

Contents

1	Bayesian model fitting applied to flow cytometry data	3
1.1	Introduction	3
1.2	Contributions to this Chapter	3
1.3	Flow cytometry and model fitting	3
1.4	ABC-Flow algorithm development	4
1.4.1	Intensity calculation	7
1.4.2	Distance Calculations	7
1.4.2.1	Kernel distance	8
1.4.2.2	Kolmogorov-Smirnov distance	12
1.4.2.3	Wald-Wolfowitz distance	13
1.5	ABC-Flow model fitting to simulated data	17
1.6	Toggle switch data collection	22
1.6.1	Circuit overview	22
1.6.2	Methods	24
1.6.2.1	<i>Escherichia coli</i> culturing conditions	24
1.6.2.2	Glycerol stock preparation	24
1.6.2.3	Revival	24
1.6.2.4	Plasmid construction	25
1.6.2.5	Polymerase Chain Reaction	25
1.6.2.6	Digestion	26
1.6.2.7	Agarose gel electrophoresis	26
1.6.2.8	Ligation	27
1.6.2.9	Transformation	27
1.6.2.10	Colony PCR	28
1.6.2.11	Sequencing	29
1.6.2.12	Inducers	29
1.6.2.13	Growth rate measurement	29

2 CONTENTS

1.6.2.14	Flow cytometry	30
1.6.2.15	Concentration assays	30
1.6.2.16	Time course assays	32
1.6.3	Results	32
1.6.3.1	pKDL071 plasmid alteration	32
1.6.3.2	Control plasmids construction	34
1.6.3.3	Growth rate investigation	35
1.6.3.4	Toggle switch concentration assays	36
1.6.3.5	Toggle switch time course assay	40
1.7	ABC-Flow used on experimental data	43
1.7.1	Toggle switch model used in ABC-Flow	43
1.7.2	ATc induction	46
1.7.3	IPTG induction	47
1.8	Discussion	48
1.9	Summary	49
Bibliography	51

1 Bayesian model fitting applied to flow cytometry data

1.1 Introduction

In this chapter I aim to fit the toggle switch model to experimental data. This chapter is organised as follows: In the first section I provide an overview of the framework developed to fit models to flow cytometry data (ABC-Flow). In the subsequent section I test ABC-Flow on simulated flow cytometry data. Next I use flow cytometry to study the toggle switch experimentally and examine the concentrations of the inducers and the time needed to flip the switch. Finally, I use ABC-Flow to fit a computational model to the experimental data acquired.

1.2 Contributions to this Chapter

The R code used to pre-process the flow cytometry obtained was provided by Alex J Fedorec. The initial development of ABC-Flow was carried out by Chris P Barnes.

1.3 Flow cytometry and model fitting

Computational modelling is well known to aid the understanding of complex systems by fitting experimental data and providing further insights and testable predictions. Experimental data is used to fit the model parameters and then the model can provide further understanding of the system and aid in the design of further experiments. Flow cytometry is used in synthetic biology for BioBrick characterisation (Kelly et al. 2009), enzyme screening (Choi et al. 2014) and industrial bioprocesses (Díaz et al. 2010) among others.

4 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

Flow cytometry data presents a challenge to computational modelling as the fluorescence intensity per cell is measured rather than number of proteins. The problem with measuring fluorescence intensity is that it is a relative and not an absolute measurement like the number or concentration of proteins in a system, which would increase the predictive power of computational models (Bower, McClintock, & Fong 2010; Cooling et al. 2010), but this type of biological data cannot be directly measured (Kelwick et al. 2014). The fluorescence intensity values can vary between experiments due to instrument settings so they can only be used in relative terms within the same experiment. Standardization of experimental methods in flow cytometry has aided the effort to reduce variability due to experimental setup (Kelly et al. 2009), but the successful conversion of fluorescence intensity to an absolute measurement of protein $\text{cell}^{-1} \text{ s}^{-1}$ has yet to be made successfully.

Another approach to the problem is converting the model output of $\text{GFP cell}^{-1} \text{ s}^{-1}$ to relative fluorescence intensity. This approach was first developed by Lillacci & Khammash (2013). The converted model output can then be compared to the data output from the flow cytometer. The fluorescent intensity measurements acquired via flow cytometry are treated as a sample from distribution of the fluorescence present in the cell (Lillacci & Khammash 2013). This means that the flow cytometry fluorescence distribution at each time point can be compared to the model fluorescence distribution. Here I expand the method developed by Lillacci & Khammash (2013) in order to be able to apply it to flow cytometry data including two fluorescent proteins simultaneously. This new framework, ABC-Flow, can be used to fit stochastic models to flow cytometry data involving two species, like the genetic toggle switch.

1.4 ABC-Flow algorithm development

The algorithm of ABC-Flow is based on the same ABC algorithm as ABC-SysBio and Stability Finder described in Sections (XXX), adapted to be used for flow cytometry data. The algorithm of ABC-Flow is outlined in Algorithm 1. The modified modules of the ABC algorithm are outlined in the sections that follow.

The user provides an SBML model file and an input file to specify the information needed to run ABC-Flow, such as the epsilon schedule and the priors to the parameters. The user must also provide a data file containing the flow cytometry data to which the model will be fitted. The data files used here were generated from .fcs files, which is the standard output of flow cytometers, using the R bioconductor

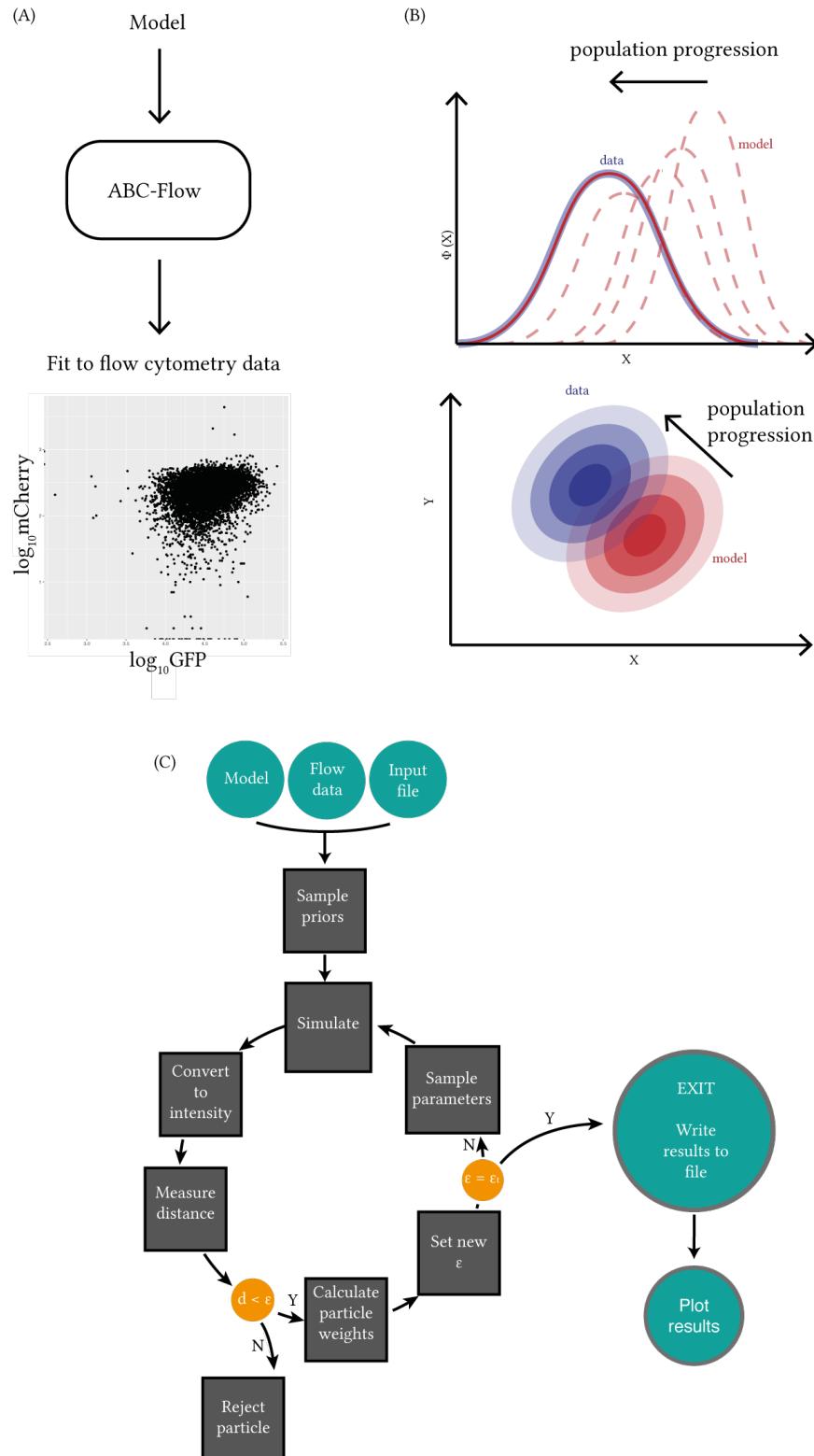


Figure 1.1 Overview of ABC-Flow. (A) ABC-Flow is used to fit models to experimental flow cytometry data. (B) The algorithm can be applied to 1D and 2D flow data. (C) ABC-Flow uses Approximate Bayesian Computation.

6 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

packages flowCore (Ellis et al. 2016b). All models are simulated stochastically using the Gillespie algorithm (Gillespie 1977). ABC-Flow simulations are implemented on GPUs. ABC-Flow is available as a Python package, and can be downloaded from <https://github.com/ucl-cssb/ABC-Flow.git>.

Algorithm 1 ABC-Flow

```

1: Initialise  $\epsilon$ 
2: population p  $\leftarrow$  1
3: if p = 1 then
4:     Sample particles ( $\theta$ ) from priors
5: else
6:     Sample particles from previous population
7:     Perturb each particle by  $\pm$  half the range of the previous population (j) to
       obtain new perturbed population (i).
8: end if
9: Simulate model using the Gillespie algorithm.
10: Convert signal to intensity:
11: for each particle do
12:     for each beta do
13:         for each timepoint do
14:             for each fluorescent protein do
15:                 Intensity =  $N\left(\text{signal} \times \mu, \sqrt{(\text{signal} \times \sigma^2)}\right)$ 
16:             end for
17:         end for
18:     end for
19: end for
20: Measure distance to data
21: Reject particles if  $d > \epsilon$ .
22: Calculate weight for each accepted  $\theta$ 
23:  $w_t^{(i)} = \begin{cases} 1, & \text{if } p = 0 \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N w_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}, \theta_t^{(i)})}, & \text{if } p \geq 0. \end{cases}$ 
24: Normalise weights
25: Repeat steps 3 - 15 until  $\epsilon \leq \epsilon_T$ 

```

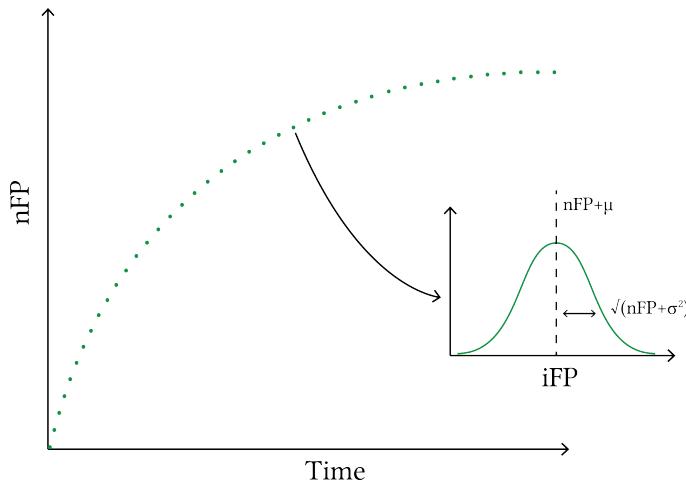


Figure 1.2 Converting the number of fluorescent proteins to the intensity (iFP) is done by drawing from a normal distribution, as shown in Equation 1.1.

1.4.1 Intensity calculation

The units of the result of the stochastic simulations is in the form of number of fluorescent proteins. On the other hand, flow cytometry data units are in the form of fluorescence intensity. For ABC-Flow, the simulation results are converted to intensity in order to be able to compare the data to the simulations. In order to do this two additional parameters are defined, intensity μ and intensity σ , for each fluorescent protein used. To convert the number of fluorescent proteins to intensity, random samples are drawn from a normal distribution:

$$X \sim N(nFP \times \mu, \sqrt{(nFP \times \sigma^2)}), \quad (1.1)$$

where nFP is the number of fluorescent proteins.

These parameters are fitted to the data along with the model parameters.

1.4.2 Distance Calculations

In order to compare the flow cytometry data to the model generated data, I had to develop a distance measure. This distance measure should be able to determine whether two datasets are sufficiently close to each other to be able to assume that they have been drawn from the same distribution. The measure should also give an estimate of how different the two data sets are, and thus get increasingly

8 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

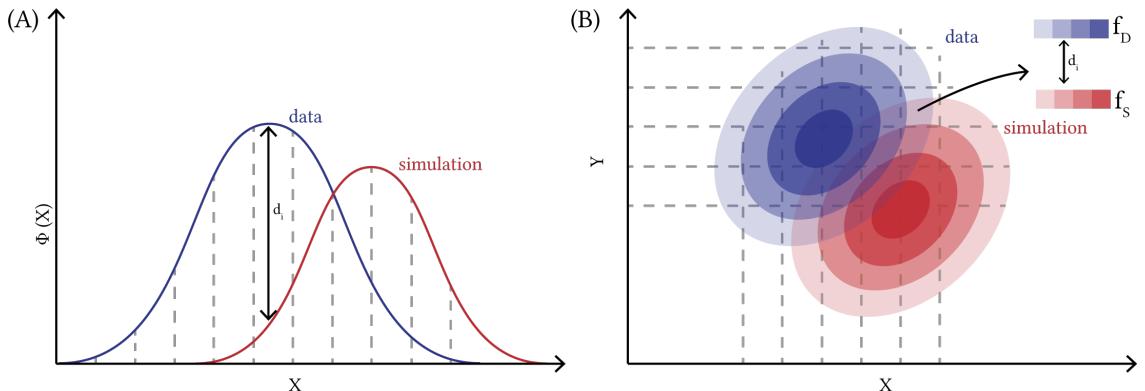


Figure 1.3 Calculating the distance between two distributions in (A) 1D and (B) 2D.

larger as two data sets are drawn from increasingly different distributions. Finally the distance measure should be applicable to one and two dimensional distributions, and be comparable between the two.

1.4.2.1 Kernel distance

In order to measure the distance between the flow cytometry data and the fitted model, the algorithm outlined in Algorithm 2 was developed. The algorithm consists of defining a grid from the minimum to the maximum value of the data. A gaussian kernel was then fit to the flow and simulated data. The distance between the two kernels is given by:

$$d = \sum_{i=x_{min}}^{x_{max}} (fD_i - fS_i)^2,$$

where fD_i is the kernel of the flow data at each value of x and fS_i the kernel of the simulated data. An illustration of the distance calculation is shown in Figure 1.3.

Algorithm 2 Distance calculation

- 1: $\text{Grid} \leftarrow \text{min}(\text{data}): \text{max}(\text{data}): \text{ngrid}$
 - 2: $kD = \text{kernel density estimation}(\text{data})$
 - 3: $kS = \text{kernel density estimation}(\text{simulations})$
 - 4: $fD = kD(\text{xx})$
 - 5: $fS = kS(\text{xx})$
 - 6: $\epsilon = \sum((fD - fS)^2)$
-

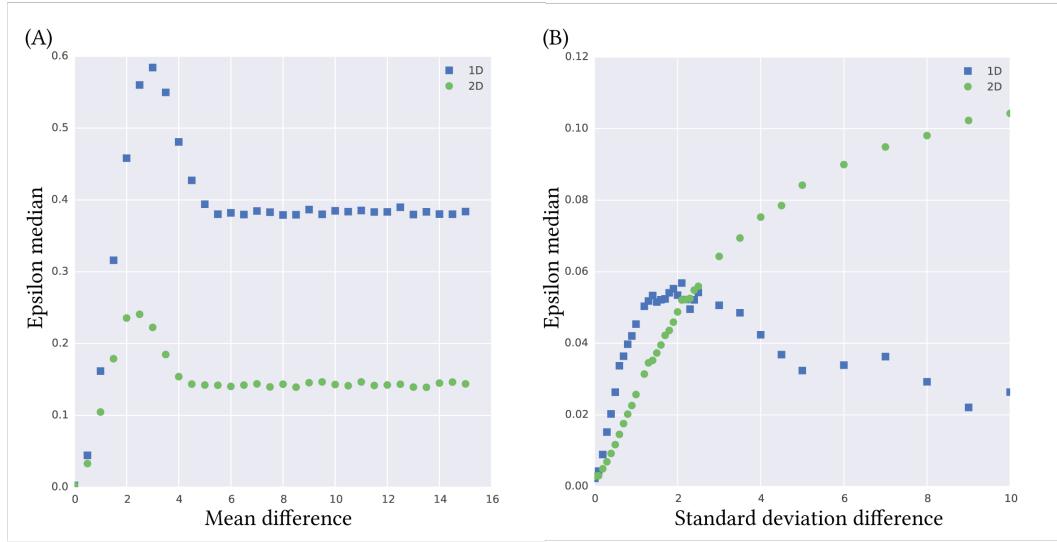


Figure 1.4 (A) The range by which epsilon varies as the difference between the mean of the distributions increases. (B) The median of the epsilon distributions varies by a small amount with increasing difference in the standard deviation of the distributions.

Prior to incorporating this distance calculation in ABC-FLow, it was tested to determine whether it is an appropriate distance to use when comparing distributions. This was done by drawing samples from two uniform distributions with varying mean and standard deviation. Algorithm 2 was then used to calculate the distance between the different distributions.

First, Algorithm 2 was tested by drawing samples from two distributions with an increasingly different mean. This is done to determine the dynamical range of the distance calculation.

From Figure 1.4 we see that the epsilon value does not increase linearly with increasing mean difference of the two distributions. As the difference between the means increases, the epsilon value reaches a peak when the difference is at 3. From that point, as the mean difference increases, epsilon values decrease until they reach a plateau at epsilon = 0.38 in the 1D case and epsilon = 0.14 in the 2D case. Next, I test the distance calculation by comparing bimodal distributions. Two bimodal distributions are generated with increasingly different mean, in 1D and 2D.

Similar to the normal distribution, for the bimodal distributions shown in Figure 1.5 we find that the epsilon values do not increase linearly. There are two peaks

10 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

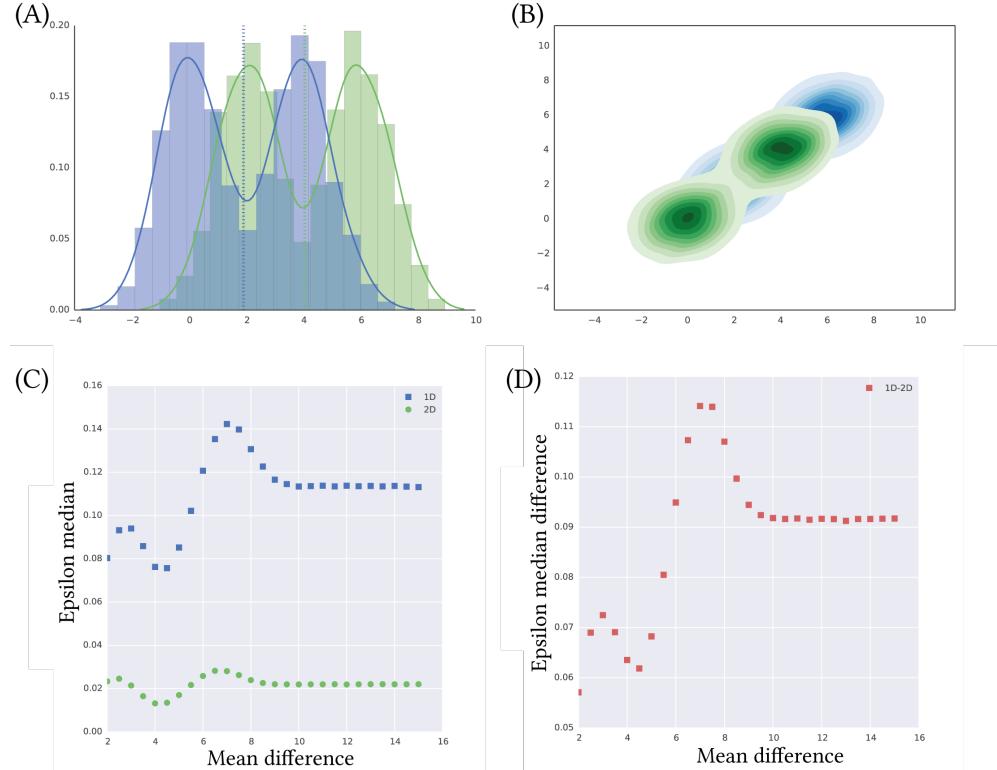


Figure 1.5 Comparing the 1D and 2D distances between bimodal distributions. (A) and (B) show samples of the bimodal distributions compared in 1D and 2D respectively with a mean difference of 4 between simulations and data. (C) The range by which epsilon median and variance varies as the difference between the mean of the distributions increases. (D) The difference between the epsilons calculated in 1D and 2D is not constant.

in the epsilon distribution, one at mean difference = 3 and one at mean difference=6. The epsilon values then decline until they reach a plateau. The epsilon values do not have a large range of values, for neither the 1D or 2D cases. We also find that the difference in the epsilon values between the 1D and 2D cases is not constant.

Finally, I study how these distance functions perform when comparing a bimodal with a normal distribution. A bimodal distribution is generated and a series of normal distributions with increasing mean, in 1D and 2D. From Figure 1.6 we find that epsilon is the lowest when the mean of the normal distribution corresponds to the μ of one of the two peaks in the bimodal distribution and the highest when there is no overlap between the distributions.

From Figures 1.4-1.6 I conclude that Algorithm 2 is not a good measure for distance to be used in ABC-Flow. If Algorithm 2 was used in order to minimize the

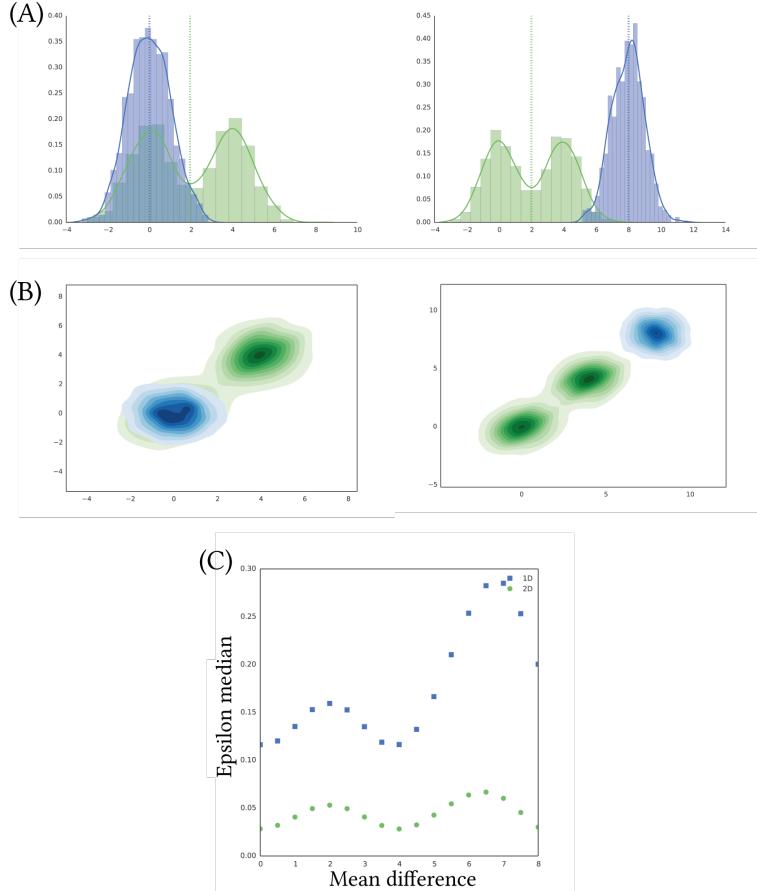


Figure 1.6 Comparing a multimodal to a normal distribution, in 1D and 2D. (A, B) The mean of the normal distribution is varied from equal to the mean of the first peak of the bimodal distribution to beyond the range of the bimodal distribution. (C) Epsilon median and variance are at the lowest when the mean of the normal distribution is equal to the mean of one of the peaks of the bimodal distribution.

distance between two distributions that start off with very different means, the distance between the two distributions will not be sufficiently minimized. This stems from the fact that ABC-FLow works by iteratively making the accepted epsilon smaller. As can be seen in Figure 1.4, if the two distributions have a large difference in the means, (>6) it would not be possible to overcome the peak that is created when the mean difference is at 3. Epsilon values increase before the decrease again, which will be a problem in ABC-Flow. Therefore a different distance calculation was developed.

1.4.2.2 Kolmogorov-Smirnov distance

In order to avoid the problems that arose from the distance calculation described in Section 1.4.2.1 I implemented a different distance calculation for ABC-Flow. I used a Python implementation of the Kolmogorov-Smirnov two sample test for the 1D case (Kolmogorov 1933). The Kolmogorov-Smirnov (KS) test is a non-parametric statistic test that determines whether two data sets were drawn from the same underlying distributions. The KS distance between two distributions is equal to the largest distance between the empirical distribution functions of the two samples, as shown in Equation 1.2.

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)| \quad (1.2)$$

For the 2D case the distance was calculated by using the 2D Kolmogorov-Smirnov two sample test. The algorithm was developed by Fasano & Franceschini (1987) and the Python implementation developed by Major (2016).

This distance calculation was tested to determine whether it is an appropriate distance function to use in ABC-Flow. Two datasets were drawn from normal distributions with increasingly different means. The KS test was then used to calculate the distance between the data sets. This was carried out in 1D and 2D. The results are shown in Figure 1.7.

The epsilons of the 1D Kolmogorov-Smirnov distance calculation increase with increasing mean difference until it reaches a plateau when the two distributions are very different. This makes it an ideal distance calculation to be used in ABC-Flow. As the epsilon threshold is lowered at each iteration the difference between the two data sets decreases. Therefore the 1D KS statistic was used in ABC-Flow.

The multidimensional Kolmogorov-Smirnov test presents a challenge, as there is no unique way to order the data points to calculate the largest distance. There are $2^d - 1$ ways of ordering the data points and defining a cumulative distribution function, where d is the number of dimensions (Lopes, Reid, & Hobson 2007). This has affected the results of the computed distance seen in Figure 1.7. The variability in the calculation of the distance between data sets originating from distributions with known distance is large relative to the range of values the calculation can take. This was further confirmed when testing this distance on simulated data, where no parameter identifiability was observed (data not shown).

To alleviate the above shortcomings of the multi-dimensional generalisation of the Kolmogorov-Smirnov test, a different distance calculation was used for the 2D

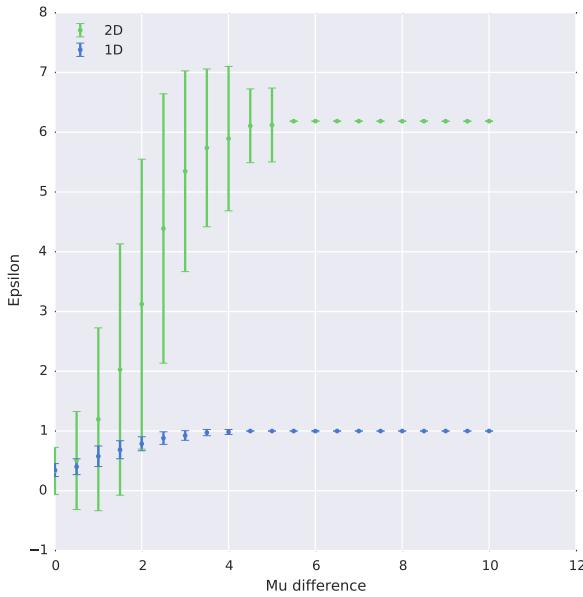


Figure 1.7 The Kolmogorov-Smirnov distance function was tested in 1D (blue) and 2D (green). Two data sets were generated with increasing mean difference, and the Kolmogorov-Smirnov two-sample test was applied to compute the distance between the two.

case. The Kolmogorov-Smirnov test for the 1D case was used in ABC-Flow as it performed well in the testing shown here.

1.4.2.3 Wald-Wolfowitz distance

For the 2D case the distance was calculated by using the multivariate Wald-Wolfowitz test (Friedman & Rafsky 1979). This is a generalisation of the Wald-Wolfowitz test proposed by (Wald & Wolfowitz 1940), a non-parametric test to determine whether two data sets were drawn from the same distribution. This test works by computing the minimum spanning tree of the pooled samples. Any edge whose nodes originated from different samples are removed, and the number of *runs* (R) is then defined by the number of disjointed subtrees (Friedman & Rafsky 1979). If the number of *runs* is small, then the null hypothesis that the two samples originated from the same distribution cannot be rejected. The quantity W for two samples, of length m and n , computed is given by:

$$W = \frac{R - 2\frac{mn}{N} - 1}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}}, \quad (1.3)$$

where $N = m + n$ and R is the number of *runs*. A Python implementation of the multivariate Wald-Wolfowitz test by Monaco (2014) was used here. This is a variation to the Wald-Wolfowitz test that can be efficiently applied to larger data sets. The Python code used is given in Appendix (XXX).

Here I test this distance calculation in a similar way as Section 1.4.2.1. First, the two data sets are drawn from increasingly different distributions, and the distance between them calculated. As shown in Figure 1.8D, the 2D distance is 0 when the difference between the μ from which the two datasets are drawn from the same distribution. The distance calculation reaches a plateau at $\epsilon = 140$ when the mean difference is 4 or larger. The 1D distance is also shown in Figure 1.8C in order to compare the two calculations, but the 1D distance was computed using the Kolmogorov-Smirnov distance described in Section 1.4.2.2.

To further study the distance calculation used in ABC-Flow, two normal distributions were simulated, with $\mu = 0$ and $\sigma = 1$ and distance between them calculated using the Kolmogorov-Smirnov test in the 1D case and the Wald-Wolfowitz test in the 2D case. Doing this multiple times, the expected variation in distance values for identical distributions can be calculated. This is the error that can be expected when measuring distance in ABC-Flow. As can be seen in Figure 1.9, the range of distance values obtained in the 1D case is small. For the 2D case, the distance values obtained vary more than in the 1D case, but it is still small relative to the range of values that the Wald-Wolfowitz test can take shown in Figure 1.8.

Using the Wald-Wolfowitz test the value of ϵ increases with increasing distance between the distributions with relatively small variability between repeats. Since the 2D Wald-Wolfowitz test performed well in the test carried out above, it was implemented in ABC-Flow as the distance function for the 2D calculations.

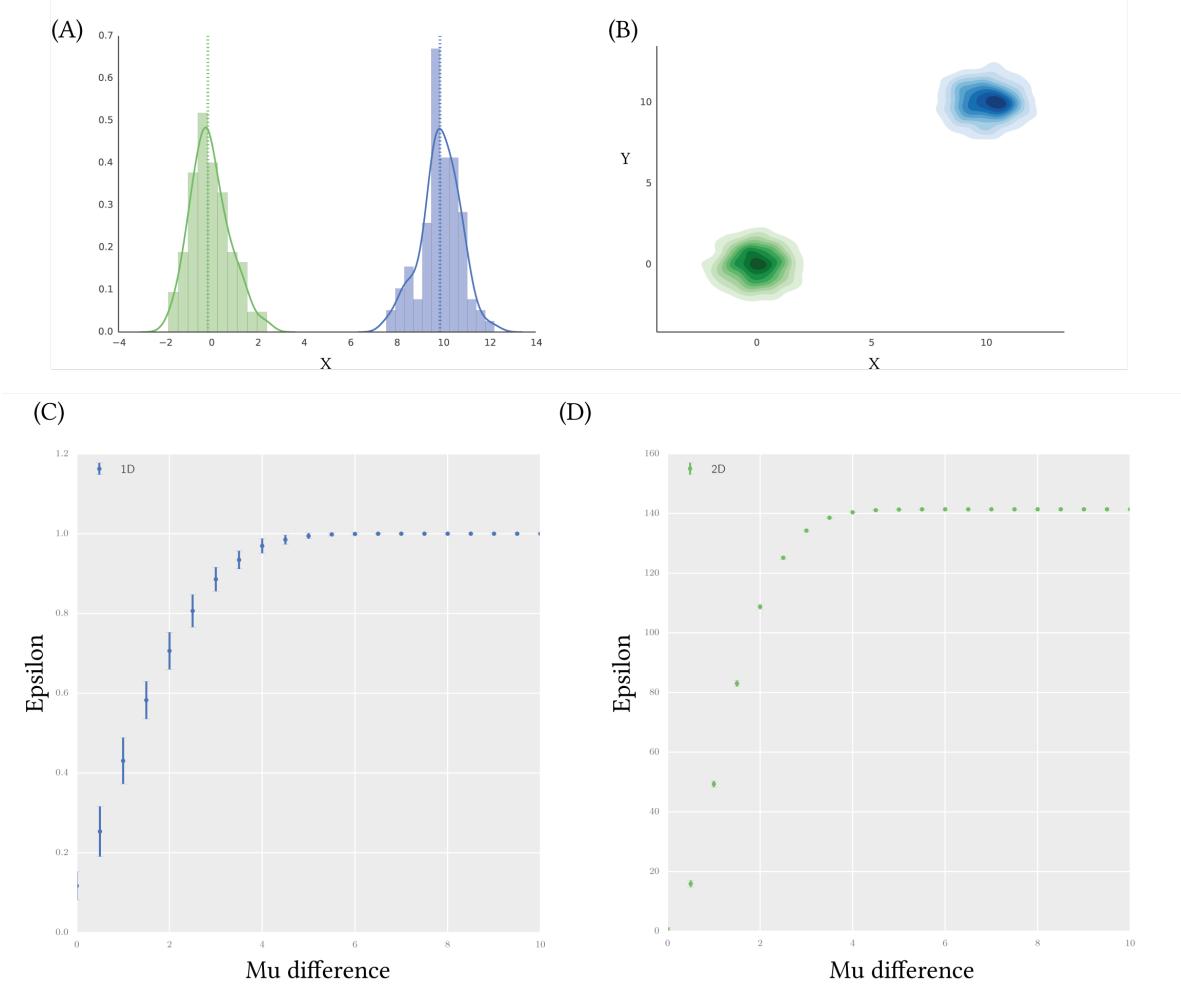


Figure 1.8 The distance calculation for data sets drawn from increasingly different distributions. Two examples are shown of distributions compared in (A) 1D and (B) 2D. (C) As the difference between the means of the two distributions increases, the distance calculation, epsilon, increases. In the 2D case (shown in green) epsilon plateaus at 2.3 and in the 1D case (shown in blue) epsilon plateaus at 1.

16 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

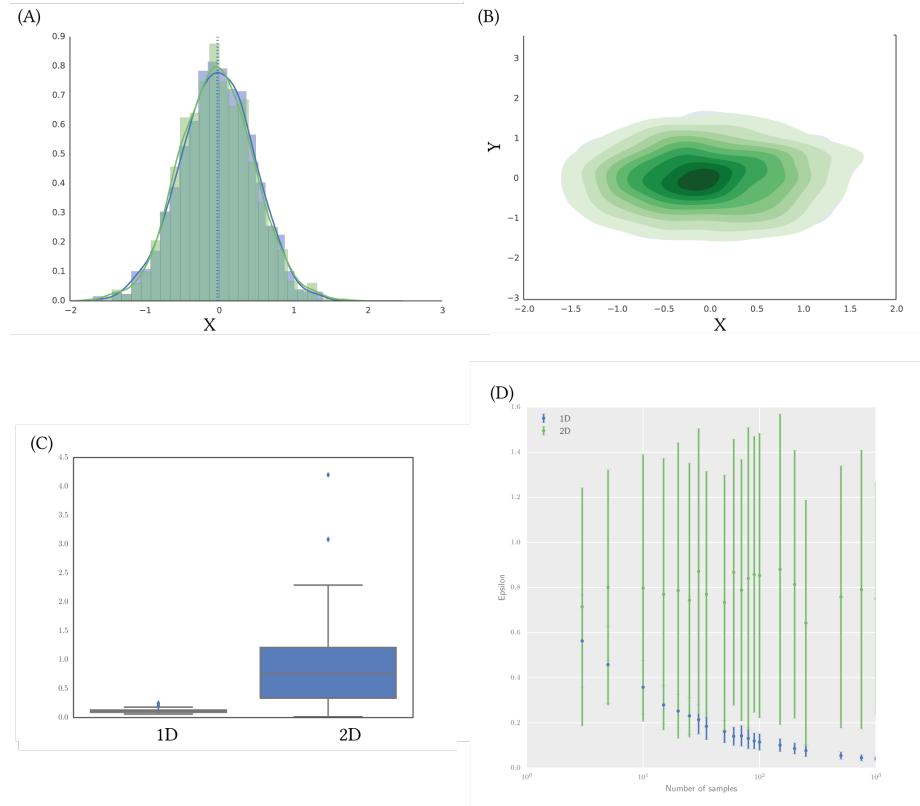


Figure 1.9 The distance between two data sets drawn from the same distribution are compared using the Kolmogorov-Smirnov two sample test. (A) in 1D and (B) in 2D. (C) The distance is calculated for 1000 data sets. A larger variation of values is found for the 2D distance calculation, but still small relative to the overall range of values. (D) As the number of samples in the datasets increase the distance calculation becomes more accurate in the 1D case. It has no effect on the 2D case.

1.5 ABC-Flow model fitting to simulated data

In this section I apply ABC-Flow to simulated data, where the parameter values used to produce the data are known. This analysis will serve as a verification test for ABC-Flow. The model used to produce the simulated data is an extension of the Gardner, Cantor, & Collins (2000) switch. The model consists of two mutually repressing transcription factors. The model used here has additional parameters allowing for gene expression to be leaky as well as include repression from an external stimulus.

In order to produce the simulated data set, an extension of the Gardner, Cantor, & Collins (2000) switch was simulated stochastically using the Gillespie algorithm (Gillespie 1977). The model used is defined by the following hazards:

$$h_1 = u \quad (1.4)$$

$$h_2 = \frac{p_1 \times p_3}{1 + p_3 + v^{p_2}} \quad (1.5)$$

$$h_3 = (1 + a) \times v \quad (1.6)$$

$$h_4 = \frac{p_4 \times p_6}{1 + p_6 + u^{p_5}}, \quad (1.7)$$

where u and v are the two proteins in the system, p_1 and p_4 represent the effective gene expression of u and v respectively, p_2 and p_5 represent the cooperativity of u and v respectively. p_3 and p_6 represent the leakiness of the promoters for each species. parameter α increases the degradation of one of the species, and simulates the addition of a repressor.

Using the time course data generated for one of the fluorescent proteins in the system, u , I use ABC-Flow to fit the model shown above, using priors centered around the parameter values used to produce the data, shown in Table 1.1. The resulting fit is shown in Figure 1.10A. In order to determine whether this is a good fit to the data, QQ plots are produced for each timepoint (Figure 1.10B). A QQ-plot is a plot where the quantiles of two distributions are plotted against each other. If the distributions are similar, the points will lie on the 45° line $x = y$ line (Wilk & Gnanadesikan 1968).

By examining the data and the fitted models shown in Figure 1.10, we see that at $\epsilon = 0.08$, there is a good fit of the model to the data using ABC-Flow. The model parameters, as well as the intensity parameters, have been fitted to simulated flow cytometry data. The model has been successfully fitted to the simulated data. This is highlighted in the QQ plots in Figure 1.10B, where the results lie in the $x=y$ line.

18 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

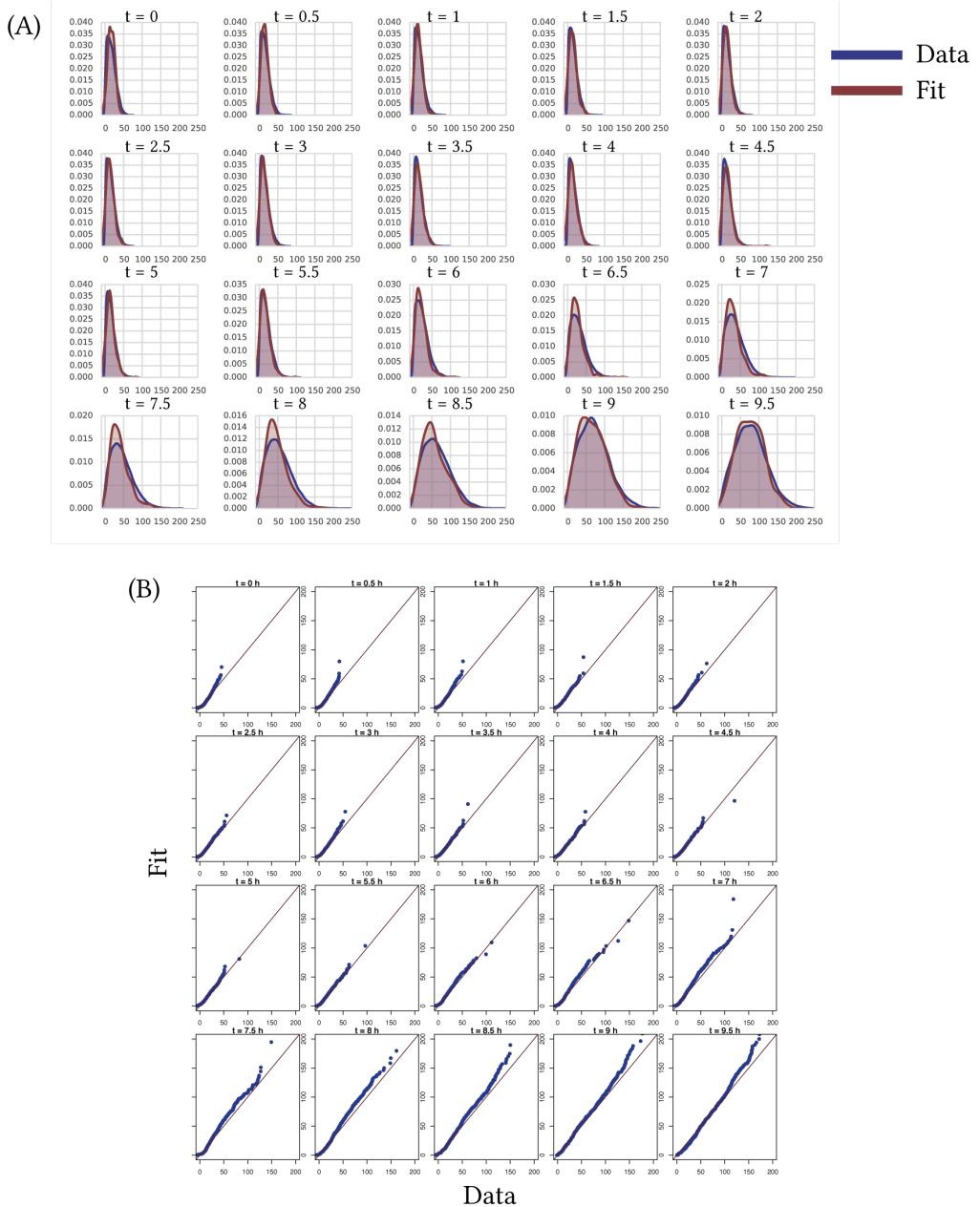


Figure 1.10 (A) 1D ABC-Flow fit (shown in blue) to data (shown in red) produced by simulating the same model. (B) QQ-plot of each time point fit. The quantile of the two distributions are plotted against each other. If the distributions are similar, the points would lie on the 45° line $x = y$, shown in red.

I further test ABC-Flow by using 2D data to fit two species of the model simultaneously. The same data was used as was used in the 1D case, but this time both species u and v were taken into account. This represents both sides of the switch model used.

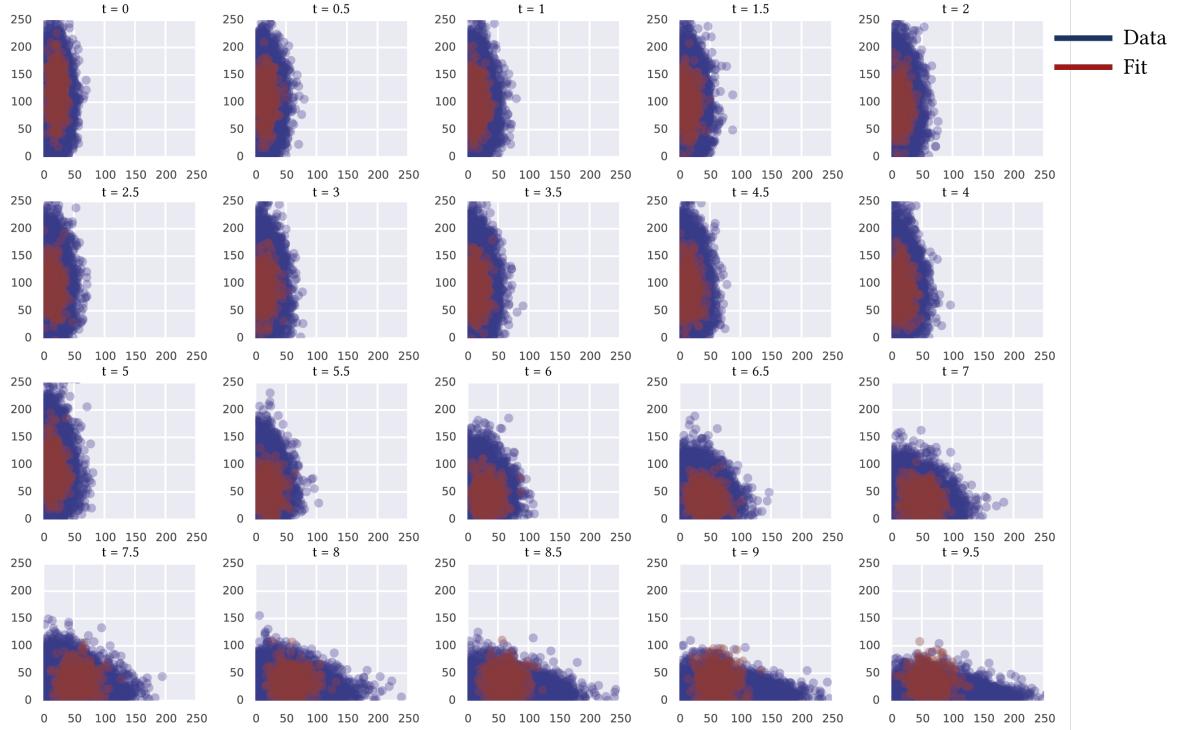


Figure 1.11 2D ABC-Flow fit (shown in blue) to data (shown in red) produced by simulating the same model.

The posterior distributions obtained from each fit are shown in Figure 1.12. We find similar posterior ranges for both the 1D and 2D fits. Both fits identified the parameters necessary to produce the simulated data. From the posteriors we find that the most constrained parameters to produce the switch behaviour in this model are parameters p_2 and p_5 , the parameters representing the cooperativity of the repressors. They both have to be equal to 2 to produce the observed behaviour in this model. We also find that α , the parameter representing the increased degradation of the repressing species due to the addition of an inducer, is tightly constrained. α is required to be small. Further, we find that μ and σ , the parameters representing the mean and standard deviation of the fluorescence intensity emitted by each fluorescent molecule to be tightly constrained.

These results demonstrate that ABC-Flow can successfully fit a computational

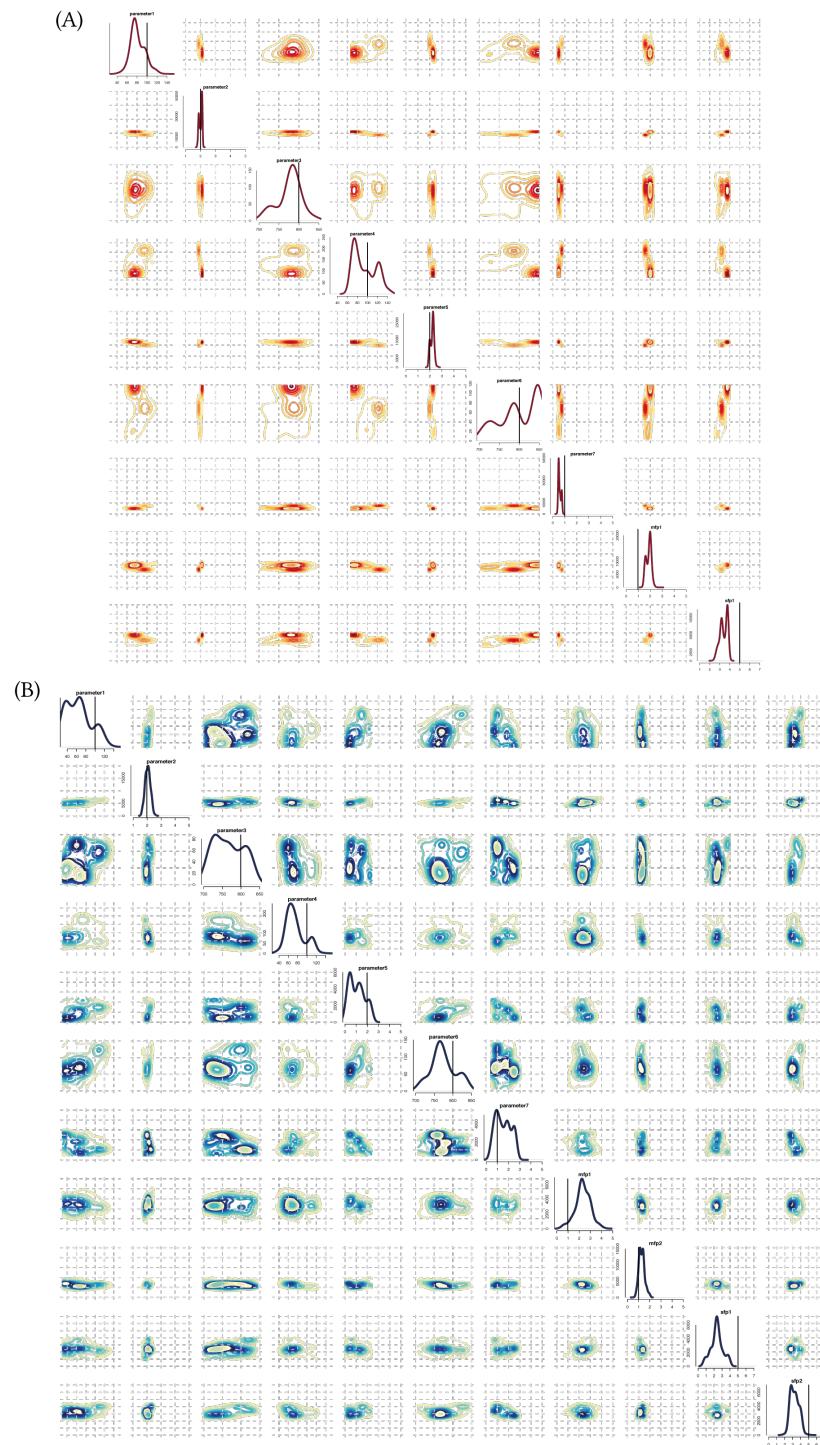


Figure 1.12 The posterior distributions of the 1D (red) 2D (blue) fits to simulated data.

Table 1.1 The priors used for the 1D and 2D ABC-FLow model fitting to simulated data

Parameters		
	1D	2D
p1	30 - 150	30 - 150
p2	1 - 5	1 - 5
p3	700 - 850	700 - 850
p4	30 - 150	30 - 150
p5	0 - 5	0 - 5
p6	700 - 850	700 - 850
p7	0 - 5	0 - 5
Species		
u	9 - 11	9 - 11
v	90 - 110	90 - 110
Intensity parameters		
mean fp1	0 - 5	0 - 5
mean fp2		0 - 5
sigma fp1	0 - 7	0 - 7
sigma fp2		0 - 7

model to flow cytometry data. It can identify the parameter values necessary to produce the observed behaviour. ABC-Flow can now be confidently applied to real flow cytometry data of the genetic toggle switch. This will allow me to fit a computational model to experimental data of the genetic toggle switch and uncover the parameters that are necessary to produce this behaviour. In the following Section I will outline the methods used to obtain the experimental data necessary and the results obtained.

1.6 Toggle switch data collection

In this section I collect experimental data on the genetic toggle switch. Using flow cytometry and the necessary inducers to flip the switch I study the switch flipping over time as well as over different inducer concentrations.

1.6.1 Circuit overview

The toggle switch plasmid I used here was provided by Litcofsky et al. (2012). All the switch components were contained in one plasmid, pKDL071. An overview of the plasmid is shown in Figure 1.13A and the sequence given in Appendix (XXX). The circuit consists of two promoters, P_{trc2} and P_{LtetO-1} (Lutz & Bujard 1997). P_{trc2} is a constitutive promoter, repressible by LacI. P_{LtetO-1} is also a constitutive promoter, repressible by TetR, as shown in Figure 1.13B. mCherry (Shaner et al. 2004) and GFP (Shimomura, Johnson, & Saiga 1962) are fluorescent proteins, that were added under the control of the same promoters as the repressors, and thus reflect the levels of TetR and LacI in the system. The plasmid contains kanamycin antibiotic resistance and is high copy (ColE1 origin of replication).

This system is capable of two states, GFP high and mCherry high. When IPTG is added to the system, it represses the repression of TetR and mCherry and thus the cells end up in the mCherry high state. When ATc is added to the system, it represses the repression of LacI and GFP and thus the cells end up in the GFP high state. If no inducer is added to the system it will randomly go to the GFP high or mCherry high states.

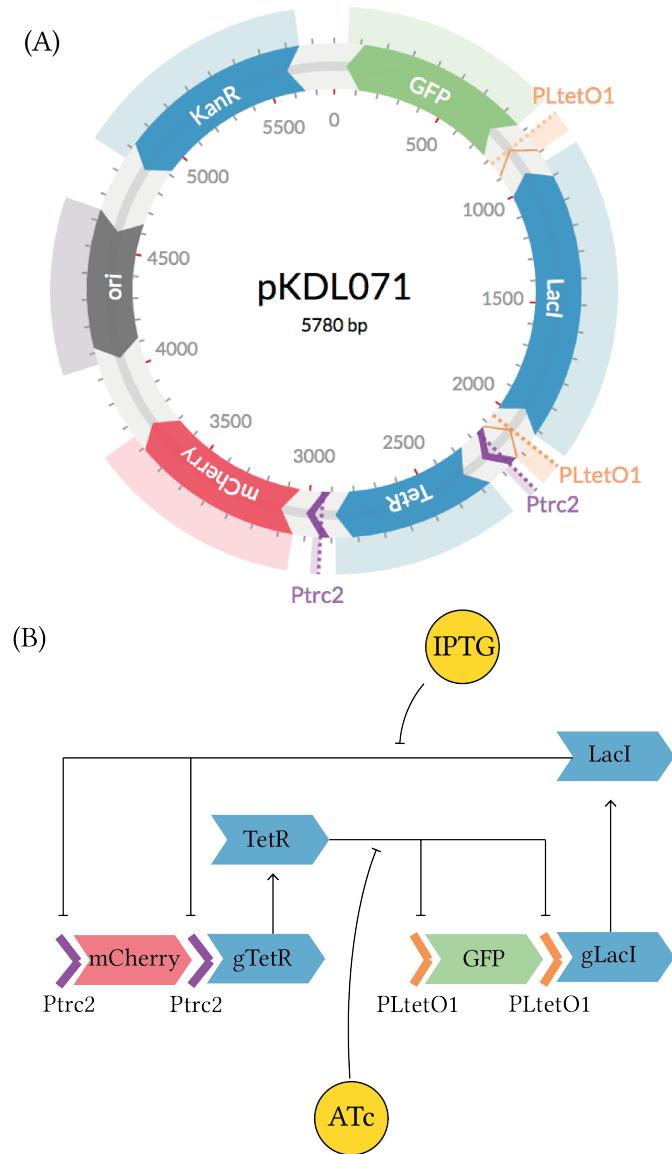


Figure 1.13 : The genetic toggle switch circuit used in this chapter. (A) The plasmid map of pKDL071, the plasmid containing the genetic toggle switch used in Litcofsky et al. (2012) (B) The interactions between each element of the circuit.

1.6.2 Methods

The toggle switch plasmid was provided by the James J Collins lab in the form of a stab culture in *E. coli* K-12 MG1655.

1.6.2.1 *Escherichia coli* culturing conditions

Lysogeny broth (LB) was made by diluting LB in deionized water to a concentration of 25 g L^{-1} and subsequently autoclaved. LB agar plates were made by adding bacteriological agar to the above solution to a concentration of 45 mg mL^{-1} before autoclaving. The solution was then cooled down to $55\text{ }^{\circ}\text{C}$ using a water bath. If antibiotic was required it was added to the correct concentration to the cooled solution. The solution was then aliquoted to plates and left to solidify in room temperature. The plates were stored in the fridge for up to 1 month.

Overnight cultures were made by picking a single colony from a static culture in an agar plate. Each colony was placed in 15 mL Falcon tubes (Fisher Scientific, MA, U.S.A) with 5 mL LB with kanamycin antibiotic at a concentration of $50\text{ }\mu\text{g mL}^{-1}$. The tubes were then screwed loosely and taped securely in order to allow for aeration. The falcon tubes were put in an incubator at $37\text{ }^{\circ}\text{C}$ with orbital shaking at 200 rpm for 12-16 hours.

1.6.2.2 Glycerol stock preparation

To preserve the transformed cultures long-term glycerol stocks were made. 5 mL LB and Kanamycin overnight cultures were made as described in Section 1.6.2.1. The cultures were kept on ice and 70 % glycerol was added to the cultures in a ratio of glycerol to culture of 1:7. These were aliquoted into cryovials and transferred to a $-80\text{ }^{\circ}\text{C}$ freezer for long-term storage.

1.6.2.3 Revival

For subsequent revival of the frozen cultures, a 1.5 mL eppendorf tube was removed from the $-80\text{ }^{\circ}\text{C}$ freezer and put on ice. Small amount was streaked onto an agar plate containing LB and kanamycin. The plates were stored in an incubator at $37\text{ }^{\circ}\text{C}$ overnight. Then the plates were sealed using parafilm and stored at $4\text{ }^{\circ}\text{C}$ for up to two weeks.

1.6.2.4 Plasmid construction

Plasmids were constructed via PCR cloning. PCR primers were chosen to add restriction enzyme sites on the 5' and 3' were needed. Following PCR amplification, the amplified DNA was purified using the Qiagen PCR cleanup kit (Qiagen, Crawley, U.K). Double digests were carried out and the desired fragment isolated via gel extraction. The relevant fragments were subsequently ligated. Following construction, each plasmid was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Crawley, U.K). Plasmid concentration was determined using the Thermo Scientific NanoDrop 1000 Spectrophotometer (Fisher Scientific, MA, U.S.A).

1.6.2.5 Polymerase Chain Reaction

In order to amplify DNA and add the restriction enzyme sites required, a Polymerase Chain Reaction (PCR) reaction was carried out with mutagenic primers. A list of primers can be found in Appendix (XXX). Q5® DNA Polymerase (NEB, MA, U.S.A) was used with its associated buffer, dNTPs and Q5® enhancer, as specified in Table 1.2. PCR reactions were run in a T100™ thermal cycler (Bio-Rad Laboratories, Inc., UK) as per the Q5® recommendations, and as outlined in Tables 1.2 and 1.3.

Table 1.2 PCR recipe

Reagent	Final concentration	50 µL reaction
Q5® buffer 5X	1X	10 µL
dNTPs	200 mM each	1 µL
Forward primer	0.5 µM	2.5 µL
Reverse primer	0.5 µM	2.5 µL
Template DNA	2 µg/50 µL	-
Q5® DNA polymerase	0.02 U µL ⁻¹	0.5 µL
Q5® enhancer	1X	10 mL
H ₂ O	-	to 50 µL

Table 1.3 Thermocycling conditions

Step	Cycles	Temperature	Time
Initiation	1	98 °C	30 s
Denaturation		98 °C	10 s
Annealing	30	72 °C	20 s
Extension		72 °C	2 min
Final extension	1	72 °C	2 min
Hold	1	4 °C	∞

1.6.2.6 Digestion

All enzymes, buffers and Bovine Serum Albumin (BSA) were supplied by NEB. Digestion controls were carried out by adding H₂O instead of DNA in the digestion reaction. Additionally, during agarose gel electrophoresis uncut plasmid was run alongside the digested plasmid in order to detect the difference.

2 µg digests were set up by mixing the plasmid with 0.5 µL of each restriction enzyme, 3 µL 10x buffer and 3 µL 10x BSA. H₂O was added to make the reaction to 20 µL. The recipe used is shown in Table 1.4. The reactions were placed in an incubator at 37 °C for 4 hours. Finally, the solutions were analysed using agarose gel electrophoresis (Section 1.6.2.7).

Table 1.4 Digestion recipe

Reagent	Volume
PstI	0.5 µL
HindIII	0.5 µL
NEB Buffer 2.1	2 µL
BSA	0.2 µL
DNA	1 µg
H ₂ O	to 20 µL

1.6.2.7 Agarose gel electrophoresis

To make a 0.8% agarose gel, 0.4 g agarose were diluted in 50 mL 1X TAE buffer. It was further dissolved by microwaving for 1-3 minutes. The solution was left to cool for 5 minutes and then 1.5 µL gel red were added. Gel trays were prepared by putting the well comb in place and taping the ends shut. The solution was then

poured into the prepared gel trays and left to solidify for 20-30 minutes at room temperature.

Agarose gel electrophoresis was carried out by placing the poured gels into the gel tanks. The tank was then flooded with 1X TAE buffer. The DNA was prepared to be analysed by adding 4 µL loading dye to 20 µL sample. A negative control was used with H₂O instead of sample. The DNA ladder of choice was prepared by adding 1 µL H₂O and 1 µL dye to 2 µL ladder. Each sample was added to a well by pipetting. The agarose gel was ran at 90 V until the dye was 80% of the way down the gel, approximately 1 hour.

To purify the fragments from the agarose gel, the gel was placed in a UV box. Using a sterile razor blade, the desired fragment was cut out and placed in a clean eppendorf tube. The DNA was isolated from the gel using the QIAquick Gel Extraction Kit.

1.6.2.8 Ligation

A ratio of 3:1 of insert to recipient plasmid was used, 1 µL T4® DNA ligase (NEB, MA, U.S.A) and 2 µL ligase buffer. H₂O was added to make the reaction up to 20 µL. The controls used for each ligation reaction, are shown in Table 1.5. Control 1 is used to detect competent cell viability, control 2 background due to uncut vector, control 3 contamination and control 4 vector re-circularization.

The ligation reactions were placed at 4 °C for 12 hours. The reactions were then placed at 65 °C for 10 minutes to heat inactivate the T4 DNA ligase enzyme. A transformation was then carried out as per Section 1.6.2.9.

Table 1.5 Ligation controls

	Control 1	Control 2	Control 3	Control 4
Vector	Uncut	✓	✓	✗
Insert	✗	✗	✗	✓
Buffer	✓	✓	✓	✓
H ₂ O	✓	✓	✓	✓
Ligase	✗	✗	✓	✓

1.6.2.9 Transformation

Thermocompetent *E.coli* Dh5α was transformed with the constructed plasmids. Each ligation reaction was added to 50 µL of thawed competent cells. The cells were sub-

sequently kept on ice for 30 minutes, then placed at a 42 °C water bath for 45 s. The cells were then placed back on ice for 15 minutes. Then 500 µL of Super Optimal broth with Catabolite repression (SOC) were added to each ligation and placed in a 37 °C shaking incubator for 3 hours. 500 µL and 50 µL were subsequently pipetted of each ligation onto petri dishes with LB agar and the appropriate antibiotic. The plates were incubated at 37 °C for 12-16 hours. Two controls were used for the transfection protocol, a positive control with no antibiotic in the LB agar and non-transfected cells and a negative control of non-transformed cells and LB agar with antibiotic. These ensure that the cells are viable and not contaminated respectively.

Finally, the number of colonies were counted on each plate. Individual colonies were then selected from each transfection and grew each separately in 5 mL LB medium for 12-16 hours at 37 °C, 200 rpm. Glycerol stocks were then prepared from each culture, as per Section 1.6.2.2.

1.6.2.10 Colony PCR

In order to determine if the fragment was successfully inserted into the vector DNA plasmid, diagnostic colony PCR was then carried out. Primers were designed that amplified the multiple cloning site of the vector DNA plasmid. These can be found in Appendix (XXX). A PCR master mix was made for the number of colonies to be amplified, 32, with an added 10% to account for pipetting error. GoTaq® Flexi DNA polymerase (Promega Corp., WI, U.S.A.) was used with its associated buffer, dNTPs and MgCl₂ and H₂O. The recipe for the master mix is shown in Table 1.6.

Table 1.6 Colony PCR master mix recipe

Reagent	Final concentration	Master mix
GoTaq® green Flexi buffer	1X	141 µL
dNTPs	200 mM each	14.1 µL
Forward primer	0.5 µM	1.4 µL
Reverse primer	0.5 µM	1.4 µL
GoTaq® Flexi polymerase	0.02 U µL ⁻¹	3.5 µL
MgCl ₂	1X	42.2 µL
H ₂ O	-	465 µL

19 µL were then added from the master mix to each PCR tube. Each of the colonies was then lifted from the transformation from the agar plate using a 20 µL pipette tip and added it to a PCR mix by mixing. The pipette tip was subsequently used to

make a scratch into a clean agar plate, and labelled it. A PCR was then carried out according to GoTaq® Flexi polymerase recommendations, and as shown in Table 1.7.

Table 1.7 Thermocycling conditions for colony PCR

Step	Cycles	Temperature	Time
Cell lysis	1	95 °C	10 minutes
Denaturation		95 °C	30 s
Annealing	35	50 °C	1 minute
Extension		72 °C	1 min
Final extension	1	72 °C	5 min
Hold	1	4 °C	∞

Finally a diagnostic agarose gel electrophoresis was carried out as outlined in Section 1.6.2.7.

1.6.2.11 Sequencing

In order to confirm plasmid identity, all plasmids were sequenced using Source Bioscience, Cambridge UK. 10 µL of each plasmid DNA were submitted at a minimum of 100 ng µL⁻¹ as per the requirements. Primer sequences were also submitted and manufactured by Source Bioscience. Primers can be found in Appendix (XXX).

1.6.2.12 Inducers

Anhydrotetracycline (ATc) solution was made by diluting ATc from Cayman Chemical Company in 100 % ethanol to a concentration of 1 mg mL⁻¹. Isopropyl-beta-D-thiogalactopyranoside (IPTG) solution was made by dissolving IPTG in deionized water to a concentration of 1 M. The solution was sterilised by passing the solution through a 0.22 µm syringe filter. Both inducers were stored in 1 mL aliquots at -20 °C.

1.6.2.13 Growth rate measurement

Plate reader analysis was carried out in order to measure the growth of *E.coli* over time. Overnight cultures were made using the method shown in Section 1.6.2.1. Overnight cultures were then diluted by a 1:1000 ratio into a 5 mL LB + kanamycin solution. The diluted cultures were grown at 37 °C with shaking at 200rpm for 1 hour. These cultures were then further diluted by a 1:100 ratio. 200 µl aliquots of the dilutions were then transferred to a clear bottom, black-walled 96-well plate.

30 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

Wells with only LB and kanamycin were also added in order to be used as blanks. The plate was then sealed using a gas permeable membrane and placed it in BMG FLUOstat OPTIMA plate reader to measure absorbance. The plate reader was set to a constant 37 °C, with 30 seconds orbital shaking at 150 rpm and 4 mm shaking width every ten minutes. Absorbance was measured at 540 nm. Data was exported as a CSV file and analysed using Python.

1.6.2.14 Flow cytometry

Flow cytometry experiments were carried out in order to get fluorescent levels in single cells. Flow cytometry allows us to gather this information for thousands of single cells. Flow cytometry data was exported as FCS files and analysed using the R bioconductor packages flowCore (Ellis et al. 2016b), flowViz (Ellis et al. 2016a) and Ggplot2 (Wickham 2009). Prior to analysis the raw data was processed to remove any debris or instrument noise detected. The data was also processed to removed any doublets, which occurs when more than one bacterial cell passes through the detector at a time. This will skew the data by including datapoint with double the fluorescent intensity that the rest of the population. The pre-processing was done by using the side scattering data. The height and the area of the sample forward scattering distribution is recorded during an experiment. The cells that lie in the diagonal where the area equals the height are single bacterial cells. If the area of the signal exceeds the height it is indicative of a doublet, or cluster of cells, and is removed from the data. This preprocessing was carried out using autoGate, developed by Fedorec (2016).

1.6.2.15 Concentration assays

Concentration assays were carried out in order to determine the concentration of each inducer (ATc and IPTG) at which the switch flips. Separate overnight cultures were prepared as per Section 1.6.2.1 with added IPTG at a concentration of 1 mM or added ATc at a concentration of 100 ng mL⁻¹ (Litcofsky et al. 2012). The cultures were then diluted by 1:1000 into fresh LB medium with varying concentrations of the opposite inducer than what the cells were grown in overnight. The concentrations used are shown in Table 1.8. For each concentration, three replicates cultures were made.

The cultures were placed in an incubator at 37 °C, 200rpm for 5 hours. The cultures were then placed in a centrifuge and spun at 13,000rpm for 5 minutes. The supernatant was discarded and replaced it with 1 mL PBS solution. The BD

Table 1.8 Concentrations used for flow cytometry assay

ATc (ng/ml)	IPTG (M)
0.05	1e-7
0.06	6e-7
0.07	1e-6
0.08	6e-6
0.09	1e-5
0.1	1e-3
1.0	0.1

32 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

LSRFortessaTM cell analyzer (Becton, Dickinson and Company) was used at the St. Mary's Flow Cytometry Core Facility at Imperial College London for flow cytometry analysis. GFP was excited using the 488 nm laser and detected using the 533/30 filter. mCherry was excited using the 561 nm laser and detected using the 620/10 filter. Data was obtained at n=10000 events per experiment.

1.6.2.16 Time course assays

Time course assays were carried out to measure the time it takes for the switch to flip to each state. Separate overnight cultures of pKDL071 were prepared as per Section 1.6.2.1 with added IPTG at a concentration of 1 mM or added ATc at a concentration of 100 ng mL⁻¹ (Litcofsky et al. 2012). Overnight cultures of pSEVA281G and pSEVA281C were also made. The cultures were then diluted by a ratio of 1:1000 into fresh LB medium. Separate cultures for each time point were made, in triplicate. For cultures grown overnight in IPTG, ATc was added at a concentration of 100 ng mL⁻¹ and for cultures grown overnight in ATc, IPTG was added at a concentration of 1 mM. All cultures were placed at 37 °C, 200rpm incubator. At 30 minutes, 1 hour and then every hour up to 6 hours flow cytometry was carried out for the corresponding cultures. Triplicates for each induction were removed from the incubator and placed in a centrifuge at 13, 000rpm for 10 minutes. The supernatant was discarded and replaced with 1 mL PBS solution. These cultures were then analysed in an AttuneTM NxT Flow Cytometer (Thermo Fisher Scientific) at University College London. GFP was excited using the 488 nm laser and detected using the 533/30 filter. mCherry was excited using the 561 nm laser and detected using the 620/10 filter. Data was obtained at n=10000 events per experiment. pSEVA281G and pSEVA281C cultures were used to set the laser voltages and pKDL071 cultures to detect the bacteria population.

1.6.3 Results

1.6.3.1 pKDL071 plasmid alteration

The pKDL071 plasmid contains all the elements of the switch. The two states of the switch are LacI high and TetR high. These are detected by using the fluorescent proteins that are controlled by the same promoters, and thus mirror the levels of LacI and TetR. The concentration of LacI can be estimated by GFP intensity and TetR concentration by mCherry intensity. In order to detect GFP and mCherry levels within each cell simultaneously, flow cytometry can be used. The lasers needed to

excite GFP and mCherry are 488 nm blue and 561 nm yellow respectively. Since the blue laser was not available for use in the BD AcuriTM C6 or the BD LSRIITM (Becton, Dickinson and Company) flow cytometers available, an alternative construct had to be made in order to be able to detect the levels of both sides of the switch.

In order to alter the switch construct to be able to detect both sides, the mCherry gene was swapped for the YFP gene. The yellow fluorescent protein is excited by the blue laser and could thus be detected using the equipment available. The YFP gene was available from BioBrick registry of standard biological parts as BBa_K592101. PCR cloning was used to introduce the flanking sequences of EcoRV and KasI restriction enzymes in the 5' and 3' ends respectively. The primers used are given in Appendix (XXX). A double digest was performed on plasmids pKDL071 and BBa_K592101, as well as positive and negative controls. Following gel extraction and ligation, the pKDL071-YFP plasmid was complete. The plasmid map is shown in Figure 1.14.

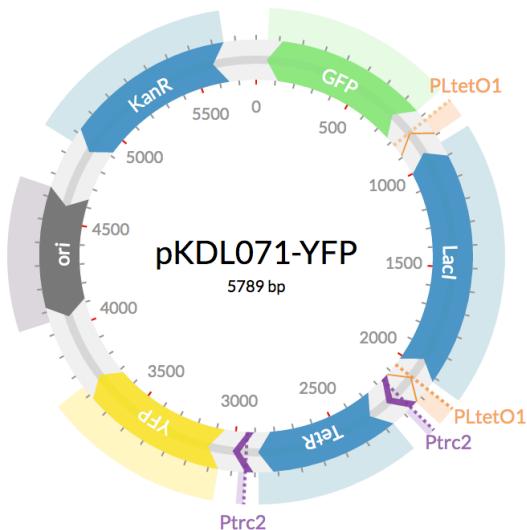


Figure 1.14 : pKDL071-YFP plasmid map.

GFP and YFP have overlapping emission spectra, which have to be compensated during flow cytometry data acquisition (Shapiro 1941). This is because the signal from GFP can be detected at the YFP detector and vice versa. Due to the high level of compensation needed to be carried out and the relatively dim signal given by the bacteria used here, the different stages of the switch, ON and OFF, could not be resolved (data not shown). In order to be able to acquire toggle switch flow cytometry data, an alternative facility was found that was able to detect GFP and

mCherry fluorescence.

1.6.3.2 Control plasmids construction

I constructed two plasmids in order to use them for the flow cytometry mCherry/GFP experiments. The first plasmid, pSEVA281G contains the promoter PLtetO-1 and GFP and the other, pSEVA281C, contains the promoter Ptrc2 and mCherry from PKDL071, shown in Figure 1.15. These two plasmids were used to determine the appropriate voltages for the lasers that excite GFP and mCherry.

pSEVA281G was constructed by digesting pKDL071 and pSEVA281 using the protocol outlined in Section 1.6.2.6. pSEVA281 is a plasmid backbone containing kanamycin resistance, a high copy origin of replication and a multiple cloning site. The digested fragments were isolated using gel purification (Section 1.6.2.7) and then ligated the isolated fragments (Section 1.6.2.8). *Escherichia coli* Dh5 α was then transformed with each plasmid (Section 1.6.2.9).

pSEVA281C was constructed via PCR cloning. PCR was carried out using the pKDL071 plasmid as a template DNA using the protocol outlined in Section 1.6.2.5. Primers were chosen so that Ptrc2 and mCherry were copied and a HindIII restriction enzyme recognition sequence added to the fragment. The rest of the cloning procedure followed as per plasmid pSEVA281G.

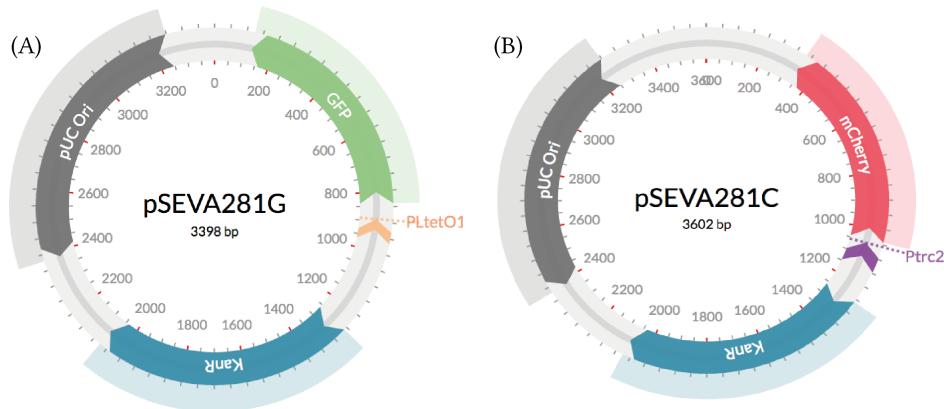


Figure 1.15 : The plasmids used to calibrate GFP and mCherry fluorescence. (A) pSEVA281G plasmid map (B) pSEVA281C plasmid map.

1.6.3.3 Growth rate investigation

I carried out a growth rate analysis to determine whether the ATc or IPTG added to pKDL071 or pSEVA281G *E. coli* cultures affected the growth of the bacteria. Cultures were grown without any inducer overnight as described in Section 1.6.2.13. Assays for the cultures were ran with and without added inducers. As can be seen in Figure 1.16, there is no difference between the conditions. The addition of either ATc or IPTG does not affect the growth rate of *E. coli* K-12 MG1655. Additionally, ATc does not affect the growth rate of *E. coli* Dh5 α . Since the addition of ATc flips the switch to the GFP high state, and IPTG to the mCherry high state, we can also conclude that the growth rate of the chassis is not affected by which side of the switch is in the high state. The growth rate of *E. coli* Dh5 α was consistently lower than that of *E. coli* K-12 MG1655.

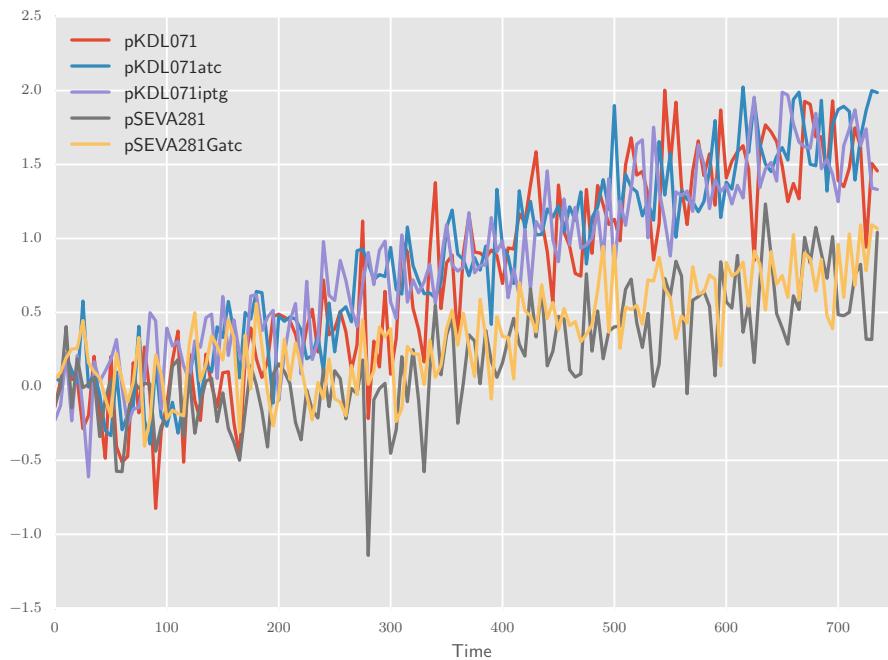


Figure 1.16 : Growth rate analysis of *E. coli* K-12 MG1655 pKDL071 and *E. coli* Dh5 α pSEVA281G cultures with and without inducers. The inducers do not affect the growth of the bacteria.

1.6.3.4 Toggle switch concentration assays

Here I aim to identify the inducer concentration at which the pKDL071 toggle switch changes state. In order to do that I carry out a concentration assay using flow cytometry, as described in Section 1.6.2.15. As can be seen in Figure 1.17A, during ATc induction the switch flips to a GFP high state when ATc concentration is at 0.09 ng mL^{-1} or higher. We observe a bimodal distribution at concentrations 0.07 ng mL^{-1} and 0.08 ng mL^{-1} , which indicates that the switching has begun at these concentrations. Thats why part of the population has switched to the GFP high state but complete switching is not observed until the concentration of ATc is at 0.09 ng mL^{-1} . In the case of IPTG induction (Figure 1.17B) we find that the switch flips to the mCherry high state when the concentration of IPTG is higher or equal to 0.001M . A decrease in GFP fluorescence is also observed. We do not observe a bimodal distribution in this case.

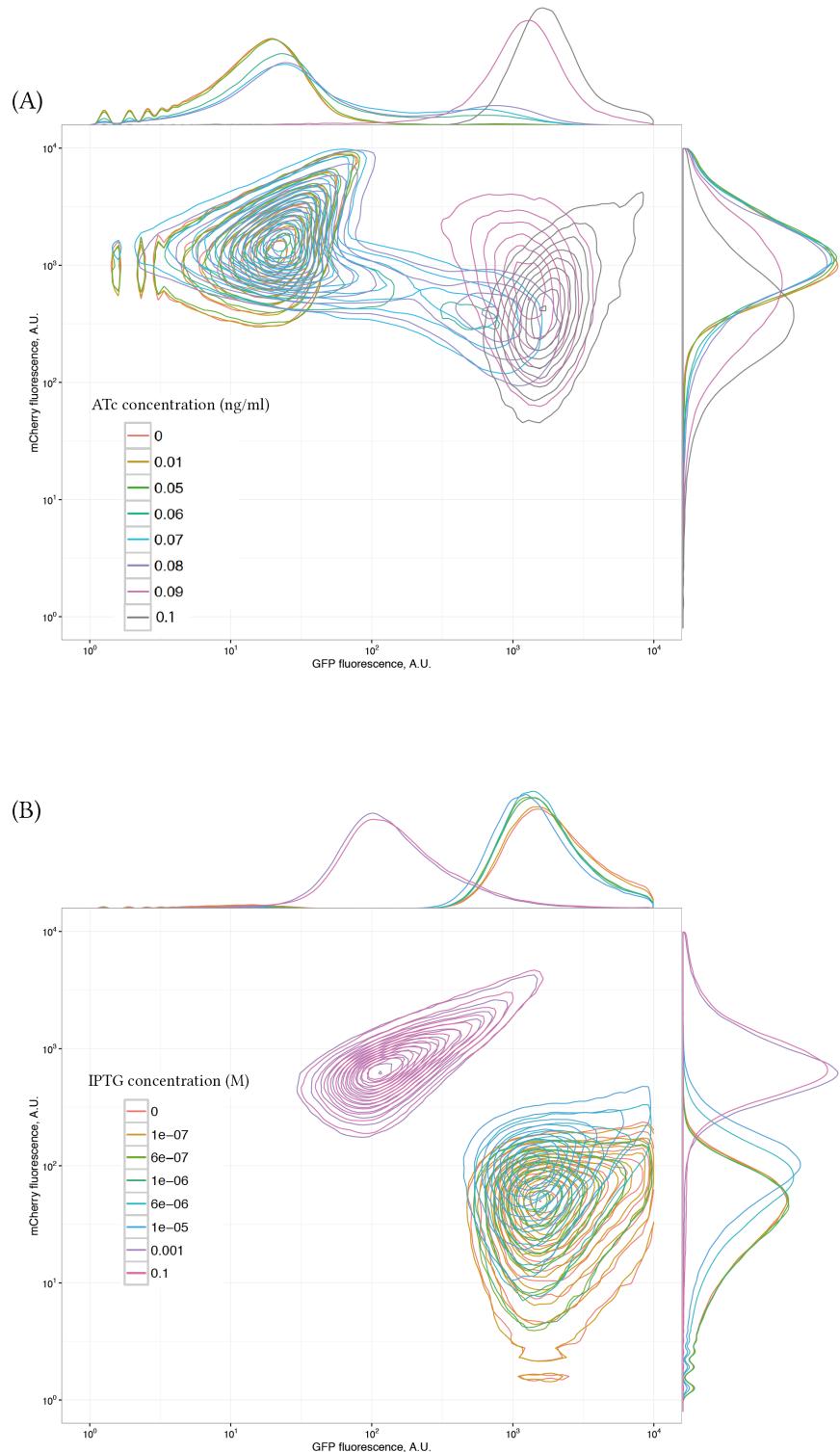


Figure 1.17 : (A) ATC induction at various concentrations (B) IPTG induction at various concentrations.

By taking into account the two induction curves of the switch turning to each high state, we can see the dynamic ranges of pKDL071 in *E.coli*. We can see in Figure 1.18 there is an approximately 100-fold change in fluorescent units during IPTG and ATc induction.

A Hill function was used to model the characterisation curves shown in Figure 1.18. The model used is the following:

$$F = P_{min} + (P_{max} - P_{min}) \frac{\left(\frac{[I]}{Kd}\right)^n}{1 + \left(\frac{[I]}{Kd}\right)^n}, \quad (1.8)$$

where F is the median fluorescent unit and [I] is the concentration of inducer. Pmin and Pmax are the minimum and maximum fluorescence respectively, and Kd and n are the dissociation constant, and Hill coefficient. I fit Hill function models using maximum likelihood estimation to the response curves. The values for parameters Pmin, Pmax, Kd, and n are 8, 1600, 0.1, 1.8 respectively for the ATc induction and 8, 700, 0.08, 2.5 for IPTG induction.

For the case of the ATc induction we observe a sharp switch between the GFP low to the GFP high state, as can be seen in the characterisation curve in Figure 1.18B. This is a clear indication of the bistability of this switch.

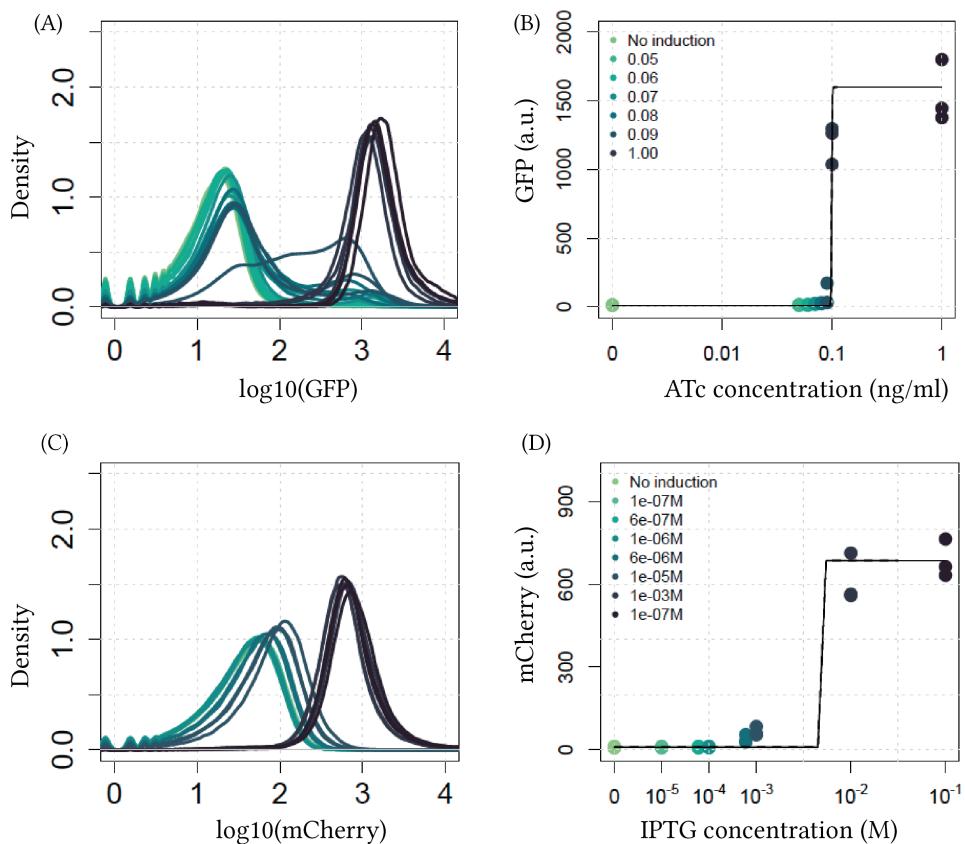


Figure 1.18 : (A, B) ATC induction of pKDL071. (C, D) IPTG induction of pKDL071.

1.6.3.5 Toggle switch time course assay

I further analysed the pKDL071 toggle switch by investigating the time it takes for it to switch from one high state to the other. To do that I used the method outlined in Section 1.6.2.16. I obtained separate time courses for the IPTG and ATc inductions.

As can be seen in Figure 1.19 pKDL071 ATc induction begins switching 1 hour after induction. Complete induction is seen at 6 hours. During the IPTG induction (Figure 1.20) we see a bimodal distribution at 4 hours, and induction is complete at 6 hours. We observe that during ATc induction there is an increase in GFP fluorescence and a decrease in mCherry fluorescence, in the case of IPTG induction the increase in mCherry fluorescence is not as prominent. A decrease in GFP fluorescence is observed during IPTG induction.

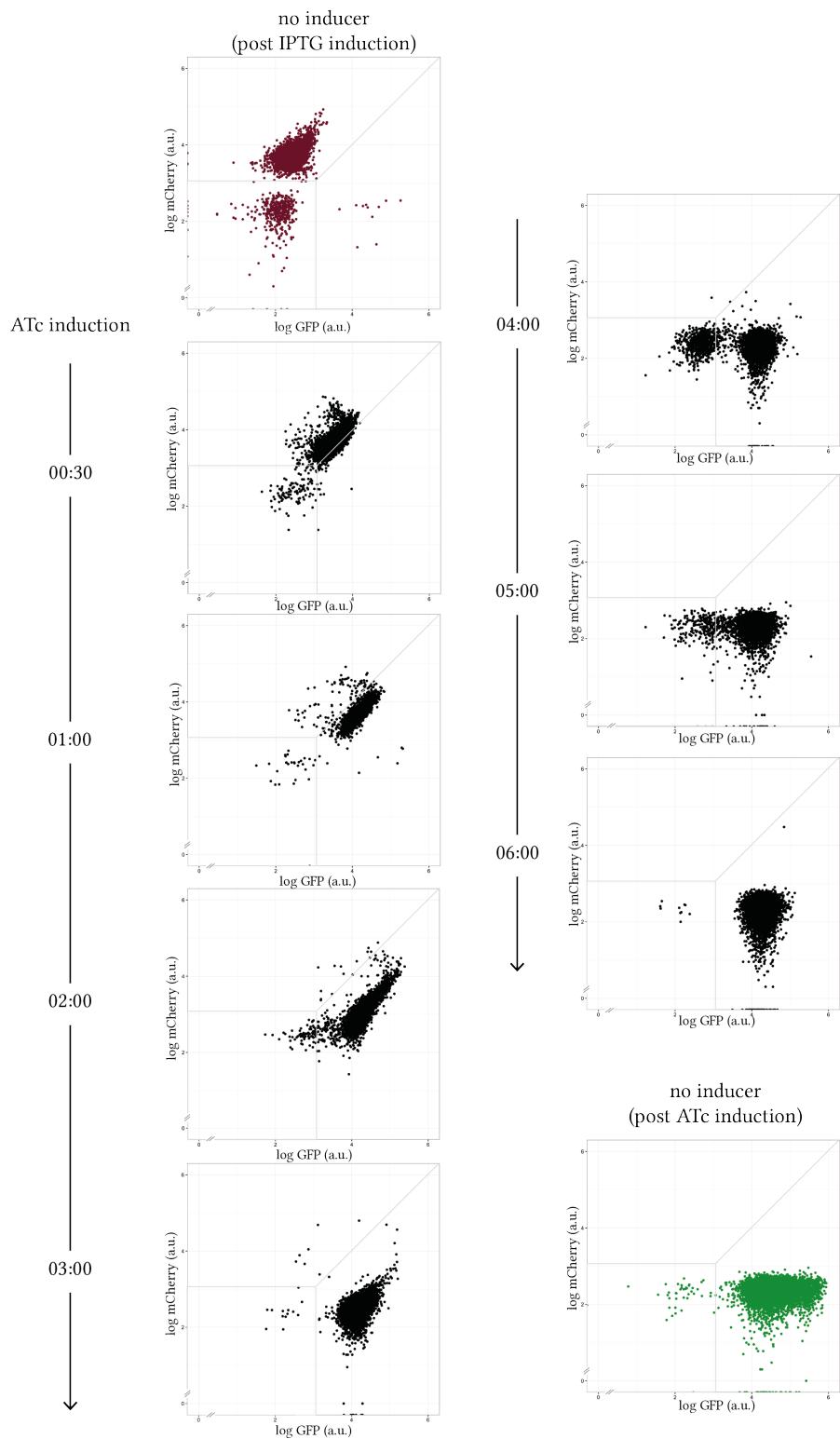


Figure 1.19 ATc induction of pKDL071 over time

42 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

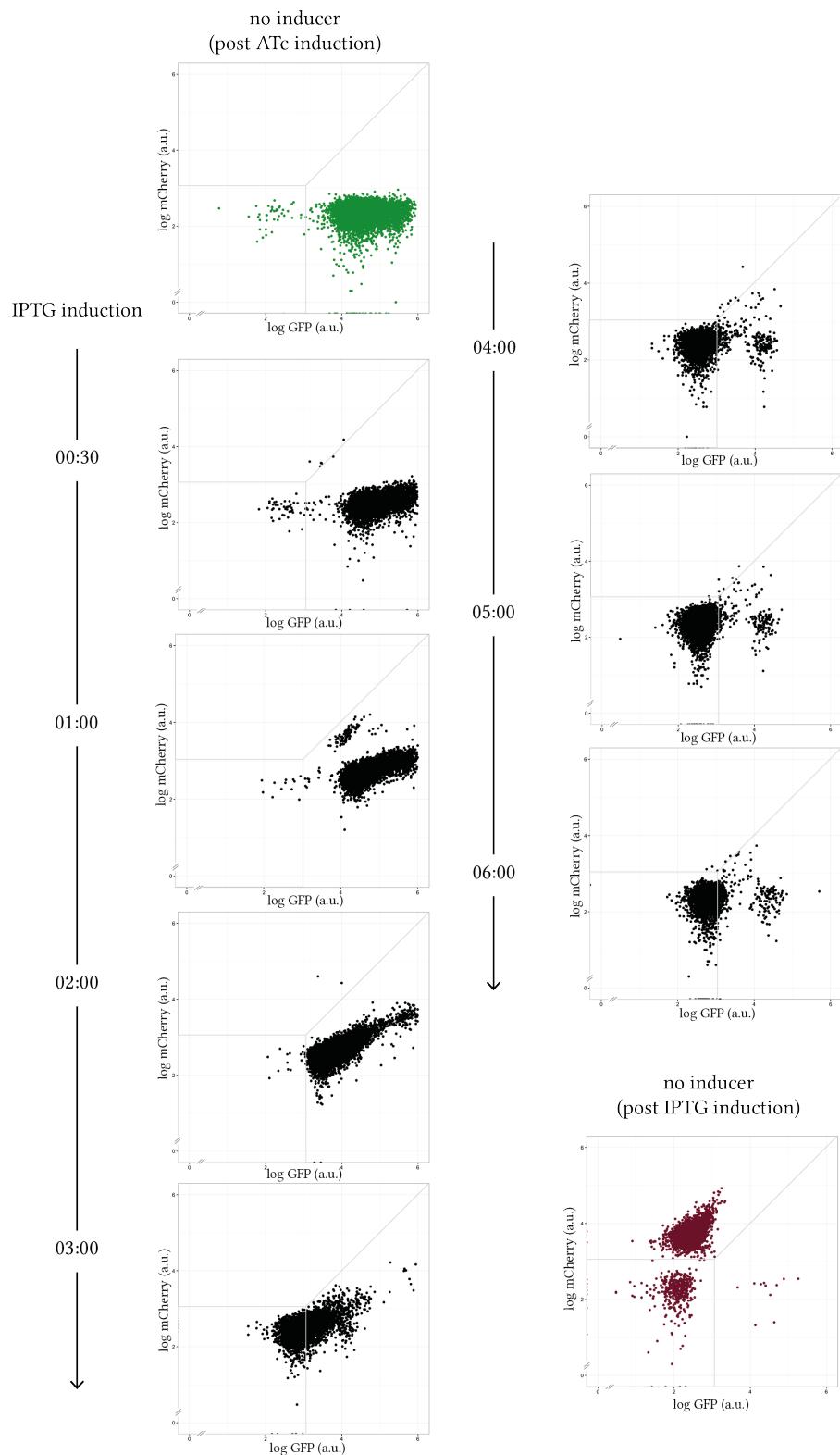


Figure 1.20 IPTG induction of pKDL071 over time

In the next section I use ABC-Flow to fit a computational model to the timecourse data obtained. Prior to fitting a model to it, I process the data by removing the unresponsive populations. This ensures that the model is fitted only to the data from cells that respond to the inducers. As seen in Figure 1.19, during the ATc induction there is an unresponsive population of cells where GFP and mCherry fluorescence are both less than 10^3 . This population is excluded from further analysis of the data. During the IPTG induction there is a population of cells that does not respond to the addition of IPTG by switching from GFP high to mCherry high. This population of cells is also excluded from further analysis.

1.7 ABC-Flow used on experimental data

In this section I apply ABC-Flow to the experimental flow cytometry data collected in Section 1.6.3.5. The data set is comprised of time course data of the Litcofsky et al. (2012) toggle switch. The two states of the switch are represented by the levels of GFP and mCherry intensity in each bacterial cell. Using ATc inducer, each cell transitions from a mCherry high state to a GFP high state.

1.7.1 Toggle switch model used in ABC-Flow

I used the extension to the Gardner, Cantor, & Collins (2000) switch described in Equations 1.4-1.7. The model used is defined by the following hazards:

$$h_1 = KD_u \times (1 + KI_u) \times GFP \quad (1.9)$$

$$h_2 = 60 \times \frac{R_u \times KL_u}{1 + KL_u + KR_u \times mCherry^2} \quad (1.10)$$

$$h_3 = KD_v \times (1 + KI_v) \times mCherry \quad (1.11)$$

$$h_4 = 60 \times \frac{R_v \times KL_v}{1 + KL_v + KR_v \times GFP^2}, \quad (1.12)$$

where GFP and mCherry represent the two fluorescent proteins in the system. The cooperativity of the repressors is represented in the model. KI_u and KI_v , KL_u KL_v , KD_u KD_v , KR_u KR_v , parameters KI_u and KI_v increase the degradation of one of the species, and simulates the addition of a repressor, IPTG or ATc respectively.

The two production hazards, h_2 and h_4 are multiplied by 60 to reflect the copy number of the toggle switch plasmid in each cell. The plasmid containing the toggle

44 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

switch used here, pKDL071, contains the ColE1 origin of replication, and thus 50-70 copies of the plasmid are present in each cell (Milo et al. 2010).

The priors used in ABC-Flow for this model are given in Table 1.9. All priors given assume a uniform distribution. The values were chosen in agreement with (Lillacci & Khammash 2013) and in reference to <http://bionumbers.hms.harvard.edu/>, the database of useful biological numbers (Milo et al. 2010).

Table 1.9 The priors used for the 1D and 2D ABC-FLow model fitting to flow cytometry data

Parameters					
Description	Symbol	Units	ATC induction	IPTG induction	
IPTG-induced mCherry degradation rate	KI_u	$h^{-1} \mu M^{-1}$			1 - 10
GFP transcription rate	R_u	molecules h^{-1}	1 - 50		1 - 50
GFP translation rate	KL_u	h^{-1}	1 - 50		1 - 50
GFP degradation	KD_u	h^{-1}	0.1 - 2		0.1 - 2
mCherry-induced GFP repression rate	KR_u	$molecules^{-1} h^{-1}$	0.016 - 1.2		0.016 - 1.2
mCherry transcription rate	R_v	$molecules h^{-1}$	1 - 50		1 - 50
mCherry translation rate	KL_v	h^{-1}	1 - 50		1 - 50
GFP-induced mCherry repression rate	KR_v	$molecules^{-1} h^{-1}$	0.016 - 1.2		0.016 - 1.2
mCherry degradation	KD_v	h^{-1}	0.1 - 2		0.1 - 2
ATC-induced GFP degradation rate	KI_v	$h^{-1} \mu M^{-1}$	1 - 10		
Species					
	GFP	$ mM$	0 - 1	100 - 1000	
	mCherry	$ mM$	100 - 1000	0 - 1	
Intensity parameters					
Mean of fluorescence of single GFP molecule	μ_{GFP}	AU	5 - 200	5 - 200	
Mean of fluorescence of single mCherry molecule	$\mu_{mCherry}$	AU	5 - 200	5 - 200	
Standard deviation of fluorescence of single GFP molecule	σ_{GFP}	AU	5 - 200	5 - 200	
Standard deviation of fluorescence of single mCherry molecule	$\sigma_{mCherry}$	AU	5 - 200	5 - 200	

1.7.2 ATc induction

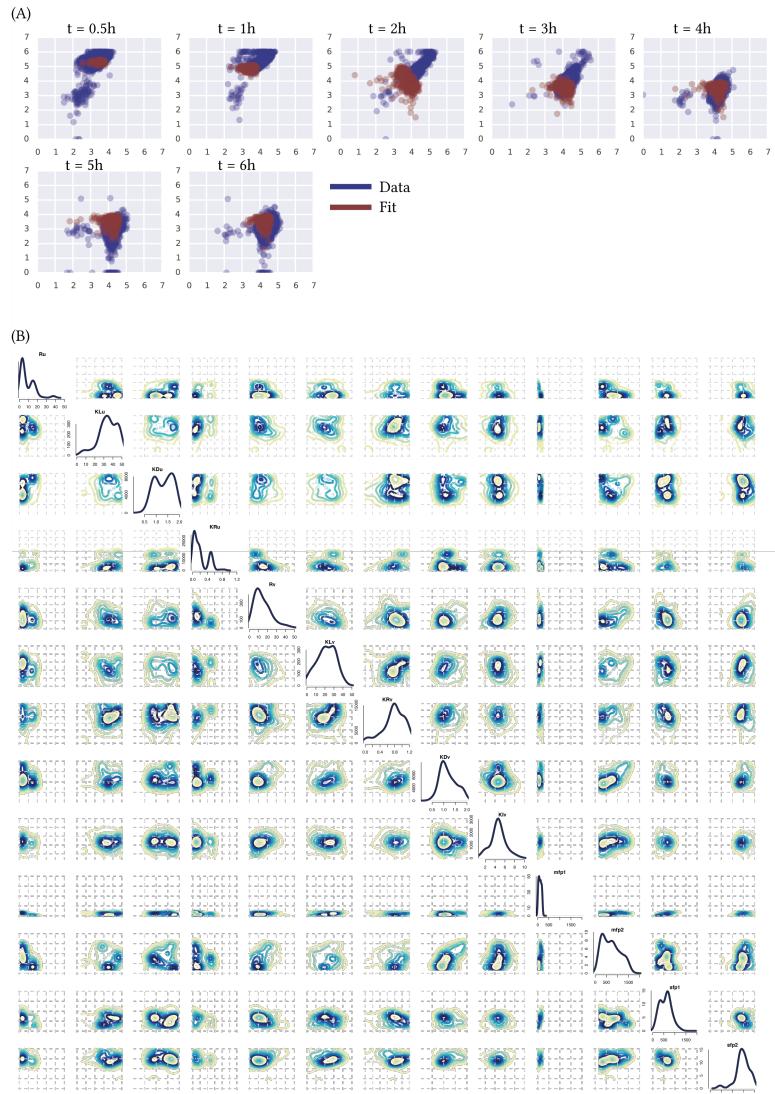


Figure 1.21 (A) The timecourse data collected in Section(XXX) (blue) and the resulting model fit from ABC-Flow (red). (B) The posterior distribution of the fitted morel.

1.7.3 IPTG induction

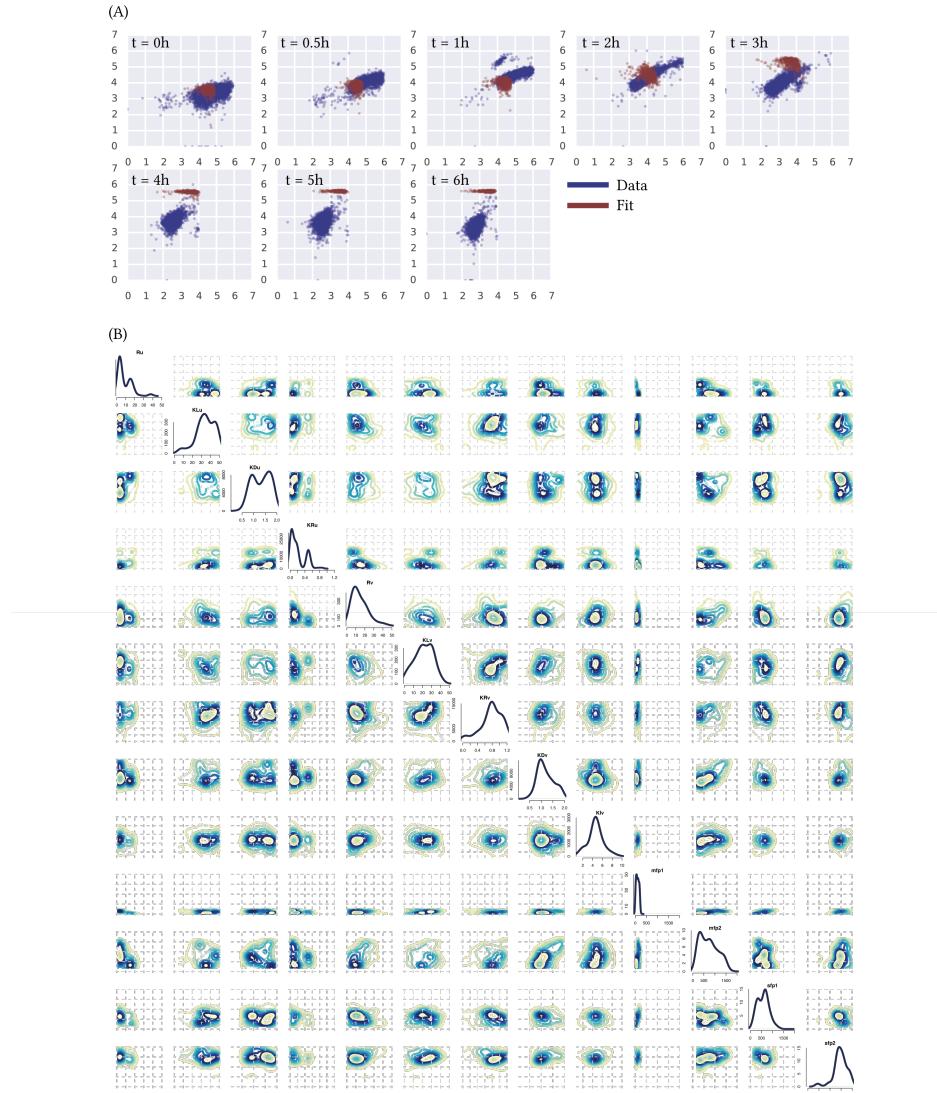


Figure 1.22

1.8 Discussion

Here I characterised the genetic toggle switch experimentally. First I study the effect of the two inducers ATc and IPTG on the growth rate of the selected chassis *E. coli* K-12 MG1655. I find that there is no detrimental effect to the bacterium by the inducers. I also find that which state the switch is on has no effect on the growth rate of the bacteria. In order for this toggle switch to be used in a synthetic biology application, it is important that both sides of the switch have an equivalent burden onto the chassis. If one of the steady states creates a larger burden and slows down the growth of the bacteria, this can create an imbalance in the population. If the toggle switch-bearing bacterial population exists in an environment with competing bacteria, for example the gut microbiome, and one of the two states creates a larger burden, this would cause the switch-bearing population to become less competitive compared to the non switch-bearing population. It is therefore crucial that the state of the switch does not affect the competitiveness of the chassis.

I further characterised the switch by determining the minimum inducer concentration necessary to change the state of the switch. I find that for ATc induction, a minimum of 0.09 ng mL^{-1} is required to cause the switch to go to a GFP high state. For IPTG induction I find that a minimum of 0.001 M is required to flip the switch to an mCherry high state. This information is critical for using this switch in other applications. Both sides of the switch are very sensitive to inducer concentrations, as the concentrations required to observe a change in fluorescence are very small.

Furthermore I find that this toggle switch, pKDL071, is faster to respond to a change in ATc concentration than to a change in IPTG concentration. For IPTG induction we observe a change in fluorescence after 3-4 hours of induction. For ATc induction we can see a difference within an hour of induction. This result is in agreement with Litcofsky et al. (2012). This difference in response times must be taken into account when using the pKDL071 switch for other applications. This difference could be due to maturation times of the fluorescent proteins. Macdonald, Chen, & Mueller (2012) found that mCherry half-maturation time is 150 mins, whereas the GFP variant used here, GFPmut3b has been especially mutated for fast action (Cormack, Valdivia, & Falkow 1996). Cormack, Valdivia, & Falkow (1996) found that whereas wild type GFP is detectable 1-2 hours after induction, GFPmut3b is detectable 8 minutes after induction. This difference could account for the different response times observed here, but further investigation is required.

ABC-Flow is similar to INSIGHT (Lillacci & Khammash 2013), as it used a Bayesian framework to fit the model to the data, and converts model output to fluorescence

intensity in order to compare the two.

1.9 Summary

In this chapter I summarised the experiments carried out for the analysis of the genetic toggle switch. I used the pKDL071 plasmid and characterised its switching behaviour over various inducer concentrations and over time. I found the concentration of each inducer necessary to flip the switch as well as the time it takes for the change to be observed. Furthermore, I investigated the effect of the inducers on the growth rate of the chassis and found that they have no effect. In the next chapter I use the data collected in the chapter to fit to the more realistic toggle switch models used in Chapter (XXX).

Bibliography

- Bower, A. G., McClintock, M. K., & Fong, S. S. (2010). ‘Synthetic biology: a foundation for multi-scale molecular biology.’ *Bioengineered Bugs* 1(5), 309–312.
- Choi, S.-L., Rha, E., Lee, S. J., Kim, H., Kwon, K., Jeong, Y.-S., Rhee, Y. H., Song, J. J., Kim, H.-S., & Lee, S.-G. (2014). ‘Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits.’ *ACS Synthetic Biology* 3(3), 163–171.
- Cooling, M. T., Rouilly, V., Misirlis, G., Lawson, J., Yu, T., Hallinan, J., & Wipat, A. (2010). ‘Standard virtual biological parts: a repository of modular modeling components for synthetic biology.’ *26*(7), 925–931.
- Cormack, B. P., Valdivia, R. H., & Falkow, S. (1996). ‘FACS-optimized mutants of the green fluorescent protein (GFP)’. *Gene* 173(1), 33–38.
- Díaz, M., Herrero, M., García, L. A., & Quirós, C. (2010). ‘Application of flow cytometry to industrial microbial bioprocesses’. *Biochemical Engineering Journal* 48(3), 385–407.
- Ellis, B., Gentleman, R., Hahne, F., Le Meur, N., Sarkar, D., & Jiang, M. (2016a). *flowViz: Visualization for flow cytometry*. R package version 1.36.2.
- Ellis, B., Haaland, P., Hahne, F., Le Meur, N., Gopalakrishnan, N., Spidlen, J., & Jiang, M. (2016b). *flowCore: flowCore: Basic structures for flow cytometry data*. R package version 1.38.2.
- Fasano, G. & Franceschini, A. (1987). ‘A multidimensional version of the Kolmogorov-Smirnov test’. *Monthly Notices of the Royal Astronomical Society* 225(1), 155–170.
- Fedorec, A. J. (2016). *autoGate*. <https://github.com/ajfedorec/autoGate.git>.
- Friedman, J. H. & Rafsky, L. C. (1979). ‘Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests’. *The Annals of Statistics*.
- Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). ‘Construction of a genetic toggle switch in Escherichia coli’. *Nature* 403(6767), 339–342.

52 BIBLIOGRAPHY

- Gillespie, D. T. (1977). 'Exact Stochastic Simulation of Coupled Chemical-Reactions'. *Journal of Physical Chemistry* 81(25), 2340–2361.
- Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., & Endy, D. (2009). 'Measuring the activity of BioBrick promoters using an in vivo reference standard.' *Journal of Biological Engineering* 3(1), 4–4.
- Kelwick, R., MacDonald, J. T., Webb, A. J., & Freemont, P. (2014). 'Developments in the tools and methodologies of synthetic biology.' *Frontiers in bioengineering and biotechnology* 2, 60–60.
- Kolmogorov, A. N. (1933). 'Sulla Determinazione Empirica di Una Legge di Distribuzione'. *Giornale dell'Istituto Italiano degli Attuari* 4, 83–91.
- Lillacci, G. & Khammash, M. (2013). 'The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations.' *Bioinformatics (Oxford, England)* 29(18), 2311–2319.
- Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., & Collins, J. J. (2012). 'Iterative plug-and-play methodology for constructing and modifying synthetic gene networks.' *Nature Methods* 9(11), 1077–1080.
- Lopes, R. H. C., Reid, I., & Hobson, P. R. (2007). 'The two-dimensional Kolmogorov-Smirnov test'. In: *International Workshop on Advanced Computing and Analysis Techniques in Physics Research*. Amsterdam, 1–12.
- Lutz, R. & Bujard, H. (1997). 'Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.' *Nucleic Acids Research* 25(6), 1203–1210.
- Macdonald, P. J., Chen, Y., & Mueller, J. D. (2012). 'Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system.' *Analytical Biochemistry* 421(1), 291–298.
- Major, S. (2016). *ndtest*. <https://github.com/syrte/ndtest.git>.
- Milo, R., Jorgensen, P., Moran, U., Weber, G., & Springer, M. (2010). 'BioNumbers—the database of key numbers in molecular and cell biology.' *Nucleic Acids Research* 38(Database issue), D750–D753.
- Monaco, J. V. (2014). 'Classification and Authentication of One-dimensional Behavioral Biometrics'. In: *International Journal of Cognitive Biometrics*, 1–8.
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). 'Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.' *Nature Biotechnology* 22(12), 1567–1572.

- Shapiro, H. M. (1941). *Practical flow cytometry*. Wiley-Liss.
- Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). 'Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*.' *Journal of Cellular and Comparative Physiology* 59, 223–239.
- Wald, A. & Wolfowitz, J. (1940). 'On a test whether two samples are from the same population'. *The Annals of Mathematical Statistics*.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.
- Wilk, M. B. & Gnanadesikan, R. (1968). 'Probability plotting methods for the analysis of data.' *Biometrika* 55(1), 1–17.