

Substrate Recognition and Mechanism Revealed by Ligand-Bound Polyphosphate Kinase 2 Structures - Supporting Information Appendix

Alice E. Parnell^{a,1}, Silja Mordhorst^{b,1}, Florian Kemper^c, Mariacarmela Giurrandino^a, Josh P. Prince^a, Nikola Schwarzer^c, Alexandre Hofer^d, Daniel Wohlwend^c, Henning J. Jessen^{d,e}, Stefan Gerhardt^c, Oliver Einsle^c, Petra C. F. Oyston^{f,g}, Jennifer N. Andexer^{b,2}, Peter L. Roach^{a,f,2}

^aChemistry, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom.

^bInstitute of Pharmaceutical Sciences, Albert-Ludwigs-University Freiburg, Albertstr. 25, 79104 Freiburg, Germany.

^cInstitute of Biochemistry, Albert-Ludwigs-University Freiburg, Albertstr. 21, 79104 Freiburg, Germany.

^dOrganic Chemistry Institute, University of Zürich (UZH), Winterthurerstr. 190, 8057 Zürich Switzerland.

^eInstitute of Organic Chemistry, Albert-Ludwigs-University Freiburg, Albertstr. 21, 79104 Freiburg, Germany.

^fInstitute for Life Sciences, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom.

^gBiomedical Sciences, Dstl Porton Down, Salisbury, Wiltshire, SP4 0JQ, United Kingdom.

¹These authors contributed equally to this work.

²To whom correspondence should be addressed: J.N.A. (jennifer.andexer@pharmazie.uni-freiburg.de) or P.L.R. (plr2@soton.ac.uk)

SI Materials and Methods

Chemicals. Unless otherwise stated, chemicals and reagents were purchased in the highest purity available.

Synthetic procedures. β,γ -Methylene adenosine 5'-triphosphate (AMPPCP) was prepared by a literature procedure (1). Adenosine 5'-pentaphosphate (A5P) was synthesized by a modification of the method of Cremosnik (2). Adenosine 5'-tetrphosphate (with 1.6 equivalents of tetra-*N*-butyl ammonium as counterion) (50 mg, 51 μ mol) was co-evaporated three times with dry toluene (1 mL) and dissolved in dry DMF (1 mL) under an atmosphere of dry nitrogen. The mixture was cooled to 0 °C and (*i*Pr)₂NP(OFm)₂ (bis((9*H*-fluoren-9-yl)methyl) diisopropylphosphoramidite, 40 mg, 77 μ mol) and 5-ethylthio-1*H*-tetrazole (13 mg, 102 μ mol) were added. The mixture was stirred at 0 °C for 10 min, then the coupling product was oxidized with *m*CPBA (*meta*-chloroperoxybenzoic acid, 77%, 23 mg, 102 μ mol) at 0 °C for 5 min. Piperidine (50 μ L) was added and the mixture was stirred for 30 min. The product was precipitated by addition of diethylether. The suspension was centrifuged and the pellets washed with diethyl ether and dried under high vacuum yielding the crude product. Purification by strong anion exchange chromatography on Q-Sepharose Fast Flow using increasing concentration of an ammonium bicarbonate eluent (product eluted at 0.3-0.5 M NH₄HCO₃) afforded pure A5P (as the ammonium salt, 12 mg, 16 μ mol, 31%) as a colorless solid.

¹H NMR (400 MHz, D₂O) δ 8.57 (s, 1H), 8.30 (s, 1H), 6.18 (d, *J* = 6.1 Hz, 1H), 4.83 (s, 1H), 4.61 (s, 1H), 4.44 (s, 1H), 4.37 – 4.20 (m, 2H). ³¹P[¹H] NMR (162 MHz, D₂O) δ -8.5 (br, 1P), -10.2 (d, *J* = 17.2 Hz, 1P), -20.6 (br, 2P), -21.00 – -21.4 (m, 1P). ³¹P NMR (162 MHz, D₂O) δ -8.5 (br, 1P), -10.2 (d, *J* = 16.5 Hz, 1P), -20.6 (br, 2P), -21.0 – -21.4 (m, 1P). ¹³C[¹H] NMR (126 MHz, D₂O) δ 155.3, 152.4, 149.1, 140.1, 118.7, 86.8, 84.1 (d, *J* = 9.0 Hz), 74.3, 70.5, 65.4 (d, *J* = 5.2 Hz). HRMS (ESI) *m/z*: calc.: 332.45693, found: 332.45745 [M-H]²⁻.

Cloning, heterologous expression and protein purification. *Ft*PPK2 was produced and purified as described previously (3). *F. tularensis ppk2* site directed mutants were generated with a modified QuikChange protocol (4) using the primers listed in Table S3. After DpnI digestion, chemically competent XL1-Blue cells were transformed with the PCR mixture, transferred onto plates containing ampicillin (100 μ g/mL) and incubated overnight (37 °C). Selected colonies were then cultured overnight in 2YT media (10 mL) with ampicillin. Plasmid DNA isolated from these cultures was sequenced to confirm the introduction of the required mutations. For protein production, *E. coli* BL21

Rosetta plysS (DE3) cells were transformed with purified mutated plasmids and the variant proteins isolated as described for the wild type (WT) enzyme (3) (Fig. S14).

Inserts for *mrppk2* WT and Asn121Asp were obtained as synthetic codon-optimized DNA from Invitrogen. These were amplified by PCR using the primers *MrPPK2*-NdeI and *MrPPK2*-HindIII (Table S3), digested with NdeI and HindIII, and ligated (T4-Quick Ligase) into similarly digested expression vector pET28a(+) (Novagene). After transformation in chemically competent *E. coli* DH5 α cells, the plasmids (pET28a-*MrPPK*-wt and pET28a-*MrPPK2*-Asn121Asp) were isolated and verified by sequencing.

Proteins were heterologously produced in *E. coli* BL21(DE3). LB medium (5 mL) containing kanamycin (50 μ g/mL) was inoculated with a single colony of a fresh transformation and incubated at 37 $^{\circ}$ C overnight. This culture was used as a 1% inoculum in LB (500 mL) with kanamycin (50 μ g/mL) and cells were cultured at 37 $^{\circ}$ C and 180 rpm up to an optical density (600 nm) of 0.5. At this point, protein production was induced by addition of IPTG (final concentration 0.2 mM) and the culture was maintained at 20 $^{\circ}$ C and 180 rpm for a further 24 h. Cells were harvested by centrifugation and stored at -20 $^{\circ}$ C until use.

For *MrPPK2* purification (Fig. S1), cells were resuspended in lysis buffer (10 mL, 40 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10% glycerol) and sonicated (Branson-Sonifier 250 R, Duty Cycle 60%, Intensity 60%, 3 x 30 s, 30 s rest) on ice. The lysate was cleared by centrifugation (3-30K Sigma, rotor 12158-H, 14811 x *g*, 4 $^{\circ}$ C, 45 min). The supernatant was applied to a Ni²⁺-NTA column (4 mL, Sepharose beads, Qiagen). The column was washed stepwise with 5 mL lysis buffer containing 5, 10, 20 and 50 mM imidazole, the proteins were eluted with imidazole concentrations from 100-500 mM and concentrated (Sartorius Vivaspin 20, cut off 30 kDa). For desalting, the proteins were applied to a PD-10 desalting column (GE Healthcare) and eluted with lysis buffer. The proteins were frozen in liquid nitrogen and stored at -20 $^{\circ}$ C.

Enzyme activity assays. The activity of *FtPPK2* was measured in reaction mixtures (1 mL) that contained final concentrations of 50 mM Tris/HCl (pH 8.0), 25 mM NaCl, 20% glycerol, 10 mM MgCl₂, 80 mM NH₄SO₄, 1 mM ADP, 0.1 mM polyP (average length 25 units, Merck chemicals), 100 nM of enzyme for variant enzymes or 10 nM of WT enzyme) and were initiated by the addition of polyP. Reactions were incubated at 30 $^{\circ}$ C and, after 30, 45, 60 and 120 min (or 30, 60, 90 and 120 s for the WT enzyme) aliquots (100 μ L) were withdrawn and the reaction quenched by the addition of EDTA (to 50 mM), followed by heating (95 $^{\circ}$ C, 5 min). Negative control assays had no polyP included. Precipitated protein was removed by centrifugation (13,200 x *g*, 5 min) and the supernatant was analyzed by ion pairing HPLC (Table S5). The time course of product formation was fitted to a linear

function to give the initial rates of reaction using GraphPad Prism which were then used to calculate specific activity and k_{cat} values. This assay method was modified to test AMP and GMP as substrates for *FtPPK2*, but no turnover to form either di- or triphosphate nucleotides was observed.

The activity of *MrPPK2* was measured in reaction mixtures (500 μ L) that contained 50 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$, 2 mM nucleotide and 5 mM polyP (Acros organics, calculated as single phosphate residues). The reaction was maintained at 37 $^{\circ}$ C and initiated by the addition of the enzyme (final concentration 1 μ M) and stopped by removing the enzyme using spin-filters (Sartorius Vivaspin 500, cut off 10 kDa). The flow-through was analyzed by HPLC (Table S5). Kinetic parameters were determined for WT *MrPPK2* and *MrPPK2* Asn121Asp variant using 0 – 5 mM nucleotides and the reaction was initiated by addition of enzyme at the following concentrations: phosphorylation of ADP to ATP, 1 μ M *MrPPK2* (WT or Asn121Asp); phosphorylation of AMP to ADP, 10 nM *MrPPK2* (WT or Asn121Asp). At selected time points, aliquots (100 μ L) were withdrawn and the reaction was quenched by addition of acetonitrile (for the phosphorylation of ADP to ATP) or by addition of EDTA (for the phosphorylation of AMP to ADP). After centrifugation (Eppendorf Minispin, rotor F45-12-11, 12044 x g, 5 min), the supernatant was analyzed by HPLC (Table S5).

As an additional control, rates of reaction for both WT enzymes and the variants *FtPPK2* Asp117Asn and *MrPPK2* Asn121Asp were analyzed using the method initially developed for the other enzyme, yielding consistent results.

Isothermal titration calorimetry (ITC). All experiments used a MicroCal iTC₂₀₀ (MicroCal, Inc.) maintained at 20 $^{\circ}$ C and stirring at 750 rpm. Experiments were carried out in triplicate using ITC experimental buffer (50 mM HEPES, pH 8.0, 10 mM $MgCl_2$, 0.3 M NaCl, 20% glycerol, 0.15 mM β -mercaptoethanol). *FtPPK2* variants were exchanged into this buffer by dialysis and/or size exclusion chromatography. Titrations began with an initial injection of 0.4 μ L, followed by 19 identical injections of 2 μ L. Data were fitted to a bimolecular binding model using Microcal Origin software. As a control and to ensure consistency of results, a titration with WT enzyme was carried out alongside experiments with *FtPPK2* variant enzymes.

Crystallization and structure determination. The crystals of *FtPPK2* complexes were all grown by hanging drop vapor diffusion at 20 $^{\circ}$ C. The precipitant well solution for the complex of *FtPPK2* with polyP (PDB ID 5LL0) was developed from an original condition from the Morpheus screen (5) and consisted of 5% glycerol/PEG 4000, 0.1 M MES/imidazole pH 6.5, 0.2 M Morpheus alcohols and crystals appeared after 2 days. The *FtPPK2* solution (15 mg/mL) contained 5 mM $MgCl_2$ and 1 mM polyP. The precipitant well solution for *FtPPK2* complexed with polyP and AMPPCPPP (PDB ID 5LLB)

consisted of 20% PEG1500 and pH 8 PCTP (propionate, cacodylate, bis-tris propane) buffer. For crystallization experiments, the protein solution contained 10 mM MgCl₂, 0.5 mM polyP and 5 mM AMPPCH₂P and crystals appeared after 5 days. The precipitant well solution for *FtPPK2* D117N variant complexed with polyP (PDB ID 5LLF) contained 32.5% MPD/PEG1000/PEG3350, 0.1 M Bicine/Tris, pH 8.5, 0.09 M NPS (nitrate, phosphate, sulfate mix) and the protein solution contained 0.5 mM polyP and 2.5 mM AMPPCH₂P and crystals appeared after 5 days. During harvesting, crystals were dipped into a solution of mother liquor adjusted to 50% glycerol and flash frozen in liquid nitrogen. Data sets were collected under cryogenic conditions (-173 °C) at the Diamond Light Source (Didcot, UK) on beamlines i24 (5LLB), i04 (5LLF) and i04-1 (5LL0). Data were processed with xia2 (6) and the structures solved by molecular replacement with either BALBES (7, 8) or PHASER (9). Models were built with COOT (10) and refined with Phenix.refine (11); ligand files were generated with ELBOW (12).

MrPPK2 (WT and variant) was crystallized by the sitting drop vapor diffusion method. In the optimized crystallization conditions, the drops consisted of 0.3 µL protein solution with a concentration of 10 mg/mL and 0.3 µL of reservoir solution, which contained 100 mM Tris/HCl pH 8.5, 300 mM lithium sulfate and 27% PEG 3350. Crystals appeared within 24 h and were transferred into a cryo-solution which differed by containing an increased PEG content of 30%. For soaking, this solution additionally contained 10 mM of MgCl₂, 10 mM polyP and 10 mM of one of the nucleotides (AMP, ADP, ATP, or A5P). The crystals were incubated for up to one hour before being flash-frozen in liquid nitrogen. X-ray data was collected at the beamlines X06SA (PXI) and X06DA (PXIII) of the Swiss Light Source (Switzerland). The datasets were processed using XDS (13) and aimless (14, 15) from the ccp4-suite (16). The structure of the unliganded protein was determined by molecular replacement with MolRep (17) using a PPK2 protomer from *Arthrobacter aurescens* (3RHF) as initial search model. The resulting model was refined with Coot (10) and autoBuster (18). The datasets from the ligand-soaked and variant crystals were solved using molecular replacement with the unliganded protein structure and then refined as described above.

Analysis of oligomerization states by means of Right Angle Static Light Scattering (RALS).

Oligomerization states were analyzed by size exclusion chromatography (SEC) combined with right angle light scattering (RALS) using a Viscotek TDA 305 (Triple Detector Array) with a GPCmax sample delivery tool. Instrument constants were determined with defined concentrations of BSA. *MrPPK2* and *FtPPK2* alone or mixed with substrates were analyzed in aliquots of 100 µl each at a sample concentration of 4 mg/ml on a Superdex 200 10/300 GL column (GE Healthcare, Germany) at a column temperature of 28 °C and a flow rate of 0.5 mL/min. *MrPPK2* was analyzed in the presence of 40 mM Tris/HCl (pH 8.0), 100 mM NaCl, and 10% (v/v) glycerol, while *FtPPK2* was analyzed in 50 mM Tris/HCl

(pH 8.0), 500 mM NaCl, and 20% (v/v) glycerol. UV₂₈₀, refractive index (RI), and RALS were monitored. For determination of molecular weights the RI and RALS signals were used in the Rayleigh equation (1) as implemented in the Viscotek software OmniSEC:

$$(1) \frac{K \cdot c}{R_\theta} = \frac{1}{Mw \cdot P_\theta} + 2A_2 \cdot c$$

with c = concentration of the macromolecule

R_θ = Rayleigh quotient of the macromolecule at incident light angle θ

Mw = weight-average of molecular mass: $\overline{Mw} = \frac{\sum M_i c_i}{c_i}$

P_θ = particle scattering function R_θ/R₀ at angle θ

K = optical constant

A₂ = second virial coefficient

K in eq. (1) is defined as

$$(2) K = \frac{2\pi^2 n_0^2}{N_A \lambda_0^4} \left(\frac{dn}{dc} \right)^2$$

with n₀ = refractive index of solvent

n = refractive index of scatterer

N_A = Avogadro constant

λ₀ = wavelength of incident LASER light

P_θ in eq. (1) is defined as

$$(3) \frac{1}{P_\theta} = 1 + \frac{16\pi^2 n_0^2 R_g^2}{3\lambda_0^2} \sin^2 \left(\frac{\theta}{2} \right)$$

with R_g being the molecule's radius of gyration

Samples that appear to contain more than one component in the 280 nm UV trace were fitted to a sum of Gaussian and Lorentzian functions using the program MagicPlot (Magicplot Systems, LLC, Saint Petersburg, Russia). The plots are shown in Fig. S12 and results are reported in Table S4.

Accession codes. Protein sequences were obtained from the Uniprot Database under accession codes Q5NEQ5 (Q5NEQ5_FRATT) for *FtPPK2* and M9XB82 (M9XB82_MEIRD) for *MrPPK2* (Fig. 2). Coordinates have been deposited in the Protein Data Bank under accession codes 5LL0, 5LLB, and 5LLF for *FtPPK2* structures and 5LC9, 5LCD, 5LDB, 5LD1, 5MAQ, 5O6K, and 5O6M for *MrPPK2* structures (Table S1).

SI Figures

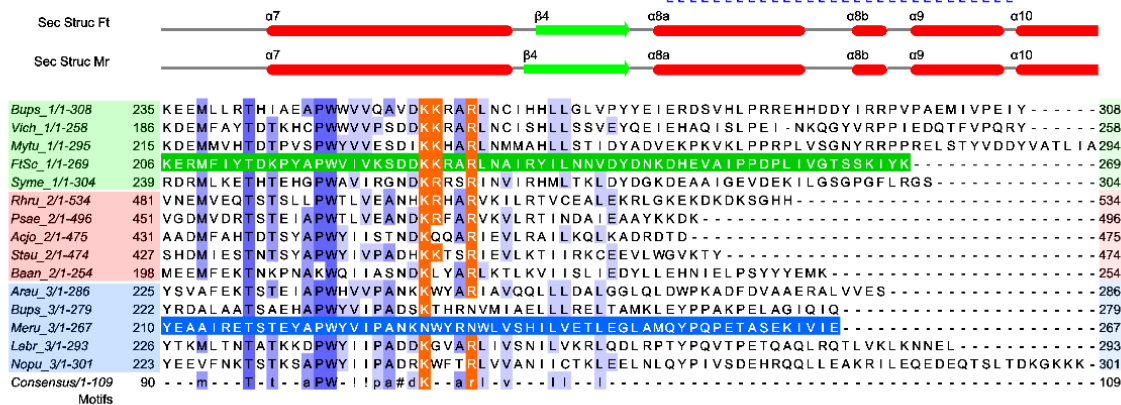
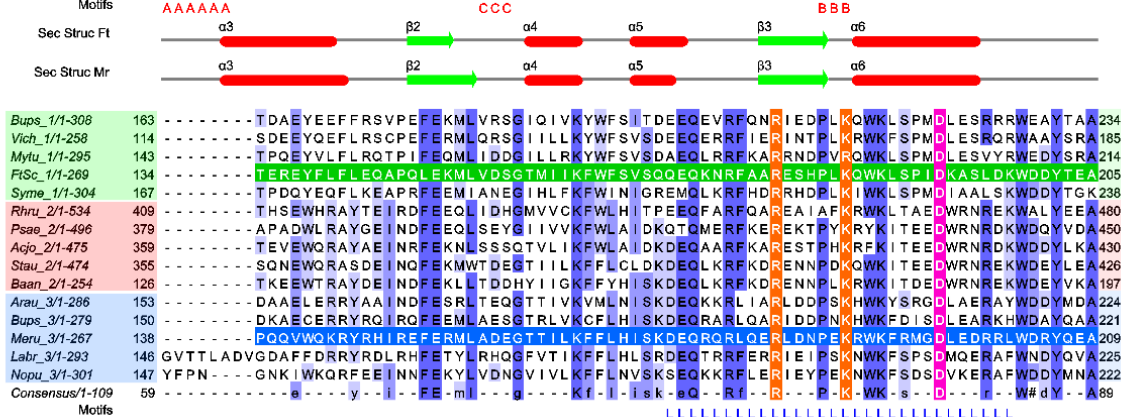
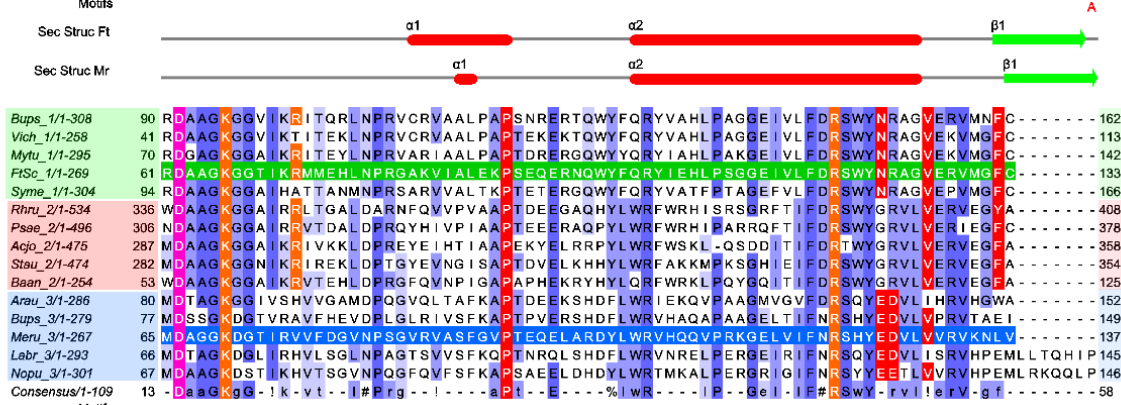
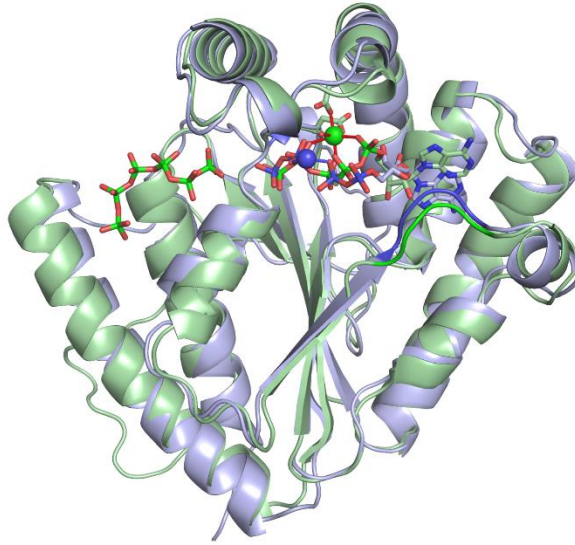
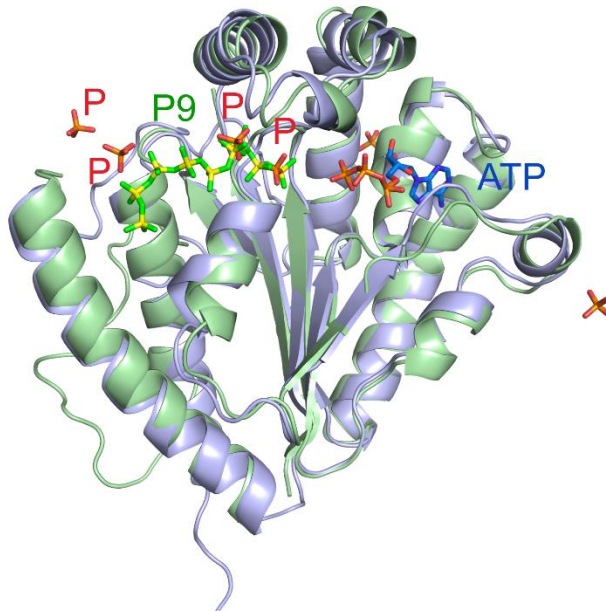


Fig. S1. Alignment of selected PPK2 sequences. Sequences were aligned with Multalin (19) and the figure prepared with Jalview (20). Individual residues are color-coded as follows: phosphate binding in orange, metal binding in pink, nucleotide-binding in red, other residues from *Ft* and *Mr* PPK2s are highlighted in green and blue, respectively. The secondary structure elements are labeled below the alignment, with α -helices in red and β -sheets in green. The Walker motifs, the lid loop and the base-binding loop are indicated by the letters A, B, L and C, respectively, below the alignment. For other sequence positions, increasingly conserved residues are colored deeper shades of blue. The sequence names and residue numbers of the PPK2 classes I, II and III are grouped on green, pink and blue backgrounds respectively. The sequences were selected from the UniProt or NCBI databases and are listed with organism name and accession codes: Class I: Bups_1, *Burkholderia pseudomallei*, Q63KT2; Vich_1, *Vibrio cholera*, C3LSY1; Mytu_1, *Mycobacterium tuberculosis*, O05877; FtSc_1, *Francisella tularensis*, Q5NEQ5; Syme_1, *Sinorhizobium meliloti*, Q92SA6. Class II: Rhru_2, *Rhodospirillum rubrum*, Q2RXW7; Psae_2, *Pseudomonas aeruginosa*, Q9HYF1; Acjo_2, *Acinetobacter johnsonii*, Q83XD3; Stau_2, *Staphylococcus epidermidis*, A0A0H2VII9; Baan_2, *Bacillus anthracis*, NP_844244.1. Class III: Arau_3, *Arthrobacter aurescens*, YP_948522.1, Bups_3, *Burkholderia pseudomallei*, Q63YJ7; Meru_3, *Meiothermus ruber*, M9XB82; Labr_3, *Lactobacillus brevis*, YP_795494.1; Nopu_3, *Nostoc punctiforme*, B2J4S4.

A



B



C

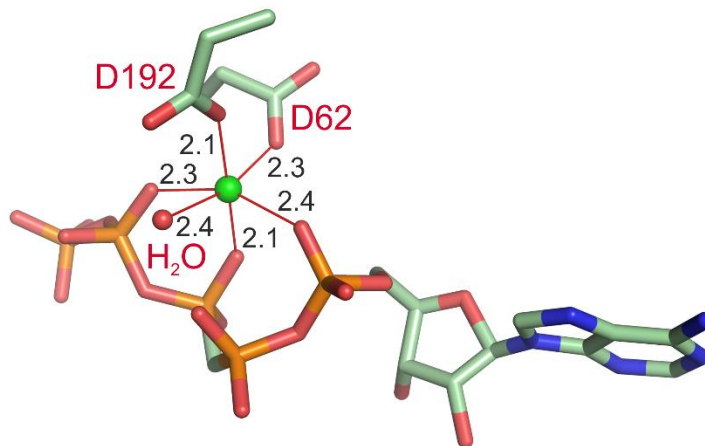


Fig. S2. Comparing *Ft* and *Mr* PPK2 structures. Mg^{2+} (large spheres) and ligand phosphorus atoms colored by structure, blue (*Mr*PPK2) and green (*Ft*PPK2). (A) Overlay of *Ft*PPK2:AMPPCPPP:PolyP₆ (green) and *Mr*PPK2:ADP:PP_i (blue) complexes. The root mean square deviation (RMSD) over 1304 atoms was 1.619 Å. (B) Overlay of *Ft*PPK2:PolyP₉ complex (polyP colored green/yellow) and *Mr*PPK2 Asn121Asp variant complexed with ATP (blue) and phosphate ions (orange/red). The four phosphate ions positioned in close proximity to the polyP binding channel are labeled with a P. (C) Magnesium ion (green) coordination and distances in the *Ft*PPK2:AMPPCPPP:PolyP₆ complex, protomer C. Distances are shown in Å.

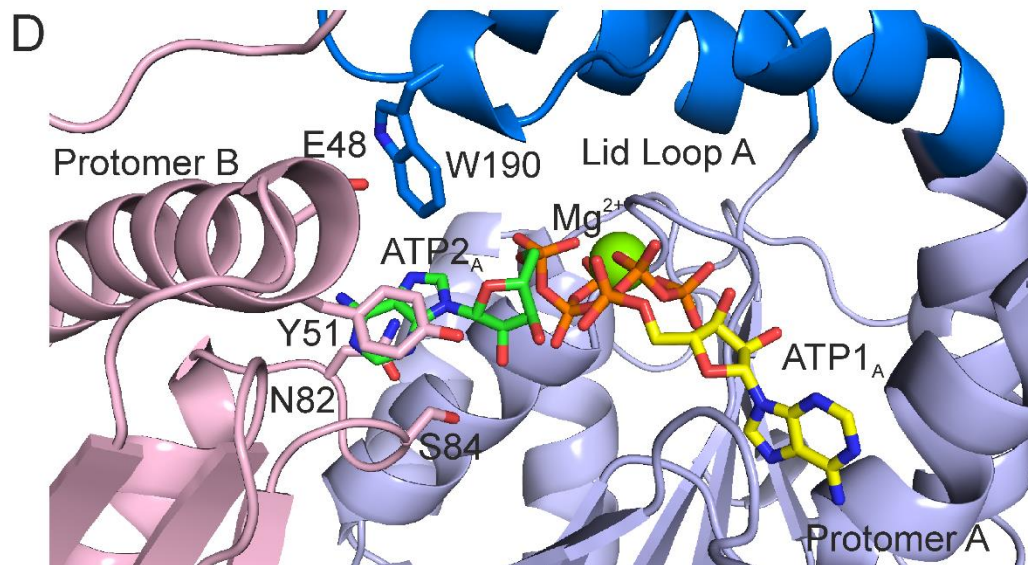
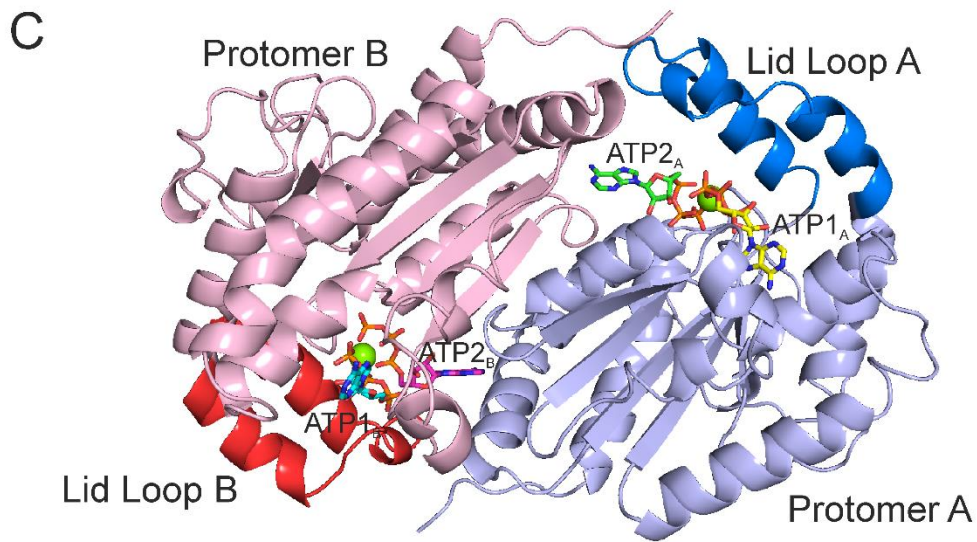
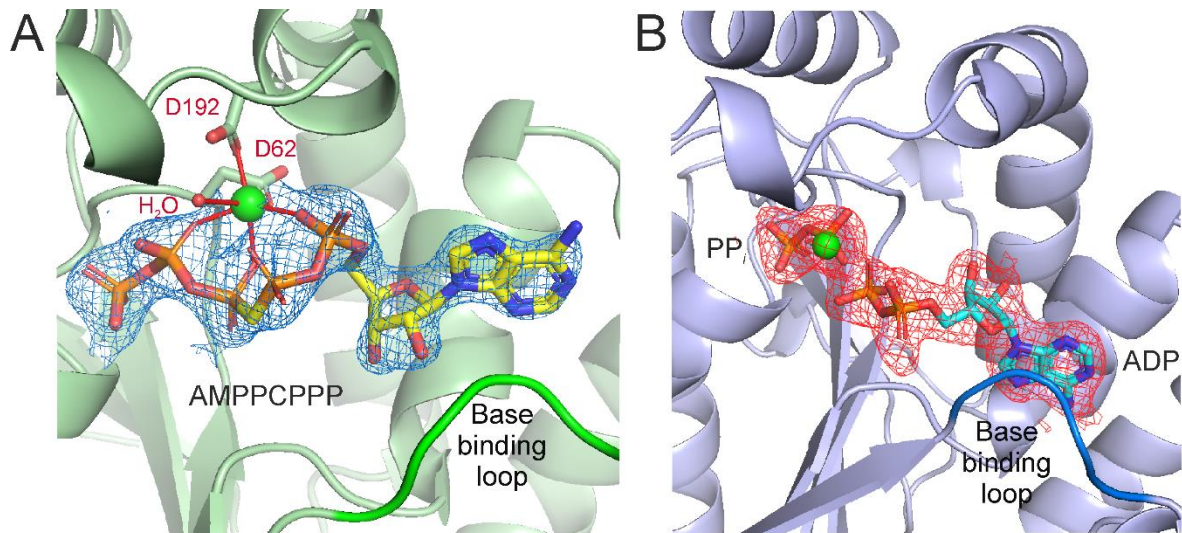


Fig. S3. Nucleotide binding in PPK2 structures. (A) Simulated annealing composite omit map ($2F_o-F_c$) around the AMPPCPPP nucleotide in complex with *Ft*PPK2, contoured at 1.5σ . (B) Simulated annealing composite omit map ($2F_o-F_c$) around the ADP and pyrophosphate (PP_i) in complex with *Mr*PPK2, contoured at 1.0σ . (C) Structure of *Mr*PPK2 with two ATP molecules bound. Two protomers A and B are shown, each with two bound ATP molecules labeled ATP1 and ATP2. The two ATP molecules associated with each protomer are labeled ATP1 and ATP2 followed by a subscript letter corresponding to the protomer (A or B). (D) Detail of ATP interactions with *Mr*PPK2. ATP1 (yellow) is bound in active site of protomer A (blue) similarly to other nucleotides (e.g., AMP and ADP); ATP2 (green) binds in a different conformation that is not observed in other complexes, with the phosphates bound to the Mg^{2+} (green sphere), but the nucleoside moiety protruding out of the active site and into the space between protomer A and protomer B (pink).

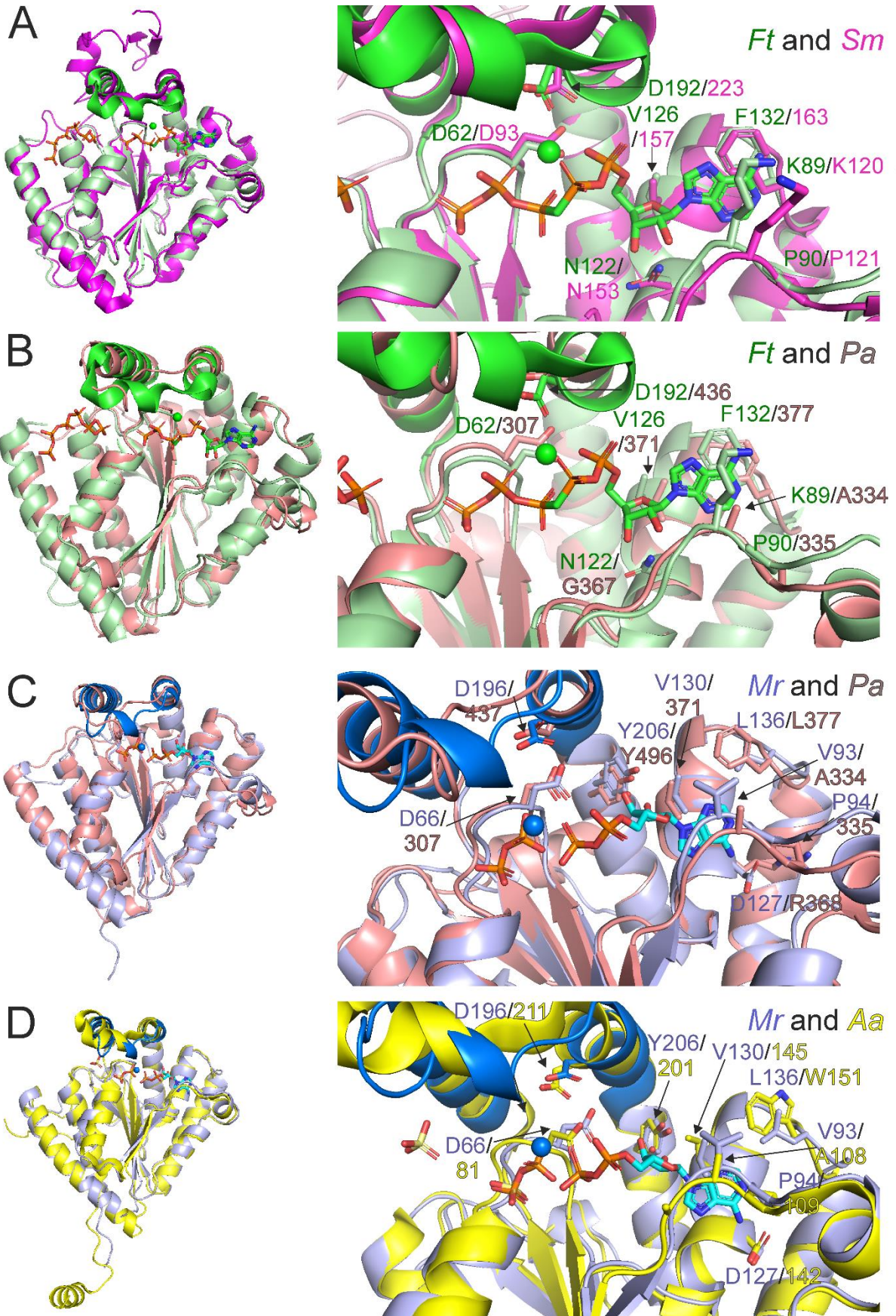
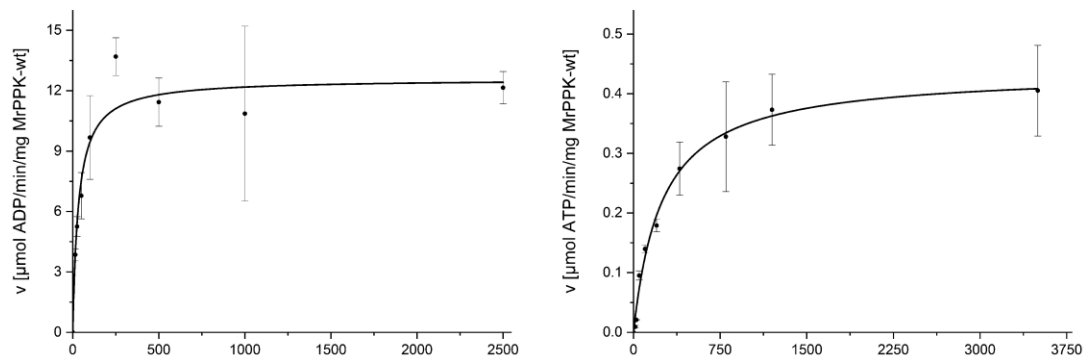


Fig. S4. Structural comparison of *Ft* and *Mr* ligand bound structures (PDBs: 5LLB and 5MAQ respectively) with unliganded PPK2s. Structures are colored as follows: *Ft*, pale green; *Mr*, pale blue; *S. meliloti* (*Sm*), magenta; *P. aeruginosa* (*Pa*), pink; *A. aurescens* (*Aa*), yellow. The lid loops and magnesium ions of *Ft* and *Mr* are bright green and bright blue, respectively. (A) Overlay of class I enzymes from *Ft* and *S. meliloti* (PDB 3CZQ). (B) Overlay *Ft* (class I) and *P. aeruginosa*, (class II, PDB 3CZP). (C) Overlay of *Mr* (class III) and *P. aeruginosa*, (class II, PDB 3CZP). (D) Overlay of *Mr* (class III) and *A. aurescens*, (class III, PDB 3RHF.).

A. *MrPPK2* WT



B. *MrPPK2* Asn121Asp

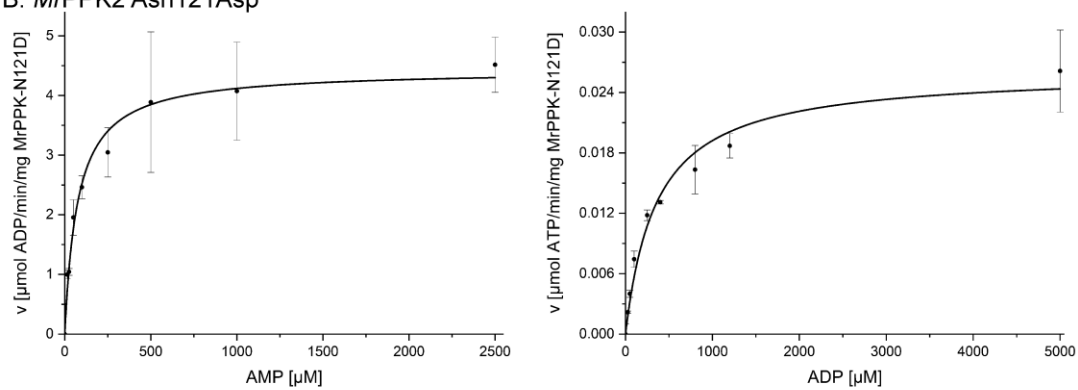


Fig. S5. Kinetic analysis of *MrPPK2* wild type and sequence variants.

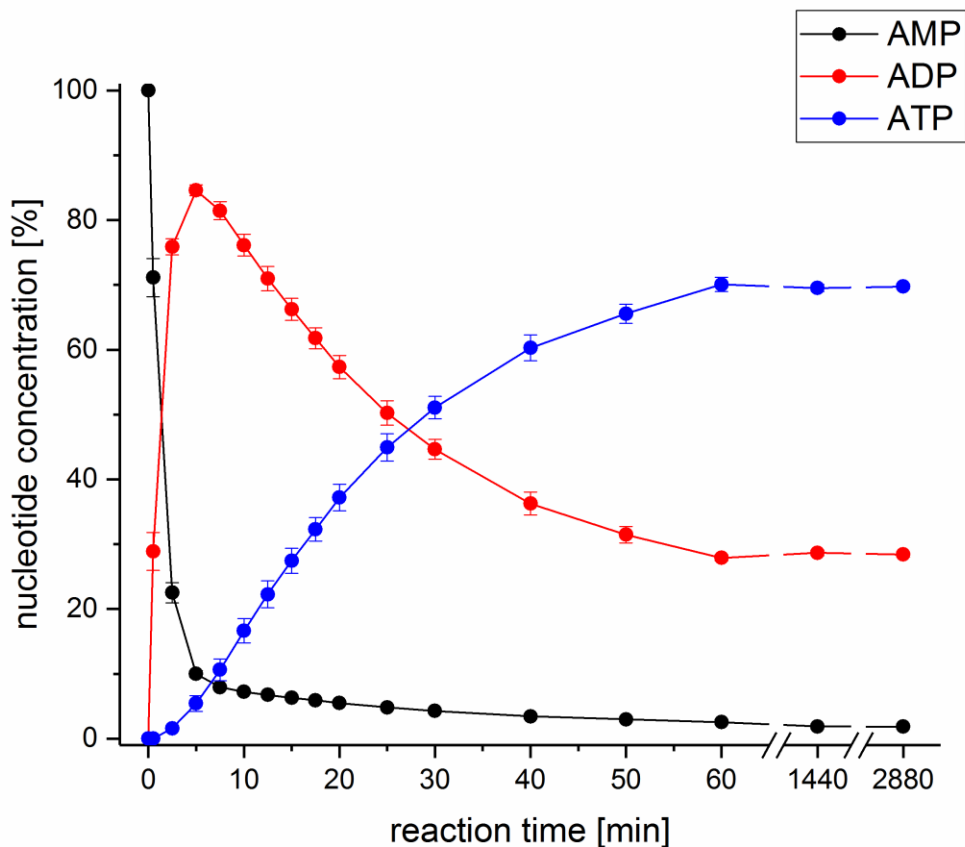
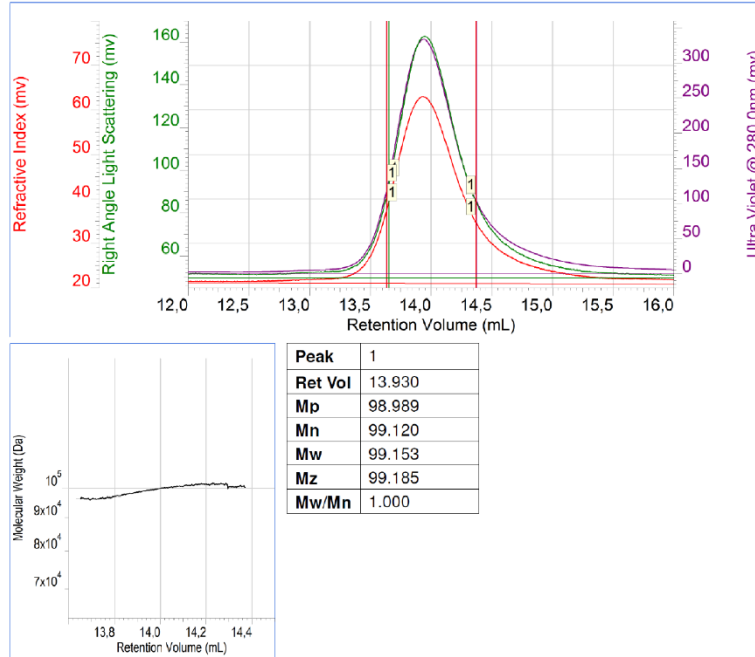


Fig. S6. Time course of AMP conversion into ADP and ATP with *MrPPK2*. 1 μ M *MrPPK2* was incubated with 1 mM AMP and 5 mM PolyP (calculated as single phosphate residues) at 37 $^{\circ}$ C as described in the Experimental Methods. At selected time points, aliquots were withdrawn and the reaction was quenched by addition of EDTA, centrifuged, and the supernatant was analyzed by HPLC. Nucleotide concentrations were determined based on HPLC calibration curves for the corresponding substances (24). Since no internal standard was added, the initial concentration of AMP (1 mM) was used as a relative reference; the sum of all nucleotide concentrations amounts to 1 mM (= 100%) at each selected time point. All reactions were performed and analyzed in triplicates.

A. MrPPK2



B. MrPPK2 plus polyP

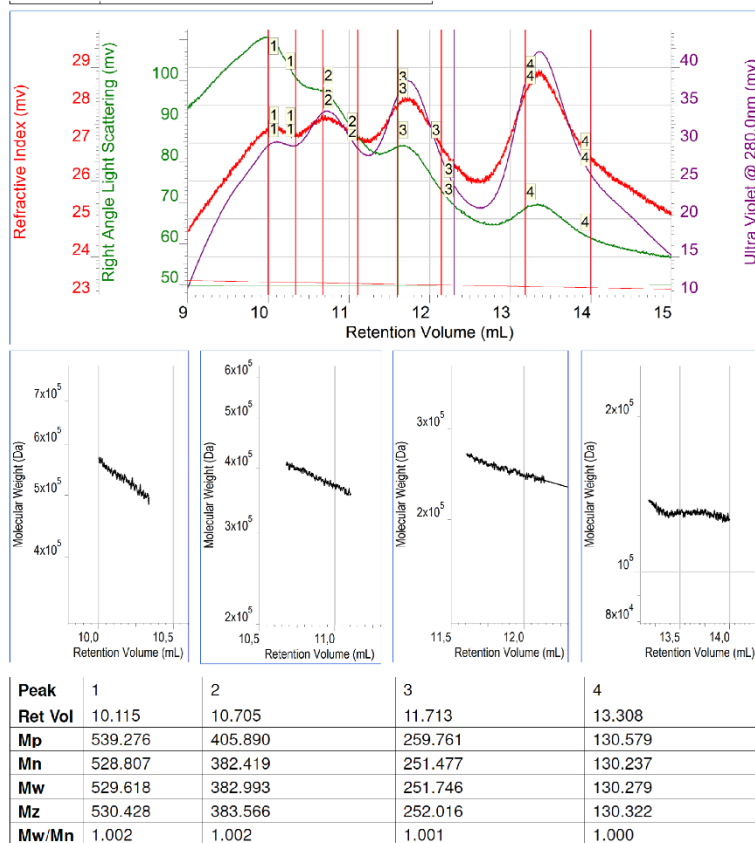
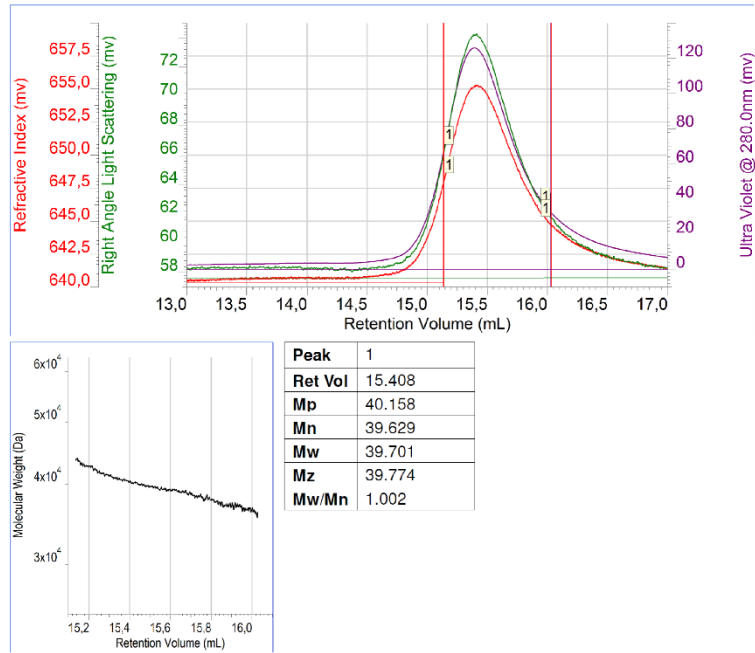


Fig. S7. Size exclusion chromatography (SEC) and right angle light scattering (RALS) analysis of *MrPPK2*. For each panel the following data is presented: top, run parameters; middle: chromatogram including Refractive Index (RI) and RALS plots with integral limits and according baselines for integration; bottom: plot of the molecular weight distribution across the peak and the according experimentally determined parameters. M_p is the molecular weight at the peak maximum, M_n is the number average, M_w the weight average, and M_z the z-average of the molecular mass. M_w/M_n is the polydispersity of the peak. Panels: (A), *MrPPK2*, (B), *MrPPK2* pre-incubated with 10 mM polyP.

A. FtPPK2



B. FtPPK2 plus polyP

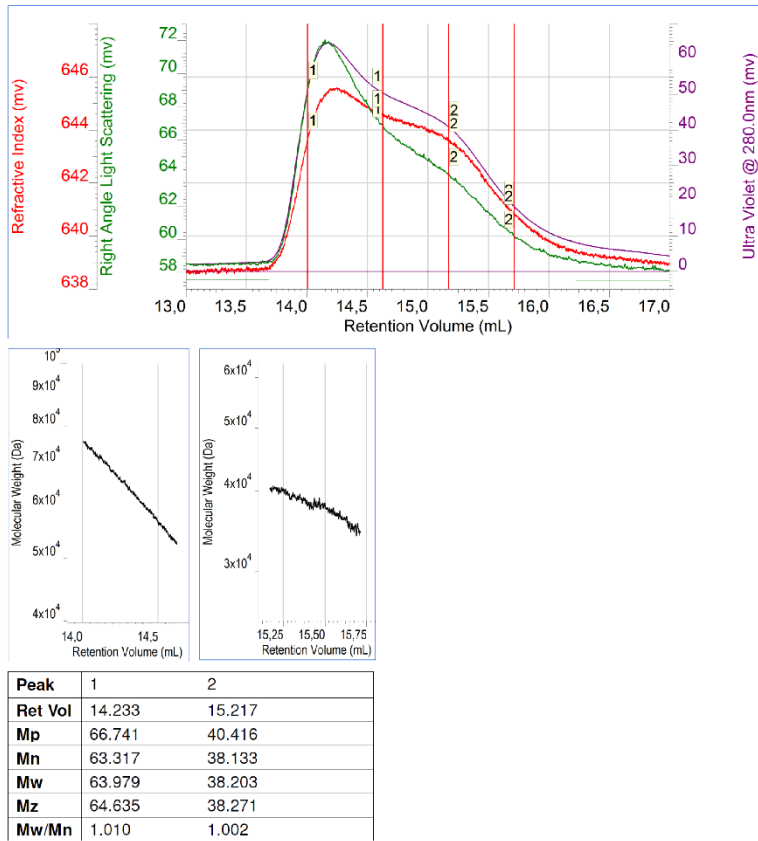
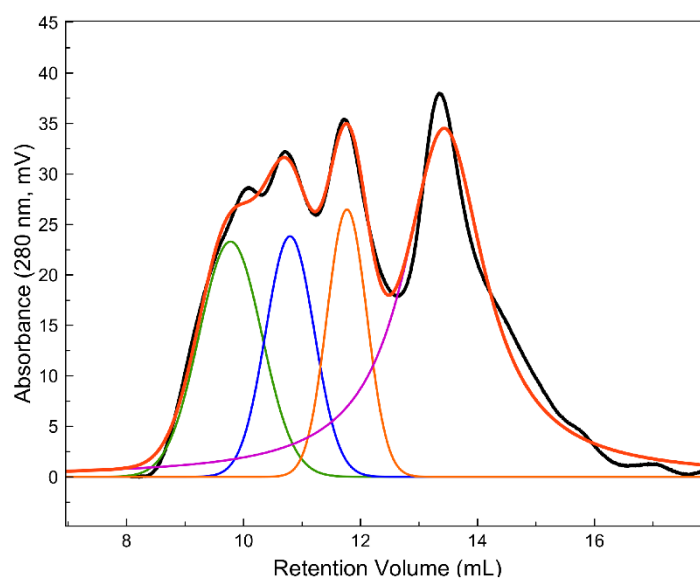


Fig. S8. Size exclusion chromatography (SEC) and right angle light scattering (RALS) analysis of *FtPPK2*. For each panel the following data is presented: top, run parameters; middle: chromatogram including Refractive Index (RI) and RALS plots with integral limits and according baselines for integration; bottom: plot of the molecular weight distribution across the peak and the according experimentally determined parameters. M_p is the molecular weight at the peak maximum, M_n is the number average, M_w the weight average, and M_z the z-average of the molecular mass. M_w/M_n is the polydispersity of the peak. Panels: (A), *FtPPK2*, (B), *FtPPK2* pre-incubated with 10 mM polyP.

A. *Mr*PPK2 WT + PolyP



B. *Ft*PPK2 WT + PolyP

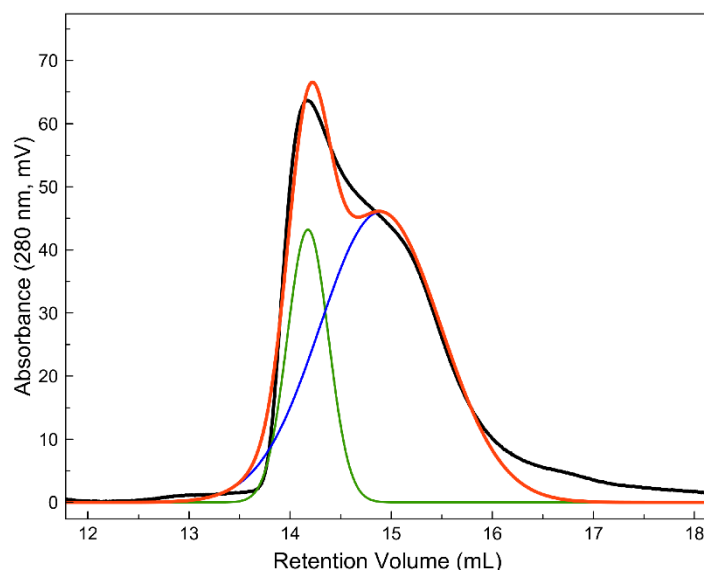


Fig. S9. Analysis of size exclusion chromatography. The experimental UV absorption traces are shown in black and were fitted to a sum function (made up of up to 4 Gaussian or Lorentzian functions) is shown in red. Curves for the individual functions are shown in green, blue, orange and pink respectively.

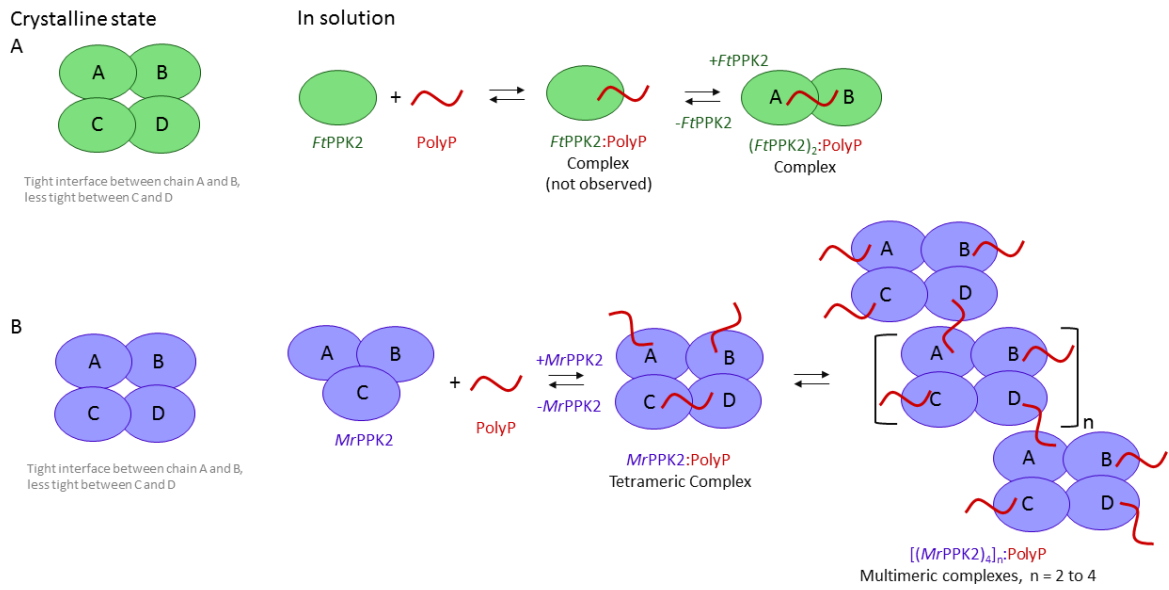
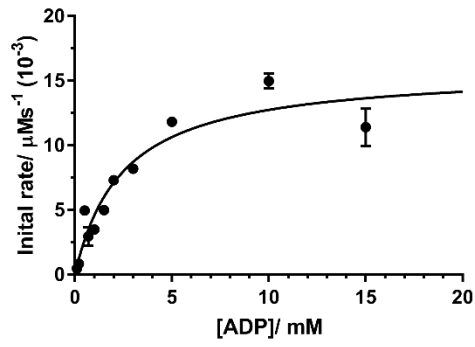
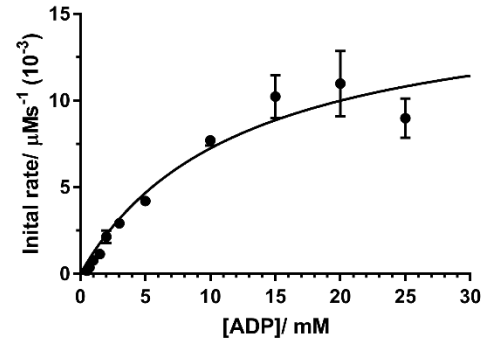


Fig. S10. Model for the oligomerization of PPK2s in the crystalline state and in solution. This includes the effects of polyP (curved red line) for the *Ft* (A) and *Mr* (B) enzymes.

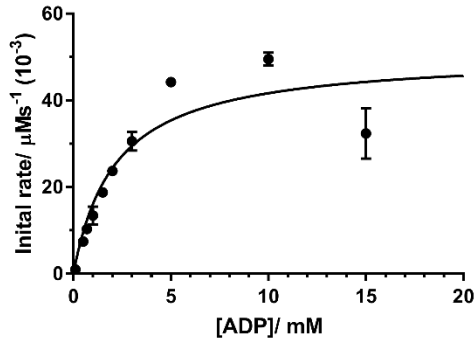
A. *FtPPK2* Asp62Ala



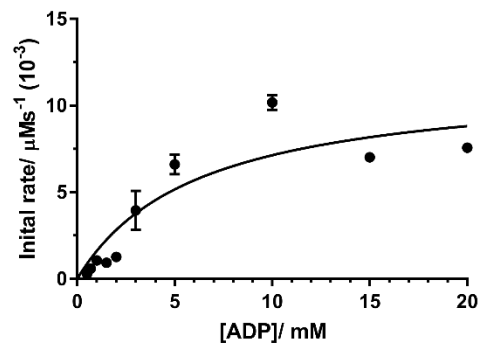
B. *FtPPK2* Lys66Ala



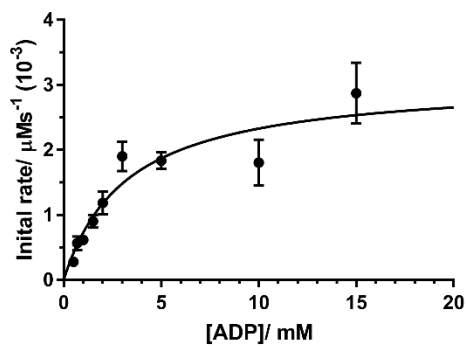
C. *FtPPK2* Asp117Asn



D. *FtPPK2* Arg118Ala



E. *FtPPK2* Arg178Ala



F. *FtPPK2* Asp192Ala

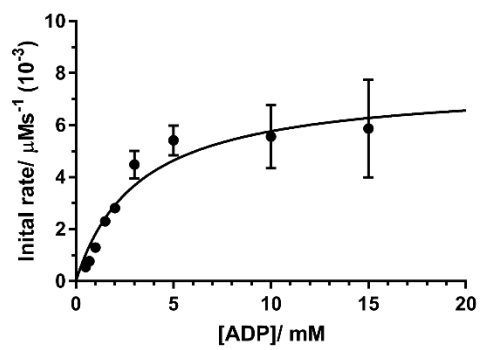


Fig. S11. Kinetic analysis of *FtPPK2* wild type and sequence variants.

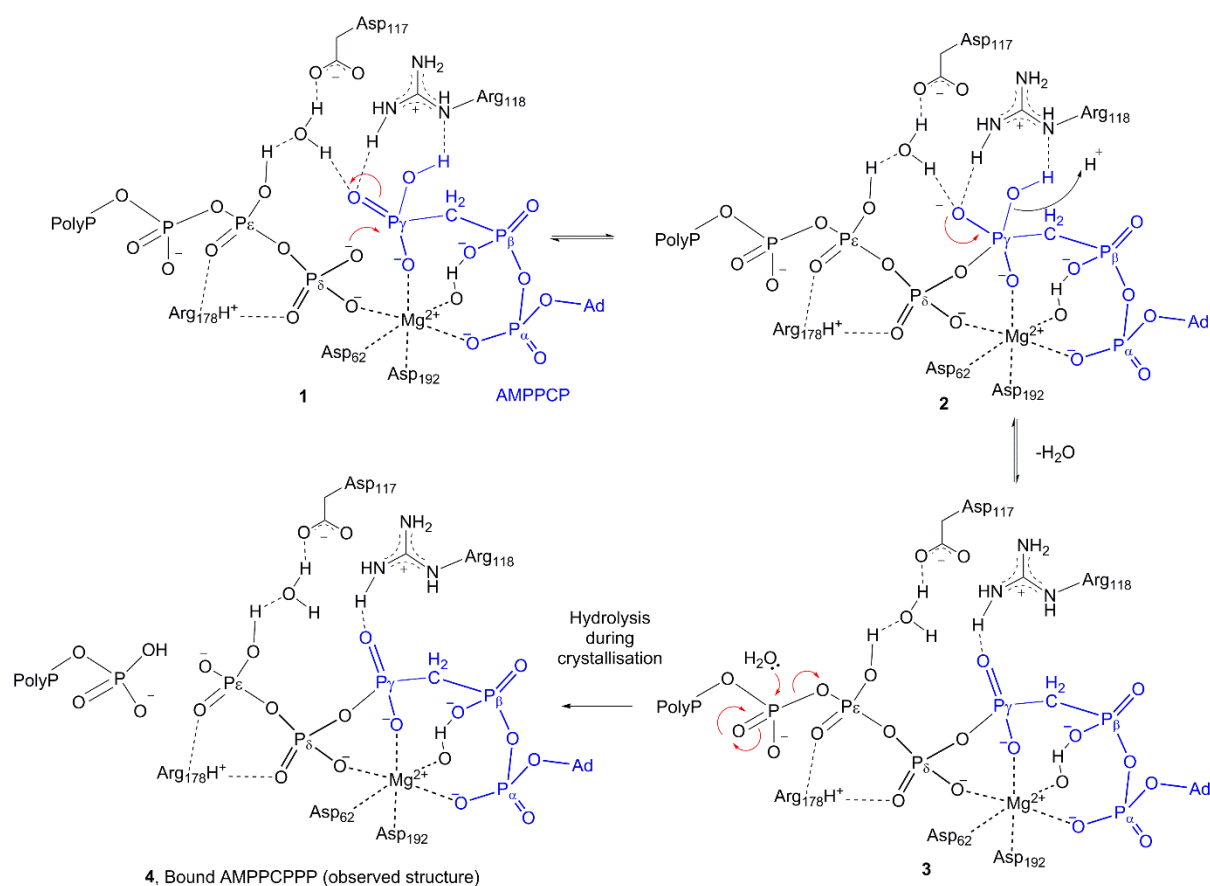


Fig. S12. Mechanistic hypothesis for reaction with AMPPCP and polyP. After binding at the enzyme active site **1**, nucleophilic attack by polyP leads to a trigonal bipyramidal intermediate **2**. The CH₂ containing phosphonate analogue is a poor leaving group (unlike the natural leaving group, ADP), but loss of water leads to compound **3**, in which the AMPPCP has become covalently linked to the polyP. Subsequent hydrolysis of the polyP chain (which may occur during the crystallization and may not directly relate to physiological catalysis) leads to the bound AMPPCPPP **4**. This figure summarizes one of several putative mechanisms that might account for the observed structure. Other alternatives include transfer of pyrophosphate by a direct substitution mechanism, or sequential transfer of two individual phosphate groups.

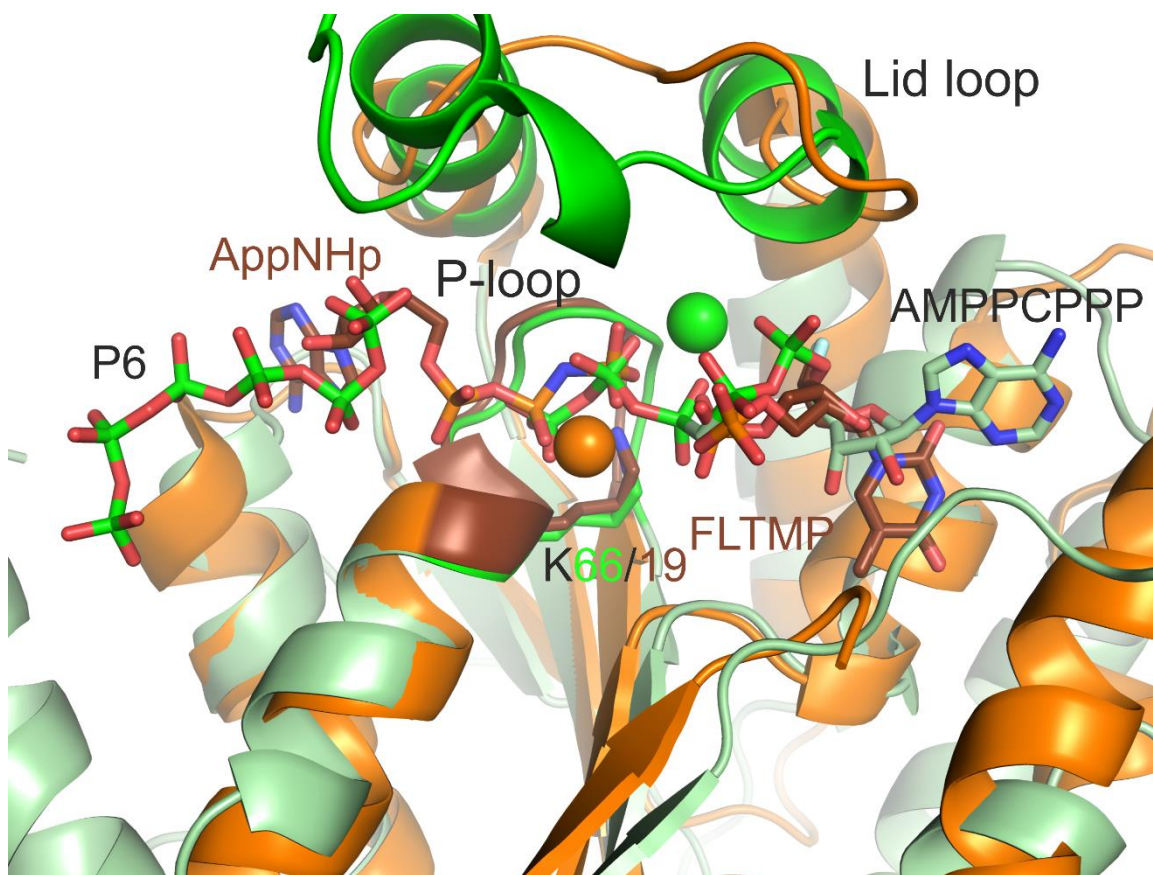


Fig. 13. Comparing *Ft*PPK2 and thymidylate kinase (TK). *Ft*PPK2 (green) complexed with polyP (P6) and AMPPCPRP (phosphorus atoms, green) is overlaid with human TK complexed with an ATP analogue, AppNHp and a thymidine analogue, 3-fluoro-3-deoxythymidine monophosphate (FLTMP) (phosphorus atoms, orange, PDB ID: 1NMY). Spheres indicate the positions of the magnesium ions. The highly similar P-loop structures, with Lys66/19 are shown in green and brown for *Ft*PPK2 and human TK respectively.

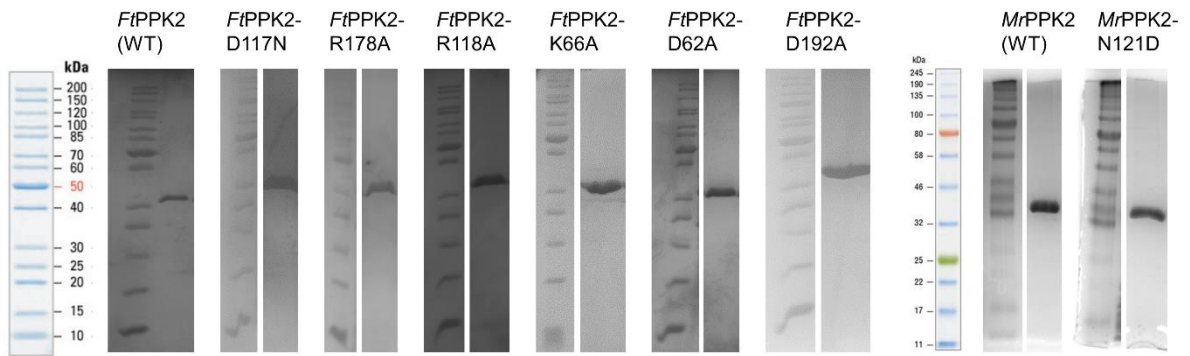


Fig. S14. SDS-PAGE analysis of PPK2s and PPK2 variants.

SI Tables

Table S1. X-ray data collection and refinement statistics. For clarity, the table is presented in two parts.

Protein^a	<i>Mr</i>	<i>Mr</i>	<i>Mr</i>	<i>Mr</i>	<i>Mr</i>
Ligands		AMP	ADP	ATP	ADP PP _i
PDB ID	5LC9	5LCD	5LDB	5LD1	5MAQ
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	164.78, 164.78, 95.0	164.4, 164.4, 95.0	165.6, 165.6, 95.2	168.0, 168.0, 95.0	166.2, 166.2, 95.0
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	45.7 - 1.90 (1.95 - 1.90)	58.14 - 2.66 (2.73 - 2.66)	45.93 - 2.30 (2.36 - 2.30)	48.25 - 2.09 (2.21 - 2.09)	117.5 - 2.45 (2.59 - 2.45)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.244 (2.809)	0.322 (2.045)	0.475 (4.413)	0.260 (1.586)	0.367 (1.509)
<i>R</i> _{meas}	0.252 (2.997)	0.329 (2.160)	0.487 (4.975)	0.274 (1.675)	0.396 (1.623)
<i>I</i> / σ <i>I</i>	19.7 (2.1)	17.3 (2.2)	10.3 (1.1)	11.7 (2.4)	8.6 (2.2)
CC1/2	0.999 (0.893)	0.998 (0.819)	0.997 (0.439)	0.998 (0.746)	0.992 (0.332)
CC*	1.000 (0.942)	0.999 (0.914)	0.999 (0.777)	1.000 (0.917)	0.998 (0.917)
Completeness (%)	100 (100)	99.0 (99.0)	100 (100)	100 (100)	100 (100)
Redundancy	52.5 (42.4)	44.0 (29.2)	27.0 (28.2)	18.4 (19.2)	13.4 (14.0)
Refinement					
No. reflections	102181(6697)	37701(2875)	59125(4537)	80333(11585)	49000 (7027)
<i>R</i> _{work} / <i>R</i> _{free}	0.172/0.193	0.199/0.247	0.177/0.210	0.163/ 0.202	0.191/0.233
No. atoms					
Protein	8687	8712	8707	8705	8641
Ligand/ion	40	65	139	287	148
Water	762	100	486	874	134
<i>B</i> -factors					
Protein	36.8	48.5	48.8	32.3	42.1
Ligand/ion	56.1	79.5	50.4	27.5	55.8
Water	40.0	32.4	40.2	35.5	22.8
R.m.s. deviations					
Bond lengths (Å)	0.010	0.008	0.010	0.010	0.01
Bond angles (°)	0.91	0.95	1.00	0.97	1.04

Protein^a	<i>Ft</i>	<i>Ft</i>	<i>Ft, D117N</i>	<i>Mr, N121D</i>	<i>Mr, N121D</i>
Ligands	PolyP ₉	PolyP ₆ AMPPCPPP	PolyP _{23,11,6}		ATP; P _i
PDB ID	5LL0	5LLB	5LLF	5O6K	5O6M
Space group	P 1 2 ₁ 1	P 1 2 ₁ 1	C 2 2 2 ₁	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions					
<i>a, b, c</i> (Å)	59.22, 145.62, 70.43	58.7, 144.91, 70.59	73.43, 165.80, 255.12	166.66, 166.66, 94.90	164.62, 164.62, 95.12
α, β, γ (°)	90, 113, 90	90, 113, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	72.81-1.96 (2.01-1.96)	72.46 - 1.92 (1.989 - 1.92)	44.16-2.20 (2.24-2.20)	47.91-2.90 (3.06-2.90)	47.53-2.30 (2.36-2.30)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.060 (0.420)	0.061 (0.555)	0.130 (2.21)	0.317 (1.79)	0.133 (1.66)
<i>R</i> _{meas}	0.197 (2.09)	0.073 (0.647)	0.146 (2.47)	0.329 (1.76)	0.136 (1.69)
<i>I</i> / σ <i>I</i>	11.1 (2.5)	13.6 (2.1)	8.06 (0.83)	7.5 (1.4)	22.1 (2.9)
CC1/2	0.998 (0.847)	0.998 (0.785)	0.995 (0.241)	0.997 (0.771)	1.000 (0.928)
CC*	0.999 (0.958)	0.999 (0.938)	0.999 (0.624)	1.000 (0.921)	1.000 (0.977)
Completeness (%)	97.1 (90.2)	99.3 (99.5)	99.5 (99.0)	99.9 (99.9)	100.0 (100.0)
Redundancy	3.3 (3.4)	3.1 (3)	5.3 (5.2)	14.0 (14.3)	26.4 (27.6)
Refinement					
No. reflections	75983 (5217)	81970 (8194)	79080 (5789)	30049 (4310)	1545304 (122783)
<i>R</i> _{work} / <i>R</i> _{free}	0.198/0.239	0.179/0.231	0.236/0.242	0.259/0.297	0.197/0.238
No. atoms					
Protein	8068	8047	8882	8666	8845
Ligand/ion	148	264	167	50	266
Water	612	672	245	6	294
<i>B</i> -factors					
Protein	46.03	58.41	52.04	19.38	56.27
Ligand/ion	60.76	79.22	63.97	28.22	67.09
Water	42.49	55.52	51.06	6.69	46.00
R.m.s. deviations					
Bond lengths (Å)	0.003	0.012	0.008	0.003	0.003
Bond angles (°)	0.79	1.03	1.07	0.569	0.600

*Values in parentheses are for highest-resolution shell. Each data set was collected from a single crystal.

Footnotes

^aProtein sources: *Ft*, *Francisella tularensis*; *Mr*, *Meiothermus ruber*; proteins have wild type sequences except where noted. Ligands: PolyP, polyphosphate of 25 residues average length was used for crystallization. The subscripts (P_n) indicate the lengths of polyphosphate chains modeled and refined in each structure.

Table S2. Structural similarity of PPK2 structures calculated using the DALI server (21). Values are RMSD (Å) and the number of residues aligned in brackets.

Protein (PDB ID, class)^{a, b}	<i>Ft</i> + AMPPCP + polyP (5LLB, I)	<i>Mr</i> + ADP + PP_i 5MAQ (5MDQ, III)	<i>P. aeruginosa</i>^c (3CZP, II)	<i>S. meliloti</i>^c (3CZQ, I)	<i>A. aureescens</i>^c 3RHF (3RHF, III)
<i>Ft</i> + AMPPCP + polyP (5LLB, I)	-	1.5 (239)	1.7 (241)	0.9 (249)	1.9 (268)
<i>Mr</i> + ADP + PP_i (5MDQ, III)	1.5 (239)	-	1.7 (241)	1.7 (249)	1.3 (268)
<i>P. aeruginosa</i> (3CZP, II)	1.7 (241)	1.7 (241)	-	1.7 (249)	1.7 (268)
<i>S. meliloti</i> (3CZQ, I)	0.9 (249)	1.7 (249)	1.7 (249)	-	1.9 (268)
<i>A. aureescens</i> (3RHF, III)	1.9 (268)	1.3 (268)	1.7 (268)	1.9 (268)	-

Notes.

^aClasses refer to the substrate specificity of PPK2 classes shown in Fig. 1A.

^bResidues were trimmed to the core PPK2 catalytic domains as follows: *P. aeruginosa* (3CZP) residues 255-496); *S. meliloti* (3CZQ), residues 50-297; *A. aureescens* (3RHF), residues 4-272.

^cAbbreviations: *P. aeruginosa*, *Pseudomonas aeruginosa*; *S. meliloti*, *Sinorhizobium meliloti*; *A. aureescens*, *Arthrobacter aureescens*.

Table S3. Oligonucleotides used in this study. Restriction sites or mutagenesis sites are in bold and underlined.

Name	Oligo sequence*
<i>Mr</i> PPK2-NdeI	5'- TATATATAC <u>CATATG</u> AAAAAATACCGCG -3'
<i>Mr</i> PPK2-HindIII	5'- TATATATA <u>AAGCTT</u> TTACTCAATAACAATTTTTTCGC -3'
<i>Ft</i> PPK2-D117N	5'- AATCGTACTTTTT <u>A</u> ATAGGTCTTGGTA -3'
<i>Ft</i> PPK2-D62A	5'- GAAGGTAGAG <u>CG</u> GCAGCTGGTA -3'
<i>Ft</i> PPK2-K66A	5'- TGCAGCTGGT <u>GCG</u> GGTGGAACT -3'
<i>Ft</i> PPK2-D192A	5'- AGCCCTATAG <u>CC</u> AAAGCATCCT -3'
<i>Ft</i> PPK2-R118A	5'- CGTACTTTTTGAT <u>GCG</u> TCTTGGTATAAT -3'
<i>Ft</i> PPK2-R178A	5'- GATTTGCTGCT <u>GCG</u> GAAAGTCATCC -3'

*Oligonucleotides were sourced from commercial suppliers.

Table S4. Summary of RALS analysis of PPK2 oligomerization.

Protein	Oligomer	Calculated MW (kDa) ^a	Without ligands		With 10 mM PolyP	
			Observed MW (kDa)	Amount of Protein ^b (%)	Observed MW (kDa)	Amount of Protein ^b (%)
<i>Ft</i>PPK2 WT	Monomer	32.9	39.7	100	38.2	24
	Dimer	65.8			64.0	76
<i>Mr</i>PPK2 WT	Monomer	33.7				
	Dimer	67.4				
	Trimer	101.1	99.2	100		
	Tetramer	134.8			130.3	19
	Octamer	269.6			251.7	15
	Dodecamer	404.4			382.9	13
	Hexadecamer	539.2			529.6	16

^aIncludes His-tag (calculated with ExPASy protparam tool (<https://web.expasy.org/protparam/>)).

^bAmount of protein (as a percentage of sample) represent the percentage of total area fitted in Fig. S9.

Table S5. Conditions for HPLC Analysis for PPK2 Kinetic Assays.

	<i>Ft</i> PPK2	<i>Mr</i> PPK2
Solid phase	Gemini C18 (150 x 4.6 mm, 5 μ)	MultoHigh 100 RP-18 (250 x 4 mm, 5 μ)
Mobile phase	0.1 M triethylamine, pH 7*	40 mM sodium acetate,
A (aqueous)		pH 4.2*
B (organic)	80% MeOH, 0.1 M triethyl- amine, pH 7*	acetonitrile
Detection wavelength	260 nm	254 nm
Flow rate	0.8 mL/min	0.5 mL/min
Equilibration	95% A	98% A
Sample injected	100 μ L	10 μ L
Gradient	24 min 95% A	10 min 98% A
	1 min 95% \rightarrow 90% A	6 min 98% \rightarrow 70% A
	10 min 90% A	2 min 70% A
	1 min 90% \rightarrow 0% A	2 min 70% \rightarrow 98% A
	5 min 100% B	10 min 98% A (re-equilibration)
	1 min 0% \rightarrow 95% A	
	10 min 95% A (re-equilibration)	

*adjusted with acetic acid.

References

1. Vogel HJ & Bridger WA (1982) Phosphorus-31 nuclear magnetic resonance studies of the methylene and fluoro analogues of adenine nucleotides. Effects of pH and magnesium ion binding. *Biochemistry* 21(2):394-401.
2. Cremosnik GS, Hofer A, & Jessen HJ (2014) Iterative synthesis of nucleoside oligophosphates with phosphoramidites. *Angew Chem Int Ed Engl* 53(1):286-289.
3. Batten LE, *et al.* (2015) Biochemical and structural characterization of polyphosphate kinase 2 from the intracellular pathogen *Francisella tularensis*. *Biosci Rep* 36(1):e00294.
4. Qi D & Scholthof KB (2008) A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. *J Virol Methods* 149(1):85-90.
5. Gorrec F (2009) The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography* 42:1035-1042.
6. Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43:186-190.
7. Langer G, Cohen SX, Lamzin VS, & Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* 3(7):1171-1179.
8. Long F, Vagin AA, Young P, & Murshudov GN (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D-Biological Crystallography* 64:125-132.
9. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40:658-674.
10. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 66:486-501.
11. Afonine PV, *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D-Biological Crystallography* 68:352-367.
12. Moriarty NW, Grosse-Kunstleve RW, & Adams PD (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D-Biological Crystallography* 65:1074-1080.
13. Kabsch W (2010) Xds. *Acta Crystallographica Section D-Biological Crystallography* 66:125-132.
14. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography* 62:72-82.
15. Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D-Biological Crystallography* 69:1204-1214.
16. Winn MD, *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* 67:235-242.
17. Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallographica Section D-Biological Crystallography* 66:22-25.
18. Bricogne G, *et al.* (2016) AutoBUSTER (Global Phasing Ltd, Cambridge, U.K.).
19. Corpet F (1988) Multiple Sequence Alignment with Hierarchical-Clustering. *Nucleic Acids Res* 16(22):10881-10890.
20. Waterhouse AM, Procter JB, Martin DMA, Clamp M, & Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.
21. Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545-549.