UPTEC X 02 021
OCT 2002
ISSN 1401-2138

MARIA ANDÉR

SmartCell – a general framework for whole-cell modeling and simulation

Master’s degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 02 021	**Date of issue 2002-10**

Author

Maria Andér

Title (English)

SmartCell – a general framework for whole-cell modeling and simulation

Title (Swedish)

Abstract

SmartCell is a framework for modeling and simulation of biological processes in whole cells. Modeling is performed in an XML-based model description format, derived from biological concepts, that has been developed as part of the framework. The simulation algorithm is based on mesoscopic kinetics, and supports diffusion and localisation of components in the cell. This work presents a first implementation of the SmartCell framework and an initial assessment of its modeling capacity and predictive power. The general conclusion is that SmartCell is highly accurate for modeling reaction kinetics and diffusion, but suffers from speed limitations for complex models including processes occurring on different time-scales.

Keywords

whole-cell modeling, computer simulation, mesoscopic, bacterial chemotaxis

Supervisors

Luis Serrano
European Molecular Biology Laboratories, Heidelberg, Germany

Examiner

Måns Ehrenberg
Dep of Cell and Molecular Biology, Uppsala university

Project name	Sponsors
SmartCell	**TTA Technotransfer AB**
Language: **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages: **58**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

SmartCell – a general framework for whole-cell modeling and simulation

Maria Andér

Sammanfattning

Datorsimuleringar har länge använts inom biologin, men det är inte förrän de senaste åren som forskare försökt simulera hela celler. SmartCell är ett program för att beskriva och studera beteendet hos en hel cell.

Det finns flera användningsområden som gör simuleringar av biologiska system intressanta. Först och främst för att undersöka om en förklarings-modell för beteendet hos ett visst system verkar riktig. Datorsimuleringar har också fördelen att egenskaper hos systemet kan mätas exakt och under mycket korta tidsintervall, samt att man undviker de praktiska problem som kan förekomma vid experimentellt arbete. Naturligtvis kommer aldrig datorsimuleringar att ersätta laboratoriearbete, men de kan fungera som ett komplement, och faktiskt ge uppslag till nya experiment.

En anledning till att datorsimuleringar på hela celler görs först nu är att modellerna blir mycket komplicerade och utvecklingen därför begränsats av tillgång till snabba datorer. En annan viktig orsak är att forskarna inte haft tillräckligt med kunskap för att modellera cellens funktioner. Båda dessa faktorer är fortfarande aktuella, och verkar begränsande på möjligheterna att göra omfattande cellsimuleringar.

I det här arbetet har möjligheterna hos SmartCell som verktyg för dator-simuleringar utvärderats. Den modell som testats är kemotaxis i bakterien *Escherichia coli*. Med hjälp av kemotaxis kan bakterien analysera sam-mansättningen av näringsämnen och gifter i sin omgivning och därifrån avgöra vilken rörelseriktning som är mest gynnsam att fortsätta simma i.

Många anser att biologin i framtiden alltmer kommer att flyttas in i datorn. Ett nytt tvärvetenskapligt forskningsfält, *systembiologin*, verkar bekräfta dessa idéer. Systembiologerna förordar en slags "holism" i biologin eftersom man poängterar att ett biologiskt system inte kan förstås annat än i sitt sammanhang och under dynamiska förhållanden. SmartCell och liknande projekt är bara början på detta systemtänkande i biologin som vi med säkerhet kommer att få se mycket mer av i framtiden.

Examensarbete 20 p, Civilingenjörsprogrammet i molekylär bioteknik

Uppsala universitet, oktober 2002

Summary

The SmartCell project was initiated in the Spring of 2000 with the aim to develop a general framework for modeling and simulation of biological processes in a whole-cell context. Specific characteristics of SmartCell is a mesoscopic, stochastic reaction kinetics, support of localisation and diffusion, a general model geometry and a highly semantic model description language. These criteria are there to allow accurate modeling and simulation of reaction networks, comprising localised components, complex dynamics and low numbers of reactant particles, such as in signal transduction pathways and genetic networks.

In this work, a first implementation of the SmartCell framework has been accomplished and an initial assessment of the modeling capacity as well as a validation of the predictive power is presented. Results show that the kinetics and diffusion algorithms are correct and converge towards the macroscopic descriptions for large numbers of molecules. A more complex system, the chemotaxis signaling network in *E coli*, has been modeled and the behaviour compared with results from previous modeling experiments. The difference in time-scales between diffusion, excitation and adaptation reactions in the chemotaxis model posed problems of computational speed. To overcome the computational overload from diffusion a simplified diffusion model was introduced, and for other fast processes assumptions on instantaneous equilibria were made. Results show that a gradient of the response regulator CheYp exists in the cell, originating from a limited rate of diffusion and the clustering of the chemotaxis receptors at one cell pole. The implications of this gradient for the behaviour of the signaling network remain to be studied.

The general conclusion from this work is that SmartCell is highly accurate in modeling reaction kinetics and diffusion, but suffers from speed limitations for complex models including processes occurring on different time-scales. The highly specific solutions present in specialised simulation programs are not available in a general framework like SmartCell and further constrain the speed of the simulations. There is also an urgent need of a graphical user interface, including a model editor, to ease the modeling process.

Contents

1 Introduction

1.1 Modeling and Simulation in Biology

The last decades have seen an intensified exploration of biological systems at the molecular level. It is the advent of new techniques that have made possible for biologists to precisely alter the structure and function of biological macromolecules by means of recombinant DNA technologies, to study macromolecules at atomic resolution and to study cells by sophisticated microscopic techniques. Structural biologists and biochemists in collaboration have explained interactions and reaction mechanisms for biological molecules. Biochemists have mapped biochemical pathways and cell biologists have studied cell morphology, dynamics and localisation of specific molecules. Geneticists have sequenced the genome of several key organisms, mapped the topology of many gene regulatory networks and recorded huge amounts of data on expression patterns in microarray experiments. The list is far from complete, but shows the enormous amount of biological information that has accumulated over years of research. However, the availability of this information is limited; it does not give a comprehensive and unified view of how cells work. Information that is stored in databases and articles is in its unprocessed form static and does not allow scientists to understand biology at the system level.

The large data sets represent a reductionist's view on biology – that a system can be divided into subsystems and understood as the sum of its components. It is natural that this view has dominated biological research the past 50 years, when research on a subcellular level has been ongoing. However, now that there is more knowledge about the details of cell biology it is again time to take a step backwards and look at the cell at a higher level in order to get an integrated perspective on cell function. Biology today face a renaissance of holism, a field often referred to as *systems biology*.

To allow integration of qualitative knowledge and quantitative data into a description of a system, a *model* is needed. In general a model is simply a collection of hypotheses and facts brought together with the aim to describe a phenomenon. It is important to note that a model can never be proved to be true, but can only be rejected as an invalid hypothesis upon failure. In fact a model will always fail at some point unless it is the system itself.

A mathematical model can, through simulations, give new insights into the dynamic properties of a biological system under different conditions. What then is a mathematical model of cell function? Both intercellular and intracellular functions can be decomposed into a network of chemical reactions, each of which can be represented by a differential equation. Therefore, a mathematical model of a biological system typically consists of a set of differential equations. Once the kinetic parameters, such as the rate constants and the initial concentrations of all components involved in the reactions, are defined the system of equations can be solved.

The complexity of a model may differ significantly, and in systems biology the models typically get very extensive since they serve to link many phenomena into one model. Because of the complexity of biological systems analytical solutions

usually do not exist, but numerical solutions can be calculated by means of computers. In addition to the lack of data one of the major constraints for the development of systems biology so far has in fact been the availability of computational power.

The most obvious application of modeling and simulation in biology is testing of hypotheses on the underlying mechanisms of cell function. This facilitates understanding how cells work, and will, in a longer time perspective, allow biologists to engineer new or altered properties of living cells. Another possibility with *in silico* studies of cells is to save time and resources. Imagine, for example, that an experimentalist has several alternative models for a specific feature of a cell. Simulations could then help guiding him towards the key experiments to perform, which will save work. Secondly some experiments are impossible to perform, e g because the dynamics is very fast or a particular molecule is hard or impossible to detect, so here simulations could be an alternative to experiments. Moreover, errors in real measurements do not exist in simulations and therefore gives more exact measurements. However, modeling and simulation can never substitute for real experiments, but could be an interesting complement in biological research.

Model building is nothing new in biology. See for example [27] for a review on the subject. However, it is only recently that one has attempted to model whole-cell behaviour taking molecular events into account. The two most successful approaches so far are the Virtual Cell project in the US [22] and the E-CELL project in Japan [33].

1.2 Aim of the Project

The Smartcell project was initiated by Luis Serrano (EMBL Heidelberg, Germany) in the Spring of 2000. Anders Kaplan, a former diploma student of Serrano's group, laid out the basic ideas and structure for SmartCell and also began implementing the program. When I started my diploma work in February 2001 the basic functionality of SmartCell was not fully implemented yet, so the first goal was to get the program up and running. The second aim was to validate SmartCell with respect to *(i)* its capacity as a modeling tool and *(ii)* the predictive power of the simulation algorithm. Naturally, the expectation was also that with time ideas for further development would appear.

2 The SmartCell Framework for Modeling and Simulation of Biological Systems

When the SmartCell project began the main goal was to develop a general software tool for modeling and simulation of complex biological systems. The overall aim is much like the aims of the Virtual Cell and E-Cell projects, but in addition SmartCell was required to support some special features. For reasons explained later SmartCell

- is based on stochastic reaction kinetics
- supports localisation
- supports diffusion
- supports reactions involving all kinds of molecules, including DNA, in the cell
- has a general model description format that resembles the semantics of everyday language and includes concepts familiar to a biologist

2.1 Stochastic Versus Continuous Reaction Kinetics

There are two alternative formalisms for mathematical descriptions of the temporal behaviour of networks of chemical reactions, the macroscopic and the mesoscopic approaches. The macroscopic approach is continuous and deterministic, describing the system in a set of differential equations, whereas the mesoscopic description is stochastic. Here the time evolution consists of a random-walk process governed by a single differential-difference equation, the master equation. Non-surprisingly the deterministic approach is more frequently used since it is more tractable to mathematical analysis, and the theory for numerical integration is well-developed for similar problems in other scientific fields. However, for chemical reaction networks a macroscopic description is not always applicable, but suffers from several weaknesses. The most important problems with using a macroscopic model for reaction kinetics is discussed below.

The physical basis of chemical reactions is collisions between molecules. This implies that the populations of molecular species can only change by integer values and therefore the time evolution of the state of the system is not continuous [17]. At high concentrations of reactants though, the chemical system can successfully be approximated by a continuum of particles. It is a well-known fact however, that some molecules in the cell, e g gene regulatory proteins and promoters of low copy genes, are present in very low amounts, which in the extreme case of chromosomal DNA sometimes implies only a single copy per cell [18]. Therefore, if for example, gene networks including low-dosage genes are part of the modeling targets, as they would in a whole-cell model, a continuous model is questionable and stochastic kinetics should be the obvious choice.

Another important weakness of the deterministic approach is the assumption that there are no fluctuations in the system. Fluctuations play an important role

in biological systems in, for example, switches between alternative regulatory pathways or in the time delay from the onset of expression until a protein reaches a threshold concentration [25]. In mesoscopic kinetics however fluctuations are an inherent property of the kinetics model.

The choice of kinetics model therefore is mainly determined by the modeling target. In the case of metabolic networks, with high concentrations of the components and stable equilibria, the continuous approximation is often useful. If the modeling target is instead a signal transduction pathway with localised components, or a genetic network with low copy numbers of reacting species and complex control mechanisms, such as positive feedback, then a stochastic model is preferable.

A major disadvantage of the mesoscopic model is that the master equation is difficult to solve both analytically and numerically. A useful alternative is however the Gillespie algorithm that circumvents the direct solution of the master equation but instead simulates chemical reactions by a Monte-Carlo approach [17]. The only assumption made in the derivation of Gillespie's algorithm is that the system is thought to be spatially homogeneous meaning that the molecules are randomly distributed in a uniform sense throughout the reaction volume. For this condition to be true it suffices to require that the elastic non-reactive collisions, that serves to randomise the positions of the molecules, occur much more frequently than the reactive collisions. This criteria should hold in any biological system consisting of many different species, with sizes much smaller than the reaction volume, reacting only with a limited number of particle types in processes with reasonably high reaction energies.

The kinetics model implemented in SmartCell is based on the mesoscopic, stochastic reaction model that is derived and verified by Gillespie in [17]. The concept "reaction rate", central to deterministic kinetics, is substituted by a reaction parameter c, or "reaction probability per unit time", in a stochastic model. However, a relation between the rate constant k and the corresponding reaction parameter c exist, which allows the model to be built from values of reaction rates measured in experiments.

In contrast to SmartCell both Virtual Cell and E-Cell are based on macroscopic kinetics represented by differential equations that are integrated numerically.

2.2 Localisation and Diffusion

Localisation, a non-uniform distribution of molecular species in the cell, is a phenomenon that occurs in prokaryotes as well as in eukaryotes. Experiments showing the importance of localisation of structural and functional proteins for cell cycle progression in *E coli* have been published [29] and disapproves the idea that the whole cytoplasm of the bacterial cell is available to all kinds of molecules. It is therefore likely that localisation plays an important role for other cellular processes as well. From localisation follows the need of transport of molecules in the cell. So far no system for active transport has been found in *E coli*, but molecules are believed to move by diffusion. For the larger eukaryotic cells however, diffusion would be too slow, and here the cytoskeleton has evolved

to mediate directed transport.

Taking localisation into account suits a stochastic reaction model for computational reasons. Since many molecules will only be present in parts of the cell this will significantly reduce the number of calculations to be performed. The underlying treatment of geometry in SmartCell makes it possible to treat diffusion as a general process. It should be noted though, that whereas diffusion processes are supported in the SmartCell modeling language, active transport is not and consequently has to be explicitly modeled.

In SmartCell localisation is implemented by assembling the cell geometry from a set of discrete building blocks, termed *sites*. The sites are either three-dimensional representing volume elements of the cytosol, surroundings or space inside organelles, two-dimensional representing membranes, one-dimensional representing DNA, or zero-dimensional representing point-sites on DNA. The resolution of the system is determined by the lattice unit, λ, which is the length of the side of each subvolume. Neighbouring sites are interconnected to permit modeling of diffusion throughout the whole cell geometry and to allow interaction between molecules in neighbouring sites of different dimensionality (eg a cytosolic protein binding to a membrane-bound protein). Sites belonging to a certain group of sites, or representing a biological structure are collected into *regions*. Basic regions defined for the *E coli* geometry are the cytosol, the cell membrane, the chromosome and the surrounding solution.

The *E coli* model is three-dimensional, but at present no visualisation program is available. However, a simple cross-section plotted in Matlab is shown in figure 1. Each square in the figure represents one cytosolic site, and the colour depth indicates the amount of a species present in that site. The plot shows the distribution of a species that is produced by a membrane-bound protein at the right end of the cell and is allowed to diffuse through the whole cytoplasm. A high turnover rate relative to the rate of diffusion produces a gradient throughout the cell.

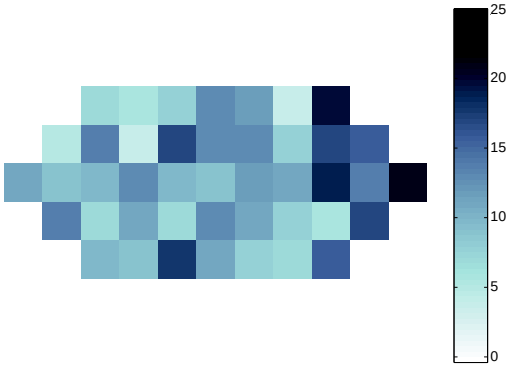


Figure 1: Section of the coli-formed SmartCell. The plot shows a snapshot of the distribution of the response regulator protein, CheY, in bacterial chemotaxis. CheY is produced in the right-hand end of the cell and diffuses throughout the whole cytoplasm.

2.3 The Model Description Format

The model description format of the SmartCell framework was developed by Anders Kaplan in his diploma work [20]. Requirements of generality, flexibility, and easy handling and understanding for a person not accustomed to programming were posed, and the general-purpose markup language XML, was chosen to meet these criteria. The specific format for a SmartCell model is presented in a specific Document Type Definition (DTD) file, defining a set of elements from which a model can be built. They include

Species - essentially describing a particle, a molecule or a larger structural unit.

DNA species - a special case of species that may contain substructures, termed DNA sites, representing genes, regulatory regions etc.

Process - the reaction concept in SmartCell. The process element holds information about reactants, products, reaction rates, catalysing enzymes, cofactors etc.

Process site - a child element of process. Defines the environment and localisation of the process.

Multi-step process - used for reactions of n identical steps with one reactant and one product. Transcription and translation are typically modeled as multi-step processes for reasons of efficiency.

Complex - a structure formed from two or more molecules.

Complex formation - the process of complex formation.

In addition, several elements associated with the simulation such as initial amounts, user-defined parameters and constraints, time-series and snapshots for monitoring species during a simulation, are also defined.

To illustrate model building in SmartCell, assume that we wish to model the binding of a ligand to its receptor according to

$$l + R \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} lR$$

where l is the free ligand present in the surroundings, R is the free receptor located in the cell membrane, $k_{on} = 10\,\mathrm{M^{-1}s^{-1}}$ is the rate constant of the association reaction and $k_{off} = 1\,\mathrm{s^{-1}}$ is the rate constant of the dissociation reaction. In terms of geometry this reaction takes place in a SmartCell environment like the one shown in figure 2.

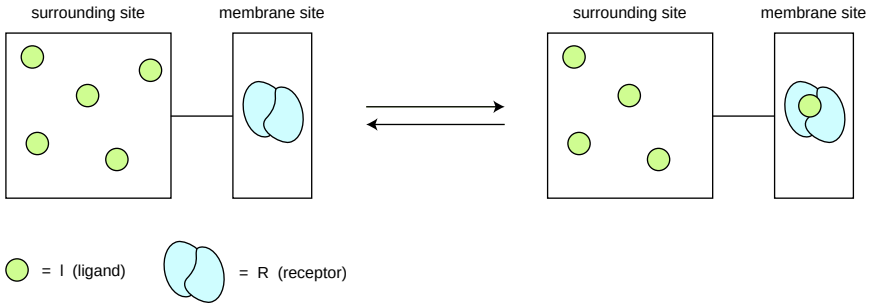


Figure 2: Geometry of the sample reaction. Note that the dimensionality of the surrounding site and the membrane site differs. The horizontal bar connecting the two sites indicates that they are physically connected with no space in between them.

In SmartCell model language the reaction would be written as follows:

```
<!-- Declaration of species -->
<species name="l" />
<species name="R" />

<!-- Definition of binding reaction -->
<process name= " l + R <=> lR " rate="1e6" reverse-rate="1" >
<process-site name="sur" region="surroundings" type="solid" >
<reactant name="l" multiplicity="1" />
<connected-to name="mem" />
</process-site>
<process-site name="mem" region="cell membrane" type="boundary">
<reactant name="R" multiplicity="1" />
<product name="lR" multiplicity="1" />
<connected-to name="sur" />
</process-site>
</process>

<!-- Initial amounts of species, default value is zero -->
<initial-amount entity="l" region="surroundings" type="solid">
1e-6
</initial-amount>

<initial-amount entity="R" region="cell membrane" type="boundary"
absolute-amount="yes">
500
</initial-amount>
```

For a detailed description of the elements and their attributes in a SmartCell model, see the SmartCell DTD-file included in [20]. See also the documentation on SmartCell [3] for a more comprehensive report on model building in SmartCell.

Note that for a more complex model typing this kind of information is tedious. A special editor is therefore required to simplify model building in practice.

2.4 The Simulation Algorithm

The simulation algorithm in SmartCell is based on the "next reaction method" described by Gibson & Bruck in [16], originating in Gillespie's simulation algorithm [17]. Instead of solving the master equation of the system, only *one* trajectory through state-space is traced during a simulation run. This implies that only one of the possible solutions is found and not the complete set of solutions. Therefore several simulations have to be performed to estimate system behaviour.

The simulation algorithm repeatedly picks the next event to occur according to the transition probabilistics of the system. In summary the algorithm can be stated as follows:

1. Set the initial numbers of molecules.
2. Calculate the probability a_i for each event i.
3. For each event i, sample a putative reaction time τ_i from an exponential distribution with parameter a_i, and add to the queue of events.
4. Pick the event with the lowest τ from the event queue.
5. Execute the event, recalculate a_i, generate a new τ_i and add it to the event queue.
6. Check dependencies and update "dirty" τ_i:s in the event queue.
7. If the event queue is not empty, go to step 4.

The term "dirty τ_i:s" refers to the reaction times associated with events whose propensity has changed upon the execution of the current event.

An interesting property of the algorithm is that no finite time-step Δt has to be introduced as an approximation of an infinitesimal dt. Thus stability considerations do not have to be taken into account, as for numerical solvers of differential equations. Instead the algorithm is in some sense self-organising and the average time-step will vary during a simulation depending on which reaction is the fastest at the moment.

However, an obvious problem with the algorithm is its long computation times for large and highly interconnected reaction networks. The phenomenon is further accentuated in systems comprising reactions with great differences in time-scales. However, if the time-scale is within the same order of magnitude for all reactions, the computation time will have an approximately linear dependence on system size, since the average degree of connectivity will usually not increase significantly for larger reaction networks.

3 Implementation of the Model Compiler and the Simulation Engine

The reader is requested to see the documentation on SmartCell for details on program design and implementation [3].

4 Validation of the Reaction Kinetics Model

To validate the implementation of the mesoscopic stochastic reaction model in SmartCell a set of test runs had to be performed, and the results of the simulations be compared with the analytical solutions of the differential equations describing the reaction. The macroscopic deterministic kinetics is, as noted above, an approximation of the system behaviour, but the stochastic and deterministic reaction models should converge for large numbers of molecules.

Which reactions had to be tested? First, it was clear that reactions involving more than two particles are highly improbable, which leaved two basic cases into which all reactions in SmartCell can be decomposed.

$$\begin{aligned} &\text{The unimolecular reaction:} && A \rightarrow B \\ &\text{The bimolecular reaction:} && A + B \rightarrow C \end{aligned}$$

If the SmartCell simulations of these two reactions converge towards the analytical solutions at high concentrations, where relative fluctuations are negligible, the corresponding equilibrium reactions should hold as well. In this work I chose to study the equilibria instead of the unidirectional reactions since they made it possible to assess fluctuations and systematic errors in the steady-state concentrations, once equilibrium was reached.

4.1 The Unimolecular Reaction

The chemical reaction under study is

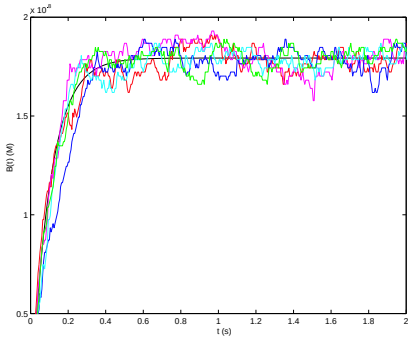
$$A \underset{k_r}{\overset{k_f}{\rightleftharpoons}} B \tag{1}$$

For this reaction the rate equation for species B and its solution is,

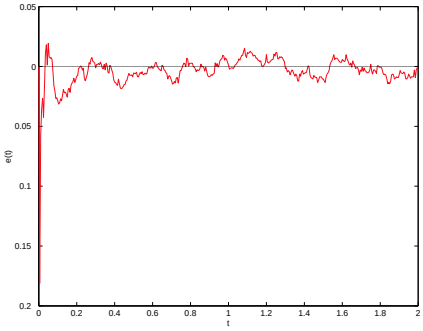
$$\begin{aligned} &\frac{db}{dt} = k_f \cdot a - k_r \cdot b \\ &b(t) = b_0 e^{-(k_f+k_r)t} + b_{eq}(1 - e^{-(k_f+k_r)t}) \\ &b_{eq} = \frac{k_f(a_0 + b_0)}{k_f + k_r} \end{aligned} \tag{2}$$

where a is the concentration of A molecules, b is the concentration of B molecules, a_0 and b_0 are the starting concentrations, b_{eq} is the concentration of B molecules at equilibrium, k_f is the rate constant for the forward reaction and k_r the rate constant for the reverse reaction.

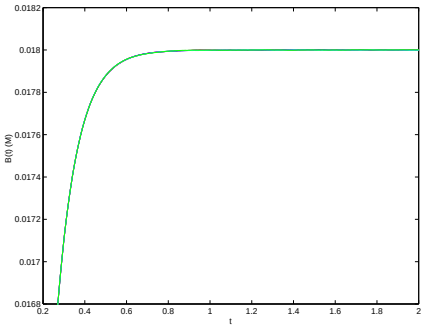
Simulations were run for two different concentration levels of the reactants, one high and one low, but with the same values for the rate constants (for exact values, see the model descriptions in section A.1.1). During the 2 second simulations the levels of species A and B were monitored, with the time evolutions $b(t)$ of B plotted in figure 3.



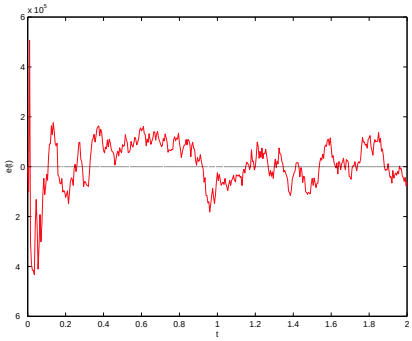
(a) Five simulations of low concentration (a_0=20 nM). The black curve shows the solution of the macroscopic kinetics model.



(b) Relative deviation from the analytical solution of the macroscopic kinetics model. Average of 20 simulations with a_0=20 nM.



(c) Five simulations of high concentrations (a_0=20 mM). Each individual curve can barely be distinguished at this magnification.



(d) Relative deviation from analytical solution for high concentration of the macroscopic kinetics model. Average of 20 simulations with a_0=20 mM. Note the difference in scale of the y-axis compared to figure 3(b).

Figure 3: Kinetics for the unimolecular reaction.

The deviation from the analytical solution for the rate equation was measured as the relative error locally (for $t < 0.1\ s$ and $t > 1.0\ s$) and as an over-all mean value for the whole simulation. The values were calculated as mean values for 20 simulations and are presented in table 1. As a comparison, the corresponding values for the five first simulations of the lower concentration are listed in table 2. As can be expected due to the lower number of product particles present, the relative error at the beginning of the simulation is more than one order of magnitude greater than the error in the equilibrium concentration. Also the values for the higher concentration more tightly follow the analytical curve. In the same way comparing the deviation for a single simulation and with that of the mean values over 20 simulations, the mean value has a smaller deviation. Calculating the average is effectively the same as raising the concentration 20 times, but serves here as a means of showing the reproducibility of the experiment.

Table 1: Relative errors compared to the analytical solution of the rate equation. The values are mean values for 20 simulations. For the absolute numbers of particles in the model, see section A.1.1.

Conc (M)	*mean ($t = 0..2\ s$))*	*start (t <0.1s)*	*equilibrium (t>1.0s)*
10^{-8}	$7.58 \cdot 10^{-3}$	$2.81 \cdot 10^{-2}$	$6.00 \cdot 10^{-3}$
10^{-2}	$7.55 \cdot 10^{-6}$	$2.50 \cdot 10^{-5}$	$4.70 \cdot 10^{-6}$

Table 2: Relative errors compared to the analytical solution of the rate equation for the first five simulations performed for low concentration of reactant ($a_0 = 10^{-8}\mu$M).

run no	*mean ($t = 0..2\ s$)*	*start (t <0.1s)*	*equilibrium (t>1.0s)*
1	$3.50 \cdot 10^{-2}$	$1.46 \cdot 10^{-1}$	$2.16 \cdot 10^{-3}$
2	$3.76 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$	$2.83 \cdot 10^{-3}$
3	$4.61 \cdot 10^{-2}$	$1.83 \cdot 10^{-1}$	$2.53 \cdot 10^{-3}$
4	$3.08 \cdot 10^{-2}$	$1.67 \cdot 10^{-1}$	$1.67 \cdot 10^{-3}$
5	$4.61 \cdot 10^{-2}$	$9.31 \cdot 10^{-2}$	$2.53 \cdot 10^{-3}$

Note in table 2 that the relative deviation for the lower concentration is not negligible in the beginning of the simulation. This clearly demonstrates that when a reaction network has species present in low amounts the stochastic effect becomes important. For higher concentrations however, the stochastic kinetics simulation algorithm converges, as expected, towards the kinetics described by the continuous rate equation.

4.2 The Bimolecular Reaction

The bimolecular reaction studied is

$$A + B \underset{k_r}{\overset{k_f}{\rightleftharpoons}} C \tag{3}$$

and the rate equation for the reaction can be written

$$\frac{dc}{dt} = k_f ab - k_r c \tag{4}$$

This differential equation also has an exact analytical solution. It is easily derived by studying the deviation from equilibrium instead of the absolute concentrations [13]. Here though, the solution is expressed in the variable c, the concentration of C molecules.

$$c(t) = c_{eq} - \frac{(c_{eq} - c_0)(a_{eq} + b_{eq} + K_D)e^{-kt}}{a_{eq} + b_{eq} + K + (c_{eq} - c_0)(1 - e^{-kt})} \tag{5}$$

in which

$$K = \frac{1}{K_D} = \frac{k_f}{k_r}$$

$$k = k_f(a_{eq} + b_{eq}) + k_r$$

$$c_{eq} = \frac{1}{2}\left[a_0 + b_0 + K \pm \sqrt{(a_0 + b_0 + K)^2 - 4a_0 b_0}\right]$$

$$a_{eq} = a_0 - c_{eq} \text{ and } \mathrm{b_{eq}} = \mathrm{b_0} - \mathrm{c_{eq}}$$

where K and K_D are the association and dissociation constants respectively.

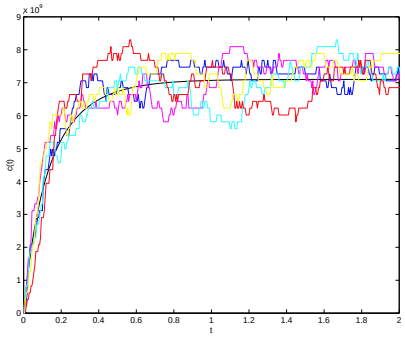
Simulations were run for three different concentrations. For the lowest concentration the rate constants were chosen differently than for the two other experiments to make the system reach equilibrium within 2 seconds. See section A.1.2 for details on model parameters.

Results from the simulation at high concentrations of reactants can be seen in figure 4. The theoretical curve falls well between the simulated curves, which is confirmed by the relative errors that are small (in the order 10^{-5}) as seen in table 3. However, for the low concentrations the deviation from the analytical solution is significant, which is shown in figure 4(a-b) and in table 4. Note the different scales of the y-axis in figure 4(b) and figure 4(d).

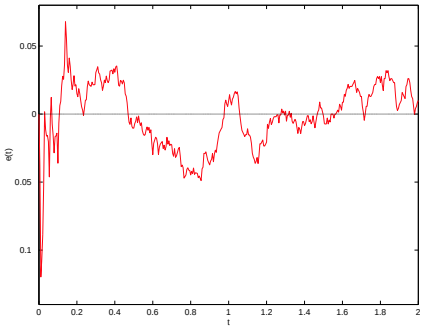
As for the unimolecular reaction the relative deviations from the analytical solutions were calculated for start and equilibrium values, and as an overall mean value. The result is presented in table 3 and 4.

Table 3: Relative errors compared to the analytical solution of the rate equation. Mean values for 20 simulations.

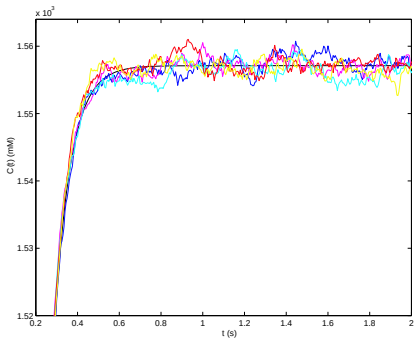
Conc (M)	*mean (t = 0..2 s)*	*start (t <0.1s)*	*equilibrium (t>1.0s)*
10^{-8}	$1.85 \cdot 10^{-2}$	$4.71 \cdot 10^{-2}$	$1.22 \cdot 10^{-2}$
10^{-3}	$1.67 \cdot 10^{-3}$	$2.76 \cdot 10^{-3}$	$1.83 \cdot 10^{-3}$
10^{-2}	$1.82 \cdot 10^{-4}$	$3.73 \cdot 10^{-4}$	$1.55 \cdot 10^{-4}$



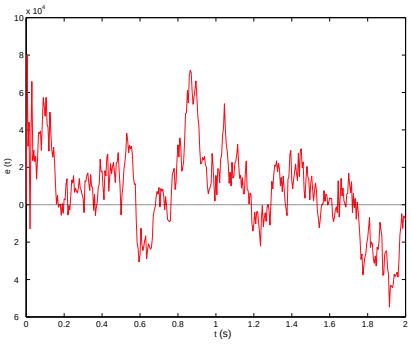
(a) Five simulations of low concentration (a_0=12 nM, b_0=10 nM). The black curve shows the solution of the macroscopic kinetics model.



(b) Relative deviation from the analytical solution of the macroscopic kinetics model. Average of 20 simulations with a_0=12 nM and b_0=10 nM.



(c) Five simulations of high concentrations (a_0=20 mM, b_0=10 mM). The black curve shows the solution of the macroscopic kinetics model.



(d) Relative deviation from analytical solution for high concentration of the macroscopic kinetics model. Average of 20 simulations with a_0=20 mM and b_0=10 mM. Note the difference in scale of the y-axis compared to figure 4(b).

Figure 4: Kinetics for the bimolecular reaction.

Table 4: Relative errors compared to the analytical solution of the rate equation for the first five simulations performed for low concentrations of reactants.

run no	*mean* ($t = 0..2$ *s*)	*start* ($t < 0.1s$)	*equilibrium* ($t > 1.0s$)
1	$5.74 \cdot 10^{-2}$	$3.44 \cdot 10^{-1}$	$3.84 \cdot 10^{-2}$
2	$7.85 \cdot 10^{-2}$	$2.17 \cdot 10^{-1}$	$6.28 \cdot 10^{-2}$
3	$7.02 \cdot 10^{-2}$	$1.12 \cdot 10^{-1}$	$7.89 \cdot 10^{-2}$
4	$1.10 \cdot 10^{-1}$	$5.55 \cdot 10^{-1}$	$7.34 \cdot 10^{-2}$
5	$7.50 \cdot 10^{-2}$	$3.22 \cdot 10^{-2}$	$5.72 \cdot 10^{-2}$

Again the deviation is significant at low concentrations of reactants, but the results for the higher concentrations indicate that the implementation of Gillespie's algorithm in SmartCell is correct.

5 Validation of the Diffusion Model

As mentioned in the introduction one of the desired features was to include diffusion in the SmartCell framework. Diffusion processes in SmartCell are treated as general processes with a rate constant determined from the diffusion constant (D) and the lattice unit (λ) of the geometry. To evaluate the accuracy of the diffusion model the simple test geometry visualised in figure 5 was designed.

There are two alternative ways of measuring the rate of diffusion. The most common approach is to start from a Dirac peak and then monitor the broadening of the peak over time. For implementation reasons a second alternative was chosen, which is to measure the average time for a single particle travelling a fixed distance.

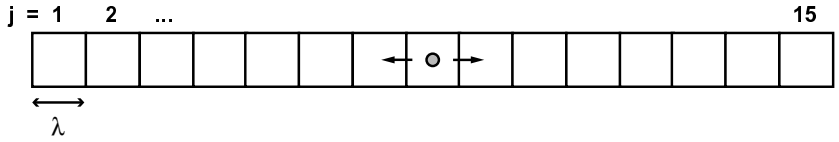


Figure 5: Diffusion model. The time required for the single particle to reach any of the boundary volumes is measured repeatedly.

The distribution of a diffusing particle is described by a normal distribution (centered around origo),

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}$$
$$\text{where } \sigma^2 = 2 \cdot \mathrm{D} \cdot \mathrm{t}$$

From the definition of the standard deviation (σ) it follows that in the case of diffusion this will be the mean traveled distance, x, during a time t. Hence, the expected time for travelling a distance x can be expressed in terms of the diffusion constant and x

$$\langle t \rangle = \frac{x^2}{2D} \tag{6}$$

For the test system in figure 5 $x = 7 \cdot \lambda = 0.7\mu m$ and the diffusion constant was specified to be $10\mu m^2 s^{-1}$ (section A.2), which gives

$$\langle t_{theor} \rangle = 0.0245\ s$$

The experimental estimate of the average time was calculated from t measured in 10^6 independent simulations

$$\langle t_{exp} \rangle = 0.0244966\ s^{-1} \approx 0.0245\ s^{-1}$$

This result indicates that the treatment of diffusion processes in SmartCell successfully predicts the behaviour described theoretically.

6 Introduction to the Bacterial Chemotaxis Signaling Network

The bacterial sensory system for chemotactic response in *E coli* is one of the most thoroughly studied signaling networks in biology. *E coli* offers good means for studying biochemistry and genetics, a fact which has resulted in a wealth of information on, amongst others, the chemotaxis signal transduction pathway; the major biochemical steps in the pathway have been identified, and most of its molecular components have been characterised with respect to both structure and function. The 30 years of research conducted on chemotaxis in *E coli* in fact have provided kinetic data for most participating reactions, which makes the pathway uniquely tractable to mathematical modeling. Therefore several attempts of modeling the chemotaxis signal transduction pathway have been made. These models serve as suitable references, when validating the predictive power of SmartCell's simulation engine as well as when assessing the usability of the model description format in the SmartCell framework.

6.1 The Chemotactic Response in Bacteria

Bacteria swim by means of 5 to 10 flagella distributed over the cell surface. A flagellum has a helical structure and is attached to a motor complex buried in the cell membrane via a flexible hook [1]. The molecular motor makes use of energy stored in the transmembrane proton gradient to rotate the flagella in either clockwise (CW) or counter-clockwise (CCW) direction. Upon counter-clockwise rotation the flagella form a bundle behind the cell which push the cell forward in a *smooth swimming* mode at a rate of approximately $20\mu\text{ms}^{-1}$. In contrast, when rotating clockwise the flagella fail to form a bundle, but induce a random *tumbling* behaviour which reorients the cell before next run. The durations of both runs and tumbles are exponentially distributed with means of 0.8 and 0.2 s respectively, in the absence of any environmental stimulus [26].

The search strategy of the bacterium is a *biased random walk*. By changing the proportions of time spent tumbling or swimming the cell can adjust its behaviour to either swim towards attractants (or away from repellents), once they are spotted, or keep on scenting for nutrients and toxic compounds. Hence attractants suppress tumbling whereas repellents promote tumbling. It has

been reported that the swimming behaviour mainly changes by extending or shortening smooth runs, whereas the average time for tumbling remains the same, a fact which could be justified by energetic considerations.

The system is however more intricate than described above, because such a simple system runs the risk of being trapped within an area of relatively low concentrations of attractant. To avoid this dysfunction the system quickly adapts by returning to basal levels of intracellular chemotaxis messengers. From experiments it is known that bacteria maintain a high sensitivity and show perfect adaptation at attractant concentrations between 2 nM and 200 mM – an amazing range of five orders of magnitude. The adaptive power arises from a memory function of the chemotaxis system, where the cell samples the present composition of the environment and compares it with that some 3-5 seconds earlier. This temporal sensing is due to the difference in time scales for the excitation reactions and the adaptation reactions. It is important to note that in this way the cell effectively responds to *changes* in attractant concentrations rather than absolute concentrations.

6.2 The Intracellular Signaling Network

The intracellular signaling pathway for chemotaxis in bacteria is an example of a common signaling theme in prokaryotes; the two-component system. All two-component systems possess a cell-surface sensory protein and an intracellular response regulator, whose activity is controlled by phosphorylation. The following discussion will be restricted to describing the pathway in *E coli* on a molecular level.

6.2.1 The Chemotaxis Signaling Network in E coli

The chemotaxis signaling pathway in *E coli* is relatively isolated from other signaling networks in the cell and consists of a network of only seven interacting proteins. They are the receptors and the proteins of the Che-family; CheA, CheW, CheY, CheB, CheR and CheZ. A schematic view of the molecular components and reactions contained within the chemotaxis signaling pathway is shown in figure 6.

The sensory function is carried out by a family of methyl-accepting chemoreceptor proteins, MCPs. In *E coli* there are five different MCPs, each responsible for mediating signals from a different set of environmental stimuli. For example Tar, which is the most abundant MCP, recognises and responds to aspartate and other amino acids, maltose via a maltose binding protein, extracellular pH, ambient temperature, as well as the repellents nickel and cobolt ions.

In vivo MCPs dimerise and on the cytoplasmic side of the plasma membrane they associate with two molecules of CheW and CheA respectively to form a stable ternary complex. The CheA protein has histidine kinase activity, and in the presence of ATP the dimer undergoes trans-autophosphorylation. CheW acts as a relay mediating conformational changes in the MCP upon ligand binding and

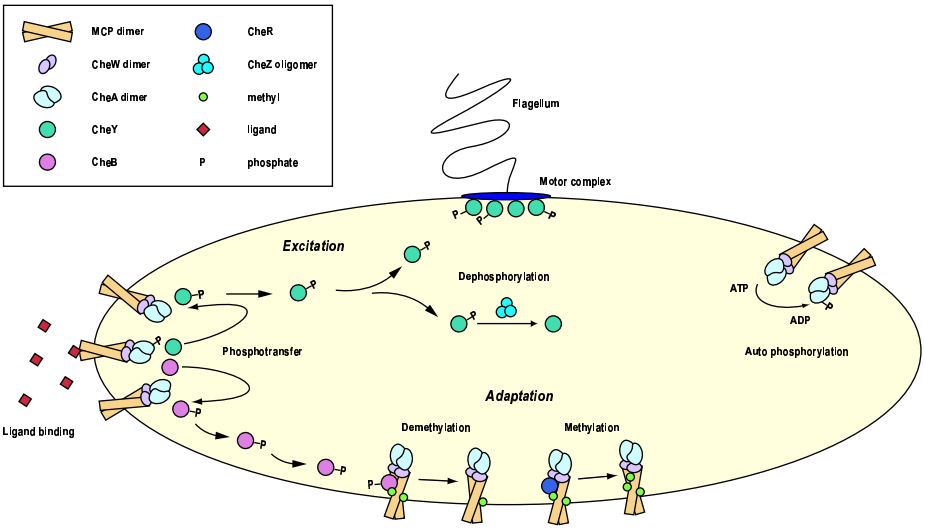


Figure 6: The chemotaxis signaling network in *E coli*.

methylation, which influences the kinase activity of CheA. The receptor complexes cluster at one of the cell poles [24], but there are still only speculations on why this localisation appears. One possible explanation is that it enables cooperative signaling, which could also partially resolve the so far unanswered question of the high signal amplification in the pathway [11].

The response regulator in bacterial chemotaxis, CheY, is a small diffusible protein which can exist in either an unphosphorylated or a phosphorylated form (CheYp). CheY undergoes spontaneous phosphorylation from the substrate PoAc, but the level of phosphorylation increases dramatically in the presence of phosphorylated CheA. Hence CheAp serves as phosphoryl-donor for CheY and the proportion of CheY and CheYp is controlled by the autokinase activity of CheA. Since the receptors are mainly located in one pole CheYp diffuses through the cytoplasm to reach its points of action, the flagellar motors (see section 6.2.3).

When an attractor molecule binds to the receptor the kinase activity of CheA drops and thereby the CheYp concentration decreases in the cell. Since attractants suppress tumbling this means that low CheYp concentrations suppress CW rotation of the flagella whereas the opposite is true for repellents.

CheY also has an intrinsic dephosphorylation activity, but the dephosphorylation is mainly controlled by a cytoplasmic phosphatase, CheZ. The activity of CheZ is regulated through CheYp controlled oligomerisation that is required for maximal activity [8]. Some evidence for CheZ CheA interaction also exist, which could be another source of regulation of the phosphatase activity [34].

6.2.2 Adaptation

After a prolonged exposure to attractant or repellent the cell returns to its pre-stimulus state, something which is due to changes in the methylation level of the receptors making them less sensitive to the stimulus. The two proteins responsible for adaptation are the methylating protein, CheR, and the demethylating protein, CheB. CheR and CheBp bind competitively to a pentapeptide "arm" on the Tar receptor that swings the adaptor enzymes close to the two glutamates that serves as sites of methylation [6].

Given a stable metabolic state with a steady supply of methyl groups in the form of S-adenosylmethionate, the activity of CheR is constant. The overall rate of methylation however is affected by the conformational state of the receptor complex, since the active phosphorylated receptor is methylated at a slower rate than the inactive unphosphorylated form [32]. The activity of CheB on the other hand is regulated by phosphorylation of its N-terminal domain. Through phosphorylation its demethylating activity increases by orders of magnitude. Just as CheY, CheB can be phosphorylated by CheA and it has been shown that CheY and CheB exhibit competitive binding to the kinase CheA [21].

From the description of the reaction network two components in the adaptation response emerges, one local permanent effect and one global transient effect. When a cell is exposed to an attractor, a higher fraction of the receptors become available to CheR-mediated methylation since the attractant causes a drop in receptor auto-phosphorylation. The result is an increase in the methylation level of the receptors, and is called a local adaptation since it works only on the occupied receptors. The second effect of ligand binding is a drop in CheBp concentration due to a decreased activity of CheA. Since CheBp is freely diffusible in the cytoplasm this will affect the whole cell, and is then part of the global adaptive response. However, the decrease in CheBp concentration is only transient and the global effect will soon disappear, so that it is only the change in methylation-state of the receptors that is preserved in a fully adapted cell.

Figure 12 gives a more qualitative understanding of the reactions responsible for adaptation. In a non-induced cell, the receptors are distributed over the states in the front plane of the grid with an average of two methyl groups per receptor dimer. When an attractor is introduced the distribution of receptors will be pushed backwards to the rear plane of the grid. Since the interaction between ligand and receptor is much faster than the signaling and adaptation reactions, the distribution over different methylation states will initially not change significantly but the reaction network now occupies a "sequestered" state. If the attractant concentration remains high the activity of the receptors will drop quickly, which with time will increase CheR's methylation activity and decrease the demethylating activity of CheBp. The reaction network has then found its new equilibrium corresponding to a shift to the left in figure 12.

6.2.3 The Flagellar Motor

Further down the chemotaxis pathway is the motor complex, with which CheYp interacts. The flagellar motor consists of Mot and Fli proteins arranged in

several layers of rings. CheYp interacts with the protein FliM and the number of CheYp molecules bound controls the direction of rotation of the flagellum.

The mechanism by which CheYp interacts with the motor complex, and thereby changes the direction of rotation of the flagella, has shown to be somewhat elusive. Two models explaining the switching of direction exist, but no consensus on one of them has been reached among scientists. In the first model, known as the FliM threshold-crossing model, the main idea is that at a certain number of CheYp molecules bound the motor switches from CCW to CW rotation. The second model, the thermal isomerisation method, instead focuses on thermal fluctuations in the motor complex where a certain level of CheYp determines the probability of the motor being in either of the two rotational states.

6.2.4 Measurements of Chemotaxis Output

The swimming behaviour, i e the output, of the signaling network can be described on a higher level in the Hill equation, which resembles the rate equation of allosteric enzymes with several cooperative binding sites. In chemotaxis the Hill equation describes the proportion of time spent swimming or rotating CCW, the swimming or rotational bias respectively.

$$bias = \frac{k}{k + [CheYp]^h} \quad k, h \, \epsilon \, \Re \tag{7}$$

where h is the Hill coefficient and k a constant determined from the bias in a non-stimulated cell.

Under steady-state conditions the wild-type bias is known to be around 0.7. The Hill coefficient for swimming bias, h, has been experimentally determined several times, with the most recent measurement giving a value of 2.5 [2]. Because of the difficulties of experimentally monitoring the swimming of single cells, the rotational bias has been introduced as a measure of bias. Instead of studying free-swimming bacteria the rotational bias is measured in so called *tethering experiments*, in which the cell is stripped of all but one flagellum and tethered to a surface by means of antibodies. This way the cell rotates rather than the flagella, and by filming a cell the time spent rotating CW or CCW can be measured. In the case of rotational bias the Hill coefficient is slightly higher, approximately 3.5, a difference that depends on the intricate way that flagella interact to control the swimming of the cell [2]. The situation is however further complicated by the phenomenon of pausing, associated with switching events for tethered cells.

The results in [2] are interesting from a simulation point of view, since they for the first time present a connection between concentration of CheYp and swimming behaviour. Earlier studies have presented a connection between attractant concentration and swimming behaviour, which comprise an extra layer of complexity in the relation.

Another parameter often encountered when reading about the chemotaxis signaling network is the *gain*. Gain is a measure of the amplification in the signaling pathway and is defined as

$$gain = \frac{\Delta\, bias}{\Delta\, fractional\ receptor\ occupancy} \tag{8}$$

Note however that this measure is ambiguous unless the concentration of inducer in the experiment is given.

6.2.5 The Chemotaxis Network in a Systems Biology Context

What is the relevance then of studying this specific reaction network in such detail and trying to model it in a computer?

What biologists have learnt is that living systems make use of similar strategies for solving the same category of problems but in different environments. This observation is supported by the way evolution works on a molecular level; functional units are duplicated, slightly modified, combined with other functional modules etc, and therefore different biological systems share many features. When it comes to chemotaxis this specific reaction network comprises general concepts in biology such as a signal transduction pathway, excitation, amplification, robustness and integration of multiple input signals on whole pathway level and for single protein complexes. Moreover many of the proteins have functional analogues in eukaryotes, so the understanding of this specific network can possibly be extended to be valid for parts of eukaryotic cell function. Hopefully studies of simple, well-defined and well-characterised networks will provide knowledge from which general patterns, or "regulatory motifs", can be observed and subsequently be applied to more complex reaction networks.

6.3 Modeling Bacterial Chemotaxis: State of the Art

Several attempts to model chemotaxis in *E coli* have already been done, the first published already in 1993. Below is a short description of four different approaches presented in chronological order.

6.3.1 Simulation of the Phosphorylation Cascade

Dennis Bray and Robert Bourret developed the first computer simulation algorithm of bacterial chemotaxis in 1992 [10]. They described the reaction network in a simplistic model consisting of only ten differential equations that were solved by numerical integration. The reason behind this stripped model was to avoid an explosion in computation time, but the trade-off was a less accurate model. For example the phosphorylation reactions were described in one step instead of association, phosphotransfer and dissociation, thereby disregarding competition between CheY and CheB. More importantly the methylation reactions were missing altogether, so the model was non-adaptive. Nevertheless, the program has been useful for predicting the behaviour of several mutants and studying the effects of altered expression levels of the component proteins of the pathway.

Bray and Bourret also discovered that the gain, a measure of the signal amplification, in the chemotaxis network could not be accurately modeled. The gain

in Bray et al's model is one order of magnitude smaller than what has been experimentally observed.

6.3.2 The First Adaptive Models

In 1997 Spiro et al presented a model of chemotaxis with adaptive power over a range of stimuli concentrations [30]. Like previous models it relied on differential equations that were numerically solved. The model accurately predicted the response to different kinds of steps and gradients in chemoattractant concentration, but once again the weakness lied in the prediction of the gain in the signal transduction pathway.

One problem with the model is that the authors used a fine-tuning strategy, changing parameters by trial and error to get a model that adapted over a range of ligand concentrations. The derived parameters however differ from the experimentally measured reaction rates and concentrations by sometimes more than an order of magnitude. A second shortcoming of Spiro's model is that it only takes the Tar receptor into account and omits association and dissociation reactions between Tar, CheW and CheA. Moreover only two methylation sites of the receptors are modeled, thereby overlooking some of the possible distribution of receptor complexes over several states.

It should be noted that already in 1995 Hauri and Ross presented an adaptive model that much resembles that of Spiro et al [19]. Compared to Spiro's model it suffered from even greater weaknesses in the accuracy of model parameters and it only adapted over a narrow range of ligand concentrations. Thus the authors failed to build a general model for adaptation, but showed that there exists a set of parameters for which the model adapts.

6.3.3 A Two-state Model of the Receptor Complexes

In 1997 Barkai and Leibler proposed a general mechanism for robust adaptation in biochemical networks, which they implemented and tested for bacterial chemotaxis [5]. The main assumption, a two-state model where the receptor complex is either active or inactive, dates back to 1984 ([4]) but is here fully developed.

The most striking feature of Barkai and Leibler's model is that it performs perfect adaptation also when model parameters are perturbed significantly. This does not imply that other features of the network stay unchanged, and the overall behaviour is therefore not robust. A good example is the time required for adaptation that varies with some model parameters.

The authors also show mathematically that the adaptive property of the network is a consequence of its connectivity and therefore does not need the fine-tuning of parameters present in earlier models. Mathematically, the demand on an adaptive system is that the steady-state value of the output does not depend on the input. Here the output is the system activity (the activity of the receptor) and the input is the ligand concentration. To meet this requirement some constraints, presented below, are imposed on the model.

Yi et al further analysed the model of Barkai and Leibler in the perspective of control theory. Interestingly they found that the model contains an integral feedback loop, which is the standard solution to designing an adaptive system in classical engineering [36]. Their analysis however posed some additional constraints on the parameters of the model, also presented in the list below.

1. CheB demethylates only active receptors.
2. The concentration of bound CheR must not depend on the ligand level.
3. The rate constants for methylation and demethylation are relatively independent of the methylation state and ligand occupancy of the receptor.
4. The Michaelis constant for CheB must be independent of the methylation state and ligand occupancy of the receptor.
5. The activity of the unmethylated receptor must be negligible relative to the methylated forms.

Comparing Spiro's model with the criteria above, one finds that it fails on criteria 3 and 4, and of course on 1 because it is not a two-state model.

The model of Barkai and Leibler is a reduced representation of the chemotaxis network omitting several processes. Among them should be noted; phosphorylation of CheY and its interaction with CheA and the motor complex, other receptor types than Tar and the formation of the receptor ternary complex with CheW and CheA, and a non-diffusible CheBp disregarding the global adaptation of the network. Also, some of the rate constants seems to be invented and not anchored in experimental data.

6.3.4 Stochastic Simulation of the Full Chemotaxis Signaling Network

In his PhD thesis Morton-Firth implemented a model including the ideas of Barkai and Leibler [26]. As in SmartCell, he used mesoscopic stochastic reaction kinetics for simulating the processes, but Morton-Firth's program, StochSim, was developed specifically for simulation of the chemotaxis reaction network.

The model of chemotaxis is extensive and incorporates reactions such as the formation of the receptor complexes and CheYp interactions with the motor complex. However, in such a huge reaction network experimental data is by mere necessity missing and some reactions are modeled in a compact and simplistic way. An important feature missing is the localisation of components to different parts of the cell, and the phosphatase activity for CheZ is treated in a semi-first order reaction mechanism. Competition between CheR and CheB for binding to the receptor is also missing since CheR is assumed to only bind the inactive form of the receptor.

7 Modeling the Chemotaxis Signaling Network in SmartCell

Looking at the complexity of the model built in Morton-Firth's thesis ([26]), it was obvious that the experiments had to take off in a stripped-down model. In this section the work going from a simple reaction network modeling the phosphorylation cascade to a more complex adaptive model will be presented. During the modeling process three major problems appeared that will be described in detail.

Several abbreviations for the molecular components of the receptor signaling network will be used in the following sections. Please see section A.3.1 for a list of notations.

7.1 An Initial Model of the Phosphorylation Cascade

To simplify the modeling problem the first attempt was to model the steady-state behaviour of the system, and therefore the reactions involved in adaptation were initially omitted. However, phosphorylation of CheB was included in order not to lose the competing behaviour with CheY for access to CheA. The model was built with Morton-Firth's (M-F) model in mind, using reactions and parameters from literature listed in [26]. The reactions and rate constants used in the first model are listed in table 8.

One conceptual change compared to M-F's model was introduced. M-F uses a simplified model of the interaction of CheY and CheB with CheA in which the phosphotransfer is included in the association-dissociation reactions. Whereas M-F assumes that the association of CheY and CheAp is diffusion limited, he uses a higher k_{on} ($5 \cdot 10^{-6}\mathrm{M}^{-1}\mathrm{s}^{-1}$, in contrast to a value of $10^{-6}\mathrm{M}^{-1}\mathrm{s}^{-1}$ that is usually considered the limit of diffusion) motivated by the fast phosphotransfer reaction and that the binding is the rate limiting step in the reaction. This treatment of the reaction seems somewhat obscure and erroneous, since the reasons given are false and the assumption can not be motivated with eg Michaelis-Menten kinetics (considering the reactions rates relative to each other). Secondly the association reaction is unidirectional in M-F's model but is modeled as an equilibrium in the SmartCell model.

For simplicity only the Tar receptor has been considered. This implies that the model disregards all kinds of cross-talk between different receptors. The association and dissociation processes of Tar, CheW and CheA has not been modeled either, but the complex is assumed to exist only in the dimeric form.

Simulations to reach a steady state were run with initial concentrations according to table 10. The eight flagella of the cell were distributed manually on the cell surface and not randomly to avoid clustering of these huge structures. This also confers stability and a higher level of reproducibility to the model. In contrast to the motor complexes the 2500 receptor complexes were clustered in the pole, leaving only one fifth of the receptor molecules in the rest of the cell membrane.

In a non-stimulated cell the CheYp concentration is known to be between 2.5 and 3 μM [12], whereas the results of the first steady-state simulations show a value about three times lower (0.76 μM). The explanation could be either a too high phosphatase activity or a too low oinase activity. Compared to the steady-state level that M-F achieves the difference is probably due to a lower rate of phosphorylation, represented by the higher value of k_{on} in M-F's model as discussed above.

The easiest and most isolated component to remodel is the phosphatase, CheZ. A first calibration showed that a CheZ concentration of about 4 μM, instead of 14.15 μM as is found in the literature, was needed to achieve a more accurate concentration of CheYp.

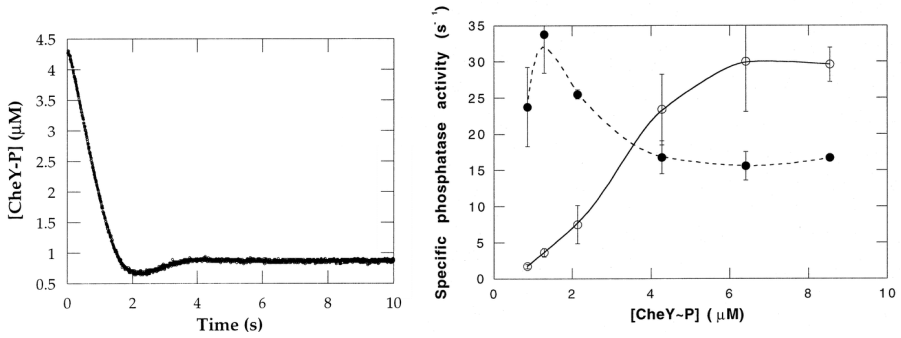
These initial test runs also demonstrated the problem with modeling the interaction between CheY and the motor complex. Even at a constant concentration of CheYp of 2.5 μM the current model failed to predict the state of the flagellar motors in terms of the average number of CheYp molecules bound. The error was small, but a"tuning" process for all the rate constants would have been difficult and time-consuming since the system consists of many mutually dependent reactions. In addition little is known about the mechanism by which they interact and therefore very sparse data is available. Most importantly though taking the state of the motor complex into account, introduces a second level of complexity into the model. It is not only the interaction between CheY and FliM that is not well-characterised, but also the interaction between different flagella. How would the states of the eight flagella be integrated to a swimming behaviour? And in the case of a single flagellum, which switching model should be applied, and with what parameters? For these reasons this part of the model was postponed until later, and instead it was decided to measure the level of CheYp alone and deduce the behaviour of the cell from that[1].

7.2 Modeling the Control of CheZ Activity

A more complex model of the kinetics of CheZ has been reported, most recently in [8]. These studies show that CheZ is present in dimeric form and is able to oligomerise. The oligomerisation process is interestingly facilitated by the presence of CheYp, and the phosphatase activity of the oligomeric form is several orders of magnitude greater than the dimeric CheZ. This leads to a characteristic delay in the onset of the phosphatase activity upon mixture with its substrate, and also an over-shoot before equilibrium is reached (figure 7(a)). This could be a means of regulating the activity of CheZ and hence avoiding a constitutively active CheZ that suppresses variations in CheYp at low concentrations (figure 7(b)). For comparison, activity measurements of a mutant, constitutively active, CheZ are also shown in figure 7(b). It is interesting to note that in the latter case activity reaches a maximum and then drops for higher concentrations.

Data in [8] provided information about the stoichiometry of the complexes and kinetic data for modeling of the reaction network. The stoichiometry of the

[1] Using results from [2] relating CheYp concentration to swimming behaviour. An interesting analysis of swimming behaviour and the state of the motor complex is however found in [26].



(a) CheZ kinetics showing the characteristic 50 ms initial delay and overshoot before equilibrium.

(b) Activity measurements for oligomerising CheZ (continuous line) and mutant constitutively active CheZ (dashed line).

Figure 7: Kinetics of dephosphorylation of CheYp by CheZ. Figures from [8].

oligomer could be deduced to be 6:3, meaning that each CheZ dimer binds one molecule of CheYp and that each oligomer consists of three CheZ dimers. The dissociation constants for the CheY-CheZ interaction (which is not oligomerising) and for the overall process of CheYp-CheZ interaction were also known. To meet the constraints posed by the kinetic data a process of curve fitting was initiated.

In order to model the delay in the regulatory effect, each of the steps in the oligomerisation had to be taken into account (otherwise the simple one-to-one model could be equally effective with two reactions; one slow for the dimeric CheZ and one fast for the oligomeric CheZ). For simple recycling of CheZ after dephosphorylation and a minimum number of reactions the following reaction scheme was considered:

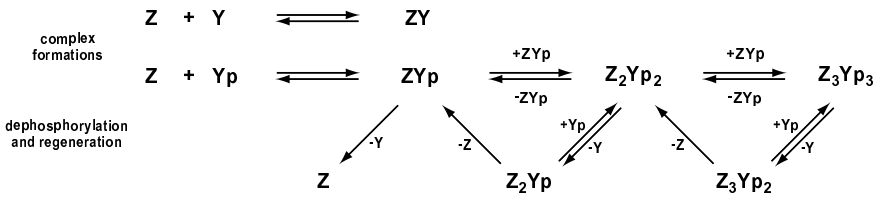
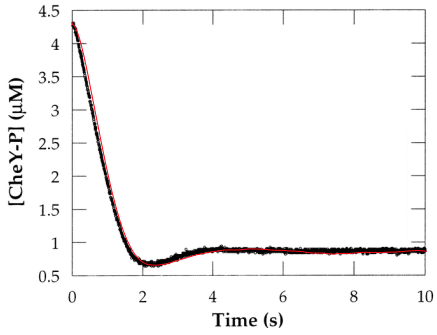


Figure 8: Reaction scheme for CheZ oligomerisation and phosphatase activity.

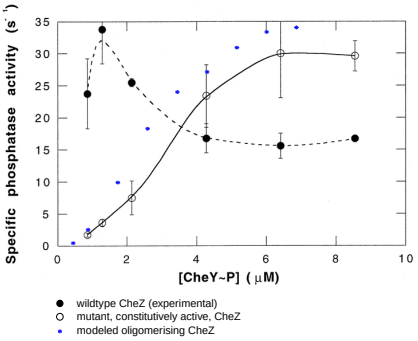
The rate constants for each of these reactions were given an initial value assuming diffusion limited reaction rates in the case of bimolecular reactions, and an estimate from kinetic data for the others. After testing a diffusion limit for CheYp and CheZ interaction the rates of association were at least one order of magnitude too low to produce the 50 ms delay shown in figure 7. The rates were increased and the curve was fitted to the experimental curve by manual

testing of different sets of parameters. The final result is shown in figure 9(a). This model follows the experimental curve quite well, but has a slight lag in the onset after the initial delay.

To check whether the activity of the modeled network followed that of its real conterpart, the activity was measured for different starting concentrations of CheYp. The experimental values are said to be measured as the initial slope of the curve directly after the delay [8]. Because of ambiguities in when the delay actually is over, the maximal slope, which usually occurred somewhere between 0.4 and 0.8 seconds, was measured instead. As is shown in figure 9(b) the shape of the curve is sigmoidal just like the experimental curve, even though the model predicts a higher activity of the enzyme at the higher concentrations of CheYp. A higher activity can be anticipated already from the results in figure 9(a) where the modeled CheZ has a slightly higher activity than what is seen in the experimental curve, something which is demonstrated in a steeper slope after a slightly longer initial delay. Because of time limits for the project it was decided to move forward using this imperfect reaction network.



(a) The red curve is the average from 10 simulations with the reaction network described in the text and appendix A.3.4. The black curve is the experimental reference from figure 7(a).

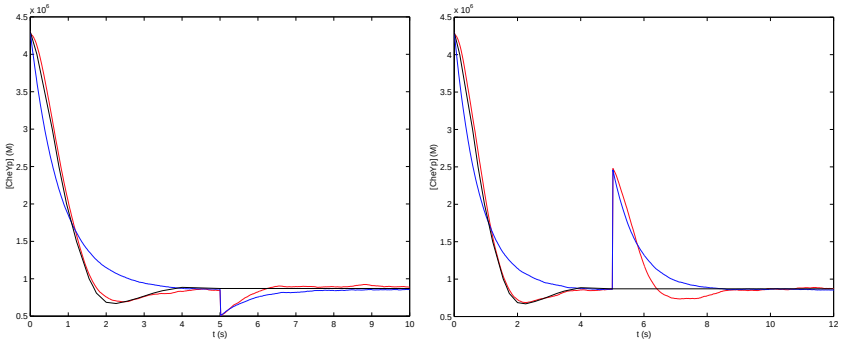


(b) Blue stars indicate activity measurements for different concentrations of CheYp. Each data point is the average of measurements from 5 simulations. The standard deviation was in the order of 10^{-14} to 10^{-15} μM, so not plotted.

Figure 9: Data from simulations of the CheZ system.

To further investigate the behaviour of the oligomerising model compared to the simpler M-F model (nonoligomerising with semi first-order kinetics), simulations were performed where a peak or a drop in CheYp was introduced. The total concentration of CheY (phosphorylated and unphosphorylated form) was kept constant, which means that effectively a shift in the equilibrium was introduced. See figure 10 for results.

These results are not particularly reliable, which is clearly seen in the differences in the red curves around the overshoot. However, from these data (and other simulations not shown here) the overshoot seems to be missing in the kinetics of the SmartCell model when going from a lower concentration of CheYp



(a) CheYp decreased momentarily to $0.5\,\mu$M at t=5.0 s. Total amount of CheY and CheYp kept constant.

(b) CheYp increased momentarily to $2.5\,\mu$M at t=5.0 s. Total amount of CheY and CheYp kept constant.

Figure 10: Kinetics of CheZ in response to step changes in CheYp concentration. Blue curve: Modeled oligomerising CheZ. Red curve: modeled non-oligomerising CheZ. Black curve: experimental reference from figure 7(a).

towards equilibrium. A comparison of the oligomerising model with the non-oligomerising model shows differences in the convexity of the curve immediately following CheYp introduction or withdrawal (better seen in the drop curve after zooming). The oligomerising model exhibits a slight delay in the onset, and therefore reaches equilibrium faster than the nonoligomerising model in the drop experiment where no overshoot occur. However, in the peak experiment the oligomerising model is slower due to the overshoot.

Worth noticing is also that the nonoligomerising model for CheZ activity used here and in M-F's work is a simplification of what actually takes place at a molecular level. The reaction is modeled as a single unidirectional reaction not taking complex formation between CheYp and CheZ into account, which means that the activity of CheZ depends linearly on CheYp concentration (compare with activity curve for mutant CheZ in figure 7(b)). This simplified model therefore becomes unphysical in the limit of high CheYp concentrations when CheZ should become saturated (as is also the case in figure 7(b) for CheYp concentrations above 2.5 μM).

After incorporating the oligomerising CheZ model into the chemotaxis reaction network simulations for different initial amounts of CheZ were run to calibrate the system. A wild-type concentration of CheYp was reached for a CheZ concentration of $0.37\,\mu$M, which is even ten times lower than for the nonoligomerising CheZ. It should be noted though that estimates of intracellular CheZ vary from $0.1\,\mu$M to $15\,\mu$M (appendix 2 in [26]).

7.3 Remodeling Diffusion in Larger Systems

Running the steady-state simulations (8 seconds of simulation time) requires around 16 hours of computation time on a 750 MHz processor in a standard PC. Knowing that adaptation for higher ligand concentrations occurs over several hundreds of seconds, one easily realises that this is too slow for practical use.

In which part of the code does then the simulation program spend most of the execution time? A profiling shows that it is in the functions where the updating of the event queue takes place. Since this is an iterative process, there exist two alternative approaches to optimise the algorithm with respect to time. Either by speeding-up each cycle (in this case for example by writing a more approximate function for calculating the logarithms), or to reduce the number of iterations, i e updates, to be performed.

A second question also arises – is there any specific process that has an average time-step much smaller than the rest, thereby constituting a majority of the updates? As might be expected, diffusion is outstanding as responsible for 99% of the executed events. In sight of this, the second approach mentioned above is probably the best choice for optimising the program, which in this case implies that the model of diffusion has to be changed.

Several suggestions on how to change the model of diffusion arose during discussions with other people in the group. Many of them included the introduction of a fixed time-step well below the average putative time for the fastest non-diffusion reaction. However, this will inevitably violate the conservation of mass in the system, unless a huge administrational function is implemented, and the central idea of a self-organising event queue is lost. The derivation and implementation of that kind of model, may it be similar to the model at hand or a distribution-based approach applied to each volume element, would be complicated and is definitely out of scope of this diploma work.

The solution to the problem that was finally implemented was to make the diffusion processes depend only on the *gradient* of the diffusing species, and only allowing transport from a higher to a lower concentration. The following reasoning was used to arrive at this implementation. Assume two neighbouring slots indexed i and j. Then the rate of diffusion from s_i to s_j is as follows

$$v_{i\to j} = \frac{D}{\lambda^2}[s_i] \tag{9}$$

and similarly, the rate of diffusion backwards,

$$v_{j\to i} = \frac{D}{\lambda^2}[s_j] \tag{10}$$

The netto flow over the boundary can then be written as

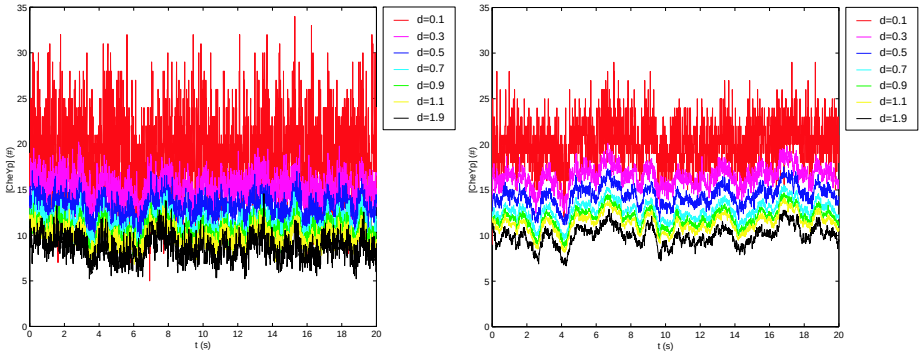
$$v_{tot} = \frac{D}{\lambda^2}[s_i] - \frac{D}{\lambda^2}[s_j] = \frac{D}{\lambda^2}([s_i] - [s_j]) \tag{11}$$

which suggests an approximation of the rate of diffusion

$$v_{i\to j} = \begin{cases} \frac{D}{\lambda^2}([s_i] - [s_j]) & [s_i] > [s_j] \\ 0 & [s_i] \leq [s_j] \end{cases} \tag{12}$$

However one should be aware of that the extremes (the fast and slow particles) will be lost, and hence some of the stochastic behaviour of the system using this model. However, a good feature of the approximate model is that it converges towards the exact model of diffusion for low numbers of molecules, which are the cases when accurate diffusion is probably most interesting to incorporate into the model.

To evaluate the impact of this simplified diffusion on the chemotaxis system compared to the previous, more accurate, model the concentration of CheYp was monitored in slices of the cell at varying distances from the pole. For results of the diffusion experiment in the chemotaxis network, see figure 1, and for a better understanding of what the cell geometry looks like, see figure 1.



(a) Exact model of diffusion.

(b) Approximate model of diffusion.

Figure 11: Distribution of CheYp in slices perpendicular to the longest axis of the cell. Legends indicate the distance from the pole (in μm).

Comparing the plots one notes that a gradient of CheYp is built up in both cases. In the case of the exact diffusion model the system is more noisy, and the concentration in the different slices overlap more than for the simple model. The execution time required for an 8 s simulation decreases significantly from 16 hours to less than 40 minutes.

Moreover, for the sake of chemotaxis, the results clearly show that according to the current model there exist a slight gradient of CheYp in the cell. Whether this has implications for the function of the network is not yet known.

7.4 Modeling an Adaptive Signaling Network

With a faster simulation process and adjusted CheZ activity it was time to model an adaptive reaction network. Again Morton-Firth's model was used as a basis for model building.

First all reactions necessary for a fully functional adaptive network had to be identified. As described in the two-state model presented by Barkai and Leibler the receptor complex is believed to exist in one out of two conformations; either inactive or active. It is only in the active state that the receptor complex is able to undergo autophosphorylation (in the following an asterisk (*) will denote the active complex).

A schematic view of the transition reactions for the receptor complex is presented in figure 12. Note that the lowest and highest methylation states are assumed to exist in only inactive and active conformations respectively. Note also that all demethylation reactions (performed by CheB) occur only for the active conformations as stated in the list of constraints on page 25. For clarity reasons only the core reactions in the switching and phosphorylation reactions are included. In addition to these receptor states variants exist, in which either CheB, CheBp, CheY or CheYp is bound to the kinase binding site of CheA. Also, CheBp and CheR bind to the Tar receptor and contribute with yet some states.

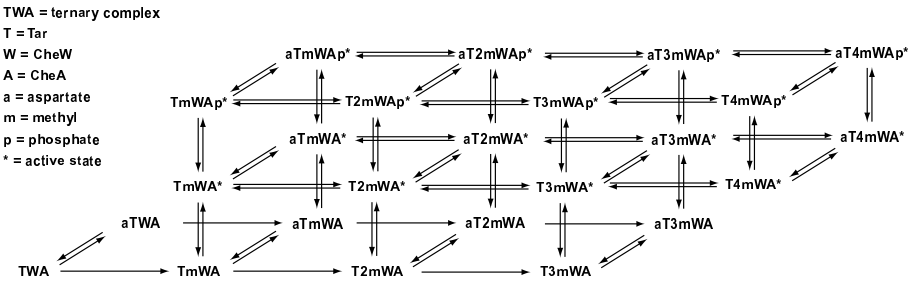


Figure 12: Receptor core reactions in the adaptive model.

Experimentally the probability for a receptor complex to exist in either state has been estimated for some levels of methylation, and after extrapolation a full set of probabilities has been developed [26]. See table 12 for data.

In some initial simulations rate constants were derived for the switching reactions ($TWA \rightleftharpoons TWA^*$) based on these probabilities. Assuming a very fast transition however resulted in a huge computational load, and therefore another modeling strategy was chosen. Instead of modeling each receptor state in separate species for active and inactive conformation they were kept in one pool of molecules and fractions of this set of species were enabled for different reactions.

An illustrating example is the binding of CheBp to the methylation binding site:

$$T2mWA^* + Bp \xrightarrow{k} T2mBpWA^*$$

The rate of the reaction is calculated as

$$v = k \cdot [T2mWA^*] \cdot [Bp]$$

If instead of using $[T2mWA^*]$ the receptor is administrated as $[T2mWA_{tot}] = [T2mWA] + [T2mWA^*]$, the rate of the reaction is calculated as

$$v = k \cdot p_a \cdot [T2mWA_{tot}] \cdot [Bp]$$

where p_a is the probability of T2mWA being in an active conformation. The only assumption made is that the equilibrium between inactive and active conformation is reached instantaneously, or at least by far much faster than the binding of CheBp to the receptor. At very low concentrations of the receptor complex the result will however be that one in effect works with fractions of single molecules.

The same reasoning was used for modeling the binding of aspartate to the receptor since that reaction is also much faster ($k_{on} = 10^9 M^{-1}s^{-1}$) than the diffusion limited reactions in the pathway. The probabilities were derived from the experimentally measured values of the K_Ds. A further complication appeared as the binding reactions coincide with the switching probabilities, so that mixed probabilities had to be derived as well. For a full description of how these probabilities were calculated, see page 53.

For the full set of reactions in the adaptive reaction network see table 14. Initial calibration simulations were performed until a steady-state was found from the initial-amounts in table 15. From the steady-state data initial amounts for the full reaction network were calculated in order to save computation time in later experiments since a calibration takes around 150 seconds of simulation time, or some 15 hours of computation time. From the full set of initial amounts the first real simulations were run.

In figure 13 the mean value of three adaptation simulations are plotted for a network both with the simple and the complex CheZ model. It is obvious that both models have adaptive power, but that none of them supports perfect adaptation (adaptation error approximately 4% in both cases). A difference between the two models, seen in the figure, is that the system with a non-oligomerising CheZ is less noisy. Comparing the excitation (the change in CheYp concentration) it is markedly higher than what Morton-Firth recorded for the same inducing aspartate concentration (figure 45 in [26]).

In figure 14 the methylation curves for the ground states of the Tar receptor are plotted. Note that immediately after the addition of aspartate T2mWA increases. This is an effect from delivery of CheBp to the cytosol as the active conformation becomes less stable. The corresponding peak in CheBp is seen in 14(b). Morton-Firth observed the same peak in CheBp but in the SmartCell

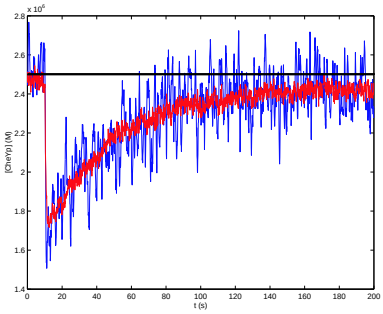
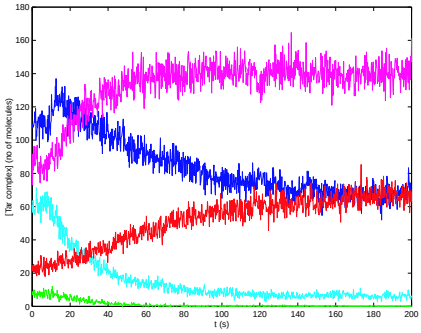


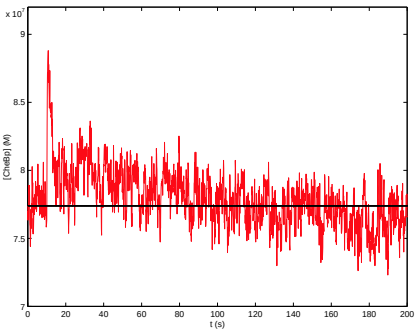
Figure 13: CheYp concentration for [asp]=10 μM at t $\geq$10 s. Both the oligomerising model (blue curve) and the nonoligomerising model (red curve) shows adaptation, but neither of them perform exact adaptation since the steady-state concentrations deviate from the starting concentration (black curve).

model the duration of the peak is ten times shorter, whereas the peak is slightly higher. This difference is probably due to different concentrations of aspartate.

Experimentally an overshoot in bias has been reported after exposure to attractant only for a shorter period of time, an impulse [28]. Some impulse experiments with saturating concentrations of aspartate were performed and it was found that the impulse has to last for approximately 10 seconds for a signal to be detected. This can be understood from the plots of the Tar complex in figure 14(a) where it is clearly seen that the methylation level – the memory of the system – does not change significantly until after around 10 seconds of exposure. Here the model is clearly inconsistent with experimental results where pulses lasting only fractions of seconds produce clear undershoots.



(a) Concentrations of the standard states of the Tar receptor; TWA (greeen), TmWA (cyan), T2mWA (blue), T3mWA (magenta) and T4mWA (red).



(b) Concentration of CheBp in the cytosol (red curve). The starting concentration is indicated by the black line.

Figure 14: Adaptive response for [asp]=10 μM at t $\geq$10 s.

The gain was calculated for the model with an oligomerising CheZ to approximately 0.1 for an inducing aspartate concentration of $10\,\mu$M. This value seems very low compared to gain parameters referenced in the literature ranging in value between 6 and 35. Note however that in equation (8) the minimal value of the denominator is 0.7 for $10\,\mu$M aspartate, and the maximal value of the nominator is 0.3, so the maximal limit of the gain parameter for this induction level is 0.43, much below the experimentally monitored value. This clearly demonstrates the problem with comparing the gain between different models when the inducing aspartate concentration is not given. The value of 1.2 for Morton-Firth's work is therefore not comparable with the result above.

8 Conclusions and Future Directions

8.1 Kinetics and Diffusion in SmartCell

From validation experiments of the kinetics and diffusion models used in SmartCell, it is clear that the algorithm converges towards the analytical solutions for large numbers of reacting particles. A recurring problem is however that the exactness of the algorithm results in long computation times for large numbers of reacting species, and therefore the stochastic approach used here better suits smaller systems with low copy numbers of each species. The diffusion algorithm in SmartCell should probably be given a new mathematical derivation with computational speed in mind. For molecules present in larger amounts this algorithm could typically be based on continuous mathematics.

The experiments with low numbers of reacting species clearly show that the kinetics is highly stochastic. This will be an important property of the simulation algorithm in the modeling of for example genetic switches regulated by transcription factors in low amounts or other on/off reactions, regulated by the build-up of a species, whose time of onset can vary significantly due to stochasticity.

8.2 Modeling Larger Systems in SmartCell

The complexity and size of the chemotaxis reaction network posed new problems to modeling in SmartCell not encountered in the smaller validation experiments. Two problems closely related to the nature of the SmartCell algorithm appeared during the project; the computational overload generated *(i)* by the diffusion model used and *(ii)* by other rapid equilibrium processes.

The current solution to the problem of modeling diffusion is only temporary, since the problem turned out much too complex to be covered in a diploma thesis. For the second problem the method suggested is to use "smart" modeling. By careful analysis of the reaction network assumptions can usually be made which reactions have to be modeled in detail and which can be simplified. Also, when there are big differences between rate constants the fast equilibria can be modeled as being reached instantaneously, as has been done in the chemotaxis model in this work. Problems with parts of the model slowing down the simulation will appear over and over again. To overcome this inherent problem the possibility of dividing the simulation algorithm into different layers treating different time-scales should be investigated. An interesting approach would be to integrate a continuous differential-equation solver into the currently used stochastic algorithm.

Another problem that appeared during the work with the chemotaxis network was the user-unfriendliness of working with the model description format in a normal text-editor such as Emacs. Each reaction had to be typed into an ordinary text file with the apparent risk of incorporating typing errors. Therefore there is an urgent need of a specific editor, with a well-designed user interface, for model building. With the generality of the model description format and

the simulation algorithm comes also the problem with a restricted possibility of developing solutions for specific problems appearing for some models. In the case of the chemotaxis reaction network much tedious work was needed to type in the same reactions for many different states of the receptors that had to be modeled as discrete species (which could be conveniently implemented as flags in a software object in a specific simulation program). A graphical user interface should typically include the possibility of defining a process valid for a set of species.

The general conclusion from modeling in SmartCell is that what you gain in generality of the framework you lose in computational speed and ease of modeling some very specific reactions.

8.3 Something about Chemotaxis

During the modeling process I also stumbled over many other problems related to the chemotaxis network, most of which have been discussed in previous work.

8.3.1 Regulation of Phosphatase Activity

As has been reported in several papers the activity of CheZ is regulated [8], but the question of how to model this control system has not yet been answered. The approach taken in this work proved unsuccessful in its current form, but the model could probably be improved by a more careful modeling and parameter tuning (ideally complemented with more experimental data). In its current state however it slows down the simulation compared to the simple model that Morton-Firth used in his thesis [26], and since no new properties emerged in the oligomerising system it is not worth the effort using it at the moment. The slightly lower sensitivity of the oligomerising system is probably not a real difference in behaviour, but an artifact from the fact that the system is more noisy than its simpler counterpart.

A new set of simulations have to be performed with adjusted values of association rates for the oligomerisation reactions of CheZ. Since the reaction parameters of the CheZ system were tuned against kinetic data from *in vitro* experiments in an aqueous environment, these have to be adjusted to be applicable for a cellular environment. The association rates derived from these experiments (listed in table A.3.4) are sometimes in the order of $10^8\,\mathrm{M}^{-1}\mathrm{s}^{-1}$, which is two orders of magnitude larger than what is considered the diffusion limited reaction rate for a bimolecular reaction taking place in the cytosol. Changing these parameters could introduce more of inertia in the system and make the response of the CheZ oligomerisation slower, which in turn could produce sharper peaks and higher gain in the chemotactic response.

The idea of a CheYp-dependent cooperative onset of the phosphatase activity is very interesting, and could be motivated from the function of the network. A delay enables full excitation and the cooperative "all-or-nothing" effect as well, so that once the CheZ response starts it is fast and no signal (i e noise) remains. Models including a regulated phosphatase activity could also, combined with

the remodeling of other parts of the network, help resolve the problem of low gain in the network.

8.3.2 Diffusion in E coli

Steady-state simulations in this work clearly shows that according to the current model there is a gradient of CheYp in an *E coli* cell, but it is still not known whether this has any function or implication for the behaviour of the chemotaxis signaling network. The localisation of CheYp is more interesting for a model incorporating the flagella and not calculating swimming behaviour in the Hill equation as must be done in the current model. For the future several experiments exploring the importance of localisation of the receptors and flagella, and changing diffusion parameters could be interesting. Elowitz et al notes in [14] that the chemotactic response time measured to be 50-200 ms is of the same order of magnitude as the time required for CheY to diffuse through the cytoplasm of *E coli*[2], a fact which supports the idea that diffusion and localisation plays a role in the intracellular chemotactic signaling network.

In figure 11 it is obvious that the gradient is sharpest in the half of the cell close to the pole where the Tar receptors are clustered. Therefore one approach speeding up the simulation, but still keeping an exact handling of diffusion where it is relevant, could be to construct a geometry with different lattice units in different parts of the geometry. The most efficient way of speeding up the program is however probably the development of a completely new algorithm for treating the diffusion processes.

8.3.3 Adaptation

The most apparent weakness of the adaptive reaction network is that is does not support perfect adaptation, but shows an adaptation error of around 4% when induced with $10\,\mu$M aspartate. Morton-Firth pointed out in his thesis that the rate of dephosphorylation of the active receptor complex is correlated to adaptation error, where the slower the dephosphorylation process the bigger the adaptation error. This is also a reason for his treatment of CheY and CheB interaction with the active phosphorylated receptor complex, in order to artificially speed up dephosphorylation. In my model I instead added an interaction between CheY and CheB with the inactive (or active) but unphosphorylated receptor, that after CheY or CheB binding could undergo autophosphorylation followed by a very fast phosphotransfer [31]. This should to some extent compensate the slower binding of CheY and CheB to the active receptor in my model compared to that of Morton-Firth.

Another possible reason behind the imperfect adaptation could lie in the modeling of CheR-Tar interaction in my model. In the data from Morton-Firth's model that I used, CheR only interacts with the inactive receptor, whereas this is not a constraint in Barkai & Leibler's nor in Yi et al's analysis of the reaction network (see the list on page 25). However, the amount of bound CheR

[2]Estimating the dimension of an *E coli* cell to be approximately $1\,\mu$m and a diffusion constant of $10\,\mu\text{m}^2\text{s}^{-1}$ for CheY according to [14].

is constant in my model, so their constraint considering CheR is not violated by this assumption. Moreover Morton-Firth's motivation for the modeling of CheR-Tar interaction is far from convincing so it would be interesting to rebuild my model including methylation processes for the active receptors too.

A peak in CheBp was observed right after the induction with aspartate as was also seen by Morton-Firth in his work. This behaviour can not be foreseen by the reaction scheme established for the chemotaxis reaction network, but comes as no surprise from the two-state model in its current form. As ligand concentration increases the overall probability of the receptor being in an active state decreases and hence much of the CheBp bound to Tar is released. This more than compensates the decrease in CheBp generation by CheA-mediated phosphorylation. The question is then, is this peak in CheBp only a model artifact, or can it be experimentally observed? If it is the real behaviour of the chemotaxis transduction pathway, the global adaptation response does not exist but is just an erroneous assumption of what the adaptation response looks like inside the cell.

Unfortunately no time was left to run extensive studies of response times, sensitivity and gain within the limits of this diploma work. Preliminary results show however that the gain in my model is approximately 0.3 for induction with 10 μM aspartate. This value is however not straightforward to compare with results from modeling or experimental work, due to the ambiguity in the definition of the gain parameter. I would suggest that the gain is redefined as "maximal gain" of the network or a "gain function" over a range of inducer concentrations in order to be a more useful measure.

Initial test runs also showed a slightly higher sensitivity of the adaption network with the non-oligomerising mechanism of CheZ. It was suggested that this difference was due to the greater noise in the complex system, which would be interesting to study further.

Both the experiments with step responses and impulse experiments indicates that a crucial component in the chemotaxis model is missing. The step responses exhibit too low a gain in current models and the impulse response is too slow and does not exhibit the overshoot that has been observed experimentally. It is tempting to speculate where this component might be. Most work so far has focused on the receptor complexes or the control of CheZ activity, e g cooperative signaling has been proposed as a means of amplification (and experimentally established in a recent paper [15]). However, this can only serve as an amplifier for small amounts of inducer when few receptors are occupied or when the receptor is of low abundance. Secondly it would be costly for the cell to have an amplification *before* CheYp production since this will lead to greater variations in CheYp concentrations and thereby a higher consumption of energy in the form of ATP. The cell would ideally let the CheYp concentration vary over the limit posed by the noise in the system to avoid an unreliable response. From this point of view it would be more beneficial for the cell to put the amplification (cooperativity) further down the network, which is at the interaction with the motor complex. Therefore I would like to see new experiments providing more information about this part of the chemotaxis reaction network.

When it comes to the impulse response however a stronger cooperative effect

at the flagellar motor is not enough. It also requires a delay in the swimming behaviour compared to the composition of the environment. This could actually be a good point for diffusion to come into play as a delaying factor. The time scale for diffusion of approximately 100 ms is not enough though to explain the overshoot, but this behaviour requires a delay in the very signal as well (a faster initiation of methylation reactions) and maybe also in the response of the flagella. It should be noted however, that the effective signal in experiments might be longer than 0.2 seconds due to aspartate that is still bound when washed away from the environment. In the model used in this work I also assume that this equilibrium is reached momentarily, so for low concentrations of aspartate and short times this kind of experiment should be performed with real modeling of the binding reactions (also due to stochasticity in the binding of few molecules of aspartate).

For the future some interesting experiments to be performed would first of all be to make full sets of test runs for response times, gain and sensitivity. The next step would be to include other receptor types into the model and find a good way of modeling cooperativity in their signaling within the limits of the SmartCell model description format. To specifically explore the possibilities in the SmartCell framework studies of the influence of localisation of receptors and flagella, diffusion of CheY etc on the behaviour should be explored.

8.4 Modeling Biological Processes in General

The main lesson learnt from the modeling process in this work has been that it is a time-consuming activity. A modeller has to be prepared to do "dirty" work, browsing through lots of articles (or databases) and simplify, rationalise and do more or less well-founded assumptions. For modeling almost any network of biological reactions only sparse data is available at present, and what is known is often unclear and the modeler has to carefully assess the quality and relevance of published data. The modeler will most certainly also stumble over many inconsistencies during data collection. Another problem with the data available on reaction rates or dissociation constants is that the measurements have usually been carried out *in vitro*, which is not always comparable with physiological conditions with respect to pH, ion-strength, environment composition etc.

A second complication inherent in the idea of systems biology lies in the difficulty of getting an overview of the problem and an intuitive feeling for how the system should behave because of the complexity of the system. A major idea with systems biology is also to study more or less unexpected, emergent properties of the system. But how is this treated when you can not trust the input data (sometimes based on guesses and wild assumptions) you feed into the model? How will you know when the result is reliable and where the inconsistencies may be? Therefore it is probably important to decompose a system into subsystems as far as possible to first get an understanding of the smaller problems and also later work in an iterative fashion remodeling different parts as the modeling process advances. Ideally this will also enable the modeler to isolate the most important or limiting empty patches in the reaction network and initiate additional experiments to be performed.

Currently much too little quantitative data is available, and in addition information residing in articles and spread over many different databases poses an administration problem. For useful access to all this information new tools for efficient data retrieval have to be developed. Also for systems biology to advance more biochemical data to perform modeling on is necessary. Unfortunately today the trend is towards qualitative analysis, like information from yeast hybrid systems elegantly provide the topology of reaction networks, but is useless for quantitative modeling unless biochemical data on rates and concentrations is added. Otherwise the modeler will be left fumbling forever for parameters in a system of too many degrees of freedom.

8.5 Future of SmartCell

This work shows that SmartCell is functional and correctly predicts basic kinetics and diffusion. However, working with SmartCell in its current state have shed light on many features missing for SmartCell to be applicable to a wider range of modeling problems and to be more user-friendly.

First of all, to make the modeling process easier in SmartCell a graphical user interface, including an editor for model files, has to be developed. For ease of use the user should be able to control activities ranging from model building to simulation and visualisation of output data within this graphical interface. As models accumulate the user could also typically choose which partial models of the cell to include in his or her reaction network. In a similar way the user would ideally be allowed to build a geometry in a drag-and-drop fashion from basic units provided by the program.

At present DNA species and multi-step events are implemented, but their functionality and the computational accuracy need to be validated in models including for example gene expression. For processes such as growth, cell division and other processes influencing the morphology of the cell other, more high-level, solutions have to be developed that act directly on the geometry of the system. Also the modeling of more complex cells, such as the compartmentalised eukaryotic cells, requires a better handling of the geometry functions in SmartCell.

When it comes to the very simulation algorithm the main problem is speed. Some speed could be gained by small changes in the current algorithm, for example by making sure that probabilities are not updated unless the limiting slot (the zero-valued slot) is changed. Usually this will not produce huge gain in speed, especially as it has administrational costs, but for solving the problem with different time-scales for different processes new solutions have to be found. A common solution to speed problems is to parallellise the code. The basic simulation algorithm in SmartCell however, does not lend itself to parallellisation due to the high degree of interconnectivity. What could be imagined however is, that for more complex systems, to handle different, relatively independent, system modules on different processors, or to let each site in the geometry be represented by a single CPU, all of which would be coordinated by a central processor, much like a physical representation of the program architecture.

A continued development of SmartCell looks promising. The originality of SmartCell lies in the general approach to modeling biological processes, both in how reactions are described and how the geometrical aspects of the system are handled. By the introduction of the multi-step processes a considerable gain in performance is expected, and thereby offering good means for modeling and simulation of genetic networks including transcription and translation processes. In combination with the stochastic kinetics and the support of diffusion and localisation, the model description format and a graphical user interface SmartCell has the potential to become a highly accurate and user-friendly tool for scientists from a wide variety of disciplines who wish to perform modeling and simulation of complex biological processes.

Acknowledgements

My sincerest gratitudes to Luis Serrano for making this Master thesis project possible. It has been a pleasure taking part in the scientific life at EMBL in general, and especially in the Serrano spirit.

I would also like to thank Johan Elvnert (TTA Technotransfer, Uppsala) and the Swedish Natural Science Research Council for financing the project, Torgny Fornstedt (Biology Education Center, Uppsala university) for coordinating the diploma work and Måns Ehrenberg (Department of Cell and Molecular Biology, Uppsala university) for advisable criticism on the manuscript and for the formal examination.

Many thanks to Anders Kaplan, my SmartCell guru, for patiently answering endless questions. Thanks Jesper Borg for discussions and sharing the joy of C++-programming, Joaquim Mendes for discussions and keeping White alive, Edward Fäldt for nice nerd discussions on program efficiency and Matlab, and Stéphane Thore for providing a biochemist's insights into enzyme kinetics.

Thanks Vicente for making the gloomiest of cave-day jjjappy. Thanks Anuschka and Kostas, my friends-in-need and the best of neighbours. Thank you, all present and former members of the Serrano group for always being friendly, helpful, and making life at EMBL so pleasant.

Merci coupcoup Genaro for the Paris trip, and daily chats on existential topics over a cup of coffee. Thanks Petra for sportive and patriotic moments, Lars for kindly overlooking long-term CD loans, Christian for showing me a bit of civil Germany, and all friends at EMBL for cheering up everyday life.

Thanks Mum and Dad for mental support in cases of emergency, Martin for introducing Jeff Buckley and Smartcruiser in my life, and Karin for those unforgettable *kras-kras* moments in Heidelberg (in absence and presence of Ritter). Thanks Erik, you are just great. In fact there would be no SmartCell at all without E A Jansson 24-hour support central. I will always be your highly devoted groupie.

Thanks and farewell SlowCell. You are my friend after all.

References

[1] B Alberts, D Bray, J Lewis, M Raff, K Roberts, and J D Watson. *Molecular biology of the cell.* Garland Publishing Inc, 3rd edition, 1994.

[2] U Alon, L Camarena, MG Surette, B Aguera y Arcas, Y Liu, S Leibler, and J B Stock. Response regulator output in bacterial chemotaxis. *EMBO Journal*, 17:4238–48, 1998.

[3] M Andér. Documentation on the SmartCell framework, version 1.0. Software documentation, European Molecular Biology Laboratories, Heidelberg, Germany, 2002.

[4] S Asakura and H Honda. Two-state model for bacterial chemoreceptor proteins. *J Mol Biol*, 176:349–67, 1984.

[5] N Barkai and S Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–17, 1997.

[6] A N Barnakov, L A Barnakova, and G L Hazelbauer. Efficient adaptational demethylation of chemoreceptors requires the same enzyme-docking site as efficient methylation. *Proc Natl Acad Sci*, 96:10667–72, 1999.

[7] Y Blat and M Eisenbach. Phosphorylation-dependent binding of the chemotaxis signal molecule CheY to its phosphatase, CheZ. *Biochemistry*, 33:629–34, 1994.

[8] Y Blat, B Gillespie, A Bren, F W Dahlquist, and M Eisenbach. Regulation of phosphatase activity in bacterial chemotaxis. *JMB*, 284:1191–99, 1998.

[9] K A Borkovich, L A Alex, and M I Simon. Attenuation of sensory receptor signaling by covalent modification. *Proc Natl Acad Sci*, 89:6756–60, 1992.

[10] D Bray, R B Bourret, and M I Simon. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Molecular Biology of the Cell*, 4:469–82, 1993.

[11] D Bray, MD Levin, and CJ Morton-Firth. Receptor clustering as a cellular mechanism to control sensitivity. *Nature*, 393:85–8, 1998.

[12] P Cluzel, M Surette, and S Leibler. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science*, 287:1652–55, 2000.

[13] M Ehrenberg. Kinetics for molecular biotechnology. Course compendia, Uppsala University, Department of Molecular Biology, 1998.

[14] M B Elowitz, M G Surette, P-E Wolf, J B Stock, and S Leibler. Protein mobility in the cytoplasm of escherichia coli. *Journal of Bacteriology*, 181:197–203, 1999.

[15] J E Gestwicki and L L Kiessling. Inter-receptor communication through arrays of bacterial chemoreceptors. *Nature*, 415:81–84, 2002.

[16] M A Gibson and J Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A*, 104:1876–89, 2000.

[17] D Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Physics*, 22:403–43, 1976.

[18] P J Halling. Do the laws of chemistry apply to living cells? *TiBS*, 14:317–18, 1989.

[19] D C Hauri and J Ross. A model of excitation and adaptation in bacterial chemotaxis. *Biophys J*, 68:708–22, 1995.

[20] Anders Kaplan. *On whole-cell modelling and simulation.* MSc thesis, Uppsala University, October 2001.

[21] J Li, RV Swanson, MI Simon, and RM Weis. The response regulators CheB and CheY exhibit competitive binding to the kinase CheA. *Biochemistry*, 34:14626–36, 1995.

[22] L M Loew and J C Schaff. The Virtual Cell: a software environment for computational cell biology. *Trends in Biotechnology*, 19:401–6, 2001.

[23] G S Lukat, W R McCleary, A M Stock, and J B Stock. Phosphorylation of bacterial response regulator proteins by low molecular weight phosphodonors. *Proc Natl Acad Sci*, 89:718–22, 1992.

[24] JR Maddock and L Shapiro. Polar location of the chemoreceptor complex in the escherichia coli cell. *Science*, 259:1717–23, 1993.

[25] H H McAdams and A Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci*, 94:814–19, 1997.

[26] Carl Jason Morton-Firth. *Stochastic simulation of cell signalling pathways.* PhD thesis, University of Cambridge, September 1998.

[27] C V Rao and A P Arkin. Control motifs for intracellular regulatory networks. *Annu Rev Biomed Eng*, 3:391–419, 2001.

[28] J E Segall, S M Block, and H C Berg. Temporal comparisons in bacterial chemotaxis. *Proc Natl Acad Sci*, 83:8987–91, 1986.

[29] L Shapiro and R Losick. Dynamic spatial regulation in the bacterial cell. *Cell*, 100:89–98, 2000.

[30] PA Spiro, JS Parkinson, and HG Othmer. A model of excitation and adaptation in bacterial chemotaxis. *Proc Natl Acad Sci*, 94:7263–68, 1997.

[31] R C Stewart. Kinetic characterization of phosphotransfer between chea and chey in the bacterial chemotaxis signal trasnduction pathway. *Biochemistry*, 36:2030–40, 1997.

[32] J B Stock and D E Koshland. Changing reactivity of receptor carboxyl groups during bacterial sensing. *J Biol Chem*, 256:10826–33, 1981.

[33] M Tomita. Whole-cell simulation: a grand challenge of the 21st century. *Trends in Biotechnology*, 19:205–10, 2001.

[34] H Wang and P Matsumara. Characterization of the CheAs/CheZ complex: a specific interaction resulting in enhanced dephophorylation activity on CheY-phosphate. *Mol Microbiol*, 19:695–703, 1996.

[35] M Welch, K Oosawa S I Aizawa, and M Eisenbach. Effects of phosphorylation, Mg^{2+}, and conformation of the chemotaxis protein CheY on its binding to the flagellar switch protein FliM. *Biochemistry*, 33:10470–76, 1994.

[36] T-M Yi, Y Huang, M I Simon, and J Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci*, 97:4649–53, 2000.

A Model Data

A.1 Kinetics Test Model

A.1.1 The Unimolecular Reaction

The test geometry was a cubic container with lattice unit $\lambda = 2\,\mu$m. Simulations were run for 2 seconds with the following parameters.

Id	a_0 (M)	b_0 (M)	$k_f(s^{-1})$	$k_b(s^{-1})$	*# particles*
low	$2.0 \cdot 10^{-8}$	0.0	9.0	1.0	$9.6 \cdot 10^{1}$
high	$2.0 \cdot 10^{-2}$	0.0	9.0	1.0	$9.6 \cdot 10^{7}$

A.1.2 The Bimolecular Reaction

The test geometry was a cubic container with lattice unit $\lambda = 2\,\mu$m. Simulations were run for 2 seconds with the following parameters.

Id	a_0 (M)	b_0 (M)	c_0 (M)	$k_f(M^{-1}s^{-1})$	$k_b(s^{-1})$	*# particles*
low	$1.2 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$	0.0	$5.0 \cdot 10^{8}$	1.0	$1.1 \cdot 10^{2}$
mid	$2.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	0.0	$1.0 \cdot 10^{2}$	$1.0 \cdot 10^{1}$	$1.4 \cdot 10^{7}$
high	$2.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	0.0	$1.0 \cdot 10^{2}$	$1.0 \cdot 10^{1}$	$1.4 \cdot 10^{8}$

A.2 Diffusion Test Model

The model geometry was a row of 15 cubic sites with lattice unit $\lambda = 0.1\,\mu$M as shown in figure 5. One particle was placed in the centre site and allowed to diffuse with a diffusion constant $D = 10\,\mu\text{m}^2\text{s}^{-1}$. The simulation was stopped as soon as the particle reached any of the boundary sites, the current simulation time recorded and a new simulation started until a total of 10^6 simulations had been accomplished.

A.3 The Chemotaxis Model

A.3.1 Notations

In the text some abbreviations for the components in the chemotaxis signaling network occur. These notations (or rules for derivation) are listed below.

Species	*Notes*
A, Y	CheA, CheY etc
Yp	CheYp
Z	CheZ dimer

Species	*Notes*
*n*Z*m*Yp	Oligomer of *n* CheZ dimers and *m=n-1,n* CheYp molecules.
TWA	The Tar receptor, CheW and CheA ternary complex (dimers of each component). Also referred to as *the (Tar) receptor complex.*
TWA*	The receptor complex in its active form.
TWAp*	The receptor complex in its active phosphorylated form (phosphate on CheA).
TWAY(p), TWAB(p)	The receptor complex with CheY(p) or CheB(p) bound to the kinase binding site.
TBpWA*, TRWA	CheBp or CheR bound to the methylating binding site on the Tar receptor.
T*n*mWA	The receptor complex of methylation level *n. n=0..4*

A.3.2 Geometry

In the simulation the geometry of the cell was chosen to be ellipsoidal, with the dimensions of an *E coli* cell (radii of 0.5 μm and 1 μm) yielding a volume of $1.208{\cdot}10^{-15}\,\mathrm{dm}^3$. The lattice unit was chosen to be 0.2 μm leading to 151 cytosolic volume elements and 212 cell membrane surface elements. A subset of the cell membrane was designated to a region *pole* (the 29 elements with the x-coordinate greater than or equal to 0.8 μm in an elliptic geometry centered around origo).

A.3.3 The Phosphorylation Cascade

The spontaneous phosphorylation

$$Y + PoAc \xrightarrow{k_f} Yp + oAc$$

is modeled as a first-order reaction

$$Y \xrightarrow{k'_f} Yp$$

assuming a constant supply of phospho-acetate (PoAc). Assuming that the reaction follows Michaelis-Menten kinetics and that [PoAc] is constant the rate of the reaction is

$$v = k_{cat}\frac{[PoAc]}{K_M + [PoAc]}[Yp] = k'_f[Yp]$$

So from the experimental values of $k'_f = 0.214\,s^{-1}$ at [PoAc]=9 mM ([8]) and $K_M = 7 \cdot 10^{-4}$ M ([23]) the first order k'_f for cellular levels of PoAc can be calculated.

Table 8: Reactions and reaction rates in the simple non-adaptive model. The unit of the rate constants is either s^{-1} or $M^{-1}s^{-1}$ depending on whether the reaction is unimolecular or bimolecular.

Reaction	k_f	k_b	*Notes*
Ligand binding			
TWA + a $\rightleftharpoons$ aTWA	$1 \cdot 10^9$	$1.2 \cdot 10^3$	$K_D = 1.2\,\mu$M [26] [10]
CheA autophosphorylation			
TWA $\rightarrow$ TWAp	15.5		[26]
aTWA $\rightarrow$ aTWAp	0.02		[26]
CheY - CheA interactions			
TWAp + Y $\rightleftharpoons$ TWApY	$1 \cdot 10^6$	1	M-F uses one-directional binding with a higher rate than what diffusion limiting poses. I prefer an equilibrium with diffusion limited k_{on} and $K_m = 6.5\,\mu$M ([31]). K_D adjusted to $1\,\mu$M (considering values for interactions between non-phosphorylated CheA).
TWA + Yp $\rightleftharpoons$ TWAYp	$1 \cdot 10^6$	8	Affinity approximately 6-fold lower than for unphosphorylated CheY in the presence of Mg^{2+} ([21]). Just made a guess from the fact that *E coli* contains high levels of Mg^{2+} ions.
CheB - CheA interactions			
TWAp + B $\rightleftharpoons$ TWAB	$1 \cdot 10^6$	3.2	M-F uses one-directional binding with higher rate with higher rate than what diffusion limiting poses. I prefer an equilibrium with diffusion limited k_{on} for $K_D = 3.2\,\mu$M. [21]
TWA + Bp $\rightleftharpoons$ TWABp	$1 \cdot 10^6$	3.2	No data on K_D available so assumed to be the same as for CheB.
Phosphotransfer			
TWApY $\rightleftharpoons$ TWAYp	650	10	[31] (k_b derived by optimisation).
TWApB $\rightleftharpoons$ TWABp	650	10	

Table 8: continued

Reaction	k_f	k_b	*Notes*
Phosphorylation reactions in cytosol			
Yp $\rightleftharpoons$ Y	0.035	$3.1 \cdot 10^{-3}$	k_f from [23]. k_b from [8], converted to unimolecular rate constant and scaled for $[PoAc]_{cell} = 10\,\mu$M.
Yp $\stackrel{Z}{\rightarrow}$ Y	$1 \cdot 10^6$		Diffusion limited bimolecular reaction assumed. [26]
Bp $\rightarrow$ B	0.35		[26]
Motor reactions: CheYp binding			
FliM*nYmYp* + *Yp* $\rightleftharpoons$ FliM*nY(m+1)Yp*			$K_D = 1.43\,\mu$M [35]
FliM*nY* + *Yp* $\rightleftharpoons$ FliM*nYYp*	$7 \cdot 10^6$	1.43	Assuming diffusion limited rate of association ($10\,\mathrm{M}^{-1}\mathrm{s}^{-1}$). Each motor is assumed to carry seven binding sites for CheY and CheYp, and again the binding to each of them is independent and is therefore additive.
FliM*nYYp* + *Yp* $\rightleftharpoons$ FliM*nY2Yp*	$6 \cdot 10^6$	2.86	
FliM*nY2Yp* + *Yp* $\rightleftharpoons$ FliM*nY3Yp*	$5 \cdot 10^6$	4.29	
FliM*nY3Yp* + *Yp* $\rightleftharpoons$ FliM*nY4Yp*	$4 \cdot 10^6$	5.72	
FliM*nY4Yp* + *Yp* $\rightleftharpoons$ FliM*nY5Yp*	$3 \cdot 10^6$	7.15	
FliM*nY5Yp* + *Yp* $\rightleftharpoons$ FliM*nY6Yp*	$2 \cdot 10^6$	8.58	
FliM*nY6Yp* + *Yp* $\rightleftharpoons$ FliM*nY7Yp*	$1 \cdot 10^6$	10.01	
Motor reactions: CheY binding			
FliM*nYmYp* + *Y* $\rightleftharpoons$ FliM*(n+1)YmYp*			$K_D = 443\,\mu$M, i e CheY binds FliM 310 times more strongly than CheYp. [35]
FliM*mYp* + *Y* $\rightleftharpoons$ FliMY*mYp*	$7 \cdot 10^6$	443	
FliMY*mYp* + *Y* $\rightleftharpoons$ FliM2Y*mYp*	$6 \cdot 10^6$	886	
FliM2Y*mYp* + *Y* $\rightleftharpoons$ FliM3Y*mYp*	$5 \cdot 10^6$	1329	The rate constants were derived by the author in the same way as for CheYp above. *m,n=0..7*
FliM3Y*mYp* + *Y* $\rightleftharpoons$ FliM4Y*mYp*	$4 \cdot 10^6$	1772	
FliM4Y*mYp* + *Y* $\rightleftharpoons$ FliM5Y*mYp*	$3 \cdot 10^6$	2215	
FliM5Y*mYp* + *Y* $\rightleftharpoons$ FliM6Y*mYp*	$2 \cdot 10^6$	2658	
FliM6Y*mYp* + *Y*			

Table 8: continued

Reaction	k_f	k_b	*Notes*
$\rightleftharpoons$ FliM7Y*mYp*	$1 \cdot 10^6$	3101	
Motor reactions:			
dephosphorylation			
FliM*nYmYp*			[26]. Same rates as when
$\rightleftharpoons$ FliM*(n+1)Y(m-1)Yp*	0.035	$3.1 \cdot 10^{-3}$	freely diffusible in cytosol.

Diffusing Species

No data was available for other species than CheY, but diffusion was assumed to be the same for all species since molecular mass makes a small contribution (inverse cubic proportionality). Here the diffusion is mainly used to let the particles move, and is probably most interesting for CheYp and CheBp that are produced locally in the cell.

Table 9: Diffusing species in the chemotaxis model

Species	D ($\mu m^2 s^{-1}$)	*Notes*
Yp	10.0	Value from [14]
Y	10.2	Value assumed to be slightly higher because no negatively charged phosphate carried by CheY.
B	10.2	
Bp	10.0	
R	10.0	
Z	10.0	

Table 10: Initial amounts of components in the chemotaxis network. The volume of the cell used in the simulations was $1.208 \cdot 10^{-15}$ dm^3.

Species	*Conc* (μM)	*Copy number/cell*
B	0.57	
Bp	1.7	
Y	15.5	
Yp	2.5	
Z	14.15	
TWA		$2.5 \cdot 10^3$
Flagella		8

A.3.4 CheZ Oligomerisation

Table 11: CheZ oligomerisation and phosphatase reactions. The unit of the rate constants is either s^{-1} or $M^{-1}s^{-1}$ depending on whether the reaction is unimolecular or bimolecular.

Reaction	k_f	k_b	*Notes*
Complex formations			
Z+Y $\rightleftharpoons$ ZY	$1 \cdot 10^4$	6.3	$K_D = 6.3 \cdot 10^{-4}$M [7]
Z+Yp $\rightleftharpoons$ ZYp	$4 \cdot 10^7$	$4.5 \cdot 10^2$	
ZYp + ZYp $\rightleftharpoons$ 2Z2Yp	$3 \cdot 10^8$	0.8	
2Z2Yp + ZYp $\rightleftharpoons$ 3Z3Yp	$3 \cdot 10^8$	0.3	
2ZYp + Yp $\rightarrow$ 2Z2Yp	$8 \cdot 10^6$		
3Z2Yp + Yp $\rightarrow$ 3Z3Yp	$1 \cdot 10^9$		
2ZYp $\rightarrow$ ZYp + Z	2		
3Z2Yp $\rightarrow$ 2Z2Yp +Z	10		
Dephosphorylation			
3Z3Yp $\rightarrow$ 3Z2Yp	$2 \cdot 10^2$		
2Z2Yp $\rightarrow$ 2ZYp	0.5		
ZYp $\rightarrow$ Z + Y	0.02		
Spontaneous phosphorylation and dephosphorylation			
Y $\rightleftharpoons$ Yp	0.214	$3.5 \cdot 10^{-2}$	$[PoAc] = 9\,\text{mM}$ [8]
	$3.14 \cdot 10^{-3}$	$3.5 \cdot 10^{-2}$	$[PoAc]_{cell} = 10\,\mu\text{M}$

A.3.5 The Adaptive Signaling Network

Receptor Complex Conformation

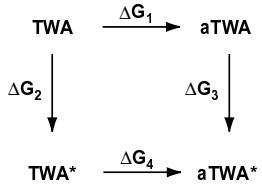
The probabilities of a receptor complex being in an inactive or an active conformation has been estimated in [26]. Knowing from experiments that the ratio of activity for receptors with 0:2:4 methyl groups is 0.0:0.489:1.0 [9] the intervening values has been derived by linear interpolation.

Table 12: Probability of receptor complexes being in active conformation. p_u=Pr(active, unoccupied), p_o=Pr(active, occupied)

Receptor	p_u	*Receptor*	p_o
TWA	0.000	aTWA	0.000
TmWA	0.246	aTmWA	0.015
T2mWA	0.489	aT2mWA	0.029
T3mWA	0.744	aT3mWA	0.515
T4mWA	1.000	aT4mWA	1.000

Ligand Receptor Interaction

The relative strength of aspartate binding to the Tar receptor in different methylation states were experimentally measured and reported in [9]. Morton-Firth used linear interpolation and a wild-type K_D of 1.2 μM to determine the K_Ds in the left column of table 13. The K_D:s for the active conformations of the receptor complex can be determined from thermodynamic considerations of the following reaction network



noting that for conservation of energy the relation

$$\Delta G_1 + \Delta G_3 = \Delta G_2 + \Delta G_4 \tag{13}$$

must hold. For the transitions in the reaction scheme either of two relations can be applied,

$$\Delta G_{1,4} = RT \ln K_D \tag{14}$$

or

$$\Delta G_{2,3} = RT \ln \frac{1-p}{p} \tag{15}$$

Combining equation (13), (14) and (15) gives (for each methylation state of the receptor complex)

$$K_D^* = \frac{p_o(1-p_u)}{p_u(1-p_o)} K_D \tag{16}$$

Table 13: Dissociation constants for aspartate binding to the Tar receptor complex.

Receptor	K_D (μM)	K_D^* (μM)	*Notes*
TWA	0.6	-	only inactive conformation
TmWA	0.9	19	
T2mWA	1.2	38	
T3mWA	7.4	7.4	
T4mWA	-	4.2	only active conformation

For modeling we need the fractions of receptors that are inactive or active irrespective of whether they are occupied or not. To derive expressions for these fractions we define x_u, x_o, x_i and x_a as the number of Tar molecules that are

unoccupied, occupied, inactive and active. We denote the conditional probabilities that a receptor is in either state as p_u and p_o, for the receptor being active given it is unoccupied or occupied, and p_i and p_a, for the receptor being occupied given it is inactive or active.

Now write down the number of molecules in each state

$$\text{inactive unoccupied:} \quad (1-p_i)x_i = (1-p_u)x_u \tag{17}$$

$$\text{active unoccupied:} \quad (1-p_a)x_a = p_u x_u \tag{18}$$

$$\text{inactive occupied:} \quad p_i x_i = (1-p_o)x_o \tag{19}$$

$$\text{active occupied:} \quad p_a x_a = p_o x_o \tag{20}$$

Combining equation (19) and (20)

$$\frac{x_a}{x_i} = \frac{p_i p_o}{p_a(1-p_o)} = K \tag{21}$$

and noting that the total number of molecules $x_{tot} = x_i + x_a$ allows us to form the desired quotients, describing the fraction of receptor molecules that are active or inactive

$$f_a = \frac{x_a}{x_{tot}} = \frac{K}{1+K} \tag{22}$$

$$f_i = 1 - f_a = \frac{1}{1+K} \tag{23}$$

Also the mixed fractions are easy to derive from these two fractions just multiplying with any of the four probabilities defined above. p_i and p_a are straightforward to derive from the K_D and K_D^* values listed, and will of course be a function of the ligand concentration.

The Adaptive Reaction Network

The full set of reactions in the adaptive reaction network are listed in table 14. Compared to the first model of the phosphorylation cascade it should be noted that the binding of CheY and CheB to unphosphorylated CheA has been added. These complexes are assumed to have the same probability of switching to an active conformation as has the pure TWA at that specific level of methylation.

Table 14: Reactions and reaction rates in adaptive model. The unit of the rate constants is either s^{-1} or $M^{-1}s^{-1}$ depending on whether the reaction is unimolecular or bimolecular. $j = 0..3$, $k = 1..4$, $l = 0..4$

Reaction	k_f	k_b	*Notes*
Ligand binding			
TWlA + a $\rightleftharpoons$ aTlWA			Implicit in parameters
TWkBpA* + a $\rightleftharpoons$ aTkBpWA*			Implicit in parameters
TWjRA + a $\rightleftharpoons$ aTjRWA			Implicit in parameters
CheA autophosphorylation			
TkmWA* $\rightarrow$ TkmWAp*	$f_a \cdot 15.5$		[26]
TkmBpWA* $\rightarrow$ TkmBpWAp*	15.5		
CheB - CheA interactions			
TlmWA + B $\rightleftharpoons$ TlmWAB	$1 \cdot 10^6$	3.2	$K_D = 3.2\,\mu$M [21]
TjmRWA + B $\rightleftharpoons$ TjmRWAB			
TkmBpWA* + B $\rightleftharpoons$ TkmBpWAB*			
CheY - CheA interactions			
TlmWA + Y $\rightleftharpoons$ TlmWAY	$1 \cdot 10^6$	2	$K_D = 2\,\mu$M
TjmRWA + Y $\rightleftharpoons$ TjmRWAY			
TkmBpWA* + Y $\rightleftharpoons$ TkmBpWAY*			
CheB - CheAp interactions			
TkmWAp* + B $\rightleftharpoons$ TkmWAp*B	$1 \cdot 10^6$	3.2	M-F uses unidirectional binding with higher rate than what diffusion limiting poses. I prefer an equilibrium with diffusion limited k_{on} and $K_D = 3.2\,\mu$M. [21]
TkmBpWAp* + B $\rightleftharpoons$ TkmBpWAp*B			
CheY - CheAp interactions			
TkmWAp* + Y $\rightleftharpoons$ TkmWAp*Y	$1 \cdot 10^6$	1	M-F uses unidirectional binding with a higher rate than what diffusion limiting poses. I prefer an equilibrium with diffusion limited k_{on} and $K_m = 6.5\,\mu$M ([31]). K_D adjusted to $1\,\mu$M (considering values for interactions between non-phosphorylated CheA).
TkmBpWAp* + Y $\rightleftharpoons$ TkmBpWAp*Y			

Table 14: continued

Reaction	k_f	k_b	*Notes*
Phosphotransfer			
T*k*mWApY* ⇌ T*k*mWAYp*	650	$f_a \cdot 10$	[31] k_b derived by optimisation.
T*k*mBpWApY* ⇌ T*k*mBpWAYp*			
T*k*mWApB* ⇌ T*k*mWABp*			
T*k*mBpWApB* ⇌ T*k*mBpWABp*			
CheBp - CheA interactions			
T*l*mWA + Bp ⇌ T*l*mWABp	$1 \cdot 10^6$	3.2	No data on K_D available so assumed to be the same as for CheB.
T*k*mBpWA* + Bp ⇌ T*k*mBpWABp*			
T*j*mRWA + Bp ⇌ T*j*mRWABp			
CheYp - CheA interactions			
T*l*mWA + Yp ⇌ T*l*mWAYp	$1 \cdot 10^6$	8	Affinity approximately six-fold lower than for the unphosphorylated CheY in the presence of Mg^{2+} ([21]). Just made a guess from the fact that *E coli* contains high levels of Mg^{2+} ions.
T*k*mBpWA* + Yp ⇌ T*k*mBpWAYp*			
T*j*mRWA + Yp ⇌ T*j*mRWAYp			
CheBp - Tar interactions			
T*k*mWA + Bp ⇌ T*k*mBpWA	$f_a \cdot 10^6$	1.25	$K_D = 1.25\,\mu$M [5], not an experimental value!
T*k*mWAB + Bp ⇌ T*k*mBpWAB			
T*k*mWAB + Bp ⇌ T*k*mBpWAB			
T*k*mWABp + Bp ⇌ T*k*mBpWABp			
T*k*mWAY + Bp ⇌ T*k*mBpWAY			
T*k*mWAYp + Bp ⇌ T*k*mBpWAYp			
Demethylation			
T*k*mBpWA* → T*(k-1)*mWA +Bp	0.11		Rate cited in [26]. Here only demethylation by the phosphorylated form of CheB is considered, since at least 70-
T*k*mBpWAB* → T*(k-1)*mWAB + Bp			
T*k*mBpWABp* → T*(k-1)*mWABp + Bp			

Table 14: continued

Reaction	k_f	k_b	*Notes*
T*k*mBpWAY* → T*(k-1)*mWAY + Bp			fold more active than CheB [6].
T*k*mBpWAYp* → T*(k-1)*mWAYp + Bp			
CheR - Tar interactions			
T*j*mWA + R ⇌ T*j*mRWA	$f_i \cdot 10^6$	0.2	$K_D = 0.2\,\mu$M [26]
T*j*mWA + R ⇌ T*j*mRWA	$f_i \cdot 10^6$	0.2	
Methylation			
T*j*mRWA → T*(j+1)*mWA + R	0.819		[26]
T*j*mRWAB → T*(j+1)*mWAB + R			
T*j*mRWABp → T*(j+1)*mWABp + R			
T*j*mRWAY → T*(j+1)*mWAY + R			
T*j*mRWAYp → T*(j+1)*mWAYp + R			
CheZ-mediated dephosphorylation See table 11.			
Phosphorylation reactions in cytosol			
Yp ⇌ Y	0.035	0.0031	k_f from [23]. k_b from [8], converted to unimolecular rate constant and scaled for $[PoAc]_{cell} = 10\,\mu$M.
Yp $\xrightarrow{Z}$ Y	$1 \cdot 10^6$		Diffusion limited bimolecular reaction assumed. [26]
Bp → B	0.35		[26]

Table 15: Initial amounts of components in the adaptive chemotaxis network. The volume of the cell used in the simulations was $1.208 \cdot 10^{-15} \mathrm{dm}^3$.

Species	*Conc* (μM)	*Copy number/cell*	*Notes*
B	0.57		
Bp	1.7		
R	0.235		
Z	4.1		Non-oligomerising model.
Z	0.264		Oligomerising model
Y	15.5		
Yp	2.5		
TWA		2500	