Variational autoencoder networks for scalable classification and analysis of protein movement in bacteria

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

Department of Bioengineering, Samueli School of Engineering and Applied Sciences, University of California, Los Angeles (UCLA)

ABSTRACT: Biofilms are polymeric matrices which allow certain species of bacteria to adhere to surfaces and survive adverse environmental conditions. In humans, pathogenic biofilm-forming bacteria can cause chronic infection in both native tissue and implants. Here, we utilize a variational autoencoder (VAE) model and accompanying data processing pipeline to understand internal protein dynamics in *Vibrio cholerae* and their subsequent effect on biofilm formation. Specifically, we used this method to observe the movement of PlzC, a protein linked to initial biofilm formation. Analysis of the superdiffusive movement of PlzC revealed highly regular oscillatory and polar motion patterns in wild-type V. *cholerae*. Using the same methodology to analyze distinctions in PlzC behavior across varying strains or steps along the biofilm pathway, the relationship between PlzC and biofilm formation can be further investigated. The VAE model and pipeline are highly generalizable and can accelerate fluorescence microscopy analysis of intracellular proteins in a wide variety of microbial species, leading to a greater understanding of signaling pathways involved in many human diseases.

KEYWORDS: *bacteria, biofilm, neural networks, variational autoencoder*

Biofilms are polymeric microbiome matrices created by bacterial colonies that provide protection against external pressures such as pH, toxins, osmolarity, shear forces, and nutrient scarcity. In infectious bacteria, biofilms can provide both antibiotic resistance and a barrier to host immune components, resulting in longer-lasting and more severe disease[1]. Prior work has shown that increased concentration levels of the secondary messenger, bis-(3′-5′)-cyclic dimeric guanosine monophosphate (c-di-GMP), reduces the motility of bacteria and enhances the production of certain polysaccharides, both essential to the initiation of biofilm formation[2] (**Figure 1**). However, the interactions and dynamics of the intricate machinery enabling biofilm formation remain unclear. The protein PlzC (VC2344) found in *Vibrio cholerae* possesses two c-di-GMP binding sites and a dissociation constant of 314 nM, indicating a high binding affinity between the two molecules and suggesting that PlzC may also play an important role in biofilm synthesis[3,10].

High-speed fluorescence microscopy is a promising option for the experimental analysis of single-cell protein trafficking, but it presents a variety of challenges[14]. The sheer scale of data (tens to hundreds of bacteria per frame across dozens of frames) makes manual analysis impossible. Furthermore, even after processing, it is often difficult to appropriately categorize data between different experimental conditions for comparison. Unsupervised machine learning models provide solutions to both of these problems in biological image processing[5,6]. Unlike supervised learning models such as convolutional neural networks or nonlinear regressions, unsupervised learning models do not require a priori labeling of data in the training set. In situations where the relationship between the independent and dependent variables is unclear, unsupervised models are able to categorize data by closely grouping samples with similar characteristics within the latent space[6].
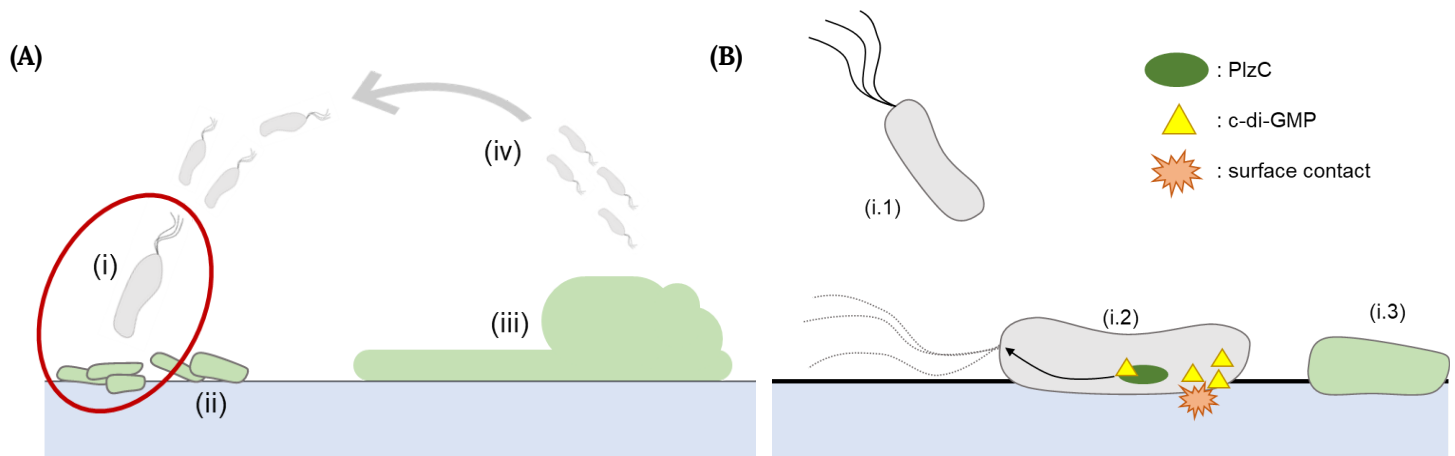
Figure 1. (A) Biofilm formation consists of (i) surface contact, sensing, and phenotypic shift, (ii) sessile bacteria multiply and begin to produce biofilm components, (iii) mature biofilm characterized by fully developed polysaccharide matrix is developed, (iv) bacteria leave the biofilm individually or in small groups and revert to a motile phenotype.
(B) A motile VC bacteria (i.1) contacts a surface, triggering the production of c-di-GMP at the point of contact. We hypothesize that PlzC acts as a shuttle to transport c-di-GMP to the poles of the bacteria (i.2), causing flagellar deactivation and causing the bacteria to enter a sessile state (i.3).

The latest developments in biofilm formation done by the Wong Lab at UCLA remark super-diffusive fluctuations of PlzC across the cell. This paper explores the use of a variational autoencoder (VAE) as a tool to analyze and classify kymographs that were obtained from high-speed, high-resolution fluorescence videos of PlzC in wild-type and different mutations of *Vibrio cholerae*. Machine learning and neural networks have had increasing promise and use in recent years due to their significant potential to eliminate the gap between the ever-increasing rate of data gathering and the comparatively stagnant rate of data analysis[15]. In particular, deep learning has provided a streamlined method for automatic image recognition which can outperform manual analysis and processing in both speed and accuracy.

A generative AI model aims to learn underlying distributions in input data. One of the most common types of generative models are variational autoencoders (VAE), which explicitly define and learn the likelihood of data[7]. VAEs are made up of an encoder and a decoder. The encoder takes in a target distribution and compresses it into a low-dimensional latent space while the decoder attempts to reconstruct the original image using only the information available in the latent space. A successfully trained network can generate latent space representations of input images and interpolate between them by using the decoder to produce new images. Currently, the analysis of kymographs, which are graphical representations of spatial position over time often used to visualize particle motion, is a significant bottleneck in our investigation of PlzC dynamics. To accelerate this process, we developed a VAE model to analyze kymographs obtained from PlzC tracking from high-speed, high-resolution fluorescence videos of different mutations of *Vibrio cholerae* (VC). We tested our VAE using kymographs of PlzC using a wild-type VC strain. The model will allow us to analyze kymographs on a mass scale and will bring us closer to understanding the role of PlzC in biofilm formation.

On a broader scale, the framework for this VAE can be used for future research involving molecular tracking and analyzing kymographs. The model can be altered and extrapolated for a wide array of research and experiments. A successful VAE for kymographs will expedite research for many future endeavors and help account for the data explosion in biological research.

MATERIALS AND METHODS

Preparation of Agar Plates and Inoculation. Bacterial cells are stored within a -80°C freezer for safe and stable storage. LB plates were prepared by adding 25g of LB broth into 500mL of DI water, letting it dissolve, and adding another 500mL of DI water. The solution was then evenly distributed in 4, 500mL bottles, each containing 3.75g of agar, and then autoclaved and placed on a shelf for future use. 10 sterilized Petri dishes were then placed in a sanitized area. LB+agar stock was then melted so that 20mL of stock can be aliquoted into each of the 10 Petri dishes and then left to dry. LB media was made by adding 25g of LB broth into 500mL of DI water, letting it dissolve, and adding another 500mL of DI water. The solution was then evenly distributed in 4, 500mL bottles. 2% LB media was made by adding 10g of NaCl, 0.2g of tryptone, and 0.1g of yeast mass into 500mL of DI water. 500mL of DI water was then added once all of the components dissolved and the solution was titrated with NaOH until a pH of 7.5 was reached. The resulting media was then evenly distributed in 4, 500mL bottles. Cells from the -80°C were inoculated into one of the agar plates and put in an incubator at 35°C for ~20 hrs, wrapped in parafilm, and stored on a benchtop.

Flow Cell Construction. Flow cells were made in the lab by using IBIDI's sticky flow channel. A coverglass was placed in 70% ethanol for at least 15 minutes to sanitize and eliminate specs of dust. Once the coverglass was cleaned, the edge of it was aligned with the edge of the sticky side of the flow cell and firmly but carefully pushed against the glue. Tweezers with tips covered by plastic tubing were used to push air bubbles trapped between the channel assembly and the coverglass. To make the flow cell assembly, two elbow joint Luers, three male lock adapters, and one female Luer lock adapter were taken from 70% ethanol-full flasks placed on a fume hood. They were dried using pressurized air and moved to a bench. 3.18 mm diameter clear tubing was cut following marks on the bench, giving us 2 long pieces of tubing and a short 1.5cm tubing piece. The elbow joint Luers were placed on each side of a clean flow cell channel. A long piece of tubing was connected to one of the elbows and the short piece of tubing was connected to the other elbow. A sterile injection port was taken from its designated location and a male Luer lock was secured into it. The female Luer lock was then attached to the other side of the injection port and the 2nd long tubing was connected to it. The remaining male Luer lock was connected upstream to the injection port for syringe connection. A 60mL syringe was filled with 50+ mL of H_2O_2 and a yellow PVDF membrane/filter was then connected to it. The filter was then connected to the male Luer lock and the whole assembly was then cleaned by running H_2O_2 at a rate of 25mL/hr for 2 hours followed by DI water at a rate of 5mL/hr for 10 hours.

Culturing of V. *cholerae* and Further Flow Cell Assembly. To make an overnight culture growth, single colonies were inoculated from the designated culture plate and placed into a test tube containing 2mL of LB media which was then placed in an incubator at 25°C for ~19 hours. 0.4mL of the cell culture was then taken in 20mL of LB media (1/50 dilution) for a first regrowth. The optical density (OD) was then measured after allowing the culture to grow for 2 hours. If the value hovered at 0.4A, 0.4 mL of cell culture was taken in 20mL of LB media for another regrowth to ensure we image cells that are at mid-log. Bubbles from the flow cell assembly were pushed away by hitting the yellow PVDF membrane/filter with a sharpie to push any bubbles in the system. The 60mL syringe was then replaced by a 140 mL autoclaved syringe filled with 2% LB media. Finally, the flow cell assembly along with the

syringe were taken to the microscope room and loaded in the microscope. 2 hours after the second regrowth, the OD was measured. If the OD hovered at 0.4 AU, the solution was then diluted to 0.03 AU and taken to the microscope room.

Fluorescence Microscopy. The Olympus IX83 inverted microscope was used for the fluorescence imaging of PlzC in the V. *cholerae cells*. A Hammamatsu qCMOS camera with an effective number of pixels of 4096 x 2304, readout speed of 120 frames/second, readout noise of 0.27 electrons, and dynamic range of 25900:1 was connected to the microscope for image collection[8]. The flow cell and media assembly were mounted on the stage of the microscope. 350uL of the diluted cell solution was injected into the flow cell through the injection port and allowed to settle into the surface of the flowcell for 20 minutes. After 20 minutes, the cells were imaged at a framerate of 20fps. Channels used for the imaging process were bright field, Cy5 (red), and YFP (yellow). The program used to control the microscope was micromanager, a plug-in for ImageJ that allowed for microscope control and initial image examination.

Image Processing. The fluorescence microscope timelapses were passed through DotsWorkflow.jl, a semi-automated processing pipeline written in Julia for image analysis. Each frame of the collected timelapses contained tens of individual cells of V. *cholerae*, some of which were motile and others sessile. The first step of the pipeline was to detect single cells of bacteria within this frame that fit the criteria for protein oscillation analysis, i.e. bacteria that were stationary over time and oriented with their long axis parallel to the flow cell, then generate timelapses of single cells cropped from the original timelapse. Cells that fit the above criteria were deemed *valid* and square-cropped with a padding of 10 pixels on all sides, then passed on to the pixel intensity extraction steps.

For each valid crop, pixel intensity values along the center contour line of the cell were recorded. To fit the center contour line to the cell, we manually selected 4-6 pixels along the longitudinal cell axis and recorded their coordinates. Then, by fitting a

cubic spline along the recorded coordinates, we calculated the path length along the spline and interpolated 20 (for 2x2 binning) or 40 (for 1x1 binning) equally-spaced pixel coordinates from one pole of the cell to the other along the cell's center axis (**Figure 2**). We then applied a Gaussian blur with a standard deviation σ = 1.5 to each frame of the timelapse and recorded the pixel intensity value per frame for each pixel interpolated from the fitted cubic spline. We termed these pixel intensities over time "traces."

1μm
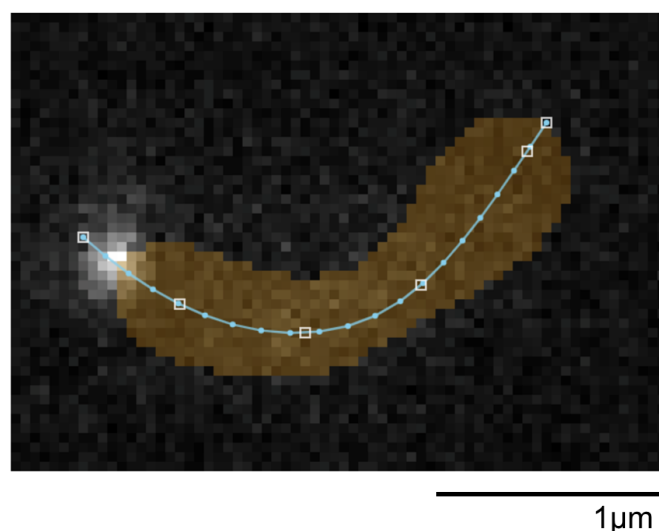
Figure 2. Isolated frame from a single cell crop of VC overlaid with predicted cell boundary (yellow) and long axis spline (blue line) computed automatically from selected points (white boxes). Note the high PlzC concentration on the leftmost pole of the VC cell.

To correct for photobleaching decay of the fluorescence imaging, we fit an average exponential decay function to the recorded traces, then demeaned the traces by recording the percent difference between the traces and the mean exponential decay function. We then used LOESS local polynomial regression with a smoothing parameter of 0.1 to remove high-frequency noise from the demeaned traces, generating a final output of smoothed traces from the DotsWorkflow.jl pipeline.

We then generated a kymograph for each crop by plotting a heatmap of the smoothed traces using Julia's CairoMakie with time (in seconds) on the x-axis and position along the longitudinal axis (in fractional position from one pole to the other). We clipped each kymograph to the length of one half-life of the fitted average exponential decay function, about 3 seconds, as the fluorescence signal yield dropped greatly after the first half-life of the photobleaching decay. These final clipped kymographs were the input for our VAE neural network.

VAE Model Development and Training. We chose to utilize the VAE functions available within the open-source PyTorch library for our machine learning model. This tensor library is an ideal option because it was designed to perform deep learning on irregular input data[4], such as that of our kymographs, and it was notably more reliable and easy to use compared to other Python libraries that we explored such as Tensorflow. Using native Pytorch linear algebra functions and convolutional neural networks, we created two methods: a variational encoder to create the lower-dimensional latent space of our data, and a decoder that attempts to reconstruct the original inputs from this latent space, reversing the dimensionality reduction. A third method initializes the actual VAE model by combining the generated encoder and decoder. Lastly, a training loop performs the optimization of this model by calculating the amount of information lost when the decoder rebuilds the input data as the sum of the mean squared error and Kullback–Leibler divergence between the two kymographs. The training loop then attempts to minimize this loss by varying encoder and decoder function parameters, thereby iteratively creating an "optimal" latent space[7] (**Figure 3A**).

To initially validate our model, we established a training set of data separate from the experimental set of data. For the training set, we used 300 artificially generated kymographs in order to train our model to classify three different hypothesized PlzC movement patterns. Accordingly, the artificial training set had an equal distribution of three possible graphical scenarios: 100 kymographs displayed oscillatory protein movement, 100 displayed polar protein movement, and 100 displayed negative or indiscernible protein movement (**Figure 3B**). We trained the model on this data, using 200 epochs and a two dimensional latent space. The results were plotted by retrieving data labels by sample ID.

Next, we trained the model using only the unlabeled, experimental kymographs. Again, we trained the model with 200 epochs. Reconstruction loss with two latent dimensions remained high, so we increased the size of the latent space to 256 dimensions. We then used UMAP dimensionality reduction to reduce this latent space to two dimensions so we could visualize the latent space and any data grouping.

RESULTS

PlzC proteins move super-diffusively within *Vibrio cholerae* cells. The diffusive speed of PlzC within VC was calculated using the Stokes-Einstein[17] relationship. We estimated the radius of PlzC via Debaye's formula (Eq. 1) and protein sequence data from UniProt (Uniprot Accession Q9KPM3) where N is the number of monomers in the polymer and b is the contour length of an amino acid[13].

$$(1) \ R^2 \ = \ \frac{Nb}{6} \ = \ 4.21 \text{nm}$$

We then substituted the calculated radius of 4.21nm into the Stokes-Einstein equation (Eq. 2) where K_B is Boltzmann constant, T is the temperature (in Kelvin), μ is the viscosity of the media, and R is the radius of the protein. Bacterial cytoplasmic viscosity was estimated to be 17.5 Pa•sec based on work by Schavemaker et al.[12]
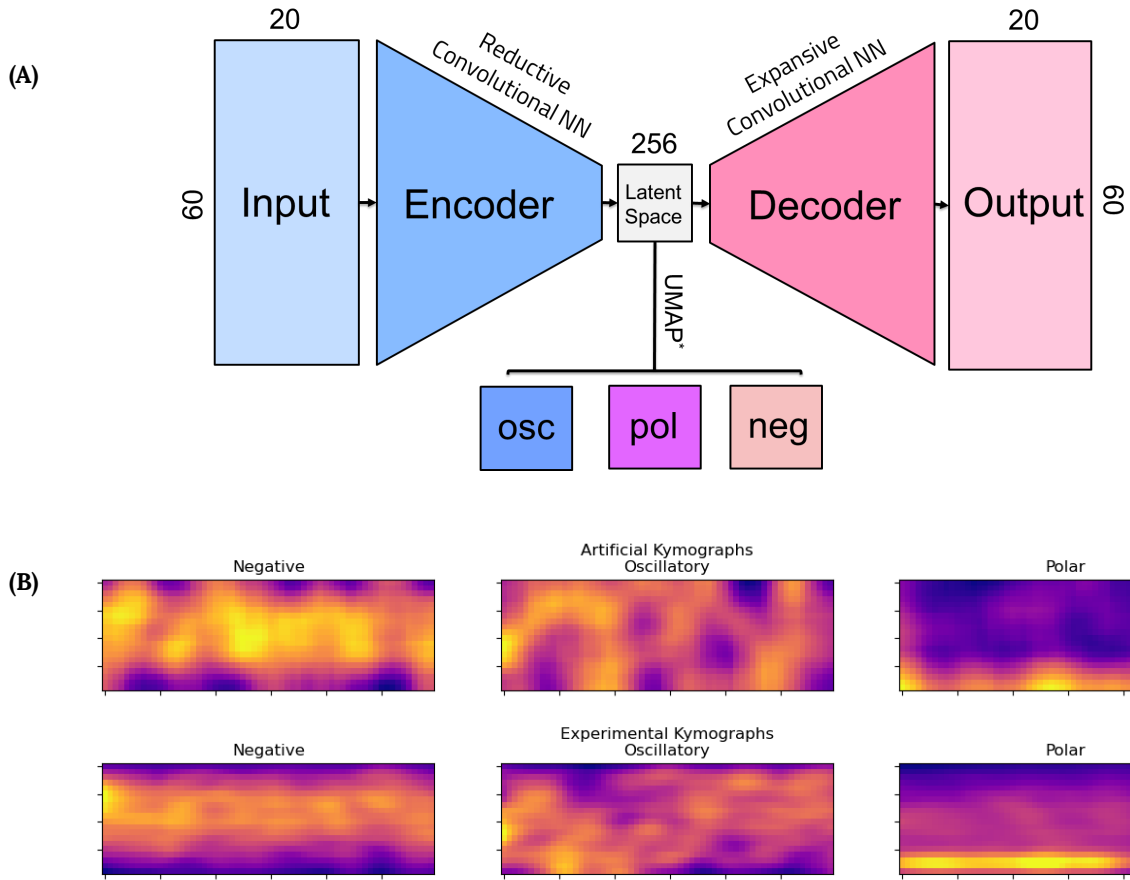
Figure 3. (A) VAE model architecture. Kymographs are fed into the encoder composed of 3 convolutional layers which successively transform kymograph by simultaneously increasing the output channels and reducing image dimensions. A linear layer compresses the input into the latent dimensions, from which the decoder (whose layers mirror those of the encoder) attempts to reconstruct the image. (B) Representative examples of negative, oscillatory, and polar artificial kymographs (top) and experimental kymographs (bottom). Contrary to initial expectations, the negative phenotype was not identified by the VAE, and the bottom left kymograph shown is actually classified as oscillatory by the VAE.

$$(2)\ D\ =\ \frac{K_b T}{6\pi\mu R}\ =\ 10\ \frac{\mu m}{s}$$

Finally, the time for PlzC to diffuse across the long axis of VC (3m) was calculated using Eq. 3 where x is the length that the molecule must travel, D is the diffusivity of the molecule and t is the time it takes for the molecule to move a distance x.

$$(3)\ t\ =\ \frac{x^2}{D}\ \approx\ 1\ \text{sec}$$

This led us to the conclusion that PlzC moving via diffusion alone would take ~1 second to travel from one end of the cell to the other.

Our VAE model revealed two dominant PlzC movement patterns. Unlike supervised learning models such as convolutional neural networks or nonlinear regressions, VAE models do not require a priori labeling of data in the training set. In situations where the relationship between the independent and dependent variables is unclear, VAEs are able to categorize data by closely grouping samples with similar characteristics within the latent space. We trained our model on 112 unlabeled wild type VC kymographs and used UMAP dimensionality reduction to visualize the encoded kymographs in the latent space of the model (**Figure 4**).

By sampling each group in the latent space, we found that PlzC adopts either upper polar, lower polar, or oscillatory behavior within the cell. Because upper and lower pole locations are assigned arbitrarily during data processing, we grouped both upper and lower polar behavior into a single polar group. Re-running the model using a range of hyperparameters and latent dimensions preserved the separation between oscillatory and polar movement patterns.

Dual regime training emphasized the regularity of PlzC movement.

To verify the accuracy of the model, we generated artificial kymographs using sinusoidally, linearly, and randomly distributed gaussian noise with the same mean and moments used to encode experimental data in the latent space. This procedure yielded kymographs visually indistinguishable from those produced using experimental data (**Figure 3B**). We then trained the VAE model on the artificial data alone (**Figure 5A**) and on the combined experimental and artificial data (**Figure 5B**) using two latent dimensions for interpretability. Across both datasets and models, the latent space encoded experimental data displayed significantly tighter grouping than the artificial data, despite the visual similarity between the two groups (**Figure 5A,B**).

Across all training methodologies, the experimental data was encoded into the same region of the latent space as the artificial oscillatory kymographs (**Figure 5B**). Even experimental kymographs which visually appeared to be polar were invariably encoded into the oscillatory region of the latent space regardless of the specific training used for the VAE model.
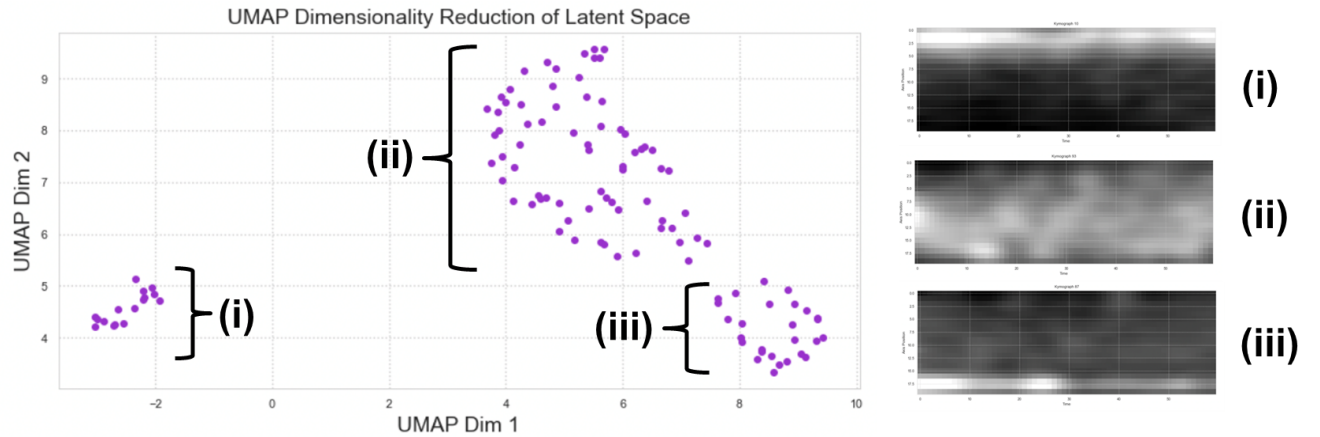

Figure 4. UMAP visualization of the latent embeddings of experimental kymographs produced by the VAE model trained on experimental data. Three primary groups were observed: (i) upper polar, (ii) oscillatory, and (iii) lower polar. Representative samples from each group are displayed to the right of the UMAP visualization.

DISCUSSION

We have shown that PlzC in V. *Cholerae* alternates between super-diffuse oscillatory movement and polar localization within the cell. Prior to building the VAE, we expected the data to fall into either oscillatory, polar, or negative movement patterns. Our VAE model indeed grouped the data into 3
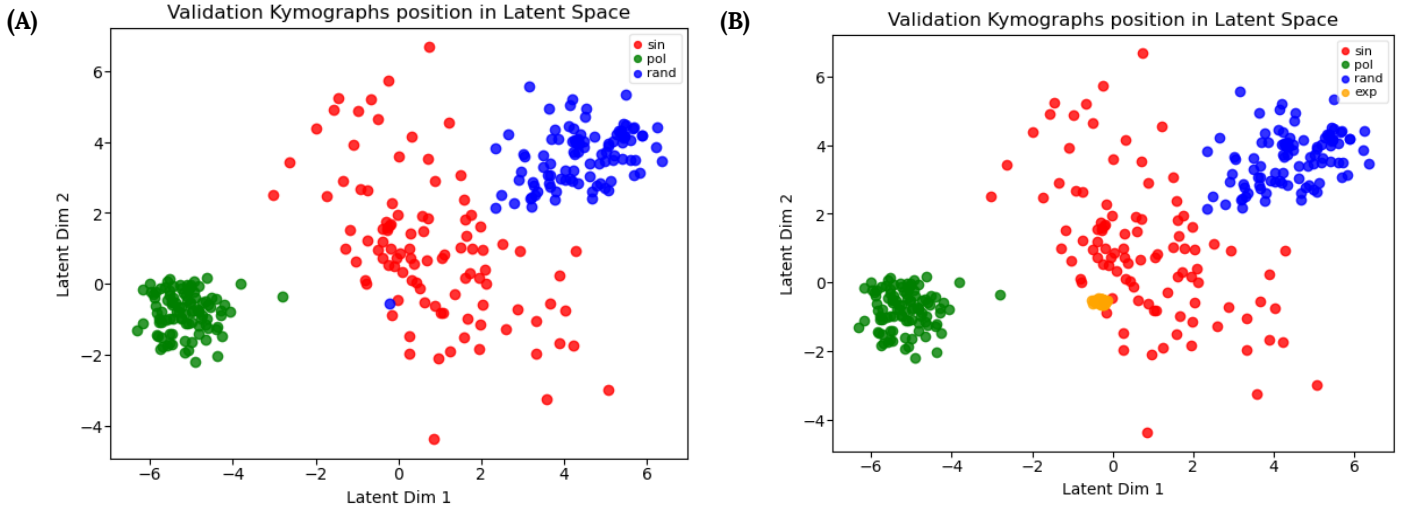
Figure 5. (A) Plotted embeddings of artificial kymographs produced by training the VAE with a combined artificial and experimental dataset and encoding down from 1200 to 2 points. Embeddings, colored according to motion type using sample ID, form clear groupings in the latent space and separate according to motion type as expected. (B) Experimental kymograph embeddings plotted with artificial data using the same encoder as plot (A). The experimental kymographs group more tightly than the artificial kymographs, but fall in the same region as the oscillatory (sin) artificial group.

groups. However, the groups corresponded to the oscillatory phenotype and two polar localizations with intensity concentrated in opposite ends of the cell. Notably, we did not observe a negative group as expected, demonstrating the advantage of using an unsupervised model as it rapidly and accurately identified the actual patterns in the data regardless of our initial hypothesis.

In addition to accelerating kymograph classification, the model is able to identify differences in data that are otherwise visually undetectable. In our comparison of the artificially generated and experimental kymographs, we found a significantly larger spread in generated data, while the experimental data concentrated on the mean of the encoding distribution in the latent space (**Figure 5**). This discrepancy in experimental and artificial variance in the latent space was preserved in models trained using different hyperparameters and different ratios of artificial to experimental training data. As such, we propose that the differing variance stems from the tight regulation of PlzC movement at the cellular level. This results in subtle differences which are not visually identifiable in the kymographs,

instead manifesting solely when the most data-rich elements of the kymograph are isolated by the VAE model. This tight regulation was not accounted for in the generation of the artificial data, hence the higher observed variance of the encoded artificial data in the latent space.

Besides classifying input data, the model has the ability to not only accurately reconstruct encoded input kymographs but also reconstruct randomly selected points within a relevant domain of the latent space. (**Figure 7**) The model was able to perform low-loss reconstruction (**Figure 7**) for both experimental (loss ≤ 23%) and artificial (loss ≤ 6%) inputs using a range of hyperparameters (**Figures 6, 7**). Together with the consistency between samples of latent space groups, these confirm the validity of the model as a rapid, scalable way to classify single cell protein movement.

It is important to note that the data processing pipeline and model itself only require fluorescence timelapses of the desired bacteria and the ability to identify a meaningful axis of molecular movement. As such, our pipeline and model present a "plug-and-play" solution for the analysis of

8

intracellular protein movement in a wide variety of single cell organisms. We plan to extend our analysis of PlzC movement by using this pipeline and model to track the protein under varying conditions in VC and also in *Pseudomonas aeruginosa*, a bacteria responsible for biofilm-related respiratory infections in cystic fibrosis patients[16].
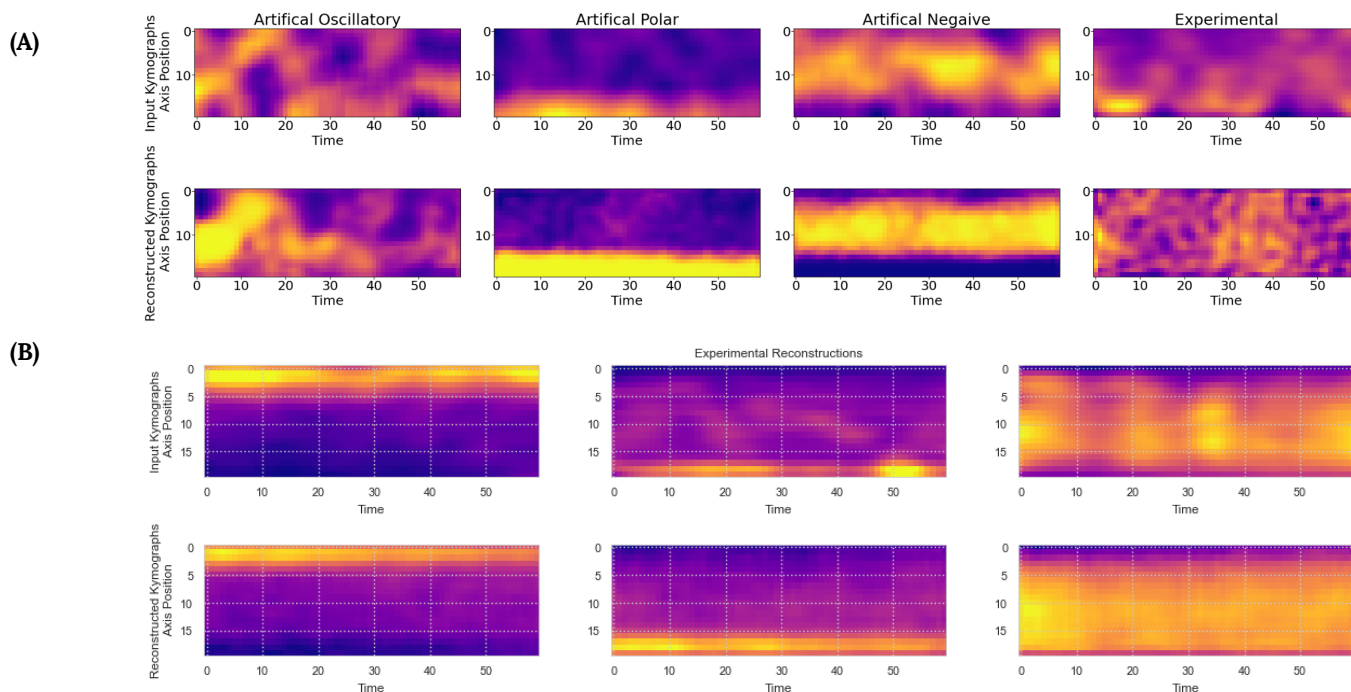
Figure 6. (A) Comparison of input kymographs (top row) and model reconstructions (bottom row) using 2 latent dimensions. Despite over 99.8% compression in the encoding phase (20x60 input data → 1x2 output data), the model is able to reconstruct both artificial and experimental kymographs with low loss (Combined MSE and KL artificial loss ≤ 6%, experimental loss ≤ 23%). (B) Training model solely with experimental data further decreases combined loss (experimental loss ≤ 5%). This results in reconstructions that are visually quite similar to the input kymographs, confirming the model's ability to encode meaningful and data-dense representations of the kymographs.
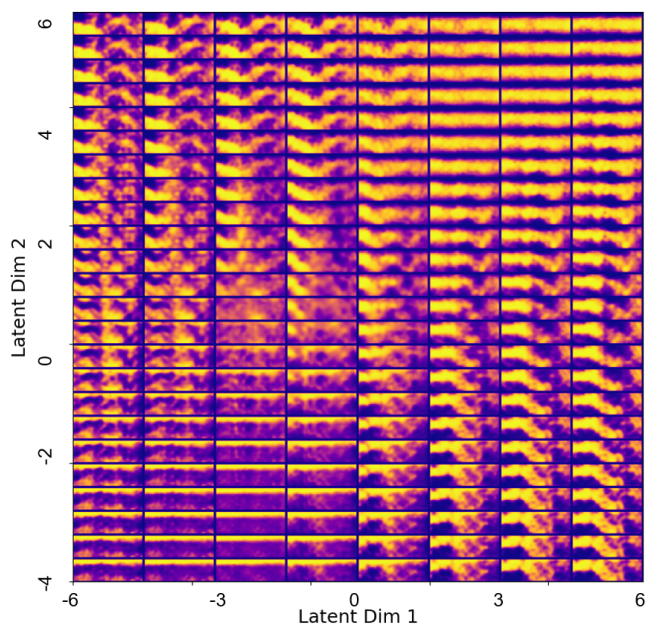
Figure 7. Decoder generated images. Equidistant pairs of points across the relevant latent space domain
$(L_1 \in [6, 6], L_2 \in [-4, 6])$ **were fed to the decoder portion of the VAE model to visualize encoding patterns. Points in the lower left region of the range produce polar kymographs while points in the upper right produce 'negative' kymographs. These reconstructions align with the embeddings of both artificial and experimental inputs as shown in Figure 5.**

CONCLUSION

Overall, our model can eliminate the bottleneck resulting from time-consuming manual analysis of

kymograph data. Furthermore, the use of this tool can be extrapolated to future single-protein tracking endeavors done by the Wong group and their collaborators in not only *Vibrio cholerae* but other bacterial species such as *Pseudomonas aeruginosa*.

The resulting VAE model and pipeline accelerated the single protein analysis needed for biofilm formation research. We were able to observe the movement of PlzC and its linkage with proteins that are associated with flagella formation and that have the ability of polar localization. The model and data processing pipeline gave us the ability to observe the movement of PlzC within the bacterial cell. This in conjunction with preliminary data such as the K_D value for PlzC and c-di-GMP complex formation allowed us to conclude that PlzC is indeed linked to biofilm formation. In the future, our model could be generalized to be used with other proteins and bacteria, as well as additional single protein tracking applications. We also will address issues with spline fitting, such as the inability to identify poles in spherical cells or struggles with fitting splines to cells with more complex shapes. We can also explore uses for artificial data created by the decoder submodel.

REFERENCES

(1) Sharma, Divakar, et al. Antibiotics versus Biofilm: An Emerging Battleground in Microbial Communities. *Antimicrobial Resistance & Infection Control* 8, no. 1, Dec. 2019, p. 76.

(2) Lin Chua, S., Liu, Y., Li, Y., Jun Ting, H., Kohli, G. S., Cai, Z., Suwanchaikasem, P., Kau Kit Goh, K., Pin Ng, S., Tolker-Nielsen, T., Yang, L., & Givskov, M. (2017). Reduced intracellular C-di-GMP content increases expression of quorum sensing-regulated genes in pseudomonas aeruginosa. *Frontiers in Cellular and Infection Microbiology* 7.

(3) Conner, J. G., Zamorano-Sánchez, D., Park, J. H., Sondermann, H., & Yildiz, F. H. (2017). The ins and outs of cyclic di-GMP signaling in vibrio cholerae. *Current Opinion in Microbiology* 36, 20–29.

(4) PYTORCH documentation. PyTorch 1.13 documentation. (n.d.). Retrieved February 6, 2023, from https://pytorch.org/docs/stable/index.html

(5) Uchida, Seiichi. "Image processing and recognition for biological images." Development, growth & differentiation vol. 55,4 (2013): 523-49. doi:10.1111/dgd.12054

(6) M. Akçakaya, B. Yaman, H. Chung and J. C. Ye, "Unsupervised Deep Learning Methods for Biological Image Reconstruction and Enhancement: An overview from a signal processing perspective," in IEEE Signal Processing Magazine, vol. 39, no. 2, pp. 28-44, March 2022, doi: 10.1109/MSP.2021.3119273.

(7) Dobilas, S. (2022, May 2). VAE: Variational autoencoders-how to employ neural networks to generate new images. Medium. Retrieved February 6, 2023, from https://towardsdatascience.com/vae-variational-autoencoders-how-to-employ-neural-networks-to-generate-new-images-bdeb216ed2c0

(8) qCMOS Cameras. Hamamatsu Photonics. (n.d.). Retrieved March 9, 2023, from https://www.hamamatsu.com/eu/en/product/cameras/qcmos-cameras.html

(9) Pratt, J.T. et al. (2007) "PILZ domain proteins bind cyclic diguanylate and regulate diverse processes in vibrio cholerae," Journal of Biological Chemistry, 282(17), pp. 12860–12870. Available at: https://doi.org/10.1074/jbc.m611593200.

(10) Conner, J.G. et al. (2017) "The ins and outs of cyclic di-GMP signaling in vibrio cholerae," Current Opinion in Microbiology, 36, pp. 20–29. Available at: https://doi.org/10.1016/j.mib.2017.01.002.

(11) Manteaux, P.-L. et al. (2016) "Adaptive physically based models in Computer Graphics," Computer Graphics Forum, 36(6), pp. 312–337. Available at: https://doi.org/10.1111/cgf.12941.

(12) Schavemaker, P.E., Boersma, A.J. and Poolman, B. (2018) "How important is protein diffusion in prokaryotes?," Frontiers in Molecular Biosciences, 5. Available at: https://doi.org/10.3389/fmolb.2018.00093.

(13) The UniProt Consortium, UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, 51, 6 Jan. 2023, p. D523–D531

(14) Aspelmeier, Timo & Egner, Alexander & Munk, Axel. (2015). Modern Statistical Challenges in High-Resolution Fluorescence Microscopy.

Annual Review of Statistics and Its Application. 2. 163-202.

(15) Xu, C., Jackson, S.A. Machine learning and complex biological data. Genome Biol 20, 76 (2019).

(16) Mulcahy LR, Isabella VM, Lewis K. Pseudomonas aeruginosa biofilms in disease. Microb Ecol. 2014 Jul;68(1):1-12. doi: 10.1007/s00248-013-0297-x. Epub 2013 Oct 6. PMID: 24096885; PMCID: PMC3977026.

(17) Miller Christina Cruickshank (1924). The Stokes-Einstein law for diffusion in solution Proc. *Royal Society London.* A106724–749 http://doi.org/10.1098/rspa.1924.0100