

Contents

| | | |
|---------|---|----|
| 1 | Bayesian model fitting applied to flow cytometry data | 3 |
| 1.1 | Introduction | 3 |
| 1.2 | Flow cytometry and model fitting | 3 |
| 1.3 | Circuit overview | 5 |
| 1.4 | Methods | 7 |
| 1.4.1 | <i>Escherichia coli</i> culturing conditions | 7 |
| 1.4.2 | Inducers | 7 |
| 1.4.3 | Glycerol stock preparation | 7 |
| 1.4.4 | Revival | 8 |
| 1.4.5 | Plasmid construction | 8 |
| 1.4.5.1 | Polymerase Chain Reaction | 8 |
| 1.4.5.2 | Digestion | 10 |
| 1.4.5.3 | Agarose gel electrophoresis | 10 |
| 1.4.5.4 | Ligation | 11 |
| 1.4.5.5 | Transformation | 11 |
| 1.4.5.6 | Colony PCR | 12 |
| 1.4.5.7 | Sequencing | 13 |
| 1.4.6 | Growth rate measurement | 13 |
| 1.4.7 | Flow cytometry | 13 |
| 1.4.7.1 | Concentration assays | 14 |
| 1.4.7.2 | Time course assays | 14 |
| 1.4.8 | ABC-Flow algorithm | 15 |
| 1.4.8.1 | Distance Calculations | 18 |
| 1.4.8.2 | Intensity calculation | 19 |
| 1.5 | Results | 19 |
| 1.5.1 | Growth rate investigation | 19 |
| 1.5.2 | Toggle switch concentration assays | 21 |

2 CONTENTS

| | | |
|---------|--|----|
| 1.5.3 | Toggle switch time course assay | 25 |
| 1.5.4 | ABC-Flow model fitting | 28 |
| 1.5.4.1 | Distance study | 28 |
| 1.5.4.2 | ABC-Flow validation using simulated data | 31 |
| 1.5.4.3 | ABC-Flow used on experimental data | 36 |
| 1.6 | Discussion | 37 |
| 1.7 | Summary | 38 |
| | Bibliography | 39 |

1 Bayesian model fitting applied to flow cytometry data

1.1 Introduction

In this chapter, I aim to study the genetic toggle switch experimentally. This chapter is organised as follows: In the first section I provide an overview of the circuit used and then outline the methods used for the experiments carried out. In the subsequent section I investigate the effect that the switch has on the growth rate of the bacteria. Then I examine the concentrations of the inducers and the time needed to flip the switch.

1.2 Flow cytometry and model fitting

Flow cytometry detects the fluorescent intensity levels in individual cells. It can also provide physical information about the size and granularity of a cell via the forward and side scattering respectively. An overview of flow cytometry is shown in Figure 1.1. A laser excites the fluorochrome present in the bacterial cells. The fluorochromes emit a signal that is detected by channels in the optics. The signals are then all collected and analysed. A sample typically consists single cell measurements of 10^4 - 10^5 cells.

Flow cytometry is used in synthetic biology, among others, for BioBrick characterisation (Kelly et al. 2009), enzyme screening (Choi et al. 2014) and industrial bioprocesses (Díaz et al. 2010). Flow cytometry is a powerful tool for synthetic biology as it can measure multiple parameters in single cells, and process up to 35,000 cells sec⁻¹ (*Attune NxT Acoustic Focusing Cytometer* 2015).

A drawback of flow cytometry is that the fluorescence intensity per cell is measured rather than number of proteins. Measuring absolute numbers of protein would

4 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

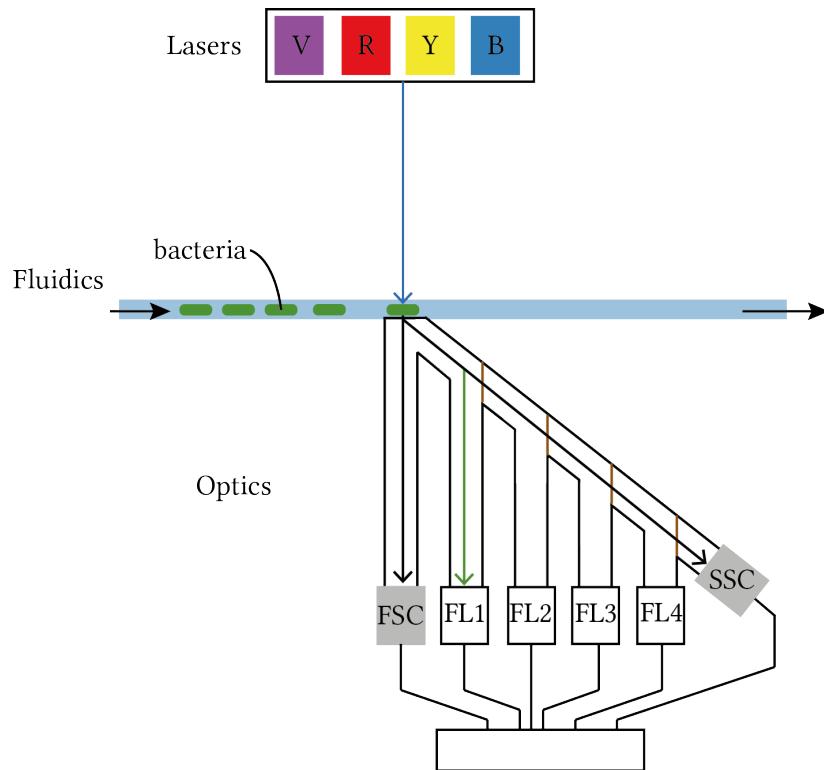


Figure 1.1 : Flow cytometry. A laser excites the fluorescent proteins present in each cell. The cytometer has up to 4 lasers, violet (V), red (R), yellow (Y) and blue (B). The detectors in the optics, FL1-4 pick up the signals. The cytometer also picks up size and granularity information via the forward scatter (FSC) and side scatter (SSC) detectors.

be ideal for model fitting, but this type of biological data cannot be directly measured (Kelwick et al. 2014). Standardization of experimental methods in flow cytometry has aided the effort to convert relative measurements, like fluorescence intensity, to absolute measurements, like $\text{GFP cell}^{-1} \text{ s}^{-1}$ (Kelly et al. 2009; Kelwick et al. 2014).

In this chapter, I developed ABC-Flow, a method to fit computational models to experimental flow cytometry data. ABC-Flow approaches the problem of absolute protein numbers by converting the model output of $\text{GFP cell}^{-1} \text{ s}^{-1}$ to relative fluorescence intensity. ABC-Flow is a python package that uses Bayesian statistics to fit the parameters of a given model to flow cytometry time course data. The algorithm and its usage is described in Section 1.4.8.

1.3 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 1.2A and the sequence given in Appendix (XXX). The circuit consists of two promoters, Ptrc2 and PLtetO-1 (Lutz & Bujard 1997). Ptrc2 is a constitutive promoter, repressible by LacI. PLtetO-1 is also a constitutive promoter, repressible by TetR, as shown in Figure 1.2B. 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.

6 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

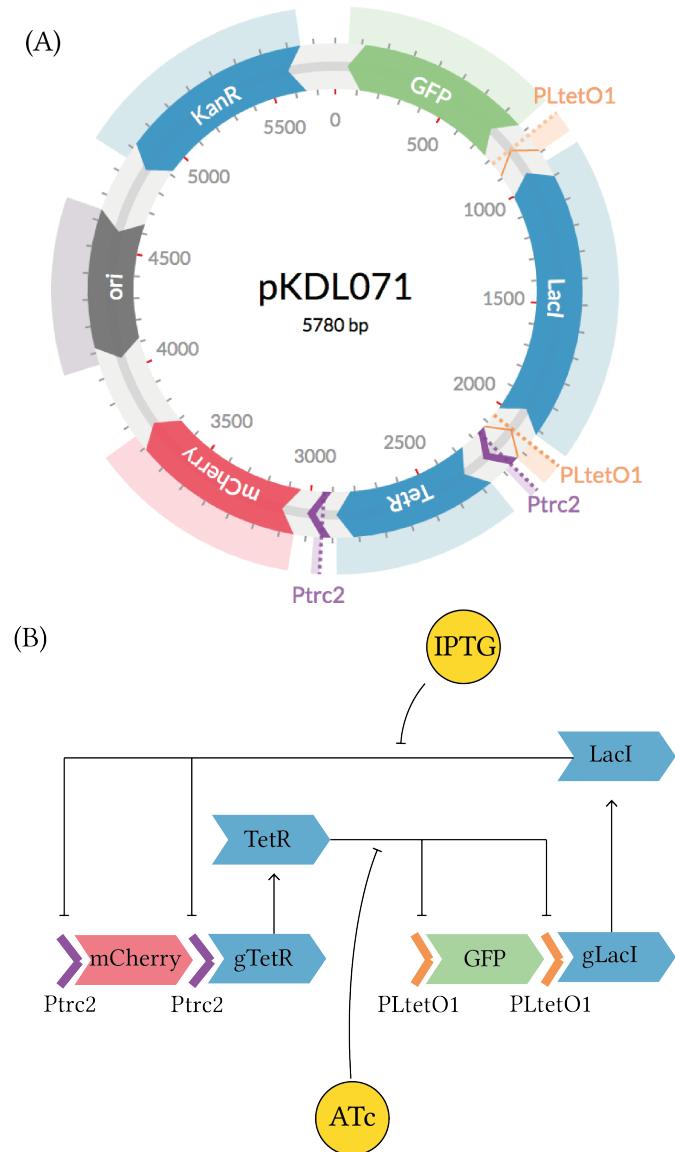


Figure 1.2 : The genetic toggle switch circuit used in this chapter. (A) The plasmid map of pKDL071, the plasmid containing the genetic toggle switch used in Litcofsky et al. (2012) (B) The interactions between each element of the circuit.

1.4 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.

1.4.1 *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.

1.4.2 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 $0.22\text{ }\mu\text{m}$ syringe filter. Both inducers were stored in 1 mL aliquots at $-20\text{ }^{\circ}\text{C}$.

1.4.3 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 1.4.1. 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.

8 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

1.4.4 Revival

For subsequent revival of the frozen cultures, a 1.5 mL eppendorf tube was removed from the -80°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°C overnight. Then the plates were sealed using parafilm and stored at 4°C for up to two weeks.

1.4.5 Plasmid construction

I constructed two plasmids in order to use them for the flow cytometry 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 1.3. 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 1.4.5.2. 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 1.4.5.3) and then ligated the isolated fragments (Section 1.4.5.4). *Escherichia coli* Dh5 α was then transformed with each plasmid (Section 1.4.5.5). pSEVA281C was constructed via PCR cloning. PCR was carried out using the pKDL071 plasmid as a template DNA using the protocol outlined in Section 1.4.5.1. Primers were chosen so that Ptrc2 and mCherry were copied and a HindIII restriction enzyme recognition sequence added to the fragment. The primers are listed in Appendix (XXX). The amplified DNA was purified using the Qiagen PCR cleanup kit (Qiagen, Crawley, U.K) and then carried out the rest of the cloning procedure as per plasmid pSEVA281G.

Following construction, each plasmid was isolated using the QIAprep Spin Mini-prep Kit (Qiagen, Crawley, U.K). Plasmid concentration was determined using the Thermo Scientific NanoDrop 1000 Spectrophotometer (Fisher Scientific, MA, U.S.A).

1.4.5.1 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 1.1. PCR reactions were run in a T100™ thermal cycler (Bio-Rad Laborat-

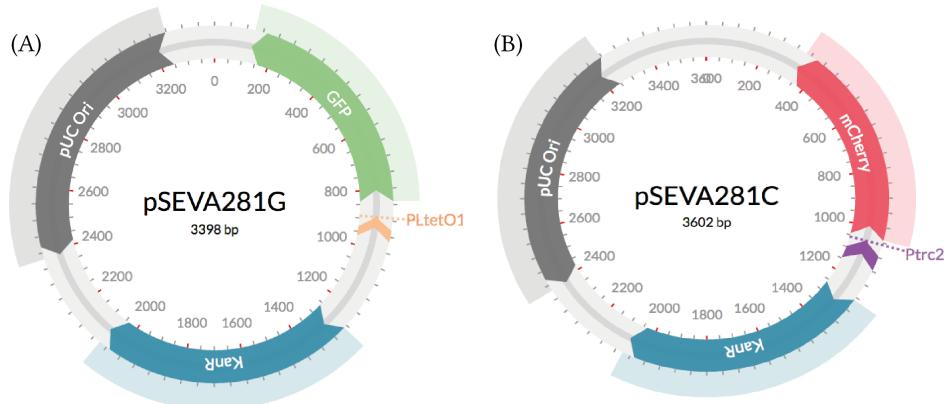


Figure 1.3 : The plasmids used to calibrate GFP and mCherry fluorescence. (A) pSEVA281G plasmid map (B) pSEVA281C plasmid map.

ories, Inc., UK) as per the Q5® recommendations, and as outlined in Tables 1.1 and 1.2.

Table 1.1 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 1.2 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 | ∞ |

1.4.5.2 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 1.3. The reactions were placed in an incubator at 37 °C for 4 hours. Finally, the solutions were analysed using agarose gel electrophoresis (Section 1.4.5.3).

Table 1.3 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 |

1.4.5.3 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.

1.4.5.4 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 1.4. 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 1.4.5.5.

Table 1.4 Ligation controls

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

1.4.5.5 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 subsequently 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.

12 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

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 1.4.3.

1.4.5.6 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 1.5.

Table 1.5 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 1.6.

Table 1.6 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 1.4.5.3.

1.4.5.7 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).

1.4.6 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 1.4.1. 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. 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.

1.4.7 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

14 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

of single cells. Flow cytometry data was exported as FCS files and analysed using the R bioconductor packages `flowCore` (`flowCore:man`), `flowViz` (`flowViz:man`) and `Ggplot2` (`ggplot2:bk`).

1.4.7.1 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 1.4.1 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 1.7. For each concentration, three replicates cultures were made.

Table 1.7 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 |

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 LSRFortessa™ 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.

1.4.7.2 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 1.4.1 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 Attune™ 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.

1.4.8 ABC-Flow algorithm

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, like the `\$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 packages `flowCore` (`flowCore:man`). ABC-Flow simulations are implemented on GPUs.

16 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

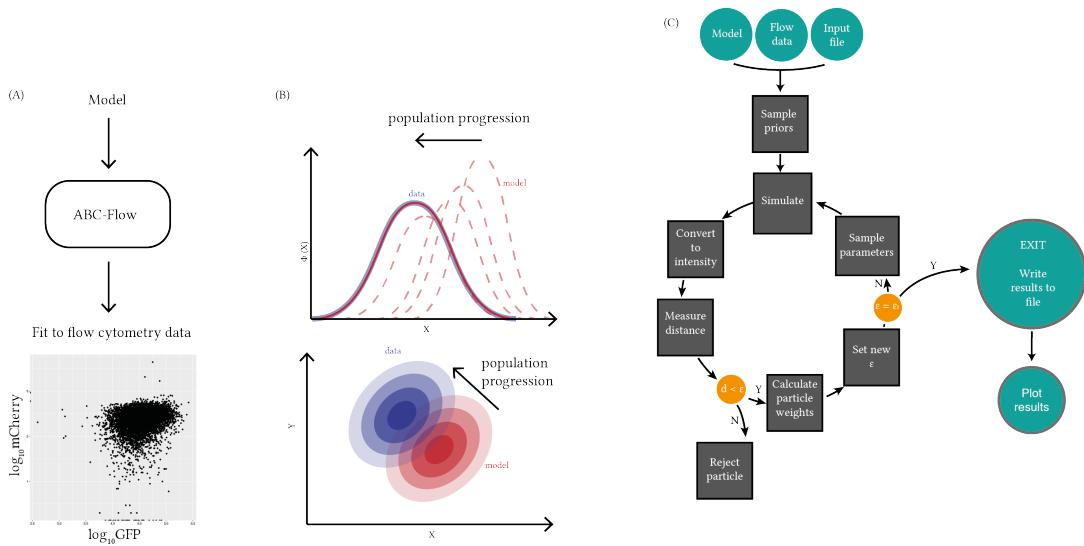


Figure 1.4 : Overview of ABC-Flow. (A) ABC-Flow is used to fit models to experimental flow cytometry data. (B) The algorithm can be applied to 1D and 2D flow data. (C) ABC-Flow uses approximate bayesian computation.

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
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  $\epsilon \leq \epsilon_T$ 

```

1.4.8.1 Distance Calculations

The distance in the 1D case was calculated using the Kolmogorov-Smirnov test (**Kolmogorov:1933**).

The Kolmogorov-Smirnov test is a non-parametric statistic test that determines whether the two distribution functions differ. The KS distance between two distributions is equal to the largest distance between the empirical distribution functions of the two samples, as illustrated in Figure 1.5 and Equation 1.1.

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)| \quad (1.1)$$

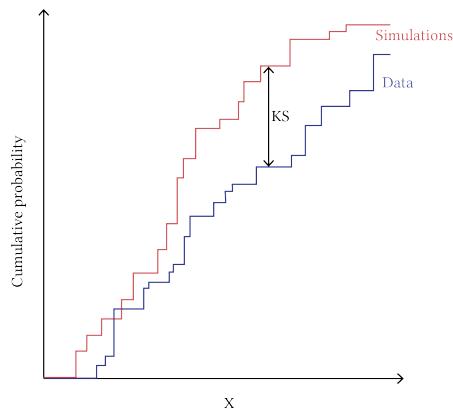


Figure 1.5 :The Kolmogorov-Smirnov distance function used to calculate the distance between distributions in the 1D case.

For the 2D case the distance was calculated in a similar way, by using an adaptation of the Kolmogorov-Smirnov distance in higher dimensions (Peacock 1983).

1.4.8.2 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:

$$X \sim N(nFP \times \mu, \sqrt{(nFP \times \sigma^2)}), \quad (1.2)$$

where nFP is the number of fluorescent proteins.

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

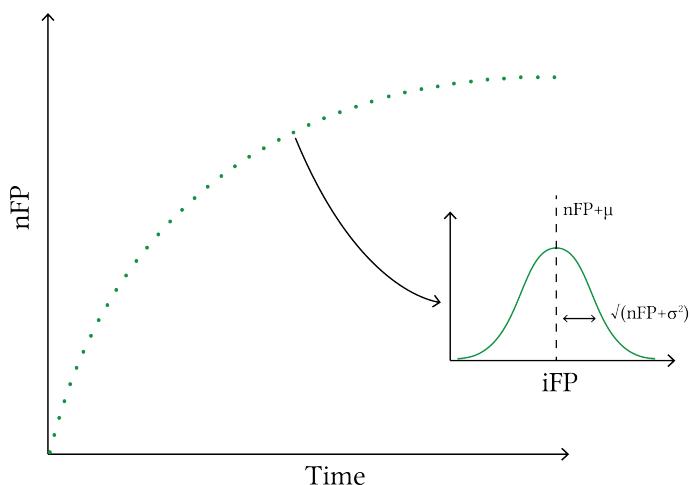


Figure 1.6 :Converting the number of fluorescent proteins to the intensity (iFP) is done by drawing from a normal distribution, as shown in Equation 1.2.

1.5 Results

1.5.1 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 1.4.6. I ran assays

20 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

for the cultures with and without added inducers. As can be seen in Figure 1.7, 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.

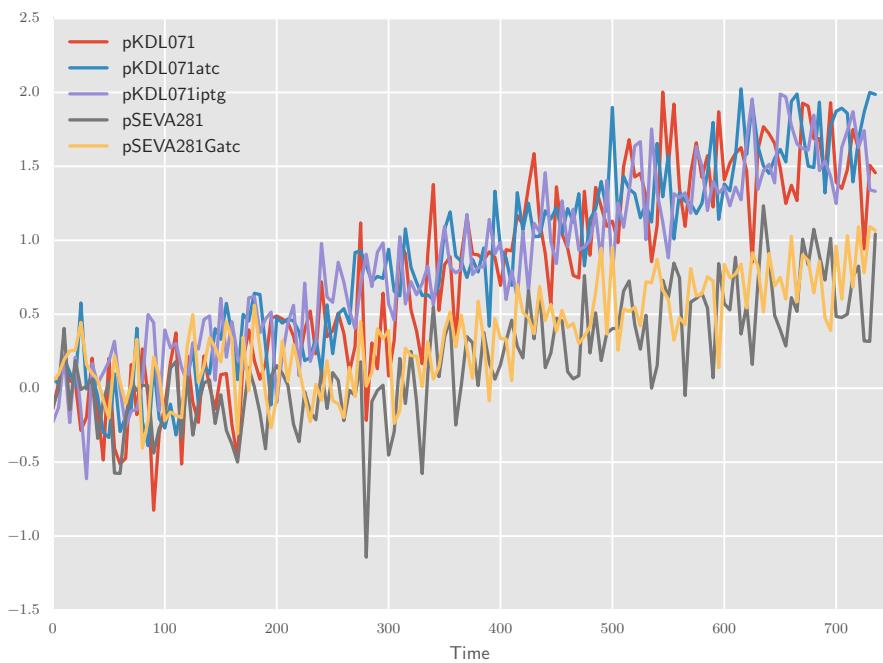


Figure 1.7 : Growth rate analysis of *E. coli* K-12 MG1655 pKDL071 and *E. coli* Dh5 α pSEVA281G cultures with and without inducers. The inducers do not affect the growth of the bacteria.

1.5.2 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 1.4.7.1. As can be seen in Figure 1.8A, 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 1.8B) 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.

22 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

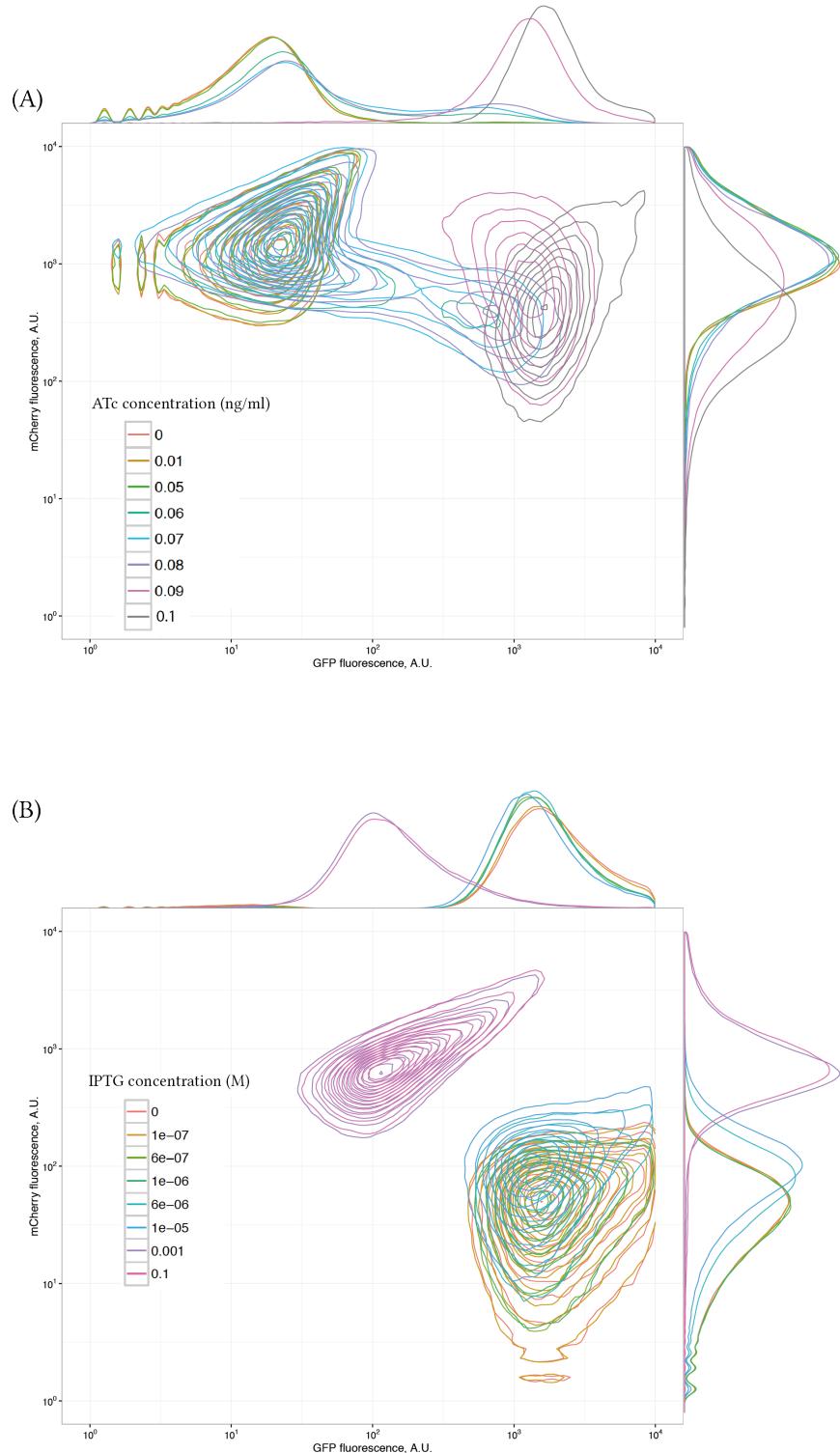


Figure 1.8 : (A) ATC induction at various concentrations (B) IPTG induction at various concentrations.

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 1.9 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 1.9. The model used is the following:

$$F = P_{min} + (P_{max} - P_{min}) \frac{\left(\frac{[I]}{K_d}\right)^n}{1 + \left(\frac{[I]}{K_d}\right)^n}, \quad (1.3)$$

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 1.9B. This is a clear indication of the bistability of this switch.

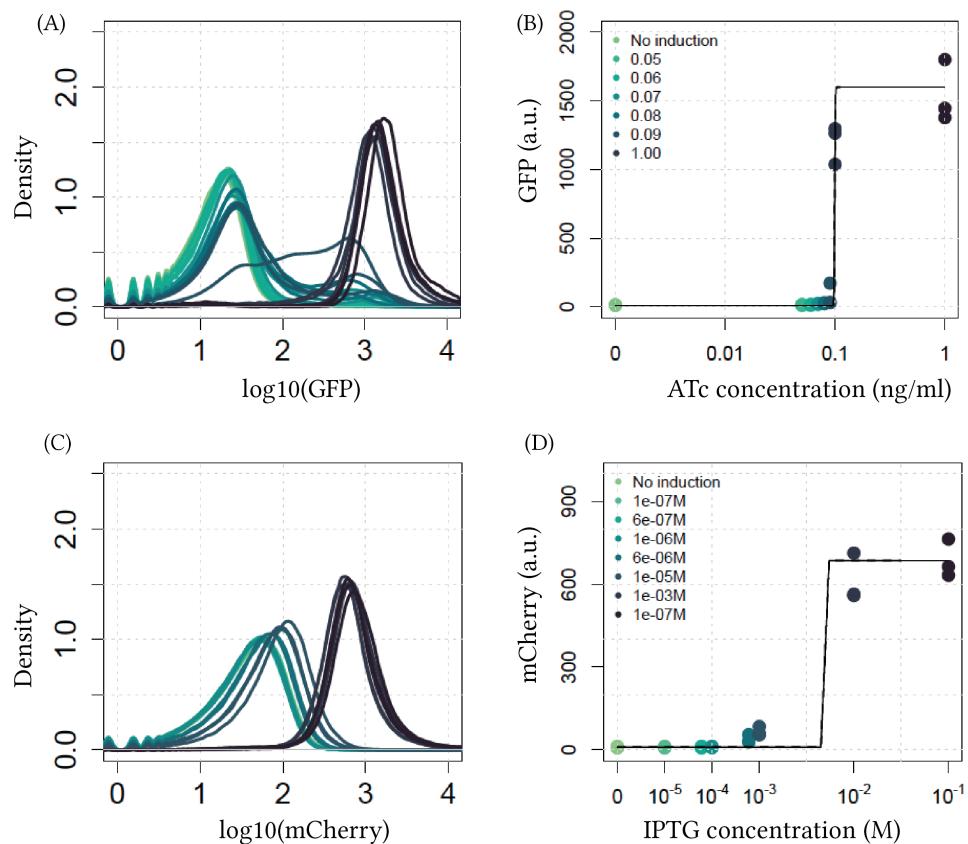


Figure 1.9 : (A, B) ATc induction of pKDL071. (C, D) IPTG induction of pKDL071.

1.5.3 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 1.4.7.2. I obtained separate time courses for the IPTG and ATc inductions.

As can be seen in Figure 1.10 pKDL071 ATc induction begins switching 1 hour after induction. Complete induction is seen at 6 hours. During the IPTG induction (Figure 1.11) 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.

26 BAYESIAN MODEL FITTING APPLIED TO FLOW CYTOMETRY DATA

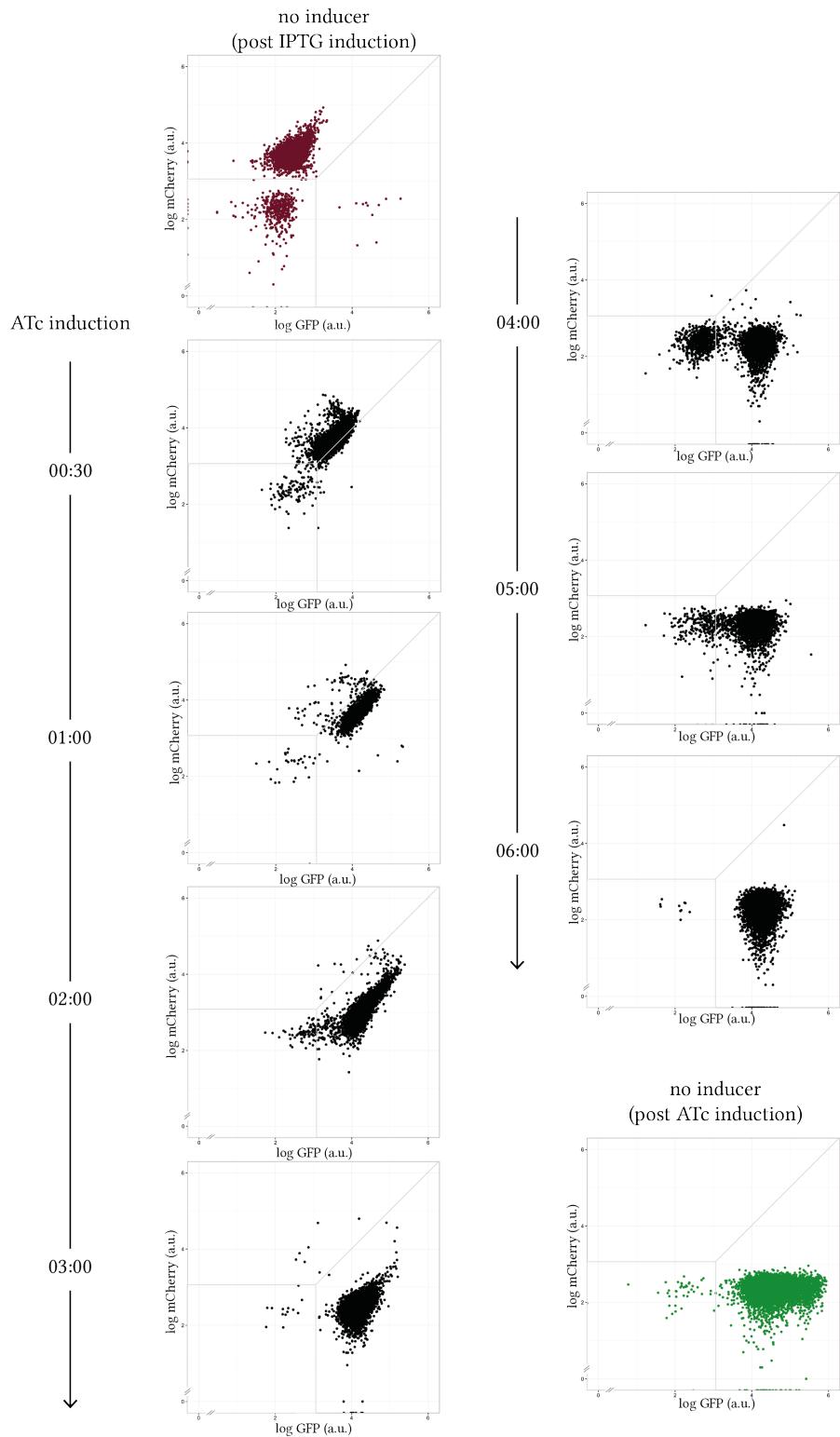


Figure 1.10 : ATc induction of pKDL071 over time

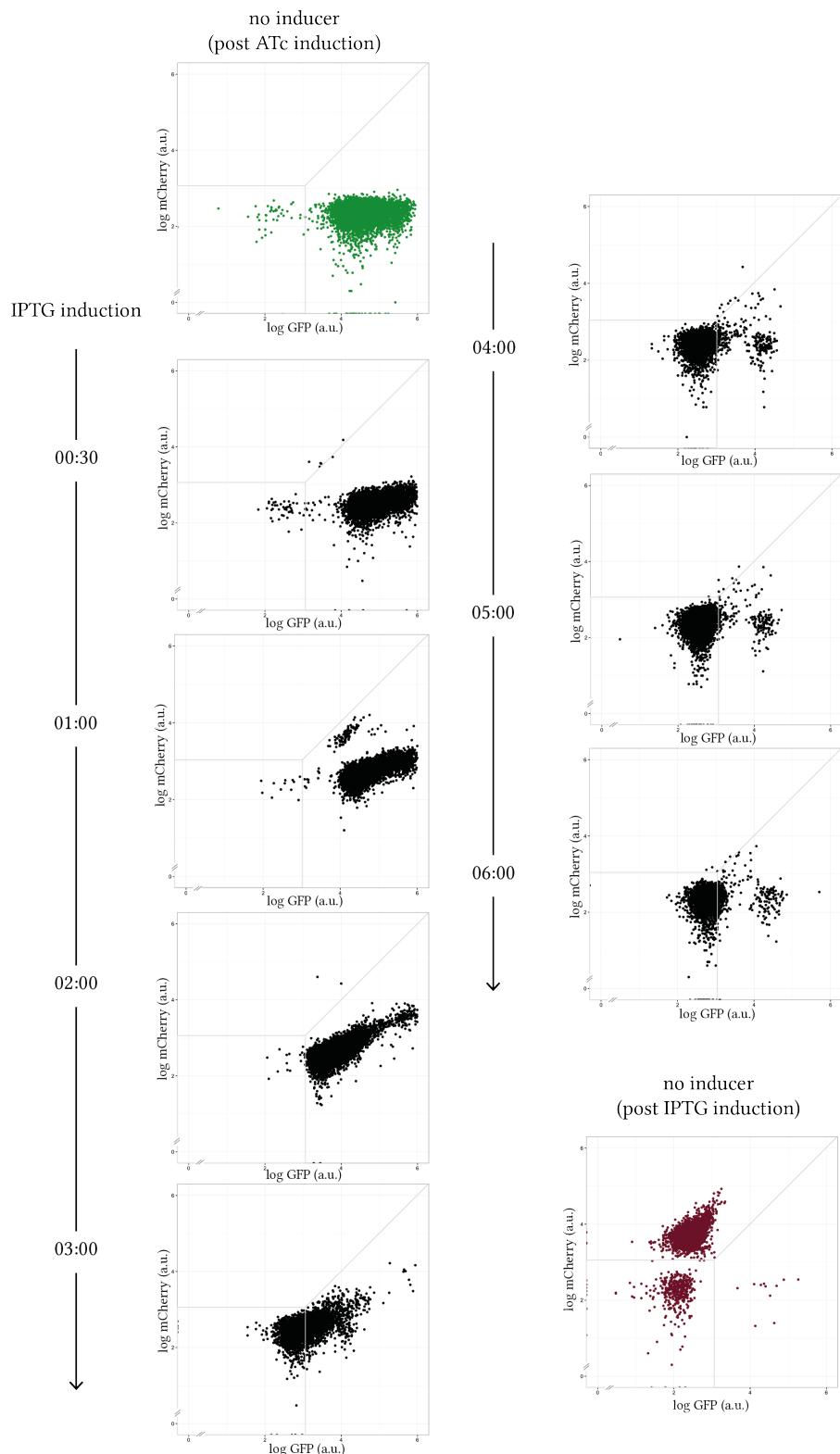


Figure 1.11 : IPTG induction of pKDL071 over time

1.5.4 ABC-Flow model fitting

In order to characterise the Litcofsky et al. (2012) toggle switch, I use the data collected in Section 1.5.3 to fit the Gardner, Cantor, & Collins (2000) toggle switch model, shown in Section (XXX). In order to do that, I use ABC-Flow as described in the Methods in Section 1.4.8.

In the following sections I first use normal distributions to study the distance functions used in ABC-Flow. Then I use a simulated data set to which I fit a toggle switch model and finally I use ABC-Flow to fit experimental data.

1.5.4.1 Distance study

Here, I simulate two normal distributions, with $\mu = 0$ and $\sigma = 1$ and measure the distance between them using the Kolmogorov-Smirnov test, as used in ABC-Flow. 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 1.12, there range of distance values obtained in the 1D case is small, as the distance between two data sets drawn from the same distribution ranges from 0.1 to 0.2. For the 2D case, the distance values obtained vary more than in the 1D case.

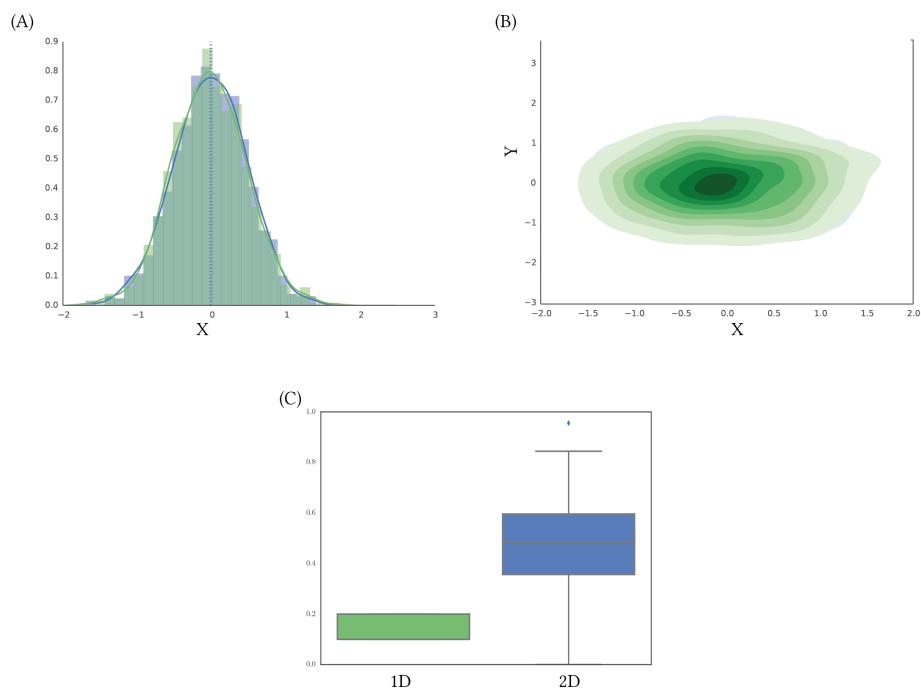


Figure 1.12 The distance between two data sets drawn from the same distribution are compared using the Kolmogorov-Smirnov two sample test. (A) in 1D and (B) in 2D. (C) The distance is calculated for 1000 data sets. A greater variation of values is found for the 2D distance calculation.

To further study the distance calculation used in ABC-Flow, the data sets are drawn from increasingly different distributions, and the distance between them calculated. As shown in Figure 1.13, the 1D distance is 0 when the difference between the μ from which the two datasets are drawn is 0.5 or less, and 1 when the difference in the μ between the two data sets is 7 or larger. In the 2D case the standard deviation of the calculated distance is larger than the 1D case. The distance calculation reaches a plateau at $\epsilon = 2.3$ when the μ difference is 3 or larger. As shown in Figure 1.13D, as the number of samples drawn from each distribution increases, the standard deviation of the distance calculations decreases in the 1D case. This trend is less obvious in the 2D case.

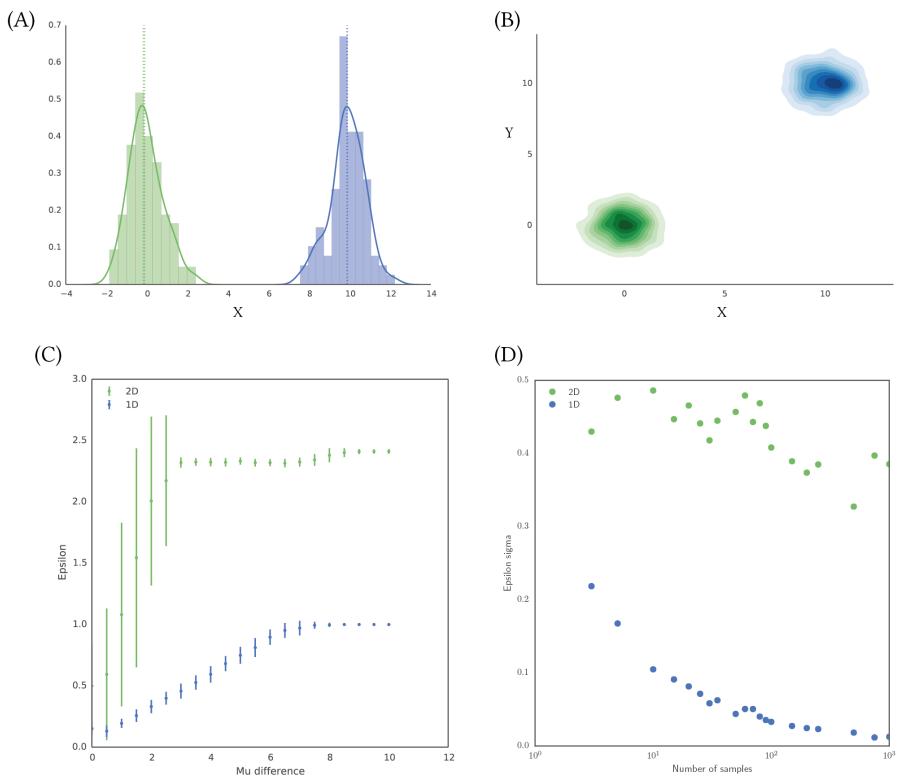


Figure 1.13 The distance calculation for data sets drawn from increasingly different distributions. Two examples are shown of distributions compared in (A) 1D and (B) 2D. (C) As the difference between the means of the two distributions increases, the distance calculation, ϵ , increases. In the 2D case (shown in green) ϵ plateaus at 2.3 and in the 1D case (shown in blue) ϵ plateaus at 1.

1.5.4.2 ABC-Flow validation using 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, representing the repression from the external repressors, IPTG and ATc.

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

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

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

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

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

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. The resulting fit is shown in Figure 1.14A. In order to determine whether this is a good fit to the data, QQ plots are produced for each timepoint (Figure 1.14B). 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).

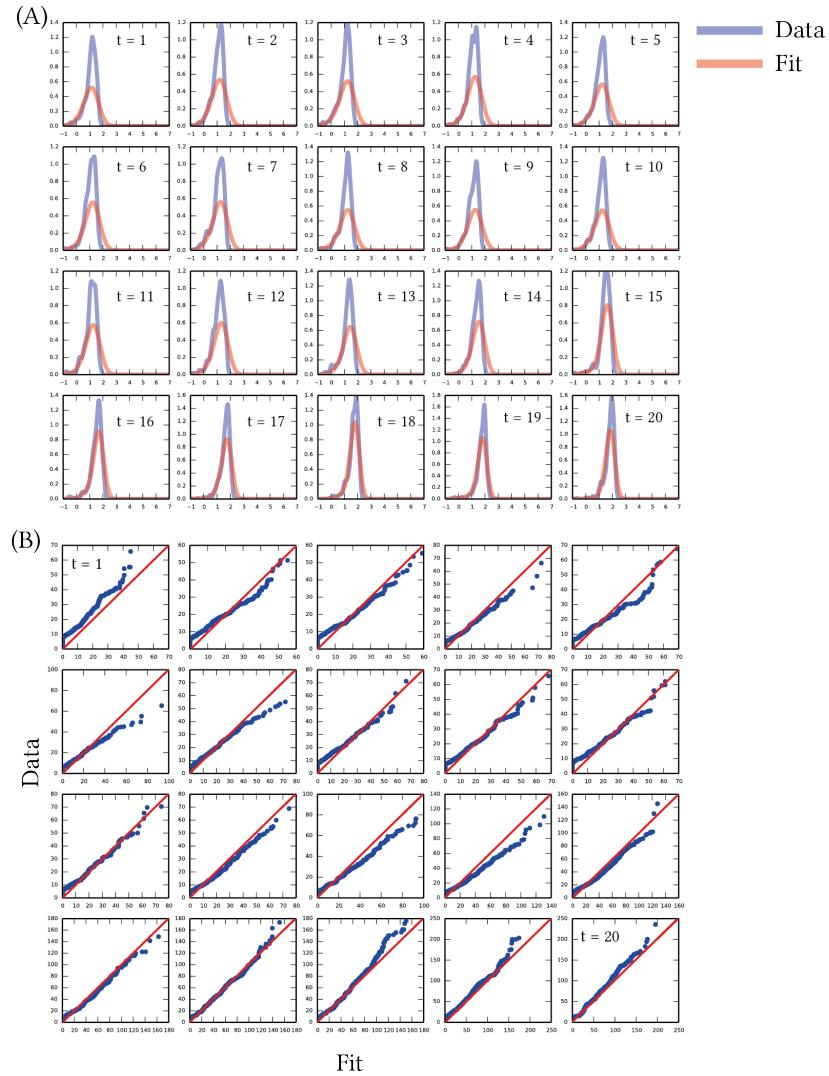


Figure 1.14 (A) 1D ABC-Flow fit (shown in blue) to data (shown in red) produced by simulating the same model. (B) QQ-plot of each time point fit. The quantile of the two distributions are plotted against each other. If the distributions are similar, the points would lie on the 45° line $x = y$, shown in red.

By examining the data and the fitted models shown in Figure 1.14, we see that at $\varepsilon = 0.1$, 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 timecourse of the model after fitting with ABC-Flow is similar to the data time course.

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. Using the same priors as in the 1D case, we can compare the two fits and investigate whether using the 1D or 2D data results in a better fit. The 1D data represents one of the marginal distributions of the data sets, as illustrated in Figure 1.15. I use the marginal distribution of u and the bivariate distribution of u and v in order to determine which one produces a better fit to the data. The resulting timecourse of the data as well as the 1D and 2D fit are shown in Figure 1.16.

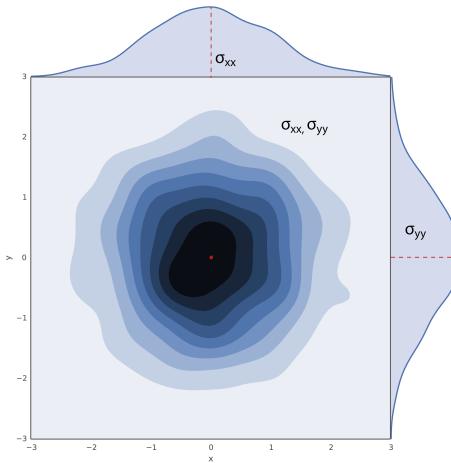


Figure 1.15 The marginal (1D) versus the bivariate (2D) distribution of the data.

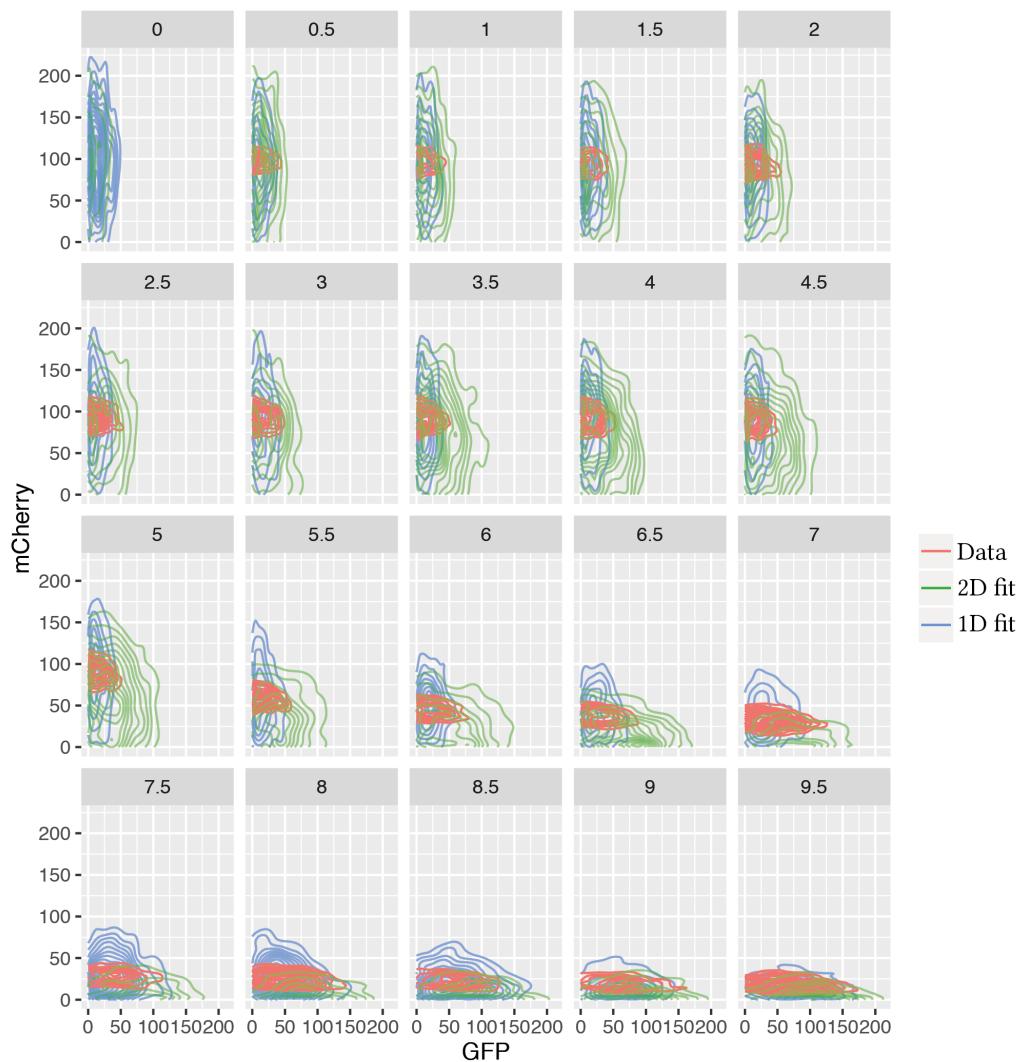


Figure 1.16 Comparing the fits obtained by using the 1D or 2D data.

The posterior distributions obtained from each fit are shown in Figure 1.17. We find similar posterior ranges for both the 1D and 2D fits.

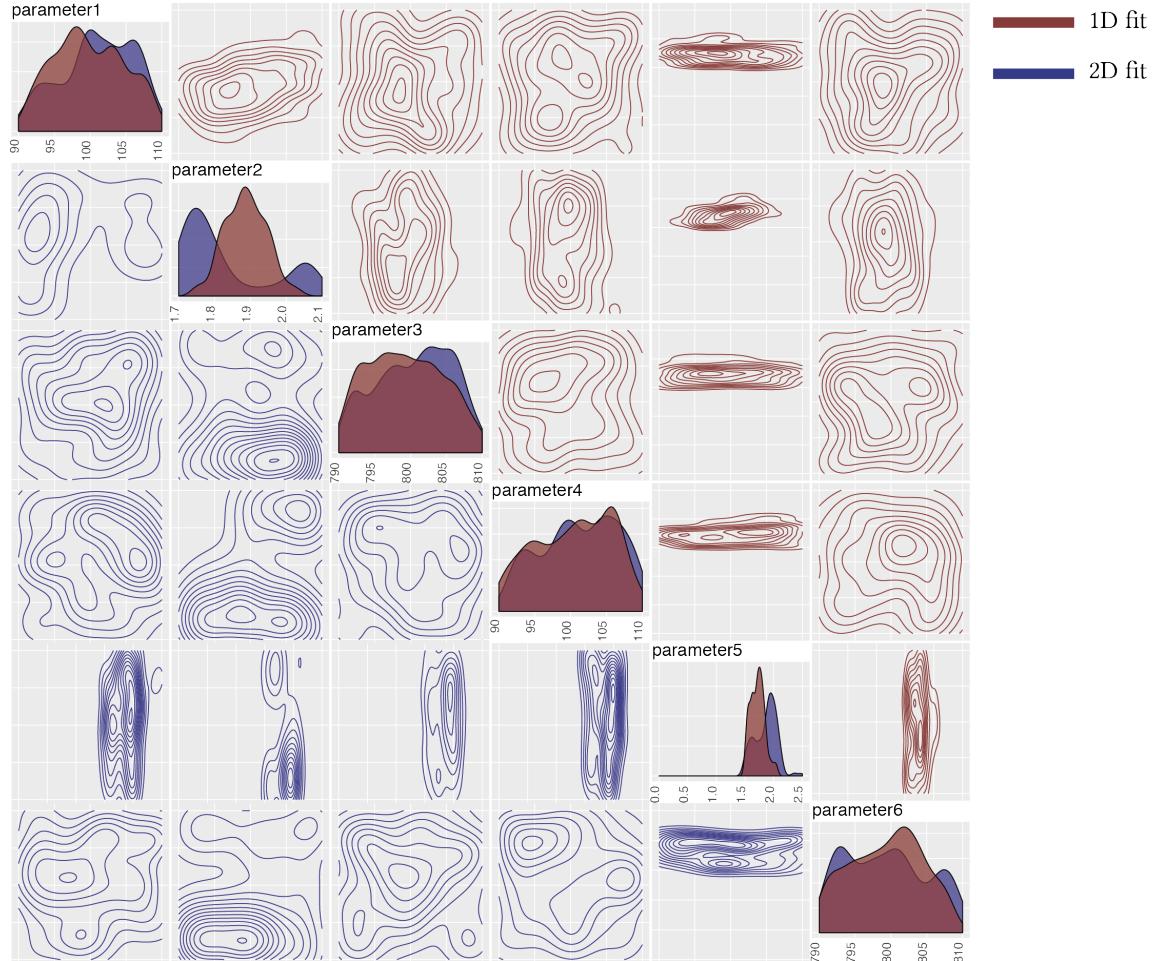


Figure 1.17 1D (red) 2D (blue) fit.

1.5.4.3 ABC-Flow used on experimental data

Next I applied ABC-Flow to the experimental flow cytometry data collected in Section 1.5.3. 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. I used the extension to the Gardner, Cantor, & Collins (2000) switch described in Equations 1.4-1.7.

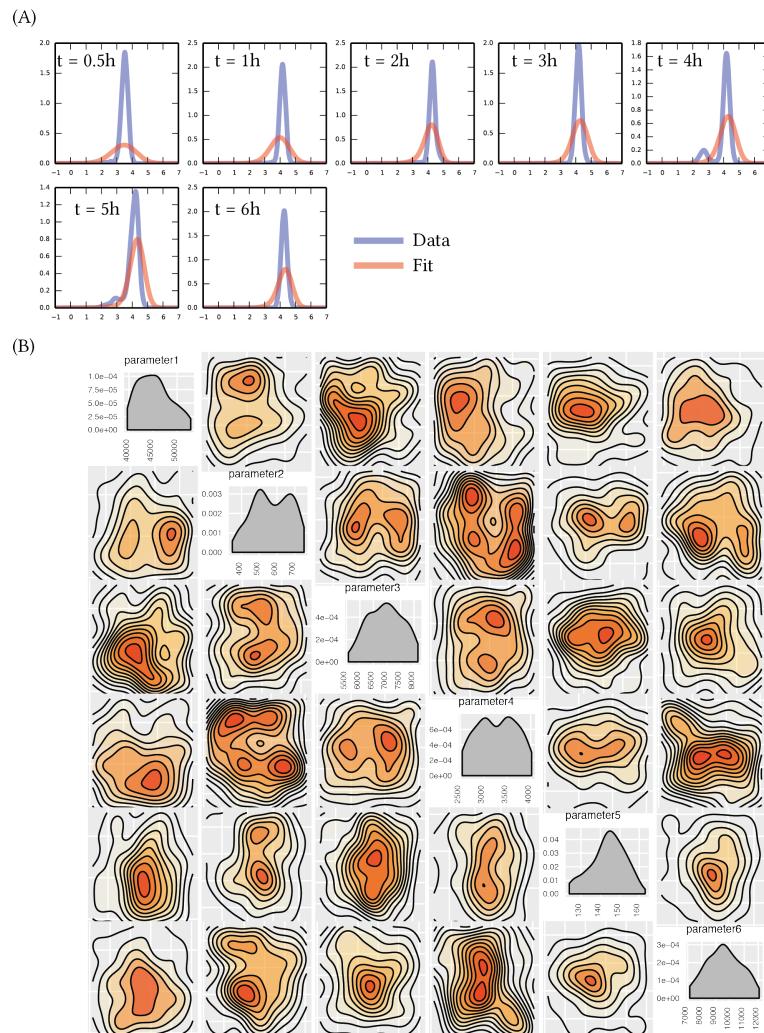


Figure 1.18

1.6 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.

1.7 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

- Attune NxT Acoustic Focusing Cytometer* (2015). CO016625.
- 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.
- 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.
- Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). ‘Construction of a genetic toggle switch in Escherichia coli’. *Nature* 403(6767), 339–342.
- 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.
- 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.
- 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.

40 BIBLIOGRAPHY

- 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.
- Peacock, J. A. (1983). ‘Two-dimensional goodness-of-fit testing in astronomy’. *Monthly Notices of the Royal Astronomical Society* 202(3), 615–627.
- 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.
- 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.
- Wilk, M. B. & Gnanadesikan, R. (1968). ‘Probability plotting methods for the analysis of data.’ *Biometrika* 55(1), 1–17.