Tracking the Diffusion of Signal Proteins in Escherichia Coli

Herbert Weidner [a]

Abstract: We describe a fast computational model to study the signal transduction in E. coli chemotaxis. It allows to trace the places and chemical reactions of individual molecules in a cell as a function of time. From the spatial resolution of the model and the known diffusion constant of the proteins, a time step of 3 µs is calculated for the simulation.

Introduction

We know very many molecular details of E. coli and a lot of graphs show which proteins interact with each other[1]. Strangely, the role of time is usually completely ignored, as if molecules far apart would react instantaneously. Attempts to trace the signal processing through many individual steps in realistic models, are rather rare.

Probably, the signal propagation and memory inside of a bacterium is carried out by state, motion and location of proteins in the cytoplasm. On the way, the proteins can change their state - either spontaneously or when they interact with other molecules. We assume that signal proteins do not attract each other to exchange information. Reactions can only happen when proteins are randomly approaching, that is to say, they are neighbors during a short period of time. Even then the transfer of the signal may not take place because they are unfavorably oriented. Fortunately, we have to trace the movement of only about 20,000 signal proteins, but even then the high computing effort requires a fast computer.

Because, as we believe, the simulation of protein migration and their interactions can contribute to the deeper understanding of bacterial signal processing, we designed a simplified model in which successive states can be calculated quickly. How exactly do molecules or molecular assemblies remember the previously measured concentration of food outside the cell? Perhaps a simulation can show how intelligent behavior results from the sequential interaction of different proteins.

A digital model of the bacterium

A typical E.coli bacterium is rod-shaped, about 2.5 µm long, has a cell diameter of 0.8 µm and contains about 2.8 million proteins. In strain RP437[2], 70,000 proteins are closely related to chemotaxis. 50,000 of them (Tsr + Tar) are receptors and remain at fixed locations in the lipid membrane. Another part, about 24,800 proteins (≈ 0.88%), are signal proteins like CheY, moving around inside the bacterium.

The aim of this study is to track the signal transmission inside the cell between the sensor proteins and the flagella motors. In contrast to the rather unrealistic assumption that the cytoplasm is a homogeneous, ideally mixed liquid whose properties can be described with differential equations, here the position of each signal protein is monitored. Molecules do not move purposefully, they constantly change their position as a result of the thermal movement. If the right partners meet at random, chemical reactions may occur. In the case of intracellular signal transmission, this can preferably be allosteric modifications which require particularly little energy.

It is impossible to describe the motion of the proteins with high precision because the computing time would be much too long. It is much simpler and faster to describe the chaotic movements by a discretization of space and time. Although the molecules move continuously, it is probably suffi-

a) 19. Apr 2017, email: herbertweidner99@gmail.com

cient to record the coarse positions at regular intervals. This is examined in more detail below.

To simplify all calculations, the cylindrical cross-section of the cell is transformed into a square, leading to a right angle 3D model: Our "bacterium" is a cuboid of length 2 µm and a square cross-section of (0.7 µm)². The whole cuboid is divided into 2.8 million cubes of edge length 7 nm, large enough to contain one typical protein. Once a time step has elapsed, each protein jumps into an adjacent cube. This digital model has a high spatial resolution, consisting of 285·100·100 cubes. A finer resolution is possible, but it extends the calculation time.

This model allows very fast calculations because the complex distance calculations (Pythagoras with calculation of the square root) are replaced by simple differences. In addition, calculations with integer numbers are always faster than with the floating point representation. The high precision of which is not required in this study.

Away from the cell walls, each cube has exactly 26 neighbors. A model is never perfect. This cube model differs from reality in several details. The molecules can only jump into one of 26 permitted directions and the distance depends on the direction. This means that the velocity of the proteins depends also on the direction - a contradiction to the usual assumption that the diffusion movement is isotropic. However, this error can be kept very small.

Consider a protein at the site (0; 0; 0). The distance to the 26 neighbors (from center to center) can only take one of three possible values 9.9 nm, 12.1 nm or 7 nm (see the first three lines of the table). Choosing the time step $\Delta t = 3$ µs (see below), we calculate the apparent speeds

frequency	distance	speed = distance/Δt	Examples: from (0;0;0) to
12	$7\sqrt{2}$ nm = 9.9 nm	3.3 mm/s	(1; 0; -1) or (-1; 1; 0)
8	$7\sqrt{3}$ nm = 12.1 nm	4.0 mm/s	(1; -1; 1) or (-1; -1; -1)
6	7 nm	2.3 mm/s	(-1; 0; 0) or (0; 0; 1)
6	14 nm	4.7 mm/s	(-2; 0; 0) or (0; 0; 2)

The large variation of the speed (3.3 ± 0.6) mm/s can be reduced if the shortest possible distance (7 nm) is occasionally doubled (see fourth line, highlighted in yellow). Then the speed variation is much lower and less direction-dependent (3.6 ± 0.3) mm/s. How realistic are these arbitrarily chosen figures? Do they match known experimental data?

Checking the diffusion constant

Molecules are never at rest. As they are continually jostled by neighboring molecules (Brownian motion), they can not move in a straight line and need many milliseconds to travel one cell length. The experimentally measurable diffusion constant is a suitable parameter to check whether the digital model correctly describes the chaotic movement of small proteins in the bacterium.

We assume that all 2.8 million proteins are uniformly distributed and each cube contains one protein (not necessarily a *signal* protein) plus a background of other particles such as H_2O, ATP or ions, whose number and movements are not of interest. No large, impermeable volumes such as ribosomes block the pathway of the signal proteins.

The diffusion constant[3] for a monomeric protein like CheY inside the a cell is $D \approx 7$ µm²/s[4]. The time scale for a particle to travel a distance L is $T = L^2/(6D)$. With the cell length of a E.coli bacterium, we get $T \approx 0.1$ s. This is amazingly much time that a signaling protein needs from the sensors to

the driving motor at the opposite end of an E.coli bacterium. Obviously, the brownian motion is very effective in preventing the tiny proteins from finding the fastest and shortest path between A and B, leaving enough time to reprogram a signal molecule when it touches other molecules.

Roughly speaking, the signal transmission through signal molecules between the sensors at one end and the drive motors at the other and takes place almost parallel to the longitudinal axis of the bacterium. Therefore, the motion of the proteins may be viewed as one-dimensional diffusion. This process is described by a simple differential equation.

$$\frac{d}{dt}c(x,t)=D\cdot\frac{d^2}{dx^2}c(x,t) \quad \text{, where}$$

- c is the concentration, the number of proteins per unit length.
- D is the diffusion constant.
- x is the position.
- t is the time.

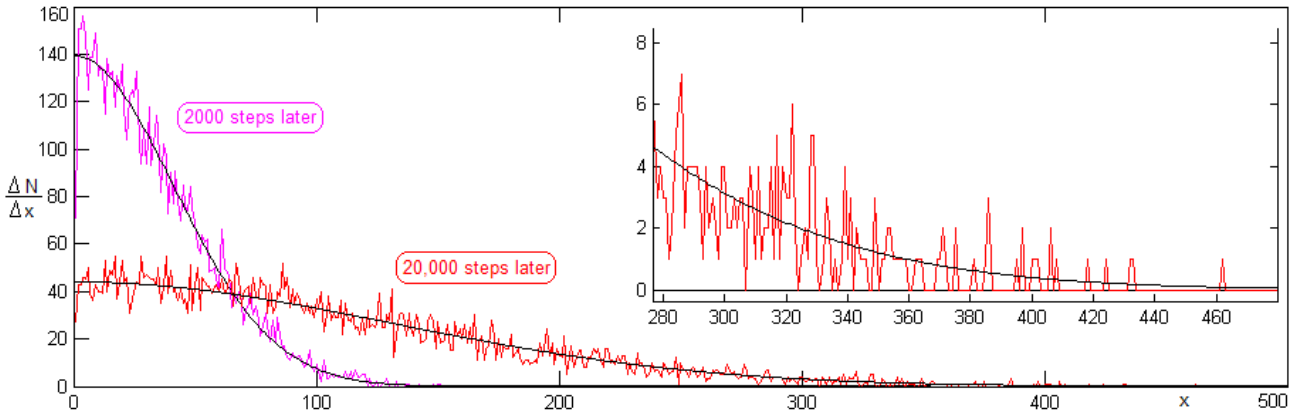
If at the time t = 0 all the N_0 proteins are concentrated near the center (x = 0) of an infinitely long bacterium, the time- and location-dependent number per unit length (Δx = 7 nm in our digital model) can be calculated with the formula

$$c_\infty(x,t)=\frac{\Delta N(x,t)}{\Delta x}=\frac{N_0}{\sqrt{4\pi Dt}}\exp\left(\frac{-x^2}{4Dt}\right) \quad \text{for} \quad -\infty<x<+\infty \quad \text{and} \quad 0\le t$$

In our model, c_∞ is the number of signal proteins in a thin slice (1·100·100) orthogonal to the x-direction.

For a model of finite length, the solution has to be adapted to the boundary conditions. To test the consistency of the digital model of a bacterium described above with experimental data, we select a total volume of 500·100·100 cubes. At the time t = 0, all N_0 = 7200 CheY signal proteins are concentrated in the volume 20·100·100 at the left edge of the cell. The proteins which drift further to the left as a result of the thermal movement are forced back into the model volume ("mirrored" would be the wrong expression because the cell wall does not act as a mirror). This doubles the number of particles to the right of this boundary. Similar happens when, after some time t_{max}, signal proteins try to leave the bacterium at the right edge of the model (x>500). The then-forced reversal can no longer be represented correctly by the simple formula above. With these limits, we get

$$c_m(x,t)=\frac{\Delta N(x,t)}{\Delta x}=\frac{2N_0}{\sqrt{4\pi Dt}}\exp\left(\frac{-x^2}{4Dt}\right) \quad \text{for} \quad 0\le x<500 \quad \text{and} \quad 0\le t<t_{max}$$



The particle density along the cell axis at different times after the start of diffusion. Magenta after 2000 time steps and red after 20000 time steps. In both cases, the best fit according to the above formula is superimposed in black color.

The very good agreement of the fitted curves (black) with the mean values of the colored simulation results confirms that the digital model of the bacterium correctly simulates the diffusion of the signal molecules. The technically necessary restriction to only 26 permitted jump directions does not cause any obvious errors.

The insertion in the upper right shows that the mathematical solution of the differential equation yields an unrealistically smooth course of the particle density (with fractions of whole molecules) which ignores the discrete nature of the molecules. It is questionable, whether this kind of idealized mathematics helps to decipher the secrets of signal transmission within a cell. The real composition of the cytoplasm differs very strongly from the idealized homogeneous medium and changes with every time step.

Now, we have to determine the value of one time step T_0 in the digital model. For x = 0, the above formula can be evaluated very easily and after 20000 time steps, one gets

$$\frac{\Delta N(0,t)}{\Delta x} = \frac{44}{7\,nm} = \frac{2 \cdot 7200}{\sqrt{4\pi \cdot D \cdot 20000\,T_0}}$$

and the result $T_0 = 3.0\,\mu s$. This time resolution exceeds by far the values used in earlier models (100 µs)[5] and allows a more detailed analysis of the protein reactions.

Experience has shown that instabilities arise when the time step T_0 exceeds about $\Delta x^2/D \approx 7$ µs. These are not to be expected with the parameters selected in the digital model.

Technical Details

Whether the simulation of a real process is time-consuming depends not only on the underlying physical formulas and the number of particles to be considered. The calculation time can be greatly reduced by means of advantageous programming.

E. coli contains 2.8 million proteins, 25,000 of which are signal proteins whose path has to be followed. Since chemical reactions (= signal transfer) can occur when two suitable molecules meet, the positions must be compared in pairs. Normal programming would be too time-consuming with 625 million comparisons. In order to achieve an acceptable working speed, this value has to be greatly reduced.

It is not necessary to compare molecules that are far away from each other. Therefore, it is useful to sort the data before each comparison. This is facilitated when the three (x, y, z) coordinates of each protein are packed into a single data word and then processed as a single unit. If the spatial resolution in the digital 3D model is limited to several hundred steps per coordinate direction, the complete coordinates of each particle inside a cell can be stored in a single 32-bit data word. In our model, the three coordinates are placed as follows in four contiguous bytes.

0xxx xxxx xxx0 yyyy yyyy yp0z zzzz zzzz

In a common computer, this quad word can be processed very fast with basic integer instructions - much faster than with three times as much floating-point arithmetic instructions. The p-bit indicates whether the CheY is phosphorylated or not. If necessary, more marker bits can be installed.

In the above example, 7200 signal proteins move 20000 times in random directions. Before each individual movement, three random numbers must be created (using a fast xorshift random number generator) and added to the current position. After the movement, it must be ensured that the protein is still inside the cell, so that the total quantity remains constant. With the help of packed coordinates and optimized assembler programming, these 144 million steps are calculated in as little as 2.7 seconds on a standard laptop.

Sometimes, due to the brown movement, possible reaction partners happen to be randomly and briefly in adjacent cubes of the model. The search costs a lot of time, but luckily it can be split into several independent tasks, which are processed by parallel threads. Each modern computer has several computer cores, which allow significant time savings when used correctly.

A complete cycle

The signal transmission in the bacterium is only understood when it can be successfully simulated with the help of many elementary steps, which are constantly repeated. A simple loop may contain the following blocks that exchange information with each other:

Begin

- Sensor activity: production of CheYp
- Brownian movement of all CheY / CheYp proteins
- Brownian movement of all CheZ proteins
- Find close neighborhoods of CheYp-CheZ proteins, react
- Find the CheYp proteins near the drive motors, let them react
- Record time and kind of interesting events

loop

Using the digital simulation, one loop-pass needs about 3300 µs and corresponds to an actual time step of 3 µs.

Perhaps the following questions can be answered by simulation: Are CheY and CheZ proteins uniformly distributed? How often do two of them approach each other sufficiently and transfer information? How long does it take to inform the drive motors to reverse their direction of rotation? How does the bacterium remember the nutrient concentration? Chemotaxis protein interaction networks can not answer these questions.

These are the questions that are now being addressed.

[1] M. Baker, P. Wolanin, J. Stock, Systems biology of bacterial chemotaxis, Current Opinion in Microbiology 2006, 9:187–192, https://www.researchgate.net/publication/7248483_Systems_biology_of_bacterial_chemotaxis
[2] M. Li, G. Hazelbauer, Cellular Stoichiometry of the Components of the Chemotaxis Signaling Complex, DOI: 10.1128/JB.186.12.3687–3694.2004, J. OF BACTERIOLOGY, June 2004, p. 3687–3694, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC419939/pdf/0244-04.pdf
[3] R. Milo, R. Phillips, Cell Biology by the numbers, Garland science, 2015, http://www.rpgroup.caltech.edu/courses/aph161/2012/files_2012/homework/CaltechJan2012.pdf
[4] M. Elowitz et. al., Protein Mobility in the Cytoplasm of E.coli, J. OF BACTERIOLOGY, Jan. 1999, p. 197–203, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC103549/pdf/jb000197.pdf
[5] K. Lipkow, S. Andrews, D. Bray, Simulated Diffusion of Phosphorylated CheY through the Cytoplasm of Escherichia coli, Journal of Bacteriology, Vol. 187, Jan. 2005, p. 45–53, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC538814/