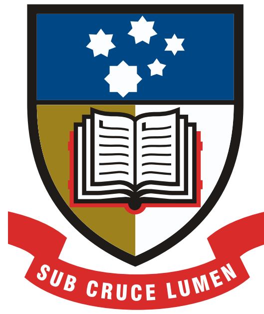

School of Biological Sciences
Discipline of Biochemistry

The Design, Synthesis and Quantitative Analysis of a Bistable Mixed Feedback Loop Gene Network

Julian Michael Juers Pietsch

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

February 2015

Abstract

Bistability, the capacity for switch-like memory, is a fundamental building block for robust behaviour in the noisy biochemical environment of a cell. Bistability has been observed experimentally in gene networks that exhibit overall positive feedback in some form; particular properties are endowed by variations on the basic network topology. The Mixed Feedback Loop (MFL) is a two-protein network that can be configured for positive feedback, and is notable since it has been observed to arise in nature more often than expected. The MFL includes an intervening protein-protein interaction to close a transcriptional feedback loop. This network architecture has been predicted to support bistable operation even without molecular cooperativity. To investigate the capabilities and features of the MFL, a synthetic bistable MFL was designed for construction in *Escherichia coli* (*E. coli*) using genetic components from bacteriophage 186. The design consists of the phage CI repressor protein inhibiting the production of its corresponding Tum antirepressor. This Tum−CI MFL prototype was first validated using a deterministic model expressly formulated for this instance of the MFL. It was then constructed in *E. coli* with dual LacZ and fluorescent reporters to permit multiple modes of measurement. Hysteresis assays — assays testing for history dependence or 'memory' of the system — were chosen as the measure of bistability, both since the bistable MFL naturally lends itself to such an assay, and since the assay simultaneously enables optimisation and setting of the switch. Measured by LacZ assay, the bistable MFL showed limited hysteresis. A detailed experimental characterisation of the network components and strains assisted in refining the data and setting bounds on model parameters. However, whilst this served to increase analytical accuracy, the deterministic model remained a poor fit of the data. When instead measuring activities in single cells by flow cytometry using the fluorescent reporter, two semi-stable sub-populations were discovered. Poor separation of the sub-populations necessitated the development of a system-specific mixture model for accurate identification of their characteristics, but the sub-population dynamics found much better agreement with the deterministic model. By building on this model with a hybrid stochastic/deterministic model, the limited hysteresis seen by LacZ assay can be explained by variation in switch robustness: the steady-state repressor concentration weights each cell's 'decision' for either of the two stable states. These results further an understanding of the core requirements for stable maintenance of epigenetic memory. The simplifications made by isolating the MFL according to the 'synthetic biology' approach allowed key features of this network motif to be determined. A deep knowledge of simple circuit structures like the MFL contributes fundamentally towards the way we understand proteins and how they fit into the complex networks that underpin the workings of life.

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Julian Pietsch Date

Acknowledgements

This thesis would not have been possible without the generous assistance, support and patience of many others.

A number of colleagues were instrumental in guiding the development of the theoretical models and analytical methods employed in this thesis. The stochastic models started their development under the guidance of Kim Sneppen at the Niels Bohr Institute in Copenhagen. Sandeep Krishna, also from the Niels Bohr Institute, provided many thoughtful discussions on modelling techniques, which helped me to refine and simplify the deterministic models. Ray Correll (originally CSIRO) provided useful advice on advanced statistical practice and methods of inference, particularly for the analysis of the bacterial growth curves.

Many, many thanks must go to my supervisors Keith Shearwin and Ian Dodd, who were always excited about my work, but equally inspiring for the way they tempered my inventive imagination with sound wisdom. Thank you for tirelessly reading through many lengthy thesis drafts and for giving me the opportunity to work on such an enjoyable project.

Many past and present members of the Shearwin lab provided valuable experimental tips, reagents when I was short, and a satisfying source of good conversation. Rachel Schubert and Linda Shearwin provided a number of foundational materials and results for the work on the Tum protein. Adam Palmer and Michael Pocock started the MFL project, providing a solid cornerstone upon which much of this thesis has been built. Barry Egan, the previous head of the lab, provided encouragement and an infectious enthusiasm for phage 186. Alexandra Ahlgren-Berg, Erin Cutts, David Priest, Danna Li, Cui Lun and Andrew Hao made sure that I always had someone to chat with and participated in many hours of discussion at the lab-ins I presented. Thank you for being there for me through the highs and lows of the thesis. Many thanks must also go to Iain Murchland, who kept me up to date with the affairs of the outside world, entertained me with plenty of cheerful banter whilst at the bench, and was always willing to hear through my (often fanciful) thoughts and ideas.

In the final throes of the thesis, Pierre Dumuid helped me to steer back on course and establish realistic goals. Many thanks for standing with me in the face of a mammoth task.

My siblings Antony and Hannah Pietsch, and Taria and Adam Schulze have always given me something to smile about and look forwards to. Thank you for all the love and support you have shown me over the past few years, and for so many laughs and dinners.

My mum and dad, Helen and Mike Pietsch, have spoiled me with their generous love and care for many, many years. Thank you for being such amazing parents.

I owe so very much to my wife Alia Pietsch. You are my inspiration, my light. And your patience and love always carried me through the PhD's darkest moments. Thank you for believing in me.

Contents

Contents i

List of Figures v

List of Tables viii

1 Introduction 1
1.1 Cellular networks drive cell behaviour . . . 2
1.2 Synthetic biology as a tool for studying network motifs . . . 4
1.2.1 Defining modules for rational circuit design . . . 5
1.2.2 Predictive models of cellular networks . . . 6
1.2.3 Bacteriophage 186: a source of new components for synthetic biology 8
1.3 The bistable MFL is an excellent candidate synthetic network . . . 10
1.4 Thesis overview . . . 12

2 Directing design of a bistable genetic circuit by mathematical modelling 15
2.1 Origin of bistability in the Tum−CI MFL . . . 16
2.2 Developing a mathematical model of the Tum−CI MFL . . . 17
2.2.1 Modelling the CI−*pR* interaction . . . 19
2.2.2 Modelling the Tum−CI interaction . . . 20
2.2.3 Deterministic free species model . . . 23
2.2.4 Deterministic total species model . . . 26
2.3 Steady-state analysis of the Tum−CI MFL model . . . 29
2.3.1 Solving the free species model at steady-state . . . 29
2.3.2 Varying the parameters . . . 34
2.4 Hysteretic behaviour . . . 38
2.4.1 Time course simulations of the Tum−CI MFL . . . 40
2.4.2 Simulating the hysteresis assay . . . 44

3 Designing and characterising a bistable Mixed Feedback Loop (MFL) 49
3.1 Designing and cloning the MFL strains . . . 51
3.1.1 Development of the preliminary Tum−CI MFL strains . . . 51
3.1.2 Introducing a fluorescence-based reporter module . . . 56
3.1.3 Shifting the range of CI expression levels . . . 59
3.2 Assaying hysteresis in the Tum−CI Mixed Feedback Loop . . . 61
3.3 Characterising the CI induction module . . . 66
3.4 Host strain characteristics . . . 77

3.4.1 Optical density measurements . . . 77
3.4.2 Growth rate . . . 79
3.4.3 Growth rates in alternative media . . . 84
3.4.4 Cell volume . . . 86
3.5 Balancing the MFL module . . . 87
3.5.1 Quantitating intracellular proteins . . . 87
3.5.2 Degradation rates of Tum and CI . . . 89
3.5.3 Production rate from pR . . . 95
3.5.4 Production rate from P_{lac} . . . 101
3.6 Chapter summary . . . 104

4 The MFL displays only weak bistability when measured over a whole population 105
4.1 Experimental limitations of the hysteretic LacZ assay . . . 105
4.1.1 Variations in optical density bias LacZ assay measurements . . . 106
4.1.2 Normalising P_{lac} induction levels to production rates improves but does not complete the picture of hysteresis . . . 108
4.2 Extending the hysteresis assay equilibration time . . . 112
4.2.1 Extending the time for equilibration brings the control strains to steady state . . . 112
4.2.2 Complete hysteresis is observed with a long equilibration time . . . 115
4.3 The deterministic model does not capture the behaviour of the MFL . . . 118
4.3.1 Searching the parameter space of the deterministic MFL model . . . 118
4.3.2 Fitting the WR-MC data set alone . . . 122
4.3.3 Fitting the combined data sets . . . 124
4.4 Chapter summary . . . 128

5 Hysteresis is obscured by stochastic switching between semi-stable states 129
5.1 Relating single-cell and whole-population measures of promoter activity . . . 130
5.1.1 Assaying gene circuit hysteresis by flow cytometry . . . 130
5.1.2 Choosing an appropriate data transformation . . . 132
5.1.3 An automated filter for selecting cell populations . . . 132
5.1.4 Morphology normalisation refines the distribution of fluorescence . . . 133
5.1.5 The curated data is suggestive of population mixing . . . 137
5.2 MFL samples within the bistable region are a mixture of two cell populations 138
5.3 The mixed population model reveals the stable states predicted for the MFL strains . . . 147
5.4 Noisy switching between sub-populations occurs throughout the hysteresis assay . . . 154
5.5 Chapter Summary . . . 158

6 Investigating noisy switching in the Tum−CI MFL by stochastic modelling 159
6.1 Stochastic modelling of gene networks . . . 160
6.2 A hybrid stochastic/deterministic model of the Tum−CI MFL . . . 167
6.3 Establishing a parameter regime for the hybrid stochastic/deterministic model 173
6.3.1 Fitting the dynamic deterministic model to a stochastic data set . . . 173
6.3.2 Optimising the magnitude of noise to reproduce observed rates of stochastic switching . . . 178

6.4 A simple stochastic model is sufficient to reproduce salient features of the Tum−CI MFL . . . 181
6.5 Improving and interpreting the stochastic model . . . 187

7 The MFL now and going forwards 189

8 Materials and Methods 193
8.1 Reagents . . . 193
8.2 General cloning methods . . . 196
8.2.1 Growth of bacteria . . . 196
8.2.2 Storage of bacterial strains . . . 197
8.2.3 Preparation and purification of DNA . . . 197
8.2.4 Polymerase Chain Reactions . . . 198
8.2.5 Analysis of DNA . . . 198
8.2.6 DNA recombination work . . . 199
8.2.7 Competent cells . . . 199
8.2.8 Sequencing . . . 200
8.2.9 Changing resistance genes for the *pR-tum* plasmid . . . 201
8.3 Strains and DNA . . . 204
8.3.1 Bacterial strains . . . 204
8.3.2 Primers . . . 205
8.3.3 Plasmids . . . 206
8.4 Assays . . . 209
8.4.1 Preparation of cell extracts . . . 209
8.4.2 Polyacrylamide gel electrophoresis of proteins and Western blotting 211
8.4.3 Quantitating concentrations of cells in culture . . . 212
8.4.4 Growth of bacteria for 96-well plate assays . . . 212
8.4.5 LacZ assay . . . 213
8.4.6 Flow cytometry . . . 214

9 Structure-function studies for Tum 217

A Fitting growth curves 251
A.1 Log-linear fits . . . 252
A.2 Gompertz fits . . . 254
A.3 Comparing the models . . . 255
A.4 Growth rate measurements . . . 257

B Scripts for analysis of flow cytometry data 259
B.1 General utility functions . . . 260
B.2 Automated selection of the main cell population . . . 262
B.3 Logicle transformation . . . 263
B.4 Morphology normalisation . . . 264
B.5 Constrained skew-*t* regression . . . 270

C Fitting deterministic time-course models to the Tum−CI MFL data 279
C.1 Deterministic simulation of the MFL in R . . . 279
C.2 Fitting the model to the LacZ assays . . . 291
C.2.1 Loading and curating the data set . . . 291

C.2.2 Setting up the model output 295
C.2.3 Defining the cost function and optimising parameters 298
C.3 Fitting the model to the flow cytometry assays 299

D The hybrid stochastic/deterministic model of the Tum−CI MFL 309
D.1 Tracking simulator state 310
D.2 The generic simulation framework 316
D.3 Classes for simulation of the MFL 324
D.3.1 mflLibrary.h 324
D.3.2 equilibration.h 334
D.3.3 models.h 345
D.4 Running the simulator 352
D.5 Adding experimental noise to stochastic simulations 359

Bibliography 363

List of Figures

1.1 Common cellular network motifs. 2
1.2 A diagrammatic representation of the bacteriophage 186 genome. 9
1.3 Gene regulatory networks that can exhibit bistability. 11

2.1 Circuit diagram for the bistable mixed feedback loop. 15
2.2 Stable states of the Tum−CI MFL. 16
2.3 Tum-CI MFL model parameters. 18
2.4 Fitting parameters for the Tum-CI interaction. 21
2.5 Bistability in the Tum−CI MFL arises as a result of the sigmoidal response of Tum production as a function of total Tum concentration. 28
2.6 Stable points of the MFL 32
2.7 Equilibrium solutions for the Tum–CI MFL as a function of total CI steady-state concentration. 33
2.8 Variation in bistable region location as a function of equilibrium parameters. . . 35
2.9 Observing variation in bistable region location as a function of production and degradation rate parameters 37
2.10 Qualitative description of hysteresis in the Tum–CI MFL 39
2.11 Determinstic time course simulations of the Tum–CI MFL 43
2.12 Determinstic hysteresis loop simulations of the Tum–CI MFL 45
2.13 Equilibration times near the points of bifurcation 47

3.1 Tum–CI MFL strain design. 50
3.2 Sequence maps for the *pR-lacZ* and *pR-tum* MFL modules. 54
3.3 Sequence maps of the plasmids used for introducing a fluorescent reporter to the MFL. 58
3.4 Sequence maps of the CI expression plasmids. 59
3.5 Comparing repression of the *pR* promoter by CI expression plasmids that utilise alternative *c*I RBSs. 60
3.6 The Tum–CI MFL shows hysteresis. 63
3.7 Sequence maps illustrating plasmid precursors to the IPTG induction reporter plasmid. 68
3.8 Induction of the P_{lac} promoter has an ultrasensitive dependence on the concentration of IPTG inducer in MFL-like strains. 70
3.9 Heteroscedasticity is reduced by using a Box-Cox transformation prior to fitting the P_{lac} induction reporter assay data with Hill curves. 72

3.10 Comparing P_{lac} promoter induction under different assay conditions in MFL-like strains. 73
3.11 Scaling the P_{lac} induction curves measured by LacZ assay to the equivalent steady-state CI concentrations. 76
3.12 Calibrating absorbance measurements at 620 nm, for cultures grown in M9 minimal media in 96-well plates, to standard optical densities at 600 nm. 78
3.13 Comparing log-linear and Gompertz fits of growth curves. 81
3.14 Growth curves for a MFL strain grown in alternative growth media. 85
3.15 Fitting the sigmoidal response of band intensity to TumHis_6 mass. 88
3.16 Using Box-Cox transformation of Western blot band intensities to derive a quantity with linear dependence on Tum mass. 88
3.17 Following Tum degradation by Western blot. 90
3.18 Measuring degradation of Tum and CI. 91
3.19 Comparing soluble and insoluble fractions of Tum at intial and final time points of the degradation assays. 92
3.20 The degradation-resistant fraction of Tum appears to be a subset of the insoluble fraction. 93
3.21 The fraction of soluble Tum appears to saturate as the production rate is increased. 94
3.22 Western blots for quantitating steady-state production of Tum from *pR*. 97
3.23 Calibrating Tum-specific band intensities on Western blots with Tum mass. . . 97
3.24 Estimates of the mass of Tum from extracts of MFL strains with pMTS-*pR*-tum^+ (pMTS-*pR*-tum^+) but without the *c*I gene. 98
3.25 Quantitating steady-state CI production from P_{lac} by Western blot for induction at 300 μM IPTG. 102
3.26 Estimating the mass of CI in the wRBS and eRBS extracts. 103
3.27 Comparing steady-state estimates of CI concentration for the MFL strains with those previously obtained for the same induction system. 104

4.1 Expressing the normal equilibration time hysteresis curves in terms of CI production rate reveals a wider putative region of bistability. 111
4.2 Extending the time for equilibration allows the control curves to reach equilibrium. 113
4.3 Complete hysteresis is observed when using the long equilibration time assay at the cost of loop collapse. 116
4.4 The steady-state model of bistability does not compare well with the long equilibration time assays. 120
4.5 The deterministic Tum−CI MFL model cannot match all features of the experimental WR-MC MFL hysteresis assays. 123
4.6 The whole-population WR-MC and ER-MC data sets can be matched up, but the additional data does not improve the model fit. 125

5.1 The bacterial cell population is easily identified using the forward and side scatter intensities. 133
5.2 The resolution between low and high fluorescence populations is poor. 134
5.3 The mean fluorescence of the cell populations overlaps well with the mean *pR* activity measured by LacZ assay. 136
5.4 Viewed as cell populations, hysteresis in the Tum−CI MFL is manifest as a history-dependent broadening of fluorescence. 137

5.5 The skew-t distribution provides a good fit of population fluorescence for the MFL control strains. 141
5.6 The interquartile range of intensity for the MFL control distributions varies as a function of the median intensity. 142
5.7 The fitted skew-t parameters of the MFL controls vary as functions of the median and interquartile range of intensity. 144
5.8 MFL distributions in the bistable region are a bimodal mixture of control-like sub-populations. 145
5.9 Bimodal cell populations are found in each of the Tum−CI MFL strains. 151
5.10 The results obtained by flow cytometry can be interpreted in terms of the steady-state deterministic model. 152
5.11 Bimodal populations are found at all the equilibration time points tested for the ER-MC Tum−CI MFL. 156

6.1 Illustrating the Gillespie algorithm. 162
6.2 Noisiness in the hybrid stochastic/deterministic model of the Tum$^-$ control strain. 172
6.3 The deterministic model matches the long equilibration time data in spite of a simplified parameter set. 177
6.4 The level of noise in CI production affects rates of stochastic switching differently to that for Tum production. 180
6.5 The hybrid stochastic/deterministic model reproduces the broad regions of bimodality observed experimentally. 182
6.6 The timing of stochastic switching in the hybrid stochastic/deterministic model of the Tum−CI MFL. 184
6.7 Sub-populations fitted to the simulated Tum−CI MFL data validate the semi-deterministic interpretation. 186

8.1 1.5% agarose gel of diagnostic digest of pMTS-*pR-tum*$^+$ plasmid with AatII, XhoI and NdeI. 202
8.2 1.5% agarose gel of diagnostic digest of *pR-tum* plasmids with NdeI/XhoI. 203
8.3 2.0% agarose gel of diagnostic digests to check for the correct origin in the *pR-tum* plasmids. 204

A.1 Well-to-well variations in 96-well plates are correlated over time. 251
A.2 Fitting the 'lacz2' data using alternative GLMs. 253
A.3 The Gompertz curve. 254
A.4 Comparing growth rate determinations between the various models of growth for the MFL strains. 256
A.5 The 'wrmc2' plate deviates significantly due to fitting biases introduced by the selection of time points. 256
A.6 MFL strain doubling times show little dependence on IPTG or assay conditions. 257
A.7 Induction reporter strain doubling times show little dependence on IPTG or assay conditions. 258

C.1 Deterministic fits of the ER-MCTum−CI MFL flow cytometry hysteresis assay. 305

List of Tables

3.1 Parameters determined for Hill fits of induction. 74
3.2 Doubling times measured for notable strains within the present thesis. 82
3.3 Growth rates measured for an MFL strain in alternative growth media. 85
3.4 Tabulating the strengths of the *pR* and P_{lac} promoters in the MFL strains. . . . 99

4.1 Comparing parameters measured experimentally or obtained from the literature with those determined by fitting the deterministic model. 127

6.1 Stochastic reactions in the Tum−CI MFL model. 170
6.2 A comparison of experimental parameter estimates and the deterministic parameters fitted to the time-course flow cytometry data. 176

8.1 Standard chemicals used in this thesis. 193
8.2 Standard buffers and growth media used in this thesis. Buffers were prepared in Milli-Q water (H_2O) unless otherwise specified. 195
8.3 Concentrations of antibiotics used in this thesis. 197
8.4 Bacterial strains used in this thesis. 204
8.5 Primers used in the course of this thesis. 205
8.6 Plasmids referred to and cloned in the course of this thesis. 206

C.1 A comparison of the deterministic parameters fitted to various subsets of the time-course flow cytometry data. 307

List of Boxes

4.1 Correcting for the dependence of LacZ units on optical density (OD_{600}) reduces variability and increases separation between the hysteresis curves. . . 107
5.1 The skew-*t* distribution. 140