Research article

A model invalidation-based approach for elucidating biological signalling pathways, applied to the chemotaxis pathway in *R. sphaeroides*

Mark AJ Roberts[1,2,3], Elias August[1,3], Abdullah Hamadeh[1,3], Philip K Maini[3,4], Patrick E McSharry[3,5], Judith P Armitage[2,3] and Antonis Papachristodoulou*[1,3]

Address: [1]Control Group, Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK, [2]Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK, [3]Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK, [4]Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK and [5]Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK

E-mail: Mark AJ Roberts - mark.roberts@bioch.ox.ac.uk; Elias August - elias_august@hotmail.com; Abdullah Hamadeh - abdullah.hamadeh@eng.ox.ac.uk; Philip K Maini - maini@maths.ox.ac.uk; Patrick E McSharry - patrick.mcsharry@maths.ox.ac.uk; Judith P Armitage - judith.armitage@bioch.ox.ac.uk; Antonis Papachristodoulou* - antonis@eng.ox.ac.uk
*Corresponding author

Abstract

Background: Developing methods for understanding the connectivity of signalling pathways is a major challenge in biological research. For this purpose, mathematical models are routinely developed based on experimental observations, which also allow the prediction of the system behaviour under different experimental conditions. Often, however, the same experimental data can be represented by several competing network models.

Results: In this paper, we developed a novel mathematical model/experiment design cycle to help determine the probable network connectivity by iteratively invalidating models corresponding to competing signalling pathways. To do this, we systematically design experiments *in silico* that discriminate best between models of the competing signalling pathways. The method determines the inputs and parameter perturbations that will differentiate best between model outputs, corresponding to what can be measured/observed experimentally. We applied our method to the unknown connectivities in the chemotaxis pathway of the bacterium *Rhodobacter sphaeroides*. We first developed several models of *R. sphaeroides* chemotaxis corresponding to different signalling networks, all of which are biologically plausible. Parameters in these models were fitted so that they all represented wild type data equally well. The models were then compared to current mutant data and some were invalidated. To discriminate between the remaining models we used ideas from control systems theory to determine efficiently *in silico* an input profile that would result in the biggest difference in model outputs. However, when we applied this input to the models, we found it to be insufficient for discrimination *in silico*. Thus, to achieve better discrimination, we determined the best change in initial conditions (total protein concentrations) as well as the best change in the input profile. The designed experiments were then performed on live cells and the resulting data used to invalidate all but one of the remaining candidate models.

Conclusion: We successfully applied our method to chemotaxis in *R. sphaeroides* and the results from the experiments designed using this methodology allowed us to invalidate all but one of the proposed network models. The methodology we present is general and can be applied to a range of other biological networks.

Background

Understanding the connectivity of signalling pathways within organisms has always been an important challenge in biological research. One approach to address this is to study individual parts *in vitro* and look at protein localisation, homologies and co-expression in order to elucidate signalling pathway connectivity. Another approach is to use genetics and mutants to attempt to work out the pathway connectivity *in vivo*.

More recently, systems biology approaches have used quantitative measurements to develop mathematical models that can be used for understanding the properties of biological signalling pathways and their connectivity [1-3]. These models are usually a result of cyclic mathematical model/experiment design iterations, which aim to yield maximum information about the system under study [4,5]. It is well known, however, that models of comparable complexity corresponding to different pathway connectivities may fit experimental data equally well, leaving the researcher with the dilemma of which model is correct [6]. The question of which model is correct is actually impossible to answer as model validation is a misnomer. The related question of which models are invalid can be answered if appropriate data are available, i.e. the inability of a model to reproduce a data set renders it invalid. Applied in this way, model invalidation can be used to reduce the number of possible models [7], hence narrowing down the number of possible network connectivities.

Previous work in model invalidation emphasised the importance of using time-varying inputs to the system under study. One method of invalidation is to apply a dynamical input and try to maximise the difference in the phase shift of two competing deterministic models [8]. The disadvantage of this approach is that phase-shift is usually difficult to quantify, especially with noisy data. Another approach is presented in [9], where the authors develop a dynamic model-based controller and an input profile that drives the system output along a prescribed target trajectory. However, this approach requires the implementation of a controller in the laboratory which may prove difficult. Other approaches for model invalidation are presented in [5,10,11], and the references therein. These lack-of-fit methods are used to invalidate models in a statistical manner, but there can be problems with this approach as it relies on large data

sets and focuses on obtaining reliable parameter estimates rather than network connectivities. There is also the issue that a wide range of model parameters could give a very similar model output and these methods would have difficulty coping with this.

In this paper we present a new method for developing mathematical models of biological signalling networks, aiming to understand biological network structure. Given a set of experimental data, models corresponding to competing network connectivies are first constructed, all of which can explain wild type data equally well. Then, experiments which maximally discriminate between models corresponding to different networks are designed systematically. These experimental results are used to invalidate models of these networks, resulting in a cyclic process which aims to produce a mathematical model corresponding to a signalling pathway structure which explains all available wild type and mutant experimental data. Our method is applicable to biological pathways for which it is possible to experimentally modify the input profile and measure the output simultaneously. One such system is the chemotaxis signalling pathway, for which tethered cell experiments can be performed in a flow cell and used to measure the response of the system to dynamic ligand concentration profiles.

Chemotaxis is the biasing of movement towards regions of higher concentrations of beneficial or lower concentrations of toxic chemicals by altering the frequency of flagella switching [12]. The signalling pathway within *E. coli* is well understood and is a simple circuit with one feedback loop [13]. The receptor in the system is a <u>M</u>ethyl-accepting <u>C</u>hemotaxis <u>P</u>rotein (MCP) that senses ligands outside the cell. Associated with the MCP is the histidine protein kinase called CheA. Binding of certain ligands to MCP decreases the auto-phosphorylation rate of CheA. CheA can transfer phosphoryl groups to two possible response regulators, CheY and CheB. CheY-P interacts with motor binding sites of the multiple *E. coli* flagella motors causing a change in direction. The receptors are constantly methylated by the action of a methyltransferase CheR, while CheB-P acts as a methylesterase to demethylate the receptor, making it less responsive to ligand binding. This creates a feedback loop, allowing for adaptation. Adaptation allows *E. coli* to react to changes of the concentration gradient of

ligands and not to changes in concentrations *per se*. Finally, CheZ acts to dephosphorylate CheY and to terminate the signal.

Chemotaxis pathways in other species are less well characterised and often contain multiple homologues of the *E. coli* system and sometimes proteins not found in the *E. coli* signalling pathway, for example, CheD [13]. Chemotaxis pathways in other species may also have a different connectivity - for example in *S. meliloti*, only one of the two CheY homologues interacts with the motor, with the other CheY acting as a phosphate sink [13,14]. Another good example is the chemotaxis pathway in *R. sphaeroides*. This bacterium has three chemotaxis operons, two of which are expressed under normal laboratory conditions. Proteins expressed from these operons have previously been shown to localise to discrete signalling clusters, one at the poles of the cell, similar to *E. coli*, and one in the cytoplasm [15]. This localisation is thought to prevent cross talk between the two clusters, allowing them to potentially signal separately. Despite these data and corresponding data on the possible phosphotransfer patterns [16], the way the signal is transmitted and integrated between the chemotaxis clusters to control flagella activity is currently unknown [17]. This is a good example of the difficulty of inferring a network structure from homology alone. The current known connectivity and protein localisation is shown in Figure 1.

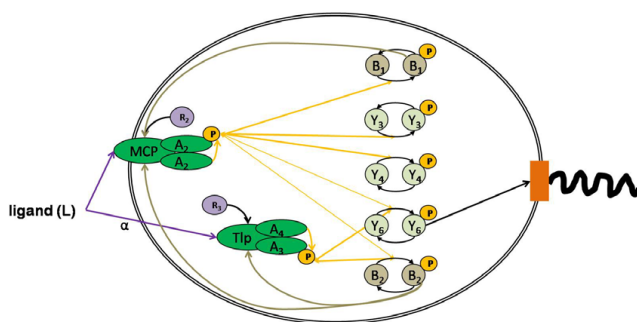In this paper, we apply our method of model invalidation to different possible and plausible connectivity

Figure 1
The chemotaxis pathway of *R. sphaeroides*. The currently known localisation and connectivity of the chemotaxis pathway in *R. sphaeroides* is shown. The localisation of the two signalling clusters, one at the pole and one in the cytoplasm was determined by GFP fusions and immunoflouresence [15]. The remaining connectivities, was determined be *in vitro* phosphotransfer measurements and by mutant data [16,21]. It is currently unknown what role CheY$_3$ and CheY$_4$ perform in the system.

structures in the chemotaxis pathway of *R. sphaeroides*. In particular, we use the model invalidation/experiment design cycle outlined above, in order to shed light on the signalling pathway in *R. sphaeroides*.

Results
A method for discriminating between competing network models
A general method for designing experiments so as to render the outputs of candidate models as different as possible is described. These experiments can then be performed in the laboratory and, when compared to the model predictions, allow the invalidation of some of the candidate models, even when the experimental data set is noisy.

Our method involves the development of ODE models corresponding to different signalling pathway connectivities, all of which can explain current wild type experimental data equally well. The models have in common all currently known interactions and differ in that each model represents a new speculative pathway connectivity. Some parameters in these models are known and some others are unknown. Thus in principle, one can develop a "set" of models with uncertain parameters for a particular signalling structure, each member of this set representing the wild type experimental data equally well. We would want to discriminate between these "sets" of models (which represent signalling pathway structures) but since this is a very hard problem, our method uses a nominal model from each set and then designs an experiment in order to discriminate between these nominal models. Once the discriminatory experiments between the nominal models are designed *in silico*, we *a posteriori* assess the discriminatory property of the stimulus by simulating the behaviour for many, randomly chosen, models within these sets in order to see whether the outputs from the two model sets remain distinguishable.

The discriminatory experiments were designed by initially determining the input profile that maximises the magnitude of the squared output difference between models over time. This is typically an optimal control problem, which is often laborious to solve and can result in an input profile that is difficult to realize experimentally [18]. In order to design the input profile in a numerically efficient manner, we used the following result from linear systems theory: for a linear time-invariant system, the input that produces the largest energy output given an input of fixed energy is a (truncated) sinusoid of a particular frequency [19]. The frequency can be obtained using a Bode plot (see Methods section), a tool which is often used in control

and systems theory [20]. Thus, we first linearised the models and then determined this particular frequency of a sinusoidal input corresponding to the error system, i.e. the difference between the two models.

If the above method is insufficient to discriminate between the models then a further method of 'mutating' the two models in biological terms or changing parameters/initial conditions (e.g. alter the initial protein concentrations) in engineering terms is applied to achieve discrimination. These changes can be tested *in silico* to determine those which will discriminate best between the models under test, before undertaking experiments. The exact nature of the perturbation that can be performed will vary with the system being investigated, but could include altering protein levels by knockout, knockdown (e.g. RNAi in eukaryotes), protein over expression, etc. Thus the space of perturbation which can be searched will be defined by what is possible to implement experimentally in the biological system under investigation.

The designed experiments can then be implemented and the data obtained from such laboratory experiments used to help us differentiate between the models under study, by invalidation. In the above procedure we used deterministic models and a worst-case input design procedure, therefore even if the data are noisy, we expected that we would still be able to invalidate some of these models.

Determining pathway connectivity in
R. sphaeroides chemotaxis
We applied our model invalidation method to elucidate unknown interconnections within the *R. sphaeroides* chemotactic signalling pathway. *In vitro* phosphorylation measurements have determined some of the internal connectivity and genetic work has shown that only $CheY_6$ and either $CheY_3$ or $CheY_4$ are required for motor switching and that $CheY_6$ can bind to the motor [21] (Figure 1). Whilst it has been shown *in vitro* that all CheYs $_{1-6}$ can interact with FliM, the rotor switch [22], it is currently unknown whether $CheY_3$ and $CheY_4$ proteins cause flagella motor switching or whether some CheYs have an effect on other parts of the signalling pathway to influence the motor indirectly, for example, by acting as a phosphate sink.

Model creation and parameter estimation
We created a number of different models representing various plausible $CheY_3$, $CheY_4$ connectivities within the chemotaxis system (Figure 2; Methods; SMBL in Additional File 1). Each of these models contains all the currently known connectivities, shown in Figure 1, but differs in which unknown connectivity is considered. To allow each of our models to represent current, wild type
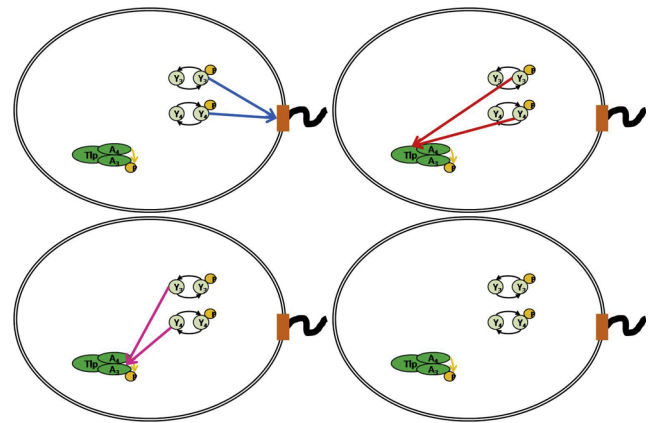

Figure 2
The models of the *R. sphaeroides* chemotaxis system.
The different connecitvities modelled for $CheY_3$ and $CheY_4$ are shown in this diagram. The coloured lines represent three models presented in this paper. In the blue model $CheY_3$-P and $CheY_4$-P bind to FliM in the motor co-operatively with $CheY_6$-P. In the red model $CheY_3$ and $CheY_4$ form a connection with the cytoplasmic cluster and act antagonistically to ligand in altering the kinase activity of $CheA_4$; therefore, the increase in $CheY_{3/4}$-P acts to increase the rate of $CheA_3$ phosphorlyation by $CheA_4$. In the magenta model $CheY_{3/4}$ interact with $CheA_3$ preventing $CheY_6$ binding and hence phosphotransfer. The fourth model corresponds to a model, in which $CheY_3$ and $CheY_4$ act as phosphate sinks only and hence do not interact with anything. This is referred to as the grey model in this paper. It should be noted that the other three models also contain this interconnection and that the difference in the grey model is that only this connection exists.

observations equally well, we fitted the unknown parameters of each model (K_{1-3}) such that the error between model prediction and wild type data is small (Figure 3). These parameters represented the activation of the receptor on the CheA (K_1), and the effect of and rate of methylation on the receptor ($K_{2,3}$). These values are currently unknown biologically. The fitting results in the models with different connectivities having different such parameter values.

We obtained the following values for the parameters for the different models (the units for values of K_1 are in μM, and in $\frac{1}{sec}$ for K_2 and K_3):

- blue: $K_1 = 1$, $K_2 = 16.5$ and $K_3 = 0.015$
- red: $K_1 = 25$, $K_2 = 3250$ and $K_3 = 0.075$
- grey: $K_1 = 1$, $K_2 = 97.5$ and $K_3 = 0.0023$
- magenta: $K_1 = 1$, $K_2 = 0.4$ and $K_3 = 0.011$

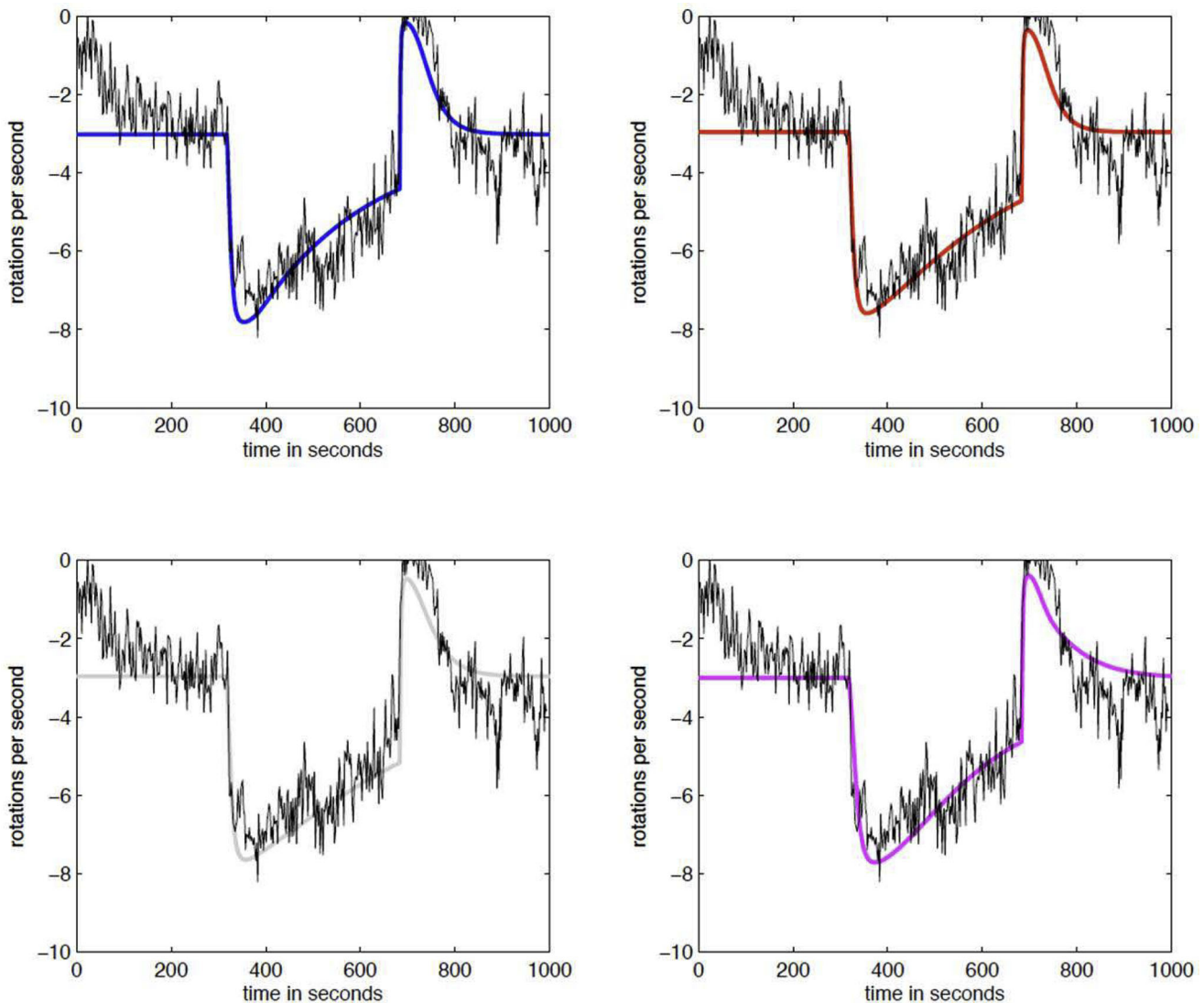We assumed that these parameters are the same for the polar and cytoplasic clusters, so $K_i = \tilde{K}_i$.

Figure 3
Fitting model parameters. The three parameters K_1 - K_3 are unknown and thus the models were fitted to wild type data in order to obtain them. WS8N (wild type) was tethered and recorded for 5 mins without, 5 mins with and 5 mins without 100 μM propionate. The coloured lines represent the best fit of each model to this data by varying the parameters K_{1-3} using the methods described in Methods section.

We used the Pearson product-moment correlation coefficient as a measure of correlation between model prediction and data. For $200 \leq t \leq 800$ seconds, we found a good correlation between data and the predictions made by the models. For example, we obtained a coefficient of 0.9466 for the blue model and of 0.9544 for the red model.

To ensure our models can fit all current experimental data we then compared the output of all models to data previously determined for the deletion of various genes within *R. sphaeroides* [21]. One possible plausible connectivity contains CheY$_3$ and CheY$_4$ acting only as phosphate sinks, similar to the roles of the multiple CheY's in *S. meliloti* [13,14]. However, the model representing this connectivity (grey model) was unable to fit mutant experimental data as it still showed chemotaxis in a CheY$_3$Y$_4$ deletion state (ΔCheY$_3$Y$_4$). Another possible connectivity where CheY$_3$ and CheY$_4$ can bind to CheA$_3$ preventing CheY$_6$ binding and hence phosphorylation was also considered (magenta model). However, this connectivity was unable to fit experimental data as it remained chemotactic in a ΔCheY$_3$Y$_4$ state (Figure 4A). To strengthen our conclusion that both
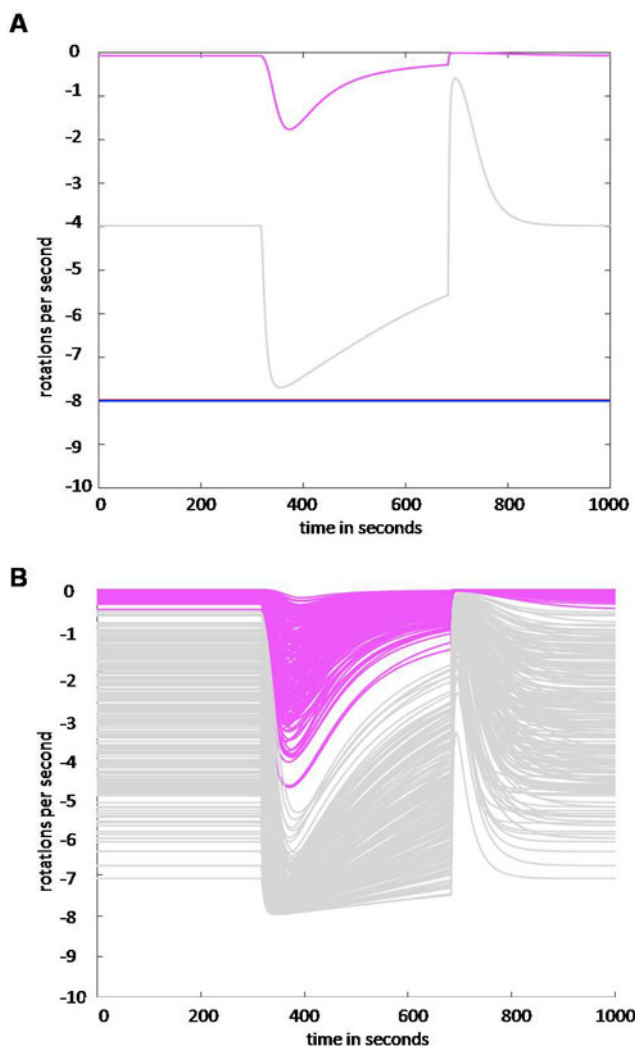
Figure 4
Simulation of ΔCheY$_3$ Y$_4$. A] Models representing the mutants lacking CheY$_3$ and CheY$_4$ were simulated. The grey and the magenta model are chemotactic (they still respond to stimulus) under these conditions and the steady state of magenta model is always close to zero. These results do not correspond to experimental observation. B] The robustness of our invalidation of the grey and magenta models was tested by running 200 simulations of the model lacking CheY$_3$ and CheY$_4$, allowing parameters k_1 - k_{14} to vary by ± 50%. None of these changes produced a non-chemotactic model.

Table 1: Reaction rates used in the *R. sphaeroides* chemotaxis models

reaction	parameter(s)	value(s)
(R$_1$) $A_2 \rightarrow A_{2p}$	k_1	0.03 s^{-1}
(R$_2$) $A_{2p+} B_1 \rightleftharpoons A_2 + B_{1p}$	k_2^+, k_2^-	0.035$(\mu Ms)^{-1}$, 0.01$(\mu Ms)^{-1}$
(R$_3$) $A_{2p} + Y_3 \rightleftharpoons A_2 + Y_{3p}$	k_3^+, k_3^-	0.065 $(\mu Ms)^{-1}$, 0
(R$_4$) $A_{2p} + Y_4 \rightleftharpoons A_2 + Y_{4p}$	k_4^+, k_4^-	0.004$(\mu Ms)^{-1}$, 0
(R$_5$) $A_{2p} + Y_6 \rightleftharpoons A_2 + Y_{6p}$	k_5^+, k_5^-	0.0006$(\mu Ms)^{-1}$, 0
(R$_6$) $A_{2p} + B_2 \rightleftharpoons A_2 + B_{2p}$	k_6^+, k_6^-	0.0035$(\mu Ms)^{-1}$, 0.01$(\mu Ms)^{-1}$
(R$_7$) $B_{1p} \rightarrow B_1$	k_7	0
(R$_8$) $Y_{3p} \rightarrow Y_3$	K_8	0.08s^{-1}
(R$_9$) $Y_{4p} \rightarrow Y_4$	K_9	0.02s^{-1}
(R$_{10}$) $Y_{6p} \rightarrow Y_6$	K_{10}	0.1s^{-1}
(R$_{11}$) $B_{2p} \rightarrow B_2$	K_{11}	0.015s^{-1}
(R$_{12}$) $A_3A_{4p} + Y_6 \rightleftharpoons A_3A_{4+} Y_{6p}$	k_{12}^+, k_{12}^-	0.1 $(\mu Ms)^{-1}$, 0
(R$_{13}$) $A_3A_{4p} + B_2 \rightleftharpoons A_3A_4 + B_{2p}$	k_{13}^+, k_{13}^-	0.006$(\mu Ms)^{-1}$, 0.07$(\mu Ms)^{-1}$
(R$_{14}$) $A_3A_4 \rightarrow A_3A_{4p}$	k_{14}	0.02s^{-1}

The reaction rates used for the models determined by fitting phosphotransfer *in vitro* experiments [16,31]. For some of these reaction rates the best fit is zero suggesting that these reactions do not occur *in vitro*.

connectivities are invalid, we ran 200 simulations of the models lacking CheY$_3$ and CheY$_4$, in which we allowed parameters k_1 to k_{14} (Table 1) to vary by ± 50%. We observed that, even when we allowed for greatly different parameter configurations, the connecitvities modelled by the magenta and grey models are still chemotactic, hence our invalidation is robust (Figure 4B). As opposed to the grey and magenta model, the red and the blue model cannot be invalidated by above deletion data.

We then designed an experiment to invalidate one or both of the remaining two models following the model-invalidation cyclic procedure described in the background section. As mentioned previously, we chose a sinusoidal input with a particular frequency to help discriminate between the models. Using a Bode plot showing the response of the difference of the outputs to a sinusoidal input with fixed amplitude, we determined an input frequency (in terms of frequency of 100 μM amplitude ligand application) in order to discriminate best between the remaining two models (Figure 5A). We chose an input period of four minutes, which was both close to the optimal in terms of discrimination and easy to implement experimentally. We then ran simulations of the time evolution of the models with alternating step inputs mimicking a sinusoidal input of this period (as this mirrors what can be implemented experimentally) to ensure that the difference in the outputs was sufficiently large to be accurately measured experimentally (Figure 5B).
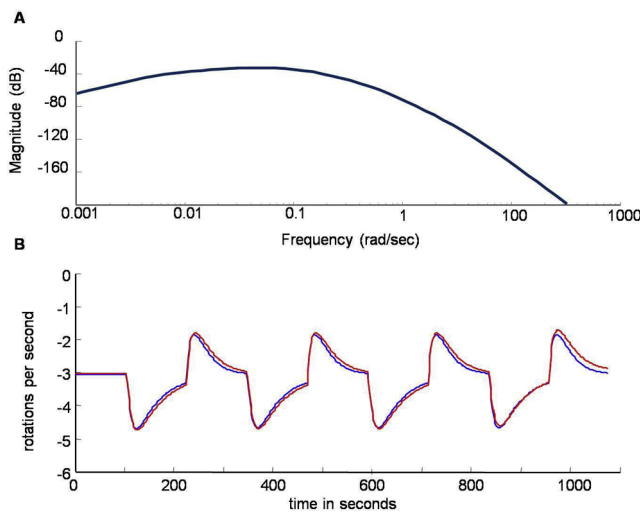
Figure 5
Model discrimination by optimal frequency. A] Bode plot determining the optimal input frequency for model discrimination. The Bode magnitude plot shows the output amplitude gain, g_ω, versus frequency ω (in decibel; that is in $20 \log(g_\omega)$). Here, the gain corresponds to the ratio of the difference between the outputs of the blue and the red model, and the input amplitude (sinusoidal input); in other words, the maximum denotes at which frequency one can expect to see the largest difference between the models. The aim was to determine the optimal input frequency to apply to the system. This was determined to have a period of 4 minutes. B] Outputs of the simulation for the behaviour of the models with this optimal frequency applied as the input frequency demonstrating that the best input perturbation alone is not enough to discriminate between the two models. The coloured lines represent the models defined in Figure 2.

We observed that the difference between the outputs of the two models under consideration was insufficient to allow discrimination between the models experimentally using this discrimination technique. Therefore, we sought experimental perturbations which resulted in a larger difference between the outputs of the different models. Possible experimental perturbations in this system involve deletions of one or more components, over expression or under expression of a protein component and growth of the bacteria in different growth conditions; the latter results in large-scale changes of the expression of the chemotaxis operons and hence protein concentrations.

Before experimental implementation, we tested the possible perturbations *in silico*. The over expression of $CheY_4$ was chosen as this resulted in a large difference in both the models' steady states and their dynamical behaviour (Figure 6A). Moreover, in order to ensure that this discrimination is not influenced by errors in determining parameter values, we confirmed that our
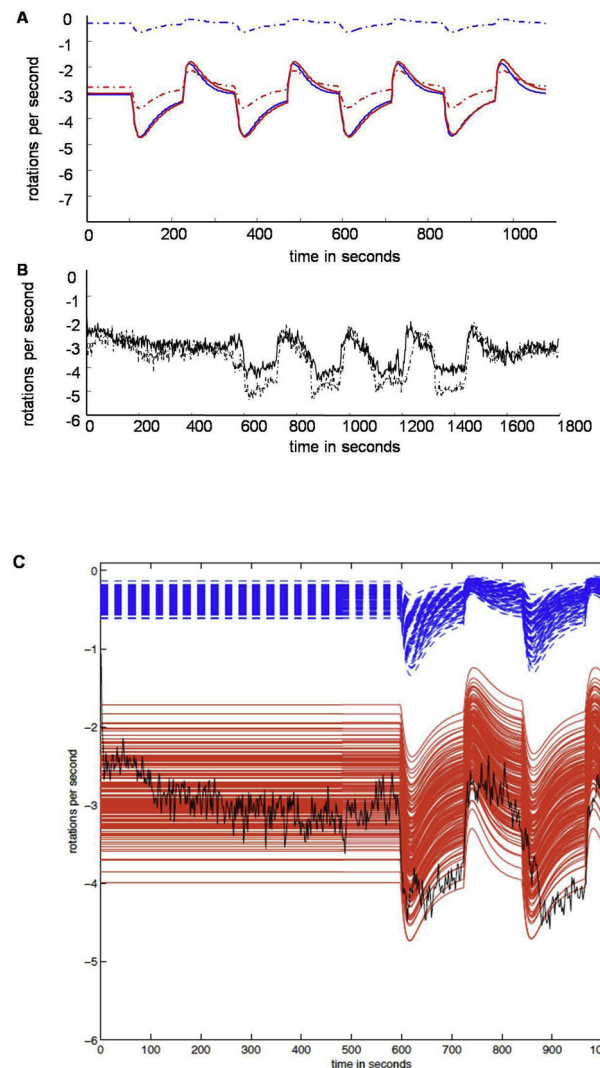

Figure 6
Discrimination by optimal initial conditions. A] Outputs of the simulations of the models with wild type (continuous lines) or levels of $CheY_4$ (total) increased to 5-times normal levels at the start of the simulation (dotted lines). The coloured lines represent the models defined in Figure 1. B] Experimental outputs from tethered cell experiments, wild type output (continuous line) and $CheY_4$ over-expressed in the cells at an average of 5-times the concentration found in wild type cells (dotted line). C] Robust invalidation of the blue model. To ensure that our invalidation was robust we ran 500 simulations assuming that $CheY_4$ is 5-fold over-expressed. We vary parameters k_1 - k_{14} by ± 15% and still observe the same clear difference between the behaviour of the blue and of the red model.

findings are relatively robust to parameter changes. We allowed all parameters to vary by ± 15% and still observed the same clear difference in the behaviour of the models (Figure 6C).

We then used IPTG to experimentally induce the over expression of $CheY_4$ from an expression plasmid. The level of $CheY_4$ over expression as a population average was verified using semi-quantitative western blotting (data not shown). We observed that when $CheY_4$ is over-expressed to five times the wild type level the tethered cell trace was similar to that of wild type cells (Figure 6B); this suggested that the model, in which the $CheY_6$ and $CheY_3$ or $CheY_4$ bind to the motor co-operatively (blue model) and do not have other interactions, is invalid, as it is unable to represent our new experimental data (Figure 6B). Since the red model, where there is an interconnection between $CheY_3$ and $CheY_4$ and the cytoplasmic cluster altering the $CheA_4$ kinase activity, can represent this data, it is the only model we cannot invalidate (Figure 6B).

Discussion

We have developed a methodology for elucidating biological signalling pathways which we applied to understanding the chemotactic signalling pathway in *R. sphaeroides*. Our method differs from many other methods currently in use in that is based on manipulation and observation of the entire biological system under *in vivo* conditions and, importantly, offers a systematic approach to model invalidation based on cyclic model development/experiment design iterations. In particular, it does not aim at designing experiments for the refinement of parameter values, but rather at identifying possible interconnection structures. By doing so, our method considers a model representing a specific structure that it aims to invalidate.

We applied our method to a real biological system and successfully managed to invalidate some potential signalling pathway network models. Thus, this method helped to more rationally consider the interconnectivity of the chemotactic signalling pathway in *R. sphaeroides* through relatively simple input-output experiments *in vivo* and helps to design rational future experiments.

The method is applicable to other biological systems but requires an experimental setup where it is possible to control the input and measure the output simultaneously. Even with this limitation the method could be used for both pathway determination and parameter determination, where multiple models with different parameters could be systematically invalidated. Our method could potentially be used to help annotate and understand signalling pathways in non-model organisms, using the information from model organisms as a guide for the first model generation step. These could then be tested and models invalidated. A parameter search could also be performed by creating models

which deviate from the model organism data and then, through model invalidation, determining which parameter sets are able to fit the experimental data. In all these cases though, the input of the system must be under experimental control and the output easily measured.

A potential limitation would be the case in which two models from different network structures produce outputs that are indistinguishable under all potential experimental conditions. The problem in this case could lie with the particular properties of the network and their discriminatory nature - i.e., whether such differences in structure are identifiable from input-output experiments [23]. Performing all calculations *in silico* before any experiments are undertaken is important in saving experimental time ensuring that experiments are only performed when the results will have discriminating property between the models under test.

In the system studied in this paper the modelling and experiment design was done on average cell traces for *R. sphaeroides* responses. The reason for this is that single cell behaviour is noisy, and in any case the model parameters we have available at the start of the procedure correspond to the average system behaviour. The total protein concentrations, for example, are the population average determined for these growth conditions. Thus the model output is for the 'average' cell hence we compared the output to the average of the tethered cell data.

When constructing the models we used *in vitro* rates as it is very difficult, if not impossible to measure these rates *in vivo*. The multiple homologues in *R. sphaeroides* would prevent a FRET assay such as the one applied to *E. coli* being used. Thus the parameters are the best estimate of the real value. To allow for this uncertainty we considered models where these parameters were varied to ensure robust invalidation. Also for simplicity in the modelling we did not consider the increased concentration of the CheAs at specific points in the cell due to clustering nor the effects of clustering. This is because the exact concentration of CheA in specific regions and the effects of clustering are currently impossible to measure experimentally. The fitted parameter K_1, which relates the activation of the ligand on the receptor, includes the effect of clustering and as such it is accounted for in our models. To ensure that this is robust we also allowed the concentration of the CheA proteins to vary sufficiently and our invalidation conclusion was still correct.

Interestingly, the connectivity that best fits the experimental data suggests that in *R. sphaeroides* $CheY_3$-P and/or $CheY_4$-P do not bind co-operatively with $CheY_6$ to

FliM. The model we have been unable to invalidate suggests that $CheY_3$-P and $CheY_4$-P form a link between the polar and cytoplasmic signalling clusters, helping transmit the signal between the two clusters. We have through modelling and comparing to experimental data invalidated a model representing the previously held hypothesis that $CheY_3$ and $CheY_4$ act only as phosphate sinks for the system. We have also invalidated a model with strict co-operative binding of CheY's at the motor and as such our technique adds to the body of knowledge on *R. sphaeroides* chemotaxis.

Conclusion
In this paper we have developed a control engineering method for elucidating biological signalling pathways and applied it to a real system. This method is based on multiple model creation and subsequent invalidation using *in silico* designed experiments. This is a general method that can be applied to other biological pathways where it is possible to control the input and measure the output in simultaneously. We used the method to invalidate all but one model for the chemotaxis signalling pathway in *R. sphaeroides* and in doing so have invalidated models of certain possible connectivities.

Methods
Strains and growth conditions
Bacterial strains and plasmids are listed in Table 2. *R. sphaeroides* strains were grown aerobically in succinate medium [24] at 30°C with shaking. When appropriate, the antibiotics nalidixic acid and kanamycin were used at 25 μgml^{-1}.

Protein over expression
pIND4-Y4 was transformed into S-17 and conjugated into *R. sphaeroides* as described previously [25]. IPTG was added to cell culture at 100 μM and the cells incubated for ~16 hours at 30°C until OD_{700} is 0.6, where the cells were then used for tethered cell analysis.

Tethered cell analysis
R. sphaeroides was grown to an OD_{700} of 0.6 and then 4 × 1 ml of cells were harvested. 3 × 1 ml were saved for western blot analysis. The remaining 1 ml was pelted and resuspended in tethering buffer (10 mM Na-HEPES pH 7.2 containing chloramphenicol at 50 μg/ml) and incubated at 30°C for 30 mins.

The cells were then tethered in a humidity chamber by incubation of 10 μl of cell suspension on a coverslip with 1 μl of anti-flagellar antibody for 30 mins.

The coverslip was then loaded onto a flow chamber and tethering buffer passed through for 5 min to remove free cells. After this period the tethered cells were observed under phase contrast at 1000 × magnification. Real time recordings were made on videotape. Tethering buffer with and without Propionate (sodium salt) was passed through the chamber at a rate of 0.09 ml/min.

The video recordings were analysed with the Hobson Bactracker (Hobson Tracking Systems) using the program Arot7. The rotation rate of the cells was measured by detecting the position of the cell every 50 ms. The data obtained were smoothed (100 points), averaged (for as many cells as available - at least 20 per graph) and plotted.

Western blotting
In order to determine protein concentrations semi quantitative western blots were employed as described previously [26].

Modelling the chemotaxis pathway in *R. sphaeroides*
Our model is split into three modules: sensing, transduction and actuation.

Sensing
We assumed the same underlying mechanisms for transmembrane (MCP) and the cytoplasmic (Tlp) receptors. The parameters of the Tlp cluster are in

Table 2: Strains and plasmids

Strain/Plasmid	Characteristics	Source
R. sphaeroides WS8N	Spontaneous nalidixic acid resistant mutant of wild type WS8	[34]
R. sphaeroides JPA421	WS8N with the *cheY₄* gene deleted by genomic replacement	[35]
E. coli S17-1 λpir	A strain capable of mobilising pAE and pAY4 into *R. sphaeroides*, Sm^R	[36]
Plasmids		
pIND4	Over expression vector with a *lac* inducible promoter, capable of replication in *R. sphaeroides*	[25]
pIND4-Y4	pAE containing *cheY₄*	[29]

Strains and Plasmids used in this study.

brackets and labelled with a tilde superscript. In the following, we list the assumptions for our model, which we adopted from the *E. coli* literature [1,6,27]:

(i) Receptors can be in different states of methylation. For simplicity, we assume that receptors are either methylated or not.

(ii) Only methylated receptors, R^m ($\tilde{R}^m$), can be in an active state, R^a ($\tilde{R}^a$).

(iii) Auto-phosphorylation of CheA$_2$-P (phosphorylation of CheA$_3$ by CheA$_4$-P) occurs only when the MCP (Tlp) receptor is active.

(iv) Binding of the ligand to a receptor inhibits its activity.

(v) CheB$_1$-P (CheB$_2$-P) binds only to active receptors in order to demethylate them.

(vi) CheR$_2$ (CheR$_3$) binds only to inactive receptors, R^i ($\tilde{R}^i$).

(vii) The number of CheR$_2$ (CheR$_3$) is constant.

We incorporated assumptions (iv), (v) and (vii) into our model as follows:

• We let the number of active receptors depend reciprocally on ligand concentrations L ($\tilde{L}$) - see below for details.

• We represented the action of CheB$_1$-P (CheB$_2$-P) through the following term: $K_2[R^a][B_{1p}]$ $\left(\tilde{K}_2[\tilde{R}^a][B_{2p}]\right)$.

• We represented the action of CheR$_2$ (CheR$_3$) by the constant reaction rate K_3 ($\tilde{K}_3$).

Using these assumptions, we represented the sensing dynamics by:

$$R^a = \frac{[R^m]}{(K_1+L)},\qquad(1)$$

$$[\dot{R}^m] = -K_2[R^a][B_{1p}] + K_3 R^i = -K_2[R^a]\left([B_{1p}] + \frac{K_3}{K_2}\right) + K_3 R^t,\qquad(2)$$

$$\tilde{R}^a = \frac{[\tilde{R}^m]}{(\tilde{K}_1+\tilde{L})} \text{ for all models but for the red model, for which:}\qquad(3)$$

$$\tilde{R}^a = \frac{[\tilde{R}^m]([Y_{3p}]+[Y_{4p}])}{\tilde{K}_1\left([Y_{3p}]+[Y_{4p}]\right)+\tilde{L}},\qquad(4)$$

$$[\dot{\tilde{R}}^m] = -\tilde{K}_2[\tilde{R}^a]\left([B_{2p}] + \frac{\tilde{K}_3}{\tilde{K}_2}\right) + \tilde{K}_3\tilde{R}^t,\qquad(5)$$

where R^t ($\tilde{R}^t$) is the total number of receptors, $R^i = R^t - R^a$ ($\tilde{R}^i = \tilde{R}^t - \tilde{R}^a$), the K_i s are unknown and

$$\tilde{L} = \alpha L + \varepsilon,\qquad(6)$$

where ε is a small constant representing disturbances at the input of the cytoplasmic sensing cluster, with nominal amplitude 0.001. We let $\alpha = 0.1$ and $R^t = \tilde{R}^t = \mu$M.

We assumed that the cytoplasmic sensing cluster senses extracellular ligand concentrations indirectly; for example, $\tilde{L}$ could be internalised attractants, a byproduct of the internalisation process or a metabolic response to it. For simplicity, we assumed an affine relationship between L and $\tilde{L}$. Note that for the red model $\tilde{R}^a$ is a function of CheY$_3$-P and CheY$_4$-P concentration levels, reflecting their effect on the activity of the cytoplasmic sensing cluster.

Transduction

Table 3 shows the copy numbers of the major proteins involved in the chemotaxis pathway of *R. sphaeroides* [28,29]. The CheA$_3$A$_4$ copy number is estimated to be the same as the CheA$_3$ copy number; this in turn was inferred from neighbouring gene expression because of the lack of a CheA$_3$A$_4$ antibody. Assuming a cell volume of 0.5 fl for *R. sphaeroides* [30], we obtained the total concentrations of the proteins in μM; this sets the maximum that can be phosphorylated.

$$[A_2] = 86 - [A_{2p}], \quad [Y_3] = 3.2 - [Y_{3p}], \quad [Y_4] = 13.2 - [Y_{4p}], \quad [(A_3A_4)] = 20 - [(A_3A_4)_p]\qquad(7)$$

$$[Y_6] = 170 - [Y_{6p}], \quad [B_1] = 76 - [B_{1p}], \quad [B_2] = 10 - [B_{2p}],\qquad(8)$$

where $[A_2]$ denotes the concentration of CheA$_2$ and all other expressions in brackets follow this notation. Assuming mass action kinetics and using the

Table 3: Average copy number of chemotaxis proteins

Number of copies of	value
CheA$_2$	26000
CheY$_3$	1000
CheY$_4$	4000
CheA$_3$A$_4$	12000
CheY$_6$	51500
CheB$_1$	23000
CheB$_2$	3000

These were determined in microaerobically grown *R. sphaeroides* determined by quantative western blotting [28,29].

dependencies given by (7) and (8), we modeled the transduction part through a set of ODEs:

$$[\dot{A}_{2p}] = [A_2](k_1^* + k_3^-[Y_{3p}] + k_4^-[Y_{4p}] + k_5^-[Y_{6p}] + k_2^-[B_{1p}] + k_6^-[B_{2p}])$$
$$- [A_{2p}](k_3^+[Y_3] + k_4^+[Y_4] + k_5^+[Y_6] + k_2^+[B_1] + k_6^+[B_2])$$

$$(9)$$

$$[\dot{Y}_{3p}] = k_3^+[Y_3][A_{2p}] - [Y_{3p}](k_8 + k_3^-[A_2]) \qquad (10)$$

$$[\dot{Y}_{4p}] = k_4^+[Y_4][A_{2p}] - [Y_{4p}](k_9 + k_4^-[A_2]) \qquad (11)$$

$$[(A_3\dot{A}_4)_p] = [(A_3A_4)](k_{14}^* + k_{12}^-[Y_{6p}] + k_{13}^+[B_{2p}]) - [(A_3A_4)_p](k_{12}^+[Y_6] + k_{13}^+[B_2])$$

$$(12)$$

$$[\dot{Y}_{6p}] = [Y_6](k_5^+[A_{2p}] + k_{12}^+[(A_3A_4)_p]) - [Y_{6p}](k_{10} + k_5^-[A_2] + k_{12}^-[(A_3A_4)])$$

$$(13)$$

$$[\dot{B}_{1p}] = k_2^+[A_{2p}][B_1] - [B_{1p}](k_7 + k_2^-[A_2]) \qquad (14)$$

$$[\dot{B}_{2p}] = [B_2](k_6^+[A_{2p}] + k_{13}^+[(A_3A_4)_p]) - [B_{2p}](k_{11} + k_6^-[A_2] + k_{13}^-[(A_3A_4)])$$

$$(15)$$

The model for the transduction part presented in this section is the same for all models except for the magenta model, for which:

$$[(A_3\dot{A}_4)_p] = [(A_3A_4)](k_{14}^* + k_{12}^-[Y_{6p}] + k_{13}^+[B_{2p}]) - [(A_3A_4)_p](k_{12}^+[Y_6] + k_{13}^+[B_2])$$
$$[(A_3\dot{A}_4)] = -[(A_3A_4)_p] - k_{15}([(A_3A_4)]([Y_{3p}] + [Y_{4p}]) - [(A_3A_4)_i])$$
$$[(A_3\dot{A}_4)_i] = k_{15}([(A_3A_4)]([Y_{3p}] + [Y_{4p}]) - [(A_3A_4)_i]),$$

where $[(A_3A_4)_i]$ denotes the case when $CheA_4$-P cannot phosphorylate $CheA_3$ due to the action of $CheY_3$ and $CheY_4$, and $0.001 \leq k_{15} \leq 1$; the latter is to say that the behaviour of the magenta model remains virtually unchanged within this parameter range of k_{15}.

We obtained the value of k_1, the reaction constant of the auto-phosphorylation of $CheA_2$, from *in vitro* experiments in the absence of the influence of receptors. However, when membrane receptors are in their active state they accelerate the auto-phosphorylation of $CheA_2$. Thus, we modified the reaction constant to $k_1^* = R^a k_1$; that is, we assumed that the *in vitro* reaction rates correspond to the case when receptors are fully active. Similarly, the phosphotransfer from $CheA_4$-P to $CheA_3$ is accelerated when cytoplasmic receptors are active and we modified the reaction constant to $k_{14}^* = \tilde{R}^a k_{14}$.

The reaction rates were obtained by fitting parameters to data from *in vitro* experiments [16,31] (Table 1).

Actuation
We denoted motor activity by M_b for the blue, M_r for the red, M_g for the grey and M_m for the magenta model. We assumed some nonspecific interaction, which does not

lead to a long lasting binding of proteins to motor sites, either between $CheY_6$-P, $CheY_3$-P and $CheY_4$-P and the motor, or only between $CheY_6$-P and the motor. For the blue model we investigated also a different type of CheY motor binding shown below in brackets. This model showed a similar behaviour to the other blue model in simulations and analysis (data not shown) and so we only discussed the findings of the first binding type. Motor activity decreases at a constant rate in the absence of the CheY's, which we assumed to be $\frac{1}{\sec}$. We modeled the behaviour of the different models as follows:

$$\dot{M}_b = [Y_{6p}]([Y_{3p}] + [Y_{4p}]) - M_b$$
$$(\dot{M}_b = [Y_{6p}]\frac{[Y_{3p}] + [Y_{4p}]}{K_4 + [Y_{3p}] + [Y_{4p}]} - M_b \text{ where } K_4 = 0.75.)$$

$$(16)$$

$$\dot{M}_r = Y_{6p} - M_r \qquad (17)$$

$$\dot{M}_g = Y_{6p} - M_g \qquad (18)$$

$$\dot{M}_m = Y_{6p} - M_m \qquad (19)$$

The output of the models is flagella activity that we can also observe in tethered cell assays. We used the following heuristic description to convert motor activity into *R. sphaeroides* body rotations, observed in the tethered cell assays (given in rot/sec):

$$\gamma_i = \frac{-1}{(S + M_i^2)}, i = b, r, g, m \qquad (20)$$

We set $S = 0.125$, which means that saturation occurs at -8 rot/sec.

Parameter fitting
In order to fit model parameters such that the model represents well the experimental data, we minimised the 2-norm of the vector, whose entries consist of the errors between data and predictions from the discretised version of the ODE models representing the individual chemical reactions investigated *in vitro*. The 2-norm of vector x is given by $\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$, where n is the length of x.

In order to fit any remaining model parameters to data from *in vivo* experiments, we simulated the model and minimised the 1-norm of the vector, whose entries are the errors between data and model predictions [32]. The 1-norm of vector x is given by $|x_1| + |x_2| + \ldots + |x_n|$

Fitting model parameters for receptor activation
Parameters K_1 - K_3 are unknown and cannot be easily measured by experiments. To determine these we fitted

the models to wild type tethering data. We performed tethered cell experiments where we applied 100 μM of attractant for 5 minutes and then removed it. We then ran a simulation of each model (Figure 3). To obtain K_is and $\tilde{K}_i$ s, we minimised the error between computer simulations and data following the online fitting procedure. For simplicity and because we are fitting the models to a single model output only, which is contaminated with noise, we let $K_i = \tilde{K}_i$.

Least squares minimisation to fit phosphotransfer parameters
We used least squares minimisation to fit all other parameters from phosphotransfer experimental data. In general, we considered a chemical reaction network with mass action kinetics of the following form:

$$\dot{x} = Af(x), x \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}, \qquad (21)$$

where $f(\cdot) \in \mathbb{R}^m$ is a vector of known monomials. Let the value of the entries of matrix A, which correspond to reaction rate constants, be unknown. What we wished to find is the entries in matrix A, given experimental data. For that purpose, we considered the following discrete-time system:

$$x(t_{q+1}) = x(t_q) + (t_{q+1} - t_q)Af(x(t_q)), \qquad (22)$$

which is the Euler discretisation of (21). Here, each $x(t_q)$ denotes a measurement at time t_q. The set of experimental measurements, which we denote by $\hat{x}$, was used to fit the unknown entries to A such as to minimise the error between the data and the model predictions, which are given by (22). We solved the following optimisation problem which minimises the 2-norm of the error between model prediction and observation (least squares minimisation):

$$\min \left\| \begin{bmatrix} \hat{x}(t_1) - \hat{x}(t_2) + (t_2 - t_1)Af(\hat{x}(t_1)) \\ \vdots \\ \hat{x}(t_{p-1}) - \hat{x}(t_p) + (t_p - t_{p-1})Af(\hat{x}(t_{p-1})) \end{bmatrix} \right\|_2, \qquad (23)$$

where p corresponds to the number of measurements.

Using phosphotransfer data [16,31] and the above method under the constraint that the rates are non-negative, we obtained values for parameters k_1 - k_{14} (Table 1).

Online fitting
Because the least squares optimisation is applied to a discretised version of a continuous time model, we applied the *method of steepest descent* online - that is, we minimise the error between simulations of the

continuous-time model and the data - to improve the fit using the values obtained before.

Experiment design
Input design
In ordered to discriminate between models by determining an input, which results in the largest difference in the outputs of the models we wished to determine the optimal input frequency. For example consider the red and the blue model, shown in Figure 2, representing different connectivities. The models are given by ordinary differential equations of the form:

$$\dot{x}_1 = f_1(x_1) + g_1(x_1, u), \ \dot{x}_2 = f_2(x_2) + g_2(x_2, u) \qquad (24)$$

Here, x_i is the vector with the different states (for example, concentrations of different proteins) of the model i, $i = 1$, 2, f_i and g_i are functions that represent the different models, and u is the input, which can be externally controlled; for example, ligand concentrations.

If $g_i(x_i, 0) = 0$ then the two models should have the same equilibrium point x^* in order to represent the data equally well:

$$f_1(x^*) = f_2(x^*) = 0$$

Moreover, we required that x^* is asymptotically stable. The measured output is given by

$$y_i = h_i(x_i), \qquad (25)$$

where h_i is the output function. In order to be able to discriminate well between two different competing models the outputs of the two should be as different as possible. Thus, we determined the input that maximised the following difference assuming all other model parameters constant:

$$\int_0^\tau (y_1 - y_2)^T (y_1 - y_2)dt. \qquad (26)$$

Here, τ denotes the duration of the experiment. To obtain an input that maximises (27) for a nonlinear system of the form (24)-(25) is difficult. However, if we relax the problem to obtaining the sinusoidal input to the system given by the linearisation of (24)-(25) that maximises (26) then it can be solved systematically as we show in the following.

Consider a linear system

$$\dot{x} = Ax + Bu, \ y = Cx, \qquad (27)$$

where A, B and C are matrices, whose entries depend on the model parameters, and $u(t) = a \sin(\omega t)$ is a

sinusoidal input with angular frequency ω and fixed amplitude a. System (27) is the so called state space representation of the model in the time domain. It is common in control systems engineering to investigate the behaviour of such a system in dependency of ω, which requires to transform the system to the so called frequency domain. If matrix A is Hurwitz (stable) then after some transient behaviour output γ is given by a sinusoidal wave that is, first, phase shifted with respect to u and second, has amplitude $\hat{a}$. Here we are only interested in the maximum of $\hat{a}$ with respect to ω.

We denoted the Laplace transform of u and γ by $U(s)$ and $Y(s)$, where s is a complex independent variable. Then,

$$Y(s) = G(s)U(s).$$

$G(s)$ is called the transfer function [33] and is given by

$$G(s) = C(sI - A)^{-1}B.$$

Note that in our analysis the only independent variable is ω. We replaced s by $j\omega$ ($s = j\omega$), where $j = \sqrt{-1}$. The Bode magnitude plot shows the value of $\frac{\hat{a}}{a}$. Thus, the maximum in the plot provides the frequency that will maximise the output to input ratio.

We linearised (9) to (20) to obtain a system of the form given by (27). We determined the frequency of the sinusoidal input that will result in the largest difference between the model outputs as described above. Finally, we simulated the experimental output of this optimal frequency.

Initial conditions design

Choosing the change or possible combination of changes that provides best discrimination between the different models is a combinatorial problem and difficult to solve efficiently. However, because for our chemotaxis models the range of possible changes was relatively small, due to being limited by what can be implemented experimentally, we performed a brute force search, checking all possibilities. The five-fold over expression of $CheY_4$ under microaerobic growth conditions yielded the best result *in silico* in discriminating between the red and the blue model.

Robustness of invalidation

In order to assess whether our invalidation conclusions are robust to parameter changes, we performed a sensitivity analysis on our models. We ran 500 simulations allowing parameters k_1 and k_{14} to vary by ± 15% (Figure 4C). Even with these variations, we were able reach the same conclusions.

Authors' contributions

MR provided the information to construct the models and performed all the microbiological experiments and data analysis. EA created and interrogated the models. EA and AH further refined the models. AP and JPA conceived of the study, and participated in its design and coordination, and helped to draft the manuscript. PKM and PEM contributed ideas to the mathematical analysis and aided the project management. MR and EA wrote the manuscript. All authors read and approved the final paper.

Additional material

Additional file 1
SBML: file containing the models used in this study.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-3-105-S1.xml]

References

1. Barkai N and Leibler S: **Robustness in simple biochemical networks.** *Nature* 1997, **387:**913–917.
2. Rao CV, Kirby JR and Arkin AP: **Design and diversity in bacterial chemotaxis: a comparative study in *Escherichia coli* and *Bacillus subtilis*.** *PLoS Biol* 2004, **2:**E49.
3. Palsson B: **The challenges of in silico biology.** *Nat Biotech* 2000, **18:**1147–1150.
4. Clemens K and Jens T: **Systems biology: experimental design.** *FEBS Journal* 2009, **276:**923–942.
5. Maksat A, Yves F-N, Jaap AK and Joke GB: **Systems biology: parameter estimation for biochemical models.** *FEBS Journal* 2009, **276:**886–902.
6. Alon U: *An introduction to systems biology: design principles of biological circuits* Boca Raton, Fla; London: Chapman & Hall/CRC; 2006.
7. Anderson J and Papachristodoulou A: **On validation and invalidation of biological models.** *BMC Bioinformatics* 2009, **10:**132.
8. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgöwer F and Gilles ED: **A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions.** *Genome Res* 2004, **14:**1773–1785.
9. Apgar JF, Toettcher JE, Endy D, White FM and Tidor B: **Stimulus design for model selection and validation in cell signaling.** *PLoS Comput Biol* 2008, **4:**e30.
10. Chen BH and Asprey SP: **On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations.** *Ind Eng Chem Res* 2003, **42:**1379–1390.
11. Faller D, Klingmuller U and Timmer J: **Simulation Methods for Optimal Experimental Design in Systems Biology.** *SIMULATION* 2003, **79:**717–725.
12. Baker M, Wolanin P and Stock JB: **Signal transduction in bacterial chemotaxis.** *Bioessays* 2006, **28:**9–22.
13. Wadhams GH and Armitage JP: **Making sense of it all: Bacterial chemotaxis.** *Nat Rev Mol Cell Biol* 2004, **5:**1024–1037.
14. Sourjik V and Schmitt R: **Phosphotransfer between CheA, CheY1, and CheY2 in the chemotaxis signal transduction chain of *Rhizobium meliloti*.** *Biochemistry* 1998, **37:**2327–2335.

15. Wadhams GH, Warren AV, Martin AC and Armitage JP: **Targeting of two signal transduction pathways to different regions of the bacterial cell.** *Mol Microbiol* 2003, **50:**763–770.
16. Porter SL and Armitage JP: **Phosphotransfer in *Rhodobacter sphaeroides* Chemotaxis.** *J Mol Biol* 2002, **324:**35–45.
17. Porter SL, Wadhams GH and Armitage JP: ***Rhodobacter sphaeroides*: complexity in chemotactic signalling.** *Trends Microbiol* 2008, **16:**251–260.
18. Bryson AE and Ho Y-C: *Applied optimal control: optimization, estimation, and control. Revised printing edn* New York; London: Hemisphere; 1975.
19. Doyle JC, Francis BA and Tannenbaum A: *Feedback control theory* New York Toronto: Macmillan Pub. Co: Collier Macmillan Canada; Maxwell Macmillan International; 1992.
20. Franklin GF, Powell JD and Emami-Naeini A: *Feedback control of dynamic systems* Upper Saddle River, NJ: Pearson Prentice Hall; 52005.
21. Porter SL, Wadhams GH, Martin AC, Byles ED, Lancaster DE and Armitage JP: **The CheYs of *Rhodobacter sphaeroides.*** *J Biol Chem* 2006, **281:**32694–32704.
22. Ferre A, de la Mora J, Ballado T, Camarena L and Dreyfus G: **Biochemical study of multiple CheY response regulators of the chemotactic pathway of *Rhodobacter sphaeroides.*** *J Bacteriol* 2004, **186:**5172–5177.
23. August E and Papachristodoulou A: **A new computational tool for establishing model parameter identifiability.** *J Comput Biol* 2009, **16:**875–885.
24. Sistrom WR: **A requirement for sodium in the growth of *Rhodopseudomonas spheroides.*** *J Gen Microbiol* 1960, **22:**778–785.
25. Ind AC, Porter SL, Brown MT, Byles ED, de Beyer JA, Godfrey SA and Armitage JP: **Inducible-Expression Plasmid for Rhodobacter sphaeroides and Paracoccus denitrificans.** *Appl Environ Microbiol* 2009, **75:**6613–6615.
26. Martin AC, Nair U, Armitage JP and Maddock JR: **Polar Localization of CheA2 in *Rhodobacter sphaeroides* Requires Specific Che Homologs.** *J Bacteriol* 2003, **185:**4667–4671.
27. Alon U, Surette MG, Barkai N and Leibler S: **Robustness in bacterial chemotaxis.** *Nature* 1999, **397:**168–171.
28. Gould M: **Chemotaxis gene expression in *Rhodobacter sphaeroides* WS8N.** *DPhil Thesis* University of Oxford; 2006.
29. Brown M: **Control of the Unidirectional Motor in *Rhodobacter sphaeroides.*** *D Phil Thesis* University of Oxford; 2009.
30. Armitage JP and Evans MCW: **Control of the protonmotive force in *Rhodopseudomonas sphaeroides* in the light and dark and its effect on the initiation of flagellar rotation.** *Biochim Biophys Acta* 1985, **806:**42–55.
31. Porter SL and Armitage JP: **Chemotaxis in *Rhodobacter sphaeroides* requires an atypical histidine protein kinase.** *J Biol Chem* 2004, **279:**54573–54580.
32. August E and Papachristodoulou A: **Efficient, sparse biological network determination.** *BMC Syst Biol* 2009, **3:**25.
33. Zhou K, Doyle JC and Glover K: *Robust and optimal control* Prentice-Hall, Inc; 1996.
34. Sockett RE, Foster JCA and Armitage JP: **Molecular biology of the *Rhodobacter sphaeroides* flagellum.** *FEMS Symp* 1990, **53:**473–479.
35. Shah DS, Porter SL, Martin AC, Hamblin PA and Armitage JP: **Fine tuning bacterial chemotaxis: analysis of *Rhodobacter sphaeroides* behaviour under aerobic and anaerobic conditions by mutation of the major chemotaxis operons and *cheY* genes.** *EMBO J* 2000, **19:**4601–4613.
36. Penfold RJ and Pemberton JM: **An improved suicide vector for construction of chromosomal insertion mutations in bacteria.** *Gene* 1992, **118:**145–146.