Universitat Autònoma de Barcelona

Information transmission through a nonlinear molecular signaling system: ErbB as a case study

Elisa Beltrán Sáez

UNIVERSITAT AUTÒNOMA DE BARCELONA

Information transmission through a nonlinear molecular signaling system: ErbB as a case study

by

Elisa Beltrán Sáez

and supervised by

Tomás Alarcón Cor

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy in Mathematics

in the

Centre de Recerca Matemàtica
Departament de Matemàtiques UAB

Programa de Doctorat en Matemàtiques

April 2019

Abstract

The ability of organisms to extract and store information from their surroundings marked a revolution in the history of life. Cells, from prokaryots to eukaryots, use specific receptors inserted in their membranes to detect extracellular molecules that cannot cross into the cell, where cell decisions are taken. Hence, those membrane receptors represent an information channel through which the environmental information can affect cell behavior and adaptation. In this Thesis, we modeled information transmission through the ErbB system, a family of receptors involved in many different cellular behaviors, such as cell proliferation or migration. Therefore, dysregulation of ErbB receptors is at the core of what can be called *information diseases*, that is, diseases that arise from the loss of the capacity to obtain and interpret extracellular information, i.e. tumor formation. Our results show a decrease in the information transmitted through the ErbB channel as the amount of ErbB receptors at the membrane increases. We considered different dynamics of the receptors and showed that the loss of information depends on the dynamics of interaction between the receptors, as well as on their interactions with the intracellular signaling machinery. In particular, we studied the interaction of active receptors with several signaling intracellular proteins and showed that the observed tendency of proteins to bind several binding sites with similar affinities translates into an increased synergy between the signaling proteins. All in all, quantifying and analysing these dependencies results in a better understanding of the dynamics and information transmission through ErbB and similar molecular systems.

Acknowledgements

Thanks to my supervisor, Tomás Alarcón for letting me in his group after my first try with a PhD. Thanks for believing in me, for letting me build a project that I feel mine and helping and supporting me along the way. (And for the final rush!)

I also want to thank the teachers who made learning much more interesting. Among them, Manuel Serra deserves a special thanks for introducing me to theoretical biology.

The list of the ways by which Guim helped me during this period is varied and probably long enough to duplicate the length of this thesis, so I will keep it short. I was very lucky to have him by my side, supporting me in the hard times and enjoying with me the good moments. My most sincere and deep thanks. I also want to thank Núria and Jaume, his parents, for adopting me.

My parents, Sara and Victor and my grandma Tere supported me and believed in me unconditionally. This is always a great source of energy and calm, both necessary for the PhD endurance race.

Thanks to my sister Mariana, who draw the ErbB receptors, and made me dream of exotic countries.

Thanks to my colleagues at CRM, who were always ready to help. The laughters at lunch time have been a good way to keep the mood up.

Thanks to my climbing partners and friends, specially to Joana, Cesc and David, with whom I shared so many evenings training to exhaustion after a day of hard work.

Finally, I want to thank the friends who have been closer during this period. I shared a house and so many beautiful moments with Guille, Marta, Anel, Carlos and Uri. They made me feel at home.

Contents

Chapter 1

Introduction

1.1 The biology: An insight into cell signaling

Most of the cells in our body have the same genetic material, yet we have hundreds of different cell types. One of the causes of this diversity is the way in which a cell's genetic information is 'read'. A myriad of molecular processes and interactions, coordinated both in space and time, take place during the acquisition of a cell fate or behavior from the information encoded in our genes. The ways in which such processes and interactions can be regulated are countless and there has been extensive work unveiling the different molecular pathways involved in cell 'decision making'.

Both the complexity arising from such a scenario and the particularities of every different case in the 'catalogue' of cell diversity makes it difficult to bring together detailed and global explanations of the whole process. It is possible nonetheless to describe the path leading from the level of the genes to level of the cell in a very general and simplified manner: the Central Dogma of Molecular Biology. Since it was proposed in 1958 by Francis Crick following the discovery of DNA's structure in 1956 by Watson and Crick, it has fallen short of an explanation to several phenomena in cell biology and an 'updated' version of it is more adequate to illustrate how cells depend on their genes.

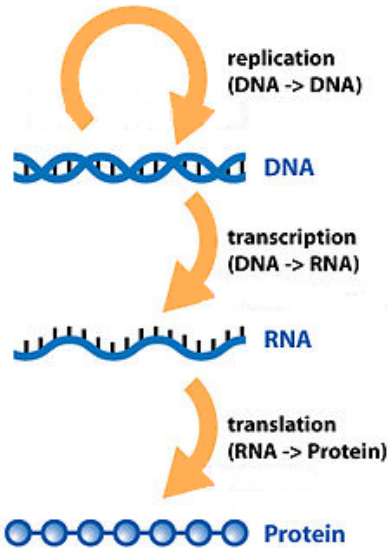


FIGURE 1.1: Schematic representation of the Central Dogma.
Figure modified from
https://commons.wikimedia.org/wiki/File:Central_Dogma_of_Molecular_Biochemistry_with_Enzymes.jpg

1.1.1 The Central Dogma revisited

The Central Dogma can be stated briefly as *DNA makes RNA and RNA makes proteins* (Fig.1.1). DNA (DeoxyriboNucleic Acid) is present in the nucleus of our cells in the form of chromosomes, and its sequence encodes our biological information. It can be copied into RNA (RiboNucleic Acid) in a process called *transcription*. RNA in turn is the template read during *translation*: the composition (and thus the structure) of a protein is translated from the genetic information encoded in RNA (copied from DNA). The reverse process is not possible.

In Crick's own words *'The Central Dogma (...) states that once information has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein'*[1].

Following Crick's definition of information, there have not yet been found instances of information transmission from proteins back to the DNA or RNA levels. But proteins do affect DNA and RNA in other ways: they control the accessibility to certain parts of the DNA and control the stability of RNA, among others. This is relevant inasmuch as the repertoire of the proteins in a cell determine the cell's

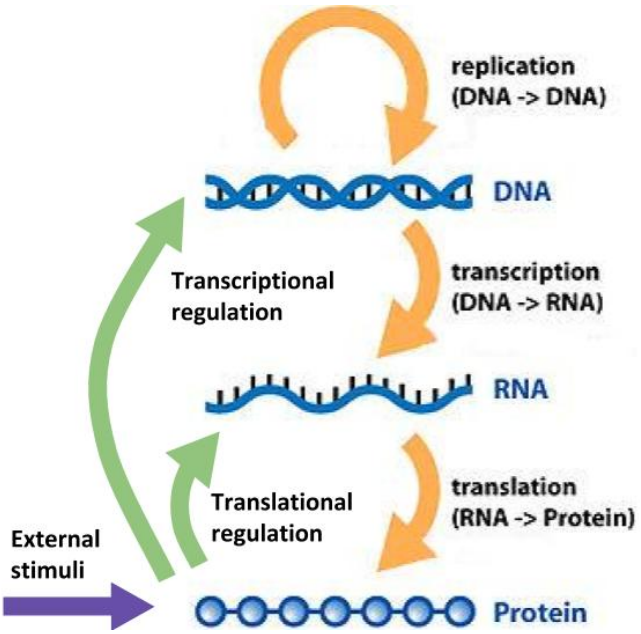


FIGURE 1.2: Central dogma revisited.
Figure modified from
https://commons.wikimedia.org/wiki/File:Central_Dogma_of_Molecular_Biochemistry_with_Enzymes.jpg

features in two ways: a) proteins are the main responsibles for a cell's activity and b) proteins regulate gene expression.

What makes cells with the same genetic information adopt different phenotypes is the way in which this genetic information is read (and thus which proteins are being synthesized), and this is in turn highly controled by proteins. Moreover, this is not a closed loop: cells need to detect and react to extracellular clues and adapt their behavior (protein state and gene expression) to the extracellular conditions (Fig 1.2).

Information gathering and processing is thus a crucial process for cells. We are familiar with such processes at the organismic scale: sensory organs of several kinds are able to detect different types of stimuli, which then will be processed by a more or less complex nervous system in order to elaborate a response. But how do cells receive and handle information? What are the 'sensory organs' of a cell? Which are the mechanisms involved in this process?

1.1.2 Basics of cell signaling

Eukariotic cells can be thought as basic units in multicellular organisms, delimited by a lipidic cell membrane that separates the extracellular environment from the

intracellular medium where cell components (organelles, molecules...) are located. The cell membrane allows the maintenance of cell homeostatis, an equilibrium in cell conditions appropriate for cell functioning. It also acts as a selective filter by letting liposoluble molecules diffuse freely through it, while acting as a barrier for polar hydrosoluble molecules.

Among all the information sources and types, information relevant to cells is often carried by organic molecules, known as ligands. Ligands can be liposolube (e.g. steroid hormones) or hydrosoluble (proteic ligands such as growth factors). Depending on this feature, their detection will be performed in different ways. Since liposoluble molecules can cross the cell membrane (and also the nuclear membrane) they do not need intermediaries for transmembrane transport. By contrast, hydrosoluble molecules cannot cross the cell membrane: their detection depends on intermediary molecules that will 'catch the relay' in the information transduction process.

1.1.3 Membrane receptors: a bridge between the exterior and the interior of a cell

Membrane receptors are proteins that are inserted in the cell membrane, spanning from the extracellular to the intracellular medium: they have an outer domain, a transmembrane domain and an intracellular domain. Their function is to detect hydrosoluble molecules that can be informative to the cell and transfer the information about such molecules to the interior of the cell, where cell decisions can be taken.

There are several types of membrane receptors which participate in information transduction, from which the receptors with tyrosine kinase activity (RTK) are the most extensively studied for many reasons. Firstly, this receptor family includes different kinds of receptors involved in many different cell responses. Furthermore, many of them are direct targets for anti-cancer drugs ([2–5]). Thus, their structure and dynamics have been the subject of research for more than 30 years and there is a substantial body of knowledge about RTKs. However, there still remain many unsolved issues about the functioning of those receptors *in vivo*, and also the phenotypic effects that altered function of the RTK may have and the mechanisms by which those phenotypic effects are expressed.

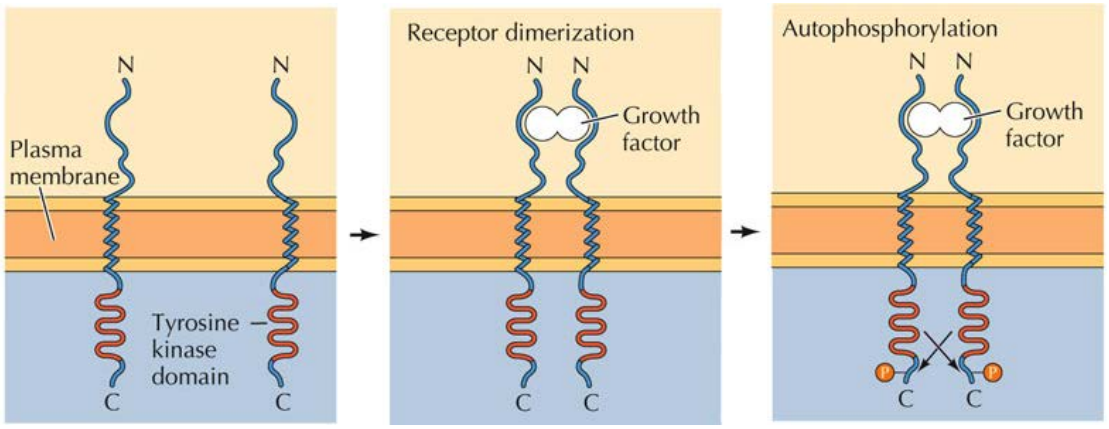


FIGURE 1.3: Schematic representation of the activation of an RTK. Figure from The Cell, GM Cooper, 4^{th} Ed.

One of the multiple ways of regulating protein activity is by means of the addition of a phosphate group to a protein's residue. This covalent modification changes the configuration of the protein, normally exposing the protein's domain in charge for a specific function (catalytic activity, binding sites...). The addition of phosphate groups is called phosphorylation and is commonly performed by the catalytic activity of a family of enzymes, called kinases, with the consumption of ATP. There are different types of kinases depending on the type of residue that is phosphorylated. In the case of RTK, the receptors are able to add phosphate groups to tyrosine residues, hence their name.

A canonical and simplified scheme of RTK operation is show in Fig.1.3.

RTKs are synthesized and inserted into the plasma membrane, where they are kept in an inactive configuration unless they bind a ligand. Upon ligand binding, the configuration of the receptor changes, which allows it to interact with other receptors to form stable dimers, where the receptors are able in turn to modify the intracellular part of each other by means of their kinase activity.

This modification of the intracellular domain by tyrosine phosphorylation creates docking sites for other proteins to bind, and it's one of the first steps in the signaling cascade that will activate, within the cell, the machinery needed to react to the ligand detected [6].

The RTK receptor family contains a lot of different types of receptors involved in many cell responses, and each subgroup has its particularities. We will focus in the Epidermal Growth Factor Receptor (EGFR) family.

1.1.4 The ErbB family

ErbB receptors are involved in a variety of cell responses, including growth, proliferation, differentiation, migration and apoptosis. This diversity is due to several combinatorial layers of interaction: at the level of the ligand, at the level of the receptors and at the intracellular level.

The ErbB family is a group of four related RTKs which are able to bind a variety of ligands (EGF, TGF, neuregulin, amphiregulin and betacatenin) with different affinities [7]. Which ligand binds to which receptor is crucial to the specificity of the responses this family of receptors are involved in.

The four different receptors of the family (ErbB1, ErbB2, ErbB3 and ErbB4) need to form dimers in order to become active, and both homodimers (dimers that contain two receptors of the same type) and heterodimers (dimers that contain two receptors of different types) have been observed experimentally ([8, 9]). The amount of dimers of each type depends on the cell type (which receptors and in which amount are being synthesized) and the relative affinities of the different combinations.

Upon dimerization the receptors can activate their partner by means of their kinase activity. The phosphorilated tyrosine residues of the receptors become then docking sites for signaling intracellular proteins that are at the top of the signaling pathways. The phosphorilation pattern in the intracellular domain of the receptors depends on the receptor sequence itself but also on the dimerization partner.

ErbB1 (also called EGFR) and ErbB4 have an extracellular domain that binds the ligand, a transmembrane domain and an intracellular domain with kinase activity, as most of the RTKs. ErbB3 receptor has all three domains but lacks kinase activity, so it needs to form heterodimers to become active.

ErbB2 is an orphan receptor -it does not bind any known ligand-, as it lacks the extracellular domain for ligand binding; this fact confers it a special role in signaling [10]. ErbB2 has been described as the preferred heterodimerization partner of all other ErbB receptors [8]. It has been detected in the membrane in different configurations: monomers, and both active and inactive dimers. We will study ErbB2 in greater detail in the next chapter.

1.1.4.1 ErbB receptors and cancer

Amidst a well organized signaling system that governs tissue structure by carefully controling cellular divisions, the ability to sustain proliferative growth is the fundamental trait of cancer cells [11]. Because of their role as growth signaling mediators, the family of ErbB receptors play a key role in cancer. In particular, ErbB2 overexpression has been reported in a number of human tumors, and it is found, as a crucial marker for poor prognosis, in about 18-25% human breast cancers [12, 13].

Without consideration of the underlying mechanisms, overexpression of ErbB2 surface molecules facilitates the formation of ErbB2 hetero- and homodimers. Because of their particular configuration, ErbB2 dimers do not need external growth signaling ligands to activate, and their excess results in sustained proliferative messaging that grants the cell with tissue-independent growth [14]. Targeted therapy focusing on ErbB2 function inhibition demonstrated positive results in early experiments [15], and it has since become an important target for cancer intervention strategies [4, 5, 16]. Its actual recognition as an oncogene - a driver of disease - has led to the development of a range of therapies (see e.g. [17] and references therein) that interfere with sustained signaling emanating from its overexpression. However, although ErbB2 targeted therapies improve the prognosis in the early stages of treatment, it is common that the patients develop resistance to the drugs within a period of one year of treatment initiation [18]. It has been proposed that the intricate structure of the ErbB signaling network is determinant to the elusive response of such receptors to treatment [19, 20].

Many open questions remain about the role of ErbB2 overexpression in cancer, and understanding its double-edged effect, as a necessary receptor for normal signaling and a driver of malignancy when overexpressed, might be key in producing novel cancer therapies. Are there viability limits in the concentration of ErbB2 relative to other receptors of the family, and if so, what external information is lost when trespassing them?

1.1.4.2 Downstream of ErbBs

As previously introduced, ErbB receptors are involved in many signaling pathways, affecting processes as diverse as growth, proliferation, differentiation, migration

and apoptosis [14]. This is possible because a plethora of intracellular proteins bind to the docking sites of the active ErbB receptors to initiate and continue the signaling process. Among such proteins, a particular set is fundamental in the pathways controlling cellular growth and proliferation.

PhosphatidylInositol 3-Kinases (PI3Ks) are enzymes with kinase activity that are involved in the PI3K/Akt/mTOR pathway, which mainly controls cell growth and proliferation [21]. As a result, disregulation of PI3Ks is often involved in tumor formation [22]. Some members of the PI3Ks family are able to bind the phosphotyrosines in ErbB receptors [23] and it has been observed that breast cancers that overexpress ErbB2 also maintain high PI3K activity [24].

Another signaling pathway involved in cell decision processes (as to whether differentiate or proliferate) is the MAP kinase pathway. Several proteins with kinase activity are involved in a chain reaction that begins at the cell membrane level and reaches the nucleus, where gene expression is modified. The first steps in this pathway involve Ras and Grb2, two proteins that are able to bind the ErbB receptors [25]. Mutations and/or overexpression of these proteins have been identified in tumors [26, 27], and overexpression of ErbB2 causes and increase in Ras/Grb2 activity and therefore in the MAP kinase signaling pathway [28]. This has been related with a change in cell behavior: cells switch from differentiation to proliferation states [29, 30].

Other canonical signaling pathways that can be triggered by ErbB activation are the PLC pathway and the JAK-STAT pathway, both involved in cell proliferation, motility and tumor formation [31–33]. Both PLC and JAK have binding sites for active ErbBs.

The interaction of the active ErbB receptors with the proteins listed in this section have been quantified in an extensive screening and published by Jones *et al.* [23].

1.1.5 Intracellular interactions

Signaling pathways controling cellular structure and function are regulated by specific molecular interactions between the different cell components as well as extracellular molecules. The stability of such interactions is shaped by structural and chemical complementarity of the molecules involved and can be characterized

by the equilibrium (or dissociation) constant of the binding and unbinding reaction at play.

Protein binding and unbinding has been studied extensively in the context of ligand-receptor interactions, giving rise to a mathematical framework to characterize them. Let us consider a ligand (L) and receptor (R) that bind to form a complex (C). The binding and unbinding reversible reaction is described by

$$L + R \underset{k_-}{\overset{k_+}{\rightleftharpoons}} C$$

where k_+ (k_-) are the binding (unbinding) constants, indicating how likely is the (un)binding reaction to occur.

According to this reaction, and following mass action kinetics [34], the complex is formed at a rate $k_{on} = k_+LR$ and it splits at a rate $k_{off} = k_-C$. The chemical equilibrium of the system happens for $k_{on} = k_{off}$, which yields a ratio between the concentrations of both free and bound receptors of

$$\frac{R}{C} = \frac{k_-}{k_+L}$$

The binding constant, k_+ has units $\mathrm{M}^{-1}\mathrm{s}^{-1}$ and the unbinding constant has units s^{-1}. $K_d = \frac{k_-}{k_+}$ is referred to as the equilibrium dissociation constant and its inverse, K_{eq}, the equilibrium constant, is a measure of the stability of the complex.

There are several bulk methods to measure the dissociation constant of a reaction [35], but it was not until the advent of single molecule experiments that individual binding and unbinding rates have been experimentally accessible [36–38]. As a result, whereas experimental dissociation constants for typical reactions can be found easily in the literature, the instances of experimental binding and unbinding constants in the literature are rare. Because similar values of K_d can be obtained from very different binding and unbinding kinetics (very fast binding and unbinding rates -fast turnover- can yield the same K_d as very slow binding and unbinding -slow turnover-) [39], the availability of experimental data of individual rates will is crucial to characterize protein interactions.

Besides the binding and unbinding reactions, covalent modification of the proteins is often needed in signaling pathways. One example of covalent modification we

are already familiar with is the addition of a phosphate group to certain residues in the intracellular part of the RTKs. There are other kinds of covalent modifications and most of them happen though a catalyzed reaction. Biological catalysts are called enzymes. Enzymes (E) are able to bind a substrate (S) and convert it to a product (P) by lowering the energy barrier between the unmodified and the modified states. The first model explaining successfully the kinetics of enzymes as a function of the rates of the different reactions (binding, unbinding and catalysis) and the substrate concentration was proposed by Michaelis and Menten [40]. The Michaelis-Menten model for enzyme kinetics is still widely used in biochemistry and related disciplines, more than 100 years after it was proposed, in order to predict the velocity of product formation as a function of the substrate concentration and the different reaction constants:

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

where $[\cdot]$ denotes a concentration and V_{max} is the maximum velocity of product formation, which happens when all the enzymes are occupied by a substrate. The Michaelis-Menten model introduces a parameter, $K_M = \frac{k_- + k_{cat}}{k_+}$, where k_{cat} is the rate at which the enzyme converts a bound substrate into a product.

Recent advances in single molecule spectroscopy methods have allowed to track the activity of a single enzyme over its activity cycles, giving the opportunity to tackle stochastic effects in the enzyme's kinetics. In the light of recent available data, it has been proposed that the unbinding reaction has a dual role: on the one hand, the unbinding rate needs to be low enough to avoid futile cycles (unbinding of the enzyme-substrate complex before product formation). However, new insights consider that, under particular conditions, such low unbinding rates result in cycles where the enzyme-substrate complex remains locked, making the process unproductive [41].

The possibility of such trade-off in unbinding kinetics poses a good example of how measuring individual rates instead of equilibrium constants might give further insight into protein interaction processes. We explore these issues in the context of ErbB-driven signaling in Chapter 3 of this thesis.

1.2 Background on mathematical modeling of the ErbB system

Besides being a focus for experimental research, the ErbB system has also been the subject of a number of modeling approaches.

The first models of ErbB receptors had their origin in classical models of enzyme kinetics aiming at understanding ligand binding, internalization and degradation of the receptors [42]. Dynamical models together with quantitative experiments proved the validity of the mathematical approach to explain the behavior of the ErbB receptor system [43]. As more and more experimental results became available thanks to the development of new biomolecular techniques, mathematical models incorporated additional layers of complexity.

This increase in complexity is illustrated by the publication in 2002 of the work by Schoeberl *et al.* [44] presented a comprehensive dynamical model of the ErbB receptors, compatible with experimental evidence. Due to the high number of species and reactions involved (it considers more than 100 reactions between over 50 species), the model can only be solved by means of numerical simulations. These modeling efforts have made it possible to predict the dependence of downstream signaling events on the receptor-ligand dynamics. In particular, it has been shown that the MAPK pathway activation depends in a non-linear way on the ErbB-ligand interaction.

Ten years later, Helikar *et al.* [45] presented an even larger model that included 245 species and over a thousand reactions. The model predicts enhanced receptor endocytosis upon Src overexpression, which has been confirmed by experimental evidence.

Other models have focused on the role of homo and heterodimerization processes in signaling. Hedricks *et al.* [46] proposed a multiscale approach where receptor trafficking and the binding and unbinding dynamics are taken into account by considering their characteristic time scales separately. From the interplay between the model and the experimental data, the authors concluded that the affinity between all receptor types is similar and that ErbB2 has a role in sustaining and amplifying ErbB response to the ligand.

Shankaran *et al.* ([47]) also studied dimerization of the ErbB receptors to predict the abundance of the different phosphorilated receptor dimers as a function of the monomer concentrations using a dynamical ODE based model of dimerization.

Zhang *et al.* [48] modeled the ErbB system by means of transfer functions to establish whether the information transmitted through the receptor system was receptor based or dimer based, concluding that the phosphorilated dimers are a better predictor of the intracellular state. This result will be used in Chapters 2 and 4 of this thesis.

Other modeling approaches have considered the different regulatory interacions (feedback loops) and their effect in the ErbB dynamics [49], and the combinatorial complexity of the ErbB network [50].

1.3 Information in biological systems

Many fields have used the methods of Information Theory to characterize communication processes since it was first proposed by Shannon in 1948 [51]. It has been widely used in biology, specially in neurophysiology [52]. Its application to molecular and cell biology has been less frequent, but a number of relevant approaches, both experimental and mathematical, have used information theory in the last decade. Refs. [53] and [54] offer good reviews on the subject. A brief introduction to the mathematics of information theory can be found in the Appendix, Section A.2.

Focusing in the molecular level, specially in signaling processes in cellular communication, information theory has emerged as an appropriate framework to study what, and to what extent, can be said about the extracellular cues from the response they elicit in the cell, or equivalently, to what extent the extracellular signals can influence the cell's behavior.

Pioneering the application of information theory to molecular processes were Bialek and Tkačik, who studied the information capacity of genetic regulatory elements and successfully applied it to the patterning of the fruit fly embryo to conclude that the patterning system achieves $\sim 90\%$ of its theoretical maximum information transmission [55, 56]. They also studied extensively the optimization of information flow in small genetic networks, where a single transcription factor protein

controls the readout of one or more genes [57–60]. The idea that information is maximized in biological processes of communication stands out remarkably along their work (see [61] for a concise summary of the mathematical formalism behind it).

Starting in 2010, a series of papers were published that addressed experimental measures of the information transduction through signaling networks, yielding interesting results on the nature of molecular information and what mechanisms can make information transmission effective. Cheong *et al.* [62] studied the limitations imposed by noise in information transmission and proposed a number of mechanisms to overcome such communication constraints. They concluded that time integration can increase the information transferred from the external ligand to the intracellular state through the Tumor Necrosis Factor (TNF) pathway. This is consistent with the result obtained by Uda *et al.* [63], which established that the information transmitted at individual time points represents a lower bound to the information transmitted along the complete time course of signaling. An experimental test for this hypothesis was presented by Selimkhanov *et al.* [64], where signaling dynamics were proven to have a key role in overcoming extrinsic noise.

Voliotis *et al.* [65] have addressed another interesting problem concerning the time modulation of the response by means of feedback loops: under particular conditions, negative feedback increases the information transfer. This results from preventing flat average response curves and thus reducing sensitivity to variation in the expression levels of the signaling molecules, all in all providing robustness to the system.

1.4 Outline of the thesis

The main goal of this thesis is to present an in-depth study of the transmission of information in the context of the cell, by means of dynamical, probabilistic and information-theoretical models. The main problem that we address is the role of membrane receptors and intracellular receptor-protein interactions in efficiently reading and transmitting the extracellular information.

We have seen along the introduction that extracellular information is often presented to the cell in the form of ligands. Ligands are detected by membrane

receptors, that undergo nonlinear dynamical interactions in the cell surface. Such dynamics of the membrane receptors is considered in detail in Chapter 2, with the aim of producing analytical expressions for the amount of active receptors upon stimulation by a ligand.

Following the receptors' reaction to an external ligand, intracellular proteins that bind the active receptors 'catch the relay' in information transmission. In Chapter 3, we address the dynamics of binding and unbinding of the intracellular proteins in order to predict the cellular state that results from a given level of receptor stimulation.

Chapter 2 and 3 are thus the necessary and consecutive links that allow us to study information transmission through the ErbB system. In Chapter 4, we use the results of Chapters 2 and 3 regarding the concentration of active receptors and the intracellular bound protein distributions to find analytical expressions for the mutual information under different signaling scenarios, thus providing a framework to understand the role and effect of the main components of the signaling system.

Finally, the results stemming from these three mathematical approaches are discussed in the Conclusions chapter, where further steps are introduced, leaving an open door for ongoing research.

Chapter 2

Mathematical models of the ErbB system

In Chapter 1, we have briefly introduced cell signaling and the molecular processes that take place when signaling is mediated by membrane receptors, as they pertain to the topic of this dissertation. In general, membrane receptors are activated upon extracellular ligand binding. Active receptors, in turn, activate intracellular proteins that will transmit the information to the interior of the cell, where cell decisions are taken. In many cases, this is a complex process that involves many different molecules and interactions.

In the case of membrane receptors with tyrosine kinase activity (RTKs), the first steps in the signaling cascade involve the interaction between inactive, monomeric receptors to form dimers (molecular complexes of two monomers) or oligomers (complexes of more than two monomers). This interaction between monomeric receptors adds to the complexity of the system: as we will see in this chapter, dimerization turns the receptors into a nonlinear channel for information transmission, which affects the amount of information can be transmitted through a channel, as we will see in Chapter 3.

In order to study such effects, first we need a model of the relevant processes in cell signaling through membrane receptors: ligand binding, receptor dimerization and activation, and activation of intracellular proteins by binding to active tyrosine residues. Initial conditions (concentrations of the different species considered

when the cell is stimulated with the ligand) are also relevant and need to be estimated or calculated for a proper analysis of the system. This involves estimating the concentration of receptors in the membrane (taken from experimental data) as well as taking into account the dynamics of the receptors in the absence of ligand (synthesis, internalization, degradation...).

In this chapter, we present a mean-field model of the dynamics of the receptors in the absence of ligand, in particular the dynamics of the ErbB2 receptor. We then continue to model the behavior of the system upon the addition of the ligand. The different processes involved in this molecular cascade have different time scales. We will explore the effect of such time scale separation and exploit it in order to turn the full model into a more tractable one that we can analyze. Last, we model the binding of intracellular proteins to the binding sites of active receptors. The state of the intracellular proteins will be used in the next chapter in order to compare the distrubitions of signaling molecules in and out of the cell.

2.1 Review of parameter values reported in the literature

When reviewing the literature searching for experimental values (or estimates) of parameters, concentrations or other quantifiable biological quantities, *BioNumbers* [66] comes to hand as an excellent tool. The aim of this web site is -as put forward by its the developers- *'(...) that the database will facilitate quantitative analysis and reasoning in a field of research where numbers tend to be soft and difficult to vouch for'.*

In general, it is difficult to obtain accurate values of the quantities one usually considers in a mathematical model of a biological system (rates, molecule numbers...). The reasons for this are twofold. Firstly, the complexity of biological systems (diversity, variability, dependence on the external conditions) make it difficult (if not impossible) to give a single estimate for the quantity of interest. On the other hand, there is often a lack of coordination between different approaches to the same problem and as such, the quantitative results of experimental studies not always match the needs of a mathematical model and viceversa. Bearing these

factors in mind, one can feel more at ease using approximate values of the different quantities to be considered in a model.

The models presented later in this chapter are systems of ordinary differential equations (ODEs) describing the changes in time of the concentration (or number of molecules) of ErbB receptors in several conformations. Their concentrations depend on several processes, each characterized by its own rate. Both the rates and concentrations used within the models we present below are taken from the literature. A review of our main sources of such data follows.

The most thoroughly studied ErbB ligand has been the Epidermal Growth Factor (EGF). Reported values of physiological EGF serum concentrations range from 60 pM to 30 nM [67], which is equivalent to a concentration of 40 to 20000 EGF molecules per μm^3. The interaction of EGF with ErbB1 has an equilibrium dissociation constant $K_d = 6.7 \cdot 10^{-10} M$, with an association rate $k_+ = 1.8 \cdot 10^8 M^{-1}$ min^{-1} and a dissociation rate $k_- = 0.12$ min^{-1}[68].

The number of ErbB1 in normal cells has been estimated to be between 50000 and 200000 receptors per cell in healthy cells, whereas it raises up to $2 - 3 \cdot 10^6$ receptors per cell in cancer cells [69]. ErbB2 is present at around 10000-60000 receptors per cell [70].

An intuitive interpretation of these number follows from the ensuing discussion. The volume of an eukaryotic cell is around 2000 μm^3. Assuming a spherical cell, the corresponding surface area is around 800 μm^2. The density of receptors at the membrane ranges from less than 1 to 400 ErbB1 molecules/μm^2 and from 10 to 100 molecules of ErbB2/μm^2.

Values regarding the receptor dimerization rates are given in [46], where the binding rates are estimated to be 10^3 min^{-1} and the unbinding rates are estimated to be 0.1 min^{-1}.

Values of the parameters characterising the interaction of intracellular proteins with phosphotyrosines are given in [23]. They provide experimental measurements of the dissociation constant between a collection of intracellular proteins to every phosphotyrosine site in ErbB receptors. High affinity interactions have $K_d \sim 10^2$nM while low affinity interactions have $K_d \sim 2 \cdot 10^3$nM. Although this study provides very extensive information about such interactions, the value of K_d is sometimes not sufficient to address the dynamics of binding and unbinding of the

intracellular proteins to the phosphotyrosine sites, as it will be analysed in the Chapter 3.

Estimated values of intracellular proteins-phosphotyrosine sites binding and unbinding constants used in ErbB models can be found in [44]. In this computational model for signaling, intracellular protein binding rates are estimated to be $10^{-2}\text{nM}^{-1}\text{min}^{-1}$ and unbinding rates are estimated to be 10 min^{-1}, which would give $K_d \sim 10^3\text{nM}$ (compatible with the data given in [23]). However, recent single molecule experiments have allowed to measure individual binding and unbinding rates. In [37], the measurements of the binding and unbinding rates of chaperonin and co-chaperonin interactions yielded values of 0.1 and 30 min^{-1}, respectively. Studies with selectin [71] and cadherins [72] yielded similar results for the unbinding rate (15 min^{-1} and 60 min^{-1}, respectively).

The values of the parameters used throughout this thesis (unless other values are specified) are summarized in Table 2.1.

Receptor dynamics			
μ_1	2400 min^{-1}	μ_2	12 min^{-1}
k_{s+}	10^7 $\text{M}^{-1}\text{min}^{-1}$	k_{s-}	10^{-3} min^{-1}
k_{1+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{1-}	0.1 min^{-1}
k_{12+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{12-}	0.1 min^{-1}
k_{2+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{2-}	10^4 min^{-1}
a_{1+}	1 min^{-1}	a_{1-}	0 min^{-1}
a_{12+}	1 min^{-1}	a_{12-}	0 min^{-1}
a_{2+}	1 min^{-1}	a_{2-}	10 min^{-1}
δ_{R1}	0.014 min^{-1}	δ_{R2}	0.0006 min^{-1}
δ_{D1}	0.1 min^{-1}	δ_{D12}	0.1 min^{-1}
$[R_{1T}]$	$5 - 20 \cdot 10^4$ cell^{-1}	$[R_{2T}]$	$1 - 6 \cdot 10^4$ cell^{-1}
$[S_T]$	pM to nM	P_{tot}	500 nM
Intracellular proteins dynamics			
k_{p+}	0.1 nM^{-1} min^{-1}	k_{p-}	30 min^{-1}

TABLE 2.1: Reference parameters used in our model analysis and simulations.

Reported values for the intracellular concentration of signaling molecules range from 0.1 to 1 μM [73].

2.2 Dynamical models of ErbB receptors

EGF-like ligands have been described to promote dimerization of monomeric receptors, but the mechanism by which dimerization happens has been a controversial issue. In the first studies on ErbB dimerization, it was hypothesized that ligands form dimers themselves, and that the interaction between receptor-bound ligands brought the monomers close together to enable receptor dimerization [74]. Later, experimental results supported an alternative hypothesis for receptor dimerization: ligand binding exposes the receptors' dimerization domains allowing direct interaction between the monomeric receptors: dimerization is ligand induced, but receptor mediated [75]. We refer to this mechanism as *ligand-induced dimerization* and we model the dynamics of the receptors according to this mechanism in Section 2.2.1.2.

Out of the four members of the ErbB receptor family, ErbB2 exhibits some noteworthy peculiarities. Firstly, it has been identified as the preferred heterodimerization partner for all the other ErbB receptors [8]. More importantly, ErbB2 does not bind any known extracellular ligand and it has been shown to present a rich behavior at the membrane even in the absence of ligands: ErbB2 has been detected at the membrane in the form of monomers, as well as in the form of homodimers (both active and inactive) in a dynamic equilibrium [76]. Why these dimers form remains an open question, though it has been hypothesized that kidnapping ErbB2 in the form of dimers avoids 'accidental' or non-specific signaling. In any case, there are several processes involving ErbB2 in the absence of ligands: synthesis and insertion in the membrane, interaction with other ErbB2 monomers, activation/deactivation, internalization, recycling and degradation.

It has recently been discovered that dimerization in the absence of ligands is not exclusive of ErbB2 but all the members of the ErbB family are in a dynamic equilibrium between monomeric, inactive dimeric and active dimeric forms at the membrane. These dimeric forms are referred to as *predimers* [77]. In contrast to the ErbB2, since the other ErbB receptors have an extracellular binding domain which avoids activation in the absence of a ligand, the active dimeric forms of ErbB1, ErbB3 and ErbB4 have poor kinase activity and are rare. The amount of receptors that are forming predimers at the cell membrane has been measured to be between 40% and 100% [78]. We devote Section 2.2.2 to model ErbB dynamics when *pre-dimerization* is the main mechanism of receptor dimerization.

The existence of both dimeric and monomeric receptors at the membrane suggests that a combination of the mechanisms listed above could be taking place at the membrane.

The dimerization mechanism is relevant to establish the membrane composition of signaling dimers. Once the dimers are activated by a ligand, they are interalized and then degraded or recycled. If we consider that internalized receptors keep on signaling (until they are degraded) with the same intensity as in the membrane and that recycled receptor are reinserted in the membrane in an inactive configuration, we can consider a one compartment model for signaling.

2.2.1 Receptor dimerization driven by ligand binding

2.2.1.1 Dynamics of the ErbB2 receptors in the absence of ligand: steady state at the membrane.

In this section, we address two issues regarding ligand-free ErbB2 dynamics: first, active dimers can signal in the absence of a ligand, and this will influence information transmission through this system, and second, the receptor composition at the membrane at the onset of signaling is a key feature that determines the activation state of the receptors that the system will reach and consequently, which (and how intensely) proteins will be recruited in the signaling process.

Consider an ErbB2 molecule through its life cycle (in the absence of ligand), from synthesis to degradation (shown schematically in Fig. 2.1). As it is synthesized, it travels to the membrane inserted in a vesicle, which will 'fuse' with the membrane. Upon fusion, the receptors will be inserted in the cell membrane, with the catalytic (kinase) site in the interior of the cell. Once at the membrane, there are many paths an ErbB2 monomer can follow: it can diffuse through the membrane until it finds another ErbB2 monomer and interact to form a dimer. While the monomer is part of the dimer, it will be switching between the two states (active and inactive), until the dimer splits or is internalized or degraded. Dimers, as well as monomers, can be internalized by means of invaginations of the cell membrane. The internalized receptors can either be recycled to the cell membrane again, or be degraded.

As shown in Fig. 2.1, ErbB2 can be found at the membrane in three configurations: monomers (R), inactive dimers (D) and active dimers (D^*). Consider also

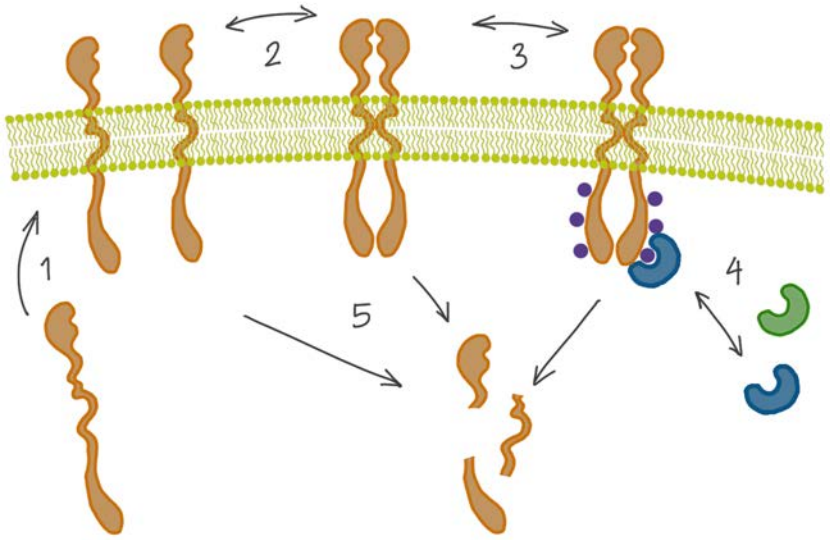


FIGURE 2.1: Simplified behavior of ErbB2 behavior in the absence of a ligand. A monomeric receptor is inserted in the membrane (1), and can diffuse until it interacts with another monomeric receptor and forms a dimer (2). Dimers switch between an active and an inactive state (3). In their active state, they can phosphorilate tyrosine residues in their intracellular domain, which will act as docking sites for many intracellular proteins (4). Bidirectional arrows indicate reversible processes.

the most relevant processes: synthesis/insertion to the membrane (with rate μ), dimerization/undimerization (with rates k_+ and k_-, respectively), activation/deactivation (with rates a_+ and a_-), internalization/degradation (with rate δ). In this model, μ and δ account for the turnover of the receptors (transitions *in* and *out* of the membrane, respectively), while the rest of the parameters are related to in-membrane dynamics.

The inactive dimers have been described to be more stable than their active dimers counterparts [76]. This has implications regarding the values of the parameters for activation and inactivation of ErbB2 homodimers: the inactivation rate will be greater than the activation rate ($k_- > k_+$). Hence, we expect the amount of inactive to be higher in the membrane at the steady state.

For simplicity and analytical tractability, we omit in this model the processes of internalization into vesicles and recycling, considering only one compartment (the cell membrane), instead of two (membrane and vesicles), which would double the number of different species of the system.

With such ingredients, we formulate the following model for the ErbB2 dynamics in the absence of a ligand:

$$\frac{dR}{dt} = \mu - 2k_+R^2 + 2k_-D - \delta_R R \tag{2.1}$$

$$\frac{dD}{dt} = k_+R^2 - k_-D - a_+D + a_-D^* - \delta_D D \tag{2.2}$$

$$\frac{dD^*}{dt} = a_+D - a_-D^* - \delta_{D^*}D^* \tag{2.3}$$

where R is the monomeric ErbB2 receptor, D is the inactive ErbB2 dimer and D^* is the active ErbB2 dimer.

The concentration of every species can be found at the steady state by solving the following set of non-linear equations.

$$\mu - 2k_+R_0^2 + 2k_-D_0 - \delta_R R_0 = 0 \tag{2.4}$$

$$k_+R_0^2 - k_-D_0 - a_+D_0 + a_-D_0^* - \delta_D D_0 = 0 \tag{2.5}$$

$$a_+D_0 - a_-D_0^* - \delta_{D^*}D_0^* = 0 \tag{2.6}$$

We first consider that the degradation rate of the monomeric form of ErbB2 is small enough ($\delta_R << 1$), we find that the state of the system at the steady state (at first order in δ_R) is

$$R_0 = \sqrt{\frac{\mu + 2k_-D_0}{2k_+}} \tag{2.7}$$

$$D_0 = \frac{k_+R_0^2 + a_-D_0^*}{k_- + a_+ + \delta_D} \tag{2.8}$$

$$D_0^* = \frac{a_+D_0}{a_- + \delta_{D^*}} \tag{2.9}$$

$$R_0 = \sqrt{\frac{\mu}{2k_+}\left(1 + \frac{k_-}{\delta_D + \frac{\delta_{D^*} a_+}{a_- + \delta_{D^*}}}\right)} \tag{2.10}$$

$$D_0 = \frac{1}{2}\frac{\mu}{\delta_{D^*}}\frac{1}{\frac{a_+}{a_- + \delta_{D^*}} + \frac{\delta_D}{\delta_{D^*}}} \tag{2.11}$$

$$D_0^* = \frac{1}{2}\frac{\mu}{\delta_{D^*}}\frac{1}{1 + \frac{\delta_D}{\delta_{D^*}}\frac{(a_- + \delta_{D^*})}{a_+}} \tag{2.12}$$

The small δ_R approximation provides a simple expression for the steady state concentration of the different species (R, D and D^*) of the system.

By relaxing the $\delta_R << 1$ approximation the analytic expression for the steady state solution (R_d) becomes more complicated:

$$R_d = \frac{1}{2}\frac{\delta_R}{2k_+}\left(1 + \frac{k_-}{\delta_D + a_+\frac{\delta_{D^*}}{a_- + \delta_{d^*}}}\right)\left(\sqrt{1 + 4\mu\frac{2k_+}{\delta_R^2}\frac{1}{1 + \frac{k_-}{\delta_D + a_+\frac{\delta_{D^*}}{a_- + \delta_{d^*}}}}} - 1\right) \tag{2.13}$$

which can be rewritten as a function of the solution we found with the small δ approximation (Eq. 2.10), R_0

$$R_d = \frac{1}{2}\frac{\delta_R}{\mu}\left(\frac{2k_+}{\mu}\right)^3 R_0\left(\sqrt{R_0^2 + \left(2\frac{\mu}{\delta_R}\right)^2} - R_0\right), \tag{2.14}$$

where $R_0 = \sqrt{\frac{\mu}{2k_+}\left(1 + \frac{k_-}{\delta_D + \frac{\delta_{D^*} a_+}{a_- + \delta_{D^*}}}\right)}$

Fig.2.2 shows the amount of each configuration at the membrane for different values of the synthesis rate of the ErbB2 receptor. We can see that, for reference values of the parameters (Table 2.1), most of the ErbB2 is in a dimeric inactive form.

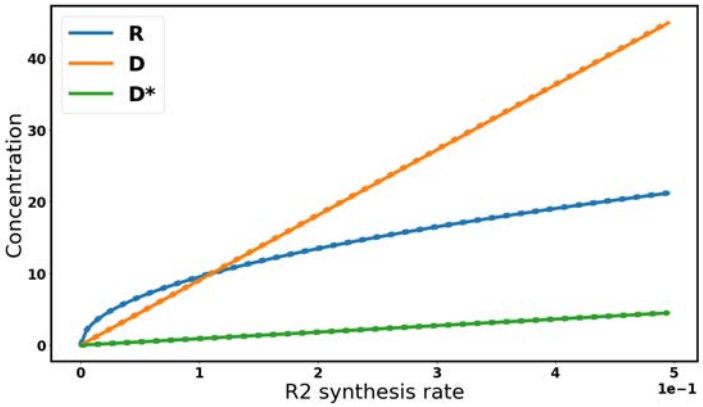


FIGURE 2.2: Amount of the different forms of R2 as a function of the synthesis rate μ. Solid lines show the results of the numerical integration of the system, dotted lines show the concentration at the steady state calculated with the small δ approximation.

2.2.1.2 Ligand-dependent receptor dynamics.

When considering ligand-driven signaling, the system becomes slightly more complicated. Ligand binding and unbinding and dimerization between the different types of monomers increase the number of reactions that are to be taken into account considerably.

Consider now the life cycle of an ErbB1 receptor in the cell (Figs. 2.3 and 2.4). Once synthesized, the receptor will be inserted in the cell membrane, with the ligand-binding domain in the extracellular environment and the kinase domain in the interior of the cell. It will not interact with other monomeric receptors until it is 'unlocked' by ligand binding. Meanwhile, it can be internalized and recycled or degraded. Once it binds a ligand molecule, it can either form dimers with other monomers of the same kind (homodimers) or with monomers of other members of the ErbB family (heterodimers). Upon dimer formation, there is activation of the receptors by addition of phosphate groups to the tyrosine residues, catalyzed by the kinase activity of the intracellular part of the receptors. Ligand-bound monomers, as well as inactive and active dimers will eventually get internalized and recycled or degraded. Bidirectional arrows in Figs 2.3 and 2.4 represent reversible processes.

The set of processes that determine the dynamics of the system of receptors described above can be modelled as a set of ODEs. For simplicity, we consider only two members of the ErbB family: ErbB1 (R_1, ligand-dependent) and ErbB2 (R_2, ligand-independent). L stands for the ligand concentration, C_1 stand for

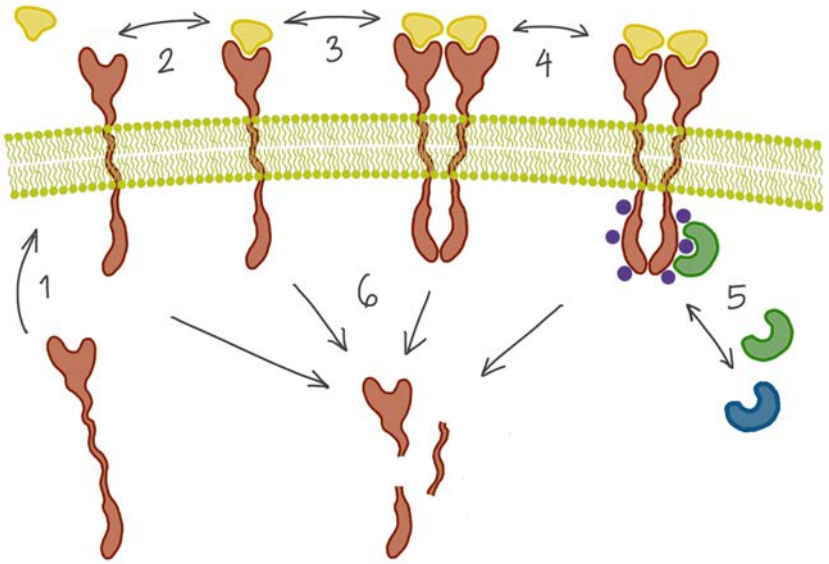


FIGURE 2.3: Schematic representation of the life cycle of ErbB1 when forming homodimers.

the ligand-bound ErbB1 receptor, D_{ij} (with $i, j = 1, 2$) are the different types of inactive dimers and D^*_{ij} are the active dimers.

$$\frac{dL}{dt} = -k_{s+}LR_1 + k_{s-}C_1 \quad (2.15)$$

$$\frac{dR_1}{dt} = \mu_1 - k_{s+}LR_1 + k_{s-}C_1 - \delta_1 R_1 \quad (2.16)$$

$$\frac{dR_2}{dt} = \mu_2 - k_{12+}C_1R_2 - 2k_{2+}R_2^2 + k_{12-}D_{12} + 2k_{2-}D_2 - \delta_2 R_2 \quad (2.17)$$

$$\frac{dC_1}{dt} = k_{s+}SR_1 - (k_{s-} + 2k_{1+}C_1 + k_{12+}R_2)C_1 + 2k_{1-}D_1 + k_{12-}D_{12} - \delta_1 D_1 \quad (2.18)$$

$$\frac{dD_1}{dt} = k_{1+}C_1^2 - (k_{1-} + a_{1+})D_1 + a_{1-}D_1^* - \delta_1 D_1 \quad (2.19)$$

$$\frac{dD_{12}}{dt} = k_{12+}R_2C_1 - (k_{12-} + a_{12+})D_{12} + a_{12-}D_{12}^* - \delta_{12} D_{12} \quad (2.20)$$

$$\frac{dD_2}{dt} = k_{2+}R_2^2 - (k_{2-} + a_{2+})D_2 + a_{2-}D_2^* - \delta_2 D_2 \quad (2.21)$$

$$\frac{dD_1^*}{dt} = a_{1+}D_1 - a_{1-}D_1^* - \delta_1 D_1^* \quad (2.22)$$

$$\frac{dD_{12}^*}{dt} = a_{12+}D_{12} - a_{12-}D_{12}^* - \delta_{12} D_{12}^* \quad (2.23)$$

$$\frac{dD_2^*}{dt} = a_{2+}D_2 - a_{2-}D_2^* - \delta_2 D_2^* \quad (2.24)$$

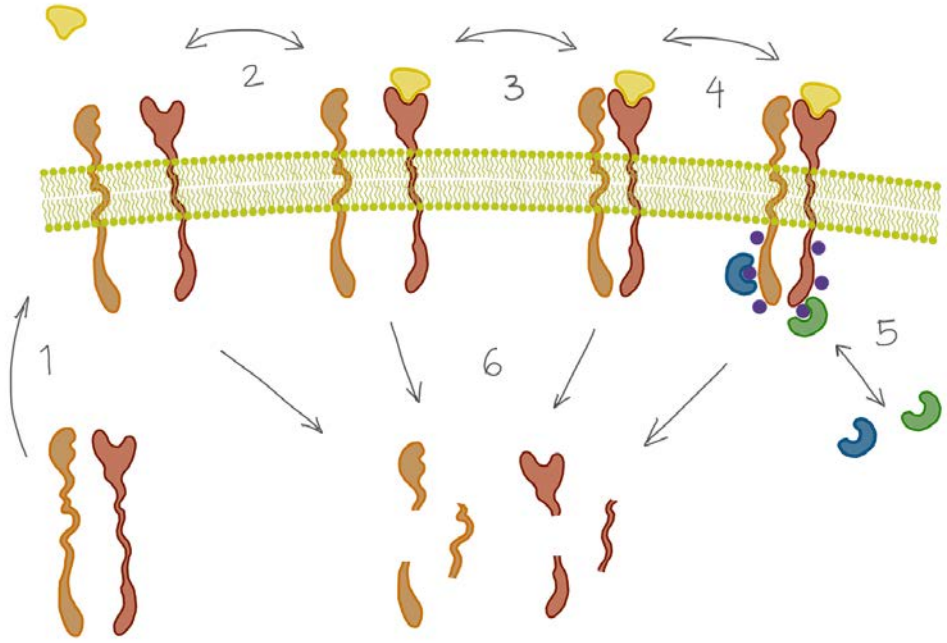


FIGURE 2.4: Schematic representation of the life cycle of ErbB1 and ErbB2 when forming heterodimers.

Note that in the absence of ligand, the ErbB1 receptor is synthesized with rate μ_1 and internalized/degraded with rate δ_1, so its steady state concentration on the membrane ($R_{1,ss}$) is $R_{1,ss} = \frac{\mu_1}{\delta_1}$. The steady state of the ErbB2 receptor was described in the previous section. We consider that the rest of species are not present before the system is exposed to ligand (at $t = 0$).

In the following, we consider a pulse of ligand, so that the time of exposure to the ligand is very short in comparison to the dynamics of the receptors.

Multiple time scales in the dynamics of the response to ligand

In this section we explore the dynamics of the response of the system of receptors when stimulated with a pulse of ligand.

In Fig. 2.5, we show that, as time progresses, the system goes through a number of different regimes, associated with different time scales (for a more detailed presentation, see Appendix, Section B.1). The only species that change in the shortest characteristic time scale (i.e. immediately upon ligand stimuation) are the ligand L and the ligand-bound receptor C_1 (Fig. 2.5 (a)); R_1 is in excess

and is not shown in the figure). The amount of the different forms of R_2 remains constant at the steady state value (different from 0) that it attained before ligand stimulation - as determined in Section 2.2.1.1. At an intermediate time scale (Fig.2.5 (b)), the concentrations of the ligand-bound receptor C_1 and the dimers D_1 and D_{12} evolve. The ErbB2 receptor R_2 and its dimeric inactive form D_2 also change: they are perturbed from their previous steady state. At this time scale, the active dimers D_1^* and D_{12}^* are not formed yet and D_2^* still remains at the steady state it attained in the absence of ligand, different from 0 (see Section 2.2.1.1. The dimers activate and they are internalized or degraded at a longer time scale (Fig. 2.5, c,d). R_2 and its different conformations reach the no-ligand steady state once the other species disappear from they system.

According to the values of the parameters proposed in the literature and reported in section 2.1 (see Table 2.1), the binding rate of the ligand is much faster than the rest of the reactions (Fig. 2.5, a). Therefore, we consider that within a very short time scale from initial stimulation with the ligand, all of it has bound a receptor so that the initial conditions at the onset of signaling given an initial ligand concentration L_0 are $L = 0$, $R_1 = R_{1,ss} - L_0 = \frac{\mu_1}{\delta_1} - L_0$ and $C_1 = L_0$ (see section B.1).

Intermediate time scale: dynamics of the dimers D_1 and D_{12}

We are interested in the amount of dimers of each type (D_1, D_{12} and D_2) formed upon ligand binding. We assume that the unbinding and activation rates of the active receptors are negligible in the time scale of dimerization (see Fig. 2.5 (b)), and Appendix, Section B.1). Within this time scale, considering only the dimerization processes, the system reduces to:

$$\frac{dR_2}{dt} = k_{12+}C_1R_2 - 2k_{2+}R_2^2 + 2k_{2-}D_2 \tag{2.25}$$

$$\frac{dC_1}{dt} = -(2k_{1+}C_1 + k_{12+}R_2)C_1 \tag{2.26}$$

$$\frac{dD_1}{dt} = k_{1+}C_1^2 \tag{2.27}$$

$$\frac{dD_{12}}{dt} = k_{12+}R_2C_1 \tag{2.28}$$

$$\frac{dD_2}{dt} = k_{2+}R_2^22 - k_{2-}D_2 \tag{2.29}$$

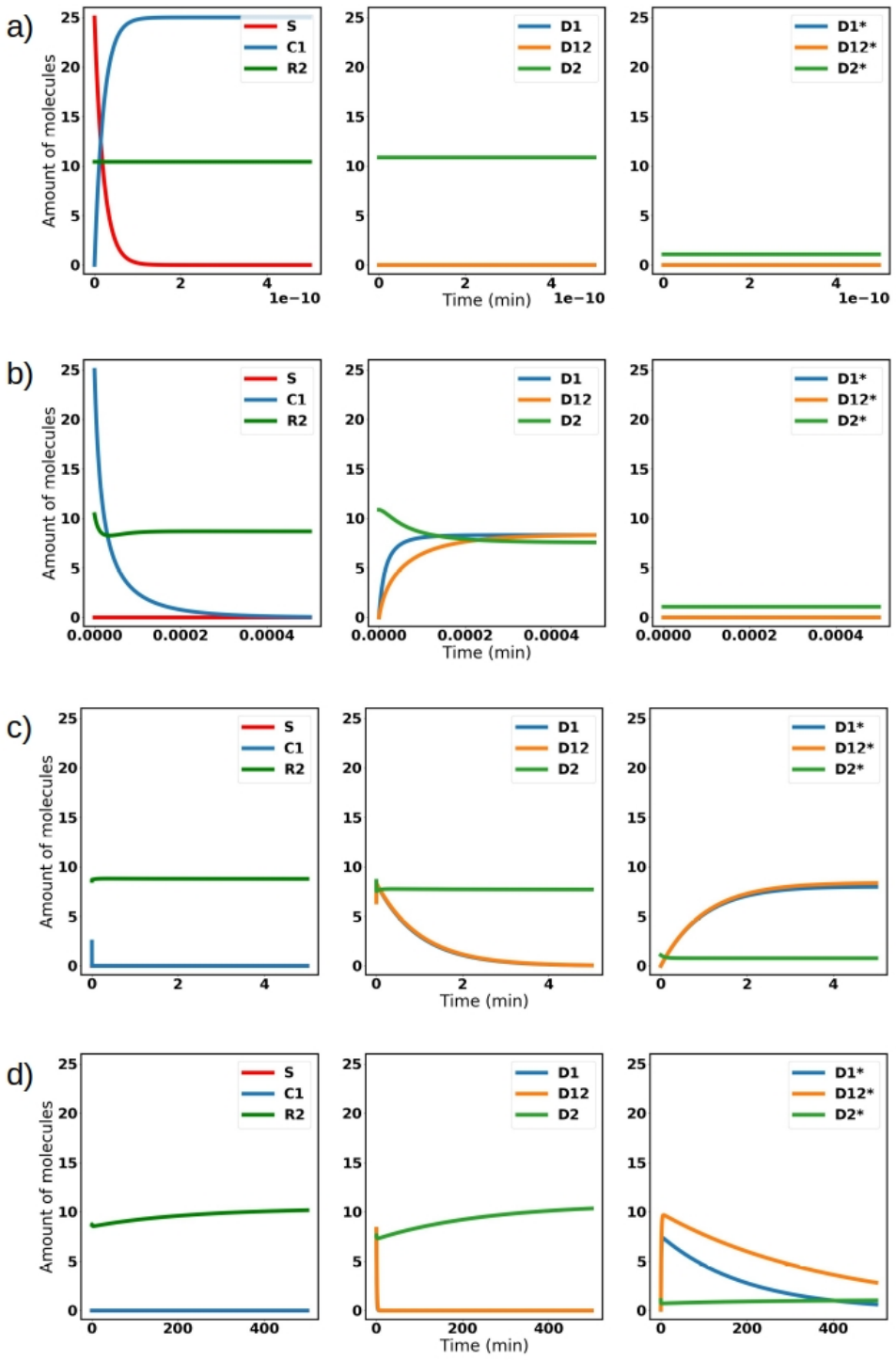


FIGURE 2.5: Time evolution of the amount of the different species in the system at different timescales. a) Short time scale (ligand binding), b) intermediate time scale (receptor dimerization), c) long time scale (receptor activation), d) very long time scale (receptor degradation). The parameters are shown in Table 2.1.

Since we are interested in the amount of dimers of each type that are formed upon stimulation of the system with a pulse of ligand, and once the dimers have been formed, the concentration of ligand bound receptor decays to 0, the concentrations of the system at the steady state cannot be determined analytically. In order to provide an analytic estimation of the steady state values of D_1 and D_{12}, we consider an alternative approach to find the amount of dimers of each type formed upon detecting a concentration of the ligand L_0, which will be needed in Chapter 4. Such analytical estimates cannot be obtained by directly tackling Eqs (2.25-2.29). Two further key assumptions here are:

- Because R_1 is in excess, all the ligand molecules bind a receptor, and hence we take the initial concentration of the ligand bound receptor (C_1) as equal to the total amount of ligand presented to the cell, L_0.
- If we suppose that the R_2 receptor reaches its steady state (presented in section 2.2.1.1) rapidly after a perturbation (see Fig. 2.5,b), then we can simplify the model presented above by considering the amount of R_2 constant (and equal to its steady state value, R_0, Eq. 2.7), and the amount of C_1 decaying because of interaction (with itself and with ErbB2).

The system is thus further simplified to:

$$\frac{dC_1}{dt} = -(2k_{1+}C_1 + k_{12+}R_0)C_1 \tag{2.30}$$

$$\frac{dD_1}{dt} = k_{1+}C_1^2 \tag{2.31}$$

$$\frac{dD_{12}}{dt} = k_{12+}R_0C_1 \tag{2.32}$$

Now we can find the amount of each type of dimers as a function of C_1:

$$\frac{dD_1}{dC_1} = \frac{k_{1+}C_1^2}{-(2k_{1+}C_1 + k_{12+}R_0)C_1} \tag{2.33}$$

$$\frac{dD_{12}}{dC_1} = \frac{k_{12+}R_0C_1}{-(2k_{1+}C_1 + k_{12+}R_0)C_1} \tag{2.34}$$

which can be integrated to get

$$D_1(C_1) = \frac{k_{12+}R_0 \log\left(2k_{1+}C_1 + k_{12+}R_0\right)}{4k_{1+}} - \frac{C_1}{2} + c_1 \tag{2.35}$$

$$D_{12}(C_1) = -\frac{k_{12+}R_0 \log\left(2k_{1+}C_1 + k_{12+}R_0\right)}{2k_{1+}} + c_{12} \tag{2.36}$$

The constants c_1 and c_{12} can be found by considering that $D_1 = 0$ and $D_{12} = 0$ when $C_1 = S_0$, so that

$$D_1(C_1) = \frac{L_0 - C_1}{2} + \frac{k_{12+}}{4k_{1+}}R_0 \log\left(\frac{2k_{1+}C_1 + k_{12+}R_0}{2k_{1+}L_0 + k_{12+}R_0}\right) \tag{2.37}$$

$$D_{12}(C_1) = \frac{k_{12+}}{2k_{1+}}R_0 \log\left(1 + \frac{2k_{1+}C_1 + k_{12+}R_0}{2k_{1+}L_0 + k_{12+}R_0}\right) \tag{2.38}$$

Now we can now take the limit of D_{12} as the amount of ligand bound receptor C_1 approaches 0 (which is the maximum amount of dimers that is formed)

$$D_{1,max} = \lim_{C_1 \to 0} D_1 = \frac{L_0}{2} - \frac{k_{12+}}{4k_{1+}}R_0 \log\left(1 + \frac{2k_{1+}L_0}{k_{12+}R_0}\right) \tag{2.39}$$

$$D_{12,max} = \lim_{C_1 \to 0} D_{12} = \frac{k_{12+}}{2k_{1+}}R_0 \log\left(1 + \frac{2k_{1+}L_0}{k_{12+}R_0}\right) \tag{2.40}$$

As depicted in Fig. 2.6 the values of D_1 and D_{12} obtained with the method described above are only a moderately accurate fit to those obtained by numerical integration of Eqs. (2.25-2.29), as they capture the tendency of D_1 to decrease and D_{12} to increase. The mismatch is due in part to the fact that the full model includes reactions (unbinding, activation and degradation) not taken into account in the analytical calculation. Also, the assumption of constant R_2 is very strong (as can be seen in Fig. 2.5 (b)). Nonetheless, the ratio between the homo- and hetero-dimers shows a good agreement between the simulations and the analytical expression (see Fig. 2.6, right), indicating that what we are missing in the analytical approximation is the loss of the dimers by other reactions, rather than the interconversion of dimer types.

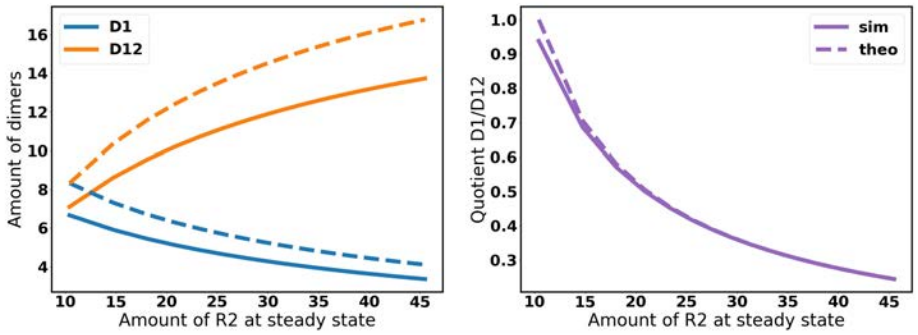


FIGURE 2.6: Left: Amount of dimers of each type as a function of the amount of monomeric R_2 at the steady state, for an initial value of 25 ligand molecules. Solid lines are computed by means of numerical integration of the whole system for different values of R_0, dotted lines are calculated by means of Eqs. 2.39 and 2.40. Left: Quotient between the amount of dimers of each type $\frac{D_1}{D_{12}}$. The parameter values are shown in Table 2.1.

We have assumed that the equilibrium between the different configurations of ErbB2 is reached instantaneously. In the Section 2.2.1.1 we calculated the steady state concentration of the different forms of ErbB2 (Eqs. 2.7-2.9). Using the steady state values for the ErbB2 subsystem, we can calculate the ratio between ligand-dependent and ligand-independent active dimers as a function of the amount of monomeric ErbB2 at the steady state.

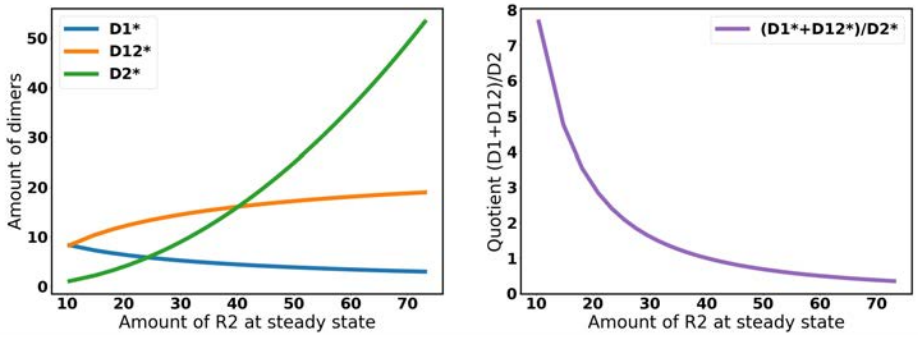


FIGURE 2.7: Left: Amount of dimers of each type as a function of the amount of monomeric R_2 at the steady state, for an initial value of 25 ligand molecules. Results of the numerical integration of the whole system for different values of R_0. Left: Quotient between the amount of informative active dimers and non-informative active dimers $\frac{D_1^*+D_{12}^*}{D_2^*}$. The parameter values are shown in Table 2.1.

Fig. 2.7 shows that the relative amount of informative dimers (D_1 and D_{12}) decreases when compared to the amount of non-informative dimers (D_2) as a function of steady state monomeric HER2 concentration. For high concentrations of ErbB2,

the amount of informative dimers is smaller than the amount of non-informative dimers (which is most likely a bad situation for signaling). Furthermore, Fig. 2.6 shows that the amount of dimers of each type depends on the initial concentration of ErbB2 monomers (determined by the steady state of the different conformations of ErbB2 at the membrane).

2.2.2 Predimerization: on the amount of homo and heterodimers in the absence of ligand

The presence of ErbB receptors in homo- and heterodimeric forms in the membrane in the absence of a ligand (not only for ErbB2 but also for other members of the ErbB family) has been reported in several studies [77], [78]. These forms are referred to as *predimers*, and their prevalence (the amount of receptors involved in a predimer) has been quantified to be between 40% and 100% of the total amount of ErbB receptors at the membrane. In this scenario, the amount of every predimer type (homo- and heterodimers) will be determined by the amount of each type of receptor and the relative affinities between them. Fig. 2.8 shows schematically the membrane composition in the absence of a ligand when predimers are formed prior to signaling.

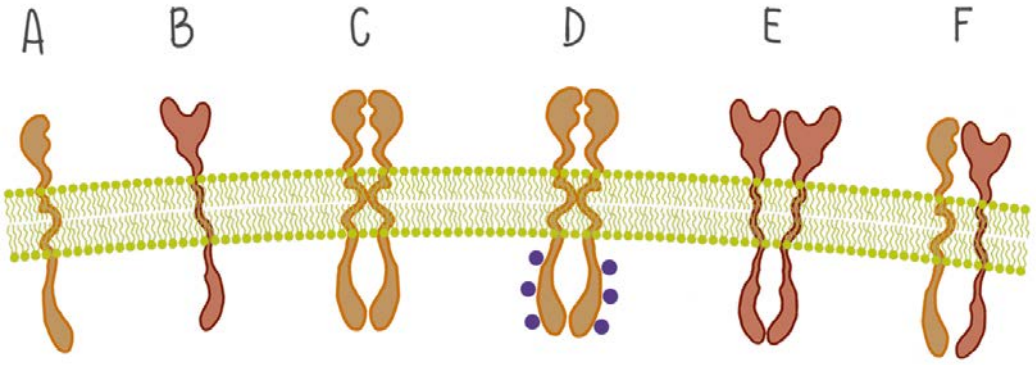


FIGURE 2.8: Membrane receptors when there is dimerization in the absence of a ligand. Receptors can be in several conformations: monomers (A and B), inactive (C) and active (D) R2 homodimers, inactive R1 homodimers (E) and inactive R1-R2 heterodimers (F).

In order to predict the predimer composition at the membrane in this new scheme, where dimers are (pre)formed before the onset of signaling, we propose the following model.

$$\frac{dR_1}{dt} = \mu_1 - 2k_{1+}R_1^2 - k_{12+}R_1R_2 + 2k_{1-}D_1 + k_{12-}D_{12} - \delta_1 R_1 \tag{2.41}$$

$$\frac{dR_2}{dt} = \mu_2 - 2k_{2+}R_2^2 - k_{12+}R_1R_2 + 2k_{2-}D_2 + k_{12-}D_{12} - \delta_2 R_2 \tag{2.42}$$

$$\frac{dD_1}{dt} = k_{1+}R_1^2 - k_{1-}D_1 - \Delta_1 D_1 \tag{2.43}$$

$$\frac{dD_{12}}{dt} = k_{12+}R_1R_2 - k_{12-}D_{12} - \Delta_{12}D_{12} \tag{2.44}$$

$$\frac{dD_2}{dt} = k_{2+}R_2^2 - k_{2-}D_2 - \Delta_2 D_2 - a_+ D_2 + a_- D_2^* \tag{2.45}$$

$$\frac{dD_2^*}{dt} = a_+ D_2 - a_- D_2^* \Delta_{2^*} \tag{2.46}$$

The equilibrium state for this system is

$$-\mu_1 + 2k_{1+}R_1^2\left(1 - \frac{k_{1-}}{(k_{1-} + \Delta_1)}\right) + k_{12+}R_1R_2\left(1 - \frac{k_{12-}}{(k_{12-} + \Delta_{12})}\right) + \delta_1 R_1 = 0 \tag{2.47}$$

$$-\mu_2 + R_2\left(2k_{2+}R_2\left(1 - \frac{k_{2-}}{(k_{2-} + \Delta_2 - \frac{a_+\Delta_{2^*}}{a_- + \Delta_{2^*}})}\right) + k_{12+}R_1\left(1 - \frac{k_{12-}}{(k_{12-} + \Delta_{12})}\right) + \delta_2\right) = 0 \tag{2.48}$$

$$D_1 = \frac{k_{1+}}{(k_{1-} + \Delta_1)}R_1^2 \tag{2.49}$$

$$D_{12} = \frac{k_{12+}}{(k_{12-} + \Delta_{12})}R_1R_2 \tag{2.50}$$

$$D_2 = \frac{k_{2+}R_2^2 + a_-D_2^*}{(k_{2-} + a_+ + \Delta_2)} \tag{2.51}$$

$$D_2^* = \frac{a_+ D_2}{a_- + \Delta_{2^*}} \tag{2.52}$$

Because the analytical analysis of this model is quite involved and the analytical steady state expressions are too complex to be informative, we proceed to consider an alternative probabilistic approach to estimate the expected number of dimers of each type. We will use the deterministic model introduce in Eqs. (2.41-2.46) to check the accuracy of the probabilistic approach.

If we consider that the binding reactions are very fast compared to the other processes affecting the system, then we expect the amount of dimers of each type to be proportional to the probability of picking the two components of a dimer from a pool of receptors of the two types, if the binding affinities for homo and heterodimers are the same [46]. Let R_{1T} be the total amount of ErbB1 and R_{2T}

the total amount of ErbB2. For large receptor populations, so that the probability of picking a given receptor type can be considered constant, we have

$$p(D_1) = \frac{1}{2}p(R_{1T})^2 = \frac{1}{2}\left(\frac{R_{1T}}{R_{1T}+R_{2T}}\right)^2, \tag{2.53}$$

$$p(D_{12}) = p(R_{1T})p(R_{2T}) = \frac{R_{1T}}{R_{1T}+R_{2T}}\frac{R_{2T}}{R_{1T}+R_{2T}}, \tag{2.54}$$

$$p(D_2) = \frac{1}{2}p(R_{2T})^2 = \frac{1}{2}\left(\frac{R_{2T}}{R_{1T}+R_{2T}}\right)^2. \tag{2.55}$$

where $p(D_1)$, $p(D_{12})$ and $p(D_2)$ are the probabilities of selecting a D_1, a D_{12} or a D_2 dimer, respectively.

From Eqs. (2.53-2.55), we can compute the expected number of dimers of each type (by multiplying the probabilities by the total number of receptors) as

$$\langle D_1\rangle = \frac{1}{2}\frac{R_{1T}^2}{R_{1T}+R_{2T}}, \tag{2.56}$$

$$\langle D_{12}\rangle = p(R_{1T})p(R_{2T}) = \frac{R_{1T}R_{2T}}{R_{1T}+R_{2T}}, \tag{2.57}$$

$$\langle D_2\rangle = \frac{1}{2}p(R_{2T})^2 = \frac{1}{2}\frac{R_{2T}^2}{R_{1T}+R_{2T}}. \tag{2.58}$$

The total amount of receptors R_{1T} and R_{2T} can be expressed as a function of the synthesis and degradation rate as $R_{1T} = \frac{\mu_1}{\delta_1}$ and $R_{2T} = \frac{\mu_2}{\delta_2}$, which yields

$$\langle D_1\rangle = \frac{1}{2}\frac{\delta_1}{\delta_2}\frac{\mu_1^2}{\mu_1\delta_2+\mu_2\delta_1}, \tag{2.59}$$

$$\langle D_{12}\rangle = \frac{\mu_1\mu_2}{\mu_1\delta_2+\mu_2\delta_1}, \tag{2.60}$$

$$\langle D_2\rangle = \frac{1}{2}\frac{\delta_2}{\delta_1}\frac{\mu_2^2}{\mu_1\delta_2+\mu_2\delta_1}. \tag{2.61}$$

In Fig. 2.10, we show the amount of dimers of each type as a function of the ErbB2 synthesis rate, μ_2, as given by Eqs. (2.59-2.61). In order to assess the accuracy of our approximation, we also compare our analytical expression with the numerical solution of the steady state (Eqs. 2.47-2.52).

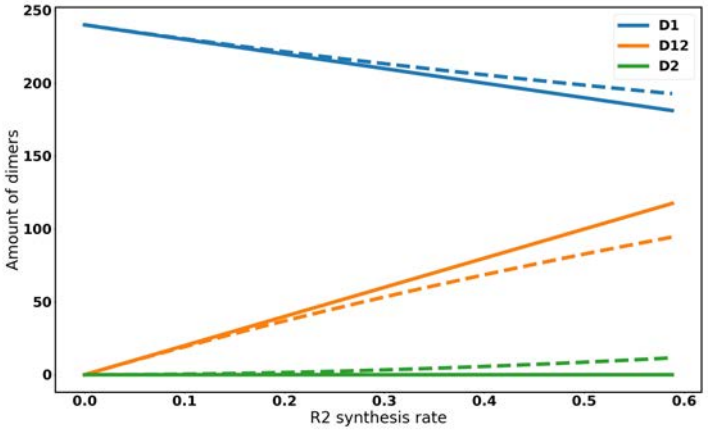


FIGURE 2.9: Steady state of the different predimers for different values of μ_2. Solid lines show the results of the numerical integration of the system described in Eqs. (2.41-2.46) at the steady state. Dashed lines show the theoretical prediction described by Eqs. (2.59-2.61). The theoretical prediction is accurate for low values of the synthesis rate (remind the ErbB2 synthesis rate has been estimated to be 0.12 min^{-1}).

As μ_2 increases, the amount of $D2$ homodimers and D_{12} heterodimers increases at the expense of a decrease in the amount of D_1 homodimers.

Using Eqs. (2.59-2.61), we can estimate the membrane dimer composition. Upon cell stimulation by a ligand, the ligand molecules bind and activate the predimers. We consider that the probability of the ligand binding a dimer is proportional to the number of binding sites for the ligand in the dimer. Because D_1 dimers have two binding sites for the ligand whereas D_{12} has only one binding site for the ligand (and assuming that there is no interaction between the ligand binding sites in a dimer so that the proportionality constant is 1), we take the probability of binding a D_1 dimer (β_1) as twice the probability of binding and activation a D_{12} dimer ($\beta_{12} = \frac{1}{2}\beta_1$). We take the probability of activation of a ligand bound dimer as 1. The probability of D_2 activation (α_2) is proportional to the equilibrium constant of the activation/inactivation reaction and does not depend on ligand stimulation.

The expected amount of active D_1, D_{12} and D_2 dimers under those assumptions is

$$\langle D_1^* \rangle = L_0 \frac{R_{1T}}{R_{1T} + R_{2T}} \tag{2.62}$$

$$\langle D_{12}^* \rangle = L_0 \frac{R_{2T}}{R_{1T} + R_{2T}} \tag{2.63}$$

$$\langle D_2^* \rangle = \alpha_2 \frac{1}{2} \frac{R_{2T}^2}{R_{1T} + R_{2T}} \tag{2.64}$$

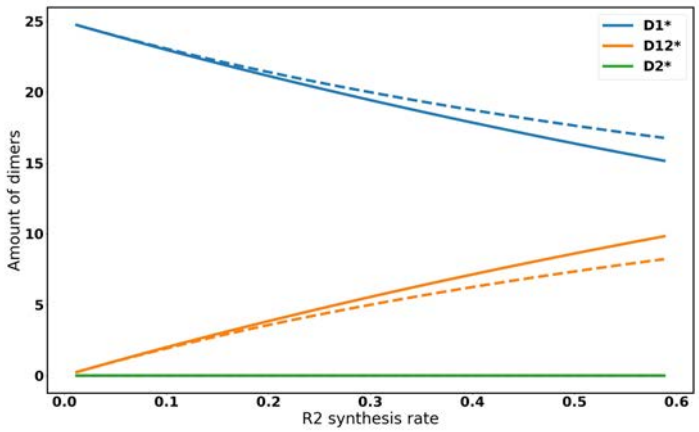


FIGURE 2.10: Amount of the different active dimers for different values of μ_2. Solid lines show the results of the numerical integration of the system described in Eq.2.41 to 2.46 at the steady state. Dashed lines show the theoretical prediction described by Eq.2.62 to 2.64. The theoretical prediction is accurate for low values of the synthesis rate (remind the ErbB2 synthesis rate has been estimated to be 0.12 min^{-1}).

2.3 Summary

In this chapter, we have analyzed the dynamics of the ErbB system under different assumptions. First, we analyzed the dynamics of the ErbB2 receptors, which displays a dynamics between different monomeric and dimeric forms even in the absence of cell stimulation by a ligand. This dynamic behavior of ErbB2 is relevant for cell signaling inasmuch the initial amount of monomeric and dimeric ErbB2 receptor is crucial to determine the amount of dimers of each type that are formed upon ligand stimulation. Next, we analyzed the dynamics of two different dimerization mechanisms proposed in the literature: ligand induced dimerization (Fig. 2.11) and predimerization (Fig. 2.12). Those two mechanisms yield a qualitatively similar dimer activation (increasing ErbB2 causes a decrease in ErbB1 homodimers

and an increase in ErbB1-ErbB2 heterodimers and ErbB2 homodimers), though there are quantitative differences. It is specially noteworthy the fact that in the case of predimerization, there is a greater range of the synthesis rate of ErbB2 for which ErbB2 homodimers is negligible, and also ErbB1 homodimers are predominant. This is due to heterodimer formation prior to cell stimulation, which decreases the amount of ErbB2 available for homodimerization. This is relevant in information transmission, as we will study in Chapter 4.

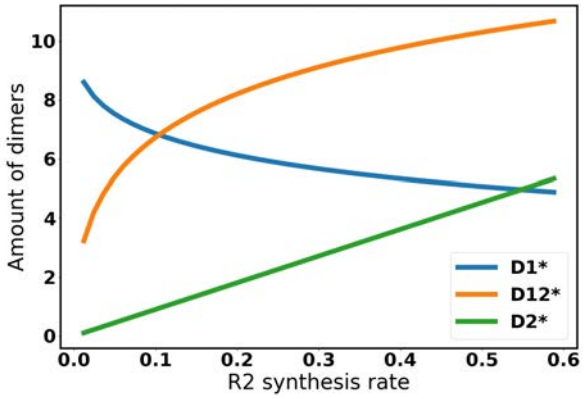


FIGURE 2.11: Amount of active dimers as a function of μ_2 for ligand induced dimerization.

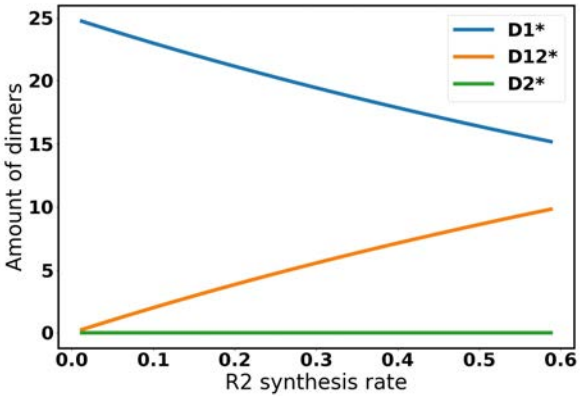


FIGURE 2.12: Amount of active dimers as a function of μ_2 for predimerization.

Chapter 3

Intracellular interactions

In Chapter 1, we learned that tyrosine kinase receptors phosphorilate tyrosine residues in their intracellular domain upon dimerization. These phosphotyrosines act as docking points for many intracellular proteins which, upon binding, get activated by the kinase activity of the receptors. A signaling cascade starts with these events that will ultimately cause a response of the cell to the stimulus.

The identity and the amount of the intracellular proteins which bind the active receptors is thought to determine the cellular response to extracellular information. Which proteins (and in which amount) bind which phosphotyrosine is determined by the aminoacid composition of the sites close to the tyrosine [25] and the likelihood and strength of the interaction is quantified by the equilibrium constant or the dissociation constant (see Chapter 1).

Jones *et al.* [23] measured the dissociation constant of several proteins to the tyrosines of the intracellular part of every member of the ErbB receptor family. We use their estimates in order to give a proxy of the intracellular state during the first steps of the signaling cascade.

Receptor activation, as we saw in Chapter 2, occurs after receptor dimerization. It is a slow process relative to the binding reactions, but slower than the degradation or internalization. As can be seen in Fig. 2.5, c) and d), the activation is relatively fast compared to internalization/degradation. Once the receptors are active, their concentration decays exponentially. Such exponential decay will be relevant for the intracellular protein dynamics, as we will see in this Chapter.

3.1 Probability distributions of the intracellular proteins: instantaneous measurement.

The values reported in the literature for the binding and unbinding constants of intracellular proteins are $\sim 10^{-1}$ nM^{-1} min^{-1} and ~ 0.1 min^{-1}, respectively [37]. The binding rate also depends on the protein concentration, which has been reported to be between 100 and 1000 nM for signaling molecules [79]. Therefore the binding rate is ~ 10 nM^{-1}min^{-1}. Taking into account the estimates of the parameter values reported in Table 2.1, we consider this to be fast compared to the activation, and especially compared to the inactivation and degradation rates of the receptors and proteins, so we take the steady state approximation for the intracellular binding and unbinding reactions: we consider that the proteins reach their steady state immediately compared to the dynamics of the active receptors (see Appendix, Section B.1).

We model earliest events of the intracellular signaling cascade triggered by receptor activation by considering that every binding site can be found in several states: free or occupied by any of the proteins that have a significant affinity for it.

We start our analysis by considering that we have only one type of intracellular protein. Then, a binding site can be in either of two states: free or bound. For many of such binding sites, the fact that it is free or bound can be modelled as a Bernouilli process, and the probability of finding x sites in a given state (free or bound) follows a binomial distribution.

For every binding site b_i and every protein P_j in a population of n_T receptors, the probability of n_{ij} sites in a bound state is

$$p(n_{ij} = k) = \frac{n_{iT}!}{(n_{iT} - k)!k!} \left(\frac{1 + A_i - K_{ij}}{1 + A_i} \right)^{n_{iT}-k} \left(\frac{K_{ij}}{1 + A_i} \right)^{k} \tag{3.1}$$

where K_{ij} is the affinity of the protein for the binding site and $A_i = \sum_{l=1}^{n_P} K_{il}$.

Because we are interested in the total number of proteins of each type that have bound an active receptor, we want to calculate $p(n_{*j})$.

$$p(n_{*j} = X) = p\left(\left(\sum_{i=1}^{n_G} n_{ij}\right) = X\right) \tag{3.2}$$

where n_{*j} is the sum of the proteins that have bound any binding site, n_B is the number of different binding sites and each n_{ij} is binomially distributed. Under appropriate conditions regarding the values of the parameters (large n and $p(=\frac{K_{ij}}{1+A_i})$ not close to 0 or 1), we can approximate the binomial distribution by a gaussian distribution according to

$$B(n,p) \sim G(\mu = np, \sigma^2 = np(1-p)) \tag{3.3}$$

where $n = n_{iT}$ and $p = \frac{K_{ij}}{1+A_i}$. Now, gaussian variables with different parameters can be added together just by adding the first and second moments characterizing the distribution: let X and Y be two independent normal distributions $X \sim G(\mu_x, \sigma_x^2)$ and $Y \sim G(\mu_y, \sigma_y^2)$, then $U = X+Y$ is $U \sim G(\mu_u = \mu_x+\mu_y, \sigma_u^2 = \sigma_x^2+\sigma_y^2)$ [80].

In our case, with two types of receptors with n_1 and n_2 binding sites respectively, the distribution of bound protein to any of the binding sites follows

$$X_j \sim G\left(\mu_x = \sum_{i\in n1,n2} n_{iT}\frac{K_{ij}}{1+A_i}, \sigma_x^2 = \sum_{i\in n1,n2} n_{iT}\frac{K_{ij}(1+A_i-K_{ij})}{(1+A_i)^2}\right) \tag{3.4}$$

Fig. 3.1 shows the comparison between the theoretical prediction by Eq. 3.4 and stochastic Gillespie simulations for the probability distribution of the number of bound proteins of a single type to active receptors.

3.1.1 Several proteins: multinomial distribution

Since the intracellular proteins are not independent -they are correlated through competition for the receptors-, we continue by considering two proteins, X_1 and X_2, that compete for the binding sites of receptors of one type.

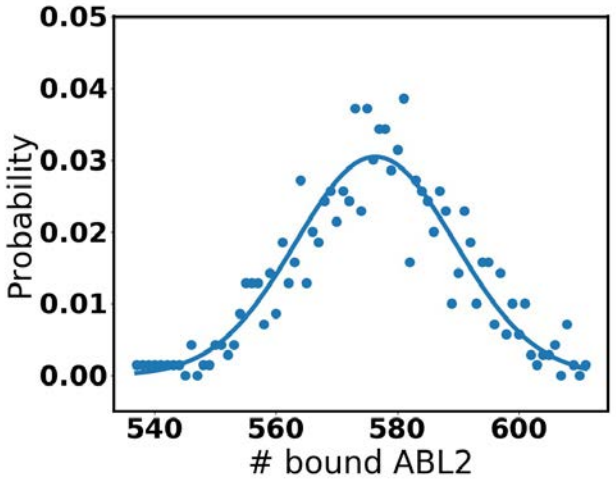


FIGURE 3.1: Probability distribution of bound intracellular protein. The solid line shows the theoretical prediction by the gaussian approximation of binomial distributions for each binding sites, the dots line show the results of stochastic Gillespie simulations of the binding and unbinding of the intracellular protein. p_1 is calculated from the affinities in [23] for ABL2.

The distribution of bound proteins of both types to one binding site is a multinomial (see Appendix, Section B.2):

$$p(X_1, X_2) = \frac{n!}{x_1!x_2!(n-x_1-x_2)!}\left(\frac{k_1}{1+k_1+k_2}\right)^{x_1}\left(\frac{k_2}{1+k_1+k_2}\right)^{x_2}\left(\frac{1}{1+k_1+k_2}\right)^{(n-x_1-x_2)} \tag{3.5}$$

We can approximate this multinomial by a multivariate normal (when n is large and $p_1 = \frac{k_1}{1+k_1+k_2}$ and $p_2 = \frac{k_2}{1+k_1+k_2}$ are not close to 0 or 1, see Appendix, Section A.1), using the normal approximation to a multinomial with parameters $\mu = \binom{np_1}{np_2}$ and $\Sigma = \binom{np_1(1-p_1) \quad -np_1p_2}{-np_1p_2 \quad np_2(1-p_2)}$. Here, n is the number of active receptors and $p_1 = \frac{k_i}{1+\sum_{j\in P} k_j}$, where k_i is the affinity of protein i to the binding site, and P is the set of considered proteins.

Once we have the multivariate normal distribution for every binding site, we can find the distribution of total protein of each type bound to the receptor by summing the multivariate normal distribution's parameters [80]. Then, the parameters are of the sum are

$$\mu_T = \begin{pmatrix} \sum_{bs} np_{1,bs} \\ \sum_{bs} np_{2,bs} \end{pmatrix} \tag{3.6}$$

$$\Sigma_T = \begin{pmatrix} \sum_{bs} np_{1,bs}(1-p_{1,bs}) & -\sum_{bs} np_{1,bs}p_{2,bs} \\ -\sum_{bs} np_{1,bs}p_{2,bs} & \sum_{bs} np_{2,bs}(1-p_{2,bs}) \end{pmatrix} \tag{3.7}$$

Fig. 3.3 shows the comparison between stochastic Gillespie simulations and the multivariate normal approximation of the number of bound proteins to active receptors when there is competition between protein types for the binding sites. In each row, we show pairs of competitors, chosen among the proteins considered more relevant in [23]. As expected, the average number of bound proteins depends on the affinity of both proteins to the binding sites.

3.2 Intracellular interactions during receptor decay

Consider the long time behavior of the system (see Fig. 2.5, d). The receptors are activated at a very fast scale, followed by a relatively slow exponential decay of the active dimers, D^*, with decay rate δ: $D^*(t) = D_0^* e^{-\delta t}$ (see Chapter 2, Fig. 2.5, d). During the period where the dimers are active, they can bind and activate intracellular proteins. The binding occurs following mass action kinetics (i.e. is proportional to the amount of reactant, $k_+ = k_{on}P$, where P is the amount of intracellular protein), and the activation of intracellular proteins occurs at rate a. If we consider that the proteins are in excess with respect to the binding sites, then the binding events can be considered to happen at constant rate, independent of the amount of intracellular proteins that have already interacted. The unbinding rate of the complex is k_-.

The probability of n binding events during the dimer's active period (T) is

$$p(n) = \int_0^{+\infty} p(n|T)p(T)dT, \tag{3.8}$$

where $p(n|T)$ is the probability of n binding events in an interval T and $p(T)$ is the probability distribution of the dimer's active time.

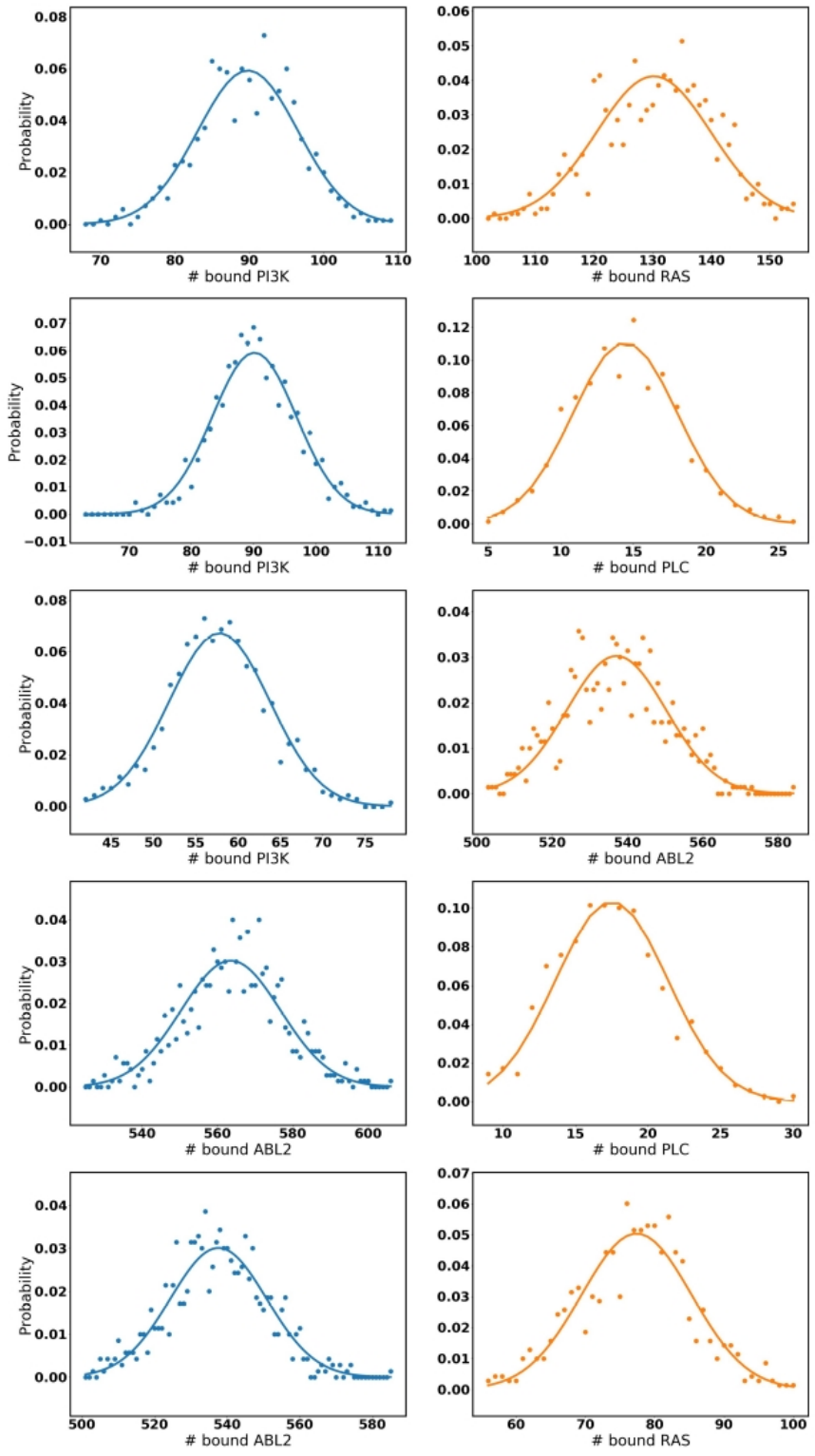


FIGURE 3.2: Probability distributions of pairs of proteins (one pair per row) that compete for the binding sites of active ErbB1, with $R1 = 100$. Solid lines show the predictions made with a multivariate normal distribution, dots show the results of stochastic Gillespie simulations of the binding and unbinding of two intracellular proteins. The mean and variance of the multivariate normal distributions are calculated from the affinity values given in [23].

3.2.1 One intracellular protein type

Consider first the simplest case: binding events to a free site are considered instantaneous, unbinding is a Poisson process with rate k_-. The instantaneous binding approximation is valid when the binding rate is very high, and this can happen because of a high affinity interaction between proteins and the active receptor, or alternatively because the concentration of the intracellular proteins is high.

Let's consider the number of unbinding events, n_U. In this scheme, $p(n_U|T)$ is given by

$$p(n_U|T) = \frac{(k_-T)^{n_U} e^{-k_-T}}{n_U!} \tag{3.9}$$

Using this expression in Eq. (3.8) and the fact that the decay of the active receptor is also a Poisson process (with rate δ), we get

$$p(n_U) = \int_0^{+\infty} \frac{(k_-T)^{n_U} e^{-k_-T}}{n_U!} \delta e^{-\delta T} dT \tag{3.10}$$

After changing variables and using the Gamma's function definition (see Appendix, Section A.4), we arrive to the following expression for $p(n_U)$

$$p(n_U) = \frac{k_-^{n_U}\delta}{(k_- + \delta)^{n_U+1}} \tag{3.11}$$

which is a geometric distribution ($p(n_U) = p^{n_U}(1-p)$) with parameter $p = \frac{k_-}{k_-+\delta}$. The average number and variance of unbinding events is therefore given by:

$$\langle n_U \rangle = \frac{k_-}{\delta} \tag{3.12}$$

$$\sigma_n^2 = \frac{k_-}{\delta^2}(k_- + \delta) \tag{3.13}$$

For an unbinding event to take place, a binding event must have happened before (a protein needs to be bound). After the last unbinding, since we consider the receptor in excess, there is an instantaneous binding. Therefore, the expected

number of binding events (and therefore the number of bound proteins) is $\langle n \rangle = \langle n_U \rangle + 1$.

Consider now a slightly more complicated situation: the binding events need not be instantaneous. The equations describing this process are

$$\frac{dP_F}{dt} = -k_+ P_F + k_- P_B \tag{3.14}$$

$$\frac{dP_B}{dt} = -k_- P_B + k_+ P_F \tag{3.15}$$

where P_F and P_B are the probabilities of the binding site being free and occupied with the protein, respectively. We can solve the equilibrium equations

$$0 = -k_+ P_F + k_- P_B \tag{3.16}$$

$$0 = -k_- P_B + k_+ P_F \tag{3.17}$$

to get the stationary probability of each state (given that $P_F + P_B = 1$)

$$P_F = \frac{k_-}{k_+ + k_-} \tag{3.18}$$

$$P_B = \frac{k_+}{k_+ + k_-} \tag{3.19}$$

Knowing the probability P_F of the binding site being free, we can now calculate the effective binding rate of each protein to the binding site (binding can only occur if the binding site is free)

$$\lambda_{in} = k_+ P_F = \frac{k_+ k_-}{k_+ + k_-} \tag{3.20}$$

Therefore the binding and unbinding can be described as a Poisson process with parameter $\lambda_{in} = \frac{k_+ k_-}{k_+ + k_-}$.

Following the same procedure than in the previous case, we find

$$p(n) = \frac{\lambda_{in}^n \delta}{(\lambda_{in} + \delta)^{n+1}} \tag{3.21}$$

The expected number and variance of binding events is

$$\langle n \rangle = \frac{\lambda_{in}}{\delta} = \frac{k_+ k_-}{\delta(k_+ + k_-)} \tag{3.22}$$

$$\sigma_n^2 = \frac{k_+ k_-}{\delta(k_+ + k_-)} \left(1 + \frac{k_+ k_-}{\delta(k_+ + k_-)}\right) \tag{3.23}$$

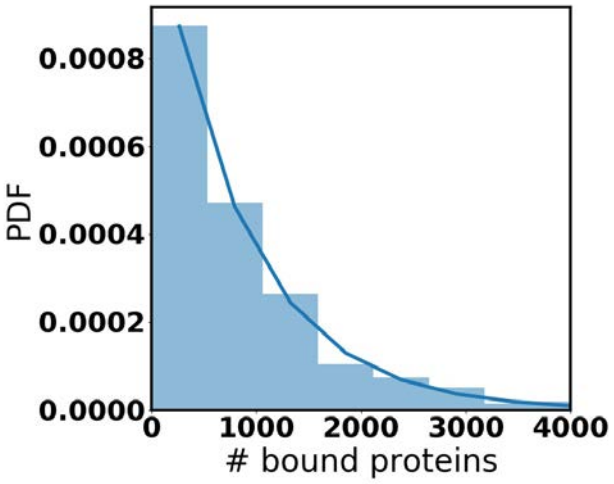


FIGURE 3.3: Probability distribution of the number of bound protein through the lifetime of an active receptor. ($\delta = 10^{-3}$ min^{-1}, $k_+ = 0.1$ nM^{-1}min^{-1}, $P = 500$ nM and $k_- = 1$ min^{-1}).

Consider now that not all binding events lead to activation, as activation is another Poisson process, with parameter α. Only the binding events that last long enough to be activated will yield an active protein. The probability of such event to occur is given by

$$p(t_a < t_u) = \frac{\alpha}{\alpha + k_-} \tag{3.24}$$

where t_a and t_u are the time elapsed before and activation or an unbinding event happen, respectively. This result allows us to calculate the average number of proteins that have been activated by a receptor while it was active as the product of Eqs. (3.22) and (3.24).

$$\langle a \rangle = \frac{k_+ k_-}{\delta(k_+ + k_-)} \frac{\alpha}{\alpha + k_-} \tag{3.25}$$

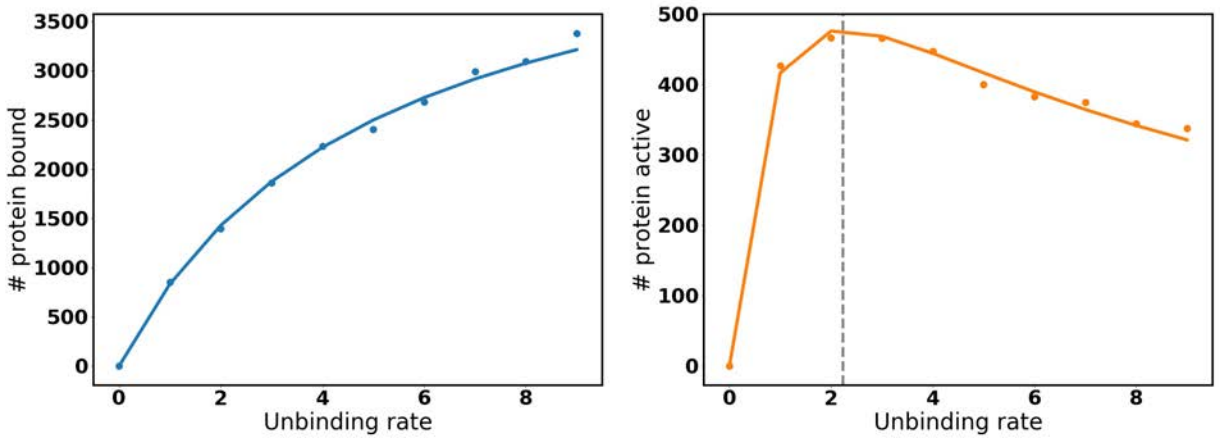


FIGURE 3.4: Expected number of bound (left) and active (right) proteins as a function of the unbinding rate. Solid lines are the results from Eqs. (3.22) and (3.25), respectively. Dots represent Gillespie simulations of the system. The dashed vertical line indicates the value $k_- = \sqrt{\frac{k_+}{\alpha}}$. The binding rate is $k_+ = 0.01$ min^{-1}. The degradation and activation rates are $\delta = 0.001$ min^{-1} and $\alpha = 1$ min^{-1} and the concentration of intracellular protein is 500 nM.

Fig. 3.5 shows the amount of proteins activated by an active receptor during its lifetime. We can see that the amount of activated proteins depends in a non-monotonous way on the unbinding rate.

If the number of active proteins is a quantity to be maximized, we can calculate the maximum of $\langle a \rangle$ with respect to any of the parameters. k_- seems to be subject to a trade-off: if it is too big, then the protein unbinds before activation; if it is too small, then it occupies the binding site so other proteins cannot bind. The maximum number of active proteins is achieved for k_- such as

$$\frac{\partial \langle a \rangle}{\partial k_-} = \frac{k_+ \alpha}{\delta} \frac{(k_+ \alpha - k_-^2)}{(k_+ + k_-)^2 (\alpha + k_-)^2} = 0 \tag{3.26}$$

which yields $k_- = \sqrt{\frac{k_+}{\alpha}}$.

3.2.2 Two intracellular protein types: competition for the binding site

Consider the situation of one binding site that can be bound by two proteins (p_1 and p_2) with affinities $K_1 = \frac{k_{1+}}{k_{1-}}$ and $K_2 = \frac{k_{2+}}{k_{2-}}$, respectively.

The binding of any of the two proteins is a stochastic process that can be described as two independent Poisson processes in a queue: the proteins can bind whenever the binding site is free, with rate k_{i+}, and they stay bound for an average time of $\frac{1}{k_{i-}}$.

The Master Equation describing such a process are

$$\frac{dP_F}{dt} = (-k_{1+} - k_{2+})P_F + k_{1-}P_1 + k_{2-}P_2 \tag{3.27}$$

$$\frac{dP_1}{dt} = -k_{1-}P_1 + k_{1+}P_F \tag{3.28}$$

$$\frac{dP_2}{dt} = -k_{2-}P_2 + k_{2+}P_F \tag{3.29}$$

where P_F, P_1 and P_2 are the probabilities of the binding site being free, occupied with protein p_1 and occupied with protein p_2, respectively. We can solve the equilibrium equations

$$0 = (-k_{1+} - k_{2+})P_F + k_{1-}P_1 + k_{2-}P_2 \tag{3.30}$$

$$0 = -k_{1-}P_1 + k_{1+}P_F \tag{3.31}$$

$$0 = -k_{2-}P_2 + k_{2+}P_F \tag{3.32}$$

to get the stationary probability of each state

$$P_F = \frac{1}{1 + K_1 + K_2} \tag{3.33}$$

$$P_1 = \frac{k_{1+} + k_{2+}}{k_{2-}(\frac{k_{1+}}{k_{2+}} + 1)} P_F \tag{3.34}$$

$$P_2 = \frac{k_{1+} + k_{2+}}{k_{1-}(\frac{k_{2+}}{k_{1+}} + 1)} P_F \tag{3.35}$$

Knowing the probability P_F of the binding site being free, we can now calculate the effective binding rate of each protein to the binding site (binding can only occur if the binding site is free)

$$\kappa_{i+} = k_{i+} P_F = \frac{k_{i+}}{1 + K_1 + K_2} \tag{3.36}$$

The probability of n binding events in a given period of time T, $P(n|T)$ for this Poisson process (for protein p_i) is

$$P(n_i|T) = \frac{(\kappa_{i+}T)^{n_i} e^{\kappa_{i+}T}}{n_i!} \tag{3.37}$$

Let's introduce now the fact that the binding site has a finite lifetime: it shows an exponential decay with rate δ. The expected lifetime of a binding site is

$$p(T) = \delta e^{-\delta T} \tag{3.38}$$

We can calculate the probability of n binding events during the lifetime of a binding site by solving

$$P_{n_i} = \int_0^{+\infty} \frac{(\kappa_{i+}T)^{n_i} e^{\kappa_{i+}T}}{n_i!} \delta e^{-\delta T} dT \tag{3.39}$$

This yields a geometric distribution

$$P_{n_i} = \frac{\kappa_{i+}^{n_i} \delta}{(\kappa_{i+} + \delta)^{n_i+1}} \tag{3.40}$$

with the following average and variance

$$\langle n_i \rangle = \frac{\kappa_{i+}}{\delta} = \frac{k_{i+}}{\delta(1+K_1+K_2)} \tag{3.41}$$

$$\sigma^2_{n_i} = \frac{k_{i+}}{\delta(1+K_1+K_2)}\left(1+\frac{k_{i+}}{\delta(1+K_1+K_2)}\right) \tag{3.42}$$

The Fano factor of the PDF in Eq. (3.40) is

$$F = \frac{\sigma^2_{n_i}}{\langle n_i \rangle} = 1 + \frac{k_{i+}}{\delta(1+K_1+K_2)} \tag{3.43}$$

According to Eq. (3.43), the presence of a competitor for the binding site reduces the Fano factor of the distribution, i.e. reduces the dispersion) with respect to a situation without competition (where the Fano factor is $F = 1 + \frac{k_{i+}}{\delta(1+K_1)}$).

Proteins that have bound a binding site can be activated. Activation is also a Poisson process, with rate α. A bound protein gets active if the activation happened before unbinding (Poisson process with rate k_{i-}). The probability of activation happening before unbinding is

$$P_{Ai} = \frac{\alpha}{\alpha + k_{i-}} \tag{3.44}$$

Then, the expected number of *active* proteins of type i is

$$\langle a_i \rangle = \frac{k_{i+}}{\delta(1+K_1+K_2)} \frac{\alpha}{\alpha + k_{i-}} \tag{3.45}$$

3.2.3 Distribution of bound proteins for several binding sites and receptors

One protein

We derived in the previous section that the number of proteins that have bound an active binding site before its degradation follows a geometric distribution:

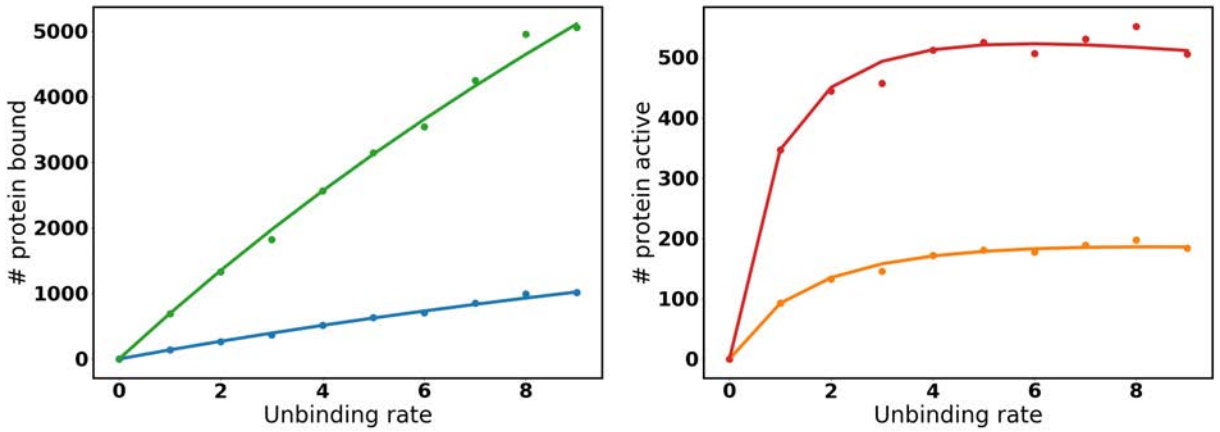


FIGURE 3.5: Expected number of bound (left) and active (right) proteins as a function of the unbinding rate for two competing proteins. Solid lines are the results from Eqs. (3.41) and (3.45), respectively. Dots represent Gillespie simulations of the system. The binding rates are $k_{1+} = 0.05$ and $k_{2+} = 0.01$ min^{-1}, respectively. The degradation and activation rates are $\delta = 0.001$ min^{-1} and $\alpha = 1$ min^{-1} and the concentration of intracellular protein is 500 nM.

$$X \sim Geo\left(\frac{k_+ k_-}{\delta(k_+ + k_-) + k_+ k_-}\right). \tag{3.46}$$

In a binary experiment (i.e. where the output can be 0 or 1), a geometric distribution describes the number of successes that are drawn before a failure. In the case under consideration, a binding event represents a success, while degradation of the receptor represents a failure. If we consider r receptors, then we are interested in the number of successes (i.e. number of bound proteins X_S) before r failures. This is described by a negative binomial distribution (see Appendix, Section A.1):

$$X_S \sim NB\left(r, \frac{k_+ k_-}{\delta(k_+ + k_-) + k_+ k_-}\right), \tag{3.47}$$

where r is the number of copies of the binding site.

The negative binomial distribution can be approximated by a gaussian ($NB(r,p) \sim G(\mu = r\frac{1-p}{p}, \sigma^2 = r\frac{1-p}{p^2}$ if r is sufficiently large or p not very small [81]. We approximate the number of bound proteins to r binding sites by a normal distribution. We then use the fact that normal distributions can be added by adding the respective means and variances [80] to estimate the total amount of protein that has bound an active receptor.

$$X_S \sim G\left(\mu = r\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}, \sigma^2 = r\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}\left(1+\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}\right)\right). \tag{3.48}$$

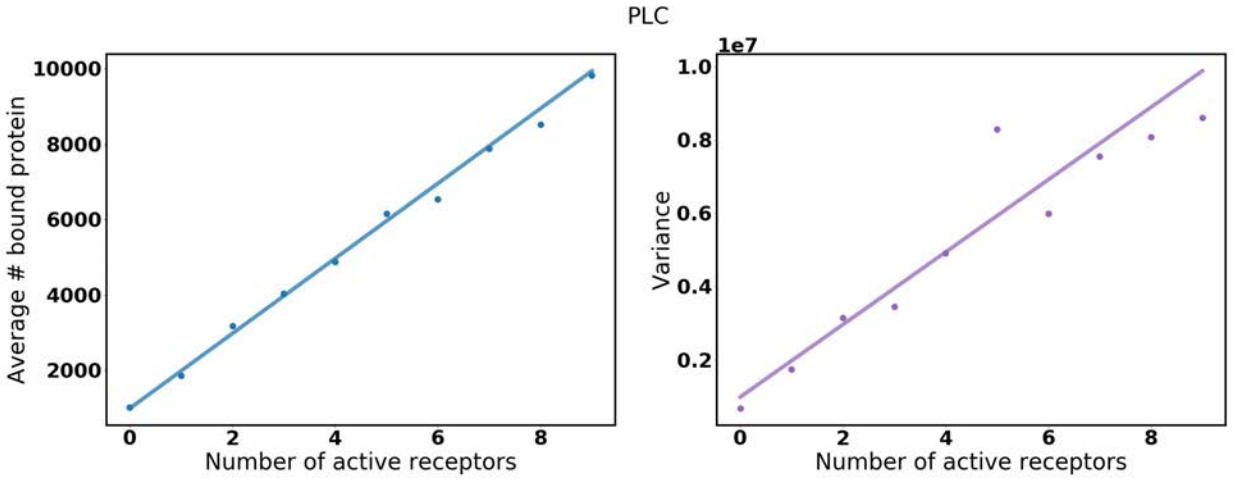


FIGURE 3.6: Mean number (left) and variance (right) of RAS proteins having bound an active receptor of type 1 through its lifetime as a function of the initial number of active receptors. Dots show the results of stochastic Gillespie simulations, solid lines show the average and variance, μ and σ^2 of the gaussian approximation described in Eq (3.48). ($\delta = 5 \cdot 10^{-2}$, $k_+ = 0.1$ and the k_- are determined by the experimental K_d as $k_- = K_d k_+$ given in [23]).

Two proteins

The generalisation to two proteins is straightforward. The distribution of two competing proteins having bound an active receptor during its lifetime is

$$X_1 \sim Geo\left(\frac{k_{1+}}{\delta(1+K_1+K_2)+k_{1+}}\right) \tag{3.49}$$

As in the previous section, when there are r copies of the same binding site, the distribution that describes the number of bound proteins is a negative binomial

$$X_{1,S} \sim NB\left(n, \frac{k_{1+}}{\delta(1+K_1+K_2)+k_{1+}}\right), \tag{3.50}$$

which can be approximated by a gaussian. The distribution for every binding site can be added, yielding a new gaussian describing the number of proteins that have bound an active receptors at any binding site:

$$X_{1,S} \sim G\left(\mu = n\sum_{bs}\frac{k_{1+,bs}}{\delta(1+K_{1,bs}+K_{2,bs})}, \sigma^2 = n\sum_{bs}\frac{k_{1+,bs}}{\delta(1+K_{1,bs}+K_{2,bs})}(1+\frac{k_{1+,bs}}{\delta(1+K_{1,bs}+K_{2,bs})})\right). \tag{3.51}$$

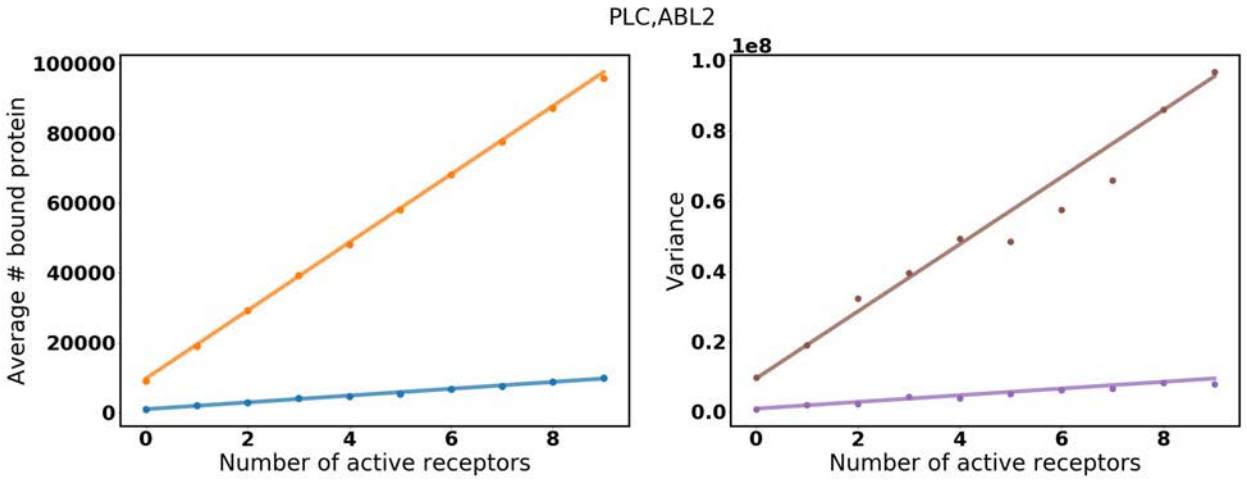


FIGURE 3.7: Average number (left) and variance (right) of bound proteins when two proteins (PLC and ABL2) are competing for the active receptors. Dots show the results of stochastic Gillespie simulations, solid lines show the average and variance, μ and σ^2 of the gaussian approximation described in Eq (3.51). ($\delta = 5 \cdot 10^{-2}$, $k_+ = 0.1$ and the k_- are determined by the experimental K_d as $k_- = K_d k_+$ given in [23]).

Figs. 3.6 and 3.7 show the comparison between the approximation made in Eqs. 3.48 and 3.51, respectively, and the results of Gillespie stochastic simulations.

3.3 Effects of competition and specialization

The binding sites in the ErbB receptors are limited compared to the number of different proteins that can bind the active receptors: ErbB1 and ErbB2 have 11 and 7 binding sites, respectively, while the number of different proteins they can bind is over 60 for both ErbB1 and ErbB2 [23].

Although not all the proteins are present in the cell at the same time (the collection of proteins expressed in the cell at a given time depends on the phenotype and other factors), it is reasonable to expect a certain degree of competition for the binding sites of the receptors, specially taking into account that most of the binding sites have few interaction partners, while others show a great diversity in the repertoire of proteins they can bind. Moreover, the affinity spectrum of the proteins for

the different ErbB binding sites shows that most of the proteins can interact with several binding sites, hinting to an underlying heterogeneous and complex interaction network [23].

The structure of this interaction network is governed by structural and chemical complementarity of the interacting parts, and this has been shaped by evolution to give rise to functional organisms.

Linking evolutionary preassures as drivers of optimization processes of the function and structure of organisms is a delicate subject, specially in complex organisms where trade-offs between different features are frequent and determining the quantity that is being optimized is a particularly difficult task. Nonetheless, the dependency of certain traits on the features of the systems can be addressed (without necessarily claiming it is subject to optimization), and this gives an insight into the factors that determine the function and structure of organisms. In this Section, we want to study how protein binding processes are affected by the structure of the protein-receptor interactions when several proteins are expressed at the same time.

With this aim in mind, we propose a toy model where two binding sites (A and B) can be bound by two proteins (p_1 and p_2). We consider that proteins are subject to structural and chemical trade-offs regarding to their interaction with their binding partners [82, 83]. Specifically, here we consider that increased complementarity for one binding site is achieved at the expense of a decrease in the complementarity of other binding sites. The rationale for this claim comes from the fact that the affinity of a protein for a binding site depends on the structural and chemical complementarity between the interacting partners [84].

We model this trade-off with a parameter, $\theta \in [0, 1]$, which indicates the degree of preference for a binding site, or equivalently, how the affinity is allocated among the binding sites. Values of θ close to 0 or 1 indicate a strong preference for one of the binding sites, while $\theta \sim 0.5$ indicates that the protein does not show a preference for any binding site. Let K_A and K_B be the affinities for binding sites A and B, respectively. Then, in our model, $K_A = \theta K_T$ and $K_B = (1-\theta)K_T$, where K_T be the total protein's affinity to be allocated between both binding sites.

With those ingredients, we can calculate the amount of proteins p_1 and p_2 that have been activated by an active receptor during its lifetime (Eq. (3.45)), and also the time that the binding receptors have been occupied.

The total amount of active protein of type i is (supposing that the protein is excess):

$$\langle a_i \rangle = \langle a_{iA} \rangle + \langle a_{iB} \rangle = \frac{\alpha}{\alpha + k_{i-}} \left(\frac{\theta_i k_{i+}}{\delta(1 + \theta_i K_i + \theta_j K_j)} + \frac{(1-\theta_i) k_{i+}}{\delta(1 + (1-\theta_i) K_i + (1-\theta_j) K_j)} \right) \tag{3.52}$$

The expected time that a binding site is occupied, τ_O can be calculated for both scenarios as the product of the probability of the binding site being occupied, $(1 - P_F)$ (Eq. (3.33)), times the expected lifetime of the binding site, $\tau = \frac{1}{\delta}$

$$\langle \tau_O \rangle = \frac{1}{\delta} \left(1 - \frac{1}{1 + \theta_i K_i + \theta_j K_j} \right) \tag{3.53}$$

We can now compare analytically two different cases:

- p_1 and p_2 show the same affinity sharing ($\theta_1 = \theta_2 = \theta$), so that:
  $K_{1A,eff} = K_1\theta$, $K_{1B,eff} = K_1(1-\theta)$ and $K_{2A,eff} = K_2\theta$, $K_{2B,eff} = K_2(1-\theta)$
- p_1 and p show opposite affinity sharing ($\theta_1 = 1 - \theta_2 = \theta$, so that:
  $K_{1A,eff} = K_1\theta$, $K_{1B,eff} = K_1(1-\theta)$ and $K_{2A,eff} = K_2(1-\theta)$, $K_{2B,eff} = K_2\theta$

Consider the first case ($\theta_i = \theta_j$):

$$\begin{aligned} \langle a_i \rangle &= \frac{\alpha}{\alpha + k_{i-}} \left(\frac{\theta_i k_{i+}}{\delta(1 + \theta_i K_i + \theta_i K_j)} + \frac{(1-\theta_i) k_{i+}}{\delta(1 + (1-\theta_i) K_i + (1-\theta_i) K_j)} \right) \\ &= \frac{k_{i+}\alpha}{\delta(\alpha + k_{i-})} \frac{1 + 2(K_i + K_j)\theta(1-\theta)}{1 + (K_i + K_j) + (K_i + K_j)^2 \theta(1-\theta)} \end{aligned} \tag{3.54}$$

The expected time that the binding sites are occupied is

$$\begin{aligned} \langle \tau_{O,c} \rangle &= \langle \tau_{O,A,c} \rangle + \langle \tau_{O,B,c} \rangle \\ &= \frac{1}{\delta} \left(1 - \frac{1}{1 + \theta K_i + \theta K_j} \right) + \frac{1}{\delta} \left(1 - \frac{1}{1 + (1-\theta) K_i + (1-\theta) K_j} \right) \end{aligned} \tag{3.55}$$

$$\langle \tau_{O,c} \rangle = \frac{1}{\delta} \frac{2 + (K_1 + K_2)}{1 + (K_1 + K_2) + \theta(1-\theta)(K_1 + K_2)^2} \tag{3.56}$$

This situation would correspond to strong competition between p_1 and p_2.

We model specialization by considering two proteins with dissimilar preferences (whenever p_i has a strong preference for binding site A, then p_j prefers binding site B), such as $\theta_i = 1 - \theta_j$. Then, the amount of active proteins is

$$\begin{aligned} \langle a_i \rangle &= \frac{\alpha}{\alpha + k_{i-}} \left(\frac{\theta_i k_{i+}}{\delta(1 + \theta_i K_i + (1-\theta_i) K_j)} + \frac{(1-\theta_i) k_{i+}}{\delta(1 + (1-\theta_i) K_i + \theta_i K_j)} \right) \\ &= \frac{k_{i+}\alpha}{\delta(\alpha + k_{i-})} \frac{1 + K_j + 2(K_i - K_j)\theta(1-\theta)}{1 + K_i K_j + (K_i + K_j) + (K_i - K_j)^2 \theta(1-\theta)} \end{aligned} \tag{3.57}$$

The expected free time is

$$\begin{aligned} \langle \tau_{O,s} \rangle &= \langle \tau_{O,A,s} \rangle + \langle \tau_{O,B,s} \rangle \\ &= \frac{1}{\delta} \left(1 - \frac{1}{1 + \theta K_i + (1-\theta) K_j} \right) + \frac{1}{\delta} \left(1 - \frac{1}{1 + (1-\theta) K_i + \theta K_j} \right) \end{aligned} \tag{3.58}$$

$$\langle \tau_{O,s} \rangle = \frac{1}{\delta} \frac{2 + (K_1 + K_2)}{1 + (K_1 + K_2) + \theta(1-\theta)(K_1 - K_2)^2 + K_1 K_2} \tag{3.59}$$

By examining Eqs. (3.56) and (3.59), we realize that comparing $\tau_{O,c}$ and $\tau_{O,s}$ is equivalent to comparing $2\theta(1-\theta)$ and $1 - 2\theta(1-\theta)$ for competition and specialization, respectively.

We realize that $2\theta(1-\theta) = 1 - 2\theta(1-\theta) = 0.5$ for $\theta = 0.5$, but this is a maximum for $2\theta(1-\theta)$ (competition) and a minimum for $1 - 2\theta(1-\theta)$ (specialization). Thus, specialization maximizes the time that the binding sites are occupied, specially for higher values of θ (more specialized proteins). But, when competition is forced upon intracellular proteins due to the limited availability of binding sites, then the condition that maximizes the time that the binding sites are occupied is $\theta = 0.5$, which means that the proteins do not show a preference for any binding site. The amount of time that binding sites are occupied correlates with a higher number of total protein in the system.

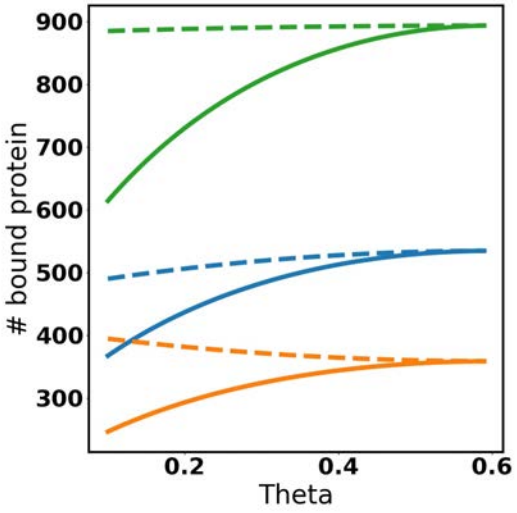


FIGURE 3.8: Amount of bound proteins as a function of θ. $k_+ = 0.1$ min^{-1}, Ptot=500, $K_{d1} = 500$ and $K_{d2} = 750$ nM^{-1}, $\delta = 10^{-3}$ min^{-1}. Blue is p_1, orange is p_2, green is the total, dashed is specialization, solid is competition.

In Fig. 3.8 we show the amount of each protein and the total amount of active protein as a function of the degree of specialization for the two cases considered (competition and specialization). For the competition scenario, the affinity distribution that maximizes the amount of active proteins is achieved for $\theta = 0.5$, which also maximizes the amount of time that the binding site are occupied. In the specialization case, the optimal θ is different for each competitors: the good competitor (bigger K) benefits from sharing its affinity equally between both binding sites, while the bad competitor (smaller K) benefits of specializing (so that it can bind a binding site free of competition).

The two cases considered above are limiting cases of a wider possibility space. In Fig. 3.9, we show the amount of active proteins and the time bound for different combinations of θ_1 and θ_2, where red and orange lines represent competition and specialization, respectively, as considered above.

Let us now consider the case of more than two binding sites. One way of quantifying the sharing of the affinity between the receptors' binding sites is by measuring the entropy of the affinity for all the binding sites a given protein can bind, H_n:

$$H_n = -\sum_i^n A_i \log A_i, \tag{3.60}$$

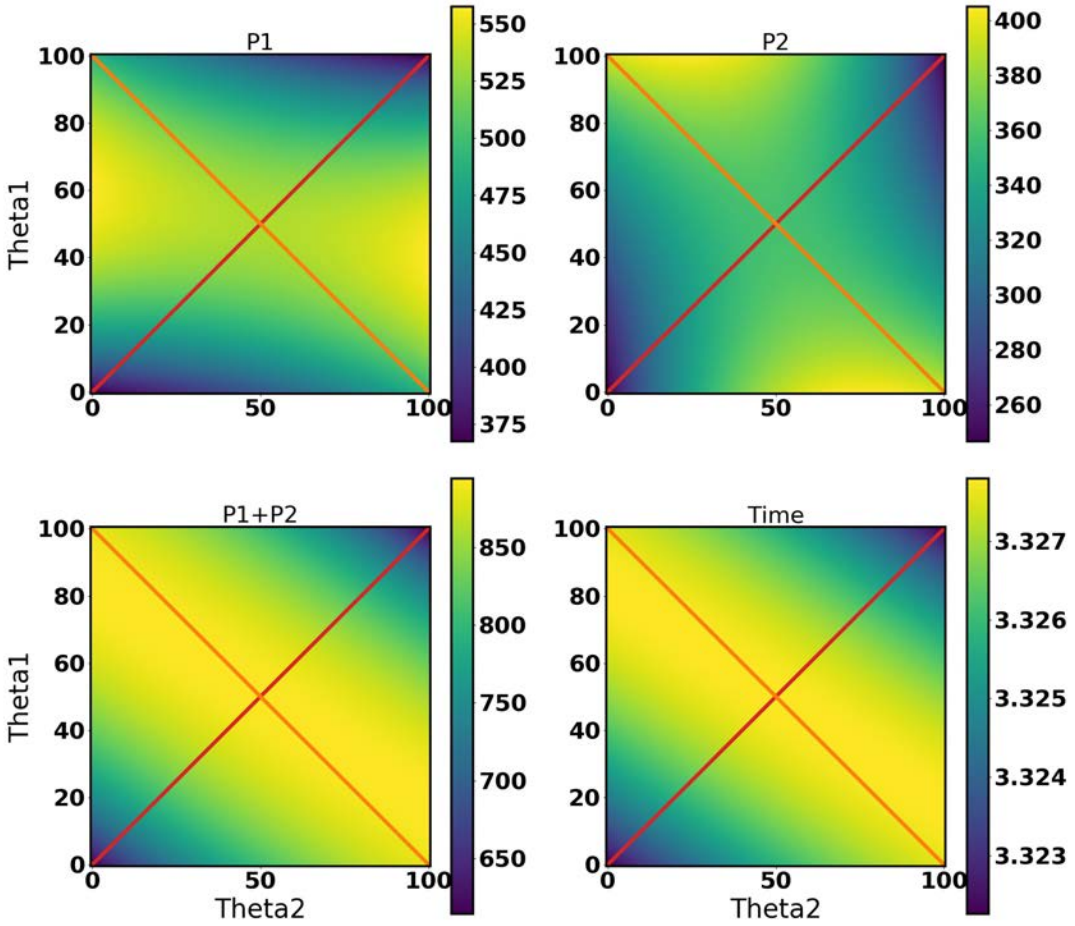


FIGURE 3.9: Amount of bound proteins (titles *P1, P2, P1+P2*) and bound time (title *Time*) of the receptor as a function of θ_1 and θ_2. $k_+ = 0.1$ min^{-1}, P$_{tot}$=500 nM, $K_{d1} = K_{d2} = 500$ nM^{-1}, $\delta = 10^{-3}$ min^{-1}. Red line shows competition, orange line shows specialization

where $A_i = \frac{K_i}{1+\sum_j^n K_j}$ is the affinity of a binding site normalised by the total affinity of the protein and the subindex n indicates the number of binding sites that the protein can bind.

In our toy model, we can calculate the number of active proteins as a function of the entropy in the affinities of each competitor (see Fig. 3.10) in order to compare the result with experimental data.

Jones *et al.* [23] published data on the affinity of a collection of proteins binding to ErbB receptors. We use this dataset in order to study the proteins' affinity distribution among binding sites. With this aim, we calculate the entropy of each protein's affinities for the different binding sites. Because the entropy depends on the number of states n (number of binding sites) as $\log n$, we normalise the

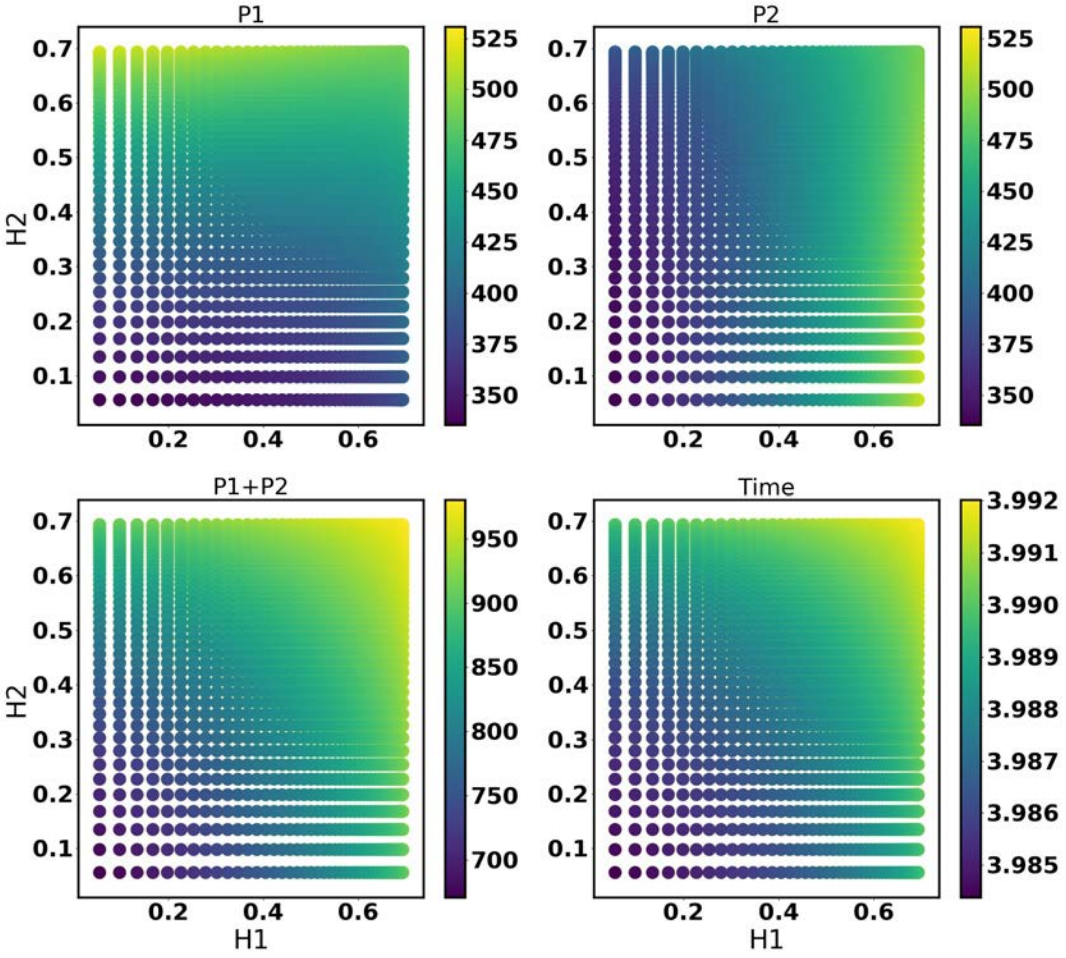


FIGURE 3.10: Amount of bound proteins (titles *P1, P2, P1+P2*) and bound time (title *Time*) as a function of H_n of each protein in a pair of competitors. $k_+ = 0.1$ min^{-1}, P$_{tot}$=500 nM, $K_{d1} = K_{d2} = 500$ nM^{-1}, $\delta = 10^{-3}$ min^{-1}

calculated entropy as $H = \frac{H_n}{\log n}$. For every pair of proteins in the dataset, we represent the pair in the space of $\frac{H_n}{\log n}$, the normalised entropy of the affinities for the different binding sites (Fig. 3.11), if the two proteins compete for at least one binding site. Each pair is then represented as a point in this space, whose coordinates correspond to the normalised affinity entropy of each member of the pair.

The entropy of real interaction partners is such that it yields a higher number of active proteins, except for certain cases, where it maximizes the amount of one or the other competitor. The reason for this could lay in the physico-chemical features of the binding site and the intracellular proteins, and/or it may have evolved in such a way to maximize the amount of proteins that are activated through ErbB membrane receptors. Such a selective preassure could arise from the fact that

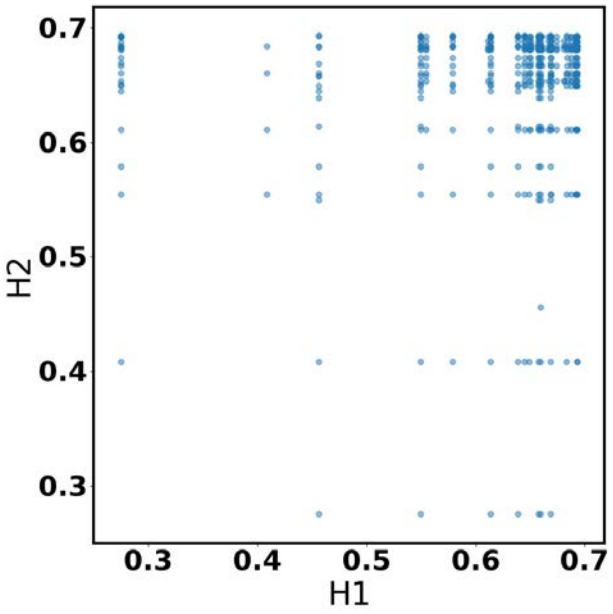


FIGURE 3.11: Proteins in [23], represented in the space of the normalised entropy of the affinities for the different binding sites. Each dot represents a pair of competitors, and its coordinates indicate the normalised affinity entropy of each member of the pair.

synthesizing intracellular proteins has a metabolic cost that would render useless if such proteins are not active in the signaling process.

Chapter 4

Information transmission through the ErbB system

In Chapter 1, we have given details as to how cells need to sense extracellular information in order to adjust their behaviour to their (micro)environment. Extracellular information needs to be transmitted across the cell membrane in order to reach the cytosol and the nucleus, where most of the cell decisions are taken. Membrane receptors serve such a purpose, acting as an information transmission channel. However, a) membrane receptors have their own dynamics at the membrane and within the cell, b) they interact with other receptors in a nonlinear way, c) they present stochasticity (due to diffusion at the membrane and to low copy numbers), and d) ligand-receptor interactions are not exclusive: one ligand can bind several kinds of receptor and vice-versa, and there is crosstalk between receptors. All these issues make the analysis of information transmission through the membrane a complex issue [54].

In this chapter, we will deal with information transmission through the cell membrane using information theory. In order to make analytical progress, we will make a number of simplifying assumptions, the scope and validity of which will be tested by means of numerical simulations. In particular, we will only consider one ligand with a given distribution and the membrane composition of the different active receptors upon ligand binding. We will derive an expression for the mutual information bewteen the input (ligand) distribution and the output (active intracellular proteins) distribution for several situations, starting with simpler situations, and moving on to more complicated cases. We will start by calculating

the mutual information for a system with only one type of receptor (R_1, which has the same distribution as the ligand) and one intracellular protein. Then, a case with two types of receptors (R_1 and R_2) and one intracellular protein will be considered. This will allow us to address the effects of nonlinearities during information transmission. Next, we will study the situation where there is only one type of receptor (R_1) with two types of intracellular proteins in order to explore the effects of multiple proteins binding to the same binding sites.

4.1 One receptor, one protein

We start our analysis by studying the mutual information between the ligand and one intracellular protein for two different cases: when the detection of the activation state of the receptors is fast, '*snap shot*-like', so that the system measures the instantaneous state of the system (Section 4.1.1), and when the detection is slow enough to consider the total binding events that happened on an active receptor (Section 4.1.2). It has been shown in [63] that the '*snap shot*-like' sensing gives a lower bound to the information obtained from temporal trajectories. Since we are interested in the effects of the different molecular components and interactions in the mutual information, rather than in the actual value of the mutual information, and given that the expression for the '*snap shot*' mutual information is simpler than the mutual information over temporal trajectories, we chose the first for the analysis of the mutual information.

Let y and x be random variables accounting for the concentration of the ligand and an intracellular protein, respectively. We first consider information transmission through ErbB1 alone. At the activation peak, the amount of active receptor is proportional to the initial amount of ligand, L_0, so in the low noise limit in the interaction between the ligand and the receptors, we can consider that the distribution of active receptor is a linear function of the ligand concentration, αy. For simplicity, we consider $\alpha = 1$.

The mutual information between the input distribution (distribution of the ligand, $\mathcal{Y}$) and the output distribution (distribution of intracellular protein, $\mathcal{X}$) is defined as

$$I(X,Y) = \int dy \int dx p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{4.1}$$

which, if we use conditional probabilities, can be rewritten as

$$I(X,Y) = H(X|Y) - H_Y(X) \tag{4.2}$$

where $H(X|Y) = \int dy p(y) \int dx p(x|y) \log_2 p(x|y)$ and $H_Y(X) = \int dy p(y) \int dx p(x|y) \log_2 p(x)$.

In the following sections, we consider gaussian input distributions, $Y \sim G(\mu_y, \sigma_y)$, so that

$$p(Y = y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\frac{(y-\mu_y)^2}{\sigma_y^2}} \tag{4.3}$$

4.1.1 Mutual information at the activation peak

As derived in Chapter 3, Section 3.1, the amount of bound proteins to each binding site at a given time is binomially distributed,

$$p(X_{bs} = k) = \frac{n_{iT}!}{(n_{iT}-k)!k!} \left(\frac{1 + A_i - K_{ij}}{1 + A_i}\right)^{n_{iT}-k} \left(\frac{K_{ij}}{1 + A_i}\right)^{k} \tag{4.4}$$

with $A_i = \sum_{l=1}^{n_P} K_{il}$.

In Chapter 3, we have determined that the PDF of the bound intracellular proteins X_1 in Eq. (4.4) is well approximated by $X \sim G(\mu = n\sum_{bs} p_{1,bs}, \sigma^2 = n\sum_{bs} p_{1,bs}(1 - p_{1,bs}))$, with n being the amount of active receptors and $p_{1,bs} = \frac{K_{bs}}{1+A_{bs}}$, K_{bs} being the equilibrium constant of the protein for binding site bs and $A_{bs} = \sum_{i\in\mathcal{P}} K_i$ the sum of the equilibrium constants of all the proteins to binding site bs.

We are interested in the amount of bound proteins at the activation peak where the amount of active receptors is maximum (which is when all ligand bound receptors are active and none has degraded), hence $R^*_{1max} = Y$. Taking this into account, $H(X|Y)$ in Eq. 4.2 is given by:

$$H(X|Y) = -\log_2 \sqrt{2\pi e \sum_{bs} p_{1,bs}(1-p_{1,bs})} - \frac{1}{2}\langle \log_2 y \rangle. \tag{4.5}$$

The expected value of a function of a random variable can be calculated by means of the method of Taylor expansions around the mean (see Appendix, Section A.3). Applying the Taylor expansion method [85] to $\langle \log_2 y \rangle$, we can rewrite Eq. (4.5) as

$$H(X|Y) = -\log_2 \sqrt{2\pi e \sum_{bs} p_{1,bs}(1-p_{1,bs})} - \frac{1}{2}\log_2 \mu_y - \frac{1}{4}\frac{\sigma_y^2}{\mu_y^2}. \tag{4.6}$$

Consider the second term $H_Y(X)$. We do not have direct knowledge of $p(x)$ (we only know the conditional distribution, Eq. (4.4)), so we need to resort to estimating $\int dy p(y) p(x|y)$ in an indirect way. This can be estimated as an infinite weighted sum of gaussians, and therefore its mean and variance can be calculated using a gaussian mixture model [86]. The average, μ', and variance, $\sigma^{2\prime}$, of the resulting normal distribution are given by:

$$\mu' = \int dy p(y) \mu_x(y) = \int dy p(y) y \sum_{bs} p_{1,bs} = \mu_y \sum_{bs} p_{1,bs} \tag{4.7}$$

$$\begin{aligned} \sigma^{2\prime} &= \int dy p(y) \sigma_x(y)^2 + \int dy p(y) \mu_x(y)^2 - \left(\int dy p(y) \mu_x(y) \right)^2 \\ &= \mu_y \sum_{bs} p_{1,bs}(1-p_{1,bs}) + \sigma_y^2 \left(\sum_{bs} p_{1,bs} \right)^2 \end{aligned} \tag{4.8}$$

Estimating $p(x)$ as $p(x) = G(\mu', \sigma^{2\prime})$, we can calculate $H_Y(X)$:

$$\begin{aligned} H_Y(X) &= \int dy \int dx p(x,y) \log_2 \frac{1}{\sqrt{2\pi}\sigma'} e^{\frac{-(x-\mu')^2}{2\sigma'^2}} \\ &= -\log_2 \sqrt{2\pi}\sigma' - \frac{1}{2\sigma'^2} \int dy \int dx p(x_1, y)(x-\mu')^2 = -\log_2 \sqrt{2\pi e}\sigma' \end{aligned} \tag{4.9}$$

Then, we obtain an analytical expression for the mutual information defined in Eq. 4.2 by combining Eqs. (4.6) and (4.9):

$$I(X,Y) = \log_2 \sqrt{\frac{\sigma'^2}{\sum_{bs} p_{1,bs}(1-p_{1,bs})}} - \frac{1}{2}(\log_2 \mu_y - \frac{1}{2}\frac{\sigma_y^2}{\mu_y^2}) \tag{4.10}$$

$$I(X,Y) = \frac{1}{2}\log_2 \left(1 + \frac{\sigma_y^2}{\mu_y} \frac{\left(\sum_{bs} p_{1,bs}\right)^2}{\sum_{bs} p_{1,bs}(1-p_{1,bs})}\right) + \left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2 \tag{4.11}$$

In order to check the accuracy of the gaussian approximation to the PDF of the number of bound protein regarding the mutual information, we have performed Gillespie stochastic simulations [87] of a system consisting of one type of receptor and one type of protein (see Appendix, Section C.2 for details). The comparison between our analytical result, Eq. (4.11), and the mutual information calculated from our numerical results is shown in Fig. 4.1. We observe that the agreement between both is very good, except in the limit of very broad ligand distributions (large Fano factor of the input distribution).

Regarding the dependence of the mutual information on the Fano factor of the input distribution, $\frac{\sigma_y}{\mu_y}$, we observe that the higher the Fano factor of the input distribution, the higher the mutual information (this is confirmed by direct stochastic simulations). This could seem counter-intuitive, as noisier input distributions yield higher mutual information. However, simulation results confirm such behavior. By inspection of Eq. (4.2), we can see that the mutual information not only depends of the input distribution but also on the square of the signal to noise ratio of the output distribution: $SNR^2 = \left(\frac{\mu}{\sigma}\right)^2 = \frac{\left(\sum_{bs} p_{1,bs}\right)^2}{\sum_{bs} p_{1,bs}(1-p_{1,bs})}$. The higher the signal to noise ratio of the output distribution, the higher the mutual information. This is achieved for high binding probabilities of the protein to several binding sites in the active receptor.

Our interpretation of this property is that in order to achieve optimal mutual information, the input distribution should be wide, allowing for a wide range of ligand concentrations (y), and the output (bound proteins) distribution should be narrow around the actual ligand value detected.

Note that the values of the mutual information obtained here are low (generally less than 1 bit). This is consistent with experimental results which found values of the mutual information ~ 1 bit [62, 63, 65].

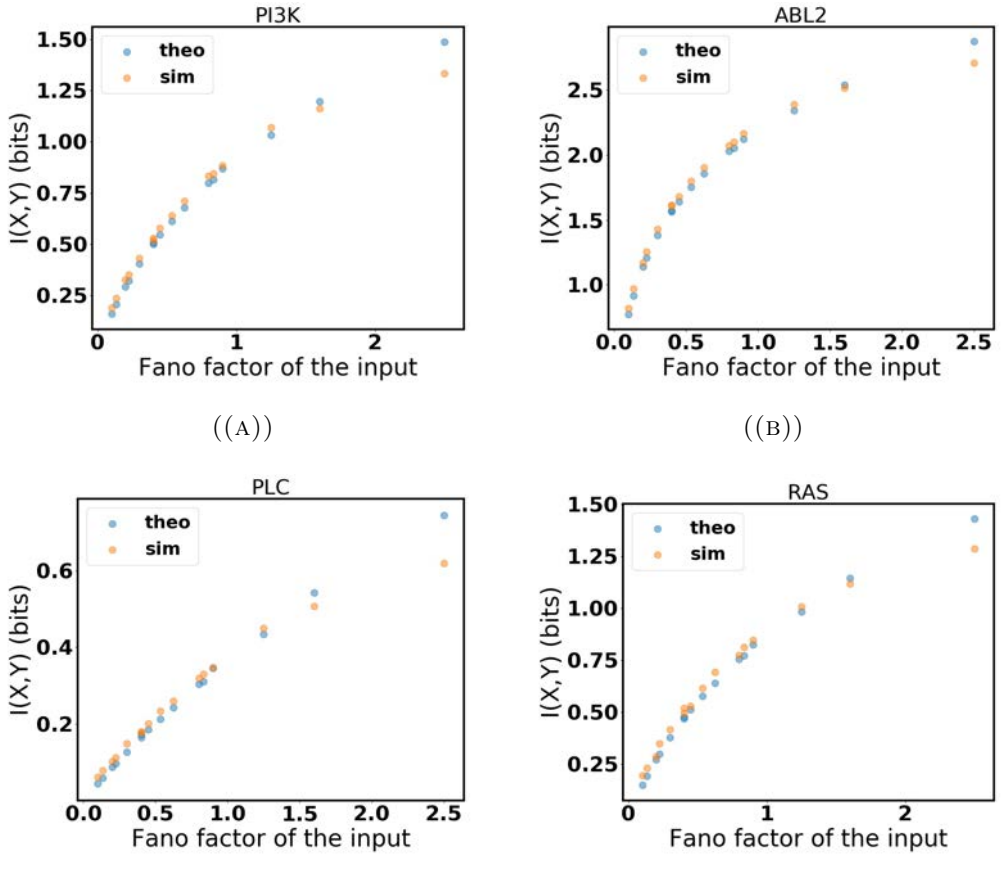


FIGURE 4.1: Mutual information as a function of the Fano factor of the input distribution for four different proteins (PI3K, ABL2, PLCγ and RAS). Blue dots represent the mutual information calculated with Eq. (4.11), orange dots are the results of Gillespie stochastic simulations. The binding probabilities p_{bs} are calculated with the values of the affinities given in [23].

4.1.2 Mutual information in the active period

It has been proposed that, measuring over an extended period of time can increase the mutual information through a biological signaling system [62]. In this section, we test whether this hypothesis applies to our system.

We will show here that whereas the analytical expression of our estimates of the mutual information is very similar to the one in the previous section, the signal to noise ratio of the output distribution (and therefore the mutual information) is rather different to the previous case.

In Chapter 3, section 3.2.3, we have shown that the number of proteins that have bound a binding site though its life time follows a geometric distribution,

and that summing over binding sites the intracellular protein distribution can be approximated as

$$X_S \sim G\left(\mu = r\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}, \sigma^2 = r\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}\left(1+\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}\right)\right) \tag{4.12}$$

where r is the number of active receptors at the onset of signaling. As before, we consider that the amount of active receptors is the same as the total amount of ligand presented to the cell, so $r \equiv y$.

The mutual information in this case reads

$$I(X,Y) = \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2}{\mu_y}\frac{\left(\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}\right)^2}{\sum_{bs}\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})}(1+\frac{k_{+,bs}k_{-,bs}}{\delta(k_{+,bs}+k_{-,bs})})}\right)+\left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2 \tag{4.13}$$

The dependence of the average and variance on y is the same as in section 4.1.1, so the functional form of the mutual information remains as in Eq. 4.11 but the coefficients change.

Instantaneous versus sustained sensing

In order to carry out a comparison between the mutual information with instantaneous versus active-period sensing, we consider a receptor and a protein which interact only through one binding site.

In this case, the mutual information with instantaneous sensing is

$$I(X,Y) = \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2}{\mu_y}\frac{\left(\frac{\frac{k_+}{k_-}}{1+\frac{k_+}{k_-}}\right)^2}{\frac{\frac{k_+}{k_-}}{1+\frac{k_+}{k_-}}\cdot\frac{1}{1+\frac{k_+}{k_-}}}\right)+\left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2 = \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2}{\mu_y}K_{eq}\right)+\left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2 \tag{4.14}$$

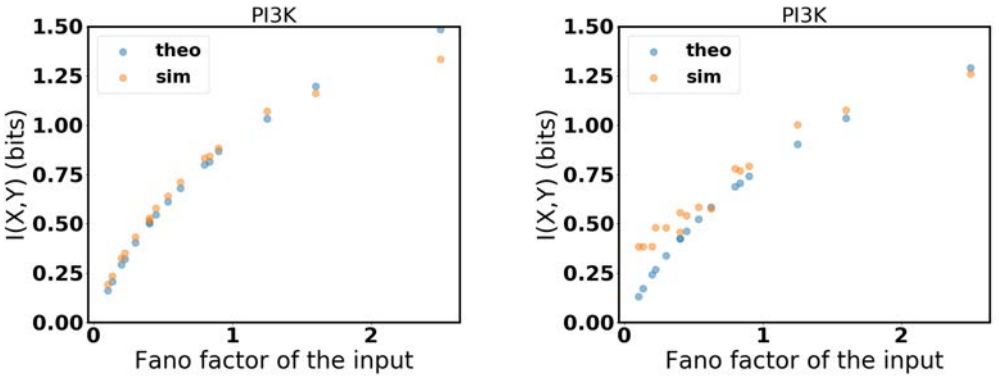
whereas for the active-period sensing, the mutual information is

$$
\begin{aligned}
I(X,Y) &= \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2}{\mu_y}\frac{\left(\frac{k_+k_-}{\delta(k_++k_-)}\right)^2}{\frac{k_+k_-}{\delta(k_++k_-)}(1+\frac{k_+k_-}{\delta(k_++k_-)})}\right)+\left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2 \\
&= \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2}{\mu_y}\frac{k_+k_-}{\delta(k_++k_-)+\frac{k_+k_-}{\delta(k_++k_-)}}\right)+\left(\frac{1}{2}\frac{\sigma_y}{\mu_y}\right)^2
\end{aligned}
\tag{4.15}
$$

Whenever $\frac{k_+k_-}{\delta(k_++k_-)+k_+k_-} > K_{eq}$, measuring over extended periods yields a higher mutual information. This is equivalent to

$$
k_-(k_- - k_+) > \delta(k_- + k_+) \tag{4.16}
$$

Eq. (4.2) holds if $k_- > k_+$ and $k_- > \delta\frac{1+K}{1-K}$. This is true for the parameter values reported in the literature and that we have used throughout this thesis (see Table 2.1).



((A)) ((B))

FIGURE 4.2: Mutual information as a function of the Fano factor of the ligand distribution for snapshot sensing (A) and active period sensing (B). Blue dots represent the mutual information calculated with Eq. (4.11) and Eq. (4.13), orange dots are the results of Gillespie stochastic simulations. Results are shown for PI3K. The binding probabilities p_{bs} are calculated with the values of the affinities given in [23].

In Fig. **??** we see that although fit between the mutual information predicted by Eq. (4.13) and the corresponding stochastic simulations is not as good as in the case of instantaneous sensing, it works fine for high Fano factor of the input distribution.

The mutual information in the active period sensing is lower than in instantaneous sensing. This is due to the faster increase in the variance of the output distribution in the active period case.

It has been hypothesized that intantaneous sensing gives a lower bound to the mutual information transmitted through a system ([63]), and that sensing over extended periods should yield higher values of the mutual information. This is true only in the case where the system is able to measure the temporal variation in the signal [88]. Our results suggest that measuring the total active proteins over a period results in a loss of mutual information due to a greater increase of the variance of the intracellular protein distribution (compared to 'snap-shot' signaling).

4.2 Two receptors, one protein: the effects of a nonlinear channel

We consider in this section the effects on mutual information of the nonlinearity in the information transmission channel associated with the interaction between different monomeric receptors to form homo and heterodimers.

As discussed in Chapter 2, different dimerization mechanisms have been described, namely, ligand induced dimerization [75] and predimerization [77]. Ligand-induced dimerization was the first to be described, but lately, with the advent of single molecule experiments, the existence of predimers in the membrane has gained support [78, 89]. In both cases, it remains to be established whether subsequent steps are determined by the amount of active monomers or active dimers, although there are studies supporting the notion that only one member of the dimer is activated upon ligand binding [90] and that the downstream activation levels are better explained by the amount of dimers of each type rather than the amount of receptors of each type in active dimers [48]. It has also been described that ErbB2 homodimers exist in the membrane in two dimeric configurations: active and inactive [76].

We start our analysis by deriving the mutual information for the predimerization mechanism. We then proceed to derive the expression for the mutual information for the ligand induced dimerization.

4.2.1 Mutual information when dimers are preformed

Consider now two types of receptors, R_1 and R_2, which form dimers. The random variable Y accounts for the ligand concentration, which follows a gaussian distribution $G(\mu_y, \sigma_y^2)$.

We consider here that activation within the dimers is directional, so that only one member of the dimer becomes active [90]. The relevant quantity in this case is not the amount of receptors that form dimers, but the amount of dimers itself [48]. We assume that ErbB2 is activated in the heterodimers, as proposed in [91].

In Chapter 2, section 2.2.2, we saw that the expected number of active dimers of each type can be approximated by

$$\langle D_1^* \rangle = L_0 \frac{R_{1T}}{R_{1T} + R_{2T}} \tag{4.17}$$

$$\langle D_{12}^* \rangle = L_0 \frac{R_{2T}}{R_{1T} + R_{2T}} \tag{4.18}$$

$$\langle D_2^* \rangle = \alpha_2 \frac{1}{2} \frac{R_{2T}^2}{R_{1T} + R_{2T}} \tag{4.19}$$

where $L_0 (\equiv y)$ is the concentration of ligand.

Because the PDF of the bound intracellular proteins can be approximated by the sum of the gaussians for every binding site (see Chapter 3, Section 3.1), the average and variance of the intracellular protein distribution can be expressed (considering only the informative dimers) in the following way

$$\langle x \rangle = n_1 S_1 + n_2 S_2 \tag{4.20}$$

$$\sigma_x^2 = n_1 S_{1-} + n_2 S_{2-}, \tag{4.21}$$

where $n_1 = \langle D_1^* \rangle$ is the number of active ErbB1 and $n_2 = \langle D_{12}^* \rangle + \langle D_2^* \rangle$ is the number of active ErbB2, $S_i \equiv \sum_{bs,1} p_{bs,i}$ and $S_{i-} \equiv \sum_{bs,i} p_{bs,i}(1 - p_{bs,i})$ for $i = 1, 2$ indicating the receptor type (ErbB1 and ErbB2, respectively) and $\sum_{bs}$ indicates the sum over binding sites. By substituing Eqs. (4.17-4.19) into Eqs. (4.20-4.21) we get

$$\langle x\rangle = y\frac{R_{1T}}{R_{1T}+R_{2T}}S_1 + y\frac{R_{2T}}{R_{1T}+R_{2T}}S_2 + \alpha_2\frac{1}{2}\frac{R_{2T}^2}{R_{1T}+R_{2T}}S_2 \tag{4.22}$$

$$\sigma_x^2 = y\frac{R_{1T}}{R_{1T}+R_{2T}}S_{1-} + y\frac{R_{2T}}{R_{1T}+R_{2T}}S_{2-} + \alpha_2\frac{1}{2}\frac{R_{2T}^2}{R_{1T}+R_{2T}}S_{2-} \tag{4.23}$$

Consider the entropy of the conditional distribution, $H(X|Y)$ (see Eq. (4.2)),

$$H(X|Y) = -\frac{1}{2}\log_2 2\pi e - \frac{1}{2}\int dyp(y)\log_2\left(y\left(\frac{R_1}{R_1+R_2}S_{1-} + \frac{R_2}{R_1+R_2}S_{2-}\right) + \frac{\alpha_2}{2}\frac{R_2^2}{R_1+R_2}S_{2-}\right) \tag{4.24}$$

where we take, for simplicity, $R_i \equiv R_{iT}$. This notation will be used throughout this Chapter.

In order to proceed further, we need to compute the variance of the output distribution $p(x)$ from the conditional distributions, $p(x|y)$, by integrating over y. As in section 4.1.1, we calculate the average μ' and variance σ'^2 of the resulting distribution by means of a mixture model [86].

$$\mu' = \int dyp(y)\mu_x(y) = \mu_y\left(\frac{R_1}{R_1+R_2}S_1 + \frac{R_2}{R_1+R_2}S_2\right) + \frac{\alpha_2}{2}\frac{R_2^2}{R_1+R_2}S_2 \tag{4.25}$$

$$\begin{aligned}\sigma^{2\prime} &= \int dyp(y)\sigma_x(y)^2 + \int dyp(y)\mu_x(y)^2 - \left(\int dyp(y)\mu_x(y)\right)^2 \\ &= \mu_y\left(\frac{R_1}{R_1+R_2}S_{1-} + \frac{R_2}{R_1+R_2}S_{2-}\right) + \frac{\alpha_2}{2}\frac{R_2^2}{R_1+R_2}S_{2-} + \sigma_y^2\left(\frac{R_1S_1+R_2S_2}{R_1+R_2}\right)^2\end{aligned} \tag{4.26}$$

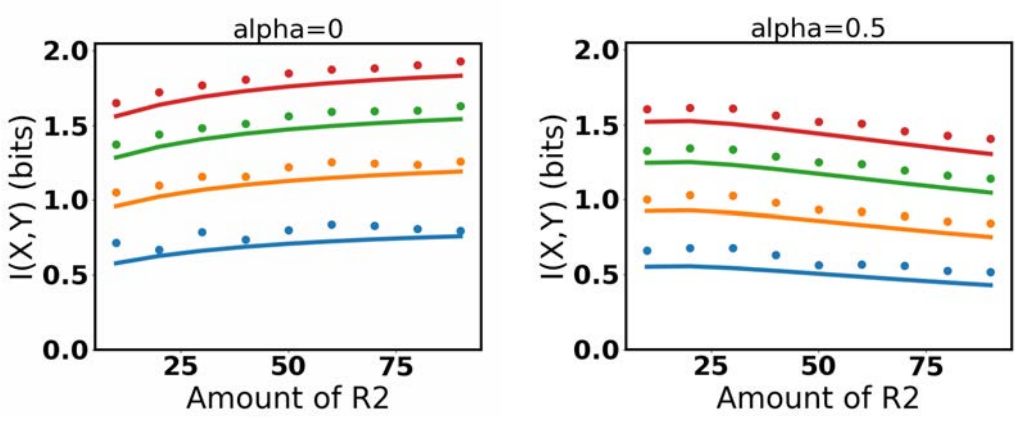
We give an estimate of the distribution of X as $G(\mu', \sigma^{2\prime})$. Using this distribution in the expression for $H_Y(X)$ we obtain

$$H_Y(X) = -\frac{1}{2}\log_2 2\pi e\left(\frac{\mu_y}{R_1+R_2}(R_1S_{1-}+R_2S_{2-}) + \frac{\alpha_2}{2}\frac{R_{2T}^2}{R_{1T}+R_{2T}}S_{2-} + \sigma_y^2\left(\frac{R_1S_1+R_2S_2}{R_1+R_2}\right)^2\right) \tag{4.27}$$

The mutual information, obtained by substracting $H_Y(X)$ from $H(X|Y)$, is

$$I(X,Y) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_y^2}{\mu_y}\frac{(R_1S_1+R_2S_2)^2}{(R_1+R_2)(R_1S_{1-}+R_2S_{2-}+\frac{\alpha_2}{2\mu_y}R_2^2)}\right) \tag{4.28}$$

In Fig. 4.3, we show the agreement between Eq. (4.28) and Gillespie stochastic simulations (see Appendix, Section C.2 for details on the simulations). ErbB2 homodimer activation reduces the mutual information transduced by the receptor system (lower valuer of $I(X,Y)$), and this effect becomes more pronounced as the amount of membrane ErbB2 receptors increases.



((A)) ((B))

FIGURE 4.3: Mutual information between the ligand and PI3K for different values of initial ErbB2 concentration and different averages of the ligand distribution ($\mu_y = 20$ (red), $\mu_y = 30$ (green), $\mu_y = 40$ (orange), $\mu_y = 50$ (blue)) when there is no activation of ErbB2 homodimers (A, $\alpha_2 = 0$) and when there is activation of ErbB2 homodimers (B, $\alpha_2 = 0.5$). The solid line was calculated according to Eq. 4.28, dots show the results of stochastic simulations. $R_1 = 50$ and the binding probabilities p_{bs} are calculated with the values of the affinities given in [23].

In order to study more thoroughly this effect, we show in Fig. 4.4 how the mutual information, Eq. (4.28) varies as we change the initial amount of ErbB2 receptors and the degree of ErbB2 homodimers activation, α_2. We explore the dependence of the mutual information on this two factors for four different combinations of the protein-binding site affinities in ErbB1 and ErbB2: low affinity for both receptors (a), high affinity for ErbB1 and low affinity for ErbB2 (b), low affinity for ErbB1 and high affinity for ErbB2 (c) and high affinity for both receptors (d). The effect of the activation of ErbB2 homodimers (increasing α_2) is detrimental in terms of mutual information for every situation considered. The effect of increasing the concentration of ErbB2 depends on the affinities of the intracellular protein for ErbB1 and ErbB2: if the protein binds preferentially to ErbB1 ((a) and (c)), then increasing the ErbB2 concentration decreases the value of the mutual information. Such effect is the consequence of the 'kidnapping' of ErbB1 by the formation of assymetric heterodimers, in which ErbB2 is activated by ErbB1. However, if the protein binds to ErbB2 preferentially ((b) and (d), then the mutual information

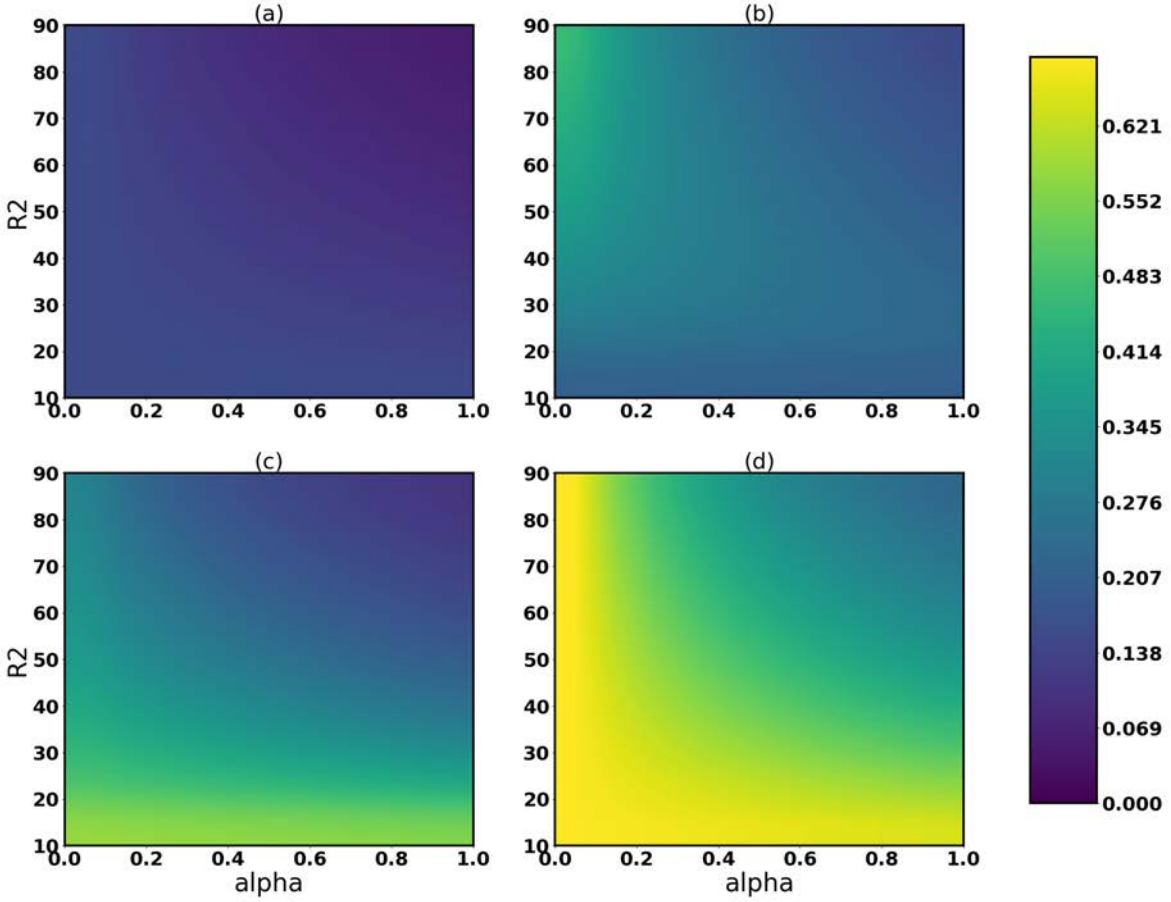


FIGURE 4.4: Mutual information as a function of the amount of ErbB2 receptors and the equilibrium constant for active ErbB2 homodimers, for different affinities of the intracellular proteins for active ErbB1 and ErbB2.

can be increased by increasing the concentration of the ErbB2 receptor, whenever α_2 remains low. For high α_2, increasing the amount of ErbB2 pronounces the effect of the formation of non informative ErbB2 homodimers and, as a consequence, the mutual information is reduced.

We conclude that the decrease of the mutual information by an overexpression of ErbB2 is affinity-dependent and it can happen through two different mechanisms:

- Activation of non informative ErbB2 homodimers (which increases as the concentration of ErbB2 increases [76]),
- 'Kidnapping' of ErbB1 in heterodimers (detrimental when the intracellular protein binds ErbB1 preferentially).

4.2.2 Mutual information for ligand-induced dimerization

When we consider ligand induced dimerization the average active ErbB1 homodimers and ErbB1-ErbB2 heterodimers can be estimated by (see Chapter 2, Section 2.2.1.2)

$$D_{1,max} = \frac{L_0}{2} - \frac{k_{12+}}{4k_{1+}} R_2 \log_2 \left(1 + \frac{2k_{1+}L_0}{k_{12+}R_2}\right) \tag{4.29}$$

$$D_{12,max} = \frac{k_{12+}}{2k_{1+}} R_2 \log_2 \left(1 + \frac{2k_{1+}L_0}{k_{12+}R_2}\right) \tag{4.30}$$

The average and the variance of the intracellular protein are (see Section 4.2.1)

$$\langle x \rangle = n_1 S_1 + n_2 S_2 \tag{4.31}$$

$$\sigma_x^2 = n_1 S_{1-} + n_2 S_{2-}, \tag{4.32}$$

where n_1 and n_2 are the number of active ErbB1 and ErbB2 receptors, respectively, specified in Eqs. 4.29 and 4.30. Then, the average and the variance of the bound protein distribution are

$$\langle x \rangle = S_1 \frac{S_0}{2} + \left(\frac{1}{2} K R_2 \log_2 \left(1 + \frac{2S_0}{KR_2}\right)\right)(S_2 - \frac{1}{2}S_1) \tag{4.33}$$

$$\sigma_x^2 = S_{1-} \frac{S_0}{2} + \left(\frac{1}{2} K R_2 \log_2 \left(1 + \frac{2S_0}{KR_2}\right)\right)(S_{2-} - \frac{1}{2}S_{1-}), \tag{4.34}$$

where $K = \frac{k_{12+}}{k_{1+}}$ and $S_0 \equiv y$.

The mutual information between the ligand and the protein distributions is defined in Eq. (4.2), and thus we need to calculate the expected value of non-linear functions of a random variable. Applying the Taylor expansion method (see Appendix, Section A.3 and reference [85]), we can calculate $H(X|Y)$ and $H_Y(X)$ and obtain the mutual information between the ligand (input) distribution and the intracellular proteins (output) distribution.

The conditional entropy $H(X|Y)$ is given by

$$H(X|Y) = -\frac{1}{2}\log_2 2\pi e - \frac{1}{2}\int dy \log_2 \frac{1}{2}\left[yS_{1-} + KR\log_2\left(1+\frac{2y}{KR}\right)\left(S_{2-}-\frac{1}{2}S_{1-}\right)\right], \tag{4.35}$$

which applying the Taylor expansion method, becomes

$$H(X|Y) = -\frac{1}{2}\log_2 2\pi e - \frac{1}{2}\log_2 \frac{1}{2}\left[\mu_y S_{1-} + KR_2\log_2\left(1+\frac{2\mu_y}{KR_2}\right)\left(S_{2-}-\frac{1}{2}S_{1-}\right)\right]. \tag{4.36}$$

The term $H_Y(X)$ involves the variance of the distribution of the intracellular protein distribution ($H_Y(X) = -\log_2\sqrt{2\pi e\sigma^{2'}}$), which we can estimate by means of a mixture model ($\sigma_x^{2'} = \int dy p_y \sigma_x^2 + \int dy p_y \mu_x^2 - \left(\int dy p_y \mu_x\right)^2$) (see reference [86]).

$$\sigma_x^{2'} = \frac{1}{2}\left(\mu_y S_{1-} + KR\log_2\left(1+\frac{2\mu_y}{KR}\right)\left(S_{2-}-\frac{1}{2}S_{1-}\right)\right) + \sigma_y^2\left(\left(\frac{S_1}{2}+\frac{\left(S_2-\frac{1}{2}S_1\right)}{1+\frac{2\mu_y}{KR_2}}\right)^2 + \left(\frac{\left(S_2-\frac{1}{2}S_1\right)}{1+\frac{2\mu_y}{KR_2}}\log(1+\frac{2\mu_y}{KR_2})\right)^2\right) \tag{4.37}$$

Then, the mutual information between the input distribution and the output distribution in the case of ligand-induced dimerization is well approximated by

$$I(X,Y) = \frac{1}{2}\log_2\left(1+\frac{\sigma_y^2\left(\left(\frac{S_1}{2}+\frac{KR_2\left(S_2-\frac{1}{2}S_1\right)}{KR_2+2\mu_y}\right)^2\right)}{\frac{1}{2}\left(\mu_y S_{1-} + KR_2\left(S_{2-}-\frac{1}{2}S_{1-}\right)\log_2\left(1+\frac{2\mu_y}{KR_2}\right)\right)}\right) \tag{4.38}$$

In Fig. 4.5, we show that the agreement between our analytical expression (Eq. 4.38) and the stochastic simulations of the system is excellent.

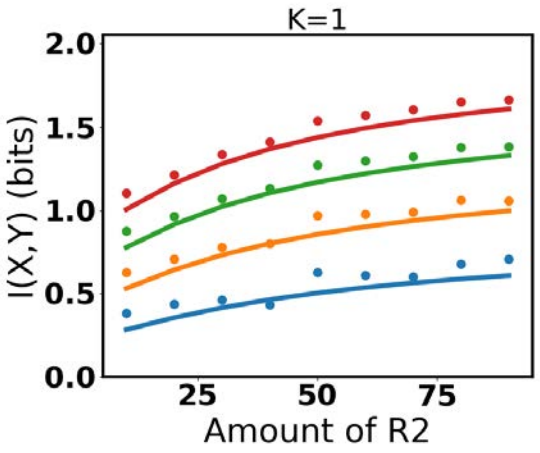


FIGURE 4.5: Mutual information for different values of initial ErbB2 concentration and different averages of the ligand distribution ($\mu_y = 20$ (red), $\mu_y = 30$ (green), $\mu_y = 40$ (orange), $\mu_y = 50$ (blue)) for ligand induced dimerization. The solid line was calculated according to Eq. 4.38, dots show the results of stochastic simulations. $R_1 = 50$, $K = 1$ and the binding probabilities p_{bs} are calculated with the values of the affinities given in [23].

Furthermore Fig. 4.6 shows that only when the intracellular protein binds ErbB1 preferentially (c), ErbB2 overexpression and an increase in heterodimerization (increase in K) are detrimental for mutual information. In all other cases, an increase in ErbB2 concentration and in the relative heterodimerization rate (K) result in higher mutual information. This is due to the fact that the total number of signaling dimers increases with heterodimer formation in the ligand induced dimerization mechanism, as proposed in [46].

Different dimerization mechanisms yield different values of the mutual information

In Fig. 4.7, we compare the mutual information corresponding to the two dimerization mechanisms compared throughout this thesis. Predimerization yields higher values of the mutual information and it is also less sensitive to changes in the levels of ErbB2. Therefore, the predimerization mechanism seems to be more efficient as an information transduction channel and it also provides robustness to the information transmission process.

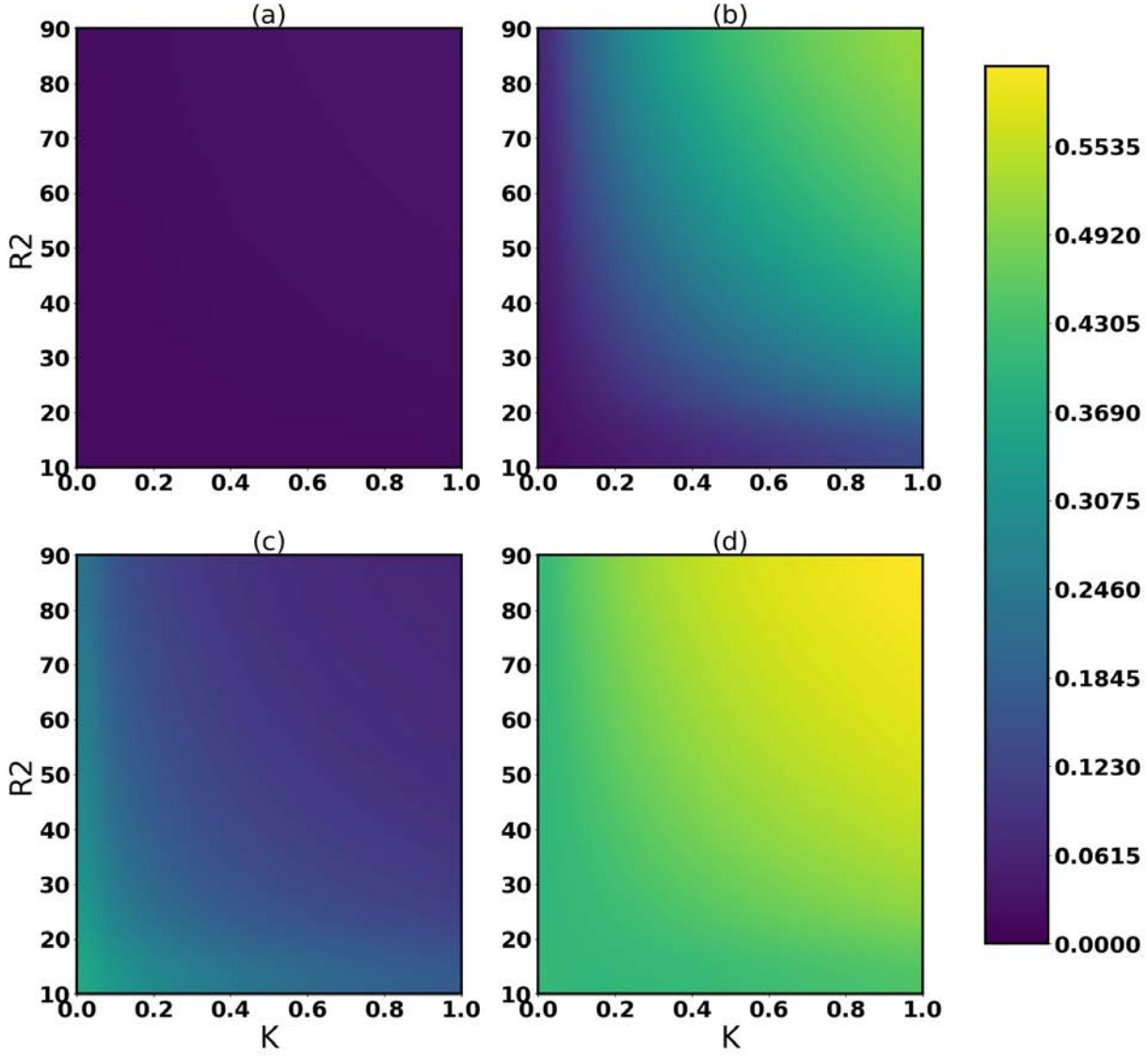
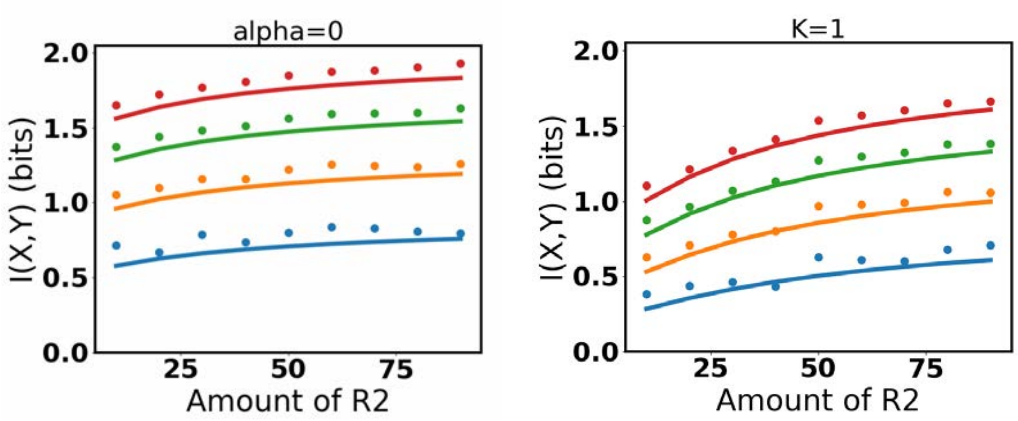


FIGURE 4.6: Mutual information between the ligand and PI3K as a function of the amount of ErbB2 receptors and $K = \frac{k_{12+}}{k_{1+}}$, for different affinities of the intracellular proteins for active R_1^* and R_2^*.

4.2.3 Effect of the receptor-protein affinity on the mutual information

In this Section, we consider the dependency of mutual information on the intracellular proteins affinities for the active receptors.



((A)) ((B))

FIGURE 4.7: Mutual information for different values of initial ErbB2 concentration and different averages of the ligand distribution ($\mu_y = 20$ (red), $\mu_y = 30$ (green), $\mu_y = 40$ (orange), $\mu_y = 50$ (blue)) for the two different dimerization mechanisms: predimerization (A) and ligand induced dimerization (B). The solid lines were calculated according to Eq. 4.28, dots show the results of stochastic simulations. $R_1 = 50$ and the binding probabilities p_{bs} are calculated with the values of the affinities given in [23].

The mutual information depends on the average and the variance of the intracellular proteins's affinity distribution. Fig. 4.8 shows the relation between the scaled variance and average (normalised by the number of binding sites that the protein can bind in the active receptor), taking the parameter values for all the proteins in Jones's database [23]. We performed an exponential fit to the data, so that we can approximate the variance as a function of the average:

$$\sigma^2 = 0.005(1 - e^{4\mu}) + 0.28 \tag{4.39}$$

Using such fit, we can explore the mutual information as a function of only S_1 and S_2 by considering $S_{1-}(S_1)$ and $S_{2-}(S_2)$.

Fig. 4.9 shows how the mutual information depends on the affinity of the intracellular protein for each receptor (ErbB1 and ErbB2), for the two different dimerization mechanisms considered. In both cases higher affinities for both receptors yield a higher mutual information. As in the previous section, predimerization gives higher values of the mutual information, and it is less sensitive to changes in the affinity of the intracellular protein to ErbB2 receptors (compatible with the reduced sensitivity to changes in the amount of ErbB2 receptors discussed in the previous section).

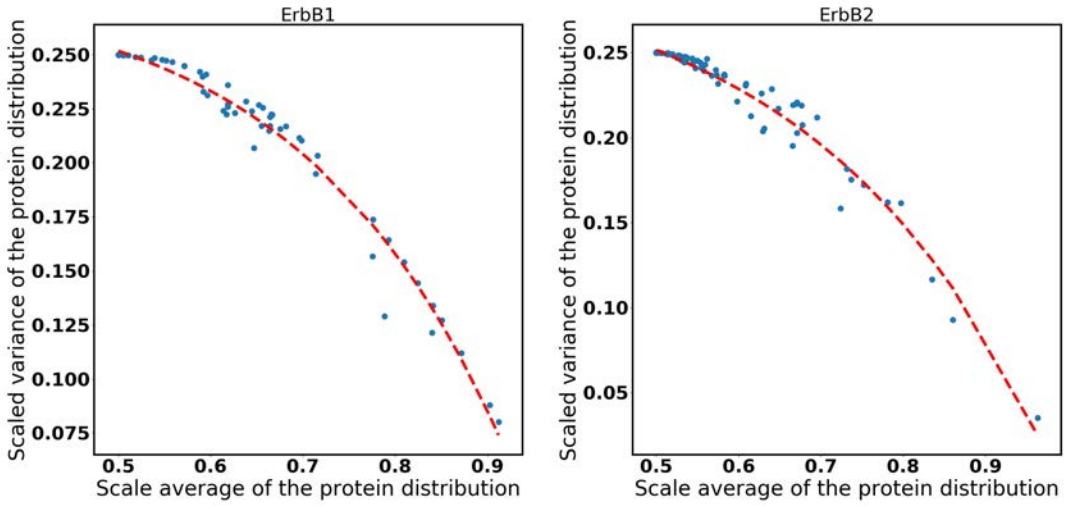
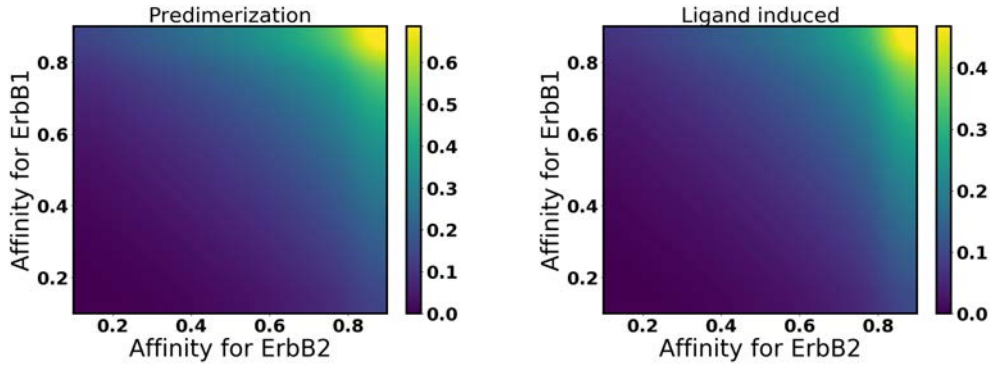


FIGURE 4.8: Variance as a function of the average of the distribution for the different proteins studied in [23]. Dots show the parameters calculated from the experimental values, the dashed line shows the fit by $y = 0.005(1 - e^{4x}) + 0.28$.

This is in contrast to what we observe from real data (shown in Fig. 4.10), if mutual information is a quantity to optimize: experimental data show that proteins that bind to ErbB receptors have binding affinities for each one that do not correspond to maxima in information transmission. This can be due to structural constraints as well as dynamical constraints (as we discussed in Chapter 3).



((A)) ((B))

FIGURE 4.9: Mutual information for different affinities of one intracellular protein for the different receptors (ErbB1 and ErbB2) and for different dimerization mechanisms (predimerization (A) and ligand induced dimerization (B)). $\mu_y = 25$, $\sigma_y^2 = 5$, $R_1 = 25$, $R_2 = 25$.

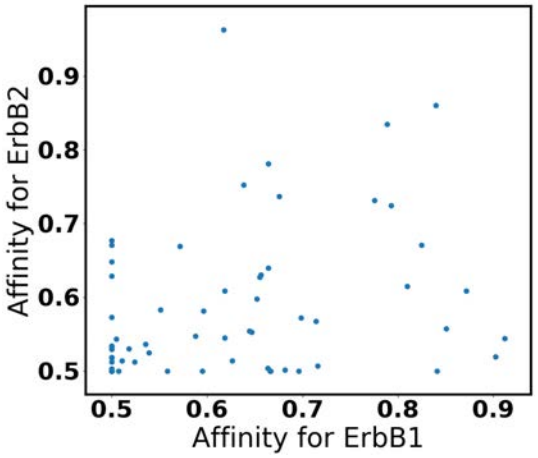


FIGURE 4.10: Proteins in [23], represented in the space of affinities for each receptor. Each dot represents a protein, and its coordinates indicate the affinity for ErbB1 and ErbB2.

4.3 One receptor, two proteins: the effects of competition.

We continue our analysis by tackling the more realistic case in which the intracellular proteins are not independent from one another: they are correlated through competition for the binding sites in active receptors. We thus move on to consider the information transmitted by two proteins that compete for the binding sites of receptors of one type.

Since the mutual information only measures correlations between two variables at a time, in this section we use multivariate mutual information [92, 93] in order to quantify the effect of having several proteins competing for the binding sites in information transmission. Multivariate mutual information measures the degree of redundancy or synergy between the two dependent variables (X_1 and X_2, the intracellular proteins in our case), that is, how much information about the independent variable (Y, the ligand) is shared by the two dependent variables. The multivariate mutual information is defined as

$$I(X_1, X_2, Y) = I(X_1, Y) - I(X_1, Y|X_2) \tag{4.40}$$

The first term in Eq. (4.40) reads

$$I(X_1, Y) = \int dy \int dx_1 p(x_1, y) \log_2 \frac{p(x_1, y)}{p(x_1)p(y)} \tag{4.41}$$

which, if we use conditional probabilities, turns into

$$I(X_1, Y) = \int dy p(y) \int dx_1 p(x_1|y) \log_2 \frac{p(x_1|y)}{p(x_1)} \tag{4.42}$$

The second term in Eq. 4.40, $I(X_1, Y|X_2)$ is given by

$$I(X_1, Y|X_2) = \int dx_2\, p(x_2) \int dy \int dx_1\, p(x_1, y|x_2) \log_2 \frac{p(x_1, y|x_2)}{p(x_1|x_2)p(y|x_2)} \tag{4.43}$$

which, after a few rearrangements, becomes

$$I(X_1, Y|X_2) = \int dy\, p(y) \int dx_2 \int dx_1\, p(x_1, x_2|y) \log_2 \frac{p(x_1, x_2|y)p(x_2)}{p(x_1, x_2)p(x_2|y)} \tag{4.44}$$

Taken together, these results allow us to express the multivariate mutual information as follows:

$$\begin{aligned} I(X_1, X_2, Y) = & \int dx_2 \int dx_1\, p(x_1, x_2) \log_2 \frac{p(x_1, x_2)}{p(x_1)p(x_2)} \\ & - \int dy\, p(y) \int dx_2 \int dx_1\, p(x_1, x_2|y) \log_2 \frac{p(x_1, x_2|y)}{p(x_1|y)p(x_2|y)} \\ & = I(X_1, X_2) - I(X_1, X_2|Y) \end{aligned} \tag{4.45}$$

In Chapter 3, we established that the distribution of bound proteins of both types to one binding site is a multinomial (see Chapter 3, Section 3.1.1):

$$p(X_1, X_2) = \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} \left(\frac{k_1}{1 + k_1 + k_2}\right)^{x_1} \left(\frac{k_2}{1 + k_1 + k_2}\right)^{x_2} \left(\frac{1}{1 + k_1 + k_2}\right)^{(n - x_1 - x_2)} \tag{4.46}$$

As we saw in Chapter 3, we can approximate this multinomial distribution by a multivariate normal distribution, using the normal approximation to a multinomial

with parameters $\mu = \binom{np_1}{np_2}$ and $\Sigma = \binom{np_1(1-p_1) \quad -np_1p_2}{-np_1p_2 \quad np_2(1-p_2)}$. Here, n is the number of active receptors and $p_1 = \frac{k_i}{1+\sum_{j\in P} k_j}$, where k_i is the affinity of protein i to the binding site, and P is the set of considered proteins.

Once we have the multivariate normal distribution for every binding site, we can find the distribution of total protein of each type bound to the receptor by summing the multivariate normal distribution's parameters. Then, the corresponding parameters are $\mu_T = \binom{\sum_{bs} np_{1,bs}}{\sum_{bs} np_{2,bs}}$ and $\Sigma_T = \binom{\sum_{bs} np_{1,bs}(1-p_{1,bs}) \quad -\sum_{bs} np_{1,bs}p_{2,bs}}{-\sum_{bs} np_{1,bs}p_{2,bs} \quad \sum_{bs} np_{2,bs}(1-p_{2,bs})}$.

We use the notation introduced in Section 4.1.1, $S_i = \sum_{bs} np_{i,bs}$ and $S_{i-} = \sum_{bs} np_{i,bs}(1 - p_{i,bs})$, where i refers to the protein type.

$I(X_1, X_2)$ in Eq. (4.45) is the mutual information of the multinomial or multivariate normal distribution $p(X_1, X_2|Y)$, which is given by (see [94])

$$I(X_1, X_2|Y) = -\frac{1}{2}\log_2(1 - \rho^2), \tag{4.47}$$

where $\rho = \frac{cov(x_1,x_2)}{\sigma_1\sigma_2} = \frac{S_{12}}{\sqrt{S_{1-}S_{2-}}}$ is the Pearson correlation coefficient. Here $S_{12} = \sum_{bs} p_{1,bs}p_{2,bs}$.

Regarding $I(X_1, X_2)$ in Eq. (4.45), we do not have direct access to the distributions $p(x_1)$, $p(x_2)$ and $p(x_1, x_2)$, but we can calculate them indirectly by integrating the conditional distributions for all y. Integrating the multivariate normal distribution with respect to y, we can calculate the new covariance matrix, which is is

$$\Sigma' = \begin{pmatrix} \mu_y S_{1-} + \sigma_y^2 S_1^2 & -\mu_y S_{12} + \sigma_y^2 S_1 S_2 \\ -\mu_y S_{12} + \sigma_y^2 S_1 S_2 & \mu_y S_{2-} + \sigma_y^2 S_2^2 \end{pmatrix}, \tag{4.48}$$

which, in turn, allows us to calculate the mutual information between X_1 and X_2 (using Eq. (4.47) with the correlation coefficient of Σ'),

$$I(X_1, X_2) = -\frac{1}{2}\log_2\left(1 - \frac{(-\mu S_{12} + \sigma^2 S_1 S_2)^2}{(\mu S_{1-} + \sigma^2 S_1^2)(\mu S_{2-} + \sigma^2 S_2^2)}\right). \tag{4.49}$$

Then, the multivariate information is

$$I(X_1, X_2, Y) = \frac{1}{2}\log_2\left(\frac{1-\rho^2}{1-\frac{(-\mu S_{12}+\sigma^2 S_1 S_2)^2}{(\mu S_{1-}+\sigma^2 S_1^2)(\mu S_{2-}+\sigma^2 S_2^2)}}\right) \tag{4.50}$$

Fig 4.11 shows the agreement of our theoretical prediction for the multivariate mutual information (Eq. (4.50)) with the mutual information calculated from stochastic simulations.

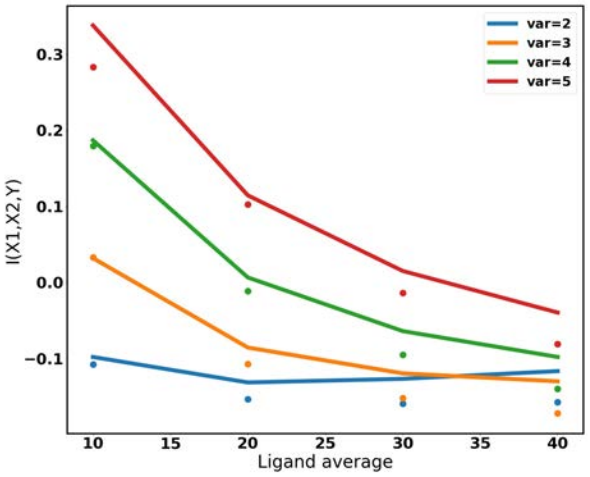


FIGURE 4.11: Multivariate mutual information as a function of the average of the input distribution for two proteins, one with 6 bs, the other with 2bs (competition for 2 bs). Note that MVMI is positive in most of the conditions.

We now consider the issue of under which conditions we observe positive multivariate mutual information, associated to redundancy, or negative multivariate mutual information, corresponding to synergy or complementarity. This is determined by the sign of $I(X_1, X_2, Y)$ (see [95] for details).

$I(X_1, X_2, Y)$ is positive if

$$(1-\rho^2) > (1-\frac{(-\mu_y S_{12}+\sigma_y^2 S_1 S_2)^2}{(\mu_y S_{1-}+\sigma_y^2 (S_1)^2)(_{2-}+\sigma_y^2 (S_2)^2)})$$

$$\frac{\sigma_y^2}{\mu_Y} > \left(\frac{(S_1)^2}{S_{1-}}+\frac{(S_2)^2}{S_{2-}}+2\frac{S_1 S_2}{S_{12}}\right)\left(\frac{S_{1-}S_{2-}}{(S_1)^2(S_2)^2}(\frac{(S_{12})^2}{S_{1-}S_{2-}-(S_{12})^2})\right)$$

$$\frac{\sigma_y^2}{\mu_y}\left(\frac{S_{1-}S_{2-}-(S_{12})^2}{(S_{12})^2}\right)\frac{1}{\left(\frac{S_{1-}}{(S_1)^2}+\frac{S_{2-}}{(S_2)^2}+2\frac{S_{1-}S_{2-}}{S_{12}S_1S_2}\right)}>1 \tag{4.51}$$

Let us define $\xi \equiv \frac{\sigma_y^2}{\mu_y}\left(\frac{S_{1-}S_{2-}-(S_{12})^2}{(S_{12})^2}\right)\frac{1}{\left(\frac{S_{1-}}{(S_1)^2}+\frac{S_{2-}}{(S_2)^2}+2\frac{S_{1-}S_{2-}}{S_{12}S_1S_2}\right)}$. In Fig. 4.12, we plot the multivariate mutual information as a function of ξ to check the validity of Eq. (4.51).

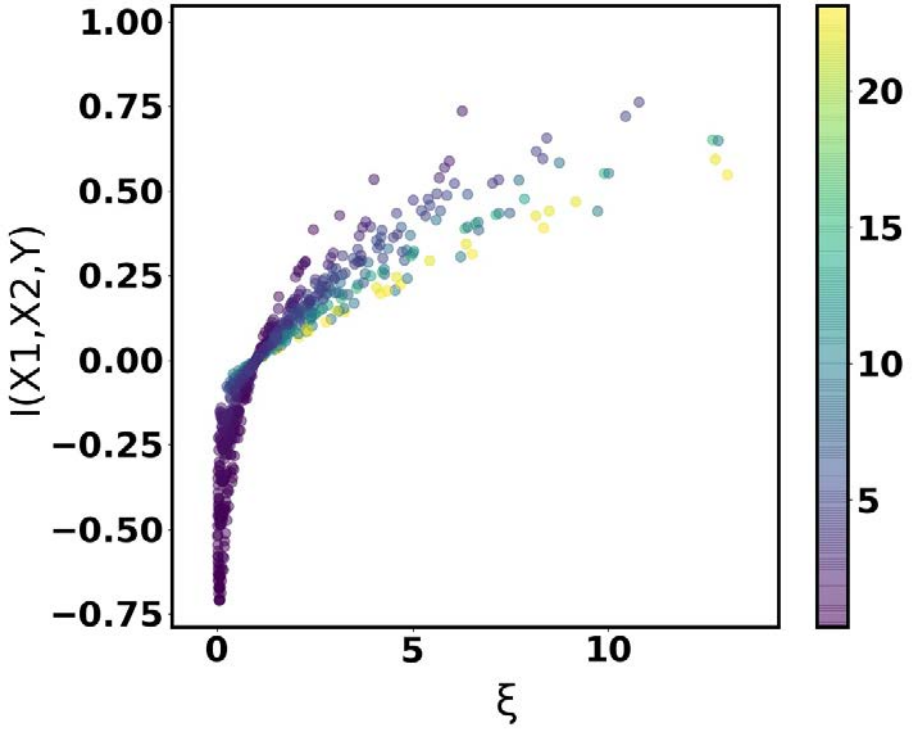


FIGURE 4.12: Multivariate mutual information as a function of ξ, for different Fano factors of the input distribution and different proteins with binding affinities taken from [23]. The color scale is related to the intensity in the competition between the proteins, as explained in the main text. Note that the multivariate mutual information is negative (synergy between intracellular proteins) for $\xi < 1$.

In Eq. 4.51, we can distinguish three factors. The first, $\frac{\sigma^2}{\mu}$ is the Fano factor of the input (ligand) distribution, which measures how noisy the input is: high values of the Fano factor are due to high variance or low mean and indicate a noisy input, whereas low values of the Fano factor are related to low variance or high mean, which come from an input with low noise.

The third factor, $\frac{1}{\left(\frac{S_{1-}}{(S_1)^2}+\frac{S_{2-}}{(S_2)^2}+2\frac{S_{1-}S_{2-}}{S_{12}S_1S_2}\right)}$ accounts for the affinities of each protein to the binding sites.

The second factor, $\left(\frac{S_{1-}S_{2-}-(S_{12})^2}{(S_{12})^2}\right)$ involves the variances and covariance of the output multivariate normal distribution, corresponding to the bound intracellular proteins, and it accounts for the correlations between different proteins, which are caused by competition for limited binding sites. Therefore is the term which includes the effect of competition on the multivariate mutual information. In Fig. 4.12, the color code represents this quantity.

Consider the second factor of Eq. 4.51 for the case of a receptor with two binding sites which both proteins can bind. We can rewrite it as

$$\begin{aligned}\left(\frac{S_{1-}S_{2-}-(S_{12})^2}{(S_{12})^2}\right) =&(p_{1A}+p_{1B})(p_{2A}+p_{2B})\left(1-\frac{p_{1A}^2+p_{1B}^2}{p_{1A}+p_{1B}}-\frac{p_{2A}^2+p_{2B}^2}{p_{2A}+p_{2B}}\right)\\ &+(p_{1A}p_{2B}-p_{1B}p_{2A})^2-(p_{1A}p_{2A}+p_{1B}p_{2B})^2\end{aligned} \tag{4.52}$$

The term $(p_{1A}p_{2B}-p_{1B}p_{2A})^2$ measures the extent of competition: if both proteins bind the same binding site with similar affinity, then this term will be very low, whereas if the binding affinity of each protein for a given binding site is very different, then this term will be high. In Fig. 4.12, it can be appreciated that negative values of the multivariate mutual information are only possible for low values of $\left(\frac{S_{1-}S_{2-}-(S_{12})^2}{(S_{12})^2}\right)$, which are achieved, in part, due to similar affinity binding for the two proteins for a binding site. Therefore we can say that competition for a binding site (at least in the case of two proteins) yields more negative values of the multivariate mutual information, thus contributing to synergy between competing proteins.

Chapter 5

Conclusions

5.1 Main results and conclusions

The ability of organisms to extract and store information from their surroundings marked a revolution in the history of life. Living systems able to make decisions, through many different ways, based on their perception of a changing environment gained key evolutionary advantages early in the appearance of complex life forms [96].

Cells, from prokaryots to eukaryots in multicellular systems, use specific receptors inserted in their membranes to detect extracellular molecules that cannot cross into the cell. In the present thesis, we have studied in detail the dynamical properties of the signaling processes of two molecular receptors of the ErbB family. Through detection of different ligands, these receptors play a key role in many cellular processes such as cell growth, proliferation, migration or apoptosis [97].

In particular, two main aspects determine the adequacy of a response to the cellular environment: the dynamics of membrane receptors and the interactions between these and intracellular molecules. Here we have studied the information flow from external ligands, throughout membrane receptors and eventually into internal molecular cascades that elicit the corresponding response. If this dynamical events can be understood as an information transmission process, what insights can mathematical modeling give us about the role played by each of the molecular components involved?

In order to explore this question, we have tracked how the input information (encoded in stimuli involving extracellular ligand) is translated into intracellular information (amount of the different active intracellular components), both characterised as probability distributions. In Chapter 2 we have modeled the dynamics of membrane receptors, the readers of the input distribution, and determined the extent of receptor activation as a response to ligand detection. The fact that intracellular proteins bind to these receptors at much faster time scales allowed us to treat their dynamics separately, as introduced in Chapter 3. In Chapter 4, both components are brought together to understand how the activation of receptors and their interactions with intracellular proteins can be predictive of the output distribution signal. This allowed us to quantify the amount of information that the output distribution read by the cell (amount of proteins of each type) carries about the input distribution of its environment (amount of ligand).

Information transmission is higher in predimerized receptors

Despite ErbB receptors represent one of the most thoroughly studied receptor families, many open questions remain about their mechanism of action and the specific roles of each member of the family. For instance, it was not until the availability of single molecule experiments that molecular configurations were described explaining the mechanisms of dimerization and activation of the ErbB receptors [98]. However, even by knowing in detail the molecular configurations that allow dimerization and activation, the prevalence of different dimerization mechanisms (such as ligand induced *vs* predimerization) remains to be assesed, with current estimates of the prevalence of predimerization ranging from 40 to 100% in ErbB membrane receptors [77, 78].

By modeling the two alternative dimerization mechanisms separately we have been able to compare the extent of information transmission for each situation, showing that the mutual information of the predimerization system is higher than in its ligand induced counterpart. If the need for effective signalling processes maximizes mutual information between external ligand and intracellular protein distributions, our modeling is indicative of why predimerization seems to be more prevalent: ErbB receptors represent a more efficient information channel when they predimerize because of two reasons. First, the mutual information through

predimerized receptors is higher than for ligand induced dimerization. Second, the mutual information is less sensitive to changes in the amount of ErbB2, which makes information transmission more robust.

The role of ErbB2 in signaling and cancer

The role of ErbB2 remains another open question, with key implications in many cellular processes. This receptor does not have any known ligand, meaning it does not specifically sense extracellular signals. Furthermore, its affinity spectrum for the intracellular proteins is different to that of the rest of ErbB receptors: it appears to be more 'promiscuous', that is, it binds a more diverse repertoire of intracellular proteins than other ErbB receptors [23].

To understand its role, current research proposes it is the preferred heterodimerization partner of the other ErbB family members [8], while amplifying and sustaining their signaling activity [46]. However, it has also been shown that its overexpression leads to self-induced growth signaling, a process leading to uncontrolled cellular proliferation at the core of carcinogenesis [11].

In Chapter 2, we analysed in the first place the dynamics of ErbB2 in the absence of ligand. ErbB2 is present in the membrane in a dynamic equilibrium between three configurations: monomers, inactive dimers and active dimers. This constitutive dimerization might be useful to buffer changes in the ErbB2 synthesis rate and it has been proposed as a mechanism for preventing ErbB2 monomers to interact in the absence of signaling [76]. Next, we analysed the coupled dynamics of ErbB1 and ErbB2 by considering two different dimerization mechanisms (ligand induced dimerization and predimerization). We have proposed two models that allow us to predict the amount of active dimers of different types (ErbB1 and ErbB2 homodimers and ErbB1-ErbB2 heterodimers). In Chapter 4, we used the predicted values of the active dimers upon ligand stimulation in order to calculate the mutual information between the ligand and the intracellular proteins. This allowed us to assess the impact of changes in the receptor membrane composition on the information transmitted through this molecular channel.

In particular, we found that the effect of changes in the receptors concentration depends on the affinity spectrum of the intracellular protein considered, and we can

draw two main conclusions about the effect of ErbB2 overexpression on information transmission.

Firstly, because of ErbB2 ligand-independent dynamics, active homodimers are formed in the absence of a ligand. This implies receptor activation that does not carry information regarding the ligand so that the associated activation of the intracellular proteins provokes a decrease in the mutual information when ErbB2 is overexpressed.

Second, we observe a decrease of mutual information for increasing concentrations of ErbB2 in proteins that bind ErbB1 preferentially (even when we do not consider ErbB2 active dimers). This is due to an increase in the formation of asymmetric heterodimers -where ErbB2 is being activated by ErbB1- when ErbB2 concentration increases. Therefore, ErbB1 is 'kidnapped' instead of being signaling as a part of an ErbB1 homodimer.

Taken together, these results regarding a decrease in the mutual information for increasing values of ErbB2 concentration is compatible with the loss of the ability to respond properly to the extracellular information carried by the ligand.

Other factors can be also causing this missbehavior, as the recruitment of a greater diversity of molecular pathways, elicited by the particular 'promiscuity' of ErbB2 [23]. One possible direction to extend this work would be by measuring the diversity of intracellular proteins recruited as a function of membrane receptors' concentrations.

Binding and unbinding rates in intracellular interactions

Another basic process in cell signaling is the binding and unbinding of intracellular proteins to active receptors, and their subsequent activation by the receptors' kinase activity. This process can be understood as an enzymatic chemical reaction where an enzyme (the receptor) binds and acts on a substrate (the intracellular proteins).

Reactions of this kind are usually modeled with Michaelis-Menten kinetics [40], which allows for a description of the rates of product formation as a function of the substrate concentration. Contrary to the assumptions involved in the Michaelis-Menten model, we included in our analysis of receptor-protein interactions the

decay of the receptor and we have realized how the limited lifetime of the receptors relates to a trade-off in the values of the rate constants. We carried out this analysis in Chapter 3.

By considering the exponential decay of receptors as a Poisson process, we have discussed the appearance of a novel trade-off in the intrinsic rates of the interactions (binding (k_+), unbinding (k_-) and activation (α)). This arises when studying the conditions, with the decay time scale at play, that maximize the amount of substrate that is converted into product. We put forward that this translates into a benefit for the cell, since synthesizing intracellular proteins has a metabolic cost that would render useless if such proteins are not active in the signaling process.

In particular, we have derived analytically an estimation for the optimal unbinding rate, $k_- = \sqrt{\frac{k_+}{\alpha}}$, that gives insight into how the intrinsic activation rate of the receptor constraints the rates of the binding and unbinding processes if we expect the maximum number of proteins to be participating in signaling.

Although there exist several experimental (bulk) techniques to measure equilibrium constants, measuring the individual binding and unbinding rates requires of single molecule techniques, which have been relatively rare. Thus, the experimental data regarding binding and unbinding constants is scarce. We have reviewed the literature in the search of such data, together with the catalytic rate of kinases. Although we did not find single molecule experimental data for the rates of a particular enzyme, we could find the binding ($k_+ \sim 0.1$ min^{-1}) and unbinding ($k_- \sim 30$ min^{-1}) rates of chaperonin and co-chaperonin (proteins that are involved in protein folding). Taking a value of the intracellular protein concentration of 500 NM (the intracellular protein concentration value used throughout this thesis). an optimal catalytic rate (if the number of active proteins is to be maximized) would be $\alpha = k_+ k_-^2 \sim 4.5 \cdot 10^4$ min^{-1} $= 7.5 \cdot 10^2$ s^{-1}. Experimental values of the catalytic rate of kinases span several orders of magnitude, from 10^{-2} to 10^6 s^{-1}, with a maximum in $k_{cat} = 10$ s^{-1} (see Fig. 5.1), so our estimate of the optimal catalytic rate is well within the experimental feasible range, close but above the average value.

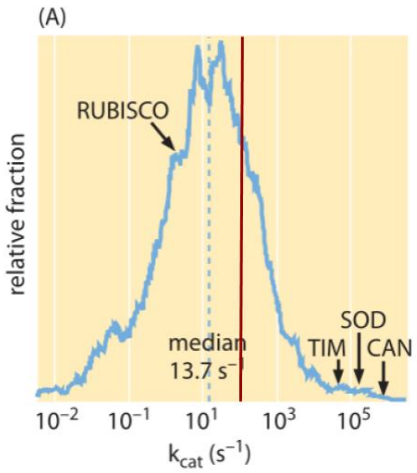


FIGURE 5.1: Distribution of experimental measures enzymatic catalytic rates (k_{cat}). Figure taken from BioNumbers [66].

Competition and specialization in binding site affinities

The binding rates of intracellular proteins to receptor binding sites, are dependent on the type of protein and binding site: protein-receptor interactions are specific. This specificity introduces novel dynamics when we understand that different proteins coexist and compete for binding site occupation. In Chapter 3 we have introduced this notion by means of understanding how binding site specificity in a population of different proteins shapes the possible occupancy levels of such sites, thus playing a key role in the efficiency of the initial stage of the signaling cascade.

In this approach, we considered several binding sites and several proteins competing for them. We considered structural constraints in the proteins and binding sites that cause that adaptation for one binding site happens in detriment of the complementarity for the other binding sites. In such a scenario, we found that the distribution of affinities that maximizes the total amount of proteins activated is such that every protein is specialized in the interaction with one binding site and does not bind the others. However, when the number of binding sites are limited and competition is forced upon proteins for the binding sites, then the affinity distribution that maximizes the efficiency of the binding and activation processes is equipartition - when the affinity is allocated equally between the binding sites without a preference for any particular binding site). We studied the allocation of the affinity for the proteins studied in [23] and found that the a high entropy for the affinity values for the different binding sites, hinting to a tendency to bind the

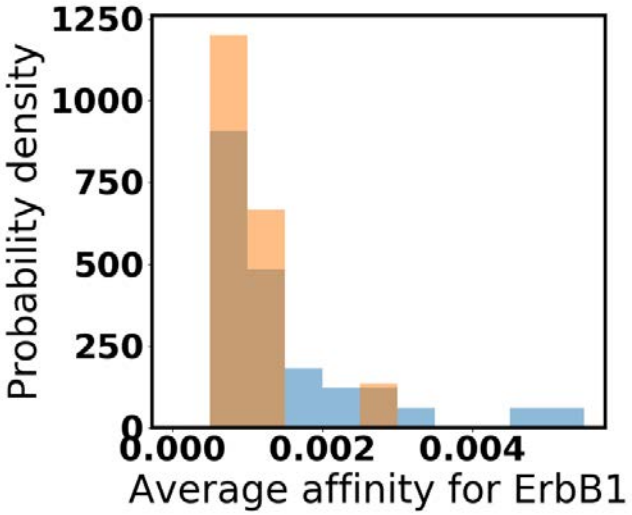


FIGURE 5.2: Average affinity for ErbB1 of the different proteins studied in [23]. The histograms represent the probability density for proteins that bind one binding site (blue) and proteins that bind two binding sites (orange).

different binding sites with similar affinity. As discussed here, such an allocation maximizes the number of proteins activated by the active receptors.

If we assume that there is a trade-off between generality and specificity, then, in a protein binding several partners, the affinity for every partner would be lower than the affinity of a protein binding a single partner, which seems to be compatible for the proteins studied in [23] for ErbB1 (see Fig. 5.2). A tendency to reduce binding rates in higher organisms has been described in [99] and a tradeoff between compexity, stability and specificity is describes in [82]. This is consistent with a broadening of the spectrum of binding partners. In a context of increasing complexity, reducing the binding affinities and increasing the repertoire of interaction partners could be a way to gain robustness.

Competition and information

Last but not least, wee also studied the effects of competition on information transmission. We focused on the multivariate mutual information as a way of quantifying the degree of synergy or redundancy between two intracellular proteins and arrived, with this approach, to the conclusion that a uniform distribution of the affinity between binding sites contributes to synergy between intracellular proteins: when the two proteins have no preference for any binding site, the information obtained by the two proteins together is greater than the information

obtained with each protein separately. This 'equipartition' of the binding affinity is obtained also as a way of maximizing the efficiency of the binding-activation processes, as we discussed in Chapter 3.

5.2 Open questions and further steps

In this doctoral thesis we have established and developed a mathematical framework to gain insight into the information transmission processes governed by ErbB receptors. As discussed in these conclusions, we have studied the dynamics at several scales within the system, from the roles of different dimerization mechanisms in receptor activation to the stochastic dynamics governing the binding of intracellular proteins to the receptors' binding sites. This approach has been the starting point to study and quantify several questions such as the role of ErbB2, protein unbinding rates or affinity distributions in cell signaling.

Nonetheless, further relevant questions regarding the ErbB system are still open, and we consider that our mathematical approach can be useful to shed further light when approaching them in the future. Firstly, several results acknowledges that noise might play a crucial role in many biological process [100], where it seems to be not only a by-product of deterministic dynamics, but itself a relevant variable. From this perspective, taking noise into account by admitting that ErbB receptors are not, as in our work, noiseless, will provide further insight into what are the specific roles of each receptor of the family.

On the other hand, we have put forward an explanation to the proliferative dynamics resulting from ErbB2 overexpression as a loss of the capacity of the receptors channel to transmit information. Knowing that this cellular process plays a determining role in tumour formation, our modeling approach has given preliminary insights by understanding how cells with overexpressed ErbB2 present a loss of mutual information and thus a poorer reading of the signals for multicellular homeostasis. However, many other factors are known to participate in this process, and our model could benefit from taking into account the fact that a great diversity of proteins are activated following ErbB2 upregulation. The role of plethora of new interacting proteins in the malignant transformation of cells remains an open area that could benefit from a mathematical approach similar to the one here discussed.

Finally, a more detailed examination of the structure of the intracellular interactions of the ErbB receptors with the first intracellular intermediaries in cell signaling and, more importantly, of the identity of the interacting partners (together with the pathways they trigger) should allow to classify the different intracellular proteins according to their affinities for one or the other receptor. Two relevant outcomes could stem from further developing this perspective. First, it would result in a better understanding of which specific interactions, and through which mechanisms, participate in each of the ErbB signaling cascades: how -and to what extent- each signaling pathway is triggered by the different receptors. This could be explored by means of information-theoretical tools as multi-channel information transmission [95]. Furthermore, when put together with experimental studies of protein expression, it would allow for the identification of proteins and pathways that are more sensitive to ErbB overexpression. This information could be used to identify cell types more likely to develop drug resistance in ErbB2-positive cancer treatments. In the light of this, strategies to circumvent the inadequate activation of such proteins could be designed based on the knowledge of their interactions with the different receptors.

All in all, by producing new results and pointing a way towards the next steps in ErbB modeling, we realize once again how the language of mathematics provides us with a powerful way of studying reality.

Appendix A

Mathematical methods

A.1 Brief summary on random variables and probability distributions

Let X be a variable representing the state of a system and $\mathcal{X}$ the set of all the possible states. We can assign a probability $p(X)$ to each state: the states with a higher probability will be more likely to be observed. In mathematical terms, we call X a random variable and $p(X)$ is its probability distribution.

Given that X must be in one of the possible states, the total probability must be 1:

$$\sum_{x \in \mathcal{X}} p_n = 1 \tag{A.1}$$

where K is the number of possible states of the system. We say that the probability is normalized.

Moments of a probability distribution

Probability distributions are functions that can be described by their position and shape. Two important quantities that characterize the position and shape of a probability distribution are the average and the variance.

The average, or expected value of the variable X, is

$$\langle X \rangle = \sum_{x \in \mathcal{X}} xp(x), \tag{A.2}$$

and the variance,

$$var(X) = \langle X^2 \rangle - \langle X \rangle^2 = \sum_{x \in \mathcal{X}} x^2 p(x) - (\sum_{x \in \mathcal{X}} xp(x))^2 \tag{A.3}$$

.

The standard deviation, $\sigma_X (= \sqrt{var(X)})$ is a measure of the broadness of the probability distribution.

$\langle X \rangle$ and $\langle X^2 \rangle$ are the first two moments of a distribution and indicate where the distribution is located and how broad it is. Higher moments of the distribution are related to other features of the distribution, such as the symmetry or other parameters describing the shape of the distribution.

Correlations

When we deal with more than one random variable, the specification of the probability distribution of each variable separately falls sometimes short: the probability of X depends on the value of Y and viceversa. When this happens, we say that the variables are correlated. If the probability of X does not depend on the values of Y, then we say that X and Y are independent.

Correlation between two random variables can be quantified in several ways, often by means of correlation coefficients. The most common one is Pearson correlation coefficient

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \tag{A.4}$$

which is a measure of the dispersion of the joint distribution of X and Y, relative to the dispersion of each variable separately.

Values of $|\rho_{X,Y}|$ close to 1 indicate a strong correlation between X and Y, whereas values of $\rho_{X,Y}$ close to 0 indicate that X and Y are not correlated.

Probability distributions used in this thesis

In this thesis, we will be dealing with some of the basic probability distributions: gaussian, binomial and poisson.

In a **normal** or **gaussian** random variable, the only non-zero moments are the first and the second. Thus, normal variables are characterized by their mean and variance: $X \sim N(\mu, \sigma^2)$, and their probability distribution follows

$$p(X = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{A.5}$$

Multivariate gaussian distributions follow the same expression, but with vectorial parameters:

$$p(X = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\Sigma^2}}. \tag{A.6}$$

with $\mu = \binom{\mu_X}{\mu_Y}$ and covariance matrix $\Sigma = \binom{\sigma_X^2 \quad cov(X,Y)}{cov(X,Y) \quad \sigma_Y^2}$.

The **binomial** distribution is a discrete probability distribution describing the discrete random variable number of successes (or failures) in binary events with probabilities p and $(1-p)$:

$$p(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \tag{A.7}$$

where n is the number of trials.

The average of a binomial distribution $B(n, p)$ is $\mu = np$ and the variance is $\sigma^2 = np(1-p)$. According to the de Moivre-Laplace theorem, in the limit of very large number of trials, the binomial distribution $B(n, p)$ can be approximated by a gaussian distribution $G(\mu = np, \sigma^2 = np(1-p)$. In particular, if $np(1-p) \to \infty$ is sufficiently large, the limiting distribution of the binomial is normal. Different rules of thumb may be used to decide whether n is large enough, and p is far enough from the extremes (0 or 1) to get a good approximation of the binomial by a normal distribution. The most widely used is $min(np, np(1-p) \geq 5$ [85].

The counterpart of the binomial distribution when we deal with several variables is the **multinomial** distributions

$$p(X = x, Y = y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y \left(1 - p_X - p_Y\right)^{(n-x-y)} \tag{A.8}$$

As for the binomial distribution, for large n and/or p far from 0 or 1, we can approximate a multinomial distribution by a multivariate gaussian distribution with $\mu = \binom{np_X}{np_Y}$ and $\Sigma = \binom{np_X(1-p_X \quad -np_Xp_Y}{-np_Xp_Y \quad np_Y(1-p_Y)}$.

The **Poisson** distribution is a discrete probability distribution arising from counting the number of events occurring independently one of each other with a constant rate, in a given interval. It reads

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{A.9}$$

It is characterized by only one parameter, λ, the average number of events in the considered interval. The mean and the variance of $Po(\lambda)$ are equal: $\mu = \sigma^2 = \lambda$. The rate at which the events occur if $\frac{\lambda}{T}$, where T is the size of the interval.

According to the Poisson limit theorem, the Poisson distribution $Po(\lambda = np)$ can be obtained from a binomial distribution $B(n, p)$ as the number of trials n goes to infinity and the expected number of successes remains fixed.

The poisson distribution $Po(\lambda)$ can be also approximated by a Gaussian $G(\mu = \lambda, \sigma^2 = \lambda)$ for big λ.

In a binary experiment (i.e. where the output can be 0 or 1), a **geometric** distribution describes the number of successes that are drawn before a failure. A geometric distribution is characterized by a single parameter, the probability of a success, p. If $X \sim Geo(p)$, then

$$p(X = x) = p^x(1 - p). \tag{A.10}$$

If we generalize to the number of successes before r failures, then the corresponding random variable follows a **negative binomial** distribution with parameters r, the number of failures, and p, the probability of a success ($X \sim NB(r, p)$). Then

$$p(X = x) = \frac{(r + x - 1)!}{(r - 1)!x!} p^x (1 - p)^r \tag{A.11}$$

The negative binomial distribution $NB(r, p)$ includes the normal distribution $N(\mu = r\frac{1-p}{p}, \sigma^2 = r\frac{1-p}{p^2})$ as a limiting case when $rp(1-p) \to \infty$.

An **exponential** distribution is described by

$$p(X = x) = \lambda e^{-\lambda x} \tag{A.12}$$

This distribution describes the time between events in a Poisson process.

A.2 Basics of information theory

Information seems to be a very intuitive concept, but very difficult to quantify. It was not until Shannon formulated the first concepts of information theory [51] that information became a quantifiable quantity. Here, we remind the basic concepts in information theory. For a detailed account on information theory, we refer the reader to the book by Thomas and Cover [101].

Entropy

Uncertainty is a very important concept in information theory. Entropy, the basic concept from which most of information theory is based, is a measure of uncertainty.

By definition, entropy is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{A.13}$$

This expression can be derived from the basic features information must have and can be understood as the degree of uncertainty we have about the value of a random variable: entropy is maximum when all the values are equiprobable, and minimum ($H(X) = 0$) when only one value is possible.

Entropy can also be defined for more than one variable in several ways.

The joint entropy is

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \tag{A.14}$$

and it measures the amount of uncertainty that we have about the value of both variables.

The conditional entropy is

$$H(Y|X) = -\sum_{x\in X} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x). \tag{A.15}$$

Mutual information

Mutual information is a measure of the amount of information that one random variable contains about another random variable, or the reduction in the uncertainty of one random variable due to the knowledge of the other.

$$I(X,Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{A.16}$$

It is a measure of correlation between two random variables.

Multivariate mutual information

A related measure is the multivariate mutual information, which quantifies the effect of considering other variables

$$I(X_1,X_2,Y) = I(X_1,Y) - I(X_1,Y|X_2). \tag{A.17}$$

If the multivariate mutual information is positive, it means that knowing X_2 does not increase the information between X_1 and Y and we say that X_2 is redundant. If the multivariate mutual information is negative, it means that the information

between X_1 and Y increases when we know the value of X_2. In this case, the new variable X_2 adds information to the system and we say that it is complementary.

A.3 Taylor expansion method

The Taylor expansion method can be used in order to approximate the moments of a function of a random variable [85].

Let $f(X)$ be a function of the random variable X. The expectation of $f(X)$ is

$$E[f(X)] = E[f(\mu_X + (X - \mu_X))] \tag{A.18}$$

where μ_X is the mean of the random variable X.

Performing a Taylor expansion of $f(x)$ around the mean to second order, Eq. (A.18) turns into

$$\begin{aligned} E[f(X)] &\approx E[f(\mu_X) + f'(\mu_X)(X - \mu_X) + \frac{1}{2}f''(\mu_X)(X - \mu_X)^2] \\ &\approx f(\mu_X) + \frac{1}{2}f''(\mu_X)\sigma_X^2 \end{aligned} \tag{A.19}$$

Similarly, for the variance of X, we have

$$Var(f(X)) = E[(f(X) - f(\mu_X))^2] \tag{A.20}$$

To second order in the Taylor expansion

$$\begin{aligned} Var(f(X)) &\approx E\left[\left(f'(\mu_X)(X - \mu_X) + \frac{1}{2}f''(\mu_X)(X - \mu_X)^2\right)^2\right] \\ &\approx f'(\mu_X)^2\sigma_X^2 + \frac{1}{4}f''(\mu_X)^2\sigma_X^2 \end{aligned} \tag{A.21}$$

This can be generalized to functions of several random variables. Let $X_1, ..., X_n$ be random variables with mean $\mu_1, ..., \mu_n$ and covariance matrix Σ and define

$X = (X_1, ..., X_n)$ and $\mu = (\mu_1, ..., \mu_n)$. Let $f(X)$ be a function of the random variables. Then (to first order in the Taylor expansion)

$$E(f(X)) \approx f(\mu) \tag{A.22}$$

and

$$Var(f(X)) \approx \sum_{i=1}^{k} (f_i'(\mu))^2 \Sigma_{ii} + 2 \sum_{i>j} f_i'(\mu) f_j'(\mu) \Sigma_{i,j} \tag{A.23}$$

A.4 Poisson distribution with an exponential parameter and the geometric distribution

Let N be a random variable describing the number of events in a Poisson process given a time interval, T. N has a Poisson distribution with parameter k (the rate at which events happen):

$$p(N = n|T) = \frac{(kT)^n e^{-kT}}{n!}. \tag{A.24}$$

Let T be a random variable with an exponential distribution with parameter δ:

$$p(T = t) = \delta e^{-\delta t}. \tag{A.25}$$

The number of events N that happen in a time interval T, when T has an exponential distribution is

$$p(N = n) = \int_0^{+\infty} \frac{(kt)^n e^{-kt}}{n!} \delta e^{-\delta t} . dt \tag{A.26}$$

In order to solve Eq. (A.26), we define $x = (k + \delta)t$ so that

$$p(N = n) = \frac{\delta k^n}{(k + \delta)^{n+1}} \int_0^{\infty} \frac{x^n e^{-x}}{n!} dt. \tag{A.27}$$

Let us remind two alternative definitions of the Gamma function [102]:

- When n is a positive integer: $\Gamma(n) = (n-1)!$
- For all complex numbers except non-positive integers: $\Gamma(z) = \int_0^{+\infty} x^{z-1}e^{-x}dx$

Using the definitions of the gamma function (Eqs. (A.4) and (A.4)), Eq. (A.27 becomes

$$p(N = n) = \frac{\delta k^n}{(k+\delta)^{n+1}}, \tag{A.28}$$

which is a geometric distribution with parameter $\frac{k}{k+\delta}$.

Appendix B

Time scale analysis

B.1 Deterministic multiscale analysis

Here, we carry out a non-dimensionalization of the system with respect to the time, in order to uncover the multiple temporal scales involved in the dynamics of the receptors.

Remind the system of two receptors, with ligand-induced dimerization:

$$\frac{dS}{dt} = -k_{s+}SR_1 + k_{s-}C_1 \tag{B.1}$$

$$\frac{dR_1}{dt} = m_1 - k_{s+}SR_1 + k_{s-}C_1 - \delta_1 R_1 \tag{B.2}$$

$$\frac{dR_2}{dt} = m_2 - k_{12+}C_1R_2 - 2k_{2+}R_2^2 + k_{12-}D_{12} + 2k_{2-}D_2 - \delta_2 R_2 \tag{B.3}$$

$$\frac{dC_1}{dt} = k_{s+}SR_1 - (k_{s-} + 2k_{1+}C_1 + k_{12+}R_2)C_1 + 2k_{1-}D_1 + k_{12-}D_{12} - \delta_1 C_1 \tag{B.4}$$

$$\frac{dD_1}{dt} = k_{1+}C_1^2 - (k_{1-} + a_{1+})D_1 + a_{1-}D_1^* - \delta_1 D_1 \tag{B.5}$$

$$\frac{dD_{12}}{dt} = k_{12+}R_2C_1 - (k_{12-} + a_{12+})D_{12} + a_{12-}D_{12}^* - \delta_{12} D_{12} \tag{B.6}$$

$$\frac{dD_2}{dt} = k_{2+}R_2^2 - (k_{2-} + a_{2+})D_2 + a_{2-}D_2^* - \delta_2 D_2 \tag{B.7}$$

$$\frac{dD_1^*}{dt} = a_{1+}D_1 - a_{1-}D_1^* - \delta_1 D_1^* \tag{B.8}$$

$$\frac{dD_{12}^*}{dt} = a_{12+}D_{12} - a_{12-}D_{12}^* - \delta_{12} D_{12}^* \tag{B.9}$$

$$\frac{dD_2^*}{dt} = a_{2+}D_2 - a_{2-}D_2^* - \delta_2 D_2^* \tag{B.10}$$

Recall also that the set of parameters used in our models is given in Table B.1.

Receptor dynamics			
μ_1	2400 min^{-1}	μ_2	12 min^{-1}
k_{s+}	10^7 $\text{M}^{-1}\text{min}^{-1}$	k_{s-}	10^{-3} min^{-1}
k_{1+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{1-}	0.1 min^{-1}
k_{12+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{12-}	0.1 min^{-1}
k_{2+}	10^3 $\text{M}^{-1}\text{min}^{-1}$	k_{2-}	10^4 min^{-1}
a_{1+}	1 min^{-1}	a_{1-}	0 min^{-1}
a_{12+}	1 min^{-1}	a_{12-}	0 min^{-1}
a_{2+}	1 min^{-1}	a_{2-}	10 min^{-1}
δ_{R1}	0.014 min^{-1}	δ_{R2}	0.0006 min^{-1}
δ_{D1}	0.1 min^{-1}	δ_{D12}	0.1 min^{-1}
$[R_{1T}]$	$5 - 20 \cdot 10^4$ cell^{-1}	$[R_{2T}]$	$1 - 6 \cdot 10^4$ cell^{-1}
$[S_T]$	pM to nM	P_{tot}	500 nM
Intracellular proteins dynamics			
k_{p+}	0.1 nM^{-1} min^{-1}	k_{p-}	30 min^{-1}

TABLE B.1: Reference parameters used in the simulations.

We define a dimensionless time $\tau = k_{s+}t$,

$$\frac{dS}{d\tau} = (-SR_1 + K_{d,s}C_1) \tag{B.11}$$

$$\frac{dR_1}{d\tau} = (-SR_1 + K_{d,s}C_1) + \frac{1}{k_{s+}}(m_1 - \delta_1 R_1) \tag{B.12}$$

$$\frac{dR_2}{d\tau} = \frac{1}{k_{s+}}\left(m_2 - k_{12+}C_1R_2 - 2k_{2+}R_2^2 + k_{12-}D_{12} + 2k_{2-}D_2 - \delta_2 R_2\right) \tag{B.13}$$

$$\frac{dC_1}{d\tau} = SR_1 - \left(K_{d,s} + 2\frac{k_{1+}}{k_{s+}}C_1 + \frac{k_{12+}}{k_{s+}}R_2 - \frac{\delta_1}{k_{s+}}\right)C_1 + 2\frac{k_{1-}}{k_{s+}}D_1 + \frac{k_{12-}}{k_{s+}}D_{12} \tag{B.14}$$

$$\frac{dD_1}{d\tau} = \frac{1}{k_{s+}}\left(k_{1+}C_1^2 - (k_{1-} + a_{1+})D_1 + a_{1-}D_1^* - \delta_1 D_1\right) \tag{B.15}$$

$$\frac{dD_{12}}{d\tau} = \frac{1}{k_{s+}}(k_{12+}R_2C_1 - (k_{12-} + a_{12+})D_{12} + a_{12-}D_{12}^* - \delta_{12}D_{12}) \tag{B.16}$$

$$\frac{dD_2}{d\tau} = \frac{1}{k_{s+}}\left(k_{2+}R_2^2 - (k_{2-} + a_{2+})D_2 + a_{2-}D_2^* - \delta_2 D_2\right) \tag{B.17}$$

$$\frac{dD_1^*}{d\tau} = \frac{1}{k_{s+}}(a_{1+}D_1 - a_{1-}D_1^* - \delta_1 D_1^*) \tag{B.18}$$

$$\frac{dD_{12}^*}{d\tau} = \frac{1}{k_{s+}}(a_{12+}D_{12} - a_{12-}D_{12}^* - \delta_{12}D_{12}^*) \tag{B.19}$$

$$\frac{dD_2^*}{d\tau} = \frac{1}{k_{s+}}(a_{2+}D_2 - a_{2-}D_2^* - \delta_2 D_2^*) \tag{B.20}$$

In the shortest time scale associated to ligand binding ($k_{s+} = 10^7$ min^{-1}), the only significative reaction is ligand binding. All the other reactions have rates of the order $r = 10^4$ min^{-1} or slower. At this time scale, the system can be simplified to

$$\frac{dS}{d\tau} = -SR_1 \tag{B.21}$$
$$\frac{dR_1}{d\tau} = -SR_1 \tag{B.22}$$
$$\frac{dR_2}{d\tau} = 0 \tag{B.23}$$
$$\frac{dC_1}{d\tau} = SR_1 \tag{B.24}$$
$$\frac{dD_1}{d\tau} = 0 \tag{B.25}$$
$$\frac{dD_{12}}{d\tau} = 0 \tag{B.26}$$
$$\frac{dD_2}{d\tau} = 0 \tag{B.27}$$
$$\frac{dD_1^*}{d\tau} = 0 \tag{B.28}$$
$$\frac{dD_{12}^*}{d\tau} = 0 \tag{B.29}$$
$$\frac{dD_2^*}{d\tau} = 0 \tag{B.30}$$

In order to study the dynamics of the system at the intermediate time scale where receptor dimerization occurs ($k_{12+} = 10^3$ min $^{-1}$), we nondimensionalize using $\tau = tk_{12+}$.

$$\frac{dS}{d\tau} = -\frac{k_{s+}}{k_{12+}}SR_1 + \frac{k_{s-}}{k_{12+}}C_1 \quad \text{(B.31)}$$

$$\frac{dR_1}{d\tau} = \frac{m_1}{k_{12+}} - \frac{k_{s+}}{k_{12+}}SR_1 + \frac{k_{s-}}{k_{12+}}C_1 - \frac{\delta_1}{k_{12+}}R_1 \quad \text{(B.32)}$$

$$\frac{dR_2}{d\tau} = \frac{m_2}{k_{12+}} - \frac{k_{12+}}{k_{12+}}C_1R_2 - 2\frac{k_{2+}}{k_{12+}}R_2^2 + K_{d,12}D_{12} + 2K_{d,2}D_2 - \frac{\delta_2}{k_{12+}}R_2 \quad \text{(B.33)}$$

$$\frac{dC_1}{d\tau} = \frac{k_{s+}}{k_{12+}}SR_1 - (\frac{k_{s-}}{k_{12+}} + 2\frac{k_{1+}}{k_{12+}}C_1 + \frac{k_{12+}}{k_{12+}}R_2 - \frac{\delta_1}{k_{12+}})C_1 + 2K_{d,1}D_1 + K_{d,12}D_{12} \quad \text{(B.34)}$$

$$\frac{dD_1}{d\tau} = \frac{k_{1+}}{k_{12+}}C_1^2 - (K_{d,1} + \frac{a_{1+}}{k_{12+}})D_1 + \frac{a_{1-}}{k_{12+}}D_1^* - \frac{\delta_1}{k_{12+}}D_1 \quad \text{(B.35)}$$

$$\frac{dD_{12}}{d\tau} = \frac{k_{12+}}{k_{12+}}R_2C_1 - (K_{d,12} + \frac{a_{12+}}{k_{12+}})D_{12} + \frac{a_{12-}}{k_{12+}}D_{12}^* - \frac{\delta_{12}}{k_{12+}}D_{12} \quad \text{(B.36)}$$

$$\frac{dD_2}{d\tau} = \frac{k_{2+}}{k_{12+}}R_2^2 - (K_{d,2} + \frac{a_{2+}}{k_{12+}})D_2 + \frac{a_{2-}}{k_{12+}}D_2^* - \frac{\delta_2}{k_{12+}}D_2 \quad \text{(B.37)}$$

$$\frac{dD_1^*}{d\tau} = \frac{1}{k_{12+}}(a_{1+}D_1 - a_{1-}D_1^* - \delta_1 D_1^*) \quad \text{(B.38)}$$

$$\frac{dD_{12}^*}{d\tau} = \frac{1}{k_{12+}}(a_{12+}D_{12} - a_{12-}D_{12}^* - \delta_{12}D_{12}^*) \quad \text{(B.39)}$$

$$\frac{dD_2^*}{d\tau} = \frac{1}{k_{12+}}(a_{2+}D_2 - a_{2-}D_2^* - \delta_2 D_2^*) \quad \text{(B.40)}$$

For the values of the parameters shown in Table 2.1, at this time scale, $\frac{k_{s-}}{k_{12+}} \sim 10^{-6}$, $\frac{m_1}{k_{12+}} \sim 10^{-1}$, $\frac{m_2}{k_{12+}} \sim 10^{-3}$, $\frac{\delta_i}{k_{12+}} \sim 10^{-6}$, $\frac{k_{i+}}{k_{12+}} \sim 1$ $K_{d,i} \sim 10^{-4}$, $K_{d,2} \sim 10$, $\frac{a_{i+}}{k_{12+}} \sim 10^{-3}$ and $\frac{a_{2-}}{k_{12+}} \sim 10^{-2}$. Neglecting all the reactions with rates slower than 10^{-1}, we get

$$\frac{dS}{d\tau} = 0 \tag{B.41}$$

$$\frac{dR_1}{d\tau} = 0 \tag{B.42}$$

$$\frac{dR_2}{d\tau} = -C_1 R_2 - 2R_2^2 + 2K_{d,2} D_2 \tag{B.43}$$

$$\frac{dC_1}{d\tau} = -(2C_1 + R_2)C_1 \tag{B.44}$$

$$\frac{dD_1}{d\tau} = C_1^2 \tag{B.45}$$

$$\frac{dD_{12}}{d\tau} = R_2 C_1 \tag{B.46}$$

$$\frac{dD_2}{d\tau} = R_2^2 - K_{d,2} D_2 \tag{B.47}$$

$$\frac{dD_1^*}{d\tau} = 0 \tag{B.48}$$

$$\frac{dD_{12}^*}{d\tau} = 0 \tag{B.49}$$

$$\frac{dD_2^*}{d\tau} = 0 \tag{B.50}$$

A third time scale is that associated to dimer activation. To explore this regime, we nondimensionalize the system with $\tau = ta_{12+}$ in order to simplify the system at the time scale of activation of the dimers.

$$\frac{dS}{d\tau} = -\frac{k_{s+}}{a_{12+}}SR_1 + \frac{k_{s-}}{a_{12+}}C_1 \quad \text{(B.51)}$$

$$\frac{dR_1}{d\tau} = \frac{m_1}{a_{12+}} - \frac{k_{s+}}{a_{12+}}SR_1 + \frac{k_{s-}}{a_{12+}}C_1 - \frac{\delta_1}{a_{12+}}R_1 \quad \text{(B.52)}$$

$$\frac{dR_2}{d\tau} = \frac{m_2}{a_{12+}} - \frac{k_{12+}}{a_{12+}}C_1R_2 - 2\frac{k_{2+}}{a_{12+}}R_2^2 + \frac{k_{12-}}{a_{12+}}D_{12} + 2\frac{k_{2-}}{a_{12+}}D_2 - \frac{\delta_2}{a_{12+}}R_2 \quad \text{(B.53)}$$

$$\frac{dC_1}{d\tau} = \frac{k_{s+}}{a_{12+}}SR_1 - (\frac{k_{s-}}{a_{12+}} + 2\frac{k_{1+}}{a_{12+}}C_1 + \frac{k_{12+}}{a_{12+}}R_2 - \frac{\delta_1}{a_{12+}})C_1 + 2\frac{k_{1-}}{a_{12+}}D_1 + \frac{k_{12-}}{a_{12+}}D_{12} \quad \text{(B.54)}$$

$$\frac{dD_1}{d\tau} = \frac{k_{1+}}{a_{12+}}C_1^2 - (\frac{k_{1-}}{a_{12+}} + \frac{a_{1+}}{a_{12+}})D_1 + \frac{a_{1-}}{D}^*_1 - \frac{\delta_1}{a_{12+}}D_1 \quad \text{(B.55)}$$

$$\frac{dD_{12}}{d\tau} = \frac{k_{12+}}{a_{12+}}R_2C_1 - (\frac{k_{12-}}{a_{12+}} + \frac{a_{12+}}{a_{12+}})D_{12} + \frac{a_{12-}}{a_{12+}}D_{12}^* - \frac{\delta_{12}}{a_{12+}}D_{12} \quad \text{(B.56)}$$

$$\frac{dD_2}{d\tau} = \frac{k_{2+}}{a_{12+}}R_2^2 - (\frac{k_{2-}}{a_{12+}} + \frac{a_{2+}}{a_{12+}})D_2 + \frac{a_{2-}}{a_{12+}}D_2^* - \frac{\delta_2}{a_{12+}}D_2 \quad \text{(B.57)}$$

$$\frac{dD_1^*}{d\tau} = \frac{a_{1+}}{a_{12+}}D_1 - \frac{a_{1-}}{a_{12+}}D_1^* - \frac{\delta_1}{a_{12+}}D_1^* \quad \text{(B.58)}$$

$$\frac{dD_{12}^*}{d\tau} = \frac{a_{12+}}{a_{12+}}D_{12} - \frac{a_{12-}}{a_{12+}}D_{12}^* - \frac{\delta_{12}}{a_{12+}}D_{12}^* \quad \text{(B.59)}$$

$$\frac{dD_2^*}{d\tau} = \frac{a_{2+}}{a_{12+}}D_2 - \frac{a_{2-}}{a_{12+}}D_2^* - \frac{\delta_2}{a_{12+}}D_2^* \quad \text{(B.60)}$$

For the values of the parameters shown in Table 2.1, at this time scale, $\frac{k_{s-}}{a_{12+}} \sim 10^{-3}$, $\frac{m_1}{a_{12+}} \sim 10^3$, $\frac{m_2}{a_{12+}} \sim 10$, $\frac{\delta_i}{a_{12+}} \sim 10^{-3}$, $\frac{k_{i+}}{a_{12+}} \sim 10^4$ $\frac{k_{i-}}{a_{12+}} \sim 10^{-1}$, $\frac{k_{2-}}{a_{12+}} \sim 10^4$, $\frac{a_{i+}}{a_{12+}} \sim 1$ and $\frac{a_{2-}}{a_{12+}} \sim 10$. Considering only reactions with rate of order 1, we get

$$\frac{dS}{d\tau} = 0 \tag{B.61}$$

$$\frac{dR_1}{d\tau} = 0 \tag{B.62}$$

$$\frac{dR_2}{d\tau} = 0 \tag{B.63}$$

$$\frac{dC_1}{d\tau} = 0 \tag{B.64}$$

$$\frac{dD_1}{d\tau} = -\frac{a_{1+}}{a_{12+}} D_1 \tag{B.65}$$

$$\frac{dD_{12}}{d\tau} = -\frac{a_{12+}}{a_{12+}} D_{12} \tag{B.66}$$

$$\frac{dD_2}{d\tau} = -\frac{a_{2+}}{a_{12+}} D_2 + \frac{a_{2-}}{a_{12+}} D_2^* \tag{B.67}$$

$$\frac{dD_1^*}{d\tau} = \frac{a_{1+}}{a_{12+}} D_1 \tag{B.68}$$

$$\frac{dD_{12}^*}{d\tau} = \frac{a_{12+}}{a_{12+}} D_{12} \tag{B.69}$$

$$\frac{dD_2^*}{d\tau} = \frac{a_{2+}}{a_{12+}} D_2 - \frac{a_{2-}}{a_{12+}} D_2^* \tag{B.70}$$

Finally, at the time scale of degradation where only the dimers are evolving we have (with $\tau = t\delta_{12}$)

$$\frac{dS}{d\tau} = \frac{1}{\delta_{12}}\left(-k_{s+}SR_1 + k_{s-}C_1\right) \quad \text{(B.71)}$$

$$\frac{dR_1}{d\tau} = \frac{1}{\delta_{12}}\left(m_1 - k_{s+}SR_1 + k_{s-}C_1 - \delta_1 R_1\right) \quad \text{(B.72)}$$

$$\frac{dR_2}{d\tau} = \frac{1}{\delta_{12}}\left(m_2 - k_{12+}C_1R_2 - 2k_{2+}R_2^2 + k_{12-}D_{12} + 2k_{2-}D_2 - \delta_2 R_2\right) \quad \text{(B.73)}$$

$$\frac{dC_1}{d\tau} = \frac{1}{\delta_{12}}\left(k_{s+}SR_1 - (k_{s-} + 2k_{1+}C_1 + k_{12+}R_2 - \delta_1)C_1 + 2k_{1-}D_1 + k_{12-}D_{12}\right) \quad \text{(B.74)}$$

$$\frac{dD_1}{d\tau} = \frac{1}{\delta_{12}}\left(k_{1+}C_1^2 - (k_{1-} + a_{1+})D_1 + a_{1-}D_1^* - \delta_1 D_1\right) \quad \text{(B.75)}$$

$$\frac{dD_{12}}{d\tau} = \frac{1}{\delta_{12}}\left(k_{12+}R_2C_1 - (k_{12-} + a_{12+})D_{12} + a_{12-}D_{12}^* - \delta_{12}D_{12}\right) \quad \text{(B.76)}$$

$$\frac{dD_2}{d\tau} = \frac{1}{\delta_{12}}\left(k_{2+}R_2^2 - (k_{2-} + a_{2+})D_2 + a_{2-}D_2^* - \delta_2 D_2\right) \quad \text{(B.77)}$$

$$\frac{dD_1^*}{d\tau} = \frac{1}{\delta_{12}}\left(a_{1+}D_1 - a_{1-}D_1^* - \delta_1 D_1^*\right) \quad \text{(B.78)}$$

$$\frac{dD_{12}^*}{d\tau} = \frac{1}{\delta_{12}}\left(a_{12+}D_{12} - a_{12-}D_{12}^* - \delta_{12}D_{12}^*\right) \quad \text{(B.79)}$$

$$\frac{dD_2^*}{d\tau} = \frac{1}{\delta_{12}}\left(a_{2+}D_2 - a_{2-}D_2^* - \delta_2 D_2^*\right) \quad \text{(B.80)}$$

For the values of the parameters shown in Table 2.1, at this time scale, $\frac{m_1}{\delta_{12}} \sim 10^6$, $\frac{m_2}{\delta_{12}} \sim 10$, $\frac{\delta_i}{\delta_{12}} \sim 1$, $\frac{k_{i+}}{\delta_{12}} \sim 10^7$ $\frac{k_{i-}}{\delta_{12}} \sim 10^2$, $\frac{k_{2-}}{\delta_{12}} \sim 10^7$, $\frac{a_{i+}}{\delta_{12}} \sim 10^3$ and $\frac{a_{2-}}{\delta_{12}} \sim 10^4$. Considering only reactions with rate of order 1, we get

$$\frac{dS}{d\tau} = 0 \tag{B.81}$$

$$\frac{dR_1}{d\tau} = -\delta_1 R_1 \tag{B.82}$$

$$\frac{dR_2}{d\tau} = -\delta_2 R_2 \tag{B.83}$$

$$\frac{dC_1}{d\tau} = -\delta_1 C_1 \tag{B.84}$$

$$\frac{dD_1}{d\tau} = -\delta_1 D_1 \tag{B.85}$$

$$\frac{dD_{12}}{d\tau} = -\delta_{12} D_{12} \tag{B.86}$$

$$\frac{dD_2}{d\tau} = -\delta_2 D_2 \tag{B.87}$$

$$\frac{dD_1^*}{d\tau} = -\delta_1 D_1^* \tag{B.88}$$

$$\frac{dD_{12}^*}{d\tau} = -\delta_{12} D_{12}^* \tag{B.89}$$

$$\frac{dD_2^*}{d\tau} = -\delta_2 D_2^* \tag{B.90}$$

B.2 Stochastic multiscale analysis

In Chapter 2, Eqs. 2.47-2.52, we consider a deterministic system with different time scales. In the slow time scale, the receptors are activated in an irreversible reaction and then degraded, following an exponential decay. During the receptors' active period, intracellular proteins can bind an unbind the receptors' binding sites. We consider the following set of stochastic reactions to give a stochastic version of the system.

$$\emptyset \xrightarrow{a_{+i} D_i} D_i^* \tag{B.91}$$

$$D_i^* \xrightarrow{\delta_{D_i^*}} \emptyset \tag{B.92}$$

$$D_i^* + \Pi \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \Pi^* \tag{B.93}$$

where D_i^* are the active receptor, Π are the unbound proteins and Π^* are the bound proteins. We consider that the proteins are in excess, so Π can be considered as constant.

The master equation describing the state of the system is

$$\frac{\partial P(D_i^*, \Pi^*, t)}{\partial t} = \sum_i (W_i(X - r_i)P(X - r_i, t) - W_i(X)P(X, t)) \tag{B.94}$$

where X is the vector of variables and r_i indicates the stoichiometry of reaction i and W_i is the rate at which reaction i happens.

In this analysis, we will consider only one binding site and one proteins. The result can be extended to several binding sites and proteins [103].

The rates of the reaction considered in Eqs. B.91-B.93 are

Rate	Reaction	Parameter value
$\mathrm{W}_1 = a_+ D_i$	$\mathrm{D}_i^* \to D_i^* + 1$	a_+=1 min^{-1}
$\mathrm{W}_2 = \delta D_i^*$	$\mathrm{D}_i^* \to D_i^* - 1$	$\delta = 10^{-3}$ min^{-1}
$\mathrm{W}_3 = k_+ D_i^* \Pi$	$\mathrm{D}_i^* \to D_i^* - 1$ $\Pi \to \Pi - 1$ $\Pi^* \to \Pi^* + 1$	$k_+ = 10^2$ min^{-1}
$\mathrm{W}_4 = k_- \Pi^*$	$\mathrm{D}_i^* \to D_i^* + 1$ $\Pi \to \Pi + 1$ $\Pi^* \to \Pi^* - 1$	$k_- = 10^1$ min^{-1}

Let's define $\epsilon = \frac{\delta}{k_+} (>> 1)$ and rescale the system with $k_+ = \frac{\kappa_+}{\epsilon}$ and $k_- = \frac{\kappa_-}{\epsilon}$.

$$w_1 = \frac{a_+ D_i^*}{\delta} \tag{B.95}$$

$$w_2 = D_i^* \tag{B.96}$$

$$w_3 = \kappa_+ D_i^* \Pi \tag{B.97}$$

$$w_4 = \kappa_- \Pi^* \tag{B.98}$$

With this rescaling, we realise that the activation and degradation of the receptor are slow reactions compared to the binding and unbinding reactions. We can rewrite Eq. (B.94) as

$$\begin{aligned}\frac{\partial P(D_i^*,\Pi^*,t)}{\partial t} =& \sum_{i=1}^{2}(w_i(X-r_i)P(X-r_i,t)-w_i(X)P(X,t)) \\ &+\frac{1}{\epsilon}\sum_{i=3}^{4}(w_i(X-r_i)P(X-r_i,t)-w_i(X)P(X,t))\end{aligned} \tag{B.99}$$

Let's define the following generating function

$$G(p,t)=\sum_X p^X P(X,t) \tag{B.100}$$

As discussed before, activation and degradation of the receptors are slow compared to the binding and unbinding reactions. At this time scale, only the active receptors, D_i^*, change over time. On the other hand, binding and unbinding happen at a fast time scale and involve the active receptors that switch between free (D_i^*) and bound (Π^*) configurations. At the fast time scale, the total amount of active receptors (free and bound) is considered constant, but Π^* changes, so we consider it as a fast variable. This allows us to express G as a sum of slow and fast processes involving slow and fast variables, respectively.

$$\frac{\partial G(p,t)}{\partial t}=H_0(p^{(S)},\partial p^{(S)})G(p^{(S)},p^{(F)})+\frac{1}{\epsilon}H_1(p^{(F)},\partial p^{(F)})G(p^{(S)},p^{(F)}), \tag{B.101}$$

where the supraindexes S, F denote slow and fast variables, respectively.

$$\epsilon\frac{\partial G(p,t)}{\partial t}=\epsilon H_0(p^{(S)},\partial p^{(S)})G(p^{(S)},p^{(F)})+H_1(p^{(F)},\partial p^{(F)})G(p^{(S)},p^{(F)}) \tag{B.102}$$

The operators H_0 and H_1 are

$$H_0=(a_+p^{(S)}-\delta)(p^{(S)}-1)\partial p \tag{B.103}$$

$$H_1=(p-1)\left(k_+\Pi\left((D^*+\Pi^*)-pq\right)-k_-q\right) \tag{B.104}$$

where $q = \partial p$.

Let us express G as a power series of ϵ

$$G = G_0 + \epsilon G_1 + (O)(\epsilon) \tag{B.105}$$

Keeping the 0^{th} order term, we have

$$H_1 G_0 = 0. \tag{B.106}$$

Due to the time scale separation, G_0 can be factorized as

$$G_0(p^{(S)}, p^{(F)}) = g_s(p^{(S)}, t) g_F(p^{(F)}) \tag{B.107}$$

This result, togehter with Eq. (B.106), yields

$$g_s(p^{(S)}, t) H_1(p_i^{(F)}, \partial p_i^{(F)}) g_F(p^{(F)}) = 0 \tag{B.108}$$

This holds for $H_1(p_i^{(F)}, \partial p_i^{(F)}) g_F(p^{(F)}) = 0$.

If we consider all the possible binding and unbinding reactions for n_G different binding sites and n_P different proteins, the operator H_1 is given by [103]

$$H_1 = \sum_{i=1}^{n_G} \sum_{j=1}^{n_P} (p_{ij} - 1) \left(k_{ij+} P_j \left(n_{iT} - \sum_{k=1}^{n_P} p_{ik} q_{ik} \right) - k_{ij-} q_{ij} \right), \tag{B.109}$$

where $q_{ij} \equiv_{ij}$.

This equation can be solved using WKB methods [103], to get

$$G(p_{ij}, t = t_{ss}) = \prod_{i=1}^{n_G} \left(\frac{1 + \sum_{k=1}^{n_P} K_{ik} p_{ik}}{1 + \sum_{k=1}^{n_P} K_{ik}} \right)^{n_{iT}} \tag{B.110}$$

This generating function corresponds to the product of n_G independent multinomials, one for each type of binding site (binding sites are supposed independent and the intracellular proteins are supposed in excess).

Appendix C

Numerical simulations

C.1 Numerical integration of ODE systems

The numerical integration of the ODE systems in Chapter 2 has been performed using the integration method 'lsoda', implemented in SciPy [104]. The values of the parameters used are listed in Table 2.1.

C.2 Stochastic Gillespie simulations

In Chapters 3 and 4, we compared our theoretical predictions with the results of stochastic Gillespie simulations [87].

We simulated different situations (considering or not protein activation and receptor decay) for populations of either one or two intracellular proteins.

In Table C.1, we list all the possible reactions that we considered in our simulations (together with the corresponding rates, that follow the law of mass action [105]). The values of the parameters are indicated in the captions of the figures where theoretical results and stochastic simulations are compared.

In Table C.1, r_{Fi} is the number of free binding sites of type i, r_{ij} is the number of binding sites of type i bound to a protein of type j, r^*_{ij} is the number of binding sites of type i bound to a protein of type j that have undergone an activation of the protein, k_+ is the binding rate, P_{tot} is the concentration of intracellular protein (assumed to be constant), k_- is the unbinding rate and δ is the degradation rate.

	Reaction	Rate
1. Binding	$\mathrm{r}_{Fi} \to r_{Fi} - 1$ $\mathrm{r}_{ij} \to r_{ij} + 1$	$k_+ P_{tot} r_{Fi}$
2. Unbinding	$\mathrm{r}_{Fi} \to r_{Fi} + 1$ $\mathrm{r}_{ij} \to r_{ij} - 1$	$k_- r_{ij}$
3. Protein activation	$\mathrm{r}_{ij} \to r_{ij} - 1$ $\mathrm{r}^*_{ij} \to r^*_{ij} + 1$	αr_{ij}
4. Active protein dissociation	$\mathrm{r}^*_{ij} \to r^*_{ij} - 1$ $\mathrm{r}_{Fi} \to r_{Fi} + 1$	$k_- r^*_{ij}$
5. Receptor degradation	$\mathrm{r}_{Fi} \to r_{Fi} - 1$	δr_{Fi}
6. Bound receptor degradation	$\mathrm{r}_{ij} \to r_{ij} - 1$	δr_{ij}
7. Bound (active) receptor degradation	$\mathrm{r}^*_{ij} \to r^*_{ij} - 1$	δr^*_{ij}

TABLE C.1: Reactions and rates of the stochastic Gillespie simulations.

Chapter 3, Section 3.1

In these sections, we described the probability distribution of the number of bound proteins at a given time. Therefore, we do not consider degradation of the receptors, nor activation of the proteins: only binding and unbinding reactions (1 and 2 in Table C.1) are considered. In order to get the computational distributions, we performed 1000 iterations of the stochastic simulation, starting with $r_{Fi} = R_T$ (with R_T being the number of active receptors) and letting the system reach equilibrium (in a time scale $\sim \frac{1}{k_+ P_{tot}}$ before recording the system's state.

Chapter 3, Section 3.2

In this section, we are interested in the number of proteins that have bound or have been activated by the active receptors during their lifetimes. We consider therefore all the reactions in Table C.1. We performed 100 iterations of the corresponding stochastic simulation, where the simulation is run until all the active receptors have been degraded. In this time course, every binding event and every activation event is recorded in order to generate the distribution of bound and active intracellular proteins.

Bibliography

[1] Francis HC Crick. On protein synthesis. In *Symp Soc Exp Biol*, volume 12, page 8, 1958.

[2] Johann S de Bono and Eric K Rowinsky. The erbb receptor family: a therapeutic target for cancer. *Trends in molecular medicine*, 8(4):S19–S26, 2002.

[3] José Baselga and Carlos L Arteaga. Critical update and emerging trends in epidermal growth factor receptor targeting in cancer. *Journal of Clinical Oncology*, 23(11):2445–2459, 2005.

[4] Esther Zwick, Johannes Bange, and Axel Ullrich. Receptor tyrosine kinase signalling as a target for cancer intervention strategies. *Endocrine-related cancer*, 8(3):161–173, 2001.

[5] Carlos L Arteaga and Jeffrey A Engelman. Erbb receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer cell*, 25(3):282–303, 2014.

[6] Mark A Lemmon and Joseph Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 141(7):1117–1134, 2010.

[7] Ami Citri and Yosef Yarden. Egf–erbb signalling: towards the systems level. *Nature reviews Molecular cell biology*, 7(7):505, 2006.

[8] Diana Graus-Porta, Roger R Beerli, John M Daly, and Nancy E Hynes. Erbb-2, the preferred heterodimerization partner of all erbb receptors, is a mediator of lateral signaling. *The EMBO journal*, 16(7):1647–1655, 1997.

[9] David J Riese and David F Stern. Specificity within the egf family/erbb receptor family signaling network. *Bioessays*, 20(1):41–48, 1998.

[10] Thomas PJ Garrett, Neil M McKern, Meizhen Lou, Thomas C Elleman, Timothy E Adams, George O Lovrecz, Michael Kofler, Robert N Jorissen, Edouard C Nice, Antony W Burgess, et al. The crystal structure of a truncated erbb2 ectodomain reveals an active conformation, poised to interact with other erbb receptors. *Molecular cell*, 11(2):495–505, 2003.

[11] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[12] Jose Baselga and Sandra M Swain. Novel anticancer targets: revisiting erbb2 and discovering erbb3. *Nature Reviews Cancer*, 9(7):463, 2009.

[13] Rita Nahta and Francisco J Esteva. Herceptin: mechanisms of action and resistance. *Cancer letters*, 232(2):123–138, 2006.

[14] Yosef Yarden and Mark X Sliwkowski. Untangling the erbb signalling network. *Nature reviews Molecular cell biology*, 2(2):127, 2001.

[15] Robert M Hudziak, Gail D Lewis, Marcy Winget, Brian M Fendly, H Michael Shepard, and A Ullrich. p185her2 monoclonal antibody has antiproliferative effects in vitro and sensitizes human breast tumor cells to tumor necrosis factor. *Molecular and cellular biology*, 9(3):1165–1172, 1989.

[16] Paul H Huang, Akitake Mukasa, Rudy Bonavia, Ryan A Flynn, Zachary E Brewer, Webster K Cavenee, Frank B Furnari, and Forest M White. Quantitative analysis of egfrviii cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proceedings of the National Academy of Sciences*, 104(31):12867–12872, 2007.

[17] Rita Nahta and Francisco J Esteva. Her-2-targeted therapy: lessons learned and future directions. *Clinical Cancer Research*, 9(14):5078–5084, 2003.

[18] Rita Nahta, Dihua Yu, Mien-Chie Hung, Gabriel N Hortobagyi, and Francisco J Esteva. Mechanisms of disease: understanding resistance to her2-targeted therapy in human breast cancer. *Nature Reviews Clinical Oncology*, 3(5):269, 2006.

[19] Alexander M Xu and Paul H Huang. Receptor tyrosine kinase coactivation networks in cancer. *Cancer research*, 70(10):3857–3860, 2010.

[20] Jayne M Stommel, Alec C Kimmelman, Haoqiang Ying, Roustem Nabioullin, Aditya H Ponugoti, Ruprecht Wiedemeyer, Alexander H Stegh, James E Bradner, Keith L Ligon, Cameron Brennan, et al. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science*, 318(5848):287–290, 2007.

[21] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2013.

[22] Kurtis E Bachman, Pedram Argani, Yardena Samuels, Natalie Silliman, Janine Ptak, Steve Szabo, Hiroyuki Konishi, Bedri Karakas, Brian G Blair, Clarence Lin, et al. The pik3ca gene is mutated with high frequency in human breast cancers. *Cancer biology & therapy*, 3(8):772–775, 2004.

[23] Richard B Jones, Andrew Gordus, Jordan A Krall, and Gavin MacBeath. A quantitative protein interaction network for the erbb receptors using protein microarrays. *Nature*, 439(7073):168, 2006.

[24] Nancy E Hynes and Gwen MacDonald. Erbb receptors and signaling pathways in cancer. *Current opinion in cell biology*, 21(2):177–184, 2009.

[25] Waltraud X Schulze, Lei Deng, and Matthias Mann. Phosphotyrosine interactome of the erbb-receptor kinase family. *Molecular systems biology*, 1(1), 2005.

[26] Johannes L Bos, Eric R Fearon, Stanley R Hamilton, Matty Verlaan-de Vries, Jacques H van Boom, Alex J van der Eb, and Bert Vogelstein. Prevalence of ras gene mutations in human colorectal cancers. *Nature*, 327(6120):293, 1987.

[27] Roger J Daly, Michele D Binder, and Robert L Sutherland. Overexpression of the grb2 gene in human breast cancer cell lines. *Oncogene*, 9(9):2723–2727, 1994.

[28] Peter W Janes, Roger J Daly, RL Sutherland, et al. Activation of the ras signalling pathway in human breast cancer cells overexpressing erbb-2. *Oncogene*, 9(12):3601–3608, 1994.

[29] Leon O Murphy and John Blenis. Mapk signal specificity: the right place at the right time. *Trends in biochemical sciences*, 31(5):268–275, 2006.

[30] CJ Marshall. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell*, 80(2): 179–185, 1995.

[31] Axel Obermeier, Ralph A Bradshaw, Klaus Seedorf, Axel Choidas, Joseph Schlessinger, and Axel Ullrich. Neuronal differentiation signals are controlled by nerve growth factor receptor/trk binding sites for shc and plc gamma. *The EMBO Journal*, 13(7):1585–1590, 1994.

[32] Heng Xie, Manuel A Pallero, Kiran Gupta, Philip Chang, Margaret F Ware, Walter Witke, David J Kwiatkowski, Douglas A Lauffenburger, Joanne E Murphy-Ullrich, and Alan Wells. Egf receptor regulation of cell motility: Egf induces disassembly of focal adhesions independently of the motility-associated plcgamma signaling pathway. *Journal of cell science*, 111(5): 615–624, 1998.

[33] Monilola A Olayioye, Iwan Beuvink, Kay Horsch, John M Daly, and Nancy E Hynes. Erbb receptor-induced activation of stat transcription factors is mediated by src tyrosine kinases. *Journal of Biological Chemistry*, 274(24): 17209–17218, 1999.

[34] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47(2):81–116, 1972.

[35] Douglas A Lauffenburger and Jennifer J Linderman. *Receptors: models for binding, trafficking, and signaling*. Oxford University Press on Demand, 1996.

[36] Antoine M Van Oijen. Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Current opinion in biotechnology*, 22(1):75–80, 2011.

[37] Takeo Miyake, Takashi Tanii, Hironori Sonobe, Rena Akahori, Naonobu Shimamoto, Taro Ueno, Takashi Funatsu, and Iwao Ohdomari. Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein- protein interaction. *Analytical chemistry*, 80(15):6018–6022, 2008.

[38] Mark P Elenko, Jack W Szostak, and Antoine M van Oijen. Single-molecule imaging of an in vitro-evolved rna aptamer reveals homogeneous ligand binding kinetics. *Journal of the American Chemical Society*, 131(29):9866–9867, 2009.

[39] Richard B Jones. Do low-affinity erbb receptor protein interactions represent the base of a cell signaling iceberg? *Expert Review of Proteomics*, 10(2):115–118, 2013. doi: 10.1586/epr.12.78.

[40] Leonor Michaelis and Maud Leonora Menten. *Die kinetik der invertinwirkung*.

[41] Shlomi Reuveni, Michael Urbakh, and Joseph Klafter. Role of substrate unbinding in michaelis–menten enzymatic reactions. *Proceedings of the National Academy of Sciences*, 111(12):4391–4396, 2014.

[42] H Steven Wiley and Dennis D Cunningham. A steady state model for analyzing the cellular binding, internalization and degradation of polypeptide ligands. *Cell*, 25(2):433–440, 1981.

[43] Boris N. Kholodenko, Oleg V. Demin, Gisela Moehren, and Jan B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *Journal of Biological Chemistry*, 274(42):30169–30181, 1999. doi: 10.1074/jbc.274.42.30169.

[44] Birgit Schoeberl, Claudia Eichler-Jonsson, Ernst Dieter Gilles, and Gertraud Müller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature biotechnology*, 20 (4):370, 2002.

[45] Tomáš Helikar, Naomi Kochi, Bryan Kowal, Manjari Dimri, Mayumi Naramura, Srikumar M Raja, Vimla Band, Hamid Band, and Jim A Rogers. A comprehensive, multi-scale dynamical model of erbb receptor signal transduction in human mammary epithelial cells. *PLoS One*, 8(4):e61757, 2013.

[46] Bart S Hendriks, Gayla Orr, Alan Wells, H Steven Wiley, and Douglas A Lauffenburger. Parsing erk activation reveals quantitatively equivalent contributions from epidermal growth factor receptor and her2 in human mammary epithelial cells. *Journal of Biological Chemistry*, 280(7):6157–6169, 2005.

[47] Harish Shankaran, Yi Zhang, Lee Opresko, and Haluk Resat. Quantifying the effects of co-expressing egfr and her2 on her activation and trafficking. *Biochemical and biophysical research communications*, 371(2):220–224, 2008.

[48] Yi Zhang, Harish Shankaran, Lee Opresko, and Haluk Resat. System theoretical investigation of human epidermal growth factor receptor-mediated signalling. *IET systems biology*, 2(5):273–284, 2008.

[49] Takashi Nakakuki, Noriko Yumoto, Takashi Naka, Mikako Shirouzu, Shigeyuki Yokoyama, and Mariko Hatakeyama. Topological analysis of mapk cascade for kinetic erbb signaling. *PLoS One*, 3(3):e1782, 2008.

[50] Michael L Blinov, James R Faeder, Byron Goldstein, and William S Hlavacek. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems*, 83 (2-3):136–151, 2006.

[51] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[52] Adrienne Fairhall, Eric Shea-Brown, and Andrea Barreiro. Information theoretic approaches to understanding circuit function. *Current opinion in neurobiology*, 22(4):653–659, 2012.

[53] Clive G Bowsher and Peter S Swain. Environmental sensing, information transfer, and cellular decision-making. *Current opinion in biotechnology*, 28: 149–155, 2014.

[54] Andre Levchenko and Ilya Nemenman. Cellular noise and information transmission. *Current opinion in biotechnology*, 28:156–164, 2014.

[55] Gašper Tkačik, Curtis G Callan Jr, and William Bialek. Information capacity of genetic regulatory elements. *Physical Review E*, 78(1):011910, 2008.

[56] Gašper Tkačik, Curtis G Callan, and William Bialek. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences*, 105(34):12265–12270, 2008.

[57] Gašper Tkačik, Aleksandra M Walczak, and William Bialek. Optimizing information flow in small genetic networks. *Physical Review E*, 80(3):031920, 2009.

[58] Aleksandra M Walczak, Gašper Tkačik, and William Bialek. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Physical Review E*, 81(4):041905, 2010.

[59] Gašper Tkačik, Aleksandra M Walczak, and William Bialek. Optimizing information flow in small genetic networks. iii. a self-interacting gene. *Physical Review E*, 85(4):041903, 2012.

[60] Thomas R Sokolowski and Gašper Tkačik. Optimizing information flow in small genetic networks. iv. spatial coupling. *Physical Review E*, 91(6):062710, 2015.

[61] Gašper Tkačik and Aleksandra M Walczak. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, 23(15):153102, 2011.

[62] Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, and Andre Levchenko. Information transduction capacity of noisy biochemical signaling networks. *Science*, 334(6054):354–358, 2011. ISSN 0036-8075. doi: 10.1126/science.1204553.

[63] Shinsuke Uda, Takeshi H Saito, Takamasa Kudo, Toshiya Kokaji, Takaho Tsuchiya, Hiroyuki Kubota, Yasunori Komori, Yu-ichi Ozaki, and Shinya Kuroda. Robustness and compensation of information transmission of signaling pathways. *Science*, 341(6145):558–561, 2013.

[64] Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, and Roy Wollman. Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373, 2014.

[65] Margaritis Voliotis, Rebecca M Perrett, Chris McWilliams, Craig A McArdle, and Clive G Bowsher. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proceedings of the National Academy of Sciences*, 111(3):E326–E333, 2014.

[66] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbersthe database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753, 2009.

[67] Mark F Ciaccio, Joel P Wagner, Chih-Pin Chuu, Douglas A Lauffenburger, and Richard B Jones. Systems analysis of egf receptor signaling dynamics with microwestern arrays. *Nature methods*, 7(2):148, 2010.

[68] Christopher M Waters, Kerby C Oberg, Graham Carpenter, and Knowles A Overholser. Rate constants for binding, dissociation, and internalization of egf: effect of receptor occupancy and ligand concentration. *Biochemistry*, 29 (14):3563–3569, 1990.

[69] Alexander Sorkin and Jason E Duex. Quantitative analysis of endocytosis and turnover of epidermal growth factor (egf) and egf receptor. *Current protocols in cell biology*, 46(1):15–14, 2010.

[70] Rebecca Worthylake, Lee K Opresko, and H Steven Wiley. Erbb-2 amplification inhibits down-regulation and induces constitutive activation of both erbb-2 and epidermal growth factor receptors. *Journal of Biological Chemistry*, 274(13):8865–8874, 1999.

[71] William Hanley, Owen McCarty, Sameer Jadhav, Yiider Tseng, Denis Wirtz, and Konstantinos Konstantopoulos. Single molecule characterization of p-selectin/ligand binding. *Journal of biological chemistry*, 278(12):10556–10561, 2003.

[72] Porntula Panorchan, Melissa S Thompson, Kelly J Davis, Yiider Tseng, Konstantinos Konstantopoulos, and Denis Wirtz. Single-molecule analysis of cadherin-mediated cell-cell adhesion. *Journal of cell science*, 119(1):66–74, 2006.

[73] Aki Fujioka, Kenta Terai, Reina E Itoh, Kazuhiro Aoki, Takeshi Nakamura, Shinya Kuroda, Eisuke Nishida, and Michiyuki Matsuda. Dynamics of the ras/erk mapk cascade as monitored by fluorescent probes. *Journal of Biological Chemistry*, 281(13):8917–8926, 2006.

[74] WJ Gullick. A new model for the interaction of egf-like ligands with their receptors: the new one-two. *European Journal of cancer*, 30(14):2186, 1994.

[75] Joseph Schlessinger. Ligand-induced, receptor-mediated dimerization and activation of egf receptor. *Cell*, 110(6):669–672, 2002.

[76] Sarel J Fleishman, Joseph Schlessinger, and Nir Ben-Tal. A putative molecular-activation switch in the transmembrane domain of erbb2. *Proceedings of the National Academy of Sciences*, 99(25):15937–15940, 2002.

[77] Rong-Hua Tao and Ichi N Maruyama. All egf (erbb) receptors have preformed homo-and heterodimeric structures in living cells. *Journal of cell science*, 121(19):3207–3217, 2008.

[78] Erik G Hofman, Arjen N Bader, Jarno Voortman, Dave J Van den Heuvel, Sara Sigismund, Arie J Verkleij, Hans C Gerritsen, and Paul MP van Bergen en Henegouwen. Ligand-induced epidermal growth factor receptor (egfr) oligomerization is kinase-dependent and enhances internalization. *Journal of Biological Chemistry*, pages jbc–M110, 2010.

[79] Stefan Legewie, Hanspeter Herzel, Hans V Westerhoff, and Nils Blüthgen. Recurrent design patterns in the feedback regulation of the mammalian signalling network. *Molecular systems biology*, 4(1):190, 2008.

[80] Don S Lemons and Paul Langevin. *An introduction to stochastic processes in physics*. JHU Press, 2002.

[81] John J Bartko. Approximating the negative binomial. *Technometrics*, 8(2): 345–350, 1966.

[82] Vikas Nanda, Sandeep V Belure, and Ofer M Shir. Searching for the pareto frontier in multi-objective protein design. *Biophysical reviews*, 9(4):339–344, 2017.

[83] Gideon Schreiber and Amy E Keating. Protein binding specificity versus promiscuity. *Current opinion in structural biology*, 21(1):50–61, 2011.

[84] Bruce E Eaton, Larry Gold, and Dominic A Zichi. Let's get specific: the relationship between specificity and affinity. *Chemistry & biology*, 2(10): 633–638, 1995.

[85] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[86] Douglas Reynolds. *Gaussian Mixture Models*, pages 827–832. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7488-4. doi: 10.1007/978-1-4899-7488-4_196.

[87] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[88] Filipe Tostevin and Pieter Rein Ten Wolde. Mutual information between input and output trajectories of biochemical networks. *Physical review letters*, 102(21):218101, 2009.

[89] Endang Purba, Ei-ichiro Saita, and Ichiro Maruyama. Activation of the egf receptor by ligand binding and oncogenic mutations: the rotation model. *Cells*, 6(2):13, 2017.

[90] Meytal Landau and Nir Ben-Tal. Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in erbb receptors. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1785(1):12–31, 2008.

[91] Michael P DiGiovanna, David F Stern, Susan M Edgerton, Steve G Whalen, Dan Moore, Ann D Thor, et al. Relationship of epidermal growth factor receptor expression to erbb-2 signaling activity and prognosis in breast cancer patients. *Journal of Clinical Oncology*, 23(6):1152–1160, 2005.

[92] Sunil Srinivasa. A review on multivariate mutual information. *Univ. of Notre Dame, Notre Dame, Indiana*, 2:1–6, 2005.

[93] William McGill. Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory*, 4(4):93–111, 1954.

[94] Reginald D. Smith. A mutual information approach to calculating nonlinearity. *CoRR*, abs/1512.00750, 2015.

[95] Malte Harder, Christoph Salge, and Daniel Polani. A bivariate measure of redundant information. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 87 1:012130, 2013.

[96] JJ Hopfield. Physics, computation, and why biology looks so different. *Journal of Theoretical Biology*, 171(1):53–60, 1994.

[97] Monilola A Olayioye, Richard M Neve, Heidi A Lane, and Nancy E Hynes. The erbb signaling network: receptor heterodimerization in development and cancer. *The EMBO journal*, 19(13):3159–3167, 2000.

[98] Yuji Teramura, Junya Ichinose, Hiroaki Takagi, Kenji Nishida, Toshio Yanagida, and Yasushi Sako. Single-molecule analysis of epidermal growth factor binding on the surface of living cells. *The EMBO journal*, 25(18): 4215–4222, 2006.

[99] Yi Y Shi, Gerald A Miller, Hong Qian, and Karol Bomsztyk. Free-energy distribution of binary protein–protein binding suggests cross-species interactome differences. *Proceedings of the National Academy of Sciences*, 103(31): 11527–11532, 2006.

[100] Howard C Berg. *Random walks in biology*. Princeton University Press, 1993.

[101] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[102] DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.22 of 2019-03-15. URL http://dlmf.nist.gov/. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[103] Núria Folguera-Blasco, Rubén Pérez-Carrasco, Elisabet Cuyàs, Javier A Menendez, and Tomás Alarcón. A multiscale model of epigenetic heterogeneity reveals the kinetic routes of pathological cell fate reprogramming. *BioRxiv*, page 452433, 2018.

[104] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/.

[105] EW Lund. Guldberg and waage and the law of mass action. *Journal of Chemical Education*, 42(10):548, 1965.