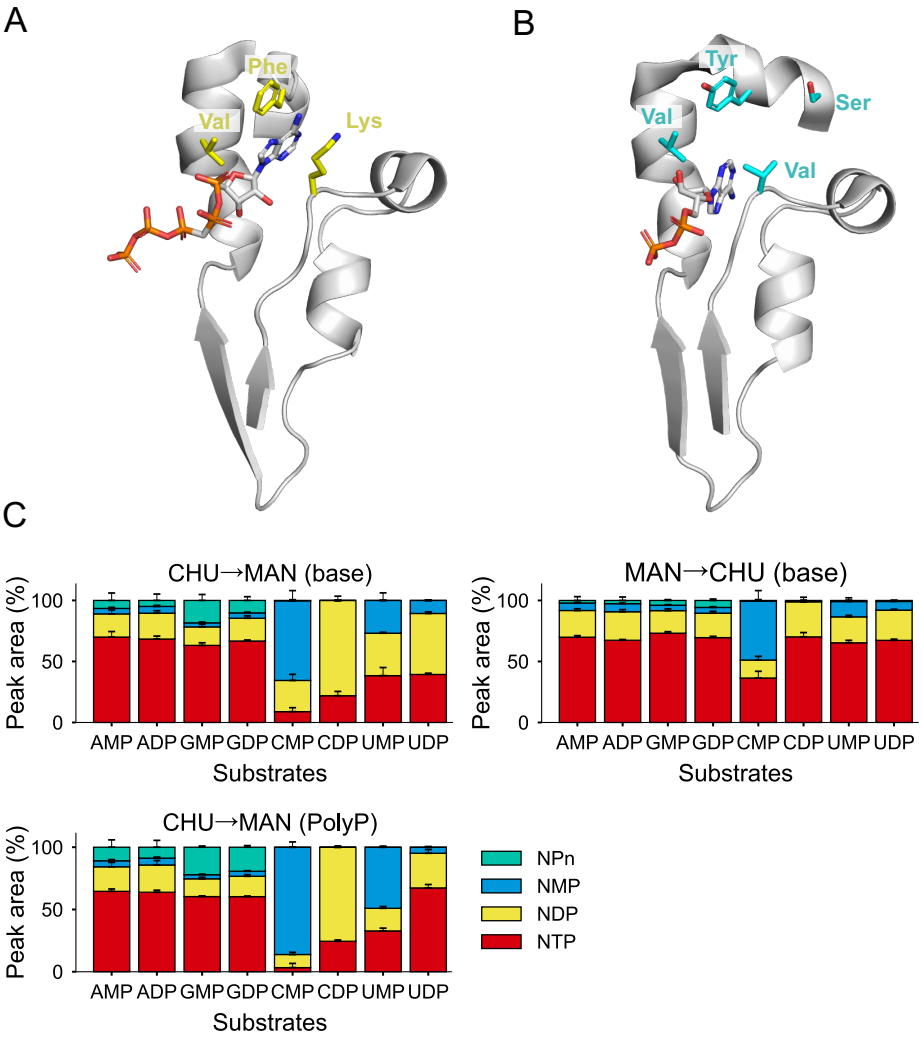
Supplementary Figure 12. Influence of C-terminal α-helix on the activity of Class III PPK2. (A) The C-terminal α-helix of one protomer rests one layer away from the active site P-loop of another protomer (left; PDB ID 6AQE). As such, the C-terminal α-helix may influence the P-loop conformation or dynamics through the lid domain (right). Thus, truncating the protein may affect nucleotide phosphorylation. Previously, it was shown that the phosphorylation of GMP and GDP was improved by the C-terminal α-helix truncation of CHU[1]. (B) Percent conversion to NTP for enzymes reacting with the indicated nucleotide (4 mM) in the presence of 10 mM Mn^{2+} using full-length (blue)

and C-terminal α-helix truncated (red) CHU, LEE, and PED. The reaction was performed with 18.9 μM enzyme, 20 mM PolyP-10, and 4 mM of indicated substrate in the standard reaction buffer. (C) The product distributions of reactions are analyzed in (B). Plotted values are the average of 4 independent runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.

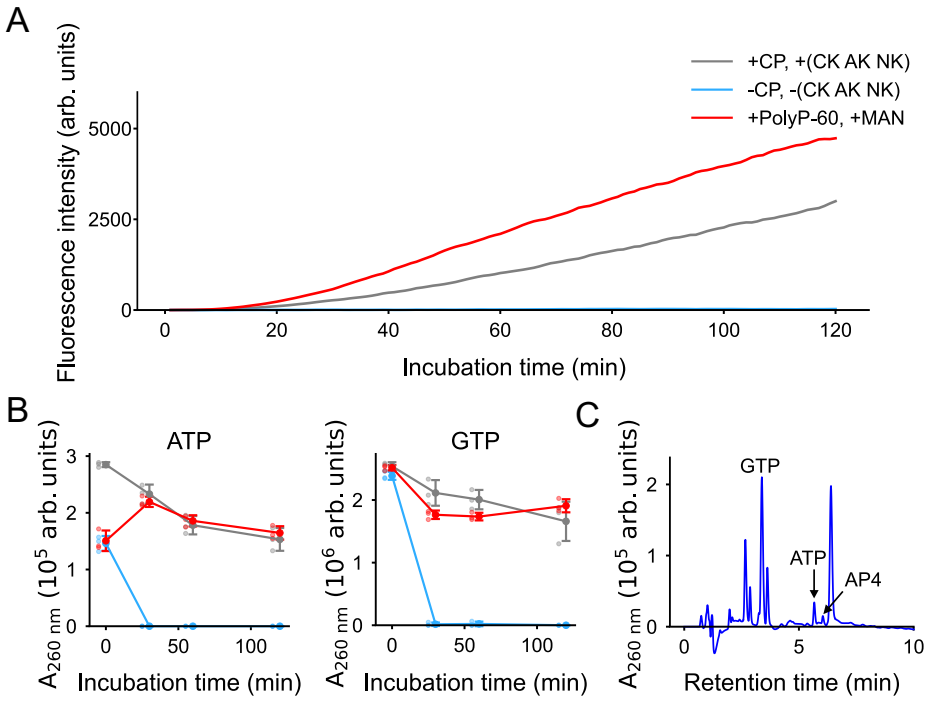


Supplementary Figure 13. Analytical ultracentrifugation (AUC) data for full-length LEE, PED, CHU, MAN, and ARE (right) and C-terminal α-helix truncated LEE, PED, and CHU (left). Data of truncated MAN and ARE were not obtained due degradation in *E. coli* and low purification yields. Raw data (scatterplot) and fitted curves (lines) of absorbance at 280 nm as a function of radius for each time

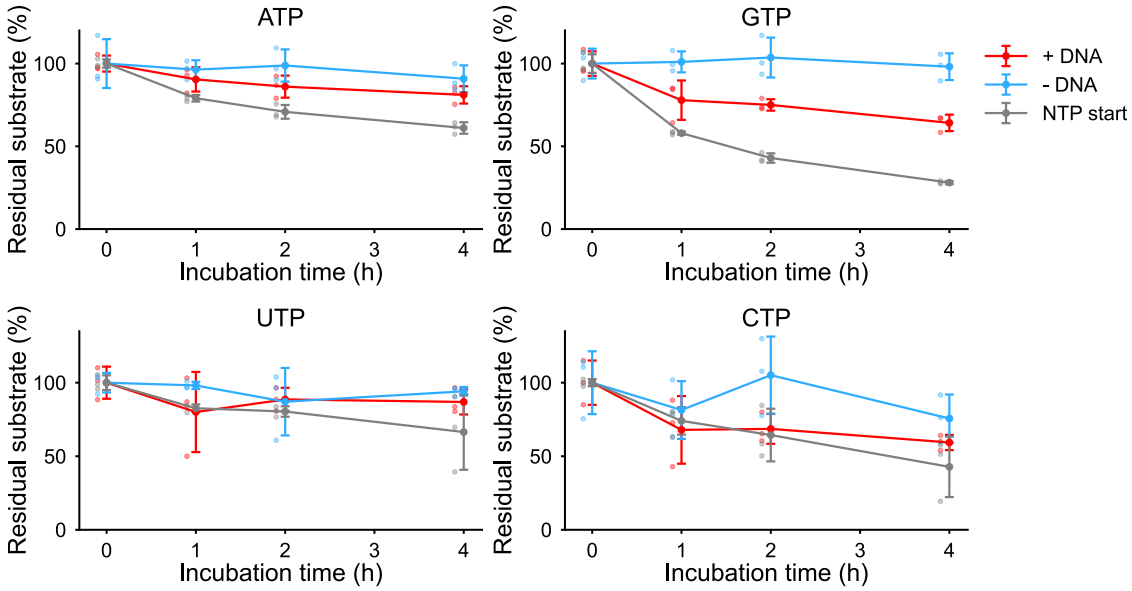
point are presented. Calculated molecular weights are given in Supplementary Table 2. Briefly, the truncation of the C-terminal α-helix altered the oligomerization state of the PPK2 enzymes: While the full-length enzymes were a dimer, truncation of the C-terminus converted the enzymes into a monomer. Source data for this figure is available in the Source Data file.



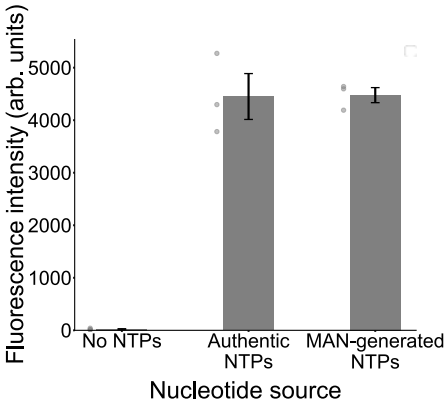
Supplementary Figure 14. NMP and NDP phosphorylation activity of MAN mutants. Nucleotide binding modes for the Class I PPK2 from *Francisella tularensis* (FRA) bound to adenosine-5′-[(β,γ)-methyleno]pentaphosphate [2] (PDB ID 5LLB, panel A) and for CHU bound to ADP (PDB ID 6AN9, panel B)[1]. MAN-CHU chimerized binding site positions (Figure 4) are shown in sticks and colored. Note that the serine of CHU has no equivalent position in FRA. The base binding position of FRA is flipped up relative to CHU, allowing for direct interaction with the chimerized binding site positions, which do not occur in the canonical binding mode. (C) Product distributions of nucleotide phosphorylation by MAN or CHU chimeras. See Figures 4C and 4E for NTP synthesis efficiency. Plotted values are the average of 4 independent runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.



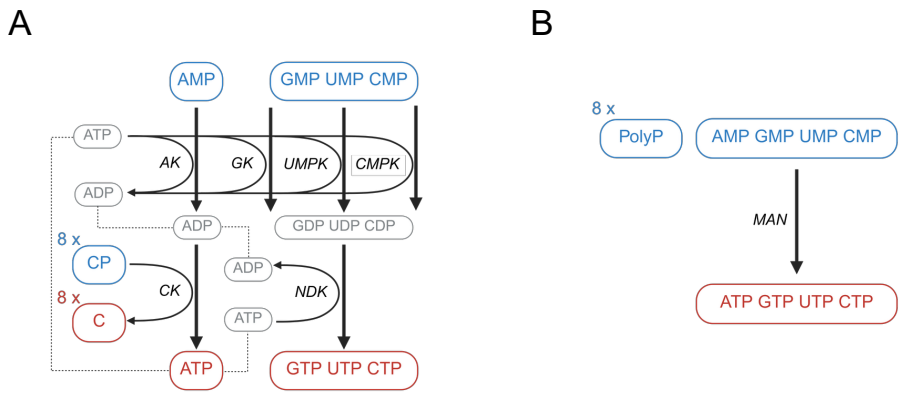
Supplementary Figure 15. Time-course analysis of MAN-mediated *in vitro* translation. (A) GFP *in vitro* translation (20 µL) was performed as described previously[3]. Using the PURExfrex 2.0 system at 37 °C, containing 0.1 mM ATP, 1 mM GTP, 1 µg GFP mRNA, and Alexa Fluor 647 for fluorescence monitoring, the reaction was performed as described below. (i) The standard system with ATP/GTP regeneration enzymes – adenylate kinase (AK), creatine kinase (CK), and nucleoside diphosphate kinase (NK) – together with 60 mM creatine phosphate (+CP, +(CK AK NK); gray curve), (ii) a system lacking these enzymes (-CP, -(CK AK NK); blue curve), and (iii) a system supplemented with 50 mM polyP-60 and 1 µM MBP-MAN instead of the conventional enzymes (+PolyP-60, +MAN; red curve). (B) Time-course of ATP and GTP peak areas measured by HPLC under the same conditions as in panel A. Reactions were quenched with trichloroacetic acid at 0, 30, 60, and 120 min. Solid lines and larger datapoints with error bars represent the average of 3 independent runs. Error bars are standard deviations. Smaller points to the left of the averaged data are the results from the individual runs. (C) HPLC chromatogram of the reaction mixture after 120 min incubation, showing the nucleotide content and composition. Note that the absence of a wavy tail at high retention times is indicative of little or no over-phosphorylation. The large peak after AP4 comes from the PURExfrex 2.0 reaction mixture. Source data for this figure is available in the Source Data file.



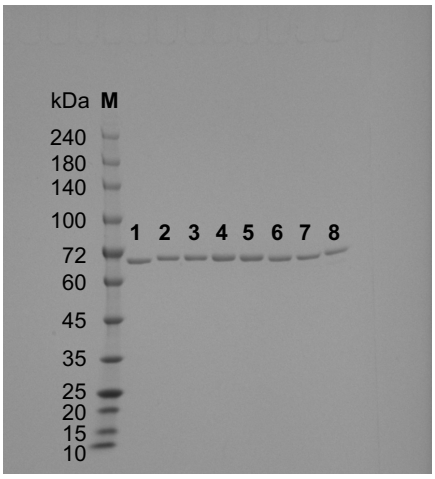
Supplementary Figure 16. Decrease of NTPs during *in vitro* transcription (IVT). Transcription of the Pepper RNA aptamer was performed using NTPs generated via the two-step synthesis with MAN and 30 mM polyP-60 (UMP and CMP for 60 min, AMP and GMP for 30 min). Reactions were quenched at 1, 2, and 4 h to measure residual NTPs. Red lines indicate reactions using synthesized NTPs, blue lines indicate reactions without DNA, and gray lines indicate transcription reactions with DNA using authentic NTPs. Solid lines and larger datapoints with error bars represent the average of 3 independent runs. Error bars are standard deviations. Smaller points to the left of the averaged data are the results from the individual runs. Source data for this figure is available in the Source Data file.



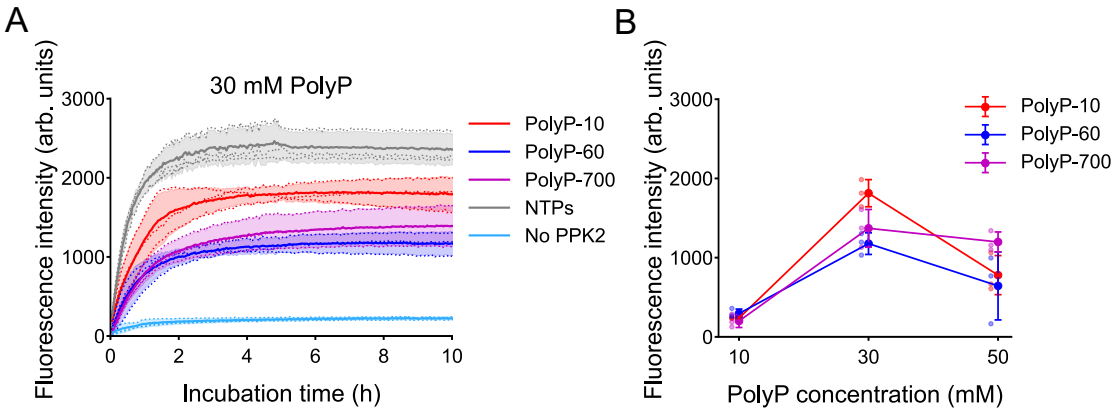
Supplementary Figure 17. *In vitro* translation using GFP mRNA synthesized from MAN-generated NTPs. GFP mRNA was transcribed from NTPs generated by MAN and PolyP-10, or from authentic NTPs, and compared to a control reaction without NTPs. Purified mRNA was used for *in vitro* translation with the PURExpress 1.0 (37 °C, 20 h). Plotted values are the average of 3 independent runs, with adjacent points indicating individual runs. Error bars are standard deviations. Source data for this figure is available in the Source Data file.



Supplementary Figure 18. NTP synthesis reactions. Substrates are shown in blue and products in red. (A) The cell-inspired reaction pathway requires six enzymes – adenylate kinase (AK), guanylate kinase (GK), CMP/dCMP kinase (CMPK), uridylate kinase (UMPK), creatine kinase (CK), and nucleoside-diphosphate kinase (NDK) – to convert various nucleoside monophosphates into the corresponding nucleoside triphosphates using creatine phosphate (CP) as a phosphoryl donor, with creatine (C) formed as a byproduct. (B) Reactions catalyzed by MAN, in which nucleoside monophosphates are phosphorylated by inorganic polyphosphate (PolyP) to yield nucleoside triphosphates.



Supplementary Figure 19. SDS-PAGE analysis of expressed Class III PPK2 constructs. Each construct carries an N-terminal 6×His tag followed by a maltose-binding protein (MBP) tag. Lane M: molecular weight marker; Lanes 1-8: Wild-type CHU, Wild-type MAN, Wild-type PED, Wild-type LEE, Wild-type ARE, Truncated CHU, Truncated PED, and Truncated Lee, respectively. 0.525 μg of total protein was loaded per lane onto a 5–20% gradient SDS-PAGE gel. Proteins were visualized by Coomassie Brilliant Blue staining. Protein migration was consistent with the theoretical molecular weights listed in Supplementary Table 2.



Supplementary Figure 20. *In vitro* transcription using NTPs produced by MAN with varying polyphosphate chain lengths and concentrations. (A) NTPs generated by MAN reactions using polyphosphates of different chain lengths (PolyP-10, PolyP-60, and PolyP-700; each at 30 mM phosphate units) were used as substrates for transcription of the Pepper RNA aptamer. Reactions were incubated at 37 °C for 10 h, and fluorescence was monitored with an Mx3005P real-time PCR thermocycler. (B) Fluorescence intensities after 8 h incubation for reactions containing polyphosphate of the indicated chain length at a concentration of 10, 30, or 50 mM phosphate units. Plotted values (solid lines) are the averages of 3 independent runs, with dashed lines in (A) and adjacent points in (B) indicating individual runs. Standard deviations are shown as shaded areas in (A) and as error bars in (B). Source data for this figure is available in the Source Data file.

Supplementary Table 1. Enzyme amino acid sequences.

Name	Uniprot ID	Protein sequence[1,2]
MAN	A0A2T5C445	MKSKQLNTDLFKVQAKQQIRLKNYDPASCNGFANRKEAEHQLKQDIKELA DLQYQLYAENKRSVLIVFQAMDAAGKDGSIRHVFSGLNPQGCKVYSFKSP SANELDHDYLWRHYNKLPPRGEIGIFNRSHYENVLITRVHPEFLLNENLP DVHATNDITTEFWERRYRQIRDFEQTLTENGTTIIKFFLHLSKDEQKKRF MERIRKPAKNWKFSSADIRERQYWDDYQHAYEQAMRATSTEEAPWYIIPA DNKWFTHVAIGNIIVETLQQMNIQMPEIS<u>SEEKDALKKAKKKLQNE</u>
PED	A0A519X5M4	MKTNTDKYLVAAKTKVKLKNFDTDYNGHLDKEAGKAELESVKEELSKFQE VLYASNAHSLLIIFQAMDAAGKDGAIAHVMSGLNPQGCQVYSFKTPSSEE YDHDFLWRHYKALPEKGRIGIHNRSHYENVLVCKVHPEYVLTENIPGYDD VKKIDDNFWKGRYESIRNFEKHLAANGVAILKFFLHVSKDEQKQRFLDRI EDPAKNWKFSAGDISERALWDDYMSAYEDALSETSTEDAPWYVIPADKKW YTRLAVSQIIAEKLKNLNLEFPKLPKEE<u>TDALEGYKKQLLSEK</u>
LEE	A0A1M5Z1E8	MKASDYQITAPIKLADLKTHEDFGKTKELEQELDEVSKKIAAIQNAMYAH GKYAALICIQGMDTAGKDSLIREVFKNFNARGVVVHSFKTPTSLELGHDY LWRHYIALPARGKVGVFNRTHYENVLVTRVHPEYILNENIPGVNDVSDID EAFWEKRFEQINNFEKHIADNGTIIFKFFLHLSKEEQKHRLLRRLNKPNK NWKFSAGDLKERALWDEYQQCYEDAINKTSKFHAPWYIIPADDKPTARLI LANILLEKLSSYKDIEEPELD<u>EKTKSNIDAYRKQLESE</u>
ARE	A0A1M6FUJ9	MNDYKVTSNISLKNTQTKVVIDDAEKQLKKLRKELGKLQDTLYAHGKYSV LVCLQGMDTAGKDSLIREVFKDFNARGVVVHSFKVPTELELKHDYLWRHY IALPARGKFGVFNRTHYENVLVTRVHPYYIMGENIPGVESVEDVNDAFWN KRFEQINNFEKHIAENGTIIFKFFLNLSKAEQKNRLLRRLRKQEKNWKFS PGDLKERKLWDKYQNCYEDAINRTSKPHAPWFTIPADDKPTARYLVAKIM LDTLKTYNNIKEPELD<u>DDIKANIEMYKQQLNNE</u>

[1]Underlined residues correspond to the C-terminal α-helix.

[2]N-terminal MBP tag not shown. Full constructs are available in the Source Data file..

Supplementary Table 2. Molecular weight and inferred oligomerization state of PPK2 enzymes by size exclusion chromatography (SEC) and analytical ultracentrifugation (AUC). Raw AUC data is presented in Supplementary Figure 13.

Enzyme	Molecular weight theoretical (kDa)	Molecular weight by SEC (kDa)[1]	MW_{SEC} / $MW_{theoretical}$	Molecular weight by AUC (kDa)	MW_{AUC} / $MW_{theoretical}$
Truncated LEE	74.5	242	3.2	88.8 ± 0.2	1.2
Truncated PED	75.1	251	3.3	89.0 ± 0.2	1.2
Truncated CHU	75.9	117	1.6	84.9 ± 0.5	1.1
Wild-type LEE	76.5	479	6.1	179.6 ± 0.2	2.3
Wild-type MAN	78.0	-	-	190.0 ± 0.5	2.4
Wild-type PED	76.8	498	6.5	179.7 ± 0.2	2.3
Wild-type CHU	78.3	501	6.4	159.3 ± 0.4	2.0
Wild-type ARE	76.3	-	-	180.7 ± 0.7	2.4

[1]Sizes are likely an over-estimation. See main text for a discussion of this point.

Supplementary Table 3. Price per 1 mmol of nucleotides and other relevant compounds available from Sigma-Aldrich.

Abbreviation	Product name	Cat. No.	Price ($/mmol)[1]
AMP	Adenosine 5'-monophosphate disodium salt	A2252	5
GMP	Guanosine 5'-monophosphate disodium salt hydrate	G8377	29
UMP	Uridine 5'-monophosphate disodium salt	U6375	20
CMP	Cytidine 5'-monophosphate disodium salt	C1006	24
ATP	Adenosine 5'-triphosphate disodium salt hydrate	A2383	44
GTP	Guanosine 5'-triphosphate sodium salt hydrate	G8877	393
UTP	Uridine 5'-triphosphate trisodium salt hydrate	U6625	378
CTP	Cytidine 5'-triphosphate disodium salt	C1506	279
PolyP	Sodium hexametaphosphate	71600	0.016
CP	Sodium creatine phosphate dibasic tetrahydrate	27920	37
PEP	Phosphoenolpyruvate	860077	55

[1]Cost calculations used 5 g of AMP, 250 g of PolyP, and 1 g for all other substrates. The cost per 1 mmol of PolyP was calculated based on the amount of one phosphate unit.

Supplementary Table 4. DNA sequence used for *in vitro* translation shown in Figure 5

Sequence of Pepper-RNA10 fusion construct[1,2]
5′GGGTCCCGGCGCCAGTGCCTGCCGAAGCAGGCCGACACGCCACGATTGGGGACCCtttTGGTC
ATGTGATCGGCGTATGCACACGGCACGCGAATAGCACAACACCCTGCGATCTCGCCGGGCAGTTA
AACAGCGAGTTCGGCCGACAAAGTATCGTGTCATCGTCGTCCCTATAGTGAGTCGTATTA

[1]T7 promoter sequence underlined.

[2]The sequence preceding 'ttt' corresponds to the Pepper sequence and the sequence following 'ttt' corresponds to the RNA10 and T7 RNA promoter sequences. RNA10 is reported to bind a phospholipid membrane[4]. The Pepper-RNA10 fusion sequence was used for IVT.

Supplementary References

1. Nocek, B. P. *et al.* Structural insights into substrate selectivity and activity of bacterial polyphosphate kinases. *ACS Catal.* **8**, 10746–10760 (2018).
2. Parnell, A. E. *et al.* Substrate recognition and mechanism revealed by ligand-bound polyphosphate kinase 2 structures. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3350–3355 (2018).
3. Wang, P.-H. *et al.* A Bifunctional Polyphosphate Kinase Driving the Regeneration of Nucleoside Triphosphate and Reconstituted Cell-Free Protein Synthesis. *ACS Synth. Biol.* **9**, 36–42 (2020).
4. Vlassov, A., Khvorova, A. & Yarus, M. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7706–7711 (2001).