Semi-Automated Bacterial Protein Dynamics Analysis

High throughput PlzC oscillation tracking in *V. Cholerae*

Aidan Boyne, Elena Tkachuk, Elsa Dubil, Exequiel Nuño, Jacob John, Jamie Stickelmaier, Julia Zhong

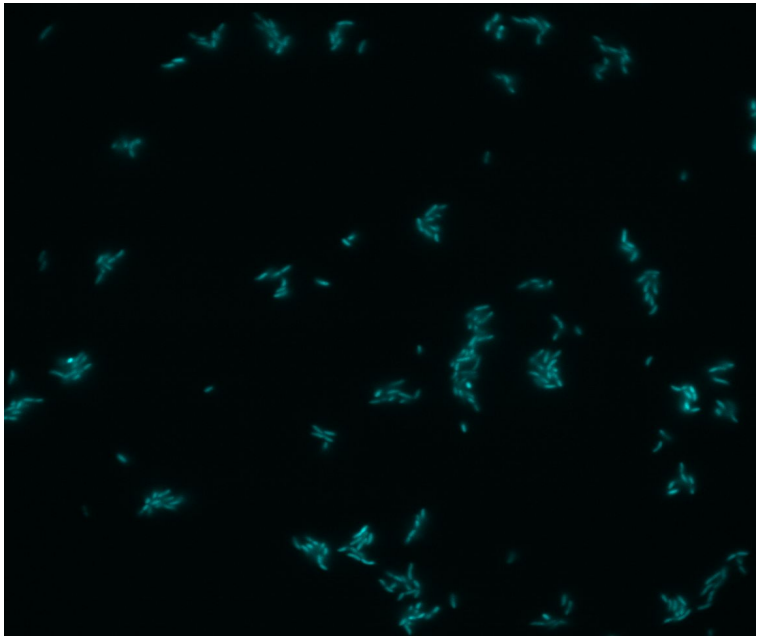
UCLA Samueli School of Engineering

Wong Lab

Biofilms are colonies of aggregating bacteria that adhere to a surface

Biofilms have a widespread impact in medicine and industry

- Allow pathogens to survive in chronic diseases
- Allow formation of resistant enclaves of bacteria on and around implants
- Affects water transport and food industries



The economic significance of biofilms

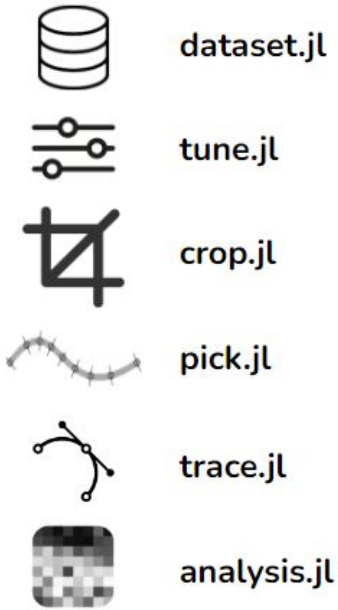
- In 2017 the global expenditure on wounds in healthcare was **$7.8T**, of which it is estimated that **$281B** corresponds to biofilms (*Nature*)
  - Annual cost of revision surgery due to biofilm-mediated infections: $7.8B globally
  - Annual market estimate of surfaces resistant to microbial contamination: $7.1B
  - Foodborne pathogens and biofilms are responsible for $78.6B lost every year
- The market of biofilms in wastewater treatment is an est. **$313B** globally
- Increased hospitalization and cleaning needs as a result of the COVID-19 pandemic have increased the market by an additional **$5T**

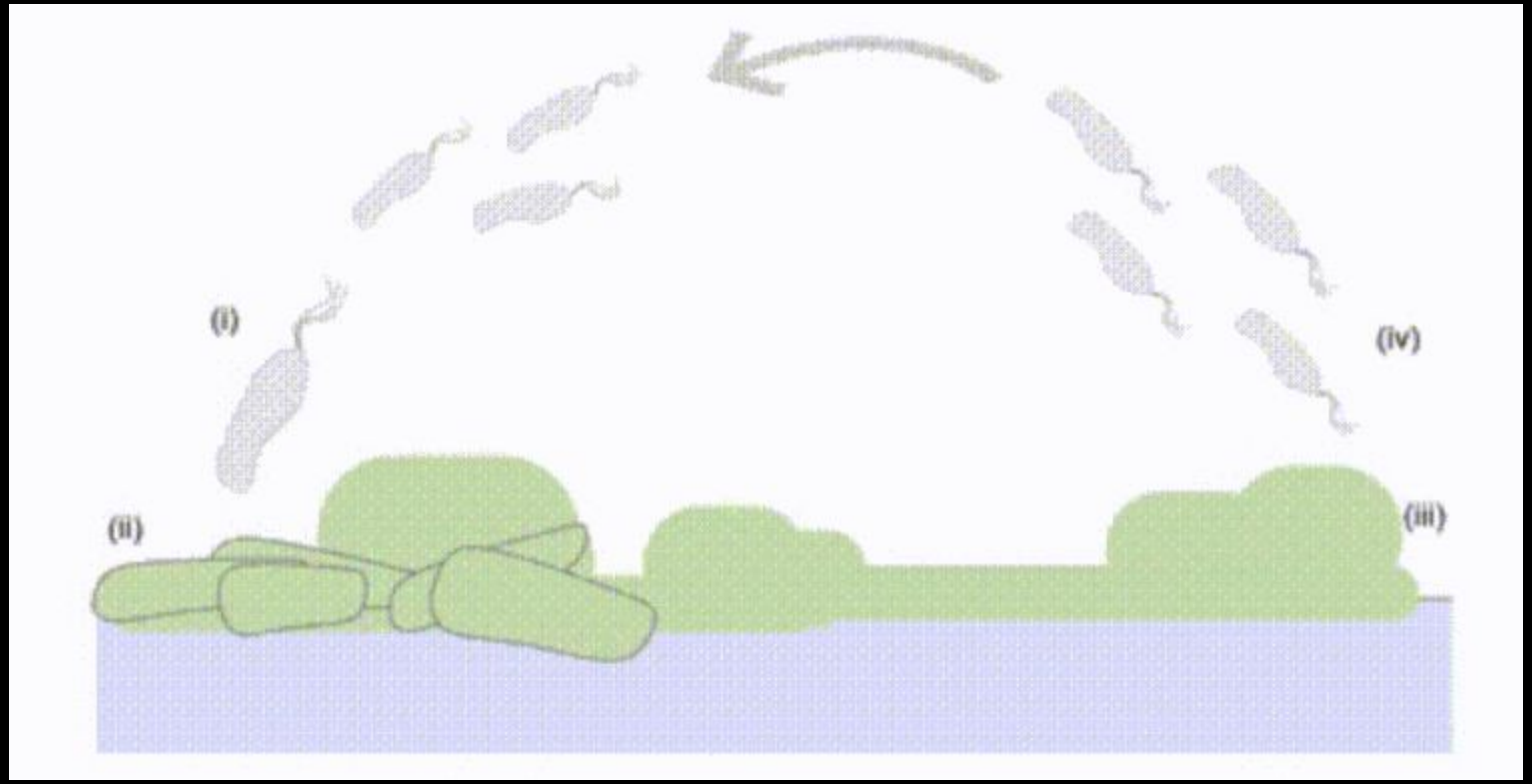
Cámara et al. *Economic significance of biofilms: A multidisciplinary and cross-sectoral challenge*

We developed a machine learning model to classify time-position plots of intracellular protein dynamics

Introducing a unique solution to a data bottleneck

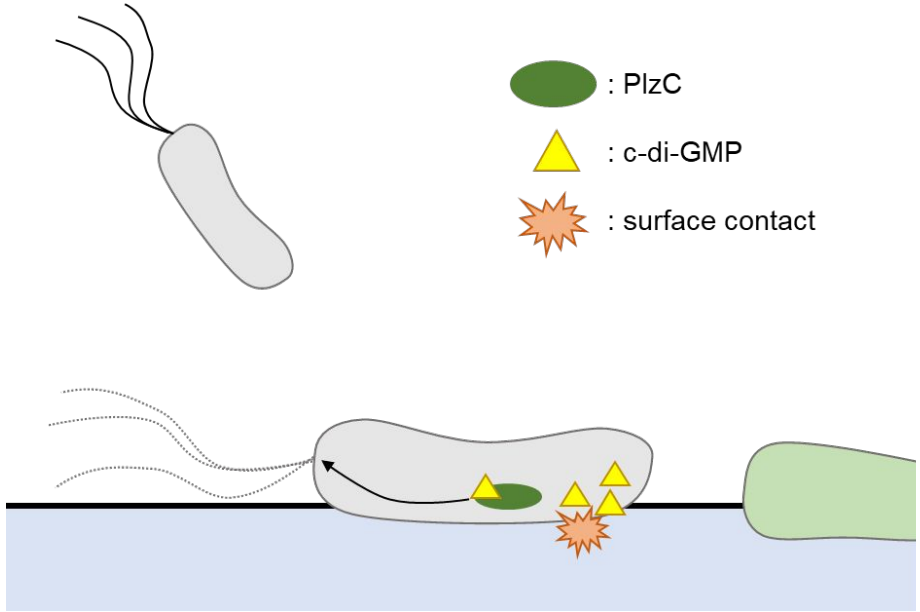
- Understanding *why* biofilms form is crucial to preventing formation
- Observing the movement of protein within bacteria allowed us to create a machine learning model
- This model can be generalized to other bacterial species and other single-protein tracking applications



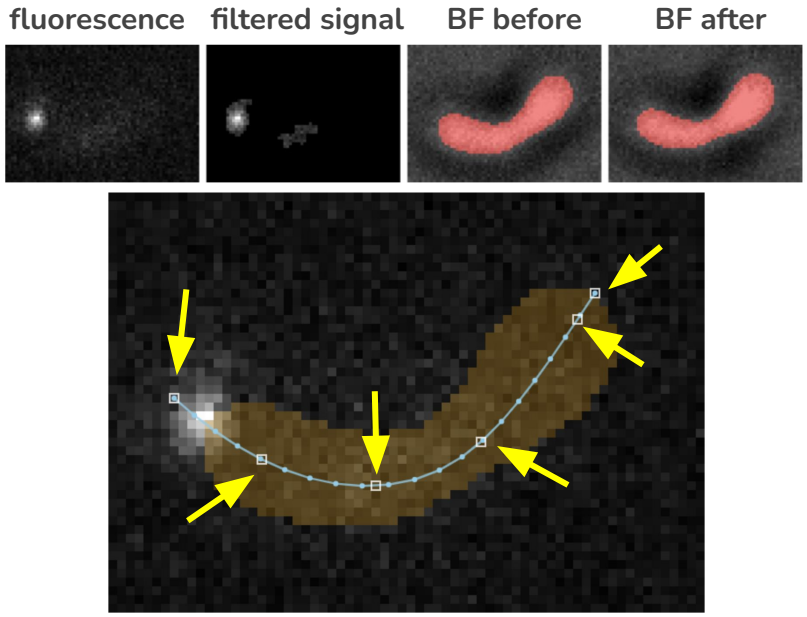


PlzC-protein has an unknown role in biofilm formation

- **Known:** c-di-GMP concentration level is related with biofilm formation
- **Hypothesis**: PlzC is a carrier protein for c-di-GMP, which is correlated with biofilm formation
- Tracking PlzC **movement** using fluorescence microscopy is essential to understanding its relationship with c-di-GMP and biofilm formation



Tracking PlzC-protein with fluorescence microscopy

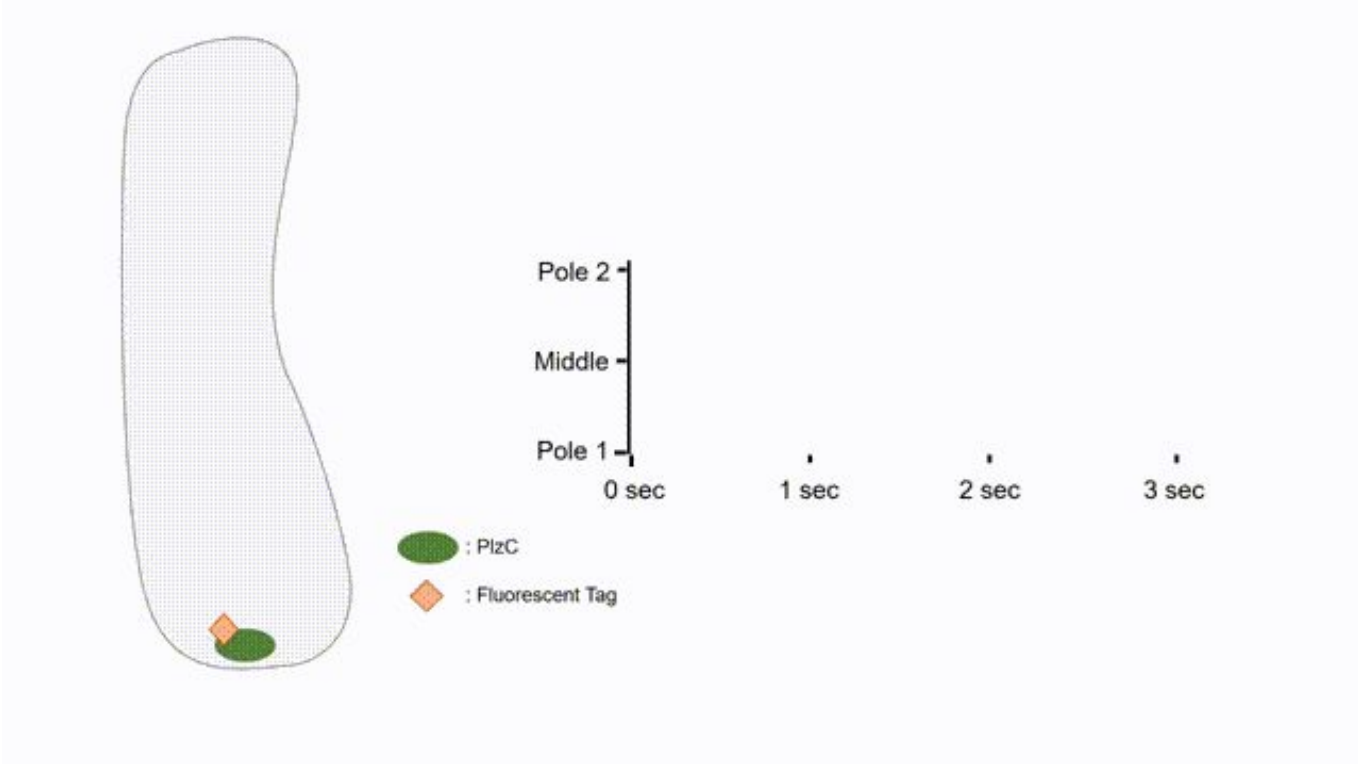


Frame #1

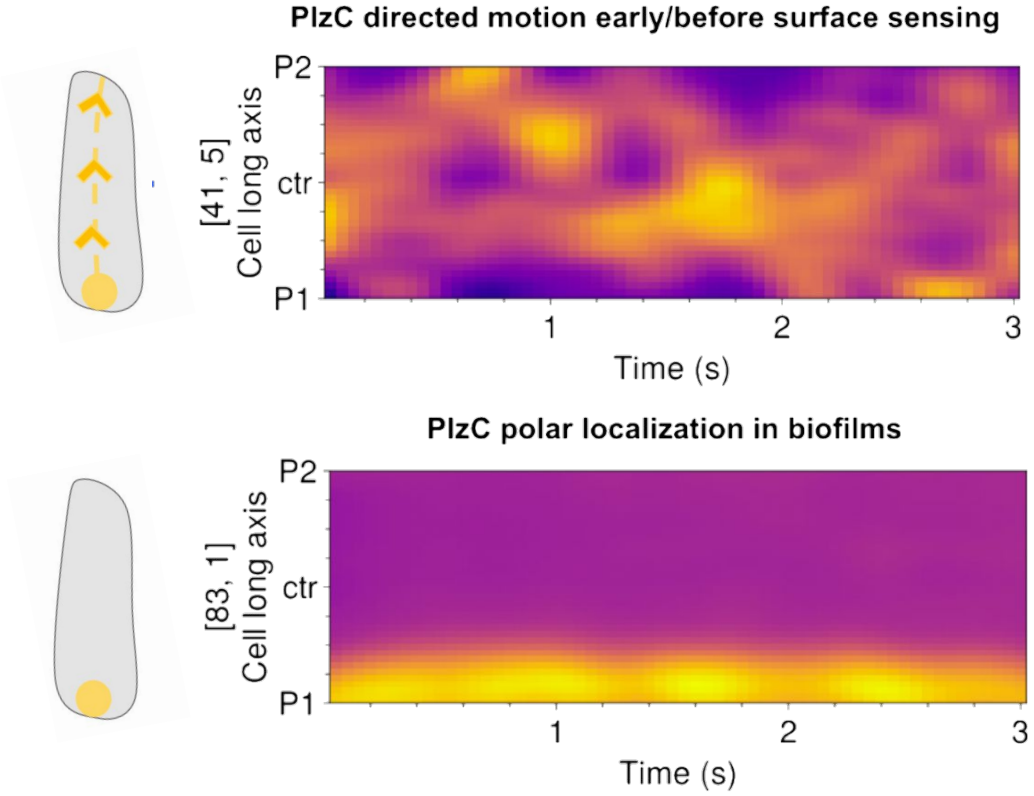
WT, PlzC-tYFP

Bacterial Sensing Group

Visualizing PlzC movement in *V. Cholerae*



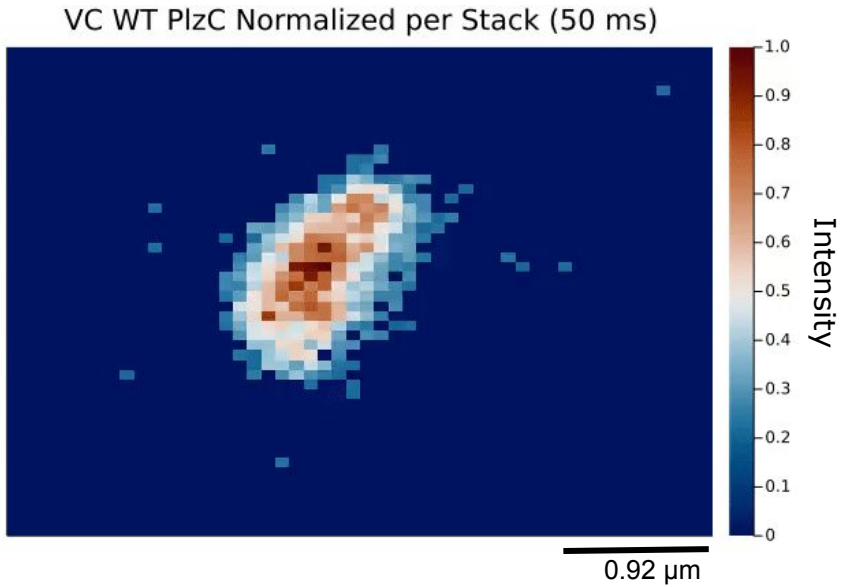
Kymographs visualize PlzC movement along splines



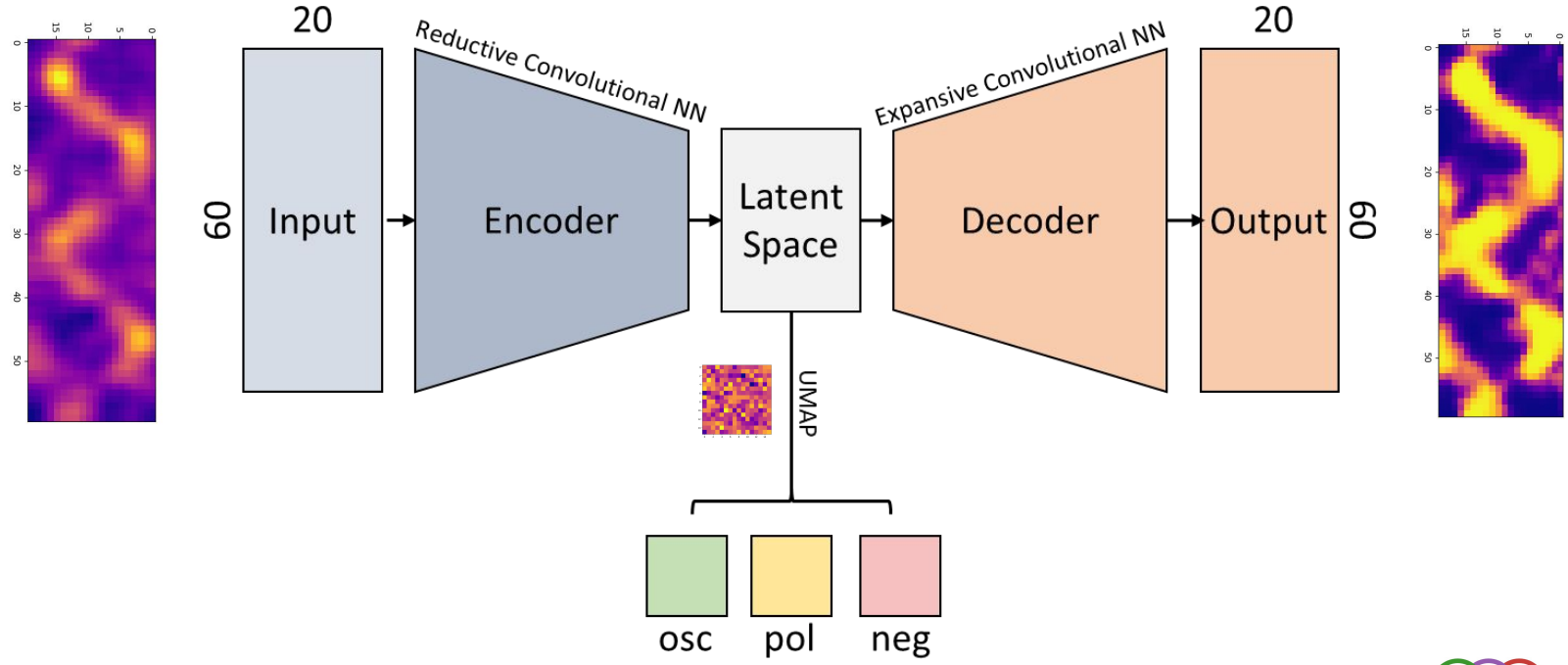
- Each cell crop has a kymograph of PlzC movement
- Classify distinctions in intracellular PlzC movement across varying strains
- **Scale of data too large and complex** datasets to find patterns manually

An automated and general workflow for large scale protein movement analysis

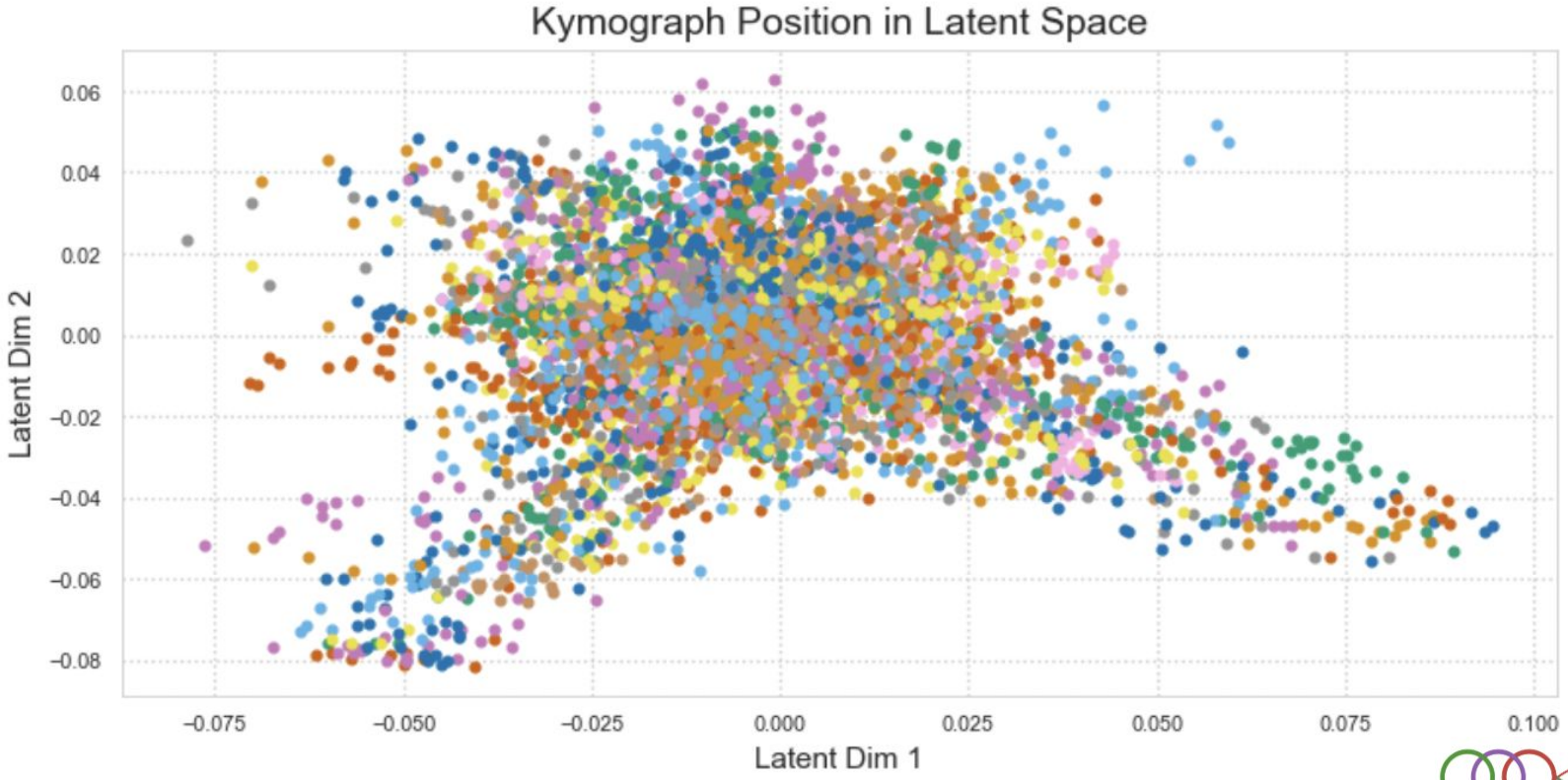
- Tracking intracellular movement of PlzC protein
- Machine learning allows for analysis of large datasets
- Observation of patterns to draw hypotheses about role in biofilm formation



Our Variational Autoencoder (VAE) allows us to classify kymographs

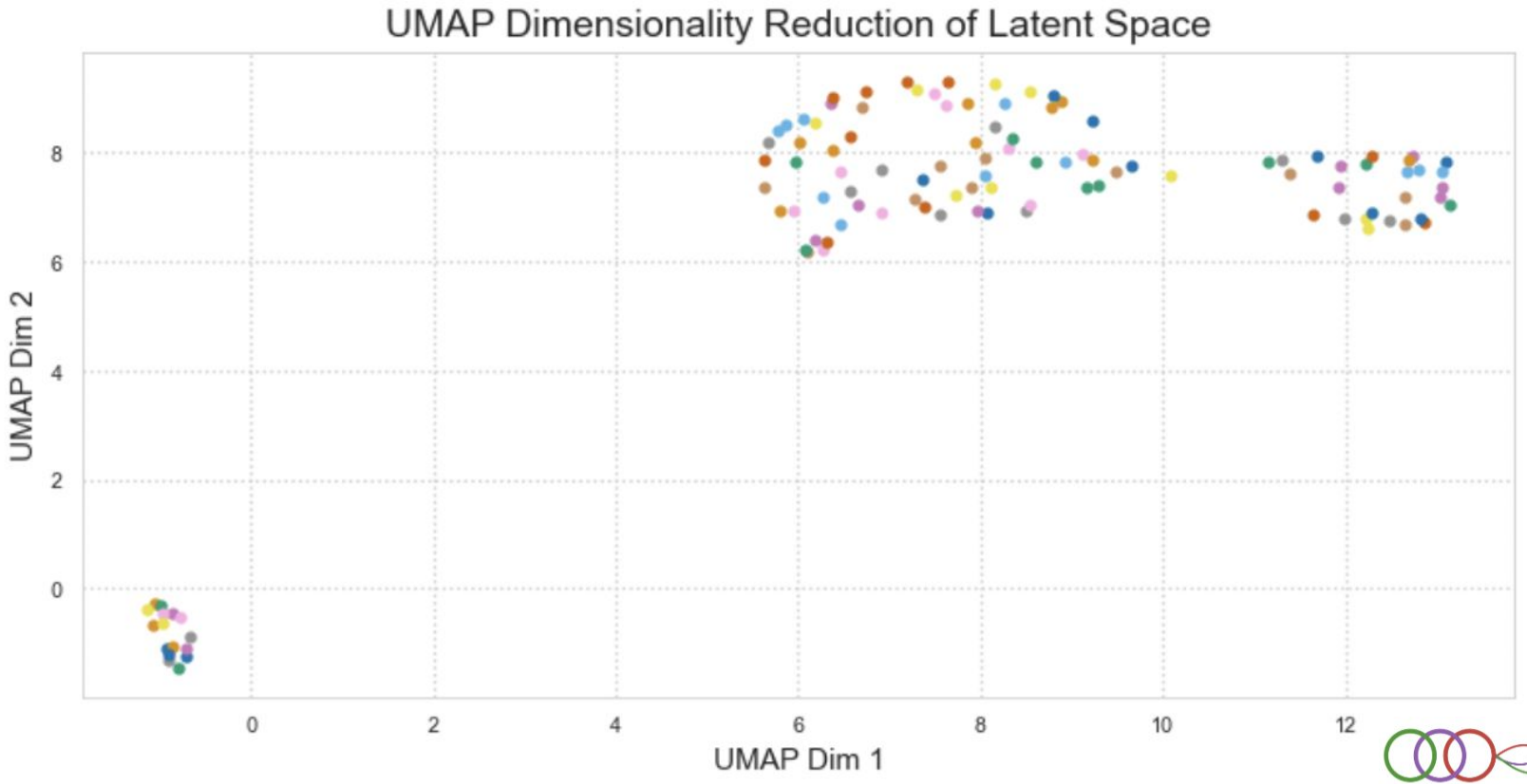


Visualizing the raw latent space is too busy



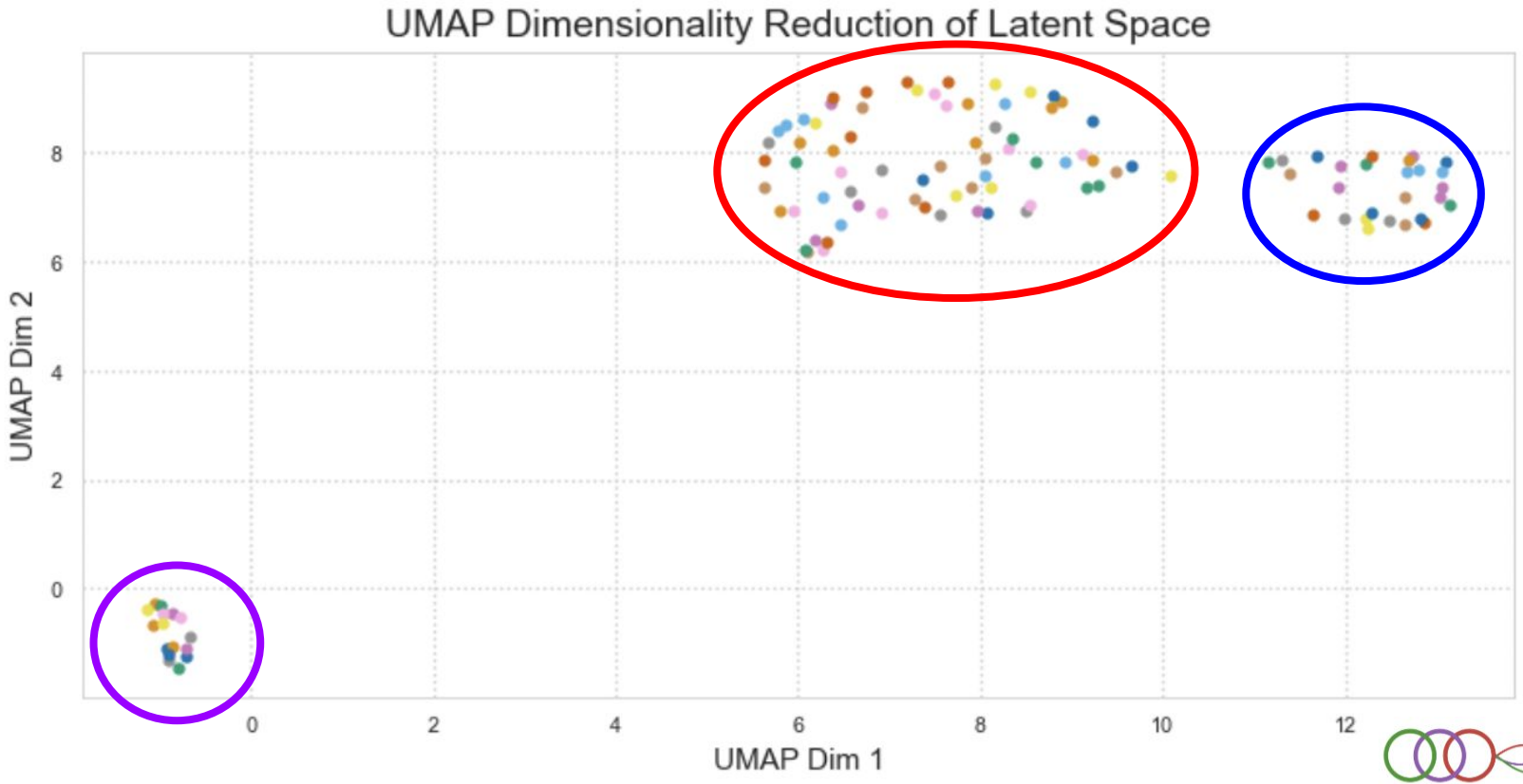
Bacterial Sensing Group

UMAP dimensionality reduction reveals grouping



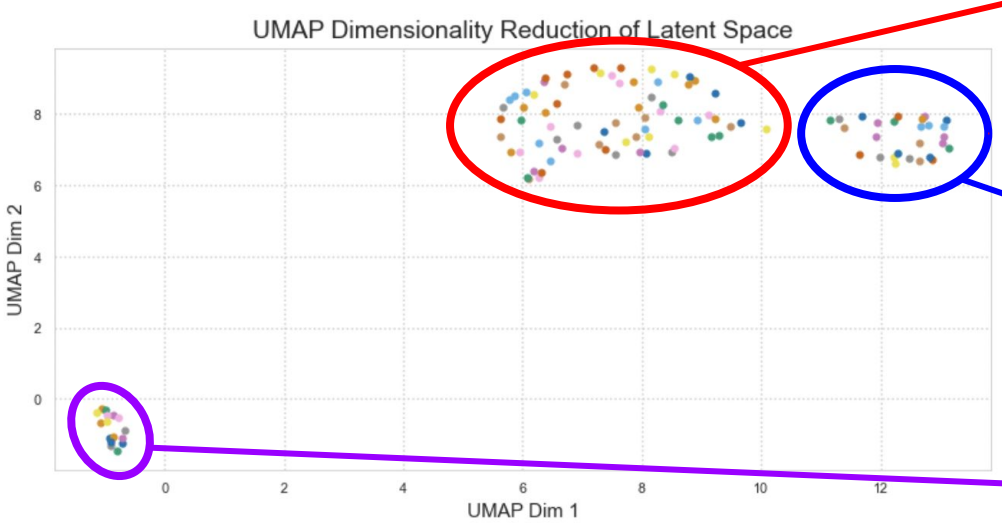
Bacterial Sensing Group

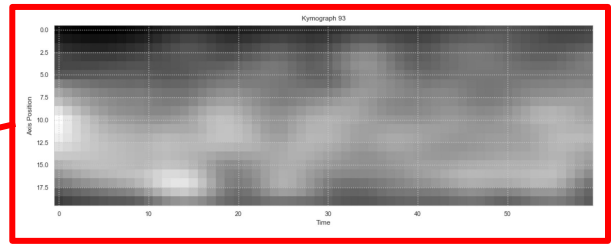
UMAP dimensionality reduction reveals grouping

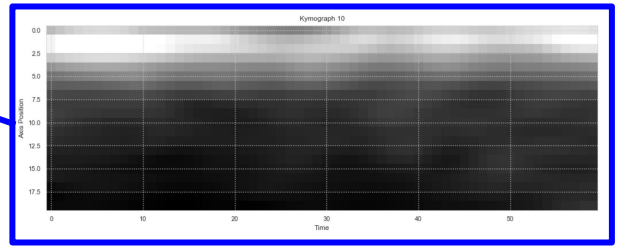


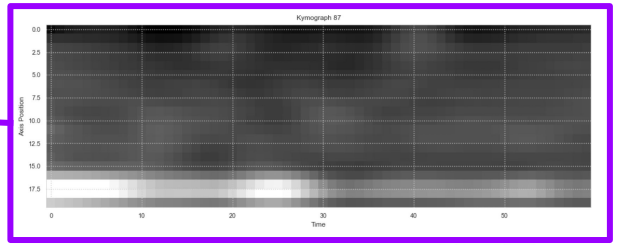
Bacterial Sensing Group

UMAP dimensionality reduction reveals grouping





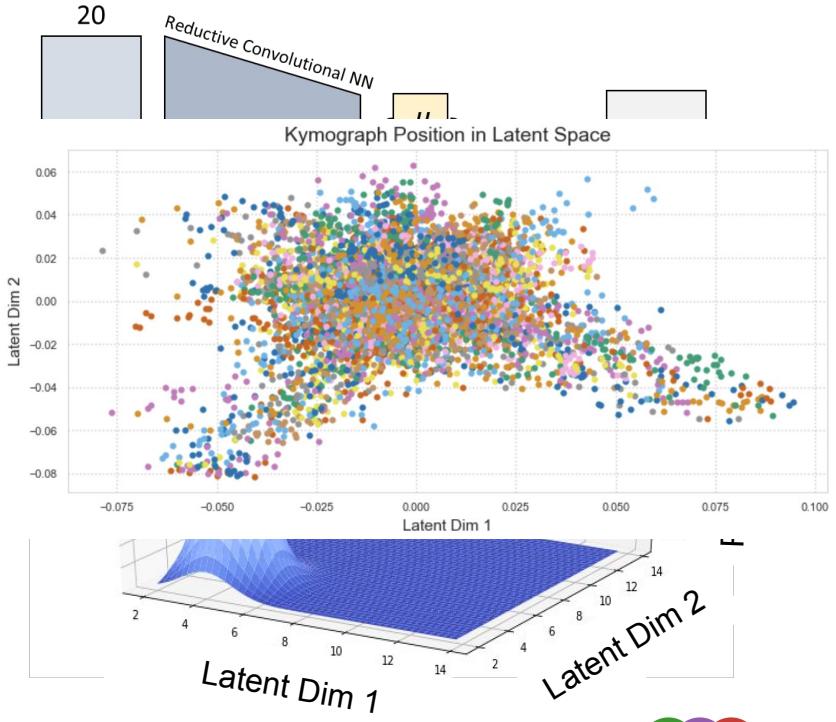




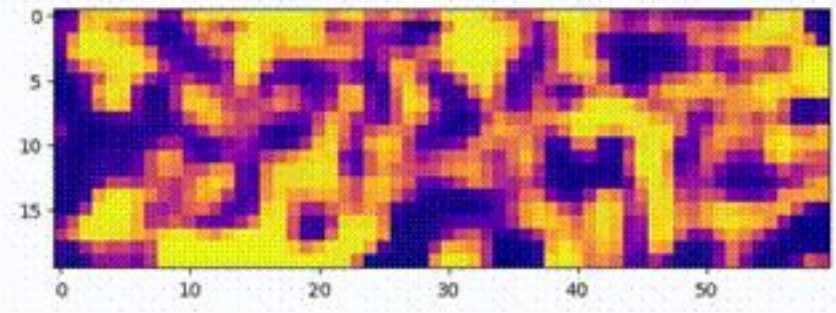
Bacterial Sensing Group

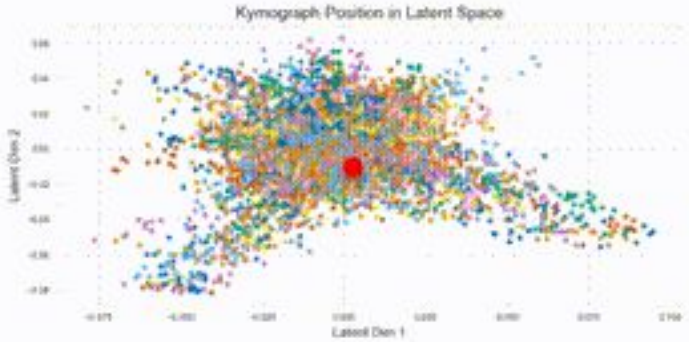
VAEs preserve latent relationships

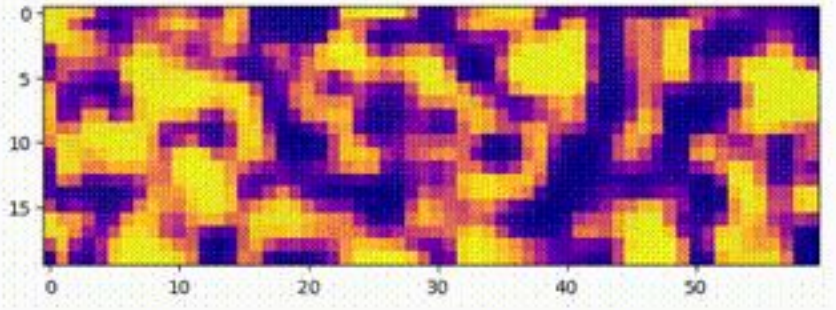
- ***Probabilistic*** representation in lower dimensional space
- Extract **essential spatial characteristics** and map to latent space

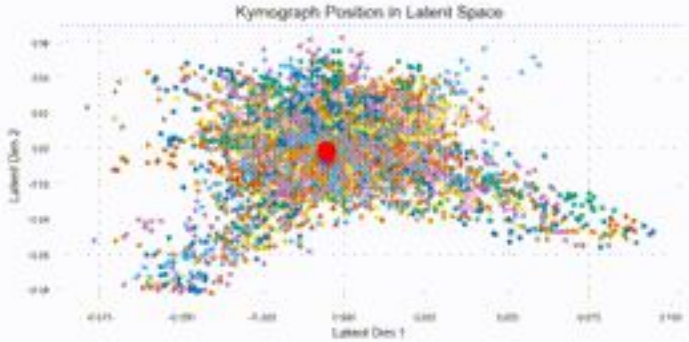


Reconstructing the latent space









Summary

- Created a tool for high-resolution protein dynamics visualization and interpretation
  - Developed to investigate molecular machinery of biofilm formation in *V. cholerae*
    - Quantified and classified differences in fluorescence kymographs of super-diffusive PlzC movement
- **Generalizable** to any form of single-molecule fluorescence tracking on the microscopic scale

Acknowledgements

Thank you to Dr. Gerard Wong, Dr. Calvin Lee, Jonathan Chen, William Schmidt, and everyone at Wong lab for the use of their lab and materials

And thank you to Dr. Gao, Salil, and Xiaoxi for their time and guidance

References

Cámara, M., Green, W., MacPhee, C. E., Rakowska, P. D., Raval, R., Richardson, M. C., Slater-Jefferies, J., Steventon, K., & Webb, J. S. (2022). Economic significance of biofilms: A multidisciplinary and cross-sectoral challenge. Npj Biofilms and Microbiomes, 8(1), 1–8. https://doi.org/10.1038/s41522-022-00306-y

Mark B. Abelson, M. D. (2012, September 6). *Of biomes, biofilm and the ocular surface*. Review of Ophthalmology. Retrieved July 4, 2022, from https://www.reviewofophthalmology.com/article/of-biomes-biofilm-and-the-ocular-surface

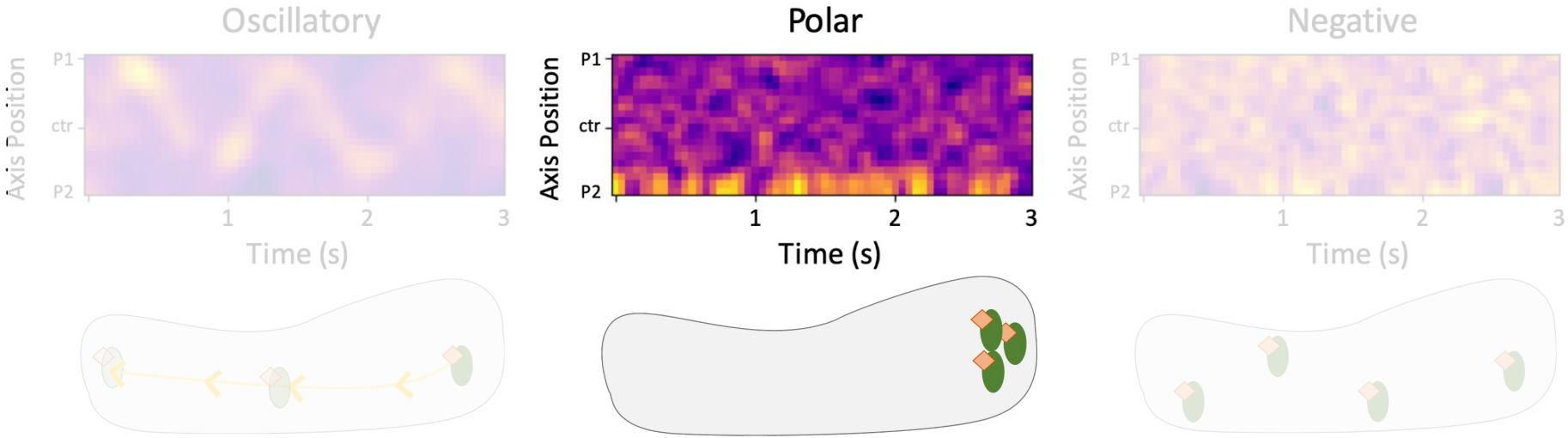
Lopez, L. (2021, April 19). *Pseudomonas aeruginosa biofilms and their partners in Crime*. Institute for Bioengineering of Catalonia. Retrieved July 6, 2022, from https://ibecbarcelona.eu/pseudomonas-aeruginosa-biofilms-and-their-partners-in-crime/

Pratt, Jason T., Rita Tamayo, Anna D. Tischler, and Andrew Camilli. “PilZ Domain Proteins Bind Cyclic Diguanylate and Regulate Diverse Processes in Vibrio Cholerae.” The Journal of Biological Chemistry 282, no. 17 (April 27, 2007): 12860–70. https://doi.org/10.1074/jbc.M611593200.

Berk, Veysel, Jiunn C. N. Fong, Graham T. Dempsey, Omer N. Develioglu, Xiaowei Zhuang, Jan Liphardt, Fitnat H. Yildiz, and Steven Chu. “Molecular Architecture and Assembly Principles of Vibrio Cholerae Biofilms.” Science 337, no. 6091 (July 13, 2012): 236–39. https://doi.org/10.1126/science.1222981.

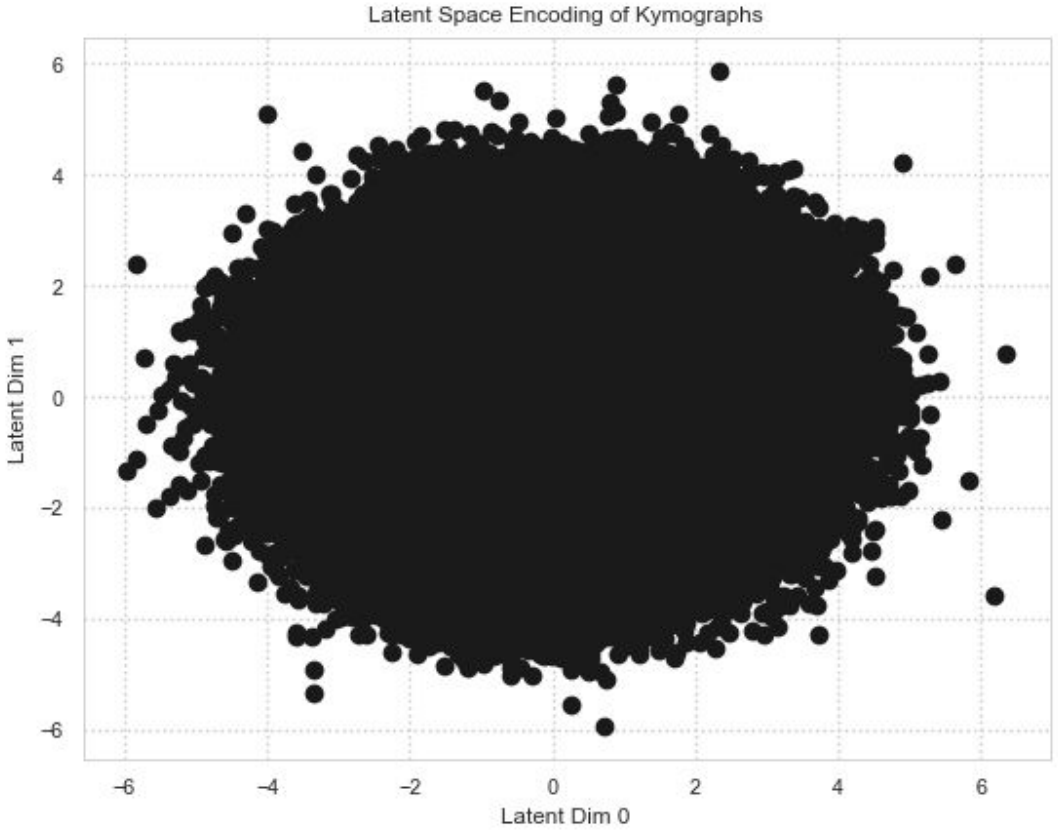
Quan Xue and M. C. Leake, "A novel multiple particle tracking algorithm for noisy in vivo data by minimal path optimization within the spatio-temporal volume," 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1158-1161, doi: 10.1109/ISBI.2009.5193263.

Going back to the biology

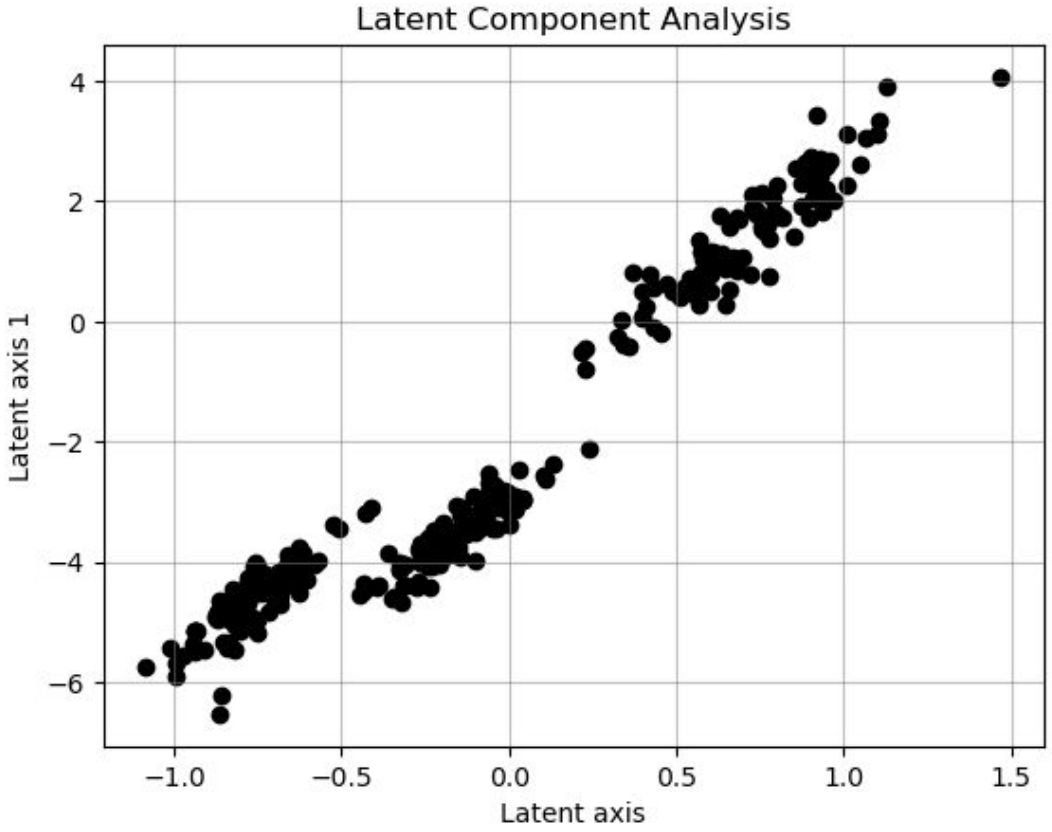


Bacterial Sensing Group

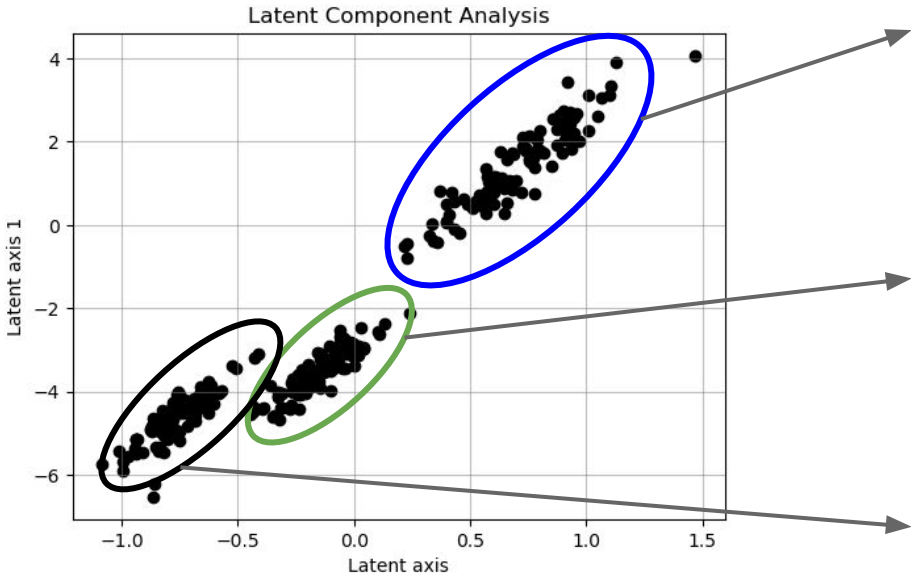
Raw VAE latent space still has too many dimensions

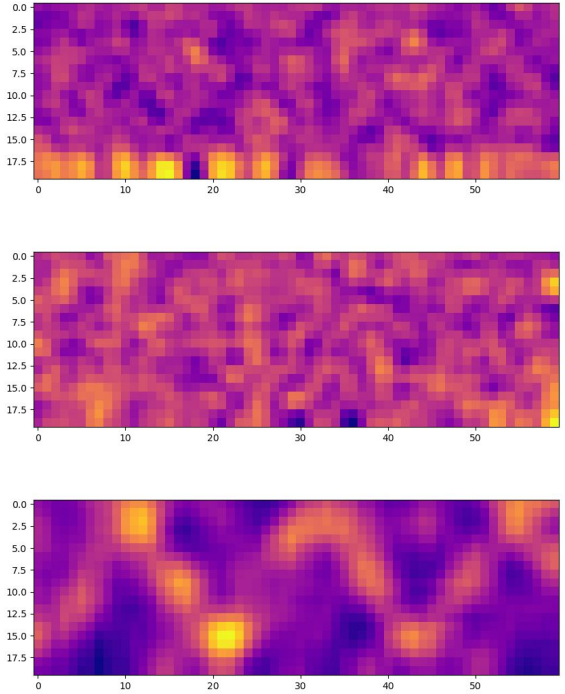


UMAP dimensionality reduction to reveal data groups



Sample points from each group using kymograph IDs



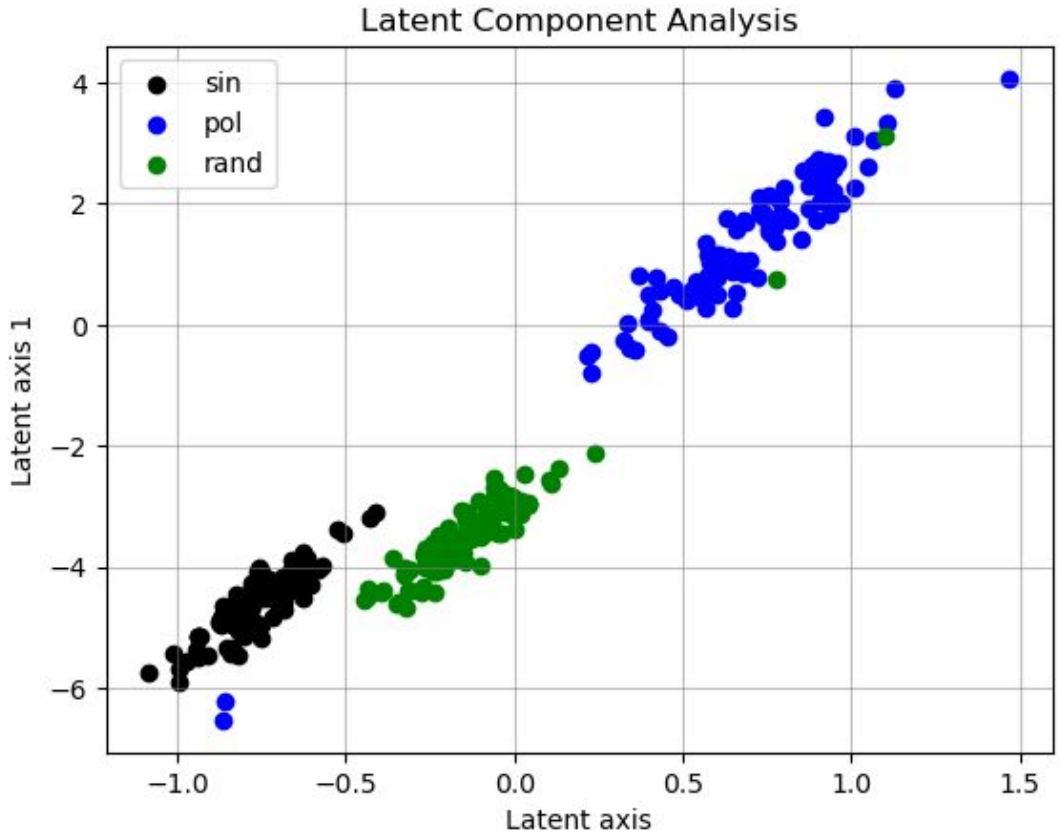


Polar

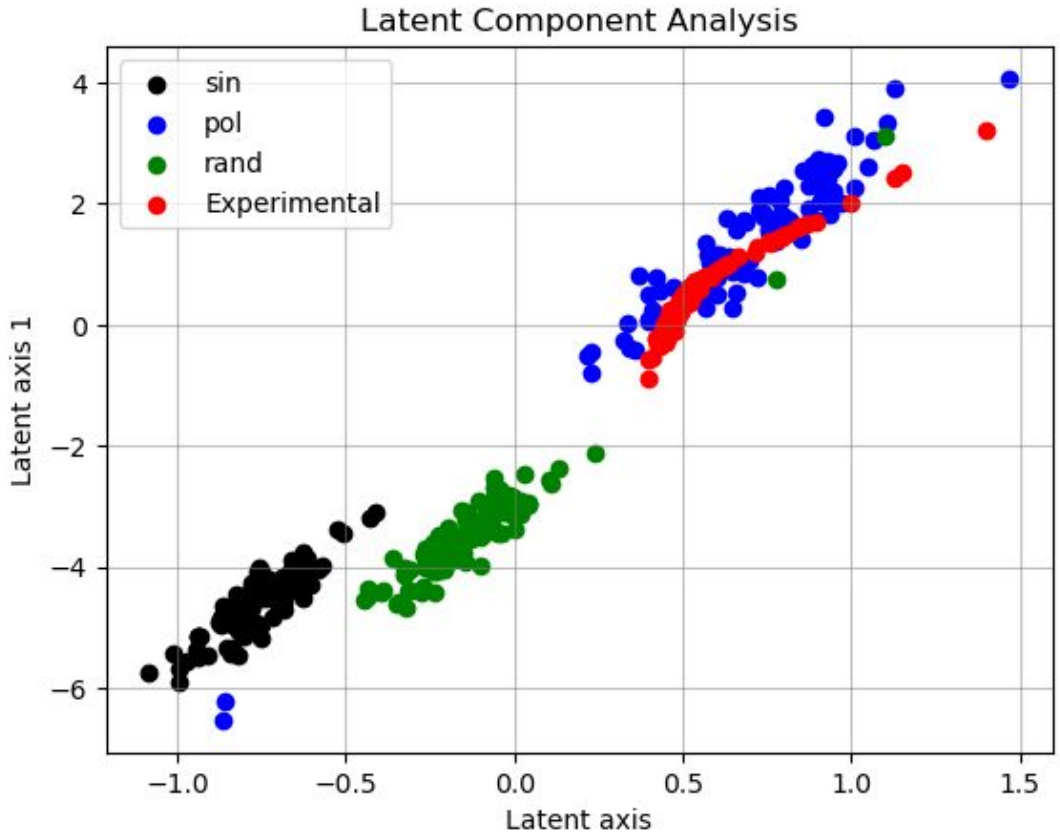
Random

Oscillatory

Label groups accordingly



Encode experimental data and look for similarities



Experimental data occupies a similar region of latent space as the **Polar Kymographs**

Wild Type PlzC is likely Polar