Published on May 30, 2024 by Arcadia Science **DOI**: 10.57844/arcadia-1e5d-e272

How can we biochemically validate protein function predictions with the deoxycytidine kinase family?

The human deoxycytidine kinase, a member of the nucleoside salvage pathway, has been studied extensively. We'll use this family to assess our structure-based protein clustering tool, ProteinCartography. We'd love feedback on how we might work with this protein for validation.

Contributors (A-Z)

Prachee Avasthi, Audrey Bell, Brae M. Bigge, Megan L. Hochstrasser, Atanas Radkov , Dennis A. Sun, Harper Wood, Ryan York

Version 1 · Mar 31, 2025

Purpose

We created ProteinCartography to computationally compare protein structures from a single family across many different species **[1]**. ProteinCartography identifies proteins similar to an input and compares the structures of each protein to every other protein to produce an interactive map with clustering information overlaid. In a previous pub,

we began formulating a plan to validate ProteinCartography by testing two foundational hypotheses: proteins within clusters will have similar functions and proteins in different clusters will have different functions **[2]**.

In this pub, we outline our ProteinCartography results for one of the protein families we've chosen to use for validation, deoxycytidine kinases, which we selected because it's been previously biochemically studied and produced results with many clear options for how to test our hypotheses **[2]**.

We're seeking feedback regarding how we might approach in-lab validation in this family, especially from those who've previously worked with deoxycytidine kinase proteins.

- This pub is part of the **platform effort**, "Functional annotation: mapping the functional landscape of proteins across biology." Visit the platform narrative for more background and context.
- This pub is part of our **validation strategy series** of pubs that starts with "A strategy to validate protein function predictions *in vitro*." We're also considering Ras GTPases as an orthogonal protein family for validation. To learn more about them, visit the accompanying pub **[3]**.
- The **ProteinCartography pipeline** used to run these analyses is available in this GitHub repo. To create the custom overlays, we used this notebook and added our custom color dictionaries, which can be found in the associated Zenodo repositories.
- The **data** associated with this pub, including ProteinCartography results for the deoxycytidine kinase family, can be found in this Zenodo repository.

Background

Why use deoxycytidine kinases?

Our initial validation of ProteinCartography is intended to test the two foundational hypotheses that proteins in the same cluster have similar structures and functions and that proteins in different clusters have differing structures and functions. To do this rapidly and in a straightforward manner, we began with proteins that had been previously biochemically characterized. We started with the 200 most well-studied human proteins **[4]**. Other factors we considered in our protein selection decision were the length of proteins and the quality of the available AlphaFold structures. The pLDDT (predicted local distance difference test), computed by AlphaFold, is a per-residue measure of the confidence of a model structure **[5]**. This score ranges from 0 to 100, with higher scores indicating greater confidence. In our case, we focused on proteins shorter than 1,280 amino acids, a length limit set by AlphaFold, and proteins with a pLDDT score higher than 80. Model structures in this pLDDT score range are typically considered high-confidence.

Taking into account each of our selection criteria **[2]**, we chose to focus on the human deoxycytidine kinase. As of this writing, there are 47 Protein Data Bank (PDB) entries for this protein, which places it among the 200 human proteins with the most solved structures. Additionally, this protein family has commercially available assay kits and it produced ProteinCartography results with clearly defined clusters that would allow us to test our foundational hypotheses (Figure 1).

What do deoxycytidine kinases do and why are they important?

Deoxycytidine kinase (dCK) has an essential role as a nucleoside kinase, critical in producing precursors for DNA synthesis **[6]**. The enzyme is crucial in the nucleoside salvage pathway, primarily phosphorylating deoxycytidine and converting it into deoxycytidine monophosphate **[7]**. The enzyme can also convert the nucleosides deoxyadenosine and deoxyguanosine to their monophosphate forms, albeit at a lower rate **[8]**. In addition to these native substrates, the dCK enzyme is essential for

activating several nucleoside analog prodrugs via phosphorylation. These analogs include anticancer drugs (cytarabine, gemcitabine, cladribine, and fludarabine) as well as antiviral drugs (lamivudine and emtracitabine) **[6]**.

Very little is known about non-human dCK homologs but they're intriguing to investigate because they could have distinct properties that might improve cancer and antiviral therapies that rely on human dCK. There's already evidence that novel human dCK homologs improve the efficacy of gene-directed enzyme prodrug therapies for cancer **[9]**. For example, a nucleoside kinase encoded by the fruit fly *Drosophila melanogaster* has broader substrate specificity, better catalytic efficiency, and improved stability **[10]** relative to its human counterpart. A truncated version of the fruit fly dCK successfully re-sensitized a drug-resistant breast cancer cell line to treatment with an anticancer nucleoside analog **[10]**. Another example is a tomato (*Solanum lycopersicum*) thymidine kinase that is highly active and less sensitive to negative feedback regulation by its reaction products **[11]**. Researchers used a combination of an anti-cancer prodrug and the tomato thymidine kinase to successfully treat malignant glioma (brain tumor) cells *in vitro* and brain tumors in mice **[12]**.

Diving into the ProteinCartography results for the deoxycytidine kinase family

Running ProteinCartography on deoxycytidine kinases

To explore the biochemical function of non-human dCK homologs, we used the ProteinCartography pipeline to find proteins that are structurally similar to the human dCK protein and group them into clusters based on that similarity. ProteinCartography uses BLAST and Foldseek to identify proteins similar to the input **[13][14]**. It compares the structures of each protein to every other protein to produce TM-scores, or structural similarity scores where a "one" indicates identical proteins **[15]**. Using these scores, the pipeline performs Leiden clustering to separate similar proteins into

clusters and reduces dimensionality to create interactive UMAP and t-SNE projections with overlays for further exploring the protein family **[16][17][18]**.

In our analysis, we used “search mode” with standard parameters and with the human dCK structure as input (UniProt ID: P27707). We requested 3,000 Foldseek hits and 7,000 BLAST hits — a total of 10,000 structures. Our run generated 2,418 unique structure hits that grouped into 12 clusters (LC00–LC11) (Figure 1). Our input protein, human dCK, is in LC04 (Figure 1 and Figure 2, A).

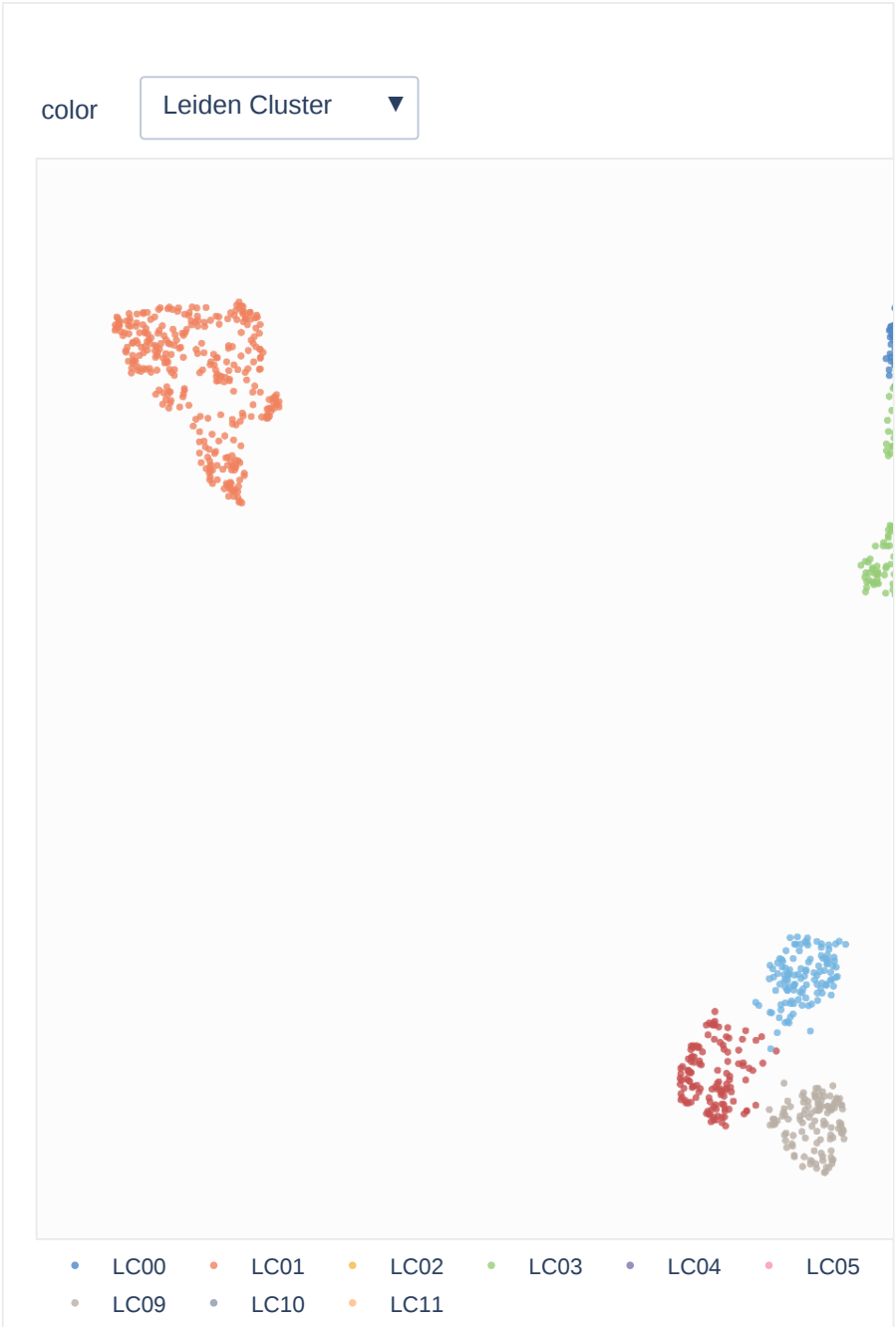


Figure 1

Interactive protein space with metadata overlays for proteins similar to human dCK.

UMAP generated by ProteinCartography for proteins identified as similar to the human dCK. Our input protein (human dCK) is in LC04, indicated by a four-pointed star. You can select different overlays via the drop-down “color” menu.

A full list of all the proteins in this analysis, plus all the aggregated information from the pipeline can be found in the aggregated features file linked below:

tsv Deoxycytidine_kinase_aggregated_features_pca_umap.tsv Download

Assessing compactness and overall quality

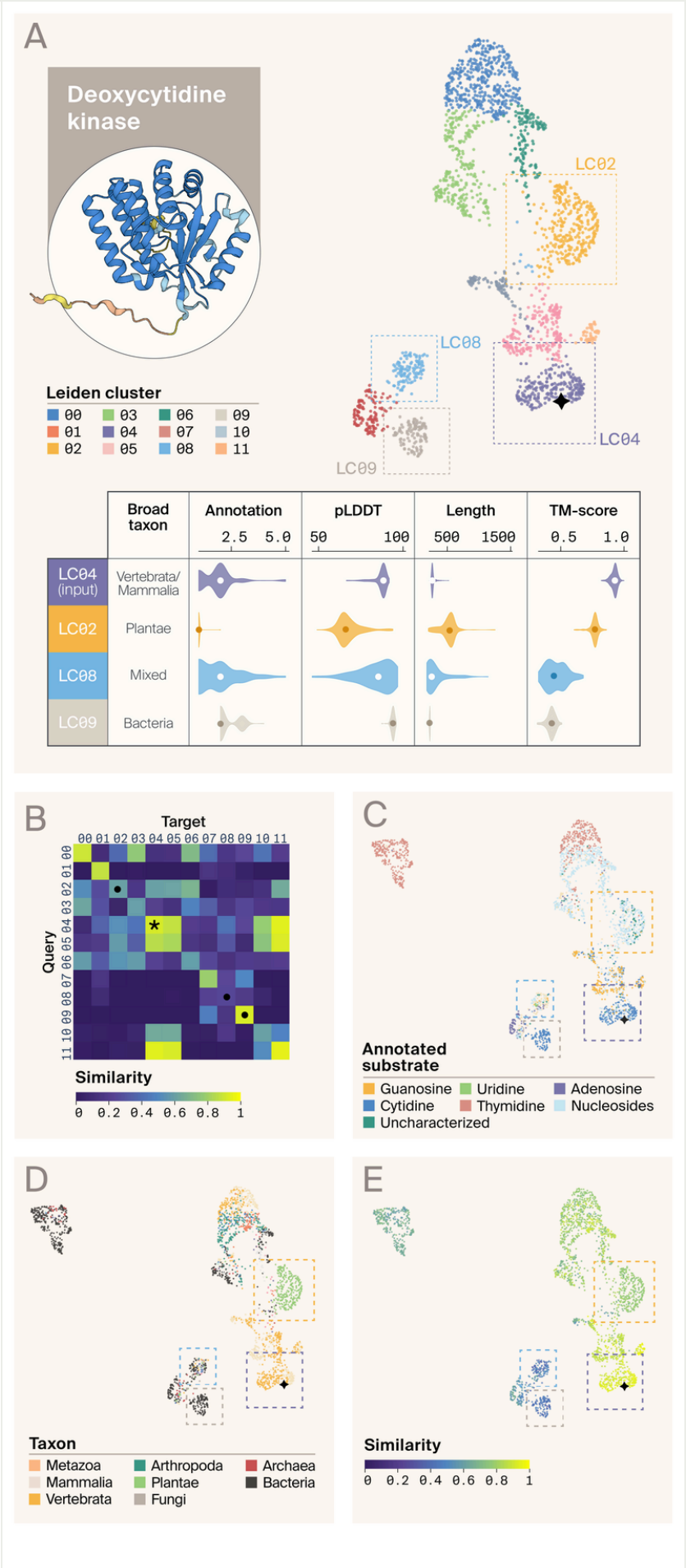
We started our analysis by exploring the Leiden cluster similarity matrix (Figure 2, B) to evaluate the quality of the protein space ProteinCartography generated. The similarity matrix displays scores calculated by comparing the mean TM-score of every structure in each cluster to every other structure in the analysis **[1]**. By looking at the similarity scores along the diagonal of the matrix, we get an idea of how tightly grouped the proteins are within each individual cluster. The average of the diagonal values is a measure we've previously described as "cluster compactness" **[1]**. The clusters in our analysis had a mean compactness score of 0.73 (average of the diagonal values in the similarity matrix). Most of the individual clusters also appear compact (a score above 0.6), in particular LC04 (score: 0.91; cluster with our input protein), LC09 (score: 0.92), and LC11 (score: 0.94) had some of the highest compactness scores (Figure 2, B). Cluster compactness represents a basic quality-control check of how well the proteins have grouped. However, given its nonlinear relationship with a number of other ProteinCartography outputs, we decided to include several clusters with low compactness in our downstream analyses to better understand the utility of cluster compactness.

As a preliminary check of the quality of the structures, we explored the distribution of mean pLDDT scores (structural confidence) and TM-scores (structural similarity) across all clusters. The pLDDT scores tell us how confident the AlphaFold structural prediction is and often low scores point to disordered regions. A score of 100 is a highly confident structure **[5]**. The majority of the structures in our dCK analysis had a pLDDT score greater than 80, except for the structures in LC02, which we discuss further below (Figure 2, A). These high scores suggest that we can be confident in the accuracy of the structural predictions. When we looked at TM-scores, which tell us how similar two structures are to each other, we saw that some structures are very similar to the input protein (TM-scores close to one), but some structures are only distantly

related (TM-scores between 0.4 and 0.5) (Figure 1 and Figure 2, A). The broad spectrum of relatedness represented enables us to more thoroughly investigate the relationship between structural similarity and function.

Exploring the data

To better understand the composition of our clusters and guide our selection process, we explored ProteinCartography's metadata overlays (Figure 1 and Figure 2, A). The metadata that we found particularly interesting for our analysis shows the distribution of taxa (broad taxonomy overlay) (Figure 2, D), length of proteins (length overlay), TM-



scores (TM-score_v_input overlay) (Figure 2, E), pLDDT scores (pLDDT overlay), and UniProt annotation scores (annotation overlay), across all of the proteins in each Leiden cluster (Figure 1).

In the following subsections, we walk through the most interesting clusters.

> **SHOW ME THE DATA**: Our full ProteinCartography analysis for the deoxycytidine kinase family is in this Zenodo repository (DOI: 10.5281/zenodo.11288250).

LC04: How does our

Figure 2

ProteinCartography outputs reveal interesting clusters of proteins with structural similarity to human dCK.

(A) The structure of human dCK, where orange indicates regions of higher disorder, alongside the UMAP projection with Leiden cluster overlay. Black diamond indicates the input protein. Note that LC01 is cropped out. Below the projection are violin plots showing the distribution of key values for each of our clusters of interest where the circles indicate the median value. White dots mean the median is below the threshold for significance, while filled-in dots denote significance in a Mann–Whitney *U* test. “Broad taxon” indicates taxonomic groups represented in each cluster. “Annotation” is the UniProt annotation confidence score, (scale: 1–5). “pLDDT” is the confidence in the AlphaFold structural prediction for each structure (scale: 0–100). “Length” is the number of amino acids in each protein. “TM-score” is the similarity of each structure to that of human dCK (scale: 0–1).

(B) Cross-cluster similarity matrix. Each box represents the average TM-score (structural similarity) when comparing all structures in one cluster to all structures in another, where a higher score means the structures are more similar. The input cluster is marked with an asterisk (*) and our clusters of interest are marked with dots (•).

(C) UMAP projection with custom overlay showing existing gene annotations. We manually sorted annotations into seven major groups based on the nucleoside or nucleoside derivative they act on and created a custom color overlay.

(D) UMAP projection with broad taxonomic groups overlaid.

(E) UMAP projection with TM-scores (compared to the input protein) overlaid. Higher TM-scores indicate higher structural similarity to human dCK.

(A, C–E) Dashed boxes mark our clusters of interest.

input protein cluster?

We began by analyzing the metadata overlays for LC04, which contains our input protein, to see whether the results seem reliable and match what we would expect for the cluster containing the input protein. We started with the broad taxonomic group overlay. ProteinCartography assigns proteins into taxonomic groups that allow for the best readability, but the taxonomic depth is not uniform. Cluster LC04 contains two dominant taxonomic groups, mammals and other vertebrates. Because our input protein is a human protein, this is reasonable. The mean length of proteins in LC04 is ~270 amino acids, which is very close to our input protein (260 amino acids), and the mean TM-score is 0.9, indicating that the proteins in this cluster adopt a fold that's highly similar to our input protein (Figure 1 and Figure 2, A). The mean pLDDT score for proteins in LC04 is 87, which confirms that the quality of the structural predictions is high and that the proteins are generally well-structured (Figure 1 and Figure 2, A). Last, the most common annotation score in this cluster is two (132 proteins out of 233 total in LC04) followed by one (78 proteins) (Figure 1 and Figure 2, A), which both suggest that existing UniProt protein annotations are of low confidence. We often observe these two annotation scores as the most common because the majority of the proteins in the UniProt database have not been biochemically characterized. Overall, these results are fairly typical for a ProteinCartography run and there were no surprises, so we're reasonably confident that the pipeline worked as we'd hoped.

LC02: Plant homologs close in structure to human dCK

By exploring the taxon distribution across the other clusters in our analysis, we found that all proteins in LC02 are in the clade Viridiplantae (Figure 1; Figure 2, A; and Figure 2, D). The proteins in this cluster have a mean length that is much higher (512 amino acids) compared to our input protein (260 amino acids) (Figure 1 and Figure 2, A). Even though the proteins in LC02 have a slightly lower mean TM-score (0.8), they should still

adopt the same fold as our input protein **[19]** (Figure 1 and Figure 2, A). The extra length of the proteins in this cluster may contribute to their lower TM-score and lower mean pLDDT score of 67. We explored the structures of a few of the individual proteins and noticed that they all have a core region with a high pLDDT score (90) that structurally aligns well with our input protein. However, that core region is flanked by unstructured portions on both the N- and C-termini, which may also contribute to the low pLDDT score for the entire protein. Similar to LC04, almost all proteins in this cluster have an annotation score of one (317 proteins out of 321 total in LC02), indicating an overall poor quality of the annotations in this cluster (Figure 1 and Figure 2, A).

LC08 and LC09: Taxonomically diverse homologs that diverge in structure from human dCK

When we explored the broad taxonomy overlay for LC08 and LC09, we found that there are highly diverse taxa represented in LC08, including Vertebrata, Bacteria, Archaea, Viridiplantae, and Arthropoda, while LC09 contains exclusively bacterial proteins (Figure 1 and Figure 2, D). The proteins in LC08 are on average longer compared to our input protein (319 amino acids vs. 260 amino acids), and this cluster also contains some very long proteins (> 1,000 amino acids) (Figure 1 and Figure 2, A). The mean length of proteins in LC09 is very uniform and most proteins are shorter than our input protein (220 amino acids vs. 260 amino acids). Finally, both LC08 and LC09 show low mean TM-scores of 0.5 and 0.4, respectively, suggesting that the proteins in these clusters have adopted a fold that is more distantly related to our input protein (Figure 1 and Figure 2, E). For both clusters, the structure quality is high, with mean pLDDT scores of 83 and 93 for LC08 and LC09, respectively, and the vast majority of the proteins (74%) have an annotation score of one or two (Figure 1 and Figure 2, A), so their annotations are lower confidence.

Overlaying annotation data

In addition to all of the overlays that the ProteinCartography pipeline outputs automatically, we can also create custom overlays to display any metadata. We manually noted which type of deoxynucleoside or deoxynucleoside derivative each protein was annotated to act on in UniProt since we noticed that not all the proteins in

our maps are kinases that are annotated as proteins that act on deoxycytidine. We overlaid this annotation data onto our Leiden cluster map (Figure 2, C).

We were curious to see if proteins annotated as acting on the same substrate would cluster together, or if perhaps proteins with certain annotations would be distributed across multiple clusters. In the case of LC04, the vast majority of the proteins were annotated as dCK (deoxycytidine kinase), the same annotation as our input protein (Figure 2, C). For LC02, the most prevalent annotation was the general annotation, “deoxynucleoside kinase,” or dNK, which could mean these proteins act on several nucleosides or that this broad annotation was used because the substrate specificity was unknown (Figure 2, C). While LC08 contained very mixed annotations, all of the proteins in LC09 were annotated as acting primarily on cytosine or cytosine derivatives (Figure 2, C). In addition to overlaying the protein annotations across the Leiden cluster map, we used ProteinCartography to generate a semantic analysis of the annotations that provides a more granular view of their distribution throughout clusters (Figure 3). For example, we can see that while the input cluster LC04 is primarily annotated as “deoxycytidine kinase,” LC09 is primarily “cytidylate kinase” (Figure 3). Additionally, we can get more detail about the mixed annotations in LC08, and see that the primary annotations are “dephospho-CoA kinase,” “uridine kinase,” and “guanylate kinase” (Figure 3).

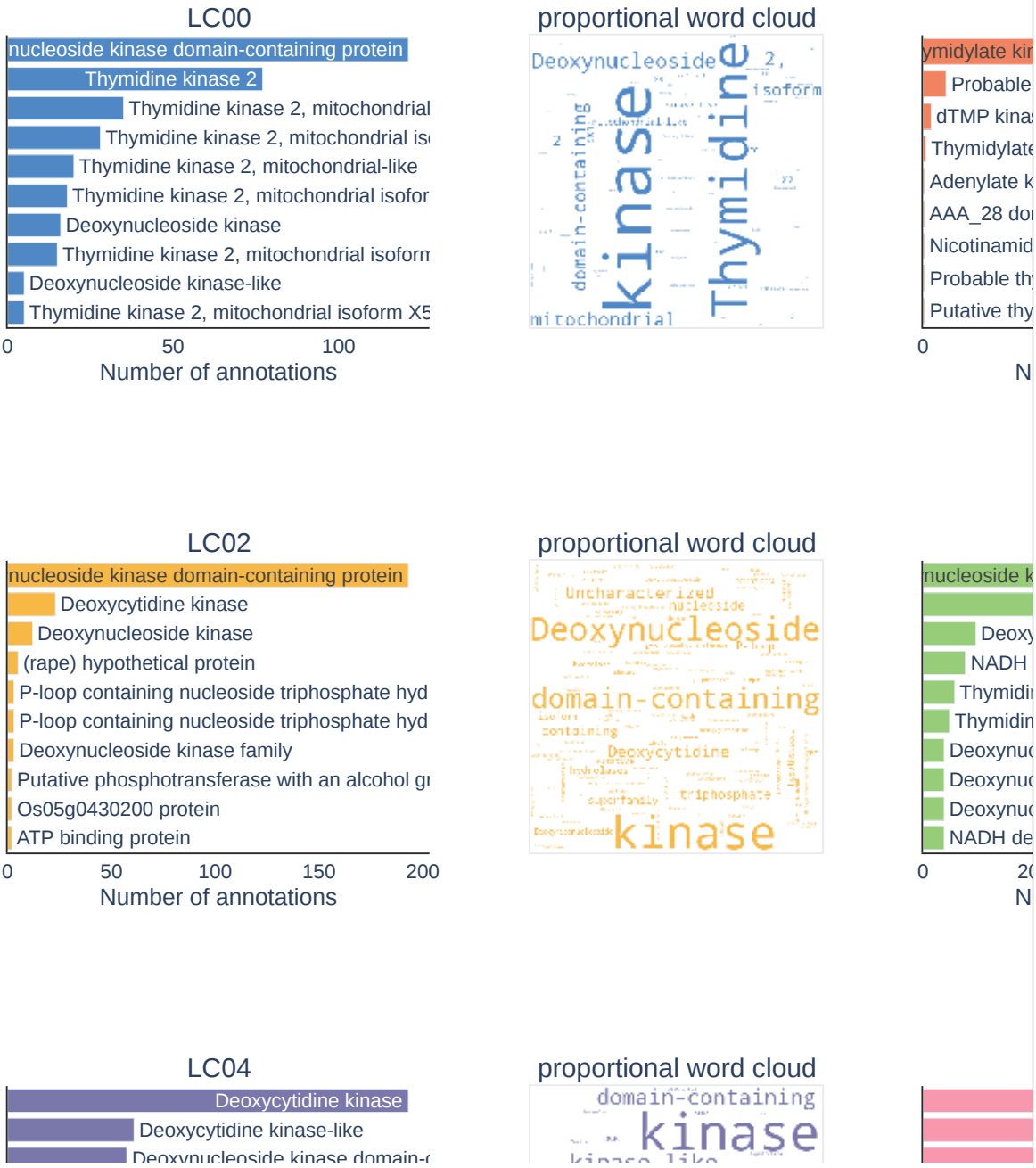


Figure 3

Interactive semantic analysis plot of the dCK proteins, showing the relationship between clustering and protein annotations.

The colors in the plot correspond to the colors of the Leiden clusters. For each cluster, the plot contains a ranked bar chart and a word cloud. The chart includes the top ten full annotation strings, while the word cloud shows the top annotation words. Hovering over a bar in the chart displays the full annotation string and the number of proteins with that annotation.

Summary

Aside from the cluster with our input protein, LC04, we find LC02, LC08, and LC09 most interesting because they contain proteins from diverse taxa and close, as well as distant, structural homologs of our input protein. We plan to use proteins from these clusters to test whether the two foundational hypotheses underlying ProteinCartography are accurate (that proteins with similar functions cluster together and those with dissimilar functions cluster separately), but we want to hear your thoughts!

What do you think?

Testing hypothesis 1: Do proteins within clusters function similarly?

Here are our ideas about how we might test this.

1. We could characterize uncharacterized proteins from the cluster containing our input protein to determine if they have the same function as the input protein (in LC04). Specifically, we plan to test the ability of proteins to phosphorylate deoxynucleoside substrates using ATP.
2. We could refine the current annotations of proteins that are annotated too broadly. In the cluster with our input protein, some proteins are annotated as the generic “deoxynucleoside kinase.” We could make this more specific by testing how these proteins interact with different substrates.

Do these seem like reasonable approaches to test this hypothesis?

Testing hypothesis 2: Do proteins in different clusters have different functions?

Here are the clusters we're considering to test this question. Each seems distinct in a different way, so we suspect that we'll find functional differences between proteins from these clusters and between these and our human input protein, which is in LC04.

1. LC02 contains exclusively plant proteins with an overall low quality of annotations. The proteins in this cluster are also longer than our input protein and contain a disordered region at each end. We could investigate whether there are functional differences between our input protein and proteins in LC02, which could be caused by the disordered region.
2. The proteins in LC08 span several distinct taxonomic clades and are only distantly related structural homologs of our input protein.
3. LC09 contains exclusively bacterial proteins that adopt a different fold from our input protein based on our structural comparisons.

1. Which of these clusters is your favorite for testing our hypothesis that proteins in different clusters have different functions?

A LC02

B LC08

C LC09

OK

How should we approach working with dCK proteins *in vitro*?

Once we select individual clusters and proteins, we'll bring them into the lab for biochemical characterization. We plan to purify each protein we select and test its ability to act on its possible substrates.

> Are there tips/tricks/challenges to biochemical analysis of dCK?

> Do you have ideas for functions of dCK that we might want to test other than or in addition to its activity as a deoxynucleoside kinase?

Additional methods

We used ChatGPT to help critique, clarify, and streamline text that we wrote.

Next steps

Now that we've selected deoxycytidine kinases as a protein family to test, we hope readers will provide feedback on the interesting clusters we identified and how to choose individual proteins for further analysis. Once selected, we'll bring these proteins into the lab for functional assays. We're planning to purify our selected proteins and run basic activity assays on each one.

While our biochemical efforts are in progress, we have a few additional computational ideas to gain insights into what we can learn from ProteinCartography clustering. We discuss these potential next steps below.

Align functional data in the literature with ProteinCartography clustering

While we plan to directly compare the function of diverse proteins from each family in our own hands, we might also be able to check our ProteinCartography clustering against empirical functional data in the literature. Do proteins with similar functional profiles cluster together? Do those known to work differently cluster apart?

This analysis should be doable, as several homologs of the human dCK enzyme have biochemical data available, including proteins from chicken **[20][21]**, frog **[20][21][22]**, worm **[23]**, arabidopsis **[24]**, fruit fly **[10][25]**, mosquito **[26]**, moth **[22]**, amoeba **[27]**, and bacteria **[28][29][30][31][32][33][34]**. There's also a review that summarizes the biochemical activity of enzymes from this family from multiple organisms **[35]**.

Learn more about clusters and individual proteins by studying specific, conserved structural features

We're broadly interested in leveraging comparative structural biology to annotate protein function. While ProteinCartography analyses rely on comparing the global protein structure, there are many other structure-based characteristics that we might consider in trying to predict function across protein families. Some of these features include secondary structural elements (like α-helices or β-sheets), surface area, hydrophobicity, electrostatics, topology, inter-protein contact networks, active sites, and potentially predicted binding sites. We're interested in comparing these features across proteins to provide more specific and accurate protein function predictions.

For example, if we start with the human dCK enzyme and determine the conservation of its structural features across many structural homologs, we may be able to predict with a higher accuracy which of these proteins have a similar function. We know that the human dCK enzyme acts not only on deoxycytidine (dC), but also on deoxyguanosine (dG) and deoxyadenosine (dA). Could we predict which other proteins act on these three nucleosides? Might we predict which proteins act on just one?

Summary

We hope that by combining our fold-based structural clustering, more specific information on structural features, and functional data from the literature, we can start to develop a more complete and predictive framework to understand protein function.

References

1. 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

2. Avasthi P, Bigge BM, Radkov A, Wood H, York R. (2024). A strategy to validate protein function predictions in vitro. https://doi.org/10.57844/ARCADIA-CAE9-96C4

3. Avasthi P, Bigge BM, Radkov A, Wood H, York R. (2024). How can we biochemically validate protein function predictions with the Ras GTPase family? https://doi.org/10.57844/ARCADIA-74AD-345F

4. Li Z, Buck M. (2021). Beyond history and “on a roll”: The list of the most well-studied human protein structures and overall trends in the protein data bank. https://doi.org/10.1002/pro.4038

5. Mariani V, Biasini M, Barbato A, Schwede T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. https://doi.org/10.1093/bioinformatics/btt473

6. Sabini E, Hazra S, Ort S, Konrad M, Lavie A. (2008). Structural Basis for Substrate Promiscuity of dCK. https://doi.org/10.1016/j.jmb.2008.02.061

7. Shewach D, Reynolds K, Hertel L. (1992). Nucleotide specificity of human deoxycytidine kinase. https://pubmed.ncbi.nlm.nih.gov/1406603/

8. Slot Christiansen L, Munch-Petersen B, Knecht W. (2015). Non-Viral Deoxyribonucleoside Kinases – Diversity and Practical Use.

https://doi.org/10.1016/j.jgg.2015.01.003

9 Munch-Petersen B, Piskur J, Søndergaard L. (1998). Four Deoxynucleoside Kinase Activities from Drosophila melanogaster Are Contained within a Single Monomeric Enzyme, a New Multifunctional Deoxynucleoside Kinase. https://doi.org/10.1074/jbc.273.7.3926

10 Larsen NB, Munch-Petersen B, Piškur J. (2014). Tomato Thymidine Kinase Is Subject to Inefficient TTP Feedback Regulation. https://doi.org/10.1080/15257770.2013.853781

11 Khan Z, Knecht W, Willer M, Rozpedowska E, Kristoffersen P, Clausen AR, Munch-Petersen B, Almqvist PM, Gojkovic Z, Piskur J, Ekstrom TJ. (2010). Plant thymidine kinase 1: a novel efficient suicide gene for malignant glioma therapy. https://doi.org/10.1093/neuonc/nop067

12 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

13 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. https://doi.org/10.1016/s0022-2836(05)80360-2

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

15 Traag VA, Waltman L, van Eck NJ. (2019). From Louvain to Leiden: guaranteeing well-connected communities. https://doi.org/10.1038/s41598-019-41695-z

16 Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE. (2019). Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. https://doi.org/10.1038/s41467-019-13055-y

17 McInnes L, Healy J, Saul N, Großberger L. (2018). UMAP: Uniform Manifold Approximation and Projection. https://doi.org/10.21105/joss.00861

18 Zhang Y. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. https://doi.org/10.1093/nar/gki524

19 Konrad A, Lai J, Mutahir Z, Piškur J, Liberles DA. (2014). The Phylogenetic Distribution and Evolution of Enzymes Within the Thymidine Kinase 2-like Gene Family in Metazoa. https://doi.org/10.1007/s00239-014-9611-6

20 Mutahir Z, Clausen AR, Andersson K, Wisen SM, Munch-Petersen B, Piškur J. (2013). Thymidine kinase 1 regulatory fine-tuning through tetramer formation.

https://doi.org/10.1111/febs.12154

21 Knecht W, Petersen GE, Munch-Petersen B, Piškur J. (2002). Deoxyribonucleoside kinases belonging to the thymidine kinase 2 (TK2)-like group vary significantly in substrate specificity, kinetics and feed-back regulation. https://doi.org/10.1006/jmbi.2001.5257

22 Skovgaard T, Uhlin U, Munch-Petersen B. (2012). Comparative active-site mutation study of human and *Caenorhabditis elegans* thymidine kinase 1. https://doi.org/10.1111/j.1742-4658.2012.08554.x

23 Clausen AR, Girandon L, Ali A, Knecht W, Rozpedowska E, Sandrini MPB, Andreasson E, Munch-Petersen B, Piškur J. (2012). Two thymidine kinases and one multisubstrate deoxyribonucleoside kinase salvage <scp>DNA</scp> precursors in <*scp*>*A*</*scp*>*rabidopsis thaliana*. https://doi.org/10.1111/j.1742-4658.2012.08747.x

24 Legent K, Mas M, Dutriaux A, Bertrandy S, Flagiello D, Delanoue R, Piskur J, Silber J. (2006). In Vivo Analysis of Drosophila Deoxyribonucleoside Kinase Function in Cell Cycle, Cell Survival and Anti-Cancer Drugs Resistance. https://doi.org/10.4161/cc.5.7.2613

25 Knecht W. (2003). Mosquito has a single multisubstrate deoxyribonucleoside kinase characterized by unique substrate specificity. https://doi.org/10.1093/nar/gkg257

26 Sandrini MPB, Söderbom F, Mikkelsen NE, Piškur J. (2007). Dictyostelium discoideum Salvages Purine Deoxyribonucleosides by Highly Specific Bacterial-like Deoxyribonucleoside Kinases. https://doi.org/10.1016/j.jmb.2007.03.053

27 Sandrini MPB, Clausen AR, On SLW, Aarestrup FM, Munch-Petersen B, Piškur J. (2007). Nucleoside analogues are activated by bacterial deoxyribonucleoside kinases in a species-specific manner. https://doi.org/10.1093/jac/dkm240

28 Carnrot C, Wehelie R, Eriksson S, Bölske G, Wang L. (2003). Molecular characterization of thymidine kinase from *Ureaplasma urealyticum*: nucleoside analogues as potent inhibitors of *mycoplasma* growth. https://doi.org/10.1046/j.1365-2958.2003.03717.x

29 Carnrot C, Vogel SR, Byun Y, Wang L, Tjarks W, Eriksson S, Phipps AJ. (2006). Evaluation of Bacillus anthracis thymidine kinase as a potential target for the development of antibacterial nucleoside analogs. https://doi.org/10.1515/bc.2006.196

30 Wang L, Westberg J, Bölske G, Eriksson S. (2001). Novel deoxynucleoside-phosphorylating enzymes in mycoplasmas: evidence for efficient utilization of

deoxynucleosides. https://doi.org/10.1046/j.1365-2958.2001.02700.x

31 Tinta T, Christiansen LS, Konrad A, Liberles DA, Turk V, Munch-Petersen B, Piškur J, Clausen AR. (2012). Deoxyribonucleoside kinases in two aquatic bacteria with high specificity for thymidine and deoxyadenosine. https://doi.org/10.1111/j.1574-6968.2012.02565.x

32 OKAZAKI R, KORNBERG, A. (1964). DEOXYTHYMIDINE KINASE OF ESCHERICHIA COLI. I. PURIFICATION AND SOME PROPERTIES OF THE ENZYME. https://pubmed.ncbi.nlm.nih.gov/14114853/

33 OKAZAKI, R, KORNBERG A. (1964). DEOXYTHYMIDINE KINASE OF ESCHERICHIA COLI. II. KINETICS AND FEEDBACK CONTROL. https://pubmed.ncbi.nlm.nih.gov/14114854/