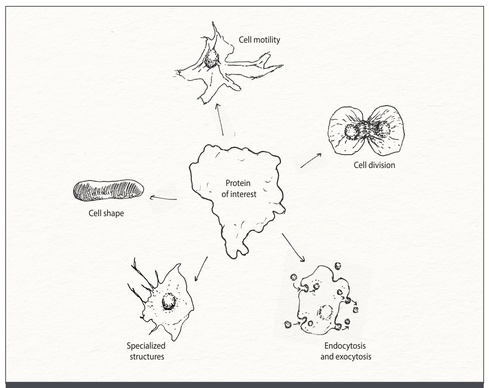
ProteinCartography: Mapping the functional landscape of proteins across biology

Presented by  Brae M Bigge 𝕏@BiggeBrae

ARCADIA

Background

How do we investigate protein families across species?



For more, read the full project narrative: **https://tinyurl.com/arcadiaannotation**

Exploring protein families before

Figure 2. BLAST results for human actin (P60709). Typical comparative protein analyses start with identifying similar proteins via protein BLAST [1].

Exploring protein families with ProteinCartography

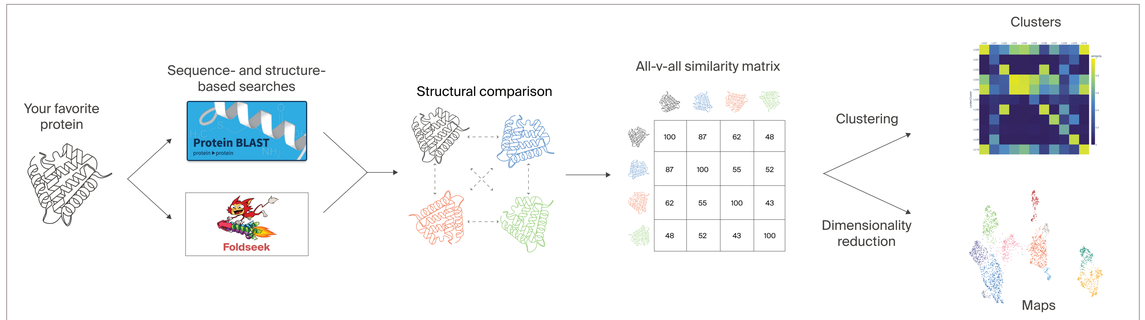


Figure 3. Overview of ProteinCartography. The pipeline searches for similar proteins using structure and function searches [1,3]. It then compares the structure of every hit to every other structure to create an all-v-all similarity matrix of structural similarity scores. It uses this matrix to cluster proteins and create interactive maps with metadata overlays [2.5,6].

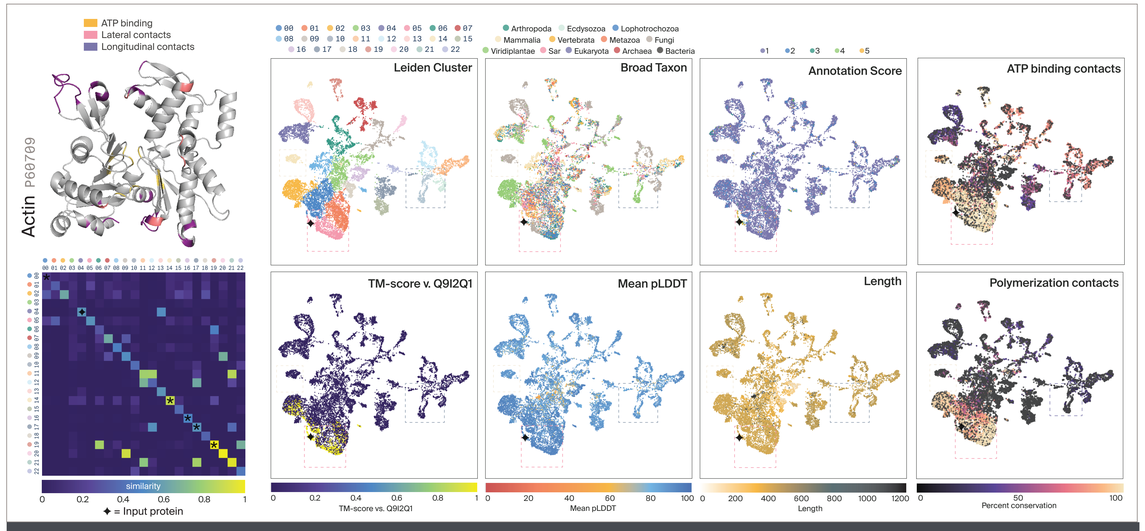


Figure 4. ProteinCartography results for actin. A) Actin with binding contacts highlighted. B) The cross-cluster similarity matrix shows the mean TM-score of all structures in each cluster versus all other proteins in each other clusters. C) The resulting map of the actin family with various metadata and analysis result overlays.

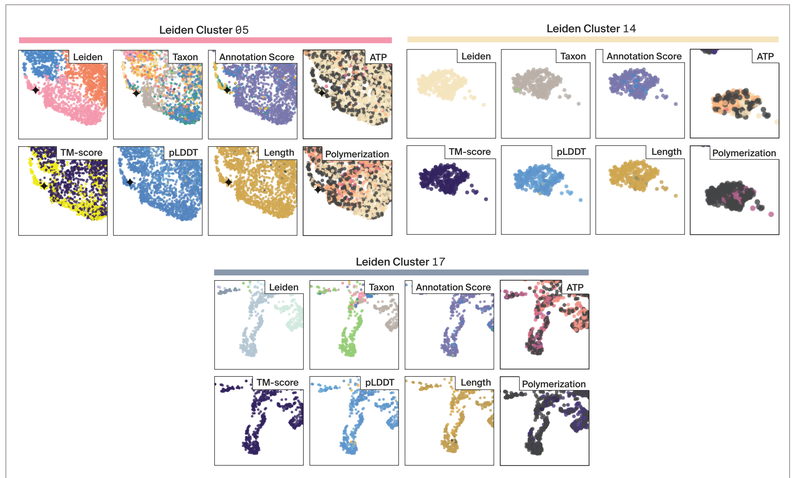


Figure 5. Closer look at clusters helps with hypothesis generation. The input protein resides in cluster 05. Cluster 14 is a group of fungal secondary actins with low conservation. Cluster 17 is a group of plant proteins with low conservation.

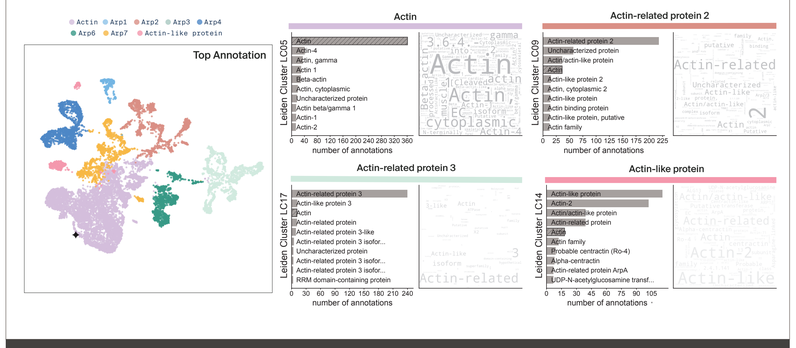


Figure 6. Semantic analysis shows proteins sorted in agreement with existing annotations, but also that "actin" is present in nearly every cluster. Bar graphs show top annotation string and word clouds show top annotation words.

Leave Feedback!

This pipeline is a work in progress — we are actively building and adding features, and we'd love your feedback!

We're particularly interested in getting feedback on the following:

- **Would this kind of pipeline be useful for your own work?**
- **How could we make it more useful for you?**
- **Do you have any recommendations for types of analyses or validation?**

Comment on the pub:

> *ProteinCartography: Comparing proteins with structure-based maps for interactive exploration*

Tweet with #PCAnalysis

bit.ly/ProteinCartography

All other published work: research.arcadiascience.com

A NOTE ON SHARING WITH US!

Part of our mission is to share as much useful research as we can.

If you choose to share a protocol or other useful information with us after viewing this poster, please understand that we may act upon this knowledge and share it when we publish our work. We publish quickly on an independent platform, so this may happen soon after you share, and we cannot wait for you to publish elsewhere.

If you decide to share anyway, yay! That's what science is all about. If your input is useful, we will include you as a contributor to the publication and explain that your role was in providing "Critical Feedback," likely with an additional description of what you shared.

tl,dr — If you're not ready for everyone to know about something, please refrain from sharing it with us.

Contributors (A–Z)

Prachee Avasthi • Conceptualization, Supervision

Brae M Bigge • Conceptualization, Data Curation, Formal Analysis, Investigation, Supervision, Visualization, Writing

Adair L Borges • Critical Feedback

Feridun Mert Celebi • Critical Feedback, Supervision, Validation

Rachel J Dutton • Critical Feedback

Galo Carcia III • Critical Feedback

Jase Gehrig • Conceptualization, Software

Megan L Hochstrasser • Editing

Cameron Dale MacQuarrie • Critical Feedback

David Q Matus • Critical Feedback

Elizabeth A McDaniel • Critical Feedback, Validation

Erin McGeever • Software

Manon Morin • Critical Feedback

Atanas Radkov • Critical Feedback, Writing

Taylor Reiter • Critical Feedback, Validation

Michael E Reitman • Critical Feedback

Dennis A Sun • Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Visualization, Writing

Emily C.P. Weiss • Critical Feedback, Validation

References

1. 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
2. Inigo Barrio Hernandez, Jingi Yeo, Jürgen Jänes, Tanita Wein, Mihaly Varadi, Sameer Velankar, Pedro Beltrao, Martin Steinegger. (2023). Clustering predicted structures at the scale of the known protein universe. https://doi.org/10.1101/2023.03.09.531927
3. 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
4. The UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. https://doi.org/10.1093/nar/gkaa1100
5. 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
6. Plotly Technologies Inc. (2015). Collaborative data science. https://plot.ly