Signalling and Crosstalk in Cytokine Pathways: Mathematical Modelling and Quantitative Analysis

Shabnam Khatibi

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Department of Electrical and Electronic Engineering
The University of Melbourne
April 2016

Produced on archival quality paper
Cancer is a leading cause of death all over the world. Focusing at the intracellular level, there are several cytokine signalling networks involved in inflammation and tumorigenesis. These pathways regulate the cell biological responses to the environment, both directly and via crosstalk. Recently, there have been several reports which have confirmed the strong relationship between inflammatory diseases and tumor development ([82, 110, 163], see [294] for a review). Two cytokines, in particular, have significant roles in wound healing, inflammation and cancer: Transforming growth factor β (TGF−β) and Interleukin-6 (IL−6). Cytokine signalling pathways are an interconnected complex system of biochemical reactions which can be represented by kinetic equations. These signalling pathways can share proteins and genes which make the intracellular signalling networks extremely complex.

Systems biology and mathematical modelling are new approaches for the study of complex systems such as intracellular signalling networks. The focus of this research is to model mathematically new descriptions of the TGF−β and IL−6 pathways based on new logistics and then integrate them into a single, robust, self-regulated model which can be used to investigate tumor development in the stomach and colon. At each level, experimental data sets were used iteratively in order to both parameterize and examine the models (e.g. the model and experimental data of Zi et al. (2011) and further model simulations have been used in Chapter 4 to parameterize our TGF−β model.).

Different approaches in Systems biology and their applications in cell signalling research are studied. TGF−β and IL−6 signalling pathways and their components are reviewed next. The previous mathematical models of TGF−β and IL−6 signalling are briefly discussed. Additionally, the role of individual signalling in cancer progression and inflammation is studied.

We developed a mathematical model which captures the details of TGF−β signalling. The detailed model consists of over 40 differential equations and highlights the necessity for the reduction and simplification methods. The TGF−β signalling model is simplified and reduced via analytical reduction methods to 6 differential equations and is further validated with experiment. For the first time an explicit negative feedback loop has been included in the model. Another contribution of the TGF−β model is that the inherent time-delays in signalling networks are incorporated in detail. In the final chapter different input patterns are studied for TGF−β signalling. Our model of TGF−β signalling
indicates that the positive feedback loop is one mechanism by which stability could be achieved. The thesis reports for the first time the coupling of the positive and negative feedback loops for TGF−β signal transduction. Furthermore, our TGF−β signalling model proposes predictions for the responses of cancer cells to TGF−β stimulation, which suggest new experimental protocols for future work.

We also developed a mathematical model that describes the IL−6 signalling system thoroughly. The large number of equations involved in this model highlights the need for simplification. Similar to TGF−β signalling model, IL−6 signalling model is simplified and reduced using mathematical methods. In order to develop a realistic model specific kinetics are used for the different reactions. Time-delays are incorporated in the IL−6 transduction mathematical model for the first time. After being validated with different experimental data, the reduced IL−6 signalling model predicts the behaviour of cancer cells in response to IL−6 stimulation. Different pulsatile ligand inputs are studied using IL−6 model and new hypotheses for the TGF−β and IL−6 signalling crosstalk are raised.

Our initial hypothesis that IL−6 signalling regulates TGF−β signalling via SMAD7, is examined using our integrated IL−6:TGF−β model. The simulations produced by the integrated model confirm the importance of the negative feedback loop of TGF−β signalling (SMAD7) via IL−6 downstream signalling, previously suggested by Jenkins et al. (2005). Various kinetic models are examined for the link between the two signalling pathways and several predictions are proposed for the pulsatile inputs and different stimulation patterns. The results of the integrated model are compared with the individual TGF−β and IL−6 models. IL−6-induced activation of SMAD7 leads to suppression of TGF−β signalling and causes double peak responses of PSMAD2 in the short-term, however, the long-term responses of the cells to TGF−β stimulation remain unchanged by IL−6 signalling. The integrated model is also validated experimentally. Conclusively, we found that the regulation of TGF−β signal transduction by IL−6 signalling occurs within the first 300 minutes after stimulation, i.e. within the transient phase of the response.

This thesis includes both theoretical and experimental work, performed by the applicant. Theoretical part of the thesis consists of designing and developing models, analytical analysis of the models and conducting numerical simulations with the numerical simulation of the models. In the experimental part, various experimental protocols were developed and examined in order to parameterize, test and validate the proposed models.
Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD except where indicated,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed,

Shabnam Khatibi
8 April 2016
Acknowledgements

I would like to express my sincere gratitude to my supervisors Professor Tony Burgess, without whom this dissertation would not have been possible, Dr. John Wagner for his wonderful guidance throughout my Ph.D. and Professor Jonathan Manton for his supports.

I also want to thank Dr. Hong-Jian Zhu for his continues assistance and invaluable advices during last four years. Additionally, I wish to acknowledge the help provided by Burgess lab members from whom I have learnt many laboratory techniques.

I take this opportunity to express gratitude to all of my friends and colleagues who have put smile on my face with their love, kindness and support.

A special thanks to my family who have always stood by me and encouraged me to pursue my dreams.
Contents

Table of Contents vii

List of Figures x

List of Tables xv

1 Introduction 1
  1.1 Motivation ........................................ 1
  1.2 Research Objectives ................................ 2
  1.3 Thesis Outline .................................... 3

2 Literature Review and Background 5
  2.0 Summary of the information included in Chapter 2 .... 5
  2.1 Systems Biology .................................... 5
  2.2 TGF−β Signalling Pathway .......................... 8
    2.2.1 Key Components of TGF−β Signalling .......... 10
    2.2.2 The Role of TGF−β Signalling in Cancer ...... 16
    2.2.3 Mathematical Approach to Modelling TGF−β Signalling . 17
  2.3 IL−6 Signalling Pathway ............................ 19
    2.3.1 Gp130 Membrane Co-receptor .................. 21
    2.3.2 JAK/STAT Pathway .............................. 22
    2.3.3 MAPK Pathway .................................. 24
    2.3.4 Inhibitory Feedback of SOCS3 ................... 24
    2.3.5 IL−6 and Cancer .................................. 25
    2.3.6 Crosstalk ......................................... 25
    2.3.7 Mathematical Approach to the Modelling IL−6 Signalling 26
  2.4 Summary ............................................. 27

3 Methods 29
  3.1 Measurement of Key Signalling Components ............ 29
    3.1.1 Cell Culture and Cell Lysis ...................... 29
    3.1.2 Western Blotting ................................ 31
3.1.3 Protein Quantitation ........................................... 32
3.2 Computer Modelling ............................................ 33
3.3 Design and Modelling Methods ................................. 33
3.3.1 Mathematical Modelling of the Intracellular Protein-Protein
Reactions .............................................................. 33
3.3.2 Reduction, Scaling and Non-dimensionalization Methods 35
3.4 Validation Methods .............................................. 40
3.4.1 Quantitation Methods .......................................... 40
3.4.2 Estimation and Curve Fitting Methods ....................... 40

4 TGF-β Receptor Model of Signal Switching in Cancer 41
4.0 Summary of the information included in Chapter 4 .......... 41
4.1 Introduction .......................................................... 42
4.2 Model Development .............................................. 44
4.2.1 The Full Model .................................................. 45
4.2.2 The Simplified Model ......................................... 48
4.3 Numerical Simulations ............................................. 51
4.4 Discussion ............................................................ 59
4.5 Materials and Methods ........................................... 60
4.6 Supplementary Material .......................................... 62
4.6.1 The Reduced Model ............................................ 62
4.6.2 Model Parameters ............................................. 69
4.6.3 Figures .............................................................. 73
4.6.4 Steady-state Analysis of the Simplified Model ............... 73

5 Modelling the IL-6 Signal Transduction System 76
5.0 Summary of the information included in Chapter 5 .......... 76
5.1 Introduction .......................................................... 76
5.2 Model Development .............................................. 78
5.3 Analysis ............................................................... 82
5.3.1 Reduction Implementation on the Receptor Part ............ 82
5.3.2 Modelling of the Nucleocytoplasmic Reactions ............. 84
5.3.3 Simplified Model ............................................... 85
5.4 Numerical Simulations and Validation .......................... 88
5.5 Discussion ............................................................ 95
5.6 Material and Methods ........................................... 96

6 TGF-β and IL-6 Signalling Crosstalk: an integrated model 98
6.0 Summary of the information included in Chapter 6 .......... 98
6.1 Introduction .......................................................... 98
6.2 Model Development ........................................ 100
6.3 Results from IL − 6:TGF−β Crosstalk Model and Discussion ... 103
6.4 Material and Methods ..................................... 110
6.5 Summary and Conclusions .................................. 112
6.6 Overall Conclusions and Future Work ...................... 113

7 Appendices ...................................................... 115
7.1 A Differential Equation Overview .......................... 115
  7.1.1 Initial Conditions and Solutions ....................... 115
  7.1.2 Ordinary VS. Partial .................................... 115
  7.1.3 Order of Differential Equations; Autonomous and Non-
          autonomous Differential Equations .................... 116
  7.1.4 Linearity VS. Non-linearity ............................. 116
  7.1.5 Systems of Differential Equations ........................ 116
  7.1.6 Qualitative Analysis; Fixed Points, Equilibrium Points and
          Stability ................................................. 118
  7.1.7 Linearization of Non-linear ODEs ....................... 118
7.2 ODE Solving Methods; Numerical Consideration .............. 119
  7.2.1 Analytical Solutions ................................... 120
  7.2.2 Numerical Solutions .................................. 121
7.3 Review of the Mathematical Models on TGF−β Signalling .... 129
7.4 Equations Describing the TGF−β and IL − 6 Signalling Systems . 145
  7.4.1 TGF−β Receptor Model ................................. 145
  7.4.2 Cytoplasmic SMAD Model ............................... 146
  7.4.3 Nuclear SMAD Model ................................... 147
  7.4.4 IL-6 Receptor Model .................................. 148
7.5 Review of the Positive Feedback Loop in TGF−β Signalling .... 149
  7.5.1 Preliminary Experimental Data Integration .............. 151

Bibliography ..................................................... 153
List of Figures

2.1 The sketch of TGF−β signalling network. Potential phosphorylation sites of the receptors are specified with empty circles attached to membrane receptor components. Arrows pointing to 6 blue dots represent degradation process. The red solid arrows originating from SMAD7:Smurf1/2 apply negative feedback on the receptor components. The dot arrows represent the transcription and translation reactions. Note that the logistics of the signalling are expanded in Figure 4.1. 10

2.2 The sketch of IL−6:Gp130:JAK:STAT signalling. Potential phosphorylation sites of the receptors are specified with empty circles attached to membrane receptor components. The dot arrows represent the transcription and translation reactions. Note that the logistics of the signalling are expanded in Figure 5.1. 21

3.1 A reversible biochemical reaction which shows the binding of two proteins. $M_1$ and $M_2$ bind together to produce $D$ with the rate of $k^+$. Similarly, $k^−$ specifies the rate of dissociation of $D$ to $M_1$ and $M_2$. 34

4.1 The full TGF−β signalling biological model. Potential phosphorylation sites of the receptors are specified with empty circles attached to R1 and R2 components. Arrows pointing to 6 blue dots represent degradation process. The red solid arrows originating from SMAD7:Smurf apply negative and/or positive feedback on the receptor components of the membrane. Oval-shaped components written in small letters represent micro-RNAs. In this figure, $S$ represents the SMAD proteins. Note that the arrow from ODC to polyamine shows a stimulatory reaction rather than conversion. 45
4.2 The full receptor model, TGF-β signal transduction. The red dashed lines which originate from phosphorylated SMAD trimer indirectly regulate the receptor levels. All the reactions from trimerization of phospho-SMAD2/3 to SMAD7 transcription and translation are reduced to the red dashed lines (see Figure 4.1 for clarification). The dotted ends of red dashed lines show that included reactions could lead to both inhibition and stimulation of their targeting reactions (demonstrating negative and positive feedback effects). In this figure S is specifically used for SMAD2/3.

4.3 TGF-β receptor signalling system. A) The schematic receptor model, TGF-β signal transduction. TGF and $\hat{S} + 3S_3$ represent the input and the output of the model, respectively. B) A Simplified Model of TGF-β signal transduction. TGF-β and $\hat{S} + S_n + 3S_3$ represent the input and the output of the model, respectively.

4.4 Total PSMAD time course for TGF-β = 5. The peak in the total PSMAD concentration occurs 50-60 min after the stimulation and corresponds to the short-term (transient) response. The constant level of PSMAD at 0.3 represents the long-term (steady-state) response of the system.

4.5 Short-term responses of PSMAD levels to different concentrations of TGF-β; transient response. The simulation time for each point in this figure is 50 min. The only parameter of the model which is being changed in this figure is the TGF-β concentration, meaning that each point has a different TGF-β input.

4.6 Long-term responses of PSMAD levels to different concentrations of TGF-β; steady-state response. The simulation time for each point in this figure is 1000 min. The only parameter of the model which is being changed in this figure is the TGF-β concentration, meaning that each point has a different TGF-β input.

4.7 Total PSMAD time course for a certain TGF-β concentration. A) simulation results for low membrane receptor concentration condition (or so called early-stage tumors) B) simulation results for high membrane receptor concentration condition (or so called late-stage tumors). These conditions were simulated via altering the receptor production rate on the membrane.

4.8 The effects of receptor concentration on the long-term response of PSMAD. A) TGF-β = 5. B) TGF-β = 2. Approximately, no difference is observed between the two panels of this figure.
4.9 The effects of SMAD concentration on the long-term response of PSMAD. A) TGF–β = 5 B) TGF–β = 2. The steady-state level of total PSMAD rises higher in Figure 4.9A than in Figure 4.9B.

4.10 The validation of the simplified model with experimental data. The dots show the level of PSMAD2 concentration obtained from experiment and the curves specify the model predictions. A) PSMAD2 time course for 0-1h on SV40-immortalized MEFs cell line stimulated with TGF–β and its corresponding blot B) PSMAD2 time course for 0-4h on SV40-immortalized MEFs cell line stimulated with TGF–β and its corresponding blot C) PSMAD2 time course for 0-4h on wild type MEFs stimulated with TGF–β and its corresponding blot.

4.11 PSMAD2 time course validation with experimental data sets from wild type MEFs. Different dot colors specify different experiments. The curve shows the model prediction of PSMAD2 dynamics.

4.12 PSMAD2 time course validation with experimental data sets from Gp130F/F MEFs. The SMAD7 level is higher in Gp130F/F MEFs and the PSMAD2 steady-state level is lower compared to wild type MEFs. Different dot colors specify different experiments. The curve shows the model prediction of PSMAD2 dynamics.

4.13 Steady-state analysis of the reduced model in the absence of feedback. A) The phase plane plots of the reduced receptor system at equilibrium for different values of c. B) The scaled total PSMAD trimer, s3, value against different values of c, where log c = 0 corresponds to c = 1.

4.14 The steady-state of the receptor system for different range of parameters. A) ε = 0.5 δ = 0(0.2)1 In these panels A,B and C c = 2^n for n = -2(1)2. B) δ = 0(0.2)1, In this panel ε changes as labeled.

4.15 The predicted effects of different concentrations of TGF–β on the short-term responses of total PSMAD when the negative feedback influences LC only.

4.16 The predicted effects of different concentrations of TGF–β on the long-term responses of total PSMAD when the negative feedback influences LC only.

5.1 Topology of the IL–6 signalling pathway model. Potential phosphorylation sites of the components are specified with empty circles. Arrows pointing to 6 blue dots represent degradation process. Oval-shaped components represent micro-RNAs. The dot arrows represent the transcription and translation reactions.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>A reduced model representing the IL-6 signal transduction network. Arrows pointing to 6 blue dots represent degradation process. The stars represent for the phosphorylated species. The dot arrows represent the transcription and translation reactions.</td>
</tr>
<tr>
<td>5.3</td>
<td>Simplified model of IL-6 signal transduction. The dashed line specifies an enzymic reaction. Arrows pointing to 6 blue dots represent degradation process. The stars represent for the phosphorylated species. The dot arrows represent the transcription and translation reactions.</td>
</tr>
<tr>
<td>5.4</td>
<td>Total STAT3 concentration in liver extracts after IL-6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves represent the fitted curve for the experimental data and the predicted values from the model, respectively. The fitted curve is plotted via trendline tool of Microsoft Excel.</td>
</tr>
<tr>
<td>5.5</td>
<td>Total STAT3 concentration time course in liver extracts after IL-6 stimulation in log timescale. The red curve is the simulation result. The figure suggests a difference between the initial level of total STAT3 and its steady-state level. The fitted curve is plotted via trendline tool of Microsoft Excel.</td>
</tr>
<tr>
<td>5.6</td>
<td>Total PSTAT3 concentration in liver extracts after IL-6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves demonstrate the fitted curve for the experimental data and the predicted values from the model, respectively. The fitted curve is plotted via trendline tool of Microsoft Excel.</td>
</tr>
<tr>
<td>5.7</td>
<td>SOCS3 concentration in liver extracts after IL-6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves show the fitted curve for the experimental data and the predicted values from the model, respectively.</td>
</tr>
<tr>
<td>5.8</td>
<td>SMAD7 concentration time course for linear and non-linear stimulation via nuclear PSTAT3.</td>
</tr>
<tr>
<td>5.9</td>
<td>SMAD7 concentration time course dynamics for IL-6 concentration input = 0.5, 100 and 400 (the concentration units are arbitrary).</td>
</tr>
<tr>
<td>5.10</td>
<td>Short-time simulation of total PSTAT3 dynamics. The blue curve shows the simulation results and the dots represent the relative experimental data for 0-1 h IL-6 stimulation.</td>
</tr>
<tr>
<td>5.11</td>
<td>Experimental and simulation results for long-time IL-6 stimulation. Different colored dots show different experiments.</td>
</tr>
<tr>
<td>5.12</td>
<td>PSTAT3 responses of IL-6 signalling system to a pulsatile input. Different curves correspond to distinct pulsation patterns.</td>
</tr>
</tbody>
</table>
5.13 PSTAT3 responses of IL−6 signalling system to different levels of membrane receptors. "m" represents the level of receptor concentrations on the membrane. .................................................. 94

6.1 The schematic diagram of TGF−β and IL−6 crosstalk. the red arrow is the focus of the current thesis. .......................................................... 100

6.2 A comparison between the PSMAD2 response of the TGF−β signalling system and the integrated IL−6:TGF−β signalling system. The integrated model output follows the TGF−β model output for the first 60 minutes, however it diverges afterwards. The steady-state PSMAD2 levels of both models are identical. ........................................ 104

6.3 The integrated model output for the double stimulation and pre-stimulation scenarios of IL−6:TGF−β signalling. ........................................ 104

6.4 PSMAD2 responses of the integrated model to different pulsatile inputs of IL−6. The TGF−β level is kept constant for the entire duration of the simulation. .............................................................. 105

6.5 The simplified TGF−β model response (proposed in Chapter 4) to pulsatile TGF−β input. ............................................................... 106

6.6 The integrated model response to pulsatile TGF−β stimulation. ........ 106

6.7 SMAD7 dynamics when the integrated model is stimulated with constant levels of IL−6 and TGF−β. .................................................. 107

6.8 Two models of the SMAD7 dynamics when the integrated model is pre-stimulated with IL−6 for 2 hours and with TGF−β an hour after IL−6 stimulation. .................................................. 107

6.9 The variation in the integrated model output as the PSTAT3/SMAD7 interaction is being reinforced gradually. Different coefficients represent different strength of the PSTAT3-induced stimulation of SMAD7. ... 108

6.10 The experimental validation of the integrated model. The blue curve is the model prediction of the double-stimulation of TGF−β and IL−6. The colored dots represent the experimental data from wild type MEFs. Each set of experimental data is specified with a color. The "cross"es represent the average of the 3 experimental replicates. ......................... 109

6.11 The representative blots of Figure 6.10. The top bands show the total PSMAD2 concentration level of the double-stimulation experiment. The bottom bands represent the loading control to which the PSMAD2 levels are normalised. .................................................. 109

7.1 The simplified cartoon of the feedbacks in TGF−β signal transduction system .................................................. 150
7.2  A) Western blot analysis of SV40-immortalized MEF cells (immortal MEFs) stimulated by TGF–β for different times, probed with PSMAD2 and β-tubulin anti-bodies. The results indicate that total PSMAD2 levels peak after 60 minutes of TGF–β stimulation. B) Western blot analysis of wild type MEF cells (MEF Gp130F/F) stimulated with TGF–β for different duration, probed with PSMAD2 and β-tubulin anti-bodies. The results indicate that total PSMAD2 levels peaks after 20 minutes of TGF–β stimulation. .......................................................... 151

7.3  Western blot analysis of wild type MEF cells (MEF Gp130F/F) stimulated with IL – 6 (10% conditioned medium) for different duration, probed with PSTAT3 and β-tubulin anti-bodies. Results indicate that total PSTAT3 levels rise gradually until 60 minutes of IL – 6 stimulation. 152
List of Tables


4.1 Components of the TGF–β receptor signalling model

4.2 Kinetic rates and binding constants of the model

4.3 The parameter values of simplified TGF–β model

5.1 List of the components of the IL–6 signalling model shown in Figure 5.1 and the standard initial conditions for the model which were constructed from the literature

5.2 Differential equations used to model the reactions of the IL–6 signalling pathway shown in Figure 5.2

5.3 Separation of the intracellular reactions based on their timescale

5.4 Kinetic rates and binding constants of the model

5.5 The parameter values of simplified IL–6 model

7.1 Comparison of all the existing models for TGF–β transduction in terms of included components and their values
Chapter 1

Introduction

1.1 Motivation

Statistics document an increasing rate of cancer incidence worldwide [368,399]. According to cancer council Australia, 1 in 3 women and 1 in 2 men are diagnosed with cancer before the age of 85 [368]. This makes cancer a major health problem for the majority of countries. The most common types of cancer in Australia are prostate, bowel, breast, melanoma and lung cancer respectively (cancer council NSW annual report). As a result, research has focused on understanding cancer and developing specific treatments.

Inflammation is reported as one of the conditions that is closely related to different cancer types (see [56,341,371] for reviews). Although it is not clear whether inflammation results from cancer or is a cancer initiator, considerable evidence confirms the links between chronic inflammation and tumor development [22,23,33,377,391]. As a result, it is essential to consider inflammatory components when investigating cancer initiators, promoter and progression.

In attempting to understand cancer, the key issue is its emerging complexity [188]. Cancer is a multi-scale, multi-variable and complicated disease that involves deregulation of distinct, connecting pathways which control fundamental cell responses [259]. One way to tackle this problem is by applying quantitative methods of Systems biology and engineering analysis to cancer research. Furthermore, quantitative modelling creates the power of prediction in this area of research.

Focusing on the intracellular level, there are several distinct signalling networks involved in cancer and inflammation that connect directly and/or indirectly [394,403]. Connections among the signalling pathways lead to a broader range of genes which are affected by mutations of the contributing components within the pathways.

TGF-β and IL-6 signalling pathways are two pleiotropic signalling systems which become deregulated in tumor development and chronic inflammation, respectively. Recent studies have proposed an indirect crosstalk between
these two signalling networks [207]. Jenkins et al. [207] have tested their proposed system experimentally and qualitatively. Particularly, IL $- 6$ affects the TGF$-\beta$ signalling negative feedback and eventually regulates its anti-proliferative pathway. This points to the need for understanding the vital role of feedback loops in designing of the regulatory networks.

The current research investigates the signalling crosstalk between TGF$-\beta$ and IL $- 6$ transduction using mathematical modelling and precise analytical analysis. We have focused on the feedback loops involved in each signalling pathway and their effects on the stability and robustness of the individual systems as well as the integrated IL $- 6$:TGF$-\beta$ network. Each model is validated using experimental data from literature and/or data derived in our lab. Furthermore, several predictions are presented which pave the way for future studies in this field.

1.2 Research Objectives

The overall goal of this research is to develop a quantitative model which describes the crosstalk between the TGF$-\beta$ and IL $- 6$ signalling pathways and demonstrates a better understanding of the role of feedback loops in regulating signalling networks. In order to achieve this, the following steps have been taken:

1. To develop a mathematical model for TGF$-\beta$ signalling

   A biological model is designed in order to capture the interactions of TGF$-\beta$ signalling. This model then is translated to a mathematical model. In this step every biochemical reaction maps to its corresponding equation.

2. To simplify and reduce the components of TGF$-\beta$ signalling model based on analytical methods

   After development of a mathematical model for TGF$-\beta$ signalling, the model is reduced and simplified to its key components. This was done by using rigorous mathematical methods and assumptions. Steady-state and equilibrium analysis help us to acquire an insight to the mechanism of the signalling system.

3. To experimentally validate our proposed TGF$-\beta$ signalling model

   Experimental data sets are used to validate the model predictions at this stage. The output of the model is the time course of a protein concentration which is specific to the signalling pathway. For TGF$-\beta$ signal transduction, we used Phospho-SMAD2/3 concentration level as the marker for TGF$-\beta$ stimulation.

4. To develop a mathematical model for IL $- 6$ signalling
We repeated similar steps in design and development of IL – 6 signalling network. By the end of this step, we have two sets of mathematical equations corresponding to TGF–β and IL – 6 signalling respectively. Each system of equations is solved in order to produce predictions from the model.

5. To simplify and reduce the components of IL – 6 signalling model based on analytical methods

Similar to TGF–β signalling model, we benefit from a series of mathematical and system analysis which leads to a simpler and more understandable model of IL – 6 signalling. The key idea here is to maintain the critical information of the system while reducing the extra details. The final output of the reduced and simplified model should be as rigorous as the full system.

6. To experimentally validate our proposed IL – 6 signalling model

The Phospho-STAT3 time course is used to experimentally validate the IL – 6 signalling model in this step. It is also important to keep the consistency in the validation steps of the two pathways.

7. To integrate the models of TGF–β and IL – 6 signalling and predict the output level, in addition to experimental validation of the integrated model

The final step is to join the two signalling models to design an integrated TGF–β and IL – 6 model. This model predictions can be tested with more experiment.

1.3 Thesis Outline

In this chapter a high-level introduction that motivates this research has been provided, following by an explicit statement of the thesis objectives. The outline of the remaining chapters is as follows.

In Chapter 2, Systems biology principles and its contribution in cytokine signalling are reviewed. The importance of mathematical modelling in studying the biological systems is explained. TGF–β and IL – 6 signalling pathways and their provided mathematical models in the literature are studied specifically, in detail. This chapter will discuss about the crosstalk between TGF–β and IL – 6 signalling pathways and their critical role in tumor progression.

In Chapter 3, the methods which are developed and used in this project are reviewed. Both experimental and theoretical methods are included in this chapter. A stepwise experimental data extraction procedure is indicated. Furthermore, the materials used in this project, including the laboratory reagents and machines, programming platforms and quantitative computer programs are described in Chapter 3.
Chapter 4 presents a framework for the design, analysis and validation of TGF–β signalling network from a Systems biology point of view. Simplification and reduction methods emphasising on the feedback loops of the signalling system are highlighted. The simulations of the final TGF–β signalling model are compared with the experimental results of the previous models in the literature. Model predictions are proposed and experimentally validated, explaining the deregulated TGF–β signal transduction in tumors.

In Chapter 5, we have repeated the design and analysis steps for IL–6 signalling pathway which lead to a simplified IL–6 mathematical model. The IL–6 signalling model explains the experimental data provided in the literature. This model is further validated with our new sets of experimental data on normal and immortal cells. Similar to the previous chapter, predictions are proposed which interpret how the components of IL–6 signalling alter in tumors.

Chapter 6 is dedicated to the integrated model of IL–6:TGF–β and investigates how IL–6 down-regulates TGF–β signalling, causing uncontrolled proliferation of the cells. The integrated model provides various simulations to predict and analyse different stimulation conditions. This chapter ultimately proposes evidence to confirm our hypothesis on the interaction between the TGF–β and IL–6 signalling systems. The contributions of the current thesis briefly recapitulate in Chapter 6 focusing on the future areas of research.
Chapter 2

Literature Review and Background

2.0 Summary of the information included in Chapter 2

Systems biology provides a unique, comprehensive approach in solving the current issues of Biology. Researchers have been attempting to apply Systems biology principles in order to understand cancer and develop treatments. Using mathematical models to capture the details of intracellular signalling pathways quantitatively, is a starting step of identifying the cancer initiators. Focusing on the close connection between tumor development and chronic inflammation, we have studied two specific signalling pathways which are often deregulated in cancer and inflammation (TGF-β and IL-6 signalling). Each signalling pathway is reviewed precisely in this chapter. The previous mathematical models, their strength and their weaknesses are analysed for individual signalling system. Finally the advantages of crosstalk studies are specified.

2.1 Systems Biology

Systems biology is a system-level understanding of living systems which often involves computational modelling and analysis of biological data [25, 240, 241, 245]. Although the study of biological systems commenced in the middle of 20th century [55, 245, 455, 470], it is only during the last decades that the developments in science and engineering have enabled us to progress the affectiveness of the systems approach to biology. What do we mean by "Systems biology": if we substitute the ambiguous "systems" word by a "network of components" and add the important "dynamic" concept, we are starting to approach the power of Systems biology [25, 250].

In general, four steps are required to gain a systems-level understanding of a biological process. The first and the most vital one is identification of the system’s structure (a topology). The structure of the system includes both the
physical structures (the actual location of implied components of the system) and the interactions between the components (the network that confines the system’s components). The interactions among the components determine the final state of the physical structure of the system and hence, must be studied in detail. However, determination of the system’s structure does not define the status of the system completely. Dynamic changes to the components of the system is an intrinsic characteristic of biological systems [245,250,283]. The second step is to understand exactly how the components vary with time. This step requires integration of both scientific and engineering principles, such as computational model development, new experimental information which is relevant to the structure and dynamics of the system. The third step: the controllers of the system and methods of control need to be determined. This step allows us to predict and simulate the output of the system. The forth step in a system-level understanding is to find a suitable method to simulate the biological system as a comprehensive, robust model which can be interrogated, perturbed and analysed over a time-frame relevant to the biology [241,245].

In the Systems biology world, the fundamental principle is to be robust [245,443,457]. Robustness is the insensitivity of biological functions to various disturbances [245]. Robustness is a characteristic of biological systems [243,245,276]. There is a close relationship between the evolution of a biological system and robustness [245,457]. If a biological system lacks the sufficient robustness, it will not survive environmental and genetic perturbations [245,372,452,458]. Only individuals with more robust properties survive sudden significant change. Robustness of a biological system cannot be identified from the components of the system only, however, robust systems have a few common characteristics by which they can be controlled e.g. possession of regulatory feedback loops [245,443]. Common mechanisms across engineering and biology which give a system robustness to different perturbations are [245]:

- Extensive system control

Extensive system control refers to negative feedback loops which provide dynamic regulation of the system and maintain the system around a stability point. Feedback loops occur when the input(s) of a system is regulated by the output continuously [126]. The feedback loop can be positive or negative depending on whether the loop enhances or dampen, respectively, the perturbations from the equilibrium point [210]. Self-regulation has been observed in bacterial chemotaxis [7,26,496]. Controlling biological systems can happen through positive feedbacks as well. Positive feedbacks create bi-stability in signal transduction and cell cycle processes which resist minor perturbations and facilitate robustness in response to environmental changes [63,122,439].

- Redundancy and phenotypic plasticity or diversity
A system can tolerate environmental change if it has alternative components and methods [245]. Phenotypic plasticity [2, 383] is the process by which a biological system changes its phenotype when reacting to perturbations caused by environmental or genetic change. Redundancy occurs when a critical component(s) of a system is duplicated in order to increase the reliability of the system [243, 245]. Reliability of a system defines the probability that the system achieves its expected function [141, 475]. In both cases systems with alternative mechanisms can overcome perturbations and are more robust [243, 245].

- **Modularity**

  Systems with different modules consist of a number of isolated parts. Perturbations occur in one module is isolated to the same module and do not affect the rest of the system. Several functions in biological evolution can increase modularity, e.g differentiation and gene mutation [125, 128, 410]. Modularity is important during the development of a biological system, as it helps to provide a more robust and more reliable topology by, for example, providing the accommodation of a foreign DNA [363] and/or limiting the developmental constrains [106, 245, 306, 446, 456].

- **Decoupling**

  By decoupling minor perturbations in the low levels of a system, the higher levels of the system can be protected. For example, Hsp90 limits the misfolding of proteins as a result of environmental stress and consequently reduces the perturbation of the gene regulation [245, 361, 372, 373].

- **Bow-tie architecture networks**

  Biological systems often have a special architecture (i.e. bow-tie organization in bacterial metabolic networks [84]) which implies intrinsic robustness [245]. What makes bow-tie networks powerful is that there is a large range of inputs and outputs connecting via a much smaller range of processes and protocols in the "core" [84, 243, 245]. In a bow-tie system, the inputs transfer to the core via a funnel and become suitably organized and processed. Hence, these systems are more flexible against the environmental changes [84]. Bow-tie systems are complex (multiple interacting components and linkages) and self-organized (systems which exhibit an overall order via their local interactions between their components i.e. biomacromolecules [54]) and are observed in metabolic pathways [288] and signal transduction [335, 336].

  Cells Mostly control their functions via the environmental cues produced by other cells [127]. Cell functions are the results of intracellular signalling many of which are identified in the last few decades [127]. Intracellular signalings networks must be precisely regulated to avoid improper responses [127]. Precision,
robustness and versatility of the cell signalling pathways originate from the crucial influences they cause [51, 127]. It has been suggested that extensive control of signalling networks can be done by using feedback loops [51, 127]. Similarly, redundancy, modularity and decoupling are often seen in different signalling systems causing high robustness [51, 380, 436, 473].

Although robustness is an essential feature of a biological system, robustness often needs to be balanced against, fragility and resource demands [243, 245]. Often, when robustness develops against a group of environmental changes, the system becomes more fragile in response to other specific perturbations. Diseases can be the direct result of the fragility of a biological system against changes from the outside. In some cases, the biological system shows resistance against therapies too [243, 245]. This can be caused by the same processes that enhance the robustness of that system against harmful perturbations [243, 245].

One example of robustness "hijacking" is cancer [244–246](and autoimmune diseases) [242, 244, 247]. It is evident that many signalling systems are deregulated in cancer [145]. In fact, changes in the intracellular signalling networks transfer a normal cell to cancerous conditions which then acquires new robustness. Tumors are highly robust, these cells resist a wide range of therapies. Tumor cells become robust via the same ways that were mentioned earlier [242, 247]. Having clarified the role of robustness in autoimmune diseases, it is important to build treatment strategies based on this robustness.

Since biological systems are complex combination of networks of components [240], crosstalk study between individual signalling systems (which arise to a cellular function) is essential. This research is focused on how two specific signalling pathways, TGF−β and IL−6, interact in order to regulate cell characteristics. Additionally, we want to study disruptions of these signalling systems in cancer. In order to achieve this, individual computational models are designed and used to predict the cellular responses in normal and tumor cells. In the final chapter, the two separate models are integrated to provide crosstalk predictions.

### 2.2 TGF−β Signalling Pathway

Transforming growth factor, TGF−β, is a member of a transforming growth factor superfamily which also includes bone morphogenic proteins signalling (BMP), Mullerian inhibitory substance, activin, inhibit, and Nodal [75, 296, 396]. Each family member controls a broad range of a cell’s biological characteristics such as: differentiation, proliferation, migration, life time and apoptosis [118, 290, 296]. Signal transduction from the TGF−β superfamily receptors is initiated at the membrane of a cell by ligand binding and a range of intracellular signalling reactions which cause gene responses in the nucleus [300, 433]. Recent studies revealed that ligand concentration, stimulation time, cell type
and even the percentage of active signalling components can affect the gene responses to TGF–β stimulation [75, 297]. Changes in TGF – β/SMAD signalling can affect the transition of normal epithelial cells to cancerous cells in the colon [168, 206, 297, 299, 330, 396]. In spite of the many studies on TGF–β signalling, there are still unanswered questions concerning the impact of signalling from this pathway on cancer cell progression. In particular, there are two essentially opposite reactions of cancer cells to TGF–β [149, 357, 503]. When cancer cells are in an early phase of malignancy, TGF–β acts as an inhibitor of proliferation and slows down cancer progression [265]. Surprisingly, when cancer cells are at a more advanced stage of malignancy, TGF–β stimulates proliferation (For review see [357, 376, 503]). Thus the TGF–β signalling has a dual behaviour towards cancer cells. In order to clarify such paradoxical behaviours, TGF–β transduction and its components need to be studied in detail. The focus of this review is to identify the key parameters in TGF–β /SMAD signalling including: the ligand TGF–β, the receptor proteins TGF–β R1 and TGF–β R2, the mediator proteins SMADs (SMAD1, 2, 3, 4, 5, 6, 7 and 8) and finally nuclear target genes such as JUN, MYC and COL4A1 (the key components of TGF–β signalling are shown in Figure 2.1).
2.2 TGF–β Signalling Pathway

Figure 2.1: The sketch of TGF–β signalling network. Potential phosphorylation sites of the receptors are specified with empty circles attached to membrane receptor components. Arrows pointing to 6 blue dots represent degradation process. The red solid arrows originating from SMAD7:Smurf1/2 apply negative feedback on the receptor components. The dot arrows represent the transcription and translation reactions. Note that the logistics of the signalling are expanded in Figure 4.1.

2.2.1 Key Components of TGF–β Signalling

- TGF–β receptors: TGF–β R1 and TGF–β R2

Defects in TGF–β signal transduction can lead to a range of diseases, including cancer [18,112,290,299,344]. Lack of sufficient levels of a functional TGF–β R2 in the cell membrane is reported as one of the main reasons of abnormal signalling in cancer cells [18,297]. There are two ways the ligand receptor complex can form. Firstly, TGF–β binds to TGF–β R2 [297], the TGF–β: TGF–β R2 complex then recruits TGF–β R1 and leads the TGF–β R2: TGF–β and TGF–βR1: TGF–βR2 complexes to combine. The second model requires the two receptor types (TGF–β R1 and
TGF−β R2) to bind the ligand, but in no particular order. The latter model occurs in BMP signal transduction [297]. Interestingly, even in the absence of TGF−β ligand, TGF−β R1 and TGF−β R2 can form an active complex when over-expressed [62, 297, 313, 447]. Though TGF−βR2 can bind independently with TGF−βR1 and consequently, phosphorylate other components in the pathway to bring about downstream signalling, no data on TGF−β responses have been reported for this particular pathway [297, 311, 396]. Two different mechanisms have been suggested for TGF−βR internalization in the presence of TGF−β. The TGF−β R:ligand complex can be internalized through a lipid raft-caveolar degradation pathway, without participating in signalling [93, 449] or, it can activate mediator proteins in the standard manner (through phosphorylation processes) and degrade after recycling to the membrane [449]. These different internalization paths will result in different TGF−βR signalling responses [194, 348]. If we are to understand the tumor-suppressive effect of TGF−β signalling versus its tumor-promoter effects, it is essential that we understand the kinetics of TGF−βR internalization and recycling [265, 449].

- TGF−β mediators: SMAD proteins

The first downstream proteins activated by the TGF−β R1 are from the SMAD family [224, 290, 297]. SMADs (Sma/mad referring to Small body sized homologues of MAD gene [297]) play a major role in TGF−β signalling transduction and its nuclear relocation. There are three categories of SMADs (a total of 8 types of SMAD proteins) activated in TGF−β signalling [178, 224, 290, 297, 330, 396]: receptor-regulated SMADs (R-SMADs) that interact directly with TGF−β family receptor kinases [290, 297], common SMADs (Co-SMADs) that bind with R-SMADs and regulate TGF−β transduction [290, 297] and inhibitory SMADs (I-SMADs) that inhibit TGF−β signalling by controlling the function of the other two groups of SMAD family proteins (R-SMADs and Co-SMADs function) [178, 224, 290, 297, 330, 396]. SMADs are regulated not only by TGF−β signalling but also by the mitogen-activated protein kinases (MAPK) signalling pathway [131, 260, 297]. SMADs appear to mediate some aspects of MAPK signalling [14, 297, 506]. SMADs are continuously shuttled between cytoplasm and nucleus independently from TGF−β induced activation [200, 386]. The cytoplasm-nucleus interchange of the SMADs is initiated through phosphorylation of the SMAD complex in the cytoplasm [30, 200, 386] and exit from the nucleus is controlled by dephosphorylation [200, 290, 386]. Activation of nuclear import receptors (importin7 and importin8) on the surface of the nucleus facilitates the phosphorylated R-SMADs to enter into the nucleus [200, 386]. It is also postulated that the dephosphorylation of R-SMADs mainly occur within the nucleus [30, 200, 386]. The TGF−β signalling system is robust, with the ability to filter rapid input and noise.
2.2 TGF–β Signalling Pathway

fluctuations [200,386]. The robustness of the TGF–β signalling system is due to the inherent delay between the input from TGF–β binding and the nuclear internalization of the SMAD complex [200,386]. Inhibitory SMADs (SMAD6 and SMAD7) control the intensity and/or duration of cellular responses to TGF–β stimulation [1,297]. One of the proposed ways that I-SMADs regulate TGF–β signalling is by binding to TGF–β family receptors in competition with R-SMADs and hence, decreasing the rate of activation of the R-SMADs [74,225,297,297,396].

- TGF–β effectors and nuclear re-localization

TGF–β superfamily signalling pathways start from ligand binding to the membrane receptors and leading to interactions in the nucleus and specific gene responses. The key parameter for measuring TGF–β signalling is the level of SMAD2/3 nuclear accumulation [18,75,511]. It has been shown that SMAD accumulation in the nucleus is directly proportional to TGF–β receptor activity [73,75,199,329]. As mentioned previously, there are eight types of SMADs in mammals. These eight genes map to four chromosomes [12]. The SMAD2, SMAD4 and SMAD7 genes which are mutated in broad range of cancers [12,143], are found on chromosome 18 (mostly around 18q21.1), while the SMAD3, SMAD5 and SMAD6 genes map to chromosome 15 [12,143]. Finally, the SMAD1 and SMAD8 genes are located on chromosome 4 and 13 [12].

Most of the SMAD family members proteins are 400-500 amino acids in length and all SMADs share two homology domains: MH1 (Mad homology 1) and MH2 (Mad homology 2) [27,290]. The SMADs have a nuclear localization signal (NLS) in the MH1 domain of the R-SMADs and this appears to be the reason that the R-SMADs enter the nucleus without necessarily binding to a Co-SMAD [12,143,262,483]; this is one of the important distinctions between R-SMADs and Co-SMADs [12,262,483]. Additionally, the MH1 domain of R-SMADs interacts with several transcription factors in the nucleus [290,486]. The MH2 domain doesn’t participate in DNA binding but also plays a role in protein association processes [12,27,290]. The SMAD linker region, which connects the MH1 and MH2 domains, is required for interactions with other signalling pathways such as, MAPK [12,297]. It is through this linker that SMAD7 binds with Smurf1/2 and exerts its TGF–β signal inhibitory role [12,35,510]. Interestingly the SMADs can bind directly to DNA with low affinity [91,319]. However, they appear to have higher binding to the transcription factors associated with specific target genes [12,14,318].

- Phospho R-SMAD/SMAD4 complex; a closer look

Co-SMADs also have the MH1 and MH2 domains connected by a linker region [179,297,397]. R-SMADs (except SMAD2) and Co-SMADs bind
DNA through the MH1 domain \([179,397]\), while binding of the SMADs to the TGF-\(\beta\) receptor occurs via the MH2 domain \([93,179]\). Two common R-SMADs which participating in TGF-\(\beta\) signal transduction are SMAD2 and SMAD3 \([297,323,396]\); both are phosphorylated in their C-terminus via TGF-\(\beta\)-R1 \([91,93,179,318,323]\). The phosphorylated R-SMADs bind with SMAD4 (a Co-SMAD) through the MH2 domains \([200,465]\). The details of the formation of the R-SMAD/Co-SMAD complex are still being debated \([116,179,323]\). It is unclear whether the structure of this complex is trimeric or dimeric \([57,116,179,205,227,290,481]\). There is evidence for formation of SMAD trimers: SMAD2/SMAD2/SMAD4, SMAD3/SMAD3/SMAD4 and SMAD2/SMAD3/SMAD4 complexes \([179,385]\). One important observation is the obligatory presence of SMAD4 in the signalling complex. Phosphorylation of the R-SMADs changes the protein structure \([133,481]\). This is more than supplying a phosphate group that normally occurs during SMAD4 activation. The difference in the responses of R-SMADs and Co-SMADs to phosphorylation processes explains why one SMAD4 can bind to several R-SMADs \([397,413,481]\).

The affinity of the R-SMAD/Co-SMAD binding influences the time that the complex spends in the nucleus \([179,386]\). Recent studies show that nuclear translocation of the complex is associated with the interactions between phosphorylated SMAD3 and importin-\(\beta\)1 \([262,483]\). The MH1 domain of SMAD2 is unable to interact with importin-\(\beta\)1 \([71,360]\). Instead, the MH2 domain of SMAD2 participates in nuclear translocation \([179,262]\). SMAD4 is imported into the nucleus either via the interaction of its MH1 domain with importin-\(\alpha\) \([351,482]\) or through the interaction via its MH2 domain with nucleoproteins CAN/Nup214 \([71,482,487]\). Also, there is also evidence claiming an important role for transcription regulators such as TAZ in binding with SMAD complexes \([182,445]\) and forcing them to be retained in the nucleus \([179,445]\). Nuclear SMAD3 uses exportin4 and GTPase Ran which allows it to recycle from the nucleus to the cytoplasm \([351,465]\). SMAD4 also uses its nuclear export sequence, in the linker region, to interact with exportin1/CRM1 and exit the nucleus \([351,465]\). The export mechanisms of SMAD2 are not yet clear \([262]\), however, there are evidences which nominate different nuclear proteins which initiate nuclear export of SMAD2/3 and terminate TGF-\(\beta\) signalling \([61]\).

- SMAD6 and SMAD7, negative regulators:

Over the last decade SMAD7 has emerged as one of the main antagonists for TGF-\(\beta\) and BMP signalling \([396]\). SMAD6 and SMAD7 negatively regulate signalling by TGF-\(\beta\) family ligands \([200,396]\). SMAD7 acts as an inhibitory factor for both TGF-\(\beta\) and BMP signalling while SMAD6 only inhibits the BMP signalling pathway \([485]\). Moreover, SMAD7 is a candidate element for mediating the crosstalk between TGF-\(\beta\) and other sig-
nalling pathways (such as BMP and Wnt) [69,324,343,492]. The inhibitory feedback for the TGF–β pathway is initiated in the nucleus by stimulation of smad7 mRNA [153,162]. The inhibitory regulator SMAD7 recruits Smurf1 (SMAD ubiquitin regulatory factor 1) and Smurf2 to increase the TGF–β receptor complex degradation [74,225]. Smurfs are ubiquitin ligases [69]. Both SMAD7 and Smurf2 can decrease the half-life of the TGF–β R1 [225]. There are two other mechanisms by which I-SMADs inhibit TGF–β signalling: SMAD7 competes with R-SMADs for binding to the TGF–β receptor and SMAD7 competes with R-SMADs for binding to Co-SMADs [216,396]. SMAD6 and SMAD7 also contain the MH2 domain, however, there is no C-terminal SXS phosphorylation site as in the R-SMADs [216]. SMAD7 binds to TGF–β receptor type1 with higher affinity than the R-SMADs [216]. The competition between R-SMADs and SMAD7 causes modulation of TGF–β signal transduction. Interaction of the N-terminus of SMAD7 with its MH2 domain leads to the most efficient inhibitory effect of SMAD7 on TGF–β R1 [166,325,492]. As mentioned before, in order to induce the receptor type1 degradation SMAD7 recruits a group of E3 ubiquitin ligases such as Smurf1/2. During this process, SMAD7 is also ubiquitininated and degraded [100,225,423]. In the absence of TGF–β, SMAD7 tends to reside in to the nucleus as it has NLS in its N-terminal [225]. TGF–β stimulation causes the SMAD7/Smurf complex to translocate to the cytoplasm [100,225,423,492]. Smurfs have a nuclear export sequence which interacts with exportin1/CRM1 [179,427]. It has been proposed that SMAD7 induces dephosphorylation of TGF–β receptor type1 via the GADD34-PP1c phosphatase [395]. Additionally, SMAD7 interferes with the action of R-SMADs on the target genes in the nucleus [504] by interfering with the binding of R-SMAD/SMAD4 to DNA [504]. There is evidence that SMAD7 may also affect the transcription of target genes directly [155,179,261,315,400,492]. SMAD7 is phosphorylated in the linker region (Ser-249 as phosphorylation site) via kinase(s), independent of TGF–β receptor [358]. Whilst this phosphorylation doesn’t influence SMAD7’s inhibitory function, it does affect the transcription processes stimulated by SMAD7 [179,358].

- TGF–β signalling target genes:

Microarray analysis has facilitated the acquisition of information on TGF–β receptor dependent gene expression [48,49,96,280,281]. Microarray analysis has led to a number of new hypotheses on potential functions for TGF–β signalling [49,96]. It is well-known that the TGF–β superfamily signalling pathways stimulate genes related to cell-cell interactions [49], however, as the following table (derived from chapter 17 of "SMAD signal transduction" by Dijke et. al. 2006 [49]) shows the microarray expression analysis reveals other epithelial gene expression signatures in response to
TGF–β stimulation.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Score</th>
<th>Cell Type</th>
<th>Biological Process associated with gene product</th>
<th>Source Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUN</td>
<td>0.75</td>
<td>epithelial</td>
<td>transcription</td>
<td>V,X,C,K,L,D</td>
</tr>
<tr>
<td>FOS</td>
<td>0.5</td>
<td>epithelial</td>
<td>transcription</td>
<td>Z,V,C,W</td>
</tr>
<tr>
<td>JUNB</td>
<td>0.5</td>
<td>epithelial, fibroblast</td>
<td>transcription</td>
<td>C,K,L,D</td>
</tr>
<tr>
<td>MYC</td>
<td>0.75</td>
<td>epithelial, fibroblast</td>
<td>cell cycle arrest</td>
<td>Z,V,C,K</td>
</tr>
<tr>
<td>SMAD7</td>
<td>0.38</td>
<td>epithelial, fibroblast, C2C12</td>
<td>signal transduction</td>
<td>Z,C,K</td>
</tr>
<tr>
<td>CTGF</td>
<td>0.38</td>
<td>epithelial, fibroblast, C2C12</td>
<td>cell-cell signalling</td>
<td>Z,C,K</td>
</tr>
<tr>
<td>TIMP3</td>
<td>0.38</td>
<td>epithelial, C2C12</td>
<td>apoptosis</td>
<td>Z,V,L</td>
</tr>
<tr>
<td>COL4A1</td>
<td>0.63</td>
<td>epithelial, fibroblast</td>
<td>cell-matrix adhesion</td>
<td>Z,X,C,K,L</td>
</tr>
<tr>
<td>ID1</td>
<td>0.5</td>
<td>epithelial, fibroblast</td>
<td>transcription</td>
<td>X,C,K,W</td>
</tr>
<tr>
<td>ID2</td>
<td>0.5</td>
<td>epithelial</td>
<td>transcription</td>
<td>Z,V,K,L</td>
</tr>
<tr>
<td>ID3</td>
<td>0.63</td>
<td>epithelial, fibroblast, C2C12</td>
<td>transcription</td>
<td>Z,V,X,K,L</td>
</tr>
</tbody>
</table>


1458 target genes have been reported to be TGF–β target genes in total 8 quality-verified studies [49]. To determine the level of robustness, Böttiger et. al. [49] scored each TGF–β target gene in Table 2.2.1 by a number between 0.25 (lowest) and 1 (highest). Score 1 means that the target gene was listed in all 8 of the cross-referenced studies and score 0.25 means the gene was named in only 2 out of 8 studies. 11 of the highly robust signatures and genes which are reported frequently are listed in Table 2.2.1. These genes include well-known TGF–β regulated proteins including the
2.2 TGF-β Signalling Pathway

Jun oncogene (JUN), FBJ osteosarcoma viral (v-fos) oncogene homolog (FOS), Jun-B proto-oncogene (JUNB), c Myc proto-oncogene (MYC), MAD homolog 7 (SMAD7), connective tissue growth factor (CTGF), TIMP metalloproteinase inhibitor 3 (TIMP3), procollagen, type IV, alpha 1 (COL4A1), and Inhibitor of DNA binding proteins (ID1, ID2, ID3). Using Gene Ontology tables, most of the known biological functions of the TGF-β signalling pathway (such as; cell cycle control, cell adhesion and cell apoptosis) are associated with ∼160 target genes [49].

2.2.2 The Role of TGF-β Signalling in Cancer

As mentioned earlier, TGF-β signalling transduction has been associated with several disorders, especially several cancers [95,161,430]. TGF-βR2 is inactivated in many gastrointestinal cancers [197,234,297]. The disruption in TGF-β signalling also occurs in response to mutations in SMAD2 and SMAD4 [18,75,297,330]. These mutations occur in different SMAD exons [297].

The role of TGF-β signalling in cancer cells (tumor-promoter or tumor-suppressor) depends on both the micro-environment and the set of cancer gene mutations in the signalling pathways [449]. For example, in vivo cancer cells compete with neighboring cells [449]. There is evidence that the duration of exposure to TGF-β might explain the dual-behaviour of cells to this signalling protein [135]. Chronic TGF-β stimulation (2-3 weeks) can induce cells to become resistant to the anti-proliferative, apoptotic effects of TGF-β signalling [135]. In other words, TGF-β may down modulate its own signalling [135]. In addition to all the other mutations that cause cancer, the duration of TGF-β/SMAD signalling appears to be a critical element that changes during the transition of normal cells to cancerous cells [511]. It has been postulated that cells respond differently to sustained and transient TGF-β stimulation [330]. Sustained exposure to TGF-β may be responsible for the growth inhibition and transient TGF-β stimulation may be required for the anti-proliferation in a certain group of tumors [330,511].

Another detectable change that often occurs when normal cells transform to cancer cells is the loss of E-cadherin [330]. This loss is a characteristic of an epithelial to mesenchymal transition (EMT) [297,330]. As a result of the complexities occur in cell responses to different TGF-β simulation and the leading role of TGF-β signal transduction in cancer development, it is essential that powerful methods, such as mathematical modelling be used to explain the role of TGF-β in the establishment of the robustness of the biochemical models proposed for signalling during the cancer suppression, promotion and progression phases [19,73].
2.2.3 Mathematical Approach to Modelling TGF–β Signalling

Although TGF–β signalling components were discovered decades ago [13, 414], the interactions which occur after activation have been difficult to understand. The signalling components change dynamically with respect to time: both the concentration and location of signalling components change within a few hours of ligand stimulation [511]. The TGF–β signalling system will be best understood by a reliable, robust model using well-designed mathematical simulations of the processes [19, 511].

Even today most biological and even biochemical studies on growth factor/cytokine signalling are qualitative (for example see [28, 85, 314]). The intrinsic complexity of biological systems gives rise to the need for computational simulation [123, 164]. Mathematical modelling is ideal for studying inherent, complex processes (such as signalling systems) [123, 164]. This is because mathematical models are amenable to computer simulations and can be systematically analyzed [123, 164]. Using mathematical modelling of the biological systems provides prediction capabilities of system’s behaviour which is otherwise difficult to obtain due to the simultaneous involvement of numerous components [123, 164].

Several methods for the mathematical modelling of cell signal transduction have been developed [75, 402]. As the concentrations and locations of signal transduction components are time variant, dynamical study of cell’s behaviour and the signalling components is required. Generally, there are two main directions in signalling pathways system biology studies (the study of interactions among components of a signalling pathway that control to the function and behaviour of the system): 1. a ‘top-down’ approach which gathers as much experimental information as is possible from the key biological parameters and outputs and simulates the signalling network bioinformatically [19, 392]; or 2. a ‘bottom-up’ approach which uses a small-scale model with quantitative data for each of the signalling components which are integrated to make a highly predictive simulation system. In short, the top-down method deals with more qualitative and static data and the bottom-up method is supported by more quantitative and dynamic data [19, 291]. The current challenge is how to bring both of these methods together [19].

The most computer representations of the cytokine signalling pathways use a set of ordinary differential equations (ODE) [230, 512]. In the ODE method, the reaction rates are representations of the mass action kinetics, and are proportional to the reactant concentrations. Generally, a system of a small number of ODE equations can be solved either precisely or via analytical approximations [90, 172]. As the number of equations increase, providing an analytical or graphical solution for the system becomes difficult. When all of the parameters for all the components are known, a reliable interpretation of dynamic behaviour of the system components can be expected [230]. The parameters for
this model need to be estimated quantitatively using mass spectrometry [472], immunoblotting [86] or protein antibody microarrays [268]. However, for complex dynamic system with so many equations, such as the cell signalling pathways, it is not possible to solve the ODEs analytically [123]. Consequently, numerical integration solutions are used and hence, sensitivity analysis is vitally important to identify the critical signalling components regulating the output of the system [293]. Numerical integration can result in large changes even in outputs from small changes in parameter values; therefore, by identifying the components to which the system is more sensitive, a more robust and reliable model of the system can be developed [177].

Alternative modelling methods such as Boolean, stochastic and spatially resolved models (e.g. PDE-based models) have been used to model cytokine signalling networks [192]. PDEs are used when attempting to model the spatial and temporal dynamics of the biochemical species. In many intracellular signalling networks the important spatial information is often incorporated in ODEs via compartmentalization [192, 289]. This approach is often used in order to simplify PDEs to ODEs [192, 289]. Stochastic modelling assists in the robustness analysis of a system [192, 289]. One way to incorporate stochastic analysis in an ODE-based model is to consider that the noise can be represented by a Gaussian distribution. The other way is to perform sensitivity analysis on the system, thus identifying which parameters influence the output of the model [192, 289]. When there is lack of kinetic information and transcription details, Boolean networks are often used in gene activation and transcription models [289]. However for larger systems, Boolean modelling becomes impractical, due to a large number of possibilities [289].

The signalling systems studied in this thesis have numerous components and involve all three spatial components: membrane, cytoplasm and nucleus. Because of the complexity of the system and number of possibilities for the reactions, it is not sensible to use a Boolean modelling. Also we have considered the most probable possibility for each reaction based on the experimental data and hence, deterministic modelling is chosen over stochastic modelling. Conceptually all the cytokine signalling models should consider spatial information of the components and thus should be modelled by PDEs. However, mathematical analyses of PDEs are not straight forward and it is more convenient to simplify PDEs to ODEs where possible. We have used spatial compartmentalization and sensitivity analysis to develop our signal transduction models in this thesis. Consequently, our systems meet the conditions of using ODE-based modelling.

We have chosen to follow an ODE modelling approach. This allows us to use experimental data (e.g. time course of protein expression) to parameterize our models, at the same time it is possible to investigate the mechanism, robustness and behaviour of the networks. The results of Chapter 4, 5 and 6, in which the proposed ODE-based models are analysed and tested experimentally, confirm
the validity of our modelling approach.

**Approaches to modelling:**

There have been at least 7 independent mathematical models of the TGF–β signalling pathway (the whole pathway or only part of it), see table 7.1 in the Appendices. Most literature models of cell signalling use differential equations to model the networks of biochemical reactions. In the current research, we have built on the most recent ODE model of TGF–β signalling provided by Zi et al. [511]. We have focused on the models with inhibitory feedback of the signalling pathway, this essential topological component seems to be missing in most of the existing models [72, 74, 308, 449, 511, 513]. The I-SMAD role as the inhibitory feedback component in TGF–β signal transduction assures the stability and robustness of the system. It raises hypothesis that the signalling system is regulated by a balance between the feed-forward and feed-back reactions, therefore to achieve more accurate simulations it is essential to simulate both the inhibitory feedback as well as the feed-forward effects (signal transduction).

SMAD7 is assumed to be a key component for the crosstalk of TGF–β signal transduction with other signalling pathways (see section 2.2.1 "SMAD6 and SMAD7, negative regulators"). A more complete description of the topology of the TGF–β signalling models, the parameter and values outputs is given in Appendices.

### 2.3 IL – 6 Signalling Pathway

The interleukin-6 (IL – 6) family of cytokines all signal from the plasma membrane to the nucleus [186,254,389,514]. These cytokines act principally by changing the function of target cells, whether it happens in a paracrine or autocrine manner [239,251]. Cytokines of the four helix bundle family, such as IL – 6, participate in the differentiation, growth and regulation of the cells of the immune and hematopoietic systems [285]. However, it should be noted that members of the IL – 6 family signal in a wide range of cell types [83, 138, 444, 476]. Signalling happens when the IL – 6 cytokine binds with its specific receptor, such as IL – 6R (also called Gp80 [448]) and then to a common signal transduction co-receptor, 130 kDa glycoprotein (Gp130) [253]. Equilibrium binding studies have documented that Gp130 increases the affinity of the interactions between the IL – 6 and IL – 6R [189].

All members of IL–6 family, IL–6, IL–11, leukaemia inhibitory factor (LIF), oncostatin M (OSM), ciliary neurotrophic factor (CNTF) and cardiotrophin-1 (CT-1) use Gp130 to transfer the signalling to the cytosol. It is also known that Gp130 induces the cytoplasm signalling by activating the Janus kinases (JAK) [187,346]. The target genes activated by IL – 6 signalling participate in differentiation, apoptosis, proliferation, inflammation and the immune responses [175].
IL – 6 target genes have been studied carefully both structurally and functionally [175].

IL – 6 family members play a significant role in the response of the body to injury and inflammation [175,431,454]. Deregulation of the signalling pathways of IL – 6 member cytokines is associated with human hematological and epithelial malignancies [109]. IL – 6 is known to both promote and inhibit tumors [251]. It has been reported that IL – 6 is secreted via tumor cell lines together with many other cell types (such as fibroblasts, endothelial cells, monocytes, keratinocytes, lymphocytes, macrophages and mast cells) [83,170,193,370,444].

IL – 6 is responsible for the regulation of many inflammatory processes [24,175,431,454]. Examples for inflammatory processes are: B cell development, macrophage activation, initiation of organ repair and the production of hepatic acute-phase protein [29,83,183,255,364]. An inflammatory disease occurs when the natural inflammation processes get deregulated [209]. One clinical treatment for inflammatory diseases is to block cytokine activities that have caused an exaggerated inflammatory response [209]. Drugs using this method have brought many clinical benefits to the treatment process for inflammatory diseases [209], however, it is essential to understand completely how/why this cytokine signalling system is distorted chronically and how this might be prevented or reversed [209].

IL – 6 signal transduction is likely to vary with respect to the strength of the initial signal i.e. the cytokine and receptor levels. In the remainder of this section, the Gp130 co-receptor and its role in IL – 6 signalling via the JAK/STAT [421] and MAPK pathways [98,401] and the negative auto-feedback via SOCS3 [15,138] will be discussed. The importance of IL – 6 transduction during the cancer progression and its effects on other signalling pathways is considered in the following sections (2.3.6 and 2.3.6). Finally, quantitation and mathematical modelling of IL – 6 signalling are introduced in a Systems biology context as the best approach to achieve an understanding of the signalling networks and their role in cancer studies.
2.3 IL – 6 Signalling Pathway

Figure 2.2: The sketch of IL – 6:Gp130:JAK:STAT signalling. Potential phosphorylation sites of the receptors are specified with empty circles attached to membrane receptor components. The dot arrows represent the transcription and translation reactions. Note that the logistics of the signalling are expanded in Figure 5.1.

2.3.1 Gp130 Membrane Co-receptor

Gp130 is found on almost all cell membranes (e.g. see [34, 160, 278, 310]). For example, cells from heart, lung, brain, kidney, spleen, colon and placenta all express Gp130 on their plasma membrane [209, 497]. Gp130 is essential for development [491], hematopoiesis [67], cell growth [81] and survival [113, 209, 497]. Gp130 regulates the binding sites for a broad range of cytokines [36, 38, 285]. In conjunction with the cytokine specific receptors, Gp130 transmits signalling for all of the IL – 6 family members (LIF, OSM, IL-6, IL-11, CNTF, CT-1, CLC, LIFR, OSMR and TCCR) [15]. Gp130 contains 6 tyrosine residues in its cytoplasmic domain, each prepares a different binding site for signalling proteins in the cytosol.

Gp130 is important for participating in a complex together with IL – 6 or IL-11 specific receptors (the IL – 6R is also known as Gp80 [181, 309]). Ligand acti-
2.3 IL – 6 Signalling Pathway

viation of the IL – 6:Gp130 complex leads to the activation of Src [432], JAK/Tyk tyrosine kinases [285] and STAT family transcriptional regulators [37, 176, 196]. Activation of JAK consequentially leads to phosphorylation of proximal tyrosine residues on Gp130 receptor [109, 209, 402]. The phosphorylated Y residues lead to binding of STAT3 and STAT1 [109, 144, 209]. Shp2 is also regulated through Y phosphorylation on Gp130 [109, 328, 402]. Shp2 contains a SH2 (Src Homology 2) domain in the N-terminal region and a Grb2 binding motif in its C-terminal region [117, 214, 453]. The activation of Shp2 is necessary for mitogen-activated protein kinase (MAPK) signal transduction to occur. MAPK activation involves the successive activation of Ras, RAF, MAPK MEK and ERK [109, 402, 428]. It is reported that Shp2 can act as an inhibitor for IL – 6 transduction via dephosphorylation of Gp130 receptors (this fact will be discussed later in this document). IL – 6 signalling activates two distinct pathways for transferring the incoming receptor-initiated signals to the nucleus: JAK/STAT and MAPK pathways [175, 176, 402]. In the following subsections these two pathways and their components are described in more detail.

2.3.2 JAK/STAT Pathway

As discussed earlier, the two of the receptors that take part in the JAK/STAT signalling pathways are Gp130 and Gp80 (also called IL-6Rβ and IL-6Rα, respectively [309]). Although Gp80 is essential for the IL – 6 transduction, it is Gp130 that activates the cytoplasmic signalling through its cytoplasmic domain [309]. This can be seen from the size of the cytoplasmic domains for each receptor. Gp80 contains only 82 amino acids in the cytoplasmic domain [309], compared to 277 amino acids in the cytoplasmic domain of Gp130 [94, 309]. Most of the intracellular signalling is through Gp130 motifs (such as the YSTV sequence for signalling through Shp2, MAPK signalling [114, 309], and the YXXQ motif for STAT signalling [114, 309]).

Research on the crystallography of the cytokine/receptor complex of IL – 6 signalling pathway suggests a sequential sequence of binding [37]. First, IL – 6 binds to the extracellular domain of IL-6Rα. Next, the IL-6/IL-6Rα complex interacts with Gp130 and finally, the resulting complex is dimerised [37, 138, 411]. This hexameric complex is ready to initiate JAK/STAT signalling.

The cytoplasmic domain of the hexameric receptor complex has the ability to interact with an intracellular tyrosine kinase (JAK). In this pathway, JAK is the key protein through which the downstream signalling is propagated [109, 181, 309]. Four members of JAK family are found in mammals: JAK1, JAK2, JAK3 and TYK2 [214, 286, 326, 415]. All, except JAK3, are constitutively associated with cytoplasmic domain of Gp130 monomers [109, 478]. Binding of IL – 6 initiates the activation of JAKs. JAK activation happens via an auto-phosphorylation process after the octamer (2IL-6/2IL-6Rα/2Gp130/2JAK) is formed [109, 175, 181, 478]. The phosphorylation of JAK triggers the phospho-
Signalling Pathway

The phosphorylation of the tyrosine residues of the two Gp130s in the complex and hence, creates binding sites for proteins containing phosphotyrosine binding domain (such as STAT transcription factors) [478]. It has been suggested that JAK1, among all the family members, is the most common kinase associated with Gp130 that triggers the STAT3 nucleocytoplasmic shuttling (reviewed by [214]). Immediately after the phosphorylation of the Gp130s, the receptor complex is ready to phosphorylate STAT3 [196,425]. Phosphorylated STAT3 separates from the receptor complex in the form of a dimer [196, 425, 478]. Phospho-STAT3 dimer is imported to the nucleus, acts as a transcription factor and loses its phosphatase there (dephosphorylation of STAT3) [232, 490].

Mammalian STAT proteins are normally categorised into 7 major groups (STAT1, 2, 3, 4, 5a, 5b and 6) [176,195,196,346,426]. There are no identified STAT-like gene in unicellular organisms [176] so STATs are assumed to be the result of the multicellular organism evolution. Reports show that STATs can be activated via different receptor tyrosine kinases such as: epidermal growth factor (EGF) receptor [137, 139, 509], fibroblast growth factor (FGF) receptor [339, 440], platelet-derived growth factor (PDGF) receptor [40, 338] and etc [46, 176, 332, 342]. All IL − 6 family members activate STAT3 and STAT1.

STATs are relatively large proteins with 750 to 850 amino acids [176]. This property lets the STAT family members interact with many different proteins through their various domains [176]. The Src homology 2 (SH2) is the domain responsible for STAT binding to the tyrosine-phosphorylated receptor motifs and also homo- and hetero- dimerization among the family members [152, 176, 398]. This variety in binding sites of STATs gives them the capability to be the common transcription factors for multiple transcription factors [176, 468, 488].

The dimerization of STATs is essential for the DNA binding. The DNA binding domain is located in the middle of the STAT molecule [176, 398]. Nuclear translocation of the STAT dimer can be followed by indirect immunofluorescence [176,342]. STAT translocation to the nucleus is transient, i.e. the STAT concentration returns to its pre-stimulus levels within 2 hours [176, 342]. A nuclear localization sequence (NLS) has not been identified for the STATs [176,307,390].

There are number of IL − 6 target genes activated by STATs:

1. APP genes [31,500]: C-reactive proteins, $\alpha_1$-antichymotrypsin [66,256], $\alpha_2$-macroglobulin [349, 466], lipopolysaccharide-binding protein [8, 387] and tissue inhibitor of metalloproteinases (TIMP) [50,231]

2. transcription factors: JunB [78], c − Fos [184, 480], interferon regulatory factor (IRF)-1 [167, 418] and CCAAT enhancer binding protein (C/EBP) [252,489]

3. Other genes: interstitial collagenase [105, 257], vasoactive intestinal peptide [77, 424], pro-opiomelanocortin [366, 442], heat-shock protein hsp90
2.3 IL-6 Signalling Pathway

2.3.3 MAPK Pathway

IL-6 stimulation activates the MAPK pathway. Mitogen-activated protein kinase (MAPK) [190, 337] is a highly conserved mechanism associated with the regulation of cell proliferation [263]. MAPK signalling can be initiated by different types of cytokine stimulation, including TGF-β [87, 420] and IL-6 family stimulations [185, 263]. MAPK activation plays a role in gene transcription, cell cycle regulation, apoptosis, protein biosynthesis and cell differentiation [229, 263].

Sh2-domain of Shp2 is used to help the phosphorylation of Tyr759 residue of activated Gp130 [39, 331, 416]. Sh2-containing tyrosine phosphatase (SHP2) is then phosphorylated by the activated JAKs [144, 185, 381, 416]. Phosphorylated Shp2 interacts with growth factor receptor bound protein 2 (Grb2) [505] that associates with son of sevenless (Sos). Sos [142] is a GDP/GTP exchanger for Ras [58]. Activation of ERK/MAPK cascade happens through the GTP form of Ras [180]. The activated ERK/MAPK signal is transferred to the nucleus, targeting the transcription factors i.e. NF-IL-6 [185, 309].

The ERK/MAPK [233] signalling pathway has been investigated in detail [214]. In this document, we focus on the IL-6/JAK/STAT signal transduction in order to emphasize on the STAT3 key role in inflammation and cancer [171].

2.3.4 Inhibitory Feedback of SOCS3

The mechanism for down-regulation of JAK/STAT signal transduction is caused by the suppressor of cytokine signalling (SOCS) proteins [176]. This family of proteins is also known as "JAK-binding proteins (JAB)" and/or "STAT-induced STAT inhibitors (SSIs)" [15, 108, 176, 322, 417]. SOCS transcription is induced by the IL-6 and LIF [303] cytokines. SOCS function as inhibitors of tyrosine phosphorylated Gp130, STAT1 and STAT3 [108, 176, 322].

The SOCS family contains 8 members: SOCS1-SOCS7 and CIS [83]. Due to their SH2 domain, SOCS are able to directly interact with the kinase domain of JAK family members [108, 176, 328], reducing their tyrosine kinase activity. Thus SOCS proteins are JAK inhibitor and provide negative feedback on the JAK/STAT signalling system [176]. The ability of SOCS to inhibit the JAKs comes from a short motif in the SH2 domain, known as the KIR (kinase inhibitory region) [16, 42]. JAK inhibition via KIR is the dominant mode of action of SOCS1 and SOCS3 in vivo [16, 42, 502].

Through the carboxy-terminal SOCS box, SOCSs target the Gp130 signalling complex for polyubiquitination and proteosomal degradation [6, 83]. The SOCS box is a C-terminal sequence of almost 40 amino acids that contains 2 conserved regions [43, 217]. It is believed that the SOCS box is essential for the stability
of the SOCS proteins [43,218]. Over-expression studies reveals that SOCS box, together with the PEST sequence of SOCS3, facilitates the degradation of the SOCS3 [17]. Experiments show that SOCSs act as the coordinators of the biological responses to the specific cytokines, as well as, their role in attenuating the cytokine signalling (by degradation of the active complexes on the membrane [201,379]) [83].

SOCSs proteins also compete with STATs for binding to the phosphorylated sites on the Gp130 receptor [322,477]. Thus SOCS3 can interact with both JAK and Gp130 simultaneously. This property creates a ternary complex with high overall affinity [16]. Consequently, if the binding cytokine signals through receptors with the same binding sites as the SOCS3 (such as $IL − 6/Gp130$), its signal transduction will be suppressed by SOCS3 [16,479]. It is also important to note that SOCS3 interacts with JAK1, JAK2 and TYK2 but not JAK3 [16]. This selectivity is because of a motif in the JAK insertion loop [16]. Through this interaction, SOCS3 is able to identify the target JAK and distinguish between them. The lack of this specific motif (motif$^{1071−1073}$ GQM) in JAK3 removes the SOCS3 feedback inhibition of its signalling [16].

### 2.3.5 $IL − 6$ and Cancer

Historically, $IL − 6$ was first discovered by Hirano and Kishimoto in 1986 as a B cell stimulating IgG production [186,327]. $IL − 6$ is produced by many tumors, for example: melanoma [214,316], multiple myeloma [214,228]and prostate cancer [214,407]. $IL − 6$ can act as autocrine and/or paracrine growth factor to stimulate tumor growth. Additionally, $IL − 6$ levels change in age-related diseases (e.g. Alzheimer’s [508]). It has been assumed that the decrease in the level of sex hormones which occurs in aging, increases the $IL − 6$ levels [111,214].

As mentioned before, because of the Gp130 involvement in the $IL − 6$ transduction, any deregulated production of $IL − 6$ can induce distributed undesired inflammation in the organs. Many diseases progress via $IL − 6$ over-expression, e.g. Alzheimer’s disease [99,333], type 2 diabetes [92,356]and some carcinomas [64,221]. A recent study reveals a direct relationship between the ulcerative colitis, developing colon cancer, over-expression of $IL − 6$ and the level of activated STAT3 [270]. Similar findings suggest that inhibition of $IL − 6$ signalling may prevent the progression of chronic inflammatory diseases and inflammation associated cancers [327].

### 2.3.6 Crosstalk

One of the features of $IL − 6$ signalling is that it interacts with other signalling pathways via its receptor, Gp130, and the alternatives in its downstream signalling (JAK/STAT and MAPK). There is evidence that $IL − 6$ crosstalks with cytokines such as BMP [493], NF-$κβ$ [31], Wnt/$β$-catenin [47]and TGF-$β$ [207].
The focus of this project is on the IL−6 and TGF−β crosstalk, with an emphasis on the role of SMAD7. This relationship is supported by the interaction between STAT3/SMAD7/TGF−β in tumors where the EGFR was over-expressed and activated [287]. Persistent activation of STAT3 stimulates the SMAD7 expression, which then desensitizes TGF−β signal transduction [207,287]. Luwor et al. (2012) showed that this relationship occurs in vivo as well as in vitro. Yao et al [495] have reported that both TGF−β and IL−6 are hyper-activated in Erlotinib-resistant cells and are essential for the survival of cancerous lung cells.

The concept of this thesis is the report on IL−6:TGF−β interactions by Jenkins et al. (2005) [207]. They used "MEF" (mouse embryonic fibroblast) cells and mutated Gp130 in order to over-express STAT3 or STAT1. Cells with STAT3-hyperactivation illustrate impaired activation and nuclear translocation of SMAD2. Therefore, the TGF−β signal transduction was down-regulated in mutated MEFs. They tested their hypothesis using a PSMAD7-luc reporter to follow the SMAD7 transcription levels in mutated cells. They have also repeated the experiment in vivo with the same result [207].

The interaction between the TGF−β and IL−6 signalling systems is very important in inflammatory, auto-immune diseases and cancers. As a result, it is essential to study both pathways quantitatively, and identify their precise possible interactions and predict ways to modulate the pathological signalling.

### 2.3.7 Mathematical Approach to the Modelling IL−6 Signalling

There are several mathematical ODE models of the JAK/STAT [70,320,402] signalling pathway. Most of the models of IL−6 signalling concentrate on MAPK signalling through Shp2 [402]. For our purposes, we have developed a mathematical model with precisely quantified components for JAK/STAT signalling.

Singh et al. [402] proposed a model of IL−6 signalling based on the literature data sets. The roles of SOCS3, Shp2 and PP2 (nuclear phosphatase) were studied subsequently by knock out studies [402]. Singh et al. [402] compared the responses of the model for perturbations of specific components in order to specify their functions. Although their model is comprehensively designed, there was no validation of the model. In addition a number of details that were overlooked in their model (such as transcription and protein synthesis). No mathematical compensations were applied to the model to compensate for these omissions. This model, notwithstanding, is one of the first attempts to develop a mathematical formulation from biological models of the IL−6 signalling. I have used some of the kinetic values and initial conditions from Sing et al. [402] for this project.

Another IL−6 model has been proposed by Huang et al. in 2010 [70]. They based their model on Singh et al. [402], but used sensitivity analysis and parameter clustering to simplify the model. They provided evidence that their simplification method was valid by comparing their results with the results of
the original Singh’s model. Their work has identified the most important and effective components in the original model, however, the validation of the model against experimental data was not attempted.

The most recent mathematical model of the IL – 6 signalling pathway was published in 2010 by Moya et al. [320]. Again, the foundation of their model was Singh’s model. They repeated the sensitivity analysis and used some new parameter estimates. They also used an implicit model of dynamics for the MAPK pathway. Finally, they applied same analysis and methods to IL-10 signal transduction. Experimental data were used in the MAPK signalling analysis. The IL – 6 and IL-10 models were used to investigate their roles in the regulation of STAT3. The paper perfectly filled the gap lacking in the mathematical and statistical analyses of the previous models, however, including MAPK signalling, all signalling pathways which activate MAPK must have been considered in the model.

Therefore, a model of the IL – 6/JAK/STAT signalling which focuses on the signalling crosstalks via activated STAT3 is not yet available. In this thesis I propose such a model and integrated IL – 6 signalling with the TGF–β model to investigate interactions between these pathways.

2.4 Summary

Intracellular signalling pathways play essential roles in regulation of the cell functions. Many diseases, such as different cancer types, are developed by the deregulation of the signalling pathways. Because of the close connection of inflammation and cancer, it is important to study the crosstalk between the inflammatory the growth signalling pathways. Additionally, the complexity of the intracellular signalling networks highlights the need for Systems biology approach and quantitative modelling of the networks. Two specific signalling pathways are studied in the current thesis: IL – 6 signalling which is up-regulated while inflammatory responses and TGF–β signalling which causes anti-growth responses. Individual studies have identified the major components involved in each signalling system qualitatively. However, no quantitative analysis and modelling are reported which emphasizes on the crosstalk between IL – 6 and TGF–β signalling pathways.

This points to the need for development of two separate mathematical models which follow Systems biology principles in order to quantitatively identify the components of each signalling system and their role in the final output of the system. Next, mathematical models are experimentally validated individually. The IL – 6 and TGF–β signalling models are combined together to develop an integrated mathematical model which inspects the crosstalk between the two pathways.

This literature review has covered four major aspects of this project:
First, the Systems biology approach and its different methods were reviewed as the main theme of studying the IL–6, TGF–β and the integrated models. Second, TGF–β signalling pathway and its key signalling components were reviewed, focusing on the role of TGF–β signalling in cancer. Previous mathematical models of TGF–β signal transduction in the literature were reviewed to specify the contribution of this project. Third, IL–6 signalling pathway and the interaction between its components were studied, emphasizing on the literature mathematical models of IL–6 signalling. Fourth, the crosstalk between IL–6 and other signalling pathways were analysed, focusing on Jenkins et al. (2005) [206] first discovery of the attenuation of TGF–β signalling via IL–6 up-regulation.
Chapter 3

Methods

• All methods in this Chapter were performed by the applicant.
• In the revised thesis, all abbreviations are either explained in full or detailed in the "Abbreviations" section on pages vii-ix.

3.1 Measurement of Key Signalling Components

MEFs were grown to confluence and the proteins extracted with detergent buffers. The method we used to quantitate the levels of SMAD2, STAT3, SMAD7, PSTAT3 and PSMAD2 was Western blotting. Where needed the experimental data set and the kinetic rates used to set initial parameters for the model were elicited from the literature [72, 74, 308, 386, 402, 449, 479, 511, 513].

Once the initial concentrations for the key signalling components were determined for MEFs, the cells were fractionated into plasma membrane, cytosol and nuclear compartments to measure the sub-cellular concentrations of each key signalling component. The proteins from cultured cells were harvested from 5 minutes to 2 days after stimulation and the levels of the key signalling components in the three subcellular compartments were measured quantitatively. The measurements are compared to the predicted values from the computer modelling.

3.1.1 Cell Culture and Cell Lysis

Mouse embryonic fibroblasts (MEFs) cells were isolated from day 13 to 15 embryos. Wild type MEFs and MEF Gp130\(^{+/\text{-}}\) [206] and SV40-immortalized MEFs [147] cells were cultured in DMEM containing 15 % FCS. The cells were typsinazed and washed with DMEM + 15 % FCS before plating. Passage 3 cells were seeded at 1×10\(^{6}\) MEFs/well in 60 mm plates for 0-4 hour, 0.5×10\(^{6}\) MEFs/well for 24 hour and 0.25×10\(^{6}\) MEFs/well for 48 hour treatment with 1 ng/ml IL – 6 (Ref 1661.F33. WEHI) and/or 5 ng/ml TGF–\(\beta\) respectively (for Gp130\(^{+/\text{-}}\) MEFs the cell numbers changes as: 1.2×10\(^{6}\) MEFs/well for 0-4 hours, 0.6×10\(^{6}\)
3.1 Measurement of Key Signalling Components

MEFs/well for 24 hours and $0.3 \times 10^6$ MEFs/well for 48 hours. After washing with cold PBS for two times, cells were lysed in ice-cold 200 ul RIPA lysis buffer, containing 1M Tris/HCL, 0.5 M EDTA, 5M NaCl, 10 % Na-Doc, 10 % TX-100, 10 % SDS, protease inhibitor 100 $\times$ and H$_2$O. The cell lysates were passed through 27 G needle for 5 times, then incubated in ice for 20 min. After incubation the samples were spun at 13,000 rpm for 30 min at 4°C. The supernatant was transferred to new tubes where 20 ul of sample was used for a BCA protein assay using sigma BCA assay kit (product No. B9643). 20 ul 5$\times$ SDS sample buffer was added to 80 ul of the lysate and the samples were heated at 95°C for 10 min. The following steps were followed to prepare the samples for Western blot experiment.

**Cell Collection**

- Aspirate (remove the medium by vacuum suction) the medium.
- Add 5 ml PBS to remove all FCS. Wash and aspirate it. Repeat this step once more to remove all the medium (FCS neutralizes the trypsin therefore any traces of media will inhibit the action of the trypsin on the cells)
- Add 2 ml TV.
- Put the plate in the incubator for about 30 sec - 1 min.
- Label cell culture flasks. (Usually 3 or 4 flasks would be needed when learning to seed cells at a particular cell density or number)
- Use 50 ml Falcon tube. Label it with cell name.
- Add 10 ml medium to the cells and transfer them all to the 50 ml tube.
- Re-suspend the mixture for about 5 times
- Spin the mixture for 5 minute and aspirate the above medium then re-suspend the pellet in 10 ml fresh medium.

**Cell Count**

- Take a 100 ul sample of the cells. Use the 1.5 ml Eppendorf tubes
- Add 30 ul of the TB liquid to 1 well of a 96 well dish.
- Mix the 100 ul aliquot (sample) of cells and add to 30 ul of the cells to the trypan blue volume.
- This is a 1:2 dilution (dilution factor should be considered when calculating cell number/ml)
- Insert 10 ul of the mixture on to the cover-slip of the hemocytometer.
- Count each of the 4 square separately and take the average.
3.1 Measurement of Key Signalling Components

- Multiply the dilution factor and the zooming factor (1e4). This is the number of cells per ml.

- Multiply the amount of the medium (2 ml TB + 10 ml FCS DME). This would be the total number of cells.

Cell Culture Set-up

- In order to have different cell concentrations in each flasks/dishes calculate the volume of the cells for different numbers of cells per ml (i.e.: 0.5e6, 1e6, 2e6, 4e6).

- Add those calculated volumes to 10 ml of medium

- Label the flasks/dishes with cell line, passage number, date, type of medium, total cell count added to the flask/dish and volume

3.1.2 Western Blotting

According to the protein concentration results from the protein assay, Novex NuPAGE® 4-12%-Bis-Tris (life technologies NP0335 Box) gels were used for loading the lysates for each time point. PSTAT3 (XP™ Rabbit mAb) antibodies were provided by Cell Signalling Technology (product No. 9145) and were used at a dilution 1:500 in 3% BSA-TBS-T. Antibody directed against SMAD7 was purchased from Santa Cruz Biotechnology and was diluted according to the manufacture’s instructions. PSMAD2 antibody (rabbit polyclonal anti-phospho-Smad2 antibody (1:1000 for Western blot)) was a gift from Prof. Peter ten Dijke (Leiden University Medical Center, Netherlands). PSTAT3 (XP™ Rabbit mAb) antibodies were provided by Cell Signalling Technology and were used at a dilution 1:500 in 3% BSA-TBS-T.

β-tubulin, Lamin B1 and transferrin receptor were used as loading control depending on the protein to be expressed. The gels were transferred onto nitrocellulose stained membrane via iBlot 2 gel transfer device (life technologies) and the membranes were scanned quantitatively using Odyssey infrared scanner (LI-COR). The following describes step by step procedure of Western blotting.

- Prepare a gel. Put it inside the western blot container.

- Use the MOPS as the liquid of the western blot if the proteins are quit big in size.

- Dilute the MOPS 1:20 (25 ml of MOPS in 500 ml of DDT).

- Pour the diluted liquid in the centre of the western blot container. Let it to redound to the sides spaces.
3.1 Measurement of Key Signalling Components

- Wash the wells of the gel using a syringe.
- Leave one of the wells for the standard sample. Here it is see-blue. Add 7 ul of the see-blue to the well.
- Use your pre boiled samples and fill all the wells in turn.
- Apply voltage to the container for about 1 hour. (The applied voltage should be between 150-180 V)
- Take the gel out and open it gently.
- Transfer the gel to the membrane and leave it about 6-8 min.
- Put the membrane in milk and cover it with aluminium foil and leave it on the shaker for a night. (we use the milk to block all the empty spaces between the layers to avoid unwanted bindings)
- Use TTBS (TBS-T) to wash the membrane. First rinse the milk. Then repeat the washing 4 times with 10 ml of TTBS following with 5 min moving on the shaker.
- Dilute 0.01 of the first anti body in 10 ml TTBS in Falcon tube.
- Add the 1st Ab to the membrane.
- Put the membrane on the shaker for another one hour.
- Another 4 time wash with 5 min shaking.
- Dilute 1ul of the second anti body in 10 ml TBS.
- Add the secondary Ab to the membrane and cover it with aluminium foil in order to avoid unwanted bindings caused by light.
- Put it on the shaker for another 1 hour.
- Repeat the 4 time wash and 5 min shaking.
- The last wash is with the 10 ml PBS to remove TTBS itself.
- Scan the result using Odyssey infrared scanner (LI-COR).

3.1.3 Protein Quantitation

The western blot images were quantitated using ImageJ 1.49p [140]. The signals for each protein were normalised using the loading controls.
3.2 Computer Modelling

The program used in this project was PYTHON 2.7 and MATLAB 7.10. PYTHON is known as a high-level multi-paradigm, object-oriented programming language. It also uses fully dynamic type system which benefits from an automatic memory management in addition to a comprehensive standard library. There are also several interpreters used for PYTHON which are suitable for almost every operating system. MATLAB is a well known and wildly used programming tool. It is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, you can analyze data, develop algorithms, and create models and applications. The curve fitting tool box of MATLAB is used for fitting the Hill equation to Figure 6, 7, 8 and 9 and deriving the Hill coefficients.

The other program that has been applied for the model development is CellDesigner [302] which is a structured diagram editor used to develop models of biochemical networks. The graphical notation system is stored in the form of Systems biology Markup Language (SBML) files but, it also supports simulations with Copasi [450]. Figures describing the model were produced via CellDesigner4.3 and Microsoft Power Point.

3.3 Design and Modelling Methods

3.3.1 Mathematical Modelling of the Intracellular Protein-Protein Reactions

Most common computer representations of cytokine signalling pathways use a set of ordinary differential equations (ODE) [20, 249]. In this method, the reaction rates are a representation of mass action kinetics and are proportional to the reactant concentrations. The ODE modelling method, when all of the components are known, is a reliable interpretation of dynamic behaviour of the system components. The parameters of this model need to be estimated quantitatively using mass spectrometry, mass immunoblotting or protein microarrays. For complex dynamic systems, such as cell signalling pathways, it is not possible to solve the ODEs analytically. Consequently, numerical integration solutions are required and hence, sensitivity analysis become more vital to identify the critical signalling components which regulate the output of the system. Furthermore, numerical integration can result in large changes even where changes in parameter values are small; therefore, by identifying the components to which the system is more sensitive, a more robust and reliable model of the system can be developed [177].

As mentioned earlier, the ODE modelling method is based on a mass-action kinetics assumption in which the rate of a reaction is defined to be proportional
3.3 Design and Modelling Methods

to the concentration of the species of the reaction. In cases where this assump-
tion breaks down- for example when spatial gradients of the species are impor-
tant or when the concentrations are so low that stochastic fluctuations in the
system can affect the steady-state and the states are not continuous, other mod-
elling methods are normally used (PDEs or stochastic modelling) [292,305].

Classical modelling of biochemical reactions decompose the complex sig-
nalling networks to a series of simple one or two direction reactions. For each
reaction (both reversible and irreversible) the changes in the concentration of
each component are described by a specific differential equation. Whenever two
successive reactions share a component, the concentration of that component
appear on the RHSs (right hand side) of the differential equations as a positive
or negative term. For a review on the modelling of intracellular reaction kinetics
see [154,408,409].

As an example for the mathematical modelling of a biochemical reaction us-
ing differential equations, a reversible protein-protein binding reaction is illus-
trated in Figure 3.1.

Figure 3.1: A reversible biochemical reaction which shows the binding of two proteins.
\(M_1\) and \(M_2\) bind together to produce \(D\) with the rate of \(k^+\). Similarly, \(k^-\) specifies the
rate of dissociation of \(D\) to \(M_1\) and \(M_2\).

The corresponding biochemical reaction can be written as:

\[ M_1 + M_2 \xrightleftharpoons[k^-]{k^+} D, \]

where \(k^-\) and \(k^+\) represent the kinetic rates of the reaction. Below are the corre-
sponding differential equations for every component in this reaction:

\[
\frac{dM_1}{dt} = -k^+ M_1 M_2 + k^- D \\
\frac{dM_2}{dt} = -k^+ M_1 M_2 + k^- D \\
\frac{dD}{dt} = k^+ M_1 M_2 - k^- D
\]

Each differential equation describes the change of the concentration of a compo-
nent with time. The kinetic rate constants appear on the RHSs.
3.3.2 Reduction, Scaling and Non-dimensionalization Methods

Complexity and the multiplicity of the components involved in a signalling network (more specifically intracellular signalling networks), often make models difficult to solve [146]. To tackle this problem, several reduction and simplification methods have been developed [115, 146]. Separating the reactions according to their time-scales and dividing the system to two subsystems (fast and slow) is one effective way of reducing the model variables and equations. In biochemistry, this method (the time-scale based reduction method) is known as the "rapid equilibrium assumption" [115, 203]. The rapid equilibrium assumption is based on the "singular perturbation" theory for nonlinear systems [272]. Both mass-action and Michaelis-Menten kinetics in biochemistry can be treated using singular perturbation methods (see [272] for more details). In an enzymic reaction, the small dimensionless parameter \( \varepsilon \), which appears in the singular perturbation format of the equations, is defined as the ratio of the initial concentrations of the enzyme and substrate [272].

In signalling networks, the system often evolves on different timescales, this allows us to benefit from the rapid equilibrium assumption in order to reduce and simplify the equations. We have used this method to decrease the complexity of our models.

One of the most important and challenging part of modelling signalling pathways is the matching of dimensions and units for the variables and kinetic rates in the model equations. To avoid mistakes in calculating the units of the variables and kinetic rates, we have used a non-dimensionalization method to substitute dimensionless variables in the equations [65,115]. Non-dimensionalization [65,115] can be used to ensure that every term in a model is consistent and all parameters are defined in appropriate units, this process is often followed by scaling of the variables and their coefficients. The parametrization of a model (specifically, the models described by differential equations) can be simplified via non-dimensionalization method. The non-dimensionalization method and its benefits are described in more detail in Chapter 4, section 4.4.

Due to the different levels of protein concentrations, detected during Western blot quantitation, it is important to use scaling methods. The experimental data are scaled to allow comparison with the simulation results.

Reduction Method

The greatest advantage of reduction methods in Systems biology is that they reduce the number of components and elements involved in the model and thus, the complexity of the model. It is possible to eliminate the components that do not significantly affect the dynamics and steady-state of the system and are difficult to measure quantitatively. The reduction method used in this thesis is "rapid equilibrium assumption". This method relies on the separation of the variables (or the reactions) in the system into "fast" and "slow". Time-scale separation
of the biochemical reactions allows one to simplify the kinetics at steady-state (or equilibrium). This section provides examples of applying rapid equilibrium assumption on biochemical systems in Chapters 4 and 5.

For instance, consider the high affinity dimerization process of \( R_2 \) (TGF–β receptor type two, see Chapter 4 for more details) that is much faster than the other TGF–β signalling reactions. In comparison to every other reaction in the TGF–β system, \( R_2 \) dimerization time scale is large enough to enable us to use the rapid equilibrium assumptions.

The time scale of each biochemical reaction can be calculated from the corresponding differential equations of the participating components. The time scale of a conversion reaction \( (A \xrightleftharpoons[k^{-}]^{k^{+}}B) \), for instance, is defined as the following:

\[
\tau = \frac{1}{k^{-} + k^{+}},
\]

which is calculated via solving the differential equation of \( A \), i.e. \( \frac{dA}{dt} = -k^{+}A + k^{-}B \), considering the fact that the total amount of \( A + B \) is constant in time. Note that \( \tau \ll 1 \) if and only if \( \begin{cases} k^{-} \gg 1 \\ or \\ k^{+} \gg 1 \end{cases} \). However, it is easier for biologists to work with off rate, \( k^{-} \). When, \( k^{-} \gg 1 \), \( \tau \) can be estimated with the inverse \( k^{-} \):

\[
\tau \simeq \frac{1}{k^{-}}.
\]

In case of \( \tau \gg T_{i} \) it is allowed to implement rapid equilibrium assumptions into our system. Here, \( T_{i} \) represents time scales of all other reactions of the system.

Our initial receptor model is illustrated in the following reactions: (In this section all primes denote the differentiation with respect to time.)

\[
R_2 + R_2 \xrightleftharpoons[K_2]{K_2} R_2 \\
R_1 + R_1 + R_2 \xrightleftharpoons[K_R]{K_R} R_{C} \tag{3.1}
\]

Equation 3.2 consists of two successive steps. In fact, the effect of \( R_1 \) dimerization is hidden in equation 3.2. In order to apply the rapid equilibrium assumption on our system, equation 3.2 should break into two steps:

\[
R_1 + R_1 \xrightleftharpoons[K_1]{K_1} R_{1\prime} \\
R_2 + R_2 \xrightleftharpoons[K_2]{K_2} R_2 \\
R_{1\prime} + R_2 \xrightleftharpoons[K_R]{K_R} R_{C} \tag{3.3}
\]
We assume all these reactions are fast. According to equations 3.3, the differential equations describing $R_1$ and $R_{1_2}$ dynamics will be as the following:

\[
R_1' = -2k^+R_1^2 + 2k^-R_{1_2} \quad (3.6)
\]

\[
R_{1_2}' = k^+R_1^2 - k^-R_{1_2} , \quad (3.7)
\]

where $r_{R_1}$ and $r_{R_{1_2}}$ represent other terms of the equations such as, production and degradation of each substance or the association and dissociation of the successive components. $k^+$ and $k^-$ represent the forward and respectively, backward kinetic rates for the $R_1$ and $R_{1_2}$ dimerization reactions. Next, we apply the conservation law on $R_1$ as a substance of the reaction 3.3:

\[
R_{1_{T}} = R_1 + 2R_{1_2} \quad (3.8)
\]

The conservation law for $R_1$ states that the total amount of $R_1$, by definition, equals to the total amount of $R_1$ as monomers adding by twice of the amount of $R_1$ as dimmers (which is twice of $R_{1_2}$). This law is true in every condition. At equilibrium, the total flux for each reaction must be zero, i.e. the right arrow flux should be equal to left arrow flux (equation 3.3):

\[
k^+R_1^2 = k^-R_{1_2}
\]

A new term is defined; $K = \frac{k^-}{k^+}$ which is called the “equilibrium constant” and has units of "concentration". Consequently, $R_{1_2}$ can be rewritten in terms of $R_1$ and $K$, $R_{1_2} = \frac{R_1^2}{K}$.

At equilibrium, the equation 3.8 can be written as:

\[
R_{1_{T}} = R_1 + 2 \frac{R_1^2}{K}
\]

Now if we differentiate this equation with respect to time (the independent variable) and subsequently we will have:

\[
R_{1_{T}}' = R_1' + 4 \frac{R_1}{K}R_1' \quad (3.9)
\]

On the other hand, $R_1'$ and $R_{1_2}'$ have already been defined via stoichiometry equations (equations 3.6 and 3.7). Substituting equations 3.6 and 3.7 into the recent equation for $R_{1_{T}}'$ (equation 3.9), we would have another description for $R_{1_{T}}'$:

\[
R_{1_{T}}' = -2k^+R_1^2 + 2k^-R_{1_2} + r_{R_1} + 2k^+R_1^2 - 2k^-R_{1_2} + 2r_{R_{1_2}} \text{ or } (3.10)
\]

\[
R_{1_{T}}' = r_{R_1} + 2r_{R_{1_2}}
\]

Combine the equations 3.9 and 3.10, a single equation will be derived:
This equation propose a differential equation for $R_1$. In fact, we have reduced a system of two coupled differential equations (equations 3.6 and 3.7) into a single differential equation for $R_1$ and hence, eliminated the direct dependence of the system on one of the variables, $R_{12}$.

The left hand side factor of equation 3.11, $(1 + 4\frac{R_1}{K})$, summarizes what happens during the dimerization process while, the right hand side describes the other reactions that affect the $R_1$ dynamics directly or through affecting $R_{12}$ dynamics, for instance $R_1$’s production and degradation reactions.

At equilibrium, the changes in the total amount of $R_1$ in time ($R_{1T}$) must be zero. Therefore, the left hand side of the equation equals zero and, it can be stated that equation below is the equilibrium condition of the dimerization reaction:

$$r_{R_1} = -2r_{R_{12}}$$

The nullclines in the slope field (phase portrait) is shown by $R_1' = 0$, thus it is only $r_{R_1} + 2r_{R_{12}}$ term that determines the nullclines. The velocity vectors longitude is proportional to $K$ through the $R_1' = (r_{R_1} + 2r_{R_{12}})/(1 + 4\frac{R_1}{K})$ equation.

The calculations above show the procedure of applying the rapid equilibrium assumptions on each fast reaction. We will benefit from this reduction method in Chapter 4 and 5 because of its advantages. These assumptions can be made only if the time scale of the reaction is remarkably greater than other reactions of the system.

Equations 3.3 and 3.4 follow the same model of dimerization. Therefore, the reductions applying to $R_2$ are the same as $R_1$. The final form of the reduced version of $R_1$ and $R_2$ dimerization would be as the following:

$$\begin{align*}
(1 + 4\frac{R_1}{K_1})R_1' &= r_{R_1} + 2r_{R_{12}} \\
(1 + 4\frac{R_2}{K_2})R_2' &= r_{R_2} + 2r_{R_{22}}
\end{align*}$$

The dimerization reaction between identical species is called "fast homodimerization" (see equations 3.3 and 3.4). If we have a closer look at equation 3.5, it is clear that it can be interpreted as a dimerization reaction but, the substances $R_{12}$ and $R_{22}$ are not of the same kind. This type of dimerization is known as "fast hetero-dimerization". In spite of the distinction between fast homo and hetero dimerization reactions, they can be treated similarly. Therefore, we apply rapid equilibrium assumption on the reaction 3.5. The differential equations describing the reaction 3.5 are:
\begin{align*}
R_{12}' &= -k^+ R_{12} R_{22} + k^- R_C + r_I \\
R_{22}' &= -k^+ R_{12} R_{22} + k^- R_C + r_{II} \\
R_C' &= k^+ R_{12} R_{22} - k^- R_C + r_C
\end{align*}

Similar to previous case (R1 dimerization) \(r_I, r_{II}\) and \(r_C\) denote the other terms of the equations. Now, we have two equations for totals:

\begin{align*}
R_{12T} &= R_{12} + R_C \\
R_{22T} &= R_{22} + R_C
\end{align*}

The final goal is to eliminate \(R_C\) dependency of these two equations (equations 3.17 and 3.18), assuming that reaction 3.5 happens relatively fast. At equilibrium we have:

\[ R_C = \frac{R_{12} R_{22}}{K}, \quad K = \frac{k^-}{k^+} \]

Substituting \(R_C\) to equations 3.17 and 3.18 we have:

\begin{align*}
R_{12T} &= R_{12} + \frac{R_{12} R_{22}}{K} \\
R_{22T} &= R_{22} + \frac{R_{12} R_{22}}{K}
\end{align*}

Now we repeat the differentiation step for both equations.

\begin{align*}
R_{12T}' &= R_{12}' + R_C' \\
R_{22T}' &= R_{22}' + R_C'
\end{align*}

We replace the terms from the equations 3.14 and 3.15:

\begin{align*}
R_{12T}' &= r_I + r_c \\
R_{22T}' &= r_{II} + r_c
\end{align*}

Now we differentiate equations 3.19 and 3.20, while \(R_C\) is replaced:

\begin{align*}
R_{12T}' &= R_{12}' + \frac{R_{22} R_{12}'}{K} + \frac{R_{12} R_{22}'}{K} \\
R_{22T}' &= R_{22}' + \frac{R_{22} R_{12}'}{K} + \frac{R_{12} R_{22}'}{K}
\end{align*}

Therefore, we write the following equations:
3.4 Validation Methods

The following methodologies are used and explained in more details in subsequent chapters (Chapter 4 (section 4.5), 5 (section 5.6) and 6 (section 6.4)).

3.4.1 Quantitation Methods

Quantification of protein expression using Western blot experiments is essential for mathematical modelling the signalling networks. ImageJ, a Java-based program, is one of the most common programs used for quantitation of blots. The blots can be converted to 8-bit grayscale in ImageJ analyses of standard Western blot figures in publications. Each row in a blot needs to be normalized to the relative loading control, however, normalization can also be performed with respect to a protein that remains constant e.g. total SMAD2/3 in TGF-β and total STAT3 in IL – 6 signalling pathways. The quantification methods are explained in [140,429].

3.4.2 Estimation and Curve Fitting Methods

In order to validate our proposed models with experimental data, we have used the curve fitting toolbox of MATLAB [438, 467]. This toolbox lets the user perform data and regression analysis and provides adjustable initial conditions and parameter values to improve the fittings [157].
Chapter 4

TGF−β Receptor Model of Signal Switching in Cancer

4.0 Summary of the information included in Chapter 4

- Background
  Transforming growth factor β, TGF−β, signalling regulates the development of embryos and tissue homeostasis in adults. In conjunction with other oncogenic changes long-term perturbation of TGF−β signalling is associated with cancer metastasis. Although TGF−β signalling can be complex, many of the signalling components are well defined by experiment; so it is possible to develop predictive models of TGF−β signalling using reduction and scaling methods. The parameterization of our TGF−β signalling model is consistent with the experimental results.

- Results
  By adding time-delayed positive feedback to the inherent time-delayed negative feedback for the TGF−β signalling - even when we used the minimal reaction set, we were able to simulate the sigmoidal, switch-like behaviour observed for the concentration dependence of long-term (> 3 hours) TGF−β stimulation. By using a “rapid equilibrium assumption” to reduce the network of TGF−β signalling reactions and the time scales of the individual reactions and the inclusion of both negative and positive feedback loops, we have developed a mathematical model for the TGF−β signalling pathway, i.e. the simplified model of TGF−β signalling. Computer simulations reveal the vital role of the coupling of positive and negative feedback loops on the regulation of the TGF−β signalling system. The incorporation of time-delays for the negative feedback loop improved the accuracy, stability and robustness of the model. This model reproduces
both the short-term and long-term switching responses for the intracellular signalling pathways at different TGF–β concentrations. We have tested the model against experimental data from MEF (mouse embryonic fibroblasts) WT, SV40-immortalized MEFs and Gp130$^{F/F}$ MEFs. The productions from the simplified model are consistent with the experimental data.

- Conclusions

The signalling feedback loops are required to model TGF–β signal transduction and its effects on normal and cancer cells. We focus on the effects of time-delayed feedback loops and their coupling to ligand stimulation in this system. The model was simplified and reduced to its key components using standard methods and the rapid equilibrium assumption. We detected differences in short-term and long-term signal switching via consideration of positive and negative feedback loops in our simplified model of TGF–β signalling. The results from the model are compared to experimental data and provide predictions for TGF–β signalling in cancer cells with different mutations.

4.1 Introduction

TGF–β is a member of the transforming growth factor superfamily, which also includes other growth factors such as bone morphogenetic proteins, Mullerian inhibitory substance, activin, inhibit and Nodal [75,296,396]. Each family member controls a broad range of cellular processes, such as differentiation, proliferation, migration, life span and apoptosis [118, 296]. TGF–β is secreted in an inactive form and sequestered in the extracellular matrix, but once activated by serine and metalloproteinases [208], TGF–β binds to cell surface TGF–β receptor complexes. The active ligand:receptor complex now initiates intracellular signalling that leads to SMAD activation and nucleocytoplasmic shuttling and, eventually, gene responses in the nucleus [300,433].

The TGF–β receptor complex is a tetramer comprised of Type 1 and Type 2 receptors that upon TGF–β binding becomes activated via autophosphorylation [178,296,471]. The activated TGF–β receptor complex is then internalized [169,301], where it phosphorylates and activates SMAD3 [178]. Activated SMAD3 then forms homotrimers, which bind to SMAD4 homotrimers and then imported into the nucleus [507]. These SMAD3:SMAD4 complexes function as transcription factors that up-regulate a number of target genes, including Jun, Fos, SNAIL1 and SMAD7, known inhibitor of TGF–β Type 1 receptors and TGF–β receptor signalling [323,451,507].

Recent studies indicate that TGF–β concentration, stimulation time, cell type and even the percentage of active signalling components can influence the
4.1 Introduction

gene responses, giving a multi-functional aspect to TGF−β signalling [75, 297]. This is of particular interest in cancer, where SMAD signalling is the most critical pathway controlling the transition of normal epithelial cells to cancerous cells in the colon [206, 297, 299, 330, 396]. In spite of the myriad studies on the TGF−β signalling pathway, there are still many unanswered questions concerning the impact of TGF−β signalling at different stages of cancer cell progression [18]. In particular, there are two opposing reactions of cancer cells to TGF−β: the proliferation of cancer cells at an early stage is inhibited by TGF−β [265], yet at more advanced stages of malignancy, proliferation of cancer cells is stimulated by this signalling protein [198].

Although TGF−β signalling components were discovered decades ago [13], the quantitative aspects, dynamics and locations of the signalling components that occur within hours of TGF−β stimulation [72, 74, 308, 386, 449, 511, 513] have been more difficult to understand. This has motivated the development of a number of mathematical models of TGF−β signalling [19, 72, 74, 308, 386, 449, 512, 513] including a recent model [511]. In a comprehensive model, Zi et al. explains the high cooperativity and discontinuous cellular response to TGF−β in terms of switch-like behaviour arising from ligand depletion. All of these models, however, lack inhibitory feedback mechanisms known to regulate the TGF−β system, including feedback through SMAD7, a key inhibitor in TGF−β signal transduction [323]. Furthermore, SMAD7 is an important component for mediating the crosstalk between TGF−β signal transduction and other cytokine signalling pathways such as IL-6 or IL-11 [206].

These models also lack the more recently discovered positive feedback in TGF−β signalling that acts by suppressing Azin1 via a microRNA (miR-433) [269]. Azin1 promotes polyamine synthesis [212, 269], which suppresses TGF−β signalling [269, 277, 345, 365]. Azin1 accomplishes this by inhibiting antizyme, thus preventing the degradation of ornithine decarboxylase (ODC) [212, 269]. ODC is essential for the biosynthesis of polyamines [212, 269] (see Figure 4.1). Interestingly, over-expression of Azin1 suppresses the expression of TGF−β and its Type 1 receptor [269]. The miR-433:Azin1:Antizyme:ODC reactions induce a positive feedback control of TGF−β signalling.

Inclusion of such feedbacks is particularly important given that it is well understood that positive and negative feedback loops, particularly in the presence of time-delays (typically found in cellular signalling systems involving gene regulation and shuttling of signalling components between compartments) can also produce both cooperativity and switch-like behaviour even in the absence of ligand depletion [120, 121, 393]. As a prelude to improving our understanding of the TGF−β signalling system we have developed a new mathematical model of the TGF−β receptor system incorporating negative feedback control via SMAD signalling, positive feedback via Azin1 and transcriptional time-delays [248, 323]. To that end, here we present a detailed model incorporating the main mechanisms of TGF−β and SMAD signalling, including feed-
backs and time-delays, but then reduce our model to a simpler system that is more amenable to both standard mathematical analysis techniques and simulation [459]. This permits us to fully characterise analytically the system at steady-state while also investigating the transient dynamics of the system in response to TGF−β signals.

4.2 Model Development

Signalling systems like the TGF−β pathway can be modelled using differential equations describing time rate of the concentration changes of the various cellular components (e.g TGF−β receptors, SMAD4) in terms of reaction rates corresponding to specific kinetic mechanisms, for example, first order or Henri-Michaelis-Menten [115]. These are shown for the TGF−β signalling system in Figure 4.1. TGF−β receptor activation starts with the dimerization of both components (TGF−β receptor type 1 and 2, called respectively R1 and R2). Extensive studies have discovered that receptor dimers initiate the signalling processes [102, 284]. The R2 dimer binds to the R1 dimer, resulting in the receptor complex RC. The RC complex binds TGF−β dimers in the medium around the cells (see Figure 4.1). This TGF−β:RC complex (LC) contains all the components essential for signalling, however, R1s are not activated (phosphorylated) so LC is not the membrane transducer of the exogenous TGF−β signal. Activation of LC requires ligand stimulated phosphorylation of R1 by R2 to produce a fully phosphorylated ligand-receptor complex (PC). PC initiates SMAD signalling via stimulation of the SMAD phosphorylation processes. The activation processes start with the phosphorylation of SMAD2/3, the formation of (PSMAD2/3)3,(SMAD4)3 [481], the translocation (Phospho-SMAD2/3)3 to the nucleus, stimulation of the SMAD7 gene and expression of the miR-433 microRNA. Accordingly, SMAD7 is transcribed, its mRNA translated and eventually the SMAD7/SMURF complex amplifies the degradation of the R1-associated membrane components (note that only SMAD3 is considered in Figure 4.1 due to its crucial role in SMAD7 and miR-433 activation). Note that the components and reactions are estimations of the most frequent occupancy among all the possibilities in biology. For instance, receptor dimerizations of type1 and 2 receptors on the membrane are reported in different orders in the literature [156,219,237,298], however, considering the significantly short time scale of the receptor dimerization reactions comparing to other intracellular reactions, altering among the different dimerization orders does not change the steady-state and equilibria of the system.
4.2 Model Development

Figure 4.1: The full TGF-β signalling biological model. Potential phosphorylation sites of the receptors are specified with empty circles attached to R1 and R2 components. Arrows pointing to 6 blue dots represent degradation process. The red solid arrows originating from SMAD7/Smurf apply negative and/or positive feedback on the receptor components of the membrane. Oval-shaped components written in small letters represent micro-RNAs. In this figure, S represents the SMAD proteins. Note that the arrow from ODC to polyamine shows a stimulatory reaction rather than conversion.

4.2.1 The Full Model

In order to simplify the intracellular reactions involving in TGF-β signalling, we have focused on the receptor components and their interactions on the membrane. Simulation of these receptor processes is achieved by collapsing the SMAD signalling interactions (e.g. nucleocytoplasmic shuttling of activated SMAD complexes and transcription and translation of feedback-associated proteins, such as SMAD7) to a single ligand dependent feedback loop that is originated from PSMAD trimer, S₃. This simplifies the initial modelling equations and focusses the feedback reactions on membrane components. The real feed-
back loop in the Full Model of TGF–β signalling is the result of a sequence of back-to-back, coupled reactions explained above (see Figure 4.1). Besides, each one of the intracellular processes happens within a specific time interval and at a defined kinetic rate. In order to summarize all the cytoplasmic and nuclear reactions in one stimulatory/inhibitory reaction, a significant time-delay needs to be included. Our model for TGF–β signalling is shown in Figure 4.2.

Altering from Figure 4.1 to Figure 4.2 a few assumptions have been made. Primarily, $\hat{S}$ is replaced with all the phosphorylated SMAD3 in the cytoplasm, while the nuclear PSMAD3 is represented via $S_n$ in Figure 4.2. SMAD4 is the common-mediator SMAD that participates in the TGF–β signalling only by interacting with PSMAD2/3. Therefore, it is possible to summarize the SMAD4 role in TGF–β signalling in $\hat{S}$. Total (PSMAD3)\textsubscript{3}(SMAD4)\textsubscript{3} concentration is represented with $S_3$ in Figure 4.2. The negative feedback cascade via SMAD7 is initiated from transcriptional active SMAD complex in the nucleus and is also reducible to $S_3$ component. However, $S_3$ is presented as a dimer in relative terms of negative feedback in the equations 4.1 in order to imitate the SMAD7:SMURF interaction. The positive feedback, on the other hand, consists of a chain of biochemical reactions which are triggered by nuclear (PSMAD3)\textsubscript{3}(SMAD4)\textsubscript{3}. As a result, Azin1:Antizyme:ODC:Polyamine reactions can be represented via a single intermediate inhibitor, i.e. $P$.

Figure 4.2: The full receptor model, TGF–β signal transduction. The red dashed lines which originate from phosphorylated SMAD trimer indirectly regulate the receptor levels. All the reactions from trimerization of phospho-SMAD2/3 to SMAD7 transcription and translation are reduced to the red dashed lines (see Figure 4.1 for clarification). The dotted ends of red dashed lines show that included reactions could lead to both inhibition and stimulation of their targeting reactions (demonstrating negative and positive feedback effects). In this figure $S$ is specifically used for SMAD2/3

Referring to the model Figure 4.2, the Full Model reactions are:
4.2 Model Development

\[
\begin{align*}
\text{R1} + \text{R1} & \xrightleftharpoons[k_1^{-}]{k_1^{+}} \text{R1}_2 \\
\text{R2} + \text{R2} & \xrightleftharpoons[k_2^{-}]{k_2^{+}} \text{R2}_2 \\
\text{R1}_2 + \text{R2} & \xrightleftharpoons[k_{RC}^{-}]{k_{RC}^{+}} \text{RC} \\
\text{RC} + \text{TGF}_2 & \xrightleftharpoons[k_{LC}^{-}]{k_{LC}^{+}} \text{LC} \\
\text{LC} & \xrightarrow[k_{PC}^{-}]{k_{PC}^{+}} \text{PC} \\
\text{PC} & \xrightarrow[k_{3}^{-}]{k_{3}^{+}} \text{...}
\end{align*}
\]

This leads to the following system of differential equations:

\[
\begin{align*}
\frac{d[\text{R1}]}{dt} &= v_1 - k_1 \cdot [\text{R1}] - k_1^{-} \cdot [N]^2 \cdot \frac{[\text{R1}]}{[\text{R1}] + K} - \\
&\quad 2k_1^{+} \cdot [\text{R1}] \cdot [\text{R1}] + 2k_1^{-} \cdot [\text{R1}_2] - k_1^{+} \cdot [P] \cdot \frac{[\text{R1}]}{[\text{R1}] + K} \\
\frac{d[\text{R2}]}{dt} &= v_2 - k_2 \cdot [\text{R2}] - 2k_2^{+} \cdot [\text{R2}] \cdot [\text{R2}] + 2k_2^{-} \cdot [\text{R2}_2] \\
\frac{d[\text{R1}_2]}{dt} &= k_1^{+} \cdot [\text{R1}] \cdot [\text{R1}] - k_1^{-} \cdot [\text{R1}_2] - k_{RC}^{-} \cdot [\text{R1}_2] \cdot [\text{R2}] + k_{RC}^{+} \cdot [\text{RC}] \\
\frac{d[\text{R2}_2]}{dt} &= k_2^{+} \cdot [\text{R1}] \cdot [\text{R1}] - k_2^{-} \cdot [\text{R2}_2] - k_{RC}^{-} \cdot [\text{R1}_2] \cdot [\text{R2}] + k_{RC}^{+} \cdot [\text{RC}] \\
\frac{d[\text{RC}]}{dt} &= k_{RC}^{+} \cdot [\text{R1}_2] \cdot [\text{R2}_2] - k_{RC}^{-} \cdot [\text{RC}] - \\
&\quad k_{LC}^{+} \cdot [\text{RC}] \cdot [\text{TGF}_2] + k_{LC}^{-} \cdot [\text{LC}] - k_{RC}^{-} \cdot [\text{RC}] - k_{RC}^{+} \cdot [N]^2 \cdot \frac{[\text{RC}]}{[\text{RC}] + K} \\
\frac{d[\text{LC}]}{dt} &= k_{LC}^{+} \cdot [\text{RC}] \cdot [\text{TGF}_2] - k_{LC}^{-} \cdot [\text{LC}] - k_{PC}^{+} \cdot [\text{LC}] + k_{PC}^{-} \cdot [\text{PC}] \\
\frac{d[\text{PC}]}{dt} &= k_{PC}^{+} \cdot [\text{LC}] - k_{PC}^{-} \cdot [\text{PC}] - k_{PC}^{-} \cdot [\text{PC}] - k_{PC}^{+} \cdot [N]^2 \cdot \frac{[\text{PC}]}{[\text{PC}] + K} \\
\frac{d[\text{S}]}{dt} &= v_S - k_S \cdot [\text{S}] - k_S^{+} \cdot [\text{PC}] \cdot \frac{[\text{S}]}{[\text{S}] + K_S} + k_S^{-} \cdot [\hat{S}] \\
\frac{d[\hat{S}]}{dt} &= k_S^{+} \cdot [\text{PC}] \cdot \frac{[\text{S}]}{[\text{S}] + K_S} - k_S^{-} \cdot [\hat{S}] - k_n^{+} \cdot [\hat{S}] + k_n^{-} \cdot [\text{S}_n] \\
\frac{d[\text{S}_n]}{dt} &= k_n^{+} \cdot [\hat{S}] - k_n^{-} \cdot [\text{S}_n] - 3k_3^{+} \cdot [\text{S}_n]^3 + 3k_3^{-} \cdot [\text{S}_3] \\
\frac{d[\text{S}_3]}{dt} &= k_3^{+} \cdot [\text{S}_n]^3 - k_3^{-} \cdot [\text{S}_3] - k_3 \cdot [\text{S}_3]
\end{align*}
\]

where \([P] = K_1 \cdot (K_1^{-} + [S_3] \cdot (t - \tau))^2\) and \([N] = [S_3] \cdot (t - \tau)\), the positive and negative feedback intermediate components, respectively.

The purpose of the full model is to document all of the reactions we considered to be significant in the regulation of TGF-\(\beta\) signalling. Analysis of the reactions determines which reactions should remain in the reduced model and
4.2 Model Development

which could be eliminated via the rapid equilibrium assumption. We will not solve the full model because it is not necessary to consider parameters which do not contribute in the determination of the steady-state of the model. Having the full model allows the readers to appreciate why and how the model is reduced.

4.2.2 The Simplified Model

The procedure of designing an effective mathematical model of the TGF−β signalling pathway incorporates several consecutive steps. The modelling was started from a very detailed, complex model of all the protein interactions and reactions on the membrane, in the cytoplasm and the nucleus (see Figure 4.1). In the full receptor model, several reactions with long time scales became summarized into feedback loops which originate from PSMAD trimer. We then simplified the receptor model to 6 differential equations. It is assumed that the R1 and R2 dynamics are similar, hence the individual components were replaced by a receptor block, R. R then become dimerized to form RC. Besides, LC and PC are assumed to to combine in one parameter, i.e. PC, since they approximately follow the same kinetics. In the SMAD activation section, an intermediate step, S, has been added so as to mimic the nuclear accumulation of phosphorylated SMAD, in comparison with the abundance of unphosphorylated SMAD in the cytosol. Note that these assumptions are used in order to provide a simplified view of the Full model (Figure 4.1). No reduction method has been used up to this point. Eventually, the reactions describing the Simplified Model could be represented as below:

\[
\begin{align*}
R + R & \xrightleftharpoons[k_{RC}]{k_{RC}} RC & R & \xrightarrow[k_{2}]{k_{2}} \cdots \\
RC + TGF_2 & \xrightarrow[k_{PC}^+]{k_{PC}^-} PC & PC & \xrightarrow[k_{PC}^{-}]{} \cdots \\
S & \xrightarrow[k_{S}]{k_{S}^+} \hat{S} & S & \xrightarrow[k_{S}^{-}]{} \cdots \\
\hat{S} & \xrightarrow[k_{n}]{k_{n}^+} S_n & S_n & \xrightarrow[k_{Sn}]{k_{Sn}^{-}} \cdots
\end{align*}
\]

Figure 4.3 highlights the assumptions we used to simplify our receptor model. Here, Figure 4.3A describes the reaction system corresponding to Figure 4.2, these reactions are used to determine the equations defining the mathematical model of TGF−β signalling. By comparing the full receptor model (Figure 4.3A) and the simplified receptor model (Figure 4.3B), the key components in the TGF−β receptor signalling model are specified. In other words, our reduction/simplification method functions similar to sensitivity analysis in keeping
the critical components in a signalling pathway.

![Figure 4.3](image)

**Figure 4.3:** TGF–β receptor signalling system. A) The schematic receptor model, TGF–β signal transduction. TGF and $\hat{S} + 3S_3$ represent the input and the output of the model, respectively. B) A Simplified Model of TGF–β signal transduction. TGF–β and $\hat{S} + S_n + 3S_3$ represent the input and the output of the model, respectively.

A set of delayed differential equations describing the Simplified Model are introduced via equations 4.3. The solutions to these equations are at equilibrium when the left hand value is zero, i.e. where the derivatives of each variable are zero with respect to time. The component $P$ represents the molecules through which the positive feedback acts on the receptors (Azin1:Antizyme:ODC:Polyamine in Figure 4.1). The positive feedback is indirect, being affected by two coupled, inhibitory processes.

To achieve the most biologically compatible and robust model of TGF–β signalling, the place of action of the individual feedback reactions contains information which affects the responses of the system. Sensitivity analysis identify LC as the negative feedback action point. SMAD7 binds to receptors and participates in the induction of E3 ubiquitin (Ub) ligase-mediated receptor ubiquitination [101,226]. Michaelis-Menten kinetics is used to model the negative feedback inhibitory function. It is been reported that polyamine depletion increases the TGF–β type 1 receptor mRNA and increases the sensitivity of cells to TGF–β-mediated growth inhibition [269,277,365]. Consequently, we have modelled the positive feedback successive reactions using two inhibitory reactions: first, the inhibition the intermediate inhibitor $P$ via miR-433 and second, the inhibition of R via $P$.

Although some cooperativity within the system originates from several dimer-
izations and trimerizations on the membrane, in the cytosol and in the nucleus, the most critical cooperativity comes from the trimerization of the Phosphorylated SMAD3 and the binding of these oligomers to the SMAD4 trimer and the consequent stimulation of the miR-433 and SMAD7 nuclear transcription. It should be noted that the trimerization of Phospho-SMADs influences cooperativity in both the positive and negative feedback loops (see Figure 4.1 for more details).

The time-delay is considered in the equations representing the dynamics of $S_3$. Hence, time-delays have been applied to both the positive and negative feedbacks. The time-delay compensates for the SMAD nucleocytoplastic shuttling and other reactions that have been consolidated in the reduced and simplified model (e.g. SMAD7 transcription and translation and the miR-433/Azin1/Antizyme/ODC reactions).

For simplification, the time-delays and amplitudes of the positive and negative feedbacks are assumed to be identical, however, it is always feasible to adjust these parameters when reliable experimental data becomes available.

\[
\frac{d[R]}{dt} = v_1 - k_1[R] - 2k_{RC}^+[R]^2 + 2k_{RC}^-[RC] - k_1^{t+}[P]\frac{[R]}{[R] + K} - k_1^{t-}[N]^2\frac{[R]}{[R] + K}
\]
\[
\frac{d[RC]}{dt} = k_{RC}^[+][R]^2 - k_{RC}^-[RC] - k_{RC}^-[RC] - k_{PC}^[+][TGF_2][RC] + k_{PC}^-[PC] - k_{RC}^[t-][N]^2\frac{[RC]}{[RC] + K}
\]
\[
\frac{d[PC]}{dt} = k_{PC}^[+][TGF_2][RC] - k_{PC}^-[PC][PC] - k_{PC}^[t-][N]^2\frac{[PC]}{[PC] + K}
\]
\[
\frac{d[S]}{dt} = v_S - k_S[S] - k_S^[t+][PC]\frac{[S]}{[S] + K_S} + k_S^[t-][\hat{S}]
\]
\[
\frac{d[\hat{S}]}{dt} = k_S^[t+][PC]\frac{[S]}{[S] + K_S} - k_S^[t-][\hat{S}] - k_n^[t+][\hat{S}] + k_n^[t-][S_n] - k_S[S]
\]
\[
\frac{d[S_n]}{dt} = k_n^[t+][\hat{S}] - k_n^[t-][S_n] - k_n[S_n]
\]

where again, $[S_3] = [S_n]^3/K_3$, $[N] = [S_3](t - \tau)$ and $[P] = K_1^2/(K_1^2 + [S_3](t - \tau))^2$.

The parameters $k_1^[t-], k_{RC}^[t-]$ and $k_{PC}^[t-]$ represent, respectively, the strength of the negative feedback on $R$, $R_C$ and $P_C$, the R1-associated membrane complexes. Although we have applied the negative feedback on $R$, $R_C$ and $P_C$ simultaneously and with identical strengths and binding constants, the feedback on $P_C$ is what produces the switching behaviour (see section 4). The positive feedback is on $R$ only, where Polyamine acts [269, 365], so that the cooperativity of the system originates from the coupling of the self-regulatory positive and negative feedback rather than from extracellular effects such as ligand dimerization or depletion.
4.3 Numerical Simulations

In the work that follows, all calculations are performed with the simplified model where concentrations are dimensionless and scaled such that \( v_1 = 1 \). The details of the scaling and the dimension-less parameters are described in section 4.4. The most interesting changes in total PSMAD concentration occur between \( O(0.1) \) and \( O(1) \), TGF–β varies between 0 and \( O(10) \) and receptor concentration varies between 0 and \( O(1) \). This is reasonable given that there is insufficient data to constrain the model, particularly for the membrane bound receptor components.

Total PSMAD concentration \( \hat{S} \) is defined as:

\[
\hat{S} + \frac{V_n}{V_c}(\hat{S}_n + [S_3])
\]

, where \( V_n \) and \( V_c \) are defined as the volume of the nucleus and the cytoplasm compartment, respectively. \( [S_3] = [S_n]^3/K_3 \) and \( [S_n] \) is calculated from the final equation of 4.3.

4.3 Numerical Simulations

Analyses of the reduced equations and scaling make it possible to study the performance of the model at steady-state in detail (also see “The Importance of Feedback” in Supplementary Material). Our model uses six coupled differential equations to represent all the reactions occurring on the membrane and within the SMAD signalling cascade.

The Zi et al. model produced sigmoidal TGF–β concentration dependence for the cellular responses to long-term stimulation [511]. The total concentration of SMAD was used as an interpretation of the final cell response. According to their results [511], the Hill coefficient of the fitted curve to the cell responses to long-time TGF–β stimulation was approximately 4.5. The Zi et al. model’s short-term (transient) responses to TGF–β followed the Hill equation with an approximate coefficient of 0.8 [511]. Zi et al. proposed that the reason for such a dramatic change in the behaviour of the system was due to a significant time-dependent ligand depletion through ligand-receptor interaction [511]. Moreover, the level of total PSMAD concentration should be higher in the short-term response comparing to the long-term. This difference in the level of responses is due to the demanding overshoot of the PSMAD concentration time course (the peak of PSMAD concentration in Figure 4.4).

In order to test our hypothesis that the positive feedback is responsible for the change of the behaviour of the system from short-term to long-term cellular responses, we repeated the simulations for the same TGF–β and stimulation times. The parameter values used to produce the following figures are shown in Table 4.3 of “Supplementary Material”. Figure 4.4 shows the predicted changes in the PSMAD concentration at different TGF–β concentrations. It is evident that the steady-state level of PSMAD is 40% less than its short-term peak
value. The predicted results shown in Figure 4.4 are consistent with the literature (e.g. [199, 511]) where the total PSMAD concentration peaks one hour after the ligand stimulation. Figure 4.5 and Figure 4.6 show the predicted transient and steady-state responses to the changes in TGF–β concentration in the system. Note that the x axis in Figure 4.5 and Figure 4.6 denote different concentration of TGF–β. Therefore, the simulation time is the same for every point in each figure. The Hill coefficients are 0.85 for the short-term and 3.87 for the long-term stimulation, i.e. similar to Zi et al. [511] (In order to compare the reproduction results see the Figure 5.A and 5.B of [511]. The parameter values are fitted to a single term (Hill coefficient) for Figures 4.5, 4.6, 4.15 and 4.16.). These results support our hypothesis that the coupling of time-delayed positive and negative feedbacks in the TGF–β signal transduction system can account for ultra-sensitivity responses to the ligand concentrations.

Figure 4.4: Total PSMAD time course for TGF–β = 5. The peak in the total PSMAD concentration occurs 50-60 min after the stimulation and corresponds to the short-term (transient) response. The constant level of PSMAD at 0.3 represents the long-term (steady-state) response of the system.

Figure 4.5: Short-term responses of PSMAD levels to different concentrations of TGF–β; transient response. The simulation time for each point in this figure is 50 min. The only parameter of the model which is being changed in this figure is the TGF–β concentration, meaning that each point has a different TGF–β input.
4.3 Numerical Simulations

Figure 4.6: Long-term responses of PSMAD levels to different concentrations of TGF–β; steady-state response. The simulation time for each point in this figure is 1000 min. The only parameter of the model which is being changed in this figure is the TGF–β concentration, meaning that each point has a different TGF–β input.

The negative feedback operates on all of the R1-associated complexes on the membrane. We suggest that it is the negative feedback through PC which regulates the system. PC is the only TGF–β-associated complex in the simplified model for TGF–β signalling. The total TGF–β ligand concentration (extracellular TGF–β, which is constant, and that which is bound within the PC complex) decreases because of the degradation of PC via the basal degradation of, and negative feedback on, PC. The saturation of the system with TGF–β flattens the TGF–β concentration response curves at high concentrations of ligand (Figure 4.5 and 4.6). In order to examine our hypothesis, we conducted a set of simulations with feedback on R and RC removed (Figures 1 and 2 in Supplementary Material). To accomplish this, $k_{f1}$ and $k_{fRC}$ are set to zero. These simulation results corroborates our initial hypothesis that negative feedback acts almost entirely through PC.

We speculate that PSMAD concentration time course in response to TGF–β stimulation is modified in cancer cells due to the possible mutations in SMADs and different receptor levels [124,235,264,462]. Figure 4.7 simultaneously shows the model results of the responses of- what we have called- early and late stage cancer cells to TGF–β stimulation. We simulated the biochemical conditions of the early-stage tumors by reducing the membrane receptor and the SMAD concentrations. More precisely, the production rates of receptors ($v_1$ in Table 4.3) and SMADs ($v_s$ in Table 4.3) were decreased. The simulation response of the total PSMAD time course in cells with lower receptor and SMAD concentrations is plotted in Figure 4.7A. A comparison of Figure 4.7A with Figure 4.4 reveals that PSMAD concentration peaks to a higher level (0.67 rather than 0.5) but flattens to a lower level at steady-state (0.13 v.s. 0.3). This result confirms the compatibility of our simplified receptor model of TGF–β signalling with both normal and cancer cell lines.
Figure 4.7: Total PSMAD time course for a certain TGF–β concentration. A) simulation results for low membrane receptor concentration condition (or so called early-stage tumors) B) simulation results for high membrane receptor concentration condition (or so called late-stage tumors). These conditions were simulated via altering the receptor production rate on the membrane.

In contrast, late-stage tumors express more responsiveness to TGF–β signalling. In order to establish late tumor environment, the receptors and SMADs levels are increased, using relative production rates. Predicted response of late tumors to TGF–β stimulation is shown in Figure 4.7B. Although total PSMAD concentration peaks at a higher level in late tumors, its steady-state level is not significantly lower than the peak, showing that the signalling remains high.

To investigate the role of receptor level in the signalling, we have simulated the behaviour of PSMAD concentration while the receptor concentration increases monotonically. Receptor production rate was increased to achieve an increase in receptor concentration. TGF–β concentration was maintained at a constant level during the experiment. This simulation was conducted for two distinct values of TGF–β: 5 and 2. The second value of TGF–β concentration is located approximately where the switch in the long-term steady-state PSMAD concentration occurs (see Figure 4.6 and Figure 4.16 in Supplementary Material). There was no distinguishable change in the PSMAD steady-state concentration when TGF–β concentration was reduced (Figure 4.8). Low receptor concen-
4.3 Numerical Simulations

The flat starts in both panels of Figure 4.8 show that the cells are insensitive to TGF−β signalling when the receptor copy numbers are very low, i.e. the situation in cancer cells. The peak determines the receptor concentration in which the highest level of signal occurs. At high concentration of the receptors, the PSMAD level reaches a saturation level and stays there as the receptor concentration increases. When the TGF−β concentration is higher (Figure 4.8A) the peak levels of PSMAD shift to the left and all the changes happen at lower receptor levels. According to the model formulation, negative feedback term is directly proportional to $-S_3^2$, while positive feedback element changes proportional to $-\frac{1}{S_3^2}$. As a result, negative feedback dominates the positive feedback at highest $S_3$ concentration (i.e. the peak value of PSMAD) and decreases the PSMAD level until it reaches a balanced and stable state (see Figure 4.8).

![Figure 4.8](image)

**Figure 4.8:** The effects of receptor concentration on the long-term response of PSMAD. A) TGF−β = 5. B) TGF−β = 2. Approximately, no difference is observed between the two panels of this figure.

We have repeated the above simulation with changing by SMAD concentration (see Figure 4.9). In neither panel of Figure 4.9 (corresponding to TGF−β = 5 and 2) is sensitivity to the TGF−β concentration evident at low SMAD levels. PSMAD concentration level increases with SMAD concentration level until the
saturation region. Moreover, decreasing the TGF−β value dramatically suppressed the signal of all SMAD concentrations. At the higher concentration of TGF−β, PSMAD levels increased even of low SMAD concentration (Figure 4.9A compared to Figure 4.9B). The distinct saturation levels of panel A and B of Figure 4.9 are relative to the distinct steady-state levels of Figure 4.5 and Figure 4.6. More specifically, PSMAD steady-state value in Figure 4.9A varies between 0.63 and 0.74, corresponding to the saturation level of PSMAD in short-term TGF−β stimulation. In Figure 4.9B instead, total PSMAD changes from 0.415 to 0.43, corresponding to PSMAD saturation level in long-term responses. Consequently, total PSMAD concentration at higher TGF−β stimulation varies with SMAD concentration levels as the system is stimulated with the ligand for a short time and vise versa.

Figure 4.9: The effects of SMAD concentration on the long-term response of PSMAD. A) TGF−β = 5 B) TGF−β = 2. The steady-state level of total PSMAD rises higher in Figure 4.9A than in Figure 4.9B.

Figure 4.10 shows that the results of the simplified model of TGF−β signalling are consistent with the experimental data from different cell lines (wild type MEFs and SV40-immortalized MEFs). Note that the blue line is not a fit of the experimental data (dots). We have compared the output of the model with the experimental data for different cell lines. Identical model parameters are used to produce Figure 4.10A, Figure 4.10B and Figure 4.7B (the blue curves).
4.3 Numerical Simulations

For the wild type MEFs (Figure 4.10C), however, we have used the same model parameters as in Figure 4.4.

![Graph showing PSMAD2 concentration over time for different conditions.]

**Figure 4.10:** The validation of the simplified model with experimental data. The dots show the level of PSMAD2 concentration obtained from experiment and the curves specify the model predictions. A) PSMAD2 time course for 0-1h on SV40-immortalized MEFs cell line stimulated with TGF−β and its corresponding blot B) PSMAD2 time course for 0-4h on SV40-immortalized MEFs cell line stimulated with TGF−β and its corresponding blot C) PSMAD2 time course for 0-4h on wild type MEFs stimulated with TGF−β and its corresponding blot

Our simplified TGF−β signalling model was tested experimentally against the data collected in our lab, by the applicant. Figure 4.11 shows the ability of the model to predict the trend of the different experimental sets. The difference between the experimental data and the simulation curves can be explained by the errors associated by the experiments. The error bars for the experimental data in Figure 4.11 suggest a broad range of values for each time point. However, the model simulation result (blue curve) provides a good approximation to the general trend of the experimental data sets. Similarly, in Figure 4.12 the sim-
A simplified model is fitted to the experimental data set from Gp130<sup>F/F</sup> MEFs [207]. In order to achieve this fit the parameters of the model had to be adjusted. The level of the SMAD7 concentration is higher in Gp130<sup>F/F</sup> MEFs due to their gene modification. As is shown in Figure 4.12, the steady-state level of PSMAD2 is lower than in Figure 4.11. Note that the error bars are smaller in Figure 4.12 for the longer time points.

Figure 4.11: PSMAD2 time course validation with experimental data sets from wild type MEFs. Different dot colors specify different experiments. The curve shows the model prediction of PSMAD2 dynamics.

Figure 4.12: PSMAD2 time course validation with experimental data sets from Gp130<sup>F/F</sup> MEFs. The SMAD7 level is higher in Gp130<sup>F/F</sup> MEFs and the PSMAD2 steady-state level is lower compared to wild type MEFs. Different dot colors specify different experiments. The curve shows the model prediction of PSMAD2 dynamics.
4.4 Discussion

The importance of TGF-β signalling in the progression of cancer heralded in a new era of cell biology research [9, 88, 350, 378]. Several models for TGF-β signalling have been proposed [72, 74, 308, 386, 449, 511, 513]. In each case the models were attempting to study the complex responses of the intracellular signalling reactions to different concentrations of TGF-β. In one of the most comprehensive mathematical models Zi et al. [511] predicted that ligand depletion contributed to the long-term response levels of PSMAD. Zi et al. suggested that at higher concentrations of TGF-β, it was not depleted in the medium; as a result there was a transfer from a transient to a switch-like response to TGF-β concentration. They also noted the possibility that negative feedback mechanisms might also contribute to the switch-like response [511].

Our TGF-β model expresses less complexity than Zi et al., but at the same time it contains the critical components that lead to robust responses to TGF-β stimulation. The robustness of our model originates from the fact that the cooperative output does not rely on the depletion of the ligand but is built into the network. It is known that time-delayed positive and negative coupled feedbacks can create robust stable signalling [120, 121, 460, 461]. In order to show the critical role of feedback loops in the signalling networks we introduced a reduced model whose steady-state and equilibria were studies in the absence and presence of a linear negative feedback. Steady-state studies with our reduced model predict that feedback considerations can lead to a major improvement in mathematical modelling of TGF-β signalling. One of the objectives of our study was to design quantitative mathematical model that not only is applicable to normal cells but also cancer cells. In cancer cells the number of TGF-β receptors decreases significantly [235, 264, 462], thus the signalling will be downregulated. The time-dependent ligand depletion model of Zi et al. [511] does not simulate the decrease in the receptor copy numbers.

We have shown that the PSMAD response of the cells is insensitive to TGF-β stimulation for low receptor concentration. This situation could be interpreted as TGF-β signal suppression in cancer cell lines. The relevant concentration of receptors via which the most signal is delivered to the cells was also predicted. Furthermore, our simulations show that the SMAD level reduction also causes a global suppression of signalling in response to TGF-β. Due to mutations of SMADs, many cancer cells have reduced levels of TGF-β signalling [124]. These results are consistent with the picture of early-stage tumors being associated with the loss of TGF-β sensitivity and the decrease of TGF-β receptor expression, a prediction that may have recently been validated in endothelial cells and several cancer cell lines [97].

However, TGF-β transduction is stimulated in late-stage tumors (termed "The TGF-β Paradox," for reviews refer to [3, 165]). Our self-regulatory model can produce simulation results consistent with both roles of TGF-β in tumori-
genesis. We simulated loss of responsiveness to TGF−β seen in early tumors via reducing the production rates of the receptors and SMADs and the reacquiring of late tumors responsiveness to TGF−β via increasing their production rates. According to our model predictions, the overshoot peak of PSMAD in response to TGF−β is higher in early tumors and the steady-state levels of PSMAD are lower generally, while in late tumors both steady-state and peak levels are higher than normal cells. Additionally, the difference between the PSMAD peak and steady-state levels becomes less in late-stage tumors to keep the signalling on for longer time. This work can be used as a guide for future experimental research on TGF−β effects on tumor progression.

In addition to reproducing the data in the literature ([511]), new experiments were designed. To validate our model for the abnormal receptor concentration level, we have conducted experiments on an immortal cell line (SV40-immortalized MEFs). Figure 4.10, Figure 4.11 and Figure 4.12 demonstrate how the experimental data on immortal cells are explained via low receptor concentration level (Figure 4.10A and Figure 4.10B) and how the signalling changes for genetically modified cell lines (Figure 4.12). Concluding Figure 4.10, Figure 4.11 and Figure 4.12, they confirm that the model is compatible with the normal and/or the abnormal conditions inside a cell. Such robustness originates from the rigorous system analysis applied on the model (see the "The Simplified Model" and "The Importance of Feedback" in "Supplementary Material").

This model provides the basis for a mathematical study of TGF−β signal transduction and its influence on cancer cells. By considering of a model where coupled, positive-negative feedback loops modulate signalling, TGF−β signal transduction can be studied more precisely using control theory analysis including system identification methods [59, 68].

### 4.5 Materials and Methods

The experimental data set and the kinetic rates used to set initial parameters for the model were elicited from the literature [72, 74, 308, 386, 449, 511, 513]. More precisely, for the initial conditions, parameter estimated values and the interpretation of some experimental we have benefitted from Zi et al. proposed model [511]. The values for all model parameters are documented in Table 4.3 of the Supplementary Material.

**Computer modelling and simulations** The programs used for these simulation where PYTHON 2.7 and MATLAB 7.10. The curve fitting tool box of MATLAB is used for fitting the Hill equation to Figure 4.5 and Figure 4.15 in Supplementary Material, and deriving the Hill coefficients.

**Mathematical and biochemical analysis** The biochemical kinetics, equilibrium analysis, feedback analysis, reduction analysis using rapid equilibrium assumption, time-delayed analysis, asymptotic expansions and sensitivity analysis have
been performed on the model [115].

**Cell culture and cell lysis** Mouse embryonic fibroblasts (MEFs) cells were isolated from day 13 to 15 embryos. Wild type MEFs, SV40-immortalized MEFs (Simian vacuolating virus 40) and Gp130\(^{F/F}\) MEFs cells were cultured in DMEM containing 15% FCS. The cells were typsinazed and washed with DMED + 10% FCS before plating. Passage 3 cells with \(1 \times 10^6\) MEFs/well were seeded in 60 mm plates for 0-4 hours, \(0.5 \times 10^6\) MEFs/well for 24 hour and \(0.25 \times 10^6\) MEFs/well for 48 hour treatment with 5 ng/ml TGF-\(\beta\) respectively (for Gp130\(^{F/F}\) MEFs the cell numbers changes as: \(1.2 \times 10^6\) MEFs/well for 0-4 hours, \(0.6 \times 10^6\) MEFs/well for 24 hours and \(0.3 \times 10^6\) MEFs/well for 48 hours). After washing with cold PBS for two times, cells were lysed in ice-cold 200 ul RIPA lysis buffer, containing 1M Tris/HCL, 0.5 M EDTA, 5M NaCl, 10 % Na Doc, 10 % TX-100, 10 % SDS, proteinase inhibitor 100 × and H_2O. The cell lysits were passed through 27 G needle for 5 times, then incubated in ice for 20 min. After incubation the samples were span at 13,000 rpm for 30 min at 4\(^\circ\)C. The supernatant was transferred to new tubes where 20 ul of samples were saved for BCA protein assay using sigma BCA assay kit (B9643). 20 ul 5× sample buffer was added to 80 ul loading sample and the samples were heated at 95\(^\circ\)C for 10 min.

**Western blotting** According to the protein concentration results from the protein assay, Novex NuPAGE® 4-12%-Bis-Tris (life technologies NP0335 Box) gels were used for loading proper amount of sample lysate for each time point. PSTAT3 (XP\(^{TM}\) Rabbit mAb) antibodies were provided by Cell Signalling Technology and were used at a dilution 1:500 in 3% BSA-TBS-T. Antibody directed against SMAD7 was provided via Santa Cruz Biotechnology and was diluted 1:1000. PSMAD2 antibody (rabbit polyclonal anti-phospho-Smad2 antibody (1:1000 for Western blot)) was a gift from Prof. Peter ten Dijke (Leiden University Medical Center, Netherlands). \(\beta\)-tubulin, actin, Lamin b1 or transferrin receptor are used as loading control depending on the protein to be expressed. Eventually, the gels were transformed onto nitrocellulose membrane via iBlot 2 gel transfer device (life technologies) and the membranes were scanned using Odyssey infrared scanner (LI-COR).

**Protein Quantitation** The western blot images were quantitated using ImageJ 1.49p. The signals of each protein were normalised to the loading control protein.
4.6 Supplementary Material

4.6.1 The Reduced Model

The full model described by equations 4.1 consists of two subsystems, separated according to the participating components. The membrane subsystem is represented by the first 7 equations and nucleocytoplasmic subsystem is described by the remaining 4 differential equations. At this stage we are focusing on the receptor subsystem only. The equations of the TGF−β system in this section (equations 4.5 and 4.6) are different from our simplified model described in Chapter 4 (equations 4.3). More specifically, we have reduced the TGF−β system differently in this section in order to describe the behaviour of the TGF−β signalling via only two differential equations and study the phase portrait of the system. In reducing the TGF−β system, the rapid equilibrium assumption was used.

As mentioned earlier in a dimerization reaction, it is almost impossible to measure $k^+$ (the forward reaction rate). Although $k^-$ (the backward reaction rate) is measurable, the easiest parameter to be quantified is $K (K = \frac{k^-}{k^+})$, i.e. the equilibrium constant. The dimerization process of R1 or R2 is much faster than the other receptor reactions. Furthermore, as discussed earlier the symbolic feedback loop was replaced with a number of relatively slow reactions. In comparison to every other reaction in the receptor system, the time scales of R1 and R2 dimerization and RC formation are fast, thus justifying the "rapid equilibrium assumption" [115,203,459].

Based on the above assumptions, we apply the conservation law to R1 and R2, so the final form of the reduced version of the R1, R2 and PC are:

$$R_{1T} = R_1 + 2R_{12} + 2RC$$
$$R_{2T} = R_2 + 2R_{22} + 2RC$$
$$PC_T = LC + PC$$

The conservation law for R1 and R2 states that the total amount of R1 and R2, by definition, equals to the total amount of R1 and R2 as monomers adding by twice of the amount of R1 and R2 as dimers (which is twice of R1 and R2). The same definition applies for PC_T.

According to the experimentalists, we know that the dimerization reactions, the binding reactions of the two dimers and the phosphorylation of the two R1s inside LC (PC association) are relatively fast comparing to the ligand binding reaction [115]. Inside the cytosol the trimerization of the phosphorylated SMAD occurs more rapidly than the SMAD phosphorylation processes. Pursuant to the rapid equilibrium assumption, at equilibrium the forward and the backward terms of the fast reactions can be considered to be equal. Therefore, the equations describing the reduced model are:
\[ \beta_{11} [R1]' + \beta_{12} [R2]' = v_1 - k_1 [R1] - 2k_{RC} [RC] - k_{f1}^- [N]^2 \frac{[R1]}{R1 + K} - 2k_{f1}^- [N]^2 \frac{[RC]}{[RC] + K} - k_{f1}^+ [P] \frac{[R1]}{[R1] + K} \]

\[ \beta_{21} [R1]' + \beta_{22} [R2]' = v_2 - k_2 [R2] - 2k_{RC} [RC] - 2k_{f2}^- [N]^2 \frac{[RC]}{[RC] + K} + \frac{k_{f2}^+ [N]^2}{[PC] + K} \]

\[ (1 + K_{PC}) [PC]' = k_{f2}^- [TGF_2] [RC] - k_{f2}^- K_{PC} [PC] - k_{PC}^- [PC] - k_{PC}^- [PC] \]

\[ [S]' = v_S - k_S [S] - k_S^+ [PC] \frac{[S]}{[S] + K_S} + k_S^- [\hat{S}] \]

\[ [\hat{S}]' = k_S^+ [PC] \frac{[S]}{[S] + K_S} - k_S^- [\hat{S}] - 3k_S^+ [\hat{S}]^3 + 3k_S^- [S_3] \]

The Importance of Feedback

This section aims to demonstrate the effects of a negative feedback loop on the steady-state of the TGF–β signalling system. The corresponding equations can be shown as the following.

\[ \frac{d[R1]}{dt} = v_1 - k_1 [R1] - 2k_{RC} [RC] - k_{f1}^- [S_3]^2 [R1] - 2k_{f2}^- [S_3]^2 [RC] \]

\[ \frac{d[R2]}{dt} = v_2 - k_2 [R2] - 2k_{RC} [RC] - 2k_{f2}^- [S_3]^2 [RC] \]

\[ \frac{d[PC]}{dt} = k_{f2}^- [TGF_2] [RC] - k_{f2}^- K_{PC} [PC] - k_{PC}^- [PC] \]

\[ \frac{d[S]}{dt} = v_S - k_S [S] - k_S^+ [PC] \frac{[S]}{[S] + K} + k_S^- [\hat{S}] \]

\[ \frac{d[\hat{S}]}{dt} = k_S^+ [PC] \frac{[S]}{[S] + K} - k_S^- [\hat{S}] \]
As is shown in equations 4.6, we have assumed that the dimerization of $R_1$, dimerization of $R_2$, association of $RC$ and the association of $PC$ are fast. Consequently, $R_{12}$, $R_{22}$, $RC$ and $LC$ are evaluated based on $R_1$, $R_2$ and $PC$. Furthermore, there is no $S_n$ defined in this system and it is assumed that $\hat{S}$ represents both nuclear and cytoplasmic phospho-SMAD.

These equations are also different from the reduced model (equations 4.5). Specifically, we have assumed that $K_{11}$, $K_{22}$ and $K_{RC}$ are large enough such that $\beta_{11} = \beta_{22} = 1$ and $\beta_{12} = \beta_{21} = 0$. Additionally, the production of $PC$ from $LC$ and trimerization of $\hat{S}$ occur relatively fast. Therefore, $[LC] = K_{PC} [PC]$ and $[S_3] = \frac{[\hat{S}]}{K_3}$ at equilibrium. In our toy model (equations 4.8), we assumed that the negative feedback loops are applied only to $R_1$ and $RC$ and follow linear kinetics (instead of Michaelis-Menten kinetics). We did not include a time-delay in the system. With these assumptions we can write the following set of equations to represent the system at equilibrium.

\[
[S] = \frac{v_S}{k_S} \\
[\hat{S}] = \frac{k_{1[PC]}^+ [LC]}{k_S} \frac{[S]}{[S] + K_S} \\
[PC] = \frac{k_{1LC}^+ [TGF_2]}{k_{PC}^- + \frac{k_{PC}^-}{K_{PC}}} \frac{[LC]}{k_{PC}^- ([S] + K)} \\
[S_3] = \frac{[\hat{S}]^3}{K_3} = \frac{1}{K_3} \left( \frac{[LC]}{K_{PC}^+ k_{PC}^- ([S] + K)} \right)^3 \\
[RC] = \frac{[R_1]^2 [R_2]^2}{K_1 K_2 K_{RC}}
\]

We have non-dimensionalized the model (equations 4.6) in order to simplify our equations without elimination of any information from the system. According to [273], scaling does not change the solution of the original system. Non-dimensionalization is often used to simplify and reduce the number of effective parameters (without changing the number of variables and equations), to analyse the behaviour of a system without considering the units of its components and to re-scale the variables whose amounts have vastly different magnitudes [76,273]. The simplifications preserve the essential features of our model. After applying the non-dimensionalization method, at equilibrium, the system in equation 4.6 can be reduced to (we have named it the toy model):

\[
0 = 1 - r_1 - c_1 r_1^2 r_2^2 - \delta r_1^{13} r_2^{12} - \epsilon_1 r_1^{14} r_2^{14} \\
0 = 1 - r_2 - c_2 r_1^2 r_2^2 - \epsilon_2 r_1^{14} r_2^{14}
\]

, with the dimension-less variables and components defined as:
Note that $k_f^{-1} = k_a$ and $k_{RC}^{-1} = k_b$. Equation 4.8 was derived after applying the reduction method explained in Chapter 3.

The system (equations 4.6) was reduced to two effective equations as functions of the concentrations of $R_1$ and $R_2$. In this section, equations 4.8 are solved and analyzed in the steady-state (equilibrium). First, we study the system in the absence of feedback ($\delta = \varepsilon_1 = \varepsilon_2 = 0$) - the simplest, symmetric system. Then, we investigate the effects of the negative, regulatory feedback loop on the equilibria of the system. This negative feedback is only applied on $R_1$ and $R_{RC}$. In fact, the importance of the positive feedback and the resulted ultra-sensitivity feature of the system are included in the "The Simplified Model". The most critical advantage of the toy model (equations 4.8) is to analytically (since we reduced all the equations to two) study the effects of a feedback loop (here negative feedback loop) on the steady-state and equilibria of a system. Based on feedback study of this system, we can speculate changes in the system’s operating point after adding complementary positive and/or negative feedbacks.

As mentioned in the introduction, previous models on TGF-β signalling lacked the consideration of the feedback loops. In other words, the steady-state analysis of TGF-β transduction in the absence of feedback corresponds to the previous mathematical models of TGF-β in the literature. In this section, we demonstrate the importance of a simple, linear negative feedback on the steady-state of the TGF-β signalling system.

The equations describing the reduced model in the absence of feedback (the no-feedback toy model) are:

$$\frac{dr_1}{dt} = 1 - r_1 - c_1 r_1^2 r_2^2$$
$$\frac{dr_2}{dt} = 1 - r_2 - c_2 r_1^2 r_2^2$$

This equation can be written as:
Here, $r_1$ and $r_2$ are the scaled values of $R_1$ and $R_2$ respectively. We have assumed $\beta_{12}$ and $\beta_{21}$ are zero, e.g. $K_{RC}$ is very large (see Table 4.3). The clearest way to appreciate the dynamics of the systems at equilibrium is to consider the phase plane. In phase plane analysis of the differential equations we normally consider the nullclines. The nullclines of $f_1$, respectively $f_2$, are defined by setting $f_1(r_1, r_2) = 0$, respectively $f_2(r_1, r_2) = 0$. The steady-state of the system is defined by a set of $(r_1, r_2)$ such that $f_1(r_1, r_2) = 0$ and $f_2(r_1, r_2) = 0$. The intersection of each pair of nullclines represents a fixed point of the system at steady-state. The pattern achieved via connecting fixed points for a set of parameters shows the equilibrium curve that describes the system at steady-state. In the absence of feedback ($\delta = \epsilon_1 = \epsilon_2 = 0$) the nullclines are given by,

$$
\begin{align*}
    r_1 &= -1 \pm \sqrt{1 + 4c_1r_2^2} \\
    r_2 &= -1 \pm \sqrt{1 + 4c_2r_1^2}
\end{align*}
$$

and the steady-state is the solution of,

$$
\begin{align*}
    1 - r_1 - c_1r_1^2r_2^2 &= 0 \\
    1 - r_2 - c_2r_1^2r_2^2 &= 0
\end{align*}
$$

which can be solved for:

$$
\frac{1 - r_2}{1 - r_1} = \frac{c_2}{c_1} \equiv c.
$$

$c$ is the slope of the curve in the $(r_1, r_2)$ phase plane. It is important to note that equation 4.16 does not define the steady-state of the system completely, though it is sufficient for plotting the phase plane (phase portrait) to visualise the behaviour of the system.

The steady-state analysis of the reduced model in the absence of feedback is summarized in Figure 4.13. Figure 4.13A shows the steady-state curves for five different values of $c$. It shows how the steady-state curve varies with $c$. Note that for all values of $c$ the steady-state curves are straight lines which pass through 1,1. Scaled total $S_3$ (here designated as $s_3$) is plotted against the logarithm of $c$ to base 2 for different values of $c$ when $\delta$ and $\epsilon$ values are zero (see Figure 4.13B). Here, $s_3$ is a representative of the output. According to the reduced model design in "The Reduced Model", $c$ is a ratio of $v_1$ and $v_2$, the production rates of
R1 and R2 respectively. Therefore, Figure 4.13B specifies how the abundance of the receptors on the membrane can upregulate/downregulate the TGF−β signalling in the absence of feedback. Experimentally, \( v_1 \) and \( v_2 \) are estimated to have similar values. So we expect the working point of the system to be in the middle part of the Figure 4.13B \((c = 1, \log c = 0)\), at which the output dramatically changes. This explains the importance of the receptor concentrations in TGF−β signalling.

Figure 4.13: Steady-state analysis of the reduced model in the absence of feedback. A) The phase plane plots of the reduced receptor system at equilibrium for different values of \( c \). B) The scaled total PSMAD trimer, \( s_3 \), value against different values of \( c \), where \( \log c = 0 \) corresponds to \( c = 1 \).

To understand the effects of feedback in this system, we then consider the toy model (equations 4.8) with its simplest negative feedback loops. For the sake of presentation, \( r_1 \) and \( r_2 \) are changed to \( x \) and \( y \), respectively. As mentioned earlier these negative feedback terms correspond to those in equations 4.5, but with the Michaelis-Menten terms linearized. The 4.8 equations can be presented by one 14th order polynomial which characterizes the full receptor system at steady-state.

\[
\varepsilon x^{14} y^{14} + \delta x^{13} y^{12} - y + cx - c + 1 = 0 \quad \varepsilon = \frac{c_2 \varepsilon_1 - c_1 \varepsilon_2}{c_1}, \quad c = \frac{c_2}{c_1} \quad (4.17)
\]

There is no straight forward procedure for solving an 14th-order equation. However, we can approximate the solution via asymptotic expansion of \( y \) as a function of \( x \). Perturbation expansion is used, assuming that the two parameters of \( \delta \) and \( \varepsilon \) are small. We have ignored the second order term for \( \delta \) and/or \( \varepsilon \) (first order approximation).

\[
y(x) = cx - c + 1 + \varepsilon x^{14}(cx - c + 1)^{14} - \delta x^{13}(cx - c + 1)^{12} + 14\varepsilon^2 x^{28}(cx - c + 1)^{27} - 26c\varepsilon\delta x^{27}(cx - c + 1)^{25} + 12\varepsilon^2 \delta^2 x^{26}(cx - c + 1)^{23} + O(n^3), \quad n \in \{\varepsilon, \delta\} \quad (4.18)
\]
The final result is illustrated in Figure 4.14 which shows how the solution from the asymptotic expansion changes with different set of values for $\delta$ and $\varepsilon$. This figure shows the deviation of the steady-state curves when negative feedback is input to the system compared to no-feedback. Feedback considerations altered the straight lines to curves which no longer pass through the $(1,1)$ point of the phase plane (see Figure 4.14A). In Figure 4.14A, the solid curves demonstrate the least and highest value of the $\delta$, while the value of $\varepsilon$ remains zero. The grey curves simply cover the range between these two limits. By comparing Figure 4.14A with Figure 4.13A (where $\delta = 0$ and $\varepsilon = 0$), it could be seen how the addition of a weak negative feedback, only on $R_1$, change the steady-state of the reduced model. A complete sensitivity analysis of the receptor system is displayed in Figure 4.14B. It is apparent that for non-zero values of $\varepsilon$ the system is less sensitive to the changes in $\delta$. Significantly, for the $c_1/c_2$ ratio of the second order of magnitude, the system is very sensitive to small changes in the feedback strength ($\varepsilon$ and $\delta$). A comparison between Figure 4.13B and Figure 4.14B reveals how the steady-state of a system can be affected via insertion of a linear negative feedback.

**Figure 4.14:** The steady-state of the receptor system for different range of parameters. A) $\varepsilon = 0.5 \delta = 0(0.2)1$ in these panels $A,B$ and $C \ c = 2^n$ for $n = -2(1)2$. B) $\delta = 0(0.2)1$, In this panel $\varepsilon$ changes as labeled.

It is important to note that the steady-state diagrams of Figure 4.14 are acquired using the asymptotic expansion estimation of the exact solution (4.18). As a result, the curves may behave differently for large values of $\varepsilon$ and $\delta$. Also note that $r_1$, $r_2$ and $s_3$ in this section represent scaled receptor variables and the level of total PSMAD trimer in the reduced model. We can follow all the scaling steps reversely to derive receptor concentrations in terms of the production/degradation rates and reaction binding constants. However, the functionality of one variable with respect to the other remains the same. We do not expect significant divergence in the actual $R_1$-$R_2$ plot (and $S_3$ plot versus $v_1$ and $v_2$ ratio) from the results shown in Figure 4.14. The changes due to negative feed-
back considerations are close to expectation. The negative feedback on R1 and RC does not extremely change the steady-state of $r_1$ and $r_2$ (note the operating point of $\log c \approx 0$ in Figure 4.14B). Such an insensitivity might be due to the lack of positive feedback and/or negative feedback on LC (refer to the numerical simulation, section 4).

### 4.6.2 Model Parameters

<table>
<thead>
<tr>
<th>Components as variables</th>
<th>Components symbols in simulation and codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGF−β receptor type 1</td>
<td>R1</td>
</tr>
<tr>
<td>TGF−β receptor type 2</td>
<td>R2</td>
</tr>
<tr>
<td>TGF−β receptor type 1 dimer</td>
<td>R1₂</td>
</tr>
<tr>
<td>TGF−β receptor type 2 dimer</td>
<td>R2₂</td>
</tr>
<tr>
<td>Receptor complex</td>
<td>RC</td>
</tr>
<tr>
<td>Ligand-receptor complex</td>
<td>LC</td>
</tr>
<tr>
<td>Completely phosphorylated ligand-receptor complex</td>
<td>PC</td>
</tr>
<tr>
<td>Cytoplasmic SMAD</td>
<td>S</td>
</tr>
<tr>
<td>Phosphorylated SMAD in the cytoplasm</td>
<td>S</td>
</tr>
<tr>
<td>Phosphorylated SMAD in the nucleus</td>
<td>$S_n$</td>
</tr>
<tr>
<td>Phosphorylated SMAD trimer</td>
<td>$S_3$</td>
</tr>
<tr>
<td>Delayed Phosphorylated SMAD trimer</td>
<td>$N$</td>
</tr>
<tr>
<td>Positive feedback intermediate inhibitor</td>
<td>$P$</td>
</tr>
<tr>
<td>Extracellular TGF−β ligand</td>
<td>TGF−β Dimer</td>
</tr>
</tbody>
</table>

Table 4.1: Components of the TGF−β receptor signalling model

<table>
<thead>
<tr>
<th>Kinetic rates</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_1$</td>
<td>Production rate of TGF−β receptor type 1</td>
<td>nMmin$^{-1}$</td>
</tr>
<tr>
<td>$k_1$</td>
<td>Degradation rate of TGF−β receptor type 1</td>
<td>min$^{-1}$</td>
</tr>
<tr>
<td>$v_2$</td>
<td>Production rate of TGF−β receptor type 2</td>
<td>nMmin$^{-1}$</td>
</tr>
<tr>
<td>$k_2$</td>
<td>Degradation rate of TGF−β receptor type 2</td>
<td>min$^{-1}$</td>
</tr>
<tr>
<td>$k_1^+$</td>
<td>Association rate of type 1 receptor homo-dimer complex</td>
<td>nM$^{-1}$min$^{-1}$</td>
</tr>
</tbody>
</table>
### SMAD Proteins

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_S$</td>
<td>Production rate of cytoplasmic SMAD</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_S$</td>
<td>Degradation rate of SMAD in the cytoplasm</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_S^+$</td>
<td>Phosphorylation rate of SMAD in the cytoplasm</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_S^-$</td>
<td>Dephosphorylation rate of SMAD in the cytoplasm</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{Sn}$</td>
<td>Degradation rate of PSMAD in the nucleus</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_n^+$</td>
<td>Import rate of PSMAD into the nucleus</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_n^-$</td>
<td>Export rate of PSMAD from the nucleus</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_3^+$</td>
<td>Association rate of PSMAD homo-trimer complex</td>
<td>$nM^{-2}min^{-1}$</td>
</tr>
<tr>
<td>$k_3^-$</td>
<td>Dissociation rate of PSMAD homo-trimer complex</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$K_{RC}$</td>
<td>The receptor complex binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_S$</td>
<td>The phosphorylation binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_3$</td>
<td>PSMAD trimer binding constant</td>
<td>$nM^2$</td>
</tr>
</tbody>
</table>

### Feedback

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{f+}$</td>
<td>Positive feedback on TGF–β receptor type 1</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{f-}$</td>
<td>Negative feedback on TGF–β receptor type 1</td>
<td>$nM^{-1}min^{-1}$</td>
</tr>
<tr>
<td>$k_{RC}$</td>
<td>Negative feedback on receptor tetramer complex</td>
<td>$nM^{-1}min^{-1}$</td>
</tr>
</tbody>
</table>
4.6 Supplementary Material

$k_{PC}^{-}$ | Negative feedback on phosphorylated ligand-receptor complex | $nM^{-1}min^{-1}$
---|---|---
$K$ | Negative feedback inhibitory binding constant | $nM$
$K_I$ | Inhibition of the intermediate inhibitor binding constant | $nM$

Table 4.2: Kinetic rates and binding constants of the model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbol</th>
<th>Literature value</th>
<th>Reference</th>
<th>Scaled data for our model</th>
</tr>
</thead>
<tbody>
<tr>
<td>R production rate</td>
<td>$v_1$</td>
<td>$0.0137 \ nM min^{-1}$</td>
<td>[462]</td>
<td>1 $nM min^{-1}$</td>
</tr>
<tr>
<td>R degradation rate</td>
<td>$k_1$</td>
<td>$0.00256 \ min^{-1}$</td>
<td>[225, 513]</td>
<td>0.2 $min^{-1}$</td>
</tr>
<tr>
<td>R positive stimulated degradation rate</td>
<td>$k_1^{f+}$</td>
<td>–</td>
<td>–</td>
<td>1 $min^{-1}$</td>
</tr>
<tr>
<td>R negative stimulated degradation rate</td>
<td>$k_1^{f-}$</td>
<td>–</td>
<td>–</td>
<td>0.4 $nM^{-1} min^{-1}$</td>
</tr>
<tr>
<td>RC association rate</td>
<td>$k_{RC}^+$</td>
<td>–</td>
<td>–</td>
<td>0.5 $nM^{-1} min^{-1}$</td>
</tr>
<tr>
<td>RC dissociation rate</td>
<td>$k_{RC}^-$</td>
<td>–</td>
<td>–</td>
<td>0.5 $min^{-1}$</td>
</tr>
<tr>
<td>RC degradation rate</td>
<td>$k_{RC}$</td>
<td>–</td>
<td>–</td>
<td>1 $min^{-1}$</td>
</tr>
<tr>
<td>RC negative stimulated degradation rate</td>
<td>$k_{RC}^{f-}$</td>
<td>$0.00256 \ nM^{-1} min^{-1}$</td>
<td>[93]</td>
<td>0.4 $nM^{-1} min^{-1}$</td>
</tr>
<tr>
<td>PC association rate</td>
<td>$k_{PC}^+$</td>
<td>–</td>
<td>–</td>
<td>1.6 $nM^{-1} min^{-1}$</td>
</tr>
<tr>
<td>PC dissociation rate</td>
<td>$k_{PC}^-$</td>
<td>–</td>
<td>–</td>
<td>0.2 $min^{-1}$</td>
</tr>
<tr>
<td>PC degradation rate</td>
<td>$k_{PC}$</td>
<td>–</td>
<td>–</td>
<td>0.4 $min^{-1}$</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{PC}^-$ (PC negative stimulated degradation rate)</td>
<td>0.4 nM$^{-1}$min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v_S$ (S production rate)</td>
<td>0.01 nMmin$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_S$ (S degradation rate)</td>
<td>0.008 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_S^+$ (S association rate)</td>
<td>0.049 nM$^{-1}$min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_S^-$ (S dissociation rate)</td>
<td>0.1 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_S$ (S degradation rate)</td>
<td>0.394 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_n^+$ (S$_n$ association rate)</td>
<td>0.156 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_n^-$ (S$_n$ dissociation rate)</td>
<td>0.739 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_S$ (S$_n$ degradation rate)</td>
<td>0.01 min$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K$ (positive/negative feedback constant)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_I$ (Inhibition of the intermediate inhibitor binding constant)</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_S$ (SMAD phosphorylation binding constant)</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R initial value</td>
<td>0.83 nM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC initial value</td>
<td>0.23 nM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S initial value</td>
<td>60.6 nM [511]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGF−β initial value</td>
<td>to be specified [511]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: The parameter values of simplified TGF−β model
4.6.3 Figures

By comparing Figure 4.15 and Figure 4.16 with Figure 4.5 and Figure 4.6 in the numerical simulation section, we confirmed that the effects of the system’s negative feedback is almost entirely on LC.

Figure 4.15: The predicted effects of different concentrations of TGF−β on the short-term responses of total PSMAD when the negative feedback influences LC only.

Figure 4.16: The predicted effects of different concentrations of TGF−β on the long-term responses of total PSMAD when the negative feedback influences LC only.

4.6.4 Steady-state Analysis of the Simplified Model

There are two possible methods to solve a system of differential equations at steady-state (or equilibrium in this case), analytical and numerical analysis. Analytical analysis defines the equilibria (whether it is single or multiple equilibrium points) explicitly. I attempted to develop analytical solutions for the equilibrium and/or steady-state of the simplified TGF−β model.

The simplified model of TGF−β consists of 6 differential equations which can be categorized into 2 subsystems (see equations 4.3). The first 3 equations...
belong to the membrane subsystem and the last 3 equations belong to the nuclear–cytoplasmic subsystem. The coupling between the two subsystems is defined such that nucleocytoplasmic subsystem feeds back on the membrane subsystem. In order to solve the steady-state of the system analytically, we equate the RHSs of the equations 4.19 to zero. This provides us with 6 algebraic equations whose solution is defines the equilibrium of the whole system. In the following calculations, the variables of the system are substituted with \( x_1 \) to \( x_6 \) respectively, for the sake of presentation.

\[
\begin{align*}
\frac{dx_1}{dt} &= v_1 - k_1 x_1 - 2k_{RC}^+ x_1^2 + 2k_{RC}^- x_2 - k_1^f \left[ P \right] \frac{x_1}{x_1 + K} - k_1^f \left[ N \right]^2 \frac{x_1}{x_1 + K} \\
\frac{dx_2}{dt} &= k_{RC}^+ x_1^2 - k_{RC}^- x_2 - k_{RC} x_2 - k_{PC}^+ \left[ TGF_2 \right] x_2 + k_{PC}^- x_3 - k_{RC}^f \left[ N \right]^2 \frac{x_2}{x_2 + K} \\
\frac{dx_3}{dt} &= k_{PC}^+ \left[ TGF_2 \right] x_3 - k_{PC}^- x_3 - k_{PC} x_3 - k_{PC}^f \left[ N \right]^2 \frac{x_3}{x_3 + K} \\
\frac{dx_4}{dt} &= v_S - k_S x_4 - k_S^+ x_3 - \frac{x_4}{x_4 + K_S} + k_S^- x_5 \\
\frac{dx_5}{dt} &= k_S^+ x_3 - \frac{x_4}{x_4 + K_S} - k_S^- x_5 - k_n^+ x_5 + k_n^- x_6 - k_S^- x_5 \\
\frac{dx_6}{dt} &= k_n^+ x_5 - k_n^- x_6 - k_S x_6
\end{align*}
\]

(4.19)

where again, \([S_3] = x_6^3/K_3\), \(x_6 = [S_3] (t - \tau)\) and \([P] = K_1^2/(K_1^2 + [S_3]^2)\).

In nucleocytoplasmic subsystem, \(x_6\) can be written in terms of \(x_5\) soling the last equation of 4.19 for its roots. Second last equation defines \(x_5\) in terms of \(x_3\) and \(x_4\). Considering the fact the two thermodynamic terms are in common between fourth and fifth equations we can also write \(x_4\) in terms of \(x_5\) and \(x_6\). All together, the nucleocytoplasmic subsystem can be represented by a single nonlinear algebraic equation:

\[
v_S - k_S x_4 = (k_n^+ + k_S^- - \frac{k_n^- k_n^+}{k_n^- + k_S^-})k_S^+ x_3 \frac{x_4}{(x_4 + K_S)(k_S^- + k_n^+ + k_n^- k_n^- k_n^- + k_S^-)}
\]

where

\[
\begin{align*}
x_6 &= \frac{k_n^+}{k_n^- + k_S^-} x_5 \\
x_5 &= \frac{k_S^+ x_3}{(x_4 + K_S)(k_S^- + k_n^+ + k_n^- k_n^- k_n^- + k_S^-)} x_4
\end{align*}
\]

(4.20)

However, reducing the algebraic equations of the membrane subsystem is more complex than the nucleocytoplasmic subsystem. Such complexity is mainly
Due to their high nonlinearity. The third equation of 4.19 suggests that \( x_2 \) can be written in terms of \( x_3 \) and \( x_6 \). Therefore, \( x_2 \) can be eliminated in the first and second equilibrium equations, giving 2 equations based on \( x_1 \), \( x_3 \) and \( x_6 \):

\[
0 = v_1 - k_1 x_1 - 2k_{RC}^+ x_1^2 + 2k_{RC}^- \frac{k_{PC}^+ + k_{PC}^-}{k_{PC}^+ [TGF_2]} x_3 + k_{PC}^- \frac{x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]}
\]

\[
-k_1^+ \frac{K_1^2 x_1}{(K_1^2 + x_6^i / K_3^2)(x_1 + K)} - k_1^- \frac{x_6^i x_1}{K_3^2 x_1 + K}
\] (4.21)

\[
0 = k_{RC}^+ x_1^2 - (k_{RC}^- + k_{RC}^+ + k_{PC} [TGF_2])
\]

\[
\left( \frac{k_{PC}^- + k_{PC}^+}{k_{PC}^+ [TGF_2]} x_3 + k_{PC}^- \frac{x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]} \right)
\]

\[
+ k_{PC}^- x_3 - k_{RC}^- \frac{x_6^i x_3}{K_3^2 \left( \frac{k_{PC}^- + k_{PC}^+}{k_{PC}^+ [TGF_2]} x_3 + k_{PC}^- \frac{x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]} \right)} + K
\],

where

\[
x_2 = \frac{k_{PC}^- + k_{PC}^+}{k_{PC}^+ [TGF_2]} x_3 + \frac{k_{PC}^- x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]}
\] (4.23)

On the other hand, equating the identical thermodynamic terms in the membrane subsystem provides another algebraic equation in terms of \( x_1 \), \( x_3 \) and \( x_6 \):

\[
2 \left( \frac{k_{RC} + k_{PC}^+ [TGF_2]}{k_{PC}^+ [TGF_2]} \right) \left( \frac{k_{PC}^- + k_{PC}^+}{k_{PC}^+ [TGF_2]} x_3 + k_{PC}^- \frac{x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]} \right)
\]

\[
- 2 \frac{k_{PC}^- x_3 + 2k_{PC}^-}{K_3^2 \left( \frac{k_{PC}^- + k_{PC}^+}{k_{PC}^+ [TGF_2]} x_3 + k_{PC}^- \frac{x_6^i x_3}{K_3^2(x_3 + K)k_{PC}^+ [TGF_2]} \right)} = (4.24)
\]

\[
v_1 - k_1 x_1 - k_{RC}^- \frac{K_1^2 x_1}{(K_1^2 + x_6^i / K_3^2)(x_1 + K)} - k_1^- \frac{x_6^i x_1}{K_3^2 x_1 + K}
\]

Equations 4.21, 4.22 and 4.24 should be solved simultaneously for \( x_1 \), \( x_3 \) and \( x_6 \) so that the equilibrium point or the equilibria of the system would be identified rigorously. However, the complexity of these algebraic equations makes it hard to introduce an analytical solution for the steady-state of the simplified system. Therefore in order to study the behaviour of the system with different inputs, numerical methods were used (instead of analytical methods). These numerical methods are constrained with the biological data. Several limitations are applied to the model parameters and the initial values of the variables. For instance, none of the components of the model can be negative since they are concentrations, kinetic rates or binding constants, or most of the cytoplasmic and nuclear variables are zero at the beginning of the stimulation. The equilibria cannot be located in the whole \( \mathbb{R}^6 \) space.
Chapter 5

Modelling the $\text{IL} - 6$ Signal Transduction System

5.0 Summary of the information included in Chapter 5

Interleukin-6 is a pleiotropic cytokine that can act as both an anti- and a pro-inflammatory regulator. Signalling via IL-6 plays a pivotal role in many the tissue regulation mechanisms including cell proliferation, apoptosis and immune responses during inflammation. Importantly, excess IL-6 signalling is associated with a diverse range of tumors. A robust mathematical model is required for predicting the likely cellular responses to IL-6 signalling. We have analysed the IL-6 signalling pathway, identified the critical components and developed the minimal systems model capable of quantitative simulation of the network in response to perturbations. Our reduced IL-6 signalling model is compatible with experimental data on changes in STAT3, phospho-STAT3 and SOCS3 levels at different sub-cellular locations in response to continuous or pulsed IL-6 stimulation. Interestingly, appropriately pulsed signalling is expected to increase phospho-STAT3 levels more than continuous stimulation with the same concentration of IL-6. The model also provides a platform for predicting the responses of the network to mutations or inhibitors of IL-6 signalling components. IL-6 induced changes in TGF-$\beta$ signalling can be predicted with the model.

5.1 Introduction

Interleukin-6 (IL-6) is a multifunction cytokine which influences both acute and chronic inflammation [134]. As a pro-inflammatory cytokine IL-6 can stimulate B cells [186]. Beside the auto-immune responses and cellular growth regulation, IL-6 signalling acts as a mediator in many chronic inflammatory diseases,
including multiple sclerosis and cancer [11,213]. Since the up-regulation of IL–6 has been reported in different types of human cancer (including renal, ovarian, breast, pancreatic and lung cancer), many anti-cancer treatments now aim to target the IL – 6 cytokine or its signalling pathways (see [202] for a review). Despite its pro-inflammatory role in chronic inflammatory diseases, IL – 6 can decrease inflammation during viral infections [353].

IL – 6-type cytokines participate in many cellular processes including differentiation, apoptosis, proliferation, inflammation and immune responses [175, 431]. IL – 6 signalling activates intracellular biochemical cycles from the plasma membrane to the nucleus. Signalling occurs when IL – 6 binds to its specific receptor IL-6R (also called Gp80) [448]. The IL – 6 :IL-6R complex then binds a common signal transduction, 130 kDa glycoprotein (Gp130) [253]. All members of IL – 6 family of cytokines (IL – 6, IL-11, leukaemia inhibitory factor (LIF), oncostatin M (OSM), ciliary neurotrophic factor (CNTF) and cardiotrophin-1 (CT-1)) use Gp130 to transfer the signalling from the cell surface to the cytoplasm [191]. All IL – 6-type cytokines induce their signalling through the activation of Janus kinases (JAK) [474]. Deregulation in the signalling pathways stimulated by IL – 6 member cytokines (e.g. IL – 6 and IL-11) is often associated with human hematological and epithelial malignancies [109]. Moreover, IL – 6 is also known as a pleiotropic cytokine that can potentially promote or inhibit tumors [251].

IL – 6 signalling interacts with other signalling pathways due to the involvement with Gp130, and the alternative downstream signalling (JAK/STAT and MAPK). There is evidence that IL – 6 crosstalks with cytokines such as, BMP [493], NF–κβ [31], Wnt/β-catenin [47]and TGF–β [207]. The focus of our study is on the IL – 6 and TGF–β crosstalk which involves the role of SMAD7 [207]. The interactions between TGF–β and IL – 6 signalling are important due to their involvement in inflammatory, auto-immune diseases and cancer. Jenkins et al. [207] used "MEF" (mouse embryonic fibroblast) cells with mutated Gp130 in order to over-express STAT3. STAT3 hyperactivated MEF cells have impaired activation and nuclear translocation of SMAD2 in response to TGF–β [207]. Both TGF–β and IL – 6 signalling are over-activated in Erlotinib-resistant cells [495] and both signalling systems are essential for the survival of lung cancer cells [495]. The crosstalk of TGF–β and IL – 6 has also been studied by Luwor et al. [287]: the STAT3/SMAD7/TGF–β interactions are also affected by the activation of the EGFR. Persistent activation of STAT3 induces SMAD7 expression, which then desensitizes TGF–β transduction [207, 287].

Considering the network of processes responsible for the regulation of IL – 6 signalling, it is important to study this signal transduction quantitatively. Since the final responses of the cells will vary with respect to the strength and timing of the initial signal, understanding the temporal changes in the IL – 6 signalling components is critical [174, 406]. Models in Systems biology are defined as a mapping from a biological system to a more accessible, immediate and con-
5.2 Model Development

As mentioned above, IL – 6 signals via two separate pathways, Jak/STAT and MAPK. The latter has been studied several times in the literature (for a review see [214]). MAPK is activated by other signalling pathways including receptor kineses such as the EGFR [185, 222, 258, 263, 375]. We examine the crosstalk between IL – 6 and TGFβ signalling [207], and focus on the STAT3/SMAD7 connection via the Jak/STAT pathway.

Based on recent data [402, 479], we have designed a detailed model of the initial phases of IL – 6 signalling which focuses on the receptor configuration and membrane reactions (Figure 5.1).
5.2 Model Development

Figure 5.1: Topology of the IL-6 signalling pathway model. Potential phosphorylation sites of the components are specified with empty circles. Arrows pointing to 6 blue dots represent degradation process. Oval-shaped components represent micro-RNAs. The dot arrows represent the transcription and translation reactions.

The IL-6 cytokine first interacts with the Gp80 receptor (also called IL-6Rα chain) [37, 176, 317, 382, 404, 405]. Then, the IL-6/Gp80 complex binds with the JAK-associated Gp130, resulting in a ligand receptor complex. The complex dimerizes and activates JAK through an autoprophosphorylation process [317]. Phosphorylated JAK subsequently phosphorylates Gp130 and leads to STAT3 binding [79, 138, 478]. After phosphorylation STAT3 dimerizes, leaves the receptor complex and starts a nucleocytoplasmic shuttling cycle that involves activation of transcription and dephosphorylation of the STAT3 in the nucleus [232, 490]. Phosphorylated Gp130 can bind SOCS3 [15] which blocks STAT3 phosphorylation. The SOCS3 gene is activated by the phosphorylated STAT3 dimer [103, 321].

The components of the IL-6 signalling pathway are listed as Table 5.1:
<table>
<thead>
<tr>
<th>Components as variables</th>
<th>Components signs in simulation and codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL – 6 receptor, Gp80</td>
<td>[Gp80]</td>
</tr>
<tr>
<td>IL – 6 receptor, Gp130</td>
<td>[Gp130]</td>
</tr>
<tr>
<td>Gp80/IL – 6 complex</td>
<td>[Gp80/L]</td>
</tr>
<tr>
<td>JAK associated Gp130</td>
<td>[Gp130/J]</td>
</tr>
<tr>
<td>Ligand-receptor complex</td>
<td>[C]</td>
</tr>
<tr>
<td>Ligand-receptor complex dimer</td>
<td>[C₂]</td>
</tr>
<tr>
<td>Ligand-receptor complex with activated JAKs</td>
<td>[CJ*]</td>
</tr>
<tr>
<td>Ligand-receptor complex with activated STAT3s</td>
<td>[CST*]</td>
</tr>
<tr>
<td>Ligand-receptor complex binding with SOCS3s</td>
<td>[CSc]</td>
</tr>
<tr>
<td>STAT3</td>
<td>[STAT3]</td>
</tr>
<tr>
<td>STAT3 dimer</td>
<td>[STAT3₂]</td>
</tr>
<tr>
<td>PSTAT3</td>
<td>[PSTAT3]</td>
</tr>
<tr>
<td>PSTAT3 dimer</td>
<td>[PSTAT3₂]</td>
</tr>
<tr>
<td>SOCS3</td>
<td>[SOCS3]</td>
</tr>
<tr>
<td>SOCS3 mRNA</td>
<td>[socs3]</td>
</tr>
<tr>
<td>Extracellular IL – 6 ligand</td>
<td>[IL – 6]</td>
</tr>
<tr>
<td>SMAD7 mRNA</td>
<td>[S7mRNA]</td>
</tr>
<tr>
<td>SMAD7</td>
<td>[S7]</td>
</tr>
</tbody>
</table>

Table 5.1: List of the components of the IL – 6 signalling model shown in Figure 5.1 and the standard initial conditions for the model which were constructed from the literature

Based on the IL–6 signalling model in Figure 5.1, the differential equations of the reaction network are of the following form. Note that we have never solved the full model of IL – 6 signalling which is described by Table 5.2, however, the table is useful for understanding the reduction methodology.
### 5.2 Model Development

<table>
<thead>
<tr>
<th>Diff. eq. based on kinetics</th>
<th>Raw diff. Eq. for simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d/dt[Gp80] = v1 - v2 - v3$</td>
<td>$d/dt[Gp80] = v_{80} - k_{a80}[Gp80] - k_{a80/L}[IL - 6][Gp80] + k_{d80/L}[Gp80/L]$</td>
</tr>
<tr>
<td>$d/dt[Gp80/L] = v3 - v4$</td>
<td>$d/dt[Gp80/L] = k_{a80/L}[IL - 6][Gp80] - k_{d80/L}[Gp80/L] - k_{ac}[Gp80/L][Gp130J] + k_{dc}[C]$</td>
</tr>
<tr>
<td>$d/dt[C] = v4 - 2v7$</td>
<td>$d/dt[C] = k_{ac}[Gp80/L][Gp130/J] - k_{dc}[C] - 2k_{a2}[C][C] + 2k_{d2}[C2]$</td>
</tr>
<tr>
<td>$d/dt[C2] = v7 - v8$</td>
<td>$d/dt[C2] = k_{a2}[C][C] - k_{d2}[C2] - k_{aJ}[C2] + k_{dJ}[CJ^*]$</td>
</tr>
<tr>
<td>$d/dt[CJ^*] = v8 - v9$</td>
<td>$d/dt[CJ^<em>] = k_{aJ}[C2] - k_{dJ}[CJ^</em>] - k_{ph}[STAT3][STAT3][CJ^<em>] + k_{dst}[CST^</em>] - k_{aCSc[CJ^*][SOCS3][SOCS3]}$</td>
</tr>
<tr>
<td>$d/dt[CST^*] = v13 - v11$</td>
<td>$d/dt[CST^<em>] = k_{ph}[CJ^</em>][STAT3][STAT3] - k_{dst}[CST^*]$</td>
</tr>
<tr>
<td>$d/dt[CSc] = v12 - v14$</td>
<td>$d/dt[CSc] = v_{ST} - k_{0CSc[CSc]} + k_{aCSc[CJ^*][SOCS3][SOCS3]}$</td>
</tr>
<tr>
<td>$d/dt[STAT3] = v15 - v16$</td>
<td>$d/dt[STAT3] = -k_{ph}[STAT3][STAT3][CJ^*] + k_{eST}[STAT3][STAT3] - k_{0ST}[STAT3]$</td>
</tr>
<tr>
<td>$d/dt[PSTAT32] = v11 - v17$</td>
<td>$d/dt[PSTAT32] = k_{dst}[CST^*] - k_{iST}[PSTAT32]$</td>
</tr>
<tr>
<td>$d/dt[PSTAT32in] = v17 - v18$</td>
<td>$d/dt[PSTAT32in] = k_{iST}[PSTAT32] - k_{dph}[PSTAT32in]$</td>
</tr>
<tr>
<td>$d/dt[STAT3in] = 2v18 - v15$</td>
<td>$d/dt[STAT3in] = 2k_{dph}[PSTAT32in] - k_{eST}[STAT3in]$</td>
</tr>
<tr>
<td>$d/dt[socs3in] = v19 - v20$</td>
<td>$d/dt[socs3in] = k_{trac}[PSTAT32in][STAT3in + K_{Sc}] - k_{emRNA}[socs3in]$</td>
</tr>
<tr>
<td>$d/dt[SOCS3] = v22 - v12$</td>
<td>$d/dt[SOCS3] = k_{bSc}[SOCS3] + k_{trac}[socs3] - 2k_{aCSc[C130^*][SOCS3][SOCS3]}$</td>
</tr>
</tbody>
</table>

Table 5.2: Differential equations used to model the reactions of the IL – 6 signalling pathway shown in Figure 5.2

The critical components and reactions for IL – 6 signalling are summarized in Figure 5.2. The dashed line represents a complex series of interactions, including genetic expression and mRNA processes (see [15] for a review) and the two headed arrows specify reversible reactions.
5.3 Analysis

5.3.1 Reduction Implementation on the Receptor Part

In accordance with to the rapid equilibrium assumption [115, 203], the membrane components of the IL-6 transduction can be reduced to the main components IL-6, Gp80 and Gp130. The relatively fast kinetic rates of the interactions on the membrane compared to the other reactions (such as import, export, transcription and translation of the intracellular proteins) enables us to apply the reduction process. It is reported that the order of magnitudes for signalling processes can be categorized as Table 5.3 [340].

<table>
<thead>
<tr>
<th>Process</th>
<th>Order of magnitude</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalling reactions</td>
<td>$&lt; 10^{6}\text{s}$</td>
<td>[150,271]</td>
</tr>
<tr>
<td>Transcriptional regulation</td>
<td>$10^2\text{s}$</td>
<td>[304,367,515]</td>
</tr>
<tr>
<td>Receptor internalization</td>
<td>$10^2\text{s}$</td>
<td>[119,211]</td>
</tr>
<tr>
<td>ligand dimerization ans receptor dimerization</td>
<td>$10^{-3}\text{s}$</td>
<td>[374]</td>
</tr>
</tbody>
</table>

Table 5.3: Separation of the intracellular reactions based on their timescale [340]

The following reactions on the membrane are fast and consequently at equilibrium these reactions can be reduced:
In IL-6:JAK:STAT signalling, Gp130/JAK are pre-associated before binding of IL−6 [15, 176]. However in the equations above, we added a pre-step to simulate Gp130 and JAK binding. Consequently, the equations 5.1 and 5.2 can be treated similarly. Based on the fast reactions above, the equilibrium values of the components can be written as:

\[
[Gp80/L] = \frac{[IL−6][Gp80]}{K_{80}}
\]

\[
[Gp130/J] = \frac{[J][Gp130]}{K_{130}}
\]

\[
[C] = \frac{[Gp80/L][Gp130/J]}{K_1} = \frac{[IL6][Gp80][J][Gp130]}{K_{80}K_{130}K_1}
\]

\[
[C_2] = \frac{[C]^2}{K_2} = \frac{[IL6]^2[Gp80]^2[J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2}
\]

\[
[CJ^*] = \frac{[C_2]}{K_2^2} = \frac{[IL6]^2[Gp80]^2[J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2^2}
\]

Now we can write the equations for the total concentrations of Gp80 and Gp130 based on the conservation law. These equations are written according to the participation of each receptor in the membrane complexes:

\[
[Gp80]_T = [Gp80] + [Gp80/L] + [C] + 2[C_2] + 2[CJ^*]
\]

\[
[Gp130]_T = [Gp130] + [Gp130/J] + [C] + 2[C_2] + 2[CJ^*]
\]

On the other hand, we can write the following equations for the derivation of the receptors with respect to time:

\[
\frac{d[Gp80]}{dt} = v_{80} - k_{80}[Gp80] - 2(k_aSCS[Gp80] + 2k_d22[C_2])
\]

\[
\frac{d[Gp130]}{dt} = v_{130} - k_{130}[Gp130] - 2(k_aSCS[Gp80] + 2k_d22[C_2])
\]

We define f_1 and f_2 as the RHSs of \(\frac{d[Gp80]}{dt}\) and \(\frac{d[Gp130]}{dt}\) in equations 5.8, respectively. By differentiating the right hand sides of the conservation equations...
(equations 5.7) and equating them with the reactions by which the total number of receptors vary \((f_1 \text{ and } f_2)\), we have:

\[
A_{11} \frac{d[Gp80]}{dt} + A_{12} \frac{d[Gp130]}{dt} = f_1 \\
A_{21} \frac{d[Gp80]}{dt} + A_{22} \frac{d[Gp130]}{dt} = f_2
\]  
(5.9)

where,

\[
A_{11} = 1 + \frac{[L]}{K_{80}} + \frac{[L][J][Gp130]}{K_{80}K_{130}K_1} + \frac{4[L]^2[Gp80][J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2}(1 + 1/K_2^*)
\]

\[
A_{21} = \frac{[L][J][Gp130]}{K_{80}K_{130}K_1} + \frac{4[L]^2[Gp80][J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2}(1 + 1/K_2^*)
\]

\[
A_{12} = \frac{[L][Gp80][J]}{K_{80}K_{130}K_1} + \frac{4[L]^2[Gp80]^2[J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2}(1 + 1/K_2^*)
\]

\[
A_{22} = 1 + \frac{[J]}{K_{130}} + \frac{[L][Gp80][J]}{K_{80}K_{130}K_1} + \frac{4[L]^2[Gp80]^2[J]^2[Gp130]^2}{K_{80}^2K_{130}^2K_1^2K_2}(1 + 1/K_2^*)
\]

and,

\[
f_1 = v_{80} - k_{\emptyset80}[Gp80] - 2R
\]

\[
f_2 = v_{130} - k_{\emptyset130}[Gp130] - 2R
\]  
(5.11)

and, \(R\) is given by:

\[
R = k_{acSc}[SOCS3]\frac{[CJ^*]}{[CJ^*] + K} + k_{dJ}[CJ^*] + k_{d2}[C_2].
\]  
(5.12)

Note that we have substituted the equilibrium values for all membrane components, equation 5.9 can be decoupled using Cramer’s rule [355].

### 5.3.2 Modelling of the Nucleocytoplasmic Reactions

According to our IL-6 signalling model, STAT3 phosphorylation occurs after binding to the IL-6:IL-6R:Gp130 complex. The biochemical reactions can be represented as:

\[
CJ^* + 2STAT3 \rightarrow CST^* \rightarrow CJ^* + PSTAT3_2
\]  
(5.13)

Assuming that the intermediate reactions are sufficiently fast, this can be reduced to:

\[
2STAT3 \xrightarrow{CJ^*} PSTAT3_2
\]  
(5.14)

\(CJ^*\) represents the enzyme that catalyses the phosphorylation process. The enzymatic reaction is modelled using Michaelis–Menten kinetics.
5.3.3 Simplified Model

The rapid equilibrium assumption [115, 203] significantly decreases the complexity of the IL - 6 signalling network. Gp130 uses a specific mechanism to phosphorylate STAT3 and transfer the signal to the nucleus: As shown in Figures 5.1 and 5.2, STAT3 becomes phosphorylated, activated and dimerized after binding with the JAK-associated Gp130 complex. Next, the phosphorylated STAT3 dimer dissociates from the membrane receptor complex. Using the mass action law for STAT3 phosphorylation adds extra complexity to the system, however, the membrane receptor complex can be considered as an enzyme and the STAT3 phosphorylation process an enzymic reaction. Michaelis–Menten kinetics replaces the mass action kinetics in Figure 5.3.

![Simplified model of IL - 6 signal transduction.](image)

The equations used for the Numerical Simulations and Validation sections are (the equations of the simplified model):

Figure 5.3: Simplified model of IL - 6 signal transduction. The dashed line specifies an enzymic reaction. Arrows pointing to blue dots represent degradation process. The stars represent for the phosphorylated species. The dot arrows represent the transcription and translation reactions.
\[
\frac{d[Gp80]}{dt} = (f_1 A_{22} - f_2 A_{12}) / (A_{11} A_{22} - A_{21} A_{12})
\]
\[
\frac{d[Gp130]}{dt} = (f_2 A_{11} - f_1 A_{21}) / (A_{11} A_{22} - A_{21} A_{12})
\]
\[
\frac{d[STAT3]_C}{dt} = v_{ST} - k_{\emptyset ST}[STAT3]_C - 2k_{phc}[CJ^*][STAT3]_C^2 / [STAT3]_C + K_2^2 - 2k_{dph}[PSTAT32]_C + k^e Vnc[STAT3]_N - k^d[STAT3]_C
\]
\[
\frac{d[PSTAT32]_C}{dt} = k_{phc}[CJ^*][STAT3]_C^2 / [STAT3]_C^2 + K_2^2 - k_{dph}[PSTAT32]_C - k^d[STAT3]_C + k^e Vnc[PSTAT32]_N
\]
\[
\frac{d[SOCS3]_C}{dt} = k_{trnsrk}^e Vnc[socs3]_N - k_{\emptyset Sc}[SOCS3]_C
\]
\[
\frac{d[STAT3]_N}{dt} = 2k_{dph}[PSTAT32]_N - 2k_{phn}[STAT3]_N^2 + k^d Vcn[STAT3]_C + k^e Vnc[STA33]_C
\]
\[
\frac{d[PSTAT32]_N}{dt} = k^d Vcn[STAT3]_C - k_{dph}[PSTAT32]_N + k_{phn}[STAT3]_N^2 - k_{phn}[PSTAT32]_N - k^d [PSTAT32]_N
\]
\[
\frac{d[socs3]_N}{dt} = -k_{\emptyset mRNA}[socs3]_N - k_{trnsrk}^e [socs3]_N + k_{trnslk}[PSTAT32]_N / [PSTAT32]_N + K_{Sc}
\]

In these equations we assumed all the import and export kinetic rates are the same for different components. Vnc and Vcn are defined as the (volume of the nucleus)/(volume of the cytoplasm) and (volume of the cytoplasm)/(volume of the nucleus), respectively. The kinetic rates and their units are listed in Tables 5.4 and 5.5

<table>
<thead>
<tr>
<th>Kinetic rates</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{80}$</td>
<td>Production rate of Gp80 receptor</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{\emptyset 80}$</td>
<td>Degradation rate of Gp80 receptor</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$v_{130}$</td>
<td>Production rate of Gp130 receptor</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{\emptyset 130}$</td>
<td>Degradation rate of Gp130 receptor</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$K_{80}$</td>
<td>Ligand:Gp80 binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_{130}$</td>
<td>Gp130:JAK binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_1$</td>
<td>Ligand-Gp80 binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_2$</td>
<td>Gp80/L:Gp130/J binding constant</td>
<td>$nM$</td>
</tr>
<tr>
<td>$K_2^*$</td>
<td>$C_2:CJ^*$ binding constant</td>
<td>$-$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STAT Proteins</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{ST}$</td>
<td>Production rate of cytoplasmic STAT3 receptor</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{\emptyset ST}$</td>
<td>Degradation rate of cytoplasmic STAT3 receptor</td>
<td>$min^{-1}$</td>
</tr>
</tbody>
</table>
5.3 Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{phc}$</td>
<td>Phosphorylation and dimerization rate of STAT3 protein in the cytoplasm</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{dph}$</td>
<td>Dephosphorylation rate of PSTAT3 protein dimer</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{phn}$</td>
<td>Phosphorylation and dimerization rate of STAT3 protein in the nucleus</td>
<td>$nM^{-1}min^{-1}$</td>
</tr>
<tr>
<td>$k_{PSn}$</td>
<td>Degradation rate of phosphorylated STAT3 dimer</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k^i$</td>
<td>Nuclear import rate of cytoplasmic complexes</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k^e$</td>
<td>Nuclear export rate of cytoplasmic complexes</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>Phosphorylation of STAT3 binding constant</td>
<td>$nM$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{trnsl}$</td>
<td>Translation rate of SOCS3</td>
<td></td>
</tr>
<tr>
<td>$k_{BSc}$</td>
<td>Degradation rate of SOCS3 in the cytoplasm</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{bsmRNA}$</td>
<td>Degradation rate of socs3 mRNA in the nucleus</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{trac}$</td>
<td>Transcription rate of socs3 mRNA</td>
<td>$nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{dJ}$</td>
<td>Basal degradation rate of $CJ^*$ on the membrane</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{aCSc}$</td>
<td>Degradation rate of $CJ^*$ stimulated with negative feedback</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$k_{d2}$</td>
<td>Basal degradation rate of $C_2$ on the membrane</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$K_{Sc}$</td>
<td>Transcription of socs3 mRNA binding constant</td>
<td>$nMmin^{-1}$</td>
</tr>
</tbody>
</table>

Table 5.4: Kinetic rates and binding constants of the model

<table>
<thead>
<tr>
<th>Parameters symbol</th>
<th>Literature value</th>
<th>Reference</th>
<th>Scaled data for our model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{80}$</td>
<td>—</td>
<td></td>
<td>0.17 $nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{d80}$</td>
<td>—</td>
<td></td>
<td>0.01 $min^{-1}$</td>
</tr>
<tr>
<td>$v_{130}$</td>
<td>—</td>
<td></td>
<td>0.15 $nMmin^{-1}$</td>
</tr>
<tr>
<td>$k_{d130}$</td>
<td>—</td>
<td></td>
<td>0.01 $min^{-1}$</td>
</tr>
<tr>
<td>$K_{80}$</td>
<td>0.5</td>
<td>[402]</td>
<td>1 $nM$</td>
</tr>
<tr>
<td>$K_{130}$</td>
<td>0.5</td>
<td>[402]</td>
<td>1 $nM$</td>
</tr>
<tr>
<td>$K_1$</td>
<td>1</td>
<td>[402]</td>
<td>1 $nM$</td>
</tr>
<tr>
<td>$K_2$</td>
<td>5</td>
<td>[402]</td>
<td>5 $nM$</td>
</tr>
<tr>
<td>$K_2^*$</td>
<td>5</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>$v_{ST}$</td>
<td>—</td>
<td></td>
<td>0.25 $nMmin^{-1}$</td>
</tr>
</tbody>
</table>
5.4 Numerical Simulations and Validation

Similar to the Chapter 4, the complexity and non-linearity of the simplified IL – 6 model make us to use numerical methods rather than analytical methods.

5.4 Numerical Simulations and Validation

We have quantitated the published experimental [479] Western blot results for STAT3, PSTAT3 and SOCS3 protein expression levels in mouse liver cells. These quantitative results were compared with the levels of these components predicted by our IL – 6 signalling model. Due to STAT3 and PSTAT3 nucleocytoplasmic shuttling, our model design requires total STAT3 concentration level (including Phosphorylated STAT3 in the cytosol and nucleus) to be constant in time. Experimental results from Wormald et al. [479] support the model assumption that total STAT3 concentration is relatively stable (see Figure 5.4). The slight decrease in the level of total STAT3 at the beginning of the simulation is due to the time-scale of the STAT3 phosphorylation process. By increasing the time-scale of phosphorylation reaction, the simulation curve converges to the total, experimental STAT3 steady-state level faster. Note that total STAT3 (monomer) concentration is defined as $[STAT3]_C + 2[PSTAT3]_C + V_{nc}([STAT3]_N + 2[PSTAT3]_N)$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{qST}$</td>
<td>$-$</td>
<td>$0.4 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{phc}$</td>
<td>$-$</td>
<td>$0.0005 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{dph}$</td>
<td>$-$</td>
<td>$0.005 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{phn}$</td>
<td>$-$</td>
<td>$0.005 \text{ nM}^{-1}\text{min}^{-1}$</td>
</tr>
<tr>
<td>$k_{qPSn}$</td>
<td>$-$</td>
<td>$0.1 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k^i$</td>
<td>$0.005 \text{ min}^{-1}$</td>
<td>[402]</td>
</tr>
<tr>
<td>$k^e$</td>
<td>$-$</td>
<td>different for different components $\text{min}^{-1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>$-$</td>
<td>$0.5 \text{ nM}$</td>
</tr>
<tr>
<td>$k_{trans}$</td>
<td>$0.01 \text{ min}^{-1}$</td>
<td>[402]</td>
</tr>
<tr>
<td>$k_{qSc}$</td>
<td>$-$</td>
<td>$0.15 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{qmRNA}$</td>
<td>$-$</td>
<td>$1 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{trac}$</td>
<td>$0.01 \text{ nM} \text{min}^{-1}$</td>
<td>[402]</td>
</tr>
<tr>
<td>$k_d$</td>
<td>$-$</td>
<td>$0.01 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{aCSc}$</td>
<td>$-$</td>
<td>$0.01 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$k_{d2}$</td>
<td>$-$</td>
<td>$0.01 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$K_{Sc}$</td>
<td>$400 \text{ nM}$</td>
<td>[402]</td>
</tr>
</tbody>
</table>

Table 5.5: The parameter values of simplified IL – 6 model
5.4 Numerical Simulations and Validation

Figure 5.4: Total STAT3 concentration in liver extracts after IL−6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves represent the fitted curve for the experimental data and the predicted values from the model, respectively. The fitted curve is plotted via trendline tool of Microsoft Excel.

In order to study the changes in STAT3 concentration level in the shorter time-scale, the simulation results as a function of STAT3 concentration have been plotted in a semi log graph (see Figure 5.5). The x axis in Figure 5.5 is a log scale and the y axis is magnified in order to elaborate the short-term variations of STAT3. The steady-state level of total STAT3 is slightly higher that its initial value. As mentioned earlier, the difference in the time-scale of the phosphorylation and dephosphorylation kinetic rates in the model is the main reason for the protein level variation.

Figure 5.5: Total STAT3 concentration time course in liver extracts after IL−6 stimulation in log timescale. The red curve is the simulation result. The figure suggests a difference between the initial level of total STAT3 and its steady-state level. The fitted curve is plotted via trendline tool of Microsoft Excel.

The levels of total PSTAT3 and SOCS3 were shown by Wormald et al. [479]. Figures 5.6 and 5.7 show the experimentally measured levels of total PSTAT3
5.4 Numerical Simulations and Validation

and SOCS3. The predicted trends for these protein concentrations simulated via our model are shown in Figures 5.6 and 5.7. Although the Wormald’s published data was not quantitated [479], we were able to use their gel images to estimate the total STAT3, total PSTAT3 and SOCS3 levels. The original photographic films were kindly provided by S. Wormald. The dots show the experimental values from Wormald et al. data [479]

Figure 5.6: Total PSTAT3 concentration in liver extracts after IL – 6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves represent the fitted curve for the experimental data and the predicted values from the model, respectively. The fitted curve is plotted via trendline tool of Microsoft Excel.

Figure 5.7: SOCS3 concentration in liver extracts after IL – 6 stimulation. The dots represent the experimental values derived from figure 1 of Wormald et al. [479]. The blue and red curves demonstrate the fitted curve for the experimental data and the predicted values from the model, respectively.

The final results for our simulations demonstrate that our simplified model of IL – 6 signalling pathway can predict the protein and phospho-protein levels accurately.
The general trend of the experimental and modelling curves for total STAT3, PSTAT3, and SOCS3 are similar. PSTAT3 measured by experiment peaks at 50 min and the peak value is almost 10 times larger than its steady-state value. The modelling predicts differences in the total concentration of PSTAT3 verses time. The simulation predicts protein levels up to 1000 min. Due to the lack of data on the absolute amount of each protein on the gels in Wormald’s paper, PSTAT3 and SOCS3 levels were normalized to the total STAT3 loadings.

Figure 5.7 shows the SOCS3 time course for both the experiment and the simulation. The simulation shows a peak signal approximately an hour after IL-6 stimulation which is consistent with the results from the experiment. In both the experimental and simulation curves, the peak value is more than 15 fold higher than the steady-state value for the concentration of SOCS3.

SMAD7 level can be predicted using our IL – 6 model. More specifically, the SMAD7 level varies proportional to the PSTAT3 dimer in the nucleus i.e. the transcription factor. Consequently, the SMAD7 concentration is defined as $[PSTAT3]^2_N / ([PSTAT3]^2_N + 1)$ in Figure 5.8 and Figure 5.9. Figure 5.8 shows time dependent changes in SMAD7 concentration for both linear and non-linear PSTAT3/SMAD7 relation. The linear relation is modeled using mass-action kinetics, while Michaelis-Menten kinetics are used to show non-linear stimulation of SMAD7 via PSTAT3 transcription factor. Furthermore, the IL – 6 ligand concentration changes the SMAD7 dynamics. Specifically, SMAD7 seems to decay later at higher ligand concentration (see Figure 5.9).

![Figure 5.8: SMAD7 concentration time course for linear and non-linear stimulation via nuclear PSTAT3.](image-url)
Figure 5.9: SMAD7 concentration time course dynamics for IL−6 concentration input = 0.5, 100 and 400 (the concentration units are arbitrary).

An advantage of a signalling model is that its predictions are not specific to a single cell line. To show the compatibility of our model we have validated the simplified IL−6 signalling model with a set of experimental data extracted from wild type mouse embryonic fibroblasts (MEFs). Figure 5.10 shows how the model simulation of PSTAT3 dynamics explains the experimental within 1 hour after IL−6 stimulation. For data extraction method of Figure 5.10 see "Material and Methods" section. Changes in PSTAT3 concentration level to long-time IL−6 stimulation is shown in Figure 5.11. Dots with the same colour belong to the same experimentally collected data set. The model simulation curve explains the experimental results more precisely for short-time rather than long-time IL−6 stimulation (see Figure 5.10 and Figure 5.11).

Figure 5.10: Short-time simulation of total PSTAT3 dynamics. The blue curve shows the simulation results and the dots represent the relative experimental data for 0-1 h IL−6 stimulation.
Another interesting experiment which can be conducted via modelling is to investigate the response of the signalling system to pulsatile IL–6 inputs. Figure 5.12 shows how PSTAT3 concentration follows different IL–6 stimulations patterns. It is noticeable that the PSTAT3 steady-state changes reversely with the "on" time increase of the stimuli. This observable phenomenon can be explained from the topology of the IL–6 signalling network (see Figure 5.3). 20 min almost equals the time for IL–6 signalling to reach the activated nuclear PSTAT3 and hence, the transcription of socs3 and activation of the feedback. The degradation of PSTAT3 happens only in the nucleus and is linearly proportional to the nuclear PSTAT3 concentration (Mass Action kinetics). However, the production rate of PSTAT3 (the STAT3 phosphorylation rate) is a nonlinear function of the cytoplasmic PSTAT3 (Michaelis–Menten kinetics). Additionally, the "off" time is not long enough to return all the signalling components to zero, including cytoplasmic and nuclear PSTAT3. Therefore, total PSTAT3 accumulates in the cell and peaks when the cytoplasmic and nuclear PSTAT3 peaks around 60 min. Since the activated receptor complex is limited, the signalling becomes attenuated gradually and total PSTAT3 concentration decreases after an hour.

As the stimulation "on" time decreases, the signalling affects the down-stream components less. The accumulated level PSTAT3 in the nucleus reduces comparing to longer "on" time stimulations, while cytoplasmic PSTAT3 production follows the same pattern. Consequently, the transient (at each pulse) and steady-state levels of total PSTAT3 increase globally. Figure 5.12 also suggests that a pulsatile stimulation pattern may provide an alternative to the production of genetically modified cell lines, such as MEF Gp130\textsuperscript{F/F}, in which the SOCS3 feedback loop is knocked out in order to over-express STAT3.
5.4 Numerical Simulations and Validation

Figure 5.12: PSTAT3 responses of IL–6 signalling system to a pulsatile input. Different curves correspond to distinct pulsation patterns.

By increasing the level of receptors concentration on the membrane, the IL–6 signalling attenuation which occurs after 1 hour, can be avoided. Therefore, the steady-state level of PSTAT3 increases by increasing the total receptor concentration (see Figure 5.13). In order to vary the membrane receptor concentration in Figure 5.13, we changed the production rates Gp80 and Gp130 simultaneously. Variable ‘m’ in Figure 5.13 is proportional to the production rates of Gp80 and Gp130 at the membrane. Additionally, the effect of SOCS3 negative feedback loop was decreased to decrease the degradation rate of both receptors.

Continual high level of total PSTAT3 concentration represents the responses of cells in chronic inflammation. Our IL–6 signalling model predicts that chronic inflammation and hence cancer occur when the level of total receptor concentration increases.

Figure 5.13: PSTAT3 responses of IL–6 signalling system to different levels of membrane receptors. "m" represents the level of receptor concentrations on the membrane.
5.5 Discussion

Understanding the IL–6 signalling pathways and its components has attracted the interest of scientists for decades [186, 327]. Due to its roles in inflammation, IL–6 signalling influences the progression of tumors, such as: melanoma [214, 316], multiple myeloma [214, 228], prostate cancer [214, 407] and colon cancer [32, 359].

A detailed mathematical models of IL–6 signalling was proposed via Singh et al. [402]. Although their model is comprehensive, no comparison with experimental data was attempted by the authors. Subsequent models of IL–6 signalling have used more mathematical and statistical analyses on the basis of Singh’s IL–6 model [70, 320], however, these more recent reports focus on MAPK signalling which adds an extra complexity to the system but ignores the major specific action of IL–6 signalling, i.e. the activation of the JAK/STAT system.

We have focused on JAK/STAT signal transduction. Our reduction method simplifies the complexity of the intracellular protein interactions and allows an extended time-scale for the simulations. This reduction method is used in biochemical engineering and non-linear dynamical systems analysis [115, 158, 422]. The final simplified model was reduced to 8 differential equations and was solved numerically. The initial values of the parameters and scaling were based on reported experimental data [464, 479].

Model predictions of changes in total STAT3, total PSTAT3 and SOCS3 concentrations compared favourably with the experimental results of Wormald et al. [479]. Small differences between the experiments and simulations may occur because of the lack of absolute quantitation from the Western blots of Wormald et al. [479]. We have also tested the model results with a set of experimental data from another cell line (MEFs) in order to demonstrate the robustness of our model to changes in the initial conditions. Our model validates and supports our parameter estimations, the model logistics and the reduction method [115]. Furthermore, using this model, predictions can be made for the pulsatile signal inputs and their influences on the ultimate steady-state of the system. Pulsatile responses to IL–6 stimulation are confirmable in future by designing a new experimental set up. The high level of IL–6 signalling in chronic inflammation and cancer was predicted by our model, which shows the robustness of this model against different perturbations. This work can be used as a foundation for an integrated model designed to include the IL–6-JAK-STAT3 signalling in more complex simulations of combined signalling pathways.

In conclusion, modelling of the IL–6 signalling network has helped us understand the importance of the negative feedback loop for the functionality and stability of the IL–6 signalling network. It is interesting to see how one signalling pathway can influence and regulate another signalling network. The importance of time-delays in such networks, caused by slow sets of reactions
or protein complex translocations is also evident in our simulations. We also predict how pulsatile stimulation of a signalling network can affect the strength and the duration of IL – 6 signalling.

5.6 Material and Methods

Three major types of data analysis have been used in this project:

- **Model design**
  In order to model the IL – 6 signalling pathways we have written ordinary differential equations for the concentration of each component using mass action or Michaelis–Menten kinetic equations [249]. The reduction method [115] was applied to the IL – 6 model using rapid equilibrium assumptions. Time-delays were considered in the equations for the PSTAT3-induced SMAD7 stimulation.

- **Computer modelling**
  The simulation of the mathematical model of IL – 6 signalling was performed via MATLAB. We used the curve fitting package from the MATLAB platform and trendline tool of Microsoft Excel in order to fit the model predictions to compare with the experimental data from Wormald et al. [479]. The IL – 6 signalling model figure (Figure 5.1) was first designed via CellDesigner [132] and later with Microsoft Power Point. The MATLAB codes are available on request.

- **Experimental data**
  The experimental data set and the kinetic rates were extracted by data analysis from Wormald et al. published in 2006 [479] in addition to a set of experimental data obtained in the following manner:

  **Cell culture and cell lysis**
  Mouse embryonic fibroblasts (MEFs) cells were isolated from day 13 to 15 embryos. Wild type MEF cells were cultured in DMEM containing 15% FCS. The cells were trypsinized and washed with DMED + 15% FCS before plating. Passage 3 cells with $1 \times 10^6$ MEFs/well were seeded in 60 mm plates for 0-1 hour, treatment with 1 ng/ml IL – 6 (Ref 1661.F33. WEHI). After washing with cold PBS for two times, cells were lysed in ice-cold 200 ul RIPA lysis buffer, containing 1M Tris/HCL, 0.5 M EDTA, 5M NaCl, 10 % Na Doc, 10 % TX-100, 10 % SDS, proteinase inhibitor 100 × and H₂O. The cell lysits were passed through 27 G needle for 5 times, then incubated in ice for 20-30 min. After incubation the samples were span at 13,000 rpm for 30 min at 4°C. The supernatant was transferred to new tubes where 20 ul of samples were saved for BCA protein assay using sigma BCA assay
kit (B9643). 20 ul 5x sample buffer was added to 80 ul loading sample and the samples were heated at 95°C for 10 min.

**Western blotting**

The protein concentrations were determined using the sigma BCA assay kit (B9643), Novex NuPAGE® 4-12%-Bis-Tris (life technologies NP0335 Box) gels were used for loading proper amount of sample lysate for each time point. PSTAT3 (XP™ Rabbit mAb) antibodies were provided by Cell Signalling Technology and were used at a dilution 1:500 in 3% BSA-TBS-T. β-tubulin is used as loading control. Eventually, the gels were transformed onto nitrocellulose membrane via iBlot 2 gel transfer device (life technologies) and the membranes were scanned using Odyssey infrared scanner (LI-COR).

**Protein Quantitation**

The western blot images were quantitated using ImageJ 1.49p. The signals of each protein were normalised to the loading control protein where possible.
Chapter 6

TGF−β and IL−6 Signalling

Crosstalk: an integrated model

6.0 Summary of the information included in Chapter 6

Crosstalk studies are inseparable parts of Systems biology approach. Systems biology studies the crosstalk between the components of a system, emphasizing on how the ultimate systems’s function is influenced by such interactions. Particularly in the case of intracellular signalling networks, in which individual signalling pathways share several components, it is crucial to study other interacting signalling pathways.

Last two chapters were dedicated to individual TGF−β and IL−6 signalling and their self-regulation. In this chapter, we introduce an "integrated" model by which we study the regulation of TGF−β signalling via IL−6 stimulation. The integrated model makes predictions for the formulation and the strength of the crosstalk between the TGF−β and IL−6 signalling pathways.

6.1 Introduction

Cell signalling in complex biological systems regulates cellular functions. Each signalling pathway targets a specific range of genes and consequently controls a set of system’s characteristics. However, signalling pathways do not function as isolated systems, rather many proteins often participate in two or more signalling pathways. The Systems biology approach can be used to study the intracellular signalling networks as an interconnected complex system, where the properties of the system are governed by several mutual interactions. Colorectal cancer, the third most common cancer in the world [362], occurs due to an accumulation of specific genetic alterations. Wnt, PGE2, EGF, TGF−β, TNF-α, IL−6 and IFNg are among the signalling pathways that interfere with the growth of
stomach or colon cancers [80, 362, 369, 412, 494, 498, 501]. All of these signalling pathways have been studied individually in the literature. In Chapters 4 and 5 we modelled and analyzed two of these signalling pathways (TGF−β and IL−6 signalling). This chapter focuses on TGF−β and IL−6 signalling crosstalk and the role of this crosstalk in gastrointestinal tumors.

TGF−β is a pleiotropic cytokine which regulates a broad range of cell’s characteristics i.e. differentiation, proliferation, migration, life time and apoptosis [118, 296]. This cytokine plays a crucial role in tumorigenesis [104, 312, 350]. Acting on early-tumors, TGF−β functions as an anti-proliferative factor that causes tumor suppression [357, 434]. Notwithstanding, late-tumors become stimulated by TGF−β signalling [107, 357]. Mathematical modelling and computational analysis of TGF−β signalling help identify the conditions and components which cause its dual-behaviour. There are reports of TGF−β signalling crosstalk with many other pathways, such as MAPK, PI3K/Akt, Wnt, Hedgehog (Hh), Notch, ILs (including IL−6), TNF−β, and IFN−γ [159]. Extensive crosstalk between individual signalling pathways and TGF−β is due to the potential ability of TGF−β signalling components (e.g. SMADs) to interact with other proteins outside their own direct signalling pathways [159]. Furthermore, TGF−β signal transduction has multiple transcriptional and non-transcriptional targets which are also influenced by other signalling pathways [159]. SMADs participate in the TGF−β signalling responsible for signal delivery from the ligand activation of the receptor to the nuclear activation events. TGF−β downstream signalling is also regulated via its inherent negative feedback originated from one of the SMAD family members, SMAD7 [155, 179, 261, 315, 400, 492].

IL−6 signalling targets genes involved in inflammation and immune responses [175, 190, 337, 421]. IL−6 signals via the JAK-STAT and MAPK signalling pathways (see Chapter 5), which makes IL−6 signalling potentially appropriate for the analysis of crosstalk with other cytokine signalling. STAT3 is the target intracellular protein which connects TGF−β and IL−6 signalings [206]. Similar to TGF−β signalling, the IL−6 system has an intrinsic negative regulatory feedback loop, SOCS3 [15, 176]. Signalling through JAK/STAT pathway and hence STAT3 level is controlled by the level of the SOCS3 protein [206].

Key parameters and mechanism of the TGF−β and IL−6 signalings crosstalk can change dramatically depending on the cell-lines, spatial location of the intracellular proteins and the developmental stage of the cells [4, 159, 402, 511]. Experimental evidence shows that TGF−β signalling can affect IL−6 signal transduction through the MAPK pathway in an autocrine loop [10, 204, 275, 437]. Additionally, TGF−β expression can increase in T helper cells via activation of STAT3 [223, 238]. A TGF−β and IL−6 signalling crosstalk via a STAT3-SMAD7 pathway interaction was first proposed via Jenkins et al. [206]. Jenkins et al. have analyzed Gp130/JAK/STAT3 signalling and the SOCS3 negative feedback in mouse gastric epithelial cells in order to determine the role of IL−6 induced
SMAD7-negative feedback on TGF–β signalling possible interactions [206]. Using the inherent feedback loops controlled by SOCS3, SMAD7 and miR-433 we have connected our previously proposed mathematical models of TGF–β and IL – 6 signalling to study the gastric tumour model described by Jenkins et al. [207].

6.2 Model Development

According to the Jenkins et al. proposed signal exchange between the TGF–β and IL – 6 signalling pathways [206], we have developed an integrated mathematical model that connects these two signal transduction systems. Our integrated model is based on our previous individual models for TGF–β and IL – 6. Figure 6.1 illustrates how the two systems combine. A positive inherent feedback on TGF–β signalling [269] was added to the recommended model suggested by Jenkins et al. The red arrow accentuates the main interaction between the two pathways.

Figure 6.1: The schematic diagram of TGF–β and IL – 6 crosstalk. the red arrow is the focus of the current thesis.

The parameters of each signalling model were re-evaluated and the initial values of variables in TGF–β and IL – 6 systems were adjusted according to experimental data in the literature [206]. The scaled parameters are mainly the kinetic rates and binding constants of the IL – 6 signalling part of the inte-
grated model. In scaling these parameters we have considered the smoothness of the curves, non-negative values for the variables, stability and robustness of the integrated model. We have compared two different kinetics models for the PSTAT3/SMAD7 interaction system: mass-action and Michaelis-Menten kinetics. PSTAT3 is considered to act as a transcription factor in the pathway and activate SMAD7 transcription (Michaelis-Menten enzymic kinetics).

The equations representing the integrated model are listed below. The negative feedback terms of the \([R]\), \([RC]\) and \([PC]\) differential equations have been changed such that the phosphorylated PSTAT3 dimer in the nucleus considered as the initiator.
where we have the followings are defined:

\[ \frac{d[Gp80]}{dt} = (f_1 A_{22} - f_2 A_{12})/(A_{11} A_{22} - A_{21} A_{12}) \]
\[ \frac{d[Gp130]}{dt} = (f_2 A_{11} - f_1 A_{21})/(A_{11} A_{22} - A_{21} A_{12}) \]
\[ \frac{d[STAT3]_C}{dt} = \nu_{ST} - k_{\emptyset ST}[STAT3]_C - 2k_{phc}[CJ^*] \frac{[STAT3]^2_C}{[STAT3]^2_C + K^2} - 2k_{dph}[PSTAT32]_C + k^eVnc[STAT3]_N - k^e[STAT3]_C \]
\[ \frac{d[PSTAT32]_C}{dt} = k_{phc}[CJ^*] \frac{[STAT3]^2_C}{[STAT3]^2_C + K^2} - k_{dph}[PSTAT32]_C - k^i[PSTAT32]_C + k^eVnc[PSTAT32]_N \]
\[ \frac{d[SOCS3]_C}{dt} = k_{trans}k^eVnc[socs3]_N - k_{\emptyset Sc}[SOCS3]_C \]
\[ \frac{d[STAT3]_N}{dt} = 2k_{dph}[PSTAT32]_N - 2k_{phn}[STAT3]^2_N - k^e[STAT3]_N + k'Vcn[STAT3]_C \]
\[ \frac{d[PSTAT32]_N}{dt} = k^iVcn[PSTAT32]_C - k_{dph}[PSTAT32]_N + p_{phn}[STAT3]^2_N - k_{\emptyset PSn}[PSTAT32]_N - k^e[PSTAT32]_N \]
\[ \frac{d[socs3]_N}{dt} = -k_{\emptyset mRNA}[socs3]_N - k_{trans}k^e[socs3]_N \]
\[ \frac{d[R]}{dt} = v_1 - k_1 [R] - 2k_{RC}^+ [R]^2 + 2k_{RC}^- [RC] - k_{RC}^{f+} [P] \frac{[R]}{[R] + K} \]
\[ \frac{d[RC]}{dt} = k_{RC}^+ [R]^2 - k_{RC}^- [RC] - k_{RC} [RC] \]
\[ \frac{d[PC]}{dt} = k_{PC}^+ [TGF_2] [RC] - k_{PC}^- [PC] - k_{PC} [PC] \]
\[ \frac{d[S]}{dt} = v_{S} - k_{S} [S] - k_{S}^+ [PC] \frac{[S]}{[S] + K_S} + k_{S}^- [\dot{S}] \]
\[ \frac{d[\dot{S}]}{dt} = k_{S}^+ [PC] \frac{[S]}{[S] + K_S} - k_{S}^- [\dot{S}] - k_{n}^- [\dot{S}] + k_{n} [S_n] - k_{S}^{-} [\dot{S}] \]
\[ \frac{d[S_n]}{dt} = k_{n}^{-} [\dot{S}] - k_{n} [S_n] - k_{S_n} [S_n] \]
6.3 Results from IL − 6:TGF−β Crosstalk Model and Discussion

Different stimulation patterns are used in this section to predict the responses of the integrated model to various perturbations. For the first experiment, the integrated system is stimulated simultaneously with IL−6 and TGF−β, i.e. "double stimulation". The levels of PSTAT3 and PSMAD2 are studied as the outputs of the integrated system when it is double stimulated. Figure 6.2 compares the PSMAD2 time course in the absence of the IL−6 stimulation with double stimulation. Due to the induced expression of SMAD7 (the negative feedback loop for the TGF−β signalling) via PSTAT3 nuclear accumulation in double stimulation experiment, the PSMAD2 concentration curve decreases to a lower level after its first peak. The lower level of PSMAD2 leads to lower nuclear SMAD7 production in TGF−β signal transduction, so the PSMAD2 level rises again, but peaks at a lower level than the first TGF−β induced peak (see Figure 6.2). The dampening of the PSMAD2 peaks is because of the loss of activated IL−6 receptors at the membrane, which leads to the reduction of the PSTAT3 signal an hour after IL−6 stimulation. The positive feedback loop for TGF−β signalling is weakened as the TGF−β signalling becomes down-regulated by SMAD7 activation. As a result, each peak of PSMAD2 is weaker than the previous peak. The steady-state level of PSMAD2 is the same in both stimulation patterns.

\[
[S_3] = [S_n]^3/K_3 \tag{6.2}
\]
\[
[N] = [S_3](t-\tau_2) \tag{6.3}
\]
\[
[P] = K_1^2/(K_1^2 + [S_3]^2(t-\tau_2)) \tag{6.4}
\]

\[
A_{11} = 1 + \frac{[L]}{K_80} + \frac{[L][J][Gp130]}{K_80K_{130}K_1} + \frac{4[L][Gp80][J]^2[Gp130]^2}{K_80K_{130}^2K_1^2K_2}(1 + 1/K_2^*) \tag{6.5}
\]
\[
A_{21} = \frac{[L][J][Gp130]}{K_80K_{130}K_1} + \frac{4[L][Gp80][J]^2[Gp130]^2}{K_80K_{130}^2K_1^2K_2}(1 + 1/K_2^*) \tag{6.5}
\]
\[
A_{12} = \frac{[L][Gp80][J]}{K_80K_{130}K_1} + \frac{4[L][Gp80][J]^2[Gp130]}{K_80K_{130}^2K_1^2K_2}(1 + 1/K_2^*) \tag{6.5}
\]
\[
A_{22} = 1 + \frac{[J]}{K_{130}} + \frac{[L][Gp80][J]}{K_80K_{130}K_1} + \frac{4[L][Gp80][J]^2[Gp130]}{K_80K_{130}^2K_1^2K_2}(1 + 1/K_2^*) \tag{6.5}
\]

\[
f_1 = v_{80} - k_{q80}[Gp80] - 2R
\]
\[
f_2 = v_{130} - k_{q130}[Gp130] - 2R
\]

\[
R = k_{aCSc}[SOC3S3][C.J^*] + k_{d.[C.J^*]} + k_{d2}[C_2]. \tag{6.7}
\]
6.3 Results from IL – 6:TGF–β Crosstalk Model and Discussion

Figure 6.2: A comparison between the PSMAD2 response of the TGF–β signalling system and the integrated IL – 6:TGF–β signalling system. The integrated model output follows the TGF–β model output for the first 60 minutes, however it diverges afterwards. The steady-state PSMAD2 levels of both models are identical.

Our second stimulation scenario includes a two-hour "pre-stimulation" of the integrated system with IL – 6 followed by the TGF–β stimulation which is started one hour after the IL – 6 stimulation. Again the total PSTAT3 and PSMAD2 levels are considered to be the outputs of the IL – 6:TGF–β crosstalk system. Figure 6.3 compares the response of the system to pre-stimulation and the double stimulation patterns. The second peak of PSMAD2 response is stronger in pre-stimulation compared to the double stimulation. When the IL – 6 signalling is terminated after 2 hours, there is a rapid decay in total PSTAT3 concentration and hence SMAD7 levels. Consequently, SMAD7 is only produced through TGF–β signalling after 2 hours and PSMAD2 levels rise to higher levels at the second peak. Again, the steady-state level of PSMAD2 for both stimulation patterns is the same.

Figure 6.3: The integrated model output for the double stimulation and pre-stimulation scenarios of IL – 6:TGF–β signalling.

We have also examined the response of the integrated model to a pulsatile
stimulation of IL – 6 (the TGF−β stimulus is constant, see Figure 6.4). The same pulsation patterns in Figure 5.12 in chapter 5 are used in this experiment. As the "on time" of the pulses increases, the PSMAD2 response trends closer to the output of the sustained response model (see Figure 6.3 for comparison). Studying Figure 5.12, it is clear that the PSTAT3 response is closer to sustained IL – 6 stimulation with pulses of longer "on" duration.

The PSMAD2 dynamics in response to short IL – 6 pulsation is comparable to the system’s output in the absence of IL – 6 signalling (see Figure 6.4 and Figure 6.2). Because of the steady PSTAT3 levels at short IL – 6 pulses (see Figure 5.12), the SMAD7 level stays high and the PSMAD2 level does not peak for a second or third time. Furthermore, the level of PSMAD2 first peak is lower in response to short IL – 6 pulses because of the higher level of the PSTAT3 peak (see Figure 5.12).

Figure 6.4: PSMAD2 responses of the integrated model to different pulsatile inputs of IL – 6. The TGF−β level is kept constant for the entire duration of the simulation.

In the next experiment, we have kept IL – 6 signalling constant and input TGF−β pulses to the integrated system. The TGF−β pulse patterns are identical to the previous experiment (IL – 6 pulsatile stimulation). Figure 6.6 demonstrates the responses of the integrated system to the TGF−β pulsatile stimulation. A comparison between Figure 6.6 and Figure 6.5 reveals that the peaks in the PSMAD2 transient response of the integrated model to IL – 6 and TGF−β stimulation appear because of the PSTAT3/SMAD7 interaction between the functions of the two proteins rather than changes to the positive feedback loop within the TGF−β pathway. Clearly, TGF−β signalling plays the main role in determining the steady-state level of PSMAD2 in the integrated model, since the equilibrium PSMAD2 levels are equivalent in both Figure 6.6 and 6.5. As expected and in a similar fashion to the previous pulsatile stimulation experiments using the integrated model, as the "on" time of the pulses increases, the model response approaches the output level of PSMAD2 for the sustained stimulation condition.
Figure 6.5: The simplified TGF-β model response (proposed in Chapter 4) to pulsatile TGF-β input.

Figure 6.6: The integrated model response to pulsatile TGF-β stimulation.

Figure 6.7 shows the SMAD7 dynamics in the double stimulation of the integrated system with both IL-6 and TGF-β. Because of the incorporation of time-delay in PSTAT3/SMAD7 and PSMAD2/SMAD7 reactions, the SMAD7 level does not start increasing instantly, i.e. there is a 10-min delay. SMAD7 concentration decreases after 60-70 min due to the lost of activated receptors in both signalling pathways. However, the SMAD7 level rises for the second time. As mentioned above, the second peak is caused mainly by the non-zero but low level of PSTAT3 produced in response to the IL-6 stimulation. Figure 6.7 compares how the different kinetics for PSTAT3/SMAD7 reactions affects the total SMAD7 dynamics.
SMAD7 dynamics is studied in a pre-stimulation experiment (see Figure 6.8). Comparing Figure 6.8 with Figure 6.7, the second peak of SMAD7 vanishes as the result of pre-stimulation. The existence of the second SMAD7 peak in the double stimulation experiment was due to the non-zero level of PSTAT3. When the system is pre-stimulated, the PSTAT3 level reduces to zero within 120 minutes (see Figure 6.8).

It is important to note that the SMAD7 level peaks around 100 min in the mass-action model of PSTAT3/SMAD7 kinetics, compared to the Michaelis-Menten kinetics. Whereas, the curve corresponding to Michaelis-Menten kinetics experiences a sudden decrease at 120 min.

In order to measure how sensitive the integrated model is to the PSTAT3/SMAD7 functional interaction, we have plotted the PSMAD2 response of the integrated model to pre-stimulation, while varying the strength of the activation of SMAD7 via PSTAT3 (see Figure 6.9). The "coefficient" in Figure 6.9 is used as a representative of the strength of the PSTAT3/SMAD7 interaction. Greater coef-
The results in Figure 6.9 support the initial hypothesis of this research, i.e. the activated form of PSTAT3 induces the down-regulation of TGF−β signalling and suppresses the PSMAD2 activation through over-expression of SMAD7, the PSMAD2 level is suppressed significantly after 150 min (see Figure 6.9). After the first peak, the effect of IL−6 signalling is eliminated and the PSMAD2 curves loose their sensitivity to PSTA3 and IL−6 signalling.

Figure 6.9: The variation in the integrated model output as the PSTAT3/SMAD7 interaction is being reinforced gradually. Different coefficients represent different strength of the PSTAT3-induced stimulation of SMAD7.

In order to test experimentally our integrated model we have developed and designed an experimental protocol. In this protocol, wild type MEFs were stimulated with IL−6 for 1 hour before their double-stimulation with TGF−β and IL−6. The one hour pre-stimulation with IL−6 acts as the time-delay used in the IL−6:TGF−β signalling connection. The 0 time point represents the initial time of the double-stimulation with TGF−β and IL−6. According to Figure 6.3, the total PSMAD2 concentration undergoes a double-peak fluctuation in the first 300 min, before reaching the steady-state. We have measured the level of total PSMAD2 for 0 to 4 hour after the double-stimulation. This experiment has been designed to test the double-peak appearance in the stimulation of TGF−β and IL−6.

Figure 6.10 shows how the predictions of the integrated model (the blue curve) are in accordance with the experimental data. Each replicates of the experiment is illustrated with a specific colour and marker shape. The points specified by a "cross" are the average of the 3 experimental sets extracted for each time point.
Figure 6.11 is the representative blot of the data shown in the Figure 6.10. It is clear that PSMAD2 level rises for the second time after its minimum at 150 min. The data points in Figure 6.10 are normalised to the loading control, e.g. actin.

![Figure 6.10: The experimental validation of the integrated model. The blue curve is the model prediction of the double-stimulation of TGF-β and IL-6. The colored dots represent the experimental data from wild type MEFs. Each set of experimental data is specified with a color. The "cross"es represent the average of the 3 experimental replicates.](image)

It is important to note that I needed to revise the time-delay parameter considered in the feedback terms of TGF-β-part of the integrated model. A proportion of the SMAD7 is induced by IL-6 signalling which also has a time-delay of 20 min and had to be added to the time-delay of the SMAD7-induced feedback terms.

![Figure 6.11: The representative blots of Figure 6.10. The top bands show the total PSMAD2 concentration level of the double-stimulation experiment. The bottom bands represent the loading control to which the PSMAD2 levels are normalised.](image)
6.4 Material and Methods

Three major types of data analysis have been used in this project:

- **Model design**
  
  In order to combine the TGF-β and IL-6 signalling pathways we have written ordinary differential equations for the concentration of each component using mass action or Michaelis–Menten kinetic equations [249]. Two different time-delays were considered in the equations for feedback loops in order to compensate for the transcription and translation processes and the connecting reaction of the two pathways.

- **Computer modelling**
  
  The simulation of the mathematical model of the integrated IL-6:TGF-β signalling was performed via MATLAB. The equations are represented by equations 6.1. The MATLAB codes are available on request. The experimental data points are produced using Microsoft Excel 2010 [352].

- **Experimental data**
  
  The experimental data set and the kinetic rates were extracted by the applicant in Walter and Eliza Hall Institute of Medical research, Burgess Lab:

  **Cell culture and cell lysis**

  Mouse embryonic fibroblasts (MEFs) cells were isolated from day 13 to 15 embryos. Wild type MEF cells were cultured in DMEM containing 15% FCS. The cells were trypsinized and washed with DMED + 15% FCS before plating. Passage 3 cells with $1 \times 10^6$ MEFs/well were seeded in 60 mm plates for 0-1 hour, treatment with 1 ng/ml IL-6 (Ref 1661.F33. WEHI). After washing with cold PBS for two times, cells were lysed in ice-cold 200 ul RIPA lysis buffer, containing 1M Tris/HCL, 0.5 M EDTA, 5M NaCl, 10 % Na Doc, 10 % TX-100, 10 % SDS, proteinase inhibitor 100 × and H2O. The cell lysits were passed through 27 G needle for 5 times, then incubated in ice for 20-30 min. After incubation the samples were spun at 13,000 rpm for 30 min at 4°C. The supernatant was transferred to new tubes where 20 ul of samples were saved for BCA protein assay using sigma BCA assay kit (B9643). 20 ul 5x sample buffer was added to 80 ul loading sample and the samples were heated at 95°C for 10 min.

  **Western blotting**

  The protein concentrations were determined using the sigma BCA assay kit (B9643), Novex NuPAGE® 4-12%-Bis-Tris (life technologies NP0335 Box) gels were used for loading proper amount of sample lysate for each time point. PSMAD2 antibody was provided by was used at a dilution 1:500 in 3% BSA-TBS-T. PSMAD2 antibody (rabbit polyclonal anti-phospho-Smad2 antibody (1:1000 for Western blot)) was a gift from Prof. Peter ten
Dijke (Leiden University Medical Center, Netherlands). β-tubulin is used as loading control. Eventually, the gels were transformed onto nitrocellulose membrane via iBlot 2 gel transfer device (Life Technologies) and the membranes were scanned using Odyssey infrared scanner (LI-COR).

**Protein Quantitation**

The western blot images were quantitated using ImageJ 1.49p. The signals of each protein were normalised to the loading control protein where possible.
6.5 Summary and Conclusions

The complexity of cancer and the growing rate of cancer diagnosis make this disease one of the greatest challenges for the health sciences worldwide. Strong correlations between cancer and other diseases of cell regulatory systems, such as chronic inflammation, increase the complexity in developing targeted treatments. Systems biology provides us a holistic approach to study biological systems and their components function-wise. Mathematical modelling, as one of methods defined in Systems biology, allows predictions for biological systems.

Two specific cytokine signalling pathways were studied individually and together, in this project: TGF–β and IL–6 signalling. Separate mathematical models were developed for TGF–β and IL–6 signalling and each system was parameterized and validated experimentally. Chapter 4 describes the model development, model analysis, model validation and model predictions for TGF–β signal transduction. The importance of feedback loops and time-delays in signal regulation are discussed in chapter 4. The contribution of a novel positive feedback loop in TGF–β signalling system and the effects of the feedback coupling in this system are studied in Chapter 4. A systematic reduction method was introduced to simplify the number of components and variables of the TGF–β signalling system without influencing the steady-state and transient response of the system. Eventually, the TGF–β model was validated with a combination of experimental data extracted from the literature and the data produced by the applicant. Developing a rigorous and precise mathematical model for TGF–β signalling equipped us with the predictions for the response of the cells in cancerous conditions (see Chapter 4).

Similar steps were repeated to develop a mathematical model for IL–6 signalling based on Systems biology principles. In Chapter 5 our methodology for designing, developing and simplifying IL–6/Gp130/JAK/STAT3 signalling is presented along with the experimental validation of the model. For the first time, the time-delay is incorporated in the equations of the IL–6 system. Two experimental data sets were used to verify our proposed mathematical model of IL–6 signalling. Chapter 5 also provides model predictions for different stimulation patterns and perturbations. These predictions introduce the idea of the crosstalk between the IL–6 and TGF–β signalling pathway via the negative feedback loop of TGF–β signalling network, which is studied in Chapter 6. In another novel prediction of Chapter 5, it is proposed that pulsatile input of IL–6 to the system can regulate the steady-state level of the output of the system. This method can be used to amplify the IL–6 signalling to eliminate signalling through SOCS3.

In order to examine our initial hypothesis concerning the suppression of TGF–β signalling by IL–6 over-stimulation, we combined our proposed models of TGF–β and IL–6 signalling and designed our integrated IL–6:TGF–β model, in Chapter 6. The previous models were re-evaluated and re-parameterized
to capture all the properties of both signalling networks. For instance, the time-delay incorporated in the feedback terms of the TGF–β signalling equations has been modified in the integrated model. The integrated model was tested against different stimulation patterns and the responses of the models were interpreted according to the network connections. The sensitivity and robustness of the integrated model against the strength of the crosstalk were studied. The simulation results from the integrated model are compatible with our assumptions and interpretation of the down-regulation of TGF–β signalling due to the hyper-activation of IL – 6 JAK/STAT signalling pathway. In the final part of this thesis, the integrated IL – 6:TGF–β signalling model is examined against experimental data from the double-stimulation of wild type MEFs. These results strongly support the biological logistics of our models, our modelling approaches, the simplification and reduction methods and the parameterization of our model.

The results from the integrated model predict that PSMAD2 levels will oscillate in the first two hours due to the IL – 6 stimulation (short-term response), while the steady-state levels (long-term response) depend on the level of the TGF–β stimulation. Specifically, the appearance of the double peak in the dynamics of total PSMAD2 concentration (which has been experimentally proven for the first time) is due to the IL – 6 signalling.

6.6 Overall Conclusions and Future Work

The work described in this thesis has focussed on the development of mathematical models of two signalling pathways (TGF–β and IL–6) that are compatible with the latest logistical analysis of these pathways. Chapter 4 and 5 provide well-designed models of TGF–β and IL – 6 signal transduction with minimum sets of equations which specify the critical reactions in these signalling pathways. In our model of TGF–β signalling, the coupling of the positive and negative feedback loops and its important effects on the regulation of the TGF–β signalling are proposed for the first time. This implies that more consideration should be made when studying systems with feedback loops e.g. signalling systems. Moreover, corporation of time-delays in order to distinguish between the fast and slow reactions is introduced for the first time in our models of TGF–β and IL – 6 signalling. In this thesis the interactions between two distinct regulatory signalling pathways are studied quantitatively for the first time, which paves the road for future studies of the relationship between chronic inflammation and tumor development.

While this thesis proposed a comprehensive study of the crosstalk between two different signalling pathways in line with Systems biology approach, many opportunities for extending the scope of this thesis remain. First, we want to experimentally explore the predictions from our model e.g. the behaviour of
the two signalling pathways with pulsatile input. This requires the development of more experimental protocols. Furthermore, more experimental data is required in order to parameterize the time-delays incorporated in the integrated model. Considering the fact that MAPK signalling is closely related to TGF$-\beta$ and IL$-6$, the immediate future step is to model MAPK signalling and study its influences on the integrated IL$-6$:TGF$-\beta$ model. Inclusion of MAPK signalling in our integrated model would make it possible to study the effects of other signalling receptors such as EGFR family members on TGF$-\beta$ and IL$-6$ signalling. Other modelling approaches (such as stochastic or PDE modelling) could also be studied and compared with the results from our ODE models.
Chapter 7

Appendices

7.1 A Differential Equation Overview

Differential equations are common tools used in mathematical modelling of physical, chemical or biological systems [41, 45, 52, 53, 236, 282]. When the components of a system vary with respect to time, differential equations are used to describe, formulate and solve the system. Examples of differential equations in different areas include mass-spring systems in Physics which follows Hooke’s law or tumor growth dynamics [45].

7.1.1 Initial Conditions and Solutions

Regardless of the type of the differential equation, there might be a set of solutions to a single differential equation. These solutions differ in their starting points. Depending on the integration constant or the arbitrary private solution, the solutions follow different curves. Occasionally, single solution among the family of solution curves is important for our purpose. The solution curves can be identified based on their "initial conditions". Initial condition is written in the form of \( y(t_0) = y_0 \) and the problems which provide initial conditions are called "initial value problems" [41, 236, 282].

7.1.2 Ordinary VS. Partial

Differential equations can be categorized into "ordinary" and "partial" differential equations. Ordinary differential equations, also known as ODEs, are defined when the dependent variable (the variable whose derivative is appeared on the left hand side of the differential equation, e.g. \( u \)) of a system is a function of a single independent variable (e.g. \( t \)) [41, 282].

\[
\begin{align*}
  u &= u(t), \\
  u' &= f(u, t)
\end{align*}
\]
Whereas, the dependent variable in partial differential equations (PDEs) is a function of multiple independent variables (e.g. \(t,s\)) [41,282].

\[
\begin{align*}
    u &= u(t_1, \ldots, t_n), \\
    f(u, t_1, \ldots, t_n, \frac{\partial u}{\partial t_1}, \ldots, \frac{\partial u}{\partial t_n}, \frac{\partial^2 u}{\partial t_1^2}, \ldots, \frac{\partial^2 u}{\partial t_n^2}, \ldots) &= 0
\end{align*}
\]  

### 7.1.3 Order of Differential Equations; Autonomous and Non-autonomous Differential Equations

Both ODEs and PDEs are further classified by the order of derivatives they contain. Order of a differential equation is specified by the number of the highest derivative on the RHS. The general form of a first-order differential equation is illustrated as \(u' = f(t,u)\), where \(f\) is a function of \(t\) and \(u\). The solution of equation 7.2 is defined to be \(u = u(t)\) whose derivative is defined as \(u' = f(t,u)\). If \(f\) does not depend directly on \(t\), then the differential equation is called "autonomous", otherwise, it is a "non-autonomous" differential equation. The autonomous equations are self-governing. This means that the dynamics described by autonomous differential equation will not change by altering the time origin (the starting point of time) [41,53,282].

### 7.1.4 Linearity VS. Non-linearity

Differential equations are classified into "linear" and "non-linear", depending on the power of the variables appearing on the RHS. If the variables appear only with power one, the differential equation is called linear, otherwise it is a non-linear differential equation [41,282].

Pendulum, for instance, is a classic mechanical system which is formulated by a nonlinear differential equation (because \(\sin(\theta)\) is not a first power in 7.5) [41,53,282].

\[
\frac{d^2 \theta}{dt^2} + \sin(\theta) = 0
\]  

### 7.1.5 Systems of Differential Equations

Many problems in science involve multiple differential equations and cannot be described by one differential equation (if they are not simplified and reduced). Depending on the number of variables, different systems of differential equations are used to describe such systems. Similar to single differential equation, systems of differential equations are linear or non-linear. As an example of a system of non-linear, first-order ODE, consider a simple protein dimerization system where two monomers combine to form a dimer:
The non-linear ODE system with two variables is of the form:

\[
M' = p - 2aM^2 + 2bD - mM
\]

\[
D' = aM^2 - bD - dD,
\]

where \(p, m, d\) are production and degradation rates.

There is no certain method or algorithm for solving non-linear systems of differential equations. It becomes more complicated when the equations that describe the system are "coupled" [41, 53, 282]. In coupled system of equations, the right hand sides of the equations are functions of two or more variables (including the variables whose derivative are appeared on the left):

\[
u' = f(t, u, v)
\]

\[
v' = g(t, u, v)
\]

All systems of ordinary differential equations can be written as systems of first-order ODEs. Most of the biological systems can be modeled with a set of first-order ordinary differential equations. That is the reason for using ODEs in order to model biological systems.

Having formulated the desired system into an appropriate model of differential equation, the next step is to find a solution, \(u(t)\), which provides the dynamics of the system. The majority of the natural systems are explained by coupled non-linear systems of differential equations and may not have analytical solutions. Analytic solutions or "closed-form" are said to be found when mathematical manipulations lead to a formula for \(u(t)\). In cases where an analytic solution cannot be found, numerical methods are used. A numerical solution is an approximate solution derived via different algorithms of approximation. Non-linear systems can often be solved through numerical methods. Qualitative methods illustrate important features of the solution without solving the system explicitly and are useful in sketching a graphical representation of the system [41, 53, 282].

Having the system solved, analytically or numerically, a number of "equilibrium points" must be found. Physical systems arising from physical phenomena often result in a number of equilibrium points. Consequent stability analysis illuminates whether the equilibrium points are stable and can be explained by a physical explanation.
7.1.6 Qualitative Analysis; Fixed Points, Equilibrium Points and Stability

In order to gain an understanding of the solution of first-order differential equations, we can sketch the "direction field" or "slope field". In direction field, the slope of each small vector (slope mark) is in the direction of \( f(t, u) \) for all points in \( t, u \)-plane. A line in the direction field which is obtained by setting \( f \) equal to a constant is called "isocline" and the lines along which the slope field is zero are "nullclines" [41, 282]. As an example, consider the system of first-order ODEs:

\[
\begin{align*}
x'_1 &= f_1(x_1, ..., x_n) \\
x'_2 &= f_2(x_1, ..., x_n) \\
&\vdots \\
x'_n &= f_n(x_1, ..., x_n)
\end{align*}
\]

The set of \( x_j \)s \((j = 1, ..., n)\) that satisfies \( f_j(x_1, ..., x_n) = 0 \) is the nullcline set of the system. The intersection where all nullclines meet is the "fixed point" or the "equilibrium point". The origin \((0, 0) \) is a fixed point of all linear systems of ordinary differential equation, while a non-linear system may have neither fixed points nor nullclines. In general, a fixed point is a state of the system where the variables are constant with time. This state is called "steady-state" [41, 282].

Stability of a system is defined based on the perturbations applied to the system near its fixed or equilibrium points. If the system converges to its equilibrium point after applying perturbation, the equilibrium is "stable". If the system deviates from its equilibrium point rapidly after applying perturbations, the equilibrium is called "unstable".

Studying the direction field of an equation is important in problems that the changes in the rate of the variables are crucial (such as equations describing cytokine signalling pathways where the system can be reduced based on the reaction rates). However, it is hard to visualize the slope field for systems with more than three time-varying variables. Therefore, this feature is not applicable to most of the biological system studies.

7.1.7 Linearization of Non-linear ODEs

As mentioned previously, most of the real world systems are non-linear and hard to solve analytically. However, if a non-linear system has an equilibrium, its stability can be assessed by studying the stability of an approximated linear system near the equilibrium point. Linear systems are easier to solve. Their behaviours are determined by their associated matrix of eigenvalues and eigenvectors. The general idea of linearization is to approximate a non-linear system by the relative linear system in a neighborhood of the equilibria and use the linear system’s features to find a proper approximated solution [41, 53, 236, 282].
Linearization starts by defining small perturbations around the equilibrium point for each variable of the system. As an example, consider the system below as the initial non-linear system:

\[
\begin{align*}
    x' &= f(x, y) \\
    y' &= g(x, y)
\end{align*}
\]

If \( x^* = (x_e, y_e) \) is an isolated equilibrium, \( u = x - x_e \) and \( v = y - y_e \) will be defined as the small perturbations around \( x^* \). Therefore, the equation can be rewritten to the form:

\[
\begin{align*}
    u' &= f(x_e + u, y_e + v) \\
    v' &= g(x_e + u, y_e + v)
\end{align*}
\]

Now, we can use the "Taylor series" to expand the right side of the equations.

\[
\begin{align*}
    u' &= f(x_e, y_e) + f_x(x_e, y_e)u + f_y(x_e, y_e)v \\
    v' &= g(x_e, y_e) + g_x(x_e, y_e)u + g_y(x_e, y_e)v
\end{align*}
\]

The first term in each equation is zero. We can represent the above equations in matrix form:

\[
\begin{pmatrix}
    u' \\
    v'
\end{pmatrix} =
\begin{pmatrix}
    f_x(x_e, y_e) & f_y(x_e, y_e) \\
    g_x(x_e, y_e) & g_y(x_e, y_e)
\end{pmatrix}
\begin{pmatrix}
    u \\
    v
\end{pmatrix}
\]

The matrix that maps \( \begin{pmatrix} u \\ v \end{pmatrix} \) to \( \begin{pmatrix} u' \\ v' \end{pmatrix} \) is called "Jacobian" matrix and contains the first partial derivatives of \( f \) and \( g \) with respect to \( x \) and \( y \). The eigenvalues of Jacobian matrix determine the stability status of the original non-linear system around the equilibrium point (only if the eigenvalues have nonzero real part) \([41,52,53,236,282]\).

### 7.2 ODE Solving Methods; Numerical Consideration

Solution methods of ordinary differential equations are controlled either by "initial conditions" or "boundary conditions". The distinction between these two conditions is in the number of points by which the subsidiary conditions are given. In initial value problems, information about a certain point (normally at \( t = 0 \)) guarantees the uniqueness of the solution, while this information is spread among a number of points (boundaries) in boundary value problems.
7.2 ODE Solving Methods; Numerical Consideration

7.2.1 Analytical Solutions

Most of linear systems are solvable via analytical methods. The usual strategy to solve linear differential equations is to reduce their order by various algorithms and solve the reduced equation using different methods, depending on the equation type [41, 52, 53, 236, 282]. Two of the most common and general methods to deal with linear systems are: 1. Fourier transform and 2. Laplace transform [41, 282].

Fourier transform enables us to solve linear boundary value problems which might have solutions in the whole real line domain. Notwithstanding, Laplace transform is able to solve linear initial value problems with the help of Fourier transform [41, 282].

As mentioned before, in order to solve a nonlinear system of differential equation an optimal linear system should be found in the first place. This linear system of equations can be solved by one of the analytical methods later. There are a few methods to approximate a nonlinear system near a point of interest. Following subsections provide brief summary on the most applicable approximation methods used to solve nonlinear systems [41, 52, 53, 236, 282].

Iterative Methods

Picard iteration (E. Picard, 1856-1941) is the simplest approximation method for solving nonlinear differential equation by a recursive analytic process. This method of approximation is founded on the basis of a classical method in mathematics called "fixed point iteration" [41, 52, 282]. Fixed point iteration finds solutions for general algebraic equations of the form \( x = g(x) \). Approximated solution \( x^* \) is generated through the iterative procedure,

\[
x_{k+1} = g(x_k), \quad k = 0, 1, 2, ...
\]

(7.6)

, provided \( |g'(x^*)| < 1 \) and the initial guess of the solution is close enough [41, 52, 282].

Picard iteration method is mostly applied to initial value problem, adopting the fixed pint iteration idea. Consider initial conditions as the following:

\[
u' = f(t, u)
\]

(7.7)

\[
u(t_0) = u_0
\]

(7.8)

We integrate both sides of the first equation to derive an equation for the solution.

\[
u(t) = u_0 + \int_{t_0}^{t} f(s, u(s))ds
\]

(7.9)

Now Picard iteration is defined based on the fixed point iteration.
7.2 ODE Solving Methods; Numerical Consideration

\[ u_{k+1}(t) = u_0 + \int_{t_0}^{t} f(s, u_k(s)) ds, \quad k = 0, 1, 2, \ldots \]  \hspace{1cm} (7.10)

In this equation \( u_0(t) \) is the initial approximation and is usually set to be the constant function \( u_0(t) = u_0 \). By applying this method, a sequence of \( u_k(t) \) is generated which converges to the absolute solution.

Technically, any solution of \( x = g(x) \) is the solution of \( f(x) = 0 \) and vice-versa. Therefore, this set of solution satisfies the below equation.

\[ x = g(x) = x - \frac{f(x)}{h(x)} \]  \hspace{1cm} (7.11)

\( h(x) \) is a non-zero arbitrary function. By setting \( h(x) \) equals to \( f'(x) \), the iterative equation is defined as:

\[ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \ldots \]  \hspace{1cm} (7.12)

Now, using fixed point iteration method, \( x_n \) will converge to the real solution, only if the initial guess is close enough to it. In truth, replacing \( h(x) \) with the first derivative of \( f \) has made the convergence as rapid as possible. In other words, the equation, known as "Newton method", is the optimal state of the equation.

Picard iteration is a functional method for acquiring the solution, however, the solution is guaranteed only if the right hand side function is regular. In any case, Using Picard iteration, the solution is still found locally. Therefore, other numerical methods are put into operation to solve initial value problem by means of computer algorithm [41, 52, 282].

7.2.2 Numerical Solutions

Up to this point, the differential equation types and some solution methods have been reviewed. Though, there are many important problems in engineering and science that are potentially unsolvable with analytical methods and it is impossible to derive from them an exact expression for the solution. As a result, alternative methods have been created to provide an accurate approximation to the solution of an initial value problem. These alternative methods are called "numerical methods" [41,52,53,236].

- Taylor Series Approach

Taylor series are the foundation of most of the numerical approximation methods since they define the function \( f \) in terms of its measured derivatives in a single point. The Taylor series of \( f(x) \) in \( x = a \) are of the form:

\[ f(a) + \frac{1}{2!} f'(a)(x - a) + \frac{1}{3!} f''(a)(x - a)^2 + \ldots = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n \]  \hspace{1cm} (7.13)
Using the first two terms of Taylor series to extend $u'(t)$ in $u'(t) = f(t, u)$ is the basis of "Euler method" for solving nonlinear differential equations. In order to increase the order of cumulative error and obtain a more precise approximation of the solution, the most sensible way is to extend the Euler formula to the third term of the Taylor series. The result is the "three term Taylor series method":

$$u_{n+1} = u_n + hf(t_n, u_n) + \frac{h^2}{2} \left[ \frac{\partial f}{\partial t} + f \frac{\partial f}{\partial u} \right] (t_n, u_n), \quad n = 0, ..., N \quad (7.14)$$

Apparently, the error of the three term Taylor series method is of the order $h^2$ which, in comparison with Euler method, makes a great difference in the approximated solution if $h$ is small.

- Explicit VS. Implicit

In numerical methods approximations, whenever a point such as $u_{n+1}$ can be approximated in a way that it only depends on the previous point approximations, the equation is solved "explicitly" and the approach is an "explicit method". Otherwise, the approximated measurement at each point depends on both itself and previous points and the approximation is acquired through an "implicit method". Implicit method such as, Newton’s method, requires solving (nonlinear) equations to find $u_{n+1}$ in terms of $u_n$. First-order implicit methods provide solutions which are "absolutely stable" while, the solutions derived via explicit methods are "conditionally stable".

As an example of implicit and explicit methods, consider "Euler method" formula (for more details on this method refer to next section):

$$u_{n+1} = u_n + hf(t_n, u_n) \quad (7.15)$$

This is in fact the explicit expression of Euler method, however, this method has implicit version as well:

$$u_{n+1} - hf(t_{n+1}, u_{n+1}) = u_n \quad (7.16)$$

Choosing between implicit and explicit numerical methods is always a trade-off. As mentioned before, the solutions via implicit methods are absolutely stable. Thus, it is sensible if one chooses the implicit instead of the explicit Euler method. However, applying implicit methods on nonlinear and complex systems often leads to nonlinear equations to be solved. Numerical methods used to solve nonlinear equations result in approximated solutions. Yet, Obtaining approximated solutions means that we have already lost the stability of our solution.
All in one, implicit methods do not induce higher stability assigning to complex nonlinear systems. Therefore, in dealing with the majority of the nonlinear models in biology it is more preferred to use explicit numerical methods, such as explicit, forth-order "Runge-Kutta".

- Error Analysis of Numerical Methods

Solving IVP (initial value problem) numerically raises a number of other properties to the solutions that need to be studied. One of these properties is the "convergence" of the solution. Convergence means how close the solution sequence \((u_1, u_2, ..., u_n)\) is to the actual solution or, how small the step size should be to give the maximum solution accuracy.

In order to quantify the convergence property, two classes of errors have been defined. The difference between the solution obtained numerically and the actual solution \(u^*_n\) is known as "global truncation error", \(E_n\).

\[
E_n = u^*_n - u_n
\]  
(7.17)

This error is caused by both the approximation formula to determine \(u_{n+1}\) used in each step and the approximated input to the approximation formula at the following step. Now, if we assume that the input in the last step is accurate, the only error is due to the approximation formula and is called "local truncated error", \(e_n\).

- Euler Method

Euler method, named after L. Euler (1707-1783), has got the simplest algorithm among numerical approximation methods. The basic idea of this method is to replace the continuous time model with an approximate discrete time model. To achieve this, the time span is divided into smaller intervals of the size \(h\).

\[
0 = t_0, t_1, t_2, ..., t_n = T \quad t_n = nh \quad n = 0, 1, 2, ..., N
\]  
(7.18)

If we use the fundamental theorem of calculus, we can write:

\[
u(t_{n+1}) - u(t_n) = \int_{t_n}^{t_{n+1}} f(t, u)dt.
\]  
(7.19)

This can be approximated by the left-hand rule as the following:

\[
u_{n+1} = u_n + hf(t_n, u_n)
\]  
(7.20)

In the last step, \(u(t_n)\) is substituted by \(u_n\) for the sake of presentation. Equation is the main Euler formula. Solution at time \(t\) is calculated based on solution at time \(t - h\) and the slope of the solution curve at the previous
approximated point. The cumulative error is bounded by $h$ over the time interval. As a result, the Euler method is of the order $h$.

Going back to our dimerization example in Section 3.3.1, we want to solve the system with Euler method $u_{n+1} = u_n + h f(t_n, u_n)$, up to two steps:

\begin{align*}
M' &= p - 2a.M^2 + 2b.D - m.M \\
D' &= a.M^2 - b.D - d.D,
\end{align*}

(7.21)
(7.22)

where $p$, $m$ and $d$ represent production rate of $M$, degradation rate of $M$ and degradation rate of $D$, respectively. First, we should set the initial conditions for $M$ and $D$. We define $M(t_0) = M_0$ and $D(t_0) = D_0$. Now, in the first step when $n = 0$ Euler equation provides:

\begin{align*}
M_1 &= M_0 + h(p - 2a.M_0^2 + 2b.D_0 - m.M_0) \\
&= (2ha)M_0^2 + (1 - mh)M_0 + (hp + 2hbD_0) \\
D_1 &= D_0 + h(a.M_0^2 - b.D_0 - d.D_0) = (ha)M_0^2 + (1 - hb - hd)D_0
\end{align*}

(7.23)

Having derived $M_1$ and $D_1$, we can obtain $M_2$ and $D_2$ successively, setting $n = 1$.

\begin{align*}
M_2 &= M_1 + h(p - 2a.M_1^2 + 2b.D_1 - m.M_1) \\
D_2 &= D_1 + h(a.M_1^2 - b.D_1 - d.D_1)
\end{align*}

(7.24)
(7.25)

Substituting $M_1$ and $D_1$ from the previous step, we can calculate the state of the system in the second step of Euler method. The consequent steps are calculated similarly. In order to avoid complex calculation, it is better to set values for the constants and the initial conditions. Here, I have used the real values for R2 dimerization in our model:

\begin{align*}
M_0 &= 8.87, \quad D_0 = 0, \quad p = 0.019, \quad m = 0.0132, \quad a = 0.01, \quad b = 0.001, \quad d = 11.8
\end{align*}

(7.26)

Substituting $h = 0.5$ in equations, we will have:

\begin{align*}
M_1 &= 8.03 \\
D_1 &= 0.39
\end{align*}

(7.27)
(7.28)

Consequently, the solutions for $n = 1$ are as the following:
7.2 ODE Solving Methods; Numerical Consideration

\[ M_2 = 7.34 \quad (7.29) \]
\[ D_2 = -3.57 \quad (7.30) \]

Similarly, next step \((n = 2)\) provides subsequent solutions:

\[ M_3 = M_2 + h(p - 2aM_2^2 + 2bD_2 - mM_2) = 6.75 \quad (7.31) \]
\[ D_3 = D_2 + h(aM_2^2 - bD_2 - dD_2) = 17.76 \quad (7.32) \]

Steps can be continued as much as it is required. Though, taking more steps and more calculations do not increase the accuracy of the Euler method; they just provide a better approximated solution curve.

- **Modified Euler Method**

  The Euler method can be more accurate, using the "trapezoidal rule".

  \[ u(t_{n+1}) - u(t_n) = \frac{h}{2} \left[ f(t_n, u_n) + f(t_{n+1}, u_{n+1}) \right]. \quad (7.33) \]

  This equation is called implicit Euler equation and its cumulative error is of the order \(h^2\). At each step of solving this equation, a nonlinear equation must be solved for \(u_{n+1}\). To avoid such a complexity, \(u_{n+1}\) at each step is replaced by \(u_{n+1}\) derived from simple Euler method. Thus, the predictor can be calculated as:

  \[ \tilde{u}_{n+1} = u_n + hf(t_n, u_n) \quad (7.34) \]

  and the corrector is in the form of:

  \[ u(t_{n+1}) - u(t_n) = \frac{h}{2} \left[ f(t_n, u_n) + f(t_{n+1}, \tilde{u}_{n+1}) \right] \quad (7.35) \]

  This predictor-corrector algorithm is known as modified Euler method, Huen’s method or second-order Runge-Kutta method. The cumulative error of modified Euler method is also of the order of \(h^2\).

- **Runge-Kutta Method (A Class of Methods)**

  The extension of Euler method is possible either by contribution of higher order of derivatives calculated or by multiplicity of evaluation of the function \(f\), at each step. The idea of the latter was first proposed by Runge in 1895. In collaboration with Heun (1900) and Kutta (1901) this method was developed to forth order and a few efforts for the fifth order was made.
Huta (1956, 1957) introduced the sixth-order of Runge-Kutta method. Recent studies are focussed on implicit Runge-Kutta methods to find numerical solutions for stiff differential equations (solutions derived through numerical methods are numerically unstable in stiff equations). This method is simple and accurate at the same time and thereupon, is the most widely used method in science and engineering.

The Runge-Kutta method is indeed a class of various methods with different calculation steps and different assumptions. They also can be explicit and implicit. The simple idea is that the average values of \( f \) accept different weights in the process of finding the next approximating point. Here, we only focus on the low-order explicit methods of standard Runge-Kutta. In general, the two-stage Runge-Kutta method can be formulated as below:

\[
    u_{n+1} = u_n + h[b_1 f(u_n) + b_2 f(u_n + h a_{21} f(u_n))] \quad (7.36)
\]

where \( t_{n+1} = t_n + h \). The constants \( b_1, b_2 \) and \( a_{21} \) should be chosen in a way that the order of the method is as high as possible. The optimal state for the first and second orders are:

\[
    First - Order : b_1 + b_2 = 1 \quad (7.37)
\]

\[
    Second - order : b_1 + b_2 = 1, \quad b_2 a_{21} = \frac{1}{2} \quad (7.38)
\]

at the same time.

Therefore, assuming \( U(t) \) as the approximated value of \( u(t) \), the method is illustrated as:

\[
    U_i(x) = u(t_0) + (t - t_0) \sum_{j=1}^{s} a_{ij} f(U_j(t)) \quad i = 1, 2, ..., s, ... \quad (7.39)
\]

\[
    U(t) = u(t_0) + (t - t_0) \sum_{j=1}^{s} b_j f(U_j(t)) \quad (7.40)
\]

The above weights are identified by the Butcher Tableau:

\[
\begin{array}{c|cccc}
    c_1 & a_{11} & a_{12} & ... & a_{1s} \\
    c_2 & a_{21} & a_{22} & ... & a_{2s} \\
    \vdots & \vdots & \vdots & \ddots & \vdots & \\
    c_s & a_{s1} & a_{s2} & ... & a_{ss} \\
    \hline
    & b_1 & b_2 & ... & b_s
\end{array} \quad (7.41)
\]
The numbers $c_i$s are defined as $c_i = \sum_{j=1}^{s} a_{ij}$ for $i = 1, 2, ..., s$. These weights can be shown in a tree graph diagram as well.

**Methods With Orders Less Than Four:**

The Butcher Tableau for order 2 with two stages is provided below:

$$
\begin{array}{ccc}
0 & | & a_{21} \\
\text{ } & | & \text{ } \\
\text{ } & | & b_1 \ b_2 \\
\end{array}
$$

(7.42)

Putting $b_1 + b_2 = 1$ and $b_2c_2 = \frac{1}{2}$, $c_2$ will be equal to $a_{21}$. Two possible cases of the tableaus meeting these conditions are:

$$
\begin{array}{ccc}
0 & | & \frac{1}{2} \\
\text{ } & | & \frac{1}{2} \\
\text{ } & | & 0 \ 1 \\
\end{array}
$$

(7.43)

and

$$
\begin{array}{ccc}
0 & | & 1 \\
\text{ } & | & 1 \\
\text{ } & | & \frac{1}{2} \ \frac{1}{2} \\
\end{array}
$$

(7.44)

Reviewing previous sections, these represent the same modified Euler method.

In order to move to third-order Runge-Kutta, three stages are essential to fill up the tableau:

$$
\begin{array}{ccc}
0 & | & a_{21} \\
\text{ } & | & \text{ } \\
\text{ } & | & a_{31} \ a_{32} \\
\end{array}
$$

(7.45)

The conditions that rise immediately are:

$$
b_1 + b_2 + b_3 = 1, \quad b_2c_2 + b_3c_3 = \frac{1}{2}, \quad b_2c_2^2 + b_3c_3^2 = \frac{1}{3}, \quad b_3a_{32}c_2 = \frac{1}{6} \quad (7.46)
$$

If we solve all these equations simultaneously, three different cases would be created, depending on the arbitrary choices of the individual constants.

The classical forth-order Runge-Kutta can be written via Butcher Tableau as below. This is the most well-known version of Runge-Kutta method. This is remarkably because $a_{31}$, $a_{41}$ and $a_{42}$ are put to zero.
The equations for the classical fourth-order Runge-Kutta are formulated as below:

\[ u_{n+1} = u_n + \frac{h}{6} [k_{n,1} + 2k_{n,2} + 2k_{n,3} + k_{n,4}], \quad n = 0, 1, 2, ..., N \]  

\[ \begin{align*}
  k_{n,1} &= f(t_n, u_n) \\
  k_{n,2} &= f(t_n + \frac{1}{2}h, u_n + \frac{1}{2}hk_{n,1}) \\
  k_{n,3} &= f(t_n + \frac{1}{2}h, u_n + \frac{1}{2}hk_{n,2}) \\
  k_{n,4} &= f(t_n + h, u_n + hk_{n,3})
\end{align*} \]  

Four functional evaluations are performed at each step in this method while, in Euler method we evaluate only one function at a time.

The \( \frac{h}{6} [k_{n,1} + 2k_{n,2} + 2k_{n,3} + k_{n,4}] \) is interpreted as the average slope, calculated in 4 different points. The error in fourth-order Runge-kutta method is of order \( h^4 \), which represents the degree of accuracy of this method compared to previous numerical methods.

It is time to return to the dimerization example and see how it is solved via fourth-order classical Runge-Kutta.

\[ M' = p - 2a.M^2 + 2b.D - m.M = f(t, M, D) \]  

\[ D' = a.M^2 - b.D - d.D = g(t, M, D) \]  

In order to simplify the calculation, we give values to equation parameters and initial conditions. Here, I have used the same values for R2 dimerization used in Euler method section:

\[ M_0 = 8.87, \quad D_0 = 0, \quad p = 0.019, \quad m = 0.0132, \quad a = 0.01, \quad b = 0.001, \quad d = 11.8 \]  

Putting \( n = 0 \) and \( h = 0.5 \), \( k_{0,i} \quad i = 1, 2, 3, 4 \) are calculated as the following:
$k_{0,1} = f(t_0, M_0, D_0)$

$= p - 2a.M_0^2 + 2b.D_0 - m.M_0 = -1.67$

$l_{0,1} = g(t_0, M_0, D_0)$

$= a.M_0^2 - b.D_0 - d.D_0 = 0.79$

$k_{0,2} = f(t_0 + \frac{1}{2}h, M_0 + \frac{1}{2}hk_{0,1}, D_0 + \frac{1}{2}hl_{0,1})$

$= p - 2a.[M_0 + \frac{1}{2}hk_{0,1}]^2 + 2b.[D_0 + \frac{1}{2}hl_{0,1}] - m.[M_0 + \frac{1}{2}hk_{0,1}] = -1.52$

$l_{0,2} = g(t_0 + \frac{1}{2}h, M_0 + \frac{1}{2}hk_{0,1}, D_0 + \frac{1}{2}hl_{0,1})$

$= a.[M_0 + \frac{1}{2}hk_{0,1}]^2 - b.[D_0 + \frac{1}{2}hl_{0,1}] - d.[D_0 + \frac{1}{2}hl_{0,1}] = -1.65$

$k_{0,3} = f(t_0 + \frac{1}{2}h, M_0 + \frac{1}{2}hk_{0,2}, D_0 + \frac{1}{2}hl_{0,2})$

$= p - 2a.[M_0 + \frac{1}{2}hk_{0,2}]^2 + 2b.[D_0 + \frac{1}{2}hl_{0,2}] - m.[M_0 + \frac{1}{2}hk_{0,2}] = -1.53$

$l_{0,3} = g(t_0 + \frac{1}{2}h, M_0 + \frac{1}{2}hk_{0,2}, D_0 + \frac{1}{2}hl_{0,2})$

$= a.[M_0 + \frac{1}{2}hk_{0,2}]^2 - b.[D_0 + \frac{1}{2}hl_{0,2}] - d.[D_0 + \frac{1}{2}hl_{0,2}] = 5.59$

$k_{0,4} = f(t_0 + h, M_0 + hk_{0,3}, D_0 + hl_{0,3})$

$= p - 2a.[M_0 + hk_{0,3}]^2 + 2b.[D_0 + hl_{0,3}] - m.[M_0 + hk_{0,3}] = -1.41$

$l_{0,4} = g(t_0 + h, M_0 + hk_{0,3}, D_0 + hl_{0,3})$

$= a.[M_0 + hk_{0,3}]^2 - b.[D_0 + hl_{0,3}] - d.[D_0 + hl_{0,3}] = -32.33$

(7.53)

$M_1$ and $D_1$ can be derived from $M_0$ and $D_0$ with the help of $k_{0,i}$s and $l_{0,i}$s:

$M_1 = M_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) = 7.31$  \hspace{1cm} (7.54)

$D_1 = D_0 + \frac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4) = 0.29$  \hspace{1cm} (7.55)

The same procedure should be followed for the successive steps ($n = 1, 2, ...$). Each $n$ represents single point in the slope plane. The curve that connects all the points gives the solution of the system with mentioned initial conditions and assumptions.

### 7.3 Review of the Mathematical Models on TGF–β Signalling

- A Rate Equation Approach to Elucidate the Kinetics and Robustness of the TGF–β Pathway, Melke et al. 2006 [308]
A minimalist model of the TGF−β family signal transduction is provided in this paper. The aim is to study the aspects of TGF−β signalling which have not been studied in detail previously, to identify the key components and modules (specifically SMAD7) in the pathway and gain the ability to predict the quantities which have not been measured yet. The signalling through ALK5 (TGF−β R1) and ALK1 (BMP signalling pathway receptor) are compared, while the effects of inhibitory feedback of SMAD7 on both signalling pathways have been modeled. Compared to the Vilar et al. [449] model, the mathematical modelling equations (17 ODE equations) of Melk et al. [308] are simple and are unable to simulate the topology of TGF−β signal transduction. However, as the first attempt to model a complex signalling pathway such as TGF−β, Melke et al. [308] succeeded to introduce a robust model and discovered that SMAD7 presence as a negative feedback for the pathway is essential to maintain this robustness. No experimental data is provided by Melke et al. [308].

- Signal Processing in the TGF−β Superfamily Ligand-Receptor Network, Vilar et al. 2006 [449]

In this paper Vilar et al. [449] mathematically model the TGF−β superfamily receptor trafficking. This model is based on 6 ODEs. In order to justify the dual-responding of cancer cells to TGF−β stimulation, it has been suggested that there are two different ways for TGF−β receptors to be internalized in response to ligand binding: receptor internalization through a lipid-raft-caveolar degradation pathway [449] in which there is no signalling initiated from the internalized receptor, and the standard internalization clathrin pathway which leads to the normal signalling degradation of the receptors. Vilar et al [449] conclude that the imbalances between the rates of the receptor internalization pathways are the key component responsible for the duality in the role of TGF−β stimulation.
7.3 Review of the Mathematical Models on TGF–β Signalling

<table>
<thead>
<tr>
<th>Parameters derived by Vilar et al. [449]</th>
<th>Parameters from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptor internalization rate (kint = 0.334 min⁻¹)</td>
<td>Experimental data sets</td>
<td>[200]</td>
</tr>
<tr>
<td>Ligand-induced receptor degradation rate (klid = 0.25 min⁻¹)</td>
<td>[313]</td>
<td></td>
</tr>
<tr>
<td>Constitutive degradation rate (kcd = 0.028 min⁻¹)</td>
<td>[93]</td>
<td></td>
</tr>
<tr>
<td>Receptor recycling rate (krec = 0.034 min⁻¹)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligand receptor complex association rate (ka = 1; ligand concentration = 0.01 nM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steady and quasi-steady-state analysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Systems theory of SMAD signalling, Clarke et al. 2009 [74]

In contrast to the Vilar et al. [449] model, the Clarke et al. TGF–β signalling model focuses on intracellular interactions rather than the receptor trafficking on the cell surface. A total of 9 ODEs are used to parameterise the TGF–β signalling system. SMAD7 inhibitory effects are not included in this model. Clark et al. [74] conclude that the system is more sensitive to the parameters regulating the R-SMAD phosphorylation and dephosphorylation processes. The importance of the role of SMAD4 in TGF–β signal transduction is pointed in this paper. It is postulated that the imbalance between the phosphorylation and dephosphorylation rates of the R-SMADs (see Chapter 2) is the critical determinant of SMAD2, 3 and 4 nuclear accumulation.
### Parameters derived by Clarke et al. [73]

<table>
<thead>
<tr>
<th>Parameters from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of receptors: 1000</td>
<td>[351]</td>
</tr>
<tr>
<td>Number of SMAD2/3 in the cytoplasm: 1.62e+5</td>
<td>[456]</td>
</tr>
<tr>
<td>Number of SMAD2/3 in nucleus: 1.8e+4</td>
<td></td>
</tr>
<tr>
<td>Number of SMAD4 in the cytoplasm: 1.2e+5</td>
<td></td>
</tr>
<tr>
<td>Number of SMAD4 in the nucleus: 3e+4</td>
<td></td>
</tr>
</tbody>
</table>

- Constraint-Based Modelling and Kinetic Analysis of the SMAD Dependent TGF–β Signalling Pathway, Zi et al 2007 [513]

In this model, both receptor trafficking and SMAD nucleocytoplasmic shuttling are included explicitly. The aim is to study the dynamic responses to TGF–β signalling via system biology approach. Steady-state analysis has been performed prior to a dynamic study of the signalling components. However, the model appears to be over-fitted. Since all of the experimental data is used for parameter estimation, no predictive analysis is compared with experimental data. This model uses one of the most comprehensive topographies for modelling the TGF–β signalling pathway (it is based on 16 ODEs) and provides access to experimental data for initial data parameterization and the dynamics of the signalling components. Zi et al. [513] develop a constraint-based modelling approach and conclude that the most effective regulator of TGF–β signalling is the balance between clatherin dependent endocytosis and non-clatherin mediated endocytosis.
<table>
<thead>
<tr>
<th>Parameters derived by Zi et al. [513]</th>
<th>Parameters from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I receptor at cell surface: 0.237 nM</td>
<td>The volume of the epithelial cells</td>
<td>[93]</td>
</tr>
<tr>
<td>Type I receptor in caveolar lipid-raft: 2.092 nM</td>
<td>Internalization, degradation and recycling rates of the receptor</td>
<td>[93, 449]</td>
</tr>
<tr>
<td>Type I receptor in early endosome: 2.06 nM</td>
<td>SMAD nucleocytoplasmic rates</td>
<td>[384]</td>
</tr>
<tr>
<td>Type II receptor at cell surface: 0.202 nM</td>
<td>PSMAD nuclear import rate</td>
<td>[486]</td>
</tr>
<tr>
<td>Type II receptor in caveolar lipid-raft: 1.778 nM</td>
<td>Total amount of SMADs</td>
<td>[173]</td>
</tr>
<tr>
<td>Type II receptor in early endosome: 1.148 nM</td>
<td>Total receptor number</td>
<td>[462]</td>
</tr>
<tr>
<td>SMAD2/3 in the cytoplasm: 492.61 nM</td>
<td>Experimental data sets (time course of stimulated PSMAD2, nuclear PSMAD2)</td>
<td>[21, 151, 200, 274]</td>
</tr>
<tr>
<td>SMAD2/3 in the nucleus: 236.45 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD4 in the cytoplasm: 1149.4 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD4 in the nucleus: 551.72 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V(HaCaT): 1.4e-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vcyt: 1.05e-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vnuc: 3.5e-13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Mathematical modelling identifies SMAD nucleocytoplasmic shuttling as a dynamic signal-interpreting system, Schmierer et al. 2008 [386]
A more developed mathematical model of TGF–β signal transduction is provided by Schmierer et al. [386]. They built a constrained mathematical model to quantify the components of TGF–β signalling pathway and predict the behaviour of a cell with an impaired SMAD2 mutant [386]. Having considered the receptor trafficking and SMAD nucleocytoplasmic shuttling, their model also includes the mechanisms of transcriptional activation by the SMAD nuclear complex, in addition to feedback control from the nucleus. However, the feedback control they use originates from an artificial small molecule (SB-431542) rather from the inherent feedback of the pathway, I-SMADs (see Chapter 2). By introducing this model, Schmierer et al. [386] derived quantitative kinetic rates and other components in TGF–β signalling system to which many subsequent modelling papers refer (such as [511]).
### Parameters derived by Schmierer et al. [386]

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Parameter from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration of receptors</td>
<td>Concentration of SMAD2 and SMAD3</td>
<td>[73]</td>
</tr>
<tr>
<td>Nucleocytoplasmic volume ratio (using experimental data)</td>
<td></td>
<td>[384]</td>
</tr>
<tr>
<td>All the kinetic rates (from optimization)</td>
<td></td>
<td>[200]</td>
</tr>
<tr>
<td>Total PSMAD2</td>
<td></td>
<td>[5]</td>
</tr>
<tr>
<td>SMAD2/3 in the cytoplasm: 82800 molecules/cell = 60.6 nM</td>
<td></td>
<td>[173]</td>
</tr>
<tr>
<td>SMAD2/3 in the nucleus: 17200 molecules/cell = 28.5 nM</td>
<td></td>
<td>[462]</td>
</tr>
<tr>
<td>SMAD4 in the cytoplasm: 69400 = 50.8 nM</td>
<td></td>
<td>[21, 151, 200, 274]</td>
</tr>
<tr>
<td>SMAD4 in the nucleus: 30600 molecules/cell = 50.8 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptors in the cytoplasm: 1400 molecules/cell = 1 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphatase in the nucleus: 600 molecules/cell = 1 nM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Quantitative Modelling and Analysis of the Transforming Growth Factor Signalling Pathway, Chung et al. 2009 [72]**

Chung et al. [72] aim to discover the changes in intracellular processes in cancer cells during exposure of TGF–β. Their final goal is to explain the duality of TGF–β stimulation (see Chapter 2 and 4). The model has deficiencies when considering the feedback role of SMAD7; however, it is one of the most complete models that study the TGF–β signalling pathway.
with 17 ODEs. A variety of different databases are used for their parameter estimations. The latter causes disruptions in the predictive nature of the model, since the experimental conditions for the data sets were different. Finally, they propose a new hypothesis to justify the paradoxical effects of TGF–β treatment for cancer cells.

<table>
<thead>
<tr>
<th>Parameters derived by Chung et al. [72]</th>
<th>Parameters from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model analysis and simulation</td>
<td>Experimental data sets (total nuclear PSMAD2 for short and long time stimulation of TGF–β, total cytoplasmic PSMAD2, total nuclear and cytoplasmic SMAD2 and total SMAD4)</td>
<td>[200,351]</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>Model validation data set</td>
<td>[274,279,351]</td>
</tr>
<tr>
<td>The comparison between normal and cancerous cells in the level of nuclear PSMAD2-SMAD4</td>
<td>Experimental data set (cancerous cell line)</td>
<td>[267]</td>
</tr>
<tr>
<td>A set of experimental data of PSMAD2 level in prostate cancer cell line</td>
<td>Most of the kinetic rates</td>
<td>[130, 148, 274, 279, 347, 384, 388, 463]</td>
</tr>
<tr>
<td>kinetic rates</td>
<td>10000 TGF–β receptors</td>
<td>[462]</td>
</tr>
<tr>
<td></td>
<td>Total SMAD2 and SMAD4: 100000 each</td>
<td>[73,384]</td>
</tr>
<tr>
<td></td>
<td>SMAD2 in the cytoplasm: 85000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMAD2 in the nucleus: 15000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMAD4 in the cytoplasm: 87000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMAD4 in the nucleus: 13000</td>
<td></td>
</tr>
</tbody>
</table>
• Quantitative analysis of transient and sustained transforming growth factor-$\beta$ (TGF-$\beta$) signalling dynamics, Zi et al. 2011 [511]

Zi et al. [511] have provided the most recent mathematical model of the TGF-$\beta$ signalling. This model uses switch-like responses as outputs of the signalling system. Their model has 19 ODEs and its main goal is to specify the dose-dependency and time-dependency of the TGF-$\beta$ signalling responses. The authors believe long-term switch-like responses to TGF-$\beta$ stimulation determine the cellular behaviours. It is because of the different sensitivities the pathway displays in response to different ligand doses and ligand stimulation time-scales. However, the assumptions used to design this model are not complete and do not explain the signalling details precisely (see Chapter 4 for more details).
### 7.3 Review of the Mathematical Models on TGF--β Signalling

<table>
<thead>
<tr>
<th>Parameters derived by Zi et al. [511]</th>
<th>Parameters from the literature</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental data sets (6 sets)</td>
<td>Initial values for the ratio and the exact values of cytoplasmic and nuclear SMAD2 and SMAD4</td>
<td>[384,386]</td>
</tr>
<tr>
<td>Model analysis and model validation</td>
<td>Volume estimation of the cell’s compartment</td>
<td>[386]</td>
</tr>
<tr>
<td>A few of the kinetic rates and initial values (estimation via SBML estimation tools)</td>
<td>Most of the kinetic rates</td>
<td>[93, 215, 225, 313,386,449,462, 513]</td>
</tr>
<tr>
<td>Two experimental data sets for validation</td>
<td></td>
<td>[386]</td>
</tr>
<tr>
<td>Receptor type I on the surface: 0.702 nM</td>
<td></td>
<td>[386]</td>
</tr>
<tr>
<td>Receptor type I in the endosome: 6.523 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptor type II on the surface: 0.201 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptor type II in the endosome: 1.440 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD2/3 in the cytoplasm: 60.6 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD2/3 in the nucleus: 28.5 nM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD4 in the cytoplasm: 50.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD4 in the nucleus: 50.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vcyt: 2.3e-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vnuc: 1e-12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In order to compare the information derived from the mathematical models above, Table 7.1 is created. Having a closer look at the data below, a large diversity is detected in the range of data provided by each model. As an example, the concentration of receptors on the cell surface varies from 1 nM to 112 nM. One explanation for such diversity may be the fact that the initial values for the components at steady-state are estimated. Depending on the demanding accuracy of the model and its relative predictions, estimation methods are different. In the explanations column, the units of the values are listed so that it is easier to compare them. The data has been aligned in a way that is more consistent with the hypothesis of our own model (see Figure 4.1). For instance, individual internalization of the TGF–β receptors has not been included and it is assumed that receptors can stimulate signal transduction only through the ligand-receptor complex on the membrane. In order to align the data with our model, I have included the receptor internalization rate in the receptor degradation rate and similarly, the receptor recycle rate in the receptor production rate (see "Explanations" column of Table 7.1). Looking through all the data provided in Table 7.1, Schmierer et al. [386] approach to data analysis and estimation is more reliable to establish the basis of our model. Schmierer et al. [386] data extraction, including protein concentrations or kinetic rates, follow a logical algorithm. Furthermore, there is no obviously out of range data detected in the following table in Schmierer et al. [386] experimental data (see Table 7.1).

<table>
<thead>
<tr>
<th>Components</th>
<th>Value</th>
<th>Models</th>
<th>Reference</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration of receptors on the membrane (in steady-state)</td>
<td>1.1 nM</td>
<td>Vilar et al. [449]</td>
<td>estimate</td>
<td>HaCaT cells, 103 receptors per cell</td>
</tr>
<tr>
<td></td>
<td>17 nM</td>
<td>Clarke et al [74]</td>
<td>estimate</td>
<td>HaCaT cells (V cyt = 9e-13, V nuc = 1e-13 ), 104 receptors per cell</td>
</tr>
<tr>
<td></td>
<td>7.5 nM</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td>HaCaT cells (V cyt = 1.05e-12, V nuc = 3.5e-13 )</td>
</tr>
<tr>
<td>Concentration of SMAD2/3 in the cytoplasm (in steady-state)</td>
<td>112 nM</td>
<td>Schmierer et al [386]</td>
<td>estimate</td>
<td>HaCaT cells ((V_{cyt} = 2.3\times10^{-12}, V_{nuc} = 10\times10^{-13})), 105 type 2 receptors per cell + 1 nM type 1 receptor per cell</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11 nM</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td>HaCaT cells ((V_{cyt} = 1.13\times10^{-12}, V_{nuc} = 3.75\times10^{-13})), 104 total receptors per cell</td>
<td></td>
</tr>
<tr>
<td>8.87 nM</td>
<td>Zi et al. 2011 [511]</td>
<td>estimate</td>
<td>HaCaT cells ((V_{cyt} = 2.3\times10^{-12}, V_{nuc} = 1\times10^{-13}))</td>
<td></td>
</tr>
<tr>
<td>300 nM</td>
<td>Clarke et al [74]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>492.61 nM</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121.2 nM</td>
<td>Schmierer et al [386]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>124.9 nM</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.6 nM</td>
<td>Zi et al. 2011 [511]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration of SMAD2/3 in the nucleus (in steady-state)</td>
<td>Melke et al. [308]</td>
<td>Vilar et al [449] study the receptor trafficking only. No data on nuclear-cytoplasmic shuttling of SMADs is provided.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vilar et al [449]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300 nM</td>
<td>Clarke et al [74]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>236.45 nM</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57 nM</td>
<td>Schmierer et al [386]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66 nM</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.5 nM</td>
<td>Zi et al. 2011 [511]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration of SMAD4 in the cytoplasm (in steady-state)</td>
<td>Melke et al. [308]</td>
<td>Vilar et al [449] study the receptor trafficking only. No data on nuclear-cytoplasmic shuttling of SMADs is provided.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration of SMAD4 in the nucleus (in steady-state)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>230 nM</td>
<td>Clarke et al [74]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1149.4 nM</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.8 nM</td>
<td>Schmierer et al [386]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>127.8 nM</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.8 nM</td>
<td>Zi et al. 2011 [511]</td>
<td>Schmierer et al [386]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 nM</td>
<td>Clarke et al [74]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>551.72 nM</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.8 nM</td>
<td>Schmierer et al [386]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57 nM</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.8 nM</td>
<td>Zi et al. 2011 [511]</td>
<td>Schmierer et al [386]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptor (type1, type2) production rate</td>
<td>Melke et al [308]</td>
<td>Vilar et al [449]</td>
<td>estimate</td>
<td>0.408 min(^{-1}) for the total receptor production rate plus the receptor recycle rate.</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>0.125, 0.25 min(^{-1})</td>
<td>Melke et al [308]</td>
<td>Vilar et al [449]</td>
<td>estimate</td>
<td>0.408 min(^{-1}) for the total receptor production rate plus the receptor recycle rate.</td>
</tr>
<tr>
<td>0.0103, 0.029 min(^{-1})</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td>0.0433 min(^{-1}) and 0.06169 min(^{-1}) for the each receptor production rate plus its recycle rate.</td>
<td></td>
</tr>
<tr>
<td>0.009 min(^{-1})</td>
<td>Chung et al [72]</td>
<td>calculation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0137, 0.019 nM(^{-1}) min(^{-1})</td>
<td>Zi et al. 2011 [511]</td>
<td>Wakefield et al. [462], calculation</td>
<td>0.01 min(^{-1}) for type 1 receptor and 0.004 min(^{-1}) for type 2 receptor</td>
<td></td>
</tr>
<tr>
<td>Receptor (type1, type2) degradation rate</td>
<td>0.0001 min(^{-1})</td>
<td>Melke et al [308]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.028 min(^{-1})</td>
<td>Vilar et al [449]</td>
<td>estimate</td>
<td>This is the constitutive degradation of the receptors. (0.361 min(^{-1}) for the rate of constitutive degradation plus internalization rate)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clarke et al [74]</td>
<td>Kavasak et al [225]</td>
<td>0.335 min(^{-1}) and 0.355 min(^{-1}) for the rate of constitutive degradation plus internalization rate of the receptors.</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>0.005, 0.025 min(^{-1})</td>
<td>Zi et al. 2007 [513]</td>
<td>0.00003 min(^{-1}) Schmierer et al [386]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00003 min(^{-1})</td>
<td>Chung et al. [72]</td>
<td>Vilar et al [449]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00256, 0.0132 min(^{-1})</td>
<td>Zi et al. 2011 [511]</td>
<td>Kavasak et al [225], Zi et al. [513]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligand-receptor complex association rate</td>
<td>Melke et al [308]</td>
<td>Vilar et al [449]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 min(^{-1})</td>
<td>Vilar et al [449]</td>
<td>The units of ligand concentration are chosen so that the association rate constant is the unit.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000012 nM(^{-1}) min(^{-1})</td>
<td>Clarke et al [74]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2197 nM(^{-2}) min(^{-1})</td>
<td>Zi et al. 2007 [513]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.11 nM(^{-1}) min(^{-1})</td>
<td>Schmierer et al [386]</td>
<td>Around 175.76 nM min(^{-1}) and 351.52 min(^{-1})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0000102 nM(^{-1}) min(^{-1})</td>
<td>Chung et al [72]</td>
<td>estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Around 0.96 min(^{-1}).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Around 0.00000816 min(^{-1}).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 7.4 Equations Describing the TGF−β and IL−6 Signalling Systems

#### 7.4.1 TGF−β Receptor Model

\[
A_{11} [R1]’ + A_{12} [R2]’ = f_1 \\
A_{21} [R1]’ + A_{22} [R2]’ = f_2
\]
7.4 Equations Describing the TGF-β and IL-6 Signalling Systems

\[ A_{11} = 1 + 4 \frac{[R1]}{K_1} + \frac{4\alpha}{K_1K_2K_{RC}} [R1] [R2]^2 \]
\[ A_{21} = 1 + 4 \frac{[R2]}{K_2} + \frac{4\alpha}{K_1K_2K_{RC}} [R1]^2 [R2] \]
\[ A_{12} = \frac{4\alpha}{K_1K_2K_{RC}} [R1] [R2]^2 \]
\[ A_{22} = \frac{4\alpha}{K_1K_2K_{RC}} [R1] [R2]^2 \]

\[ f_1 = v_1 - k_1 [R1] - 2 \left( k_{RC} + k_{LRC^*} \frac{[TGF]}{K_{LRC}K_{LRC^*}} + k_b [SMAD7]_C \right) \frac{[R1]^2 [R2]^2}{K_1K_2K_{RC}} - k_a [SMAD7]_C [R1] \]

\[ f_2 = v_2 - k_2 [R2] - 2 \left( k_{RC} + k_{LRC^*} \frac{[TGF]}{K_{LRC}K_{LRC^*}} + k_b [SMAD7]_C \right) \frac{[R1]^2 [R2]^2}{K_1K_2K_{RC}} \]

Note that the activated receptor complex, \([LRC^*]\), is given by:

\[ [LRC^*] = \frac{[TGF][R1]^2[R2]^2}{K_1K_2K_{RC}K_{LRC}K_{LRC^*}} \]

\[ \alpha = \left( 1 + \frac{[TGF]}{K_{LRC}} + \frac{[TGF]}{K_{LRC}K_{LRC^*}} \right) \cdot \]

7.4.2 Cytoplasmic SMAD Model

\[ \frac{d[SMAD2]_C}{dt} = v_2 - k_2 [SMAD2]_C - k_2^* [LRC^*] \frac{[SMAD2]_C}{[SMAD2]_C + K_2^*} - k_2^* [SMAD2]_C + k_2^* V_N^N [SMAD2]_N \]
\[ \frac{d[SMAD3]_C}{dt} = v_3 - k_3 [SMAD3]_C - k_3^* [LRC^*] \frac{[SMAD3]_C}{[SMAD3]_C + K_3^*} - k_3^* [SMAD3]_C + k_3^* V_N^N [SMAD3]_N \]

\[ A_{11} \frac{d[SMAD2^*]_C}{dt} + A_{12} \frac{d[SMAD3^*]_C}{dt} + A_{13} \frac{d[SMAD4]_C}{dt} = f_1 \]
\[ A_{21} \frac{d[SMAD2^*]_C}{dt} + A_{22} \frac{d[SMAD3^*]_C}{dt} + A_{23} \frac{d[SMAD4]_C}{dt} = f_2 \]
\[ A_{31} \frac{d[SMAD2^*]_C}{dt} + A_{32} \frac{d[SMAD3^*]_C}{dt} + A_{33} \frac{d[SMAD4]_C}{dt} = f_3 \]

\[ \frac{d[SMAD7]_C}{dt} = v_7 - k_7 [SMAD7]_C + k_7^* V_N^N [SMAD7]_N \]
7.4 Equations Describing the TGF–β and IL–6 Signalling Systems

\[ A_{11} = 1 + 9 \frac{[\text{SMAD2}]^2_C}{K_2} + 9 \frac{[\text{SMAD2}]^2_C [\text{SMAD4}]^3_C}{K_2 K_4 K_{24}} \]
\[ A_{13} = 9 \frac{[\text{SMAD2}]^3_C [\text{SMAD4}]^2_C}{K_2 K_4 K_{24}} \]
\[ A_{22} = 1 + 9 \frac{[\text{SMAD3}]^2_C}{K_3} + 9 \frac{[\text{SMAD3}]^2_C [\text{SMAD4}]^3_C}{K_3 K_4 K_{34}} \]
\[ A_{23} = 9 \frac{[\text{SMAD3}]^3_C [\text{SMAD4}]^2_C}{K_3 K_4 K_{34}} \]
\[ A_{31} = 9 \frac{[\text{SMAD2}]^3_C [\text{SMAD4}]^3_C}{K_2 K_4 K_{24}} \]
\[ A_{32} = 9 \frac{[\text{SMAD3}]^3_C [\text{SMAD4}]^3_C}{K_3 K_4 K_{34}} \]
\[ A_{33} = 1 + 9 \frac{[\text{SMAD4}]^2_C}{K_4} + 9 \frac{[\text{SMAD2}]^3_C [\text{SMAD4}]^3_C}{K_2 K_4 K_{24}} + 9 \frac{[\text{SMAD3}]^3_C [\text{SMAD4}]^3_C}{K_3 K_4 K_{34}} \]

\[ f_1 = k_2^* [\text{LRC}^*] \frac{[\text{SMAD2}]_C}{[\text{SMAD2}]_C + K_2^*} - 3k_{24}^i \frac{[\text{SMAD2}]^3_C [\text{SMAD4}]^3_C}{K_2 K_4 K_{24}} \]
\[ f_2 = k_3^* [\text{LRC}^*] \frac{[\text{SMAD3}]_C}{[\text{SMAD3}]_C + K_3^*} - 3k_{34}^i \frac{[\text{SMAD3}]^3_C [\text{SMAD4}]^3_C}{K_3 K_4 K_{34}} \]
\[ f_3 = v_4 - (k_4 + k_4^i) [\text{SMAD4}]_C + k_4^i v_N^C [\text{SMAD4}]_N - 3k_{24}^i \frac{[\text{SMAD2}]^3_C [\text{SMAD4}]^3_C}{K_2 K_4 K_{24}} \]
\[-3k_{34}^i \frac{[\text{SMAD3}]^3_C [\text{SMAD4}]^3_C}{K_3 K_4 K_{34}} \]

Note that \( A_{12} = A_{21} = 0 \).

7.4.3 Nuclear SMAD Model

\[
\frac{d}{dt} [\text{SMAD2}]_N = k_2^* [\text{SMAD2}]_N + k_2^i v_N^C [\text{SMAD2}]_C - k_2^e [\text{SMAD2}]_N
\]
\[
\frac{d}{dt} [\text{SMAD3}]_N = k_3^* [\text{SMAD3}]_N + k_3^i v_N^C [\text{SMAD3}]_C - k_3^e [\text{SMAD3}]_N
\]

\[
A_{11} \frac{d}{dt} [\text{SMAD2}^*]_N + A_{12} \frac{d}{dt} [\text{SMAD3}^*]_N + A_{13} \frac{d}{dt} [\text{SMAD4}]_N = f_1
\]
\[
A_{21} \frac{d}{dt} [\text{SMAD2}^*]_N + A_{22} \frac{d}{dt} [\text{SMAD3}^*]_N + A_{23} \frac{d}{dt} [\text{SMAD4}]_N = f_2
\]
\[
A_{31} \frac{d}{dt} [\text{SMAD2}^*]_N + A_{32} \frac{d}{dt} [\text{SMAD3}^*]_N + A_{33} \frac{d}{dt} [\text{SMAD4}]_N = f_3
\]

\[
\frac{d}{dt} [\text{SMAD7}]_N = v_T + \left( v_A + v_B \right) \frac{[\text{STAT3}]^2_N}{[\text{STAT3}]^2_N + K_B^2} \frac{[\text{SMAD3}^*_3 \cdot \text{SMAD4}^*]_N^2}{[\text{SMAD3}^*_3 \cdot \text{SMAD4}^*]_N^2 + K_A^2} - k_T [\text{SMAD7}]_N
\]
7.4 Equations Describing the TGF-β and IL-6 Signalling Systems

\[ A_{11} = 1 + 9 \frac{[\text{SMAD2}]^2_N}{K_2} + 9 \frac{[\text{SMAD2}]^2_N [\text{SMAD4}]^3_N}{K_2 K_4 K_{24}} \]
\[ A_{13} = 9 \frac{[\text{SMAD2}]^3_N [\text{SMAD4}]^2_N}{K_2 K_4 K_{24}} \]
\[ A_{22} = 1 + 9 \frac{[\text{SMAD3}]^2_N}{K_3} + 9 \frac{[\text{SMAD3}]^2_N [\text{SMAD4}]^3_N}{K_3 K_4 K_{34}} \]
\[ A_{23} = 9 \frac{[\text{SMAD3}]^3_N [\text{SMAD4}]^2_N}{K_3 K_4 K_{34}} \]
\[ A_{31} = 9 \frac{[\text{SMAD2}]^2_N [\text{SMAD4}]^3_N}{K_2 K_4 K_{24}} \]
\[ A_{32} = 9 \frac{[\text{SMAD3}]^2_N [\text{SMAD4}]^3_N}{K_3 K_4 K_{34}} \]
\[ A_{33} = 1 + 9 \frac{[\text{SMAD4}]^2_N}{K_4} + 9 \frac{[\text{SMAD2}]^3_N [\text{SMAD4}]^2_N}{K_2 K_4 K_{24}} + 9 \frac{[\text{SMAD3}]^3_N [\text{SMAD4}]^2_N}{K_3 K_4 K_{34}} \]

\[ f_1 = -k_2^* [\text{SMAD2}]^*_N + 3k_2^{24} V_N^C \frac{[\text{SMAD2}]^*_I [\text{SMAD4}]^*_I}{K_2 K_4 K_{24}} \]
\[ f_2 = -k_3^* [\text{SMAD3}]^*_N + 3k_3^{34} V_N^C \frac{[\text{SMAD3}]^*_I [\text{SMAD4}]^*_I}{K_3 K_4 K_{34}} \]
\[ f_3 = k_4^* V_N^C [\text{SMAD4}]_I - k_5^* [\text{SMAD4}]_N + 3k_4^{24} V_N^C \frac{[\text{SMAD2}]^*_I [\text{SMAD4}]^*_I}{K_2 K_4 K_{24}} + 3k_3^{34} V_N^C \frac{[\text{SMAD3}]^*_I [\text{SMAD4}]^*_I}{K_3 K_4 K_{34}} \]

Note that \( A_{12} = A_{21} = 0 \), and that the transcription factor complex \([\text{SMAD3}]_I \cdot [\text{SMAD4}]_I \) is given by:

\[[\text{SMAD3}]_I \cdot [\text{SMAD4}]_I = \frac{[\text{SMAD3}]^*_I [\text{SMAD4}]^*_I}{K_3 K_4 K_{34}}.\]

### 7.4.4 IL-6 Receptor Model

\[ A_{11}[80]' + A_{12}[130]' = f_1 \]
\[ A_{21}[80]' + A_{22}[130]' = f_2 \]

\[ A_{11} = 1 + \frac{[L]}{K_{80}} + \frac{[L][J][130]}{K_{80} K_{130} K_1} + 4 \frac{[L]^2 [80][J]^2 [130]^2}{K_{80}^2 K_{130}^2 K_1^2 K_2}(1 + 1/K_2^*) \]
\[ A_{21} = \frac{[L][J][130]}{K_{80} K_{130} K_1} + 4 \frac{[L]^2 [80][J]^2 [130]^2}{K_{80}^2 K_{130}^2 K_1^2 K_2}(1 + 1/K_2^*) \]
\[ A_{12} = \frac{[L][80][J]}{K_{80} K_{130} K_1} + 4 \frac{[L]^2 [80]^2 [J]^2 [130]}{K_{80}^2 K_{130}^2 K_1^2 K_2}(1 + 1/K_2^*) \]
\[ A_{22} = 1 + \frac{[J]}{K_{130}} + \frac{[L][80][J]}{K_{80} K_{130} K_1} + 4 \frac{[L]^2 [80]^2 [J]^2 [130]}{K_{80}^2 K_{130}^2 K_1^2 K_2}(1 + 1/K_2^*) \]
7.5 Review of the Positive Feedback Loop in TGF−β Signalling

This document investigates the existence and importance of a positive feedback on the TGF−β signal transduction via SMADs. TGF−β signalling pathway and its importance in cell responses have been known for decades. Previous mathematical and computational models of this pathway have noticed one or two inherent negative feedbacks (SMAD7 and TMEPAI) by which help to explain the long-term responses of cells to TGF−β stimulation. System analysis suggests detection of oscillation in TGF−β signalling with negative delayed feedbacks. These oscillations have not been reported experimentally. These clues have provided us a new hypothesis: there might be a (merely strong) positive feedback to TGF−β signalling which couples with the SMAD7 negative feedback. We speculate that interactions between the positive and negative feedback can explain the non-oscillatory behaviour of the cell response (here, PSMAD2) at both short and long times after stimulation by TGF−β.

We have introduced a recently identified microRNA, miR-433, to initiate a positive feedback via a complex series reactions. It has been known that TGF−β is capable of regulating the expression of different groups of microRNAs. However, it was not clear which microRNA TGF−β regulates. Li et al. [269] have introduced miR-433 as the targeted microRNA via TGF−β in fibrotic kidneys. Using invivo and invitro studies, they have found antizyme inhibitor 1 (Azin1) is regulated by miR-433 in renal injuries.

miR-433 is reported to have two SMAD-binding sites (SBS)s at its 5′-end in the genomes of human, mouse and rat (at -1845 bp and - 582 bp) [269]. These SBSs, in addition to the binding of SMAD3 to SBS2 of miR-433 in mouse embryonic fibroblasts, these also appears to be a physical interaction between SMAD3 with miR-433. Li et al. show that treatment of TECs (thyroid epithelial cells) with TGF−β decreases the expression of Azin1 at the RNA and protein levels [269]. It is also reported that Azin1 affects the regulation of ECM expression in liver fibrosis [269, 354]. Applying knockdown experiments to TECs and analysis with the luciferase reporter assays, Li et al. proposed an inverse relationship between Azin1 and miR-433 expression. More precisely, miR-433 targets the 3′-untranslated region of Azin1 and negatively regulates its expression.
The role of Azin1 in promotion of polyamine synthesis has been known [212,269]. Indeed, Azin1 binds to the antizyme and inhibits its function, including the initiation of ornithine decarboxylase (ODC) degradation [212,269]. ODC is essential for biosynthesis of polyamines [212,269]. Depletion of polyamine activates TGF−β signalling [269,277,345,365]. We postulate that the Azin1 expression maintains cellular polyamine concentration, suppressing of TGF−β-induced expression of fibrotic markers [269]. Reports show that over-expression of Azin1 suppresses the expression levels of TGF−β and its type 1 receptor while, it has no effects on the receptor type 2 expression levels [269].

A diagram of the positive stimulatory feedback for TGF−β signal transduction is illustrated in figure 1. Since this positive feedback includes multiple stimulatory/inhibitory reactions, transcription/translation and possible change of compartments for some components, a noticeable time-delay needs to be included in the model.

The discovery of such a positive feedback would be an important factor for understanding the pathway modelling of TGF−β. We believe it is now possible to describe the dual behaviour of the cancer cell responses to TGF−β stimulation with a mathematical model.

Figure 7.1: The simplified cartoon of the feedbacks in TGF−β signal transduction system
7.5 Review of the Positive Feedback Loop in TGF–β Signalling

7.5.1 Preliminary Experimental Data Integration

Figure 7.2: A) Western blot analysis of SV40-immortalized MEF cells (immortal MEFs) stimulated by TGF–β for different times, probed with PSMAD2 and β-tubulin anti-bodies. The results indicate that total PSMAD2 levels peak after 60 minutes of TGF–β stimulation. B) Western blot analysis of wild type MEF cells (MEF Gp130^{F/F}) stimulated with TGF–β for different duration, probed with PSMAD2 and β-tubulin anti-bodies. The results indicate that total PSMAD2 levels peaks after 20 minutes of TGF–β stimulation.
7.5 Review of the Positive Feedback Loop in TGF–β Signalling

Figure 7.3: Western blot analysis of wild type MEF cells (MEF Gp130F/F) stimulated with IL–6 (10% conditioned medium) for different duration, probed with PSTAT3 and β-tubulin antibodies. Results indicate that total PSTAT3 levels rise gradually until 60 minutes of IL–6 stimulation.

The first part of the experimental data extraction was obtained at the Ludwig Institute for Cancer Research, Parkville branch, Epithelial laboratory. Mouse embryo fibroblast cells (SV40-immortalized and WT MEFs) were analyzed using, protein assays and Western blot methods. Figure 7.2 shows the Western blot results of total PSMAD2 measured in the two cell lines. These cells were stimulated with TGF–β for different times. Stimulation activates the TGF–β signalling from the membrane to the nucleus of a cell. By altering the stimulation duration we can develop a time course for the PSMAD2 dynamic changes in the cells. A similar protocol has been followed for detecting total PSTAT3 in different cell lines, activating IL–6 signalling pathway. The results of western blot analysis for wild type MEFs and the relative time course are shown in Figure 7.3. These experimental results are for the evaluation of our model.

Clearly, it is possible to detect PSMAD2 and PSTAT3 in MEF cells. Western blotting is not the most precise method for quantification of the experimental data; however, it does provide initial estimate of the PSMAD2 levels in a cell line.
Bibliography


[284] K. Luo and H. Lodish. Signaling by chimeric erythropoietin-TGF-β receptors: homodimerization of the cytoplasmic domain of the type I TGF-β receptor and heterodimerization with the type II receptor are both required for intracellular signal transduction. The EMBO journal, 15(17):4485, 1996.


Control of hepatocyte proliferation and survival by Fgf receptors is essential for liver regeneration in mice. *Gut*, 2014. doi:10.1136/gutjnl-2014-307874.


(CRM1)-dependent nuclear export of Smad ubiquitin regulatory factor 1 (Smurf1) is essential for negative regulation of transforming growth factor-β signaling by Smad7. *J Biol Chem*, 278(12):10716–21, 2003.


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Khatibi, Shabnam

Title:
Signalling and crosstalk in cytokine pathways: mathematical modelling and quantitative analysis

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/91092