Mathematical models of cell signalling in heterogeneous populations

Léa Sta

Department of Mathematics

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

January, 2023

The candidate confirms that the work submitted is her own except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below.

The candidate confirms that appropriate credit has been given where reference has been made to the work of others.

Joint publications

The mathematical analysis in Chapter 3, the method of Section 4.1 and the study of the general receptor-ligand system in Section 4.3 has been refereed and accepted for publication in SIAM as follows:

- **Sta, L.**, Adamer, M., and Molina-París, C. (2022). Algebraic study of receptor-ligand systems: a dose-response analysis. arXiv preprint arXiv:2206.13364.

Part of the work in Chapter 3 and Chapter 4 (mainly Sections 4.2 and 2.10) has been submitted as follows:

- **Sta, L.**, Voisinne, G., Cotari, J., Adamer, M., Molina-Paris, C., and Altan-Bonnet, G. (2022). Tuning of cytokine signaling through imbalanced abundances of receptors and kinases. bioRxiv.

A manuscript with the work conducted in Chapter 5 is in preparation.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764698.

First and foremost, I would like to thank my supervisors, Carmen Molina-París, Grant Lythe and Martín López-García for their patience and guidance throughout the past four years. I am very appreciative of their invaluable advice. I am also extremely grateful for the mathematical biology group they formed in Leeds, it really contributed to make my time in England enjoyable. I also want to thank them for sharing their extended scientific network, which allowed me to make international collaborations. Thank you for all the opportunities I was given during this PhD, the QuanTII meetings, the conferences and the public outreach sessions.

I want to thank Grégoire Altan-Bonnet for directing my interest to most of the specific questions I attempted to answer in this thesis, and for giving me the opportunity to visit his laboratory at the NIH, USA. This work would not have been the same without his guidance and continuous enthusiasm.

I would like to express my gratitude to Michael Adamer for his constant reassurance and his patience when answering all my questions related to algebra and paper writing. Working with him really made me a better scientist.

I give thanks to Guillaume Voisinne and Jesse Cotari for providing the data set of the first chapter and spending so much time explaining the biological mechanisms to me. In particular, I want to thank Guillaume for the time he took to answer all my questions related to data fitting.

I give thanks to the PhD students of the mathematical biology group in Leeds for their friendship, their emotional support and technical help. Working and travelling with them was always such a pleasure. They are the greatest team I could have wished for and I do hope we will stay in touch for more hikes and highly specific linguistic discussions.

From QuanTII, I want to thank the other ESRs for their companionship. We lived the thesis adventure together and despite the difficult COVID time, our interactions were always as if we knew each others for a long time. I also want to give special thanks to Jessica Brennan for her patience and help for administrative and travel-related issues.

Finally, I would like to thank my family and my friends for their eternal support. They always manage to cheer me up and I am so grateful for their presence in my life. I also want to thank Pim for being there since the beginning and the UK blues dance community for making me feel at home on the dancefloor even when I knew no one.

Abstract

Immune cells express thousands of receptors on their membrane surface to sense their environment and communicate with each other. Receptors bind specifically to extra-cellular molecules called ligands. The binding of a ligand to its receptor initiates an intra-cellular signalling cascade which leads to the control of cellular fate, such as division, death, migration or differentiation. As every cell expresses a different number of receptors, each cell responds differently to a given ligand. First motivated by seemingly paradoxical experimental observations on the interleukin-7/interleukin-7 receptor (IL-7/IL-7R) receptor-ligand system, this thesis investigates how receptor copy numbers impact the cell's response, as measured by the amplitude and the half-maximal effective concentration (or EC_{50}). In particular, deterministic mathematical models of various receptor-ligand systems are developed. For each model, making use of algebraic tools, such as Gröbner bases, analytic expressions for the amplitude and the EC_{50} are computed. Such expressions allow one to identify precisely how a cell's response depends on the receptor core structure, namely receptor chain copy numbers and receptor architecture. They also reduce numerical errors and facilitate parameter inference, as demonstrated by the fitting of two IL-7R models to the motivating experimental data set. The results obtained are generalised to a larger family of receptor-ligand systems, for which the amplitude is computed without the use of advanced algebraic tools. Finally, as the immune system relies on the coordination of many cells to fight pathogens, the complex relationship between the cell population dynamics and the receptor copy number distribution in the cellular population is examined. To

this end, agent-based models of increasing complexity, which model the competition for interleukin-2 (IL-2) within the T cell population, are constructed, by adding stochastic cellular events one at a time. A mathematical description of each model is provided, which enables us, when possible, to derive the desired receptor copy number distribution (in this case for the IL-2 receptor).

Abbreviations

ABC	Approximate Bayesian computation
ABM	Agent-based model
CRN	Chemical reaction network
EC_{50}	half-maximal effective concentration
IC_{50}	half-maximal inhibitory concentration
IEK	Intra-cellular extrinsic kinase
IL-n	Interleukin n
IL-nR	Interleukin n receptor
ILC	Innate lymphoid cell
JAK	Janus kinase
RTK	Receptor tyrosine kinase
SRLK	Sequential receptor-ligand with intra-cellular extrinsic kinase
STAT	Signal transducer and activator of transcription
T regs	Regulatory T cells

Mathematical notation

$\mathbb{C}$	Set of complex numbers
$\mathbb{Q}$	Set of rational numbers
$\mathbb{N}$	Set of natural numbers, comprising zero
$\mathbb{N}^*$	Set of natural numbers without zero
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of positive real numbers
z^T	Transpose of vector (or matrix) z

Preface

During the first three years of the thesis, I was a Marie Skłodowska-Curie early stage researcher in the Innovative Training Network: Quantitative T cell Immunology and Immunotherapy (QuanTII). This network involves mathematicians, physicists, biologists and computer scientists from around the world, from academia and industry partners. It aimed to develop novel quantitative understanding in the compartmentalisation and dynamics of T cells, T cell receptor (TCR) repertoires in health and disease in blood and tissues, and T cell immunotherapies. Fifteen early stage researcher positions were open across Europe, from which three were for Leeds-based fellows. I received the project entitled "Signals determining naive T cell maintenance". Started in January 2019, this thesis took its most important turn when I went to Dr. Grégoire Altan-Bonnet's laboratory at the NIH (Bethesda, MA, USA) in June 2019. During this one-month secondment, I had the opportunity to discover wet laboratory work, thus understanding better how biologists acquire the data mathematicians are using to parametrise their models. But mainly, it was an occasion for me to create a long-term collaboration with Grégoire Altan-Bonnet. Indeed, he directed my interest to the analysis of the amplitude and EC_{50} of receptor-ligand systems and encouraged me to explore with analytical methods how the dose-response curve depends on receptor expression levels (or copy numbers). He showed me the data and the first IL-7R model described in Chapter 3, and explained to me how his team could not recreate the EC_{50} experimental behaviour with this model. Back to Leeds, I was introduced to the Gröbner bases, an algebraic tool that has been essential in this thesis.

During the Covid-19 pandemic, I pursued my thesis at my parents' home in France. This remote-working time offered an opportunity to develop online collaborations. I started writing a paper with Grégoire Altan-Bonnet and two of his ex-collaborators, Jesse Cotari and Guillaume Voisinne. I was in contact with a post-doctoral researcher in Zurich, Michael Adamer, whose background is in algebraic biology. He was incredibly helpful with algebraic questions around the use of Gröbner bases. These two collaborations led to the work presented in Chapters 3 and 4 and resulted in two papers: Sta *et al.* (2022a) (accepted for publication) and Sta *et al.* (2022b) (submitted).

In June 2021, Grégoire Altan-Bonnet contacted me and our team in Leeds with another research question: can we develop mathematical models that take into account the distribution of receptor expression levels in a cell population (that may compete with another cell population)? Can we try to understand how this distribution and the immune response at the population level are related? These questions led to the work of Chapter 5 that I started in November 2021. The agent-based model of this chapter reached its final form in March 2022. This work allowed me to complete a separate and brief study, that I started in September 2019 on the competition between T cells and innate lymphoid cells for IL-7, which I include in Chapter 5.

Contents

List of Figures

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

1.1 Biological introduction

1.1.1 Overview of the immune system

The immune system is a complex network of cells, proteins and organs that defend the body against invaders (collectively called *pathogens*) such as bacteria, viruses, and parasites, while preserving its own cells. It means being able to identify intruders (and discriminate between proteins naturally present in the body and non-self proteins) and scale the response to the pathogen type and dose. Indeed, an infection has to be fought with the most efficient weapon available while minimising the damage caused to the neighbouring cells and tissue (Sompayrac, 2019).

There exist two defence strategies against an invasion. The *innate immune system* is composed of barriers, such as skin and mucous, which prevent pathogens from entering the body, and specific cells which can quickly eliminate the majority of bacterial infections. For instance, in vertebrates, a certain type of cells, called *macrophages*, detect most bacteria and "eat" them (phagocytosis), thus removing the intruder from the body (Sompayrac, 2019). All living organisms, even plants, fungi, insects or primitive multi-cellular organisms, possess an innate immune system (Murphy *et al.*, 2011). In vertebrates, it is completed with the *adaptive immune system* (or *acquired immune system*). This other defence strategy is highly specific to each particular invader. Indeed, it keeps track of the pathogens the

body encountered to respond quicker and more efficiently in case of a re-infection. Vaccination relies on this memory mechanism. The adaptive immune response is carried out by white blood cells known as *lymphocytes*, in particular B cells (which mature in the bone marrow) and T cells (which mature in the thymus). The mathematical models of this thesis mainly focus on T cell communication and T cell population dynamics.

1.1.2 T cells at work

T cells sense their environment and communicate between each other through thousands of receptors at their membrane surface, which bind specifically to extra-cellular molecules, called ligands. Receptors can be composed of multiple molecules: proteins that cross the cell membrane, known as *trans-membrane chains*, and intra-cellular molecules that bind to these chains. The binding of a receptor to its ligand induces an intra-cellular cascade of chemical reactions, called *downstream signalling* (or *signal*), from the surface of the cell to its nucleus, which alters the cell's gene expression and regulates the cell's fate such as migration, proliferation, death, or differentiation (De Belly *et al.*, 2022; Maxwell & Webb, 2008; Uings & Farrow, 2000).

In absence of threat (*homeostasis*), there exists a subset of T cells, called *naive cells*, that have never been part of an immune response. This T cell population is maintained in part thanks to survival signals such as the binding of the ligand called interleukin-7 (IL-7) to its specific receptor (the interleukin-7 receptor or IL-7R)[1] (Akdis *et al.*, 2011). To detect pathogens, naive T cells express a particular type of receptor called T cell receptor (TCR)(Sompayrac, 2019). All the TCRs expressed by one cell[2] are identical and only bind to a set of specific proteins. In order to recognise the multitude of pathogens that naturally exists, the body creates a large number of T cells, each with a different TCR structure, thanks to quasi-random somatic gene recombination (Sompayrac, 2019) (it selects, randomly edits, and combines together different variants of the three main genes that give the TCR repertoire its diversity). This can lead up to 4.5×10^{15} (Murugan *et al.*,

[1] Note that IL-7 is also necessary for T cell development in the thymus (Hong *et al.*, 2012).
[2] Each T cell expresses about 3×10^4 identical TCRs (Varma, 2008).

2012) possible TCRs in the whole repertoire, which is more than the number of stars in our galaxy (about 1×10^{11} to 4×10^{11} stars (Franck *et al.*, 2001)).

When a pathogen enters the body (supposing it passed the defences of the innate immune system), it, first, must be detected by a special category of cells, called antigen-presenting cells (APCs). These cells present pieces of antigens (the molecules that flags the pathogen as an invader: it can be proteins, peptides, lipids,...) to a T cell's TCRs. If the antigen molecules bind to the TCRs (*i.e.* the TCRs' structure matches the antigens' structure) of the T cell, and if enough TCRs are bound to antigens (the signal must be stronger than a certain threshold: multiple TCRs must signal in the same time), the T cell switches from a resting state to an activated state (through a signal that changes its gene expression, as explained above). The activation of a T cell triggers its machinery which will eventually clear the pathogen from the body. This machinery depends on the T cell type: an activated killer T cell will kill (by inducing apoptosis, *i.e* assisted suicide) virus-infected cells, helper T cells will start producing the proteins that will support the other immune cells during the infection (Sompayrac, 2019), *etc.* One of the most important proteins secreted by a subset of activated T cells is the interleukin-2 (IL-2). This ligand binds to the interleukin-2 receptor (IL-2) composed of three trans-membrane chains (γ_c, IL-2Rβ and IL-2Rα) and several intra-cellular molecules (Rochman *et al.*, 2009). Note that naive T cells do not produce IL-2 and do not express the complete interleukin-2 receptor (Il-2R): the IL-2Rα chain is missing[1]. While only a certain type of T cells start secreting IL-2 once activated, all activated T cells express IL-2Rα (and thus bind to IL-2 with a high affinity) (Shipkova & Wieland, 2012). The binding of IL-2 to the IL-2R triggers signals that induce activated T cells to divide (Akdis *et al.*, 2011). Cell division is the key process by which a local immune response (a cell detects an intruder) can become a global immune response (multiple cells, maybe in different parts of the body, fight against the pathogen). Indeed, when a cell divides, it makes copies of itself. That is, the number of T cells expressing TCRs that are able to detect the intruder is increasing. Furthermore, the secretion of IL-2 in the extra-cellular medium, in addition to other co-stimulatory signals, will activate

[1]The incomplete IL-2R may bind to IL-2 but with a very low affinity. Thus, the binding of IL-2 to the IL-2R deprived of IL-2Rα is neglected in the mathematical models of this thesis.

certain neighbouring T cells which, in turn, will start to produce and consume IL-2, and divide (Sompayrac, 2019). Thus, an increasing number of cells joins the force to fight the intruder, until it is eventually cleared from the body. At the end of an immune response (when the pathogens has been eliminated), some activated T cells become memory T cells (thus keeping track of the invader) and the others die. We will see in Chapter 5 how a subset of T cells can become activated in absence of pathogens and secrete IL-2. This T cell activation at homeostasis, when controlled, is essential to maintain an effective immune system.

1.1.3 Receptor-ligand systems and modelling

The human body consists of more than 3×10^{13} cells (Bianconi *et al.*, 2013), each of them ensuring crucial functions. To guarantee proper functioning, cells need to communicate with each other and sense their environment. They do so by expressing thousands of surface receptors, just like T cells, which binds to ligands. Receptors can be classified into different families, based on how they respond to signalling molecules. The two families discussed in this thesis are receptor tyrosine kinases (RTKs) (Du & Lovly, 2018) and cytokine receptors signalling through associated Janus kinases (JAKs) (Leonard *et al.*, 2019). These two families of receptors are similar in many regards. Both RTKs and JAK-associated receptors are activated by ligand-induced dimerisation, *i.e.*, the ligand brings together the two protein chains which form the full receptor[1]. This dimerisation leads to phosphorylation[2] of the receptor intra-cellular domains. In spite of their similarities, these two receptor families have one notable difference in the mechanisms by which they initiate downstream signalling: while RTKs have intrinsic kinase activity, cytokine receptors do not contain intrinsic kinase domains. Instead cytokine receptors make use of Janus family tyrosine kinases (a type of extrinsic intra-cellular kinase) and signal, in part, by the *activation of signal transducer*

[1] Note that, in this thesis, we call a *dimer* a set of two receptor chains. A *heterodimer* (or *heterodimeric receptor*) is a dimer composed of two different receptor chains, and a *homodimer* (or *homodimeric receptor*) is a dimer with two identical chains.

[2] *Phosphorylation* is the attachment of a phosphate group to a molecule or protein (molecule A). This process can be triggered by another molecule (molecule B) and we say that molecule A is *phosphorylated* (by molecule B). A molecule can be phosphorylated by itself and we say that it has been *autophosphorylated*. A molecule A which is phosphorylated, is usually written pA.

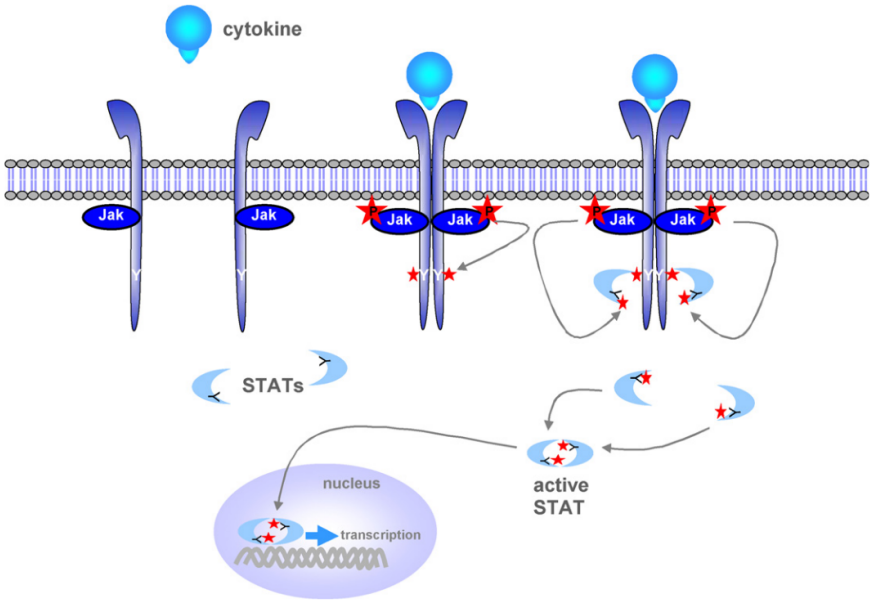


Figure 1.1: Diagram of the JAK/STAT signalling pathway initiated by cytokine molecules binding to a generic receptor dimer, reproduced from Ref. Haan *et al.* (2006) with permission from Elsevier. The red stars attached to a molecular species indicate that the species is phosphorylated.

and activator of transcription (STAT) proteins (Lin & Leonard, 2019). The interaction between the JAK molecule and the activator of transcription STAT is called the JAK/STAT signalling pathway[1] (O'Shea *et al.*, 2015; Steelman *et al.*, 2004; Villarino *et al.*, 2015) and is illustrated in Figure 1.1 for a generic receptor dimer. As described by Haan *et al.* (2006), when the two chains of a receptor are brought together by a cytokine (ligand), the JAK molecules they are bound to trans-autophosphorylate themselves. The tyrosine residues of the intra-cellular tail of the receptor chains are then phosphorylated by the JAKs and will act as docking sites for other intra-cellular proteins such as STAT molecules. STAT molecules bind to the receptor chains, become phosphorylated themselves and then dissociate. Two phosphorylated STAT (pSTAT) molecules will then form a dimer in the cell cytoplasm, migrate to the nucleus and act as transcription factor, *i.e.*, read the part of the DNA which will determine cellular fate. For example,

[1]We call signalling pathway the cascade of chemical reactions from the surface of a cell to its nucleus, which was triggered by the binding of a ligand to its receptor. The JAK/STAT pathway is the name of one of the multiple existing pathways.

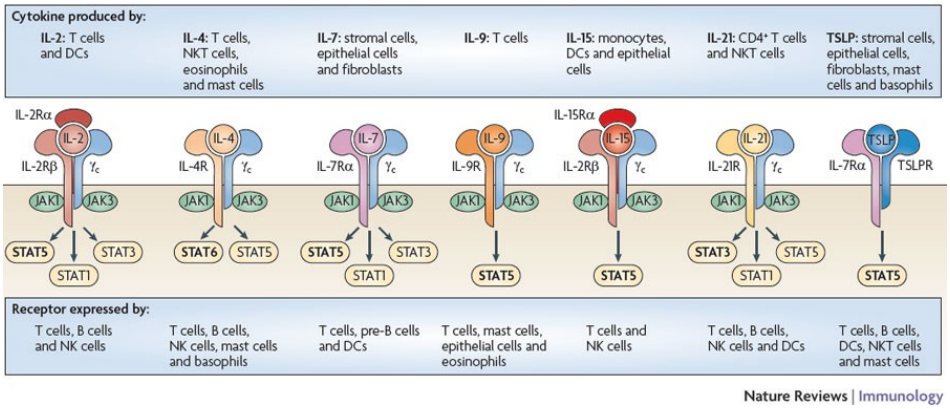


Figure 1.2: The common gamma chain family, comprising receptors for interleukin-2, 4, 7, 9, 15 and 21. Figure taken from Ref. Rochman *et al.* (2009) (reproduced with permission from Springer Nature).

the common gamma chain family of cytokine receptors, comprising receptors for interleukin-2 (IL-2), 4, 7, 9, 15 and 21 (Akdis *et al.*, 2011; Lin & Leonard, 2018; Rochman *et al.*, 2009), signals through the JAK/STAT pathway (see Figure 1.2). This family of receptors shares the eponymous common gamma chain (written γ_c or sometimes γ) which signals through the intra-cellular extrinsic Janus kinase molecule, JAK3. All receptors in this family complete the signalling core of the receptor with a second chain associated with JAK1: IL-4, IL-7, IL-9 and IL-21 make use of IL-4Rα, IL-7Rα, IL-9Rα and IL-21Rα, respectively. IL-2 and IL-15 share the JAK1-associated IL-2Rβ chain and include a third chain for specificity and enhanced sensitivity (IL-2Rα and IL-15Rα, respectively). This thesis mainly focuses on IL-7 and IL-2 receptors (IL-7R and IL-2R), which signal through STAT5 phosphorylation (Lin & Leonard, 2019). The amount of phosphorylated STAT can often be used as the experimental measure of the intra-cellular response generated by the cytokine stimulus. For instance, the quantity of phosphorylated STAT5 (pSTAT5) is a measure of the biological effect induced by the binding of IL-2 or IL-7 to their respective receptor.

Receptor-ligand interactions are essential in cell-to-cell communication, as is the case for immune cell populations (Farhat *et al.*, 2021), and thus, a large body of literature has been devoted to the experimental and theoretical study of

cell signalling dynamics (De Belly *et al.*, 2022; Feinerman *et al.*, 2010; Gonnord *et al.*, 2018; Lauffenburger & Linderman, 1996; Leonard *et al.*, 2019; Molina-París *et al.*, 2013; Park *et al.*, 2019; Ring *et al.*, 2012; Rochman *et al.*, 2009; Wiley *et al.*, 2003), in particular cytokine signalling (Altan-Bonnet & Mukherjee, 2019; Lin & Leonard, 2018; Raeber *et al.*, 2018; Villarino *et al.*, 2015). The study of a receptor-ligand system generally relies on the analysis of its dose-response (or concentration-effect) curve, which describes the relation between ligand concentration and the biological effect (or cellular response) it generates when binding its specific receptor (Lambert, 2004; Lauffenburger & Linderman, 1996; Maxwell & Webb, 2008). For example, the quantity of pSTAT5 as a function of the IL-2 concentration is a dose-response curve of the IL-2/IL-2R system. Mathematical models of receptor-ligand systems have been developed to compute a dose-response curve, under the assumption that a biological effect is proportional to the number of ligand-bound receptors (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Molina-París *et al.*, 2013; Wiley *et al.*, 2003).

Given a dose-response curve, two quantities (or metrics) have been defined to characterise the properties of the ligand-receptor system under consideration. These metrics are: the amplitude, which is defined as the difference between the maximal and minimal response, and the half-maximal effective concentration (or EC_{50}), which is the concentration of ligand required to induce an effect corresponding to 50% of the amplitude (Lambert, 2004; Lauffenburger & Linderman, 1996; Maxwell & Webb, 2008). In pharmacology, the amplitude is said to be a measure of the efficacy of the ligand, and the EC_{50}, a measure of the potency (or sensitivity) of the ligand (for a given receptor) (Dushek *et al.*, 2011; Lambert, 2004; Maxwell & Webb, 2008). More mathematical definitions of the amplitude and EC_{50}, as well as an illustration, will be given in Section 2.4.

Exploiting the controlled environment of *in vitro* experiments, most cell signalling studies focus on the estimation of the affinity constant for a given receptor-ligand system, and the quantification of biochemical on and off rates for the binding and unbinding, respectively, of receptor and ligand molecules. Mathematical models of cell signalling often consist in fitting the dose response curve to experimental data using statistical methods (Chen *et al.*, 2013; Finlay *et al.*, 2020; Gesztelyi *et al.*, 2012; Goutelle *et al.*, 2008; Jiang & Kopp-Schneider, 2014; Keshtkar *et al.*, 2021;

Li *et al.*, 2015; Suriyatem *et al.*, 2017). Other cell signalling mathematical models can study receptor trafficking, *i.e.*, internalisation, recycling or degradation of the bound receptor-ligand complex (Fallon & Lauffenburger, 2000; Molina-París *et al.*, 2013; Park *et al.*, 2019; Wilmes *et al.*, 2021), the signalling pathway (Anders *et al.*, 2020; McKeithan, 1995), or the competition of cells (Feinerman *et al.*, 2010) or receptors (Higuera *et al.*, 2021) for the same ligand. As diverse as cell signalling studies can be, most of them consider that cells of the same population behave in a similar manner, for instance by expressing the same number of receptors (or receptor constituents). However, recent single-cell studies have shown that cells present heterogeneous expression levels of receptor copy numbers. Not only does the copy number depend on the cell type, but receptor copy numbers vary strongly between isogenic[1] cells of the same type (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Feinerman *et al.*, 2010). Given the heterogeneity of receptor copy numbers across and within cell types, it is timely to understand how a cell's response to a given ligand depends on the expression level (or copy number) of its receptor. This quantification will be a first step to account for the variability of receptor expression levels when designing and studying receptor-ligand models (both from an experimental and mathematical perspective) (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Gonnord *et al.*, 2018; Ring *et al.*, 2012). The heterogeneity in cellular responses induced by the variability in receptor expression levels of a given cell population may significantly impact the cell population dynamics. The link between the receptor copy number distribution among the cell population and the population dynamics must also be better understood.

1.2 Objectives of the thesis

As previously mentioned, cells of different type, but also isogenic cells, express heterogeneous levels of receptors. This heterogeneity might have great implications in modulating, for instance, the immune response to a pathogen or in cancer, due to the diversity of potential cellular responses in the population. A natural conjecture would be that increasing the abundance of a receptor chain

[1] Isogenic cells all have identical genes.

increases the cell's response, reflected in an increase in amplitude (it increases the number of fully-formed receptors) and decrease in EC_{50} (increases the sensitivity of the ligand). This behaviour has been observed in RTKs of cancerous cells (Bache *et al.*, 2004; Du & Lovly, 2018; Eladdadi & Isaacson, 2008; Regad, 2015). However, Gregoire Altan-Bonnet, Guillaume Voisinne and Jesse Cotari conducted experiments that showed that increased γ_c abundance decreased the amplitude and the sensitivity of IL-7-induced cellular response in a population of T cells. In Chapter 3 of this thesis, I present the experimental results that lead to this seemingly paradoxical observation, and propose three mathematical models of the IL-7R in an attempt to explain it. The first model is a simple IL-7 receptor model in which the molecular complex able to induce the signal is formed sequentially. This model also allows the formation of complexes deprived of the extrinsic kinase and thus, unable to signal. The formation of these "dummy" complexes can explain the paradox only partially and variations of the first model are investigated. This chapter demonstrates that, contrary to popular belief, dose-response curves are not univariate functions of the concentration of ligand but instead are greatly influenced by the copy number of receptor constituents. In this chapter, I also compute the analytic expressions of the amplitude and EC_{50} of the first and last models, making use of Gröbner bases. The method employed to compute such expressions is summarised in Chapter 4. Amplitude and EC_{50} are key quantities commonly used in pharmacodynamic modelling, yet a comprehensive mathematical investigation of the behaviour of these two metrics is still outstanding for most receptor-ligand systems. Analytic expressions of these quantities show directly the dependence of the amplitude and EC_{50} on the parameters of the system and facilitate parameter exploration. In Chapter 4, I apply the method to additional examples of receptor-ligand systems, thus exploring other receptor configurations of biological and immunological relevance. The amplitude and EC_{50} expressions of some models are very similar, which led me to study a generalised receptor-ligand system with n trans-membrane chains. The application of the method to these different receptor architectures also showed its limitations: some models are too complex to be solved with the Gröbner basis method. In the last section of Chapter 4, I investigate whether I can analyse the amplitude and EC_{50} of more complex models by computing the

amplitude and EC_{50} expressions of simpler sub-models. Chapters 3 and 4 focus on the response of a single cell but do not explore the consequences of receptor copy number variability on the dynamics of a population of T cells. In turn, these chapters do not take into account the impact of the population dynamics[1] on the receptor distribution. I propose to explore this complex relationship in Chapter 5, by constructing agent-based models of the competition for IL-2 within a T cell population. I start with a simple and fully deterministic base model of the competition for this cytokine within a population with a constant number of cells. Then, I add stochastic cellular rules one at a time, such as death or division events, thus building more complicated models. For each model, I provide a mathematical description and try to determine the shape of the IL-2R copy number distribution at any time. Finally, I suggest several extensions of these models, comprising the competition for IL-2 between two different T cell populations. A Python code for the most complicated model (which encompasses the other models) is provided.

Overall, this thesis explores the following questions: How does cell signalling depend on molecular copy numbers? How can we develop mathematical models that capture the cellular heterogeneity in copy numbers? How can we model the difference of response to similar stimuli? After recapitulating, in Chapter 2, the mathematical tools and methods that will be employed in this thesis, I show, in Chapter 3, that the heterogeneity in molecular copy number may impact cell signalling in a counter-intuitive manner. Chapter 4 presents theoretical results which show that the receptor architecture can change the way the amplitude and EC_{50} of the model vary in response to receptor copy number differences. Finally, in Chapter 5, I combine mathematical descriptions and qualitative observations from numerical simulations to analyse agent-based models of the competition for IL-2 within a T cell population. I determine (when possible) how cellular dynamics events change the distribution of IL-2 receptor copy number expression within the cell population. I conclude this thesis in Chapter 6.

[1] For instance, what is the impact of the death of a cell on the receptor distribution? The death of one hundred cells? The arrival of new cells?

Chapter 2

Mathematical background

This chapter provides a background in chemical reaction network theory, perturbation theory and stochastic processes, as well as an overview of some related mathematical tools and methods which will be used in this thesis. Sections 2.3, 2.4 and 2.5 are directly adapted from Ref. Sta *et al.* (2022a).

2.1 Ordinary differential equations and systems

A major part of the work of this thesis consists of deterministic modelling, in particular the application of ordinary differential equations (ODEs) to represent biological mechanisms. Models consisting of a set of ODEs are commonly used in biological sciences to describe many types of systems such as predator-prey models (Diz-Pita & Otero-Espinar, 2021), viral infection models (Liao *et al.*, 2020; Locke *et al.*, 2021) or molecular models (Fallon & Lauffenburger, 2000). In this section, largely inspired by Ref. Allen (2007), I provide some basic definitions and results on (systems of) differential equations that will be used in this thesis.

2.1.1 Ordinary differential equations

Let us first recall the definition of an ordinary differential equation.

Definition 1 (Ordinary differential equation)**.** An ordinary differential equation

of order n (or n^{th} order differential equation) is an equation of the form

$$f(x, \frac{dx}{dt}, \frac{d^2x}{dt^2}, \ldots, \frac{d^nx}{dt^n}, t) = 0,$$

where x is a function of t (which represents time in this thesis) and is n times differentiable. If this differential equation does not depend explicitly on t, *i.e.*, we have

$$f(x, \frac{dx}{dt}, \frac{d^2x}{dt^2}, \ldots, \frac{d^nx}{dt^n}) = 0,$$

then it is said to be *autonomous*; otherwise it is *non-autonomous*.

If an n^{th} order differential equation can be expressed as follows

$$\frac{d^nx}{dt^n} + a_1(t)\frac{d^{n-1}x}{dt^{n-1}} + \ldots + a_{n-1}(t)\frac{dx}{dt} + a_n(t)x = g(t), \tag{2.1}$$

where the coefficients $a_i(t)$, $i = 1, \ldots, n$, and $g(t)$ are not functions of x (nor any derivative of x), then it is referred to as *linear*. Otherwise it is called *non-linear*. If for all t, $g(t) = 0$, then differential equation (2.1) is called *homogeneous*, and *non-homogeneous* otherwise. Note that the coefficients $a_i(t)$, $i = 1, \ldots, n$, and $g(t)$ are always supposed to be real valued.

First-order differential equations

In this thesis, we will consider first-order linear differential equations, which have the following form:

$$\frac{dx}{dt} + a(t)x = g(t). \tag{2.2}$$

The functions $a(t)$ and $g(t)$ are assumed to be continuous. This equation can be solved making use of the *method of the variation of parameters*. We first solve the homogeneous equation

$$\frac{dx}{dt} + a(t)x = 0,$$

to obtain the general solution of the differential equation: for each time t,

$$x(t) = Ke^{-A(t)}, \tag{2.3}$$

where K is a constant and $A'(t) = a(t)$. To obtain the solution of (2.2), one now assumes that K is a function of t and substitutes the general solution in equation (2.2), which yields an expression for the derivative of $K(t)$, $K'(t)$:

$$K'(t) = g(t)e^{A(t)}.$$

After integration, one finds an expression for the function $K(t)$,

$$K(t) = \int g(u)e^{A(u)}du + c,$$

where c is a constant of integration. The expression for $K(t)$ replaces the constant K in equation (2.3). Suppose the initial condition is $x(t_0) = x_0$. Making use of this initial condition, one can determine the constant of integration, c, and obtain the solution of (2.2). That is, for each time t, we have

$$x(t) = e^{-\int_{t_0}^{t} a(s)ds}\left(x_0 + \int_{t_0}^{t} e^{\int_{t_0}^{u} a(s)ds} g(u)du\right). \tag{2.4}$$

Second-order differential equations and Wronskian

Homogeneous second-order linear differential equations will also be found in this thesis. They are of the following form:

$$\frac{d^2x}{dt^2} + a_1(t)\frac{dx}{dt} + a_2(t)x = 0. \tag{2.5}$$

Often, this ODE is completed by a set of two initial conditions, for instance $x(t_0) = x_0$ and $\frac{dx}{dt}(t_1) = x_1$. Assume that $a_1(t)$ and $a_2(t)$ are continuous functions of t. In general, the explicit analytic solution of (2.5) cannot be found. However, if one knows a solution of (2.5) (one may find an obvious solution), then the Wronskian can be used to find all the solutions of the system.

Definition 2 (Wronskian)**.** The Wronskian of two univariate differentiable functions $u(t)$ and $v(t)$ is defined as follows. For all t,

$$W(u, v)(t) = u(t)v'(t) - u'(t)v(t),$$

where $u'(t)$ and $v'(t)$ are the derivative of $u(t)$ and $v(t)$, respectively.

Suppose y_1 and y_2 are solutions of (2.5). Then, their Wronskian obeys the first-order differential equation:

$$\begin{aligned} W'(y_1, y_2)(t) &= y_1'(t)y_2'(y) + y_1(t)y_2''(t) - y_1''(t)y_2(t) - y_1'(t)y_2'(t) \\ &= y_1(t)\left[-a_1(t)y_2'(t) - a_2(t)y_2(t)\right] - \left[-a_1(t)y_1'(t) - a_2(t)y_1(t)\right]y_2(t) \\ &= -a_1(t)W(y_1, y_2)(t), \end{aligned}$$

where, the second step was obtained by substituting the second derivatives by their expression from equation (2.5). This ordinary linear, autonomous and homogeneous differential equation can be solved to obtain a general expression for the Wronskian:

$$W(y_1, y_2)(t) = Ke^{A(t)},$$

where $A'(t) = -a_1(t)$ and K is a constant. Now, assume that we know one of the solutions of the differential equation (2.5), say y_2. Then, following the definition of the Wronskian, any other solution of (2.5), y_1, satisfies the following first-order differential equation:

$$y_1'(t) - \frac{y_2'(t)}{y_2(t)}y_1(t) = -\frac{W(y_1, y_2)(t)}{y_2(t)}.$$

For this method to work, we need $W(y_1, y_2)(t) \neq 0$ for all t, *i.e.*, the Wronskian never vanishes. In that case the functions y_1 and y_2 are linearly independent. However, if the Wronskian vanishes, and under the conditions that the functions y_1 and y_2 are analytic[1](which is always the case in this thesis), then y_1 and y_2 are linearly dependent (Bocher, 1901) and there exists a constant c such that $y_1 = cy_2$.

[1]A univariate real-valued function, $f(x)$, is said to be *analytic* if it is locally given by a convergent power series. That is, $f(x)$ is analytic on an open set D if for any given $x_0 \in D$, one can write $f(x) = \sum_{n=0}^{+\infty} \alpha_n (x - x_0)^n$, where α_0, $\alpha_1, \ldots$ are real numbers and the series is convergent (*i.e.*, $\lim_{N \to +\infty} \sum_{n=0}^{N} \alpha_n (x - x_0)^n < +\infty$) for x in the neighbourhood of x_0.

In some specific cases, an exact solution of the second-order differential equation (2.5) can be found. For instance, if $a_1(t)$ and $a_2(t)$ are constant (written a_1 and a_2), then the number of roots of the associated characteristic polynomial, $r^2 + a_1 r + a_2$, determines the shape of the solution:

- if the characteristic polynomial has two real distinct roots, r_1 and r_2, the solution of (2.5) is
$$x(t) = Ae^{r_1 t} + Be^{r_2 t}.$$
This expression can also be rewritten in term of sum of cosh and sinh.

- If the characteristic polynomial has a repeated real root, r_0, the solution of (2.5) is
$$x(t) = (At + B)e^{r_0 t}.$$

- If the characteristic polynomial has two complex conjugate roots, $r_1 = a + ib$ and $r_2 = a - ib$, then the solution of (2.5) is
$$x(t) = e^{at}(A\cos(bt) + B\sin(bt)).$$

In the three cases, A and B are constants to be determined with the initial conditions.

2.1.2 First-order systems of ordinary differential equations

Most of the time, more than one variable is considered in mathematical models. For instance in an infectious disease model, one could consider the number of infected patients, the number of patients which have never been infected and the number of recovered people (SIR model in Allen (2007)). In that case, we consider a system of differential equations.

Definition 3 (First-order system of ordinary differential equations)**.** A first-order system of ordinary differential equations satisfies
$$\frac{dX}{dt} = F(X(t), t),$$

where the vectors $X(t) = (x_1(t), \ldots, x_n(t))^T$, $F = (f_1, \ldots, f_n)^T$ and each $f_i \equiv f_i(x_1(t), \ldots, x_n(t), t)$.

The system is said to be *autonomous* if F does not depend explicitly on t. Otherwise, it is said to be *non-autonomous*. A system defined as in Definition 3 is said to be *linear* if each ODE of the system can be expressed as

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{i,j}(t)x_j + b_i(t),$$

for $i = 1, \ldots, n$. Otherwise it is said to be *non-linear*. Non-linear systems of ODEs are very common in biology due to the interaction between species. A common example found in this thesis is the case in which two species x_1 and x_2 bind to form a third species x_3. Then, the ODE which determines the evolution of the concentration of each of this species will have a term proportional to $x_1(t)x_2(t)$ under the mass action law[1]. Finally, if the system is linear and $b_i(t) = 0$ for all i and at any t, then the system is said to be *homogeneous*, and *non-homogeneous* otherwise.

Steady state

One important type of solution for an autonomous system of differential equations is the constant solution, X^* , known as the steady state solution, which satisfies

$$F(X^*) = 0.$$

This solution is obtained by setting all the derivatives (left side) to 0 in a system of ODEs and solving the resulting system of equations. A steady state can be biologically interpreted as a state in which the amount of each species considered in the model is not changing anymore. Some models may have several steady states. We call *non-trivial* steady state a steady state X^* which is not the null vector.

[1]More details will follow in Section 2.3.

Stability of the steady state

An important question about steady states is whether they are *locally stable*, *asymptotically stable* or *unstable*. A steady state is locally stable if a solution starting close to the steady state solution, X^*, remains close to X^* as $t \to +\infty$. A steady state is asymptotically stable if it is locally stable and if a solution which starts close to the steady state solution, X^*, approaches X^* as $t \to +\infty$. A steady state that meets neither of these conditions is said to be unstable. In this thesis, we will study the stability of steady states of systems of ODEs by computing the eigenvalues of the Jacobian matrix associated with the system.

Definition 4 (Jacobian matrix)**.** Consider a system of ODEs as defined in Definition 3. Then the associated Jacobian matrix is

$$\mathcal{J}(X) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(X) & \dots & \frac{\partial f_1}{\partial x_n}(X) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(X) & \dots & \frac{\partial f_n}{\partial x_n}(X) \end{pmatrix}$$

A steady state solution, X^*, is asymptotically stable if and only if all of the eigenvalues of the Jacobian matrix evaluated at this steady state, $\mathcal{J}(X^*)$, have a real negative part. If one or more eigenvalues have a positive real part, the steady state is unstable. If any eigenvalue has a real part equal to zero, the stability analysis is said to be inconclusive.

2.2 An introduction to Gröbner bases

In this thesis, I make use of Gröbner bases (Buchberger, 1965, 2006) to solve systems of polynomial equations. In this section, inspired by Cox *et al.* (2013), I give an overview of the important definitions related to the polynomial ring and introduce the Gröbner bases formally. I then explain how they are employed in this thesis to solve polynomial systems.

2.2.1 The polynomial ring

Gröbner bases are defined in the polynomial ring which is commutative[1]. First, let us recall the definition of a commutative ring.

Definition 5 (Commutative ring)**.** A *commutative ring* consists of a set R and two binary operations $+$ and $\cdot$ defined on R such that

1. $(a+b)+c = a+(b+c)$ and $(a\cdot b)\cdot c = a\cdot(b\cdot c)$ for all $a, b, c \in R$ (associativity).
2. $a + b = b + a$ and $a \cdot b = b \cdot a$ for all $a, b \in R$ (commutativity).
3. $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in R$ (distributivity).
4. There are $0, 1 \in R$ such that $a + 0 = a \cdot 1 = a$ for all $a \in R$ (identities).
5. Given $a \in R$, there exists $b \in R$ such that $a + b = 0$ (additive inverses).

Now, let us describe the polynomial ring in more detail. We first recall the general definition of a polynomial, starting by introducing the notion of monomials.

Definition 6 (Monomial)**.** A *monomial* in $x_1, \ldots, x_n$ is a product of the form $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$, where the exponents α_i, $i = 1, \ldots, n$, are non negative integers. The *degree* of a monomial is $\alpha_1 + \ldots + \alpha_n$.

For a less cumbersome notation, the abbreviation $\underline{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$, where $\alpha = (\alpha_1, \ldots, \alpha_n)$ will be used. We let $|\alpha| = \alpha_1 + \ldots + \alpha_n$ be the degree of the monomial $\underline{x}^\alpha$. We are now ready to formally define polynomials of n variables.

Definition 7 (Polynomial)**.** A polynomial f in $x_1, \ldots, x_n$ with coefficients in a field $\mathbb{K}$ (typically $\mathbb{Q}$ or $\mathbb{R}$ in this thesis), is a finite linear combination of monomials. A polynomial f is written in the form

$$f = \sum_\alpha b_\alpha \underline{x}^\alpha, \quad b_\alpha \in \mathbb{K},$$

where the sum is over a finite number of n-tuples $\alpha = (\alpha_1, \ldots, \alpha_n)$. We call b_α the *coefficient* of monomial $\underline{x}^\alpha$. If $b_\alpha \neq 0$, then $b_\alpha \underline{x}^\alpha$ is called a *term* of f. The *degree* of polynomial $f \neq 0$ is the maximum $|\alpha|$ such that $b_\alpha \neq 0$. The degree of the zero polynomial, $f = 0$, is undefined.

[1]Note that the notion Gröbner bases have been generalised to non-commutative rings but it is out of the scope of this thesis.

It can be shown that the set of all polynomials of n variables with coefficients in $\mathbb{K}$, with addition and multiplication, satisfies the axioms of a commutative ring. This ring is called *the polynomial ring* and denoted by $\mathbb{K}[x_1, \ldots, x_n]$. Now that the polynomial ring is defined, let us introduce an ordering relationship on this set. We first recall the definition of a total ordering.

Definition 8 (Total ordering)**.** A total ordering is a relation $\leq$ on some set X, which satisfies the following for all $a, b, c \in X$:

- $a \leq a$ (reflexivity),
- if $a \leq b$ and $b \leq$c then $a \leq c$ (transitivity),
- if $a \leq b$ and $b \leq a$ then $a = b$ (anti-symmetry),
- $a \leq b$ or $b \leq a$ (strongly connected or total).

With this in place, we can now introduce one of the most important notions of this section: monomial ordering.

Definition 9 (Monomial ordering (or monomial order))**.** A *monomial ordering* on $\mathbb{K}[x_1, \ldots, x_n]$ is a relation $\geq$ on the set of monomials $\underline{x}^\alpha$, $\alpha \in \mathbb{N}^n$, such that :

1. $\geq$ is a total ordering on $\mathbb{N}^n$,
2. For all $\alpha, \beta, \gamma \in \mathbb{N}^n$, if $\alpha \geq \beta$, then $\alpha + \gamma \geq \beta + \gamma$
3. $\geq$ is a well-ordering on $\mathbb{N}^n$ (this means that every non-empty subset of $\mathbb{N}^n$ has a smallest element with respect to $\geq$).

One first example of an ordering on n-tuples, which is a monomial ordering, is the lexicographic order.

Definition 10 (Lexicographic (lex) order)**.** Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$ be in $\mathbb{N}^n$. We say $\alpha \geq_{lex} \beta$ if the leftmost nonzero entry of the vector $\alpha - \beta \in \mathbb{N}^n$ is positive, *i.e.*, if $\alpha - \beta$ is in the form $(0, \ldots, 0, a_1, \ldots, a_r)$ with $a_1 \geq 0$. With such a monomial ordering on $\mathbb{K}[x_1, \ldots, x_n]$, we have $\underline{x}^\alpha \geq_{lex} \underline{x}^\beta$.

Consider the variables $x_1, \ldots, x_n$ of any polynomial of $\mathbb{K}[x_1, \ldots, x_n]$. Then, in term of the set of monomials $\underline{x}^\alpha$ defined earlier, we have $x_1 = \underline{x}^{(1,0,\ldots,0)}$, $x_2 =$

$\underline{x}^{(0,1,\ldots,0)}$, *etc.*... Thus, the variables $x_1, \ldots, x_n$ are ordered in the usual way in the lex ordering:

$$(1, 0, \ldots, 0) \geq_{lex} (0, 1, 0, \ldots, 0) \geq_{lex} \ldots \geq_{lex} (0, \ldots, 0, 1),$$

so $x_1 \geq_{lex} x_2 \geq_{lex} \ldots \geq_{lex} x_n$. Note that for a general case of n variables, there exist $n!$ lex orders [1].

Example 11 (Lex order (example inspired by Cox *et al.* (2013)))**.** Consider the multivariate polynomial, $Q = 4xy^2z + 4z^2 - 5x^3 + 7x^2z^2$. Let us choose a lex monomial order such that $x \geq_{lex} y \geq_{lex} z$. In term of the monomial notation $\underline{x}^{\alpha=(\alpha_1,\alpha_2,\alpha_3)} = x^{\alpha_1}y^{\alpha_2}z^{\alpha_3}$, we have $xy^2z = \underline{x}^{(1,2,1)}$, $z^2 = \underline{x}^{(0,0,2)}$, $x^3 = \underline{x}^{(3,0,0)}$ and $x^2z^2 = \underline{x}^{(2,0,2)}$. We have $(3, 0, 0) \geq_{lex} (2, 0, 2) \geq_{lex} (1, 2, 1) \geq_{lex} (0, 0, 2)$, so the monomials of Q are ordered as follows: $x^3 \geq_{lex} x^2z^2 \geq_{lex} xy^2z \geq_{lex} z^2$.

We might want to take the degree of the monomials into account and order monomials of bigger degree first. One way to do this is the graded lexicographic order (or grlex).

Definition 12 (Graded lex order)**.** Let $\alpha, \beta \in \mathbb{N}^n$. We say that $\alpha \geq_{grlex} \beta$ if $|\alpha| \geq |\beta|$ or, $|\alpha| = |\beta|$ and $\alpha \geq_{lex} \beta$.

Note that the choice of a lex order is essential to define a grlex order.

Example 13 (Grlex order (example inspired by Cox *et al.* (2013)))**.** Consider again the multivariate polynomial, $Q = 4xy^2z + 4z^2 - 5x^3 + 7x^2z^2$. Let us choose a grlex monomial order such that $x \geq_{lex} y \geq_{lex} z$. The monomials of Q are ordered by decreasing degree. To order the monomials, xy^2z and x^2z^2, which have the same degree, we use the chosen lex order. We have $x^2z^2 \geq_{lex} xy^2z$. Thus, under such grlex order, the monomials of Q are ordered as follows: $x^2z^2 \geq_{grlex} xy^2z \geq_{grlex} x^3 \geq_{grlex} z^2$.

Other orderings, such as graded reverse lex order, exist but we do not introduce them in this thesis. A monomial ordering allows one to define the leading monomial,

[1] There is one lex order to each way the n variables can be ordered. There exist $n!$ way of ordering n elements. For example, consider a set $\{a, b, c\}$ of three elements, then there are $3! = 6$ lexicographic orders on this set: $\{a \leq b \leq c, a \leq c \leq b, b \leq a \leq c, b \leq c \leq a, c \leq a \leq b, c \leq b \leq a\}$.

the leading coefficient and the leading term of any polynomial of $\mathbb{K}[x_1, \ldots, x_n]$. Let us first define the multidegree of a polynomial of $\mathbb{K}[x_1, \ldots, x_n]$.

Definition 14 (Multidegree)**.** Let $f = \sum_\alpha b_\alpha \underline{x}^\alpha$ be a non-zero polynomial of $\mathbb{K}[x_1, \ldots, x_n]$ and let $\geq$ be a monomial order. Then, the *multidegree* of f is $multideg(f) = \max \left\{ \alpha \in \mathbb{N}^n \middle| b_\alpha \neq 0 \right\}$, where the maximum is taken with respect to the ordering $\geq$.

With this multidegree, we can define the following three notions:

Definition 15 (Leading coefficient)**.** Let $f = \sum_\alpha b_\alpha \underline{x}^\alpha$ be a non-zero polynomial of $\mathbb{K}[x_1, \ldots, x_n]$ and let $\geq$ be a monomial order. Then, the *leading coefficient* of f is $LC(f) = b_{multideg(f)}$.

Definition 16 (Leading monomial)**.** Let $f = \sum_\alpha b_\alpha \underline{x}^\alpha$ be a non-zero polynomial of $\mathbb{K}[x_1, \ldots, x_n]$ and let $\geq$ be a monomial order.Then, the *leading monomial* of f is $LM(f) = \underline{x}^{multideg(f)}$.

Definition 17 (Leading term)**.** Let $f = \sum_\alpha a_\alpha \underline{x}^\alpha$ be a non-zero polynomial of $\mathbb{K}[x_1, \ldots, x_n]$ and let $\geq$ be a monomial order. Then, the *leading term* of f is $LT(f) = LC(f) \cdot LM(f)$.

The determination of the leading term, coefficient and monomial are dependent on the choice of the monomial ordering.

Example 18 (Leading term of a multivariate polynomial depends on monomial ordering)**.** When choosing the grlex monomial ordering of Example 13, the leading term of polynomial $Q = 4xy^2z + 4z^2 - 5x^3 + 7x^2z^2$ is $7x^2z^2$, where 7 is the leading coefficient and x^2z^2 the leading monomial. If we choose the lex monomial ordering of Example 11, then the leading term of Q is $-5x^3$.

The determination of the leading term is a crucial step in the division and reduction algorithms (Cox *et al.*, 2013). Since it is highly dependent on the choice of monomial order, such ordering must be chosen carefully. A "bad" choice may complicate the computation and make the algorithm less efficient.

Now that we recalled these important definitions, we are equipped to introduce Gröbner bases.

2.2.2 Definition of Gröbner bases

A Gröbner basis is a particular kind of generating set of an ideal in a polynomial ring. We first recall the definition of an ideal.

Definition 19 (Ideal)**.** Let $(R, +, \cdot)$ be a commutative ring. A subset $I \subseteq R$ is an *ideal* if it satisfies:

1. $0 \in I$.
2. If $a, b \in I$, then $a + b \in I$.
3. If $a \in I$ and $b \in R$, then $b \cdot a \in I$.

In other words, an ideal is a subset of a commutative ring containing zero and closed under internal addition and external multiplication. Note that ideals can be generated by a set of elements of the ring.

Definition 20 (Ideal generated by a set)**.** Let $(R, +, \cdot)$ be a commutative ring and $f_1, \ldots, f_s \in R$. Then

$$I = \langle f_1, \ldots, f_s \rangle = \left\{ \sum_{i=1}^{s} h_i f_i \middle| h_1, \ldots, h_s \in R \right\}$$

is an ideal called the *ideal generated by* $\{f_1, \ldots, f_s\}$. We also say that $\{f_1, \ldots, f_s\}$ is a *generating set* of I, and for $i = 1, \ldots, s$, f_i is a *generator* of I.

Finally, let us introduce the notion of an ideal of leading terms.

Definition 21 (Ideal of leading terms)**.** Let $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ be an ideal other than $\{0\}$, and fix a monomial ordering on $\mathbb{K}[x_1, \ldots, x_n]$. Then we denote by $LT(I)$ the set of leading terms (with respect to the given monomial order) of non-zero elements of I:

$$LT(I) = \left\{ LT(f) \middle| f \in I \setminus \{0\} \right\}.$$

We denote by $\langle LT(I) \rangle$ the ideal generated by the elements of $LT(I)$. The ideal $\langle LT(I) \rangle$ is called the *ideal of leading terms*.

Observe that $\langle LT(I) \rangle$ is impractical to build as we need to consider all the polynomials of I and take their leading term. If I is generated by a finite set

$\{f_1, \ldots, f_s\}$, we would rather consider the ideal generated by the leading terms of the generators, $\langle LT(f_1), \ldots, LT(f_s)\rangle$. However, $\langle LT(I)\rangle$ and $\langle LT(f_1), \ldots, LT(f_s)\rangle$ may be different ideals (but we do have $\langle LT(f_1), \ldots, LT(f_s)\rangle \subseteq \langle LT(I)\rangle$ (Cox *et al.*, 2013)).

Example 22 (Case for which $\langle LT(f_1), \ldots, LT(f_s)\rangle \neq \langle LT(I)\rangle$. Example taken from Cox *et al.* (2013))**.** Consider $I = \langle f_1, f_2\rangle = \langle x^3 - 2xy, x^2y - 2y^2 + x\rangle$ and use the grlex ordering on monomials in $\mathbb{K}[x, y]$. We have

$$x^2 = xf_2 - yf_1,$$

so $x^2 \in I$. Thus, $LT(x^2) = x^2$ is a linear combination of the leading terms of the elements of I and we have $x^2 \in \langle LT(I)\rangle$. However, $LT(f_1) = x^3$ and $LT(f_2) = x^2y$. As x^2 cannot be expressed as a linear combination of x^3 and x^2y, we have $x^2 \notin \langle LT(f_1), LT(f_2)\rangle$.

It is however possible to find a generating set $\{f_1, \ldots, f_s\}$ of I such that $\langle LT(f_1), \ldots, LT(f_s)\rangle = \langle LT(I)\rangle$. We first have the following result:

Theorem 23 (Hilbert basis theorem)**.** Every ideal I of the polynomial ring is finitely generated. In other words, there exists a finite subset $G = \{g_1, \ldots, g_s\} \subseteq I$ such that $I = \langle g_1, \ldots, g_s\rangle$.

The proof of this theorem (Cox *et al.*, 2013) establishes that we can always find a generating set $\{g_1, \ldots, g_s\}$ of an ideal I such that $\langle LT(g_1), \ldots, LT(g_s)\rangle = \langle LT(I)\rangle$. This special generating set is called a Gröbner basis.

Definition 24 (Gröbner basis)**.** Fix a monomial order on the polynomial ring $\mathbb{K}[x_1, \ldots, x_n]$. A finite subset $G = \{g_1, \ldots, g_s\}$ of an ideal $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$, $G \neq \{0\}$, is said to be a *Gröbner basis* if

$$\langle LT(g_1), \ldots, LT(g_s)\rangle = \langle LT(I)\rangle.$$

As a convention, the empty set $\emptyset$ is the Gröbner basis of the zero ideal $\{0\}$.

More informally, a set $G = \{g_1, \ldots, g_s\} \subseteq I$ is a Gröbner basis of I if and only if the leading term of any element of I is divisible by one of the leading terms of the elements of G, $LT(g_i)$. For a given monomial order, every ideal, I, of the

polynomial ring $\mathbb{K}[x_1, \ldots, x_n]$ has a Gröbner basis and this is a generating set of I. For a given monomial order, an ideal $I \neq \{0\}$ can have multiple Gröbner bases, but has only one *reduced Gröbner basis*.

Definition 25 (Reduced Gröbner basis)**.** A *reduced Gröbner basis* for a polynomial ideal I is a Gröbner basis, G, of I such that

1. $\forall g \in G,\ LC(g) = 1$,
2. $\forall g \in G$, no monomial of g is divisible by the leading term of the other elements of G.

Note that, in this thesis, if we fix a lex monomial ordering and compute a Gröbner basis, we will say that we compute a *lex Gröbner basis*. Equipped with these formal definitions, one can apply Gröbner bases to solve systems of polynomial equations.

2.2.3 Use of Gröbner bases

Gröbner basis computation is a practical tool for solving systems of polynomial equations. Consider a set of n polynomials, $P_1, \ldots, P_n$ of $\mathbb{K}[x_1, \ldots, x_n]$. One wants to find the solutions of the system $P_1 = 0, \ldots, P_n = 0$, *i.e.*, find the n-tuples $(x_1^*, \ldots, x_n^*)$ such that we have

$$\begin{aligned} P_1(x_1^*, \ldots, x_n^*) &= 0, \\ &\vdots \\ P_n(x_1^*, \ldots, x_n^*) &= 0. \end{aligned} \tag{2.6}$$

The set of n-tuples which satisfies system (2.6), is called an affine variety.

Definition 26 (Affine variety)**.** Let $f_1, \ldots, f_s$ be polynomials of $\mathbb{K}[x_1, \ldots, x_n]$. Then we call an *affine variety* the set of points (n-tuples) at which all the polynomials $f_1, \ldots, f_s$ vanish:

$$V(f_1, \ldots, f_s) = \left\{ (x_1^*, \ldots, x_n^*) \in \overline{\mathbb{K}}^n \middle| f_i(x_1^*, \ldots, x_n^*) = 0, \quad \forall i \in \{1, \ldots, s\} \right\}.$$

We denote by $\overline{\mathbb{K}}$ the *algebraic closure* of $\mathbb{K}$. That is, the roots of every non-constant polynomial with coefficients in $\mathbb{K}$ lie in $\overline{\mathbb{K}}$. We note that $\mathbb{K} \subseteq \overline{\mathbb{K}}$.

Example 27 (Algebraic closure of $\mathbb{R}$)**.** Every non-constant polynomial of $\mathbb{R}[x]$ admits a root in the complex field $\mathbb{C}$. Thus, the algebraic closure of $\mathbb{R}$ is $\overline{\mathbb{R}} = \mathbb{C}$.

Such a definition of an affine variety shows that we are studying the common complex roots of a given system of polynomial equations. In this thesis, we focus on real roots, in particular positive roots[1].

When system (2.6) has a finite number of solutions, the associated variety is a finite set of points. We call such a variety *zero-dimensional*.

The problem of solving the system (2.6) is equivalent to the problem of finding all the points of the associated affine variety $V(P_1, \ldots, P_n)$. Now, $I = \langle P_1, \ldots, P_n \rangle$ defines an ideal of $\mathbb{K}[x_1, \ldots, x_n]$. We can define a variety associated to this ideal.

Definition 28 (Variety associated to an ideal)**.** Let $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ be an ideal. We denote by $V(I)$ the following set

$$V(I) = \left\{ (x_1^*, \ldots, x_n^*) \in \overline{\mathbb{K}}^n \middle| f(x_1^*, \ldots, x_n^*) = 0 \ \forall f \in I \right\}.$$

We have the following result (Cox *et al.*, 2013).

Proposition 29. $V(I)$ is an affine variety. In particular, if $I = \langle f_1, \ldots, f_s \rangle$, then $V(I) = V(f_1, \ldots, f_s)$.

It results that if the system (2.6) has a finite number of solutions, $V(I)$ is finite and we have the following definition.

Definition 30 (Zero-dimensional ideal)**.** Let I be an ideal of $\mathbb{K}[x_1, \ldots, x_n]$. We say that I is *zero-dimensional* if the associated variety, $V(I)$ is a finite set.

It follows from Proposition 29 that $V(I)$ can be described by any generating set of I. Let $G = (g_1, \ldots, g_s)$ be a Gröbner basis of I. Then we have $V(P_1, \ldots, P_n) = V(I) = V(g_1, \ldots, g_s)$. As a result, the system of polynomial equations associated to G, $g_1 = 0, \ldots, g_s = 0$ has the same roots as the system (2.6).

[1] The problem of finding the positive real roots of a system of polynomial equations is usually a hard problem.

Now suppose that $I = \langle P_1, \ldots, P_n \rangle$ is a zero-dimensional ideal (or equivalently that system (2.6) admits a finite set of solutions). Note that all the polynomial systems considered in this thesis will admit a finite set of solutions. Choosing a lex monomial order, a Gröbner basis, $G = \{g_1, \ldots, g_s\}$, of I, has a special form:

$$\begin{aligned} &g_1(x_1, \ldots, x_n), \\ &g_2(x_1, \ldots, x_{n-1}), \\ &\vdots \\ &g_s(x_1). \end{aligned}$$

That is, the associated system of polynomial equations $(g_1 = 0, \ldots, g_s = 0)$ is a *triangular system*. Thus, it can be solved iteratively, starting by solving $g_s(x_1) = 0$ with one variable solving methods, then solving $g_{s-1}(x_1, x_2) = 0$ substituting x_1 with a similar procedure, *etc.* These roots are also the roots of the system (2.6).

The computation of Gröbner bases is implemented in most computer algebra software or libraries, such as Mathematica (Wolfram Research, Inc., 2019) or Macaulay2 (Grayson & Stillman) which will be used in this thesis. They make use of modified and improved versions of the first algorithm proposed to compute Gröbner bases, the Buchberger algorithm (Buchberger, 1965, 2006). Informally speaking, this algorithm is a multi-dimensional generalisation of the Gaussian elimination and the Euclidean algorithm. The computation of Gröbner bases can be computationally complex, making its practical use to solve polynomial systems limited. However, in this thesis, many successful cases are highlighted.

2.3 A brief introduction to chemical reaction network theory

In this section, I briefly summarise the relevant notions of chemical reaction network theory (CRNT). In this thesis, a chemical reaction network (CRN), $\mathcal{N}$, is viewed as a multi-set $\mathcal{N} = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, where $\mathcal{S}$ is the set of species, $\mathcal{C}$ the set of

complexes, and $\mathcal{R}$ the set of reactions. We note that in the context of a CRN, a "complex" is a linear combination of species and need not be a "biological functional unit", which we refer to as a *biological complex*. We denote, whenever useful, a biological complex formed by species X and Y as $X : Y$, where the colon denotes the physical bond between X and Y. The order of species in the biological complex is irrelevant, *i.e.,* $X : Y = Y : X$.

Example 31 (Heterodimeric receptor tyrosine kinase)**.** Consider the heterodimeric receptor tyrosine kinase (RTK). This receptor is composed of two different single chains, X_1 and X_2, which dimerise to form the full receptor (biological complex) $Y_1 = X_1 : X_2$. The ligand, L, then binds to this receptor to form the biological complex, $Y_2 = L : Y_1$, which will induce a signal. A simple heterodimeric receptor tyrosine kinase (RTK) model has a species set $\mathcal{S} = \{X_1, X_2, Y_1, Y_2\}$, a complex set $\mathcal{C} = \{X_1 + X_2, Y_1, Y_2\}$, and a reaction set $\mathcal{R} = \{X_1 + X_2 \to Y_1, Y_1 \to X_1 + X_2, Y_1 \to Y_2, Y_2 \to Y_1\}$. In this thesis the ligand concentration (L) is taken to be an input parameter (*i.e.*, L is constant over time) and, hence, it does not feature as a separate chemical species in the species set $\mathcal{S}$. I want the reader to notice how the set of complexes, $\mathcal{C}$, in a CRN, is not solely composed of biological complexes, and that the set of species, $\mathcal{S}$, includes all the biological complexes and not only the single chain components (receptor chains that are not bound to any complex, such as X_1 and X_2).

We can associate a reaction graph to every CRN $\mathcal{N}$, by letting the vertex set be $\mathcal{C}$ and the (directed) edge set $\mathcal{R}$. There exists a class of important CRNs defined by their network reversibility.

Definition 32 (Network reversibility)**.** Let $\mathcal{N}$ be a CRN with its associated reaction graph $\mathcal{G}(\mathcal{C}, \mathcal{R})$. An edge between C_i and $C_j \in \mathcal{C}$ exists if $C_i \to C_j \in \mathcal{R}$. If for every edge $C_i \to C_j \in \mathcal{R}$, the edge $C_j \to C_i \in \mathcal{R}$ also exists, then the network is called *reversible*. If for every edge, $C_i \to C_j \in \mathcal{R}$, a directed path exists going back from C_j to C_i, then the network is called *weakly reversible*. All reversible networks are also weakly reversible.

A general reaction from complex C_i to complex C_j can be written as

$$C_i = \sum_{k=1}^{n} \alpha_{ik}\, X_k \to \sum_{k=1}^{n} \alpha_{jk}\, X_k = C_j, \tag{2.7}$$

where the sum is over the set of n species $(X_1, X_2, \ldots, X_n)$. The elements of non-negative integer vectors, $\alpha_i = (\alpha_{i1}, ..., \alpha_{in})^T$ and $\alpha_j = (\alpha_{j1}, ..., \alpha_{jn})^T$, are the *stoichiometric coefficients*. One can define the *reaction vector*, given by $r = \alpha_j - \alpha_i$. For a CRN with n species and m reactions we can now define the $n \times m$ matrix of all reaction vectors, Γ, such that $\Gamma = (r_1, \ldots, r_m)$. This matrix is called the *stoichiometric matrix*.

Example 33 (Heterodimeric RTK continued)**.** The heterodimeric RTK model has $m = 4$ reactions and $n = 4$ species. Its reaction graph is given by

$$X_1 + X_2 \rightleftharpoons Y_1 \rightleftharpoons Y_2.$$

The model is reversible and its reaction vectors are $r_1 = (-1, -1, 1, 0)^T$, $r_2 = (1, 1, -1, 0)^T$, $r_3 = (0, 0, -1, 1)^T$ and $r_4 = (0, 0, 1, -1)^T$. The stoichiometric matrix is

$$\Gamma = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

To derive dynamical properties from the static description so far provided, we make use of the law of mass action kinetics (Horn & Jackson, 1972). First, we assign a *rate constant*, $k_j \in \mathbb{R}^+$, with $j \in \{1, \ldots, m\}$, to each reaction in the network. Second, we denote the concentration of species X_i by the time-dependent function $x_i(t)$. With this notation, we associate to every complex $C_i = \sum_{k=1}^{n} \alpha_{ik}\ X_k$, a monomial (product of the concentrations of the species X_k, $k = 1, \ldots, n$, each concentration elevated to the power of their associated stoichiometric coefficient) as follows

$$\underline{x}^{\alpha_i} \equiv x_1(t)^{\alpha_{i1}} \ldots x_n(t)^{\alpha_{in}}, \tag{2.8}$$

where n is the number of species in the network. Note that, in equation (2.7) the complex C_i (complex on the left-hand side) is called the *reactant complex*. The *reaction rate* of a reaction is the associated concentration product of its reactant complex multiplied by the rate constant. That is, for a reaction $C_i \to C_j$, if we write k_i its reaction constant and $\underline{x}_i^{\alpha}$ the monomial associated to C_i, the reaction

rate is $k_i \underline{x}_i^{\alpha}$. Let us write, $x(t) = (x_1(t), ..., x_n(t))^T$, the vector of the concentration of species at time t. We also define the *flux vector*, $\mathcal{R}(x)$, the $m \times 1$ column vector of all reaction rates. The ordinary differential equations (ODEs) governing the dynamics of the reaction network are given by

$$\frac{dx}{dt} = \Gamma \mathcal{R}(x), \tag{2.9}$$

where Γ is the stoichiometric matrix (defined above). We note that the reaction rate of the i^{th} reaction is the i^{th} row in $\mathcal{R}(x)$.

From (2.9) we can also deduce the conserved quantities of the reaction network. That is, if a column vector exists, $z \in \mathbb{Z}^n$, such that $d(z^T x)/dt = z^T \Gamma \mathcal{R}(x) = 0$, the quantity $z^T x$ is conserved. Consequently, the set of such vectors z defines a basis for the space of conserved quantities. In this way, conservations induce linear relations between the variables. Informally, we say that a molecular species, X_i, is conserved if its total number of molecules, which we write N_i, is constant. N_i is determined by the initial conditions.

Example 34 (Heterodimeric RTK continued)**.** The dynamical system associated with the heterodimeric RTK model is given by

$$\begin{aligned}\frac{dx}{dt} = \frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} &= \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} k_1 x_1 x_2 \\ q_1 y_1 \\ k_2 y_1 \\ q_2 y_2 \end{pmatrix} \\ &= \begin{pmatrix} -k_1 x_1 x_2 + q_1 y_1 \\ -k_1 x_1 x_2 + q_1 y_1 \\ k_1 x_1 x_2 - (q_1 + k_2) y_1 + q_2 y_2 \\ k_2 y_1 - q_2 y_2 \end{pmatrix},\end{aligned}$$

with k_i as the reaction constants of the forward chemical reactions ($\rightharpoonup$) and q_i the reaction constants of the backward reactions ($\leftharpoondown$). One can note that vectors $z_1 = (1, 0, 1, 1)^T$ and $z_2 = (0, 1, 1, 1)^T$ satisfy $d(z_1^T x)/dt = 0$ and $d(z_2^T x)/dt = 0$. That is, a basis for the conservation equations is given by the linear relations, valid for any time t, $x_1(t) + y_1(t) + y_2(t) = N_1$ and $x_2(t) + y_1(t) + y_2(t) = N_2$, where N_1 (resp. N_2) is the total number of X_1 (resp. X_2) in the system. These

relations imply that the total amount of the species X_1 (resp. X_2) is conserved by adding the amounts of the bound states of the molecule (Y_1 and Y_2, respectively) to the amount of free molecule X_1 (resp. X_2).

We can now define the biologically meaningful steady states of a CRN.

Definition 35. A vector x^* is a *biologically meaningful steady state* if $\Gamma R(x^*) = 0$ and $x_i^* \geq 0 \; \forall i \in \{1, \ldots, n\}$. Note that for this steady state not to be trivial, we want at least one of the x_i^* to be strictly positive. We will call *strictly positive biological meaningful steady states*, biological meaningful steady states for which $x_i^* > 0 \; \forall i \in \{1, \ldots, n\}$.

A useful connection between the static network structure (defined earlier) and the existence (and stability) of unique biologically meaningful steady states can be made via deficiency theory (Feinberg, 1987).

Definition 36 (Deficiency)**.** Let $\mathcal{N}$ be a CRN with ℓ connected components in the reaction graph[1] and $\eta = \dim \operatorname{span}(r_1, \ldots, r_m)$ be the dimension of the span of the reaction vectors. The deficiency of $\mathcal{N}$ is then given by

$$\delta = |\mathcal{C}| - \ell - \eta.$$

The notion of network deficiency leads to one of the fundamental theorems of CRNT, the *Deficiency Zero Theorem* (Feinberg, 1987), which connects the network structure to the dynamics of a CRN.

Theorem 37 (Deficiency zero theorem (Feinberg, 1987))**.** Let $\mathcal{N}$ be a weakly reversible CRN with $\delta = 0$. Then the network has a unique strictly positive biologically meaningful steady state for every set of initial conditions, and this steady state is asymptotically stable.

With certain additional conditions on the reaction rates (see Refs. Dickenstein & Pérez Millán (2011); Feinberg (1989)) that will be true in this thesis, biologically meaningful steady states are *detailed balanced*. This means that for every reaction

[1] Given a graph $\mathcal{G}$, two vertices of $\mathcal{G}$ belong to the same connected component if there exists a (undirected) path between these two vertices. If all the vertices of $\mathcal{G}$ are connected (that is, there exists a path between any pair of vertices of $\mathcal{G}$) then $\mathcal{G}$ has a unique connected component. For examples, the graph $A \to B \leftarrow C$ has a unique connected component. The graph $A \to B \leftarrow C \quad D \to E$ has two connected components.

of the form (2.7), the steady states satisfy

$$K(\underline{x}^*)^{\alpha_i} = (\underline{x}^*)^{\alpha_j}, \tag{2.10}$$

where the number $K = k/q$, the ratio of the rate constants of the forward and backward reactions, is called the *affinity constant* of the reaction. We write $(\underline{x}^*)^{\alpha_i}$ the monomial associated to complex C_i in which the concentration of the species are at steady state: $(\underline{x}^*)^{\alpha_i} = (x_1^*)^{\alpha_{i1}} \dots (x_n^*)^{\alpha_{in}}$.

Substituting the detailed-balanced steady state equations into the conservation relations yields a system of polynomials equations. This polynomial system will be one of the main interest in Chapters 3 and 4. In these chapters, as we always study the systems at steady state, we will omit the * to simplify the notation. As we focus on the study of systems of polynomial equations, we can use the techniques developed in the field of computational algebra and algebraic geometry (Cox *et al.*, 2013). Such methods have also been successfully applied to many topics in chemical reaction network theory, see *e.g.,* Refs. Dickenstein *et al.* (2019); Gross *et al.* (2016); Sadeghimanesh & Feliu (2019). In this thesis, we will solve, when possible, such polynomial systems to obtain exact expressions for the steady state concentrations of any species of the CRN, making use of Gröbner bases (Buchberger, 1965, 2006).

Example 38 (Heterodimeric RTK continued)**.** The heterodimeric RTK model has 3 complexes, 1 connected component and the dimension of the span of the reaction vectors is 2; hence, $\delta = 3 - 1 - 2 = 0$. Since the network is reversible, we know from Theorem 37 that there exists exactly one stable positive steady state for each set of initial conditions. By finding the steady state of system (2.9), one can show that we have indeed the following detailed-balanced steady state $y_1^* = (k_1/q_1)x_1^*x_2^*$ and $y_2^* = (k_2/q_2)y_1^*$. The combination of the conservation equations at steady state, $x_1^* + y_1^* + y_2^* = N_1$ and $x_2^* + y_1^* + y_2^* = N_2$, and the detailed-balanced steady state equations gives a system of polynomial equations. The variables are the concentration of the single chain components at steady state,

x_1^* and x_2^*:

$$\begin{aligned} 0 &= -N_1 + x_1^* + \frac{k_1}{q_1} x_1^* x_2^* + \frac{k_2}{q_2}\frac{k_1}{q_1} x_1^* x_2^*, \\ 0 &= -N_2 + x_2^* + \frac{k_1}{q_1} x_1^* x_2^* + \frac{k_2}{q_2}\frac{k_1}{q_1} x_1^* x_2^*. \end{aligned} \tag{2.11}$$

Note that, as the biological complex Y_2 is actually formed by association of the ligand L and the biological complex Y_1, the constant k_2 is proportional to the concentration of ligand L. Consequently, I will write $(\bar{k}_2/q_2)L$ instead of (k_2/q_2) in polynomials system (2.11), where $\bar{k}_2 = k_2/L$ is the reaction constant of $L + Y_1 \to Y_2$ (this is equivalent to the notation of k_2 as the reaction constant of $Y_1 \to Y_2$).

2.4 Signalling function: amplitude and EC_{50}

In Chapters 3 and 4, we will want to closely investigate pharmacological properties of receptor-ligand systems, rather than the steady state structure of the models. In particular, we want to study the dose-response (or concentration-effect) curve of the system, which describes the relation between ligand concentration and the biological effect (or cellular response) it generates when binding its specific cell surface receptor. As mentioned in Chapter 1, a lot of effort has been devoted to explore the steady state structure of chemical reaction networks. In this thesis, we make use of algebraic methods to explore the dose-response of receptor-ligand systems. To do so, we start with the definition of signalling complex. We note that in most biological instances the signalling complex is formed by all the subunit chains that make up the full receptor, intra-cellular kinases and the specific ligand (Cotari *et al.*, 2013b; Dushek *et al.*, 2011; Feinerman *et al.*, 2010; Gonnord *et al.*, 2018; Janes & Lauffenburger, 2013; Leonard *et al.*, 2019; Uings & Farrow, 2000).

Definition 39. The signalling complex of a receptor-ligand system is the biological complex which induces a biological response.

We are now equipped to define the signalling function and the dose-response curve.

Definition 40. We define the *signalling function*, $\sigma : \mathbb{R}^+ \to \mathbb{R}^+$, $L \mapsto \sigma(L)$, as the univariate function which assigns to a given value of ligand concentration, L, the number (or concentration) of signalling complexes formed at steady state. The dose-response curve is the corresponding plot of the signalling function.

We note that in this thesis we will not distinguish between number (or concentration) of signalling complexes since one can be obtained from the other if we know the volume of the system and Avogadro's number.

The specific choice of σ will depend on the receptor-ligand system under consideration. The signalling function associated to each model studied in this thesis will be defined in the corresponding section when needed. In the models proposed in this thesis, the signalling function will be a product of the steady state values (numbers) of sub-unit chains that make up the full receptor, intra-cellular kinases, affinity constants of the reactions involved, and ligand concentration. This together with equations (2.8) and (2.9) indicate that the signalling function will always be algebraic[1].

Example 41 (Heterodimeric RTK continued from Section 2.3)**.** In the heterodimeric RTK model, the biological complex which induces a signal is Y_2. That is, the signalling function is given by $\sigma(L) \equiv y_2^*$. The detailed-balance steady state gives $y_2^* = K_2 K_1 L x_1^* x_2^*$, where we wrote $K_1 = \frac{k_1}{q_1}$, $K_2 = \frac{\bar{k}_2}{q_2}$. Hence, the signalling function is a product of the concentration (at steady state) of the single chains components that form the signalling complex and the affinity constants of the reactions involved in its formation. Solving the system of polynomial equations described in the previous example, one can obtain expressions for x_1^* and x_2^* and thus an expression for σ.

Next, we define a central object of study in this thesis; namely, the amplitude of the signalling function, often referred as efficacy in the pharmacology literature (Maxwell & Webb, 2008).

Definition 42. The *amplitude* of the signalling function, A, is the difference between the maximum and the minimum of σ; that is, $A \equiv max(\sigma) - min(\sigma)$.

[1] An algebraic function is a function that can be defined as the root of a polynomial equation. The product of algebraic functions is an algebraic function. A formal definition is recalled in Chapter 4 (Def. 78).

We note that when $\min(\sigma) = 0$, which is the case considered in this thesis ($\min(\sigma) = \sigma(0) = 0$), the amplitude is given by the maximum of the signalling function. If, in addition, the dose-response curve attains its maximum at large concentrations (for instance, when the dose-response curve is a sigmoid), we have

$$A = \lim_{L \to +\infty} \sigma(L). \tag{2.12}$$

The amplitude provides information about the magnitude of the intra-cellular response to the stimulus, L. The larger the amplitude is, the larger the response variability will be. The amplitude is always bounded by the number of molecules available. However, this bound is often not tight (Pérez Millán & Dickenstein, 2015). To quantify the sensitivity of the model to the stimulus, *i.e.,* the potency of the ligand L, we introduce the *half-maximal effective concentration*, EC_{50}.

Definition 43. The *half-maximal effective concentration*, or EC_{50}, is the ligand concentration L^* which satisfies $\sigma(L^*) = min(\sigma) + \frac{max(\sigma) - min(\sigma)}{2} = min(\sigma) + \frac{A}{2}$ (when σ is increasing).

We say that the EC_{50} is inversely proportional to ligand potency; namely, the lower the EC_{50}, the higher the potency (or sensitivity) of the ligand. Figure 2.1 illustrates the amplitude and the EC_{50} of a sigmoid dose-response curve (A) when its minimum is zero: increasing the amplitude shifts up the maximum of the curve and results in greater efficacy (B), and decreasing the EC_{50} shifts the dose-response curve to the left and increases the potency of the ligand (C).

The amplitude and the EC_{50} are generally inferred from the fitting of a dose-response curve to an equation. When the dose-response curve is a sigmoid, it is generally fitted to a Hill equation:

Definition 44 (Hill equation)**.** The *Hill equation* is defined as follows:

$$f_{\text{Hill}} : L \mapsto A \frac{L^n}{(L_{50})^n + L^n},$$

where n is called the *Hill coefficient*, A is the amplitude of the curve and L_{50} is the EC_{50}.

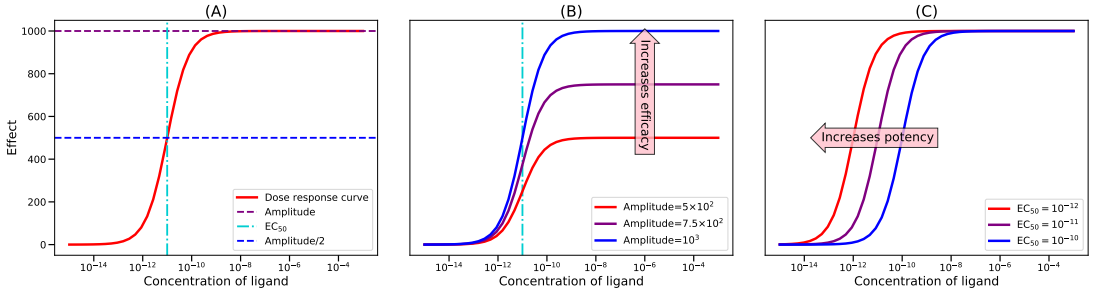


Figure 2.1: Sigmoid dose-response curve: number of signalling complexes formed at steady state, $\sigma(L)$, as a function of the concentration of ligand L (arbitrary units). (A) The maximum value of $\sigma(L)$ defines the amplitude, since the minimum of the curve is 0. The EC_{50} is the concentration of ligand which corresponds to half the amplitude. (B) Three dose-response curves with the same EC_{50} value and different amplitudes. Increasing the amplitude shifts up the maximum of the curve and increases the efficacy of the ligand. (C) Three dose-response curves with the same amplitude and different EC_{50} values. Decreasing the EC_{50} shifts the dose-response curve to the left and increases the potency of the ligand.

The parameters n, L_{50} and A are inferred from the fitting process. In this thesis, we prefer to fit a sigmoidal dose-response curve to a sigmoid function

$$f_{\text{sigmoid}} : L \mapsto \frac{A}{1 + e^{-n(L-L_{50})}},$$

where A is the amplitude, n a slope coefficient and L_{50} the EC_{50}. Note that the Hill equation and the sigmoid equation are mathematically equivalent (as $L^n = e^{n \log L}$).

2.5 Perturbation theory

A well-known difficulty with the lex Gröbner basis method to solve polynomial systems, and polynomial equations in general, is that there is usually no analytic solution when the degree of a univariate polynomial is greater than four. This result is known as the Abel-Ruffini theorem (Żołądek, 2000). Therefore, in order to make progress, we need to resort to either numerical computations or analytic approximations. Since receptor-ligand systems are often characterised by a sigmoidal dose-response curve, at least to calculate the amplitude, the only

quantity of interest is the limit of the signalling function at infinity. In order to calculate this limit (where possible analytically, otherwise numerically) we make use of perturbation theory for polynomial equations.

Greatly inspired by the Dover book written by Simmonds and Mann (Simmonds & Mann, 2013), this section reviews some notions of perturbation theory and justifies the steps of the method used to compute the analytic amplitude expression in Sections 3.4.1 and 4.2.6. We start by defining an asymptotic expansion.

Definition 45 (Asymptotic expansion)**.** Let f be a univariate real-valued function and $\epsilon, a_0, \ldots, a_n \in \mathbb{R}$. We say that

$$\sum_{n=0}^{N} a_n \, f_n(\epsilon),$$

is an *asymptotic expansion* of f in ϵ if:

- the family of functions $\{f_n\}_{n=0,\ldots,N+1}$ is a *gauge sequence*, *i.e.*, $f_n(\epsilon) = o(f_{n-1}(\epsilon))$ as $\epsilon \to 0$, for $n = 0, \ldots, N+1$, and
- $f(\epsilon) - \sum_{n=0}^{N} a_n \, f_n(\epsilon) = \mathcal{O}(\epsilon^{N+1})$ as $\epsilon \to 0$.

The numbers $a_0, \ldots, a_N$ are called the *coefficients* of the asymptotic expansion.

Note that, here, we used the Landau notation:

Definition 46 (Little o notation)**.** Let f and g be two real-values univariate functions and $x_0 \in \mathbb{R}$. We write $f(x) = o(g(x))$ as $x \to x_0$ if and only if

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

We supposed that $g(x) \neq 0$ for x in the neighbourhood of x_0.

Definition 47 (Big O notation)**.** Let f and g be two real-values univariate functions and $x_0 \in \mathbb{R}$. We write $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$ if and only if

$$\exists \, d, C > 0, \forall x, |x - x_0| < d \Rightarrow |f(x)| \leq C|g(x)|.$$

.

That is, an asymptotic expansion is a formal series [1] of functions which provides an approximation of a given function (as the argument of the function tend to a specific point) when truncated. The core of perturbation theory is the notion of asymptotic expansion and the following fundamental theorem.

Theorem 48 (Fundamental theorem of perturbation theory)**.** Consider an asymptotic expansion of a function f in ϵ with coefficients $a_0, \ldots, a_N$. If it is a power series which satisfies

$$a_0 + a_1\epsilon + a_2\epsilon^2 + \ldots + a_N\epsilon^N + \mathcal{O}(\epsilon^{N+1}) = 0,$$

for any sufficiently small ϵ, and the coefficients a_i are independent of ϵ, then we have

$$a_0 = a_1 = \ldots = a_N = 0.$$

Note that if an asymptotic expansion is a power series, we call it a *regular expansion*. If the expansion has non-integer powers, then it is called a *singular expansion*. Theorem 48 only applies to regular expansions.

We are now ready to study the behaviour of the root of a univariate polynomial. Let $n \in \mathbb{N}^*$. We consider a univariate polynomial, $P_\epsilon(x)$, of degree n, in the variable x, with coefficients which depend on the parameter ϵ, and suppose that such polynomial can be re-written in the following form

$$\begin{aligned} P_\epsilon(x) = (1 + b_0\epsilon + c_0\epsilon^2 + \ldots) + A_1\epsilon^{\nu_1}(1 + b_1\epsilon + c_1\epsilon^2 + \ldots)x + \ldots \\ + A_n\epsilon^{\nu_n}(1 + b_n\epsilon + c_n\epsilon^2 + \ldots)x^n, \end{aligned} \tag{2.13}$$

where for each i, ν_i is a rational number, A_i, b_i, c_i, ... are real constants, $(1 + b_i\epsilon + \ldots)$ is a regular asymptotic expansion of the general form

$$a_0 + a_1\epsilon + \ldots + a_N\epsilon^N + \mathcal{O}(\epsilon^{N+1}).$$

We are interested in the behaviour of the roots of $P_\epsilon(x)$ when $\epsilon \to 0$. That is,

[1] A formal series is an infinite sum considered independently from any notion of convergence.

we want to determine the coefficients of the asymptotic expansion of the roots of $P_\epsilon(x)$, as $\epsilon \to 0$. Note that given a polynomial, we cannot always find regular expansions of its roots. Instead, we are looking for singular expansions first. We will "make" each expansion regular by a change of variable so we can apply Theorem 48 and find the coefficients of the expansion. For such a polynomial, $P_\epsilon(x)$, we have the following result.

Theorem 49. Each root of a polynomial (2.13) is of the form

$$x(\epsilon) = \epsilon^p \omega(\epsilon), \quad \omega(0) \neq 0 \text{ (not identically 0)}, \tag{2.14}$$

where ω is a continuous function of ϵ for ϵ sufficiently small and $p \in \mathbb{Q}$.

The proof of this theorem (see Ref. Simmonds & Mann (2013)) gives a method to study the asymptotic behaviour of the roots of polynomial (2.13).

Method. We first aim to determine what we call the *proper values* of $p \in \mathbb{Q}$ (as defined in Theorem 49). Let $P_\epsilon(x)$ be a polynomial that can be written as in equation (2.13). Let p be a rational number and x a root of $P_\epsilon(x)$. Let us replace x by $\epsilon^p\omega(\epsilon)$ in $P_\epsilon(x)$. We can re-write the polynomial as follows

$$\begin{aligned} P_\epsilon(\epsilon^p\omega(\epsilon)) = Q_\epsilon(\omega) &+ \epsilon(b_0 + b_1 A_1 \epsilon^{\nu_1+p}\omega(\epsilon) + \ldots + b_n A_n \epsilon^{\nu_n+np}\omega(\epsilon)^n) \\ &+ \epsilon^2(c_0 + c_1 A_1 \epsilon^{\nu_1+p}\omega(\epsilon) + \ldots + c_n A_n \epsilon^{\nu_n+np}\omega(\epsilon)^n) + \ldots, \end{aligned} \tag{2.15}$$

where

$$Q_\epsilon(\omega) = 1 + A_1\omega(\epsilon)\epsilon^{\nu_1+p} + \ldots A_n\omega(\epsilon)^n\epsilon^{\nu_n+np}.$$

Thus, $Q_\epsilon(\omega)$ collects all the terms with the lowest power of ϵ for each power of $\omega(\epsilon)$. Now, if we have to satisfy $\omega(0) \neq 0$, we have the following result:

Proposition 50. At least two of the exponents in set $E = \{0, \nu_1 + p, \ldots, \nu_n + np\}$ must have identical minimal values.

Proof. Suppose that $\nu_k + kp$ is the unique smallest exponent of E. Then we have

$$\epsilon^{-(\nu_k+kp)} P_\epsilon(\epsilon^p\omega(\epsilon)) \underset{\epsilon\to 0}{\sim} A_k\omega(0)^k.$$

Since $\omega(0) \neq 0$ and $P_\epsilon(\epsilon^p \omega(\epsilon)) = 0$ by hypothesis, this implies that $A_k = 0$. That is, there is no term in $\epsilon^{\nu_k + kp}$ in $Q_\epsilon(\omega)$ and $\nu_k + kp \notin E$, which is a contradiction with the initial hypothesis. So E has more than one minimal value. □

We now proceed by finding all the values of p and their associated minimal exponent to form the set $\{(p_j, e_j)\}_{j=1,\dots,m}$. To this end, we follow a graphical algorithm which indicates when two or more components of E have equal minimal values:

1. On a plane (p, q), draw the lines $q = \nu_j + jp$, $j = 1, \dots, n$ and the line $q = 0$.

2. From the right, for p sufficiently large, the smallest exponent is 0. As p decreases (one can imagine a fictive vertical line moving from right to left), there will be a first point where at least two lines intersect ($q = 0$ and another one). Let us call this point $(p_1, 0)$. One line will have the largest slope, n_1.

3. Let the fictive vertical line keep moving to the left and follow this line of slope n_1 until the next intersection (p_2, e_2). Find the new intersected line with the largest slope n_2.

4. Continue until there is no other intersection. The last intersection involves the line with the largest slope of all the lines n.

We apply this method on two examples and illustrate the algorithm in Sections 3.4 and 4.2.6. This algorithm finds all the intersection points of the lines of equation $q = \nu_j + jp$, $j = 0, \dots, n$ and $q = 0$ that are on the lower envelope of these lines. In this way, we have generated a set of pairs $\{(p_j, e_j)\}_{j=1,\dots,m}$ corresponding to each intersection we encountered. Each of these intersection points is called a *branch* of the asymptotic behaviour of the roots of our original polynomial $P_\epsilon(x)$. Now let us define for each branch j, the scaled polynomial $T_\epsilon^{(j)}$, as follows:

$$T_\epsilon^{(j)}(\omega(\epsilon)) = \epsilon^{-e_j} P_\epsilon(\epsilon^{p_j} \omega(\epsilon)). \tag{2.16}$$

We can re-write $T_\epsilon^{(j)}(\omega(\epsilon))$ as a sum of two polynomials

$$T_\epsilon^{(j)}(\omega(\epsilon)) = T_0^{(j)}(\omega(\epsilon)) + E_\epsilon^{(j)}(\omega(\epsilon)),$$

where $E_0^{(j)}(\omega(\epsilon)) = 0$ and the coefficients of $T_0^{(j)}(\omega(\epsilon))$ do not depend on ϵ explicitly. By multiplying P_ϵ by ϵ^{-e_j}, we extracted the dominant part of $P_\epsilon(\epsilon^{p_j}\omega(\epsilon))$, *i.e.*, the term that will determine the behaviour of $P_\epsilon(\epsilon^{p_j}\omega(\epsilon))$ as $\epsilon \to 0$. The solutions of $T_0^{(j)}(\omega(\epsilon)) = 0$, for each j, will give the values of $\omega(\epsilon)$ which, when substituted into $x = \epsilon^{p_j}\omega(\epsilon)$, will give the n roots of $P_\epsilon(x)$. Now, to apply Theorem 48, and find the coefficients of the asymptotic expansion of the roots of $P_\epsilon(x)$, the non-zero roots of $T_\epsilon^{(j)}(\omega(\epsilon))$ (approached by the roots of $T_0^{(j)}(\omega(\epsilon))$ as $\epsilon \to 0$) need to be regular (*i.e.*, can be approached by a regular expansion) but this is not necessarily the case. Indeed, ν_j or (p_j, e_j) may be non-integer rationals or $T_0^{(j)}(\omega(\epsilon))$ may have repeated roots. To obtain regular expansions, we introduce the new variable β such that:

$$\epsilon = \beta^{q_j}, \tag{2.17}$$

where q_j is the least common denominator (lcd) of the set of exponents $\{0, \nu_1 + p_j, \ldots, \nu_n + np_j\}$. Finally, we form:

$$R_\beta^{(j)}(\omega(\beta)) = T_j(\omega(\beta), \beta^{q_j}) = \beta^{-q_j e_j} P(\beta^{q_j p_j}\omega(\beta), \beta^{q_j}), \tag{2.18}$$

where $T_j(\omega(\epsilon), \epsilon) = T_\epsilon^{(j)}(\omega(\epsilon))$ and $P(\omega(\epsilon), \epsilon) = P_\epsilon(\omega(\epsilon))$. The polynomial $R_\beta^{(j)}(\omega(\beta))$ has the same roots as the polynomial $T_\epsilon^{(j)}(\omega(\epsilon))$ but its non-zero roots have a regular expansion in β of the form

$$\omega(\beta) = a_0 + a_1\beta + \ldots + a_N\beta^N + \mathcal{O}(\beta^{N+1}).$$

By substituting this expansion into $R_\beta^{(j)}(\omega)$ and applying the fundamental theorem of perturbation theory (Theorem 48), we find an expression for $a_0, a_1, \ldots$. We then come back to x with the transformation $x = \beta^{q_j p_j}\omega(\beta)$ for each branch. In practice we explore each branch one by one and can eliminate those which are irrelevant (for instance when we have a negative root, since in our case the roots of the polynomials are concentrations of species, or $\omega(0) = 0$).

The above discussion can be summarised algorithmically as follows.

1. Replace the variable x by $\epsilon^p\omega(\epsilon)$ in $P_\epsilon(x)$, assuming $\omega(0) \neq 0$. One obtains a polynomial of the form

$$P_\epsilon(\epsilon^p\omega(\epsilon)) = Q_\epsilon(\omega) + \epsilon(\ldots) + \ldots.$$

2. Write the set of exponents for Q_ϵ: $E = \{0, \nu_1 + p, \ldots, \nu_n + np\}$.

3. Determine the pairs, (p_j, e_j), of proper values and minimal exponents following the graphical algorithm described above. Each pair corresponds to an asymptotic branch to explore.

4. For each branch j:
   - 4.1. Define $T_\epsilon^{(j)}(\omega) = \epsilon^{-e_j} P_\epsilon(\epsilon^{p_j}\omega)$.
   - 4.2. Introduce β such that $\epsilon = \beta_j^q$, where $q_j = \text{lcd}(0, \nu_1+p, \ldots, \nu_n+np)$, and form $R_\beta^{(j)}(\omega) = T_{\beta_j^q}^{(j)}(\omega)$.
   - 4.3. In $R_\beta^{(j)}(\omega) = 0$, substitute ω by a regular expansion $\omega(\beta) = a_0 + a_1\beta + \ldots + a_N\beta^N + \mathcal{O}(\beta^{N+1})$.
   - 4.4. Apply the fundamental theorem of perturbation theory to obtain an analytic expression for $a_0, a_1, \ldots$. Usually at this step, we can discriminate whether this branch is relevant (see example 3.4).
   - 4.5. Find the asymptotic expansion of the root of the original polynomial, P_ϵ, by $x = \beta^{q_j p_j}\omega(\beta)$.

In this thesis we are mainly interested in the first non-zero coefficient of the regular expansion of ω since it drives the behaviour of the root of P_ϵ in the limit $\epsilon \to 0$.

2.6 Estimation of the number of positive real roots of a polynomial

It is often useful to estimate the number of positive or negative real roots of a polynomial. For instance, the number of biologically meaningful steady states of a system of ODEs may be related to the number of positive real roots of a polynomial. For another instance, the stability of a particular steady state of a system depends on whether the characteristic polynomial of the Jacobian matrix of the system (evaluated at the steady state) has positive real roots.

2.6.1 Descartes' rule and Budan's theorem

The Descartes' rule of signs (Curtiss, 1918) provides a theoretical upper bound on the number of positive or negative real root of a given polynomial.

Theorem 51 (Descartes' rule of signs)**.** A univariate polynomial $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ with real coefficients cannot have more positive real roots than the sequence $a_0, \ldots, a_n$ has variations of sign.

This rule has been later refined to state that the number of variations of sign is either equal to the number of positive real roots, or else exceeds it by an even number. The maximum number of negative real roots can be obtained with the same rule by considering $f(-x)$.

Example 52. Consider the univariate polynomial equation $f(x) = x^3 - 5x^2 + 3x + 9$. The signs of the coefficients (ordered by decreasing monomial degrees) are $(+, -, +, +)$. There are two sign changes thus polynomial $f(x)$ admits 2 or 0 positive real roots. We have $f(-x) = -x^3 - 5x^2 - 3x + 9$: the signs of the coefficients of $f(-x)$ are $(-, -, -, +)$. There is only one sign variation which means that polynomial $f(x)$ has one real negative roots. Polynomial $f(x)$ can be factorised as $f(x) = (x-3)^2(x+1)$ which shows that $f(x)$ has 2 positive roots and one negative root.

Sometimes, we might prefer to estimate the number of positive real roots of a polynomial in an interval different from $[0, +\infty)$. To this end, we use Budan's theorem (Akritas, 1982), which is a generalisation of Descartes' rule.

Theorem 53 (Budan's theorem)**.** Consider a univariate polynomial $f(x)$ with real coefficients. Consider a real number, h and write V_h the number of sign variations of the sequence of the coefficients of polynomial $f_h(x) \equiv f(x+h)$. Finally, let us define a half-open interval $[p, q)$, with $p < q \in \mathbb{R}$ and write $\#f_{p,q}$ the number of roots of f in this interval, with their multiplicity. Then, $V_p - V_q - \#f_{p,q}$ is a non-negative even integer.

Example 54. Consider the univariate polynomial $f(x) = x^3 - 2x^2 - 5x + 6$, and let us find the number of roots of this polynomial in interval $(0, 2)$. We have $V_0 = 2$ and as $f(x+2) = x^3 + 4x^2 - x - 4$, $V_2 = 1$. Hence, Budan's theorem asserts that f has a unique root ($V_0 - V_2 = 1$) in interval $(0, 2)$. Note that this

polynomial can be factorised as $(x-1)(x+2)(x-3)$, which confirms the result.

2.6.2 Routh–Hurwitz criterion

The stability analysis of a steady state of a system of ODEs usually consists of analysing the characteristic polynomial of the Jacobian matrix of the system, evaluated at said steady state. Descartes' rule is a useful theorem to define a necessary condition of stability: if the polynomial characteristic has all its coefficients of the same sign, then the system may be stable. On the contrary, if there is at least one sign variation, then the system is unstable. The Routh–Hurwitz criterion (Anagnost & Desoer, 1991) defines a sufficient condition for the stability of a steady state when the polynomial characteristic of the associated Jacobian matrix has positive coefficients. Let us first define the Routh table of a polynomial.

Definition 55 (Routh table)**.** Let $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ be a univariate polynomial of degree n, where a_i, $i = 0, \ldots, n$ are real strictly positive coefficients. The Routh table of $f(x)$ is defined (when possible, *i.e.*, none of the denominators is 0) as follows:

a_0	a_2	a_4	$\ldots$
a_1	a_3	a_5	$\ldots$
$b_1 = \dfrac{a_1a_2 - a_3a_0}{a_1}$	$b_2 = \dfrac{a_1a_4 - a_5a_0}{a_1}$	$b_3 = \dfrac{a_1a_6 - a_7a_0}{a_1}$	$\ldots$
$c_1 = \dfrac{b_1a_3 - b_2a_1}{b_1}$	$c_2 = \dfrac{b_1a_5 - b_3a_1}{b_1}$	$\ldots$	$\ldots$

With this in place, we can define the Routh–Hurwitz criterion.

Theorem 56 (Routh–Hurwitz criterion)**.** Let $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ be a polynomial of degree n, where a_i, $i = 0, \ldots, n$ are real strictly positive coefficients. If the elements of the first columns of the Routh table of $f(x)$ are all of the same sign (and not 0), then all the zeros of polynomial $f(x)$ are in the left half-plane (*i.e*, have a negative real part). In that case, we say that polynomial $f(x)$ satisfies the Routh–Hurwitz criterion.

Let us compute the Jacobian matrix, $\mathcal{J}(X)$, of a system of ODEs and evaluate it at a steady state, X^*. If the characteristic polynomial of the Jacobian matrix $\mathcal{J}(X^*)$

satisfies the Routh–Hurwitz criterion, then the steady state, X^* is asymptotically stable.

2.7 Stochastic processes

In deterministic models, the random fluctuations (such as the difference in receptor expression levels between cells of a population) from one individual to another are ignored. That is, all individuals are assumed to be identical. One way to take these fluctuations into account is to use stochastic models. In this section, inspired by Allen (2010), I introduce the main tools from probability theory that are used in this thesis. I also introduce some useful results on two specific stochastic processes: the pure death process and the simple birth and death process.

2.7.1 Probability theory

Here, I give an overview of some basic definitions of probability theory and describe the probability distributions used in this thesis.

Random variable, mean, variance and probability functions.

First, let us recall the definition of a sample space and random variable.

Definition 57 (Sample space)**.** The *sample space* of a given random experiment is the set of all the possible outcomes of this experiment.

Note that in this section, we make use of the usual notations in probability theory and ω now denotes an element of the sample space.

Definition 58 (Random variable)**.** A *random variable* X is a real-valued function defined on a sampled space Ω,

$$X : \Omega \mapsto \mathbb{R}.$$

A random variable X is said to be a *discrete random variable* if its support $S_X = \left\{ x \in \mathbb{R} \middle| X(\omega) = x \text{ for } \omega \in \Omega \right\}$ is finite (or countably finite). If S_X is uncountably infinite, it is said that X is a *continuous random variable*.

Example 59 (Discrete random variable)**.** Suppose we toss a single coin. The outcomes of this experiment are head (H) or tail (T). Thus, the sample space is $\Omega = \{H, T\}$. A discrete random variable associated with this experiment could return 1 if head is obtained and 0 otherwise.

In this thesis, we will write $\mathbb{P}(A)$ the probability that the event A happens. For instance in the previous example, the probability that we obtain tail when tossing the coin will be denoted $\mathbb{P}(T)$, or, as a function of the random variable X defined above, $\mathbb{P}(X = 0)$. Let us now introduce useful functions used in probability theory. We start with the cumulative distribution function.

Definition 60. (c.d.f.) The *cumulative distribution function* (c.d.f.) of a random variable X is the function $F : \mathbb{R} \mapsto [0, 1]$ defined by

$$F(x) = \mathbb{P}(X \leq x).$$

It can be shown that F is non-decreasing, right continuous and satisfies

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to +\infty} F(x) = 1.$$

The cumulative distribution function describes how probabilities accumulate.

For a discrete random variable, we define the probability mass function.

Definition 61 (p.m.f.)**.** Suppose X is a discrete random variable with c.d.f. F. The *probability mass function* (p.m.f) is defined as

$$f(x) = \mathbb{P}(X = x).$$

The p.m.f. and the c.d.f are linked through the following relationship:

$$F(x) = \sum_{a_i < x} f(a_i)$$

where $\{a_i\}_i$ is a collection of elements of the support of X, S_X, and $F(x) = 0$ for any $x < \inf_i(a_i)$.

The continuous version of the p.m.f. is the probability density function.

Definition 62 (p.d.f.)**.** Consider a continuous random variable, X with c.d.f. F.

If there exists a non-negative integrable function $f : \mathbb{R} \mapsto [0, +\infty)$ such that

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

then this function is called the *probability density function* (p.d.f.) of X. It follows that, if F is differentiable,

$$\frac{dF}{dx}(x) = f(x) \quad \text{for } x \in S_X.$$

Making use of the p.d.f., one can compute the probability associated with the outcome of any event. Suppose A a random event ($A \subseteq \Omega$) and X an associated continuous random variable with p.d.f. f. We write $A_X = \left\{ x \in \mathbb{R} \middle| X(\omega) \text{ for } \omega \in A \right\}$. Then

$$\mathbb{P}(A) = \mathbb{P}(X \in A_X) = \int_{A_X} f(x)dx.$$

Often, we want to compute the expected value or the standard deviation of a random variable. These concepts help characterise the p.d.f. of the random variable.

Definition 63 (Expected value)**.** Suppose X a continuous random variable with p.d.f. f. Its expected value (or expectation), denoted $\mathbb{E}(X)$ is given by

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf(x)dx.$$

Suppose X a discrete random variable with p.m.f. f, defined on the space $S_X = \{a_i\}_{i=1}^{+\infty}$. Then, its expected value is defined as

$$\mathbb{E}(X) = \sum_{i=1}^{+\infty} a_i f(a_i).$$

The expected value of X is a weighted average and corresponds to the mean of the possible values that X can take.

The variance and standard deviation of a random variable are defined as follows:

Definition 64 (standard deviation and variance)**.** Suppose X a random variable with mean $\mu_X = \mathbb{E}(X)$. The variance of X is defined as $Var(X) \equiv \mathbb{E}([X - \mu_X]^2)$. Its standard deviation, $\sigma(X)$ is the square root of the variance.

Note that, in this section, σ denotes the standard deviation, as commonly denoted in probability theory (in this section, σ is not a signalling function).

Some common distributions

I now recall some well-known distributions that are used in this thesis. We start with the continuous uniform distribution which will be used as a non-informative prior in Bayesian inference (see Section 2.8).

Definition 65 (Continuous uniform distribution)**.** A random variable X is said to follow a *continuous uniform distribution* with parameters $-\infty < a < b < +\infty$, and we write $x \sim \mathcal{U}(a, b)$, if its probability density function (see Figure 2.2(a)) is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} .$$

Another very common continuous distribution is the normal distribution.

Definition 66 (Normal distribution)**.** A random variable X is said to follow a *normal distribution* with parameters μ and σ, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$, if its probability density function (see Figure 2.2(b)) is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < +\infty.$$

The parameters μ and σ^2 are the mean and variance of X respectively.

Biological quantities (which are positive quantities) may be normally distributed. However, many measurements show a more skewed distribution, with a low mean and a large variance. Such a distribution often fit a log-normal distribution (Koch, 1966):

Definition 67 (Log-normal distribution)**.** A random variable $X = \log(Y)$ is said to follow a *log-normal distribution* with parameters μ and σ, and we write

$X \sim \log \mathcal{N}(\mu, \sigma^2)$, if its probability density function (see Figure 2.2(c)) is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right), \quad 0 < x < +\infty.$$

The parameters μ and σ^2 are the mean and variance of the normally distributed variable Y. The mean of X is $e^{\mu+\frac{\sigma^2}{2}}$, its variance is $[e^{\sigma^2}-1]e^{2\mu+\sigma^2}$.

I also recall the definition of the Poisson distribution which is used when simulating birth processes in Chapter 5.

Definition 68 (Poisson distribution)**.** A discrete random variable X is said to follow a *Poisson distribution* with parameter λ, and we write $X \sim \mathcal{P}(\lambda)$, if its probability mass function (see Figure 2.2(d)) is

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N}.$$

Parameter λ is also the mean and the variance of X

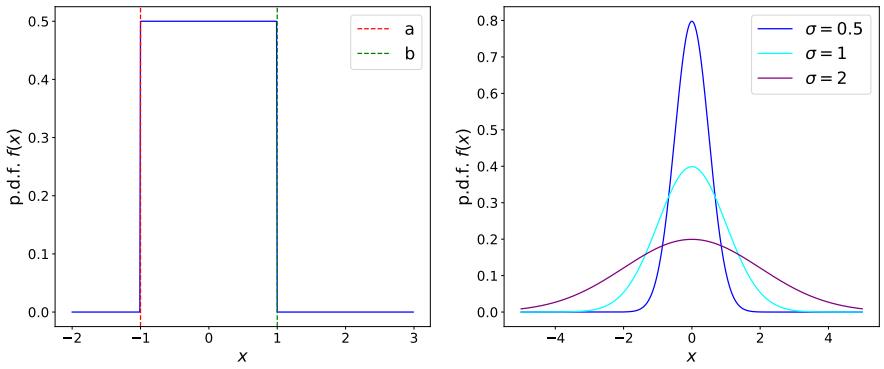
Finally, I give the definition of the Gumbel distribution.

Definition 69 (Gumbel distribution)**.** A discrete random variable X is said to follow a *Gumbel distribution* with parameters m and β, and we write $x \sim Gumbel(m, \beta)$, if its probability density function (see Figure 2.2(e)) is

$$f(x) = \frac{1}{\beta} e^{-(z+e^{-z})}.$$

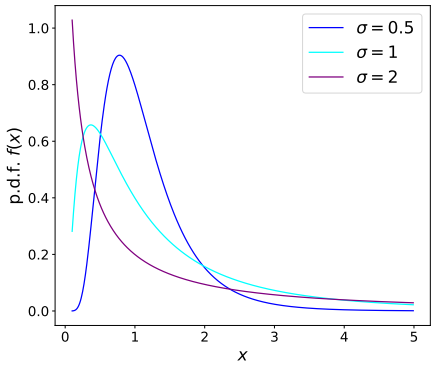
where $z = \dfrac{x-m}{\beta}$. The mean of such a distribution is given by $m + \gamma_e\beta$ (where γ_e is the Euler–Mascheroni constant) and its standard deviation is $\dfrac{\pi\beta}{\sqrt{6}}$.

Note that we call *mode* of a probability distribution the maximum value of the p.d.f. (or p.m.f. for discrete variables). In other words, the mode is the value of the distribution that is the most likely to be sampled.

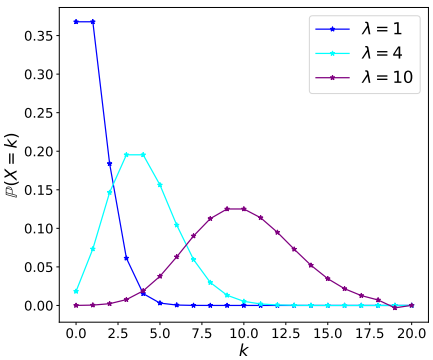


(a) Uniform

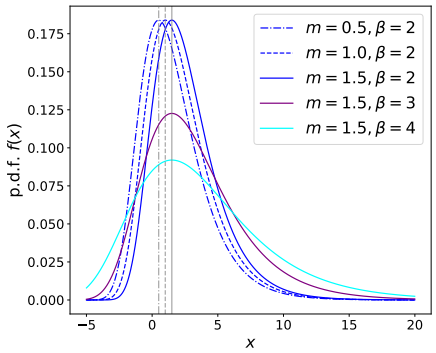
(b) Normal



(c) Log-normal



(d) Poisson



(e) Gumbel

Figure 2.2: P.d.f. (p.m.f. for the Poisson distribution) of the different probability distributions described in this section. (a) Uniform distribution for $a = -1$, $b = 1$. (b) Normal distribution for $\mu = 0$ and different values of σ. (c) Log-normal distribution for $\mu = 0$ and different values of σ. (d) Poisson probability mass function for different values of λ. The function is defined only at integer values of k; the connecting lines are guides for the eye. (e) Gumbel probability density function for different parameter values: fixed β, different m (blue), and fixed m, different β (solid lines). The value of m is also indicated with vertical grey lines (dash-dotted line for $m = 0.5$, dashed line for $m = 1$ and solid line for $m = 1.5$).

2.7.2 Branching process: pure death process

A stochastic process is defined as a collection of random variables. Branching processes are stochastic processes describing a population in which each individual, independently, follows the same set of rules. In Chapter 5, I will make use of some results from the analysis of the continuous-time branching process called *pure death process*.

Consider a population of individuals which have the same death rate μ. That is, any individual alive at time t has probability $\mu\Delta t$ of dying before $t + \Delta t$ as $\Delta t \to 0$. Thus, the probability that an individual survives during this time interval is $1 - \mu\Delta t$. Writing $e^{-\mu\Delta t} = 1 - \mu\Delta t + o(\Delta t)$, the probability that one individual survives to time t is $e^{-\mu t}$.

Kolmogorov equations. Let $N(t)$ denote the number of individuals alive at time t and let $\mathbb{P}(N(t) = n)$ define the probability that exactly n individual survived at time t, $n < N(0) \equiv N_0$. For simplification, we write $p_n(t) = \mathbb{P}(N(t) = n)$ in this section. Then we have:

$$\begin{aligned}
p_{N_0}(t) &= e^{-N_0\mu t}, \\
p_{N_0-1}(t) &= N_0 e^{-(N_0-1)\mu t}(1 - e^{-\mu t}), \\
p_{N_0-2}(t) &= \binom{N_0}{2} e^{-(N_0-2)\mu t}(1 - e^{-\mu t})^2, \\
&\vdots \\
p_{N_0-j}(t) &= \binom{N_0}{j} e^{-(N_0-j)\mu t}(1 - e^{-\mu t})^j, \\
&\vdots \\
p_0(t) &= (1 - e^{-\mu t})^{N_0}.
\end{aligned}$$

Hence, p_n satisfies the *forward Kolmogorov equations*:

$$\begin{aligned}
\frac{dp_n}{dt} &= -n\mu p_n + \mu(n+1)p_{n+1} \quad \text{for } n < N_0, \\
\frac{dp_{N_0}}{dt} &= -\mu N_0 p_{N_0}.
\end{aligned} \tag{2.19}$$

From this equation, we can compute the mean number of individuals at time t:

$$\mathbb{E}(N(t)) = \sum_{n=0}^{+\infty} n p_n(t). \tag{2.20}$$

Indeed,

$$\begin{aligned}
\frac{d\mathbb{E}(N)}{dt} &= \sum_{n=0}^{+\infty} n \frac{p_n}{dt} \\
&= \sum_{n=1}^{N_0} n \frac{p_n}{dt} \\
&= \sum_{n=1}^{N_0-1} n[-n\mu p_n + \mu(n+1)p_{n+1}] - N_0^2 \mu p_{N_0} \\
&= -\sum_{n=1}^{N_0} \mu n^2 p_n + \sum_{n=2}^{N_0} \mu(n-1) n p_n \\
&= -\mu \sum_{n=1}^{N_0} n^2 p_n + \mu \sum_{n=2}^{N_0} n^2 p_n - \mu \sum_{n=2}^{N_0} n p_n(t) \\
&= -\mu p_1(t) - \mu \sum_{n=2}^{N_0} n p_n \\
&= -\mu \mathbb{E}(N).
\end{aligned}$$

That is, the mean number of individuals alive at time t satisfies an ordinary differential equation:

$$\frac{d\mathbb{E}(N)}{dt} = -\mu \mathbb{E}(N). \tag{2.21}$$

Hence, $\mathbb{E}(N(t)) = N_0 e^{-\mu t}$.

Time of extinction This paragraph is inspired by Ref. Solomon *et al.* (1971). Let τ_i be the time to first death in a population of initial size i. Then, for $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(\tau_i \leq t) &= 1 - \mathbb{P}(\tau_i > t) \\
&= 1 - \mathbb{P}(N(t) = i : N(0) = i) \\
&= 1 - e^{-i\mu t}.
\end{aligned} \tag{2.22}$$

Its density function is

$$f_i(t) = \frac{d}{dt}\mathbb{P}(\tau_i \leq t) = i\mu e^{-i\mu t}. \tag{2.23}$$

Thus,

$$\begin{aligned}\mathbb{E}(\tau_i) &= \int_0^{+\infty} i\mu t e^{-i\mu t} dt \\ &= \frac{1}{i\mu}.\end{aligned} \tag{2.24}$$

The time to extinction τ of a population of initial size N_0 is given by the sum of the τ_is for $0 < i \leq N_0$: $\tau = \tau_1 + ... + \tau_{N_0}$. Hence, the mean time to extinction of a pure death process is given by:

$$\mathbb{E}(\tau) = \sum_{i=1}^{N_0} \mathbb{E}(\tau_i) = \frac{1}{\mu}\sum_{i=1}^{N_0}\frac{1}{i} \approx \frac{1}{\mu}(\log(N_0) + \gamma_e) \tag{2.25}$$

where γ_e is the Euler–Mascheroni constant. Note that the centred distribution (distribution's mean brought back to 0) of τ is a Gumbel distribution.

Proof. For any time t, the c.d.f. of the random variable τ is

$$\begin{aligned}\mathbb{P}(\tau < t + \mathbb{E}(\tau)) &= (1 - e^{-\mu(t+\mathbb{E}(\tau))})^{N_0} \\ &= (1 - \frac{e^{-\mu t - \gamma_e}}{N_0})^{N_0}.\end{aligned} \tag{2.26}$$

When $N_0 \to +\infty$, we obtain

$$\lim_{N_0 \to +\infty} \mathbb{P}(\tau < t + \mathbb{E}(\tau)) = e^{-e^{-\mu t - \gamma_e}} \tag{2.27}$$

which is the c.d.f. of a Gumbel distribution $Gumbel(m, \beta)$ with parameters $m = \frac{-\gamma_e}{\mu}$ and $\beta = \frac{1}{\mu}$. □

2.7.3 Time of extinction of a simple birth and death process

In Chapter 5, I will also make use of some results from the analysis of the continuous-time branching process called *simple birth and death process*. In particular, I will be interested in characterising the distribution (mode and mean)

of the time to extinction of a population of individuals that follows a birth and death process with death rate μ and birth rate λ. That is, any individual alive at time t has probability $\mu\Delta t$ of dying before $t + \Delta t$ as $\Delta t \to 0$. In addition, during this small time interval Δt, each individual has a probability $\lambda\Delta t$ to give birth to a new individual. We assume that $\mu > \lambda$ so that extinction is certain (Allen, 2010).

Let us write $p_0(t)$ the probability that such a population, which started with only one individual, is extinct at time t. We have (Allen, 2010):

$$p_0(t) = \frac{\mu - \mu e^{(\mu-\lambda)t}}{\lambda - \mu e^{(\mu-\lambda)t}}. \tag{2.28}$$

Now, consider a population starting with $N_0 > 1$ individuals. We write $N(t)$ the number of individuals at time t and τ_{N_0} the time to extinction of the population. Then, we have

$$\mathbb{P}(\tau_{N_0} < t : N(0) = N_0) = p_0(t)^{N_0}. \tag{2.29}$$

The distribution of time to extinction is characterised by its p.d.f.:

$$f_{\tau_{N_0}}(t) = \frac{d}{dt}(p_0(t)^{N_0}) = N_0 p_0(t)^{N_0-1} p_0'(t), \tag{2.30}$$

where p_0' is the first derivative of p_0.

Mode of the distribution The mode of this distribution is the maximum of the p.d.f.. We compute this maximum by finding t such that

$$\frac{df_{\tau_{N_0}}}{dt}(t) = N p_0(t)^{N_0-2}(p_0(t)p_0''(t) + p_0'(t)^2(N_0 - 1)) = 0. \tag{2.31}$$

The first and second derivative of p_0 are:

$$p_0'(t) = \frac{\mu(\mu-\lambda)^2 e^{(\mu-\lambda)t}}{(\lambda - \mu e^{(\mu-\lambda)t})^2}, \tag{2.32a}$$

$$p_0''(t) = \frac{\mu(\mu-\lambda)^3 e^{(\mu-\lambda)t}(\lambda + \mu e^{(\mu-\lambda)t})}{(\lambda - \mu e^{(\mu-\lambda)t})^3}. \tag{2.32b}$$

Note that the denominator of p_0'' is negative so is p_0''. The equation $p_0(t)p_0''(t) + p_0'(t)^2(N_0 - 1) = 0$, can be simplified into a polynomial of degree 2 in $X = e^{(\mu-\lambda)t}$:

$$\lambda + (\mu - \lambda)N_0 X - \mu X^2 = 0. \tag{2.33}$$

Solving this polynomial, we obtain:

$$t_{mode} = \frac{1}{\mu - \lambda} \log\left(\frac{(\mu - \lambda)N_0 + \sqrt{4\lambda\mu + (\mu - \lambda)^2 N_0^2}}{2\mu}\right). \tag{2.34}$$

This is the mode of the distribution of time to extinction of a population, with N_0 individuals at $t = 0$, that follows a simple birth and death process. When $N_0 \to +\infty$, we obtain:

$$t_{mode} \underset{N_0 \to +\infty}{\longrightarrow} \frac{1}{\mu - \lambda} \log\left(\frac{\mu - \lambda}{\mu} N_0\right). \tag{2.35}$$

Expected time to extinction I also provide expressions to compute the mean time to extinction, $\bar{\tau}_{N_0}$, of a population following a simple birth and death process with initial size N_0. The expected time to extinction satisfies the following relationship (Allen, 2010):

$$\bar{\tau}_{N_0} = N_0\lambda(1 + \bar{\tau}_{N_0+1}) + N_0\mu(1 + \bar{\tau}_{N_0-1}) + (1 - N_0(\mu + \lambda))(1 + \bar{\tau}_{N_0}). \tag{2.36}$$

Note that we have $\bar{\tau}_0 = 0$. Making use of this recursive relation, we can obtain an analytic expression for τ_m, $m > 1$ (Allen, 2010):

$$\bar{\tau}_{N_0} = \begin{cases} \frac{1}{\mu} + \sum_{i=2}^{+\infty} (\frac{\lambda}{\mu})^{i-1} \frac{1}{\mu i} & \text{if } N_0 = 1 \\ \bar{\tau}_1 + \sum_{s=1}^{N_0-1} (\frac{\lambda}{\mu})^s \sum_{i=s+1}^{+\infty} (\frac{\lambda}{\mu})^{i-1} \frac{1}{\mu i} & \text{for } N_0 > 1 \end{cases}. \tag{2.37}$$

By recognizing that $-\log(1 - x) = \sum_{n=1}^{+\infty} \frac{x^n}{n}$ for $|x| < 1$, τ_1 can be re-written as:

$$\bar{\tau}_1 = \frac{1}{\mu} + \sum_{i=2}^{+\infty} (\frac{\lambda}{\mu})^{i-1} \frac{1}{\mu i} = -\frac{1}{\lambda} \log(1 - \frac{\lambda}{\mu}). \tag{2.38}$$

2.8 Approximate Bayesian computation: rejection algorithm

When developing mathematical models of biological systems, one of the central question is to find the parameter values that allow the model to best reproduce the experimental data. In this section, I introduce a Bayesian method to infer parameters of mathematical models (Toni *et al.*, 2009), called *rejection algorithm*. This method is based on Bayes' theorem (for instance introduced in Ref. Blitzstein & Hwang (2015) and in many other texts), which relates the conditional probabilities of two events A and B:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)},$$

where $\mathbb{P}(A|B)$ describes the probability that the random event A happens knowing that the random event B occurred. In statistical inference, this theorem is formulated as:

$$\pi(\theta|D) = \frac{\pi(\theta)\pi(D|\theta)}{\int_\theta \pi(D|\theta)\pi(\theta)d\theta}$$

where θ represents the set of the parameters of the model and D is the observed data. In this equation, $\pi(\theta)$ is known as the *prior distribution*, $\pi(D|\theta)$ is the *likelihood* of observing the data D given parameters θ, and $\pi(\theta|D)$ is called the *posterior distribution*. The integral in the denominator is just a normalisation constant that can be ignored. Thus, we often write:

$$\pi(\theta|D) \propto \pi(\theta)\pi(D|\theta). \tag{2.39}$$

The prior distribution encodes the prior beliefs of the user about the parameters. For instance, the user may know that their parameters lie in a small value interval and thus choose an *informative* prior distribution (such as a normal or log-normal distribution) which gives a higher density on this region of the parameter space. On the contrary, a user with little to no knowledge of the parameters might prefer choosing a *non-informative* prior distribution, for instance a uniform distribution over a large interval of values. The user aims to evaluate the posterior distribution, which according to equation (2.39) requires the computation of the likelihood. Unfortunately, it is often difficult to compute for mathematical models.

2. MATHEMATICAL BACKGROUND

The rejection algorithm, which belongs to the family of approximate Bayesian computation (ABC) methods, allows one to infer posterior distributions exploiting the computational efficiency of modern simulation techniques by replacing the calculation of the likelihood with a comparison between the observed and simulated data (Toni *et al.*, 2009). The algorithm goes as follows:

Algorithm 1 ABC rejection algorithm (Toni *et al.*, 2009)

Choose the desired posterior sample size N, the acceptance threshold ϵ, the distance measure $d(\cdot, \cdot)$ and set $n = 0$.
while $n < N$ **do**
    Sample the parameters θ^* from the prior distribution $\pi(\theta)$.
    Simulate a dataset D^* from $\pi(D|\theta^*)$.
    if $d(D, D^*) \leq \epsilon$ **then**
        Accept the parameter set θ^*
        Set $n = n + 1$.
    end if
end while

A wordy version of this algorithm could be as follows. First, sample parameters according to the prior distribution and simulate the mathematical model with these parameters as inputs. Then, compute the distance between the experimental data set D and the model output D^*. If the distance is small enough, then accept the parameters, *i.e.*, store them for later. These steps are repeated as many times as necessary to reach a posterior of size N (in practice, the user fixes a maximum number of simulation $N_{sim} \gg N$ and hopes that the algorithm will manage to find N parameter sets for which the distance is small enough). The set of all the accepted parameters defines the posterior distribution. We say that the algorithm was a success, when the posterior distribution is significantly narrower than the prior distribution: the true parameters have a high probability to lie in the small interval defined by this posterior. When the posterior distribution presents a large standard deviation, the user might want to choose a smaller acceptance threshold or a more informative prior. However these two solutions may not always lead to a narrower posterior distribution. Some data set might not allow the user to learn about specific parameters in the model. Finally, the algorithm might also converge slowly (*i.e.*, the algorithm needs many occurrences, N_{sim}, to find N

parameter sets for which the distance is smaller than the acceptance threshold), especially when the parameter space to explore is large (many parameters or prior distributions spanning a large interval). More computationally efficient algorithms have been developed, such as ABC–sequential Monte Carlo (Toni *et al.*, 2009), but are not used in this thesis as I never infer more than four parameters at once.

Note that in this thesis, instead of choosing a threshold ϵ, I compute the distance value $d(D, D^*)$ for the N_{sim} simulations and select the N parameter sets that return the N smallest distance values.

2.9 Agent-based modelling

In ODEs models and branching processes, all individuals of a population follow the same rules and have similar attributes. This assumption, that the population is homogeneous, is usually unrealistic as considerable differences can be observed from one individual to an other: for instance, two cells of a same population can express a different number of a specific receptor at their surface (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Feinerman *et al.*, 2010). They can also behave very differently depending on their local environment (maybe one cell has more access to resources than another,...). This apparent heterogeneity in a given population is difficult to incorporate in ODEs models as each individual should be represented by a different variable. An *agent-based model (ABM)* can be used, instead, to model heterogeneous populations of interacting agents (Bauer *et al.*, 2009; Truong *et al.*, 2022). An ABM is a stochastic, discrete-event and discrete-time numerical model (An *et al.*, 2009) in which each individual of the population (for instance a cell or a bacterium), also known as *agent*, has a set of specific attributes and obeys his own set of rules that might depend on its local environment (which can be a 2D or 3D space, a local protein concentration, *etc.*...). Agent-based modelling thus allows one to model macroscopic phenomena (at the population scale) emerging from the aggregated outcome of the agents' properties, behaviours and interactions (Borgonovo *et al.*, 2022). ABMs have been widely used to model biological systems such as tumour environment (An *et al.*, 2009), cancer cell's response to a treatment (Truong *et al.*, 2022) or decision making in ecology (DeAngelis & Diaz, 2019)... Indeed, a great advantage of agent-based modelling is that their definition is

easy to understand: ABMs allow the user to model almost any biological system intuitively as we specify the set of rules for each agent and make them interact, mimicking reality. Such models are also flexible as it is usually easy to modify a rule.

Agent-based modelling thus allows researchers to examine systems whose behaviour cannot be entirely described analytically by equations (Borgonovo *et al.*, 2022). However, results of an ABM are, in general, hard to interpret, mainly because there is no closed-form expression that links the inputs and outputs of the model. For this reason, modellers rely on numerical observations and often struggle to assess whether the result of the ABMs are robust and their conclusion valid (Borgonovo *et al.*, 2022). In addition, despite being particularly suitable for parallel programming, an ABM can be a very complex network (the model can be multiscale and spatio-temporal) resulting in long simulation times. For this reason, it is not always possible to parametrise such models from experimental observations. Due to the difficulty to calibrate ABMs, they are often seen as qualitative tools, only offering exploratory observations. A lot of recent studies have been devoted to finding systematic and exhaustive methods of robust sensitivity analysis of agent-based models (Borgonovo *et al.*, 2022; Ten Broeke *et al.*, 2016; Thiele *et al.*, 2014). These methods usually require significant computational resources (time and power) and may not be applicable to all agent-based models. When conducting such as study of an ABM, one has to define the goal of the sensitivity analysis, *i.e.*, the outputs to observe and measure to assess the analysis. Determining a relevant output may be difficult, depending on the system. For example, in one of the agent-based model described in Chapter 5, I measure the average number of receptors expressed by cells of two different cohorts. I use this criterion to determine for which parameter values my model switches from a specific regime to another. While this criterion is relevant for some sets of parameters, I show that it is unfortunately not the case for all parameter values. That is, for a robust analysis, another criterion should be determined. Finally, to test the robustness and stability of agent-based models, one may want to vary non-parametric elements, such as behavioural rules (Borgonovo *et al.*, 2022), which complicates the process.

In this thesis, we keep the agent-based models simple enough so that they can

be described, at least partially, mathematically and thus provide a mechanistic understanding of the system.

2.10 Guide for Chapters 3 and 4

This section does not describe any mathematical background but provide a summary of the mathematical models of receptor-ligand systems (chemical reaction schemes) studied in Chapters 3 and 4 and the main results of their analysis (the expression of their amplitude and EC_{50}, when computed). We advise the reader to use this section as a guide when reading both these chapters.
The receptors considered in the models are composed of one, two or three trans-membrane chains (γ, α and β), with or without extrinsic kinase (denoted JAK). One of the receptor-ligand model describes the competition for the γ chain between the main receptor and an extra decoy chain denoted R. The concentrations of γ and α chains are written x and y, respectively, in the mathematical expressions. The concentration of R in Chapter 3 and the concentration of the β chain in Chapter 4 will be denoted w in the mathematical expressions. The concentration of the extrinsic kinase is denoted z in mathematical expressions and the ligand concentration is written L (L is also used to denote the species in the chemical reactions). The notation $N_\bullet$ will denote the total number of the chain or kinase of concentration $\bullet$ (for instance, N_x denotes the total number of γ chain which concentration is written x). The affinity constants of the chemical reactions considered are denoted by K_i, K_i' or K_i'', $i = 1, 2, 3$. This notation, partially summarised in Table 2.1, will be recalled in Chapters 3 and 4. The modelling assumptions for the different models will be stated in the relevant sections.

We provide tables which recapitulate all the specific receptor-ligand systems

Description	Species	Concentration	Total quantity
Ligand	L	L	
Primary trans-membrane chain	γ	x	N_x
Secondary trans-membrane chain	α	y	N_y
Extrinsic intra-cellular kinase chain	JAK	z	N_z

Table 2.1: Summary of the common notation used in Chapters 3 and 4.

studied in chapters 3 and 4: models with intrinsic kinase activity (RTK) are summarised in Table 2.2, and models that require an intra-cellular extrinsic kinase to signal (IEK) can be found in Table 2.3. Table 2.4 recapitulates the analytic results obtained for the amplitude and EC_{50} of receptor-ligand systems with extrinsic kinase. The results on the amplitude and EC_{50}/IC_{50} (the definition of the IC_{50} can be found in Section 4.5) of RTK models are summarised in Tables 2.5 and 2.6, respectively. Note that the main reason why an expression for the amplitude or the EC_{50} could not be obtained is indicated in the summarising tables with the appropriate symbol (★, ♠, ♣) as follows[1]:

★ A Gröbner basis of the polynomial system describing the model (or of the augmented polynomial system for the computation of the EC_{50}) could not be computed within a reasonable time,

♠ The dose-response curve was bell-shaped and the computation of its maximum (amplitude) was intractable,

♣ The amplitude or the EC_{50} is a root of a polynomial of degree greater than 2. If the polynomial degree is less than 5, the roots may be computed exactly (resulting in long complicated expressions), but selecting the positive real root that let the other variables of the system positive is potentially hard.

.

[1]These descriptions will make sense while reading Chapters 3 and 4.

Model	Chemical reactions
Monomeric RTK	$L + \gamma \rightleftharpoons L:\gamma \qquad K_3$
Homodimeric RTK A	$\gamma + \gamma \rightleftharpoons \gamma:\gamma \qquad K_2$ $L + \gamma:\gamma \rightleftharpoons L:\gamma:\gamma \qquad K_3$
Homodimeric RTK B	$L + \gamma \rightleftharpoons L:\gamma \qquad K_3'$ $\gamma + L:\gamma \rightleftharpoons L:\gamma:\gamma \qquad K_2'$
Homodimeric RTK AB	$\gamma + \gamma \rightleftharpoons \gamma:\gamma \qquad K_2$ $L + \gamma:\gamma \rightleftharpoons L:\gamma:\gamma \qquad K_3$ $L + \gamma \rightleftharpoons L:\gamma \qquad K_3'$ $\gamma + L:\gamma \rightleftharpoons L:\gamma:\gamma \qquad K_2'$
Heterodimeric RTK A	$\gamma + \alpha \rightleftharpoons \gamma:\alpha \qquad K_2$ $L + \gamma:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_3$
Heterodimeric RTK B	$\gamma + L \rightleftharpoons L:\gamma \qquad K_3'$ $\alpha + L:\gamma \rightleftharpoons L:\gamma:\alpha \qquad K_2'$
Heterodimeric RTK C	$L + \alpha \rightleftharpoons L:\alpha \qquad K_3''$ $\gamma + L:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_2''$
Heterodimeric RTK AB	$\gamma + \alpha \rightleftharpoons \gamma:\alpha \qquad K_2$ $L + \gamma:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_3$ $\gamma + L \rightleftharpoons L:\gamma \qquad K_3'$ $\alpha + L:\gamma \rightleftharpoons L:\gamma:\alpha \qquad K_2'$
Heterodimeric RTK AC	$\gamma + \alpha \rightleftharpoons \gamma:\alpha \qquad K_2$ $L + \gamma:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_3$ $\alpha + L \rightleftharpoons L:\alpha \qquad K_3''$ $\gamma + L:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_2''$
Heterodimeric RTK BC	$\alpha + L \rightleftharpoons L:\alpha \qquad K_3''$ $\gamma + L:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_2''$ $\gamma + L \rightleftharpoons L:\gamma \qquad K_3'$ $\alpha + L:\gamma \rightleftharpoons L:\gamma:\alpha \qquad K_2'$
Heterodimeric RTK ABC	$\gamma + \alpha \rightleftharpoons \gamma:\alpha \qquad K_2$ $L + \gamma:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_3$ $\gamma + L \rightleftharpoons L:\gamma \qquad K_3'$ $\alpha + L:\gamma \rightleftharpoons L:\gamma:\alpha \qquad K_2'$ $\alpha + L \rightleftharpoons L:\alpha \qquad K_3''$ $\gamma + L:\alpha \rightleftharpoons L:\gamma:\alpha \qquad K_2''$

Table 2.2: Chemical reaction schemes of the RTK models. Monomeric RTK, homodimeric RTK A and heterodimeric RTK A are analysed in Section 4.2.1, 4.2.2 and 4.2.3, respectively. The other models are studied in Section 4.5.

Model	Chemical reactions
Monomeric with IEK (Section 4.2.4)	$JAK + \gamma \rightleftharpoons JAK:\gamma \quad K_1$ $L + \gamma:JAK \rightleftharpoons L:\gamma:JAK \quad K_3$ $L + \gamma \rightleftharpoons L:\gamma \quad K_3$
Homodimeric with IEK (Section 4.2.5)	$JAK+\gamma \rightleftharpoons JAK:\gamma \quad K_1$ $JAK:\gamma+\gamma \rightleftharpoons \gamma:\gamma:JAK \quad K_2$ $JAK:\gamma+JAK:\gamma \rightleftharpoons JAK3:\gamma:\gamma:JAK \quad K_2$ $\gamma+\gamma \rightleftharpoons \gamma:\gamma \quad K_2$ $L+\gamma:\gamma \rightleftharpoons L:\gamma:\gamma \quad K_3$ $L+\gamma:\gamma:JAK \rightleftharpoons L:\gamma:\gamma:JAK \quad K_3$ $L+\gamma:JAK:\gamma:JAK \rightleftharpoons L:\gamma:JAK:\gamma:JAK \quad K_3$
Heterodimeric with IEK (Section 3.2)	$JAK + \gamma \rightleftharpoons JAK:\gamma \quad K_1$ $\alpha + JAK:\gamma \rightleftharpoons \alpha:\gamma:JAK \quad K_2$ $\gamma + \alpha \rightleftharpoons \alpha:\gamma \quad K_2$ $L + \alpha:\gamma \rightleftharpoons L:\alpha:\gamma \quad K_3$ $L + \alpha:\gamma:JAK \rightleftharpoons L:\alpha:\gamma:JAK \quad K_3$
Heterodimeric with IEK and additional R (Section 3.4)	$JAK + \gamma \rightleftharpoons JAK:\gamma \quad K_1$ $\alpha + JAK:\gamma \rightleftharpoons \alpha:\gamma:JAK \quad K_2$ $\gamma + \alpha \rightleftharpoons \alpha:\gamma \quad K_2$ $L + \alpha:\gamma \rightleftharpoons L:\alpha:\gamma \quad K_3$ $L + \alpha:\gamma:JAK \rightleftharpoons L:\alpha:\gamma:JAK \quad K_3$ $R + \gamma \rightleftharpoons R:\gamma \quad K_4$ $R + \gamma:JAK \rightleftharpoons R:\gamma:JAK \quad K_4$
Trimeric with IEK (Section 4.2.6)	$JAK + \gamma \rightleftharpoons JAK:\gamma \quad K_1$ $\beta + \gamma:JAK \rightleftharpoons \beta:\gamma:JAK \quad K_2$ $\beta + \gamma \rightleftharpoons \beta:\gamma \quad K_2$ $\alpha + \beta:\gamma:JAK \rightleftharpoons \alpha:\beta:\gamma:JAK \quad K_3$ $\alpha + \beta:\gamma \rightleftharpoons \alpha:\beta:\gamma \quad K_3$ $L + \alpha:\beta:\gamma:JAK \rightleftharpoons L:\alpha:\beta:\gamma:JAK \quad K_4$ $L + \alpha:\beta:\gamma \rightleftharpoons L:\alpha:\beta:\gamma \quad K_4$

Table 2.3: Chemical reaction schemes of the models with IEK.

Model	Amplitude	EC_{50}
Monomeric with IEK	$\frac{K_1 z}{1+K_1 z} N_x$	$\frac{1}{K_3}$
Homodimeric with IEK	Not obtained ★	Not obtained ★
Heterodimeric with IEK	$\frac{K_1 z}{1+K_1 z} M$	$M \frac{1+K_2(N_x+N_y-M)+\sqrt{1+K_2^2(N_y-N_x)^2+2K_2(N_x+N_y-M)}}{K_2K_3(M-2N_x)(M-2N_y)}$
Heterodimeric with IEK and additional R	$\frac{K_1 z}{1+K_1 z} M$	Solution of a polynomial of degree 3 independent on K_1 and N_z (and condition for $x>0$ and $y>0$) ♣
Trimeric with IEK	$\frac{K_1 z}{1+K_1 z} \min(N_x, N_y, N_w)$	Solution of a polynomial of degree 4 independent on K_1 and N_z (and condition for $x>0$, $y>0$ and $w>0$) ♣

Table 2.4: Amplitude and EC_{50} expression (when obtained) of the models with intra-cellular extrinsic kinase (IEK). We wrote: $z = \frac{-1+K_1(N_z-N_x)+\sqrt{\Delta_1}}{2K_1}$ and $M = \min(N_x, N_y)$

Model	Amplitude
Monomeric	N_x
Homodimeric RTK A	$\frac{N_x}{2}$
Homodimeric RTK B	$\frac{(-1+\sqrt{1+2K_2' N_x})^2}{4K_2'}$
Homodimeric RTK A+B	$\in (0, \frac{1+2K_2' N_x - \sqrt{1+4K_2' N_x}}{2K_2'})$ ♠
Heterodimeric RTK A	$\min(N_x, N_y)$
Heterodimeric RTK B	$\frac{1+K_2'(N_y+N_x)-\sqrt{1+K_2'^2(N_y-N_x)^2+2K_2'(N_x+N_y)}}{2K_2'}$
Heterodimeric RTK C	$\frac{1+K_2''(N_y+N_x)-\sqrt{1+K_2''^2(N_x-N_y)^2+2K_2''(N_x+N_y)}}{2K_2''}$
Heterodimeric RTK A+B	$\frac{1+K_2'(N_y+N_x)-\sqrt{1+K_2'^2(N_y-N_x)^2+2K_2'(N_x+N_y)}}{2K_2'}$
Heterodimeric RTK A+C	$\frac{1+K_2''(N_y+N_x)-\sqrt{1+K_2''^2(N_x-N_y)^2+2K_2''(N_x+N_y)}}{2K_2''}$
Heterodimeric RTK B+C	$\in (0, \min(N_x, N_y)] \cap (0, \frac{K_3'+K_3''+K_2'K_3'(N_x+N_y)-\sqrt{(K_3'+K_3''+K_2'K_3'(N_x+N_y))^2-4K_2'^2K_3'^2N_xN_y}}{K_2'K_3'})$ ♠
Heterodimeric RTK A+B+C	$\in (0, \min(N_x, N_y)] \cap (0, \frac{K_3'+K_3''+K_2K_3(N_x+N_y)+\sqrt{(K_3'+K_3''+K_2K_3(N_x+N_y))^2-4K_2^2K_3^2N_xN_y}}{K_2K_3})$ ♠

Table 2.5: Summary of the analytic expressions of amplitude (or any knowledge about this quantity) for the different RTK models. We write $\in (a, b)$ when the amplitude of the model ranges between a and b but we did not obtain an analytic expression.

Model	EC_{50}	IC50
Monomeric	$\frac{1}{K_3}$	N/A
Homodimeric RTK A	$\frac{1+2K_2N_x+\sqrt{1+4K_2N_x}}{2K_2K_3N_x}$	N/A
Homodimeric RTK B	$\frac{3+K_2'N_x+3\sqrt{1+2K_2'N_x}-K_3'\sqrt{\Delta_B^{ho}}}{2K_3}$	$\frac{3+K_2'N_x+3\sqrt{1+2K_2'N_x}+K_3'\sqrt{\Delta_B^{ho}}}{2K_3}$
Homodimeric RTK A+B	Solution of a polynomial of degree 3 with condition $x>0$ ♣	Solution of a polynomial of degree 3 with condition $x>0$ ♣
Heterodimeric RTK A	$M\frac{1+K_2(N_x+N_y-M)+\sqrt{1+K_2^2(N_y-N_x)^2+2K_2(N_x+N_y-M)}}{K_2K_3(M-2N_x)(M-2N_y)}$	N/A
Heterodimeric RTK B	$\frac{4}{K_3'(1+K_2'(N_x+N_y)+3\sqrt{1+K_2'^2(N_x-N_y)^2+2K_2'(N_x+N_y)})}$	N/A
Heterodimeric RTK C	$\frac{4}{K_3''(1+K_2''(N_x+N_y)+3\sqrt{1+K_2''^2(N_y-N_x)^2+2K_2''(N_x+N_y)})}$	N/A
Heterodimeric RTK A+B	$\mathcal{A}\frac{1+K_2(N_x+N_y-\mathcal{A})+\sqrt{\Delta_{ec}^{het}(K_3')}}{\mathcal{A}^2K_2K_3+4K_2K_3N_xN_y-2\mathcal{A}(K_3'+K_2K_3(N_x+N_y))}$	N/A
Heterodimeric RTK A+C	$\mathcal{A}\frac{1+K_2(N_x+N_y-\mathcal{A})+\sqrt{\Delta_{ec}^{het}(K_3'')}}{\mathcal{A}^2K_2K_3+4K_2K_3N_xN_y-2\mathcal{A}(K_3''+K_2K_3(N_x+N_y))}$	N/A
Heterodimeric RTK B+C	$\frac{-p+\sqrt{p^2-16\mathcal{A}^2K_3'K_3''}}{4\mathcal{A}K_3'K_3''}$	$\frac{-p-\sqrt{p^2-16\mathcal{A}^2K_3'K_3''}}{4\mathcal{A}K_3'K_3''}$
Heterodimeric RTK A+B+C	Solution of a polynomial of degree 3 with conditions $x>0$, $y>0$ ♣	Solution of a polynomial of degree 3 with conditions $x>0$, $y>0$ ♣

Table 2.6: Summary of the analytic expressions of the EC_{50} (or any knowledge about this quantity) for the different RTK models. We wrote $M=\min(N_x,N_y)$, $\Delta_B^{ho}=\frac{2(7+9\sqrt{1+2K_2'N_x})+K_2'N_x(K_2'N_x+6(4+\sqrt{1+2K_2'N_x}))}{K_3'^2}$, $\Delta_{ec}^{het}(K)=1+K_2^2(N_x-N_y)^2+2K_2(N_x+N_y)+\frac{2K_2(K-K_3)}{K_3}\mathcal{A}$ and $p=K_2'K_3'(A-2N_y)(2N_x-A)+2\mathcal{A}(K_3'+K_3'')$. Note that $N_x+N_y-M=\max(N_x,N_y)$. We wrote N/A when the quantity does not exist for the model. Finally, in each row, $\mathcal{A}$ denotes the amplitude of the corresponding model.

Chapter 3

Tuning of IL-7 signalling through imbalanced abundances of receptors and kinases

This chapter, for the most part, will focus on interleukin-7 (IL-7) and its receptor (IL-7R) (Cotari *et al.*, 2013b; Gonnord *et al.*, 2018; Leonard *et al.*, 2019; Palmer *et al.*, 2008; Park *et al.*, 2019; Rochman *et al.*, 2009). Interleukin-7 is a cytokine involved in T cell development, survival and homeostasis (Akdis *et al.*, 2011; Ma *et al.*, 2006). Its receptor, IL-7R, is displayed on the surface of T cells and is composed of two trans-membrane chains: the common gamma chain (denoted by γ_c) and the specific high affinity chain IL-7Rα (denoted by α when there is no ambiguity) (Gonnord *et al.*, 2018; Ma *et al.*, 2006; Molina-París *et al.*, 2013; Park *et al.*, 2019). This cytokine receptor does not contain intrinsic kinase domains. Thus, it makes use of Janus family tyrosine kinases (JAKs) and signals in part by the activation of signal transducer and activator of transcription (STAT) proteins (Lin & Leonard, 2019). The gamma chain binds to the intra-cellular extrinsic Janus kinase molecule, JAK3. Binding of IL-7 to the dimeric JAK3-bound IL-7 receptor, defined as $\alpha : \gamma_c : JAK3$, initiates a series of biochemical reactions from the membrane of the cell to its nucleus, which in turn lead to a cellular response. For the IL-7R system, the STAT protein preferentially activated is STAT5 (Lin & Leonard, 2019). The amount of phosphorylated STAT5 (which

is released when IL-7 binds its receptor) can be used as the experimental measure of the intra-cellular response generated by the IL-7 stimulus.

As mentioned in the introduction of this thesis, previous single-cell studies have shown that isogenic cells present heterogeneous expression levels of receptor constituents copy numbers (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Feinerman *et al.*, 2010). Prior studies also demonstrated that cytokine receptor signalling can be significantly altered by the precise abundance of the molecular signalling components (Cotari *et al.*, 2013b). Since the cytokine receptors of the common gamma chain family are signalling through association with an intra-cellular Janus kinase (JAK), their co-optation of soluble kinases could elicit large non-genetic variability of ligand detection within a population of cells. Grégoire Altan-Bonnet, Jesse Cotari and Guillaume Voisinne used single-cell measurements to explore the impact of the relative abundance of signalling components in terms of variable sensitivity (EC_{50}) and amplitude in response to extra-cellular cytokines. In particular, they focused on the common gamma chain family and explored a conjecture derived from their initial finding of γ_c chain competition (Cotari *et al.*, 2013b); namely, that increases in γ_c chain abundance would increase the number of fully-formed receptors and the sensitivity to the γ_c family cytokines. Furthermore, by analogy to the role of RTK upregulation in cancer (Bache *et al.*, 2004; Du & Lovly, 2018; Eladdadi & Isaacson, 2008; Regad, 2015), a natural expectation would be that γ_c chain upregulation would lead to stronger maximal signal upon ligand binding. After all, one could naturally think that increasing the abundance of a crucial receptor component, such as γ_c chain, increases the number of fully formed receptors, and thus increases cytokine uptake and cell's response. This work, however, stems from experiments examining the impact of γ_c chain abundance on T cell responses to IL-7, which immediately led to the paradoxical observation that increased abundance of the γ_c chain in fact decreased both the sensitivity and the maximal response to IL-7. I present the experimental work in Section 3.1.

In an attempt to resolve this paradox, I developed and analysed three IL-7R mathematical models, making use of the notation of chemical reaction network theory introduced in Section 2.3. In the first model, described in Section 3.2, the signalling complex $L : \gamma_c : \alpha : JAK3$ is formed sequentially. This model

also includes the formation of kinase-deprived complexes, $\alpha : \gamma_c$, called "dummy" receptors, which can bind to the ligand L (here IL-7) but are unable to signal. The formation of "dummy" complexes $L : \gamma_c : \alpha$ can explain the observed decrease in amplitude as the γ_c abundance increases. However, the model is unable to explain the EC_{50} behaviour and thus needs to be augmented. We conjectured that the increase in EC_{50} with increasing γ_c abundance could be explained by one of the two following mechanisms: an allostery effect or the competition for the γ_c chain between the different receptors of the common gamma chain family. I explore the first mechanism in Section 3.3, in which I developed the allostery model. This mathematical model is a variation of the first IL-7R model in which the binding of the kinase JAK3 to the γ_c chain is assumed to change the binding affinity of the α chain to the complex $\gamma_c : JAK3$, or the ligand L to $JAK3 : \gamma_c : \alpha$. The second mechanism is tested in Section 3.4. To take into consideration that the γ_c chain is shared with other cytokine receptors (Rochman *et al.*, 2009), I modify the first IL-7R model to include an additional receptor chain which has the ability to bind to the γ_c chain or the complex $\gamma_c : JAK3$ to form decoy receptors. This extra chain is unknown *a priori* and could be for instance IL-2Rβ, IL-4Rα or the pre-formed complex IL-2Rβ:IL-2Rα. It prevents the γ_c chain from binding IL-7Rα and form IL-7R. This last model seems to be able to reproduce both amplitude and EC_{50} behaviours. Note that all the systems studied in this section are at steady state and we considered the ligand L to be in excess, *i.e.*, we consider L as a parameter of the models instead of a variable.

This study shows that the dose-response curve of a receptor-ligand system is not a univariate function of the ligand concentration but, instead, depends on the receptor chain abundances. It is particularly explicit when I make use of Gröbner bases to compute the analytic expressions of the steady state, amplitude and EC_{50} for the first and third IL-7R models. Analytic expressions directly show the dependence on parameters and facilitate parameter exploration, which will prove to be very useful when fitting the models to experimental data. I summarise the findings and conclude in section 3.5.

3.1 A paradoxical observation: increasing the availability of γ_c chains decreases IL-7 induced T cell response

In this section, I present the experimental work and data analysis conducted by Jesse Cotari, Guillaume Voisinne and Grégoire Altan-Bonnet and published in Sta *et al.* (2022b). Making use of flow cytometry, they analysed the responsiveness of murine T cells to γ_c cytokines (in particular IL-7) as measured by STAT5 phosphorylation. Let us first briefly introduce flow cytometry and cell-to-cell variability analysis so that a reader with no biological background can understand how the data have been acquired.

3.1.1 Flow cytometry and cell-to-cell variability analysis

Flow cytometry is a technique routinely used nowadays to measure physical and chemical characteristics of a population of cells (or particles) (Picot *et al.*, 2012; Rieseberg *et al.*, 2001). Cells in suspension in a tube or plate are passing one by one through one or more laser beams in a machine called flow cytometer. Cells are often first treated and labelled with antibodies and fluorescent markers so light is absorbed and re-emitted differently from the laser emitter. The flow cytometer then measures scattered light at several angles and fluorescence emission. The data, acquired and analysed by a computer, are then used to sort cells according to their characteristics. The whole flow cytometry process is summarised in Figure 3.1 from Ref. Picot *et al.* (2012).

Flow cytometry data show that a population of isogenic cells are never truly identical (Cotari *et al.*, 2013a). They may vary in expression of key proteins or do not produce signal at the same ligand concentration. This variability means that cells do not respond homogeneously to a given stimulus. The measure of cell-to-cell variability (CCV) accounts for the heterogeneity of cell populations but does not connect this variability to downstream cell response variability. Fortunately, flow cytometry allows the simultaneous measurement of protein abundance and pathway activation (for instance abundance of IL-7Rα and phosphorylated STAT5). Cell-to-cell variability analysis (CCVA) is a computational method which allows to

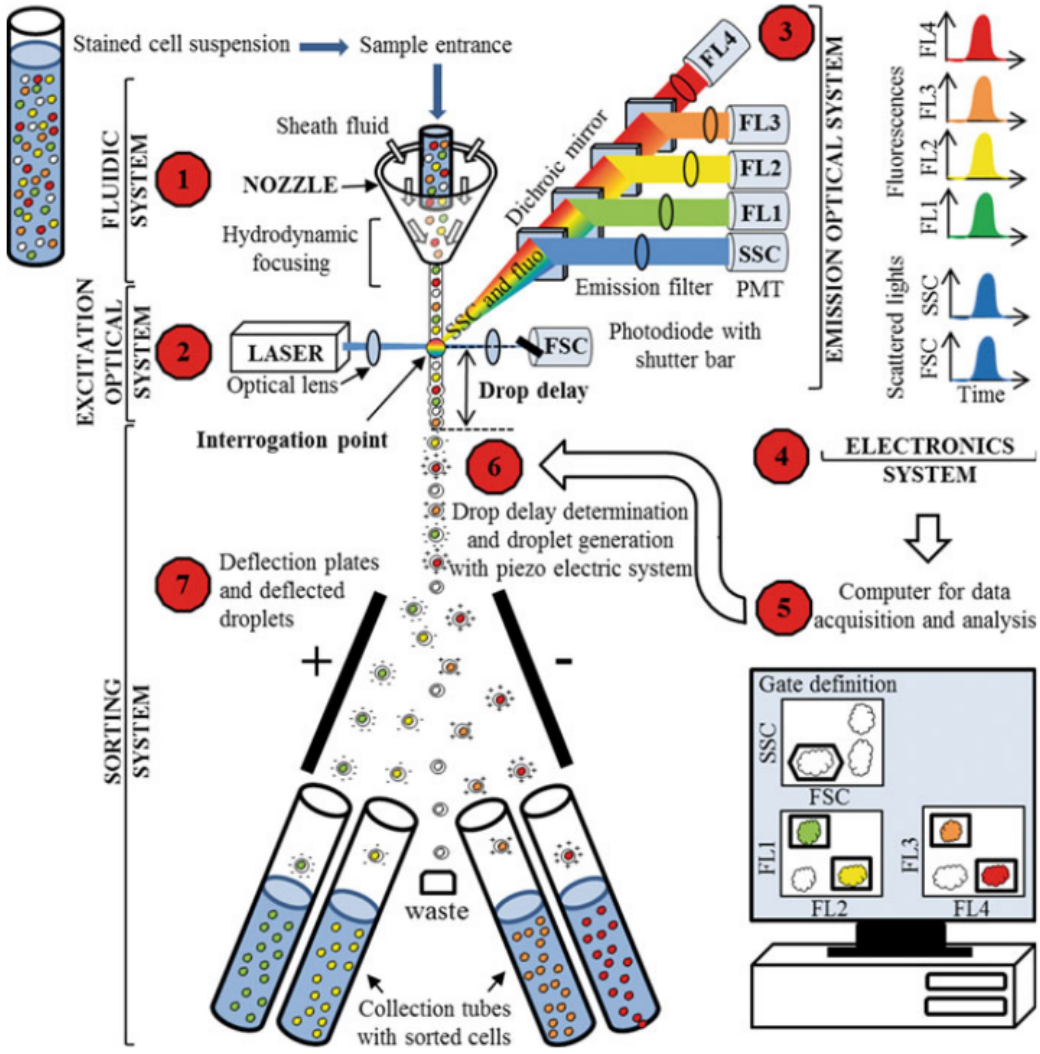


Figure 3.1: Scheme describing flow cytometry method, taken from Ref. Picot *et al.* (2012) (reproduced with permission from Springer Nature).

correlate the variability in protein abundance and the variability in cell response to a given stimulus, using flow cytometric data (see Figure 3.2).

Cell-to-cell variability analysis (CCVA) For each dose of stimulus (for instance ligand concentration), cells of similar protein abundance are gathered (data binning) and associated to a certain fluorescence intensity level. For each group of cells (bin), a dose-response curve is computed by plotting the fluorescence level of the bin as a function of the concentration of stimulus (see Figure 3.2). From this dose-response curve, amplitude and EC_{50} can be calculated. This method bypasses the fact that dose-response curves of a single cell are technically impossible (cells have a memory of their previous stimulation and adapt accordingly) (Cotari *et al.*, 2013a).

3.1.2 Data analysis

After cell transfection[1], which artificially increased γ_c abundance above normal physiological levels in some cells, Jesse Cotari, Guillaume Voisinne and Gregoire Altan-bonnet made use of flow cytometry and CCVA to analyse the response of murine T cells to IL-7. Guillaume Voisinne conducted the data analysis. The complete dataset contains 432,380 cells corresponding to 8 experimental conditions (7 IL-7 doses ranging from 10pM to 10 nM and the control "no-cytokine" condition) and for which the pSTAT5 fluorescence intensity along with the absolute number of JAK3, IL-7Rα, IL-2Rα and γ_c were quantified. Calibrated data were log-transformed and subdivided into bins corresponding to different total numbers of JAK3, IL-7Rα and γ_c proteins per cell denoted as N_z, N_y and N_x, respectively. For the N_y and N_z variables, bins of width 0.2 were created in the interval ranging from the 5% to the 95% quantile of log-transformed values. For the N_x variable, bins of width 0.25 were created in the entire range (from min to max) of log-transformed values. Only bins containing at least 5 cells for each experimental condition were selected for further analysis. For each bin, making use of the 8 experimental conditions, a dose-response curve has been computed (sigmoid) and

[1]Cell transfection is the process of deliberately introducing foreign DNA into a cell.

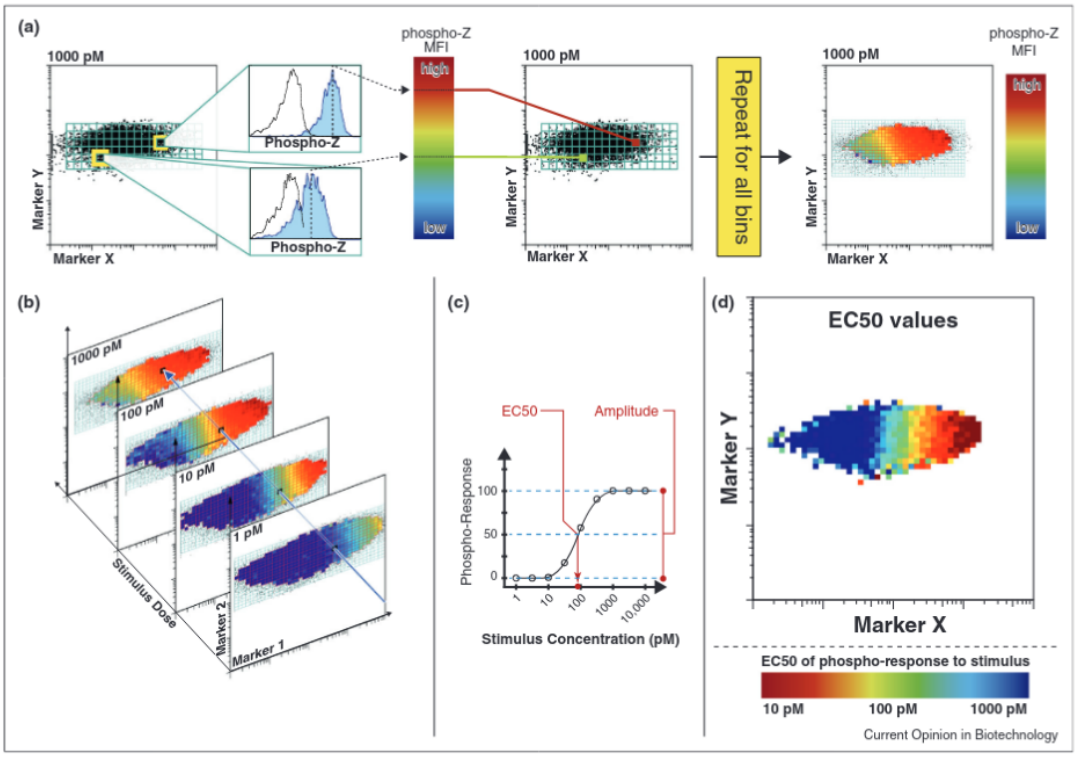


Figure 3.2: Scheme describing the cell-to-cell variability analysis (CCVA) method, taken from Ref. Cotari *et al.* (2013a) with permission from Elsevier. (a) Cells with similar protein abundances are grouped together and associated to a fluorescence intensity level (binning). (b) The process described in figure (a) is repeated for every dose of stimulus. (c) For each bin, the dose-response curve is computed by plotting the fluorescence intensity level as a function of the dose. (d) From dose-response curves, amplitude and EC_{50} can be computed. We thus obtain these pharmacological quantities for each bin.

fitted to a Hill equation:

$$pSTAT5(L) = \beta_1 + \beta_2 \times \frac{L}{(EC_{50} + L)}, \tag{3.1}$$

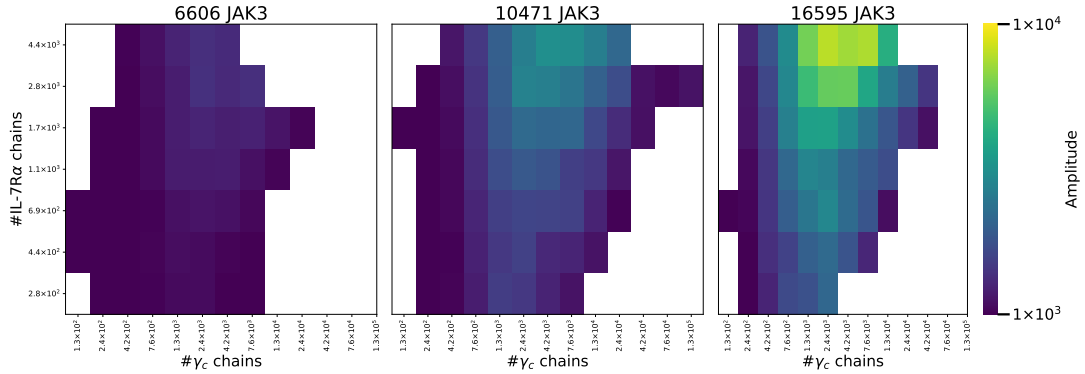
where L denotes the extra-cellular IL-7 concentration. Values of the β_1, β_2 and EC_{50} parameters were estimated using a non-linear least-square analysis. Guillaume Voisinne minimised the sum of differences between predicted and measured mean log-transformed pSTAT5 values weighted by the standard error of the mean. From this analysis, he selected the bins for which convergence towards a minimum was achieved and for which the estimated EC_{50} parameter was higher than $10^{-13}M$ with a 95% confidence interval within $[10^{-20}M, 10^{-8}M]$.

This final data set, that will be used to fit the models of this chapter, consists of 172 bins, *i.e.*, 172 values of amplitude and EC_{50}, each corresponding to a certain level of JAK3, IL-7Rα and γ_c. The colormaps of the amplitude and EC_{50} of these bins are displayed in Figure 3.3.

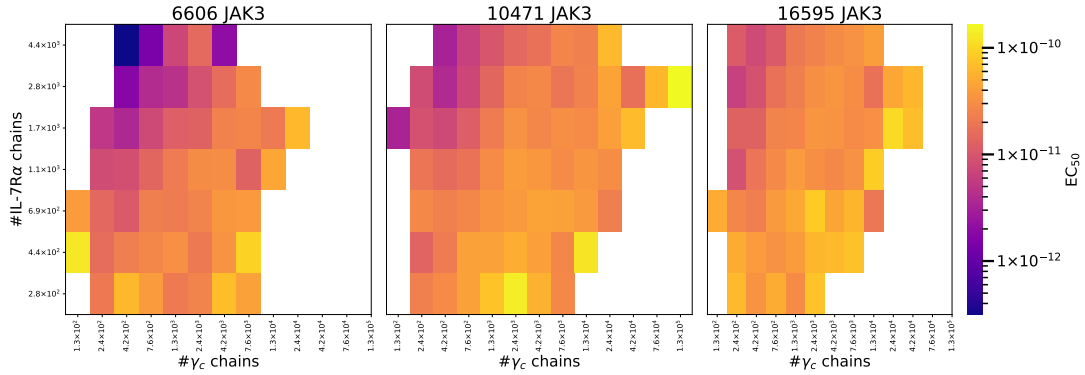
3.1.3 A counter-intuitive result

At a given stimulus, single-cell resolution shows that STAT5 phosphorylation correlates monotonically with the abundance of IL-7Rα and JAK3. However, a greater abundance of γ_c chains corresponds to diminished pSTAT5 levels (see Figure 3.4). The amplitude and EC_{50} maps (Figure 3.3) show a strikingly complex relationship between γ_c levels and IL-7 response. The EC_{50} map indicates that a greater abundance of γ_c increases EC_{50} values. The amplitude map exposes an even more complex relationship between γ_c expression levels and IL-7 response. A clear trend can be observed for any abundance of IL-7Rα, in which increasing the abundance of the γ_c chain first increases and then decreases the amplitude. This overall trend explains the reduced pSTAT5 levels at a high IL-7 dose (10 nM) observed in Figure 3.4.

Experimental results show a seemingly paradoxical observation: high abundance of γ_c is found to reduce the responsiveness to IL-7, reflected both in a smaller



(a) Experimental amplitude



(b) Experimental EC_{50}

Figure 3.3: Complete map of the experimental amplitude and EC_{50} for different levels of signalling components. Each rectangle corresponds to one bin of the data set.

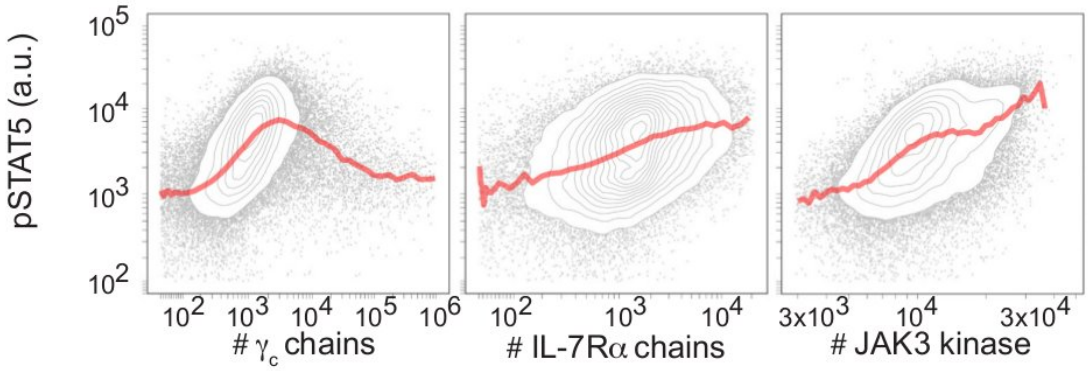


Figure 3.4: Figure from Sta *et al.* (2022b). Distribution of pSTAT5 in response to 10 nM of IL-7, for different abundances of IL-7Rα, JAK3, and γ_c after cell-to-cell variability analysis. Each grey dot represents a cell, the grey circles are level lines and red line represents the geometric mean of pSTAT5 for each level of receptor component.

amplitude and a greater EC_{50} of the pSTAT5 dose-response curve (diminished signal associated with the increased abundance of a crucial component of the signalling pathway have also been observed in Ref. Prabakaran *et al.* (2014)). In an attempt to explain this paradox, I propose, in the following sections, three mathematical models of IL-7 receptor-mediated signalling. The experimental conditions in which the data were obtained justify several approximations that make the modelling more tractable. First, since IL-7 signalling reaches equilibrium within 10 minutes of exposure (Cotari *et al.*, 2013a; Vogel *et al.*, 2016), the models assume that all biochemical reactions reach a steady-state to calculate the abundance of each complex. This simplification allows us to make use of experimentally-determined equilibrium constants (K) to quantify the interaction between receptor components in the membrane (Pillet *et al.*, 2010), rather than kinetic rates for the forward and backward biochemical reactions (on and off rates), which have not been experimentally determined. Second, the parameters defining STAT5 phosphorylation have not been determined. In previous work (Cotari *et al.*, 2013b), Jesse Cotari, Guillaume Voisinne and Gregoire Altan-Bonnet demonstrated that the abundance of pSTAT5 correlated linearly with the number of IL-7 molecules bound to the cell surface, hence we assumed that the fraction of phosphorylated STAT5 molecules and the number of IL-7Rα : γ_c :IL-7 complexes,

σ, follow the affine relation:

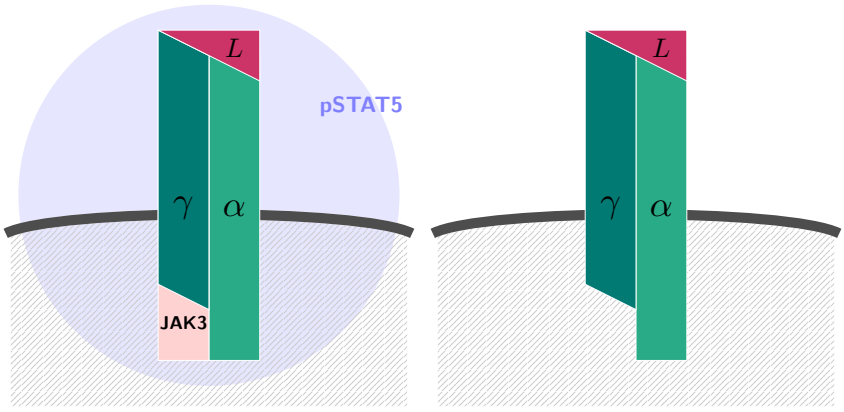
$$pSTAT5 = \xi_1 \sigma + \xi_2, \tag{3.2}$$

where ξ_1 and ξ_2 are parameters to be determined. Finally, since IL-7Rα depends on JAK1 (mentioned in the introduction (Chapter 1)) for signalling, a similar balance would exist for these proteins. However, though they tested multiple antibodies, Jesse Cotari, Guillaume Voisinne and Gregoire Altan-Bonnet could not find one suitable for the flow cytometric analysis of JAK1. Consequently, the modelling efforts focus on the effects of IL-7Rα, JAK3, and γ_c abundances, which could be experimentally quantified, and the number of phosphorylated STAT5 molecules was estimated as proportional to the number of fully-formed signalling complexes IL-7Rα : γ_c :IL-7:JAK3, calculated from the equilibrium amounts of its molecular components.

3.2 A first IL-7R model (model 1)

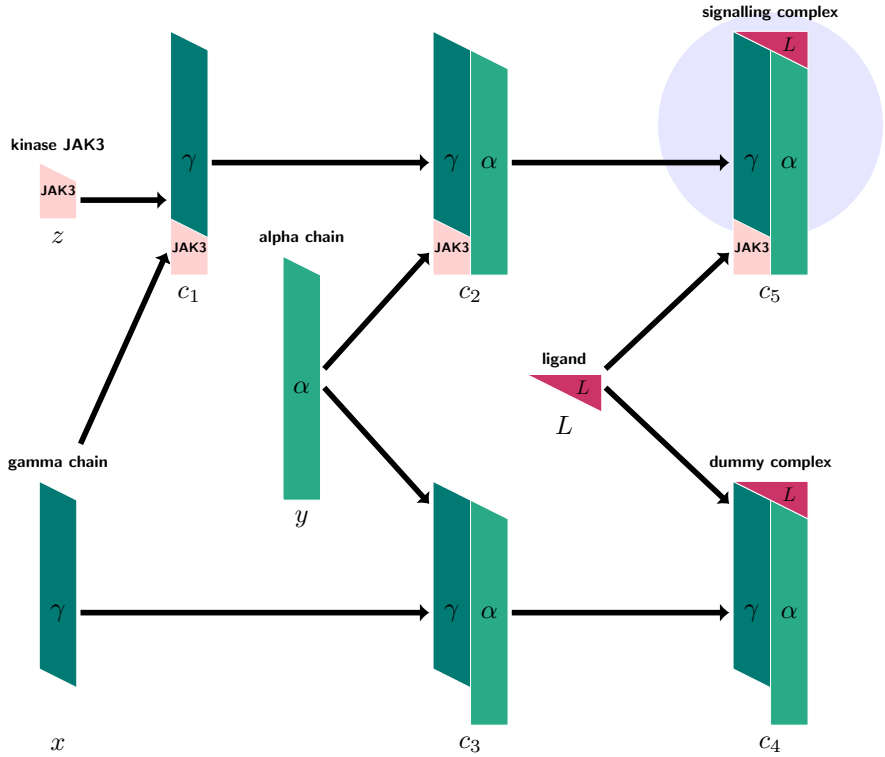
The experimental results present a paradox: high abundance of the γ_c chain, an essential IL-7 receptor component, is found to reduce the responsiveness to this cytokine, reflected both in a smaller signalling amplitude and a greater EC_{50}. To resolve this, we considered the fact that γ_c does not have intrinsic kinase ability, but rather depends on the kinase activity of a receptor-associated JAK3 to activate downstream signalling (Leonard *et al.*, 2019) (Fig. 3.5(a)). As such, Gregoire Altan-Bonnet and his team hypothesised that increased abundance of the γ_c chain alone could exert a dominant negative effect by acting as "dummy" subunits, binding to IL-7Rα chains to form signalling-deficient receptors (Fig. 3.5(b)). Such a negative effect of over-abundance has been previously reported in the case of specific T cells, which requires an associated downstream kinase (Lck, the lymphocyte specific protein tyrosine kinase) (Nakayama *et al.*, 1993).

To test this hypothesis, the mathematical model of IL-7 signalling presented in Ref. Cotari *et al.* (2013b) is modified to account for the binding of JAK3 to the γ_c chain.



(a) Signalling IL-7 complex

(b) Dummy IL-7 complex



(c) IL-7R model: sequential chemical reaction scheme

Figure 3.5: First IL-7R model (model 1): (a) The IL-7 receptor is composed of the trans-membrane γ_c and α chains. The γ_c chain can bind the intra-cellular downstream kinase JAK3. When the ligand, IL-7, binds the full receptor, it phosphorylates STAT5. (b) Model 1 allows the formation of "dummy" complexes: IL-7 bound IL-7R complexes, devoid of JAK3, which are unable to induce intra-cellular signalling. (c) IL-7 bound IL-7R complexes with JAK3 are able to induce intra-cellular signalling, and thus, are called "signalling" complexes. IL-7R "dummy" and signalling complexes are formed sequentially. The mathematical notation used in this section is shown below each molecule or complex. Note that complexes c_2 and c_3 are also referred to as signalling and "dummy" receptors, respectively.

3.2.1 Model description

Consider a model in which the IL-7R signalling complex (illustrated in Figure 3.5(a), where the hatched area determines the intra-cellular environment) is formed sequentially, one molecule at a time; the γ_c chain binds to the kinase, JAK3, then the α chain binds to the complex formed by γ_c and JAK3 (similarly to the crosslinking model in Ref. Ho & Harrington (2010)). Finally, the ligand, IL-7, also denoted by L, binds to the signalling receptor composed of γ_c, α and $JAK3$. The model also includes the formation of "dummy" receptors, which do not involve the kinase JAK3 (Fig. 3.5(b)). Figure 3.5(c) illustrates the sequential formation of the signalling and "dummy" complexes. The reaction scheme for this model, which we will refer to as model 1, is as follows

$$\begin{array}{lcll} \gamma_c + JAK3 & \rightleftharpoons & \gamma_c : JAK3, & K_1, \\ \alpha + \gamma_c : JAK3 & \rightleftharpoons & \alpha : \gamma_c : JAK3, & K_2, \\ \alpha + \gamma_c & \rightleftharpoons & \alpha : \gamma_c, & K_2, \\ L + \alpha : \gamma_c & \rightleftharpoons & L : \alpha : \gamma_c, & K_3, \\ L + \alpha : \gamma_c : JAK3 & \rightleftharpoons & L : \alpha : \gamma_c : JAK3, & K_3, \end{array} \tag{3.3}$$

where for $i = 1, 2, 3$, K_i is the affinity constant of the appropriate reaction. We assumed that the binding (or not) of JAK3 to γ_c has no influence on the binding of the other receptor chains (no allostery): that is the chemical reactions involved in the formation of the "dummy" or signalling complexes have the same affinity constants. One can show that this system has deficiency zero and is reversible (see Section 2.3). Therefore, for every set of rate constants and initial conditions, there exists exactly one positive steady state. Moreover, this positive steady state is in detailed balance. We remind the reader that we assume mass action kinetics to determine reaction rates. We denote the concentration of γ_c, α, JAK3 and IL-7 by x, y, z, and L, respectively. The reaction rate for the forward/backward reaction ($\rightharpoonup/\leftharpoondown$) is given by k_i and q_i, respectively, for $i = 1, 2, 3$. We note that $K_i = k_i/q_i$. The concentrations of the product complexes of the forward reactions are denoted by c_i in order of appearance (see Figure 3.5(c)). We can now write down the ordinary differential equations (ODEs) associated to the system of reactions (3.3):

$$\frac{dx}{dt} = -k_1 xz + q_1 c_1 - k_2 xy + q_2 c_3,$$

$$\begin{aligned}
\frac{dy}{dt} &= -k_2 y c_1 + q_2 c_2 - k_2 x y + q_2 c_3, \\
\frac{dz}{dt} &= -k_1 x z + q_1 c_1, \\
\frac{dc_1}{dt} &= k_1 x z - q_1 c_1 - k_2 y c_1 + q_2 c_2, \\
\frac{dc_2}{dt} &= k_2 y c_1 - q_2 c_2 - k_3 c_2 L + q_3 c_5, \\
\frac{dc_3}{dt} &= k_2 x y - q_2 c_3 - k_3 c_3 L + q_3 c_4, \\
\frac{dc_4}{dt} &= k_3 c_3 L - q_3 c_4, \\
\frac{dc_5}{dt} &= k_3 c_2 L - q_3 c_5.
\end{aligned} \tag{3.4}$$

We note that the ligand is assumed to be in excess and its concentration is thus a parameter of the model. A suitable basis for the conservation equations is

$$\begin{aligned}
N_x &= x + c_1 + c_2 + c_3 + c_4 + c_5, \\
N_y &= y + c_2 + c_3 + c_4 + c_5, \\
N_z &= z + c_1 + c_2 + c_5,
\end{aligned} \tag{3.5}$$

that is, single chain molecules are conserved since we do not consider the generation or degradation of molecules. The constants N_x, N_y and N_z represent the total copy number of γ_c, α and JAK3 molecules per cell, respectively. Detailed balance leads to the following steady state equations:

$$\begin{aligned}
c_1 &= K_1 x z, \\
c_2 &= K_2 y c_1, \\
c_3 &= K_2 x y, \\
c_4 &= K_3 L c_3, \\
c_5 &= K_3 L c_2.
\end{aligned} \tag{3.6}$$

Substituting the steady state equations into the conservation equations, we obtain

a system of polynomials

$$\begin{aligned} 0 &= -N_x + x + K_1xz + K_2K_1xyz + K_2xy + K_3K_2Lxy + K_3K_2K_1Lxyz, \\ 0 &= -N_y + y + K_2K_1xyz + K_2xy + K_3K_2Lxy + K_3K_2K_1Lxyz, \\ 0 &= -N_z + z + K_1xz + K_2K_1xyz + K_3K_2K_1Lxyz. \end{aligned} \tag{3.7}$$

3.2.2 Mathematical analysis: steady state, amplitude and EC_{50} expression

The polynomial system (3.7) can be solved numerically for a particular set of parameter values. However, an analytic solution will provide greater insight and will allow us to derive expressions for the amplitude and the EC_{50}.

Analytic computation of the steady state

We make use of Macaulay2 (Grayson & Stillman) to compute a lex Gröbner basis for this model, which will lead to a triangular set of polynomials [1], as follows:

$$0 = z^2 + \frac{[1 + K_1(N_x - N_z)]}{K_1}z - \frac{N_z}{K_1}, \tag{3.8a}$$

$$0 = y^2 + \frac{[1 + K_2(K_3L + 1)(N_x - N_y)]}{K_2(K_3L + 1)}y - \frac{N_y}{K_2(K_3L + 1)}, \tag{3.8b}$$

$$0 = x - \frac{1}{N_x}yz - \frac{(N_x - N_z)}{N_x}y - \frac{(N_x - N_y)}{N_x}z - \frac{N_x(N_x - N_y - N_z) + N_yN_z}{N_x}. \tag{3.8c}$$

Equation (3.8c) gives

$$x = \frac{(N_x - N_y + y)(N_x - N_z + z)}{N_x} = \frac{N_x - N_y + y}{1 + K_1z},$$

where the last equality follows from equation (3.8a). Solving the system (3.8) and selecting the biologically meaningful solution, we obtain an analytic expression for the number of free (unbound) JAK3, α and γ_c molecules at steady state

$$z = \frac{-1 + K_1(N_z - N_x) + \sqrt{\Delta_1}}{2K_1}, \tag{3.9a}$$

[1] Example code is provided in Appendix A.

$$y = \frac{-1 + K_2(N_y - N_x)(K_3L + 1) + \sqrt{\Delta_2}}{2K_2(K_3L + 1)}, \tag{3.9b}$$

$$x = \frac{N_x - N_y + y}{1 + K_1 z}, \tag{3.9c}$$

where we have introduced

$$\Delta_1 = 4K_1N_z + \left[K_1(N_x - N_z) + 1\right]^2,$$

and

$$\Delta_2 = 4K_2N_y\left(K_3L + 1\right) + \left[K_2(N_x - N_y)(K_3L + 1) + 1\right]^2.$$

We study the dose-response curve of this model given by the number of signalling complexes, $L : \gamma_c : \alpha : JAK3$, per cell at steady state as a function of L. The signalling function, $\sigma(L)$, is given by the steady state equations (3.6):

$$\sigma(L) \equiv c_5 = K_3K_2K_1Lxyz. \tag{3.10}$$

.

Analytic computation of the amplitude

A simple inspection of the behaviour of (3.10) shows that the dose-response curve is a sigmoid, such that $\sigma(0) = 0$. Therefore the amplitude A is given by the asymptotic behaviour of the signalling function as follows:

$$A \equiv \lim_{L \to +\infty} \sigma(L). \tag{3.11}$$

We will prove this result rigorously for a more general class of models in Section 4.3.

We first notice that z is independent of L. We now compute the product xy (at steady state) as follows

$$xy = \frac{(N_x - N_y)y + y^2}{1 + K_1 z}.$$

From equation (3.8b) we can replace the polynomial in y of degree two by an

expression linear in y:

$$(N_x - N_y)y + y^2 = \frac{N_y - y}{K_2(K_3L + 1)}.$$

Thus, we obtain the following analytic expression for the signalling function:

$$\sigma(L) = K_3K_2K_1Lxyz = \frac{K_1z}{(1 + K_1z)}\frac{K_3L}{(K_3L + 1)}(N_y - y). \tag{3.12}$$

Since $\frac{K_3L}{1+K_3L} \to 1$ when $L \to +\infty$, we need to study the expression $N_y - y$ in this limit. We have

$$N_y - y = \frac{(N_y + N_x)K_2(K_3L + 1) + 1 - \sqrt{\Delta_2}}{2K_2(K_3L + 1)}, \tag{3.13}$$

where

$$\Delta_2 = K_2^2(K_3L + 1)^2(N_x - N_y)^2 + 2K_2(K_3L + 1)(N_x + N_y) + 1.$$

Keeping to lowest order in $\mathcal{O}(\frac{1}{L})$ we obtain

$$\begin{aligned} N_y - y &= \frac{1 + (N_x + N_y)K_2(K_3L + 1) - K_2(K_3L + 1)|N_x - N_y|(1 + \mathcal{O}(\frac{1}{L}))}{2K_2(K_3L + 1)}, \\ &= \frac{N_x + N_y - |N_x - N_y|}{2} + \mathcal{O}(\frac{1}{L}). \end{aligned} \tag{3.14}$$

Finally, noticing that

$$\frac{N_x + N_y - |N_x - N_y|}{2} = \min(N_x, N_y),$$

we obtain the amplitude

$$A = \min(N_x, N_y)\frac{K_1z}{1 + K_1z}, \tag{3.15}$$

where z is the analytic expression obtained in (3.9). This result indicates that the amplitude of this model is the total number of the limiting trans-membrane chain modulated by a factor, valued in the interval $[0, 1]$, which only depends on K_1,

N_x and N_z.

Analytic computation of the EC_{50}

We now determine the EC_{50} by finding the value of L_{50} such that

$$\sigma(L_{50}) = \frac{A}{2} = K_1 K_2 K_3 L_{50} x_{50} y_{50} z_{50}, \tag{3.16}$$

where x_{50}, y_{50} and z_{50} are the steady state expressions found in (3.9) evaluated at $L = L_{50}$. Let us write $M = \min(N_x, N_y)$. We make use of the expression for $\sigma(L)$, the signalling function described in (3.12), and equation (3.13), to isolate the square root in equation (3.16).

$$\sqrt{\Delta_2} = K_2(K_3 L_{50} + 1)(N_x + N_y) + 1 - \frac{K_2(K_3 L_{50} + 1)^2 M}{K_3 L}, \tag{3.17}$$

with $M = \min(N_x, N_y)$. We square the equation to remove the root and simplify the expression to obtain

$$\begin{aligned} 0 = {} & 4K_2^2(K_3 L_{50} + 1)^2 N_x N_y + K_2^2(K_3 L_{50} + 1)^4 \frac{M^2}{K_3^2 L_{50}^2} \\ & - 2\frac{K_2^2(K_3 L_{50} + 1)^3 M(N_x + N_y)}{K_3 L_{50}} - 2\frac{K_2(K_3 L_{50} + 1)^2 M}{K_3 L_{50}}. \end{aligned} \tag{3.18}$$

Since we are looking for a positive value of L_{50}, we divide by $K_2(K_3 L_{50} + 1)^2$ and rewrite the previous expression as follows:

$$\begin{aligned} 0 = {} & 4K_2 K_3^2 L_{50}^2 N_x N_y + K_2(K_3 L_{50} + 1)^2 M^2 \\ & - 2K_2(K_3 L_{50} + 1) M(N_x + N_y) K_3 L_{50} - M K_3 L_{50}. \end{aligned} \tag{3.19}$$

It leads to a polynomial of degree 2 in L_{50},

$$0 = M^2 K_2 + 2K_3 L_{50} M(-1 + K_2(M - N_x - N_y)) + K_3^2 K_2 L_{50}^2 (M - 2N_x)(M - 2N_y). \tag{3.20}$$

The discriminant of this polynomial is positive:

$$\Delta = [1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - M)]4K_3^2M^2, \tag{3.21}$$

so that there are two potential solutions (both roots are positive):

$$\begin{aligned} L_{50}^+ &= M\frac{1 + K_2(N_x + N_y - M) + \sqrt{1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - M)}}{K_2K_3(M - 2N_x)(M - 2N_y)}, \\ L_{50}^- &= M\frac{1 + K_2(N_x + N_y - M) - \sqrt{1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - M)}}{K_2K_3(M - 2N_x)(M - 2N_y)}. \end{aligned} \tag{3.22}$$

Two solutions exist since by squaring equation (3.17) we lose the positive steady state (x, y, z, L) hypothesis. Substituting these expressions back into the steady state equations shows that only L_{50}^+ leads to a biologically meaningful solution $(L, x, y, z > 0)$. The relevant analytic expression of the EC_{50} is given by

$$EC_{50} = M\ \frac{1 + K_2(N_x + N_y - M) + \sqrt{1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - M)}}{K_2K_3(M - 2N_x)(M - 2N_y)} \tag{3.23}$$

with $M = \min(N_x, N_y)$.

This result shows that the EC_{50} value for this system is independent of the kinase, since the parameters K_1 and N_z are absent in the previous expression.

Alternatively, we now propose a more algebraic method to derive the analytic expression of the EC_{50}. We compute a lex Gröbner basis for the augmented system of polynomials consisting of the steady state equations (3.7) and

$$K_1K_2K_3Lxyz\ (1 + K_1z) - \frac{MK_1z}{2} = 0, \tag{3.24}$$

where this time x, y, z, and L are variables. The resulting triangular system describes directly L_{50} and x_{50}, y_{50}, z_{50} (defined in equation (3.16)), *i.e.* the EC_{50} and x, y, z at $L = EC_{50}$:

$$0 = L^2 + \frac{2M[-1 + K_2(M - N_x - N_y)]}{K_2K_3(M - 2N_x)(M - 2N_y)}L + \frac{M^2}{K_3^2(M - 2N_x)(M - 2N_y)}, \tag{3.25a}$$

$$0 = z^2 + \frac{1 + K_1(N_x - N_z)}{K_1}z - \frac{N_z}{K_1}, \tag{3.25b}$$

$$0 = y - \frac{K_3(M-2N_x)(M-2N_y)}{2M}L + \frac{K_2(2N_x-M)+2}{2K_2}, \tag{3.25c}$$

$$0 = x - \frac{K_3(M-2N_x)(M-2N_y)(N_x-N_z+z)}{2MN_x}L + \frac{[2+K_2(2N_y-M)](N_x-N_z+z)}{2K_2N_x}. \tag{3.25d}$$

Solving (3.25a) and selecting the solution for which y and x, given by equations (3.25c) and (3.25d), respectively, are positive yields the final result, in agreement with (3.23). The use of the algebraic method is more elegant as it gives directly the correct EC_{50} expression.

3.2.3 Model validation

The analytic study of this IL-7R mathematical model shows that the abundance of JAK3 has an influence on the amplitude but not on the EC_{50}, which is in accordance with the experimental results (see Figure 3.3).

A short numerical analysis has been conducted to explore how γ_c abundance impacts the cytokine response (Figure 3.6). Model 1 demonstrated that the over-abundance of γ_c chains could indeed inhibit the IL-7 signalling response and thus reduce its amplitude (see Figure 3.6(a)). Namely, both the number of signalling complexes and non-signalling (or "dummy") complexes grow linearly as the number of γ_c chains increases (Figure 3.6(b) shows the number of signalling and "dummy" complexes at high ligand concentration, so that the number of signalling complexes is the amplitude). When the number of γ_c equals the number of IL-7Rα chains, the amplitude of the IL-7 response (number of signalling complexes at high ligand concentration) reaches its maximum. As the abundance of γ_c exceeds that of IL-7Rα, the number of signalling complexes decreases and that of "dummy" complexes, in turn, increases. We observe (proof in a more general case in Section 4.3) that the sum of the "dummy" and signalling complexes is always equal to the total copy number of the limiting component (γ_c during the linear growth and IL-7Rα after the maximum amplitude has been reached). This implies that once the number of γ_c chains exceeds that of IL-7Rα, all IL-7Rα chains are ligand– and γ_c–bound to form either a signalling or a "dummy" complex. When the number of γ_c chains (N_x) exceeds that of JAK3 (N_z), the number of signalling complexes

(here the amplitude A) goes back to zero rather quickly (proportionally to the inverse of the total number of γ_c chains).

Proposition 70. The amplitude decreases proportionally to $\frac{1}{N_x}$ as $N_x \to +\infty$.

Proof. Consider $h = \frac{1}{K_1(N_x - N_z)}$. Then,

$$z = \frac{-1 - \frac{1}{h} + \frac{1}{h}\sqrt{1 + 2h + h^2(1 + 4K_1N_z)}}{2K_1}.$$

As $h \to 0$,

$$\begin{aligned} z &= \frac{-1 - \frac{1}{h} + \frac{1}{h}[1 + \frac{1}{2}(2h + h^2(1 + 4K_1N_z)) - \frac{1}{8}(2h + h^2(1 + 4K_1N_z))^2 + o(h^4)]}{2K_1} \\ &= \frac{-1 - \frac{1}{h} + \frac{1}{h}(1 + h + \frac{1+4K_1N_z-1}{2}h^2 + o(h^2))}{2K_1} \\ &= N_z h + o(h). \end{aligned}$$

Back to the original notation, as $N_x \to +\infty$, we obtain

$$z \underset{N_x \to +\infty}{\sim} \frac{N_z}{K_1(N_x - N_z)} \underset{N_x \to +\infty}{\sim} \frac{N_z}{K_1 N_x}.$$

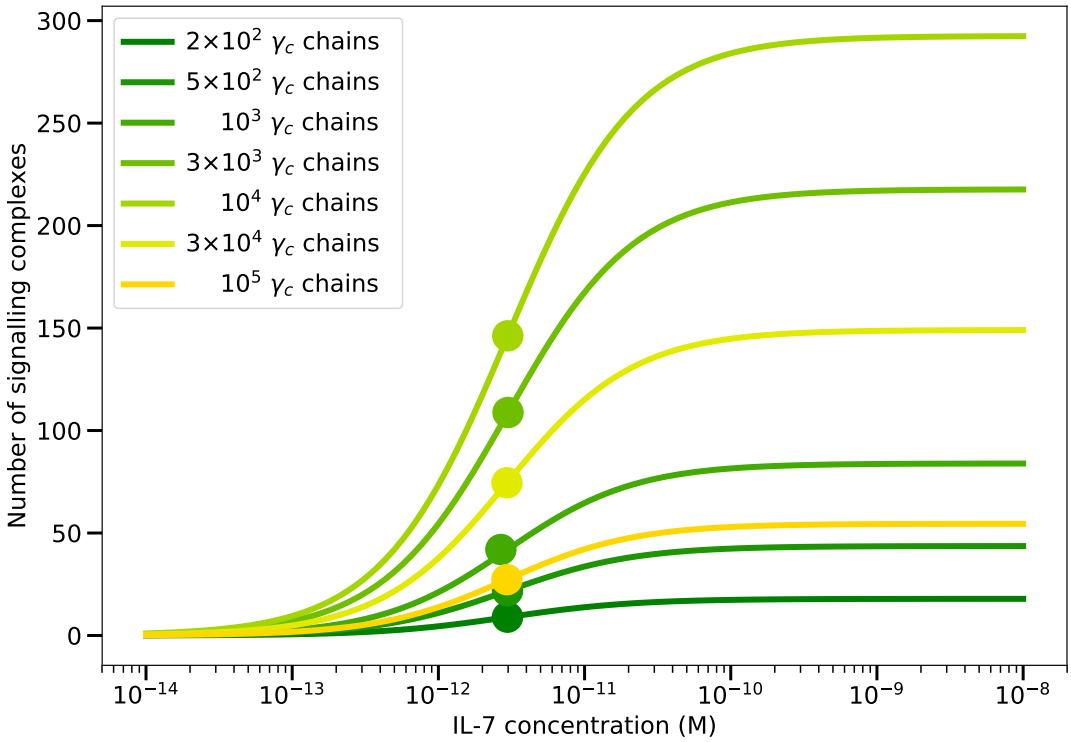
Thus, as $N_x \to +\infty$ (and so $\min(N_x, N_y) = N_y$),

$$A = \frac{K_1 z}{K_1 z + 1} N_y \underset{N_x \to +\infty}{\sim} \frac{N_z N_y}{N_x}.$$

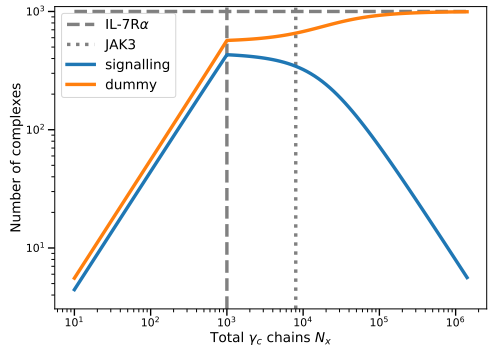
□

These observations shows that model 1 seems to be a good candidate to reproduce the experimental results on the amplitude, thus explaining the decrease in amplitude observed in excess of γ_c chains by the formation of "dummy" complexes. The EC_{50}, however, seems to be a more complicated feature that is not captured by this model: the modelled EC_{50} seems rather flat where we expect an increase of this quantity with the increase of γ_c chains abundance (Figure 3.6(c)).

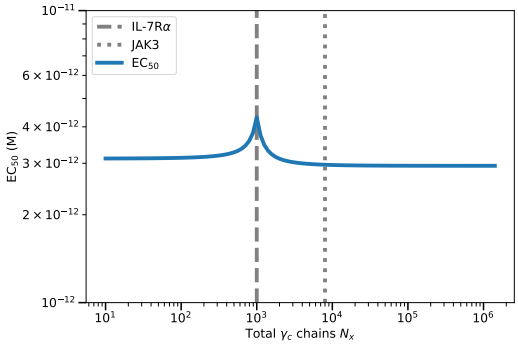
We now fit the model to the experimental data making use of approximate Bayesian computation (ABC) and infer the affinity constants K_1, K_2 and K_3. An introduction to ABC can be found in Section 2.8.



(a) Dose-response curve for different γ_c chains abundance



(b) Signalling and "dummy" complexes at high ligand concentration



(c) Modelled EC_{50}

Figure 3.6: Numerical exploration of the impact of γ_c chain abundance on the IL-7 response in model 1: (a) Dose-response curve for different γ_c abundances. The amplitude (the plateau of the dose-response curve) increases then decreases with the increase of γ_c abundance. The EC_{50} (dots) reminds mostly unchanged. (b) Number of signalling and "dummy" complexes at high ligand concentration ($L = 10^{-3}$ M). (b) The modelled EC_{50} is mostly flat. For this figure, the parameters has been fixed to 10^3 IL-7Rα chains per cell (as determined in Ref. Waickman *et al.* (2020)), 8×10^3 JAK3 molecules per cell (mean JAK3 abundance measured experimentally), $K_1 = 10^{-4}$, $K_2 = 17 \times 10^{-3}$ and $K_3 = 34 \times 10^{10}\text{M}^{-1}$. Values of K_2 and K_3 were taken from Ref. Cotari *et al.* (2013b).

Approximate Bayesian computation and results

Before inferring the parameters of the model, one has to make sure that we can compare the model output and the experimental data. Indeed, for a given parameter set, solving system (3.7) returns the steady state number of unbound single chain components. One can then multiply these quantities to obtain the number of signalling complexes formed (Equation (3.10)). However, the original experimental data set is composed of STAT5 phosphorylation level. As mentioned in Section 3.1, in a previous work (Cotari *et al.*, 2013b), Jesse Cotari, Guillaume Voisinne and Gregoire Altan-Bonnet demonstrated that the abundance of pSTAT5 correlated linearly with the number of IL-7 molecules bound to the surface of the cells. Hence, for any ligand concentration L, the fraction of phosphorylated STAT5 molecules, $pSTAT5(L)$, and the number of signalling complexes, $\sigma(L)$ follow an affine relation with parameters ξ_1 and ξ_2 for any ligand concentration L, as introduced in equation (3.2):

$$pSTAT5(L) = \xi_1 \sigma(L) + \xi_2,$$

where pSTAT5 is a STAT5 phosphorylation level measured in the experiment for given γ_c, IL-7Rα and JAK3 abundances, and $\sigma(L)$ is the number of signalling complexes (Equation (3.10)) returned by the model evaluated at the same chain abundances, for the ligand concentration L. Since we expect to only compare the experimental and modelled amplitude and EC_{50}, let us investigate how this affine transformation on σ affects the two pharmacological quantities. The amplitude A is defined as $A = \max(\sigma) - \min(\sigma)$. Thus, the amplitude experimentally measured, A', is proportional to the modelled amplitude, A:

$$\begin{aligned} A' &= \max(pSTAT5) - \min(pSTAT5) \\ &= \xi_1 \max(\sigma) + \xi_2 - (\xi_1 \min(\sigma) + \xi_2) \\ &= \xi_1 (\max(\sigma) - \min(\sigma)) \\ &= \xi_1 A. \end{aligned} \tag{3.26}$$

One can also show that the EC_{50} is independent of the parameters ξ_1 and ξ_2. Indeed, by definition of the EC_{50} (see Section 2.4), we have:

$$\begin{aligned} & pSTAT5(EC_{50}) = \frac{A'}{2} + \min(pSTAT5) \\ \iff \quad & \xi_1\sigma(EC_{50}) + \xi_2 = \frac{\xi_1 A}{2} + \xi_1 \min(\sigma) + \xi_2 \\ \iff \quad & \xi_1\sigma(EC_{50}) = \xi_1(\frac{A}{2} + \min(\sigma)) \\ \iff \quad & \sigma(EC_{50}) = \frac{A}{2} + \min(\sigma). \end{aligned} \tag{3.27}$$

Hence, to compare the modelled and experimental amplitude and EC_{50}, one only needs the parameter ξ_1 to rescale the modelled amplitude.

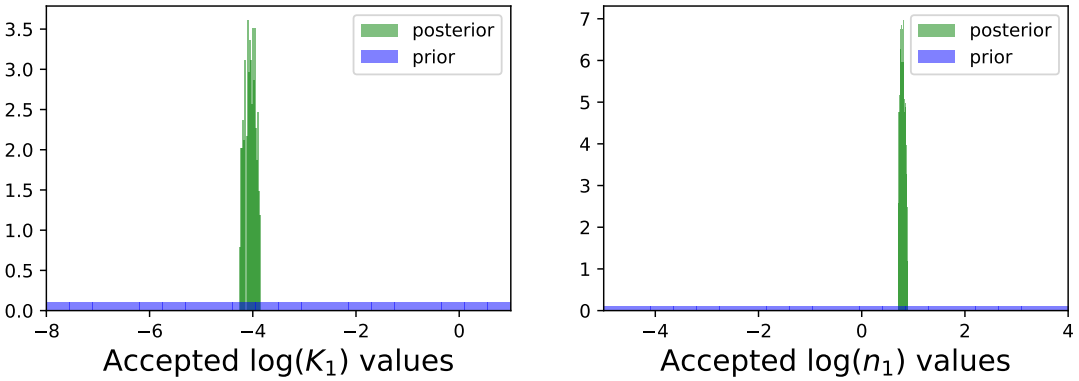
To perform ABC, one has to simulate the model multiple times, each time exploring a new parameter set. Bayesian computation can thus be numerically cumbersome. However, here, the use of the analytic expression of the amplitude and EC_{50}, obtained in Section 3.2.2, dramatically reduces computation time and increases numerical precision. Indeed, one can just substitute the parameter values in the analytic expression, instead of computing, for each set of parameter values, the dose-response curve, the amplitude and finally the EC_{50}. Moreover, the mathematical analysis shows that the amplitude and the EC_{50} do not depend on the same parameters: the amplitude only depends on K_1 and ξ_1 while only K_2 and K_3 are present in the EC_{50} expression. This separation of parameters facilitates parameter exploration and inference, as we can conduct ABC on the amplitude and EC_{50} separately. It also means that we infer only two parameters at a time, which effectively increases the convergence speed of the parameter inference (*i.e.*, we can sample fewer parameter values in the prior to observe a significant posterior).

I decided not to make any assumption on the parameters and chose uniform prior distributions on the logarithm of the parameters: $\log(K_1) \sim \mathcal{U}(-8, 1)$, $\log(K_2) \sim \mathcal{U}(-10, 10)$, $\log(K_3) \sim \mathcal{U}(8, 12)$ and $\log(\xi_1) \sim \mathcal{U}(-4, 4)$. Note that I chose the range of these distributions in accordance with biologically relevant values (for instance, I expect K_2 and K_3 to be around 10^{-4} and $10^{11}\mathrm{M}^{-1}$, respectively,

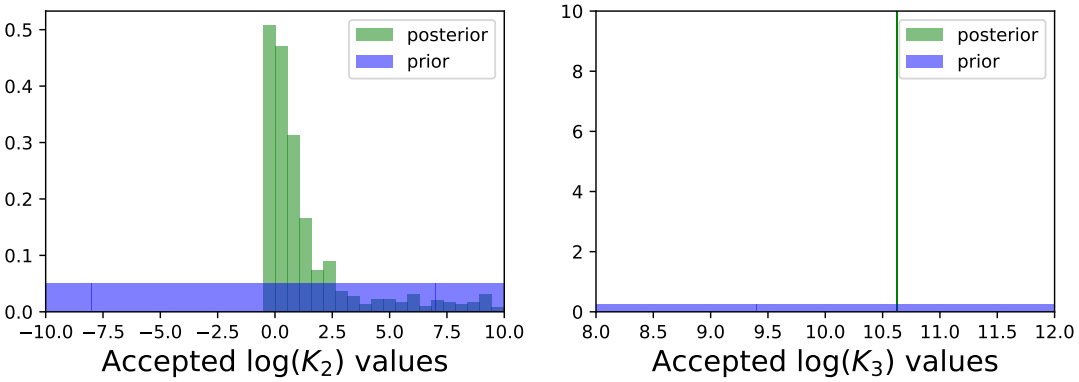
as determined in Ref. Cotari *et al.* (2013b)). I also adapted the ranges of the distributions after several test simulations to ensure not to obtain cropped posterior distributions. For each simulation, I sample a parameter set according to the uniform prior distributions defined above. Then, for each bin i of the data set, I extract the γ_c chain, IL-7Rα and JAK3 abundance levels which set the values for N_x, N_y and N_z, respectively, and simulate the model (making use of the analytic expressions obtained in the previous section for the amplitude and EC_{50}). The following distance d^2 is computed:

$$d^2 = \sum_{i=1}^{N_{bin}} (model[i] - data[i])^2, \tag{3.28}$$

where $N_{bin} = 172$ is the number of bins, $model[i]$ corresponds to the modelled quantity (amplitude or EC_{50}) with the abundances of bin i and $data[i]$ is the amplitude or EC_{50} computed from the data set for bin i. I repeat these steps five million times and obtain a list of five million distance values corresponding to the five million different parameter sets tested. I select the 10^3 parameter sets that minimise the distance to obtain a posterior distribution. The normalised posterior distributions resulting from the two separated ABC (one fitting the amplitude and one fitting the EC_{50}) are shown in Figure 3.7. The Bayesian computation returns narrow distributions for all the parameters, with an extremely thin distribution for K_3 (the y-scale has been limited to 10 so that the prior distribution is observable but the peak goes up to 600) and a larger posterior for K_2. The correlations graphs (Figure 3.8) shows that K_1 and ξ_1 are correlated, which is expected as ξ_1 and K_1 are directly multiplied. On the contrary, it seems that K_2 and K_3 are totally uncorrelated. One has to be cautious when interpreting the posterior distributions of K_2 and K_3 as the model is unable to fit the experimental EC_{50} (Figure 3.9(b)). Indeed, the modelled EC_{50} is rather flat and does not reproduce the experimentally observed increase in sensitivity. Model 1 is nonetheless able to reproduce with a good accuracy the experimental amplitude (Figure 3.9(a)). For a more compact comparison, Figure 3.10 shows the colormap of the modelled amplitude and EC_{50} alongside with the data colormap provided in Section 3.1. The modelled quantities in Figure 3.9 and 3.10 were generated with the parameter values that



(a) Amplitude parameters

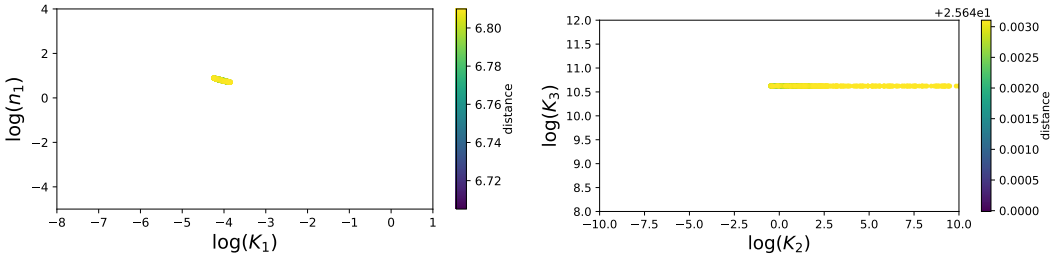


(b) EC_{50} parameters

Figure 3.7: Normalised prior (blue) distribution of the 5×10^6 parameter samples and posterior distribution (green) of the 10^3 best parameter values for model 1.

yield the smallest distance value: $K_1 = 10^{-4.04}$, $\xi_1 = 10^{0.79}$, $K_2 = 10^{-0.06}$ and $K_3 = 10^{10.62}\text{M}^{-1}$.

The parameter values obtained in this section vary slightly from those obtained by Guillaume Voisinne ($K_1 = 10^{-4.5}$, $K_2 = 10^{-0.09}$, $K_3 = 10^{10.6}\text{M}^{-1}$, $\xi_1 = 10^{0.9}$) and published in Ref. Sta *et al.* (2022b). This difference is explained by the use of different fitting strategies. Here we make use of ABC and, thanks to the analytic study, separated the fitting of the amplitude and EC_{50}. Guillaume Voisinne, however, found the amplitude and EC_{50} numerically after computation of the dose-response curve, and fitted these quantities together making use of a



(a) Amplitude parameters

(b) EC_{50} parameters

Figure 3.8: Correlation plots of the 10^3 best parameter values for model 1 as a function of the distance value.

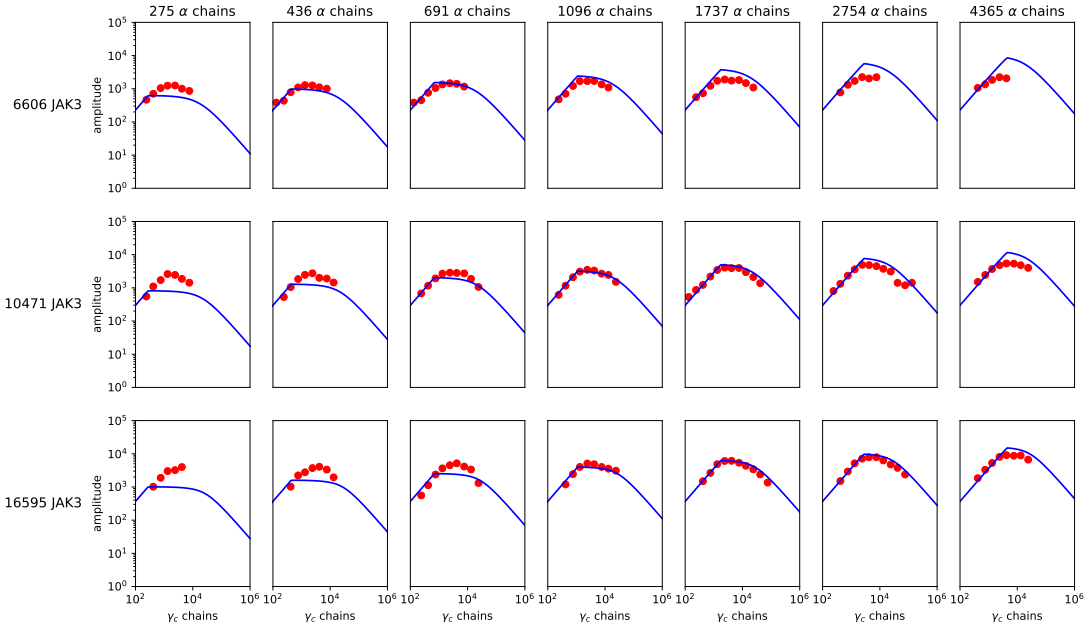
least-square method. The distance he minimised was

$$d^2_{GV} = \sum_{i=1}^{N_{bin}} \left(EC_{50}^{model}[i] - EC_{50}^{data}[i]\right)^2 + \sum_{i=1}^{N_{bin}} \left(\text{amplitude}^{model}[i] - \text{amplitude}^{data}[i]\right)^2, \tag{3.29}$$

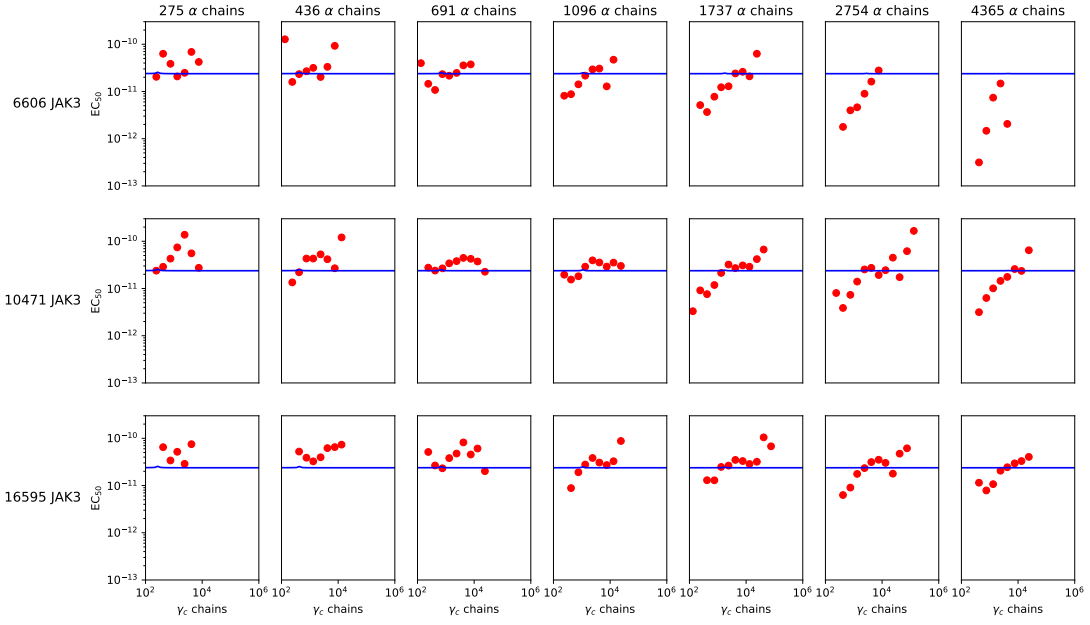
where the amplitude and EC_{50} values were in log-scale. To compare my work to his, I also fitted the parameters of model 1 by minimising d^2_{GV} in the ABC. I obtained parameter values very similar to the ones already obtained in this section (see Figure 3.11). In the three cases (ABC with separation of the amplitude and EC_{50}, Guillaume Voisinne's version and ABC with amplitude and EC_{50} fitted together), the parameter values obtained are of the same order of magnitude.

Conclusions

As demonstrated through Figures 3.9 and 3.10 in particular, model 1 is able to reproduce the amplitude behaviour but is unable to predict the observed effect of γ_c abundance on EC_{50}. This suggests that model 1 needs to be augmented to account for the peculiar increase of EC_{50} for cells with over-abundant expressions of γ_c chain. To this end, I examine two variations of this model. First, I explore how an hypothetical allosteric change induced by JAK3 binding to γ_c could limit binding of IL-7Rα to γ_c and the formation of the signalling IL-7 receptor complex. Then, I explore and analyse the case of model 1 in which we account for the fact that γ_c is a chain shared between many interleukin receptors (Rochman *et al.*, 2009).

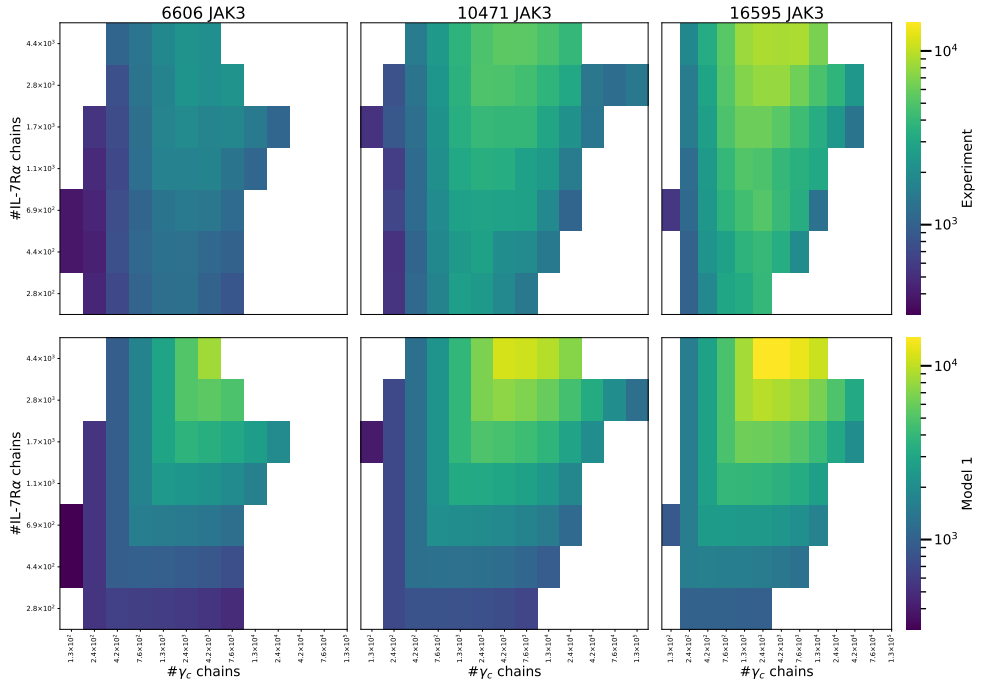


(a) Experimental and modelled amplitude (model 1)

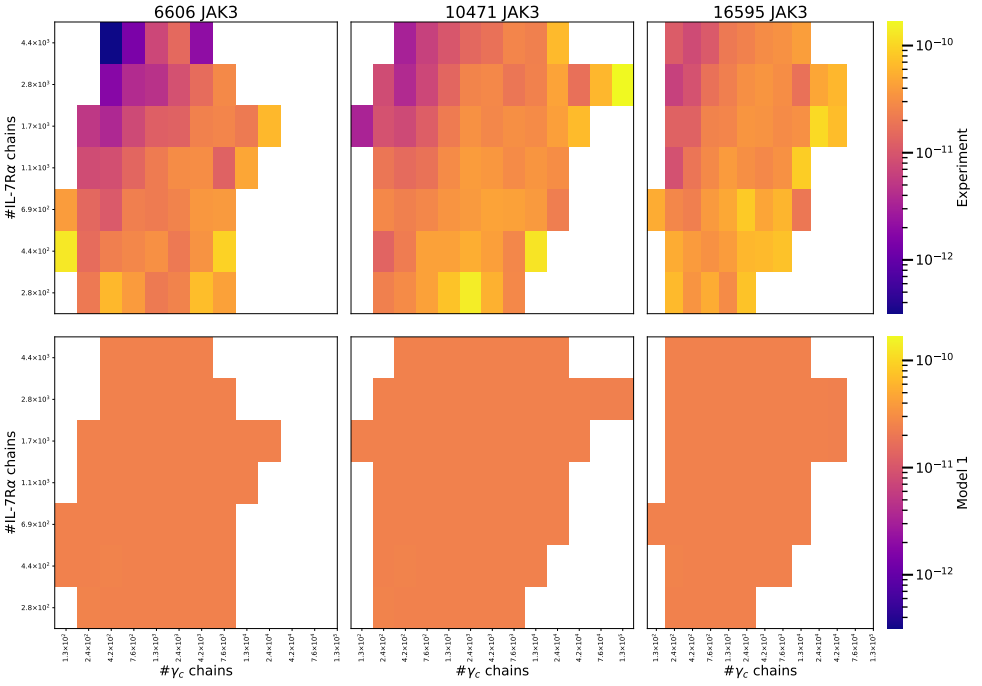


(b) Experimental and modelled EC_{50} (model 1)

Figure 3.9: Comparison of the experimental (red dots) and modelled amplitude and EC_{50} (model 1 in blue) for different levels of signalling components. The modelled quantities were generated with the best parameter values (those which minimise the distance): $K_1 = 10^{-4.04}$, $\xi_1 = 10^{0.79}$, $K_2 = 10^{-0.06}$ and $K_3 = 10^{10.62}\mathrm{M}^{-1}$.



(a) Experimental and modelled amplitude (model 1)



(b) Experimental and modelled EC_{50} (model 1)

Figure 3.10: Complete map of the experimental and modelled amplitude and EC_{50} (model 1) for different levels of signalling components. The modelled quantities were generated with the best parameter values (those which minimise distance): $K_1 = 10^{-4.04}$, $\xi_1 = 10^{0.79}$, $K_2 = 10^{-0.06}$ and $K_3 = 10^{10.62}\mathrm{M}^{-1}$.

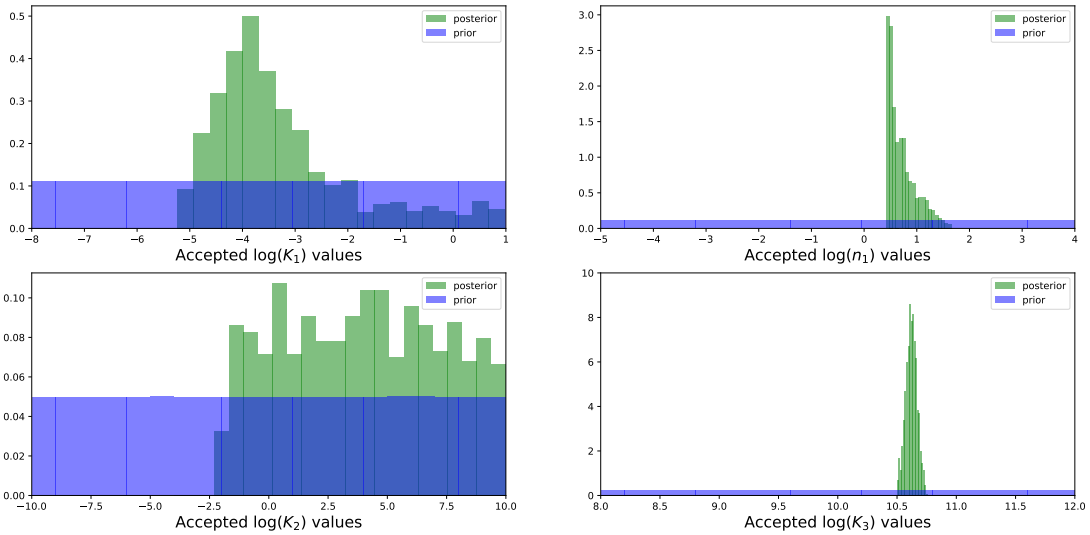


Figure 3.11: Posterior distribution of ABC minimising d^2_{GV} (Equation (3.29)). The best parameter values obtained are $K_1 = 10^{-4.08}$, $K_2 = 10^{3.62}$, $K_3 = 10^{10.6}\text{M}^{-1}$, $\xi_1 = 10^{0.81}$. Note that the posterior for K_2 is not significant. Prior distributions were chosen uniform in the interval showed in the figure.

3.3 IL-7R model with allostery

To increase the EC_{50}, the ability of the γ_c chain to form IL-7 signalling receptors must be reduced. The first hypothesis explored is that the formation of "dummy" complexes is faster than that of the signalling ones. This results in the monopolisation of the γ_c chain by non-signalling receptors.

3.3.1 Model description

Model 1 is revised to allow for allostery; that is, the affinity constants of the chemical reactions forming the "dummy" and signalling complexes are not equal anymore. The affinity constant for the reaction between the α and the γ_c chain is different from that of α and the complex formed by γ_c and JAK3 ($K_2 \neq K_2'$) and the constant for the binding of IL-7 to the "dummy" receptor is different from that of IL-7 to the signalling receptor ($K_3 \neq K_3'$). The reaction scheme is as follows:

$$\begin{array}{lcll} \gamma_c + JAK3 & \rightleftharpoons & \gamma_c : JAK3, & K_1, \\ \alpha + \gamma_c : JAK3 & \rightleftharpoons & \alpha : \gamma_c : JAK3, & K_2, \\ \alpha + \gamma_c & \rightleftharpoons & \alpha : \gamma_c, & K_2', \\ L + \alpha : \gamma_c & \rightleftharpoons & L : \alpha : \gamma_c, & K_3', \\ L + \alpha : \gamma_c : JAK3 & \rightleftharpoons & L : \alpha : \gamma_c : JAK3, & K_3, \end{array} \tag{3.30}$$

where the K_i and K_i' are the affinity constants. Figure 3.5 still illustrates this model, which we call allostery model. Proceeding similarly to model 1, a polynomial system, combination of steady state and conservation equations, can be written:

$$\begin{aligned} 0 &= -N_x + x + K_1xz + K_2K_1xyz + K_2'xy + K_3'K_2'Lxy + K_3K_2K_1Lxyz, \\ 0 &= -N_y + y + K_2K_1xyz + K_2'xy + K_3'K_2'Lxy + K_3K_2K_1Lxyz, \\ 0 &= -N_z + z + K_1xz + K_2K_1xyz + K_3K_2K_1Lxyz. \end{aligned} \tag{3.31}$$

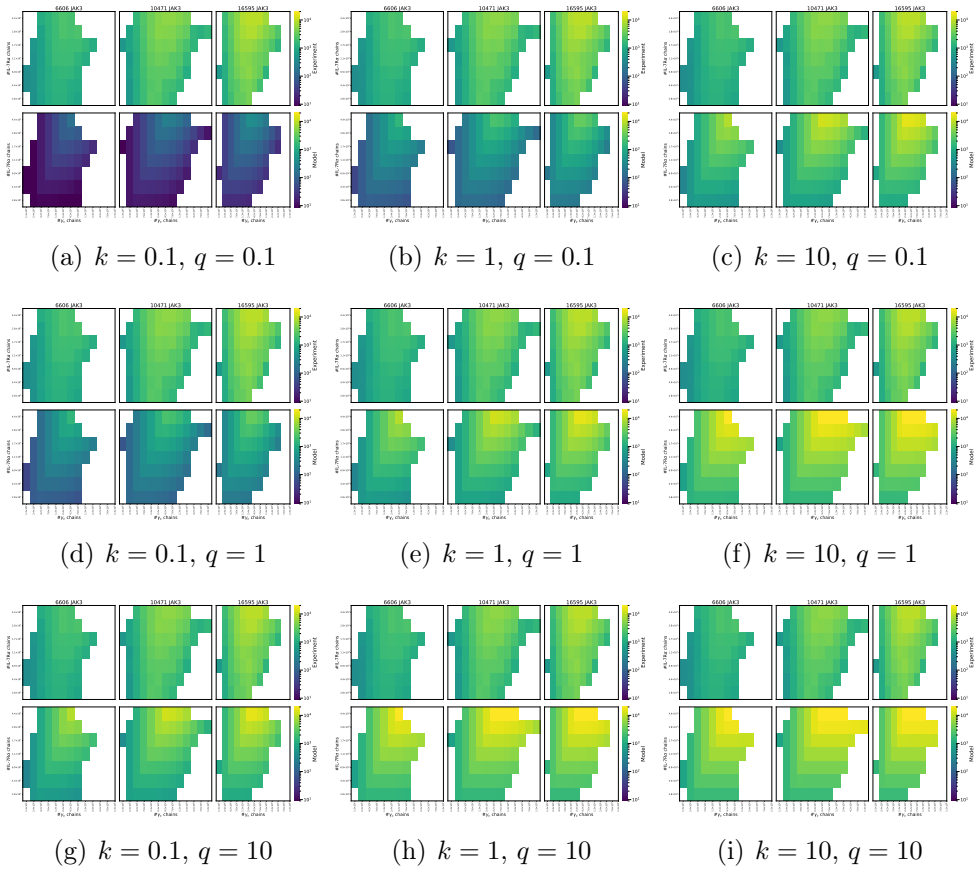
One could compute a Gröbner basis of this polynomial system and proceed to an analytic study of the allostery model. However, the Gröbner basis obtained is more complicated (see Appendix B) than the basis obtained for model 1, which makes the mathematical analysis more tedious. The numerical exploration (see below) shows that this model has little chance to be a good candidate to reproduce the experimental EC_{50} behaviour. Thus, I decided not to continue the mathematical study, and only make use of (3.31) to numerically compute the amplitude and EC_{50} of this model.

In the following section, the amplitude and EC_{50} of the allostery model are computed as follows. First, for a given set of parameters (K_1, K_2, K_3, K_2', K_3', N_x, N_y, N_z), and a fixed ligand concentration L, polynomial system (3.31) is solved to obtain the steady state expression of the unbound single chain components (x, y and z). These quantities are multiplied to obtain the number of formed signalling complexes $\sigma(L) = K_3K_2K_1Lxyz$. Keeping the same parameter set, this step is repeated for a range of ligand concentrations to obtain a dose-response curve. This curve is a sigmoid: it reaches a plateau at high concentration which is the amplitude of the model. The dose-response curve is then fitted to a Hill equation (Gesztelyi *et al.*, 2012; Goutelle *et al.*, 2008) to compute the EC_{50}.

3.3.2 Numerical exploration

To numerically explore whether the allostery model is able to reproduce the experimental amplitude and EC_{50} expression, I fixed all the parameters of the model ($K_1 = 10^{-4.04}$ and $\xi_1 = 10^{0.79}$ as inferred for model 1 (even though the amplitude of this model is *a priori different* from model 1), $K_2' = 17 \times 10^{-3}$ and $K_3' = 34 \times 10^{10}$ M $^{-1}$ as computed in Cotari *et al.* (2013b)) but varied the ratios, k and q, between the affinity constants of the reactions forming signalling and "dummy" complexes: $K_2 = k \times K_2'$ and $K_3 = q \times K_3'$. I computed the amplitude and the EC_{50} for k and q equal to 0.1, 1 and 10 (see Figure 3.12 and 3.13). Note that the graphs on the diagonal ($q = k$) of these figures are, by definition, outputs of model 1. Figure 3.12 shows that the allostery model is able to reproduce the experimental amplitude behaviour: the modelled amplitude increases then decreases as γ_c chain abundance increases (x-axis). However, as shown in Figure 3.13, the EC_{50} of the allostery model displays more intriguing behaviours. Indeed, for $q = 0.1$ (ignoring the diagonal graphs), the EC_{50} seems to decrease slightly for high abundance of γ_c chains. On the contrary, for $q = 10$ (ignoring the diagonal graphs), the EC_{50} seems to slightly increase (but, once again, only for high abundance of γ_c chains). In the other configurations, the EC_{50} seems rather flat. Surprisingly, the ratio k does not seem to have any influence on the EC_{50}. This result is unexpected as one could conjecture that diminished K_2 or K_3 values (*i.e.*, $q < 1$ or $k < 1$), compared to K_2' and K_3', would prioritise the formation of "dummy" complexes and reduce IL-7-induced response: the EC_{50} would increase rather than decrease. Similarly, it is natural to expect that larger binding rates implied in the formation of signalling complexes (*i.e.*, $q > 1$ or $k > 1$) would prioritise the formation of such complexes and thus decrease rather than increase the EC_{50}. This model confirms that the EC_{50} is not an intuitive quantity.

To summarise, this model with biochemical allostery still reproduces the observed amplitude and there exists a set of parameters where the EC_{50} does indeed increase with the abundance of γ_c overall. However the general pattern of the EC_{50} did not agree with the observed experimental evidence: the modelled EC_{50} increased only once the amount of γ_c was really large, while the experimental EC_{50} seemed to present a regular increase.

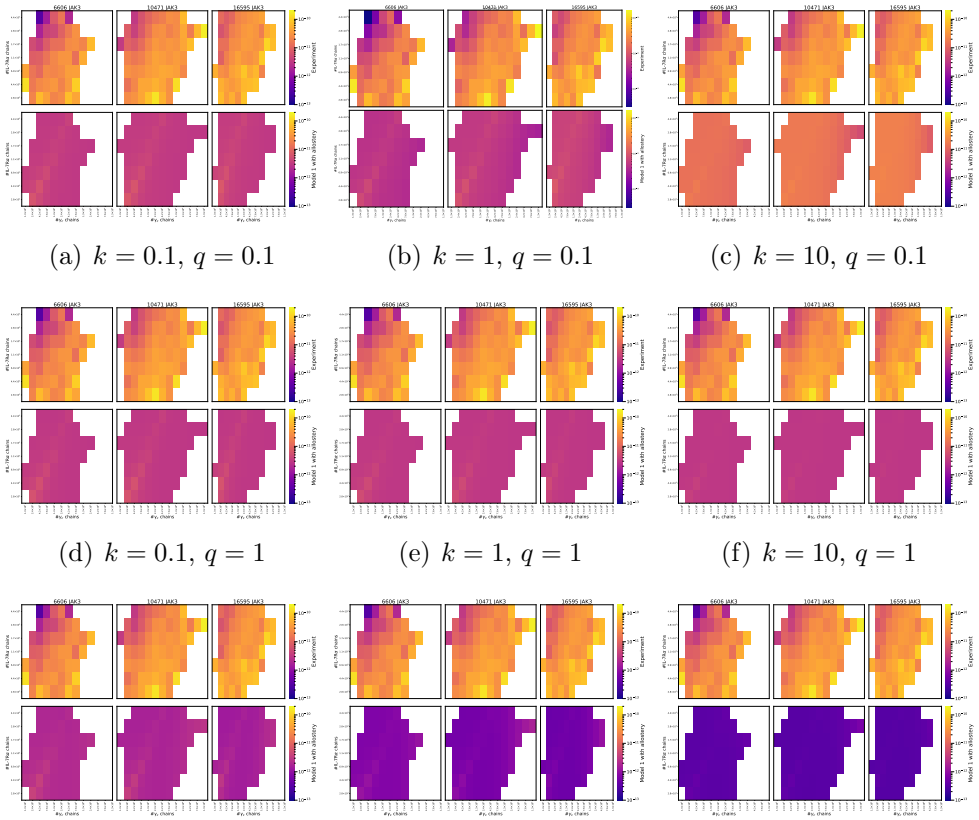


(a) $k = 0.1$, $q = 0.1$ (b) $k = 1$, $q = 0.1$ (c) $k = 10$, $q = 0.1$

(d) $k = 0.1$, $q = 1$ (e) $k = 1$, $q = 1$ (f) $k = 10$, $q = 1$

(g) $k = 0.1$, $q = 10$ (h) $k = 1$, $q = 10$ (i) $k = 10$, $q = 10$

Figure 3.12: Numerical exploration of the allostery model. The modelled amplitude (bottom row of each graph) is computed and compared to the experimental amplitude (top row of each graph) for different values of k and q defined as $K_2 = k \times K_2'$ and $K_3 = q \times K_3'$. We set $K_1 = 10^{-4.04}$ and $\xi_1 = 10^{0.79}$ as inferred in the previous section (even though the amplitude of this model is *a priori different*), $K_2' = 17 \times 10^{-3}$ and $K_3' = 34 \times 10^{10}$ M $^{-1}$ as computed in Cotari *et al.* (2013b).



(a) $k = 0.1$, $q = 0.1$ (b) $k = 1$, $q = 0.1$ (c) $k = 10$, $q = 0.1$

(d) $k = 0.1$, $q = 1$ (e) $k = 1$, $q = 1$ (f) $k = 10$, $q = 1$

(g) $k = 0.1$, $q = 10$ (h) $k = 1$, $q = 10$ (i) $k = 10$, $q = 10$

Figure 3.13: Numerical exploration of the allostery model. The modelled EC_{50} (bottom row of each graph) is computed and compared to the experimental EC_{50} (top row of each graph) for different values of k and q defined as $K_2 = k \times K_2'$ and $K_3 = q \times K_3'$. We set $K_1 = 10^{-4.04}$ as inferred in the previous section, $K_2' = 17 \times 10^{-3}$ and $K_3' = 34 \times 10^{10}$ M $^{-1}$ as computed in Cotari *et al.* (2013b).

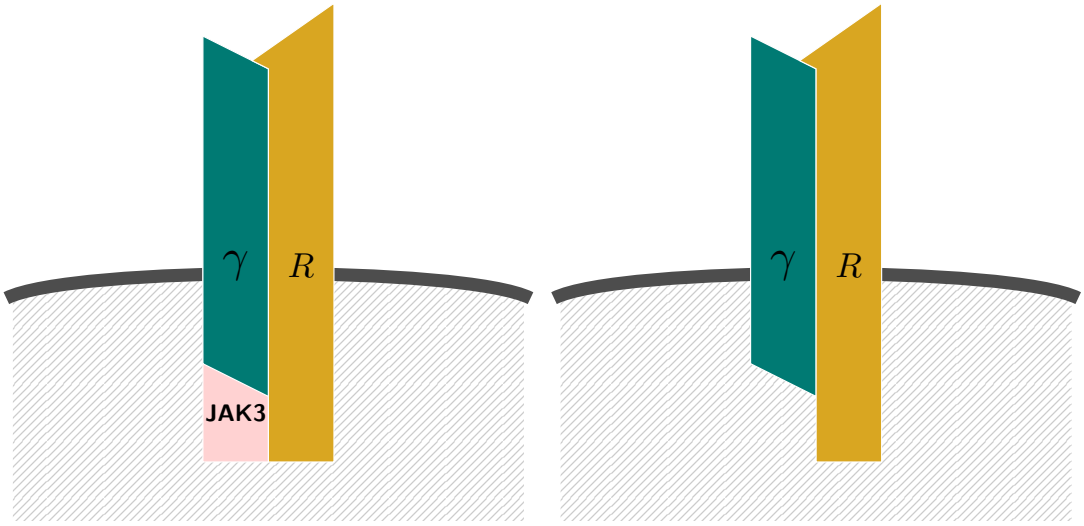
3.4 IL-7R model with an additional subunit (model 2)

The previous hypothesis of allostery, differences in binding caused by interactions between the components of the IL-7 receptor complex, did not lead to a mathematical model which could reproduce the experimental data. In this section, I modify model 1 to account for the fact that the γ_c chain is shared between multiple cytokine receptors (Rochman *et al.*, 2009). Indeed, the presence of other pre-formed receptors could limit the amount of γ_c chain available to form signalling IL-7 complexes.

3.4.1 Model description

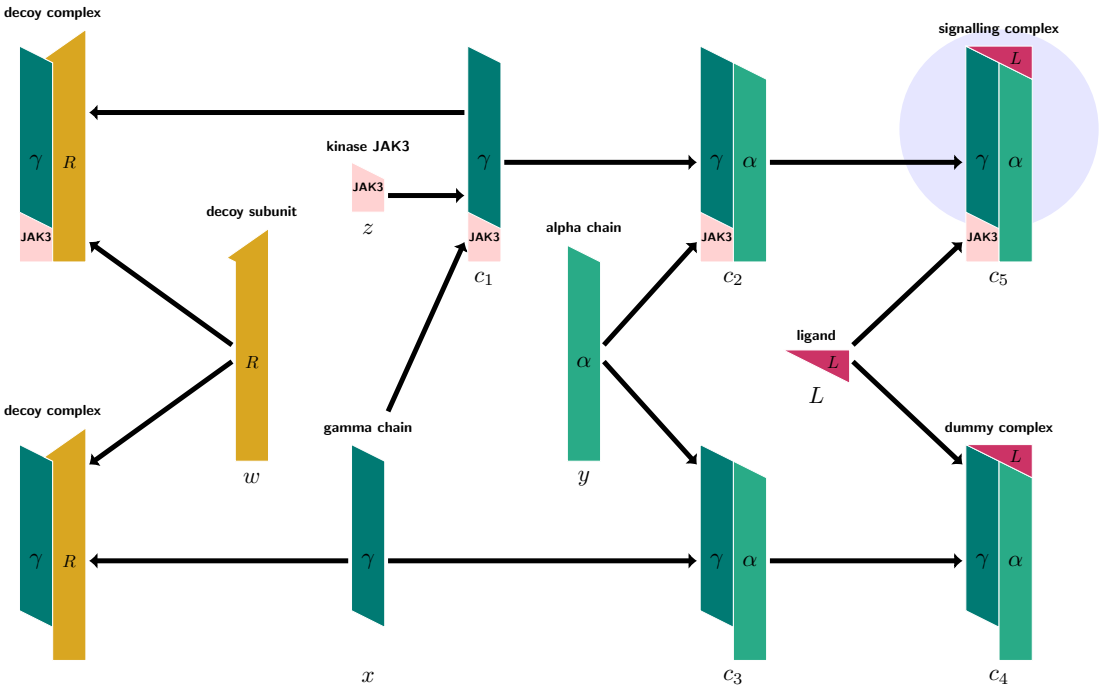
The first IL-7R model (model 1) described the IL-7 receptor system without any consideration to the fact that the γ_c chain is shared with other cytokine receptors (Rochman *et al.*, 2009). I now account for this competition by including in model 1 an additional receptor chain, R, which could bind to the γ_c chain, or the complex $\gamma_c : JAK3$ to form decoy receptor complexes (see Figure 3.14(a) and Figure 3.14(b), where the hatched area indicates the cytoplasmic region). Decoy receptors prevent the formation of signalling complexes by sequestering the γ_c chain away from IL-7Rα. Here, R could, for instance, be IL-2Rβ, IL-2Rβ:IL-2Rα or IL-4Rα. Alternatively, this additional chain could account for the set of all the receptor chains that bind to γ_c (except IL-7Rα). Similarly to model 1, we assume no allostery so that the binding rates related to the formation of the signalling and the "dummy" complexes are the same. The resulting reaction scheme of this new model, which we call model 2 (summarised in Figure 3.14(c)), is given by

$$
\begin{array}{lcll}
JAK3 + \gamma_c & \rightleftharpoons & JAK3 : \gamma_c & K_1, \\
\alpha + JAK3 : \gamma_c & \rightleftharpoons & \alpha : \gamma_c : JAK3 & K_2, \\
\alpha + \gamma_c & \rightleftharpoons & \alpha : \gamma_c & K_2, \\
L + \alpha : \gamma_c & \rightleftharpoons & L : \alpha : \gamma_c & K_3, \\
L + \alpha : \gamma_c : JAK3 & \rightleftharpoons & L : \alpha : \gamma_c : JAK3 & K_3, \\
R + \gamma_c & \rightleftharpoons & R : \gamma_c & K_4, \\
R + JAK3 : \gamma_c & \rightleftharpoons & R : \gamma_c : JAK3 & K_4.
\end{array}
$$



(a) Decoy receptor with kinase

(b) Decoy receptor without kinase



(c) Second IL-7R model sequential chemical reaction scheme

Figure 3.14: IL-7R model with an additional receptor subunit (model 2): The signalling and "dummy" complexes are the same as in model 1. This second model allows the formation of decoy receptors: (a) with the kinase JAK3, (b) or without the kinase. (c) The IL-7R "dummy" and signalling complexes are formed sequentially. Decoy complexes can be formed to prevent the formation of signalling or "dummy" complexes. The mathematical notation used in this section is annotated below each subunit or complex.

We use w to describe the concentration of the additional chain R and the notation of model 1 for the other variables. Similarly to model 1, we write the system of ODEs describing the time evolution for each complex and then derive (a basis for) the conservation and steady state equations. Combining them, we obtain the following polynomial system:

$$\begin{aligned}
0 &= -N_x + x + K_2xy + K_1xz + K_2K_1xyz + K_3K_2Lxy + K_3K_2K_1Lxyz \\
&\quad + K_4xw + K_1K_4xwz, \\
0 &= -N_y + y + K_2xy + K_2K_1xyz + K_3K_2Lxy + K_3K_2K_1Lxyz, \\
0 &= -N_z + z + K_1xz + K_2K_1xyz + K_3K_2K_1Lxyz + K_1K_4xwz, \\
0 &= -N_w + w + K_4xw + K_4K_1xwz,
\end{aligned} \tag{3.32}$$

where N_w is the additional conserved quantity.

Despite its mathematical analysis being more complex than the one of model 1, Model 2 seems to be able to reproduce the experimental amplitude and EC_{50} behaviour (see Section 3.4.3). Its analytic study is thus pursued.

3.4.2 Mathematical analysis of the amplitude and EC_{50}

We compute a lex Gröbner basis of polynomials (3.32) to obtain the following triangular system:

$$0 = K_1z^2 + z[1 + K_1(N_x - N_z)] - N_z, \tag{3.33a}$$
$$0 = Ay^3 + By^2 + Cy + D, \tag{3.33b}$$
$$0 = [K_2K_4(1 + K_3L)N_xN_y]\,x + (Ay^2 + By + C + K_4N_y)(N_x - N_z + z), \tag{3.33c}$$
$$0 = [K_2K_4(1 + K_3L)N_y]\,w + Ay^2 + By + [K_2(1 + K_3L) - K_4]N_y, \tag{3.33d}$$

where

$$\begin{aligned}
A &= -K_2(1 + K_3L)[K_2(1 + K_3L) - K_4], \\
B &= K_4 - K_2(1 + K_3L)[1 + K_4(N_w - N_x + 2N_y) + K_2(1 + K_3L)(N_x - N_y)], \\
C &= N_y[-2K_4 + K_2(1 + K_3L)(1 + K_4(N_w - N_x + N_y))],
\end{aligned}$$

$$D = K_4 N_y^2.$$

Solving (3.33a) gives the number of free JAK3 molecules per cell at steady state, z; solving (3.33b) gives the number of free (unbound) α chains per cell at steady state, y; and substituting y and z into (3.33c) and (3.33d) gives the remaining steady states. We obtain the following implicit steady state expressions for the number of free (unbound) chains:

$$\begin{aligned} z =& \frac{-1 + K_1(N_z - N_x) + \sqrt{[1 + K_1(N_x - N_z)]^2 + 4N_z K_1}}{2K_1}, \\ x =& -\frac{(Ay^2 + By + C + K_4 N_y)(N_x - N_z + z)}{K_2 K_4 (1 + K_3 L) N_x N_y}, \\ w =& -\frac{Ay^2 + By + [K_2(1 + K_3 L) - K_4] N_y}{K_2 K_4 (1 + K_3 L) N_y}. \end{aligned} \tag{3.34}$$

The problem now reduces to finding the positive real roots of (3.33b). As (3.33b) is a polynomial of degree three, we could, in principle, find an exact analytic solution. However, such a solution might not be very conclusive. Instead, we show how perturbation theory can be used to obtain the amplitude of the dose-response. In this model, the signalling complex is still $L : \alpha : \gamma_c : JAK3$. The signalling function is given by

$$\sigma(L) \equiv K_3 K_2 K_1 L x y z. \tag{3.35}$$

In Section 4.3.3 we will show that, for this model, the maximum of σ is attained in the limit $L \to +\infty$. Hence, we have

$$A \equiv \lim_{L \to +\infty} \sigma(L). \tag{3.36}$$

Combining (3.33a), written as $N_x - N_z + z = \frac{N_x}{1 + K_1 z}$, and (3.33b), we obtain a reduced expression for the product xy

$$xy = \frac{N_y - y}{K_2(1 + K_3 L)(1 + K_1 z)}, \tag{3.37}$$

which allows us to rewrite the amplitude as

$$A = \lim_{L\to+\infty} \frac{K_1 z}{(1+K_1 z)} \frac{K_3 L}{(K_3 L+1)} (N_y - y). \tag{3.38}$$

We note that z is independent of L and, therefore, to compute the amplitude we only need to find the behaviour of y as $L \to +\infty$.

Perturbation theory to determine y as $L \to +\infty$

We now apply the method described in Ref. Simmonds & Mann (2013) and summarised in Section 2.5. Let $\epsilon = \frac{1}{L}$ and define the polynomial P_ϵ as follows:

$$P_\epsilon(y) \equiv P_2(y)\epsilon^2,$$

where P_2 is the polynomial (3.33b). We added a factor of ϵ^2 to remove any negative powers of ϵ in P_2. We obtain the polynomial

$$P_\epsilon(y) = A_\epsilon y^3 + B_\epsilon y^2 + C_\epsilon y + D_\epsilon, \tag{3.39}$$

where

$$\begin{aligned}
A_\epsilon =& -K_2(\epsilon+K_3)[K_2(\epsilon+K_3)-\epsilon K_4],\\
B_\epsilon =& K_2^2K_3^2(N_y-N_x) - \epsilon K_2K_3[1+2K_2N_x-2K_2N_y+K_4(N_w-N_x+2N_y)]\\
&+\epsilon^2(K_4-K_2(1+K_2N_x-K_2N_y+K_4(N_w-N_x+2N_y))),\\
C_\epsilon =& \epsilon N_y[K_2K_3(1+K_4(N_w-N_x+N_y))\\
&+\epsilon(K_2-2K_4+K_2K_4(N_w-N_x+N_y))],\\
D_\epsilon =& K_4N_y^2\epsilon^2.
\end{aligned}$$

We now replace y by $\epsilon^p\omega(\epsilon)$ with $\omega(0) \neq 0$ and $p \in \mathbb{Q}$, according to theorem 49. We obtain

$$P_\epsilon(\epsilon^p\omega(\epsilon)) = A_{p,\epsilon}\omega^3 + B_{p,\epsilon}\omega^2 + C_{p,\epsilon}\omega + D_{p,\epsilon}, \tag{3.40}$$

where

$$A_{p,\epsilon} = -\epsilon^{3p}K_2(\epsilon+K_3)(K_2(\epsilon+K_3)-\epsilon K_4),$$

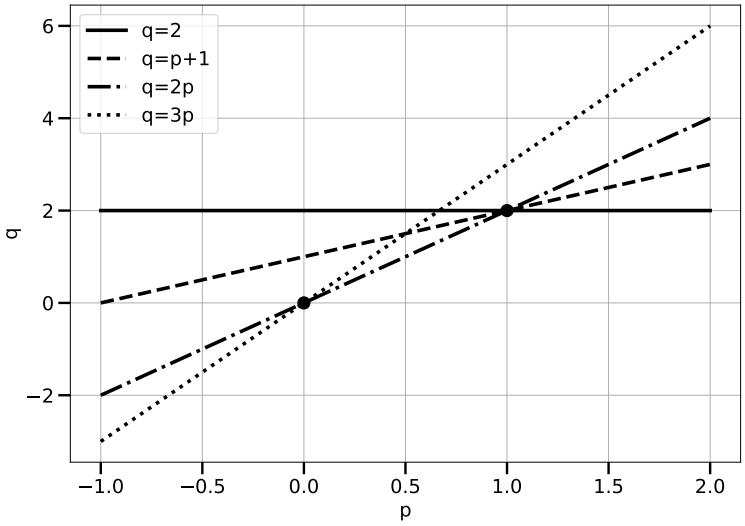


Figure 3.15: The lines defined in set E and the proper values, (black dots), computed following the graphical algorithm described in Section 2.5.

$$
\begin{aligned}
B_{p,\epsilon} =& \epsilon^{2p}(K_2^2K_3^2(N_y - N_x) - \epsilon K_2K_3(1 + 2K_2(N_x - N_y) + K_4(N_w - N_x + 2N_y)) \\
&+ \epsilon^2(K_4 - K_2(1 + K_2N_x - K_2N_y + K_4(N_w - N_x + 2N_y)))), \\
C_{p,\epsilon} =& \epsilon^{1+p}N_y(K_2K_3(1 + K_4(N_w - N_x + N_y)) \\
&+ \epsilon(K_2 - 2K_4 + K_2K_4(N_w - N_x + N_y))), \\
D_{p,\epsilon} =& K_4N_y^2\epsilon^2.
\end{aligned}
$$

The smallest exponents in the previous equation are

$$E = \{2, 1 + p, 2p, 3p\}.$$

We note that 0 is not in E because we multiplied P_2 by ϵ^2. Applying the graphical algorithm detailed in Section 2.5, we find the proper values $(0, 0)$ and $(1, 2)$ (see Figure 3.15). We investigate these two branches.

Branch (0,0). We make use of the notation in Section 2.5, to define

$$T_\epsilon^{(1)}(\omega) \equiv \epsilon^0 P_\epsilon(\omega\epsilon^0).$$

The least common denominator of {2,1,0,0} is $q_1 = 1$. Therefore in accordance with the notation of Section 2.5

$$\epsilon = \beta,$$

and the polynomial $R_\beta^{(1)}$ defined as

$$R_\beta^{(1)}(\omega) \equiv T_\epsilon^{(1)}(\omega),$$

is the polynomial P_ϵ. It means that we have $y = \omega$ and we can directly carry out a regular perturbation expansion.

Let us write the asymptotic expansion $y = y_0 + y_1\epsilon + y_2\epsilon^2 + \ldots$ and replace it in $P_\epsilon(y)$. Since $P_\epsilon(y) = 0$, by the fundamental theorem of perturbation theory (Theorem 48) we obtain a system of equations in y_0, $y_1, \ldots$, which can be solved. The first equation of the system is given by

$$-K_2^2 K_3^2 y_0^2 (N_x - N_y + y_0) = 0. \tag{3.41}$$

We are only interested in non-negative values of y_0, since we want y to be biologically meaningful. We also require $\omega(0) = y(0) = y_0 \neq 0$ (not identically zero) from Theorem 49. Thus, solving equation (3.41), we obtain $y_0 = N_y - N_x$ if $N_y > N_x$ and $y_0 = 0$ otherwise. Assuming $y_0 = 0$ (*i.e.*, $N_x \geq N_y$), we solve the next order equation

$$K_4 N_y^2 + K_2 K_3 N_y (1 + K_4(N_w - N_x + N_y)) y_1 + K_2^2 K_3^2 (N_y - N_x) y_1^2 = 0. \tag{3.42}$$

We select the positive root of this polynomial and obtain an expression for y_1 when $y_0 = 0$. Thus, we have

$$y_1 = N_y \frac{1 + K_4(N_w - N_x + N_y) + \sqrt{\Delta_{y_1}}}{2K_2 K_3 (N_x - N_y)}, \tag{3.43}$$

where we wrote $\Delta_{y_1} = 1 + 2K_4(N_w + N_x - N_y) + K_4^2(N_w - N_x + N_y)^2$. Equation (3.43) shows that $y_1 > 0$ when $N_x \geq N_y$. Hence, $y \approx \epsilon y_1$ converges to zero from above and, therefore, represents a biologically meaningful solution. We can

conclude that

$$\lim_{\epsilon\to 0}(N_y - y) = \left\{ \begin{array}{ll} N_x, & \text{if } N_y > N_x, \\ N_y, & \text{otherwise.} \end{array} \right\} = \min(N_x, N_y). \tag{3.44}$$

Branch (1,2). On this branch, and again following the notation of Section 2.5, we define

$$T_\epsilon^{(2)}(\omega) \equiv \epsilon^{-2} P_\epsilon(\epsilon^1 \omega).$$

The least common denominator of $\{2,1+1,0+1,0+1\}$ is $q_2 = 1$, so $R_\beta^{(2)}$ is the same polynomial as $T_\epsilon^{(2)}$. Since $y = \omega\epsilon$, we have $N_y - y \underset{\epsilon\to 0}{\sim} N_y$. Furthermore, when replacing ω by an asymptotic expansion $\omega_0 + \omega_1\epsilon + \ldots$ in $T_\epsilon^{(2)}$ and applying the fundamental theorem of perturbation theory (Theorem 48), we obtain the same equation for w_0 as for y_1 in the previous branch (see equation (3.42)):

$$K_4 N_y^2 + K_2 K_3 N_y (1 + K_4(N_w - N_x + N_y))\omega_0 + K_2^2 K_3^2 (N_y - N_x)\omega_0^2 = 0\ . \tag{3.45}$$

We have $y_1 = \omega_0$. In other words, at large, but finite $L = 1/\epsilon$, the convergence behaviour of the two branches is identical. This agrees with Theorem 37 which states that there is only one positive solution for each set of reaction constants and initial conditions. In conclusion, we find that $N_y - y = \min(N_x, N_y)$, which gives the following expression for the amplitude

$$A \equiv \frac{K_1 z}{1 + K_1 z} \min(N_x, N_y), \tag{3.46}$$

with z defined in (3.34). As the steady state concentration of JAK3, z, is the same in the IL-7R model with or without the extra chain R, the same expression has been obtained for the amplitude of both models 1 and 2.

Computation of the EC_{50}

Since we did not compute analytic expressions for each steady state concentration, the EC_{50} expression has to be obtained by computing a Gröbner basis of the polynomial system (3.32) augmented by the polynomial

$$K_3 K_2 K_1 L x y z (1 + K_1 z) - \frac{K_1 z M}{2} = 0, \tag{3.47}$$

considering x, y, z and L as variables, with $M = \min(N_x, N_y)$. The lex Gröbner basis obtained for this system is:

$$0 = K_1 z^2 + (1 + K_1(N_x - N_z))z - N_z, \tag{3.48a}$$

$$0 = K_3 a L^3 + A_L L^2 + B_L L + C_L, \tag{3.48b}$$

$$0 = y + \frac{-aL^2 + B_y L + C_y}{2K_2(K_2 - K_4)M^2}, \tag{3.48c}$$

$$0 = w + \frac{aL^2 + B_w L + C_w}{2K_4(K_2 - K_4)M^2}, \tag{3.48d}$$

$$0 = x + \frac{aL^2 + B_x L + C_x}{2K_2K_4M^2(1 + K_1 z)}, \tag{3.48e}$$

where we wrote:

$$\begin{aligned}
a &= K_2^2 K_3^2 (M - 2N_x)(M - 2N_y)^2, \\
A_L &= K_2 K_3^2 M(M - 2N_y)(-2 + 3K_2 M - K_4(M + 2N_w - 2N_x) \\
&\quad - 2K_2(2N_x + N_y)), \\
B_L &= K_3 M^2 [2K_4 \\
&\quad + K_2(-2 + 3K_2 M + 2K_4(-M - N_w + N_x + N_y) - 2K_2(N_x + 2N_y))], \\
C_L &= K_2(K_2 - K_4)M^3, \\
B_y &= -\frac{A_L}{K_3}, \\
C_y &= M^2(-2K_4 + K_2(2 + K_4(M + 2N_w - 2N_x) - 2K_2(M - N_x - N_y))), \\
B_w &= -2K_2 K_3 M(M - 2N_y)(1 + K_4 N_w + K_2(N_x + N_y - M)), \\
C_w &= K_2 M^2(K_2(M - 2N_y) - 2K_4 N_w), \\
B_x &= -K_2 K_3 M(M - 2N_y)(2 + K_4(M + 2N_w - 2N_x) - 2K_2(M - N_x - N_y)), \\
C_x &= M^2(2K_4 + K_2(K_2 - K_4)(M - 2N_y)).
\end{aligned}$$

The polynomial (3.48a) is expected to be independent of the ligand concentration, L. The EC_{50} expression is the real positive root of polynomial (3.48b) at which x, y and w (obtained via polynomials (3.48e), (3.48c) and (3.48d), respectively) are positive. The polynomial (3.48b) reflects the parameter dependence of the EC_{50}: since the parameters K_1 and N_z are not present in its coefficients, one can

affirm that the EC_{50} is, once again, independent of the kinase, JAK3. Thus, the problem of computing the EC_{50} is reduced to solving a univariate polynomial (equation (3.48b)). In comparison, before any algebraic manipulation was possible, the polynomial system (3.32) had to be solved multiple times to obtain the dose-response curve (a sigmoid), which was then fitted to a Hill equation. Finally, the EC_{50} was computed from the fitted parameters of the Hill curve. Alternatively, if one wanted to apply the Gröbner basis-free method of Section 3.2, one would have to solve the polynomial (3.33b) in y (which is possible in theory), find its positive real solution (which is potentially hard), substitute the expression of y into $\sigma(L)$, and then solve for the EC_{50}.

3.4.3 Model validation

A short numerical exploration shows that, this time, the increase of γ_c abundance does increase the EC_{50}, under the condition that the abundance of the additional chain R is proportional to a certain power of the abundance of γ_c, $N_w = N_x^p$ with $p > 1$ (see Figure 3.16). Indeed, a fixed amount of R or an amount proportional ($p = 1$) to N_x is not sufficient to observe changes in the EC_{50} compared to model 1. Figure 3.17 shows the amplitude and EC_{50} variations with the increase of γ_c chain abundance for $p = 1.2$ and $K_4 = 10^{-2}$. The mathematical analysis conducted in the previous section shows that the EC_{50} is independent of JAK3 abundance, in accordance with the experimental observations. This analysis also showed that the amplitudes of model 1 and 2 are the same, which guarantees that model 2 will be as accurate as model 1 to fit the amplitude. Model 2, thus, seems to be a good candidate to reproduce both the experimental amplitude and EC_{50} behaviours. I now make use of ABC to fit model 2 to the data in order to infer the values of the affinity constants and the power p which links N_x and N_w.

Approximate Bayesian computation (ABC)

In Section 3.4.2, I reduced the problem of computing the EC_{50} to solving a univariate polynomial of degree 3 (equation (3.48b)) and selecting the real positive root of this polynomial at which x, y and w (obtained via polynomials (3.48e), (3.48c) and (3.48d), respectively) are positive. Unfortunately, due to the huge

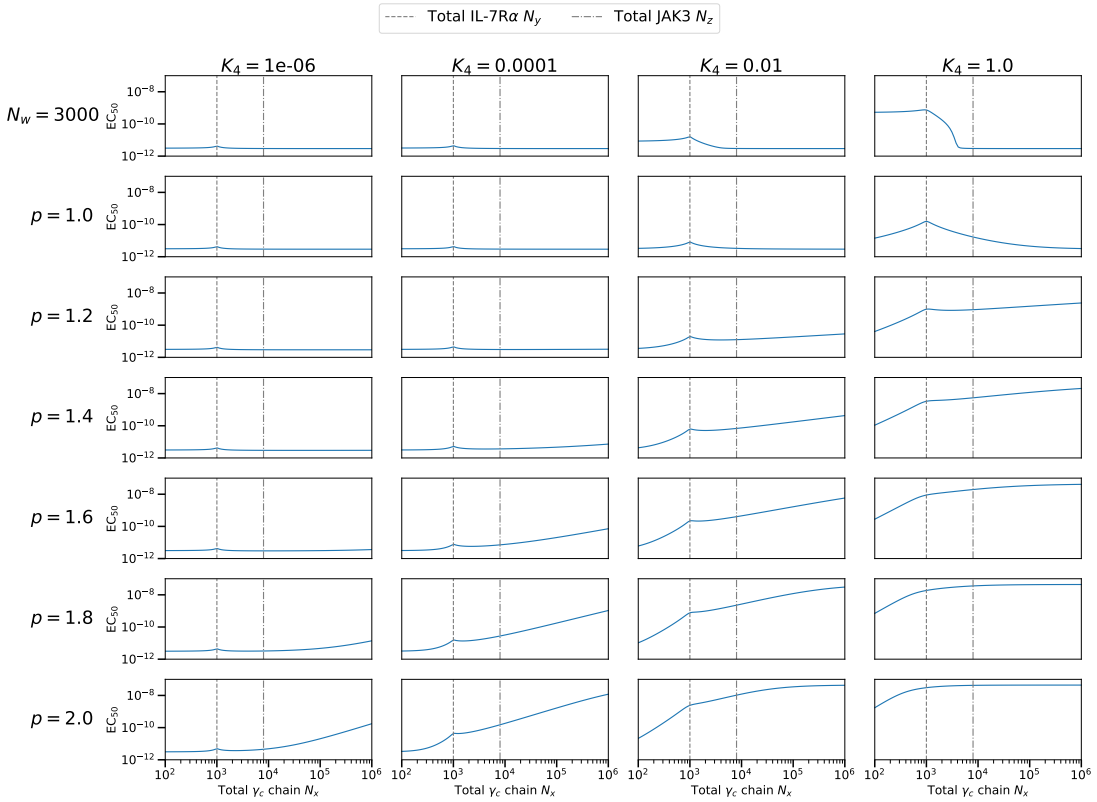
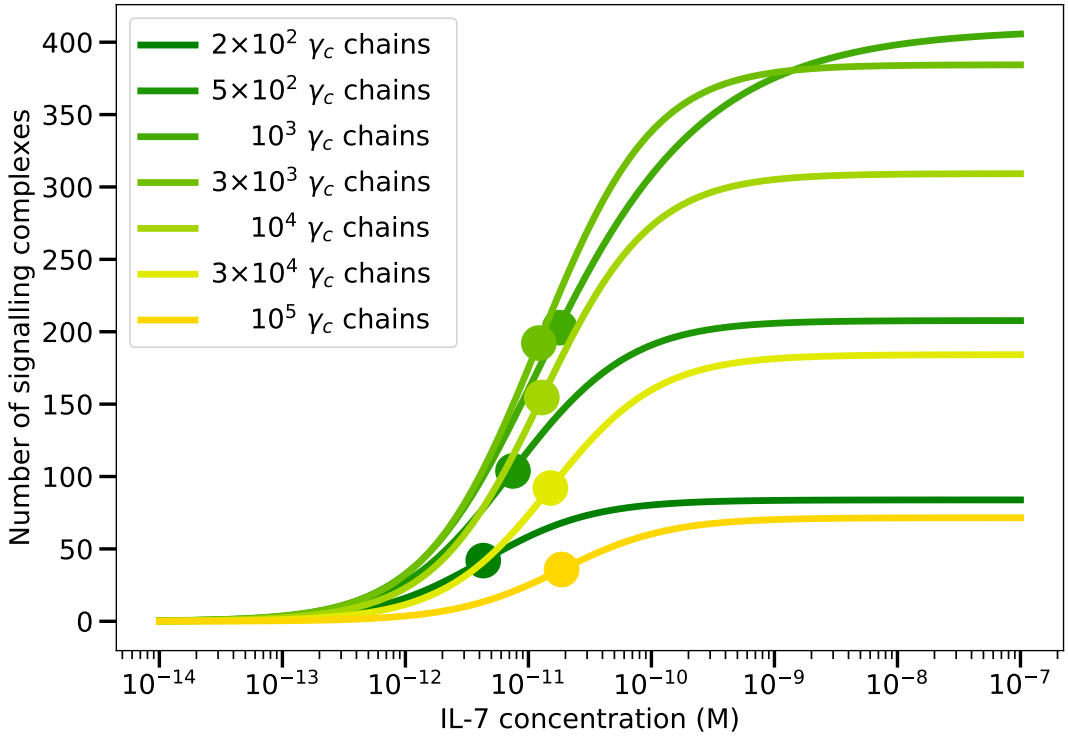
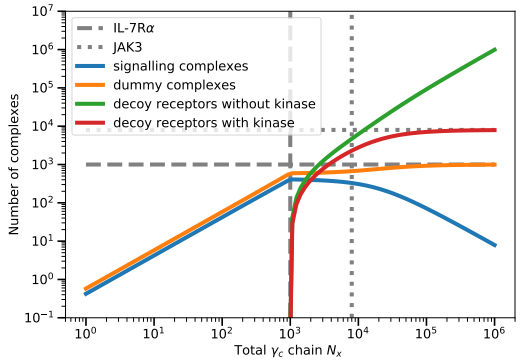


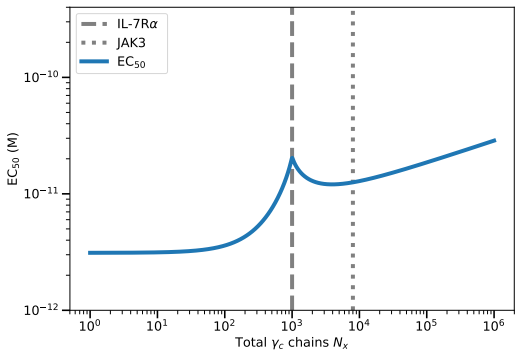
Figure 3.16: Numerical exploration of the EC_{50} of model 2 for different values of K_4 and $N_w = N_x^p$ as a function of the total γ_c abundance N_x. The other parameters were fixed: $K_1 = 10^{-4.04}$, $K_2 = 17 \times 10^{-3}$, $K_3 = 34 \times 10^{10}\mathrm{M}^{-1}$, $N_y = 10^3$ and $N_z = 8 \times 10^3$. Increasing the abundance of γ_c chains increases the EC_{50} only when $p > 1$.



(a) Dose-response curve for different γ_c chains abundance



(b) Signalling and "dummy" complexes at high ligand concentration



(c) Modelled EC_{50}

Figure 3.17: Numerical exploration of the impact of γ_c chain abundance on the IL-7 response in model 2: (a) Dose-response curve for different γ_c abundances. The amplitude (the plateau of the dose-response curve) increases then decreases with the increase of γ_c abundance. The EC_{50} (dots) increases as γ_c abundance increases (dots shift to the right). (b) Number of signalling and "dummy" complexes at high ligand concentration ($L = 10^{-3}$ M). (b) The modelled EC_{50} increases as γ_c abundance increases. For this figure, the parameters has been fixed to 10^3 IL-7Rα chains per cell, 8×10^3 JAK3 molecules per cell, $K_1 = 10^{-4.04}$, $K_2 = 17 \times 10^{-3}$ and $K_3 = 34 \times 10^{10}\text{M}^{-1}$, $K_4 = 10^{-2}$ and $p = 1.2$. Values of K_2 and K_3 were taken from Ref. Cotari *et al.* (2013b).

difference in order of magnitude of the parameters involved in the coefficients of polynomial (3.48b) (for instance, $K_3 = 10^{10}\mathrm{M}^{-1}$ and $K_1 = 10^{-4}$), the problem of finding the roots of this polynomial is ill-conditioned, *i.e.*, the roots of the polynomial are very sensitive to small perturbations (such as numerical errors) to the coefficients (this is a common problem in polynomial root finding (Boyd, 2002; Tsai, 2014)). This makes it difficult to find the positive real roots of polynomial (3.48b) that also let the other variables positive and in biologically relevant ranges (for instance we want $N_x \geq x > 0$ as N_x is the total γ_c abundance available)[1]. To fix this issue, one could try to normalise the coefficients of polynomial (3.48b) or decompose the polynomial into a Chebychev basis according to the method described in Boyd (2002). This is, however, out of the scope of this thesis. To simulate model 2 and perform the approximate Bayesian computation, I thus make use of more traditional methods: computing a dose-response curve by solving system (3.32) for different ligand concentrations and fitting the curve to a sigmoid equation to derive the amplitude and EC_{50}. Despite the fact that I do not make use of polynomial (3.48b), the mathematical analysis conducted in the previous section is still very useful. Indeed, I showed that the amplitude and the EC_{50} of model 2 depend on different separated subsets of parameters: the amplitude only depends on K_1 while the EC_{50} depends on all the other parameters. The pharmacological quantities can thus be fitted separately. Furthermore, the amplitude of model 1 and 2 are the same. Hence, the amplitude does not need to be fitted again and we can fix K_1 to the value found when fitting model 1 ($K_1 = 10^{-4.04}$). Finally, the separation of the parameters also allows us to only compare the modelled and experimental EC_{50}, independently of the parameters ξ_1 and ξ_2 defined in (3.2).

Once again, I have decided to not make any assumption on the parameters and chose uniform prior distributions on the logarithm of the parameters (except for p): $\log(K_2) \sim \mathcal{U}(-7, 6)$, $K_3 \sim \mathcal{U}(9, 14)$, $\log(K_4) \sim \mathcal{U}(-8, 4)$ and $p \sim \mathcal{U}(0.5, 2)$. The ABC algorithm goes as follows. For each simulation, I sample a parameter set according to the prior distributions defined above. Then, for each bin i of the data set, I extract the γ_c chain, IL-7Rα and JAK3 abundances (which define the values for N_x, N_y and N_z, respectively). For these values of N_x, N_y and

[1] I tried to compute the roots of polynomial (3.48b) making use of the function *roots* from the Python library Numpy (Oliphant, 2006)

N_z and for different ligand concentrations L (typically L ranges from 10^{-15}M to 10^{-6} M), I numerically solve the polynomial system (3.32) to obtain the steady state number of unbound single chains x, y, z and w. The number of signalling complexes formed, $\sigma(L) = K_3K_2K_1Lxyz$, is computed by multiplying the steady state values of the single chain components. The values of σ at the different ligand concentrations define the dose-response curve, which is fitted to a sigmoid equation to obtain the EC_{50}. I am now able to compute the distance d^2 between the modelled and experimental EC_{50}:

$$d^2 = \sum_{i=1}^{N_{bin}} (model[i] - data[i])^2, \tag{3.49}$$

where $N_{bin} = 172$ is the number of bins, $model[i]$ corresponds to the modelled EC_{50} with the abundances of bin i and $data[i]$ is the EC_{50} computed from the data set for bin i. I repeat these steps five million times and obtain a list of five million distance values corresponding to the five million different parameter sets tested. I select the 10^3 parameter sets that minimise the distance d^2 to obtain a posterior distribution. The normalised posterior distributions for parameters K_2, K_3, K_4 and p, resulting from the ABC fitting the EC_{50} are shown in Figure 3.18. The posterior distributions of all the parameters are narrower than the prior distributions. In particular, the posterior distributions of K_2, K_4 and p present a clear peak which means that the best parameter value most likely falls around this mode. The posterior distribution of K_3, however, does not have an obvious mode and so the value of K_3 was not inferred accurately. The parameter sensitivity analysis, conducted by Michael Adamer and published in Ref. Sta *et al.* (2022b), shows that K_3 is one of the least important parameters in model 2: this might be the reason why we struggle to infer it. The correlation plots (Figure 3.19) show that K_2 and K_3, as well as p and K_4 seem correlated while the other pairs of parameters are mostly independent. Figure 3.20 shows the modelled EC_{50} computed with the best parameters values determined by this ABC ($K_2 = 10^{-5.16}$, $K_3 = 10^{13.2}\text{M}^{-1}$, $K_4 = 10^{-4.098}$ and $p = 1.41$) along with the EC_{50} computed with the experimental data set. It shows that model 2 is much better than model 1 at reproducing the experimental EC_{50} behaviour. Figure 3.21 summarises the similar

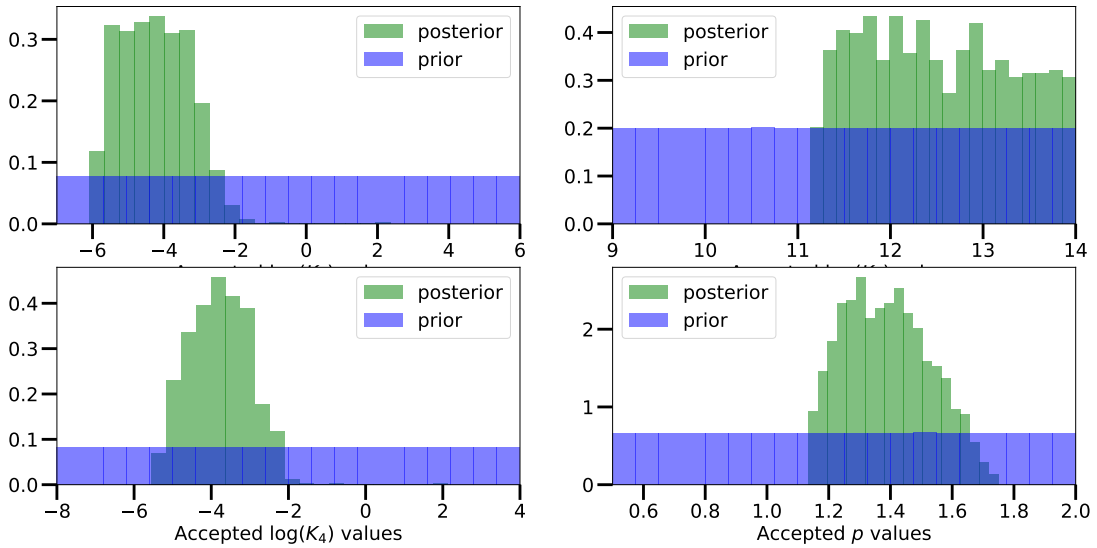


Figure 3.18: Normalised prior (blue) distribution of the five million parameter samples and posterior distribution (green) of the 10^3 best parameters for model 2.

behaviour of the modelled and experimental amplitude and EC_{50}.

Conclusions

The addition of the extra chain R to model 1 did not modify the amplitude expression but changed the modelled EC_{50} behaviour. As shown in Figure 3.20, model 2 can now reproduce the experimental EC_{50} increase with γ_c chain abundance, while predicting the observed amplitude behaviour as accurately as model 1. The parameter values inferred with Bayesian computation in this section vary from the ones found by Guillaume Voisinne in Ref. Sta *et al.* (2022b), mainly because we did not choose the same fitting strategy. The difference in parameter values between those of this section and his, is never more than one order of magnitude.

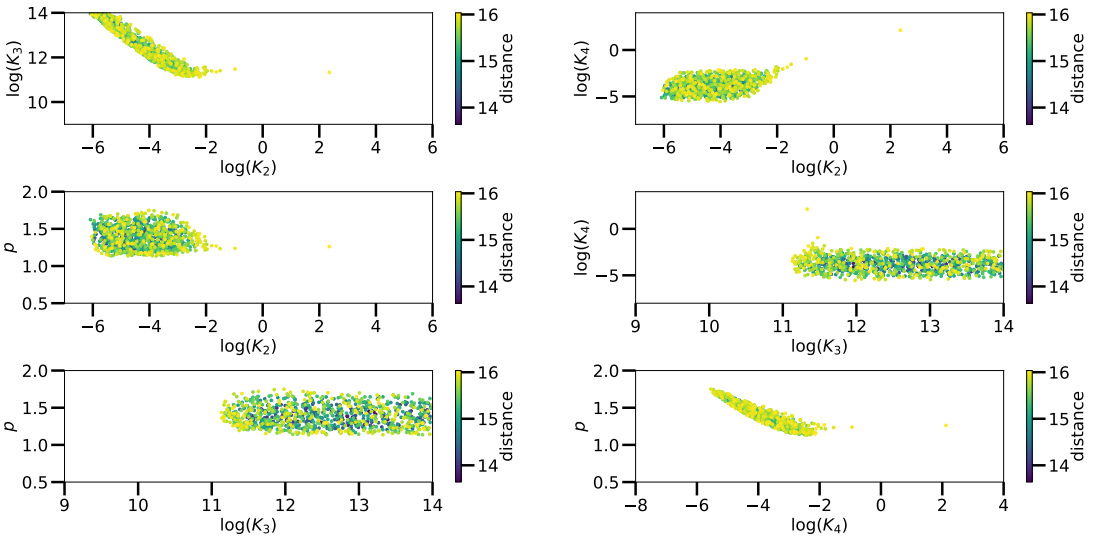


Figure 3.19: Correlation plots of the 10^3 best parameters according to the distance value, resulting from the fitting of model 2 to the experimental EC_{50}.

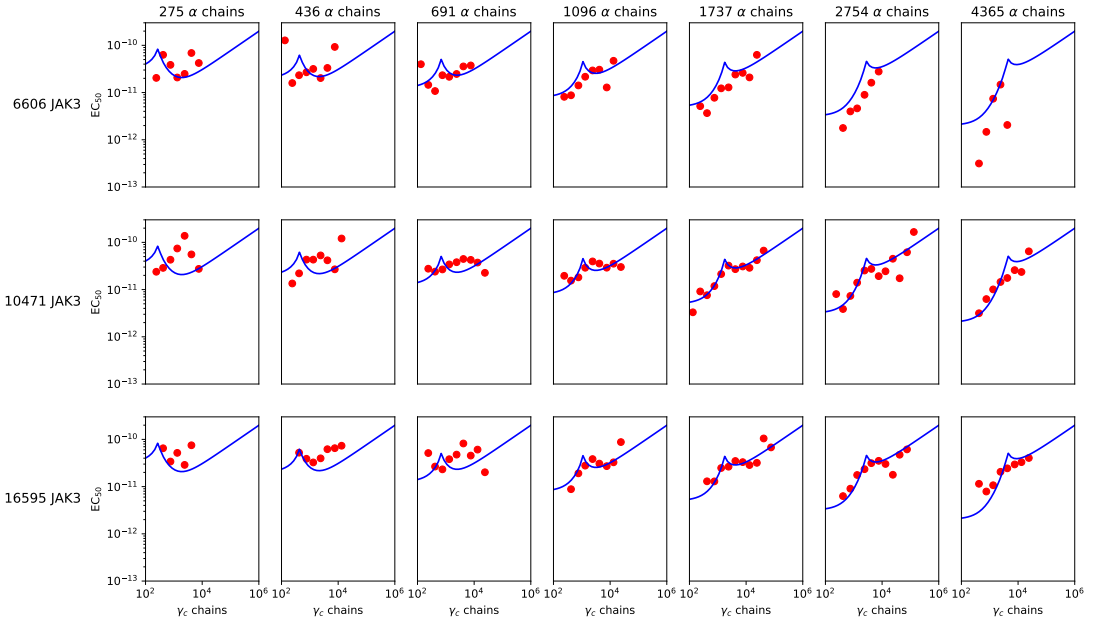
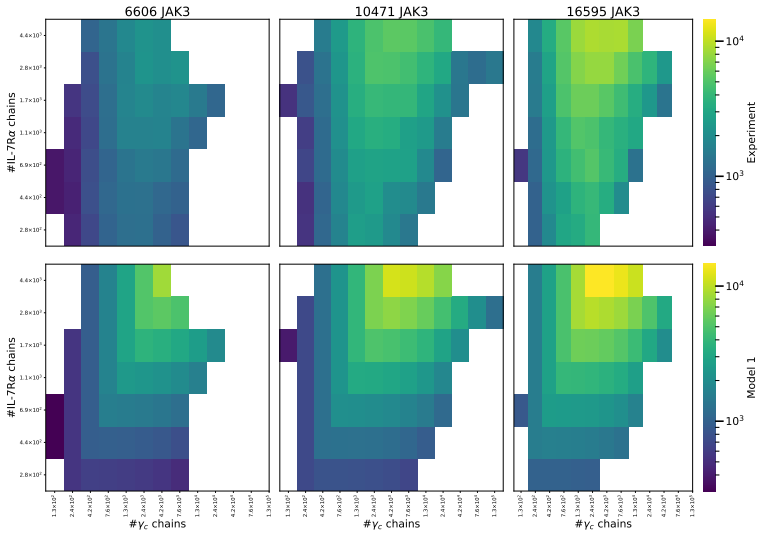
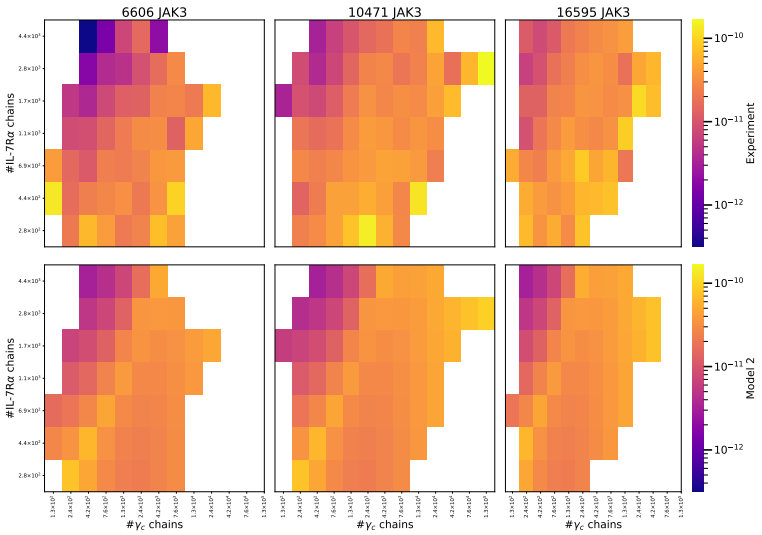


Figure 3.20: Comparison of the experimental (red dots) and modelled EC_{50} (model 2 in blue) for different receptor chain abundances. The modelled quantities were generated with the best parameters (those which return the smallest distance): $K_2 = 10^{-5.16}$, $K_3 = 10^{13.2}\mathrm{M}^{-1}$, $K_4 = 10^{-4.098}$ and $p = 1.41$.



(a) Experimental and modelled amplitude (model 2)



(b) Experimental and modelled EC_{50} (model 2)

Figure 3.21: Complete map of the experimental and modelled amplitude and EC_{50} (model 2) for different levels of receptor chain abundances. The modelled quantities were generated with the best parameters (those which return the smallest distance):$K_2 = 10^{-5.16}$, $K_3 = 10^{13.2}\text{M}^{-1}$, $K_4 = 10^{-4.098}$ and $p = 1.41$.

3.5 Discussion

In this chapter, I proposed three IL-7R models to elucidate the seemingly paradoxical observation that increased abundances of γ_c chain, a crucial component of the IL-7R, decreased the IL-7-induced T cell response, reflected both in a smaller amplitude and a greater EC_{50}. The models explain the amplitude dynamics by the formation of signalling-deficient ("dummy") complexes that did not co-opt intra-cellular kinases. At high γ_c chain abundances, all IL-7Rα chains are γ_c- and ligand-bound, forming "dummy" complexes instead of signalling receptors. The introduction, in model 2, of the γ_c- binding protein, R, whose abundance correlates with that of γ_c, allows to reproduce the experimental EC_{50} behaviour. This protein, R, simulates the competition for the γ_c chain between the receptors of the γ_c family, hence specific competition for the formation of the complete IL-7 receptor accounts for the EC_{50} increase in the IL-7 response.

To further explore the imprint of γ_c abundance on cytokine responses, Guillaume Voisinne, Jesse Cotari and Grégoire Altan-Bonnet repeated their experimental measurements and analysis for the STAT5 phosphorylation responses triggered by IL-2 and IL-15 stimulation. In both cases, the effect of γ_c on the amplitude of response to the respective cytokine was nearly identical to that seen following IL-7 treatment (see Figure 3.22). Most interestingly, γ_c chain abundance did not significantly affect EC_{50} values for either IL-2 or IL-15.
Making use of these additional experimental results and a previous IL-2R model (Cotari *et al.*, 2013b), Guillaume Voisinne and Jesse Cotari concluded that the extra chain, R, could not be IL-2Rβ, and suggested that R would act only on the IL-7 receptor (Sta *et al.*, 2022b). Precise identification of this protein is still pending.

The work of this chapter demonstrates that optimal signalling necessitates a precise balance of protein abundances (for instance, model 1 and 2 show that the maximum amplitude is attained when the number of γ_c chains equals the number of IL-7Rα chains). The requirement for abundance balance between kinases and receptor subunit chains might provide built-in protection against aberrant activation, as observed during oncogenic transformation (Shtiegman *et al.*, 2007).

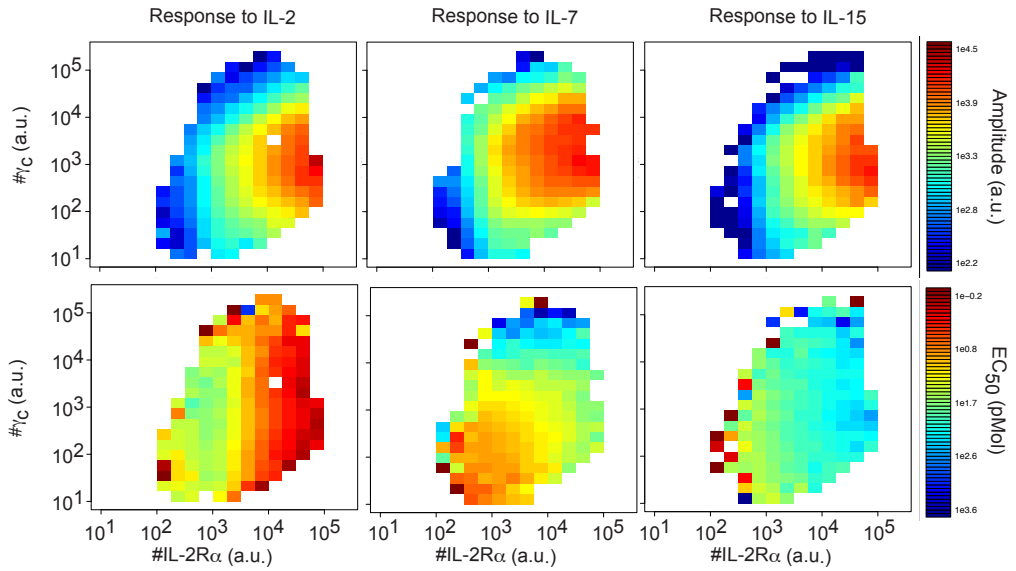


Figure 3.22: Experimental maps of amplitude (top) and EC_{50} (bottom) for the pSTAT5 responses to IL-2, IL-7 and IL-15, for different expression levels of IL-2Rα and γ_c. Note how similar the amplitude values are for different cytokines, while the EC_{50} have very different dependencies with the abundances of IL-2Rα and γ_c. Figure taken from Ref. Sta *et al.* (2022b).

3. TUNING OF IL-7 SIGNALLING THROUGH IMBALANCED ABUNDANCES OF RECEPTORS AND KINASES

Robust input/output response does limit functional variability but also protects against enhancement of receptor signalling and its dysregulated accumulation of downstream survival signals. This work also showed that while the behaviour of the amplitude seems to be similar across several cytokine-receptor systems, the sensitivity of the cellular response, as measured by EC_{50}, is a non-intuitive, cytokine-specific quantity which should motivate further experimental and theoretical analysis.

Chapter 4

Algebraic analysis of receptor-ligand systems and generalisations

As mentioned in Chapter 1, the study of a receptor-ligand system generally relies on the analysis of its dose-response curve and in particular of the key pharmacological quantities such as the amplitude and the EC_{50}. Previous work has shown that the estimation of the amplitude and the EC_{50} from experimental data is often possible, although strong bias and errors might be introduced (Badrick *et al.*, 2023). Usually one starts with a data set where the number (or concentration) of receptor-ligand signalling complexes formed is measured for different values of the ligand concentration. Then, the estimation of the amplitude and the EC_{50} is turned into a regression problem by assuming a functional relationship in the data set and fitting a parametric curve. A simple first approach is to plot experimental values (corresponding to a measurable variable which quantifies cellular response, such as pSTAT5 as used in Chapter 3) as a function of ligand concentration. The amplitude and the EC_{50} are then read directly from a curve formed by interpolation of the data points. Since the EC_{50} is likely to fall between two data points, a geometrical method (Alexander *et al.*, 1999) can be used for an accurate determination. Nowadays, many software packages can compute the amplitude and the EC_{50} from the data set making use of statistical methods, which consist in finding the "best-fit" equation to the dose-response curve. The most common shape of the dose-response curve is a sigmoid, and thus, can be fitted with the Hill equation (Gesztelyi *et al.*, 2012; Goutelle *et al.*, 2008). However, other functions

are also possible, such as a logistic equation (Chen *et al.*, 2013; Li *et al.*, 2015), a log-logistic equation (Jiang & Kopp-Schneider, 2014; Suriyatem *et al.*, 2017), or the Emax model (Macdougall, 2006; Thomas, 2006). An asymmetrical sigmoid equation is sometimes needed for better precision (Chen *et al.*, 2013; Suriyatem *et al.*, 2017). The amplitude and the EC_{50} are parameters of these equations and can thus, be directly inferred from the fitting process. When a data set does not follow the strictly increasing pattern of these Hill-like functions, then more complex functions, such as bell-shaped curves (Rovati & Nicosia, 1994), or multi-phasic curves (Di Veroli *et al.*, 2015) can be used. It is important to note that even though these empirical regression methods allow one to quantify two key receptor-ligand metrics, amplitude and EC_{50}, they do not offer any mechanistic insights for the receptor-ligand system under consideration. To this end, mathematical models can be used to describe the receptor-ligand system at a molecular level; that is, mathematical models consider the biochemical reactions which initiate a cellular response (Eftimie *et al.*, 2016; Lauffenburger & Linderman, 1996; Tyson *et al.*, 2003; White *et al.*, 2022; Wiley *et al.*, 2003). The challenge in such models is finding analytic, ideally closed-form, expressions for the amplitude and the EC_{50}. Due to the non-linear nature of the biochemical reactions involved, this poses a significant and practical challenge. As seen in Chapter 3, the advantages of having analytic (or closed-form) expressions of the amplitude and the EC_{50} for a large class of receptor-ligand systems are many: (i) they allow to quantify their dependence on receptor copy numbers, (ii) they facilitate mathematical model validation and parameter exploration, and (iii) they reduce computational cost. To the best of my knowledge such expressions have been obtained in a few instances: closed or open bi-molecular receptor-ligand systems (Gabrielsson *et al.*, 2018), monomeric receptors (Mack *et al.*, 2008), ligand-induced dimerisation (White *et al.*, 2022) or ternary complexes (Douglass Jr *et al.*, 2013). More complicated receptor-ligand models have been studied with chemical reaction network theory (CRNT) (Feinberg, 1987; Otero-Muras *et al.*, 2017; Shiu, 2010), but CRNT has thus far, been focused on the analysis of the steady state of the system (*i.e.*, existence and number of steady states and their stability). Yet, as demonstrated in Chapter 3, CRNT is an essential and useful framework to start any mathematical investigation of the amplitude and the EC_{50}.

Another aspect which can be effectively addressed by mechanistic mathematical modelling is the effect of internal or external perturbations to the state of a cell. For example, in single-cell experiments or even repetitions of bulk experiments (Cotari *et al.*, 2013b; Feinerman *et al.*, 2010), the experimental conditions can never be replicated exactly. This leads to noise not only in the measured quantities, but also in the reaction mechanisms themselves. This variation can be captured in mathematical models which encode parameters such as affinity constants or total copy number of constituent molecular species. An analytical study of the dependency of pharmacologically relevant quantities, such as amplitude and EC_{50}, on the reaction parameters can facilitate *in silico* drug design (Moraga *et al.*, 2017). While amplitude and EC_{50} are widely employed to characterise biological phenomena, the manner in which they depend on the parameters of the receptor-ligand model is not fully understood. Thus, improved understanding of these relationships could provide novel biological and computational insights.

In Chapter 3, I computed the analytic expression of the amplitude and EC_{50} of two IL-7R models making use of a tool from computational algebraic geometry called Gröbner basis. The expressions obtained showed that the amplitude and EC_{50} of the models were not depending of the same parameters and could be fitted independently. This result, coupled with the use of the analytic expressions when possible, dramatically reduced the computational cost and time of model simulation and fitting. I recapitulate the key steps of the method used to compute the analytic expressions of the amplitude and EC_{50} in Section 4.1. The IL-7R models were simple enough to illustrate the algorithm, and thus, to derive analytic expressions for the key pharmacological quantities, yet complex enough to show its limitations, as additional mathematical operations, like perturbation theory, were necessary. The previous chapter focused on cytokine receptors which display a particular receptor architecture. Many other receptor configurations exist, such as RTK receptors (Du & Lovly, 2018). In Section 4.2, I apply the method and compute the amplitude and EC_{50} expressions for other receptor-ligand systems. In particular, I consider monomeric receptors, homodimeric and heterodimeric receptors, and trimeric receptors, requiring, or not, a downstream kinase to signal. All these systems are studied at steady-state and the ligand is assumed to be

in excess. These additional examples yield two main results: (i) the amplitude and EC_{50} of some models are very similar, (ii) if the model is too complex, the method is unable to compute analytic expressions of these key quantities. As a consequence of the first remark, I generalise this result by studying a family of receptor-ligand models in Section 4.3 and 4.4. In response to the second remark, I investigate, in Section 4.5, whether it is possible to learn about the amplitude and EC_{50} of complex models (computing the analytic expression or finding upper and lower bounds) by computing these quantities for simpler sub-models.
The models studied in this chapter (and in Chapter 3) are summarised in the tables in Section 2.10. These tables also recapitulate the analytic expressions of amplitude and EC_{50} obtained for each of these models.

4.1 A method to compute analytic expressions of the steady state, amplitude and EC_{50} of receptor-ligand systems

From the two IL-7R examples studied in Section 3.2 and 3.4, we devise a general algorithm to compute analytic expressions of the steady state, amplitude and EC_{50} for some receptor-ligand binding models when the ligand L is considered in excess. The key steps are described in Table 4.1.

One of the crucial parts of the proposed algebraic algorithm is the amplitude computation. Usually, we have the simplification that $\min(\sigma) = \sigma(0) = 0$, however, finding $\max(\sigma)$ can be challenging. For certain classes of models we have

$$\lim_{L \to +\infty} \sigma(L) = \max(\sigma),$$

which greatly reduces the calculation. We can now either solve the Gröbner basis from step 3 directly, to obtain analytic expressions of the steady state concentrations of the single chains components, or use perturbation theory as outlined in Section 3.4. In the final step, if an exact expression for the EC_{50} cannot be computed, *i.e.,* the univariate polynomial in L has a large degree, one

already reduces the cost of the EC_{50} computation compared to the naive approach. This polynomial also indicates the dependencies of the EC_{50} on the parameters.

Key steps

1) Write the mass action kinetics set of ODEs for the system under consideration.
2) Obtain the polynomial system by combining the steady state and conservation equations.
3) Compute a lex Gröbner basis of the polynomial system obtained in step 2.
4) Define the signalling function $\sigma(L)$.
5) Compute the amplitude by finding the extreme values of σ:

$$\text{Amplitude} = \max(\sigma) - \min(\sigma).$$

6) Compute a lex Gröbner basis of the polynomial system obtained in step 2 augmented by the equation

$$\sigma(L) - \left[\frac{\text{Amplitude}}{2} + \min(\sigma)\right] = 0,$$

with the ligand concentration, L, considered as an additional variable. This additional equation corresponds to definition 43 (in Chapter 2) of the EC_{50}.
7) Find the positive roots of the univariate polynomial in L of the Gröbner basis obtained in step 6. The root which allows the other variables of the polynomial system to be positive is the EC_{50}.

Table 4.1: Key steps of the algebraic method to compute analytic expressions for the amplitude and EC_{50} of receptor-ligand systems

4.2 Exploring different receptor architectures

The unexpected modulation of cytokine responsiveness based on receptor abundances, unveiled in the previous chapter, led us to consider the overall biological impact of such an arrangement. The method described in the previous section can be applied to more general biochemical designs of receptors, connecting

extra-cellular signals to intra-cellular responses. In this section, I consider several possible arrangements of receptor with or without intrinsic kinase activity with one (monomeric), two (dimeric) or three (trimeric) trans-membrane chains. For every model, we consider a primary chain called γ, equivalent to the common gamma chain γ_c in the previous chapter. For monomeric and homodimeric receptors, this is the only chain. Heterodimeric receptors are also composed of a secondary chain called α. Finally, I consider a trimeric receptor with a primary chain γ, a secondary chain β and third chain α, by analogy with the IL-2 receptor introduced in Chapter 1 (the steady state of a trimeric receptor model with three identical chains has been studied in Ref. Ho & Harrington (2010)). All the receptor-ligand systems of this section have a similar structure and follow similar assumptions to the models in Chapter 3: the signalling complex is formed sequentially, the ligand is considered in excess, models with an intra-cellular extrinsic kinase (IEK) form signalling and "dummy" complexes with the same binding rates (no allostery) and the systems are at steady state. We also assume mass-action kinetics. All the models are reversible and deficiency 0, thus admit a unique positive steady state.

Notation Here, we use notation similar to the previous chapter. In the chemical reactions, γ, α and JAK describe the primary chain, secondary chain and IEK, respectively, when present in the model. In mathematical equations, x denotes the unbound (free) primary chain concentration, y denotes the unbound secondary chain concentration and z the unbound IEK concentration. The ligand in chemical reactions and its concentration in mathematical equations are both denoted by L. N_x, N_y and N_z stand for the total number of γ chain, α chain and IEK molecules per cell, respectively. Finally, since only reversible reactions are considered, k_i denotes the forward reaction constant ($\rightharpoonup$) and q_i the backward reaction constant ($\leftharpoondown$). We write $K_i = \dfrac{k_i}{q_i}$ for the affinity constants. To stay consistent across all the models studied, K_1 will describe the affinity constant of the binding of the IEK to the primary chain. The constant K_2 will denote the binding of the primary (bound to the IEK or not) and secondary chain in the heterodimeric receptor case and the binding of two primary chains in the homodimeric case. Finally, K_3 will stand for the affinity constant of the binding of the ligand to the full receptor (with or without IEK). These consistent choices should allow amplitude and EC_{50}

comparisons.
We remind that in this thesis we will not distinguish between number (or concentration) of signalling complexes since one can be obtained from the other if we know the volume of the system and Avogadro's number.

We first consider monomeric (one chain), homodimeric (two identical chains) and heterodimeric (two different chains) RTK receptors. For instance, these three receptor configurations are the different forms of the epidermal growth factor receptor (EGFR or ErbB-1), a member of the ErbB family, a sub-family of RTK receptors. Indeed, the ligand, EGF, can bind to monomers or pre-formed inactive homodimers of the EGFR (Maruyama, 2014; Yarden & Schlessinger, 1987). Alternatively, the EGFR can pair with another ErbB receptor family, such as ErbB-2/Her2/neu to create heterodimers. In the second part of this section, we will consider the same receptor architectures but, this time, assuming they do not possess intrinsic kinases and need intra-cellular extrinsic kinases (IEK) to signal. Note that the heterodimeric receptor model with IEK has been studied in Section 3.2 as one of the IL-7R model.

4.2.1 Monomeric receptor model (RTK)

The first receptor structure considered is a monomeric receptor [1] with intrinsic kinase activity. This monomeric receptor is composed of a unique chain, γ (see Figure 4.1). The sole chemical reaction considered is the binding of the γ chain with the ligand:

$$L + \gamma \quad \rightleftharpoons \quad L:\gamma \qquad\qquad K_3 = \frac{k_3}{q_3}.$$

The signalling complex is $L:\gamma$ and we denote its concentration by c. The analysis of this system and computation of its amplitude and EC_{50} is straightforward and does not require the computation of Gröbner bases. Indeed, the ordinary

[1] Note that the work in this section differs from the study conducted in Mack *et al.* (2008) as we assume that the ligand is in excess.

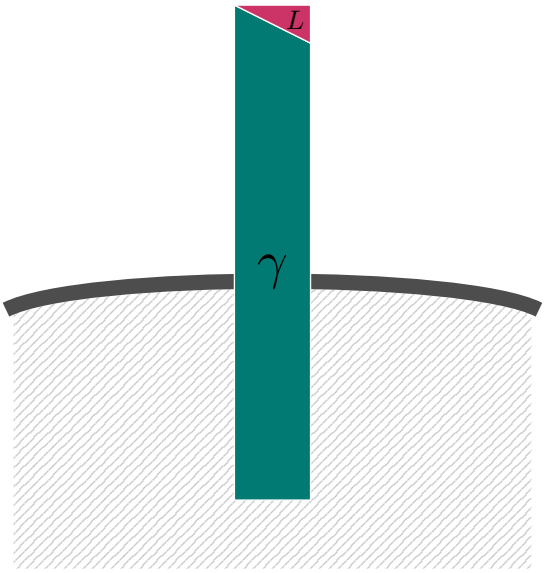


Figure 4.1: The monomeric receptor with intrinsic kinase activity is composed of only one chain, γ. The binding of the ligand, L, to this chain forms the signalling complex. The hatched area determines the intra-cellular environment.

differential equations associated with this system

$$\begin{aligned} \frac{dx}{dt} &= -k_3 L x + q_3 c, \\ \frac{dc}{dt} &= k_3 L x - q_3 c, \end{aligned} \tag{4.1}$$

lead to an obvious equation which translates the conservation of the γ chain:

$$N_x = x + c. \tag{4.2}$$

Since the system is considered at equilibrium, the following detailed-balance steady state equation holds:

$$c = \frac{k_3}{q_3} L x. \tag{4.3}$$

Combining the steady state and conservation equations, we obtain a polynomial equation:

$$0 = -N_x + x + K_3 L x. \tag{4.4}$$

Solving this polynomial yields the steady state expression of the number of unbound γ chains per cell:

$$x = \frac{N_x}{1 + K_3 L}. \tag{4.5}$$

The amplitude can now be computed. The signalling function is defined as

$$\sigma(L) \equiv K_3 L x = \frac{K_3 L N_x}{1 + K_3 L}. \tag{4.6}$$

The graph of this function, the dose-response curve, is a sigmoid. Thus, the amplitude A is attained at high concentration of ligand:

$$A \equiv \lim_{L \to +\infty} \sigma(L) = N_x. \tag{4.7}$$

In this case, the amplitude is the maximum number of signalling complexes that can be formed, as there are N_x gamma chains available. The EC_{50} can be computed by finding L^* such that

$$\sigma(L^*) = \frac{A}{2}, \tag{4.8}$$

which gives:

$$EC_{50} = \frac{1}{K_3}. \tag{4.9}$$

One can notice that the signalling function is a Hill function with Hill coefficient $n = 1$:

$$\sigma(L) = N_x \frac{L^n}{\left(\frac{1}{K_3}\right)^n + L^n}.$$

The amplitude and EC_{50} are then directly read on the expression by identification.

4.2.2 Homodimeric RTK model

Consider a homodimeric receptor with intrinsic kinase, composed of two identical chains γ. The ligand binds to this dimer to form a signalling complex (see Figure

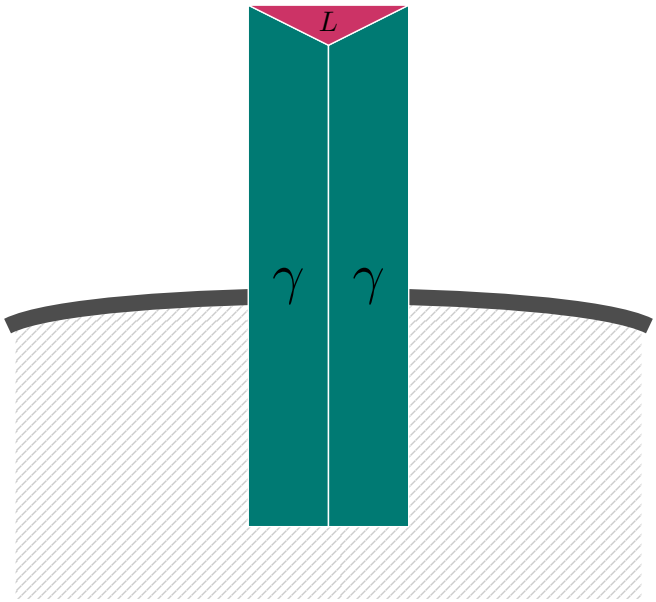


Figure 4.2: The homodimeric receptor RTK is composed of two identical chains γ. The ligand, L, binds to the dimer $\gamma:\gamma$ to form the signalling complex. The hatched area determines the intra-cellular environment.

4.2). The reaction scheme writes as follows:

$$\begin{aligned} \gamma+\gamma \quad &\rightleftharpoons \quad \gamma:\gamma \qquad && K_2=\frac{k_2}{q_2},\\ L+\gamma:\gamma \quad &\rightleftharpoons \quad L:\gamma:\gamma \qquad && K_3=\frac{k_3}{q_3}. \end{aligned}$$

In the next equations, the concentration of complex $\gamma:\gamma$ is denoted c_1 and c_2 denotes the concentration of the signalling complex $L:\gamma:\gamma$. The system of differential equations associated to the model is:

$$\begin{aligned} \frac{dx}{dt} &= -2k_2x^2+2q_2c_1,\\ \frac{dc_1}{dt} &= k_2x^2-q_2c_1-k_3Lc_1+q_3c_2,\\ \frac{dc_2}{dt} &= k_3Lc_1-q_3c_2. \end{aligned} \tag{4.10}$$

Noticing that $\frac{dx}{dt}+2\frac{dc_1}{dt}+2\frac{dc_2}{dt}=0$ yields the following conservation equation:

$$N_x = x + 2c_1 + 2c_2. \tag{4.11}$$

Detailed-balance gives the steady-state equations:

$$\begin{aligned} c_1 &= K_2 x^2, \\ c_2 &= K_3 L c_1. \end{aligned} \tag{4.12}$$

Finally, by substituting the steady state equations in the conservation equation, we obtain the following polynomial describing the model:

$$0 = -N_x + x + 2K_2(K_3 L + 1)x^2. \tag{4.13}$$

The number of unbound gamma chains per cell at the steady state is given by the only positive root of this polynomial (only biologically meaningful root).

$$x = \frac{-1 + \sqrt{1 + 8K_2(1 + K_3 L)N_x}}{4K_2(1 + K_3 L)}. \tag{4.14}$$

The signalling function is defined as

$$\sigma(L) \equiv K_3 K_2 L x^2 = K_3 K_2 L \left[\frac{-1 + \sqrt{1 + 8K_2(1 + K_3 L)N_x}}{4K_2(1 + K_3 L)} \right]^2. \tag{4.15}$$

Since $\sigma(L = 0) = 0$, the amplitude is the maximum of σ. Writing for conciseness $\Delta = 1 + 8K_2(1 + K_3 L)N_x$, the derivative with respect to L of σ is:

$$\frac{d\sigma}{dL} = K_3(-1 + \sqrt{\Delta}) \frac{1 + 8K_2 N_x - \sqrt{\Delta} + K_3 L(-1 + 8K_2 N_x + \sqrt{\Delta})}{16K_2(1 + K_3 L)^3 \sqrt{\Delta}}. \tag{4.16}$$

The numerator of the fraction can be re-written as:

$$1 + 8K_2 N_x - \sqrt{\Delta} + K_3 L(-1 + 8K_2 N_x + \sqrt{\Delta}) = \Delta - \sqrt{\Delta} + K_3 L(\sqrt{\Delta} - 1). \tag{4.17}$$

Since $\Delta > 1$, $\dfrac{d\sigma}{dL}$ is always positive. Thus, σ is an increasing function. Furthermore,

it admits a finite positive limit when $L \to +\infty$. The amplitude is this limit:

$$A \equiv \lim_{L \to +\infty} \sigma(L) = \frac{N_x}{2}. \tag{4.18}$$

The EC_{50} can be obtained by finding L_{50} such that

$$\sigma(L_{50}) = \frac{A}{2} = \frac{N_x}{4}. \tag{4.19}$$

Following the method described in Section 4.1, consider the ensuing polynomial system

$$\begin{aligned} 0 &= -N_x + x + 2K_2(K_3 L + 1)x^2, \\ 0 &= \frac{N_x}{4} - K_3 K_2 L x^2, \end{aligned} \tag{4.20}$$

where L and x are variables, and compute a Gröbner basis of this new system. Making use of Macaulay2 (Grayson & Stillman), we obtain the following triangular system

$$\begin{aligned} 0 &= L^2 - \frac{1 + 2K_2 N_x}{N_x K_2 K_3} + \frac{1}{K_3^2}, \\ 0 &= x - \frac{N_x K_3}{2} L + \frac{K_2 N_x + 1}{2K_2}, \end{aligned} \tag{4.21}$$

which we solve to obtain two values for L and two associated values for x:

$$\begin{aligned} L_1 &= \frac{1 + 2K_2 N_x + \sqrt{1 + 4K_2 N_x}}{2K_2 K_3 N_x} \quad \text{and} \quad x_1 = \frac{-1 + \sqrt{1 + 4K_2 N_x}}{4K_2} > 0, \\ L_2 &= \frac{1 + 2K_2 N_x - \sqrt{1 + 4K_2 N_x}}{2K_2 K_3 N_x} \quad \text{and} \quad x_2 = \frac{-1 - \sqrt{1 + 4K_2 N_x}}{4K_2} < 0. \end{aligned} \tag{4.22}$$

The root x_1 is the only biologically meaningful solution and represents the number of unbound γ chains at the steady state when $L = L_1 = EC_{50}$. In conclusion,

$$EC_{50} = \frac{1 + 2K_2 N_x + \sqrt{1 + 4K_2 N_x}}{2K_2 K_3 N_x}. \tag{4.23}$$

4.2.3 Heterodimeric RTK model

We now consider a heterodimeric receptor model with intrinsic kinase (note that this model has been introduced in Section 2.3). The ligand binds to a dimer

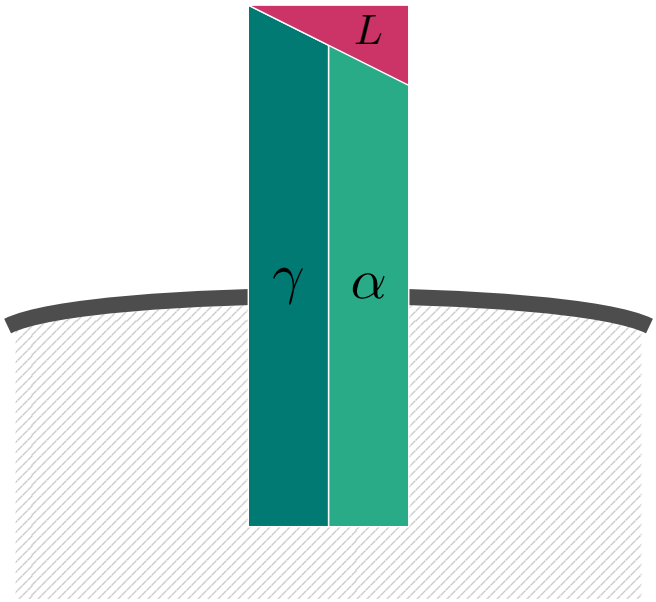


Figure 4.3: The heterodimeric RTK is composed of two trans-membrane chains γ and α. The binding of the ligand, L, to the complex formed by these two chains forms the signalling complex. The hatched area determines the intra-cellular environment.

composed of the primary and secondary chains, thus forming the signalling complex $L : \alpha : \gamma$ illustrated in Figure 4.3. The reaction scheme can be summarised as follows:

$$\begin{aligned} \gamma + \alpha \quad &\rightleftharpoons \quad \gamma : \alpha \qquad && K_2 = \frac{k_2}{q_2}, \\ L + \gamma : \alpha \quad &\rightleftharpoons \quad L : \gamma : \alpha \qquad && K_3 = \frac{k_3}{q_3}. \end{aligned}$$

We proceed as in the previous examples: write the ordinary differential equations associated to the reaction scheme, derive the conservation and steady state equations and finally obtain a polynomial system (this example is detailed in Chapter 2):

$$\begin{aligned} 0 &= -N_y + y + K_2 xy + K_3 K_2 L xy, \\ 0 &= -N_x + x + K_2 xy + K_3 K_2 L xy. \end{aligned} \tag{4.24}$$

Computing a Gröbner basis of (or reorganising by hand) this system yields the following triangular system to solve:

$$0 = -N_x + (1 + K_2(N_y - N_x)(1 + K_3L))x + K_2(K_3L + 1)x^2, \tag{4.25a}$$

$$0 = y - x - N_y + N_x. \tag{4.25b}$$

The expression of x is obtained by solving (4.25a) and selecting the positive root

$$x = \frac{-1 - K_2(K_3L + 1)(N_y - N_x) + \sqrt{\Delta}}{2K_2(K_3L + 1)}, \tag{4.26}$$

where $\Delta = (1 + K_2(K_3L + 1)(N_y - N_x))^2 + 4N_xK_2(K_3L + 1)$. An expression for y can be found by solving equation (4.25b) knowing x, but is not relevant for the computation of the amplitude and EC_{50}. Indeed, the signalling function, defined as:

$$\sigma(L) \equiv K_3K_2Lxy, \tag{4.27}$$

can be rewritten as follows

$$\sigma(L) = \frac{K_3L(N_x - x)}{K_3L + 1}, \tag{4.28}$$

where we made used of equation (4.25b). The derivative with respect to L of σ is

$$\frac{d\sigma}{dL} = \frac{N_x - x}{(K_3L + 1)^2} - \frac{K_3L}{1 + K_3L}\frac{dx}{dL}, \tag{4.29}$$

where

$$\frac{dx}{dL} = -K_3\frac{(1 + K_2(1 + K_3L)(N_x + N_y) - \sqrt{\Delta})}{2K_2(1 + K_3L)^2\sqrt{\Delta}}. \tag{4.30}$$

Since $1 + K_2(1 + K_3L)(N_x + N_y) \geq \sqrt{\Delta} \geq 1$, $\frac{dx}{dL}$ is negative. Hence, $\frac{d\sigma}{dL} > 0$ and σ is an increasing function of L. As $\sigma(0) = 0$, the amplitude of the heterodimeric RTK model is the maximum of the signalling function, which is reached at high

ligand concentration. As $L \to +\infty$, we have

$$\begin{aligned} x &= o\left(\frac{1}{L}\right) && \text{if } N_x < N_y, \\ x &= N_x - N_y + o\left(\frac{1}{L}\right) && \text{if } N_y < N_x. \end{aligned} \tag{4.31}$$

Thus, the amplitude of the model is:

$$A \equiv \lim_{L \to +\infty} \sigma(L) = \min(N_x, N_y). \tag{4.32}$$

The EC_{50} is computed by finding L_{50} such that:

$$\sigma(L_{50}) = \frac{\min(N_x, N_y)}{2}. \tag{4.33}$$

Once again, following the method described in Section 4.1, we compute a Gröbner basis of the following system where we consider x, y and L as variables:

$$\begin{aligned} -N_y + y + K_2xy + K_3K_2Lxy, \\ -N_x + x + K_2xy + K_3K_2Lxy, \\ \frac{A}{2} - K_3K_2Lxy. \end{aligned} \tag{4.34}$$

Making use of the software Macaulay2 (Grayson & Stillman), the resulting Gröbner basis yields the following system of polynomial equations:

$$\begin{aligned} 0 &= L^2 - 2A\frac{1 + K_2(N_x + N_y - A)}{K_2K_3(A - 2N_x)(A - 2N_y)}L + \frac{A^2}{K_3^2(A - 2N_x)(A - 2N_y)}, \\ 0 &= y - K_3\frac{(A - 2N_x)(A - 2N_y)}{2A}L + \frac{1}{K_2} - \frac{A}{2} + N_x, \\ 0 &= x - K_3\frac{(A - 2N_x)(A - 2N_y)}{2A}L + \frac{1}{K_2} - \frac{A}{2} + N_y. \end{aligned} \tag{4.35}$$

We obtain two expressions for L:

$$L_1 = A\frac{1 + K_2(N_x + N_y - A) + \sqrt{\Delta}}{K_2K_3(A - 2N_x)(A - 2N_y)}, \tag{4.36}$$

$$L_2 = A\frac{1 + K_2(N_x + N_y - A) - \sqrt{\Delta}}{K_2K_3(A - 2N_x)(A - 2N_y)}, \tag{4.37}$$

and associated expressions for x and y:

$$\begin{aligned} x_1 &= \frac{-1 + K_2(N_x - N_y) + \sqrt{\Delta}}{2K_2}, \quad y_1 = \frac{-1 + K_2(N_y - N_x) + \sqrt{\Delta}}{2K_2}, \\ x_2 &= \frac{-1 + K_2(N_x - N_y) - \sqrt{\Delta}}{2AK_2K_3}, \quad y_2 = \frac{-1 + K_2(N_y - N_x) - \sqrt{\Delta}}{2K_2}, \end{aligned} \tag{4.38}$$

where we wrote $\Delta = 1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - A)$. Only positive expressions for x, y and L are biologically relevant. Since x_2 and y_2 are negative, L_1 is the EC_{50} of the model:

$$EC_{50} = A\frac{1 + K_2(N_x + N_y - A) + \sqrt{1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y - A)}}{K_2K_3(A - 2N_x)(A - 2N_y)}, \tag{4.39}$$

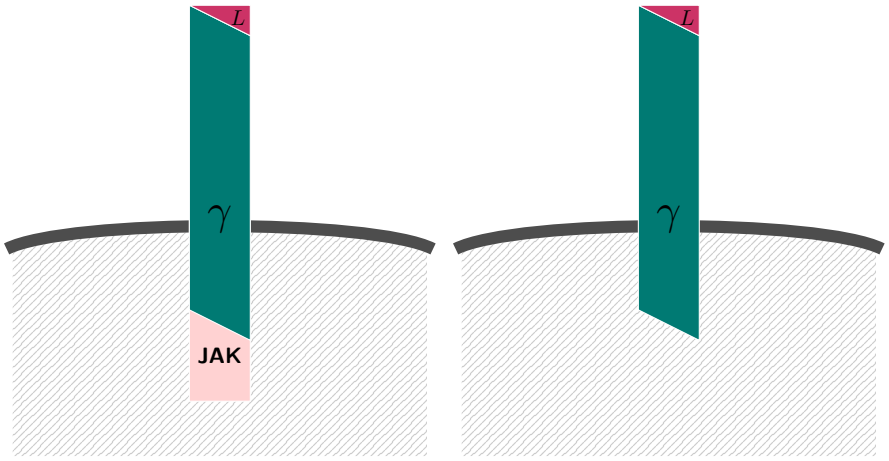
where $A = \min(N_x, N_y)$. One can notice that this EC_{50} expression is the same as the EC_{50} expression obtained for the IL-7R model (heterodimeric receptor with IEK) of Section 3.2.

4.2.4 Monomeric receptor model with IEK

Now, consider the three previous configurations but, this time, the kinase is extrinsic to the receptor chains. Signalling complexes as well as "dummy" complexes (receptors deprived of kinase which cannot induce any signal) can be formed. Note that the heterodimeric model with IEK was studied in Section 3.2. We remind the reader that we assume that there is no allostery (signalling and "dummy" complexes are formed with the same binding rates).

The monomeric receptor model with intra-cellular extrinsic kinase (IEK) can form a signalling complex (composed of the primary chain, the kinase and the ligand) and a "dummy" complex (composed of the ligand and primary chain only), as illustrated in Figure 4.4. The reaction scheme is as follows:

$$\begin{array}{lcll} JAK & + \ \gamma & \rightleftharpoons \ JAK:\gamma & K_1, \\ L & + \ \gamma:JAK & \rightleftharpoons \ L:\gamma:JAK & K_3, \\ L & + \ \gamma & \rightleftharpoons \ L:\gamma & K_3. \end{array}$$



(a) Signalling complex (b) Dummy complex

Figure 4.4: Monomeric model with intra-cellular extrinsic kinase: A kinase can bind to the unique chain γ of the monomeric receptor thus forming a signalling receptor. The ligand, L, can bind to the signalling receptor to form a signalling complex (left) or directly to a γ chain to form a "dummy" complex unable to signal (right). The hatched area determines the intra-cellular environment.

Writing the associated ordinary differential equations, steady state and conservations equations can be derived and combined to obtain the following polynomial system:

$$\begin{aligned} 0 &= -N_x + x + K_1xz + K_3Lx + K_3K_1Lxz, \\ 0 &= -N_z + z + K1xz + K_3K_1Lxz. \end{aligned} \tag{4.40}$$

Computing a Gröbner basis of (or rearranging by hand) this system leads to a triangular system

$$\begin{aligned} 0 &= -N_z + z + K_1(N_x - N_z)z + K_1z^2, \\ 0 &= N_x - N_z + z - x(1 + K_3L), \end{aligned} \tag{4.41}$$

which can be solved to obtain analytic expressions for the copy numbers of the receptors component chains at steady state:

$$\begin{aligned} z &= \frac{-1 + K_1(N_z - N_x) + \sqrt{4K_1N_z + (1 + K_1(N_x - N_z))^2}}{2K_1}, \\ x &= \frac{N_x - N_z + z}{1 + K_3L}. \end{aligned} \tag{4.42}$$

The signalling function is defined as

$$\sigma(L) \equiv K_3 K_1 L x z. \tag{4.43}$$

One can notice from equation (4.41) that z is independent of L and that $N_x - N_z + z = \dfrac{N_x}{1 + K_1 z}$. Thus, σ can be re-written as

$$\sigma(L) = \frac{K_3 L}{1 + K_3 L} \frac{K_1 z}{1 + K_1 z} N_x. \tag{4.44}$$

Similarly to the monomeric model with intrinsic kinase, one can recognise a Hill function with Hill coefficient $n = 1$. By identification, the amplitude is

$$A = \frac{K_1 z}{1 + K_1 z} N_x, \tag{4.45}$$

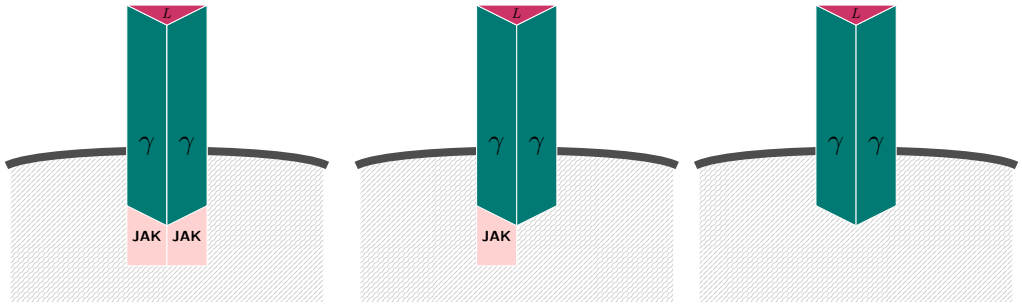
and the EC_{50} is

$$EC_{50} = \frac{1}{K_3}. \tag{4.46}$$

This time, if the binding of the kinase to the main chain γ is not fast enough (if K_1 is small), "dummy" complexes will be formed and the maximum response will not reach its theoretical maximum, N_x. Instead, the signalling function only reached a fraction, $\dfrac{K_1 z}{1 + K_1 z}$, of it. One can also notice that the EC_{50} of this system is the same as the EC_{50} of the monomeric model with intrinsic kinase described in Section 4.2.1.

4.2.5 Homodimeric receptor model with IEK

Next, I consider the homodimeric receptor with extrinsic downstream kinase. The receptor is composed of two γ chains, each of them binding to a kinase. Thus, two non-signalling ("dummy") complexes can be formed: a complex composed of the two γ chains but no kinase and a complex formed of a γ chain bound to a kinase and a γ chain without kinase (see Figure 4.5). The chemical reactions are



(a) Signalling complex (b) Dummy complex 1 (c) Dummy complex 2

Figure 4.5: Homodimeric receptor model with intra-cellular extrinsic kinase: the receptor is composed of two γ chains (none, one or both can be bound to a kinase). The ligand, L, binds to this homodimer. If both chains are bound to a kinase, the binding of the ligand forms a signalling complex (left). Otherwise, it forms a complex unable to signal called "dummy" complex (middle and right). The hatched area determines the intra-cellular environment.

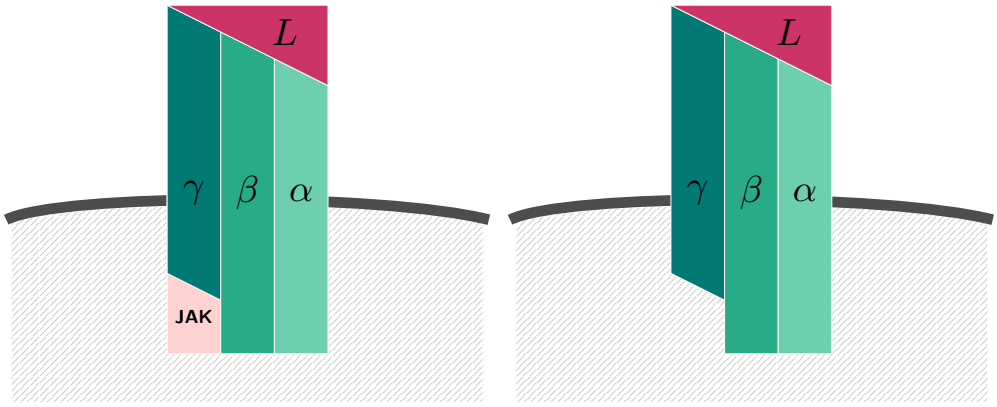
as follows:

$$\begin{array}{lcll}
JAK+\gamma & \rightleftharpoons & JAK:\gamma & K_1,\\
JAK:\gamma+\gamma & \rightleftharpoons & \gamma:\gamma:JAK & K_2,\\
JAK:\gamma+JAK:\gamma & \rightleftharpoons & JAK3:\gamma:\gamma:JAK & K_2,\\
\gamma+\gamma & \rightleftharpoons & \gamma:\gamma & K_2,\\
L+\gamma:\gamma & \rightleftharpoons & L:\gamma:\gamma & K_3,\\
L+\gamma:\gamma:JAK & \rightleftharpoons & L:\gamma:\gamma:JAK & K_3,\\
L+\gamma:JAK:\gamma:JAK & \rightleftharpoons & L:\gamma:JAK:\gamma:JAK & K_3.
\end{array}$$

Proceeding following the method described in Section 4.1, one can write the differential equations associated to the model. Combining the steady state and conservation equations, we obtain the following polynomial system:

$$\begin{aligned}
0 =& -N_x + x + K_1xz + 2(K_1K_2x^2z + K_2K_1^2x^2z^2 + K_2x^2 + K_3K_2Lx^2 \qquad (4.47a)\\
&+ K_3K_2K_1Lx^2z + K_3K_2K_1^2Lx^2z^2),\\
0 =& -N_z + z + K_1xz + K_1K_2x^2z + 2K_1^2K_2x^2z^2 + K_3K_2K_1Lx^2z \qquad (4.47b)\\
&+ 2K_3K_2K_1^2Lx^2z^2.
\end{aligned}$$

The Gröbner basis of this polynomial system is composed of very long polynomials (the first univariate polynomial is shown in Appendix C). The signalling function



(a) Signalling complex (b) Dummy complex

Figure 4.6: Trimeric receptor model with intra-cellular extrinsic kinase: the receptor is composed of three different chains, γ, β and α. The ligand, L, binds to this trimer: if the receptor is bound to the kinase, it forms a signalling complex (a). The ligand bound to the receptor deprived of kinase forms a "dummy" complex (b). The hatched area determines the intra-cellular environment.

$K_3K_2K_1Lx^2z$ is a sigmoid and one could use perturbation theory to compute the amplitude. However due to the length and high degree (degree 6 for the first polynomial) of the polynomials of the Gröbner basis, the analytic study is intractable and numerical analysis is preferred.

4.2.6 Trimeric receptor model with IEK

Finally, we consider a receptor composed of three chains (γ, β and α), such as IL-2R or the IL-15R (Rochman *et al.*, 2009), which signals through the downstream kinase JAK. Similarly to the models previously described in this thesis, signalling and "dummy" complexes can be formed (see Figure 4.6) with the same affinity constants (we assume no allostery), and the receptor constituents are assembled

sequentially. The reaction scheme is summarised follows:

$$\begin{array}{lcll}
JAK & + \gamma & \rightleftharpoons JAK:\gamma & K_1, \\
\beta & + \gamma:JAK & \rightleftharpoons \beta:\gamma:JAK & K_2, \\
\beta & + \gamma & \rightleftharpoons \beta:\gamma & K_2, \\
\alpha & + \beta:\gamma:JAK & \rightleftharpoons \alpha:\beta:\gamma:JAK & K_3, \\
\alpha & + \beta:\gamma & \rightleftharpoons \alpha:\beta:\gamma & K_3, \\
L & + \alpha:\beta:\gamma:JAK & \rightleftharpoons L:\alpha:\beta:\gamma:JAK & K_4, \\
L & + \alpha:\beta:\gamma & \rightleftharpoons L:\alpha:\beta:\gamma & K_4.
\end{array}$$

Note that for this model, we denoted by K_4 the affinity constant of the binding of the ligand with the receptor, and K_2 and K_3 the binding constants of the secondary chain and third chain, respectively. Combining conservation and steady state equations obtained from the differential equations associated to this reaction scheme, we obtain the following polynomial system:

$$\begin{aligned}
0 &= -N_z + z + K_1xz + K_2K_1xwz + K_3K_2K_1xyzw + K_4K_3K_2K_1Lxyzw, \\
0 &= -N_x + x + K_1xz + K_2xw + K_2K_1xwz + K_3K_2K_1xywz + K_3K_2xwy \\
&\quad + K_4K_3K_2K_1Lxywz + K_4K_3K_2Lxyw, \\
0 &= -N_y + y + K_3K_2K_1xyzw + K_3K_2xwy + K_4K_3K_2Lxyw \\
&\quad + K_4K_3K_2K_1Lxywz, \\
0 &= -N_w + w + K_2xw + K_2K_1xwz + K_3K_2K_1xyzw + K_3K_2xwy \\
&\quad + K_4K_3K_2Lxwy + K_4K_3K_2K_1Lxyzw.
\end{aligned} \tag{4.48}$$

In addition to the usual notation, we denote by w the concentration of unbound β chains at steady state and N_w its total copy number. Let us compute the amplitude of the trimeric receptor model with IEK. Note that a more elegant computation of this quantity will follow from the generalisation of Section 4.3 (SRLK model with $n = 3$). Making use of Macaulay2 (Grayson & Stillman), we compute a Gröbner basis of this polynomial system and obtain:

$$0 = Aw^4 + Bw^3 + Cw^2 + Dw + E, \tag{4.49a}$$

$$0 = -N_z + (1 + K_1(N_x - N_z))z + K_1z^2, \tag{4.49b}$$

$$0 = N_w - N_y + (-1 + K_2(N_w - N_x))w - K_2w^2 + y, \tag{4.49c}$$

$$0 = x + \frac{(N_z - N_x - w)(N_x - N_z + z)}{N_x}, \tag{4.49d}$$

where

$$A = K_2^2 K_3(1 + K_4 L),$$
$$B = -K_2 K_3(1 + K_4 L)(-1 + 2K_2(N_w - N_x)),$$
$$C = K_2 + K_2 K_3(1 + K_4 L)(K_2 N_w^2 + N_x + K_2 N_x^2 - 2N_w(1 + K_2, N_x) + N_y),$$
$$D = 1 + K_2(N_w - N_x)(-1 + K_3(1 + K_4 L)(N_w - N_y)),$$
$$E = -N_w.$$

One can notice from equation (4.49b) that $\frac{N_x - N_z + z}{N_x} = \frac{1}{1 + K_1 z}$. It follows that the biologically meaningful solutions of equations (4.49b), (4.49c) and (4.49d), are:

$$z = \frac{-1 + K_1(N_z - N_x) + \sqrt{\Delta_1}}{2K_1}, \tag{4.50a}$$
$$x = \frac{N_x - N_w + w}{1 + K_1 z}, \tag{4.50b}$$
$$y = N_y - N_w + (1 + K_2(N_x - N_w))w + K_2 w^2, \tag{4.50c}$$

where $\Delta_1 = (1 + K_1(N_x - N_z))^2 + 4K_1 N_z$ and w, kept unknown, is a positive real root of (4.49a). We define the signalling function of this model as

$$\sigma(L) \equiv K_4 K_3 K_2 K_1 L x y z w. \tag{4.51}$$

The concentration of JAK at steady state, z, is, once again, independent from the ligand concentration, L. Replacing the expression of x and y in σ, we obtain a signalling function expression which only depends on w:

$$\sigma(L) = K_2 K_3 K_4 L \frac{K_1 z}{1 + K_1 z} P_w(w), \tag{4.52}$$

where we defined P_w as

$$P_w(w) \equiv w(N_x - N_w + w)(N_y - N_w + (1 + K_2(N_x - N_w))w + K_2 w^2). \tag{4.53}$$

A simple inspection of (4.52) shows that the signalling function is a sigmoid and $\sigma(0) = 0$. The amplitude is, thus, the limit of σ at high concentration (this will later be proven in Section 4.3):

$$\begin{aligned} A &= \lim_{L\to+\infty} \sigma(L) \\ &= K_4K_3K_2\frac{K_1 z}{1+K_1 z}\lim_{L\to+\infty}(LP_w(w)). \end{aligned} \tag{4.54}$$

The number of unbound β chains at steady state, w, satisfies a polynomial of degree 4 (equation (4.49a)). We could, in principle, find an exact analytic solution of this quartic function. However, such a solution will probably be overly complicated and thus, not be very informative. To compute the amplitude, we only need to find the behaviour of w as $L \to +\infty$. To this end and analogously to Section 3.4, we apply the perturbation theory method described in Ref. Simmonds & Mann (2013) and summarised in Section 2.5.

We define $P_\epsilon(w)$ as the polynomial in equation (4.49a) where we replaced L by $\frac{1}{\epsilon}$. We replace w by $\epsilon^p\omega(\epsilon)$ where $\omega(0) \neq 0$ according to Theorem 49 (Section 2.5). We obtain:

$$P_\epsilon(\epsilon^p\omega) = A_\epsilon\omega^4 + B_\epsilon\omega^3 + C_\epsilon\omega^2 + D_\epsilon\omega + E_\epsilon, \tag{4.55}$$

where

$$\begin{aligned} A_\epsilon &= \epsilon^{4p-1}(\epsilon K_2^2K_3 + K_2^2K_3K_4), \\ B_\epsilon &= \epsilon^{3p-1}(-\epsilon K_2K_3(-1+2K_2(N_w-N_x)) - K_2K_3K_4(-1+2K_2(N_w-N_x))), \\ C_\epsilon &= \epsilon^{2p-1}(K_2K_3K_4(K_2N_w^2 + N_x + K_2N_x^2 - 2N_w(1+K_2N_x) + N_y) \\ &\quad + \epsilon K_2(1+K_3(K_2N_w^2 + N_x + K_2N_x^2 - 2N_w(1+K_2N_x) + N_y))), \\ D_\epsilon &= \epsilon^{p-1}(\epsilon(1+K_2(N_w-N_x)(-1+K_3(N_w-N_y))) \\ &\quad + K_2K_3K_4(N_w-N_x)(N_w-N_y)), \\ E_\epsilon &= -N_w. \end{aligned}$$

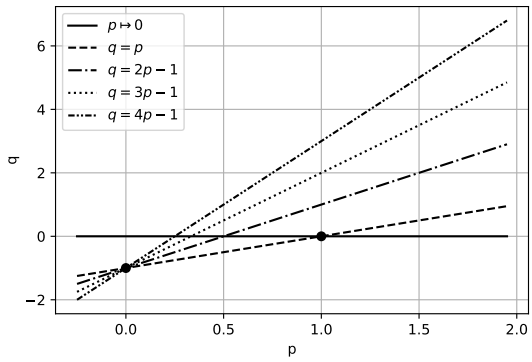


Figure 4.7: The list E and the proper values (black dots).

The smallest exponents of ϵ in the previous equation are:

$$E = \{0, p-1, 2p-1, 3p-1, 4p-1\}. \tag{4.56}$$

Applying the graphical algorithm detailed in the mathematical background Section 2.5, we find the proper values $(0, -1)$ and $(1, 0)$ (see Figure 4.7). We investigate these two branches. Note that the following calculations were obtained with the help of the symbolic computation software Mathematica (Wolfram Research, Inc., 2019) when necessary.

Branch $(0, -1)$**:** We make use of the notation in Section 2.5 to define

$$T_\epsilon^{(1)}(\omega) = \epsilon P_\epsilon(\omega). \tag{4.57}$$

The least common denominator of $(0, -1, -1, -1, -1)$ is $q_1 = 1$. Therefore in accordance with the notation in Section 2.5, $\epsilon = \beta$ and

$$R_\beta^{(1)}(\omega) \equiv T_\epsilon^{(1)}(\omega).$$

We can directly carry out a regular perturbation expansion, $\omega(\epsilon) = w_0 + w_1\epsilon + ...$, of $w = \epsilon^0\omega(\epsilon)$.

Let us write the asymptotic expansion $w = \omega = w_0 + w_1\epsilon + ...$ and replace it in $T_\epsilon^{(1)}(\omega)$. Since $P_\epsilon(w) = 0$, by the fundamental theorem of perturbation theory (Theorem 48), we obtain a system of equations in w_0, $w_1, \ldots$, which can be solved.

The first equation of the system is given by:

$$w_0 K_2 K_3 K_4 (w_0 - N_w + N_x)(w_0 - N_w + w_0 K_2 (w_0 - N_w + N_x) + N_y) = 0, \quad (4.58)$$

which yields four solutions for w_0:

$$w_{01} = 0, \quad (4.59a)$$

$$w_{02} = N_w - N_x, \quad (4.59b)$$

$$w_{03} = \frac{-1 + K_2(N_w - N_x) + \sqrt{(1 + K_2(N_x - N_w))^2 - 4K_2(N_y - N_w)}}{2K_2}, \quad (4.59c)$$

$$w_{04} = \frac{-1 + K_2(N_w - N_x) - \sqrt{(1 + K_2(N_x - N_w))^2 - 4K_2(N_y - N_w)}}{2K_2}. \quad (4.59d)$$

To be biologically relevant, the value of $w_0 = \lim_{L \to +\infty} w$ must be real and non-negative. Moreover, w_0 must satisfy additional constraints so that x, y (obtained, respectively, by equations (4.50b) and (4.50c)) are also positive at high concentration.

Case $w_0 = w_{01}$: If $w_0 = w_{01} = 0$, then

$$\begin{aligned} x &\underset{L \to +\infty}{\sim} \frac{N_x - N_w}{1 + K_1 z}, \\ y &\underset{L \to +\infty}{\sim} N_y - N_w, \\ w &\underset{L \to +\infty}{\sim} \frac{w_1}{L}. \end{aligned} \quad (4.60)$$

We first need to assume $N_x > N_w$ and $N_y > N_w$ so that x and y are positive. The coefficient w_1 is found by replacing w by its asymptotic expansion in $T_\epsilon^{(1)}$, assuming $w_0 = 0$ and applying the fundamental theorem of perturbation theory (Theorem 48). We obtain:

$$w_1 = \frac{N_w}{K_2 K_3 K_4 (N_w - N_x)(N_w - N_y)}. \quad (4.61)$$

Under the conditions $N_x > N_w$ and $N_y > N_w$, w_1 is positive so x, y and w are positive at high concentration. Thus,

$$w_0 = 0 \text{ when } N_x > N_w \text{ and } N_y > N_w. \tag{4.62}$$

Case $w_0 = w_{02}$: If $w_0 = w_{02} = N_w - N_x$, then we have

$$\begin{aligned} x &\underset{L\to+\infty}{\sim} \frac{w_1}{1+K_1 z}, \\ y &\underset{L\to+\infty}{\sim} N_y - N_x, \\ w &\underset{L\to+\infty}{\sim} N_w - N_x, \end{aligned} \tag{4.63}$$

Thus, we need to assume $N_w > N_x$ and $N_y > N_x$ so that w and y are positive. We obtain w_1 by replacing w by its expansion in $T_\epsilon^{(1)}$ assuming $w_0 = N_w - N_x$ and applying the fundamental theorem of perturbation theory (Theorem 48). We obtain:

$$w_1 = \frac{-N_x}{K_2 K_3 K_4 (N_w - N_x)(N_x - N_y)}. \tag{4.64}$$

Under the conditions $N_y > N_x$ and $N_w > N_x$, w_1 is positive. Thus,

$$w_0 = N_w - N_x \text{ when } N_w > N_x \text{ and } N_y > N_x. \tag{4.65}$$

Case $w_0 = w_{03}$ or $w_0 = w_{04}$: Let us first note that w_{03} and w_{04} are real when what is inside the square root $\Delta \equiv (1+K_2(N_x-N_w))^2 - 4K_2(N_y-N_w)$ is positive.

In both cases, we have

$$\begin{aligned} w &\underset{L\to+\infty}{\sim} w_0, \\ x &\underset{L\to+\infty}{\sim} \frac{N_x - N_w + w_0}{1+K_1 z}, \\ y &\underset{L\to+\infty}{\sim} \begin{cases} \dfrac{w_1\sqrt{\Delta}}{L} & \text{if } w_0 = w_{03} \\ -\dfrac{w_1\sqrt{\Delta}}{L} & \text{if } w_0 = w_{04} \end{cases}. \end{aligned} \tag{4.66}$$

To be biologically relevant, we need simultaneously $w_0 > 0$, $N_x - N_w + w_0 > 0$ and $y > 0$. Let us investigate the first two conditions. To shorten the notation, we

write $b = K_2(N_x - N_w)$ and re-write Δ as $\Delta = (1+b)^2 - 4K_2(N_y - N_w)$. Thus,

$$w_0 = \frac{-1 - b \pm \sqrt{\Delta}}{2K_2}, \tag{4.67a}$$

$$N_x - N_w + w_0 = \frac{-1 + b \pm \sqrt{\Delta}}{2K_2}, \tag{4.67b}$$

where $\pm$ is a plus sign if $w_0 = w_{03}$ and minus sign if $w_0 = w_{04}$.

To obtain $w_0 > 0$ and $N_x - N_w + w_0 > 0$, we need:

$$\sqrt{\Delta} > 1 + b \qquad \text{and} \qquad \sqrt{\Delta} > 1 - b \qquad \text{if } w_0 = w_{03}, \tag{4.68}$$

$$-1 - b > \sqrt{\Delta} \qquad \text{and} \qquad -1 + b > \sqrt{\Delta} \qquad \text{if } w_0 = w_{04}. \tag{4.69}$$

If $b > 0$ ($N_x > N_w$), then $1 + b > 1 > 1 - b$, so the stricter condition in (4.68) is $\sqrt{\Delta} > 1 + b$, which is satisfied when $N_w > N_y$. If $b < 0$ ($N_w < N_x$), we have $1 + b < 1 < 1 - b$, so the stricter condition in (4.68) is $\sqrt{\Delta} > 1 - b$, which is satisfied when $N_x > N_y$. Note that if $b > 0$, then $-1 - b < 0$, and if $b < 0$, then $-1 + b < 0$. Since we want $\Delta > 0$ and so $\sqrt{\Delta} > 0$, condition (4.69) can never be satisfied. Thus, w_{04} is not an option for w_0.

Finally, let us study the sign of w_1 (to study the sign of y). We obtain w_1 as a function of w_0 by replacing w by its asymptotic expansion in $T_\epsilon^{(1)}$ and applying the fundamental theorem of perturbation theory (Theorem 48):

$$w_1(w_0) = \frac{N_w + w_0(-1 - K_2(w_0 - N_w + N_x)(1 + K_3 y(w_0)))}{K_2 K_3 K_4 Q(w)}, \tag{4.70}$$

where we defined

$$\begin{aligned} Q(w) \equiv\ & 4w_0^3 K_2 + 3w_0^2(1 + 2K_2(N_x - N_w)) \\ & + 2w_0(K_2(N_w - N_x)^2 + N_x - 2N_w + N_y) + (N_w - N_x)(N_w - N_y), \end{aligned}$$

and $y(w_0)$ is the expression for y obtained in (4.50c) evaluated at $w = w_0$. Notice that $y(w_{03}) = 0$. In this case ($w_0 = w_{03}$), the numerator of w_1 can be simplified

into:

$$N_w + w_0(-1 + K_2(N_w - N_x)) - K_2 w_0^2 = N_y - y(w_0) = N_y. \tag{4.71}$$

Substituting $w_0 = w_{03}$ in (4.70), we obtain thanks to Mathematica (Wolfram Research, Inc., 2019):

$$w_1(w_0 = w_{03}) = \frac{2K_2 N_y}{K_3 K_4 \sqrt{\Delta}(1 - \sqrt{\Delta} + K_2(N_w + N_x - 2N_y))}. \tag{4.72}$$

Now that we obtained an expression for w_1, let us check it is indeed positive (so that $y > 0$). The sign of w_1 is the sign of $1 - \sqrt{\Delta} + K_2(N_w + N_x - 2N_y)$. We have

$$\begin{aligned}
1 - \sqrt{\Delta} + K_2(N_w + N_x - 2N_y) > 0 &\iff 1 + K_2(N_w + N_x - 2N_y) > \sqrt{\Delta} \\
&\implies (1 + K_2(N_x + N_w - 2N_y))^2 > \Delta \\
&\iff (N_x + N_w - 2N_y)^2 > (N_x - N_w)^2 \\
&\iff 2N_x N_w + 4N_y^2 + 4N_y N_w + 4N_w N_x > -2N_x N_w \\
&\iff (N_y - N_x)(N_y - N_w) > 0.
\end{aligned}$$

The last inequality is always true when we assume $N_y < N_x$ and $N_y < N_w$ which were the conditions so that $w_0 > 0$ and $N_x - N_w + w_0 > 0$. Thus, when $N_y < N_w$ and $N_y < N_x$,

$$w_0 = \frac{-1 + K_2(N_w - N_x) + \sqrt{(1 + K_2(N_x - N_w))^2 - 4K_2(N_y - N_w)}}{2K_2}. \tag{4.74}$$

We can now recapitulate the values of w_0 and w_1 and their domain of definition for this branch in table 4.2.

Domain	w_0	w_1
$N_x, N_y > N_w$	0	$\frac{N_w}{K_2K_3K_4(N_w - N_x)(N_w - N_y)}$
$N_y, N_w > N_x$	$N_w - N_x$	$\frac{-N_x}{K_2K_3K_4(N_w - N_x)(N_x - N_y)}$
$N_x, N_w > N_y$	w_{03}	$\frac{2K_2N_y}{K_3K_4\sqrt{\Delta}(1 - \sqrt{\Delta} + K_2(N_w + N_x - 2N_y))}$

Table 4.2: Values of w_0 and w_1 with their domain of definition for the branch $(0, -1)$. We wrote $\Delta = (1 + K_2(N_x - N_w))^2 - 4K_2(N_y - N_w)$ and $w_{03} = \frac{-1 + K_2(N_w - N_x) + \sqrt{(1 + K_2(N_x - N_w))^2 - 4K_2(N_y - N_w)}}{2K_2}$.

Branch $(1, 0)$**:** We now study the other branch obtained by the perturbation theory method. Once again,we make use of the notation in Section 2.5 to define

$$T_\epsilon^{(2)}(\omega) = P_\epsilon(\epsilon\omega). \tag{4.75}$$

The least common denominator of $(0, -1+1, -1+2, -1+3, -1+4)$ is $q_2 = 1$. Therefore in accordance with the notation in Section 2.5, $\epsilon = \beta$ and

$$R_\beta^{(2)}(\omega) \equiv T_\epsilon^{(2)}(\omega).$$

We can directly carry out a regular perturbation expansion. We replace ω by an expansion $a_0 + a_1\epsilon + a_2\epsilon^2 + ...$ in $T_\epsilon^{(2)}$ and apply the fundamental theorem of perturbation theory (Theorem 48). In this branch, since we have $w = \epsilon\omega$, $w \to 0$ when $\epsilon \to 0^+$ (*i.e.*, when $L = \frac{1}{\epsilon} \to +\infty$). Furthermore, $a_0 = w_1$, where w_1 is the coefficient of the asymptotic expansion obtained in the previous branch. In other words, at large, but finite $L = 1/\epsilon$, the convergence behaviour of the two branches is identical. This agrees with Theorem 37 which states that there is only one positive solution for each set of reaction constants and initial conditions.

Back to the amplitude: The behaviour of the signalling function in the limit of $L \to +\infty$ is a constant which determines the amplitude A:

$$\sigma(L) \underset{L\to+\infty}{\sim} K_4K_3K_2K_1zLx_\infty y_\infty w_\infty \equiv \text{constant} \equiv A, \tag{4.76}$$

where x_∞, y_∞ and w_∞ are the behaviours of x, y and w, respectively, when $L \to +\infty$. They are recapitulated in equations (4.60), (4.63) and (4.66).

When $N_x, N_y > N_w$ ($w_0 = 0$), we obtain:

$$\begin{aligned} A(w_0 = 0) &= K_4K_3K_2\frac{K_1}{1+K_1z}\frac{L(N_x-N_w)(N_y-N_w)N_w}{K_2K_3K_4(N_w-N_x)(N_w-N_y)L} \\ &= \frac{K_1z}{1+K_1z}N_w. \end{aligned} \tag{4.77}$$

When $N_y, N_w > N_x$ ($w_0 = N_w - N_x$), we obtain:

$$\begin{aligned} A(w_0 = N_w - N_x) &= K_4K_3K_2\frac{K_1z}{1+K_1z}\frac{-N_xL(N_y-N_x)(N_w-N_x)}{K_2K_3K_4L(N_w-N_x)(N_x-N_y)} \\ &= \frac{K_1z}{1+K_1z}N_x. \end{aligned} \tag{4.78}$$

When $N_x, N_w > N_y$, *i.e.*, when

$$w_0 = \frac{-1+K_2(N_w-N_x)+\sqrt{(1+K_2(N_x-N_w))^2-4K_2(N_y-N_w)}}{2K_2},$$

we obtain:

$$\begin{aligned} A(w_0 = w_{03}) &= K_4K_3K_2\frac{K_1z}{1+K_1z}\frac{w_0(N_x-N_w+w_0)Lw_1\sqrt{\Delta}}{L} \\ &= \frac{K_1z}{1+K_1z}\frac{2K_2^2N_yw_0(N_x-N_w+w_0)}{(1-\sqrt{\Delta}+K_2(N_w+N_x-2N_y))}. \end{aligned} \tag{4.79}$$

As previously, let us write $b = K_2(N_x - N_w)$ for concise notation. We compute

$w_0(N_x - N_w + w_0)$ from the expressions derived in equation (4.67):

$$\begin{aligned}\frac{(-1+\sqrt{\Delta})-b}{2K_2}\frac{(-1+\sqrt{\Delta})+b}{2K_2} &= \frac{(-1+\sqrt{\Delta})^2-b^2}{4K_2^2}\\ &= \frac{1+\Delta-2\sqrt{\Delta}-b^2}{4K_2^2}\\ &= \frac{2+b^2+2b-4K_2(N_y-N_w)-b^2-2\sqrt{\Delta}}{4K_2^2}\\ &= \frac{1-\sqrt{\Delta}+K_2(N_x+N_w-2N_y)}{2K_2^2}.\end{aligned} \tag{4.80}$$

Substituting this expression into the amplitude, we obtain:

$$A(w_0 = w_{03}) = \frac{K_1 z}{1+K_1 z}N_y. \tag{4.81}$$

Merging the three cases together, we obtain the analytic expression of the amplitude for the model of the trimeric receptor with IEK:

$$A = \frac{K_1 z}{1+K_1 z}\min(N_x, N_y, N_w), \tag{4.82}$$

where $z = \dfrac{-1+K_1(N_z-N_x)+\sqrt{\Delta_1}}{2K_1}$ has been computed in equation (4.50a). Notice that a very similar formula has been obtained for the amplitude of the monomeric and heterodimeric receptor models with IEK (Section 3.2 and 4.2.4). Indeed, the amplitude of these models is the total quantity of the limiting component (receptor chain with the smallest abundance) modulated by a ratio, $\dfrac{K_1 z}{1+K_1 z}$, that takes a value between 0 and 1. This ratio only depends on the total number of JAK and γ molecules, and the affinity constant of their binding, K_1.

Computation of the EC_{50}: Since we did not compute analytic expressions for each steady state concentration, the EC_{50} expression has to be obtained by computing a Gröbner basis of the polynomial system (4.49) augmented by the

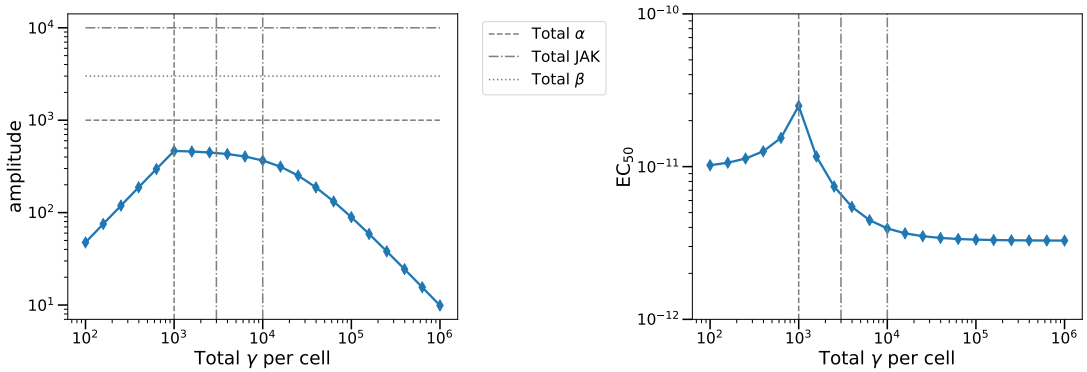


Figure 4.8: Amplitude (left) and EC_{50} (right) of the model of the trimeric receptor with IEK as a function of the total number of γ chains, for 10^3 α chains, 3×10^3 β chains and 10^4 JAK molecules per cells, $K_1 = 10^{-4.04}$ (as determined in the previous chapter), $K_2 = 3.3 \times 10^{-4}$, $K_3 = 3.7 \times 10^{-4}$ and $K_4 = 67 \times 10^{10} \mathrm{M}^{-1}$ as determined in Ref. Cotari *et al.* (2013b) for the IL-2 receptor. The amplitude was computed making use of the analytic expression (4.82). The EC_{50} was obtained by fitting the dose-response curve to a sigmoid equation. The dose-response curve was derived by solving system (4.48) for different ligand concentrations.

polynomial

$$K_4 K_3 K_2 K_1 z L x y w (1 + K_1 z) - \frac{K_1 z M}{2} = 0, \tag{4.83}$$

considering x, y, z, w and L as variables, with $M = \min(N_x, N_y, N_w)$. The Gröbner basis, which can be found in Appendix D, is composed of a polynomial of degree 4 in L, independent of N_z and K_1. The EC_{50} of this model is a positive root of this polynomial. Thus, the EC_{50} of the trimeric receptor model is, like the EC_{50} of the other models with IEK, independent of the kinase.

The amplitude increases then decreases as γ abundance increases, while the EC_{50} seems mostly independent from this quantity, especially at high γ concentration (see Figure 4.8). In light of the additional experiment presented in Figure 3.22, the trimeric receptor model of this section seems to be a good candidate to characterise the IL-2/IL-2R and IL-15/IL-15R systems. Note that Figure 4.8 was generated with the affinity constant values determined for the IL-2R in Cotari *et al.* (2013b). The value of K_1 inferred in Chapter 3 stands for the trimeric receptor model as long as the quantity of β chain is not limiting (as in that case,

the amplitude expressions of the IL-7R model and the trimeric receptor model are the same).

4.2.7 Discussion

The amplitude and EC_{50} of different receptor configurations (monomeric, homo- or heterodimeric with intra-cellular extrinsic or intrinsic kinase and trimeric receptor with extrinsic kinase) have just been computed when possible, following the method presented in Section 4.1. The calculated expressions are summarised in the tables in Section 2.10.

Variability and robustness of receptor-ligand systems

Making use of the analytic amplitude expressions previously computed, we comprehensively explore the impact of varied abundances of receptor onto cell's response and compare the different models of this section [1]. For every receptor system I considered the up/downregulation of the primary receptor chain γ. The mean abundance of the secondary receptor chain, α, was set to 10^3 proteins per cell and if a downstream kinase (IEK) is required, I set its level to 10^4 proteins per cell. These values are chosen similar to the average IL-7Rα and JAK3 abundances used in the simulations of the IL-7R models of Chapter 3. Finally, I took the following (no allostery) affinity constant values: $K_1 = 10^{-4.04}$ (as inferred in Chapter 3), $K_2 = 17 \times 10^{-3}$, $K_3 = 34 \times 10^{10}\mathrm{M}^{-1}$ (as determined in Ref. Cotari *et al.* (2013b)). The theoretical results and numerical explorations indicate a striking difference for the effect of primary receptor abundance depending on the receptor arrangement. Indeed, when the primary chain is the only one in the receptor (monomeric or homodimeric RTK), increasing the abundance of that chain proportionally increases the amplitude. This results in a large potential for enhanced signalling (Fig. 4.9). As always, the less abundant chain determines the theoretical maximum of the amplitude. However, this maximum is not attained for receptors with downstream kinase due to the formation of "dummy" complexes

[1] Note that the trimeric receptor model is excluded from the following analysis as the definition of the affinity constants is different and its RTK version is not considered in this manuscript.

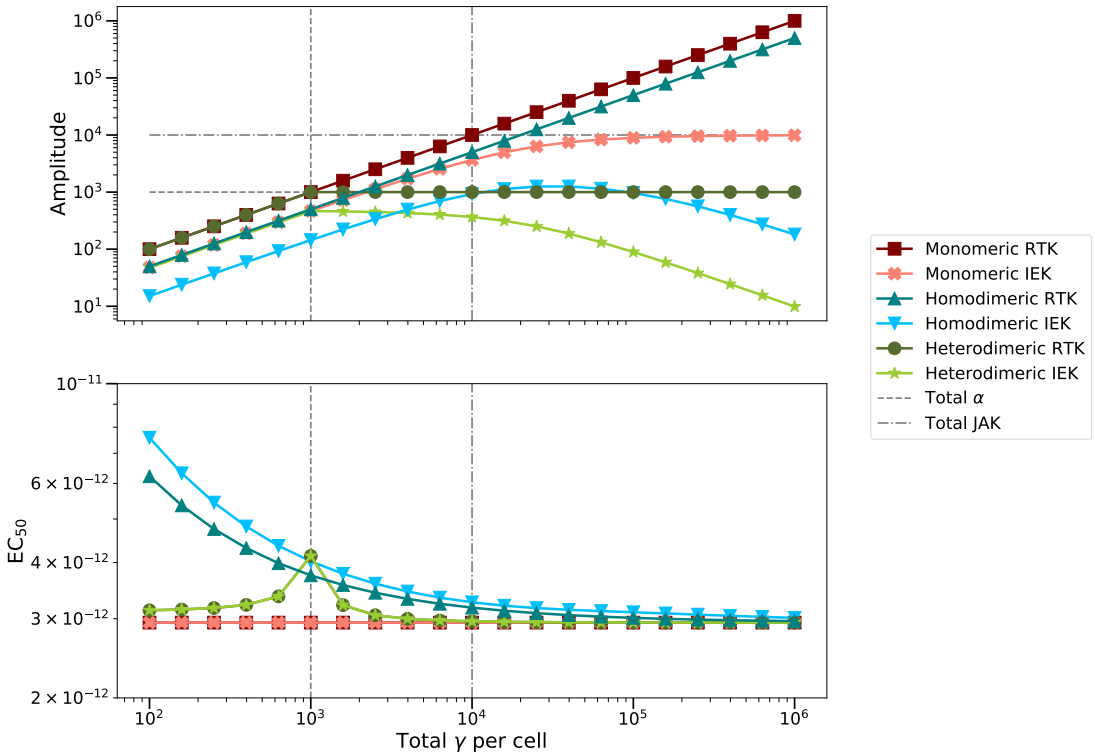


Figure 4.9: Amplitude (top) and EC_{50} (bottom) of the monomeric, homodimeric and heterodimeric RTK and with intra-cellular extrinsic kinase (IEK), for 10^3 α chains and 10^4 JAK molecules per cells, $K_1 = 10^{-4.04}$ (as determined in the previous chapter), $K_2 = 17 \times 10^{-3}$ and $K_3 = 34 \times 10^{10}\mathrm{M}^{-1}$.

(this is a direct illustration of how the ratio $\dfrac{K_1 z}{1 + K_1 z}$ modulates the amplitude of receptors with IEK). Homodimeric and heterodimeric receptors with IEK present a loss of signalling when the number of primary chains increases, thus, rendering the tuning of the response by the primary chain abundance more complex than RTKs.

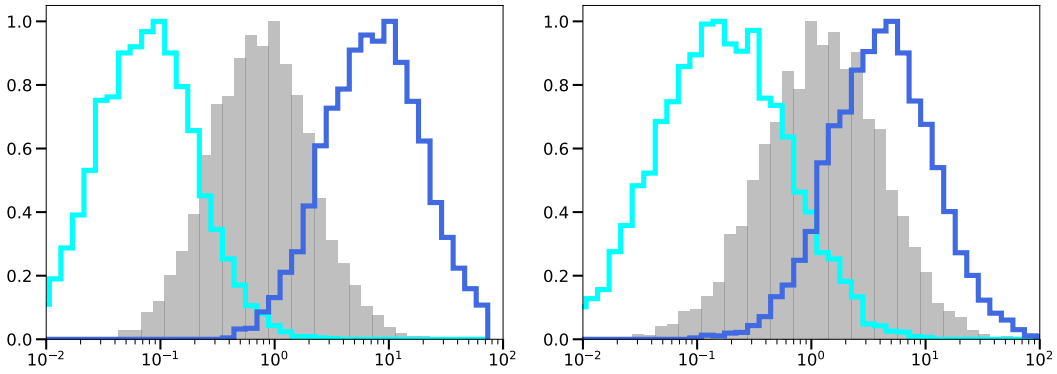
I also simulated cellular populations for each of the arrangements, with the distribution of each protein described by a log-normal distribution with a standard deviation of 1 (Fig. 4.10). The abundance of the secondary receptor chain (if present) was set to have a mean of 10^3 and downstream kinase a mean of $10^{3.5}$. We then simulated populations with a 10-fold up/downregulation of the primary chain, with a mean of 10^4 for the wild type (reference distribution). Homodimeric

RTKs or monomeric RTKs had the greatest variability of amplitude, matching the distribution of the receptor exactly (as expected from the amplitude expressions). Homodimeric receptors with IEK and heterodimeric RTKs had the least variability, maintaining amplitude at a moderately high level. Heterodimeric receptors with IEK had larger variability, but had lower amplitude when upregulating and downregulating the primary chain by 10-fold. Monomeric IEK had a larger variation in amplitude when downregulating than when upregulating the primary chain. These results indicated that the composition of a receptor's signalling core could have strong effects on the variability of cellular responses to extra-cellular ligands. The γ_c family's arrangement tends to allow for large variability, but leads to lower amplitudes (see Chapter 3 and Ref. Cotari *et al.* (2013b)). Heterodimeric receptors with intrinsic kinase ability (RTK), such as the receptor for insulin, IGF-1, or homodimeric receptors with downstream kinases (with IEK), such as the IL-6 receptor, tend to have intermediate but consistent amplitude. Homodimeric receptors with associated kinase activity (RTK), such as EGFR, are highly variable in EC_{50} and their amplitude does not saturate (Shi *et al.*, 2016; Shtiegman *et al.*, 2007).

We also computed the EC_{50} for each of the receptor configurations. We observed that the EC_{50} of the monomeric models are the same, the EC_{50} of heterodimeric models are also equal (as expected from the theoretical analysis), and the EC_{50} of the homodimeric models are similar (Fig. 4.9). This led us to conjecture that the variability of the EC_{50} with the abundance of the primary chain depends on the manner the trans-membrane receptor is built and is rather independent of its intra-cellular activity. Further experimental work will be required to identify a molecular mechanism to test such conjecture. We also observe that the EC_{50} for all receptor configurations tends to a constant for large values of γ abundance. This value is the dissociation constant of the ligand to the receptor $\frac{1}{K_3}$.

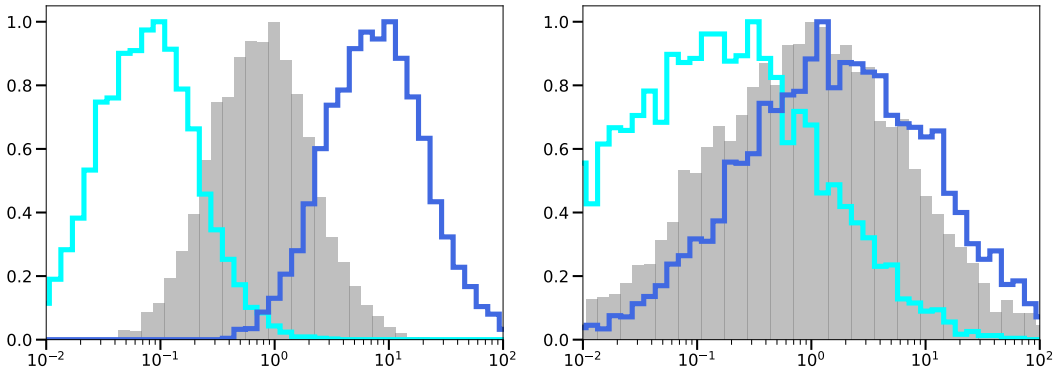
Three interesting remarks

From the analytic computations and numerical simulations of this section, one can make three remarks.



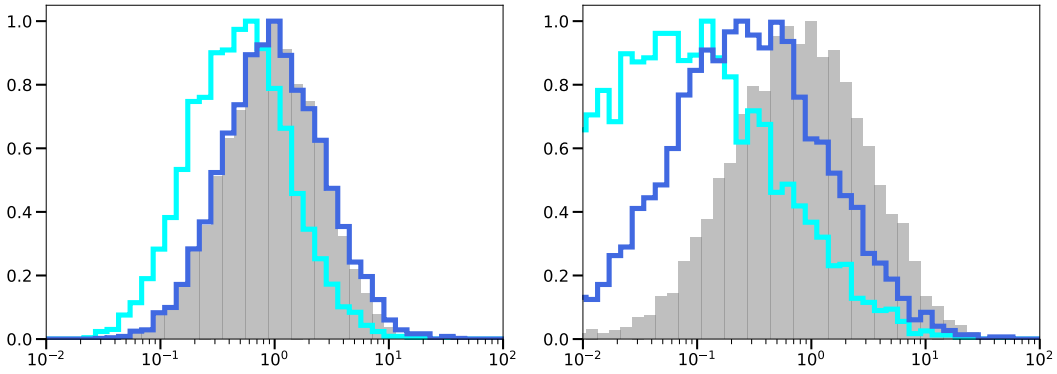
(a) Monomeric RTK

(b) Monomeric IEK



(c) Homodimeric RTK

(d) Homodimeric IEK



(e) Heterodimeric RTK

(f) Heterodimeric IEK

Figure 4.10: Normalised and centred amplitude distributions for each receptor configuration when the primary chain, γ, is upregulated (dark blue) or downregulated (cyan) by 10-fold (compared to the grey distribution (wild type)).

Remark 71 (Same amplitude form for several models)**.** The amplitudes of monomeric and heterodimeric receptor models with intrinsic kinase (RTK models) are the lowest total number of all the receptor constituents. The amplitudes of the monomeric, heterodimeric (Section 3.2) and trimeric receptor models with IEK, as well as the amplitude of the heterodimeric receptor model with IEK and additional subunit (Section 3.4), are of the form $\frac{K_1 z}{1+K_1 z}N$ where

$$z = \frac{-1 + K_1(N_z - N_x) + \sqrt{4K_1 N_z + (1 + K_1(N_x - N_z))^2}}{2K_1}$$

and N is the total number of the limiting component (trans-membrane chain that has the smallest total abundance).

Remark 72 (Same EC_{50} for models with extrinsic or intrinsic kinases)**.** The EC_{50} of the two monomeric models are the same. Similarly, the EC_{50} of the heterodimeric model RTK and the heterodimeric model with IEK (IL-7R model described in Section 3.2) are the same.

Remark 73 (Limitation of the method)**.** The method described in Section 4.1 can effectively compute a closed-form expression of the amplitude and, in some cases with a little extra work, the EC_{50}. However, the model has to be simple enough so that the Gröbner bases computed are tractable (in terms of length) and solvable (the degree of the polynomials of the basis is low).

In response to Remark 71, in Section 4.3, I introduce a family of models with n trans-membrane chains which is a generalisation of the models with IEK, and derive the general amplitude expression for the models of this family without making use of Gröbner bases. The analysis of this family also finally proves that the limit of the signalling function of the models in Chapter 3 and the trimeric receptor model of Section 4.2.6 is indeed the amplitude. This study is completed in Section 4.4, in which I also explore the consequences of Remark 72. Finally, in Section 4.5 and as a consequence of Remark 73, I investigate whether we can still gain insights on the amplitude or EC_{50} of complex models (for which the method of Section 4.1 fails to provide analytic expressions) by computing these quantities for simpler sub-models.

4.3 General results on a family of receptor-ligand models: SRLK models

In spite of the general applicability of the method outlined in Section 4.1, we still have to make the assumption that the computed limit of the signalling function coincides with its amplitude. In this section we show that this is indeed the case for a wide class of receptor-ligand systems. An analytic closed-form expression for the amplitude follows with little extra work. The EC_{50} can then be studied making use of key steps 6 and 7 in Section 4.1. We start by giving an abstract generalisation of the examples from Sections 4.2.4, 3.2 and 4.2.6.

Definition 74 (SRLK model)**.** We call a *sequential receptor-ligand model with extrinsic kinase (SRLK)* a receptor-ligand model with the following properties:

- The receptor is composed of n different trans-membrane chains, $X_1, \ldots, X_n$, which bind sequentially,

$$X_1 : \ldots : X_{i-1} + X_i \rightleftharpoons X_1 : \ldots : X_i \quad \text{for all } i \in \{2, \ldots, n\}.$$

- X_1 can bind reversibly to an intra-cellular extrinsic kinase Z.
- The signalling receptor is given by $Z : X_1 : \ldots : X_n$ and the "dummy" receptor by $X_1 : \ldots : X_n$.
- The extra-cellular ligand, L, binds reversibly to the signalling (or "dummy") receptor, forming the signalling (or "dummy") complex $Z : X_1 : \ldots : X_n : L$ (or $X_1 : \ldots : X_n : L$).

The biochemical reaction network for a general SRLK model is given by

$$\begin{array}{lcll}
Z & + & X_1 & \rightleftharpoons \quad Z:X_1 & K_0, \\
X_2 & + & Z:X_1 & \rightleftharpoons \quad Z:X_1:X_2 & K_1, \\
X_2 & + & X_1 & \rightleftharpoons \quad X_1:X_2 & K_1', \\
\vdots & & \vdots & \quad\vdots & \vdots \\
X_{i+1} & + & Z:X_1:\ldots:X_i & \rightleftharpoons \quad Z:X_1:\ldots:X_{i+1} & K_i, \\
X_{i+1} & + & X_1:\ldots:X_i & \rightleftharpoons \quad X_1:\ldots:X_{i+1} & K_i', \\
\vdots & & \vdots & \quad\vdots & \vdots \\
X_n & + & Z:X_1:\ldots:X_{n-1} & \rightleftharpoons \quad Z:X_1:\ldots:X_n & K_{n-1}, \\
X_n & + & X_1:\ldots:X_{n-1} & \rightleftharpoons \quad X_1:\ldots:X_n & K_{n-1}', \\
L & + & Z:X_1:\ldots:X_n & \rightleftharpoons \quad Z:X_1:\ldots:X_n:L & K_n, \\
L & + & X_1:\ldots:X_n & \rightleftharpoons \quad X_1:\ldots:X_n:L & K_n',
\end{array} \tag{4.84}$$

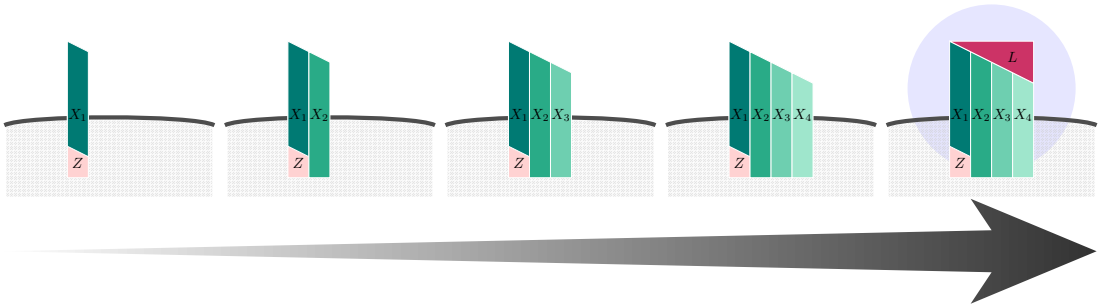
where the K_i (or K_i') are the affinity constants related to the formation of the signalling (or "dummy") complex. Figure 4.11 illustrates the formation of the signalling and "dummy" complexes in an SRLK model with $n = 4$ trans-membrane chains. We assume the system at steady state and that the ligand is in excess. In what follows we refer to these two assumptions as the *experimental hypotheses*.

We write z (or x_i) for the steady state concentration of unbound chain Z (or X_i). We also use L to denote the ligand concentration. Finally, N_z (or N_i) denotes the total copy number per cell of the species Z (or X_i). An SRLK model satisfying the experimental hypotheses is then described by the following polynomial system:

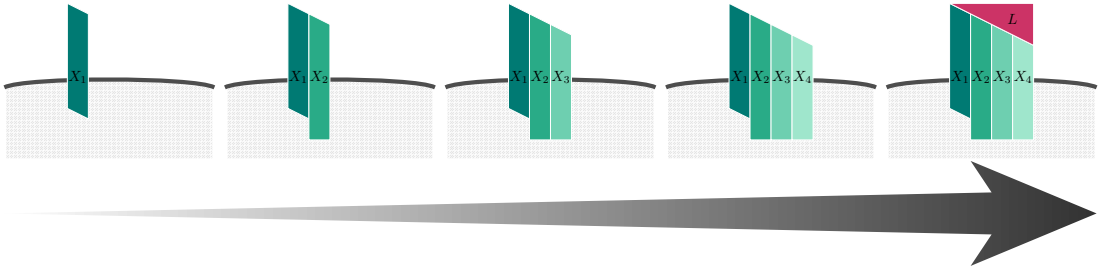
$$\begin{aligned}
N_z &= z + K_0 z\left(x_1 + K_1 x_1 x_2 + \ldots + K_1 \ldots K_{n-1} x_1 \ldots x_n + \ldots K_n x_1 \ldots x_n L\right) \\
&= z + K_0 z\left[x_1 + \sum_{j=2}^{n}\left(\prod_{l=1}^{j-1} K_l x_l x_j\right) + L\prod_{j=1}^{n} K_j x_j\right], \qquad (4.85a) \\
N_1 &= x_1 + K_1' x_1 x_2 + \ldots + K_1' \ldots K_{n-1}' x_1 \ldots x_n + K_1' \ldots K_n' L x_1 \ldots x_n \qquad (4.85b) \\
&\quad + K_0 z\left(x_1 + K_1 x_1 x_2 + \ldots + K_1 \ldots K_{n-1} x_1 \ldots x_n + K_1 \ldots K_n x_1 \ldots x_n L\right),
\end{aligned}$$

for $i = 2, \ldots, n-1$:

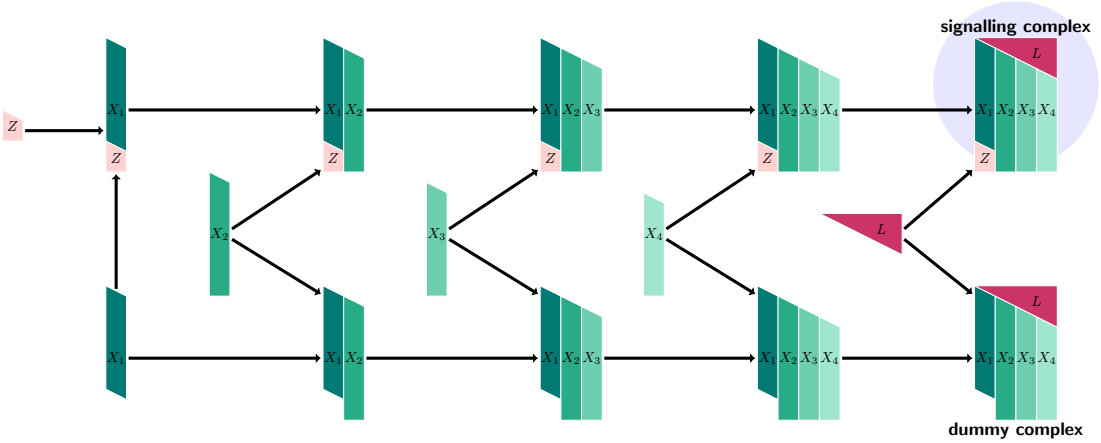
$$\begin{aligned}
N_i &= x_i \qquad (4.85c) \\
&\quad + K_1' \ldots K_{i-1}' x_1 \ldots x_i + \ldots + K_1' \ldots K_{n-1}' x_1 \ldots x_n + K_1' \ldots K_n' x_1 \ldots x_n L
\end{aligned}$$



(a) Sequential formation of the signalling complex



(b) Sequential formation of the "dummy" complex



(c) Sequential chemical reaction scheme of a SRLK model

Figure 4.11: SRLK model with $n = 4$ trans-membrane chains: (a) (resp. (b)) Sequential formation of the signalling (resp. dummy) complex. (c) Summarising scheme of the sequential formation of the signalling (resp. "dummy") complex: the chain X_1 binds first to the intra-cellular kinase Z (only for the signalling complex). Next, the chain X_2 binds to the complex $Z : X_1$ (resp. X_1) and X_3 binds to $Z : X_1 : X_2$ (resp. $X_1 : X_2$). Then, X_4 binds to $Z : X_1 : X_2 : X_3$ (resp. $X_1 : X_2 : X_3$). Finally, the ligand L finally binds to the signalling receptor $Z : X_1 : X_2 : X_3 : X_4$ (resp. "dummy" receptor $X_1 : X_2 : X_3 : X_4$), thus forming the signalling (resp. "dummy") complex.

$$
\begin{aligned}
&+ K_0 z(K_1 \dots K_{i-1} x_1 \dots x_i + \dots + K_1 \dots K_{n-1} x_1 \dots x_n \\
&+ K_1 \dots K_n x_1 \dots x_n L) \\
&= x_i + \sum_{j=i}^{n} \left(\prod_{l=1}^{j-1} K'_l x_l x_j + K_0 z \prod_{l=1}^{j-1} K_l x_l x_j \right) + L \prod_{j=1}^{n} K'_j x_j + K_0 z L \prod_{j=1}^{n} K_j x_j, \\
N_n &= x_n + K_0 \dots K_{n-1} z x_1 \dots x_n + K'_1 \dots K'_{n-1} x_1 \dots x_n \\
&+ K'_1 \dots K'_n L x_1 \dots x_n + K_0 \dots K_n z L x_1 \dots x_n \, .
\end{aligned} \tag{4.85d}
$$

We note that many results in this section can be further simplified under the additional hypothesis of no allostery.

Definition 75. There is *no allostery* in an SRLK model if $K_i = K'_i$ for all $i = 1, \dots, n$.

Finally, we formally define the signalling and dummy functions for this class of models.

Definition 76. For an SRLK model under the experimental hypotheses the signalling function, $\sigma(L)$, is the number of signalling complexes formed as a function of the ligand concentration, L, and can be written as follows

$$\sigma(L) = K_0 z L \prod_{i=1}^{n} K_i x_i.$$

Similarly, the dummy function, $\delta(L)$, is the number of "dummy" complexes formed as a function of the ligand concentration, L, and can be written as follows

$$\delta(L) = L \prod_{i=1}^{n} K'_i x_i.$$

Note that the IL-7R model of Section 3.2 is one example of an SRLK model and the definition of signalling function given in Section 2.4 is equivalent. We now introduce the notion of a limiting component.

Definition 77. The species, X_j, which has the smallest total copy number of molecules

$$0 < N_j < N_i, \ \forall i \neq j,$$

is the limiting component of the system. If there are multiple limiting components, $X_{j_1}, \ldots, X_{j_r}$, then

$$0 < N_{j_1} = \ldots = N_{j_r} < N_i, \ \forall i \notin \{j_1, \ldots, j_r\}.$$

If the signalling function attains its maximum for large values of the ligand concentration, then, since by definition $\sigma(0) = 0$, the amplitude of such model is given by

$$A \equiv \lim_{L \to +\infty} \sigma(L).$$

In this section we present some general results for $\lim_{L \to +\infty} \delta(L)$ and $\lim_{L \to +\infty} \sigma(L)$ applicable to SRLK models.

4.3.1 Asymptotic study of the steady states

While it is difficult to find closed-form expressions of the steady states for general receptor-ligand systems, in what follows we show that considerable progress can be made for the specific case of SRLK models. In this section we describe the behaviour of the concentrations, x_i, in the limit $L \to +\infty$. First, we recall the definition and a property of algebraic functions.

Definition 78. A univariate function $y = f(x)$ is said to be algebraic if it satisfies the polynomial equation:

$$y^m + R_{m-1}(x) y^{m-1} + \cdots + R_0(x) = 0, \qquad (\dagger)$$

where the $R_i(x)$ are rational functions of x, *i.e.,* are of the form $\frac{p(x)}{q(x)}$, where p and q are polynomial functions and $q(x) \neq 0$ for all $x \in \mathbb{R}$.

Remark 79. Note that the polynomial (†) has m solutions. These solutions are called the branches of an algebraic function and one often specifies a particular branch.

Since we are interested in the limit behaviour, the following lemma proves useful.

Lemma 80. Any bounded, continuous solution of (†) defined on $\mathbb{R}$ has a finite limit at $+\infty$ (and $-\infty$).

Proof. Multiply (†) by the common denominator of the R_i and let $x = \epsilon^{-1}$ to obtain

$$\underbrace{\left[\prod_{i=0}^{m} \tilde{q}_i(\epsilon)\right]}_{\tilde{r}_m(\epsilon)} \tilde{y}^m + \tilde{r}_{m-1}(\epsilon)\tilde{y}^{m-1} + \cdots + \tilde{r}_0(\epsilon) = 0,$$

with $\tilde{y} = \tilde{f}(\epsilon)$. We have now recast the original problem into the form of equation (2.13). By Theorem 49 we know that an expansion for the roots exists and we note that the points of $f(x)$ as $x \to +\infty$ correspond to the points of $\tilde{f}(\epsilon)$ as $\epsilon = 0$. Note that, since all real $f(x)$ are bounded, so are the real $\tilde{f}(\epsilon)$. Therefore all real $\tilde{f}(0)$ are finite and equal to the limits $\lim_{x\to+\infty} f(x)$. A unique limit is chosen by specifying a branch of $f(x)$. The proof for $x \to -\infty$ follows *mutatis mutandis*. □

With this background in place, we can now proceed to study SRLK models in detail. We start by showing that in steady state the signalling and the dummy functions have a positive limit when L tends to $+\infty$.

Lemma 81. The signalling and the dummy functions of an SRLK model satisfying the experimental hypotheses admit a finite limit when $L \to +\infty$ and this limit is positive.

Proof. The function σ (or δ) are algebraic functions bounded on $\mathbb{R}$ between 0 and $\min(N_z, N_1, \ldots, N_n)$ (or $\min(N_1, \ldots, N_n)$) so they admit a finite limit when $L \to +\infty$. Let us denote this limit by c_σ (or c_δ). We know that c_σ and c_δ are non-negative because σ and δ are, by definition, products of non-negative functions.

Consider $c_\delta = 0$. Then since $\sigma(L) = K_0 z \prod_{i=1}^{n} \frac{K_i}{K_i'}\delta(L)$, we have $c_\sigma = 0$ (we note that z being also an algebraic function, z also admits a finite limit when $L \to +\infty$). Since δ converges to 0, we need

$$\prod_{i=1}^{n} x_i \underset{L\to+\infty}{\sim} \frac{C_n}{L^p}, \tag{4.86}$$

with C_n a positive constant and $p > 1$. We recall and rewrite polynomial (4.85d):

$$N_n = x_n + K_0 z \prod_{i=1}^{n-1} K_i \prod_{i=1}^{n} x_i + \prod_{i=1}^{n-1} K_i' \prod_{i=1}^{n} x_i + \delta(L) + \sigma(L). \tag{4.87}$$

Assuming (4.86) when $L \to +\infty$ in (4.87), we obtain:

$$\lim_{L \to +\infty} x_n = N_n,$$

and so we must have

$$\prod_{i=1}^{n-1} x_i \underset{L \to +\infty}{\sim} \frac{C_{n-1}}{L^p},$$

with $p > 1$ and C_{n-1} a positive constant. Passing to the limit in polynomial (4.85c) for $i = n - 1$, we obtain

$$\lim_{L \to +\infty} x_{n-1} = N_{n-1}.$$

We repeat the process for every conservation equation (4.85c) of the species X_i and we obtain

$$\forall i = 1, \ldots, n, \ \lim_{L \to +\infty} x_i = N_i,$$

which is a contradiction with equation (4.86). So $c_\delta > 0$.

Now, consider $c_\sigma = 0$. Then since $\sigma(L) = K_0 z \prod_{i=1}^{n} \frac{K_i}{K_i'} \delta(L)$, z has to tend to 0. However, when passing to the limit $L \to +\infty$ in equation (4.85a), we obtain

$$N_z = \lim_{L \to +\infty} (z + K_0 z x_1 + \ldots + \sigma(L)) = 0,$$

which is a contradiction.

Conclusion: $c_\sigma > 0$ and $c_\delta > 0$. □

An equivalent result holds for the steady state concentration of the kinase.

Lemma 82. In an SRLK model under the experimental hypotheses, the concentration of the extrinsic intra-cellular kinase Z admits a positive finite limit, $c_z > 0$, when $L \to +\infty$.

Proof. The concentration of kinase z being an algebraic function bounded on $\mathbb{R}$ between 0 and N_z, it admits a finite limit c_z when $L \to +\infty$. We know that $c_z \geq 0$ because z is a concentration. We now prove that $c_z > 0$. Since δ converges to a positive constant when $L \to +\infty$, we must have

$$\prod_{i=1}^{n} x_i \underset{L\to+\infty}{\sim} \frac{c_d}{L},$$

where c_d is a positive constant. Since σ also admits a finite limit when $L \to +\infty$, it means that

$$z \prod_{i=1}^{n} x_i \underset{L\to+\infty}{\sim} \frac{c_s}{L},$$

where c_s is a positive constant. So z has to satisfy

$$z \underset{L\to+\infty}{\sim} c_z,$$

where $c_z = \dfrac{c_s}{c_d}$ is a positive constant. □

In the particular case of no allostery, we can write an explicit expression of the limit of z, c_z.

Lemma 83. Consider an SRLK model which satisfies the experimental hypotheses. If we assume no allostery, then the steady state value of the extrinsic intra-cellular kinase, z, is given by

$$z = \frac{-1 + K_0(N_z - N_1) + \sqrt{\Delta_z}}{2K_0}, \tag{4.88}$$

where

$$\Delta_z = (1 + K_0(N_1 - N_z))^2 + 4K_0N_z.$$

Proof. We assumed no allostery so $K_i = K_i'$ for all $i = 1, \ldots, n$. Equation (4.85a)

gives:

$$N_z - z = K_0 z \left(x_1 + \sum_{j=2}^{n} \left(\prod_{l=1}^{j-1} K_l x_l x_j \right) + L \prod_{j=1}^{n} K_j x_j \right).$$

By substituting this equality in equation (4.85b), we obtain:

$$N_1 = N_z - z + \frac{N_z - z}{K_0 z},$$

so z is a positive root of the polynomial

$$-N_z + z(1 + K_0(N_1 - N_z)) + K_0 z^2,$$

with L-independent coefficients. The two possibilities are:

$$z_1 = \frac{-1 + K_0(N_z - N_1) + \sqrt{4K_0 N_z + (1 + K_0(N_1 - N_z))^2}}{2K_0},$$
$$z_2 = \frac{-1 + K_0(N_z - N_1) - \sqrt{4K_0 N_z + (1 + K_0(N_1 - N_z))^2}}{2K_0}.$$

The expression z_1 is always positive, while z_2 is always negative. Hence z_1 is the steady state kinase concentration, z. □

By Lemma 83, z is independent of L (thus, $c_z = z$) and only depends on K_1, N_1 and N_z. Note that this result is equivalent to the one obtained in Section 3.2 for the first IL-7R model. Finally, we study the behaviour of the concentration x_i in the limit $L \to +\infty$. We first give bounds to the asymptotic dependency of x_i on L.

Lemma 84. Let us consider an SRLK model which satisfies the experimental hypotheses. Then no concentration x_i behaves proportionally to L^q, $q > 0$ or $\frac{1}{L^p}$, $p > 1$ when $L \to +\infty$.

Proof. Lemma 82 affirms that z tends to a positive constant when $L \to +\infty$. In order for σ or δ to converge to a positive constant as stated in lemma 81, we need

$$\prod_{i=1}^{n} x_i \underset{L \to +\infty}{\sim} \frac{c}{L}, \tag{4.89}$$

where c is a positive constant. Since the concentrations $x_1, \ldots, x_n$ are bounded functions (between 0 and their respective N_i), it is impossible to have for any $i = 1, \ldots, n$, $x_i \underset{L\to+\infty}{\sim} c_i L^q$ with c_i constant and $q > 0$. From equation (4.89) it follows that it is impossible to have any $x_i \underset{L\to+\infty}{\sim} \frac{c_i}{L^p}$ for $p > 1$. □

We can now state the main theorem of this section.

Theorem 85. We consider an SRLK model which satisfies the experimental hypotheses. If there exists a unique limiting component X_{i_0}, then

$$x_{i_0} \underset{L\to+\infty}{\sim} \frac{c_{i_0}}{L},$$

and for all $i = 1, \ldots, n, i \neq i_0$,

$$x_i \underset{L\to+\infty}{\sim} c_i,$$

where c_{i_0} and c_i are positive constants.

Proof. Since the concentrations x_i are algebraic functions (with coefficients in $\mathbb{R}$) bounded on $\mathbb{R}$, they admit a non-negative limit when $L \to +\infty$.

We know that we need

$$\prod_{i=1}^{n} x_i \underset{L\to+\infty}{\sim} \frac{c}{L}, \tag{4.90}$$

with c a positive constant, so that σ and δ converge when $L \to +\infty$. Lemma 82 shows that z tends to a positive constant when $L \to +\infty$. Thus, it follows from equation (4.90) and Lemma 84 that at least one of the x_i must tend to 0. We will prove that the only concentration that can tend to 0 is x_{i_0}, and so $x_{i_0} \underset{L\to+\infty}{\sim} \frac{c_{i_0}}{L}$, with c_{i_0} a constant.

1) There exists at least one chain X_j whose concentration tends to 0. The conservation equation of X_j described in equation (4.85c) is:

$$N_j = x_j + \prod_{i=1}^{j} K_i' x_i + K_0 z \prod_{i=1}^{j} K_i x_i + \prod_{i=1}^{j+1} K_i' x_i + K_0 z \prod_{i=1}^{j+1} K_i x_i + \ldots + \delta(L) + \sigma(L).$$

When $L \to +\infty$, we obtain

$$N_j = \lim_{L\to+\infty} \delta(L) + \lim_{L\to+\infty} \sigma(L).$$

We cannot form more "dummy" or signalling complexes than the number of molecules available. Since X_{i_0} is the limiting component, we have

$$\delta(L) + \sigma(L) \leq N_{i_0}, \quad \forall L.$$

This yields in the limit $L \to +\infty$, $N_j \leq N_{i_0}$. By hypothesis this implies that $j = i_0$ and so X_j is our limiting component X_{i_0}.

2) Reciprocally, if x_{i_0} tends to a positive constant when $L \to +\infty$, then there exists at least one x_j, $j \neq i_0$ such that $x_j \to 0$ when $L \to +\infty$. The limit when $L \to +\infty$ of equation (4.85c) (set to $i = j$) gives

$$\lim_{L \to +\infty} [\delta(L) + \sigma(L)] = N_j.$$

However, since we also have $\delta + \sigma \leq N_{i_0}$, we obtain when taking the limit, $N_j \leq N_{i_0}$, which is a contradiction with the fact that X_{i_0} is the only limiting component.

Conclusion: X_{i_0} is limiting if and only if its concentration tends to 0, and we have

$$x_{i_0} \underset{L \to +\infty}{\sim} \frac{c_{i_0}}{L},$$

and for $i \neq i_0$,

$$x_i \underset{L \to +\infty}{\sim} c_i,$$

where c_{i_0} and c_i are positive constants. □

Corollary 86. If an SRLK model, which satisfies the experimental hypotheses, has multiple limiting components, $X_{i_1}, \ldots, X_{i_r}$, $i_1 < \ldots < i_r$, then

$$x_{i_1} \underset{L \to +\infty}{\sim} \frac{c_{i_1}}{L^{p_1}}, \ldots, x_{i_r} \underset{L \to +\infty}{\sim} \frac{c_{i_r}}{L^{p_r}},$$

where $c_{i_1}, \ldots c_{i_r}$ are positive constants and $p_1 = \ldots = p_r = \dfrac{1}{r}$. The concentrations of the non-limiting components, x_i, (for $i \notin \{i_1, \ldots, i_r\}$) tend to positive constants, $c_i > 0$.

Proof. If X_{i_1} and X_{i_2} are limiting components, they are the only ones whose concentrations, x_{i_1} and x_{i_2}, tend to 0 when $L \to +\infty$. From equation (4.90) we

can write

$$x_{i_1} \underset{L\to+\infty}{\sim} \frac{c_{i_1}}{L^{p_1}},$$
$$x_{i_2} \underset{L\to+\infty}{\sim} \frac{c_{i_2}}{L^{p_2}},$$

with c_{i_1} and c_{i_2} constants and $p_1, p_2 > 0$, such that $p_1 + p_2 = 1$.

From system (4.85), we have:

$$N_{i_1} = x_{i_1} + \sum_{j=i_1}^{n}\left(\prod_{l=1}^{j-1} K'_l x_l x_j + K_0 z \prod_{l=1}^{j-1} K_l x_l x_j\right) + L\prod_{j=1}^{n} K'_j x_j + K_0 z L \prod_{j=1}^{n} K_j x_j,$$
$$N_{i_2} = x_{i_2} + \sum_{j=i_2}^{n}\left(\prod_{l=1}^{j-1} K'_l x_l x_j + K_0 z \prod_{l=1}^{j-1} K_l x_l x_j\right) + L\prod_{j=1}^{n} K'_j x_j + K_0 z L \prod_{j=1}^{n} K_j x_j.$$

Since X_{i_1} and X_{i_2} are limiting components, we have $N_{i_1} = N_{i_2}$ and, if $i_1 < i_2$, we obtain

$$N_{i_1} = N_{i_2} \iff x_{i_1}\left(1 + \sum_{j=i_1}^{i_2-1}\left(\prod_{l=1}^{j-1} K'_l \prod_{l=1, l\neq i_1}^{j} x_l + K_0 z \prod_{l=1}^{j-1} K_l \prod_{l=1, l\neq i_1}^{j} x_l\right)\right) = x_{i_2}. \tag{4.91}$$

Since all the x_i, with $i \neq i_1$, $i \neq i_2$, tend to a positive constant when $L \to +\infty$, we have

$$1 + \sum_{j=i_1}^{i_2-1}\left(\prod_{l=1}^{j-1} K'_l \prod_{l=1, l\neq i_1}^{j} x_l + K_0 z \prod_{l=1}^{j-1} K_l \prod_{l=1, l\neq i_1}^{j} x_l\right) \underset{L\to+\infty}{\sim} C,$$

where C is a positive constant. Thus, we obtain the behaviour of the left side of equation (4.91)

$$x_{i_1}\left(1 + \sum_{j=i_1}^{i_2-1}\left(\prod_{l=1}^{j-1} K'_l \prod_{l=1, l\neq i_1}^{j} x_l + K_0 z \prod_{l=1}^{j-1} K_l \prod_{l=1, l\neq i_1}^{j} x_l\right)\right) \underset{L\to+\infty}{\sim} \frac{C c_{i_1}}{L^{p_1}}.$$

Since the right side is given by

$$x_{i_2} \underset{L\to+\infty}{\sim} \frac{c_{i_2}}{L^{p_2}},$$

it results in $p_1 = p_2 = \dfrac{1}{2}$.

If there are r limiting components $x_{i_1}, \ldots, x_{i_r}$, $i_1 < \ldots < i_r$, then we have

$$\begin{aligned} x_{i_1} &\underset{L\to+\infty}{\sim} \frac{c_{i_1}}{L^{p_1}}, \\ &\vdots \qquad\qquad \vdots \\ x_{i_r} &\underset{L\to+\infty}{\sim} \frac{c_{i_r}}{L^{p_r}}, \end{aligned}$$

with $c_{i_1}, \ldots, c_{i_r}$ positive constants, and $p_1, \ldots, p_r > 0$ such that $p_1 + \ldots + p_r = 1$. We proceed in the same way as for the case of two limiting components and we obtain $p_1 = \ldots = p_r = \dfrac{1}{r}$. □

4.3.2 Asymptotic study of the signalling and dummy functions

The previous section presented numerous small results which give insight into the steady state behaviour of SRLK receptor-ligand systems. We are now in a position to combine these results to state and prove the main theorem, which gives closed-form formulæ for the limits of the signalling and dummy functions.

Theorem 87. Consider an SRLK model which satisfies the experimental hypotheses. Let us write $X_{i_1}, \ldots, X_{i_r}$ as the limiting components and denote N_{i_0} as the total amount of any limiting component, i.e., $N_{i_0} \equiv N_{i_1} = \ldots = N_{i_r}$. The limit of the signalling function is given by

$$\lim_{L\to+\infty} \sigma(L) = \frac{\prod_{i=1}^n K_i K_0 c_z}{\prod_{i=1}^n K'_i + \prod_{i=1}^n K_i K_0 c_z} N_{i_0},$$

and the limit of the dummy function is

$$\lim_{L\to+\infty} \delta(L) = \frac{\prod_{i=1}^n K'_i}{\prod_{i=1}^n K'_i + \prod_{i=1}^n K_i K_0 c_z} N_{i_0},$$

where

$$c_z = \lim_{L\to+\infty} z.$$

Proof. By definition of σ and δ we have:

$$\delta(L) + \sigma(L) = L \prod_{i=1}^{n} x_i \left(\prod_{i=1}^{n} K_i' + K_0 z \prod_{i=1}^{n} K_i \right),$$

which implies that

$$\lim_{L \to +\infty} [\delta(L) + \sigma(L)] = \lim_{L \to +\infty} \left(L \prod_{i=1}^{n} x_i \left(\prod_{i=1}^{n} K_i' + K_0 z \prod_{i=1}^{n} K_i \right) \right).$$

Using the limit properties and since everything converges, we obtain:

$$\lim_{L \to +\infty} [\delta(L) + \sigma(L])] = \lim_{L \to +\infty} \left(L \prod_{i=1}^{n} x_i \right) \left(\prod_{i=1}^{n} K_i' + K_0 \lim_{L \to +\infty} (z) \prod_{i=1}^{n} K_i \right).$$

However, Theorem 85 states that x_{i_0} tends to 0 when $L \to +\infty$. Thus, equation (4.85c) when $i = i_0$ at $L \to +\infty$ gives

$$N_{i_0} = \lim_{L \to +\infty} [\delta(L) + \sigma(L)].$$

Consequently since $z \to c_z > 0$ from lemma 82, we obtain:

$$\lim_{L \to +\infty} (L \prod_{i=1}^{n} x_i) = \frac{N_{i_0}}{\prod_{i=1}^{n} K_i' + K_0 c_z \prod_{i=1}^{n} K_i}.$$

We substitute this limit into the expression of σ and δ and obtain the desired expressions. □

Under the assumption of no allostery, these expressions can be further simplified.

Corollary 88. Consider the system of Theorem 87 and assume there is no allostery. Denote the limiting components by $X_{i_1}, \ldots, X_{i_r}$ and denote N_{i_0} as the total amount of any limiting component, i.e., $N_{i_0} \equiv N_{i_1} = \ldots = N_{i_r}$. The limit of the signalling function is

$$\lim_{L \to +\infty} \sigma(L) = \frac{K_0 z}{1 + K_0 z} N_{i_0},$$

and the limit of the dummy function is

$$\lim_{L\to+\infty} \delta(L) = \frac{N_{i_0}}{1+K_0 z},$$

with z given by equation (4.88) in Lemma 83.

Proof. Since there is no allostery, we have $K_i = K_i'$ for all i. Lemma 83 states that z is independent of L, thus $c_z = z$. Applying these statements in the expressions of the previous theorem, we obtain the expressions of this corollary. □

From the previous expressions we observe that the limit of the signalling and dummy functions are equal to the total copy number of the limiting component, N_{i_0}, multiplied by a term which is bounded between 0 and 1. This term only depends on the affinity constant K_0 and the steady state concentrations of the kinase, Z and X_1. In order to relate the above limits back to biologically meaningful quantities, all there is left to show is that the explicit expression of the limit of σ is in fact the amplitude of the system. Since $\sigma(0) = 0$, let us first note that the amplitude is equal to the maximum of σ. Under the no allostery assumption, we can show mathematically that this maximum is the limit of σ when $L \to +\infty$. To this end, the following lemma is needed.

Lemma 89. Consider an SRLK model under the experimental hypotheses. If there is no allostery, then we have

$$\sup \sigma(L) = \lim_{L\to+\infty} \sigma(L).$$

Proof. Since X_{i_0} is the limiting component, we know from Theorem 85 that its concentration tends to 0 when $L \to +\infty$. We have

$$\delta + \sigma \leq N_{i_0} = \lim_{L\to+\infty} (\delta + \sigma). \tag{4.92}$$

In the no allostery case, z is independent of L and we have $\sigma = K_0 z \delta$. Thus, equation (4.92) gives

$$(1 + K_0 z)\delta \leq (1 + K_0 z) \lim_{L\to+\infty} \delta.$$

Hence, we can conclude

$$\lim_{L\to+\infty} \delta = \sup \delta,$$

and

$$\lim_{L\to+\infty} \sigma = \sup \sigma.$$

□

The supremum here is attained and is a maximum. Thus, the amplitude for a SRLK receptor-ligand system when there is no allostery is the limit of σ described in Corollary 88. This result is the generalisation of the examples discussed in Sections 4.2.4, 3.2 and 4.2.6. Indeed, the amplitudes of the monomeric model with IEK (Section 4.2.4), the IL-7R model of Section 3.2 and the trimeric receptor model of Section 4.2.6 can be recovered by setting $N_{i_0} = N_x$, $N_{i_0} = \min(N_x, N_y)$ and $N_{i_0} = \min(N_x, N_y, N_w)$, respectively. We now have also rigorously shown that the limit of the signalling function is indeed the amplitude.

4.3.3 SRLK models with additional subunit receptor chains

As hinted in Section 3.4, the IL-7R model with the additional sub-unit receptor chain is part of a larger group of models which are an extension of the SRLK family. Therefore, our previous results can be extended to this type of models. Again, we start by giving an abstract definition of the extended SRLK family of models.

Definition 90 (Extended SRLK model)**.** An *extended SRLK model* is an SRLK model where we assume that each intermediate complex, $Z : X_1 : \ldots : X_i$ (or $X_1 : \ldots : X_i$), for $i = 1, \ldots, n$ can bind to an extra chain, Y_i, with an affinity constant K_{y_i} (or K'_{y_i}), to form a decoy complex $Z : X_1 : \ldots : X_i : Y_i$ (or $X_1 : \ldots : X_i : Y_i$). The addition of a sub-unit chain of the kind Y_i prevents the binding of ligand to the receptor, and thus, does not allow the formation of signalling or "dummy" complexes.

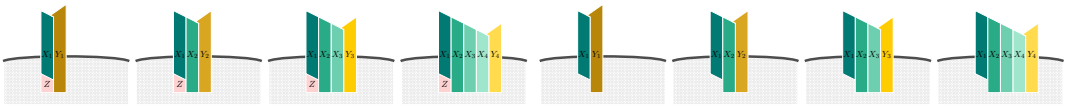
The chemical reaction network for an extended SRLK model is given by:

$$
\begin{array}{lcll}
Z & + X_1 & \rightleftharpoons Z : X_1 & K_0, \\
Y_1 & + X_1 & \rightleftharpoons X_1 : Y_1 & K'_{y_1}, \\
Y_1 & + Z : X_1 & \rightleftharpoons Z : X_1 : Y_1 & K_{y_1}, \\
X_2 & + Z : X_1 & \rightleftharpoons Z : X_1 : X_2 & K_1, \\
X_2 & + X_1 & \rightleftharpoons X_1 : X_2 & K'_1, \\
Y_2 & + X_1 : X_2 & \rightleftharpoons X_1 : X_2 : Y_2 & K'_{y_2}, \\
Y_2 & + Z : X_1 : X_2 & \rightleftharpoons Z : X_1 : X_2 : Y_2 & K_{y_2}, \\
\vdots & \vdots & \vdots & \vdots \\
X_{i+1} & + Z : X_1 : \ldots : X_i & \rightleftharpoons Z : X_1 : \ldots : X_{i+1} & K_i, \\
X_{i+1} & + X_1 : \ldots : X_i & \rightleftharpoons X_1 : \ldots : X_{i+1} & K'_i, \\
Y_i & + X_1 : \ldots : X_i & \rightleftharpoons X_1 : \ldots : X_i : Y_i & K'_{y_i}, \\
Y_i & + Z : X_1 : \ldots : X_i & \rightleftharpoons Z : X_1 : \ldots : X_i : Y_i & K_{y_i}, \\
\vdots & + \vdots & \vdots & \vdots \\
Y_n & + X_1 : \ldots : X_n & \rightleftharpoons X_1 : \ldots : X_n : Y_n & K'_{y_n}, \\
Y_n & + Z : X_1 : \ldots : X_n & \rightleftharpoons Z : X_1 : \ldots : X_n : Y_n & K_{y_n}, \\
L & + Z : X_1 : \ldots . X_n & \rightleftharpoons Z : X_1 : \ldots : X_n : L & K_n, \\
L & + X_1 : \ldots . X_n & \rightleftharpoons X_1 : \ldots : X_n : L & K'_n.
\end{array}
$$

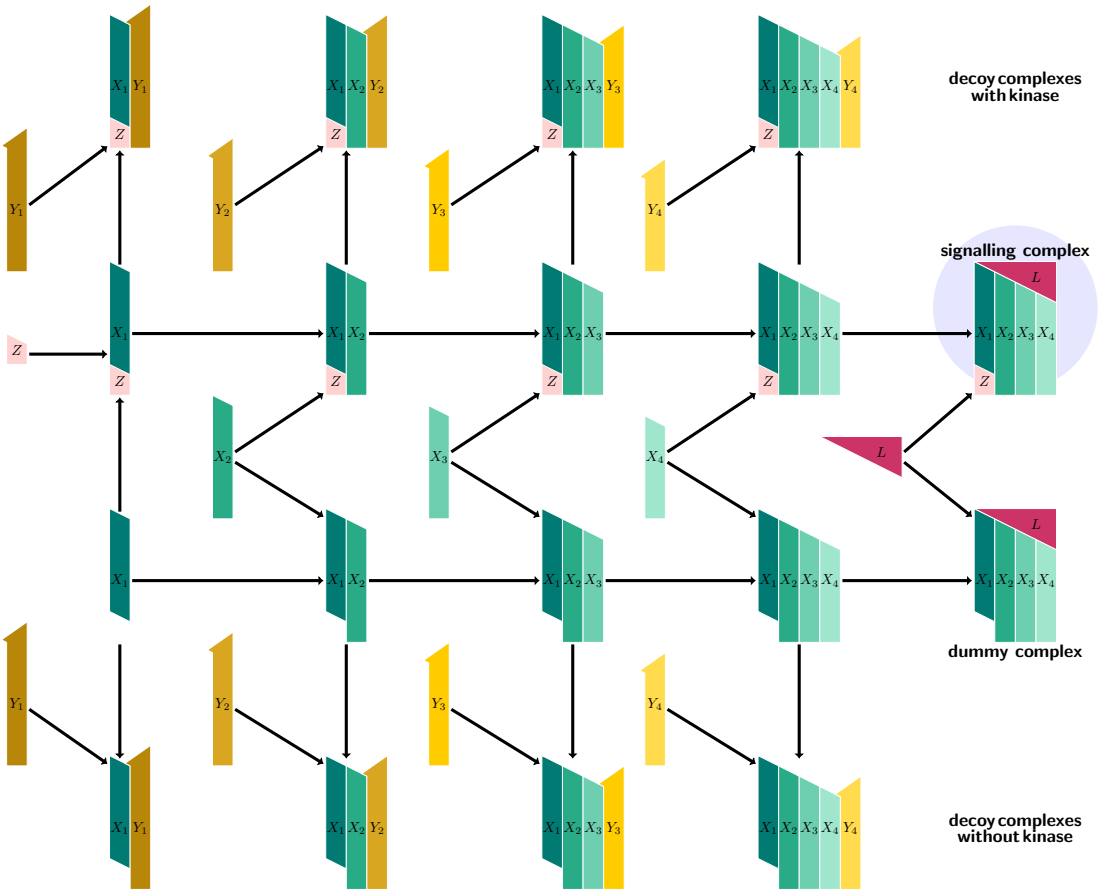
where K_i, K'_i, K_{y_i} and K'_{y_i} denote the affinity constants. Figure 4.12(a) and Figure 4.12(b) show the decoy complexes of an extended SRLK receptor-ligand system with $n = 4$ trans-membrane chains. The signalling and "dummy" complexes are built sequentially similarly to the classic SRLK model (see Figure 4.11 and Figure 4.12(c)).

We note that while we assume all the X_i to be different species, we allow that $Y_i = Y_j$ or $Y_i = \emptyset$, as long as for $i = 1, \ldots, n$, $Y_i \notin \{X_1, \ldots, X_n, Z, L\}$. We assume that the receptor-ligand system is in steady state and the ligand is in excess. We further assume that the concentration of the species Y_i (which we write y_i) are all bounded. We could consider the case when the Y_i are in excess, and thus, treat their concentration as a parameter of the model, or assume that the number of Y_i molecules is conserved. We refer to these assumptions as the *extended experimental hypotheses*.

The signalling and dummy functions of classic and extended SRLK receptor-ligand systems are equivalent (see Definition 76). The polynomial system describing an extended SRLK model under the extended experimental hypotheses is



(a) Decoy complexes with kinase (b) Decoy complexes without kinase



(c) Sequential chemical reaction scheme of an extended SRLK

Figure 4.12: Extended SRLK model with $n = 4$ trans-membrane chains: (a) An additional subunit Y_i can bind to each intermediate signalling complex $Z : X_1 : ... : X_i$, to form decoy complexes with kinase. (b) The subunit Y_i can also bind to the intermediate "dummy" complexes $X_1 : X_2 : ... : X_i$, forming decoy complexes without kinase. (c) Summarising scheme of the sequential formation of the signalling and "dummy" complexes. At each step, their formation can be interrupted by the binding of a subunit Y_i to the intermediate complex, forming a decoy complex.

given by

$$N_z = z + K_0 z \left(x_1(1 + K_{y_1} y_1) + \sum_{j=2}^{n} \left((1 + K_{y_j} y_j) \prod_{l=1}^{j-1} K_l x_l x_j \right) \right) + \sigma(L), \tag{4.93a}$$

$$N_1 = x_1(1 + K'_{y_1} y_1) + \sum_{j=2}^{n} \left((1 + K'_{y_j} y_j) \prod_{l=1}^{j-1} K'_l x_l x_j \right) + \delta(L) \tag{4.93b}$$
$$+ K_0 z \left(x_1(1 + K_{y_1} y_1) + \sum_{j=2}^{n} \left((1 + K_{y_j} y_j) \prod_{l=1}^{j-1} K_l x_l x_j \right) \right) + \sigma(L),$$

$$\vdots$$

$$N_i = x_i + \sum_{j=i}^{n} \left((1 + K'_{y_j} y_j) \prod_{l=1}^{j-1} K'_l x_l x_j + K_0 z (1 + K_{y_j} y_j) \prod_{l=1}^{j-1} K_l x_l x_j \right)$$
$$+ \delta(L) + \sigma(L), \quad \text{for } i = 2, \ldots, n-1, \tag{4.93c}$$

$$\vdots$$

$$N_n = x_n + K_0 z \, (1 + K_{y_n} y_n) \prod_{j=1}^{n-1} K_j x_j x_n + (1 + K'_{y_n} y_n) \prod_{j=1}^{n-1} K'_j x_j x_n + \delta(L) + \sigma(L). \tag{4.93d}$$

This system of polynomials is completed by the conservation equations of the species Y_i, for $i = 1, \ldots, n$, if we assume they are conserved.

We can extend the notion of no allostery to the extended models.

Definition 91. An extended SRLK model is said to be under the assumption of *no allostery* if for each $i = 1, \ldots, n$, $K_i = K'_i$ and $K_{y_i} = K'_{y_i}$.

With these expanded definitions, we can extend the results previously obtained for the SRLK receptor-ligand systems.

Theorem 92. The theorems and lemmas previously true for the SRLK models are true for the extended SRLK models under the same (extended) hypotheses.

Proof. The concentrations y_i are bounded ($0 \leq y_i \leq N_{y_i}$) algebraic function on $\mathbb{R}$, and therefore admit a limit when $L \to +\infty$. As the expressions of σ and δ are not modified, the addition of the Y_i variables to an SRLK model, assuming the

extended experimental hypotheses, does not change the proofs of the previous lemmas and theorems. □

4.3.4 A few examples of (extended) SRLK models

In spite of some presumably strong modelling assumptions, the (extended) SRLK models can describe a broad range of cytokine-receptor systems. The monomeric receptor model with IEK of Section 4.2.4, the IL-7R model described in Section 3.2 and the trimeric receptor model of Section 4.2.6 are SRLK models with $n = 1$, $n = 2$ and $n = 3$ trans-membrane chains, respectively. The IL-7R model with the additional subunit described in Section 3.4 is an extended SRLK model. In this section, we provide examples of other interleukin-signalling systems which are part of the (extended) SRLK family.

A number of interleukin receptors share different molecular components. For instance, cytokine receptors of the common gamma family (comprising the receptors for IL-2,4,7,9,15 and 21 (Rochman *et al.*, 2009)) share the common gamma chain, γ. In addition the IL-2 and IL-15 receptors share the sub-unit chain, IL-2Rβ. The competition for these sub-unit chains can be mathematically described with an extended SRLK model, as follows.

Example 93 (Extended SRLK model: IL-2/IL-2R model with formation of IL-7R and IL-15R)**.** Let us suppose we want to study the formation of IL-2/IL-2R complexes taking into account the competition for the γ chain between IL-2Rβ and IL-7Rα, and the competition for the complex γ :IL-2Rβ between the sub-units IL-2Rα and IL-15Rα. We can use an extended SRLK model with

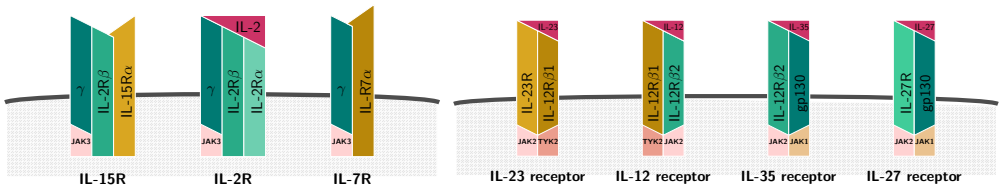
$$\{Z, X_1, X_2, X_3, L\} = \{\text{JAK3}, \gamma, \text{IL-2R}\beta, \text{IL-2}\alpha, \text{IL-2}\}$$

and

$$\{Y_1, Y_2, Y_3\} = \{\text{IL-7R}\alpha, \text{IL-15R}\alpha, \emptyset\}.$$

This example is illustrated in Figure 4.13(a).

A further extended SRLK example is that of the IL-12 family of receptors, which share multiple components (Vignali & Kuchroo, 2012), and each of which is composed of two trans-membrane sub-unit chains. The IL-12 receptor is composed



(a) IL-2/IL-2R model with IL-7R and IL-15R competition

(b) IL-12 receptor family

Figure 4.13: (a) Illustration of example 93: IL-2R, IL-7R and IL-15R competing for the common gamma chain and IL-2Rβ. The IL-2R is composed of three sub-unit chains: the gamma chain, IL-2Rβ and IL-2Rα. IL-15R is composed of the gamma chain, IL-2Rβ and the specific chain IL-15Rα. The IL-7R is composed of the gamma chain and IL-7Rα. All these receptors signal through the Janus kinase JAK3. (b) Illustration of example 94: 1. Competition for IL-12Rβ1 between the IL-12 and the IL-23 receptors. We assume that IL-23R and IL-12Rβ2 are already bound to their associated extrinsic kinase JAK2; 2. Competition for IL-12Rβ2 between the IL-12 and IL-35 receptors. We consider the complexes IL-12Rβ1:TYK2 and gp130:JAK1 already pre-formed; 3. Competition for gp130 between the IL-35 and the IL-27 receptors. We consider the complexes IL-12Rβ2:JAK2 and IL-27R:JAK2 already pre-formed.

of the sub-unit chains IL-12Rβ1 and IL-12Rβ2. The IL-23 receptor signals via the IL-23R chain and the IL-12Rβ2 chain. The IL-27R (also known as WSX-1) and glycoprotein 130 (gp130) form the IL-27 receptor. Finally, IL-12Rβ2 and gp130 form the IL-35 receptor. The sub-unit chains gp130, IL-12Rβ1 and IL-12Rβ2 bind to a kinase from the JAK family (JAK1, TYK2 and JAK2, respectively). This competition can be described with extended SRLK models as follows.

Example 94 (Extended SRLK models: IL-12R family)**.** We provide three examples of extended SRLK systems which characterise the competition for receptor sub-units between receptors of the IL-12 family (see Figure 4.13(b)).

1. Suppose we want to study the IL-12-induced signalling process taking into account the competition for IL-12Rβ1. We can use an extended SRLK model with

$$\{Z, X_1, X_2, L, Y_1, Y_2\} = \{\text{TYK2}, \text{IL-12R}\beta 1, \text{IL-12R}\beta 2^*, \text{IL-12}, \text{IL-23R}^*, \emptyset\}.$$

2. To study IL-35-induced signalling taking into account the competition for

IL-12Rβ2, we can use an extended SRLK model with

$$\{Z, X_1, X_2, L, Y_1, Y_2\} = \{\text{JAK2}, \text{IL-12R}\beta 2, \text{gp130}^*, \text{IL-35}, \text{IL-12R}\beta 1^*, \emptyset\}.$$

3. An extended SRLK model with

$$\{Z, X_1, X_2, L, Y_1, Y_2\} = \{\text{JAK1}, \text{gp130}, \text{IL-27R}^*, \text{IL-27}, \text{IL-12R}\beta 2^*, \emptyset\}$$

can describe the IL-27-induced signalling, when there is competition for the sub-unit chain gp130 with the IL-35 receptor.

Above we have made use of the notation X^* to denote the pre-formed complex composed of the receptor chain X and its intra-cellular extrinsic kinase (TYK2 for IL-12Rβ1, JAK1 for gp130 and JAK2 for all the others).

4.4 Investigating the EC_{50} of SRLK models

4.4.1 Amplitude of SRL models

The SRLK models are a generalisation of the models with IEK and no allostery studied in the previous sections of this thesis. Similarly, we can define a family of models that generalise the RTK models. Let us call this new family *sequential receptor-ligand (SRL)* systems. Making use of the same notation as for the SRLK systems, the biochemical reaction network for a general SRL model (n trans-membrane chains) is given by

$$\begin{array}{lcllll}
X_2 & + & X_1 & \rightleftharpoons & X_1 : X_2 & K_1, \\
\vdots & & \vdots & & \vdots & \vdots \\
X_{i+1} & + & X_1 : \ldots : X_i & \rightleftharpoons & X_1 : \ldots : X_{i+1} & K_i, \\
\vdots & & \vdots & & \vdots & \vdots \\
X_n & + & X_1 : \ldots : X_{n-1} & \rightleftharpoons & X_1 : \ldots : X_n & K_{n-1}, \\
L & + & X_1 : \ldots : X_n & \rightleftharpoons & X_1 : \ldots : X_n : L & K_n,
\end{array} \tag{4.94}$$

We consider the same notation, hypotheses and definitions as for the SRLK models. Thus, the signalling function is

$$\sigma(L) \equiv K_1 \dots K_n L x_1 \dots x_n, \tag{4.95}$$

and a SRL model with n trans-membrane chains at the steady state is described by the following polynomial system:

$$N_1 = y_1 + K_1 x_1 x_2 + \dots + K_1 \dots K_{n-1} x_1 \dots x_n + K_1 \dots K_n L x_1 \dots x_n, \tag{4.96a}$$

$$\vdots$$

$$N_i = x_i + K_1 \dots K_{i-1} x_1 \dots x_i + \dots \tag{4.96b}$$
$$+ K_1 \dots K_{n-1} x_1 \dots x_n + K_1 \dots K_n x_1 \dots x_n L,$$

$$\vdots$$

$$N_n = x_n + K_1 \dots K_{n-1} x_1 \dots x_n + K_1 \dots K_n L x_1 \dots x_n. \tag{4.96c}$$

By following a similar process to that of the SRLK models [1], we can show that the amplitude of SRL models is the total number of the limiting component:

$$A_{SRL} = \min(N_1, ..., N_n). \tag{4.97}$$

With this final result, we generalised Remark 71 for receptors composed of different trans-membrane chains, with or without downstream kinase.

4.4.2 EC_{50} of SRL and SRLK models

Remark 72 points out that the EC_{50} for SRLK (with no allostery) and SRL models with $n = 1$ and $n = 2$ chains are the same. In other words, the EC_{50} of these

[1] (1) Prove that $\lim_{L \to +\infty} \sigma(L) = c_\sigma > 0$. (2) Show that no x_i tends to 0 faster than $\frac{1}{L}$ nor grows as a power of L when $L \to +\infty$. (3) Show that the concentration of the limiting component tends to 0 as fast as $\frac{1}{L}$ and that the concentrations of the other trans-membrane chains tend to a positive constant. (4) Show that $\lim_{L \to +\infty} (\sigma(L)) = \min(N_1, ...N_n)$. (5) Prove that this limit is the amplitude of the system.

SRLK models is independent of the kinase z (or JAK). In Section 4.2.6, we show that this is also the case for the heterotrimeric receptor. I conjecture that the EC_{50} of SRLK models with no allostery is independent of the kinase z (*i.e.*, independent from N_z and K_0) for any trans-membrane chains number, n. Unfortunately, I did not manage to prove this result rigorously. In this section, however, I advance a sketch of the proof that justifies my intuition.

To distinguish SRL and SRLK models, I write y_i the concentration of the trans-membrane chain X_i in the SRL model[1] for $i = 1, \dots, n$. To find the EC_{50}, we want to solve the following problem for the SRLK:

Problem 95 (Finding the EC_{50} of SRLK systems)**.** Find L such that

$$K_0 z K_1 \dots K_n L x_1 \dots x_n = \frac{K_0 z}{(1 + K_0 z)} \frac{\min(N_1, \dots, N_n)}{2}.$$

The EC_{50} of SRL models can be found by solving:

Problem 96 (Finding the EC_{50} of SRL systems)**.** Find L such that

$$K_1 \dots K_n L y_1 \dots y_n = \frac{\min(N_1, \dots, N_n)}{2}.$$

In Lemma 83, we computed an expression for z that is independent of the ligand concentration, L. Thus, we can remove equation (4.85a) from system (4.85) and consider the remaining polynomials as a system of n variables $x_1, ..., x_n$, in which z is replaced by its expression as a function of the parameters. Now, by a change of basis $x_1(1 + K_0 z) = y_1$ and $y_i = x_i$, this modified polynomial system becomes system (4.96). With the same change of basis, Problems 95 and 96 are equivalent. This results that to study the EC_{50} of SRLK systems, it is sufficient to study the corresponding SRL system. It follows that the EC_{50} of SRLK systems is independent of the kinase (independent from N_z and K_0).

In order to write a rigorous proof, I would need to justify the removal of the first equation in system (4.85), as well as the change of basis with more precise algebraic tools. I will also probably need to show that SRL and SRLK systems admit a unique EC_{50}. The existence of such quantity follows from the definition

[1] In other words, we re-write the polynomial system (4.96) with y_i instead x_i, i=1,...,n.

of such biochemical networks. The uniqueness of the EC_{50}, however, is a difficult question as I cannot show that the signalling function σ is monotonic.

4.5 Exploring the combination of simple models: thermodynamic cycles

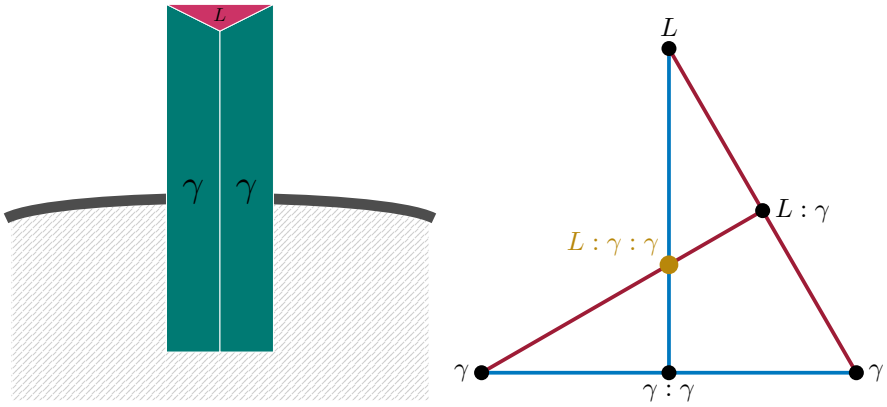
The method presented in Section 4.1 allows one to compute the analytic expression for the amplitude and EC_{50} of a receptor-ligand model when the model is simple enough (so the degree of the polynomials of the Gröbner basis is not too high). For more complicated models, we managed to reduce the cost of the computation of the amplitude and EC_{50} but the analytic study is intractable very quickly. For some models (as demonstrated with the homodimeric model with IEK in Section 4.2.5), the computation of the Gröbner basis is a challenge and the result is composed of very long polynomials of high degree that are very impractical to analyse. Thus, the domain of application of the method presented in Section 4.1 seems limited to the computation of pharmacological quantities of simple models. Now, suppose one could decompose a complex model into sub-models for which the analytic amplitude (resp. EC_{50}) can be computed. Could one link the amplitude (resp. EC_{50}) expression of the sub-models to the amplitude (resp. EC_{50}) of the complex model? Here, I investigate this question by studying simple models, which I later combine to artificially create a more complex model.

In this section, I consider the different orders to form the signalling complex of the homodimeric and heterodimeric receptors with intrinsic kinase activity (RTK). The homodimeric RTK model described in Section 4.2.2 assumes that the ligand binds to the dimer composed of two γ chains to form the complex $L : \gamma : \gamma$. In Section, 4.2.3, I described a heterodimeric RTK model in which the ligand binds to the dimer composed of the primary and secondary chains, γ and α, to form the signalling complex $L : \gamma : \alpha$. The assembling order of these models will be referred to as *order A* or *model A*. We now examine different sequential assembling orders, obtained by transposition (permutation of two elements) of species in the chemical reactions of models A. There is one alternative model of the homodimeric

RTK A, which will be called homodimeric RTK B, obtained by permutation of the ligand with one of the γ chains. Two alternative models of the heterodimeric RTK A, called heterodimeric RTK B and C, can be obtained by permuting the ligand with the primary or the secondary chain. Models A and their variants (which I will also refer to as *base models*) are simple enough to compute analytic expressions of their amplitude and EC_{50} with the method described in Section 4.1. A base model assembles the signalling complex sequentially by adding one species after another in a unique way. I later combine these base models together to create more complicated models which I will call *combined models*. Thus I obtain complex models which can be decomposed into sub-models (the base models). By definition, the combined models display thermodynamic cycles: the signalling complex can be formed by several routes of reactions, and the detailed-balance steady state equations enforce constraints on the affinity constants. A combined model of models A and B will be denoted model AB. The models studied in this section are recapitulated in Table 2.2.

Biologically, it is very hard to know exactly in which order a signalling complex is formed, though crystallography techniques can eliminate certain possibilities by examining the binding site structures of the proteins (Boulanger *et al.*, 2003; Stauber *et al.*, 2006; Wang *et al.*, 2005). Most likely, the different orders described by the base models co-exist and the preferred order to form the signalling complex is the one constituted of the chemical reactions with the highest association constants (higher association rate/slower dissociation rate). Note that the signalling complex could also be formed by association of several pre-formed complexes (for instance, the IL-2R/IL-2 complex, L :IL-2Rα :IL-2Rβ : γ_c, could be formed by association of the pre-formed complexes L :IL-2Rα and γ_c :IL-2Rβ). This is however out of the scope of this section.

We make use of the notation and hypotheses described in Section 4.2, and assume mass-action law. The ligand L is always assumed to be in excess, *i.e.*, it is considered as a parameter of the model. All the models of this section are at steady state and are deficiency 0: they admit a unique biologically meaningful steady state.



(a) Signalling complex (b) Model configurations

Figure 4.14: (a) The homodimeric RTK signalling complex is the association of the receptor (composed of two identical γ chains) and the ligand L. (b) The signalling complex (in yellow) can be formed in two manners corresponding to two base models: the receptor, $\gamma : \gamma$, is formed first and the ligand binds to the receptor (homodimeric RTK A model, blue path) or the ligand binds first to one of the γ chain and the other γ chain binds to the complex $\gamma : L$ (homodimeric RTK B model, red path). The combined model homodimeric RTK AB follows blue and red paths.

4.5.1 Half-maximal inhibitory concentration or IC_{50}

Before starting any computation, let us define the concept of half-maximal inhibitory concentration:

Definition 97 (Half-maximal inhibitory concentration)**.** Let A denote the amplitude of the dose-response curve function σ. The *half-maximal inhibitory concentration*, or IC_{50} is the ligand concentration L^* which satisfies $\sigma(L^*) = min(\sigma) + \dfrac{max(\sigma) - min(\sigma)}{2} = min(\sigma) + \dfrac{A}{2}$ when σ is decreasing.

The IC_{50} is very similar to the EC_{50}, except, it is defined when the dose-response curve is decreasing, *i.e.*, increasing the ligand concentration inhibits the cell's response. In this section, we will encounter bell-shaped dose-response curves: the equation $\sigma(L^*) = min(\sigma) + \dfrac{max(\sigma) - min(\sigma)}{2} = min(\sigma) + \dfrac{A}{2}$ admits two solutions, the smallest is the EC_{50} and the largest is the IC_{50}.

4.5.2 Homodimeric RTK models

In this subsection, we study different ways of forming the homodimeric RTK signalling complex, composed of two identical chains γ and the ligand L, $\gamma:\gamma:L$ (see Figure 4.14(a)). It can be formed in two different manners: the γ chains dimerise then bind to the ligand, or the ligand binds first to one γ chain before binding to the other chain (see Figure 4.14(b)). These two orders can also co-exist.

Homodimeric RTK B

Let us consider the homodimeric RTK model of Section 4.2.2, which we refer to as homodimeric RTK A. By permuting a γ chain with the ligand in the chemical reactions, we obtain the homodimeric RTK model B. Its chemical reaction network is as follows:

$$\begin{array}{lcll} L+\gamma & \rightleftharpoons & L:\gamma & K_3', \\ L:\gamma+\gamma & \rightleftharpoons & L:\gamma:\gamma & K_2'. \end{array}$$

Note that I distinguish the affinity constants of model B from model A using primes. The index 3 still refers to the binding of the ligand L to a chain or complex. Writing the ordinary differential equations associated to the chemical reaction scheme and combining the steady state and conservation equations, I obtain the following polynomial, which describes model B:

$$0 = -N_x + (1 + K_3'L)x + 2K_2'K_3'Lx^2. \tag{4.98}$$

The positive solution of this polynomial, which represents the number of unbound γ chains per cell at steady state, is:

$$x = \frac{-1 - K_3'L + \sqrt{(1+K_3'L)^2 + 8K_2'K_3'LN_x}}{4K_2'K_3'L}. \tag{4.99}$$

Computation of the amplitude: The signalling complex is $L:\gamma:\gamma$, thus, the signalling function of this model is:

$$\sigma(L) \equiv K_3'K_2'Lx^2, \tag{4.100}$$

where x is the steady state expression found above. Since $\sigma(0) = 0$, the amplitude is defined as the maximum of this function. Unfortunately, contrary to the previous models of this chapter, the maximum of σ is not its limit when $L \to +\infty$ (in fact, $\sigma \to 0$). We find the maximum of σ by computing its derivative. For concise notation, we write

$$\Delta \equiv (1 + K_3'L)^2 + 8N_xK_2'K_3'L, \tag{4.101}$$

and the numerator of x:

$$f \equiv -1 - K_3'L + \sqrt{\Delta}. \tag{4.102}$$

The signalling function can be re-written with the new notation:

$$\sigma(L) = \frac{f^2}{16K_2'K_3'L}. \tag{4.103}$$

We compute the derivative of σ:

$$\begin{aligned}
\frac{d\sigma}{dL}(L) &= \frac{f}{16K_2'K_3'L^2}(2\frac{df}{dL}L - f) \\
&= \frac{f}{16K_2'K_3'L^2}(-2K_3'L + \frac{2K_3'(1 + K_3'L) + 8N_xK_2'K_3'}{\sqrt{\Delta}}L + 1 + K_3'L - \sqrt{\Delta}) \\
&= \frac{f}{16K_2'K_3'L^2\sqrt{\Delta}}(K_3'^2L^2 - 1 + (1 - K_3'L)\sqrt{\Delta}) \\
&= -\frac{(-1 + K_3'L)f^2}{16K_2'K_3'L^2\sqrt{\Delta}}.
\end{aligned} \tag{4.104}$$

The equation $\frac{d\sigma}{dL}(L) = 0$ has a unique solution:

$$L_a = \frac{1}{K_3'}. \tag{4.105}$$

Note that $\frac{d\sigma}{dL}$ is positive on $[0, \frac{1}{K_3'})$ and negative on $(\frac{1}{K_3'}, +\infty)$ so $\sigma(L_a)$ is a maximum. The dose-response curve increases then decreases (a short numerical simulation shows it is in fact a bell-shaped curve (see Figure 4.17(a) in Section 4.5.4)). This result is not surprising considering the model: as the availability

of ligand increases, more and more signalling complexes are formed. As it keeps increasing, there is enough ligand (alternatively one could interpret it as: the binding of the ligand to the γ chain is fast enough) to bind each γ chain and form $L:\gamma$ complexes only, thus decreasing the number of signalling complexes until none are formed anymore. Finally, the amplitude of the homodimeric receptor model B (denoted A^{ho_B}), is given by $\sigma(L_a)$:

$$A^{ho_B} \equiv \sigma(L_a) = \frac{(-1+\sqrt{1+2K_2'N_x})^2}{4K_2'}. \tag{4.106}$$

Remark 98. An analytic expression for the steady state and the amplitude of this model have already been computed in Ref. White *et al.* (2022). However, White *et al.* considered another signalling function. As a result, they did not compute the EC_{50} expression corresponding to the model of this section. Since the derivation of the EC_{50} and IC_{50} directly follows from the computation of the amplitude, and as it is the first time that a bell-shaped dose-response curve is encountered in this thesis, I chose to detail the computation of the steady state and amplitude of this model.

Computation of the EC_{50} and IC_{50}: Since the dose-response curve of this model is bell-shaped, we expect two solutions to the equation:

$$\sigma(L) = \frac{A^{ho_B}}{2}. \tag{4.107}$$

The EC_{50} is the solution on the ascendant part of the curve, the IC_{50} the solution on the descendant part of the dose-response curve. To compute their expression, one could solve directly equation (4.107) or compute a Gröbner basis as described in the method of Section 4.1. This time, I chose the first option, but the two methods yield the same result. For concise notation, let us write $\mathcal{A} \equiv -1+\sqrt{1+2K_2'N_x}$. Replacing x and the amplitude by their respective expression, equation (4.107) can be re-written as:

$$\frac{(-1-K_3'L+\sqrt{(1+K_3'L)^2+8K_2'K_3'LN_x})^2}{2K_3'L} = \mathcal{A}^2. \tag{4.108}$$

Isolating the square-root and squaring the equation to remove it, we obtain after several simplifications the following polynomial in L:

$$K_3'^2 L^2 - K_3' L \left(\frac{16 K_2'^2 N_x^2}{2\mathcal{A}^2} + \frac{\mathcal{A}^2}{2} - 2 - 4K_2' N_x \right) + 1 = 0. \tag{4.109}$$

Replacing $\mathcal{A}$ by its modified definition,

$$\mathcal{A} = \frac{(-1 + \sqrt{1 + 2K_2' N_x})(1 + \sqrt{1 + 2K_2' N_x})}{1 + \sqrt{1 + 2K_2' N_x}} = \frac{2K_2' N_x}{1 + \sqrt{1 + 2K_2' N_x}},$$

we obtain the following simplified polynomial in L:

$$K_3'^2 L^2 - K_3' L \left(K_2' N_x + 3(1 + \sqrt{1 + 2K_2' N_x}) \right) + 1 = 0, \tag{4.110}$$

which has two positive roots corresponding to the expression of the IC_{50} and the EC_{50}:

$$\begin{aligned} EC_{50}^{ho_B} &= \frac{K_2' N_x + 3(1 + \sqrt{1 + 2K_2' N_x}) - \sqrt{-4 + (K_2' N_x + 3(1 + \sqrt{1 + 2K_2' N_x}))^2}}{2K_3'}, \\ IC_{50}^{ho_B} &= \frac{K_2' N_x + 3(1 + \sqrt{1 + 2K_2' N_x}) + \sqrt{-4 + (K_2' N_x + 3(1 + \sqrt{1 + 2K_2' N_x}))^2}}{2K_3'}. \end{aligned} \tag{4.111}$$

Homodimeric RTK AB

Now, consider that the ways of forming the signalling complex $L : \gamma : \gamma$ in homodimeric RTK A and B models are possible. The chemical reaction scheme of the combined model, called homodimeric model AB, is as follows:

$$\begin{array}{lcll} \gamma + \gamma & \rightleftharpoons & \gamma : \gamma & K_2, \\ L + \gamma : \gamma & \rightleftharpoons & L : \gamma : \gamma & K_3, \\ L + \gamma & \rightleftharpoons & L : \gamma & K_3', \\ L : \gamma + \gamma & \rightleftharpoons & L : \gamma : \gamma & K_2'. \end{array}$$

Writing the ordinary differential equations describing the system, we derive the conservation equation

$$N_x = x + [L:x] + 2[x:x] + 2[L:x:x], \tag{4.112}$$

and the steady state equations

$$\begin{aligned} [x:x] &= K_2 x^2, \\ [L:x] &= K_3' L x, \\ [L:x:x] &= K_3' K_2' L x^2, \\ [L:x:x] &= K_3 K_2 L x^2, \end{aligned} \tag{4.113}$$

where $[X:Y]$ denotes the concentration of complex $X:Y$. The existence of a detailed-balanced steady state enforces conditions on the product of affinity constants:

$$K_2 K_3 = K_2' K_3'. \tag{4.114}$$

Combining the steady state relations and the conservation equation, we obtain the following polynomial:

$$0 = -N_x + (1 + K_3' L)x + 2K_2(K_3 L + 1)x^2. \tag{4.115}$$

The positive solution of this polynomial represents the number of unbound γ chains at steady state:

$$x = \frac{-1 - K_3' L + \sqrt{(1 + K_3' L)^2 + 8N_x K_2(K_3 L + 1)}}{4K_2(K_3 L + 1)}. \tag{4.116}$$

Computation of the amplitude: The signalling function is defined by the number of signalling complexes $L:\gamma:\gamma$ formed at steady state as a function of the ligand concentration:

$$\sigma(L) \equiv K_3 K_2 L x^2 = K_3' K_2' L x^2. \tag{4.117}$$

As in the case in the homodimeric model B, the signalling function of this model does not attain its maximum when $L \to +\infty$ (in fact, $\sigma(L) \to 0$ and a short

numerical simulation shows that the dose-response curve seems to be bell-shaped). Thus we need to compute the derivative of σ to compute its maximum (*i.e.*, the amplitude since $\sigma(0) = 0$). We want to find L_a such that

$$\frac{d\sigma}{dL}(L_a) = 0, \tag{4.118}$$

and the amplitude will be $\sigma(L_a)$. Unfortunately, even with the use of a symbolic computation software such as Mathematica (Wolfram Research, Inc., 2019), the computation of L_a and the amplitude results in long and over-complex expressions which are not useful for further analytic study or even for any computational use (I provide the Mathematica notebook in Appendix E).

Computation of the EC_{50}/IC_{50}: Without the amplitude, we cannot derive any explicit analytic expression for the EC_{50} and IC_{50}. However, if we write $\mathcal{A}$ the unknown amplitude of this model, we can reduce the computation of the EC_{50} and IC_{50} to solving a polynomial of degree 3, supposing we obtained the amplitude numerically. Indeed, by computing a Gröbner basis of polynomial (4.115) and the additional equation

$$2K_2K_3Lx^2 - \mathcal{A} = 0, \tag{4.119}$$

considering L and x as variables, we obtain the following polynomial system:

$$\begin{aligned} 0 = {} & L^3 + 2\frac{-K_2K_3\mathcal{A}^2 - K_2K_3N_x^2 + \mathcal{A}(K_3' + 2K_2K_3N_x)}{\mathcal{A}K_3'^2}L^2 \\ & + \frac{1 + 4K_2(N_x - \mathcal{A})}{K_3'^2}L - \frac{2\mathcal{A}K_2}{K_3K_3'^2}, \end{aligned} \tag{4.120a}$$

$$0 = x + \frac{K_3K_3'^2\mathcal{A}(N_x - \mathcal{A})L^2 + pL + q\mathcal{A}}{2K_2\mathcal{A}(\mathcal{A}(K_3' - K_3) + K_3N_x)}, \tag{4.120b}$$

where

$$\begin{aligned} p = {} & 2\mathcal{A}^3K_2K_3^2 - 2K_2K_3^2N_x^3 + 2\mathcal{A}K_3N_x(K_3' + 3K_2K_3N_x) \\ & + \mathcal{A}^2(-2K_3K_3' + K_3'^2 - 6K_2K_3^2N_x), \\ q = {} & 2\mathcal{A}^2K_2K_3 + K_3N_x(1 + 2K_2N_x) + \mathcal{A}(K_3' - K_3(1 + 4K_2N_x)). \end{aligned}$$

Polynomial (4.120a) is of degree 3 so has at least one (and no more than three) positive real root. By decreasing order of monomial degree, its coefficient signs are $(+, ?, +, -)$ with the second coefficient sign being unknown *a priori*. The roots of polynomial (4.120a) are potential EC_{50} and IC_{50} expressions of the model. Only roots that lead to a positive value of x (which is obtained by solving polynomial (4.120b)) are actually EC_{50} and IC_{50}. Since we have $\sigma(0) = 0$ and $\lim_{L\to+\infty} \sigma(L) = 0$, we expect an even number of positive real roots of polynomial (4.120a) which lead to a positive value of x (in fact exactly two because of the degree of polynomial (4.120a)). Thus according to Descartes'rule (see Section 2.6.1) of sign, the second coefficient of polynomial (4.120a),

$$-K_2K_3\mathcal{A}^2 - K_2K_3N_x^2 + \mathcal{A}(K_3' + 2K_2K_3N_x), \tag{4.121}$$

must be negative.

Alternatively, the EC_{50} and IC_{50} can be found numerically by numerically solving the equation $\sigma(L^*) = \dfrac{\mathcal{A}}{2}$, making use of the analytic expression for σ. The largest solution will be the IC_{50} and the smallest the EC_{50}.

Note that in absence of closed-form amplitude expression, the analytic study is very limited: the computation of the Gröbner basis does not show the dependency on the parameters of the pharmacological quantities. Moreover, numerically solving polynomial (4.120a) might be difficult if the problem is ill-conditioned (see example in Section 3.4)

Refining the upper bound for the amplitude: As equation (4.121) must be negative, it provides an interval in which the amplitude $\mathcal{A}$ ranges. Indeed, equation (4.121) is a polynomial of degree 2 in $\mathcal{A}$ with a negative leading coefficient. Thus $\mathcal{A}$ has to range in interval $(0, A^-) \cup (A^+, +\infty)$ where

$$\begin{aligned} A^- &= \frac{1 + 2K_2'N_x - \sqrt{1 + 4K_2'N_x}}{2K_2'}, \\ A^+ &= \frac{1 + 2K_2'N_x + \sqrt{1 + 4K_2'N_x}}{2K_2'}, \end{aligned} \tag{4.122}$$

are the roots of the polynomial defined by equation (4.121). By definition, the amplitude has also to be smaller than N_x and we notice that $A^+ > N_x > A^-$. Hence, $\mathcal{A} \in (0, A^-)$. This interval, however, is only slightly more precise [1] than the natural upper bound N_x (which is also the amplitude of model A). Since

$$A^- - A^{ho_B} = \frac{K_2' N_x + \sqrt{1 + 2K_2' N_x} - \sqrt{1 + 4K_2' N_x}}{2K_2'} > 0,$$

the interval $(0, A^-)$ is too large to compare the amplitude of model AB with the amplitude of model B.

4.5.3 Heterodimeric RTK models

We now study different ways of forming the heterodimeric RTK signalling complex, composed of a primary chain γ, a secondary chain α and the ligand L, $\gamma : \alpha : L$ (see Figure 4.15(a)). It can be formed in three different manners: (A) The ligand binds to the dimerised receptor $\alpha : \gamma$; or (B) the ligand binds first to the γ chain before binding to the α chain; or (C) the ligand binds first to the α chain (see Figure 4.15(b)). Two of these orders, or all of them, can also co-exist.

Heterodimeric RTK B and C

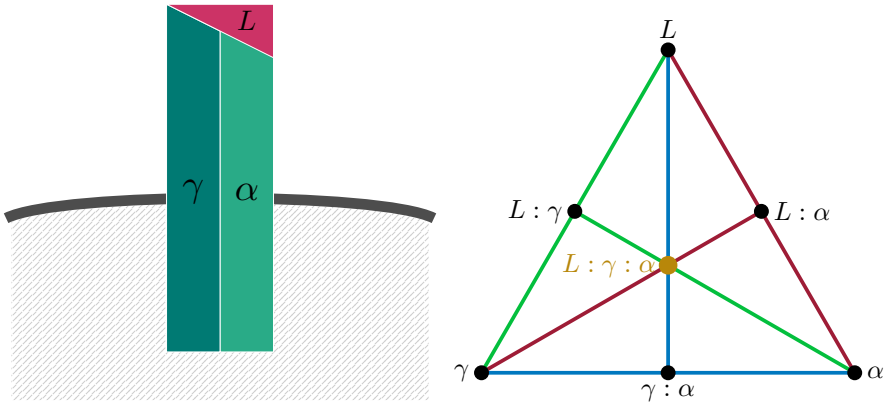
Let us consider the heterodimeric RTK model of Section 4.2.3, which we refer to as heterodimeric RTK A. By permuting the α chain and the ligand in the chemical reactions, I obtain the heterodimeric RTK B model. Its chemical reaction network is as follows:

$$\begin{array}{lcll} \gamma + L & \rightleftharpoons & L : \gamma & K_3', \\ L : \gamma + \alpha & \rightleftharpoons & L : \gamma : \alpha & K_2'. \end{array}$$

The steady state of this model is described by the following polynomial system:

$$\begin{aligned} 0 &= -N_x + (K_3' L + 1)x + K_3' K_2' L x y, \\ 0 &= -N_y + y + K_3' K_2' L x y. \end{aligned} \tag{4.123}$$

[1] It, of course, depends on the parameter values. But the difference between N_x and A^- is negligible (compared to N_x) for the parameter values used in the numerical simulations of this section.



(a) Signalling complex (b) Model configurations

Figure 4.15: (a) The heterodimeric RTK signalling complex is the association of the receptor (composed of a γ chain and a α chain) and the ligand L. (b) The signalling complex (in yellow) can be formed in three manners corresponding to three base models: the receptor is formed first and the ligand binds to the receptor (heterodimeric RTK A model, blue path), the ligand binds first to the γ chain and the α binds to the complex $\gamma : L$ (heterodimeric RTK B model, green path) or the ligand binds first to the α chain and the γ chain binds to the complex $\alpha : L$ (heterodimeric RTK C, red path). The combined model heterodimeric RTK AB follows blue and green paths, model heterodimeric RTK AC follows the blue and red paths and the model heterodimeric BC follows the green and red paths. The combined model heterodimeric RTK ABC follows the blue, red and green paths.

Reorganising the polynomials by hand or computing a Gröbner basis, we obtain a triangular system easier to solve:

$$\begin{aligned} 0 &= -N_x + (1 + K_3' L(1 + K_2'(N_y - N_x)))x + K_2' K_3' L(K_3' L + 1)x^2, \\ 0 &= N_x - N_y - (1 + K_3' L)x + y. \end{aligned} \tag{4.124}$$

Solving the first polynomial and selecting the positive solution we obtain an expression for the number of unbound γ chains at the steady state:

$$x = \frac{-1 - K_3' L(1 + K_2'(N_y - N_x)) + \sqrt{\Delta}}{2K_2' K_3' L(K_3' L + 1)}, \tag{4.125}$$

where we wrote $\Delta \equiv (1 + K_3' L + K_2' K_3' L(N_y - N_x))^2 + 4N_x K_2' K_3' L(K_3' L + 1)$. The number of unbound α chains at steady state can be obtained through the second polynomial knowing x.

Computation of the amplitude: The signalling function σ is the number of signalling complexes $L : \gamma : \alpha$ as a function of the concentration of ligand:

$$\sigma(L) \equiv K_3' K_2' L x y. \tag{4.126}$$

Note that polynomial system (4.123) gives $K_3' K_2' L x y = N_x - (1 + K_3' L)x$. By replacing x by its expression (Eq. (4.125)), σ can be re-written:

$$\sigma(L) = \frac{1 + K_3' L(1 + K_2'(N_y + N_x)) - \sqrt{\Delta}}{2K_2' K_3' L}. \tag{4.127}$$

Since $\sigma(0) = 0$ the amplitude is defined as the maximum of σ. Additionally, since we have

$$\frac{d\sigma}{dL} = \frac{\sigma}{L\sqrt{\Delta}} > 0,$$

and know that σ is a bounded function by definition, the amplitude is:

$$\begin{aligned} A^{het_B} &\equiv \lim_{L \to +\infty} \sigma(L) \\ &= \frac{1 + K_2'(N_y + N_x) - \sqrt{1 + K_2'^2(N_y - N_x)^2 + 2K_2'(N_x + N_y)}}{2K_2'}. \end{aligned} \tag{4.128}$$

Computation of the EC_{50}: To find the EC_{50}, let us compute a Gröbner basis of the polynomial system (4.123) augmented by the equation

$$K_3' K_2' L x y - \frac{A^{het_B}}{2} = 0. \tag{4.129}$$

Making use of the software Macaulay2 (Grayson & Stillman), I obtain:

$$0 = L - \frac{2A^{het_B}}{K_3'((A^{het_B})^2 K_2' + 4N_x N_y K_2' - 2A^{het_B}(1 + K_2'(N_x + N_y)))}, \tag{4.130a}$$

$$0 = y - N_y + \frac{A^{het_B}}{2}, \tag{4.130b}$$

$$0 = x - N_x + A^{het_B}(\frac{1}{2} - \frac{1}{K_2(A^{het_B} - 2N_y)}). \tag{4.130c}$$

The solution of equation (4.130a) yields the expression for the EC_{50} as a function of the amplitude:

$$L = \frac{2A^{het_B}}{K_3' \left[(A^{het_B})^2 K_2' + 4N_x N_y K_2' - 2A^{het_B}(1 + K_2'(N_x + N_y))\right]}. \tag{4.131}$$

Let us note that the denominator of L is a polynomial of degree 2 in A^{het_B} and let us factorise it :

$$\begin{aligned} &K_3' \left[(A^{het_B})^2 K_2' + 4N_x N_y K_2' - 2A^{het_B}(1 + K_2'(N_x + N_y))\right] \\ &= K_2' K_3' \left(A^{het_B} - \frac{b + \sqrt{d}}{K_2'}\right)\left(A^{het_B} - \frac{b - \sqrt{d}}{K_2'}\right). \end{aligned} \tag{4.132}$$

where we wrote $b = 1 + K_2'(N_x + N_y)$ and $d = b^2 - 4N_x N_y K_2'^2$. We can also use these notation to describe the amplitude:

$$A^{het_B} = \frac{b - \sqrt{d}}{2K_2'}. \tag{4.133}$$

Replacing this expression into equation (4.132), we obtain:

$$K_2' K_3' \left(A^{het_B} - \frac{b + \sqrt{d}}{K_2'}\right)\left(A^{het_B} - \frac{b - \sqrt{d}}{K_2'}\right) = \frac{K_3'}{4K_2'}\left(\sqrt{d} - b\right)\left(-b - 3\sqrt{d}\right). \tag{4.134}$$

Finally, merging equations (4.131), (4.133) and (4.134) together, we obtain the final expression for the EC_{50}:

$$EC_{50}^{het_B} = \frac{4}{K_3' \left(1 + K_2'(N_x + N_y) + 3\sqrt{1 + K_2'^2(N_x - N_y)^2 + 2K_2'(N_x + N_y)}\right)}. \tag{4.135}$$

The heterodimeric RTK C model is the heterodimeric RTK B model where the role of γ and α (which have the same mathematical role) is interchanged. The chemical reaction scheme is as follows:

$$\begin{array}{lcll} L + \alpha & \rightleftharpoons & L:\alpha & K_3'', \\ L:\alpha + \gamma & \rightleftharpoons & L:\gamma:\alpha & K_2''. \end{array}$$

We obtain the amplitude and EC_{50} expression of this model by replacing K_2' and K_3' in the corresponding expressions of model B by K_2'' and K_3'' (note that these expressions were already symmetric in N_x and N_y). Thus by analogy with heterodimeric RTK model B, the expressions of the amplitude and EC_{50} of heterodimeric RTK model C are:

$$A^{het_C} = \frac{1 + K_2''(N_y + N_x) - \sqrt{1 + K_2''^2(N_x - N_y)^2 + 2K_2''(N_x + N_y)}}{2K_2''}, \tag{4.136}$$

and

$$EC_{50}^{het_C} = \frac{4}{K_3'' \left(1 + K_2''(N_x + N_y) + 3\sqrt{1 + K_2''^2(N_y - N_x)^2 + 2K_2''(N_x + N_y)}\right)}. \tag{4.137}$$

Heterodimeric RTK AB and AC

Consider now that the signalling complex can be formed in two different ways, following the orders of models heterodimeric RTK A and homodimeric RTK B. The chemical reaction scheme is as follows:

$$\begin{array}{lcll} \gamma + \alpha & \rightleftharpoons & \gamma:\alpha & K_2, \\ L + \gamma:\alpha & \rightleftharpoons & L:\gamma:\alpha & K_3, \\ \gamma + L & \rightleftharpoons & L:\gamma & K_3', \\ L:\gamma + \alpha & \rightleftharpoons & L:\gamma:\alpha & K_2'. \end{array}$$

Writing the steady state equations, the positive detailed-balanced steady state only exists if the following condition is enforced:

$$K_3K_2 = K_3'K_2'. \tag{4.138}$$

Combining the conservation and steady state equations with the detailed balance constraint, we obtain the following polynomial system which describes the model:

$$\begin{aligned} 0 &= -N_x + (1 + K_3'L)x + K_2xy + K_3K_2Lxy, \\ 0 &= -N_y + y + K_2xy + K_3K_2Lxy. \end{aligned} \tag{4.139}$$

We compute a Gröbner basis of these two polynomials (or the polynomials are simple enough to be reorganised by hand) and obtain the following triangular system:

$$0 = -(1 + K_3'L)N_y + [1 + K_3'L + K_2(1 + K_3L)(N_x - N_y)]y + K_2(1 + K_3L)y^2, \tag{4.140a}$$

$$0 = N_y - N_x + (1 + K_3'L)x - y. \tag{4.140b}$$

We select the positive solution of the Gröbner basis to obtain the number of unbound α chains (y) and γ chains (x) per cell at the steady state:

$$\begin{aligned} y &= \frac{-[1 + K_3'L + K_2(1 + K_3L)(N_x - N_y)] + \sqrt{\Delta}}{2K_2(K_3L + 1)}, \\ x &= \frac{y - N_y + N_x}{1 + K_3'L}, \end{aligned} \tag{4.141}$$

where we wrote

$$\Delta = (1 + K_3'L)^2 + K_2^2(1 + K_3L)^2(N_x - N_y)^2 + 2K_2(K_3L + 1)(K_3'L + 1)(N_x + N_y).$$

Computation of the amplitude: The signalling complex is $L : \gamma : \alpha$ and so the signalling function is defined as:

$$\sigma(L) \equiv K_3K_2Lxy = K_3'K_2'Lxy. \tag{4.142}$$

One can notice that

$$\begin{aligned} K_2xy &= \frac{K_2y(y-N_y+N_x)}{1+K_3'L} \\ &= \frac{K_2(1+K_3L)y^2+K_2(1+K_3L)(N_x-N_y)y}{(1+K_3'L)(1+K_3L)} \\ &= \frac{(1+K_3'L)(N_y-y)}{(1+K_3'L)(1+K_3L)}. \end{aligned}$$

The last simplification came from polynomial (4.140a). Thus, σ can be re-written as

$$\sigma(L) = \frac{K_3L}{1+K_3L}(N_y-y). \tag{4.143}$$

Since $\sigma(0) = 0$, the amplitude is the maximum of σ. Let us prove that the amplitude is the limit of σ when $L \to +\infty$. We first observe that:

Proposition 99. σ is a monotonically increasing function.

Proof. Notice that

$$\frac{d(N_y-y)}{dL} = \frac{(K_3-K_3')(N_y-y)}{(1+K_3L)\sqrt{\Delta}}. \tag{4.144}$$

Hence,

$$\begin{aligned} \frac{d\sigma}{dL} &= \frac{K_3(N_y-y)}{(1+K_3L)^2} + \frac{K_3L}{1+K_3L}\frac{d(N_y-y)}{dL} \\ &= \frac{K_3(N_y-y)}{\sqrt{\Delta}(1+K_3L)^2}((K_3-K_3')L+\sqrt{\Delta}). \end{aligned} \tag{4.145}$$

If $K_3 \geq K_3'$, then $\dfrac{d\sigma}{dL} > 0$ and the proof is finished. If $K_3' > K_3$, let us prove that the term $(K_3-K_3')L+\sqrt{\Delta}$ is always positive. Suppose $(K_3-K_3')L+\sqrt{\Delta} \leq 0$,

$$\begin{aligned} \sqrt{\Delta} \leq (K_3'-K_3)L \implies & \qquad \Delta \leq (K_3'-K_3)^2L^2 \\ \iff & \qquad aL^2+bL+c \leq 0, \end{aligned}$$

where

$$\begin{aligned} a &= K_2^2K_3^2(N_x-N_y)^2 + 2K_3K_3'K_2(N_x+N_y) + K_3(2K_3'-K_3), \\ b &= 2(K_3' + K_2^2K_3(N_x-N_y)^2 + K_2(K_3+K_3')(N_x+N_y), \end{aligned}$$

$$c = 1 + K_2^2(N_x - N_y)^2 + 2K_2(N_x + N_y).$$

The three coefficients are positive because $2K_3' > K_3$ by hypothesis, so the polynomial $aL^2 + bL + c$ does not have any real root. Since its leading coefficient a is positive, $aL^2 + bL + c > 0$. This is a contradiction. Thus, $(K_3 - K_3')L + \sqrt{\Delta}$ is always positive. □

Since σ is an increasing bounded (by definition) function, its maximum is reached when $L \to +\infty$. The limit of y when $L \to +\infty$ is

$$\lim_{L \to +\infty}(y) = \frac{-K_3' - K_2K_3(N_x - N_y) + \sqrt{\Delta_\infty}}{2K_2K_3}, \tag{4.146}$$

where

$$\Delta_\infty = K_3'^2 + K_3^2K_2^2(N_x - N_y)^2 + 2K_2K_3K_3'(N_x + N_y).$$

Thus the amplitude is

$$A^{het_{AB}} \equiv \frac{K_3' + K_2K_3(N_x + N_y) - \sqrt{\Delta_\infty}}{2K_2K_3}. \tag{4.147}$$

Finally, since we have $K_2K_3 = K_2'K_3'$, the amplitude can be re-written and we obtain the same expression as the amplitude of heterodimeric model B:

$$A^{het_{AB}} = A^{het_B} = \frac{1 + K_2'(N_y + N_x) - \sqrt{1 + K_2'^2(N_y - N_x)^2 + 2K_2'(N_x + N_y)}}{2K_2'}. \tag{4.148}$$

Computation of the EC_{50}: The EC_{50} is obtained by finding L^* such that $\sigma(L^*) = \frac{A^{het_{AB}}}{2}$. We compute a Gröbner basis of the polynomial system (4.139) augmented with the equation

$$K_3K_2Lxy - \frac{A^{het_{AB}}}{2} = 0, \tag{4.149}$$

and consider x, y and L as variables. For simplification of the notation, we write the amplitude $A^{het_{AB}} = \mathcal{A}$ in the following system. We obtain a new system of

polynomials:

$$0 = L^2 + \frac{2\mathcal{A}(-1 - K_2(N_x + N_y) + \mathcal{A}K_2)}{p}L + \frac{K_2\mathcal{A}^2}{K_3 p}, \tag{4.150a}$$

$$0 = y - \frac{p}{2\mathcal{A}K_2}L + \frac{2 - \mathcal{A}K_2 + 2K_2N_x}{2K_2}, \tag{4.150b}$$

$$0 = x + \frac{K_2(\mathcal{A} - 2N_y)p}{\mathcal{A}q}L + \frac{r}{q}, \tag{4.150c}$$

where

$$\begin{aligned} p &= \mathcal{A}^2K_2K_3 + 4K_2K_3N_xN_y - 2\mathcal{A}(K_3' + K_2K_3(N_x + N_y)),\\ q &= 2K_2(\mathcal{A}(K_3' - K_3) + 2K_3N_y),\\ r &= \mathcal{A}^2K_2K_3 - 2\mathcal{A}(K_3 - K_3' + 2K_2K_3N_y) + 4K_3N_y(1 + K_2N_y). \end{aligned}$$

We first show a required result.

Proposition 100. p is always a positive quantity.

Proof. Indeed, p is a polynomial of degree 2 in $\mathcal{A}$ which has two positive real roots according to Descartes'rule of sign, and a positive leading coefficient. Its roots are

$$\begin{aligned} A^+ &= \frac{K_3' + K_2K_3(N_x + N_y) + \sqrt{\Delta_\infty}}{K_2K_3},\\ A^- &= \frac{K_3' + K_2K_3(N_x + N_y) - \sqrt{\Delta_\infty}}{K_2K_3} = 2A_{AB}^{het}. \end{aligned} \tag{4.151}$$

Thus since p is always evaluated at the amplitude $A^{het_{AB}}$ defined in equation (4.147) and $A^{het_{AB}} < A^- < A^+$, we have $p > 0$ for all parameter values. □

Polynomial (4.150a) admits two roots:

$$\begin{aligned} L_1 &= \mathcal{A}\frac{1 + K_2(N_x + N_y - \mathcal{A}) + \sqrt{\Delta_{ec}}}{p},\\ L_2 &= \mathcal{A}\frac{1 + K_2(N_x + N_y - \mathcal{A}) - \sqrt{\Delta_{ec}}}{p}, \end{aligned}$$

where

$$\Delta_{ec} = 1 + K_2^2(N_x - N_y)^2 + 2K_2(N_x + N_y) + \frac{2K_2(K_3' - K_3)}{K_3}\mathcal{A}.$$

Note that the expressions of L_1 and L_2 were simplified knowing that $\mathcal{A}$ and p are positive. The two roots are positive (proof by Descartes'rule and by noting that the amplitude is always lower than N_x or N_y by definition). The EC_{50} is the positive root of polynomial (4.150a) which leads to positive values of y and x (defined by polynomials (4.150b) and (4.150c), respectively). Evaluating (4.150b) at L_1 and L_2 yields

$$y_{L=L_1} = \frac{-1 - K_2(N_x - N_y) + \sqrt{\Delta_{ec}}}{2K_2} > 0,$$
$$y_{L=L_2} = -\frac{1 + K_2(N_x - N_y) + \sqrt{\Delta_{ec}}}{2K_2} < 0.$$

Hence,

$$EC_{50}^{het_{AB}} = A^{het_{AB}} \frac{1 + K_2(N_x + N_y - A^{het_{AB}}) + \sqrt{\Delta_{ec}}}{p}, \tag{4.152}$$

where $p = (A^{het_{AB}})^2 K_2 K_3 + 4K_2 K_3 N_x N_y - 2A^{het_{AB}}(K_3' + K_2 K_3(N_x + N_y))$.

We define the heterodimeric model AC by combining the models A and C. Model AC is model AB in which the roles of α and γ are interchanged. The chemical reaction scheme is as follows:

$$\begin{array}{lcll} \gamma + \alpha & \rightleftharpoons & \gamma : \alpha & K_2, \\ L + \gamma : \alpha & \rightleftharpoons & L : \gamma : \alpha & K_3, \\ \alpha + L & \rightleftharpoons & L : \alpha & K_3'', \\ L : \alpha + \gamma & \rightleftharpoons & L : \gamma : \alpha & K_2''. \end{array}$$

From a purely mathematical point of view, models AB and AC are identical. Thus, by replacing K_2' and K_3' by K_2'' and K_3'', respectively, in the amplitude and EC_{50} expressions of the previous model (which were already symmetric in N_x and

N_y), we obtain the amplitude and EC_{50} expressions of model AC:

$$A^{het_{AC}} = A^{het_C} = \frac{1 + K_2''(N_y + N_x) + \sqrt{1 + K_2''^2(N_x - N_y)^2 + K_2''(N_x + N_y)}}{2K_2''}, \tag{4.153}$$

$$EC_{50}^{het_{AC}} = A^{het_{AC}} \frac{1 + K_2(N_x + N_y - A^{het_{AC}}) + \sqrt{\Delta_{ec}}}{p}, \tag{4.154}$$

where

$$\Delta_{ec} = 1 + K_2^2(N_y - N_x)^2 + 2K_2(N_x + N_y) + \frac{2K_2(K_3'' - K_3)}{K_3} A^{het_{AC}},$$
$$p = (A^{het_{AC}})^2 K_2 K_3 + 4K_2 K_3 N_x N_y - 2A^{het_{AC}}(K_3'' + K_2 K_3(N_x + N_y)).$$

Heterodimeric RTK BC

Consider now the case in which the signalling complex can be formed following the orders of models B and C. The chemical reaction scheme of model heterodimeric RTK BC is:

$$\begin{array}{lcll} \alpha + L & \rightleftharpoons & L:\alpha & K_3'', \\ L:\alpha + \gamma & \rightleftharpoons & L:\gamma:\alpha & K_2'', \\ \gamma + L & \rightleftharpoons & L:\gamma & K_3', \\ L:\gamma + \alpha & \rightleftharpoons & L:\gamma:\alpha & K_2'. \end{array}$$

A positive detailed-balanced steady state only exists if we enforce the following constraint on the product of affinity constants:

$$K_3' K_2' = K_3'' K_2''. \tag{4.155}$$

Combining steady state and conservation equations, we obtain the following polynomial system describing the model BC:

$$\begin{aligned} 0 &= -N_x + (1 + K_3' L)x + K_3' K_2' L x y, \\ 0 &= -N_y + (1 + K_3'' L)y + K_3' K_2' L x y. \end{aligned} \tag{4.156}$$

This polynomial system can be reorganised (by hand or by computing a Gröbner basis) into the following triangular system:

$$\begin{aligned} 0 &= -(1+K_3'L)N_y + [1+K_3''L \\ &\quad + K_3'L(1+K_3''L+K_2'(N_x-N_y))]y + K_2'K_3'L(1+K_3''L)y^2, \end{aligned} \tag{4.157a}$$

$$0 = -N_x + N_y + x(1+K_3'L) - y(1+K_3''L). \tag{4.157b}$$

Solving polynomial (4.157a) and selecting the positive solution, we obtain an expression for the number of unbound α chains per cell at steady state y:

$$y = \frac{-1-K_3''L-K_3'L(1+K_3''L+K_2'(N_x-N_y))+\sqrt{\Delta}}{2K_2'K_3'L(1+K_3''L)}, \tag{4.158}$$

where

$$\Delta = [1+K_3''L+K_3'L(1+K_3''L+K_2'(N_x-N_y))]^2+4K_2'K_3'L(1+K_3''L)(1+K_3'L)N_y.$$

The number of unbound γ chains at steady state is given by

$$x = \frac{N_x - N_y + (1+K_3''L)y}{1+K_3'L}. \tag{4.159}$$

The number of signalling complexes $L:\gamma:\alpha$ as a function of the ligand dose defines the signalling function σ:

$$\sigma(L) \equiv K_3'K_2'Lxy. \tag{4.160}$$

One can notice that

$$\begin{aligned} K_3'K_2'Lxy &= \frac{K_2'K_3'L(1+K_3''L)y^2 + K_3'K_2'Ly(N_x-N_y)}{1+K_3'L} \\ &= \frac{(1+K_3'L)N_y - (1+K_3'L)(1+K_3''L)y}{1+K_3'L}. \end{aligned}$$

The last step comes from polynomial (4.140a). Hence, σ can be re-written in a

simpler expression:

$$\sigma(L) = N_y - (1 + K_3'' L)y. \tag{4.161}$$

Computation of the amplitude: From equations (4.156), one can notice that $\lim_{L\to+\infty} \sigma(L) = 0$ so the amplitude is not the limit of the signalling function at high concentration. A short numerical simulation shows that the dose-response curve is bell-shaped. This implies that the analytical expression of the amplitude has to be computed through the derivative by finding L_a such that

$$\frac{d\sigma}{dL}(L_a) = 0, \tag{4.162}$$

and finally the amplitude is defined as

$$A^{het_{BC}} \equiv \sigma(L_a). \tag{4.163}$$

Unfortunately, Mathematica (Wolfram Research, Inc., 2019) cannot solve the equation $\frac{d\sigma}{dL}(L_a) = 0$ (mostly due to the fact that $\frac{d\sigma}{dL}$ is a long and complicated expression (see Appendix F)).

Computation of the EC_{50}: The EC_{50} computation can be reduced to the computation of zeros of a polynomial, by computing a Gröbner basis of the polynomial system (4.156) augmented with the equation

$$2K_3'K_2'Lxy - \mathcal{A} = 0, \tag{4.164}$$

where x, y, L are the variables. We wrote $\mathcal{A}$ the unknown amplitude of the model to shorten the notation. We obtain:

$$0 = L^2 + \frac{p}{2\mathcal{A}K_3'K_3''}L + \frac{1}{K_3'K_3''}, \tag{4.165a}$$

$$0 = y + \frac{\mathcal{A}K_3''}{K_2'(2N_x - \mathcal{A})}L + \mathcal{A}\left(\frac{1}{2} + \frac{K_3''}{K_2'K_3'(2N_x - \mathcal{A})}\right) - N_y, \tag{4.165b}$$

$$0 = x + \frac{\mathcal{A}K_3''}{K_2'(2N_y - \mathcal{A})}L + \mathcal{A}\left(\frac{1}{2} + \frac{1}{K_2'(2N_y - \mathcal{A})}\right) - N_x, \tag{4.165c}$$

where

$$p = K_2' K_3' (\mathcal{A} - 2N_y)(2N_x - \mathcal{A}) + 2\mathcal{A}(K_3' + K_3'').$$

Positive real roots of polynomial (4.165a) that lead to y and x (obtained by solving polynomials (4.165b) and (4.165c), respectively) positive are EC_{50} and IC_{50} expressions of the model. Since $\sigma(0) = 0$ and $\lim_{L \to +\infty} \sigma(L) = 0$, we expect an even number of solutions of the equation

$$\sigma(L^*) = \frac{\mathcal{A}}{2}.$$

Since polynomial (4.165a) is of degree 2, it has no more than two roots. Thus, p must be negative so the two roots are positive real numbers. The smallest root corresponds to the EC_{50} and the largest root is the IC_{50} of model heterodimeric RTK BC. We obtain these expressions as functions of the unknown amplitude $\mathcal{A}$ (but which can easily be computed numerically):

$$\begin{aligned} EC_{50}^{het_{BC}} &= \frac{-p - \sqrt{p^2 - 16\mathcal{A}^2 K_3' K_3''}}{4\mathcal{A} K_3' K_3''}, \\ IC_{50}^{het_{BC}} &= \frac{-p + \sqrt{p^2 - 16\mathcal{A}^2 K_3' K_3''}}{4\mathcal{A} K_3' K_3''}, \end{aligned} \tag{4.166}$$

where p was previously defined as

$$p = K_2' K_3' (\mathcal{A} - 2N_y)(2N_x - \mathcal{A}) + 2\mathcal{A}(K_3' + K_3'').$$

Attempt to refine the upper bound of the amplitude: By definition, the amplitude is a positive number lower than N_x and N_y. Since p must be negative, the amplitude must range in interval

$$A^{het_{BC}} \in ((0, A^-) \cup (A^+, +\infty)) \cap (0, \min(N_x, N_y)]$$

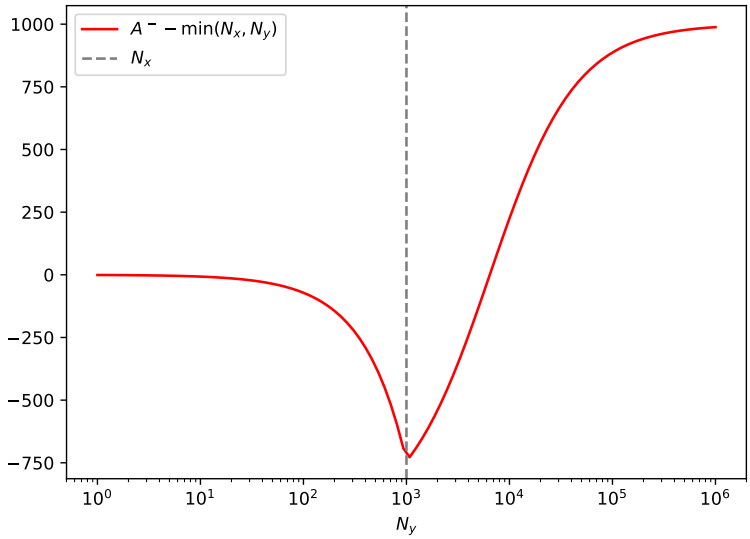


Figure 4.16: Plot of $N_y \mapsto A^- - \min(N_x, N_y)$ for $N_x = 10^3$, $K_2' = 17 \times 10^{-3}$,$K_3' = 34 \times 10^{10}\mathrm{M}^{-1}$, $K_3'' = 10^2 \times K_3'$. The function does not have a constant sign.

where A^+ and A^- are the roots of the polynomial of degree 2 with a negative leading coefficient defined by p:

$$\begin{aligned} A^+ &= \frac{K_3' + K_3'' + K_2'K_3'(N_x + N_y) + \sqrt{\Delta_{BC}}}{K_2'K_3'}, \\ A^- &= \frac{K_3' + K_3'' + K_2'K_3'(N_x + N_y) - \sqrt{\Delta_{BC}}}{K_2'K_3'}, \end{aligned} \tag{4.167}$$

where

$$\Delta_{BC} = (K_3' + K_3'' + K_2'K_3'(N_x + N_y))^2 - 4K_2'^2K_3'^2N_xN_y.$$

We have $A^+ > N_x$ and $A^+ > N_y$ so the amplitude must range in $(0, A^-) \cap (0, \min(N_x, N_y)]$. However, the sign of $A^- - \min(N_x, N_y)$ is not constant (see Figure 4.16) so we cannot refine the interval of the amplitude.

Heterodimeric RTK ABC

Finally, consider now the case in which the signalling complex can be assembled in three different ways, following the formation orders of models heterodimeric RTK

A, B and C. The chemical reaction scheme of model heterodimeric RTK ABC is:

$$\begin{array}{lcll} \gamma+\alpha & \rightleftharpoons & \gamma:\alpha & K_2, \\ L+\gamma:\alpha & \rightleftharpoons & L:\gamma:\alpha & K_3, \\ \gamma+L & \rightleftharpoons & L:\gamma & K_3', \\ L:\gamma+\alpha & \rightleftharpoons & L:\gamma:\alpha & K_2', \\ \alpha+L & \rightleftharpoons & L:\alpha & K_3'', \\ L:\alpha+\gamma & \rightleftharpoons & L:\gamma:\alpha & K_2''. \end{array}$$

This model presents two thermodynamic cycles and the existence of a positive detailed-balanced steady state is conditioned by the following constraints on the product of the affinity constants:

$$K_3K_2 = K_3'K_2' = K_3''K_2''. \tag{4.168}$$

Model heterodimeric RTK ABC is described by the following polynomial system:

$$\begin{aligned} 0 &= -N_x + (1+K_3'L)x + K_2xy + K_3K_2Lxy, \\ 0 &= -N_y + (1+K_3''L)y + K_2xy + K_3K_2Lxy. \end{aligned} \tag{4.169}$$

Reorganising this polynomial system by hand or computing a Gröbner basis, the system (4.169) can be re-written as a triangular system:

$$\begin{aligned} 0 &= -(1+K_3'L)N_y + [1 + K_3'L + K_3''L + K_3'K_3''L^2 \\ &\quad + K_2(1+K_3L)(N_x - N_y)]y + K_2(1+K_3L)(1+K_3''L)y^2, \\ 0 &= -N_x + N_y + (1+K_3'L)x - (1+K_3''L)y. \end{aligned} \tag{4.170}$$

We solve these polynomials and select the positive solution for y to obtain an expression for the numbers of unbound α chains and γ chains at steady state (y and x, respectively):

$$\begin{aligned} y &= \frac{-1 - K_3'L - K_3''L - K_3'K_3''L^2 - K_2(1+K_3L)(N_x - N_y) + \sqrt{\Delta_{ABC}}}{2K_2(1+K_3L)(1+K_3''L)}, \\ x &= \frac{N_x - N_y + (1+K_3''L)y}{1+K_3'L}, \end{aligned} \tag{4.171}$$

where

$$\begin{aligned}\Delta_{ABC} = &[1 + K_3'L + K_3''L + K_3'K_3''L^2 + K_2(1 + K_3L)(N_x - N_y)]^2 \\ &+ 4K_2(1 + K_3L)(1 + K_3''L)(1 + K_3'L)N_y.\end{aligned}$$

Computation of the amplitude: The number of signalling complexes $L : \alpha : \gamma$ as a function of the ligand concentration defines the signalling function:

$$\sigma(L) \equiv K_3K_2Lxy. \tag{4.172}$$

From equations (4.169), we see that $\lim_{L\to+\infty} \sigma(L) = 0$ (in fact, a short numerical simulation shows that the dose-response curve is bell-shaped), thus the amplitude has to be calculated by computing the derivative of the signalling function. Unfortunately, once again, the derivative of σ is an over-complex expression which is not useful for further analytic computation. However, since we obtained an analytic expression of σ, we can compute its maximum (the amplitude) numerically very easily.

Computation of the EC_{50}: If we write $\mathcal{A}$ the unknown amplitude of model ABC (which can be obtained numerically), one can reduce the computation of the EC_{50} and the IC_{50} to solving a polynomial of degree 3. Indeed, by computing a Gröbner basis to the polynomial system (4.169) augmented with the equation

$$2K_3K_3Lxy - \mathcal{A}, \tag{4.173}$$

where we considered x, y and L as variables, we obtain the following new polynomial system:

$$0 = L^3 + \frac{p}{2\mathcal{A}K_3'K_3''}L^2 + \frac{1 + K_2(N_x + N_y - \mathcal{A})}{K_3'K_3''}L - \frac{\mathcal{A}K_2}{2K_3K_3'K_3''}, \tag{4.174a}$$

$$0 = y + f(N_x, K_3'')L^2 + \frac{q}{2\mathcal{A}K_2(\mathcal{A}(K_3 - K_3'') - 2K_3N_x)}L + g(N_x, K_3''), \tag{4.174b}$$

$$0 = x + f(N_y, K_3')L^2 + \frac{q - 2K_3(N_y - N_x)p}{2\mathcal{A}K_2(\mathcal{A}(K_3 - K_3') - 2K_3N_y)}L + g(N_y, K_3'). \tag{4.174c}$$

We wrote

$$\begin{aligned} p &= -\mathcal{A}^2 K_2 K_3 - 4K_2 K_3 N_x N_y + 2\mathcal{A}(K_3' + K_3'' + K_2 K_3(N_x + N_y)), \\ q &= -K_2 K_3^2 \mathcal{A}^3 + 2(-K_3' K_3'' + K_3 K_3' + K_3 K_3'' + K_2 K_3^2(2N_x + N_y))\mathcal{A}^2 \\ &\quad - 4K_3 N_x (K_3' + K_3'' + K_2 K_3(N_x + 2N_y))\mathcal{A} + 8K_2 K_3^2 N_x^2 N_y, \end{aligned}$$

and defined the functions:

$$\begin{aligned} f :(N, K) &\mapsto \frac{K_3 K_3' K_3''(\mathcal{A} - 2N)}{K_2(\mathcal{A}(K_3 - K) - 2K_3 N)}, \\ g :(N, K) &\mapsto \frac{\mathcal{A}^2 K_2 K_3 + 4K_3 N(1 + K_2 N) - 2\mathcal{A}(K_3 - K + 2K_2 K_3 N)}{2K_2(\mathcal{A}(K - K_3) + 2K_3 N)}. \end{aligned}$$

The signs of the coefficients of polynomial (4.174a) by decreasing degree are $(-, ?, +, -)$ as the sign of p seems unclear *a priori*. According to Descartes' rule of sign, it means that polynomial (4.174a) has one or three positive real roots. However since the dose-response curve reaches 0 for $L = 0$ and $L \to +\infty$, the equation $\sigma(L^*) = \dfrac{\mathcal{A}}{2}$ must have an even number of solutions. It results that $p < 0$ and polynomial (4.174a) admits two positive real roots which lead to y and x (obtained by solving polynomials (4.174b) and (4.174c), respectively) positive. The lowest root is the expression for the EC_{50} and the larger root is the expression for the IC_{50} of model ABC. Though we can use Cardano's formulæ to obtain analytic expressions of these roots, the resulting expressions are complex and long, which is not useful for further analysis.

Alternatively, since we obtained an analytic expression for σ, the EC_{50} and IC_{50} can be found by numerically solving $\sigma(L^*) = \dfrac{\mathcal{A}}{2}$.

Attempt to refine the upper bound of the amplitude: Similar to the heterodimeric model BC, as $p < 0$, the amplitude must range in the interval $(0, \min(N_x, N_y)] \cup (0, A^-) \cup (A^+, +\infty)$, where

$$\begin{aligned} A^+ &= \frac{K_3' + K_3'' + K_2 K_3(N_x + N_y) + \sqrt{\Delta_{BC}}}{K_2 K_3}, \\ A^- &= \frac{K_3' + K_3'' + K_2 K_3(N_x + N_y) + \sqrt{\Delta_{BC}}}{K_2 K_3}, \end{aligned} \tag{4.175}$$

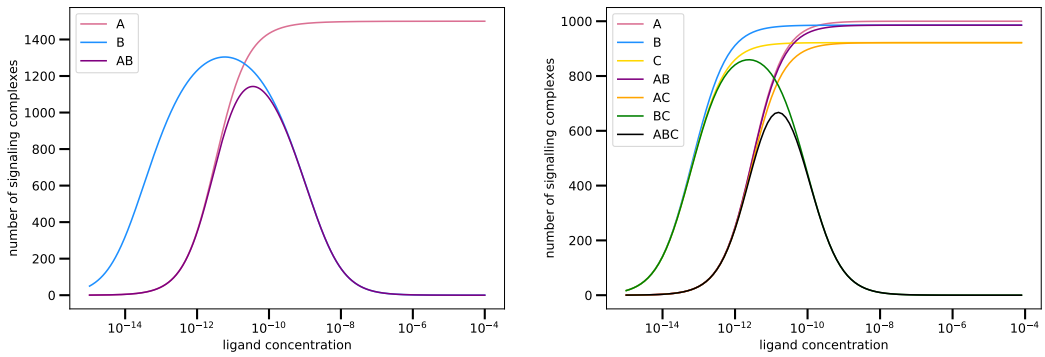
and $\Delta_{BC} = (K_3' + K_3'' + K_2K_3(N_x + N_y))^2 - 4K_2^2K_3^2N_xN_y$ have been defined in the previous section (we remind the reader that we have $K_2K_3 = K_2'K_3'$). Since $A^+ - \min(N_x, N_y) > 0$,the amplitude ranges in the interval $(0, \min(N_x, N_y)] \cap (0, A^-)$. We demonstrated in the previous section that the function $N_y \mapsto A^- - \min(N_x, N_y)$ does not have a constant sign (see Figure 4.16) and the amplitude upper bound cannot be refined.

4.5.4 Discussion

In the previous sections, I attempted to compute the analytic expression of the amplitude and EC_{50} of homodimeric and heterodimeric RTK models, ranging from relatively simple systems to more complex models with several thermodynamic cycles. Such expressions could not be obtained for most complex models as new complications arose (such as non-sigmoidal dose-response curves, see Figure 4.17). I then sought to find tighter bounds for the amplitude of complex models, with very little success. The analytic expressions that I managed to obtained were not as interesting as I expected since they were not really comparable. Nonetheless, I showed that the amplitude of heterodimeric models AB and AC were equal to the amplitude of models B and C, respectively. To gain further insights, I conduct a comprehensive numerical exploration which compares the pharmacological quantities of the different models of this section for a broad range of parameter values.

Comparing pharmacological quantities:

Plotting the dose-response curves of the different models studied in this section yields the following observation (see Figure 4.17 for a fixed set of parameter values): the amplitude of a combined model seems lower than the amplitudes of the models the combined model stems from, both in the heterodimeric and homodimeric cases. We also observe that the EC_{50} of model AB (resp. AC) seems to range between the EC_{50} of models A and B (resp. A and C). However, the value of the EC_{50} of heterodimeric model BC does not seem to fall between the EC_{50} of models B and C. To numerically investigate these rough observations, I



(a) Homodimeric RTK models

(b) Heterodimeric RTK models

Figure 4.17: Dose response curves for homodimeric (a) and heterodimeric (b) RTK models for the following parameters: $K_2 = 17 \times 10^{-3}$, $K_3 = 34 \times 10^{10}\text{M}^{-1}$, $K_3' = \dfrac{K_3}{2}$, $K_3'' = 3 \times K_3$, $N_x = 3 \times 10^3$ and $N_y = 10^3$.

implemented the following algorithm that I describe in the heterodimeric case which is the most general configuration:

1. Define ranges for the parameters involved N_x, N_y, $\lambda = \dfrac{K_3'}{K_3}$ and $\mu = \dfrac{K_3''}{K_3}$. We call L_x, L_y, L_λ and L_μ the list of values of the parameters N_x, N_y, λ and μ, respectively.

2. For each model X, $X = A, B, C, AB, AC, BC$ or ABC, define the multi-dimensional arrays Amp^X, EC^X and IC^X (when relevant) such that the element Amp^X_{ijkl} (respectively, EC^X_{ijkl}, IC^X_{ijkl}) is the amplitude (respectively, the EC_{50}, the IC_{50}) of model X when $N_x = L_x[i]$, $N_y = L_y[j]$, $\lambda = L_\lambda[k]$ and $\mu = L_\mu[l]$, where we wrote $L[p]$ the p-th element of the list L.

3. We define a model order $[A, B, AB, AC, BC, ABC]$ and the model X_p will define the p-th model of this ordered list. We then compare each array Amp^{X_p} (respectively, EC^{X_p}, IC^{X_p}) with each other, element by element, and create the matrix M^{amp} (respectively, M^{ec}, M^{ic}) as follows. The element M^{amp}_{pq}(row p, column q) (resp. M^{ec}_{pq}, M^{ic}_{pq}) is equal to:

   - 1 if for all i, j, k, l, $Amp^{X_p}_{ijkl} - Amp^{X_q}_{ijkl} > 0$ (respectively, $EC^{X_p}_{ijkl} - EC^{X_q}_{ijkl} > 0$, $IC^{X_p}_{ijkl} - iC^{X_q}_{ijkl} > 0$)

- -1 if for all i, j, k, l, $Amp_{ijkl}^{X_p} - Amp_{ijkl}^{X_q} < 0$ (respectively, $EC_{ijkl}^{X_p} - EC_{ijkl}^{X_q} < 0$, $IC_{ijkl}^{X_p} - iC_{ijkl}^{X_q} < 0$)

- 0 if for all i, j, k, l, $Amp_{ijkl}^{X_p} - Amp_{ijkl}^{X_q} \approx 0$ (respectively, $EC_{ijkl}^{X_p} - EC_{ijkl}^{X_q} \approx 0$, $IC_{ijkl}^{X_p} - iC_{ijkl}^{X_q} \approx 0$). Note that we say that a quantity $Q - P$ is ≈ 0 if $|Q - P| < atol$ and $\dfrac{|Q - P|}{|Q|} < rtol$ where $atol$ (resp. $rtol$) is a given absolute (resp. relative) tolerance.

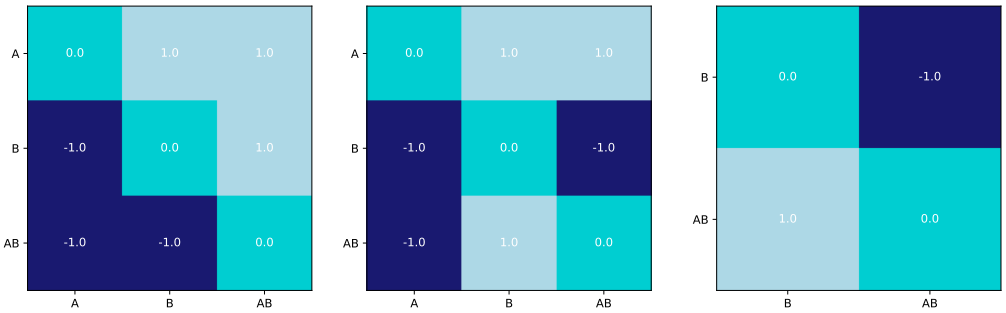
- 3 if none of the previous statements is true.

The diagonal of the matrices M will always be 0. For instance, the element M_{12}^{amp} will be equal to 1 if the amplitude of the model A is greater than the amplitude of the model B for all the parameter values $L_x, L_y, L_\mu, L_\lambda$.

We vary the parameters (N_x and λ for the homodimeric case, N_x, N_y, λ and μ for the heterodimeric case) in the following ranges: N_x and N_y in the interval $[10, 10^6]$, λ and μ in the interval $[10^{-2}, 10^2]$. The values of the absolute and relative tolerances were determined after typical values for the amplitude, EC_{50} and IC_{50}: $atol = 10^{-3}$M for the amplitude, $atol = 10^{-17}$M for the EC_{50} and $atol = 10^{-16}$M for the IC_{50}. For the three quantities, $rtol = 10^{-3}$.

We show the matrices M for the homodimeric case in figure 4.18 and the results for the heterodimeric cases in figure 4.19. In the homodimeric case, we observe:

$$\begin{aligned} A^{ho_{AB}} &< A^{ho_A}, A^{ho_B}, \\ EC_{50}^{ho_A} &> EC_{50}^{ho_{AB}} > EC_{50}^{ho_B}, \\ IC_{50}^{ho_{AB}} &> IC_{50}^{ho_B}. \end{aligned} \tag{4.176}$$



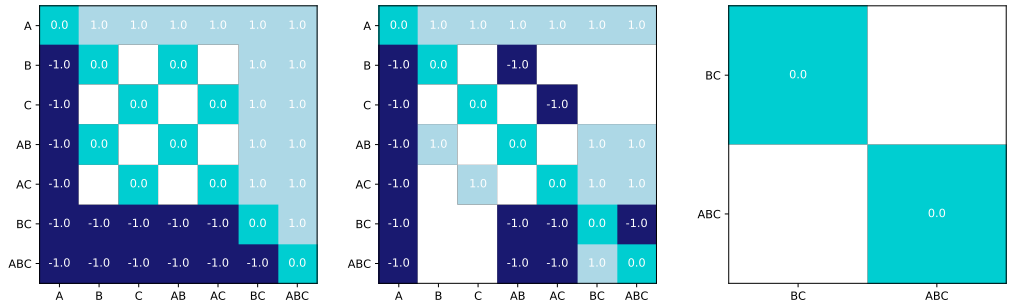
(a) Comparison of amplitudes (b) Comparison of EC_{50}s (c) Comparison of IC_{50}s

Figure 4.18: Amplitude (a), EC_{50} (b) and IC_{50} (c) comparison matrices for the homodimeric RTK models. A case is light blue (value 1) (resp. dark blue (value -1)) if the quantity (amplitude or EC_{50} or IC_{50}) of the model of the row is greater (resp. lower) than the quantity of the model of the column for all the parameter values tested. The case is turquoise (value 0) if the quantity of the model of the row and the model of the column are equal. The parameters chosen were $K_2 = 17 \times 10^{-3}$, $K_3 = 34 \times 10^{10}\mathrm{M}^{-1}$.

In the heterodimeric case, we observe:

$$\begin{aligned}
A^{het_{C+B}} &< A^{het_C}, A^{het_B}, \\
A^{het_{AB}} = A^{het_B} &< A^{het_A}, \\
A^{het_{AC}} = A^{het_C} &< A^{het_A}, \\
A^{het_{ABC}} &< A^{het_{BC}}, A^{het_{AB}}, A^{het_{AC}}, A^{het_A}, A^{het_B}, A^{het_C}, \\
EC_{50}^{het_A} &> EC_{50}^{het_{AB}} > EC_{50}^{het_B}, \\
EC_{50}^{het_A} &> EC_{50}^{het_{AC}} > EC_{50}^{het_C}, \\
EC_{50}^{het_{AB}}, EC_{50}^{het_{AC}} &> EC_{50}^{het_{ABC}} > EC_{50}^{het_{BC}}.
\end{aligned} \tag{4.177}$$

In summary, as hinted, we observe that the amplitude of a combined model is lower than the amplitude of its submodels. Unfortunately, we could not prove it analytically. From a biological point of view, this observation is not surprising. Indeed, in combined models, we allow the formation of intermediate complexes composed of a chain (present in limited quantity) and the ligand (which is in excess), such as $L : \gamma$ or $L : \alpha$. At any concentration of ligand, such intermediate complexes are formed but not enough chains are present to complete them into



(a) Comparison of amplitudes (b) Comparison of EC_{50}s (c) Comparison of IC_{50}s

Figure 4.19: Amplitude (a), EC_{50} (b) and IC_{50} (c) comparison matrices for the heterodimeric RTK models. A case is light blue (value 1) (resp. dark blue (value -1)) if the quantity (amplitude or EC_{50} or IC_{50}) of the model of the row is greater (resp. lower) than the quantity of the model of the column for all the parameter values tested. The case is turquoise (value 0) if the quantity of the model of the row and the model of the column are equal. The case is blank if no pattern has been detected (dummy value 3 in the algorithm description). The parameters chosen were $K_2 = 17 \times 10^{-3}$, $K_3 = 34 \times 10^{10}\mathrm{M}^{-1}$.

signalling complexes. Thus, we always form fewer signalling complexes in a combined model than in a base model. Our rough observation on EC_{50} seems also confirmed but, once again, we do not have an analytic proof.

Conclusions

Overall, in this section, I computed analytic expressions for the amplitude and EC_{50} of some of the models, which thus reduces computational cost and shows parameter dependency. However, this section was mostly unsuccessful to solve the initial issue. Apart from the amplitudes of heterodimeric models AB and AC which were equal to the amplitudes of models B and C, respectively, we were unable to compare (analytically) the combined model quantities to their sub-models attributes. The reasons are numerous: we could not always compute the amplitude or EC_{50} expressions because of the complexity of the model or because the dose-response was not sigmoidal (thus, we could not make the useful approximation that the limit of the signalling function is the amplitude). Finally,

as already investigated in CRNT, combining models together is not a trivial task as it is almost unable to conserve basic properties of chemical reaction networks (Gross *et al.*, 2020). Combining simple models in the hope of noticing a relationship between the amplitude or EC_{50} of the combined and base models was rather optimistic.

4.6 Summary and discussion

First illustrated by two examples in Chapter 3, I proposed, in this chapter, a method to compute analytic expressions for two relevant pharmacodynamic metrics, the amplitude and the EC_{50}, for receptor-ligand systems. The method starts with the computation of a Gröbner basis for the polynomial system of the receptor-ligand system in steady state. As shown in the different examples of this chapter (and previous chapter), the derivation of the amplitude is easier when the maximum of the dose-response curve is attained at large ligand concentration (for instance when the dose-response curve is a sigmoid). In that case, the amplitude is the limit of the signalling function when the ligand concentration tends to infinity. When the model is simple enough, the polynomial system, simplified by the computation of the Gröbner basis, can be solved iteratively to obtain an analytic expression for the steady state. From these expressions, it is then relatively straightforward to compute the amplitude (*i.e.*, the limit of the signalling function at large values of the ligand concentration) and the EC_{50}. For more complex models, getting such steady state expressions can be more challenging. In some cases (see Sections 3.4 and 4.2.6), perturbation theory can be used to derive the expression for the amplitude. Computing another Gröbner basis can dramatically simplify the calculation of the EC_{50}, and in turn display how it depends on the parameters of the model.

However, solving the computed Gröbner bases is not always possible as one may obtain large degree polynomials, depending on the systems' complexity. I tried to circumvent this issue, in Section 4.5, by investigating the relationship between the amplitude and EC_{50} of models and their sub-models. This work introduced a lot of new challenges such as the computation of amplitudes when the maximum response is not the asymptotic behaviour of the dose-response curve. For instance,

the computation of the maximum for bell-shaped dose-response curves (which has been done for simple models in Refs. Douglass Jr *et al.* (2013); Mack *et al.* (2008)) may involve the computation of the derivative of the signalling function, which can be laborious even with the use of symbolic software, as demonstrated in Section 4.5. Finally, the method of Section 4.1 often requires additional mathematical tools or knowledge, such as perturbation theory, which makes it rather a challenge to be used by those who are not mathematically trained. In spite of the (sometimes, complicated) calculations that the method requires, I believe that analytic expressions of the pharmacological metrics characterising simple receptor-ligand systems may provide significant advantages when studying such biological systems.

I also introduced a family of receptor-ligand systems, called SRLK, in which the signalling complex, composed of a kinase, a ligand and n trans-membrane sub-unit chains, is built sequentially. These models could also form "dummy" and decoy complexes, similarly to the IL-7R models of Chapter 3, which the SRLK family encompasses. By manipulating the polynomials describing the SRLK models, I was able to derive an analytic expression of the amplitude under the no allostery assumption. I also showed that the maximum of the dose-response curve for both the IL-7R models of Chapter 3 (and the trimeric receptor model of Section 4.2.6) was indeed the amplitude of the models. Despite relatively strong assumptions, I believe that the SRLK approach can be used to model a broad range of biochemical systems, such as receptor competition in interleukin signalling. The analytic expressions obtained for the amplitude could improve our understanding of biological mechanisms requiring a fine tuning of cytokine signalling such as cancer treatment (Spolski *et al.*, 2017) or cytokine storm control (Fajgenbaum & June, 2020; Savarin & Bergmann, 2018). I showed in Section 4.3.4 how the SRLK models can account for the competition for the gamma chain between the IL-2 family of receptors and the competition for receptor components between the IL-12 family of receptors. However, many receptors signal through different configurations. IL-35, for instance, can signal through homodimerisation of gp130 or IL-12Rβ2 (Collison *et al.*, 2012). It has been shown that IL-6, a cytokine implied in cytokine storms (Chen *et al.*, 2021; Fajgenbaum & June, 2020), signals through an hexameric structure composed of two IL-6Rα chains and two gp130

molecules (Boulanger *et al.*, 2003). Furthermore, it seems that the ligand IL-6 first binds to the IL-6Rα chain before any association with gp130 (Boulanger *et al.*, 2003). Thus, one could imagine other general receptor models that may involve any of the following: 1) homo-oligomerisation (when two trans-membrane chains X_i are identical), 2) other orders of signalling complex formation (non-sequential orders or for instance, if the ligand is not the final sub-unit to be bound), 3) multiple kinases (including kinases binding to other sub-unit chains, such as JAK1 which binds to IL-7Rα (Park *et al.*, 2019)), or 4) a more detailed JAK-STAT pathway (most cytokine receptors activate multiple STAT molecules, whose copy numbers tune the immune response elicited (Lin & Leonard, 2019)).

Conclusion of Chapters 3 and 4

In the current and prior chapters, I computed the analytic expression of the amplitude and EC_{50} for specific receptor-ligand models, but also for general families of such systems, such as SRLK models in Section 4.3. Analytic expressions for the amplitude and the EC_{50} offer mechanistic insight for the receptor-ligand systems under consideration, allow one to quantify the parameter dependency of these two key variables, and can facilitate model validation and parameter exploration as illustrated in great detail in Chapter 3. This work demonstrated that, while being commonly considered as univariate functions of the ligand concentration, the amplitude and EC_{50} are, instead, multivariate functions, highly depending on the abundances of every protein composing the signalling receptor, and even, sometimes, protein abundances of other receptors. The results of this chapter also showed that the receptor architecture can have a great influence on these two key pharmacological quantities. Indeed, we demonstrated that the quantitative effects of receptor chain upregulation can be vastly different, depending on the elements of a receptor's signalling core. This study provides a theoretical and quantitative framework with which to interpret the potential functional significance of receptor up/downregulation during lymphocyte differentiation (Kalia *et al.*, 2010; Voisinne *et al.*, 2015), oncogenesis (Du & Lovly, 2018) or drug treatment (Vogel *et al.*, 2016).

With this work I hope to have initiated, or renewed, an interest for the algebraic

analysis of receptor-ligand systems. Finally, I believe the results presented in these chapters are a first step to account for the variability of receptor expression levels when designing and studying receptor-ligand models (both from an experimental and mathematical perspective) (Cotari *et al.*, 2013b; Farhat *et al.*, 2021; Gonnord *et al.*, 2018; Ring *et al.*, 2012).

Chapter 5

Agent-based models of the competition for IL-2 between T regs and self-activated T cells

5.1 Introduction

The immune system is a "team effort" which involves many different cells working together to ensure proper defence against pathogens (Sompayrac, 2019). In homeostasis (in absence of pathogens/threat), key cell populations must be maintained to ensure a prompt and appropriate immune response when under attack. This means being able to identify invaders (*i.e.*, discriminate between proteins present naturally in the body (self-peptides) and non-self proteins) and scale the response to the pathogen type and dose. The maintenance of an effective immune system, or its response to a pathogen, results from the coordination of each cell's behaviour, which depends on the cell's local environmental conditions (protein stimuli, space for division, ...) and cell's attributes (receptor expression level, receptor structure, exhaustion level, ...). The previous two chapters have focused on understanding how a cell's response to a given ligand depends on the expression level of its receptor. I demonstrated that, depending on the receptor structure, the individual cell's response can greatly vary (and sometimes in a non-intuitive manner) when increasing the cell's receptor (or receptor chain) expression levels. This variability in an individual cell's response may modulate the overall cell population dynamics

and, thus, be an essential mechanism that gives the immune system its flexibility. For instance, the competition for IL-2 between cells that express various levels of IL-2R is one of the key factors to switch from homeostasis (the resting state in the absence of pathogen) or a local inflammation, to a more global immune response (Amado *et al.*, 2013; Oyler-Yaniv *et al.*, 2017). This competition is also essential, at homeostasis, to maintain an effective immune system. Let me explain this mechanism in greater detail.

As mentioned in Chapter 1, T cells express special receptors, called T cell receptors (TCRs), which can bind to bits of non-self proteins (foreign peptides), thus recognising a specific pathogen. All the TCRs expressed by one cell are identical. The body uses quasi-random somatic gene recombination to generate a large number of T cells, each with a different TCR structure, to be able to recognise the multitude of pathogens that naturally exist. However, this genetic recombination does not always lead to efficient cells. Before being released in the organs and lymph nodes (which we refer to as periphery), T cells mature in the thymus where their TCRs are tested against a variety of self-peptides. Cells with TCRs which fail to react to these peptides, or which react too strongly to the test proteins, undergo apoptosis (assisted suicide) (McCaughtry & Hogquist, 2008). Cells with TCRs of intermediate binding affinity will survive, pursue their development and eventually be released in the periphery (see Figure 5.1). However, not all peptides that exist in the body (self-peptides) are present in the thymus. Consequently, some cells, particularly sensitive to self-peptides, escape the thymic selection and move to the periphery (Bouneaud *et al.*, 2000; Zehn & Bevan, 2006). These cells can become activated in absence of pathogens, triggered by the binding of self-peptides to their TCRs: this mechanism is called *autoimmunity*. In absence of regulation, these *self-activated* (or *self-reactive*) T cells can proliferate and trigger an immune response: they try to eliminate the self-peptide, causing pathological and/or functional damage to the organ or tissue containing the target protein. This provokes clinical symptoms and is referred to as *autoimmune disease*. Some autoimmune diseases are very common (Davidson & Diamond, 2001; Marrack *et al.*, 2001), such as type-1 diabetes (Atkinson *et al.*, 2014), rheumatoid arthritis (Weyand & Goronzy, 2021) or psoriasis (Bowcock &

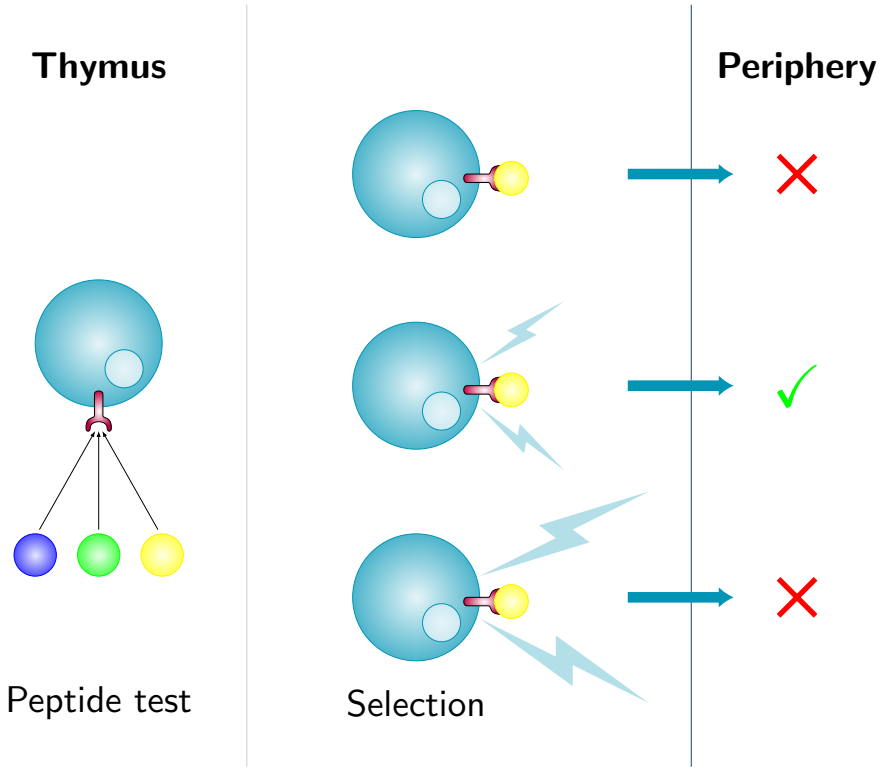


Figure 5.1: T cell selection in the thymus: cells that react too poorly or too strongly to self-peptides undergo apoptosis. Only cells that bind to self-peptides with intermediate affinity are released in the periphery.

Krueger, 2005), and while most are not fatal (mainly thanks to treatments), they might cause major disruptions in the host's life. Other autoimmune diseases, such as giant cell myocarditis (Rosenstein *et al.*, 2000), can be lethal. Developing new treatments for these conditions requires to understand the underlying mechanisms of immune diseases, which necessitates biological experiments and the development of mathematical models.

In healthy vertebrates, the regulation of self-activated T cells operates in lymph nodes and spleen (both places are referred to as secondary lymphoid organ) which contain specialised T cells, called *regulatory T cells* (T regs)[1]. These cells are mainly defined by the expression of the forkhead box P3 transcription factor (Foxp3) and are essential to maintain immunological homeostasis: the

[1]In the rest of this thesis, we refer to as conventional T cell, any T cell that is not a regulatory T cell.

absence of T regs precipitates T cell-mediated autoimmune diseases in mice and humans (Kim *et al.*, 2007a; Lahl *et al.*, 2007; Sakaguchi *et al.*, 1995). T regs depend on interleukin-2 (IL-2) for their survival (Abbas *et al.*, 2018; Liu *et al.*, 2015; Owen *et al.*, 2018) and, as expression of Foxp3 prevents the production of IL-2 (Amado *et al.*, 2013), they rely on IL-2 producing cells, such as self-reactive conventional T cells, to maintain their population at homeostasis (Hemmers *et al.*, 2019; Lee *et al.*, 2012; Owen *et al.*, 2018). It is believed that, in a healthy individual, the competition for IL-2 between self-activated T cells and T regs regulates the two cell populations at homeostasis (Busse *et al.*, 2010; Feinerman *et al.*, 2010; Höfer *et al.*, 2012): T regs prevent the inflammation from spreading in the organ by controlling the diffusion of IL-2 produced by self-reactive cells (Oyler-Yaniv *et al.*, 2017), and thus limit the proliferation of such cells; the production of IL-2 maintains the population of T regs at physiological levels which prevents autoimmune diseases. This competition is reinforced by the IL-2/IL-2R feedback dynamics: the binding of IL-2 to its receptor induces the upregulation of IL-2Rα expression and inhibits IL-2 secretion (for IL-2 producing cells) (Feinerman *et al.*, 2010; Kim *et al.*, 2006). The competition for IL-2 between regulatory and self-reactive conventional T cells is a nice and self-contained example of population dynamics (as IL-2 is produced and consumed by T cells). Many mathematical models have been proposed to describe this mechanism (Busse *et al.*, 2010; Feinerman *et al.*, 2010; Higuera *et al.*, 2017; Reynolds *et al.*, 2014; Voisinne *et al.*, 2015; Wong *et al.*, 2021), but the impact of cell-to-cell variability in receptor expression levels on the system was never the primary focus. In addition to this competition between two T cell populations, there exists a competition for IL-2 within the conventional T cell population (Höfer *et al.*, 2012). Conventional T cells that express more IL-2Rα signal stronger than others. It leads them to express even more IL-2Rα and thus achieve stronger signalling capabilities. At the same time, they produce less and less IL-2. A continuum level of IL-2Rα within the conventional T cell population may lead to a split between IL-2 consumers and producers (Höfer *et al.*, 2012).

In this chapter, I investigate how the heterogeneity in IL-2 receptor expression levels among the self-activated conventional T cell population affects the population dynamics, and how, in turn, the population dynamics (and in particular

cellular events such as death or division of a cell) impacts the IL-2R[1] distribution among the population. To take into account each cell's individuality, agent-based modelling is a suitable computational tool. However, agent-based models (ABMs) tend to be complex (the models can have a lot of parameters, be spatio-temporal and multi-scale, have multiple intricate rules...) and computationally intensive. Consequently, ABMs can be hard to study and thus have the reputation to be rather uninformative as they do not provide mechanistic insights. To alleviate the problem, I propose, in this chapter, to mathematically describe simple agent-based models of the competition for IL-2 within the activated conventional T cell population.
This chapter starts with a deterministic two-attribute system which describes the IL-2/IL-2R dynamics: a cell that consumes[2] IL-2, upregulates its IL-2R expression level. However, each cell is assumed to secrete IL-2 at a constant rate. I, first, analyse this two-attribute system in the case of a population with a constant number of cells. Later, stochastic cellular events, such as death, immigration/activation or division, will be added to construct more complicated agent-based models. A combination of mathematical analysis and numerical simulations is necessary to understand how the population dynamics and the receptor distribution among the conventional T cell population are related.

The models of this chapter describe interactions between events with timescales of days or weeks (death and division of cells, modelled as independent stochastic events) and receptor upregulation dynamics on timescales of minutes or hours (modelled deterministically with a pair of differential equations for each cell). The even faster timescale, of the release and re-absorption of IL-2 molecules, is not explicitly modelled but replaced by the assumption that, summed over activated conventional T cells, the total rate of IL-2 production is equal to the total rate of consumption. Further discussion of parameter values will be

[1] In this chapter, we will not distinguish between IL-2R and IL-2Rα, as IL-2R deprived of IL-2Rα has a low binding affinity to IL-2.

[2] In this chapter, I refer to as IL-2 absorption (or consumption) the whole (not modelled) process of IL-2/IL-2R binding, receptor internalisation, signal transduction and receptor recycling. We do not consider any cytokine or receptor degradation. Once ligand-bound, a receptor is immediately internalised and recycled at the surface for another ligand to bind to.

provided at the end of this chapter. A Python code of all the agent-based models of the competition for IL-2 discussed in this chapter is available at https://github.com/leasta/ABM_thesis (more information on how the models are simulated can be found in Appendix J).

5.2 Deterministic two-attribute dynamics

First, consider a population of N activated conventional T cells and assume this number, N, to be constant. We define, for each cell, $T \in \{1, \ldots, N\}$, two time-dependent attributes: the current IL-2R expression level of the cell at time t, denoted by $r_T(t)$, and the quantity of IL-2 the cell accumulated since it entered the cell pool (here, the beginning of the simulation, $t = 0$) up to time t, denoted $i_T(t)$. Thus, for each cell T, $i_T(t)$ is a monotonically increasing function of t. Note that the cell's attribute at time t, $r_T(t)$ (respectively, $i_T(t)$) is a real-valued quantity that is proportional to the number of receptors expressed (respectively, number of molecules of IL-2 absorbed) by cell T. We will sometimes refer to $r_T(t)$ as the number of receptors of the cell, by abuse of terminology[1]. The dimension of the attributes, $r_T(t)$ and $i_T(t)$ will be denoted by $[r_T]$ and $[i_T]$. Along with the time-dependent attributes, $i_T(t)$ and $r_T(t)$, associated with individual cells, we construct the population quantities $I(t)$ and $R(t)$ given by

$$I(t) = \sum_{T=1}^{N} i_T(t), \tag{5.1a}$$

$$R(t) = \sum_{T=1}^{N} r_T(t), \tag{5.1b}$$

which represent the total amount of IL-2 accumulated up to time t and total number of receptors expressed at time t by the entire cell population, respectively.

[1] In the simplest interpretation, the variables r_T and i_T are dimensionless because both are numbers of molecules. However, the justification for treating them as real numbers is not simply that they take large values, but also by analogy with experimental measurements. Flow cytometry assigns a fluorescence value to each cell that is proportional to the number of copies of the molecule in question on that cell's surface. Measurements of the number of molecules of a given type inside an individual cell are also indirect. Fortunately, because the differential equations (5.2) are linear in r_T and i_T, they hold regardless their dimension.

Each cell produces extra-cellular IL-2 with a constant rate p. That is, at each time, t, pN IL-2 is released in the extra-cellular medium for all the cells to compete for. We assume that each cell consumes IL-2 proportionally to its current IL-2R expression level, $r_T(t)$, and that the entirety of the pN IL-2 produced is consumed at each time t. Thus, at any time t, each cell, T, receives the fraction, $\frac{r_T(t)}{R(t)}$, of the pN extra-cellular IL-2 available. In addition, each cell T upregulates its IL-2R expression level proportionally to the quantity of IL-2 it accumulated up to time t ($i_T(t)$), with an upregulation rate u. Hence, the dynamics of the attributes of one cell T, $r_T(t)$ and $i_T(t)$. can be described as the following system of ODEs:

$$\frac{di_T}{dt} = pN\frac{r_T}{R}, \tag{5.2a}$$

$$\frac{dr_T}{dt} = ui_T. \tag{5.2b}$$

Note that by summing equation (5.2a) over the cell population, we recover the quasi steady state approximation that the total rate of IL-2 production (pN) is equal to the total rate of consumption:

$$\sum_T \frac{di_T}{dt} = pN. \tag{5.3}$$

For each cell T, we have $i_T(0) = 0$ and, to model the heterogeneity in receptor expression level, we suppose that the initial value $r_T(0)$ is drawn from a log-normal distribution with parameters m_0 and σ_0:

$$r_T(0) \sim \log \mathcal{N}(m_0, \sigma_0^2).$$

A scheme of the model can be found in Figure 5.2. The only stochastic element of this model is the initial condition of system (5.2) for each cell, T[1]. The rest of the two-attributes model is entirely deterministic and an analytic expression for

[1]From one simulation of the model to another, we have different initial conditions (drawn from the log-normal distribution $\log \mathcal{N}(m_0, \sigma_0^2)$) for each cell. However, for a given simulation of the model, once the initial conditions fixed, the system is entirely deterministic: the receptor dynamics of each cell is governed by (5.2).

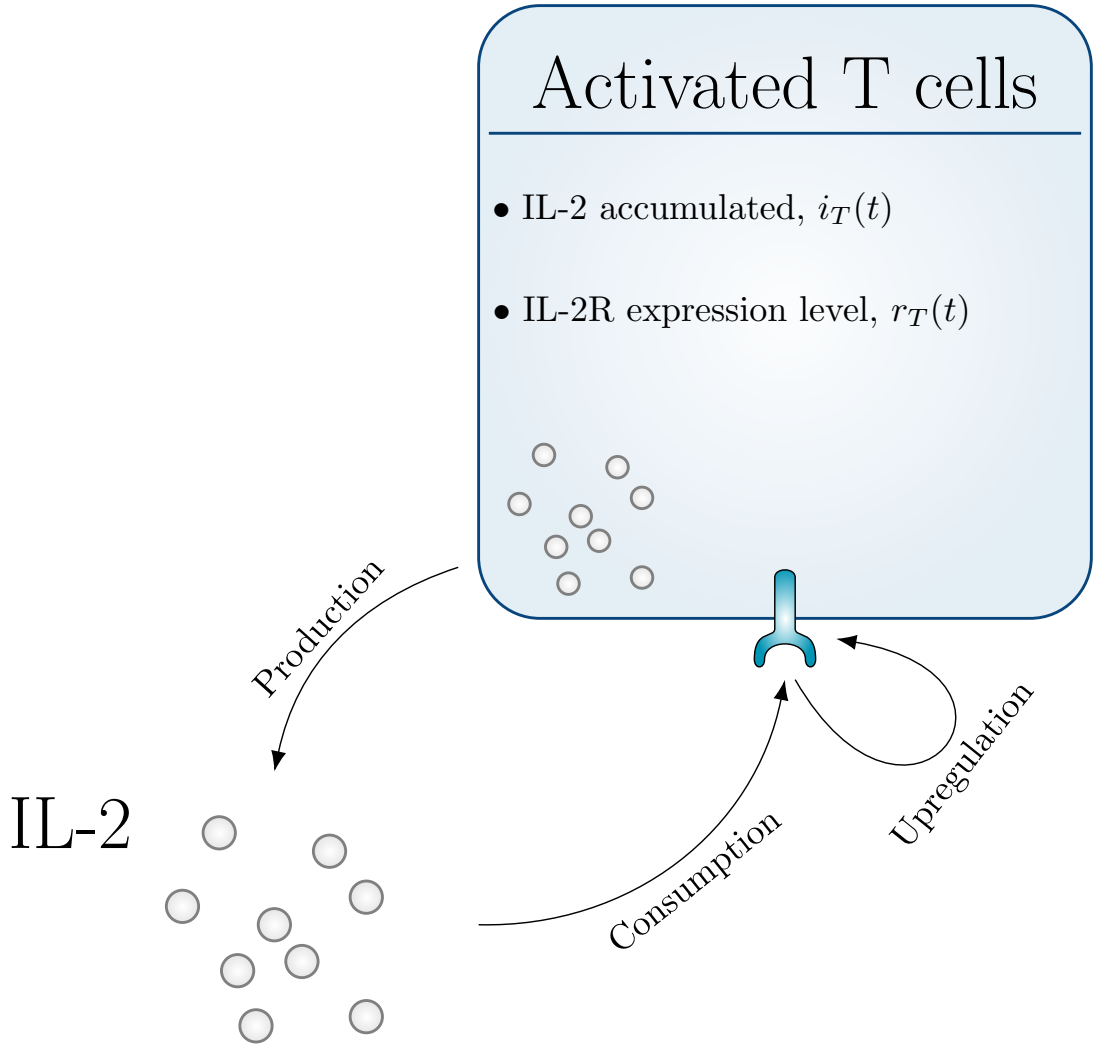


Figure 5.2: Scheme of the model in which the population size of self-activated conventional T cells is constant (fully deterministic model): T cells produce IL-2, which they consume proportionally to their current IL-2R expression level, $r_T(t)$. They upregulate their IL-2R expression proportionally to the level of IL-2 consumed since their entrance in the cell pool ($t = 0$ here), $i_T(t)$.

$i_T(t)$ and $r_T(t)$ can be found for any time t. Let us first determine a more explicit expression for the population variable, $R(t)$.

5.2.1 Determining the population variable $R(t)$

Since the total number of cells, N, is constant over time, one can easily differentiate R twice. The first derivative, R', is:

$$\begin{aligned} R'(t) &= \sum_{T=1}^{N} \frac{dr_T}{dt}(t) \\ &= u \sum_{T=1}^{N} i_T(t), \end{aligned} \tag{5.4}$$

where we replaced $\dfrac{dr_T}{dt}$ by its expression in (5.2). Since $i_T(0) = 0$, we obtain the initial value of R':

$$R'(t=0) = 0. \tag{5.5}$$

From equation (5.3), we obtain the expression of the second derivative R'':

$$\begin{aligned} R''(t) &= u \sum_{T=1}^{N} \frac{di_T}{dt} \\ &= upN. \end{aligned} \tag{5.6}$$

Finally, by integrating (5.6), and because of (5.5), we obtain for all t:

$$R(t) = \frac{upN}{2} t^2 + R_0, \tag{5.7}$$

where we wrote $R(0) = \sum_{T=1}^{N} r_T(0) = R_0$.

5.2.2 Analytic expressions of $i_T(t)$ and $r_T(t)$ for any cell T

Now that we obtained an explicit expression for $R(t)$, we can differentiate equation (5.2a). For any activated conventional T cell, T, we have

$$\frac{d}{dt}\left(R \frac{di_T}{dt}\right) = pN \frac{dr_T}{dt} \iff R \frac{d^2 i_T}{dt^2} + R' \frac{di_T}{dt} = upN i_T,$$

where we replaced $\frac{dr_T}{dt}$ by its expression in equation (5.2b). The function $i_T(t)$ thus satisfies the following system composed of an ODE and its initial condition:

$$\begin{cases} \frac{d^2 i_T}{dt^2} + \frac{upN}{R}\frac{di_T}{dt}t - \frac{upN}{R}i_T(t) = 0 \\ i_T(0) = 0 \end{cases} . \tag{5.8}$$

One can notice that the identity function $i(t) = t$ is an obvious non-zero solution of system (5.8). Suppose that i_1 and i_2 are solutions of (5.8). We define the Wronskian $W(i_1, i_2)$ of these two functions as, for all t,

$$W(i_1, i_2)(t) \equiv i_1(t)i_2'(t) - i_1'(t)i_2(t).$$

Because of the initial condition of system (5.8), the Wronskian is always equal to 0 (see Section 2.1). As a consequence, the solutions of (5.8) are linearly dependent (Bocher, 1901) and for any cell T, for all t, we can write

$$i_T(t) = K_T t, \tag{5.9}$$

where K_T is a constant to be determined, *a priori* different for each cell T. Making use of the above expression for i_T and integrating equation (5.2b), we obtain an analytic expression for $r_T(t)$. For any cell T, for any time t,

$$r_T(t) = \frac{uK_T}{2}t^2 + r_T(0). \tag{5.10}$$

To determine the constant K_T, we substitute the expression of $r_T(t)$, $i_T(t)$ and $R(t)$ in equation (5.2a) and obtain for any cell T and time t:

$$K_T = \frac{NpK_T ut^2 + 2r_T(0)Np}{uNpt^2 + 2R_0}. \tag{5.11}$$

In particular, when $t = 0$, we obtain for any cell T,

$$K_T = pN\frac{r_T(0)}{R_0}. \tag{5.12}$$

The constant K_T thus represents the fraction of IL-2 consumed by the cell T

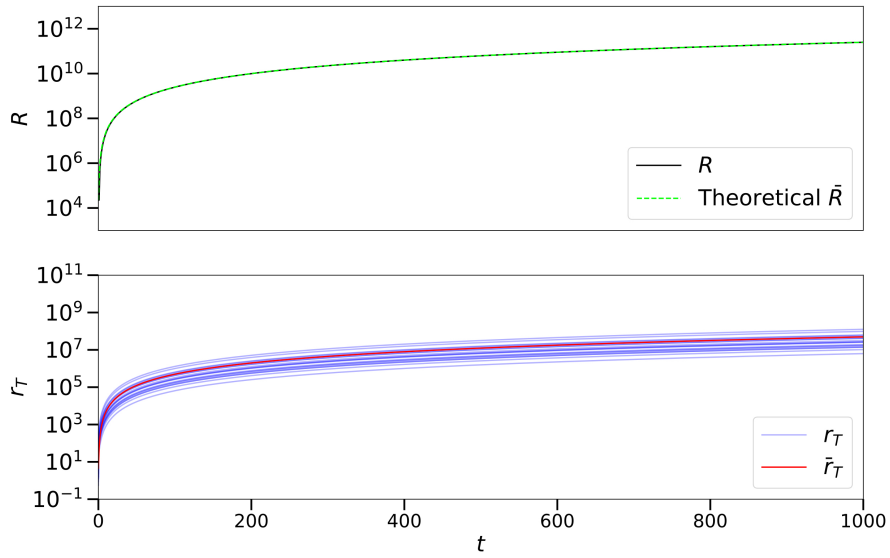


Figure 5.3: Deterministic case: time evolution plots of the population variable, $R(t)$, alongside its theoretical expression (top); and (bottom) IL-2R expression levels of 0.3% randomly chosen cells compared to the average theoretical expression (red), $\bar{r}_T(t) = \bar{r_0}\left(\frac{upN}{2R_0}t^2 + 1\right)$, where $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$. The parameter values used for this simulation are: $N = 5 \times 10^3$ cells, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[r_T]$/day/cell, $m_0 = 1$ and $\sigma_0 = 1$. The time step of the simulation was $\Delta t = 1$ days.

during the first absorption. Finally, substituting the expression of K_T yields the final expressions for $i_T(t)$ and $r_T(t)$. For any cell T and at any time t, we have

$$i_T(t) = pN\frac{r_T(0)}{R_0}t, \tag{5.13a}$$

$$r_T(t) = \frac{upN}{2}\frac{r_T(0)}{R_0}t^2 + r_T(0). \tag{5.13b}$$

We simulated the model with 5000 cells and plotted $R(t)$, and $r_T(t)$ for 0.3% of the cell population[1] as a function of time in Figure 5.3. Notice that $r_T(t) = \frac{r_T(0)}{R_0}R(t)$

[1]The model was simulated with the same code (described in appendix J) as for the other agent-based models that will be developed in this chapter, setting the cellular event rates (such as death or division rate) to 0.

at all t, so that $\frac{r_T(t)}{R(t)}$ is in fact constant over time. That is, each cell increases its receptor expression while maintaining its initial proportion, $\frac{r_T(0)}{R_0}$, within the population (see individual receptor trajectories in Figure 5.3 (bottom graph)).

5.2.3 Distribution of $r_T(t)$ and $i_T(t)$ at any time t

We can factorise $i_T(t)$ and $r_T(t)$ by $r_T(0)$:

$$\begin{aligned} i_T(t) &= r_T(0)\frac{pNt}{R_0}, \\ r_T(t) &= r_T(0)(\frac{upN}{2R_0}t^2 + 1). \end{aligned}$$

The term R_0 is the total IL-2R expression level (sum over all T cells) at $t = 0$. Though being a sum of random variables, R_0 is a constant in each simulation of the agent-based model[1]. Since $r_T(0)$ follows a log-normal distribution, $\log(r_T(0))$ follows a normal distribution. Thus, the logarithms of $i_T(t)$ and $r_T(t)$ are the sum of a normally distributed variable and a constant:

$$\begin{aligned} \log(i_T(t)) &= \log(r_T(0)) + \log(\frac{pNt}{R_0}) && \forall t > 0, \\ \log(r_T(t)) &= \log(r_T(0)) + \log(\frac{upN}{2R_0}t^2 + 1) && \forall t. \end{aligned} \tag{5.14}$$

As a consequence, $\log(i_T(t))$ and $\log(r_T(t))$ are normally distributed. Hence, for any conventional T cell, T, at any time t, $i_T(t)$ and $r_T(t)$ are log-normally distributed with mean $m_0 + \log\left(\frac{pN}{R_0}t\right)$ and $m_0 + \log\left(\frac{upN}{2R_0}t^2 + 1\right)$, respectively:

$$\begin{aligned} i_T(t) &\sim \log\mathcal{N}\left(m_0 + \log\left(\frac{pN}{R_0}t\right), \sigma_0^2\right) && \forall t > 0, \\ r_T(t) &\sim \log\mathcal{N}\left(m_0 + \log\left(\frac{upN}{2R_0}t^2 + 1\right), \sigma_0^2\right) && \forall t. \end{aligned} \tag{5.15}$$

At any time t (only $t > 0$ for $i_T(t)$), $i_T(t)$ and $r_T(t)$ distributions are the shifted initial log-normal distribution of $r_T(0)$ (see Figure 5.4). We have $i_T(t = 0) = 0$

[1] R_0 is different from one simulation to another. However, for a given simulation, it is a constant.

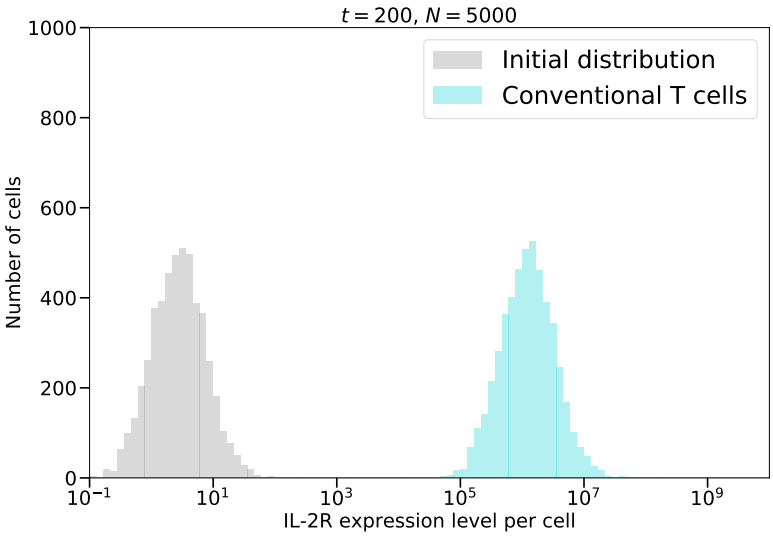


Figure 5.4: Deterministic case: distribution of IL-2R expression level among the cell population from the simulation of Figure 5.3 at $t = 200$ days. The distribution (blue) is shifted to the right compared to the initial distribution (in grey).

for any cell T.

5.3 Hybrid dynamics with cell death (only)

In this section, we consider the previous deterministic two-attributes model and introduce the first stochastic cellular event in the dynamics by now assuming that T cells may die with death rate μ: the number of living cells may be said to follow a pure death process. The model is recapitulated in Figure 5.5. Death events occur at discrete times. The time between two successive death events is a random variable drawn from an exponential distribution.

The number of cells, $N(t)$, and the population variables $R(t)$ and $I(t)$ are now stochastic time-dependent quantities and we will consider their ensemble averages, which we denote using the overbar. Our first such quantity is the mean number of activated T cells at time t, given by:

$$\bar{N}(t) = N(0)e^{-\mu t}. \tag{5.16}$$

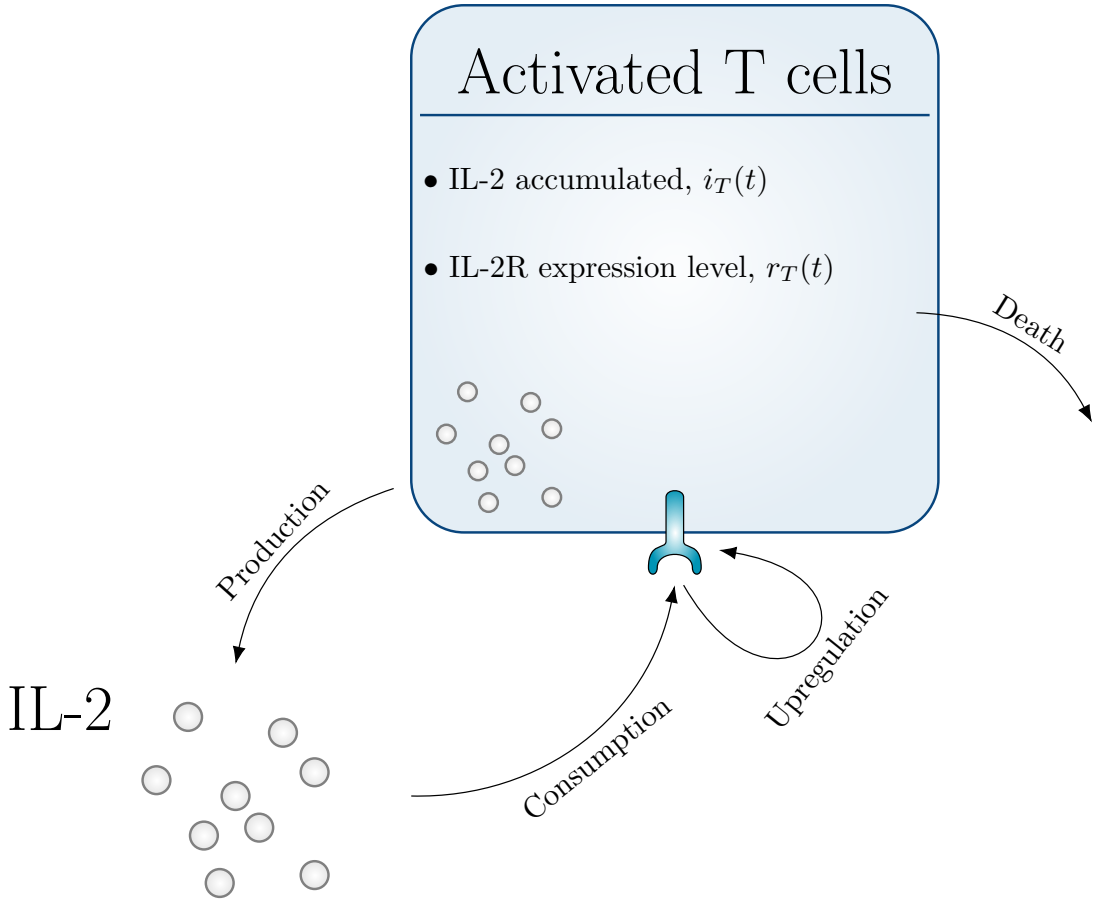


Figure 5.5: Scheme of the agent-based model in which activated conventional T cells which may die with death rate μ. Activated T cells produce and consume IL-2, and upregulate their IL-2R expression level, according to the two-attributes dynamics described in the previous section.

Contrary to the computation conducted for the previous deterministic model, this time, the differentiation of $R(t) = \sum_{T=1}^{N(t)} r_T(t)$ is troublesome as the number of cells, $N(t)$, is randomly changing over time. We can, however, by considering $I(t)$ alongside of $R(t)$, calculate the corresponding expressions for $\bar{I}(t)$ and $\bar{R}(t)$.

5.3.1 Computation of $\bar{I}(t)$ and $\bar{R}(t)$

During a small time interval, $[t, t+\Delta t]$, each cell has probability $\mu\Delta t$ of dying. The probability that no cell dies is $1-\mu N(t)\Delta t$. When a cell T dies, the IL-2 it absorbed before its death, $i_T(t)$, is lost to the cell population. Thus, as $\Delta t \to 0$, for any t,

$$I(t+\Delta t) = \begin{cases} I(t) + pN(t)\Delta t + \mathcal{O}(\Delta t^2) & \text{with probability } 1-\mu N(t)\Delta t, \\ \text{and for each cell } T, & \\ I(t) + pN(t)\Delta t - i_T(t) + \mathcal{O}(\Delta t^2) & \text{with probability } \mu\Delta t. \end{cases} \tag{5.17}$$

Thus, the ensemble average $\bar{I}(t+\Delta t)$ is given by

$$\begin{aligned} \bar{I}(t+\Delta t) &= \left(\bar{I}(t) + p\bar{N}(t)\Delta t\right)\left(1-\mu\bar{N}(t)\Delta t\right) \\ &\quad + \left[\bar{I}(t) + p\bar{N}(t)\Delta t\right]\mu\bar{N}(t)\Delta t - \frac{\bar{I}(t)}{\bar{N}(t)}\mu\bar{N}(t)\Delta t + \mathcal{O}(\Delta t^2) \\ &= \bar{I}(t) + \left(p\bar{N}(t) - \mu\bar{I}(t)\right)\Delta t + \mathcal{O}(\Delta t^2). \end{aligned}$$

That is, we obtain a differential equation for $\bar{I}$:

$$\frac{d\bar{I}}{dt} = p\bar{N} - \mu\bar{I}. \tag{5.18}$$

Since $I(0) = 0$, its solution is

$$\bar{I}(t) = pN(0)te^{-\mu t}. \tag{5.19}$$

We proceed similarly for $R(t)$. When a randomly chosen cell T dies, its IL-2 receptors are lost to the cell population. Thus as $\Delta t \to 0$, for any t,

$$R(t+\Delta t) = \begin{cases} R(t) + uI(t)\Delta t + \mathcal{O}(\Delta t^2) & \text{with probability } 1 - \mu N(t)\Delta t, \\ \text{and for each cell } T, & \\ R(t) + uI(t)\Delta t - r_T(t) + \mathcal{O}(\Delta t^2) & \text{with probability } \mu\Delta t. \end{cases} \tag{5.20}$$

We obtain the following differential equation for $\bar{R}$:

$$\frac{d\bar{R}}{dt} = u\bar{I} - \mu\bar{R}. \tag{5.21}$$

Substituting the expression for $\bar{I}$ and solving the differential equation, we obtain:

$$\bar{R}(t) = (\frac{upN(0)}{2}t^2 + R_0)e^{-\mu t}, \tag{5.22}$$

where $R_0 = R(t = 0)$.

In this hybrid dynamics model, the average variable $\bar{R}(t)$ is equal to the corresponding variable in the purely deterministic model multiplied by the exponential in the mean population function (5.16).

5.3.2 Analytic expressions for $i_T(t)$ and $r_T(t)$

In practice, the hybrid dynamics of each realisation of the model can be described as intervals of deterministic dynamics interrupted by discrete events (here, an event is the death of a cell). In the intervals between events, $N(t)$ and $R(t)$ are constant and system (5.2) is valid.

However, how to link the expressions for $i_T(t)$ and $r_T(t)$, derived in equations (5.13) and valid only during the small time interval between two population stochastic event (death of a cell), and the continuous functions $i_T(t)$ and $r_T(t)$, defined for the whole lifetime of cell T, remains unclear. Heuristically, we can solve system (5.2) in which we replaced $R(t)$ and $N(t)$ by the average expression $\bar{R}(t)$ and $\bar{N}(t)$, respectively:

$$\frac{di_T}{dt} = p\bar{N}\frac{r_T}{\bar{R}}, \tag{5.23a}$$

$$\frac{dr_T}{dt} = ui_T, \tag{5.23b}$$

with the same initial conditions:

$$\begin{aligned} i_T(0) &= 0, \\ r_T(0) &\sim \log \mathcal{N}(m_0, \sigma_0^2). \end{aligned}$$

In that case, we can proceed similarly to Section 5.2. Differentiating equation (5.23a), we obtain for each alive activated conventional cell, T:

$$\frac{d}{dt}(\bar{R}\frac{di_T}{dt}) = p\bar{N}\frac{r_T}{dt} + p\frac{d\bar{N}}{dt}r_T.$$

By replacing $\frac{dr_T}{dt}$ by its expression in equation (5.23b), the previous equation is equivalent to:

$$\bar{R}\frac{d^2i_T}{dt^2} + \bar{R}'\frac{di_T}{dt} = up\bar{N}i_T - \mu p\bar{N}r_T, \tag{5.24}$$

where $\bar{R}'$ is the derivative of $\bar{R}$. Isolating r_T in equation (5.23a) yields:

$$r_T = \frac{\bar{R}}{p\bar{N}}\frac{di_T}{dt}, \tag{5.25}$$

which we substitute in equation (5.24) to obtain:

$$\widetilde{R}e^{-\mu t}\frac{d^2i_T}{dt^2} + \widetilde{R}'e^{-\mu t}\frac{di_T}{dt} - up\bar{N}i_T = 0, \tag{5.26}$$

where we defined for all t, $\widetilde{R}(t) \equiv \frac{upN(0)}{2}t^2 + R_0$ as the expression for $R(t)$ in the fixed population case. Finally, by replacing $\bar{N}$ by its expression and simplifying the exponential, we obtain the ordinary differential equation that describes i_T (during the lifetime of cell T):

$$\widetilde{R}\frac{d^2i_T}{dt^2} + \widetilde{R}'\frac{di_T}{dt} - \widetilde{R}''i_T = 0, \tag{5.27}$$

where $\widetilde{R}''$ is the second derivative of $\widetilde{R}$. Considering the initial conditions, for each cell T, $i_T(t)$ satisfies the differential equation obtained for $i_T(t)$ in the fixed

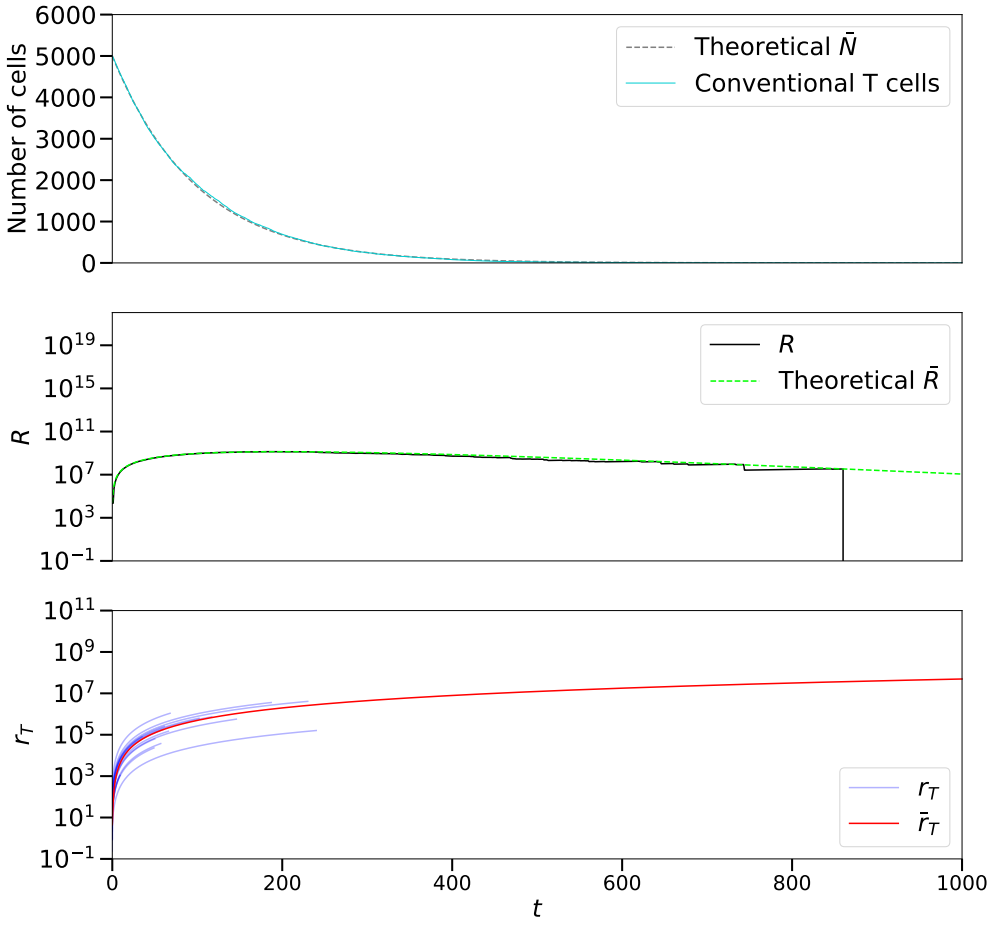


Figure 5.6: ABM with death only: time evolution plots of the number of cells, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells compared to the theoretical average expression (red), $\bar{r}_T(t) = \bar{r_0}\left(\frac{upN(0)}{2R_0}t^2 + 1\right)$, where $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$. The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\mu = 0.01$ /day, $m_0 = 1$ and $\sigma_0 = 1$. The simulation started with $N(0) = 5 \times 10^3$ cells and the time step was $\Delta t = 1$ day. In the bottom plot, a blue curve ends when the corresponding T cell dies.

population case (system (5.8)). The domain of definition of this system, however, is the lifetime of cell T and not the whole simulation time. We solved this system in Section 5.2. Hence, for any cell T at any time t (during its lifetime), we have (see Figure 5.6)

$$i_T(t) = pN(0)\frac{r_T(0)}{R_0}t, \tag{5.28a}$$

$$r_T(t) = \frac{upN(0)}{2}\frac{r_T(0)}{R_0}t^2 + r_T(0). \tag{5.28b}$$

As a consequence, $r_T(t)$ and $i_T(t)$ follow the distributions described in (5.15):

$$\begin{aligned} i_T(t) &\sim \log\mathcal{N}\left(m_0 + \log\left(\frac{pN(0)}{R_0}t\right), \sigma_0^2\right) && \forall\ 0 < t < t_T^{out}, \\ r_T(t) &\sim \log\mathcal{N}\left(m_0 + \log\left(\frac{upN(0)}{2R_0}t^2 + 1\right), \sigma_0^2\right) && \forall\ 0 \leq t < t_T^{out}, \end{aligned}$$

where t_T^{out} is the time at which the cell T dies. Note that if the cell did not die at the end of the simulation (at $t = t_{max}$), then $t_T^{out} = t_{max}$. Plotting the distributions of IL-2R expression and accumulated IL-2 levels for each time step of the simulation (and with the x-axis in log-scale), we observe normal distributions moving to the right and flattening as the population of cells is going extinct (see Figure 5.7 for the receptor distribution).

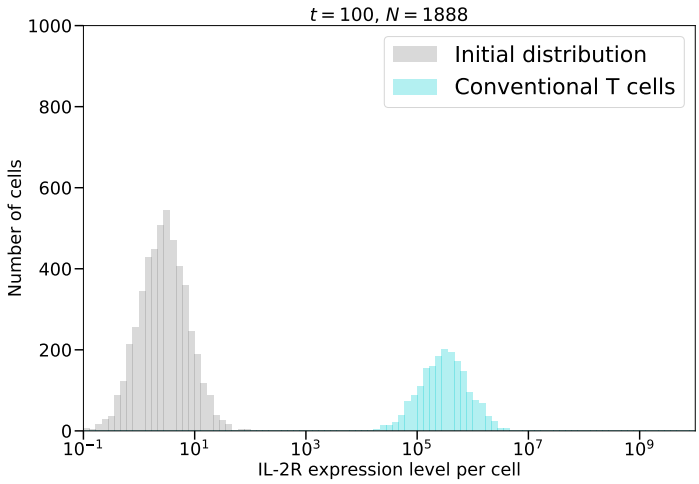


Figure 5.7: ABM with death only: distribution of IL-2R expression level among the cell population from the simulation of Figure 5.6 at t=100 (blue). The receptor distribution is shifted to the right and flattened, compared to the initial distribution (grey). At $t = 100$ days, there are 1888 cells left in the cell pool.

5.4 Death and activation hybrid system: two regimes and a stochastic steady state

We introduce a second type of population stochastic event (see Figure 5.8): we now suppose that new cells may enter the activated conventional cell pool, by being activated from a naive state (not modelled)[1]. In combination with cell death (death rate μ), this has the effect of driving the population to a stochastic steady state.

5.4.1 Mathematical analysis

If new cells can enter the pool of activated conventional T cells with rate α, the mean number of cells satisfies:

$$\frac{d\bar{N}}{dt} = \mu\bar{N} + \alpha. \tag{5.29}$$

[1]In classic branching processes theory, this activation event would be called immigration.

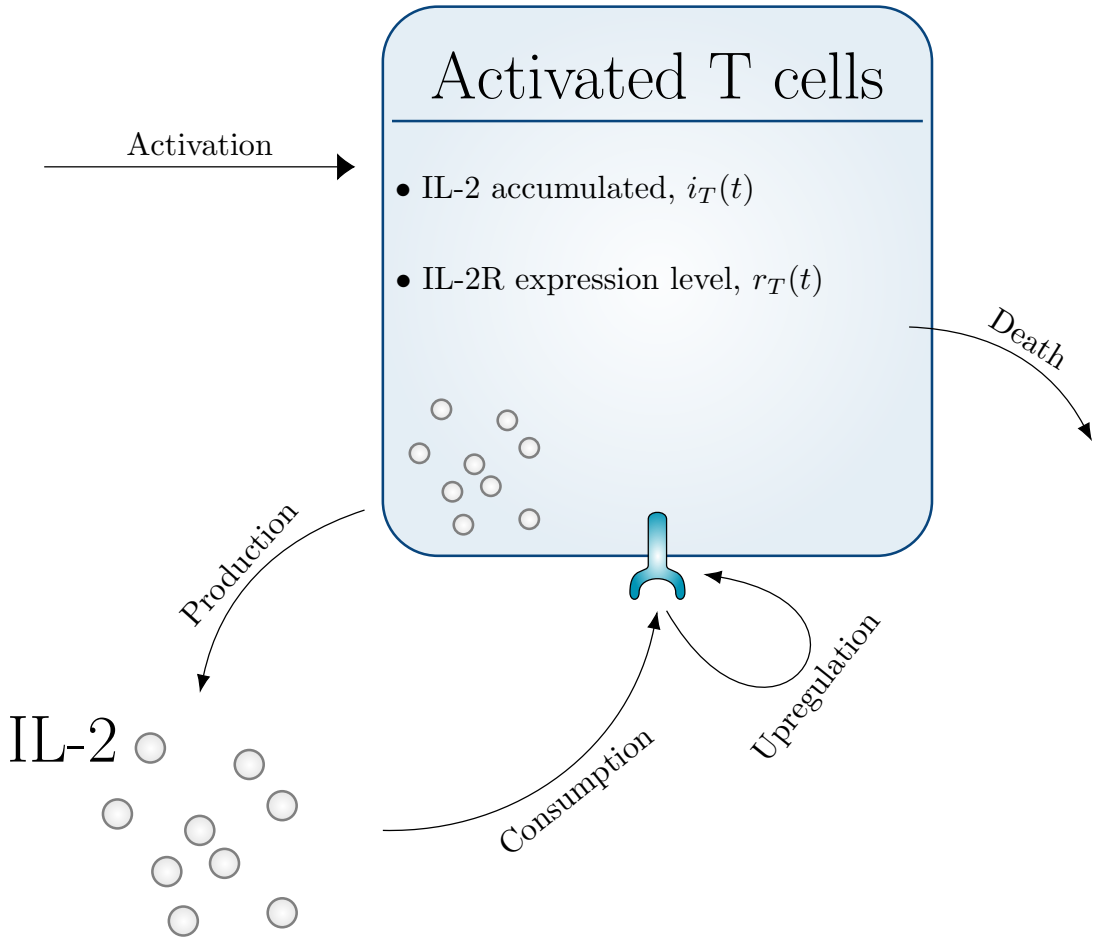


Figure 5.8: Death and activation hybrid system: activated T cells produce and consume IL-2. They also upregulate their IL-2R expression. New cells enter the pool at rate α (activation from the naive state) and any cell may die with death rate μ.

This differential equation can be solved and we obtain the mean number of cells at any time t:

$$\bar{N}(t) = \left(N(0) - \frac{\alpha}{\mu}\right) e^{-\mu t} + \frac{\alpha}{\mu}, \tag{5.30}$$

where $\bar{N}(0) = N(0)$ is the initial number of cells in the activated pool. In the steady state, there are on average $N_\infty \equiv \dfrac{\alpha}{\mu}$ cells in the cell pool. We investigate the properties of the population of cells in the steady state, starting with the mean quantities $\bar{I}(t)$ and $\bar{R}(t)$.

Expressions for $\bar{I}(t)$ and $\bar{R}(t)$

We proceed similarly to the previous model and consider a small time step interval $[t, t+\Delta t]$. During this time interval, each cell has probability $\mu\Delta t$ of dying. A new cell, T, can enter the cell pool with probability $\alpha\Delta t$. This cell, arriving at $t_T^{in} \in [t, t+\Delta t]$ will receive a certain number of IL-2R receptors, $r_T(t_T^{in}) = r_T^0$, drawn from the log-normal distribution $\log \mathcal{N}(m_0, \sigma_0^2)$, that has mean $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$. The probability that no cell dies or becomes activated is $\rho = 1 - \mu N(t)\Delta t - \alpha\Delta t$. Thus, as $\Delta t \to 0$, for any t,

$$R(t+\Delta t) = \begin{cases} R(t) + uI(t)\Delta t + \mathcal{O}(\Delta t^2) & \text{with probability } \rho, \\ R(t) + uI(t)\Delta t + r_T^0 + \mathcal{O}(\Delta t^2) & \text{with probability } \alpha\Delta t, \\ \text{for each cell } T, & \\ R(t) + uI(t)\Delta t - r_T(t) + \mathcal{O}(\Delta t^2) & \text{with probability } \mu\Delta t. \end{cases} \tag{5.31}$$

A new cell starts with no absorbed cytokine ($i_T(0) = 0$), so the differential equation for $\bar{I}(t)$, obtained in (5.18), remains unchanged. Thus, $\bar{R}(t)$ and $\bar{I}(t)$ satisfy the following system of differential equations:

$$\frac{d\bar{R}}{dt} = u\bar{I} - \mu\bar{R} + \alpha\bar{r_0}, \tag{5.32a}$$

$$\frac{d\bar{I}}{dt} = p\bar{N} - \mu\bar{I}. \tag{5.32b}$$

Let us now solve these differential equations to specify $\bar{I}(t)$ and $\bar{R}(t)$. Solving the homogeneous equation,

$$\frac{d\bar{I}}{dt} = -\mu\bar{I},$$

leads to the general solution

$$\bar{I}(t) = K_I e^{-\mu t},$$

where K_I is a constant to be determined. To find a particular solution of equation (5.32b), we use the method of variation of parameters. We obtain:

$$K_I'(t)e^{-\mu t} = p\left(N(0) - \frac{\alpha}{\mu}\right)e^{-\mu t} + p\frac{\alpha}{\mu},$$

which by integration yields:

$$K_I(t) = p\left(N(0) - \frac{\alpha}{\mu}\right)t + p\frac{\alpha}{\mu^2}e^{\mu t} + K_1,$$

where K_1 is a constant to determine. Finally, considering the initial condition $\bar{I}(0) = I(0) = 0$, we derive a general expression for the average total quantity of absorbed IL-2 at any time t:

$$\bar{I}(t) = \left(p\left(N(0) - \frac{\alpha}{\mu}\right)t - \frac{p\alpha}{\mu^2}\right)e^{-\mu t} + \frac{p\alpha}{\mu^2}. \tag{5.33}$$

We now determine an expression for the average total receptor copy number $\bar{R}(t)$. The general solution of the homogeneous equation,

$$\frac{d\bar{R}}{dt} = -\mu\bar{R},$$

is $\bar{R}(t) = K_R e^{-\mu t}$, where K_R is a constant to determine. Making use of the method of variation of parameters to find a particular solution of equation (5.32a), we obtain:

$$K_R'(t)e^{-\mu t} = u\bar{I}(t) + \alpha\bar{r_0}.$$

Substituting $\bar{I}(t)$ by its expression obtained in equation (5.33) gives:

$$K_R'(t) = u\left(p\left(N(0) - \frac{\alpha}{\mu}\right)t - \frac{p\alpha}{\mu^2}\right) + u\frac{p\alpha}{\mu^2}e^{\mu t} + \alpha\bar{r_0}e^{\mu t}.$$

Integration of $K'_R(t)$ yields

$$K_R(t) = u\left(p\left(N(0) - \frac{\alpha}{\mu}\right)\frac{t^2}{2} - \frac{p\alpha}{\mu^2}t\right) + u\frac{p\alpha}{\mu^3}e^{\mu t} + \frac{\alpha\bar{r_0}}{\mu}e^{\mu t} + K_2, \tag{5.34}$$

where K_2 is a constant to determine. We have:

$$\bar{R}(t) = K_R(t)e^{-\mu t}.$$

Finally, considering the initial condition $\bar{R}(0) = R(0) = R_0$, we find

$$K_2 = R_0 - \frac{\alpha\bar{r_0}}{\mu} - \frac{up\alpha}{\mu^3},$$

and obtain an expression for $\bar{R}(t)$ at any time t:

$$\bar{R}(t) = \left[pu\left(\left(N(0) - \frac{\alpha}{\mu}\right)\frac{t^2}{2} - \frac{\alpha}{\mu^2}t\right) + R_0 - R_\infty\right]e^{-\mu t} + R_\infty, \tag{5.35}$$

where R_∞ is the mean value of $R(t)$ at steady state:

$$R_\infty = \frac{\alpha}{\mu}\left(\bar{r_0} + \frac{up}{\mu^2}\right). \tag{5.36}$$

Computing $i_T(t)$ and $r_T(t)$

The attributes of each cell T of the current cell pool satisfy, heuristically, the following system of differential equations:

$$\frac{di_T}{dt} = p\bar{N}\frac{r_T}{\bar{R}}, \tag{5.37a}$$

$$\frac{dr_T}{dt} = ui_T. \tag{5.37b}$$

This system is the same as for the previous models. However, the initial conditions are not defined at $t = 0$ anymore but at the time the cell T enters the pool, t_T^{in} (which can be 0 if the cell is present at the beginning of the simulation):

$$i_T(t_T^{in}) = 0,$$
$$r_T(t_T^{in}) \sim \log \mathcal{N}(m_0, \sigma_0^2).$$

We can decouple system (5.37) and obtain:

$$\bar{R}\frac{d^2 i_T}{dt^2} + (\bar{R}' - \bar{R}\frac{\bar{N}'}{\bar{N}})\frac{di_T}{dt} - up\bar{N}i_T = 0, \tag{5.39a}$$

$$\frac{d^2 r_T}{dt^2} - up\bar{N}\frac{r_T}{\bar{R}} = 0. \tag{5.39b}$$

We note that $\bar{R}(t)$ and $\bar{N}(t)$ are composed of an exponential and a constant term. Thus, the exponential cannot be simplified and the differential equations cannot be solved in the general case.

Two regimes

The computation of the average steady state value of the total number of IL-2R (equation (5.36)) shows that

$$\frac{R_\infty}{N_\infty} = \bar{r_0} + \frac{up}{\mu^2}.$$

This ratio is the long-term average receptor expression level per cell. If

$$\bar{r_0} \gg \frac{up}{\mu^2},$$

then the average long-term receptor expression level of a cell during its lifetime differs little from the value it was given when the cell entered the pool. On the contrary if

$$\bar{r_0} \ll \frac{up}{\mu^2},$$

then on average, cells significantly upregulate their receptor expression during their lifetime. These two regimes exhibit different dynamics that are only captured by numerical simulations. We will see later that, in fact, only a minority of cells live long enough to massively upregulate their receptor expression level. These cells, however, are the main contributors of the population variable $R(t)$.

In the following sections, I investigate the two regimes in greater details, making use of numerical simulations and analytic results. For the simulations of this section, I used a fixed time step $\Delta t = 1$. For the parameter values considered, this time step is small enough to not distort the dynamics (see Appendix H). Let

us start with the *egalitarian regime* (case $\bar{r_0} \gg \frac{up}{\mu^2}$) as it has the most intuitive outcome.

5.4.2 Egalitarian regime

Suppose that the cells entering the activated conventional T cell pool express, on average, $\bar{r_0} \gg \frac{up}{\mu^2}$ IL-2 receptors. In this regime, if $\bar{r_0}$ is big enough, we even observe no receptor upregulation (see Figure 5.9; a receptor upregulation can be observed for a smaller $\bar{r_0}$, see Appendix G). All cells consume IL-2 and upregulate their IL-2R at the same speed (no trajectories are crossing each other in Figure 5.10). We simulate the model with $N(0) = N_\infty$. The average ensemble $\bar{R}(t)$ reaches its steady state, R_∞ (defined in equation (5.36)), rapidly. As illustrated in Figure 5.9, we observe that $R(t)$ and $N(t)$ remain very close to their average value $\bar{R}(t)$ and $\bar{N}(t)$. That is, R_∞ is a good approximation of $R(t)$ and $N_\infty = \frac{\alpha}{\mu}$ is a good approximation of $N(t)$, at any time t.

Approximating $N(t)$ and $R(t)$ by their average steady state value to find $r_T(t)$

Setting $\bar{R}(t) = R_\infty$ and $\bar{N}(t) = N_\infty$ for all t, we can solve equation (5.39b). The characteristic polynomial of this differential equation has two roots $r^* \equiv \sqrt{\frac{puN_\infty}{R_\infty}}$ and $-r^*$. Thus, the solution of the differential equation is of the form:

$$r_T(t) = Ae^{r^*t} + Be^{-r^*t}, \tag{5.40}$$

where A and B are constants to be determined with the initial conditions. Let us write r_T^0 the receptor expression level of cell T when it entered the pool at time $t = t_T^{in}$. We write t_T^{out} the time at which this cell will die ($t_T^{out} = t_{max}$ if the cell did not die at the end of the simulation). For this cell, we have $r_T(t_T^{in}) = r_T^0$ and $\frac{dr_T}{dt}(t_T^{in}) = 0$ which yields:

$$\begin{aligned} Ae^{r^*t_T^{in}} + Be^{-r^*t_T^{in}} &= r_T^0, \\ Ae^{r^*t_T^{in}} - Be^{-r^*t_T^{in}} &= 0. \end{aligned} \tag{5.41}$$

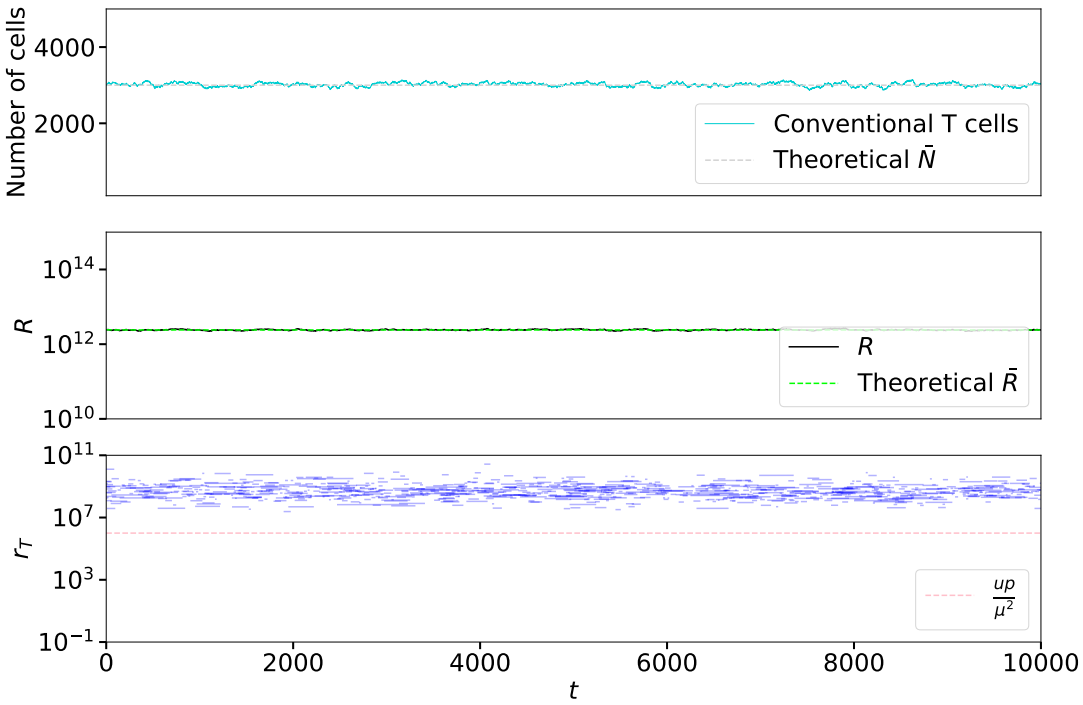


Figure 5.9: Time evolution plots of the number of cells, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells. The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\mu = 0.01$ /day, $m_0 = 20$ and $\sigma_0 = 1$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.

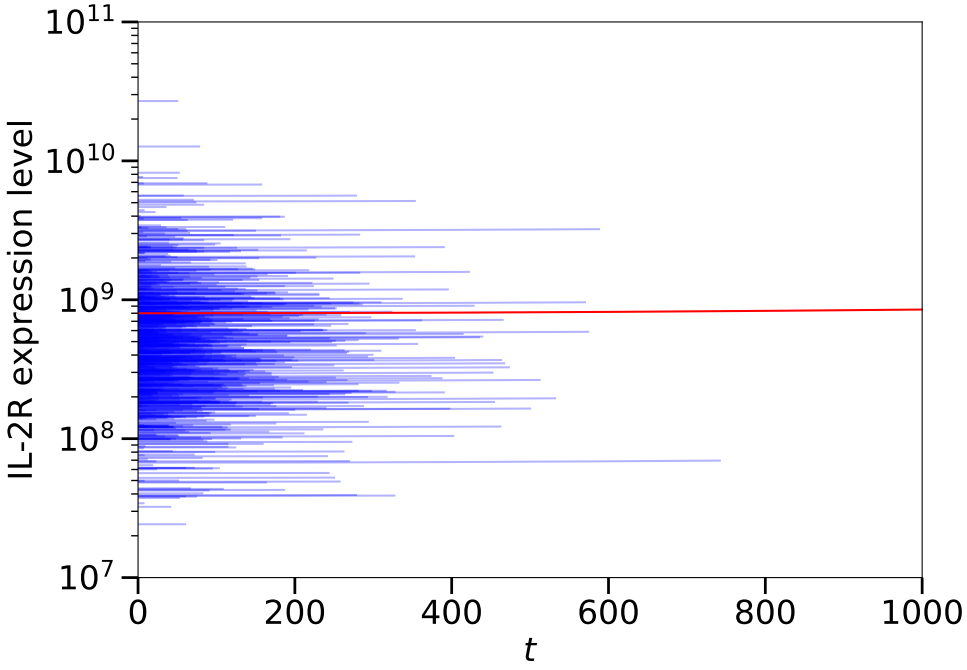


Figure 5.10: Individual receptor upregulation trajectories from the simulation shown in Figure 5.9, reported to the cell's activation time, t_T^{in}. No trajectories are crossing each other: they all follow the expression obtained in equation (5.42), that has mean $\bar{r_0}\cosh(r^*t)$, where $\bar{r_0} = \mathbb{E}(r_T^0) = e^{m_0+\frac{\sigma_0^2}{2}}$.

Solving this system, we obtain $A = \frac{r_T^0}{2}e^{-r^*t_T^{in}}$ and $B = \frac{r_T^0}{2}e^{r^*t_T^{in}}$. Thus, for any cell T and for all $t \in [t_T^{in}, t_T^{out})$:

$$r_T(t) = r_T^0 \cosh\left(r^*(t - t_T^{in})\right). \tag{5.42}$$

We can derive an expression for $i_T(t)$ from equation (5.37b) and obtain for any cell T for all $t \in [t_T^{in}, t_T^{out})$:

$$i_T(t) = \frac{r_T^0 r^*}{u}\sinh\left(r^*(t - t_T^{in})\right). \tag{5.43}$$

Figure 5.10 shows that the receptor upregulation dynamics of each cell does follow the expression obtained in equation (5.42).

5.4.3 Gerontocracy: the oldest cells deprive newer cells of cytokine

Suppose now that the cells enter the activated conventional T cell pool with, on average, $\bar{r_0} \ll \frac{up}{\mu^2}$ IL-2 receptors.

Distinguishing two cell cohorts

Suppose we simulate the model with time step Δt from $t = 0$ to $t = t_{max}$. We start the simulation with a certain number of cells that produce and consume IL-2, and upregulate their receptor expression level. The cells that enter the pools at a time $t > 0$ express a significantly lower number of receptors compared to the cells that were present at the beginning of the simulation and had time to upregulate their receptor expression level. The initial cohort is thus responsible for the majority of the cytokine intake and deprives the newer cells of IL-2. As a consequence, these newer cells cannot upregulate their receptor expression level. Once the last cell of the initial cohort dies, let us say at $t = t_1$, many cytokine molecules become suddenly available to be consumed by the cells that entered the pool between $t = 0$ and $t = t_1$. These cells can now increase their receptor copy number and start depriving subsequent cells of cytokine. Thus, at any time but $t = 0$, the cell population is split in two cohorts: the *new cohort* and the *old cohort*. The cells of the old cohort express high receptor expression levels and prevent the cells of the new cohort from increasing their receptor copy number by hoovering all the IL-2 produced. That is, we call this regime a *gerontocracy*. When the last cell of the old cohort dies, we observe a *cohort switch*: the new cohort becomes the old cohort, *i.e.*, it will be composed of the oldest cells of the pool and will deprive the newer cells (the fresh new cohort) of cytokine (see Figure 5.11). This split of the cell population, between a cohort that express a high number of IL-2R and an other cohort that express very few receptors, is consistent with the observations of Höfer *et al.* (2012) mentioned in the introduction.

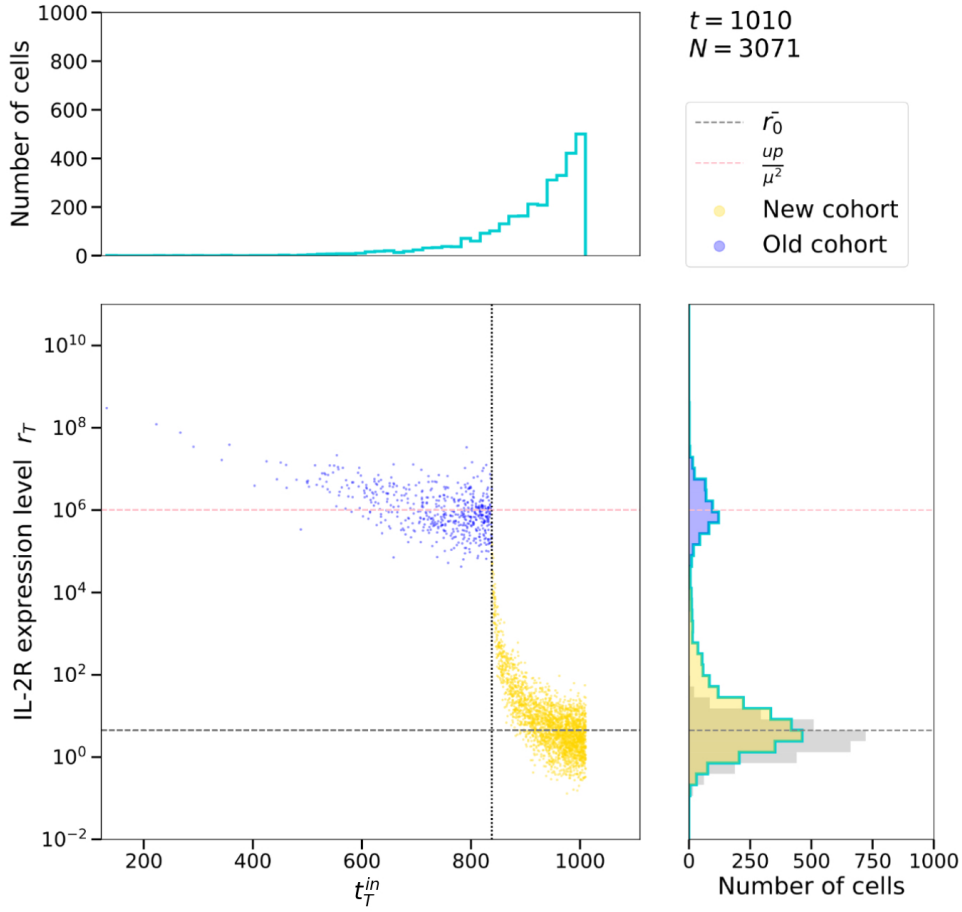


Figure 5.11: Scatterplot, at $t = 1010$, of the receptor expression level of each cell as a function of its date of activation (time at which it entered the cell pool, t_T^{in}), with joint distributions, in the gerontocracy regime. The vertical black dotted line represents the time at which the last cohort switch occurred: cells which joined the pool before this time belong to the old cohort (blue), cells that became activated after this time constitute the new cohort (yellow). At the cohort switch, cells of the previous new cohort (current old cohort) suddenly upregulated their IL-2R expression level, thus creating a "discontinuity" in the scatterplot and a bimodal receptor distribution (turquoise line on the right). As a consequence, cells of the current old cohort express a much higher number of IL-2R than cells of the new cohort. Each mode of the receptor distribution corresponds to the IL-2R distribution of one cohort. When almost all the cells of the old cohort died, the second mode is imperceptible. We indicated the initial IL-2R distribution in grey. The simulation was conducted with the following parameter values: $\mu = 0.01$ /day, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\sigma_0 = 1$, $m_0 = 1$, an initial population of $N(0) = \dfrac{\alpha}{\mu}$ cells and a time step $\Delta t = 1$ day.

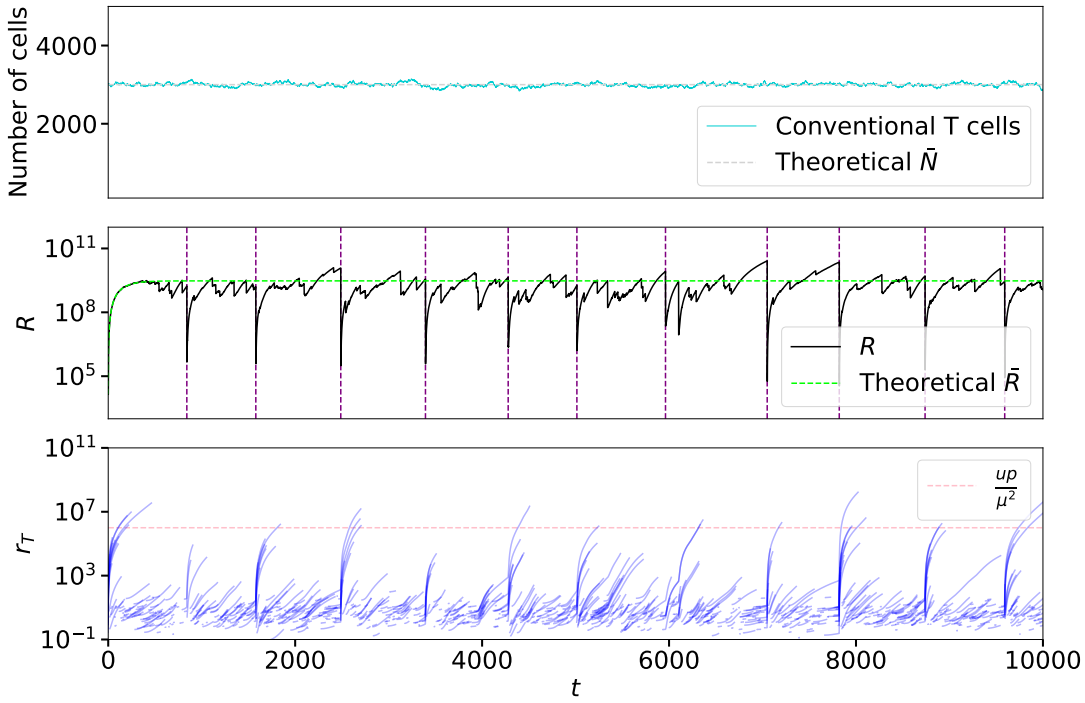


Figure 5.12: Time evolution plots of the number of cells, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top); population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of about 0.3% randomly chosen cells. Cohort switches (indicated by vertical dashed purple lines) coincide with $R(t)$'s sharp drops and the sudden receptor expression upregulation of certain cells. The parameter values used for this simulation were: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\mu = 0.01$ /day, $m_0 = 1$ and $\sigma_0 = 1$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.

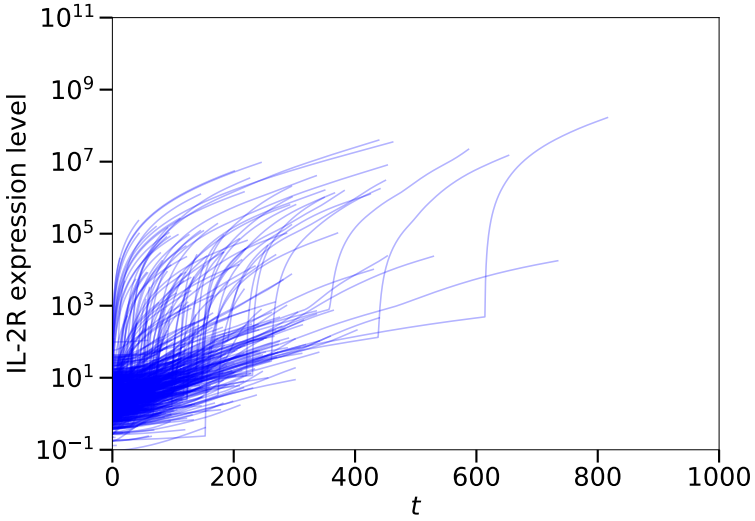


Figure 5.13: Individual receptor upregulation trajectories from the simulation shown in Figure 5.12, reported to the cell's activation time, t_T^{in}.

Cohort switches occur regularly

Cohort switches occur regularly and generally coincide with sharp $R(t)$'s drops and sudden upregulation of receptor expression in certain cells (see Figures 5.12 and 5.13). Indeed, when the last cell of the old cohort dies, R suddenly loses its major contributor (resulting in a sudden decrease). When the new cohort becomes the old cohort, cells of the recent old cohort suddenly upregulate their IL-2R expression level (because of the sudden IL-2 abundance resulting from the death of the cells that were the main consumers).

Since the old cohort follows a pure death process, its mean time to extinction,

$$\bar{\tau} \equiv \frac{1}{\mu} \sum_{i=1}^{N_\infty} \frac{1}{i}, \tag{5.44}$$

is the average time length at which a cohort switch happens. The centred distribution of cohort switch occurrences is a Gumbel distribution (see Section 2.7.2 for the definition and proof) with parameters $m = \dfrac{-\gamma_e}{\mu}$ and $\beta = \dfrac{1}{\mu}$. The constant γ_e is the Euler–Mascheroni constant. This distribution is illustrated in Figure 5.14.

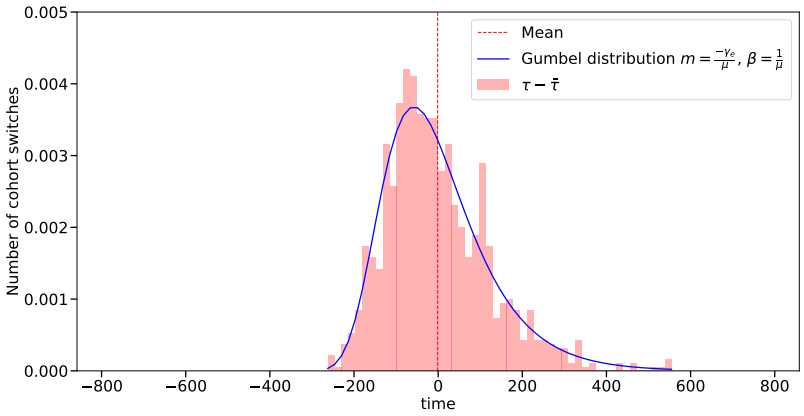


Figure 5.14: Normalised distribution of the time between cohort switches, τ, over a simulation until $t_{max} = 10^6$ days, centred around the mean time to extinction $\bar{\tau} = \frac{1}{\mu}\sum_{i=1}^{N_\infty}\frac{1}{i}$. The red dashed vertical line represents the average time of this distribution, which is to be compared with $\bar{\tau}$ (as this line falls around the 0 of the x-axis, the mean of this distribution is approximatively $\bar{\tau}$). This distribution follows a Gumbel distribution $Gumbel(m, \beta)$ with parameters $m = \frac{-\gamma_e}{\mu}$ and $\beta = \frac{1}{\mu}$. The simulation was conducted with the following parameter values: $\mu = 0.01$ /day, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\sigma_0 = 1$, an initial population of $N(0) = \frac{\alpha}{\mu}$ cells and a time step $\Delta t = 1$ day.

5.4.4 Additional investigations

In this section, I explore the transition between the two regimes, investigate what happens if the initial cell distribution is wide, and study the receptor distribution of the cell population over the whole simulation.

Bifurcation: when do we change from one regime to another?

We observed that this model has two different regimes, depending on whether the typical cell entering the pool receives a IL-2R expression level much larger or much lower than an average upregulation level $\frac{up}{\mu^2}$. A few questions remain unanswered: what happens when $\bar{r_0} \approx \frac{up}{\mu^2}$? Is the transition between the two regimes gradual or do we observe a binary switch when $\bar{r_0}$ passes the threshold $\frac{up}{\mu^2}$? To investigate these issues, we compute the mean IL-2R expression level of the old and new cohorts ($\bar{r}_{old}$ and $\bar{r}_{new}$, respectively) at a certain time t^* between the last two cohort switches of the simulation, for different $\bar{r_0}$ values. We chose this time to be $t^* = \frac{t_2 - t_1}{8}$ where t_1 and t_2 are the time of the penultimate and last cohort switches of the simulation, respectively. The definition of t^* was chosen such that the old cohort contains enough cells at each simulation, so that $\bar{r}_{old}$ does not vary too significantly when repeating a simulation. In the egalitarian regime (large $\bar{r_0}$), we anticipate these two means, $\bar{r}_{old}$ and $\bar{r}_{new}$, to be very close. On the contrary, in the gerontocracy regime (low $\bar{r_0}$), we expect the mean IL-2R expression levels of the two cohorts to be clearly distinguishable. We numerically computed $\bar{r}_{old}$ and $\bar{r_{new}}$ for 200 increasing values of $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$ (we varied m_0 and fixed $\sigma_0 = 1$), from 10 to 10^9 $[r_T]$/cell. For each $\bar{r_0}$ value, we repeated the simulation 2000 times and averaged $\bar{r}_{old}$ and $\bar{r}_{new}$ over these occurrences. The result is displayed in Figure 5.15. As expected, when $\bar{r_0} \ll \frac{up}{\mu^2}$, the two means are well distinct, showing that the cell population is divided in two cohorts. On the contrary, when $\bar{r_0} \gg \frac{up}{\mu^2}$, the two curves are superposed, meaning that the two cohorts express similar levels of IL-2R. Around $\frac{up}{\mu^2}$, we can still distinguish the two cohorts, though the difference in IL-2R expression levels is smaller. This suggests that $\frac{up}{\mu^2}$ is a good characteristic value to separate the two regimes but is not a hard threshold: the transition between the two regimes is gradual. This is

confirmed by short numerical simulations for $\bar{r_0} \approx 2\frac{up}{\mu^2}$, $\bar{r_0} \approx \frac{up}{\mu^2}$ and $\bar{r_0} \approx \frac{up}{75\mu^2}$ (see Appendix G). These simulations show that, as m_0 decreases, the system gradually starts to behave like in the gerontocracy regime: R becomes noisier and an increasing number of individual r_T trajectories (reported to their date of activation) are crossing each other.

Larger initial standard deviation σ_0

In the previous simulations, we fixed σ_0 and varied m_0 to explore the different ranges of $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$. However, not only is σ_0 the major contributor of $\bar{r_0}$ but it is also the main parameter that determines the width of the distribution (the variance of a log-normal distribution with parameters m_0 and σ_0 is $(e^{\sigma_0^2} - 1)e^{2m_0 + \sigma_0^2}$). When considering $\log(\bar{r_0})$, σ_0 is the parameter that determines the width of the normal distribution. Let us investigate the impact of the standard deviation on the dynamics.

I first repeated the computational regime transition analysis for larger fixed values of σ_0 ($\sigma_0 = 3$ and 5), and obtained graphs (not shown) very similar to Figure 5.15 (note that the first m_0 values were negative to accommodate the lower $\bar{r_0}$ values). At low $\bar{r_0}$, the average receptor expression levels of the two cohorts are distinct, at high $\bar{r_0}$, the mean of the two cohorts are equal.

To investigate further the influence of σ_0 on the dynamics, I simulated the model for a low value of $\bar{r_0}$, determined by a negative m_0 and a large σ_0. For instance, one could choose $m_0 = -2$ and $\sigma = 3$, or $m_0 = -10$ and $\sigma_0 = 5$, which both lead to $\bar{r_0} \approx 12$ $[r_T]$/cell, while choosing parameters such that $\frac{up}{\mu^2} = 10^6$ $[r_T]$/cell. The time evolutions of $R(t)$ and the individual r_T trajectories show that the system seems to behave like in the gerontocracy regime (described when $\sigma_0 = 1$): namely, $R(t)$ is noisy and the individual trajectories cross each others. However, one can observe that the sharp drops of $R(t)$ are more numerous than when $\sigma_0 = 1$ and seem uncorrelated from the cohort switches (see Figure 5.16). The scatterplot (individual receptor copy number as a function of the time at which the cell entered the pool, see Figure 5.17) shows that cohort switches are smoother than when $\sigma_0 = 1$: the receptor distribution is never bimodal and no discontinuity (such as observed in Figure 5.11) is observed. Instead, when the last

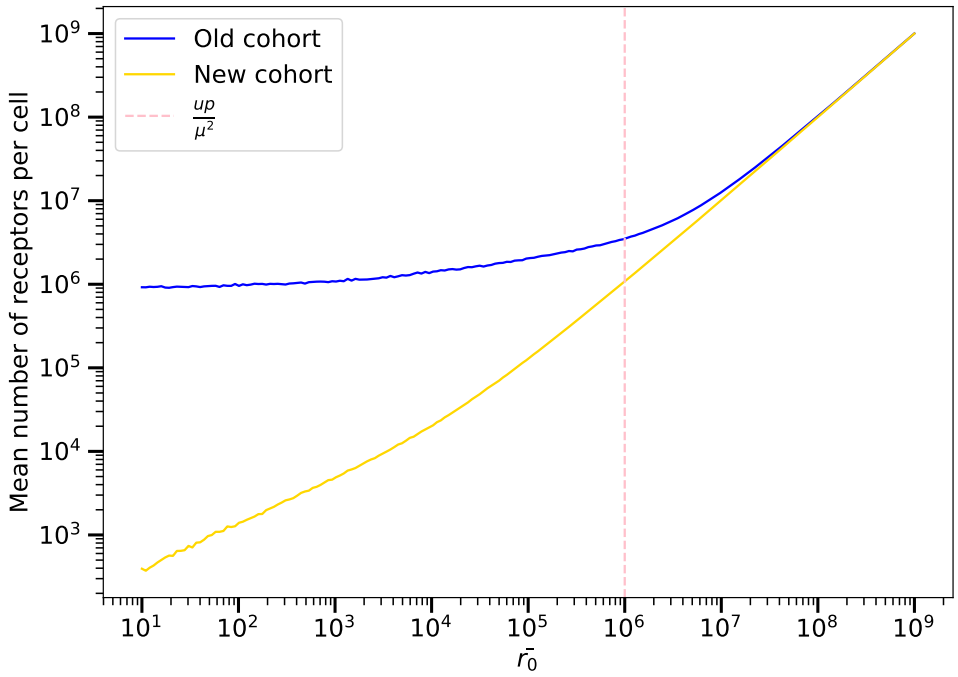
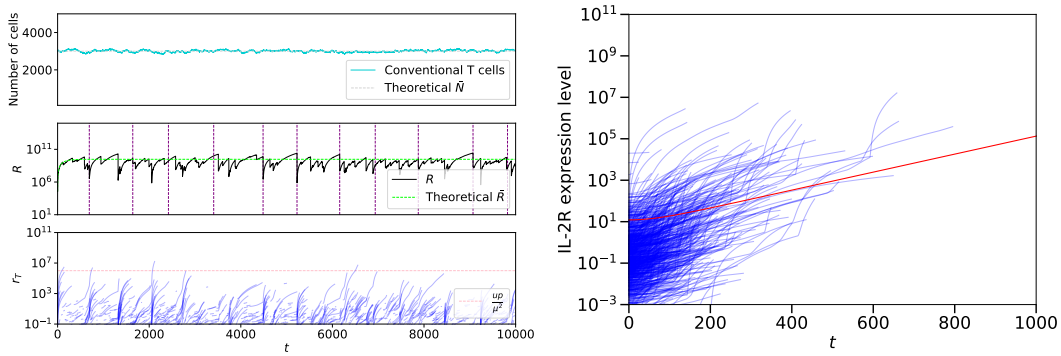


Figure 5.15: Regime transition graph: Mean number of receptors of the old (blue) and new (yellow) cohorts as a function of $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$. The curves represent the mean over 2000 simulations of the mean number of receptors among a cell cohort at t^*. The time t^* was arbitrarily defined as $\frac{t_2 - t_1}{8}$ where t_2 is the time of the last cohort switch of the simulation and t_1 the time of the penultimate cohort switch of the simulation. To obtain these graphs, the standard deviation of the receptor distribution at activation was fixed at $\sigma_0 = 1$. The initial mean number of receptors, m_0, was changed such that $\bar{r_0}$ varied between 10 and 10^9 $[r_T]$/cell. For each value of $\bar{r_0}$, the simulation was run for the following parameters: $t_{max} = 3000$ days, $\Delta t = 1$ day, $\mu = 0.01$ /day, $\alpha = 30$ cells/day, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell and $N(0) = \frac{\alpha}{\mu}$.

cell of the old cohort dies, we observe a gradual split of the receptor distributions of the two cohorts, thus explaining the two distinct means on the regime transition graph (Figure 5.15). These two distributions, however, significantly overlap, which keeps the whole distribution unimodal and almost unchanged.
I conjecture that, in this case, as the receptor distribution of each cohort is wide, $R(t)$ is more sensitive to the death of cells that expressed a high receptor level than to the extinction of a whole cohort (like when $\sigma_0 = 1$). Indeed, as illustrated in Figure 5.17, some cells of the old cohort express about 10^9 receptors while others express about 10^3 receptors. The death of a cell expressing about 10^9 receptors will have a greater impact on $R(t)$ (and make more IL-2 available for the other cells to compete for, resulting in a greater receptor upregulation of the other cells) than the death of a cell with about 10^3 IL-2R. That is, if the first cell dies, $R(t)$ might suddenly decrease. However, if the second cell dies, even if it is the last cell of the old cohort, we might not observe any significant change in the dynamics, as there are other cells expressing much more receptors in the pool.

Now, for a large $\bar{r_0}$ (for instance $m_0 = 3, \sigma_0 = 5$, which yields $\bar{r_0} \approx 5\frac{up}{\mu^2}$ with the usual parameter values), the system seems to behave like in the egalitarian regime (case $\sigma_0 = 1$), though we observe some individual receptor trajectories crossing each others (see Figure 5.18(b)). We also note that $R(t)$ seems noisy and underestimated (which will be discussed in the next paragraph). Despite a large $\bar{r_0}$ value, we notice that the majority of cells start with a receptor expression level much lower than $\frac{up}{\mu^2}$ (see Figure 5.18(a)). Let us call the population of cells expressing a high receptor level ($r_T(0) \gg \frac{up}{\mu^2}$) the *dominant cohort*. As new cells with this characteristic are constantly created, the dominant cohort never goes extinct. Thus, at any time, cells of the dominant cohort express a high receptor level and prevent the other cells from upregulating their own receptor level significantly. Indeed, cells that do not enter the cell pool with enough receptors, will never increase their receptor expression enough to compete with the cells of the dominant cohort. The high receptor level of cells of the dominant cohort also ensures that $R(t)$ is maintained around its theoretical value R_∞: we "mimic" the egalitarian regime observed when $\sigma_0 = 1$. On average, both the old



(a) Time evolution

(b) Relative trajectories

Figure 5.16: (left) Time evolution plots of the number of cells, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells. Cohort switches are indicated with the vertical dashed purple lines. (right) Individual receptor upregulation trajectories from the simulation, reported to the cell's activation time, t_T^{in}. The mean of the expression obtained in equation (5.42), $\bar{r_0}\cosh(r^*t)$, is also plotted (red line). The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\mu = 0.01$ /day, $m_0 = -2$ and $\sigma_0 = 3$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.

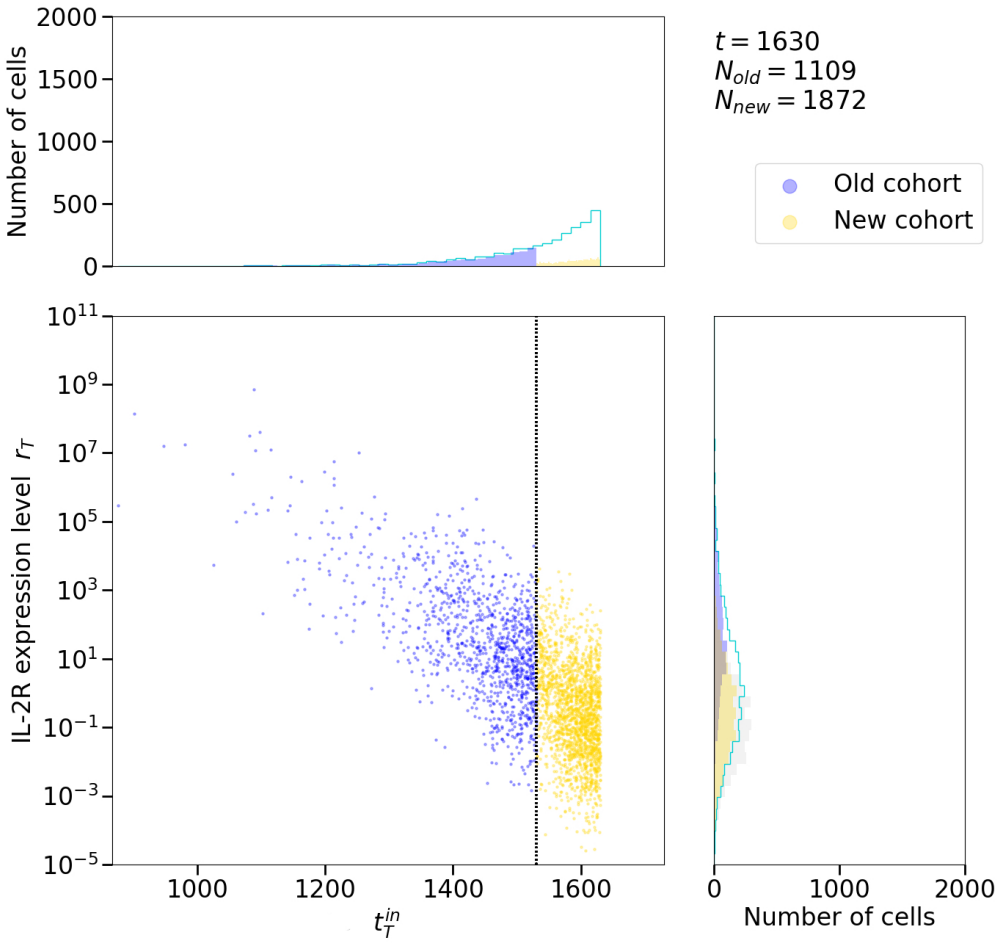
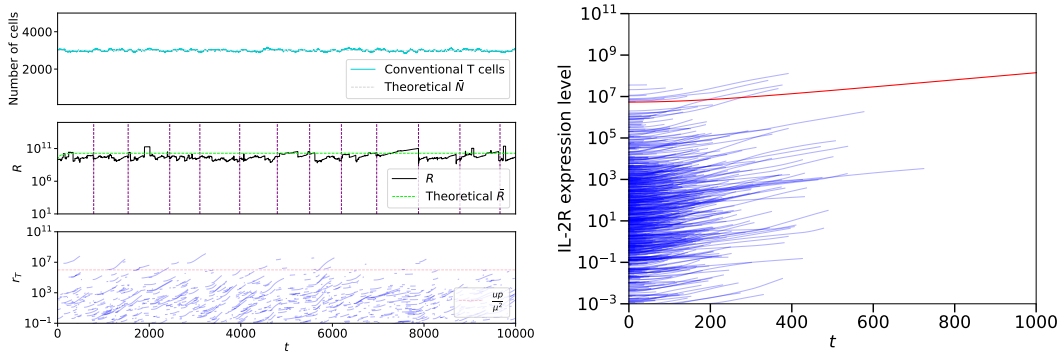


Figure 5.17: Scatterplot, at $t = 1630$, of the receptor expression level of each cell as a function of its date of activation (time at which it entered the cell pool, t_T^{in}), with joint distributions, when $\sigma_0 = 3$ and $m_0 = -2$. The vertical black dotted line represents the time at which the last cohort switch occurred: cells which joined the pool before this time belong to the old cohort (blue), cells that became activated after this time constitute the new cohort (yellow). In this configuration, the IL-2R distribution (turquoise line) is unimodal, despite the distribution of the two cohorts having distinct modes. These two distributions significantly overlap. We indicated the initial IL-2R distribution in grey. The simulation was conducted with the following parameter values: $\mu = 0.01$ /day, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, an initial population of $N(0) = \dfrac{\alpha}{\mu}$ cells and a time step $\Delta t = 1$ day.



(a) Time evolution

(b) Relative trajectories

Figure 5.18: (left) Time evolution plots of the number of cells, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells. Cohort switches are indicated with the vertical dashed purple lines. (right) Individual receptor upregulation trajectories from the simulation, reported to the cell's activation time, t_T^{in}. The mean of the expression obtained in equation (5.42), $\bar{r_0}\cosh(r^*t)$, is also plotted (red line). The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells//day, $\mu = 0.01$ /day, $m_0 = 3$ and $\sigma_0 = 5$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.

and the new cohorts present similar receptor expression levels (as observed in Figure 5.15), though huge disparities can be observed in each cohort (because cells of the dominant cohort can belong to the new and the old cohorts).

As mentioned, I observe that if σ_0 is too large (typically $\sigma_0 > 5$), the model clearly underestimates the population variable $R(t)$ (its average value is lower than its theoretical value). I conjecture that this is because the mean of the initial log-normal distribution, $\bar{r_0}$, is largely underestimated. Indeed, when sampling from the log-normal distribution, points at the tail of the distribution are the major contributors of the mean while being less likely to be sampled. If the sample size is not big enough, we might not sample enough points in this tail, resulting in a distribution with a mean value much smaller than theoretically predicted (see Figure 5.19).

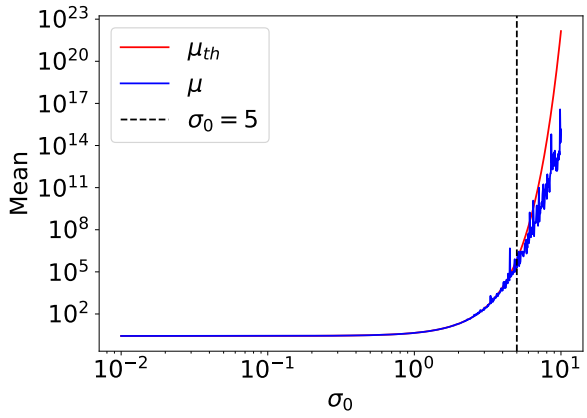


Figure 5.19: The mean of a sample of size $N = 10^5$, drawn from a log-normal distribution, $\log \mathcal{N}(m_0, \sigma_0^2)$, μ (blue), is compared to the theoretical mean of this distribution (red), $\mu_{th} = e^{m_0 + \frac{\sigma_0^2}{2}}$, for $m_0 = 1$ and different values of σ_0. When $\sigma_0 > 5$, the mean is underestimated.

Distribution of receptor expression levels over the whole simulation

Now, suppose that we simulate the model for a long time (for instance $t_{max} = 10^6$ days, with a time step Δt) and at each time, t, we sample some cells alive at time t. What is the receptor distribution of these sampled cells? As long as every cell have the same probability to be selected, the sampling method does not matter. Here I chose to sample the cells that died at each time step. That is, each cell has a probability $\mu \Delta t$ to be sampled. I fix $\sigma_0 = 1$ again. The receptor distribution of the n cells sampled during the simulation is shown in Figure 5.20 for both regimes. This distribution is compared to the initial receptor distribution (re-scaled to size n) and the relative contributions of each cohort.

First, we notice that, in the egalitarian regime ($m_0 = 20$), the receptor distribution of the sampled cells is very similar to the initial distribution, as expected from the observations in the previous sections: cells from both cohorts behave similarly and their long-term receptor expression level differs very little from their initial value. In the gerontocracy regime ($m_0 = 1$), however, the distribution is slightly shifted to the right compared to the initial distribution and displays a long tail[1]. The mode of the distribution corresponds to the contribution

[1]Note that we describe the distributions when the x-axis is in logarithmic scale.

of cells of the new cohort which were prevented from upregulating their IL-2R expression level by cells of the old cohort. The latter cells substantially increased their IL-2R expression and are responsible for the tail of the distribution.

In both cases, the mean number of receptors is $\bar{r_0}+\frac{up}{\mu^2}$, as expected. Figure 5.20 confirms that, in the gerontocracy regime, only the old cohort actually upregulates its receptor expression during its lifetime to a value much larger than its initial value. That is, a minority of cells drives the average number of receptors up so that it equals $\bar{r_0} + \frac{up}{\mu^2}$.

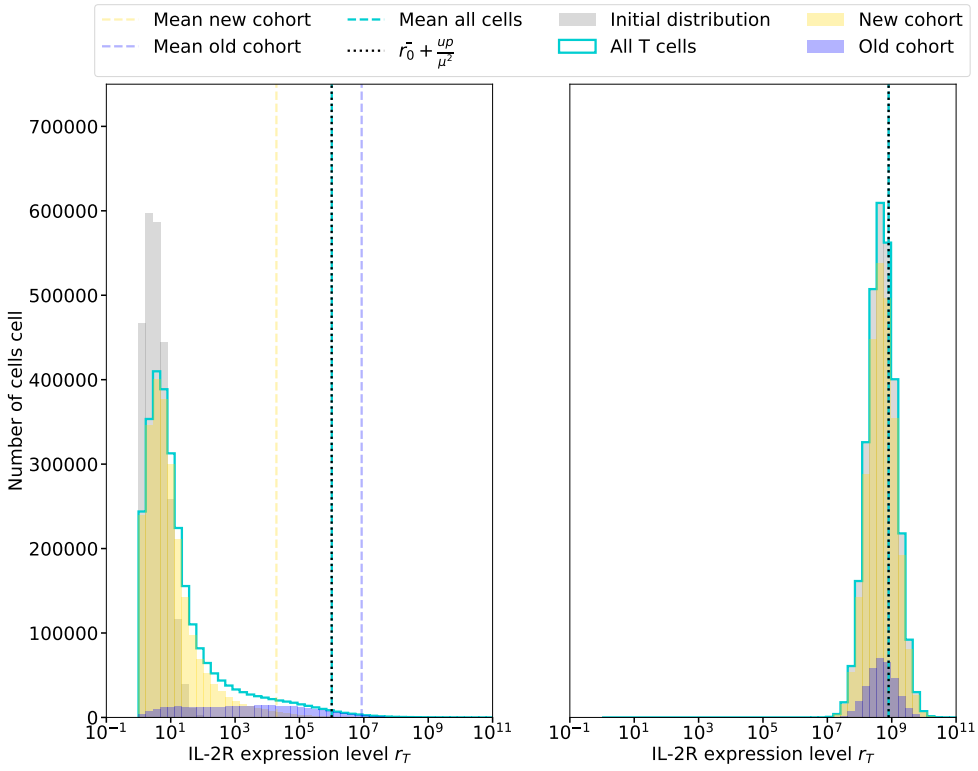


Figure 5.20: Distribution of IL-2R expression level of the cells that died during the simulation (in turquoise) when the average initial receptor is low ($m_0 = 1$, left) and large ($m_0 = 20$, right) compared to the potential upregulation $\bar{r_0} + \frac{up}{\mu^2}$. This distribution is compared with the initial log-normal distribution (grey) and the receptor distributions of the old and new cohorts (blue and yellow, respectively). When $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$ is low (left), the receptor distribution presents a long tail corresponding to old cells which lived long enough to upregulate massively their IL-2R expression. In the other regime (right), cells of the old and new cohorts died with similar IL-2R expression levels (closed to their initial level). The mean number of receptors expressed at death of the whole population (dashed turquoise line) is equal to $\bar{r_0} + \frac{up}{\mu^2}$ (dotted black line). In the egalitarian regime, the mean number of receptor of both cohorts is also equal to this value. The parameter values used for the simulations were: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\mu = 0.01$ /day, and $\alpha = 30$ cells/day, $\sigma_0 = 1$, $\Delta t = 1$ day and $t_{max} = 10^6$ days.

5.5 Death, activation and division

Let us now introduce a third type of stochastic event (see Figure 5.21): cell division (division rate λ). That is, at each time, a cell has a certain probability to be replaced by two cells with different attributes[1]. In combination to cell death (death rate μ) and cell activation (activation rate α), the population can grow exponentially or is driven to a stochastic steady state. We will study in more details the latter case.

5.5.1 Mathematical analysis

We start by specifying the average number of cells $\bar{N}(t)$ which satisfies the following differential equation

$$\frac{d\bar{N}}{dt} = \alpha + (\lambda - \mu)\bar{N}. \tag{5.45}$$

Considering the initial conditions $\bar{N}(t=0) = N(0)$, this differential equation can be solved and we obtain:

$$\bar{N}(t) = (N(0) - \frac{\alpha}{\mu - \lambda})e^{(\lambda-\mu)t} + \frac{\alpha}{\mu - \lambda}. \tag{5.46}$$

This equation shows that:

- If $\mu > \lambda$, there exists a steady state in which, on average, there are $N_\infty \equiv \frac{\alpha}{\mu - \lambda}$ cells in the pool.
- If $\lambda > \mu$, the population grows exponentially.

Note that if $\mu = \lambda$, we observe a linear growth and $\bar{N}(t) = \alpha t + N(0)$. In the rest of this section, we assume that $\mu > \lambda$ as we are interested in population growth control.

Consider a population of $N(t)$ cells alive at time t. Each cell has probability $\mu\Delta t$ of dying or probability $\lambda\Delta t$ of dividing before time $t + \Delta t$. In addition,

[1] This process is similar to a birth event in which at any time step, an individual has a certain probability to give birth to a new individual. However, here, we prefer talking about division, as the attribute values of both the mother and daughter cells will be updated after this event.

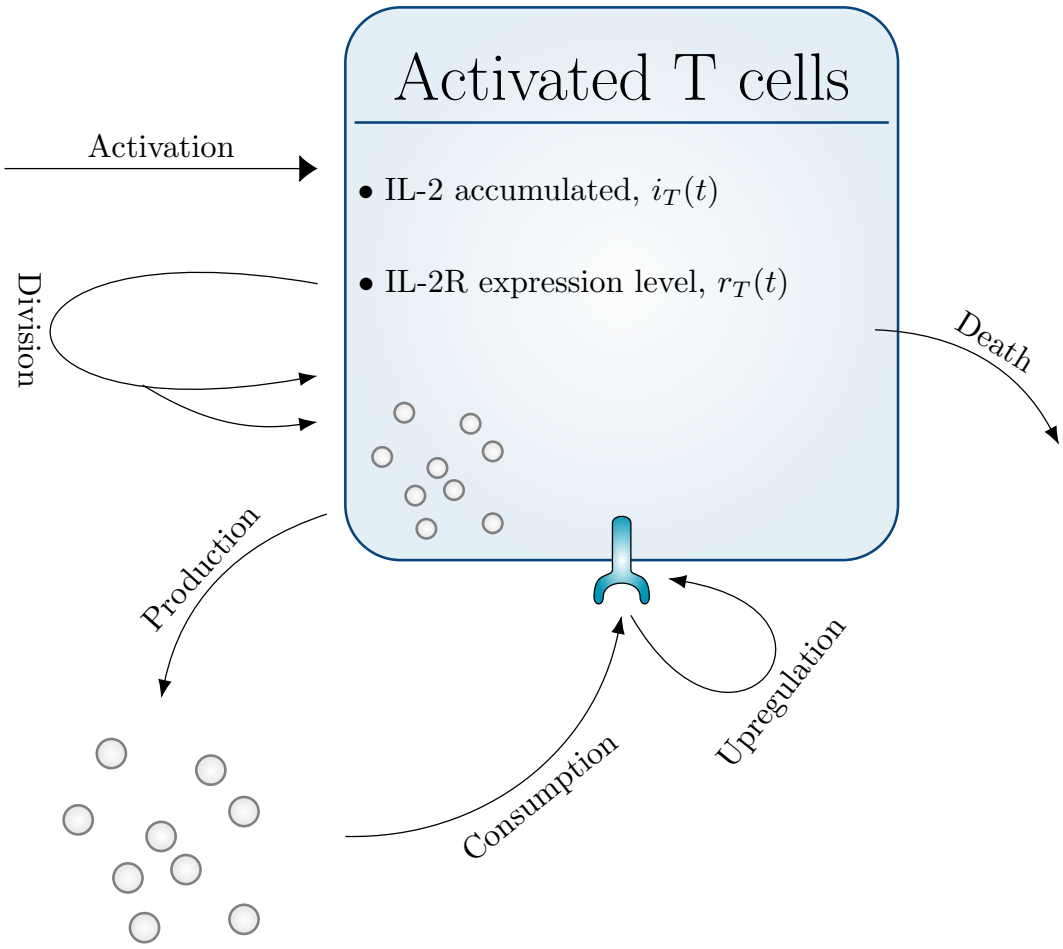


Figure 5.21: Activation, death and division process: activated T cells produce and consume IL-2. They also upregulate their IL-2R expression. New cells enter the pool at rate α and cells can die with death rate μ. Cells can divide with rate λ. Daughter cells inherit half of the receptors of their mother (but no IL-2) and enter immediately the activated cell pool.

during this small time interval $[t, t+\Delta t]$, a new cell can enter the pool (a naive cell becomes activated) with probability $\alpha\Delta t$. The probability that none of the events mentioned above happen during Δt is $1-\mu N(t)\Delta t-\lambda N(t)\Delta t-\alpha\Delta t$. As established in the previous section, the entrance of a new cell in the pool does not affect the total IL-2 absorbed $I(t)$. We assume that the two daughter cells receive half the IL-2R expression level of their mother but receive no IL-2. Thus, the total IL-2R expression level in the pool is conserved under division and the differential equation satisfied by $\bar{R}(t)$ is the equation (5.32a) determined in the previous section. However, since daughter cells do not receive any cytokine, the quantity of IL-2 absorbed by the mother cell before division is lost to the population: that is, I is decreased by this amount. Hence, as $\Delta t \to 0$,

$$I(t+\Delta t) = \begin{cases} I(t)+pN(t)\Delta t+\mathcal{O}(\Delta t^2) & \text{with probability } \rho, \\ \text{for each cell } T, & \\ I(t)+pN(t)\Delta t-i_T(t)+\mathcal{O}(\Delta t^2) & \text{with probability } (\mu+\lambda)\Delta t, \end{cases} \tag{5.47}$$

where $\rho = 1-\mu N(t)\Delta t-\lambda N(t)\Delta t$. The system of differential equations satisfied by $\bar{I}(t)$ and $\bar{R}(t)$ is:

$$\frac{d\bar{R}}{dt} = u\bar{I}-\mu\bar{R}+\alpha\bar{r_0}, \tag{5.48a}$$

$$\frac{d\bar{I}}{dt} = p\bar{N}-(\mu+\lambda)\bar{I}. \tag{5.48b}$$

Expression for $\bar{I}(t)$:

We solve equation (5.48b), with initial condition $\bar{I}(0)=0$. The solution of the homogeneous equation is:

$$I(t) = K_I e^{-(\mu+\lambda)t}, \tag{5.49}$$

where K_I is an unknown constant. We make use of the method of variation of parameters, considering K_I as a function of time, and we obtain:

$$K_I'(t) = p\bar{N}(t)e^{(\mu+\lambda)t}. \tag{5.50}$$

Integration yields

$$K_I(t) = \frac{p}{2\lambda}\left(N(0) - \frac{\alpha}{\mu-\lambda}\right)e^{2\lambda t} + \frac{\alpha p}{(\mu-\lambda)(\mu+\lambda)}e^{(\mu+\lambda)t} + C_I, \tag{5.51}$$

where C_I is a constant to be determined. Thus,

$$\begin{aligned}\bar{I}(t) &= K_I(t)e^{-(\lambda+\mu)t} \\ &= \frac{p}{2\lambda}\left(N(0) - \frac{\alpha}{\mu-\lambda}\right)e^{(\lambda-\mu)t} + \frac{\alpha p}{(\mu-\lambda)(\mu+\lambda)} + C_I e^{-(\lambda+\mu)t}.\end{aligned} \tag{5.52}$$

The initial condition, $\bar{I}(0) = 0$, gives an expression for C_I:

$$C_I = \frac{p\left(\alpha - (\mu+\lambda)N(0)\right)}{2\lambda(\mu+\lambda)}.$$

Finally, we obtain:

$$\bar{I}(t) = \frac{pl\left[\eta e^{(\lambda-\mu)t} + \zeta e^{-(\mu+\lambda)t} + 2\lambda\alpha\right]}{2\lambda(\mu^2-\lambda^2)}, \tag{5.53}$$

where we wrote

$$\begin{aligned}\eta =& (\lambda+\mu)\left((\mu-\lambda)N(0) - \alpha\right), \\ \zeta =& (\mu-\lambda)(\alpha - (\mu+\lambda)N(0)).\end{aligned}$$

Expression for $\bar{R}(t)$:

We now solve equation (5.48a). We are looking for a solution of the form

$$\bar{R}(t) = K_R(t)e^{-\mu t},$$

which substituted in (5.48a) yields:

$$\begin{aligned}K_R'(t) &= u\bar{I}(t)e^{\mu t} + \alpha\bar{r_0}e^{\mu t} \\ &= up\frac{\eta e^{\lambda t} + \zeta e^{-\lambda t} + 2\lambda\alpha e^{\mu t}}{2\lambda(\mu^2-\lambda^2)} + \alpha\bar{r_0}e^{\mu t}.\end{aligned} \tag{5.54}$$

After integration, we obtain

$$K_R(t) = \frac{\mathcal{A}e^{\lambda t} + \mathcal{B}e^{\mu t} + \mathcal{C}e^{-\lambda t}}{2\lambda^2(\mu^2 - \lambda^2)\mu} + C_R, \tag{5.55}$$

where we wrote

$$\begin{aligned}
\mathcal{A} &= pu\mu(\lambda + \mu)((\mu - \lambda)N(0) - \alpha),\\
\mathcal{B} &= 2\alpha\lambda^2((\mu^2 - \lambda^2)\bar{r_0} + pu),\\
\mathcal{C} &= pu\mu(\mu - \lambda)((\mu + \lambda)N(0) - \alpha),
\end{aligned}$$

and C_R is an integration constant that will be determined later. Thus,

$$\begin{aligned}
\bar{R}(t) &= K_R(t)e^{-\mu t}\\
&= \frac{\mathcal{A}e^{(\lambda-\mu)t} + \mathcal{B} + \mathcal{C}e^{-(\lambda+\mu)t}}{2\lambda^2(\mu^2 - \lambda^2)\mu} + C_R e^{-\mu t}.
\end{aligned} \tag{5.56}$$

Making use of the initial condition $\bar{R}(0) = R_0$, we find an expression for C_R

$$C_R = R_0 - \frac{\mathcal{A} + \mathcal{B} + \mathcal{C}}{2\lambda^2\mu(\mu^2 - \lambda^2)}, \tag{5.57}$$

which we substitute in (5.56) to obtain:

$$\bar{R}(t) = \frac{\mathcal{A}e^{(\lambda-\mu)t} + \mathcal{B} + \mathcal{C}e^{-(\lambda+\mu)t} - (\mathcal{A} + \mathcal{B} + \mathcal{C})e^{-\mu t}}{2\lambda^2\mu(\mu^2 - \lambda^2)} + R_0 e^{-\mu t}. \tag{5.58}$$

One can notice that when $t \to +\infty$, if $\mu > \lambda$, $\bar{R}(t)$ tends to a constant R_∞:

$$\begin{aligned}
\lim_{t\to+\infty} \bar{R}(t) &= \frac{\mathcal{B}}{2\lambda^2\mu(\mu^2 - \lambda^2)}\\
&= \frac{2\alpha\lambda^2((\mu^2 - \lambda^2)\bar{r_0} + pu)}{2\lambda^2\mu(\mu^2 - \lambda^2)}\\
&= \frac{\alpha}{\mu}\left(\bar{r_0} + \frac{pu}{\mu^2 - \lambda^2}\right) \equiv R_\infty.
\end{aligned} \tag{5.59}$$

Finally, let us write $\bar{R}(t)$ as a function of R_∞:

$$\begin{aligned}\bar{R}(t) &= \frac{\mathcal{A}e^{-\mu t}\left(e^{\lambda t}-1\right)+\mathcal{C}e^{-\mu t}\left(e^{-\lambda t}-1\right)}{2\lambda^2\mu(\mu^2-\lambda^2)}+(R_0-R_\infty)e^{-\mu t}+R_\infty \\ &= pu\left[\frac{e^{\lambda t}-1}{2\lambda^2}\left(N(0)-\frac{\alpha}{(\mu-\lambda)}\right)+\frac{e^{-\lambda t}-1}{2\lambda^2}\left(N(0)-\frac{\alpha}{(\mu+\lambda)}\right)\right]e^{-\mu t} \\ &\quad +(R_0-R_\infty)\,e^{-\mu t}+R_\infty.\end{aligned} \tag{5.60}$$

Note that when $\lambda \to 0$, we recover the expression obtained for $\bar{R}(t)$ in Section 5.4 (equation (5.35)).

Proof. Indeed, as $\lambda \to 0$, we have

$$\frac{e^{-\lambda t}-1}{2\lambda^2} = \frac{-t}{2\lambda}+\frac{t^2}{4}+o(1),$$

and

$$\frac{e^{\lambda t}-1}{2\lambda^2} = \frac{t}{2\lambda}+\frac{t^2}{4}+o(1),$$

as well as

$$\frac{\alpha}{\mu-\lambda} = \frac{\alpha}{\mu}\frac{1}{1-\frac{\lambda}{\mu}} = \frac{\alpha}{\mu}\left(1+\frac{\lambda}{\mu}+o(\lambda)\right),$$

and

$$\frac{\alpha}{\mu+\lambda} = \frac{\alpha}{\mu}\frac{1}{1+\frac{\lambda}{\mu}} = \frac{\alpha}{\mu}\left(1-\frac{\lambda}{\mu}+o(\lambda)\right).$$

Thus, as $\lambda \to 0$, we obtain

$$\begin{aligned}&\frac{e^{\lambda t}-1}{2\lambda^2}\left(N(0)-\frac{\alpha}{(\mu-\lambda)}\right)+\frac{e^{-\lambda t}-1}{2\lambda^2}\left(N(0)-\frac{\alpha}{(\mu+\lambda)}\right) \\ &= -\frac{\alpha t}{\mu^2}+\left(N(0)-\frac{\alpha}{\mu}\right)\frac{t^2}{2}+o(\lambda).\end{aligned}$$

Finally,

$$\lim_{\lambda\to 0} R_\infty = \frac{\alpha}{\mu}\left(\bar{r_0}+\frac{up}{\mu^2}\right),$$

which is the expression obtained for R_∞ in the previous section in equation (5.36). □

Dynamics of $i_T(t)$ and $r_T(t)$

Heuristically, the two attributes of a cell T, $i_T(t)$ and $r_T(t)$, satisfy the system of ODEs (5.37). However, the initial conditions are different. Let us write t_T^{in}, the time at which the cell T entered the cell pool, whether it has been activated or it is the daughter of a cell, T_{mother}, that just divided. Then, we have:

$$i_T(t_T^{in}) = 0, \tag{5.61a}$$

$$\begin{aligned} r_T(t_T^{in}) &\sim \log \mathcal{N}(m_0, \sigma_0^2) && \text{if the cell just became activated,} \\ &= r_{T_{\text{mother}}}(t_T^{in})/2 && \text{otherwise.} \end{aligned} \tag{5.61b}$$

System (5.37) can be decoupled to obtain system (5.39). Similarly to the previous model, the expression of $\bar{R}(t)$ is also a sum of a constant and an exponential. Thus, it may not be possible to obtain expressions for $i_T(t)$ and $r_T(t)$ in the general case.

Distribution of time between cohort switches

As for the study of the model without division, let us distinguish the new and old cell cohorts. As each cell can divide into two cells, we assume that daughter cells remain in the cohort of their mother: a cell of the old (respectively, new) cohort gives two cells that belong to the old (respectively, new) cohort. That way, the cells of the old cohort follow a pure birth and death process with death rate μ and birth rate λ (see Section 2.7.3). As we assumed $\mu > \lambda$, the extinction of this cohort is certain. At steady state, the time to extinction of such population follows a Gumbel distribution with parameters $m = \frac{1}{\mu-\lambda}\log(N_\infty(\frac{\mu-\lambda}{\mu}))$ and $\beta = \frac{\bar{\tau}-m}{\gamma_e}$, where γ_e is the Euler–Mascheroni constant and $\bar{\tau}$ is given by equation (2.37) in Section 2.7.3. Such distribution also describes the time between two cohort switches (see Figure 5.22).

5.5.2 Numerical observations

Once again, we resort to numerical simulations to understand the stochastic dynamics of the system. Simulations of the model with division were carried out for varying values of the division rate λ both in the egalitarian ($m_0 = 20$,

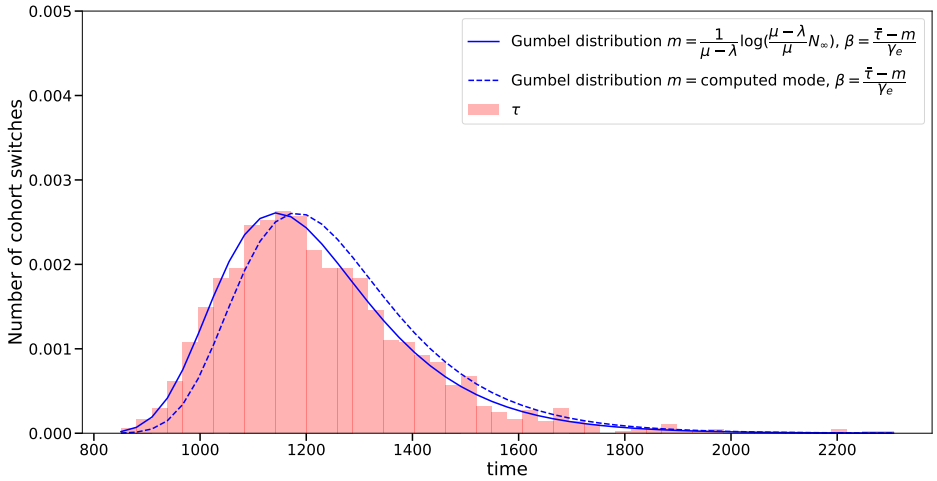


Figure 5.22: In the agent-based model with division, the time between two cohort switches (which is the time to extinction of a pure birth and death process) follows a Gumbel (blue solid line) distribution with parameter $m = \frac{1}{\mu-\lambda}\log(N_\infty(\frac{\mu-\lambda}{\mu}))$ and $\beta = \frac{\bar{\tau}-m}{\gamma_e}$, where γ_e is the Euler–Mascheroni constant and $\bar{\tau}$ is given by equation (2.37) in Section 2.7.3. The model was simulated with the following parameter values: $\mu = 0.01$ /day, $\lambda = 0.003$ /day, $\alpha = 30$ cells/day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\Delta t = 1$ day, $t_{max} = 2 \times 10^6$ days, $N(0) = \frac{\alpha}{\mu-\lambda}$, $\sigma_0 = 1$ and $m_0 = 1$. Note that $N(0) = N_\infty$ was too large to compute $\bar{\tau}$ making use of the expression introduced in Section 2.7.3. Here, to plot the Gumbel distribution, the numerical value of $\bar{\tau}$ (mean of the distribution in red) was used. The Gumbel distribution with parameter m being equal to the actual mode of this distribution has also been added (dashed line).

$\sigma_0 = 1$) and gerontocracy ($m_0 = 1$, $\sigma_0 = 1$) cases. Note that these two regimes are now defined by comparison of $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$ with $\frac{up}{\mu^2 - \lambda^2}$ [1]. We fixed the other parameters to the following values: $\mu = 0.01$ /day, $\alpha = 30$ cells/day, $u = 10$ $[r_T]/[i_T]$/day and $p = 10$ $[i_T]$/day/cell. The division rate, λ, takes the values 0.001 /day, 0.003 /day and 0.007 /day, corresponding to low, moderate and high division rate, respectively.

We looked at the evolution of the receptor distribution of each cohort and observed three types of population events:

A - A cohort switch immediately followed by a sudden receptor upregulation of the cells of the new cohort (see Figure 5.25 below),

B - A cohort switch without any sudden receptor upregulation of the cells of the new cohort (see Figure 5.26 below),

C - A sudden receptor upregulation of the cells of the new cohort without cohort switch (see Figure 5.27 below).

These population events affect the overall cell population dynamics. In some cases, we may observe that both cohorts are characterised by very different receptor distributions (typically the cells of the old cohort express a receptor level significantly higher than the cells of the new cohort). However, in other cases, the receptor distributions of the two cohorts significantly overlap. That is, for most parameter values tested, the two cases of homeostasis can be observed:

1. The two cell cohorts have distinct receptor distributions (a small distribution overlap may be observed),

2. The two cell cohorts are mixed and indistinguishable (receptor distributions overlap significantly or completely).

In the model without division, in the gerontocracy regime ($\sigma_0 = 1$), the homeostasis of type 1 was the only one observed and was characterised by a receptor

[1] I chose to keep the names *gerontocracy* for $\bar{r_0} \ll \frac{up}{\mu^2 - \lambda^2}$ and *egalitarian regime* for $\bar{r_0} \gg \frac{up}{\mu^2 - \lambda^2}$ by analogy with the model without division. However, we will see that, the oldest cells are not always the cells expressing the highest number of receptors anymore.

distribution of the entire cell population which was bimodal, each mode matching the mode of the receptor distribution of a cohort. On the contrary, in the egalitarian regime, the receptor distribution of the cell population was unimodal because the cohort distributions overlapped. Here, we will see that the dynamics is more complicated as cohort distributions may be bimodal (and thus one cohort could be responsible for the multimodality of the receptor distribution of the entire cell population). Note that the definitions of the two types of homeostasis and the three population events are purely qualitative, based on the observation of the evolution of the scatterplots (see output description in Appendix J and Figures 5.24, 5.25, 5.26 and 5.27,below).

Let us now explore in more detail the two different regimes: $\bar{r_0} \ll \frac{up}{\mu^2 - \lambda^2}$ and $\bar{r_0} \gg \frac{up}{\mu^2 - \lambda^2}$. Note that figures illustrating this section come from simulations with a fixed time step $\Delta t = 10$. A simulation with this time step, while seemingly large, displays the same dynamics as a simulation with a smaller Δt (see Appendix I). In the following qualitative descriptions, we will call *family* the set of an initial cell which entered the pool after activation, and its descendants. The first cell is said to be *generation* 0. When the cell divides, its daughters are *generation* 1. The daughters of the daughters are *generation* 2 and so on. A family goes extinct when all the descendants of the initial cell died. We will refer to as *last descendant* the cells at the end of the family tree (since a cell that divided does not exist anymore, the last descendant are the only cells currently alive in the simulation). Finally, we refer to as *ancestors* of a cell T, all the cells of the previous generations of the family of T that lead to this cell. Figure 5.23 recapitulates these definitions. When mentioning the age of a family, we refer to the average generation number of the last descendants.

Egalitarian regime ($\bar{r_0} \gg \frac{up}{\mu^2 - \lambda^2}$)

In the egalitarian case, for the three levels of division, we do not observe any receptor upregulation. The two cell cohorts always express about the same level of receptors (homoeostasis of type 2), and cohort switches do not change the receptor distribution (we observe a sequence of population events of type B, as described

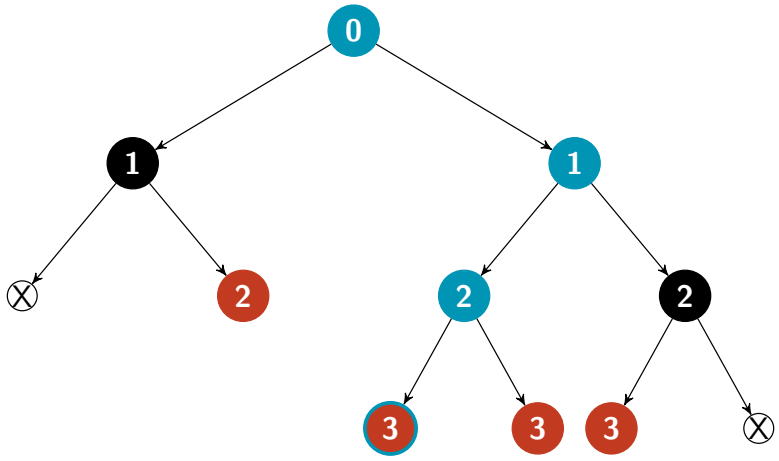


Figure 5.23: Example of a family of cells. The number inside each cell of the family indicates its generation. The initial cell entered in the cell pool with rate α (generation 0). It then divided with rate λ to give two daughter cells (generation 1). These cells, in turn, divided and their daughters died (crossed circles), remained in the cell pool or divided as well. The cells at the end of the tree (that are still alive) are the last descendants (red). The ancestors of the last descendant circled in blue are indicated in blue.

before). More interestingly, when the division rate is high ($\lambda = 0.007$), we observe a downward shift of the receptor distribution of each cell cohort (the cells of the old cohort can even express a lower receptor level than cells of the new cohort, see Figure 5.24). This is due to the fact that each daughter cell only receives half of the mother cell receptor copy number. At each division, the receptor level is divided by two: as there is very little receptor upregulation, the older the cell family is, the lower the receptor level expressed by the last descendants.

Gerontocracy ($\bar{r_0} \ll \frac{up}{\mu^2 - \lambda^2}$)

In the gerontocracy case, all population events can occur and the two homeostasis types are observed (see Table 5.1 for an example). At low division rate ($\lambda = 0.001$ /day for $\mu = 0.01$ /day), we see that, most of the time, the receptor distributions of the two cohorts are distinct, and we mostly observe a sequence of population events of type A and C. At moderate division rate ($\lambda = 0.003$ /day for $\mu = 0.01$ /day), we seem to alternate between sequences of population events of type A and C (A-C-A-C cycle) and sequences of events of type B and C (B-C-B-C cycle).

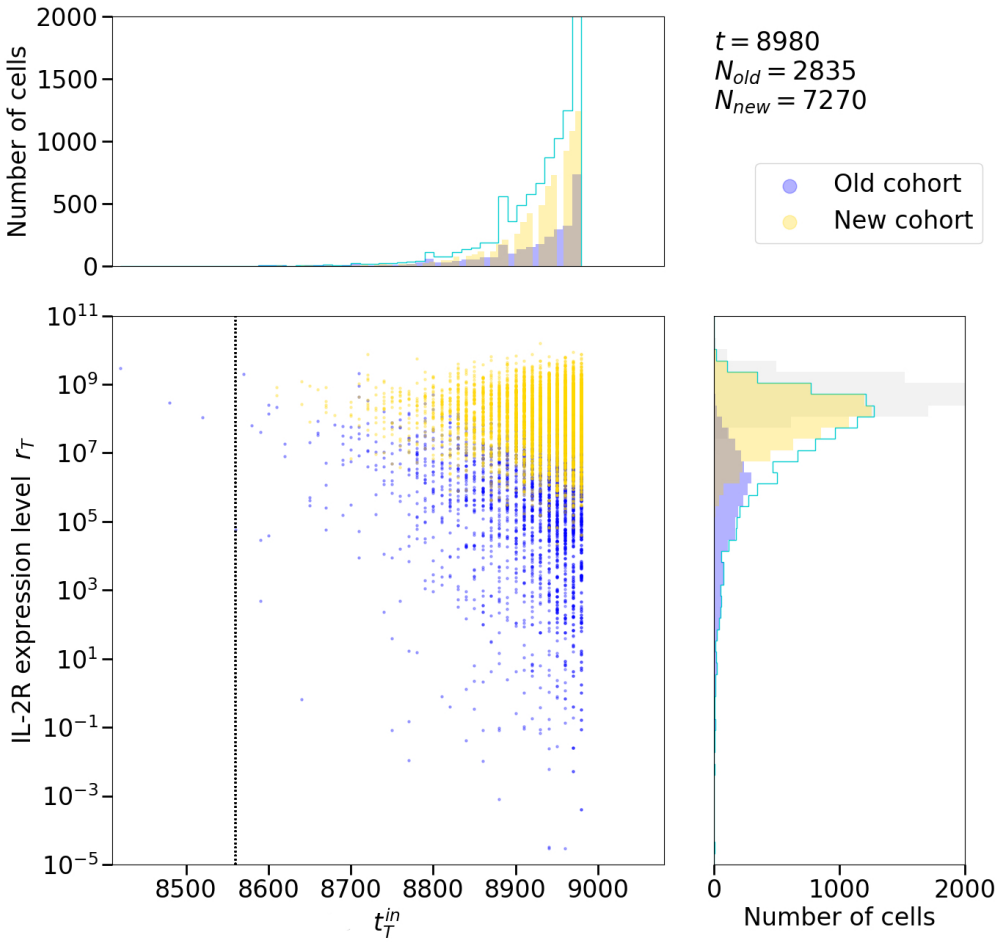


Figure 5.24: Scatterplot at $t = 8980$ of individual receptor levels as a function of the time at which the cell entered the pool, with joint distributions (receptor distribution is on the right), for a simulation with the following parameter values: $m_0 = 20$, $\sigma_0 = 1$, $\mu = 0.01$ /day, $\lambda = 0.007$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu - \lambda}$ and $\Delta t = 10$ days. The cells of the old cohort are indicated in blue, cells of the new cohort are in yellow. We observe that both receptor distributions are wider than the initial one (in grey), and shifted down. The two distributions also overlap significantly and the two sub-populations appear to be mixed (homeostasis of type 2). The last cohort switch is indicated by the dotted line. This cohort switch did not induce any visible receptor upregulation of any of the cohort.

$\lambda(/day)$	Events
0.001	ACACAAACAAACACA
0.003	ACACCBCBCBBCBCCBCB
0.007	CBCBCCA

Table 5.1: Sequence of events observed during a simulation of the agent-based model with division, death and activation, when $\sigma_0 = 1$, $m_0 = 1$ and $\mu = 0.01$ /day. Other parameter values were $\Delta_t = 10$ days, $t_{max} = 10000$ days, $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day and $N(0) = N_\infty = \frac{\alpha}{\mu - \lambda}$. The letters in red indicate when the two cohorts were undistinguishable (homeostasis of type 2). Letters in black indicate that the two cohort had separated receptor distributions. Note that sometimes, it is not trivial to distinguish which type of homeostasis or population event we are observing. This table is provided for illustration purposes.

Usually, during a A-C-A-C cycle, the receptor distributions of the two cohorts are distinct. However, if the cells of the new cohort managed to upregulate their receptor levels above a certain threshold (which need to be determined), we might switch to a B-C-B-C cycle. In such cycle, we may alternate between the two types of homeostasis. Finally, at high division rate ($\lambda = 0.007$ /day for $\mu = 0.01$ /day), we mostly observe a sequence of population events of type B and C. The receptor distribution of both cohorts (but especially the old cohort) can be significantly wider than the initial distribution (see Figure 5.26). This is due to the fact that cell families live up to a large generation number. For $\lambda = 0.007$ /day (both in the gerontocracy and egalitarian regimes), we can go up to generation 50. Some cells managed to upregulate massively their receptor expression level resulting in descendants with a very large number of receptors (compared to the mean receptor copy number of its cohort). Other cells express very few receptors, because their ancestors did not upregulate enough their receptor expression level to compensate the fact that each daughter receives only half the receptor level of their mother.

The A-C-A-C or B-C-B-C cycles may be explained as follows. Once in a while, a cell, or a group of cells, of the old cohort will die or divide which results in the cells of the old cohort not expressing enough receptors to deprive of cytokine the cells of the new cohort. Consequently, cells of the new cohort suddenly gain access to IL-2 and upregulate massively their receptor expression level (event of type

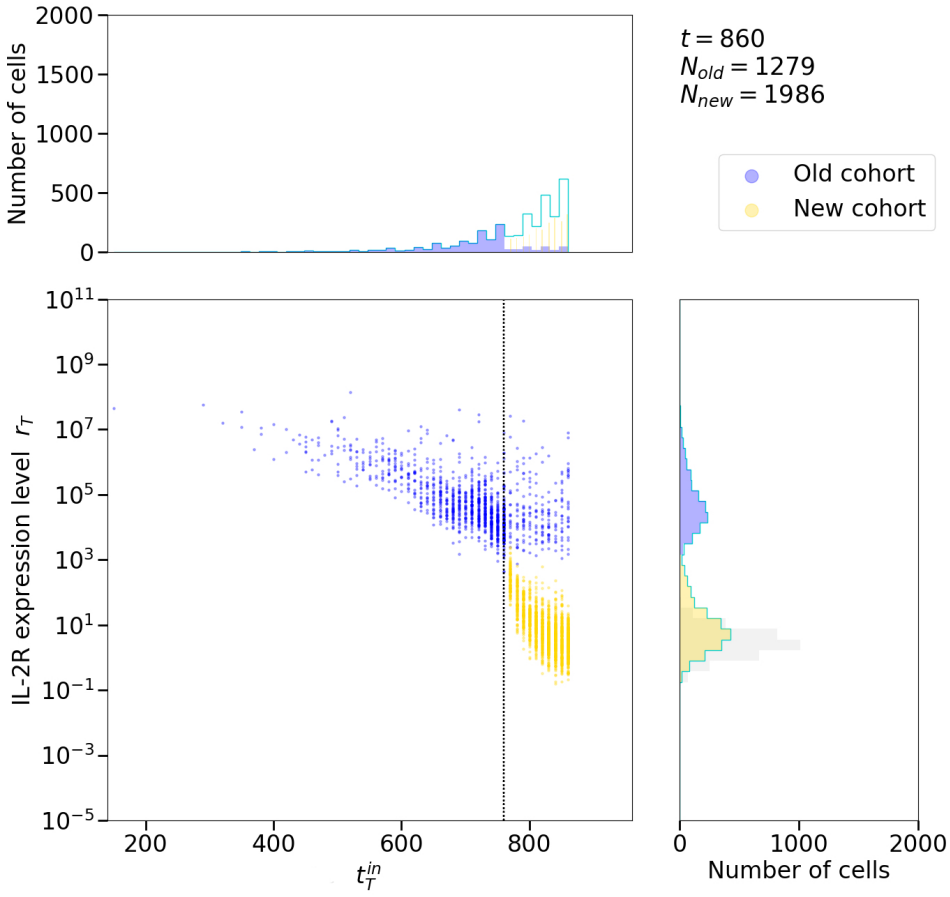


Figure 5.25: Scatterplot at $t = 860$ of individual receptor levels as a function of the time at which the cell entered the pool, with joint distributions (receptor distribution is on the right), for a simulation with the following parameter values: $m_0 = 1$, $\sigma_0 = 1$, $\mu = 0.01$ /day, $\lambda = 0.001$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu - \lambda}$ and $\Delta t = 10$ days. The cells of the old cohort are indicated in blue, cells of the new cohort are in yellow. The vertical black dotted line indicates the last cohort switch. We observe that the cells of the old cohort that did not divide yet (cells on the left of the black line) express a higher receptor expression level than cells of the new cohort (there is a "discontinuity" in the scatterplot). That is, the cohort switch induced a receptor upregulation of the previous new cohort (current old cohort) (population event of type A). Daughters of the cells of the old cohort (blue dots on the right of the dotted line) still express high receptor levels. Cells of the new and the old cohort express very different receptor expression levels and are well distinct on the scatter plot (homeostasis of type 1). The receptor distribution of the whole cell population is bimodal, each mode matching the mode of the receptor distribution of one cohort.

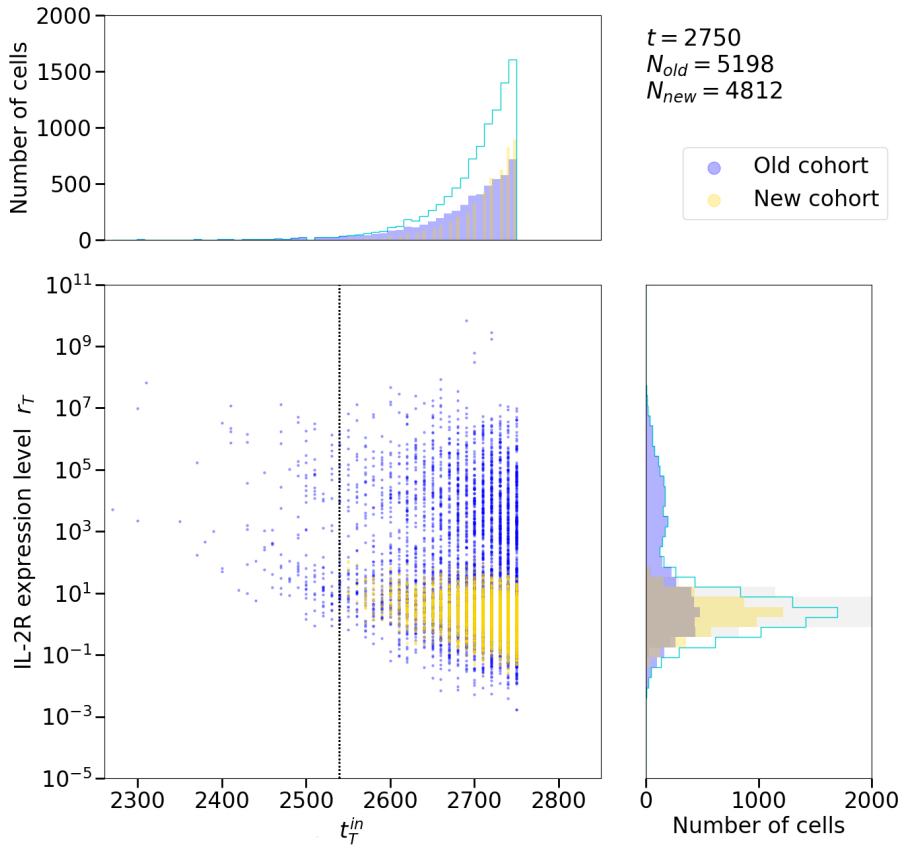


Figure 5.26: Scatterplot at $t = 2750$ of individual receptor levels as a function of the time at which the cell entered the pool, with joint distributions (receptor distribution is on the right), for a simulation with the following parameter values: $m_0 = 1$, $\sigma_0 = 1$, $\mu = 0.01$ /day, $\lambda = 0.007$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu - \lambda}$ and $\Delta t = 10$ days. The cells of the old cohort are indicated in blue, cells of the new cohort are in yellow. The last cohort switch is indicated by the vertical black dotted line. We observe that some cells of the old cohort that did not divide yet (left of the black line), express about the same receptor expression level as cells of the new cohort (there is no "discontinuity"). That is, the last cohort switch did not induce any significant receptor upregulation (population event of type B). Despite the fact that the receptor distributions of the two cohort do not have the same mode, they overlap significantly (homeostasis of type 2). We also note that in this specific case (high division rate), the receptor distribution is wider than the initial distribution (grey): some cell families lived long enough so that the last descendant expresses many (because its ancestors had time to upregulate massively their receptor expression) or very few receptors (because each descendant only received half the receptors of the mother and could not consume enough IL-2 to upregulate).

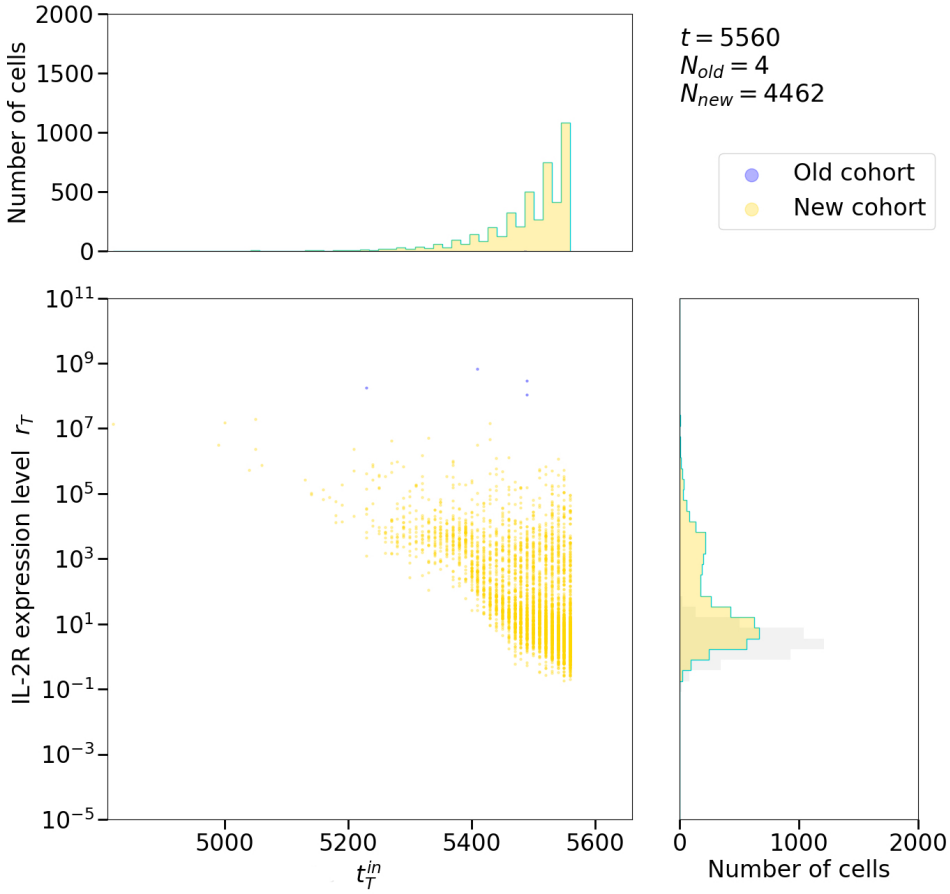


Figure 5.27: Scatterplot at $t = 5560$ of individual receptor levels as a function of the time at which the cell entered the pool, with joint distributions (receptor distribution is on the right), for a simulation with the following parameter values: $m_0 = 1$, $\sigma_0 = 1$, $\mu = 0.01$ /day, $\lambda = 0.003$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu - \lambda}$ and $\Delta t = 10$ days. The cells of the old cohort are indicated in blue, cells of the new cohort are in yellow. We observe that the cells of the new cohort are upregulating their receptor expression level before the cohort switch (there are four cells left in the old cohort): this is an illustration of population event of type C. The upregulation started around $t = 5400$, where there is a "discontinuity" in the scatterplot. This results in a bimodal and wider receptor distribution of the new cohort. Note that the cells of the old cohort still express a significantly higher receptor level than most the cells of the new cohort: the two cohorts are distinct (homeostasis of type 2).

C). When the last cell of the old cohort dies, if this cell was expressing enough receptors to impact the cells of the new cohort, then the cells of the new cohort increase their receptor expression level again (event of type A). If the last cell of the old cohort was not impacting the cells of the new cohort, then its death does not affect them significantly: we observe a cohort switch without sudden receptor upregulation (event of type B).

Let us point out that a bimodal receptor distribution of the entire cell population does not always mean that there are two distinct cohorts: during a population event of type C, both modes correspond to the modes of the distribution of the new cohort (see Figure 5.27). After a succession of population events of type B and C, the two modes can be caused by the old cohort distribution (see Figure 5.26).

Finally, we note that for the three division levels, the population variable $R(t)$ displays sharp drops that seem to coincide with a sudden receptor upregulation of the cells of the new cohort (population event of type A or C). These drops, however, are less significant (smaller amplitude) than in absence of division. I conjecture that the variations of $R(t)$ are more gradual due to the sequence of events C and A (A-C-A-C cycle) or B and C (B-C-B-C cycle). Indeed, in a B-C-B-C cycle, $R(t)$ suddenly decreases at events C only. I conjecture that, since during an event of type C some cells of the old cohort remain in the system, the decrease is smaller than if all the cells of this cohort had died (to be confirmed quantitatively). During a A-C-A-C cycle, $R(t)$ decreases at event C (when the cells of the old cohort suddenly stop depriving the cells of the new cohort) and at event A (when the last cell of the old cohort dies and cytokine become available to cells of the new cohort). Between each of these events, as cells of the new cohort upregulate their receptor expression level (first very suddenly and then more gradually), $R(t)$ is brought back to a value around R_∞. As a result, instead of decreasing significantly only at cohort switches, $R(t)$ decreases in two steps (and comes back to its steady state value between each decrease), resulting in even smaller drops.

5.6 Extending the IL-2 competition model

In this section, I propose two extensions of the agent-based model described in Section 5.5. While I provide a mathematical framework of the extended models, I did not conduct any substantial mathematical analysis of these systems. A Python code of the extended models is provided at https://github.com/leasta/ABM_thesis (see Appendix J). Let us start by describing the process of cytokine deprivation-induced death (or commonly called starvation) of activated conventional T cells.

5.6.1 Cellular events depending on cell variables: starvation

In the previous models, the population rules (death, activation and division) affected the IL-2R distribution by removing or adding cells to the population compartment. On the contrary, the IL-2 secretion and uptake, and the upregulation of its receptor, had no consequence on the population dynamics. Thus, to evaluate the impact of the IL-2R distribution on the population dynamics, cellular events that depend on the individual cell attributes (here i_T and r_T) must be considered. For instance, a cell deprived of cytokine may have an increased probability of dying.

Cytokine deprivation (by restriction of the concentration of available extracellular proteins or by limitation of the number of receptors expressed by the cells) is a well-studied biological mechanism, which occurs naturally or can be created for a biological experiment. For instance, at the end of an infection, decreased IL-2 concentration induces apoptosis of the majority of conventional T cells (cytokine withdrawal-induced cell death (CWID))(Duke & Cohen, 1986; Larsen *et al.*, 2017; Strasser & Pellegrini, 2004), thus bringing back its population to homeostatic levels. More related to the models of this chapter, IL-2 deprivation-induced apoptosis is the mechanism by which regulatory T cells maintain conventional T cell population to physiological levels at homeostasis (Pandiyan *et al.*, 2007). In a lab, to investigate the role of a specific protein, biologists may block its production or prevent its receptor binding by treating cells *in vitro* or mice *in vivo* with certain antibodies. They can also use mutant mice that do not produce this particular protein or its associated receptor. Both mechanisms artificially induce

cytokine deprivation and thus may lead to the death of essential cell populations. For instance, Oliveira (2013) proposes a method to assess, *in vitro*, human T cell apoptosis induced by IL-2 starvation, mimicking the end phase of an immune response.

Following these examples, it is natural to consider cytokine deprivation-induced cell death when designing models of IL-2 competition. Consider the model of Section 5.5 and now suppose that a cell, T, that absorbed less IL-2 than a certain threshold, θ_{starv}, has a higher probability of dying. That is, during a small time interval $[t, t+\Delta t]$, each cell of the system may die with probability $\mu\Delta t$. However, each cell, T, that satisfies

$$i_T(t) < \theta_{starv}, \tag{5.62}$$

may die with probability $(\mu+\nu)\Delta t$, where we wrote ν the starvation rate.

The mathematical analysis of models including cellular events dependent on cell's attributes, such as this starvation mechanism, may become very complicated, if not intractable. The dynamics of the model with starvation is not obvious to conjecture, especially since the dynamics of the model described in Section 5.5 is already not simple. Let us, once again, discriminate the cell population in two cohorts, as described in Section 5.4 and, to simplify the dynamics, consider the gerontocracy regime in absence of division. Cells of the old cohort deprive cells of the new cohort of IL-2. As a result, the latter cells have a higher probability of dying, which reduces the size of the new cohort. This decreased number of cells in the new cohort may have two consequences: fewer cells are competing for IL-2 and fewer cells live long enough to be part of the old cohort. However, a reduced old cohort may lead to more IL-2 available for the other cells to compete for. Consequently, cells of the new cohort may upregulate their IL-2R faster and thus stop starving. The next old cohort will then be larger and less IL-2 will be available which will trigger starvation of the new cohort again. This "circular" conjectured dynamics may be highly dependent on the choice of θ_{starv} and the other parameter values. Indeed, θ_{starv} may be chosen such as the cells of the old cohort never starve but cells of the new cohort always do. Further investigation, including numerical analyses, is needed.

5.6.2 Competition model between regulatory and activated conventional T cells

As mentioned in the introduction of this chapter, self-activated T cells are not the only cells consuming IL-2 in the periphery at homeostasis. Regulatory T cells also express the IL-2R and the competition for IL-2 between the two cells populations is the key to prevent autoimmunity (Höfer *et al.*, 2012). Let us, now, consider an agent-based model of the competition for IL-2 between regulatory and self-reactive T cells. We keep the notation of the one-population models ($N(t)$ for the number of cells at time t, p for production rate, u for upregulation rate, *etc.*...) and distinguish the parameters for regulatory or conventional T cells by adding an index, r and c, respectively. Hence, $N_c(t)$ and $N_r(t)$ denote the number of conventional self-activated and regulatory T cells at time t, respectively; μ_c (respectively, μ_r) is the death rate of conventional T cells (respectively, regulatory T cells), *etc.*... Both cell types upregulate their IL-2R expression level proportionally to their IL-2 consumption (cytokine accumulated), but only self-reactive conventional T cells produce the cytokine. For each cell, T, we define the time-dependent attributes $r_T(t)$ and $i_T(t)$ which correspond, respectively, to the number of receptor and quantity of accumulated IL-2 of cell T. These attributes satisfy the following system of equations, for any cell T:

$$\frac{di_T}{dt} = c_* p N_c(t) \frac{r_T(t)}{R(t)}, \tag{5.63a}$$

$$\frac{dr_T}{dt} = u_* i_T(t), \tag{5.63b}$$

$$\tag{5.63c}$$

where the index $*$ must be replaced by c if T is a conventional self-reactive cell or r if T is a regulatory T cell. To take into account the possibility that both cells may not consume IL-2 at the same speed (for instance, one cell type may recycle its receptors slower), we introduced the parameters c_c and c_r which correspond to consumption rates. Note that, this time, the definition of $R(t)$ slightly changes:

$$I(t) = \sum_{T=1}^{N_c(t)+N_r(t)} i_T(t), \tag{5.64a}$$

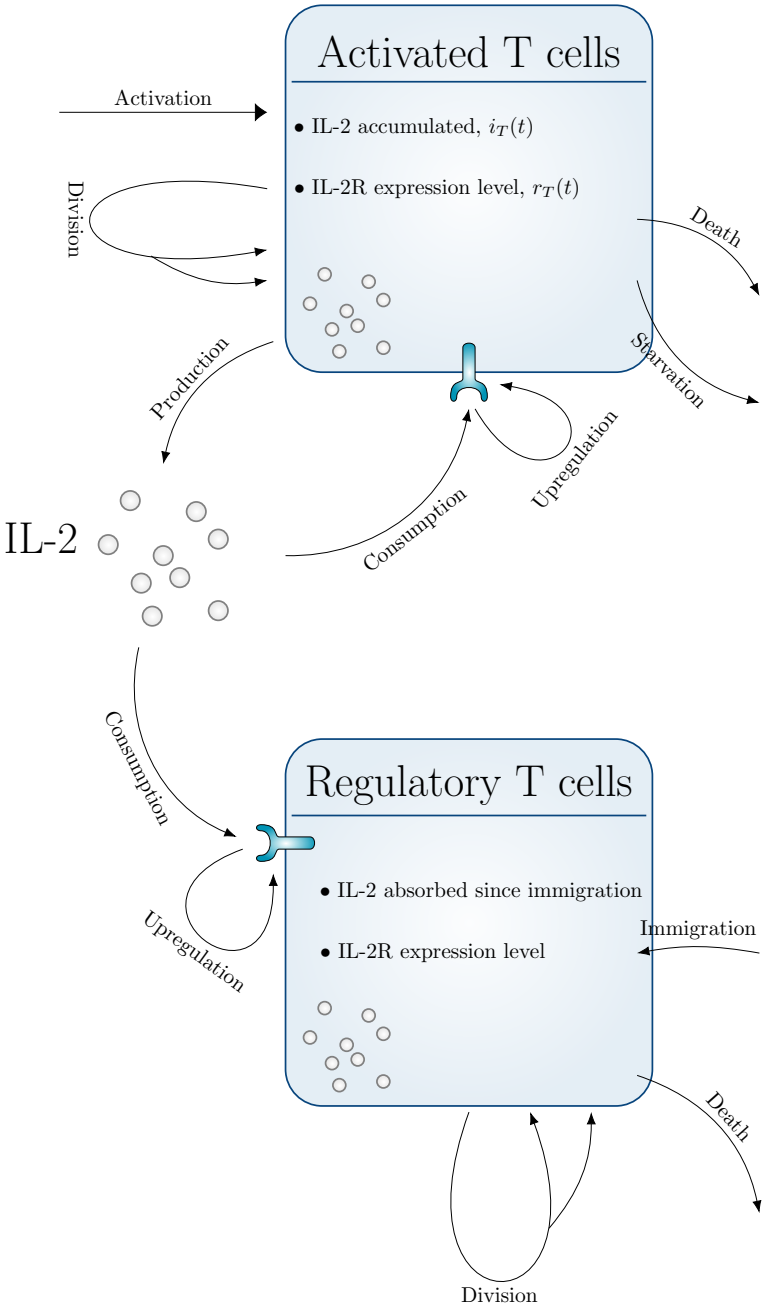


Figure 5.28: Scheme of the agent-based model of the competition for IL-2 between regulatory (bottom) and self-activated conventional (top) T cells. IL-2 production and absorption, and IL-2R upregulation are deterministic events. Division of regulatory T cells and starvation of conventional T cells are stochastic and dependent on the cell's attributes. Other cellular events are stochastic and do not depend on the cell considered.

$$R(t) = c_c \sum_{T=1}^{N_c(t)} r_T(t) + c_r \sum_{T=1}^{N_r(t)} r_T(t). \tag{5.64b}$$

Finally, for any cell, T, entering the pool at $t = t_T^{in}$ (by activation or from the division of a mother cell, T_{mother}), the initial conditions of system (5.63) are:

$$i_T(t_T^{in}) = 0, \tag{5.65a}$$

$$\begin{aligned} r_T(t_T^{in}) &\sim \log \mathcal{N}(m_{0*}, \sigma_{0*}^2) && \text{if the cell just became activated,} \\ &= r_{T_{\text{mother}}}(t_T^{in})/2 && \text{otherwise,} \end{aligned} \tag{5.65b}$$

where, once again, the index $*$ must be replaced by c if T is a conventional self-reactive cell or r if T is a regulatory T cell.

Hence, the IL-2R/IL-2 dynamics for both populations of cells is very similar to the other models of this chapter. However, population rules for regulatory T cells may be different from the rules for self-reactive conventional cells. We assume that, during a small time interval $(t, t + \Delta t)$, new cells, from any of these populations, may enter the pool with probability $\alpha_* \Delta t$. Each cell has probability $\mu_* \Delta t$ of dying. A conventional T cell may divide or starve as in Sections 5.5 and 5.6.1. Several studies observed that, in absence of IL-2, the population of regulatory T cells falls below physiological levels (Abbas *et al.*, 2018; Humrich *et al.*, 2010; Létourneau *et al.*, 2009; Sakaguchi *et al.*, 1995). We thus assume that a regulatory T cell may need a certain amount of IL-2, θ_{div} to divide. That is, if for any regulatory T cell, T, we have

$$i_T(t) > \theta_{div}, \tag{5.66}$$

then the cell T may divide with probability $\lambda_r \Delta t$ in the small time interval $(t, t + \Delta t)$. We assume that regulatory T cells do not starve. The competition model is recapitulated in Figure 5.28.

The initial goal of this model is to understand in what settings we observe the two possible scenarios: proliferation of self-reactive conventional T cell subsets (immune response), or their extermination (or, at least, regulation) by being starved of IL-2 by regulatory T cells (homeostasis). A special attention should

be given to the receptor distribution of self-reactive cells in presence or absence of regulatory T cells. A distinct receptor distribution shape corresponding to a particular event or state of the system, could be a helpful information for biologists. Finally, when the population of self-reactive cells escapes the control of regulatory T cells, the “time of escape" should be investigated, measured and better understood.

5.7 Alternative system: modelling the competition for IL-7 between ILCs and T cells

Competition for a cytokine between two cell populations, such as the competition for IL-2 between two T cell sub-populations, is common in immunology. Another example is the competition for IL-7 between T cells and innate lymphoid cells (ILCs) in lymph nodes (Cherrier *et al.*, 2020; Martin *et al.*, 2017; Sheikh & Abraham, 2019). ILCs are tissue-resident cells that mirror the functions of T cells (Eberl *et al.*, 2015). These cells react more promptly than T cells to pathogens (Kang & Coles, 2012; Vivier *et al.*, 2018) but do not proliferate in reaction to the invasion. Instead, they produce a range of cytokines that direct the developing immune response into one that is adapted to the original insult (pathogen type and dose) (Eberl *et al.*, 2015). ILCs have been classified into several subsets (Sonnenberg & Hepworth, 2019) and this section focuses on ILC2s and ILC3s which require IL-7 for their development and maintenance at homeostasis (Cherrier *et al.*, 2020; Kang & Coles, 2012). IL-7 is a critical survival factor for T cells produced by a small number of specialised tissue cells (Cherrier *et al.*, 2020). It binds to the IL-7R which is composed of the IL-7Rα chain and the γ_c chain. IL-7/IL-7R binding induces downregulation of IL-7Rα[1] expression in T cells. However, it appears that ILCs are resistant to this downregulation (Martin *et al.*, 2017). Hence, a small number of ILCs can seriously limit the availability of IL-7 for T cells.

[1]Note that in this section, we will not distinguish the full IL-7 receptor and the IL-7Rα chain. We will write IL-7R or IL-7Rα to denote the same thing.

In this section, we discuss the possibility of adapting the system proposed by Park *et al.* (2019) and Reynolds *et al.* (2013) to model competition for IL-7 between T cells, into an agent-based model similar to the ones presented in the rest of this chapter. We modify this system to account for the competition with ILCs. Note that Mathematica (Wolfram Research, Inc., 2019) has sometimes been used to speed-up cumbersome steady-state computations of this section. First, let us recapitulate the average dynamics described in Park *et al.* (2019) and Reynolds *et al.* (2013), and understand how we can adapt this system to ILCs' behaviour.

5.7.1 One-population model: altruistic or egoistic model

Consider a fixed-size population of N cells expressing the IL-7 receptor. The cytokine IL-7, constantly produced with rate ϕ, binds to the receptor with rate k_{on}. The receptor-ligand complex dissociates with rate k_{off}. We assume that free (unbound) receptors (respectively, bound receptors) are degraded with rate σ_r (respectively, σ_c). Finally, we assume that the signal elicited by receptor-ligand binding is proportional to the number of receptor-ligand complexes formed (constant ψ) and can be degraded with rate χ. With this notation in place, we define a system of ODEs describing the time evolution of the average dynamics of the concentration of extra-cellular IL-7, $I(t)$, the average number of free receptors per cell, $R(t)$, the average number of bound receptors (complexes) per cell, $C(t)$, and the amount (per cell) of signal produced by the binding of IL-7 to its receptor, $S(t)$. The system is as follows:

$$\frac{dI}{dt} = \phi + N(k_{\text{off}}C - k_{\text{on}}IR), \tag{5.67a}$$

$$\frac{dR}{dt} = -k_{\text{on}}IR + k_{\text{off}}C - \sigma_r R + f(k_s, S)\xi, \tag{5.67b}$$

$$\frac{dC}{dt} = k_{\text{on}}IR - k_{\text{off}}C - \sigma_c C, \tag{5.67c}$$

$$\frac{dS}{dt} = \psi C - \chi S. \tag{5.67d}$$

The function f (and the parameter k_s) tunes the receptor synthesis (rate ξ) by making it dependent or independent on the quantity of signal elicited by

IL-7/IL-7R binding. The different functions f studied in this section are compared in Figure 5.29 below.

IL-7R altruistic dynamics in T cells

In Park *et al.* (2019) and Reynolds *et al.* (2013), we have

$$f : (k_s, S) \mapsto \frac{1}{1 + \frac{S}{k_s}}. \tag{5.68}$$

Thus, system (5.67) describes the average altruistic IL-7R dynamics in T cells. This dynamics is called *altruistic* as T cells that received enough IL-7-mediated survival signal downregulate their membrane IL-7R expression level. As a result, cells that did not receive enough signal are prioritised for IL-7 signalling. Here, f is a decreasing function of S (see Figure 5.29) between 0 and 1. Parameter k_s is a characteristic parameter of S: when $S \ll k_s$, the receptor synthesis is independent of signalling; when $S \gg k_s$, we have perfect altruism as receptor synthesis is fully inhibited. In that case, as receptors are forming IL-7/IL-7R complexes or are degraded but no new receptors are created, we observe a decreased in unbound membrane receptor level. With this definition of f, the system admits a unique steady state, which has been computed in Park *et al.* (2019):

$$I^* = \frac{\phi\sigma_r(k_{\text{off}} + \sigma_c)(k_s N\sigma_c\chi + \phi\psi)}{k_{\text{on}}\sigma_c(k_s N\sigma_c\chi(N\xi - \phi) - \phi^2\psi)}, \tag{5.69a}$$

$$R^* = \frac{1}{N\sigma_r}\left(\frac{k_s N^2\sigma_c\xi\chi}{k_s N\sigma_c\chi + \psi\phi} - \phi\right), \tag{5.69b}$$

$$C^* = \frac{\phi}{N\sigma_c}, \tag{5.69c}$$

$$S^* = \frac{\phi\psi}{N\sigma_c\chi}. \tag{5.69d}$$

This steady state is positive ($I^* > 0$, $R^* > 0$, $C^* > 0$ and $S^* > 0$) under the condition:

$$\phi < \phi_{threshold} = \frac{\sqrt{k_s N^2\sigma_c\chi(k_s\sigma_c\chi + 4\psi\xi)} - k_s N\sigma_c\chi}{2\psi}. \tag{5.70}$$

This condition means that a large extra-cellular concentration of IL-7 will reduce the number of free IL-7R per cell to 0 (mathematically, to a negative receptor number), even with receptor synthesis on.

Stability analysis The Jacobian matrix of system (5.67) at steady state is:

$$\mathcal{J}(I^*, R^*, C^*, S^*) = \begin{pmatrix} -Nk_{\text{on}}R^* & -Nk_{\text{on}}I^* & Nk_{\text{off}} & 0 \\ -k_{\text{on}}R^* & -k_{\text{on}}I^* - \sigma_r & k_{\text{off}} & \frac{-k_s\xi}{(k_s+S^*)^2} \\ k_{\text{on}}R^* & k_{\text{on}}I^* & -k_{\text{off}} - \sigma_c & 0 \\ 0 & 0 & \psi & -\chi \end{pmatrix}.$$

A short computation with Mathematica (Wolfram Research, Inc., 2019) shows that the expressions of the eigenvalues of the Jacobian matrix are long and complicated, and their sign is not trivial. However, the study of the determinant and characteristic polynomial of this matrix demonstrates that the Jacobian matrix has no eigenvalues which are real positive, or equal to zero.

Proof. The determinant of the Jacobian matrix evaluated at the steady state,

$$\det(\mathcal{J}(I^*, R^*, C^*, S^*)) = \chi k_{\text{on}} R^* N \sigma_c \sigma_r,$$

is positive. It results that 0 is not an eigenvalue of the matrix.

Additionally, we can study the roots of the characteristic polynomial of the Jacobian matrix at steady state:

$$\chi_{\mathcal{J}} = Ax^4 + Bx^3 + Cx^2 + Dx + E, \tag{5.71}$$

where

$$\begin{aligned} A &= 1, \\ B &= \chi + k_{\text{off}} + k_{\text{on}}I^* + k_{\text{on}}R^*N + \sigma_c + \sigma_r, \\ C &= \chi k_{\text{off}} + \chi k_{\text{on}}I^* + \chi k_{\text{on}}R^*N + \chi(\sigma_c + \sigma_r) \\ &\quad + k_{\text{on}}I^*\sigma_c + k_{\text{on}}R^*N\sigma_c + k_{\text{off}}\sigma_r + k_{\text{on}}R^*N\sigma_r + \sigma_r\sigma_c, \\ D &= \chi k_{\text{on}}I^*\sigma_c + \chi k_{\text{on}}R^*N\sigma_c + \chi k_{\text{off}}\sigma_r + \chi k_{\text{on}}R^*N\sigma_r + \chi\sigma_r\sigma_c \end{aligned}$$

A	C	E	0
B	D	0	0
$\frac{BC-AD}{B}$	E	0	0
$\frac{BCD-AD^2+B^2E}{BC-AD}$	0	0	0
E	0	0	0

Table 5.2: Routh table of the characteristic polynomial (5.71).

$$+ k_{\text{on}} R^* N \sigma_c \sigma_r + \frac{k_s k_{\text{on}} I^* \psi \xi}{(k + S^*)^2},$$
$$E = \chi k_{\text{on}} R^* N \sigma_c \sigma_r.$$

All the coefficients of the characteristic polynomial are positive so according to Descartes' rule, the Jacobian matrix has no positive real eigenvalue. □

These results, however, do not exclude the possibility of complex eigenvalues with positive real part or on the imaginary axis. The Routh table (see Section 2.6.2) associated to the characteristic polynomial of $\mathcal{J}$ is described in Table 5.2. The roots of (5.71) have a negative real part if and only if

$$BC > D, \tag{5.72}$$

and

$$\frac{BCD - D^2}{B^2} > E. \tag{5.73}$$

Unfortunately, these two conditions are not informative in the general case. The stability analysis of the steady state remains inconclusive.

One could also suppose that extra-cellular IL-7 can be degraded. This hypothesis complicates the computation of the steady state (see Appendix K) and so won't be considered for the rest of this section.

IL-7R egoistic dynamics in ILCs

Contrary to T cells, ILCs do not downregulate their IL-7R expression when receiving IL-7R-mediated signal (Martin *et al.*, 2017). Hence, a cell that received enough survival signal will continue binding to IL-7, without any consideration

for the other cells that may struggle to survive. Here, I propose two functions f to model this *egoistic* dynamics.

Linear egoism I first assume that ILCs actually upregulate their IL-7R expression level when receiving IL-7R-mediated signal. This upregulation is linear, without any limit (carrying capacity), similar to the competition models for IL-2 described previously in this chapter. We define the function f as follows:

$$f : (k_s, S) \mapsto \frac{S}{k_s}. \tag{5.74}$$

Here, the more a cell signals through IL-7/IL-7R binding, the more it synthesises membrane IL-7R. A simulation of the system, as an agent-based model, will probably show a minority of cells hoovering all the intra-cellular cytokine, and depriving the other cells (the cells that do not express enough IL-7R to compete) of IL-7.

In that case, the system is linear in each variable and the non-trivial steady state is easy to obtain by substitution:

$$I^* = \frac{(k_{\text{off}} + \sigma_c)\chi, k_s\sigma_r}{k_{\text{on}}(\psi\xi - \chi\sigma_c k_s)}, \tag{5.75a}$$

$$R^* = \frac{\phi(\psi\xi - \chi\sigma_c k_s)}{\chi\sigma_c N k_s \sigma_r}, \tag{5.75b}$$

$$C^* = \frac{\phi}{N\sigma_c}, \tag{5.75c}$$

$$S^* = \frac{\psi\phi}{\chi N\sigma_c}. \tag{5.75d}$$

Every variable at steady state is positive as long as

$$\psi\xi - \chi\sigma_c k_s > 0. \tag{5.76}$$

This constraint has multiple interpretations, depending on the parameter we want to isolate. Note that all these parameters characterise intra-cellular mechanisms and thus, may not be tunable.

Following the process used for the stability analysis of the altruistic model[1], we can show that, under condition (5.76), the Jacobian matrix of the system with linear egoism has a strictly positive determinant and that the coefficients of its characteristic polynomial are all positive. As a consequence, the Jacobian matrix does not admit any positive real, nor zero, eigenvalues. As, once again, the Routh-Hurwitz criterion is not informative, we cannot exclude the possibility of complex eigenvalues with positive real part. The stability analysis of the steady state remains inconclusive.

Non-linear egoism I also propose a non-linear egoism function, inspired from the altruistic model (see Figure 5.29). Now suppose that the function f is:

$$f : (k_s, S) \mapsto \frac{1}{1 + \frac{k_s}{S}}. \tag{5.77}$$

This time, f is an increasing function of S, meaning that an increased signal will increase IL-7R synthesis (until a certain carrying capacity). Parameter k_s is, once again, a characteristic parameter of S: when $k_s \ll S$, membrane receptor synthesis is independent of signalling; when $k_s \gg S$, there is no synthesis. Note that the latter case does not model egoism but altruism.

This model admits the following non-trivial steady state:

$$I^* = \frac{(k_{\text{off}} + \sigma_c)\sigma_r(k_s\chi N\sigma_c + \phi\psi)}{k_{\text{on}}\sigma_c(-k_s N\sigma_c\chi - \phi\psi + \psi\xi N)} \tag{5.78a}$$

$$R^* = \frac{\phi}{\sigma_r N}\frac{-k_s\chi N\sigma_c - \phi\psi + \psi\xi N}{k_s\chi N\sigma_c + \phi\psi} \tag{5.78b}$$

$$C^* = \frac{\phi}{N_c\sigma_c} \tag{5.78c}$$

$$S^* = \frac{\psi\phi}{\chi N\sigma_c} \tag{5.78d}$$

The steady state solutions are positive as long as

$$-k_s\chi N\sigma_c - \phi\psi + \psi\xi N > 0, \tag{5.79}$$

[1]The Jacobian matrix and its characteristic polynomial differ very little from the altruistic case as only the expression of f changed.

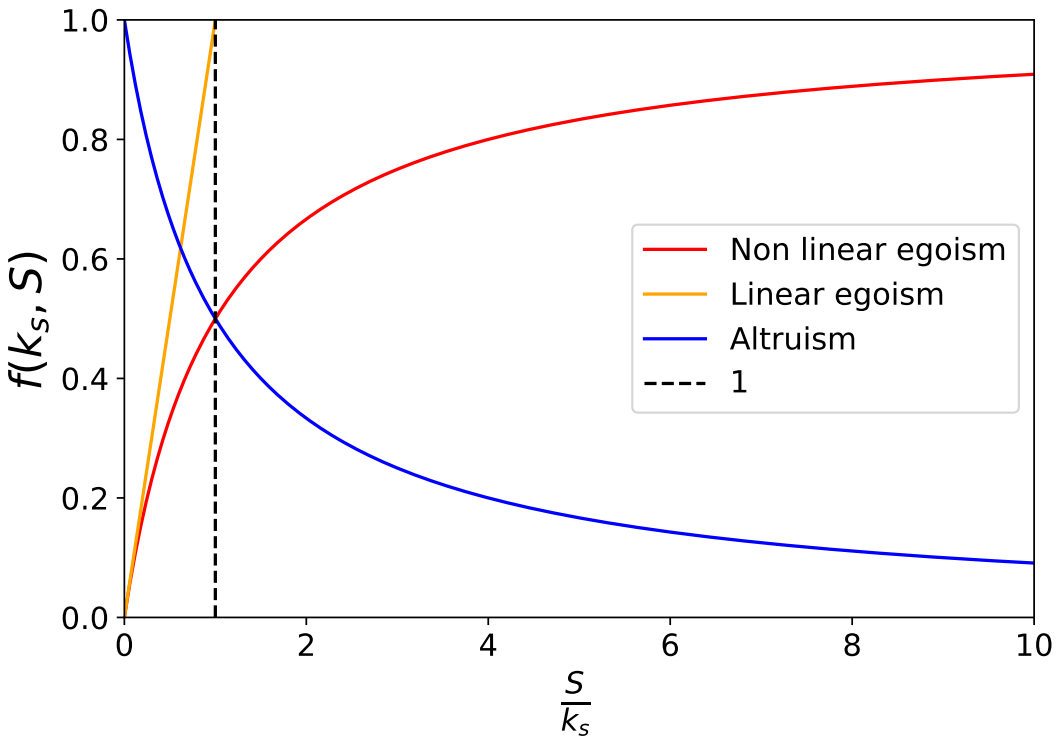


Figure 5.29: Comparison of functions, f, used to define the one-population models of the competition for IL-7. As soon as the cell receives some signal elicited by IL-7 binding its receptor, down/up regulation of IL-7R expression level starts. In the altruistic or non-linear egoistic cases, the change in IL-7R expression slows down as the amount of signal increases.

which, re-written as

$$\frac{-k_s \chi N \sigma_c + \psi \xi N}{\psi} > \phi,$$

can be interpreted as a constraint on IL-7 production.

Following the process used for the stability analysis of the altruistic model, we can show that the Jacobian matrix does not admit any positive real, nor zero, eigenvalues. We cannot, however, exclude the possibility of complex eigenvalues with positive real part. Once again, the stability analysis of the steady state remains inconclusive.

Adapting the system of ODEs into an ABM

To create an agent-based model, one cannot directly replace the two-attribute dynamics introduced in Section 5.2 by one of these ODEs systems. First, here, each system describes the time-evolution of the extra-cellular ligand concentration, while we assumed this quantity to be at steady state (production equal to consumption) in the previous models. Second, $R(t)$, $C(t)$ and $S(t)$ are average quantities (per cell) and do not take into account cells' specificities. Let us attempt to adapt system (5.67) into an agent-based model.

For any cell, $\bullet$, let us define three attributes: its current number of free receptors $r_\bullet(t)$, its current number of bound receptors, $c_\bullet(t)$, and its current quantity of signal, $s_\bullet(t)$. It is easy to link these attributes to the average variables:

$$R(t) = \frac{1}{N(t)} \sum_\bullet r_\bullet(t), \tag{5.80a}$$

$$C(t) = \frac{1}{N(t)} \sum_\bullet c_\bullet(t), \tag{5.80b}$$

$$S(t) = \frac{1}{N(t)} \sum_\bullet s_\bullet(t), \tag{5.80c}$$

where $N(t)$ denotes the current number of cells. Note that if $r_\bullet(0)$ (I suppose $c_\bullet(0) = 0$ and $s_\bullet(0) = 0$) is drawn from a log-normal distribution (similar to the models of the competition for IL-2), then $R(t)$, $C(t)$ and $S(t)$ are random variables. We also introduce the variable $I(t)$, which represents the current quantity of IL-7 present in the extra-cellular medium. Its dynamics is described by the following ODE:

$$\begin{aligned} \frac{dI}{dt} &= \phi + N\left(k_{\text{off}} C - k_{\text{on}} I R\right) \\ &= \phi + k_{\text{off}} \sum_\bullet c_\bullet - k_{\text{on}} I \sum_\bullet r_\bullet. \end{aligned} \tag{5.81}$$

For each cell $\bullet$, its attributes satisfy the following system of ODEs:

$$\frac{dr_\bullet}{dt} = -\sigma_r r_\bullet + f(k_s, s_\bullet)\xi + k_{\text{off}} c_\bullet - k_{\text{on}} I r_\bullet, \tag{5.82a}$$

$$\frac{dc_\bullet}{dt} = -\sigma_c c_\bullet - k_{\text{off}} c_\bullet + k_{\text{on}} I r_\bullet, \tag{5.82b}$$

$$\frac{ds_\bullet}{dt} = \psi c_\bullet - \chi s_\bullet. \tag{5.82c}$$

In addition to this deterministic dynamics, stochastic cellular events which are independent on cell's characteristics, such as death or immigration, are added. These events will affect $N(t)$, which, in turn, will impact each of these cell attributes (via $I(t)$). Note that in this system, the cell's response is tracked with the signal attribute, $s_\bullet(t)$. A starvation or division process as introduced in Section 5.6, may compare $s_\bullet(t)$ to a threshold. To model the competition between ILCs and T cells, we might want to combine the altruistic system with one of the egoistic models (linear egoism or non-linear egoism). Let us detail the competition in the following section.

5.7.2 Competition models

We combine the altruistic model (for T cells) with one of the egoistic systems (for ILCs) to model the competition for IL-7 between ILCs and T cells. The extra-cellular IL-7 is available for both cell types. We denote by the index 1 (respectively, 2) variables and parameters related to T cells (respectively, ILCs). We kept the same notation for the parameters as in the one-population systems, except that we renamed k_s to k_1 for T cells and k_2 for ILCs. The system is as follows:

$$\begin{aligned}
\frac{dI}{dt} &= \phi + N_1(k_{\text{off}}C_1 - k_{\text{on}}IR_1) + N_2(k_{\text{off}}C_2 - k_{\text{on}}IR_2), && \text{(5.83a)}\\
\frac{dR_1}{dt} &= -k_{\text{on}}IR_1 + k_{\text{off}}C_1 - \sigma_{r1}R_1 + \frac{1}{1+\frac{S_1}{k_1}}\xi_1, && \text{(5.83b)}\\
\frac{dC_1}{dt} &= k_{\text{on}}IR_1 - k_{\text{off}}C_1 - \sigma_{c1}C_1, && \text{(5.83c)}\\
\frac{dS_1}{dt} &= \psi_1 C_1 - \chi_1 S_1, && \text{(5.83d)}\\
\frac{dR_2}{dt} &= -k_{\text{on}}IR_2 + k_{\text{off}}C_2 - \sigma_{r2}R_2 + f(k_2, S_2)\xi_2, && \text{(5.83e)}\\
\frac{dC_2}{dt} &= k_{\text{on}}IR_2 - k_{\text{off}}C_2 - \sigma_{c2}C_2, && \text{(5.83f)}\\
\frac{dS_2}{dt} &= \psi_2 C_2 - \chi_2 S_2. && \text{(5.83g)}
\end{aligned}$$

We assumed that the receptor-ligand binding and dissociation constants are the same for both cells (rates k_{on} and k_{off}, respectively). Here $f(k_2, S_2)$ is the linear egoistic function ($f(k_2, S_2) = \frac{S_2}{k_2}$) or the non-linear egoistic function ($f(k_2, S_2) = \frac{1}{1+\frac{k_2}{S_2}}$). Note that in both cases (linear or non-linear egoism), there exists a steady state for which $R_2^* = 0$, $C_2^* = 0$, and $S_2^* = 0$, which is similar to the steady state computed for the altruistic model:

$$I^* = \frac{(k_{\text{off}} + \sigma_{c1})\phi\sigma_{r1}(N_1 k_1 \sigma_{c1}\chi_1 + \psi_1\phi)}{k_{\text{on}}\sigma_{c1}(k_1 N_1^2 \sigma_{c1}\chi_1\xi_1 - \phi k_1 N_1 \sigma_{c1}\chi_1 - \psi_1\phi^2)}, \tag{5.84a}$$

$$R_1^* = \frac{k_1 N_1^2 \sigma_{c1}\chi_1\xi_1 - \phi k_1 N_1 \sigma_{c1}\chi_1 - \psi_1\phi^2}{\sigma_{r1} N_1 (N_1 k_1 \sigma_{c1}\chi_1 + \psi_1\phi)}, \tag{5.84b}$$

$$C_1^* = \frac{\phi}{N_1\sigma_{c1}}, \tag{5.84c}$$

$$S_1^* = \frac{\psi_1\phi}{N_1\sigma_{c1}\chi_1}. \tag{5.84d}$$

This steady state is biologically meaningful under the condition that:

$$\phi < \phi_{threshold} = \frac{\sqrt{k_1 N_1^2 \sigma_{c1}\chi_1(k_1\sigma_{c1}\chi_1 + 4\psi_1\xi_1)} - k_1 N_1 \sigma_{c1}\chi_1}{2\psi_1}. \tag{5.85}$$

ILCs and T cells competing for IL-7: linear egoism

Let f be the function:

$$f : (k_2, S_2) \mapsto \frac{S_2}{k_2},$$

and assume that none of the variables are equal to 0. We can derive the steady state of all the variables as a function of C_1^*. Equations (5.83d) and (5.83g) (with the derivative set to 0) give

$$S_1^* = \frac{\psi_1}{\chi_1} C_1^*, \tag{5.86}$$

$$S_2^* = \frac{\psi_2}{\chi_2} C_2^*, \tag{5.87}$$

and equations (5.83c) and (5.83f) (with derivative set to 0) give

$$k_{\text{on}} I^* R_1^* = (k_{\text{off}} + \sigma_{c1}) C_1^*, \tag{5.88}$$

$$k_{\text{on}} I^* R_2^* = (k_{\text{off}} + \sigma_{c2}) C_2^*. \tag{5.89}$$

Substituting these expressions in (5.83b) and (5.83e) yields:

$$R_1^* = \frac{\xi_1 k_1 \chi_1 - \sigma_{c1} k_1 \chi_1 C_1^* - \sigma_{c1} \psi_1 (C_1^*)^2}{\sigma_{r1}(k_1 \chi_1 + \psi_1 C_1^*)}, \tag{5.90}$$

$$R_2^* = \frac{(\psi_2 \xi_2 - \sigma_{c2} \chi_2 k_2) C_2^*}{\chi_2 k_2 \sigma_{r2}}. \tag{5.91}$$

The expression of R_2^* substituted in (5.89) allows the computation of an expression for the steady state concentration of extra-cellular cytokine:

$$I^* = \frac{(k_{\text{off}} + \sigma_{c2}) \chi_2 k_2 \sigma_{r2}}{k_{\text{on}} (\psi_2 \xi_2 - \sigma_{c2} \chi_2 k_2)}. \tag{5.92}$$

Finally, we find the expression of C_2^* as a function of C_1^* by substituting $k_{\text{off}} C_1^* - k_{\text{on}} I^* R_1^* = \sigma_{c1} C_1^*$ and $k_{\text{off}} C_2^* - k_{\text{on}} I^* R_2^* = \sigma_{c2} C_2^*$ (from equations (5.88) and (5.89)) in equation (5.83a) (derivative at 0). We obtain:

$$C_2^* = \frac{\phi - N_1 \sigma_{c1} C_1^*}{N_2 \sigma_{c2}}. \tag{5.93}$$

The variable C_1^* is the root of the following polynomial:

$$\mathcal{A} x^2 + \mathcal{B} x + \mathcal{C} = 0, \tag{5.94}$$

where

$$\begin{aligned} \mathcal{A} =& \psi_1 [\sigma_{r1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1} k_{\text{on}} I^*], \\ \mathcal{B} =& k_1 \chi_1 [\sigma_{r_1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1} k_{\text{on}} I^*], \\ \mathcal{C} =& - k_{\text{on}} I^* k_1 \chi_1 \xi_1, \end{aligned}$$

which has been obtained by substituting the expression of R_1^* in equation (5.88). Coefficients $\mathcal{A}$ and $\mathcal{B}$ are positive, $\mathcal{C}$ is negative. Hence, according to Descartes' rule, this polynomial admits a unique positive real solution (the only one that is

biologically meaningful):

$$C_1^* = \frac{-k_1\chi_1[\sigma_{r_1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1}k_{\text{on}}I^*] + \sqrt{\Delta_{C_1}}}{2\psi_1[\sigma_{r_1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1}k_{\text{on}}I^*]}, \tag{5.95}$$

where we wrote

$$\begin{aligned}\Delta_{C_1} = &k_1^2\chi_1^2[\sigma_{r_1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1}k_{\text{on}}I^*]^2 \\ &+ 4\xi_1 k_1\chi_1 k_{\text{on}}I^*\psi_1[\sigma_{r_1}(k_{\text{off}} + \sigma_{c1}) + \sigma_{c1}k_{\text{on}}I^*].\end{aligned}$$

Note that this steady state is biologically relevant under the following conditions:

$$C_1^* < \frac{\phi}{N_1\sigma_{c1}} \qquad \text{because we want } C_2^* > 0, \tag{5.96}$$

$$\psi_2\xi_2 - \sigma_{c2}\chi_2 k_2 > 0 \qquad \text{because we want } I^* > 0. \tag{5.97}$$

Note that, contrary to the one-population model, the IL-7 production must be greater than a certain threshold to ensure a biologically meaningful steady state.

ILCs and T cells competing for IL-7: non-linear egoism

Let us now consider the competition model with non-linear egoism. Let f be the function

$$f : (k_2, S_2) \mapsto \frac{1}{1 + \frac{k_2}{S_2}}.$$

To find positive steady states, we solve

$$0 = \phi + N_1(k_{off}C_1 - k_{on}IR_1) + N_2(k_{off}C_2 - k_{on}IR_2), \tag{5.98a}$$

$$0 = -k_{on}IR_1 + k_{off}C_1 - \sigma_{r1}R_1 + \frac{k_1}{k_1 + S_1}\xi_1, \tag{5.98b}$$

$$0 = k_{on}IR_1 - k_{off}C_1 - \sigma_{c1}C_1, \tag{5.98c}$$

$$0 = \psi_1 C_1 - \chi_1 S_1, \tag{5.98d}$$

$$0 = -k_{on}IR_2 + k_{off}C_2 - \sigma_{r2}R_2 + \frac{S_2}{k_2 + S_2}\xi_2, \tag{5.98e}$$

$$0 = k_{on}IR_2 - k_{off}C_2 - \sigma_{c2}C_2, \tag{5.98f}$$

$$0 = \psi_2 C_2 - \chi_2 S_2, \tag{5.98g}$$

and suppose that no variable is equal to 0.

Equations (5.98d) and (5.98g) give

$$S_1^* = \frac{\psi_1}{\chi_1} C_1^*, \tag{5.99}$$

$$S_2^* = \frac{\psi_2}{\chi_2} C_2^*. \tag{5.100}$$

Equations (5.98c) and (5.98f) give

$$k_{on} I^* R_1^* = (k_{off} + \sigma_{c1}) C_1^*, \tag{5.101}$$

$$k_{on} I^* R_2^* = (k_{off} + \sigma_{c2}) C_2^*. \tag{5.102}$$

Substituting these results into (5.98a) we obtain:

$$C_2^* = \frac{\phi - N_1 \sigma_{c1} C_1^*}{N_2 \sigma_{c2}}. \tag{5.103}$$

By substituting (5.99) and (5.101) in (5.98b), we obtain:

$$R_1^* = \frac{-\sigma_{c1} \psi_1 (C_1^*)^2 - \sigma_{c1} \chi_1 k_1 C_1 + k_1 \xi_1 \chi_1}{\sigma_{r1} (\chi_1 k_1 + \psi_1 C_1^*)}. \tag{5.104}$$

Similarly, making use of equations (5.100) and (5.102) in (5.98e), we obtain:

$$R_2^* = \frac{-\sigma_{c2} \psi_2 (C_2^*)^2 + C_2^* (\psi_2 \xi_2 - \sigma_{c2} \chi_2 k_2)}{\sigma_{r2} (\chi_2 k_2 + \psi_2 C_2^*)}. \tag{5.105}$$

We also have

$$I^* = \frac{(k_{\text{off}} + \sigma_{c1}) C_1^*}{k_{\text{on}} R_1^*}. \tag{5.106}$$

Combining (5.101) and (5.102) together, we have:

$$(k_{off} + \sigma_{c2}) C_2^* R_1^* = (k_{off} + \sigma_{c1}) C_1^* R_2^*. \tag{5.107}$$

Replacing R_1^*, R_2^* and C_2^* by their expression as a function of C_1^* in the previous equation yields the following polynomial of degree 3 in C_1^*:

$$\mathcal{A}(C_1^*)^3 + \mathcal{B}(C_1^*)^2 + \mathcal{C}C_1^* + \mathcal{D} = 0, \tag{5.108}$$

where

$$\begin{aligned}
\mathcal{A} =& N_1\psi_1\psi_2\sigma_{c1}\mathcal{K},\\
\mathcal{B} =& -[\phi\psi_1\psi_2 - \chi_1 k_1 N_1\psi_2\sigma_{c1} + \chi_2 k_2 N_2\psi_1\sigma_{c2}]\mathcal{K} - N_2\psi_1\psi_2\sigma_{c2}\sigma_{r1}\xi_2(k_{off}+\sigma_{c1}),\\
\mathcal{C} =& \chi_1 k_1(\sigma_{c2}\sigma_{r1}(k_{off}+\sigma_{c1})(\phi\psi_2 + \chi_2 k_2 N_2\sigma_{c2} - N_2\psi_2\xi_2)\\
& - \sigma_{c1}\sigma_{r2}(k_{off}+\sigma_{c2})(\phi\psi_2 + \chi_2 k_2 N_2\sigma_{c2} + N_1\psi_2\xi_1)),\\
\mathcal{D} =& \chi_1 k_1\sigma_{r2}\xi_1(k_{off}+\sigma_{c2})(\phi\psi_2 + \chi_2 k_2 N_2\sigma_{c2}),
\end{aligned}$$

where we wrote $\mathcal{K} = (\sigma_{c1}\sigma_{r2}(k_{off}+\sigma_{c2}) - \sigma_{c2}\sigma_{r1}(k_{off}+\sigma_{c1}))$. The signs of the coefficients are not obvious and may depend on the parameters values. All we know is that $\mathcal{D} > 0$. Thus, we might have 0 to 3 positive steady states (with no variables equal to 0). Note that to obtain a biologically meaningful steady state, we are looking for a root C_1^* that ensures that $R_1^* > 0$, $R_2^* > 0$ and $C_2^* > 0$. To have $C_2^* > 0$ we want

$$C_1^* < \frac{\phi}{N_1\sigma_{c1}}.$$

The condition $R_2^* > 0$ means that

$$C_1^* > \frac{N_2\sigma_{c2}\chi_2 k_2 - \psi_2 N_2\xi_2 + \psi_2\phi}{N_1\sigma_{c1}\psi_2}.$$

Finally, to ensure $R_1 > 0$ we want

$$\frac{-\sigma_{c1}k_1\chi_1 + \sqrt{(\sigma_{c1}k_1\chi_1)^2 + 4k_1\xi_1\psi_1\chi_1\sigma_{c1}}}{2\sigma_{c1}\psi_1} > C_1^*,$$

and

$$C_1^* > 0 > \frac{-\sigma_{c1}k_1\chi_1 - \sqrt{(\sigma_{c1}k_1\chi_1)^2 + 4k_1\xi_1\psi_1\chi_1\sigma_{c1}}}{2\sigma_{c1}\psi_1}.$$

Thus, the competition model with non-linear egoism has between 1 and 4 steady states: the steady state with "no ILCs" and three possible steady state

with both types of cell that correspond to the solutions of the polynomial (5.108).

Now suppose that $\sigma_{r1} = \sigma_{r2} = \sigma_r$ and $\sigma_{c1} = \sigma_{c2} = \sigma_c$. Equation (5.107) can be simplified:

$$R_1^* C_2^* = C_1^* R_2^*, \tag{5.109}$$

as well as the coefficients of polynomial (5.108). In that case, C_1^* is the positive real root of a polynomial of degree 2:

$$AX^2 + BX + C = 0, \tag{5.110}$$

where

$$\begin{aligned} A &= -N_2\psi_1\psi_2\sigma_c\xi_2, \\ B &= -\chi_1 k_1 \psi_2 \sigma_c (N_1\xi_1 + N_2\xi_2), \\ C &= \chi_1 k_1 (\phi\psi_2 + \chi_2 k_2 N_2 \sigma_c)\xi_1. \end{aligned}$$

If the unique positive root of this polynomial satisfies the conditions that ensure $R_1^* > 0$, $R_2^* > 0$ and $C_2^* > 0$, then we found a positive steady state for the system, with no variables equal to 0.

Adaptation as an agent-based model

The work conducted for the one-population model can be adapted for the competition model. The extra-cellular cytokine concentration, $I(t)$, satisfies the following ODE:

$$\frac{dI}{dt} = \phi + k_{\text{off}} \sum_{\bullet} c_\bullet - k_{\text{on}} I. \sum_{\bullet} r_\bullet \tag{5.111}$$

For each cell $\bullet$, we define the same three attributes as for the one-population models, $r_\bullet(t)$, $c_\bullet(t)$ and $s_\bullet(t)$, which satisfy the following system of ODEs:

$$\frac{dr_\bullet}{dt} = -\sigma_{rn} r_\bullet + f_n(k_n, s_\bullet)\xi_n + k_{\text{off}} c_\bullet - k_{\text{on}} I r_\bullet, \tag{5.112a}$$

$$\frac{dc_\bullet}{dt} = -\sigma_{cn} c_\bullet - k_{\text{off}} c_\bullet + k_{\text{on}} I r_\bullet, \tag{5.112b}$$

$$\frac{ds_\bullet}{dt} = \psi_n c_\bullet - \chi_n s_\bullet, \tag{5.112c}$$

where $n = 1$ if the cell $\bullet$ is a T cell and $n = 2$ if it is an ILC. I defined

$$f_1(k_1, s_\bullet) = \frac{1}{1 + \frac{s_\bullet}{k_1}},$$

and

$$f_2(k_2, s_\bullet) = \frac{1}{1 + \frac{k_2}{s_\bullet}}$$

or

$$f_2(k_2, s_\bullet) = \frac{s_\bullet}{k_2},$$

depending on the egoism type considered. Cellular stochastic events, such as death or division, affect $N_1(t)$ and $N_2(t)$ which then impact the other variables.

5.8 Discussion and conclusion

In this chapter, we aimed to characterise the IL-2R distribution among a cell population and to understand its relationship with cellular stochastic events. Since complicated agent-based models do not always provide mechanistic insights, we focused on the mathematical description of simpler models first. In particular, we started with a simple and fully deterministic model of the IL-2/IL-2R dynamics within a population with a constant number of cells, and added the stochastic cellular events one at a time.

To make the mathematical analysis of these models tractable, we made strong simplifications of the actual biology. First, we assumed that the per cell IL-2 production is constant. In reality, the IL-2 secretion may decrease when a IL-2 receptor binds this cytokine (Feinerman *et al.*, 2010). We also ignored any receptor trafficking (such as receptor internalisation, degradation and recycling) and signal transduction. Instead, we hid all of these mechanisms under the term "cytokine consumption". Heuristically, we assumed that receptors are never degraded (while the cell is alive) and, once they bind to a molecule of IL-2, they are immediately internalised and recycled back to the cell surface. We also did not implement any limit on the number of receptors a cell can express, which might lead to unrealistic

high receptor expression levels. Luckily, it seems that the rules of our model created a "natural" receptor carrying capacity as no cells expressed more than 10^{11} receptors (except in the fully deterministic case). We assumed that the IL-2 absorbed by a cell is accumulated and never degraded until the cell dies or divides. Introducing such degradation may reduce the "natural" receptor carrying capacity of any cell. In the case of cellular events depending on local variables, such as the starvation or T regs division described in Section 5.6, such intra-cellular cytokine degradation may allow these cellular events to be switched on and off as a cell accumulates and loses intra-cellular cytokine. We also assumed that cells, that do not express IL-2Rα (*i.e.*, naive cells), do not consume IL-2, and thus were not explicitly taken into account in the model. One could also note that we assumed the cell division to be instantaneous. In reality, before division, a cell enters a cell cycle which lasts for a couple of hours (Belluccini *et al.*, 2022; De Boer & Perelson, 2005; Hogan *et al.*, 2013). However, we have little information on the behaviour of a cell undergoing such process regarding its IL-2/IL-2R dynamics, and thus chose not to include it in the model. Finally, we did not consider any spatial feature, even though it has been shown that regulatory T cells cluster around IL-2 producing cells, which regulates IL-2 diffusion in the organ and contains the inflammation (Amado *et al.*, 2013; Liu *et al.*, 2015; Oyler-Yaniv *et al.*, 2017; Wong *et al.*, 2021).

We completed our mathematical analysis, which is, by essence, valid for any parameter value, with qualitative observations from numerical simulations. Consequently, we had to assign a value to each parameter of the model. In a mouse, a T cell lives for about a month (den Braber *et al.*, 2012). A human T cell, however, may live for more than a year (den Braber *et al.*, 2012). However, activated T cells must have a much shorter lifetime, about a few days or weeks, as they are supposed to exist only as long as needed to clear an infection. In this chapter, we assumed that the stochastic events considered in our models were happening on a timescale of days. Ref. Burroughs *et al.* (2006) proposes a mathematical model of IL-2 consumption by T cells and T regs in which they give the death rate the following value $\mu = 0.1$ /day to 0.01/day (following measurements from Ref. Michie *et al.* (1992)). We kept this value for our models and gave a sensible value (on the same timescale) to the division rate, λ. The value of parameter α

should depend on the total influx of cells considered. Indeed, whether we consider a blood sample, a lymph node or a whole mouse, we do not consider the same fraction of cells of an individual and thus the total influx must be adapted to this fraction (the thymus releases self-reactive cells at a rate that is measured for the whole body). In the examples of this chapter, we chose $\alpha = 30$ cells/day, which allows us to have about 10^3 to 10^4 cells at steady state. Considering more than 10^4 cells at steady-state significantly increases the simulation time of the models. In Ref. Burroughs *et al.* (2006), the input rate (of all T cells) varies between 0 and 10^4 cells/mL/day. Note that the self-reactive T cells represent only a very small fraction of all T cells at homeostasis (there are about 10^6 T cells in a hypothetical spherical skin-draining lymph node of diameter 2mm (Catron *et al.*, 2004; Kim *et al.*, 2007b)). Finally, if an activated T cell lives for a couple of days, then the IL-2/IL-2R dynamics should happen within hours. We gave to p and u the following values: $p = 10$ $[i_T]$/day/cell and $u = 10$ $[r_T]/[i_T]$/day. Changing all the parameters by the same factor should not change the dynamics.

In absence of stochastic events or in presence of cell death only, we managed to derive an expression for the receptor distribution at any time of the simulation. When adding other stochastic events (activation then division), the dynamics became more complicated and the mathematical description could not capture the entire behaviour of the system. For instance, while the cells follow the same rules, phenomena of inequalities arise (such as the emergence of groups of cells that express very different levels of receptors or absorbed different levels of cytokine), which were not prescribed by the mathematics. Nonetheless, with a mixture of mathematical analysis and numerical simulations, we managed to significantly improve our understanding of the dynamics of the model. In particular, we managed to explain the sharp noisy variations displayed by the population variables and clarify why the receptor distribution was always changing despite the system being at a stochastic steady state.
When adding the activation event, for some parameter regimes, we observed the emergence of two cell populations. The cells that were the first to enter the pool (old cohort) express high receptor levels and deprive the newer cells (new cohort) of cytokine, which prevent them from upregulating their receptor expression level.

That is, the cells that express a low receptor expression level keep expressing very few IL-2R, while the other cells keep expressing a high number of receptors. When the old cohort goes extinct, the cells of the new cohort suddenly and massively upregulate their receptor expression level. In a "rags to riches" manner, the new cohort becomes the old cohort and starts to deprive the newer cells of cytokine. In another parameter regime, we observe a more egalitarian dynamics in which all cells express about the same level of IL-2R and behave similarly. The addition of the division event complicates the dynamics. Indeed, such receptor "inheritance" mechanism blurs the distinction between cells with high and low receptor expression levels, as we may observe cells with both low and high receptor levels in a family of cells of the old cohort, as well as in a family of the new cohort. This illustrates that criteria of interest used to analyse one model (such as the discrimination between old and new cohorts in the model with death and activation only) may not be relevant anymore when adding or removing stochastic events.

Despite our improved understanding of the relationship between the IL-2R distribution and the population dynamics, we could not link a particular shape of the receptor distribution, at a certain time t of the simulation, to a particular event or state of the system. Indeed, we showed that a bimodal receptor distribution might have three different explanations in the model with division. The receptor distribution of cells sampled at any time of the simulation (such as Figure 5.20) may be more stable but more investigation is needed [1].

The original ambition of this project extended beyond the work conducted in the first sections of this chapter: the "end goal" was to model the competition for IL-2 between regulatory and self-reactive conventional T cells, taking into account cellular events depending on individual cell characteristics. We provided a mathematical framework and a Python code for this final model, but no detailed mathematical analysis. As the competition for a protein between two cell populations is a common mechanism in the immune system, we could extend our agent-based model to other systems. For instance, the competition for IL-7

[1] Note that such sampling might not be feasible in a biological experiment. Thus any comparison between experimental data and mathematical predictions may not be possible.

between T cells and innate lymphoid cells is interesting because the two cell populations do not have the same receptor dynamics: IL-7-induced signal downregulates IL-7R expression in T cells but not innate lymphoid cells. A deterministic model for the competition for IL-7 between T cells was introduced in Refs. Park *et al.* (2019); Reynolds *et al.* (2013). Similar models for the competition for IL-7 within the innate lymphoid cell populations and between T cells and innate lymphoid cells have been developed in Section 5.7. These deterministic systems were, then, adapted into agent-based model frameworks. However, the structure of these agent-based models is quite different from the agent-based models developed for the IL-2 competition, as the cytokine extra-cellular concentration was taken into account.

Models of competition for a cytokine might be helpful for therapeutic insights. Indeed, understanding the mechanisms by which one population can overtake the other is necessary to develop potential treatments for immune-related diseases. For instance, a low dose of IL-2 therapy seems to prioritise regulatory T cell proliferation (over self-reactive proliferation), which ameliorates autoimmune diseases caused by lack of T regs (Abbas *et al.*, 2018; Trotta *et al.*, 2018). A well-suited mathematical model could help determine the dose by identifying the settings in which T regs are overtaken by self-reactive T cells. Taking into account the diversity inside each cell population allows a greater modelling plasticity and may be the only way to model the subtle biological mechanisms at work in the immune system.

Chapter 6

Concluding remarks

The work conducted in this thesis examined the complex relationship between individual cell signalling, cell population dynamics and cell-to-cell variability in receptor expression levels in a population. The non-intuitive aspect of this relationship was first observed experimentally: increasing the abundance of an essential component of the IL-7R decreases the cell's response to IL-7, reflected in an increasing EC_{50} and a non-monotonic amplitude (increase then decrease). The mathematical models of the IL-7R/IL-7 receptor-ligand system, developed in Chapter 3, explained this seemingly paradoxical observation by the formation of non-signalling complexes. Making use of algebraic tools, such as Gröbner bases, analytic expressions of the amplitude and EC_{50} of these models were computed. Such expressions reduced computational costs (computational time and number of computation) and facilitated the fitting of the models to the data set, by showing how these quantities depend on the parameters of the system. The algebraic method was, then, applied to different receptor-ligand systems, in order to explore the impact of the receptor architecture on the variations of the amplitude and EC_{50} in response to the increased abundance of an essential receptor component. The amplitude expression of a general family of receptor-ligand systems has also been computed, this time without making use of advanced algebraic tools. This work, described in the first two research chapters of this thesis, demonstrated the importance of balancing the abundances of the receptor core chains for optimal signalling. In Chapter 5, I developed and proposed a mathematical description of several agent-based models of the competition for IL-2 within the T cell population.

In particular, I examined the impact of cellular event such as immigration, death or cell division, on the IL-2R expression level distribution. The results were not always predictable, even for simple models. In particular, phenomena of inequality arose in the cell population, some cells depriving the others of cytokine. We also observed that the receptor distribution, for some parameter regimes, was constantly changing, despite the cell population being at steady state. These changes were explained making use of numerical simulations.
The work conducted in these three chapters shows that competition is prevalent in biological systems and must be taken into account in mathematical models. At the cellular level, the competition for a receptor chain between non-signalling receptors (incomplete signalling receptors or other cytokine receptors) has a great influence on the cell's response, both in terms of amplitude and EC_{50}. At the cell population level, the competition for the ligand is responsible for the modulation of important mechanisms in the immune system, as well as for the maintenance of key cell populations. It can also be responsible for splitting the cell population into many sub-populations, each of them expressing different average receptor numbers and playing different roles in the immune response or homeostasis.

One of the main challenges in mathematical modelling is to be realistic enough so that the main features of the biological systems are reproduced, but simple enough to have tractable and insightful mathematical descriptions. Despite the apparent simplicity of the models proposed in this thesis, they demonstrate that the quantitative effects of receptor chain upregulation can be vastly different, depending on the elements of a receptor's signalling core. This work provides a theoretical and quantitative framework with which to interpret the potential functional significance of receptor up/downregulation during lymphocyte differentiation (Kalia *et al.*, 2010; Voisinne *et al.*, 2015), oncogenesis (Du & Lovly, 2018) or drug treatment (Vogel *et al.*, 2016). It also sheds light on the importance of investigating the cellular level to model precise population systems. Further work could combine the cellular and populations models of this thesis. Indeed, eventual non-monotonic relationships between the cell's receptor copy number and its ligand-mediated response are not considered in the current models of Chapter 5. Fortunately, as demonstrated by the experimental work described in

Section 3.5, the IL-2-mediated amplitude and EC_{50} are monotonic functions of the number of IL-2Rα chains. Namely, increasing the abundance of IL-2Rα chains increases the cell's response, as assumed in the agent-based models of this thesis. It has been measured that T regs have lower IL-2 EC_{50} than conventional T cells (Trotta *et al.*, 2018). That is, for the same dose of IL-2, the T regs response will be greater than the conventional T cell response, on average. Combined with the experimental work described in Section 3.5, this tells us that the average IL-2Rα number in T regs is greater than in activated conventional T cells. This result, useful to calibrate the model of the competition for IL-2 between these two cell populations, is consistent with other experimental studies that directly measured the receptor numbers of conventional and regulatory T cells (Abbas *et al.*, 2018). In the case of non-monotonic variations of the cell's response when increasing the availability of an essential receptor chain component, a cellular model may be a useful first step to guide the development of more precise population models.

References

Abbas, A.K., Trotta, E., R. Simeonov, D., Marson, A. & Bluestone, J.A. (2018). Revisiting il-2: Biology and therapeutic prospects. *Science immunology*, **3**, eaat1482. 220, 281, 302, 305

Akdis, M., Burgler, S., Crameri, R., Eiwegger, T., Fujita, H., Gomez, E., Klunker, S., Meyer, N., O'Mahony, L., Palomares, O. *et al.* (2011). Interleukins, from 1 to 37, and interferon-γ: receptors, functions, and roles in diseases. *Journal of allergy and clinical immunology*, **127**, 701–721. 2, 3, 6, 65

Akritas, A.G. (1982). Reflections on a pair of theorems by budan and fourier. *Mathematics Magazine*, **55**, 292–298. 42

Alexander, B., Browse, D., Reading, S. & Benjamin, I. (1999). A simple and accurate mathematical method for calculation of the ec50. *Journal of pharmacological and toxicological methods*, **41**, 55–58. 119

Allen, L.J. (2007). *Introduction to mathematical biology*. Pearson/Prentice Hall. 11, 15

Allen, L.J. (2010). *An introduction to stochastic processes with applications to biology*. CRC press. 44, 53, 54

Altan-Bonnet, G. & Mukherjee, R. (2019). Cytokine-mediated communication: a quantitative appraisal of immune complexity. *Nature Reviews Immunology*, 1. 7

Amado, I.F., Berges, J., Luther, R.J., Mailhé, M.P., Garcia, S., Bandeira, A., Weaver, C., Liston, A. & Freitas, A.A. (2013). Il-2 coordinates

il-2–producing and regulatory t cell interplay. *Journal of Experimental Medicine*, **210**, 2707–2720. 218, 220, 299

An, G., Mi, Q., Dutta-Moscato, J. & Vodovotz, Y. (2009). Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **1**, 159–171. 57

Anagnost, J.J. & Desoer, C.A. (1991). An elementary proof of the routh-hurwitz stability criterion. *Circuits, systems and signal processing*, **10**, 101–114. 43

Anders, A., Ghosh, B., Glatter, T. & Sourjik, V. (2020). Design of a mapk signalling cascade balances energetic cost versus accuracy of information transmission. *Nature communications*, **11**, 3494. 8

Atkinson, M.A., Eisenbarth, G.S. & Michels, A.W. (2014). Type 1 diabetes. *The Lancet*, **383**, 69–82. 218

Bache, K.G., Slagsvold, T. & Stenmark, H. (2004). Defective downregulation of receptor tyrosine kinases in cancer. *The EMBO journal*, **23**, 2707–2712. 9, 66

Badrick, T., Ward, G. & Hickman, P. (2023). The effect of the immunoassay curve fitting routine on bias in troponin. *Clinical Chemistry and Laboratory Medicine (CCLM)*, **61**, 188–195. 119

Bauer, A.L., Beauchemin, C.A. & Perelson, A.S. (2009). Agent-based modeling of host–pathogen systems: The successes and challenges. *Information sciences*, **179**, 1379–1389. 57

Belluccini, G., López-García, M., Lythe, G. & Molina-París, C. (2022). Counting generations in birth and death processes with competing erlang and exponential waiting times. *Scientific Reports*, **12**, 1–20. 299

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F. *et al.* (2013). An estimation of the number of cells in the human body. *Annals of human biology*, **40**, 463–471. 4

Blitzstein, J.K. & Hwang, J. (2015). *Introduction to probability*. Crc Press Boca Raton, FL. 55

Bocher, M. (1901). Certain cases in which the vanishing of the wronskian is a sufficient condition for linear dependence. *Transactions of the American Mathematical Society*, **2**, 139–149. 14, 226

Borgonovo, E., Pangallo, M., Rivkin, J., Rizzo, L. & Siggelkow, N. (2022). Sensitivity analysis of agent-based models: a new protocol. *Computational and Mathematical Organization Theory*, **28**, 52–94. 57, 58

Boulanger, M.J., Chow, D.c., Brevnova, E.E. & Garcia, K.C. (2003). Hexameric structure and assembly of the interleukin-6/il-6 α-receptor/gp130 complex. *Science*, **300**, 2101–2104. 181, 215

Bouneaud, C., Kourilsky, P. & Bousso, P. (2000). Impact of negative selection on the t cell repertoire reactive to a self-peptide: a large fraction of t cell clones escapes clonal deletion. *Immunity*, **13**, 829–840. 218

Bowcock, A.M. & Krueger, J.G. (2005). Getting under the skin: the immunogenetics of psoriasis. *Nature Reviews Immunology*, **5**, 699–711. 218

Boyd, J.P. (2002). Computing zeros on a real interval through chebyshev expansion and polynomial rootfinding. *SIAM Journal on Numerical Analysis*, **40**, 1666–1682. 111

Buchberger, B. (1965). Ein algorithmus zum auffinden der basiselemente des restklassenringes nach einem nulldimensionalen polynomideal. *PhD thesis, Universitat Insbruck*. 17, 26, 31

Buchberger, B. (2006). Bruno buchberger's phd thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of symbolic computation*, **41**, 475–511. 17, 26, 31

Burroughs, N., de Oliveira, B.M.P.M. & Pinto, A.A. (2006). Regulatory t cell adjustment of quorum growth thresholds and the control of local immune responses. *Journal of theoretical biology*, **241**, 134–141. 299, 300

Busse, D., de la Rosa, M., Hobiger, K., Thurley, K., Flossdorf, M., Scheffold, A. & Höfer, T. (2010). Competing feedback loops shape il-2 signaling between helper and regulatory t lymphocytes in cellular microenvironments. *Proceedings of the National Academy of Sciences*, **107**, 3058–3063. 220

Catron, D.M., Itano, A.A., Pape, K.A., Mueller, D.L. & Jenkins, M.K. (2004). Visualizing the first 50 hr of the primary immune response to a soluble antigen. *Immunity*, **21**, 341–347. 300

Chen, L.Y., Biggs, C.M., Jamal, S., Stukas, S., Wellington, C.L. & Sekhon, M.S. (2021). Soluble interleukin-6 receptor in the covid-19 cytokine storm syndrome. *Cell Reports Medicine*, **2**, 100269. 214

Chen, Z., Bertin, R. & Froldi, G. (2013). Ec50 estimation of antioxidant activity in dpph assay using several statistical programs. *Food chemistry*, **138**, 414–420. 7, 120

Cherrier, M., Ramachandran, G. & Golub, R. (2020). The interplay between innate lymphoid cells and t cells. *Mucosal Immunology*, **13**, 732–742. 282

Collison, L.W., Delgoffe, G.M., Guy, C.S., Vignali, K.M., Chaturvedi, V., Fairweather, D., Satoskar, A.R., Garcia, K.C., Hunter, C.A., Drake, C.G. *et al.* (2012). The composition and signaling of the il-35 receptor are unconventional. *Nature immunology*, **13**, 290–299. 214

Cotari, J.W., Voisinne, G. & Altan-Bonnet, G. (2013a). Diversity training for signal transduction: leveraging cell-to-cell variability to dissect cellular signaling, differentiation and death. *Current opinion in biotechnology*, **24**, 760–766. xvii, 68, 70, 71, 74

Cotari, J.W., Voisinne, G., Dar, O.E., Karabacak, V. & Altan-Bonnet, G. (2013b). Cell-to-cell variability analysis dissects the plasticity of signaling of common γ chain cytokines in t cells. *Science signaling*, **6**, ra17–ra17. 7, 8, 32, 57, 65, 66, 74, 75, 86, 87, 89, 96, 97, 98, 110, 116, 121, 150, 151, 153, 216

Cox, D., Little, J. & OShea, D. (2013). *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media. 17, 20, 21, 23, 25, 31

Curtiss, D. (1918). Recent extentions of descartes' rule of signs. *Annals of Mathematics*, 251–278. 42

Davidson, A. & Diamond, B. (2001). Autoimmune diseases. *New England Journal of Medicine*, **345**, 340–350. 218

De Belly, H., Paluch, E.K. & Chalut, K.J. (2022). Interplay between mechanics and signalling in regulating cell fate. *Nature Reviews Molecular Cell Biology*, **23**, 465–480. 2, 7

De Boer, R.J. & Perelson, A.S. (2005). Estimating division and death rates from cfse data. *Journal of computational and applied mathematics*, **184**, 140–164. 299

DeAngelis, D.L. & Diaz, S.G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Frontiers in Ecology and Evolution*, **6**, 237. 57

den Braber, I., Mugwagwa, T., Vrisekoop, N., Westera, L., Mögling, R., de Boer, A.B., Willems, N., Schrijver, E.H., Spierenburg, G., Gaiser, K. *et al.* (2012). Maintenance of peripheral naive t cells is sustained by thymus output in mice but not humans. *Immunity*, **36**, 288–297. 299

Di Veroli, G.Y., Fornari, C., Goldlust, I., Mills, G., Koh, S.B., Bramhall, J.L., Richards, F.M. & Jodrell, D.I. (2015). An automated fitting procedure and software for dose-response curves with multiphasic features. *Scientific reports*, **5**, 1–11. 120

Dickenstein, A. & Pérez Millán, M. (2011). How far is complex balancing from detailed balancing? *Bulletin of mathematical biology*, **73**, 811–828. 30

Dickenstein, A., Millan, M.P., Shiu, A. & Tang, X. (2019). Multistationarity in structured reaction networks. *Bulletin of Mathematical Biology*, **81**, 1527–1581. 31

Diz-Pita, É. & Otero-Espinar, M.V. (2021). Predator–prey models: A review of some recent advances. *Mathematics*, **9**, 1783. 11

Douglass Jr, E.F., Miller, C.J., Sparer, G., Shapiro, H. & Spiegel, D.A. (2013). A comprehensive mathematical model for three-body binding equilibria. *Journal of the American Chemical Society*, **135**, 6092–6099. 120, 214

Du, Z. & Lovly, C.M. (2018). Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular cancer*, **17**, 1–13. 4, 9, 66, 121, 215, 304

Duke, R. & Cohen, J. (1986). Il-2 addiction: withdrawal of growth factor activates a suicide program in dependent t cells. *Lymphokine research*, **5**, 289–299. 277

Dushek, O., Aleksic, M., Wheeler, R.J., Zhang, H., Cordoba, S.P., Peng, Y.C., Chen, J.L., Cerundolo, V., Dong, T., Coombs, D. *et al.* (2011). Antigen potency and maximal efficacy reveal a mechanism of efficient t cell activation. *Science signaling*, **4**, ra39–ra39. 7, 32

Eberl, G., Colonna, M., Di Santo, J.P. & McKenzie, A.N. (2015). Innate lymphoid cells: a new paradigm in immunology. *Science*, **348**, aaa6566. 282

Eftimie, R., Gillard, J.J. & Cantrell, D.A. (2016). Mathematical models for immunology: current state of the art and future research directions. *Bulletin of mathematical biology*, **78**, 2091–2134. 120

Eladdadi, A. & Isaacson, D. (2008). A mathematical model for the effects of her2 overexpression on cell proliferation in breast cancer. *Bulletin of mathematical biology*, **70**, 1707–1729. 9, 66

Fajgenbaum, D.C. & June, C.H. (2020). Cytokine storm. *New England Journal of Medicine*, **383**, 2255–2273. 214

Fallon, E.M. & Lauffenburger, D.A. (2000). Computational model for effects of ligand/receptor binding properties on interleukin-2 trafficking dynamics and t cell proliferation response. *Biotechnology progress*, **16**, 905–916. 8, 11

Farhat, A.M., Weiner, A.C., Posner, C., Kim, Z.S., Orcutt-Jahns, B., Carlson, S.M. & Meyer, A.S. (2021). Modeling cell-specific dynamics and regulation of the common gamma chain cytokines. *Cell reports*, **35**, 109044. 6, 7, 8, 57, 66, 216

Feinberg, M. (1987). Chemical reaction network structure and the stability of complex isothermal reactors—i. the deficiency zero and deficiency one theorems. *Chemical engineering science*, **42**, 2229–2268. 30, 120

Feinberg, M. (1989). Necessary and sufficient conditions for detailed balancing in mass action systems of arbitrary complexity. *Chemical Engineering Science*, **44**, 1819–1827. 30

Feinerman, O., Jentsch, G., Tkach, K.E., Coward, J.W., Hathorn, M.M., Sneddon, M.W., Emonet, T., Smith, K.A. & Altan-Bonnet, G. (2010). Single-cell quantification of il-2 response by effector and regulatory t cells reveals critical plasticity in immune response. *Molecular systems biology*, **6**, 437. 7, 8, 32, 57, 66, 121, 220, 298

Finlay, D.B., Duffull, S.B. & Glass, M. (2020). 100 years of modelling ligand–receptor binding and response: A focus on gpcrs. *British journal of pharmacology*, **177**, 1472–1484. 7

Franck, S., Block, A., Bloh, W., Bounama, C., Garrido, I. & Schellnhuber, H.J. (2001). Planetary habitability: is earth commonplace in the milky way? *Naturwissenschaften*, **88**, 416–426. 3

Gabrielsson, J., Peletier, L.A. & Hjorth, S. (2018). Lost in translation: what's in an ec50? innovative pk/pd reasoning in the drug development context. *European journal of pharmacology*, **835**, 154–161. 120

Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B. & Tosaki, A. (2012). The hill equation and the origin of quantitative pharmacology. *Archive for history of exact sciences*, **66**, 427–438. 7, 95, 119

REFERENCES

Gonnord, P., Angermann, B.R., Sadtler, K., Gombos, E., Chappert, P., Meier-Schellersheim, M. & Varma, R. (2018). A hierarchy of affinities between cytokine receptors and the common gamma chain leads to pathway cross-talk. *Science Signaling*, **11**, eaal1253. 7, 8, 32, 65, 216

Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M. & Maire, P. (2008). The hill equation: a review of its capabilities in pharmacological modelling. *Fundamental & clinical pharmacology*, **22**, 633–648. 7, 95, 119

Grayson, D.R. & Stillman, M.E. (????). Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/. 26, 79, 130, 133, 139, 193, 327, 337

Gross, E., Harrington, H.A., Rosen, Z. & Sturmfels, B. (2016). Algebraic systems biology: a case study for the wnt pathway. *Bulletin of mathematical biology*, **78**, 21–51. 31

Gross, E., Harrington, H., Meshkat, N. & Shiu, A. (2020). Joining and decomposing reaction networks. *Journal of mathematical biology*, **80**, 1683–1731. 213

Haan, C., Kreis, S., Margue, C. & Behrmann, I. (2006). Jaks and cytokine receptors—an intimate relationship. *Biochemical pharmacology*, **72**, 1538–1546. xvii, 5

Hemmers, S., Schizas, M., Azizi, E., Dikiy, S., Zhong, Y., Feng, Y., Altan-Bonnet, G. & Rudensky, A.Y. (2019). Il-2 production by self-reactive cd4 thymocytes scales regulatory t cell generation in the thymus. *Journal of Experimental Medicine*, **216**, 2466–2478. 220

Higuera, L.d.l., López-García, M., Lythe, G. & Molina-París, C. (2017). Il-2 stimulation of regulatory t cells: A stochastic and algorithmic approach. In *Modeling cellular systems*, 81–105, Springer. 220

Higuera, L.d.l., López-García, M., Lythe, G. & Molina-París, C. (2021). Ctla-4-mediated ligand trans-endocytosis: A stochastic model. *Mathematical, Computational and Experimental T Cell Immunology*, 257–280. 8

Ho, K.L. & Harrington, H.A. (2010). Bistability in apoptosis by receptor clustering. *PLoS computational biology*, **6**, e1000956. 77, 124

Höfer, T., Krichevsky, O. & Altan-Bonnet, G. (2012). Competition for il-2 between regulatory and effector t cells to chisel immune responses. *Frontiers in immunology*, **3**, 268. 220, 245, 279

Hogan, T., Shuvaev, A., Commenges, D., Yates, A., Callard, R., Thiebaut, R. & Seddon, B. (2013). Clonally diverse t cell homeostasis is maintained by a common program of cell-cycle control. *The Journal of Immunology*, **190**, 3985–3993. 299

Hong, C., Luckey, M.A. & Park, J.H. (2012). Intrathymic il-7: the where, when, and why of il-7 signaling during t cell development. In *Seminars in immunology*, vol. 24, 151–158, Elsevier. 2

Horn, F. & Jackson, R. (1972). General mass action kinetics. *Archive for rational mechanics and analysis*, **47**, 81–116. 28

Humrich, J.Y., Morbach, H., Undeutsch, R., Enghard, P., Rosenberger, S., Weigert, O., Kloke, L., Heimann, J., Gaber, T., Brandenburg, S. *et al.* (2010). Homeostatic imbalance of regulatory and effector t cells due to il-2 deprivation amplifies murine lupus. *Proceedings of the National Academy of Sciences*, **107**, 204–209. 281

Janes, K.A. & Lauffenburger, D.A. (2013). Models of signalling networks–what cell biologists can gain from them and give to them. *J Cell Sci*, **126**, 1913–1921. 32

Jiang, X. & Kopp-Schneider, A. (2014). Summarizing ec50 estimates from multiple dose-response experiments: A comparison of a meta-analysis strategy to a mixed-effects model approach. *Biometrical Journal*, **56**, 493–512. 7, 120

Kalia, V., Sarkar, S., Subramaniam, S., Haining, W.N., Smith, K.A. & Ahmed, R. (2010). Prolonged interleukin-2rα expression on virus-specific cd8+ t cells favors terminal-effector differentiation in vivo. *Immunity*, **32**, 91–103. 215, 304

Kang, J. & Coles, M. (2012). Il-7: the global builder of the innate lymphoid network and beyond, one niche at a time. In *Seminars in immunology*, vol. 24, 190–197, Elsevier. 282

Keshtkar, E., Kudsk, P. & Mesgaran, M.B. (2021). Perspective: common errors in dose–response analysis and how to avoid them. *Pest Management Science*, **77**, 2599–2608. 7

Kim, H.P., Imbert, J. & Leonard, W.J. (2006). Both integrated and differential regulation of components of the il-2/il-2 receptor system. *Cytokine & growth factor reviews*, **17**, 349–366. 220

Kim, J.M., Rasmussen, J.P. & Rudensky, A.Y. (2007a). Regulatory t cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nature immunology*, **8**, 191–197. 220

Kim, P.S., Lee, P.P. & Levy, D. (2007b). Modeling regulation mechanisms in the immune system. *Journal of theoretical biology*, **246**, 33–69. 300

Koch, A.L. (1966). The logarithm in biology 1. mechanisms generating the log-normal distribution exactly. *Journal of theoretical biology*, **12**, 276–290. 47

Lahl, K., Loddenkemper, C., Drouin, C., Freyer, J., Arnason, J., Eberl, G., Hamann, A., Wagner, H., Huehn, J. & Sparwasser, T. (2007). Selective depletion of foxp3+ regulatory t cells induces a scurfy-like disease. *The Journal of experimental medicine*, **204**, 57–63. 220

Lambert, D. (2004). Drugs and receptors. *Continuing education in anaesthesia critical care & pain*, **4**, 181–184. 7

Larsen, S.E., Voss, K., Laing, E.D. & Snow, A.L. (2017). Differential cytokine withdrawal-induced death sensitivity of effector t cells derived from distinct human cd8+ memory subsets. *Cell death discovery*, **3**, 1–8. 277

Lauffenburger, D.A. & Linderman, J. (1996). *Receptors: models for binding, trafficking, and signaling*. Oxford University Press. 7, 120

Lee, H.M., Bautista, J.L., Scott-Browne, J., Mohan, J.F. & Hsieh, C.S. (2012). A broad range of self-reactivity drives thymic regulatory t cell selection to limit responses to self. *Immunity*, **37**, 475–486. 220

Leonard, W.J., Lin, J.X. & O'Shea, J.J. (2019). The γc family of cytokines: basic biology to therapeutic ramifications. *Immunity*, **50**, 832–850. 4, 7, 32, 65, 75

Létourneau, S., Krieg, C., Pantaleo, G. & Boyman, O. (2009). Il-2–and cd25-dependent immunoregulatory mechanisms in the homeostasis of t-cell subsets. *Journal of Allergy and Clinical Immunology*, **123**, 758–762. 281

Li, J.L., Liu, X.Y., Xie, J.T., Di, Y.L. & Zhu, F.X. (2015). A comparison of different estimation methods for fungicide ec50 and ec95 values. *Journal of Phytopathology*, **163**, 239–244. 8, 120

Liao, L.E., Carruthers, J., Smither, S.J., Team, C.V., Weller, S.A., Williamson, D., Laws, T.R., García-Dorival, I., Hiscox, J., Holder, B.P. *et al.* (2020). Quantification of ebola virus replication kinetics in vitro. *PLoS computational biology*, **16**, e1008375. 11

Lin, J.X. & Leonard, W.J. (2018). The common cytokine receptor γ chain family of cytokines. *Cold Spring Harbor perspectives in biology*, **10**, a028449. 6, 7

Lin, J.X. & Leonard, W.J. (2019). Fine-tuning cytokine signals. *Annual Review of Immunology*, **37**, 295–324. 5, 6, 65, 215

Liu, Z., Gerner, M.Y., Van Panhuys, N., Levine, A.G., Rudensky, A.Y. & Germain, R.N. (2015). Immune homeostasis enforced by co-localized effector and regulatory t cells. *Nature*, **528**, 225–230. 220, 299

Locke, M., Lythe, G., López-García, M., Muñoz-Fontela, C., Carroll, M. & Molina-París, C. (2021). Quantification of type i interferon inhibition by viral proteins: Ebola virus as a case study. *Viruses*, **13**, 2441. 11

Ma, A., Koka, R. & Burkett, P. (2006). Diverse functions of il-2, il-15, and il-7 in lymphoid homeostasis. *Annu. Rev. Immunol.*, **24**, 657–679. 65

Macdougall, J. (2006). Analysis of dose–response studies—e max model. In *Dose finding in drug development*, 127–145, Springer. 120

Mack, E.T., Perez-Castillejos, R., Suo, Z. & Whitesides, G.M. (2008). Exact analysis of ligand-induced dimerization of monomeric receptors. *Analytical chemistry*, **80**, 5550–5555. 120, 125, 214

Marrack, P., Kappler, J. & Kotzin, B.L. (2001). Autoimmune disease: why and where it occurs. *Nature medicine*, **7**, 899–905. 218

Martin, C.E., Spasova, D.S., Frimpong-Boateng, K., Kim, H.O., Lee, M., Kim, K.S. & Surh, C.D. (2017). Interleukin-7 availability is maintained by a hematopoietic cytokine sink comprising innate lymphoid cells and t cells. *Immunity*, **47**, 171–182. 282, 286

Maruyama, I.N. (2014). Mechanisms of activation of receptor tyrosine kinases: monomers or dimers. *Cells*, **3**, 304–330. 125

Maxwell, S.R. & Webb, D.J. (2008). Receptor functions. *Medicine*, **36**, 344–349. 2, 7, 33

McCaughtry, T.M. & Hogquist, K.A. (2008). Central tolerance: what have we learned from mice? In *Seminars in immunopathology*, vol. 30, 399–409, Springer. 218

McKeithan, T.W. (1995). Kinetic proofreading in t-cell receptor signal transduction. *Proceedings of the national academy of sciences*, **92**, 5042–5046. 8

Michie, C.A., McLean, A., Alcock, C. & Beverley, P.C. (1992). Lifespan of human lymphocyte subsets defined by cd45 isoforms. *Nature*, **360**, 264–265. 299

Molina-París, C., Reynolds, J., Lythe, G. & Coles, M.C. (2013). Mathematical model of naive t cell division and survival il-7 thresholds. *Frontiers in immunology*, **4**, 434. 7, 8, 65

Moraga, I., Spangler, J.B., Mendoza, J.L., Gakovic, M., Wehrman, T.S., Krutzik, P. & Garcia, K.C. (2017). Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers. *Elife*, **6**, e22882. 121

Murphy, K., Travers, P., Walport, M. *et al.* (2011). *Janeway's immunobiology*. Taylor & Francis. 1

Murugan, A., Mora, T., Walczak, A.M. & Callan Jr, C.G. (2012). Statistical inference of the generation probability of t-cell receptors from sequence repertoires. *Proceedings of the National Academy of Sciences*, **109**, 16161–16166. 2

Nakayama, T., Wiest, D.L., Abraham, K.M., Munitz, T.I., Perlmutter, R.M. & Singer, A. (1993). Decreased signaling competence as a result of receptor overexpression: overexpression of cd4 reduces its ability to activate p56lck tyrosine kinase and to regulate t-cell antigen receptor expression in immature cd4+ cd8+ thymocytes. *Proceedings of the National Academy of Sciences*, **90**, 10534–10538. 75

Oliphant, T.E. (2006). *A guide to NumPy*, vol. 1. Trelgol Publishing USA. 111

Oliveira, J.B. (2013). Evaluation of il-2-withdrawal-induced apoptosis in human t lymphocytes. In *Immune Homeostasis*, 25–31, Springer. 278

O'Shea, J.J., Schwartz, D.M., Villarino, A.V., Gadina, M., McInnes, I.B. & Laurence, A. (2015). The jak-stat pathway: impact on human disease and therapeutic intervention. *Annual review of medicine*, **66**, 311–328. 5

Otero-Muras, I., Yordanov, P. & Stelling, J. (2017). Chemical reaction network theory elucidates sources of multistability in interferon signaling. *PLoS computational biology*, **13**, e1005454. 120

Owen, D.L., Mahmud, S.A., Vang, K.B., Kelly, R.M., Blazar, B.R., Smith, K.A. & Farrar, M.A. (2018). Identification of cellular sources of il-2 needed for regulatory t cell development and homeostasis. *The Journal of Immunology*, **200**, 3926–3933. 220

Oyler-Yaniv, A., Oyler-Yaniv, J., Whitlock, B.M., Liu, Z., Germain, R.N., Huse, M., Altan-Bonnet, G. & Krichevsky, O. (2017). A tunable diffusion-consumption mechanism of cytokine propagation enables plasticity in cell-to-cell communication in the immune system. *Immunity*, **46**, 609–620. 218, 220, 299

Palmer, M.J., Mahajan, V.S., Trajman, L.C., Irvine, D.J., Lauffenburger, D.A. & Chen, J. (2008). Interleukin-7 receptor signaling network: an integrated systems perspective. *Cellular & molecular immunology*, **5**, 79–89. 65

Pandiyan, P., Zheng, L., Ishihara, S., Reed, J. & Lenardo, M.J. (2007). Cd4+ cd25+ foxp3+ regulatory t cells induce cytokine deprivation–mediated apoptosis of effector cd4+ t cells. *Nature immunology*, **8**, 1353–1362. 277

Park, J.H., Waickman, A.T., Reynolds, J., Castro, M. & Molina-París, C. (2019). Il7 receptor signaling in t cells: A mathematical modeling perspective. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **11**, e1447. 7, 8, 65, 215, 283, 284, 302

Pérez Millán, M. & Dickenstein, A. (2015). Implicit dose-response curves. *Journal of Mathematical Biology*, **70**, 1669–1684. 34

Picot, J., Guerin, C.L., Le Van Kim, C. & Boulanger, C.M. (2012). Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology*, **64**, 109–130. xvii, 68, 69

Pillet, A.H., Lavergne, V., Pasquier, V., Gesbert, F., Thèze, J. & Rose, T. (2010). Il-2 induces conformational changes in its preassembled receptor core, which then migrates in lipid raft and binds to the cytoskeleton meshwork. *Journal of molecular biology*, **403**, 671–692. 74

Prabakaran, S., Gunawardena, J. & Sontag, E. (2014). Paradoxical results in perturbation-based signaling network reconstruction. *Biophysical journal*, **106**, 2720–2728. 74

Raeber, M.E., Zurbuchen, Y., Impellizzieri, D. & Boyman, O. (2018). The role of cytokines in t-cell memory in health and disease. *Immunological reviews*, **283**, 176–193. 7

Regad, T. (2015). Targeting rtk signaling pathways in cancer. *Cancers*, **7**, 1758–1784. 9, 66

Reynolds, J., Coles, M., Lythe, G. & Molina-París, C. (2013). Mathematical model of naive T cell division and survival IL-7 thresholds. *Frontiers in Immunology*, **4**, 434. 283, 284, 302

Reynolds, J., Amado, I.F., Freitas, A.A., Lythe, G. & Molina-París, C. (2014). A mathematical perspective on $CD4^+$ T cell quorum-sensing. *Journal of Theoretical Biology*, **347**, 160–175. 220

Rieseberg, M., Kasper, C., Reardon, K.F. & Scheper, T. (2001). Flow cytometry in biotechnology. *Applied microbiology and biotechnology*, **56**, 350–360. 68

Ring, A.M., Lin, J.X., Feng, D., Mitra, S., Rickert, M., Bowman, G.R., Pande, V.S., Li, P., Moraga, I., Spolski, R. *et al.* (2012). Mechanistic and structural insight into the functional dichotomy between il-2 and il-15. *Nature immunology*, **13**, 1187–1195. 7, 8, 216

Rochman, Y., Spolski, R. & Leonard, W.J. (2009). New insights into the regulation of t cells by γ c family cytokines. *Nature Reviews Immunology*, **9**, 480–490. xvii, 3, 6, 7, 65, 67, 91, 99, 138, 175

Rosenstein, E.D., Zucker, M.J. & Kramer, N. (2000). Giant cell myocarditis: most fatal of autoimmune diseases. In *Seminars in arthritis and rheumatism*, vol. 30, 1–16, Elsevier. 219

Rovati, G.E. & Nicosia, S. (1994). Lower efficacy: interaction with an inhibitory receptor or partial agonism? *Trends in pharmacological sciences*, **15**, 140–144. 120

Sadeghimanesh, A. & Feliu, E. (2019). Gröbner bases of reaction networks with intermediate species. *Advances in Applied Mathematics*, **107**, 74–101. 31

Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. (1995). Immunologic self-tolerance maintained by activated t cells expressing il-2 receptor alpha-chains (cd25). breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *The Journal of Immunology*, **155**, 1151–1164. 220, 281

Savarin, C. & Bergmann, C.C. (2018). Fine tuning the cytokine storm by ifn and il-10 following neurotropic coronavirus encephalomyelitis. *Frontiers in immunology*, **9**, 3022. 214

Sheikh, A. & Abraham, N. (2019). Interleukin-7 receptor alpha in innate lymphoid cells: more than a marker. *Frontiers in Immunology*, **10**, 2897. 282

Shi, T., Niepel, M., McDermott, J.E., Gao, Y., Nicora, C.D., Chrisler, W.B., Markillie, L.M., Petyuk, V.A., Smith, R.D., Rodland, K.D. *et al.* (2016). Conservation of protein abundance patterns reveals the regulatory architecture of the egfr-mapk pathway. *Science Signaling*, **9**, rs6–rs6. 153

Shipkova, M. & Wieland, E. (2012). Surface markers of lymphocyte activation and markers of cell proliferation. *Clinica chimica acta*, **413**, 1338–1349. 3

Shiu, A.J. (2010). *Algebraic methods for biochemical reaction network theory*. University of California, Berkeley. 120

Shtiegman, K., Kochupurakkal, B., Zwang, Y., Pines, G., Starr, A., Vexler, A., Citri, A., Katz, M., Lavi, S., Ben-Basat, Y. *et al.* (2007). Defective ubiquitinylation of egfr mutants of lung cancer confers prolonged signaling. *Oncogene*, **26**, 6968–6978. 116, 153

Simmonds, J.G. & Mann, J.E. (2013). *A first look at perturbation theory*. Courier Corporation. 36, 38, 103, 141

Solomon, D. *et al.* (1971). Time to extinction of a pure death process. 51

Sompayrac, L.M. (2019). *How the immune system works*. John Wiley & Sons. 1, 2, 3, 4, 217

Sonnenberg, G.F. & Hepworth, M.R. (2019). Functional interactions between innate lymphoid cells and adaptive immunity. *Nature Reviews Immunology*, **19**, 599–613. 282

Spolski, R., Gromer, D. & Leonard, W.J. (2017). The γ c family of cytokines: fine-tuning signals from il-2 and il-21 in the regulation of the immune response. *F1000Research*, **6**. 214

Sta, L., Adamer, M. & Molina-París, C. (2022a). Algebraic study of receptor-ligand systems: a dose-response analysis. *arXiv preprint arXiv:2206.13364*. x, 11

Sta, L., Voisinne, G., Cotari, J., Adamer, M., Molina-París, C. & Altan-Bonnet, G. (2022b). Tuning of cytokine signaling through imbalanced abundances of receptors and kinases. *bioRxiv*. x, xviii, 68, 74, 90, 112, 113, 116, 117

Stauber, D.J., Debler, E.W., Horton, P.A., Smith, K.A. & Wilson, I.A. (2006). Crystal structure of the il-2 signaling complex: paradigm for a heterotrimeric cytokine receptor. *Proceedings of the National Academy of Sciences*, **103**, 2788–2793. 181

Steelman, L., Pohnert, S., Shelton, J., Franklin, R., Bertrand, F. & McCubrey, J. (2004). Jak/stat, raf/mek/erk, pi3k/akt and bcr-abl in cell cycle progression and leukemogenesis. *Leukemia*, **18**, 189–218. 5

Strasser, A. & Pellegrini, M. (2004). T-lymphocyte death during shutdown of an immune response. *Trends in immunology*, **25**, 610–615. 277

Suriyatem, R., Auras, R.A., Intipunya, P. & Rachtanapun, P. (2017). Predictive mathematical modeling for ec50 calculation of antioxidant activity and antibacterial ability of thai bee products. *Journal of Applied Pharmaceutical Science*, **7**, 122–133. 8, 120

Ten Broeke, G., Van Voorn, G. & Ligtenberg, A. (2016). Which sensitivity analysis method should i use for my agent-based model? *Journal of Artificial Societies and Social Simulation*, **19**, 5. 58

Thiele, J.C., Kurth, W. & Grimm, V. (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using netlogo and r. *Journal of Artificial Societies and Social Simulation*, **17**, 11. 58

Thomas, N. (2006). Hypothesis testing and bayesian estimation using a sigmoid e max model applied to sparse dose-response designs. *Journal of biopharmaceutical statistics*, **16**, 657–677. 120

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, **6**, 187–202. 55, 56, 57

Trotta, E., Bessette, P.H., Silveria, S.L., Ely, L.K., Jude, K.M., Le, D.T., Holst, C.R., Coyle, A., Potempa, M., Lanier, L.L. *et al.* (2018). A human anti-il-2 antibody that potentiates regulatory t cells by a structure-based mechanism. *Nature medicine*, **24**, 1005–1014. 302, 305

Truong, V.T., Baverel, P.G., Lythe, G.D., Vicini, P., Yates, J.W. & Dubois, V.F. (2022). Step-by-step comparison of ordinary differential equation and agent-based approaches to pharmacokinetic-pharmacodynamic models. *CPT: pharmacometrics & systems pharmacology*, **11**, 133–148. 57

Tsai, E. (2014). A method for reducing ill-conditioning of polynomial root finding using a change of basis. 111

Tyson, J.J., Chen, K.C. & Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current opinion in cell biology*, **15**, 221–231. 120

Uings, I. & Farrow, S. (2000). Cell receptors and cell signalling. *Molecular Pathology*, **53**, 295. 2, 32

Varma, R. (2008). Tcr triggering by the pmhc complex: valency, affinity, and dynamics. *Science signaling*, **1**, pe21–pe21. 2

Vignali, D.A. & Kuchroo, V.K. (2012). Il-12 family cytokines: immunological playmakers. *Nature immunology*, **13**, 722–728. 175

Villarino, A.V., Kanno, Y., Ferdinand, J.R. & O?Shea, J.J. (2015). Mechanisms of jak/stat signaling in immunity and disease. *The Journal of Immunology*, **194**, 21–27. 5, 7

Vivier, E., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J.P., Eberl, G., Koyasu, S., Locksley, R.M., McKenzie, A.N., Mebius, R.E. *et al.* (2018). Innate lymphoid cells: 10 years on. *Cell*, **174**, 1054–1066. 282

Vogel, R.M., Erez, A. & Altan-Bonnet, G. (2016). Dichotomy of cellular inhibition by small-molecule inhibitors revealed by single-cell analysis. *Nature communications*, **7**, 1–10. 74, 215, 304

Voisinne, G., Nixon, B.G., Melbinger, A., Gasteiger, G., Vergassola, M. & Altan-Bonnet, G. (2015). T cells integrate local and global cues to discriminate between structurally similar antigens. *Cell reports*, **11**, 1208–1219. 215, 220, 304

Waickman, A.T., Keller, H.R., Kim, T.H., Luckey, M.A., Tai, X., Hong, C., Molina-París, C., Walsh, S.T. & Park, J.H. (2020). The cytokine receptor il-7rα impairs il-2 receptor signaling and constrains the in vitro differentiation of foxp3+ treg cells. *Iscience*, **23**, 101421. 86

Wang, X., Rickert, M. & Garcia, K.C. (2005). Structure of the quaternary complex of interleukin-2 with its α, ß, and γc receptors. *Science*, **310**, 1159–1163. 181

Weyand, C.M. & Goronzy, J.J. (2021). The immunology of rheumatoid arthritis. *Nature immunology*, **22**, 10–18. 218

White, C., Rottschäfer, V. & Bridge, L. (2022). Insights into the dynamics of ligand-induced dimerisation via mathematical modelling and analysis. *Journal of Theoretical Biology*, **538**, 110996. 120, 185

Wiley, H.S., Shvartsman, S.Y. & Lauffenburger, D.A. (2003). Computational modeling of the egf-receptor system: a paradigm for systems biology. *Trends in cell biology*, **13**, 43–50. 7, 120

Wilmes, S., Jeffrey, P.A., Martinez-Fabregas, J., Hafer, M., Fyfe, P.K., Pohler, E., Gaggero, S., López-García, M., Lythe, G., Taylor, C. *et al.* (2021). Competitive binding of stats to receptor phospho-tyr motifs accounts for altered cytokine responses. *Elife*, **10**, e66014. 8

Wolfram Research, Inc. (2019). Mathematica, Version 12.0. Champaign, IL. 26, 142, 146, 188, 202, 283, 285, 329, 343

Wong, H.S., Park, K., Gola, A., Baptista, A.P., Miller, C.H., Deep, D., Lou, M., Boyd, L.F., Rudensky, A.Y., Savage, P.A. *et al.* (2021). A local regulatory t cell feedback circuit maintains immune homeostasis by pruning self-activated t cells. *Cell*, **184**, 3981–3997. 220, 299

Yarden, Y. & Schlessinger, J. (1987). Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. *Biochemistry*, **26**, 1443–1451. 125

Zehn, D. & Bevan, M.J. (2006). T cells with low avidity for a tissue-restricted antigen routinely evade central and peripheral tolerance and cause autoimmunity. *Immunity*, **25**, 261–270. 218

Żołądek, H. (2000). The topological proof of abel-ruffini theorem. *Topological Methods in Nonlinear Analysis*, **16**, 253–265. 35

Appendix A

Macaulay2 code to compute Gröbner bases

Most Gröbner bases of this thesis have been computed making use of Macaulay2 (Grayson & Stillman). We provide the code to compute the Gröbner basis of the IL-7R model described in Section 3.2.

```
R = frac(QQ[Nx,Ny,Nz,L,K1,K2,K3])[x,y,z,MonomialOrder=>Lex]

I = ideal(
- Nx + x + K2*x*y + K1*x*z + K2*K_1*x*y*z
        + K3*K2*L*x*y + K3*K2*K1*L*x*y*z,
- Ny + y + K2*x*y + K2*K_1*x*y*z
        + K3*K2*L*x*y + K3*K2*K1*L*x*y*z,
- Nz + z + K1*x*z + K2*K_1*x*y*z + K3*K2*K1*L*x*y*z
)

g = gens gb I
```

A. MACAULAY2 CODE TO COMPUTE GRÖBNER BASES

Appendix B

Gröbner basis for the steady state of the allostery model

I include here the Gröbner basis of polynomial system (3.31) computed with Mathematica (Wolfram Research, Inc., 2019). Solving this system of polynomial equations would give the number of unbound γ_c, IL-7Rα and JAK3 chains at steady state. In Mathematica, K_2' and K_3' are written $K2p$ and $K3p$ respectively.

B. GRÖBNER BASIS FOR THE STEADY STATE OF THE ALLOSTERY MODEL

```
In[1]:= GroebnerBasis[
  {-Nx + K1 * x * z + K2 * K1 * x * y * z + K2p * x * y + K3p * K2p * L * x * y + K3 * K2 * K1 * L * x * y * z + x,
   -Ny + K2 * K1 * x * y * z + K2p * x * y + K3p * K2p * L * x * y + K3 * K2 * K1 * L * x * y * z + y,
   -Nz + z + K1 * z * x + K1 * K2 * x * y * z + K3 * K2 * K1 * L * x * y * z}, {x, y, z}]
```

Out[1]= {K2p^2 Nz^2 + 2 K2p^2 K3p L Nz^2 + K2p^2 K3p^2 L^2 Nz^2 − K1 K2 Nz z + K1 K2p Nz z −
2 K2p^2 Nz z − K1 K2 K3 L Nz z + K1 K2p K3p L Nz z − 4 K2p^2 K3p L Nz z − 2 K2p^2 K3p^2 L^2 Nz z −
K1 K2 K2p Nx Nz z − K1 K2p^2 Nx Nz z − K1 K2 K2p K3 L Nx Nz z − K1 K2 K2p K3p L Nx Nz z −
2 K1 K2p^2 K3p L Nx Nz z − K1 K2 K2p K3 K3p L^2 Nx Nz z − K1 K2p^2 K3p^2 L^2 Nx Nz z − K1 K2 K2p Ny Nz z +
K1 K2p^2 Ny Nz z − K1 K2 K2p K3 L Ny Nz z − K1 K2 K2p K3p L Ny Nz z + 2 K1 K2p^2 K3p L Ny Nz z −
K1 K2 K2p K3 K3p L^2 Ny Nz z + K1 K2p^2 K3p^2 L^2 Ny Nz z + 2 K1 K2 K2p Nz^2 z + K1 K2p^2 Nz^2 z +
2 K1 K2 K2p K3 L Nz^2 z + 2 K1 K2 K2p K3p L Nz^2 z + 2 K1 K2p^2 K3p L Nz^2 z + 2 K1 K2 K2p K3 K3p L^2 Nz^2 z +
K1 K2p^2 K3p^2 L^2 Nz^2 z + K1 K2 z^2 − K1 K2p z^2 + K2p^2 z^2 + K1 K2 K3 L z^2 − K1 K2p K3p L z^2 +
2 K2p^2 K3p L z^2 + K2p^2 K3p^2 L^2 z^2 + K1^2 K2 Nx z^2 − K1^2 K2p Nx z^2 + K1 K2 K2p Nx z^2 + K1 K2p^2 Nx z^2 +
K1^2 K2 K3 L Nx z^2 + K1 K2 K2p K3 L Nx z^2 − K1^2 K2p K3p L Nx z^2 + K1 K2 K2p K3p L Nx z^2 +
2 K1 K2p^2 K3p L Nx z^2 + K1 K2 K2p K3 K3p L^2 Nx z^2 + K1 K2p^2 K3p^2 L^2 Nx z^2 + K1^2 K2 K2p Nx^2 z^2 +
K1^2 K2 K2p K3 L Nx^2 z^2 + K1^2 K2 K2p K3p L Nx^2 z^2 + K1^2 K2 K2p K3 K3p L^2 Nx^2 z^2 + K1 K2 K2p Ny z^2 −
K1 K2p^2 Ny z^2 + K1 K2 K2p K3 L Ny z^2 + K1 K2 K2p K3p L Ny z^2 − 2 K1 K2p^2 K3p L Ny z^2 +
K1 K2 K2p K3 K3p L^2 Ny z^2 − K1 K2p^2 K3p^2 L^2 Ny z^2 + K1^2 K2^2 Nx Ny z^2 − K1^2 K2 K2p Nx Ny z^2 +
2 K1^2 K2^2 K3 L Nx Ny z^2 − K1^2 K2 K2p K3 L Nx Ny z^2 − K1^2 K2 K2p K3p L Nx Ny z^2 + K1^2 K2^2 K3^2 L^2 Nx Ny z^2 −
K1^2 K2 K2p K3 K3p L^2 Nx Ny z^2 − K1^2 K2 Nz z^2 + K1^2 K2p Nz z^2 − 4 K1 K2 K2p Nz z^2 − 2 K1 K2p^2 Nz z^2 −
K1^2 K2 K3 L Nz z^2 − 4 K1 K2 K2p K3 L Nz z^2 + K1^2 K2p K3p L Nz z^2 − 4 K1 K2 K2p K3p L Nz z^2 −
4 K1 K2p^2 K3p L Nz z^2 − 4 K1 K2 K2p K3 K3p L^2 Nz z^2 − 2 K1 K2p^2 K3p^2 L^2 Nz z^2 −
K1^2 K2^2 Nx Nz z^2 − 3 K1^2 K2 K2p Nx Nz z^2 − 2 K1^2 K2^2 K3 L Nx Nz z^2 − 3 K1^2 K2 K2p K3 L Nx Nz z^2 −
3 K1^2 K2 K2p K3p L Nx Nz z^2 − K1^2 K2^2 K3^2 L^2 Nx Nz z^2 − 3 K1^2 K2 K2p K3 K3p L^2 Nx Nz z^2 −
K1^2 K2^2 Ny Nz z^2 + K1^2 K2 K2p Ny Nz z^2 − 2 K1^2 K2^2 K3 L Ny Nz z^2 + K1^2 K2 K2p K3 L Ny Nz z^2 +
K1^2 K2 K2p K3p L Ny Nz z^2 − K1^2 K2^2 K3^2 L^2 Ny Nz z^2 + K1^2 K2 K2p K3 K3p L^2 Ny Nz z^2 +
K1^2 K2^2 Nz^2 z^2 + 2 K1^2 K2 K2p Nz^2 z^2 + 2 K1^2 K2^2 K3 L Nz^2 z^2 + 2 K1^2 K2 K2p K3 L Nz^2 z^2 +
2 K1^2 K2 K2p K3p L Nz^2 z^2 + K1^2 K2^2 K3^2 L^2 Nz^2 z^2 + 2 K1^2 K2 K2p K3 K3p L^2 Nz^2 z^2 + K1^2 K2 z^3 −
K1^2 K2p z^3 + 2 K1 K2 K2p z^3 + K1 K2p^2 z^3 + K1^2 K2 K3 L z^3 + 2 K1 K2 K2p K3 L z^3 − K1^2 K2p K3p L z^3 +
2 K1 K2 K2p K3p L z^3 + 2 K1 K2p^2 K3p L z^3 + 2 K1 K2 K2p K3 K3p L^2 z^3 + K1 K2p^2 K3p^2 L^2 z^3 +
K1^2 K2^2 Nx z^3 + 3 K1^2 K2 K2p Nx z^3 + 2 K1^2 K2^2 K3 L Nx z^3 + 3 K1^2 K2 K2p K3 L Nx z^3 +
3 K1^2 K2 K2p K3p L Nx z^3 + K1^2 K2^2 K3^2 L^2 Nx z^3 + 3 K1^2 K2 K2p K3 K3p L^2 Nx z^3 + K1^3 K2^2 Nx^2 z^3 +
2 K1^3 K2^2 K3 L Nx^2 z^3 + K1^3 K2^2 K3^2 L^2 Nx^2 z^3 + K1^2 K2^2 Ny z^3 − K1^2 K2 K2p Ny z^3 + 2 K1^2 K2^2 K3 L Ny z^3 −
K1^2 K2 K2p K3 L Ny z^3 − K1^2 K2 K2p K3p L Ny z^3 + K1^2 K2^2 K3^2 L^2 Ny z^3 − K1^2 K2 K2p K3 K3p L^2 Ny z^3 −
2 K1^2 K2^2 Nz z^3 − 4 K1^2 K2 K2p Nz z^3 − 4 K1^2 K2^2 K3 L Nz z^3 − 4 K1^2 K2 K2p K3 L Nz z^3 −
4 K1^2 K2 K2p K3p L Nz z^3 − 2 K1^2 K2^2 K3^2 L^2 Nz z^3 − 4 K1^2 K2 K2p K3 K3p L^2 Nz z^3 −
2 K1^3 K2^2 Nx Nz z^3 − 4 K1^3 K2^2 K3 L Nx Nz z^3 − 2 K1^3 K2^2 K3^2 L^2 Nx Nz z^3 + K1^3 K2^2 Nz^2 z^3 +
2 K1^3 K2^2 K3 L Nz^2 z^3 + K1^3 K2^2 K3^2 L^2 Nz^2 z^3 + K1^2 K2^2 z^4 + 2 K1^2 K2 K2p z^4 + 2 K1^2 K2^2 K3 L z^4 +
2 K1^2 K2 K2p K3 L z^4 + 2 K1^2 K2 K2p K3p L z^4 + K1^2 K2^2 K3^2 L^2 z^4 + 2 K1^2 K2 K2p K3 K3p L^2 z^4 +
2 K1^3 K2^2 Nx z^4 + 4 K1^3 K2^2 K3 L Nx z^4 + 2 K1^3 K2^2 K3^2 L^2 Nx z^4 − 2 K1^3 K2^2 Nz z^4 −
4 K1^3 K2^2 K3 L Nz z^4 − 2 K1^3 K2^2 K3^2 L^2 Nz z^4 + K1^3 K2^2 z^5 + 2 K1^3 K2^2 K3 L z^5 + K1^3 K2^2 K3^2 L^2 z^5,

$$
\begin{aligned}
&- K1\,K2\,Nz + K1\,K2p\,Nz - K2p^2\,Nz - K1\,K2\,K3\,L\,Nz + K1\,K2p\,K3p\,L\,Nz - 2\,K2p^2\,K3p\,L\,Nz - K2p^2\,K3p^2\,L^2\,Nz - \\
&\quad K1\,K2\,K2p\,Nx\,Nz - K1\,K2\,K2p\,K3\,L\,Nx\,Nz - K1\,K2\,K2p\,K3p\,L\,Nx\,Nz - K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,Nz + \\
&\quad K1\,K2\,K2p\,Nz^2 + K1\,K2\,K2p\,K3\,L\,Nz^2 + K1\,K2\,K2p\,K3p\,L\,Nz^2 + K1\,K2\,K2p\,K3\,K3p\,L^2\,Nz^2 - \\
&\quad K1\,K2\,K2p\,Nz\,y + K1\,K2p^2\,Nz\,y - K1\,K2\,K2p\,K3\,L\,Nz\,y - K1\,K2\,K2p\,K3p\,L\,Nz\,y + 2\,K1\,K2p^2\,K3p\,L\,Nz\,y - \\
&\quad K1\,K2\,K2p\,K3\,K3p\,L^2\,Nz\,y + K1\,K2p^2\,K3p^2\,L^2\,Nz\,y + K1\,K2\,z - K1\,K2p\,z + K2p^2\,z + K1\,K2\,K3\,L\,z - \\
&\quad K1\,K2p\,K3p\,L\,z + 2\,K2p^2\,K3p\,L\,z + K2p^2\,K3p^2\,L^2\,z + K1^2\,K2\,Nx\,z - K1^2\,K2p\,Nx\,z + K1\,K2\,K2p\,Nx\,z + \\
&\quad K1\,K2p^2\,Nx\,z + K1^2\,K2\,K3\,L\,Nx\,z + K1\,K2\,K2p\,K3\,L\,Nx\,z - K1^2\,K2p\,K3p\,L\,Nx\,z + K1\,K2\,K2p\,K3p\,L\,Nx\,z + \\
&\quad 2\,K1\,K2p^2\,K3p\,L\,Nx\,z + K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,z + K1\,K2p^2\,K3p^2\,L^2\,Nx\,z + K1^2\,K2\,K2p\,Nx^2\,z + \\
&\quad K1^2\,K2\,K2p\,K3\,L\,Nx^2\,z + K1^2\,K2\,K2p\,K3p\,L\,Nx^2\,z + K1\,K2p^2\,K3p\,L^2\,Nx^2\,z + K1\,K2\,K2p\,Ny\,z - \\
&\quad K1\,K2p^2\,Ny\,z + K1\,K2\,K2p\,K3\,L\,Ny\,z + K1\,K2\,K2p\,K3p\,L\,Ny\,z - 2\,K1\,K2p^2\,K3p\,L\,Ny\,z + \\
&\quad K1\,K2\,K2p\,K3\,K3p\,L^2\,Ny\,z - K1\,K2p^2\,K3p^2\,L^2\,Ny\,z + K1^2\,K2^2\,Nx\,Ny\,z - K1^2\,K2\,K2p\,Nx\,Ny\,z + \\
&\quad 2\,K1^2\,K2^2\,K3\,L\,Nx\,Ny\,z - K1^2\,K2\,K2p\,K3\,L\,Nx\,Ny\,z - K1^2\,K2\,K2p\,K3p\,L\,Nx\,Ny\,z + K1^2\,K2^2\,K3^2\,L^2\,Nx\,Ny\,z - \\
&\quad K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,Ny\,z - K1^2\,K2\,Nz\,z + K1^2\,K2p\,Nz\,z - 3\,K1\,K2\,K2p\,Nz\,z - K1\,K2p^2\,Nz\,z - \\
&\quad K1^2\,K2\,K3\,L\,Nz\,z - 3\,K1\,K2\,K2p\,K3\,L\,Nz\,z + K1^2\,K2p\,K3p\,L\,Nz\,z - 3\,K1\,K2\,K2p\,K3p\,L\,Nz\,z - \\
&\quad 2\,K1\,K2p^2\,K3p\,L\,Nz\,z - 3\,K1\,K2\,K2p\,K3\,K3p\,L^2\,Nz\,z - K1\,K2p^2\,K3p^2\,L^2\,Nz\,z - K1^2\,K2^2\,Nx\,Nz\,z - \\
&\quad 2\,K1^2\,K2\,K2p\,Nx\,Nz\,z - 2\,K1^2\,K2^2\,K3\,L\,Nx\,Nz\,z - 2\,K1^2\,K2\,K2p\,K3\,L\,Nx\,Nz\,z - 2\,K1^2\,K2\,K2p\,K3p\,L\,Nx\,Nz\,z - \\
&\quad K1^2\,K2^2\,K3^2\,L^2\,Nx\,Nz\,z - 2\,K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,Nz\,z - K1^2\,K2^2\,Ny\,Nz\,z + K1^2\,K2\,K2p\,Ny\,Nz\,z - \\
&\quad 2\,K1^2\,K2^2\,K3\,L\,Ny\,Nz\,z + K1^2\,K2\,K2p\,K3\,L\,Ny\,Nz\,z + K1^2\,K2\,K2p\,K3p\,L\,Ny\,Nz\,z - K1^2\,K2^2\,K3^2\,L^2\,Ny\,Nz\,z + \\
&\quad K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Ny\,Nz\,z + K1^2\,K2^2\,Nz^2\,z + K1^2\,K2\,K2p\,Nz^2\,z + 2\,K1^2\,K2^2\,K3\,L\,Nz^2\,z + \\
&\quad K1^2\,K2\,K2p\,K3\,L\,Nz^2\,z + K1^2\,K2\,K2p\,K3p\,L\,Nz^2\,z + K1^2\,K2^2\,K3^2\,L^2\,Nz^2\,z + K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Nz^2\,z + \\
&\quad K1^2\,K2\,z^2 - K1^2\,K2p\,z^2 + 2\,K1\,K2\,K2p\,z^2 + K1\,K2p^2\,z^2 + K1^2\,K2\,K3\,L\,z^2 + 2\,K1\,K2\,K2p\,K3\,L\,z^2 - \\
&\quad K1^2\,K2p\,K3p\,L\,z^2 + 2\,K1\,K2\,K2p\,K3p\,L\,z^2 + 2\,K1\,K2p^2\,K3p\,L\,z^2 + 2\,K1\,K2\,K2p\,K3\,K3p\,L^2\,z^2 + \\
&\quad K1\,K2p^2\,K3p^2\,L^2\,z^2 + K1^2\,K2^2\,Nx\,z^2 + 3\,K1^2\,K2\,K2p\,Nx\,z^2 + 2\,K1^2\,K2^2\,K3\,L\,Nx\,z^2 + 3\,K1^2\,K2\,K2p\,K3\,L\,Nx\,z^2 + \\
&\quad 3\,K1^2\,K2\,K2p\,K3p\,L\,Nx\,z^2 + K1^2\,K2^2\,K3^2\,L^2\,Nx\,z^2 + 3\,K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,z^2 + K1^3\,K2^2\,Nx^2\,z^2 + \\
&\quad 2\,K1^3\,K2^2\,K3\,L\,Nx^2\,z^2 + K1^3\,K2^2\,K3^2\,L^2\,Nx^2\,z^2 + K1^2\,K2^2\,Ny\,z^2 - K1^2\,K2\,K2p\,Ny\,z^2 + 2\,K1^2\,K2^2\,K3\,L\,Ny\,z^2 - \\
&\quad K1^2\,K2\,K2p\,K3\,L\,Ny\,z^2 - K1^2\,K2\,K2p\,K3p\,L\,Ny\,z^2 + K1^2\,K2^2\,K3^2\,L^2\,Ny\,z^2 - K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Ny\,z^2 - \\
&\quad 2\,K1^2\,K2^2\,Nz\,z^2 - 3\,K1^2\,K2\,K2p\,Nz\,z^2 - 4\,K1^2\,K2^2\,K3\,L\,Nz\,z^2 - 3\,K1^2\,K2\,K2p\,K3\,L\,Nz\,z^2 - \\
&\quad 3\,K1^2\,K2\,K2p\,K3p\,L\,Nz\,z^2 - 2\,K1^2\,K2^2\,K3^2\,L^2\,Nz\,z^2 - 3\,K1^2\,K2\,K2p\,K3\,K3p\,L^2\,Nz\,z^2 - \\
&\quad 2\,K1^3\,K2^2\,Nx\,Nz\,z^2 - 4\,K1^3\,K2^2\,K3\,L\,Nx\,Nz\,z^2 - 2\,K1^3\,K2^2\,K3^2\,L^2\,Nx\,Nz\,z^2 + K1^3\,K2^2\,Nz^2\,z^2 + \\
&\quad 2\,K1^3\,K2^2\,K3\,L\,Nz^2\,z^2 + K1^3\,K2^2\,K3^2\,L^2\,Nz^2\,z^2 + K1^2\,K2^2\,z^3 + 2\,K1^2\,K2\,K2p\,z^3 + 2\,K1^2\,K2^2\,K3\,L\,z^3 + \\
&\quad 2\,K1^2\,K2\,K2p\,K3\,L\,z^3 + 2\,K1^2\,K2\,K2p\,K3p\,L\,z^3 + K1^2\,K2^2\,K3^2\,L^2\,z^3 + 2\,K1^2\,K2\,K2p\,K3\,K3p\,L^2\,z^3 + \\
&\quad 2\,K1^3\,K2^2\,Nx\,z^3 + 4\,K1^3\,K2^2\,K3\,L\,Nx\,z^3 + 2\,K1^3\,K2^2\,K3^2\,L^2\,Nx\,z^3 - 2\,K1^3\,K2^2\,Nz\,z^3 - \\
&\quad 4\,K1^3\,K2^2\,K3\,L\,Nz\,z^3 - 2\,K1^3\,K2^2\,K3^2\,L^2\,Nz\,z^3 + K1^3\,K2^2\,z^4 + 2\,K1^3\,K2^2\,K3\,L\,z^4 + K1^3\,K2^2\,K3^2\,L^2\,z^4,
\end{aligned}
$$

$$
\begin{aligned}
&- K2p\,Nz - K2p\,K3p\,L\,Nz + K2p\,z + K2p\,K3p\,L\,z + K1\,K2p\,Nx\,z + K1\,K2p\,K3p\,L\,Nx\,z + K1\,K2\,Ny\,z - \\
&\quad K1\,K2p\,Ny\,z + K1\,K2\,K3\,L\,Ny\,z - K1\,K2p\,K3p\,L\,Ny\,z - K1\,K2\,Nz\,z - K1\,K2p\,Nz\,z - \\
&\quad K1\,K2\,K3\,L\,Nz\,z - K1\,K2p\,K3p\,L\,Nz\,z - K1\,K2\,y\,z + K1\,K2p\,y\,z - K1\,K2\,K3\,L\,y\,z + \\
&\quad K1\,K2p\,K3p\,L\,y\,z + K1\,K2\,z^2 + K1\,K2p\,z^2 + K1\,K2\,K3\,L\,z^2 + K1\,K2p\,K3p\,L\,z^2 + K1^2\,K2\,Nx\,z^2 + \\
&\quad K1^2\,K2\,K3\,L\,Nx\,z^2 - K1^2\,K2\,Nz\,z^2 - K1^2\,K2\,K3\,L\,Nz\,z^2 + K1^2\,K2\,z^3 + K1^2\,K2\,K3\,L\,z^3,
\end{aligned}
$$

$$
\begin{aligned}
&K2\,Nz - K2p\,Nz + K2\,K3\,L\,Nz - K2p\,K3p\,L\,Nz + K2\,K2p\,Nx\,Nz + K2\,K2p\,K3\,L\,Nx\,Nz + K2\,K2p\,K3p\,L\,Nx\,Nz + \\
&\quad K2\,K2p\,K3\,K3p\,L^2\,Nx\,Nz - K2\,K2p\,Nz^2 - K2\,K2p\,K3\,L\,Nz^2 - K2\,K2p\,K3p\,L\,Nz^2 - K2\,K2p\,K3\,K3p\,L^2\,Nz^2 + \\
&\quad K2\,K2p\,Nz\,y - K2p^2\,Nz\,y + K2\,K2p\,K3\,L\,Nz\,y + K2\,K2p\,K3p\,L\,Nz\,y - 2\,K2p^2\,K3p\,L\,Nz\,y +
\end{aligned}
$$

B. GRÖBNER BASIS FOR THE STEADY STATE OF THE ALLOSTERY MODEL

$$
\begin{aligned}
&K2\,K2p\,K3\,K3p\,L^2\,Nz\,y - K2p^2\,K3p^2\,L^2\,Nz\,y - K2\,z + K2p\,z - K2\,K3\,L\,z + K2p\,K3p\,L\,z - K1\,K2\,Nx\,z + \\
&K1\,K2p\,Nx\,z - K2\,K2p\,Nx\,z - K1\,K2\,K3\,L\,Nx\,z - K2\,K2p\,K3\,L\,Nx\,z + K1\,K2p\,K3p\,L\,Nx\,z - \\
&K2\,K2p\,K3p\,L\,Nx\,z - K2\,K2p\,K3\,K3p\,L^2\,Nx\,z - K1\,K2\,K2p\,Nx^2\,z - K1\,K2\,K2p\,K3\,L\,Nx^2\,z - \\
&K1\,K2\,K2p\,K3p\,L\,Nx^2\,z - K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx^2\,z - K1\,K2^2\,Nx\,Ny\,z + K1\,K2\,K2p\,Nx\,Ny\,z - \\
&2\,K1\,K2^2\,K3\,L\,Nx\,Ny\,z + K1\,K2\,K2p\,K3\,L\,Nx\,Ny\,z + K1\,K2\,K2p\,K3p\,L\,Nx\,Ny\,z - K1\,K2^2\,K3^2\,L^2\,Nx\,Ny\,z + \\
&K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,Ny\,z + K1\,K2\,Nz\,z - K1\,K2p\,Nz\,z + 2\,K2\,K2p\,Nz\,z + K1\,K2\,K3\,L\,Nz\,z + \\
&2\,K2\,K2p\,K3\,L\,Nz\,z - K1\,K2p\,K3p\,L\,Nz\,z + 2\,K2\,K2p\,K3p\,L\,Nz\,z + 2\,K2\,K2p\,K3\,K3p\,L^2\,Nz\,z + K1\,K2^2\,Nx\,Nz\,z + \\
&2\,K1\,K2\,K2p\,Nx\,Nz\,z + 2\,K1\,K2^2\,K3\,L\,Nx\,Nz\,z + 2\,K1\,K2\,K2p\,K3\,L\,Nx\,Nz\,z + 2\,K1\,K2\,K2p\,K3p\,L\,Nx\,Nz\,z + \\
&K1\,K2^2\,K3^2\,L^2\,Nx\,Nz\,z + 2\,K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,Nz\,z + K1\,K2^2\,Ny\,Nz\,z - K1\,K2\,K2p\,Ny\,Nz\,z + \\
&2\,K1\,K2^2\,K3\,L\,Ny\,Nz\,z - K1\,K2\,K2p\,K3\,L\,Ny\,Nz\,z - K1\,K2\,K2p\,K3p\,L\,Ny\,Nz\,z + K1\,K2^2\,K3^2\,L^2\,Ny\,Nz\,z - \\
&K1\,K2\,K2p\,K3\,K3p\,L^2\,Ny\,Nz\,z - K1\,K2^2\,Nz^2\,z - K1\,K2\,K2p\,Nz^2\,z - 2\,K1\,K2^2\,K3\,L\,Nz^2\,z - \\
&K1\,K2\,K2p\,K3\,L\,Nz^2\,z - K1\,K2\,K2p\,K3p\,L\,Nz^2\,z - K1\,K2^2\,K3^2\,L^2\,Nz^2\,z - K1\,K2\,K2p\,K3\,K3p\,L^2\,Nz^2\,z - \\
&K2\,K2p\,y\,z + K2p^2\,y\,z - K2\,K2p\,K3\,L\,y\,z - K2\,K2p\,K3p\,L\,y\,z + 2\,K2p^2\,K3p\,L\,y\,z - K2\,K2p\,K3\,K3p\,L^2\,y\,z + \\
&K2p^2\,K3p^2\,L^2\,y\,z - K1\,K2\,z^2 + K1\,K2p\,z^2 - K2\,K2p\,z^2 - K1\,K2\,K3\,L\,z^2 - K2\,K2p\,K3\,L\,z^2 + \\
&K1\,K2p\,K3p\,L\,z^2 - K2\,K2p\,K3p\,L\,z^2 - K2\,K2p\,K3\,K3p\,L^2\,z^2 - K1\,K2^2\,Nx\,z^2 - 2\,K1\,K2\,K2p\,Nx\,z^2 - \\
&2\,K1\,K2^2\,K3\,L\,Nx\,z^2 - 2\,K1\,K2\,K2p\,K3\,L\,Nx\,z^2 - 2\,K1\,K2\,K2p\,K3p\,L\,Nx\,z^2 - K1\,K2^2\,K3^2\,L^2\,Nx\,z^2 - \\
&2\,K1\,K2\,K2p\,K3\,K3p\,L^2\,Nx\,z^2 - K1^2\,K2^2\,Nx^2\,z^2 - 2\,K1^2\,K2^2\,K3\,L\,Nx^2\,z^2 - K1^2\,K2^2\,K3^2\,L^2\,Nx^2\,z^2 - \\
&K1\,K2^2\,Ny\,z^2 + K1\,K2\,K2p\,Ny\,z^2 - 2\,K1\,K2^2\,K3\,L\,Ny\,z^2 + K1\,K2\,K2p\,K3\,L\,Ny\,z^2 + K1\,K2\,K2p\,K3p\,L\,Ny\,z^2 - \\
&K1\,K2^2\,K3^2\,L^2\,Ny\,z^2 + K1\,K2\,K2p\,K3\,K3p\,L^2\,Ny\,z^2 + 2\,K1\,K2^2\,Nz\,z^2 + 2\,K1\,K2\,K2p\,Nz\,z^2 + \\
&4\,K1\,K2^2\,K3\,L\,Nz\,z^2 + 2\,K1\,K2\,K2p\,K3\,L\,Nz\,z^2 + 2\,K1\,K2\,K2p\,K3p\,L\,Nz\,z^2 + 2\,K1\,K2^2\,K3^2\,L^2\,Nz\,z^2 + \\
&2\,K1\,K2\,K2p\,K3\,K3p\,L^2\,Nz\,z^2 + 2\,K1^2\,K2^2\,Nx\,Nz\,z^2 + 4\,K1^2\,K2^2\,K3\,L\,Nx\,Nz\,z^2 + 2\,K1^2\,K2^2\,K3^2\,L^2\,Nx\,Nz\,z^2 - \\
&K1^2\,K2^2\,Nz^2\,z^2 - 2\,K1^2\,K2^2\,K3\,L\,Nz^2\,z^2 - K1^2\,K2^2\,K3^2\,L^2\,Nz^2\,z^2 - K1\,K2^2\,z^3 - K1\,K2\,K2p\,z^3 - \\
&2\,K1\,K2^2\,K3\,L\,z^3 - K1\,K2\,K2p\,K3\,L\,z^3 - K1\,K2\,K2p\,K3p\,L\,z^3 - K1\,K2^2\,K3^2\,L^2\,z^3 - K1\,K2\,K2p\,K3\,K3p\,L^2\,z^3 - \\
&2\,K1^2\,K2^2\,Nx\,z^3 - 4\,K1^2\,K2^2\,K3\,L\,Nx\,z^3 - 2\,K1^2\,K2^2\,K3^2\,L^2\,Nx\,z^3 + 2\,K1^2\,K2^2\,Nz\,z^3 + \\
&4\,K1^2\,K2^2\,K3\,L\,Nz\,z^3 + 2\,K1^2\,K2^2\,K3^2\,L^2\,Nz\,z^3 - K1^2\,K2^2\,z^4 - 2\,K1^2\,K2^2\,K3\,L\,z^4 - K1^2\,K2^2\,K3^2\,L^2\,z^4,
\end{aligned}
$$

$$
\begin{aligned}
&-Nz - K2p\,Nz\,y - K2p\,K3p\,L\,Nz\,y + z + K1\,Nx\,z - K1\,Nz\,z + K2p\,y\,z + K2p\,K3p\,L\,y\,z + K1\,K2\,Nx\,y\,z + \\
&K1\,K2\,K3\,L\,Nx\,y\,z - K1\,K2\,Nz\,y\,z - K1\,K2\,K3\,L\,Nz\,y\,z + K1\,z^2 + K1\,K2\,y\,z^2 + K1\,K2\,K3\,L\,y\,z^2,
\end{aligned}
$$

$$
\begin{aligned}
&K2\,K3\,Nz - K2p\,K3p\,Nz + K2\,K2p\,K3\,Nx\,Nz + K2\,K2p\,K3\,K3p\,L\,Nx\,Nz - K2\,K2p\,K3\,Nz^2 - \\
&K2\,K2p\,K3\,K3p\,L\,Nz^2 + K2\,K2p\,K3\,Nz\,y - K2p^2\,K3p\,Nz\,y + K2\,K2p\,K3\,K3p\,L\,Nz\,y - K2p^2\,K3p^2\,L\,Nz\,y - \\
&K2\,K3\,z + K2p\,K3p\,z - K1\,K2\,K3\,Nx\,z - K2\,K2p\,K3\,Nx\,z + K1\,K2p\,K3p\,Nx\,z - K2\,K2p\,K3\,K3p\,L\,Nx\,z - \\
&K1\,K2\,K2p\,K3\,Nx^2\,z - K1\,K2\,K2p\,K3\,K3p\,L\,Nx^2\,z - K1\,K2^2\,K3\,Nx\,Ny\,z + K1\,K2\,K2p\,K3\,Nx\,Ny\,z - \\
&K1\,K2^2\,K3^2\,L\,Nx\,Ny\,z + K1\,K2\,K2p\,K3\,K3p\,L\,Nx\,Ny\,z + K1\,K2\,K3\,Nz\,z + 2\,K2\,K2p\,K3\,Nz\,z - \\
&K1\,K2p\,K3p\,Nz\,z + 2\,K2\,K2p\,K3\,K3p\,L\,Nz\,z + K1\,K2^2\,K3\,Nx\,Nz\,z + 2\,K1\,K2\,K2p\,K3\,Nx\,Nz\,z + \\
&K1\,K2^2\,K3^2\,L\,Nx\,Nz\,z + 2\,K1\,K2\,K2p\,K3\,K3p\,L\,Nx\,Nz\,z + K1\,K2^2\,K3\,Ny\,Nz\,z - K1\,K2\,K2p\,K3\,Ny\,Nz\,z + \\
&K1\,K2^2\,K3^2\,L\,Ny\,Nz\,z - K1\,K2\,K2p\,K3\,K3p\,L\,Ny\,Nz\,z - K1\,K2^2\,K3\,Nz^2\,z - K1\,K2\,K2p\,K3\,Nz^2\,z - \\
&K1\,K2^2\,K3^2\,L\,Nz^2\,z - K1\,K2\,K2p\,K3\,K3p\,L\,Nz^2\,z - K2\,K2p\,K3\,y\,z + K2p^2\,K3p\,y\,z - K2\,K2p\,K3\,K3p\,L\,y\,z + \\
&K2p^2\,K3p^2\,L\,y\,z - K1\,K2\,K2p\,K3\,Nx\,y\,z + K1\,K2\,K2p\,K3p\,Nx\,y\,z + K1\,K2\,K2p\,K3\,Nz\,y\,z - \\
&K1\,K2\,K2p\,K3p\,Nz\,y\,z - K1\,K2\,K3\,z^2 - K2\,K2p\,K3\,z^2 + K1\,K2p\,K3p\,z^2 - K2\,K2p\,K3\,K3p\,L\,z^2 - \\
&K1\,K2^2\,K3\,Nx\,z^2 - 2\,K1\,K2\,K2p\,K3\,Nx\,z^2 - K1\,K2^2\,K3^2\,L\,Nx\,z^2 - 2\,K1\,K2\,K2p\,K3\,K3p\,L\,Nx\,z^2 - \\
&K1^2\,K2^2\,K3\,Nx^2\,z^2 - K1^2\,K2^2\,K3^2\,L\,Nx^2\,z^2 - K1\,K2^2\,K3\,Ny\,z^2 + K1\,K2\,K2p\,K3\,Ny\,z^2 - \\
&K1\,K2^2\,K3^2\,L\,Ny\,z^2 + K1\,K2\,K2p\,K3\,K3p\,L\,Ny\,z^2 + 2\,K1\,K2^2\,K3\,Nz\,z^2 + 2\,K1\,K2\,K2p\,K3\,Nz\,z^2 +
\end{aligned}
$$

$$
\begin{aligned}
&2\,K1\,K2^2\,K3^2\,L\,Nz\,z^2 + 2\,K1\,K2\,K2p\,K3\,K3p\,L\,Nz\,z^2 + 2\,K1^2\,K2^2\,K3\,Nx\,Nz\,z^2 + 2\,K1^2\,K2^2\,K3^2\,L\,Nx\,Nz\,z^2 - \\
&K1^2\,K2^2\,K3\,Nz^2\,z^2 - K1^2\,K2^2\,K3^2\,L\,Nz^2\,z^2 - K1\,K2\,K2p\,K3\,y\,z^2 + K1\,K2\,K2p\,K3p\,y\,z^2 - \\
&K1\,K2^2\,K3\,z^3 - K1\,K2\,K2p\,K3\,z^3 - K1\,K2^2\,K3^2\,L\,z^3 - K1\,K2\,K2p\,K3\,K3p\,L\,z^3 - 2\,K1^2\,K2^2\,K3\,Nx\,z^3 - \\
&2\,K1^2\,K2^2\,K3^2\,L\,Nx\,z^3 + 2\,K1^2\,K2^2\,K3\,Nz\,z^3 + 2\,K1^2\,K2^2\,K3^2\,L\,Nz\,z^3 - K1^2\,K2^2\,K3\,z^4 - K1^2\,K2^2\,K3^2\,L\,z^4,
\end{aligned}
$$

$$
\begin{aligned}
&-Ny + Nz + y + K2p\,Nx\,y + K2p\,K3p\,L\,Nx\,y - K2p\,Ny\,y - K2p\,K3p\,L\,Ny\,y + K2p\,y^2 + K2p\,K3p\,L\,y^2 - \\
&z - K1\,Nx\,z + K1\,Nz\,z - K1\,z^2,\ -Nz + z + K1\,Nx\,z - K1\,Ny\,z - K1\,Nz\,z + K1\,y\,z + K1\,K2\,Nx\,y\,z + \\
&K1\,K2\,K3\,L\,Nx\,y\,z - K1\,K2\,Ny\,y\,z - K1\,K2\,K3\,L\,Ny\,y\,z + K1\,K2\,y^2\,z + K1\,K2\,K3\,L\,y^2\,z + K1\,z^2,
\end{aligned}
$$

$$
\begin{aligned}
&-K2p\,K3p\,Nz + K2p\,K3p\,z + K1\,K2p\,K3p\,Nx\,z + K1\,K2\,K3\,Ny\,z - K1\,K2p\,K3p\,Ny\,z - K1\,K2\,K3\,Nz\,z - \\
&K1\,K2p\,K3p\,Nz\,z - K1\,K2\,K3\,y\,z + K1\,K2p\,K3p\,y\,z - K1\,K2\,K2p\,K3\,Nx\,y\,z + K1\,K2\,K2p\,K3p\,Nx\,y\,z + \\
&K1\,K2\,K2p\,K3\,Ny\,y\,z - K1\,K2\,K2p\,K3p\,Ny\,y\,z - K1\,K2\,K2p\,K3\,y^2\,z + K1\,K2\,K2p\,K3p\,y^2\,z + \\
&K1\,K2\,K3\,z^2 + K1\,K2p\,K3p\,z^2 + K1^2\,K2\,K3\,Nx\,z^2 - K1^2\,K2\,K3\,Nz\,z^2 + K1^2\,K2\,K3\,z^3,
\end{aligned}
$$

$$
\begin{aligned}
&K2\,Nx - K2p\,Nx + K2\,K3\,L\,Nx - K2p\,K3p\,L\,Nx - K2\,Ny + K2p\,Ny - K2\,K3\,L\,Ny + K2p\,K3p\,L\,Ny + K2p\,Nz + \\
&K2p\,K3p\,L\,Nz - K2\,x + K2p\,x - K2\,K3\,L\,x + K2p\,K3p\,L\,x + K2\,y - K2p\,y + K2\,K3\,L\,y - K2p\,K3p\,L\,y - K2p\,z - \\
&K2p\,K3p\,L\,z - K1\,K2\,Nx\,z - K1\,K2\,K3\,L\,Nx\,z + K1\,K2\,Nz\,z + K1\,K2\,K3\,L\,Nz\,z - K1\,K2\,z^2 - K1\,K2\,K3\,L\,z^2,
\end{aligned}
$$

$$
\begin{aligned}
&-Nx + Ny + x - y + K1\,x\,z,\ -Nx + Ny + Nz + x - y - K2\,Nx\,y - K2\,K3\,L\,Nx\,y + \\
&K2\,Ny\,y + K2\,K3\,L\,Ny\,y + K2\,x\,y + K2\,K3\,L\,x\,y - K2\,y^2 - K2\,K3\,L\,y^2 - z,
\end{aligned}
$$

$$
\begin{aligned}
&K2\,K3\,Nx - K2p\,K3p\,Nx - K2\,K3\,Ny + K2p\,K3p\,Ny + K2p\,K3p\,Nz - K2\,K3\,x + K2p\,K3p\,x + K2\,K3\,y - \\
&K2p\,K3p\,y + K2\,K2p\,K3\,Nx\,y - K2\,K2p\,K3p\,Nx\,y - K2\,K2p\,K3\,Ny\,y + K2\,K2p\,K3p\,Ny\,y - K2\,K2p\,K3\,x\,y + \\
&K2\,K2p\,K3p\,x\,y + K2\,K2p\,K3\,y^2 - K2\,K2p\,K3p\,y^2 - K2p\,K3p\,z - K1\,K2\,K3\,Nx\,z + K1\,K2\,K3\,Nz\,z - K1\,K2\,K3\,z^2\}
\end{aligned}
$$

B. GRÖBNER BASIS FOR THE STEADY STATE OF THE ALLOSTERY MODEL

Appendix C

Gröbner basis for the steady state of the model of homodimeric receptor with IEK

The Gröbner basis of the polynomial system describing the homodimeric receptor model with IEK is so long that Mathematica crashes everytime I want to save the notebook. Instead, I present the first polynomial of the basis, which is a univariate polynomial of degree 6.

C. GRÖBNER BASIS FOR THE STEADY STATE OF THE MODEL OF HOMODIMERIC RECEPTOR WITH IEK

```
In[1]:= FullSimplify[
 CoefficientList[-K1 Nx - K3 L Nx + K1 x + K3 L x - K1^2 Nx x - 2 K1 K3 L Nx x + K1^2 K3 L Nx x + K1^2 Nz x +
   K1 K3 L Nz x + K1^2 x^2 + 2 K1 K2 x^2 + 2 K1 K3 L x^2 - K1^2 K3 L x^2 + 2 K2 K3 L x^2 + 2 K1 K2 K3 L x^2 +
   2 K2 K3^2 L^2 x^2 - K1^2 K3 L Nx x^2 + K1^3 K3 L Nx x^2 - 2 K1 K2 K3 L Nx x^2 + 2 K1^2 K2 K3 L Nx x^2 -
   2 K1 K2 K3^2 L^2 Nx x^2 + 2 K1^2 K2 K3^2 L^2 Nx x^2 + 2 K1^3 K2 Nx^2 x^2 + 4 K1^3 K2 K3 L Nx^2 x^2 +
   2 K1^3 K2 K3^2 L^2 Nx^2 x^2 + 2 K1^2 K2 Nz x^2 + K1^2 K3 L Nz x^2 - K1^3 K3 L Nz x^2 + 2 K1 K2 K3 L Nz x^2 +
   2 K1^2 K2 K3 L Nz x^2 + 2 K1 K2 K3^2 L^2 Nz x^2 - 4 K1^3 K2 Nx Nz x^2 - 4 K1^2 K2 K3 L Nx Nz x^2 -
   4 K1^3 K2 K3 L Nx Nz x^2 - 4 K1^2 K2 K3^2 L^2 Nx Nz x^2 + 2 K1^3 K2 Nz^2 x^2 + 4 K1^2 K2 K3 L Nz^2 x^2 +
   2 K1 K2 K3^2 L^2 Nz^2 x^2 + 2 K1^2 K2 x^3 + K1^2 K3 L x^3 - K1^3 K3 L x^3 + 6 K1 K2 K3 L x^3 - 2 K1^2 K2 K3 L x^3 +
   6 K1 K2 K3^2 L^2 x^3 - 4 K1^2 K2 K3^2 L^2 x^3 - 3 K1^3 K2 Nx x^3 - 2 K1^2 K2 K3 L Nx x^3 - 4 K1^3 K2 K3 L Nx x^3 -
   2 K1^2 K2 K3^2 L^2 Nx x^3 - K1^3 K2 K3^2 L^2 Nx x^3 + 3 K1^3 K2 Nz x^3 + 7 K1^2 K2 K3 L Nz x^3 -
   K1^3 K2 K3 L Nz x^3 + 7 K1^2 K2 K3^2 L^2 Nz x^3 - 4 K1^3 K2 K3^2 L^2 Nz x^3 + K1^3 K2 x^4 + 4 K1^2 K2 K3 L x^4 -
   2 K1^3 K2 K3 L x^4 + 4 K1 K2^2 K3 L x^4 - 4 K1^2 K2^2 K3 L x^4 + 4 K1^2 K2 K3^2 L^2 x^4 - 3 K1^3 K2 K3^2 L^2 x^4 +
   8 K1 K2^2 K3^2 L^2 x^4 - 8 K1^2 K2^2 K3^2 L^2 x^4 + 4 K1 K2^2 K3^3 L^3 x^4 - 4 K1^2 K2^2 K3^3 L^3 x^4 -
   7 K1^3 K2^2 Nx x^4 - K1^2 K2^2 K3 L Nx x^4 - 20 K1^3 K2^2 K3 L Nx x^4 - 2 K1^2 K2^2 K3^2 L^2 Nx x^4 -
   19 K1^3 K2^2 K3^2 L^2 Nx x^4 - K1^2 K2^2 K3^3 L^3 Nx x^4 - 6 K1^3 K2^2 K3^3 L^3 Nx x^4 + 6 K1^3 K2^2 Nz x^4 +
   10 K1^2 K2^2 K3 L Nz x^4 + 8 K1^3 K2^2 K3 L Nz x^4 + 20 K1^2 K2^2 K3^2 L^2 Nz x^4 - 2 K1^3 K2^2 K3^2 L^2 Nz x^4 +
   10 K1^2 K2^2 K3^3 L^3 Nz x^4 - 4 K1^3 K2^2 K3^3 L^3 Nz x^4 + 5 K1^3 K2^2 x^5 + 5 K1^2 K2^2 K3 L x^5 +
   10 K1^3 K2^2 K3 L x^5 + 10 K1^2 K2^2 K3^2 L^2 x^5 + 5 K1^3 K2^2 K3^2 L^2 x^5 + 5 K1^2 K2^2 K3^3 L^3 x^5 +
   6 K1^3 K2^3 x^6 + 2 K1^2 K2^3 K3 L x^6 + 22 K1^3 K2^3 K3 L x^6 + 6 K1^2 K2^3 K3^2 L^2 x^6 + 30 K1^3 K2^3 K3^2 L^2 x^6 +
   6 K1^2 K2^3 K3^3 L^3 x^6 + 18 K1^3 K2^3 K3^3 L^3 x^6 + 2 K1^2 K2^3 K3^4 L^4 x^6 + 4 K1^3 K2^3 K3^4 L^4 x^6, x]]
Out[1]= {-(K1 + K3 L) Nx, K1 + K3 L + K1 K3 L (-2 Nx + Nz) + K1^2 ((-1 + K3 L) Nx + Nz),
 2 K2 K3 L (1 + K3 L) + K1^3 (2 K2 K3^2 L^2 Nx^2 + K3 L (1 + 4 K2 Nx) (Nx - Nz) + 2 K2 (Nx - Nz)^2) +
  2 K1 (K2 + K3 L + K2 K3 L (1 - (1 + K3 L) Nx + Nz + K3 L Nz (1 + Nz))) +
  K1^2 (1 + 2 K2 K3^2 L^2 Nx (1 - 2 Nz) + 2 K2 Nz + K3 L (-1 + Nx (-1 + K2 (2 - 4 Nz)) + Nz + K2 Nz (2 + 4 Nz))),
 -K1 (-6 K2 K3 L (1 + K3 L) + K1 (-K3 L + K2 (1 + K3 L) (-2 + K3 L (4 + 2 Nx - 7 Nz))) +
     K1^2 (3 K2 Nx + K3 L (1 + K2 (4 + K3 L) Nx) + K2 (1 + K3 L) (-3 + 4 K3 L) Nz)),
 -K1 K2 (1 + K3 L) (-4 K2 K3 L (1 + K3 L) + K1 K3 L (-4 + K2 (1 + K3 L) (4 + Nx - 10 Nz)) +
     K1^2 (-1 + 7 K2 Nx + K3 L (3 + K2 (13 + 6 K3 L) Nx) + 2 K2 (1 + K3 L) (-3 + 2 K3 L) Nz)),
 5 K1^2 K2^2 (1 + K3 L)^2 (K1 + K3 L), 2 K1^2 (K2 + K2 K3 L)^3 (K3 L + K1 (3 + 2 K3 L))}
```

Appendix D

Gröbner basis for the computation of the EC_{50} of the trimeric receptor model

I compute the Gröbner basis (for the EC_{50}) of the trimeric receptor model making use of Macaulay2 (Grayson & Stillman),(P_1, P_2, P_3, P_4, P_5), and adapted the output to Mathematica. The following notebook shows the polynomials which compose the reduced Gröbner basis computed. The solutions of polynomial P_1 are potential expressions for the EC_{50} of the trimeric model. This polynomial is independent on K_1 and N_z (see extracted coefficients at the end of the notebook).

D. GRÖBNER BASIS FOR THE COMPUTATION OF THE EC_{50} OF THE TRIMERIC RECEPTOR MODEL

```
P1[L_] := L^4 +
  ((-4 * Nx * Ny * K2 * K3 * M - 8 * Nx * Nw * K2 * K3 * M - 4 * Ny * Nw * K2 * K3 * M + 6 * Nx * K2 * K3 * M^2 +
        4 * Ny * K2 * K3 * M^2 + 6 * Nw * K2 * K3 * M^2 - 4 * K2 * K3 * M^3 - 4 * Nx * K2 * M - 4 * Nw * K2 * M +
        4 * K2 * M^2 - 4 * M) / (8 * Nx * Ny * Nw * K2 * K3 * K4 - 4 * Nx * Ny * K2 * K3 * K4 * M -
        4 * Nx * Nw * K2 * K3 * K4 * M - 4 * Ny * Nw * K2 * K3 * K4 * M + 2 * Nx * K2 * K3 * K4 * M^2 +
        2 * Ny * 8 * K2 * K3 * K4 * M^2 + 2 * Nw * K2 * K3 * K4 * M^2 - K2 * K3 * K4 * M^3)) * L^3 +
  ((8 * Nx * Ny * K2 * K3^2 * M^2 + 4 * Ny^2 * K2 * K3^2 * M^2 + 4 * Nx * Nw * K2 * K3^2 * M^2 +
        8 * Ny * Nw * K2 * K3^2 * M^2 - 6 * Nx * K2 * K3^2 * M^3 - 12 * Ny * K2 * K3^2 * M^3 -
        6 * Nw * K2 * K3^2 * M^3 + 6 * K2 * K3^2 * M^4 + 4 * Nx * K2 * K3 * M^2 +
        8 * Ny * K2 * K3 * M^2 + 4 * Nw * K2 * K3 * M^2 - 8 * K2 * K3 * M^3 + 4 * K2 * M^2 + 4 * K3 * M^2) /
     (16 * Nx * Ny^2 * Nw * K2 * K3^2 * K4^2 - 8 * Nx * Ny^2 * K2 * K3^2 * K4^2 * M -
        16 * Nx * Ny * Nw * K2 * K3^2 * K4^2 * M - 8 * Ny^2 * Nw * K2 * K3^2 * K4^2 * M +
        8 * Nx * Ny * K2 * K3^2 * K4^2 * M^2 + 4 * Ny^2 * K2 * K3^2 * K4^2 * M^2 +
        4 * Nx * Nw * K2 * K3^2 * K4^2 * M^2 + 8 * Ny * Nw * K2 * K3^2 * K4^2 * M^2 -
        2 * Nx * K2 * K3^2 * K4^2 * M^3 - 4 * Ny * K2 * K3^2 * K4^2 * M^3 -
        2 * Nw * K2 * K3^2 * K4^2 * M^3 + K2 * K3^2 * K4^2 * M^4)) * L^2 +
  ((-2 * Nx * K3 * M^3 - 4 * Ny * K3 * M^3 - 2 * Nw * K3 * M^3 + 4 * K3 * M^4 - 4 * M^3) /
     (16 * Nx * Ny^2 * Nw * K3 * K4^3 - 8 * Nx * Ny^2 * K3 * K4^3 * M - 16 * Nx * Ny * Nw * K3 * K4^3 * M -
        8 * Ny^2 * Nw * K3 * K4^3 * M + 8 * Nx * Ny * K3 * K4^3 M^2 + 4 * Ny^2 * K3 * K4^3 * M^2 +
        4 * Nx * Nw * K3 * K4^3 * M^2 + 8 * Ny * Nw * K3 * K4^3 * M^2 - 2 * Nx * K3 * K4^3 * M^3 -
        4 * Ny * K3 * K4^3 * M^3 - 2 * Nw * K3 * K4^3 * M^3 + K3 * K4^3 * M^4)) * L +
  ((M^4) / (16 * Nx * Ny^2 * Nw * K4^4 - 8 * Nx * Ny^2 * K4^4 * M - 16 * Nx * Ny * Nw * K4^4 * M -
       8 * Ny^2 * Nw * K4^4 * M + 8 * Nx * Ny * K4^4 * M^2 + 4 * Ny^2 * K4^4 * M^2 +
       4 * Nx * Nw * K4^4 * M^2 + 8 * Ny * Nw * K4^4 * M^2 - 2 * Nx * K4^4 * M^3 -
       4 * Ny * K4^4 * M^3 - 2 * Nw * K4^4 * M^3 + K4^4 * M^4))
P2[w_, L_] := w + ((-32 * Nx * Ny^3 * Nw * K3 * K4^3 + 16 * Nx * Ny^3 * K3 * K4^3 * M +
        48 * Nx * Ny^2 * Nw * K3 * K4^3 * M + 16 * Ny^3 * Nw * K3 * K4^3 * M -
        24 * Nx * Ny^2 * K3 * K4^3 * M^2 - 8 * Ny^3 * K3 * K4^3 * M^2 -
        24 * Nx * Ny * Nw * K3 * K4^3 * M^2 - 24 * Ny^2 * Nw * K3 * K4^3 * M^2 +
        12 * Nx * Ny * K3 * K4^3 * M^3 + 12 * Ny^2 * K3 * K4^3 * M^3 + 4 * Nx * Nw * K3 * K4^3 * M^3 +
        12 * Ny * Nw * K3 * K4^3 * M^3 - 2 * Nx * K3 * K4^3 * M^4 - 6 * Ny * K3 * K4^3 * M^4 -
        2 * Nw * K3 * K4^3 * M^4 + K3 * K4^3 * M^5 - 32 * Nx * Ny^2 * Nw * K4^3 +
        16 * Nx * Ny^2 * K4^3 * M + 32 * Nx * Ny * Nw * K4^3 * M + 16 * Ny^2 * Nw * K4^3 * M -
        16 * Nx * Ny * K4^3 * M^2 - 8 * Ny^2 * K4^3 * M^2 - 8 * Nx * Nw * K4^3 * M^2 -
        16 * Ny * Nw * K4^3 * M^2 + 4 * Nx * K4^3 * M^3 + 8 * Ny * K4^3 * M^3 +
        4 * Nw * K4^3 * M^3 - 2 * K4^3 * M^4) / (4 * M^3)) * L^3 +
  ((16 * Nx * Ny^3 * K2 * K3^2 * K4^2 + 16 * Nx * Ny^2 * Nw * K2 * K3^2 * K4^2 +
        16 * Ny^3 * Nw * K2 * K3^2 * K4^2 - 32 * Nx * Ny^2 * K2 * K3^2 * K4^2 * M -
        16 * Ny^3 * K2 * K3^2 * K4^2 * M - 16 * Nx * Ny * Nw * K2 * K3^2 * K4^2 * M -
        32 * Ny^2 * Nw * K2 * K3^2 * K4^2 * M + 20 * Nx * Ny * K2 * K3^2 * K4^2 * M^2 +
        28 * Ny^2 * K2 * K3^2 * K4^2 * M^2 + 4 * Nx * Nw * K2 * K3^2 * K4^2 * M^2 +
        20 * Ny * Nw * K2 * K3^2 * K4^2 * M^2 - 4 * Nx * K2 * K3^2 * K4^2 * M^3 -
```

```
        16 * Ny * K2 * K3^2 * K4^2 * M^3 - 4 * Nw * K2 * K3^2 * K4^2 * M^3 +
        3 * K2 * K3^2 * K4^2 * M^4 + 32 * Nx * Ny^2 * K2 * K3 * K4^2 +
        32 * Nx * Ny * Nw * K2 * K3 * K4^2 + 32 * Ny^2 * Nw * K2 * K3 * K4^2 -
        48 * Nx * Ny * K2 * K3 * K4^2 * M - 32 * Ny^2 * K2 * K3 * K4^2 * M -
        16 * Nx * Nw * K2 * K3 * K4^2 * M - 48 * Ny * Nw * K2 * K3 * K4^2 * M +
        16 * Nx * K2 * K3 * K4^2 * M^2 + 40 * Ny * K2 * K3 * K4^2 * M^2 + 16 * Nw * K2 * K3 * K4^2 * M^2 -
        12 * K2 * K3 * K4^2 * M^3 + 16 * Nx * Ny * K2 * K4^2 + 16 * Ny * Nw * K2 * K4^2 +
        16 * Ny^2 * K3 * K4^2 - 8 * Nx * K2 * K4^2 * M - 16 * Ny * K2 * K4^2 * M -
        8 * Nw * K2 * K4^2 * M - 16 * Ny * K3 * K4^2 * M + 8 * K2 * K4^2 * M^2 +
        4 * K3 * K4^2 * M^2 + 16 * Ny * K4^2 - 8 * K4^2 * M) / (4 * K2 * K3 * M^2)) * L^2 +
   ((-8 * Nx * Ny^2 * K2 * K3^3 * K4 - 8 * Ny^3 * K2 * K3^3 * K4 - 8 * Ny^2 * Nw * K2 * K3^3 * K4 +
        8 * Nx * Ny * K2 * K3^3 * K4 * M + 20 * Ny^2 * K2 * K3^3 * K4 * M +
        8 * Ny * Nw * K2 * K3^3 * K4 * M - 2 * Nx * K2 * K3^3 * K4 * M^2 - 14 * Ny * K2 * K3^3 * K4 * M^2 -
        2 * Nw * K2 * K3^3 * K4 * M^2 + 3 * K2 * K3^3 * K4 * M^3 - 16 * Nx * Ny * K2 * K3^2 * K4 -
        24 * Ny^2 * K2 * K3^2 * K4 - 8 * Nx * Nw * K2 * K3^2 * K4 - 16 * Ny * Nw * K2 * K3^2 * K4 +
        12 * Nx * K2 * K3^2 * K4 * M + 40 * Ny * K2 * K3^2 * K4 * M + 12 * Nw * K2 * K3^2 * K4 * M -
        16 * K2 * K3^2 * K4 * M^2 - 8 * Nx * K2 * K3 * K4 - 24 * Ny * K2 * K3 * K4 -
        8 * Nw * K2 * K3 * K4 + 20 * K2 * K3 * K4 * M - 8 * K2 * K4 - 8 * K3 * K4) / (4 * K2 * K3^2 * M)) * L +
   ((4 * Ny^2 * K3^2 - 4 * Ny * K3^2 * M + K3^2 * M^2 + 4 * Nx * K3 + 8 * Ny * K3 - 6 * K3 * M + 4) / (4 * K3))
P3[z_] := z^2 + ((Nx * K1 - Nz * K1 + 1) / (K1)) * z + (-Nz) / (K1)
P4[y_, L_, z_] :=
 y + ((-16 * Nx * Ny^2 * Nw * K4^3 + 8 * Nx * Ny^2 * K4^3 * M + 16 * Nx * Ny * Nw * K4^3 * M +
        8 * Ny^2 * Nw * K4^3 * M - 8 * Nx * Ny * K4^3 * M^2 - 4 * Ny^2 * K4^3 * M^2 -
        4 * Nx * Nw * K4^3 * M^2 - 8 * Ny * Nw * K4^3 * M^2 + 2 * Nx * K4^3 * M^3 +
        4 * Ny * K4^3 * M^3 + 2 * Nw * K4^3 * M^3 - K4^3 * M^4) / (2 * M^3)) * L^3 +
   ((4 * Nx * Ny^2 * K2 * K3 * K4^2 + 8 * Nx * Ny * Nw * K2 * K3 * K4^2 + 4 * Ny^2 * Nw * K2 * K3 * K4^2 -
        8 * Nx * Ny * K2 * K3 * K4^2 * M - 4 * Ny^2 * K2 * K3 * K4^2 * M - 4 * Nx * Nw * K2 * K3 * K4^2 * M -
        8 * Ny * Nw * K2 * K3 * K4^2 * M + 3 * Nx * K2 * K3 * K4^2 * M^2 + 6 * Ny * K2 * K3 * K4^2 * M^2 +
        3 * Nw * K2 * K3 * K4^2 * M^2 - 2 * K2 * K3 * K4^2 * M^3 + 4 * Nx * Ny * K2 * K4^2 +
        4 * Ny * Nw * K2 * K4^2 - 2 * Nx * K2 * K4^2 * M - 4 * Ny * K2 * K4^2 * M - 2 * Nw * K2 * K4^2 * M +
        2 * K2 * K4^2 * M^2 + 4 * Ny * K4^2 - 2 * K4^2 * M) / (K2 * K3 * M^2)) * L^2 +
   ((-4 * Nx * Ny * K2 * K3^2 * K4 - 2 * Ny^2 * K2 * K3^2 * K4 - 2 * Nx * Nw * K2 * K3^2 * K4 -
        4 * Ny * Nw * K2 * K3^2 * K4 + 3 * Nx * K2 * K3^2 * K4 * M + 6 * Ny * K2 * K3^2 * K4 * M + 3 * Nw *
          K2 * K3^2 * K4 * M - 3 * K2 * K3^2 * K4 * M^2 - 2 * Nx * K2 * K3 * K4 - 4 * Ny * K2 * K3 * K4 -
        2 * Nw * K2 * K3 * K4 + 4 * K2 * K3 * K4 * M - 2 * K2 * K4 - 2 * K3 * K4) / (K2 * K3^2 * M)) * L +
   ((2 * Nx * K3 + 2 * Ny * K3 + 2 * Nw * K3 - 3 * K3 * M + 4) / (2 * K3))
P5[x_, L_, z_] := x + ((-32 * Nx * Ny^3 * Nw * K3 * K4^3 + 16 * Nx * Ny^3 * K3 * K4^3 * M +
        48 * Nx * Ny^2 * Nw * K3 * K4^3 * M + 16 * Ny^3 * Nw * K3 * K4^3 * M -
        24 * Nx * Ny^2 * K3 * K4^3 * M^2 - 8 * Ny^3 * K3 * K4^3 * M^2 -
        24 * Nx * Ny * Nw * K3 * K4^3 * M^2 - 24 * Ny^2 * Nw * K3 * K4^3 * M^2 +
        128 Nx * Ny * K3 * K4^3 * M^3 + 12 * Ny^2 * K3 * K4^3 * M^3 + 4 * Nx * Nw * K3 * K4^3 * M^3 +
        12 * Ny * Nw * K3 * K4^3 * M^3 - 2 * Nx * K3 * K4^3 * M^4 - 6 * Ny * K3 * K4^3 * M^4 -
        2 * Nw * K3 * K4^3 * M^4 + K3 * K4^3 * M^5 - 32 * Nx * Ny^2 * Nw * K4^3 +
```

D. GRÖBNER BASIS FOR THE COMPUTATION OF THE EC_{50} OF THE TRIMERIC RECEPTOR MODEL


```
      16 * Nx * Ny ^ 2 * K4 ^ 3 * M + 32 * Nx * Ny * Nw * K4 ^ 3 * M + 16 * Ny ^ 2 * Nw * K4 ^ 3 * M -
      16 * Nx * Ny * K4 ^ 3 * M ^ 2 - 8 * Ny ^ 2 * K4 ^ 3 * M ^ 2 - 8 * Nx * Nw * K4 ^ 3 * M ^ 2 -
      16 * Ny * Nw * K4 ^ 3 * M ^ 2 + 4 * Nx * K4 ^ 3 * M ^ 3 + 8 * Ny * K4 ^ 3 * M ^ 3 +
      4 * Nw * K4 ^ 3 * M ^ 3 - 2 * K4 ^ 3 * M ^ 4) / (4 * Nx * M ^ 3)) * z * L ^ 3 +
((16 * Nx * Ny ^ 3 * K2 * K3 ^ 2 * K4 ^ 2 + 16 * Nx * Ny ^ 2 * Nw * K2 * K3 ^ 2 * K4 ^ 2 +
      16 * Ny ^ 3 * Nw * K2 * K3 ^ 2 * K4 ^ 2 - 32 * Nx * Ny ^ 2 * K2 * K3 ^ 2 * K4 ^ 2 * M -
      16 * Ny ^ 3 * K2 * K3 ^ 2 * K4 ^ 2 * M - 16 * Nx * Ny * Nw * K2 * K3 ^ 2 * K4 ^ 2 * M -
      32 * Ny ^ 2 * Nw * K2 * K3 ^ 2 * K4 ^ 2 * M + 20 * Nx * Ny * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 2 +
      28 * Ny ^ 2 * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 2 + 4 * Nx * Nw * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 2 +
      20 * Ny * Nw * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 2 - 4 * Nx * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 3 -
      16 * Ny * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 3 - 4 * Nw * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 3 +
      3 * K2 * K3 ^ 2 * K4 ^ 2 * M ^ 4 + 32 * Nx * Ny ^ 2 * K2 * K3 * K4 ^ 2 +
      32 * Nx * Ny * Nw * K2 * K3 * K4 ^ 2 + 32 * Ny ^ 2 * Nw * K2 * K3 * K4 ^ 2 -
      48 * Nx * Ny * K2 * K3 * K4 ^ 2 * M - 32 * Ny ^ 2 * K2 * K3 * K4 ^ 2 * M -
      16 * Nx * Nw * K2 * K3 * K4 ^ 2 * M - 48 * Ny * Nw * K2 * K3 * K4 ^ 2 * M +
      16 * Nx * K2 * K3 * K4 ^ 2 * M ^ 2 + 40 * Ny * K2 * K3 * K4 ^ 2 * M ^ 2 + 16 * Nw * K2 * K3 * K4 ^ 2 * M ^ 2 -
      12 * K2 * K3 * K4 ^ 2 * M ^ 3 + 16 * Nx * Ny * K2 * K4 ^ 2 + 16 * Ny * Nw * K2 * K4 ^ 2 +
      16 * Ny ^ 2 * K3 * K4 ^ 2 - 8 * Nx * K2 * K4 ^ 2 * M - 16 * Ny * K2 * K4 ^ 2 * M -
      8 * Nw * K2 * K4 ^ 2 * M - 16 * Ny * K3 * K4 ^ 2 * M + 8 * K2 * K4 ^ 2 * M ^ 2 +
      4 * K3 * K4 ^ 2 * M ^ 2 + 16 * Ny * K4 ^ 2 - 8 * K4 ^ 2 * M) / (4 * Nx * K2 * K3 * M ^ 2)) * z * L ^ 2 +
((-8 * Nx * Ny ^ 2 * K2 * K3 ^ 3 * K4 - 8 * Ny ^ 3 * K2 * K3 ^ 3 * K4 - 8 * Ny ^ 2 * Nw * K2 * K3 ^ 3 * K4 +
      8 * Nx * Ny * K2 * K3 ^ 3 * K4 * M + 20 * Ny ^ 2 * K2 * K3 ^ 3 * K4 * M + 8 * Ny * Nw * K2 * K3 ^ 3 * K4 * M -
      2 * Nx * K2 * K3 ^ 3 * K4 * M ^ 2 - 14 * Ny * K2 * K3 ^ 3 * K4 * M ^ 2 - 2 * Nw * K2 * K3 ^ 3 * K4 * M ^ 2 +
      3 * K2 * K3 ^ 3 * K4 * M ^ 3 - 16 * Nx * Ny * K2 * K3 ^ 2 * K4 - 24 * Ny ^ 2 * K2 * K3 ^ 2 * K4 -
      8 * Nx * Nw * K2 * K3 ^ 2 * K4 - 16 * Ny * Nw * K2 * K3 ^ 2 * K4 + 12 * Nx * K2 * K3 ^ 2 * K4 * M +
      40 * Ny * K2 * K3 ^ 2 * K4 * M + 12 * Nw * K2 * K3 ^ 2 * K4 * M - 16 * K2 * K3 ^ 2 * K4 * M ^ 2 -
      8 * Nx * K2 * K3 * K4 - 24 * Ny * K2 * K3 * K4 - 8 * Nw * K2 * K3 * K4 +
      20 * K2 * K3 * K4 * M - 8 * K2 * K4 - 8 * K3 * K4) / (4 * Nx * K2 * K3 ^ 2 * M)) * z * L +
((4 * Ny ^ 2 * K3 ^ 2 - 4 * Ny * K3 ^ 2 * M + K3 ^ 2 * M ^ 2 + 8 * Ny * K3 + 4 * Nw * K3 - 6 * K3 * M + 4) /
    (4 * Nx * K3)) * z +
((-32 * Nx ^ 2 * Ny ^ 3 * Nw * K3 * K4 ^ 3 + 32 * Nx * Ny ^ 3 * Nw * Nz * K3 * K4 ^ 3 +
      16 * Nx ^ 2 * Ny ^ 3 * K3 * K4 ^ 3 * M + 48 * Nx ^ 2 * Ny ^ 2 * Nw * K3 * K4 ^ 3 * M +
      16 * Nx * Ny ^ 3 * Nw * K3 * K4 ^ 3 * M - 16 * Nx * Ny ^ 3 * Nz * K3 * K4 ^ 3 * M -
      48 * Nx * Ny ^ 2 * Nw * Nz * K3 * K4 ^ 3 * M - 16 * Ny ^ 3 * Nw * Nz * K3 * K4 ^ 3 * M -
      24 * Nx ^ 2 * Ny ^ 2 * K3 * K4 ^ 3 * M ^ 2 - 8 * Nx * Ny ^ 3 * K3 * K4 ^ 3 * M ^ 2 -
      24 * Nx ^ 2 * Ny * Nw * K3 * K4 ^ 3 * M ^ 2 - 24 * Nx * Ny ^ 2 * Nw * K3 * K4 ^ 3 * M ^ 2 +
      24 * Nx * Ny ^ 2 * Nz * K3 * K4 ^ 3 * M ^ 2 + 8 * Ny ^ 3 * Nz * K3 * K4 ^ 3 * M ^ 2 +
      24 * Nx * Ny * Nw * Nz * K3 * K4 ^ 3 * M ^ 2 + 24 * Ny ^ 2 * Nw * Nz * K3 * K4 ^ 3 * M ^ 2 +
      12 * Nx ^ 2 * Ny * K3 * K4 ^ 3 * M ^ 3 + 12 * Nx * Ny ^ 2 * K3 * K4 ^ 3 * M ^ 3 +
      4 * Nx ^ 2 * Nw * K3 * K4 ^ 3 * M ^ 3 + 12 * Nx * Ny * Nw * K3 * K4 ^ 3 * M ^ 3 -
      12 * Nx * Ny * Nz * K3 * K4 ^ 3 * M ^ 3 - 12 * Ny ^ 2 * Nz * K3 * K4 ^ 3 * M ^ 3 -
      4 * Nx * Nw * Nz * K3 * K4 ^ 3 * M ^ 3 - 12 * Ny * Nw * Nz * K3 * K4 ^ 3 * M ^ 3 -
      2 * Nx ^ 2 * K3 * K4 ^ 3 * M ^ 4 - 6 * Nx * Ny * K3 * K4 ^ 3 * M ^ 4 - 2 * Nx * Nw * K3 * K4 ^ 3 * M ^ 4 +
```

```
      2 * Nx * Nz * K3 * K4^3 * M^4 + 6 * Ny * Nz * K3 * K4^3 * M^4 + 2 * Nw * Nz * K3 * K4^3 * M^4 +
      Nx * K3 * K4^3 * M^5 - Nz * K3 * K4^3 * M^5 - 32 * Nx^2 * Ny^2 * Nw * K4^3 +
      32 * Nx * Ny^2 * Nw * Nz * K4^3 + 16 * Nx^2 * Ny^2 * K4^3 * M + 32 * Nx^2 * Ny * Nw * K4^3 * M +
      16 * Nx * Ny^2 * Nw * K4^3 * M - 16 * Nx * Ny^2 * Nz * K4^3 * M -
      32 * Nx * Ny * Nw * Nz * K4^3 * M - 16 * Ny^2 * Nw * Nz * K4^3 * M - 16 * Nx^2 * Ny * K4^3 * M^2 -
      8 * Nx * Ny^2 * K4^3 * M^2 - 8 * Nx^2 * Nw * K4^3 * M^2 - 16 * Nx * Ny * Nw * K4^3 * M^2 +
      16 * Nx * Ny * Nz * K4^3 * M^2 + 8 * Ny^2 * Nz * K4^3 * M^2 + 8 * Nx * Nw * Nz * K4^3 * M^2 +
      16 * Ny * Nw * Nz * K4^3 * M^2 + 4 * Nx^2 * K4^3 * M^3 + 8 * Nx * Ny * K4^3 * M^3 +
      4 * Nx * Nw * K4^3 * M^3 - 4 * Nx * Nz * K4^3 * M^3 - 8 * Ny * Nz * K4^3 * M^3 -
      4 * Nw * Nz * K4^3 * M^3 - 2 * Nx * K4^3 * M^4 + 2 * Nz * K4^3 * M^4) / (4 * Nx * M^3)) * L^3 +
  ((16 * Nx^2 * Ny^3 * K2 * K3^2 * K4^2 + 16 * Nx^2 * Ny^2 * Nw * K2 * K3^2 * K4^2 +
      16 * Nx * Ny^3 * Nw * K2 * K3^2 * K4^2 - 16 * Nx * Ny^3 * Nz * K2 * K3^2 * K4^2 -
      16 * Nx * Ny^2 * Nw * Nz * K2 * K3^2 * K4^2 - 16 * Ny^3 * Nw * Nz * K2 * K3^2 * K4^2 -
      32 * Nx^2 * Ny^2 * K2 * K3^2 * K4^2 * M - 16 * Nx * Ny^3 * K2 * K3^2 * K4^2 * M -
      16 * Nx^2 * Ny * Nw * K2 * K3^2 * K4^2 * M - 32 * Nx * Ny^2 * Nw * K2 * K3^2 * K4^2 * M +
      32 * Nx * Ny^2 * Nz * K2 * K3^2 * K4^2 * M + 16 * Ny^3 * Nz * K2 * K3^2 * K4^2 * M +
      16 * Nx * Ny * Nw * Nz * K2 * K3^2 * K4^2 * M + 32 * Ny^2 * Nw * Nz * K2 * K3^2 * K4^2 * M +
      20 * Nx^2 * Ny * K2 * K3^2 * K4^2 * M^2 + 28 * Nx * Ny^2 * K2 * K3^2 * K4^2 * M^2 +
      4 * Nx^2 * Nw * K2 * K3^2 * K4^2 * M^2 + 20 * Nx * Ny * Nw * K2 * K3^2 * K4^2 * M^2 -
      20 * Nx * Ny * Nz * K2 * K3^2 * K4^2 * M^2 - 28 * Ny^2 * Nz * K2 * K3^2 * K4^2 * M^2 -
      4 * Nx * Nw * Nz * K2 * K3^2 * K4^2 * M^2 - 20 * Ny * Nw * Nz * K2 * K3^2 * K4^2 * M^2 -
      4 * Nx^2 * K2 * K3^2 * K4^2 * M^3 - 16 * Nx * Ny * K2 * K3^2 * K4^2 * M^3 -
      4 * Nx * Nw * K2 * K3^2 * K4^2 * M^3 + 4 * Nx * Nz * K2 * K3^2 * K4^2 * M^3 +
      16 * Ny * Nz * K2 * K3^2 * K4^2 * M^3 + 4 * Nw * Nz * K2 * K3^2 * K4^2 * M^3 +
      3 * Nx * K2 * K3^2 * K4^2 * M^4 - 3 * Nz * K2 * K3^2 * K4^2 * M^4 +
      32 * Nx^2 * Ny^2 * K2 * K3 * K4^2 + 32 * Nx^2 * Ny * Nw * K2 * K3 * K4^2 +
      32 * Nx * Ny^2 * Nw * K2 * K3 * K4^2 - 32 * Nx * Ny^2 * Nz * K2 * K3 * K4^2 -
      32 * Nx * Ny * Nw * Nz * K2 * K3 * K4^2 - 32 * Ny^2 * Nw * Nz * K2 * K3 * K4^2 -
      48 * Nx^2 * Ny * K2 * K3 * K4^2 * M - 32 * Nx * Ny^2 * K2 * K3 * K4^2 * M -
      16 * Nx^2 * Nw * K2 * K3 * K4^2 * M - 48 * Nx * Ny * Nw * K2 * K3 * K4^2 * M +
      48 * Nx * Ny * Nz * K2 * K3 * K4^2 * M + 32 * Ny^2 * Nz * K2 * K3 * K4^2 * M +
      16 * Nx * Nw * Nz * K2 * K3 * K4^2 * M + 48 * Ny * Nw * Nz * K2 * K3 * K4^2 * M +
      16 * Nx^2 * K2 * K3 * K4^2 * M^2 + 40 * Nx * Ny * K2 * K3 * K4^2 * M^2 +
      16 * Nx * Nw * K2 * K3 * K4^2 * M^2 - 16 * Nx * Nz * K2 * K3 * K4^2 * M^2 -
      40 * Ny * Nz * K2 * K3 * K4^2 * M^2 - 16 * Nw * Nz * K2 * K3 * K4^2 * M^2 -
      12 * Nx * K2 * K3 * K4^2 * M^3 + 12 * Nz * K2 * K3 * K4^2 * M^3 + 16 * Nx^2 * Ny * K2 * K4^2 +
      16 * Nx * Ny * Nw * K2 * K4^2 - 16 * Nx * Ny * Nz * K2 * K4^2 - 16 * Ny * Nw * Nz * K2 * K4^2 +
      16 * Nx * Ny^2 * K3 * K4^2 - 16 * Ny^2 * Nz * K3 * K4^2 - 8 * Nx^2 * K2 * K4^2 * M -
      16 * Nx * Ny * K2 * K4^2 * M - 8 * Nx * Nw * K2 * K4^2 * M + 8 * Nx * Nz * K2 * K4^2 * M +
      16 * Ny * Nz * K2 * K4^2 * M + 8 * Nw * Nz * K2 * K4^2 * M - 16 * Nx * Ny * K3 * K4^2 * M +
      16 * Ny * Nz * K3 * K4^2 * M + 8 * Nx * K2 * K4^2 * M^2 - 8 * Nz * K2 * K4^2 * M^2 +
      4 * Nx * K3 * K4^2 * M^2 - 4 * Nz * K3 * K4^2 * M^2 + 16 * Nx * Ny * K4^2 -
      16 * Ny * Nz * K4^2 - 8 * Nx * K4^2 * M + 8 * Nz * K4^2 * M) / (4 * Nx * K2 * K3 * M^2)) * L^2 +
```

D. GRÖBNER BASIS FOR THE COMPUTATION OF THE EC_{50} OF THE TRIMERIC RECEPTOR MODEL


```
((-8*Nx^2*Ny^2*K2*K3^3*K4 - 8*Nx*Ny^3*K2*K3^3*K4 -
      8*Nx*Ny^2*Nw*K2*K3^3*K4 + 8*Nx*Ny^2*Nz*K2*K3^3*K4 +
      8*Ny^3*Nz*K2*K3^3*K4 + 8*Ny^2*Nw*Nz*K2*K3^3*K4 +
      8*Nx^2*Ny*K2*K3^3*K4*M + 20*Nx*Ny^2*K2*K3^3*K4*M +
      8*Nx*Ny*Nw*K2*K3^3*K4*M - 8*Nx*Ny*Nz*K2*K3^3*K4*M -
      20*Ny^2*Nz*K2*K3^3*K4*M - 8*Ny*Nw*Nz*K2*K3^3*K4*M -
      2*Nx^2*K2*K3^3*K4*M^2 - 14*Nx*Ny*K2*K3^3*K4*M^2 -
      2*Nx*Nw*K2*K3^3*K4*M^2 + 2*Nx*Nz*K2*K3^3*K4*M^2 +
      14*Ny*Nz*K2*K3^3*K4*M^2 + 2*Nw*Nz*K2*K3^3*K4*M^2 +
      3*Nx*K2*K3^3*K4*M^3 - 3*Nz*K2*K3^3*K4*M^3 - 16*Nx^2*Ny*K2*K3^2*K4 -
      24*Nx*Ny^2*K2*K3^2*K4 - 8*Nx^2*Nw*K2*K3^2*K4 -
      16*Nx*Ny*Nw*K2*K3^2*K4 + 16*Nx*Ny*Nz*K2*K3^2*K4 +
      24*Ny^2*Nz*K2*K3^2*K4 + 8*Nx*Nw*Nz*K2*K3^2*K4 +
      16*Ny*Nw*Nz*K2*K3^2*K4 + 12*Nx^2*K2*K3^2*K4*M +
      40*Nx*Ny*K2*K3^2*K4*M + 12*Nx*Nw*K2*K3^2*K4*M -
      12*Nx*Nz*K2*K3^2*K4*M - 40*Ny*Nz*K2*K3^2*K4*M -
      12*Nw*Nz*K2*K3^2*K4*M - 16*Nx*K2*K3^2*K4*M^2 +
      16*Nz*K2*K3^2*K4*M^2 - 8*Nx^2*K2*K3*K4 - 24*Nx*Ny*K2*K3*K4 -
      8*Nx*Nw*K2*K3*K4 + 8*Nx*Nz*K2*K3*K4 + 24*Ny*Nz*K2*K3*K4 +
      8*Nw*Nz*K2*K3*K4 + 20*Nx*K2*K3*K4*M - 20*Nz*K2*K3*K4*M - 8*Nx*K2*K4 +
      8*Nz*K2*K4 - 8*Nx*K3*K4 + 8*Nz*K3*K4)/(4*Nx*K2*K3^2*M))*L +
 ((4*Nx*Ny^2*K3^2 - 4*Ny^2*Nz*K3^2 - 4*Nx*Ny*K3^2*M + 4*Ny*Nz*K3^2*M +
      Nx*K3^2*M^2 - Nz*K3^2*M^2 + 8*Nx*Ny*K3 + 4*Nx*Nw*K3 - 8*Ny*Nz*K3 -
      4*Nw*Nz*K3 - 6*Nx*K3*M + 6*Nz*K3*M + 4*Nx - 4*Nz)/(4*Nx*K3))
```

```
FullSimplify[CoefficientList[P1[L], L]]
```

Out[ ]=
$$\left\{\frac{M^4}{K4^4\,(M-2\,Nw)\,(M-2\,Nx)\,(M-2\,Ny)^2},\ \frac{2\,M^3\,(-2+K3\,(2\,M-Nw-Nx-2\,Ny))}{K3\,K4^3\,(M-2\,Nw)\,(M-2\,Nx)\,(M-2\,Ny)^2},\right.$$
$$\left(2\,M^2\,\left(2\,K3+K2\,\left(2+2\,K3\,\left(-2\,M+Nw+Nx+2\,Ny\right)+K3^2\,\left(3\,M^2-3\,M\,\left(Nw+Nx+2\,Ny\right)+\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\left.2\,\left(Nw\,Nx+2\,\left(Nw+Nx\right)\,Ny+Ny^2\right)\right)\right)\right)\right)\Big/\left(K2\,K3^2\,K4^2\,\left(M-2\,Nw\right)\,\left(M-2\,Nx\right)\,\left(M-2\,Ny\right)^2\right),$$
$$\frac{2\,M\,\left(2+K2\,\left(2\,K3\,M^2-M\,\left(2+3\,K3\,\left(Nw+Nx\right)+2\,K3\,Ny\right)+2\,\left(Nw+Nx+2\,K3\,Nw\,Nx+K3\,\left(Nw+Nx\right)\,Ny\right)\right)\right)}{K2\,K3\,K4\,\left(M^3-8\,Nw\,Nx\,Ny-2\,M^2\,\left(Nw+Nx+8\,Ny\right)+4\,M\,\left(Nx\,Ny+Nw\,\left(Nx+Ny\right)\right)\right)},$$
$$\left.1\right\}$$

Appendix E

Mathematica notebook for the computation of the amplitude of homodimeric RTK model A+B

Here, I provide the Mathematica (Wolfram Research, Inc., 2019) notebook that I tried to use to compute the amplitude of homodimeric RTK model A+B. The affinity constant K_3' is written $K3p$.

E. MATHEMATICA NOTEBOOK FOR THE COMPUTATION OF THE AMPLITUDE OF HOMODIMERIC RTK MODEL A+B

```
In[1]:= x[L_] := (-1 - K3p * L + Sqrt[8 * Nx * K2 * (1 + K3 * L) + (1 + K3p * L) ^ 2]) / (4 * K2 * (K3 * L + 1))

In[ ]:= sig[L_] := K2 * K3 * L * x[L] * x[L]

In[ ]:= FullSimplify[D[sig[L], L]]
```

$$\text{Out}[\ ]= -\frac{K3^2\, L \left(1 + K3p\, L - \sqrt{(1 + K3p\, L)^2 + 8\, K2\, (1 + K3\, L)\, Nx}\right)^2}{8\, K2\, (1 + K3\, L)^3} + \frac{K3 \left(1 + K3p\, L - \sqrt{(1 + K3p\, L)^2 + 8\, K2\, (1 + K3\, L)\, Nx}\right)^2}{16\, K2\, (1 + K3\, L)^2} + \frac{K3\, L \left(-K3p + \frac{K3p + K3p^2\, L + 4\, K2\, K3\, Nx}{\sqrt{(1 + K3p\, L)^2 + 8\, K2\, (1 + K3\, L)\, Nx}}\right) \left(-1 - K3p\, L + \sqrt{(1 + K3p\, L)^2 + 8\, K2\, (1 + K3\, L)\, Nx}\right)}{8\, K2\, (1 + K3\, L)^2}$$

In[ ]:=
```
Solve[D[sig[L], L] == 0, L]
```

Out[ ]=

$$\left\{\left\{L \to -\frac{-2\,K3\,K3p+3\,K3p^2}{3\,K3\,K3p^2}-\left(2^{1/3}\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)\right)\Big/\left(3\,K3\,K3p^2\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right)+\frac{1}{3\times 2^{1/3}\,K3\,K3p^2}\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right\},$$

$$\left\{L \to -\frac{-2\,K3\,K3p+3\,K3p^2}{3\,K3\,K3p^2}+\left(\left(1+i\sqrt{3}\right)\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)\right)\Big/\left(3\times 2^{2/3}\,K3\,K3p^2\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right)-\frac{1}{6\times 2^{1/3}\,K3\,K3p^2}\left(1-i\sqrt{3}\right)\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right\},$$

$$\left\{L \to -\frac{-2\,K3\,K3p+3\,K3p^2}{3\,K3\,K3p^2}+\left(\left(1-i\sqrt{3}\right)\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)\right)\Big/\left(3\times 2^{2/3}\,K3\,K3p^2\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right)-\frac{1}{6\times 2^{1/3}\,K3\,K3p^2}\left(1+i\sqrt{3}\right)\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx+\sqrt{\left(4\left(-K3^2\,K3p^2+6\,K3\,K3p^3-9\,K3p^4\right)^3+\left(-2\,K3^3\,K3p^3+18\,K3^2\,K3p^4+54\,K3\,K3p^5-54\,K3p^6+216\,K2\,K3^2\,K3p^4\,Nx\right)^2\right)}\right)^{1/3}\right\}\right\}$$

E. MATHEMATICA NOTEBOOK FOR THE COMPUTATION OF THE AMPLITUDE OF HOMODIMERIC RTK MODEL A+B

Appendix F

Mathematica notebook for the computation of the amplitude of heterodimeric RTK model B+C

The signalling function of the heterodimeric RTK model B+C is $\sigma(L) = N_y - (1 + K_3'' L)y$ where y is the steady state concentration of unbound α chains. Thus, $\frac{d\sigma}{dL}(L) = K_3''(y + L\frac{dy}{dL})$. Here, I provide the analytic expression of $\frac{dy}{dL}$ as computed thanks to Mathematica. This expression is rather long, even in an supposedly simplified form. Mathematica cannot find L such that $\frac{d\sigma}{dL}(L) = 0$.

F. MATHEMATICA NOTEBOOK FOR THE COMPUTATION OF THE AMPLITUDE OF HETERODIMERIC RTK MODEL B+C

```
In[1]:= y[L_] := -((1 + K3p L + K3pp L (1 + K3p L + K2pp (Nx - Ny)) -
          Sqrt((1 + L (K3p + K3pp + K3p K3pp L + K2pp K3pp (Nx - Ny)))^2 +
            4 K2pp K3pp L (1 + K3p L) (1 + K3pp L) Ny)) / (2 K2pp K3pp L (1 + K3pp L)))

In[3]:= FullSimplify[D[y[L], L]]

Out[3]= 1/(2 K2pp L^2 (1 + K3pp L)^2) (L (1 + K3p L + K3pp L (1 + K3p L + K2pp (Nx - Ny)) -
          Sqrt((1 + L (K3p + K3pp + K3p K3pp L + K2pp K3pp (Nx - Ny)))^2 +
            4 K2pp K3pp L (1 + K3p L) (1 + K3pp L) Ny)) +
        1/K3pp (1 + K3pp L) (1 + K3p L + K3pp L (1 + K3p L + K2pp (Nx - Ny)) -
          Sqrt((1 + L (K3p + K3pp + K3p K3pp L + K2pp K3pp (Nx - Ny)))^2 +
            4 K2pp K3pp L (1 + K3p L) (1 + K3pp L) Ny)) - 1/K3pp L (1 + K3pp L)
        (K3p + K3p K3pp L + K3pp (1 + K3p L + K2pp (Nx - Ny)) - (K3p + K3p^2 L (1 + K3pp L) (1 + 2 K3pp L) +
            K3p K3pp L (4 + 2 K2pp (Nx + Ny) + 3 K3pp L (1 + K2pp (Nx + Ny))) +
            K3pp (1 + K2pp (Nx + Ny) + K3pp L (1 + K2pp^2 (Nx - Ny)^2 + 2 K2pp (Nx + Ny)))) /
          (Sqrt((1 + L (K3p + K3pp + K3p K3pp L + K2pp K3pp (Nx - Ny)))^2 +
              4 K2pp K3pp L (1 + K3p L) (1 + K3pp L) Ny))))
```

Appendix G

Agent-based model with activation and death: switching from egalitarian regime to gerontocracy as m_0 decreases

Figures G.1 and G.2 show that the dynamics observed in the egalitarian regime is still observed in the case of $\bar{r_0} \approx 2\frac{up}{\mu^2}$ ($m_0 = 14, \sigma_0 = 1$). In that case, the receptor distribution over the whole simulation is also similar to the one in Figure 5.20 in the egalitarian case. The dynamics (with the usual parameter values) in the case of $m_0 = 11$ ($\bar{r_0} \approx \frac{up}{\mu^2}$) and $m_0 = 9$ ($\bar{r_0} \approx \frac{up}{75\mu^2}$) is shown in Figures G.3 and G.4. As m_0 decreases ($\sigma_0 = 1$ fixed), more and more r_T trajectories (when reported to their activation time) are crossing each other (even though the majority still follows the the mean of the expression obtained in equation (5.42)), and R becomes noisier.

G. AGENT-BASED MODEL WITH ACTIVATION AND DEATH: SWITCHING FROM EGALITARIAN REGIME TO GERONTOCRACY AS M_0 DECREASES

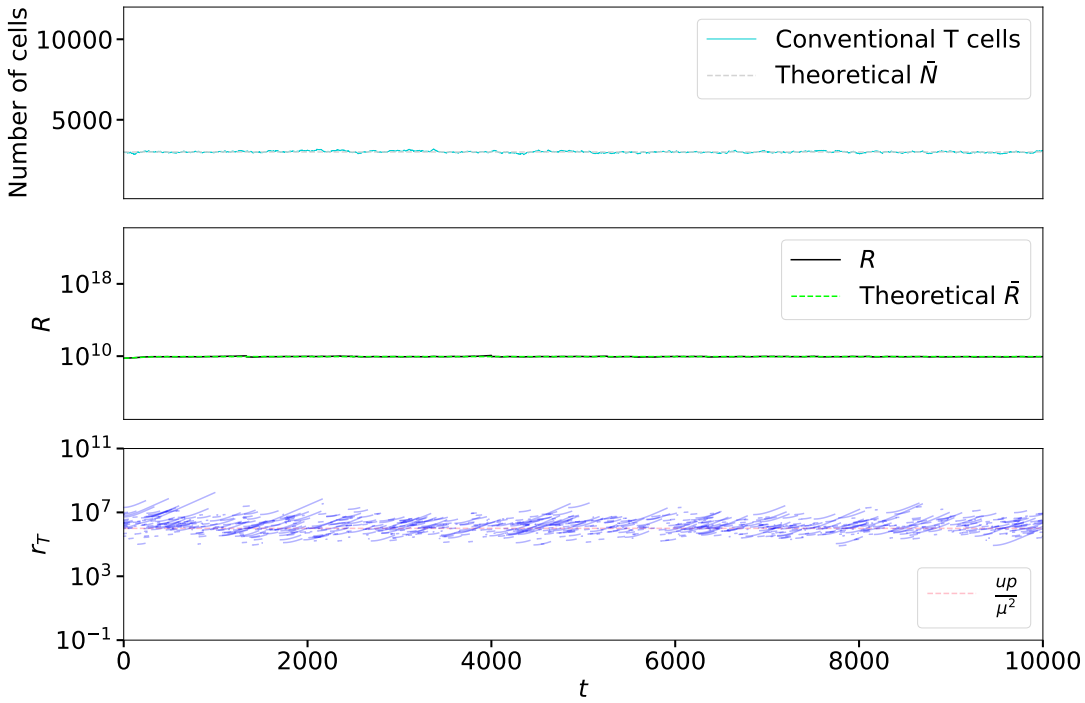


Figure G.1: Time evolution plots of the cell number, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells. The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$, $\mu = 0.01$ /day, $m_0 = 14$ and $\sigma_0 = 1$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.

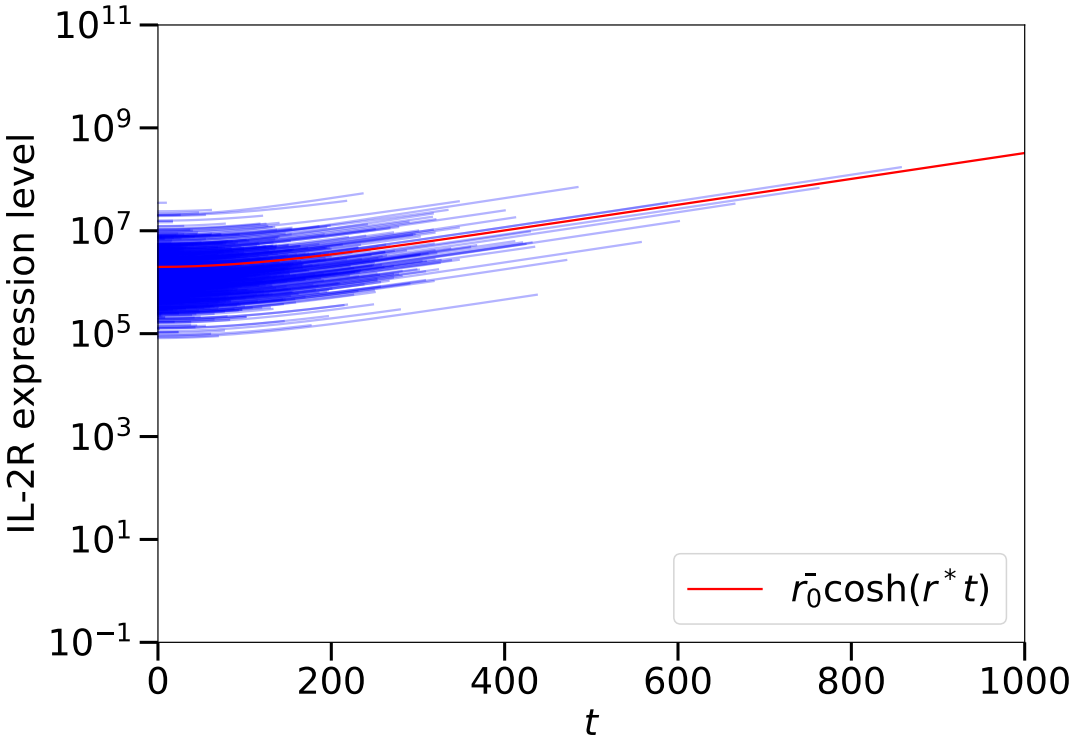
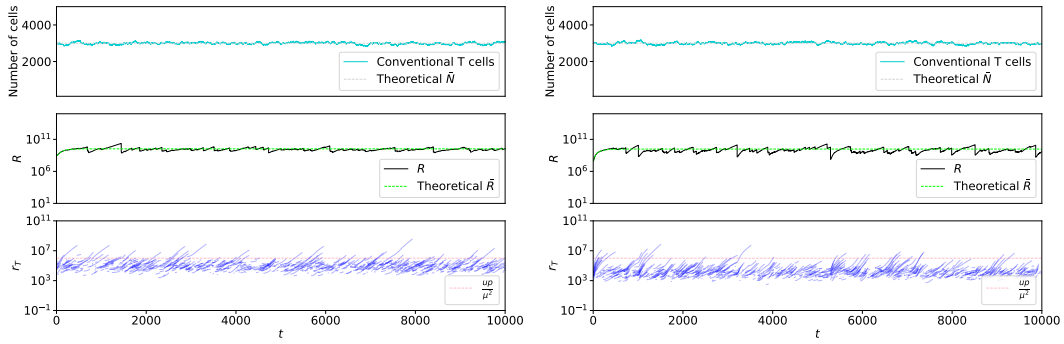


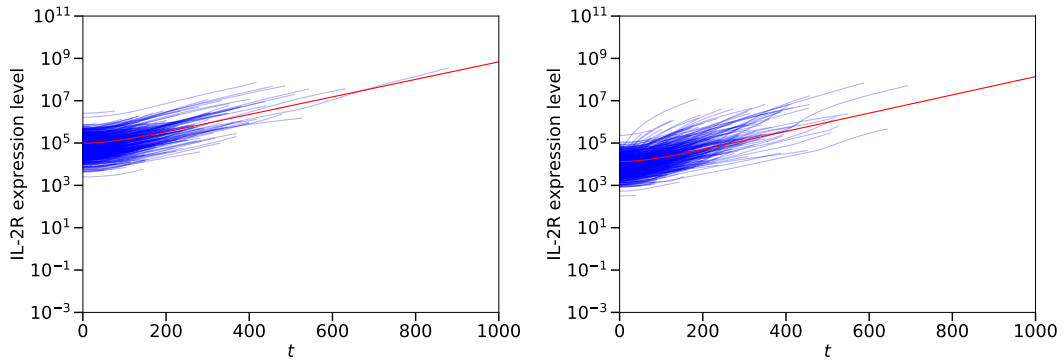
Figure G.2: Individual receptor upregulation trajectories from the simulation shown in Figure G.1, reported to the cell's activation time, t_T^{in}. No trajectories are crossing each other: they all follow the expression obtained in equation (5.42) (red line).

G. AGENT-BASED MODEL WITH ACTIVATION AND DEATH: SWITCHING FROM EGALITARIAN REGIME TO GERONTOCRACY AS M_0 DECREASES



(a) $m_0 = 11$ (b) $m_0 = 9$

Figure G.3: Time evolution plots of the cell number, $N(t)$, and its theoretical average expression, $\bar{N}(t)$ (top), population variable $R(t)$ alongside its theoretical average expression, $\bar{R}(t)$ (middle); and (bottom) IL-2R expression level of 0.3% randomly chosen cells. The parameter values used for this simulation are: $u = 10$ $[r_T]/[i_T]$/day, $p = 10$ $[i_T]$/day/cell, $\alpha = 30$ cells/day, $\mu = 0.01$ /day, and $\sigma_0 = 1$. The simulation started with $N(0) = \dfrac{\alpha}{\mu}$ cells and the time step was $\Delta t = 1$ day.



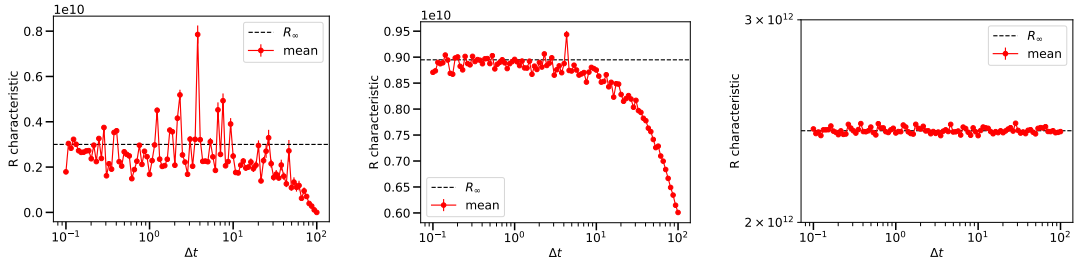
(a) $m_0 = 11$ (b) $m_0 = 9$

Figure G.4: Individual receptor upregulation trajectories from the simulation shown in Figures G.3, reported to the cell's activation time, t_T^{in}. As m_0 decreases (i,e. $\bar{r_0}$ decreases), more and more trajectories are crossing each other. Most of them still follow the mean of the expression obtained in equation (5.42), $\bar{r_0}\cosh(r^*t)$ (red line).

Appendix H

Agent-based model with activation and death: Δt analysis

To make sure we simulated the agent-based model (with death and activation only) with a time step, Δt, small enough, I computed, at steady state, the mean value (over one simulation) of $R(t)$ for simulations with different time steps (from 0.1 to 100) for the different regimes (small, moderate or large m_0). The result, displayed in Figure H.1 shows that from $\Delta t = 10$, the average value of $R(t)$ deviates from the theoretical mean steady state value R_∞. Thus, we can affirm that $\Delta_t = 1 < 10$ is a reasonable time step for the simulations.



(a) $m_0 = 1, \sigma_0 = 1$ (b) $m_0 = 14, \sigma_0 = 1$ (c) $m_0 = 20, \sigma_0 = 1$

Figure H.1: Each dot represents the mean of $R(t)$ throughout a simulation of the agent-based model with activation and death with the corresponding time step, Δt (in days), given on the x-axis. The other parameter values were: $\mu = 0.01$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu}$ and $t_{max} = 5000$ days. This mean value is to be compared with its theoretical value R_∞ (black dashed line).

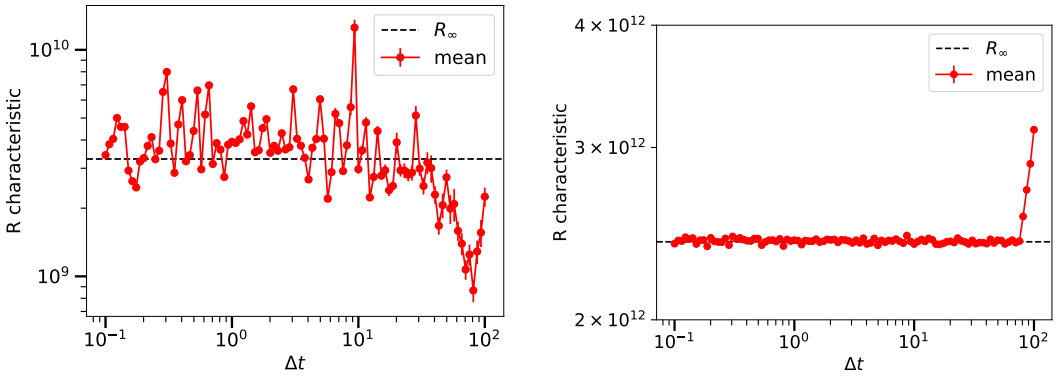
H. AGENT-BASED MODEL WITH ACTIVATION AND DEATH: ΔT ANALYSIS

Appendix I

Agent-based model with activation, death and division: Δt analysis

To make sure we simulated the agent-based model (with death, activation and division) with a time step, Δt, small enough, I computed, at steady state, the mean values of $R(t)$ for simulations with different time steps (from 0.1 to 100) for the two different regimes (m_0 small or large). The result, displayed in Figure I.1 shows that from $\Delta t = 30$, the average value of $R(t)$ deviates from the theoretical mean steady state value R_∞. Thus, we can affirm that $\Delta_t = 10 < 30$ is a reasonable time step for the simulations.

I. AGENT-BASED MODEL WITH ACTIVATION, DEATH AND DIVISION: ΔT ANALYSIS



(a) $m_0 = 1, \sigma_0 = 1$ (b) $m_0 = 20, \sigma_0 = 1$

Figure I.1: Each dot represents the mean of $R(t)$ throughout a simulation of the agent-based model with activation and death with the corresponding time step, Δt (in days), given on the x-axis. The other parameter values were: $\mu = 0.01$ /day, $\lambda = 0.003$ /day, $p = 10$ $[i_T]$/day/cell, $u = 10$ $[r_T]/[i_T]$/day, $\alpha = 30$ cells/day, $N(0) = \frac{\alpha}{\mu - \lambda}$ and $t_{max} = 5000$ days. This mean value is to be compared with its theoretical value R_∞ (black dashed line).

Appendix J

General agent-based model: Python code description

I provide a Python code to simulate the models of competition for IL-2 discussed in Chapter 5, comprising the extensions of Section 5.6. The code, available at https://github.com/leasta/ABM_thesis, does not distinguish the old and new cohorts.

J.1 Code structure and output

I define two classes of cells: conventional T cells, `Tconv`, and regulatory T cells, `Treg`. Both cell types have the same attributes but follow different rules.

J.1.1 Structure

To make the code readable, I split it into different files. The code is organised as follows:

- *params.py* defines the T cells parameters,
- *cells.py* defines the classes, `Tconv` and `Treg`, and their attributes,
- *functions.py* defines all the functions that rule the T cells behaviour (from activation and death to cytokine consumption and receptor upregulation),

- *globals.py* defines the populations of conventional and regulatory T cells as global variables (`profile_Tc` and `profile_Tr` respectively),
- *main.py* simulates the model from $t = 0$ to $t = t_{max}$ by updating, at each time step, the global profiles. It also creates the output plots.

In the following sections, I will detail the content of each file.

J.1.2 Output

The *main.py* routine creates a graph similar to Figure 5.13: it plots $N(t)$, $R(t)$ and a sample of $r_T(t)$ trajectories as a function of time. When the analytic expression for $\bar{N}(t)$ and $\bar{R}(t)$ is known, it adds it to the graph in dashed lines. If the parameter `animation` is set to `True`, the routine saves, at each time step, a scatterplot with joint distributions similar to Figure 5.11. One can then watch the animated simulation by opening any of these figures and press the right arrow on the keyboard to visualise the next ones [1].

J.2 File description

In this section, I detail the content of each file of the code. Especially, I introduce the parameters and describe the main functions.

J.2.1 Parameters

The parameters of the model are defined in the file *params.py*. Table J.1 recapitulates the notation of this code, which, most of the time, matches with the mathematical notation introduced in Chapter 5. Additionally, this file computes the value of $\bar{r_0} = e^{m_0 + \frac{\sigma_0^2}{2}}$ for the two cell populations (since m_0 and σ_0 can be different for both populations). All parameters (except `m0_T*`) need to be positive but the code won't stop the user from choosing negative parameter values. Finally,

[1] One could create an animated GIF from the set of all the scatterplots. However when merging more than a hundred figures, the resulting file is heavy and takes a long time to be created and to be opened. Thus, I chose to not create it and use the sequence of figures instead.

Python	Math	Description
`dt`	Δt	Time step
`tmax`	t_{max}	Final time of the simulation
`animation`		Boolean: save a scatterplot at each step if `True`
`N*`	N_*	Initial number of cells of the population
`m0_T*`	m_{0_*}	Parameter of the initial receptor distribution (mean of the normal distribution)
`sig0_T*`	σ_{0_*}	Parameter of the initial receptor distribution (standard deviation of the normal distribution)
`u_T*`	u_*	Receptor upregulation rate
`p_Tc`	p	Cytokine production rate
`mu_T*`	μ_*	Death rate
`a_T*`	α_*	Activation rate
`lb_T*`	λ_*	Division rate
`starv_T*`	ν	Starvation rate
`thrstarv_Tc`	θ_{starv}	Starvation threshold
`thrdiv_Tr`	θ_{div}	Division threshold
`c_T*`	c_*	Cytokine consumption rate
`track_T*`		Probability that a cell is tracked during the simulation

Table J.1: Notation for the parameters of the agent-based model of the competition for IL-2 between conventional and regulatory T cells. Conventional and regulatory T cells have *a priori* different parameter values. To this end, in the python code, I denote any quantity, `X`, (parameter, list or profile) related to conventional (resp. regulatory) T cells by `X_Tc` (resp. `X_Tr`). In this table, I chose to describe the parameters in the general case and the asterisk, `*`, replaces `c` or `r`. When we consider the competition within the conventional T cells only, `*` replaces `c` in the python notations and should be ignored in the mathematical notations.

the parameter `animation` is a boolean that must be set to `False` if one wants a code which runs faster, and does not want the scatterplots.

J.2.2 Cell classes

The file *cells.py* defines two similar cell classes: conventional (called `Tconv`) and regulatory T cells (called `Treg`). Both classes have two attributes: the IL-2R level expressed by the cell, `r`, and the amount of IL-2 it accumulated since it entered the cell pool, `i`. When created, a cell receives a certain number of receptors drawn from a log-normal distribution with parameters `m0_Tc` and `sig0_Tc` for conventional T cells, or `m0_Tr` and `sig0_Tr` for regulatory T cells. Both cell types receive no IL-2. The attributes `i` and `r` (i_T and r_T in Chapter 5) will be updated during the simulation, thanks to two update class functions `consume` and `IL2R_upregulation`. Additionally, each cell has a certain probability to be tracked during the simulation. If a cell is tracked, then its boolean attribute, `track`, will be set to `True` and the values taken by its attribute `r` at each time t will be stored in the attribute `rlist`. The time during which r is updated (from the time at which the cell enters the pool until it dies or divides (or the simulation ends)), is stored in attribute `timelist`. For any cell, the initial receptor value and the time at which it entered the cell pool are stored in `rlist` and `timelist`, respectively. Two functions related to this tracking are defined: `update_list` updates the attributes `rlist` and `timelist`. The function `reset_tracking` resets these lists when the cell divides. The two daughter cells will need to be selected again to be tracked. The initialisation of a conventional T cell is as follows (the initialisation of a regulatory T cell is similar):

```
class Tconv():
        def __init__(self,t):
                self.r = np.random.lognormal(m0_Tc,sig0_Tc)
                self.i=0
                self.rlist=[self.r]
                self.timelist=[t]
                ran=np.random.random()
                if ran<tracked_Tc:
                        self.track=True
                else:
```

```
self.track=False
```

J.2.3 Functions

The file *functions.py* contains four important functions.

IL-2 consumption and IL-2R upregulation

The function `IL2consumption_IL2production_IL2Rupregulation` computes the quantity of IL-2 available for the cells to consume at time t, calculates the populations variable $R(t)$ and updates the cell attributes `i` and `r` following the two-attributes dynamics described in section 5.2 (with a forward-Euler scheme).

Stochastic population events: Activation

The function `activation` determines the number (sampled from a Poisson distribution) of new cells entering the two cell pools during the time interval `dt` and updates the corresponding global profiles.

Stochastic population events: Death or division

The function `death_division_starvation` selects cells that will die, divide or do nothing during the small time interval $[t, t+\texttt{dt}]$.

First, we determine the cells that will undergo a stochastic event (death or division) at time t. To this end, we attribute to each cell a number drawn from a uniform distribution. If this number is lower than the product

$$\texttt{dt} \times (\text{division rate} + \text{death rate}),$$

then the cell will die or divide. Note that we first need to determine the actual death rate of each conventional T cell, which is $\mu_c + \nu$ if the cell did not absorb enough IL-2 (see starvation process described in Section 5.6), μ_c otherwise. Similarly, for regulatory T cells, the division rate is `lb_Tc` if it absorbed enough IL-2, 0 otherwise. Each of these cells will then randomly die or divide, in accordance with their death and division rate. If a cell dies, it will be removed from its global profile. If a cell, T, divides, a new cell of the same type (`Tconv` or `Treg`), T_d,

will be created. The receptor number `r` of both cells will then be set to half the value expressed by cell T when it divided, `T.r`. That is, attribute `rlist` of both cells will be re-initialised at `[T.r/2]`. Attribute `timelist` will be reset to `[t]` (division occurred at time t). The daughter cell, T_d, is added to the corresponding profile (`profile_Tc` if it is a `Tconv`, `profile_Tr` if it is a `Treg`).

Theoretical expression of the average ensembles

The function `theoretical_NR` returns the average ensemble values of the population variables $N(t)$ and $R(t)$ evaluated at each time-step of the simulation, when it was computed theoretically in Chapter 5. It returns vectors of negative values otherwise.

J.2.4 Initialisation of global variables

The initialisation function of cell profiles is defined in *globals.py*. It takes `Nc` and `Nr`, the initial number of conventional and regulatory T cells, as arguments and establishes two lists (as global variables), `profile_Tc` and `profile_Tr` containing `Nc` conventional T cells and `Nr` regulatory T cells, respectively. These cells were initialised as described in Section J.2.2

J.2.5 Main routine

The file, *main.py*, contains the main routine. First, it initialises the global profiles, `profile_Tc` and `profile_Tr`, by calling the function in *globals.py*. It also initialises the scatterplot for the eventual animation. Once this in place, it calls the function `animateAndSave`. This function first creates a folder (called `results_dir` in the following piece of code) in which all the figures will be saved. It also copies the file *params.py* in this folder for the record. Then, this function initialises a couple of lists where all the important variables will be stored for the time of the simulation. Finally, if `animation==True`, it saves the scatterplot at $t = 0$. With this in place, it updates the lists initialised for each time step as follows [1]:

[1]The arguments of the function `advance` have been replaced by `*args`. I have also replaced individual outputs of this function by the generic word `outputs`. Finally, I have

```
output_list=[]
timenow=0
while timenow<tmax:
        timenow+=dt
        outputs=advance(*args, timenow)
        output_list.append(outputs)
        if animation==True:
                animateGraph(timenow)
                file_name = "animation"+'%s'%timenow+'.png'
                plt.savefig(results_dir + file_name)
```

The function `advance` simulates the agent-based model for each time step, calling the functions defined in *functions.py*, `death_division_starvation`, `immigration` and `IL2consumption_IL2production_IL2Rupregulation` in this order. The function `animateGraph` updates the scatterplot initialised prior the update routine.

Once the simulation of the agent-based model for $t = 0, \ldots, t_{max}$ is over, a graph of the time evolution of the number of cells ($N_c(t)$, $N_r(t)$ and $N(t) = N_c(t)+N_r(t)$), the population variable $R(t)$ and individual r_T trajectories of `track_Tc` and `track_Tr` randomly chosen cells, is created.

J.3 Different sub-cases

Sub-models studied in Chapter 5 can be simulated with this code by setting the following parameter values: `Nr=0`, `a_Tr=0`. In addition, to obtain each sub-model in particular, the following parameter values must be chosen:

- Deterministic model: `mu_Tc=0`, `a_Tc=0`, `lb_Tc=0` and `starv_Tc=0`.
- Hybrid dynamics with death only: `a_Tc=0`, `lb_Tc=0` and `starv_Tc=0`.
- Hybrid dynamics with death and activation: `lb_Tc=0` and `starv_Tc=0`.
- Hybrid dynamics with death, activation and division: `starv_Tc=0`.

created a generic list `output_list` to store these general outputs. In the code, there is a separated list for each output.

For all these cases, theoretical expressions for $\bar{N}(t)$ and $\bar{R}(t)$ were computed and are added to the output graph. Note that in this case, as there is only one cell population, the parameter `c_Tc` does not affect the dynamics. However, it will rescale $R(t)$. To reproduce the figures of Chapter 5, one has to set `c_Tc=1`. We remind that the cohort discrimination is not included in this code version.

Appendix K

Altruistic model with degradation of free ligand

Here, we consider the altruistic model described in Section 5.7.1 and suppose that the extra-cellular ligand can be degraded with rate σ_i. The system of ODEs is as follows:

$$\frac{dI}{dt} = \phi + N(k_{\text{off}}C - k_{\text{on}}IR) - \sigma_i I, \tag{K.1a}$$

$$\frac{dR}{dt} = -k_{\text{on}}IR + k_{\text{off}}C - \sigma_r R + \frac{1}{1+\frac{S}{k_s}}\xi, \tag{K.1b}$$

$$\frac{dC}{dt} = k_{\text{on}}IR - k_{\text{off}}C - \sigma_c C, \tag{K.1c}$$

$$\frac{dS}{dt} = \psi C - \chi S. \tag{K.1d}$$

This system has a unique positive steady state (I^*, R^*, C^*, S^*).

Proof. Setting all the derivatives to zero, the equation (K.1c) gives :

$$k_{\text{on}}I^*R^* = (k_{\text{off}} + \sigma_c)C^*, \tag{K.2}$$

and equation (K.1d) gives

$$S^* = \frac{\psi C^*}{\chi}. \tag{K.3}$$

K. ALTRUISTIC MODEL WITH DEGRADATION OF FREE LIGAND

Substituting (K.2) in equation (K.1a), we obtain:

$$I^* = \frac{\phi - \sigma_c N C^*}{\sigma_i}. \tag{K.4}$$

This equations shows that we need to suppose $\frac{\phi}{\sigma_c N} > C^*$ to obtain a positive steady state. Substituting the expression for I^* in equation (K.2) gives:

$$R^* = \frac{(k_{\text{off}} + \sigma_c) C^* \sigma_i}{k_{\text{on}}(\phi - \sigma_c N C^*)} \tag{K.5}$$

Finally, substituting the expressions obtained for I^*, R^* and S^* in (K.1b) (with the derivative set to 0), we obtain:

$$-\sigma_c C^* - \frac{(k_{\text{off}} + \sigma_c) C^* \sigma_i \sigma_r}{k_{\text{on}}(\phi - \sigma_c N C^*)} + \frac{k_s \chi \xi}{k_s \chi + \psi C^*} = 0. \tag{K.6}$$

Since we suppose $\frac{\phi}{\sigma_c N_c} > C^* > 0$, this equation can be simplified to obtain a polynomial of degree 3 in C^*. Thus, C^* is a root of the following polynomial:

$$AX^3 + BX^2 + CX + D = 0 \tag{K.7}$$

where

$$\begin{aligned}
A =& k_{\text{on}} N \psi \sigma_c^2 \\
B =& -\phi k_{\text{on}} \sigma_c \psi + \sigma_c^2 N k_{\text{on}} k_s \chi - \sigma_i \sigma_r \psi (k_{\text{off}} + \sigma_c) \\
C =& -\phi k_{\text{on}} \sigma_c k_s \chi - \sigma_c N k_{\text{on}} k_s \chi \xi - \sigma_i \sigma_r k_s \chi (k_{\text{off}} + \sigma_c) \\
D =& k_s \chi \xi \phi k_{\text{on}}
\end{aligned}$$

We note that A and D are positive coefficients, C is negative. According to Descartes' Rule, whatever is the sign of B, polynomial (K.7) has 2 or 0 positive real root and exactly one negative real root. Let us show that there exists a unique positive value for C^* such that the steady state is positive.

We expressed the other variables at the steady state as functions of C^* (Equations (K.4), (K.5) and (K.3)). We supposed $\frac{\phi}{\sigma_c N} > C^*$, so that $I^* > 0$ and $R^* > 0$. Let us apply Budan's theorem (see Section 2.6.1) and find the number of roots the polynomial (K.7) in interval $(0, \frac{\phi}{\sigma_c N})$. Replacing X by $X + \frac{\phi}{N \sigma_c}$ in polynomial

a	b	c	d	number of sign variations
+	+	+	-	1
+	+	-	-	1
+	-	-	-	1
+	-	+	-	3

Table K.1: Sign variations of the coefficients of polynomial (K.7) when X was substituted by $X + \frac{\phi}{N\sigma_c}$. We wrote the sequence of coefficients by decreasing associated monomial degree a, b, c, d.

(K.7), yields a polynomial with a positive leading coefficient and a negative last coefficient (coefficient associated to the monomial of degree 0). The signs of the two coefficients in between are not trivial but do not need to be determined. Indeed, whatever are their signs, there is always at least one sign variation in the sequence of coefficients (see Table K.1). As a consequence, polynomial (K.7) always admits, at least, one positive real root in interval $(0, \frac{\phi}{N\sigma_c})$. We also showed that polynomial (K.7) admits only two positive real roots (so it cannot admits 3 roots in interval $(0, \frac{\phi}{N\sigma_c})$). Hence, (K.7) has exactly one real root in interval $(0, \frac{\phi}{\sigma_c N})$. The altruistic model with degradation of free ligand has a unique positive steady state. □

Note that we can show that the Jacobian matrix of system (K.1) has a strictly positive determinant and that its characteristic polynomial admits no real positive roots. As a consequence, the Jacobian matrix does not admit any real positive, nor zero, eigenvalues. We cannot, however, exclude the possibility of complex eigenvalues with positive real part. The stability analysis of the steady state remains inconclusive.