

WestminsterResearch

<http://www.westminster.ac.uk/research/westminsterresearch>

Developing stochastic models for spatial inference: bacterial chemotaxis.

Yoon-Dong Yu¹

Yoonjoo Choi¹

Yik-Ying Teo²

Andrew R. Dalby^{1*}

¹ Department of Statistics, University of Oxford

² The Wellcome Trust Centre for Human Genetics, University of Oxford

*Andrew Dalby now works within the School of Life Sciences, University of Westminster

This is an electronic copy of an article originally published as:

Yu, Yoon-Dong and Choi, Yoonjoo and Teo, Yik-Ying and Dalby, Andrew R. (2010) Developing stochastic models for spatial inference: bacterial chemotaxis. PLoS ONE, 5 (5). e10464. doi: 10.1371/journal.pone.0010464.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk

Developing Stochastic Models for Spatial Inference: Bacterial Chemotaxis

Yoon-Dong Yu¹, Yoonjoo Choi¹, Yik-Ying Teo², Andrew R. Dalby^{1*}

¹ Department of Statistics, University of Oxford, Oxford, United Kingdom, ² The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Abstract

Background: Biological systems are inherently inhomogeneous and spatial effects play a significant role in processes such as pattern formation. At the cellular level proteins are often localised either through static attachment or via a dynamic equilibrium. As well as spatial heterogeneity many cellular processes exhibit stochastic fluctuations and so to make inferences about the location of molecules there is a need for spatial stochastic models. A test case for spatial models has been bacterial chemotaxis which has been studied extensively as a model of signal transduction.

Results: By creating specific models of a cellular system that incorporate the spatial distributions of molecules we have shown how the fit between simulated and experimental data can be used to make inferences about localisation, in the case of bacterial chemotaxis. This method allows the robust comparison of different spatial models through alternative model parameterisations.

Conclusions: By using detailed statistical analysis we can reliably infer the parameters for the spatial models, and also to evaluate alternative models. The statistical methods employed in this case are particularly powerful as they reduce the need for a large number of simulation replicates. The technique is also particularly useful when only limited molecular level data is available or where molecular data is not quantitative.

Citation: Yu Y-D, Choi Y, Teo Y-Y, Dalby AR (2010) Developing Stochastic Models for Spatial Inference: Bacterial Chemotaxis. PLoS ONE 5(5): e10464. doi:10.1371/journal.pone.0010464

Editor: Eshel Ben-Jacob, Tel Aviv University, Israel

Received: November 17, 2009; **Accepted:** April 5, 2010; **Published:** May 13, 2010

Copyright: © 2010 Yu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: During his DPhil studies, Y-DY is supported by the Sulsong Foundation of South Korea. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dalby@stats.ox.ac.uk

Introduction

Biological systems are by their very nature heterogeneous [1,2]. The cell contains many different compartments and even in the “fluid” cytosol many protein molecules are localised through interactions with the cytoskeleton or with the membrane/membrane bound proteins [3–5]. It is also possible that molecules within cells could be localised dynamically by reaction-diffusion processes, which are known to establish spatial organisation in chemical and biological systems [6,7].

Biological systems can also be characterised by their robustness and their ability to deal with both internal (intrinsic) and external (extrinsic) fluctuations [8,9]. Organisms need to be resistant to fluctuations in environmental conditions so that they can maintain developmental and regulatory control, otherwise external perturbations could have serious consequences for development and for maintaining homeostasis.

In order to build more realistic models of biological processes we need to build models that incorporate stochastic and spatial effects [10]. One solution to this modelling problem has been the creation of agent based or mesoscopic models [11,12]. In these models the system components are treated as objects whose behaviour is described by a set of physical rules, while the rest of the cellular components are ignored and treated as contributing to the background in which the simulations take place. One of the weaknesses of this approach is that it is difficult to use the models

for inference. Many potential models can be constructed with different sets of parameters or using variations on the physical rules used to describe the behaviour of the agents. It is not possible to construct an analytical framework that can be used to optimise the model parameters as would be the case in models that use ordinary differential equations where a likelihood method could be used [13].

For mesoscopic and agent based methods it is not possible to construct a likelihood function for the parameters and so a trial and error approach has to be taken, where successive models are built with different sets of parameters that are tested against the experimental data by measuring the closeness of fit. For this approach to work two criteria have to apply;

1. A series of parameter values have to be constructed that reflect some prior knowledge of the system.
2. The output of the model has to be evaluated against the experimental data.

A further implicit requirement is that different models should be distinguishable by the criteria used for the evaluation of the different models. For ordinary differential equation models, parameter scan methods have been developed and phase plane analysis can be used to examine the dynamic regimes for different parameter values [14,15]. In cases where parameters have not been determined experimentally phenomenological models can be

constructed where the properties of the model can be determined. In the case of mesoscopic models increased realism comes at a cost of computational time and so there is a need to keep the number of simulations that need to be carried out to a minimum.

For mesoscopic models, model evaluation can be particularly difficult in cases where there are a large number of parameters or where the model is very complex and the experimental data only measures a small number of variables. In these cases the model is likely to be either over-fitted or the models will be indistinguishable unless there is a very sensitive scoring of models.

Bacterial Chemotaxis

Bacterial Chemotaxis is the process by which bacteria sense gradients of specific chemicals, which can either act as attractants or repellents [16–18]. Attractants are usually sources of nutrients for the bacterium. In the case of the bacterium *Escherichia coli* the swimming behaviour of the cell is controlled by an alternating sequence of linear swimming behaviour punctuated by periods of tumbling motions. These two behaviours are controlled by the bacterial flagella motors that cause the whip like flagella to form a coordinated bundle which acts like a semi-rigid propeller, when they rotate counter-clockwise or for the bundle to break and tumbling to begin when they turn in a clockwise direction.

The rotation of the bacterial flagella motor is controlled by a signal transduction pathway that starts from the chemical receptors embedded in the cellular membrane that detect the attractant and repellent molecules. The signal is then passed through a series of intermediate species that are either localised close to the receptor array or that are freely diffusing in the cytoplasm. The key diffusive molecule is known as CheY which can be found in a phosphorylated and unphosphorylated state. The binding of the phosphorylated state CheY-P to the flagella motors increases the probability of a transition from counter-clockwise to clockwise rotation of the motor. So that a higher concentration of CheY-P in the cell promotes tumbling.

Figure 1 shows a schematic representation of the chemotaxis network. In this case only one of the five types of cellular receptors is shown, this corresponds to the aspartate receptor Tar. Aspartate is an example of an attractant molecule, along with sugars and

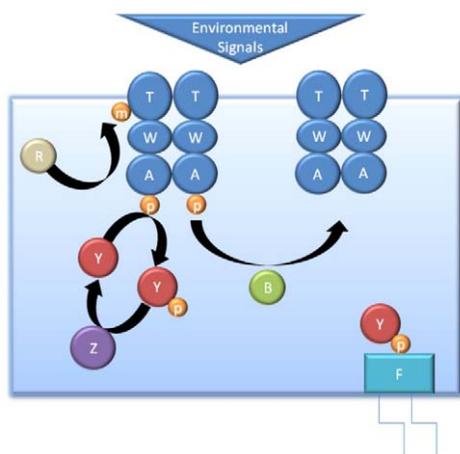


Figure 1. Schematic representation of the *E. coli* chemotaxis system. The Tar receptors are labelled T, CheW, CheY, CheA, CheZ, CheB and CheR are labelled as W, Y, A, B, Z and R respectively and the flagellar motor (FliM) is labelled as F. Phosphate and methyl groups are shown as orange circles labelled p or m respectively. doi:10.1371/journal.pone.0010464.g001

most amino acids (leucine is an exception). CheY is phosphorylated by the activated form of the kinase CheA and dephosphorylated by the phosphatase CheZ. Binding of attractants to the receptor decreases the rate of CheY phosphorylation reducing the concentration of CheY-P and therefore reducing the tumbling rate.

An important feature of chemotaxis is that it detects gradients, so that in an environment where there is an equal concentration of attractant in every direction the bacterium will resume tumbling as there is no preferred direction to take. The system is very sensitive to differences in gradient and can detect differences of one molecule per cell volume per micron. Once the cell is again in a homogeneous environment it returns to its equilibrium tumbling frequency, which is the same regardless of the environmental level of attractant.

The fluctuations between counter-clockwise and clockwise rotations for the flagella motor have been measured in single cell studies for unstimulated cells, where there are no attractant or repellent molecules in the bacterial media [19]. This experimental data can be used to construct a model of the equilibrium state of the system, where we can investigate the localisation of the signal transduction pathway components. Experimental studies using green fluorescent protein labelled CheZ molecules have shown that the phosphatase is localised to the receptor array, but there is a need to independently quantify the localisation more precisely [20–23].

In this work we use mesoscopic models of *E. coli* chemotaxis to infer the localisation of CheZ. We have developed a statistical analysis of the output of the stochastic simulations that allows the different models to be distinguished and the localisation parameters to be determined.

Results and Discussion

The movement of a single **unstimulated** (i.e. in the absence of either repellent or attractant) bacterial cell has been recorded over 170 minutes at 0.01 second intervals. The time series has been recorded as a binary time series where -1 represents counter-clockwise rotations of the flagella motor and 1 represents clockwise rotations. The time series can be broken down into different distributions for the time intervals in clockwise and counter-clockwise motions. Korobkova and co-workers showed that the clockwise intervals are characterised by an exponential like distribution and the counter-clockwise intervals by a long-tailed deviation from the exponential distribution [19].

The stochastic simulations are based upon the experimentally determined values of the kinetic parameters for the series of chemical reactions equations that form the signal transduction network. These parameters are not part of the experimentally determined validation set and there is a non-linear relationship between them, that we are trying to determine through the simulations. The parameters that we wish to infer are the volume in which the CheZ molecules are localised (α) and the fraction of CheZ molecules that are localised within this volume (β). An initial simulation (Simulation 0) was run with no CheZ localisation at all, the other twelve simulations were carried out using the CheZ localisations given in table 1.

In order to make the comparison with the experimental data the switching function from counter-clockwise to clockwise rotation of the flagella motors has to be calculated from the concentration of CheY-P molecules. This step in the reaction cannot be modelled explicitly due to the lack of experimentally determined kinetic parameters, and so this is done using a threshold transformation as described by Emonet and co-workers.

Table 1. CheZ localisation parameters for the simulations.

β	0.25	0.5	0.75	1
α				
0.01	Simulation 1	Simulation 2	Simulation 3	Simulation 4
0.05	Simulation 5	Simulation 6	Simulation 7	Simulation 8
0.1	Simulation 9	Simulation 10	Simulation 11	Simulation 12

Where α is the fraction of the cell volume around the anterior array in which the CheZ molecules are localised and β is the fraction of CheZ molecules localised in this volume.

doi:10.1371/journal.pone.0010464.t001

$$\text{Switching Threshold} = \langle \text{CheY-P} \rangle - 0.55\sigma(\text{CheY-P}) \quad (1)$$

Where $\langle \text{CheY-P} \rangle$ is the mean number of CheY-P molecules and $\sigma(\text{CheY-P})$ is the standard deviation.

Verifying the Equilibrium Behaviour

With any stochastic simulation it is important to verify that the system is reproducing the expected variation and that the variability is not divergent. In this case the cells are unstimulated and so the signal transduction network should be at equilibrium. This implies that the variation in the levels of CheY-P should not show any significant upward or downward trend over long periods of time, but should fluctuate around a mean value. Figure 2 shows the variation in CheY-P concentration for the model where CheZ is completely delocalised. The switching threshold is shown in blue. The trace has been smoothed over a sliding window of 0.3 seconds. In this case the number of CheY-P molecules does not show any underlying linear trend and the simulation can be assumed to have reached equilibrium. This is not the case for the model where CheZ is completely localised, where there is an underlying linear trend and the system does not achieve

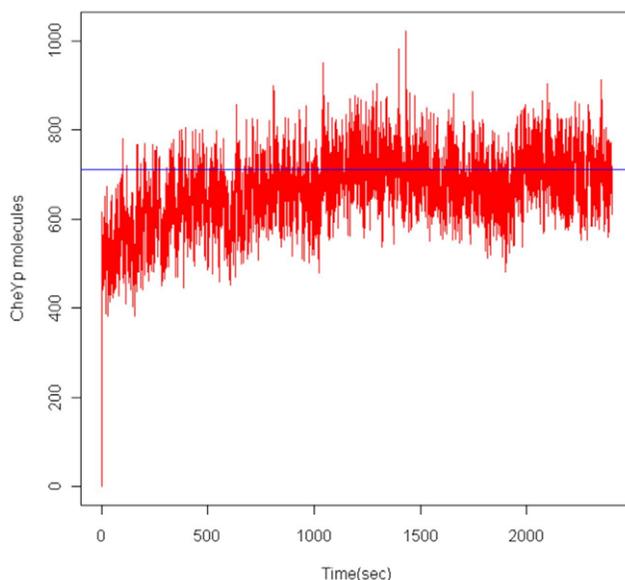


Figure 2. The variation in number of CheY-P molecules for the freely diffusing CheZ model.

doi:10.1371/journal.pone.0010464.g002

equilibrium, and so the complete rigid localisation of CheZ is not consistent with our model.

From all of the choices of CheZ localisation parameter the simulations that did not converge to equilibrium and so were discounted from further statistical analysis were simulations 3,4,5,6 and 8.

One of the weaknesses of mesoscopic models is the computational resources that are required to carry out the calculations. As these calculations are inherently stochastic it is important to verify that the output correctly samples the possible outcomes. There are two possible approaches that have been used extensively in the molecular dynamics literature. The first is to run a single simulation for a very long time periods so that the sampling is as complete as possible and the second is to run many different short simulations and to average them [24]. In this case as the system is at equilibrium and we are measuring time series data the first, single simulation approach is more appropriate but we can verify this by checking the variability of the switching threshold from the simulations, in order to demonstrate the robustness of the method. The mean and the sample standard deviation for the switching thresholds of ten repeats of Simulation 0, where CheZ is freely diffusing are 721.8893 seconds and 0.8981 seconds respectively. This indicates that our simulator is robust and that the true mean at the 95% level can be found in the confidence interval between 721.2468 and 722.5317 seconds for the completely delocalised model.

Preliminary Comparison of the Completely Localised and Freely Diffusing Models of CheZ Localisation

Figure 3 shows the counter-clockwise and clockwise interval distributions for the model where CheZ is diffusing freely. In the experimental data the counter-clockwise intervals vary from 0 to 179.13 seconds, whilst that of the clockwise intervals ranges from 0 to 2.77 seconds. Previous studies have put the two sets of interval data on the same scale as counter-clockwise switching intervals of longer than 20 seconds are rare events, and so we restricted our analysis to the range 0 to 20 seconds.

An important feature of the counter-clockwise distribution is the presence of a peak in the 0 to 1.5 second region. This bimodality should arise naturally from the simulations if they are to be reliable

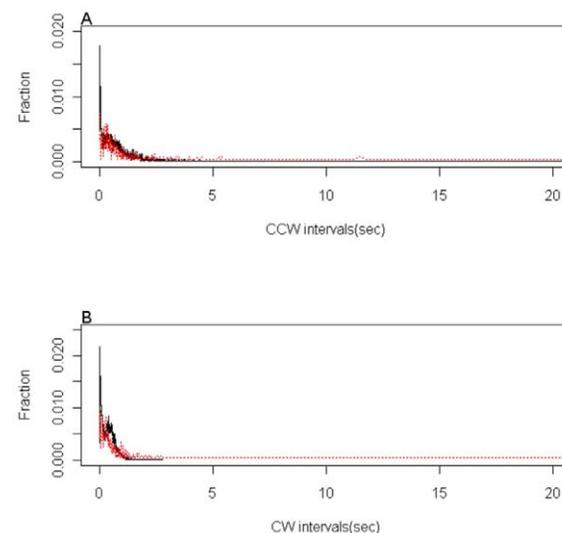


Figure 3. The counter-clockwise and clockwise interval distributions for the freely diffusing CheZ model.

doi:10.1371/journal.pone.0010464.g003

models of the signal transduction network. Almost all of the simulations except for those with a very high degree of CheZ localisation exhibit this bimodality. A complete breakdown of the results are given in table 2. These show that our simulations can robustly reproduce the experimental data. The finding that the total localisation of CheZ cannot reproduce the empirical switching behaviour, agrees with the experimental finding that modified bacteria where CheZ cannot be localised still retain chemotactic activity [21]. Ideally we could further corroborate our model if time series data for the swimming of these modified organisms was available, so that we could check our results for the switching behaviour in the case of totally delocalised CheZ against the actual observed behaviour.

The clockwise distribution only contains a single peak at 0.5 seconds. In the unlocalised CheZ simulations there is a slight shift of this peak to 0.3 seconds, but the change is much larger for the completely localised model. By considering these extreme models it indicates that complete localisation is less consistent with our model than a freely diffusive model but it is necessary to build a more detailed statistical analysis to infer the actual degree of localisation.

Statistical Evaluation of the Simulated Interval Distributions

The simplest method for comparing an observed distribution to its underlying distribution is the use of the quantile-quantile or Q-Q plot. The Q-Q plot shows how the quantiles (percentiles for example) of the two distributions match to each other. The larger the number of quantiles that are compared the higher the resolution of the comparison. For a perfect match all of the points should lie on the diagonal. As Korobkova and co-workers pointed out in their earlier work, there will be expected variation in the switching frequency distributions of different *E.coli* cells [19]. We therefore have to consider the experimental data as a sample from the population switching distributions. We can then use this data to determine point estimators of the parameters for an appropriate population switching distribution, which will have the form that best fits the experimental data. In the case of the clockwise interval

distribution this follows an exponential distribution. The Q-Q plot can therefore be used to measure the deviations between the simulations and experimental data and the exponential distribution. In the case of the counter-clockwise interval distribution there is a long tailed deviation from the exponential distribution but the underlying distribution is still exponential.

Point estimations were made using Maximum Likelihood Estimation for the rate λ are made for both the counter-clockwise and clockwise intervals of the simulated data. For example in the freely diffusing case these are estimated to be $\hat{\lambda}_{CCW} = 0.5643$ seconds and $\hat{\lambda}_{CW} = 1.2944$ seconds.

Figure 4 shows the Q-Q plot for the freely diffusing CheZ model comparing the quantiles for the simulated data to that of the fitted exponential function. For the freely diffusing CheZ simulation the median of the counter-clockwise interval distribution is shifted 0.6 seconds. While over 50% of the data points lie on the diagonal there is a significant amount of deviation for intervals longer than 5.7 seconds, which represents the long-tailed deviation from exponential. The short clockwise interval distribution is better described by an exponential function and so more of the points lie on the diagonal.

Locally Weighted Scatterplot Smoothing

As mentioned previously there is often a lack of simulation and experimental data because of the computational cost of simulations and the resource costs of collecting the experimental data. In this case only a single empirical sampling of the swimming motions is available. By creating a smoothed distribution for the simulated and experimental data we can compare directly the different interval distributions. The smoothed distributions can be used to calculate the 95% confidence intervals. If there is a complete overlap between the fitted simulated distributions and the experimentally determined confidence intervals then the two will share the same true fraction mean at the 5% significance level. The degree of overlap will reflect the closeness of the simulation to the experimentally observed values.

Smoothing was carried out for the simulated and experimental data using LOESS fittings with different bandwidths, $h(x) = 0.8, 0.6, 0.4$ and 0.2 seconds and using different polynomial

Table 2. Summary of the simulation results.

Simulation	Existence of the Chemical Equilibrium	Statistical Agreement	Figures
0	yes	yes	3A and 3B
1	yes	yes	5A and 5B
2	yes	yes	5C and 5D
3	no	-	-
4	no	-	-
5	no	-	-
6	no	-	-
7	yes	yes	5E and 5F
8	no	-	-
9	yes	yes	5G and 5H
10	yes	yes	5I and 5J
11	yes	yes	5K and 5L
12	yes	no	5M and 5N

The absence of a chemical equilibrium is indicated by a long term trend in the switching behaviour of the system.

doi:10.1371/journal.pone.0010464.t002

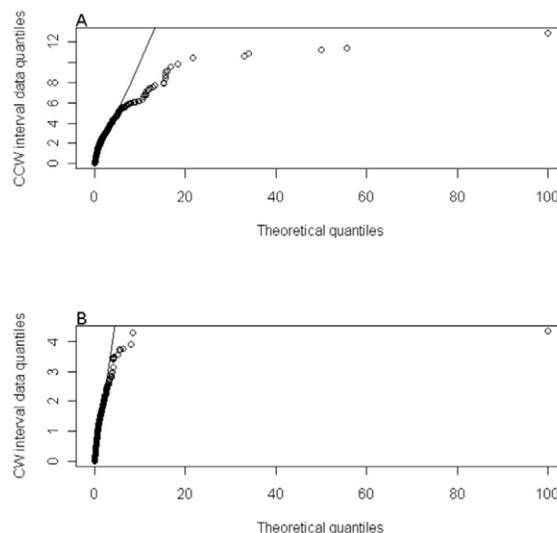


Figure 4. The Q-Q plot for the freely diffusing CheZ model, comparing the simulated quantiles against an exponential function fitted to the experimental data.

doi:10.1371/journal.pone.0010464.g004

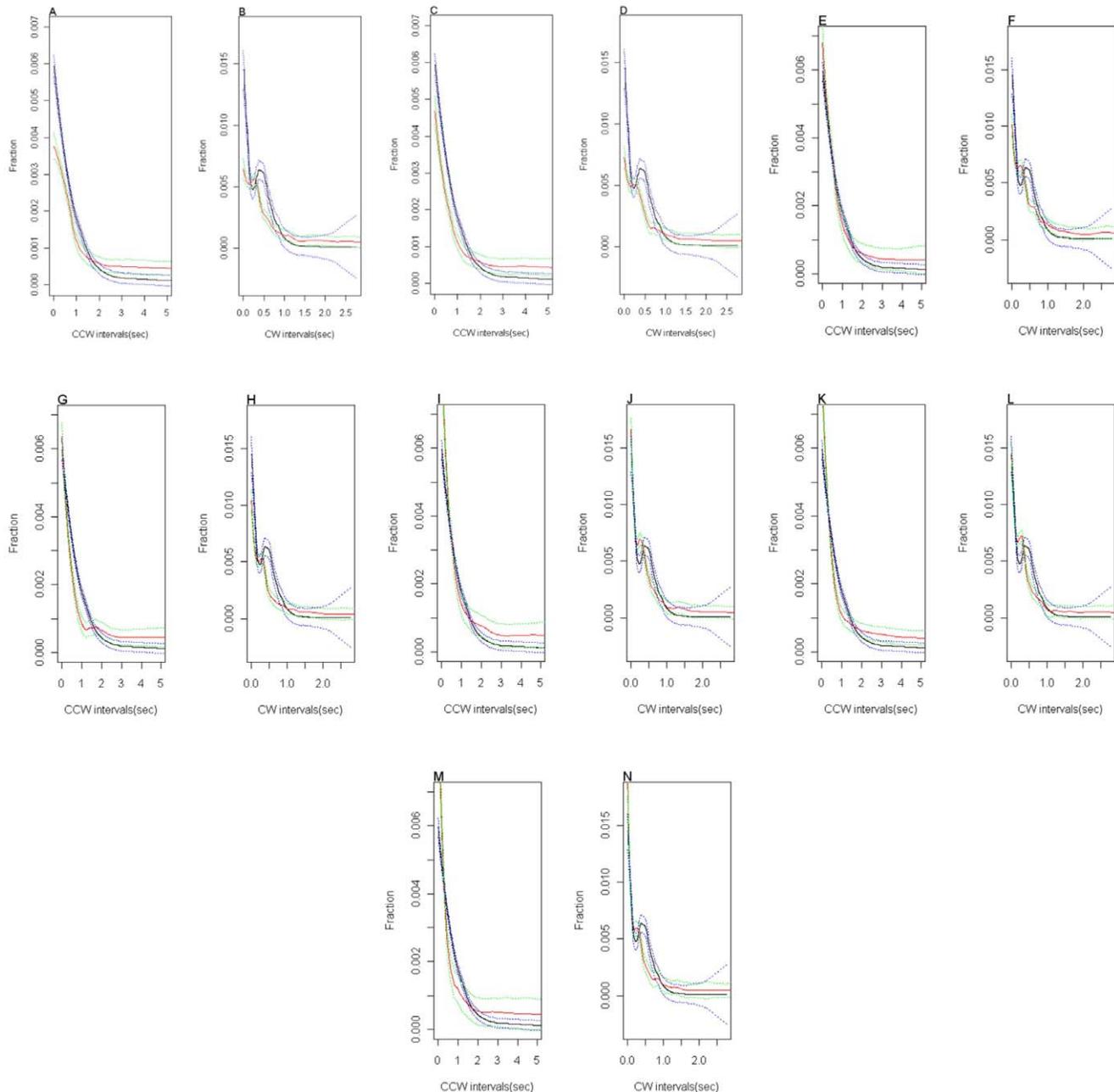


Figure 5. The LOESS fitting of the clockwise and counter-clockwise distributions that achieved equilibrium. The letters correspond to the simulations given in table 2.
doi:10.1371/journal.pone.0010464.g005

degrees such as local quadratic, cubic, linear and constant. Bandwidth and polynomial degree have to be selected carefully so as not to over-smooth the data. Figure 5 shows the LOESS fitting of the clockwise and counter-clockwise distributions from the simulations that achieved an equilibrium distribution, and the corresponding measures of statistical closeness for each simulation are given in table 3. There is excellent agreement between the two traces in many cases and the bimodality of the counter-clockwise interval distribution is also often reproduced.

The LOESS fittings are sensitive enough to be able to distinguish between the different sets of localisation parameters. These results confirm that the absence of CheZ localisation does

not diminish chemotaxis (Simulation 0 the model with freely diffusing CheZ). This result agrees with the empirical findings for cells with a truncated form of CheZ which cannot be localised. They also show that the statistically closest simulation to the experimentally observed distributions was simulation 2. In this case the amount of the CheZ localisation will be about 50% of the total number of CheZ molecules in the 1% cell volume around the anterior array of the *E.coli* cell, which agrees well with Manson's empirical estimation. However, we can also infer that the optimum extent of CheZ localisation at the anterior array can depend on the volume in which the CheZ is localised and that when the volume for the CheZ localisation is 1% of the whole cell volume, the

Table 3. Statistical closeness between the simulations and the experimental data.

Simulation	%Statistical Closeness	
	CCW Intervals	CW Intervals
0	90.3	81.3
1	93.1	80.6
2	100	89.2
7	94.1	80.6
9	98.4	87.4
10	94.3	83.8
11	95.3	86

doi:10.1371/journal.pone.0010464.t003

amount of CheZ localisation is between 25% and 50% of the total number of CheZ molecules. For a 5% cell volume for CheZ localisation it is inferred that 75% of the total CheZ molecules are present, and in the case of 10% cell volume for the CheZ localisation it is between 25% and 75% of the total number of CheZ molecules. Further simulations could be used to increase the accuracy of the inference by sampling the parameter space more finely.

Conclusions

Using a mesoscopic model of the bacterial chemotaxis signal transduction network we have been able to infer the localisation of CheZ.

Exploratory analysis using the smoothed distribution of counter clockwise and clockwise switching intervals distinguished between parameter sets for the volume of localisation (V_A) and the number of localised molecules ($CheZ_A$), and show that incorrect values can lead to unrealistic simulations in which simulated systems do not reach the chemical equilibrium. In addition, using Q-Q plot analysis we can show that only systems that converge to equilibrium also capture the same underlying statistical distributions as that of the experimental data.

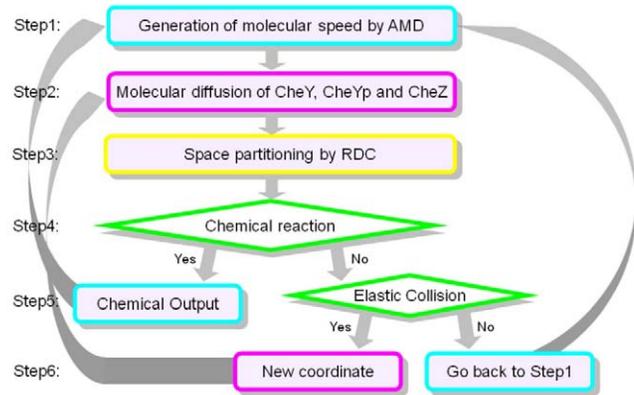
Through statistical comparisons of the chosen simulations and the experimental data, we have shown that chosen simulations from with our new algorithm agree closely with the empirically determined mean swimming responses of an observed wild-type cell in a medium in which no attractant was present. The statistically closest simulation to the experimental data was when 50% of the CheZ molecules were localised in 0.04% of the cell volume around the receptor arrays. Considering the counter clockwise intervals from 0s to 20 seconds, about 95.63% of the true fraction means of the simulated data are close to the experimental data. In the case of the clockwise intervals from 0s to 2.77 seconds, about 84.66% of the true fraction means of two populations are statistically close.

This is in very good agreement with the experimentally estimated degree of CheZ localisation from green fluorescent protein labelled experiments (M. Manson personal communication).

Materials and Methods

Simulation Algorithm

Simulations were carried out with our own implementation of the Andrews and Bray version of the Smoluchowski algorithm written in Python. Figure 6 shows the overall outline of the

**Figure 6.** An outline of the MESMAX algorithm.

doi:10.1371/journal.pone.0010464.g006

algorithm. The Andrews and Bray approach allows longer time-steps to be used than in the original algorithm by using effective molecular radii that indicate the volume swept out by the diffusive process in the longer time-step. These modified radii are particularly important in calculating if reactions have taken place.

The differences between this algorithm and that used within Smoldyn are the use of a Maxwell-Boltzmann distribution for the sampling of the molecular velocities, and the use of Recursive Dimensional Clustering (RDC) to simplify the collision detection process [25]. RDC uses a space partitioning algorithm to divide the simulation into volumes that contain a smaller number of objects which are then checked for collisions rather than trying the brute force collision detection for the entire simulation space.

There are a variety of possible boundary conditions that can be applied. In this case a mixture of periodic boundary conditions and impermeable surfaces was used. If during the course of a time-step the molecules make contact with the surface then it is reflected like light from a mirror by an inert impermeable surface. Any molecules that diffuse past a boundary are transferred across the system following a straight line path over the course of the time-step.

The time-step for the calculations was set to be 0.1 milliseconds as a compromise in giving the highest degree of accuracy without increasing the computational over-head unnecessarily.

Simulation Set-up

Phosphate is much smaller than the other molecules in the simulation and so it can be treated as a dimensionless point particle. The masses of CheA, CheY (and CheY-P), CheZ and FliM are 71, 14, 24 and 38 kDa respectively. Only CheY, CheY-P and CheZ are mobile in the cytoplasm and so their diffusion constants are needed for the simulation. CheY and CheY-P have a diffusion constant of $10 \mu\text{m}^2/\text{s}$ and CheZ has a diffusion constant of $6 \mu\text{m}^2/\text{s}$ [3]. The radii of CheA, CheY (and CheY-P), CheZ and FliM were estimated to be 5.5nm, 2.3nm, 3.9nm and 2.7nm respectively using the PSA program within JOY [26].

Figure 7 shows the *E.coli* simulation system. An *E.coli* cell is approximately cylindrical with a length of $2.5 \pm 0.6 \mu\text{m}$ and a diameter of $0.88 \pm 0.9 \mu\text{m}$ [27]. There can be a considerable difference in the ratio between length and diameter depending on growth conditions. For the simulations this is approximated to a rectangular box with a length of $2.5 \mu\text{m}$ and a width and height of $0.78 \mu\text{m}$. The 6700 CheA molecular dimers are located in a radial array of area $267.8 \mu\text{m}^2$, 20nm from the inside surface of the anterior cell wall. This corresponds to the active portion of the

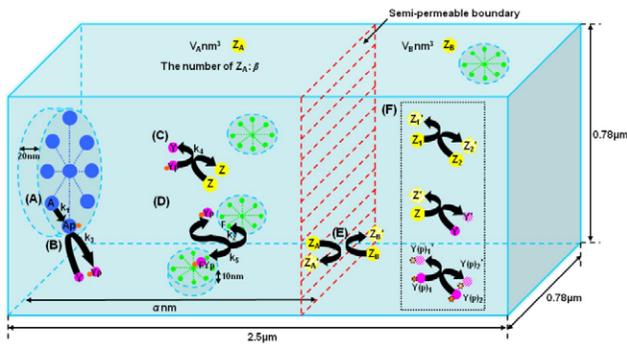
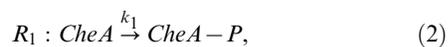


Figure 7. The *E. coli* simulation system.
doi:10.1371/journal.pone.0010464.g007

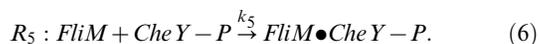
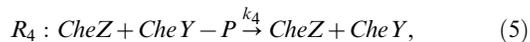
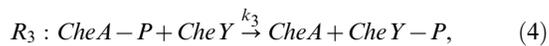
receptor-CheA complex that is responsible for signal transduction from the membrane receptors. Four rings each 45nm in diameter and made of 34 regularly spaced FliM monomers are placed on the long side walls of the cell, positioned 10nm from the inside surface. Each ring is placed randomly on a different side wall [28]. The 1600 CheZ dimers are located according to the localisation parameters either in a volume close to the anterior ray or freely diffusing in the cytoplasm [29]. The cytoplasm is then seeded with 8200 freely diffusing CheY monomers [29].

The set of reaction equations are given below and the rate constants are given in table 4.

There are two unimolecular chemical reactions:



There are three bimolecular reactions:



Statistical Methods

We employed the use of exploratory procedures to examine the underlying structure between the experimental and simulated data. Since reactions within a system will eventually reach a chemical equilibrium, we assess its existence within our simulations by assessing constancy of variance and amplitude of the smoothed stochastic time evolutions of CheY-P molecules with a sliding average of width 0.3s. Subsequently we verify that our simulation results have the same underlying statistical features as that of the

References

- Ridgway D, Broderick G, Lopez-Campistrous A, Ru'Aini M, Winter P, et al. (2008) Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophysical Journal* 94: 3748–3759.
- Lewontin R (2000) *The Triple Helix*. CambridgeMassachusetts: Harvard University Press.
- Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *Journal of Bacteriology* 181: 197–203.
- Mochly-Rosen D (1995) Localization of protein kinases by anchoring proteins: A theme in signal transduction. *Science* 268: 247–251.

Table 4. Table of kinetic rate constants.

Rate Constant	Value	Reference
k_1	0.4 s^{-1}	[32]
k_2	20 s^{-1}	[33]
k_3	$1 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$	[34]
k_4	$1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$	[29]
k_5	$5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$	[33]

The literature values for the rate constants used in the reaction equations.
doi:10.1371/journal.pone.0010464.t004

experimental data using a comparison of the percentiles of the theoretical and simulated distributions using a quantile-quantile (Q-Q) plot. Theoretical distributions for the counter-clockwise and clockwise interval data are generated from appropriate exponential distributions with rate parameters estimated using the maximum likelihood approach within the R statistical environment [30].

We estimate the true mean patterns of the experimental and simulated data by the use of locally weighted scatterplot smoothing (LOESS) using the loess function in R [30,31]. LOESS fittings were carried out at different bandwidths, $h(x) = 0.8, 0.6, 0.4$ and 0.2 , and varying polynomial degrees such as local quadratic, cubic, linear and constant for both the simulations and the experimental data. Here, the bandwidth and the polynomial degree are chosen to give the best fit to the data so as not to over-smooth it.

By applying LOESS in order to compare the simulated and experimental cases, we can compute approximate 95% confidence intervals for the LOESS fits for each dataset. If these confidence intervals contain all of the LOESS estimates for the experimental data, this will mean that the simulation and the experimental data statistically have the same true fraction mean at significance level $\alpha = 0.05$. However, if it does not contain all the LOESS estimates of the experimental data, we can still say that true fraction means of two populations are statistically close without loss of generality so long as their approximate confidence intervals overlap. In our paper, the degree of the statistical closeness between two populations is represented by the overlapped proportion of their approximate confidence intervals. This gives a more quantitative measure of the degree of agreement between the two distributions.

Acknowledgments

The authors would like to thank Michael D. Manson for his helpful discussions regarding the experimental data for the localisation of CheZ and to Steven S. Andrews for advice on the stochastic and spatial modelling approach.

Author Contributions

Conceived and designed the experiments: ARD. Performed the experiments: YDY YC. Analyzed the data: YDY YYT. Contributed reagents/materials/analysis tools: YC. Wrote the paper: YDY ARD.

5. Janmey PA (1998) The cytoskeleton and cell signaling: Component localization and mechanical coupling. *Physiological Reviews* 78: 763–781.
6. Luby-Phelps K (2000) Cytoarchitecture and physical properties of cytoplasm: Volume, viscosity, diffusion, intracellular surface area, volume 192.
7. Goryachev AB, Pokhilko AV (2008) Dynamics of cdc42 network embodies a turing-type mechanism of yeast cell polarity. *FEBS Letters* 582: 1437–1443.
8. Elowitz MB, Levine AJ, Siggia ED, Swain PD (2002) Stochastic gene expression in a single cell. *Science* 297: 1183–1186.
9. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 94: 814–819.
10. Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys Biol* 1: 137–151.
11. Paton R, Gregory R, Vlachos C, Saunders J, Wu H (2004) Evolvable social agents for bacterial systems modeling. *IEEE Transactions on Nanobioscience* 3: 208–216.
12. Emonet T, Macal CM, North MJ, Wickershamand CE, Cluzel P (2005) Agentcell: A digital single-cell assay for bacterial chemotaxis. *Bioinformatics* 21: 2714–2721.
13. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6: 187–202.
14. Mendes P, Kell DB (1998) Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* 14: 869–883.
15. Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Research* 13: 2467–2474.
16. Koshland DE (1981) Biochemistry of sensing and adaptation in a simple bacterial system. *Annual Review of Biochemistry* 50: 765–782.
17. Wadhams GH, Armitage JP (2004) Making sense of it all: Bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* 5: 1024–1037.
18. Yi T, Andrews BW, Iglesias PA (2007) Control analysis of bacterial chemotaxis signaling. *Methods in Enzymology* 422: 123–140.
19. Korobkova E, Emonet T, Villar J, Shimizu T, Cluzel P (2004) From molecular noise to behavioural variability in a single bacterium. *Nature* 428: 574–578.
20. Sourjik V, Berg H (2000) Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. *Molecular Microbiology* 37: 740–751.
21. Cantwell B, Draheim R, Weart R, Nguyen C, Stewart R, et al. (2003) CheZ phosphatase localizes to chemoreceptor patches via CheA-short. *Journal of Bacteriology* 185: 2354–2361.
22. Liberman L, Berg H, Sourjik V (2004) Effect of chemoreceptor modification on assembly and activity of the receptor-kinase complex in *Escherichia coli*. *Journal of Bacteriology* 186: 6643–6646.
23. Rao C, Kirby J, Arkin A (2005) Phosphatase localization in bacterial chemotaxis: Divergent mechanisms, convergent principles. *Physical Biology* 2: 148–158.
24. Yaffee RA, McGee M (2000) *An Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*. London, UK: Academic Press.
25. DeLoura M (2001) *Games Programming Gems 2*. Boston/Massachusetts: Charles River Media.
26. Mizuguchi K, Deane C, Blundell T, Johnson M, Overington J (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics* 14: 617–623.
27. Darnton N, Turner L, Rojevsky S, Berg H (2007) On torque and tumbling in swimming *Escherichia coli*. *Journal of Bacteriology* 189: 1756–1764.
28. Lipkow K, Andrews S, Bray D (2005) Simulated diffusion of phosphorylated CheY through the cytoplasm of *Escherichia coli*. *Journal of Bacteriology* 187: 45–53.
29. Li M, Hazelbauer G (2004) Cellular stoichiometry of the components of the chemotaxis signaling complex. *Journal of Bacteriology* 186: 3687–3694.
30. R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org>. ISBN 3-900051-07-0.
31. Cleveland WS, Grosse E, Shyu WM (1992) *Local Regression Models*. New York: Wadsworth. pp 309–376.
32. Francis N, Levit M, Shaikh T, Melanson L, Stock J, et al. (2002) Subunit organization in a soluble complex of tar, CheW, and CheA by electron microscopy. *Journal of Biological Chemistry* 277: 36755–36759.
33. Sagi Y, Khan S, Eisenbach M (2003) Binding of the chemotaxis response regulator CheY to the isolated, intact switch complex of the bacterial flagellar motor. lack of cooperativity. *Journal of Biological Chemistry* 278: 25867–25871.
34. Stewart R, Jahreis K, Parkinson J (2000) Rapid phosphotransfer to CheY from a CheA protein lacking the CheY-binding domain. *Biochemistry* 39: 13157–13165.