The role of extrinsic noise in biomolecular information processing systems: an *in silico* analysis

Submitted for the degree of Doctor of Philosophy

by

Oleg Lenive
Imperial College London

September 2014

Typeset in Times by LaTeX

This thesis is the product of my own work, which builds on the work of other as indicated by the references in the text.

No part of this dissertation has already been, or is currently being submitted by the author for any other degree or diploma or other qualification.

This dissertation does not exceed 100,000 words, excluding appendices, bibliography, footnotes, tables and equations. It does not contain more than 150 figures.

This work is supported by a Biotechnology and Biological Sciences Research Council grant and completed in the Department of Life Sciences at Imperial College, London.

Abstract

The intrinsic stochasticity of biomolecular systems is a well studied phenomenon. Less attention has been paied to other sources of variability, so called extrinsic noise. While the precise definition of extrinsic noise depends on the system in question, it affects all cells and its significance has been demonstrated experimentally.

Information theory provides a rigorous mathematical framework for quantifying both the amount of information available to a signalling system and its ability to transmit this information. Intracellular signal transduction remains a relatively unexplored frontier for the application of information theory.

In this thesis, we rely on a metric called mutual information to quantify information flow in models of biochemical signalling systems. After briefly discussing the theoretical background and some of the practical difficulties of estimating mutual information in Chapter 2, we apply it in the context of simplified models of intracellular signalling, referred to as motifs. Using a comprehensive set of two-node motifs we explore the effects of extrinsic noise, model parameters and various combinations of interaction, on the system's ability to transmit information about an input signal, represented by a telegraph process. Our results illustrate the importance of the system's response time and demonstrate a trade-off in transmitting information about the current state of the input or its average intensity over a period of time.

In Chapter 4, we address the problem of determining the magnitude of extrinsic noise in the presence of intrinsic stochasticity. Using the Approximate Bayesian Computation - sequential Monte Carlo algorithm, together with published experimental data, we infer parameters describing extrinsic noise in a model of *E. coli* gene expression.

Lastly, in Chapter 5, we construct and analyse models of bacterial two-component signalling, bringing together insights gleaned from earlier work. The results show how the abundances of different molecular species in the system may transmit information about the input signal despite its stochas-

tic nature and considerable variation in the numbers of protein molecules present.

Contents

List of Figures

Acknowledgements

Acknowledgements

Firstly, I thank my supervisors Professor Michael Stumpf and Dr J Krishanan for giving me the privilege of working on this project. It has been a much more intellectually enriching experience than I could have possibly imagined before starting.

I thank Paul Kirk, Sarah Filippi, Maxime Huvet and Heather Harrington for their advice and helpful answers to questions on many topics, from mathematics to computational practicalities.

A special thank you to Dr Suhail Islam without whose early guidance and continued support, the work described in this thesis would not have been possible.

I would like to thank all the member of the theoretical systems biology group, past and present, John Pinney, Shiobhan Mc Mahon, Delphine Rolando, Justina Zurauskiene, Angelique Ale, Juliane Liepe, Ann Babtie, Rodrigo Liberal, Daniel Silk, Aron Sim, Virginia Fairclough, Rob Johnson, Phoebe Jones, Dominic Smith, Chris Barnes, Michal Komorowski, Tina Toni, Tom Thorne, Adam MacLean, Xia Sheng, Nathan Harmston, Waqar Ali and Ryan Topping. You have always been friendly, lively and supportive. I have really enjoyed the time I've spent with you.

I also thank Dr Christine Batchelor for helpful discussion and an outside perspective on the field, as well as her help with proofreading.

A huge thanks to my parents for their continued moral and material support throughout my time as a PhD student.

Finally, I thank the Biotechnology and Biological Sciences Research Council for funding the project.

Chapter 1

Introduction

Recently there has been increased interest in the application of ideas from information theory to problems in molecular biology. The range of topics is broad, examples include sequence comparison [1], gene regulatory networks [2, 3], tumor necrosis factor signalling [4], positional information during embryogenesis [5] and experimental design [6]. In this thesis, we explore the implications of considering a cellular signalling pathway as a communication channel. To do this, we construct models of biomolecular signal transduction and quantify information flow through these systems.

Complex networks of molecular interactions facilitate information gathering and processing at the cellular level. A detailed understanding of inter- and intra-cellular signalling underpins advances in medicine, biotechnology and other fields. The disciplines of molecular biology and biochemistry are built on a foundation of decades of experimental work: observations of the behaviour of biochemical systems in various settings, both *in vitro* and *in vivo*, characterisation, cataloging and classifying the structures, sequences and functions of various molecular components. Modern high-throughput techniques, from sequencing methods to automated fluorescence microscopy, have provided further insight while also creating new challenges for data analysts. The discipline of systems biology, including mathematical and computational modelling of biomolecular signalling, has developed in parallel with the experimental sciences.

Arguably, we can not hope to attain a satisfactory understanding of both quantitative and qualitative aspects of cellular signalling without resorting to computational models. Such models force us to make our assumptions explicit, test our explanations of experimental observations, and have the potential to reveal previously unexpected behaviour.

They may also be a source of new hypotheses and suggest directions for further experiments. However, mathematical and *in silico* modelling of cellular signalling processes is challenging due to the size and complexity of the networks as well as incomplete knowledge of their biophysical properties.

Computational models of biomolecular processes range from Boolean representations of large genetic interaction networks [7] to atomic-scale modelling of biophysical processes [8]. The models presented in this thesis fall between these two scales and are focused on representing the function of individual molecular species with no consideration given to the nature of the chemical reactions involved or to their molecular structures. We also do not consider the effects of inhomogeneous spatial distributions of molecules within a cell. Such spatial effects are thought not to play a significant role in the functions of systems examined in this work. Although, in other signalling and information processing systems, such as those involved in *Drosophila* embryo morphogenesis, spatial concentration gradients are a critical part of the process [9]. Detailed mechanistic models are usually restricted to small portions of the network, dictated by our interpretation of the system's function and its experimental tractability. However, even at this level, we lack a complete understanding of the interplay between various biochemical processes.

Mechanistic models of cellular signalling usually fall into one of two categories, deterministic or stochastic. Typically, deterministic models consist of systems of ordinary differential equations (ODEs) which can be solved numerically. The trajectory of species concentrations in time can be interpreted as representing the average abundance of the species. Deterministic modelling approaches, such as the linear noise approximation [10] or methods based on expansion of moments [11], can also be used to represent variability arising from stochasticity in the real system. Stochastic differential equation models assume that the system's trajectory consists of two components, a drift term describing the mean of the trajectory and a diffusion term, usually a Wiener process, which accounts for the apparent noise in the system. The above methods use continuous variables to represent species abundances under the assumption that the numbers of molecules involved are sufficiently large for the influence of individual reactions to be negligible.

A more explicit representation of the stochasticity in a molecular system can be achieved by employing dynamic Monte Carlo (DMC) simulation methods, that use discrete variables to explicitly represent the numbers of molecules in the system. The models used

with such methods consist of sets of discrete events, which may represent individual chemical reactions or more complex processes. The probability of a particular event occurring, also known as the hazard, may depend on the numbers of species in the system as well as the rate parameter association with the event. Computationally, DMC simulation may be carried by various algorithms [12], the most well know of which is the Gillespie algorithm [13].

Deterministic and stochastic models have been widely used in systems biology to gain insight into cellular function and decision making. Relatively simple ODE models of bimolecular signalling processes are capable of displaying a range of non-linear dynamics [14–16]. Meanwhile, experimental studies [17, 18] showing that there is considerable cell-to-cell variability in the abundance of many protein and mRNA species continue to motivate the use of stochastic models.

Regardless of the modelling framework used, a number of questions remain pertinent. How reliable is the signalling given that the system may display non-linear dynamics and be subject to sources of noise in the form of intrinsic stochasticity and cel-to-cell variability? What features of the signalling system determine how well it transmits information about the signal? To try to answer them, we turn to ideas from information theory.

1.1 Application of information theory to cellular signalling

In his 1948 paper [19] Shannon lays the foundations for the field of information theory. He derived his results in the context of communication systems which can take on a number of discrete or continuous states. These states can be transmitted from the input to the output of the system with some probability of a change, or error, during the process. Shannon's work provided a mathematical framework for modern communication and data compression technologies. His findings have since been applied in a wide range of disciplines including computer science, physics, economics and statistics [20]. In the biological sciences, information theory has been widely used by neurophysiologists to quantify relationships between neural stimuli and responses [21, 22]. The application of information theoretic ideas to signal processing by biomolecular systems has been less common. Examples include references [2–5, 23–29].

Biomolecular signalling systems have evolved to facilitate the transmission of information about their environment in order to regulate a cell's responses. These processes include:

- Control of gene expression by changes in transcriptional activator and/or repressor concentrations.
- The response of downstream targets of signalling cascades to changes in receptor activity.
- Changes in metabolite concentrations in response to changes in the concentrations of allosteric regulators of metabolic enzymes.

Much of Shannon's original work is concerned with the transmission of a signal over a noisy communication channel. Noise in biomolecular systems can be broadly divided into two categories, intrinsic and extrinsic. Intrinsic noise refers to the variability produced by stochastic processes such as individual chemical reactions or discrete events such as the synthesis or degradation of a macromolecule. Models which are simulated using DMC methods take this stochasticity into account. Conversely, extrinsic noise represents variability due to processes that are not explicitly described in such models. This variability may be in the form of molecule copy numbers or reaction rates. For example, a single strain of *E. coli* cells grown in the same culture displays considerable cell to cell variability, which can not be accounted for solely by intrinsic noise [17]. There is no single, universally accepted, method for modelling extrinsic noise. One approach is to assume that variation in extrinsic noise is much slower than the dynamics of the process being modelled. In this case, extrinsic noise may be represented by perturbations of parameters and initial conditions before each simulation of the model.

To determine the effectiveness of signal transduction through a noisy channel we need to quantify the relationship between the input signal and the system's response. An information theoretic metric called mutual information (MI) provides a statistical measure of interdependence between random variables. One advantage of this metric is that it captures non-linear relationships. While MI may be calculated for any pair of variables, we are particularly interested the MI between variables which are considered to be the inputs and outputs of signalling pathways.

When attempting to make conclusions about information flow through biomolecular systems, there are a number of broader questions which should be addressed. There

is no obvious evolutionary reason why information transmission and information processing should be separated in a biomolecular system. Further, it is unclear whether the amount of information transmitted by a signalling pathway is limited by its biophysical properties or whether there was simply no selective pressure to evolve a pathway capable of transmitting more information. It is not always desirable to maximise the overall MI between input and output variables. For example, adaptive sensing systems such as those involved in bacterial chemotaxis [30] or in light sensitive cells of the eye [31, 32] have evolved to transmit information about relative changes in an input signal. Similarly, in bet hedging situations [33], imperfect transmission of information about an input variable may be selected for.

To tackle some of these complications, we will consider information transmission about particular features of an input signal. We begin by quantifying MI for simple abstract models representing an input signal and its effect on one or downstream cellular components. Next we use single cell data from *E. coli* to try to infer the magnitude of extrinsic noise affecting gene expression. In the final chapter we construct a stochastic model of a bacterial two-component signalling system being expressed from its operon, subject to extrinsic noise. We then estimate the MI between an input signal and the system's response.

Chapter 2

Theoretical Background

2.1 Information Theory

Entropy - quantifying uncertainty

Shannon introduced the concept of entropy, $H(p_i)$, of a discrete probability distribution of events with outcome probabilities given by $p_1, ..., p_n$. He qualified his choice of measure by stating that it should conform to three properties. Namely that H should: (1) be continuous with respect to outcome probability (p_i); (2) increase monotonically with n when all outcomes are equally likely (i.e. $p_i = 1/n$); (3) be a weighted sum of the entropies of combined and nested outcomes. To illustrate the last condition, consider a set of events with outcome probabilities $p_1, ..., p_n$ and let p_1 itself be a set of events with outcome probabilities $q_1, ..., q_m$. The metric should satisfy

$$H(q_1, ..., q_m, p_2, ..., p_n) = H(p_1, ..., p_n) + p_1 H(q_1, ..., q_m).$$

Shannon noted the analogy between his chosen measure and the measure used in statistical mechanics to quantify disorder, diversity, or the number of possible arrangements of a system [19, 20]. Shannon's entropy plays a central role in quantifying uncertainty and information. For a random variable X, which may take values x_1 to x_n with probabilities $p(x_i)$, the entropy H is defined as,

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i). \tag{2.1}$$

The quantity H may be interpreted as a measure of uncertainty in the choice of x. The base of the logarithm used, determines the units, and may be chosen arbitrarily. Commonly used units include bits (base 2) and nats (natural logarithm). Entropy is related to expectation (E) as,

$$H(X) = E \log \frac{1}{p(X)}.$$

This metric may also be viewed as the expected amount of information carried by the outcome of $X \sim p(X)$ about $p(X)$, or even as the average level of surprise that an observer may experience in response to the outcome of $X \sim p(X)$. The entropy of a random variable also corresponds to the average length (e.g. in bits) of the record required to completely describe the outcome of the random variable using the most optimal encoding protocol [20].

An intuitive example of the relationship between entropy, uncertainty and the probabilities of outcomes can be illustrated by considering the entropy of a biased coin flip. Let

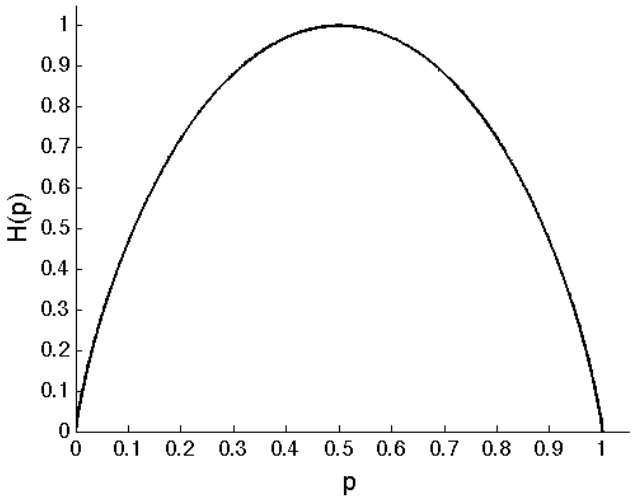

Figure 2.1: Biased coin flip. The entropy (H) is maximised when the uncertainty about the outcome is at its greatest ($p = 0.5$) and reaches zero when the outcome is certain [20].

the probability of heads be p. $H = 0$ if there is only one possible outcome, $p = 0$ or $p = 1$. H is maximised when $p = 0.5$. Intuitively, this corresponds to 1 bit of information about the outcome. For other values of p, either the heads or the tails outcome is more likely, resulting in lower uncertainty ($H < 1$ bit) [20]. These observations have direct bearing on results involving the telegraph process, discussed in subsequent chapters. In general, H is maximised when all outcomes are equally likely ($p(x_i) = 1/n$).

An analagous quantity, known as differential entropy, may be defined if X is a continuous variable,

$$H(X) = -\int_{-\infty}^{\infty} p(x) log p(x) dx. \tag{2.2}$$

However, differential entropy does not retain all the properties of entropy for discrete variables. In particular, it may be negative or zero. It can not be interpreted as the average amount of information required to encode the outcomes of a random variable or as the information contained in a random variable.

Given two discrete random variables, X and Y, which can take on n and m possible values respectively, the entropy of the joint distribution of X and Y is given by,

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log p(x_i, y_j). \tag{2.3}$$

If X and Y are independent $H(X, Y) = H(X) + H(Y)$, otherwise $H(X, Y) \leq H(X) + H(Y)$.

The conditional entropy of a random variable $X|Y$ is similarly given by,

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log p(x_i|y_j). \tag{2.4}$$

The joint and conditional entropies are related to each-other as, $H(X, Y) = H(X) + H(X|Y) = H(Y) + H(Y|X)$. This is known as the chain rule for entropy.

Mutual information

One may want to quantify the extent of interdependence between two variables. This may be done using mutual information (MI), defined as,

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \tag{2.5}$$

$I(X; Y)$ can be thought of as a measure of the decrease in uncertainty about X given knowledge of the outcome of Y and *vice versa*. MI can be expressed in terms of the joint and marginal distributions of two discrete variables,

$$I(X; Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \tag{2.6}$$

The MI between two continuous random variables can also be defined as

$$I(X; Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy. \tag{2.7}$$

Unlike entropy, MI for continuos random variables must be non-negative. It retains its meaning as the decrease in uncertainty about one variable given knowledge of the other.

The MI between two independently distributed variables is zero, while the MI between two perfectly correlated variables is equal to their entropy. Unlike correlation, MI takes into account non-linear relationships between variables but provides no description of any trends in the relationship (as opposed to positive or negative correlation). It also does not indicate any causal relationship between variables. An analytical result links correlation and MI for the specific case of two jointly Gaussian variables,

$$I(X; Y) = -\frac{1}{2} \log(1 - \rho^2) \tag{2.8}$$

where ρ is the Pearson correlation coefficient. This result can be used to make direct comparisons between numerical estimates of MI and the true value.

MI is related to the Kullback-Leibler (KL) divergence [20]. Specifically, it is the KL divergence between the joint distribution of the two variables and the product of their marginal distributions,

$$I(X;Y) = D_{KL}(p(x,y)||p(x)p(y)). \tag{2.9}$$

While KL divergence is not symmetric in general, MI is symmetric.

As with conditional entropy, MI can be conditioned on a third variable Z so that,

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z).$$

In terms of the joint and marginal probabilities,

$$I(X;Y|Z) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l} p(x_i,y_j,z_k)\log\frac{p(z_k)p(x_i,y_j,z_k)}{p(x_i,z_k)p(y_j,z_k)}.$$

This quantity remains symmetric with respect to X and Y.

The maximum amount of information that can be transmitted through a noisy channel depends not only on the nature of the channel but also on the input signal distribution. However the highest amount of information that can be transmitted through a given channel may remain limited by properties of the channel. This limit is known as the channel capacity (C) and is defined as,

$$C = \sup_{p(x)} I(X;Y).$$

If the coin flip example above is instead considered to be a binary channel, transmitting information about the state of the coin, its channel capacity ($C = 1$bit) is reached when $p = 0.5$. In other words, the expected amount of information transmitted by this channel is maximised when the entropy of the input signal is maximised.

It is intuitive that the information content of a signal transmitted through an imperfect communication channel can not be increased by subsequent processing. More formally, given a Markov chain of three variables $X \to Y \to Z$,

$$I(X;Z) \leq I(X;Y).$$

This is known as the data processing inequality [20].

2.2 Mutual Information

Previous applications to biochemical systems

In previous work involving the estimation of MI in models of biomolecular systems [2, 4, 23–28, 34] investigators have often relied on the assumption that the variables of interest follow multivariate normal distributions. One advantage of such an approach is that the MI for a bivariate Gaussian distribution can be obtained using the analytical expression.

In references [2, 27, 35], the authors consider models of small genetic networks. They rely on the assumption of multivariate normal distributions at steady state, together with a fixed prior limit on the maximum number of molecules in the system, to optimise model parameters for a high MI between an input signal and the output of the system.

We take an alternative approach to obtaining a distribution of outputs at steady state. Rather than representing system variability as Gaussian noise around a single steady state, we consider a set of steady states produced by perturbation of the system parameters. This corresponds to our model of extrinsic noise.

In this work, we aim to obtain reasonable estimates of MI between variables which do not come from a multivariate Gaussian distributions. However, we begin by using samples obtained from bivariate Gaussian distributions to assess and compare the performance of various estimators of MI.

Rather than trying to find the maximum MI (channel capacity) given a set of constraints, we explore parameter space with the aim of elucidating the relationship between model parameter space and MI between the input signal and the model's output.

Estimating MI between continuous variables

Our aim is to obtain an estimate of the MI between two continuous variables given a limited number of samples from the joint probability distribution. Figure 2.2 illustrates four common approaches to tackling this problem. We briefly outline each approach before considering some of the practical challenges and justifying our choice of estimator for the subsequent work.

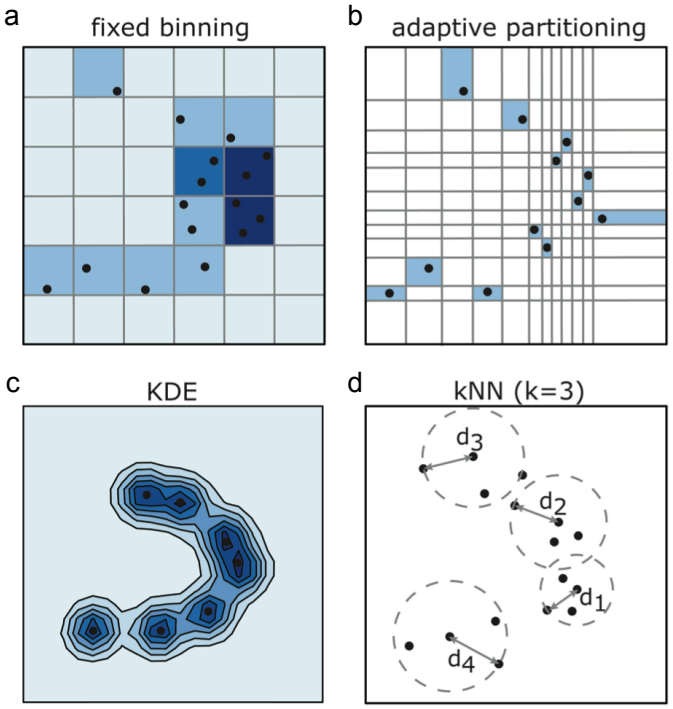

Figure 2.2: Four approaches for estimating MI given a limited number of samples from the joint distribution of two continuous variables. (**a**) The plug-in estimator with fixed bin sizes. (**b**) Adaptive partitioning. (**c**) Kernel density estimation. (**d**) k-nearest neighbour statistics. Figure taken from [36].

The plug-in estimator

A natural approach is to partition the joint probability distribution space into equally sized bins (Figure 2.2a). The probability density at the centre of each bin is then assumed to be proportional to the number of data points found in that bin. Marginal probability distributions are calculated by summing the rows or columns of the joint probability matrix. This is known as the "plug-in", "empirical" or "naïve" estimator [37].

Adaptive partitioning

An alternative approach is to allow the dimensions of each bin, or partition, to vary according to the available data (Figure 2.2b). Potential advantages of this approach include decreased sensitivity to outliers and a finer resolution in regions where sample density is high.

Given a sample of N points from from a joint distribution $f(X, Y)$, the support can be partitioned into a grid of N_E^2 rectangular elements of variable lengths along the x- and y-axes [38]. The lengths may be chosen so that the data points are uniformly distributed among the rows and columns of the grid. Specifically, the probability of a data point being in the i-th column or j-th row is given by,

$$P(i) = P(j) = \frac{1}{N_E}$$

For independent variables, the expected number of elements in each grid element, given by $E(i, j) = NP(i)P(j)$, can be used to decide on the appropriate choice of N_E. Ideally N should be divisible by N_E. $I(X, Y)$ may then be calculated by applying the formula for MI between discrete variables (Equation 2.6).

The Fraser-Swinney algorithm [39] provides an alternative method of adaptive partitioning. Rather than considering the number of elements in each bin, the algorithm begins with a fixed partitioning of the sample space and recursively segregates data into smaller bins by dividing existing bins in half along the x- and y-axes. This process requires some criterion for determining when a bin should not be sub-divided further. In the Fraser-Swinney algorithm, the criterion is that the sample distribution in a bin appears uniform as determined by a χ-squared test.

k-nearest neighbour estimators

Methods for estimating MI based on k-nearest neighbour (kNN) statistics have been developed by Kraskov *et al* [40]. Their approach builds on previous kNN based methods for estimating entropy. Given a metric for measuring the distance between samples X_i, a kNN based approach involves obtaining an expression for the probability distribution ($P(\epsilon)$) of the distance (ϵ) between X_i and its k-th nearest neighbour. By considering the true probability density function to be approximately constant over the region containing the kNNs for each X_i, an expression relating entropy to ϵ was obtained. MI can then be calculated as the difference between the marginal and joint entropies (Equation 2.5). The authors concluded that smaller values of k decrease systematic bias of the estimate while larger k values help reduce the variance of the estimate.

Kernel density estimators

A kernel density estimator (KDE) does not rely on binning data. Instead, each sampled data point (X_i) contributes to a continuous approximation ($\hat{f}$) of the true probability

distribution (f).

$$\hat{f}(x, X_i; t) = \frac{1}{N} \sum_{i=1}^{N} K(x, X_i; t) \tag{2.10}$$

where t determines the scale of the kernels and is known as the bandwidth. In the case of the commonly used, Gaussian kernel

$$K(x, X_i; t) = \frac{1}{\sqrt{2\pi t}} e^{(x-X_i)^2/2t}. \tag{2.11}$$

One-dimensional Gaussian KDEs are applicable to the estimation of MI between a continuos and a discrete variable. This method also naturally extend to the two-dimensional case for estimating MI between two continuous variables. In some rare cases it may be possible to plug the density estimate into Equation 2.7 and solve the resulting equation analytically. However, a more general computational approach is to calculate the estimated probability density at each point of a grid and apply Equation 2.6 to obtain the estimate of MI.

In subsequent work we rely on a kernel density estimator developed by Botev *et al* [41]. First, we briefly consider the practical challenges faced when using the plug-in estimator and compare its performance to the kernel density estimator of Botev *et al*.

2.3 Practicalities of estimating MI

The accuracy of the plug-in method is determined by the quality of the estimate of the joint probability distribution. In the limit of an infinite number of data points and infinitely small bin sizes, this approach converges to the true value of the MI. However, given realistic constraints, there are three main factors which determine the accuracy of the plug-in method:

1. The support over which the joint probability density is estimated.
2. The number of available data points (N).
3. The number of bins (m).

Selecting an appropriate support for the joint probability density distribution minimises systematic errors cause by truncation of the distribution. In the case of the plug-in

estimator, this may simply mean making the support sufficiently broad to include all the available data points.

The relationship between m and N is non-trivial. If the selected bin size is large, the estimated joint probability density will be a coarse approximation to the true density and is likely to lead to an underestimate of MI. Conversely, if there are too few data points for the chosen number of bins, the estimated joint probability density will contain spurious features which are likely to result in an overestimate of MI.

A further complication comes in the form of a systematic bias which can be reduced by increasing m but not by increasing N as illustrated in Figure 2.3a. The cause of systematic bias is that the curvature of the probability density manifold is poorly estimated when assuming that the number of samples in the bin corresponds to density at the centre of the bin.

Although the number of data points which can realistically be obtained by simulating from a model far exceeds that produced by experimental studies, it is not practical to rely solely on brute force, repeated simulations, to obtain accurate estimates of MI.

Discretisation of continuous variables creates a bias in the MI estimate. A number of methods of correcting for this bias have been proposed, including Miller-Madow, shrink entropy, and Schumann Grassberger [37].

Diffusion based kernel density estimation

One of the draw-backs of the simplest implementation of the Gaussian kernel density estimator is that the estimated density is not reconciled with known boundary conditions of the true distribution. Botev *et al* [41] developed a density estimator that is consistent near the boundary by treating the estimated density as the probability density of a diffusion process. The authors began with the observation that the Gaussian kernel density estimate (Equation 2.10) corresponds to the solution of a partial differential equation,

$$\frac{\partial}{\partial t}\hat{f}(x;t) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\hat{f}(x;t)$$

which describes the state of a diffusion process after time t given initial conditions,

$$\hat{f}(x;0) = \frac{1}{N}\sum_{i=1}^{N}\delta(x - X_i)$$

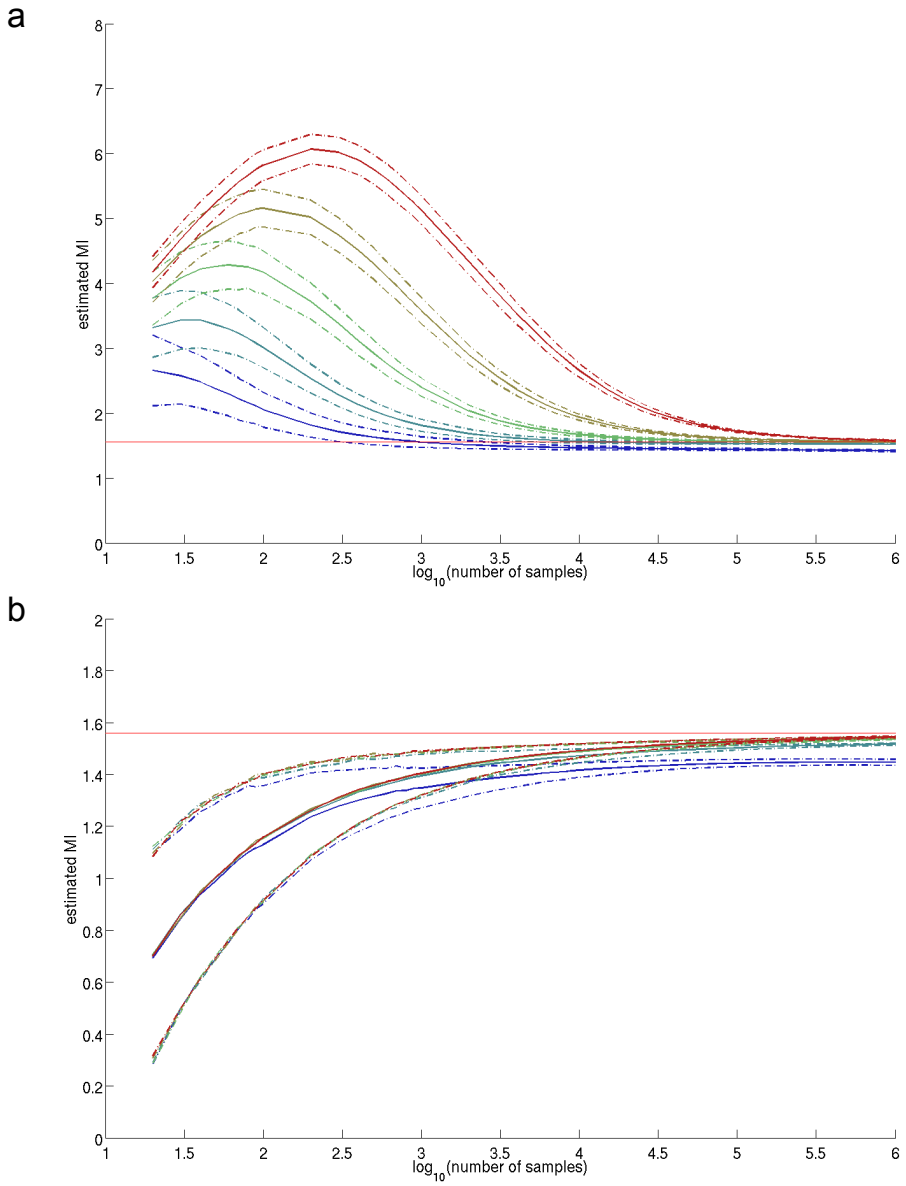

Figure 2.3: Dependence of different estimators of MI on the number of available data points and the discretisation of the joint distribution space. Means of 1000 independent estimates of MI (solid lines) are plotted against the number of samples used for each estimate. Dashed lines indicate a distance of two standard deviations from the mean. Samples were taken from a bivariate Gaussian distribution with a correlation coefficient of 0.9406. Estimates were made using (**a**) the plug-in method and (**b**) the kernel density estimator. For each estimator the joint probability distribution is represented numerically by 2^5 (blue), 2^6 (dark green), 2^7 (light green), 2^8 (yellow) 2^9 (red) discrete bins, for the plug-in method or mesh points, for the kernel density estimator.

where δ is Dirac's delta function. They demonstrate that for sufficiently small values of t this is equivalent to the standard Gaussian kernel density estimator and as $t \to \infty$ the estimated distribution approaches a uniform distribution over the bounded region. To tackle the problem of selecting an appropriate value of t, the authors developed a non-parametric bandwidth selection algorithm for use with their Gaussian kernel density estimator.

We employ the kernel density estimator developed by Botev *et al* to obtain an estimate of the joint probability distribution of two variables based on a limited number of samples. The distribution is represented numerically by a discrete mesh of 2^n points. The density at each mesh point is normalised before being used to calculate the MI estimate according to Equation 2.6.

Choice of estimator

Figure 2.3b provides an example of how the MI estimate for a bivariate Gaussian distribution is affected by varying the number of available data points and the number of mesh points (n).

Figure 2.3 illustrates some of the practical challenges faced when the joint probability distribution of interest is unknown. For very large numbers of samples the accuracy of the plug-in method increases as the number of bins is increased. However, for a fixed number of samples, increasing the number of bins may lead to a decrease in the accuracy of the estimate. Most saliently, it is unclear *a priori*, what combinations of m and N are appropriate for a given distribution. The KDE based estimator suffers from an analogous problem but with respect to the number of mesh points used to approximate the joint probability distribution. The use of a more fine mesh (higher n) requires additional computational time. The key difference is that an inappropriate choice of bin sizes when using the plug-in estimator leads to large errors in bins which contain few samples. These errors create spurious complexity in the estimate of the joint probability distribution, resulting in an inflated estimate of MI. The adaptive KDE algorithm compensates for this by picking a higher bandwidth for lower values of N. Thus, in the example shown in Figure 2.3, as n increases, the accuracy of the MI estimate improves regardless of N. For this reason, we rely on the adaptive KDE estimator in the subsequent work.

2.4 Parameter Inference and Likelihood

When constructing a mathematical model of a system or process, one is presented with two basic types of choices. Firstly, an abstract, mathematical description of the process is required. Secondly, numeric values for model parameters need to be selected. Here, we focus on the problem of inferring model parameters when experimental measurements and a model of the system are available.

One common approach is to employ maximum likelihood estimation (MLE). When using MLE, we are assuming that there exists a parameter set that corresponds to the 'true' parameter set of the real system. Thus, the observed data represents a sample from a probability distribution which is a function of the true parameter set. To proceed with the MLE, we need to define a likelihood function,

$$\mathcal{L}(\theta; D) = p(D|\theta),$$

where θ is the model parameter set and D corresponds to experimental observations or data. The likelihood function, or simply "likelihood", allows us to compare parameter sets and identify those that have a greater chance of corresponding to the true parameter. The aim is then to find the parameter set, $\hat{\theta}$, that has the highest likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\theta; D).$$

In practice, this problem can rarely be solved analytically. Instead, numerical optimisation techniques are often used to explore parameter space and attempt to identify the global maximum of the likelihood [42]. One of the main challenges is to find the global maximum in a realistic amount of computational time without becoming trapped in a local maximum. Whether the numerical approach will succeed in a particular case depends on the method used and the shape of the likelihood function. A detailed discussion of optimisation methods is beyond the scope of this chapter, for a detailed discussion of the topic see reference [43].

Alternatively the problem of model parameter inference can be considered in the Bayesian framework. The relationship between the available data (D) and a set of model parameters (θ) can be described using Bayes' theorem,

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{2.12}$$

where $p(\theta|D)$ is the (inferred) posterior, $p(D|\theta)$ is the likelihood, $p(\theta)$ describes our (prior) belief or knowledge about the possible parameter ranges, and $p(D)$ is termed

the evidence. The goal of Bayesian parameter inference is to obtain the (posterior) distribution of model parameters which are plausible given the available data.

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Since we are interested in the relative likelihood that a particular parameter produces model behaviour which is in agreement with the data, we do not need to calculate $p(D)$ [44]. In other words, we may consider all the available data to be of equal importance.

2.5 Approximate Bayesian Computation

For most models of biomolecular systems, it is difficult or impossible to obtain an analytical expression for the likelihood [45]. Here we discuss how, so called, likelihood-free methods based on approximate Bayesian computation (ABC) may be used to circumvent explicit calculation of the likelihood and instead provide samples of model parameters that fall within credible intervals of the true parameter value.

To achieve this, model parameter sets, also referred to as particles, are sampled from a prior distribution and used to obtain realisations from the model. These simulated outcomes are then compared to the data using an appropriately chosen distance metric d and tolerance ϵ, which determine how stringently the posterior is approximated [46].

The rejection sampler

A relatively simple but robust ABC method, called the rejection sampler [46–49], involves sampling parameters directly from the prior distribution (π) and simulating from the model (M) until a sufficient number (N) of particles that fall within ϵ of the data are obtained (Algorithm 1). As every particle is sampled independently, the ABC rejection algorithm does not run the risk of becoming trapped in local minima. The price of this robustness is that only a small proportion of particles are accepted. Thus, the computational time needed for this algorithm becomes prohibitive as model complexity and the number of model parameters increase.

ABC Markov chain Monte Carlo

To tackle the problem of low acceptance rates, an alternative ABC algorithm, known as Markov chain Monte Carlo (MCMC), has been developed [47–49]. The ABC-MCMC

Algorithm 1 ABC rejection algorithm

Inputs: M, π, N, ϵ
Outputs: Set of N accepted particles

$i \leftarrow 1$
while $i \leq N$ **do**
    Sample a new candidate particle $\theta^* \sim \pi(\theta)$
    Simulate from the model to obtain $x = f(\theta^*, M)$
    if $d(D, x) \leq \epsilon$ **then**
        Accept particle
        $i \leftarrow i + 1$
    end if
end while

algorithm is related to the Metropolis-Hastings algorithm, which produces dependent samples from a target distribution. Both algorithms start from an initial sample or point in parameter space and proceed in a step-wise fashion to create a Markov chain of particles. Each new particle (θ^*) is either the same as the previous particle (θ_i) or corresponds to a sample from a proposal distribution $q(\theta|\theta_i)$. The Metropolis-Hastings algorithm requires us to be able to evaluate a function $g(\theta)$, that is proportional to the target distribution (Algorithm 2). For a symmetric proposal distribution $q(\theta_i|\theta^*) = q(\theta^*|\theta_i)$, candidate

Algorithm 2 The Metropolis-Hastings algorithm

Inputs: f, N, q
Outputs: Set of N samples from $g(\theta)$

$i \leftarrow 1$
Initialise θ_i
while $i \leq N$ **do**
    Sample new candidate particle $\theta^* \sim q(\theta|\theta_i)$

$$\alpha \leftarrow \min\left(1, \frac{g(\theta^*)q(\theta_i|\theta^*)}{g(\theta_i)q(\theta^*|\theta_i)}\right)$$

    Set $\theta_{i+1} \leftarrow \theta^*$ with probability α, otherwise $\theta_{i+1} \leftarrow \theta_i$
    $i \leftarrow i + 1$
end while

particles are accepted according to the ratio of the target distribution at the two points θ^* and θ_i.

In the ABC-MCMC algorithm, our aim is to obtain samples that approximate the posterior distribution. However, while we can evaluate the function describing the prior, we do not have a function which is proportional to the posterior. As with the ABC rejection method, a distance metric and tolerance are used to constrain the set of accepted particles and approximate the posterior (Algorithm 3). The sampled particles form a

Algorithm 3 ABC-MCMC algorithm

Inputs: M, π, N, ϵ, q
Outputs: Set of N accepted particles

$i \leftarrow 1$
Initialise θ_i
while $i \leq N$ **do**
    Sample new candidate particle $\theta^* \sim q(\theta|\theta_i)$
    Simulate from the model to obtain $x = f(\theta^*, M)$
    if $d(D, x) \leq \epsilon$ **then**

$$\alpha \leftarrow \min\left(1, \frac{\pi(\theta^*)q(\theta_i|\theta^*)}{\pi(\theta_i)q(\theta^*|\theta_i)}\right)$$

        With probability α, set $\theta_{i+1} \leftarrow \theta^*$ and $j \leftarrow j + 1$.
    else
        $\theta_{i+1} \leftarrow \theta_i$
    end if
    $i \leftarrow i + 1$
end while

Markov chain and the algorithm comes to a halt once the desired number of particles ($i = N$) have been accepted into the approximation of the posterior.

The stationary distribution of accepted particles is $\pi(\theta|d(D, x) \leq \epsilon)$. However, unlike the ABC rejection algorithm, ABC-MCMC produces dependent samples. One practical concern is that, if the posterior distribution contains more than one local minimum, the Markov chain may be confined to one of the minima for a disproportionate number of steps. A related problem may occur if the Markov chain enters a region of parameter space where there is little overlap between the proposal and posterior distributions. In this case, the probability of accepting a new candidate particle may become very low, reducing the efficiency of the algorithm [48].

ABC sequential Monte Carlo

To overcome some of the practical challenges of estimating posterior parameter distributions using the ABC-MCMC algorithm, Sisson *et al* [48] and Toni *et al* [49] developed an alternative approach, based on sequential Monte Carlo (SMC). The ABC-SMC algorithm generates a series of particle populations that converge on the region of higher likelihood. The initial population is sampled from the prior distribution. Subsequent populations are obtained by generating new particles from the preceding population and accepting or rejecting them according to how well they reproduce the data, as measured by a distance metric d. The acceptance thresholds for each new population (ϵ_t) become increasingly stringent ($\epsilon_t < \epsilon_{t-1}$) so that the final population represents a sample from the target distribution. To generate a new particle (θ^{**}), a particle (θ^{*}) is sampled from the previous population according to the particle weights, then perturbed according to a perturbation kernel K_t.

The sequence of acceptance threshold distances $\epsilon_1, ..., \epsilon_T$, referred to as the ϵ schedule, can be determined in advance, as in Algorithm 4, or adjusted dynamically based on the distances associated with particles in the preceding population. In practice, the performance and efficiency of the ABC-SMC algorithm are dependent on the choice of the ϵ schedule and the perturbation kernel [50]. However, the ABC-SMC algorithm has less of a tendency to become trapped in a region of low likelihood than the ABC-MCMC algorithm.

Of the three ABC methods described here, ABC-SMC has emerged as the most powerful and widely used approach. It continues to be a subject for further research and finds applications in a variety of settings [6, 51–53]. While the ABC-SMC approach has also been extended to tackle the problem of model selection [49], in this work we only use it for parameter inference.

Algorithm 4 ABC-SMC algorithm with a pre-defined ϵ schedule

Inputs: M, π, N, $\epsilon_1, ..., \epsilon_T$, K_t
Outputs: Set of m populations of N accepted particles

$i \leftarrow 1$
$t \leftarrow 0$
while $t \leq T$ **do**
  if $t = 0$ **then**
    while $i \leq N$ **do**
      Sample a new particle $\theta^* \sim \pi(\theta)$
      Simulate from the model to obtain $x = f(\theta^*, M)$
      if $d(D, x) \leq \epsilon_t$ **then**
        Accept particle $\theta^{(i,t)} \leftarrow \theta^*$
        $\omega^{(i,t)} \leftarrow 1$
        $i \leftarrow i + 1$
      end if
    end while
  else
    while $i \leq N$ **do**
      Sample θ^* from $\{\theta^{(j,t-1)}\}_{1 \leq j \leq N}$ with probability $\{\omega^{(j,t-1)}\}_{1 \leq j \leq N}$.
      Perturb to obtain a new particle $\theta^{**} \sim K_t(\theta|\theta^*)$
      Simulate from the model to obtain $x = f(\theta^{**}, M)$
      if $d(D, x) \leq \epsilon_t$ **then**
        Accept particle $\theta^{(i,t)} \leftarrow \theta^{**}$

$$\omega^{(i,t)} \leftarrow \frac{\pi(\theta^{(i,t)})}{\sum_{j=1}^{n} \omega^{(j,t)} K_t(\theta^{(i,t)}|\theta^{(j,t-1)})}$$

        $i \leftarrow i + 1$
      end if
    end while
  end if
  Normalise particle weights
  $t \leftarrow t + 1$
end while

Chapter 3

Extrinsic noise in ODE models of small signalling motifs

3.1 Introduction

Intracellular signal transduction and processing takes place through changes in the abundances and activities of specific molecular species. Natural or synthetic biochemical signalling pathways and regulatory systems are often described in terms of networks where the nodes represent species and directed edges describe interactions between them. Several connected nodes can be referred to as a motif. Alon and co-workers have proposed a definition of network motifs based on the relative frequency of their appearance in realistic networks as compared to a suitable statistical null model of network organization [54]. Among many other applications this was applied to the *E. coli* transcriptional regulation network. Different motif concepts have been introduced in the recent literature and here the term is used to describe any small network of connected nodes. The particular relevance of motifs (and this has driven the choice of definition used here) is that they represent functional units that can be joined together to form larger networks. In a bottom-up perspective there is the hope that understanding the dynamics of individual motifs will translate (at least partially) to larger systems composed of such motifs.

Variants of similar motifs may appear in homologous systems [55] or as a result of convergent evolution in distantly related species which have evolved similar solutions to the problems of signal transduction [56, 57]. This motivates the search for general trends and conclusions regarding the dynamical properties of broad classes of small

networks and interaction types. More generally, the ability of cellular systems to sense their environment and respond appropriately to any changes in it have moved into the focus of evolutionary, but also synthetic biology. In the former context we can associate differential reproductive success to differences in how reliably and quickly an organism parses its environment and initiates appropriate responses. In the latter context, we may be interested in having reliable biosensors, which rely on potentially complex molecular machines in order to sense and indicate the presence or absence of target stimuli.

In order to fully appreciate the ability of molecular systems to process and transmit information, we set out to model the dynamics of simple motifs and how differences between cells affect cell populations. Here we will specifically consider cases where the molecular abundances inside cells are sufficiently high that stochastic effects can be ignored. In practice this is often thought to be the case in eukaryotic signalling systems [58] and we can therefore employ deterministic, ordinary differential equation (ODE) models. These, of course, allow for complex and non-linear behaviour, but even for relatively simple system dynamics the link between dynamical behaviour and information transmission capability can give rise to surprisingly rich behaviour.

We use mutual information to quantify the differences in information transmission along the simplest motifs. Information theory provides a very general mathematical framework for this analysis, and has recently gained great prominence in molecular and cell biology. Information here is defined probabilistically as the "level of surprise upon seeing a given instance of a random variable." In real-world systems, however, as we discuss below, even populations of genetically identical cells exhibit considerable levels of cell-to-cell variability. The causes of such variability are manifold and are only beginning to be understood: differences in cell-cycle stage, availability of ribosomes, proteasomes etc, cause cells to respond differently to identical stimuli as can the local microenvironment, particularly when extracellular signalling is important. Examples of processes that may be affected by this include cell differentiation in multicellular organisms and the transmission of signals between neurones [59].

Here we investigate the role that such cell-to-cell variability has on the transduction of simple signals along simple motifs. To represent extrinsic noise, we perturb the ODE model parameters using a Gaussian kernel. In doing this, we depart from the commonly used assumption that the input and output of the model are jointly Gaussian. We then assess the influence of model parameters on the system's ability to robustly transmit information about an external signal by estimating the mutual information (MI) between

the system's states and properties of the input signal.

By perturbing model parameters, rather than adding noise to the observed output of the model, we acknowledge that the effects of extrinsic noise are subject to the non-linear dynamics of the system. This is also what differentiates extrinsic noise from measurement error.

We begin by considering a constant input signal, which can take values of either zero or one. The output corresponds to the steady state of the system given the input signal. We determine which motifs are most likely to robustly transmit information about the input. Next, we investigate how signal transduction is affected when the input is changing dynamically and note that, in this regime, information transmission is dominated by the response time of the system. Finally, we compare each motif's ability to transmit information about the current state of the signal with its ability to extract information about the average signal intensity. Both are dominated by the response time of the system. We observe a trade-off between transmitting information about the current state of the input signal and its average intensity over a period of time. Different, and sometimes conflicting, requirements are placed on the model's parameters when trying to achieve close to maximum information transmission in these two scenarios.

3.2 Motifs subject to extrinsic noise

Modelling signalling motifs

Two common modelling formalisms for relating the structure of a network to an ODE model are mass action and Hill kinetics. These regulate the way in which the rate of change in one species can be explained by interactions with another. The mass action kinetic model of a single node with an activating input is described by,

$$\frac{dX}{dt} = a_X + b_X S - \delta_X X \tag{3.1}$$

where X is the amount of gene product, S is the strength of the input signal, a_X is the basal expression rate in the absence of an input signal, b_X represents the strength of the promoter and δ_X is the degradation rate of the gene product. The corresponding Hill equation model, which is typically used to describe cooperative effects in enzymatic reactions, but which also serves as a model for gene activation and repression, for a

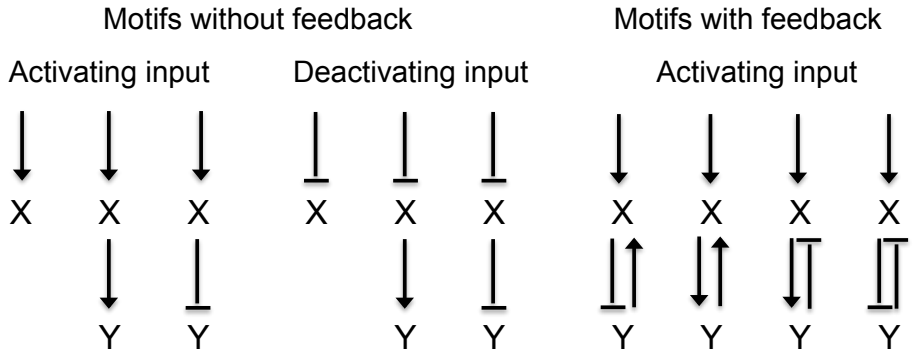

Figure 3.1: The set of motifs considered in this study. Nodes (X or Y) represent genes while edges represent either activating (arrows) or repressing (blunt arrows) interactions.

single node with an activating input is given by,

$$\frac{dX}{dt} = a_X + b_X \frac{(S/K_{SX})^{H_{SX}}}{1 + (S/K_{SX})^{H_{SX}}} - \delta_X X \tag{3.2}$$

and for a deactivating input,

$$\frac{dX}{dt} = a_X + b_X \frac{1}{1 + (S/K_{SX})^{H_{SX}}} - \delta_X X \tag{3.3}$$

We choose to use models based on Hill kinetics to account for the fact that intracellular biochemical processes must become saturated at some level due to the limited amount of material in a cell. Each node corresponds to a state variable in an ODE model.

Figure 3.1 illustrates the combinations of interconnections between nodes considered here. Edges represent either activating or repressive interactions. Nodes which are subject to regulation from two different sources, in this case node X for each of the motifs with feedback, can be regulated in two different ways. In the first case, both input signals are required to elicit a response. This is represented in the ODE model by the multiplication of terms representing the incoming edges,

$$\frac{dX}{dt} = a_X + b_X \frac{(S/K_{SX})^{H_{SX}}}{(1 + (S/K_{SX})^{H_{SX}})} \frac{(Y/K_{YX})^{H_{YX}}}{(1 + (Y/K_{SX})^{H_{YX}})} - \delta_X X \tag{3.4}$$

Alternatively, the incoming signals may act independently, in an additive fashion,

$$\frac{dX}{dt} = a_X + b_X \left(\frac{(S/K_{SX})^{H_{SX}}}{(1 + (S/K_{SX})^{H_{SX}} + (Y/K_{YX})^{H_{YX}})} + \frac{(Y/K_{YX})^{H_{YX}}}{(1 + (S/K_{SX})^{H_{YX}} + (Y/K_{YX})^{H_{YX}})} \right) - \delta_X X \tag{3.5}$$

The two mechanisms have been previously described as AND and OR respectively, by analogy to electronic circuits [60].

The Hill kinetic model is relatively simple and generic. It has previously been used to represent regulatory interactions between genes [23, 61, 62], as well as other biochemical processes such as cooperative enzymatic reactions and changes in the concentrations of covalently modified species in a signalling pathway [14, 63]. While in principle a two-node model may also be used to describe transcription and translation, here we abstract away the mechanistic detail of gene expression to focus on the effects of different combinations of regulatory interactions between genes.

Extrinsic noise

Biomolecular signal transduction involves non-linear, dynamic responses in the presence of noise. The noise prevalent at molecular levels can be broadly categorised as being either intrinsic or extrinsic. Even in a genetically homogeneous population of cells, the copy numbers of individual proteins may vary considerably. Stochastic simulation algorithms have often been used to account for variation in the number of species comprising the system of interest. Such variability, due to random effects at the level of molecular interactions has been termed 'intrinsic noise'. In contrast, 'extrinsic noise' has emerged as a catch-all term used to describe variability due to factors that are not explicitly accounted for in the model. These may include differences in the copy numbers of transcriptional and/or translational machinery, ribosomes, proteasomes, concentrations of the relevant metabolites or inhomogeneities in the extracellular environment that affect cells differently.

Investigations of both prokaryotic and eukaryotic gene expression [17, 18] have demonstrated that Poisson production and degradation of individual mRNA and protein molecules is not sufficient to account for the overall observed variability in a population of otherwise identical cells. Extrinsic noise is one potential mechanism that could generate higher than Poissonian variability at the population level. In the context of mathematical modelling, extrinsic noise may conveniently be represented as variability in the parameter values of the model. In this approach a set of simulations with varying rate parameters represents a sample from a population of cells. Here, we model extrinsic noise by sampling each rate parameter of our model from a truncated Gaussian distribution (N), limiting model parameters to non-negative values. The mean of this distribution

(θ) represents the population average and will be referred to as the 'nominal' parameter value. The parameters used for each simulation ($\hat{\theta}$) are then obtained according to,

$$\hat{\theta}_i \sim N(\theta_i, (\eta\theta_i)^2) \tag{3.6}$$

where θ_i is the i^{th} nominal parameter value, $\hat{\theta}_i$ is referred to as the 'perturbed' parameter, η is the magnitude of extrinsic noise, and i is an index over all the model parameters which are subject to such extrinsic noise. The noise affecting each parameter is assumed to be uncorrelated to noise in other parameters.

3.3 Results

Steady state response

We begin by considering the steady states of nodes X and Y in response to one of two constant input signal states for the case when $H = 1$. In order to explore model parameter space the remaining nominal parameters are sampled from a $\log_{10}$-uniform prior ranging from 10^{-2} to 10^{2}. The ODEs describing the behaviour of the system are solved analytically to obtain samples from the joint probability distribution of the input and model output at steady state.

Estimates of $I(S;X)$ are calculated from an approximation of the joint probability distribution obtained using a one-dimensional kernel density estimator. To determine how many samples from the joint probability distribution are required we tested the adaptive KDE estimator on a range of bivariate Gaussian distributions with known parameters, and thus know MI values (see Chapter 2). We conclude that 10^4 samples are sufficient to obtain reasonable estimates of MI based on the joint input-output distributions produced by these models. Figure 3.2a shows an example of the distributions of steady states obtained at a node subject to an activating or repressing input signal. The full set of motifs analysed here is shown in Figure 3.3a.

Information carried by steady state outputs

Due to extrinsic noise, a distribution of steady state outputs is produced in response to each input. In this regime the motif's ability to transmit information about the signal is determined by the extent of overlap between these distributions. This, in turn, depends

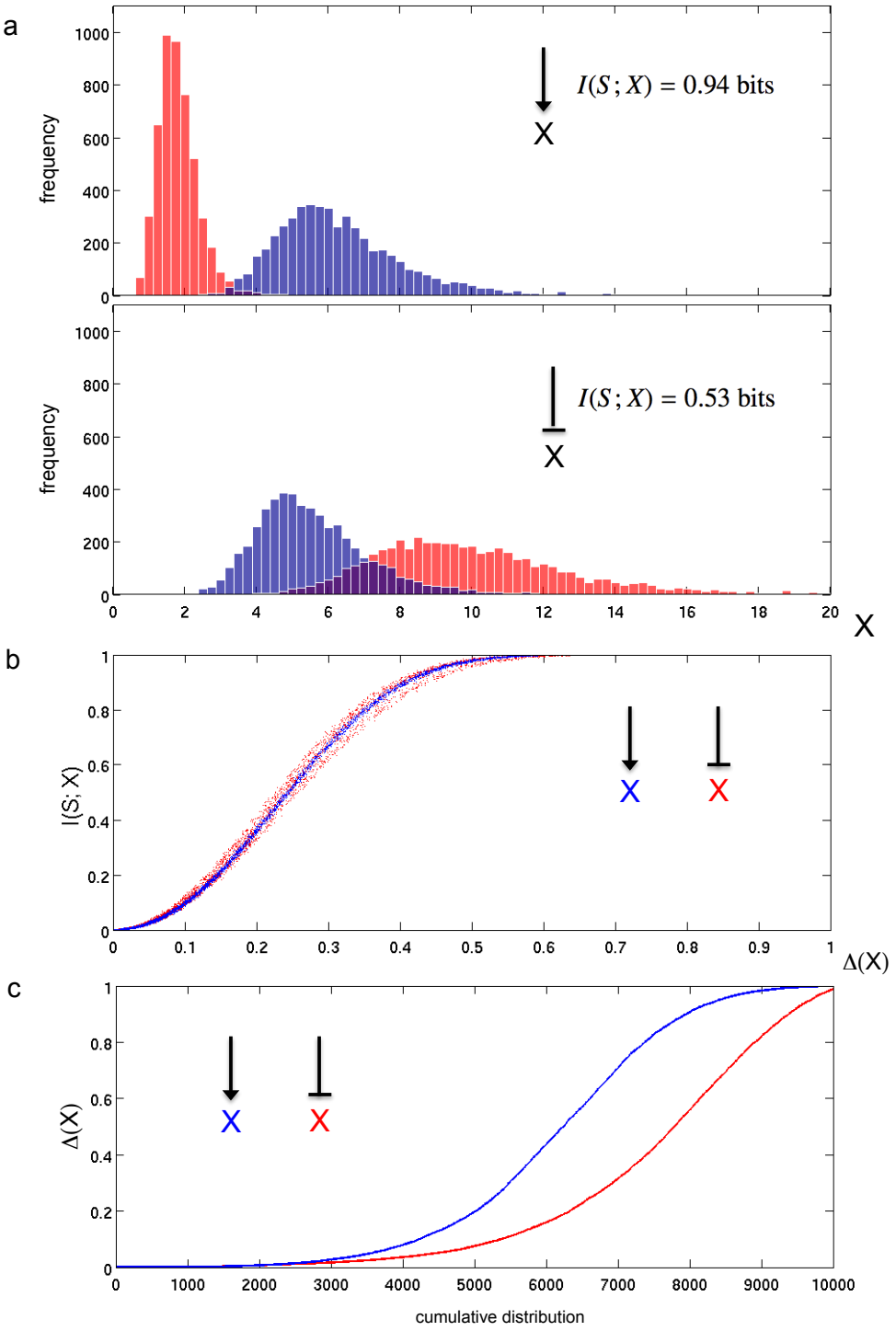

Figure 3.2: Relationship between output distributions and MI. (**a**) Output distributions produced by an activating (top) or repressing (bottom) input signal affecting a single node. The signal takes on one of two possible value, $S = 0$ (red) or $S = 1$ (blue). Histograms show 10^4 steady state output values (X) produced by perturbing the same nominal parameter set ($a_X = 2.77$, $b_X = 12.53$, $K_{SX} = 0.84$, $H_{SX} = 1$, $\sigma = 0.2$). (**b**) Scatter plot demonstrating the relationship between Δ and $I(S;X)$ for the node X subject to activating (blue) and repressing (red) inputs with $\sigma = 0.1$. (**c**) Cumulative distribution plot of $\Delta(X)$ values obtained for the activating (blue) and repressing (red) inputs.

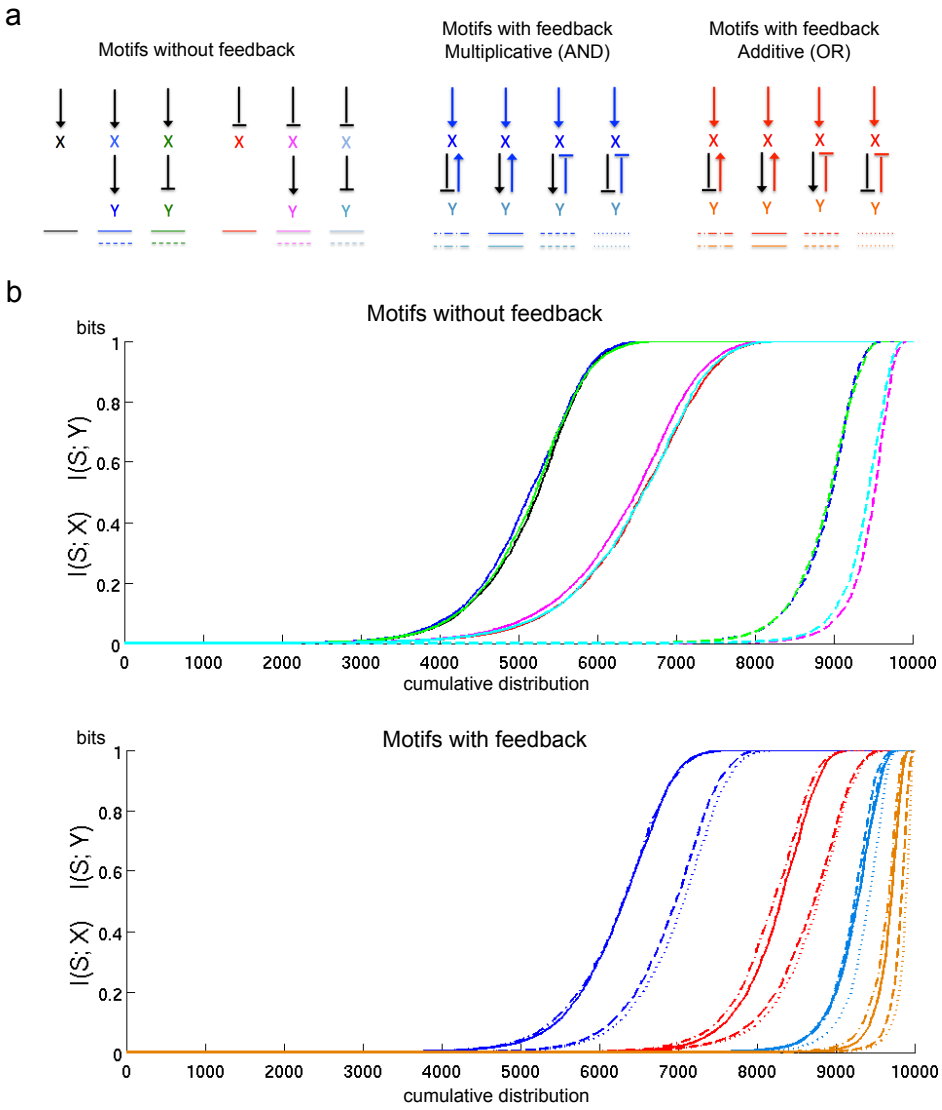

Figure 3.3: Information contained in steady state distributions. (**a**) Edges between nodes represent either activating (arrows) or repressive (blunt arrows) interactions. For nodes with two incoming edges there are two possible regulatory mechanisms, multiplicative (blue) or additive (red). Horizontal lines under each motif indicate the line types and colours used to represent the cumulative distributions of $I(S;X)$ (top line) and $I(S;Y)$ (bottom line) in **b**. (**b**) Cumulative distributions of the mutual information between the chosen input signal (S) and the steady state response at each node (X or Y). Nominal parameters were perturbed using a Gaussian kernel with $\sigma = 0.1$.

on the sensitivity of the model's steady state to the various parameters. The relative difference between steady state responses (Δ) at each node provides a convenient measure of the system's ability to produce distinct outputs:

$$\Delta(X) = \left| \frac{X_1 - X_0}{\max(X_0, X_1)} \right|, \tag{3.7}$$

where X_0 and X_1 are the steady state outputs of the model for a given nominal parameter set with $S_0 = 0$ and $S_1 = 1$ respectively. There is a clear relationship between Δ and $I(S;X)$ or $I(S;Y)$ (Figure 3.2b). However, an analytical expression for this relationship is difficult to obtain.

The main conclusions that can be drawn from the observed distributions of MI estimates and Δ values are as follows. Firstly, motifs containing activating interactions tend to produce more distinct (or distinguishable) distributions than those involving deactivating interactions (Figures 3.2c and 3.3) and thus are more likely to yield higher values of MI. Secondly, multiplicative (AND) interactions are more likely to produce distinct outputs, and thus also higher MIs, than additive (OR) interactions (Figure 3.3). Finally, adding feedback acts to decrease the probability of obtaining a high $I(S;X)$ or $I(S;Y)$ for the explored parameter range (Figure 3.3). Increasing the magnitude of extrinsic noise makes high MI values less likely, but does not change the relative ordering of the motifs with respect to their ability to transmit information about the input.

Response to a dynamically changing input

Up to this point we have only considered steady state inputs and outputs. To investigate how a temporally changing signal can be transmitted, we look to the telegraph process. This is a Markov jump process which can take one of two possible states. The probability of switching from one state to the other is determined by two constant rate parameters, k_{on} and k_{off}. We first generate the input signal trajectory by simulating a realisation of the telegraph process. This trajectory is then used to obtain a piece-wise numerical solution of the ODE model (Figure: 3.4). An independent telegraph process trajectory is generated for every simulation of the model.

Responding to signal state

We begin by considering the relationship between the state of the telegraph process (S) after some time (T_m) and the state of either the first (X) or second (Y) node of the motif

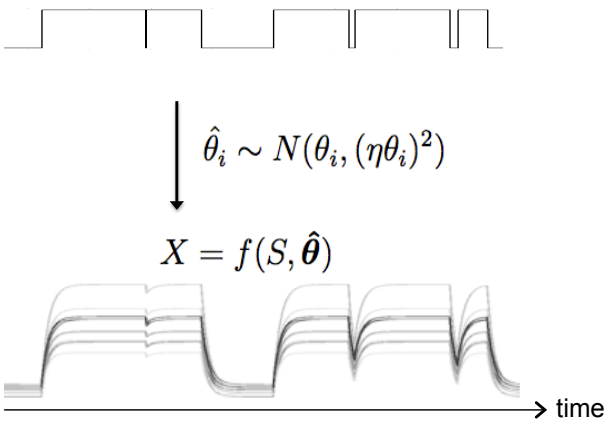

Figure 3.4: An example of how extrinsic noise may lead to different output trajectories given the same input trajectory. In this illustration, the same telegraph process trajectory is used for different instances of extrinsic noise. In the remainder of this work, a new input telegraph process trajectory is sampled for each simulation.

at the same time point (Figure 3.4).

Solving the model ODEs numerically also allows us to include perturbation of the Hill coefficient as part of our model of extrinsic noise. In addition, we repeat the simulations and analysis using fixed nominal Hill coefficient values of 2 and 4 and find that applying noise to the Hill coefficient does not cause qualitative changes in the relationship between MI and model parameters. Since the input signal takes on values of either $S = 0$ or $S = 1$, increasing the Hill coefficient exaggerates the importance of the K_m parameters. For higher Hill coefficients a lower K_m is required to produce sufficiently different outputs. As a result of this, models with a higher Hill coefficient are less likely to exhibit a high $I(S;X)$ or $I(S;Y)$ for the signal values used here.

A comparison of the cumulative probability distributions of the MI estimates shows that high MI values are less likely with a dynamically changing input. This is due to the added constraint of the output needing to respond sufficiently rapidly to the stochastic switching of the input. However, the relative order of motifs with respect to their tendency to yield a high MI values remains the same (Figure 3.5).

For all the tested motifs the degradation parameters are clear limiting factors for achieving a high $I(S;X)$. An example of this is shown in Figure 3.6. In the case of motifs involving repressive interactions the K_m parameter also constrains the maximum mutual information. However, the causes of these two dependences are different. Two features

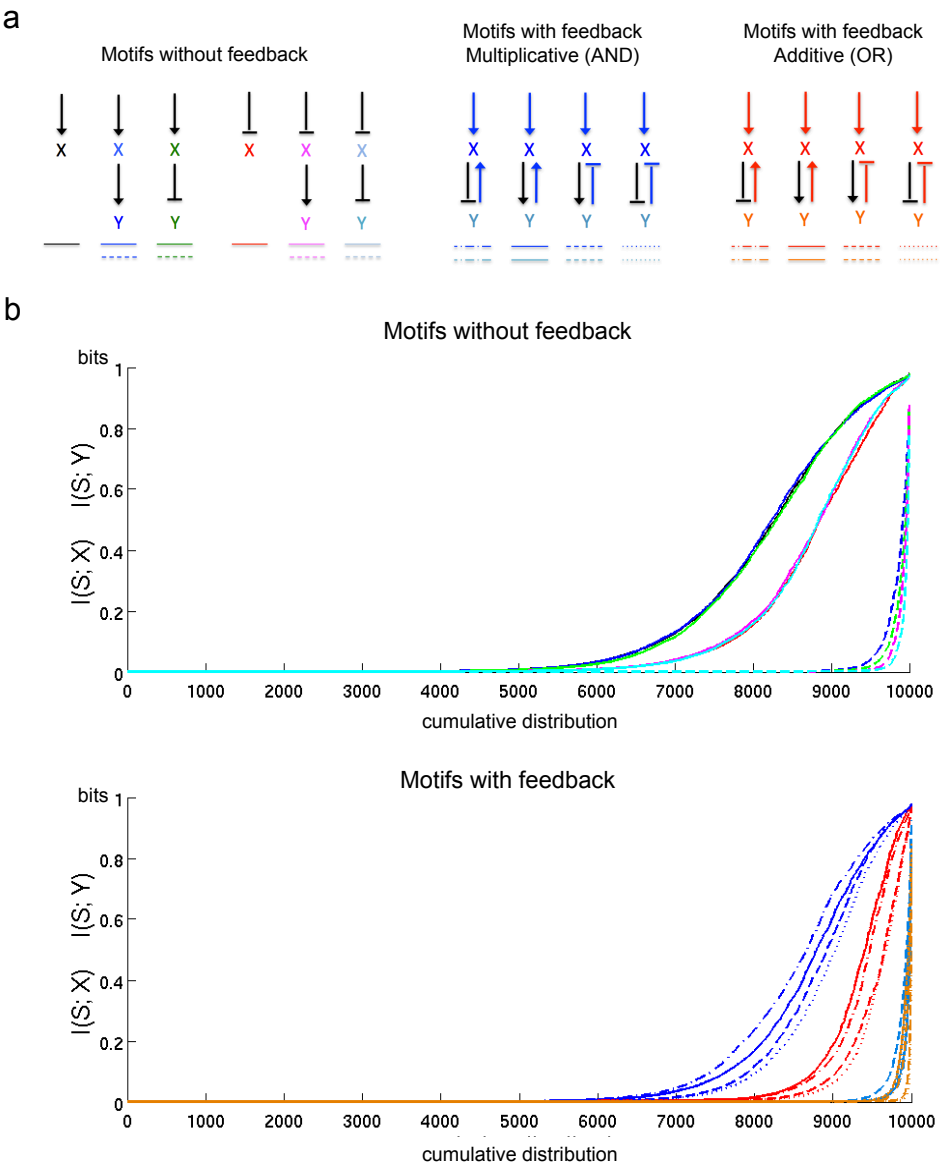

Figure 3.5: Information about a dynamically changing signal. (**a**) Edges between nodes represent either activating (arrows) or repressive (blunt arrows) interactions.(**b**) Cumulative distributions of the mutual information between the final state of the dynamically changing (S) and the final response at each node (X or Y). Nominal parameters were perturbed using a Gaussian kernel with $\sigma = 0.1$.

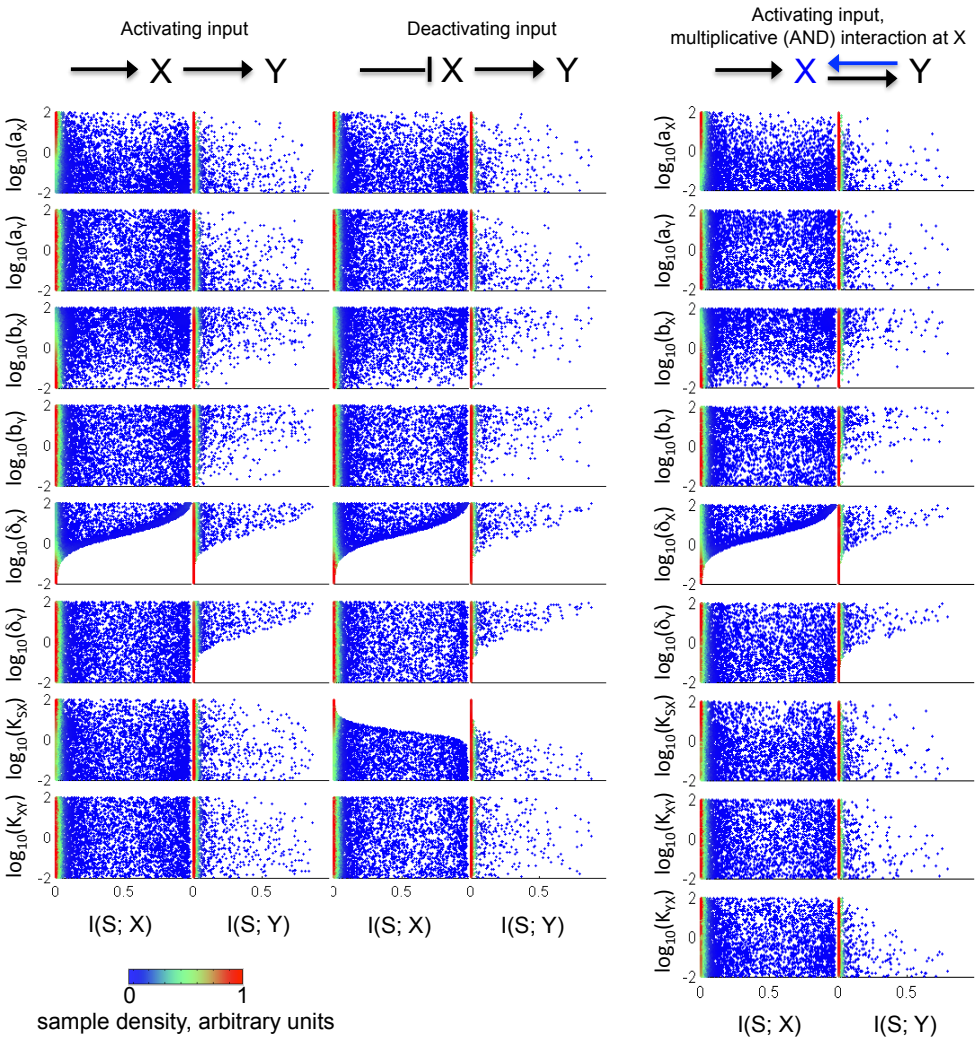

Figure 3.6: The relationships between $I(S;X)$ or $I(S;Y)$ and nominal parameter values for three, two-node motifs. Scatter plots of 10^4 sampled nominal parameters against the corresponding MI estimate. Redder hues indicate higher density.

of the responding system affect its ability to transmit information about the current state of a telegraph process, the time taken to respond to input changes, and the extent of overlap between output distributions. While the degradation rates determine the former, the remaining parameters influence the latter. Lower values of basal activation and higher promoter strengths favour a higher $I(S;X)$, although the significance of these parameters is much less than that of degradation rates and K_m values.

Response Time

For a single node, the time taken for the output to respond to a switch in the input state is determined solely by its degradation rate if the output is governed by an ODE of the following form,

$$\frac{dX}{dt} = A(\theta, S) - \delta X \tag{3.8}$$

Where A is the function determining the production rate, θ is the vector of model parameters, S is the state of the telegraph process and δ is the degradation rate. Consider node X without feedback.

$$\frac{dX}{dt} = a_X + b_X \frac{(S/K_{SX})^{H_{SX}}}{(1 + (S/K_{SX})^{H_{SX}})} - \delta_X X \tag{3.9}$$

At steady state $X = \frac{A}{\delta_X}$ where $A = a_X + b_X \frac{S}{S+K_{SX}}$.

Given an in initial steady state $X = X_0$ with $S = S_0$, let a step change in signal from S_0 to S occur at time $t = 0$.

If $X_0 = \frac{A_0}{\delta_X} = \left(a_X + b_X \frac{S_0}{S_0+K_{SX}}\right)/\delta_X$, equation (3.9) can be integrated to give,

$$X = \frac{1}{\delta_X}\left(A - (A - A_0)e^{-\delta_X t}\right) \tag{3.10}$$

Let the response time (τ) be the time taken for X to change from $\frac{A_0}{\delta_X}$ to $\frac{A_0}{\delta_X} + \epsilon\left(\frac{A}{\delta_X} - \frac{A_0}{\delta_X}\right)$ where $0 \leq \epsilon \leq 1$ is an arbitrary threshold indicating proximity to the new steady state. Using equation (3.10) the following expression for τ can be obtained,

$$\tau = -\frac{ln(1-\epsilon)}{\delta_X} \tag{3.11}$$

In the absence of extrinsic noise, the response time depends only on the degradation parameter δ_X. This is consistent with the observation that numerically estimateed MI values are strongly dependent on the degradation rates for each node. While the relationship between model parameters and the response time will be more complicated for

motifs involving feedback, high degradation rates are nonetheless a prerequisite for high $I(S\,;X)$ or $I(S\,;Y)$.

Responding to the "on" / "off" ratio

A temporally varying input can encode information in a variety of ways. As well as considering the strength of the input state at a given time point, we are interested in the relative amount of time spent in either the "on" or "off" state.

The telegraph process may be parametrised in terms of the ratio of probabilities of switching events (R) and the overall frequency with which events occur (ϕ),

$$R = \frac{k_{on}}{k_{on} + k_{off}}$$

$$\phi = 2\frac{k_{on}k_{off}}{k_{on} + k_{off}}$$

To investigate how $I(R;X)$ and $I(R;Y)$ are affected by different motif connectivities, parameters and interaction mechanisms, we repeat the above parameter sampling procedure with $R \sim U(0,1)$. In order to keep the mean number of switching events per unit time (ϕ) constant regardless of the sampled value of R, the switching rates are calculated as functions of R,

$$k_{off} = \frac{\phi}{2R}$$

$$k_{on} = \frac{k_{off}R}{1-R}$$

Changing the characteristics of the signal alters its average information content. The entropy of S is maximised when $R = 0.5$ and the current state of the signal contains, on average, one bit of information ($H(S) = 1$). While it is possible to calculate the, so called, differential entropy of a continuous random variable, this quantity does not have the same meaning as the entropy of a discrete random variable. The differential entropy may be negative and in this case $H(R) = 0$. Thus, $H(R)$ does not provide an upper limit for the amount of information that may be encoded in R. In fact, a continuous random variable may be used to encode an infinite amount of information. Conveniently, MI does not lose its meaning for the case of two continuous variables. It remains constrained to non-negative values. The amount of information about R that may be represented by some other variable is determined by the joint probability distribution of the two variables. To obtain MI estimates for two continuous variables, we employ a two-dimensional kernel density estimator.

First, consider the effects of sampling R on the transmission of information about $S = S(R)$. R corresponds to the probability of the signal being in the "on" state at a given point in time. When the value of R is fixed,

$$H(S) = -\sum_{i=0}^{1} p(S_i)\log(p(S_i)) = -\left(R\log(R) + (1-R)\log(1-R)\right)$$

In the case when R comes from a set of N discrete values,

$$H(S) = -\sum_{j=1}^{N} p(R_j)\Big(R_j\log(R_j) + (1-R_j)\log(1-R_j)\Big)$$

with $p(R_j) = 1/N$ if the distribution is uniform. For a continuous uniform distribution, $N \to \infty$ and $p(R_j) = (b-a)/N = \Delta R_j$.

$$\begin{aligned} H(S) &= -\lim_{N\to\infty}\sum_{j=1}^{N}\Delta R_j\Big(R_j\log(R_j) + (1-R_j)\log(1-R_j)\Big) \\ &= -\int_a^b R\log(R) - (1-R)\log(1-R)dR \end{aligned} \tag{3.12}$$

For $R \sim U(0,1)$, $H(S) = \frac{1}{2}$nats ≈ 0.721bits. Although the expectation $E[R] = 0.5$ corresponds to the previously fixed $R = 0.5$, the average amount of information encoded in S is lower for variable R. Since S remains a discrete variable, $H(S)$ still provides an upper limit on the information that may be encoded by it. However, this observation does not take into account limits on the response time of the system. In fact, the observed maximum $I(S;X)$ in the dynamic system under varying R is always less than half a nat (Figure 3.7).

We now turn our attention to the MI between the states of each node and R. To transmit the optimal amount of information about R, the system must integrate input states over a period of time. This can be achieved if the system does not reach a steady state between signal switching events. After some time, the distribution of X or Y will carry information about the relative amount of time spent in each state (R). Slower response times (lower degradation rates) result in higher $I(R;X)$ or $I(R;Y)$, as can be seen in the example in Figure 3.8.

However, a rapidly responding system may still transmit some information about R. In other words, knowing something about the ratio may also tell us something about the likely state of the signal at a given time. If the input signal state at a given time point is

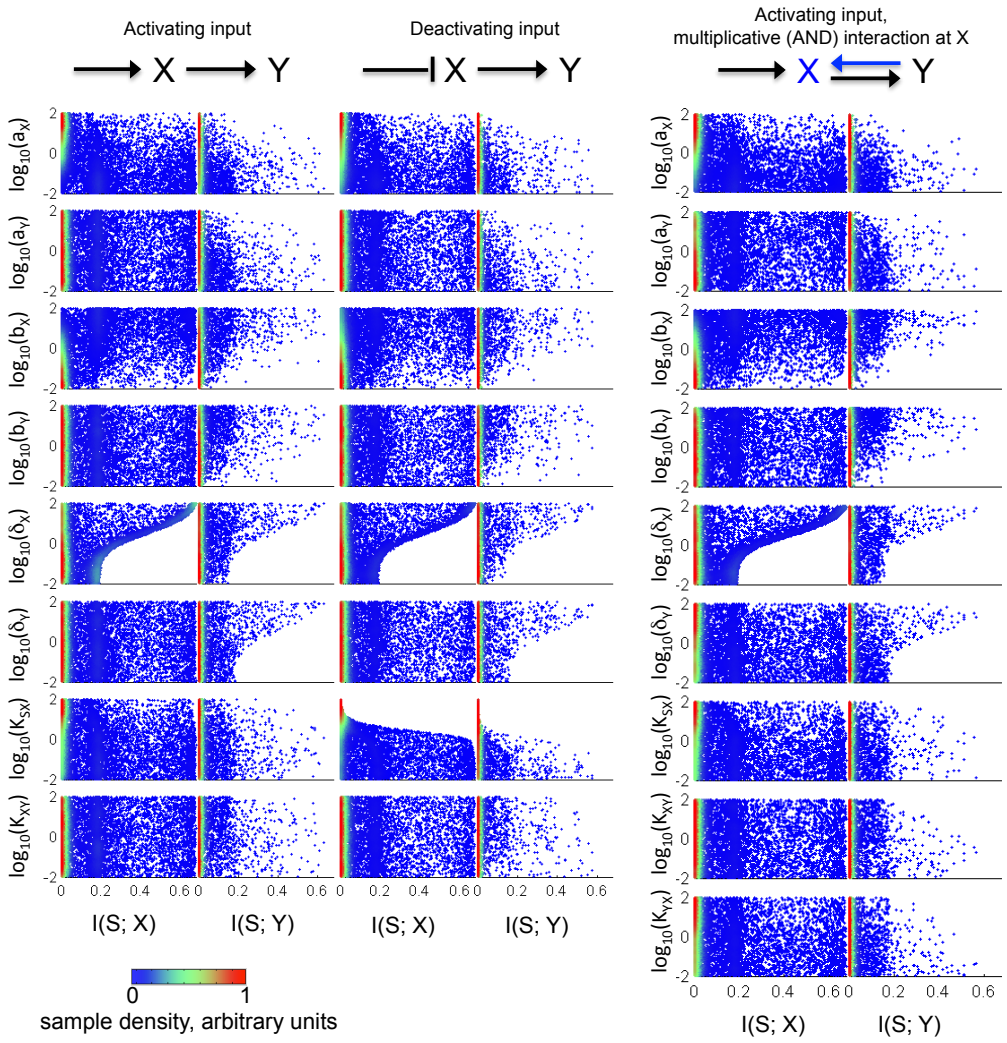

Figure 3.7: The relationships between $I(S;X)$ or $I(S;Y)$ and nominal parameter values under the condition that $R \sim U(0,1)$. Scatter plots of 10^4 sampled nominal parameters against the corresponding MI estimate. Redder hues indicate higher density.

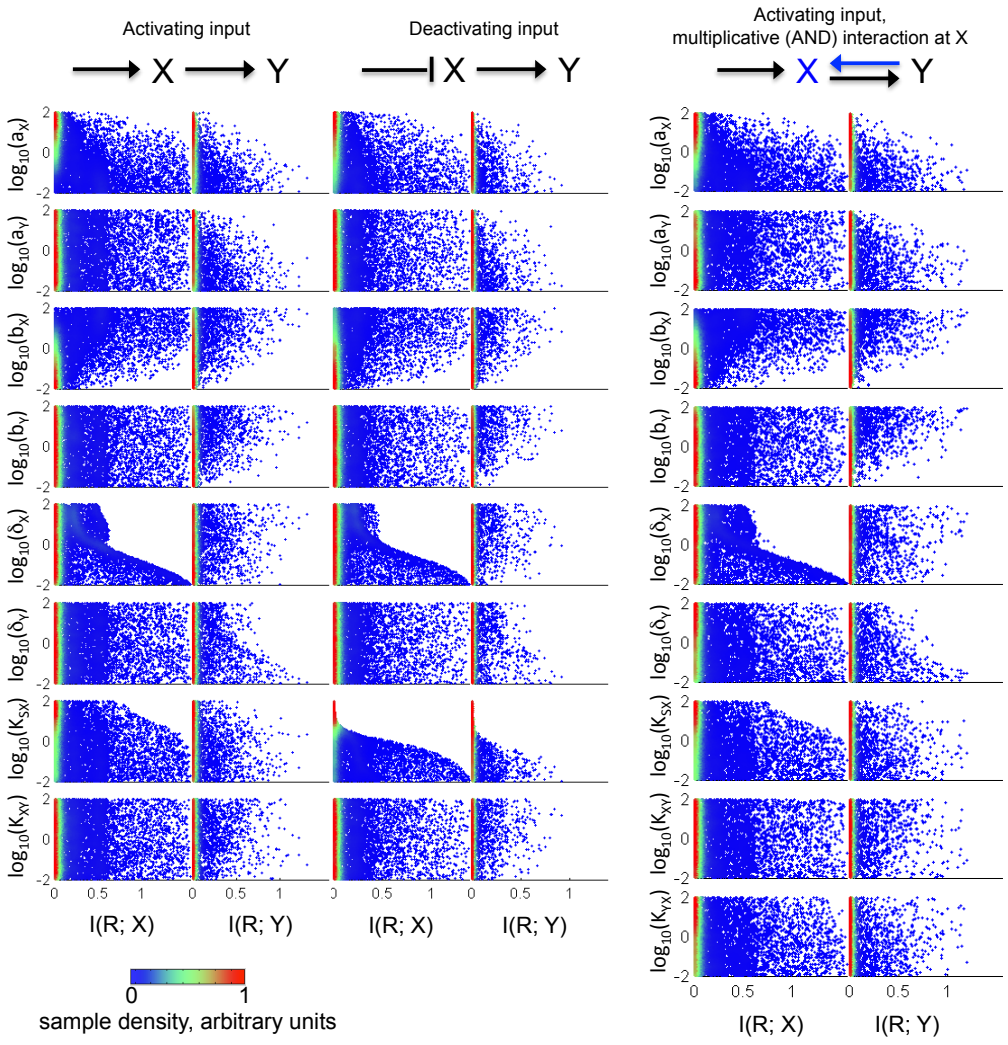

Figure 3.8: The relationships between $I(R;X)$ or $I(R;Y)$ and nominal parameter values for three, two-node motifs. Scatter plots of 10^4 sampled nominal parameters against the corresponding MI estimate. Redder hues indicate higher density.

"on", it is more likely that R is high and *vice versa*. This also explains why systems with very low degradation rates (δ_X or δ_Y) are able to represent up to 0.2 bits of information about S (Figure 3.7).

There is a trade off between retrieving information about the current state of the signal and information about the recent trend for the signal to be "on" or "off" (Figure: 3.9). It is possible for X or Y to robustly represent the current state of the signal, transmitting up to 0.5 nats of information about S, while carrying a similar amount of information about the ratio. However, transmitting close to the optimum amount of information about R requires a low δ_X or δ_Y, which precludes accurate representation of S. A signalling motif is able to retrieve close to the optimum amount of information about both S and R provided that it is relayed by separate nodes. For example, if node X carries information about S, node Y is free to average this information over a period of time, and thus, carry information about R.

3.4 Discussion

A values of $I(S;Y)$ or $I(R;Y)$ can been seen as measures of the signalling system's ability to robustly extract and transmit information about specific properties of the input signal. The observation that $I(S;Y) \leq I(S;X)$ for the steady state case can be explained in terms of the data processing inequality. In the steady state limit, the signal and the system nodes form a Markov chain, $S \rightarrow X \rightarrow Y$ as S can only influence the steady state of Y via the steady state of X. Although in the case of the dynamically changing input, the (Markovian) conditions for the data processing inequality are not satisfied, the qualitative behaviour of the system still resembles that observed for the steady state case. Specifically, X is a bottleneck through which information about S must pass before reaching Y.

The distributions of steady state outputs for equivalent parameter sets demonstrate that activation and repression are not equivalent processes when extrinsic noise is taken into account. The cumulative distributions of MI values indicate that all of the examined motifs are capable of transmitting close to the maximum possible amount of information about the state of the signal. Added complexity in the form of feedback leads to a decrease in the likelihood of a high MI. This is caused by the increase in the size of the parameter space of the system leading to a decrease in the fraction of parameter space that produces a high MI. This observation suggests that added complexity may

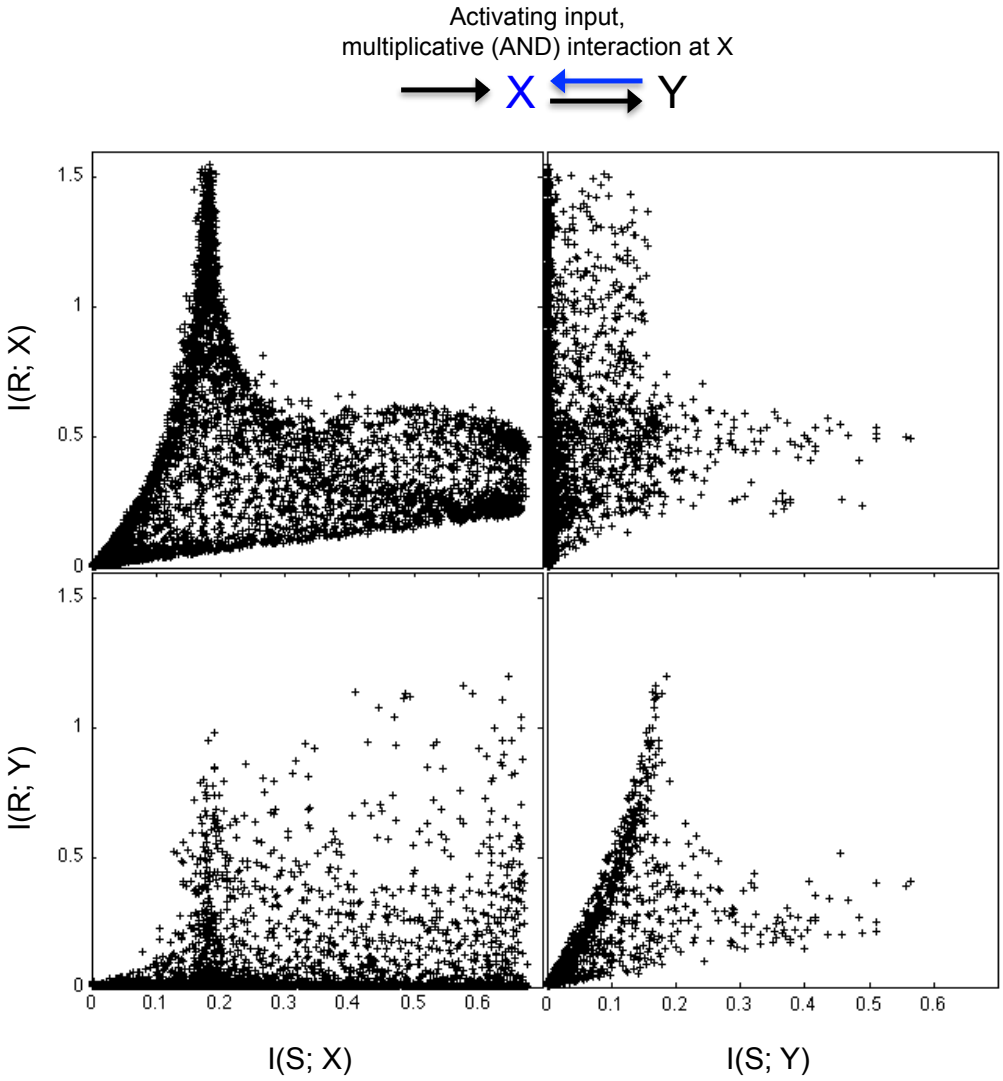

Figure 3.9: The trade-off between obtaining information about the current state of the signal and the average "on"/"off" time of the signal. Scatter plots of 10^4 sampled nominal parameters for the two node motif with multiplicative feedback and all activating interactions.

not necessarily increase the robustness of signal transmission. However, we do not try to equate the natural selection process with simple exploration of parameter space. The cumulative distributions of MI merely hint at the role that complexity may play in the evolutionary process. Another caveat to this is that these models only consider a single channel for information transition. In actual cellular signalling systems additional robustness may be provided by, so called, cross-talk between signalling pathways.

Extracting information about the ratio of times spent in each state may be relevant for situations in which downstream cellular processes are too slow to respond to every change in the signal. Alternatively, it may be advantageous to extract information about underlying properties of the signal. For example, the binding and unbinding of a transmembrane receptor to other cells in the environment may appear as a binary signal, however the relevant information may be the concentration of cells, rather than the activity of the receptor *per se*.

Here, we choose k_{on} and k_{off} parameters of the telegraph process so that the frequency of switching events remains the same. This is to exclude behaviour which may be dependent on the frequency of switching. However, this is not to say that frequency dependent behaviour has no biological relevance, but rather to consider it as a separate topic. The importance of frequency dependent behaviour has previously been demonstrated in both natural [64] and synthetic signalling pathways [65].

Systems with both a high $I(S;X)$ and high $I(R;Y)$ can be described as both transmitting and processing the input signal. Alternatively they can be viewed as transmitting information about different properties of the signal. However, it is not possible for a single node to transmit the optimum amount of information about both S and R. Obtaining information about both aspects of the signal, necessitates a division of labour between different components of the signalling pathway. Extracting an optimal amount of information about R requires loss of information about S. Hence parameter sets which yield both high $I(R;X)$ and high $I(S;Y)$ are not observed.

The significance of this work stems from the fact that it does not rely on any implicit assumptions of linearity, such as the linear noise approximation, and constitutes an exhaustive exploration of parameter space for 14 motifs representing signalling systems. The key conclusion is that a signalling motif's topology may predispose it to transmit more information about the input but all the tested topologies have the potential of achieving the maximum signal transmission fidelity.

Much attention has previously been focused on the MI between concentrations or numbers of species at a given time point [2–4, 34]. Rather than focusing on quantifying the exact number of bits transmitted, it is arguably more pertinent to consider what kind of information is relevant in each particular case. The answer to this question carries implications for how the signalling or regulatory system in question extracts information about its environment.

Chapter 4

Inferring extrinsic noise in *E. coli* gene expression

4.1 Introduction

Experiments have demonstrated the presence of considerable cell-to-cell variability in mRNA and protein numbers [66–70] and slow fluctuations on timescales similar to the cell cycle [17, 71]. Broadly speaking, there are two plausible causes of such variability. One is the inherent stochasticity of biochemical processes which are dependent on small numbers of molecules. The other relates to differences in numbers of protein, mRNA, metabolites and other molecules available for each reaction or process within a cell, as well as any heterogeneity in the physical environment of the cell population. These sources of variability have been dubbed as "intrinsic noise" and "extrinsic noise", respectively.

One of the earliest investigations into the relationship between intrinsic and extrinsic noise employed two copies of a protein with different fluorescent tags, expressed from identical promoters equidistant from the replication origin in *E. coli* [72]. By quantifying fluorescence for a range of expression levels and genetic backgrounds the authors concluded that intrinsic noise decreases monotonically as transcription rate increases while extrinsic noise attains a maximum at intermediate expression levels. Other studies have considered extrinsic noise in the context of a range of cellular processes including the induction of apoptosis [73]; the distribution of mitochondria within cells [74]; and progression through the cell cycle [75]. From a computational perspective, extrin-

sic variability has been modelled by linking the perturbation of model parameters to perturbation of the model output using the Unscented Transform [76].

Taniguchi *et al* [17] carried out a high-throughput quantitative survey of gene expression in *E. coli*. By analysing images from fluorescent microscopy they obtained discrete counts of protein and mRNA molecules in individual *E. coli* cells. They provided both the measurements of average numbers of protein and mRNA molecules in a given cell and also the measurements of cell-to-cell variability of molecule numbers. The depth and scale of their study revealed the influence of extrinsic noise on gene expression levels. The authors demonstrated that the measured protein number distributions can be described by Gamma distributions, the parameters of which can be related to the transcription rate and protein burst size [77]. To quantify extrinsic noise they consider the relationship between the means and the Fano factors of the observed protein distributions. They also illustrate how extrinsic noise in protein numbers may be attributed to fluctuations occurring on a timescale much longer than the cell cycle.

Here we aim to describe extrinsic noise at a more detailed, mechanistic, level using a stochastic model of gene expression. Such a description calls for quantitative inference of the model's parameters. We achieve this by relying on the data made available by Taniguchi *et al* and employing approximate Bayesian computation (ABC). One difficulty that arises when trying to investigate the extent and effect of extrinsic noise is that it is difficult to separate it from intrinsic noise. To overcome this confounding effect, the parameters of our model come in two varieties. Firstly, reaction rate parameters describe the probability of events occurring per unit of time. These correspond to the reaction rate parameters of a typical stochastic model which accounts for intrinsic noise. Secondly, noise parameters describe the variability in reaction rate parameters caused by the existence of extrinsic noise. This approach allows us to simultaneously infer the rate parameters and the magnitude of extrinsic noise.

Stochastic simulation and ABC inference methods are both computationally costly endeavours. In this particular case, the experimental data corresponds to snapshots of the system at a single time point. Thus, a complete temporal trajectory of the system is not necessary to carry out comparisons with the data. This allows us to make the problem computationally tractable. To this end, we develop a model-specific simulation method which takes advantage of the Poissonian relationship between the number of surviving protein molecules produced from a given mRNA molecule and its lifetime, under certain assumptions.

4.2 Modelling gene expression

A simple model of gene expression may represent the processes of transcription and translation using mass-action kinetics to describe production and degradation of various species as pseudo-first order reactions. Such a model may be simulated stochastically to take into account the intrinsic variability of processes involving low numbers of molecules. In the simplest version of this model, mRNA is produced from the promoter at a constant rate. However, such Poissonian mRNA production is often not sufficient to account for the variability in mRNA numbers measured experimentally in both prokaryotic and eukaryotic cells. In addition to this, for many genes, transcription appears to occur in bursts rather than at a constant rate. These characteristics of gene expression have been observed in organisms as diverse as bacteria [17], yeast [69], amoeba [67] and mammals [68]. One model of gene expression that takes this into account is the, so called, two-state model.

The two-state promoter model

In the two-state model of gene expression, a gene's promoter is represented as either active or inactive [70, 78]. Here we use a variant of the two-state model with the inactive state corresponding to a lower transcription rate rather than no transcription at all. For each state of the promoter, transcription events at that promoter are represented by a Poisson process with rate parameter corresponding to the transcription rate. Biochemical processes such as transcription factor binding or reorganisation of chromatin structure may account for the existence of several distinct levels of promoter activity. However, which factors play a dominant role in the apparent switching, remains an unanswered question.

The Gillespie algorithm [13] may be used to simulate all the reactions represented by this model and obtain a complete trajectory of the system through time. However, in this case we are only interested in the number of molecules present at the time of measurement. We use a model-specific stochastic algorithm (Algorithm 5) which allows us to reduce the number of computational steps required to obtain a single realisation from the model.

The following reactions, represented using mass-action kinetics, comprise the two-state

model:

$$
\begin{aligned}
\texttt{inactive-promoter} &\xrightarrow{k_{\text{on}}} \texttt{active-promoter} \\
\texttt{active-promoter} &\xrightarrow{k_{\text{off}}} \texttt{inactive-promoter} \\
\texttt{inactive-promoter} &\xrightarrow{k_0} \texttt{inactive-promoter} + \texttt{mRNA} \\
\texttt{active-promoter} &\xrightarrow{k_1} \texttt{active-promoter} + \texttt{mRNA} \\
\texttt{mRNA} &\xrightarrow{k_2} \texttt{mRNA} + \texttt{Protein} \\
\texttt{mRNA} &\xrightarrow{d_1} \emptyset \\
\texttt{Protein} &\xrightarrow{d_2} \emptyset
\end{aligned}
$$

The propensity functions (hazards) for each of the above reactions are listed below:

$$
\begin{aligned}
h_0 &= k_{\text{on}}[\texttt{inactive-promoter}] \\
h_1 &= k_{\text{off}}[\texttt{active-promoter}] \\
h_2 &= k_0[\texttt{inactive-promoter}] \\
h_3 &= k_1[\texttt{active-promoter}] \\
h_4 &= k_2[\texttt{mRNA}] \\
h_5 &= d_1[\texttt{mRNA}] \\
h_6 &= d_2[\texttt{Protein}]
\end{aligned}
$$

Here the square brackets refer to the number of molecules of a species rather than its concentration.

The model presented here relies on a number of assumptions about the process of gene expression. Firstly, that the production of mRNA and protein can be described sufficiently well by pseudo-first-order reactions. Secondly, that degradation of mRNA and protein can be described as an exponential decay. In a bacterial cell, mRNA molecules are degraded enzymatically and typically have a half-life on the scale of several minutes. The half-life of protein molecules usually exceeds the time required for cell growth and division during the exponential growth phase. Thus, dilution due to partitioning of protein molecules between daughter cells tends to be the dominant factor in decreasing the number of protein molecules. Here we do not build an explicit model of cell division,

instead the decrease in protein numbers is approximated by an exponential decay. Finally, it is assumed that there is no feedback mechanism by which the number of mRNA or protein molecules produced by the gene affects its promoter switching, transcription or translation rates.

Representing extrinsic noise

Here we model extrinsic noise using a similar approach to that in the previous chapter on models of two-node regulatory motifs. Extrinsic noise is represented by using a Gaussian kernel to perturb the reaction rate parameters before each simulation of the model. The effect of extrinsic noise on each reaction is assumed to be independent. The reaction rates associated with a particular gene are termed nominal parameters (θ_n).

$$\theta_n = [k_{\text{on}}, k_{\text{off}}, k_0, k_1, k_2, d_1, d_2]$$

The values determining the magnitude of the perturbation are termed the noise parameters (η).

$$\eta = [\eta_{k_{\text{on}}}, \eta_{k_{\text{off}}}, \eta_{k_0}, \eta_{k_2}, \eta_{d_1}, \eta_{d_2}]$$

Together they comprise the full parameter set for the model $\theta = [\theta_n, \eta]$.

In the case of the two-state model of a single gene, each θ_n has a corresponding extrinsic noise parameter with the exception that the basal transcription rate (k'_0) is defined as a fraction of the active transcription rate (k'_1) so the two reaction rates are subject to the same perturbation (η_{k_0}) before each simulation. This is motivated by the idea that extrinsic factors affecting the transcription rate do not depend on the state of the promoter. The parameters used to generate a single realisation from the two-state model are obtained by sampling from $f(\mu, \sigma)$. Where f is a truncated normal distribution, restricted to non-negative values by rejection sampling with μ and σ being the mean and standard

deviation of the corresponding normal distribution.

$$
\begin{aligned}
k'_{\text{on}} &\sim f\left(k_{\text{on}}, k_{\text{on}}\eta_{k_{\text{on}}}\right) \\
k'_{\text{off}} &\sim f\left(k_{\text{off}}, k_{\text{off}}\eta_{k_{\text{off}}}\right) \\
k'_1 &\sim f\left(k_1, k_1\eta_{k_1}\right) \\
k'_0 &= k_0 k'_1 \\
k'_2 &\sim f\left(k_2, k_2\eta_{k_2}\right) \\
d'_1 &\sim f\left(d_1, k_1\eta_{d_1}\right) \\
d'_2 &\sim f\left(d_2, k_2\eta_{d_2}\right)
\end{aligned}
$$

The final time point of each simulation represents the number of mRNA and protein molecules in a single cell at the time of measurement.

Simulation procedure

In order to reduce the computational cost of each simulation, rather than using Gillespie's direct method to simulate the entire trajectory of mRNA and protein numbers, we employed Algorithm 5 to obtain samples of the numbers of mRNA and protein molecules at the time of measurement (t_m). First, a realisation of the telegraph process is used to obtain the birth and decay times of mRNA molecules. These are then used to sample the number of protein molecules that were produced from each mRNA molecule and survived until t_m. This procedure makes use of the Poissonian relationship between the life time of an individual mRNA molecule and the number of surviving protein molecules that were produced from it. This relationship is derived in Appendix 6.4 and its use is illustrated in Figure 4.1. The final output is the number of both mRNA (M) and protein (P) molecules present in the system at t_m.

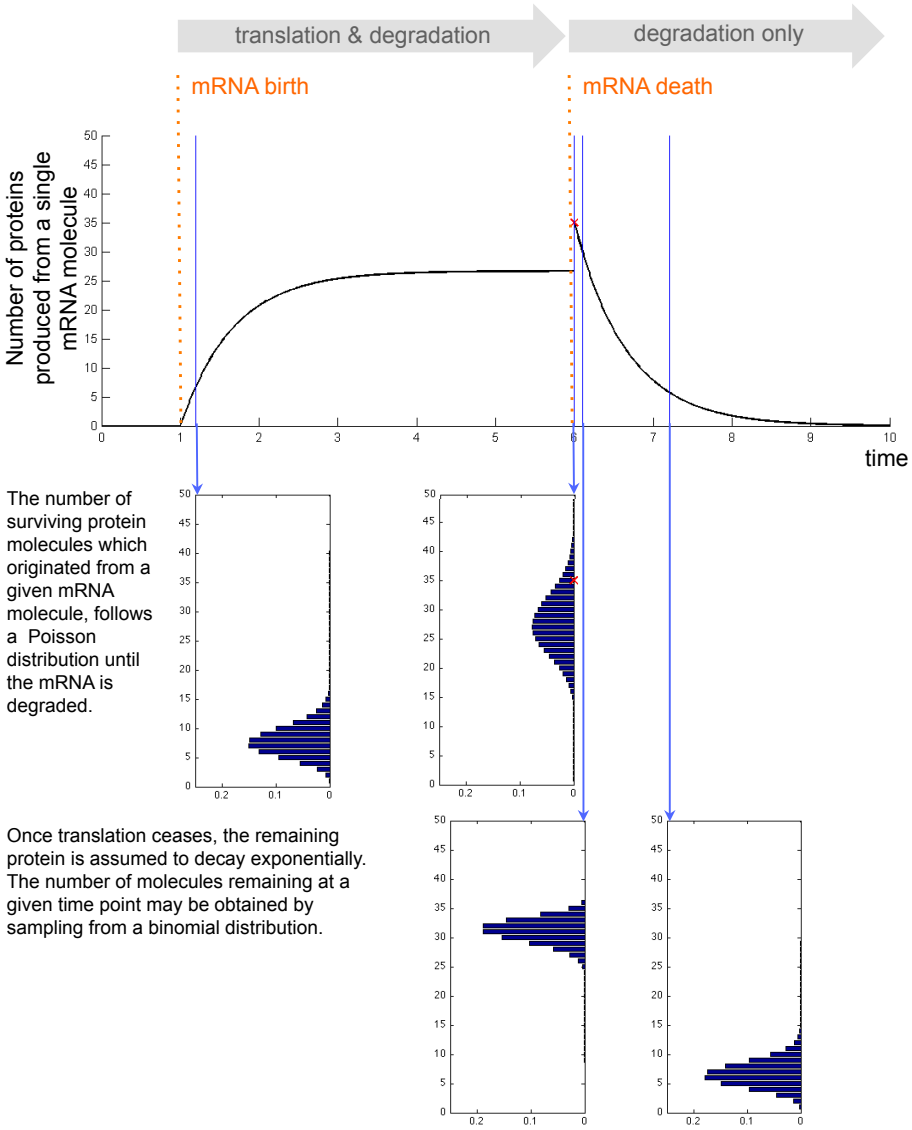

Figure 4.1: An illustration of how the birth and death times of an mRNA molecule are used to obtain the number of proteins that were produced from it and then survived until the time at which mRNA and protein numbers were measured.

Algorithm 5 Simulation of the two-state model

Inputs: θ_n, η, t_m
Outputs: M, P

Obtain perturbed parameters using the nominal (θ_n) and noise (η) parameters.
Stage one: simulate mRNA production subject to an underlying telegraph process.
$S \leftarrow 1$ with probability $k'_{\text{on}}/(k'_{\text{on}} + k'_{\text{off}})$), otherwise $S \leftarrow -1$ ▷ Select the initial state of the telegraph process.
$t \leftarrow 0$ ▷ Initialise simulation time.
$M_b \leftarrow 0$ ▷ Initialise the number of mRNA molecules produced.
$i \leftarrow 1$ ▷ Initialise index of mRNA molecules.
while $t < t_m$ **do**
    if S = -1 **then**
        $k_S \leftarrow k'_{\text{on}}$
        $k_m \leftarrow k'_0$
    else
        $k_S \leftarrow k'_{\text{off}}$
        $k_m \leftarrow k'_1$
    end if
    $\tau \sim Exp(k_S)$ ▷ Sample the time until the next switching event.
    if $t + \tau > t_m$ **then** ▷ Ensure that $t + \tau$ does not exceed the final time point.
        $\tau \leftarrow t_m - t$
    end if
    $M_\tau \sim Poisson(\tau k_m)$ ▷ Sample the number of mRNA molecules produced.
    $M_b \leftarrow M_b + M_\tau$
    while $i \leq M_b$ **do**
        $u_i \sim Uniform(t, t + \tau)$ ▷ Sample birth times for each mRNA.
        $i \leftarrow i + 1$
    end while
    $t \leftarrow t + \tau$
    $S \leftarrow -S$
end while

Stage two: simulate mRNA degradation; protein production and degradation.

$M \leftarrow 0$ ▷ Initialise the number of mRNA molecules at t_m.

$P \leftarrow 0$ ▷ Initialise the number of protein molecules at t_m.

$i \leftarrow 1$

while $i \leq M_b$ **do** ▷ For each mRNA molecule that was produced:

  $v \sim Exp(d_1')$ ▷ Sample the time until mRNA decay.

  $T_l \leftarrow min(u_i + v; t_m) - u_i$ ▷ Calculate mRNA lifetime.

  $P_l \sim Poisson\left(\frac{k_2'}{d_2'}(1 - e^{-d_2' T_l})\right)$
  ▷ Sample the number of surviving proteins at time point $u_i + v$.

  $T_d \leftarrow t_m - min(u_i + v; t_m)$ ▷ Time since mRNA decay.

  if $T_d = 0$ **then**

    $M \leftarrow M + 1$ ▷ mRNA survived until t_m.

    $P \leftarrow P + P_l$

  else

    $P_s \sim Binomial(P_l, e^{-d_2' T_d})$
    ▷ Sample the number of surviving proteins at time t_m.

    $P \leftarrow P + P_s$

  end if

  $i \leftarrow i + 1$

end while

Taniguchi *et al* [17] used images of about a thousand cells to obtain estimates of mean mRNA numbers, mRNA Fano factors, mean protein numbers and protein number variances. For this reason, we use 10^3 simulation runs when calculating summary statistics. The experimental measurements of mRNA lifetimes are compared directly to the mRNA degradation rate parameter (d_1) in the model by assuming that lifetimes correspond to the inverse of the decay rate.

The run time of each simulation largely depends on the number of mRNA molecules produced during the simulated time period and thus on the model parameters. In the dataset considered here, mRNA numbers in the cell at any one time are in the low tens while the corresponding protein may be present in hundreds or thousands of copies. There is a significant computational advantage to using algorithm (5) in this context. Nonetheless, inferring the posterior distribution for a given gene can take several days or even weeks of computational time on a modern processor core.

Inference procedure

We use an ABC-SMC algorithm to infer plausible parameter sets for the two-state model based on the experimental data. The inference procedure is similar to that employed by [49, 50, 79], as described in Algorithm 6.

For the distance metric, d, we take the Euclidean distance between the logarithms of each type of experimental measurement (D_i) and the corresponding simulation results (x_i):

$$d(D, x) = \sqrt{\sum_{i=1}^{i=5} (\log D_i - \log x_i)^2}$$

$$D = \left[\mu_{mRNA}, \frac{\sigma^2_{mRNA}}{\mu_{mRNA}}, \mu_{prot}, \sigma^2_{prot}, \tau^{-1}_{mRNA}\right]$$

Where μ_{mRNA} is the mean number of mRNA molecules; $\sigma^2_{mRNA}/\mu_{mRNA}$ is the Fano factor of the mRNA distribution; μ_{prot} is the mean number of protein molecules; σ^2_{prot} is the variance of the protein distribution and τ^{-1}_{mRNA} gives the exponential decay rate constant for mRNA degradation based on the measured mRNA lifetime (τ_{mRNA}).

$$x = \left[\mu_M, \frac{\sigma^2_M}{\mu_M}, \mu_P, \sigma^2_P, d_1\right]$$

Where μ_M is the mean number of mRNA molecules; σ^2_M/μ_M is the Fano factor of the mRNA distribution; μ_P is the mean number of protein molecules; σ^2_P is the variance of the protein distribution and d_1 corresponds to the nominal mRNA degradation rate. The first sampled population of particles (population zero in Algorithm 6), provides a benchmark for the choice of ϵ values in the next population. Since we have no knowledge of the distribution of distances until a set of particles is sampled, all particles are accepted in the first population. For subsequent populations, ϵ values are chosen such that the probability of acceptance with the new ϵ value is equal to q_t. The vector q is chosen prior to the simulation. This allows for larger decreases in ϵ in the first few populations while keeping the actual epsilon values used, a function of the distances (g) in the previous population. New populations are sampled until the final epsilon value is reached $\epsilon_f = 0.1$. To obtain θ^* from θ we use a uniform perturbation kernel:

$$\theta^* \sim U(\theta - \mu_{t-1}, \theta + \mu_{t-1})$$

where μ_{t-1} is the vector of standard deviations of each parameter in the previous population.

Algorithm 6 ABC SMC with summary statistics

Inputs: π, N, ϵ_f
Outputs: Set of populations of N accepted particles

$i \leftarrow 1$
$t \leftarrow 0$
$q \leftarrow [0.01, 0.05, 0.25, 0.75, ..., 0.75]$
Initialise ϵ vector.
while $\epsilon > \epsilon_f$ **do**
    if $t = 0$ **then**
        while $i \leq N$ **do**
            Sample a new θ from π.
            Simulate from the model 10^3 times according to Algorithm 5.
            Calculate summary statistics, x, from the simulation outputs.
            if $d(D, x) < \epsilon$ **then**
                Accept particle.
                $\omega^{(i,t)} \leftarrow 1$
                $i \leftarrow i + 1$
            end if
        end while
    else
        while $i \leq N$ **do**
            Sample θ from $\{\theta^{(j,t-1)}\}_{1 \leq j \leq N}$ with probability $\{\omega^{(j,t-1)}\}_{1 \leq j \leq N}$.
            Perturb θ to obtain θ^*.
            Simulate from the model 10^3 times according to Algorithm 5.
            Calculate summary statistics, x, from the simulation outputs.
            if $d(D, x) < \epsilon$ **then**
                Accept particle

$$\omega^{(i,t)} \leftarrow \frac{\pi(\theta^{(i,t)})}{\sum_{j=1}^{n} \omega^{(j,t)} K_t(\theta^{i,t} | \theta^{(j,t-1)})}$$

                $i \leftarrow i + 1$
            end if
        end while
    end if
    Normalise weights.
    $t \leftarrow t + 1$
    Set ϵ such that $Pr(g_i \leq \epsilon_i) = q_t$
end while

Use of experimental data

We focus on a subset of 87 genes from the published data set from [17]. These are all the genes for which, in addition to protein numbers, the experimental data include both fluorescence *in situ* hybridization (FISH) measurements [80] of mRNA numbers and mRNA lifetimes measurements obtained using RNAseq [81]. We are able to directly use the summary statistics provided in the supplementary data of [17]. We note that these genes are not a random sample from the set of all genes and exhibit higher than average expression levels.

To identify model parameters for which the two-state model, with extrinsic noise, is able to reproduce the experimental measurements, we carry out Bayesian inference using an ABC-SMC algorithm that compares summary statistics from simulated and experimental data [79]. Specifically we used the following summary statistics: (1) the mean numbers of mRNA molecules; (2) the Fano factors of mRNA molecule distributions; (3) the mean numbers of protein molecules; (4) the variances of protein molecule numbers; and (5) mRNA lifetimes converted to expontial decay rate parameters. The distributions of these summary statistics are shown in Figure 4.2. We assume that the summary statistics correspond to steady state expression levels for each gene. While there is no guarantee that this is the case for every gene, the majority of genes are unlikely to be undergoing major changes in their expression level given that the cells are in a relatively constant environment.

Parameter prior

The telegraph process may be parametrized in terms of the ratio of probabilities of switching events (k_r) and the overall frequency with which events occur (k_f):

$$k_r = \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}}$$

$$k_f = 2\frac{k_{\text{on}}k_{\text{off}}}{k_{\text{on}} + k_{\text{off}}}$$

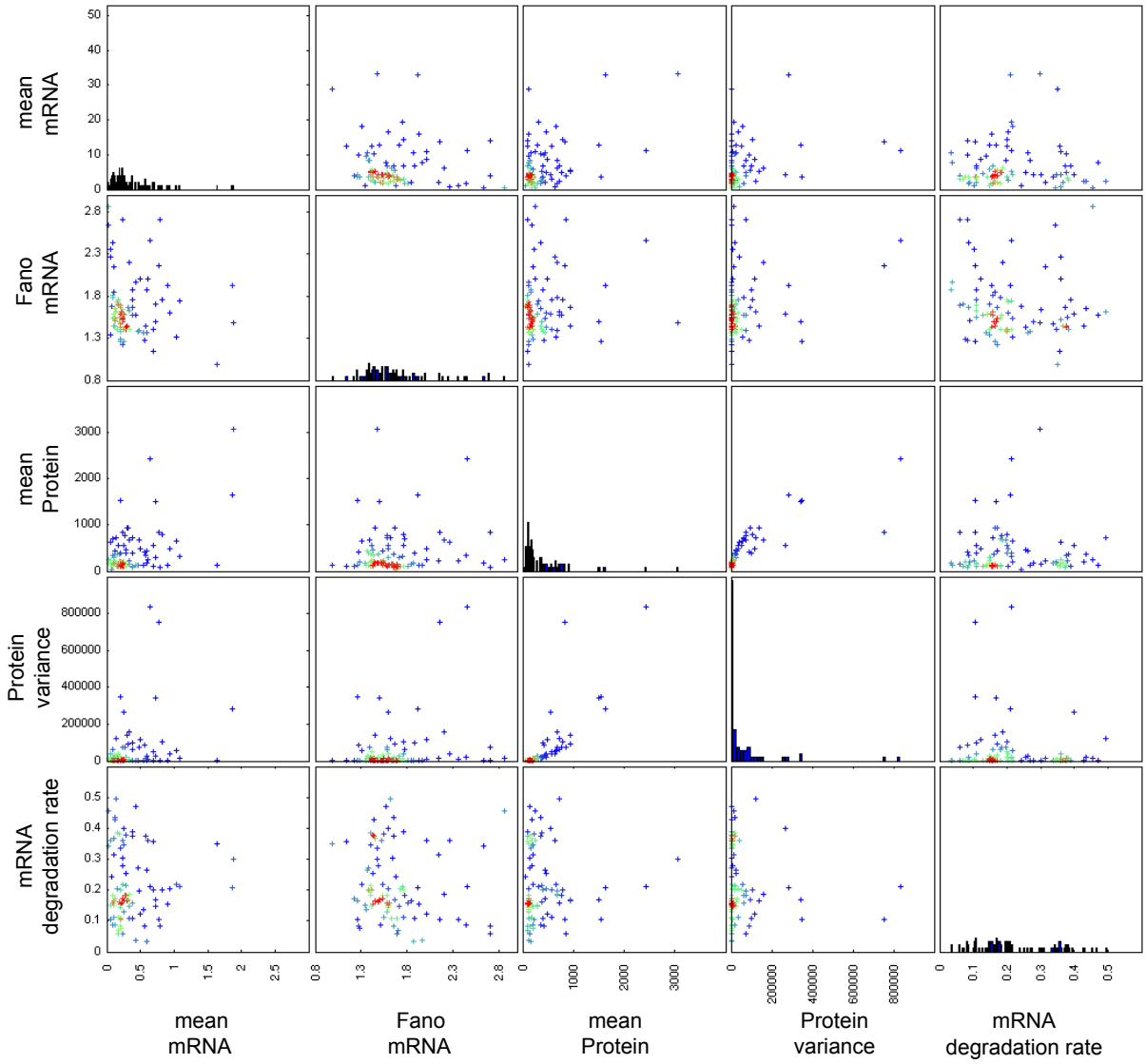

Figure 4.2: Experimentally measured summary statistics. Each point on the scatter plots is an estimate of the corresponding summary statistic or mRNA lifetime experimental measurement. Histograms on the diagonal all use the same vertical axis as the top left most histogram.

To obtain θ, the vector of parameters used in the ABCSMC inference procedure (Algorithm 6), rate and noise parameters are sampled from the following uniform priors,

$$
\begin{aligned}
k_r &\sim U(0, 1) \\
k_f &\sim U(0, 0.1) \\
k_0 &\sim U(0, 1) \\
k_1 &\sim U(0, 1) \\
k_2 &\sim U(0, 10) \\
d_1 &\sim U(0.01, 0.6) \\
d_2 &\sim U(0.0005, 0.05) \\
\eta_{k_{\text{on}}} &\sim U(0, 0.5) \\
\eta_{k_{\text{off}}} &\sim U(0, 0.5) \\
\eta_{k_1} &\sim U(0, 0.4) \\
\eta_{k_2} &\sim U(0, 0.4) \\
\eta_{d_1} &\sim U(0, 0.4) \\
\eta_{d_2} &\sim U(0, 0.4).
\end{aligned}
$$

The parameters for the telegraph process, sampled from the prior as k_r and k_f, are converted to k_{on} and k_{off} before being passed to the simulation algorithm (Algorithm 5) as follows,

$$
\begin{aligned}
k_{\text{off}} &= \frac{k_f}{2k_r} \\
k_{\text{on}} &= \frac{k_{\text{off}} k_r}{1 - k_r}.
\end{aligned}
$$

Rate parameters k_r and k_0 as well as the noise parameters (η) are unit-less. The remaining parameters have units $1s^{-1}$.

To ensure that M and P are from a distribution close to equilibrium, simulation duration is set depending on the nominal degradation rates for mRNA (d_1) and protein (d_2),

$$
t_m = L(d_1^{-1} + d_2^{-1})
$$

where t_m is the final time point and L is a constant chosen arbitrarily to indicate the desired proximity to the steady state distribution. Here we use $L = 5$.

4.3 Results and discussion

Posterior distributions of parameters

We begin our analysis by examining the posterior distributions of parameters obtained for each gene using the ABC-SMC inference procedure. The simulated summary statistics converged to within the desired threshold of the experimental measurements for 86 out of 87 genes. The inferred posterior for the one remaining gene converged relatively slowly and we chose to terminate the process after 30 days of computational time. Figure 4.3 shows a contour plot of the distribution of summary statistics and the mRNA degradation rate, obtained from particles in the final ABC-SMC population for a typical gene (*dnaK*). Other examples can be seen in Figures 1, 2, 3 and 4. We begin with a discussion of features of the posterior parameter distributions, that are common to most genes. Next, we examine the relationships between model parameters and summary statistics of the model outputs. Lastly, we carry out a type of sensitivity analysis on the inferred posteriors.

In the two-state model, the switching of the promoter between active and inactive states is described by a telegraph process which can be parametrised either in terms of the switching reaction rates (k_{on} and k_{off}) or in terms of the on/off bias (k_r) and frequency of switching events (k_f). The simulation algorithm takes parameters in the form of k_{on} and k_{off}. However, the effects of k_r and k_f on the observed mRNA distribution may be interpreted more directly and intuitively.

For the majority of genes the k_0 and k_r parameters are relatively small. This appears to be a prerequisite for a high Fano factor of the mRNA distribution and the mean marginal inferred values of these parameters are negatively correlated with Fano factors across all 86 genes as discussed below. A low switching rate combined with a low basal expression rate ensures that there are two distinct mRNA expression levels. This in turn produces a larger variance in measured mRNA counts and results in Fano factor values well above one. Conversely, genes for which mRNA production appears to be more Poissonian were inferred to have basal mRNA production rates close to one, i.e. similar to the active mRNA production rates. In other words, these genes appear to be constitutively active. Here again, we point out that the two-state promoter model provides a convenient abstraction and a hypothesis for explaining the super-Poissonian variance in mRNA copy number [70, 78]. However, based on these observations it is difficult

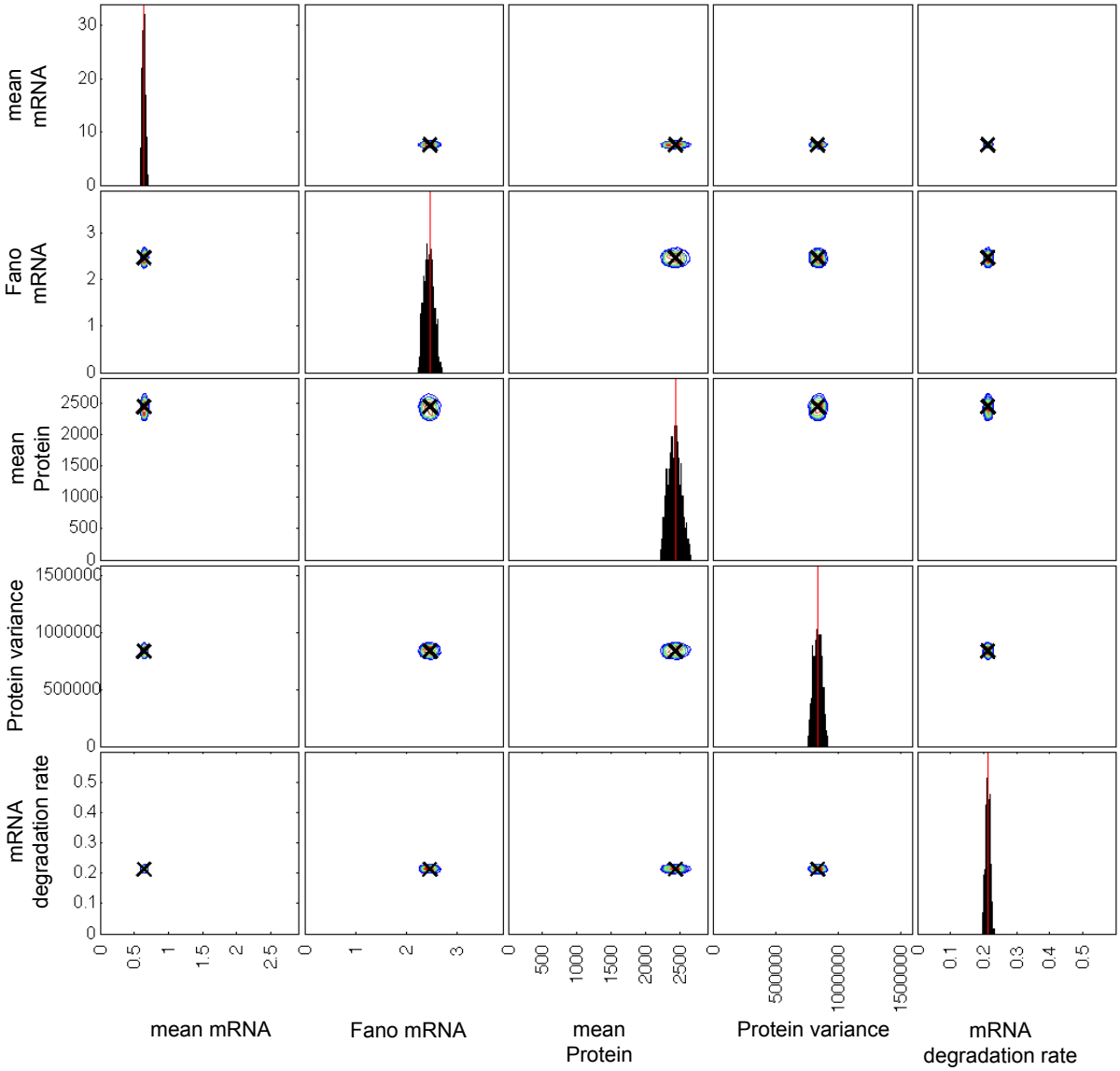

Figure 4.3: Posterior distribution of summary statistics for the gene *dnaK*. Contour plots indicating the density of points with the corresponding summary statistic for each particle in the final population.

to determine whether a model with more states or some other more elaborate regulatory model, would not be more appropriate. Our preliminary attempts at carrying out the inference procedure with a one-state model indicate that extrinsic noise alone does not explain the observed mRNA distributions without also producing unacceptably high variability in protein numbers.

Our initial inference attempts used only the summary statistics from the data. We observed that the production and degradation rate parameters for mRNA (k_1 and d_1) and protein (k_2 and d_2) tended to be positively correlated in the posterior parameter distributions of many genes. This is due to limited identifiability of model parameters since different combinations of rates may produce similar steady state expression levels. The problem is partly alleviated by using data on the mRNA degradation rate to constrain the range of acceptable values for d_1 for each gene (Figure 4.4). In addition, the model of extrinsic noise prevents extremely high rate parameter values from being accepted since they would result in greater extrinsic noise. Despite these limitations, our approach does provide an indication of the possible range of extrinsic noise values that can account for the observed variability in mRNA and protein numbers.

Although the posterior summary statistics (and mRNA degradation rate) are reasonably well constrained and distinct for each gene, the distributions of model parameters may still be relatively broad (Figure 4.4). There are a number of reasons for this. Firstly, changes in parameters associated with active transcription and translation, as well as degradation rates, are more easily inferred than parameters describing switching between promoter states, basal transcription or extrinsic noise. In particular, when the production and degradation rates for the same species are subjected to differing extrinsic noise parameters, the inference procedure struggles to resolve between the different source of extrinsic noise. This explains the correlation between the means of inferred extrinsic noise parameters (Figure 4.5). Such correlations between extrinsic noise parameters are not observed in the posterior of each gene or when taking the single particle with the highest weight from the final population of each gene as in Figure 4.6.

A comparison of Figues 4.5 and 4.6 suggests that a certain level of extrinsic noise is expected for all genes. However, the extrinsic noise may affect various combinations of rate parameters and it may not be possible to discern if, for example, the production rate or the degradation rate is more affected by extrinsic variability. While our inference procedure does not indicate a distinctive lower boundary for the amount of extrinsic noise affecting each reaction rate, there is usually an upper limit to the inferred noise

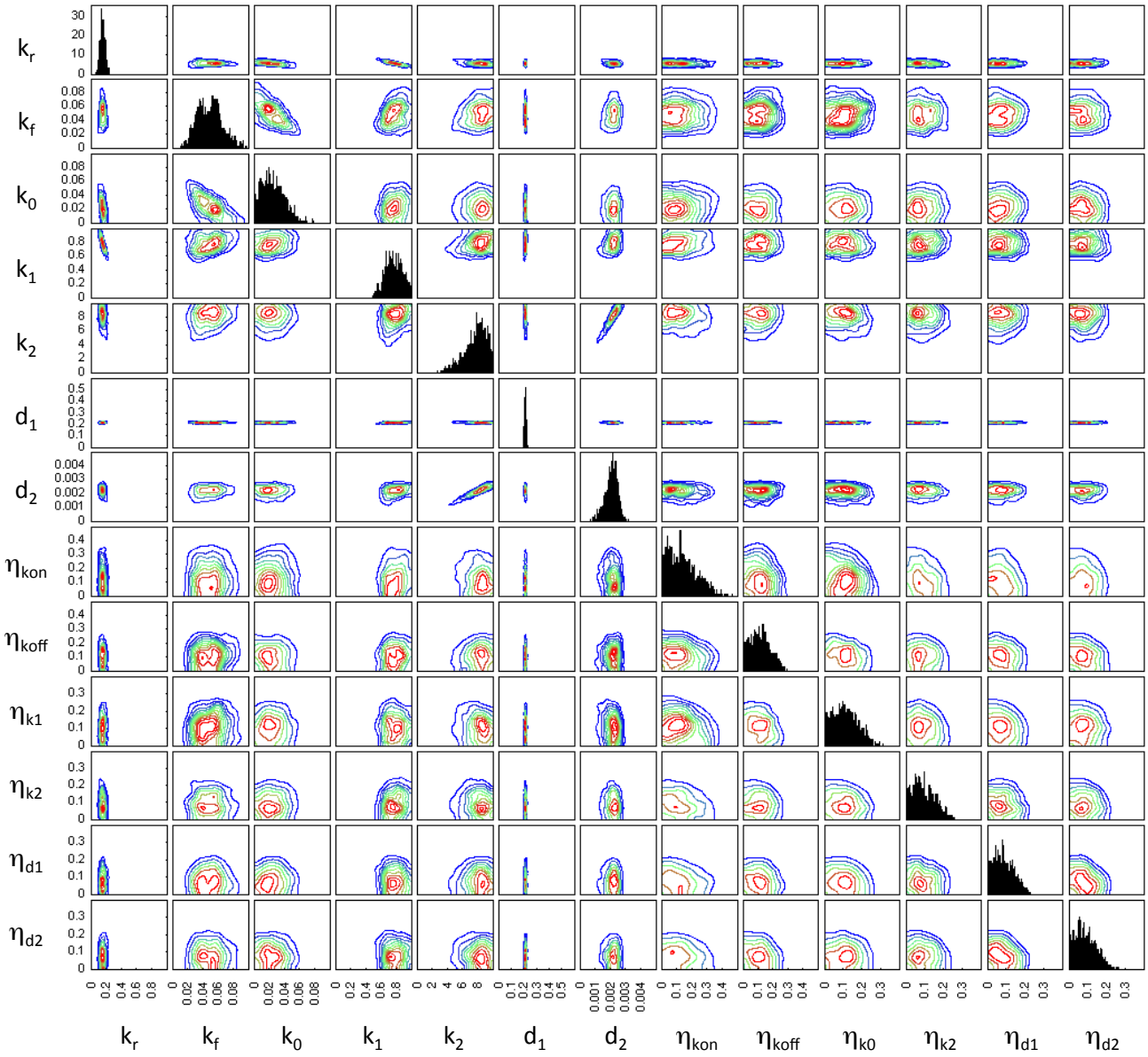

Figure 4.4: Posterior distribution of model parameters for the gene *dnaK*. Contour plots indicating the density of points with the corresponding parameter values for each particle in the final population.

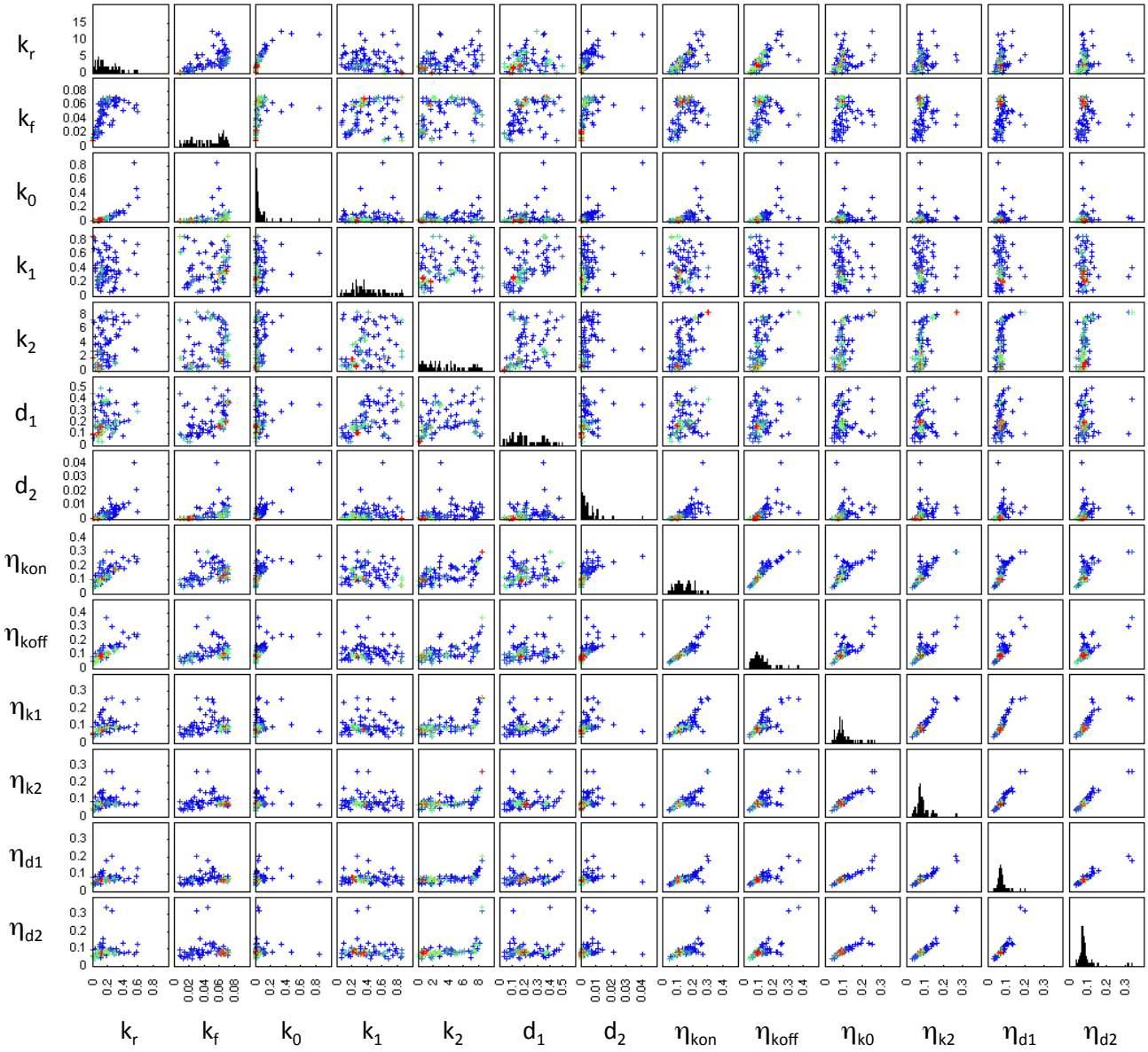

Figure 4.5: Relationships between means of the marginal parameter posteriors. Scatter plots of the means of the marginal distributions of parameter posteriors are shown for all pairs of parameters. Each point corresponds to a gene.

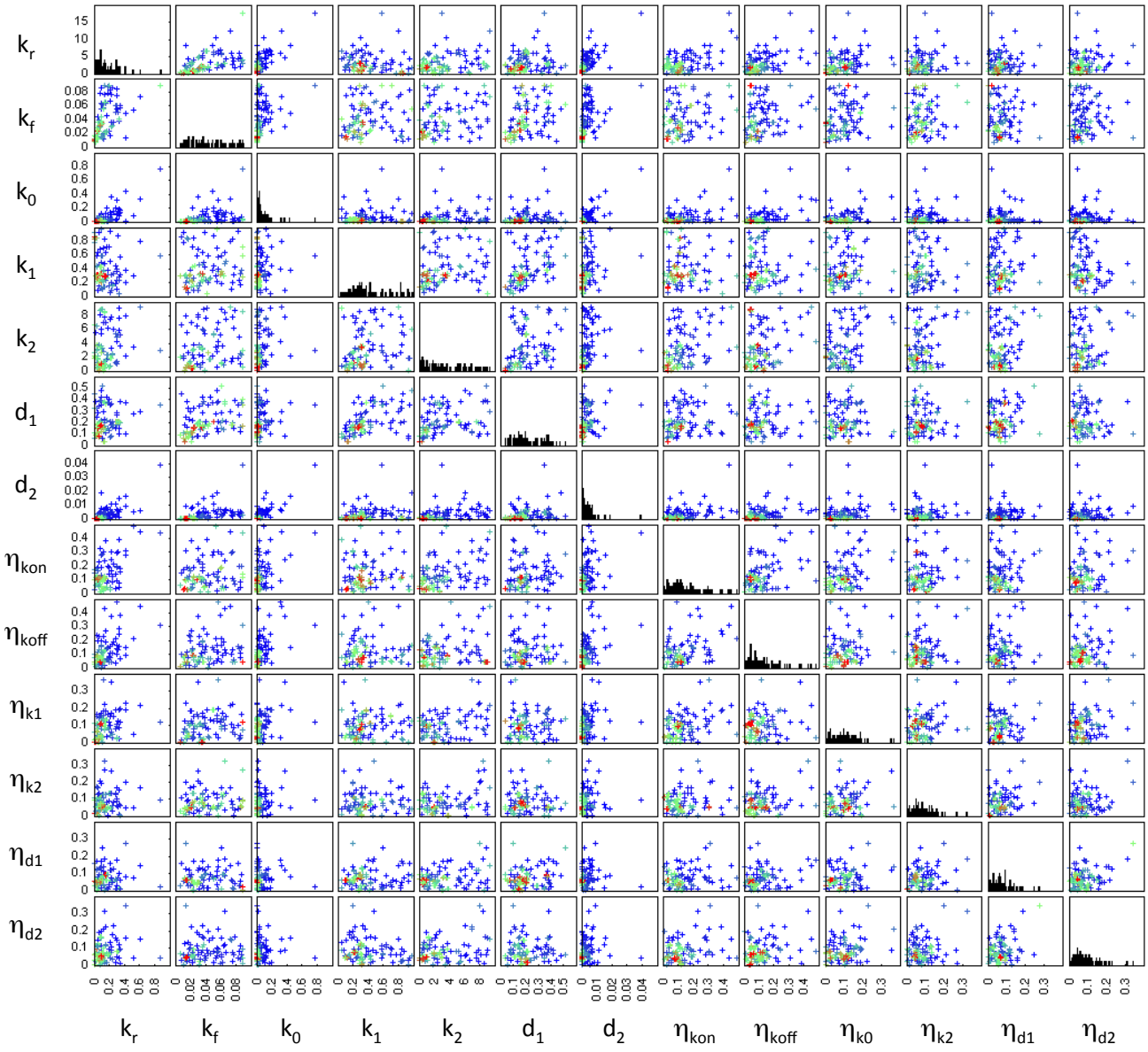

Figure 4.6: Relationships between the heaviest particles. Scatter plots of the particles with the highest weight in the final ABC-SMC population, shown for all pairs of parameters. Each point corresponds to a particle from the inferred posterior of one gene.

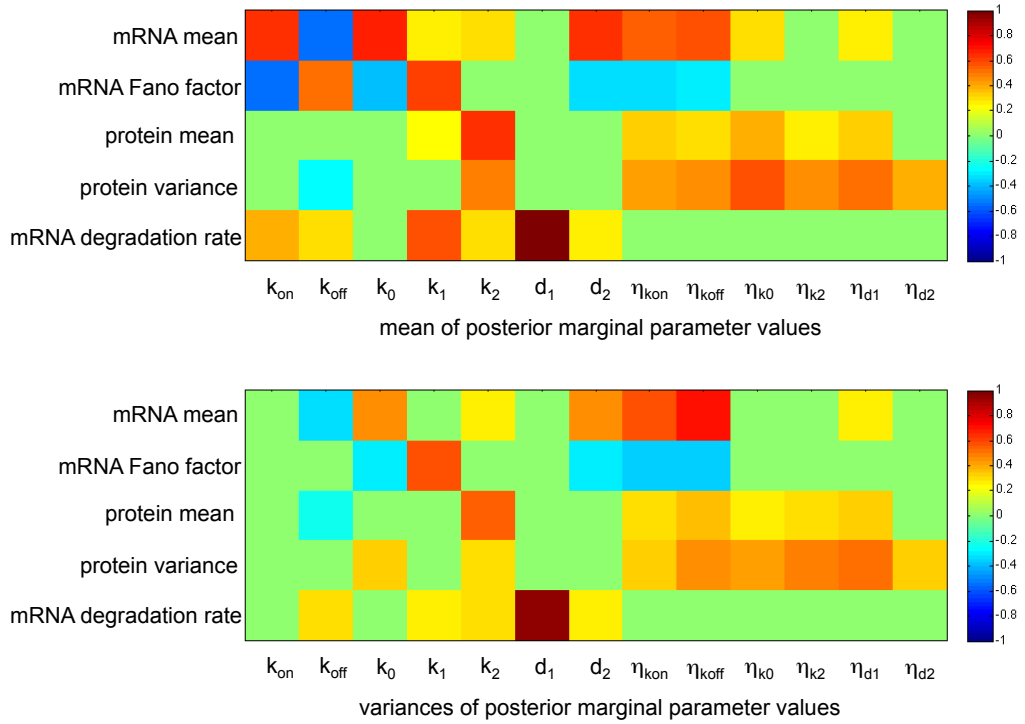

Figure 4.7: Heat maps of correlation coefficients between parameters and summary statistics. Heat maps are of the correlation coefficients calculated between experimentally obtained summary statistics and the mean (top) or the variance (bottom) of the marginal posterior for each model parameter. Correlation coefficients for which the associated p-values are greater than 0.05, after correcting for multiple testing using the Benjamini-Hochberg method [82], are treated as zero for plotting purposes.

parameters ranges. The extrinsic noise parameters for most genes are below 0.2 (Figure 4.6), however, for some genes, $\eta_{k_{\text{on}}}$ and $\eta_{k_{\text{off}}}$ have relatively broad posterior marginal distributions.

To better understand the relationship between model parameters and observed patterns of gene expression, we look for correlations between means and variances of the inferred marginal parameters of each gene and the summary statistics used in the inference procedure (Figure 4.7). As expected, the correlation between the measured mRNA degradation rate, calculated from mRNA lifetime, and the inferred mRNA degradation rate parameter of the model, is close to one.

The promoter switching rate parameters, k_{on} and k_{off} display positive and negative correlation with the mean mRNA number, respectively. They have the opposite relationship with the mRNA distribution Fano factor. This is consistent with the idea that distinct levels of transcription are required to account for the observed mRNA Fano factors. The corresponding extrinsic noise parameters $\eta_{k_{\text{on}}}$ and $\eta_{k_{\text{off}}}$ are positively correlated with mRNA abundance. However, the means and variances of the marginal distributions of these parameters are negatively correlated with the Fano factor of the mRNA distribution. This indicates that when promoter switching is affected by higher extrinsic noise, the mRNA distribution becomes more Poissonian as the effect of the two distinct promoter states is averaged out.

Curiously, the mean and variance of the protein degradation rate (d_2) are positively correlated with mean mRNA number and negatively correlated with the mRNA Fano factor. Unlike the translation rate (k_2), it shows no significant correlation with the mean or variance of the protein number.

Parameter stiffness/sloppiness

There are two complementary approaches to investigating the sensitivity of a modelled system to its parameters or inputs. One approach is to consider a single point in parameter space and study how the model responds to infinitesimal changes in parameters. This local approach usually involves calculating the partial derivatives of the model output with respect to the parameters of interest. Alternatively, one may consider how the model behaviour varies within a region of parameter space by sampling parameters and observing model behaviour. Regardless of the method used, different linear combinations of parameters will affect the model output to varying degrees [83]. Gutenkunst *et al* [84] coined the terms "stiff" and "sloppy" to describe these differences. They defined a Hessian matrix,

$$H^{\chi^2}_{i,j} \equiv \frac{d^2\chi^2}{d\log\theta_i d\log\theta_j}$$

where χ^2 provides a measure of model behaviour, such as the average squared change in the species time course. By considering the eigenvalues of this Hessian, λ_i, the authors were able to quantify the (local) responsiveness of the system to a given change in parameters. Conceptually, moving along a stiff direction in parameter space causes a large change in model behaviour, conversely moving along a sloppy direction results in comparatively little effect on the output of the system.

Secrier *et al* [85] later demonstrated how these ideas can be applied to the analysis of posterior distributions obtained by ABC methods. Principal component analysis (PCA) may be used to approximate the log posterior density using a multivariate normal (MVN) distribution. They showed that the eigenvalues of the covariance matrix, s_i, of this MVN distribution are related to the eigenvalues of the Hessian as $\lambda_i = 1/s_i$.

To assess the the stiffness/sloppiness of the inferred parameters we carry out PCA of the covariance matrices of log posterior distributions for each gene. In interpreting the results of the PCA we assume that the posterior distribution is, in practice, unimodal. The principal components (eigenvectors), v, and the corresponding loadings (eigenvalues), s, provided by the PCA are then used to obtain the eigen-parameters, q, as

$$q_i = s_i v_i.$$

We calculate the projections of each parameter, θ_i, onto each eigen-parameter, q_j, as

$$c_{i,j} = \theta_i \cdot q_j.$$

As a measure of the overall sloppiness of each parameter, l, we use the sum of the contributions of each parameter to the eigen-parameters, $l_i = \sum_j c_{i,j}$. This can also be thought of as the sum of the projections of each principal component onto the parameter, weighted by the fraction of total variance explained by each of the principal components.

Having obtained a measure of the sloppiness of each parameter, for each gene, we carry out hierarchical clustering [86] of genes and parameters using a Euclidean distance metric for both (Figure 4.8).

The majority of genes show a similar pattern of parameter stiffness/sloppiness. The most distinctive and the second most distinctive clusters consist of just two genes each, *yiiU* with *aceE* and *cspE* with *map*, respectively. These four genes are distinguished by unusually sloppy promoter activity ratio, k_r, and promoter switching frequency, k_f, parameters. The pair *yiiU* and *aceE* display a high ratio of protein variance to protein mean (Fano factor) and are stiff with regard to the protein degradation rate noise parameter η_{d_1}. *cspE* also has a high Fano factor of the protein distribution while *map* has an unusually low mRNA Fano factor. What these four genes appear to have in common is that the variability in their protein numbers is difficult to explain based solely on the mRNA variability. Thus, a higher level of extrinsic noise is inferred to account for the observed variability. Since these genes comprise a small minority, it may be that their

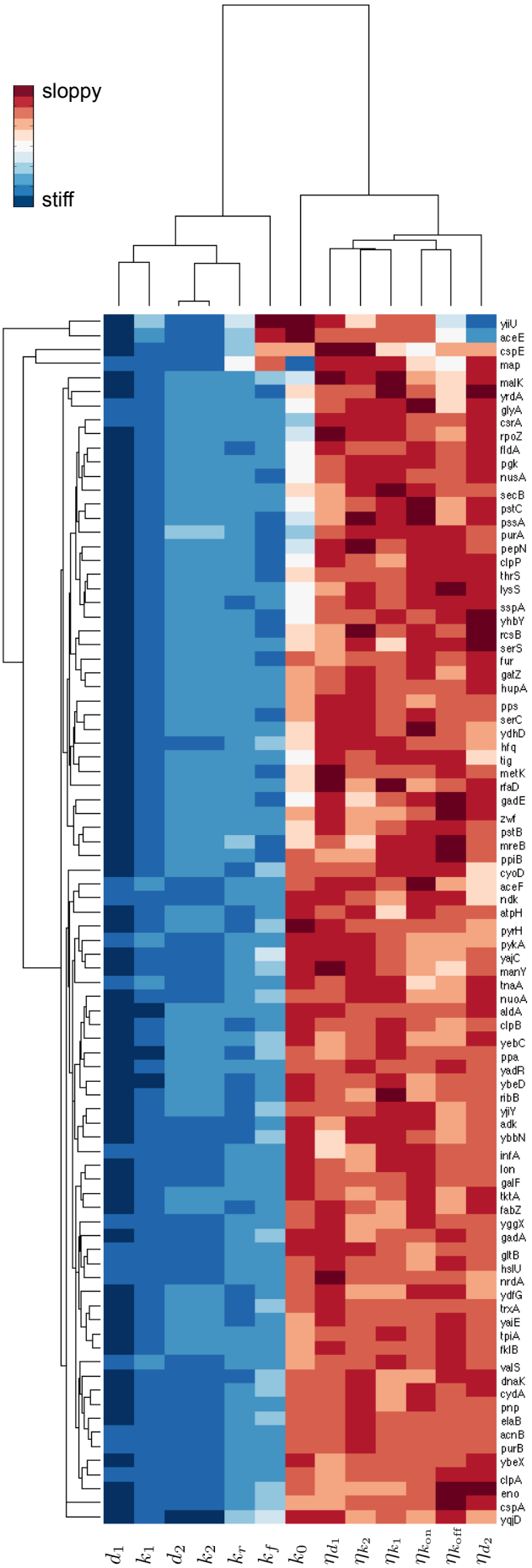

Figure 4.8: Clustering of genes and inferred posteriors according to parameter sloppiness. Clustergram showing a heat map of parameter sloppiness for each gene. Dendograms above and to the left of each heat map display the hierarchical tree obtained using a Euclidean distance metric.

expression is subject to regulatory mechanisms that are not well approximated by the two state model. The remaining majority of genes are broadly divided into two similar groups which differ mostly in the sloppiness of k_0.

The noise and rate parameters segregate into two clusters with the noise parameters generally being sloppier than the rate parameters (Figure 4.8). The least sloppy parameter is the mRNA degradation rate (d_1). This is not surprising since it was used, together with the molecule number summary statistics, to infer the posterior distribution. Of the rate parameters, the basal transcription rate (k_0) is the sloppiest and often approaches the noise parameters in its sloppiness. Since this parameter is defined as a fraction of the active transcription rate (k_1), its relative sloppiness should not be equated to a lack of importance. For most genes the marginal posterior of k_0 is largely constrained to the lower half of its prior distribution, $U(0, 1)$. The only exception being the gene *map* for which the measured mRNA Fano factor was close to one and the marginal posterior of k_0 is in the top half of the prior range. The mean of the marginal posterior of k_0 is negatively correlated with the mRNA Fano factor across all genes (Figure 4.7). The two other parameters that influence the mRNA Fano factor, k_r and k_f, are the next sloppiest rate parameters.

4.4 Conclusions

Cell-to-cell variability in genetically homogeneous populations is a ubiquitous phenomenon [87–89]. Attempts to quantify it are complicated by the difficulty of assigning it to a single cellular process or any one experimentally measurable variable. It can also be difficult to distinguish between the intrinsic stochasticity of biochemical processes in the short term and longer term variations which may have been inherited from previous cell generations.

By including a representation of extrinsic noise in our model of gene expression we attempt to infer the extent to which the rates of biochemical processes can vary between cells while still producing the experimentally measured mRNA and protein variability. We demonstrate the usefulness of an efficient method for exact stochastic simulation of the two-state model of gene expression. This model is necessary to explain the experimentally measured mRNA variation (Fano factor) and is capable of describing the majority of the observed data. We show that the amount of extrinsic noise affecting most genes appears to be limited, but non-negligible.

The exact simulation method described here occupies a niche between those cases when only samples from the steady state mRNA distribution of the two-state model [68, 90, 91] are required and cases when an approximation to the protein distribution [77, 92] is sufficient. The computational advantages of the simulation method described here are limited to specific conditions, such as, low numbers of mRNA molecules and higher numbers of protein molecules. The most limiting factor of this simulation method is that it is not applicable to models in which the protein products affect upstream processes such as promoter activity, transcription or translation. The addition of such interactions would mean that the assumptions used in deriving the Poissonian relationship between the number of surviving protein molecules produced form a given mRNA molecule and mRNA's lifetime would no longer be satisfied. Perhaps an approximate algorithm could be developed on the basis of algorithm (5) to handle such situations. Alternatively, the tau-leaping algorithm [93] may be more appropriate for models involving these kinds of feedback interactions. Algorithm (5) could, however, be naturally extended to models involving regulatory interactions between non-coding RNAs as the simulation of that part of the model is equivalent to Gillespie's exact algorithm. Although here we use summary statistics of mRNA and protein number measurements, the simulation method is also applicable to cases where a direct comparison between sample distributions, for example using the Hellinger distance, is required.

The inferred extrinsic noise parameters will also include the effects of regulatory mechanisms that are not well described by the two-state model. In this sense, our definition of noise becomes blurred with our ignorance about the regulatory interactions involved in the expression of each gene. Nonetheless, the biochemical mechanisms governing gene expression in a given species are shared between many genes. This is in agreement with our observation that, for most genes, inferred model parameters show similar patterns of sloppyness. If we are able to refine our understanding of the shared aspects of gene expression, we may be able to improve our understanding of both the nature of the noise affecting it, and the regulatory mechanisms controlling it. In practice this may mean finding a mechanistic explanation for the two-state model or further refining it to achieve a better agreement between simulations and experimental results.

The *in silico* approach used here not only relied on, but was inspired by the experimental work of Tanaguchi *et al* [17]. As the resolution of high throughput experimental techniques and the quantity of data they generate continues to increase, more complete observations of cellular processes may begin to yield data amenable to statistical analysis

and inference of extrinsic noise. These may in turn require other modelling, computational and theoretical approaches which would not rely on the assumptions and simplifications that we make in this work [94]. Models involving interactions between multiple genes, mediated by the protein products of those genes, require more general simulation methods. If sufficient computational capacity is available, they could be simulated using Gillespie's direct method [13]. Alternatively, approximate methods using, for example, the unscented transform [95] or moment expansion [11], could be employed.

Chapter 5

Signal transduction by two-component systems

5.1 Two-component signalling systems

A common mechanism by which cells sense and respond to their environment revolves around membrane bound receptors coupled to downstream biochemical processes. Bacterial two-component systems (TCS) are an example of such a sensing system. A TCS is comprised of a transmembrane histidine kinase (HK) receptor and a response regulator (RR) protein. TCSs are ubiquitous in bacteria. As many as 50000 proteins belonging to TCSs have been identified on the basis of sequence analysis [96]. They are responsible for regulating a plethora of cellular functions including metabolism, virulence, stress responses and sporulation. The *E. coli* K-12 genome encodes 30 different TCSs responsible for sensing a wide range of environmental variables, from osmolarity (EnvZ/OmpR), to phosphate levels (PhoR/PhoB). The well studied chemotaxis proteins also comprise an atypical variant of a TCS [97]. Similarities in TCSs between bacterial species can be attributed to both horizontal gene transfer and co-evolution [98]. Homologues of TCSs have also been identified in *Saccharomyces cerevisiae* [99] and *Arabidopsis thaliana* [100], however, eukaryotic TCSs are comparatively rare.

The basic steps involved in TCS signal transduction are as follows. The receptor HK consumes ATP to auto-phosphorylate a conserved histidine residue. The HK can interact with a cognate RR protein. If the RR is unphosphorylated at a specific, conserved, aspartate residue, the phosphate group may be transferred from the HK to the RR. In the

event that the RR is phosphorylated, the HK can act as a phosphatase, dephosphorylating the RR. Phosphorylation controls RR activity. The active RR may behave as an enzyme, as is the case with the chemotaxis protein CheB. However, it is more common for the RR to act as a transcription factor.

TCSs can be divided into two distinct categories, orthodox and non-orthodox (Figure 5.1), based on the domains present in the HK [96]. The typical, orthodox HK consists of a sensing domain linked to a conserved kinase core. The RR is comprised of a receiver (REC) domain linked to an effector domain, such as a DNA binding domain. Non-orthodox HKs possess two additional domains, a REC domain and a histidine phosphotransfer (HPt) domain. Rather than directly transferring the phosphate from the kinase core to the REC domain of the RR, as is the case with orthodox systems, non-orthodox HKs first pass the phosphate to an aspartate residue of their own REC domain and then to a histidine on the HPt domain. Such modular structure may be what allowed TCSs to diversify their functions during the course of evolution. In the future, TCSs may provide a toolbox of interchangeable components for synthetic biologists [101–103].

HKs and RRs are usually thought of as acting in cognate pairs to perform a specific function. In fact, 23 of the *E. coli* K-12 cognate pairs are co-localised on the chromosome [98]. Nonetheless, so called, cross-talk between different cognate pairs exists and may play a significant role in cellular signalling. In addition, interactions between TCSs can be mediated by accessory proteins and allosteric regulators [97].

While signal transduction via phosphotransfer from the HK to the RR is a core feature of TCSs, the details of the signalling mechanism may differ between systems. The HK can respond to a stimulus either by altering its kinase activity (e.g. CpxA, LuxN), its phosphatase activity (e.g. KdpD) or both (e.g. NtrB, PhoQ) [96]. In turn, the active RR can function as a monomer (e.g. Spo0F, CheY, CheB), a dimer (e.g. FixJ, OmpR, KdpE, PhoB, Spo0A) or a even as an oligomer (NtrC, UhpA) [104]. It may also exist as a mixture of dimers and monomers irrespective of its phosphorylation state, as is the case with PhoP [105].

5.2 Modelling TCS phosphotransfer

Bacterial TCSs are relatively well characterised compared to many other biomolecular signalling pathways. Despite this, the task of building a model of TCS signalling

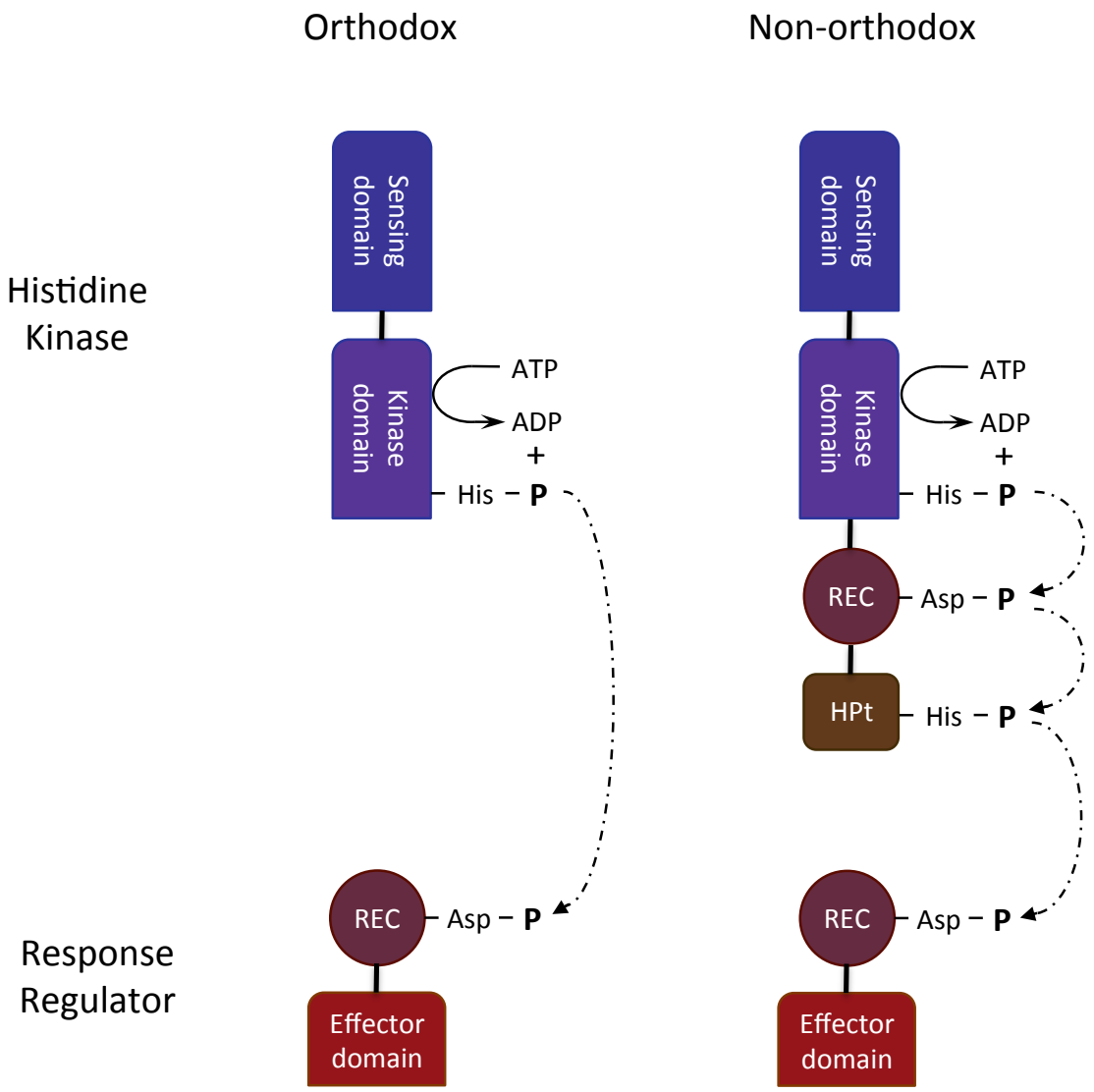

Figure 5.1: A schematic illustration of orthodox and non-orthodox TCS signalling. The input signal is detected by the signalling domain on the HK. Information about the signal is transmitted by changes in the phosphorylation state of conserved histidine (His) or aspartate (Asp) residues on the HK and RR. The HK's kinase domain consumes ATP to phosphorylate a histidine residue. In orthodox TCSs this phosphate can be transferred directly to the REC domain of the RR. Non-orthodox HKs contain two additional domains, REC and HPt. The phosphate is transferred to these domains before reaching the RR. Figure based on reference [96].

is fraught with uncertainties. As discussed above, the details of the signalling process vary between particular TCSs. While there are many studies which provide experimental measurements, these studies focus on a range of different strains, TCSs, molecular species, or processes. Even experiments involving the same TCS from the same bacterial strain may have been carried out under different conditions, *in vitro* or *in vivo*. For example, some experiments rely on measurements of a reporter gene's expression level in live bacterial cells [106] while methods such as surface plasmon resonance can provide information about binding affinities but require purified sample of the relevant macromolecules [105].

Previous modelling studies include the use of deterministic models to compare orthodox and non-orthodox TCSs [107] as well as investigations into ultrasensitivity and bistability exhibited by atypical histidine kinases [108]. Stochastic models have provided insight into the effects of varying the kinetic parameters of the system on the qualitative nature of its response to the signal [109].

The aims of the work presented in this chapter are two-fold. Firstly, to develop a plausible model of TCS signalling based on published data about rate parameters and to connect it to the gene expression model described in the previous chapter. Secondly, to use this model to provide an illustrative example of how approaches rooted in information theory may be applied in the modelling of cellular signalling.

We begin by considering three closely related models describing interactions between the HK and the RR. We compare the behaviours of the deterministic versions of these models with the numbers of protein molecules remaining fixed. Although many of the parameters used in these models are derived from experimental measurements of the PhoR/PhoB TCS, these model are not necessarily a realistic representation of that particular TCS. Subsequently we focus on one of these models in a less idealised setting by carrying out stochastic simulation of the TCS together with its operon and a downstream gene.

Model 1

Model 1 consists of 14 reactions (appendix B.1) and forms the core of the more complex models, 2 (appendix B.2) and 3 (appendix B.3). The dimerisation and autophosphoryla-

tion of HK is described by the following reactions.

$$\begin{aligned} 2\mathrm{HK} &\xrightarrow{c_1} \mathrm{HK_2} & c_1 &= 0.0125s^{-1} \\ \mathrm{HK_2} &\xrightarrow{c_{-1}} 2\mathrm{HK} & c_{-1} &= 0.005s^{-1} \\ \mathrm{HK_2} &\xrightarrow{c_k} \mathrm{HK_2p} & c_k &= 0.025s^{-1} \\ \mathrm{HK_2p} &\xrightarrow{c_{-k}} \mathrm{HK_2} & c_{-k} &= 0.0077s^{-1} \end{aligned}$$

Experimental studies of the EnvZ HK *in vitro* have indicated that at equilibrium, the dimeric form (HK_2) of the receptor is favoured [110]. The precise values of c_1 and c_{-1} were chosen arbitrarily but remain consistent with this observation. The parameter c_k represents the autophosphorylation rate of the HK dimer. This is the parameter which is altered in response to a signal affecting the receptor kinase activity. In the absence of any external signal we set $c_k = 0$. The value of c_k when a signal is present can also be interpreted as the strength of the external signal. The magnitude of the chosen values of c_k and c_{-k} is consistent with *in vivo* work on the PhoR/PhoB system [111] and also, crudely, with the timescales of PhoR phosphorylation *in vitro* [112].

A more pronounced response to the signal can be obtained by increasing c_k. While HK autophosphorylation carries an energetic cost for the cell, alternative ways of increasing responsiveness to the signal, such as synthesising more HK proteins may also be costly.

We chose to focus on TCSs in which the active form of the RR is the dimer (RRp_2). The following reactions describe background phosphorylation of the RR and dimerisation of the phosphorylated form.

$$\begin{aligned} \mathrm{RR} &\xrightarrow{c_a} \mathrm{RRp} & c_a &= 8.67e-05s^{-1} \\ \mathrm{RRp} &\xrightarrow{c_g} \mathrm{RR} & c_g &= 0.00026s^{-1} \\ 2\mathrm{RRp} &\xrightarrow{c_2} \mathrm{RRp_2} & c_2 &= 0.00166s^{-1} \\ \mathrm{RRp_2} &\xrightarrow{c_{-2}} 2\mathrm{RRp} & c_{-2} &= 0.02158s^{-1} \end{aligned}$$

In the absence of phosphatase activity from its cognate HK (PhoR), PhoB becomes phosphorylated *in vivo*. The autodephosphorylation rate of the RR (c_g) corresponds to the experimental measurement for PhoB [111]. The background phosphorylation rate c_a was then chosen so that $1/3$ of the RR is phosphorylated at equilibrium in the absence

of HK. The parameter c_2 was chosen arbitrarily but falls within the range of plausible protein association rates [113], c_{-2} was them chosen to give a $K_d = 13\ \mu$M as measured for Spo0A [104].

TCS signalling requires the temporary association of the HK and RR proteins. Depending on whether each protein is phosphorylated or not, there are four possible combinations of pairs. Model 1 only includes the reactions necessary for phosphate transfer and the dephosphorylation of phosphorylated RR by the unphosphorylated HK dimer.

$$\begin{aligned}
\mathrm{HK_2{:}RR} &\xrightarrow{c_{-3a}} \mathrm{HK_2 + RR} & c_{-3a} &= 0.00498 s^{-1} \\
\mathrm{HK_2p + RR} &\xrightarrow{c_{3b}} \mathrm{HK_2p{:}RR} & c_{3b} &= 0.00166 s^{-1} \\
\mathrm{HK_2 + RRp} &\xrightarrow{c_{3c}} \mathrm{HK_2{:}RRp} & c_{3c} &= 0.00166 s^{-1} \\
\mathrm{HK_2{:}RRp} &\xrightarrow{c_{-3c}} \mathrm{HK_2 + RRp} & c_{-3c} &= 0.00498 s^{-1} \\
\mathrm{HK_2p{:}RR} &\xrightarrow{c_t} \mathrm{HK_2{:}RRp} & c_t &= 0.0279 s^{-1} \\
\mathrm{HK_2{:}RRp} &\xrightarrow{c_p} \mathrm{HK_2{:}RR} & c_p &= 0.0087 s^{-1}
\end{aligned}$$

Protein association rates were chosen as before. The dissociation rates c_{-3a} and c_{-3c} are picked assuming that all HK:RR complexes have a $K_d = 3\ \mu$M [111] regardless of phosphorylation state. Phosphate transfer (c_t) and phosphatase (c_p) rates are in accordance with *in vivo* measurements for PhoR/PhoB [111]. As with a signal affecting the HK kinase activity, phosphatase activity is treated as being on when the signal is present and zero when signal is absent.

Experimental measurements indicate that RR is more abundant than HK by an order of magnitude and that this ratio is maintained over a range of expression levels [111]. We simulate the system with 100 and 1000 molecules of HK and RR, respectively. Figure 5.2 shows the trajectory of $\mathrm{RRp_2}$ in model 1 in response to signals affecting kinase activity, phosphatase activity or both. Model 1 shows a greater response when the signal affects phosphatase activity (c_p) than when it affects kinase activity (c_k) (Figure 5.2). The parameter c_p directly influences the rate of removal of RRp from the system. In contrast, c_k affects the production of $\mathrm{HK_2p}$ so its influence on RR_p abundance is attenuated by the dephosphorylation rate of $\mathrm{HK_2p}$ (c_{-k}) as well as the rate of phosphate transfer from HK to RR (c_t).

To better understand the significance of various model parameters we carried out sensi-

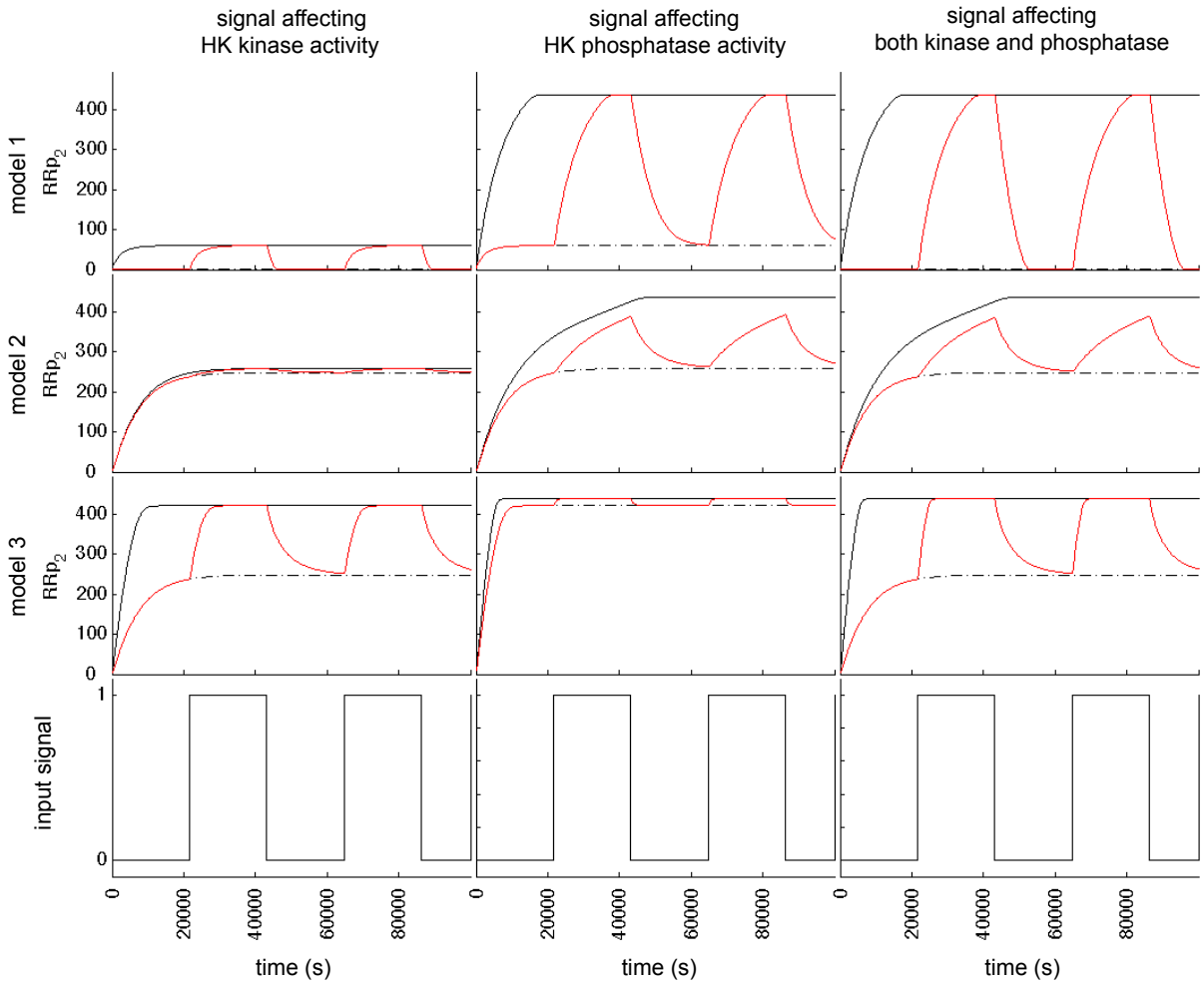

Figure 5.2: Trajectories of RRp_2 for each of the three models with three different mechanisms for responding to the input signal. Each row corresponds to one of the three models. Each column shows the response to a signal type. Trajectories are plotted for a constitutively active signal (solid black lines), no signal (dashed black lines) and an alternating signal (red lines). First column, HK kinase activity represented by the parameters c_k and c_l (model 3 only) is zero when the signal is zero (off) and $0.025s^{-1}$ when the signal is one (on). Second column, HK phosphatase activity is $0.087s^{-1}$ when the signal is zero and zero when the signal is one. Third column, both HK kinase and HK phosphatase activities are affected by the signal.

tivity analysis using the StochSens package [114]. This package is an implementation of an alternative approach to the typical local sensitivity analysis. Rather than considering the derivative of the model's output with respect to the model parameters (θ), StochSens focuses on the changes in the probability distribution of the model's output (ψ). This approach is more appropriate when the system in question is stochastic since its output is not defined by a single observable value but by a probability distribution. StochSens places this approach in a determinist setting by using the linear noise approximation to approximate ψ and quantifies how it changes using the Fisher Information Matrix, defined as,

$$I_F(\theta) = \mathbb{E}\left(\frac{\delta \log \psi(X, \theta)}{\delta \theta}\right)^2$$

where X is the model's output.

We calculate the local sensitivity of the model at steady state, using the parameter values as listed above. Visual inspection of the model output trajectories indicates that 90000 seconds of simulated time is more than sufficient for the steady state to be reached. Since the amount of active RR is considered to be the output of the TCS in response to a signal, we are particularly interested in the sensitivity of RRp_2 with respect to model parameters at steady state. In model 1, RRp_2 is most sensitive to the background phosphorylation rate of RR (c_a) and the rate of association of RRp with HK_2 (c_{3c}) (Figure 5.3). The next, most influential parameters are those governing HK autophosphorylation (c_k) and HK phosphatase activity (c_p). These are the parameters upon which the input signal acts.

Model 2

Model 2 builds on model 1 by including the unproductive binding of unphosphorylated HK to uphosphorylated RR.

$$HK_2 + RR \xrightarrow{c_{3a}} HK_2{:}RR \qquad\qquad c_{3a} = 0.00166 s^{-1}$$

The association rate c_{3a} was chosen to be the same as for other reactions which produce a complex containing the HK and RR proteins. It is entirely plausible that the affinity of HK for phosphorylated and unphosphorylated RR differ. Models 1 and 2 can be considered to be two extreme cases of this. The formation of the unphosphorylated complex leads to a decrease in the abundance of free HK_2, effectively inhibiting both the kinase and phosphatase activities of HK. The abundance of RRp_2 increases (Figure

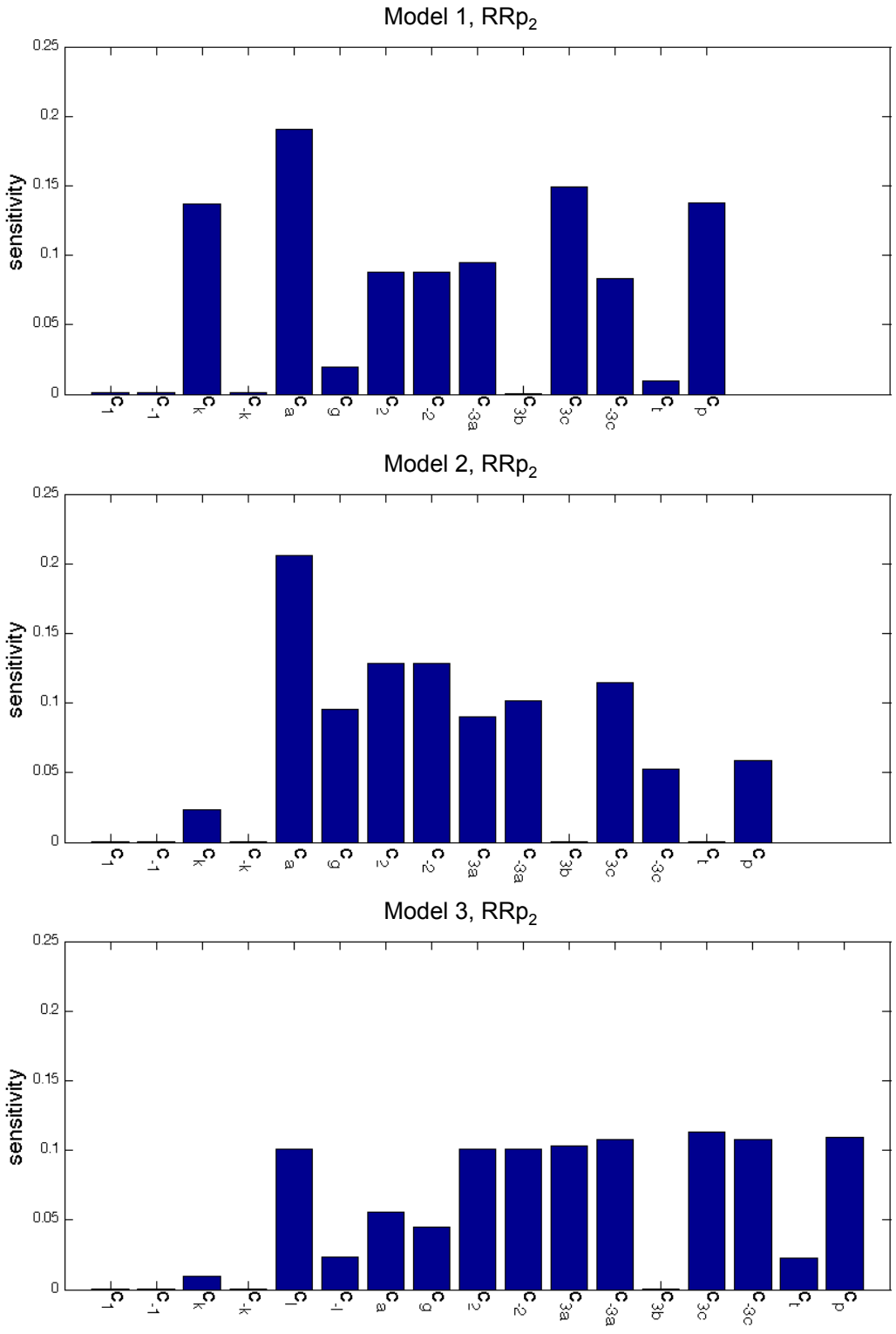

Figure 5.3: Sensitivity analysis of RRp_2 with respect to model parameters for each of the three models. Sensitivity analysis was carried out using the StochSens package. Models were simulated until a steady state was reached before sensitivity values were calculated.

5.2) but the system becomes much less sensitive to signals affecting c_k (Figure 5.3) since almost all the HK is bound to either RR or RRp. The equilibrium state of model 2 is even more sensitive to the basal phosphorylation (c_a) and dephosphorylation (c_g) rates of RR and RRp than that of model 1.

Model 3

In addition to the reactions making up model 2, model 3 includes two more reactions. These reactions allow for the phosphorylation and dephosphorylation of HK in complex with RR.

$$\mathrm{HK_2{:}RR} \xrightarrow{c_l} \mathrm{HK_2p{:}RR} \qquad c_l = 0.025s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_{-l}} \mathrm{HK_2{:}RR} \qquad c_{-l} = 0.0077s^{-1}$$

In this model, signals affecting HK kinase activity alter both the c_k and c_l parameters so that the signal is sensed by HK dimers bound to unphosphorylated RR. Since most HK in the system at any one time is in complex with RR, this feature of the model has a significant effect on the behaviour of the system. Model 3 is much more responsive to a signal affecting HK kinase activity. Sensitivity analysis confirms that this due to the autophosphorylation of HK in complex with RR, with c_l being a more influential parameter than c_k (Figure 5.3). Conversely, the response of model 3 to changes in HK phosphatase activity is considerably lower than that of the other two models. (Figure 5.2). Removing the need for HK:RR dissociation puts the phosphatase activity of HK in more direct competition with its kinase activity. Curiously, the rate of phosphate transfer from HK to RR (c_t) from does not appear to be a very influential parameter in any of the three models at steady state. However, model 3 is the one most sensitive to it.

5.3 Modelling the TCS operon

For all three models the amount of $\mathrm{RRp_2}$ depends on the initial amounts of HK and RR in the system. This raises two question. Firstly, how can a TCS robustly transmit a signal given that HK and RR protein numbers fluctuate due to both intrinsic and extrinsic noise? Secondly, given that the active RR regulates gene expression, does the inherent noisiness of the process of regulating gene expression swamp any information about the input signal?

To try to answer these questions we constructed a model (model 4, appendix B.4) which includes the production of HK and RR from an operon that is also subject to regulation by RRp_2. In addition, we model a gene which is regulated by RRp_2 and produces a protein (X) that is not involved in TCS signalling. As well as simulating the stochastic model using the Gillespie algorithm, we perturbed the reaction rate parameters of the two-state model governing gene expression before each simulation run, to represent extrinsic noise as discussed in Chapter 4.

This model includes all the reactions from model 3 as well as the following reactions

describing promoter activity, transcription and translation.

$$\text{promoter-off} \xrightarrow{k_{\text{on}}} \text{promoter-on} \qquad k_{\text{on}} = 0.0431s^{-1}, \eta_{k_{\text{on}}} = 0.0672$$

$$\text{promoter-on} \xrightarrow{k_{\text{off}}} \text{promoter-off} \qquad k_{\text{off}} = 0.4479s^{-1}, \eta_{k_{\text{off}}} = 0.0428$$

$$\text{RRp}_2 + \text{promoter-off} \xrightarrow{c_{\text{on}}} \text{RRp}_2\text{-promoter-on} \qquad c_{\text{on}} = 0.000332s^{-1}$$

$$\text{RRp}_2 + \text{promoter-on} \xrightarrow{c_{\text{on}}} \text{RRp}_2\text{-promoter-on} \qquad c_{\text{on}} = 0.000332s^{-1}$$

$$\text{RRp}_2\text{-promoter-on} \xrightarrow{c_{\text{off}}} \text{RRp}_2 + \text{promoter-off} \qquad c_{\text{off}} = 0.006s^{-1}$$

$$\text{RRp}_2\text{-promoter-on} \xrightarrow{k_1} \text{mRNA} + \text{RRp}_2\text{-promoter-on} \qquad k_1 = 0.434s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{promoter-on} \xrightarrow{k_1} \text{promoter-on} + \text{mRNA} \qquad k_1 = 0.434s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{promoter-off} \xrightarrow{k_0 k_1} \text{promoter-off} + \text{mRNA} \qquad k_0 k_1 = 0.025649s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{mRNA} \xrightarrow{d_1} \emptyset \qquad d_1 = 0.396s^{-1}, \eta_{d_1} = 0.0441$$

$$\text{mRNA} \xrightarrow{k_{2\text{HK}}} \text{HK} + \text{mRNA} \qquad k_{2\text{HK}} = 0.87416s^{-1}, \eta_{k_2} = 0.0369$$

$$\text{mRNA} \xrightarrow{k_{2\text{RR}}} \text{RR} + \text{mRNA} \qquad k_{2\text{RR}} = 8.7416s^{-1}, \eta_{k_2} = 0.0369$$

$$\text{HK} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2 \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{p} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RRp} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RRp}_2 \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{:RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{p:RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{:RRp} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

Parameters describing the binding of RRp_2 to the promoter (c_{on} and c_{off}) are based on experimental measurements for the KdeE RR [115]. In this model we assume that only the active form of the RR binds to the promoter region and that transcriptional activity is in the more active state as long as it is bound. Dissociation of the RR puts the promoter

in the inactive state but does not preclude it from stochastically switching to the active state.

The parameters regulating promoter switching (k_{on} and k_{off}), basal and active transcription rates (k_0 and k_1), the translation rate of RR ($k_{2\text{RR}}$) and the noise parameters were obtained by taking the particle with the highest weight from the inferred posterior distribution of parameters for gene *rcsB*, which encodes the RR for the RcsC/RcsB TCS [106] (see Chapter 4). The translation rate of HK ($k_{2\text{HK}}$) was set to be one tenth of that of RR to account for the ratio of their abundances. We use equivalent reactions with the same parameters to describe the promoter activity, transcription and translation of the gene encoding protein X.

$$\text{promoter(X)-off} \xrightarrow{k_{\text{onX}}} \text{promoter(X)-on}$$

$$k_{\text{onX}} = 0.0431 s^{-1}, \eta_{k_{\text{on}}} = 0.0672$$

$$\text{promoter(X)-on} \xrightarrow{k_{\text{offX}}} \text{promoter(X)-off}$$

$$k_{\text{offX}} = 0.4479 s^{-1}, \eta_{k_{\text{off}}} = 0.0428$$

$$\text{RRp}_2 + \text{promoter(X)-off} \xrightarrow{c_{\text{onX}}} \text{RRp}_2\text{promoter(X)-on} \qquad c_{\text{onX}} = 0.000332 s^{-1}$$

$$\text{RRp}_2 + \text{promoter(X)-on} \xrightarrow{c_{\text{offX}}} \text{RRp}_2\text{promoter(X)-on} \qquad c_{\text{offX}} = 0.000332 s^{-1}$$

$$\text{RRp}_2\text{promoter(X)-on} \xrightarrow{c_{\text{offX}}} \text{RRp}_2 + \text{promoter(X)-off} \qquad c_{\text{offX}} = 0.006 s^{-1}$$

$$\mathrm{RRp_2promoter(X)\text{-}on} \xrightarrow{k_{1X}} \mathrm{mRNA(X)} + \mathrm{RRp_2promoter(X)\text{-}on}$$
$$k_{1X} = 0.434 s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{promoter(X)\text{-}on} \xrightarrow{k_{1X}} \mathrm{promoter(X)\text{-}on} + \mathrm{mRNA(X)}$$
$$k_{1X} = 0.434 s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{promoter(X)\text{-}off} \xrightarrow{k_{0X} k_{1X}} \mathrm{promoter(X)\text{-}off} + \mathrm{mRNA(X)}$$
$$k_{0X} k_{1X} = 0.025649 s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{mRNA(X)} \xrightarrow{d_{1X}} \emptyset$$
$$d_{1X} = 0.396 s^{-1}, \eta_{d_1} = 0.0441$$
$$\mathrm{mRNA(X)} \xrightarrow{k_{2X}} \mathrm{mRNA(X)} + \mathrm{protein(X)}$$
$$k_{2X} = 8.7416 s^{-1}$$
$$\mathrm{protein(X)} \xrightarrow{d_{2X}} \emptyset$$
$$d_{2X} = 0.0035 s^{-1}, \eta_{d_2} = 0.1664$$

From the perspective of cellular signalling, we are most interested in the behaviours of $\mathrm{RRp_2}$ and protein X. The former represents the TCSs response to the input signal while the latter servers and example of how this response may be propagated to downstream processes. Figure 5.4 shows an example of the trajectories of these two species.

5.4 Quantifying information flow

We can estimate the mutual information (MI) between the state of the input signal and various outputs of the model to quantify how well information about the signal is being transmitted. As discussed previously (Chapters 2 and 3), the MI will depend on properties of the signal as well as the model. For the input signal we use a telegraph process with on and off switching rates of 10^{-5} s^{-1}. We begin by estimating the MI between the state of the signal and the numbers of each species present in the system at the final simulation time point (90000 seconds).

While the number of molecules of a particular species present at a given time point provides an obvious measure of the behaviour of the system, it will not necessarily be the feature that carries the most information about the signal. In an idealised system, the

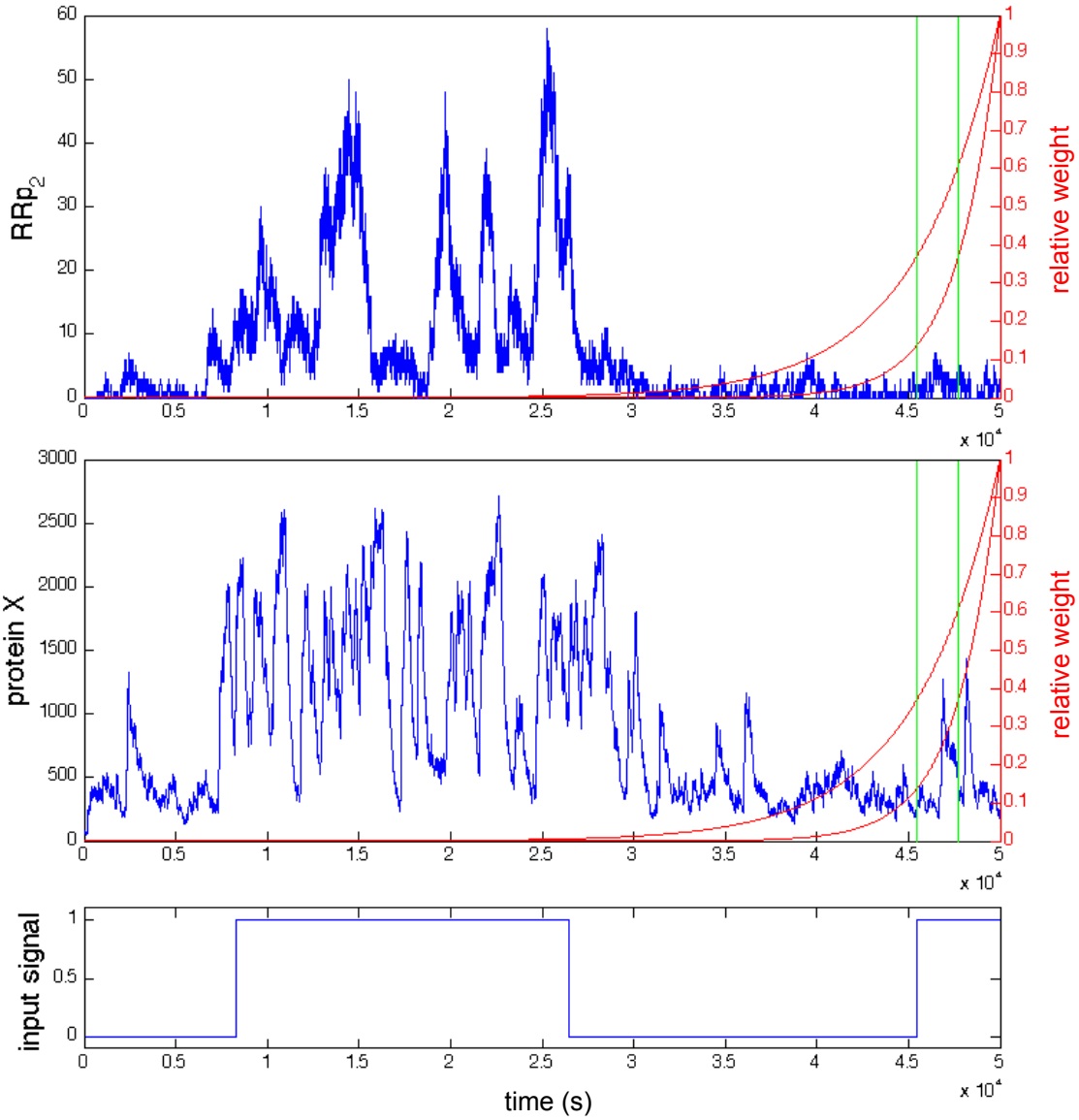

Figure 5.4: Example trajectories of RRp_2 and protein X in response to an input signal affecting HK kinase activity. The red and green lines illustrate how the weights used to calculate average molecule numbers depend on time relative to the final simulation time point.

complete trajectory of a particular species or set of species could be tracked to extract the maximum amount of information about the signal. However, the cell itself contains no mechanism for recording or "reading" such temporal trajectories. Similarly, there is no biological mechanism for counting the numbers of molecules present in the cell at a given time. Instead, cellular processes which respond to the signal do so by cumulatively aggregating the effects of indirect influences of the signal over a period of time. In reality, such processes are stochastic and the details are specific to each signalling mechanism or system.

Here we look at two toy examples of how information may be extracted from the temporal trajectory without resorting to analysing the whole trajectory or restricting information gathering to a single time point. The number of molecules of each species does not change during the time period between reactions (τ). Thus, we can take a uniformly weighted average (A_u) of the number of molecules of a given species over a period of time (T seconds).

$$A_u = \frac{1}{T} \sum_{i=1}^{N} \tau_i S_i$$

where i is an index over all (N) the time segments contained within T and S_i is the number of molecules of the species of interest during the i-th time period. This approach is a natural extension of considering just the number of species at a single time point. However, the temporal dynamics of cellular signalling do not conform to such abstractions. The influence of past molecule numbers on subsequent events will, in reality, not be uniform over an arbitrary time period. Complex dynamics such as multistability or ultrasensitivity may dictate how past information is integrated.

Our second example is motivated by a relatively simple scenario in which the "memory" about the molecule number decays exponentially over time. A biological example of this could be a system in which S is an enzyme producing a metabolite that decays via a first order reaction. Our exponential average A_e then corresponds to the expected abundance of the metabolite, given a temporal trajectory of the number of molecules of S.

$$A_e = \frac{1}{C} \sum_{j=1}^{M} w_j S_j$$

where j is an index over all time segments, C is a normalising constant, S_j is the molecule count from the j-th time period and w_j is the weight given to S_j.

$$C = \int_0^{t'_0} e^{-\lambda t'} dt'$$

where λ is the rate parameter of exponential decay, t' is the time until the final time point and t'_0 is the earliest time point from which A_e is calculated.

$$w_i = \frac{1}{\lambda}\left(e^{-\lambda t'_j} - e^{-\lambda t'_{j-1}}\right)$$

where t'_j and t'_{j-1} correspond to the end and beginning respectively, of the j-th time segment. This type of weighting is illustrated in Figure 5.4.

We obtain estimates of MI between the state of the signal and molecule numbers at the final time point. For species which do not exceed 255 molecules in any of the final time points, we use the plug-in estimator of MI. For more abundant species we use the one-dimensional kernel density estimator described in Chapter 2. To estimate the MI between the signal state and the uniform average A_u or the exponential average A_e we also use the one dimensional kernel density estimator. The MI estimates are based on 10000 simulation runs. We calculate A_u and A_e for a range of time periods (T) and exponential decay parameters (λ) respectively. Figure 5.5a shows how the MI estimates depend on the time period over which the average is taken for RRp_2 and for protein X. We note that the values of T or λ which yield the highest MI are different for each species. For some species, such as RRp_2, MI is highest for relatively small values of T or $1/\lambda$. In such cases there is little benefit in averaging the species abundance over a period of time and the MI between the signal and the number of molecules at that time point is almost as high as the best average. For other species, for example protein X, using the average species abundance over a period of time produces a noticeable increase in MI. Figure 5.5b shows a comparison of the highest MI values obtained using each averaging method.

5.5 Discussion

The TCS signalling mechanism

The three models of TCS signalling demonstrate how relatively minor mechanistic differences can have a significant impact on the way the system responds to a particular kind of input signal. While it is not possible to say which of the mechanisms, if any, best resemble actual TCS systems based purely on computational modelling, they do provide testable hypotheses for potential experimental studies. Whether the kinetics of HK autophosphorylation are significantly affected by complex formation with the RR

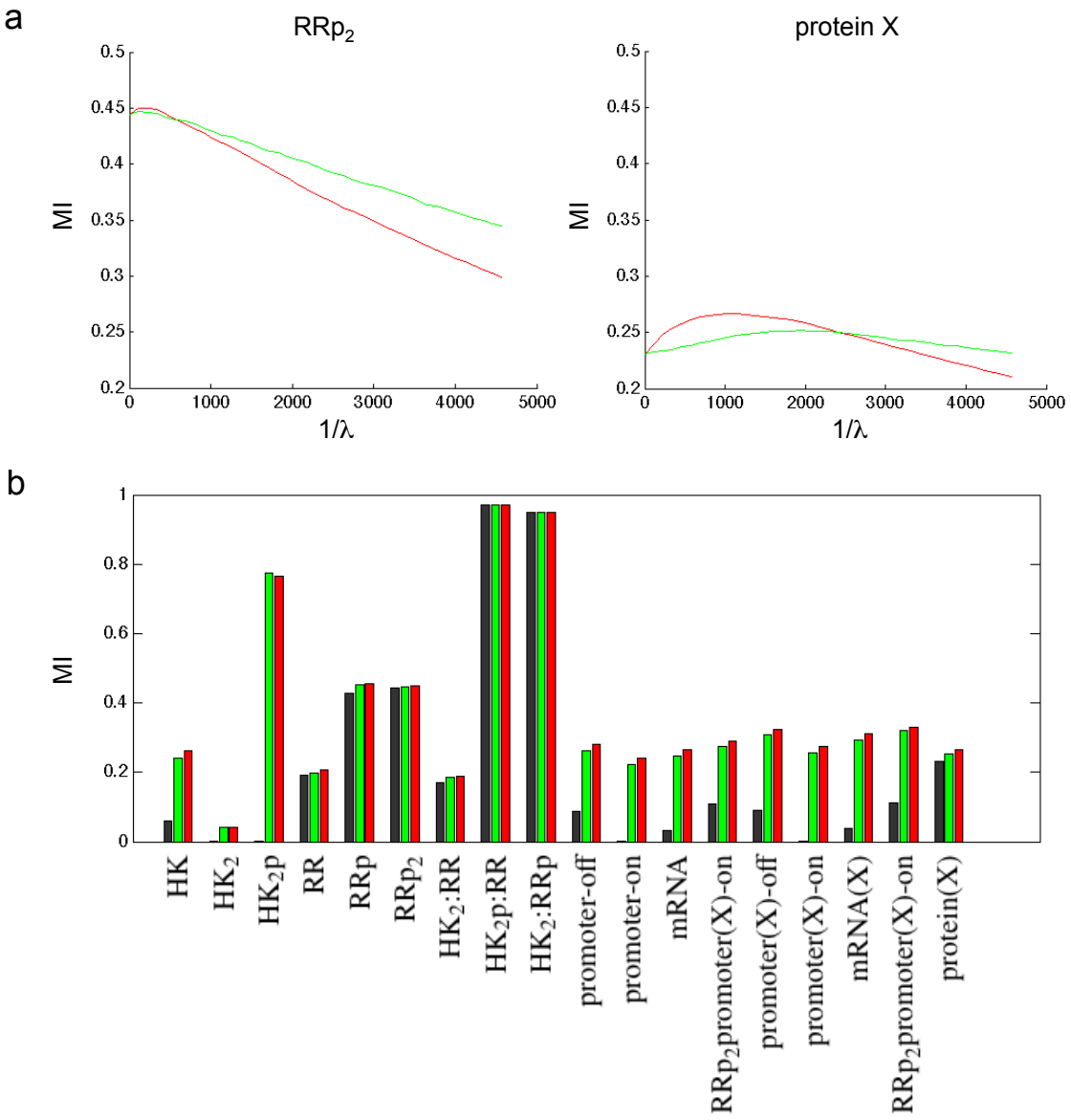

Figure 5.5: Extracting information from species trajectories. (**a**) MI between final signal state and molecule number averages obtained using either a uniform (green) or an exponential (red) weighting plotted as a function of the averaging parameter $(1/\lambda)$. (**b**) Bar graph showing the highest MI values obtained between the final signal state and the molecule number at the final time point (black) or the average number of molecules using a uniform (green) or exponential (red) weighting.

may determine whether the system is more responsive to changes in HK kinase or HK phosphatase activities. Models 2 and 3 illustrate two possible alternatives. These observations are related to effects, previously termed, retroactivity [116] and load-induced modulation [117], which have been investigated using other models of cellular signalling. The sensitivity of the models to parameters governing complex formation and dissociation (Figure 5.3) suggests that the affinity of protein interactions may be a feature of the system that allows for fine tuning, either through natural selection or genetic engineering.

The focus on extrinsic noise and information transmission makes our modelling efforts complementary to other recently published TCS models [107–109]. While increasing the model complexity by extending it to include transcription and translation further increases uncertainty about the quantitative fidelity of these models, the qualitative behaviour shown here appears to be plausible and illustrates both the degree of noise that needs to be taken into consideration and the potential for robust signalling in what may appear to be an extremely noisy system.

Extracting information

MI can be thought of as measuring the rate of information transfer. If the cellular machinery could "read" the number of molecules present in the system at regular time intervals, $I(S,X)$ could be interpreted directly as a rate of information transfer. However, in reality, a cell does not have access to the complete trajectory over any period of time. Instead, downstream processes could act as a kind of memory, responding to the state of the system and recording a moving average.

From the point of view of extracting more information about the input signal state, the benefits of averaging molecule counts over a period of time vary depending on characteristics of the trajectory of the species in question as well as the trajectory of the input signal. When averaging over longer time periods, information about the state of the signal will be lost as the chance of the signal changing during the relevant time period increases. This loss of information may be counterbalanced by the benefit of averaging out the intrinsic stochasticity, or intrinsic noise, in the system. Promoter states and mRNA molecule counts provide examples of the benefit of taking an average over time (Figure 5.5b). For these species there is very little MI between the signal state and the number of molecules at the final time point. Nonetheless, information about the signal

state is not lost. Instead, more of it is encoded in the temporal changes to molecule numbers. For a more abundant species, such as protein X, taking the average over a period of time does not bring as much of an improvement. For signals that change on a much slower time scale, averaging may be more advantageous even when the species molecule count is high. It may allow downstream processes to accumulate information about the input signal over a longer period of time, leading to a more precise response.

Higher MI values could be obtained by optimising model and signal parameters but this was not the goal here. The parameters were chosen primarily with the aim of obtaining biologically plausible values, based on current literature, and not with the aim of attaining a particular MI. We also note that other parameter values may drive the model into alternative regimes. For example increasing the affinity of the RR for the promoter can result in the system being very slow to respond to a decrease in the input signal. The signal was then chosen arbitrarily to provide an illustrative case for the kind of features of biomolecular signal transduction that should be taken into consideration and may be relevant in the real system.

In reality the relationship between the signal being sensed and the response of the system is likely to be much more complex. If the expression of TCS components is significantly altered by a prolonged stimulus, the system may become more sensitive to the signal over time. Thus, the rate of information transmission about the signal state may increase or decrease in response to longer term changes in the signal.

Chapter 6

Conclusion

6.1 Defining noise

Information theory provides a holistic approach for understanding cellular signalling. However, it does not necessarily yield simple answers. Instead we are faced with difficult questions about the precise nature of signalling mechanisms, the probability distribution of signals being sensed by the cell and what it is that we mean when we talk about noise. In this work, we have always represented extrinsic noise as a relatively small Gaussian perturbation of the model parameters. The precise definition of extrinsic noise depends on the system in question and the particular context of the model. Consider, for example, a rapidly acting signalling pathway which needs to reliably interpret changes in the environment despite the presence of cell-to-cell variability. The most relevant model of extrinsic noise may be the distribution of the numbers of relevant proteins within each cell. In this case, a gamma distribution [17] may be a better model for noise than a truncated Gaussian. The extent of correlation between the numbers of components of the signalling pathway also becomes an important question. Alternatively if the system in question is responsible for regulating homeostatic processes within the cell, some of the cell-to-cell variability can no longer be considered as noise and instead becomes part of the signal. A concrete example of this is the cell size checkpoint in the yeast cell cycle [118].

6.2 Implications

Information in context

The problem of picking an appropriate context when estimating the entropy of a variable is an old one in information theory [119]. The context determines the probabilities of outcomes and thus the entropy. One example is the entropy of a piece of text. Should it be calculated based on the frequency of letters or words in the relevant language? Even the boundaries of a language are fuzzy. The same problem exists in the biological context. It becomes most obvious when trying to define a probability distribution for an environmental variable sensed by a cell or organism. For simplicity, the input signals considered in this work are limited to two possible states. This permitted us to abstract away not only uncertainty about the signal distribution but also a layer of complexity created by the non-linear relationship between an external signal and the response of a receptor complex [108, 120]. An applied analysis of information transmission by a cellular signalling pathway calls for a statistical analysis of the environmental variable being sensed by it as well as the output of the sensory system. Nonetheless, even without knowledge of the true probability distribution of the signal in the environment, an information theoretic approach may be taken to carry out a comparative analysis of different signalling pathways (or models) subject to a proposed input signal distribution.

Motifs

Our work with motifs illustrates the importance of the timescales of fluctuations in the signal in relation to the response time of the signal transduction system. From the perspective of transmitting information there is no fundamental difference between activation and deactivation. However, the response time of the system may be largely determined by deactivation or degradation rates. The practical implication of this observation is that when investigating biological signalling it is worth considering not only the dynamics of the response to the appearance of a signal but also to its removal.

The importance of protein degradation rates for signalling dynamics [121] and the propagation of noise [122] has been noted previously. Here we only consider first order degradation or dilution but even this simple model highlights its importance for signal transduction. Protein degradation in eukaryotic cells is facilitated by specialised cellular machinery [123] and complex regulatory mechanisms [124]. What effect do these have

on the response times of signalling systems? How do they affect the 'memory' of the system and its ability to integrate out intrinsic stochasticity? Whether these questions have general answers or are only meaningful in the context of a particular signalling system remains to be resolved.

Inferring noise

Distinguishing between intrinsic and extrinsic noise can be problematic both conceptually and practically. Conceptually, what is considered to be extrinsic depends on the definition of the system in question. Practically, the observed distribution of a measurable variable is an amalgamation of the effects of both types of noise. Considering noise in gene expression is convenient from a conceptual point of view. In Chapter 5 we demonstrate how approximate Bayesian computation in combination with a sufficiently fast simulation method may be used to tackle the practical problem inferring the origin and magnitude of extrinsic noise.

Bacterial signalling systems

By studying a stochastic model of a two-component signalling system we are able to show how such a system is capable of transmitting information about an extracellular signal despite considerable variability in the the number of protein molecules caused by both intrinsic and extrinsic noise. A more general observation from this piece of work is that the amount of information that can be extracted by cellular signalling systems over time depends on how the temporal trajectory of molecule numbers is processed by downstream components. This point relates back to the tradeoff between extracting information about the value of a variable at a given time point as opposed to information about its average value over a period of time as observed for models of signalling motifs.

Information content of trajectories

When considered together, the results presented in Chapters 3 and 5 suggest that intrinsic and extrinsic noise create different obstacles for achieving effective signal transduction. Extrinsic noise may place an upper limit on the amount of information transmitted. An apparently noisy time series may still contain information, masked by intrinsic stochasticity. This information may potentially be extracted by a downstream process which

measures and remembers the system state. For a cell, extracting this information incurs a cost. This cost may be born metabolically, by synthesising more components of a signalling pathway and allowing a larger number of energy consuming reactions or processes to occur. It may also be born by sacrificing response time and relying on the memory of the system to integrate out the noise in the signal over a longer period of time.

While terms such as "measure" and "remember" may appear anthropomorphic, they can be defined precisely for specific biomolecular systems, in terms of changes in molecule numbers, reaction rates, concentration gradients or other quantifiable biochemical properties. As with the precise definition of extrinsic noise, their meaning depends on the context.

6.3 Ideas for future work

The modelling of motifs could be extended to more complex topologies with higher numbers of nodes. This would probably need to be done in a more selective fashion as the number of possible motif topologies increases sharply with the number of nodes in the motif. The computational cost of each simulation also increases. Perhaps a more interesting approach would be to consider a selection of model and parameter combinations chosen on the basis of their qualitative temporal dynamics. Non-linear dynamics such as adaptation, oscillation and chaotic behaviour could be studied to determine their effect on MI. On a similar note, one could consider a range of more complex temporal signal trajectories and then attempt to find motif and parameter combinations capable of extracting information about particular properties of the signal. The representation of signalling motifs using stochastic models may provide another avenue for future investigations.

The work presented in Chapter 4 could be continued in the same vein by trying to find more efficient methods of simulating other biologically relevant models. On the parameter inference front, additional types of experimental data, such as measurements of protein degradation rate, could potentially be used to further constrain posterior distribution of both rate and noise parameters. The two-state model used here can be further extended to explicitly include cell division with binomial partitioning of molecules between the daughter cells. If this were to be done, measurements of cell cycle duration

and chromosomal duplication times, as well as the variability in these values, may prove useful in constraining posterior parameter distributions.

Theoretically, the magnitude of extrinsic noise affecting different genes should be similar. It may be worthwhile to try to find a vector of noise parameters that can satisfy a particular distance metric for the majority of genes. There is no obvious correct way of doing this and the high dimensionality of the model parameter space may prove problematic. However if the prior distribution of noise parameters could be constrained or even fixed to a point value, the inference of reaction rate parameters may require fewer computational steps. One caveat to this is that the two-state model will not necessarily be applicable to genes which are undergoing changes in their expression level.

One natural extension of the work done on the model of a two-component system presented in Chapter 5 may be to consider the effect of dimerisation of the response regulator. This could involve constructing alternative models in which the response regulator dimerises regardless of its phosphorylation state. Dephosphorylation of the dimeric form could also be included. Further studied could consider the effects of oligomerisation of the response regulator. Model validation could potentially be carried out using quantitative data about an input signal, such as the concentration of a particular molecule in the growth media, coupled to measurements of the TCS's activity. MI depends on the distributions of both the response and the input signal. Obtaining a sufficient number of samples by varying the input signal experimentally and measuring the system's output may be too costly as discussed above. However, if the distribution of input signals is known and a satisfactory model of the system is available, simulated data could be used to obtain estimates of the MI between the signal and the response.

6.4 Applying information theory

From a practical perspective, the application of estimating mutual information (MI) is limited by the number of data points obtained in experiments. Traditional experimental techniques in molecular biology, such as gel electrophoresis or quantitative PCR, often yield only tens of data points. This makes it impractical to use such measurements to estimate MI directly. Difficulties also arise from the potentially complex relationship between the number of data points and parameters of the estimator such as the number of bins or the bandwidth in a kernel density estimator (see Chapter 2). One approach for overcoming these issues, is to use the available data to construct a plausible model

of the system. This model could then be used to obtain sufficient samples from the joint probability distribution.

With the continued development of high-throughput experimental methods the quantity of data generated continues to increase. One notable example is flow cytometry experiments in which the fluorescence intensity from two different markers is measured simultaneously [125]. Such experiments can provide measurements from thousands of individual cells in a sample in a matter of minutes. Each one of these measurements could effectively correspond to a point in the joint probability distribution of variables quantifying the abundances of the two different markers. Perhaps MI could be a useful measure of their potential for identifying cell lineages or different regimes of cellular function [126]. In this case, estimates of MI may be a suitable substitute for a parametric model of the relationship between such variables.

More generally, information theoretic measures should not be seen as a substitute for parametric models of signalling systems. Building a model may be beneficial for exposing our underlying assumptions about the mechanism in question and the manner in which information is transmitted through the system. These models need not necessarily be detailed mechanistic models. The key point is that the model should capture the behaviour that is most relevant for signal transduction. It is worth making an effort to determine whether the properties of the model that are thought to describe information flow, correspond to those which encode a significant proportion of this information in the real system. These features may be molecule numbers or concentrations. Alternatively information may be encoded in the timescales or frequencies with which biochemical changes occur.

One of the strengths of Shannon's formulation of information is that it disassociates the concepts of information and meaning. Estimates or calculations of entropy or mutual information for any given variable(s) do not require a justification in of themselves. However, when the results of quantifying information flow through signalling systems are interpreted, this quantity is imbued with meaning by the interpreter. This is where caution should be taken to make assumptions explicit.

References

[1] O. Penner, P. Grassberger, M. Paczuski, Sequence alignment, mutual information, and dissimilarity measures for constructing phylogenies., PLoS ONE 6 (2011) e14373.

[2] G. Tkačik, A. M. Walczak, W. Bialek, Optimizing information flow in small genetic networks. III. A self-interacting gene, Physical Review E 85 (2012) 041903.

[3] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, I. Golding, General properties of transcriptional time series in Escherichia coli, Nature Genetics 43 (2011) 554–560.

[4] R. Cheong, A. Rhee, C. Wang, I. Nemenman, Information transduction capacity of noisy biochemical signaling networks., Science 334 (2011) 354–358.

[5] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, W. Bialek, Positional information, in bits., Proceedings of the National Academy of Sciences 110 (2013) 16301–16308.

[6] J. Liepe, S. Filippi, M. Komorowski, M. P. H. Stumpf, Maximizing the information content of experiments in systems biology, PLoS Computational Biology 9 (2013) e1002888.

[7] A. Pomerance, E. Ott, M. Girvan, W. Losert, The effect of network topology on the stability of discrete state models of genetic control, Proceedings of the National Academy of Sciences 106 (2009) 8209–8214.

[8] T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation., Nature (2009).

[9] M. C. Gibson, Bicoid by the numbers: quantifying a morphogen gradient., Cell 130 (2007) 14–16.

[10] J. Elf, M. Ehrenberg, Fast evaluation of fluctuations in biochemical networks with the linear noise approximation., Genome Research (2003) 2475–2484.

[11] A. Ale, P. Kirk, M. P. H. Stumpf, A general moment expansion method for stochastic kinetic models, The Journal of Chemical Physics 138 (2013) 174101.

[12] M. A. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, The Journal of Physical Chemistry A 104 (2000) 1876–1889.

[13] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, Journal of computational physics 22 (1976) 403–434.

[14] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, C. Tang, Defining network topologies that can achieve biochemical adaptation, Cell 138 (2009) 760–773.

[15] P. J. Ingram, M. P. H. Stumpf, J. Stark, Network motifs: structure does not determine function., BMC Genomics 7 (2006) 108.

[16] Y. Fu, T. Glaros, M. Zhu, P. Wang, Z. Wu, J. J. Tyson, L. Li, J. Xing, Network topologies and dynamics leading to endotoxin tolerance and priming in innate immune cells., PLoS Computational Biology 8 (2012) e1002526.

[17] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells., Science 329 (2010) 533–538.

[18] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, J. S. Weissman, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise., Nature 441 (2006) 840–846.

[19] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal XXVII-3 (1948) 379–423.

[20] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, Inc., 1991.

[21] N. Brunel, J.-P. Nadal, Mutual information, Fisher information, and population coding, Neural Computation 10 (1998) 1731–1757.

[22] R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy–a model-free measure of effective connectivity for the neurosciences., Journal of computational neuroscience 30 (2011) 45–67.

[23] A. Tiwari, O. A. Igoshin, Coupling between feedback loops in autoregulatory networks affects bistability range, open-loop gain and switching times, Physical biology 9 (2012) 055003.

[24] G. Tkacik, C. G. Callan, W. Bialek, Information flow and optimization in transcriptional regulation, Proceedings of the National Academy of Sciences 105 (2008) 12265–12270.

[25] G. Tkačik, T. Gregor, W. Bialek, The role of input noise in transcriptional regulation, PLoS ONE 3 (2008) e2774.

[26] G. Tkačik, C. G. Callan, W. Bialek, Information capacity of genetic regulatory elements., Physical review. E, Statistical, nonlinear, and soft matter physics 78 (2008) 011910.

[27] G. Tkačik, A. Walczak, W. Bialek, Optimizing information flow in small genetic networks, Physical Review E 80 (2009) 031920.

[28] G. Tkačik, A. M. Walczak, Information transmission in genetic regulatory networks: a review., Journal of physics. Condensed matter : an Institute of Physics journal 23 (2011) 153102.

[29] T. J. Kobayashi, Implementation of dynamic Bayesian decision making by intracellular kinetics., Physical review letters 104 (2010) 228104.

[30] M. A. J. Roberts, A. Papachristodoulou, J. P. Armitage, Adaptation and control circuits in bacterial chemotaxis., Biochemical Society Transactions 38 (2010) 1265–1269.

[31] T. Reuter, Fifty years of dark adaptation 1961-2011., Vision research 51 (2011) 2243–2262.

[32] D. Heitzler, G. Durand, N. Gallay, A. Rizk, S. Ahn, J. Kim, J. D. Violin, L. Dupuy, C. Gauthier, V. Piketty, P. Crépieux, A. Poupon, F. Clément, F. Fages, R. J. Lefkowitz, E. Reiter, Competing G protein-coupled receptor kinases balance G protein and β-arrestin signaling., Molecular Systems Biology 8 (2012) 590.

[33] T. J. Perkins, P. S. Swain, Strategies for cellular decision-making, Molecular Systems Biology 5 (2009).

[34] E. Ziv, I. Nemenman, C. H. Wiggins, Optimal signal processing in small stochastic biochemical networks., PLoS ONE 2 (2007) e1077.

[35] A. M. Walczak, G. Tkačik, W. Bialek, Optimizing information flow in small genetic networks. II. Feed-forward interactions, Physical Review E 81 (2010) 041905.

[36] S. S. Mc Mahon, A. Sim, S. Filippi, R. Johnson, J. Liepe, D. Smith, M. P. H. Stumpf, Information theory and signal transduction systems: From molecular information processing to network inference., Seminars in cell & developmental biology 35 (2014) 98–108.

[37] P. E. Meyer, F. Lafitte, G. Bontempi, minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information., BMC Bioinformatics 9 (2008) 461.

[38] C. Cellucci, A. Albano, P. Rapp, Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms, Physical Review E 71 (2005) 066208.

[39] A. Fraser, H. Swinney, Independent coordinates for strange attractors from mutual information, Physical Review A 33 (1986) 1134–1140.

[40] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Physical Review E 69 (2004).

[41] Z. I. Botev, J. F. Grotowski, D. P. Kroese, Kernel density estimation via diffusion, The Annals of Statistics 38 (2010) 2916–2957.

[42] P. Krik, T. Thorne, M. Stumpf, Model selection in systems and synthetic biology, Current opinion in biotechnology 24 (2013) 767–774.

[43] J. John E. Dennis, R. B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations, Prentice-Hall, Inc., 1983.

[44] J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, M. P. H. Stumpf, A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation., Nature Protocols 9 (2014) 439–456.

[45] C. P. Robert, Bayesian computational tools, arXiv.org (2013).

[46] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, M. W. Feldman, Population growth of human Y chromosomes: a study of Y chromosome microsatellites., Molecular biology and evolution 16 (1999) 1791–1798.

[47] P. Marjoram, J. Molitor, V. Plagnol, S. Tavare, Markov chain Monte Carlo without likelihoods, Proceedings of the National Academy of Sciences 100 (2011) 15324–15328.

[48] S. Sisson, Fan, M. Tanaka, Sequential Monte Carlo without likelihoods, Proceedings of the National Academy of Sciences 104 (2007) 7.

[49] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems., Journal of the Royal Society, Interface / the Royal Society 6 (2009) 187–202.

[50] S. Filippi, C. P. Barnes, J. Cornebise, M. P. H. Stumpf, On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo, Statistical Applications in Genetics and Molecular Biology 12 (2013) 87–107.

[51] K. Scranton, J. Knape, P. de Valpine, An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model, Ecology 95 (2014) 1418–1428.

[52] P. Fearnhead, D. Prangle, Constructing summary statistics for approximate Bayesian computation: semiautomatic approximate Bayesian computation, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74 (2012) 419–474.

[53] P. Marjoram, A. Zubair, S. V. Nuzhdin, Post-GWAS: where next? More samples, more SNPs or more biology?, Heredity 112 (2013) 79–88.

[54] S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of Escherichia coli, Nature Genetics 31 (2002) 64–68.

[55] K. Wuichet, I. B. Zhulin, Origins and diversification of a complex signal transduction system in prokaryotes., Science Signaling 3 (2010) ra50.

[56] M. Babu, N. Luscombe, L. Aravind, M. Gerstein, S. Teichmann, Structure and evolution of transcriptional regulatory networks, Current Opinion in Structural Biology 14 (2004) 283–291.

[57] G. C. Conant, A. Wagner, Convergent evolution of gene circuits., Nature Genetics 34 (2003) 264–266.

[58] J. J. Hughey, T. K. Lee, M. W. Covert, Computational modeling of mammalian signaling networks, Journal of Chemical Physics 2 (2009) 194–209.

[59] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, Molecular cell biology, 4th edition, W.H.Freeman & Co Ltd, 2000.

[60] U. Alon, Network motifs: theory and experimental approaches, Nature Reviews Genetics 8 (2007) 450–461.

[61] S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif., Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 11980–11985.

[62] T. Kuhlman, Z. Zhang, M. H. Saier, T. Hwa, Combinatorial transcriptional control of the lactose operon of Escherichia coli, Proceedings of the National Academy of Sciences 104 (2007) 6043–6048.

[63] N. A. Shah, C. A. Sarkar, Robust network topologies for generating switch-like cellular responses., PLoS Computational Biology 7 (2011) e1002085.

[64] K. A. Fujita, Y. Toyoshima, S. Uda, Y.-i. Ozaki, H. Kubota, S. Kuroda, Decoupling of receptor and downstream signals in the Akt pathway by its low-pass filter characteristics., Science Signaling 3 (2010) ra56.

[65] P. Wei, W. W. Wong, J. S. Park, E. E. Corcoran, S. G. Peisajovich, J. J. Onuffer, A. Weiss, W. A. Lim, Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells, Nature 488 (2012) 384–388.

[66] I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, Real-Time kinetics of gene activity in individual bacteria, Cell (2005).

[67] J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, Transcriptional pulsing of a developmental gene, Current biology 16 (2006) 1018–1025.

[68] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS biology 4 (2006) e309.

[69] D. Zenklusen, D. R. Larson, R. H. Singer, Single-RNA counting reveals alternative modes of gene expression in yeast, Nature Structural & Molecular Biology 15 (2008) 1263–1271.

[70] R. Z. Tan, A. van Oudenaarden, Transcript counting in single cells reveals dynamics of rDNA transcription - Tan - 2010 - Molecular Systems Biology - Wiley Online Library, Molecular Systems Biology (2010).

[71] N. Rosenfeld, Gene regulation at the single-Cell level, Science 307 (2005) 1962–1965.

[72] M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell, Science 297 (2002) 1183–1186.

[73] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, P. K. Sorger, Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis, Nature 459 (2009) 428–432.

[74] I. G. Johnston, B. Gaal, R. P. das Neves, T. Enver, F. J. Iborra, N. S. Jones, Mitochondrial variability as a source of extrinsic cellular noise, PLoS Comput. Biol. (2012).

[75] B. B. Kaufmann, A. van Oudenaarden, Stochastic gene expression: from single molecules to the proteome, Current Opinion in Genetics & Development 17 (2007) 107–112.

[76] T. Toni, B. Tidor, Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology, PLoS Comput. Biol. (2013).

[77] L. Cai, N. Friedman, X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, Nature 440 (2006) 358–362.

[78] A. Raj, A. van Oudenaarden, Nature, nurture, or chance: Stochastic gene expression and its consequences, Cell 135 (2008) 216–226.

[79] C. P. Barnes, S. Filippi, M. Stumpf, T. Thorne, Considerate approaches to constructing summary statistics for ABC model selection - Springer, Statistics and Computing (2012).

[80] J. M. Levsky, R. H. Singer, Fluorescence in situ hybridization: past, present and future., Journal of cell science 116 (2003) 2833–2838.

[81] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nature Reviews Genetics 10 (2009) 57–63.

[82] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, JSTOR: Journal of the Royal Statistical Society. Series B (Methodological) 57 (1995) 289–300.

[83] K. Erguler, M. P. H. Stumpf, Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models, Molecular BioSystems 7 (2011) 1593.

[84] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models., PLoS Computational Biology 3 (2007) 1871–1878.

[85] M. Secrier, T. Toni, M. P. H. Stumpf, The ABC of reverse engineering biological signalling systems, Molecular BioSystems 5 (2009) 1925.

[86] Z. Bar-Joseph, D. K. Gifford, T. S. Jaakkola, Fast optimal leaf ordering for hierarchical clustering, Bioinformatics 17 (2001) S22–S29.

[87] J. Kacmar, A. Zamamiri, R. Carlson, N. R. Abu-Absi, F. Srienc, Single-cell variability in growing Saccharomyces cerevisiae cell populations measured with automated flow cytometry., Journal of biotechnology 109 (2004) 239–254.

[88] T. L. Yuan, G. Wulf, L. Burga, L. C. Cantley, Cell-to-Cell variability in PI3K protein level regulates PI3K-AKT pathway activity in cell populations, Current biology 21 (2011) 173–183.

[89] B. Li, L. You, Predictive power of cell-to-cell variability, Quantitative Biology (2013).

[90] J. Peccoud, B. Ycart, Markovian modeling of gene-Product synthesis, Theoretical population biology (1995).

[91] A. R. Stinchcombe, C. S. Peskin, D. Tranchina, Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression, Physical Review E 85 (2012) 061919.

[92] V. Shahrezaei, P. S. Swain, Analytical distributions for stochastic gene expression, Proceedings of the National Academy of Sciences 105 (2008) 17256–17261.

[93] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, Journal of Chemical Physics 115 (2001) 1716–1733.

[94] G. Lillacci, M. Khammash, The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations, Bioinformatics 29 (2013) 2311–2319.

[95] D. Silk, P. D. W. Kirk, C. P. Barnes, T. Toni, A. Rose, S. Moon, M. J. Dallman, M. P. H. Stumpf, Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes., Nature Communications 2 (2011) 489.

[96] R. Gao, A. M. Stock, Biological insights from structures of two-component proteins., Annual Review of Microbiology 63 (2009) 133–154.

[97] K. Jung, L. Fried, S. Behr, R. Heermann, Histidine kinases and response regulators in networks, Current Opinion in Microbiology 15 (2012) 118–124.

[98] X. Sheng, M. Huvet, J. W. Pinney, M. P. H. Stumpf, Evolutionary characteristics of bacterial two-component systems., Advances in experimental medicine and biology 751 (2012) 121–137.

[99] T. Maeda, S. M. Wurgler-Murphy, H. Saito, A two-component system that regulates an osmosensing MAP kinase cascade in yeast, Nature 369 (1994) 242–245.

[100] I. Hwang, J. Sheen, Two-component circuitry in Arabidopsis cytokinin signal transduction, Nature 413 (2001) 383–389.

[101] Y. Göpel, B. Görke, Rewiring two-component signal transduction with small RNAs, Current Opinion in Microbiology 15 (2012) 132–139.

[102] A. Möglich, R. A. Ayers, K. Moffat, Design and signaling mechanism of light-regulated histidine kinases., Journal of molecular biology 385 (2009) 1433–1444.

[103] D. Albanesi, M. Martín, F. Trajtenberg, M. C. Mansilla, A. Haouz, P. M. Alzari, D. de Mendoza, A. Buschiazzo, Structural plasticity and catalysis regulation of a thermosensor histidine kinase., Proceedings of the National Academy of Sciences 106 (2009) 16185–16190.

[104] R. J. Lewis, D. J. Scott, J. A. Brannigan, J. C. Ladds, M. A. Cervin, G. B. Spiegelman, J. G. Hoggett, I. Barák, A. J. Wilkinson, Dimer formation and transcription activation in the sporulation response regulator Spo0A., Journal of molecular biology 316 (2002) 235–245.

[105] P. Perron-Savard, G. De Crescenzo, H. Le Moual, Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent., Microbiology (Reading, England) 151 (2005) 3979–3987.

[106] I. Virlogeux, H. Waxin, C. Ecobichon, J. O. Lee, M. Y. Popoff, Characterization of the rcsA and rcsB genes from Salmonella typhi: rcsB through tviA is involved in regulation of Vi antigen synthesis., Journal of bacteriology 178 (1996) 1691–1698.

[107] C. P. Barnes, D. Silk, X. Sheng, M. P. H. Stumpf, Bayesian design of synthetic biological systems, Proceedings of the National Academy of Sciences of the United States of America 108 (2011) 15190–15195.

[108] M. Amin, S. L. Porter, O. S. Soyer, Split histidine kinases enable ultrasensitivity and bistability in two-component signaling networks, PLoS Computational Biology 9 (2013) e1002949.

[109] A. M. Kierzek, L. Zhou, B. L. Wanner, Stochastic kinetic model of two component system signalling reveals all-or-none, graded and mixed mode stochastic switching responses , Molecular BioSystems 6 (2010) 531–542.

[110] O. Ashenberg, K. Rozen-Gagnon, M. T. Laub, A. E. Keating, Determinants of homodimerization specificity in histidine kinases., Journal of molecular biology 413 (2011) 222–235.

[111] R. Gao, A. M. Stock, Probing kinase and phosphatase activities of two-component systems in vivo with concentration-dependent phosphorylation profiling., Proceedings of the National Academy of Sciences 110 (2013) 672–677.

[112] K. Yamamoto, K. Hirao, T. Oshima, H. Aiba, R. Utsumi, A. Ishihama, Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli., The Journal of biological chemistry 280 (2005) 1448–1456.

[113] M. Schlosshauer, D. Baker, Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness., Protein science : a publication of the Protein Society 13 (2004) 1660–1669.

[114] M. Komorowski, J. Zurauskiene, M. P. H. Stumpf, StochSens–matlab package for sensitivity analysis of stochastic chemical systems, Bioinformatics 28 (2012) 731–733.

[115] A. Narayanan, S. Kumar, A. N. Evrard, L. N. Paul, D. A. Yernool, An asymmetric heterodomain interface stabilizes a response regulator-DNA complex., Nature Communications 5 (2014) 3282.

[116] D. Del Vecchio, A. J. Ninfa, E. D. Sontag, Modular cell biology: retroactivity and insulation, Molecular Systems Biology 4 (2008).

[117] P. Jiang, A. C. Ventura, E. D. Sontag, S. D. Merajver, A. J. Ninfa, D. Del Vecchio, Load-induced modulation of signal transduction networks., Science Signaling 4 (2011).

[118] J. J. Turner, J. C. Ewald, J. M. Skotheim, Cell size control in yeast, Current biology 22 (2012) R350–R359.

[119] J. Gleick, The Information: A history, a theory, a flood, HarperCollins UK, 2011.

[120] L. Huang, C. Q. Pan, B. Li, L. Tucker-Kellogg, B. Tidor, Y. Chen, B. C. Low, Simulating EGFR-ERK signaling control by scaffold proteins KSR and MP1 reveals differential ligand-sensitivity co-regulated by Cbl-CIN85 and endophilin, PLoS ONE 6 (2011) e22933.

[121] N. E. Buchler, U. Gerland, T. Hwa, Nonlinear protein degradation and the function of genetic circuits., Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 9559–9564.

[122] M. Komorowski, J. Miekisz, M. Stumpf, Decomposing noise in biochemical signaling systems highlights the role of protein degradation, Biophysical Journal (2013).

[123] F. Luciani, C. K. mir, M. Mishto, M. Or-Guil, R. J. de Boer, A mathematical model of protein degradation by the proteasome, Biophysj 88 (2005) 2422–2432.

[124] D. Komander, M. Rape, The ubiquitin code, Annual Review of Biochemistry 81 (2012) 203–229.

[125] B. Mahata, X. Zhang, A. A. Kolodziejczyk, V. Proserpio, L. Haim-Vilmovsky, A. E. Taylor, D. Hebenstreit, F. A. Dingler, V. Moignard, B. Gttgens, W. Arlt, A. N. McKenzie, S. A. Teichmann, Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis, Cell reports 7 (2014) 1130–1142.

[126] D. Hebenstreit, A. Deonarine, M. M. Babu, S. A. Teichmann, Duel of the fates: the role of transcriptional circuits and noise in CD4+ cells., Current opinion in cell biology 24 (2012) 350–358.

Inferring extrinsic noise in *E. coli* gene expression

Derivation of the Poisson parameter

Consider the following system of reactions.

$$\emptyset \xrightarrow{\lambda} X$$
$$X \xrightarrow{\rho} Y$$
$$X \xrightarrow{\delta} \emptyset$$
$$Y \xrightarrow{\gamma} \emptyset$$

The system is allowed to run, starting with zero tokens of X and Y, from time t_0 until time t_m. The following table lists events of interest which can occur in the system and the symbols used to represent the number of times each even occurs in the time interval $T = t_m - t_0$.

Symbol	Event being counted
i	X is produced
j	X survived until time t_m
$\hat{i}$	X does not survive until time t_m (i.e. X is degraded or converted to Y)
$\hat{j}$	X is degraded
k	X is converted into Y
l	Y survives until time t_m

Deriving probability distributions for the numbers of events

A standard result is that the number of events which occur over a time period will have a Poisson distribution if the events are occurring with a uniform probability in a time interval. This applies to the production of X.

$$P(i) = \frac{(\lambda T)^i}{i!} e^{-\lambda T} \tag{1}$$

Where $P(i)$ is the probability that X is produced i times during the time period T.

Finding the number of surviving tokens of X

In order to find the probability distribution of the number of X tokens after the time period T we consider the probability that j tokens of X remain given that i tokens of X were produced.

$$P(j|i) = \binom{i}{j} p^j (1-p)^{i-j} \tag{2}$$

Where p is the probability that X survives given that it was produced during the time period T.

Let $P(j)$ be the probability that j tokens of X were produced during T and survived regardless of the number (i) of X tokens produced.

$$P(j) = \sum_{i=j}^{\infty} P(i)P(j|i) \tag{3}$$

Using equations 1 and 2.

$$P(j) = \sum_{i=j}^{\infty} \frac{(\lambda T)^i}{i!} e^{-\lambda T} \binom{i}{j} p^j (1-p)^{i-j}$$

$$= e^{-\lambda T} p^j \sum_{i=j}^{\infty} \frac{(\lambda T)^i}{i!} \frac{i!}{j!\,(i-j)!} p^j (1-p)^{i-j}$$

$$= \frac{e^{-\lambda T} p^j}{j!\,(1-p)^j} \sum_{i=j}^{\infty} \frac{q^i}{(i-j)!}$$

Where $q = \lambda T(1-p)$.
Considering just the summation term.

$$\sum_{i=j}^{\infty} \frac{q^i}{(i-j)!} = \frac{q^j}{0!} + \frac{q^{j+1}}{1!} + \frac{q^{j+2}}{2!} + \frac{q^{j+2}}{2!} + ... + \frac{j+i}{(i-j)!} + ...$$

$$= q^j \left(1 + \frac{q}{1} + \frac{q^2}{2!} + \frac{q^3}{3!} + ... + \frac{q^i}{(i-j)!} + ...\right)$$

$$= q^j \sum_{n=0}^{\infty} \frac{q^n}{n!} = q^j e^q$$

$$= (\lambda T)^j (1-p)^j e^{\lambda T(1-p)}$$

Thus, the expression for $P(j)$ can be written as:

$$P(j) = \frac{e^{-r} p^j (\lambda T)^j (1-p)^j e^{\lambda T(1-p)}}{j!\,(1-p)^j}$$

$$P(j) = \frac{(p\lambda T)^j}{j!} e^{-p\lambda T} \tag{4}$$

$P(j)$ has a Poisson distribution with the parameter $p\lambda T$.

Next, we need to find p. The probability that X was produced during a infinitesimally short time interval dt is given by dt/T. That is to say that X is produced once at some point during T and the probability of X production is uniform over T. The probability of a token of X produced at time t surviving until time t_m follows an exponential decay with a rate δ'. In this case $\delta' = \rho + \delta$. Thus the probability of a token of X being produced at time t and surviving until t_m is given by

$$\frac{dt}{T} e^{-\delta'(t_m - t)}$$

The overall probability that X survives given that it was produced can be found by integrating over the time period T.

$$p = \int_{t_0}^{t_m} \frac{dt}{T} e^{-\delta'(t_m - t)}$$

$$p = \frac{1}{\delta' T}\left(1 - e^{-\delta' T}\right) \tag{5}$$

Where $T = t_m - t_0$.

The numbers of the other events are also Poisson distributed. Similarly to $P(j)$, the number of tokens of X which did not survive ($P(\hat{i})$) will also be Poisson distributed.

$$P(\hat{i}) = \frac{p_{\hat{i}} \lambda T}{\hat{i}!} e^{-p_{\hat{i}} \lambda T}$$

Where $p_{\hat{i}} = (1 - p) = 1 - \frac{1}{\delta' T}\left(1 - e^{\delta' T}\right)$.

Also, for $P(k)$, the distribution of the number of tokens of Y produced

$$P(k) = \frac{(p_k \lambda T)^k}{k!} e^{-p_k \lambda T}$$

with $p_k = \frac{\rho}{\rho + \delta}(1 - p)$.

Examples of parameter posterior distributions

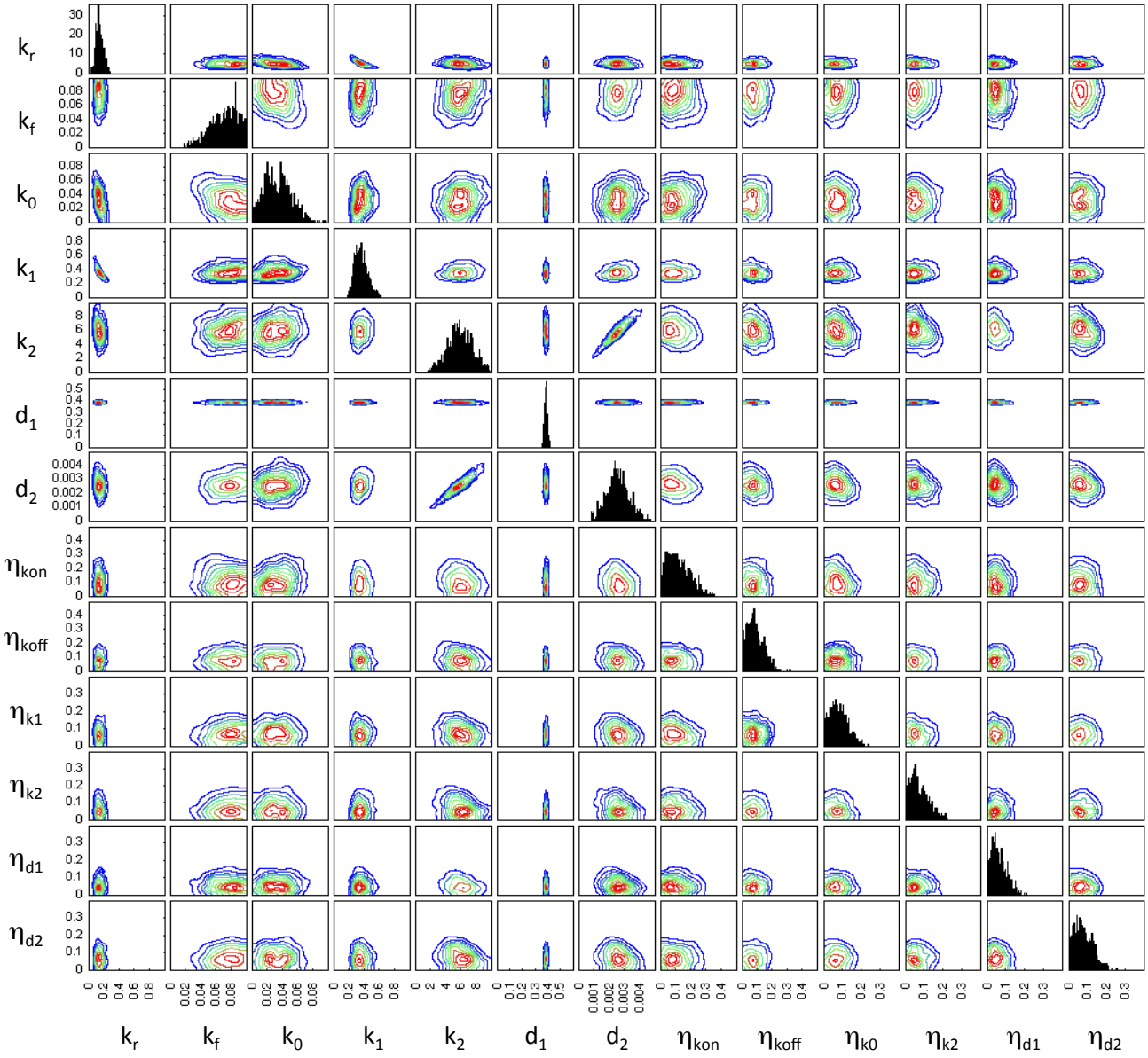

Figure 1: Posterior distribution of model parameters for the gene *rcsB*. Contour plots indicating the density of points with the corresponding parameter values for each particle in the final population.

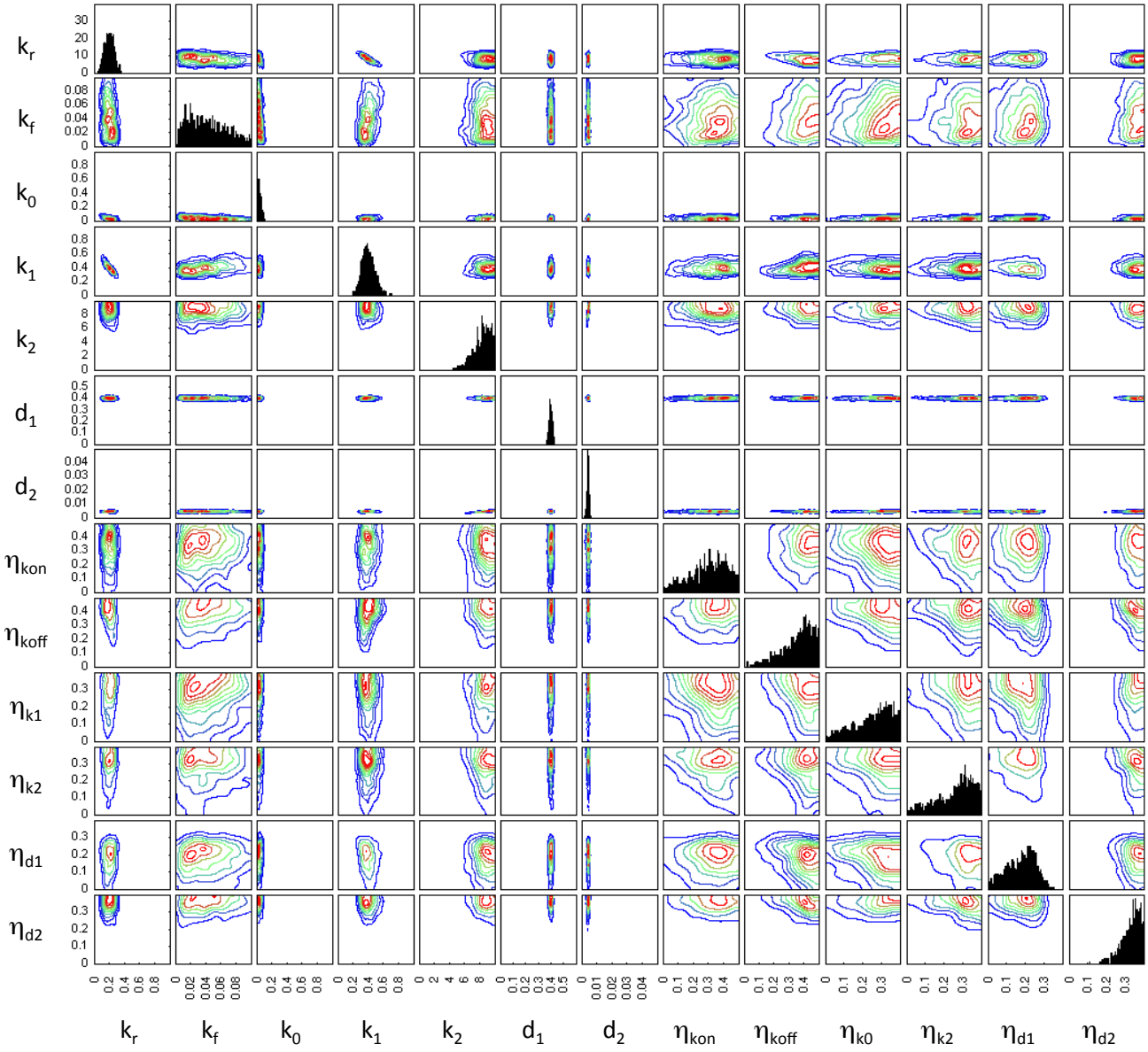

Figure 2: Posterior distribution of model parameters for the gene *yiiU*. Contour plots indicating the density of points with the corresponding parameter values for each particle in the final population.

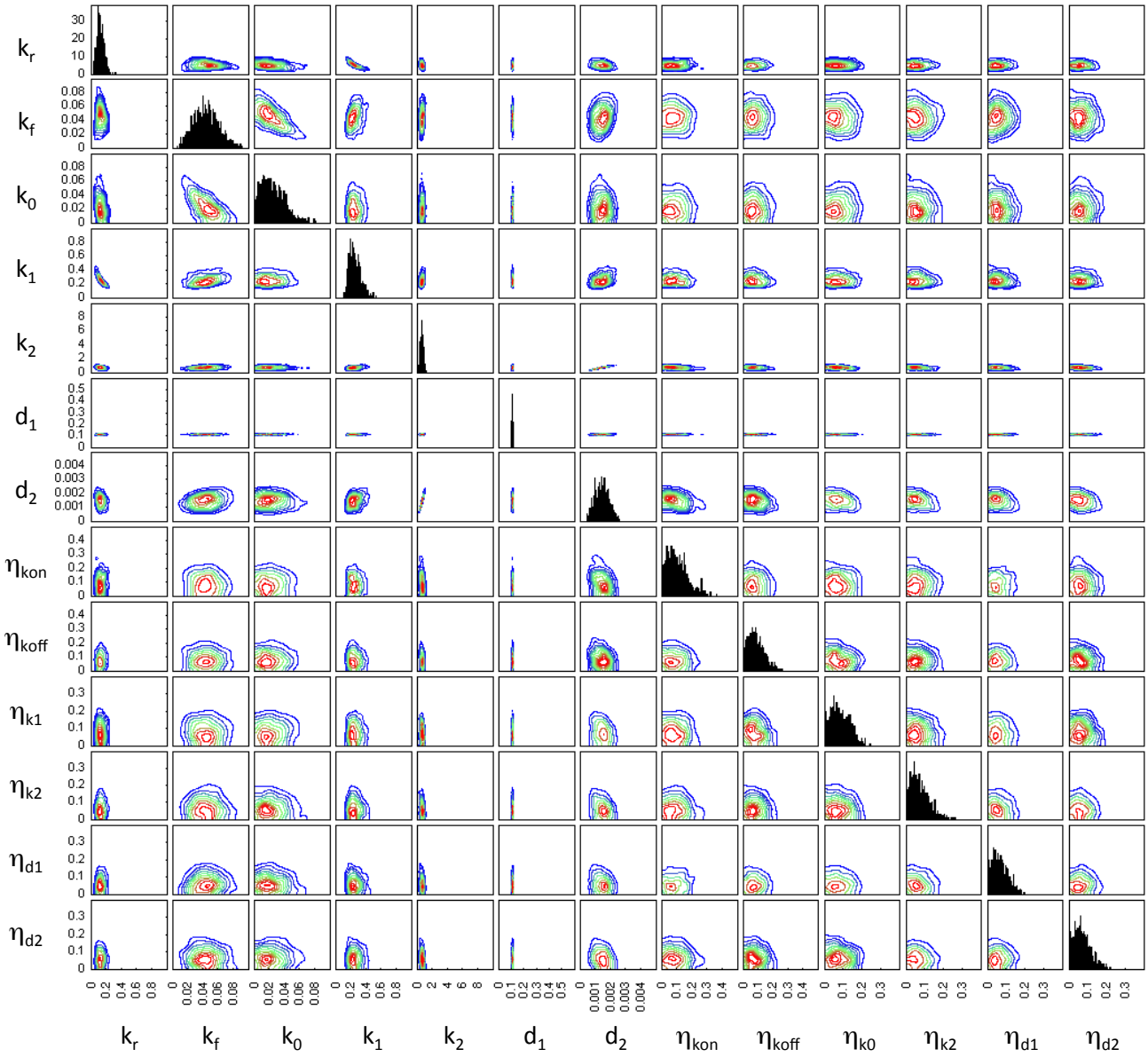

Figure 3: Posterior distribution of model parameters for the gene *yebC*. Contour plots indicating the density of points with the corresponding parameter values for each particle in the final population.

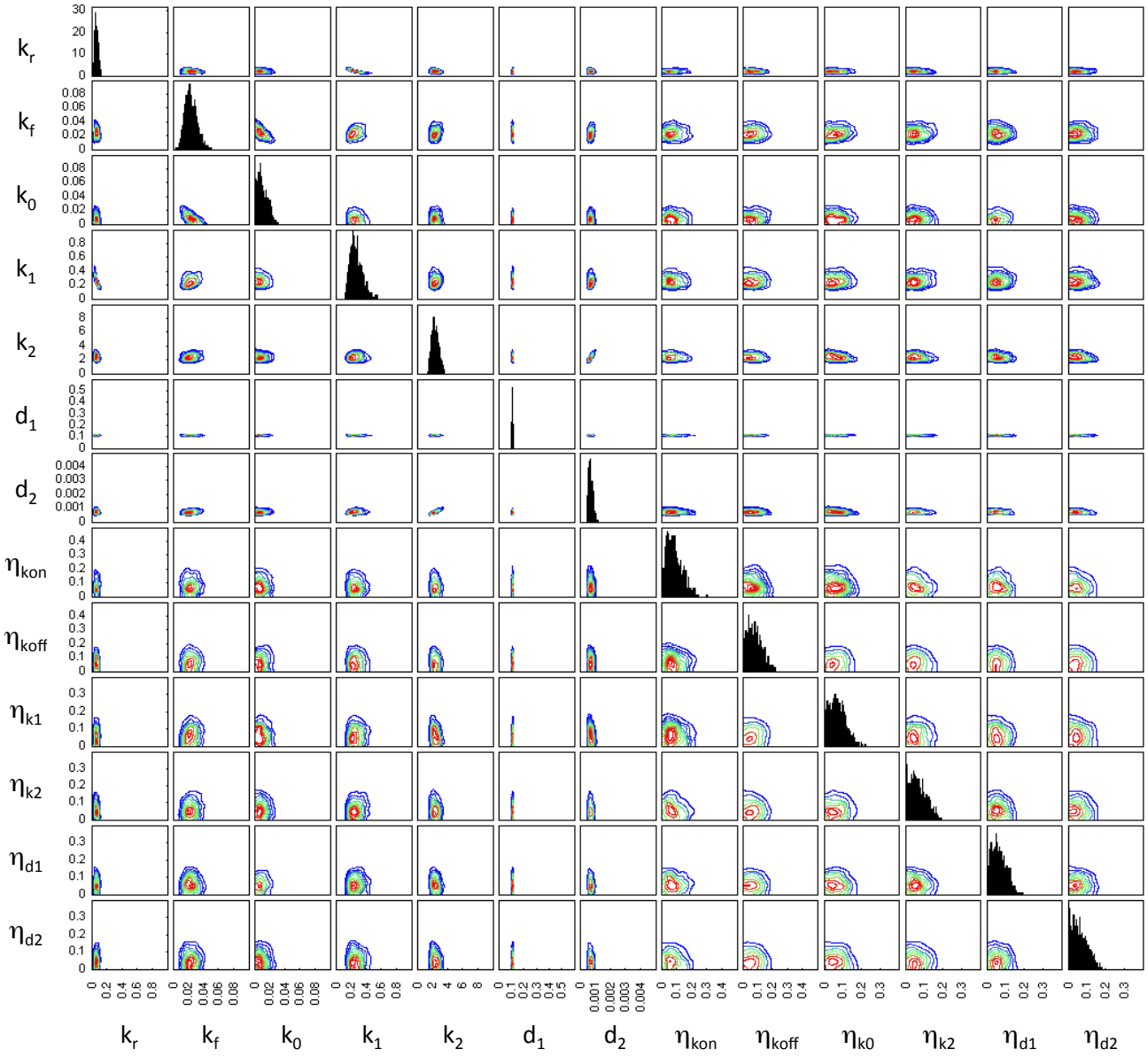

Figure 4: Posterior distribution of model parameters for the gene *eno*. Contour plots indicating the density of points with the corresponding parameter values for each particle in the final population.

Signal transduction by two-component systems

Model 1

$$2\mathrm{HK} \xrightarrow{c_1} \mathrm{HK_2} \qquad c_1 = 0.0125 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_{-1}} 2\mathrm{HK} \qquad c_{-1} = 0.005 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_k} \mathrm{HK_2p} \qquad c_k = 0.025 s^{-1}$$

$$\mathrm{HK_2p} \xrightarrow{c_{-k}} \mathrm{HK_2} \qquad c_{-k} = 0.0077 s^{-1}$$

$$\mathrm{RR} \xrightarrow{c_a} \mathrm{RRp} \qquad c_a = 8.67e-05 s^{-1}$$

$$\mathrm{RRp} \xrightarrow{c_g} \mathrm{RR} \qquad c_g = 0.00026 s^{-1}$$

$$2\mathrm{RRp} \xrightarrow{c_2} \mathrm{RRp_2} \qquad c_2 = 0.00166 s^{-1}$$

$$\mathrm{RRp_2} \xrightarrow{c_{-2}} 2\mathrm{RRp} \qquad c_{-2} = 0.02158 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_{-3a}} \mathrm{HK_2} + \mathrm{RR} \qquad c_{-3a} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p} + \mathrm{RR} \xrightarrow{c_{3b}} \mathrm{HK_2p{:}RR} \qquad c_{3b} = 0.00166 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RRp} \xrightarrow{c_{3c}} \mathrm{HK_2{:}RRp} \qquad c_{3c} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_{-3c}} \mathrm{HK_2} + \mathrm{RRp} \qquad c_{-3c} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_t} \mathrm{HK_2{:}RRp} \qquad c_t = 0.0279 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_p} \mathrm{HK_2{:}RR} \qquad c_p = 0.0087 s^{-1}$$

Model 2

$$2\mathrm{HK} \xrightarrow{c_1} \mathrm{HK_2} \qquad c_1 = 0.0125 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_{-1}} 2\mathrm{HK} \qquad c_{-1} = 0.005 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_k} \mathrm{HK_2p} \qquad c_k = 0.025 s^{-1}$$

$$\mathrm{HK_2p} \xrightarrow{c_{-k}} \mathrm{HK_2} \qquad c_{-k} = 0.0077 s^{-1}$$

$$\mathrm{RR} \xrightarrow{c_a} \mathrm{RRp} \qquad c_a = 8.67e-05 s^{-1}$$

$$\mathrm{RRp} \xrightarrow{c_g} \mathrm{RR} \qquad c_g = 0.00026 s^{-1}$$

$$2\mathrm{RRp} \xrightarrow{c_2} \mathrm{RRp_2} \qquad c_2 = 0.00166 s^{-1}$$

$$\mathrm{RRp_2} \xrightarrow{c_{-2}} 2\mathrm{RRp} \qquad c_{-2} = 0.02158 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RR} \xrightarrow{c_{3a}} \mathrm{HK_2{:}RR} \qquad c_{3a} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_{-3a}} \mathrm{HK_2} + \mathrm{RR} \qquad c_{-3a} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p} + \mathrm{RR} \xrightarrow{c_{3b}} \mathrm{HK_2p{:}RR} \qquad c_{3b} = 0.00166 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RRp} \xrightarrow{c_{3c}} \mathrm{HK_2{:}RRp} \qquad c_{3c} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_{-3c}} \mathrm{HK_2} + \mathrm{RRp} \qquad c_{-3c} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_t} \mathrm{HK_2{:}RRp} \qquad c_t = 0.0279 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_p} \mathrm{HK_2{:}RR} \qquad c_p = 0.0087 s^{-1}$$

Model 3

$$2\mathrm{HK} \xrightarrow{c_1} \mathrm{HK_2} \qquad c_1 = 0.0125 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_{-1}} 2\mathrm{HK} \qquad c_{-1} = 0.005 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_k} \mathrm{HK_2p} \qquad c_k = 0.025 s^{-1}$$

$$\mathrm{HK_2p} \xrightarrow{c_{-k}} \mathrm{HK_2} \qquad c_{-k} = 0.0077 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_l} \mathrm{HK_2p{:}RR} \qquad c_l = 0.025 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_{-l}} \mathrm{HK_2{:}RR} \qquad c_{-l} = 0.0077 s^{-1}$$

$$\mathrm{RR} \xrightarrow{c_a} \mathrm{RRp} \qquad c_a = 8.67e-05 s^{-1}$$

$$\mathrm{RRp} \xrightarrow{c_g} \mathrm{RR} \qquad c_g = 0.00026 s^{-1}$$

$$2\mathrm{RRp} \xrightarrow{c_2} \mathrm{RRp_2} \qquad c_2 = 0.00166 s^{-1}$$

$$\mathrm{RRp_2} \xrightarrow{c_{-2}} 2\mathrm{RRp} \qquad c_{-2} = 0.02158 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RR} \xrightarrow{c_{3a}} \mathrm{HK_2{:}RR} \qquad c_{3a} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_{-3a}} \mathrm{HK_2} + \mathrm{RR} \qquad c_{-3a} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p} + \mathrm{RR} \xrightarrow{c_{3b}} \mathrm{HK_2p{:}RR} \qquad c_{3b} = 0.00166 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RRp} \xrightarrow{c_{3c}} \mathrm{HK_2{:}RRp} \qquad c_{3c} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_{-3c}} \mathrm{HK_2} + \mathrm{RRp} \qquad c_{-3c} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_t} \mathrm{HK_2{:}RRp} \qquad c_t = 0.0279 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_p} \mathrm{HK_2{:}RR} \qquad c_p = 0.0087 s^{-1}$$

Model 4

$$2\mathrm{HK} \xrightarrow{c_1} \mathrm{HK_2} \qquad c_1 = 0.0125 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_{-1}} 2\mathrm{HK} \qquad c_{-1} = 0.005 s^{-1}$$

$$\mathrm{HK_2} \xrightarrow{c_k} \mathrm{HK_2p} \qquad c_k = 0.025 s^{-1}$$

$$\mathrm{HK_2p} \xrightarrow{c_{-k}} \mathrm{HK_2} \qquad c_{-k} = 0.0077 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_l} \mathrm{HK_2p{:}RR} \qquad c_l = 0.025 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_{-l}} \mathrm{HK_2{:}RR} \qquad c_{-l} = 0.0077 s^{-1}$$

$$\mathrm{RR} \xrightarrow{c_a} \mathrm{RRp} \qquad c_a = 8.67e-05 s^{-1}$$

$$\mathrm{RRp} \xrightarrow{c_g} \mathrm{RR} \qquad c_g = 0.00026 s^{-1}$$

$$2\mathrm{RRp} \xrightarrow{c_2} \mathrm{RRp_2} \qquad c_2 = 0.00166 s^{-1}$$

$$\mathrm{RRp_2} \xrightarrow{c_{-2}} 2\mathrm{RRp} \qquad c_{-2} = 0.02158 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RR} \xrightarrow{c_{3a}} \mathrm{HK_2{:}RR} \qquad c_{3a} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RR} \xrightarrow{c_{-3a}} \mathrm{HK_2} + \mathrm{RR} \qquad c_{-3a} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p} + \mathrm{RR} \xrightarrow{c_{3b}} \mathrm{HK_2p{:}RR} \qquad c_{3b} = 0.00166 s^{-1}$$

$$\mathrm{HK_2} + \mathrm{RRp} \xrightarrow{c_{3c}} \mathrm{HK_2{:}RRp} \qquad c_{3c} = 0.00166 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_{-3c}} \mathrm{HK_2} + \mathrm{RRp} \qquad c_{-3c} = 0.00498 s^{-1}$$

$$\mathrm{HK_2p{:}RR} \xrightarrow{c_t} \mathrm{HK_2{:}RRp} \qquad c_t = 0.0279 s^{-1}$$

$$\mathrm{HK_2{:}RRp} \xrightarrow{c_p} \mathrm{HK_2{:}RR} \qquad c_p = 0.0087 s^{-1}$$

$$\text{promoter-off} \xrightarrow{k_{\mathrm{on}}} \text{promoter-on} \qquad k_{\mathrm{on}} = 0.0431 s^{-1}, \eta_{k_{\mathrm{on}}} = 0.0672$$

$$\text{promoter-on} \xrightarrow{k_{\mathrm{off}}} \text{promoter-off} \qquad k_{\mathrm{off}} = 0.4479 s^{-1}, \eta_{k_{\mathrm{off}}} = 0.0428$$

$$\mathrm{RRp_2} + \text{promoter-off} \xrightarrow{c_{\mathrm{on}}} \mathrm{RRp_2}\text{-promoter-on} \qquad c_{\mathrm{on}} = 0.000332 s^{-1}$$

$$\mathrm{RRp_2} + \text{promoter-on} \xrightarrow{c_{\mathrm{on}}} \mathrm{RRp_2}\text{-promoter-on} \qquad c_{\mathrm{on}} = 0.000332 s^{-1}$$

$$\mathrm{RRp_2}\text{-promoter-on} \xrightarrow{c_{\mathrm{off}}} \mathrm{RRp_2} + \text{promoter-off} \qquad c_{\mathrm{off}} = 0.006 s^{-1}$$

$$\mathrm{RRp_2}\text{-promoter-on} \xrightarrow{k_1} \mathrm{mRNA} + \mathrm{RRp_2}\text{-promoter-on} \qquad k_1 = 0.434 s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{promoter-on} \xrightarrow{k_1} \text{promoter-on + mRNA} \qquad k_1 = 0.434s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{promoter-off} \xrightarrow{k_0 k_1} \text{promoter-off + mRNA} \qquad k_0 k_1 = 0.025649s^{-1}, \eta_{k_1} = 0.0516$$

$$\text{mRNA} \xrightarrow{d_1} \emptyset \qquad d_1 = 0.396s^{-1}, \eta_{d_1} = 0.0441$$

$$\text{mRNA} \xrightarrow{k_{2\text{HK}}} \text{HK + mRNA} \qquad k_{2\text{HK}} = 0.87416s^{-1}, \eta_{k_2} = 0.0369$$

$$\text{mRNA} \xrightarrow{k_{2\text{RR}}} \text{RR + mRNA} \qquad k_{2\text{RR}} = 8.7416s^{-1}, \eta_{k_2} = 0.0369$$

$$\text{HK} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2 \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{p} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RRp} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{RRp}_2 \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{:RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{p:RR} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{HK}_2\text{:RRp} \xrightarrow{d_2} \emptyset \qquad d_2 = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$

$$\text{promoter(X)-off} \xrightarrow{k_{\text{onX}}} \text{promoter(X)-on}$$

$$k_{\text{onX}} = 0.0431s^{-1}, \eta_{k_{\text{on}}} = 0.0672$$

$$\text{promoter(X)-on} \xrightarrow{k_{\text{offX}}} \text{promoter(X)-off}$$

$$k_{\text{offX}} = 0.4479s^{-1}, \eta_{k_{\text{off}}} = 0.0428$$

$$\text{RRp}_2 + \text{promoter(X)-off} \xrightarrow{c_{\text{onX}}} \text{RRp}_2\text{promoter(X)-on} \qquad c_{\text{onX}} = 0.000332s^{-1}$$

$$\text{RRp}_2 + \text{promoter(X)-on} \xrightarrow{c_{\text{offX}}} \text{RRp}_2\text{promoter(X)-on} \qquad c_{\text{offX}} = 0.000332s^{-1}$$

$$\text{RRp}_2\text{promoter(X)-on} \xrightarrow{c_{\text{offX}}} \text{RRp}_2 + \text{promoter(X)-off} \qquad c_{\text{offX}} = 0.006s^{-1}$$

$$\mathrm{RRp_2promoter(X)\text{-}on} \xrightarrow{k_{1X}} \mathrm{mRNA(X)} + \mathrm{RRp_2promoter(X)\text{-}on}$$
$$k_{1X} = 0.434s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{promoter(X)\text{-}on} \xrightarrow{k_{1X}} \mathrm{promoter(X)\text{-}on} + \mathrm{mRNA(X)}$$
$$k_{1X} = 0.434s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{promoter(X)\text{-}off} \xrightarrow{k_{0X}k_{1X}} \mathrm{promoter(X)\text{-}off} + \mathrm{mRNA(X)}$$
$$k_{0X}k_{1X} = 0.025649s^{-1}, \eta_{k_1} = 0.0516$$
$$\mathrm{mRNA(X)} \xrightarrow{d_{1X}} \emptyset \qquad d_{1X} = 0.396s^{-1}, \eta_{d_1} = 0.0441$$
$$\mathrm{mRNA(X)} \xrightarrow{k_{2X}} \mathrm{mRNA(X)} + \mathrm{protein(X)} \qquad k_{2X} = 8.7416s^{-1}$$
$$\mathrm{protein(X)} \xrightarrow{d_{2X}} \emptyset \qquad d_{2X} = 0.0035s^{-1}, \eta_{d_2} = 0.1664$$