Published on Oct 06, 2023 by Arcadia Science **DOI**: 10.57844/arcadia-ac10-23e7

Discovering shared protein structure signatures connected to polyphosphate accumulation in diverse bacteria

Only some bacteria accumulate substantial amounts of polyphosphate (polyP). We thought that despite sequence divergence, polyP synthesis enzymes in these bacteria might have similar structures. We found this is sometimes true but doesn't fully explain the phenomenon.

Contributors (A-Z)

Prachee Avasthi, Brae M. Bigge, Feridun Mert Celebi, Jase Gehring, Megan L. Hochstrasser, Elizabeth A. McDaniel, Austin H. Patton, Taylor Reiter, Dennis A. Sun

Version 2 · Mar 31, 2025

Purpose

Polyphosphate is an important polymer for diverse organisms, specifically for bacterial stress response, pathogen virulence, and basic metabolism. In wastewater treatment

plants, specific microbial lineages remove phosphorus from the water by taking in orthophosphate [$(PO_4)^{3-}$] and polymerizing it into chains of polyphosphate (polyP). At a later treatment stage, these phosphorus-filled cells are removed from the water. This process is crucial for preventing eutrophication of the downstream water and maintaining environmental standards. However, identifying which microbes perform specific polyP accumulation activities in wastewater is challenging. Namely, just because a given bacterium encodes enzymes that catalyze polyP formation does not mean that the bacterium contributes meaningfully to polyP accumulation in wastewater **[1]**. This lack of predictability hinders rational engineering approaches to make the wastewater treatment process more reliable. While there could be many explanations for differing polyP accumulation phenotypes, we wondered if structural differences in polyP-polymerizing enzymes might explain this observation.

We recently developed a tool called ProteinCartography that uses protein structural similarity to identify homologous protein families **[2]**, and we thought this polyP puzzle could be an interesting test case. We hypothesized that regardless of sequence divergence, bacteria with enhanced polyP accumulation would have highly similar structures of the polyphosphate kinase PPK1, which catalyzes polyP formation, since protein structure tends to be indicative of protein function **[3]**. We first used ProteinCartography to cluster all PPK1 structures and compare them to the PPK1 protein structure from a bacterium, Accumulibacter, that we know is important for polyP accumulation in wastewater. We then explored support for our hypothesis using different metrics and visualizations, such as comparing sequence and structural similarity and phylogenetic distance against the Accumulibacter PPK1 protein.

We found examples of high PPK1 protein structural similarity within pathogenic bacteria that are phylogenetically related to Accumulibacter, and which also display enhanced polyP accumulation as part of their virulence and stress response mechanisms. Additionally, we found examples of high PPK1 structural similarity between lineages that are distantly related and are either important or abundant in the wastewater treatment process. This suggests that this method could serve as an initial screening step to prioritize lineages to be tested for polyP activity. However, these PPK1 similarity trends weren't universal compared to other experimentally verified polyP-accumulating organisms in wastewater. Overall, making useful inferences with this approach is highly dependent on curating polyP trait data, which is only available for a handful of bacterial lineages in wastewater. However, even based on this limited trait data, we were still able to come up with novel protein candidates and species that could be experimentally tested for validation purposes.

While we don't have plans to follow up on these findings for translational purposes, we think these findings may be useful to groups specifically studying phosphorus removal in wastewater treatment plants, or more broadly, to those interested in general stress responses in bacteria. This work may also be interesting to those curious about the types of insights that can be gained by exploring structural homologs of a protein of interest.

- This pub is part of the **platform effort**, "Annotation: Mapping the functional landscape of protein families across biology." Visit the platform narrative for more background and context.
- **Data** from this pub is available in Zenodo.
- All associated **code** is available in this GitHub repository.

Background and goals

Inorganic polyphosphates (polyP) are polymers of orthophosphate [$(PO_4)^{3-}$] and are ubiquitous across the tree of life, from bacteria to higher-order eukaryotes. Polyphosphates span numerous essential functions in prokaryotes across varying contexts, such as involvement in basic metabolism, sensing/responding to environmental changes, stress responses, and virulence and host immune evasion **[4] [5]**. Nearly all sequenced bacteria have the genetic repertoire for taking up inorganic phosphorus and forming chains of polyP, catalyzed by the PPK polyphosphate kinases **[6]**. Since most eukaryotes form polyP through different genetic pathways than in prokaryotes **[7][8]**, the PPK enzymes have been of particular interest as an antibiotic target for pathogens such as *Acinetobacter baumannii, Mycobacterium tuberculosis,* and *Pseudomonas aeruginosa* **[9][10][11]**. Some archaea also possess PPK enzymes, but it is unknown if they contribute significantly to environmental polyP cycling **[12]**.

Not only is polyphosphate accumulation important with respect to human pathogens, it also plays a critical role in the process of wastewater treatment. The goal of wastewater treatment is to remove inorganic nutrients such as nitrogen and phosphorus to prevent downstream eutrophication, where excessive nutrients lead to freshwater ecosystem imbalance and harmful algal blooms **[13]**. In modern-day

wastewater treatment plants, this process depends on specific microbial lineages present in wastewater, which accumulate phosphorus and are eventually removed from the water **[14]**.

Engineering these systems to improve efficiency of phosphorus removal is tricky because it's not yet clear which microbes contribute the most to polyP accumulation. It's not even clear how to predict whether a given microbe will accumulate a lot of polyP or very little — almost all bacteria have genes for phosphate polymerization machinery, but there isn't a clear correlation between sequence and accumulation activity. That said, we do know about a few groups of bacteria that accumulate high levels of polyP. As its name suggests, *Candidatus* Accumulibacter phosphatis (hereafter referred to as Accumulibacter) is a model polyphosphate-accumulating organism in wastewater within Pseudomonadota (previously Proteobacteria). *Tetrasphaera* spp. within the Actinobacteria are also abundant in Danish wastewater treatment plants and contribute to polyphosphate cycling **[15][16][17]**. Many other microbes are important in wastewater treatment as a whole, but it's not known which participate heavily in phosphate accumulation. Additionally, outside of wastewater, certain bacterial lineages store substantial amounts of intracellular polyphosphate in response to stress **[18][19]**.

Why some bacteria seem to be good at accumulating polyP and others aren't remains an open question. While there could be numerous explanations for this, such as gene expression differences, copy number variation, metabolic dynamics, etc., we decided to explore this question through the lens of protein sequence and predicted protein structural similarity. We hypothesized that regardless of sequence divergence or phylogenetic distance, bacteria that exhibit enhanced polyphosphate accumulation in different contexts may have highly similar PPK1 protein structures. We decided to:

1. Compare the sequences and structures of approximately 28,000 PPK1 proteins to that of the Accumulibacter PPK1 protein (since we know this bacterial lineage has high levels of polyP accumulation).
2. Look for signatures of potential convergent evolution of protein structure, which could reveal mechanistic clues about phosphate polymerization. We sought to do this by searching for examples of high structural similarity of PPK1 proteins in taxa that are either distantly related to Accumulibacter, or that we do not expect to have high structural similarity based on phylogenetic distance.

3. Construct general frameworks for integrating protein sequence and structural similarity metrics with phylogenetic comparisons, so that in the longer-term, we might perform these types of analyses for other proteins in a high-throughput and reproducible fashion.

The approach

We used the PPK1 protein from Accumulibacter as a query to compare sequence and structural similarity to all other PPK1 proteins retrieved from UniProt. To assess how phylogenetic distance connects to both sequence and structural similarity, we inferred a phylogeny of PPK1 sequences from Pseudomonadota, the phylum in which Accumulibacter is classified. From this tree, we calculated the patristic (i.e. phylogenetic) distance and compared it among protein sequences and structures. By comparing phylogenetic distance to protein sequence and structural similarity, we sought to find proteins that were highly similar in structure (and presumably function), yet highly evolutionarily distant from the Accumulibacter PPK1. Species with such proteins may have thus convergently evolved the ability to accumulate polyP.

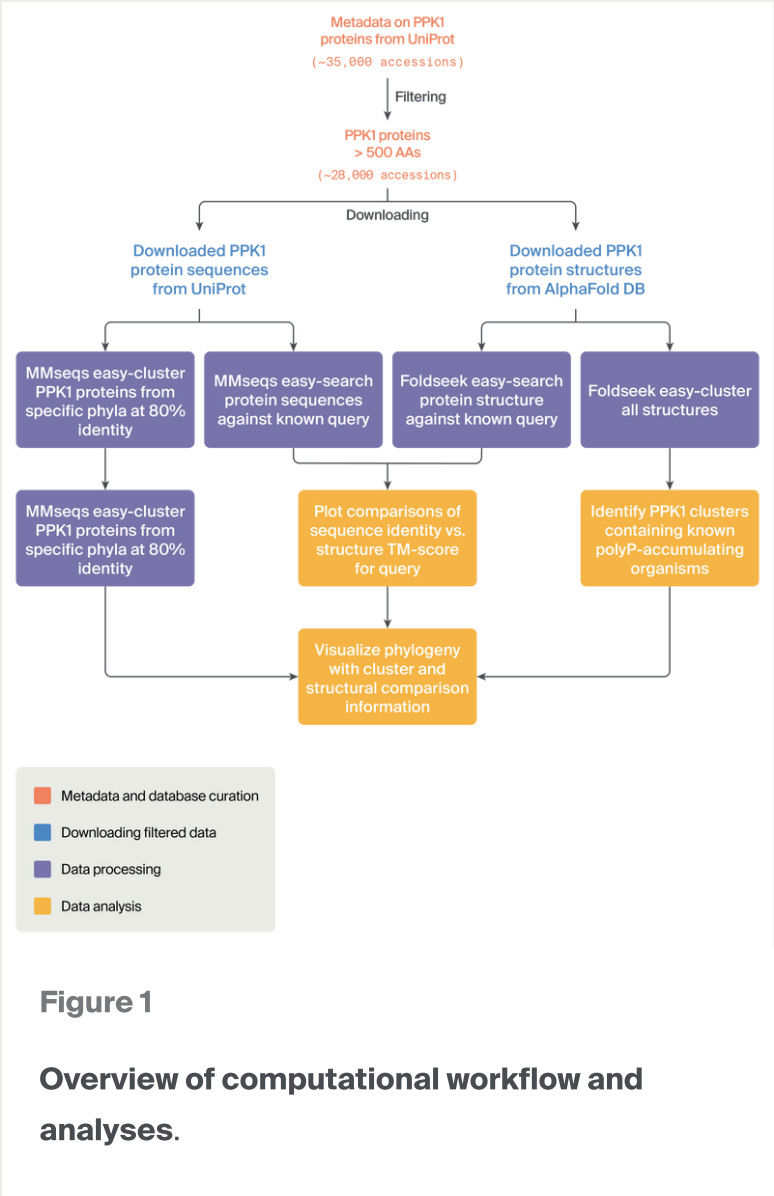

Figure 1

Overview of computational workflow and analyses.

Metadata and database curation

We first collected metadata for approximately 35,000 accessions annotated as PPK1 in bacteria and archaea in UniProt (Figure 1). This included information about protein length, assigned functional annotation, and taxonomic information for the organism. We then selected all proteins larger than 500 amino acids (AAs) to filter out short proteins such as incomplete clone sequences or incorrectly annotated sequences. We chose this filter based on plotting the distribution of protein lengths from all PPK1 entries from UniProt, and a length of greater than 500 AAs was sufficient to remove incorrectly annotated proteins or short clone sequences. This resulted in approximately 28,000 accessions that we were confident were annotated as PPK1. We curated metadata with the tidyverse R package (version 2.0) **[20]**. For each accession,

we downloaded the protein sequence from UniProt and the protein structure from the AlphaFold database (version 4) **[21]**. We've provided a TSV file of the metadata for the resulting ~28,000 accessions and gathered protein sequences and structures in this Zenodo archive **[22]**.

> **SHOW ME THE DATA**: You can access all the PPK1 protein sequences, structures, and metadata that we used, plus the MMseqs2 and Foldseek results, result tables, and files for phylogenetic inference on Zenodo (DOI: 10.5281/zenodo.8378182).

Preprocessing PPK1 protein sequences and structures

Since Accumulibacter is a hallmark polyphosphate-accumulating organism in wastewater, we wanted to compare all PPK1 protein sequences and structures to the Accumulibacter PPK1. We used the PPK1 protein (UniProt accession A0A369XMZ4) from the *Candidatus* Accumulibacter phosphatis UW-LDO-IC strain, which is now reclassified as *Candidatus* Accumulibacter meliphilus UW-LDO **[23][24]** (GenBank genome accession GCA_003332265.1). First, we clustered all PPK1 structures using Foldseek (version 6.29) with `foldseek easy-cluster` **[25]** within the ProteinCartography pipeline **[2]**. We then created a Nextflow workflow that runs both `mmseqs easy-search` with MMseqs (version 14.7) **[26]** and `foldseek easy-search` that performs all-v-all pairwise sequence and structure comparisons for all PPK1 sequences or structures against the Accumulibacter PPK1 and plots the results.

Data analysis and visualization

We used results from `mmseqs easy-search` and `foldseek easy-search` to plot the comparison of protein sequence similarity to TM-score for all PPK1 proteins against the Accumulibacter PPK1 using the R packages tidyverse (version 2.0) and ggpubr (version 0.6.0) **[27]**. TM-score is a metric for measuring the topological similarity of two protein structures, where scores range from 0–1 and a score of 1 is a perfect match between the two structures **[28]**. We plotted and overlaid pairwise comparisons of

protein sequence similarity and structural similarity for each PPK1 query compared to the Accumulibacter PPK1 with the corresponding phylum as the color.

For highlighting specific comparisons to the Accumulibacter PPK1 structure, we used the notebook `explore-ppk1-structures.ipynb` to visualize the alignment of two protein structures with Biopython (version 1.81) **[29]** and the py3Dmol (version 2.0.1) package **[30]** using PDB files as inputs. For each comparison, we took screenshots of the structure alignment from the notebook.

To investigate the phylogenetic distribution of sequences within the Pseudomonadota phylum (in which Accumulibacter is classified), we inferred a phylogenetic tree of a reduced set of Pseudomonadota PPK1 sequences. To obtain this reduced set of PPK1 sequences, we first clustered sequences at 80% identity using `mmseqs easy-cluster` , appending PPK1 sequences for Accumulibacter, *Neisseria gonorrhoeae* strain ATCC 700825 [Q5FAJ0], *Pseudomonas aeruginosa* strain ATCC 15692 [P0DP44], *Acinetobacter baumannii* 83444 [A0A829RFS7], and *Ralstonia solanacearum* strain UW386 [A0A5B7U1Z3]. We also included an outgroup PPK1 sequence from *Streptomyces coelicolor* to root the tree. We created an alignment of approximately 1,500 sequences with MUSCLE (version 5.1) **[31]** and a phylogenetic tree inferred with FastTree 2 (version 2.1.11) **[31]**.

We inspected and rooted the tree using iTOL **[32]**, and visualized in Empress v1.2.0 **[33]**. In the HTML viewer of Empress, we added two metadata rings for each representative sequence to show sequence similarity and structure similarity (TM-score) for each query compared to the Accumulibacter PPK1. Finally, we compared phylogenetic distance for these representative sequences to pairwise sequence identity and TM-score compared to Accumulibacter PPK1. We read the tree in Newick format into R using the ape package (version 5.7) **[34]**, calculated the patristic distance (sum of branch lengths between two terminal branches and their common ancestor node) with the adephylo package (version 1.11) **[35]**, and plotted into an interactive HTML plot with Plotly (version 4.10.2) **[36]**.

Additional methods

We used ChatGPT to write and clean up code. We also used it to suggest wording ideas, then we picked which parts to use.

All the **code** we generated and used for the pub is available in this GitHub repository (DOI: 10.5281/zenodo.8412197), including a workflow for making protein sequence and structural comparisons to a query, a Jupyter notebook for overlaying structures, and visualization scripts.

The results

SHOW ME THE DATA: You can access all the PPK1 protein sequences, structures, and metadata that we used, plus the MMseqs2 and Foldseek results, result tables, and files for phylogenetic inference on Zenodo.

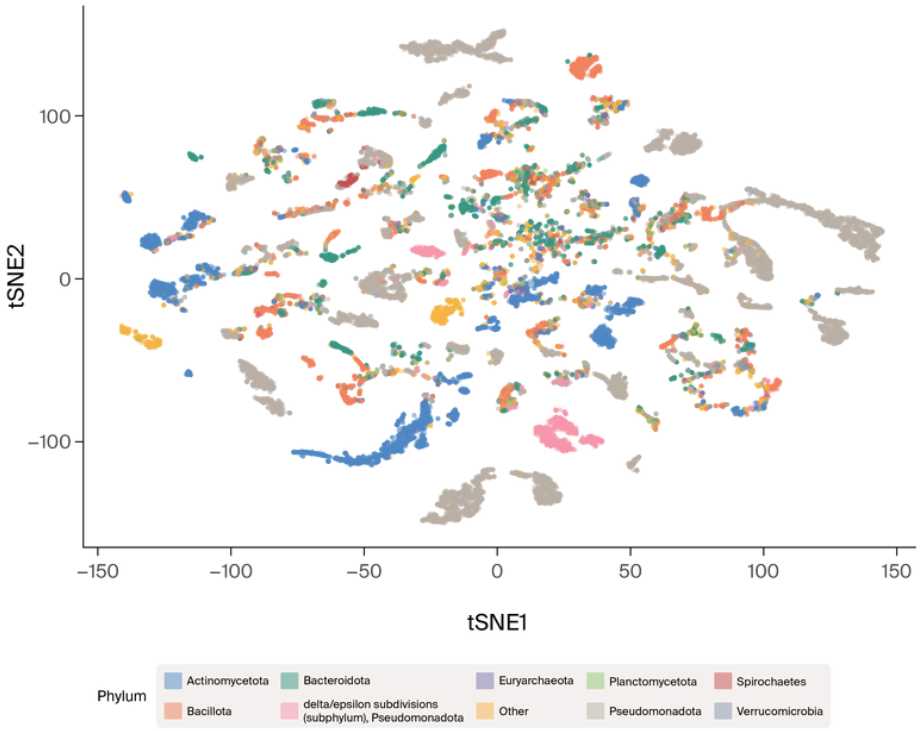

Figure 2

Clustering of all PPK1 structures using `foldseek easy-cluster` and plotted in two-dimensional space with TSNE.

Points are colored by phylum provided with the UniProt metadata, where only the top nine most frequent phyla are colored and all other phyla are represented as “other.”

We sought to test the hypothesis that phosphate-polymerizing PPK1 enzymes from bacteria that we know to be effective polyP accumulators have more similar protein structures than expected given their sequence divergence. If supported this hypothesis would suggest that we may predict whether uncharacterized species accumulate high levels of polyP. We predicted that we’d find proteins with divergent sequences that are still structurally similar to the Accumulibacter PPK1 protein.

We first clustered all ~28,000 PPK1 structures and labeled the clusters with phylum information (Figure 2). We inspected clusters that contain Accumulibacter PPK1 structures: SC59, SC21, SC13. We found a few proteins within those clusters that have high TM-scores (i.e. their structures are very similar to the Accumulibacter PPK1), but which come from other phyla. These include *Nitrospira sp.* [A0A3C1Z3C9], *Gemmatimonadetes sp*. [A0A7Y2B3S7] and *Methanomassiliicoccus sp.* [A0A847T1M7]

(compare their structures in Figure 3). We were encouraged that the first two taxa are bacterial lineages that are either important or abundant in wastewater and freshwater **[37][38]**. *Methanomassiliicoccus* spp. are methanogenic archaea important for anaerobic wastewater treatment processes and production of methane. It is still largely unknown how or if methanogenic archaea contribute to polyphosphate accumulation in wastewater even though they have the genetic potential **[12]**. PPK1 proteins from additional microbes cluster with the Accumulibacter PPK1, but we don't have data on their polyphosphate phenotypes. These results highlight that our approach could be useful in screening for candidate polyP-accumulating bacteria, which could then be verified through wet-lab experiments.

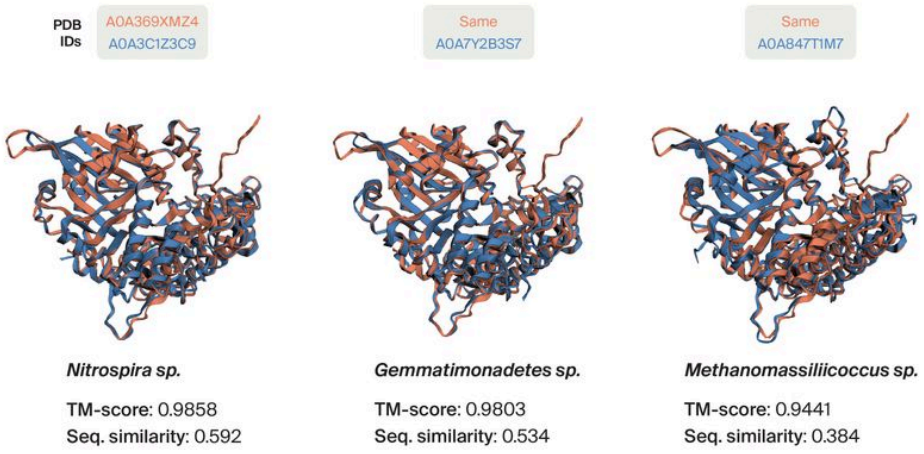

Figure 3

Structural comparisons of Accumulibacter PPK1 to PPK1 structures from other phyla that are significant in wastewater treatment processes.

Accumulibacter PPK1 structures are colored in orange and query structures in blue.

We were also interested in examples where proteins have high structural similarity but low sequence similarity, which could suggest convergent evolution of structure. Alternatively, this could suggest that structural similarity of PPK1 is dictated by local, rather than global sequence similarity. To explore this, we compared all PPK1 protein sequences and structures to our model phosphate polymerizing enzyme, the Accumulibacter PPK1 (Figure 4). We were reassured to find that all pairwise TM-score comparisons to the Accumulibacter PPK1 were 0.8 and above, as current practice is to

treat a TM-score above 0.5 as sufficient for inferring the same fold and assigning an annotation to a protein **[39]**. This high structural conservation of all queries is likely due to us prefiltering accessions greater than 500 AAs to ensure we made comparisons to correctly annotated PPK1 proteins.

As expected, the general trend is that with decreasing PPK1 sequence identity, protein structural alignment (represented by TM-score) also decreases. However, there is a plateau of decreasing protein sequence similarity but fairly high structural similarity, specifically for sequences within Pseudomonadota (Figure 4, grey points). This suggests that there are indeed proteins with similar protein structure despite dissimilar sequence composition.

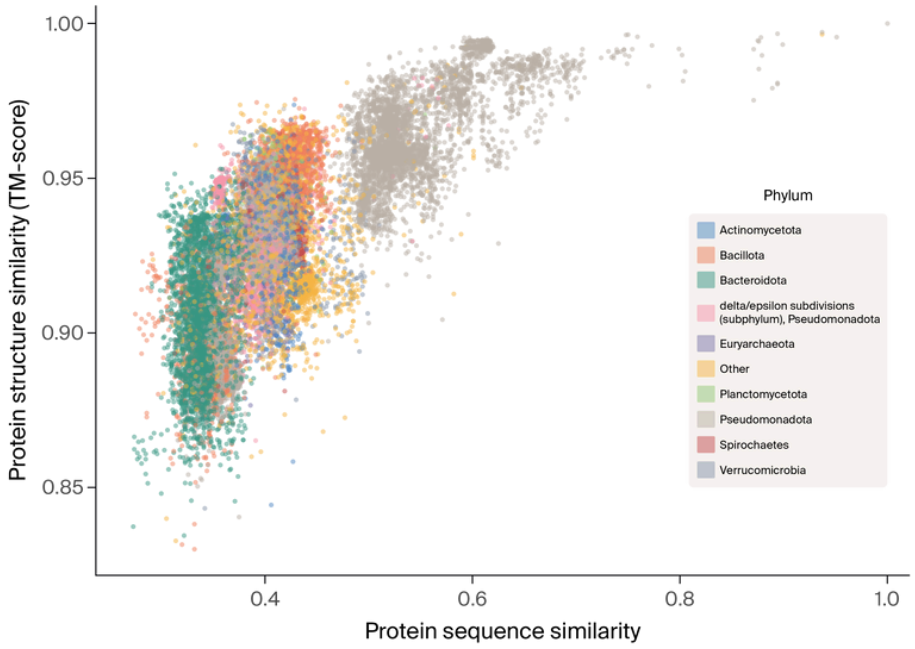

Figure 4

Pairwise all-v-all comparison of protein sequence and structural similarity (TM-score) to the Accumulibacter PPK1 reference protein.

We calculated pairwise protein sequence similarity against Accumulibacter PPK1 with `mmseqs easy-search` and calculated pairwise protein structure similarity against Accumulibacter PPK1 with `foldseek easy-search`. Colors for phylum match with Figure 2 and only the most frequent nine phyla are displayed, with all others represented as “other.” These phylum designations were directly pulled from UniProt — organisms within “Pseudomonadota, delta/epsilon subdivisions (subphylum)” were previously considered part of Deltaproteobacteria, and are sometimes now considered part of the overall “Pseudomonadota” phylum or other groups, are therefore grouped separately in UniProt.

To test if PPK1 structures convergently evolved among distantly related taxa, we inferred a tree for 1,500 representative Pseudomonadota PPK1 sequences. We overlaid the phylogenetic tree with each PPK1 TM-score compared to the Accumulibacter PPK1 and labeled a handful of organisms known to exhibit enhanced

polyP accumulation (Figure 5). We then used the phylogeny of PPK1 sequences to obtain the patristic distance among sequences, a measure of evolutionary distance defined as the sum of branch lengths separating two proteins in the tree. We compared the patristic distance to both the protein sequence identity and structure alignment to the Accumulibacter PPK1 (Figure 6). Unsurprisingly, there is a consistent decrease in protein sequence similarity as phylogenetic distance increases for all sequences compared to the Accumulibacter PPK1 (Figure 6). Notably, the shape of the pattern differs when we plot phylogenetic distance versus structural similarity (TM-score). That is, whereas sequence similarity drops off consistently with increasing phylogenetic distance before plateauing, protein structure is conserved at greater phylogenetic distances before eventually dropping off sharply (Figure 6). This aligns with the thinking that protein structures evolve slower and overall more conserved than protein sequences, but emphasizes a need for additional assessment of the extent to which we expect TM-score and sequence similarity to correspond.

Figure 5

Phylogenetic tree of representative Pseudomonadota PPK1 sequences.

We constructed this phylogenetic tree by first clustering all Pseudomonadota PPK1 sequences at 80% identity with `mmseqs easy-cluster` , aligning with MUSCLE, and constructing the tree with FastTree 2. We visualized the tree within Empress, where we made it ultrametric. The metadata inner ring represents pairwise structural similarity (TM-score) of the query protein to the Accumulibacter PPK1 structure, and the outer ring represents pairwise sequence similarity of the query protein to the Accumulibacter PPK1 sequence. We've highlighted specific examples of organisms within this phylum that are known to exhibit enhanced polyphosphate accumulation, and taxa colors match Figure 6.

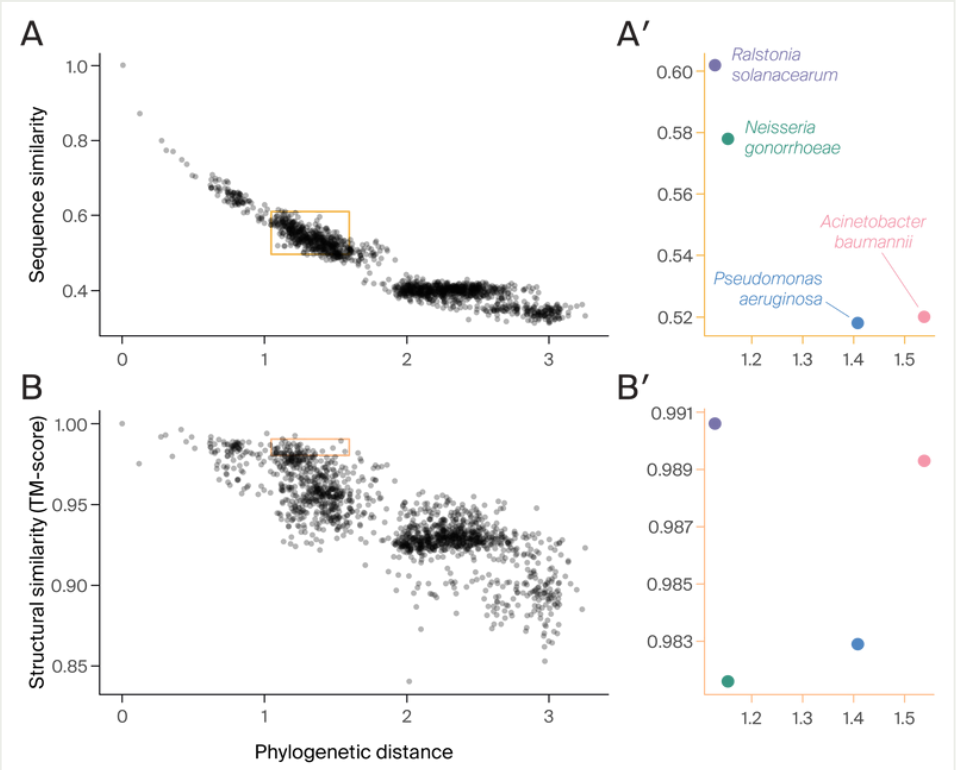

Figure 6

Comparisons of phylogenetic distance (patristic distance) versus protein sequence structure similarity for representative Pseudomonadota PPK1 proteins.

Colors of specific examples match those in Figure 5. Boxes in A and B correspond to approximate areas shown in A′ and B′.

Click to view an interactive version of this figure in a new tab. In the interactive, you can hover over a point to see the statistics and taxonomy for the organism.

Based on knowledge of human pathogens where polyphosphate accumulation is important for virulence and in looking at the results as a whole, the most striking data points were in *Neisseria gonorrhoeae* strain ATCC 700825 [Q5FAJ0], *Pseudomonas aeruginosa* strain ATCC 15692 [P0DP44], *Acinetobacter baumannii* 83444 [A0A829RFS7], and *Ralstonia solanacearum* strain UW386 [A0A5B7U1Z3] (Figure 5 and Figure 6), where each protein had a > 0.98 TM-score compared to the Accumulibacter PPK1. The first three organisms are human pathogens in which polyphosphate accumulation is linked to virulence. Some strains of *Neisseria gonorrhoeae* accumulate large amounts of polyphosphate granules on the exterior of

the cell into a pseudo-capsule and this is connected to human immune system evasion **[40]**. *Pseudomonas aeruginosa* causes infections in immunocompromised individuals, and *ppk1* knockouts lead to deficiencies in biofilm formation, motility, and quorum sensing **[41]**. *Acinetobacter baumannii* is a multi-drug resistant bacterium that causes nosocomial infections, and inhibition of PPK1 by repurposed drugs led to decreased biofilm formation, surface motility, and overall virulence **[42]**. *Ralstonia solanacearum* is a plant pathogen that causes bacterial wilt disease in crops like potatoes and tomatoes **[43]**, where biofilm formation, motility, and quorum sensing are important virulence factors for surviving in the nutrient-poor xylem of plants **[44][45]**.

Overall, these results highlight that this comparative approach to integrating protein structural predictions with phylogenetics could identify patterns of convergent evolution and functional importance across diverse bacterial lineages within the contexts of human health, agriculture, and biotechnological applications. Creating explicit statistical tests for correlating sequence and structural similarity and looking for phylogenetic outliers of this ratio will help us narrow down protein and species candidates for further validation.

Caveats

From these results, we've generated interesting hypotheses about the structural conservation of PPK1 across diverse bacteria, specifically in those that are known to accumulate large amounts of polyphosphate. Subsequent wet-lab experiments would be needed to validate whether protein structures with similar TM-scores indeed have similar activities or phenotypes related to polyphosphate accumulation, but this approach provides a starting place to test in the lab.

Interestingly, we did not find the same level of high similarity between PPK1 protein structures from Accumulibacter and *Tetrasphaera* spp. (average TM-score of 0.931 between five *Tetrasphaera* PPK1 proteins), even though these are the two main, experimentally verified bacterial lineages that contribute to polyphosphate accumulation in wastewater. If structural similarity and assessed PPK1 function were perfectly correlated, we would have expected that *Tetrasphaera* spp. would have the highest structural similarity to the Accumulibacter PPK1. However, the five *Tetrasphaera* spp. PPK1 proteins fell into the SC22, SC29, and SC39 clusters. Interestingly within these clusters also were important lineages in the wastewater treatment process such as other methanogenic archaeal lineages including

Methanomicrobiales, and several *Gemmatimonadetes* spp. Additionally, the *Tetrasphaera* clusters also contained several *Cyanobacteria* lineages such as the marine *Prochlorococcus, Synechococcus,* and *Leptolyngbya*. Although these lineages did not fall in the same clusters as Accumulibacter or have as much protein structure similarity to the Accumulibacter PPK1 as expected, this could suggest that several, different protein structures evolved and converged in different lineages that could be connected to increased polyphosphate accumulation under certain conditions.

Additionally, we restricted our analysis to comparisons of only the PPK1 protein, but PPK2 or copy number variation of PPK family proteins can contribute to enhanced polyphosphate accumulation, as they do in *Pseudomonas aeruginosa* **[46][47]**. Follow-up to this work could include co-clustering of PPK1 along with PPK2 for bacterial lineages that contain both to connect to polyphosphate accumulation phenotypes.

Key takeaways

- Querying ~28,000 PPK1 proteins against the Accumulibacter PPK1 resulted in highly similar comparisons to PPK1 protein structures in other lineages important in the wastewater treatment process and human pathogens where polyphosphate accumulation is an important virulence trait
- Searching for examples of high structural similarity of PPK1 proteins in distantly related taxa provided cases to test for potential convergent evolution of the protein structure
- More broadly, we can start connecting protein structure and phylogenetic comparisons together to generate more informed hypotheses about the evolutionary patterns of protein families, as well as harnessing novel or efficient protein functions that can be re-engineered for biotechnological applications.

Next steps

We believe that polyP accumulation and the PPK1 protein could be a good test case as we continue developing our platform, both computationally and in the lab. We could interrogate why certain proteins end up in certain structural clusters by performing domain analyses to look for common motifs within clusters. With more trait

information, we could start to compare PPK1 structures from high vs. low polyP-accumulating bacteria to identify key structural features required for efficient polyP formation.

As we build out our platform workflows, we are actively looking for proteins that are biologically interesting and allow for quick experimental validation of our computational predictions. Since there are many existing assays for quantifying polyphosphate in the lab **[48]**, we believe we could potentially build off our results with PPK to test subsequent *in silico* tools and eventually test hypotheses with wet-lab validation.

We're curious to hear what tools and approaches you'd like to see us explore next for connecting protein structure comparisons to phylogenetic metrics, and we're open to ideas for other proteins that could be better test cases for our development efforts.

References

1. Fernando EY, McIlroy SJ, Nierychlo M, Herbst F-A, Petriglieri F, Schmid MC, Wagner M, Nielsen JL, Nielsen PH. (2019). Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman–FISH. https://doi.org/10.1038/s41396-019-0399-7

2. Avasthi P, Bigge BM, Celebi FM, Cheveralls K, Gehring J, McGeever E, Mishne G, Radkov A, Sun DA. (2024). ProteinCartography: Comparing proteins with structure-based maps for interactive exploration. https://doi.org/10.57844/ARCADIA-A5A6-1068

3. Illergård K, Ardell DH, Elofsson A. (2009). Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. https://doi.org/10.1002/prot.22458

4. Rao NN, Gómez-García MR, Kornberg A. (2009). Inorganic Polyphosphate: Essential for Growth and Survival. https://doi.org/10.1146/annurev.biochem.77.083007.093039

5 Achbergerová L, Nahálka J. (2011). Polyphosphate - an ancient energy source and active metabolic regulator. https://doi.org/10.1186/1475-2859-10-63

6 Nocek BP, Khusnutdinova AN, Ruszkowski M, Flick R, Burda M, Batyrova K, Brown G, Mucha A, Joachimiak A, Berlicki Ł, Yakunin AF. (2018). Structural Insights into Substrate Selectivity and Activity of Bacterial Polyphosphate Kinases. https://doi.org/10.1021/acscatal.8b03151

7 Denoncourt A, Downey M. (2021). Model systems for studying polyphosphate biology: a focus on microorganisms. https://doi.org/10.1007/s00294-020-01148-x

8 Zhang H, Gómez-García MR, Shi X, Rao NN, Kornberg A. (2007). Polyphosphate kinase 1, a conserved bacterial enzyme, in a eukaryote, *Dictyostelium discoideum*, with a role in cytokinesis. https://doi.org/10.1073/pnas.0706847104

9 Gautam LK, Sharma P, Capalash N. (2021). Attenuation of Acinetobacter baumannii virulence by inhibition of polyphosphate kinase 1 with repurposed drugs. https://doi.org/10.1016/j.micres.2020.126627

10 Shahbaaz M, Nkaule A, Christoffels A. (2019). Designing novel possible kinase inhibitor derivatives as therapeutics against Mycobacterium tuberculosis: An in silico study. https://doi.org/10.1038/s41598-019-40621-7

11 Neville N, Roberge N, Ji X, Stephen P, Lu JL, Jia Z. (2021). A Dual-Specificity Inhibitor Targets Polyphosphate Kinase 1 and 2 Enzymes To Attenuate Virulence of Pseudomonas aeruginosa. https://doi.org/10.1128/mbio.00592-21

12 Paula FS, Chin JP, Schnürer A, Müller B, Manesiotis P, Waters N, Macintosh KA, Quinn JP, Connolly J, Abram F, McGrath JW, O'Flaherty V. (2019). The potential for polyphosphate metabolism in Archaea and anaerobic polyphosphate formation in Methanosarcina mazei. https://doi.org/10.1038/s41598-019-53168-4

13 Schindler DW. (1977). Evolution of Phosphorus Limitation in Lakes. https://doi.org/10.1126/science.195.4275.260

14 Seviour RJ, Mino T, Onuki M. (2003). The microbiology of biological phosphorus removal in activated sludge systems. https://doi.org/10.1016/s0168-6445(03)00021-4

15 Otieno J, Kowal P, Mąkinia J. (2022). The Occurrence and Role of Tetrasphaera in Enhanced Biological Phosphorus Removal Systems. https://doi.org/10.3390/w14213428

16 Kristiansen R, Nguyen HTT, Saunders AM, Nielsen JL, Wimmer R, Le VQ, McIlroy SJ, Petrovski S, Seviour RJ, Calteau A, Nielsen KL, Nielsen PH. (2012). A

metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological phosphorus removal. https://doi.org/10.1038/ismej.2012.136

17 Breiland AA, Flood BE, Nikrad J, Bakarich J, Husman M, Rhee T, Jones RS, Bailey JV. (2018). Polyphosphate-Accumulating Bacteria: Potential Contributors to Mineral Dissolution in the Oral Cavity. https://doi.org/10.1128/aem.02440-17

18 Rijal R, Cadena LA, Smith MR, Carr JF, Gomer RH. (2020). Polyphosphate is an extracellular signal that can facilitate bacterial survival in eukaryotic cells. https://doi.org/10.1073/pnas.2012009117

19 Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. (2019). Welcome to the Tidyverse. https://doi.org/10.21105/joss.01686

20 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. (2021). Highly accurate protein structure prediction with AlphaFold. https://doi.org/10.1038/s41586-021-03819-2

21 McDaniel E. (2023). Discovering shared protein structure signatures connected to polyphosphate accumulation in diverse bacteria. https://doi.org/10.5281/ZENODO.8378182

22 Petriglieri F, Singleton CM, Kondrotaite Z, Dueholm MKD, McDaniel EA, McMahon KD, Nielsen PH. (2022). Reevaluation of the Phylogenetic Diversity and Global Distribution of the Genus "*Candidatus*Accumulibacter." https://doi.org/10.1128/msystems.00016-22

23 Camejo PY, Oyserman BO, McMahon KD, Noguera DR. (2019). Integrated Omic Analyses Provide Evidence that a "Candidatus Accumulibacter phosphatis" Strain Performs Denitrification under Microaerobic Conditions. https://doi.org/10.1128/msystems.00193-18

24 van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. (2023). Fast and accurate protein structure search with Foldseek. https://doi.org/10.1038/s41587-023-01773-0

25 Steinegger M, Söding J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. https://doi.org/10.1038/nbt.3988

26 Kassambara A (2023). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0, https://rpkgs.datanovia.com/ggpubr/.

27 Zhang Y, Skolnick J. (2004). Scoring function for automated assessment of protein structure template quality. https://doi.org/10.1002/prot.20264

28 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. https://doi.org/10.1093/bioinformatics/btp163

29 Rego N, Koes D. (2014). 3Dmol.js: molecular visualization with WebGL. https://doi.org/10.1093/bioinformatics/btu829

30 Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. https://doi.org/10.1371/journal.pone.0009490

31 Letunic I, Bork P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. https://doi.org/10.1093/nar/gkab301

32 Cantrell K, Fedarko MW, Rahman G, McDonald D, Yang Y, Zaw T, Gonzalez A, Janssen S, Estaki M, Haiminen N, Beck KL, Zhu Q, Sayyari E, Morton JT, Armstrong G, Tripathi A, Gauglitz JM, Marotz C, Matteson NL, Martino C, Sanders JG, Carrieri AP, Song SJ, Swafford AD, Dorrestein PC, Andersen KG, Parida L, Kim H-C, Vázquez-Baeza Y, Knight R. (2021). EMPress Enables Tree-Guided, Interactive, and Exploratory Analyses of Multi-omic Data Sets. https://doi.org/10.1128/msystems.01216-20

33 Paradis E, Schliep K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. https://doi.org/10.1093/bioinformatics/bty633

34 Jombart T, Balloux F, Dray S. (2010). *adephylo*: new tools for investigating the phylogenetic signal in biological traits. https://doi.org/10.1093/bioinformatics/btq292

35 Plotly Technologies Inc. (2015). Collaborative data science. https://plot.ly

36 Mehrani M-J, Sobotka D, Kowal P, Ciesielski S, Makinia J. (2020). The occurrence and role of Nitrospira in nitrogen removal systems. https://doi.org/10.1016/j.biortech.2020.122936

37 Mujakić I, Piwosz K, Koblížek M. (2022). Phylum Gemmatimonadota and Its Role in the Environment. https://doi.org/10.3390/microorganisms10010151

38 Manca B, Buffi G, Magri G, Del Vecchio M, Taddei AR, Pezzicoli A, Giuliani M. (2023). Functional characterization of the gonococcal polyphosphate pseudo-capsule. https://doi.org/10.1371/journal.ppat.1011400

39 Rashid MH, Rumbaugh K, Passador L, Davies DG, Hamood AN, Iglewski BH, Kornberg A. (2000). Polyphosphate kinase is essential for biofilm development, quorum sensing, and virulence of *Pseudomonas aeruginosa*. https://doi.org/10.1073/pnas.170283397

40 Peeters N, Guidot A, Vailleau F, Valls M. (2013). <*scp*>*R*</*scp*>*alstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era. https://doi.org/10.1111/mpp.12038

41 Lowe-Power TM, Khokhani D, Allen C. (2018). How Ralstonia solanacearum Exploits and Thrives in the Flowing Plant Xylem Environment. https://doi.org/10.1016/j.tim.2018.06.002

42 Kang Y, Liu H, Genin S, Schell MA, Denny TP. (2002). *Ralstonia solanacearum* requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. https://doi.org/10.1046/j.1365-2958.2002.03187.x

43 Zhang H, Ishige K, Kornberg A. (2002). A polyphosphate kinase (PPK2) widely conserved in bacteria. https://doi.org/10.1073/pnas.262655199

44 Christ JJ, Willbold S, Blank LM. (2020). Methods for the Analysis of Polyphosphate in the Life Sciences. https://doi.org/10.1021/acs.analchem.9b05144