Applying machine learning to the flagellar motor for biosensing

Tom J. Zajdel, Andrew Nam, Jove Yuan, Vikram R. Shirsat, Behzad Rad, and Michel M. Maharbiz

***Abstract*— *Escherichia coli* detects and follows chemical gradients in its environment in a process known as chemotaxis. The performance of chemotaxis approaches fundamental biosensor speed and sensitivity limits, but there have been relatively few attempts to incorporate the response into a functional biosensor. Toward that end, we have developed software to process digital microscope images of a large number of tethered *E. coli* responding to different chemical perturbations. Upwards of fifty cells can be recorded in one experiment, allowing for rapid labeling of the chemotactic responses of multiple cells. After we collected hundreds of wild-type chemotactic *E. coli* motor responses to dilutions of aspartate and leucine, we trained a support vector classifier (SVC) to estimate the order of magnitude of aspartate concentration between $0\,\mathrm{M}$, $100\,\mathrm{nM}$, and $1\,\mu\mathrm{M}$ with a single cell classification subset accuracy of 69%. We trained another SVC to differentiate between aspartate and leucine with a single cell classification subset accuracy of 83%. Using a majority-vote method on a bacterial population of size N, estimates have 95% confidence for $N = 27$ bacteria for concentration detection and $N = 9$ bacteria for chemical differentiation. These methods are a step towards adaptable chemotaxis-based biosensing.**

I. Introduction

During chemotaxis, *Escherichia coli* biochemically averages measurements from hundreds of receptors and adjusts the probability that the cell will change course or continue swimming in the same direction [1]. The canonical chemotaxis system consists of a sensory module with a rapid response time (0.1-2 sec) and an adaptation module that lags behind (30-60 sec). The result is a fast response with sensitivity to signals across five orders of dynamic range. The detection limit for the amino acid L-aspartate is $3.2\,\mathrm{nM}$, while large step changes in chemoattractant concentration affect motility within $0.3\,\mathrm{sec}$ [2].

With the high speed and sensitivity of bacterial chemotaxis, it is an attractive system to incorporate into a biosensor. However, surprisingly few studies have attempted to draw inferences based on observed *E. coli* behavior. One study observed swimming trajectories [3] while another used fluorescence resonance energy transfer (FRET) to directly report chemoreceptor kinase activity [4], which is an invasive single-cell technique that requires a sensitive microscope. Additionally, high variability between individual bacteria in a population requires responses to be recorded in large numbers, which is cumbersome.

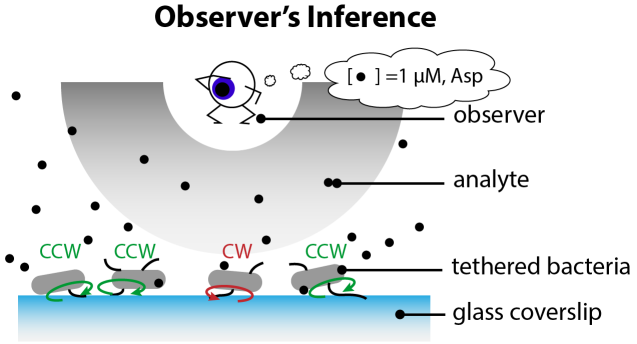

Fig. 1. The *E. coli* chemotaxis system rapidly adjusts motor bias in response to analytes and then slowly adapts to background conditions. Observation of motor behavior of a large population of bacteria enables inference of analytes in their environment.

In this work, we utilize an established flagellar stub tethering protocol [5] to monitor the chemotactic motor state of a large number of *E. coli* flagellar motors. A microscope is used to observe the motor response of tethered bacteria to dilution series of two amino acid analytes: L-aspartate (Asp), a chemoattractant; and L-leucine (Leu), a chemorepellent. The software we have developed includes an image processing algorithm that extracts the motor's direction as a function of time for the recorded bacteria. The data are then used to train support vector classifiers (SVCs) [6] that can reliably detect different concentrations of Asp as well as differentiate between Asp and Leu. The SVCs are scored with 3-fold cross-validation and inference confidence is characterized. These methods are a step towards the development of chemotaxis-based biosensors, which are attractive due to the adaptability of chemoreceptor systems, demonstrated by ongoing efforts to engineer novel chemoreceptors [7]. Such study also informs attempts to incorporate chemotactic bacteria as the front-end for electrochemical biosensors for microbiorobotics [8].

II. Methods

A. Strains and growth conditions

The strain used in this study, SYC12, derived from the wild type chemotactic strain *E. coli* RP437 [9], genomically expresses a sticky flagellin mutant [10]. These bacteria were grown to mid-exponential phase and their flagella were sheared to sticky stubs, following protocols published elsewhere [5].

B. Flagellar tethering and optical microscopy

The sheared bacterial suspension was injected into a disposable flowcell (Sticky Slide VI 0.2 Luer, Integrated

*Research was supported by the Office of Naval Research (award number N000141310551). Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. M.M. Maharbiz is a Chan Zuckerberg Biohub investigator.

T.J. Zajdel*, A. Nam, J. Yuan, and V.R. Shirsat are with the Department of Electrical Engineering and Computer Sciences, University of California - Berkeley, CA 94720 USA (*corresponding author: zajdel@eecs.berkeley.edu)

B. Rad is a Senior Scientific Engineering Associate at the Molecular Foundry, Lawrence Berkeley National Laboratory, CA 94720 USA

M.M. Maharbiz is a Professor with the Departments of Electrical Engineering and Computer Sciences and Bioengineering, University of California - Berkeley, CA 94720 USA and is an investigator with the Chan Zuckerberg Biohub, San Francisco, CA 94158 USA

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

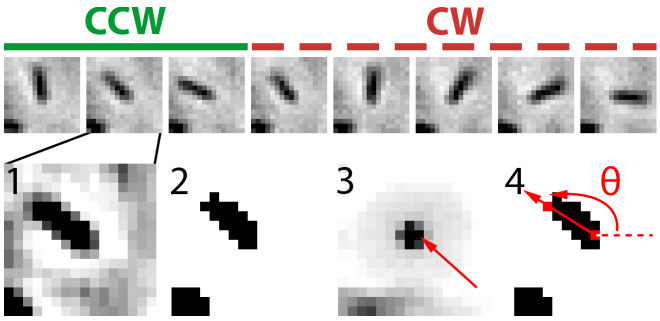

Fig. 2. Example time lapse microscopy of a tethered spinner cell (top). The time between frames is 32 ms. A switch in motor direction is evident between the third and fourth frame. In order to extract the cell's orientation over time, the following image processing is done on each frame: (1) smoothing and unsharp mask contrast enhancing, (2) isodata thresholding, (3) averaging to find tether point, (4) flood fill to find farthest point and calculate heading.

Bio Diagnostics, Martinsried, Germany) attached to a glass coverslip (VWR International, Radnor, PA). As supplied, the coverslips had a mildly hydrophobic surface to which the flagellar stubs readily adhered. The bacteria were allowed to tether to the glass for twenty minutes. After this time, 100 µL of motility medium (10 mM KPO_4, 0.1 mM EDTA, 1 mM L-methionine, pH 7.0) was gently flowed through to remove untethered bacteria.

Phase contrast images were taken with an EMCCD camera (iXon+, Andor, Belfast, Northern Ireland) at 40X magnification with a 32 m sec sampling period and a square imaging field of 550 by 550 µm. Motility medium with dissolved Asp or Leu (concentrations between 100 nM to 1 mM) was flushed through the flowcell and imaging immediately commenced for 120 seconds. To give the bacteria time to return to their baseline adaptation, the channel was flushed with clean motility medium and was left for 10 minutes before the introduction of a new condition.

C. Data processing

1) Image segmentation and processing: The image segmentation and feature extraction algorithm is pictured in Fig. 2. Microscopy image streams were first segmented by an ImageJ macro [11]. Each image was denoised by replacing each pixel with an average of its 3 by 3 neighborhood. Then, an unsharp mask was applied to strengthen the borders between the bacteria and their background. Then the image was binarized using an isodata thresholding algorithm and despeckled by a 3 by 3 median filter.

A Python script was used to extract rotation information and compute motor behavior features of interest. The tether point of each spinning cell was detected semi-automatically: a user selected a spinning cell, the binarized image was averaged across 100 frames, and the pixel with the lowest average brightness was deemed the tether point. For each frame, a flood fill originating at the tether point determined the extent of the cell body. The angle between the tether point and the farthest point was recorded as the cell's heading.

Once a large number of heading time series were collected for each experimental condition, features of interest were extracted from the bacteria in aggregate. Each heading trace was smoothed by a 250-ms moving average filter. Then, the rotational velocity was calculated by convolving the one-dimensional derivative kernel $k = [0.5, 0, -0.5]$ across the heading. One example trace is shown in Fig. 3. Finally, all the calculated speed traces were passed through a quality control program, which presented the user with a hysteresis-thresholded prediction of cell direction. This allowed the user to reject poor traces, and, if necessary, adjust the thresholds to ensure a good quality match between the calculated trace and the labeled motor directions.

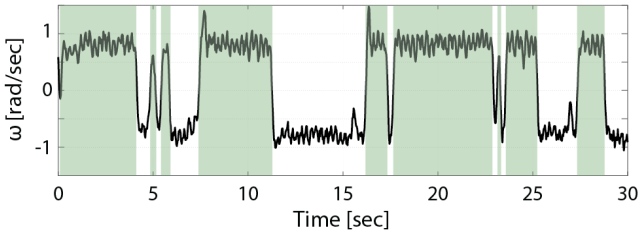

Fig. 3. Sample rotational velocity series for a tethered bacterium cell in motility medium. Counterclockwise intervals, where rotational velocity is positive, are shaded in green.

2) Feature extraction: The features that we extracted from these traces were the motor bias B and number of switches N_s. Each of these features were calculated for every bacterium in each condition. The motor bias was computed as the ratio of the total time spent rotating in the counterclockwise (CCW) direction over the length of the analysis window T. Hence, motor bias is the average time the motor is rotating in the CCW direction and ranges between 0 and 1. A higher bias typically indicates the presence of a chemoattractant signal, since longer CCW rotation results in longer runs that bias the cell's movement toward the signal [12]. Since B is averaged over a period of time, two cells may have the same bias despite having different switching profiles. To further disambiguate responses, the number of motor switches N_s was also tracked.

D. Condition inference and confidence

1) Support vector classifiers for per-bacterium labeling: Two classification problems were considered in this study: determining Asp concentration and differentiating between Asp and Leu. For each case, a support vector classifier (SVC) was trained to label individual bacteria as belonging to one group or another. We used the Python scikit-learn library [13] to train and evaluate our SVCs. We studied trends in the average population results to identify candidate features to train our SVCs toward these goals. We performed 3-fold cross-validation on each model and calculated the average subset accuracy, which corresponds to the probability that a given bacterium is classified correctly.

2) Confidence from a population of bacteria: Since individual bacteria have high variance in their responses, a practical observer must measure the response of a number of bacteria to have confidence in its inference. A cautious observer watching N bacteria responding to the same conditions could require at least $N/2$ of the population to have the same classification to make a statement about the conditions. If the classifier has a per-bacterium success rate of p, then the number of successfully labeled bacteria for a given population N will

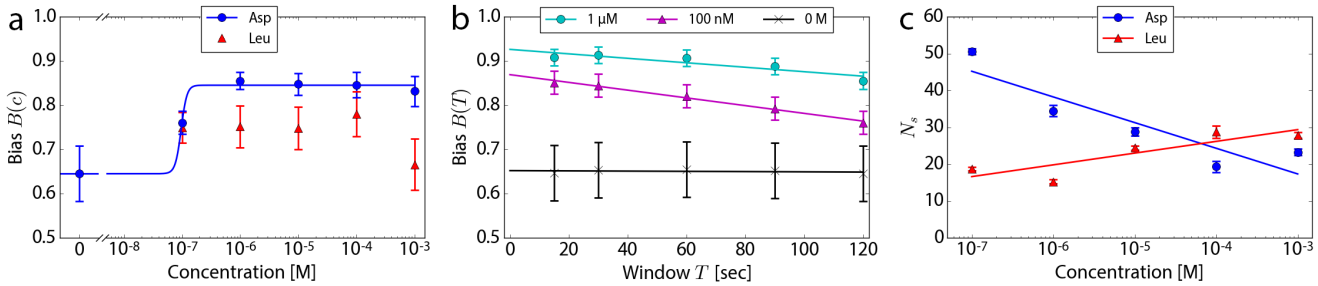

Fig. 4. (a) $B(c)$ for Asp and Leu with analysis window $T = 120$sec. (b) $B(T)$ for Asp at 1 µM, 100 nM, and 0 M with linear fit lines. (c) $N_s(c)$ for Asp and Leu with analysis window $T = 120$sec.

follow the binomial distribution. If we denote our confidence C as probability of $N/2$ being correctly labeled is given by:

$$C = 1 - F(N/2; p, N) \tag{1}$$

In Equation (1), $F(x; p, N)$ is the cumulative distribution function for the binomial distribution evaluated at x. When $x = N/2$, this term represents the probability that the majority of bacteria are not correctly classified. As the population size N increases, the confidence C also increases, from 0 to 1.

III. Results and Discussion

A. Motor feature distributions shift with time and concentration due to motor adaptation

The experimental conditions tested and the number of good traces from each condition are listed in Table I. These results came from six experimental repeats for each condition taken on three different days. A two-minute experiment yielded a maximum of 50 usable motor traces, but a typical yield was between 20 to 30 traces.

TABLE I

Bacteria count per condition used in analysis.

	Concentration [M]					
	0	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
Asp	146	126	70	207	57	139
Leu	146	132	89	72	41	41

Our survey of motor responses revealed candidate features for our SVCs. Fig. 4 plots $B(c)$ (at $T = 120$sec) and $B(T)$ for all bacteria exposed to Asp. Motor bias versus Asp concentration followed a sigmoid relationship. This response saturates at higher concentrations because Asp is a known chemoattractant with a very low detection threshold. Wild type cells quickly adapt to their environment, but the kinetics of this adaptation vary with the level of perturbation. This can be seen in the linear fits to $B(T)$ in Fig. 4. $B(c)$ was also plotted for the chemorepellent Leu, but a sigmoid relationship could not be established. It has been reported that although Leu is a chemorepellent, it acts as an attractant at low concentrations [14], which complicates analysis. As a result, concentration prediction was not attempted for Leu in this initial study.

We also fit parameters to the trends describing N_s as a function of concentration as shown in Fig 4. The trend established in our tests was different between Asp and Leu. A linear trend results for both Asp and Leu, with N_s decreasing with increasing $\log_{10}(c)$ for Asp and N_s increasing with increasing $\log_{10}(c)$ for Leu. Therefore, two samples could be differentiated into Asp or Leu by observing $N_s(c)$ over several dilutions.

B. SVC training and validation

1) Determining Asp concentration: E. coli is very sensitive to Asp concentration so its motor bias remains saturated near 1 µM, even after 120 sec. Therefore, motor bias is not a good feature to use to differentiate Asp concentrations between 1 µM and 1 mM. However, the $B(T)$ response is distinct between 0 M, 100 nM, and 1 µM. We randomly selected 120 cells labeled as either '0 M', '100 nM', or '1 µM' to populate the label vector $\mathbf{y}_{120\times 1}$. Given the distinct linear trends in $B(T)$, we populated our feature matrix $\mathbf{X}_{120\times 4}$ with rows consisting of the $B(T)$ of the corresponding cell in the label vector at $T = 30, 60, 90, 120$ sec:

$$\mathbf{X}_{120\times 4} = \begin{bmatrix} B(30\text{s}) & B(60\text{s}) & B(90\text{s}) & B(120\text{s}) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

The data was split into three groups of 40 for 3-fold cross-validation, and the SVC was trained with a polynomial kernel of degree 3. The computed subset accuracy (average fraction of correct labels) was 69%, reflecting the high variance in individual cell response. The confusion matrix is presented in Fig. 5, demonstrating the best accuracy (77%) was for 1 µM Asp, the highest concentration used in the classification. Careful optimization of data acquisition could increase the accuracy further, but inherent variability in gene expression will always produce a range of sensitivities in a given population.

2) Differentiating between Asp and Leu: E. coli responds differently to Asp and Leu since the former is a chemoattractant and the latter is a chemorepellent. Leu generally induces more motor switching in *E. coli* as its concentration is increased, while Asp suppresses motor switching as its concentration is increased. We randomly selected 90 cells labeled as either 'Asp' or 'Leu' to populate the label vector $\mathbf{y}_{90\times 1}$. Given the distinct trends in $N_s(c)$, we populated our feature matrix $\mathbf{X}_{90\times 3}$ with rows consisting of $N_s(c)$ of the corresponding cell in the label vector at $c = 10^{-7}, 10^{-5}, 10^{-3}$ M:

$$\mathbf{X}_{90\times 3} = \begin{bmatrix} N_s(10^{-7}\text{M}) & N_s(10^{-5}\text{M}) & N_s(10^{-3}\text{M}) \\ \vdots & \vdots & \vdots \end{bmatrix}$$

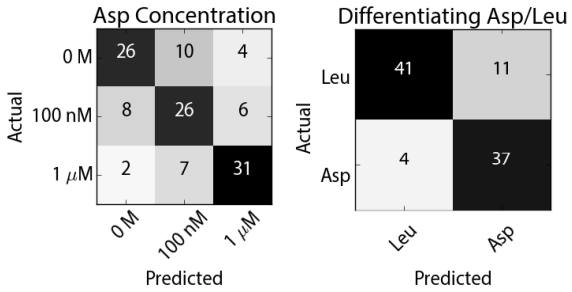

Fig. 5. The confusion matrices resulting from cross-validation of the classifiers.

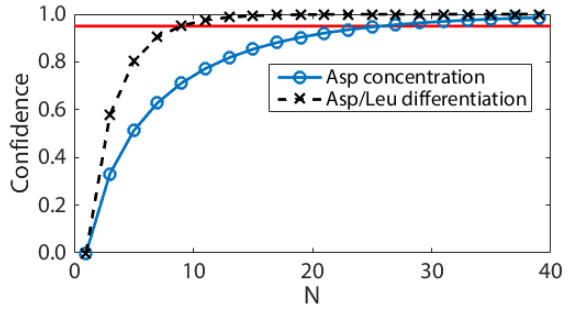

Fig. 6. Confidence versus the number of cells classified, assuming that the majority vote is accepted. A confidence of 0.95 (marked with horizontal red line) occurs at $N = 27$ when determining Asp concentration and $N = 9$ when differentiating between Asp and Leu.

The data was broken into three groups of 30 for 3-fold cross-validation, and the SVC was trained with a linear kernel. The subset accuracy was 83%. The confusion matrix is presented in Fig. 5 and reflects this high subset accuracy.

C. Number of cells required for high prediction confidence

Following Equation (1) and the subset accuracies achieved by the SVCs, the majority-label confidence was calculated as a function of number of bacteria N. A confidence of 95% is achieved at $N = 27$ for determining Asp concentration and $N = 9$ for Asp/Leu differentiation. These population sizes are easily achieved in one experiment, and only 120 seconds of data need to be recorded. However, incorporating more labels would likely require more data to properly train the SVCs and require a higher N, more features, or multiple strains of *E. coli* to properly assess conditions.

D. Discussion

These preliminary results provide a framework for using observed flagellar motor responses to infer environmental changes that can be readily expanded by biological engineering. Synthetic biologists have designed chimeric chemoreceptors that confer new sensing ability to *E. coli* [7]. Receptors can be knocked out, resulting in orthogonal responses from different cell populations, and could make it easier to differentiate between chemicals. Protein engineering techniques allow for fine tuning of the sensitivity of chemotaxis to different concentration ranges [15], expanding the range of detectable concentrations beyond those in this study. The speed of the motor could also be tracked, allowing for real time tracking of the proton-motive force, which responds to a number of other environmental signals and powers the motor's rotation [16]. In short, this method can be readily extended by building onto ongoing work in several domains.

IV. Conclusion

In this paper, we propose using observations of the bacterial flagellar motor to identify and quantify bioanalytes. The chemotaxis system couples chemoreceptor measurements to the motor response, which we characterized for a large number of cells using our image processing suite. We then demonstrated inferring aspartate concentration and differentiating between leucine and aspartate by training SVCs to classify motor responses. Since the chemotaxis system is readily adaptable, there is potential for bottom-up design of new biosensors that work on this principle.

Acknowledgment

We thank Dr. Jasmine Nirody, Dr. Moshe Baruch, Dr. Caroline Ajo-Franklin, and Dr. Yoshiyuki Sowa for helpful discussions during the preparation of this work. We thank Debleena Sengupta, Alexander Walczak, and Victor Tieu for early contributions to methods used in this work. SYC12 was a gift from Dr. Yoshiyuki Sowa at Hosei University. Bacterial culture and microscopy were performed at the Molecular Foundry at Lawrence Berkeley National Laboratory. Michel M. Maharbiz is a Chan Zuckerberg Biohub investigator.

References

[1] V. Sourjik and N. S. Wingreen, "Responding to chemical gradients: bacterial chemotaxis," *Curr. Opinion in Cell Biology*, vol. 24, no. 2, pp. 262–268, 2012.
[2] T. Sagawa *et al.*, "Single-cell *E. coli* response to an instantaneously applied chemotactic signal," *Biophysical Journal*, vol. 107, no. 3, pp. 730–739, 2014.
[3] J.-B. Masson *et al.*, "Noninvasive inference of the molecular chemotactic response using bacterial trajectories," *PNAS*, vol. 109, no. 5, pp. 1802–1807, 2012.
[4] D. Clausznitzer *et al.*, "Predicting chemical environments of bacteria from receptor signaling," *PLoS Comput. Biol.*, vol. 10, no. 10, 2014.
[5] Y. Inoue, "Rotation measurements of tethered cells," *The Bacterial Flagellum: Methods and Protocols*, pp. 163–174, 2017.
[6] C.-W. Hsu *et al.*, "A practical guide to support vector classification," 2003.
[7] S. Bi *et al.*, "Engineering hybrid chemotaxis receptors in bacteria," *ACS synthetic biology*, vol. 5, no. 9, pp. 989–1001, 2016.
[8] T. J. Zajdel *et al.*, "Towards a biohybrid sensing platform built on impedance-based bacterial flagellar motor tachometry," in *2017 IEEE BioCAS*, Oct 2017, pp. 292–295.
[9] J. S. Parkinson and S. E. Houts, "Isolation and behavior of escherichia coli deletion mutants lacking chemotaxis functions." *Journal of bacteriology*, vol. 151, no. 1, pp. 106–113, 1982.
[10] G. Kuwajima, "Construction of a minimum-size functional flagellin of escherichia coli." *Journal of bacteriology*, vol. 170, no. 7, pp. 3305–3309, 1988.
[11] C. A. Schneider *et al.*, "Nih image to imagej: 25 years of image analysis," *Nature methods*, vol. 9, no. 7, p. 671, 2012.
[12] S. M. Block *et al.*, "Impulse responses in bacterial chemotaxis," *Cell*, vol. 31, no. 1, pp. 215–226, 1982.
[13] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[14] H. Mao *et al.*, "A sensitive, versatile microfluidic assay for bacterial chemotaxis," *Proc. Natl. Acad. Sci.*, vol. 100, no. 9, pp. 5449–5454, 2003.
[15] B. A. Mello and Y. Tu, "Quantitative modeling of sensitivity in bacterial chemotaxis: the role of coupling among different chemoreceptor species," *PNAS*, vol. 100, no. 14, pp. 8223–8228, 2003.
[16] T. J. Zajdel *et al.*, "Probing the dynamics of the proton-motive force in *E. coli*," in *Proc. of IEEE SENSORS*, 2014, pp. 1764–1767.