Computational Geometric Analysis for *C. elegans* Trajectories on Thermal and Salinity Gradient

Yuetan Chu

The Department of Physics, University of Science and Technology of China, Hefei, China
Email: ytchu@mail.ustc.edu.cn

Abstract

Elegans are one of the best model organisms in neural researches, and tropism movement is a typical learning and memorizing activity. Based on one imaging technique called Fast Track-Capturing Microscope (FTCM), we investigated the movement regulation. Two movement patterns are extracted from various trajectories through analysis on turning angle. Then we applied this classification on trajectory regulation on the compound gradient field, and theoretical results corresponded with experiments well, which can initially verify the conclusion. Our breakthrough is performed computational geometric analysis on trajectories. Several independent features were combined to describe movement properties by principal composition analysis (PCA) and support vector machine (SVM). After normalizing all data sets, no-supervising machine learning was processed along with some training under certain supervision. The final classification results performed perfectly, which indicates the further application of such computational analysis in biology researches combining with machine learning.

Keywords

C. elegans, Tropism, Trajectories Classification, Computational Geometric Analysis, PCA

1. Introduction

When an animal experiences an environment at one point, it can form a kind of memory for its future behaviors. This movement pattern requires the animal a mechanism of navigating, along with conditions memorizing and decision-making based on its previous experiences. It is a process that includes learning, memorizing, and performing [1]. There are many possible trajectories

when an animal chasing for its favorable environment, however, specific characteristics of these trajectories have not been invested thoroughly. Combining some geometric features with certain neural activities guiding these motions can be meaningful work. The nematode is an ideal model organism. Due to the conservation of neural circuit structures in evolution, researches on the mechanism of nematode learning can easily be extended to other more complex organisms. The nematode does not have an innate preferred temperature. Instead, *C. elegans* cultivated at a given temperature will remember this temperature (Tc) and then migrate toward it when on a thermal gradient [2]. The same pattern applies to the salinity gradient as well [3]. What we would like to research is the corresponding between elegans movements with their neural activities shown by fluorescence imaging. There are several apparent parts within the neural activities [4] [5], and we want to divide the elegans movement into several parts as well and do further analysis on the trajectory geometry, which shows its strong application in many other fields [6]. However, there is not any similar analysis on elegans trajectories, especially in such conditions with extra stimulation. In this paper, we mainly include the following parts: statistics on turning angle distributions, which suggest the possibility that there exist several different movement modes: application of such classification in the movement regulation on the compound fields; principle composition analysis (PCA) on trajectories; results of auto-clustering algorithms performed on trajectories classification.

We take the standard cultivating technology for all elegans [7]. On the experimental platform, the temperature is set at 16°C on the left edge and 25°C on the right edge. In this way, elegans may be inclined to move to the left edge instead of random walking. Similarly, the salinity field is a rectangular area ranging from 100 mmol/mL NaCl on the right side to 50 mmol/mL NaCl on the left side. Elegans may take different strategies or their combinations when navigating, while some of them may hesitate within our recording time, namely not showing a determined orientation [8]. An imaging technology called Fast Track-Capturing Microscope (FTCM) [9] is applied to our research to record trajectories. In each set of experiments, 6 - 8 healthy nematodes were picked out from NGM and washed with NGM buffer (the same composition as the NGM plate, but without agar) in order to wash away bacterial food, and then we put them into the platform (set the salinity and temperature gradient field in advance) to start the experiment. The movement trajectories of nematodes are captured with FTCM at a rate of 2 frames per second for a total of 25 minutes. Discrete data sets were obtained by normal particle tracking and shape analysis programs written in Labview (National Instruments, Austin) and MATLAB (Mathworks, Natick). In the experiment, we managed to keep the beginning point of each trajectory consistent for all elegans, and the middle point of the platform was selected as the optimal origin for all trajectories. The relative error can be smaller than 50 unit length. As a result, within a certain error range that we can ignore to a great extent, we would regard all elegans starting at the same point, the base point (0, 0) in the following research.
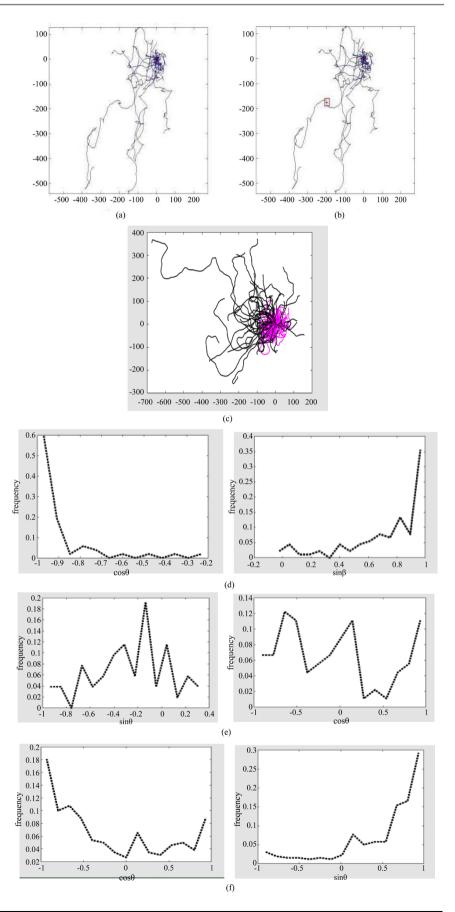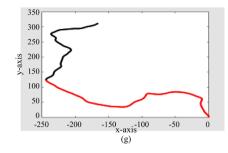
2. Results

The Trajectories of Elegans Containing Two Different Categories

Our analytical pipeline for computational geometry began with the collection of 2D trajectories. We obtained 2D-FTCM images of elegans trajectories (Figure 1(a)). Our analysis was based on more than one thousand trajectories data sets on each kind of gradient field from FTCM. These trajectories were documented with a sequence of discrete points, such as (x_1, y_1), (x_2, y_2). Especially, (x_n, y_n) represented the endpoint of one trajectory and (x_0, y_0) represented the original point of one trajectory, namely $(0, 0)$. Unit length in this system is reset, not according to millimeter nor any other normal length unit. We next used programs written in MATLAB to remove self-intersection in small scales (smaller than 50 × 50 scale) (Figure 1(b)) to make the trajectories more smooth, while maintaining self-intersection parts on a larger scale. The intersection on a small scale, called "small turn", meant a perturbation in motion. Instead of a general direction, this perturbation represented some tiny random neural activities and could not be included in general trajectory analysis conveniently.

Before further analysis, we should point out that although we expect a definite direction for elegans to move toward, there were still lots of elegans in the hesitation and sleepy states within twenty-five minutes (Figure 1(c)). Twenty-five minutes might not be enough for all elegans to wake up thoroughly from the original environment with a given temperature and then decide which direction should go forward. Trajectories of these elegans, which always behave with a relatively small distance, were not included in the following analysis because they just acted as a random walk, presenting little regulation [7]. However, how to define and select this kind of trajectories could be challenging. If the criteria were the final distance of the trajectory, which meant when the final distance of one trajectory is less than particular value, the trajectory should be regarded as a track without tendency, it could be controversial to define such a distance. However, we gave a method to determine such tracks with principal component analysis (PCA) automatically (Later in this passage).

To describe these tracks, we firstly regarded movement as Brownian motion [5]. The movement of elegans can be described in this pattern: one nematode collected information from the current point, such as temperature and salinity, and then compared it with previous information. This comparison may guide it to moving to the next step. For example, one nematode was inclined to move to somewhere cooler with lower temperatures. When reaching a point with a higher temperature than before, it may manage to get away in an opposite direction. All these strategies can be depicted in a summary that elegans may run forward for a small distance and then turn an angle. This pattern was just similar to Brownian motion. Every step was about 1 mm, equivalent to the length of one nematode. In this method, our first step was to investigate the distribution of every turning angle. Angles can be written as

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 1. Basic images of elegans trajectories in salinity gradient (a) Original images of 2D-FTCM trajectories. All trajectories are set starting at the nearly same point. (b) One typical example of self-intersection is denoted in the red frame. (c) Trajectories without definite orientation are presented in the pink color, while other normal ones are presented in black color. (d), (e) Two different types of turning angles distribution. One is cosine-value gathering pattern and the other is sine-value gathering pattern. (f) An example of trajectory with no clear cosine-value gathering nor clear sine-value gathering pattern. Note the value of frequency. (g) A trajectory consists with both HM and LM. HM part is marked in red and LM part is marked in black.

$$\theta = \arccos\left(\frac{\Delta x}{\sqrt{\Delta x^2 + \Delta y^2}}\right) = \arcsin\left(\frac{\Delta y}{\sqrt{\Delta x^2 + \Delta y^2}}\right) \tag{1}$$

$$\Delta x = x_{i+1} - x_i \tag{2}$$

$$\Delta y = y_{i+1} - y_i \tag{3}$$

However, different description methods came to different results (**Figure 1(d)**, **Figure 1(e)**). For certain trajectories, when we documented the angle in the form of sine, absolute value distribution concentrated between 0.8 and 1 (the proportion was more than 70 percent), while absolute value distribution in cosine only presented uniform distribution, or in other words, random distribution. For other trajectories, such concentration regulation could only be revealed in value of cosine, while there was no determined regulation in absolute value distribution in the sine method. This regulation meant the possibility that these trajectories could be further sorted automatically by criteria based on these angle distributions. We can regard these two kinds of trajectories as two different patterns, called horizontal mode (HM) and longitudinal mode (LM). Almost all trajectories can be categorized into these two modes. However, there was also a small proportion of trajectories (about 5 percent) not presenting any concentration regulation in distribution (**Figure 1(f)**). Both value distribution in sine and cosine presented relatively uniform distribution. In general, these particular trajectories consist of a combination of two movement modes (**Figure 1(g)**).

3. Movement in the Compound Gradient Field

The compound field is composed of a horizontal thermal gradient field and a vertical salinity gradient field. The temperature gradient and salinity gradient in the compound field are consistent with the previous single gradient field. The cool side is set at the left edge and the suitable salinity side is set at the lower

edge. It is obvious that the elegans may be inclined to the third quadrant (**Figure 2(a)**).

For the convenience of following expressions, we first define the final trajectories angle in the elegans trajectories. The angle, which we record as θ, is defined as the following:

$$\theta = \begin{cases} \arctan\left(\dfrac{y_n}{x_n}\right) & \text{The final point falls in the first quadrant} \\ \arctan\left(\dfrac{y_n}{x_n} + \pi\right) & \text{The final point falls in the second, third quadrant} \\ \arctan\left(\dfrac{y_n}{x_n} + 2\pi\right) & \text{The final point falls in the fourth quadrant} \end{cases} \quad (4)$$

Under this definition, we can guarantee all angles are in the range [0, pi/2]. We then define $N(\theta)$, which equals to the ratio of the number of trajectories which the final angle is less than θ to the total number. In this method, we can determine at which angle the elegans are most inclined to.
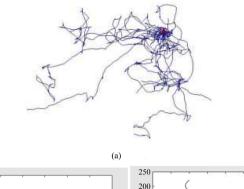
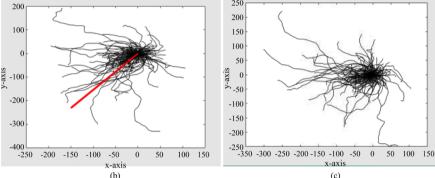$$\frac{dN(\theta)}{\theta} = 0 \quad \text{when } \theta = \theta_{est} \quad (5)$$

θ_{est} the most movement angle, can describe the general orientation of all trajectories. We applied this method to trajectories in the compound field. Instead of 225°, the middle line in the third quadrant as our common sense, θ_{est} equals to 237° (**Figure 2(b)**). We then wondered whether there is anything differences in computational geometric features between trajectories in compound field and single gradient field. After rotating whole images to make the most movement orientation parallel to negative x-axis, figures show little distinction from trajectories in single gradient fields (**Figure 2(c)**).

The most reasonable explanation of deviation of θ_{est} from our common sense might be different weight of intention which movement strategy elegans, HM or LM, may take in thermal and salinity gradient field. After specific statistics, the ratio of the number of elegans which take HM strategies to the left which take LM strategy is 1.425 in salinity gradient, while the number is 0.543 in the thermal gradient. The ratio to total trajectories number is (N (HM), N (LM)) = (35.19%, 64.81%) for thermal gradient and (N (HM), N (LM)) = (58.76%, 41.24%) for salinity gradient. This quantity strongly means elegans could prefer to take LM in thermal gradient and HM in salinity gradient. What's more, we also calculated the average coordinate value for final points respectively for trajectories with HM pattern and LM pattern, which is written as $\left(\langle x_{Hn}\rangle, \langle y_{Hn}\rangle\right)$ for HM and $\left(\langle x_{Ln}\rangle, \langle y_{Ln}\rangle\right)$ for LM. As the symbol of general movement orientation vector, $\left(\overline{x}_n, \overline{y}_n\right)$ is certainly different for elegans in different gradient.

For elegans in thermal gradient

$$\begin{cases} \left(\langle x_{Hn}\rangle, \langle y_{Hn}\rangle\right) = \left(-174.33, 29.87\right) \\ \left(\langle x_{Ln}\rangle, \langle y_{Ln}\rangle\right) = \left(-56.28, 21.84\right) \end{cases}$$

(a)

(b)   (c)

Figure 2 Trajectories in compound gradient field. (a) Original images of 2D-FTCM trajectories. (b) The line of most movement angle is shown in the figure with red line. In this direction, there is the most number of trajectories. (c) After rotation, there is no obvious difference with images in the single gradient.

For elegans in salinity gradient

$$\begin{cases} \left(\langle x_{Hn}\rangle, \langle y_{Hn}\rangle\right) = \left(-187.86, -9.23\right) \\ \left(\langle x_{Ln}\rangle, \langle y_{Ln}\rangle\right) = \left(-53.02, -16.15\right) \end{cases}$$

Considering different weight in HM and LM, we can combine these two vectors in HM and LM into one general orientation vector.

For elegans in thermal gradient

$$\begin{cases} x_{\text{thermal}} = -174.33 \times 35.19\% - 56.28 \times 64.81\% = -97.82 \\ y_{\text{thermal}} = 29.87 \times 35.19\% + 21.84 \times 64.81\% = 54.67 \end{cases}$$

For elegans in salinity gradient

$$\begin{cases} x_{\text{salinity}} = -187.86 \times 58.76\% - 53.02 \times 41.24\% = -132.25 \\ y_{\text{salinity}} = -9.23 \times 58.76\% - 16.15 \times 41.24\% = -12.08 \end{cases}$$

Then we introduced influence factor due to different impact from thermal and salinity gradient. The influence factor is defined as following

$$\beta = \frac{\eta_{\text{HM in thermal}}}{\eta_{\text{LM in thermal}}} = \frac{58.76\%}{35.19\%} = 1.67 \tag{6}$$

Combining two general vectors in two fields when the influence factor taken into considering, we can get a standard vector, (−0.82, −1.42), to measure movement orientation in compound field. Angle of this vector is 239.995°, which is
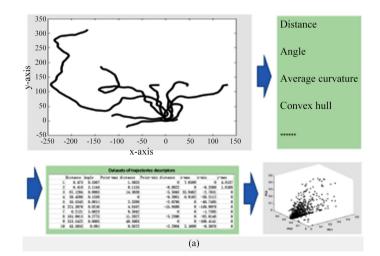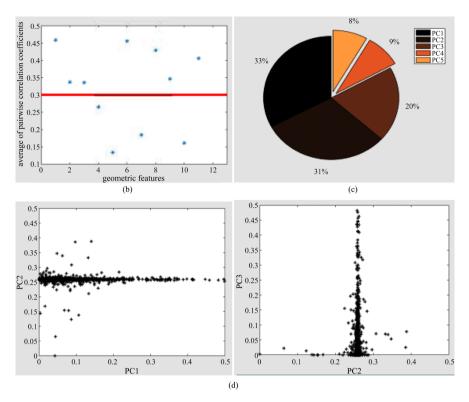
according with 237˚ to a great extend.

4. A Method for Quantifying Geometry Feature of Trajectories

The method aims to sort these trajectories automatically according to criteria based on their morphological features (**Figure 3(a)**). Ultimately, we obtained 11 descriptors of trajectories [6]. These descriptors include both basic shape features (eg, final distance, final trajectories angle) and more complex shape features (eg, convex hull, mean curvature). We next investigated whether the multiple descriptors would be useful for representing data set for a large trajectory population with continuous morphological variables. From our initial analysis of independence among descriptors, we selected six based on two criteria. First, we selected two descriptors that reflect principal structural features (final distance and final trajectories angle). Second, four other descriptors that showed high independence (the averages of pairwise correlation coefficients were <0.3) were selected (**Figure 3(b)**). Considering symmetry, they are chosen as features including convex hull, maximum x-coordinate value, and maximum slope value within one trajectory.

We present our method in datasets of salinity trajectories, while the result is similar to datasets under other gradients. Principal component analysis (PCA) was performed with these five descriptors to obtain trajectory distribution in the new space. The first third features (principal components 1-3, PC 1-3)covered about 85% of the variance in the data (**Figure 3(c)**). Principal components are normalized linear combinations of the original descriptors and reflect specific morphological properties of trajectories.

We then investigated the distribution of trajectories in the feature space with axes corresponding to PC1, PC2 and PC3 (**Figure 3(d)**). All values of each combination are processed with max-min normalization. The point distribution is quite concentrated, which indicates there are some strong common features among all the trajectories.
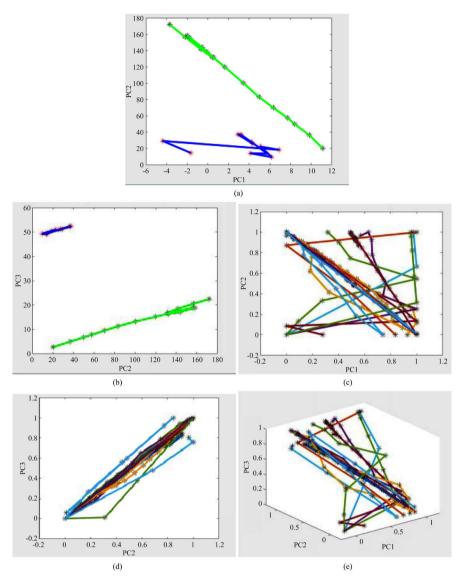

(a)

Figure 3. Principal component analysis (PCA) was performed with geometric features of all trajectories [8]. (a) Processing of trajectories geometric analysis. Several geometric features can be extracted from original images, and the data sets are further analyzed by the techniques of dimensional deduction and automatic classification by machine learning. (b) Averages of correlation coefficients of each feature. The features from left to right respectively are final distance, final angle, minimum convex hull, maximum convex hull, maximum x-coordinate value, minimum x-coordinate value, maximum y-coordinate value, minimum y-coordinate value, average curvature, maximum and minimum slope value in one trajectories. Considering symmetry, maximum y-coordinate value is not included in the following studying because this value should be zero theoretically. (c) Proportion of each component by PCA. The first third features covered about 85%, which means an ideal result of our method. (d) The distribution of trajectories in the feature space. The axes are set as PC1 with PC2, and PC2 with PC3.

We selected two trajectories, on the thermal gradient field, one of which is LM pattern and another is HM pattern, with dimension more than 2×200, which suggests it had a long period of morphological evolution. We then used the combination of features obtained by PCA to describe their transition. The first pair of features is PC1 and PC2 (**Figure 4(a)**) and the second pair of features are PC2 and PC3 (**Figure 4(b)**). One can hardly distinguish two patterns from this relationship. Furthermore, we mapped the trajectories of shape transitions of several individual trajectories in feature space (**Figure 4(c)**, **Figure 4(d)**) and generated a 3D map depicting the behavior of the spine population (**Figure 4(e)**).
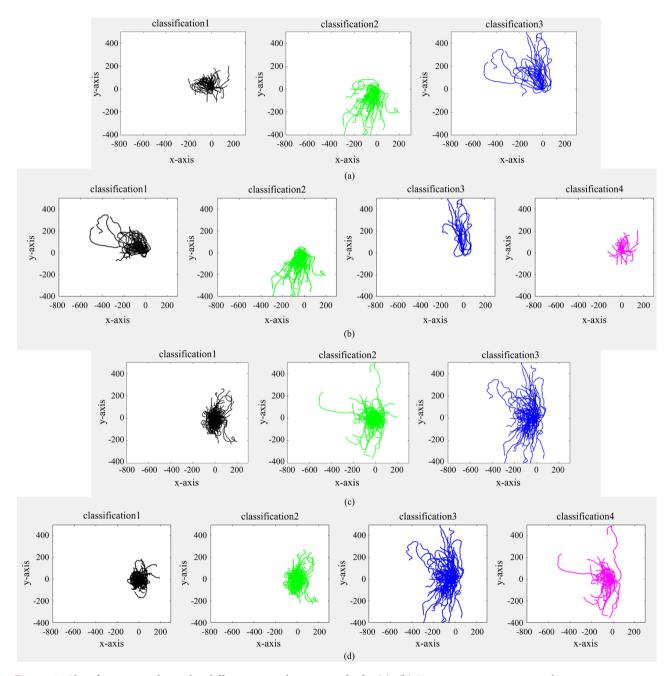
The next step is to utilize automatically classification by machine learning. and the first step should be normalization. There are two main methods of normalization dealing with trajectories. One is transforming all trajectories into

Figure 4. Morphological characteristics transition in the feature space. (a), (b) Two single evolution trajectories are shown in the picture. The green one represents HM pattern, while the blue one represents LM pattern. (c), (d), (e) Several trajectories evolution trajectories shown in 2D and 3D feature space.

standard vectors, for example, with a dimension of 2 × 400. The longest dimension of datasets of one trajectory is no more than 2 × 350, so the size 2 × 400 can guarantee we will not lose any information. Lagrangian interpolation was used when replenishing the vector. Every three adjacent points in the datasets were selected to construct a parabola to predict coordinate information except for these known points. Another method is to use those certain computational geometric features as well as PCA combinations. Similarly, every value should be processed with max-min normalization. In this way, we can apply unsupervised machine learning to each processed data set. However, there were always some special trajectories which belong to a category of their own. There is no meaning when one category only contains one example. Naturally, we first classified these

trajectories artificially and then took these examples as the test set, while using the classified trajectories as the training set to supervise and train a simple linear model.

The results are shown as follows (Figure 5), the effect dividing all trajectories into three or four groups can be both satisfactory. As for the normalization on trajectory vector, which is divided into three groups (Figure 5(a)), the first classification represented those elegans, which did not have a clear movement orientation, namely a short trajectory distance. This method can help solve the

Figure 5. Classification results under different normalization methods. (a), (b) Trajectories vectors normalization, one is set as three groups and another is four. (c), (d) Geometric features normalization, and the classification results are similar.

above problem about these trajectories automatically. The second and third classifications have definite different movement directions. One is moving up along the y-axis to a great extend while another is moving down approximately. The results of four groups are shown below (**Figure 5(b)**). The self-intersection part is separated successfully, as well as the trajectories along y-axis. Furthermore, the first classification indicates HM pattern excitingly, where the general direction is approximately parallel to negative x-axis. This result verifies our previous conclusion. When it comes to the geometric features normalization, the classification can be better. In the three-group classification system, self-intersection, HM as well as LM can be all separated clearly (**Figure 5(c)**). What is more accurate is that in the four-group system, algorithms can also distinguish those trajectories walking in an opposite direction (**Figure 5(d)**). Because their trajectories distances are similar to self-intersection, we did not consider them originally. However, machine learning can separate them very well. This result suggests further potential and strength of machine learning in biology research.

5. Conclusion

In this paper, we mainly invest the movement regulation of *C. elegans* on the experimental platform with a salinity and temperature gradient. We perform computational geometric analysis on thousands of trajectories. Different statistical methods show obviously different distributions on the turning angles of trajectories, and we extract two movement modes according to such results. We then verify this classification mainly from two viewpoints. One is its application on the prediction of movement direction on the compound field, which matched with the experiment well. The other verification is performed by auto-classification by clustering algorithms. Although there exist various advantages and disadvantages in each method, all final results show an apparent classification, which can correspond with the previous assumption. On the one hand, our conclusion can be a solid basement for further elegans fluorescence imaging, which aims to invest more on elegans' neural activities. On the other hand, it suggests a potential application of machine learning in biological researches.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Basu, J. and Siegelbaum, S.A. (2015) The Corticohippocampal Circuit, Synaptic Plasticity, and Memory. *Spring Harbor Perspectives in Biology*, **7**, a021733.

[2] Hedgecock, E.M. and Russell, R.L. (1975) Normal and Mutant Thermotaxis in the Nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, **72**, 4061-4065. https://doi.org/10.1073/pnas.72.10.4061

[3] Hawk, J.D. and Calvo, A.C. (2017) Integration of Plasticity Mechanisms within a

Single Sensory Neuron of *C. elegans* Actuates a Memory. *Neuron*, **97**, 356-367.
https://doi.org/10.1016/j.neuron.2017.12.027

[4]   Ventimiglia, D. and Bargmann, C.I. (2017) Diverse Modes of Synaptic Signaling, Regulation, and Plasticity Distinguish Two Classes of *C. elegans* Glutamatergic Neurons. *eLife*6, **6**, e31234. https://doi.org/10.7554/eLife.31234

[5]   Iino, Y. and Yoshida, K. (2009) Parallel Use of Two Behavioral Mechanisms for Chemotaxis in *Caenorhabditis elegans*. *Journal of Neuroscience*, **29**, 5370-5380.
https://doi.org/10.1523/JNEUROSCI.3633-08.2009

[6]   Kashiwagi, Y., Higashi, T., Obashi, K., *et al.* (2019) Computational Geometry Analysis of Dendritic Spines by Structured Illumination Microscopy. *Nature Communications*, **10**, Article Number: 1285. https://doi.org/10.1038/s41467-019-09337-0

[7]   Brenner, S. (1974) The Genetics of *Caenorhabditis elegans*. *Genetics*, **77**, 71-94.

[8]   Adachi, T., *et al.* (2010) Reversal of Salt Preference Is Directed by the Insulin/PI3K and Gq/PKC Signaling in *Caenorhabditis elegans*. *Genetics*, **186**, 1309-1319.
https://doi.org/10.1534/genetics.110.119768

[9]   Lorensen, W.E. and Cline, H.E. (1987) Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *ACM SIGGRAPH Computer Graphics*, **21**, 163-169.
https://doi.org/10.1145/37402.37422