# Polyphosphate-dependent synthesis of ATP and ADP by the family-2 polyphosphate kinases in bacteria

Boguslaw Nocek[a], Samvel Kochinyan[b], Michael Proudfoot[b], Greg Brown[b], Elena Evdokimova[a,b], Jerzy Osipiuk[a], Aled M. Edwards[a,b], Alexei Savchenko[a,b], Andrzej Joachimiak[a,c,1], and Alexander F. Yakunin[b,1]

[a]Midwest Center for Structural Genomics and Structural Biology Center, Department of Biosciences, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Argonne, IL 60439; [b]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada M5G 1L6; and [c]Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637

Edited by Robert Haselkorn, University of Chicago, Chicago, IL, and approved September 30, 2008 (received for review August 1, 2008)

**Inorganic polyphosphate (polyP) is a linear polymer of tens or hundreds of phosphate residues linked by high-energy bonds. It is found in all organisms and has been proposed to serve as an energy source in a pre-ATP world. This ubiquitous and abundant biopolymer plays numerous and vital roles in metabolism and regulation in prokaryotes and eukaryotes, but the underlying molecular mechanisms for most activities of polyP remain unknown. In prokaryotes, the synthesis and utilization of polyP are catalyzed by 2 families of polyP kinases, PPK1 and PPK2, and polyphosphatases. Here, we present structural and functional characterization of the PPK2 family. Proteins with a single PPK2 domain catalyze polyP-dependent phosphorylation of ADP to ATP, whereas proteins containing 2 fused PPK2 domains phosphorylate AMP to ADP. Crystal structures of 2 representative proteins, SMc02148 from *Sinorhizobium meliloti* and PA3455 from *Pseudomonas aeruginosa*, revealed a 3-layer $\alpha/\beta/\alpha$ sandwich fold with an $\alpha$-helical lid similar to the structures of microbial thymidylate kinases, suggesting that these proteins share a common evolutionary origin and catalytic mechanism. Alanine replacement mutagenesis identified 9 conserved residues, which are required for activity and include the residues from both Walker A and B motifs and the lid. Thus, the PPK2s represent a molecular mechanism, which potentially allow bacteria to use polyP as an intracellular energy reserve for the generation of ATP and survival.**

crystal structure | mutagenesis | Walker motif | AMP phosphorylation | ADP phosphorylation

Inorganic polyphosphate (polyP) is a linear polymer composed of tens to hundreds of orthophosphate residues ($P_i$) linked by the energy-rich phosphoanhydride bonds, and it is found in all prokaryotes and eukaryotes (1–3). PolyP has numerous biological functions that include substitution for ATP in kinase reactions, acting as an energy source or storage reservoir of $P_i$, chelating metals, and adjusting cellular physiology during growth, development, stress, starvation, and virulence (1, 4). Bacteria with low intracellular polyP levels show defective responses to various stress conditions, biofilm formation, quorum sensing, motility, and other virulence properties (3, 5). Depending on the physiological state of the bacterium, the intracellular concentration of polyP is 0.1–200 mM. The polyP level is controlled mainly by the activity of several polyphosphatases and 2 families of polyP kinases, PPK1 and PPK2 (1, 6–9).

PPK1 (EC 2.7.4.1; InterPro IPR003414) is responsible for the synthesis of most of the cellular polyP, using the terminal phosphate of ATP as substrate (10, 11). The only characterized PPK2 (IPR005660) enzyme, PA0141 from *Pseudomonas aeruginosa*, uses polyP as a substrate to generate GTP from GDP (12). PPK2 shows no sequence similarity to PPK1 and is distinguished by much higher polyP utilization activity (13). The PPK2 enzyme PA0141 from *P. aeruginosa* preferentially phosphorylates GDP to GTP, whereas its affinity for ADP is 2.5 times lower (12, 13). Because the expression of PA0141 in *P. aeruginosa* cells is increased >100 times during the stationary growth phase, it has been suggested that this enzyme functions in the generation of GTP for the synthesis of alginate, an exopolysaccharide essential for the virulence of *P. aeruginosa* (12). In the social slime mold *Dictyostelium discoideum*, a different PPK was found (DdPPK2) with sequence and properties similar to that of actin-related proteins (Arps) (14). DdPPK2 is a complex of 3 Arps (Arp1, Arp2, and Arpx) and resembles the actin family in molecular weight, sequence, and filamentous structure. Remarkably, DdPPK2 can polymerize into an actin-like filament concurrent with the synthesis of polyP (14). Presently, there are 1,380 PPK1-like and 722 PPK2-like sequences in the sequenced microbial genome databases (InterPro database; www.ebi.ac.uk/interpro) with many genomes containing both PPK1 and PPK2. The PPK1 enzymes represent attractive drug targets (13), and the presence of PPK2-like genes in the genomes of a wide spectrum of pathogens coupled with their absence in the human genome suggests that PPK2s could also be targets for therapeutic intervention.

Here, we present the results of structural and biochemical characterization of 2 groups of PPK2 proteins containing 1 or 2 fused PPK2 domains. We have demonstrated that the first group, which contains a single PPK2 domain, catalyzes the polyP-dependent phosphorylation of nucleoside diphosphates (ADP, GDP) to nucleoside triphosphates, whereas the proteins with 2 fused PPK2 domains phosphorylate nucleoside monophosphates (AMP, GMP) to the respective diphosphates. Crystal structures of 2 representative proteins were solved, and site-directed mutagenesis identified the conserved residues important for catalysis.

## Results and Discussion

**Two Groups of PPK2 Enzymes.** A BLAST search of the sequenced genomes using the sequence of the biochemically characterized PPK2 enzyme PA0141 from *P. aeruginosa* as a query revealed that many genomes encode 2 or 3 PPK2 paralogs, most of which contain a single PPK2 domain ≈230 residues in length (1-domain PPK2). In addition, some genomes show the presence of a longer gene (496–544 residues) with 2 fused PPK2 domains, perhaps produced by a gene duplication event (2-domain PPK2). For example, the genome of *P. aeruginosa* encodes 2 1-domain PPK2 (PA0141 and PA2428) and 1 2-domain PPK2 (PA3455) proteins, whereas *Sinorhizobium meliloti* has 3 genes encoding only 1-domain PPK2 proteins (SMc02148, SMa0172, and SMa0670). Sequence analysis of 2-domain PPK2 proteins revealed that their C-terminal domain is more conserved and shows higher similarity to the sequences of

Author contributions: B.N., A.M.E., A.S., and A.F.Y. designed research; B.N., S.K., M.P., G.B., E.E., J.O., and A.F.Y. performed research; B.N., S.K., M.P., G.B., E.E., J.O., A.M.E., A.S., A.J., and A.F.Y. analyzed data; and B.N., A.M.E., A.S., A.J., and A.F.Y. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3CZP and 3CZQ).

[1]To whom correspondence may be addressed. E-mail: andrzejj@anl.gov or a.iakounine@utoronto.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0807563105/DCSupplemental.

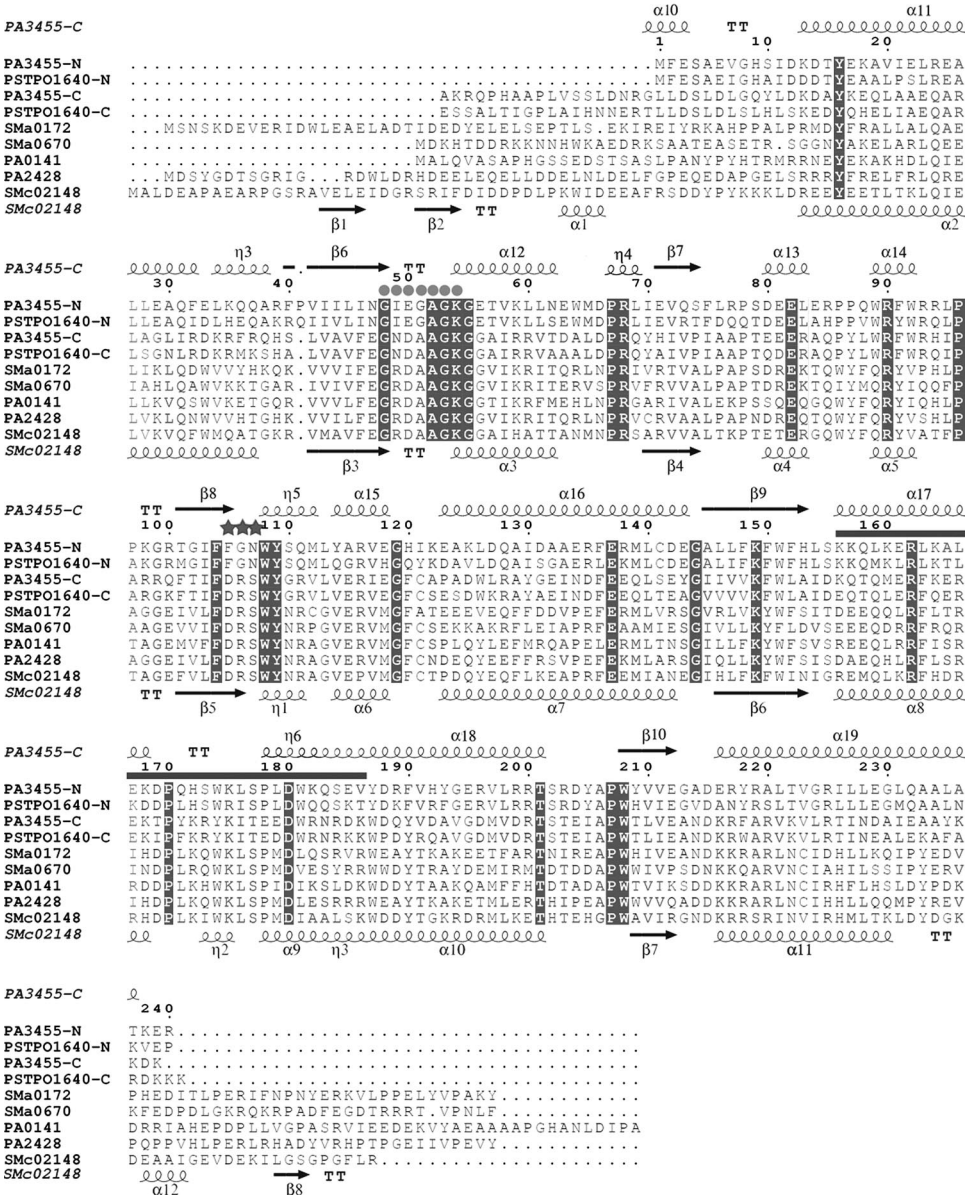


**Fig. 1.** Structure-based sequence alignment of the PPK2 domains of the proteins containing 1 or 2 PPK2 domains. Residues conserved in all PPK2 proteins are highlighted in black. The secondary structure elements of the PA3455 C-terminal PPK2 domain and SMc02148 are shown above and below the alignment, respectively. The Walker A motif is designated by balls, the Walker B motif is indicated by asterisks, and the lid module is indicated by the thick line. The single PPK2 domain proteins (short PPK2) comprise SMc02148 (Q92SA6), PA2148 (Q9I154), PA0141 (Q9I6Z1), SMa0670 (Q92ZU4), and SMa0172 (Q930V2). The 2 PPK2 domain proteins (long PPK2) include PA3455 (Q9HYF1; PA3455-N, N-terminal PPK2 domain; PA3455-C, C-terminal PPK2 domain), and PSPTO1640 (Q886D9; PSPTO1640-N, N-terminal domain; PSPTO1640-C, C-terminal domain).

1-domain PPK2 proteins than their N-terminal domains (Fig. 1). Previous bioinformatics analysis indicated that PPK2 proteins belong to the large superfamily of P-loop kinases and represent a distinct family of predicted kinases closely related to nucleotide kinases (15). P-loop kinases are characterized by the presence of 2 conserved sequence motifs (Walker A and Walker B) and the lid module. The Walker A motif (or P-loop, GXXXXGK) binds the $\beta$- and $\gamma$-phosphates of ATP, whereas the conserved Asp of the Walker B motif coordinates a $Mg^{2+}$ cation, which is also bound to the $\beta$- and $\gamma$-phosphates of ATP (15).

**Enzymatic Activity of Purified PPK2 Proteins.** We overexpressed in *Escherichia coli* and purified 2 2-domain PPK2 proteins (PA3455 from *P. aeruginosa* and PSPTO1640 from *P. syringae* pv. tomato) and 6 1-domain PPK2 proteins (PA2428 from *P. aeruginosa*, SMc02148, SMa0172, and SMa0670 from *S. meliloti*, Atu0418 from *Agrobacterium tumefaciens*, and RPA4569 from *Rhodopseudomonas palustris*). With polyP (polyP$_{12–13}$) as a phosphodonor, all of the 1-domain PPK2 proteins exhibited polyP-dependent ADP phosphorylation activity and generated ATP (2.3–13.7 $\mu$mol/min per mg of protein) [supporting information (SI) Fig. S1]. In contrast, both 2-domain PPK2 proteins catalyzed the polyP-dependent phosphorylation of AMP and produced ADP as a final product (39.0–40.0 $\mu$mol/min per mg of protein) (Fig. S1). Recently, the polyP:AMP phosphotransferase PAP from *Acinetobacter johnsonii* 210A has been shown to phosphorylate AMP in a polyP-dependent reaction (16). This protein shares 40% sequence identity with PA3455 and contains 2 fused PPK2 domains, indicating that it is a 2-domain PPK2 protein.

Both groups of PPK2 proteins were also capable of phosphorylating GMP to GDP (PA3455 and PSPTO1640) or GDP to GTP (SMc02148, SMa0172, and SMa0670), but this activity was ≈3 times lower than that with adenosine nucleotides (Table 1, data shown for PA3455 and SMc02148). Interestingly, the previously characterized PA0141 has been shown to prefer GDP over ADP as a substrate (12), indicating that short PPK2 proteins can have different substrate preferences (ADP or GDP). PA3455 also exhibited significant activity with dAMP (8.8–9.9 $\mu$mol/min per mg of protein), dGMP (2.0 $\mu$mol/min per mg of protein), IMP (2.7–2.8 $\mu$mol/min per mg of protein), and XMP (1.1–1.6 $\mu$mol/min per mg of protein). Thus, 2-domain PPK2 proteins are polyP-dependent AMP kinases

**Table 1. Kinetic parameters of PA3455 and SMc02148 with various substrates**

| Variable substrate | Constant substrate | $K_m$, μM | $k_{cat}$, $s^{-1}$ | $k_{cat}/K_m$, $M^{-1}\cdot s^{-1}$ |
|---|---|---|---|---|
| PA3455 wt* | | | | |
| Poly-P | AMP | 62 ± 4.4 | 37.7 ± 1.2 | $6.1 \times 10^5$ |
| Poly-$P_{(3)}$ | AMP | 110 ± 20 | 3.4 ± 1.3 | $0.3 \times 10^5$ |
| Poly-P | GMP | 25 ± 2.0 | 6.7 ± 0.2 | $2.7 \times 10^5$ |
| Poly-$P_{(3)}$ | GMP | 55 ± 15 | 1.9 ± 0.1 | $0.4 \times 10^5$ |
| AMP | Poly-P | 820 ± 40 | 44.6 ± 0.8 | $0.5 \times 10^5$ |
| GMP | Poly-P | 1,550 ± 120 | 13.4 ± 0.4 | $0.9 \times 10^4$ |
| $Mg^{2+}$ | AMP/Poly-P | 940 ± 20 | 40.1 ± 0.4 | $0.4 \times 10^5$ |
| $Mg^{2+}$ | GMP/ Poly-P | 2,060 ± 150 | 13.0 ± 0.5 | $0.6 \times 10^4$ |
| PA3455-C‡ | | | | |
| Poly-P | AMP | 18 ± 3.0 | 2.3 ± 0.1 | $1.3 \times 10^5$ |
| Poly-P | GMP | 26 ± 3.6 | 0.4 ± 0.02 | $1.5 \times 10^4$ |
| AMP | Poly-P | 3,340 ± 700 | 6.4 ± 0.7 | $0.2 \times 10^4$ |
| GMP | Poly-P | 1,220 ± 90 | 0.4 ± 0.01 | $0.3 \times 10^3$ |
| $Mg^{2+}$ | AMP/Poly-P | 3,670 ± 200 | 4.1 ± 0.2 | $0.1 \times 10^4$ |
| $Mg^{2+}$ | GMP/Poly-P | 2,050 ± 480 | 0.3 ± 0.04 | $0.2 \times 10^3$ |
| SMc02148 wt | | | | |
| Poly-P | ADP | 21 ± 3.9 | 8.6 ± 0.2 | $4.1 \times 10^5$ |
| ADP | Poly-P | 32 ± 4.1 | 7.6 ± 0.01 | $2.4 \times 10^5$ |
| GDP | Poly-P | 520 ± 70 | 0.8 ± 0.03 | $0.2 \times 10^4$ |
| $Mg^{2+}$ | ADP/Poly-P | 4,200 ± 300 | 8.3 ± 0.3 | $0.2 \times 10^4$ |
| SMc02148 W194A | | | | |
| Poly-P | ADP | 50 ± 20 | 2.3 ± 0.1 | $4.6 \times 10^4$ |
| ADP | Poly-P | 380 ± 30 | 1.4 ± 0.03 | $0.4 \times 10^4$ |
| SMc02148 Y233A | | | | |
| Poly-P | ADP | 20 ± 6.1 | 5.0 ± 0.2 | $2.5 \times 10^5$ |
| ADP | Poly-P | 110 ± 10 | 2.1 ± 0.04 | $1.9 \times 10^4$ |

Poly-$P_{(12–13)}$ from Sigma was used in all experiments (if not stated otherwise).
*Full-length WT protein (496 amino acids).
†C-terminal PPK2 domain (amino acids 270–496).

(PPK2M or PPK2 monophosphate-specific), whereas 1-domain PPK2 enzymes use polyP to phosphorylate ADP or GDP (PPK2D or PPK2 diphosphate specific).

To compare the reaction requirements of the 2 PPK2 groups, we used PA3455 and SMc02148 as representative enzymes for the PPK2M and PPK2D proteins, respectively. Reverse reactions (dephosphorylation of ADP or ATP in the presence of polyP) were not detected for both proteins, indicating that, like PA0141 and the *A. johnsonii* PAP (13, 16), the PPK2 proteins function preferentially in the direction of the polyP-driven nucleotide phosphorylation. Both PA3455 and SMc02148 had a broad pH optimum (pH 8.0–9.5) and showed no activity in the absence of a divalent metal cation. $Mg^{2+}$ was the most effective metal for both proteins, whereas low activity was observed with $Co^{2+}$ or $Ni^{2+}$ and no activity was observed in the presence of $Mn^{2+}$ or $Ca^{2+}$ (data not shown). Both enzymes showed comparable catalytic efficiencies ($k_{cat}/K_m$), but PA3455 had a higher affinity to $Mg^{2+}$ and higher activity, whereas SMc02148 exhibited higher affinity to polyP and ADP (Table 1). PA3455 (but not SMc02148) was also active with tri-polyP (polyP$_3$) as a phosphodonor, but its activity and substrate affinity were lower than those with polyP$_{12–13}$ (Table 1). Depending on the culture conditions, the intracellular concentrations of ADP and AMP in bacteria can vary from 0.15 to 2.51 mM (17, 18). These concentrations are close to the substrate affinity ($K_m$) of both PPK2 enzymes for these nucleotides (Table 1), suggesting that they can efficiently phosphorylate these nucleotides in vivo.

**Crystal Structures of PA3455 and SMc02148.** The crystal structures of PA3455 and SMc02148 were determined by single anomalous diffraction (SAD) phasing using seleno-L-methionine (Se-Met)-substituted enzymes to 2.0- and 2.2-Å resolution, respectively. The structure of PA3455 revealed a homodimeric organization with each monomer containing 2 PPK2 domains (residues 1–238 and 259–495) connected by a flexible linker (residues 238–258, disordered in the structure) (Fig. 2*A*). The structure of SMc02148 showed the presence of 4 PPK2 monomers in the asymmetric unit arranged as a D2 tetramer (Fig. 2*B*). The tetrameric arrangement is very similar to that of the 4 PPK2 domains of the PA3455 dimer (which looks like a pseudotetramer) (Fig. 2). Both the PA3455 dimer and SMc02148 tetramer have the shape of a rectangular box. The oligomeric organization of both proteins is consistent with the results of the gel-filtration analysis, which showed that PA3455 was a dimer (97 kDa) and SMc02148 was a tetramer (124.5 kDa) in solution. A Dali search for PA3455 and SMc02148 structural homologs identified >30 probable or known thymidylate kinases as the closest structural homologs. The top 3 hits were the same for

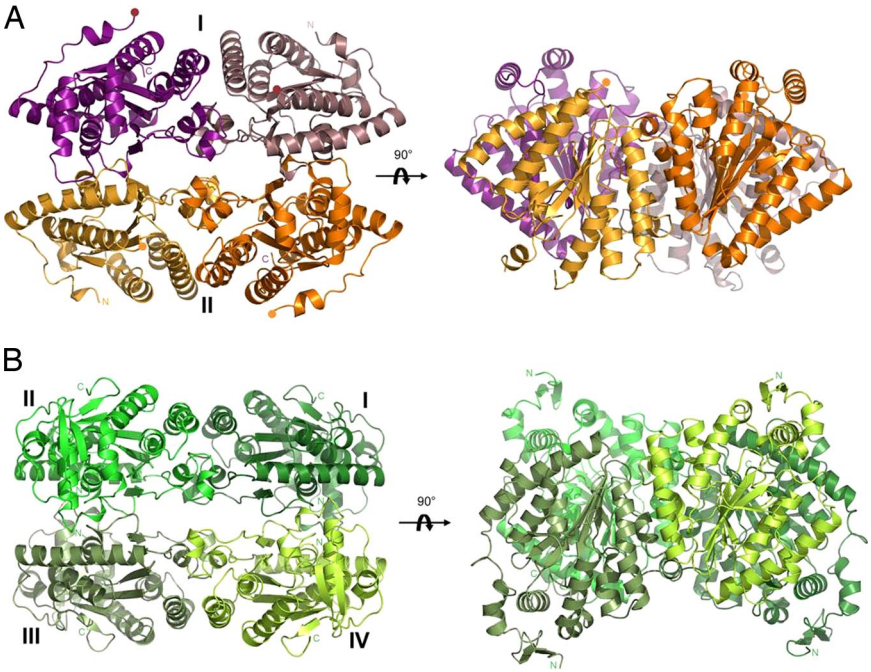


**Fig. 2.** Overall crystal structures of the PA3455 dimer (*A*) and SMc02148 tetramer (*B*). For both proteins, 2 views related by a 90° rotation are shown. PPK2 domains are shown in different colors, and the Roman numerals designate protein monomers. The proteins show a very similar packing with the dimensions 88 × 68 × 50 Å (PA3455) and 85 × 67 × 63 Å (SMc02148).

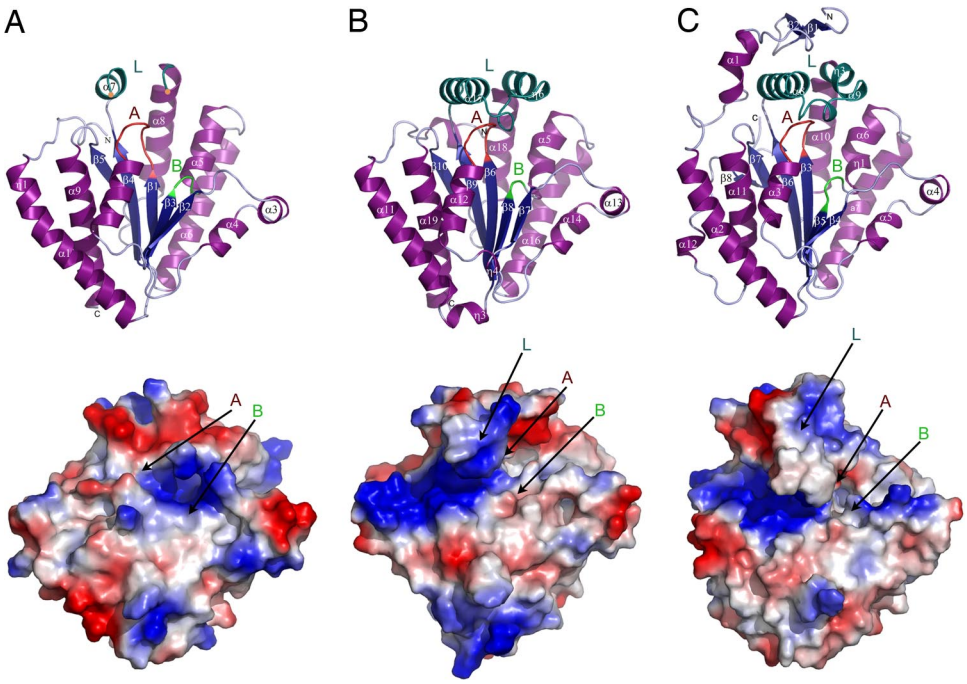


**Fig. 3.** Structures of the PPK2 domains of PA3455 and SMc02148. (*Upper*) Ribbon diagrams of the N-terminal PPK2 domain of PA3455 (*A*), C-terminal PPK2 domain of PA3455 (*B*), and SMc02148 monomer (*C*). The secondary structure elements are shown in different colors ($\alpha$-helices, dark magenta; $\beta$-strands, dark blue; loops, light blue) and are labeled. The Walker A loop is colored in red, the Walker B loop in green, and the lid module helices in dark green, and they are indicated by the capital letters (A, B, and L, respectively). Two yellow dots in 4A designate the boundaries of the disordered part of the lid module of the PA3455 N-terminal PPK2 domain. (*Lower*) Surface charge distribution of the same PPK2 domains showing the presence of the extended positively charged patch on the left side of the PA3455 C-terminal PPK2 domain (*B*) and SMc02148 (*C*), which is absent in the PA3455 N-terminal domain (*A*). The surface charge distribution was determined by using PyMOL (http://pymol.sourceforge.net).

both proteins: ST1543 from *Sulfolobus tokodaii* [Protein Data Bank (PDB) ID code 2PLR; *Z*-score 12.3 and 14.4, rmsd 2.9 and 3.1 Å], Tmk from *Staphylococcus aureus* (PDB ID code 2CCG; *Z*-score 11.9 and 13.5, rmsd 3.6 Å), and DTYMK from human (PDB codes 1NMY and 1NN0; *Z*-score 11.5 and 13.5, rmsd 3.2 and 3.4 Å).

The structures of both the N- and C-terminal PPK2 domains of PA3455 and the SMc02148 monomer are arranged into a 3-layer $\alpha/\beta/\alpha$ sandwich with the central 5-stranded (in PA3455) or 6-stranded (in SMc02148) parallel $\beta$-sheet flanked by $\alpha$-helices on both sides and on the top (Fig. 3). The C-terminal PPK2 domain of PA3455 overlays well with the N-terminal domain (*Z*-score 9.5, rmsd of equivalent C$\alpha$ atoms 1.56 Å) and with SMc02148 (*Z*-score 10.8, rmsd 1.44 Å) (Fig. S2). The central $\beta$-sheet of the PA3455 C-terminal PPK2 domain is flanked by 3 long helices ($\alpha$11, $\alpha$12, $\alpha$19) on one side and 5 shorter helices ($\alpha$13, $\alpha$14, $\alpha$15, $\alpha$16, $\alpha$18) on the other side. On the top, the central $\beta$-sheet is covered by the lid comprised of 2 helices ($\alpha$17 and $\eta$6) and containing several conserved basic residues (Arg-419, Arg-423, Lys-429, Lys-432, Lys-443) (Fig. 3). SMc02148 has a very similar lid module ($\alpha$8, $\alpha$9, and $\eta$6), which is covered by the small N-terminal extension domain ($\beta$1, $\beta$2, and $\alpha$1) on the top. The lid module of the N-terminal PPK2 domain of PA3455 was only partially modeled because of missing density, suggesting significant conformational flexibility of this module like in thymidylate kinases (19). In contrast, the lid modules of the PA3455 C-terminal PPK2 domain and SMc02148 were well resolved. This difference might be attributed to the presence of small-molecule ligands trapped in these proteins below the lid. Structure refinement of SMc02148 and the C-terminal PPK2 domain of PA3455 revealed the presence of several electronic densities, which were interpreted as glycerol, malonic acid, and formic acid based on their shape and presence in the crystallization solution. In both structures, these ligands interact with the side chains and backbone atoms of the conserved residues from the lid and both Walker motifs (Ala-308, Gly-310, Lys-311, Gly-312, Asp-362, and Arg-423 in PA3455 and Gly-96, Lys-97, Arg-209, and Lys-218 in SMc02148) (Fig. S3). The coordination of these ligands perhaps imitates the location of the phosphate groups of the substrate (probably polyP) bound to the PPK2 domain and suggests the position of the PPK2 active site (under the lid module near the 2 Walker loops). Below the lid, 3 structures hold 2 very similar loops containing the Walker motifs A and B (Fig. 3).

**The N-Terminal PPK2 Domain of PA3455 Is Catalytically Inactive.** The Walker B loop of the PA3455 N-terminal PPK2 domain contains Phe-104 and Gly-105 instead of a conserved Asp–Arg motif (Fig. 1) and a different distribution of positively charged residues on the protein surface (Fig. 3*A*), suggesting that this domain might be catalytically inactive. To verify this idea, we cloned the individual N- and C-terminal PPK2 domains of PA3455 into a protein expression vector and purified them. Both of the PA3455 C-terminal fragments (containing residues 261–496 and 270–496) catalyzed polyP-dependent phosphorylation of AMP to ADP. When compared with the native (full-length) PA3455, the C-terminal PPK2 domain (residues 270–496) exhibited reduced catalytic efficiency ($k_{cat}/K_m$) and affinity ($K_m$) to AMP and $Mg^{2+}$ (Table 1). In contrast, the N-terminal PPK2 domain of PA3455 (residues 1–270) did not show any detectable activity in both forward (polyP-dependent AMP or GMP phosphorylation) and reverse (ADP- or GDP-dependent polyP phosphorylation) reactions. Moreover, no enzymatic activity was observed upon incubation of the N-terminal domain of PA3455 with polyP and other canonical riboxyribonucleoside or deoxyribonucleoside triphosphates, diphosphates, or monophosphates. The lack of the enzymatic activity in the PA3455 N-terminal PPK2 domain is consistent with the absence of the Walker B Asp–Arg motif, which has been shown to play an important role in the catalytic mechanism of thymidylate kinases (15, 20). A protein surface analysis also revealed that the N-terminal domain does not have the large positively charged patch on the left side of the domain, which is likely to be involved in the binding of polyP (Fig. 3*A*). Finally, alanine replacement mutagenesis of 3 residues of the C-terminal PPK2 domain (Asp-307, Lys-311, and Asp-437) in the native PA3455 produced catalytically inactive proteins, indicating that the C-terminal PPK2 domain of PA3455 is responsible for its enzymatic activity. Thus, the N-terminal domain of PA3455 and probably other PPK2M proteins seems to be catalytically inactive and might be responsible for the protein dimerization because we found that the individually expressed C-terminal domain exists mainly as a monomer in solution (17 kDa).

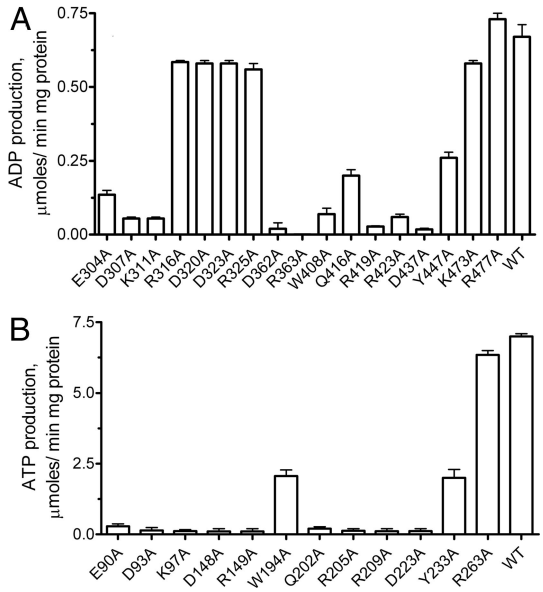


**Fig. 4.** Alanine replacement mutagenesis of the PA3455 C-terminal PPK2 domain (*A*) and SMc02148 (*B*). Enzymatic activity of purified mutant proteins was measured as polyP-dependent phosphorylation of AMP to ADP (for the C-terminal PPK2 domain of PA3455, residues 261–496) or ADP to ATP (for SMc02148). Experimental conditions were as described in *Materials and Methods*.

**Alanine Replacement Mutagenesis of PA3455 and SMc02148.** Structure-based sequence alignment of PA3455 and SMc02148 with several other PPK2 proteins identified 24 strictly conserved amino acid residues (Fig. 1). The overall structural similarity of PPK2 domains to thymidylate kinases suggests that the active site of PPK2s is located near the 2 Walker motifs. Alanine replacement mutagenesis of the conserved residues of the PA3455 C-terminal PPK2 domain and SMc02148 revealed that 9 mutant proteins have a greatly reduced or very low activity, indicating that these residues are important for catalysis (Fig. 4). As expected, the mutagenesis of the Walker B Asp–Arg motif (D362A and R363A in PA3455 and D148A and R149A in SMc02148) produced inactive proteins. Alanine replacement of 3 charged residues located close to or within the Walker A motif (E304A, D307A, and K311A in PA3455; E90A, D93A, and K97A in SMc02148) resulted in a complete (or almost complete) loss of enzymatic activity (Fig. 4). In addition, mutagenesis of the conserved residues of the lid module (R419A and D437A in PA3455 and R205A and D223A in SMc02148) produced proteins with low or no activity, indicating that, like in thymidylate kinases, the lid module plays an important role in the activity of PPK2 domains.

**Potential Mechanisms of Substrate Binding and Catalysis by PPK2 Enzymes.** Structural similarity between the PPK2 domains and thymidylate kinases and the conservation of the key catalytic residues of thymidylate kinases in PPK2 enzymes suggest that these enzymes have a common evolutionary origin and catalytic mechanism. Our model of the PPK2 catalytic mechanism is based on the mechanisms proposed for thymidylate kinases from *E. coli* (Tmk) and *Mycobacterium tuberculosis* ($TMPK_{Mtub}$) (19, 21) and is supported by the results of our mutagenic studies (Fig. 4 and Table 1). Overlay experiments using structures of thymidylate or adenylate kinase complexes with substrates or inhibitors (PDB ID codes 1NMY or 1ZIN) and SMc02148 (Fig. S4) suggest that in PPK2 enzymes the nucleotide substrate (ADP or AMP) binds to the protein on the right side of the Walker A-containing loop in the pocket between 2 $\alpha$ helices ($\alpha$5 and $\alpha$6 in SMc02148 and $\alpha$14 and $\alpha$15 in PA3455) (Fig. 5). This pocket contains the side chains of several conserved residues that are likely to contribute to substrate binding. In SMc02148, the adenine base might be sandwiched between the side chains of Val-157 and Phe-163 and coordinated through hydrogen bonding with Glu-90, Tyr-130, Arg-133, Tyr-134, and Glu-158. This suggestion is supported by the greatly reduced activity of SMc02148 E90A and PA3455 E304A (Fig. 4). Like in *E. coli* Tmk (19), the side chain of the Walker A motif Asp-93 (Asp-307 in PA3455) probably interacts with the ribose 3′-hydroxyl, whereas its 2′-hydroxyl might form a hydrogen bond to Tyr-233 (Tyr-447 in PA3455), which is consistent with the reduced affinity of SMc02148 Y233A to ADP and low activity of D93A (and PA3455 D307A and Y447A) (Table 1 and Fig. 4). The side chain of the conserved Walker B Asp-148 (Asp-362 in PA3455) might coordinate the $Mg^{2+}$ ion, which might also be bound to the terminal phosphate oxygen of AMP or ADP as in the *M. tuberculosis* $TMPK_{Mtub}$ (21). We suggest that the adjacent Arg-149 (Arg-363 in PA3455) of the Walker B motif interacts with the terminal phosphate groups of AMP (or ADP) and polyP (possibly in concert with $Mg^{2+}$) (Fig. 5). The substrate phosphate groups are likely to interact with the side chain of the conserved Arg-111 from the adjacent monomer (Arg-325 in PA3455), which is positioned close to Asp-148 and Arg-149 of the Walker B motif (3.5 Å). These interactions are important for the neutralization of the electrostatic repulsion between the anionic substrates and their proper alignment for phosphoryl transfer as suggested earlier for the *M. tuberculosis* thymidylate kinase (22). The key role of this Arg residue in the PPK2 reaction is supported by our mutagenesis results (Fig. 4). In addition, the polyP oxygens are predicted to interact with the side

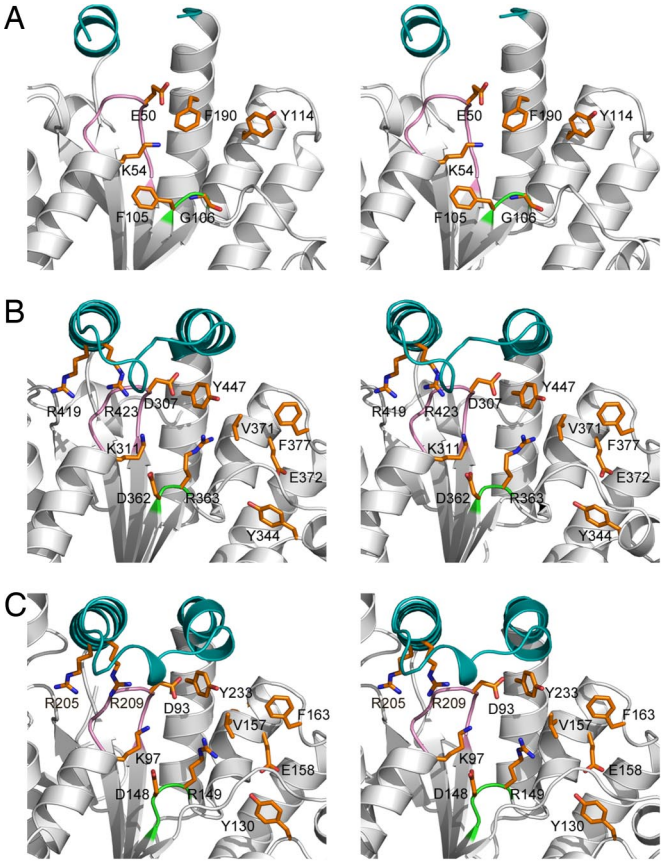


**Fig. 5.** Close-up stereoview of the active sites of PA3455 and Mc02148. (*A*) The PA3455 N-terminal domain. (*B*) The PA3455 C-terminal domain. (*C*) SMc02148. The secondary structure elements are shown in light gray with the colored catalytic modules: the Walker A loop in violet, the Walker B loop in green, and the lid in teal. The side chains of the conserved residues are shown as orange sticks and labeled.

chains of the conserved basic residues of the Walker A motif (Lys-97 in SMc02148 and Lys-311 in PA3455), the lid module (Arg-205, Arg-209 in SMc02148 and Arg-419, Arg-423 in PA3455) (Fig. 5), and the predicted polyP binding site (shown as a blue area on the left from the lid in Fig. 3 *B* and *C*). Like in the proposed catalytic mechanism of thymidylate kinases (21), we suggest that in PPK2 enzymes a Walker A carboxylate residue (Asp-93 in SMc02148 and Asp-307 in PA3455) acts as a general base attracting 1 hydroxyl proton from the terminal phosphate group of the nucleotide (AMP or ADP) and preparing it to attack the terminal phosphate of polyP.

## Conclusion

In most prokaryotes, polyP is produced by the PPK1 family of PPKs, which use ATP as a phosphodonor (1, 3, 5). The presence of the energy-rich phosphoanhydride bond in polyP suggests that energy conservation might be one of the molecular functions of polyP in the cell (1, 3, 5). Our results, together with 2 previous studies (13, 16), demonstrated that the enzymes of the PPK2 family of PPKs function preferentially as polyP-dependent nucleotide kinases. The proteins with 2 PPK2 domains phosphorylate AMP to ADP, whereas single PPK2 domain kinases use polyP to phosphorylate ADP or GDP. Thus, by the combined action of these 2 groups of PPK2 enzymes, AMP can be converted to ATP. Although it has been suggested that intracellular polyP is not likely to be able to support the high growth rate of microbial cells per se (1), the polyP-driven synthesis of ADP and ATP by the PPK2 enzymes might be important for the survival of microbial cells under conditions of stress or pathogenesis.

## Materials and Methods

**Protein Overexpression, Purification, and Site-Directed Mutagenesis.** The cloning of the genes encoding PA3455, SMc02148, and other PPK2 proteins (PAPTO1640, PA2428, SMa0172, SMa0670, Atu0418, and RPA4569) into the modified pET15b was carried out as described (23). The proteins were expressed as a fusion with an N-terminal $His_6$ tag in the *E. coli* strain BL21 (DE3) and purified as described (23, 24). Seleno-methionine (SeMet)-labeled proteins were produced by overexpression in the *E. coli* methionine auxotroph strain B834 (DE3) (Novagen). Alanine replacement mutagenesis of PA3455 and SMc02148 was performed as described (24).

**Enzymatic Assays.** polyP-dependent nucleoside monophosphate (AMP, GMP) kinase activity of PPK2 proteins was determined by using the modified enzyme-coupled assay with pyruvate kinase and lactate dehydrogenase (25). Reaction mixtures (0.2 mL) contained 50 mM Tris·HCl (pH 8.0), 10 mM NaCl, 10 mM $MgCl_2$, 5 mM AMP (or GDP), 0.25 mM polyP ($polyP_{12–13}$; Sigma), 1 mM phospho*enol*pyruvate, 1 mM NADH, 1.9 units of pyruvate kinase (Sigma), 2.5 units of L-lactate dehydrogenase (Sigma), and 0.2–0.5 $\mu$g of enzyme. PolyP-dependent phosphorylation of nucleoside diphosphates (ADP, GDP) was measured by using the modified enzyme-coupled assay with hexokinase and glucose 6-phosphate dehydrogenase (26) by using a reaction mixture (0.2 mL) containing 50 mM Tris·HCl (pH 8.0), 10 mM NaCl, 12 mM $MgCl_2$, 5 mM ADP (or GDP), 0.25 mM polyP ($polyP_{12–13}$; Sigma), 5 mM glucose, 1 mM NAD, 0.6 units of hexokinase (Sigma), 0.6 units of glucose 6-phosphate dehydrogenase (Sigma), and 0.2–0.5 $\mu$g of enzyme. PolyP-dependent phosphorylation of other nucleotides and the reverse PPK2 reaction (ATP- or ADP-dependent phosphorylation of polyP) (16) were analyzed by using an HPLC-based assay (AXXI-CHROM ODS column with 5-$\mu$m-size particles, 4.6 $\times$ 250 mm; Cole Scientific) with detection at 254 nm as described (27). Kinetic parameters were determined by nonlinear curve fitting using GraphPad Prism software (version 4.00 for Windows). TLC analysis of the PPK2 reaction products was performed with cellulose plates (Sigma) as described (28).

**Protein Crystallization and Structure Determination.** Crystals of the SeMet-labeled PA3455 and SMc02148 were grown by the hanging-drop vapor-diffusion method at 18°C as described (28). PA3455 was crystallized in the presence of 0.1 M Hepes-K (pH 7.0), 5% Tacsimate, and 10% PEG MME 5K, whereas the crystals of SMc02148 were grown in the presence of 0.1 M Bis-Tris (pH 7.5), 0.1 M Na-formate, 0.1 M Li-sulfate, and 0.3 M NDSB 211. Diffraction data were collected at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source (Argonne, IL) (29). The data were processed with HKL-3000 (30). The structures were determined by SAD phasing, density modification, and initial protein model building as implemented in the HKL-3000 software package (30). Initial models were further processed with the Arp/Warp software, and the final models were built with the program COOT (31) and refined with the program REFMAC5 (32). The data collection and refinement statistics are shown in Table S1.

**ACKNOWLEDGMENTS.** We thank all members of the Structural Proteomics in Toronto Center and Structural Biology Center at Argonne National Laboratory for help in conducting these experiments. This work was supported by Genome Canada (through the Ontario Genomics Institute), National Institutes of Health Grant GM074942, and the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357.

1. Kornberg A, Rao NN, Ault-Riche D (1999) Inorganic polyphosphate: A molecule of many functions. *Annu Rev Biochem* 68:89–125.
2. Kulaev IS, Vagabov VM (1983) Polyphosphate metabolism in microorganisms. *Adv Microb Physiol* 24:83–171.
3. Brown MR, Kornberg A (2008) The long and short of it: Polyphosphate, PPK, and bacterial survival. *Trends Biochem Sci* 33:284–290.
4. Rashid MH, *et al.* (2000) Polyphosphate kinase is essential for biofilm development, quorum sensing, and virulence of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 97:9636–9641.
5. Brown MR, Kornberg A (2004) Inorganic polyphosphate in the origin and survival of species. *Proc Natl Acad Sci USA* 101:16085–16087.
6. Wurst H, Kornberg A (1994) A soluble exopolyphosphatase of *Saccharomyces cerevisiae*. Purification and characterization. *J Biol Chem* 269:10996–11001.
7. Rao NN, Kornberg A (1996) Inorganic polyphosphate supports resistance and survival of stationary-phase *Escherichia coli*. *J Bacteriol* 178:1394–1400.
8. Andreeva NA, Kulakovskaya TV, Kulaev IS (1998) Purification and properties of exopolyphosphatase isolated from *Saccharomyces cerevisiae* vacuoles. *FEBS Lett* 429:194–196.
9. Lichko L, Kulakovskaya T, Kulaev I (1998) Membrane-bound and soluble polyphosphatases of mitochondria of *Saccharomyces cerevisiae*: Identification and comparative characterization. *Biochim Biophys Acta* 1372:153–162.
10. Ahn K, Kornberg A (1990) Polyphosphate kinase from *Escherichia coli*: Purification and demonstration of a phosphoenzyme intermediate. *J Biol Chem* 265:11734–11739.
11. Kumble KD, Ahn K, Kornberg A (1996) Phosphohistidyl active sites in polyphosphate kinase of *Escherichia coli*. *Proc Natl Acad Sci USA* 93:14391–14395.
12. Ishige K, Zhang H, Kornberg A (2002) Polyphosphate kinase (PPK2), a potent, polyphosphate-driven generator of GTP. *Proc Natl Acad Sci USA* 99:16684–16688.
13. Zhang H, Ishige K, Kornberg A (2002) A polyphosphate kinase (PPK2) widely conserved in bacteria. *Proc Natl Acad Sci USA* 99:16678–16683.
14. Gomez-Garcia MR, Kornberg A (2004) Formation of an actin-like filament concurrent with the enzymatic synthesis of inorganic polyphosphate. *Proc Natl Acad Sci USA* 101:15876–15880.
15. Leipe DD, Koonin EV, Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. *J Mol Biol* 333:781–815.
16. Shiba T, *et al.* (2005) Polyphosphate:AMP phosphotransferase as a polyphosphate-dependent nucleoside monophosphate kinase in *Acinetobacter johnsonii* 210A. *J Bacteriol* 187:1859–1865.
17. Bhattacharya M, *et al.* (1995) Single-run separation and detection of multiple metabolic intermediates by anion-exchange high-performance liquid chromatography and application to cell pool extracts prepared from *Escherichia coli*. *Anal Biochem* 232:98–106.
18. Buchholz A, Takors R, Wandrey C (2001) Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. *Anal Biochem* 295:129–137.
19. Lavie A, *et al.* (1998) Structural basis for efficient phosphorylation of 3′-azidothymidine monophosphate by *Escherichia coli* thymidylate kinase. *Proc Natl Acad Sci USA* 95:14045–14050.
20. Lavie A, *et al.* (1997) Structure of thymidylate kinase reveals the cause behind the limiting step in AZT activation. *Nat Struct* Biol 4:601–604.
21. Fioravanti E, *et al.* (2003) *Mycobacterium tuberculosis* thymidylate kinase: Structural studies of intermediates along the reaction pathway. *J Mol Biol* 327:1077–1092.
22. Li de la Sierra I, *et al.* (2001) X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95-Å resolution. *J Mol Biol* 311:87–100.
23. Zhang RG, *et al.* (2001) Structure of *Thermotoga maritima* stationary-phase survival protein SurE: A novel acid phosphatase. *Structure (London)* 9:1095–1106.
24. Proudfoot M, *et al.* (2008) Biochemical and structural characterization of a novel family of cystathionine $\beta$-synthase domain proteins fused to a Zn ribbon-like domain. *J Mol Biol* 375:301–315.
25. Huo X, Viola RE (1996) Functional group characterization of homoserine kinase from *Escherichia coli*. *Arch Biochem Biophys* 330:373–379.
26. Marina A, *et al.* (1998) Carbamate kinase from *Enterococcus faecalis* and *Enterococcus faecium*: Cloning of the genes, studies on the enzyme expressed in *Escherichia coli*, and sequence similarity with *N*-acetyl-L-glutamate kinase. *Eur J Biochem* 253:280–291.
27. Chen H, *et al.* (1997) High-performance liquid chromatographic determination of (-)-$\beta$-D-2-aminopurine dioxolane and (-)-$\beta$-D-2-amino-6-chloropurine dioxolane, and their metabolite (-)-$\beta$-D-dioxolane guanine in monkey serum, urine, and cerebrospinal fluid. *J Chromatogr B Biomed Sci Appl* 691:425–432.
28. Savchenko A, *et al.* (2007) Molecular basis of the antimutagenic activity of the house-cleaning inosine triphosphate pyrophosphatase RdgB from *Escherichia coli*. *J Mol Biol* 374:1091–1103.
29. Rosenbaum G, *et al.* (2006) The Structural Biology Center 19ID undulator beamline: Facility specifications and protein crystallographic results. *J Synchrotron Radiat* 13:30–45.
30. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M (2006) HKL-3000: The integration of data reduction and structure solution, from diffraction images to an initial model in minutes. *Acta Crystallogr D* 62:859–866.
31. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
32. Murshudov GN, *et al.* (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D* 55:247–255.