Bioinformatics, 2024, **40(4)**, btae170
https://doi.org/10.1093/bioinformatics/btae170
Advance Access Publication Date: 29 March 2024
Original Paper

Bioimage informatics

MotGen: a closed-loop bacterial motility control framework using generative adversarial networks

BoGeum Seo[1], DoHee Lee[2], Heungjin Jeon[3], Junhyoung Ha [2,*], SeungBeum Suh [2,*]

[1]Department of Mechanical Engineering, Seoul National University, 08826 Seoul, Republic of Korea
[2]Center for Healthcare Robotics, Korea Institute of Science & Technology, 02792 Seoul, Republic of Korea
[3]Infection Control Convergence Research Center, Chungnam National University, 34134 Daejeon, Republic of Korea

*Corresponding authors. Center for Healthcare Robotics, Korea Institute of Science & Technology, 5 Hwarang-ro 14-gil, 02792 Seoul, Republic of Korea. E-mails: jhha@kist.re.kr (J.H.) and keenhurt81@kist.re.kr (S.S.)

Associate Editor: Hanchuan Peng

Abstract

Motivation: Many organisms' survival and behavior hinge on their responses to environmental signals. While research on bacteria-directed therapeutic agents has increased, systematic exploration of real-time modulation of bacterial motility remains limited. Current studies often focus on permanent motility changes through genetic alterations, restricting the ability to modulate bacterial motility dynamically on a large scale. To address this gap, we propose a novel real-time control framework for systematically modulating bacterial motility dynamics.

Results: We introduce MotGen, a deep learning approach leveraging Generative Adversarial Networks to analyze swimming performance statistics of motile bacteria based on live cell imaging data. By tracking objects and optimizing cell trajectory mapping under environmentally altered conditions, we trained MotGen on a comprehensive statistical dataset derived from real image data. Our experimental results demonstrate MotGen's ability to capture motility dynamics from real bacterial populations with low mean absolute error in both simulated and real datasets. MotGen allows us to approach optimal swimming conditions for desired motility statistics in real-time. MotGen's potential extends to practical biomedical applications, including immune response prediction, by providing imputation of bacterial motility patterns based on external environmental conditions. Our short-term, in-situ interventions for controlling motility behavior offer a promising foundation for the development of bacteria-based biomedical applications.

Availability and implementation: MotGen is presented as a combination of Matlab image analysis code and a machine learning workflow in Python. Codes are available at https://github.com/bgmseo/MotGen, for cell tracking and implementation of trained models to generate bacterial motility statistics.

1 Introduction

Living organisms, including flagellated bacteria, can respond quickly to the surrounding environmental stimuli and control their taxis behaviors as their fundamental survival strategy. For example, bacteria exhibit biased random walk through precisely controlling rotary motors, which enables either run or tumble movement by rotating the helical filament clockwise or counterclockwise, respectively. These repeated modes of motility behavior enable bacteria to change direction in space in a favorable direction. Bacteria also show promising characteristics of being utilized as therapetuic agents, especially for cancer therapeutics, as they preferentially accumulate and proliferate in the tumoral region where the immune surveillance is not active (Zhou *et al.* 2018). Thus, the bacteria's intrinsic locomotive elements make bacteria a good platform for a delivery system. In addition to the bacteria's unique intrinsic locomotive elements, bacteria can be designed specifically as immune stimulators by utilizing their intrinsic pathogen-associated molecular patterns. Hence, when bacteria's unique actuation and immune stimulator are combined, the traditionally unreachable necrotic tumoral regions are now accessible and even have greater therapeutic output without external driving forces (Suh *et al.* 2018).

Even though bacteria demonstrate great potential, it is essential to secure precise locomotive control for better performance as therapeutic agents. Many seminal studies to enhance understanding of bacterial locomotion include analysis of a single cell or a whole bacteria population using microscopic imaging analysis and mathematical modeling (Sebag *et al.* 2015). For instance, the bacterial locomotion was varied via genetic engineering for phenotypical variations like flagella shape, number, and length on motility behaviors or differences in bacterial strains (e.g. *Escherichia coli, Bacillus Subtilius*) which significantly affected swimming behaviors (Turner *et al.* 2016). Bacterial motility behaviors are also governed by environmental conditions. For example, flagella motors adapt to persistent stimulus, and bacteria exhibit different motility behaviors under different pH (Maurer *et al.* 2005) and temperature levels (Rudenko *et al.* 2019). Even though genetically or chemically engineered bacteria provide various swimming modifications, these bacteria experience permanent and thereby irreversible changes. Considering versatile biological system, finding methods to control bacteria under short-term in-situ interventions under different environments for adaptive modes of motion are imperative.

Received: 20 December 2023; **Revised:** 2 March 2024; **Editorial Decision:** 25 March 2024; **Accepted:** 27 March 2024

Since the desired bacterial behavior is linked to the local environments and the feedback system between response regulators and flagella motors, constructing frameworks that link the environmental conditions and output behavior can provide predictions about the optimal strategies for the motility control of the bacteria (Emonet *et al.* 2005). Despite the vast literature on motility studies, we still have little access to the precise and versatile control strategies of the swimming motility of organisms in realistic biological environments. The numerical and analytical analysis of the linkage between external stimuli and motility performance has been based on a large dataset of genomic (Ramoneda *et al.* 2023) and proteomic sequences (Shin *et al.* 2021), yet only a little from experimental data. Developing experimented image-based datasets with quantifiable motility features is a difficult process as experimental determination often includes technical limitations and empirical optimization that are laborious, time-consuming, and expensive (Dubay *et al.* 2023). Therefore, such a dataset with quantifiable motility features with labeling is unavailable. A real data-based dataset is needed to find a confounding factor in the environment governing bacterial motility, particularly under physiological conditions. Since identifying the precise correlation between environmental conditions and specific motility characteristics through simple comparative analysis of humans is difficult, a machine learning algorithm employment is desired (Greener *et al.* 2021). While neural network models are used for their exceptional ability to handle complex nonlinear correlations, the model must capture the stochastic nature of bacterial moving behavior. Therefore, we used the Generative Adversarial Networks (GANs) that can generate synthetic data mimicking real data's stochastic features by training two models simultaneously through an adversarial process (Goodfellow *et al.* 2014). The recent advances of GAN are adopted in various biological fields, including image analysis (Wang *et al.* 2022), omics (Ahmed *et al.* 2021) and data augmentation (Kruitbosch *et al.* 2021) for better simulations and prediction tasks. The advances are from the essence of GAN's ability to capture the true biological features intact, which motivated us to further enhance and improve bacterial motility prediction.

In this study, we developed MotGen (motility generator) to investigate the impact of physiological pH and temperature conditions on bacterial motility behaviors under physiological conditions. Bacterial motility performance was characterized by recording and analyzing bacteria's trajectories at constant input surrounding stimuli. From the extracted motility features, we have identified key parameters that exert significant influence and can be effectively utilized as controllable variables within the motility control framework. To create simulated synthetic data, we directly used selected motility features as a low dimensional motility vector to train the MotGen. Using the generative model, we could extract information about the underlying behavior process from the swimming trajectories to examine the consequences of physiological environmental stimuli on bacterial motility parameters and model a control system for the operational surrounding system of desired motility behaviors. Our results suggest that our model can qualitatively reproduce the bacterial motility statistics focusing on physiological environments. We further confirmed the behavioral feedback system of open-loop and closed-loop control framework by comparing the pH and temperature changes with the model-estimated bacterial motility behaviors in a real-time manner. Additionally, we verified the ideal operational regime for practical biomedical applications of bacteria-mediated immune stimulation assay (Fig. 1).

To the best of our knowledge, for the first time, we present the real-time, versatile, and machine learning based control framework of bacteria motility, verified with the practical biological experimental setups. Using MotGen, we could easily obtain desired environmental conditions corresponding to the specific motility features. Our novel approach proposed here will serve as a stepping-stone for the development of precise motility control of bacteria for practical use in the study of biomedical applications including *in vivo* applications.

2 Materials and methods

2.1 Feature extraction from bacteria trajectories

All motility characterization parameters were obtained by computational analysis of recorded experimental video using custom MATLAB code (Mathworks, USA). Bacterial trajectories were tracked by comparing adjacent frames using an image analysis library 'object tracking software 2003' developed by the Berg group (Harvard). The algorithm identifies all moving objects that enter the field of view for the entire frame using the Image Analysis Toolbox, labels each identified object with the ID number, and saves a list of particles' locations (positions and frame numbers). The x–y coordinates of each bacterial trajectory were identified from the object tracking results, and key motility parameters regarding the bacteria's running and tumbling states were calculated to characterize the bacterial motility (Fig. 2). The bacteria often leave the depth of field as they move along the z-axis. This technical limitation was optimized by additional user modifications that compared moving cells in adjacent frames to connect cells based on the absolute distance, heading angles, and velocity.

The running and tumbling events (swimming bacteria altering course) were determined using the previously established 35°/point criterion (Turner *et al.* 2016). When the angle change is $\leq$35°/point for not $<$3 consecutive points, the bacterium is considered in a running state and for the tumbling state, when the angle is $>$35° for not $<$2 consecutive points. After identifying the run and tumble state of bacterial trajectories, we calculated the instantaneous velocity and angular change (°/point) following the methodology depicted in Fig. 2. Based on these calculations, we determined the overall average velocity, as well as the linear velocity and angular velocity for the top 10% of trajectories in each run and tumble state. Additional motility parameters such as the path length, displacement, directionality (Sahari *et al.* 2012), and the moving ratio were supplemented to better characterize detailed dynamic features of bacteria (Turner *et al.* 2016, Masuzzo *et al.* 2017).

The bacteria's total traveled length and the distance between the starting point and the endpoint were used to distinguish between motile and nonmotile bacteria. The suspended nonmotile bacteria are common observations that are often eliminated for motility analysis. To avoid tracking artifacts, tracks with a total displacement $<$20 μm and a displacement between two points $<$15 μm were considered nonmotile bacteria and eliminated. These described features are partly

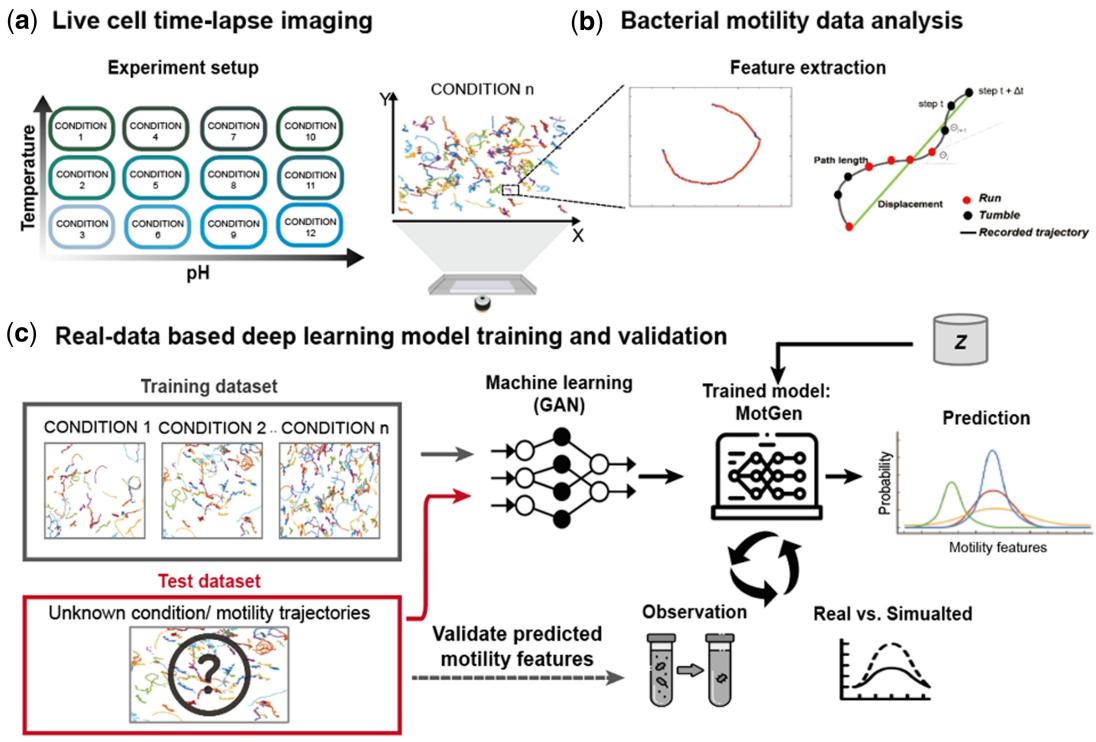

Figure 1. Schematic overview of the bacterial motility control workflow. (a) Experimental setup, (b) bacterial motility data analysis, and (c) real-data based deep learning model training and validation process.

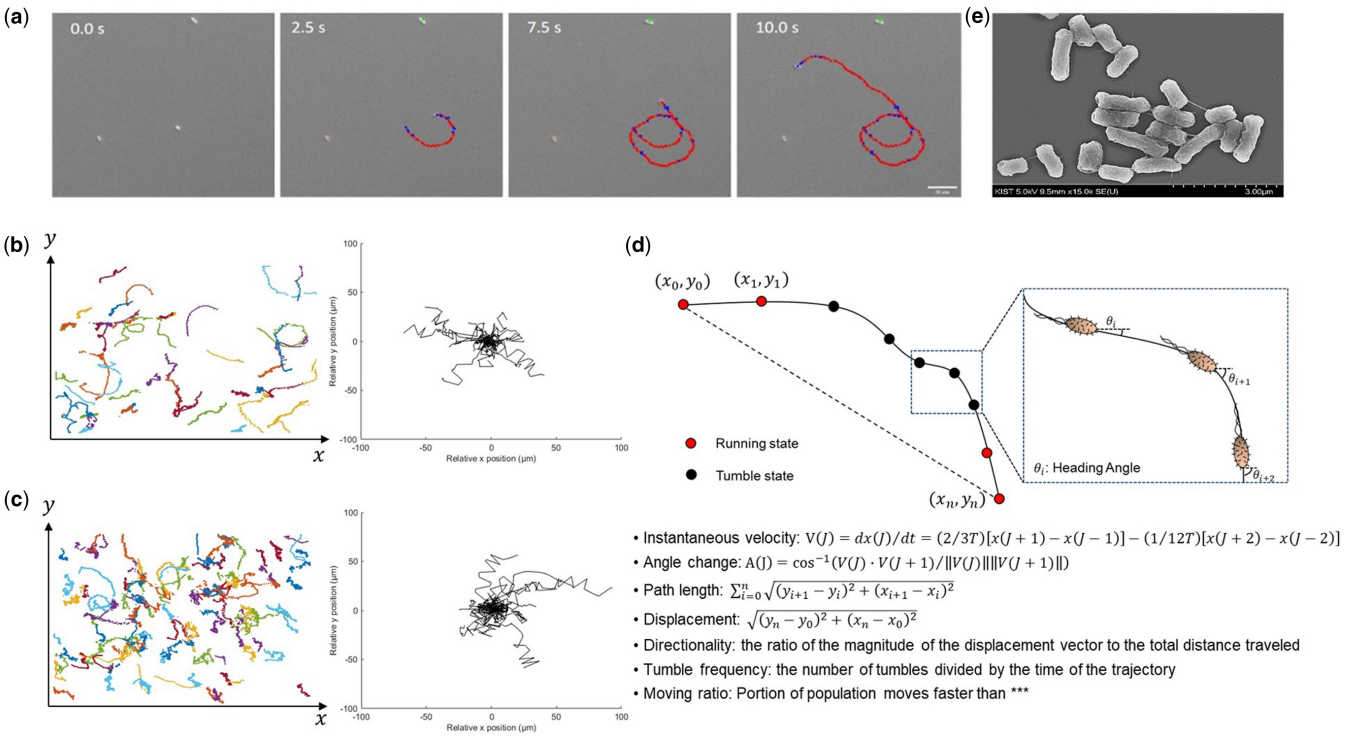

Figure 2. Object tracking result for image analysis. (a) Snapshots of a single bacterium's trajectory marked with running (red) and tumbling (blue) events; scale bar 20 μm. (b) and (c) Comparison of the two representatives tracked *E. coli* Nissle under the same temperature (37°) but different pH media: (b) pH 5, (c) pH 6. On the left, different colors of trajectory show the connected positions of a swimming bacterium. On the right, a few trajectories are randomly chosen and plotted at the origin. (d) Motility parameters extracted from a bacterium trajectory in the 2D space. (e) Scanning electron micrograph of flagellated *E. coli* Nissle. Scale Bar, 3 μm.

taken from seminal studies and some are newly designed and optimized. The overall statistical information on tracked motile bacteria under different experimental conditions was presented in Supplementary Table S1.

2.2 Training generative adversarial network architecture using bacterial trajectory data

Modeling bacteria mobility is challenging because they exhibit stochastic behaviors. For this reason, we utilize a generative model to capture the randomness of the bacteria mobility with its statistical characteristics. The nine motility parameters, extracted from the 9730 trajectories attained through bacteria motility characterization, were used as the training data to train the MotGen. The architecture of both networks consists of nine linear layers (Supplementary Fig. S3) and Leaky ReLU activations for all layers along with the Adam Optimizer with tuned hyperparameters to enhance the training process. The trained Generator network can be used as the generative model to create a new set of nine motility parameters by giving the environmental parameters (i.e. pH and temperature levels) along with a set of random numbers (i.e. latent vector) to the Generator. Once the model finishes training, the generated result is compared with a portion (10%) of training data and cross validated with newly experimented data in given conditions.

2.3 Statistical analysis

Statistical analyses such as the two-way ANOVA test and unpaired *t*-test were used to obtain *P*-values for data comparisons ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$). The data obtained from the 2D motility assay with pH and temperature change were performed ($n > 3$) to have enough dataset. The data for the immune response ELISA test was performed ($n = 3$), and the same statistical tests were performed.

3 Results and discussion

3.1 Characterizations of controllable motility variable determination

The motility characteristics based on the modes of bacterial movement between running and tumbling states significantly impact overall motility trends. For example, a flagellated bacteria alters running and tumbling mode by changing the bundle of flagella to change directions towards better living conditions (Fig. 2a and e). In this study, we examined in more detail the effect of environmental conditions on bacterial motility at various external temperatures and pH values, and apparent differences in motility patterns were found when physical or chemical environments change. The swimming track of *E. coli* Nissle and each swimming and tumbling point were reconstructed from their positions in successive frames in the film-recorded experiments (Fig. 2d). To verify the quality of image analysis output, a few motility parameters were compared with previously published data. For all testing conditions, the average swimming velocity of 25–30 μm/s was consistent with previously published data (Berg and Anderson 1973). The moving ratio of each environmental condition after the exclusion of nonmotile bacteria was consistent with each other, suggesting that the data-obtaining method was consistent with each repeated experiment and consisted of enough trajectory samples for all conditions to avoid possible data bias concerns. Nevertheless, the general motility pattern switches when the chemical environment changes using pH variations. For example, at pH 5, the swimming trajectories were generally smooth and slightly curved, while at pH 6, the trajectories showed frequent tumbling with frequent angular change as well as stronger curvatures (Fig. 2b and c). Bacteria exposed to higher pH surroundings showed lower displacement and path length but higher tumbling frequency, so more trajectories are centered at the origin (Fig. 2c). Interestingly, the pH levels are not linear rated to each motility feature, and each of them manifests different increasing or decreasing patterns. For instance, when the surrounding pH increases, the tumbling frequency decreases linearly, but the running velocity decreases and then increases (Fig. 3a and e).

The temperature change analysis shows that path length, displacement, running velocity, and angular velocity were significantly altered. The effect of increasing temperature led to much higher path length but a decrease in running velocities, which is expected as the FlhD level is down-regulated in the febrile temperature (Rudenko *et al.* 2019). Therefore, when the total swimming path length increased with increased running angular velocity and the tumbling angular velocity, the bacteria's beginning and end positions were further apart, which explains why these factors significantly differed with the temperature changes (Fig. 3b–d). This movement is consistent with the Ornstein-Uhlenbeck process, where the tumbling event starts by decreasing the running velocity value (Uhlenbeck and Ornstein 1930). However, it is important to note that the running velocities or angular velocities do not increase linearly with respect to temperature increase. In fact, for running and tumbling velocities, 25°C showed the highest activity, and 37°C showed the lowest activity (Fig. 3a and d).

For a different pH suspension of *E. coli* Nissle, the impact on tumbling-related motility parameters was much apparent. The changes in tumbling angular velocities and the tumbling frequency were most apparent when the lower and higher pH motility media were compared. This association is more evident at higher temperatures as the bacteria are known to have increased running velocity with a high tumbling rate (Dubay *et al.* 2022). Interesting to note that the tumbling frequency peaks almost up to 0.4 around 37°C at lower pH, possibly from an abrupt temperature change (Maeda *et al.* 1976). When bacteria were exposed to different pH conditions, the displacement and path lengths were significantly less different. For instance, the pH 8 condition shows the most statistical differences compared to the other pH levels, possibly from the increase in intracellular proton concentration interfering with the torque-generating units and leading to a strong correlation with path length and displacement parameters (Maurer *et al.* 2005).

By comparing each motility performance's mean and the variance that should be consistent and varied, we found motility features that exhibit nonlinear relationships with surrounding environments. In fact, changes in both pH and temperature will have a compound impact on motility features. For example, the angular and running velocities are expected to vary when the surrounding swimming conditions change. The other parameters, such as path length and displacement also change as a coupling effect. These complex, nonlinear relationships of running and tumbling-related parameters to the surroundings are reasons why we need a deep learning analysis to find stronger associations and systemically understand the bacteria behavior in our bodies. Therefore, we selected motility features to be trained for

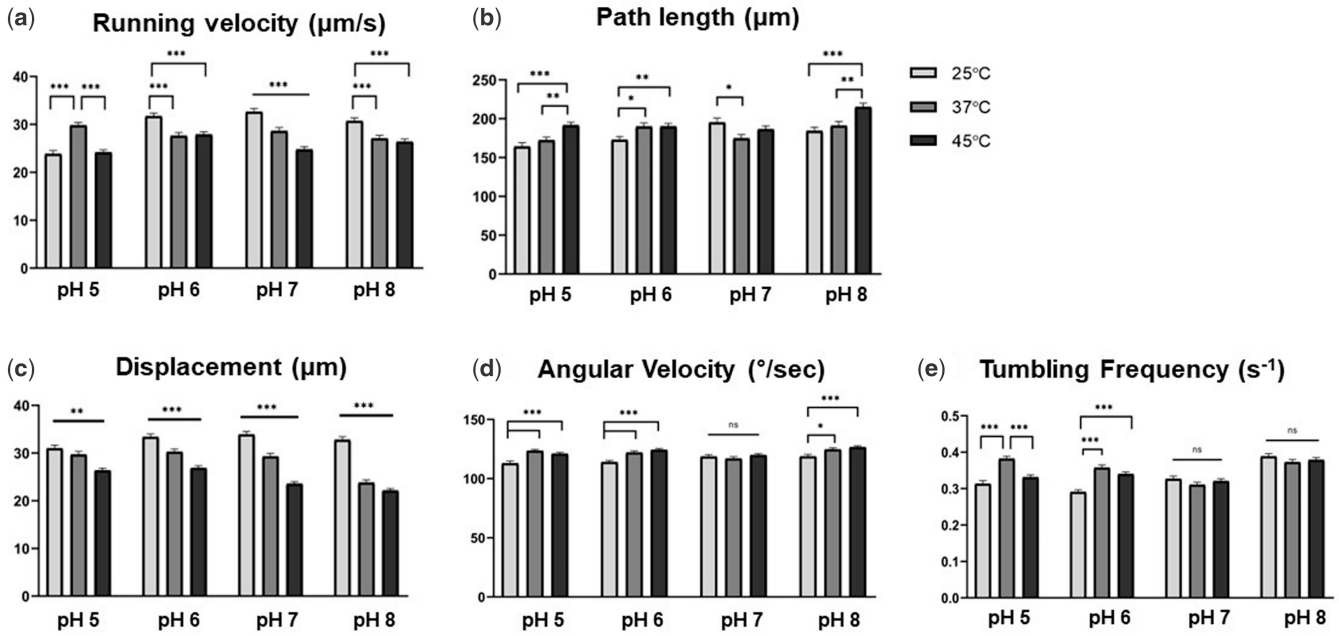

Figure 3. Statistics of moving bacterial motility. (a) Running velocity, (b) path length, (c) displacement, (d) angular velocity, and (e) tumbling frequency. The unpaired *t*-test and pairwise comparisons were performed and the data showing means and SEM. The representative controllable motility parameters are shown with $*P$-value < 0.05.

MotGen, which will interpolate the bacterial motility pattern under this environmental system.

3.2 Training and validation of generative adversarial networks

Since training GAN is known to be a difficult process, prone to common known problems such as mode collapse and gradient vanishing, choosing appropriate model hyperparameters is crucial in training quality. The data input, network architecture, activation function, and other hyperparameters were experimentally determined through validation process. First, our training dataset was labeled with the twelve environmental conditions as a context label to perform task-incremental learning for enhancing training process (de Ven *et al.* 2022). Also, unlike seminal studies on bacterial motility modeling which use simulated data, our approach significantly reduced computational efforts and training time by solely utilizing extracted stochastic quantities from real-data experiments (Mencattini *et al.* 2020, Zou *et al.* 2022). The GAN training data, collected from real experimental data, must be evenly distributed to mitigate biases. Supplementary Figure S1 illustrates the consistent skewed or symmetric behaviors observed across all environmental conditions (Dufour et al. 2016, Turner *et al.* 2016). Subsequently, MotGen successfully simulates these distributions, validated against both training and test data, as exemplified by path length comparisons across temperature ranges (Supplementary Fig. S2). The large dataset extracted from real experiments led to design of multi-layer perceptrons (MLPs) with higher feature dimensions of the hidden layers for a larger expressiveness. Also, the latent dimension's vector size (=300) was chosen empirically to capture the input data's characteristic by comparing the loss function. To enhance training quality, we used Rectified Linear unit (ReLU) activation function instead of conventional sigmoid or hyperbolic tangent functions. The batch normalization was applied for all layers to increase training rate while stabilizing the training process.

For the validation, we used the standard adversarial loss function in a Min-Max objective along with the generative loss, Kullback-Leibler divergence (KLD), and a 3D data plot of the input and synthesized data. The traditional min–max loss function refers to the simultaneous optimization of the probability distribution of the generated data and the real input data, which can be achieved by minimizing the generator's loss and maximizing of the discriminator's loss. Both generator and discriminator show typical training characteristic, where the generator loss stabilizes while the discriminator loss decreases gradually, and approximately converged around 1.34 around epoch 200 (Fig. 4a).

However, such loss function cannot represent the training quality as the loss function for the generator is prone for saturation problem and therefore should be monitored with KLD, which quantifies statistical distance using divergence scores between two probability distributions (Kullback and Leibler 1951). When the KLD values decrease, the divergence between the discriminator and generator also decreases as the discriminator cannot distinguish the real and simulated data, and our data shows decrease in KLD values corresponding with falling discriminator loss (Fig. 4b). The KLD values are sufficient indicator for typical GAN issues like nonconvergence, mode collapse, and diminished gradients but we additionally checked changes in gradients and saddle points to confirm the training process of adversarial networks (Fig. 4b). We tracked a latent variable's transformation process through the repeated generating and adversarial process to have the desired transforming gradient as the gradients will not change when mode collapse occurs (Thanh-Tung and Tran 2020). The KLD values (*y*-axis) with the number of epochs (*x*-axis) and few representative gradient changes were visualized to show the change in global minimum as training progressed. Around epoch 200, the global minimum

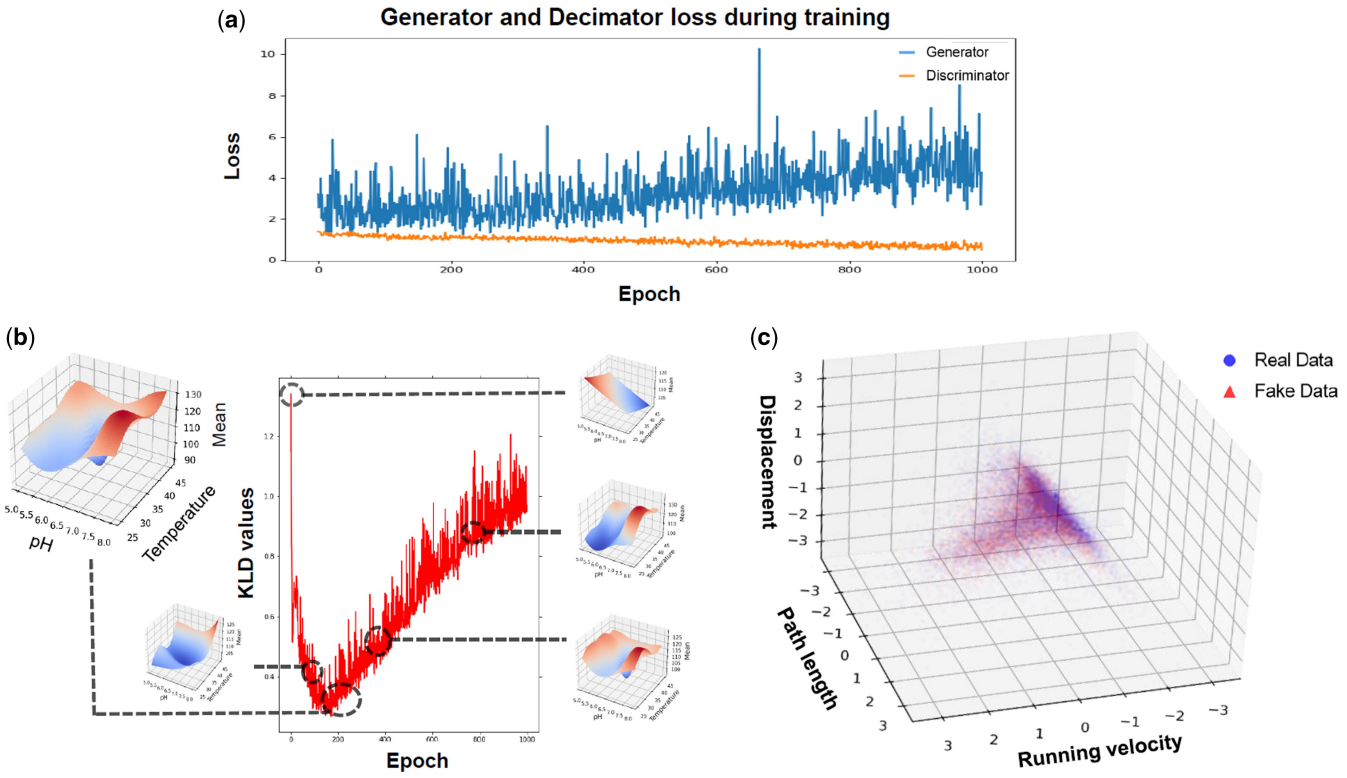

Figure 4. MotGen training (a) The evolution of the loss function from the generator and discriminator are shown in blue and orange respectively with the number of epochs (*x*-axis). (b) The comparison between KLD (*y*-axis) values for the number of epochs and gradients of the latent space along the training process. As training progresses, until the early-stopping point, the KLD values decrease, and the gradients change accordingly. (c) The plot of generated versus real data of selected motility parameters. The generated data shows a similar data manifold as the blue dot representing the empirical data.

gradient, the KLD values, and the loss function are ideal and stabilized and considered as early-stopping criteria for the training model without any GAN training issues (Fig. 4b).

Lastly, we verified our trained model by simultaneously fitting the real experimental data and the generated data by plotting in 3D space. When appropriately parameterized, the samples from real data and the samples produced by the MotGen will be nearly identical. The real data (blue dots) and the generated data (red dots) are fitted well with each other confirming that our final model can quantitatively reproduce the motility parameters of given environmental input conditions without any common training failure like a mode collapse (Fig. 4c). The well-trained generator model for bacterial motility parameters is used to create a large set of datasets for predictions using feedback systems and biological models.

3.3 Open-loop control system

To verify that our trained model can generate the nonlinear motility characteristics of given environmental conditions, we designed an open-loop system with a fixed pH level (pH 5 and 8) and varied temperature levels (Fig. 5a). Through this experimental setup, we demonstrated how the dynamic relationships of temperatures govern a specific motility parameter without any feedback system. We observed nonlinear increase in pH 5 and a linear increase in pH 8, as expected from the previous characterization result (Fig. 5b). Finally, we generated bacterial motility data based on conditions of the open-loop setting. Our well-trained MotGen well captured the nonlinearity of running velocities in different pH models. The model also simulated fast convergence of path length features under both pH conditions. The real experimented data from the open-loop control setup and the simulated data match quite closely and are not significantly different, suggesting that our generator model is well-trained to capture substantial bacterial motility behavior under various conditions and suitable to further investigate the behavior of bacterial motility under practical biological experimental setup.

3.4 Microbial immune response prediction based on the bacterial motility pattern

Bacteria have specific molecular patterns such as lipopolysaccharides from the outer membrane and flagellin from the flagella that can trigger chronic inflammation by increasing exposure towards host cells and induce invasion ability (Chaban *et al.* 2015). Tumor necrosis factor α (TNF-α) is often used as indicators of inflammation which is upregulated in a tumor microenvironment triggering leukocyte activation. Since the active immune response is directly proportional to bacterial swimming as a function of flagellar rotation (Lovewell *et al.* 2011), we hypothesized that MotGen could contribute to better prediction of the immune activation process using different motility patterns (Fig. 6a).

To investigate the extent of macrophages' response to bacteria with different motility trends, we exposed the tuned-motile bacteria of four different environmental conditions to the macrophages and analyzed the cytokine concentrations. As shown in Fig. 6b, the bacteria-conditioned group exhibited a significant increase compared to the same conditioned

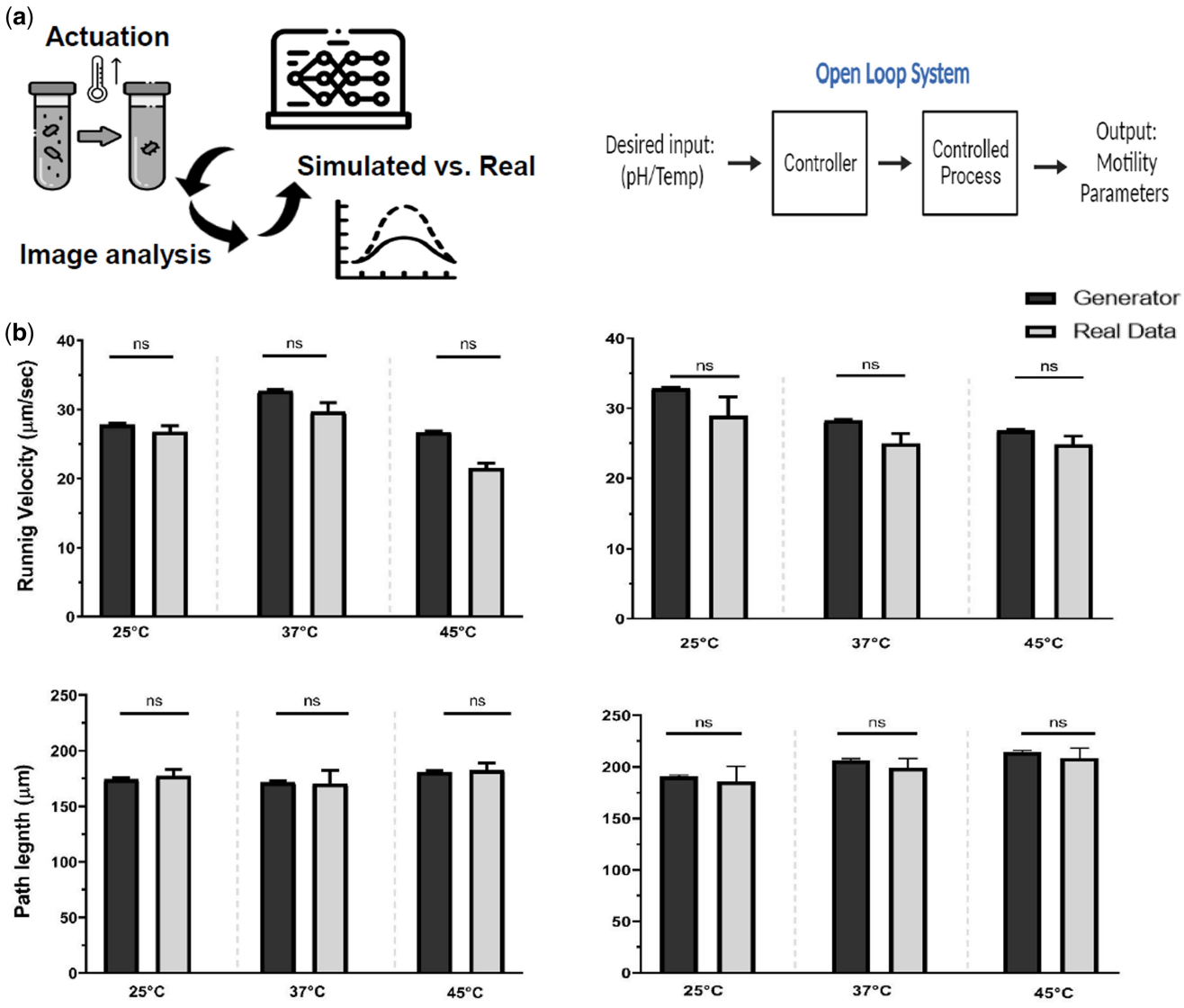

Figure 5. Open loop control system to predict and verify bacterial motility characteristics. (a) Schematic of open-loop control system. A sample population of bacteria was analyzed periodically to monitor the motility characteristics while the temperature is increased at a constant rate (0.5°C/min). (b) The motility behaviors of the increased temperature in an open-loop setting are compared between two different pH conditions: pH 5 (left) and pH 8 (right) and showing means and SEM.

control group. The increase of cytokine release from bacteria interaction is anticipated. More interestingly, bacteria under different ranges of pH media consist of different motility patterns and have different impacts on cytokine release. According to our generated data, the higher the pH level, the more running and tumbling activities increase, ultimately stimulating more macrophages. Each bacterium's different running and tumbling state is known to have a different flagellar bundle formation that leads to different tumble frequencies and distinct trajectory paths leading to higher instigation (Horstmann *et al.* 2017). Also, the bacteria exposed to lower pH levels (pH <7) have shorter path lengths and running and tumbling velocities, which lead to relatively moderate cytokine release (Fig. 6b). These motility patterns are consistent with the analyzed and predicted overall motility trends in given environments. The cytokine analysis revealed that different modes of bacteria motile levels with different tumbling and running patterns can induce different amounts of pro-inflammatory cytokine, indicating that contact frequency of immune activators affects cytokine secretion.

Phagocytic events are another indication of immune response to microbial invasion as a complementing process from the contact frequency of microbial and inflammatory phases. Once the LPS-stimulated macrophages are exposed to the microbial, the following anti-inflammatory and phagocytic process occurs (Fadok *et al.* 1998). Previous studies have confirmed that phagocytic cells responded directly to the flagellar torque of the bacteria and less phagocytosis was observed in less-active bacteria (Lovewell *et al.* 2011). Compared to the samples of a noninfected control, cells exposed to the pH 5 media consisting of low-tumbling bacteria had fewer phagocytic events than those exposed to the pH 7 media consisting of high-tumbling bacteria. The different levels of phagocytosis from the different tumbling levels of swimming bacteria are clearly visible by comparing the fluorescent signal intensities of bacteria interior of the macrophage cells (Fig. 6c). Phagocytic activities were calculated

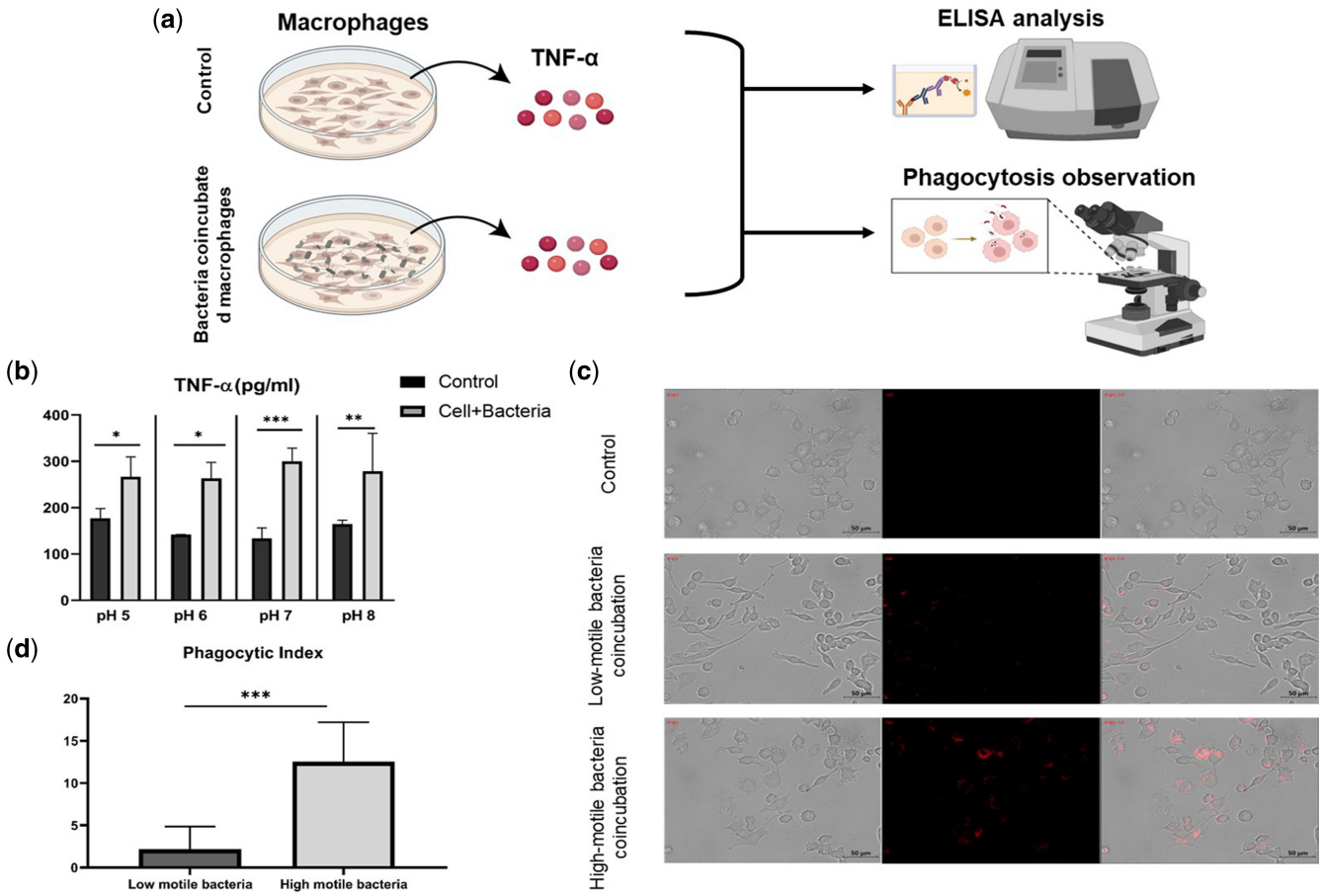

Figure 6. Response of macrophage cells to MotGen-predicted motility patterned bacteria. (a) Effects of tuned-motile bacteria on immune response by macrophages. (b) Macrophage response was analyzed using ELISA assay in response to a different mode of bacterial motility. (c) The composite microscopic images show the activated macrophages after phagocytosis of red fluorescent bacteria. (d) Phagocytic index of macrophages co-incubated with different pH media of swimming bacteria.

using the mean number of intensities per phagocytic cells (Barbuddhe *et al.* 1998). The phagocytic index of the cells co-incubated with low-motile bacteria samples (2.19 ± 1.68) showed significant differences from those co-incubated with high-motile bacteria samples (12.5 ± 4.66). These inflammatory events indicate that even though not perfectly explained; there is a clear indication that a distinct motility behavior results in differences in microbial invasion and thus explains the different cytokine release. Therefore, we are optimistic that our model can be utilized for a tunable interaction of bacteria with host immune factors to better predict host-based therapeutic approaches to intracellular pathogens.

3.5 Feedback closed-loop control system

A living cell is difficult to control as the system requires high dimensionality with full consideration of noise and uncertainty from many internal and external factors (Dufour *et al.* 2014). Nonetheless, the Generative Adversarial Network can mimic high-order distribution modeling and is often applied to the control system requiring highly ordered data. Thereby, we can find the desired operational environmental conditions when the desired motility parameter is given by fully integrating complex relations of multiple input parameters to the feedback control system using the generator model. To this aim, we set the experimental parameters (pH and temperature levels) as MotGen predicts the experimental setup in each repeated feedback loop, and the output motility feature is returned back to the system for minimizing system errors (Fig. 7a).

The feedback controller begins with the randomly selected initial motility distributions from the actual experimented data and tuning towards desired motility values by using the backpropagation algorithm to obtain a new environmental condition as a new input for the feedback system. We used the Adam optimizer function, the most common first-order gradient-based stochastic optimization, to calculate the gradients for the backpropagation process. We empirically tuned the hyper-parameter conditions and found the best learning rate ($lr = 1e{-}5$) and weight decay ($wd = 1e{-}6$). Because the system will have a slight oscillation, to minimize this error disturbance, choosing appropriately small step sizes can minimize the cost function effectively. To demonstrate our closed-loop system, we began with the initial condition of pH 6.0 and 37° motility parameters and set the desired goal for the directionality parameter to increase by 20% while running velocity to decrease by 15%. After the process of the backpropagation algorithm, the inverse Jacobian matrix is used to return the new temperature and pH values to update for another feedback loop. Each feedback loop's step size is different and chosen empirically because the model follows the local gradient; e.g. when the step size is too large, the returned values experience some fluctuations that are out of our experimental condition range. With the suitable step size

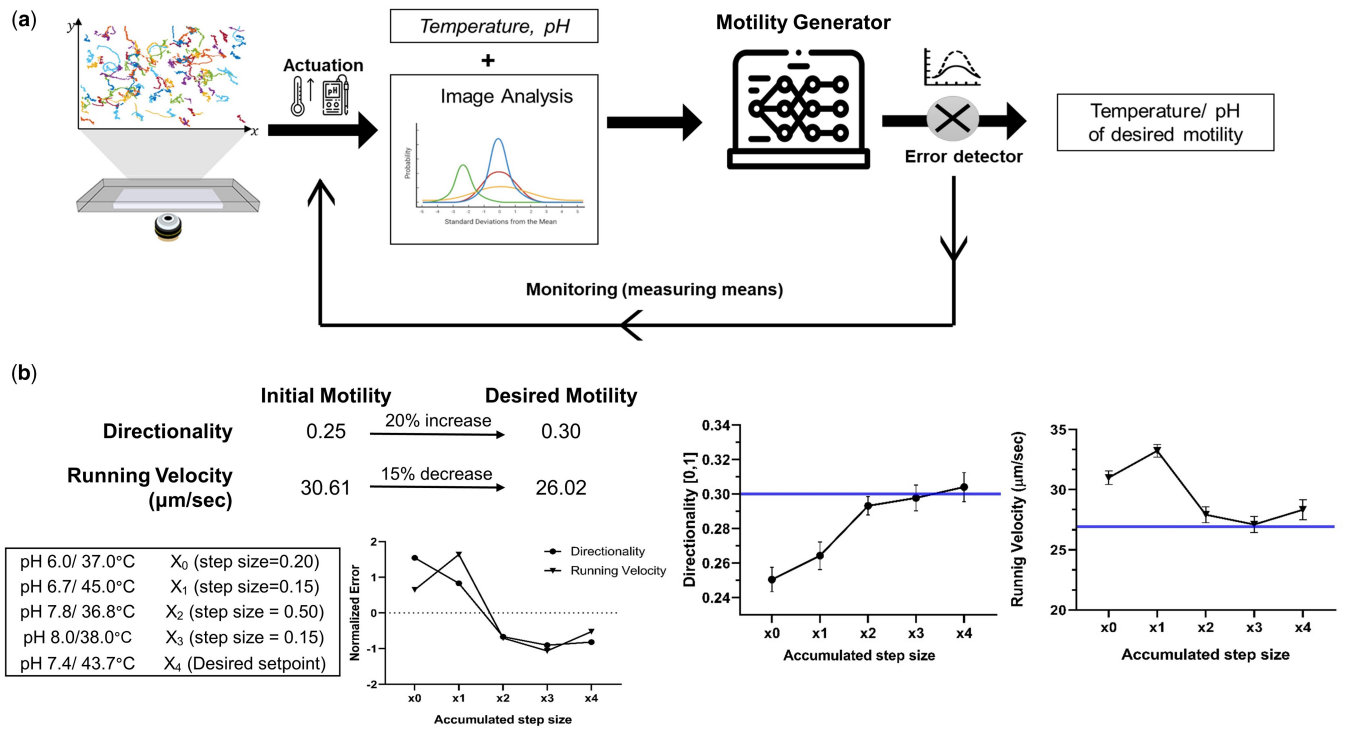

Figure 7. Closed-loop system to control bacterial motility in real-time. (a) Schematic of closed loop system. A set of input motility features are used to simulate the anticipated motility characteristics within a closed-loop environment with a defined target setpoint. (b) Starting from the initial condition, the resultant motility features (directionality and running velocity) of each repeated feedback iteration is plotted. Target setpoints are marked with blue lines. The error of each iteration was normalized.

values and multiple feedback iterations, the model returned pH and temperature conditions that are closer to the target motility values (Fig. 7b). The system error, which is the difference between the input value and the feedback value, is returned to the controller. After four iterations, the system error decreased close to zero.

In conclusion, we successfully demonstrated our quantitative predictive model and the experimental validation in the feedback control system for accurately predicting and controlling multiple motility parameters. In addition, the fact that changes in swimming conditions that do not require bacterial genetic changes can accurately and predictively alter bacterial motility characteristics will help further research into practical applications for controlling bacterial movement under biological settings in real-time.

4 Conclusion

To date, the diversity of motility behavior studies is increasing, yet the full behavioral diversity has yet to be elucidated. Because bacterial motility is an important indication of local environment information, as well as an indication of the entire population, the study on individual and collective behavior of bacterial motility is important. In this paper, we simultaneously varied pH and temperature to resemble a human physiological system and characterized *E. coli* Nissle motility behaviors. Through the image analysis, we found that pH and temperature can have independent and compound impacts on certain motility features. The direct comparison between quantitatively mapped motility parameters of each environmental condition was analyzed via machine learning-driven analysis and modeled using MotGen. By training the generative model to mimic the distribution of the bacterial motility parameters, we could independently predict certain environmental conditions that affect bacterial motility performance parameters' mean and variance. Our trained model was verified using the open-loop control system by comparing the real and simulated data. The relation between the controlled-motility bacteria and the immune cells demonstrates the promising biomedical application of machine learning-based analysis to living organisms. Finally, predictive and controllable bacteria motility systems are possible under the different physiological systems via a closed-looped feedback control system.

Through MotGen, we are optimistic to demonstrate the possibility of tuning the bacterial motility without altering the genetic variations and predicting swimming conditions under physiological systems to enhance the precise motility control of bacteria for practical biomedical applications. MotGen innovatively integrates two pivotal aspects in bioinformatics: firstly, it utilizes real biological data to inform the design of the GAN network, ensuring a robust model grounded in actual biological phenomena that are relevant to physiological conditions; secondly, it achieves closed-loop control of real biological systems, enabling an understanding of the intricate relationships governing bacterial behavior and facilitating accurate predictions of their dynamic responses. Notably, the effect of multiple external stimuli on motility features was predictable. Our study on the ecological role of motility behaviors meets an important avenue for future research that helps us to utilize bacteria in bioengineering fully.

Acknowledgements

The bacteria strain *E.coli* Nissle 1917 (Ardeypharm GmbH, Herdecke, Germany) used for the immune activation assay

was integrated with a gene cassette coding for a red fluorescent protein (smURFP and HO-1) by Yuhyun Ji and Hak Suk Chung from Korea Institute of Science and Technology. B.S. is grateful for financial support from Hyundai Motor Chung Mong-Koo Foundation.

Author contributions

B.S. conducted data acquisition, conceptualization, methodology, experimental testing, analyzed the results, and wrote the manuscript. D.L. did experimental testing. H.J confirmed methodology. J.H. supervised validation and reviewed the manuscript. S.S helped conceptualization, methodology, formal analysis, and supervised writing.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported by grants from Korea Institute of Science and Technology [2E32981] and from the National Research Foundation of Korea [2022M3H4A1A03076638].

References

Ahmed KT, Sun J, Cheng S *et al.* Multi-omics data integration by generative adversarial network. *Bioinformatics* 2021;**38**:179–86. https://doi.org/10.1093/bioinformatics/btab608.

Barbuddhe S, Malik S, Gupta LK. Effect of in vitro monocyte activation by listeria monocytogenes antigens on phagocytosis and production of reactive oxygen and nitrogen radicals in bovines. *Vet Immunol Immunopathol* 1998;**64**:149–59. https://doi.org/10.1016/s0165-2427(98)00129-9.

Berg HC, Anderson RA. Bacteria swim by rotating their flagellar filaments. *Nature* 1973;**245**:380–2. https://doi.org/10.1038/245380a0.

Chaban B, Hughes HV, Beeby M. The flagellum in bacterial pathogens: for motility and a whole lot more. *Semin Cell Dev Biol* 2015;**46**:91–103. Elsevier.

de Ven GMV, Tuytelaars T, Tolias AS. Three types of incremental learning. *Nat Mach Intell* 2022;**4**:1185–97. https://doi.org/10.1038/s42256-022-00568-3.

Dubay MM, Johnston N, Wronkiewicz M *et al.* Quantification of motility in bacillus subtilis at temperatures up to 84°C using a submersible volumetric microscope and automated tracking. *Front Microbiol* 2022;**13**:836808. https://doi.org/10.3389/fmicb.2022.836808.

Dubay MM, Acres J, Riekeles M *et al.* Recent advances in experimental design and data analysis to characterize prokaryotic motility. *J Microbiol Methods* 2023;**204**:106658. https://doi.org/10.1016/j.mimet.2022.106658.

Dufour YS, Gillet S, Frankel NW *et al.* Direct correlation between motile behavior and protein abundance in single cells. PLoS Comput Biol 2016;**12**:e1005041.

Dufour YS, Fu X, Hernandez-Nunez L *et al.* Limits of feedback control in bacterial chemotaxis. *PLoS Comput Biol* 2014;**10**:e1003694. https://doi.org/10.1371/journal.pcbi.1003694.

Emonet T, Macal CM, North MJ *et al.* AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics* 2005;**21**:2714–21. https://doi.org/10.1093/bioinformatics/bti391.

Fadok VA, Bratton DL, Konowal A *et al.* Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-beta, PGE2, and PAF. *J Clin Invest* 1998;**101**:890–8. https://doi.org/10.1172/jci1112.

Goodfellow I, Pouget-Abadie J, Mirza M *et al.* Generative adversarial nets. In: *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 2, 2014, 2672–2680.

Greener JG, Kandathil SM, Moffat L *et al.* A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2021;**23**:40–55. https://doi.org/10.1038/s41580-021-00407-0.

Horstmann JA, Zschieschang E, Truschel T *et al.* Flagellin phase-dependent swimming on epithelial cell surfaces contributes to productive salmonella gut colonisation. *Cell Microbiol* 2017;**19**:e12739.

Kruitbosch HT, Mzayek Y, Omlor S *et al.* A convolutional neural network for segmentation of yeast cells without manual training annotations. *Bioinformatics* 2021;**38**:1427–33. https://doi.org/10.1093/bioinformatics/btab835.

Kullback S, Leibler RA. On information and sufficiency. *Ann Math Statist* 1951;**22**:79–86.

Lovewell RR, Collins RM, Acker JL *et al.* Step-wise loss of bacterial flagellar torsion confers progressive phagocytic evasion. *PLoS Pathog* 2011;**7**:e1002253. https://doi.org/10.1371/journal.ppat.1002253.

Maeda K, Imae Y, Shioi JI *et al.* Effect of temperature on motility and chemotaxis of *Escherichia coli*. *J Bacteriol* 1976;**127**:1039–46. https://doi.org/10.1128/jb.127.3.1039-1046.1976.

Masuzzo P, Huyck L, Simiczyjew A *et al.* An end-to-end software solution for the analysis of high-throughput single-cell migration data. *Sci Rep* 2017;**7**:42383. https://doi.org/10.1038/srep42383.

Maurer LM, Yohannes E, Bondurant SS *et al.* ph regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* k-12. *J Bacteriol* 2005;**187**:304–19. https://doi.org/10.1128/JB.187.1.304-319.2005.

Mencattini A, Giuseppe DD, Comes MC *et al.* Discovering the hidden messages within cell trajectories using a deep learning approach for in vitro evaluation of cancer drug treatments. *Sci Rep* 2020;**10**:7653. https://doi.org/10.1038/s41598-020-64246-3.

Ramoneda J, Stallard-Olivera E, Hoffert M *et al.* Building a genome-based understanding of bacterial ph preferences. *Sci Adv* 2023;**9**:eadf8998. https://doi.org/10.1126/sciadv.adf8998.

Rudenko I, Ni B, Glatter T *et al.* Inefficient secretion of anti-sigma factor flgm inhibits bacterial motility at high temperature. *iScience* 2019;**16**:145–54. https://doi.org/10.1016/j.isci.2019.05.022.

Sahari A, Headen D, Behkam B. Effect of body shape on the motile behavior of bacteria-powered swimming microrobots (bacteriabots). *Biomed Microdevices* 2012;**14**:999–1007. https://doi.org/10.1007/s10544-012-9712-1.

Sebag AS, Plancade S, Raulet-Tomkiewicz C *et al.* A generic methodological framework for studying single cell motility in high-throughput time-lapse data. *Bioinformatics* 2015;**31**:i320–i328.

Shin J-E, Riesselman AJ, Kollasch AW *et al.* Protein design and variant prediction using autoregressive generative models. *Nat Commun* 2021;**12**:2403. https://doi.org/10.1038/s41467-021-22732-w.

Suh S, Jo A, Traore MA *et al.* Nanoscale bacteria-enabled autonomous drug delivery system (nanobeads) enhances intratumoral transport of nanomedicine. *Adv Sci (Weinh)* 2018;**6**:1801309. https://doi.org/10.1002/advs.201801309.

Thanh-Tung H, Tran T. Catastrophic forgetting and mode collapse in GANs. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 2020, 1–10, https://doi.org/10.1109/IJCNN48605.2020.9207181.

Turner L, Ping L, Neubauer M *et al.* Visualizing flagella while tracking bacteria. *Biophys J* 2016;**111**:630–9. https://doi.org/10.1016/j.bpj.2016.05.053.

Uhlenbeck GE, Ornstein LS. On the theory of the brownian motion. *Phys Rev* 1930;**36**:823–41. https://doi.org/10.1103/PhysRev.36.823.

Wang J, Tabassum N, Toma TT *et al.* 3D GAN image synthesis and dataset quality assessment for bacterial biofilm. *Bioinformatics* 2022;**38**:4598–604. https://doi.org/10.1093/bioinformatics/btac529.

Zhou S, Gravekamp C, Bermudes D *et al.* Tumour-targeting bacteria engineered to fight cancer. *Nat Rev Cancer* 2018;**18**:727–43. https://doi.org/10.1038/s41568-018-0070-z.

Zou Z, Liu Y, Young Y-N *et al.* Gait switching and targeted navigation of microswimmers via deep reinforcement learning. *Commun Phys* 2022;**5**:158. https://doi.org/10.1038/s42005-022-00935-x.

Published by Oxford University Press 2024.
Bioinformatics, 2024, 40, 1–11
https://doi.org/10.1093/bioinformatics/btae170
Original Paper